Group I introns and Homing Endonucleases in T-even-like Bacteriophages

Linus Sandegren

Department of Molecular Biology and Functional Genomics
Stockholm University
Stockholm 2004

Doctoral Thesis 2004
Department of Molecular Biology and Functional Genomics
The Arrhenius Laboratories for Natural Sciences
Stockholm University
SE-106 91 Stockholm
Sweden

ISBN 91-7265-925-4

Abstract

Homing endonucleases are rare-cutting enzymes that cleave DNA at a site near their own location, preferentially in alleles lacking the homing endonuclease gene (HEG). By cleaving HEG-less alleles the homing endonuclease can mediate the transfer of its own gene to the cleaved site via a process called homing, involving double strand break repair. Via homing, HEGs are efficiently transferred into new genomes when horizontal exchange of DNA occurs between organisms.

Group I introns are intervening sequences that can catalyse their own excision from the unprocessed transcript without the need of any proteins. They are widespread, occurring both in eukaryotes and prokaryotes and in their viruses. Many group I introns encode a HEG within them that confers mobility also to the intron and mediates the combined transfer of the intron/HEG to intronless alleles via homing.

Bacteriophage T4 contains three such group I introns and at least 12 freestanding HEGs in its genome. The majority of phages besides T4 do not contain any introns, and freestanding HEGs are also scarcely represented among other phages.

In the first paper we looked into why group I introns are so rare in phages related to T4 in spite of the fact that they can spread between phages via homing. We have identified the first phage besides T4 that contains all three T-even introns and also shown that homing of at least one of the introns has occurred recently between some of the phages in Nature. We also show that intron homing can be highly efficient between related phages if two phages infect the same bacterium but that there also exists counteracting mechanisms that can restrict the spread of introns between phages.

In the second paper we have looked at how the presence of introns can affect gene expression in the phage. We find that the efficiency of splicing can be affected by variation of translation of the upstream exon for all three introns in T4. Furthermore, we find that splicing is also compromised upon infection of stationary-phase bacteria. This is the first time that the efficiency of self-splicing of group I introns has been coupled to environmental conditions and the potential effect of this on phage viability is discussed.

In the third paper we have characterised two novel freestanding homing endonucleases that in some T-even-like phages replace two of the putative HEGs in T4. We also present a new theory on why it is a selective advantage for freestanding, phage homing endonucleases to cleave both HEG-containing and HEG-less genomes.

“Phages really are amazing creatures. You can learn all there is to know about their ways in a month, yet after a hundred years they can still surprise you at a pinch.”

Table of contents

List of papers

This thesis is based on the following original articles and manuscripts, which will be referred to by their roman numerals.

I **Sandegren. L.**, Sjöberg. B-M.
Distribution, Sequence Homology and Homing of Group I Introns among T-even-like Bacteriophages. Evidence for recent transfer of old introns.
Journal of Biological Chemistry **279** (21) pp. 22218-22227 (2004)

II **Sandegren. L.**, Sjöberg. B-M.
Self-splicing of the Bacteriophage T4 Group I Introns is Affected by the Growth of the Infected Bacterium and Requires Efficient Translation of the Pre-mRNA In Vivo.
Manuscript

III **Sandegren. L.**, Nord. D., Sjöberg. B-M.
Two Genes Encoding Novel Homing Endonucleases Replace the Putative Homing Endonuclease Genes mobC and mobE in Several T4-related Phages.
Manuscript

Abbreviations

DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
mRNA	Messenger RNA
rRNA	Ribosomal RNA
tRNA	Transfer RNA
HEG	Homing Endonuclease Gene
ORF	Open Reading Frame

Nucleotides:

ATP	Adenosine Triphosphate
dATP	Deoxyadenosine Triphosphate
CTP	Cytidine Triphosphate
dCTP	Deoxycytidine Triphosphate
GTP	Guanosine Triphosphate
dGTP	Deoxyguanosine Triphosphate
TTP	Thymidine Triphosphate
dTTP	Deoxythymidine Triphosphate
UTP	Uridine Triphosphate

dNTP	any Deoxyribonucleoside Triphosphate
dNDP	any Deoxyribonucleoside Diphosphate
dNMP	any Deoxyribonucleoside Monophosphate

Genes:

nrdA	Aerobic Ribonucleotide Reductase, Large Subunit
nrdB	Aerobic Ribonucleotide Reductase, Small Subunit
nrdD	Anaerobic Ribonucleotide Reductase
nrdG	Anaerobic Ribonucleotide Reductase, Activase
td	Thymidylate Synthase

E. coli	*Escherichia coli*
B. subtilis	*Bacillus subtilis*

Introduction

Concerning phages

This thesis is largely concerned with Phages, and from its pages a reader may discover much of their character and a little of their history. Bacteriophages, viruses that infect and kill bacteria (Greek, *phage* - "to eat"), have been studied and used as tools in molecular genetics and microbiology ever since they were discovered in the beginning of the 1900:s. They have aided researchers in the elucidation of many of the fundamental molecular mechanisms of life, e.g. the confirmation that DNA is the hereditary material (Hershey and Chase, 1952), the finding of restriction and modification of DNA (Luria and Human, 1952), the structure of DNA (Watson and Crick, 1953) the definition of the gene structure (Benzer, 1955), the identification of mRNA as the messenger of DNA (Brenner et al., 1961) and the definition of the genetic code (Crick et al., 1961). At the time of their discovery all phages were generally thought of as identical. Now we know that there exist a vast number of different bacteriophages that are genetically and morphologically distinct. Traditionally, phages have been categorised according to their morphology and host range and it is not until now, with the increasing number of complete phage genomes sequenced, that comparisons can be made with regards to the genetic history and relationships of phages.

T-even bacteriophages

Taxonomically the T-even bacteriophages belong to the family *Myoviridae* (Phages with contractile tails) among *dsDNA phages* (International Committee on Taxonomy of Viruses, www.ncbi.nlm.nih.gov/ICTVdb/). Morphologically they are recognised by their rather complex structure with an icosahedral head and a long contractile tail ended with six tail fibres (Fig. 1). The exact origin of the T-even bacteriophages is somewhat uncertain, but they have been traced back to "the Phage Group" in the lab of Max Delbrück in the 1940:s (Abedon, 2000). Delbrück and colleagues were concerned with the

problem that all people working with bacteriophages at that time had their own, more or less isogenic, phage strains and it was very hard to compare results. This problem originates in the early belief that all bacteriophages were identical. Therefore the Phage group made a collection of seven lytic phages that infected the routinely used intestinal bacterium *Escherichia coli B*, and that they found were easy to work with at 37°C. These phages were called T1-T7 (T for type) (Demerec and Fano, 1945). Out of these, the T-even phages, T2, T4 and T6 were found to be morphologically, antigenically and genetically very similar, and proved to be very useful for biochemical and genetic studies. T4 and T6 probably originates from a mixture of phages supplied to the Phage Group by Dr. Tony L. Rakieten (Demerec and Fano, 1945), and were most likely isolated from sewage (Abedon, 2000). T2 dates back to about 1927 and Dr. Jacques Bronfenbrennen. It was probably isolated from faeces and has been known by a number of different names (P28, α, TI, γ, PC). Today there exists (at least) two commonly used "sub-strains" of T2, T2L and T2H, the former is likely an isolate of γ from Salvador Luria and the latter an isolate by Hersey and coworkers (Abedon, 2000). Of the other reference phages from Delbrück´s lab, T1 and T5 are related and belong to *Siphoviridae* (Phages with long non-contractile tails) and T3 and T7 are related and belong to *Podoviridae* (Phages with short tails).

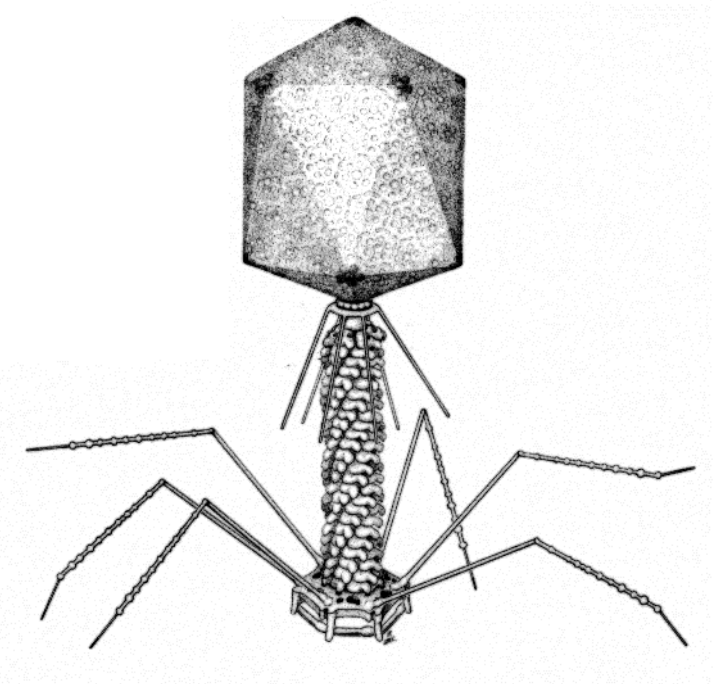


Figure 1. Bacteriophage T4. Figure adapted from (Eiserling and Black, 1994) and republished with permission of American Society for Microbiology; permission conveyed through Copyright Clearance Center, Inc.

Today there is a vast number of isolates of bacteriophages with morphological or genetic similarity to the T-even phages, most of them isolated from faeces from patients or from sewage. For clarity I will throughout this thesis refer to T2, T4 and T6 as the T-even phages and to later isolates as T-even-*like* phages. The practise of classifying phages morphologically has led to some confusion when genetic relationships have begun to be revealed. "T4-like phages" are classically phages that morphologically resemble T4. However, this class includes a very wide range of phages with different bacterial hosts and very little genetic similarity. With increasing amounts of sequencing data being collected, subgroups of the T4-like phages have been proposed. The phages closest related to T4, including T2 and T6, are frequently called T-even phages (although I prefer to separate the "original" T-evens and T-even-*like* phages). More distantly related phages have been named pseudo-T-evens and schizo-T-evens accordingly (Desplats et al., 2002; Monod et al., 1997; Tetart et al., 2001). Comparative studies of the additional genome sequences now being completed for phages related to T4 will likely shed more light on the relationships and phylogenetic history of this group of phages.

Bacteriophage T4

Most of what is known about the T-even, and T-even-like bacteriophages has come from studies of T4. As mentioned earlier, the functions of many genetic mechanisms were originally elucidated in T4 and several of its proteins are used routinely in molecular biology today, due to the vast biochemical knowledge we have about them. T4 has a genome comprising 168903 base pairs (bp), one of the largest known for bacteriophages. It is now completely sequenced (Kutter et al., 1993) and contains nearly 300 probable genes, 289 encoding proteins, 8 tRNAs and two other small RNA species of unknown function (Miller et al., 2003b). Half of the proposed genes still have no described function and no homology to any genes in GenBank. Only 62 genes are essential under standard laboratory conditions while the rest can be deleted without abolishing phage growth (Miller et al., 2003b). However, these "non-essential" genes most likely provide important functions to the phage

during growth in more natural conditions or during transition between different growth conditions or different hosts.

The T4 DNA contains modified bases, probably to avoid degradation by host restriction systems and to make a distinction between its own DNA and the host DNA that is rapidly degraded by phage endonucleases early in infection (reviewed in (Carlson et al., 1994)). Instead of cytosine, T4 uses 5-hydroxymethylcytosine (HmC) that is further almost completely glucosylated postreplicationally. T4 DNA is also highly enriched in A-T base pairs (65.5%) (Miller et al., 2003b) compared to *E. coli* (49%) (De Ley, 1970), and the phage genes contain a codon usage bias which is complemented for by the eight tRNAs encoded by the phage which are normally only scarcely expressed by the host (Mosig, 1994). The modification of cytosines has also been proposed to increase the double strand stability of the A-T rich genome (Miller et al., 2003b).

The majority of past and present work on T4 and the T-even bacteriophages is done with *E. coli* as host, but if this is the preferred wild type host is not known. T4 has been shown to infect several different enteric bacteria such as *Klebsiella*, *Shigella*, *Salmonella* and *Proteus* (Ackermann and Krisch, 1997; Dawes, 1976) and it can replicate in a number of different gram-negative bacteria even though infection is impaired (Wais and Goldberg, 1969). It should be kept in mind that there are a vast number of potential host bacteria in the mammalian gut that have avoided characterisation due to problems of fulfilling their specific growth requirements when performing lab isolations from faeces.

Overview of T4 development

The lifecycle of bacteriophage T4 is outlined in figure 2. Infection by T4 and its close relatives is very efficient approaching 100% plating efficiency under standard laboratory conditions (Goldberg et al., 1994). Adsorption to the bacterial cell wall is through binding to receptors on the cell. Initially the distal tips of the six tail fibers bind cooperatively but reversibly to the lipopolysaccharide (LPS) of the cell surface ((Stent and Wollman, 1952) and reviewed in (Goldberg et al., 1994)). Reversible binding by the tail fibers allow the phage to “wander” over the cell

surface until it reaches a baseplate recognition site. A variety of major outer membrane proteins of *E. coli* can work as receptors for permanent phage binding (Eddy, 1992; Schwarz et al., 1983). The hyper variable tips of the tail fibers determine the host range of the T-even-like phages (Beckendorf, 1973; Beckendorf et al., 1973). After receptor recognition, the base plate at the end of the phage tail makes contact with the outer membrane and the cell wall, the tail is contracted and gp5 (a lysosyme located in the base plate) makes a hole through which the phage DNA is injected (Kao and McClain, 1980).

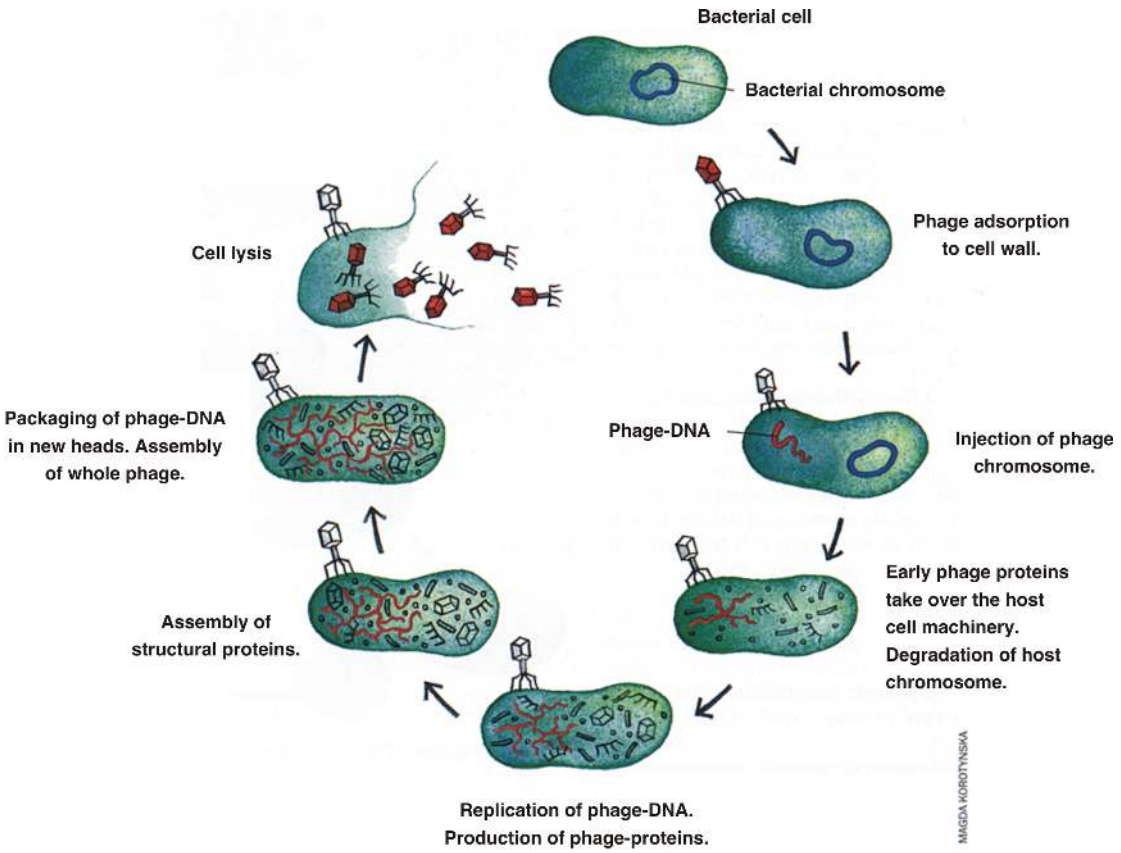


Figure 2. Bacteriophage T4 infection cycle. Picture modified from (Carlson, 2000) and republished with permission from Magdalena Korotynska.

Once inside the cell, the phage DNA is recognised by the host RNA polymerase and early genes are transcribed (see T4 transcriptional regulation). Among early genes are found genes for the phage take over of the cell machinery, genes for production of DNA precursors for phage DNA synthesis and genes for regulation of middle and late gene expression. During early gene expression phage endonucleases are produced that degrade the host chromosome, thereby completely stopping

bacterial gene expression (reviewed in (Carlson et al., 1994)). Early genes are followed by expression of middle genes that predominantly produces regulatory proteins, and then by late genes coding for structural phage proteins for building new phage particles and loading of the phage chromosome into the new heads. Assembly of the structural proteins into new heads, tails and other structures occurs mainly by auto assembly guided by chaperones such as GroEL from the host and gp31 from the phage (Keppel et al., 1990). The final stage is the packaging of the phage genomes into the heads by terminase enzymes. This takes place by head-full packaging of slightly more than one genome equivalent generating circularly permuted chromosomes with terminal redundancy (reviewed in (Murialdo, 1991)).

Transcriptional regulation

Under laboratory conditions a T4 infection in logarithmically growing *E. coli* in rich medium at 37°C takes about 25 minutes from injection of the phage DNA to lysis of the host cell and release of typically 200-300 progeny phage. This rapid production of new phage is dependent on a temporally regulated infection cycle (Koch and Hershey, 1959). T4 gene regulation shows an ordered expression and repression of genes during the infection. As mentioned above, three different types of genes can be distinguished, early, middle and late. Regulation of gene expression is mainly accomplished at the transcriptional level. All transcription of phage genes is done by the host core RNA polymerase but with a changing set of modifications throughout the infection cycle (see below).

Early genes are transcribed immediately after the phage has injected its DNA into the cell. The early promoters differ from *E. coli* σ^{70}-promoters but still contain a -35 box and a -10 box (Wilkens and Ruger, 1994) and can compete with the host promoters for RNA polymerase binding. A protein called Alt is present in 30-50 copies in the phage head and injected into the cell together with the phage DNA (Horvitz, 1974a; Horvitz, 1974b). It catalyses ADP-ribosylation of one of the subunits of the RNA polymerase (Rohrer et al., 1975). This modification seems to strengthen the activation of transcription at T4 early promoters compared

to *E. coli* promoters (Koch et al., 1995; Wilkens and Ruger, 1996; Wilkens et al., 1997).

