The transcription programme of the protein-primed halovirus SH1

Kate Porter,1 Brendan E. Russ,2 Ji Yang2 and Michael L. Dyall-Smith3

1Biota Holdings Ltd, 10/585 Blackburn Road, Notting Hill, Victoria 3168, Australia
2Department of Microbiology and Immunology, University of Melbourne, 3052, Australia
3Max-Planck Institute for Biochemistry, Department of Membrane Biochemistry, Martinsried 82152, Germany

SH1 is the only reported isolate of a spherical halovirus, a dominant morphotype in hypersaline lakes. The virus lytically infects the haloarchaeon Haloarcula hispanica, and carries a 30.9 kb linear dsDNA genome that, in a previous study, was proposed to contain 56 protein-coding genes, probably organized into between four and eight operons. In the present study, these predictions were directly tested by determining the orientations and lengths of virus transcripts using systematic RT-PCR and primer extension. Seven major transcripts were observed that together covered most of the genome. Six transcripts were synthesized from early in infection (1 h post-infection; p.i.) onwards, while transcript T6 was only detected late in infection (5–6 h p.i.). No transcripts were detected in the inverted terminal repeat sequences or at the extreme right end of the genome (ORFs 55–56). Start points for the major transcripts were mapped by primer extension and corresponded closely to the 5’ termini determined by RT-PCR. Between 1 and 4 h p.i., transcripts usually terminated not far beyond the end of their last coding ORF, but late in infection, transcripts from the same promoters often terminated at more distal points, resulting in much of the genome being transcribed from both strands. Since many of these transcripts are complementary, RNA–RNA interactions are likely, and may play a role in regulating viral gene expression. Puromycin blockage of post-infection protein synthesis significantly altered the levels of certain virus transcripts, indicating that de novo protein synthesis is essential for the correct regulation of SH1 gene expression.

INTRODUCTION

Viruses are ubiquitous, extremely diverse and occur in large numbers throughout the biosphere (Hendrix, 2002; Hendrix et al., 1999). In aqueous environments, virus concentrations are typically 10-fold or higher than that of the microbial cells, upon which they effect a range of selective pressures (Bouver & del Giorgio, 2007; Weinbauer, 2004). While marine and fresh waters have been relatively well studied, far less is known about hypersaline waters, such as salt lakes, where salt concentrations hover around saturation. Under such harsh conditions the microbial diversity is severely restricted, mainly to extremely halophilic Archaea (Family Halobacteriaceae).

Nevertheless, the cell densities are often very high and the virus concentrations even higher (Benloch et al., 2002; Burns et al., 2004; Guixa-Boixareu et al., 1996; Oren et al., 1997). The majority of known archaeal viruses are not head–tail particles; instead, they encompass a range of novel groups and include many with unique morphologies (Prangishvili et al., 2006). The predominant virus-like particles in salt lakes are spindle-shaped and spherical (Guixa-Boixareu et al., 1996; Oren et al., 1997), but very few of these have been isolated and described (Dyall-Smith et al., 2003). On the other hand, head–tail particles appear to represent a minor proportion, yet have been the most commonly isolated. In non-extreme environments, such as marine and fresh waters, the great majority of bacteriophages are those with head–tail morphologies (Caudovirales) (Ackermann, 2007; Hendrix et al., 1999).

Halovirus SH1 is, to our knowledge, the only published isolate of a spherical halovirus, and was isolated in this laboratory from a salt lake sample taken in Western Australia (Porter et al., 2005). It is lytic, and infects Haloarcula hispanica, an experimentally convenient and genetically tractable haloarchaeon (Cline & Doolittle, 1992;
Woods et al., 1999). The virus replication cycle takes 5–6 h, and although cell lysis may occur at this time, delayed lysis is common, with cells continually shedding virus for several hours before eventually lysing (Porter et al., 2005). SH1 particles are ~70 nm in diameter, with a fragile outer protein layer and an internal membrane layer, somewhat resembling the structure of several characterized virus groups, including tectiviruses (Bamford et al., 2005; Fauquet et al., 2005; Jääniloja et al., 2008). A 3D model of SH1 has recently been published (Jääniloja et al., 2008) and shows a dramatic array of surface spikes. The genome of SH1 is a GC-rich (68.4%), linear dsDNA of 30 898 bp, with terminal inverted repeat sequences (ITRs) and 5′ terminal proteins (Bamford et al., 2005; Porter et al., 2005; Porter & Dyall-Smith, 2008). It is believed to be replicated by a protein-primed DNA polymerase, but the genome does not appear to encode any homologue of a known polymerase, and presumably uses a host enzyme.

