Molecular Microbiology (2008) 70(3), 549–553 ■
doi:10.1111/j.1365-2958.2008.06434.x
First published online 25 September 2008

MicroCommentary

A shifty stop for a hairy tail

Adam S. Olia and Gino Cingolani*
Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University, Syracuse, NY 13210 USA.

Summary

The tail apparatus of the bacteriophage SPP1 is an extraordinary ~1600-Å-long molecular machine. The tail mediates attachment of the virus to the host surface receptor, as well-as ejection of the viral genome into the host. The distal tip of the tail binds the extracellular ectodomain of the *Bacillus subtilis* receptor YueB, while the tail tube provides a conduit to funnel the viral genome into the host. This process, which culminates with the ejection of the ~44 kb of viral DNA across the thick, cell envelope of the Gram-positive bacterial cell, takes place in a time scale of seconds to minutes and represents a remarkable example of biotransformation. In this issue of *Molecular Microbiology*, Auzat *et al.* provide compelling evidence that the two major structural proteins of the SPP1 tail, gp17.1 (~19.1 kDa) and gp17.1* (~28 kDa), share a common N-terminal sequence, and that gp17.1* is generated by a translational frameshift in the gene 17.1. The extra domain fused to gp17.1* is synthesized by a +1 programmed translational frameshift at the end of gene 17.1, which leads to the synthesis of approximately one gp17.1* for every three equivalents of gp17.1. This finding extends our current knowledge of translational frameshifts and provides a framework to understand how *Siphoviridae* phages like SPP1 have developed long-tail machines using only two major structural proteins.

Introduction

Bacteriophages represent the most abundant form of life on earth. They have evolved to infect bacteria present in any natural reservoir, such as soil, water and even the intestines of animals. Consistent with their simple yet efficient organization, phages have developed formidable mechanisms to penetrate the bacterial cell envelope and inject their genetic material into the host. In tailed bacteriophages such as SPP1 and T4, the tail apparatus is primarily involved in scanning the host cell surface and driving ejection of the viral genome. As a result of the tremendous morphological and chemical differences in the host cell surface, phage tails vary tremendously in size, complexity and mechanisms of function. The tail morphology also provides a simple, intuitive way to classify tailed bacteriophages into three families: *Podoviridae* (e.g. P22, phi29, epsilon15), *Siphoviridae* (e.g. SPP1, λ) and *Myoviridae* (e.g. T4) (Ackermann, 2003). *Podoviridae,* like P22, have short, non-contractile tails that efficiently penetrate the cell envelope of Gram-negative bacteria. In P22, perhaps the best characterized *Podoviridae*, the phage first attaches to the host via the tailspike gp9, which hydrolyses the O-antigenic repeating units of the *Salmonella* lipopolysaccharide. This is followed by the penetration of the host cell envelope by the tail needle gp26 (Andrews *et al.*, 2005), which is ejected into the host during infection (Israel, 1977), likely opening a pore in the host cell envelope. In *Siphoviridae* such as λ or SPP1, the tail is ~1350- to 2600-Å-long (Katsura and Hendrix, 1984; Pedulla *et al.*, 2003; Vegge *et al.*, 2005; McGrath *et al.*, 2006) and does not contract during infection. In contrast, in *Myoviridae,* like the classical bacteriophage T4, the tail is long, complex and contractile. For T4, probably the best structurally characterized bacteriophage, structural information has been obtained for both the extended and contracted states. Comparison of the two structures shows that tail contraction is associated with a dramatic shortening of the tail sheath to about one-third of its original length. This contraction projects the tail tube by about half of its total length and induces its rotation by 345°, allowing it to penetrate the periplasmic space of the host (Leiman *et al.*, 2003; Rossmann *et al.*, 2004; Kostyuchenko *et al.*, 2005).

