The structure of bacteriophage T5 injection stop signal.

John Davison

INRA de Versailles, Route de St Cyr, 78026 Versailles, France

email : jrndavison@gmail.com

Keywords:

Phage injection, phage genomes, phage evolution, DNA degradation, RecBCD nuclease, DnaA protein, injection stop signal, DNA protection, DNA packaging, terminase.

Abstract

Escherichia coli bacteriophage T5 differs from most phages in that it injects its genome in two steps: First Step Transfer, FST, corresponding to leftmost 7.9% of the genome and Second Step Transfer, SST, corresponding to the remainder of the genome. Expression of genes A1 and A2 is required for SST. DNA injection stops at a site known as the injection stop signal (iss) which is a cis acting site located in the untranslated region of the Left Terminal Repeat (LTR). The *iss* region is extremely complicated with many repeats, inverted repeats and palindromes. This report compares the *iss* regions of 21 T5 related phages and shows that all have a common conserved structure including a series of 8 DnaA boxes arranged in a highly specific manner; reminiscent of the origin of replication (*oriC*) of the host. DnaA protein, which binds DnaA boxes is a mostly membrane bound, leading to the suggestion that injection stops at *iss* due to binding to DnaA protein at the membrane. A new model of the mechanism of T5 injection stop and SST injection re-start is suggested.

1. Introduction

Bacteriophage T5 was one of the original seven “T-phages” isolated by Max Delbrück and his collaborators in the 1940s. The molecular biology of bacteriophage T5 has been reviewed (1) and other reviews specifically covered the First Step Transfer (FST) region of phage T5 (2, 3). T5 has a linear genome of 121,752 bp with large terminal repeats (Left Terminal Repeat, LTR and Right Terminal Repeat, RTR) each comprising 10,139 bp (4). T5 differs from most phages in that it injects its DNA in two steps; firstly, the First Step Transfer (FST) which is the left-most 7.9% and then, following pre-early gene expression, the remainder of the genome (Second Step Transfer, SST). The expression of two genes, *A1* and *A2*, is necessary for the SST DNA injection. The FST region contains the injection stop signal (*iss*) which is the locus at which DNA injection stops (2, 5). The FST DNA can be separated from the SST DNA by mechanical shearing of the T5/bacterium complex, usually in the presence of cloramphenicol to prevent protein synthesis (2). It was hypothesized that *iss* could be due to a special nucleotide sequence on the DNA. The position of the region containing *iss* was calculated and then cloned and sequenced. The *iss* locus lies within the left end of a 1662 bp non-coding region which contains 97 stop codons and extends from 9194 to 10856. The *iss* region is rich in a bewildering array direct repeats, inverted repeats and palindromes, (4, 5) and these were shown to be conserved, and in the same order, in other T5-like phages (3).

The genetic characterization of *iss* is difficult since it is an essential *cis*-acting site. The phenotype of mutants in this region would be expected to be either non-conditional lethal or wild-type; depending on whether or not an essential region was affected. Thus, in both cases, mutants could not be phenotypically identified and for 50 years there seemed no way to resolve this problem. However, recent interest in bacteriophages as therapeutic agents and food preservation agents has resulted in the isolation and nucleotide sequencing of many new T5-like phages, from different environmental sources, and capable of infecting a variety of bacterial hosts. These T5-like phages have widely diverged as visualized by the extensive deletions and substitutions and mutations within their genomes.

Since *iss* is an essential *cis*-acting site, it is likely to have been preserved in evolution, even between widely divergent T5-like genomes. This review investigates the similarity of the sequences in the *iss* region among 21 T5-like phages, particularly with regard to the conservation of the direct repeats, inverted repeats and palindromes that form part of the *iss* region.

All T5-like phages conserve the *A1* and *A2* genes that are responsible for SST following the injection pause at the *iss* region. It is tempting to suppose (4) that A1 and A2 proteins interact, directly or indirectly, with *iss* to relieve the injection stop and thus allow second step transfer. However, there is no evidence for this and this assumption will be questioned in the discussion section.

This article further documents the similarities and differences between the FST coding regions of the T5-like phages, with the object of defining those genes essential for the multiple phenotypes encoded by the pre-early genes.

The author is now retired from laboratory activity and the reported data are thus entirely *in silico.* The purpose of this article is to report these new findings and to make to suggestions and hypotheses that can be experimentally tested by others.

2. Materials and Methods

2.1. Nucleotide Sequence Analysis

Nucleotide sequence analysis was performed using classical methods such as NCBI Blast analysis, EMBOSS, Galaxy Fuzznuc, and EBI Clustal.

2.2. Nomenclature and the T5 genetics

Nomenclature is important in understanding the genetic map of T5 shown in Figure. 1. The left terminal repeat (LTR) of 10,139 bp is identical in nucleotide sequence to the right terminal repeat (RTR). T5 DNA is transferred in two steps: the First Step Transfer (FST) and the Second Step Transfer (SST). The FST region is contained between the left terminus of the molecule and the point of shear of the phage/bacterium complex in a blender and is estimated about 7.9% of the genome (roughly 9618 bp). The *iss* sequence prevents SST transfer; presumably by an *iss* DNA/bacterium interaction. The second step transfer (SST) region comprises the remainder of the unique part of the genome, as well as the RTR and a small part of the right end of the LTR (between *iss* and the right end of the LTR).

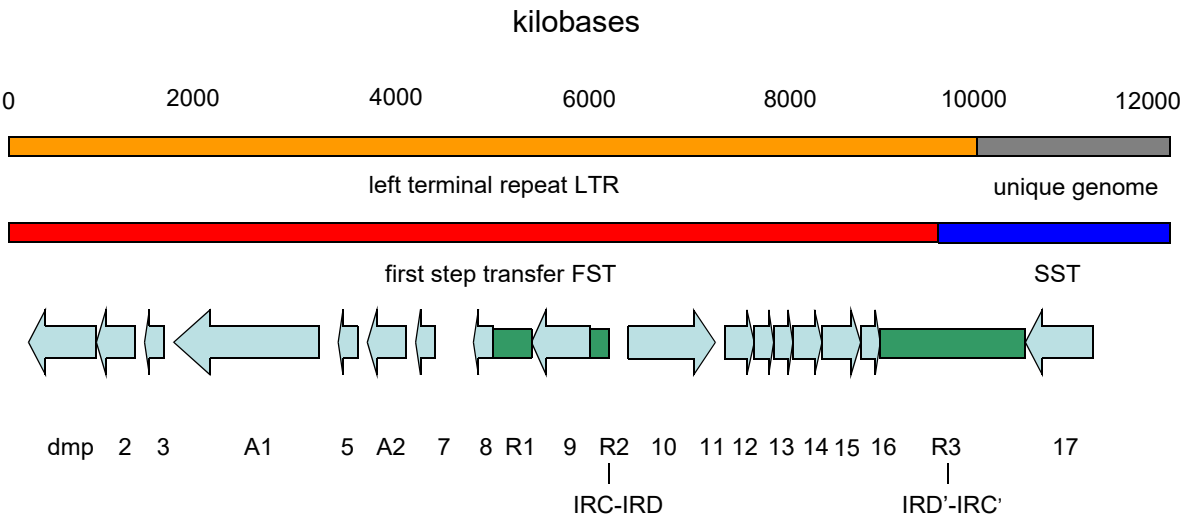

Figure 1. Genetic map of the LTR of T5. Only the left 12,000 bp of the 121,750 bp phage T5 genome are shown. The top line shows the position of the LTR (orange). The line 2 shows the FST (red) and the left end of the SST (dark blue). Line 3 shows the positions of the ORFs (pale blue) corresponding to the Paris-Orsay nucleotide sequence (AY692264.1). Line 3 also shows the positions of the repeat regions R1, R2 and R3 (green). R2 and R3 respectively contain the inverted repeats IRC-IRD and IRD'-IRC' as indicated. Repeat regions R1 and R2 are separated by the insertion of a HNH endonuclease gene (ORF 9). This insertion is not present in most T5-like phages.

2.3. T5-like phages sequences used in this study

The collection of 21 T5-like phages sequences used in this study is shown in Table 1. All of these phages carry the *dmp*, *A1* and *A2* genes, in this order, in the left end of the LTR. They all also contain the large untranslated region containing the *iss* locus in the right end of the LTR; discussed in detail below (4). Some true T5-like phage isolates are nearly identical, and, in these cases, only

one example of each was chosen for analysis. Thus, chee24 is very similar to pork27, pork2, saus47N, saus111K, poul124; while chee130_1 is almost identical to saus132, poul149, chee158, cott162 and saus176N. Similarly, DT57C is almost identical to DT571/2 in the LTR region.

Genbank contains many more sequences that have homology to T5 but only those in Table 1 were retained as being true T5-like phages. Others seem to be the product of sequencing errors and/or sequence assembly errors; and have (e.g.) the *dmp*, *A1*, and *A2* genes in the late region of the genome or have late genes in the LTR. This seems due to the fact that sequence assembly software has difficulty in processing the presence of the large terminal repeats.

2.4. The nucleotide sequence of the first step transfer region

There are differences between the 3 published T5 nucleotide sequences (Hangzhou AY587007.1 Paris-Orsay AY692264.1, and Moscow AY543070.1) and these were analyzed in detail by Davison (3). These differences make it difficult to interpret the data for the identification and nomenclature of the ORFs in the FST region. In particular, the Hangzhou sequence is missing 75 bp, starting at coordinate 5599, compared to Paris-Orsay and Moscow sequences. This, and other smaller differences, results in a lack of a complete correlation of the Hangzhou ORFs compared to the Paris-Orsay and Moscow sequences (which are very similar, though not identical). Consequently, the Paris-Orsay AY692264.1 sequence coordinates have been used for ORF analysis. However, the Hangzhou (AY587007.1) nucleotide sequence publication contains an invaluable

Table 1. Phage nucleotide sequences used in this study

Phage Name	**NCBI Accession Number**	**Reference**
Phage T5 (Moscow)	AY543070.1	unpublished
Phage T5 (Hangzhou)	AY587007.1	(4)
Phage T5 (Paris Orsay)	AY692264.1	unpublished
Enterobacteria phage EPS7	CP000917.1	(6)
Yersinia phage phiR201	HE956708.2	unpublished
Enterobacteria phage SPC35	HQ406778.1	(7)
Escherichia phage AKFV33	HQ665011.1	(8)
Escherichia phage FFH1	KJ190157.1	(9)
Salmonella phage Stitch	KM236244.1	(10)
Enterobacteria phage DT57C	KM979354.1	(11)
Salmonella phage Shivani	KP143763.1	(12)
Escherichia phage chee24	MF431730.1	(13)
Escherichia phage chee130_1	MF431736.1	(13)
Salmonella phage BSP22A	KY787212.1	unpublished
Escherichia phage Gostya9	MH203051.1	(14)
Pectobacterium phage My1	JX195166.1	(15)
Pectobacterium phage DU_PP_V	MF979564.1	unpublished
Klebsiella phage IME260	KX845404.2	unpublished
Klebsiella phage Sugarland	MG459987.1	(16)
Providentia phage PR1	KY363465.1	(17)
Salmonella phage Sw2	MH631454.1	unpublished
Salmonella phage Seafire	MK050846.1	(18)
Escherichia phage BF23	Mei Liu, Pers comm.	unpublished
Phage T5 injection stop signal	M16226.1	(5)

Table 1. Phage nucleotide sequences used in this study

detailed analysis of the regions containing repeats, inverted repeats and palindromes and these coordinates are used for discussion of the repeat regions.

