J. Mol. Biol. (1998) **277**, 749–755

JMB

COMMUNICATION

Equine Infectious Anemia Virus Transactivator is a Homeodomain-type Protein

Dieter Willbold[1]*, Armin U. Metzger[1], Heinrich Sticht[1] Karl Christian Gallert[2], Renate Voit[3], Norbert Dank[2], Peter Bayer[4] Gerhard Krauss[2], Roger S. Goody[4] and Paul Rösch[1]

[1]*Lehrstuhl für Biopolymere Universität Bayreuth D-95440 Bayreuth, Germany*

[2]*Lehrstuhl für Biochemie Universität Bayreuth D-95440 Bayreuth, Germany*

[3]*Deutsches Krebsforschungszentrum Molekularbiologie der Zelle D-69120 Heidelberg, Germany*

[4]*Max-Planck-Institut für Molekulare Physiologie Dortmund, Germany*

**Corresponding author*

Lentiviral transactivator (Tat) proteins are essential for viral replication. Tat proteins of human immunodeficiency virus type 1 and bovine immunodeficiency virus form complexes with their respective RNA targets (Tat responsive element, TAR), and specific binding of the equine anemia virus (EIAV) Tat protein to a target TAR RNA is suggested by mutational analysis of the TAR RNA. Structural data on equine infectious anemia virus Tat protein reveal a helix-loop-helix-turn-helix limit structure very similar to homeobox domains that are known to bind specifically to DNA. Here we report results of gel-shift and footprinting analysis as well as fluorescence and nuclear magnetic resonance spectroscopy experiments that clearly show that EIAV Tat protein binds to DNA specifically at the long terminal repeat Pu.1 (GTTCCTGTTTT) and AP-1 (TGACGCG) sites, and thus suggest a common mechanism for the action of some of the known lentiviral Tat proteins *via* the AP-1 initiator site. Complex formation with DNA induces specific shifts of the proton NMR resonances originating from amino acids in the core and basic domains of the protein.

Keywords: EIAV; Tat; protein-DNA-complex; NMR; homeodomain

Introduction

Lentiviruses, such as the human immunodeficiency virus type-1 and 2 (HIV-1, HIV-2), the equine infectious anemia virus (EIAV), and the visna virus (VV) cause slow, progressive diseases in their hosts. Gene expression of these viruses is controlled by Tat proteins, potent transactivators that generally comprise less than 130 amino acid residues. These proteins are mandatory components of the replication cycle of immunodeficiency viruses, including HIV, and are thus ideal targets for therapeutic interference with the host diseases.

Tat proteins can conventionally be divided into three major classes according to sequence similarities. Bovine immunodeficiency virus (BIV) Tat (b-Tat), HIV-1 and HIV-2 Tat (h1-Tat, h2-Tat), and simian immunodeficiency virus (SIV) Tat (s-Tat) comprise one class of these proteins. Sequences of proteins from this class may be subdivided into an NH_2-terminal, a cysteine-rich, a core, a basic, and a glutamine-rich domain (Jones & Peterlin, 1994; Kingsman & Kingsman, 1996; Dorn *et al.*, 1990; Willbold *et al.*, 1994). Immunodeficiency virus Tat proteins show high homology in the cysteine-rich region, contrasting EIAV protein (e-Tat) which completely lacks the cysteine-rich sequence (NH_2-terminal domain, Leu1 to Asn34; core domain, Tyr35 to Tyr49; basic domain, Arg55 to Arg63). Thus, e-Tat protein comprises the single membered second class of Tat proteins. Sequence homology between all these class one and two Tat proteins is extremely high in the core sequence and moder-

Present addresses: K. C. Gallert, Hoeschst AG, Corporate Research and Technology, H824, D65926 Frankfurt, Germany; P. Bayer, A. U. Metzger, MRC, Cambridge, CB2 2QH, Great Britain.

Abbreviations used: EIAV, equine infectious anemia virus; HIV-1, HIV-2, human immunodeficiency virus type-1 and 2; VV, visna virus; Tat, lentiviral transactivator; TAR, Tat responsive element; BIV, bovine immunodeficiency virus; SIV, simian immunodeficiency virus; TFE, trifluoroethanol; HLHTH, helix-loop-helix-turn-helix; LTR, long terminal repeat; ds, double-stranded.

