Mutational Analysis of the Terminal Protein Tpg of Streptomyces Chromosomes: Identification of the Deoxynucleotidylation Site

Chien-Chin Yang1*, We-Chi Sun1, Wan-Yu Wang1, Chi-Hung Huang2, Fang-Shy Lu1, Shu-Min Tseng1, Carton W. Chen3*

1 Department of Chemistry, Chung-Yuan Christian University, Chung-li, Taiwan, 2 Institute of Biotechnology, National Taipei University of Technology, Taipei, Taiwan, 3 Department of Life Sciences, Institute of Genome Sciences, National Yang-Ming University, Shih-Pai, Taipei, Taiwan

Abstract

The linear chromosomes and linear plasmids of Streptomyces are capped by terminal proteins (TPs) covalently bound to the 5' ends of the DNA. The TPs serve as primers for DNA synthesis that patches in the single-stranded gaps at the telomeres resulting from the bi-directional replication ('end patching'). Typical Streptomyces TPs, designated Tpgs, are conserved in sequence and size (about 185 amino acids), and contain a predicted helix-turn-helix domain and a functional nuclear localization signal. The Tpg-encoding gene (tpg) is often accompanied by an upstream gene tap that encodes an essential telomere-associating protein. Five lone tpg variants (not accompanied by tap) from various Streptomyces species were tested, and three were found to be pseudogenes. The lone tpg variant on the SLP2 plasmid, although functional, still requires the presence of tap on the chromosome for end patching. Using a combination of in vitro deoxynucleotidylation, physical localization, and genetic analysis, we identified the threonine at position 114 (T114) in Tpg of Streptomyces lividans chromosome as the deoxynucleotidylated site. Interestingly, T114 could be substituted by a serine without destroying the priming activity of Tpg in vitro and in vivo. Such T114S substitution is seen in and a number of pseudogenes as well as functional Tpgs. T114 lies in a predicted coil flanked by two short helixes in a highly hydrophilic region. The location and structural arrangement of the deoxynucleotidylated site in Tpg is similar to those in the TPs of phage ø29 and adenoviruses. However, these TPs are distinct in their sequences and sizes, indicating that they have evolved independently during evolution. Using naturally occurring and artificially created tpg variants, we further identified several amino acid residues in the N-terminus and the helix-turn-helix domain that were important for functionality.

Introduction

Soil bacteria of the genus Streptomyces possess linear chromosomes and linear plasmids that are capped by terminal proteins (TPs) covalently bound to the 5' ends of the DNA [1,2]. Replication of these linear replicons is accomplished in two steps: (i) bidirectional replication initiated from an internal origin, which results in single-stranded gaps at the 5' end; and (ii) patching of the single-stranded gaps by TP-primed DNA synthesis [3]. The TPs remain covalently attached to the telomeres of the linear Streptomyces replicons after replication.

The patching TP-primed DNA synthesis in Streptomyces differs from the replicative TP-primed replication involved in the replication of adenoviruses [4] and phage ø29 [5], which have been extensively studied. In end patching, only about 300 nt of single-stranded gaps need to be filled [3], whereas replicative TP-primed synthesis duplicates the whole replicons end-to-end.

In Streptomyces, most TPs are highly conserved in sequences and size (about 185 amino acids). These conserved TPs are designated archetypal TPs, and are encoded by tpg gene, which typically lie downstream of a tap gene in the same operon [6,7]. Tap is also essential for replication of linear Streptomyces replicons. It binds specifically to a secondary structure formed by the single-stranded 5' overhang during replication, and presumably recruits Tpg to the telomere location for the end patching reaction [8].

The tap-tpg operon generally lies in the terminal region of the Streptomyces chromosomes. A few linear plasmids also contain the tap-tpg operon or a lone tpg homolog. In addition, apparent tpg pseudogenes are found in sequenced Streptomyces genomes. The first lone tpg homolog was discovered in the 15.4-kb terminal sequences shared by the S. lividans chromosome and the right end of SLP2 [9]. This homolog, designated tpgSLP2.38, exhibited atypical codon usage and encoded a product (183-amino acid) with a number of deletions and insertions. It was presumably a pseudogene. Subsequently, more putative tpg pseudogenes have been found in the linear plasmids and chromosomes in many Streptomyces species, for example, S. avermitilis [10], S. ambofaciens [11], S. griseus [12,13], and S. violaceoruber [14].
The N-terminal halves of Tpgs contain a number of predicted functional and structural domains [7]: (i) a helix-turn-helix (HTH) domain overlapping a segment similar to part of the thumb domain of HIV reverse transcriptase; (ii) a mono-partite nuclear localization signal (NLS) that is functional in targeting nuclei of human [15] and plant [16] cells, but is not essential for end patching [15]; (iii) an amphiphilic β-strand that may be involved in protein-protein interactions or protein-membrane interactions. Interactions between the Streptomyces telomeres in vivo have been recently demonstrated [17].

SCPI, a 350-kb linear plasmid in Streptomyces coelicolor A3(2), encodes its own TP, Tpc [18]. Tpc shares no homology with Tpgs, and is significantly larger (259 amino acids). The coding gene, tpc, is also downstream from a gene tac that is essential for replication of SCPI. Tpc, like Tpgs, contains an HTH domain in the N-terminal region and a functional bi-partite NLS in the central region [15].

Recently, Yang et al. [19] demonstrated in vitro deoxynucleotidylation of Tpgs of S. lividans, in which dCMP, the first nucleotide at the 5' ends of the S. lividans chromosome, is specifically covalently attached to Tpgs in the in vitro reactions. The dCMP was attached to a Thr residue of Tpgs. There are 11 Thr residues substituted by a Ser without destroying the deoxynucleotidylation of Tpgs in vivo. Examination of five lone TPG homologs identified three of them to be pseudogenes. Interestingly all these pseudogenes have the T114S substitution, and therefore the defect must lie somewhere else. Moreover, we identified several amino acid residues in the T114 region [15].

