Mutational analysis of a human immunodeficiency virus type 1 Tat protein transduction domain which is required for delivery of an exogenous protein into mammalian cells

Jinseu Park,1 Jiyoon Ryu,1 Kyeong-Ae Kim,1 Hak Joo Lee,1 Jae Hoon Bahn,1 Kyuhyung Han,1 Eui Yul Choi,1 Kil Soo Lee,1 Hyeok Yil Kwon2 and Soo Young Choi1

1 Division of Life Sciences, Hallym University, Chunchon 200-702, Korea
2 Department of Physiology, College of Medicine, Hallym University, Chunchon 200-702, Korea

The human immunodeficiency virus type 1 (HIV-1) Tat protein transduction domain (PTD), which contains a high proportion of arginine and lysine residues, is responsible for highly efficient protein transduction through the plasma membrane. To identify the role of the PTD sequence motif in transduction, various deletions and substitutions were introduced into the PTD. Tat–green fluorescent protein (GFP) fusion proteins, containing various lengths of the Tat PTD, were expressed and the extent of their transduction into mammalian cells was analysed by Western blot analysis and fluorescence microscopy. Deletion analysis of PTD mapped to a nine amino acid motif (residues 49–57: RKKRRQRRR) sufficient for transduction. Further deletion of this Tat basic domain either at the N terminus or at the C terminus significantly decreased transduction efficiency. The transduction efficiencies of GFPs fused to nine consecutive lysine (9Lys–GFP) or arginine (9Arg–GFP) residues were similar to that of Tat(49–57)–GFP. The transduced proteins localized to both the nucleus and the cytosol, as assessed by confocal microscopy and Western blot analysis of subcellular fractions from transduced cells. Thus, the availability of recombinant GFP fusion proteins facilitates the simple and specific identification of protein transduction mediated by these peptide sequences. The modified PTD sequences designed in this study may provide useful tools necessary for delivering therapeutic proteins/peptides into cells.

Introduction

Delivering proteins with therapeutic potential into cells is difficult because of their size and biochemical properties. Thus, it has been difficult to utilize such proteins as therapeutic drugs (Egleton & Davis, 1997). Therefore, the therapeutic application of proteins could be achieved by the development of delivery vectors that are capable of the efficient delivery of various proteins into cells.

Human immunodeficiency virus type 1 (HIV-1) Tat protein, which can be secreted from infected cells, has the ability to enter neighbouring cells through the plasma membrane and accumulate in the cell (Frankel & Pabo, 1988; Ensoli et al., 1993; Green & Loewenstein, 1988; Mann & Frankel, 1991). Due to this property, whole Tat protein or part of it has been tested because of its ability to deliver several proteins, including ovalbumin, β-galactosidase and horseradish peroxidase into cells (Fawell et al., 1994; Watson & Edwards, 1999). A basic domain of the Tat protein rich in arginine and lysine residues, called the protein transduction domain (PTD), has been identified as being responsible for the ability to traverse the plasma membrane. It has also recently been shown to serve as a carrier to direct the uptake of heterologous proteins into cells by generating genetic in-frame PTD fusion proteins (Jin et al., 2001; Kwon et al., 2000; Nagahara et al., 1998; Vocero-Akbani et al., 1999). Furthermore, it has recently been reported that β-galactosidase, fused to the basic domain of HIV-1 Tat, was delivered in the biologically active form to all tissues including the brain when injected intraperitoneally into mice (Schwarze et al., 1999).

Other proteins, such as the homeodomain of Drosophila antennapedia (Antp) and the herpes simplex virus type 1 VP22 transcription factor, have also been shown to enter cells in culture (Joliot et al., 1991; Elliott & O’Hare, 1997; Han et al., 2000). Small regions of these proteins, ranging from 11 to 34
**Table 1. Oligonucleotides used for construction of plasmids encoding a series of PTD variants, the corresponding amino acid sequences and the resulting GFP fusion proteins**