In a previous study, the SH1 genome was sequenced, and protein-primed analyses used to predict a total of 56 virus ORFs (Bamford et al., 2005). The dense packing and arrangement of these ORFs strongly suggested that they were transcribed as operons, and the authors proposed a minimum of four, and potentially six to eight, operons. In the current study, we have experimentally tested the proposed operon structure of halovirus SH1 by mapping the transcripts and promoters used at different times in the infection cycle, and used puromycin blockage of protein synthesis to detect regulation of virus gene expression.

**METHODS**

**Media.** Media relevant to this study were prepared according to established protocols, as described in the online resource The HaloHandbook (http://www.haloarchaea.com/resources/halohandbook/). A 30 % (v/v) concentrated stock solution of artificial salt water (SW) containing (per litre): 240 g NaCl, 30 g MgCl2·6H2O, 35 g MgSO4·7H2O, 7 g KCl, 0.5 g CaCl2·2H2O, adjusted to pH 7.5 using a small volume (~2 ml) of 1 M Tris–Cl (pH 7.5). From this stock, halovirus diluent (HVD) and modified growth medium (MGM) containing 18 or 23 % (w/v) SW were prepared as described previously (Nuttall & Dyall-Smith, 1993; Porter et al., 2005).

**Strains, growth conditions and virus purification.** The haloarchaeon H. hispanica (ATCC 33960) (Juez et al., 1986) was grown in 18 or 23 % (w/v) MGM, as described previously (Porter et al., 2005). Halovirus SH1 (GenBank accession no. AY950802) was propagated in H. hispanica cells, as previously described (Porter et al., 2005).

For virus purification, unlysed H. hispanica cells and cell debris were removed from the culture by low-speed centrifugation (Sorval GSA; 6000 r.p.m., 30 min, 10 °C). Virus particles were then pelleted from the supernatant (Beckman SW28; 23 000 r.p.m., 13 h, 10 °C), onto a cushion of 30 % (v/v) sucrose, in HVD. The pellet was resuspended in HVD and purified by linear rate zonal 5–70 % (v/v) sucrose (in HVD) gradient centrifugation (Beckman SW28; 23 000 r.p.m., 2 h, 10 °C). The virus band was collected and recentrifuged to equilibrium in a solution containing 1.3 g CsCl ml⁻¹, in HVD (Beckman 70Ti; 60 000 r.p.m., 20 h, 10 °C). The virus band was collected, diluted in HVD, and the virus pelleted (Beckman SW55; 35 000 r.p.m., 75 min, 10 °C). The virus pellet was resuspended in HVD.

**Oligonucleotide primers.** Oligonucleotide primers were purchased from GeneWorks. Oligonucleotide primers referred to in the text are listed in Table 1. A full list of the primers used in this study is provided in Supplementary Table S1 available with the online version of this paper.

**Nucleic acid extractions.** Virus DNA was extracted from purified virus preparations using SDS and proteinase K, as described previously (Bath et al., 2006).

Infected-cell RNA was prepared as follows: Early exponential phase cells of H. hispanica were infected with SH1 at a m.o.i. of 50 (1 h, 37 °C). Under these conditions, the percentage of infected cells was expected to be 100 % (Porter et al., 2005). Cells were washed of unadsorbed virus by three cycles of centrifugation and resuspension in 10 ml fresh 18 % (w/v) MGM. After the final resuspension, cells were diluted in 100 ml medium. In some cases, puromycin (100 µg ml⁻¹; Sigma-Aldrich) was added to the infected cultures (Stolt & Zillig, 1994). Cultures were then incubated at 37 °C, with slow shaking (100 r.p.m.), and 10 ml samples were taken at hourly intervals and the cells were pelleted by centrifugation (13 000 g, 2 min, 4 °C). RNA was extracted using the RNeasy Mini kit (Qiagen), according to the manufacturer’s specifications. To prevent DNA contamination, 10 µg RNA was treated with 2 U DNase 1 (RNase-free) (New England Biolabs) and 40 U Recombinant RNasin RNase Inhibitor (Promega) for 30 min at 37 °C. DNase I was inactivated at 75 °C for 10 min.

**RT-PCR.** For first-strand cDNA synthesis, 0.5 µg virus-specific primer was incubated with ~1.5 µg RNA at 70 °C for 5 min, and was then chilled on ice for 5 min. Reverse transcription was performed at 50 °C for 1 h using 30 U AMV reverse transcriptase, with 40 U Recombinant RNasin RNase Inhibitor and 1 mM dNTPs (Promega). cDNA was purified using the QIAquick PCR Purification kit (Qiagen), and was resuspended in 30 µl RNase-free H2O.

A 3 µl volume of the cDNA preparation was used as template for PCR using GoTaq Green Master Mix (Promega), according to the manufacturer’s specifications. Template was denatured at 95 °C for 5 min; followed by 25–35 cycles of denaturation at 95 °C for 1 min, annealing for 30 s, and extension at 72 °C for 30 s; and a final extension at 72 °C for 10 min. cDNA levels were compared by dilution of the template, followed by 25 cycles of amplification.

