doi:10.1016/j.jmb.2003.11.035 J. Mol. Biol. (2004) 335, 1151–1171

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Genomic Analysis of Bacteriophages SP6 and K1-5, an Estranged Subgroup of the T7 Supergroup

D. Scholl[1]*, J. Kieleczawa[2], P. Kemp[3], J. Rush[4], C. C. Richardson[4] C. Merril[1], S. Adhya[5] and I. J. Molineux[3,6]

[1]*Section of Biochemical Genetics, The National Institute of Mental Health NIH, 9000 Rockville Pike Bethesda, MD 20895, USA*

[2]*Wyeth/Genetics Institute Cambridge, MA 02140, USA*

[3]*Microbiology and Molecular Genetics, University of Texas Austin, TX 78712, USA*

[4]*Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA*

[5]*Section of Developmental Genetics, The National Cancer Institute, NIH, Bethesda, MD 20892, USA*

[6]*Institute for Cell and Molecular Biology, University of Texas, Austin, TX 78712 USA*

**Corresponding author*

We have determined the genome sequences of two closely related lytic bacteriophages, SP6 and K1-5, which infect *Salmonella typhimurium* LT2 and *Escherichia coli* serotypes K1 and K5, respectively. The genome organization of these phages is almost identical with the notable exception of the tail fiber genes that confer the different host specificities. The two phages have diverged extensively at the nucleotide level but they are still more closely related to each other than either is to any other phage currently characterized. The SP6 and K1-5 genomes contain, respectively, 43,769 bp and 44,385 bp, with 174 bp and 234 bp direct terminal repeats. About half of the 105 putative open reading frames in the two genomes combined show no significant similarity to database proteins with a known or predicted function that is obviously beneficial for growth of a bacteriophage. The overall genome organization of SP6 and K1-5 is comparable to that of the T7 group of phages, although the specific order of genes coding for DNA metabolism functions has not been conserved. Low levels of nucleotide similarity between genomes in the T7 and SP6 groups suggest that they diverged a long time ago but, on the basis of this conservation of genome organization, they are expected to have retained similar developmental strategies.

Keywords: T7 supergroup; *Salmonella typhimurium* SP6; *Escherichia coli* K1-5; genomic analysis; comparative genomics

Introduction

The characterized members of the T7 group of phages are currently represented as a close-knit family, all of which are very similar to the prototype phage T7. The genome sequences of coliphages T7 and T3, the yersiniophages ϕYeO3-12 and ϕA1122 revealed that all essential and many non-essential genes are colinear.[1–4] Several other likely members of this group were described by heteroduplex mapping or biochemical and genetic criteria and appear to be similar in genetic organization to T7.[5–10] The *Pseudomonas putida* phage gh-1 is also related to T7 but has diverged more extensively.[11] Surprisingly, a prophage in the *P. putida* KT2440 genome contains homologues to most essential genes of T7.[12] It is not known if the prophage is inducible or whether its genes retain function.

Genome sequences of more distant T7 relatives include *Vibrio parahaemolyticus* VpV262, *Pseudomonas aeruginosa* ϕKMV, roseophage SIO1, and cyanophage p60.[13–16] These phages are morphologically similar to T7 but some lack the phage-specific

Present address: J. Rush, Cell Signaling Technology, Beverly, MA 01915, USA.

Abbreviations used: RNAP, RNA polymerase; TR, terminal repeat; dsDNA, double-stranded DNA; ORF, open reading frame.

E-mail address of the corresponding author: dscholl@usa.net

RNA polymerase (RNAP) that was once considered the hallmark feature of the T7 group. The SIO1 genome also lacks the direct terminal repeats that are characteristic of a T7-like mode of DNA replication. Collectively, these phages make up the T7 supergroup, of which the T7 group is a subset.[13] There are insufficient numbers of these less closely related members of the T7 family to conclude much about their evolution or even their strategy of infection.

SP6 was first described in 1961 as a lytic phage infecting F^- strains of *Salmonella typhimurium* LT2; the phage is known mainly for its highly specific single subunit RNAP often used in gene expression experiments.[17] SP6 has been considered to be a member of the T7 group, although notable differences are known.[18] The sequence of the SP6 RNAP is less similar to T7 RNAP than the latter is, for example, to T3 RNAP, and although SP6 primase is functionally related to T7 primase-helicase, helicase activity of the SP6 protein has not been demonstrated.[19,20] Also unlike T7, an SP6 tail protein was found to be similar to the tailspike of the otherwise unrelated temperate salmonella phage P22, an observation that suggests that more extensive genetic mosaicism is present in SP6 than among the closer relatives of T7.[21]

K1-5 infects *Escherichia coli* strains that produce either the K1 or the K5 polysaccharide capsule and encodes two tail proteins that confer this "dual" adsorption specificity.[22] An SP6-like promoter was found upstream of the K1-5 tail genes. Additionally, the K1-5 RNA polymerase was found to be nearly identical with the SP6 RNA polymerase, suggesting that the two phages may be related. It was noted that SP6 and K1-5 had similar "cassette-like" structures in the region of the genomes that encodes the tail fiber genes, suggesting that the phages acquired their particular host specificity by horizontal gene transfer.

Here, we report the complete genome sequences of SP6 and K1-5. Genome-wide comparisons reveal that SP6 and K1-5 are very closely related, only a few *ORFs* or obvious regulatory elements are present in only one of the two genomes. Limited sequence information on phage K1E[23] (D.S., unpublished data) indicates that it too may be very closely related to SP6/K1-5, suggesting that the three phages belong to a close-knit family. Although the overall genome organization of the SP6 and T7 groups appear similar, gene order is not conserved among the DNA metabolism genes, and the majority of SP6 or K1-5 genes in this region show little, if any, sequence similarity to the comparable T7 region. More surprisingly, sequence similarity searches of the SP6 and K1-5 genomes failed to reveal counterparts to several essential T7 genes. We term these phages the SP6 group, as another subset of the T7 supergroup. As such, they are more closely related to the T7 group than to more distant members of the supergroup. It appears that the T7 group and SP6 group have diverged from a common ancestor largely by vertical changes, although it is apparent that horizontal events also occurred.

Results and Discussion

SP6 and K1-5 virion morphology

Electron micrographs of negatively stained SP6, K1-5, and T7 virions are shown in Figure 1. Both SP6 and K1-5 have icosahedral heads with a face-to-face diameter of 65–68 nm, slightly larger than that of T7 (61–63 nm). The larger size of SP6 and K1-5 heads may reflect their larger capsid proteins, which in turn may be necessary to accommodate their longer genomes (see below). An internal core structure is present in both SP6 and K1-5, comparable to that found in T7 heads, but only two of the three T7 proteins that comprise the core have a sequence homologue in the other phages. Like T7, a short, stubby tail is present; additional appendages are apparent, perhaps reflecting the cell envelope hydrolytic activities carried by both SP6 and K1-5 virions.

SP6 and K1-5 genome overview

The SP6 genome is a single, linear, double-stranded DNA (dsDNA) molecule of 43,769 bp that is flanked by 174 bp terminal repeats (TR). The genome has a G + C content of 47.23%, significantly lower than its host *S. typhimurium* LT2 (G + C content of 52.22%). The K1-5 genome is also a single, linear, dsDNA molecule of 44,385 bp flanked by slightly longer 234 bp TRs. The K1-5 genome has a G + C content of 45.25%, which is significantly lower than that of *E. coli* MG1655 (50.79%). Base-pairs 1–45 of the K1-5 TR are 87% identical with base-pairs 5–48 of SP6, and the last 26 bp of the TR are 92% identical; however, the internal portion of the K1-5 TR exhibits little sequence similarity with that of SP6.

The 5′ ends of the non-transcribed strand (by the phage RNAP, see below) are defined as the left end of the genomes. On the basis of a combination of BLAST searches and visual inspection, we have assigned 52 open reading frames (ORFs) on both the SP6 and K1-5 genomes. The genomes are close-packed; 92% of the genomes are occupied by

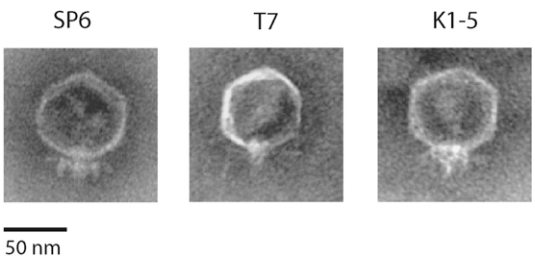

Figure 1. Negatively stained electron micrographs of SP6, K1-5 and T7, showing the morphology typical of *Podoviridae*.

coding sequence. Almost all K1-5 ORFs have SP6 counterparts, only the tail-associated proteins that degrade extracellular polysaccharide and a few small ORFs of unknown function are restricted to one or the other in one phage genome. All predicted genes are transcribed from one strand. The longest potential *ORF* on the opposite strand contains 219 and 201 codons for SP6 and K1-5, respectively, neither of which is preceded by a recognizable Shine-Dalgarno sequence and are therefore unlikely to be expressed. Most non-coding regions contain a recognizable putative regulatory signal. The longest non-coding regions are 399 bp for SP6 and 369 for K1-5, both of which lie between the left terminal repeat and a putative host promoter. Short overlaps between two coding regions are frequent, with the longest coding overlap being 90 and 87 bases between *ORF19* and *ORF20* in SP6 and K1-5, respectively. One UUG and two GUG initiation codons are annotated for each phage genome, all other initiation codons are AUG and, with the exception of ORF27 and K1-5 ORF11, all are preceded by good Shine-Dalgarno sequences.

One of the hallmarks of the known genome sequences of the T7 group is that there is a bias against many type II restriction enzyme sites; selection for site avoidance seems to be the primary mechanism of defense against type II enzymes. This selection has been less intense for K1-5 and SP6. The T7 genome lacks sites for both *Eco* RI and *Eco* RV, both of which are statistically expected to be present ten times. The K1-5 genome lacks *Eco* RI sites but contains 12 *Eco* RV sites, whereas SP6 contains six and 16 sites, respectively, for these enzymes. There are only two sites in T7 for *Eco* RII, and the recognition site of *dcm* methyltransferase, but K1-5 and SP6 contain ten and 18 sites, respectively. About 80 sites are expected statistically for this sequence (5′-CCWGG). The sequence 5′-GATC is the recognition site for a number of restriction enzymes and is the target for the *dam* methyltransferase, which is present in both *E. coli* and *S. typhimurium*. *dam* recognition sites have been selected against in the genomes of T7, T3, and the yersiniophages ϕA1122 and ϕYeO3-12. On statistical grounds, the 5′-GATC sequence is expected to occur about 150 times in each genome, but its distribution in the close-knit T7 group actually ranges from three to ten. There has been less bias against 5′-GATC sequences in both the K1-5 and SP6 genomes, which contain 29 and 53 sites, respectively.

The SP6 genome does, however, contain one clear example of restriction site avoidance that is equivalent to the insensitivity of T7 to restriction by *Eco* P15.[24] The type III enzyme *Sty*LT I recognizes the asymmetric sequence 5′-CAGAG/GTCTC-5′ but restriction requires two copies of the sequence in a head-to-head orientation.[25] As an asymmetric pentamer, this sequence is expected 85 times in the SP6 genome, but is actually present only 12 times. Furthermore, all 12 copies of 5′-CAGAG are on the same strand, i.e. in a head-to-tail orientation, and the *Sty*LT I enzyme cannot therefore restrict SP6 DNA. In contrast, the K1-5 genome contains 102 copies of 5′-CAGAG that are divided equally between the two DNA strands.