<table>
<thead>
<tr>
<th>Amino acid sequence of PTD*</th>
<th>Nucleotide sequence (sense/antisense pairs, 5’ to 3’)+</th>
<th>GFP fusion proteins‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>39GRKKRRQRRR37</td>
<td>TAGCGCCAGAAGAAGCGAGACAGCGACGAAGAC</td>
<td>Tat(48–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>39RKRRQRRRR37</td>
<td>TAGGAGAACGCGACAGCGACGAAGAC</td>
<td>Tat(49–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>59KKRRQRRR37</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(50–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>51RRQRRRR37</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(51–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>50KKRR35</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(52–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>50KRRQ24</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(53–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>50KRRQR55</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(54–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>KKKKKK24</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(55–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>RRRRRRR24</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>Tat(56–57)–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td>RRRRRRRR</td>
<td>TAAGAGACGAGACAGCGACGAAGAC</td>
<td>9Lys–GFP</td>
</tr>
<tr>
<td></td>
<td>TCGAGTCTTCGCTGTCTCCGCTTCTCTCCGCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCGAGACGAGAGGCGAGACGAGAGC</td>
<td>9Arg–GFP</td>
</tr>
</tbody>
</table>

* Numbers indicate amino acid positions, in the context of HIV-1 Tat.
† Oligonucleotide sequences of PTD variants engineered between the His tag region and the GFP gene are shown.
‡ Removal of the C- or N-terminal amino acid of the basic domain in a sequential manner results in GFP fusion proteins with different lengths of the Tat basic domain. These fusion proteins were designated according to the amino acid positions in the Tat protein.

amino acids in length, were able to cross the lipid bilayer of cells either alone or fused to various polypeptides or oligonucleotides (reviewed in Kwon et al., 2000; Schwarze & Dowdy, 2000; Schwarze et al., 2000; Lindgren et al., 2000; Derossi et al., 1998).

Although it has been suggested that transduction occurs in a receptor- and transporter-independent fashion that appears to target the lipid bilayer directly, the mechanism of transduction mediated by HIV-1 PTD needs to be elucidated (Derossi et al., 1998). The cellular internalization of homeoproteins and homeodomains occurs at 4 °C and 37 °C and cannot be saturated, which suggests a transduction mechanism that is energy- and receptor-protein-independent (Joliot et al., 1991; Derossi et al., 1994, 1996). Therefore, transduction by these proteins is independent of the classical endocytosis pathway, and it shows such characteristics as high efficiency, non-cell-type specificity and low toxicity.

In the present study, we have defined the sequence requirements for the HIV-1 PTD by deletion analysis and described the contribution of basic amino acid sequences to transduction activity by substitution of HIV-1 PTD with nine consecutive arginines or lysines.

### Methods

#### Construction of expression vectors.

The complete green fluorescent protein (GFP) gene sequence was amplified from plasmid pEGFP–C2 (Clontech) by PCR using Pfu DNA polymerase (Clontech). The sense primer was 5’ CTCAGAGTGAAGAAGCCGAAGACGCTG 3’ and the antisense primer was 5’ GGATCCTTACTTGTACAGCTCGTTCCATGCAGAG 3’. The PCR product was digested with XhoI–BamHI and subcloned into XhoI–BamHI sites of pET15b (Invitrogen) to construct pGFP, which expresses the GFP fusion protein without the basic domain of HIV-1 Tat. Clones with an expected 0.7 kbp insert were selected using XhoI–BamHI restriction analysis and analysed by sequencing (Sambrook et al., 1989).

pTat–GFP was constructed in the following manner to express the basic domain (amino acids 48–57) of HIV-1 Tat as a fusion protein with GFP. First, two oligonucleotides were synthesized and annealed to generate a double-stranded oligonucleotide, encoding the nine amino acids from the basic domain of HIV-1 Tat. The sequences were (top strand) 5’ TAGGAGAACGCGACAGCGACGAAGAC and (bottom strand) 5’ TCGAGTCTTCGCTGTCTCCGCTTCTCTCC 3’. The double-stranded oligonucleotide was directly ligated into the NdeI–XhoI-digested pGFP, in-frame with the C-His open reading frame to generate the Tat–GFP expression plasmid, pTat–GFP. This plasmid, encoding the basic domain (amino acids 48–57) of HIV-1 Tat fused with
GFP was modified such that oligonucleotides corresponding to a series of PTD variants were annealed and inserted between the His tag region and the GFP gene of pGFP. The sequences of the oligonucleotides cloned into the plasmid were confirmed using a fluorescence-based automated sequencer (model 373A; Applied Biosystems). The oligonucleotides used in the construction of the plasmids with a series of PTD variants, the corresponding amino acids and the resulting GFP fusion proteins are summarized in Table 1.