A number of controls were performed to ensure the accuracy of the RT-PCR method. To avoid mis-priming on SH1 RNA during cDNA synthesis, primer sequences were first examined in silico against the SH1 genome sequence using the program AmplifX V1.4.0 (http://ifrj.nord.univ-mrs.fr/AmplifX). For each experiment, three specific control reactions were included. First, a positive control to show the PCR components were functioning. This had the same additions, except that the template was SH1 DNA. Second, to check for mis-priming on host cell cDNA (that may have been synthesized during reverse-transcription), SH1-specific PCRs were performed on cDNA made from reverse-transcribed, uninfecected H. hispanica RNA (using the same SH1-specific primers). This was a negative control. Third, a control for contamination by SH1 DNA used a template of RNA extracted from SH1-infected H. hispanica, but which had not undergone reverse transcription (negative control).

Additional controls were also performed. To confirm that the correct PCR products were generated, five randomly chosen products were sequenced. All gave the expected sequence (data not shown). To check that cDNA was not being randomly generated, a non-SH1-specific primer, 1RTF, was used to reverse transcribe cDNA from SH1-infected H. hispanica RNA. The cDNA was then screened with six randomly chosen SH1-specific PCRs. All screens produced no
product (data not shown). As an additional control to ensure that cDNA was not being randomly generated, reverse transcription of cDNA from SH1-infected *H. hispanica* RNA was performed with no primer. The cDNA was then screened with six randomly chosen SH1-specific PCRs. All screens produced no product (data not shown). Transcripts were numbered from left to right along the genome.

**Primer extension assays.** Primer extension was performed essentially as described by Ausubel *et al.* (1994). Primer (30 μM) was labelled with 30 μCi (1.1 MBq) [*γ-32P*]ATP using 2 U T4 polynucleotide kinase (New England Biolabs). cDNA was synthesized from ~10 μg RNA using 15 μCi (0.55 MBq) labelled primer, as above. A reference ladder was synthesized from ~10 ng pTM24 (sequence shown in Supplementary Fig. S1), using 30 μCi (1.1 MBq) labelled primer bgalpEXT. Samples were then precipitated and analysed on a 6 % (w/v) premixed acrylamide/bis/urea sequencing gel (Ambion). The gel was electrophoresed at 60 W and products were exposed to a phosphoimager plate overnight.

**Bioinformatic analyses.** In silico analysis for primer mis-priming was performed using AmplifX V1.4.0 (http://ifrjr.nord.univ-mrs.fr/AmplifX?lang=en). Cumulative AT and GC skews were produced using the DNA Base Composition Analysis Tool (http://molbiol-tools.ca/Jie_Zheng).