Architecture of bacteriophage SPP1 tail apparatus

Since its identification in 1968 (Riva *et al.*, 1968), the *Bacillus subtilis* bacteriophage SPP1 has been one of the most studied and better characterized viruses infecting

Accepted 26 August, 2008. *For correspondence. E-mail cingolag@upstate.edu; Tel. (+1) 315 464 8744; Fax (+1) 315 464 8750.

Gram-positive bacteria. The phage genome consists of ~44 kb (Alonso *et al*., 1997) with partially circularly permuted ends (Morelli *et al*., 1979; Tavares *et al*., 1992). Like other members of the *Siphoviridae* family, SPP1 has a long and flexible tail apparatus attached to an icosahedral capsid, ~650 Å in diameter. The tail apparatus serves as a conduit for the ejection of viral DNA from the viral capsid into the cell host and is attached immediately below the phage connector by the head–tail joining protein gp17. In SPP1, the connector is defined as the trimeric complex formed by the portal protein gp6 and the head completion proteins, gp15 and gp16 (Orlova *et al*., 2003).

The structural organization of the SPP1 non-contractile tail is known from electron microscopic studies (Plisson *et al*., 2007). The tail is built by a proteinaceous tail tube, which is flexibly joined to an adsorption tip, at the end distal from the connector. The tip, 310-Å-long and 95-Å-wide, contains the receptor binding domain and a beta-strand-rich protein, gp21, which is structurally similar to the phage P22 tailspike (Steinbacher *et al*., 1994) and is likely important for degradation of the host cell wall during infection. The tip is connected to the tail tube by gp19.1, which allows bending of the tip with respect to the tube, enhancing the ability of the tip to scan for receptors on the host surface. The most noticeable component of the phage tail, the tail tube, has a diameter of ~112 Å and is ~1600-Å-long. It is composed of ~40 rings stacked vertically and rotated ~21° relative to each other. The tube is built by two major tail proteins (MTP), gp17.1 and gp17.1*, in an approximate ratio 3:1. The diameter of the tail tube prior to DNA ejection is ~110 Å, and it has an internal lumen ~56 Å in diameter, occupied by multiple copies of the tape measure protein, gp18.

Frameshifts in viruses

Frameshifting is a well-described process in which the ribosome begins reading out of frame at a designed point in an mRNA message. A random translational frameshift is an exceedingly rare event, as, normally, this would quickly lead to missense and nonsense mutations in the polypeptide chain. However, a variety of signals and mRNA secondary structures can enhance this rare process and increase the frequency of frameshifting to more than 80%, leading to what is usually referred to as a 'programmed' frameshift. Frameshifts can be programmed for one of two purposes: either to produce a specific ratio of two products or to add a level of regulation, where the frequency of frameshifting is variable and controlled *in trans*. Viral frameshifted proteins tend to be of the set ratio variety and function to produce structural proteins, while most non-viral examples favour a regulatory role.

When altering the correct reading frame of an mRNA, two frameshifts are more likely: −1 frameshift, where the ribosome slips one nucleotide backwards and +1 frameshift (more rare), where the ribosome jumps one nucleotide (Fig. 1). The most widespread translational frameshift is the −1 frameshift, which occurs in many viruses (and phages) as well-as in viral elements, such as transposons. This frameshift, in which the ribosome slips back one nucleotide out of its initial frame, is caused by a dual-slippage event at a heptanucleotide site of the form X.XXY.YYZ, where the triplets indicate the Frame 0 reading frame. In this case, the ribosome can slip back by one base, and both the A-site and P-site still maintain two non-wobble position base-pairings. In this scenario, the P-site slips from XXY to XXX and the A-site from YYZ to YYY (Jacks *et al*., 1988; Ivanov and Atkins, 2007). This phenomenon is greatly enhanced when the second codon in the heptanucleotide (the YYZ) generated by the −1 frameshift is a rare codon, which causes the ribosome to stall while it waits for the new aminoacyl-tRNA to occupy the A-Site (Stahl *et al*., 2002).