SP6 and K1-5 coding sequences

There is insufficient conservation, both in the order of genes that have obvious counterparts with the close-knit members of the T7 family of phages and in the number of genes that are unrelated by sequence to the T7 family, to maintain the T7 gene numbering system. Genes (*ORFs*) have therefore been numbered sequentially from the left genome end. Table 1 shows those that have T7 homologues. *ORFs* that are present in one genome only and where the other contains no genetic information have been assigned non-integral numbers. In assigning numbers we have coordinated with the authors of the SP6 genome annotation (Genbank accession number AY288927), who obtained their sequence independently of this work. The two genome sequences agree completely.

Most identifiable ORFs are homologues of T7 family genes (Table 1). Like T7, the SP6 and K1-5 genomes can be divided roughly into three regions containing the early, middle, and late genes. A putative terminator that is similar to T7 TE lies immediately downstream of ORF7, the gene for RNAP, and defines the end of the early region. The middle regions encode several putative DNA metabolism functions, it ends after *ORF24*, a DNA ligase, and before the first recognizable structural gene, *ORF29*, which codes for the head-tail connector. Unlike T7, the beginning of the late region cannot be defined by the presence of a consensus late promoter. Curiously, in SP6 but not K1-5, a promoter-like sequence overlaps the 5′ end of *ORF30*; however the sequence is not predicted to be an active promoter (see Table 2, below).

Predicted ORFs and regulatory elements for both SP6 and K1-5 are shown in Table 1. Two easily recognizable ORFs are present in the early regions of both genomes. Both SP6 and K1-5 code for a putative anti-restriction protein, ORF3, that shares similarity with T7 gp0.3. Sequence similarity between the SP6 and K1-5 ORF3 is only 45% but T7 and T3 gp0.3, which are known to have the same role *in vivo*, are completely different in their primary amino acid sequence.[26] ORF3 is likely an active anti-restriction protein, as SP6 grows equally well on *S. typhimurium* LT2 with or without its natural type I restriction systems (I.J.M., unpublished results); comparable information is not available for K1-5. The other easily identified early gene codes for the phage RNAP (ORF7), whose sequence is 85% identical between SP6 and K1-5. The SP6 enzyme has been well characterized.[19,27,28] Both phages encode several other putative ORFs in the early region whose function cannot be predicted; aside from ORF5, which is conserved only by the predicted number of amino acids and

Table 1. SP6 and K1-5 coding regions and regulatory elements

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
	S 1–174							Terminal repeat
	K 1–234							
	S 174–180							CJ terminator
	K 234–240							
	S 576–611							Host promoter
	K 609–643							
0.2	S 711–884	TAACTGAGGTTACTATC**ATG**	57					No K1-5 homologue
0.6	S 998–1099	CTTAATGAGGTGATACT**ATG**	33					No K1-5 homologue
1	S 1096–1275	GCTAGGGGAGGTGCTGA**ATG**	59	75				
	K 781–960	TGAATGAGGTATTAACT**ATG**	59					
2	S 1268–1477	ACTGAAGGAGATAATCC**ATG**	69	24				
	K 960–1184	TTGAGGGAGGGTAAATA**ATG**	74					
	S 1564–1624							Potential RNase III recognition site
	K 1491–1566							
3	S 1637–1996	ACTAAGAGGAAATAGAA**ATG**	119	45	T7 0.3 (4.6e-39)	T7 0.3 (1.8e-21)	gp0.3 (77/48)	Anti-type I restriction
	K 1575–1907	TAAATGAGGATTAAATC**ATG**	110					
4	S 1997–2113	AAGAAGAGGACGAATAA**ATG**	38	70				
	K 1910–2023	GAAGAGGAAGAATAAGG**ATG**	37					
	S 2104–2126							Cryptic phage promoter
	K 2014–2036							
	S 2125–2162							Potential RNase III recognition site
	K 2034–2079							
5	S 2181–2372	AATTTGATGAGGCGATT**ATG**	63	Insignificant				
	K 2100–2288	TACTAGGAGTTGAGATT**ATG**	62					
	S 2385–2407							Cryptic phage promoter
	K 2323–2345							
6	S 2435–3319	CATATGGAGAACATACC**ATG**	294	67				
	K 2353–3237	TATATAAGGTGATTACT**ATG**	294					
	S 3324–3366							Potential RNase III recognition site
	K 3236–3278							
7	S 3392–6016	CCGCATGAGGATACAAG**ATG**	874	85	Several phage	Several phage	gp1 (47/45)	Phage-specific RNA polymerase
	K 3301–5928	CACTAGGAGTAAACAAG**ATG**	875		RNAP	RNAP		
	S 6035–6055							Early region terminator TE for host RNAP
	K 5936–5959							
	S 6118–6140							Phage promoter
	K 6020–6042							
	S 6287–6309							Phage promoter

(*continued*)

Table 1 Continued

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
	K 6204–6226							
8	S 6340–6468	TCATTGGAGAATTAATC**ATG**	42	70	T7 gp1.1 (0.0013)	T7 gp1.1 (8.7e-6)	gp1.1 (45/52)	Conserved in T7 group, but non-essential and no known function
	K 6252–6380	ACTTTGGAGATTTAACC**ATG**	42					
9	S 6472–8457	TGTGGGAGTACTAAGTT**ATG**	661	92	Several phage primase/helicase	Several phage primase/helicase	gp4 (38/38)	Primase
	K 6384–8372	CGTGGGAGTTCTAAGTT**ATG**	662					
	S 8411–8433							Phage promoter embedded in *ORF9*
	K 8320–8342							
10	S 8601–9212	ACTAACT*GGAG*ATTATT**ATG**	203	43				
	K 8448–9131	ATAACTGGAGATTGATT**ATG**	227					
	S 9112–9144							Phage promoter embedded in *ORF10*
	K 9010–9032							
10.5	K 9142–9360	TATCTAAGGAGTTAATC**ATG**	72					No SP6 homologue
11	S 9199–9369	CGTTACGGAGGAAACCT**ATG**	56	66				
	K 9350–9559	CAGCTTCAT*GGT*GAAAT**ATG**	69					
12	S 9431–9592	TAATTGGAGATACATAA**ATG**	53	60				
	K 9618–9770	TGATTGGGAGGTATTAA**ATG**	50					
13	S 9579–12128	TGAAACGGAGTTAATGT**ATG**	849	89	Roseophage SIO1 DNAP (7.2×10^{-35})	Roseophage SIO1 DNAP (4.0×10^{-32})	gp5 (40/37)	DNA polymerase
	K 9757–12297	TGAGACGGAGTTAATGT**ATG**	846					
13.5	K 12297–12413	TGGAAGGAAACACACTG**ATG**	38					No SP6 homologue
14	S 12128–12226	TGGAAGGAGACACACTA**ATG**	32	88				
	K 12413-12511	GAGGAGGGTTACGATTG**ATG**	32					
	S 12541–12563							Phage promoter
	K 12580–12602							
15	S 12629–13006	ATATAGGAGATATAAAC**ATG**	125	91				
	K 12667–13044	ATATAGGAGATATAAAT**ATG**	125					
15.5	K 13057–13353	AATAAGGGGGGGTTATG**ATG**	98					No SP6 homologue
	S 13005–13027							Phage promoter
	K 13351–13373							
16	S 13216–14022	ACAGGAGGAATAAATTA**ATG**	268	85				
	K 13523–14329	AATAGGAGACTAAATAA**ATG**	268					
17	S 14036–14254	TTAATCAAGGAGGTTTA**ATG**	72	68				
	K 14339–14557	AATATTAAGGAGAAAGG**ATG**	72					
	S 14286–14308							Phage promoter
	K 14590–14612							
18	S 14359–14727	AGAAGGAGAATTTAAAT**ATG**	122	92				
	K 14663–15034	CAAAAGGAGAATTTAAT**ATG**	123					
	S 14734–14759							Putative terminator
	K 15094–15119							
19	S 14792–15100	TTATTGGAGAATGAATT**ATG**	102	49				
	K 15152–15424	TTATTGGAGAATGAATT**ATG**	90					

(*continued*)

Table 1 Continued

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
20	S 15010–16038	TTAAAGGAGTGGCCTAC**ATG**	342	83	*P. aeruginosa* phage PaP3 exonuclease (1.3×10^{-12})	*P. aeruginosa* phage PaP3 exonuclease (1.0×10^{-10})	gp6 (insignificant)	Exonuclease
	K 15337–16365	CTGAAGGAGTGGCCTAC**ATG**	342					
21	S 16023–16433	TGGGGAGATGATGATTC**ATG**	136	87	T4 endonuclease VII (5.2×10^{-8})	T4 endonuclease VII (5.2×10^{-8})	gp3 (insignificant)	Holliday-junction endonuclease
	K 16350–16760	TGGGGAGATGATGCGGA**ATG**	136					
22	S 16426–17433	AAAGAAGGAGGGTAAAC**TTG**	335	94				
	K 16753–17760	AAAGAAGGAGGTTAAAG**TTG**	335					
	S 17428–17450							Phage promoter
	K 17755–17777							
23	S 17534–18004	GAATGGAGGAATTAATT**ATG**	156	63 (95% for C-terminal 80 aa)	T7 gp 1.7 (1.6×10^{-20})	T7 gp1.7 (1.3×10^{-20})	gp1.7 (38/37); 63% identity over C-terminal half	Only the C terminus of gp1.7 is conserved in the T7 group
	K 17861–18364	TGAATGGAGGATTCAAT**ATG**	167					
24	S 18004–18951	ATTGATGAGGATCTCTA**ATG**	315	80	Uncultured crenarchaeote 74A4 ATP-dependent ligase (0.13)	Uncultured crenarchaeote 74A4 ATP-dependent ligase (0.030)	gp1.3 (insignificant)	DNA ligase
	K 18366–19313	TGATGAGGATATTTGAT**ATG**	315					
25	S 18923–19135	GGTTGGAGAGTTGCGCC**ATG**	70	71				
	K 19285–19500	AGCGGGAGAATTAAGAC**ATG**	71					
26	S 19101–19274	GAAAGTGGAGGAGTTGT**ATG**	57	58				
	K 19463–19633	GAAGGTTGACGAATTAT**ATG**	56					
27	S 19271–19732	CAGTTTGTGGTGCAGCC**ATG**	153	75	Ralstonia acetyltransferase ($12. \times 10^{-6}$)	Roseophage SIO1 gp26.2 (1.7×10^{-9})		
	K 19630–20157	GCACAACTGGTGGTCCA**ATG**	175		Roseophage SIO1 gp26.2 (3.8×10^{-5})			
28	S 19742–19951	AGTTTTAAGGAGATAAC**ATG**	69	82				
	K 20167–20373	AGTTCTAAGGAGATAAC**ATG**	68					
	S 19951–19973							Cryptic phage promoter
29	S 19953–21500	TAGCTTAGGAGTTTGAT**ATG**	515	91	*P. putida* KT2440 prophage (3.1×10^{-58})	*P. putida* KT2440 prophage (1.2×10^{-55})	gp8 (34/33)	Head-tail connector.
	K 20375–21925	TAGCTTGA*AGGTG*TAAT**ATG**	516					Found as virion protein
30	S 21504–22406	TGAAGGAGGGTTAATTA**GTG**	300	70	Cyanophage p60 protein (4.3×10^{-8})	Cyanophage p60 protein (7.5×10^{-7})	gp9 (Insignificant)	Scaffolding protein
	K 21925–22845	GAACTTAAGGAGGCGTAA**ATG**	306					
	S 22412–22434							Phage promoter
	K 22851–22874							