3. Results

3.1. Phylogenetic relationships of T5-like phages

A phylogeny analysis (Figure 2) using the Dmp protein shows that most T5-like phages form a single clade (Type1) which includes T5, BF23, FFH1, AKFV33, chee24, Shivani, Gostya9, DT57C, EPS7, BSP22A, chee130-1, Stitch, phiR201, SW2 and Seafire. However Type 2 phages (My1, and DU_PP-V) and Type 3 phages (IME260 and Sugarland) are on separate branches that are evolutionarily distant from each other and from Type 1 phages. A fourth clade (Type 4) has a single member PR1. Confirmation of these separate types of T5-like phages is seen by Clustal analysis for other sequences (not shown).

Members of the Type 1 group have many differences between themselves, yet the difference between Type 1 and Types 2+3 is more profound since the entire region from ORFp011 to ORFp015 is missing and is replaced by other nucleotide sequences containing ORFs that are read in the opposite direction to those of T5.

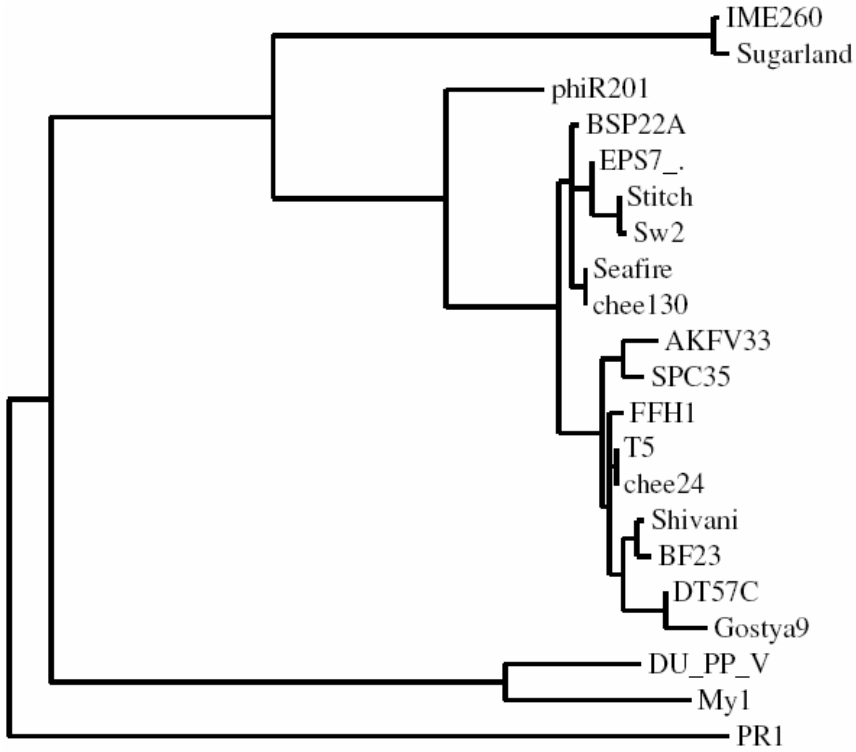

Figure 2. Phylogenetic relationships of the T5-like phages. The relationships were determined with the deoxynucleoside 5'-monophosphatase amino acid sequences of the corresponding phages, using the Phylogeny.fr program.

The only Type 4 phage, PR1, is very different from the others. Oliveira et al (17) suggest that PR1 forms to a new species within the *Siphoviridae* family" and suggest the creation of a new genus within the *Siphoviridae* family, named 'PR1virus. PR1 lacks homology in with the repeat region containing *iss* (the main subject of this review) and thus is not included further in this analysis.

3.2. The ORFs in the T5 FST region

It was interesting to compare the ORFs in the FST region since this would give an indication of which FST ORFs are essential for phage growth. It was further important to establish the similarities and differences between these phages for comparison with the similarities and

differences found in the non-translated *iss* region. As will be noted in the discussion, the non-translated *iss* region is more rigorously maintained than are many of the translated coding regions.

The original mutant hunts (2) identified only 2 genes (*A1* and *A2*) essential for phage viability in the LTR. Another gene *dmp* (encoding a deoxynucleoside 5'-monophosphatase), is not essential for growth and thus was not found as a conditional lethal mutant. These genes are shown in Figure 1, together with the other putative ORFs in the T5 FST region.

It is not excluded that other essential genes could have been missed in early mutant searches, particularly since the searches were not extensive and since some FST ORFs are quite small. In particular, the T5 FST region seems to have more functions than identified genes to accomplish them: injection stop, restart of phage injection, degradation of host DNA (but not T5 DNA), shut-off of host RNA and protein synthesis, shut-off of pre-early genes, protection against restriction enzymes and CRISPR and protection against RecBCD nuclease. Thus, the FST region may contain unidentified genes that may facilitate these functions. For this reason, it was interesting to compare the ORFs of different T5-like phages since important ORFs are likely to be conserved, while non-essential ORFs could be lost. A previous study (3) found that only genes *dmp, A1, A2* and T5p007 were common to all the 10 T5-like phage nucleotide sequences available at that time. The present study confirms this conclusion and extends it to all 21 T5-like phages (Table 2).

3.2.1. The *dmp* gene

The *dmp* gene is always the first (left-most) gene on the genome and is present in all 21 T5-like phages examined. This is curious since viable *dmp* mutants of T5 have been isolated that completely lack the deoxynucleoside 5'-monophosphatase activity, as well as the Dmp protein (19). Thus it might have been expected that some T5-like phage would lack this non-essential gene; yet this is not the case. It is further unclear why T5-phages code for a 5'-deoxyribonucleotidase, since its usefulness in T5 replication is not obvious and most non-T5 bacteriophages do not code for this enzyme. Mutants in *dmp* showed delayed early enzyme production and delayed DNA synthesis, though phage yields were eventually normal. In contrast to wild type T5, *dmp* mutants reincorporated host nucleotides into T5 DNA. The Dmp enzyme has a short lifetime and inactivates before T5 DNA synthesis starts. It is possible that Dmp could protect against an excess of toxic products of host DNA degradation. It is also possible that *dmp* may be an essential gene in the wild (i.e. outside of controlled laboratory conditions).

3.2.2. The *A1* gene

A1 mutants do not inject SST DNA, cause degradation of host DNA, or shut-off of host RNA and protein synthesis or shut-off of T5 pre-early gene activity. The A1 and A2 proteins form a heterodimer which is consistent with a role of both of these proteins in SST injection. A1 protein is associated with membranes where it accounts for up to 10% of the newly synthesized membrane proteins following infection (20). Membrane association is consistent with a role in SST injection since, as shown below, it is likely that the FST DNA is membrane bound.

A1 mutants do not degrade host DNA, suggesting perhaps a nuclease activity. However, apart from one preliminary report (21), no nuclease activity has been reported for the A1 protein or any other pre-early protein. One possibility, discussed in detail below, is that *A1* codes for a site-specific restriction endonuclease. It seems unlikely that a search for a site specific restriction enzyme has been conducted since this would necessitate a specific DNA substrate carrying the appropriate restriction sites, special analysis by gel electrophoresis and perhaps adding ATP and *S*-adenosylmethionine to the reaction mixture. The restriction enzyme hypothesis for the A1 protein, whereby restriction sites are present in the host DNA but not in the FST DNA, would explain the old enigma of how T5 degrades host DNA without degrading its own DNA.

(Finally, one report showed an interaction of the A1 protein with host RNA polymerase, possibly modifying the promoter specificity of this enzyme (22). However other authors failed to observe these interactions with A1 protein (23). Still a third report (24), using phage BF23, indicated

Phage ORF	T5	EPS7	PhiR201	SPC35	AKFV33	FFH1	Stitch	DT57C	BF23	Shivani	chee24	chee130_1	BSP22A	Gostya9	Sp3	SW3	Seafire	My1	DU_PP_V	IME260	Sugarland
T5p001 dmp 244 aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T5p002 131 aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
T5p003 92 aa	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T5p004 A1 543 aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T5p005£ 65aa	+	-	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	+	-	-
T5p006 A2 135aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T5p007 83aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T5p008 67 aa	+	-	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+	+	+
T5p009 172 aa	+	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
T5p010 336 aa	+	+	+	+	+	+	+	-	+	+	+	c	+	-	-	+	+	c	c	+	+
T5p011 87 aa	+	+	+	+	+	+	+	-	+	-	+	+	+	-	+	+	+	-	-	-	-
T5p012 59 aa	+	+	+	+	+	+	-	+	-	+	-	-	+	+	+	-	+	-	-	-	-
T5p013 70 aa	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-
T5p014 114 aa	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	-	-	-	-
T5p015 92 aa	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T5p016 67 aa	+	+	-	+	+	+	+	-	+	-	-	-	+	-	-	+	-	+	+	-	-
End of LTR																					
T5p017 294 aa	+	-	+	-	+	-	-	-	-	+	-	-	-	+	-	+	-	-	-	+	-

Table 2. Presence of T5 LTR protein homologs in other T5-like phages. Amino acid sequence homology searches were performed for the predicted proteins homologous to the those of the Paris-Orsay sequence (AY692264.1) using BlastP program of the NCBI. Symbols are: + = present; - = absent, c = only the C-terminal present.

that the host RNA polymerase interacted with the BF23 equivalent of the A2 protein (rather than *A1*). Since none of these three conflicting observations were confirmed or followed-up, and in view of the contradictory nature of the results, it is difficult to evaluate their respective credibilities.)