There are extensive cases of −1 frameshifts in viruses including the T7 gene 10, phiX174 lysis gene, retrovirus gag-pol, the yeast L-A virus coat-pol protein and the coronavirus F1-F2 protein (Witte and Baltimore, 1978; Felsenstein and Goff, 1988; Levin *et al*., 1993). Viruses and phages use a variety of mechanisms such as Shine-Dalgarno-like sequences, stop codons or 3′ pseudoknots and stem-loops to stall the ribosome, which is the main factor influencing the efficiency of frameshifts (for an example, see Levin *et al*., 1993). In bacteriophage λ, a −1 frameshift at the interface between gene T and gene G is used to produce the gene product G-T at a frequency of ~4%. The λ frameshift is located between the bases encoding residues 13–14, at a sequence G.GGA.AAG, where the Gly-Lys product is produced in both the Frame 0 as well-as the Frame −1. The mechanism seems to be induced by a pause in the ribosome caused by a rare cysteine codon immediately 3′ to the lysine codon. This might lead to the ribosome pausing with the $tRNA^{Gly}$ in the P-site, the $tRNA^{Lys}$ in the A-site, and then slipping back by one nucleotide, changing only the wobble base of the two tRNAs (Levin *et al*., 1993).

In contrast to −1 frameshifts, there are very few examples of +1 frameshifts in viruses, in which the ribosome slips forward one base. In the bacteriophage PSA, a +1 frameshift occurs in the *cps* gene at the sequence ACA.CCC.UCC.G, which is followed by a pseudoknot at the 3′ of the shift site (Zimmer *et al*., 2003). This frameshift results in a frequency of frameshifting (and generation of the longer product) of 13.2%, which is *exactly* the ratio of proteins needed to form pentons and hexons of a T = 7 phage (55 penton proteins and 360 hexon proteins) (Zimmer *et al*., 2003). This demonstrates

