Virology 509 (2017) 140–145

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Natural history of a viral cohesive end site: *cosN* of the λ-like phages

Nicole Vahanian[1], Choon Seok Oh, Jean Sippy, Michael Feiss*

Department of Microbiology, Roy J. and Lucille A. Carver College of Medicine University of Iowa, Iowa City, IA 52242, USA

ARTICLE INFO

Keywords:
DNA packaging
DNA recognition
Virus assembly
Terminase
Viral DNA maturation

ABSTRACT

The base pairs of *cosN*, the site where the 12 base-long cohesive ends are generated in λ-like phages, show partial-two fold rotational symmetry. In a bioinformatic survey, we found that the *cosN* changes in 12 natural *cosN* variants are restricted to bp 6-to-12 of the cohesive end sequence. In contrast, bp 1–5 of the cohesive end sequence are strictly conserved (13/13), as are the two bp flanking the left nicking site (bp −2 and −1). The bp flanking the right nick site (bp 13 and 14) are conserved in 12 of 13 variants. Five *cosN* variants differing by as many as five bp were used to replace lambda's *cosN*. No significant effects of the *cosN* changes on λ's virus yield were found. In sum, bp −2 to 5 are critical *cosN* function, and bp 6–12 of the cohesive end sequence are not critical for terminase recognition or virus fitness.

1. Introduction

The virion DNA of many tailed bacteriophages has a unique sequence and cohesive ends. Cohesive ends are short complementary single-stranded extensions, either 3′- or 5′-ended, that anneal upon entry of the viral DNA into a host cell. Annealing circularizes the viral DNA. At late times in infection, rolling circle replication and recombination produce linear and circular DNA multimers, called concatemers, that are processed into unit-length virion DNAs with cohesive ends. During DNA processing, the viral terminase's endonuclease cleaves concatemeric DNA, creating the cohesive ends. Cleavage occurs as part of the DNA packaging process, i.e., the packaging of the viral DNA into the empty capsid shell precursor called the prohead. Endonuclease processing is also used by the herpes viruses, which also make specific cuts along concatemeric DNA (Nadal et al., 2010; Tong and Stow, 2010), and the *pac* bacteriophages, which use the headful DNA packaging strategy (Casjens and Gilcrease, 2009).

There are some obvious functional constraints for viruses that use cohesive ends. First, during a viral infection, the cohesive ends need to anneal quickly following entry of the DNA into the cell, as dsDNA ends are vulnerable to attack by host exonucleases. The cohesive ends should not fold back on themselves, i.e., not be self-complementary. Second, the cohesive ends should remain annealed long enough for the host ligase to seal the nicks. Third, the site of the annealed cohesive ends functions as a substrate when the packaging enzyme, terminase, regenerates cohesive ends by introducing appropriately placed staggered nicks, followed by strand separation [reviewed in (Catalano et al., 1995)].

Phage cohesive ends are reported as short as 7 nucleotides long, so perhaps 7 nucleotides may be at or near the minimum length (Ellis and Dean, 1985). Although there is an obvious similarity between Type II restriction enzyme target sites and phage cohesive end sites, there are major differences. Phage cohesive ends are longer, because, unlike the restriction targets, cohesive ends must efficiently anneal at physiological temperatures. Cohesive ends also lack strict two-fold rotational symmetry, avoiding self-annealing into hairpin structures. Also, strict cohesive end symmetry could potentially lead to formation of cruciform structures in vivo, leading to attack by resolvases (Eykelenboom et al., 2008; Cromie et al., 2000)

In phage λ and its relatives, cohesive ends are regenerated by the introduction of staggered nicks at *cosN*, the site containing the cohesive end sequence. Here we consider $cosN^{\lambda}$ to include the 16 bp-long sequence that includes the 12 bp cohesive end sequence plus the 2 bp flanking the two nick sites (Fig. 1). The cohesive ends are 12 base-long extensions of the 5′ ends of the DNA strands. The viral DNA packaging enzyme, terminase, nicks *cosN* with exquisite accuracy. Terminase is a large protein consisting of large (TerL) and small (TerS) subunits.[2] The λ terminase protomer is a $TerS_2TerL_1$ heterotrimer; heterotrimers further associate into a tetramer of heterotrimers (Maluf et al., 2005, 2006). TerS is a small DNA binding protein that forms a cylindrical oligomer (Casjens, 2011; Oliveira et al., 2013; Rao and Feiss, 2015). TerL contains two activities, the C-terminal endonuclease, and the N-