3.2.3. *A2* gene

Like *A1* mutants, *A2* mutants do not inject SST DNA. A2 protein has been purified and was shown to be a dimer that binds double stranded DNA (25, 26). It associates with A1 protein, which is a membrane binding protein.

Cloning of the *A2* gene in phage λ shows it has no deleterious effect on λ replication and the λ-T5 $A2^+$ hybrid is able to complement a T5 $A2^-$ mutant (27). Like phage T5, a λ phage carrying the $A2^+$ gene is inhibited in *E. coli* carrying the *ColIb* plasmid. This shows that only the A2 protein is responsible for this abortive infection (28).

3.2.4. ORF T5p007

Apart from *dmp*, *A1* and *A2*, the other FST ORFs have no known mutants or attributable phenotypes. ORF T5p007 is present in all T5-like phages, including Types 2 and 3, and thus could be a candidate for an essential gene; coding perhaps for the restriction insensitivity protein.

3.2.5. ORF T5p003 and T5p015

T5p003 (92 aa) and T5p015 (92 aa) are unique to T5 and are absent in all other T5-like phages.

3.2.6. ORF T5p005 and T5p008

T5p 005 (65aa) and T5p008 (67 aa) are present in all Type 1 phages (except EPS7, DT57C and Gostya9) and thus could be candidates for a useful, but non-essential, genes. T5p008 is also present in Types 2 and 3 T5-like phages, while T5p 005 is only present in Type 1.

3.2.7. ORF T5p009 HNH endonuclease gene

In T5 Repeat Regions 1 and 2 (Figure 1) are separated by the insertion of a HNH endonuclease gene (T5p009, 172 aa). Such HNC nuclease insertions are found at various loci in many different bacteriophages, including elsewhere in the T5 genome. Among the T5-like phages, only Shivani and FFH1 carry HNH at this particular locus. Thus, the other T5-like phages provide an interesting evolutionary flashback to a time before the T5 acquired this insert. The insertion of the T5 HNH endonuclease gene was a complicated event: it is a 604 bp insertion that replaces a 261 bp region present in most T5-like phages.

3.2.8. T5p010

T5p010 (336 aa; sometimes known in Genbank as the 37 kD pre-early protein) is present in most T5-like Type 1 phages (except in the deletion phages DT57C and Gostya9). It is also present in Type 3 phages (Sugarland and IME260) but curiously it is in the reverse orientation. In Type 2 phages My1 and DU_PP_V it is also in the reverse orientation but, even more curiously, only the C-terminal coding half of the gene is present. Similarly, in chee130_1 only the C-terminal half of the gene is present (though it is in the 'normal' T5 orientation). The conservation of at least part of the T5p010 gene could indicate that it has importance to the T5 life cycle, though its complete absence in Gostya9 and DT57C seems to argue against this.

3.2.9. T5p011 and T5p012

T5p011 and T5p012 are missing from many T5-like phages and can be considered to be non-essential.

3.2.10. The relationships between the ORFs of T5-like phages

The T5 coding region 1 to 6022, corresponding to the rightward transcribed genes (Figure 1), is generally conserved, with several variations, in most T5-like phages. The presence, or absence, of these ORFs is shown in Table 2.

In contrast, the coding region 6209 to 9194 between Repeat Region 2 and Repeat Region 3 of T5 shows highly divergent nucleotide sequences, with many deletions/insertions, between the different T5-like phages. In this region, phage T5 has almost no DNA homology to the Type 2 (My1 and DU_PP_V) and Type 3 T5-like phages (IME260, and Sugarland). Furthermore, it is transcribed from the opposite strand.

Phage DT57C is instructive since it has the smallest LTR (8081 bp) containing only 11 ORFs (compared to only 16 in T5 and 22 in phage My1) due to a deletion (compared to T5) that removes T5p008, T5p009, T5p010, and T5p011. Gostya9 similarly has a large deletion in this region.

Phage My1 (Type 2) is also interesting since it has the longest LTR (12854 bp) containing 22 ORFs. Many of these ORFs are unique to My1, or are shared only with the other Type 2 phage DU_PP_V. Both My1 and DU_PP_V lack T5 ORFs T5p011 through T5p015 though they retain T5p016. ORF My1_018 is curious, being homologous to ORFs from DU_PP_V, but also to Type 1 phages chee24, Shivani, SW2, EPS7, and Stitch; but having no homolog in T5.

3.3. The structure of the untranslated FST Repeat Regions

The LTR of T5 carries 3 untranslated repeat regions:
1) Repeat Region 1, (629 bp) 4682 to 5310
2) Repeat Region 2, (188 bp) 6022 to 6209
3) Repeat Region 3, (1.2 kb) 9194 to 10409
These 3 repeat regions are highly conserved with the same order of repeats, inverted repeats and palindromes between almost all of the T5-like phages indicating that they serve some important functions (3). It is instructive to contrast the high level of conservation of the untranslated repeat regions to the much lower level conservation of many LTR ORFs, among the T5-like phages; where only *dmp*, *A1*, *A2* and ORFp007 are consistently conserved.

Repeat Region 1 is present in almost all T5-like phages and contains IRA and IRA' and IRB. However, its locus is too near to the left end of the LTR to be considered as a candidate for *iss* which should be in Repeat Region 3.

Repeat Region 2 contains the IRC, IRD and IRB'. Repeat region 2 is also not a prime candidate for the *iss* signal; which should be located in Repeat Region 3. On the other hand, it contains inverted repeats IRCspacerIRD (6022 to 6152) which are nearly identical to IRD'spacerIRC' in Repeat Region 3 and this latter is a probable candidate for *iss*. In the Discussion section, it will be argued that both IRCspacerIRD and IRD'spacerIRC' are part of the *iss*, despite the large distance between them.

Repeat Regions 1 and 2 are separated in phage T5, Shivani and FFH1 by the insertion of the HNH endonuclease gene. However, most T5-like phages do not carry this insert and it is obvious that Repeat Regions 1 and 2 were originally a single repeat region in the ancestor of T5-like phages.

3.3.1. Repeat Region 3: Repeats, Inverted Repeats and Palidromes

Repeat Region 3, (1.2 kb) 9194 to 10409 is mostly in the LTR but overruns into the unique genome by 270 bp. Repeat Region 3 is extremely complicated, rather like a Matryoshka doll, or Russian nesting doll, where one structure hides another, which in turn hides another and then another (Figure 3). It contains palindrome palC, palindrome palD, inverted repeat IRD', direct repeat DRE1, direct repeat DRB copy 1, direct repeat DRB copy 2, inverted repeat IRC', direct repeat DRE2, direct repeat C copy 1, palindrome palE, direct repeat DRC copy 2, direct repeat DRC

copy 3, palindrome palH, palindrome palG, direct repeat DRC copy 4, palindrome palI, palindrome palJ, 9 copies of direct repeat DRD, and direct repeat DRA copy 2, and copy 3.

Repeat Region 3 is in the correct position, just to the left of the shear site, to contain the *iss* site and is shown in Figure 3. It comprises at least 3 major kinds of repeat sequences:
1) IRD'spacerIRC' (9332 to 9462, 131 bp), 2) two DRE repeats (66 bp), 3) a series of nine DRD repeats.

3.3.2. Repeats Inverted Repeats and Palindromes in Repeat Region 3

It was shown (5) that the Repeat Region 3 contains a multitude of repeats, inverted repeats and palindromes and this was confirmed by sequencing of the entire genome (4). However, during the course of the present investigation some new repeats and inverted repeats were discovered. The use of the word 'spacer' in the naming of these repeats is an attempt to maintain the IRD' and IRC' nomenclature of Wang et al (4). It should not be interpreted to believe that the 'spacer' is without function. Indeed, evidence is given below that the spacer and the inverted repeats contain functionally important elements.

3.3.2.1. IRD'spacerIRC' and IRCspacerIRD 131bp inverted repeats

IRD'spacerIRC' in Repeat Region 3

This inverted repeat comprises the IRD' and IRC' inverted repeats as well as the 52 bp spacer between them. This region contains PalD, DRG3, and DRG4.

>AY587007.1: c9462-9332 Reverse complement
AGTGAGAAACACTGTTATTTAAGTTATCCACAGACTTATCAACAGCACGATTCGC
CTTGACAAAATCCTAGCCATTTTGAAGCGTTATCAACAGACTTATCCCCAGGTTA
TCCTACTGTATATTTATACAG

IRCspacerIRD 131bp inverted repeats in Repeat Region 2

IRCspacerIRD comprises the IRC and IRD inverted repeats as well as the 52 bp spacer between them:

>AY587007.1: 6022-6152 Bacteriophage T5
AGTGAGAAACACTATCATTTAAGTTATCCACAGCCTTATACATTTTGTAAATTAAC
TTGACACAATAGGAAGGCTTTATAGCACTTTTACACAGACTTATCCACAGGTTAT
TCCACTGTATAAATATACAG

The 131 bp IRCspacerIRD region has close homology to the 131 bp IRD'spacerIRC' (Figures 3 and 4) region except that it is inverted in orientation. It is located in Region 2 (3180 bp to the left of IRD'spacerIRC') yet, despite this distance, it will be argued below that it nonetheless part of the *iss* signal. The fact that these large inverted repeats (IRD'spacerIRC' and IRDspacerIRC) are present in all T5-like phages, and also separated by several thousand bp, reinforces this viewpoint. The IRD'spacerIRC' and IRDspacerIRC are found, at comparable positions, in all typical T5-like phages (and indeed phiR201 contains an additional copy of IRD'). However, the atypical T5-like phages Type 2 and Type 3 phages are again exceptions: Type 2 phages (My1 and DU_PP_V) carry IRD'spacerIRC' but IRDspacer and only half of IRC, while Type 3 phages (IME260 and Sugarland) carry IRD'spacerIRC' and IRD, but not IRC.

The spacer distance between IRD' and IRC' and between IRC and IRD is always 52 bp for all T5-like phages (see below). The significance of this is unknown, but the double helix makes one complete turn about its axis every 10.4–10.5 base pairs. Thus, they are separated by 5 (52/10.4) helix turns which may ensure that certain features are always on the same side of the DNA (for example for protein binding). The spacer region is also a region of DNA bending which again may indicate protein binding.

In T5, the 131 bp inverted repeats IRCspacerIRD are closely homologous to the 131 bp inverted repeats IRD'spacerIRC' with a separation distance of 3180 bp. Inverted repeats IRCspacerIRD and IRC'spacerIRD' are found in all typical T5-like phages though the separation

distance varies widely. For DST57C it is only 1851bp, while for BSP22A it is 4258 bp. Thus, this separation distance does not seem critical for whatever function these regions serve.

