Structural and Functional Characterization of Human Immunodeficiency Virus tat Protein

STEVEN RUBEN,1 ANN PERKINS,1 RICHARD PURCELL,2 KEITH JOUNG,3 REY SIA,3 ROBERT BURGHOFF,3 WILLIAM A. HASELTINE,3 AND CRAIG A. ROSEN1*

Department of Molecular Oncology, Roche Institute of Molecular Biology,1 and Department of Protein Chemistry, Hoffmann-La Roche Inc., Roche Research Center,2 Nutley, New Jersey 07110, and Department of Pathology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 021153

Received 21 July 1988/Accepted 19 September 1988

Site-directed mutagenesis was used to identify functional domains present within the human immunodeficiency virus (HIV) tat protein. Transient cotransfection experiments showed that derivatives of tat protein with amino acid substitutions either at the amino-terminal end or at cysteine residue 22, 27, or 25 were no longer able to transactivate HIV long terminal repeat-directed gene expression. Incubation of Tat expressed in Escherichia coli with zinc demonstrated that both authentic Tat and cysteine mutation derivatives could form metal-protein complexes. The tat proteins that contained alterations within the cluster of positively charged amino acid residues retained their ability to transactivate gene expression, albeit at markedly reduced levels. Indirect immunofluorescence showed that the authentic tat protein and the amino-terminal and cysteine substitution mutants all localized in the nucleus, with accumulation being most evident in the nucleolus. In contrast, nuclear accumulation was greatly reduced with the basic-substitution mutations. Consistent with this result, a fusion protein that contained amino acids GRKKR, derived from the basic region, fused to the amino-terminal end of β-galactosidase also accumulated within the nucleus. These results demonstrate that the 14-kilodalton tat protein contains at least three distinct functional domains affecting localization and transactivation.

Materials and Methods

Construction of expression vectors. For expression of Tat in eucaryotic cells, the vector pictured in Fig. 2A was used. Plasmid pSVEX contains the simian virus 40 (SV40) early-region promoter, followed by a multiple-restriction-site polylinker (obtained from plasmid SP6, Promega Biotech) and SV40 polyadenylation signals obtained from plasmid pSVGT5-Ra8G (16). For expression of authentic Tat, the Sall-BamHI fragment extending from nucleotides 5332 to 8020 of strain HXBC2 was inserted into the Sall-BglII site present on plasmid pSVEX. For expression of mutant tat proteins, the same cloning strategy was followed.

For procaryotic expression, the XhoII-HindIII tat fragment of strain HXBC2 was cloned into the procaryotic expression vector pDS56/RBS11-1 (2) shown in Fig. 2B.

For expression of β-galactosidase fusion proteins, synthetic oligonucleotides encoding the amino acids pictured in Fig. 2C were synthesized with 5′ HindIII and 3′ BamHI overhangs. Oligonucleotides were ligated to a BamHI-Sall DNA fragment containing the lacZ gene of Escherichia coli, and the resultant DNA fragment was cloned into the HindIII-Sall site of plasmid pSVEX.

