Expression, Purification and Characterization of CAEV−Tat

Sabine Schwarz, Silke Hoffmann and Paul Rösch

Lehrstuhl für Struktur und Chemie der Biopolymere

Universität Bayreuth, 95440 Bayreuth, Germany

Phone +49−921−553540 Fax +49−921−553544 E−mail sabine.schwarz@uni−bayreuth.de

According to their different mechanisms lentiviral transcription transactivating proteins can be divided into two subgroups: Tat proteins from human, simian and bovine immunodeficiency viruses (HIV−Tat, SIV-Tat, BIV−Tat) − along with the Tat protein from Equine Infectious Anemia Virus (EIAV−Tat) − function in an RNA sequence dependent manner, while Tat proteins of the so−called 'small ruminant lentiviruses' (SRLV) − Caprine Arthritis Encephalitis Virus (CAEV) and Ovine Maedi−Visna−Virus (OMVV) − are supposed to interact with the cellular transcription factors Jun and Fos to activate viral transcription (1,2).

Various solution structures of Tat proteins belonging to the first group have already been determined by NMR spectroscopy (3). However, SRLV Tat proteins have not been structurally characterized so far. To provide the amounts of protein necessary for NMR experiments and to enable isotope labelling for heteronuclear NMR experiments a bacterial expression system for CAEV−Tat, an 87 amino acid SRLV−Tat protein, was established.

13 of the 15 arginine codons of the CAEV *tat* gene are AGG or AGA, the codons used least frequently by *Escherichia coli.* To prevent inefficient translation due to the presence of these rare codons, CAEV *tat* was coexpressed with *argU*, the gene encoding the tRNA that recognizes these two arginine codons. In addition, several rare isoleucine, leucine, glycine and proline codons were changed to the corresponding codons commonly used by *E. coli* by site−directed mutagenesis. The modified CAEV *tat* gene was expressed in *E. coli* with high yields.

An easy and efficient purification protocol which involves a cation exchange chromatography step at pH 8.0 was established for CAEV Tat. The identity of the purified protein was confirmed by N−terminal amino acid sequencing.

Preliminary far−UV CD and one−dimensional ^{1}H NMR spectra revealed CAEV−Tat to be a structured, predominantly α−helical protein in accordance with secondary structure prediction.

INTRODUCTION

Caprine Arthritis Encephalitis Virus (CAEV) and the closely related ovine Maedi Visna Virus (OMVV) are both members of the Lentivirus subfamily of retroviruses. Together they are referred to as 'small ruminant Lentiviruses' (SRLV). Both cause chronic progressive inflammatory diseases like rheumatoid arthritis, leukoencephalomyelitis, pneumonitis and mastitis in infected goats and sheep after an incubation period of several years (4).

Like all Lentiviruses CAEV and OMVV encode a 10 kDa transactivator protein Tat, which is necessary for activating viral transcription. 87 amino acid CAEV− and 94 amino acid OMVV−Tat are homologous proteins which share over 52 % sequence identity (Figure 1A). There is no significant sequence similarity to other lentiviral Tat proteins.

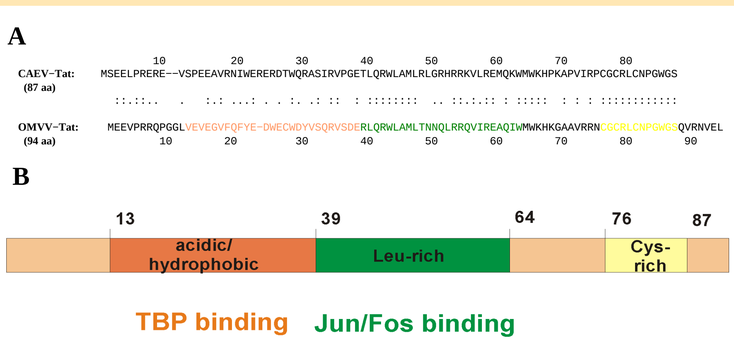


Figure 1A: Alignment of SRLV−Tat proteins **(':'** indicates identical amino acids, **'.'** indicates similar amino acids)

Figure 1B: Sequence domains of OMVV−Tat

While the Tat proteins serve a common function in all Lentiviruses, there are differences in their mechanisms of transactivation: Tat proteins from human, simian and bovine immunodeficiency viruses (HIV, SIV, BIV respectively) − along with the Tat protein form Equine Infectious Anemia Virus (EIAV) − interact with a Tat activation region (TAR), a stem−loop structure located 3' of the transcription initiation site on the viral mRNA (1). In contrast, SRLV do not have a TAR−like structure and their Tat proteins do not bind RNA. OMVV−Tat was shown to bind to the cellular transcription factors Fos and Jun which target the resulting complex to an AP−1 site 5' of the transcription initiation site in the viral long terminal repeat (LTR). Located at the AP−1 site Tat recruits the TATA box binding protein (TBP) resulting in enhanced transcription initiation. Figure 1B shows the sequence domains of OMVV−Tat relevant for interactions with Fos, Jun and TBP (2).