In this study, combining biochemical and genetic approaches, we identified T114 at the C-terminus of Tpgs as the attachment site of the telomere DNA. Interestingly, this residue may be substituted by a Ser without destroying the deoxynucleotidylation function of Tpg in vitro and the priming function in vivo. Examination of five lone tpg homologs identified three of them to be pseudogenes. Interestingly all these pseudogenes have the T114S substitution, and therefore the defect must lie somewhere else. Moreover, we identified several amino acid residues in the neighborhood of T114, the N-terminus, and the helix-turn-helix (HTH) domain of Tpgs that were important for the functionality of Tpgs.

Materials and Methods

Bacterial Cultures and Molecular Manipulations

Genetic manipulations of E. coli and Streptomyces were performed according to the methods of Kieser et al. [22]. E. coli BL21(DE52) (Stratagene) and S. lividans MR04 [23] were used for expression of protein. pLUS980 plasmid containing a 4.5-kb BglII fragment spanning the tap-tpg operon from S. coelicolor cosmID11 [24] was obtained from Chia-Hui Ke.

Protein Expression in E. coli BL21(DE52) and Purification

All the proteins used here were constructed in vector pet-15b (Novagen) where His6 is tagged to the N-terminal, or in pet-22b to the C-terminal with the purpose of purification, or vector pSET A (Invitrogen) without His6-tagging. Protein expression was induced by IPTG (isopropyl-b-D-1-thiogalactopyranoside). The cells were centrifuged and sonicated in A8 buffer (50 mM Tris-HCl pH8, 20 mM NaCl, 10% glycerol) and the His6-tagged protein purified using nickel-bound resin (Biomab) according to the protocols provided by the manufacturer. Insoluble His6-tagged proteins were purified in denatured state using nickel-bound resin and eluted in Elution buffer containing 6 M urea, refolded in Elution buffer by step-wise dialysis, and stored at −20°C in Elution buffer containing 5% DMSO (dimethyl sulfoxide). The concentrations of proteins were determined by the Bradford method [25]. Tpg without the His6-tag was purified by electrophoresis in 12% SDS-PAGE as described previously [19].

In vitro Deoxynucleotidylation and Purification of the Labeled Tpg

In vitro deoxynucleotidylation of untagged Tpg in cell extracts with alpha-[32P]-dCMP was based on the procedure described previously [19]. The labeled proteins were purified by SDS-PAGE or by immuno-precipitation using protein A-Sepharose and anti-Tpg antibody. With this procedure, Tpg with various lengths of deoxynucleotides were recovered.

An alternative procedure employed purified His6-TpgSli, His6-TapSli, and His6-tagged DinB1 DNA polymerase (product of gene SCO1330 of S. coelicolor). A 20-μl mixture, containing 0.3 μg each of His6-TpgSliSli, His6-TapSli, and His6-DinB1, 0.1 pmole of denatured 99-bp telomere DNA of S. lividans chromosome (prepared by PCR), 2.5 mM ATP, 10 mM Tris-HCl (pH7.5), 7 mM MgCl2, 0.1 mM dithiothreitol, and 0.17 μM alpha-[32P]-dCTP was incubated at 25°C for 20 min. The reaction was stopped by trichloroacetic acid (TCA) precipitation in the presence of 1 μg of yeast tRNA as carriers. The labeled TP was eluted from the gel by cracking and soaking three times in 200 μl of 50 mM ammonium bicarbonate, 5% β-mercaptoethanol and 0.1% SDS, and precipitated with TCA.

Phosphoamino Acids Analysis

The procedures of Pargellis et al. [26] and Garcia et al. [27] were followed with minor modifications. Purified alpha-[32P]-dCMP-labeled Tpg protein was treated in 50 μl of 5.7 N HCl at 110°C for 1–2 h. The hydrolysates were dried and resuspended in 10 μl of water. A sample containing about 50 cpm was mixed with non-radioactive standards (1 μg each phosphoserine, phosphothreonine, and phosphotyrosine; Sigma), and spotted on a cellulose thin layer plate (Merck) and subjected to two-dimensional electrophoresis in the Hunter Thin Layer Electrophoresis System (C.B.S. Scientific Company). The first dimension was carried out at pH 1.9 at 1.5 kV for 20 min and the second dimension in pH 3.5 at 1.5 kV for 16 min. The internal standards (marked with dotted circles) were visualized by spraying with 0.25% ninhydrin in acetone, and the radioactivity was imaged by autoradiography.

Fragmentation of the TP-dCMP Adduct

Alpha-[32P]-dCMP-labeled TpgSliSli was isolated from a polyacrylamide gel by elution and precipitation with TCA, and cleaved with CNBr (Sigma) according to the published procedure [28] with minor modifications. The reaction products were separated on a 16% polyacrylamide gel formulated for the analysis of small peptides [29], followed by autoradiography without drying.

For preparative digestion with endoprotease LysC (Roche), alpha-[32P]-dCMP-labeled TpgSliSli was isolated using protein A-Sepharose beads, precipitated with TCA, and cleaved with LysC at 25°C for 20 h in a 50-μl solution containing [32P]-labeled TP-dCMP, 1 μg enzyme, 25 mM Tris-HCl (pH 8.8), 1 mM EDTA and 0.1 M or 0.3 M of urea, and stopped by vacuum drying. The products were analysis by SDS-PAGE and autoradiography.

Mass Spectrometric Analysis

Native dCMP-TP complexes for the purpose of mass spectrometric analysis were gained from linear plasmid pLUS980L harbored in Streptomyces through four steps of purification as described [19], whereas TpgC was just separated by 12% SDS-PAGE because it locates alone against a clear background. Mass spectrometric analysis was operated at either Genomics...
Center, National Yang Ming University (ESI and MALDI) or Molecular Medicine Research Center, Chang-Gung University (MALDI).

**Construction of Mutated tpg Genes**

Two-step PCR was employed to create mutations in \( \text{tg} \). The altered sequences were confirmed by sequencing before being used.

**Construction of Linear Plasmids and Test for Linearity of the DNA**

Linear plasmids were constructed following the general procedure of Qin et al. [30]. The linearity of the plasmid DNA in the transformants was confirmed by restriction digestion and Southern hybridization. To facilitate manipulation of the \( \text{tg} \) sequence, an \( \text{Ndel} \) site was added immediately upstream of and including the initiation codon of \( \text{tg} \) on pLUS980. The resulting plasmid was designated pLUS980(Nd). Linear plasmids were generated from these plasmids and their derivative using the previously described procedure [30,31], i.e., linearization of the plasmid DNA by \( \text{Asel} \) digestion in the \( \text{E. coli} \) vector sequence followed by transformation of \( \text{Streptomyces} \).

