Impaired Striatal Akt Signaling Disrupts Dopamine Homeostasis and Increases Feeding
A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette

Richard P. Bonocora¹, Qinglu Zeng¹,², Ethan V. Abel³, and David A. Shub⁴

Department of Biological Sciences, University at Albany, State University of New York, Albany, NY 12222

Edited* by Marlene Belfort, Wadsworth Center, Albany, NY, and approved August 23, 2011 (received for review May 25, 2011)

Since its initial description more than two decades ago, the ribosome bypass (or “hop”) sequence of phage T4 stands out as a uniquely extreme example of programmed translational frameshifting. The gene for a DNA topoisomerase subunit of T4 has been split by a 1-kb insertion into two genes that retain topoisomerase function. A second 50-nt insertion, beginning with an in-phase stop codon, is inserted near the start of the newly created downstream gene 60. Instead of terminating at this stop codon, approximately half of the ribosomes skip 50 nucleotides and continue translation in a new reading frame. However, no functions, regulatory or otherwise, have been imputed for the truncated peptide that results from termination at codon 46 or for the bypass sequence itself. Moreover, how this unusual mRNA organization arose and why it is maintained have never been explained. We show here that a homing endonuclease (MobA) is encoded in the insertion that created gene 60, and the mobA gene together with the bypass sequence constitute a mobile DNA cassette. The bypass sequence provides protection against self-cleaveage by the nuclease, whereas the nuclease promotes horizontal spread of the entire cassette to related bacteriophages. Group I introns frequently provide protection against self-cleaveage by associated homing endonucleases. We present a scenario by which the bypass sequence, which is otherwise a unique genetic element, might have been derived from a degenerate group I intron.

In most of the phages of the T4 superfamily, the large subunit of DNA topoisomerase is a single polypeptide encoded by gene 39. However, in phage T4 this gene is disrupted by a 1-kb insertion, creating a truncated gene 39 plus a new gene (gene 60) that encodes the remainder of the topoisomerase subunit (1, 2) (Fig. 1). Furthermore, the C terminus of the truncated gene 39 and the N terminus of the newly created gene 60 each contain additional amino acid residues (43 and 30, respectively) that were not present in the original gene 39 homologs. These two independently translated proteins interact with the small subunit (encoded by gene 52) to create an enzymatically active topoisomerase (4). A second, 50-nt insertion beginning with an in-phase stop codon, is inserted into the newly created downstream gene 60 (2). Instead of terminating at this stop codon, approximately half of the ribosomes skip 50 nucleotides and continue translation in a new reading frame (5). Features involved in this remarkable process include: the stop codon, the codon at which translation resumes, a portion of the pre-hop nascent peptide, a stem-loop at the start of the bypassed sequence, a Shine/Dalgarno-like sequence within the bypassed RNA, and the structure of the bypassed mRNA (5–8).

The genome sequence of phage T4 (GenBank accession no. NC_000866) indicated that the insertion into gene 39 contains two ORFs: an apparent H-N-H homing endonuclease pseudo-gene (mobA), whose inferred translation product terminates after only 37 codons (Fig. 2), and gene 60.1, which overlaps the start of gene 60 and encodes a basic protein of 126 amino acids with no relatives in the databases. Interestingly, a screen of independently isolated phages related to T4 revealed several (5/37) having the same architecture in their large subunit topoisomerase genes. Although DNA sequences of the ~1-kb insertions were not provided, each newly created gene 60 contained a ribosome bypass sequence identical to that of T4 (9, 10).

Results and Discussion

MobA Is a Site-Specific Endonuclease That Nicks T2 DNA Close to the Insertion Site of the T4 Bypass Sequence. Thinking that one of the insertions initially identified by Repoila et al. (9) might encode an intact mobA gene, we sequenced the PCR products of the gene 39/60 regions of two of them (phages Pol and SKX, provided by Henry Krisch, Laboratory of Microbiology and Molecular Genetics, CNRS, Toulouse, France) and resequenced this region from phage T4. Surprisingly, the insertions in all three phages are almost identical, encoding a single ORF of 271 amino acids (including both the mobA H-N-H domain and g60.1) that is highly similar to the other phage T4 H-N-H family homing endonucleases (Fig. 2). The SKX and Pol mobA sequences are identical; the only differences from the T4 sequence are synonymous G-to-A third position substitution in codon 156 and an A-to-G substitution at the second position of codon 169, changing threonine to alanine.