(*continued*)

Table 1 Continued

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
31	S 22486–23691	TAAGGAGATTATAATAC**ATG**	401	85	*P. putida* KT2440 prophage protein (0.0064)	T7 capsid protein (0.3)	gp10 (30/30)	Major capsid protein.
	K 22925–24133	TAAGGAGATTATAATAC**ATG**	402					Found as virion protein
	S 23697–23729							Phage RNAP terminator Tϕ
	K 24142–24166							
32	S 23744–24484	TTTGTTTAGGAGGATTC**ATG**	246	86	*P. putida* KT2440 prophage tail protein (7.9×10^{-12})	*P. putida* KT2440 prophage tail protein (3.6×10^{-9})	gp11 (32/32)	Tail protein
	K 24188–24925	TTTTGTTAGGAGGATTC**ATG**	245					Found as virion protein
33	S 24484–26895	TGGGAGAGTTATCGCTA**ATG**	803	82	*P. putida* KT2440 prophage tail protein (2.0×10^{-65})	*P. putida* KT2440 prophage tail protein (4.8×10^{-76})	gp12 (36/35)	Tail protein.
	K 24925–27327	CCGTGGGAGGATCGTTA**ATG**	800					Found as virion protein
34	S 26896–27615	CTAAGAGGAGAGTGTAA**ATG**	239	64	*P. putida* phage gh-1 internal virion (1.6×10^{-5})	*P. putida* phage gh-1 internal virion (0.43)	gp14 (Insignificant)	Internal head protein. Found as virion protein
	K 27327–28049	ACCAAAAGGAGGGTCTA**ATG**	240					
35	S 27616–30552	GAACAAGAGGAGTTTAA**ATG**	978	43			gp15 (Insignificant)	Internal head protein with murein hydrolase activity.
	K 28049–30997	GGAACAA*GAGGT*ATTTA**ATG**	982					Found as virion protein
	S 29393–29414							ϕ promoter (SP6; embedded in *ORF35*)
	K 31037–31059							
	S 30550–30604							Potential RNase III processing site
	K 30995–31051							
36	S 30619–34431	TTAACTAAGGAGGTAAC**ATG**	1270	40	*P. putida* KT2440 prophage (1.3×10^{-10})	*P putida* KT2440 prophage (3.2×10^{-13})	gp16 (weak/36)	Internal head protein, T7 gp16 homologue but lacks murein hydrolase activity.
	K 31062–34247	GGCACTATGGAGGAATT**ATG**	1061					Found as virion protein
37	S 34431–35390	GACTTATTGGAGGATTG**ATG**	319	71	*P. putida* KT2440 prophage (.0015)	*P. putida* KT2440 prophage (0.00066)	gp17 (Insignificant)	Tail fiber
	K 34247–35209	GATGAACTGGAGGATTG**ATG**	320					Found as virion protein
38	S 35399–35596	ACTTTATAGGAGGTAAG**ATG**	65	78	T7 gp17.5 (.008)	T7 gp17.5 (.008)	gp17.5 (31/31)	Holin
	K 35219–35413	CAGCATAAGGAGGTGGG**ATG**	64					

(*continued*)

Table 1 Continued

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
39	S 35580–35879	ATGGAAGGAGAAGAAGC**GTG**	99	77			gp18 (30/insignificant)	Small terminase subunit
	K 35397–35696	GATTTGGAGGGAGAAGC**GTG**	99					
	K 35488–35510							Phage promoter embedded in *ORF39* (not in SP6)
40	S 35879–37774	TTTGCTAAGGAGGCTTA**ATG**	631	80	*P. putida* KT2440 prophage (6.0×10^{-78})	T7 maturase (5.4×10^{-77})	gp19 (45/46)	Large terminase subunit
	K 35696–37594	TTCACTAAGGAGGATTG**ATG**	632					
	S 37773–37795							Phage promoter
	K 37593–37615							
41	S 37921–38220	ATAGATAAGGAGTAATA**ATG**	99	36				
	K 37756–38034	ATAATAGGAGAGAGACA**ATG**	92					
42	S 38235–38534	TAAGAAGGAGATAACAT**ATG**	99	49				DD-peptidase similarity
	K 38053–38343	ATAAGGAGGAAGATTAC**ATG**	96					
42.3	S 38544–38789	GCCGCTAAGGAGGCTAG**ATG**	81					No K1-5 homologue
43	S 38982–39332	GACTTACTGGAGTGATA**ATG**	116	79				Muramoyl pentapeptide carboxypeptidase similarity
	K 38346–38690	CCAGCAGGGGTATAATA**ATG**	114					
44	S 39444–39593	AAATGGAGAGCATTACG**ATG**	49	88				
	K 38799–38948	ATGGAAAGGGAGATGCA**ATG**	49					
45	S 39671–39880	CCAGAAGGAGGCGCAAC**ATG**	69	75				Lipoprotein similarity
	K 39026–39235	AGAGAAGGAGGCACAAA**GTG**	69					
	S 39909–39931							Phage promoter
	K 39265–39287							
46	S 40011–41663	TAGTAATAGGAGGTTTA**ATG**	550		SP6 tailspike			Tailspike protein. Similarity does not include tail-binding domain. Found as virion protein
46	K 39367–41265	AAGTTAGGAGATAGC**ATG**	632			K1-5 K5 lyase		K5 Lyase. Found as virion protein
	S 41669–41689							Stem-loop, putative terminator (inefficient)
	K 41273–41294							
	S 41698–41738							Potential RNase III recognition site
	K 41303–41342							
47	S 41747–43261	TCAGAAAAGGAGGTGAC**ATG**	504					Endogalacturonase similarity. Found as virion protein

(*continued*)

Table 1 Continued

ORF	Position	Translation initiation region	Amino acid residues	Identity (%)	Most significant BLASTp similarity (SP6)	Most significant BLASTp similarity (K1-5)	Identity with T7 homologue (SP6/K1-5) (%)	Function/comments
47	K 41351–43786′	TCAGAAAAGGAGGTTAC**ATG**	811			K1-5 endosialidase		K1 endosialidase. Found as virion protein
48	S 43269–43412	TATTCAAGTGAGGCTGT**ATG**	47	87				K5 lyase, K1 endosialidase similarity
	K 43764–43913	CAAGTTAGAGGAGATGT**ATG**	50					
49	S 43466–43543	TAATCAAAGGAGATAAC**ATG**	25	24				
	K 43984–44112	ATTCCAAAGGAGGTAAC**ATG**	42					
	S 43596–43769							Terminal repeat
	K 44151–44385							

Only BLASTp search results that indicated some logical function are included. Amino acid alignments were done using GCG gapped comparisons.

Table 2. Promoters and promoter-like elements for SP6 RNA polymerase

Transcription start	Promoter sequence[a]	Comments
+1		
6135	ATTTAAGGGACACTATAGGACTA	
6304	ATTTACGTGACACTATTGAACTA	
8428	AATTAGGTGACACTATAGAAGAG	Embedded in *ORF9*
9129	TATTACGTGACACTATAGGTGGG	Embedded in *ORF10*
12559	ATTTAGGTGACACTATAGGAGGA	
13022	AATTAGGTGACACTATAGAACAA	
14303	ATTTAGGGGACACTATAGAAGGG	
17445	ATTTAAGTGACACTATAGAACAA	
22430	ATTTAGGTGACACTATAGAAGGG	Consensus
29410	CGCTAGGTGACACTATAGGTGCC	Embedded in *ORF35*
37790	AATTAGGGGACACTATAGAAGGA	
39927	ATTTAGGTGACACTATAGAATAG	
	ATTTAGGTGACACTATAGAAGRR	**Consensus**
Transcription start (+1)	Promoter-like sequence[b]	Comments[b]
2121	AGTTATGTGACACTATAAGATGA	+1A
2402	TCATTCGTGGCACTATGTGAAAT	+1T, − 8G, − 13T
19968	ATATGCAAGACACTATACTTGAG	+1C, − 11A; embedded in *ORF29*

[a] Underlined bases at −12, −11, and −10 indicate the likely specificity difference between SP6 and K1-5 RNAP.
[b] + 1(A, T, C), −8G, −11A, and −13T have weak activity.[35]

genome position, these proteins share from 24% to 70% sequence identity. SP6 *ORF0.2* and *ORF0.6*, the first genes likely to be expressed after infection, have no counterparts in the K1-5 genome and their functions cannot be predicted from database homology searches.

The middle region of SP6 and K1-5 encodes several highly conserved ORFs that are not found in the T7 family of phages and whose functions yet cannot be predicted. However, three K1-5 proteins, ORF10.5, ORF13.5, and ORF15.5, have no SP6 homologues. Like T7, many of the genes in this region encode DNA metabolism functions but the relative order of the SP6 and K1-5 genes is different from that of the close T7 relatives. SP6 and K1-5 ORF8 are similar to T7 gp1.1, a protein that is conserved in many members of the T7 family but that has no known biological function. *ORF9* codes for a primase, which at 92% amino acid identity is one of the most conserved proteins encoded by the two phages. *ORF13* codes for DNA polymerase (DNAP), which at 89% identity is also highly conserved. Both proteins exhibit about 50% sequence identity with their T7 homologues. Three other putative DNA metabolism genes have been identified that are highly conserved between SP6 and K1-5 but are not closely related to their presumed T7 counterparts: ORF20 is most closely similar to *P. aeruginosa* phage P3 exonuclease, ORF21 to phage T4 endonuclease VII, and ORF24 is predicted to be an ATP-dependent DNA ligase. ORF23 is related by amino acid sequence to T7 gp1.7. T7 gene *1.7* is expressed early in the class II region of T7, the protein is known to be beneficial for phage growth but its function has not been defined.[29] However, unlike T7 gene *1.7*, SP6 and K1-5 *ORF23* is located near the end of the middle gene cluster. Only the 3′ half of gene *1.7* is conserved in the T7 group, suggesting that the C-terminal half of gp1.7 contains the important function.[4] In agreement with this idea, overall sequence identity between SP6 and K1-5 ORF23 is 63% (Table 1) but their C-terminal 80 residues are 95% identical. Similarly, the overall sequence identity of each protein with T7 gp1.7 is ~37.5% but this value rises to 63% for the C-terminal region.