Expression and purification of Tat–GFP fusion proteins.
BL21 E. coli (Pharmacia) transformed with plasmids encoding GFP or Tat–GFP fusion proteins was grown overnight at 37 °C in LB broth supplemented with 100 μg/ml ampicillin. The overnight culture was diluted tenfold with fresh LB medium and cultured at 37 °C with shaking at 250 r.p.m. until OD₆₀₀ = 1.0. Protein expression was induced by adding IPTG to a final concentration of 0.5 mM for 4 h. To prepare the denatured Tat–GFP fusion proteins, the induced cells were harvested and lysed by sonication in a binding buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris–HCl, pH 7–9) containing 6 M urea and protease inhibitors (20 mg/ml soybean trypsin inhibitor, 2 mg/ml aprotinin, 5 mg/ml leupeptin, 100 mg/ml PMSF). After removal of the cell debris by centrifugation, the clarified cell extract was then loaded on to a nickel Sepharose column. Proteins were eluted by an elution buffer (1 M imidazole, 20 mM Tris–HCl, pH 7–9), followed by desalting with a PD10 column (Amersham). The native Tat–GFP fusion protein was obtained by the same procedure but without a denaturing agent. The protein concentration in each fraction was quantified by densitometric analysis after separation by SDS–PAGE, using bovine serum albumin (BSA) as a standard. The protein concentrations were determined with a Bradford protein assay (Bio-Rad), and using bovine serum albumin (BSA) as a standard. The protein concentration in each fraction was determined with a Bradford protein assay (Bio-Rad), and using bovine serum albumin (BSA) as a standard. The protein concentration in each fraction was determined with a Bradford protein assay (Bio-Rad), and using bovine serum albumin (BSA) as a standard.

Western blot analysis. Cell lysates were prepared by lysing monolayer cells on a six-well plate with a lysis buffer (125 mM Tris–HCl,
Results

Construction of Tat–GFP fusion protein expression vectors

Previous studies have implicated the basic domain of HIV-1 Tat, the PTD, as being involved in protein transduction into cells. This basic domain consists of six arginine residues, two lysine residues and one glutamine residue. In the present study, we used GFP as a model protein to confirm that this domain can mediate the transduction of foreign protein into cells and to identify the sequence requirements of the HIV-1 Tat basic domain for transduction activity.

To construct a vector for expression of Tat–GFP fusion protein, the initial coding sequence of GFP was amplified by PCR and inserted into a PET15b vector using an N-terminal histidine hexapeptide (6-His), which allowed us to purify and concentrate recombinant proteins from the bacterial lysates by affinity chromatography using a metal-chelating matrix. Analysis of the purified Tat–GFP proteins by SDS–PAGE showed that they apparently migrated at a higher molecular mass than did the GFP protein without a Tat sequence (Fig. 1B). The purified products were shown to react with a rabbit polyclonal antibody to GFP by Western blot analysis (data not shown).

To investigate whether or not the denatured Tat(48–57)–GFP was transduced and restored to biological activity in the cells, various concentrations of denatured Tat(48–57)–GFP protein were added to HeLa cells for 1 h, and the protein levels and fluorescence intensity were analysed by Western blot analysis and fluorescence microscopy, respectively. As shown in Fig. 2(A), the level of transduced protein inside the cells increased and fluorescence signals were detected in a dose-dependent manner (Fig. 2B). Unlike Tat–GFP, GFP without PTD could not be transduced into the cells. These data indicate that the denatured Tat(48–57)–GFP
Mutational analysis of the HIV-1 Tat PTD

Deletion analysis of the HIV-1 basic domain to determine the sequence required for transduction activity

The ten amino acid peptide corresponding to residues 48–57 of the HIV-1 Tat protein was used as a starting reference, since it contains the complete basic domain and serves as the PTD (Nagahara et al., 1998). In order to determine the minimum length required for transduction, a set of deletions in the HIV-1 PTD was systematically introduced into the Tat–GFP fusion protein (Table 1). A series of Tat–GFP fusion proteins was prepared and tested for their ability to mediate intracellular delivery.

The transduction activity of Tat(49–57)–GFP was indistinguishable from that of Tat(48–57)–GFP (Fig. 3A, compare lanes 9 and 10), whereas Tat(50–57)–GFP and Tat(51–57)–GFP appeared to be transduced at slightly lower levels than Tat(48–57)–GFP (Fig. 3A, compare lanes 2, 8 and 9). These results suggested that deletion of the amino acid residue corresponding to Gly-48 did not affect the transduction efficiency, and that the minimum basic domain sequences for the efficient transduction was from 49 to 57 in the PTD.