Various \( \text{tg} \) homologs, including the putative pseudogenes, were generated by PCR and used to substitute \( \text{tg} \) on pLUS980 or pLUS980(Nd) through a series of restriction, subcloning, and ligation manipulations such that \( \text{tg} \) in the end products was precisely replaced by these sequences without any alteration of the upstream translation start codon.

To test the linearity of plasmid DNA, genomic DNA was isolated from \( \text{Streptomyces} \) host (mostly MR04) harboring pLUS980, pLUS980(Nd), or their derivatives, digested with \( \text{SacI} \) or \( \text{Spel} \) (which cuts uniquely in these plasmids), electrophoresed in agarose gel, and subjected to mass spectrometric analysis by ESI or MALDI ionization. An apparent molecular weight of about 9 kD (Fig. 1B). The largest \( \text{SacI} \) polypeptide fragment, \( \text{A}_{91} \sim \text{K}_{156} \), of Tpg\( ^{\text{Si}} \) was shown to have a molecular weight of 7.0 kD. We interpreted the increased apparent molecular weight to result from the attachment of the dCMP residue(s). A \( \text{K}_{156} \) was a subset of \( \text{L}_{45} \sim \text{I}_{183} \), which contained seven Thr residues. These results suggested that the deoxynucleotidylated Thr residue lay in the C-terminal region of Tpg\( ^{\text{Si}} \).

To narrow down the location of the deoxynucleotidylated Thr in the C-terminal region, we employed mass spectrometry on trypsin-digested Tpg\( ^{\text{Si}} \) to identify the trypsin fragment that contained the deoxynucleotidylated Thr by its altered molecular weight. Tpg\( ^{\text{Si}} \)-capped DNA was isolated by guanidine HCl-CsCl density gradient centrifugation followed by glass bead binding, and the DNA was trimmed by benzonase, an endonuclease that hydrolyzes DNA into 2- to 3-bp fragments (according to the manufacturer’s specification). The resulting Tpg\( ^{\text{Si}} \) was isolated by SDS-PAGE, digested with trypsin, and subjected to mass spectrometric analysis by ESI or MALDI ionization. It was anticipated that, if a fragment contained the deoxynucleotidylated Thr, the fragment with the native molecular weight would be absent from the sample (Fig. 1C).

In the C-terminal region, the only expected trypsin fragment that was not detected by either ESI or MALDI was the \( \text{L}_{105} \sim \text{R}_{118} \) fragment. A fragment with a mass corresponding to that of dCMP\( _2 \text{L}_{105} \sim \text{R}_{118} \) was detected in the MALDI spectrum. In contrast, trypsin digestion of Tpg\( ^{\text{Si}} \) protein produced in \( \text{E. coli} \) (presumably without deoxynucleotidylase) gave rise to the expected \( \text{L}_{105} \sim \text{R}_{118} \) trypsin fragment in both ESI and MALDI spectra. These results indicated that the \( \text{L}_{105} \sim \text{R}_{118} \) fragment contained the deoxynucleotidylated Thr. The \( \text{L}_{105} \sim \text{R}_{118} \) fragment lies in the \( \text{A}_{91} \sim \text{K}_{156} \) LysC fragment and contains two Thr residues, T108 and T114.

**Identification of the Deoxynucleotidylation Site at T114**

Site-directed mutagenesis was subsequently employed to replace T108 and T114 with different residues in \( \text{tg} \) to test the functionality of the resulting mutant Tpg\( ^{\text{Si}} \). To do this, the mutant \( \text{tg} \) gene was placed on pLUS980 or its derivative pLUS980(Nd) (Fig. 2A). These two plasmids contained a linear plasmid sequence consisting of a pair of telomeres of the \( \text{S. lividans} \) chromosome and the deoxynucleotidylated Thr residue.

The **Prediction of Hydropathy of Proteins**

Protein hydropathy plots were conducted using the Kyte-Doolittle algorithm [34] implemented at the University of Virginia FASTA Server (http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?tm = misc1).

**Results**

**Localization of the Deoxynucleotidylation Site in a C-terminal Fragment of Tpg\( ^{\text{Si}} \)**

*In vitro* deoxynucleotidylated TPs typically selects specifically the nucleotide corresponding to the first nucleotide at the 5' end of the replicons, such as dCMP for adenovirus-2 [35], dAMP for phage \( \phi \text{29} \) [21], dCMP for linear plasmid pAL105 of *Arthrobacter nitroguajaciensis* [36], dGMP for linear plasmid SCP1 of *Streptomyces* (Tsai, H.-H, unpublished data), and dCMP for the *S. lividans* chromosome [19]. In the case of \( \phi \text{29} \), the *in vitro* deoxynucleotidyl-
capped telomeres, designated pLUS980L and pLUS980(Nd)L, respectively (Fig. 2B).

*S. lividans* MR04, in which the *tapSli-tpgSli* operon was deleted along with large stretches of terminal DNA from the chromosome [23], was chosen as the host. If the *tpg* variant on pLUS980 and pLUS980(Nd) were defective, transformation of MR04 using the *Ase*I-linearized plasmid DNA would produce either no transformants, or a few transformants harboring only circular plasmids. The latter would result from the plasmids that had escaped *Ase*I digestion or linear fragments that had circularized in the transformants [37]. The topology of the plasmids in the transformants was determined by restriction digestion. For example, *Spe*I, which cuts singly in pLUS980 and pLUS980(Nd) DNA, would produce two fragments of 5.1 and 8.0 kb from the linear plasmid DNA, but would produce a single larger fragment from the circular plasmid DNA (Fig. 2C).