We tentatively define the late region as beginning with the first clearly recognizable structural gene (*ORF29*), which codes for the head-tail connector. Several of the proteins encoded in this region have been identified as structural components of the virion (Figure 3). The coding sequence for ORF29 is followed immediately by that for the scaffolding protein (ORF30), major capsid protein (ORF31), and two tail proteins (ORF32 and ORF33). ORF34 has closest similarity to an internal virion protein of the T7 relative gh-1. The two largest ORFs encoded by SP6 and K1-5 are likely to be internal core proteins that function in the initial stages of infection.[30] ORF35 has recently been shown to hydrolyze peptidoglycan *in vitro*.[31] SP6 ORF35 and ORF36 show less similarity to their K1-5 counterparts than most other major proteins. ORF37 is a short (relative to other members of the T7 group) tail fiber protein, and ORF38 is most closely related to T7 gp17.5, a typical class II holin with two expected transmembrane domains.[32] The following two ORFs are predicted to be the two subunits of terminase, the enzyme that packages replicated DNA into proheads. ORF39 exhibits weak similarity to T3 and ϕA1122 gp18, whereas ORF40 is closely related to the large terminase subunit common to all current members of the T7 group, and to cyanophage P60.

Downstream of the gene for the large terminase subunit lies a phage promoter and several ORFs that, in addition to those discussed below, code for small proteins with sequence similarities to various cell envelope-hydrolytic activities. These ORFs probably do not retain enzymatic activities, they may be remnants of recombination events that caused a host range switch.

The most striking difference between K1-5 and SP6 are the genes whose products likely form tail appendages. K1-5 encodes both a lyase (ORF46) that allows it to infect *E. coli* strains that display the K5 polysaccharide capsule and an endosialidase (ORF47) that confers specificity to *E. coli* K1 capsulated strains. It has been shown that both proteins are part of a single virion.[22] Neither the endosialidase nor the lyase is present in SP6 which, instead, encodes a protein (ORF46) that is closely related to the tailspike protein of the salmonella phages ST46T and P22. P22 tailspike is an endorhamnosidase involved in degrading the *Salmonella* O-antigen, and it seems likely that SP6 ORF46 has the same activity. However, unlike P22, which adsorbs to only smooth strains of *S. typhimurium*, SP6 grows well on both smooth and rough strains. Immediately downstream of the SP6 tailspike gene lies ORF47, whose product exhibits some similarity to a cell-wall hydrolase. The first 11 amino acid residues of SP6 ORF47 are identical with those of the K1-5 endosialidase but beyond that there is no sequence similarity. ORF47 is a structural protein and we suggest that it is also part of the SP6 tail structure, and that it confers specificity to another, unknown, bacterial host.

One notable feature of the SP6 and K1-5 genomes is that two-thirds of the predicted ORFs in the early and middle regions have no matches with clear biological significance to any database protein. In contrast, almost all the predicted late ORFs, which are known or expected to have structural or morphogenetic roles, show sequence similarity to proteins with functions that are anticipated to have a beneficial function during phage growth. A second feature is that 48 of the 105 predicted ORFs shown in Table 1 are less than 100 amino acid residues in length. Of these, only the holin (ORF38) and small terminase subunit (ORF39) have predicted activities. ORF8 (T7 gp1.1) is also highly conserved in the T7 group but has no known biochemical activity.

SP6 and K1-5 regulatory elements

Consistent with the assignment of ORF40 as a T7-related large terminase subunit, K1-5 and SP6 have extensive terminal repeats, suggesting that they have similar schemes for DNA replication and packaging. We have annotated one putative sigma70 promoter for early gene expression (Table 1), additional host promoters are likely but need to be identified experimentally.

Most SP6 and K1-5 transcription occurs using their respective phage-encoded RNAPs. Several SP6 promoters have been identified and cloned and transcriptional activity with respect to promoter sequence has been examined comprehensively by saturation mutagenesis.[33,34] On the basis of this information we have identified 12 phage RNAP promoters in the SP6 genome. Only the promoter upstream of ORF31 conforms exactly to the consensus. Three promoter-like sequences that contain one or more nucleotide changes at conserved positions, including + 1G that is required for SP6 RNAP transcription initiation *in vitro*, were found but are likely not to be functional.

K1-5 RNAP has not been studied biochemically but we predict the promoter sequence to be similar to SP6. As such, we have identified 15 promoter or promoter-like elements in the genome (Table 3), four match the consensus exactly. Most K1-5

Table 3. Promoter-like elements in K1-5

Transcription start	Promoter-like sequence	Comments[a]
+1		
2031	ATTTAGCTGACACTATAAGAGAA	+1A
2340	ATTTACTTAACACTATATAAGGT	+1T, − 9A
6037	ATTTACCGGACACTATAGGATAG	
6221	ATTTACCGGACAGTATAGATAAG	−5G
8337	ATTTACCGGATACTATAGAAGAG	−7T; embedded in *ORF9*
9027	ATTTTGCCGACACTATAGAAGGC	embedded in *ORF10*
12597	ATTTACTGGACACTATAGAAGGA	Consensus
13369	AATTACTAGACACTATAGAACAA	
14607	ATTTACTGGACACTATAGAAGAG	Consensus
17772	ATTTAGTTGACACTATAGAACAA	
22869	ATTTACTGGACACTATAGAAGGG	Consensus
31054	TTATGATAGGCACTATGGAGGAA	−8G
35505	CAATACTGGACACTATAGAAGAA	Embedded in *ORF39*
37610	AATTACTGGACACTATAGAAGGA	
39282	ATTTACTGGACACTATAGAAGAG	Consensus
	ATTTACYG**GACACTATAGAAGRR**	Consensus

Underlined bases at −12, −11, and −10 indicate the likely specificity difference between K1-5 and SP6 RNAP.
[a] By analogy with SP6 promoters,[35] +1(A, T), −5G, −7T, −8G, and −9A are predicted to have weak activity.[35]

promoters are in positions equivalent to those in SP6, but the SP6 promoter-like sequence within *orf29* and the K1-5 promoter upstream of *orf40* have no counterpart in the other genomes. In addition, the SP6 promoter upstream of *orf36* is embedded within *orf35*, whereas in K1-5 it lies within a non-coding region. Like its SP6 counterparts, the K1-5 promoter-like sequences at positions 2031 and 2340 lack the consensus + 1G by analogy with SP6[34] and are thought unlikely to be active promoters. The assignment of two other promoters should also be regarded as tentative. At position 8337, a promoter-like sequence has −7T, another at position 31054 has −8G; neither SP6 nor T7 RNAP tolerate changes well at these positions of their promoters.[34,35] However, the equivalent SP6 promoters are expected to be active *in vivo*.[34]

The consensus promoter sequences for SP6 and K1-5 differ at positions −10, −11, −12 (Tables 2 and 3). The SP6 consensus −12G and −10T are not major determinants of promoter activity *in vivo* or *in vitro* but any substitution at −11G is inactivating.[34] All K1-5 promoters have a pyrimidine at −11, suggesting that SP6 RNAP will not transcribe K1-5 DNA and *vice versa*. A comparable specificity of the phage RNAP for its cognate promoters is seen throughout the T7 group; for example, T7 RNAP does not recognize T3 promoters and *vice versa*. However, a single amino acid change in T7 RNAP is sufficient to switch its specificity to T3 promoters, and a change at −11 in a T7 promoter is sufficient to allow its recognition by T3 RNAP.[36,37] Two changes at positions −9 and −8 of the SP6 promoter to those of a T7 promoter allow recognition by T7 RNAP *in vitro*.[38] All SP6 and most K1-5 promoters have the sequence 5′-GA at these positions, suggesting that the K1-5 genome cannot be transcribed efficiently by T7 RNAP.

Transcription termination of host RNA polymerase in both phages probably occurs immediately downstream of ORF7, the phage RNAP. SP6 nucleotides 6035 through 6063 and K1-5 nucleotides 5936 through 5966 can be drawn as similar stem-loop structures followed by a T-rich segment. The two sequences are both about 66% identical with the T7 early terminator TE and the predicted secondary structures are also essentially the same. A likely terminator for phage RNAP lies downstream of *orf31*, which codes for the major capsid protein in both phages. This is the equivalent position to the characterized T7 major terminator Tϕ and, as in T7, termination is unlikely to be 100% efficient because there is no promoter between the terminator and the two downstream tail genes, *orf32* and *orf33*. Both ORF32 and ORF33 are found in mature virions (Figure 3) and are expected to constitute the stubby tail. By analogy with T7, both proteins should be essential for infectivity.

We tentatively identified additional terminators in both the SP6 and K1-5 genomes (Table 1). One lies downstream of *orf18* in the middle region of the genomes, another downstream of *orf46*, coding for SP6 tailspike or K1-5 lyase. Neither putative terminator can be fully efficient, as genes immediately downstream would not then be expressed. ORF20 and ORF21 are, respectively, an exo- and endonuclease. Both proteins are expected to be required for phage growth, their T7 counterparts are the essential gp6 and gp3. Furthermore, ORF47 of both phages has been shown to be a virion protein (Figure 3). Visual inspection of the genome sequences reveals that both SP6 and K1-5 contain counterparts to the T7 CJ pause/terminator element.[39–41] This element lies immediately downstream of the concatemer junction in replicating DNA, which corresponds to immediately downstream of the left terminal repeat in genomic DNA. A nine base sequence corresponding to CJ is highly conserved in members of the T7 and SP6 groups, although it is not found in the more distant relatives ϕKMV and VpV262 (Figure 4). Five to seven bases downstream of CJ is a conserved CTCC sequence that includes the site of termination at T7 CJ.[40] Pausing at CJ is enhanced greatly by lysozyme and is known to be important for T7 DNA packaging.[39] Interestingly, however, neither SP6 nor K1-5 code for a protein with significant sequence similarity to T7 lysozyme.

Comparison of SP6 and K1-5 with T7

The T7 group of phages whose genome sequences are known include the coliphages T7 and T3, the yersiniophages ϕYeO3-12 and ϕA1122, and the *P. putida* phage gh-1. The genetic organization of these phages is co-linear and most T7 proteins have obvious sequence homologues coded by the other genomes. T7 has been shown to recombine with T3 and with ϕYeO3-12, and it has been suggested that ϕYeO3-12 and ϕA1122 are ancestral to T3.[2,4,42–44] SP6 and K1-5 have diverged much further from T7 than these other phages, and they have diverged further from each other than members of the T7 group have from T7 (or themselves). Nevertheless, although substantial sequence divergence has occurred, most SP6 and K1-5 gene products remain clearly related to each other. Even when sequence similarities are not apparent, many SP6 gene products have comparably sized counterparts in K1-5, and presumed regulatory elements occupy equivalent positions on the two genomes. We therefore designate SP6 and K1-5 as members of the SP6 group, which likely contains the phages K1E and K5. The latter phages maintain the morphology, phage promoter sequence (upstream of the tail genes) and tail fiber endosialidase or lyase of K1-5.

Similarities between counterpart proteins of the SP6 and T7 groups are modest in number, 13 ORFs show about 30–52% amino acid identity but one (SP6 ORF3) shows 77% (Table 1). However, these include many essential T7 proteins and the biology of the SP6 and T7 groups are likely to prove similar. The major patterns of the T7 lifecycle: transcription by a phage-specific RNAP, DNA metabolism by phage-encoded DNA polymerase,

primase, and DNA maturase, together with the common structural components of the virion suggest that the SP6 group can be compared directly to the T7 group.