In contrast to early transcription, middle and late transcription are dependent upon synthesis of phage proteins. Among the early T4 genes produced are four important transcriptional regulators, ModA, ModB, AsiA and MotA. ModA and B are two additional ADP-ribosylating enzymes that further modify the subunits of the core RNA polymerase (Goff, 1974; Mosig et al., 1998). MotA is an activator of transcription from middle promoters (Guild et al., 1988). Middle promoters lack the –35 box and instead have an "extended" -10 box (Hughes and Mathee, 1998; Stitt and Hinton, 1994). A sequence in front of this extended -10 box, about 30 base pairs upstream of the transcription initiation site, is recognised and bound by MotA. AsiA is an anti-σ^{70} factor that forms hetero-dimers with σ^{70}, abolishing the -35-box recognition and thereby lowering the activation of transcription at early promoters (reviewed in (Hughes and Mathee, 1998)). At the same time the modified RNA polymerase can interact with the MotA protein via contacts with AsiA, shifting transcription from early to middle promoters. In this way MotA, bound to the upstream site, effectively replace the -35 region in recognition of middle promoters (Hinton et al., 1996).

During middle transcription additional regulatory proteins for late gene expression are made. The key player in late transcription is the phage encoded σ-factor gp55 that binds to the RNA polymerase core instead of σ^{70}. Normally, *E. coli* σ^{70} binds stronger to the RNA polymerase than gp55 and will thus outcompete it. AsiA however, acting as a true anti-σ^{70} factor, weakens the interaction between σ^{70} and the core subunits of RNA polymerase and shifts the preference to gp55 binding (Williams et al., 1994). Late promoters, like middle promoters, lack a recognisable -35 box and instead only have a special T4 late -10 box (Christensen and Young, 1982). Late transcription is therefore also dependent on additional activator proteins, especially RNA polymerase binding protein gp33 and DNA polymerisation sliding clamp protein gp45, to direct the RNA polymerase to the correct promoters. No late expression takes place without DNA replication and this is regulated mainly via the actions of gp45 (Epstein et al., 1964). It is the active, ATP dependent, loading of the sliding clamp protein gp45 onto DNA that is thought to generate the

coupling of late transcription to DNA replication (Fu et al., 1996). Interaction between gp45 sliding along the DNA, and gp33 and gp55 bound to the RNA polymerase directs transcriptional initiation to the late promoters.

All T4 genes do not have a recognisable promoter and several are transcribed as operons from a single promoter (Miller et al., 2003b). There are also cases where late proteins are transcribed on early and middle transcripts but not expressed until late in infection due to repression by mRNA processing and translational control. One characterised mechanism to avoid translation of late genes from polycistronic early transcripts is the presence of an RNA hairpin in front of some late genes, that masks the translational initiation region at the start of these genes. The hairpin cannot form on transcripts from the correct, late promoter since the late promoter is located within the hairpin region (Macdonald et al., 1984; McPheeters et al., 1986). Such repression of translation of late genes expressed on extended earlier transcripts has been recognised for the homing endonuclease genes within the *nrdB* and *td* introns in T4 (Gott et al., 1988) see Paper I.

Nucleotide metabolism

Bacteriophage T4 devotes a large part of its genome to genes coding for enzymes used in nucleotide metabolism. All but one of the proteins needed for making the building blocks of phage DNA are encoded by the phage itself. The only host enzyme used is nucleoside diphosphate kinase needed for phosphorylation of dNDPs to dNTPs (Moen et al., 1988). Nucleotides for production of the phage genome can come from two different sources, *de novo* synthesis of dNTPs from ribonucleotides or reutilization of dNMPs from the degraded host genome (the salvage pathway), as shown in figure 3. Most of the dNTPs needed for the 200-300 new T4 genomes produced under optimal conditions comes from *de novo* synthesis. The complete degradation of the *E. coli* genome can provide nucleotides for less than 20 phage genomes. An exponentially growing *E. coli* cell may contain 3-4 genome equivalents (Bremer and Dennis, 1987) so the upper limit of phage genomes produced by the salvage pathway would be 80.

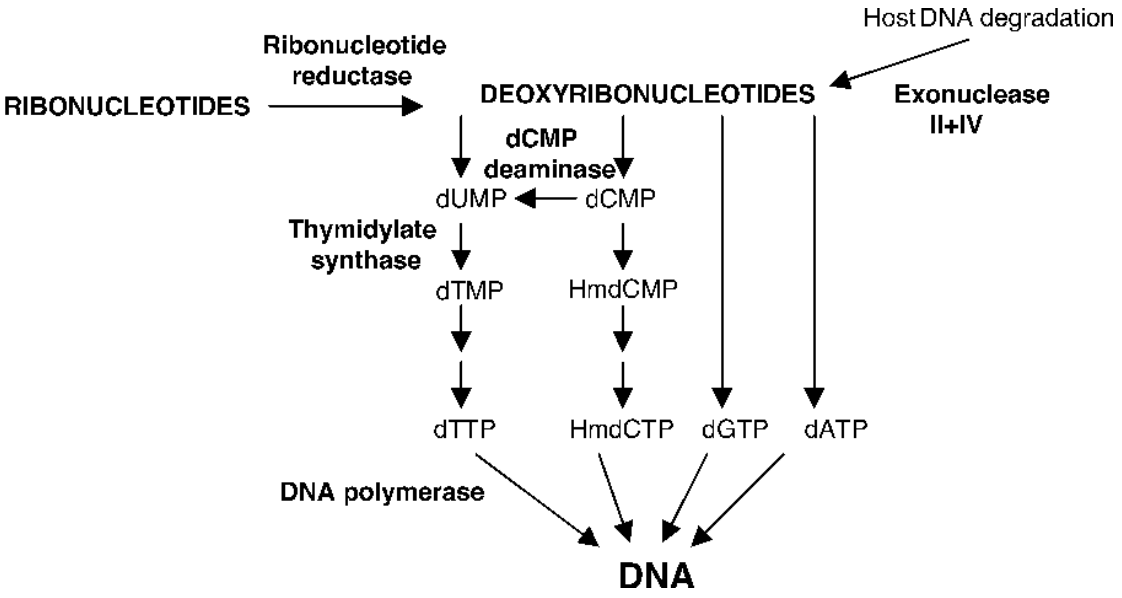


Figure 3. Nucleotide metabolism in bacteriophage T4.

Production of dNTPs by the *de novo* pathway has been proposed to occur by a large multienzyme system, the dNTP synthetase complex (Allen et al., 1983; Allen et al., 1980; Greenberg et al., 1994; Mathews, 1993a; Reddy et al., 1977). A 1.5 MDa complex shown to include most of the enzymes required for complete production of dNTPs from ribonucleotides has been isolated from T4-infected cells (Moen et al., 1988). The dNTP synthetase complex is also thought to be physically interacting with the replication machinery at the very site of replication, thereby channelling the newly synthesised dNTPs directly to the DNA polymerase (Mathews, 1993a; Mathews, 1993b; Wheeler et al., 1996). Production of ribonucleotide reductase, the enzyme catalysing the reduction of ribonucleotides to deoxyribonucleotides, appears to be the limiting factor in assembly and initiation of dNTP synthesis by the enzyme complex (Chiu et al., 1982; Tseng et al., 1990). Bacteriophage T4 has two different ribonucleotide reductases, one functioning during aerobic growth and the other during anaerobic growth. The aerobic enzyme is built up by two different homodimeric subunits R1 (*nrdA*) and R2 (*nrdB*) (Berglund, 1975), while the anaerobic enzyme functions as a single homodimer (encoded by *nrdD*) (Young et al., 1994) that also requires an activator protein encoded by *nrdG*.

DNA replication

T4 encodes all but one of the proteins needed for replicating its genome. The only host enzyme used in replication is the RNA polymerase in order to synthesise RNA primers for initiation of leading strand replication (Luder and Mosig, 1982). DNA replication in T4 is initiated in two different ways. The first replication fork is initiated at one of four major origins of replication (ori) oriA, oriE, oriF and oriG (Kreuzer and Alberts, 1985; Kreuzer and Morrical, 1994; Menkens and Kreuzer, 1988). After the first round of replication, initiation of new replication forks from the oris is repressed and instead occurs via recombination intermediates initiated from the DNA ends of previously replicated chromosomes (Dannenberg and Mosig, 1983; Luder and Mosig, 1982). The major pathway for this recombination-dependent replication is through strand invasion by single-stranded, 3´ chromosomal ends that lead to new replication forks (reviewed in (Mosig, 1998)). This is a very efficient way of solving the problem with replication of the ends of the linear chromosomes and results in branched concatemers of T4 genomes that are later used for packaging into the heads. The circularly permuted chromosomes of T4 makes this recombination dependent replication very efficient and the amount of replication during infection is ten times higher than in the uninfected cell (Werner, 1968). Replication is terminated when the phage protein gp2 is expressed which binds to free chromosome ends and represses new recombination initiation events (Lipinska et al., 1989).

Phage assembly and Lysis

The final stages of T4 infection are the assembly of new phage particles and lysis of the host cell. Almost half of the genes in T4 codes for phage structural proteins or proteins involved in the assembly of the structures (Miller et al., 2003b). There are 24 proteins involved in head morphogenesis (reviewed in (Black et al., 1994)). The head is made up of the shell proteins gp23 and gp24 that are assembled on a scaffold made of proteins gp21 and gp22. When the scaffold is covered gp21 proteolytically cleaves the other proteins, degrading the scaffold thus forming a prohead that is ready for DNA packaging. DNA packaging is initiated by the endonucleolytic generation of packable chromosome ends

by the terminase complex (gp16, 17, 17', 17'') (Franklin et al., 1998; Franklin and Mosig, 1996). The DNA is loaded into the head by a head full mechanism that requires energy by ATP hydrolysis. Approximately 3% more than a complete genome is loaded generating the terminal redundancy of the chromosome ends (Streisinger et al., 1964). When the head is full the terminase complex cleaves the genome and the head assembly is completed by addition of gp13, gp14 and Wac that make up the "whiskers" and the attachment site for the tail.

The tail is made up of a baseplate and a two-layer cylinder. The inner layer of the tail cylinder (the tail tube) is made up of gp19 while the outer layer (the tail sheet) is made up of gp18 (Dickson, 1974; King and Mykolajewycz, 1973). The inside of the tail contains a passage for the DNA upon injection (Smith and Aebi, 1976). Assembly of the head and the tail occurs independently and after completion they are joined together in a reaction that can occur spontaneously *in vitro* (Coombs and Arisaka, 1994). The baseplate is a complex structure made up of 15 different proteins (reviewed within (Miller et al., 2003b)) among others gp12 making up the short tail fibers that anchor the phage irreversibly to the cell upon infection and gp5 that forms a needle structure that punctures through the outer cell membrane upon tail contraction (Kanamaru et al., 2002) and that also contain a lysosyme-activity that degrades a hole in the cell wall for passage into the cell. The baseplate is also the attachment site for the long tail fibers that are the primary adsorption organelles for the phage (Kellenberger et al., 1965).

The long tail fibers consist of proteins gp34, gp35, gp36 and gp37 with gp34 and gp37 forming the "legs" and gp35 and gp36 forming the "joints" (reviewed in (Wood et al., 1994)). The C-terminal part of gp37 (forming the distal tips of the tail fibers) is hypervariable among T-even-like phages and determines the host range of the phage by receptor recognition (Hashemolhosseini et al., 1994a; Hashemolhosseini et al., 1994b; Montag et al., 1990).

After the assembly of the phage particles the host cell is lysed via the action of the T4 lysosyme gpe and the T4 holin gpt (Mukai et al., 1967; Streisinger et al., 1961). A pore is generated in the inner membrane by gpt through which gpe can migrate and attack the cell wall from within and

the cell is disrupted, releasing the new phage particles. If additional phage attack an infected cell lysis is delayed (lysis inhibition) and infection can be prolonged for up to several hours (reviewed in (Abedon, 1994)). It is not known what signals this delay in gpt synthesis.

T-even-like phage genomics

As of the 29 of July 2004 there were 236 phage genome sequences registered in GenBank, 31 of which belong to the Myoviridae and 6 of which are T4-like phages. In addition, the complete genome sequences of Pseudo-T-even phages RB43 and PHG31 have been determined by the Tulane Phage Sequencing Group but not yet deposited in GenBank. The emergence of whole genome sequences of T4-related phages opens up for a more extensive analysis of the phylogenetic history and evolution of these phages.

Table1. Genomic data of T4-related phages

Phage (bacterial host)	Class	Genome size (bp)	tRNA	HMC	% G+C
T4 (*E. coli*)	T-even	168903	8	+	35,3
RB69 (*E. coli*)	T-even-like	167903	2	?	37,7
RB43 (*E. coli*)	Pseudo-T-even	180500	1	-	43,2
RB49 (*E. coli*)	Pseudo-T-even	164018	0	-	40,4
44RR (*Aeromonas sp.*)	Pseudo-T-even	173591	16	?	43,9
PHG31 (*Aeromonas sp.*)	Pseudo-T-even	172965	15	?	43,9
KVP40 (*Vibrio sp.*)	Schizo-T-even	244835	29	-	42,6
AehI (*Aeromonas sp.*)	Schizo-T-even	233234	21	?	42,8

Based on sequence similarity the T4-type phages can be divided into different groups (Tetart et al., 2001). The original T-even phages T2, T4 and T6 are very closely related at the primary sequence level as was initially determined by genetic mapping (Russell, 1967) and by DNA-duplex studies (Kim and Davidson, 1974). T-even-like phages typically share 80-90% DNA sequence identity with T4 (Cowie et al., 1971; Loayza et al., 1991; McPheeters et al., 1988; Sandegren and Sjöberg, 2004; Selick et al., 1993; Tetart et al., 2001). Pseudo-T-even phages have little DNA sequence similarity to T4 but they typically have 50-80% amino acid sequence identity in homologous proteins and the schizo-T-evens have around 50% amino acid identity to T4 (Desplats et al., 2002; Monod et al.,

1997; Tetart et al., 2001). Exo-T-evens such as the cyanophage S-PM2 only moderately resembles the T4-morphology (Desplats et al., 2002) and in those proteins where homology to T4 can be detected there is about 30% identity at the protein level indicating that they are only distantly related to the T-even phages (Hambly et al., 2001). Phages of all these groups appear to have similar genome organisation where groups of homologous genes are found in the same order on the chromosome (Hambly et al., 2001; Matsuzaki et al., 1999; Russell, 1967). However, even among closely related phages there are reorganisations where genes or blocks of genes have been inserted or deleted in the genomes in between otherwise conserved genes (Loayza et al., 1991; Miller and Jozwik, 1990; Repoila et al., 1994; Sandegren and Sjöberg, 2004; Selick et al., 1993). The region encoding the structural phage proteins of the head and tail appears to be the most conserved among the T4-type phages (Hambly et al., 2001; Matsuzaki et al., 1999; Miller et al., 2003b; Monod et al., 1997; Tetart et al., 2001) with the order of genes 18-23 being conserved even between T4 and the exo-T-even phage S-PM2 (Hambly et al., 2001). This has led to the suggestion that the T4-type phages have a set of essential genes that have coevolved together (Desplats et al., 2002) and that the optional sequences in between such conserved regions have been included via gene exchange between distantly related T4-type phages (Repoila et al., 1994). This is further corroborated by the fact that a large percentage of the genes in every new T4-type phage genome sequenced have no homologs to the other members of this group or to any other sequence in the databases (30% RB49 (Desplats et al., 2002), 65% KVP40 (Miller et al., 2003b).

The size of the phage head is what determines how much DNA it can contain and in line with this the Schizo-T-even phages have larger genomes corresponding to their slightly elongated heads (Tetart et al., 2001). Another feature that vary among the T4-type phages is the use of modified bases. Several of the pseudo-T-even and schizo-T-even phages lack the genes for dCMP hydroxymethylation and DNA glucosylation and in line with this their DNA is also cleaved by restriction endonucleases that are not able to cleave modified T4 DNA (Matsuzaki et al., 1992; Miller et al., 2003b; Monod et al., 1997). The number of phage encoded tRNAs also differ from none in RB49 to 29 in KVP40 (Table 1) probably reflecting different needs for complementing tRNAs for translation in different phages and bacterial hosts.

Another interesting fact that has emerged from whole genome sequence analysis is the differences in transcriptional regulation that apparently exists between T4, pseudo-T-even phage RB49 and schizo-T-even phage KVP40 (Desplats et al., 2002; Miller et al., 2003b). The two latter phages appear to use early promoters very similar to *E. coli*-like σ^{70} promoters instead of the extended early promoters used by T4. Furthermore, there are no homologs of the T4 *motA*, *asiA*, *modA* or *modB* genes in RB49 (Desplats et al., 2002) or KVP40 (Miller et al., 2003b) and no middle promoters are found throughout their genomes indicating that this mode of transcriptional regulation is absent in these phages. In contrast, homologs of all proteins important for late transcription in T4 are present in RB49 and KVP40. The variable occurrence in related phages of genetic systems central to T4 further strengthens the view that differences between the T4-type phages can occur via exchange of whole genetic modules (Botstein and Herskowitz, 1974; Repoila et al., 1994).

Group I introns

Splicing is the post-transcriptional (RNA splicing), and in some cases post-translational (protein splicing) removal of intervening, non-coding sequences from within a gene. Splicing at the RNA level can occur by different mechanisms, Group I-, Group II-, Group III- and nuclear mRNA introns all splice via two consecutive transesterification reactions closely involving the intron RNA while splicing of nuclear tRNA introns and archaeal introns occur via the action of endonuclease proteins followed by ligation. This division of introns into different groups is based both on their mechanism of splicing and on conserved intron motifs. Group I and group II introns are also called self-splicing introns because many of these introns can catalyse their own excision *in vitro* without any help from proteins.

The first self-splicing introns found were the group I introns. These were shown by Cech et al. to be able to splice *in vitro* in the total absence of proteins (Cech et al., 1981). All group I introns share conserved structural motifs (Davies et al., 1982; Michel et al., 1982) and utilise the same catalytic mechanism.

Splicing mechanism

A unifying feature of group I introns is the splicing reaction that takes place through two consecutive transesterification reactions (Fig. 4). The first step is catalysed by the binding of an exogenous guanosine nucleoside or nucleotide at a conserved guanosine-binding site in helix P7 in the intron core (Michel et al., 1989a) This free guanosine initiates the splicing reaction by a nucleophilic attack of its 3' hydroxyl group on the phosphorus atom at the 5' splice site (Cech et al., 1981). Activation of the cleavage is dependent upon Mg^{2+} or Mn^{2+} ions that are positioned at the catalytic site (Grosshans and Cech, 1989; Piccirilli et al., 1993; Steitz and Steitz, 1993; Weinstein et al., 1997). The guanosine forms a 3', 5' phosphodiester bond to the 5' end of the intron, and the 3' hydroxyl of the last nucleotide in the upstream exon is free to make a nucleophilic attack on the 3' splice site, ligating the two exons together and releasing the intron (Price, 1987). This mechanism of two consecutive transesterifications is similar in group II and nuclear mRNA splicing, but in those cases it is the 2' hydroxyl group of an internal adenosine in the intron that initiates the first nucleophilic attack (reviewed in (Cech, 1990)).