SRLV Tat proteins have not been structurally characterized so far (3). To provide the high amounts of protein necessary for structural studies with NMR a high−yield bacterial expression system along with a simple and efficient purification protocol for CAEV−Tat was established.

RESULTS

Due to differences in the codon usages of *E. coli* and CAEV, expression levels of CAEV *tat* remained very low in *E. coli* (Figure 2B, lane 2). To achieve an efficient translation of CAEV *tat* several rare isoleucine, leucine, glycine and proline codons were changed to the corresponding codons frequently used by *E.coli.* The modified CAEV *tat* was coexpressed with *argU,* which encodes the tRNA recognizing AGG and AGA codons. These two triplets are the codons used least frequently by *E.coli* but are frequently used in CAEV (13 of CAEV−Tat's 15 arginines are encoded by them). Figure 2A shows the so modified CAEV *tat* gene. Expression levels could be dramatically improved (Figure 2B, lane 4).

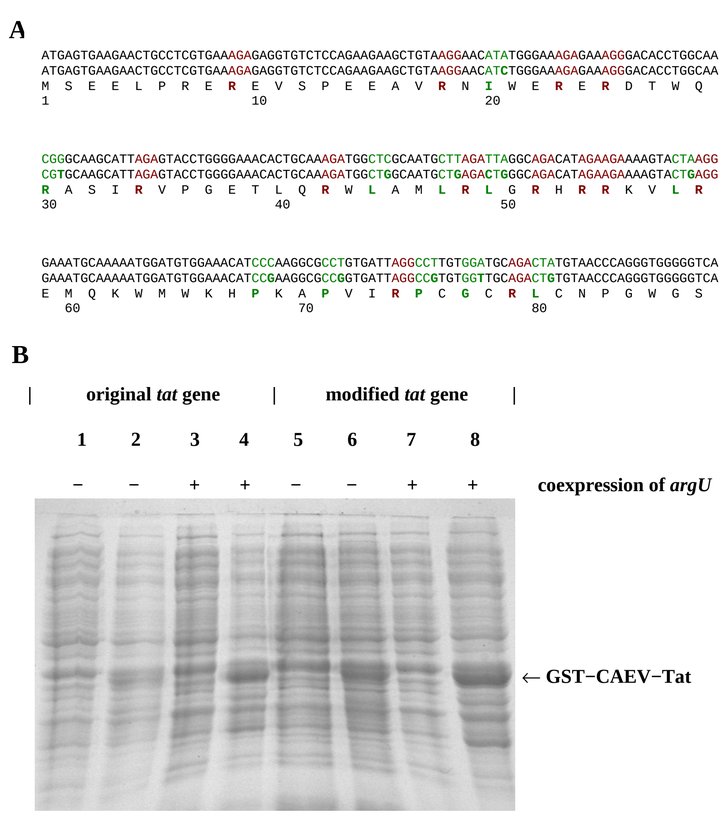


Figure 2: Modifications of CAEV *tat*

A: The original (top) and the modified CAEV *tat* (bottom) nucleic acid sequences are shown. Modified codons are printed in red, AGG/AGA codons are printed in green.

B: Original and modified CAEV *tat* genes were expressed as GST−fusion proteins in *E. coli* BL21 with (+) or without (−) coexpression of *argU* (lanes 1, 3, 5, 7: before induction with 1mM IPTG, lanes 2, 4, 6, 8: four hours after induction with 1 mM IPTG).

Initially CAEV−Tat was expressed as fusion protein with several different affinity tags (for example GST−, His−, Strep− or CBD−tags were used). Purification attempts based on an affinity chromatography step and subsequent cleavage of the fusion protein remained unsuccessful. Finally we expressed CAEV−Tat without any affinity tag at all and developed an efficient one−step purification protocol based on a cation exchange chromatography step at pH 8.0 (Figure 3). CAEV−Tat containing fractions were pooled, desalted by dialysis against ammonium acetate and lyophilized. Cells grown in minimal medium yielded 8 mg CAEV−Tat per liter bacterial culture.

