Multiple Controls Regulate the Expression of mobE, an HNH Homing Endonuclease Gene Embedded within a Ribonucleotide Reductase Gene of Phage Aeh1

Ewan A. Gibb and David R. Edgell*

Department of Biochemistry, Schulich School of Medicine and Dentistry, The University of Western Ontario, London, ON N6A 5C1, Canada

Received 4 March 2007/Accepted 13 April 2007

Mobile genetic elements have the potential to influence the expression of genes surrounding their insertion site upon invasion of a genome. Here, we examine the transcriptional organization of a ribonucleotide reductase operon (nrd) that has been invaded by an HNH family homing endonuclease, mobE. In Aeromonas hydrophila phage Aeh1, mobE has inserted into the large-subunit gene (nrdA) of aerobic ribonucleotide reductase (RNR), splitting it into two smaller genes, nrdA-a and nrdA-b. This gene organization differs from that in phages T4, T6, RB2, RB3, RB15, and LZ7, where mobE is inserted in the nrdA-nrdB intergenic region. We present evidence that the expression of Aeh1 mobE is regulated by transcriptional, posttranscriptional, and translational controls. An Aeh1-specific late promoter drives expression of mobE, but strikingly the mobE transcript is processed internally at an RNase E-like site. We also identified a putative stem-loop structure upstream of mobE that sequesters the mobE ribosome binding site, presumably acting to down regulate MobE translation. Moreover, our transcriptional analyses indicate that the surrounding nrd genes of phage Aeh1 are expressed by a different strategy than are the corresponding phage T4 genes and that transcriptional readthrough is the only mechanism by which the promoterless Aeh1 nrdB gene is expressed. We suggest that the occurrence of multiple layers of control to limit the expression of mobE late in the Aeh1 infection cycle is an adaptation of Aeh1 to reduce any effects on expression of the surrounding nrd genes early in phage infection when RNR function is critical.

Homing endonucleases are a distinctive class of site-specific yet sequence-tolerant DNA endonucleases that promote the mobility of themselves and surrounding genomic sequence to genomes lacking the endonuclease by a process termed homing (reviewed in reference 3). Homing endonuclease genes are often found within self-splicing group I or II introns (3, 30) or inteins (20, 21), but many bacterial and phage genomes encode a large number of so-called freestanding endonucleases, i.e., the genomes carry endonuclease genes that are not obviously encoded within self-splicing elements (13, 29, 40, 54). Experimental evidence to date suggests that freestanding endonucleases are also mobile genetic elements, promoting their spread to genomes lacking the endonuclease by a double-strand break (DSB) repair pathway termed intronless homing (4, 28, 33, 51). Database surveys of sequenced bacterial and phage genomes have revealed that freestanding endonucleases are more abundant than intron- or intein-encoded versions, which is particularly evident in the T-even-like phages (29, 40, 44, 54). Phage T4, for instance, is infested with 15 homing endonucleases, representing ~10% of the coding potential of the genome; 13 of these endonucleases are freestanding (40).

Whereas the mobility pathways of intron-encoded and freestanding endonucleases are well described (4, 33, 41, 46), comparatively little is known regarding the regulation of homing endonuclease expression. In particular, many freestanding endonuclease genes in phage genomes lack recognizable promoters or ribosome binding sites (RBSs), raising the question of how these genes are expressed. Furthermore, because many freestanding endonuclease genes are inserted between conserved and functionally critical phage genes (13, 40, 48), the impact of endonuclease gene insertion on the transcriptional regulation of neighboring genes is an outstanding question.

Transcriptional regulation plays a critical role in T-even phage development by determining the temporal order in which phage genes are expressed postinfection (reviewed in reference 40). Phage T4, for instance, executes a well-documented takeover of the transcriptional machinery of Escherichia coli by subverting the host RNA polymerase to transcribe phage genes preferentially over E. coli genes (40). Three temporal classes of transcripts, regulated by early, middle, and late phage promoters, orchestrate the synthesis of phage genes. Among the phage genes that are transcribed early after infection are those whose products are involved in the synthesis of precursors for DNA replication (6, 8), including the nrdA and nrdB genes encoding the large and small subunits of the class la aerobic ribonucleotide reductase (RNR), respectively (59, 60). In phage T4-infected cells, transcription of T4 nrdA and nrdB is tightly coordinated to ensure the maximal level of RNR activity early in T4 infection before the onset of DNA replication (59, 60).

Interestingly, the well-conserved nrd genomic region of T-even-like phage is a common target of homing endonucleases, as evidenced by the occurrence of both intron-encoded and freestanding homing endonuclease genes in a number of T-
even phages (39, 40, 44, 51, 52). Of particular interest is the freestanding HNH family endonuclease gene mobE, inserted in the nrdA-nrdB intergenic region of phages T4, T6, RB2, RB3, RB15, and LZ7 (44, 51). Genetic evidence suggests that mobE spreads between genomes, as crosses between phages containing mobE and phages lacking mobE revealed a >95% frequency of inheritance of mobE in progeny (33, 51). MobE likely possesses a recognition and cleavage site within or near the nrdA or nrdB coding region.

We recently described a novel gene arrangement created by the insertion of the mobE gene into the nrdA coding region of *Aeromonas hydrophila* T-even-like phage Aeh1 (18). The insertion fragments the Aeh1 nrdA gene at the active site, creating two smaller genes (*nrdA-a* and *nrdA-b*) that each encode active-site residues of RNR. The *mobE* insertion is not a self-splicing intron or intein and, despite the absence of splicing, is not inactivating for NrdA function. We showed that the NrdA-a and NrdA-b proteins form a complex with the small subunit NrdB protein, reconstituting a functional class Ia RNR activity by creating a composite active site with each protein providing functionally critical residues.

Here, we investigate the Aeh1 nrd operon with the goal of elucidating the regulation of mobE from the standpoint of its effect on expression of the surrounding nrd genes. Our transcriptional data, the first for phage Aeh1, suggest that regulation of the nrd and mobE genes employs a different strategy than does that of the corresponding genes of phage T4. Furthermore, we present evidence that expression of Aeh1 mobE is subject to strong negative regulation that limits MobE function to late in the Aeh1 infective cycle. We suggest that the multiple layers of control that regulate mobE function are adaptations of phage Aeh1 to limit the consequences of the mobE insertion into a critical gene of nucleotide metabolism.

**Materials and Methods**

**Bacterial strains, phage, and growth conditions.** *Escherichia coli* DH5α was used for plasmid construction and propagation, while strain INVαF' (Invitrogen) was used for the cloning of 5'-RLM-RACE (RNA ligase-mediated rapid amplification of cDNA ends) products. *E. coli* strains were grown in LB medium supplemented with the appropriate antibiotics (ampicillin, 100 μg/ml; kanamycin, 50 μg/ml). *A. hydrophila* strain C-1 was used to propagate bacteriophage Aeh1 in tryptic soy broth medium (EMD Bioscience) as previously described (18).