Site-directed mutagenesis. To obtain single-stranded DNA corresponding to the coding sequence of tat, a Sall-BamHI fragment encompassing nucleotides 5332 to 8020 of strain HXBC2 was cloned into the replicative form (RF) of phage M13mp9 (New England Biolabs). Infected cells were grown to saturation, and single-stranded DNA was isolated from phage. Oligonucleotides used in mutagenesis were synthesized on an Applied Biosystems DNA synthesizer. Oligonucleotides were deprotected with ammonium hydroxide at 55°C overnight, isolated on 20% polyacrylamide gels containing 7 M urea, eluted overnight in 1 M triethylamine buffer, pH 7.4, and desalted over SEP-PAK C14 columns (Waters Associates). Labeled probes for hybridization...
screening were made by treating oligonucleotides with kinase in the presence of $\gamma^{32}$P-labeled ATP to a specific activity of 100 to 300 cpm/pg. For mutagenesis, the procedure followed was adapted from that of Nesbit and Beilharz (17). Briefly, oligonucleotide-directed mutagenesis was performed by annealing the desired kinase-treated primer and universal M13 sequencing primer (New England Biolabs) to the appropriate single-stranded M13 DNA template, extending the primer strand with DNA polymerase I large fragment, and ligating the synthesized strand with T4 DNA ligase to generate covalently closed, circular, double-stranded RFs with the 1-base-pair (bp) mismatch. To reduce the background resulting from residual unconverted single-stranded DNA, the mixture of covalently closed circular DNA and single-stranded DNA was phenol-chloroform extracted, ethanol precipitated, and incubated with PstI and SacI for 3 h to generate the fragment containing the tat gene with the 1-bp mismatch. This mixture was then added to the Pst-SacI-digested plasmid pTZ 18U (U.S. Biochemicals) and incubated overnight with T4 DNA ligase. The ligation reaction mix was transformed into competent E. coli JM101, and one-third of the transformation mixture was plated on 5-bromo-4-chloro-3-indolyl-β-D-galactoside indicator plates [to check for the presence of inserts in the pTZ18U plasmid], and the remaining two-thirds were grown overnight in 10 ml of LB broth. If a large percentage of white colonies were evident, DNA was extracted from the overnight culture and used for a second transformation. The resultant colonies were screened for the presence of the desired mutation. To screen for the presence of a single-base-pair mismatch, the oligonucleotide used in creating the mutant was treated with kinase and $\gamma^{32}$P]ATP and used to screen a set of possible mutants. Ninety-six-well plates containing 200 µl of LB freezing medium were inoculated with the individual colonies, which were grown overnight and replica plated onto LB agar plates containing ampicillin (100 µg/ml). Plates were incubated at 37°C overnight, and colonies were lifted with Whatman 541 paper.

The hybridization procedure used was adapted from that of Wood et al. (26), with tetramethyl ammonium chloride used in the washings. The use of tetramethyl ammonium chloride eliminates differences in the bond energies of A-T and G-C pairs and allows calculation of the temperature of dissociation ($T_d$) directly from the length of DNA without considering base composition. Four different filters for each set of clones were washed at four different temperatures, above and below the predicted $T_d$. Clones that remained hybridized under stringent temperatures were selected for dideoxy chain termination sequencing (22). Fragments from clones containing the desired insert were then cloned into the expression vector pSVEX.

**Cell transfection and CAT assays.** The functional activity of each tat mutant was assessed by chloramphenicol acetyltransferase (CAT) assays following transfection of HeLa cells. Cells were seeded 24 h prior to transfection at a density of 10⁶ per 100-mm dish. Transfection was achieved with a modified DEAE-dextran–calcium phosphate protocol. Briefly, the desired amount of DNA was mixed with 475 µl of transfection solution (137 mM NaCl, 5 mM KCl, 0.3 mM Na₂HPO₄, 7H₂O, 2.4 mM Tris chloride [pH 7.4] plus 1/100 volume of MgCl₂–CaCl₂ solution [10 mg of MgCl₂ and CaCl₂ per ml] plus 10 µl of DEAE solution (10 mg of DEAE-dextran per ml in water) and incubated for 15 min at room temperature. Medium was removed from plates, and cells were washed once and then overlaid with fresh medium and serum. At 48 h following transfection, the medium was removed, and the cells were washed with phosphate-buffered saline (PBS) and suspended in 150 µl of 250 mM Tris (pH 8.0). Cells were subjected to three freeze-thaw cycles (−70 and 37°C, respectively) and centrifuged to remove cell debris. Then, 35 µl of the cell extract was incubated with 5 µl of chloramphenicol reaction mix (24 mM acetyl coenzyme A, 0.1 µCi of [14C]chloramphenicol) at 37°C. The reaction was quenched with the addition of ethyl acetate, and the organic layer was removed, lyophilized, suspended in 20 µl of ethyl acetate, and spotted onto thin-layer chromatography plates (9). Autoradiograms were produced by overnight exposure of the dried plate.