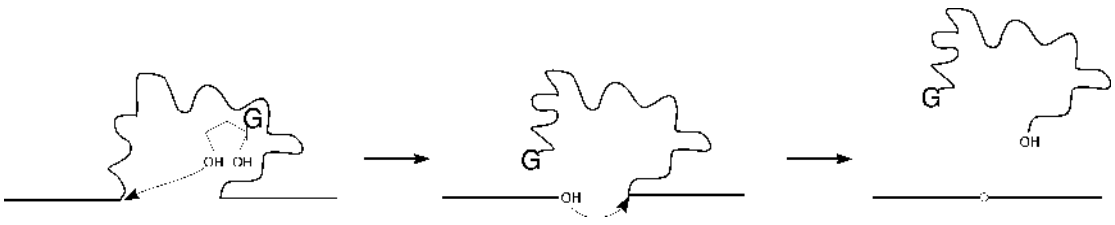


Figure 4. Splicing mechanism for group I introns. Straight lines are exons, wavy lines are introns. Drawn essentially as in (Cech, 1990). See text for details.

Structure

All group I introns share several regions of conserved nucleotides that build up the core region of the intron shown in the secondary structure scheme (Fig. 5) (Michel et al., 1982). The secondary structure model contains nine paired regions (Davies et al., 1982; Michel et al., 1982) that are, with the exception of P2, found in all group I introns so far. Additional sequences, making up extra, paired regions frequently exist and group I introns are divided into subgroups according to the presence of such additional regions (Michel and Westhof, 1990).

Although the secondary structures of large RNAs are fairly straightforward to determine via free-energy minimizations and covariation analysis, tertiary structures have been much harder to elucidate. This is primarily because of the difficulties of forming good crystals of the negatively charged RNA for X-ray crystallographic studies. However, more and more knowledge has been gathered about large RNA structures with group I introns as general favourites. Michel and Westhof (1990) constructed a 3-dimensional group I intron model based on sequence covariations of the 87 group I introns known at the time (Michel and Westhof, 1990). They predicted a tightly folded structure with stacked helices held together by a number of tertiary interactions (see below).

In 1996 Cate et al. presented the structure of the P4-P6 domain of the *Tetrahymena* rRNA intron, at 2.8 Å (Cate et al., 1996a; Cate et al., 1996b). The P4-P6 region has been shown to fold separately from the rest of the intron (Murphy and Cech, 1993). The crystal structure of this domain confirmed many of the ideas earlier proposed by Michel and Westhof and also revealed novel structural motifs for the folding of large catalytic RNAs. A crystal structure of almost the whole *Tetrahymena* intron at 5 Å resolution was later solved (Golden et al., 1998). Even though this structure is not at atomic resolution it shows the predicted tightly folded globular structure with two sets (P3-P9 and P4-P6) of coaxially stacked helices packed against each other with P3-P9 forming a bent pseudocontinous helix wrapped around the P4-P6 domain (Golden et al., 1998). The P1 helix is missing in the structure but the positioning of P1 with the 5' splice site at the guanosine binding site within P7 as in the

Michel and Westhof model fits well with the crystallographic data and does not require any reorganisation of the domains (Golden et al., 1998).

The tight packing of the RNA helices is mediated both by binding of positively charged metal ions, mainly Mg^{2+}, that compensates for the negatively charged phosphate backbone of the RNA (Celander and Cech, 1991; Christian and Yarus, 1993; Downs and Cech, 1996; Zarrinkar and Williamson, 1996) and by a number of tertiary interactions between different regions of the intron, some of which are depicted in figure 5B. A common motif of large RNA structures is the GNRA tetraloop (Costa and Michel, 1995; Michel et al., 1989b; Michel and Westhof, 1990). Several such motifs are found in group I introns where the terminal bases of the loop makes a tertiary contact with the minor groove of another helix such as interactions P11, L5-P6 and L2-P8 (see Fig 5B) (Cate et al., 1996a; Costa and Michel, 1995; Jaeger et al., 1991; Michel and Westhof, 1990; Murphy and Cech, 1994). Tetraloops are frequently found in optional regions of the different subgroups of group I introns and several different folds appear to have evolved to stabilise the conserved core of the intron (Cate et al., 1996a; Michel and Westhof, 1990). Another important set of interactions between the two intron domains are the triple helix regions that are formed at the junctions between P6-P7 (J6/7) and P3-P4 (J3/4) and that are thought to help orient the P3-P9 and P4-P6 domains relative to each other (Fig. 5B) (Downs and Cech, 1994; Michel et al., 1990; Zarrinkar and Williamson, 1996).

The core of the intron is made up of the contact region between the two separately folding P4-P6 and P3-P9 domains with the guanosine binding site located in the P7 helix (Michel et al., 1989a) (see Fig. 5). Binding of the guanosine is through hydrogen bonding to an invariant guanine residue and is facilitated via a special stacking interaction of an unpaired base in the P7 helix (Ehrenman et al., 1989; Michel et al., 1989a). Alignment of the 5' and 3' splice sites with the guanosine binding site for the catalytic step is facilitated via base pairing of the internal guide sequence (IGS) that is made up by sequences in the P1 loop and sequences around the 3' splice site (Davies et al., 1982). The P9.0 interaction (Burke, 1989; Burke et al., 1990; Michel et al., 1989a; Michel and Westhof, 1990) brings together the 3' splice site and the guanosine binding site, determining the position of the 3' splice site, and together with the P10 interaction (Michel

et al., 1989a) aligns the two splice sites in the correct orientation (Burke et al., 1990) (Fig. 5C). The P1 and P2 helices together with the 3' end of the intron is positioned in a cleft created between the P4-P6 and P3-P9 domains (Golden et al., 1998; Michel and Westhof, 1990; Wang et al., 1993).

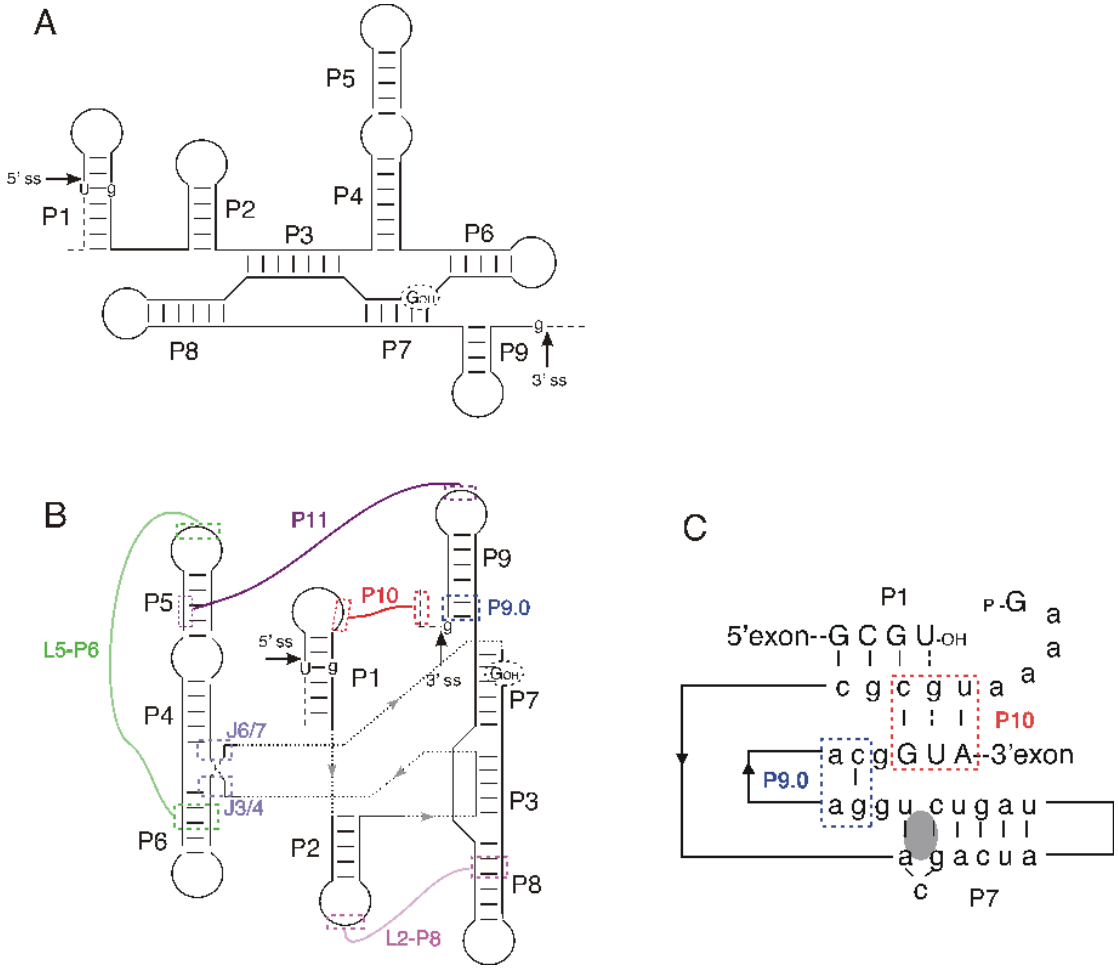


Figure 5 A) The "old" secondary structural representation showing the paired regions conserved among group I introns. Paired regions are numbered P1-P9 and the splice-sites are indicated with arrows. Exon sequences involved in structural contacts around the splice-sites are drawn as dotted lines. The G-binding site in P7 is indicated. Drawn according to (Cech, 1990). **B)** The "new" secondary structure representation (Cech et al., 1994) with tertiary interactions depicted by coloured boxes and lines. **C**) Folding of the T4 *nrdB* intron sequence at the catalytic site just after the first transesterification step (Drawn essentially as the *Tetrahymena* intron in (Burke et al., 1990)). P9.0 and P10 are boxed and the guanosine-binding site is shaded. Exon sequences and the bound guanosine are in upper case and intron sequences in lower case.

Proteins involved in group I intron splicing

Although many group I introns are capable of folding and splicing *in vitro* there are a number of proteins that have been shown to make splicing more efficient *in vivo*, mainly by mediating correct and rapid folding of the RNA. Large RNAs such as group I introns face two folding problems, they may get kinetically trapped in missfolded conformations and not reach the correct catalytic fold or the correct fold may be thermodynamically unstable once it is reached (Herschlag, 1995).

Many group I and group II introns encode proteins on open reading frames within the intron (see below). Several such proteins have been shown to possess maturase activity, stabilising the active fold of the respective intron (reviewed in (Lambowitz et al., 1999)). For group I introns, maturases have predominantly been found associated with yeast group I introns (Lambowitz et al., 1999; Solem et al., 2002). However, among the second type of catalytic introns, the group II introns, maturases encoded within the introns are much more common.
Apart from the intron-associated proteins there are several nuclear or host encoded proteins shown to enhance splicing by stabilising the active fold of the introns. *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase CYT-18 (Collins and Lambowitz, 1985; Mannella et al., 1979; Wallweber et al., 1997) and the CBP2 protein of *Saccharomyces* yeasts (McGraw and Tzagoloff, 1983) have both been shown to promote splicing of several endogenous group I introns as well as group I introns from other organisms. CYT-18 can promote splicing of several group I introns (Guo and Lambowitz, 1992; Mohr et al., 1994; Mohr et al., 1992) by binding to the P4-P6 domain and provide a scaffold for stabilising the interactions needed for assembly of the P3-P9 domain (Caprara et al., 1996a; Caprara et al., 1996b; Caprara et al., 2001; Saldanha et al., 1996; Saldanha et al., 1995; Waldsich et al., 2002). CBP2 has instead been shown to bind on the opposite side of the intron, close to the P1 binding site, and to stabilise the catalytically active conformation (Weeks and Cech, 1995a; Weeks and Cech, 1995b; Weeks and Cech, 1996).

Other proteins enhance splicing of group I introns by resolving missfolded conformations thereby aiding the RNA in folding correctly. Such proteins that only help RNA folding and are not needed once the active conformation is reached are called RNA chaperones (Herschlag,

1995). The best characterised RNA chaperones are *E. coli* proteins StpA and ribosomal protein S12. Both proteins bind RNA non-specifically and have been shown to enhance splicing of the T4 *td* intron *in vitro* (Coetzee et al., 1994; Zhang et al., 1995). *In vivo* studies confirm that StpA and S12 facilitate splicing by resolving tertiary contacts thereby enabling the intron to refold into its active conformation (Clodi et al., 1999; Semrad and Schroeder, 1998; Waldsich et al., 2002). The list of proteins with RNA chaperone activity on group I introns is likely to become longer since assays for their identification both *in vitro* and *in vivo* now exist (Clodi et al., 1999; Herschlag et al., 1994; Zhang et al., 1995).

Group I intron distribution

Group I introns are widespread being found in both Eukaryotes and Bacteria as well as in both eukaryotic and prokaryotic viruses while so far there are no group I introns found in Archaea. They have been found in mitochondrial and chloroplast genomes and in nuclear genomes of unicellular eukaryotes and in all three major RNA species: mRNA, rRNA and tRNA. Although widespread the distribution of group I introns is irregular, with differences in occurrence between closely related species. The fact that group I introns can be found in a specific gene of one organism and be absent from the same gene in closely related species have raised the question whether these differences have occurred through differential loss or gain of introns. The finding that several of these introns are mobile (see below) lends strong support to the latter theory.

Homing endonucleases and intron mobility

A large number of group I introns have been shown to be mobile. This was first described for the *omega* intron in the large-subunit rRNA gene of yeast mitochondria (Jacquier and Dujon, 1985). The *omega* intron was found to copy itself into intronless copies of the gene, converting them to intron-containing, and the mobility was shown to be fully dependent on the expression of an endonuclease encoded in an open reading frame within the intron (Jacquier and Dujon, 1985; Macreadie et al., 1985). Accordingly, introns lacking such a homing endonuclease gene (HEG) or

having a defective HEG are not mobile. The transfer of a HEG-containing intron always occur to the same sequence in the recipient gene as the intron-insertion site in the donor allele, therefore this process has been called "homing" (Dujon, 1989). Homing endonucleases have been found associated with both group I and group II introns as well as with inteins (protein splicing elements that are removed post-replicationally) and as freestanding genes between other genes.

Homing is initiated by the cleavage by the homing endonuclease. Most group I intron-encoded homing endonucleases only cleave intronless alleles while intron-containing alleles are immune to cleavage since the intron interrupts the recognition site of the homing endonuclease. After cleavage, the freestanding or group I intron encoded endonuclease appears to take little or no part in the rest of the homing process (Bell-Pedersen et al., 1990; Eddy and Gold, 1992). Instead it is the recombination/repair machinery of the cell that via different pathways (Mueller et al., 1996) use an intron-containing copy of the gene as template when repairing the cut, thus copying the intron/HEG into the cognate site of the repaired allele. In contrast, group II intron homing utilises a different mechanism called “retrohoming” (Curcio and Belfort, 1996) where the excised intron reverse-splices into the cut allele mediated by a maturase activity of the homing endonuclease protein followed by reverse transcription into DNA.

In figure 6 the proposed major pathway for freestanding and group I intron homing in bacteriophage T4 is depicted (Mosig, 1998). The cut DNA is partially degraded by exonucleases generating single-stranded 3’ ends (a) that can then invade the donor allele (b) and initiate replication-dependent copying of the ORF-containing intron to the repaired DNA-strand (c-d). The resolution of the double Holliday-junction intermediates (e) can produce different recombination cross-over products, only one of which is shown in figure 6 (f).

Homing endonucleases have been divided into four major families based upon structural and functional properties and named after their sequence motifs LAGLIDADG, GIY-YIG, H-N-H and His-Cys box (Belfort and Roberts, 1997). Although there are differences among the classes in the mechanism of action (some work as dimers, others in association with

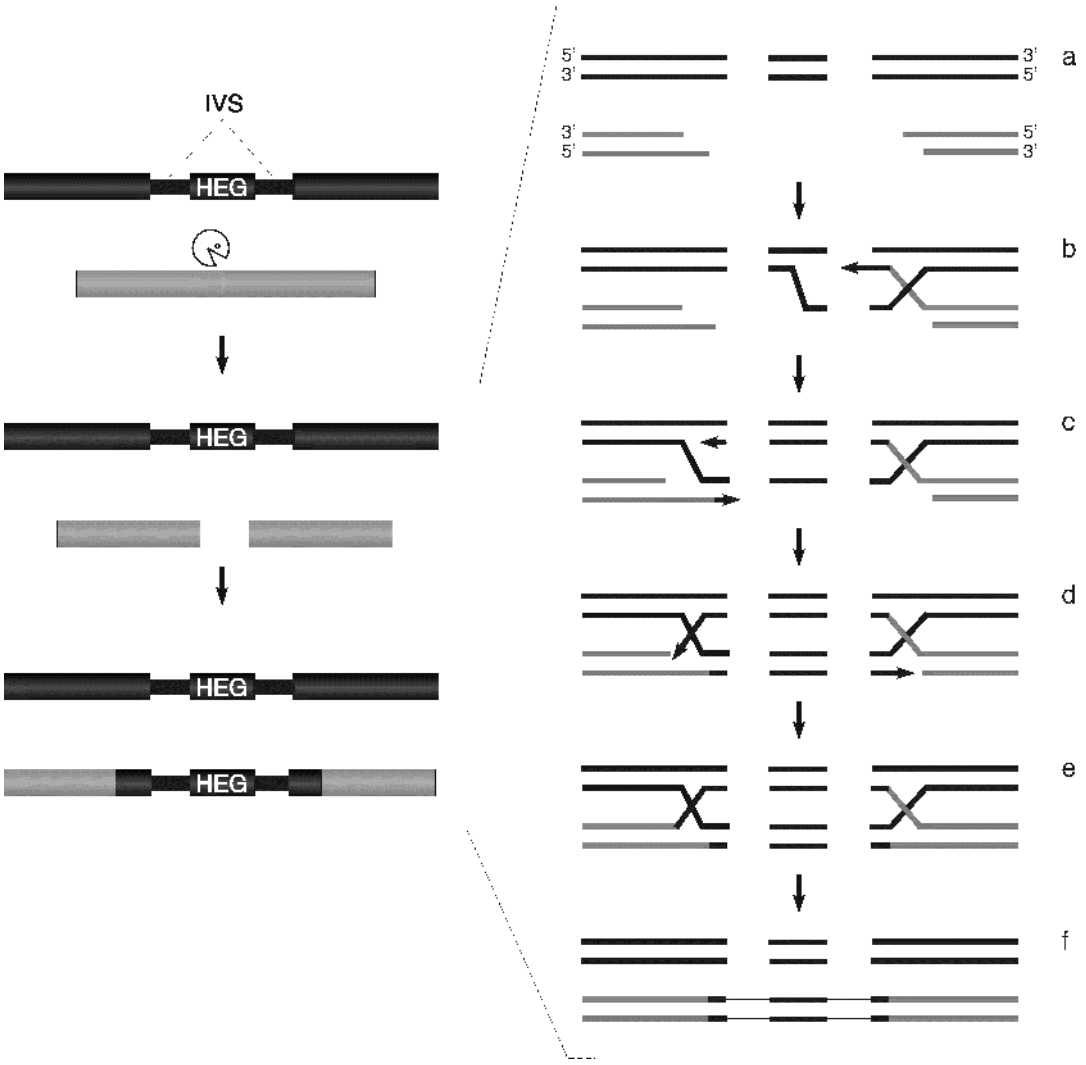


Figure 6. The Double Strand Break Repair mechanism for intron homing drawn essentially as in (Mosig, 1998). Black lines indicate donor sequences, grey lines indicates recipient sequences. See text for mechanistic details.

additional proteins and even as RNA-protein complexes (Zimmerly et al., 1995)) the generation of double stranded cuts at, or near, the recognition sequence is a unifying feature. The recognition sequences for cleavage are much longer than for type II restriction enzymes, spanning up to 40 base pairs of DNA (Bryk et al., 1995). A specific nomenclature for naming homing endonucleases has been adapted from the nomenclature for restriction endonucleases. A prefix describing the locality of the HEG (I- for intron-encoded, PI- for intein-encoded, and F- for freestanding) is followed by a three-letter genus-species designation followed by a Roman numeral to distinguish enzymes from the same organism (e.g. the *omega* homing endonuclease is called I–SceI, the first intron-encoded HEG found in *Saccharomyces cerevisiae*, for latest updates of nomenclature see (Roberts et al., 2003)).