**Isolation of Aeh1 genomic DNA.** Genomic DNA was extracted from 500 μl of a 2 × 10^11 PFU/ml phage stock. The phage were mixed with equal volumes of phenol-chloroform, mixed for 5 min by inverting the tube repeatedly, and spun for 5 min at 5,000 × g. This was repeated four times before ethanol precipitation and resuspension in 50 μl Tris-EDTA buffer (10 mM Tris-Cl, pH 8, 1 mM EDTA). PCR. All amplification reactions were performed with a Biometra Thermocycler programmed for 35 cycles with annealing temperatures specific for each primer pair. Products were amplified from Aeh1 genomic DNA (gDNA) using Taq DNA polymerase (New England Biolabs [NEB]) and purified using a QIAGEN PCR purification kit or purified from agarose gels using a QIAGEN gel purification kit according to the manufacturer’s instructions. Primers were designed manually using the Aeh1 genome sequence (NC_005260), and a complete list of all primers used can be found in Table 1.

**RNA extraction and purification.** Samples (3 ml) of Aeh1-infected cultures were taken at various times postinfection, while the 0-min sample was uninfected bacteria. Immediately after collection, samples were suspended in an equal volume of RNA Later (Ambion) and stored on ice. Infected cells were pelleted by centrifugation for 5 min at 5,000 rpm at 4°C. RNA extractions for the time course were performed simultaneously by resuspending pelleted cells in Tris-EDTA buffer supplemented with 400 μg/ml lysozyme. RNA was extracted using a QIAGEN RNEasy minikit according to the manufacturer’s instructions. RNA was eluted in 80 μl RNase-free water, and a 5-μl aliquot was analyzed on a 1% agarose gel to determine RNA quality and integrity. To remove contaminating DNA, RNA was treated with TURBO DNase (Ambion) according to the manufacturer’s protocol, ethanol precipitated, and resuspended in RNase-free water.

**RT-PCR.** Reverse transcription-PCR (RT-PCR) was performed using 5 μl of total RNA and 20 pmol of primer. Reaction mixtures were incubated at 25°C for 5 min, 37°C for 60 min, and 23°C for 10 min with Moloney murine leukemia virus reverse transcriptase (NEB) in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, and 2 mM dithiothreitol (DTT). A 5-μl aliquot of each reverse transcription reaction mixture served as the template for the amplification of cDNA. PCRs were performed using *Taq* DNA polymerase (NEB) in supplied buffer. Cycling conditions for the amplification of cDNA were as follows: 94°C for 45 s, 42°C for 45 s, and 72°C for 60 s, for 35 cycles. Ampiclons were resolved on 1% agarose gels in 1× Tris-borate-EDTA buffer (89 mM Tris, 89 mM borate, and 2 mM EDTA).

**Primer extension.** Primers (Table 1) used for primer extension and cycle sequencing reactions were 5’ labeled using T4 polynucleotide kinase (PNK). Reaction mixtures consisting of 20 pmol primer, 125 μCi [γ-^32P]ATP, and 10 μl PNK (NEB) were incubated at 37°C for 20 min in supplied buffer. PNK was heat inactivated at 90°C for 3 min. The 5’ terminal of early and late transcripts were determined by annealing 1 pmol of radiolabeled primer to 10 μg total RNA in 5 μl primer extension buffer (50 mM Tris-HCl, 50 mM KCl, 10 mM MgCl₂, 10 mM DTT, 1 μM [each] deoxynucleotides triphosphate [dNTP], 0.5 mM spermidine). The mixture was denatured for 3 min at 90°C and then hybridized for 20 min at 51.7°C. Reverse transcription was carried out at 42°C for 1 h in primer extension buffer (as above) with 2 mM sodium pyrophosphate, 200 μM Moloney murine leukemia virus reverse transcriptase (NEB), and 20 U RNase inhibitor (NEB). The reaction mixtures were ethanol precipitated, resuspended in 5 μl H₂O, and digested with 5 μl of Rnase H in 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl₂, and 10 mM DTT at 37°C for 20 min. The reactions were stopped with 10 μl stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). Cycle sequencing reactions (USB) were performed on the corresponding PCR fragment of the Aeh1 genome using the same end-labeled primer as used for the primer extension reaction. The sequencing products were resolved on a 6% (wt/vol) denaturing polyacrylamide gel (19:1 acrylamide-bisacrylamide) and visualized using a PhosphorImager (GE Healthcare).

**Northern hybridization.** Total RNA (5 μg) was glyoxalated using glyoxal sample load dye (Ambion) and resolved on 1% agarose gel in 1× TBE buffer (10 mM PIPES [piperazine-N₂,N₂-bis(2-ethanesulfonic acid]), 30 mM Bis-Tris, 10 mM EDTA). Using downward alkaline transfer, the RNA was fixed to a positively charged Biodyne nylon membrane (Pall Corporation) according to the methods of Sambrook and Russell (90a). Membranes were soaked in 20 mM Tris-HCl (pH 8.0) prior to prehybridization. Prehybridization was carried out for 2 h at 68°C in 0.5 M sodium phosphate (pH 7.2), 7% (wt/vol) sodium dodecyl sulfate (SDS), 1 mM EDTA (pH 7.0). Radiolabeled probes were generated using the Nick Translation System (Invitrogen) with 1 μg of gel-purified PCR template amplified using the primers listed above to the manufacturer’s instructions. A 500-bp double-stranded DNA probe (80 ng) was denatured at 100°C for 5 min before hybridization at 42°C overnight. Blots were washed once in 0.1 × SSC (1 × SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.1% SDS at 42°C and three times in 0.5 × SSC-0.1% SDS at 68°C. The membrane was air dried briefly and wrapped in Saran Wrap. Images were visualized using a PhosphorImager (GE Healthcare).

**5’ RLM-RACE.** Tobacco acid phosphatase (TAP; Epicentre) was used to remove the γ and β phosphates from 5’ termini of 13 μg RNA in 50 mM sodium acetate (pH 5.0), 0.1% 1-mercaptoethanol, 1 mM EDTA, and 0.01% Triton X-100. Non-TAP-treatment control reactions were performed by replacing TAP with nuclease-free water. An RNA adapter (DE-193) was ligated to the 5’ terminal of 2.5 μg of TAP-treated or non-TAP-treated RNA using T4 RNA ligase (NEB) in supplied buffer. Ligated RNA was then purified from excess, unligated adapter oligonucleotide using RNAeasy minicolumns (QIAGEN) according to the manufacturer’s instructions. RT-PCR was carried out as described above using DE-198 (for nrdA-a) or DE-200 (for mobE) and a 5-μl aliquot of 5’ adapter-ligated RNA. Gene-specific amplification of cDNA was carried out as described above using DE-196/DE-198 (for nrdA-a 5’ termini) or DE-196/DE-200 (for mobE 5’ termini) and the following cycling conditions: 94°C for 30 s, 50°C (nrdA-a) or 58°C (mobE) for 30 s, and 72°C for 60 s, for 35 cycles. A 5-μl aliquot of this reaction mixture was used as a template for nested PCR using primers RT-197/RT-199 (for nrdA-a) or DE-197/DE-201 (for mobE termini). Cycling conditions for nested PCR were as follows: 94°C for 30 s, 50°C (nrdA-a) or 58°C (mobE) for 30 s, and 72°C for 60 s, for 35 cycles. Amplicons were gel purified and cloned into pCR2.1 (Invitrogen). Ten positive clones were selected and sequenced.