Firstly, T108 and T114 were individually mutated to Ser in TpgSli. Interestingly, both the resulting plasmids, pLUS980-T108S (Fig. S1 in Supporting Information) and pLUS980-T114S (Fig. 2C, lane 6), could replicate in linear form in MR04, indicating that neither of the mutations inactivated TpgSli. Similarly, the T108S and T114S mutations were individually created in *tpg*SLP2.19, and again these mutations did not inactivate TpgSLP2.19 (data not shown). These results suggested that either none of these Thr residues was the deoxynucleotidylation site, or the deoxynucleotidylated Thr could be substituted by a Ser without losing its function. Substitutions of four other Thr residues at 101, 123, 143 and 176 positions in the C-terminal region also did not destroy the ability of Tpg to support replication of the linear plasmid (Fig. S2 in Supporting Information).

Next, T108 and T114 of TpgSli were substituted by three other residues – Ala, Cys, and Tyr. None of these substitutions at T108 inactivated TpgSli. In contrast, the substitutions at T114 gave rise to transformants at very low frequencies (about three orders of magnitude lower), which harbored only circular plasmids (Fig. 2C, lanes 2, 4, 5). Substitution of T114 by a His, which might provide an amino group to form a covalent bond with the nucleotide [38], also failed to support replication of the linear plasmid (Fig. 2C, lane 3). These results eliminated the role of T108 as the site of deoxynucleotidylation, leaving T114 as the final candidate.

If T114 was the deoxynucleotidylation site in vivo, the Ser residue that substituted it in the T114S mutants would be expected to be deoxynucleotidylated, and this was tested in an *in vitro* deoxynucleotidylation assay using His-tagged TpgSli and TpgSli-T114S proteins produced in *E. coli*. In the *in vitro* deoxynucleotidylation reaction, approximately equal weights of TpgSli and TpgSli-T114S were labeled by radioactive dCMP. The alpha-[32P]-dCMP labeled TpgSli and TpgSli-T114S were subjected to acid hydrolysis, and chromatography on cellulose thin layer plate [19]. The results showed that the labeled dCMP was attached to a Ser residue on TpgSli-T114S (Fig. 3 middle panel), while it was attached to a Thr residue of TpgSli (Fig. 3, left panel). The efficiency of deoxynucleotidylation for TpgSli-T114S was approximately 30% of that for TpgSli (Fig. 3, right panel). These results...
supported the idea that T114 was the deoxynucleotidylation site in vivo and that it could be functionally substituted by Ser albeit possibly with a reduced efficiency of deoxynucleotidylation.

In the TpgSli sequence, T114 is followed closely by two negatively charged Asp residues (D115 and D117 in TpgSli). In the ø29 TP, the deoxynucleotidylated S232 is also flanked by two negatively charged Asp (D231) and Glu (E233). Introduction of a D115A mutation destroyed the Tpg function, whereas a D117A mutation had no effect on the Tpg function (Fig. S3B, Supporting Information). Interestingly, D115 could be substituted by a Glu (D115E mutation) but not by a Asn (D115N mutation) without destroying the Tpg function (Fig. S3B, Supporting Information), indicating that a negative charge there is important.

The residue upstream of T114 is also a Ser. A S113A mutation had no effect on the Tpg functionality (Fig. S1 in Supporting Information).

Testing Tpg Variants for Support of Replication of Linear Plasmids

Examination of the aa sequences of 17 Tpg homologs (including the conceptual translation products of the putative pseudogenes; Fig. 4) revealed that 12 contained T114 and five contained S114. Of the five S114-containing tpg homologs, tpgSli, which encodes a product of only about one half length (95 aa), is most likely a pseudogene. The other four S114-containing homologs, tpgSLP2.38, tpgpFRL1.6, tpgSAP1_11, and tpgpSV2.102, along with the T114-containing tpgSLP2.19 were tested for their ability to support the replication of linear plasmids. All these tpg homologs are present on a linear plasmid unaccompanied by a tap homolog. Of these, tpgSLP2.38, tpgSAP1_11, and tpgpSV2.102 could not support replication of linear plasmids in MR04 (Fig. S4 in Supporting Information).
indicating that they were indeed pseudogenes. $\text{tpg}_{\text{SLP2.19}}$ and $\text{tpg}_{\text{pFRL1.6}}$ could support replication of linear plasmids in MR04 (Fig. S4 in Supporting Information). It was probably not surprising that these two genes are functional, because the encoded Tpg homologs are relatively conserved and the codon usage is typical for *Streptomyces* (high G/C at the third positions of the codons). However, $\text{tpg}_{\text{pFRL1.6}}$ contains a 13-amino acid extension at the N-terminus. Interestingly, the pseudogene product $\text{tpg}_{\text{SAP1}_1}$ is highly similar to $\text{tpg}_{\text{pFRL1.6}}$ in sequence including the 13-amino acid extension at the N-terminus. The defect of this product was likely to be due to one or more amino acid substitutions elsewhere.

$\text{tpg}_{\text{SLP2.19}}$, when placed downstream of $\text{tap}$ on pLUS980L, was presumably transcribed from the same promoter as $\text{tap}$. To test whether the lone $\text{tpg}_{\text{SLP2.19}}$ on SLP2 was expressed, and, if so, whether it alone was sufficient for supporting the replication of SLP2, attempts were made to use SLP2tsr, an SLP2 derivative with $\text{tsr}$ inserted in Tn$^{4811}$[39], to transform MR04 ($\text{tap-tpg}$) and ZX7 (parent of MR04, $\text{tap-tpg}^+$). Thior transformants were readily obtained in ZX7, but not in MR04. This indicated that either $\text{tpg}_{\text{SLP2.19}}$ was not expressed on SLP2, or the participation of $\text{tap}$ was required as shown in the replication of the *Streptomyces* chromosomes [8]. To check this, $\text{tap}_{\text{Sli}}$ was inserted into the $\varnothing\text{C31}$ att site on the MR04 chromosome using the integrative plasmid pSET152 [40]. The resulting strain, CK03, could then be transformed successfully with SLP2tsr. This result indicated that $\text{tpg}_{\text{SLP2.19}}$ on SLP2 was functional, and the presence of $\text{tap}$ elsewhere was required for the replication of SLP2.

### Analysis of the N-terminal Region of Tpg

Compared to most other Tpgs, the pseudogene product $\text{tpg}_{\text{pSV2.102}}$ lacks eight amino acid residues at the N-terminal (Fig. 4). To examine the importance of these eight amino acids, three mutations in $\text{tpg}_{\text{pSV2.102}}$ were created with a deletion of eight (residue 2 to 9), four (2 to 5), and two (2 to 3) amino acids from the N-terminus. The defect of this product was likely to be due to one or more amino acid substitutions elsewhere.