Group I introns in bacteriophage T4

Bacteriophage T4 contains three group I introns, one in the aerobic ribonucleotide reductase gene (*nrdB*) (Gott et al., 1986; Sjöberg et al., 1986), one in the anaerobic ribonucleotide reductase gene (*nrdD*) (formerly known as *sunY*) (Gott et al., 1986; Young et al., 1994), and one in the gene coding for thymidylate synthase (*td*) (Chu et al., 1984). The sequences of these three introns are remarkably similar compared to the overall sequences of group I introns indicating that the T4 introns have a common origin. Apart from the structural elements always present in group I introns, the T4 introns have additional regions and form their own subgroup (IA2) among the group I introns. The P7 helix is followed by two additional stem loops, P7.1 and P7.2, and helix P9 is made up of three stem loops P9, P9.1 and P9.2 instead of one.

Given the high degree of similarity between the three T4 introns a remarkable difference in the *nrdD* intron structure compared to the other two T4 introns is the lack of a P2 helix (Xu and Shub, 1989). This is especially strange since this region is thought to form a tertiary interaction with P8, positioning the 5' splice site at the G-binding site in the catalytic centre (Michel and Westhof, 1990). The fact that this helix is

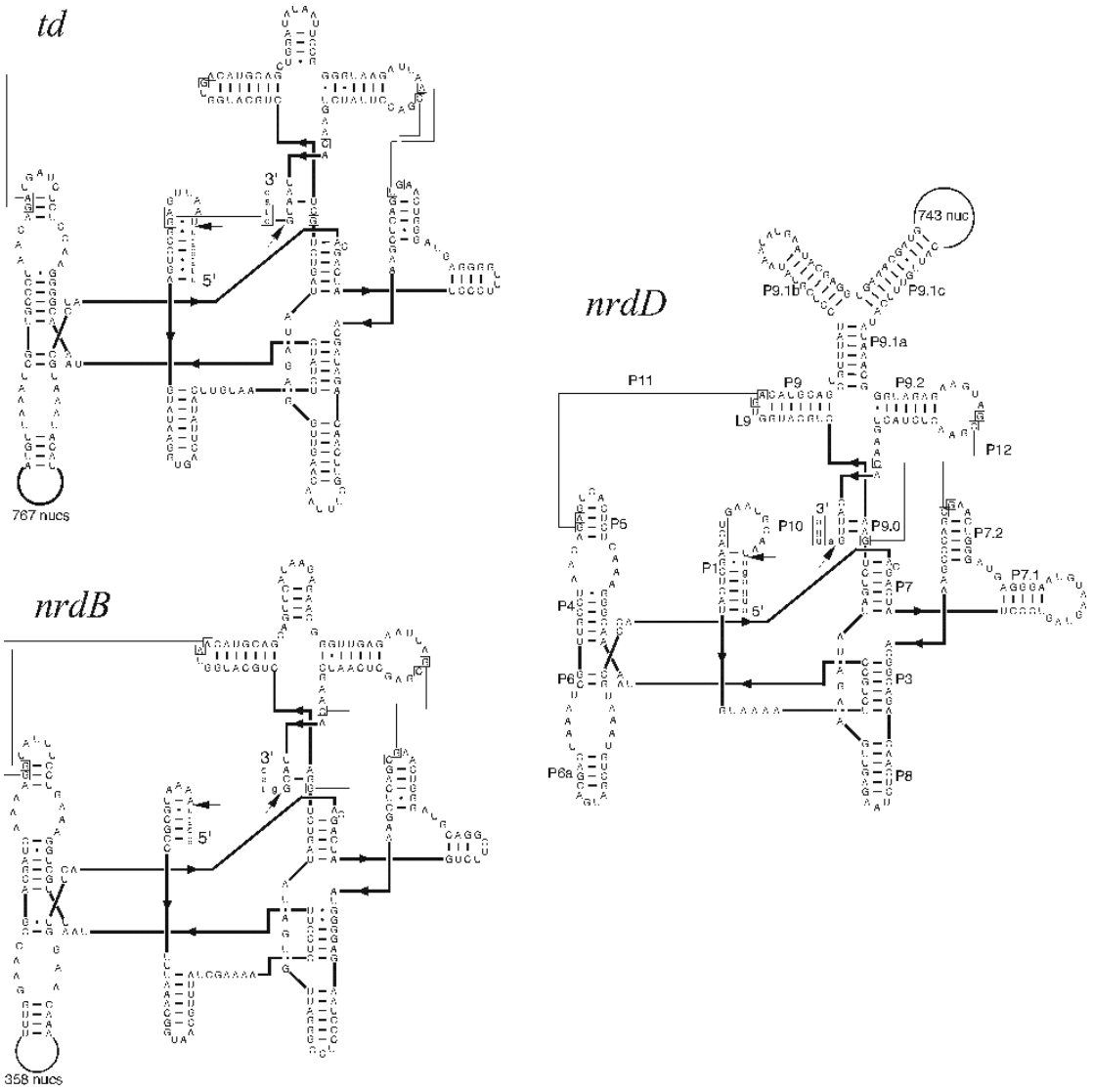


Figure 7. Secondary structure models for the T4 group I introns.

absent in some other group I introns implies that different folds have evolved to accomplish the 5' splice site positioning. In addition to the regions forming the catalytic structures of the intron, each of the T4 introns contains a long open reading frame (ORF) within the intron. These ORFs are situated in the peripheral loop of P6a in the *td* and *nrdB* introns and in the loop of P9.1 in the *nrdD* intron (Fig. 7). The ORFs are 735 (*td*), 774 (*nrdD*) and 291 (*nrdB*) bases long and have for the *td* and *nrdD* introns been shown to encode homing endonucleases, while the *nrdB* intron ORF is a remnant of a homing endonuclease gene (see below).

T4 intron-encoded homing endonucleases

Although all three T4 introns contain homing endonuclease ORFs only the *td* and *nrdD* introns are mobile (Quirk et al., 1989a; Quirk et al., 1989b). The ORF products of the *td* and *nrdD* introns have been shown to be homing endonucleases (called I-TevI and I-TevII respectively) with recognition sequences spanning the intron insertion sites (Bell-Pedersen et al., 1990; Bell-Pedersen et al., 1989) while the *nrdB* intron carries a non-functional version of such an endonuclease (Eddy and Gold, 1991). The lack of mobility of the *nrdB* intron is due to a 491 base pairs deletion representing almost two thirds of the original *nrdB* HEG. The closely related phage RB3 has been found to contain an almost identical *nrdB* intron that contains an ORF of 807 base pairs (Eddy and Gold, 1991). This ORF encodes a homing endonuclease (I-TevIII) with a recognition sequence at the *nrdB* intron insertion site. Strangely though, this intron has not been found to home in laboratory experiments (Eddy and Gold, 1991). In contrast to the homology between the three introns in T4, the HEGs are highly divergent (Shub et al., 1988). I-TevI and I-TevII belong to the GIY-YIG family but I-TevIII, based on the RB3 sequence, belongs to the H-N-H family (Eddy and Gold, 1991). From this it seems that the HEGs have different origins and have invaded the introns at different times. An indication that the T4 HEGs have resided in the phage genome for a long time is that they display the same codon usage and high A-T content as T4 and and also have highly T-even specific promoters and a precise transcriptional regulation system (Edgell et al., 2000). To hinder the HEGs from being translated from the unspliced pre-mRNA, which

would disturb intron splicing, a stem-loop structure is situated upstream of the HEG, covering the translational initiation site (Gott et al., 1988; Shub et al., 1987). Instead, the HEGs are translated from transcripts generated from their own T4 promoters. I-TevI and I-TevII have late promoters and I-TevIII has both a middle and a late promoter (Guild et al., 1988; Kassavetis et al., 1986). Expression of the homing endonucleases late in infection is probably advantageous since multiple copies of the phage genome, potential targets of homing, are present at that time.

Distribution of introns among T-even-like phages

As for many other group I introns, the distribution of the T4 introns is very irregular (Eddy, 1992; Pedersen-Lane and Belfort, 1987; Quirk et al., 1989b; Sandegren and Sjöberg, 2004). The *td* intron is the one most widely present in the T-even phage family, being found in T4, T6 and the more recently isolated phages RB3, LZ2, TuIa and U5 (Chu et al., 1984; Eddy, 1992; Sandegren and Sjöberg, 2004). The *nrdB* intron has only been found in T4, RB3 and U5 (Eddy and Gold, 1991; Sandegren and Sjöberg, 2004), and the *nrdD* intron only in T4 and U5 (Gott et al., 1986; Sandegren and Sjöberg, 2004). The distribution of group I introns among the T-even-like bacteriophages is discussed more thoroughly in the section describing Paper I.

Freestanding homing endonucleases in T4

Apart from the intron-encoded homing endonucleases T4 also contains several freestanding homing endonuclease genes inserted in between other genes. Sharma et al. (1992) recognised that five previously uncharacterised T4 genes (*segA-E,* for similar to endonucleases of group I introns) share the GIY-YIG sequence motif with the I-TevI endonuclease and intron encoded homing endonucleases of fungal mitochondria (Sharma et al., 1992) and they also showed that SegA has endonuclease activity *in vitro* with specificity to make a double strand cut within its neighbour gene *uvsX* (Sharma et al., 1992; Sharma and Hinton, 1994). Also SegC and SegE proteins have been shown to have endonuclease activity

(Kadyrov et al., 1997; Shcherbakov et al., 2002) and reports of unpublished data (Kadyrov et al., 1997) states that SegB and SegD also show site-specific endonuclease activity. In addition, two more genes have been added to the *seg*-family in T4, *segF* (previously gene 69 (Belle et al., 2002)) and *segG* (previously gene 32.1 (Liu et al., 2003)) both of which encode proteins that possess site-specific endonuclease activity. Of these, SegA, SegE and SegF cleaves the HEG-containing alleles in T4 although SegE cleaves RB30 DNA and SegF T2 DNA more efficiently (Belle et al., 2002; Kadyrov et al., 1997). SegC is the only one of these homing endonucleases reported not to cleave T4 DNA (Shcherbakov et al., 2002). In this respect the freestanding homing endonucleases seem to differ from their intron-encoded counterparts where the intron-containing allele is immune to cleavage due to the disruption of the endonuclease recognition sites by the introns. The apparent preference of the *seg*-genes for cleaving the genomes of related phages instead of the T4 genome has been shown to explain the concept of localized marker exclusion in mixed infections between T4 and related phages (Belle et al., 2002; Liu et al., 2003). In mixed infections between T2 and T4, T2 genetic markers are generally present at only 10-20% frequency in the progeny and some T2 loci are even less represented with <1% being transmitted to the offspring (Russell and Huskey, 1974). Belle et al. (2002) showed that the exclusion of T2 gene 56 was attributable to the presence of *segF* next to the T4 gene 56 and that the cleavage and following double strand break repair results in the predominance of T4 gene 56 and *segF* in the progeny (Belle et al., 2002). If this is a general feature of the *seg*-genes, having multiple homing endonuclease genes throughout the genome would give T4 a clear selective advantage in mixed infections with other phages by being able to cleave their genomes.

A second group of genes in T4 that have homology to homing endonucleases are the *mob*-genes (*mobA-E*, for similarity to mobile endonucleases) that share the H-N-H motif with the I-TevIII endonuclease and several mobile endonucleases of group I and group II introns (Kutter et al., 1995). None of the *mob*-genes have yet been reported to have endonuclease or homing properties (Miller et al., 2003b). This will be discussed further in the section describing Paper III.

Introns in other bacteriophages

The three introns represented in T4 are the only group I introns found in the T-even-like bacteriophages or any other *E. coli* phages. However, a growing number of group I introns are now being found in other phages, infecting gram-positive bacteria. Goodrich-Blair and co-workers have shown that a group I intron in the gene coding for DNA polymerase is abundant among the *Bacillus subtilis* HMU phages (Goodrich-Blair et al., 1990; Goodrich-Blair and Shub, 1994). HMU phages SPO1, SP82, 2C and ϕe all contain a group I intron at the same position in their DNA polymerase gene while in an additional phage (SP8), also belonging to this family, no intron was found (Goodrich-Blair et al., 1990). Although the sample size is small, this intron appears to be widespread in the HMU phage population in contrast to the scattered distribution of introns in the T-even-like phages. Interestingly, although the intron sequences of these phages are highly similar their H-N-H endonuclease ORFs have diverged substantially having only 43-70% amino acid identity (Goodrich-Blair and Shub, 1994). These homing endonucleases also have the unusual property of cleaving only one strand of their substrate and they cleave both intronless and intron-containing alleles (Goodrich-Blair and Shub, 1996). They also appear to have evolved differences in their recognition sequences so that SPO1 endonuclease cleaves SP82 intron-containing DNA and vice versa, with competition between phage sequences in mixed infections as a consequence (Goodrich-Blair and Shub, 1996; Landthaler et al., 2004). The unrelated *Bacillus thuringiensis* phage Bastille contain a group I intron at the exact same site in the DNA polymerase gene as the HMU phages but the intron sequence differs substantially and is instead more similar to the introns in the *Staphylococcus* phage Twort (see below) (Landthaler and Shub, 2003). However, the Bastille intron encodes a H-N-H homing endonuclease (I-BasI) homologous to the SPO1 and SP82 homing endonucleases I-HmuI and I-HmuII. I-BasI also nicks only one strand of the DNA target but it only cleaves intronless alleles (Landthaler and Shub, 2003).

Two other *B. subtilis* phages have been shown to contain introns. The virulent β22 phage has an intron inserted in the thymidylate synthase gene (Bechhofer et al., 1994). The β22 intron is inserted 21 base pairs further downstream in the *td* gene compared to the insertion site of the T4 *td* intron (Bechhofer et al., 1994). Like the T4 *nrdB* intron it has, what

appears to be, only a fragment of a homing endonuclease ORF but it is inserted in the P8 loop and not in the P6 loop as in the T4 *td* intron, indicating that these introns have been invaded by different HEGs on separate occasions.

The temperate *Bacillus subtilis* prophage SPβ contains two group I introns, one in each of the genes coding for the two subunits (b*nrdE* and b*nrdF*) of a class Ib ribonucleotide reductase, and in addition also an intein (a protein splicing element that is removed post-translationally) in the b*nrdE* gene (Lazarevic et al., 1998). In related prophages however, there are different intron/intein configurations (see Table 3) with variation in the number and positions of intervening sequences (Lazarevic, 2001). In total there are five different intron versions in the b*nrdE* and b*nrdF* genes among the seven prophage-containing *Bacillus* strains screened, with high degree of similarity (>98% identity) between introns in the same position and with 60 to 70% identity between all introns.

Two group I introns have been reported in bacteriophages belonging to the taxa *Siphoviridae*, infecting a completely different group of bacteria, the lactic acid bacteria. The lytic *Lactobacillus delbrueckii* bacteriophage LL-H was shown to have a group I intron in the gene *terL* (Mikkonen and Alatossava, 1995), coding for the large subunit of the terminase protein responsible for cutting the phage genome into genome size pieces during packaging into the phage capsids. The second intron was found in a gene of unknown function in the temperate lactococcal bacteriophage r1t (van Sinderen et al., 1996).

The phage containing the largest number of introns is the *Staphylococcus aureus* bacteriophage Twort that was shown to have at least five group I introns (Landthaler and Shub, 1999). Three group I introns (without HEGs) were shown to reside in the same gene (*orf142*), of unknown function (Landthaler and Shub, 1999). Two additional group I introns are inserted in the phage *nrdE* gene (Landthaler et al., 2002). One of the *nrdE* introns has an ORF inserted in the P6 loop that encodes a H-N-H homing endonuclease (I-TwoI) that nicks one strand of the intronless allele. The other *nrdE* intron has a 106 nucleotides insertion in the P6 loop that may be a remnant of a HEG.

A large screen for group I introns of *Streptococcus thermophilus* bacteriophages showed that half of the 62 phages screened contained a group I intron in the lysin gene (Foley et al., 2000). All introns were found to contain an ORF with the H-N-H motif and in five phages this ORF has suffered deletions most likely rendering them non-functional. Sequencing of the introns from a subset of the phages showed that they are almost identical even though the phages were isolated in different parts of the world (Foley et al., 2000).

The latest additions to phage group I introns were found via the complete genome sequence of the Staphylococcal phage K (O'Flaherty et al., 2004). This phage contains three introns, one in the lysin gene (*lys*-I1) and the other two in the DNA polymerase gene (*pol*-I2 and *pol*-I3). All three introns contain a HEG belonging to the H-N-H family.

An interesting fact is that all group I introns found in phages belong to the same structural subgroup (IA2) of the group I introns. Only three other group I introns (two in *Chlamydomonas eugametos* and one in *Chlamydomonas reinhardtii*) belong to this group, of the more than 2000 group I introns characterised (For updated compilations of group I introns see the RNA secondary structure database at: http://www.rna.icmb.utexas.edu/). This indicates that all the phage introns share a common ancestor. Furthermore, there is a clear bias towards genes coding for proteins involved in DNA synthesis being interrupted by introns in phages, with nucleotide synthesis genes (ribonucleotide reductases and thymidylate synthase) and DNA polymerases being clearly over-represented. This is discussed further in Paper II.