**Southern hybridization.** Aeh1 total DNA (5 μg) was digested with restriction enzymes, separated on agarose gels and blotted to nylon membranes as described above. Membranes were probed using cDNA fragments of the Aeh1 mobE gene as probes.
TABLE 1. Primers used in this study

<table>
<thead>
<tr>
<th>Name</th>
<th>Purpose</th>
<th>Oligonucleotide sequence (5′-3′)</th>
<th>Position in Aeh1 genome (AY266303.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE-3</td>
<td>Northern probe; mobE PCR</td>
<td>GGAATTTCCATATGAATTATAGAAAGAATACGCG</td>
<td>43162–43139 + Ndel site</td>
</tr>
<tr>
<td>DE-4</td>
<td>Northern probe; mobE PCR</td>
<td>CGGAGTCTCATATGCTGGAGTCTATTCGCG</td>
<td>42424–42466 + BamHI site</td>
</tr>
<tr>
<td>DE-15</td>
<td>Northern probe; ndrA-a PCR</td>
<td>GGAGGACATGGTTAATGAGAATATAGTGG</td>
<td>44630–44666 + Pml site</td>
</tr>
<tr>
<td>DE-16</td>
<td>Northern probe; ndrA-a PCR</td>
<td>GGGATCTTTGATATCTCAGAGACTTATGAC</td>
<td>43207–43232 + BamHI site</td>
</tr>
<tr>
<td>DE-25</td>
<td>RT-PCR; ndrA-a/mobE junction; cycle sequencing; mobE flanking region</td>
<td>TCGTTGATGGAATTCCAGCG</td>
<td>43397–43378</td>
</tr>
<tr>
<td>DE-26</td>
<td>RT-PCR; ndrA-a/mobE junction; cycle sequencing; mobE flanking region</td>
<td>GATAGTCAAGACGTTATCG</td>
<td>42268–42287</td>
</tr>
<tr>
<td>DE-109</td>
<td>RT-PCR; ndrA-a/mobE junction</td>
<td>TCACCTCCCATACATTGTTG</td>
<td>43054–43073</td>
</tr>
<tr>
<td>DE-110</td>
<td>Primer extension; ndrA-a; cycle sequencing; ndrA-a flanking region</td>
<td>ATCGATTGAGTTCGAGCAAG</td>
<td>44533–44552</td>
</tr>
<tr>
<td>DE-111</td>
<td>RT-PCR; ndrA-a 5′ flanking region; cycle sequencing; ndrA-a flanking region</td>
<td>TCTCAGCTGAGAATTCCAG</td>
<td>44887–44868</td>
</tr>
<tr>
<td>DE-128</td>
<td>Primer extension; mobE</td>
<td>TTTGTAACCGGAGATCTCC</td>
<td>43092–43111</td>
</tr>
<tr>
<td>DE-131</td>
<td>RT-PCR; ndrA-b/mobE junction; RPA probe template; ndrA-b PCR</td>
<td>TGAAGCCTGCTATCTTCTGAGTATTTGAC</td>
<td>43120–41191 + SacI site</td>
</tr>
<tr>
<td>DE-143</td>
<td>3′ RACE; ndrA-b PCR</td>
<td>CGGGAGCTCTATATGCTGGAGTCTATTTGAC</td>
<td>43120–41191 + SacI site</td>
</tr>
<tr>
<td>DE-145</td>
<td>RT-PCR; ndrB PCR; Northern probe; ndrB PCR</td>
<td>TCGGATTCTATGCTCTTGAGTATTTGAC</td>
<td>43120–41191 + SacI site</td>
</tr>
<tr>
<td>DE-146</td>
<td>3′ RACE; ndrB PCR1; Northern probe; ndrB PCR</td>
<td>GATAGTCAAGACGTTATCG</td>
<td>40375–40394</td>
</tr>
<tr>
<td>DE-153</td>
<td>3′ RACE; ndrB nested PCR</td>
<td>CGTGCGAAGAATGCTGCAT</td>
<td>42071–42052</td>
</tr>
<tr>
<td>DE-164</td>
<td>RT-PCR; ndrB 3′ flanking region; 3′ RACE; ndrB nested PCR</td>
<td>CAGAAATCGAGAAGATGCTGCAT</td>
<td>40609–40677</td>
</tr>
<tr>
<td>DE-165</td>
<td>RT-PCR; ndrB 3′ flanking region; 3′ RACE; ndrB nested PCR</td>
<td>CAGAAATCGAGAAGATGCTGCAT</td>
<td>40609–40677</td>
</tr>
<tr>
<td>DE-193</td>
<td>RNA adaptor</td>
<td>GCUCGUAGUAGGAAGUAAGAAGACACGUCAUGUUGC</td>
<td>NA</td>
</tr>
<tr>
<td>DE-194</td>
<td>3′ RACE; adaptor-specific oligonucleotide</td>
<td>TTTTCAATGAAAGCAGCAGAAGACG</td>
<td>NA</td>
</tr>
<tr>
<td>DE-195</td>
<td>3′ RACE; adaptor-specific oligonucleotide</td>
<td>CAAACGATGTTGTATCATTCAGGCC</td>
<td>NA</td>
</tr>
<tr>
<td>DE-196</td>
<td>5′ RACE; adaptor-specific oligonucleotide</td>
<td>CAGTGATGAGGATAGTAAAGCAGTACG</td>
<td>NA</td>
</tr>
<tr>
<td>DE-198</td>
<td>5′ RACE; ndrA specific cDNA synthesis, PCR1</td>
<td>TTCTTATCATGATACGCTGC</td>
<td>44291–44310</td>
</tr>
<tr>
<td>DE-199</td>
<td>5′ RACE; ndrA nested PCR</td>
<td>TACATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-204</td>
<td>RPA probe template; ndrA-b PCR2</td>
<td>TACATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-205</td>
<td>RPA probe template; ndrA-b PCR2</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-206</td>
<td>RPA probe template; ndrA-b PCR1</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-207</td>
<td>RPA probe template; ndrA-b PCR1</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-208</td>
<td>RPA probe template; ndrB PCR2</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-209</td>
<td>RPA probe template; ndrB PCR2</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-211</td>
<td>5′ RACE; mobE-specific cDNA synthesis, PCR1</td>
<td>ATCATTTGCGAGCAGAGC</td>
<td>44384–44403</td>
</tr>
<tr>
<td>DE-212</td>
<td>5′ RACE; mobE nested PCR</td>
<td>TCTTTTTTGGAGAATGCTACTCCAG</td>
<td>42266–42280</td>
</tr>
<tr>
<td>DE-214</td>
<td>RT-PCR; mobE/ndrA-b junction</td>
<td>GAAACTCTATAGGCTTGTTG</td>
<td>42651–42632</td>
</tr>
</tbody>
</table>

* NA, not applicable.