Finally, (as discussed in detail below), each inverted repeat (IRC, IRD, IRD' and IRC') contains two DnaA boxes, one of which (in each case) overlaps into the spacer.

3.3.2.2. DRE 66 bp direct repeats

There are two large newly identified (66 bp) repeat regions immediately to the right of the IRC' repeat:

DRE1 >AY587007.1: 9455-9520
TTCTCACTTCCGTTTTTAAATACGAATCATTATCATTCGTATTCCGCTTTTTAAATGAG
AATCATT

DRE1 contains PalE, PalF, DRC1 and DRF1.

DRE2 >AY587007.1: 9559-9624
TTCTCATTCCCTTTCTTAAATGAGAATCATTATCATTTGCATTTCACTTTTTAAATGAG
ATTGATT

DRE2 contains PalG, DRC3, PalH, DRC4, DRF3 and part of PalI.

3.3.2.3. DRF direct repeats

DRF1 23bp >AY587007.1: 9495-9517
ATTCCGCTTTTTAAATGAGAATC
DRF2 23bp >AY587007.1: 9518-9540
ATTCCGCTTTTTAAATGAGAATC
DRF3 23 bp >AY587007.1: 9564-9586
ATTCCCTTTCTTAAATGAGAATC
DRF4 20bp >AY587007.1: 9599-9618
ATTTCACTTTTTAAATGAGA

3.3.2.4.DRB and DRG direct repeats

DRB1 23 bp >AY587007.1: 9379-9357
CTGGGGATAAGTCTGTTGATAAC 9379-9357

DRB2 23 bp >AY587007.1: 9379-9357
CTGTTGATAAGTCTGTGGATAAC 9418-9440

DRB1 and DRB2 (23 bp) repeats were reported (4) in both IRD' and IRC'. This is curious since IRD' and IRC' are inverted repeats, paired with IRD and IRC, and thus, logically, there should be 2 DRB sites in these latter too. A verification found four additional repeats, in IRC, IRD, IRD' and IRC' respectively. However, the new repeats are shorter (18 bp compared to 23 bp for DRB) and, while they overlap the DRB repeats, they no longer correspond to them. They have thus been named DRG1 through DRG4. The positions of the new DRG repeats are shown in Figure 3. Thus, all 4 sequences IRC, IRD, IRD' IRC' carry a common component; DRG, which may imply a common ancestry. As will be shown below, the DRB repeats each two carry DnaA boxes.

DRG1 17 bp >AY587007.1: 6044-6061
GTTATCCACAGCCTTAT
DRG2 18bp >AY587007.1: 6117-6134
CTTATCCACAGGTTATTC

DRG3 18bp >AY587007.1: c9379-9362
GTTATCAACAGACTTATC
DRG4 18bp >AY587007.1: c9440-9423
GTTATCCACAGACTTATC

3.3.3. DRE repeats

The newly defined 66 bp repeats, DRE1 and DRE2 are present twice in Repeat Region 3 (together with the repeats and palindromes they contain). DRE1 begins immediately to the right of IRD'spacerIRC' and is separated from DRE2 by a 37bp region containing the 25 bp repeat DRC2 (Figure 2). The DRE repeats are present, at least partially, in all T5-like Types 1, 2 and 3.

3.3.4. Multiple DRD repeats

The right end of the T5 LTR contains 9 copies of DRD (TGTGCAAATCCGA), 6 of which are in the LTR and 3 nearby in the left end of the unique DNA (Figure 3). Because of their positions, too far the right of the LTR, they are not considered candidates for *iss*. All T5-like phages (Types 1, 2 and 3) also have multiple copies of Direct Repeat D (DRD) in this region, indicating some important function such as the generation of the unique ends of the T5 genome or DNA packaging (3). However, against this hypothesis is the fact that there are three additional DRDs at coordinates 12073, 79020 and 79036 where they could not be involved in these processes.

3.3.5. DnaA boxes

It was shown by Heusterspreute et. al. in 1987 (5), that the IRD'spacerIRC' region contains one consensus DnaA box TTATCCACA, and also contains 3 variant DnaA boxes; TTATCAACA (present twice) and TTATCCCCA (present once) (Figure 3). For convenience these variant sites will be collectively referred to below as DnaA* boxes.

This observation prompted a search for DnaA boxes and DnaA* boxes in the same positions in other T5-like phages. The results are shown in Figure 4. It can be seen that consensus DnaA boxes (following the definition of Schaper and Messer (29) TT(A)/(T)TNCACA), as well as DnaA* boxes, are present in the IRD'spacerIRC' region of all T5-like phages.

Figure 4 also shows the IRCspacerIRD (drawn as the reverse complement of the IRD'spacerIRC' region) and it can be seen that DnaA boxes and DnaA* boxes are also common in the IRCspacerIRD region.

It is interesting that the DnaA and DnaA* boxes are always in pairs separated by 3 bp (which is reminiscent of the clustering of DnaA sites around the oriC region). They always in the same position relative to the two IRD'spacerIRC' and IRCspacerIRD inverted repeats with one half in the IR and the other half of other in the spacer (Figure 4). Indeed, almost all T5-like phages carry eight DnaA boxes or DnaA*boxes in the FST region: with 2 in IRD', 2 in IRC', 2 in IRC and 2 in IRD.

Of the total of 153 DnaA boxes shown in Figure 4, 50% (76/153) are consensus DnaA boxes (TTATCCACA, 71; TTATTCACA, 4; TTTTACACA, 3; TTATGCACA, 2) and 77/153 are DnaA* boxes (the most common being TTATCAACA, 21 and TTATTAACA, 17) which differ from the consensus DnaA boxes by only 1 or 2 bp. However it should be noted that consensus site was based upon the binding of naturally occurring DnaA boxes around *oriC*, *rpoH* and *mioC*. Thus, it is not known whether the DnaA* boxes would bind DnaA protein more weakly or more strongly. In addition, DnaA protein binding properties concerns not only the nucleotide sequence but also ATP, the context of the neighboring nucleotides, the interaction with bending proteins IHF and FIS and cooperativity between DnaA proteins.

The presence of the 3 deletions in IRC of Type 1 phages (T5, Stitch and SW2) seems to indicate that not all eight DnaA and DnaA* are absolutely essential. This is supported by the

observation that Type 2 and 3 phages lack part (or all) of IRC (respectively). It is noteworthy that the missing DnaA boxes are always in the IRC region.

DNA replication origins have been shown to be near the centre of the T5 DNA molecule (30, 31, 32). It thus seems unlikely that the DnaA boxes in the FST DNA are concerned with DNA replication. Furthermore, T5 replication functions are carried by the SST region and thus are not expressed during the FST period.

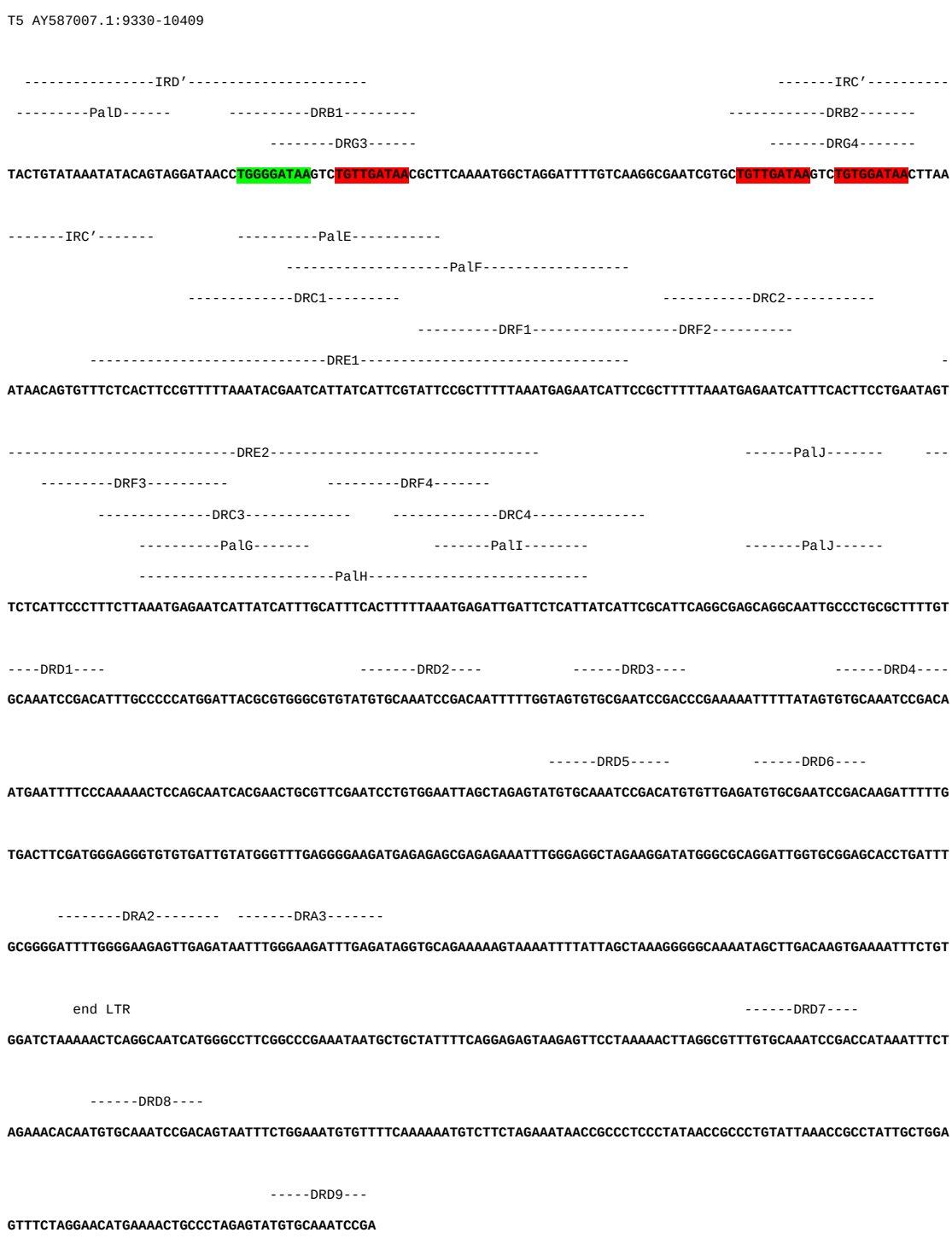

Figure 3. The repeats, inverted repeats and palindromes in Repeat Region 3 of T5.The nucleotide sequence of the untranslated region (from coordinates 9330 to 10409 of AY587007.1) is shown together with the repeats, inverted repeats and palindromes it contains. The 131 inverted repeat IRD'spacerIRC' (9332 to 9462) comprises IRD', IRC' and the region between them. The other partner of this inverted repeat is the 131 bp IRCspacerIRD (coordinates 6022 to 6152), not present on this Figure, but is shown on Figure 4). The approximately calculated shear-point for the FST would be at

coordinate 9618 in DRE2. The positions of the DnaA (green) and DnaA* (red) protein binding sites are shown (see also Figure 4). Because of the strand orientation, the DnaA protein binding sites are shown as the reverse complement. The end of the LTR is indicated.