On the basis of sequence comparison, only two early functions, antirestriction and RNAP, are predicted to be conserved between the SP6 and T7 groups. In T7, the conditionally essential genes *1.2* (anti-dGTPase) and *1.3* (DNA ligase), are expressed both early after infection and as class II or middle genes.[45] Neither SP6 or K1-5 contains obvious counterparts to T7 gene *1.2*, although both *S. typhimurium* and *E. coli* encode a deoxyguanosine triphosphatase (*dgt*). Over-expression of *dgt* lowers intracellular levels of dGTP and inhibition of the dGTPase by T7 gp1.2 is required for phage growth.[46] It is expected, therefore, that SP6 and K1-5 would not grow in cells over-expressing *dgt*. The female host-specificity of SP6 may be due, in part, to the lack of a gene *1.2* homologue.[17,47,48] SP6 and K1-5 ORF24 (DNA ligase) is located close to the late gene cluster and is not predicted to be expressed early. Whether SP6 or K1-5 DNA ligase is essential, as in T4, or only conditionally essential, as in T7 and T3, is not known.

The cluster of class II or DNA metabolism genes is arranged differently in the SP6 group, relative to T7 (Figure 2). It is not known whether gene order is important in this region, but moving the RNAP gene to different locations on the genome is known to cause significant changes in the T7 growth-cycle.[29] SP6 and K1-5 encode close relatives (ORF13 and ORF9, respectively) to both T7 DNA polymerase and primase-helicase. However, helicase activity could not be demonstrated with the SP6 enzyme.[20] The third phage protein necessary for *in vitro* T7 DNA replication, SSB, has no obvious homologue in the SP6 or K1-5 genomes. This is surprising, since T7 SSB interacts with both T7 DNAP and primase-helicase and is essential for phage growth *in vivo*.

A second gene that was expected from our knowledge of the T7 life-cycle but has not been recognized by sequence inspection in the SP6 and K1-5 genomes is a homologue to T7 gene *2*. In infected *E. coli* B or K-12 cells the failure to inhibit host RNAP leads to premature degradation of T7 DNA, particularly near the left genome end.[49] SP6-infected cells are known to contain an inhibitor of bacterial RNAP (S. Nechaev & K. Severinov, personal communication) but no protein with similarity to either T7 gp0.7 or gp2 has been assigned. There are also no sequence homologues to T7 lysozyme encoded in the SP6 and K1-5 genomes. T7 gp3.5 lysozyme regulates RNAP activity at the level of both initiation and termination.[39–41,50,51] In particular, termination at the CJ terminator requires lysozyme.[39–41] *In vivo*, T7 gene *3.5* is necessary for normal DNA replication, although null mutants make a burst about one-third that of wild-type, yielding pinpoint plaques at low efficiency.[52,53] Mutations that suppress this growth defect lie in gene *1*, T7 RNAP. They make the enzyme hypersensitive to inhibition by lysozyme, and they restore normal levels of DNA replication and packaging.[53,54] The major effect of the gene *1* mutation is likely to improve termination efficiency at CJ. As CJ is conserved in both the SP6 and K1-5 genomes, it is expected that their RNAP is more prone to terminate at CJ than is wild-type T7 RNAP.

Morphologically, phages in the T7 and SP6 groups are very similar; overall, their structural genes have the same genetic organization and many share sequence similarity (Table 1). However, like phage gh-1,[11] there is only one capsid protein, the programmed translational frameshift that provides the minor capsid protein T7 (and T3) gp10B does not occur in SP6 or K1-5. In addition, the T7 head protein gp6.7 and the tail protein gp7.3, both of which are thought to be essential for infectivity although their functions are not known in detail, have no obvious counterparts in SP6 or K1-5. Similarly, SP6 and K1-5 have no sequence homologue to T7 gene *13*, which is required for incorporation of gp6.7 into the phage head (P.K. & J.M., unpublished results).

Two of the three proteins that form the T7 internal core structure (gp14, gp15 and gp16) are related by sequence to SP6 and K1-5 ORF34 and ORF36. However, a comparable large size is the only common feature of ORF35 with T7 gp15. Interestingly, SP6 and K1-5 ORF35 hydrolyze peptidoglycan *in vitro*, whereas in T7 this activity is associated with gp16.[30,31] Furthermore, unlike T7 gp16, ORF35 is predicted to have lysozyme rather than lytic transglycosylase activity.

The presence of an internal core structure in SP6 and K1-5 particles, the sequence similarity of two core proteins with those of T7, and the presence of a murein hydrolase in a core protein, all suggest that SP6 and K1-5 eject the core proteins into the cell at the initiation of infection. The T7 core proteins help transport the leading end of the phage genome into the cell, and it seems likely that SP6 and K1-5 ORFs 34, 35, and 36 have a similar role.[55–57] If this is correct, the conserved location of the phage RNA polymerase gene near the leading end of the genome further suggests that, again like T7, internalization of the SP6 and K1-5 genomes into the cell is largely catalyzed by transcription.

ORF32 and ORF33 correspond to T7 gp11 and gp12, the proteins that constitute the short, stubby T7 tail. Both proteins are part of the virions (Figure 3) and the SP6 and K1-5 tail structures are similar to that of T7 (Figure 1). ORF37 is also a structural protein of the phage (Figure 3) and is weakly related in sequence to the tail fiber protein of the T7-like prophage in *P. putida* KT2440. Many phages, including T7, have tail fiber proteins with high levels of sequence similarity to the prophage protein and the majority of these proteins contain more than 550 amino acid residues. SP6 and K1-5 ORF37 proteins are only about half that size and sequence similarities to the prophage protein are confined to the N-terminal region. This is the part

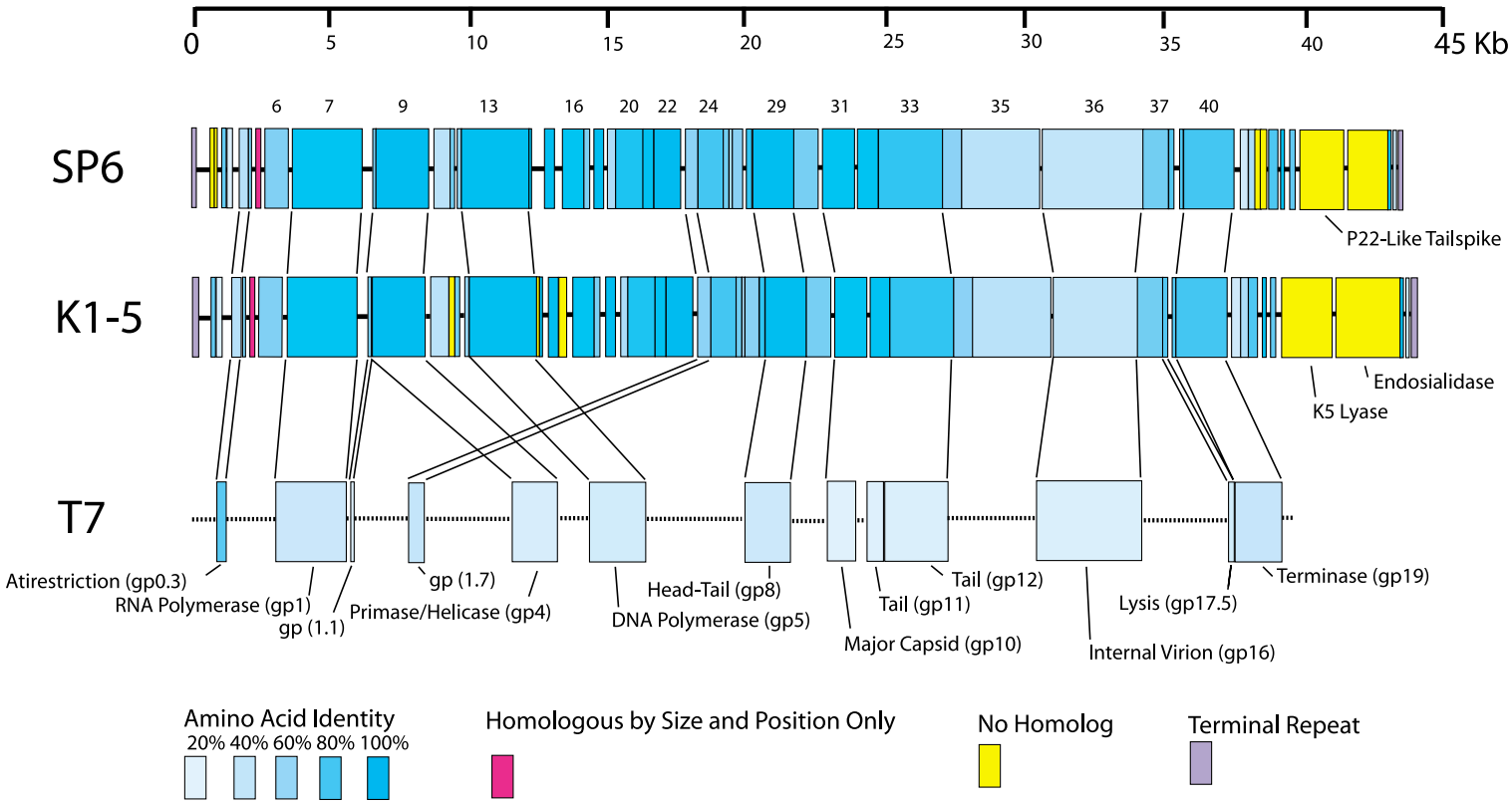

Figure 2. Genome alignment of SP6, K1-5 and T7. The percentage amino acid identities of similar ORFs are depicted in a color scale. The ORF numbers refer just to SP6 and K1-5 and not to T7.

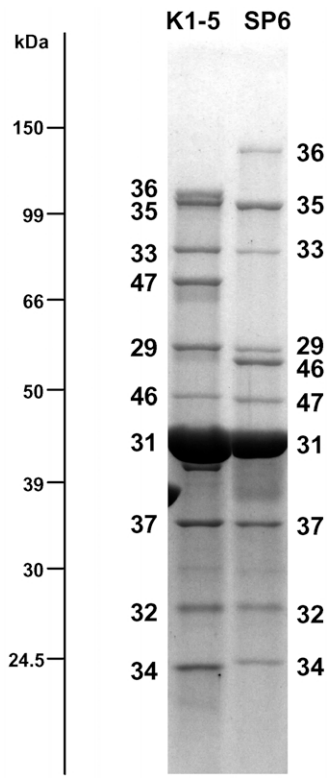

Figure 3. SDS-PAGE analysis of SP6 and K1-5 particles. Bands were excised and identified by N-terminal sequencing. Bands are indicated by ORF number, the scale (kDa) was determined using protein standards electrophoresed in adjacent lanes.

of the T3 and T7 tail fiber that attaches to the tail.[58,59] Thus, the C-terminal domain of the T3 or T7 tail fiber, which binds to the cell-surface lipopolysaccharide receptor, is missing from SP6 and K1-5 ORF37.

K1-5 binds to both the K1 and K5 capsulated strains of *E. coli*, each virion contains both the lyase and the endosialidase necessary to degrade the polysaccharide capsules.[22] Both proteins can be visualized by SDS-PAGE as virion components (Figure 3). Although the predicted size of the ORF46 lyase is 66.7 kDa (705 residues) and that of the ORF47 endosialidase is 90.4 kDa (811 residues), both proteins are thought to be processed near their C termini. By comparing sequence alignments of several endosialidases and lyases Mühlenhoff *et al*. predicted the mature form of K1-5 lyase to be 52.1 kDa (505 residues) and that of the endosialidase 78.4 kDa (705 residues).[60] Both values are consistent with the electrophoretic migration of virion proteins on SDS-PAGE (Figure 3).