**3′ RLM-RACE.** An RNA adapter (DE-193) was ligated to the 3′ terminus of 13 µg of total RNA using T4 RNA ligase (NEB) in supplied buffer. Ligated RNA was purified from excess adapter oligonucleotide using RNAasy minicolumns (Qiagen) according to the manufacturer’s instructions. RT-PCR was carried out as described above using DE-194 and a 5′-32P labeled RNA. Gene-specific amplification of cDNA was carried out as described above using DE-194/DE-143 (for ndrA-b termini) or DE-194/DE-146 (for ndrB termini) and the following cycling conditions: 94°C for 30 s, 55°C for 30 s, and 72°C for 60 s, for 35 cycles. A 5′-32P labeled RNA, synthesized using T7 RNA polymerase, was used as a template for nested PCR using primers DE-195/DE-153 (for ndrA-b termini) or DE-195/DE-164 (for ndrB termini). Cycling conditions for nested PCR were as follows: 94°C for 30 s, 50°C for 30 s, and 72°C for 60 s, for 35 cycles. Amplicons were gel purified, cloned into pCR2.1 (Invitrogen), and sequenced.

**RNase protection.** RNase protection assays (RPAs) were performed according to the manufacturer’s instructions (Ambion). The antisense probe template was generated by PCR from genomic Aeh1 DNA using primers DE-132/DE-131 (ndrA-b termini) and DE-206/DE-207 (ndrB termini) with the following cycling conditions: 94°C for 45 s, 45°C for 15 s, and 72°C for 60 s, for 35 cycles. A T7 promoter and additional nonhomologous sequence were incorporated by a second round of PCR using the above antisense template and primers DE-204/DE-205 (ndrA-b termini) or DE-208/DE-209 (ndrB termini) using the following cycling conditions: 94°C for 30 s, 58°C for 30 s, and 72°C for 60 s, for 35 cycles. The labeled RNA probes were transcribed in 50-µl volumes consisting of 50 µCi [α-32P]UTP, 1 µl PCR template, 50 U T7 RNA polymerase (NEB), and supplied buffer. The reaction mixture was incubated at 37°C for 2 h before 2 U of Turbo-DNase (Ambion) was added, and the incubation continued for 30 min. The resultant RNA probes were gel purified from a 5% denaturing polyacrylamide gel. Total RNA (7.5 µg) was hybridized overnight with purified RNA probes (28,000 cpm of each probe) at 42°C in supplied hybridization buffer. Control reactions used 7.5 µg of yeast RNA. Hybridized probe was digested with an RNase A-T1 mixture in supplied digestion buffer for 30 min. RNAses were inactivated, and the protected RNA was precipitated. The sample was analyzed by electrophoresis through a 6% denaturing polyacrylamide gel. Images were visualized on a PhosphorImager, and the amount of transcriptional readthrough and termination was estimated using ImageQuant software (GE Healthcare).

**Promoter predictions.** Early and late phase Aeh1 promoters were predicted by extracting 100 bp upstream and downstream of the start codon of Aeh1 genes that are homologous to phage T4 genes that are transcribed at early and late times post-T4 infection. A training model was generated for Aeh1 early- and late-transcribed genes using the Gibbs Motif Sampler in recursive sampler mode.
This training model was then used in an unbiased search to scan the Aeh1 genome sequence for early and late promoters using DSCAN (31, 43). A user-generated program converted the Gibbs DSCAN output into an alignment file that was used to generate sequence logos (11).

RESULTS

Overview of regulatory elements in the Aeh1 nrd genomic region. The class Ia aerobic RNR genes of phage Aeh1 are found in an operon-like arrangement, with each gene in the same transcriptional orientation and each gene preceded by an RBS (Fig. 1A) (48). The most significant difference in the organization of this region between Aeh1 and other T-even-like phages is the location of mobE, an HNH family homing endonuclease gene (29). In phage Aeh1, mobE is inserted in the Aeh1 nrdA gene, splitting the large subunit gene into two smaller genes, nrdA-a and nrdA-b (48), whereas in phages T4, T6, RB2, RB3, RB15, and LZ7, mobE is inserted between the nrdA and nrdB coding regions (33, 48, 51).

To investigate the transcriptional organization of the Aeh1 nrd genomic region, we identified putative promoters based on our own and published computational predictions (44). We identified two potential promoters, an early promoter upstream of the Aeh1 nrdA-a gene and a late promoter upstream of the Aeh1 mobE gene (Fig. 1A and B). The Aeh1 nrdA-a early promoter is comprised of two boxes separated by 17 nucleotides (nt), with a UP element upstream of the box (Fig. 1B) (16, 50). The box differs from the canonical T4 early promoter in that a Gi is preferred at position −37 rather than a T (44). The late promoter upstream of mobE is very similar to consensus T4 late promoters (9, 40), except for a slight preference for a C at position −25 (Fig. 1B), which is atypical for T4 late...
promoters. We could not identify an early or late promoter sequence upstream of the \textit{nrdB} gene. Significantly, no middle-like promoter in the \textit{Aeh1} genome could be identified by bioinformatics methods, consistent with previous reports indicating that \textit{Aeh1} does not possess the middle-mode transcriptional machinery found in phages T4, RB69, and 44RR (44).

We examined the regions downstream of each gene in the \textit{Aeh1 nrd} operon for Rho-independent transcriptional terminators (25, 49, 63) and found putative terminators downstream of genes \textit{50}, \textit{nrdA-b}, and \textit{nrdB} (Fig. 1A and C). The \textit{nrdA-b} and \textit{nrdB} terminators are 5 to 6 nt downstream of each gene's stop codon, while the stop codon of gene \textit{50} lies within the loop region of the predicted terminator. All three terminators consist of putative 5-bp stems with a 4-nt tetraloop (Fig. 1C). Although the stem structures are not similar in sequence, the tetraloop sequences of the two terminators are identical. A short 4-nt poly(U) tract follows the \textit{nrdA-b} terminator, while longer tracts follow the terminator predicted downstream of genes \textit{50} [6-nt poly(U) tract] and \textit{nrdB} [10-nt poly(U) tract].