$\text{tpg}_{\text{SLP2.19}}$ was not expressed on SLP2, or the participation of $\text{tap}$ was required as shown in the replication of the *Streptomyces* chromosomes [8]. To check this, $\text{tap}_{\text{Sli}}$ was inserted into the eC31 att site on the MR04 chromosome using the integrative plasmid pSET152 [40]. The resulting strain, CK03, could then be transformed successfully with SLP2tsr. This result indicated that $\text{tpg}_{\text{SLP2.19}}$ on SLP2 was functional, and the presence of $\text{tap}$ elsewhere was required for the replication of SLP2.

**Figure 4. Tpg sequence analysis.** The sequences of Tpg homologs encoded by various *Streptomyces* chromosomes and linear plasmids, including translation products of pseudogenes, are aligned. The chromosome-encoded Tpgs are designated by three-letter abbreviations of the species (Sil, *S. lividans*; Sav, *S. avermitilis*; Sro, *S. rochei*; Ssc, *S. scabies*), the plasmid-encoded Tpgs by the plasmid names and the pseudogenes by the designations in the sequence databases or publications. Tpgs encoded by the same replicon are distinguished by their gene designations. Sources of the sequences are: *S. coelicolor* chromosome [7], *S. lividans* chromosome [6], *S. avermitilis* chromosome and SAP1 plasmid [10], pSV2 plasmid (pSV2.82) in *S. violaceoruber* (GenBank accession number NC_004934), pFRL1 plasmid in *Streptomyces* sp. FR1 [45], *S. rochei* chromosome and pSLA2-L and pSLA2-M plasmids [46], SLP2 plasmid in *S. lividans* [7], *S. scabies* chromosome (http://www.sanger.ac.uk/Projects/S_scabies/), pSC2 plasmid in *S. clavuligerus* (GenBank accession number AY392421). The amino acid numbering is that of $\text{tpg}_{\text{Sli}}$. The lengths (in amino acid residues) of the Tpgs are indicated at the right. Conceptually translated products of pseudogenes proven in this study (SAP1.1, SLP2.38, and pSV2.102) and the apparent pseudogene, Sav_39 (with deletions of about 90 amino acids), are marked by asterisks. The structural and functional domains previously identified are indicated at the bottom: HIV reverse transcriptase (RT) domain; helix-turn-helix (HTH) domain; nuclear localization signal (NLS); and amphiphilic beta-sheet. Substitution mutations investigated in this study are placed above the sequence. Deletions are indicated by ‘D’. Multiple substitutions are boxed together. Mutations that inactivated Tpg are shown in red; those that did not, in green. doi:10.1371/journal.pone.0056322.g004
In addition to the N-terminal truncation, the 47th residue (based on Tpg	extsuperscript{6ii}) of Tpg	extsuperscript{102}, located in the ‘turn’ of the predicted HTH motif of Tpg, is a Thr, while Ala or Ile are present at this position in the other Tpgs (Fig. 4). To test the possible effect of the Thr substitution, a V47T mutation was introduced into Tpg	extsuperscript{6ii}. The resulting Tpg	extsuperscript{6ii} was functional, indicating that the V47T substitution had not effect on the functionality of Tpg	extsuperscript{6ii}.

Tpgs possess a putative helix-turn-helix (HTH) motif despite the fact that the in vitro DNA-binding specificity appears to be low [6]/Yi-Hong Chen, unpublished results). The atypical TP, Tpc, also contains a HTH domain with a different sequence [18]. To examine the importance of the predicted HTH domain in Tpg sequences (Fig. 4), several amino acids in the helix domains of Tpg	extsuperscript{6ii} were substituted by different residues to lower their potential to form HTH (Table 1). In the first helix of the HTH motif, the G37A and A40P mutations (which reduced the probability of HTH formation to 25%) did not cause a defective Tpg. Quadruple mutation, R18S-A26V-G37A-M44V, generated by error-prone PCR, inactivated Tpg	extsuperscript{6ii} (data not shown). Of these four mutations, G37A and M44V were in the HTH motif. Tpg	extsuperscript{6ii} containing the G37A-M44V double mutation had a 71% probability of HTH formation and exhibited no effect. Therefore, the defect caused by the quadruple mutations must be attributed to the other two upstream mutations. Moreover, an A42D mutation (insignificant probability of HTH formation) inactivated Tpg	extsuperscript{6ii} (Fig. S6, Supporting Information). Perhaps the collapse of the first helix might not be critical, but the defect was caused by repulsion of DNA by the negative charge of the Asp residue. These results are summarized in Table 1 and Fig. 4.

The second helix of the HTH motif is more hydrophilic. The R34P and Y53F mutations (71% and 90% probability of HTH formation, respectively) resulted in a defective Tpg	extsuperscript{6ii} (Fig. S6, Supporting Information). Possibly the positively charged R34 was involved in interaction with a phosphate group on the DNA. A V32D mutation in this region that introduced a negatively charged amino acid and reduced the predicted probability of HTH formation to 71% also resulted in a defective Tpg	extsuperscript{6ii} (Fig. S6, Supporting Information).

These results indicated the importance of the HTH motif of Tpg	extsuperscript{6ii} in end patching. This is in contrast to the NLS motif that immediately follows the HTH motif, which may be mutated without affecting replication of the linear plasmids [15].

