Published online 14 May 2008

Nucleic Acids Research, 2008, Vol. 36, No. 10 ***e62***
doi:10.1093/nar/gkm1170

Terminal proteins of *Streptomyces* chromosome can target DNA into eukaryotic nuclei

Hsiu-Hui Tsai[1], Chih-Hung Huang[2], Alan M. Lin[1] and Carton W. Chen[1,*]

[1]Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Shih-Pai, Taipei 112 and [2]Institute of Biotechnology, National Taipei University of Technology, Taipei 106, Taiwan

Received November 2, 2007; Revised December 18, 2007; Accepted December 19, 2007

ABSTRACT

Streptomyces **species are highly abundant soil bacteria that possess linear chromosomes (and linear plasmids). The 5′ ends of these molecules are covalently bound by terminal proteins (TPs), that are important for integrity and replication of the telomeres. There are at least two types of TPs, both of which contain a DNA-binding domain and a classical eukaryotic nuclear localization signal (NLS). Here we show that the NLS motifs on these TPs are highly efficient in targeting the proteins along with covalently bound plasmid DNA into the nuclei of human cells. The TP-mediated nuclear targeting resembles the inter-kingdom gene transfer mediated by Ti plasmids of** ***Agrobacterium tumefaciens*****, in which a piece of the Ti plasmid DNA is targeted to the plant nuclei by a covalently bound NLS-containing protein. The discovery of the nuclear localization functions of the** ***Streptomyces*** **TPs not only suggests possible inter-kingdom gene exchanges between** ***Streptomyces*** **and eukaryotes in soil but also provides a novel strategy for gene delivery in humans and other eukaryotes.**

INTRODUCTION

The linear chromosomes and plasmids of *Streptomyces* species are capped by terminal proteins (TPs) at the 5′ ends of the DNA (1). The TP provides protection against exonuclease attack on the DNA, and functions as a primer for DNA synthesis to patch the single-stranded gaps at the 3′ ends during replication (2).

Several *Streptomyces* TPs have been isolated or identified from genome sequences. Most of them (designated Tpg) are highly conserved in sequences and size (184–185 aa) (3,4). On *Streptomyces* chromosomes, the *tpg* gene forms an operon with a *tap* gene, which encodes another protein essential for end-patching DNA synthesis (5). Bao and Cohen (5) showed that the Tap protein of *S. coelicolor* (Tap^{Sco}) interacts with Tpg of *S. coelicolor* (Tpg^{Sco}) and the single-stranded telomere DNA, and proposed that Tap^{Sco} recruits and positions Tpg^{Sco} at the telomere during its replication. In an *in vitro* system, Yang *et al.* (6) demonstrated that Tpg^{Sco} was specifically deoxynucleotidylated by dCMP (the first nucleotide of the *S. coelicolor* chromosome) at a Thr residue.

A monopartite nuclear localization signal (NLS) motif was predicted downstream of and adjacent to a helix-turn-helix DNA-binding domain at the N-terminus of Tpg^{Sco} and the identical Tpg of *S. lividans*, Tpg^{Sli} (4). As more Tpg homologs were identified, NLS motifs were also found in all of them except perhaps that of pSCL2 plasmid and the predicted pseudogene products (Fig. 1A). All these NLS motifs contain the consensus K(K/R)X(K/R) sequence for the basic core of monopartite NLS (7). Initially, this discovery was regarded as fortuitous, because: (*i*) NLSs often overlap with a DNA-binding domain, and sometimes are used for DNA binding (8,9); and (*ii*) a nuclear localization function of a TP would appear incongruous in streptomycetes that lack a nucleus.

Recently, a novel TP (designated Tpc) encoded by linear plasmid SCP1 of *S. coelicolor* was isolated and characterized (10). Tpc is distinct from Tpgs in both aa sequence and size (259 versus 184–185 aa), and represents the product of convergent evolution. Tpc also contains a predicted NLS (Fig. 1B), which, however, differs from that on Tpg^{Sco} and Tpg^{Sli} in being separate from the DNA-binding domain, and in being bipartite. The finding of two distinct types of NLSs in two different types of TP suggested that their occurrences were not coincidental, and that they serve a real biological function.

In this study, we showed that the NLSs on both types of TP (Tpgs and Tpc) are functional in nuclear targeting. When fused to a triple green fluorescence protein concatemer (EGFP3), they could target the fusion protein into human nuclei. These TPs could also carry covalently attached DNA into the nuclei. TPs with a mutation in NLS are defective in nuclear localization, but remain competent in supporting end-patching and capping of linear replicons. This suggests that the nuclear targeting function of TPs has evolved independently of the end-patching function. All these findings indicate that the

*To whom correspondence should be addressed. Tel: +1 886 2 2826 7040; Fax: +1 886 2 2826 4930; Email: cwchen@ym.edu.tw