Free-standing homing endonucleases (i.e., those not encoded within a group I intron or intein) cleave the DNA of close relatives lacking the endonuclease gene, usually in a gene adjacent to the site of endonuclease gene insertion (12–14), and are copied into the recipient genome via recombination-mediated DNA repair, a process that has been called “homing” (for recent reviews, see refs. 14 and 15). Therefore, we tested the protein product of the mobA gene for endonuclease activity on DNA from phage T2, which has an uninterrupted gene 39.

The mobA gene of phage T4 was expressed in vitro using cell-free extracts and also in vivo by induction of plasmid expression vector pMobA. Incubation of either protein preparation with a

Author contributions: R.P.B., Q.Z., and D.A.S. designed research; R.P.B. and Q.Z. performed research; E.V.A. contributed new reagents/analytic tools; R.P.B., Q.Z., and D.A.S. analyzed data; and R.P.B., Q.Z., and D.A.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

†R.P.B. and Q.Z. contributed equally to this work.

‡Present address: Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139.

§Present address: Thomas Jefferson University, 233 South 10th St., Philadelphia, PA 19107.

To whom correspondence should be addressed. E-mail: shub@albany.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1107633108/-/DCSupplemental.

After this work was completed, the phage T4 genomic sequence was redetermined (accession no. HM_137666) with mobA sequence identical to that presented here (11). The inaccuracies in accession NC_000866 were undoubtedly due to the common practice at that time of cloning genes in unregulated plasmids before sequencing. Genes like mobA, whose products are highly toxic to E. coli, were likely to have suffered inactivating mutations at the cloning step.
PCR-amplified region of T2 gene 39 generated a specific nick on the template (bottom) strand (Figs. 3 and 4B), but the proteins were inactive on the homologous T4 sequence (Fig. 4B). Introduction of a UAG stop codon at Tyr195 (Fig. 2) abolished activity (Fig. 3). To confirm that this activity was indeed encoded in T4 (and not in some other phage contaminant), we used DNA from a mutant T4 phage in our collection that has a deletion (saΔ9) (16) located distantly from the g39/60 locus as template for in vitro protein synthesis. This protein product displayed similar nuclease activity on T2 DNA as when our wild-type T4 DNA was used to program the reaction (Fig. 3). Nucleotide resolution mapping of the cleavage position showed that MobA cleaves T2 DNA 19 bp upstream of the insertion site of the bypass sequence in T4 gene 60 (Fig. 4A and C.).

**Function of the Bypass Sequence Is Prevention of Self-Cleavage by MobA.** The close association of the cleavage site and the bypass insertion site suggested that the bypass sequence might play...
 Horizontal Gene Transfer.

Origin of the Bypass Sequence. The results presented here provide an explanation for the function (prevention of self-cleavage by MobA) and persistence (spread by horizontal gene transfer) of the ribosomal bypass sequence. However, we still have no convincing evidence for how this remarkable element came to be inserted into gene 60 in the first place. Although programmed translational frameshifts of one or a few nucleotides have been repeatedly encountered in bacteria and eukaryotes, nothing remotely close to the 50-nt ribosomal hop in T4 gene 60 has been substantiated in any other biological system (20). One possible scenario, consistent with the rarity of bypassing, is suggested by similarities of this system to the recently described phenomenon of “collaborative homing” of homing endonucleases and group I introns (13).