### RESULTS

**Fine mapping of SH1 transcripts**

To determine the operon structure of SH1, cDNAs of specific virus transcripts were systematically sized and mapped to the virus genome using RT-PCR (see Methods). An example of this strategy, shown in Fig. 1, summarizes the results for transcript T5, covering ORFs 23–34, for which the genome annotation suggested that these ORFs were likely to be transcribed as one operon. A cDNA copy of the predicted mRNA was prepared using SH1-infected *H. hispanica* RNA as template, and primer 19786R, which is complementary to a sequence (nucleotides 19 786–19 806) near the end of ORF 34. After reverse transcription, the length of the resulting cDNA was established by successive PCRs that scanned from the primer site Table 1.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′–3′)</th>
<th>Location (nt)</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>242R</td>
<td>GGGTATCCTGCTCCTAAAGG</td>
<td>242–259; 30 631–30 648</td>
<td>ITR</td>
</tr>
<tr>
<td>665R</td>
<td>GTTTGCCCAACTTAAATGG</td>
<td>684–665</td>
<td>ORF 3</td>
</tr>
<tr>
<td>935R</td>
<td>GACACACCTGACGCTGAGG</td>
<td>953–935</td>
<td>ORF 4</td>
</tr>
<tr>
<td>3124R</td>
<td>GTTCGAGGTCGACGCTGACG</td>
<td>3143–3124</td>
<td>ORF 12</td>
</tr>
<tr>
<td>3416R</td>
<td>GCCCTACCCAGGAGGAGG</td>
<td>3343–3416</td>
<td>ORF 13</td>
</tr>
<tr>
<td>10471F</td>
<td>TCCTCCTCCCTCAGTTGG</td>
<td>10 452–10 471</td>
<td>ORF 21</td>
</tr>
<tr>
<td>11387F</td>
<td>GTCTTGAATCCTCCGATCGG</td>
<td>11 367–11 387</td>
<td>ORF 22</td>
</tr>
<tr>
<td>11438R</td>
<td>TCACTCCGGAATGAGATGTG</td>
<td>11 457–11 438</td>
<td>ORF 23–22</td>
</tr>
<tr>
<td>11457F</td>
<td>CATCCTTATCGGAGGTGG</td>
<td>11 438–11 457</td>
<td>ORF 22–23</td>
</tr>
<tr>
<td>11637R</td>
<td>GTGCTTACGACGTGCTG</td>
<td>11 654–11 675</td>
<td>ORF 23</td>
</tr>
<tr>
<td>11675R</td>
<td>GCCCTGACATGAGGTCTCC</td>
<td>11 694–11 675</td>
<td>ORF 23</td>
</tr>
<tr>
<td>11694F</td>
<td>GGGAGATGTCGATGAG</td>
<td>11 657–11 694</td>
<td>ORF 23</td>
</tr>
<tr>
<td>12286R</td>
<td>AAGTCGAAGTACGACTGACC</td>
<td>12 305–12 286</td>
<td>ORF 24</td>
</tr>
<tr>
<td>13251F</td>
<td>GAACTGCGACCTTTCCTCCA</td>
<td>13 232–13 251</td>
<td>ORF 26</td>
</tr>
<tr>
<td>13518R</td>
<td>CTGTGACCTTGTCTCAGG</td>
<td>13 537–13 518</td>
<td>ORF 27</td>
</tr>
<tr>
<td>16585F</td>
<td>GACGCTGTGCGAGAGCAG</td>
<td>16 566–16 585</td>
<td>ORF 29</td>
</tr>
<tr>
<td>17398R</td>
<td>GACATGAGAGGACGGCG</td>
<td>17 417–17 398</td>
<td>ORF 30</td>
</tr>
<tr>
<td>17790F</td>
<td>CAGTTTACGTGATGAG</td>
<td>17 771–17 790</td>
<td>ORF 31</td>
</tr>
<tr>
<td>18012R</td>
<td>GTGTGAACTTGAGGTAG</td>
<td>18 031–18 012</td>
<td>ORF 32</td>
</tr>
<tr>
<td>18128R</td>
<td>GCTGTGTCGTAGGGGTAG</td>
<td>18 148–18 128</td>
<td>ORF 32</td>
</tr>
<tr>
<td>19437F</td>
<td>GTGGAACGGACGGTTA</td>
<td>19 418–19 437</td>
<td>ORF 33</td>
</tr>
<tr>
<td>19786R</td>
<td>GACTCCGAGTACCTGCTG</td>
<td>19 806–19 786</td>
<td>ORF 34</td>
</tr>
<tr>
<td>25874F</td>
<td>GCCCAACCTTCTCCTCCCT</td>
<td>25 854–25 874</td>
<td>ORF 49</td>
</tr>
<tr>
<td>26523F</td>
<td>GATTCGAGGAGCAGCAG</td>
<td>26 506–26 523</td>
<td>ORF 50.5</td>
</tr>
<tr>
<td>26960F</td>
<td>GAGACGGCGGGGTAGCCG</td>
<td>26 940–26 960</td>
<td>ORF 51</td>
</tr>
<tr>
<td>27573F</td>
<td>GGGACCTCGGCGGCTGAG</td>
<td>27 556–27 573</td>
<td>ORF 53</td>
</tr>
<tr>
<td>27835F</td>
<td>GTGTGACGCGACCCGAGCC</td>
<td>27 815–27 833</td>
<td>ORF 54</td>
</tr>
<tr>
<td>30390F</td>
<td>GTACACGCTGGCGAGGTGG</td>
<td>30 370–30 390</td>
<td>ORF 35</td>
</tr>
<tr>
<td>1RTF</td>
<td>CGACAGTATGAGTGG</td>
<td>1871–1887</td>
<td>ORF 55</td>
</tr>
<tr>
<td>bgalpEXT</td>
<td>GCCATCTGACTGAATACGTCTCC</td>
<td>pTM24†</td>
<td></td>
</tr>
</tbody>
</table>

*Designed to target the HF2 genome sequence (GenBank accession no. AF222060).
†Refer to Supplementary Fig. S1. All primers except bgalpEXT and 1RTF are designed to target the SH1 genome sequence (GenBank accession no. AY950802).
Many transcripts were very long, and most were produced during the early (1–2 h p.i.), middle (3–4 h p.i.) and late stages (5–6 h p.i.) of the infection cycle. In some cases it was not possible to distinguish by RT-PCR between transcripts and at different times during infection. The ITRs, i.e. prior to nucleotide 859 and after nucleotide 27 883, were also regions where no transcripts could be detected, on either strand, at any time between 1 and 6 h p.i.

The ITRs, i.e. prior to nucleotide 859 and after nucleotide 27 883, were also regions where no transcripts could be detected, on either strand, at any time between 1 and 6 h p.i.