© 2008 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

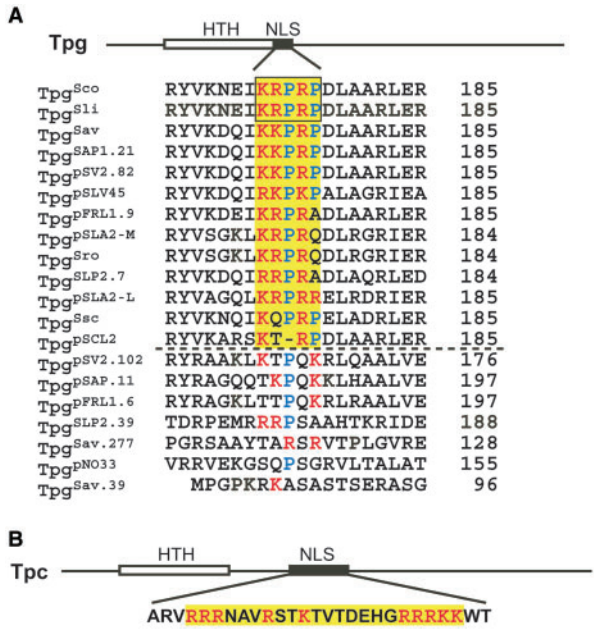

Figure 1. Potential NLS sequences in TPs of *Streptomyces*. (**A**) The archetypal Tpg family. The plasmid-encoded Tpgs are designated by the plasmid names, and the chromosome-encoded Tpgs are designated by three-letter abbreviations of the species (Sav, *S. avermitilis*; Sco, *S. coelicolor*; Sli, *S. lividans*; Sro, *S. rochei*; Ssc, *S. scabies*). For those Tpgs that are encoded by the same replicon, they are distinguished by their gene numbers. Sources of the sequences are: *S. coelicolor* chromosome (4), *S. lividans* chromosome (3), *S. avermitilis* chromosome and SAP1 plasmid (38), pSV2 plasmid (pSV2.82) in *S. violaceoruber* (GenBank accession number NC_004934), pSLV45 plasmid *S. lavendulae* (39), pFRL1 plasmid in *Streptomyces* sp. FR1(40), *S. rochei* chromosome and pSLA2-L and pSLA2-M plasmids (41), SLP2 plasmid in *S. lividans* (4), *S. scabies* chromosome (http://www.sanger.ac.uk/Projects/S_scabies/), pSCL2 plasmid in *S. clavuligerus* (GenBank accession number AAQ93595), pNO33 plasmid in *S. albulus* (GenBank accession number YP_170689). The locations of the DNA-binding HTH domain and potential NLS sequences are depicted by the open and filled box, respectively, on the prototype Tpg^{Sco}. The potential NLS in various Tpgs (except the putative pseudogene products) are shaded in yellow. Within this region, the basic aa's are in red and Pro in blue. The length (in aa) of the Tpg proteins is indicated at the right. Conceptually translated products of proven pseudogenes ($Tpg^{SLP2.39}$; Yang, C.-C., unpublished results) or putative pseudogenes (widely divergent sequence and/or anomalous length) are placed below the dashed line. (**B**) Tpc of SCP1 plasmid. The labels are as in (A).

nuclear targeting of the TP-capped linear replicons of *Streptomyces* is biologically significant, and may mediate inter-kingdom gene transfer in soil.

MATERIALS AND METHODS

Growth and genetic manipulations of bacterial cultures and plasmids

Bacterial cultures and plasmids are listed in Table 1. Basic microbiological and molecular biological procedures were according to Kieser *et al.* (11) and Sambrook *et al.* (12). *S. lividans* TK64 (13) and MR04 (14) was used for propagation of *Streptomyces* plasmids. Mutations in cloned genes were generated by site-directed mutagenesis by PCR.

Construction of EGFP3 fusion proteins

Tpg and Tpc genes and their NLS-deleted sequences were obtained by PCR, and oligonucleotides containing the NLS sequences were commercially synthesized. These sequences were inserted between the *Sac*I and *Eco*RI sites upstream of EGFP3 (encoding a triple green fluorescence protein concatemer under the control of the CMV immediate-early promoter) on pEGFP3 (15) to generate TP-EGFP3 and NLS-EGFP3 fusion proteins, respectively.

Construction of mini linear plasmids

The 3.4 kb *Sac*I-*Hin*dIII fragment spanning the tap^{Sco}-tpg^{Sco} operon was generated by PCR, and inserted between the *Sac*I and *Hin*dIII sites of pLUS966 to generate pLUS986. The tap^{Sco}-tpg^{Sco} operon containing the K3A or R4A mutations was created by PCR using appropriate primer sets (listed in Supplementary Data), and used to replace the tap^{Sco}-tpg^{Sco} operon on pLUS986 to generate pLUS986(K3A) and pLUS986(R4A), respectively.

Mini linear plasmid pLUS892L containing the *tas* and *tpc* genes, the pSLA2 ARS, and a pair of SCP1 telomeres and its circular progenitor pLUS892 were described previously (10). The *tac-tpc* sequence containing the ΔARVRRR) mutation in *tpc* was created by PCR using appropriate primer sets (listed in Supplementary Data), and used to replace the corresponding *Hin*dIII fragment of pLUS892 to generate pLUS892 Δ(ARVRRR). *Hin*dIII-linearized pLUS966 and *Ase*I-linearized pEGFP3 were filled in by DNA polymerase I to create blunt ends, and ligated by T4 DNA ligase to create pLUS966-EGFP3.

Generation of linear plasmids from the circular progenitor plasmids followed the general procedure of Qin and Cohen (16). The circular plasmid DNA was linearized by *Ase*I in the ColE1 vector sequence and used to transform *S. lividans* TK64 or MR04. Linear plasmids were isolated from thiostrepton-resistant transformants, and confirmed by restriction analysis.

Transfection of human cell cultures

HeLa and HEK 293T human cell lines were grown in DMEM medium supplemented with 10% (vol/vol) fetal bovine serum. They were transfected using lipofectamine according to the procedure specified by the manufacturer (Invitrogen). Fluorescent transfected cells were scored under a fluorescence microscope. To prepare TP-capped linear plasmid DNA for transfection, *Streptomyces* cultures containing the plasmid were grown in YEME medium to exponential phase, harvested by centrifugation, treated with lysozyme (1 mg/ml) at 37° for 30 m, and osmotically lyzed by dilution in 10 vol of TE buffer. The lysate was electrophoresed in 0.8% agarose gel containing 0.05% SDS. Linear plasmid DNA was visualized by ethidium bromide staining and eluted electrophoretically.