### Table 1. Effects of mutations in the putative HTH region of Tpg	extsuperscript{6ii} on the probability of HTH formation and ability to support replication of linear plasmids.

| Mutation | Sequence	extsuperscript{1} | % Probability of HTH | Plasmid linearity	extsuperscript{2} |
|----------|-----------------------------|----------------------|-------------------------------------|
| Wild-type | KGTKAVAQMLRVSQRTVERYVK     | 100                  | +                                   |
| G37A     | KGTKAVAQMLRVSQRTVERYVK     | 25                   | +                                   |
| G37A M44V| KGTKAVAQMLRVSQRTVERYVK     | 71                   | +                                   |
| A40P     | KGTKVAQMLRVSQRTVERYVK      | 25                   | +                                   |
| A42D     | KGTKAVQMLRVSQRTVERYVK      | Insignificant        | –                                   |
| V47T     | KGTKAVAQMLRVSQRTVERYVK     | 90                   | +                                   |
| V52D     | KGTKAVAQMLRVSQRTDERVK      | 71                   | –                                   |
| R54P     | KGTKAVAQMLRVSQRTVEPKV      | 71                   | –                                   |
| Y55F     | KGTKAVAQMLRVSQRTVEVK       | 90                   | –                                   |

	extsuperscript{1}The substituted amino acid residues are underlined.

	extsuperscript{2}Support of linear plasmid replication: ‘+’, yes; ‘–’, no.

doi:10.1371/journal.pone.0056322.t001

Discussion

In this study, we have identified T114 to be the site of deoxynucleotidylation of Tpg	extsuperscript{6ii}. A substitution of T114 with Tyr, Cys, or His inactivated Tpg	extsuperscript{6ii}. A substitution with Ser did not, which is not surprising, because the functional lone Tpg	extsuperscript{6FRD1} also contains a Ser at this position (Fig. 4). For o29 TP, a substitution of the deoxynucleotidylation residue S232 with Thr inactivates the TP [41], but a substitution with Cys produces a TP with a reduced (7%) efficiency of deoxynucleotidylation [42].

It is noteworthy that the amino acid residue immediate upstream of T114 is also a Ser. When T114 became defective in deoxynucleotidylation through a His/Ala/Cys/Tyr substitution, S113 could not serve as the deoxynucleotidylation site as S114 could. Thus, it appeared that, while Ser might serve as a deoxynucleotidylation, the deoxynucleotidylation is specific at position 114.

The different effects of T114Y and T114S substitutions are interesting. While both Tyr and Ser possess a priming hydroxyl group for DNA synthesis, the inability of Tyr to serve the priming function may be due to its bulkier and relatively more hydrophobic side chain. Secondary structure prediction by the methods of Garnier, et al. [33] places T114 in a coiled domain flanked by two short helix segments in Tpg	extsuperscript{6ii} (Fig. 5 top). This is similar to the situations of the TPs of o29 and adenoviruses. That the deoxynucleotidylated Ser residues in these two viral TPs lie in a predicted coil segment flanked by two short helices has been previously noticed [21]. In the case of the TPs of o29, this prediction was confirmed in the TP-DNA polymerase c-crystal structure [44].

Hydropathy prediction showed that T114 lay in a highly hydrophilic region, presumably close to the surface of the protein (Fig. 5 bottom). The deoxynucleotidylation S232 of o29 [45] and S557 of adenovirus [20] also lie in highly hydrophilic regions (Fig. 5 bottom). In contrast to these, the deoxynucleotidylation Y190 of PRD1 TP lies in a b-structure enclosed in a hydrophobic area [45].

Despite these similarities in geometric properties, there is no similarity in sequence and size among the TPs in the different systems. The patching TP-primed DNA synthesis in Streptomyces also differs from the replicative TP-primed DNA synthesis (o29, adenovirus, and PRD1) in that the former requires a proper DNA template for in vitro deoxynucleotidylation [19], whereas the latter does not. Moreover, the telomere sequences in the former consist
of extensive palindromic sequences with potentials of forming complex secondary structures, which is not found in the telomeres of the latter. It follows that the telomere associating protein, Tap, that recognizes the secondary structure formed by the 3’ overhang at the telomeres is also absent from the replicative TP-primed synthesis systems. All these considerations, plus the use of different DNA polymerases by these systems [43], lead to a conclusion that these different TP-primed DNA synthesis systems have evolved independently, recruiting different proteins with a suitable priming function in the process. Even among Streptomyces, there are atypical, heterologous TPs capping linear plasmids and linear chromosomes, one of which (Tap of SCP1) has been characterized [18].

In addition to the variations in TP families in Streptomyces, the frequent occurrence of tpg pseudogenes is also remarkable for an essential housekeeping gene. In this study, we have demonstrated three defective tpg pseudogenes out of five candidates. It is unlikely that the pseudogenes have arisen by gene duplication, and the absence of an accompanying tap gene argues against it. More likely, the lone tpg homologs were acquired through horizontal transfer from a plasmid. The findings of the tpg pseudogenes in the terminal regions of the linear replicons of Streptomyces, where exchanges with other linear replicons are frequent, also support this notion. The abundance of tpg pseudogenes suggests that such horizontal events are relatively frequent.

SLP2 is interesting in that it contains two lone tpg homologs, the functional tpgSLP2.19 and the pseudogene tpgSLP2.30. Although tpgSLP2.19 is functional, a tap gene present elsewhere is nonetheless required for the replication of SLP2. This finding answers the previous observations that SLP2 could not replicate in certain S. lividans mutants whose chromosomales had suffered circularization and deletions (in the terminal regions where the tap-tpg operon resides) [44]. Apparently, tap was the chromosomal gene in the deletion that was required for replication of SLP2. The requirement of tap for replication of SLP2 is the same for that of the linear chromosomes in Streptomyces [8]. The evolutionary significance of the emergence of lone functional tpg is obscure.

Supporting Information

Figure S1 Topology of pLUS980(Nd) derivatives containing substitutions at T108 or T114 of tpg. (A) Physical maps of plasmids pLUS980 and pLUS980(Nd); (B) Genomic DNA containing a pLUS980 derivative (except lanes 2 and 3) was isolated from MRO4, digested with SpI (Sp), and subjected to agarose gel electrophoresis. The four substitutions at T108 and at T114 are indicated by the substituting amino acids. Lane M, 1 kb DNA ladders as size markers. Lanes 1–5, plasmid DNA (purified or in total genomic DNA) of known topology and size serving as controls and markers; 1, SpI-digested genomic DNA from MRO4 containing a pLUS980 derivative (fragment size: 8.0 and 5.1 kb); 2, AseI-digested pLUS980 (largest fragment 14.3 kb); 3, SpI-digested pLUS980 (15.6 kb); lane 4: SpI-digested genomic DNA from MR04 containing SpI-digested pLUS980(Nd) (8.0 and 5.1 kb); 5, SpI-digested genomic DNA from MR04 containing pLUS980(Nd) with a S113A substitution in tpg (8.0 and 5.1 kb). The sizes (in kb) of the SpI fragments are indicated. The sizes of the circularized plasmid DNA derived from the four linearized pLUS980(Nd) derivatives with a substitution at T114 varied from transformants to transformants, depending on the end joining sites. Shown here are representative cases. (TIFF)