* Corresponding author.
E-mail addresses: michael-feiss@uiowa.edu, feissm@gmail.com (M. Feiss).
[1] Present address: School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana 47405-1701, USA.
[2] The λ genes for TerS and TerL are *Nu1* and *A*, resp., and the phage specific names for the proteins are gpNu1 and gpA, respectively. For simplicity, here we use the generic designations TerS and TerL. We designate the specific λ proteins as $TerS^{\lambda}$ and $TerL^{\lambda}$.

http://dx.doi.org/10.1016/j.virol.2017.06.016
Received 28 April 2017; Received in revised form 8 June 2017; Accepted 11 June 2017

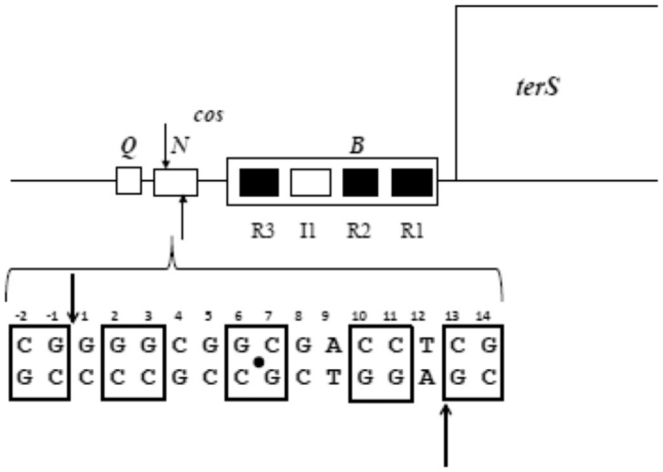

Fig. 1. Phage λ DNA packaging signals and the *terS* gene. Upper: *cos* and *terS* gene organization in λ-like phages. *cosN* is the 22 bp-long nicking site where TerL introduces staggered nicks (vertical arrows) to generate the cohesive ends of virion DNA. *cosB* contains three repeats of the TerS binding sequence, R3, R2, and R1, as well as I1, a binding site for the *E. coli* bending protein IHF. *cosN* and *cosB* are required for initiation of DNA packaging. *cosQ* is required, along with *cosN*, for termination of DNA packaging. Lower: $cosN^{\lambda}$. Vertical arrows indicate nick positions, boxed bp show two-fold rotational symmetry, and the dot indicates the center of symmetry. In the standard λ bp numbering system, bp 1 is the first base of the left cohesive end. For the present work, the bp 5′ to the left nick site are designated −2 and −1.

terminal ATPase center (Rao and Feiss, 2015; elToro et al., 2016; Hang et al., 2000; Mitchell and Rao, 2004; Mitchell et al., 2002). The ATPase powers translocation of viral DNA into the prohead. TerL is not an intrinsically accurate endonuclease, and efficient and accurate nicking at the proper positions in *cosN* requires that TerL be anchored by the binding of TerS at *cosB*, a complex site adjacent to *cosN* (Catalano et al., 1995). *cosB* is critical for the process of generating cohesive ends. Cohesive end-containing virion DNAs are produced during the DNA packaging process. DNA packaging is initiated when protomers assemble on a *cos* along a concatemer (Yang et al., 2017). Nicks are introduced at *cosN*, the cohesive ends are separated, and terminase remains bound to the *cosB* containing DNA end. This DNA-terminase complex docks on the portal vertex of a prohead, and the powerful translocation ATPase powers the DNA into the prohead. When the translocating motor encounters the next *cos* along the concatemer, packaging is terminated by cleavage of the downstream *cosN*. Packaging is processive – the terminase that completes packaging of the first chromosome remains bound to the *cosB*-containing end of the next chromosome, docks on a prohead and sponsors translocation of the second viral DNA in the sequence, and so on. In this scenario, initiation is an infrequent event, which may be due to the limited amount of terminase that is made and the efficiency of assembly at *cos*. In contrast, the events of termination and packaging of successive DNAs along the chromosome are proposed to be efficient (Catalano et al., 1995).

The λ *cosN* sequence shows partial two-fold rotational symmetry, including the nick positions and 2 bp flanking the nick positions (Fig. 1). An intriguing question is whether *cosN*'s symmetry is significant for function, i.e., for recognition by TerL. An early study indicated *cosN* is functionally asymmetric (Xu and Feiss, 1991a). That is, a G→T mutation at the bp flanking the left nick site, designated here as the −1 bp of *cosN*, reduced *cosN* nicking, but the reciprocal C→A change at the right, i.e., at bp 13, was innocuous. Similar asymmetric results were obtained for the G2C/C11G and G3C/C10G symmetric pairs of mutations. The burst size of a G6A, C7G double mutant was not significantly different from that of λ cos^{+} (Xu and Feiss, 1991a). Multiple bisulfite-induced C→T mutations in the right half of *cosN* had modest effects on virus yield, while mutations affecting the left half caused severe defects.

