



HAL
open science

Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes

David Bikard, Céline Loot, Zeynep Baharoglu, Didier Mazel

► **To cite this version:**

David Bikard, Céline Loot, Zeynep Baharoglu, Didier Mazel. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. *Microbiology and Molecular Biology Reviews*, 2010, 74 (4), pp.570-588. 10.1128/MMBR.00026-10 . pasteur-02505407

HAL Id: pasteur-02505407

<https://pasteur.hal.science/pasteur-02505407>

Submitted on 1 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Folded DNA in action: hairpin formation and biological functions in prokaryotes

David Bikard^{1,2}, Céline Loot^{1,2}, Zeynep Baharoglu^{1,2} and Didier Mazel^{1,2}

¹ Institut Pasteur, Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, F-75015 Paris, France

² CNRS, URA2171, F-75015 Paris, France

Correspondence should be addressed to Didier Mazel

(mazel@pasteur.fr) Tel 33 1 40 61 32 84, Fax 33 1 45 68 88 34

Running title: Biological functions of hairpin DNA in prokaryotes

Subject category: Genome stability & dynamics, Microbiology & Pathogens

1		
2	INTRODUCTION.....	4
3	DNA HAIRPIN FORMATION.....	5
4	Hairpin formation from ssDNA	5
5	Formation of ssDNA through horizontal gene transfer.	5
6	(i) Conjugation.....	5
7	(ii) Transformation	7
8	(iii) Phage infection.....	7
9	Macromolecule synthesis and repair.	7
10	(i) Transcription	7
11	(ii) Replication.	8
12	(iii) DNA repair.....	8
13	Single-strand binding proteins.	9
14	Cruciform extrusion.....	9
15	Mechanism of cruciform extrusion.	9
16	Regulation of cruciform extrusion.....	10
17	Effect of cruciform extrusion on DNA topology dynamics.	11
18	Genetic instability of inverted repeats.....	12
19	DNA HAIRPIN BIOLOGICAL FUNCTION	13
20	Hairpins and replication origins.....	13
21	Priming on single strand.	13
22	(i) G4 type priming.....	13
23	(ii) ϕ X174-type priming.	14
24	(iii) Filamentous phage type priming.....	14
25	Double-strand DNA replication.	14
26	(i) Chromosomal and theta replication.	15
27	(ii) Strand displacement replication.	15
28	(iii) Rolling circle replication (RCR).	16
29	Hairpins and transcription	16
30	Hairpin promoters.	16
31	Promoter inhibition through cruciform extrusion.....	17

1	Hairpins and conjugation	18
2	Hairpins and recombination	18
3	The single-stranded CTX phage of <i>Vibrio cholerae</i>	19
4	The IS200/IS605 insertion sequence family.....	19
5	The IS91 insertion sequence.....	19
6	Integrans.	20
7	Other hairpin DNA: phage packaging, retrons, etc.....	20
8	Single-stranded phage packaging.....	20
9	Retrons.....	20
10	HAIRPIN FORMATION: CRUCIFORM EXTRUSION VS. SINGLE-STRANDED HAIRPIN	21
11	PROTEIN / HAIRPIN RECOGNITION	22
12	Mimicry: subverting the host proteins.....	22
13	Protein recognition of hairpin features	23
14	Strand selectivity	24
15	EVOLUTION OF HAIRPINS WITH BIOLOGICAL FUNCTIONS.....	25
16	Single-stranded DNA, stress and horizontal transfer.....	25
17	On the origins of folded DNA binding proteins.....	27
18	RCR Rep proteins, relaxases and IS608 transposase.....	27
19	Integron integrases.	28
20	N4 vRNAP.	28
21	CONCLUSION	29
22	ACKNOWLEDGMENTS	30
23	REFERENCES.....	30
24	FIGURES LEGENDS	44
25	SUMMARY	47
26		
27		

INTRODUCTION

The B-helix form of DNA proposed by Watson and Crick accounts for most of DNA's behavior in the cell. Nevertheless, it is now obvious that DNA isn't always present in this canonical structure, but can also form alternative structures such as Z-DNA, cruciforms, triple-helix H-DNA, quadruplex G4-DNA and slipped-strand DNA (147). This review focuses on DNA hairpins, i.e. DNA with intrastrand base pairing, their functions and properties, in light of the specific behavior of DNA in horizontal gene transfer between bacterial cells.

Hairpin structures can be formed by sequences with inverted repeats (IRs), also termed palindromes, following two main mechanisms. Firstly, in several cellular processes, DNA is single-stranded (ssDNA); for instance, on the lagging strand of replication, during DNA repair or, more importantly, during rolling circle replication, bacterial conjugation, natural transformation and virus infection. ssDNA is not simply a transient inert state of DNA, but can fold into secondary structures recognized by proteins, notably involved in site-specific recombination, transcription and replication. A second mechanism is the formation of hairpins from double-stranded DNA (dsDNA) as a cruciform, i.e. two opposite hairpins extruding through intrastrand base pairing from a palindromic sequence. The existence of cruciforms was already hypothesized soon after Watson and Crick's discovery (119): negative supercoiling of double-stranded DNA (dsDNA) could provide free energy to stabilize cruciforms that would otherwise be unstable. Cruciforms then attracted much attention in the 1980's when their existence was experimentally assessed *in vitro* under natural superhelical densities (117). But most studies at that time rejected their possible implication in cellular processes because of the slow kinetics of cruciform formation, which made them theoretically very unlikely to occur *in vivo* (25, 130). Nonetheless, this point of view was revised when techniques revealing cruciforms *in vivo* were developed and biological functions involving DNA secondary structures were discovered.

There are three ways in which DNA hairpins can interact with proteins and impact cell physiology. (i) Cruciform formation modifies the coiling state of DNA (143), which is known to affect the binding of

1 regulatory proteins for transcription, recombination and replication (26)(54); (ii) the DNA-protein interaction
2 can be inhibited if a hairpin overlaps a protein recognition site (65). (iii) Proteins can directly recognize and
3 bind DNA hairpins (8, 48, 97, 100, 139).

4 We describe here the cellular processes leading to DNA hairpin formation, biological functions
5 involving hairpins, and the mechanisms of protein-hairpin recognition. Finally, we try to shed light on the
6 evolution of folded DNA with biological functions and their cognate proteins.

7 **DNA HAIRPIN FORMATION**

8 **Hairpin formation from ssDNA**

9 The production of a large amount of single-stranded DNA (ssDNA) in the cell occurs mainly during the
10 entry of exogenous DNA, macromolecular synthesis and repair. The three mechanisms of DNA uptake,
11 namely, natural transformation, conjugation and, occasionally, bacteriophage infection, involve the
12 production of ssDNA (Figure 1). The processes of replication and transcription also involve the unwinding of
13 duplex DNA; finally, DNA repair can lead to the production of large quantities of ssDNA. The amount of
14 single strand available, its lifetime and the bound proteins are different properties of these processes that may
15 affect the possibility of hairpins to fold.

16 **Formation of ssDNA through horizontal gene transfer.**

17 **(i) Conjugation.** Conjugation is the process by which one bacterium can actively transfer DNA to a
18 neighboring cell (Figure 1). The mechanism of conjugation is conserved across all described systems. A
19 protein called relaxase binds and nicks a cognate origin-of-transfer site (*oriT*). This reaction results in a
20 complex between the relaxed plasmid and the relaxase (together with accessory factors), called the
21 relaxosome. Only the strand that is covalently bound by the relaxase is transferred to the recipient cell as
22 ssDNA. The transferred strand (T-strand) is excreted from the donor cell through the type IV secretion

1 system and the relaxase then directs recircularization of the T-strand in the recipient cell (for a
2 comprehensive review, see (4)). Two main families of conjugative elements have been described: self-
3 transmissible plasmids and “integrative and conjugative elements” (ICEs). ICEs cannot autonomously
4 replicate and are thus carried by chromosomes or other replicons. These elements are able to excise
5 themselves as circular intermediates through the action of a recombinase/excisionase and are then transferred
6 following the same mechanism. In the recipient cell, they can be integrated through homologous
7 recombination or through the action of a site-specific recombinase (14, 70). The length of the DNA molecule
8 that is transferred is usually the size of the whole conjugative element (usually <200kb).

9 Occasionally, chromosomal DNA can be transferred. This happens when conjugative plasmids are
10 integrated into the chromosome, a famous example being the plasmid F/Hfr system (96, 137). Alternatively,
11 the conjugation functions carried by ICEs can also promote transfer of chromosomal or plasmid DNA, as
12 demonstrated for the SXT element in *Vibrio cholerae* (60). In this case, the length of the transferred strand is
13 limited by the conjugation bridge strength and the contact time between the bacteria. Since the time of early
14 genetic mapping through Hfr conjugation of the *Escherichia coli* chromosome by Nelson, we have learned
15 that it takes about 100 min to transfer the whole *E.coli* chromosome (4.6 Mb) (111). Although very long
16 DNA fragments can be transferred, the average length of ssDNA region is unknown. Indeed, the ssDNA
17 length and its lifetime depend on the speed of complementary strand synthesis in the recipient strain. The
18 only direct data available comes from microscopy experiments enabling visualization of complementary
19 strand synthesis and showing that synthesis already starts within 5 min after the donor and recipient cells are
20 mixed (6). Nevertheless, the number of ssDNA replication origins is unknown in most cases. Single-stranded
21 origins of replication have been studied in the case of rolling-circle replication, which is discussed later (II.1).
22 The fact that specific origins of replication have evolved for initiation of complementary strand synthesis
23 suggests that this process does not happen easily at random sequences. It is therefore unlikely that
24 complementary strand synthesis is initiated at numerous loci. Conjugation thus massively produces ssDNA
25 and conjugative plasmids are probably a place of choice for the evolution of functions where hairpins are
26 involved. Indeed, the very process of conjugation, for instance, implies DNA secondary structures (48) (see

1 “Hairpin and conjugation”).

2 **(ii) Transformation.** Bacterial competence for natural transformation is a physiological state that
3 permits uptake and incorporation of naked exogenous DNA (Figure 1). Many Gram-negative bacteria
4 (species of *Haemophilus*, *Neisseria*, *Helicobacter*, *Vibrio* and *Acinetobacter*) as well as Gram-positive
5 bacteria (species of *Bacillus*, *Mycobacterium* and *Streptomyces*) are capable of natural competence. In all
6 cases, one strand of the transformed DNA is degraded, providing the energy for transport of the
7 complementary strand across the cytoplasmic membrane (20). Some bacteria have been shown to fragment
8 exogenous DNA so that they take only small bits, while others can take up long DNA molecules (37).
9 Monitoring of ssDNA fate during transformation in *Streptococcus pneumoniae* revealed that ssDNA does
10 not subsist in the cell more than 15 min (104). Globally, the length of the incoming DNA and the lifetime of
11 ssDNA in the recipient cell are probably shorter than for conjugation. The entering single strand is protected
12 from the action of nucleases essentially by the binding of SSB (22), whereas, during conjugation, the relaxase
13 is covalently bound to the T-strand, effectively protecting it from exonucleases. However, in some bacteria
14 including *B.subtilis* and *S.pneumoniae*, a protein named DprA has been found to bind the incoming ssDNA,
15 protecting it from both endo- and exonucleases and facilitating further homologous recombination (108). All
16 in all, during transformation, ssDNA is not long-lived in the cell; it is either quickly integrated into the
17 chromosome through homologous recombination or it is degraded.

18 **(iii) Phage infection.** Single-stranded phages encapsidate their genome and deliver it to newly infected
19 cells in this form (Figure 1). The maximum amount of DNA that can be transferred is equivalent to the size
20 of the phage genome (generally <10 kb), but here again, little is known about the timing of complementary
21 strand synthesis. Nevertheless, hairpins have been found to play important roles at all steps of ssDNA phage
22 life cycles, from synthesis of the complementary strand (88, 144) to phage DNA encapsidation (125) (see the
23 part “DNA hairpin biological functions”).

24 **Macromolecule synthesis and repair.**

25 **(i) Transcription.** RNA synthesis requires the opening of the DNA duplex. The size of the

1 transcription bubble ranges between 12 and 25 bp, covered by the transcription complex (44). This small
2 opening leaves very little room for secondary structure formation, and transcription is thus unlikely to foster
3 hairpin formation. On the contrary, the transcription bubble needs to unfold hairpins that it may encounter so
4 as to enable production of the correct transcripts by the RNA polymerase (RNAP).

5 **(ii) Replication.** In contrast to transcription, DNA synthesis produces large amounts of ssDNA. Firstly,
6 the replication initiation step often requires melting of a large DNA region around the origin of replication.
7 Multiple hairpins have been found to play important roles at replication origins (16, 99) (see the part
8 “Hairpins and replication origins”). Secondly, lagging strand replication is not continuous and an ssDNA
9 loop is formed to place the DNA in the correct orientation for DNA polymerase. Half of the replication loop
10 consists of nascent Okazaki fragment and the other half of ssDNA extruded by the helicase (Figure 2). In
11 E.coli, Okazaki fragments are 1 kb to 2 kb nucleotides long, and the replication fork speed is about 1 kb.s⁻¹
12 in optimal conditions (78). The lifetime of ssDNA should thus be on the order of a second. Evidence that
13 inverted repeats (IRs) can fold into stable hairpins in vivo during replication came from the observation that
14 large and perfect IRs are genetically unstable on plasmids in E.coli. Indeed, they are the cause of mismatched
15 alignment or slippage during replication (91, 131). In particular, deletions of IRs occur preferentially on the
16 lagging strand (138).

17 Finally, a special mode of replication, called rolling circle replication (RCR), involves unwinding of
18 the full lagging strand into ssDNA (75). Multiple hairpins have been found to play important roles in RCR
19 (82, 83, 85, 113) (see Figure 5).

20 **(iii) DNA repair.** A major source of ssDNA in the cell is through DNA repair. Double-strand breaks
21 are processed by the RecBCD enzyme which produces ssDNA tails through its exonuclease activity. These
22 ssDNA tails can then be bound by RecA and may be involved in homologous strand invasion and replication-
23 dependent repair (79, 80, 86). Double-strand breaks can be caused by many agents, including ionizing
24 radiation, UV light and oxygen radicals, but in normally growing cells as well, double-strand breaks are
25 formed in almost every cell cycle as a consequence of replication through imperfect DNA templates (for a

1 comprehensive review, see (36)).