Figure S2 Topology of pLUS980(Nd) derivatives containing a T101S, T143S, T176S, or T123S substitution in tpg. (A) Physical maps of plasmids pLUS980 and pLUS980(Nd). (B) Isolated genomic DNA was digested with SacI (Sa) except for that in lane 2, which was not enzyme digested. The sizes (in kb) of pLUS980(Nd)L (T143S) and the SaI fragments are indicated. (TIFF)
Figure S3 Topology of the pLUS980(Nd) derivatives containing D115A and D117A mutations. (A) Physical maps of plasmids pLUS980 and pLUS980(Nd). (B) Genomic DNA from transformants of pLUS980(Nd) and its derivatives linearized by Acl digestion. Lane 1, pLUS980(Nd). Lane 2, pLUS980(Nd) containing D115A mutation (no linear plasmid present). Lane 3, pLUS980(Nd) containing D117A mutation. Lane 4, pLUS980(Nd) containing D115E mutation. Lane 5, pLUS980(Nd) containing D115N mutation (no linear plasmid present). The sizes of the circularized plasmid DNA derived from the transformants varied from transformants to transformants, depending on the end joining sites. Shown here are representative cases. (TIFF)

Figure S4 Topology of pLUS980 derivatives containing tpg homologs from linear plasmids. (A) Physical maps of plasmids pLUS980 and pLUS980(Nd). (B) Genomic DNA isolated from MR04 transformants of the pLUS980 derivatives, and electrophoresed with or without prior restriction digestion with SacI or SpeI as indicated.  \( \eta_{tpg} \) was substituted by the following homologs in the pLUS980 derivatives: \( \eta_{tpg}^{pSV2.102} \), \( \eta_{tpg}^{pSV2.072} \), \( \eta_{tpg}^{pSV2.052} \), \( \eta_{tpg}^{pSV2.102} \), \( \eta_{tpg}^{pSV2.072} \), \( \eta_{tpg}^{pSV2.052} \), \( \eta_{tpg}^{FRL1.6} \), \( \eta_{tpg}^{AP1-11} \), \( \eta_{tpg}^{pSV2.102} \), \( \eta_{tpg}^{pSV2.072} \), \( \eta_{tpg}^{pSV2.052} \), \( \eta_{tpg}^{FRL1.6} \), \( \eta_{tpg}^{AP1-11} \), and \( \eta_{tpg}^{pSV2.102} \). The sizes of the circularized plasmid DNA derived from the transformants varied from transformants to transformants, depending on the end joining sites. Shown here are representative cases. (TIFF)

Figure S5 Topology of pLUS980(Nd) derivatives containing N-terminal deletions in tpg. (A) Physical maps of plasmids pLUS980 and pLUS980(Nd). (B) Genomic DNA from three independent transformants of MR04 was digested with SacI (Sa). Lanes 1–3, deletion of 2 aa’s (residues 2–3); lanes 4–6, deletion of 4 aa’s (residues 2–5); lane 7–9, deletion of 8 aa’s (residues 2–9); lane 10, pLUS980(Nd) DNA. The sizes (in kb) of the SacI fragments are indicated. The sizes of the circularized plasmid DNA derived from the transforming linear plasmid DNA varied from transformants to transformants, depending on the end joining sites. Shown here are representative cases. (TIFF)

Acknowledgments

We thank Professor David A. Hopwood for reading the manuscript and suggesting improvements.

Author Contributions

Conceived and designed the experiments: CCY CWC. Performed the experiments: CCY WGS WWY FSL. Analyzed the data: CCY CWC. Contributed reagents/materials/analysis tools: CHH. Wrote the paper: CCY CWC.