SP6 has no homologues to either the endosialidase or lyase proteins but virions contain two proteins that are predicted to interact with polysaccharide. One receptor for SP6 is expected to be the *S. typhimurium* O-antigen, as the phage encodes a homologue (ORF46) of the P22 gp9 tailspike. The latter protein is an endorhamnosidase that cleaves the α1,3-*O*-glycosidic bond between the repeating tetrasaccharide units of O-antigen.[61] Interestingly, P22 adsorbs to only smooth *S. typhimurium* LT2 strains, but SP6 infects both smooth and rough LT2 strains with equal efficiency (J.M., unpublished results). Thus, SP6 must contain a receptor distinct from O-antigen. A likely second receptor-binding protein of SP6 is ORF47, which exhibits sequence similarity to a putative VirG protein of *Y. pestis* and to a fungal endopolygalacturonase. Whether ORF47 is an enzyme that can degrade part of the bacterial cell envelope is not known.

It was noted that the K1-5 (and K1E) endosialidase lacked the N-terminal domain, relative to the K1F endosialidase.[21] The missing domain is analogous to the tail-binding domain of the T7 tail fiber, raising the question of how the K1-5 enzyme is attached to the head. Similarly, the SP6 endorhamnosidase lacks the equivalent N-terminal head-binding domain of the P22 tailspike. We speculate that the enzymes are attached to the K1-5 and SP6 particles through an interaction with ORF37, which contains sequences corresponding to the tail-binding domain of the T7 tail fiber gp17 but which lacks the C-terminal lipopolysaccharide-binding domain. Interestingly, the T7 tail fiber and the K1F endosialidase are trimeric proteins, whereas the K1E endosialidase (which is almost identical with K1-5 ORF47) was purified as a hetero-oligomeric complex that contains an unidentified 38 kDa protein.[59,62–64] The size of this protein corresponds quite well with the predicted sequence and electrophoretic mobility of K1-5 and SP6 ORF37 (35.5 kDa, Figure 3).

Evolution of the SP6 group

Their analysis of the marine vibriophage VpV262 (Figure 4) genome led Hardies *et al*. to define the T7 supergroup as the phages that contain a T7-like head structure module.[13] The term supergroup thus encompasses phages that share some features of T7 and its closest relatives but lack some hallmark features of T7, for example, a phage-specific RNAP. SP6 and K1-5 are much more closely related to T7 than the marine phages of the supergroup but are clearly well-separated from the close-knit

	← TR	↓ CJ	
SP6	….GTGGGCGTGGGCT	ATCTGTT	CGTTTGCTCCGCTT
K1-5	….GAGGGCGTGGGCT	ATCTGTT	CCTTTGCTCCTCAC
T7	….TCTCTGTGTCCCT	ATCTGTT	ACAGTCTCCTAAAG
øA1122	….CTCCTTGTCCCCT	ATCTGTT	ACAGCCTCCTAAAG
T3	….TCTATGTGTCCCT	ATCTGTT	AGCCCCTTAAAGTA
øYeO3-12	….CCTATGTGTCCCT	ATCTGTT	AGCCACCCTTAAAG
K11	….TCTCTGTGTCCCT	ATCTGTT	GGTACTCATTAAGT
gh-1	….CCTATCTGTCCCT	ATCTGTA	AGTCCTCCTTGAGT

Figure 4. Alignment of sequences of various phages of the SP6 and T7 groups at the CJ pause/terminator element. TR is the right end, marked by the down arrow, of the terminal repetition in concatemeric DNA. Underlined residues in CJ are conserved among all phages. Underlined residues in the right part of the sequences refer to the conserved sequence that includes the site of termination in T7 DNA (double underline).[40]

group of the coliphages T7 and T3, the yersiniophages ϕA1122 and ϕYeO3-12, the *P. putida* T7-like prophage and its lytic phage gh-1. Aside from the anti-type I restriction protein ORF3, which in SP6 is 77% identical with T7 gp0.3, most K1-5 and SP6 ORFs exhibit less than 40% amino acid sequence identity with their T7 counterparts. The order of genes involved in DNA metabolism, lysis, and adsorption has not been conserved between the SP6 and T7 groups. However, T7 growth seems robust to significant changes in gene order,[29] and this difference between the groups may not translate into differences in growth strategy. Nevertheless, we define the SP6 group of phages (SP6, K1-5, and probably K1E) as a distinct subgroup within the T7 supergroup. DNA sequence identity between the SP6 and K1-5 genomes, as determined by the GCG program GAP, is about 67%. The genetic organization of the two phages is almost identical, and even for those genes where sequence similarity cannot be discerned, most of the predicted ORFs are comparable in size. In almost all cases, computer searches for sequence similarity using a K1-5 ORF give the best hit to the counterpart SP6 protein, and *vice versa*. Excluding the tail enzymes, which confer host-cell specificity and are not related by sequence, the level of amino acid identity between proteins of known function extends from a high of 92% for ORF9 primase to a low of 40% for ORF36.

SP6 and K1-5 have diverged somewhat more within their subgroup compared to the members of the T7 subgroup. The longest stretches of nucleotide identity between genomes are 74 bp, within ORF13, coding for the DNAP, and 73 bp, most of which specifies a putative transcription terminator distal to ORF18. There are also two 62 bp identities: one lies within *orf9*, primase, the second is relatively A + T-rich and extends from the 3′end of ORF14 into a non-coding, unspecified region. No other identity greater than 50 bp is present in the SP6 and K1-5 genomes. For comparison, the coliphages T7 and T3 share a common 177 bp stretch and contain five regions of >100 nucleotide identity.[2] The divergence of SP6 and K1-5 undoubtedly has been accentuated by their utilizing different hosts, which themselves are thought to have separated about 100 million years ago.[65] The distributions of *Sty*LT I sites in the two phage genomes testify to the selection imposed to remove those sites in the SP6 genome. Nevertheless, it is hard to imagine how an ancestor of SP6, which may have been susceptible to *Sty*LT I restriction as K1-5 DNA is today, could have successively lost nearly 100 sites. Perhaps the restriction system was once part of a mobile element that swept only slowly through the *S. typhimurium* serovar, or perhaps the usual host for SP6 was non-restricting and it encountered the restriction system only occasionally.

The predicted promoter sequences of the SP6 and K1-5 RNAP are different, which must reflect differences in the phage RNAP. Dietz *et al*. also found two different promoter specificities associated with four capsule-specific coliphages.[66] These phages were not characterized by sequence analysis but two of them hybridized weakly to T7 DNA, and may thus be related to K1-5. It is not obvious how promoter specificities change, although a single amino acid change is sufficient to switch T7 RNAP from T7 to T3 promoter specificity.[36] A single nucleotide change within the promoter is also sufficient to allow T7 RNAP to transcribe from a T3 promoter and *vice versa*,[37] but the presence of many promoters scattered across the genome makes it difficult to imagine how different RNAP—promoter combinations evolve. The minimum number of phage promoters necessary for good (albeit not optimal) growth is not known but it is clear that both T7 and T3 deletion mutants lacking some class II promoters grow well in the laboratory and would probably grow well in many environments if the missing gene products were available. As another example, wild-type T3 contains a T7-specific gene *17* promoter that is not utilized during infection.[67] Thus coinfection by different T7-group or SP6-group phages could yield a recombinant with a number of promoters that are not immediately recognized by the RNAP. Such a recombinant would likely evolve either to "correct" each promoter separately or to expand the promoter specificity of the RNAP. Continued evolution could then lead to all promoters becoming of the same type, followed by a regain of single promoter specificity by the RNAP. However, coinfections by T7-like phages whose RNAPs exhibit different promoter specificities may result in mutual exclusion.[18] Nevertheless, rare viable recombinants (e.g. T7 × T3, T7 × ϕYeO3-12) have been isolated in the laboratory, and they may be formed in the environment.[2,42–44] Viable recombinants between two phages may even be obtained in circumstances where neither parent can grow. For example, T7 grows very poorly, if at all, on most K1 strains. However, by adding endosialidase to degrade the capsule and expose the cell surface, T7 can adsorb and replicate (D.S., unpublished results). Thus, the only barrier to T7 infection on these cells is the capsule. However, if the source of endosialidase is the phage K1-5, a capsulated strain may be infected almost simultaneously by both K1-5 and T7, and recombination between the genomes could then occur.

The T7 group consists of obligate lytic phages that, because of their very close similarity, have been thought to evolve primarily by vertical descent and only secondarily by recombination following almost simultaneous infection of a cell by two related phages. However, in several instances database searches using SP6 or K1-5 ORFs as query sequences revealed a related protein from a prophage in *P. putida* KT2440.[68] This prophage contains an integrase function but also contains homologues to almost all T7 genes (the host transcription inhibitors gp0.7 and gp2, and the gp17.5

holin being notable exceptions). No repressor that could prevent phage gene expression is obvious but host RNAP promoters that could transcribe the phage RNAP are also missing, and thus in the prophage state there is not likely to be substantial phage gene expression. Nevertheless, it would be surprising if all the prophage genes have retained function, since many T7 genes are toxic to *E. coli* when cloned. Nevertheless, the very presence of an intact T7-like prophage in a host chromosome should provide a superinfecting lytic phage more opportunities for diversification by recombination than by coinfection.

With the exceptions that there are no homologues to T7 genes *13* and *18.5/18.7*, it is notable that the SP6 and K1-5 structural and morphogenetic genes are colinear with those of the T7 group between the head-tail connector (ORF29, T7 gene *8*) through the genes for the two terminase subunits (ORF39 and ORF40, T7 genes *18* and *19*). The latter are the last essential genes of T7 to be expressed. However, SP6 and K1-5 have additional tail enzyme "modules" which, relative to T7, appear to have been inserted near the right genome end that allow phage growth on strains that produce extracellular polysaccharide. The modules include K1-5 endosialidase and lyase, and SP6 tailspike endorhamnosidase and the putative endopolygalacturonase. The genes for these proteins form a variant of "morons", first defined in the lambdoid phages HK97 and HK022.[69] The SP6 and K1-5 genes are preceded by a phage, rather than a host, promoter and are followed by regions of predicted RNA secondary structure that could serve as transcriptional terminators and/or RNase III recognition sites. The modular cassette structure of the tail enzymes elicits the suggestion that these phages readily change host specificity by acquiring new enzyme activities, probably by illegitimate recombination. In addition to the major tail enzymes, there are several small ORFs located between the terminase and the SP6 endorhamnosidase or K1-5 lyase genes, all of which exhibit weak amino acid sequence similarities to various bacterial enzymes involved in biosynthesis or turnover of cell envelope macromolecules. These small genes are likely remnants of past recombination events among ancestral phages that had different host specificities. In accord with this idea, there is significant sequence similarity between ORF41, 42, and 43. SP6 ORF41 and 42 share 31% identity over their entire length, whereas ORF42 and ORF43 share 42% identity over about half their length.