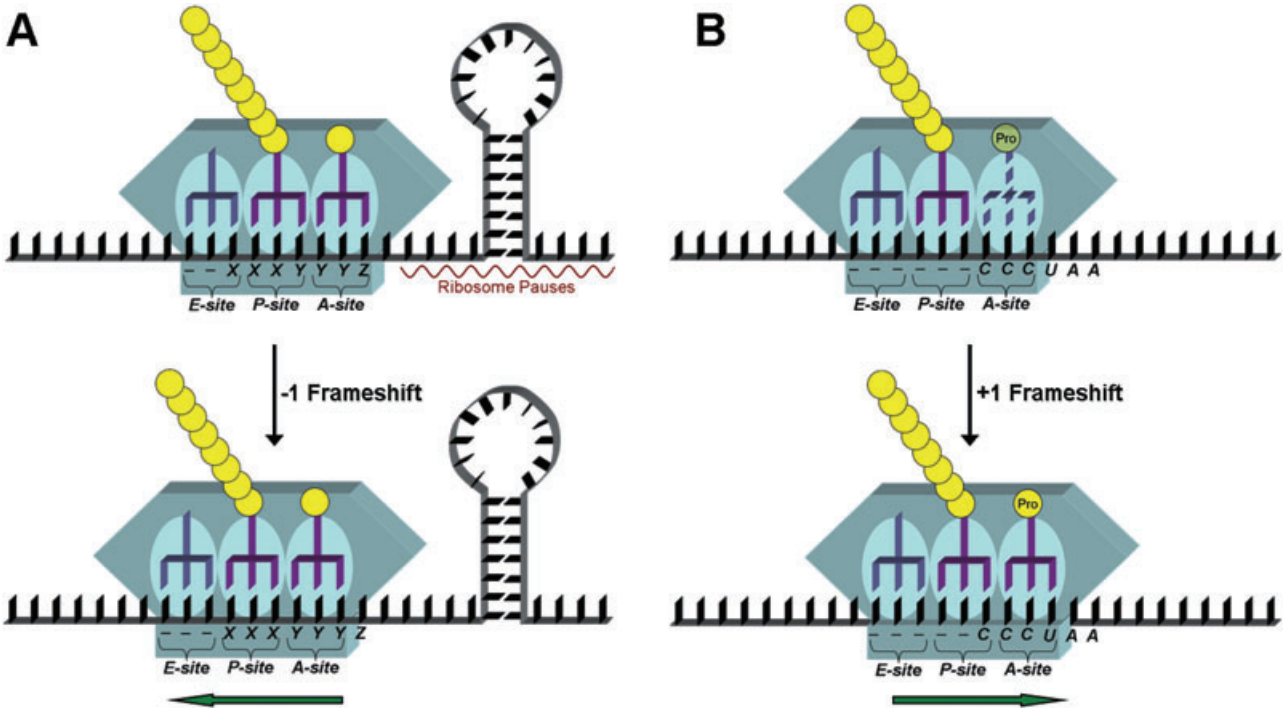

Fig. 1. Common mechanisms of viral frameshift.
A. In the standard model of a –1 frameshift, the ribosome encounters the sequence X.XXY.YYZ and is paused usually as a result of a 3′ secondary structure in the mRNA (e.g. a stem-loop or pseudoknot). During this pause, the ribosome slips backwards one base, with both the A-site and P-site re-annealing, changing only the base pairing of the wobble base.
B. In the +1 frameshift, the ribosome pauses immediately before a stop codon, normally waiting for a rare $tRNA^{Pro}$ (shown as dashed) at the proline 'CCC' codon. This leads the ribosome to move forward one base, and in the case of the rare $tRNA^{Pro}$, the A-site finds the more common 'CCU' codon rather than the rare 'CCC'.

the incredible level of fine-tuning possible in these kinds of systems. In addition to the frameshift in the *cps* gene, there is another +1 frameshift in the *tsh* protein, which utilizes the shifty-stop mechanism, induced by the sequence ACA.CCC.UGA, where the rare CCC proline codon pauses the ribosome, leading to a 25% frequency of frameshifting (a 1:3 ratio of long : short protein product) (Zimmer *et al*., 2003).

In the phage Q54, a fusion between the MTP and the receptor-binding protein occurs because of a +1 shifty stop, resulting in ~25% abundance of the fusion protein as compared with the MTP alone. This may be due to the CCC proline, which is five times less abundant than the CCU codon, which also encodes proline. Additionally, there are two stem-loops present immediately following the stop (and shift site) of the Q54 orf23, one of which overlaps the ribosomal binding site of the shifted frame (orf24), resulting in lack of translation initiation for the orf24 gene (Fortier *et al*., 2006).

Programmed translational frameshift in SPP1 MTP gp17

In this issue of *Molecular Microbiology*, Auzat *et al*. provide compelling evidence that the two major structural proteins of the SPP1 tail, gp17.1 and gp17.1* have a common N-terminal sequence and that gp17.1* is further extended by a programmed translational frameshift. This frameshift extends gp17.1 (~19.1 kDa) by ~10 kDa, to yield gp17.1* (~28 kDa). The authors demonstrate that the frameshift takes place at the second of two rare proline codons (CCC). Slippage of the ribosome by +1 nucleotide leads to the in frame translation of ~80 codons of the open reading frame *orf17.2*, which extends downstream of the gene 17.1 stop codon by 87 amino acids. To demonstrate the critical role of the second proline codon in the translational frameshift, Auzat *et al*. introduced mutations at positions 1 and 2 of the first rare proline codon, with the goal of replacing the rare CCC codon with a more abundant codon. Such substitutions yielded single amino-acid changes in gp17.1*, suggesting that the first CCC codon is neither necessary nor sufficient for the translational frameshift. In contrast, a single-point mutation in the third base of the second codon (CCC→CCT) completely abolished production of gp17.1*, confirming that the frameshift is caused by the second rare CCC codon and the presence of the adjacent in frame open reading frames *orf17.2*. By expressing the SPP1 gene 17.1 and *orf 17.2* in a plasmid (pPT25), the authors elegantly reproduced the