References

1. Kirby R, Chen CW (2011) Genome architecture. In: Dyson P, editor. Strepomyces: Molecular Biology and Biotechnology. Norfolk, UK: Caister Academic Press. p 5–26.
2. Chen CW (2007) Strepomyces Linear Plasmids: Replication and Telomeres. In: Minhardt F, Klassen R, editors. Microbial Linear Plasmids. Berlin, Heidelberg: Springer-Verlag. 33–61.
3. Chaconas G, Chen CW (2005) Replication of linear bacterial chromosomes: No longer going around in circles. In: Higgins NP, editor. The Bacterial Chromosome. Washington, D.C.: American Society for Microbiology. 525–539.
4. Liu H, Naismith JH, Hay RT (2003) Adenovirus DNA replication. Curr Top Microbiol Immunol 272: 131–164.
5. Salas M (1991) Protein-primed DNA replication. Annu Rev Biochem 60: 37–91.
6. Bao K, Cohen SN (2001) Terminal proteins essential for the replication of linear plasmids and chromosomes in Strepomyces. Genes Dev 15: 1518–1527.
7. Yang C-C, Huang C-H, Li C-Y, Tsay Y-G, Lee S-C, et al. (2002) The terminal proteins of linear plasmids and chromosomes: A novel class of replication priming proteins. Mol Microbiol 43: 397–405.
8. Bao K, Cohen SN (2003) Recruitment of terminal protein to the ends of Strepomyces linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. Genes Dev 17: 774–785.
9. Huang CH, Chen CW, Tsai HH, Chen C, Lin YS, et al. (2003) Linear plasmid SLP2 of Strepomyces lividans is a composite replicon. Mol Microbiol 47: 1563–1576.
10. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, et al. (2003) Complete genome sequence and comparative analysis of the industrial microorganism Strepomyces avermitilis. Nat Biotechnol 21: 526–531.
11. Chouhet F, Aigle B, Gallois A, Mangenet S, Gerbaud C, et al. (2006) Evolution of the terminal regions of the Strepomyces linear chromosome. Mol Biol Evol 23: 2361–2369.
12. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikunoya M, et al. (2008) Genome sequence of the streptomycin-producing microorganism Strepomyces griseus IFO 13350. J Bacteriol 190: 4050–4060.
13. Suzuki H, Marushima K, Ohnishi Y, Horinouchi S (2008) A novel pair of terminal protein and telomere-associated protein for replication of the linear chromosome of Strepomyces griseus IFO13350. Biosci Biotechnol Biochem 72: 2973–2980.
14. Spatz K, Kohin H, Redenbach M (2002) Characterization of the Strepomyces cattleyaenei SANK55570 plasmids pSV1 and pSV2. FEMS Microbiol Lett 213: 85–92.
15. Tsai H-H, Huang C-H, Lin AM, Chen CW (2008) Terminal proteins of Strepomyces chromosome can target DNA into eukaryotic nuclei. Nucleic Acids Res 36: e62.
16. Mao T (2010) Function of the nuclear localization signals of Strepomyces terminal proteins in plants [M. S. Thesis]. Taipei: National Yang-Ming University. 64 p.
17. Tsai H-H, Huang C-H, Tesmer I, Eric DA, Chen CW (2011) Linear Strepomyces plasmids form superhelical circles through interactions between their terminal proteins. Nucleic Acids Research 39: 2163–2174.
18. Huang C-H, Tsai H-H, Tsai Y-G, Chien Y-N, Wang S-L, et al. (2007) The telomere system of the Strepomyces linear plasmid SCP1 represents a novel class. Molecular Microbiology 63: 1710–1718.
19. Yang CC, Chen YH, Tsai HH, Huang CH, Huang TW, et al. (2006) In vitro deoxynucleotidylation of the terminal protein of Strepomyces linear chromosomes. Appl Environ Microbiol 72: 7959–7961.
20. de Jong RN, van der Vliet PC (1999) Mechanism of DNA replication in eukaryotic cells: Cellular host factors stimulating adenovirus DNA replication. Gene 236: 1–12.
21. Hermoso JM, Mendez E, Soriano F, Salas M (1965) Location of the active residue involved in the linkage between the terminal protein and the DNA of phage φ29. Nucleic Acids Res 13: 7715–7720.
22. Kioser T, Bibb M, Buttner MJ, Chater KF, Hopwood DA (2000) Practical Strepomyces Genetics. Norwich: The John Innes Foundation. 613 p.
23. Redenbach M, Fleet F, Pennel W, Gluckcr I, Rauland U, et al. (1995) The Strepomyces lividans Tn5 chromosome contains a 1 Mb deletogenic region flanked by two ampalatable regions. Mol Gen Genet 241: 255–262.
24. Redenbach M, Kioser HM, Dynapaitie D, Eichner A, Callum J, et al. (1996) A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Strepomyces coelicolor A3(2) chromosome. Molecular Microbiology 21: 77–96.
25. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, et al. (1994) Current Protocols in Molecular Biology: Current Protocols.
26. Pargellis CA, Nunes-Duby SE, de Vargas LM, Landy A (1988) Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. J Biol Chem 263: 7678–7685.

27. García P, Hermoso JM, García JA, García E, López R, et al. (1986) Formation of a covalent complex between the terminal protein of pneumococcal bacteriophage Cp-1 and 5′-dAMP. J Virol 58: 31–35.

28. Coligan JE, Dunn BM, Plough HL, Speicher DW, Wingfield PT, editors (1995) Current Protocols in Protein Science: John Wiley.

29. Westermeier J (1990) Electrophoresis in Practice. Weinheim, Germany.

30. Qin Z, Cohen SN (1998) Replication at the telomeres of the Streptomyces linear plasmid pSLA2. Mol Microbiol 28: 893–904.

31. Shiffman D, Cohen SN (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.

32. Dodd IB, Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res 18: 5019–5026.

33. Garnier J, Gibrat JF, Robson B (1996) GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 266: 540–553.

34. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophatic character of a protein. J Mol Biol 157: 105–132.

35. Smart JE, Stillman BW (1982) Adenovirus terminal protein precursor. Partial amino acid sequence and the site of covalent linkage to virus DNA. J Biol Chem 257: 13499–13506.

36. Kolkenbrock S, Fetzer N (2010) Identification and in vitro deoxynucleotidylation of the terminal protein of the linear plasmid pAL1 of Alkabacter nitroguajacolicus Ru61a. FEMS Microbiol Lett 304: 169–176.

37. Qin Z, Cohen SN (2002) Survival mechanisms for Streptomyces linear replicons after telomere damage. Mol Microbiol 45: 785–794.

38. Billy E, Hess D, Hofsteenge J, Filipowicz W (1999) Characterization of the adenylation site in the RNA 3′-terminal phosphate cyclase from Escherichia coli. J Biol Chem 274: 34955–34960.

39. Hsu CC, Chen CW (2010) Linear plasmid SLP2 is maintained by partitioning, intrahyphal spread, and conjugal transfer in Streptomyces. J Bacteriol 192: 307–315.

40. Bierrmann M, Logan R, O’Brien K, Seno ET, Rao RN, et al. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116: 43–49.

41. Garmendia C, Salas M, Hermoso JM (1983) Site-directed mutagenesis in the DNA linking site of bacteriophage φ29 terminal protein: isolation and characterization of a Ser232Thr mutant. Nucleic Acids Res 16: 5727–5740.

42. Garmendia C, Hermoso JM, Salas M (1990) Functional domain for priming activity in the phage φ29 terminal protein. Gene 90: 73–79.

43. Tsai HH, Niu HW, Yang CC, Chen CW (2011) Translesion-synthesis DNA polymerases participate in replication of the telomeres in Streptomyces. Nucleic Acids Res.

44. Lin Y-L (1998) Involvement of the terminal sequence of the Streptomyces chromosome in the maintenance of linear plasmids [M. S. thesis]. Taipei: National Yang Ming University. 69 p.

45. Zhang R, Yang Y, Fang P, Jiang C, Xu L, et al. (2006) Diversity of telomere palindromic sequences and replication genes among Streptomyces linear plasmids. Appl Environ Microbiol 72: 5729–5733.

46. Kinashi H, Fujii S, Hatani A, Kurokawa T, Shinkawa H (1998) Physical mapping of the linear plasmid pSLA2-L and localization of the eryCII and actI homologs. Biosci Biotechnol Biochem 62: 1892–1897.