We also identified a putative stem-loop structure immediately upstream of the \textit{mobE} AUG codon in the intergenic region separating \textit{nrdA-a} and \textit{mobE} (Fig. 1D). This stem-loop structure does not possess features characteristic of Rho-independent terminators. Rather, this RNA hairpin has a predicted role in regulating translation of MobE, as the \textit{mobE} RBS is sequestered within the hairpin. The predicted \textit{mobE} late promoter is positioned such that late transcripts would not include sufficient sequence to form a stable stem-loop structure and sequester the \textit{mobE} RBS.

**Promoters of two temporal classes regulate expression of the \textit{Aeh1 nrd} operon.** We used primer extension analysis to confirm the predicted \textit{Aeh1} early and late promoters upstream of the \textit{nrdA-a} and \textit{mobE} genes, respectively. To map early transcripts upstream of the \textit{nrdA-a} gene, we isolated total RNA from \textit{A. hydrophila} before \textit{Aeh1} infection and at various times post-\textit{Aeh1} infection. Primer extension analysis revealed a transcript that initiated at nucleotide C-24 relative to the \textit{nrdA-a} gene (Fig. 2A). This transcript was detected as early as 1 min postinfection (Fig. 2B, control, lane 2) but not in uninfected \textit{A. hydrophila} extracts (Fig. 2B, control, lane 1). Early transcripts persisted over the time course of the phage infection and remained detectable at 50 min postinfection (Fig. 2B, control, lane 8). To confirm the initiating nucleotide, we used primer extension analysis to map the 5′-RLM-RACE analysis of \textit{nrdA-a} initiated transcripts aligned with the genomic sequence (gDNA).

**FIG. 2.** Mapping of transcript initiation sites upstream of the \textit{nrdA-a} gene. (A) Primer extension mapping of the transcription start site using primer DE-110 and RNA isolated 20 min post-\textit{Aeh1} infection. An aliquot of the primer extension reaction mixture was electrophoresed alongside a sequencing ladder of the \textit{nrdA-a} upstream region. The initiating nucleotide is identified by an asterisk. (B) (Top) Primer extension analysis of transcription initiation at the \textit{nrdA-a} promoter at various times post-\textit{Aeh1} infection. (Bottom) Primer extension analysis of RNA isolated after rifampin treatment of \textit{Aeh1}-infected cells. Rifampin was added 5 min post-\textit{Aeh1} infection (indicated by a plus sign). The sequence of the \textit{nrdA-a} upstream region is shown below with positions of the −35 and −10 boxes of the early promoter and transcriptional start site indicated relative to the \textit{nrdA-a} ATG codon. The RBS is underlined. (C) Partial sequences of four clones from 5′-RLM-RACE analysis of \textit{nrdA-a} initiated transcripts aligned with the genomic sequence (gDNA).

Downloaded from http://jb.asm.org/ on October 31, 2017 by guest
scriptional initiation, we treated Aeh1-infected cells with rifampin at 5 min postinfection to inhibit the host RNA polymerase before isolating RNA for primer extension analysis (Fig. 2B, rifampin treated). Because Aeh1 does not encode its own RNA polymerase (44, 48), transcription of phage genes is dependent on the host RNA polymerase. Primer extension analysis with rifampin-treated RNA revealed an extension product that accumulated until 10 min post-Aeh1 infection.

**FIG. 3.** Mapping of transcript initiation sites upstream of the **mobE** gene. (A) Representative primer extension mapping of the transcription start site using primer DE-120, electrophoresed alongside a sequencing ladder generated with the same primer. The sequence of the **mobE** upstream region and the structure of the predicted stem loop are indicated to the right of the gel. Potential initiating nucleotides (T-23 and G-22) are indicated by arrows, while the G-10 and T-9 sites are indicated by asterisks. (B) Experimental outline to distinguish initiated from processed transcripts using TAP as elaborated in the text. P, 5′ phosphate; open rectangle, RNA adaptor oligonucleotide. (C) Agarose gel of 5′-RLM-RACE reactions using RNA treated with (+) or without (−) TAP. Aliquots of the mixture from the final nested PCR step were electrophoresed alongside a pBR322/AluI ladder. (D) Sequences of five clones corresponding to each of the 5′-RLM-RACE products in panel B. Only partial sequences of the clones, and putative RNase E sites are indicated. (E) Summary of transcript mapping and 5′-RLM-RACE data indicating that late-initiated transcripts would not include sufficient sequence to form a regulatory hairpin that sequesters the **mobE** RBS.

VOL. 189, 2007
REGULATION OF HOMING ENDONUCLEASE **mobE** EXPRESSION
4653

on October 31, 2017 by guest
http://jb.asm.org/ Downloaded from
Transcription initiation sites upstream of the mobE gene were also determined using primer extension analysis (Fig. 3A). Aeh1-specific transcripts initiating upstream of mobE were detected at 15 min postinfection but not at earlier time points, consistent with a predicted phage-specific late promoter (Fig. 3A, compare lanes 3 to 6 with lanes 1 and 2). Similar to the early transcripts, the late transcripts also persisted over the time course of the phage infection (Fig. 3A, lane 6). Two potential initiation sites were mapped to T-23 and G-22 relative to the mobE ATG codon, respectively (Fig. 3A). Two other primer extension products that map to G-10 and T-9 appear at 5 min post-Aeh1 infection and persist throughout the time course of the infection. The early appearance of the G-10 and T-9 extension products is inconsistent with initiation from a late promoter. Furthermore, the G-10 and T-9 sites are located near the RBS upstream of mobE, suggesting that these transcripts would not be translated efficiently.

**The mobE transcript is processed at an RNase E-like site.**

To confirm the initiating nucleotide at the mobE late promoter, we used 5’ RLM-RACE to identify the 5’ end of the transcript and to determine if the G-10 and T-9 sites mapped by primer extension resulted from posttranscriptional processing. Posttranscriptional processing can be detected by 5’ RLM-RACE because treatment of RNA with TAP removes the γ and β phosphates from RNA with a 5’-triphosphate end but does not remove the α phosphate (5, 19). In the subsequent ligation step of the 5’ RLM-RACE procedure, only RNA molecules with a single (α) 5’ phosphate are substrates for T4 RNA ligase (Fig. 3B). Thus, initiating transcripts that possessed a 5’ triphosphate would be detected only in RNA samples that were TAP treated, whereas RNA molecules that possessed a single 5’ phosphate would be detected in both TAP− and TAP+ samples. As seen in Fig. 3C, two bands of ~280 bp (band A) and ~220 bp (band B) were amplified from RNA samples that were treated with TAP prior to 5’ RLM-RACE (Fig. 3C, lane 1). In contrast, a single band of ~220 bp was amplified from RNA that was not treated with TAP prior to 5’ RLM-RACE (Fig. 3C, lane 2). To determine the 5’ ends of each of the amplified fragments, the three bands were separately excised, cloned, and sequenced.