0022–2836/98/140749–07 $25.00/0/mb981636

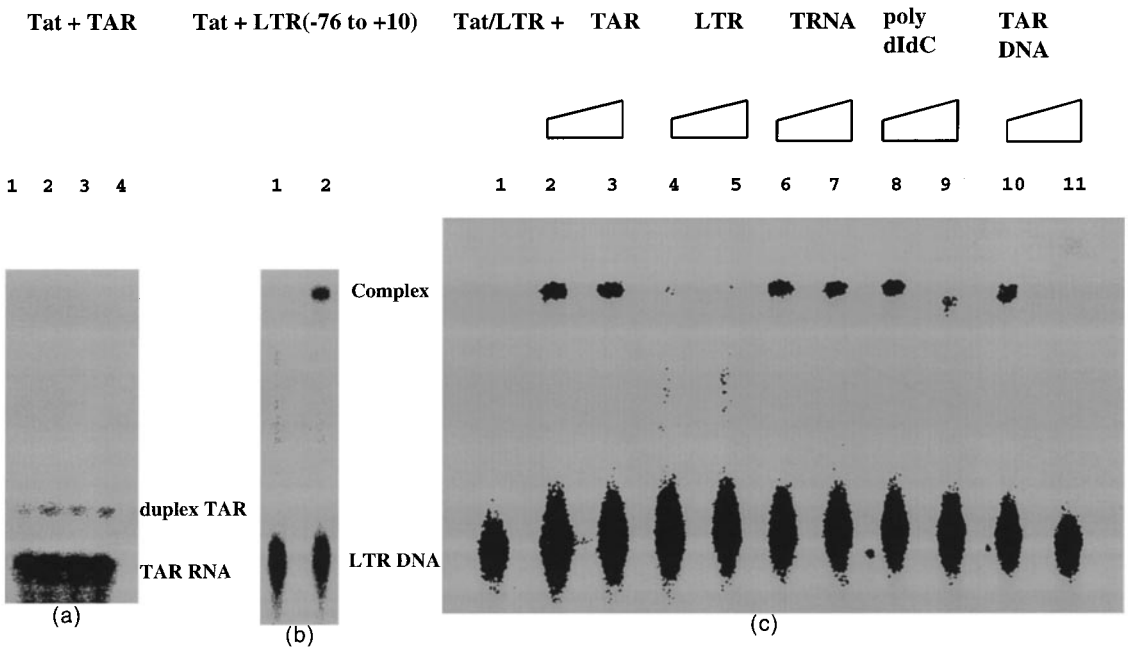

Figure 1. Nucleic acid binding specificity of the EIAV-Tat protein was analysed by gel mobility-shift assays. The 86-mer dsLTR (−76 to +10) oligonucleotide and the EIAV-TAR RNA were labelled with [γ-^{32}P]ATP and T4 polynucleotide kinase. ^{32}P-labelled e-TAR RNA and 86-mer dsLTR (−76 to +10) oligonucleotide, respectively, was incubated with e-Tat in 25 μl reactions containing 100 nM ^{32}P-labelled e-TAR RNA or 86-mer dsLTR, 50 mM Tris, 20 mM KCl, 5 mM DTT, 5 mM $MgCl_2$, and 0.1% (v/v) Triton X-100 adjusted to pH 7.5 with HCl. After ten minutes incubation at room temperature, the free 86-mer dsLTR or EIAV-TAR RNA were separated from the bound nucleic acid by electrophoresis in 7.5% (w/v) non-denaturing polyacrylamide gels containing 44.5 mM Tris base, 44.5 mM boric acid, and 0.1 % Triton X-100 adjusted to pH 8.3 with HCl (Gala & Schmitz, 1978). Electrophoresis was for two hours at 200 V and 15°C. Competition binding assays between dsLTR (−76 to +10) and a set of nucleic acids (e-Tar RNA, tRNA from yeast, poly(dIdC) and EIAV TAR dsLTR) were performed as above with 250 nM e-Tat and competing nucleic acid concentrations in 40 and 300-fold mass (tRNA, poly(dIdC) or molar (e-tar, dsLTR) excess over labelled dsLTR (−76 to +10). Twenty-five nucleotide EIAV TAR RNA (GCACUCAGAUUCUGCGGUCUGAGUC) was synthesized by the standard T7 polymerase methods. DNA oligomers (MWG, Munic, Germany) and poly(dIdC) (Sigma, Germany) were commercial products. e-Tat was prepared in an *Escherichia coli* system according to the standard procedure (Rosin-Arbesfeld *et al.*, 1994). (a) Gel mobility shift assay of 100 nM TAR RNA with 0, 0.5, 1.0 and 2.5 μM e-Tat, respectively (lanes 1, 2, 3, 4). (b) Gel mobility shift assay of 100 nM dsLTR (−76 to +10) free (lane 1), and with 500 nM e-Tat (lane 2). (c) Competition binding assay. In gel retardation assays with a DNA sequence representing the EIAV LTR from positions −76 to +10, it could be seen that binding to LTR (−76 to +10) was tight and stable against competition with 40-fold and 300-fold molar excess TAR RNA (lanes 2 and 3), 40-fold and 300-fold mass excess yeast tRNA (lanes 6 and 7), and poly(dIdC) (lanes 8 and 9), and TAR DNA (lanes 10 and 11). Further, k_D was estimated to be in the range of 100 nM. Lanes 4 and 5 represent a control experiment with a 40 and 300-fold molar excess of LTR (−76 to +10).

ately high in the basic, that is the Arg and Lys-rich, sequence. The latter is the nucleic acid interaction sequence (Dorn *et al.*, 1990). h1-Tat and e-Tat also display highly similar conformational features (Willbold *et al.*, 1994; Bayer *et al.*, 1995). A third, entirely different class of Tat proteins not sharing sequence homologies with the members of the first two classes is formed by the Tat proteins of CAEV (c-Tat) and visna virus (v-Tat) (Pyper *et al.*, 1986). These two Tat proteins share high homology throughout their sequence.