**Indirect immunofluorescence.** COS-7 cells were plated at a density of 10⁶ cells per 35-mm dish 24 h prior to transfection. Cells were transfected with 1 µg of plasmid DNA by the DEAE-dextran transfection procedure as described previously (5), except that the chloroquine step was omitted. At 48 h posttransfection, the cells were washed with PBS, fixed in 3% paraformaldehyde in PBS (pH 7.8) for 30 min at room temperature, washed for 5 min in PBS plus 10 mM glycine, incubated for 5 min with 1% Triton X-100 in PBS, washed again for 5 min, incubated with 25 mM glycine for 30 min, washed, and finally blocked with 3% goat serum (Vector Laboratories) in Western antibody buffer (1% bovine serum albumin in PBS, pH 6.5, with 0.5 M NaCl and 0.05% Tween-20) for at least 1 to 5 h at room temperature. Additional nontransfected cells were treated with the same procedure up to this point, and the tat antibody prepared against a peptide corresponding to amino acids 67 to 87 was preabsorbed to the cells during incubation. After blocking, the preabsorbed tat antibody was diluted in Western antibody buffer to a concentration of 1:300 and added to the cells. Cells were incubated at 4°C overnight. The following day, cells were washed twice with PBS containing glycine for 10 min each and then incubated in the dark with rhodamine isothiocyanate-conjugated goat anti-rabbit immunoglobulin G secondary antibody diluted 1:100 in Western antibody buffer plus 1% goat serum with gentle shaking at room temperature for 1 h. Cells were washed three times with PBS containing glycine. Cover slips were mounted, and cells were viewed with a light fluorescence microscope. A similar procedure was used for localization of the β-galactosidase fusion proteins, except that preabsorption of antibody was not done.

**Zinc blotting and autoradiography.** The procaryotic tat expression vectors were grown to an OD₆₀₀ of 0.6, and tat expression was induced by addition of 1 mM isopropylthio-galactopyranoside (IPTG). After an additional 4 h of growth, cells were centrifuged and lysed, and proteins were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were electrophoretically transferred to nitrocellulose with a Bio-Rad Transblot apparatus. After transfer to nitrocellulose, the filters were washed in metal-binding buffer and probed for 1 h with $^{65}$ZnCl₂ as described by Shiff et al. (23).

**RESULTS**

To identify functional domains within the tat protein, site-directed mutagenesis was used to introduce single-amino-acid substitutions. Substitutions were introduced within a truncated form of tat protein encoded by the first exon. Previous studies have shown that the first 58 amino acids are sufficient for full activity (24). Mutations introduced within the tat gene are shown in Fig. 1. Initially, three regions were chosen for mutagenesis: the amino-terminal...
domain, the cysteine-rich domain, and amino acids present within the positively charged domain. For mutagenesis, a genomic fragment spanning nucleotides 5332 to 8020 from plasmid HXB2C2 was subcloned into the RF of phage M13. Single-stranded DNA was prepared and mutagenesis was performed as described in the Materials and Methods section. Following identification of clones containing the desired mutation, the DNA encoding the tat region was subcloned into the eucaryotic and procaryotic expression vectors shown in Fig. 2. The functional activity and subcellular localization of each Tat derivative were assessed as described below.