Further deletion of Lys-50 resulted in a slight decrease of transduction activity, while deletion of Lys-51 dramatically abolished transduction activity (Fig. 3A, B). In a similar manner, sequential deletions of the C-terminal amino acids (Arg-57, Arg-56, Arg-55 and Gln-54) produced lower efficiencies of transduction activity. The results analysed by fluorescence microscopy were consistent with those transduction activities shown in the Western blot analysis (Fig. 3C). In general, Tat–GFP fusion protein with a full-sized basic domain was translocated into both the nucleus and the cytoplasm, while GFP fusion proteins with a truncated basic domain were present in the cell at a low level. Therefore, we concluded that the basic domain from 49 to 57 is required for efficient transduction activity, whereas further deletions of amino acids from the N or C terminus of the Tat basic domain lead to significant loss of transduction activity.

Transduction activity with basic amino acids modified to polylysine or polyarginine

The basic domain from 49 to 57 of HIV-1 Tat consists of six arginine residues, two lysine residues and one glutamine residue. This suggests that the positively charged amino acids within this domain may influence transduction efficiency. To
investigate this possibility, we substituted the basic domain with polylsine (9Lys–GFP) or polyarginine (9Arg–GFP) peptides.

To determine the transduction activity of Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP, various concentrations of each fusion protein were added to the culture media of HeLa cells for 1 h and transduction analysed by Western blotting and fluorescence microscopy. As shown in Fig. 4, all GFP fusion proteins were successfully delivered into the cells with similar transduction efficiency, whereas control GFP was not delivered into the cells. Transduced proteins appeared to locate either in...
the cytoplasm or in the nucleus. To clarify further the subcellular localization of transduced proteins in the cells, nuclear and cytosolic fractions were prepared from cells transduced with Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP and analysed by Western blotting using anti-GFP, anti-actin or anti-PARP antibodies. As shown in Fig. 5, Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP were detected at similar intensity in the nucleus as well as in the cytoplasm of transduced cells, whereas control GFP was not detected in the cells. When the intracellular localization of transduced proteins was visualized by confocal microscopy, Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP were found to be present in both the nucleus and the cytosol (Fig. 6). The distribution of 9Lys–GFP and 9Arg–GFP was very similar to that of Tat(49–57)–GFP, although neither polylysine nor polyarginine contains a typical nuclear-localization signal. To confirm that Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP can localize to the nucleus, transduced cells were double-stained with the nucleus-specific marker PI. Dual-colour detection of the various GFP fusion proteins and PI demonstrated that Tat(49–57)–GFP, 9Lys–GFP and 9Arg–GFP localized to the nucleus, while GFP alone did not (Fig. 6). Taken together, these results indicated that substituting the HIV-1 basic domain with nine arginines or nine lysines retained full transduction activity and did not affect the subcellular localization of transduced proteins.

Discussion

Previous studies have demonstrated that the HIV-1 Tat PTD directs the transduction of heterologous proteins into cells and is sufficient for transduction in the absence of other regions of Tat. In this study, taking advantage of the ability of Tat–GFP to be detected by fluorescence microscopy in cells, efforts were made to delineate sequence requirements for the transduction of heterologous protein through the plasma membrane of cells mediated by the HIV-1 Tat basic domain. The green fluorescent protein from the jellyfish *Aequorea victoria* is one of the most widely used reporter proteins in the analysis of protein targeting or trafficking in cells (Tsien, 1998). GFP fusion proteins with serially deleted or substituted HIV-1 Tat basic domain were prepared and analysed for their ability to mediate intracellular delivery. The minimal sequence required for efficient transduction was localized to the basic domain from residues 49 to 57 of HIV-1 Tat. Further deletions revealed that both N- and C-terminal amino acids of the Tat basic domain (49–57) peptide were important for efficient transduction activity. In addition, GFP fused to nine consecutive lysines or arginines was demonstrated to retain full transduction activity with an efficiency similar to that of Tat(49–57)–GFP. Since the HIV-1 Tat basic domain mainly consists of positively charged amino acids, it may interact with anionic components on the surface of the cell membrane. Polysulfonated compounds like heparin have been shown to be potent inhibitors of transduction mediated by the HIV-1 basic domain (Rusnati et al., 1998). The involvement of cell-surface heparan sulfate (HS) proteoglycans in the transduction of HIV-1 Tat protein was recently demonstrated, firstly by using cells genetically defective in the biosynthesis of fully sulfated HS and the competitive inhibition of Tat uptake by soluble heparin, and secondly by treatment with glycosaminoglycan lyases specifically degrading HS chains (Tyagi et al., 2001).