3.4. The terminase substrate and packaging

T5 differs from circularly permuted phages (such as T4) in that it has unique ends. These ends must be generated by the terminase enzyme: the large and small subunits of which have been identified by sequence homology (33). However, the terminase substrate (defined as the DNA loci on which the terminase acts to produce the unique ends and packaging into phage particles) has not been identified (3). Some indication of the terminase substrate could be obtained by examination of the nucleotide sequences around the ends of the genome; including the left and right termini and the junction of the right end of the unique genome and the RTR. A previous investigation using 10 different T5-like phages suggested that the DRD repeat, which is present in multiple copies in the region just left of the right end of the T5 LTR (Figure 2), might be involved in the terminase substrate and the generation of genome ends (3). The present study confirms that DRD repeats are present, in this position, in all 21 T5-like phages but gives no evidence for their role in the terminase substrate. Furthermore, there are two DRD repeats elsewhere in the T5 genome at coordinates 79019 to 79048 where they are no likely to be involved in the generation of genome ends. Thus, the identity of the terminase substrate remains unclear.

3.5. Rolling circle formation

Rolling circle formation in T5 is not well understood but has been presumed to be initiated by some kind of general recombination between LTR and the RTR. Such a recombination would necessarily involve the loss of one complete terminal repeat and indeed the circles observed by electron microscopy correspond to the whole T5 DNA molecule minus one terminal repeat (23, 25). During the generation of complete genome molecules from the rolling circle concatemer it is usually considered that the terminal repeats are duplicated before packaging. There is no experimental evidence for this, and no mechanism has been postulated. An alternative to terminal repeat duplication is that only one unit genome is packaged for every two genomes on the rolling circle; which seems wasteful.

4. Discussion

4.1 Events following T5 infection

4.1.1 Phage injection and injection stop

Phage injection takes place from the left hand end of the molecule (34) and is initiated by the simple interaction of the bacteriophage with the host outer-membrane receptor protein FhuA (previously called TonA) which is concerned with ferrichrome transport and also acts as a receptor for phages T1 and ϕ80 (35). *In vivo*, DNA transfer stops; when only 7.9% has been injected. However FhuA seems not to be responsible for the injection stop since complete genome injection can be demonstrated *in vitro* using proteoliposomes containing FhuA protein (36). This is further supported by the observation that a different T5-like phage, BF23, which also undergoes two-step injection, uses a different receptor; BtuB, the *E. coli* vitamin B12 outer membrane transport protein (37). Phages DT57C and EPS7 also uses the BtuB receptor (6, 38). Other T5-like phages described in this article infect very different bacterial hosts and possibly also use different receptors.

Thus the question as to the host protein responsible for the injection stop is unsolved. However, the presumptive *iss* regions of T5, and all of its relatives, carries DnaA boxes to which DnaA protein binds (Figure 4). DnaA protein is ubiquitous among almost all bacteria and is found attached

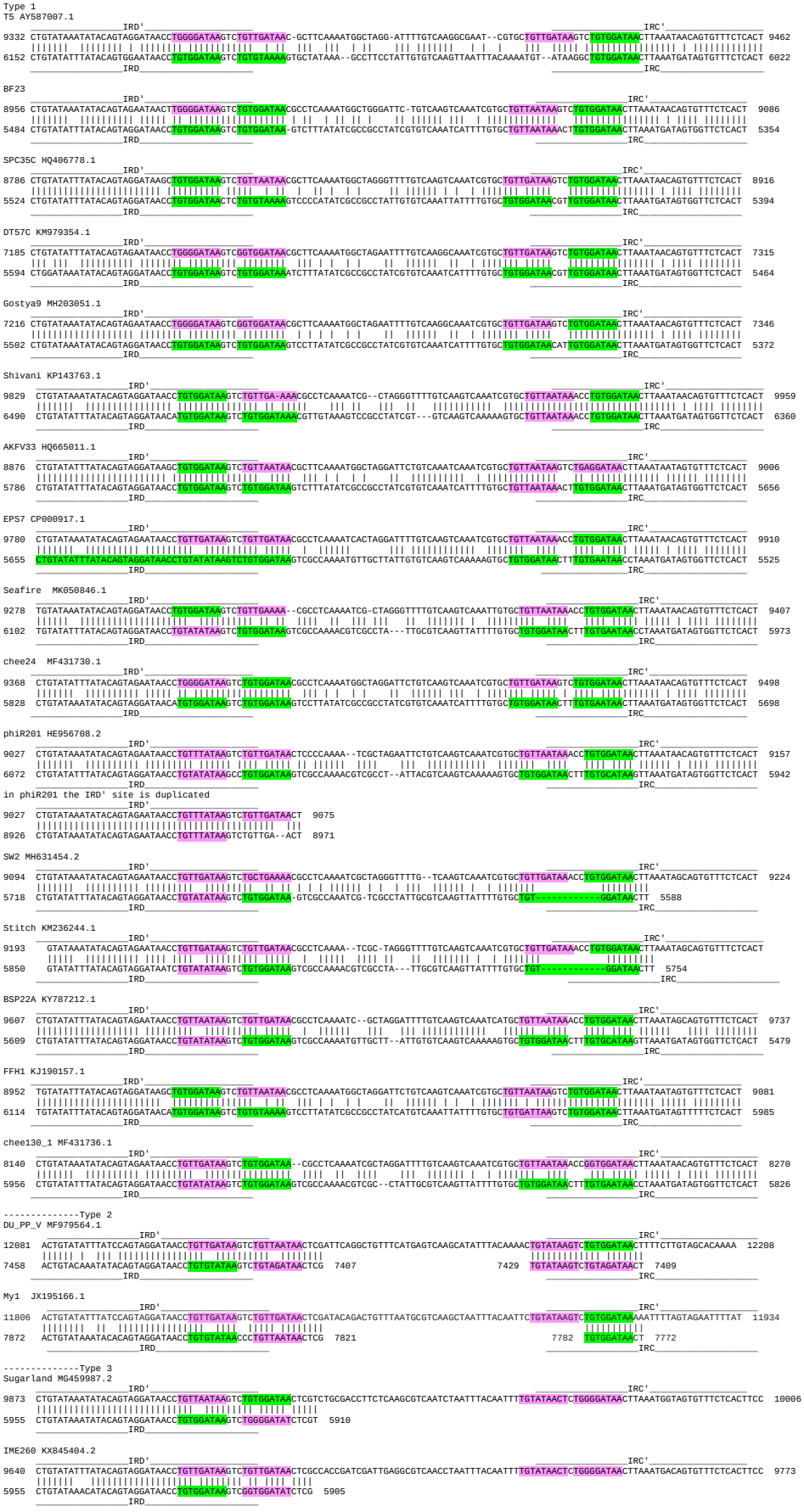

Figure 4. DnaA protein binding sites in the IRD'spacerIRC' and IRCspacerIRD inverted repeats. The inverted repeats IRD'spacerIRC' (top line of each phage) is drawn in the classical direction, 5' to 3'; from left to right. The inverted repeat IRCspacerIRD inverted repeats (bottom line of each phage) is drawn as the reverse complement 3' to 5', from right to left. Figure 4 is drawn in this way only to illustrate the homology and does not imply that these sequences exist in this way *in vivo*. Since these are inverted repeats the two sequences are thus largely homologous. The phage sequences are derived from those in Table 1 and the coordinates of each inverted repeat are shown at the beginnings and ends of the lines. Gaps in the sequences indicate non-homology due to missing nucleotides. The DnaA and DNA* bindings sites are shown. Consensus DnaA binding sites TT(A)/(T)TNCACA (TTATCCACA, 71; TTATTCACA, 4; TTTTACACA, 3; TTATGCACA, 2) are shown in green and non-consensus DnaA binding sites (for example TTATCAACA, 21 and TTATTAACA, 17) are shown in pink. It should be noted that, because of the strand orientation, the DnaA protein binding sites and the DnaA* protein binding sites are drawn as the reverse complement; so that, for example, TGTGGATAA is equivalent to TTATCCACA but on the complementary strand.

to the plasma membrane and forms helical structures along the longitudinal axis of the cell (39). The DnaA protein has a 35-fold higher density in the plasma membrane than in the cytosol. As will be discussed below, it seems possible that FST injection halts because the membrane bound DnaA protein binds to the DnaA boxes of the *iss* region.

4.1.2 DNA degradation

Within the first few minutes of T5 infection the host DNA is degraded to nucleotides and then to free bases which are excreted from the cell. Host DNA degradation also occurs in *A2* mutants, but not in *A1* mutants; suggesting that the A1 protein is involved in host DNA degradation. However, no nuclease activity has been reported to be associated with the A1 protein or any other FST protein. The rapid degradation of host DNA also raises the question of how T5 DNA itself escapes this degradation; particularly since T5 DNA has no special characteristics that distinguish it from that of the host. One possibility (discussed below) is that the *A1* gene encodes a restriction enzyme active on sites in the *E. coli* genome; while the T5 FST DNA carries no such sites. Under this hypothesis complete host degradation would be accomplished by the host RecBCD nuclease (next section).

4.1.3 Protection against RecBCD exonuclease

To quote Dillingham and Kowalczykowski (40) "any phage that exposes free DNA ends as part of its life cycle must find a means to evade destruction by RecBCD". Unlike the restriction insensitivity function (see below), protection against RecBCD nuclease cannot be explained by newly synthesized proteins coded by the FST region. The left end of T5 DNA enters the cell first and should immediately be exposed to RecBCD nuclease before any FST proteins could be synthesized. Many phages such as Mu, P22 and T4 have a terminal protection protein attached to the ends of their DNA in the phage particle and this is injected into the host along with the DNA. It was suggested previously (3) that T5 may also carry a similar protection protein but this has not yet been investigated by testing the infectivity of T5 DNA extracted without the use of denaturing agents such as phenol.