The diversity of cell-surface polysaccharide in enterobacteria is immense. There are 167 O-serotypes and 80 polysaccharide K antigens known for *E. coli* alone.[70] While some of these capsules are associated with pathogenicity, for example in colonizing the gut, they can serve as barriers to phages. Resistance to phage infection is likely to be the primary driving force for capsule structure diversification. Capsule-specific phages have evolved to overcome these outermost layers of the cell; in turn, cells have diversified to present chemically different polysaccharide structures. The "arms race" between bacteria and phages could provide selective pressure for the evolution of new polysaccharide capsules and potentially the emergence of new pathogens.

Similarly, over 1000 distinct *Salmonella* O-antigen variants have been described.[71] Phage P22 recognizes only three of these but the majority of the *Salmonella* typing phages are related to P22, differing mainly in adsorption spectrum, immunity, and susceptibility to superinfection exclusion systems.[72,73] SP6 has a broader host range than these P22-like phages as it grows also on rough strains of LT2. Whether SP6 grows on smooth *S. typhimurium* strains that are resistant to P22 by O-antigen variation has not been tested.

Additional comments

Polysaccharide-degrading enzymes carried on phage virions have been well studied.[21–23,61,64,74–85] K1-5 was the first phage described where two distinct activities have been found on a single virion.[22] Presumably these enzyme activities originated from bacterial hosts, although there is only limited nucleotide similarity between the phage K5 lyase and the bacterial chromosome-encoded lyase *elmA*.[86] Phage P22, and presumably SP6, degrade the smooth portion of the lipopolysaccharide layer of *S. typhimurium* LT2, and coliphages similar to P22 that infect smooth *E. coli* strains have been isolated.[87] However, most laboratory strains of *E. coli* are rough and, consequently, most attention has been paid to rough-specific coliphages. This quirk of history in phage biology may have led to an under-appreciation of the role of virion-associated, polysaccharide-degrading enzymes associated with the degradation of polysaccharide capsules.

Three enzyme activities have now been found as part of K1-5 virions. In addition to the O-antigen or K-antigen degrading activities, both SP6 and K1-5 virions contain an enzyme that degrades peptidoglycan. This enzyme activity is common to most dsDNA phage virions.[31] It is of interest that by sequence analysis SP6 and K1-5 virions contain a lysozyme activity, whereas members of the T7 group contain a lytic transglycosylase. Furthermore, the murein hydrolase activity in SP6 and K1-5 is part of ORF35, the positional counterpart to T7 gp15, whereas it is part of gp16 in the T7 group. Therefore, the acquisition of murein hydrolase activity must have occurred after the two phage groups had separated. The ability to degrade peptidoglycan is not essential for phages that infect Gram-negative hosts but its presence extends the conditions under which a successful infection can occur.[30,31,88] The fusion of distinct murein hydrolase activities to different virion structural components in the T7 and SP6 groups argues that the enzyme was not present in a common ancestral phage and was acquired by

both modern groups under a selection for an increased efficiency of infection.

Materials and Methods

SP6 was purified by CsCl density-gradient centrifugation and genomic DNA was isolated by extraction with phenol. Sequence data were obtained by a combination of primer walking from known SP6 sequences using genomic DNA as a template and by shotgun cloning random fragments. The genome ends were determined by a combination of sequencing to the genome end, cloning and sequencing the terminal repeat regions, and by ligating the genome to linearized, blunt-ended pUC19. Ligating the genome to itself to form circular molecules or concatemers was unsuccessful, as judged by PCR and DNA sequencing across the repeat region using appropriate primers. Furthermore, obtaining clones of the terminal repeats was far more difficult than anticipated. The left genome end, in particular, exhibited sequence variability between the different methods and independent repetitions of the same method on different phage stocks. While the sequence presented is thought to represent the majority, it is possible that a percentage of genomes within a population have different genome ends.

K1-5 DNA was also isolated from phage purified by CsCl density-gradient centrifugation and extraction with phenol. Genomic DNA was used directly as a template for sequencing by primer walking using known sequences as starting points. Both strands were sequenced completely with sufficient overlap to have threefold confirmation. The ends were confirmed by ligating phage DNA to mp19 RF DNA cut with *Sma*I and by performing PCR on the ligation mix using primers specific to mp19 and to sequences near to either end of the phage DNA. The PCR product was then sequenced to locate the exact ends of the genome.

Annotation for both phages were done using a combination of BLAST[89] searches and visual inspection. Database searches were performed on individual ORFs (BLASTP) as well as the entire genome (BLASTX and BLASTN). Amino acid sequence alignments were performed using GCG-Lite.[90]

Sequences and annotation

The Genbank accession number for the SP6 genome is AY370673 and that for the K1-5 genome is AY370674. During the course of these studies we discovered that the SP6 genome sequence had been determined by another group[91] (Genbank Accession number AY288927). The two sequences agree completely, a result that provides confidence in the accuracy of the two approaches to DNA sequencing. The annotation of the SP6 genome determined as part of this work has been coordinated with that of Dobbins *et al.*,[91] it differs mainly in the predicted start sites for ORFs that have no defined function.

Acknowledgements

This work was supported, in part, by United States Public Health Service grants AI-06045 (to C.C.R.) and GM32095 (to I.J.M.). We thank Kunio Nagashima for assistance with electron microscopy and Klause Linse for assistance with protein sequencing.

References

1. Dunn, J. J. & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477–535.
2. Pajunen, M. I., Elizondo, M. R., Skurnik, M., Kieleczawa, J. & Molineux, I. J. (2002). Complete nucleotide sequence and likely recombinatorial origin of bacteriophage T3. *J. Mol. Biol.* **319**, 1115–1132.
3. Pajunen, M. I., Kiljunen, S. J., Söderholm, M. E-L. & Skurnik, M. (2001). Complete genomic sequence of the lytic bacteriophage ϕYeO3-12 of *Yersinia enterocolitica* serotype O:3. *J. Bacteriol.* **183**, 1928–1937.
4. Garcia, E., Elliott, J. M., Ramanculov, E., Chain, P. S. G., Chu, M. C. & Molineux, I. J. (2003). The genome sequence of *Yersinia pestis* bacteriophage ϕA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J. Bacteriol.* **185**, 5248–5262.
5. Hausmann, R. (1988). The T7 group. In *Bacteriophages* (Calendar, R., ed.), vol. 1, pp. 259–289, Plenum Press, New York.
6. Studier, F. W. (1979). Relationships among different strains of T7 and among T7-related bacteriophages. *Virology*, **95**, 70–84.
7. Hyman, R. W., Brunovskis, I. & Summers, W. C. (1973). DNA base sequence homology between coliphages T7 and ϕII and between T3 and ϕII as determined by heteroduplex mapping in the electron microscope. *J. Mol. Biol.* **77**, 189–196.
8. Hyman, R. W., Brunovskis, I. & Summers, W. C. (1974). A biochemical comparison of the related bacteriophages T7, ϕI, ϕII, W31, H, and T3. *Virology*, **57**, 189–206.
9. Brunovskis, I., Hyman, R. W. & Summers, W. C. (1973). *Pasturella pestis* bacteriophage H and *Escherichia coli* bacteriophage ϕII are nearly identical. *J. Virol.* **11**, 306–313.
10. Korsten, K. H., Tomkiewicz, C. & Hausmann, R. (1979). The strategy of infection as a criterion for phylogenetic relationships of non-coli phages morphologically similar to phage T7. *J. Gen. Virol.* **43**, 57–73.
11. Kovalayova, I. V. & Kropinski, A. M. (2003). The complete genomic sequence of lytic bacteriophage gh-1 infecting pseudomanas putida–evidence for close relationship to the T7 group. *Virology*, **311**, 305–315.
12. Weinel, C., Nelson, K. E. & Tummler, B. (2002). Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ. Microbiol.* **4**(12), 809–818.
13. Hardies, S. C., Comeau, A. M., Serwer, P. & Suttle, C. A. (2003). The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. *Virology*, **310**, 359–371.
14. Lavigne, R., Bourkaltseva, M. V., Robben, J., Sykilinda, N. N., Kurochkina, L. P., Grymonprez, B. *et al.* (2003). The genome of bacteriophage ϕKMV: a T7-like lytic phage infecting *Pseudomonas aeruginosa*. *Virology*, **312**, 49–59.
15. Rohwer, F., Segall, A., Steward, G., Seguritan, V., Breitbart, M., Wolven, F. & Azam, F. (2000). The

complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. *Limnol. Oceanogr.* **45**, 408–418.
16. Feng, C. & Lu, J. (2002). Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl. Environ. Microbiol.* **68**, 2589–2594.
17. Zinder, N. D. (1961). A bacteriophage specific for F^- *Salmonella* strains. *Science*, **133**, 2069–2070.
18. Molineux, I. J. (1999). The T7 family of bacteriophages. In *Encyclopedia of Molecular Biology* (Creighton, T. E., ed.), pp. 2495–2507, Wiley, New York.
19. Kotani, H., Ishizaki, Y., Hiraoka, N. & Obayashi, A. (1987). Nucleotide sequence and expression of the cloned gene of bacteriophage SP6 RNA polymerase. *Nucl. Acids Res.* **15**, 2653–2664.
20. Tseng, T. Y., Frick, D. N. & Richardson, C. C. (2000). Characterization of a novel DNA primase from the *Salmonella typhimurium* bacteriophage SP6. *Biochemistry*, **39**, 1643–1653.
21. Scholl, D., Adhya, S. & Merril, C. (2002). Bacteriophage SP6 is closely related to phages K1-5, K5, and K1E but encodes a tail protein very similar to that of the distantly related P22. *J. Bacteriol.* **184**, 2833–2836.
22. Scholl, D., Rogers, S., Adhya, S. & Merril, C. (2001). Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. *J. Virol.* **75**, 2509–2515.
23. Long, G. S., Bryant, J. M., Taylor, P. W. & Luzio, J. P. (1995). Complete nucleotide sequence of the gene encoding bacteriophage E endosialidase: implications for K1E endosialidase structure and function. *Biochem. J.* **309**, 543–550.
24. Meisel, A., Bickle, T. A., Kruger, D. H. & Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, **355**, 467–469.
25. Dryden, D. T. F., Murray, N. E. & Rao, D. N. (2001). Nucleoside triphosphate-dependent restriction enzymes. *Nucl. Acids Res.* **29**, 3728–3741.
26. Studier, F. W. & Movva, N. R. (1976). SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.* **19**, 136–145.
27. Butler, E. T. & Chamberlin, M. J. (1982). Bacteriophage SP6-specific RNA polymerase. *J. Biol. Chem.* **257**, 5772–5778.
28. Kassavetis, G. A., Butler, E. T., Roulland, D. & Chamberlin, M. J. (1982). Bacteriophage SP6-specific RNA polymerase. *J. Biol. Chem.* **257**, 5779–5788.
29. Endy, D., You, L., Yin, J. & Molineux, I. J. (2000). Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl Acad. Sci. USA*, **97**, 5375–5380.
30. Moak, M. & Molineux, I. J. (2000). Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Mol. Microbiol.* **37**, 345–355.
31. Moak, M. & Molineux, I. J. (2003). Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol. Microbiol.* In the Press.
32. Wang, I., Smith, D. L. & Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* **54**, 799–825.
33. Brown, J. E., Klement, J. F. & McAllister, W. T. (1986). Sequences of three promoters for the bacteriophage SP6 RNA polymerase. *Nucl. Acids Res.* **14**, 3521–3526.
34. Shin, I., Kim, J., Cantor, C. R. & Kang, C. (2000). Effects of saturation mutagenesis of the phage SP6 promoter on transcription activity, presented by activity logos. *Proc. Natl Acad. Sci. USA*, **97**, 3890–3895.
35. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, **39**, 10419–10430.
36. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. (1992). Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228**, 506–515.
37. Klement, J. F., Moorefield, M. B., Jorgensen, E., Brown, J. E., Risman, S. & McAllister, W. T. (1990). Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. *J. Mol. Biol.* **215**, 21–29.
38. Lee, S. S. & Kang, C. (1993). Two base pairs at -9 and -8 distinguish between the bacteriophage T7 and SP6 promoters. *J. Biol. Chem.* **268**, 19299–19304.
39. Zhang, X. & Studier, F. W. (1997). Mechanism of inhibition of T7 RNA polymerase by T7 lysozyme. *J. Mol. Biol.* **269**, 964–981.
40. Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J. & McAllister, W. T. (1997). Mutant bacteriophage T7 RNA polymerases with altered termination properties. *J. Mol. Biol.* **269**, 28–40.
41. Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J. & McAllister, W. T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **280**, 201–213.
42. Beier, H., Golomb, M. & Chamberlin, M. (1977). Isolation of recombinants between T7 and T3 bacteriophages and their use in *in vitro* transcriptional mapping. *J. Virol.* **21**, 753–756.
43. Beier, H. & Hausmann, R. (1974). T7 × T3 phage crosses leading to recombinant RNA polymerases. *Nature (London)*, **251**, 538–540.
44. Molineux, I. J., Mooney, P. Q. & Spence, J. L. (1983). Recombinants between bacteriophages T7 and T3 which productively infect F-plasmid-containing strains of *Escherichia coli*. *J. Virol.* **46**, 881–894.
45. Studier, F. W. & Dunn, J. J. (1982). Organization and expression of bacteriophage T7 DNA. *Cold Spring Harbor Symp. Quant. Biol.* **42**, 999–1007.
46. Huber, H. E., Beauchamp, B. & Richardson, C. C. (1988). *Escherichia coli* dGTP triphosphohydrolase is inhibited by gene *1.2* protein of bacteriophage T7. *J. Biol. Chem.* **263**, 13549–13556.
47. Molineux, I. J. & Spence, J. L. (1984). Virus-plasmid interactions: mutants of bacteriophage T3 that abortively infect plasmid F-containing (F^+) strains of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **81**, 1465–1469.
48. Molineux, I. J. (1991). Host–parasite interactions: recent developments in the genetics of abortive phage infections. *New Biol.* **3**, 230–236.
49. De Wyngaert, M. A. & Hinkle, D. C. (1980). Characterization of the defects in bacteriophage T7 DNA synthesis during growth in the *Escherichia coli* mutant *tsnB*. *J. Virol.* **33**, 780–788.
50. Kumar, A. & Patel, S. S. (1997). Inhibition of T7 RNA polymerase transcription initiation and transition from initiation to elongation are inhibited by T7