2 It has also been shown that when replication forks encounter a lesion, the replication of the lagging and
3 leading strands can be uncoupled in order to bypass the lesion, leaving ssDNA gaps on the damaged strand
4 (55, 90, 116). These gaps are around 1 kb in length and can be processed by RecA-mediated recombinational
5 repair (Figure 2).

6 **Single-strand binding proteins.** In all these processes, ssDNA in the cell is not left naked. Several
7 proteins bind ssDNA without sequence specificity. The most important ones are the RecA and single-strand
8 binding (SSB) protein. SSB coats any ssDNA present in the cell and prevents intrastrand pairing, i.e. hairpin
9 formation. The RecA protein also binds ssDNA forming a straight nucleoproteic filament. RecA can then
10 promote strand invasion of homologous dsDNA and catalyze recombination (79). Furthermore, SSB directs
11 RecA binding to ssDNA (81, 122). Recent single molecule studies have shown how tetrameric SSB can
12 spontaneously migrate along ssDNA, melting unstable hairpins while stimulating RecA filament elongation
13 (124).

14 Although ssDNA is present on many occasions in the cell, hairpin formation is strongly constrained by
15 SSB and RecA binding. Proteins that ensure their function through hairpin binding are thus in competition
16 with SSB and RecA for substrate availability. Hairpins that are formed need to be stable enough to resist SSB
17 melting and coating. For instance, it was demonstrated that SSB can inhibit the activity of the plasmid pT181
18 RepC protein at secondary cleavage sites on ssDNA, but not at its primary binding site (76) (see the part
19 “Rolling Circle Replication”).

20

21 **Cruciform extrusion**

22 **Mechanism of cruciform extrusion.** The formation of DNA hairpins in the cell does not
23 necessarily require the production of ssDNA. Extrusion of cruciforms occurs through the opening of the

1 DNA double helix to allow intrastrand base pairing. Base opening in relaxed DNA is both infrequent and
2 transient. However, negatively supercoiled DNA molecules are much more active, because their topology
3 facilitates both large- and small-scale opening of the double helix (42). Two main mechanisms for cruciform
4 extrusion have been proposed (Figure 3)(92). The first (type S) implies small-scale melting of the double
5 helix at the dyad of the IR (~10bp). This small opening allows a few bases to pair with their cognate base in
6 the repeat. The stem can then be elongated through branch migration, which is also facilitated by negative
7 supercoiling. The other mechanism (type C) involves the melting of a large region, which is favored by
8 nearby AT-rich sequences. This large melting would allow hairpins to fold on both strands leading to
9 cruciform formation (Figure 3). The S-type mechanism is highly dependent on the IR sequence (it is favored
10 by the AT-rich sequence at the dyad), and works in physiological ionic conditions (133). On the other hand,
11 C-type extrusion takes place in low-salt solutions and is highly dependent on the presence of AT-rich
12 neighbor sequences, but should theoretically be suppressed at physiological ion concentrations (110).
13 Nevertheless, this mechanism could possibly take place in DNA regions with propensities to undergo
14 substantial denaturation, such as replication origins.

15

16 **Regulation of cruciform extrusion.** Cruciforms were extensively studied in the 1980's when
17 techniques enabling their observation *in vitro* were developed, such as S1 sensitivity and 2D electrophoresis.
18 Although cruciform extrusion can be energetically favorable under moderate superhelical densities, the slow
19 kinetics of cruciform extrusion raises questions as to their relevance *in vivo* (25). However, several
20 techniques later developed led to the demonstration of cruciform formation *in vivo* under natural superhelical
21 densities (33, 34, 64, 112). In particular, cruciforms that were tuned to fold stably at different superhelical
22 densities have even been used to measure the natural superhelix densities of plasmids. *In vivo* cross-linking
23 with psoralen demonstrated that the propensity of an IR to fold into a cruciform strongly depends on its
24 sequence and context, and that some IRs can exist as cruciforms at levels as high as 50% in plasmids in
25 living *E.coli* cells (148, 149).

1 Nevertheless, most reported cruciform detection involved artificial conditions favoring hairpin
2 extrusion: small loops, IR in AT-rich regions, perfect palindromes with AT-rich centers and GC-rich stems,
3 *topA* background or salt shock to increase supercoiling (131, 149, 150). Random IRs do not seem to fold
4 cruciforms at significant rates under average *in vivo* supercoiling. However, many factors may transiently
5 increase local superhelical density to a critical level sufficient for cruciform extrusion (see review (118)).
6 Biological processes such as transcription and replication may generate local and temporal domains of
7 supercoiling on circular DNA (34, 93, 128). Indeed, during replication and transcription, enzymes alter the
8 structure of DNA, such that additional twists are added (positive supercoiling) or subtracted (negative
9 supercoiling). Negative supercoiling favors the unwinding of the DNA double helix, which is required for
10 initiation of transcription and replication processes (59, 120). As transcription proceeds, DNA in front of the
11 transcription machinery becomes positively supercoiled, and DNA behind becomes negatively supercoiled.
12 Similarly, during replication, strand separation by the helicase leads to positive supercoiling of the duplex
13 ahead of the fork (see review (128)).

14 Changes in supercoiling in response to external and/or internal stimuli could also play a significant role
15 in the formation and stability of cruciforms. In *E.coli*, superhelicity has been shown to vary considerably
16 during cell growth and to change under different growth conditions (7, 68). Moreover, topology analysis of
17 reporter plasmids isolated from strains where the SOS stress response regulon is constitutively expressed
18 revealed higher levels of negative supercoiling (98). Finally, the level of superhelicity is known to be
19 variable between bacterial strains. For instance, the average supercoiling density of a pBR322 reporter
20 plasmid extracted from mid-log cultures of WT *Salmonella* is 13% lower ($\sigma=-0.060$) than that from *E.coli*
21 ($\sigma=-0.069$) (18).

22 **Effect of cruciform extrusion on DNA topology dynamics.** The positioning of IRs within
23 topological domains appears to be another parameter that influences cruciform extrusion. Studies involving
24 visualization of the cruciform on supercoiled plasmids through atomic force microscopy have shown that
25 extrusion is favored when IRs are positioned at the apex of a plectonemic supercoil (115). Furthermore,

1 cruciforms can exist in two distinct conformations, an X-type conformation and a planar conformation. In the
2 X-type conformation, the cruciform arms form an acute angle and the main DNA strand is sharply bent,
3 whereas in the planar conformation, the arms are present at an angle of 180° (129). It has been shown that
4 the rest of the DNA molecule is deeply affected by the conformation adopted by the cruciform. X-type
5 cruciforms tend to localize at the apex of the plectonemic supercoil and restrict slithering of the molecule, i.e.
6 they reduce the possibility of distant sites coming into contact. Environmental conditions, such as salt
7 concentration and protein binding, are factors influencing the conformation choice. For instance, the RuvA
8 protein tetramer which binds to the Holliday junction at the base of cruciforms forces them into a planar
9 conformation in which the constraints upon DNA movements are relieved (129). It has thus been proposed
10 that cruciform extrusion may act as a molecular switch that can control DNA transactions between distant
11 sites. Such long-range contacts are known to be essential in many cellular processes, including site-specific
12 recombination, transposition or control of gene expression through DNA-loop formation (1, 46, 94, 127).

13

14

Genetic instability of inverted repeats

15 It was quickly noticed that long palindromes are impossible to maintain in vivo (for a review, see (91)),
16 either because they are not genetically stable and will be partially mutated or deleted, or because they are not
17 viable, i.e. the molecule carrying them cannot be replicated (24). It is assumed that instability and inviability
18 are caused by the inability of the replication fork to process secondary structures that are too stable, and by
19 the presence of proteins destroying these structures. In particular, the SbcCD enzyme can cleave hairpins
20 forming on ssDNA, leading to double-strand breaks that are then repaired by recombination (17, 27). This
21 leads to constraints on the size and perfection of the inverted repeats that can be maintained in vivo.
22 Typically, a size of 150-200 bp is a limit for IRs, although the presence of mismatches and spacers between
23 the repeats strongly improves their maintenance. However, a mutation mechanism was identified, which
24 tends to restore perfection to quasi-palindromes during chromosomal replication (38). The model proposes

1 that during replication, the nascent DNA strand dissociates from its template strand, forming a partial hairpin
2 loop structure. The nascent strand is then extended by DNA synthesis from the hairpin template, forming a
3 more fully paired hairpin. IRs are thus balanced between a mechanism that tends to perfection and the fact
4 that perfect IRs are not genetically stable.

5

6

DNA HAIRPIN BIOLOGICAL FUNCTION

7

Hairpins and replication origins

8 Hairpins play an essential and common role in replication initiation. Indeed, they have been found to be
9 indispensable for initiation of complementary strand synthesis on single-stranded phages as well as for
10 replication of dsDNA replicons, in particular, during rolling circle replication.

11 **Priming on single strand.** The first evidence for the role of DNA hairpins in a biological function
12 came from the early studies of the primosome. The inability of DNA polymerases to initiate *de novo*
13 replication makes the independent generation of a primer necessary (78). The primosome is a complex of
14 proteins which carries out this priming through *de novo* synthesis of a small RNA whose 3' end can be used
15 by the DNA polymerase as a starting point. The role of RNA in priming DNA replication was discovered
16 primarily through studies of single-stranded phages, notably G4 and ϕ X174 (88, 144). Single-stranded
17 phages are delivered to the infected cells and have evolved diverse mechanisms for priming synthesis of the
18 complementary strand, but all the strategies described to date involve DNA hairpins.

19 **(i) G4 type priming.** Phage G4 carries, in the region of replication initiation, three hairpins with stems
20 of 5 to 19 bp and loops of 4 to 8 bases. Early models invoked these structures as recognition sites for the
21 primase, DnaG (88). However, it was later shown that none of these hairpins are required for DnaG to initiate
22 primer synthesis in the absence of SSB in *E.coli* (135). The hairpins seem, in fact, to direct the binding of
23 SSB so that primase recognition site 5'-CTG-3' is exposed (134). This mechanism is likely to be at stake for a

1 large number of G4-like phages, including $\alpha 3$, St-1 and ϕK . This is an illustration of how hairpins can direct
2 protein binding and structure an ssDNA region (Figure 4).

3 (ii) **$\phi X174$ -type priming.** Although $\phi X174$ is a close relative of G4, the priming mechanism leading
4 to complementation of ssDNA cannot be realized by DnaG alone. The PriA protein, which is now known to
5 play a major role in stalled replication fork restart, was first identified as an essential component of the
6 $\phi X174$ primosome (144). It catalyzes priming from a specific primosome assembly site (PAS) which can
7 adopt a stable secondary structure (5). However, it is now clear that the main PriA substrates are not PAS
8 sites but D-loops and R-loops encountered during replication, DNA repair and recombination events. It has
9 thus been proposed that PAS sequences have evolved to mimic the natural targets of PriA (103). A stem-loop
10 formed on a single strand can indeed be viewed as a branched structure between a double strand and two
11 single-strand components (a Y-fork). PriA was recently shown to bind Y-forks (136). This is an illustration
12 of hairpins that have evolved to be recognized by a host protein, to direct primosome assembly (Figure 4).

13 (iii) **Filamentous phage type priming.** In the case of the M13 phage and other filamentous phages (f1
14 and fd), synthesis of the complementary strand is primed neither by DnaG nor PriA, but by the host RNA
15 polymerase (RNAP) holoenzyme containing the sigma70 subunit which synthesizes a 20 nt long RNA primer
16 (57, 71). The RNAP recognizes a double hairpin structure mimicking a promoter with a -35 and a -10 box
17 (56) (Figure 4). Here again, hairpins have evolved to be recognized by a host protein. Hairpins recognized by
18 the RNAP have now been associated with several functions (see 2.a).

19 **Double-strand DNA replication.** The first step in dsDNA replication is the melting of a region
20 where the replication priming complex can load. This melting event is favored, with some exceptions, by a
21 complex of proteins (DnaA for the chromosome, or Rep for plasmids), which binds the DNA (usually at
22 direct repeats: DnaA boxes or iterons) and bends it (72, 77, 109). This bending promotes DNA melting, but
23 also formation of alternative DNA structures.

24 A common feature of many origins of replication is the presence of inverted repeats (IRs). The

1 extrusion of IRs as cruciforms is energetically more favorable than the simple DNA melting and is thus very
2 likely to occur, absorbing a part of the strain generated. Furthermore, when DNA melting actually occurs
3 (which is favored by AT- rich regions present in most *oris*) IRs are free to fold into hairpins. There is thus
4 ample opportunity at origins of replication for a DNA structure to arise and interact with proteins.

5 As a matter of fact, hairpins have also been shown to play essential roles in primosome assembly in
6 dsDNA replication. The generation of a primer occurs in two major ways: opening of the DNA double helix
7 followed by RNA priming (chromosomal, theta and strand displacement replications) or cleavage of one of
8 the DNA strands to generate a 3'-OH end (rolling-circle replication (RCR)) (35, 75). In both mechanisms,
9 cases where hairpins play essential roles have been described.

10 **(i) Chromosomal and theta replication.** The DnaA protein plays a central role in the replication of
11 the bacterial chromosome and of several plasmids. It is involved in the control of replication initiation,
12 unwinding of the helix and recruitment of the priming complex (for a review see (109)). It has been proposed
13 that in some replication origins, a hairpin structure carrying a DnaA box folds in the region unwound by
14 DnaA itself. This hairpin, named M13-A, is at the core of the ABC priming mechanism first described for the
15 R6K plasmid (101). M13-A is specifically bound by DnaA, which then recruits DnaB, DnaC and finally
16 initiates RNA priming. This mechanism was later proposed to occur at the *E.coli* origin of replication (16),
17 and putative M13-A hairpins are present in a large number of theta-replicating plasmids.

18 Inverted repeats other than M13-A and called single-stranded initiators (*ssi*) are often present at
19 replication origins and can be involved in RNA priming. In the same way that filamentous phages prime
20 complementary strand synthesis, the F plasmid origin of replication has a hairpin (*ssiD* or *Frpo*) recognized
21 by *E.coli* RNAP which synthesizes an RNA primer (100). Other *ssi* have been isolated from a variety of
22 plasmids and shown to use a ϕ X174 type priming involving PriA (for a review, see (99)).