The identity of the purified protein was confirmed by N−terminal amino acid sequencing.

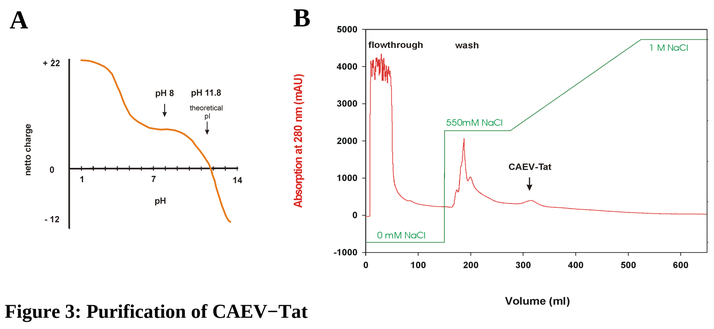
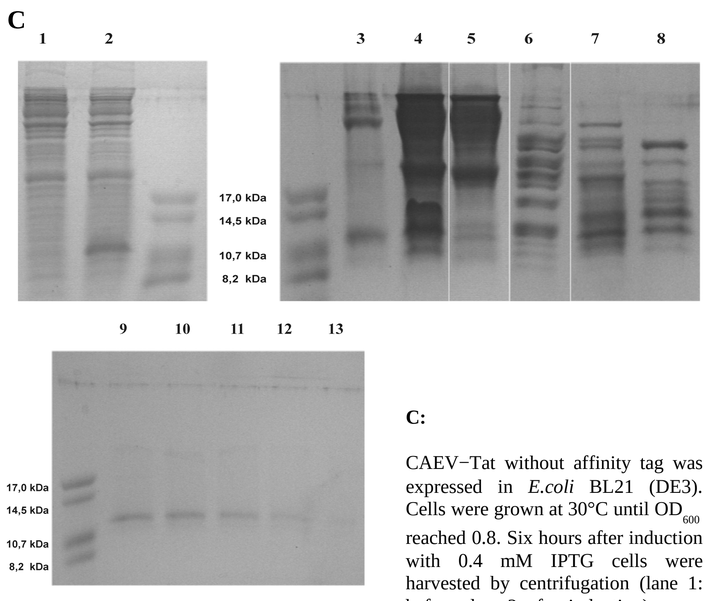


Figure 3: Purification of CAEV−Tat

A: Titration curve of CAEV−Tat calculatedaccording to Bellon (5). Due to its unusual high pI of 11.8 CAEV−Tat is positively charged at pH 8.0.

B: Elution profile of a cation exchange chromatography run at pH 8.0 using a 25 ml CM−Sepharose Fast Flow column. Bound cellular proteins can be eluted with start buffer containing 550 mM NaCl. Pure CAEV−Tat elutes at higher NaCl concentrations.



C:

CAEV−Tat without affinity tag was expressed in *E.coli* BL21 (DE3). Cells were grown at 30°C until OD_{600} reached 0.8. Six hours after induction with 0.4 mM IPTG cells were harvested by centrifugation (lane 1: before, lane 2: after induction).

The cell extract was divided into an insoluble (lane 3) and a soluble fraction (lane 4), which was subsequently used for cation exchange chromatography. Lane 5 shows the column flowthrough, lanes 6 to 8 show different fractions of the 550 mM NaCl washing step. In lanes 9 to 13 some fractions with pure CAEV−Tat are shown, which could be eluted using NaCl concentrations higher than 550 mM.

Purified CAEV−Tat was characterized using far−UV CD and one−dimensional ^{1}H NMR spectroscopy (Figure 4). Spectra analysis revealed CAEV−Tat as a structured, mainly α−helical protein. Secondary structure prediction using the Psipred algorithm (6) proposes three α−helices: One helix comprises the region homologous to OMVV−Tat's TBP binding domain, the two other helices are predicted for the leucine rich region which is conserved among SRLV−Tat proteins and is responsible for interactions with Jun and Fos in OMVV−Tat (Figure 4C).