**Subcellular localization of the tat protein.** Prior to analysis of the mutant tat proteins, the subcellular localization of Tat was determined. Previously, Hauber et al. demonstrated that Tat is localized in the nucleus (11). To determine the subcellular localization of Tat, COS cells were transfected with plasmid pSVtat and subjected to indirect immunofluorescence analysis 48 h posttransfection. In this vector system (Fig. 2A), a high level of transient gene expression is achieved through replication of plasmid DNA containing the SV40 origin of replication by SV40 T antigen which is constitutively expressed in the COS cell line. As shown in Fig. 3A and B, the tat protein was localized in the nucleus, in accord with previous findings. Moreover, the majority of tat protein was sequestered within the nucleolus under these conditions. Nucleolar staining was also apparent in the absence of accumulation in the nucleoplasm. Experiments done in parallel with a c-fos expression vector or a nucleus-targeted β-galactosidase fusion protein (see below) gave the expected nuclear localization and nucleolar exclusion, indicating that the nucleolar localization observed with the tat protein was not an artifact of the procedure.

**Several functional domains are present in the tat protein.** The activity of the individual mutant tat derivatives was assessed by using transient cotransfection assays with plasmid pU3R-III, which contains the bacterial cat gene under the control of the HIV LTR (20). Transfections were performed in HeLa cells, and CAT activity was measured at 48 h posttransfection. As expected, cotransfection of pU3R-III with plasmid pSVtat, which carries the first exon of tat, led to stimulation of LTR-directed cat gene expression (Fig. 4). Transactivation was completely abolished with the amino-terminal (plasmid pEPV) and cysteine substitutions mutations (C22, C25, C27, and C37). Similar results were obtained following cotransfection of human T4 lymphoid Jurkat cells and CHO cells (not shown). Loss of transactivation with these mutant proteins suggests a structural defect in a functional domain, since expression and proper subcellular localization of the mutant proteins were maintained (see below).

In contrast with the above findings, cotransfection with plasmids containing substitutions or deletions of the basic residues (50, 55, and 50–51) gave two phenotypes. At saturating amounts of expression vector DNA, activity was 10 to 20% of that observed with wild-type DNA (Fig. 4). However, at nonsaturating amounts of DNA, the activity of the proteins expressed from plasmids pR55 and pKK5051 was markedly reduced. For example, in the range from 1 to 0.1 μg, plasmid pKK5051, which contains the Lys-Lys to Ser-Gly substitution, retained only 1/100 the activity of an equivalent amount of plasmid pSVtat (data not shown).

**Subcellular localization of mutant tat proteins.** To identify the subcellular localization of the individual mutant proteins, plasmids were transfected into COS cells and the subcellular localization of Tat was examined by indirect immunofluorescence at 48 h posttransfection. Both the amino-terminal substitution mutant, expressed from plasmid pEPV, and the cysteine substitution mutants were localized within the
FIG. 3. Indirect immunofluorescence staining of COS-7 cells expressing authentic tat protein. COS cells were transfected with 1 μg of plasmids pSVtat (A), pC22 (B), and pC27 (C) and stained 48 h posttransfection as described in the text. Phase-contrast (left) and fluorescence (right) photomicrographs are shown. Arrows mark the perimeter of the cell.
nucleus, with accumulation most evident in the nucleolus. Representative photomicrographs are shown in Fig. 3. The proteins that contained substitutions at the basic residues, although expressed, as indicated by their ability to transactivate, could not be localized accurately by indirect immunofluorescence, perhaps because the level of accumulation in the nucleus was below the level of detection with our antisera.

Identification of nuclear localization signals. Localization of Tat within the nucleus raised the possibility that this protein contained a distinct nuclear localization sequence. The diminished ability of the mutant proteins with substitution of the basic residues to transactivate gene expression and the lack of detectable protein in the nucleus suggested the presence of a potential nuclear localization signal within this region.