Table 2. Intron-containing genes in bacteriophages

Phage	Gene	Intron subgroup	HEG	Reference
T4 (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	(Chu et al., 1984)
	nrdB	IA2	(P6) I-TevIII (H-N-H, partial)	(Gott et al., 1986; Sjöberg et al., 1986)
	nrdD	IA2 (no P2)	(P9.1) I-TevII (GIY-YIG)	(Gott et al., 1986; Young et al., 1994)
T6 (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	
RB3 (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	(Eddy, 1992; Sandegren and Sjöberg, 2004)
	nrdB	IA2	(P6) I-TevIII (H-N-H)	(Eddy and Gold, 1991)
LZ2 (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	(Eddy, 1992; Sandegren and Sjöberg, 2004)
TuIa (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	(Eddy, 1992; Sandegren and Sjöberg, 2004)
U5 (*E. coli*)	*td*	IA2	(P6) I-TevI (GIY-YIG)	(Sandegren and Sjöberg, 2004)
	nrdB	IA2	(P6) I-TevIII (H-N-H, partial)	(Sandegren and Sjöberg, 2004)
	nrdD	IA2 (no P2)	(P9.1) I-TevII (GIY-YIG, partial)	(Sandegren and Sjöberg, 2004)
SPO1 (*B. subtilis*)	DNA-pol	IA2	(P8) I-HmuI (H-N-H)	(Goodrich-Blair et al., 1990)
SP82 (*B. subtilis*)	DNA-pol	IA2	(P8) I-HmuII (H-N-H)	(Goodrich-Blair and Shub, 1994)
2c (*B. subtilis*)	DNA-pol	IA2	(P8) I-HmuII (H-N-H)	(Goodrich-Blair and Shub, 1994)
ϕ (*B. subtilis*)	DNA-pol	IA2	(P8) (H-N-H)	(Goodrich-Blair and Shub, 1994)
β22 (*B. subtilis*)	*td*	IA2	(P8) (GIY-YIG, partial)	(Bechhofer et al., 1994)
Bastille (*B. thuringiensis*)	DNA-pol	IA2? (no P2)	(P8) I-BasI (H-N-H)	(Landthaler and Shub, 2003)
SPβ prophage (*B. subtilis*)	*bnrdE-I*	IA2	-	(Lazarevic et al., 1998)
	bnrdF-I	IA2	(P6) YosQ (H-N-H)	(Lazarevic et al., 1998)
M1918 prophage (*B. subtilis*)	*bnrdE-I*	IA2	-	(Lazarevic, 2001)
	bnrdF-I	IA2	(P6) (GIY-YIG)	(Lazarevic, 2001)
M1321 prophage (*B. subtilis*)	*bnrdE-I1*	IA2	-	(Lazarevic, 2001)
	bnrdE-I2	IA2	-	(Lazarevic, 2001)
M135 prophage (*B. subtilis*)	*bnrdE-I*	IA2	-	(Lazarevic, 2001)
BSG40 prophage (*B. subtilis*)	*bnrdE-I1*	IA2?	(P6) ORF732 (GIY-YIG)	(Lazarevic, 2001)
	bnrdE-I2	IA2	(P6) (GIY-YIG)	(Lazarevic, 2001)
	bnrdF-I	IA2	(P6) (GIY-YIG)	(Lazarevic, 2001)
Twort (*S. aureus*)	*Orf 142* I1	IA2	-	(Landthaler and Shub, 1999)
	Orf 142 I2	IA2	-	(Landthaler and Shub, 1999)
	Orf 142 I3	IA2	-	(Landthaler and Shub, 1999)
	nrdE-I1	IA2	(P6) 106 nt remnant?	(Landthaler et al., 2002)
	nrdE-I2	IA2	(P6) I-TwoI (H-N-H)	(Landthaler et al., 2002)
LL-H (*Lactobacillus*)	*terL*			(Mikkonen and Alatossava, 1995)
r1t (*Lactococcus*)	*orf40*			(van Sinderen et al., 1996)
S3b+27 more (*S. thermophilus*)	lysin	IA2 (no P2)	(P8) (H-N-H)	(Foley et al., 2000)
DT1+S92+S93 (*S. thermophilus*)	lysin	IA2 (no P2)	(P8) (H-N-H, partial)	(Foley et al., 2000)
Sfi16A (*S. thermophilus*)	lysin	IA2 (no P2)	(P8) (H-N-H, partial)	(Foley et al., 2000)
ST64 (*S. thermophilus*)	lysin	IA2 (no P2)	(P8) (H-N-H, partial)	(Foley et al., 2000)
Phage K (*Staphylococcus sp.*)	lysin I1		I-KsaI (H-N-H)	(O'Flaherty et al., 2004)
	DNApol I2		I-KsaII (H-N-H?)	(O'Flaherty et al., 2004)
	DNApol I3		I-KsaII (H-N-H)	(O'Flaherty et al., 2004)

Present studies

Paper I: Distribution, Sequence Homology and Homing of Group I Introns among T-even-like Bacteriophages. Evidence for recent transfer of old introns.

Before this study was initiated it was known that group I introns were only scarcely found among phages closely related to T4 (Eddy, 1992; Pedersen-Lane and Belfort, 1987; Quirk et al., 1989b). In 32 T-even-like phages examined, Sean Eddy found that besides T4 only T6, LZ2, RB3 and TuIa contain the *td* intron and that RB3 also contains a *nrdB* intron (Eddy, 1992). This is somewhat unexpected since the *td* and *nrdD* introns of T4 are mobile (Quirk et al., 1989a) and the I-TevIII homing endonuclease of the RB3 *nrdB* intron at least can initiate cleavage (Eddy and Gold, 1991). This spoke against a general selective advantage for introns in phages, and a T2 construct with the *td* and *nrdD* introns did not display any growth advantage over intronless T2 in one-step growth experiments in rich medium (Quirk et al., 1989a). Eddy also attempted to test if the introns had any effect on the burst size of the phage by measuring the burst sizes of T4 with and without introns during growth in different media (Eddy, 1992). His conclusion from this was that any effect (positive or negative) of introns under the conditions tested was smaller than could be detected in such experiments (the inherent variation between experiments was typically ± 20%).

Our aim in this paper was to investigate if the introns found in different T-even-like phages indeed have been spread between the phages and to see if the sequences of the introns and of the *td*, *nrdB* and *nrdD* genes of phages with and without introns could give some indications to why introns are not more frequent among T-even-like phages. No intron sequences were available other than of the T4 introns and the RB3 *nrdB* intron and no complete sequences of any of the *td*, *nrdB* and *nrdD* genes from T-even-like phages other than T4 had been determined previously. During the course of this study the Tulane Phage Sequencing Group led by Jim Karam sequenced several genomes of phages, more or less closely related to T4. This gave us the opportunity to include complete sequences

of the three genes from phages outside the T-even-like group. No introns were found in any of the genomes sequenced.

In our amplification of the three genes from phages with and without introns we included the previously uncharacterised T-even-like phage U5. This phage was isolated at a microbiology course at Uppsala University and was kindly provided to us together with several of the RB-phages by Karin Carlson, Department of Cell and Molecular Biology, Uppsala University. To our surprise, U5 contains all three introns and is the first phage isolate besides T4 shown to do so. Like RB3, the U5 *nrdB* intron contains a much longer I-TevIII gene but the U5 I-TevIII has, like its T4 counterpart, suffered a deletion although only of two base pairs. This is however enough to shift the reading frame and since it occurs early in the gene it most likely disrupts its function. In addition to the frame shift deletion there is also a 6 base pairs deletion exactly spanning the late promoter of the I-TevIII gene (Fig. I1).

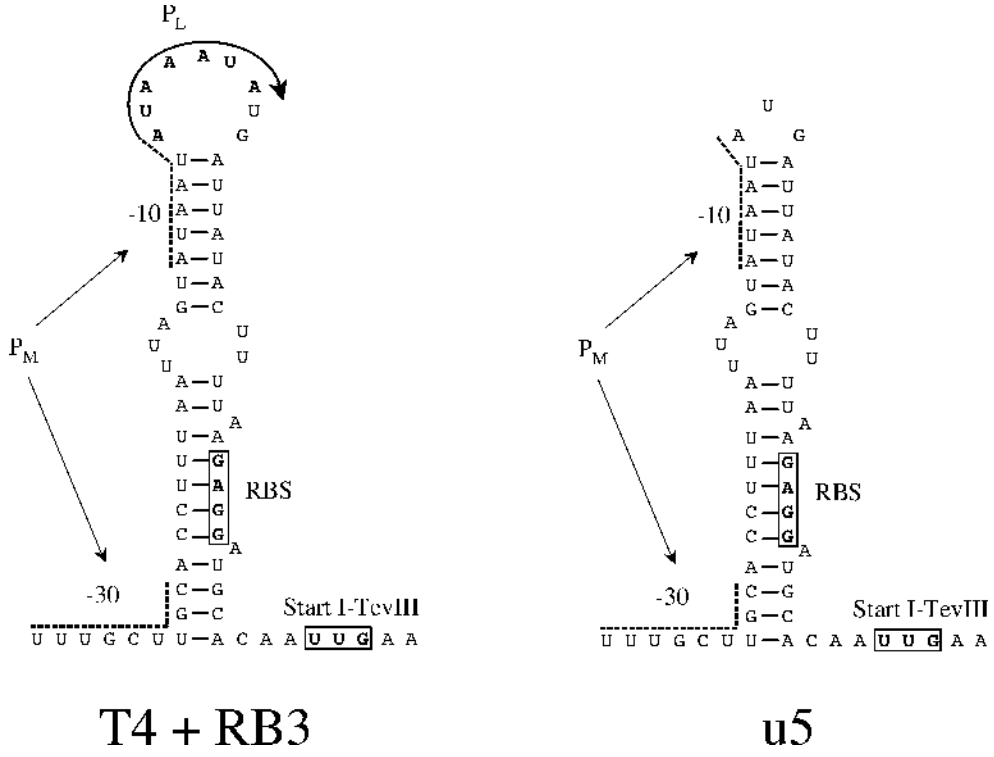


Figure I1. Deletions in U5 I-TevIII promoter region compared to T4 and RB3.

Interestingly, the I-TevII gene in the U5 *nrdD* intron has also suffered a 2 base pairs deletion early in its reading frame. It is tempting to speculate that expression of I-TevII and I-TevIII in some way is detrimental to the phage and that these deletions have enabled U5 to remain within the population. If a homing endonuclease can cleave at additional sites in a genome, to which it cannot home, it is possible that this has a negative effect on phage viability. If this is a general trait of homing endonucleases it may be an explanation to their scattered appearance among phages. The toxicity of the T4 intron-encoded homing endonucleases to *E. coli* when cloned on other than low copy plasmids is generally thought to come from cleavage of the bacterial genome (Quirk et al., 1989a). However, T4 is an example of a phage that contains multiple HEG-containing introns and at least 12 freestanding genes related to homing endonucleases some of which have been shown to possess endonuclease activity and to induce cleavage at multiple sites in the T4 genome (see paper III). This may reflect an increased tolerance to double strand breaks in T4 (Belle et al., 2002).

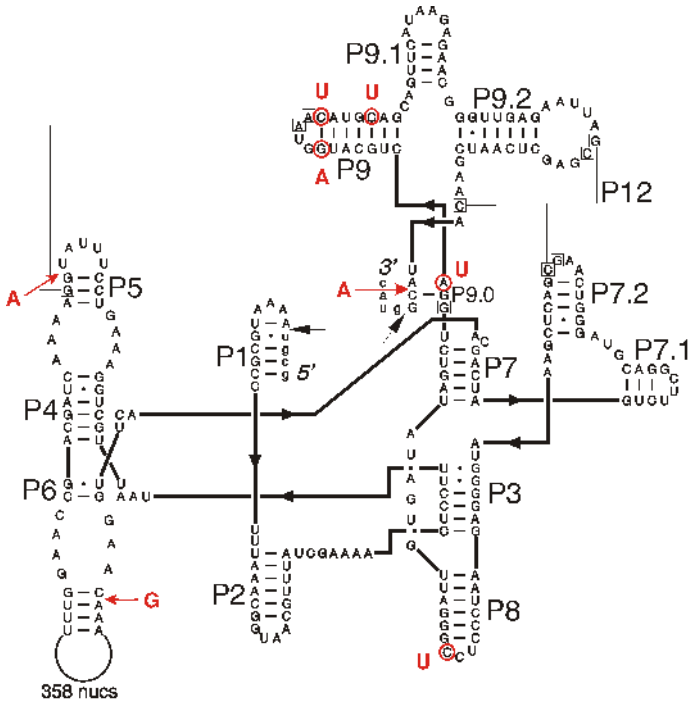
When comparing the sequences of the intron-containing *td*, *nrdB* and *nrdD* genes there is a striking difference between the amount of sequence variation in the exons and in the introns between all intron-containing phages and especially between American phage isolates T4, T6, RB3 and LZ2. Among the American phages there is only one nucleotide difference throughout the *td* intron/HEG region while there is around 10% variation in *td* exon sequences between the phages. Although the catalytic parts of the introns may have a low mutation rate due to functional constraints the coding regions of the HEGs should accumulate at least synonymous changes over time. The lack of sequence variation between HEGs of cognate introns together with the scattered distribution of the introns on the gene-based phylogenies (Fig. 3 in paper I) strongly indicate that the introns have been spread horizontally between the phages, and from the lack of variation between the *td* introns of American isolates the acquisition of the *td* introns appears to have been very recent among these phages.

The positions of the nucleotide substitutions within catalytic regions of the *nrdB* and *td* introns between phages are shown in figure I2 (there are no differences in catalytic regions between the T4 and U5 *nrdD* introns).

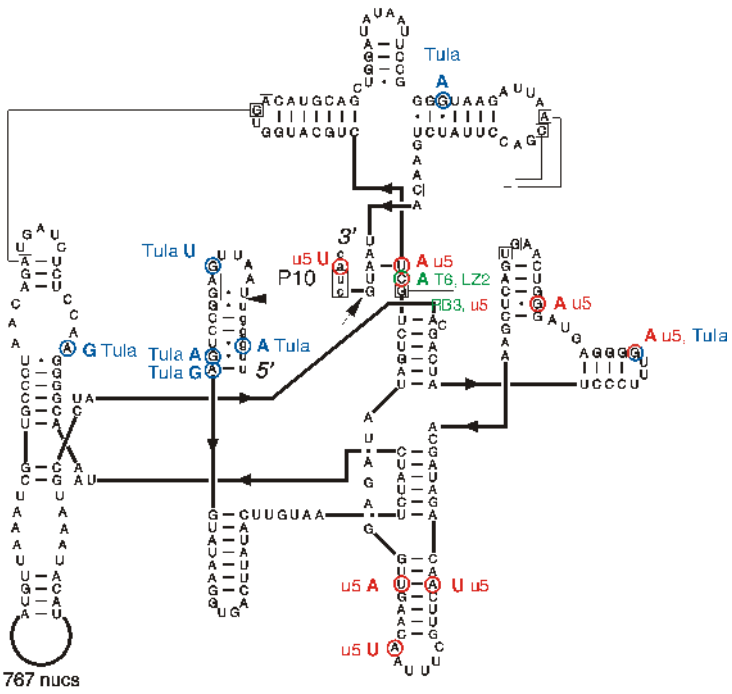
None of the seven new introns sequenced in this study show any variations that can be predicted to negatively affect splicing. Differences occurring in nucleotides involved in base pairing interactions are either accompanied by covariation at the corresponding site, keeping the interaction intact, or introduce an accepted non Watson-Crick base pair (most often G-U).

Interestingly, a large part of the changes in the intron structural parts are situated in regions P1, P9.0 and P10, known to be important in splice site selection. The *td* sequence of TuIa differs from its T4 counterpart at seven positions in regions of intron secondary structure. Four of these occur in the internal guide sequence, made up of helix P1 and the P10 interaction. P1 is the major determinant of the 5´ splice site while the P10 interaction is responsible for keeping the region of 5´ cleavage in close proximity to the 3´ splice site for ligation of the two exons (Burke, 1989; Davies et al., 1987). The P1-helix is maintained intact in spite of these changes, showing a high tolerance to variations given that base pairing is maintained. A G→U change in the P1-loop of the *td* intron in TuIa, and an A→U change of the third base of *td* exon2 in U5, both facilitate the possible formation of stronger P10 interactions in these phages. Together with the variations seen in P1 this means that a large part of the variations between the different phage strains occur in regions involved in the 5´ splice site selection. These variations may be adaptations in order to avoid cryptic splicing due to variations in exon sequences between the phages.

Another interesting set of variations is seen in the P9.0 pairing of the *td* and *nrdB* introns. In most group I introns the P9.0 interaction is made up of two base pairs between nucleotides preceding the P7 helix and the penultimate nucleotides at the 3' splice site (Burke et al., 1990; Michel and Westhof, 1990). In T4 only the *nrdD* intron contains a P9.0 interaction of two Watson-Crick base pairs while it is made up of only one base pair in both *td* and *nrdB* introns (see Fig. I2). In U5, however, the *nrdB* intron has an A→U substitution, compared to T4, in the J7/9 linking region and thus can form a two base pair P9.0. This feature is shared by the RB3 *nrdB* intron (Eddy and Gold, 1991). In the *td* intron, the P9.0 of T4 is made up of a single A-U base pair while in T6, U5, RB3 and LZ2 the U is preceded by an A, instead of the C in T4 and TuIa. This A can then base



nrdB intron



td intron

Figure I2. Sequence variations in catalytic parts among the *td* and *nrdB* introns. Changes are colour-coded for each phage. Changes in *nrdB* are identical for both RB3 and U5.

pair with the penultimate U of the intron forming a P9.0 interaction of two base pairs. However, U5 has an A instead of a U preceding P9 and will thus still only have a one base pair P9.0. Opposing A:s are also found in both T4 *nrdB* and U5 *td* introns next to the single identifiable Watson-Crick base pair. Introns that entirely lack recognizable P9.0 pairings often have opposing A:s on both strands in this region (Michel and Westhof, 1990), indicating that non-Watson-Crick A-A base pairs may constitute the P9.0 interaction. Thus, all T-even-like phage introns except the *td* intron in T4 and TuIa probably form two base pair P9.0 interactions. This further corroborates the view that variations in the interactions important for splice-site selection are especially frequent among the phage introns.

Our screen of homing between phages in mixed infections showed, as expected, that homing most easily occurs between closely related phages. The few sequence variations around the intron insertion sites in the *nrdB* and *nrdD* introns among the T-even-like phages indicate that it is not a lack of homing sites that has limited the spread of these introns among phages. However, the higher degree of variation between *td* alleles together with the presence of an optional gene (*td*.1) flanking the *td* gene appear to lower the homing efficiency of the *td* intron between some of the phages. Furthermore, the process of general exclusion can reduce the efficiency of intron transfer but as was shown for the *td* intron of RB3 in crosses with T4ΔIVS, homing can overcome this exclusion to some extent and from the few progeny that obtained the intron it can be transferred with increased efficiency in subsequent mixed infections.
In conclusion we find that intron homing has occurred between T-even-like phages in Nature and can be very efficient between closely related phages but that sequence divergence of regions flanking the intron insertion sites, optional occurrence of flanking genes and the process of exclusion can lower the efficiency of homing. This together with the apparent rapid degeneration of HEG-function that we find in some of the introns may limit the spread of introns among the T-even-like phages. However, once an intron succeeds in transferring to a more distantly related phage, subsequent homing to phages related to the new host will be rapid.