Intron-encoded homing endonucleases are protected from cleaving their own genomes by disruption of the DNA-binding/cleavage site by the intron (15) (Fig. 5A), whereas most intronless (or free-standing) homing enzymes have sequence polymorphisms at their target sites that prevent cleavage (17) (Fig. 5B). In collaborative homing, an intron (lacking its own homing endonuclease gene) is inserted into the target sequence of a free-standing homing endonuclease, thereby preventing self-cleavage by the enzyme (13) (Fig. 5C). When a related sensitive DNA sequence is cleaved, both the intron and endonuclease gene are incorporated into the recipient during recombination-dependent DNA repair. The relationship between mobA and the bypass sequence appears to be a variation of the collaborative homing theme (Fig. 5D), which raises the interesting possibility that the bypass may once have been a functional group I intron.

Despite the lack of sequence relatedness between the bypass sequence and any known intron, there are suggestive similarities in their structures and insertion sites. Like the bypass sequence, the 5′ ends of many bacteriophage introns begin with an in-frame stop codon, which is part of the base-paired stem (P1) with which all group I introns begin (21, 22). Furthermore, like the insertion sites of most group I introns (12, 23–25), the bypass sequence is inserted into a highly conserved sequence of an essential protein, the Mg2+-binding region at the active site of type II topoisomerases (26) (Fig. 1B).

These considerations, together with the results presented here, lead us to propose the following hypothesis as a plausible scenario for origin of the bypass. It is possible that the bypass sequence was derived from a group I intron that engaged in collaborative homing with free-standing endonuclease MobA. The intron may have suffered a large deletion that prevented splicing but, adventitiously, the conformation of the remaining RNA (8) may have permitted ribosomal bypassing. Unlike most other “essential” T4 genes, conditional lethal mutations in gene 39 still allow a significant yield of progeny phage under non-permissive conditions (27). So, despite the negative selection...
implied by reduced translation efficiency, the bypass could have been maintained because of its essential relationship with mobA, with continued selection favoring increased efficiency of ribosome hopping. In this context, it is interesting to note that a portion of the nascent gp60 peptide chain upstream of the bypass sequence, which is required for efficient ribosome hopping (6, 7), is not homologous to a segment of the original gene 39, but is contributed as part of the 1-kb insertion that includes mobA (Fig. 1A).

mobA is one example of a homing endonuclease-like gene that disrupts the coding sequence of an essential gene without destroying its function (1), and two other examples have recently been described in phages belonging to the T4 superfamily. In one case, a GIY-YIG family endonuclease gene disrupts DNA polymerase (28, 29); in the other case, a H-N-H family endonuclease gene disrupts the large subunit of ribonucleotide reductase (30). In all three cases, the products of the split genes reunite to form enzymatically active proteins. In the ribonucleotide reductase case, this requires peptide sequences that had been added to the ends of the split proteins at the site of the endonuclease gene insertion (31). The C terminus of T4 gp39 and the N terminus of gp60 also have extensive amino acid sequences (43 and 30 residues, respectively) that are absent in the intact homologs and it has been suggested (31) that these may also play a role in assembly of the split proteins into an active enzyme.

Table 1. Homing frequency

<table>
<thead>
<tr>
<th>Expression plasmid*</th>
<th>Experiment</th>
<th>0</th>
<th>0.00002</th>
<th>0.0002</th>
<th>0.002</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMobA</td>
<td>1</td>
<td>97</td>
<td>99</td>
<td>94</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>88</td>
<td>98</td>
<td>92</td>
<td>54</td>
</tr>
<tr>
<td>pF-Cphi</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>pBAD</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

*T2 progeny plaques (of 100 picked) that hybridize to the bypass sequence probe MobA25.

*Cells also contain the donor plasmid pT4 (39-60ΔmobA).