As shown in Fig. 3D, the sequences of five clones corresponding to the larger of the two bands (band A) in the TAP+ sample were in agreement with the primer extension analysis that mapped the initiating nucleotide to G-22. Surprisingly, the 5’ ends of sequences of five clones from the smaller (band B) of the two amplified products in the TAP+ sample all mapped to an A-U-rich region within the mobE coding region, ~60 nt away from G-22. The 5’ ends of clones corresponding to the single amplified product in the TAP− sample also mapped to the same A-U-rich region (Fig. 3D).

Collectively, the 5’ RLM-RACE results show that the initiating nucleotide of the mobE late promoter is G-22 and that the other two potential transcript initiation sites (G-10 and T-9) mapped by primer extension analyses are not true initiation sites and likely result from reverse transcriptase pausing at secondary structures in the region of the regulatory hairpin upstream of mobE. Furthermore, the lack of 5’ RLM-RACE products that map to the G-10 and T-9 sites suggests that these sites do not represent posttranscriptional processing products with 5’ monophosphates. These data also indicate that late-initiating transcripts would not include sufficient sequence to form a stem-loop structure to sequester the mobE RBS (Fig. 3D).
Moreover, the amplification of a product in the TAP sample was unexpected, because only RNA transcripts that have been internally processed would possess a single 5-phosphate. Our data suggest that the mobE transcript is processed at an A-U-rich region that is adjacent to a hairpin, features that are characteristic of an RNase E processing site (15).

The nrd and mobE genes are transcribed on a polycistronic mRNA. To determine the sizes of transcripts that initiate at the early and late promoters, we used Northern hybridization with probes corresponding to the nrdA-a, mobE, and nrdB genes, respectively (Fig. 4). In all three instances, the radiolabeled probe detected a band of ~4 kb as early as 5 min postinfection, which remained detectable at >20 min postinfection (Fig. 4A, B, and C). This transcript is of sufficient length to carry the nrdA-a, mobE, nrdA-b, and nrdB genes on a polycistronic message and is similar in size to a predicted transcript of 4.3 kb. Detection of mobE-specific late transcripts is complicated by the fact that the mobE gene is also present on transcripts that initiate upstream of nrdA-a throughout the course of the phage infection (Fig. 2B). For each gene-specific probe, additional hybridizing bands were also observed. None of the bands, however, matched the predicted sizes of transcripts initiating from either the early promoter upstream of nrdA-a or the late promoter upstream of mobE and terminating at the nrdA-b or nrdB terminator. Likewise, transcripts that initiated at predicted early promoters upstream of genes 52 and 51 would have produced transcripts larger than 4.2 kb (Fig. 1A), which were not observed with any probe. It is unlikely that the additional bands resulted from spurious hybridization of the probes to A. hydrophila rRNA, because no signals were observed with RNA isolated before Aeh1 infection of A. hydrophila (lane 0 of each panel).

The Northern blot analyses suggested that the nrdA-a, mobE, nrdA-b, and nrdB genes were present on a common transcript, but the presence of additional hybridizing bands suggested that RNA processing or cross-hybridization might be responsible for the additional signals observed. To confirm that these genes are cotranscribed, we performed a series of RT-PCR experiments. Using primers designed against the nrdA-a, mobE, and nrdA-b coding regions, we amplified RT-PCR products of the predicted sizes with RNA isolated 30 min postinfection (Fig. 5B and C). This result indicates that the nrdA-a/mobE and mobE/nrdA-b genes are cotranscribed. Similarly, we showed that the nrdA-b and nrdB genes are cotrans-
scribed using two different primer sets (Fig. 5E and G). Collectively, the results of the RT-PCRs support the Northern blot analyses in showing that the nrdA-a, mobE, nrdA-b, and nrdB genes are cotranscribed.

Expression of the nrdB gene is dependent on transcriptional readthrough. The Northern blot and RT-PCR experiments also served to determine the functionality of the predicted transcriptional terminators in the nrd operon. In particular, the presence of a ~4-kb hybridizing band in Northern blots (Fig. 4), coupled with the amplification of a RT-PCR product spanning the junction of the nrdA-b and nrdB genes (Fig. 5E), suggested that the transcriptional terminator 3′ to the nrdA-b gene allowed a significant amount of transcriptional readthrough. Likewise, we were able to amplify RT-PCR products using primers that flanked the predicted transcriptional terminator downstream of the nrdB gene (Fig. 5D) and primers that flanked the predicted terminator between genes 50 and nrdA-a (Fig. 5A). RT-PCR, however, is a highly sensitive method that may detect rare transcriptional readthrough events that are otherwise not detectable by Northern hybridization.

To validate the functionality of the predicted terminators, we used 3′ RLM-RACE to map the 3′ ends of transcripts in the nrd genomic region (Fig. 6A). The sequences of 6 out of 10 clones revealed a termination event at the poly(U) tract immediately downstream of the nrdA-b terminator (Fig. 6B). Similarly, 4 out of 10 clones revealed transcriptional termination at the nrdB terminator, immediately following the poly(U) tract (Fig. 6C). In both cases, we amplified shorter products that mapped to sites 5′ to the predicted terminators, likely representing transcripts that were degraded during purification or transcripts that were partially processed from the 3′ end. These results, nonetheless, indicate that the transcriptional terminators downstream of nrdA-b and nrdB are functional.

We were interested in estimating the amount of transcriptional readthrough at the nrdA-b terminator, because transcriptional readthrough is the only mechanism by which the promoterless nrdB gene could be expressed. To estimate the level of transcriptional readthrough, we used an RPA with a probe that spanned the 3′ end of nrdA-b, the transcriptional terminator and intergenic region, and the 5′ end of nrdB (Fig. 6A). The probe also included 29 nt of nonhomologous sequence, which would allow us to distinguish full-length undigested probe (290 nt) from protected probe that corresponded to transcriptional readthrough (261 nt) or termination events (~133 nt based on 3′-RLM-RACE data). As seen in Fig. 6D, with RNA isolated at 5, 10, and 15 min postinfection, 85% of the protected fragment was 261 nt in length, indicative of transcription readthrough, whereas only 15% of the protected fragment was of the length predicted for termination.

Similarly, we used RNase protection to estimate the amount of transcriptional readthrough at the nrdB terminator (Fig. 6E). With RNA isolated at 15 min post-Aeh1 infection, 93% of the protected fragment was of the size expected for termination events (111 nt), with only 7% of the protected fragment of the size expected for a readthrough event (197 nt). Similar ratios of readthrough to termination events were found for RNA isolated at 20 and 40 min postinfection. This result indicating efficient termination at the end of nrdB is in stark contrast to that observed for the nrdA-b terminator, which indicated a significant amount of readthrough.