Interaction of h1-Tat and HIV-1 TAR (h1-TAR) is specific, and the complex has a dissociation constant in the nanomolar range (Churcher *et al.*, 1993). h1-TAR consists of a stem-loop structure with a very typical uridine-rich bulge. RNAs with secondary structure elements similar to h1-TAR are found in the 5′ ends of mRNAs from HIV-2, SIV, and BIV. These RNA sequences are believed to represent the respective TAR elements. Recent structural investigations of peptides comprising the basic sequence regions of h1-Tat (Aboul-ela *et al.*, 1995; Metzger *et al.*, 1996) and b-Tat (Chen & Frankel, 1995; Puglisi *et al.*, 1995; Ye *et al.*, 1995) in complex with their target TAR elements show that the peptide adopts an extended β-type structure, and that h1-Tat and b-Tat bind to the bulge and stem regions of their target TAR RNA. A functional h1-Tat, however, can be composed with the basic sequence substituted by the α-helical recognition sequence of HIV-1 Rev (regulator of virion activity) protein, although the flexibility of the nucleic acid region seems to be of major importance (Tan *et al.*, 1993).

A stem-loop-structure at the 5′ end of the EIAV-mRNAs that, however, lack the characteristic bulge

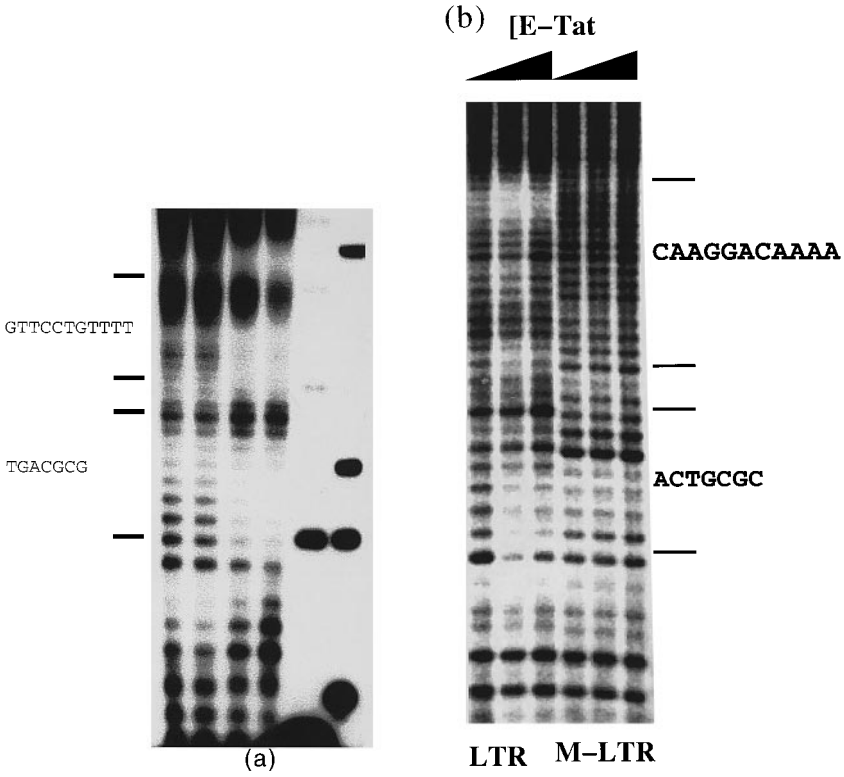

Figure 2. Mapping of e-Tat DNA binding region by polyacrylamide gel nuclease protection assay (DNase I footprinting). (a) Four different reactions (each 25 µl volume) were performed in presence of 0 (lane 1), 10 µM (lane 2), 50 µM (lane 3), and 100 µM (lane 4) e-Tat, respectively. Lanes 5 and 6 show molecular mass standards (lane 5 DNA oligomers in multiples of ten base-pairs; lane 6, 12, 20, 24, 44 base-pair DNA oligomers). The nucleotide sequence of LTR (−76 to +10) (EMBL Data Bank: gb_lv:eiavcg; accession number: M16575; (Chiu *et al.*, 1985) is indicated to the right of the autoradiograph. AP-1 and Pu.1 recognition sites are depicted as bars. A 50 nM double strand LTR (−76 to +10) DNA was incubated for 15 minutes at 37°C in 50 mM Tris-HCl (pH 7.0), 5 mM dithiothreitol, 20 mM KCl and 5 mM $MgCl_2$ at 37°C. After an additional five minutes incubation at 30°C the reaction was started by addition of 0.02 unit DNase I. After one minute the reaction was stopped and the products were separated on a 12% (w/v) denaturing polyacrylamide gel. (b) A control experiment for the DNase I footprint analysis (a) was performed using the 86-mer double-stranded LTR (from −76 to +10, 5′ end-labelled) as a probe. The mutant LTR DNA probe (MLTR), which was used as a control, was identical to the LTR probe except that the wild-type Pu.1 site (5′-GTTCCTGTTTT-3′) was exchanged against 5′-CAAGGACAAAA-3′, and the wild-type AP-1 site (5′-TGACGCG-3′) against 5′-ACTGCGC-3′. In a standard reaction 25 µl of 50 nM double-stranded LTR or MLTR were incubated either with no protein or with 10 to 100 µM in 50 mM Tris-HCl (pH 7.0), 5 mM DTT, 20 mM KCl, 5 mM $MgCl_2$, for 15 minutes at 30°C and five minutes at room temperature, followed by addition of 0.1 unit DNase (Boehringer). Incubation was continued for exactly one minute at room temperature. After phenol extraction and ethanol precipitation, the samples were analysed on denaturing 12% polyacrylamide gels.