Fig. 4. MobA cleavage of phage DNA depends on the absence of the bypass sequence. (A) Determination of the cleavage site of MobA on T2 gene 39 DNA. Substrate DNA amplified with end-labeled bottom strand primer: (+) digested or (−) undigested with in vitro-synthesized MobA. The same end-labeled oligonucleotide was used to prime dideoxy sequencing reactions on cloned T2 DNA. All samples were subjected to electrophoresis in the same polyacrylamide gel. Sequence and position of cleavage (arrow) are shown on the left. (B) DNAs were amplified by PCR from plasmid templates, with either the top (Top) or bottom (Bot) strand primer end-labeled with 32P. Amplified DNA was digested (+) or not digested (−) with MobA and subjected to denaturing gel electrophoresis next to molecular size markers (M). Target DNAs were from wild type T2 and T4 with the bypass sequence deleted (T4hop), and T2 with the bypass sequence inserted (T2hop). (C) The DNA sequence of a portion of T2 gene 39 is shown, with the position of cleavage (CS) by MobA on the bottom strand indicated by an arrow. The coding strand of the homologous region of T4 gene 60 is shown above, with the position of insertion of the 50-nt bypass sequence (HOP) indicated by an arrow. Sequence differences are shown in boldface type. The translated amino acid sequence is shown above the nucleotide sequence.

Fig. 5. Modes of protection against self-cleavage by homing endonucleases. (A) Homing endonuclease gene (HEG) encoded within a group I intron. The intron interrupts the sensitive binding/cleavage site (O). (B) HEG inserted intercistronically, resistance due to sequence differences at the binding/cleavage site (○). (C) HEG inserted intercistronically. Resistance is due to group I intron inserted into sensitive binding/cleavage site (●). (D) mobA inserted within an ancestral gene 39. Resistance is due to insertion of bypass sequence (Hop) within the sensitive binding/cleavage site. Group I introns, dark bars; homing endonuclease genes, shaded boxes; target genes without insertions, clear boxes; intergenic spaces, thin lines.
We have shown here that mobA can be spread by horizontal transmission, and it is likely that other similar constructs will also be shown to practice homing. Recent metagenomic analysis of environmental DNA samples has yielded additional examples of genes split by insertion of putative homing endonucleases (32), indicating that the infectious splitting of essential proteins may be a more common phenomenon than previously thought.

Materials and Methods

Oligonucleotides. Restriction sites are underlined in the sequences listed below.

ym001: AAAATTACCGCTCATAAGGATACCTGCTAGGAGACA
TATGATTAGAAAATACATAATGTAG
ynn03: GAGGATTTAATAGAACGACGAACATT
ym004: CCGAAAACCAAAGGAGATCTAAGCATGA
T2T4-3: AAAAATGCAAGAGTTTCCAAAGGACGACTCC
T4idhp1: TCTGAGGCTTATTGTCTCTGCTCCTGTTAGTAA
T4idhp2: TCCAGCACGAGAAGAAATAGAATACCTCCTACATCGTCCCTCG
T4idhp4: CTGTATATGACGAGGAGCATTAGGTAGTTATCTAACTTT
GTTCTCCTTGCTG
T4idhp2: TCCAAATACCTCTATTAGGTATTTTTTAGATTAGATACCGG
GCTAACATCTAGTGTACGACCTGCA
(50-nt bypass sequence is in italics);
T4idhp4: TATATAGGATCCAAAGGAAATGCTGCTATTAAACATG
T4dhop1: TCTGCAAGTACGAGGGTACGCTTACCTATAG
T4dhop2: TCCAGCACGAGAAGAAATAGAATACCTCCTACATCGTCCCTCG
T4dhop4: CTGTATATGACGAGGAGCATTAGGTAGTTATCTAACTTT
GTTCTCCTTGCTG

Plasmid Construction. For expression of MobA, T4 mobA was amplified by PCR with KOD HiFi DNA polymerase (Novagen) using primers EVA7 and EVA8, digested with PciI and EcoRI and ligated into similarly digested pBAD/Myc-HisA vector (Invitrogen) to create pmobA. The 50-bp ribosomal bypass from phage T4 was inserted in a fragment of T2 gene 39 by the SOElig PCR method (33) with KOD HiFi DNA polymerase using primers T2T4-3, T2hop2, T2hop1, and T2hop2. The product was digested with PstI and SacI and ligated into a pBSM13+ (Stratagene) to generate pT2hop. pT4dhop, which has the ribosomal bypass sequence removed from a fragment of the T4 gene 60, was created in a similar fashion using primers T2T4-3, T4idhp4, T4idhop2, and T4idhop1.