4.1.4 Turn-off of host RNA and protein and FST RNA and protein synthesis

Both host and phage protein and RNA synthesis are turned off after about 5 minutes. The A1 protein (but not the A2 protein) is necessary for this turn-off.

4.1.5 Restriction Insensitivity

Restriction insensitivity was discussed at length in a previous review (3, 41, 42, 43) and will not be repeated here. However, recent article (44) investigated the effect of the CRISPR-cas system on T5. It was found that T5 was insensitive to *E. coli* CRISPR-cas strains targeting the early and late regions of T5. However, T5 was sensitive to *E. coli* CRISPR-cas strains targeting the FST region. Though the authors were unaware of the earlier *EcoRI* restriction insensitivity papers, this CRISPR

situation exactly parallels the results with restriction by *EcoRI*; whereby *EcoRI* sites in the early and late regions are insensitive to *EcoRI* restriction and *EcoRI* sites in the FST region were cleaved by *EcoRI* (41, 42, 43). This suggests that CRISPR-cas may, perhaps, be added to the list of nucleases that are protected by a T5 pre-early function.

4.2. Repeat region 3 and the injection stop signal

The FST region of T5 is defined by the site of shear of T5 infected cells in a blender. The usual estimate of the FST region is left-most 7.9% (roughly 9618 bp from the left); which falls within DRD2 repeat (Figure 3) and iss is thought to be to the left of this shear site (5). Inspection of this region reveals the large inverted repeat IRD'spacerIRC' (131 bp) and the immediately adjacent 66 bp DRE1 and DRE2 repeats, as shown in Figure 3. These are, individually, or together, likely candidates for iss and are present in all T5-like phages. This high degree of conservation is all the more remarkable given the almost complete lack of conservation of the ~3 kb coding region immediately to the left of Repeat Region 3 (see above and Fig. 1).

The IRD'spacerIRC'-DRE cluster is extremely complex containing within it many repeats and palindromes shown in Figure 3. It is impossible to guess the function(s) of these, though their extreme conservation, among 21 different Types 1, 2 and 3 T5-like phages, suggests their importance.

4.3. Repeat region 2 and the injection stop signal

Repeat region 2 begins at 6022 and ends at 6209 and thus is too far leftwards from the shear site 9618 to be considered as *iss*. Yet the 131 bp inverted repeat IRD'spacerIRC' of Repeat Region 3 is also present, in inverted order, as IRCspacerIRD, in Repeat Region 2. This cannot be a coincidence since IRCspacerIRD is present at similar positions in all T5-like phages. This, almost inevitably, leads to the conclusion that IRCspacerIRD and IRD'spacerIRC' are part of a functional complex and that they must both be part of the *iss* injection stop process. This hypothesis is reinforced by the finding of common elements between IRC and IRD and IRD'and IRC'; namely the DnaA boxes discussed below.

One observation may elucidate this situation: while investigating the FST injection of T5, Rogers and Rhoades (45) observed two FST bands on the gels. The first and largest corresponded to the classical FST band of about 9618 bp, but a second band was present at about 5846 bp which (given the possible error measurements in 1979) is very close to the position of IRC at 6022. The authors did not have an explanation for this band. However, this raises the intriguing possibility that the FST DNA may attach to the membrane in two places; at IRD'spacerIRC'and at IRCspacerIRD (coordinates ~9618 and ~6022.)

4.4. The Classical model of T5 Injection

The classical model (2) for the two-step injection of T5 is that injection stops (for unknown reasons) at *iss* and then resumes (for unknown reasons) under the action of the newly synthesized *A1* and *A2* products. While this model has stood the test of time (50 years!), and no other model has been proposed, it is unsatisfactory since it simply describes the phenomenon.

The classical model makes only one prediction, namely that 100% of the genome enters the host cell intact. Two studies investigated the replication of T5 DNA, using electron microscopy, and both observed circular DNA molecules carrying replication forks and the length of the circle was that of the T5 DNA minus the length of one terminal repeat. This confirms the idea that circular molecules are formed by some kind of recombination of the LTR and RTR. However, whereas Bourguignon et al. (24) observed full length linear T5 DNA carrying replication loops, Everett (26) also observed linear molecules carrying replication loops but with a length 10 to 12% shorter than T5 DNA. Everett stated *"Our most unexpected finding was that the contour lengths of peak II DNA and of the slow sedimenting form (ssf) were on average about I0 to I2 % shorter than that of mature*

T5 virion DNA", and concluded "*it appears unlikely that the ssf is a true intermediate in the pathway of T5 DNA replication*". In view of the fundamental disagreement between these two results, the question of whether intact full length T5 DNA enters the cell remains unresolved.

A second major problem with the classical model is that it makes no attempt to reconcile the facts that the A1 protein (directly or indirectly) degrades the host DNA (but spares T5 DNA) and yet A1 protein (with A2 protein) is also required for SST injection. These ambiguities must be resolved and a new hypothesis for T5 injection may thus be needed.

4.5. A new model for T5 injection stop

The first step towards construction of a new model for T5 phage injection is to determine how injection stops and, to this end, the present article has characterized the injection stop signal (*iss*) by comparing the appropriate *iss* regions, for 21 different T5-like phages. A word of caution is, however, necessary in this approach. Only phages T5 and BF23, that have been studied genetically and biochemically and mutants of these phages can complement each other. However two-step injection has been verified only in these two phages. It is thus an assumption, based on the common structure of the *iss*, regions that all T5-like phages use two-step injection.

The present study reveals the vast complexity of the IRD'spacerIRC'-DREa-DREb region, which contains within itself a multitude of overlapping repeats, inverted repeats and palindromes. The fact that the untranslated *iss* region is highly conserved, between different phage genomes that are not themselves well conserved in their coding regions, attests to its importance. Its locus, just to the left of the shear site, suggests it is likely to be (wholly or partly) the *iss* region. The conservation of this region suggests that cannot it be greatly altered without destruction of the *iss* function. Furthermore the present study has demonstrated the presence of 8 precisely organized putative DnaA boxes; one pair of DnaA boxes ie associated with each IRC, IRD, IRD' and IRC'. Each FST region of all T5-like phages has eight DnaA boxes (with rare exceptions; including T5 itself, which has only 7, Figure 4) are arranged in pairs and separated by 3 nucleotides; which is reminiscent of the precise clustering of DnaA sites around the *oriC* region.

It has been shown that, while DnaA protein is purified from the cytosol as a soluble protein, the vast majority is bound to the plasma membrane, with a 35-fold higher density in close proximity to the cell membrane than in the cytosol (46). It remains unknown whether these DnaA proteins would interact with the DnaA boxes of T5 which must pass within intimate proximity during the injection process. This could perhaps be tested using a temperature sensitive DnaA mutant at the non-permissive temperature.

The presence of 8 DnaA boxes in the IRCspacerIRD and the IRD'spacerIRC' regions of the FST DNA, together with the numerous repeats, and palindromes, could cause considerable bending of the DNA when bound to DnaA protein (and perhaps other proteins such a IHF and FIS). It could even form a knot, together with the IRCspacerIRD regions which could physically prevent transfer.

Given this hypothesis as to why FST injection stops, it is now necessary to know how SST injection restarts under the influence of the A1 and A2 proteins.

4.6. The A1 protein restriction enzyme hypothesis

The classical model of T5 two-step injection ignores the inconvenient truth, which was already know at that time, that the A1 protein is probably an endonuclease since it causes degradation of the host DNA. Indeed early experiments reported an endonuclease that cleaved host DNA to acid insoluble material though the *A1* gene product was not specifically identified (21). T5 FST DNA is not degraded, indicating that it is protected from the nuclease that destroys host DNA. The most satisfying explanation is that the A1 protein attacks sites on the host DNA that are not present on the FST DNA. This is a perfect description of a restriction enzyme that would cleave restriction sites on the host but not the FST DNA which lacks these sites.

The hypothesis that A1 protein is a restriction endonuclease explains host degradation but does not explain how it detaches the *iss* signal from the membrane bound DnaA proteins to which it is attached; since there is no reasonable mechanism whereby a restriction enzyme could do this. This

then leaves the mystery of how then SST injection restarts. This new model suggests the A1 restriction enzyme (possibly in cooperation with the *A2* gene product) also cleaves the FST DNA at a restriction site to the right of the site of blockage of the membrane bound FST DNA. By cleaving the FST in this way, injection is free to continue, leaving the FST DNA behind still attached to the DnaA protein in the membrane.

Restriction enzymes often cleave palindromic sequences and it should be noted (Figure 2) that there are several palindromes in the appropriate region of the T5 LTR region (PalD, PalE, PalF, PalG, PalH, PalI and PalJ). It should also be remembered that these palindromes are also present in the RTR. Thus cleavage of the LTR restriction site should also imply cleavage of the RTR restriction site when it enters the cell on the SST DNA. Restriction enzymes often cleave palindromic sequences to give cohesive ends and the presence of homologous cohesive ends in the LTR and the RTR, following cleavage, could facilitate circle formation in preparation for rolling circle replication.

It is counterintuitive to imagine that a phage would simply abandon about 7.5 % of its genome, yet it should be remembered that this would happen anyway. At the time of circle formation it has always been considered that the LTR and the RTR fuse by recombination with the result that the equivalent of one terminal repeat is lost (24, 26). Thus the only difference is one of timing: one terminal repeat is lost earlier than previously imagined.

4.7. The role of A2 protein

A2 protein is also needed, together with A1 protein, for SST transfer. A2 polypeptide has a molecular weight of 15000 has been purified and shown to be a DNA binding dimer (25). The A1 and A2 proteins are known to bind together and to be membrane associated. A2 protein does not play a role in host DNA degradation. The function of the A2 protein in second step transfer is thus unknown and any suggestion would be pure speculation.

Lanni speculated that A2 protein could have a role to protect T5 FST DNA from the A1 nuclease. Another possibility, suggested here, is that A2 protein could modify the A1 restriction enzyme, to which it is bound, to change its specificity to be able to recognize the restriction site located in the FST DNA. According to this hypothesis the restriction site recognized by the A1/A2 complex would be more specific than the sites on the host DNA recognized by A1 alone. A possible extension of this hypothesis would be that the A1/A2 complex could have an active role in circle formation by joining the cohesive ends of restriction cleavage sites on the LTR and the incoming RTR. This would be site-specific recombination. Such a function could be analogous to the *int* protein of phage λ which catalyses λ DNA integration to form a prophage; yet, when combined with *xis* protein, *int* specificity changes and causes prophage excision.