translational frameshift that yields gp17.1 and gp17.1* in *Escherichia coli*. As *E. coli* and *B. subtilis* share similar codon usage, the codon CCC is rare in both bacteria and causes the ribosome to slow down if repeated in tandem. Remarkably, overexpression of the rare tRNA for proline suppresses production of gp17.1*, suggesting that the stalling of the ribosome at the 'hungry' CCC codon is essential for the ribosome to slide +1 nucleotide on the messenger RNA. The ability to reproduce the translational frameshift *in trans* using a plasmid, allowed the authors to engineer SPP1 phages that selectively assemble only gp17.1 or gp17.1*. These phages are ~6.8 MDa heavier and ~4.6 MDa lighter respectively, than wild-type SPP1 bacteriophage and present striking morphological differences. Tails assembled exclusively by gp17.1* are 'hairy' as compared with both wild-type and gp17.1-only tails. The protrusions irradiating from the tail tube that causes the rough surface appearance seen by electron microscopy represent the C-terminal domain of gp17.1* translated as a result of the frameshift at codon 177. It has been suggested that the extra domain inserted into gp17.1* adopts an immunoglobulin-like fold and may have a role in carbohydrate binding (Fraser *et al*., 2006; 2007).

Conclusion – what have we learned?

The data presented by Auzat *et al*. (2008) reveal an elegant mechanism to generate structural complexity and likely functional versatility within the tail apparatus of the bacteriophage SPP1, while maintaining an efficient and minimal genome. The +1 programmed translational frameshift ensures production of two proteins in a relative stoichiometry of 3:1. In the absence of a translational frameshift, the virus would have to carry a second gene for gp17.1* with a weaker promoter to enable lower-level production with respect to the gp17.1 gene product. However, a separate gene for gp17.1* would dramatically reduce the efficiently of tail self-assembly and likely lead to a decreased number of infectious virions. In fact, assuming that in the wild-type SPP1 phage, gp17.1 and gp17.1* co-assemble translationally, the existence of separate gene products for gp17.1 and gp17.1* would favour the homotypic interaction gp17.1 : gp17.1 and gp17.1* : gp17.1* instead of the heterotypic assembly of gp17.1 : gp17.1*. In the absence of dedicated chaperons, the former scenario would likely make the SPP1 tail assembly and phage maturation dramatically less efficient. Likewise, a separate gene for gp17.1* would result in an enlargement of the viral genome to encode redundant information. Interestingly, the insertion domain of gp17.1* confers a 'hairy' appearance to the phage tail, as assessed by negative stain electron microscopy. It will be interesting to determine if phages lacking the gp17.1* protein undergo conformational changes in the tail tube similar to those seen for wild-type SPP1 upon binding to the surface receptor YueB (Plisson *et al*., 2007). In conclusion, the data presented by Auzat *et al*. (2008) add a new step in the direction of understanding the assembly and architecture of the SPP1 non-contractile tail apparatus. This remarkably dynamic machine is capable of propagating the signal for ejection from the cell surface receptor to the connector across ~1600 Å of tail tube. This unprecedented long-distance signal transduction commits the virus to a new cycle of infection and leads to ejection of the viral genome into the host.

References

Ackermann, H.W. (2003) Bacteriophage observations and evolution. *Res Microbiol* **154:** 245–251.

Alonso, J.C., Luder, G., Stiege, A.C., Chai, S., Weise, F., and Trautner, T.A. (1997) The complete nucleotide sequence and functional organization of *Bacillus subtilis* bacteriophage SPP1. *Gene* **204:** 201–212.

Andrews, D., Butler, J.S., Al-Bassam, J., Joss, L., Winn-Stapley, D.A., Casjens, S., and Cingolani, G. (2005) Bacteriophage P22 tail accessory factor gp26 is a long triple-stranded coiled-coil. *J Biol Chem* **280:** 5929–5933.

Auzat, I., Dröge, A., Weise, F., Lurz, R., and Tavares, P. (2008) Origin and function of the two major tail proteins of bacteriophage SPP1. *Mol Microbiol* (in press).