is thought to represent the EIAV TAR (e-TAR) element (Carvalho & Derse, 1991; Hoffmann & White, 1995). Mutational studies show that this RNA sequence modulates e-Tat-mediated transactivation (Carvalho & Derse, 1991), but no direct evidence of e-Tat binding to this element is available. It is thus suggested that e-Tat interaction with e-TAR requires additional cellular factors (Carvalho & Derse, 1991).

The third class of Tat proteins, c-Tat and v-Tat, are not known to bind nucleic acids without host cell factors. Instead, they are supposed to act on DNA sequence elements upstream of the transcription start *via* DNA-binding proteins. Requirement of c-Tat for viral replication is still under discussion (Harmache *et al.*, 1995).

Several models for Tat action are discussed: h1-Tat and h2-Tat act directly on the transcription complex (Keen *et al.*, 1996; Carroll *et al.*, 1991), and presence of these Tat proteins in proximity of the transcription complex may be achieved by direct binding to their target TAR sequences. Antitermination activity is suggested for h1-Tat, s-Tat, and b-Tat action (Greenblatt *et al.*, 1993). Indirect interaction of h1-Tat with upstream DNA is proposed (Jeang *et al.*, 1993). c-Tat is thought to act on AP-1 elements (Kalinski *et al.*, 1994), and v-Tat binds to upstream AP-1 and AP-4 DNA sequence elements, possibly by recruiting cellular bZip type factors such as jun/fos type proteins (Carruth *et al.*, 1996).

Three-dimensional structural models determined on the basis of nuclear magnetic resonance (NMR) data of the 75 amino acid residue e-Tat (Willbold *et al.*, 1994) and the 86 amino acid residue h1-Tat (Bayer *et al.*, 1995) show common features such as the fold of the highly conserved core sequence region. No stable elements of regular secondary structure were found in either protein (Willbold *et al.*, 1994; Bayer *et al.*, 1995). Three different sequence regions of e-Tat, however, show a tendency towards helical structure (Willbold *et al.*, 1994), in obvious contrast to h1-Tat, where helices were never observed. The helical elements of e-Tat are stabilized in trifluoroethanol (TFE) containing solution, where particularly the basic sequence region forms a regular α-helix (Sticht *et al.*, 1994).

Under these circumstances, the protein clearly shows a helix-loop-helix-turn-helix (HLHTH) secondary structure motif, with the third helix formed by the basic domain (Sticht *et al.*, 1994). The HLHTH motif is very well known from studies of homeotic proteins (Gehring *et al.*, 1994), and similarities between the e-Tat protein and other nucleic acid binding domains such as homeobox protein domains have been pointed out earlier (Rösch & Willbold, 1996).

In order to further investigate the possibility that e-Tat belongs to the structure family of homeotic DNA-binding proteins we studied the interaction