**Extended viral transcripts**

Late in infection, transcriptional read-through of terminators produced longer transcripts. Since many SH1 operons face each other, many of the extended transcripts (denoted by an ‘e’ suffix, e.g. T2e) are complementary to the coding transcripts that they overlap (Fig. 2c). Excepting transcript T7, SH1 produced transcripts complementary to all of the coding mRNAs identified. The level of read-through varied between transcripts and at different times during infection. In some cases it was not possible to distinguish by RT-PCR the levels of shorter and longer transcripts originating from the same promoter. For example, T1 and T1e were expressed at similar levels early in infection (Fig. 2c), and may represent only the longer transcript (T1e), but late in infection there were ~100-fold higher levels of T1 relative to T1e, indicating that T1 termination was occurring, and both transcripts were produced. The situation was clearer in the case of T5/T5e, where T5 was produced early in infection, while transcript T5e was only produced late in infection, and at a ~10-fold
lower level compared with T5. The difference in timing and relative levels can be interpreted as weak read-through transcription of the T5 terminator to a more distal terminator at or near promoter P6.

Late in infection, transcripts T2e–T4e and T6e were produced at ~10-fold higher levels compared with the complementary transcripts that they overlap (Fig. 2c). As these transcripts would be partially or completely complementary (e.g. T1 and T2e; T5 and T6e), they could bind to form dsRNA, and so block translation. In this way, the extended transcripts may play a regulatory role in mRNA translation.

**Location of SH1 promoters by primer extension**

Using primer extension, the upstream regions of the previously predicted operons were examined for potential
transcription start sites (Fig. 2a) (Bamford et al., 2005), i.e. upstream of ORFs 1, 12, 13, 22, 23 and 50 were mapped. Primers used for each experiment are given in Table 2. For each panel, lane 1 is a size ladder and lane 2 is the primer extension product. Ladder positions for 100, 150 and 200 nt are indicated at the left of each gel panel by triangles (open, grey and filled, respectively). The extension product lengths are indicated at the right edge of each gel panel. Sizes are in nt.

Table 2. Summary of primer extension results using infected cell RNA

<table>
<thead>
<tr>
<th>ORF</th>
<th>Promoter</th>
<th>Primer</th>
<th>Primer extension product</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 1</td>
<td></td>
<td>242R</td>
<td>–</td>
</tr>
<tr>
<td>ORF 3</td>
<td></td>
<td>665R</td>
<td>–</td>
</tr>
<tr>
<td>ORF 4</td>
<td>P4</td>
<td>935R</td>
<td>Product of 95 nt (nucleotide 858)</td>
</tr>
<tr>
<td>ORF 12</td>
<td>P3</td>
<td>3124R</td>
<td>Product of 115 nt (nucleotide 3028)</td>
</tr>
<tr>
<td>ORF 13</td>
<td>P2</td>
<td>3416R</td>
<td>Product of 117 nt (nucleotide 3317)</td>
</tr>
<tr>
<td>ORF 22</td>
<td>P1</td>
<td>11387F</td>
<td>Product of 67 nt (nucleotide 11 434)</td>
</tr>
<tr>
<td>ORF 23</td>
<td>P5</td>
<td>11637R</td>
<td>Product of 113 nt (nucleotide 11 541)</td>
</tr>
<tr>
<td>ORF 32</td>
<td></td>
<td>18012R</td>
<td>–</td>
</tr>
<tr>
<td>ORF 49</td>
<td></td>
<td>25874F</td>
<td>–</td>
</tr>
<tr>
<td>ORF 50</td>
<td>P6</td>
<td>26523F</td>
<td>Product of 54 nt (nucleotide 26 560)</td>
</tr>
<tr>
<td>ORF 51</td>
<td></td>
<td>26960F</td>
<td>–</td>
</tr>
<tr>
<td>ORF 53</td>
<td></td>
<td>27573F</td>
<td>–</td>
</tr>
<tr>
<td>ORF 54</td>
<td></td>
<td>27833F</td>
<td>–</td>
</tr>
<tr>
<td>ORF 55</td>
<td></td>
<td>30390F</td>
<td>–</td>
</tr>
<tr>
<td>ORF 56</td>
<td></td>
<td>242R</td>
<td>–</td>
</tr>
</tbody>
</table>

Fig. 3. Identification of SH1 transcriptional start sites. Using primer extension, the transcription start sites for SH1 ORFs 4, 12, 13, 22, 23 and 50 were mapped. Primers used for each experiment are given in Table 2. For each panel, lane 1 is a size ladder and lane 2 is the primer extension product. Ladder positions for 100, 150 and 200 nt are indicated at the left of each gel panel by triangles (open, grey and filled, respectively). The extension product lengths are indicated at the right edge of each gel panel. Sizes are in nt.