23 **(ii) Strand displacement replication.** The best described example of strand displacement replication is
24 plasmid RSF1010. The plasmid-encoded RepC protein binds to iterons and unwinds the DNA in a region
25 carrying two single-stranded initiators (*ssiA* & *ssiB*). These sequences are IRs which fold into hairpins. The

1 secondary structures of these hairpins and parts of their sequences have been shown to be essential for
2 replication (106). The current model states that plasmid-encoded RepB primase specifically recognizes *ssiA*
3 and *ssiB* and primes continuous replication from these sequences (61-63). However, it is not clear whether
4 *ssiA* and *ssiB* fold when the region is largely single-stranded or whether they extrude as a cruciform, thanks
5 to the action of RepC.

6 **(iii) Rolling circle replication (RCR).** RCR is widely present among plasmids and viruses (including
7 the filamentous phages previously mentioned), with the model being plasmid pT181 (for a review see, (75)).
8 The plasmid-encoded Rep protein binds to the double-stranded origin of replication (*dso*) and bends the
9 DNA, producing a strain leading to the extrusion of a hairpin carrying the Rep nicking site. This structure
10 was among the first cruciforms probed in vivo (113). Rep nicks DNA in the hairpin and becomes covalently
11 attached to the 5' phosphate (Figure 5). The free 3'-OH end serves as the primer for leading strand synthesis.
12 No synthesis occurs on the lagging strand until it is completely unwound by the helicase and released as
13 ssDNA. The synthesis of the complementary strand is then initiated at the single-strand origin (*sso*). Four
14 classes of *sso* have been described (*ssoA*, *ssoW*, *ssoT* and *ssoU*). These classes have little nucleotide sequence
15 homology, but share structural features (82) necessary for their recognition by the host RNA polymerase
16 which primes complementary strand synthesis (82, 84, 85).

17 **Hairpins and transcription**

18 There are essentially three ways in which hairpins and cruciforms can affect transcription. (i) The
19 extrusion of a cruciform dramatically reduces the local supercoiling of DNA. Since superhelical density is
20 known to affect the activity of promoters, cruciform extrusions in promoter regions could reduce their
21 activity (142). (ii) A cruciform could prevent proteins from binding to their cognate site if it overlaps the
22 extruding sequence. (iii) RNA polymerases or transcription factors could recognize hairpins present on
23 ssDNA or extruded from dsDNA. Since there is as yet no documented case for the first possibility, only the
24 two other mechanisms are discussed here.

25 **Hairpin promoters.** We have already discussed how the RNAP can recognize hairpin promoters to

1 prime DNA replication (rolling circle replication / filamentous phage type priming / F plasmid replication).
2 The RNAP primes F plasmid replication through recognition of the *Frpo* hairpin, but under certain
3 conditions, it can produce transcripts longer than the one needed for priming and express the downstream
4 genes (100).

5 Accordingly, transcription from a structured single-stranded promoter was suggested to occur during
6 conjugative DNA transfer for several *oriT*-associated genes of enterobacterial conjugative plasmids, namely
7 *ssb*, *psiB* and sometimes *ardA*. Considering that conjugation consists of ssDNA entry into the recipient cell,
8 the product of these genes - respectively single-strand binding, anti-SOS and anti-restriction - could be
9 needed for maintaining the plasmid in the recipient. Indeed, the transcriptional orientation of these genes,
10 always on the leading strand, means that the transferred strand is destined to be the transcribed strand (21).
11 Moreover, conjugative induction of these first loci so as to enter the recipient bacterium was shown to be
12 transfer-dependent (69). The burst of activity observed shortly after initiation of conjugation led to the
13 proposal that this early transcription could be mediated by the presence of a secondary structure in the
14 transferred ssDNA (3, 114) that mimics an RNA polymerase promoter recognized by the *Frpo* sigma factor
15 (100).

16 Other hairpin promoters which are not involved in priming have been described. Notably, the N4 virion
17 carries three hairpin promoters specifically recognized by the virion RNA polymerase (vRNAP) and used to
18 direct the transcription of the phage early genes (Figure 6). Upon infection of *E.coli*, the N4 double-stranded
19 DNA injected into the cell is supercoiled by the host DNA gyrase, which leads to the extrusion of hairpin
20 promoters as cruciforms (28, 29).

21 **Promoter inhibition through cruciform extrusion.** Early studies have shown how an artificial
22 IR overlapping a promoter can regulate transcription by superhelix-induced cruciform formation (64).
23 Although promoters usually have higher activity with increasing superhelix density, such a promoter has a
24 lower expression level at high superhelix density because of the extrusion of the IR as a cruciform preventing
25 RNAP binding. It has also been shown that the N4 hairpin placed between the -10 and -35 boxes of the *rrnB*

1 P1 promoter can repress its activity in a supercoil-dependent manner (28). DNA cruciform extrusion seems
2 likely to be a mechanism for the regulation of genes repressed by supercoiling. However, it is not clear how
3 common this mechanism of regulation is, since no compelling natural example has been reported. The *bgl*
4 operon promoter, which presents a 13bp IR, was first thought to be a natural example of such regulation
5 (132). However, it was later shown that no cruciform is required to account for its supercoiling-dependent
6 repression (15).

7 **Hairpins and conjugation**

8 IRs are present in a majority of origins of transfers (*oriT*) (40). The best described is the origin of
9 transfer of R388, where an IR named IR2 located 5' to the nicking site plays an essential role (49).
10 Conjugation occurs as follow: DNA is nicked at the *oriT* and bound covalently by the plasmid-encoded
11 relaxase protein TrwC. The T-strand is then unwound through rolling circle replication and transferred to the
12 recipient cell. Although the folding of IR2 into a hairpin is not required for the initial nicking of the *oriT*, the
13 recircularization of the T-strand requires folding of IR2 into a hairpin specifically recognized by the relaxase
14 (48).

15 In addition to IR2, other IRs important for transfer efficiency are present in R388 *oriT* (95), but their
16 exact role remains to be elucidated. It is not yet known whether their sequence or structure is important. They
17 probably help adapt *oriT* into a potentially active state through cruciform formation.

18 Two relaxases other than TrwC have been crystallized: the F plasmid relaxase TraI (31) and the R1162
19 plasmid relaxase MobA (107). Although they show poor sequence homology to TrwC, the 3D structure of all
20 this relaxases is very similar. These enzymes are evolutionarily homologous to certain identical mechanisms
21 of action.

22 **Hairpins and recombination**

23 To date, there are three compelling examples of recombination systems using DNA hairpins as
24 substrates: the CTX phage recombination site, the IS200/IS605 insertion sequence family, and integron *attC*

1 recombination sites.

2 **The single-stranded CTX phage of *Vibrio cholerae*.** CTX is a single-stranded phage involved
3 in *V. cholerae* virulence. In lysogenic phase, it integrates the *V. cholerae* chromosome I or II at its respective
4 *dif1* and *dif2* sites. Chromosomal *dif* sites are recombination sites recognized by the XerCD protein complex
5 which solves concatemers and allows proper chromosome segregation. CTX enters the infected cells as
6 ssDNA, and the single-stranded form is directly integrated into one of the chromosomes (140). The *attP*
7 recombination site of CTX carries a ~150 bp forked hairpin, which is homologous to *dif* sites (Figure 7). The
8 phage uses this hairpin to hijack the host XerCD protein complex which catalyzes a strand exchange between
9 *attP* and the *dif* site (30).

10 **The IS200/IS605 insertion sequence family.** The mechanism of transposition of the recently
11 discovered IS200/IS605 insertion sequence family greatly differs from systems already described, in
12 particular those using DDE transposase catalysis (50). The best studied representative of this family, IS608,
13 was originally identified in *H. pylori* (74). It presents at its ends short palindromes recognized as hairpins by
14 the TnpA transposase. “Top strands” of the two IS ends are nicked and joined together by TnpA a few base
15 pairs away from the hairpins (19 nt upstream from the left hairpin and 10 nt downstream from the right
16 hairpin) (8, 53). TnpA then catalyzes the formation of a single-stranded transposon circle intermediate which
17 is then inserted specifically into a single-stranded target. This target site is not recognized directly by TnpA,
18 but by four bases at the foot of the hairpin in the transposition circle (Figure 8 and (52)) that realize
19 unconventional base pairing with the ssDNA target sequence.

20 **The IS91 insertion sequence.** IS91 is a member of an insertion sequence family displaying a
21 unique mechanism of transposition. The IS91 transposase is related to replication proteins of RCR plasmids.
22 IS91 transposition involves an ssDNA intermediate generated in a rolling circle fashion (105). Short
23 palindromes have been identified in the regions essential for transposition just a few base pairs away from the
24 recombination sites. Their exact functions have not been studied. Nevertheless, striking similarities between
25 these regions, RCR plasmids *dso* and conjugation *oriTs* suggest that these palindromes might fold into

1 hairpins recognized by the IS91 transposase.

2 **Integrans.** Integrans are natural recombination platforms able to stockpile, shuffle and differentially
3 express gene cassettes. Discovered by virtue of their importance in multiple antibiotic resistances, they were
4 later identified in 10% of sequenced bacterial chromosomes, where they can contain hundreds of cassettes
5 (11). The cassettes are generally single ORFs framed by *attC* recombination sites (121). When expressed, the
6 integron integrase can recombine *attC* sites leading to excision of a circular cassette. Such a cassette can then
7 be integrated at a primary recombination site named *attI*. *attC* recombination sites have been shown to be
8 recognized and recombined by the integrase only as hairpins (Figure 9) (12, 102). A surprising feature of
9 *attC* hairpins is their huge polymorphism. Their stem length ranges from 54 to 80 bp and their loop length
10 from 3 to 80 bp. Highly conserved mismatches known to be involved in hairpin recognition by the integrase
11 are also present (12, 13) (see the part “Strand selectivity”).

12 **Other hairpin DNA: phage packaging, retrons, etc.**

13 **Single-stranded phage packaging.** The single-stranded filamentous phages (f1, fd, M13, Ike)
14 contain IRs that can fold into hairpins. We have already described the hairpins involved in complementary
15 strand synthesis, but the largest hairpin identified on these genomes is the packaging signal (PS) recognized
16 in translocation of ssDNA into the virion capsid. This hairpin is probably recognized by the phage
17 transmembrane protein pI and determines the orientation of DNA within the particle (125). Both the structure
18 and sequence determinants of the PS-hairpin are required for its function (126).

19 **Retrons.** Retrons are DNA sequences found in the genomes of a wide variety of bacteria (89). They
20 code for a reverse transcriptase similar to that produced by retroviruses and other types of retro-elements.
21 They are responsible for synthesis of an unusual satellite DNA called msDNA (multicopy single-stranded
22 DNA). msDNA is a complex of DNA, RNA and probably protein. It is composed of a small single-stranded
23 DNA linked to a small single-stranded RNA molecule folded together into a secondary structure. msDNA is
24 produced in many hundreds of copies per cell (89). Whether msDNA are selfish elements or play a role in the
25 cell remains to be discovered.

1 large spacer sequences (80 bp) between the repeats were also able to fold cruciform structures. Interestingly,
2 it was noted that the recombinogenic strand of *attC* sites is always found on the leading strand of replication
3 in natural integrons. Under such conditions, the most probable pathways for *attC* hairpin formation are
4 through ssDNA generated by repair or cruciform extrusion. It has been observed that integron cassettes are
5 particularly AT-rich (102), which could favor *attC* site extrusion following a C-type mechanism. Although it
6 is not yet known whether the SOS response triggers IS608 movements, we know that this is the case for other
7 classes of insertion sequences such as IS10 (2, 39).

8 To summarize, large perfect IRs can presumably fold into cruciforms but are genetically unstable
9 because of their propensity to hinder replication and be cleaved by SbcCD. Small perfect (or almost-perfect)
10 IRs can fold into cruciforms only when their sequence and context allow it. The N4 promoters and pT181
11 plasmid origin of replication are examples of such IRs with biological functions. Imperfect IRs are
12 genetically more stable regardless of their size, but fold into cruciforms only rarely. They could still be
13 involved in biological functions that take place at low frequencies such as integrons or IS608 recombination.
14 Alternatively, imperfect IRs present in topologically constrained regions such as replication origins could
15 also fold into cruciforms, which might be the case for the M13-A hairpin and for the *ssi* present in some
16 origins of replication. Note that these hairpins are specifically bound by cognate proteins that could stabilize
17 cruciforms.

18 **PROTEIN / HAIRPIN RECOGNITION**

19 **Mimicry: subverting the host proteins**

20 Some of the hairpins described in the literature have evolved to mimic the "natural" target of the
21 protein they interact with. The PAS sequences of single-stranded phages mimic Y-forks that are recognized
22 by PriA. The *sso* of RCR plasmids, the *Frpo* hairpin and the filamentous phages priming hairpins all mimic
23 promoters recognized by the host RNAP. The M13-A hairpin mimics a natural *dnaA* box and the CTX *attP*

1 has been observed: the protein recognizes one strand and not the other. In light of the hairpin/protein
2 interactions described above, it is easy to understand how proteins discriminate between the two strands.
3 They all show base-specific interactions with bases either in the loop, at the single-stranded base of the stem
4 or with extrahelical bases. Any of these interactions can account for strand selectivity. Some of these systems
5 appear to have good reason to process one strand and not the other. The N4 virion needs to initiate
6 transcription in the right direction. Recombination of the wrong strand for integron cassettes would lead to
7 their integration in the wrong direction, where they could not be transcribed. Finally, if a different strand of
8 IS608 is recognized at each end of the transposon, this would lead to the junction of the top strand with the
9 bottom strand, a configuration that cannot be processed further and is likely to be lethal. Therefore, one
10 strand had to be chosen and the other strongly discriminated against.

11 **EVOLUTION OF HAIRPINS WITH BIOLOGICAL FUNCTIONS**

12 A variety of hairpins have been selected to be recognized by host proteins, especially in single-stranded
13 phages and plasmids. The single-stranded nature of DNA during transfer of mobile elements drove the
14 evolution of secondary structures able to hijack the host cell machinery. The use of host priming proteins,
15 host RNAP or even host recombinases incites single-stranded phages not to bring additional proteins with
16 them and still be processed into a replicative form. Similarly, when a quick reaction is required upon transfer,
17 ssDNA hairpins are the best elements for driving the response, as exemplified by the hairpin promoters
18 present on several conjugative plasmids. We first discuss how horizontal gene transfer, the presence of
19 ssDNA in the cell and the SOS response are interrelated. Secondly, we briefly review the origin of those
20 proteins that have evolved to specifically use hairpin DNA as their substrate.