DISCUSSION

One of the most striking differences in the genomic organization of the nrdA-nrdB region of T-even-like phage is the presence or absence of homing endonuclease genes (39, 44, 48, 51). Here, we focus on the transcriptional organization of the Aeh1 nrd genes, as the mobE insertion has created a unique opportunity to examine the integration of a mobile endonuclease gene into an operon that is under stringent transcriptional control. Overall, we find that the Aeh1 nrd and mobE genes are regulated by a different transcriptional strategy than the corresponding T4 genes and that transcriptional readthrough plays a critical role in the regulation of the Aeh1 nrdB gene. Furthermore, we find evidence that Aeh1 mobE is subject to negative regulation, which limits MobE function to late in the Aeh1 infective cycle.

The genome sequence of phage Aeh1 identified homologs of T4 proteins that function to direct the host RNA polymerase to recognize early and late phage promoters preferentially over host promoters (44, 47). One critical difference, however, was the lack of middle-promoter-like promoters in Aeh1 and the absence of the middle-mode transcription factor, MotA, from the genome sequence, suggesting that Aeh1 does not possess a class of transcripts analogous to T4 middle transcripts. In our examination of the Aeh1 nrd operon, we identified promoters upstream of nrdA-a and mobE that were active at early and late time points, respectively, but were unable to identify any promoters analogous to T4 middle promoters. Significantly, primer extension analysis of the nrdA-a early promoter showed that the promoter remained active over the course of the phage infection and that the initiating nucleotide for transcription was C-22. Both of these observations are in contrast to T4 early promoters, which usually initiate at an A nucleotide (32, 61) and which are active 1 to 3 min post-T4 infection (35, 40). Likewise, the late promoter upstream of Aeh1 mobE is active 15 min postinfection, a significant delay compared with T4 late promoters, which are active 7 min postinfection (35, 40). Our transcriptional data are similar to those found for phages SP-M2 and RB49, which like Aeh1 possess only two transcriptional classes (early and late) and lack the middle-mode transcription machinery (10, 12, 37).

With respect to transcriptional regulation of the nrd genes, the most significant difference between T4 and Aeh1 is the lack of a middle-promoter-like promoter upstream of nrdB in Aeh1 (summarized in Fig. 7). In phage T4, expression of the nrdA and nrdB genes is regulated such that the NrdA protein appears ~1 to 2 min before the NrdB protein (reviewed in reference 23). Synthesis of NrdB is therefore the rate-limiting step in the onset of dNTP synthesis, which occurs ~5 min postinfection. The expression of nrdA is controlled by two early promoters and one late promoter (59, 60). A middle promoter was identified upstream of nrdA by bioinformatic methods (40) and recently confirmed by transcript mapping (56, 60). An immediate-early promoter is located upstream of the frd gene, ~4.1 kb from nrdA. Transcripts (Tu) from this promoter extend through frd, td, and nrdA and are detected as early as 2 min postinfection. At ~3 min postinfection, transcripts (T3) initiate from a weak early promoter immediately upstream of nrdA (60). These transcripts, however, could not be capped by guanylyl transferase and thus may represent products of a
FIG. 6. Transcription termination and readthrough at Rho-independent terminators downstream of the nrdA-b and nrdB genes. (A) Schematic of the nrdA-b and nrdB genes with the positions of probes used for RPAs. The sizes of protected fragments corresponding to termination and readthrough are indicated. (B and C) 3’-RLM-RACE sequence results of transcripts terminating at the nrdA-b and nrdB terminators, respectively. Partial sequences of cloned 3’-RLM-RACE products are aligned with the genomic sequences (gDNA) corresponding to the 3’ regions of the nrdA-b and nrdB genes, respectively. The stem-loop structure of each terminator is indicated on the genomic sequence. (D and E) RPAs showing the ratio of readthrough transcription to termination at various times at the nrdA-b and nrdB terminators, respectively. For each protection assay, aliquots of the reaction mixtures were electrophoresed alongside a labeled 100-bp ladder (M). The percentage of product corresponding to termination or readthrough is indicated below each gel. −, probe only; +, probe digested with RNase.
posttranscriptional processing event of the Tu transcript (60). Approximately two-thirds of transcripts terminate at a Rho-independent terminator immediately downstream of \textit{nrdA}, while the remaining transcripts continue through \textit{mobE} and \textit{nrdB} to create a “deoxyribonucleotide operon.” However, this long polycistronic message is not likely to represent a significant source of the NrdB protein, because the distance between \textit{nrdA} and the 3’ end of \textit{nrdB} is such that NrdB would not be translated until \( \sim 8 \) min postinfection (due to the length of the transcript and rate of translation), well after the onset of dNTP synthesis in phage-infected cells. In addition, splicing of the group I intron from \textit{nrdB} transcripts delays NrdB translation (45, 55). Thus, the T4 \textit{nrdB} gene is also under the control of a middle promoter that is active \( \sim 3 \) min postinfection (59), ensuring the appearance of the NrdB protein \( \sim 1 \) min after the appearance of the NrdA protein.

Transcription of the Aeh1 \textit{nrd} genes occurs by a different strategy, in part because the \textit{td} and \textit{frd} genes of Aeh1 are \( \sim 160 \) kb distant from the \textit{nrd} genes (48). Our results indicate that a \( \sim 4 \)-kb transcript initiates from an early promoter upstream of \textit{nrdA-a} that is sufficient in length to include the \textit{nrdA-a}, \textit{mobE}, \textit{nrdA-b}, and \textit{nrdB} genes. In addition, RT-PCR experiments indicate that the \textit{nrdA-a} gene is present on another transcript that likely initiates at one of two predicted early promoters upstream of gene 52 or 51. Transcripts that initiate at gene 52 or 51 could conceivably extend through the entire \textit{nrd} operon, but we could not detect hybridizing bands of \( > 4.2 \) kb by Northern analysis as would be expected if these transcripts extended through the \textit{nrd} genes to the Rho-independent terminators downstream of \textit{nrdA-b} or \textit{nrdB}. At \( \sim 15 \) min after Aeh1 infection, transcripts also initiate at a late promoter upstream of \textit{mobE} and presumably extend through \textit{nrdA-b} and \textit{nrdB}. Inter-
estingly, we do not observe a corresponding reduction in transcription from the nrdA-a early promoter, which complicates detection of late-time-specific mobE transcripts by Northern analysis because the mobE gene is present on both early and late messages. Nonetheless, our data show that transcription of the Aeh1 nrd operon is continuous throughout the infective cycle, representing a departure from transcription of the T4 nrd genes that are subject to stricter temporal control.

Studies of phage T4-infected cells have implicated the NrdB protein as the key regulator in the appearance of RNR activity (59, 60). For phage Aeh1, however, it is tempting to speculate that synthesis and posttranslational assembly of the NrdA-a and NrdA-b proteins into a complex are the rate-limiting step in the appearance of RNR activity, rather than synthesis of the NrdB protein. This hypothesis may account for the presence of a transcriptional terminator downstream of nrdA-b that would prevent an accumulation of nrdB message and protein before assembly of the NrdA-a/NrdA-b heterodimer. The late promoter upstream of mobE, which is active 15 min postinfection, may be analogous to the T4 nrdB middle promoter and act to increase Aeh1 nrdB message and protein levels to coincide with the formation of the NrdA-a/NrdA-b heterodimer.