Table 1. Bacterial cultures and plasmids used in this study

Strain/plasmid	Relevant genotype/description	Source/reference
S. lividans TK64	*pro-2 str-6* SLP2$^-$ SLP3$^-$	(13)
S. lividans MR04	*pro-2 str-6 rec-46* Δ*dndA* Δ(*tap*^{Sli}-*tpg*^{Sli}) SLP2$^-$ SLP3$^-$	(14)
pEGFP3	Plasmid containing a reporter gene EGFP3 (encoding triple EGFP protein concatemer) under the control of the CMV immediate-early promoter P_{CMVIE}	(15)
pEGFP3-Tpg^{Sco}	pEGFP3 containing *tpg*^{Sco} fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-Tpg^{Sco}ΔNLS5C	pEGFP3 containing *tpg*^{Sco} with a deletion of KRPRP fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3- Tpg^{Sco}ΔNLS10C	pEGFP3 containing *tpg*^{Sco} with a deletion of EIKRPRPDLA fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-NLS5C	pEGFP3 containing KRPRP fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-NLS10C	pEGFP3 containing EIKRPRPDLA (from Tpg^{Sco}) fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-NLS10C(E1A)	pEGFP3 containing AIKRPRPDLA fused to the N-terminus of EGFP3	Fig. 2B; this study
pEGFP3-NLS10C(I2A)	pEGFP3 containing EAKRPRPDLA fused to the N-terminus of EGFP3	Fig. 2B; this study
pEGFP3-NLS10C(K3A)	pEGFP3 containing EIARPRPDLA fused to the N-terminus of EGFP3	Fig. 2B; this study
pEGFP3-NLS10C(R4A)	pEGFP3 containing EIKAPRPDLA fused to the N-terminus of EGFP3	Fig. 2B; this study
pEGFP3-NLS10C(R6A)	pEGFP3 containing EIKRPAPDLA fused to the N-terminus of EGFP3	Fig. 2B; this study
pEGFP3-Tpg^{Sav}	pEGFP3 containing *tpg*^{Sav} fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-NLS10A	pEGFP3 containing QIKKPRPDLA (from Tpg^{Sav}) fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-NLS10R	pEGFP3 containing KLKRPRQDLR (from Tpg$^{pSLA2\text{-}M}$/Tpg$^{pSLA2\text{-}L}$) fused to the N-terminus of EGFP3	Fig. 2A; this study
pEGFP3-Tpc	pEGFP3 containing Tpc fused to the N-terminus of EGFP3	Fig. 3; this study
pEGFP3-NLS27S	pEGFP3 containing the 27-aa NLS (from Tpc) fused to the N-terminus of EGFP3	Fig. 3; this study
pEGFP3-TpcΔ(ARVRRR)	pEGFP3 containing TpcΔ(ARVRRR) fused to the N-terminus of EGFP3	Fig. 3; this study
pEGFP3-TpcΔ(RRRKKWT)	pEGFP3 containing TpcΔ(RRRKKWT) fused to the N-terminus of EGFP3	Fig. 3; this study
PLUS986L	linear plasmid containing the *tap*^{Sco}-*tpg*^{Sco} operon, pSLA2 ARS and a pair of *S. lividans* chromosomal telomeres	Fig. 4A; this study
pLUS986(K3A)	pLUS986 with the K3A mutation in the NLS of Tpg^{Sco}	Fig. 4; this study
pLUS986(K3A)L	linear version of pLUS986(K3A)	Fig. 4; this study
pLUS986(K3A)-EGFP3	pLUS986(K3A) containing EGFP3 under the control of the CMV immediate-early promoter P_{CMVIE}	This study
pLUS986(K3A)-EGFP3L	linear version of pLUS986(K3A)-EGFP3	This study
pLUS986(R4A)	pLUS986 with the R4A mutation in the NLS of Tpg^{Sco}	Fig. 4; this study
pLUS986(R4A)L	linear version of pLUS986(R4A)	Fig. 4; this study
pLUS892	plasmid containing the *tac-tpc* region of SCP1, pSLA2 ARS of and a pair of SCP1 telomeres	(10)
pLUS892L	linear version of pLUS892	Fig. 4A; (10)
pLUS892(ΔARVRRR)	pLUS892 containing the (ΔARVRRR) mutation in *tpc*	Fig. 4; this study
pLUS892(ΔARVRRR)L	linear version of pLUS892(ΔARVRRR)	Fig. 4; this study
pLUS966	recombinant circular plasmid containing the terminal 372 bp of the *S. lividans* chromosome, the *tsr* (thiostrepton resistance) gene and a 6.4 kb autonomously replicating sequence of pSLA2	(4)
pLUS966-EGFP3	pLUS966 with pEGFP3 sequence inserted at the *Hin*dIII site	Fig. 5A; This study
pLUS966-EGFP3L	linear version of pLUS966-EGFP3	Fig. 5B; This study

NLS prediction

NLSs were predicted using the PredictNLS server (http://cubic.bioc.columbia.edu/predictNLS) based on Cokol *et al.* (9).

RESULTS

Tpgs and their NLSs are functional in nuclear targeting

To test the nuclear localization function, we fused the Tpg^{Sco} and Tpg^{Sav} sequences to the reporter gene EGFP3 (encoding a triple green fluorescence protein concatemer) under the control of the CMV immediate-early promoter (P_{CMVIE}) on pEGFP3 (15) (Fig. 1A), and introduced the constructed plasmids, pEGFP3-Tpg^{Sco} and pEGFP3-Tpg^{Sav}, into HeLa and/or HEK 293T human cells by transfection using the lipofectamine procedure. Green fluorescence was produced in ~80% of the transfected cells in 16 h after transfection, and the fluorescence accumulated in the nuclei of these cells (Fig. 2A). In contrast, in cells transfected by the pEGFP3 vector, fluorescence was present mainly in the cytosol. These results indicated that the Tpg^{Sco} and Tpg^{Sav} sequences were functional in nuclear targeting.

To test the role of the putative NLS of Tpg^{Sco} in nuclear localization, the predicted NLS motif (KRPRP) was fused to the N-terminus of EGFP3 (Fig. 2A). HeLa cells transfected by the resulting plasmid, pEGFP3-NLS5C, displayed green fluorescence mainly in the cytosol. On the other hand, Tpg^{Sco} with a deletion of the pentapeptide lost its nuclear localization function when fused to EGFP3 (pEGFP3-Tpg^{Sco}ΔNLS5C; Fig. 2A). When an expanded NLS motif-containing decapeptide, EIKRPRPDLA, was fused to EGFP3 (pEGFP3-NLS10C), the fused protein was concentrated in the nuclei of transfected HeLa cells (Fig. 2A). EGFP3 fused to Tpg^{Sco} lacking this decapeptide was localized mainly in the cytosol (pEGFP3-Tpg^{Sco}ΔNLS10C). These results indicated that the