Felsenstein, K.M., and Goff, S.P. (1988) Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *J Virol* **62:** 2179–2182.

Fortier, L.C., Bransi, A., and Moineau, S. (2006) Genome sequence and global gene expression of Q54, a new phage species linking the 936 and c2 phage species of Lactococcus lactis. *J Bacteriol* **188:** 6101–6114.

Fraser, J.S., Yu, Z., Maxwell, K.L., and Davidson, A.R. (2006) Ig-like domains on bacteriophages: a tale of promiscuity and deceit. *J Mol Biol* **359:** 496–507.

Fraser, J.S., Maxwell, K.L., and Davidson, A.R. (2007) Immunoglobulin-like domains on bacteriophage: weapons of modest damage? *Curr Opin Microbiol* **10:** 382–387.

Israel, V. (1977) E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. *J Virol* **23:** 91–97.

Ivanov, I.P., and Atkins, J.F. (2007) Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. *Nucleic Acids Res* **35:** 1842–1858.

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* **55:** 447–458.

Katsura, I., and Hendrix, R.W. (1984) Length determination in bacteriophage lambda tails. *Cell* **39:** 691–698.

Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., Arisaka, F., Mesyanzhinov, V.V., and Rossmann, M.G. (2005) The tail structure of bacteriophage T4 and its mechanism of contraction. *Nat Struct Mol Biol* **12:** 810–813.

Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* **60:** 2356–2370.

Levin, M.E., Hendrix, R.W., and Casjens, S.R. (1993) A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. *J Mol Biol* **234:** 124–139.

McGrath, S., Neve, H., Seegers, J.F., Eijlander, R., Vegge, C.S., Brondsted, L., *et al.* (2006) Anatomy of a lactococcal phage tail. *J Bacteriol* **188:** 3972–3982.

Morelli, G., Fisseau, C., Behrens, B., Trautner, T.A., Luh, J., Ratcliff, S.W., *et al.* (1979) The genome of *B. subtilis* phage SSP1: the topology of DNA molecules. *Mol Gen Genet* **168:** 153–161.

Orlova, E.V., Gowen, B., Droge, A., Stiege, A., Weise, F., Lurz, R., *et al.* (2003) Structure of a viral DNA gatekeeper at 10 A resolution by cryo-electron microscopy. *EMBO J* **22:** 1255–1262.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* **113:** 171–182.

Plisson, C., White, H.E., Auzat, I., Zafarani, A., Sao-Jose, C., Lhuillier, S., *et al.* (2007) Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *EMBO J* **26:** 3720–3728.

Riva, S., Polsinelli, M., and Falaschi, A. (1968) A new phage of *Bacillus subtilis* with infectious DNA having separable strands. *J Mol Biol* **35:** 347–356.

Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., and Leiman, P.G. (2004) The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* **14:** 171–180.

Stahl, G., McCarty, G.P., and Farabaugh, P.J. (2002) Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci* **27:** 178–183.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265:** 383–386.

Tavares, P., Santos, M.A., Lurz, R., Morelli, G., de Lencastre, H., and Trautner, T.A. (1992) Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. *J Mol Biol* **225:** 81–92.

Vegge, C.S., Brondsted, L., Neve, H., McGrath, S., van Sinderen, D., and Vogensen, F.K. (2005) Structural characterization and assembly of the distal tail structure of the temperate lactococcal bacteriophage TP901-1. *J Bacteriol* **187:** 4187–4197.

Witte, O.N., and Baltimore, D. (1978) Relationship of retrovirus polyprotein cleavages to virion maturation studied with temperature-sensitive murine leukemia virus mutants. *J Virol* **26:** 750–761.

Zimmer, M., Sattelberger, E., Inman, R.B., Calendar, R., and Loessner, M.J. (2003) Genome and proteome of Listeria monocytogenes phage PSA: an unusual case for programmed +1 translational frameshifting in structural protein synthesis. *Mol Microbiol* **50:** 303–317.