4.8. The role of IRCspacerIRD

The above model has considered that the *iss* signal is near and to the left of the shear site and thus IRD'spacerIRC' is considered a likely candidate for *iss*. While this is probably correct, it does not exclude a role of IRCspacerIRD (which is 3180 bp to the left) in the *iss* injection stop signal. IRD'spacerIRC' and IRCspacerIRD could both have a role as *iss*, whereby the DnaA boxes they contain could bind the FST DNA to the membrane to form a Gordian knot (perhaps including the repeats and palindromes they contain) that prevents injection. The only solution, to be rid of the knot, is to cut the DNA.

4.9. The new restriction enzyme/DnaA protein model.

The new model is summarized in Figure 5 and shows *iss* region attached to the membrane bound DnaA protein (Figure 1a). The *A1* coded restriction endonuclease cleaves the host DNA thereby causing host DNA degradation. The host DNA degradation is completed by the action of the RecBCD nuclease. This degradation would shut-off of host messenger RNA and protein synthesis. The A1/A2 heterodimer also cleaves restriction sites upstream (rightwards) of the *iss* signal,

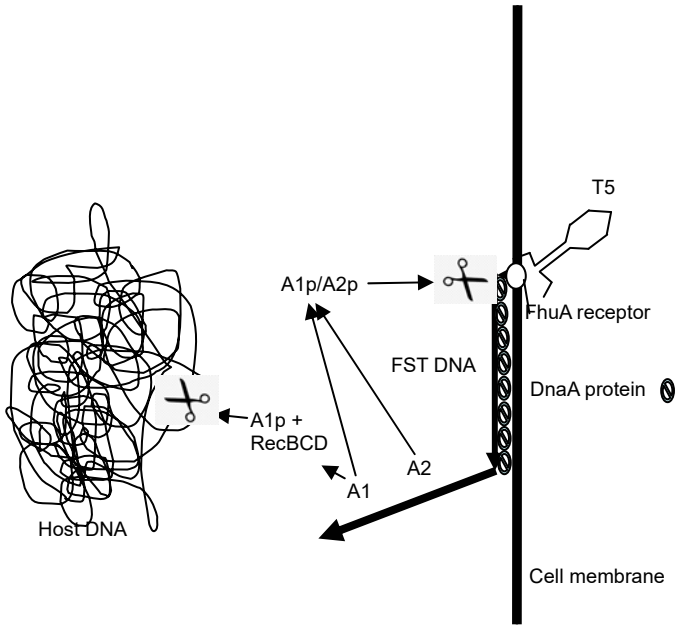

Figure 5a

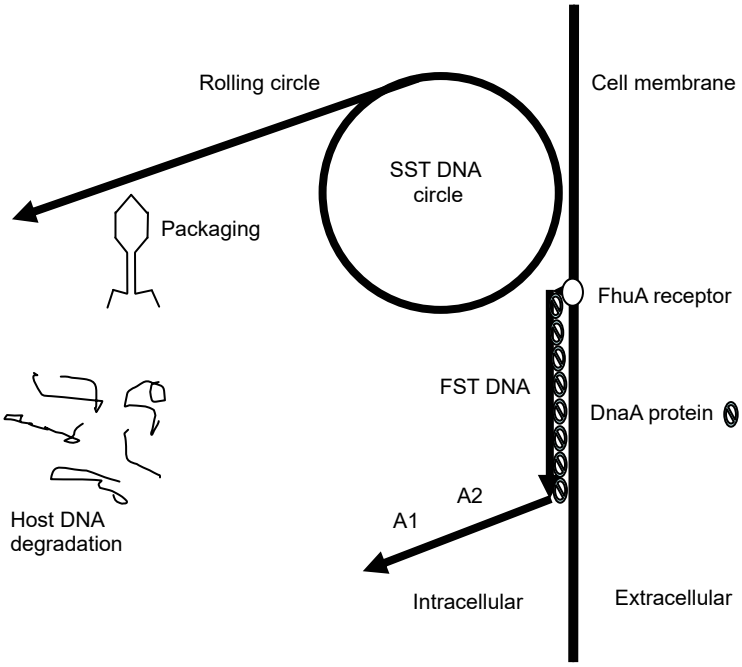

Figure 5b

Figure 5. First and second step injection of the T5 genome. The vertical black line represents the membrane with the interior of the cell to the left and the outside to the right. Figure 5a: This figure is nor drawn to scale. The T5 phage is bound to the FhuA receptor (open circle) and has injected the FST DNA (black arrow). The FST DNA is bound by its DnaA protein binding sites to the DnaA protein (barred circles) in the membrane. The A1 protein is a restriction endonuclease that cleaves the host DNA. Complete host DNA degradation is achieved by the host RecBCD nuclease. Figure 5b: This figure is nor drawn to scale and the rolling circle in is, in reality, about ten times the size of the FST DNA. The A1/A2 heterodimer endonuclease cleaves the FST DNA at a specific restriction site beyond the blockage caused by the DnaA protein binding sites and permits injection to continue. When the homologous restriction site on the RTR DNA arrives inside the cell, it is similarly cleaved. The two restriction sites join either by their cohesive ends or by site specific

recombination catalyzed by the A1/A2 dimer. This results in circle formation that protects the phage DNA from degradation by RecBCD nuclease. This circle formation initiates DNA replication and eventually rolling circle formation which forms concatameric DNA. The concatameric DNA is then processed and packaged into phage particles by the terminase.

permitting the transfer of the SST DNA; by removing the stop caused by the attachment of the DnaA boxes to membrane-bound DnaA protein (Figure 1b). The A1/A2 heterodimer also cleaves the corresponding site on the incoming RTR DNA and facilitates circle formation either by virtue of the cohesive ends of the restriction cleaved ends or by site specific recombination. DNA replication leads to rolling circle formation and then to concatenate formation. The concatenate is then processed and packaged into phage particles.

4.10. Predictions of the new restriction site model

Naturally, this new restriction site model of T5 two-step injection must be too simplistic since it assigns no role to the conserved multitude of repeats (inverted repeats and palindromes) in the *iss* region. It seems unlikely that these conserved structures are not involved in the injection stop or re-start processes.

The restriction site model predicts the generation of a FST DNA fragment as a normal part of T5 infection. Curiously, strong support for the restriction site model, comes from a 50 year old (1968) experiment by Lanni (2), who was the first scientist to observe two step injection of T5 DNA and its dependence on the synthesis *A1* and *A2* gene products. In one control experiment she verified the DNA from cells infected with T5 but neither blended nor treated with cloramphenicol. As expected high molecular weight phage T5 DNA was present. However, a fragment corresponding in size to FST DNA (reference 2; p230 Figure 2) was also clearly present. The authors did not attempt to assign a biological origin of this FST fragment, but instead suggested DNA breakage during extraction (47) leaving unclear why breakage of the large T5 DNA molecule would generate a fragment of FST size.

The presence of this FST fragment demonstrates that the FST fragment does not only derive from blending (since no blending was used) but is also a natural part of the infection process. This result is incompatible with the classical model whereby the T5 DNA molecule is injected intact. In contrast, the presence of this FST fragment is a prediction of the restriction enzyme model outlined above.

The new model takes into account the nuclease properties of the A1 protein as well as the presence of precisely orientated DnaA boxes within the *iss* region. It provides a different viewpoint of two-step injection, and, unlike the standard model, it makes predictions that can be tested. Using this insight, it should be possible to test for restriction endonuclease activity in the early stages of T5 infection. The additional prediction that T5 DNA has two restriction sites for the A2 modified A1 restriction endonuclease, one in each terminal repeat, could also be tested.

It would be interesting to investigate whether T5 infection of a thermo-sensitive DnaA mutant, at the non-permissive temperature, would give a two-step injection or whether all of the DNA would enter the cell in one singe step; which would probably be lethal for both the phage and the host.

Finally, this report illustrates the difficulty of using genetics to investigate cis-acting essential functions where mutation would result in a lethal phenotype. A possible solution is to use a heterologous system. For example, it may be possible to clone the *iss* region into a small mobilizable plasmid vector (such as pJRD215 (48)). If so, perhaps the *iss* region would prevent conjugal transfer of the plasmid to another host and this may be a way to select mutants in the *iss* region. Such a possibility may open up the *iss* region to a new kind of genetic analysis.

5. References

1 Sayers JR. 2006. Bacteriophage T5. In: Calendar, R. (ed) The Bacteriophages. Oxford University Press, NY.; p 268-76. ISBN 0-19-514850-9

2 Lanni YT. 1968. First-step-transfer deoxyribonucleic acid of bacteriophage T5. Bacteriol Rev *32*:227–42. PMID:4879238

3 Davison, J. 2015. Pre-early functions of bacteriophage T5 and its relatives. Bacteriophage *5*e1086500. PMID:26904381 http://doi: 10.1080/21597081.2015.1086500.

4 Wang J, Jiang, Y, Vincent, M, Sun, Y, Yu H, Wang J, Bao Q, Kong H, Hu S. 2005. Complete genome sequence of bacteriophage T5. Virology *332*:45–65. PMID:15661140; http://doi:10.1016/j.virol.2004.10.049

5 Heusterspreute M, Ha-Thi V, Tournis-Gamble, S, Davison J. 1987.The first-step transfer-DNA injection-stop signal of bacteriophage T5. Gene *52*:155-164.

6 Hong J, Kim K-P, Heu S, Lee SJ, Adhya S, Ryu S. 2008. Identification of host receptor and receptor-binding module of a newly sequenced T5-like phage EPS7. FEMS Microbiol Lett 289:202-209. PMID:19025561; http://dx.doi.org/10.1111/j.1574-6968.2008.01397.x

7 Kim M, Ryu S. 2011. Characterization of a T5-Like Coliphage, SPC35, and Differential Development of Resistance to SPC35 in *Salmonella enterica* Serovar *Typhimurium* and Escherichia coli. Appl. Environ. Microbiol 77:2042-2050. PMID:21257810; http://dx.doi.org/ 10.1128/AEM.02504-10

8 Niu YD, Stanford K, Kropinski AM, Ackermann HW, Johnson RP, She YM, Ahmed R, Villegas A, McAllister TA. 2012. *Escherichia* phage bV_EcoS_AKFV33, complete genome. Genomic, Proteomic and Physiological Characterization of a T5-like Bacteriophage for Control of Shiga Toxin-Producing *Escherichia coli* O157:H7; *PLoS ONE 7:*. E34585 PMID: 22514640; http://dx.doi.org/10.1371/journal.pone.0034585

9 Hong Y, Pan Y, Harman NJ, Ebner PD. 2014. Complete genome sequences of two *Escherichia coli* O157:H7 phages effective in limiting contamination of food products. Genome Announ *2*:e00519-14. doi:10.1128/genomeA.00519-14.