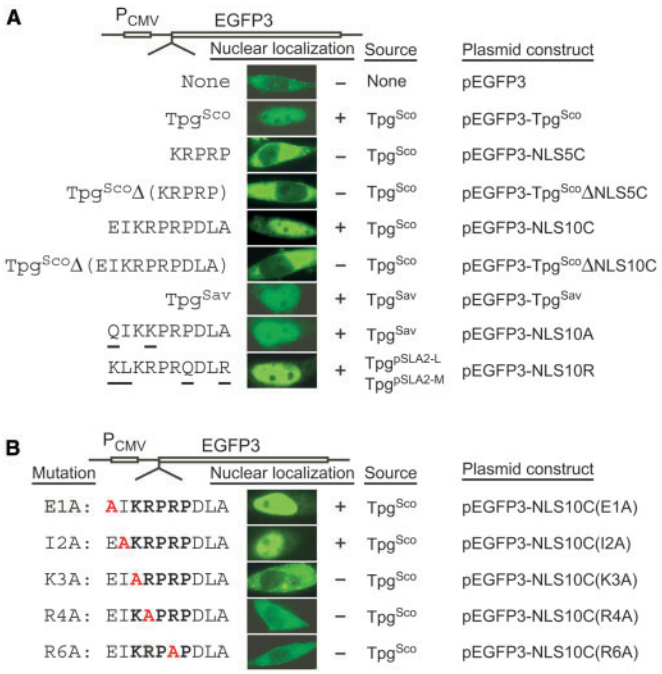

Figure 2. Nuclear localization function of Tpgs. (**A**) NLS in Tpgs. The vector used for transfection was pEGFP3. The Tpg sequences fused in-frame to the N-terminus of EGFP3 are listed to the left. The residues in Tpg^{Sav} and $Tpg^{pSLA2\text{-}L}$/$Tpg^{pSLA2\text{-}M}$ that differ from those in Tpg^{Sco} are underlined, and the mutant residue is in red. The nuclear localization test results ('–', negative; '+', positive) are shown in the fluorescent microscopic images of a representative transfected cell in the middle. The original source of the sequence and the plasmid construct are listed to the right. (**B**) 'Alanine Scan' mutant variants of Tpg^{Sco}. The mutant decapeptides are listed to the left, and the introduced alanine is in red.

EIKRPRPDLA decapeptide was necessary and sufficient for the nuclear targeting function of Tpg^{Sco}.

The putative NLS motif-containing decapeptide QIKKPRPDLA in Tpg^{Sav}, which differs from that in Tpg^{Sco} by two aa residues (Fig. 1A), could also target EGFP3 into the nuclei (pEGFP3-NLS10A; Fig. 2A). Moreover, a putative NLS motif-containing decapeptide, KLKRPRQDLR, in $Tpg^{pSLA2\text{-}M}$ and $Tpg^{pSLA2\text{-}L}$ of *S. rochei* (3,17), which differs from that in Tpg^{Sco} by four aa residues (Fig. 1A), was also competent for nuclear localization (pEGFP3-NLS10R, Fig. 2A). These results suggested that the predicted NLS sequences in all of the Tpgs (except perhaps that of pSCL2) might be functional.

The 'alanine scanning' mutation procedure (18) was employed to test the functionality of the decapeptide EIKRPRPDLA of Tpg^{Sco}. Changes of the first two aa of the decapeptide to A (E1A and I2A mutations) did not affect the nuclear localization function (Fig. 2B). Alteration of the third (K3A), fourth (R4A) or sixth (R6A) aa to A blocked nuclear localization. This result confirmed the essential role of key basic aa in the NLS.

The bipartite NLS in Tpc is also functional

Tpc, the TP of the SCP1 plasmid of *S. coelicolor*, is distinct from Tpgs in both sequence and size (10). Its central region contains a predicted bipartite NLS motif of

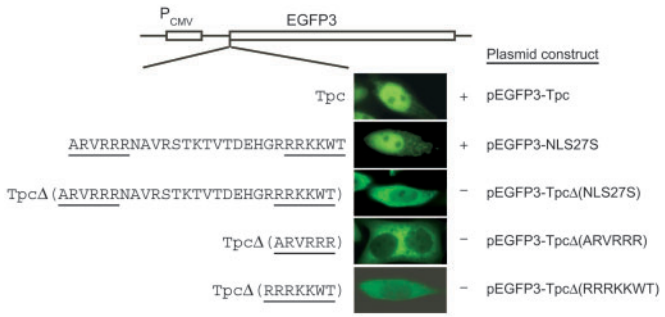

Figure 3. Bipartite NLS in Tpc. The sequences fused in-frame to EGFP3 are listed at the left. The two NLS clusters are underlined. The plasmid constructs are listed to the right. The nuclear localization results ('–', negative; '+', positive) are shown in the fluorescent microscopic images of a representative transfected cell.

a common [RK]{3,}?x{8,16}[RK]{4,}? pattern found in nearly 200 nuclear proteins (9). Tpc-EGFP3 fusion protein (pEGFP3-Tpc) was concentrated in the nuclei of transfected HeLa cells, indicating that Tpc was also capable of nuclear targeting (Fig. 3). A 27-aa polypeptide spanning the predicted NLS motif of Tpc (pEGFP3-NLS27S) was sufficient for targeting the fused EGFP3 to the nuclei of HeLa cells. This 27-aa polypeptide contains two separate putative basic aa clusters—ARVRRR and RRRKKWT. Deletion of this polypeptide from Tpc on pEGFP3-TpcΔ(NLS27S) blocked nuclear localization. Deletion of either of the basic clusters on pEGFP3-TpcΔ(ARVRRR) and pEGFP3-TpcΔ(RRRKKWT) also blocked nuclear localization (Fig. 3), confirming the bipartite nature of this NLS.

NLS-defective TP is functional in replicating linear plasmids

While the NLS motifs in the TPs function in nuclear localization, are they also important for replication of the linear replicons? To answer this question, linear plasmids were constructed following the procedure of Qin and Cohen (16). First, an *E. coli* plasmid pLUS986 was constructed that contained the tap^{Sco}-tpg^{Sco} operon and an autonomously replicating sequence (ARS) from linear plasmid pSLA2 (19) flanked by a pair of *S. lividans* chromosomal telomeres. Such replication-proficient sequences containing telomeres, when linearized at the bracketing adventitious DNA and introduced by transformation into *Streptomyces*, can generate functional linear plasmids (16,20). The Δ(tap^{Sli}-tpg^{Sli}) mutant MR04 of *S. lividans* (14) transformed with pLUS986 DNA that had been linearized by *Ase*I digestion (at the *E. coli* vector sequence) harbored an11.7 kb linear plasmid (designated pLUS986L) with the expected size and the expected *Sac*I restriction fragments (Fig. 4A). In this assay, linear plasmids that do not encode a functional TP and necessary accessory protein(s) cannot replicate in MR04, and only circular form may be found in transformants. Circular pLUS986 DNA possessing a single *Sac*I site would produce only a single (linear) *Sac*I fragment on digestion. Next, the K3A and R4A mutations in NLS (above) were each introduced into the tpg^{Sco} gene on pLUS986 to give rise to plasmids pLUS986(K3A) and