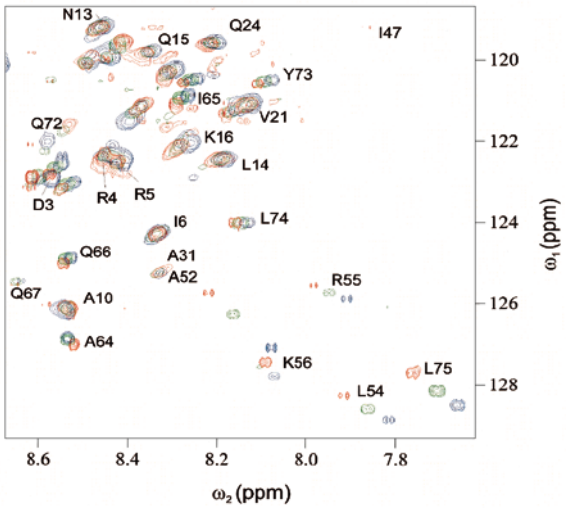

Figure 3. Superposition of heteronuclear single quantum coherence spectra of e-Tat (red), Pu.1/e-Tat (green), and AP-1/e-Tat (blue). Nuclear magnetic resonance spectra were recorded at 291 K with ^{15}N-labelled e-Tat that was prepared from an *E. coli* expression system grown on minimal media with $^{15}HN_4Cl$ as the sole nitrogen source. All samples were in 20 mM potassium phosphate buffer (pH 6.4), 100 mM NaCl. The concentrations were 1.0 mM e-Tat (e-Tat alone, red coloured spectrum), 1.0 mM e-Tat with 1.0 mM Pu.1 DNA double strand obtained from oligonucleotides TAAGTTCCTGTTTTT and AAAAACAGGAACTTA (e-Tat with Pu.1 DNA, green coloured spectrum) and 1.2 mM e-Tat with 1.2 mM AP-1 DNA double strand (obtained from oligomers TTTGTGACGCGTTA and TAACGCGTCACAAA). Spectra were obtained with a two-dimensional ^{1}H-^{15}N correlation *via* double INEPT transfer recorded in a phase-sensitive mode employing time proportional phase incrementation. Water solvent signal was suppressed using a spin lock. Fluorescence spectroscopy measurements were done at 25°C on a SLM Smart 8000 spectrofluorimeter (Colora, Lorch, Germany) equipped with a PH-PC 9635 photomultiplier. Reactions were performed in a buffer containing 100 mM NaCl and 50 mM Tris-HCl (pH 7.0) in a total volume of 1 ml with 100 nM e-Tat (TAR RNA titration) and 1 μM e-Tat (LTR (−76 to +10) titration). For tyrosine fluorescence, the samples were excited at 274 nm, and emission was measured at 303 nm (slit widths of 1 and 6 nm, respectively). The data were evaluated with the program Grafit (Erithacus software) after correction for dilution and the inner filter effect with a quadratic equation describing the binding equilibrium.

of e-Tat with the supposed TAR recognition sequence and various DNA sequences by biochemical and spectroscopic methods.

In a first set of experiments, gel retardation and competition assays with e-TAR and various DNA fragments were performed. These experiments showed very low affinity of e-Tat to e-TAR (Figure 1(a)), the dissociation constant (k_D) estimated to be in the millimolar range. Thus, e-Tat does not bind directly to the proposed TAR element. e-Tat and the corresponding TAR DNA form a weak complex that shows a k_D value of the order of several μM, as estimated roughly from these experiments, and TAR DNA in the complex with e-Tat could be competed with excess poly(dIdC) (data not shown). Thus, although these experiments showed that e-Tat binds to TAR DNA with much higher affinity than it does to e-TAR, it became clear that TAR DNA was not the natural target site of e-Tat as complexes between specific DNA binding proteins, such as homeodomains, and their target sequences, are much tighter, with dissociation constants typically in the nanomolar range (Gehring *et al.*, 1994).

It is well documented that e-Tat does have transcriptional activity, and that DNA sequence elements are important for this activity (Maury *et al.*, 1994; Dorn & Derse, 1988; Sherman *et al.*, 1989). A potential DNA target sequence would be upstream, but proximal to transcription start at position +1. The upstream region of the EIAV LTR contains several target sites for DNA-binding proteins, such as AP-1, Pu.1, Sp-1, and NFκB (Carvalho & Derse, 1993). It is known that these sites are recognized by proteins with helical DNA binding domains, such as GCN4, jun/fos, and others (Sitlani & Crothers, 1996). Gel retardation assays clearly showed that e-Tat binds to EIAV LTR (−76 to +10) (Figure 1(b)). The LTR (−76 to +10) DNA could not be competed with even a 300-fold excess of e-TAR RNA (Figure 1(c)). The same was true for tRNA and, to a lesser extent, poly(dIdC) and TAR DNA. The dissociation constant for the complex could be estimated to be approximately 100 nM from these experiments, a value that was corroborated by fluorescence titration assays with e-Tat and EIAV LTR (−76 to +10).

DNase I protection assays were performed in order to find out whether the EIAV LTR (−76 to +10) indeed contained a specific e-Tat binding site in the upstream region and, if so, to pin down the exact binding site of e-Tat within LTR. These assays were done with LTR (−76 to +10), a sequence containing the AP-1 and Pu.1 target sites (Figure 2). Increasing protection against nuclease hydrolysis of the phosphodiester bonds could be observed at the AP-1 and Pu.1 sites in the presence of increasing concentrations of e-Tat. This clearly showed that e-Tat does not only bind DNA in the LTR region, but, in addition, that this binding is very specific to the AP-1 and Pu.1 sites.

Two and three-dimensional NMR spectra of the e-Tat/AP-1 and the e-Tat/Pu.1 complexes are of excellent quality and allow tentative assignment of several resonances in the heteronuclear single quantum coherence spectra (Figure 3). A number of resonances, most of them tentatively assigned to protons of amino acid residues in the core and basic domains of e-Tat, are shifted upon DNA complex formation. Moreover, complexation by Pu.1 DNA and AP-1 DNA causes the same resonances to shift, and the shift direction is identical for these resonances, an observation which might hint to similar modes of interaction with both sites. The shifts are much more pronounced, however, in the AP-1 complex, possibly due to tighter binding. Our earlier suggestion of e-Tat being a homeodomain type protein (Rösch & Willbold, 1996) would be in accord with these studies.