The presence of a ~4-kb transcript shown by Northern blot analyses using either the nrdA-a or the nrdB gene as a probe suggested that the Rho-independent terminator downstream of nrdA-b is inefficient and allows a significant amount of readthrough, although our 3'-RLM-RACE data indicated that some transcription events terminate at this point. The inefficiency of this terminator highlights another key difference between the T4 and Aeh1 nrd genes, namely, that expression of the Aeh1 nrdB gene is solely dependent on readthrough transcription at the nrdA-terminator, from transcripts that initiate either at the nrdA-a early or at the mobE late promoter. Transcriptional readthrough has been reported in T4 and in some instances is the only mechanism of expression of promoterless or “orphaned” genes (27). We estimate from RNase protection assays that 85% of transcripts read through the Aeh1 nrdA-a terminator at all time points sampled post-Aeh1 infection. In contrast, termination is very efficient at the terminator downstream of the nrdB gene, with 93% of transcripts terminating at this point at all times sampled. The inefficiency of the nrdA-b terminator correlates with the short 4-nt poly(U) tract that follows the stem-loop structure. The length of the poly(U) tract has been shown to be critical in directing efficient termination in a number of experimental systems (1, 2, 62). Conversely, the efficiency of the nrdB terminator is positively correlated with the longer, 10-nt poly(U) tract that follows the stem-loop structure.

Few freestanding endonuclease genes in phage have been characterized in detail, but experimental evidence to date suggests that freestanding endonuclease genes have coevolved with the phage genome to minimize their impact on gene structure and function (33). The controls that we describe here for Aeh1-carried mobE include a late-regulated promoter that drives expression of mobE and a putative stem-loop structure that is predicted to sequester the mobE RBS in early transcripts that initiate upstream at the nrdA-a early promoter, presumably limiting MobE translation. This regulatory stem-loop structure would form only in early transcripts that extend through mobE, because transcripts that initiate at the late promoter upstream of mobE do not include enough RNA sequence to form a stable stem-loop to sequester the mobE RBS. The transcriptional and translation controls described for mobE are similar to those known for a number of phage T4 genes (24, 36, 38), including genes for the T4 intron-encoded endonucleases I-TevI, I-TevII, and I-TevIII (14, 22). Moreover, mobE appears to be subject to negative regulation in the form of posttranscriptional processing. We mapped an RNase E-like site in the mobE coding region, immediately upstream of a predicted hairpin that is characteristic of RNase E sites (15). RNase E is involved in posttranscriptional processing of a number of phage T4 genes (58), the most relevant to this work being the freestanding GIY-YIG endonuclease gene segG, which lies upstream of gene 32 in phage T4 (34). Interestingly, SegG-induced DSBs were detected in phage genomic DNA by Southern blot analysis (33), suggesting that RNase E processing does not abolish translation of the SegG message. It remains to be determined if RNase E is responsible for posttranscriptional processing of the mobE message and if the processing affects MobE protein levels. However, we have previously isolated a complex of the NrdA-a/NrdA-b/NrdB proteins from Aeh1-infected cells (18), indicating that posttranscriptional processing has little effect on translation of the nrdA-a, nrdA-b, and nrdB genes that are cotranscribed with mobE.

Thus, a prediction from our results is that Aeh1 mobE would be functional only at late time points during phage infection. It is possible that limiting translation from the mobE RBS on early transcripts maximizes translation of the nrdA-a, nrdA-b, and nrdB coding regions at a stage in the Aeh1 infective cycle when RNR function is critical. Alternatively, limiting MobE function to late in the infection cycle may correlate with the completion of DNA replication in Aeh1-infected cells and the availability of genome equivalents to facilitate repair of DSBs generated during mobE homing. Tight regulation of MobE expression may also correlate with the nonspecific DNA cleavage activity of the endonuclease, as we have shown elsewhere that overexpression of MobE from an inducible plasmid-based promoter in E. coli is extremely toxic (E. A. Gibb and D. R. Edgell, unpublished data). Late expression of endonuclease function during phage infection may be a general pattern, as transcription of T4 intron-encoded and freestanding endonucleases is also restricted to middle or late stages of phage infection (22, 35).

The mechanisms governing the regulation of T4 mobE have not been studied in detail, but it is interesting that T4 mobE does not apparently possess the same transcriptional and translational controls as does Aeh1 mobE. Transcript mapping of the T4 nrdA-nrdB region failed to locate a promoter that specifically drives expression of mobE but did identify a Rho-independent terminator within mobE that functions to regulate the expression of the upstream nrdA gene (59). These data suggest that expression of T4 mobE is dependent on readthrough transcription from the nrdA terminus, which occurs in one out of every three transcripts (59). Translation of T4 MobE, however, is likely an infrequent event because no readily identifiable RBS lies upstream of the mobE AUG codon (40), and thus translation may be dependent on translational coupling with the upstream nrdA gene.

In summary, our data show that expression of Aeh1 mobE is
temporally regulated by a late promoter that is active ~15 min postinfection. We also provide evidence for posttranscriptional processing of the mobE transcript by RNase E. In addition, a putative stem-loop structure sequesters the mobE RBS, presumably preventing its translation from early transcripts. Our primer extension data show that late transcripts would not include sufficient sequence to form the regulatory stem-loop, likely freeing the RBS and facilitating translation of MobE at late time points. It is tempting to speculate that these controls have evolved as a response to the mobE invasion of the Aehl ndA gene, because similar controls do not appear to exist for the mobE gene in phage T4 that is located in the ndA-ndB intergenic region. A recent survey of sequenced T-even-like phage genomes for genes that had undergone lateral gene transfer found only a single candidate, the ndA gene (17). This may not be surprising given the prevalence of mobile endonucleases in the ndA-ndB region of a number of T-even-like phages and the possibility that homing endonucleases can shuffle DNA between genomes due to coevolution of flanking sequence that accompanies an endonuclease-mediated mobility event (7, 26, 33, 42). Thus, the RNR genes of phage may prove to be a useful model for studying the evolution and function of a critical enzyme of nucleotide metabolism and also provide insight into the mechanism(s) by which mobile endonucleases integrate into transcriptional programs with minimal impact on the regulation of RNR function.

ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Institutes of Health Research (MOP77779) to D.R.E. We thank Gavin Wilson for assistance with promoter predictions and David Hanford for discussion and reading of the manuscript.

REFERENCES


2. Abe, H., and H. Aiba. 1996. Differential contributions of two elements of rho-independent terminator to transcription termination and mRNA stabi-


6. Berglund, O. 1975. Ribonucleoside diphosphate reductase induced by bac-


Tsong, M. J., P. He, J. M. Hilfinger, and G. R. Greenberg. 1990. Bacterio-