Paper II: Self-splicing of the Bacteriophage T4 Group I Introns is Affected by the Growth of the Infected Bacterium and Requires Efficient Translation of the Pre-mRNA *In Vivo*.

Ever since the T4 introns were found it has been speculated about how they affect the phage viability. Initially it was argued that in order to be maintained in the phage genome the introns must have a positive effect (probably through regulation of the intron containing genes) on phage viability (Eddy, 1992; Goodrich-Blair et al., 1990; Gott et al., 1986; Shub, 1991; Shub et al., 1988; Shub et al., 1987) However, when it was found that introns are very rare among the T-even-like phages and that the introns are/have been mobile the view that they are merely selfish genetic elements that confer no advantage to the phage prevailed. So far there have been no reports of any effect, positive or negative, on phage viability that can be coupled to intron presence.

In 1998 Semrad and Schroeder published a paper showing that splicing of the T4 *td* intron was highly dependent upon translation of the upstream exon (Semrad and Schroeder, 1998). The effect on *td* splicing was shown to be due to the formation of an interfering tertiary interaction between a region of the upstream exon and the 3´splice site of the intron. During normal translation the movement of ribosomes along the pre-mRNA disrupts this interaction but if translation is hindered by addition of antibiotics or by introduction of stop codons upstream of the interacting region, formation of the interaction will inhibit splicing. Since this could make the introns a regulatory "switch" for regulation of the intron-containing genes we decided to test if the dependence on translation of the upstream exon for efficient splicing was a common theme for the T4 introns and if limited growth conditions could affect splicing by this mechanism. We found through a computer-based complementarity screen with *nrdB* exon and intron sequences, that a highly similar interacting region exists early in the *nrdB* exon (see Fig. 3A, Paper II). Several other potential interacting regions were also found. By measuring splicing in the absence of translation by introduction of stop codons or addition of chloramphenicol, that inhibits the elongation step of translation, we were able to show that splicing of the *nrdB* intron also is dependent upon

translation of the upstream exon. This was however, not primarily due to the interaction with similarity to the one in *td* but other regions, closer to the intron, such as an alternative P1 stem (see Fig. 3B, Paper II) were responsible for most of the reduction. Deletion of the exon regions potentially forming such interfering interactions restored the efficiency of splicing to close to wild type levels even in the absence of translation.

A number of potential interfering regions with complementarity to intron sequences can be predicted for all three T4 introns and we predict that this is a general feature of introns in translated genes. In line with this, we also showed that reduction of splicing efficiency occurred for all three T4 introns during phage infection of cells that had been treated with chloramphenicol and in which translation therefore was abolished. A similar effect of chloramphenicol on splicing was noted by Belfort and coworkers already in 1985 but they ascribed the effect to the inhibition of translation of an unknown protein factor that was needed for splicing *in vivo* (Belfort et al., 1985). So far no such protein factor essential for T4-intron splicing has been found, although several *E. coli* proteins that enhance group I intron splicing have been characterised (see section on Proteins involved in group I intron splicing in the introduction).

Since burst size experiments have large intrinsic variations (Eddy, 1992) we choose to measure the effect of growth condition on splicing directly by limited primer extension. Using a plasmid-encoded T4 *nrdB* gene containing the intron, we found a substantial reduction in splicing efficiency when the bacterial culture entered stationary phase and after prolonged growth in stationary phase splicing efficiency was down to 40% from the initial 90% during logarithmic growth. This effect on splicing was also seen for all three T4 introns upon infection of bacteria in stationary phase. Although we have so far not definitely shown that it is the reduction of translation in stationary phase that reduce the efficiency of splicing, this is the first time that environmental effects such as growth conditions have been shown to affect splicing of phage introns. The reduction of splicing when the infected cell is in a starving state was the favourite predicted scenario for a regulatory role of the introns (Eddy, 1992; Goodrich-Blair et al., 1990; Gott et al., 1986; Shub, 1991; Shub et al., 1988; Shub et al., 1987). By lowering the *de novo* production of ribonucleotides for DNA synthesis, energy consumption could be reduced

and the survival of the infected bacterium prolonged. If too much energy is spent on nucleotide synthesis the bacterium might die before any viable phage has been produced. Prolonged infection might be a good way for a highly infectious phage like T4 to stay dormant until there are new host cells available to infect. Further experiments are needed in order to evaluate a possible effect of introns on the survival rate of infected bacteria.

Paper III: Two Genes Encoding Novel Homing Endonucleases Replace the Putative Homing Endonuclease Genes *mobC* and *mobE* in Several T4-related Phages.

During the work on the distribution of group I introns among the T-even-like phages (Paper I) we noted that there were differences in the flanking genes of all three intron-associated genes *td*, *nrdB* and *nrdD*. We therefore performed an independent screen of these regions from 20 T-even-like phages that confirmed the presence of the *mobE* gene between *nrdA* and *nrdB* in five other phages, the absence of *mobC* gene between *nrdD* and *nrdG* in all phages except T4 and the presence of a previously unreported gene not present in T4 (we named it *td*.1 see Paper I) upstream of the *td* gene in many of the phages. The *mobC* and *mobE* genes belong to a family of five T4 genes called the *mob*-genes (for similarity to mobile endonucleases (Kutter et al., 1995)) that share homology with the H-N-H homing endonucleases. The *mob*-genes are putative freestanding HEGs (i.e. not associated with introns). In addition to the five *mob*-genes, T4 also contains seven freestanding *seg*-genes (for similarity to endonucleases encoded by group I introns (Sharma et al., 1992)) containing the GIY-YIG motif, four of which (*segA* (Sharma et al., 1992; Sharma and Hinton, 1994), *segE* (Kadyrov et al., 1997), *segF* (Belle et al., 2002) and *segG* (Liu et al., 2003)) have been shown to encode proteins with homing endonuclease activity. Table III1 shows a compilation of the distribution of freestanding homing endonuclease genes among T4-like phages.

Our sequence of the T4 *mobE* gene differed from the previously reported sequences of this region by two frame shifts that brought two proposed

open reading frames into one continuous open reading frame, with the start codon overlapping the stop codon of *nrdA* and the stop codon overlapping the start codon of *nrdB* (see Fig. III1). All five additional *mobE* genes that we found in the screen (in T6, RB2, RB3, RB15 and LZ7 respectively) were inserted between *nrdA* and *nrdB* in exactly the same manner. Since all our sequences were obtained by direct sequencing of PCR fragments from genomic DNA this strongly suggests that the original sequence of *mobE* (obtained from a cloned genome fragment (Sjöberg et al., 1986; Tseng et al., 1988)) was of a non-functional gene and that expression of full-length MobE protein is toxic to *E. coli*, a feature shared with several experimentally proven homing endonucleases (Kowalski and Derbyshire, 2002). In agreement with this we have been unable to clone full-length *mobE*.

Even though most T-even-like phages lack *mobC* and *mobE*, several phages have unrelated genes at the same sites. Phages T6, RB3 and LZ2 all have a gene related to the *seg*-genes in place of *mobC* between *nrdD* and *nrdG*. We have named this gene *segH* as it is the eighth *seg*-gene described. *In vitro* expressed SegH protein displayed homing endonuclease-like activity by introducing site-specific, double-strand cuts in the end of the *nrdD* gene, a little more than eighty base pairs upstream of the *segH* start codon. Similarly, phage U5 has an open reading frame of 544 codons between the *nrdA* and *nrdB* genes, in place of the T4 *mobE* gene. The product of this gene also showed homing endonuclease activity by introducing a double-strand cut in the downstream *nrdB* gene. The U5 gene has no homology to any genes presently in GenBank and we have named it *hef* (for homing endonuclease-like function).

Although we have tested DNA targets from several different phages and tested several different buffer conditions we have still not been able to detect either double-, or single-strand cleavage in the regions of the *mobC* and *mobE* genes by either T4, T6 or RB3 MobE proteins or T4 MobC *in vitro*. This could mean that they either are non-functional or that we have not tested the correct conditions/target sequences for cleavage. It may also be that they only recognise and cleave the modified DNA (containing glucosylated hydroxymethyl-cytosine) that many T-even-like phages have.

Table III1. Freestanding homing endonucleases in T4-like bacteriophages

Phage	*mobA*	*mobB*	*mobC*	*mobD*	*mobE*	*segA*	*segB*	*segC*	*segD*	*segE*	*segF (gp69)*	*segG (gp32.1)*
T2			-[1]		-[1]	-[8]					-[2,3]	-[6]
T4	+	+	+	+	+	+	+	+	+	+	+	+
T6			*segH*[1]		+[1]	-[8]					-[2,3]	-[6]
RB2			-[1]		+[1]							
RB3			*segH*[1]		+[1]	-[8]						+[6]
RB6												+[6]
RB8												+[6]
RB9												+[6]
RB10												+[6]
RB14			-[1]		-[1]							
RB15			-[1]		+[1]	-[8]					-[2,3]	
RB18												-[6]
RB23			-[1]		-[1]							
RB26			-[1]			-[8]						
RB27			-[1]		-[1]							-[6]
RB30										-[5]		
RB32			-[1]		-[1]							-[6]
RB49	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]
RB51			-[1]							-[5]		
RB61			-[1]									
RB69	-[4]	-[4]	-[4]	-[4]	-[4]	-[4,8]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]
RB70												-[6]
LZ1			-[1]		-[1]							
LZ2			*segH*[1]		-[1]							
LZ5											-[2,3]	
LZ7			-[1]		+[1]							
LZ11						-[8]						
U5			-[1]		*hef*[1]							
TuIa			-[1]			-[8]						
TuIb												-[6]
AehI	-[4]	-[4]	-[4]	-[4]	"+[4]"	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	-[4]
RR44						-[4]	-[4]	-[4]	-[4]	-[4]	-[4]	
Mi												+[6]
PST												+[6]
FSα												+[6]
SV76												-[6]
KVP40	-[7]	-[7]	-[7]	-[7]	-[7]	-[7]	-[7]	-[7]	"+[7]"	-[7]	-[7]	-[7]

Refs. 1 – Our work; 2 – (Belle et al., 2002); 3 – (Mosig et al., 2001); 4 – genome sequence; 5 - (Kadyrov et al., 1997).; 6 – GenBank sequences/(Liu et al., 2003); 7 – (Miller et al., 2003a); 8 - (Sharma and Hinton, 1994)

To investigate if the novel homing endonucleases SegH and Hef can promote homing of their genes, and to further investigate the reported unidirectional inheritance of *mobE*, we performed mixed infections with phages with and without these genes. We were able to show that all three versions of the *segH* gene (T6, RB3 and LZ2) are homing proficient in crosses with T2, but due to the general exclusion of other phages by U5 we have still not been able to show if it is the activity of Hef that promotes the strong local exclusion of the *nrdA-nrdB* region of recipient phages in crosses with U5. By using T6 as donor phage for studies of *mobE* homing we were able to avoid the problem of general exclusion that we previously have experienced with T4 *mobE*. In a screen of the flanking *nrdB* region from T6/T2 progeny that had inherited the *mobE* gene we found that three of 13 phages had a recombination site in the region of *nrdB* that we could screen. The sequence variation between T2 and T6 in the first half of the *nrdB* gene would restrict homing-induced recombination events to the second half of the gene, which we have only been able to screen a small part of. Nonetheless, the occurrence of recombination points in the *nrdB* gene of these progeny phages is encouraging and screening of an increased region of the *nrdB* gene is in progress.

SegH and Hef were shown to cleave both HEG-containing and HEG-less alleles. Cleavage of the own DNA may be lethal to the host phage perhaps explaining the limited occurrence of HEGs in phage genomes. However, limited cleavage activity at sites in the own genome may be tolerated since it, at least in T4, induces recombination dependent replication. We propose that it may also increase the propensity of the HEG to spread to new sites in a genome, as discussed in Paper III. Relocations have been proposed to be the only way for a HEG to avoid degeneration and stay viable over evolutionary time (Goddard and Burt, 1999) and the high frequency of deleterious deletions that we observe in intron-encoded HEGs (Paper I) suggests that the selective pressure to remove unnecessary sequences is high in phages.

Although insertion of freestanding HEGs between genes may be tolerated in many cases since the HEG does not interrupt the reading frames of the nearby genes, the insertion of *mobE* between the *nrdA* and *nrdB* genes, that code for subunits of the same protein, has affected the expression of

them. Tseng and coworkers (1990) showed that coordinated expression of the two proteins, R1 (*nrdA*) and R2 (*nrdB*), constituting the T4 aerobic ribonucleotide reductase occur by a set of shifts in promoter usage and transcriptional termination/anti-termination (Tseng et al., 1990). As shown in figure III1, expression of *nrdA* and *nrdB* occurs both from separate promoters and from a common promoter as part of an operon together with dihydrofolate reductase (*frd*), thymidylate synthase (*td*) and possibly endonuclease II (*denA*) (Tseng et al., 1990).

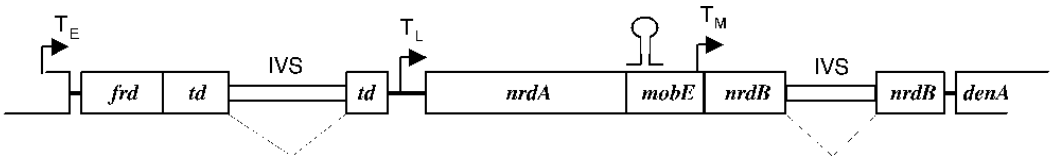


Figure III1. Operon of nucleotide synthesis genes. Promoters (E = early, M = middle, L = late), and the transcriptional terminator (stem-loop) before *nrdB* are shown. Dotted lines indicate exons that are spliced together. Drawn essentially as in (Tseng et al., 1988).

Early in infection *nrdA* is expressed from the polycistronic transcript generated from an early promoter located upstream of *frd*. Initially, transcripts from this promoter do not include the *nrdB* gene, downstream of *nrdA*, due to a transcriptional termination site located 50 base pairs after the *nrdA* stop codon. Instead, *nrdB* is transcribed from a MotA dependent, middle promoter just in front of the gene (Tseng et al., 1990). T4 anti-termination factors are produced at about 5 minutes after infection, and then transcription from the early, *frd* promoter reaches the *nrdB* gene. At approximately the same time transcription starts from a late promoter in front of *nrdA*, and this gives a coordinated expression of the ribonucleotide reductase genes during replication (Tseng et al., 1990). These alternations in the expression of *nrdA* and *nrdB* most likely influence the production of dNTPs by the dNTP synthetase complex.

Intriguingly, both the transcriptional terminator downstream of *nrdA* and the middle promoter in front of *nrdB* are situated within the *mobE* gene (Fig. III1). The complicated regulatory scheme of *nrdA* and *nrdB*

coexpression must therefore have been adopted upon insertion of *mobE* in the T4 genome. Phages lacking *mobE* only have a 50 base pairs intergenic region between *nrdA* and *nrdB* without obvious terminator/promoter sequences (our unpublished observation) and coordinated expression of R1 and R2 proteins is likely to occur from the promoters upstream of *nrdA* in these phages. Yet, how the extensive sequence variation in the N-terminal part of the RB3 *mobE* gene (which lacks the *nrdB* middle promoter while having the upstream terminator) and the insertion of the much longer *hef* gene in U5 affect expression of the *nrdA* and *nrdB* genes remains to be investigated.

Future studies

Several interesting follow-up projects have emerged from the data that this thesis is based upon.

To determine if it is indeed the reduction of translation during stationary growth that affects the splicing efficiency of the T4 introns we are deleting the major intron-interacting regions in the *nrdB* exon1 (Δ45-129;ΔCry-P1). If splicing of such a construct is increased during stationary growth in comparison to the wild type T4 *nrdB* transcript this would be a strong indication that the reduction of splicing that we have seen is indeed coupled to the upstream exon of *nrdB* and not due to any factors acting on splicing in trans.

It would also be of interest to analyse if the reduction of splicing in stationary phase bacteria will yield any difference in the survival rate of bacteria infected with intron-containing phages compared to intronless phages. An alternative way to detect an effect of introns upon phage production would be to measure the production of viable phage in stationary phase without natural lysis by sampling an infected culture over several hours/days, lyse the cells with chloroform and measure the amount of viable phage that was present inside the bacterium.

Further biochemical and bioinformatic studies on the Hef protein is under way to determine if it represents a new type of freestanding homing endonuclease or if it is a distant relative to one of the previously characterised families. *In vivo* cleavage studies of both Hef and SegH could shed more light on whether they have any preference for cleavage of HEG-less alleles. Homing and *in vivo* cleavage studies of *mob*-genes are also under way to determine if they really encode homing endonucleases. The *mobE* gene is so far the best candidate since it appears to promote homing.

We are also planning to do homing studies with the *td*.1 gene that is situated between the *frd* and *td* genes in most T-even-like phages that we have screened but that is absent in T4. Preliminary studies have shown that the sequence of *td*.1 is almost identical between all phages except RB23 that contain a completely unrelated gene at the same position. No sequence homologies to any genes in the databases have been found for these two genes so far.

Acknowledgements

I would like to thank the people without which help and support this thesis would not have been possible:

My supervisor **Britt-Marie Sjöberg** for sharing her great knowledge of science and for always letting me try my own ideas.

My parallellslalompolare in the lab **David Nord** for all the help with phage crosses, massive PCR screenings, lunches, loftning, football-coaching, Zombie movies and for all the Great Wars in space and elsewhere.

The ”old” PhD students **Jessica**, **Maja**, **Monica**, **Jasmine** and **Martin** for always helping out during the first years. **Margareta Karlsson** my first lab-supervisor. **Ant**, **Pernilla**, **Alex**, **Hanna**, **Tove**, **Anna**, **Patrick**, **Jossan**, **Annika** and all the “new” PhD students and other personnel at the Department of Molecular Biology & Functional Genomics for making it a very nice place to work in.

All the nice “scientific” people I have had the opportunity to meet on conferences over the years (in order of appearance) **Larry Gold** who told me to stay away from the phage introns because they were not worth working with (didn’t follow that advise), **Jack Szostak** for nice discussion and Lagavulin, **Jennifer Doudna** whose crystal structure of P4-P6 got me interested in group I introns, **John Roth** who has told me twice to stay away from ppGpp strains (did follow that advise!), **Matt Goddard** for all help with HEG phylogenetic theories, **Sean Eddy** for a nice discussion about introns in phages, **Betty Kutter** for a very nice T4 meeting 2003, **Chris Mathews** for advice on Paper I, **Jim Karam** for sharing ideas on phage genomes, **Deborah Hinton, David Shub** and **David Edgell** for nice discussions on T4 freestanding HEGs and **Steve Abedon** for being the MicroDude.