21

22 **Single-stranded DNA, stress and horizontal transfer**

23 We have seen that hairpin formation in the cell is most likely to occur in the presence of ssDNA in the

1 cell. Intracellular single-stranded DNA triggers the SOS response (Figure 10). ssDNA is the substrate for
2 RecA polymerization. The formation of a RecA nucleofilament on ssDNA stimulates self-cleavage of the
3 general repressor LexA, leading to its inactivation. Promoters from the SOS regulon, controlling mostly
4 DNA repair, recombination and mutagenic polymerases, are than de-repressed (Figure 10).

5 SOS is thus induced when an abnormal amount of ssDNA is present in the cell. The formation of
6 hairpins from ssDNA is thus likely to occur in a context where the SOS response is activated. Induction of
7 the SOS response is often synonymous with stress. This happens, for example, when the cell tries to replicate
8 damaged DNA, causing replication forks to stall (141). Another source of ssDNA comes from DNA intake
9 by horizontal gene transfer and phage infection. For instance, conjugative transfer of R plasmids -
10 conjugative plasmids carrying multiple resistances - has been shown to induce the SOS stress response in the
11 recipient cell, except when an anti-SOS factor is encoded by the plasmid (*psiB*, already mentioned in the part
12 “Hairpins and transcription”) (Z. Baharoglu, D. Bikard, D. Mazel, submitted for publication). Interestingly,
13 the expression of these anti-SOS genes is under control of ssDNA promoters, i.e. of hairpin substrates.

14 Furthermore, in the case of integrons, expression of the integrase (*intI*) has recently been shown to be
15 controlled by SOS (51). Some antibiotics are known to induce the SOS response in Gram-negative and
16 Gram-positive bacteria (73). These antibiotics, such as quinolones, trimethoprim and beta-lactams, were
17 tested and found to be inducers of expression of the *intI* promoter. This is certainly a way for integrons to
18 "know" when potential substrates are present in the cell and to recombine them. Indeed, the induction of SOS
19 during conjugative transfer of R plasmids results in induction of the integrase, allowing genome
20 rearrangements in the recipient bacterium (Z. Baharoglu, D. Bikard, D. Mazel, submitted for publication).
21 Furthermore, integrons are often found on conjugative plasmids and may well take advantage of the single-
22 stranded transfer to acquire cassettes and spread horizontally. Similarly, for IS608, specific integration into
23 the ssDNA substrate has been proposed as a mechanism for targeting mobile elements and ensuring
24 interbacterial spread (53).

25 Not only does the SOS response promote genetic rearrangments, but it also induces horizontal gene

1 transfer. It is known, for instance, that stress can induce competence in some bacteria (23) (Figure 10).
2 Another effect of SOS induction is the derepression of genes involved in the single-stranded transfer of
3 integrating conjugative elements (ICEs), such as SXT from *V.cholerae*, which is a ~100 kb ICE that transfers
4 and integrates the recipient bacteria's genome, conferring resistance to several antibiotics (9). Moreover,
5 different ICEs are able to combine and create their own diversity in a RecA-dependent manner (i.e. using
6 homologous recombination, which is also induced by SOS) (45, 145). As for R plasmids, SXT transfer was
7 observed to induce SOS in *V.cholerae*. Finally, some lysogenic phages are also known to induce their lytic
8 phase under stressful conditions (43). One might thus see the use of ssDNA by integrons and other
9 recombination systems as a mechanism for evolving: diversity is generated under stressful conditions.

10 **On the origins of folded DNA binding proteins**

11 While, in many examples described above, one can see that hairpins evolved to subvert the host
12 machinery, in other instances, proteins evolved to specifically and sometimes exclusively recognize hairpin
13 structures. This is the case for the RCR Rep proteins, the relaxases of conjugative elements, the transposase
14 of IS608, the integron integrases and phage N4 vRNAP. Where do these proteins come from and what
15 pushed them to recognize ssDNA rather than dsDNA?

16 **RCR Rep proteins, relaxases and IS608 transposase.** Interestingly, the IS608 transposase as
17 well as conjugative relaxases have been found to be structurally similar to RCR Rep proteins (123). All of
18 these proteins have in common the use of a tyrosine residue to covalently bind DNA. The Rep proteins
19 belong to a vast superfamily spanning eubacteria, archae and eukaryotes (67). The superfamily is
20 characterized by two sequence motifs: an HUH motif (histidine-hydrophobic residue-histidine) presumed to
21 ligate a Mg²⁺ ion and required for nicking, and a YxxxY motif where the tyrosines (Y) bind the DNA
22 covalently, with one of the tyrosine being optional. All these proteins thus probably have a common ancestor,
23 ancient enough to account for the diversity of their functions and their spread among the kingdoms of life.
24 The ability to bind hairpin DNA might have been an important feature in early stages of life when single-

1 stranded DNA might have been more widely present. In this instance, the relaxases of conjugative plasmids
2 obviously need to recognize ssDNA features to process the ssDNA in the recipient cell. Recombination of
3 ssDNA by the IS608 transposase is probably a way to target mobile elements and to ensure their spread.
4 Finally, the reason why RCR plasmid Rep proteins would recognize hairpins rather than the more stable
5 dsDNA is probably that origins of replications need to be strongly negatively coiled to unwind the double
6 helix, and under these conditions, hairpins can be the most stable conformation of DNA.

7 **Integron integrases.** Integron integrases (IntI) are also tyrosine recombinases covalently binding
8 DNA. However, they are not related to the Rep protein superfamily. The closest relatives to integron
9 integrases are the XerCD proteins. However, IntI proteins carry an additional domain, compared to XerCD.
10 This domain is involved in binding of the extrahelical bases of the *attC* hairpins that are essential for strand
11 selectivity (13, 97). It would be tempting to speculate that integrons diverged from a single-stranded CTX-
12 like phage which already used XerCD to recombine hairpin DNA. This special feature of ssDNA
13 recombination would then have been selected to form an evolving recombination platform, thanks to its
14 ability to sense both stressful conditions and the occurrence of horizontal gene transfer.

15 **N4 vRNAP.** N4 vRNAP is an evolutionarily highly divergent member of the T7 family of RNAPs
16 (32). N4 vRNAP and T7 RNAP recognize their promoter with similar domains and motifs. However, N4
17 vRNAP recognizes a hairpin, whereas T7 RNAP recognizes dsDNA. The difference lies in the domain
18 interacting with the hairpin loop. It displays substantial architectural complexity and base-specific
19 interactions for N4 vRNAP, whereas the same domain in its counterpart just fits an AT-rich DNA sequence
20 without base recognition (19). The reason why the N4 phage has evolved to transcribe several genes only
21 from cruciform promoters is unclear. It is likely a way for the virion to sense the coiling state of DNA in the
22 cell, which is known to be modified during the cell cycle and is particularly negative during the SOS stress
23 response (98).

CONCLUSION

The use of DNA hairpins in biological processes is ubiquitous in prokaryotes and their viruses. How do these hairpins arise from duplex DNA? Numerous cellular processes lead to the formation of ssDNA, notably replication and the mechanisms of horizontal gene transfer, but also DNA damage and repair. Furthermore, the implication of cruciform DNA has been demonstrated at the RCR *dso* and for N4 phage promoters. Nevertheless, functions associated with cruciforms do not seem to be widely spread due to the slow kinetics of cruciform formation. However, cruciforms might play a role in special cases, but the difficulty of probing them *in vivo* makes these events underestimated. In eukaryotes, cruciform binding proteins have recently been identified and are suggested to play a major role in genome translocation (87) and replication initiation (146).

Not surprisingly, single-stranded phages have been found to use DNA hairpins at almost every step of their life cycle: complementary strand synthesis, replication, integration into the host chromosome and packaging. But hairpins play a role in the replication of a much larger number of elements, probably including the origin of replication of *E.coli*.

A striking feature is the opportunism of single-stranded DNA in subverting host machinery. The three different mechanisms of complementary strand synthesis have evolved hairpins directing priming by three different host proteins (DnaG, PriA, RNAP) in three different ways. Another example of opportunistic use of host machinery is the CTX phage which integrates *V. cholerae* chromosome I through a hairpin mimicking the XerCD recombination site. Also, the variety of hairpins recognized by the RNAP, either for replication priming or for transcription leads to the perception of ssDNA as evolutionarily very flexible.

Finally, the evolution of functions involving ssDNA is deeply intertwined with horizontal gene transfer, response to stress and genome plasticity. Horizontal gene transfers lead to ssDNA production and involve functions requiring hairpins. Together with stresses that also generate ssDNA, they activate the SOS response and trigger systems involved in genome plasticity, some of which use hairpin DNA, such as IS608 or integrons. To close the loop, the SOS response can trigger more horizontal transfer, notably through

1 activation of natural transformation, ICE conjugation and lysogenic phages.

2 The cases discussed above illustrate at least three different families of proteins in which specific
3 hairpin binding activities have independently evolved. It thus seems quite easy both for proteins to evolve
4 hairpin binding activity and for hairpins to evolve in such a way that they can exploit host proteins. Hairpin
5 recognition can be seen as a way for living systems to expand the repertoire of information storage in DNA
6 beyond the primary base sequence. These hairpin recognition examples illustrate how DNA can carry
7 information via its conformation. Finally, this review is probably not exhaustive, as new functions in which
8 folded DNA plays a role most likely remain to be discovered.

9 ACKNOWLEDGMENTS

10 The authors acknowledge Marie-Eve Val and Pr Adam Wilkins for helpful discussions. This study was
11 carried out with financial assistance from the Institut Pasteur, the Centre National de la Recherche
12 Scientifique (CNRS-URA 2171), the French National Research Agency (ANR-08-MIE-016), the EU (NoE
13 EuroPathoGenomics, LSHB-CT-2005-512061) and the Fondation pour la Recherche Médicale (programme
14 “équipe FRM” 2007).

16 REFERENCES

- 17 1. **Adhya, S.** 1989. Multipartite genetic control elements: communication by DNA loop. *Annu Rev*
18 *Genet* **23**:227-50.
- 19 2. **Aleshkin, G. I., K. V. Kadzhaev, and A. P. Markov.** 1998. High and low UV-dose responses in
20 SOS-induction of the precise excision of transposons tn1, Tn5 and Tn10 in *Escherichia coli*. *Mutat*
21 *Res* **401**:179-91.
- 22 3. **Althorpe, N. J., P. M. Chilley, A. T. Thomas, W. J. Brammar, and B. M. Wilkins.** 1999.
23 Transient transcriptional activation of the IncI1 plasmid anti-restriction gene (*ardA*) and SOS
24 inhibition gene (*psiB*) early in conjugating recipient bacteria. *Mol Microbiol* **31**:133-42.

- 1 4. **Alvarez-Martinez, C. E., and P. J. Christie.** 2009. Biological diversity of prokaryotic type IV
2 secretion systems. *Microbiology and molecular biology reviews: MMBR* **73**:775.
- 3 5. **Arai, K., and A. Kornberg.** 1981. Unique primed start of phage phi X174 DNA replication and
4 mobility of the primosome in a direction opposite chain synthesis. *Proceedings of the National*
5 *Academy of Sciences of the United States of America* **78**:69.
- 6 6. **Babic, A., A. B. Lindner, M. Vulic, E. J. Stewart, and M. Radman.** 2008. Direct Visualization
7 of Horizontal Gene Transfer. *Science* **319**:1533.
- 8 7. **Balke, V. L., and J. D. Gralla.** 1987. Changes in the linking number of supercoiled DNA
9 accompany growth transitions in *Escherichia coli*. *J Bacteriol* **169**:4499-506.
- 10 8. **Barabas, O., D. R. Ronning, C. Guynet, A. B. Hickman, B. Ton-Hoang, M. Chandler, and**
11 **F. Dyda.** 2008. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-
12 directed target site selection. *Cell* **132**:208.
- 13 9. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal
14 dissemination of antibiotic resistance genes. *Nature* **427**:72-4.
- 15 10. **Benham, C. J., A. G. Savitt, and W. R. Bauer.** 2002. Extrusion of an imperfect palindrome to
16 a cruciform in superhelical DNA: complete determination of energetics using a statistical
17 mechanical model. *J Mol Biol* **316**:563-81.
- 18 11. **Boucher, Y., J. E. Koenig, H. W. Stokes, and M. Labbate.** 2007. Integrons: mobilizable
19 platforms that promote genetic diversity in bacteria. *Trends in microbiology* **15**:301.
- 20 12. **Bouvier, M., G. Demarre, and D. Mazel.** 2005. Integron cassette insertion: a recombination
21 process involving a folded single strand substrate. *Embo J* **24**:4356-67.
- 22 13. **Bouvier, M., M. Ducos-Galand, C. Loot, D. Bikard, and D. Mazel.** 2009. Structural features
23 of single-stranded integron cassette attC sites and their role in strand selection. *PLoS Genet*
24 **5**:e1000632.
- 25 14. **Burrus, V., and M. K. Waldor.** 2004. Shaping bacterial genomes with integrative and
26 conjugative elements. *Res Microbiol* **155**:376-86.
- 27 15. **Caramel, A., and K. Schnetz.** 1998. Lac and lambda repressors relieve silencing of the
28 *Escherichia coli* bgl promoter. Activation by alteration of a repressing nucleoprotein complex.
29 *Journal of molecular biology* **284**:875.