In summary, the mapped start sites corresponded closely to the 5' ends of transcripts mapped by RT-PCR, while the lack of detectable start sites upstream of ORFs 1, 2, 55 and 56 is consistent with the RT-PCR data, which did not show transcription across ORFs 1–3, 55 and 56 (above, Fig. 2c). The primer extensions for transcripts T2, T3 and T4 were all performed on the same RNA preparation, at the same time, and run on the same gel (Fig. 3). The results indicate that all three promoters (P2, P3 and P4) are utilized, and that the levels of transcripts T2 and T3 are approximately equal, while the levels of T4 are slightly less.

**SH1 promoters and repeated sequence elements nearby**

The regions upstream of the mapped 5' transcription start sites are shown in Fig. 4. For ORF 4, there may be an alternative transcription initiation site. Likely promoter elements are indicated and, for comparison, the consensus haloarchaeal promoter sequence is given below. Potential TATA box motifs can be seen for the promoters upstream of ORFs 12, 13, 22 and 23, although in some motifs only five convincing bases can be identified. TATA boxes are difficult to identify upstream of the start sites near ORFs 4 or 50,5, although both of these promoters contain a novel element (WW) at −11/−10, as noted by Brenneis et al. (2007).

Start codons for ORFs 12, 13, 22 and 23 are only 0–5 nt from their respective transcription start sites, and the mRNAs for these ORFs are either leaderless or too short to contain a Shine–Dalgarno sequence.

The sequences around the major virus promoters were examined for repeated sequences that may represent factor binding sites regulating promoter activity. The reverse-complement of the sequence CCAACATTCTTA occurs twice, 30 nt apart, on either side of promoter P1 between ORFs 22 and 23 (overlined in Fig. 4). There is an additional imperfect repeat, CCACCAACTCC, 29 nt leftward of the first copy, on the opposite strand (nucleotides 11 393–11 403) (SH1 GenBank accession no. AY950802). Promoter P3 (between ORFs 11 and 12) has a similar shorter sequence, CCACCAATT, nearby (underlined in Fig. 4);
Hairpin–loop, inverted repeat sequence that could form a strong potential transcript T7, immediately upstream of ORF 54 is an ORFs 13, 20 and 33. While no promoter was identified for tracts. The latter sequence motif also occurs singly within CACACGA (underlined in Fig. 4), flanked by poly-G two closely spaced direct repeats of the sequence upstream. Near promoter P6, upstream of ORF 50.5, are 131 nt downstream, while an imperfect copy appears 94 nt 1 nt from the 5 end of the transcript is an AUG codon 9 of two directly repeated copies of the sequence TCCCGGCCAATT. All of these repeats contain the same hexameric motif CCAATT (bold type above), which is also present 60 nt upstream of the P4 promoter. This motif has a highly uneven distribution along the SH1 genome, being found 14 times between nucleotides 1 and 11,474 but only once thereafter (at nucleotide 24,391). The sequence CGACGGAGGAGGGCCGG occurs close to the P4 promoter (−13 to −25 in Fig. 4) and is repeated 131 nt downstream, while an imperfect copy appears 94 nt upstream. Near promoter P6, upstream of ORF 50.5, are two closely spaced direct repeats of the sequence CACACGA (underlined in Fig. 4), flanked by poly-G tracts. The latter sequence motif also occurs singly within ORFs 13, 20 and 33. While no promoter was identified for transcript T7, immediately upstream of ORF 54 is an inverted repeat sequence that could form a strong potential hairpin–loop, CggTccgGctCTGTCGTTccGCGGCGCG. A transcript for ORF 55 was not detected, but upstream of the start codon was the sequence GCGGCCGTT, repeated twice between nucleotides 30,555 and 30,573, and not found elsewhere in the genome.

Reannotation of the SH1 genome

The results of this study can be used to refine the previously published annotation of the SH1 genome (Bamford et al., 2005). Transcription across ORFs 1 and 56 was not detectable. These ORFs occur within the ITR sequences, a very uncommon location for protein-coding regions in the genomes of other protein-primed viruses (Supplementary Table S2). In addition, the first 48 aa of the predicted proteins are encoded by the ITR sequences, and are identical, but afterwards they differ significantly in length and sequence. Finally, the isoelectric points (IEPs) of their predicted protein products are very high (>11) compared with most haloarchaeal proteins, which average around 5 (Baliga et al., 2004; Bolhuis et al., 2006; Falb et al., 2005; Ng et al., 2000). Together, these features indicate that ORFs 1 and 56 are unlikely to be used to encode SH1 proteins. Transcription across ORFs 2 and 3 was also not detected and their predicted proteins have very high IEPs (>12).