All friends that have made these years great: **Andy**, **Dabuel**, **Milla**, **Amy, Elin**. **Oskar** for fishing at the Big Essing. **Richard** and **Bobby** for the early years, and indeed the late ones too. **Dr. S** for hosting all the crayfish parties. **Q**, **Dr. T**, **Dr. Skåne**, **Dr. M**. **Dr. Martin** for all the Whisky and

Ford Fairlane evenings. **Stina** the Pors provider. **X-björn**, **Niklas** "Typhoon" **Holmström**, **Christian**, **Mikaela**, **Dr. Tina**, **Dan**, **Susanne**, **Love, Hanna**. **Kristina** for nearly always booking the innebandy. **Hubert**, **Stefan, Malin**, **Linda**. **Calle** for all the battle-Sundays at KI. **Dr. Martin S** for fishing, walking and talking. **Maria** & **Henrik** for great midsummer-, and glögg-parties. And all the footballers, Loftare, NF and MoleQul people.

My Parents **Gunnar** and **Britt-Mari** for all their support when the bacteria/phages/genes and other lab-stuff don't do as they are supposed to. My sister **Ellinor**, her husband **Atte** and little "**Udo**". Mina morföräldrar **Bertil** och **Irma** (och **Sussie**). My aunt **Britta,** and **Göran**. Annas parents **Åke** and **Birgitta** for always being so nice and for great dinners with a lot of great food. **Astrid**, **Jonas**, **Maja** and **Lukas**. The Italian side **Inga-Britt**, **Enzo**, **Ettore** and **Monica** for a great time in Rome after my licentiate defence.

Anna, for all her love and for being Världens bästa.

Several other persons/institutions have also made a difference during this work (in no particular order): AIK, J.R.R. Tolkien, Mark Knopfler, BK Kanon, The Loft, The Macallan, Jorms fiskecamp, X-björns, Springbanks, Tullfritt, gamla Lantisgrillen, bastun under Arrhe, 106,7 Rockklassiker, Gula Villan, Bono, BB King, Laphroaigh, Urga, Natural Selection, and many many more…..

Inspiration and energy for the completion of this work has come from innumerable cans of Coca Cola and on successful lab days from Single Malt Whisky.

References

Abedon, S. T. (1994). Lysis and the Interaction between Free Phage and Infected Cells. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 397-405.

Abedon, S. T. (2000). The murky origin of Snow White and her T-even dwarfs. Genetics *155*, 481-486.

Ackermann, H. W., and Krisch, H. M. (1997). A catalogue of T4-type bacteriophages. Arch Virol *142*, 2329-2345.

Allen, J. R., Lasser, G. W., Goldman, D. A., Booth, J. W., and Mathews, C. K. (1983). T4 phage deoxyribonucleotide-synthesizing enzyme complex. Further studies on enzyme composition and regulation. J Biol Chem *258*, 5746-5753.

Allen, J. R., Reddy, G. P., Lasser, G. W., and Mathews, C. K. (1980). T4 ribonucleotide reductase. Physical and kinetic linkage to other enzymes of deoxyribonucleotide biosynthesis. J Biol Chem *255*, 7583-7588.

Bechhofer, D. H., Hue, K. K., and Shub, D. A. (1994). An intron in the thymidylate synthase gene of *Bacillus* bacteriophage beta 22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame. Proc Natl Acad Sci U S A *91*, 11669-11673.

Beckendorf, S. K. (1973). Structure of the distal half of the bacteriophage T4 tail fiber. J Mol Biol *73*, 37-53.

Beckendorf, S. K., Kim, J. S., and Lielausis, I. (1973). Structure of bacteriophage T4 genes 37 and 38. J Mol Biol *73*, 17-35.

Belfort, M., Pedersen-Lane, J., West, D., Ehrenman, K., Maley, G., Chu, F., and Maley, F. (1985). Processing of the intron-containing thymidylate synthase (*td*) gene of phage T4 is at the RNA level. Cell *41*, 375-382.

Belfort, M., and Roberts, R. J. (1997). Homing endonucleases: keeping the house in order. Nucleic Acids Res *25*, 3379-3388.

Bell-Pedersen, D., Quirk, S., Clyman, J., and Belfort, M. (1990). Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res *18*, 3763-3770.

Bell-Pedersen, D., Quirk, S. M., Aubrey, M., and Belfort, M. (1989). A site-specific endonuclease and co-conversion of flanking exons associated with the mobile td intron of phage T4. Gene *82*, 119-126.

Belle, A., Landthaler, M., and Shub, D. A. (2002). Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev *16*, 351-362.

Benzer, S. (1955). Fine structure of a genetic region in bacteriophage. Proc Natl Acad Sci U S A *41*, 344-354.

Berglund, O. (1975). Ribonucleoside diphosphate reductase induced by bacteriophage T4. III. Isolation and characterization of proteins B1 and B2. J Biol Chem *250*, 7450-7455.

Black, L., Showe, M., and Steven, A. (1994). Morphogenesis of the T4 Head. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 218-258.

Botstein, D., and Herskowitz, I. (1974). Properties of hybrids between Salmonella phage P22 and coliphage lambda. Nature *251*, 584-589.

Bremer, H., and Dennis, P. P. (1987). Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, F. C. Neidhardt, ed. (Washington DC, American Society for Microbiology), pp. 1527-1542.

Brenner, S., Jacob, F., and Meselson, M. (1961). Unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 41-46.

Bryk, M., Belisle, M., Mueller, J. E., and Belfort, M. (1995). Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. J Mol Biol *247*, 197-210.

Burke, J. M. (1989). Selection of the 3'-splice site in group I introns. FEBS Lett *250*, 129-133.

Burke, J. M., Esherick, J. S., Burfeind, W. R., and King, J. L. (1990). A 3' splice site-binding sequence in the catalytic core of a group I intron. Nature *344*, 80-82.

Caprara, M. G., Lehnert, V., Lambowitz, A. M., and Westhof, E. (1996a). A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. Cell *87*, 1135-1145.

Caprara, M. G., Mohr, G., and Lambowitz, A. M. (1996b). A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core. J Mol Biol *257*, 512-531.

Caprara, M. G., Myers, C. A., and Lambowitz, A. M. (2001). Interaction of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) with the group I intron P4-P6 domain. Thermodynamic analysis and the role of metal ions. J Mol Biol *308*, 165-190.

Carlson, K. (2000). Virus mot bakterier. Forskning och Framsteg, 8-13.

Carlson, K., Raleigh, E., and Hattman, S. (1994). Restriction and Modification. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 369-381.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996a). Crystal structure of a group I ribozyme domain: principles of RNA packing. Science *273*, 1678-1685.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996b). RNA tertiary structure mediation by adenosine platforms. Science *273*, 1696-1699.

Cech, T. R. (1990). Self-splicing of group I introns. Annu Rev Biochem *59*, 543-568.

Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994). Representation of the secondary and tertiary structure of group I introns. Nat Struct Biol *1*, 273-280.

Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. Cell *27*, 487-496.

Celander, D. W., and Cech, T. R. (1991). Visualizing the higher order folding of a catalytic RNA molecule. Science *251*, 401-407.

Chiu, C. S., Cook, K. S., and Greenberg, G. R. (1982). Characteristics of a bacteriophage T4-induced complex synthesizing deoxyribonucleotides. J Biol Chem *257*, 15087-15097.

Christensen, A. C., and Young, E. T. (1982). T4 late transcripts are initiated near a conserved DNA sequence. Nature *299*, 369-371.

Christian, E. L., and Yarus, M. (1993). Metal coordination sites that contribute to structure and catalysis in the group I intron from *Tetrahymena*. Biochemistry *32*, 4475-4480.

Chu, F. K., Maley, G. F., Maley, F., and Belfort, M. (1984). Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc Natl Acad Sci U S A *81*, 3049-3053.

Clodi, E., Semrad, K., and Schroeder, R. (1999). Assaying RNA chaperone activity *in vivo* using a novel RNA folding trap. Embo J *18*, 3776-3782.

Coetzee, T., Herschlag, D., and Belfort, M. (1994). Escherichia coli proteins, including ribosomal protein S12, facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones. Genes Dev *8*, 1575-1588.

Collins, R. A., and Lambowitz, A. M. (1985). RNA splicing in *Neurospora* mitochondria. Defective splicing of mitochondrial mRNA precursors in the nuclear mutant cyt18-1. J Mol Biol *184*, 413-428.

Coombs, D. C., and Arisaka, F. (1994). T4 Tail Structure and Function. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 259-281.

Costa, M., and Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. EMBO J *14*, 1276-1285.

Cowie, D. B., Avery, R. J., and Champe, S. P. (1971). DNA homology among the T-even bacteriophages. Virology *45*, 30-37.

Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. Nature *192*, 1227-1232.

Curcio, M. J., and Belfort, M. (1996). Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. Cell *84*, 9-12.

Dannenberg, R., and Mosig, G. (1983). Early intermediates in bacteriophage T4 DNA replication and recombination. J Virol *45*, 813-831.

Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A., and Scazzocchio, C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. Nature *300*, 719-724.

Davies, R. W., Waring, R. B., and Towner, P. (1987). Internal guide sequence and reaction specificity of group I self-splicing introns. Cold Spring Harb Symp Quant Biol *52*, 165-171.

Dawes, J. (1976). Bacterial characteristics determining the potential host range of bacteriophage T4. J Bacteriol *127*, 1024-1025.

De Ley, J. (1970). Reexamination of the association between melting point, buoyant density, and chemical base composition of deoxyribonucleic acid. J Bacteriol *101*, 738-754.

Demerec, M., and Fano, U. (1945). Bacteriophage-resistant mutants in *Escherichia coli*. Genetics *30*, 119-136.

Desplats, C., Dez, C., Tetart, F., Eleaume, H., and Krisch, H. M. (2002). Snapshot of the genome of the pseudo-T-even bacteriophage RB49. J Bacteriol *184*, 2789-2804.

Dickson, R. C. (1974). Protein composition of the tail and contracted sheath of bacteriophage T4. Virology *59*, 123-138.

Downs, W. D., and Cech, T. R. (1994). A tertiary interaction in the *Tetrahymena* intron contributes to selection of the 5' splice site. Genes Dev *8*, 1198-1211.

Downs, W. D., and Cech, T. R. (1996). Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. RNA *2*, 718-732.

Dujon, B. (1989). Group I introns as mobile genetic elements: facts and mechanistic speculations-a review. Gene *82*, 91-114.

Eddy, S. (1992) Introns in the T-even bacteriophages. Thesis, University of Colorado, Boulder.

Eddy, S. R., and Gold, L. (1991). The phage T4 *nrdB* intron: a deletion mutant of a version found in the wild. Genes Dev *5*, 1032-1041.

Eddy, S. R., and Gold, L. (1992). Artificial mobile DNA element constructed from the EcoRI endonuclease gene. Proc Natl Acad Sci U S A *89*, 1544-1547.

Edgell, D. R., Belfort, M., and Shub, D. A. (2000). Barriers to intron promiscuity in bacteria. J Bacteriol *182*, 5281-5289.

Ehrenman, K., Schroeder, R., Chandry, P. S., Hall, D. H., and Belfort, M. (1989). Sequence specificity of the P6 pairing for splicing of the group I *td* intron of phage T4. Nucleic Acids Res *17*, 9147-9163.

Eiserling, A. F., and Black, W. L. (1994). Pathways in T4 morphology. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 209-212.

Epstein, R. H., Bolle, A., Steinberg, C., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R., Susman, M., Denhardt, C., and Lielausis, I. (1964). Physiological studies of conditional lethal mutants of bacteriophage T4D. Cold Spring Harbor Symp Quant Biol *28*, 375-392.

Foley, S., Bruttin, A., and Brussow, H. (2000). Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. J Virol *74*, 611-618.

Franklin, J. L., Haseltine, D., Davenport, L., and Mosig, G. (1998). The largest (70 kDa) product of the bacteriophage T4 DNA terminase gene 17 binds to single-stranded DNA segments and digests them towards junctions with double-stranded DNA. J Mol Biol *277*, 541-557.

Franklin, J. L., and Mosig, G. (1996). Expression of the bacteriophage T4 DNA terminase genes 16 and 17 yields multiple proteins. Gene *177*, 179-189.

Fu, T. J., Sanders, G. M., O'Donnell, M., and Geiduschek, E. P. (1996). Dynamics of DNA-tracking by two sliding-clamp proteins. EMBO J *15*, 4414-4422.

Goddard, M. R., and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci U S A *96*, 13880-13885.

Goff, C. G. (1974). Chemical structure of a modification of the *Escherichia coli* ribonucleic acid polymerase alpha polypeptides induced by bacteriophage T4 infection. J Biol Chem *249*, 6181-6190.

Goldberg, E., Grinius, L., and Letellier, L. (1994). Recognition, Attachment, and, Injection. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 347-356.

Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998). A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. Science *282*, 259-264.

Goodrich-Blair, H., Scarlato, V., Gott, J. M., Xu, M. Q., and Shub, D. A. (1990). A self-splicing group I intron in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. Cell *63*, 417-424.

Goodrich-Blair, H., and Shub, D. A. (1994). The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. Nucleic Acids Res *22*, 3715-3721.

Goodrich-Blair, H., and Shub, D. A. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. Cell *84*, 211-221.

Gott, J. M., Shub, D. A., and Belfort, M. (1986). Multiple self-splicing introns in bacteriophage T4: evidence from autocatalytic GTP labeling of RNA *in vitro*. Cell *47*, 81-87.

Gott, J. M., Zeeh, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M., and Shub, D. A. (1988). Genes within genes: independent expression of phage T4 intron open reading frames and the genes in which they reside. Genes Dev *2*, 1791-1799.

Greenberg, G. R., He, P., Hilfinger, J., and Tseng, M.-J. (1994). Deoxyribonucleoside Triphosphate Synthesis and Phage T4 DNA Replication. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 14-27.

Grosshans, C. A., and Cech, T. R. (1989). Metal ion requirements for sequence-specific endoribonuclease activity of the *Tetrahymena* ribozyme. Biochemistry *28*, 6888-6894.

Guild, N., Gayle, M., Sweeney, R., Hollingsworth, T., Modeer, T., and Gold, L. (1988). Transcriptional activation of bacteriophage T4 middle promoters by the MotA protein. J Mol Biol *199*, 241-258.

Guo, Q., and Lambowitz, A. M. (1992). A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core. Genes Dev *6*, 1357-1372.

Hambly, E., Tetart, F., Desplats, C., Wilson, W. H., Krisch, H. M., and Mann, N. H. (2001). A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc Natl Acad Sci U S A *98*, 11411-11416.

Hashemolhosseini, S., Holmes, Z., Mutschler, B., and Henning, U. (1994a). Alterations of receptor specificities of coliphages of the T2 family. J Mol Biol *240*, 105-110.

Hashemolhosseini, S., Montag, D., Kramer, L., and Henning, U. (1994b). Determinants of receptor specificity of coliphages of the T4 family. A chaperone alters the host range. J Mol Biol *241*, 524-533.

Herschlag, D. (1995). RNA chaperones and the RNA folding problem. J Biol Chem *270*, 20871-20874.

Herschlag, D., Khosla, M., Tsuchihashi, Z., and Karpel, R. L. (1994). An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. EMBO J *13*, 2913-2924.

Hershey, A. D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol *36*, 39-56.

Hinton, D. M., March-Amegadzie, R., Gerber, J. S., and Sharma, M. (1996). Characterization of pre-transcription complexes made at a bacteriophage T4 middle promoter: involvement of the T4 MotA activator and the T4 AsiA protein, a sigma 70 binding protein, in the formation of the open complex. J Mol Biol *256*, 235-248.

Horvitz, H. R. (1974a). Bacteriophage T4 mutants deficient in alteration and modification of the *Escherichia coli* RNA polymerase. J Mol Biol *90*, 739-750.

Horvitz, H. R. (1974b). Control by bacteriophage T4 of two sequential phosphorylations of the alpha subunit of *Escherichia coli* RNA polymerase. J Mol Biol *90*, 727-738.

Hughes, K. T., and Mathee, K. (1998). The anti-sigma factors. Annu Rev Microbiol *52*, 231-286.

Jacquier, A., and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell *41*, 383-394.

Jaeger, L., Westhof, E., and Michel, F. (1991). Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. J Mol Biol *221*, 1153-1164.

Kadyrov, F. A., Shlyapnikov, M. G., and Kryukov, V. M. (1997). A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages. FEBS Lett *415*, 75-80.

Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., and Rossmann, M. G. (2002). Structure of the cell-puncturing device of bacteriophage T4. Nature *415*, 553-557.

Kao, S. H., and McClain, W. H. (1980). Roles of bacteriophage T4 gene 5 and gene s products in cell lysis. J Virol *34*, 104-107.

Kassavetis, G. A., Zentner, P. G., and Geiduschek, E. P. (1986). Transcription at bacteriophage T4 variant late promoters. An application of a newly devised promoter-mapping method involving RNA chain retraction. J Biol Chem *261*, 14256-14265.

Kellenberger, E., Bolle, A., Boydelatour, E., Epstein, R. H., Franklin, N. C., Jerne, N. K., Reale-Scafati, A., and S'Echaud, J. (1965). Functions and Properties Related to the Tail Fibers of Bacteriophage T4. Virology *26*, 419-440.

Keppel, F., Lipinska, B., Ang, D., and Georgopoulos, C. (1990). Mutational analysis of the phage T4 morphogenetic 31 gene, whose product interacts with the *Escherichia coli* GroEL protein. Gene *86*, 19-25.

Kim, J. S., and Davidson, N. (1974). Electron microscope heteroduplex study of sequence relations of T2, T4, and T6 bacteriophage DNAs. Virology *57*, 93-111.

King, J., and Mykolajewycz, N. (1973). Bacteriophage T4 tail assembly: proteins of the sheath, core and baseplate. J Mol Biol *75*, 339-358.

Koch, G., and Hershey, A. D. (1959). Synthesis of phage-precursor proteins in bacteria infected with T2. J Mol Biol *1*, 260-276.

Koch, T., Raudonikiene, A., Wilkens, K., and Ruger, W. (1995). Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of *E. coli* RNA polymerase modulates T4 "early" transcription. Gene Expr *4*, 253-264.

Kowalski, J. C., and Derbyshire, V. (2002). Characterization of homing endonucleases. Methods *28*, 365-373.

Kreuzer, K. N., and Alberts, B. M. (1985). A defective phage system reveals bacteriophage T4 replication origins that coincide with recombination hot spots. Proc Natl Acad Sci U S A *82*, 3345-3349.

Kreuzer, K. N., and Morrical, S. W. (1994). Initiation of DNA Replication. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 28-42.

Kutter, E., Gachechiladze, K., Poglazov, A., Marusich, E., Shneider, M., Aronsson, P., Napuli, A., Porter, D., and Mesyanzhinov, V. (1995). Evolution of T4-related phages. Virus Genes *11*, 285-297.