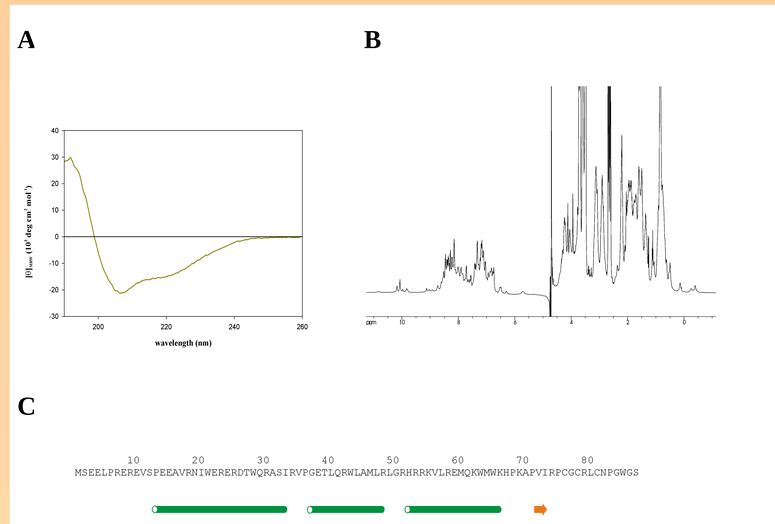


Figure 4: Spectroscopic characterization and secondary structure prediction

A: Far−UV CD spectrum of CAEV−Tat (10 µM in 50 mM potassium phosphate, pH 5.8) recorded on a Jasco J−810 spectropolarimeter. Minima at 208 and 222 nm and an intense signal at 195 nm are characteristic for α−helical proteins.

B: One−dimensional ^{1}H NMR spectrum of CAEV−Tat (100 µM in 100 mM potassium phosphate, pH 5.5, 100 mM NaCl, 5 mM DTT, 10 % (v/v) D_2O) recorded on a Bruker Avance DRX600 spectrometer. Several Methyl proton resonances shifted to high field are indicating a folded protein. Little dispersion of amide proton resonances and αH resonances shifted to high field are typical for α−helical proteins.

C: Secondary structure prediction using Psipred (6). Green barrels indicate predicted α−helices, orange arrows indicate predicted β−strands.

Large quantities of CAEV−Tat are now available for further structural studies with NMR. Figure 5 shows a NOESY spectrum of CAEV−Tat, an important step on the way to a high−resolution structure. The crowded fingerprint region of this spectrum makes it necessary to apply heteronuclear experiments. Isotopically labelled CAEV−Tat necessary for these experiments will easily be produced by growing *E. coli* cells in ^{15}N− or ^{15}N/^{13}C−enriched minimal media.

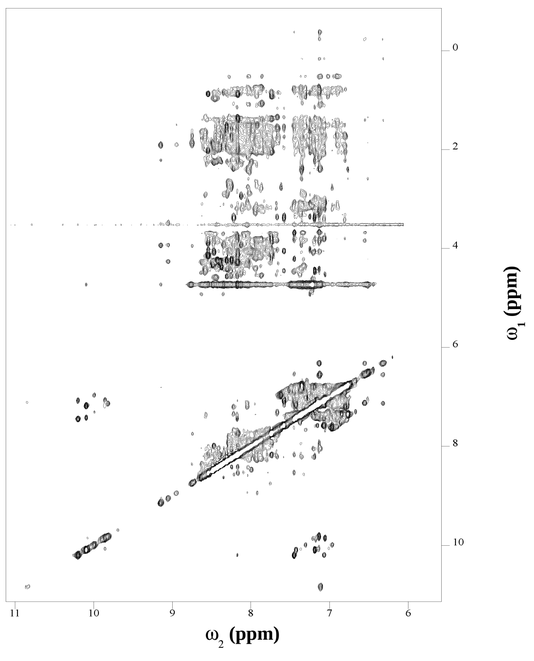


Figure 5: NOESY spectrum of CAEV−Tat

560 µM CAEV−Tat in 100 mM potassium phosphate, pH 5.5, 100 mM NaCl, 5 mM DTT, 10 % (v/v) D_2O recorded on a Bruker DRX600 spectrometer.

ACKNOWLEDGEMENT

This work was supported by Graduiertenkolleg 'Biosynthese der Proteine und Regulation ihrer Aktivität'.

REFERENCES

[1] Taube et al. (1999) Virology *264*, 245−253

[2] Morse et al. (1999) J. Virol. *73*, 37−45

[3] Willbold et al. (1996) Eur. J. Biochem. *240*, 45−52
Boehm (1998) Dissertation Universität Bayreuth

[4] Pépin et al. (1998) Vet. Res. *29*, 341−367

[5] Bellon (1994) Equilibres multiples

[6] Jones (1999) J. Mol. Biol. *292*, 195−202