RT-PCR determined the 5′ end of the long T4 transcript to be near the start of ORF 4, and primer extension on T4 mRNA refined this to a position occurring a few codons within ORF 4. To accommodate the shorter than expected transcript, the next most likely start codon (in the same reading frame) would be a GUG, 92 nt downstream at nucleotides 948–950, a position that occurs at the end of a predicted strong stem–loop structure.

ORF 12 begins 19 nt downstream of the transcription start point determined in this study, and while this arrangement is compatible with the published annotation, the first three bases of the transcript form an AUG start codon for an ORF (nucleotides 3028–3180) that largely overlaps ORF 12 but in a different reading frame. This alternative ORF has a calculated IEP of 5.32, similar to the average for haloarchaeal proteins, while the IEP of the ORF 12 protein is very high (IEP=11.7). Which of these two ORFs is utilized by SH1 will need to be determined.

The transcriptional start site of ORF 22 was determined to be downstream of the annotated start codon, making this ORF shorter than previously predicted. It may start at a downstream GUG start codon beginning at nucleotide 11,339, or perhaps use a UUG start codon beginning at nucleotide 11,429 (Fig. 4).

The transcriptional start site for ORF 50 was located 42 nt downstream of the annotated GUG translation start codon. Just 1 nt from the 5′ end of the transcript is an AUG codon that begins a new ORF (ORF 50.5, nucleotides 26,559–26,443), which terminates just past an AUG start codon for a now shortened version of ORF 50 (ORF 50s, nucleotides 26,446–25,937). Within ORF50.5, there are no potential start codons in-frame with ORF 50 until nucleotide 25,937, so the arrangement of ORF 50.5–ORF 50s appears the most likely.

SH1 ORF 55 is intriguing, as it appears to be a legitimate coding ORF yet transcription across it was not detected.
While the predicted protein shows no sequence similarity to known or hypothetical protein sequences, it is long (865 aa), has a low calculated IEP that is characteristic of haloarchaeal proteins, and is conserved in a related virus (K. Porter and M. L. Dyall-Smith, unpublished data). The cumulative GC skew plot of the genome is also consistent with this ORF being transcribed. Perhaps ORF 55 is transcribed only under specific conditions, or at a very early time in infection, or the mRNA has a very short half-life.

The influence of puromycin on virus transcription

The antibiotic puromycin prevents protein synthesis by binding to ribosomes, and can be used to show whether the transcription of virus genes is dependent on the translated product(s) of other genes (Azzam & Algranati, 1973; Matsunaga et al., 2001). It has been used to study halovirus φH1 gene expression (Gropp et al., 1989, 1992). As shown in Fig. 2(d), puromycin blocked the late expression of transcripts T1 and T5, while upregulating their transcription during the mid-phase of infection. Levels of T2e–T4e and T6e were lower and occurred earlier, while the appearance of T7 was delayed. The blocking of transcript T5 late in infection also resulted in the absence of transcript T5e, but this was not the case with T1 and T1e, for which the counter-transcript covering ORFs 4–16 was detected (T1e). This may indicate the use of an additional promoter. There was no effect on the expression of transcripts T2, T3, T4 or T6.

DISCUSSION

The SH1 genome was sequenced in 2005, and proposed to contain a minimum of four, and potentially six to eight, functional operons (Bamford et al., 2005). In the present study, the protein-coding ORFs of SH1 were indeed found to be transcribed as long, polycistronic mRNAs that cover most of the genome. Excluding the ‘extended’ transcripts, there are seven major mRNAs that contain between 4 and 16 ORFs each, largely matching the operon predictions of Bamford et al. (2005). There were two distinct phases of transcription, early–middle and late, as well as changes in relative levels over the course of infection. Early/late transcription, with no distinctive middle phase, has been observed in other viruses, including archaeal viruses SIRV1 and SIRV2 (Kessler et al., 2004).

Surprisingly, transcripts encoding the capsid proteins were synthesized from the earliest period examined, 1 h p.i., and increased in relative amounts over the infection cycle. The structural protein genes and the putative packaging ATPase (ORF 17) were most highly transcribed late in infection (5–6 h p.i.), when capsid assembly and genome packaging would be expected to be maximal. Three promoters were located over the region of ORFs 4–16, producing overlapping transcripts that all terminated just after ORF 16. All three transcripts included ORF 13, encoding viral protein (VP)1, the largest capsid protein. The thermophilic archaeal virus SSV1 shows a similar behaviour, with many genes being transcribed as operons from single 5’ start sites, but with terminator read-through resulting in the formation of several RNA species (Fröls et al., 2007; Reiter et al., 1987). The archaeal satellite virus pSSVx also utilizes transcriptional read-through of terminators (Contursi et al., 2007).