Kutter, E., Guttman, B., Batts, D., Peterson, S., Djavakhishvili, T., Stidham, T., Arisaka, F., Mesyanzhinov, V., Ruger, W., and Mosig, G. (1993). Genomic map of bacteriophage T4. In Genetic maps., S. J. O´Brien, ed. (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).

Lambowitz, A. M., Caprara, M. G., Zimmerly, S., and Perlman, P. S. (1999). Group I and Group II Ribozymes as RNPs: Clues to the Past and Guides to the Future. In The RNA World, R. F. Gesteland, T. R. Cech, and J. F. Atkins, eds. (New York, Cold Spring Harbor Laboratory Press), pp. 451-485.

Landthaler, M., Begley, U., Lau, N. C., and Shub, D. A. (2002). Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of *Staphylococcus aureus* phage Twort. Nucleic Acids Res *30*, 1935-1943.

Landthaler, M., Lau, N. C., and Shub, D. A. (2004). Group I intron homing in *Bacillus phages* SPO1 and SP82: a gene conversion event initiated by a nicking homing endonuclease. J Bacteriol *186*, 4307-4314.

Landthaler, M., and Shub, D. A. (1999). Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes. Proc Natl Acad Sci U S A *96*, 7005-7010.

Landthaler, M., and Shub, D. A. (2003). The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the *Bacillus thuringiensis* phage Bastille. Nucleic Acids Res *31*, 3071-3077.

Lazarevic, V. (2001). Ribonucleotide reductase genes of *Bacillus* prophages: a refuge to introns and intein coding sequences. Nucleic Acids Res *29*, 3212-3218.

Lazarevic, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C., and Karamata, D. (1998). Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPbeta. Proc Natl Acad Sci U S A *95(4)*, 1692-1697.

Lipinska, B., Rao, A. S., Bolten, B. M., Balakrishnan, R., and Goldberg, E. B. (1989). Cloning and identification of bacteriophage T4 gene 2 product gp2 and action of gp2 on infecting DNA *in vivo*. J Bacteriol *171*, 488-497.

Liu, Q., Belle, A., Shub, D. A., Belfort, M., and Edgell, D. R. (2003). SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. J Mol Biol *334*, 13-23.

Loayza, D., Carpousis, A. J., and Krisch, H. M. (1991). Gene 32 transcription and mRNA processing in T4-related bacteriophages. Mol Microbiol *5*, 715-725.

Luder, A., and Mosig, G. (1982). Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. Proc Natl Acad Sci U S A *79*, 1101-1105.

Luria, S. E., and Human, M. L. (1952). A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol *64*, 557-569.

Macdonald, P. M., Kutter, E., and Mosig, G. (1984). Regulation of a bacteriophage T4 late gene, soc, which maps in an early region. Genetics *106*, 17-27.

Macreadie, I. G., Scott, R. M., Zinn, A. R., and Butow, R. A. (1985). Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. Cell *41*, 395-402.

Mannella, C. A., Collins, R. A., Green, M. R., and Lambowitz, A. M. (1979). Defective splicing of mitochondrial rRNA in cytochrome-deficient nuclear mutants of *Neurospora crassa*. Proc Natl Acad Sci U S A *76*, 2635-2639.

Mathews, C. K. (1993a). The cell-bag of enzymes or network of channels? J Bacteriol *175*, 6377-6381.

Mathews, C. K. (1993b). Enzyme organization in DNA precursor biosynthesis. Prog Nucleic Acid Res Mol Biol *44*, 167-203.

Matsuzaki, S., Kuroda, M., Kimura, S., and Tanaka, S. (1999). Vibriophage KVP40 and coliphage T4 genomes share a homologous 7-kbp region immediately upstream of the gene encoding the major capsid protein. Arch Virol *144*, 2007-2012.

Matsuzaki, S., Tanaka, S., Koga, T., and Kawata, T. (1992). A broad-host-range vibriophage, KVP40, isolated from sea water. Microbiol Immunol *36*, 93-97.

McGraw, P., and Tzagoloff, A. (1983). Assembly of the mitochondrial membrane system. Characterization of a yeast nuclear gene involved in the processing of the cytochrome b pre-mRNA. J Biol Chem *258*, 9459-9468.

McPheeters, D. S., Christensen, A., Young, E. T., Stormo, G., and Gold, L. (1986). Translational regulation of expression of the bacteriophage T4 lysozyme gene. Nucleic Acids Res *14*, 5813-5826.

McPheeters, D. S., Gosch, G., and Gold, L. (1988). Nucleotide sequences of the bacteriophage T2 and T6 gene 32 mRNAs. Nucleic Acids Res *16*, 9341.

Menkens, A. E., and Kreuzer, K. N. (1988). Deletion analysis of bacteriophage T4 tertiary origins. A promoter sequence is required for a rifampicin-resistant replication origin. J Biol Chem *263*, 11358-11365.

Michel, F., Ellington, A. D., Couture, S., and Szostak, J. W. (1990). Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. Nature *347*, 578-580.

Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989a). The guanosine binding site of the *Tetrahymena* ribozyme. Nature *342*, 391-395.

Michel, F., Jacquier, A., and Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie *64*, 867-881.

Michel, F., Umesono, K., and Ozeki, H. (1989b). Comparative and functional anatomy of group II catalytic introns-a review. Gene *82*, 5-30.

Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol *216*, 585-610.

Mikkonen, M., and Alatossava, T. (1995). A group I intron in the terminase gene of *Lactobacillus delbrueckii subsp. lactis* phage LL-H. Microbiology-Uk *141*, 2183-2190.

Miller, E. S., Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Durkin, A. S., Ciecko, A., Feldblyum, T. V., White, O., Paulsen, I. T., Nierman, W. C., *et al*. (2003a). Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J Bacteriol *185*, 5220-5233.

Miller, E. S., and Jozwik, C. E. (1990). Sequence analysis of conserved *regA* and variable orf*43.1* genes in T4-like bacteriophages. J Bacteriol *172*, 5180-5186.

Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003b). Bacteriophage T4 genome. Microbiol Mol Biol Rev *67*, 86-156.

Moen, L. K., Howell, M. L., Lasser, G. W., and Mathews, C. K. (1988). T4 phage deoxyribonucleoside triphosphate synthetase: purification of an enzyme complex and identification of gene products required for integrity. J Mol Recognit *1*, 48-57.

Mohr, G., Caprara, M. G., Guo, Q., and Lambowitz, A. M. (1994). A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the *Tetrahymena* ribozyme. Nature *370*, 147-150.

Mohr, G., Zhang, A., Gianelos, J. A., Belfort, M., and Lambowitz, A. M. (1992). The *Neurospora* CYT-18 protein suppresses defects in the phage T4 *td* intron by stabilizing the catalytically active structure of the intron core. Cell *69*, 483-494.

Monod, C., Repoila, F., Kutateladze, M., Tetart, F., and Krisch, H. M. (1997). The genome of the pseudo T-even bacteriophages, a diverse group that resembles T4. J Mol Biol *267*, 237-249.

Montag, D., Hashemolhosseini, S., and Henning, U. (1990). Receptor-recognizing proteins of T-even type bacteriophages. The receptor-recognizing area of proteins 37 of phages T4 TuIa and TuIb. J Mol Biol *216*, 327-334.

Mosig, G. (1994). Synthesis and Maturation of T4-Encoded tRNAs. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 182-185.

Mosig, G. (1998). Recombination and recombination-dependent DNA replication in bacteriophage T4. Annu Rev Genet *32*, 379-413.

Mosig, G., Colowick, N. E., and Pietz, B. C. (1998). Several new bacteriophage T4 genes, mapped by sequencing deletion endpoints between genes 56 (dCTPase) and dda (a DNA-dependent ATPase-helicase) modulate transcription. Gene *223*, 143-155.

Mosig, G., Gewin, J., Luder, A., Colowick, N., and Vo, D. (2001). Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. Proc Natl Acad Sci U S A *98*, 8306-8311.

Mueller, J. E., Clyman, J., Huang, Y. J., Parker, M. M., and Belfort, M. (1996). Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. Genes Dev *10*, 351-364.

Mukai, F., Streisinger, G., and Miller, B. (1967). The mechanism of lysis in phage T4-infected cells. Virology *33*, 398-404.

Murialdo, H. (1991). Bacteriophage lambda DNA maturation and packaging. Annu Rev Biochem *60*, 125-153.

Murphy, F. L., and Cech, T. R. (1993). An Independently Folding Domain of RNA Tertary Structure within the *Tetrahymena* Ribozyme. Biochemistry *32*, 5291-5300.

Murphy, F. L., and Cech, T. R. (1994). GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. J Mol Biol *236*, 49-63.

O'Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G. F., and Ross, R. P. (2004). Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content. J Bacteriol *186*, 2862-2871.

Pedersen-Lane, J., and Belfort, M. (1987). Variable occurrence of the *nrdB* intron in the T-even phages suggests intron mobility. Science *237*, 182-184.

Piccirilli, J. A., Vyle, J. S., Caruthers, M. H., and Cech, T. R. (1993). Metal ion catalysis in the *Tetrahymena* ribozyme reaction. Nature *361*, 85-88.

Price, J. V. (1987). Origin of the phosphate at the ligation junction produced by self-splicing of *Tetrahymena thermophila* pre-ribosomal RNA. J Mol Biol *196*, 217-221.

Quirk, S. M., Bell-Pedersen, D., and Belfort, M. (1989a). Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. Cell *56*, 455-465.

Quirk, S. M., Bell-Pedersen, D., Tomaschewski, J., Rüger, W., and Belfort, M. (1989b). The inconsistent distribution of introns in the T-even phages indicates recent genetic exchanges. Nucleic Acids Res *17*, 301-315.

Reddy, G. P., Singh, A., Stafford, M. E., and Mathews, C. K. (1977). Enzyme associations in T4 phage DNA precursor synthesis. Proc Natl Acad Sci U S A *74*, 3152-3156.

Repoila, F., Tetart, F., Bouet, J. Y., and Krisch, H. M. (1994). Genomic polymorphism in the T-even bacteriophages. EMBO J *13*, 4181-4192.

Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K.*, et al.* (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res *31*, 1805-1812.

Rohrer, H., Zillig, W., and Mailhammer, R. (1975). ADP-ribosylation of DNA-dependent RNA polymerase of *Escherichia coli* by an NAD+: protein ADP-ribosyltransferase from bacteriophage T4. Eur J Biochem *60*, 227-238.

Russell, R. L. (1967) Speciation among the T-even bacteriophages. Thesis, California Institute of Technology, Pasadena, California.

Russell, R. L., and Huskey, R. J. (1974). Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. Genetics *78*, 989-1014.

Saldanha, R., Ellington, A., and Lambowitz, A. M. (1996). Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection. J Mol Biol *261*, 23-42.

Saldanha, R. J., Patel, S. S., Surendran, R., Lee, J. C., and Lambowitz, A. M. (1995). Involvement of *Neurospora* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicing. Biochemistry *34*, 1275-1287.

Sandegren, L., and Sjöberg, B. M. (2004). Distribution, sequence homology and homing of group I introns among T-even like bacteriophages. Evidence for recent transfer of old introns. J Biol Chem *279*, 22218-22227.

Schwarz, H., Riede, I., Sonntag, I., and Henning, U. (1983). Degrees of relatedness of T-even type *E. coli* phages using different or the same receptors and topology of serologically cross-reacting sites. Embo J *2*, 375-380.

Selick, H. E., Stormo, G. D., Dyson, R. L., and Alberts, B. M. (1993). Analysis of five presumptive protein-coding sequences clustered between the primosome genes, 41 and 61, of bacteriophages T4, T2, and T6. J Virol *67*, 2305-2316.

Semrad, K., and Schroeder, R. (1998). A ribosomal function is necessary for efficient splicing of the T4 phage thymidylate synthase intron *in vivo*. Gene Develop *12*, 1327-1337.

Sharma, M., Ellis, R. L., and Hinton, D. M. (1992). Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proc Natl Acad Sci U S A *89(14)*, 6658-6662.

Sharma, M., and Hinton, D. M. (1994). Purification and characterization of the SegA protein of bacteriophage T4, an endonuclease related to proteins encoded by group I introns. J Bacteriol *176*, 6439-6448.

Shcherbakov, V., Granovsky, I., Plugina, L., Shcherbakova, T., Sizova, S., Pyatkov, K., Shlyapnikov, M., and Shubina, O. (2002). Focused genetic recombination of bacteriophage T4 initiated by double-strand breaks. Genetics *162*, 543-556.

Shub, D., A. (1991). The antiquity of group I introns. Curr Opin Genet Dev *1(4)*, 478-484.

Shub, D. A., Gott, J. M., Xu, M. Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J., and Belfort, M. (1988). Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc Natl Acad Sci U S A *85*, 1151-1155.

Shub, D. A., Xu, M. Q., Gott, J. M., Zeeh, A., and Wilson, L. D. (1987). A family of autocatalytic group I introns in bacteriophage T4. Cold Spring Harb Symp Quant Biol *52*, 193-200.

Sjöberg, B. M., Hahne, S., Mathews, C. Z., Mathews, C. K., Rand, K. N., and Gait, M. J. (1986). The bacteriophage T4 gene for the small subunit of ribonucleotide reductase contains an intron. EMBO J *5*, 2031-2036.

Smith, P. R., and Aebi, U. (1976). Studies of the structure of the T4 bacteriophage tail sheath. I. The recovery of three-dimensional structural information from the extended sheath. J Mol Biol *106*, 243-271.

Solem, A., Chatterjee, P., and Caprara, M. G. (2002). A novel mechanism for protein-assisted group I intron splicing. RNA *8*, 412-425.

Steitz, T. A., and Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. Proc Natl Acad Sci U S A *90*, 6498-6502.

Stent, G. S., and Wollman, E. L. (1952). On the two step nature of bacteriophage absorption. Biochim Biophys Acta *8*, 260-269.

Stitt, B., and Hinton, D. (1994). Regulation of middle mode transcription. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 142-160.

Streisinger, G., Edgar, R. S., and Denhardt, G. H. (1964). Chromosome Structure in Phage T4. I. Circularity of the Linkage Map. Proc Natl Acad Sci U S A *51*, 775-779.

Streisinger, G., Mukai, F., Dreyer, W. J., Miller, B., and Horiuchi, S. (1961). Mutations affecting the lysozyme of phage T4. Cold Spring Harb Symp Quant Biol *26*, 25-30.

Tetart, F., Desplats, C., Kutateladze, M., Monod, C., Ackermann, H. W., and Krisch, H. M. (2001). Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J Bacteriol *183*, 358-366.

Tseng, M. J., He, P., Hilfinger, J. M., and Greenberg, G. R. (1990). Bacteriophage T4 *nrdA* and *nrdB* genes, encoding ribonucleotide reductase, are expressed both separately and coordinately: characterization of the *nrdB* promoter. J Bacteriol *172*, 6323-6332.

Tseng, M. J., Hilfinger, J. M., Walsh, A., and Greenberg, G. R. (1988). Total sequence, flanking regions, and transcripts of bacteriophage T4 *nrdA* gene, coding for alpha chain of ribonucleoside diphosphate reductase. J Biol Chem *263*, 16242-16251.

van Sinderen, D., Karsens, H., Kok, J., Terpstra, P., Ruiters, M. H., Venema, G., and Nauta, A. (1996). Sequence analysis and molecular characterization of the temperate *Lactococcal* bacteriophage r1t. Mol Microbiol *19*, 1343-1355.

Wais, A. C., and Goldberg, E. B. (1969). Growth and transformation of phage T4 in Escherichia coli B-4, Salmonella, Aerobacter, Proteus, and Serratia. Virology *39*, 153-161.

Waldsich, C., Masquida, B., Westhof, E., and Schroeder, R. (2002). Monitoring intermediate folding states of the *td* group I intron in vivo. EMBO J *21*, 5281-5291.

Wallweber, G. J., Mohr, S., Rennard, R., Caprara, M. G., and Lambowitz, A. M. (1997). Characterization of *Neurospora* mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF. RNA *3*, 114-131.

Wang, J. F., Downs, W. D., and Cech, T. R. (1993). Movement of the guide sequence during RNA catalysis by a group I ribozyme. Science *260*, 504-508.

Watson, J. D., and Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. Nature *171*, 737.

Weeks, K. M., and Cech, T. R. (1995a). Efficient protein-facilitated splicing of the yeast mitochondrial bI5 intron. Biochemistry *34*, 7728-7738.

Weeks, K. M., and Cech, T. R. (1995b). Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5' splice site domain. Cell *82*, 221-230.

Weeks, K. M., and Cech, T. R. (1996). Assembly of a ribonucleoprotein catalyst by tertiary structure capture. Science *271*, 345-348.

Weinstein, L. B., Jones, B. C., Cosstick, R., and Cech, T. R. (1997). A second catalytic metal ion in group I ribozyme. Nature *388*, 805-808.

Werner, R. (1968). Distribution of growing points in DNA of bacteriophage T4. J Mol Biol *33*, 679-692.

Wheeler, L. J., Ray, N. B., Ungermann, C., Hendricks, S. P., Bernard, M. A., Hanson, E. S., and Mathews, C. K. (1996). T4 phage gene 32 protein as a candidate organizing factor for the deoxyribonucleoside triphosphate synthetase complex. J Biol Chem *271*, 11156-11162.

Wilkens, K., and Ruger, W. (1994). Transcription from early promoters. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 132-141.

Wilkens, K., and Ruger, W. (1996). Characterization of bacteriophage T4 early promoters *in vivo* with a new promoter probe vector. Plasmid *35*, 108-120.

Wilkens, K., Tiemann, B., Bazan, F., and Ruger, W. (1997). ADP-ribosylation and early transcription regulation by bacteriophage T4. Adv Exp Med Biol *419*, 71-82.

Williams, P. K., Kassavetis, A. G., Herendeen, R. D., and Geiduschek, P. (1994). Regulation of late gene expression. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 161-175.

Wood, W., Eiserling, F., and Crowther, R. (1994). Long Tail Fibers:Genes, Proteins, Structure, and Assembly. In Molecular Biology of Bacteriophage T4., J. D. Karam, ed. (Washington, DC, American Society for Microbiology), pp. 282-290.

Xu, M. Q., and Shub, D. A. (1989). The catalytic core of the *sunY* intron of bacteriophage T4. Gene *82*, 77-82.

Young, P., Öhman, M., Xu, M. Q., Shub, D. A., and Sjöberg, B. M. (1994). Intron-containing T4 bacteriophage gene *sunY* encodes an anaerobic ribonucleotide reductase. J Biol Chem *269*, 20229-20232.

Zarrinkar, P. P., and Williamson, J. R. (1996). The kinetic folding pathway of the *Tetrahymena* ribozyme reveals possible similarities between RNA and protein folding. Nat Struct Biol *3*, 432-438.

Zhang, A., Derbyshire, V., Salvo, J. L., and Belfort, M. (1995). *Escherichia coli* protein StpA stimulates self-splicing by promoting RNA assembly in vitro. RNA *1*, 783-793.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S., and Lambowitz, A. M. (1995). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell *83*, 529-538.