- 1 16. **Carr, K. M., and J. M. Kaguni.** 2002. Escherichia coli DnaA protein loads a single DnaB helicase
2 at a DnaA box hairpin. The Journal of biological chemistry **277**:39815.
- 3 17. **Chalker, A. F., D. R. Leach, and R. G. Lloyd.** 1988. Escherichia coli sbcC mutants permit
4 stable propagation of DNA replicons containing a long palindrome. Gene **71**:201.
- 5 18. **Champion, K., and N. P. Higgins.** 2007. Growth rate toxicity phenotypes and homeostatic
6 supercoil control differentiate Escherichia coli from Salmonella enterica serovar Typhimurium. J
7 Bacteriol **189**:5839-49.
- 8 19. **Cheetham, G. M., and T. A. Steitz.** 1999. Structure of a transcribing T7 RNA polymerase
9 initiation complex. Science **286**:2305-9.
- 10 20. **Chen, I. s., P. J. Christie, and D. Dubnau.** 2005. The ins and outs of DNA transfer in bacteria.
11 Science (New York, N.Y.) **310**:1456.
- 12 21. **Chilley, P. M., and B. M. Wilkins.** 1995. Distribution of the ardA family of antirestriction genes
13 on conjugative plasmids. Microbiology (Reading, England) **141**:2157.
- 14 22. **Claverys, J. P., B. Martin, and P. Polard.** 2009. The genetic transformation machinery:
15 composition, localization, and mechanism. FEMS Microbiol Rev **33**:643-56.
- 16 23. **Claverys, J. P., M. Prudhomme, and B. Martin.** 2006. Induction of competence regulons as a
17 general response to stress in gram-positive bacteria. Annu Rev Microbiol **60**:451-75.
- 18 24. **Collins, J., G. Volckaert, and P. Nevers.** 1982. Precise and nearly-precise excision of the
19 symmetrical inverted repeats of Tn5; common features of recA-independent deletion events in
20 Escherichia coli. Gene **19**:139.
- 21 25. **Courey, A. J., and J. C. Wang.** 1983. Cruciform formation in a negatively supercoiled DNA may
22 be kinetically forbidden under physiological conditions. Cell **33**:817.
- 23 26. **Cozzarelli, N. R., and J. C. Wang.** 1990. DNA Topology and Its Biological Effects, vol. 20. Cold
24 Spring Harbor Laboratory Pr.
- 25 27. **Cromie, G. A., C. B. Millar, K. H. Schmidt, and D. R. Leach.** 2000. Palindromes as substrates
26 for multiple pathways of recombination in Escherichia coli. Genetics **154**:513-22.
- 27 28. **Dai, X., M. B. Greizerstein, K. Nadas-Chinni, and L. B. Rothman-Denes.** 1997. Supercoil-
28 induced extrusion of a regulatory DNA hairpin. Proc Natl Acad Sci U S A **94**:2174-9.
- 29 29. **Dai, X., and L. B. Rothman-Denes.** 1998. Sequence and DNA structural determinants of N4

- 1 virion RNA polymerase promoter recognition. *Genes & Development* **12**:2782.
- 2 30. **Das, B., J. Bischerour, M.-E. Val, and F. o.-X. Barre.** 2010. Molecular keys of the tropism of
3 integration of the cholera toxin phage. *Proceedings of the National Academy of Sciences of the*
4 *United States of America* **107**:4377.
- 5 31. **Datta, S., C. Larkin, and J. F. Schildbach.** 2003. Structural Insights into Single-Stranded DNA
6 Binding and Cleavage by F Factor TraI. *Structure* **11**:1369.
- 7 32. **Davydova, E. K., I. Kaganman, K. M. Kazmierczak, and L. B. Rothman-Denes.** 2009.
8 Identification of bacteriophage N4 virion RNA polymerase-nucleic acid interactions in
9 transcription complexes. *J Biol Chem* **284**:1962-70.
- 10 33. **Dayn, A., S. Malkhosyan, D. Duzhy, V. Lyamichev, Y. Panchenko, and S. Mirkin.** 1991.
11 Formation of (dA-dT)_n cruciforms in *Escherichia coli* cells under different environmental
12 conditions. *J Bacteriol* **173**:2658-64.
- 13 34. **Dayn, A., S. Malkhosyan, and S. M. Mirkin.** 1992. Transcriptionally driven cruciform
14 formation in vivo. *Nucleic Acids Res* **20**:5991-7.
- 15 35. **del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas.** 1998.
16 Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* **62**:434-64.
- 17 36. **Dillingham, M. S., and S. C. Kowalczykowski.** 2008. RecBCD enzyme and the repair of
18 double-stranded DNA breaks. *Microbiology and molecular biology reviews: MMBR* **72**:642.
- 19 37. **Dubnau, D.** 1999. DNA uptake in bacteria. *Annu Rev Microbiol* **53**:217-44.
- 20 38. **Dutra, B. E., and S. T. Lovett.** 2006. Cis and trans-acting effects on a mutational hotspot
21 involving a replication template switch. *J Mol Biol* **356**:300-11.
- 22 39. **Eichenbaum, Z., and Z. Livneh.** 1998. UV light induces IS10 transposition in *Escherichia coli*.
23 *Genetics* **149**:1173-81.
- 24 40. **Francia, M. V., A. Varsaki, M. P. Garcillán-Barcia, A. Latorre, C. Drinas, and F. de la**
25 **Cruz.** 2004. A classification scheme for mobilization regions of bacterial plasmids. *FEMS*
26 *microbiology reviews* **28**:79.
- 27 41. **Francia, M. V., J. C. Zabala, F. de la Cruz, and J. M. Garcia-Lobo.** 1999. The IntI1 integron
28 integrase preferentially binds single-stranded DNA of the *attC* site. *Journal of Bacteriology*
29 **181**:6844-6849.

- 1 42. **Furlong, J. C., K. M. Sullivan, A. I. Murchie, G. W. Gough, and D. M. Lilley.** 1989. Localized
2 chemical hyperreactivity in supercoiled DNA: evidence for base unpairing in sequences that
3 induce low-salt cruciform extrusion. *Biochemistry* **28**:2009.
- 4 43. **Galkin, V. E., X. Yu, J. Bielnicki, D. Ndjonka, C. E. Bell, and E. H. Egelman.** 2009. Cleavage
5 of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *J Mol Biol* **385**:779-
6 87.
- 7 44. **Gamper, H. B., and J. E. Hearst.** 1982. A topological model for transcription based on
8 unwinding angle analysis of E. coli RNA polymerase binary, initiation and ternary complexes. *Cell*
9 **29**:81-90.
- 10 45. **Garriss, G., M. K. Waldor, and V. Burrus.** 2009. Mobile antibiotic resistance encoding
11 elements promote their own diversity. *PLoS Genet* **5**:e1000775.
- 12 46. **Gellert, M., and H. Nash.** 1987. Communication between segments of DNA during site-specific
13 recombination. *Nature* **325**:401-4.
- 14 47. **Gleghorn, M. L., E. K. Davydova, L. B. Rothman-Denes, and K. S. Murakami.** 2008.
15 Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA
16 polymerase. *Molecular cell* **32**:707.
- 17 48. **Gonzalez-Perez, B., M. a. Lucas, L. A. Cooke, J. S. Vyle, Fernando, and G. Moncali  n.**
18 2007. Analysis of DNA processing reactions in bacterial conjugation by using suicide
19 oligonucleotides. *The EMBO journal* **26**:3847.
- 20 49. **Guasch, A., M. Lucas, G. Moncalian, M. Cabezas, R. Perez-Luque, F. X. Gomis-Ruth, F.**
21 **de la Cruz, and M. Coll.** 2003. Recognition and processing of the origin of transfer DNA by
22 conjugative relaxase TrwC. *Nat Struct Biol* **10**:1002-10.
- 23 50. **Gueguen, E., P. Rousseau, G. Duval-Valentin, and M. Chandler.** 2005. The transpososome:
24 control of transposition at the level of catalysis. *Trends Microbiol* **13**:543-9.
- 25 51. **Guerin, E., J. Barb  , G. Cambray, M.-C. c. Ploy, N. Sanchez-Alberola, S. Campoy, I.**
26 **Erill, Sandra, B. Gonzalez-Zorn, and D. Mazel.** 2009. The SOS response controls integron
27 recombination. *Science (New York, N.Y.)* **324**:1034.
- 28 52. **Guynet, C., A. Achard, B. T. Hoang, O. Barabas, A. B. Hickman, F. Dyda, and M.**
29 **Chandler.** 2009. Resetting the site: redirecting integration of an insertion sequence in a

- 1 predictable way. *Molecular cell* **34**:612.
- 2 53. **Guynet, C., A. B. Hickman, O. Barabas, F. Dyda, M. Chandler, and B. Ton-Hoang.** 2008.
3 In vitro reconstitution of a single-stranded transposition mechanism of IS608. *Mol Cell* **29**:302-
4 12.
- 5 54. **Hatfield, G. W., and C. J. Benham.** 2002. DNA topology-mediated control of global gene
6 expression in *Escherichia coli*. *Annual review of genetics* **36**:175.
- 7 55. **Heller, R. C., and K. J. Marians.** 2006. Replication fork reactivation downstream of a blocked
8 nascent leading strand. *Nature* **439**:557.
- 9 56. **Higashitani, A., N. Higashitani, and K. Horiuchi.** 1997. Minus-strand origin of filamentous
10 phage versus transcriptional promoters in recognition of RNA polymerase. *Proceedings of the*
11 *National Academy of Sciences of the United States of America* **94**:2909.
- 12 57. **Higashitani, N., A. Higashitani, Z. W. Guan, and K. Horiuchi.** 1996. Recognition
13 mechanisms of the minus-strand origin of phage f1 by *Escherichia coli* RNA polymerase. *Genes*
14 *to cells: devoted to molecular & cellular mechanisms* **1**:829.
- 15 58. **Higashitani, N., A. Higashitani, A. Roth, and K. Horiuchi.** 1992. SOS induction in
16 *Escherichia coli* by infection with mutant filamentous phage that are defective in initiation of
17 complementary-strand DNA synthesis. *J Bacteriol* **174**:1612-8.
- 18 59. **Hirose, S., and K. Matsumoto.** 2005. Possible roles of DNA supercoiling in transcription, p.
19 138-143. *In* T. Ohshima (ed.), *DNA conformation and transcription*, vol. XII. Springer.
- 20 60. **Hochhut, B., J. Marrero, and M. K. Waldor.** 2000. Mobilization of plasmids and chromosomal
21 DNA mediated by the SXT element, a constant found in *Vibrio cholerae* O139. *Journal of*
22 *bacteriology* **182**:2043.
- 23 61. **Honda, Y., H. Sakai, H. Hiasa, K. Tanaka, T. Komano, and M. Bagdasarian.** 1991.
24 Functional division and reconstruction of a plasmid replication origin: molecular dissection of the
25 oriV of the broad-host-range plasmid RSF1010. *Proceedings of the National Academy of Sciences*
26 *of the United States of America* **88**:179.
- 27 62. **Honda, Y., H. Sakai, and T. Komano.** 1988. Two single-strand DNA initiation signals located in
28 the oriV region of plasmid RSF1010. *Gene* **68**:221.
- 29 63. **Honda, Y., H. Sakai, T. Komano, and M. Bagdasarian.** 1989. RepB' is required in trans for

- 1 the two single-strand DNA initiation signals in oriV of plasmid RSF1010. Gene **80**:155.
- 2 64. **Horwitz, M. S., and L. A. Loeb.** 1988. An E. coli promoter that regulates transcription by DNA
3 superhelix-induced cruciform extrusion. Science **241**:703-5.
- 4 65. **Horwitz, M. S., and L. A. Loeb.** 1988. An E. coli promoter that regulates transcription by DNA
5 superhelix-induced cruciform extrusion. Science (New York, N.Y.) **241**:703.
- 6 66. **Huang, X., F. J., and J. D. Helmann.** 1997. sigma factor mutations affecting the sequence-
7 selective interaction of RNA polymerase with -10 region single-stranded DNA. Nucleic acids
8 research **25**:2603.
- 9 67. **Ilyina, T. V., and E. V. Koonin.** 1992. Conserved sequence motifs in the initiator proteins for
10 rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and
11 archaeobacteria. Nucleic Acids Res **20**:3279-85.
- 12 68. **Jaworski, A., N. P. Higgins, R. D. Wells, and W. Zacharias.** 1991. Topoisomerase mutants
13 and physiological conditions control supercoiling and Z-DNA formation in vivo. J Biol Chem
14 **266**:2576-81.
- 15 69. **Jones, A. L., P. T. Barth, and B. M. Wilkins.** 1992. Zygotic induction of plasmid *ssb* and *psiB*
16 genes following conjugative transfer of *Incl1* plasmid Collb-P9. Molecular microbiology **6**:605.
- 17 70. **Juhas, M., D. W. Crook, and D. W. Hood.** 2008. Type IV secretion systems: tools of bacterial
18 horizontal gene transfer and virulence. Cellular microbiology **10**:2377.
- 19 71. **Kaguni, J. M., and A. Kornberg.** 1982. The rho subunit of RNA polymerase holoenzyme
20 confers specificity in priming M13 viral DNA replication. The Journal of biological chemistry
21 **257**:5437.
- 22 72. **Katayama, T., S. Ozaki, K. Keyamura, and K. Fujimitsu.** Regulation of the replication cycle:
23 conserved and diverse regulatory systems for DnaA and oriC. Nat Rev Microbiol **8**:163-70.
- 24 73. **Kelley, W. L.** 2006. Lex marks the spot: the virulent side of SOS and a closer look at the LexA
25 regulon. Mol Microbiol **62**:1228-38.
- 26 74. **Kersulyte, D., B. Velapatino, G. Dailide, A. K. Mukhopadhyay, Y. Ito, L. Cahuayme, A. J.**
27 **Parkinson, R. H. Gilman, and D. E. Berg.** 2002. Transposable element ISHp608 of
28 *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion
29 specificity. J Bacteriol **184**:992-1002.

- 1 75. **Khan, S. A.** 2005. Plasmid rolling-circle replication: highlights of two decades of research.
2 Plasmid **53**:126-36.
- 3 76. **Koepsel, R. R., and S. A. Khan.** 1987. Cleavage of single-stranded DNA by plasmid pT181-
4 encoded RepC protein. Nucleic acids research **15**:4085.
- 5 77. **Konieczny, I.** 2003. Strategies for helicase recruitment and loading in bacteria. EMBO Rep
6 **4**:37-41.
- 7 78. **Kornberg, A., and T. A. Baker.** 1992. DNA Replication. W.H. Freeman & Company.
- 8 79. **Kowalczykowski, S. C.** 1994. In vitro reconstitution of homologous recombination reactions.
9 Experientia **50**:204.
- 10 80. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer.**
11 1994. Biochemistry of homologous recombination in Escherichia coli. Microbiol Rev **58**:401-65.
- 12 81. **Kowalczykowski, S. C., and R. A. Krupp.** 1987. Effects of Escherichia coli SSB protein on the
13 single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that
14 SSB protein facilitates the binding of RecA protein to regions of secondary structure within
15 single-stranded DNA. J Mol Biol **193**:97-113.
- 16 82. **Kramer, M. G., M. Espinosa, T. K. Misra, and S. A. Khan.** 1999. Characterization of a single-
17 strand origin, ssoU, required for broad host range replication of rolling-circle plasmids. Molecular
18 microbiology **33**:466.
- 19 83. **Kramer, M. G., M. Espinosa, T. K. Misra, and S. A. Khan.** 1998. Lagging strand replication of
20 rolling-circle plasmids: specific recognition of the ssoA-type origins in different gram-positive
21 bacteria. Proceedings of the National Academy of Sciences of the United States of America
22 **95**:10505.
- 23 84. **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1998. Lagging-strand replication from the ssoA
24 origin of plasmid pMV158 in Streptococcus pneumoniae: in vivo and in vitro influences of
25 mutations in two conserved ssoA regions. Journal of bacteriology **180**:83.
- 26 85. **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1997. Plasmid rolling circle replication:
27 identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I
28 for lagging strand synthesis. The EMBO journal **16**:5784.
- 29 86. **Kreuzer, K. N.** 2005. Interplay between DNA replication and recombination in prokaryotes.