It has been suggested that the transduction of protein into cells could be mediated by the HIV-1 basic domain due simply to an inherent characteristic of specific amino acid sequences such as arginine or lysine (Schwarze et al., 2000). This hypothesis is supported by our result with substitutions of the HIV-1 basic domain by nine arginine or nine lysine residues. The substituted 9Arg–GFP and 9Lys–GFP were each shown to transduce as efficiently into cells as Tat–GFP. When nine arginine or nine lysine residues were fused with other heterologous proteins, such as superoxide dismutase (Park et al., 2002), catalase (Jin et al., 2001), *E. coli* nitroreductase and p21CIP1/WAF1 (unpublished results), these fusion proteins were also efficiently transduced into cells. These homogeneous polyarginine and polylysine peptides may be non-immunogenic when administered *in vivo*. Therefore, the protein-transduction domains substituted with arginines or lysines as the sole carrier vehicle may have advantages over HIV-1 PTD itself.

Recently, Wender et al. (2000) examined the transduction efficiency of truncated versions of the Tat basic domain and of nine arginine peptides that were directly fluorescent-labelled. However, in the present study, the potential of various oligopeptides as PTDs was studied as forms genetically fused to heterologous protein. This approach may more closely reflect eventual ‘pharmaceutical’ usage for protein delivery. We found that the relative transduction efficiencies of fused proteins with different length of PTD were similar to those of free oligopeptides with a corresponding number of basic residues, although there were some differences in the degree of loss between the free and the fused forms. We found that 9Arg–GFP was transduced into the cells with a similar efficiency to Tat(49–57)–GFP, while Wender et al. (2000) observed that a fluorescent-labelled nine-arginine peptide was more efficient at entering cells than Tat(49–57) (RKKRRQRRR). This discrepancy in the transduction efficiency may derive from the properties of the fusion target protein, including the degree of unfolding, polarity and molecular shape of the protein.

When the Tat–GFP fusion proteins were observed by fluorescence microscopy and confocal microscopy, they were translocated into both the cytoplasm and the nucleus. It has been suggested that this cluster of nine basic amino acids, Tat(49–57), serves as a nuclear-localization signal (NLS) (Efthymiadis et al., 1998; Ruben et al., 1989). Although the glycine residue at the 48 position was initially reported to serve as part of the NLS, it is clear that this residue is required for neither transduction nor nuclear localization. This was
demonstrated by Tat{49–57}–GFP, which contained no glycine residue, being transduced into cells and localized both in the cytoplasm and the nucleus with an efficiency similar to that of Tat{48–57}–GFP (Fig. 5). Like Tat{49–57}–GFP, the translocation of 9Lys–GFP and 9Arg–GFP was also observed in the nucleus, as well as in the cytosol (Fig. 6).

The molecular mechanisms of the transduction of heterologous protein mediated by the HIV-1 basic domain through the cell membrane are not yet characterized. In general, the denatured proteins containing HIV-1 Tat PTD transduce more efficiently into cells than do correctly folded proteins (Kwon et al., 2000; Nagahara et al., 1998; Schwarze et al., 1999). It has been suggested that the increased transduction efficiency of denatured proteins may result from reduced structural constraints, which allow proteins to pass more easily through the cell membrane (Nagahara et al., 1998). Supporting evidence comes from a previous study, which found that unfolding was necessary to traverse the cellular membrane when Tat was used to ferry dihydrofolate reductase into cells (Bonifaci et al., 1995). Further evidence has been provided by a recent study on the structural characterization of HIV-1 Tat showing that the basic region of HIV-1 Tat is well exposed to solvent (Peloponese et al., 2000). Therefore, these results imply that the conformation of proteins may be critical in the efficient transduction of heterologous proteins into cells.

In summary, GFP fusion proteins containing the various protein-transduction domains developed in this study facilitated a comparative analysis of transduction efficiency and provided useful tools for studying the mechanism of transduction mediated by the basic domain of the HIV-1 Tat protein. In addition, these GFP fusion proteins have the potential to be used in a variety of applications, including studies of in vivo and in vitro protein delivery.

This work was supported by grants from the National Research Laboratory Program (M1-9911-00-0025) of the Korean Ministry of Science and Technology, and in part by the grant from the Hallym Academy of Sciences at Hallym University (1995), Korea.

References


Received 16 November 2001; Accepted 4 January 2002