Regions important for nuclear localization have been identified for several viral and cellular proteins (3, 12). Although no definitive structure has been identified, a number of continuous positively charged residues are often present within the localization sequence (3, 12). Therefore, the ability of the basic residues of Tat to function as a nuclear localization or accumulation signal was examined. For these studies, we tested the ability of the basic residues to direct the heterologous 116,000-kDa β-galactosidase protein, encoded by the lacZ gene of E. coli, to the nucleus. A similar approach has been used to identify other nuclear targeting signals (12). β-Galactosidase fusion proteins with the general structure shown in Fig. 2C were constructed in the SV40-driven expression vector. Subcellular localization was determined 48 h posttransfection of COS cells by indirect immunofluorescence with anti-β-galactosidase antibody. As shown in Fig. 5, the fusion protein expressed from plasmid pt1-βGal, which contains residues GRKKR of Tat, was observed predominantly in the nucleus. In contrast, the fusion protein expressed from plasmid pt2-βGal, which contains the second stretch of positively charged amino acids including the sequence QRRP, was for the most part localized in the cytoplasm. Similar results were obtained with a staining procedure that measured functional β-galactosidase activity (not shown).

Monomeric form of Tat can form metal-protein complexes. The presence of a cysteine-rich region within the tat protein is reminiscent of putative zinc- or metal-binding fingers present in some transcriptional regulatory proteins (10, 14, 21). The cysteines present in Tat have the potential to form different structural configurations with metal in either the monomeric or dimeric form. Recently, Frankel et al. have shown that Tat can form a stable dimer configuration in the presence of divalent cations (9).

To examine the ability of the monomeric form of Tat and the cysteine substitution mutants to bind metal, the DNA fragments encoding these mutations were excised from the eucaryotic expression vector and cloned into the procaryotic expression vector pictured in Fig. 2. A similar level of Tat expression was observed with each Tat derivative following IPTG induction in E. coli (Fig. 6A). The ability of each protein to bind metal was addressed by electrophoretic transfer of protein to nitrocellulose membranes, followed by incubation with $^{65}$ZnCl$_2$. This procedure has been used to identify other zinc-binding proteins (23). As illustrated in Fig. 6B, both the authentic and mutant tat proteins retained their ability to bind zinc under these conditions. Thus, it is unlikely that loss of biological activity reflects the ability of the monomeric form of Tat to bind zinc.

**DISCUSSION**

In the present study, site-directed mutagenesis was used to identify functional domains within the HIV tat protein. The tat protein was localized predominantly in the nucleus, in accord with previously published results (11). Moreover, accumulation within the nucleolus was observed. The nucleolar localization is likely to be real, since parallel experiments performed with a c-fos protein and nuclear β-galactosidase expression vector gave clear nuclear staining and nucleolar exclusion. The nucleolar localization of Tat may provide clues towards understanding its mechanism of action, which at present remains unclear. Since at least one component of the mechanism of Tat function is thought to involve posttranscriptional events, localization within the nucleolus would be consistent with mechanisms that involve mRNP formation or RNA processing or transport. Similar mechanisms have been proposed for another HIV transacting protein, Rev (regulator of virion protein expression, previously referred to as arttirs), which also accumulates in the nucleolus (unpublished data). Alternatively, if transactivation occurs via an indirect mechanism, one may speculate that intermediates of this pathway are present in the nucleolus.

By alteration of individual amino acids, three potential functional domains within the tat protein were identified. The first region critical for Tat function is present at the amino-terminal end. Deletion of the first three amino acids led to normal subcellular localization, although the ability to transactivate was lost. It is noteworthy that this substitution removed one of the three negatively charged (glutamic acid) residues present in the protein.