The results presented here may shed new light on several data from biochemical and molecular biology experiments. Deletion of LTR (−76 to −41) dramatically reduces response of gene expression to e-Tat, although the LTR still contains the TATA box (Dorn & Derse, 1988). Responsiveness to e-Tat depends on LTR (−50 to −17), but additional upstream sequences are needed to restore e-Tat response. The downstream TAR sequence element is necessary for e-Tat action (Sherman *et al.*, 1989). EIAV LTR and HIV-1 LTR are responsive to their respective Tat protein, and HIV-1 LTR still is h1-Tat responsive when the U3 region is replaced by the EIAV U3 region. In contrast, EIAV LTR with the U3 region of HIV-1 shows high basal activity but lacks e-Tat response ability (Maury *et al.*, 1994), in accord with a model proposed for Tat protein action. HIV-1 LTR with the EIAV U3 region still contains the h1-Tat binding site I in the R region and the terminator site in R. EIAV LTR with the HIV-1 U3 region shows, however, high basal activity due to the strong enhancer elements in HIV-1 LTR. As HIV-1 LTR does not contain a Pu.1 site, and as the HIV-1 LTR AP-1 site is much

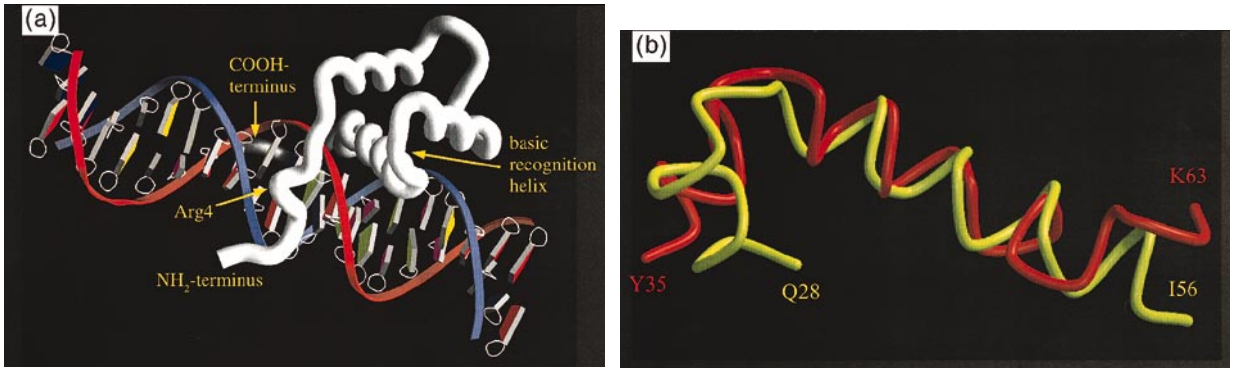

Figure 4. (a) Model of an e-Tat DNA complex based on the structure of the engrailed homeodomain (Kissinger *et al.*, 1990). The protein is shown in a tube presentation and NH_2 and COOH terminus are indicated. (b) Overlay of the helix-turn-helix motif of e-Tat in 40% TFE containing solution (red) and the engrailed homeodomain (yellow). This Figure was prepared using the programs GRASP (Nicholls, 1993), MOLSCRIPT (Kraulis, 1991) and Raster3D (Meritt & Murphy, 1994). Restraints for the modelling of the complex were deduced from sequence alignment (based on functional similarities) of e-Tat with the engrailed homeodomain, for which the crystal structure of the DNA complex is known at a resolution of 2.8 Å (Kissinger *et al.*, 1990; Protein Data Bank code 1HDD). According to the degree of homology, restraints were deduced for C^{α} atoms only or for all heavy atoms that were identical between corresponding amino acid residues. The orientation of the NH_2-terminal amino acid residues R4, R5 and T9 was fixed in e-Tat, because these residues correspond to R3, R5 and T6 of the engrailed homeodomain which form key interactions to the minor groove of the DNA in the complex. Restraints for the relative position at the C^{α} atoms of amino acid residues A10 to Q24 and R32 to A64 were extracted from the crystal structure of the complex. These residues correspond to A7 to N21 and Y25 to K57 of the engrailed homeodomain, comprising the NH_2-terminal helix and the HTH motif of this protein. Additional restraints were used to fix the side-chain orientation of Q38, F41, L42, L45, G46, A52, L54, R55, K57-Q60, L62, and A64. This selection includes those residues which form DNA contacts in the engrailed homeodomain DNA complex or which correspond to the pattern of conserved hydrophobic residues characteristic for homeodomains. The calculation of the e-Tat structure in the complex is similar to one described previously (Sticht *et al.*, 1997). All calculations were carried out using a modified version of the *ab initio* simulated annealing protocol that is supplied with the X-PLOR program package (Brünger, 1993). The calculation started from templates with randomized backbone torsion angles. Restraints deduced on the basis of homology to the engrailed homeodomain were included in the calculation using harmonic restraints with force constants of 10 kcal mol^{-1} $Å^{-2}$ and 5 kcal mol^{-1} $Å^{-2}$ for the backbone and side-chain heavy atoms, respectively. The protocol included 50 ps of molecular dynamics at 1000 K and a slow cooling phase of 100 ps during which the system was cooled from 1000 K to 100 K. Reduced van der Waals radii and a less stringent weight on geometry were used for the first half of the high-temperature (conformational search) phase. Electrostatic and hydrogen bonding interactions were not considered explicitly, and the non-bonding interactions were represented by a simple repulsive energy term. A total of 20 structures was calculated and the structure showing the lowest energy was used for docking to a *B*-form DNA (containing AP-1 site). The resulting structure was subjected to a restrained energy minimization yielding the final model of the complex.

further upstream at −330, the presumed e-Tat binding site close to the transcription start is lost, and thus the e-Tat response is defunct. Striking differences in essential residues of the core sequences between e-Tat and h1-Tat were noticed very early (Dorn *et al*., 1990), and exchange of only part of the highly conserved core sequence regions between h1-Tat (LCFITKALGISY) and e-Tat (LCFLRSLGIDY) leads to almost complete loss of transactivation activity of LTR-controlled gene expression (Carroll *et al*., 1991). e-Tat Q60R, LDASL50A mutation and insertion of a PW sequence in front of L50 lead to a loss of activity of more than 90% (Derse & Newbold, 1993). Interestingly, these variations destroy highly conserved spacing of the hydrophobic residues in the proposed helix-turn-helix motif (Gehring *et al*., 1994; Rösch & Willbold, 1996).