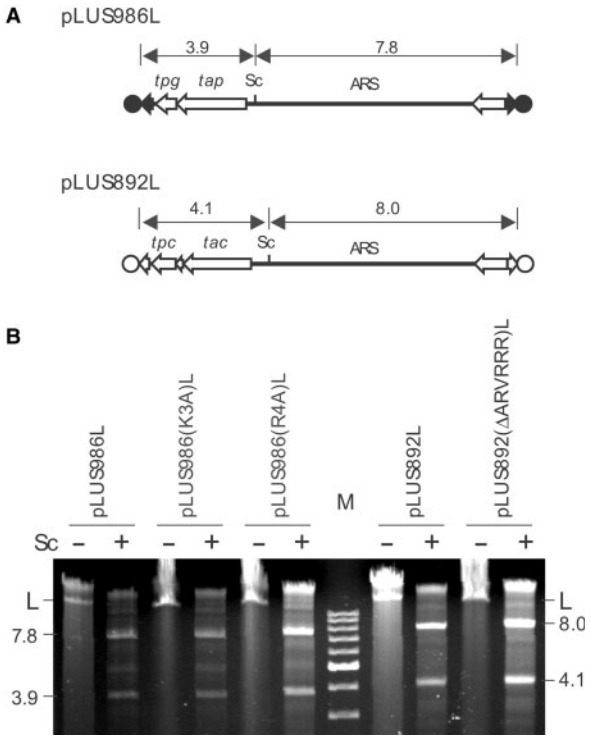

Figure 4. NLS-defective TPs are functional in replication. (**A**) Mini linear plasmids constructed. pLUS986L contains a pair of *S. lividans* telomeres (filled arrows) capped by the Tpg^{Sco} proteins (filled circles) encoded by the tpg^{Sco} it carries. pLUS986(K3A)L and pLUS986(R4A)L are pLUS986L containing the K3A and R4A mutation in the NLS of Tpg^{Sco} (Fig. 2B), respectively. pLUS892 contains a pair of SCP1 telomeres (terminal open arrows) capped by Tpc proteins (open circles) encoded by the *tpc* it carries (10). pLUS892(ΔARVRRR)L is a derivative of pLUS892L containing the Δ(ARVRRR) mutation in the NLS of Tpc (Fig. 3). The *Sac*I (Sc) restriction site and the size of the expected restriction fragments are indicated. ARS, autonomously-replicating sequence of pSLA2. (**B**) NLS-defective TPs cap mini linear plasmids. *S. lividans* strains (see text) were transformed by the circular progenitor of the five mini linear plasmids, pLUS986, pLUS986(K3A), pLUS986(R4A), pLUS892, and pLUS892(ΔARVRRR) that had been digested by *AseI*. Thiostrepton-resistant transformants were isolated, and DNA extracted from the transformants, digested by *Sac*I ('+', with digestion; '−'without digestion), and electrophoresed. The number and size of the *Sac*I fragments of the plasmids were as expected from the linear plasmids. Circular plasmids would give only a single *Sac*I fragment.

pLUS986(R4A), respectively. Transformation of MR04 with these plasmids linearized by *Ase*I also produced linear plasmids [designated pLUS986(K3A)L and pLUS986(R4A)L, respectively] with the expected size and *Sac*I fragments (Fig. 4B). These results indicate that a functional NLS in Tpg^{Sco} is not necessary for performing the end patching role.

Using the same procedure, the Δ(ARVRRR) mutation was introduced into the *tpc* gene on a mini linear plasmid, pLUS892L (10), which contained the essential *tac* and *tpc* gene pair, the pSLA2 ARS, and a pair of SCP1 telomeres. The resulting plasmid, pLUS892(ΔARVRRR)L, capped by the NLS-defective Tpc also replicated as a linear DNA in TK64 (Fig. 4B). These results indicated that a functional NLS in Tpc was also not essential for end patching.

The finding that two different classes of *Streptomyces* TPs contains different types of NLS motifs, which are functional in nuclear targeting but not required for replication, indicates that the nuclear localization functions have not emerged coincidentally, but have evolved convergently in two different systems for an identical biological role.

TPs carried covalently bound DNA into the nuclei

To determine whether TP may lead covalently linked DNA into the nuclei a linear plasmid, pLUS966-EGFP3L, was constructed that contained a pair of *S. lividans* telomeres and the EGFP3 gene under the control of the CMV promoter (Fig. 5A and B). In cells transfected with the Tpg^{Sco}-capped pLUS966-EGFP3L DNA, transient expression of EGFP3 was observed after 6 h in HeLa (Fig. 5C) and HEK 293T cells and reached a maximum of ~70% (Fig. 5D). In comparison, in transfection by the progenitor circular plasmid pLUS966-EGFP3 and proteinase K-treated pLUS966-EGFP3L DNA, fluorescent cells were seen after 12 h and reached a lower maximum (~60%).

The high efficiency of gene delivery by the Tpg^{Sco}-capped linear DNA might be due to either active nuclear targeting conferred by the TP or protection against cellular exonuclease attack. To resolve this issue, EGFP3 was placed on pLUS986(K3A)L, which was capped by an NLS-defective (K3A) Tpg^{Sco}. Transfection using the resultant plasmid produced similar results as the circular DNA and uncapped linear DNA in both HeLa and HEK 293T cells, i.e. later appearance and lower numbers of fluorescent transformants than the linear DNA capped by normal Tpg^{Sco} (Fig. 5D). This result indicated the higher efficiency of delivery by Tpg^{Sco} was mainly conferred by the active nuclear targeting function of its NLS.

DISCUSSION

New NLS motifs in Tpgs

We have demonstrated that the predicted monopartite NLS in Tpg^{Sco}, Tpg^{Sli}, Tpg^{Sav} and $Tpg^{pSLA2\text{-}L}$/$Tpg^{pSLA2\text{-}M}$ were functional in nuclear targeting either as separate domains or as part of the Tpg proteins. The predicted bipartite NLS motif on Tpc, which is found in nearly 200 nuclear proteins, is also functional in nuclear targeting, and the basic residues in both of its two clusters are essential for function.

The laboratory-observed nuclear targeting function of the NLSs in Tpgs and Tpc, which is not essential for the end-patching function, strongly suggests a role in nature. However, there is no nucleus in bacteria, hence what would be the biological function of these NLSs?