- 1 Annual review of microbiology **59**:43.
- 2 87. **Kurahashi, H., H. Inagaki, T. Ohye, H. Kogo, T. Kato, and B. S. Emanuel.** 2006.
- 3 Chromosomal translocations mediated by palindromic DNA. *Cell Cycle* **5**:1297-303.
- 4 88. **Lambert, P. F., D. A. Waring, R. D. Wells, and W. S. Reznikoff.** 1986. DNA requirements at
- 5 the bacteriophage G4 origin of complementary-strand DNA synthesis. *J. Virol.* **58**:450.
- 6 89. **Lampson, B. C., M. Inouye, and S. Inouye.** 2005. Retrons, msDNA, and the bacterial
- 7 genome. *Cytogenetic and genome research* **110**:491.
- 8 90. **Langston, L. D., and M. O'Donnell.** 2006. DNA replication: keep moving and don't mind the
- 9 gap. *Mol Cell* **23**:155-60.
- 10 91. **Leach, D. R.** 1994. Long DNA palindromes, cruciform structures, genetic instability and
- 11 secondary structure repair. *Bioessays* **16**:893-900.
- 12 92. **Lilley, D. M.** 1985. The kinetic properties of cruciform extrusion are determined by DNA base-
- 13 sequence. *Nucleic Acids Res* **13**:1443-65.
- 14 93. **Liu, L. F., and J. C. Wang.** 1987. Supercoiling of the DNA template during transcription. *Proc*
- 15 *Natl Acad Sci U S A* **84**:7024-7.
- 16 94. **Liu, Y., V. Bondarenko, A. Ninfa, and V. M. Studitsky.** 2001. DNA supercoiling allows
- 17 enhancer action over a large distance. *Proc Natl Acad Sci U S A* **98**:14883-8.
- 18 95. **Llosa, M., S. Bolland, and F. de la Cruz.** 1991. Structural and functional analysis of the origin
- 19 of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related
- 20 IncN plasmid R46. *Mol Gen Genet* **226**:473-83.
- 21 96. **Low, K. B.** 1972. Escherichia coli K-12 F-prime factors, old and new. *Bacteriological reviews*
- 22 **36**:587.
- 23 97. **MacDonald, D., G. Demarre, M. Bouvier, D. Mazel, and D. N. Gopaul.** 2006. Structural
- 24 basis for broad DNA specificity in integron recombination. *Nature* **440**:1157-1162.
- 25 98. **Majchrzak, M., R. P. Bowater, P. Staczek, and P. Parniewski.** 2006. SOS repair and DNA
- 26 supercoiling influence the genetic stability of DNA triplet repeats in Escherichia coli. *J Mol Biol*
- 27 **364**:612-24.
- 28 99. **Masai, H., and K. Arai.** 1996. DnaA- and PriA-dependent primosomes: two distinct replication
- 29 complexes for replication of Escherichia coli chromosome. *Frontiers in bioscience: a journal and*

- 1 virtual library **1:d48**.
- 2 100. **Masai, H., and K. Arai.** 1997. Frpo: a novel single-stranded DNA promoter for transcription and
3 for primer RNA synthesis of DNA replication. *Cell* **89**:897.
- 4 101. **Masai, H., N. Nomura, and K. Arai.** 1990. The ABC-primosome. A novel priming system
5 employing dnaA, dnaB, dnaC, and primase on a hairpin containing a dnaA box sequence. *J Biol*
6 *Chem* **265**:15134-44.
- 7 102. **Mazel, D.** 2006. Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4**:608-20.
- 8 103. **McGlynn, P., A. A. Al-Deib, J. Liu, K. J. Marians, and R. G. Lloyd.** 1997. The DNA replication
9 protein PriA and the recombination protein RecG bind D-loops. *Journal of molecular biology*
10 **270**:212.
- 11 104. **Mejean, V., and J. P. Claverys.** 1984. Use of a cloned DNA fragment to analyze the fate of
12 donor DNA in transformation of *Streptococcus pneumoniae*. *J Bacteriol* **158**:1175-8.
- 13 105. **Mendiola, M. V., I. Bernales, and F. de la Cruz.** 1994. Differential roles of the transposon
14 termini in IS91 transposition. *Proc Natl Acad Sci U S A* **91**:1922-6.
- 15 106. **Miao, D. M., Y. Honda, K. Tanaka, A. Higashi, T. Nakamura, Y. Taguchi, H. Sakai, T.**
16 **Komano, and M. Bagdasarian.** 1993. A base-paired hairpin structure essential for the
17 functional priming signal for DNA replication of the broad host range plasmid RSF1010. *Nucleic*
18 *Acids Res* **21**:4900-3.
- 19 107. **Monzingo, A. F., A. Ozburn, S. Xia, R. J. Meyer, and J. D. Robertus.** 2007. The structure of
20 the minimal relaxase domain of MobA at 2.1 Å resolution. *Journal of molecular biology* **366**:165.
- 21 108. **Mortier-Barriere, I., M. Velten, P. Dupaigne, N. Mirouze, O. Pietrement, S. McGovern, G.**
22 **Fichant, B. Martin, P. Noirot, E. Le Cam, P. Polard, and J. P. Claverys.** 2007. A key
23 presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming
24 ssDNA to RecA. *Cell* **130**:824-36.
- 25 109. **Mott, M. L., and J. M. Berger.** 2007. DNA replication initiation: mechanisms and regulation in
26 bacteria. *Nature reviews. Microbiology* **5**:343.
- 27 110. **Murchie, A. I., and D. M. Lilley.** 1987. The mechanism of cruciform formation in supercoiled
28 DNA: initial opening of central basepairs in salt-dependent extrusion. *Nucleic Acids Res*
29 **15**:9641-54.

- 1 111. **Nelson, T. C.** 1951. Kinetics of genetic recombination in *Escherichia coli*. *Genetics* **36**:162.
- 2 112. **Noirot, P., J. Bargonetti, and R. P. Novick.** 1990. Initiation of rolling-circle replication in
3 pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proceedings of the*
4 *National Academy of Sciences of the United States of America* **87**:8560.
- 5 113. **Noirot, P., J. Bargonetti, and R. P. Novick.** 1990. Initiation of rolling-circle replication in
6 pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc Natl Acad Sci U*
7 *S A* **87**:8560-4.
- 8 114. **Nomura, N., H. Masai, M. Inuzuka, C. Miyazaki, E. Ohtsubo, T. Itoh, S. Sasamoto, M.**
9 **Matsui, R. Ishizaki, and K. Arai.** 1991. Identification of eleven single-strand initiation
10 sequences (ssi) for priming of DNA replication in the F, R6K, R100 and ColE2 plasmids. *Gene*
11 **108**:15.
- 12 115. **Oussatcheva, E. A., J. Pavlicek, O. F. Sankey, R. R. Sinden, Y. L. Lyubchenko, and V. N.**
13 **Potaman.** 2004. Influence of global DNA topology on cruciform formation in supercoiled DNA. *J*
14 *Mol Biol* **338**:735-43.
- 15 116. **Pages, V., and R. P. Fuchs.** 2003. Uncoupling of leading- and lagging-strand DNA replication
16 during lesion bypass in vivo. *Science* **300**:1300-3.
- 17 117. **Panayotatos, N., and R. D. Wells.** 1981. Cruciform structures in supercoiled DNA. *Nature*
18 **289**:466-70.
- 19 118. **Pearson, C. E., H. Zorbas, G. B. Price, and M. Zannis-Hadjopoulos.** 1996. Inverted repeats,
20 stem-loops, and cruciforms: significance for initiation of DNA replication. *J Cell Biochem* **63**:1-
21 22.
- 22 119. **Platt, J. R.** 1955. POSSIBLE SEPARATION OF INTERTWINED NUCLEIC ACID CHAINS BY
23 TRANSFER-TWIST. *Proceedings of the National Academy of Sciences of the United States of*
24 *America* **41**:181.
- 25 120. **Pruss, G. J., and K. Drlica.** 1989. DNA supercoiling and prokaryotic transcription. *Cell* **56**:521-
26 3.
- 27 121. **Recchia, G. D., and R. M. Hall.** 1995. Gene cassettes: a new class of mobile element.
28 *Microbiology* **141**:3015-3027.
- 29 122. **Reddy, M. S., M. B. Vaze, K. Madhusudan, and K. Muniyappa.** 2000. Binding of SSB and

- 1 RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA
2 polymerization on single-stranded DNA. *Biochemistry* **39**:14250-62.
- 3 123. **Ronning, D. R., C. Guynet, B. Ton-Hoang, Z. N. Perez, R. Ghirlando, M. Chandler, and F.**
4 **Dyda.** 2005. Active site sharing and subterminal hairpin recognition in a new class of DNA
5 transposases. *Mol Cell* **20**:143-54.
- 6 124. **Roy, R., A. G. Kozlov, T. M. Lohman, and T. Ha.** 2009. SSB protein diffusion on single-
7 stranded DNA stimulates RecA filament formation. *Nature* **461**:1092-7.
- 8 125. **Russel, M., N. A. Linderoth, and A. Sali.** 1997. Filamentous phage assembly: variation on a
9 protein export theme. *Gene* **192**:23.
- 10 126. **Russel, M., and P. Model.** 1989. Genetic analysis of the filamentous bacteriophage packaging
11 signal and of the proteins that interact with it. *Journal of virology* **63**:3284.
- 12 127. **Schleif, R.** 2000. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet*
13 **16**:559-65.
- 14 128. **Schwartzman, J. B., and A. Stasiak.** 2004. A topological view of the replicon. *EMBO Rep*
15 **5**:256-61.
- 16 129. **Shlyakhtenko, L. S., P. Hsieh, M. Grigoriev, V. N. Potaman, R. R. Sinden, and Y. L.**
17 **Lyubchenko.** 2000. A cruciform structural transition provides a molecular switch for
18 chromosome structure and dynamics. *Journal of molecular biology* **296**:1169.
- 19 130. **Sinden, R. R., S. S. Broyles, and D. E. Pettijohn.** 1983. Perfect palindromic lac operator DNA
20 sequence exists as a stable cruciform structure in supercoiled DNA in vitro but not in vivo.
21 *Proceedings of the National Academy of Sciences of the United States of America* **80**:1797.
- 22 131. **Sinden, R. R., G. X. Zheng, R. G. Brankamp, and K. N. Allen.** 1991. On the deletion of
23 inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform
24 formation in vivo. *Genetics* **129**:991-1005.
- 25 132. **Singh, J., M. Mukerji, and S. Mahadevan.** 1995. Transcriptional activation of the *Escherichia*
26 *coli* bgl operon: negative regulation by DNA structural elements near the promoter. *Molecular*
27 *microbiology* **17**:1085.
- 28 133. **Sullivan, K. M., and D. M. Lilley.** 1987. Influence of cation size and charge on the extrusion of
29 a salt-dependent cruciform. *J Mol Biol* **193**:397-404.

- 1 134. **Sun, W., and G. N. Godson.** 1998. Structure of the Escherichia coli primase/single-strand DNA-
2 binding protein/phage G4oric complex required for primer RNA synthesis. Journal of molecular
3 biology **276**:689.
- 4 135. **Swart, J. R., and M. A. Griep.** 1993. Primase from Escherichia coli primes single-stranded
5 templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins.
6 Template sequence requirements based on the bacteriophage G4 complementary strand origin
7 and Okazaki fragment. The Journal of biological chemistry **268**:12970.
- 8 136. **Tanaka, T., T. Mizukoshi, K. Sasaki, D. Kohda, and H. Masai.** 2007. Escherichia coli PriA
9 protein, two modes of DNA binding and activation of ATP hydrolysis. The Journal of biological
10 chemistry **282**:19917.
- 11 137. **Tatum, E. L., and J. Lederberg.** 1947. Gene Recombination in the Bacterium Escherichia coli.
12 Journal of bacteriology **53**:673.
- 13 138. **Trinh, T. Q., and R. R. Sinden.** 1991. Preferential DNA secondary structure mutagenesis in the
14 lagging strand of replication in E. coli. Nature **352**:544-7.
- 15 139. **Val, M.-E., M. Bouvier, J. Campos, D. Sherratt, F. o. Cornet, D. Mazel, and F. o.-X. Barre.**
16 2005. The single-stranded genome of phage CTX is the form used for integration into the
17 genome of Vibrio cholerae. Molecular cell **19**:559.
- 18 140. **Val, M. E., M. Bouvier, J. Campos, D. Sherratt, F. Cornet, D. Mazel, and F. X. Barre.** 2005.
19 The single-stranded genome of phage CTX is the form used for integration into the genome of
20 Vibrio cholerae. Mol Cell **19**:559-66.
- 21 141. **Walker, G. C.** 1996. The SOS Response of Escherichia coli. Escherichia coli and Salmonella.
22 Neidhardt, FC Washington DC American Society of Microbiology **1**:1400-1416.
- 23 142. **Wang, J. C., and A. S. Lynch.** 1993. Transcription and DNA supercoiling. Current opinion in
24 genetics & development **3**:764.
- 25 143. **White, J. H., and W. R. Bauer.** 1987. Superhelical DNA with local substructures. A
26 generalization of the topological constraint in terms of the intersection number and the ladder-
27 like correspondence surface. Journal of molecular biology **195**:205.
- 28 144. **Wickner, S., and J. Hurwitz.** 1975. Association of phi X174 DNA-Dependent ATPase Activity
29 with an Escherichia coli Protein, Replication Factor Y, Required for in vitro Synthesis of phi X174

- 1 DNA. Proceedings of the National Academy of Sciences **72**:3342.
- 2 145. **Wozniak, R. A., D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, G. Garriss, C.**
3 **Dery, V. Burrus, and M. K. Waldor.** 2009. Comparative ICE genomics: insights into the
4 evolution of the SXT/R391 family of ICEs. PLoS Genet **5**:e1000786.
- 5 146. **Zannis-Hadjopoulos, M., W. Yahyaoui, and M. Callejo.** 2008. 14-3-3 cruciform-binding
6 proteins as regulators of eukaryotic DNA replication. Trends in biochemical sciences **33**:44.
- 7 147. **Zhao, J., A. Bacolla, G. Wang, and K. M. Vasquez.** 2010. Non-B DNA structure-induced
8 genetic instability and evolution. Cell Mol Life Sci **67**:43-62.
- 9 148. **Zheng, G., D. W. Ussery, and R. R. Sinden.** 1991. Estimation of superhelical density in vivo
10 from analysis of the level of cruciforms existing in living cells. J Mol Biol **221**:122-9.
- 11 149. **Zheng, G. X., T. Kochel, R. W. Hoepfner, S. E. Timmons, and R. R. Sinden.** 1991.
12 Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at
13 specific sites in DNA of living cells. J Mol Biol **221**:107-22.
- 14 150. **Zheng, G. X., and R. R. Sinden.** 1988. Effect of base composition at the center of inverted
15 repeated DNA sequences on cruciform transitions in DNA. J Biol Chem **263**:5356-61.
- 16
17
18

FIGURES LEGENDS

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Figure 1. DNA uptake and production of ssDNA in the cell. The three mechanisms: conjugation, natural transformation and infection by ssDNA phage lead the production of large amount of ssDNA in the cell.