The observation of a direct e-Tat–DNA interaction allows modelling of the corresponding complex that is based on the crystal structure of the engrailed homeodomain-DNA complex (Kissinger *et al*., 1990). This modelling was intended to check whether the steric requirements of a HTH motif are met in e-Tat, and whether the resulting model has a structure similar to that reported previously for e-Tat in 40% TFE containing solution (Sticht *et al*., 1994).

The quality of the modelled protein structure (Figure 4(a)) was assessed by PROCHECK (Laskowski *et al*., 1993) analysis or by energetic considerations. The van der Waals energy of the model (−291 kcal mol^{-1}) calculated using the full CHARMM potential function (Brooks *et al*., 1983) is only slightly higher than that measured for the engrailed homeodomain (−340 kcal mol^{-1}). No gross steric overlaps were detected for the side-chains of the residues of the putative HTH motif of e-Tat, confirming that the spatial requirements of the e-Tat side-chains in that region are in agreement with the presence of a HTH motif in this sequence region. Figure 4(b) shows length and arrangement of the helices to be similar to that reported for the HTH motif in 40% TFE containing solution (Sticht *et al*., 1994), underlining the intrinsic ability of e-Tat to adopt this structural motif. Furthermore, the model shows favourable van der Waals and electrostatic interactions after the restrained energy minimization and no steric overlap with *B*-form DNA.

c-Tat can bind to an upstream DNA sequence *via* host factors (Kalinski *et al*., 1994). In EIAV, we could now find indications that an upstream DNA sequence may function as a Tat target site. In these cases, the common theme may be binding of Tat proteins to an upstream DNA target. Detailed knowledge of the three-dimensional structure of the e-Tat-DNA complex, however, determined either by NMR or X-ray crystallography, is the necessary prerequisite to finally decide how close the analogy of e-Tat and homeodomain-type proteins is on a conformational level.

Acknowledgements

The European Union, the German-Israeli Foundation, the Deutsche Forschungsgemeinschaft, and the Fonds der Chemischen Industrie provided financial support. We are grateful to Drs A. Gazit, K.-T. Jeang, M. B. Peterlin, R. Rosin-Arbesfeld and A. Yaniv for helpful discussions.

References

Aboul-ela, F., Karn, J. & Varani, G. (1995). The structure of human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* **253**, 313–332.

Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R. & Rösch, P. (1995). Structural studies of HIV-1 Tat protein. *J. Mol. Biol.* **247**, 529–535.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Brünger, A. T. (1993). *X-PLOR Version 3.1,* Howard Hughes Medical Institute and Yale University, New Haven, CT..

Carroll, R., Martarano, L. & Derse, D. (1991). Identification of lentivirus Tat functional domains through generation of equine infectious anemia virus-human immunodeficiency virus type-1 *tat* gene chimeras. *J. Virol.* **65**, 3460–3457.

Carruth, L. M., Norse, B. A. & Clements, J. E. (1996). The leucine domain of the visna virus Tat protein mediates targeting to an AP-1 site in the viral long terminal repeat. *J. Virol.* **70**, 4338–4344.

Carvalho, M. & Derse, D. (1991). Mutational analysis of the equine infectious anemia virus Tat-responsive element. *J. Virol.* **65**, 3468–3474.

Carvalho, M. & Derse, D. (1993). Physical and functional characterization of transcriptional control elements in the equine infectious anemia virus promoter. *J. Virol.* **67**, 2064–2074.

Chen, L. & Frankel, A. D. (1995). A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. *Proc. Natl Acad. Sci. USA,* **92**, 5077–5081.

Chiu, I. M., Yaniv, A. & Dahlberg, J. E. (1985). Nucleotide sequence evidence for relationship of aids virus and lentiviruses. *Nature,* **317**, 366–368.

Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, J. G., Gait, M. J. & Karn, J. (1993). High affinity binding of TAR RNA by the human immunodeficiency virus type-1 Tat protein requires base-pairs in the RNA stem and amino acids flanking the basic region. *J. Mol. Biol.* **230**, 90–110.

Derse, D. & Newbold, S. H. (1993). Mutagenesis of EIAV Tat reveals structural features essential for transcriptional activation and TAR element recognition. *Virology,* **194**, 530–536.

Dorn, P. L. & Derse, D. (1988). *Cis*-acting and *trans*-acting regulation of gene expression of equine infectious anemia virus. *J. Virol.* **62**, 3522–3526.