Plasmid pmobATS4 was constructed by amplification of the T2 gene 39 with primers MobA10 and MobA11 and cloned in pBSM13+. pT4(39–60)mobA was constructed as the donor plasmid for homing experiments. It contains all of gene 39 and all but the last seven codons of gene 60 of phage T4, to provide sequence homology for recombination with phase T2. To prevent the lethality that accompanies its uncontrolled expression, the plasmid also contains an in-frame deletion of mobA lacking codons 15–132: T4 gene 39 to codon 14 of mobA was amplified with primers g39.5 and MobA7, and amplification from codon 133 of mobA to codon 153 of gene 60 was with primers MobA8B and MobA6. These products were digested with BamH1 and ligated, and the purified product was digested with PstI and XbaI and ligated into similarly digested pACYC184.

In Vitro Expression of MobA Protein. Templates for in vitro protein synthesis were obtained by PCR amplification of T4 mobA with Taq DNA polymerase (Fermentas), according to the manufacturer’s specifications, using primers ym001 and ym003. The purified product was used to direct MobA expression in vitro with TNT Quick Coupled Transcription/Translation Systems (Promega) as specified by the manufacturer.

In Vivo Expression of MobA Under Induction by l-Arabinose. A culture of E. coli LMG194 (Invitrogen) containing pMobA was grown to OD600 = 0.5 in RM medium (1% M9 Salts, 2% Casamino Acids (Difco), 0.2% glucose, 1 mM MgCl2) containing 100 μg/mL ampicillin, induced by adding l-arabinose to a final concentration of 10 μM, and allowed to grow for 5 h. Cells were harvested by centrifugation at 6,000 × g for 20 min, resuspended in equal volume ice-cold lysis mixture (50 mM Tris HCl, pH 7.2, 1 mM EDTA, 1 mM PMSF, 2 μg/mL leupeptin, and 200 mM KCl), sonicated to complete lysis and used directly in endonuclease assays.

Generation of DNA Substrates. Individually 5’ end-labeled targets were generated by 5’ end labeling (with [γ-32P]ATP and T4 polynucleotide kinase) one of the oligonucleotides before use in PCR together with an unlabeled primer partner (34). PCR was performed with Taq DNA polymerase as described above.

Endonuclease Assays. Endonuclease assay reactions (10 μL) containing 2 μL crude cell extract and 4 μL DNA substrate were incubated at 30 °C for 30 min in 0.05 M NaCl, 0.05 M Tris (pH 7.5), and 0.5 μg poly(dI-dC). Reaction products were extracted with an equal volume of phenol before separation on a 4% denaturing polyacrylamide gel.

Cloning Site Mapping. The location of the MobA cleavage site was mapped, as previously described (34), on the bottom strand of the T2 gene 39 using substrate amplified from plasmid pmobATS4 with labeled M13fwd primer and unlabeled M13rev primer. PCR amplification, 5’ end-labeling, and endonuclease conditions were the same as described above. Cleavage products were separated by electrophoresis on 4% denaturing polyacrylamide gel next to sequencing ladders generated using the same template DNA and labeled primers.

Plasmid to Phage Homing. Plasmid-to-phage homing assays were performed as described by Quirk et al. (18). Briefly, E. coli LMG194 cells harboring plasmids pmobA (donor) and either pMobA, pF-CphI, (35) or pBAD/Myc-HisA (the empty expression plasmid) were induced with increasing concentrations of l-arabinose on plates at 30 °C for 2 h. Serial dilutions of T2 phage were spotted onto the plates and incubated overnight at 30 °C. The dilution that had almost cleared was scraped off, mixed with 1 mL buffered saline, shaken at 37 °C for 1 h, and diluted for single plaques on H plates. Plaques were transferred to a gridded lawn, and growing plasmid products were detected by plaque hybridization (17) using oligo tels 24a that were separated by electrophoresis on 4% denaturing polyacrylamide gel next to sequencing ladders generated using the same template DNA and labeled primers.