Figure 2. Hairpin formation during replication. Hairpins can fold on the ssDNA formed by the discontinuous replication of the lagging strand, or on ssDNA gaps remaining after lesion bypass.

Figure 3. Mechanisms of cruciform extrusion. In the C-type pathway, a substantial region of dbDNA is denatured allowing the folding of the whole hairpins on both strands in one step. In the S-type pathway, a small region is denatured (~10bp) allowing the folding of a small hairpin that can then be elongated through branch migration.

Figure 4. Priming of replication on ssDNA hairpins. G4-type priming: hairpins structure the region, directing the binding of SSB and allowing access to the dnaG primase. Phi174-type priming: a ssDNA hairpin forms a Y-fork recognized by PriA which directs the formation of a primosome. Filamentous phage type priming: an hairpin mimicking a promoter is recognized by the RNA polymerase (RNAP), which synthesizes an RNA primer for replication.

Figure 5. Rolling Circle Replication. A)The Rep protein binds an hairpin formed by double-stranded origin (dso) and extruded from dbDNA as a cruciform. Rep nicks DNA and covalently binds the 5'-end, leaving a 3'-end for replication to proceed. The leading strand is replicated while the lagging strand is

1 extruded and remains single-stranded until the single-stranded origin (sso) is reached. The RNA polymerase
2 (RNAP) binds the sso hairpin and synthesizes an RNA primer for replication. B) The pT181 dso in cruciform
3 conformation. C) The pT181 sso as folded by the mFOLD software.

4

5 **Figure 6. N4 virion hairpin promoters.** The three promoters of N4 controlling the expression of
6 the early genes as cruciform structures.

7

8 **Figure 7. The *V. cholerae* chromosome I dif site and the CTX phage hairpin.** The CTX
9 *attP* region folds into a forked hairpin mimicking the *V. cholerae dif1*. This enables the CTX phage to use the
10 host XerC/D recombinase to catalyze its integration into the chromosome.

11

12 **Figure 8. Organization of IS608 and Overall Transposition Pathway (adapted from**
13 **Guynet, 2009 (52)).** (A) Organization. *tnpA* and *tnpB* open reading frames (light and dark arrows,
14 respectively). Left end (LE) and right end (RE) (red and blue boxes, respectively). (B) Sequence of LE and
15 RE. Sequence and secondary structures, IP_L and IP_R, at the LE and RE IS608 are shown. Left and right
16 tetranucleotide cleavage sites are boxed in black (C_L) and underlined in blue (C_R). C_R forms part of IS608,
17 whereas C_L does not. B_L and B_R are shown on a red and blue background, respectively. Position of cleavage
18 and of formation of the 5' phosphotyrosine TnpA-DNA intermediate (vertical arrows). (C) Transposition
19 pathway. (i) Schematized IS608 with IP_L and IP_R, left (TTAC; C_L) and right (TCAA; C_R) cleavage sites. (ii)
20 Formation of a single-strand transposon circle intermediate with abutted left and right ends. The transposon
21 junction (TCAA) and donor joint (TTAC) are shown. (iii) Pairing with the target (TTAC) and cleavage
22 (vertical arrows). (iv) Inserted transposon with new left and right flanks (dotted black lines).

23

24 **Figure 9. Recombination between an attC site hairpin of an integron cassette and a**

1 **double-stranded attI site.** The first recombination steps (a-c) between the folded attC site and the dsDNA
2 attI site are identical to classical recombination steps catalyzed by other tyrosine recombinases. (b) Four
3 integrase monomers bind to the core sites (the proper strand of the attC site being recognized through a
4 specific binding with the extrahelical G). Binding to structural determinants makes the pink monomers
5 inactive, leaving to the green monomers the possibility to realize the first strand exchange (c). The pseudo-
6 holliday junction formed cannot be resolved by a second strand exchange as occurs with classical tyrosine
7 recombinases. The current model is that replication is involved to solve the junction in a process that remains
8 to be understood (d).

9

10 **Figure 10. ssDNA: at the crossroads of horizontal gene transfer, the SOS response and**
11 **genetic rearrangements.** 1) Conjugation, transformation, phage infection and environmental stress lead
12 to the production of ssDNA in the cell. 2) The RecA proteins binds ssDNA and triggers the self-cleavage of
13 LexA (brown circles). 3) The SOS regulon is derepressed, recombinases are expressed (orange triangles), and
14 DNA coiling is modified. 4) Increased supercoiling leads to cruciform formation. 5) Induction of IS
15 transposition and integron recombination.6) ICE conjugation, lysogenic phages and natural competence are
16 induced.

SUMMARY

1
2 Structured forms of DNA with intrastrand pairing are generated in several cellular processes and are
3 involved in biological functions. These structures may arise on single-stranded DNA (ssDNA) produced
4 during replication, bacterial conjugation, natural transformation or during viral infections. Furthermore,
5 negatively supercoiled DNA can extrude inverted repeats as hairpins in structures called cruciforms. Whether
6 they are on ssDNA or as cruciforms, hairpins can modify the access of proteins to DNA and, in some cases,
7 they can be directly recognized by proteins. Folded DNA has been found to play an important role in
8 replication, transcription regulation and recognition of the origins of transfer in conjugative elements. More
9 recently, they were shown to be used as recombination sites. Many of these functions are found on mobile
10 genetic elements likely to be single-stranded, including viruses, plasmids, transposons and integrons, thus
11 giving some clues as to the manner in which they might have evolved. We review here, with special focus on
12 prokaryotes, the functions in which DNA secondary structures play a role and the cellular processes giving
13 rise to them. Finally, we attempt to shed light on the selective pressures leading to the acquisition of
14 functions for DNA secondary structures.

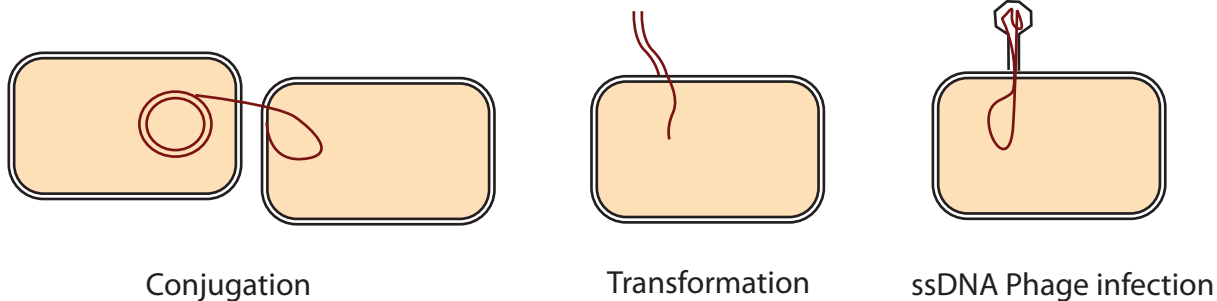


Figure 1. DNA uptake and production of ssDNA in the cell. The three mechanisms: conjugation, natural transformation and infection by ssDNA phage lead the production of large amount of ssDNA in the cell.

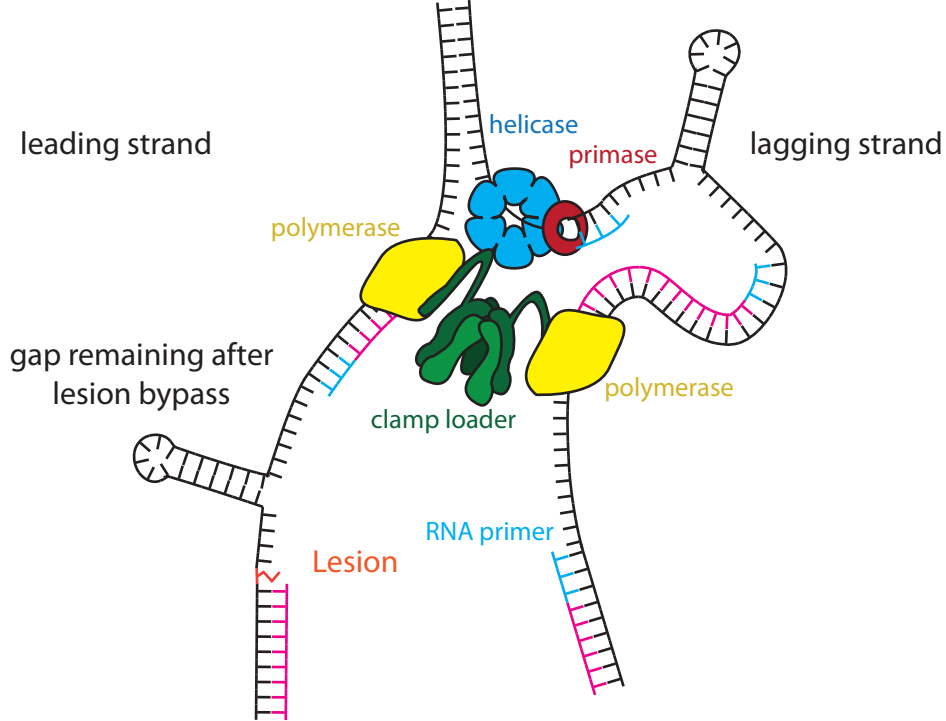


Figure 2. Hairpin formation during replication. Hairpins can fold on the ssDNA formed by the discontinuous replication of the lagging strand, or on ssDNA gaps remaining after lesion bypass.

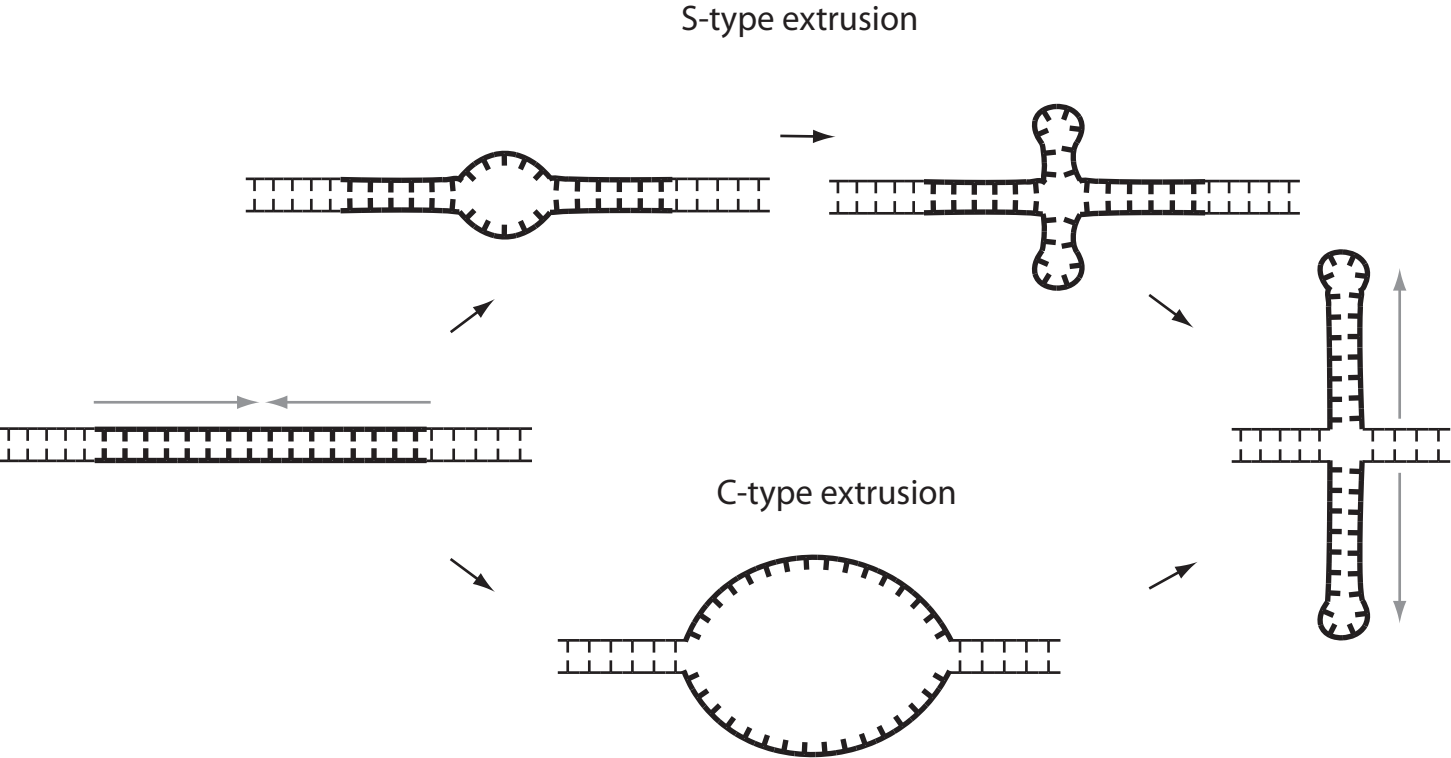


Figure 3. Mechanisms of cruciform extrusion. In the C-type pathway, a substantial region of dbDNA is denatured allowing the folding of the whole hairpins on both strands in one step. In the S-type pathway, a small region is denatured (~10bp) allowing the folding of a small hairpin that can then be elongated through branch migration.