A very early molecular study found that the DNAs of λ-like phages 21, 424, and 434 could form mixed dimers with λ DNA through cohesive end annealing (Baldwin et al., 1966). Subsequent sequencing showed that the cohesive ends of λ, 21 and ϕ80 are identical. Another early study showed that the cohesive ends of λ-like phage ϕD326 differed from λ's *cosN*, aka $cosN^{\lambda}$, at bp positions 9 and 12 (Murray et al., 1975). More recent findings show that there is considerable natural variation in *cosN* in the λ-like phages (Ravin et al., 2000; Kropinski et al., 2007). ϕD326 and phage N15 have the same *cosN* sequence, which differ from $cosN^{\lambda}$ at bp 9 and 12 (Ravin et al., 2000, 1998). *Salmonella* phage Gifsy-1's *cosN* differs from $cosN^{\lambda}$ at positions 8, 9 and 11 (Kropinski et al., 2007). These observations led us to wonder about the extent of *cosN* sequence variation in nature. Extensive data exist in the bacterial genome sequences of the NCBI Genomes data, since λ-like prophages are ubiquitous.

2. Results

2.1. Bioinformatics of cosN sites of λ-like phages and prophages

To identify λ-like prophages, three BLASTn (Altschul et al., 1997) searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1997) were done using λ DNA sequence extending from bp 48,473 to bp 299 as the query sequence.[3] The query sequence starts at the first bp of *cosQ* and extends through *cosN* and *cosB*, to a point early in the TerS-encoding *Nu1* gene. The searches of the NCBI genomes database were done in May-June of 2016. The first search was of the virus taxid (taxid 10239). To search for prophages, two additional searches were done (Table 1). A prokarya search queried the prokarya (taxid 2) excluding *E. coli* (taxid 562), and an *E. coli* search queried the *E. coli* taxid. We included all sequences that matched the query sequence from *cosQ* through bp 50, which contains the highly conserved I2 segment that extends from bp 18 to bp 40 (Oh et al., 2016). The virus search generated 9 phages with three different *cosN* sequences. From the prokarya search, 9 different *cosN* sequences were found. The two most common sequences matched those of phages λ (74 matches) and Gifsy-1 (71 matches). The *E. coli* search generated an additional 3 *cosN* variants. All the different *cosN* sequences found are listed in Table 1, which also includes the identified viruses. Many of the *cosN* sequences recovered were in incomplete prophages. A concern for incomplete prophages is that, since the prophage segment is not viable, prophage sequences are not under selection during lytic virus growth. Hence these *cosN* sequences might contain bp changes resulting from genetic drift that occurred after the prophage suffered a lethal deletion event. To ask about prophage genetic drift we examined the nearby *cosQ* sequence to see if bp differences had occurred. Of 1521 *cosQ* sequences examined, none had a bp change. Because *cosQ* is at the beginning of the query sequence, we verified that the search program would have identified such changes. Although this is a limited study, our results suggest that genetic drift is not frequent in defective prophages, further suggesting that the defective prophages in our sample became defective relatively recently. Prophage *cosN* sites were found with as many as 5 bp differences from $cosN^{\lambda}$, with the differences confined primarily to the right half of the cohesive end sequence, namely bp 6 through 12. Single changes were found at bp 13 and 14. Markedly different distributions of *cosN* sequences were obtained for the two bacterial genome searches. In the case of the prokarya (without *E. coli*) search, 82% of the *cosN* sites were roughly equally distributed between the $cosN^{\lambda}$ and $cosN^{Gifsy\text{-}1}$ sequences. In contrast, for the *E. coli* search, 97.

[3] In the Sanger sequence (Sanger et al., 1982). Nucleotide sequence of bacteriophage lambda DNA. *Journal of Molecular Biology* 162: 729), bp 138 and 139 are G residues. Re-sequencing reveals only one G residue, so we eliminated the extra G from our query sequence. Sequence designations used in this paper are as in Sanger sequence, with bp 1 being the first base of the left cohesive end. For simplicity, the two bp to the left of bp 1 are designated as bp −1 and −2 (Fig. 2).