Nuclear transport of proteins bearing an NLS is mediated by the importin α/β heterodimer in eukaryotes. It is possible that Tpgs and Tpc interact with a similar system and perform an unknown biological process in *Streptomyces*. To investigate this possibility, we used the importins α and β sequences from human, mouse, *Drosophila melanogaster* and *Arabidopsis thaliana* as query in blastp and psi-blast searches against the

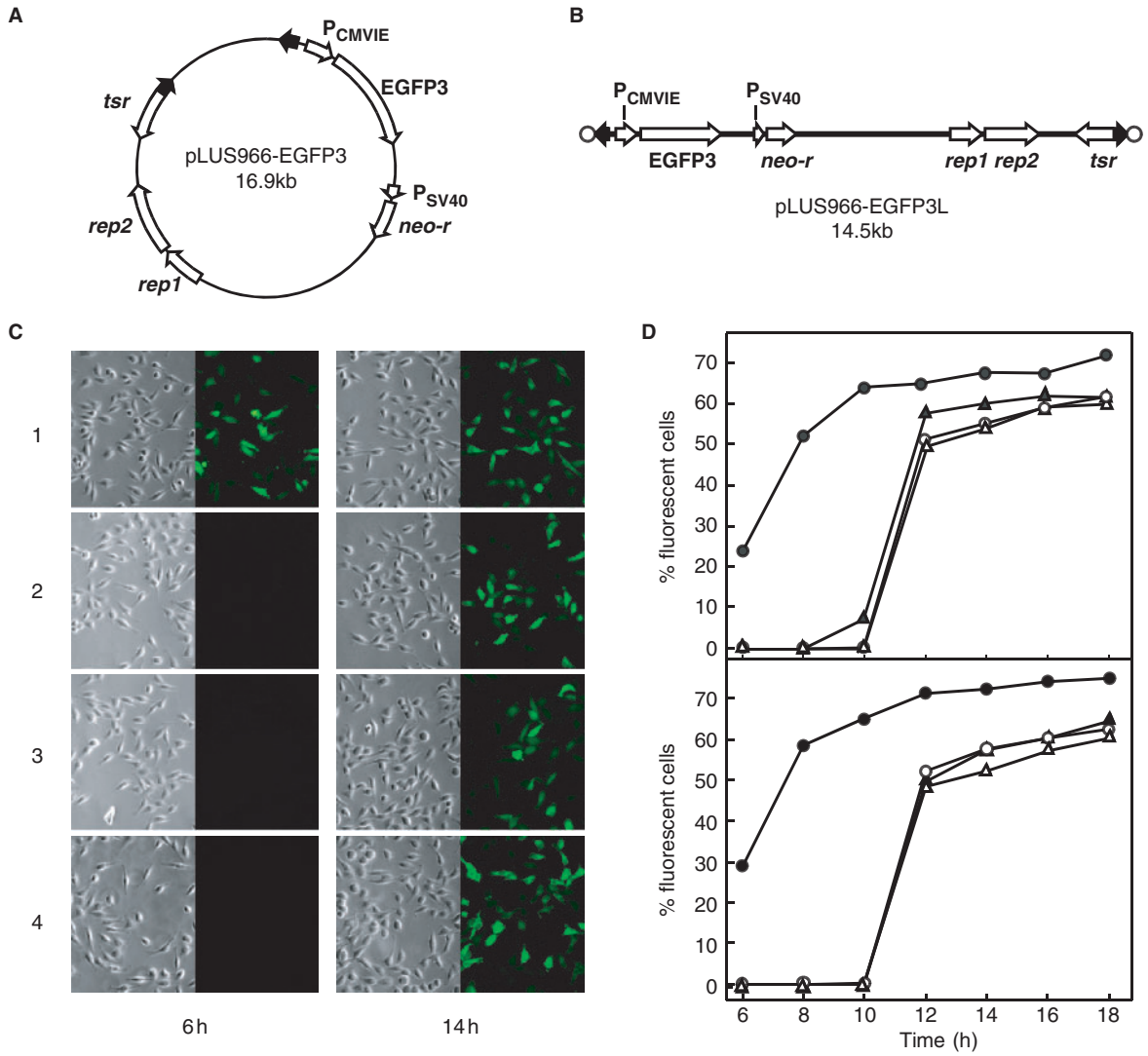

Figure 5. Delivery of TP-capped DNA into the nuclei. (**A**) pLUS966-EGFP3 containing a linear plasmid sequence with an EGFP3 gene under the control of the CMV promoter. The promoters and genes are indicated by the open arrows, and the *S. lividans* telomeres by filled arrows. *tsr*, thiostrepton resistance gene; *rep1* and *rep2*, replication genes of pSLA2 (19); *neo-r*, neomycin resistance gene. (**B**) Linear plasmid pLUS966-EGFP3-L capped by Tpg^{Sco} (filled circles) obtained by transformation of *S. lividans* by *Ase*I-linearized pLUS966-EGFP3 DNA. (**C**) Transfer and expression of EGFP3 in transfected cells. After transfection, fluorescent cells were photographed and counted under fluorescence microscope. Representative bright-field and fluorescence photographs of transfected HeLa cells at 6 and 14 h after transfection are shown. 1, pLUS966-EGFP3L DNA; 2, pLUS986(K3A)-EGFP3L DNA; 3, proteinase K-treated pLUS966-EGFP3L DNA; 4, pLUS966-EGFP3 DNA. (**D**) Comparison of efficiency of EGFP3 transfer and expression. Upper panel, HeLa cells; lower panel, HEK 293T cells. Fluorescent cells were counted after transfection. Filled circles, pLUS966-EGFP3L DNA; open circles, pLUS986(K3A)-EGFP3L DNA; open triangles, proteinase K-treated pLUS966-EGFP3L DNA; filled triangles, pLUS966-EGFP3 DNA.

S. coelicolor and *S. avermitilis* genomic databases. No significant homologous hit (*E*-values < 0.1) was found. Therefore, either the NLS motifs in Tpgs and Tpc interact with a heterologous system in *Streptomyces*, or they interact with a system outside of *Streptomyces*.

TP-mediated transfer is similar to T-DNA transfer

TP-mediated DNA transfer is analogous to the transfer of T-DNA by *Agrobacterium tumefaciens*. During conjugation with a plant cell, a Ti plasmid-encoded VirD2 protein in *A. tumefaciens* nicks at a border of the T-DNA sequence and remains covalently bound to the 5′ end. A rolling circle-type replication initiated at the nick, followed by a second nick, removes a single-strand stretch of T-DNA, which is transported into the plant cell. Inside the plant cells, the T-DNA is bundled by another T-DNA-encoded protein, VirE2, and led by the VirD2 protein into the plant nuclei, where integration takes place. VirD2 contains a monopartite and a bipartite NLS, which are required for nuclear targeting (21). VirD2 is attached to the T-DNA at a Tyr residue (22), whereas Tpg^{Sco} is attached to the *Streptomyces* DNA at a Thr residue (6).