lysozyme *via* a ternary complex with RNA polymerase and promoter DNA. *Biochemistry*, **36**, 13954–13962.
51. Huang, J., Villemain, J., Padilla, R. & Sousa, R. (1999). Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription. *J. Mol. Biol.* **293**, 457–475.
52. Silberstein, S., Inouye, M. & Studier, F. W. (1975). Studies on the role of bacteriophage T7 lysozyme during phage infection. *J. Mol. Biol.* **96**, 1–11.
53. Zhang, X. (1995). T7 RNA polymerase and T7 lysozyme: genetic, biochemical and structural analysis of their interaction and multiple roles in T7 infection. PhD dissertation SUNY, Stonybrook, New York.
54. Moffatt, B. A. & Studier, F. W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, **49**, 221–227.
55. García, L. R. & Molineux, I. J. (1996). Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. *J. Bacteriol.* **178**, 6921–6929.
56. Struthers-Schlinke, J. S., Robins, W. P., Kemp, P. & Molineux, I. J. (2000). The internal head protein gp16 of bacteriophage T7 controls DNA ejection from the virion. *J. Mol. Biol.* **301**, 35–45.
57. Molineux, I. J. (2001). No syringes please, ejection of T7 DNA from the virion is enzyme-driven. *Mol. Microbiol.* **40**, 1–8.
58. Kato, H., Fujisawa, H. & Minagawa, T. (1986). Subunit arrangement of the tail fiber of bacteriophage T3. *Virology*, **153**, 80–86.
59. Steven, A. C., Trus, B. L., Maizel, J. V., Unser, M., Parry, D. A., Wall, J. S. *et al.* (1988). Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. *J. Mol. Biol.* **200**, 351–365.
60. Mühlenhoff, M., Stummeyer, K., Grove, M., Seuerborn, M. & Gerardy-Schahn, R. (2003). Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. *J. Biol. Chem.* **278**, 12634–12644.
61. Iwashita, S. & Kanegasaki, S. (1976). Enzymic and molecular properties of base-plate parts of bacteriophage P22. *Eur. J. Biochem.* **65**, 87–94.
62. Hallenbeck, P. C., Vimr, E. R., Yu, F., Bassler, B. & Troy, F. A. (1987). Purification and properties of a bacteriophage-induced *endo-N*-acetylneuraminidase specific for poly-alpha-2,8-sialosyl carbohydrate units. *J. Biol. Chem.* **262**, 3553–3561.
63. Tomlinson, S. & Taylor, P. W. (1985). Neuraminidase associated with coliphage E that specifically depolymerizes the *Escherichia coli* K1 capsular polysaccharide. *J. Virol.* **55**, 374–378.
64. Gerardy-Schahn, R., Bethe, A., Brennecke, T., Muhlenhoff, M., Eckhardt, M., Zeising, S. *et al.* (1995). Molecular cloning and functional expression of bacteriophage PK1E-encoded endoneuraminidase Endo NE. *Mol. Microbiol.* **16**, 441–450.
65. Ochman, H., Elwyn, S. & Moran, N. A. (1999). Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA*, **96**, 12638–12643.
66. Dietz, A., Andrejauskas, E., Messerschmid, M. & Hausmann, R. (1986). Two groups of capsule-specific coliphages coding for RNA polymerases with new promoter specificities. *J. Gen. Virol.* **67**, 831–838.
67. Rosa, M. D. & Andrews, N. C. (1981). Phage T3 contains an exact copy of the 23 base-pair phage T7 RNA polymerase promoter sequence. *J. Mol. Biol.* **147**, 41–53.
68. Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A. *et al.* (2002). Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799–808.
69. Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F. & Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *J. Mol. Biol.* **299**, 27–51.
70. Whitfield, C. & Roberts, I. S. (1999). Structure, assembly, and regulation of expression of capsules in *E. coli*. *Mol. Microbiol.* **31**, 1307–1319.
71. Orskov, F. & Orskov, I. (1983). Summary of a workshop on the clone concept in the Epidemiology, taxonomy, and evolution of the *Enterobacteriaceae* and other bacteria. *J. Infect. Dis.* **148**, 346–357.
72. Schicklmaier, P., Wieland, T. & Schmieger, H. (1999). Molecualr characterization and module composition of P22-related *Salmonella* phage genomes. *J. Biotechnol.* **73**, 185–194.
73. Schmieger, H. (1999). Molecular survey of the *Salmonella* phage typing scheme of Anderson. *J. Bacteriol.* **181**, 1630–1635.
74. Altmann, F., Kwiatkowski, B. & Stirm, S. (1986). A bacteriophage-associated glycanase cleaving β-pyranosidic linkages of 3-deoxy-D-mannose-2-octulosonic acid. *Biochem. Biophys. Res. Commun.* **136**, 329–335.
75. Bayer, M. E., Thurow, H. & Bayer, M. H. (1979). Penetration of the polysaccharide capsule of *Escherichia coli* (Bi161/42) by bacteriophage K29. *Virology*, **94**, 95–118.
76. Bayer, M. E., Takeda, K. & Uetake, H. (1980). Effects of receptor destruction by *Salmonella* bacteriophages (ϵ15 and c341. *Virology*, **105**, 328–337.
77. Bessler, W., Fehmel, F., Freund-Mölbert, E., Knüferman, H. & Stirm, S. (1975). *Escherichia coli* capsule bacteriophages. IVFree capsule depolymerase 29. *J. Virol.* **15**, 976–984.
78. Gupta, D. S., Jann, B. & Jann, K. (1983). Enzymatic degradation of the capsular K5-antigen of *E. coli* by coliphage K5. *FEMS Microbiol. Letters*, **16**, 13–17.
79. Hynes, W., Hancock, L. & Fischetti, V. (1995). Analysis of a second bacteriophage hyaluronidase gene from *Streptococcus pyogenes*: evidence for a third hyaluronidase involved in extracellular enzymatic activity. *Infect. Immun.* **63**, 3015–3020.
80. Kwiatkowski, B., Boschek, B., Thiele, H. & Stirm, S. (1982). Endo-*N*-acetylneuraminidase associated with bacteriophage particles. *J. Virol.* **43**, 697–704.
81. Kwiatkowski, B., Boschek, B., Thiele, H. & Stirm, S. (1983). Substrate specificity of two bacteriophage-associated endo-*N*-acetylneuraminidases. *J. Virol.* **45**, 367–374.
82. Machida, Y., Miyake, K., Hattori, K., Yamamoto, S., Kawase, M. & Iijima, S. (2000). Structure and function of a novel coliphage-associated sialidase. *FEMS Microbiol. Letters*, **182**, 333–337.
83. Nimmich, W. (1997). Degradation studies on *Escherichia coli* capsular polysaccharides. *FEMS Microbiol. Letters*, **153**, 105–110.
84. Petter, J. G. & Vimr, E. R. (1993). Complete nucleotide sequence of the bacteriophage K1F tail gene encoding endo-*N*-acylneuraminidase (endo-*N*) and comparison to an endo-*N* homolog in bacteriophage PK1E. *J. Bacteriol.* **175**, 4354–4363.
85. Rieger-Hüg, D. & Stirm, S. (1981). Comparative study of host capsule depolymerases associated with *Klebsiella* bacteriophages. *Virology*, **113**, 363–378.
86. Clarke, B. R., Esumeh, F. & Roberts, I. S. (2000). Cloning, expression, and purification of the K5 capsular polysaccharide lyase (KflA) from coliphage K5:

evidence for two distinct K5 lyase enzymes. *J. Bacteriol.* **182**, 3761–3766.
87. Dhillon, T. S., Poon, A. P. W., Chan, D. & Clark, A. J. (1998). General transducing phages like *Salmonella* phage P22 isolated using a smooth strain of *Escherichia coli* as host. *FEMS Microbiol. Letters*, **161**, 129–133.
88. Rydman, P. S. & Bamford, D. H. (2000). Bacteriophage PRD1 DNA entry uses a viral membrane-associated transglycosylase activity. *Mol. Microbiol.* **37**, 356–363.
89. Atschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
90. Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.
91. Dobbins, A. T., George, M., Jr, Basham, D. J., Ford, M. E., Houtz, J. M., Pedulla, M. L. *et al.* (2004). Complete genomic sequence of the virulent *Salmonella* bacteriophage SP6. *J. Bacteriol.* In the Press.

Edited by M. Gottesman

(Received 9 September 2003; received in revised form 13 November 2003; accepted 18 November 2003)