The second set of mutations were made within the highly conserved cysteine-rich region. Mutations were made in which cysteines 22, 27, and 37 were changed to serine and cysteine 25 was changed to arginine. With each of these proteins, normal expression and proper subcellular localization were retained, but transactivation was abolished. It has been proposed that the cysteine-rich domain has the potential to form a zinc finger common to several other transcriptional regulatory proteins (10, 14, 21). Frankel et al. have proposed an alternative structure that is distinct from a zinc-binding finger and is involved in metal-linked dimer formation (9). Regardless of the configuration of the cysteines in the native molecule, our results clearly emphasize
FIG. 5. Localization of \( \beta \)-galactosidase fusion proteins. COS cells were transfected with plasmids pT1-\( \beta \)Gal (A and B) and pT2-\( \beta \)Gal (C). At 48 h posttransfection, cells were fixed and permeabilized, and the subcellular localization of the fusion proteins was determined by indirect immunofluorescence with anti-\( \beta \)-galactosidase antibody. No phase-contrast micrograph is shown for the cells in panel B.
FIG. 6. Zinc-65 binds to Tat. *E. coli* cells expressing Tat were lysed, and proteins were separated by SDS-PAGE. Proteins were electrophoretically transferred to nitrocellulose. The nitrocellulose filter was first treated with 0.5% ZnCl₂ and then visualized by autoradiography (B). The filter was then washed extensively to remove zinc and blotted with anti-tat antibody plus conjugated goat anti-rabbit alkaline phosphatase (A). Lanes: Control *E. coli* lysate (lane 1), control lysate obtained from *E. coli* expressing authentic tat protein (lane 2), pC22 (lane 3), pC25 (lane 4), pC27 (lane 5), pC37 (lane 6). Arrows indicate tat protein.

The importance of the individual cysteine residues for biological function. In addition, the ability of Tat and the cysteine substitution derivatives to form metal-protein complexes indicates that loss of function cannot be attributed to the ability of the monomeric form of Tat to bind metal under these conditions. Further analysis of the cysteine region, in particular the ability of the mutant cysteine substitution derivatives to form metal-linked dimers, is in progress.

The last group of tat mutations centered on the stretch of positively charged amino acids present between residues 48 and 56. Although Tat function was retained with these alterations, the activity of the individual mutants was markedly diminished. Mutant pKKS50I had only minor activity compared with the wild-type tat protein. Although protein was expressed, as evidenced by transactivation, indirect immunofluorescence failed to detect substantial levels of Tat in the nucleus. No activity was detected with plasmid pK50-S, which has a stop codon at amino acid 50. Thus, residues between positions 50 and 58 are required for function (24).

To examine further the role of the basic residues, their ability to function as a nuclear localization signal was examined. We found that a β-galactosidase fusion protein that contained the sequence GRKKR from the tat protein at the amino-terminal end of β-galactosidase accumulated within the nucleoplasm. In contrast, mutant proteins carrying amino-terminal fusions with the QRRP sequence corresponding to the second cluster of basic residues within Tat were, for the most part, localized in the cytoplasm. Therefore, it seems likely that the GRKKR sequence, or a part of this sequence, functions as a nuclear localization signal. This result is not surprising, considering that stretches of basic residues have been reported to serve a similar function in other nuclear proteins (12).

Since each of the proteins with mutations in the basic stretch demonstrated some degree of transactivation, it is likely that a proportion of tat protein accumulates within the nucleus. As each of the Tat derivatives contained only one or two amino acid alterations, it is possible that residues present in the second cluster of charged amino acids may substitute for those deleted within the true localization signal, although in themselves they do not serve the same function. It is noteworthy that the β-galactosidase fusion proteins accumulated within the nucleus and were excluded from the nucleolus, whereas the tat protein accumulated within the nucleolus. This raises the possibility that Tat contains a distinct nucleolar localization signal or that the intrinsic structure of the protein controls localization.

In summary, our results suggest that the small 14-kDa tat protein is a complex protein with at least three distinct functional domains. Further studies on the precise role of the individual functional regions should aid in our understanding of the mechanism of transactivation.

ACKNOWLEDGMENTS

We thank Nahum Sonnenberg and Sophie Roy for the gift of tat antibody, Reiner Gentz for plasmid pDS56RBS11-1, Richard Chiczonite, Paulo Ferreira, and James Hemstead for advice on immunofluorescence, Tom Curran for the c-fos vector, Robert Crowl for the *E. coli* lacZ gene, and Bryan Cullen for helpful discussion.

W.A.H. was supported by Public Health Service grant CA42098 from the National Institutes of Health.

LITERATURE CITED