Transcript T6 was the only late SH1 transcript, appearing after 4 h p.i. In many bacteriophages, the late genes encode virus structural proteins, but T6, encompassing ORFs 35–50.5 (see below), includes no known structural proteins, and only one ORF that shows significant similarity to a known sequence. The ORF 48 predicted protein is similar in size and sequence to halovirus φCh1 protein p72 (GenBank accession no. AAM88745), although the function of the latter is unknown. The putative promoter for T6 does not display a consensus TATA motif, but there are repeat elements nearby that may be involved in late activation. There are reported examples of archaeal viruses that use protein activators of both viral (Iro et al., 2007) and host cell origin (Kessler et al., 2006).

An extraordinary feature of SH1 transcription was the degree of transcription from the non-coding strand. This is particularly evident late in infection, when many transcripts read-through their early phase terminators to more distal sites, resulting in counter-transcripts for most of the coding transcripts. These counter-transcripts do not contain convincing ORFs and are unlikely to encode functional proteins, but may be a form of translational regulation. In some cases the levels of counter-transcripts were low relative to their cognate coding transcripts, and any complementary binding between them would not be expected to be significant. The detection of such transcripts may be a reflection of the sensitive nature of RT-PCR. However, late in infection (5–6 h p.i.), not only were the levels of T1 and T5 high, but so were those of their counter-transcripts. The stability and abundance of these transcripts, as well as the potential interactions between complementary virus transcripts and their effects on SH1 gene expression, are areas for future study. The regulatory effects of dsRNA and interference RNA in halovirus replication have been well documented. Earlier work with halovirus φH has demonstrated the regulatory role of antisense RNA along with processing of dsRNA by a novel and previously unknown RNase of haloarchaea (Stolt & Zillig, 1993). Archaeal satellite virus pSSVx also makes use of anti-sense RNAs (Contursi et al., 2007). In the latter example, translation is blocked by modifying the 5’ ends of mRNAs, although even if dsRNAse processing alters the 3’ untranslated regions of viral mRNAs, alterations in mRNA stability could still affect translation (Brenneis et al., 2007).

SH1 promoters P1 and P5 are 49 nt apart and arranged back-to-back, an organizational pattern that is frequently observed in virus genomes and commonly used to ensure that the activities of nearby promoters, such as in
switching between lytic and temperate programmes of gene expression, are co-regulated (Beck & Warren, 1988). In such cases, the intergenic regions often bind regulatory proteins that alter the activity of one or both promoters. For example, in phage φ29 the outward-facing promoters A3 and A2(b,c) are 219 nt distant, and are regulated cooperatively by virus proteins p4 and p6, which bind at four sites between them and activate promoter A3 while repressing A2(b,c) (Badia et al., 2006; Camacho & Salas, 2001; Meijer et al., 2001). In the temperate halovirus φH, the back-to-back rep gene and T4 promoters are much closer, being just 63 nt apart. Here, rep expression was shown to be repressed directly by transcription from the T4 promoter, without the participation of regulatory virus proteins (Stolt & Zillig, 1994). The RT-PCR results of the present study indicate that even though the P1 and P5 promoters are even more closely spaced than in the φH example, they do not appear to interfere with each other, as their activities parallel each other, with transcription rising at the same time from low to high levels late in infection. The role of the sequence repeats near these promoters remains to be determined.

Puromycin blockade of protein synthesis altered the timing of some early–middle SH1 transcripts, advancing transcripts T1, T5, T2e–T4e and T6e, while delaying T7. The behaviour of T2e–T4e may be explained by anti-termination of the T2–T4 transcripts. Others were unaffected (T2–T4), as was the major late transcript T6. The unaltered transcription pattern of T6 suggests that protein synthesis during infection is not required for regulation of its promoter (P6). Other effects of puromycin were the reduced or undetectable levels of specific transcripts late in infection, such as T1 and T5, as well as the extended transcripts T2e–T6e. As the extended transcripts T2e–T4e and T6e were unaffected, although their ‘parent’ transcripts (T2–T4 and T6) were not, this suggests that they may have a separate promoter, rather than being extensions of the original transcripts. The changes seen in the overall pattern of transcription were complex, suggesting that two or more interacting regulatory mechanisms are involved.

In summary, the genes of halovirus SH1 are organized into a relatively small number of operons with tightly regulated promoters. Late in infection, transcription is characterized by extensive counter-transcription that may influence the translation of mRNAs. More detailed studies of gene regulation in this virus will not only allow us to improve our understanding and annotation of the SH1 genome, but also allow a deeper understanding of archaeal (particularly haloarchaeal) transcriptional behaviour.

ACKNOWLEDGEMENTS

The support of the Department of Microbiology and Immunology (University of Melbourne) is greatly appreciated. B.E.R. was supported by Melbourne Research Scholarships. K. P. was supported by an Australian Postgraduate Award.

REFERENCES


K. Porter and others


Edited by: H-P. Klenk