Both being soil bacteria, agrobacteria and streptomycetes may have more genetic interactions than have been noted. First, Kelly and Kado (23) recently reported

that T-DNA may be transferred and integrated by *Agrobacterium* into the chromosome of *Streptomyces*. Second, the linear plasmid SLP2 of *S. lividans* contains a pair of homologs of two hypothetical genes, *ymg* and *yme*, which are present on an octopine-type Ti plasmid of *A. tumefaciens* in an identical arrangement (24). Codon usage analysis suggests that these two genes were horizontally acquired, perhaps through gene exchanges between a linear plasmid of *Streptomyces* and the Ti-plasmid of *A. tumefaciens*.

We propose that TP-capped linear DNA of *Streptomyces*, like the Ti plasmids of *A. tumefaciens*, is also involved in inter-kingdom gene transfer in soil. The target of such proposed transfer is not clear. *Streptomyces* species are highly abundant in soil, and the likely eukaryotic targets for transfer include plants and fungi. Such transfer, if real, would be of great evolutionary and ecological significance.

The existence of many genes of bacterial origin in the genomes of plants and other eukaryotes (25) has been suggested to result from horizontal gene transfer mediated by bacterial systems such as the Ti-plasmids (26). The TP-capped linear DNA system may also be involved in such inter-kingdom gene transfer.

The *Streptomyces* TP as a gene delivery tool

The Ti-plasmids of *Agrobacterium* are the most effective gene delivery tool in plant biotechnology (review in 27). They have also been adopted for gene transfer for other targets such as human nuclei (28) and mammalian mitochondria (29). In addition, NLSs have been used in various systems to aid non-viral gene delivery in human gene therapy (reviewed by 30–32). In these schemes, the positively charged NLSs are coupled with the negatively charged DNA, or covalently coupled with a carrier component/condensing agent or the phosphate–sugar backbone of the DNA.

The disadvantage of non-covalent coupling of NLS-containing peptides is that dissociation of the complex can occur during intracellular trafficking. The non-specific covalent coupling of NLS peptides to plasmid DNA does not markedly enhance nuclear uptake or increase reporter gene expression, while the covalent attachment of NLS peptides may inhibit intended gene expression. To prevent such inhibition, NLS peptides are coupled with specific locations in the plasmid DNA—primarily at the termini. Zanta *et al*. (33) reported a 10- to 1000-fold increase (depending on the cell types used) in gene expression using a linear DNA construct with a SV40-derived NLS peptide coupled with one of the hairpin ends compared with the DNA construct without the NLS peptide cap. However, similar attempts (for example, 34,35) met with little or no success.

In comparison to the existing NLS-aided gene delivery systems, the TP-capped *Streptomyces* replicons offer an efficient alternative. The TP caps offer both an active nuclear localization function and protection from cellular exonucleases, which is one of the gene delivery barriers (36). TP capping is biological and complete, and does not require elaborate physical and chemical procedures. Only standard molecular cloning techniques in *E. coli* and *Streptomyces* are required for production of TP-capped linear DNA for transfection. The size of inserts is limited by that of the cloning systems. *E. coli* plasmid vectors generally can accommodate inserts of tens of kb, and linear *Streptomyces* plasmids reach 1 Mb. Such a promising gene delivery system requires vigorous study and development. Notably, while this article was being prepared, the idea of using TP-capped linear DNA as 'a potential new strategy for assembly of synthetic therapeutic gene vector' was proposed by Tolmachov and Coutelle (37).

ACKNOWLEDGEMENTS

We thank David Hopwood for critical reading of the manuscript and suggestions for improvement. This study was funded by research grants from National Science Council, R. O. C. to C.-H.H. (NSC94-2311-B027-002) and C.W.C. (NSC94-2321-B010-005, NSC94-2321-B010-002), a grant (Aim for the Top University Plan) from the Ministry of Education, R. O. C. to C.W.C., and a National Health Research Institute postdoctoral fellowship award (PD91006N) to C.-H.H. Funding to pay the Open Access publication charges for this article was provided by National Science Council, R. O. C.

Conflict of interest statement. None declared.

REFERENCES

1. Chen,C.W. (1996) Complications and implications of linear bacterial chromosomes. *Trends Genet*., **12**, 192–196.
2. Chang,P.C. and Cohen,S.N. (1994) Bidirectional replication from an internal origin in a linear *Streptomyces* plasmid. *Science*, **265**, 952–954.
3. Bao,K. and Cohen,S.N. (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev*., **15**, 1518–1527.
4. Yang,C.-C., Huang,C.-H., Li,C.-Y., Tsay,Y.-G., Lee,S.-C. and Chen,C.W. (2002) The terminal proteins of linear *Streptomyces* chromosomes and plasmids: a novel class of replication priming proteins. *Mol. Microbiol*., **43**, 297–305.
5. Bao,K. and Cohen,S.N. (2003) Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev*., **17**, 774–785.
6. Yang,C.C., Chen,Y.H., Tsai,H.H., Huang,C.H., Huang,T.W. and Chen,C.W. (2006) *In vitro* deoxynucleotidylation of the terminal protein of *Streptomyces* linear chromosomes. *Appl. Environ. Microbiol*., **72**, 7959–7961.
7. Hodel,M.R., Corbett,A.H. and Hodel,A.E. (2001) Dissection of a nuclear localization signal. *J. Biol. Chem*., **276**, 1317–1325.
8. LaCasse,E.C. and Lefebvre,Y.A. (1995) Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res*., **23**, 1647–1656.
9. Cokol,M., Nair,R. and Rost,B. (2000) Finding nuclear localization signals. *EMBO Rep*, **1**, 411–415.
10. Huang,C.-H., Tsai,H.-H., Tsay,Y.-G., Chien,Y.-N., Wang,S.-L., Cheng,M.-Y., Ke,C.-H. and Chen,C.W. (2007) The telomere system of the *Streptomyces* linear plasmid SCP1 represents a novel class. *Mol. Microbiol*., **63**, 1710–1718.
11. Kieser,T., Bibb,M., Buttner,M.J., Chater,K.F. and Hopwood,D.A. (2000) *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich.