10 Grover JM, Luna AJ, Wood TL, Chamakura K.R, Kuty Everett GF. 2015. Complete Genome of *Salmonella enterica Serovar Typhimurium* T5-Like Siphophage Stitch. Genome Announc *3*: e01435-14. Published 2015 Feb 5. doi:10.1128/genomeA.01435-14

11 Golomidova A.K, Kulikov EE, Prokhorov NS, Guerrero-Ferreira RC, Ksenzenko VN, Tarasyan KK Letarov A.V. 2015. Complete genome sequences of T5-related *Escherichia coli* bacteriophages DT57C and DT571/2 isolated from horse feces. *Arch. Virol 160*: 3133–3137.

12 Piya, D.; Xie, Y.; Hernandez, A.C.; Everett, G.F.K. 2015: Complete Genome of *Salmonella enterica serovar Typhimurium* Siphophage Shivani. Genome Announc *3(1).* PMID:25720685; http://dx.doi.org/ 10.1128/genomeA.01443-14

13 Sváb D, Falgenhauern L, Rohdebn M, Szabón J, Chakraborty T, Tóth I. 2018 Identification and Characterization of T5-Like Bacteriophages Representing Two Novel Subgroups from Food Products. Front Microbiol *9:202*. doi: 10.3389/fmicb.2018.00202.

14 Golomidova AK, Kulikov EE, Babenko VV, Ivanov PA, Prokhorov NS, Letarov AV. 2019. *Escherichia coli* bacteriophage Gostya9, representing a new species within the genus T5 virus. Arch Virol *164(3)*:879-884. doi: 10.1007/s00705-018-4113-2. Epub 2018 Dec 1.

15 Lee DH, Lee JH, Shin H, Ji S, Roh E, Jung K, Ryu S, Choi J, Heu S. 2012. Complete Genome Sequence of *Pectobacterium carotovorum subsp. Carotovorum.* Bacteriophage My1. J. Virol: *86 (20)*:11410-11411. PMID:22997426; http://dx.doi.org/: 10.1128/JVI.01987-12.

16 Erickson SG, Lessor L, O'Leary C.J, Gill JJ, Liu, M. 2018. Complete Genome Sequence of *Klebsiella pneumoniae* Siphophage Sugarland. Microbiol Resour Announc. 7(19*)*: e01014-18. doi:10.1128/MRA.01014-18.

17 Oliveira H, Pinto G, Hendrix H, Noben JP, Gawor J, Kropinski AM, Łobocka M, Lavigne R, Azeredo J. 2017. Lytic *Providencia rettgeri* Virus of Potential Therapeutic Value Is a Deep-

Branching Member of the T5 virus Genus. Appl Environ Microbiol *83(23):*e01567-17. doi:10.1128/AEM.01567-17.

18 Hartman S, Zeng C, O'Leary C, Newkirk H, Kongari R, Gill J, Liu M. 2019. Complete genome sequence of *Salmonella enterica* serovar *Enteritidis* siphophage Seafire. Microbiol Res. Announc 8:e01167-19. https://doi.org/10.1128/MRA.01167-19.

19 Mozer TJ, Thompson RB, Berget SM, Warner HR. 1977. Isolation and characterization of a bacteriophage T5 mutant deficient in deoxynucleoside 5'-monophosphatase activity. J Virol: *24*:642–650. PMID: 335083

20 Duckworth DH, Dunn G.B, McCorquodale DJ. 1976 Identification of the gene controlling the synthesis of the major bacteriophage T5 membrane protein. J Virol, 18:542-549. PMID:775127 PMCID: PMC515580.

21 Fielding PE, Lunt MR. 1969. A new deoxyribonuclease activity from bacteria infected with T5 bacteriophage. FEBS Lett 12;5(3):214-217. PMID 11947280, DOI 10.1016/0014-5793(69)80335-2.

22 McCorquodale DJ, Chen CW, Joseph MK, Woychik R. 1981. Modification of RNA Polymerase from *Escherichia coli* by Pre-Early Gene Products of Bacteriophage T5. J Virol 40 (3): 958-962. PMID:7033565 PMCID: PMC256710.

23 Szabo C, Moyer RW. 1975. Purification and properties of a bacteriophage T5-modified form of *Escherichia coli* RNA polymerase. J. Virol 15:1042-1046.

24 Kikuchi S Yoshinari K. Ishimaru H, Mizobuchi K. 1988. Regulation of the Temporal Synthesis of Proteins in Bacteriophage BF23-Infected Cells. J. Virol 62:4569-4576 PMCID: PMC253568 PMID: 3054152.

25 Snyder CE Jr, Benzinger RH. 1981. Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene *A2* protein. J Virol *40*:248–257. PMID 7288924

26 Snyder CE Jr. 1984. Bacteriophage T5 gene *A2* protein alters the outer membrane of *Escherichia coli.* J Bacteriol 60:1191–1195. PMID:6389511.

27 Brunel, F.; Davison, J.; Merchez, M. 1979 Cloning of bacteriophage T5 DNA fragments in plasmid pBR322 and bacteriophage lambda *gtWES*. Gene 8:53-68; PMID:535730.

28 Davison J, Brunel F, Merchez MA. 1979. A new host-vector system allowing selection for foreign DNA inserts in bacteriophage lambda *gtWES*. Gene, 8:69-80. PMID:231542.

29 Schaper S, Messer W. 1995. Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. J Biol Chem. *270(29)*:17622-6. PMID:7615570.

30 Bourguignon GJ, Sweeney TK, Delius, H. 1976. Multiple origins and circular structures in replicating T5 bacteriophage DNA. J Virol 18:245-59. PMID:1255872.

31 Everett RD, Lunt MR 1980. DNA replication of bacteriophage T5. 2. Structure and properties of the slow sedimenting form of intracellular T5 DNA. J Gen Virol, 47:133–150. PMID:7365461.

32 Everett RD 1981. DNA replication of bacteriophage T5. 3. Studies on the structure of concatemeric T5 DNA. J Gen Virol, 52:25–38. PMID:7264606.

33 Ponchon L, Boulanger P, Labesse G., Letellier 2006. L. The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily. A bioinformatics analysis validated by a functional study on bacteriophage T5. J Biol Chem 281:5829-5836. PMID:16377618; http://dx.doi.org/ 10.1074/jbc.M511817200.

34 Shaw AR, Davison J.1979. Polarized injection of the bacteriophage T5 chromosome. J Virol *30*:933-935. PMID:480474.

35 Böhm J, Lambert O, Frangakis AS, Letellier L, Baumeister W, and Rigaud JL. 2001. FhuA-mediated phage genome transfer into liposomes. A cryo-electron tomography study. Curr. Biol.11:1168–1175. PMID:11516947 http://doi: http://dx.doi.org/10.1016/S0960-9822(01)00349-9.

36 Bertin A, de Frutos, M, Letellier L. 2011. Bacteriophage-host interactions leading to genome internalization. Curr Opin Microbiol. 14(4):492-6. doi: 10.1016/j.mib.2011.07.010.

37 Krauel V, Heller KJ 1991. Cloning, sequencing, and recombinational analysis with bacteriophage BF23 of the bacteriophage T5 oad gene encoding the receptor-binding protein. J Bacteriol. *173*:1287-97. PMID:1825083.

38 Golomidova AK, Kulikov EE, Prokhorov NS, Guerrero-Ferreira RC, Knirel YA, Kostryukova ES, Tarasyan KK, Letarov AV. 2016. Branched Lateral Tail Fiber Organization in T5-Like Bacteriophages DT57C and DT571/2 is Revealed by Genetic and Functional Analysis. Viruses. 8(1):pii: E26. doi: 10.3390/v8010026.

39 Saxena R. Fingland N. Patil D, Sharma AK, Crooke E. 2013. Crosstalk between DnaA protein, the initiator of Escherichia coli chromosomal replication, and acidic phospholipids present in bacterial membranes. Int. J. Mol. Sci 14:8517–8537. doi: 10.3390/ijms14048517.

40 Dillingham MS, Kowalczykowski SC. 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev 72:642-71. PMID:19052323.

41 Davison J, Brunel F. 1979. Restriction insensitivity in bacteriophage T5. I. Genetic characterization of mutants sensitive to *EcoRI* restriction. J Virol. 29:11-16. PMID:430589.

42 Davison J, Brunel F. 1979. Restriction insensitivity in bacteriophage T5. II. Lack of *EcoRI* modification in T5+ and T5ris mutants. J Virol. 29:17-20. PMID: 430591.

43 Brunel, F, Davison J. 1979. Restriction insensitivity in bacteriophage T5. III. Characterization of *EcoRI*-sensitive mutants by restriction analysis. J Mol Biol 128:527-543. PMID:374742 http://dx.doi.org/10.1016/0022-2836(79)90291-2.

44 Strotskaya A, Savitskaya E, Metlitskaya A, Morozova N, Datsenko KA, Semenova, E, Severinov K. 2017. The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res *45(4)*:1946-1957. doi: 10.1093/nar/gkx042.PMID:28130424.

45 Rogers SG, Hamlett NV, Rhoades M. 1979. Interruption-deficient mutants of bacteriophage T5. II. Properties of a mutant lacking a specific interruption. J Virol *29*:726–734. PMID:430607.

46 Saxena R, Fingland N. Patil D, Sharma AK, Crooke E. 2013. Crosstalk between DnaA protein, the initiator of *Escherichia coli* chromosomal replication, and acidic phospholipids present in bacterial membranes. Int J Mol Sci. 14:8517–8537. PMCID: PMC3645759. PMID:23595001.

47 Théry-Lanni Y. McCorquodale D.J. Wilson CM. 1964 Molecular aspects of DNA transfer from phage T5 to Host Cells: II. Origin of first-step-transfer DNA fragments. J. Molec Biol 10:19-27. DOI:10.1016/s0022-2836(64)80024-3.

48 Davison J, Heusterspreute M, Chevalier N, Ha-Thi V, Brunel F. 1987. Vectors with restriction site banks. V. pJRD215, a wide-host-range cosmid vector with multiple cloning sites. Gene, 51:275-280. PMID: 3036654.