Dorn, P. L., DaSilva, L., Martarano, L. & Derse, D. (1990). Equine infectious anemia virus Tat: insights into the structure, function, and evolution of lentiviral transactivator proteins. *J. Virol.* **64**, 1616–1624.

Galas, A. & Schmitz, D. J. (1978). DNase footprinting: a simple method for detection of protein-DNA binding specificity. *Nucl. Acids Res.* **5**, 3157–3170.

Gehring, W. J., Affolter, M. & Bürglin, T. (1994). Homeodomain proteins. *Annu. Rev. Biochem.* **63**, 487–526.

Greenblatt, J., Nodwell, J. R. & Mason, S. W. (1993). Transcriptional antitermination. *Nature,* **364**, 401–406.

Harmache, A., Vitu, C., Russo, P., Bouyac, M., Hieblot, C., Peveri, P., Vigne, R. & Suzan, M. (1995). The caprine arthritis encephalitis virus tat gene is dispensable for efficient viral replication *in vitro* and *in vivo*. *J. Virol.* **69**, 5445–5454.

Hoffman, D. W. & White, S. W. (1995). NMR analysis of the *trans*-activation response (TAR) RNA element of equine infectious anemia virus. *Nucl. Acids Res.* **23**, 4058–4065.

Jeang, K. T., Chun, R., Lin, N. H., Gatignol, A., Glabe, C. G. & Fan, H. (1993). *In vitro* and *in vivo* binding of human immunodeficiency virus type-1 Tat protein and SP1 transcription factor. *J. Virol.* **67**, 6224–62333.

Jones, K. A. & Peterlin, B. M. (1994). Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**, 717–743.

Kalinski, H., Mashiah, P., Rotem, D., Orzech, Y., Sherman, L., Miki, T., Yaniv, A., Gazit, A. & Tronick, S. R. (1994). Characterization of cDNAs species encoding the Tat protein of caprine arthritis encephalitis virus. *Virology,* **204**, 828–834.

Keen, N. J., Gait, M. J. & Karn, J. (1996). Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. *Proc. Natl Acad. Sci. USA,* **93**, 2505–2510.

Kingsman, S. M. & Kingsman, A. J. (1996). The regulation of human immunodeficiency virus type-1 gene expression. *Eur. J. Biochem.* **240**, 491–507.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell,* **63**, 579–590.

Kraulis, P. K. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Maury, W. J., Carpenter, S., Graves, K. & Chesebro, B. (1994). Cellular and viral specificity of equine infectious anemia virus transactivation. *Virology,* **200**, 632–642.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0–a program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869–873.

Metzger, A. U., Schindler, T., Willbold, D., Kraft, M., Steegborn, C., Volkmann, A., Frank, R. W. & Rösch, P. (1996). Structural rearrangements on HIV-1 Tat (32-72) TAR complex formation. *FEBS Letters,* **384**, 255–259.

Nicholls, A. (1993). GRASP: graphical representation and analysis of surface properties. Columbia University, New York.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science,* **270**, 1200–1203.

Pyper, J. M., Clements, J. E., Gonda, M. A. & Narayan, O. (1986). Sequence homology between cloned caprine arthritis encephalitis virus and visna virus two neurotropic lentiviruses. *J. Virol.* **58**, 665–670.

Rösch, P. & Willbold, D. (1996). Is EIAV Tat protein a homeodomain? *Science,* **272**, 1672.

Rosin-Arbesfeld, R., Mashiah, P., Willbold, D., Rösch, P., Tronick, S. R., Yaniv, A. & Gazit, A. (1994). Biological activity and intracellular location of the Tat protein of equine infectious anemia virus. *Gene,* **150**, 307–311.

Sherman, L., Yaniv, A., Lichtman-Pleban, H., Tronick, S. R. & Gazit, A. (1989). Analysis of regulatory elements of the equine infectious anemia virus and caprine arthritis–encephalitis virus long terminal repeats. *J. Virol.* **63**, 4925–4931.

Sitlani, A. & Crothers, D. M. (1996). Fos and Jun do not bend the AP-1 recognition site. *Proc Natl Acad. Sci. USA,* **93**, 3248–3252.

Sticht, H., Willbold, D., Ejchart, A., Rosin-Arbesfeld, R. & Rösch, P. (1994). Trifluoroethanol stabilizes a helix-turn-helix motif in equine infectious anemia virus transactivator protein. *Eur. J. Biochem.* **225**, 855–861.

Sticht, H., Gallert, K.-C., Krauss, G. & Rösch, P. (1997). Homology modeling of adenylosuccinate synthetase from *Saccharomyces cerevisiae* reveals a possible binding region for single stranded DNA. *J. Biomol. Struct. Dynam.* **14**, 667–675.

Tan, R., Chen, L., Buettner, J. A., Hudson, D. & Frankel, A. D. (1993). RNA recognition by an isolated α-helix. *Cell,* **73**, 1031–1040.

Willbold, D., Rosin-Arbesfeld, R., Sticht, H., Frank, R. & Rösch, P. (1994). Structure of the equine infectious anemia virus Tat protein. *Science,* **264**, 1584–1587.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.* **2**, 827–832.

Edited by J. Karn

(Received 26 November 1997; accepted 14 January 1998)