12. Sambrook,J., MacCallum,P. and Russel,D. (2001) *Molecular Cloning*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
13. Hopwood,D.A., Kieser,T., Wright,H.M. and Bibb,M.J. (1983) Plasmids, recombination, and chromosomal mapping in *Streptomyces lividans* 66. *J. Gen. Microbiol.*, **129**, 2257–2269.
14. Redenbach,M., Flett,F., Piendl,W., Glocker,I., Rauland,U., Wafzig,O., Leblond,P. and Cullum,J. (1993) The *Streptomyces lividans* 66 chromosome contains a 1 Mb deletogenic region flanked by two amplifiable regions. *Mol. Gen. Genet.*, **241**, 255–262.
15. Fujii,T., Tamura,K., Copeland,N.G., Gilbert,D.J., Jenkins,N.A., Yomogida,K., Tanaka,H., Nishimune,Y., Nojima,H. and Abiko,Y. (1999) Sperizin is a murine RING zinc-finger protein specifically expressed in haploid germ cells. *Genomics*, **57**, 94–101.
16. Qin,Z. and Cohen,S.N. (1998) Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. *Mol. Microbiol.*, **28**, 893–904.
17. Mochizuki,S., Hiratsu,K., Suwa,M., Ishii,T., Sugino,F., Yamada,K. and Kinashi,H. (2003) The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol. Microbiol.*, **48**, 1501–1510.
18. Cunningham,B.C. and Wells,J.A. (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, **244**, 1081–1085.
19. Chang,P.C., Kim,E.S. and Cohen,S.N. (1996) *Streptomyces* linear plasmids that contain a phage-like, centrally located, replication origin. *Mol. Microbiol.*, **22**, 789–800.
20. Shiffman,D. and Cohen,S.N. (1992) Reconstruction of a *Streptomyces* linear plasmid replicon from separately cloned DNA fragment: existence of a cryptic origin of circular replication within the linear plasmid. *Proc. Natl Acad. Sci. USA*, **89**, 6129–6133.
21. Tinland,B., Koukolikova-Nicola,Z., Hall,M.N. and Hohn,B. (1992) The T-DNA-linked Vird2 protein contains two distinct functional nuclear localization signals. *PNAS*, **89**, 7442–7446.
22. Vogel,A.M. and Das,A. (1992) Mutational analysis of *Agrobacterium tumefaciens* virD2: tyrosine 29 is essential for endonuclease activity. *J. Bacteriol.*, **174**, 303–308.
23. Kelly,B.A. and Kado,C.I. (2002) *Agrobacterium*-mediated T-DNA transfer and integration into the chromosome of *Streptomyces lividans*. *Mole. Plant Pathol.*, **3**, 125–134.
24. Huang,C.H., Chen,C.Y., Tsai,H.H., Chen,C., Lin,Y.S. and Chen,C.W. (2003) Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol. Microbiol.*, **47**, 1563–1576.
25. Brown,J.R. (2003) Ancient horizontal gene transfer. *Nat. Rev. Gen.*, **4**, 121–132.
26. Broothaerts,W., Mitchell,H.J., Weir,B., Kaines,S., Smith,L.M.A., Yang,W., Mayer,J.E., Roa-Rodriguez,C. and Jefferson,R.A. (2005) Gene transfer to plants by diverse species of bacteria. *Nature*, **433**, 629–633.
27. Gelvin,S.B. (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.*, **67**, 16–37.
28. Kunik,T., Tzfira,T., Kapulnik,Y., Gafni,Y., Dingwall,C. and Citovsky,V. (2001) Genetic transformation of HeLa cells by *Agrobacterium*. *PNAS*, **98**, 1871–1876.
29. Yoon,Y.G. and Koob,M.D. (2005) Transformation of isolated mammalian mitochondria by bacterial conjugation. *Nucleic Acids Res.*, **33**, e139.
30. Munkonge,F.M., Dean,D.A., Hillery,E., Griesenbach,U. and Alton,E.W. (2003) Emerging significance of plasmid DNA nuclear import in gene therapy. *Adv. Drug Deliv. Rev.*, **55**, 749–760.
31. van der Aa,M.A., Mastrobattista,E., Oosting,R.S., Hennink,W.E., Koning,G.A. and Crommelin,D.J. (2006) The nuclear pore complex: the gateway to successful nonviral gene delivery. *Pharmaceut. Res.*, **23**, 447–459.
32. Cartier,R. and Reszka,R. (2002) Utilization of synthetic peptides containing nuclear localization signals for nonviral gene transfer systems. *Gene Ther.*, **9**, 157–167.
33. Zanta,M.A., Belguise-Valladier,P. and Behr,J.P. (1999) Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. *Proc. Natl Acad. Sci. USA*, **96**, 91–96.
34. Van der Aa,M.A., Koning,G.A., d'Oliveira,C., Oosting,R.S., Wilschut,K.J., Hennink,W.E. and Crommelin,D.J. (2005) An NLS peptide covalently linked to linear DNA does not enhance transfection efficiency of cationic polymer based gene delivery systems. *J. Gene Med.*, **7**, 208–217.
35. Tanimoto,M., Kamiya,H., Minakawa,N., Matsuda,A. and Harashima,H. (2003) No enhancement of nuclear entry by direct conjugation of a nuclear localization signal peptide to linearized DNA. *Bioconjugate Chem.*, **14**, 1197–1202.
36. Lechardeur,D., Sohn,K.J., Haardt,M., Joshi,P.B., Monck,M., Graham,R.W., Beatty,B., Squire,J., O'Brodovich,H. and Lukacs,G.L. (1999) Metabolic instability of plasmid DNA in the cytosol: a potential barrier to gene transfer. *Gene Ther.*, **6**, 482–497.
37. Tolmachov,O. and Coutelle,C. (2007) Covalent attachment of multifunctional chimeric terminal proteins to 5′ DNA ends: a potential new strategy for assembly of synthetic therapeutic gene vectors. *Med. Hypothes.*, **68**, 328–331.
38. Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol.*, **21**, 526–531.
39. Hosted,T.J., Wang,T. and Horan,A.C. (2004) Characterization of the *Streptomyces lavendulae* IMRU 3455 linear plasmid pSLV45. *Microbiol.*, **150**, 1819–1827.
40. Zhang,R., Yang,Y., Fang,P., Jiang,C., Xu,L., Zhu,Y., Shen,M., Xia,H., Zhao,J., Chen,T. *et al.* (2006) Diversity of telomere palindromic sequences and replication genes among *Streptomyces* linear plasmids. *Appl. Environ. Microbiol.*, **72**, 5728–5733.
41. Kinashi,H., Fujii,S., Hatani,A., Kurokawa,T. and Shinkawa,H. (1998) Physical mapping of the linear plasmid pSLA2-L and localization of the *eryAI* and *actI* homologs. *Biosci. Biotechnol. Biochem.*, **62**, 1892–1897.