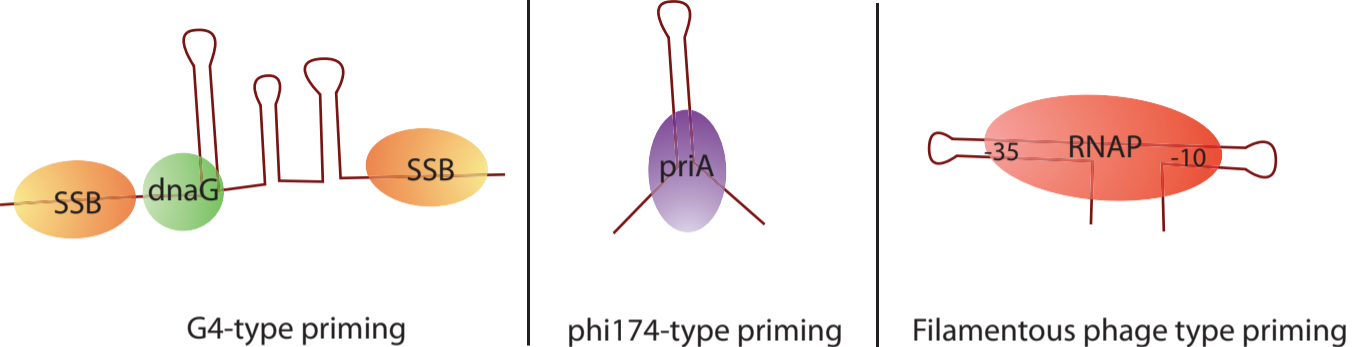
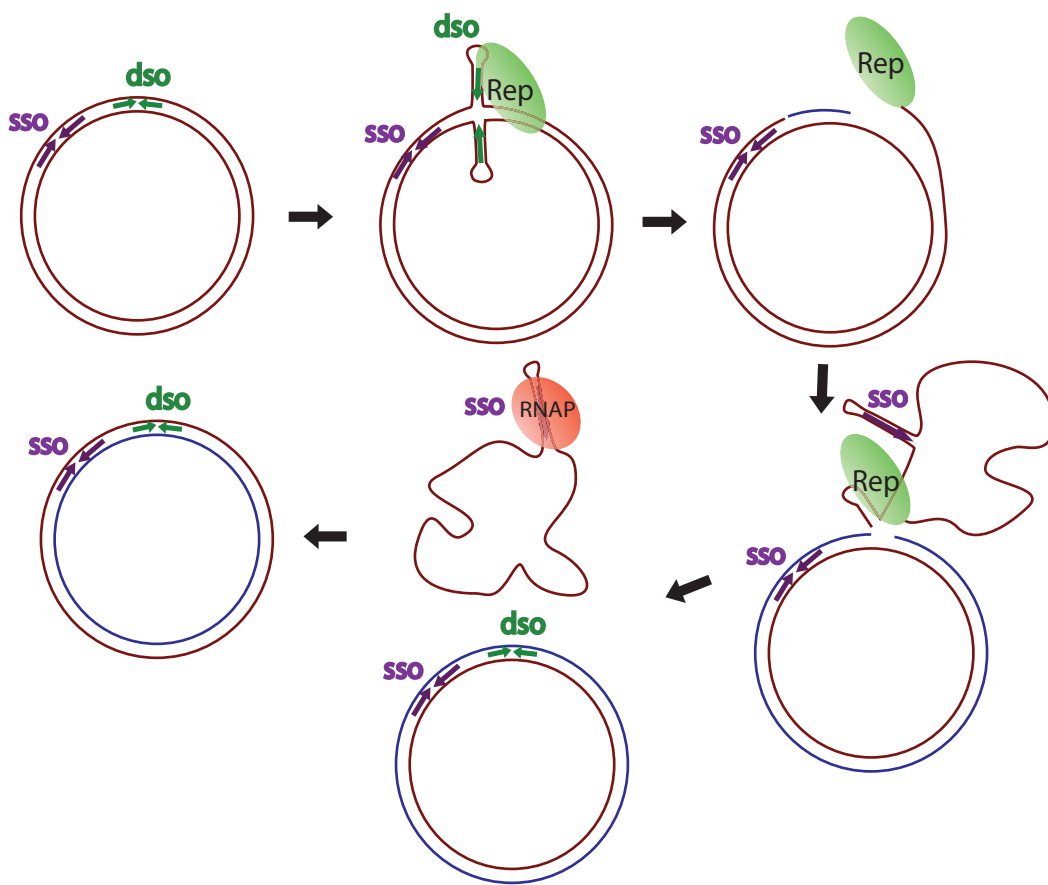
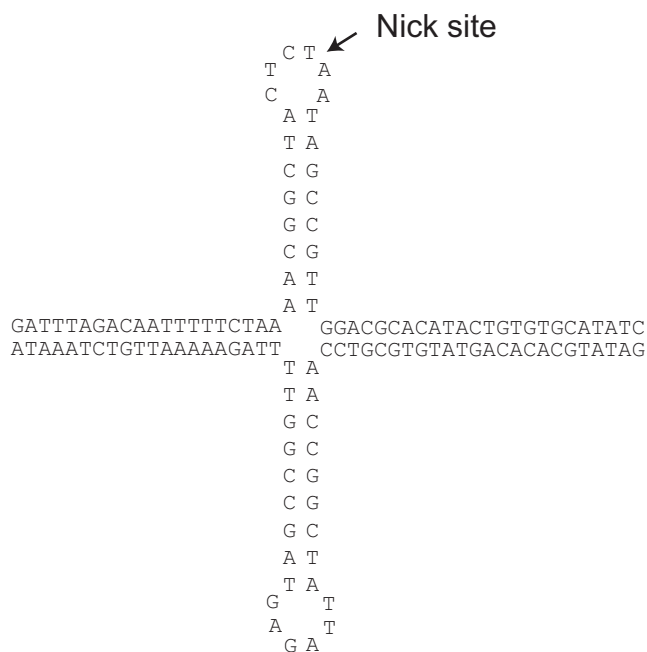


Figure 4. Priming of replication on ssDNA hairpins. G4-type priming: hairpins structure the region, directing the binding of SSB and allowing access to the dnaG primase. Phi174-type priming: a ssDNA hairpin forms a Y-fork recognized by PriA which directs the formation of a primosome. Filamentous phage type priming: an hairpin mimicking a promoter folds and is recognized by the RNA polymerase (RNAP), which synthesizes an RNA primer for replication.

A



B



C

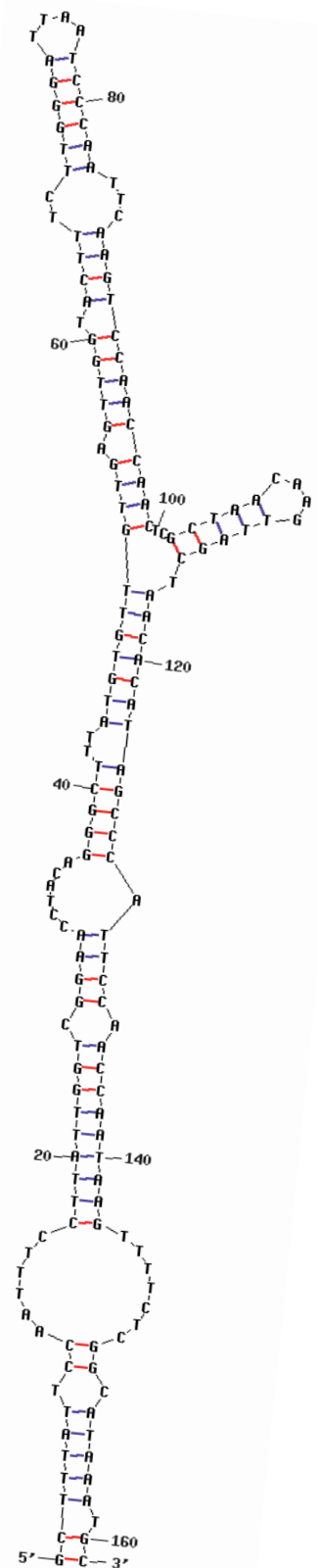


Figure 5. Rolling Circle Replication. A) The Rep protein binds an hairpin formed by double-stranded origin (dso) and extruded from dbDNA as a cruciform. Rep nicks DNA and covalently binds the 5'-end, leaving a 3'-end for replication to proceed. The leading strand is replicated while the lagging strand is extruded and remains single-stranded until the single-stranded origin (sso) is reached. The RNA polymerase (RNAP) binds the sso hairpin and synthesizes an RNA primer for replication. B) The pT181 dso in cruciform conformation. C) The pT181 sso as folded by the mFOLD software.

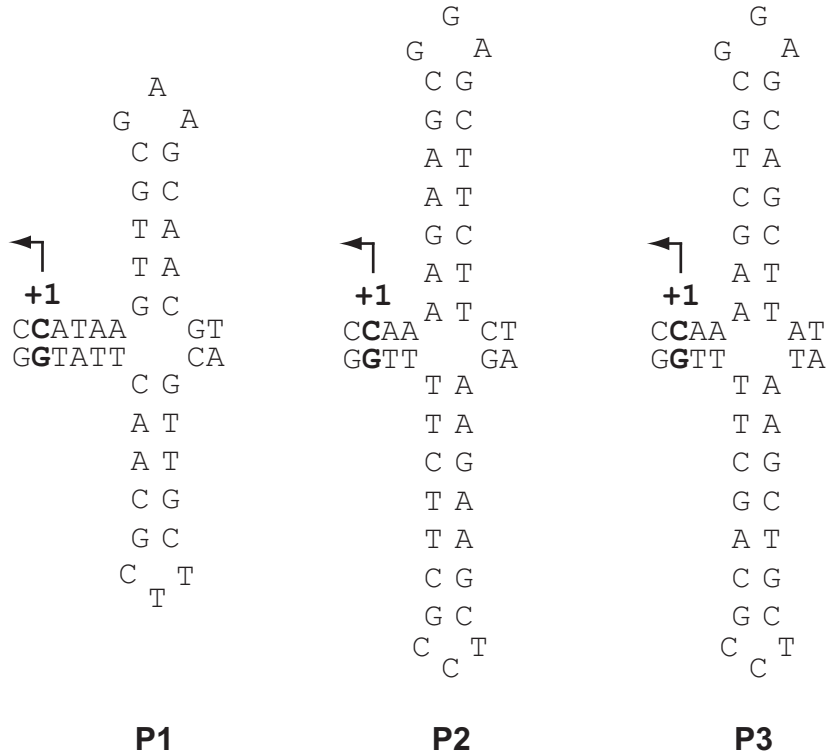
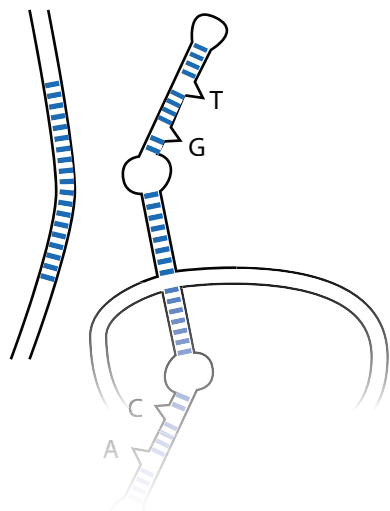
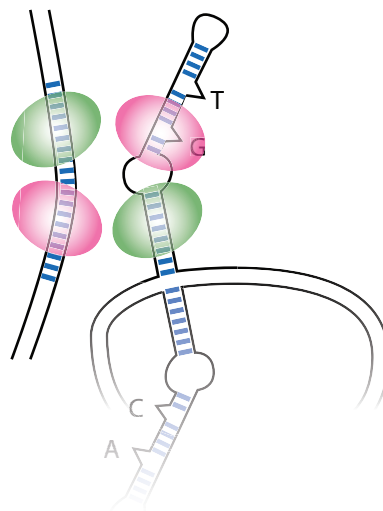


Figure 6. N4 virion hairpin promoters. The three promoters of N4 controlling the expression of the early genes as cruciform structures.

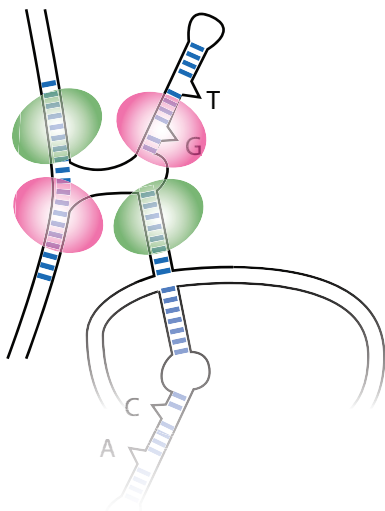
(a)



(b)



(c)



(d)

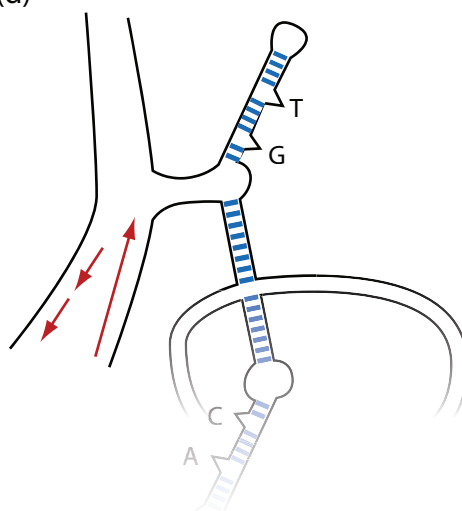


Figure 9. Recombination between an attC site hairpin of an integron cassette and a double-stranded attI site.

The first recombination steps (a-c) between the folded attC site and the dsDNA attI site are identical to classical recombination steps catalyzed by other tyrosine recombinases. (b) Four integrase monomers bind to the core sites (the proper strand of the attC site being recognized through a specific binding with the extrahelical G). Binding to structural determinants makes the pink monomers inactive, leaving to the green monomers the possibility to realize the first strand exchange (c). The pseudo-holliday junction formed cannot be resolved by a second strand exchange as occurs with classical tyrosine recombinases. The current model is that replication is involved to solve the junction in a process that remains to be understood (d).

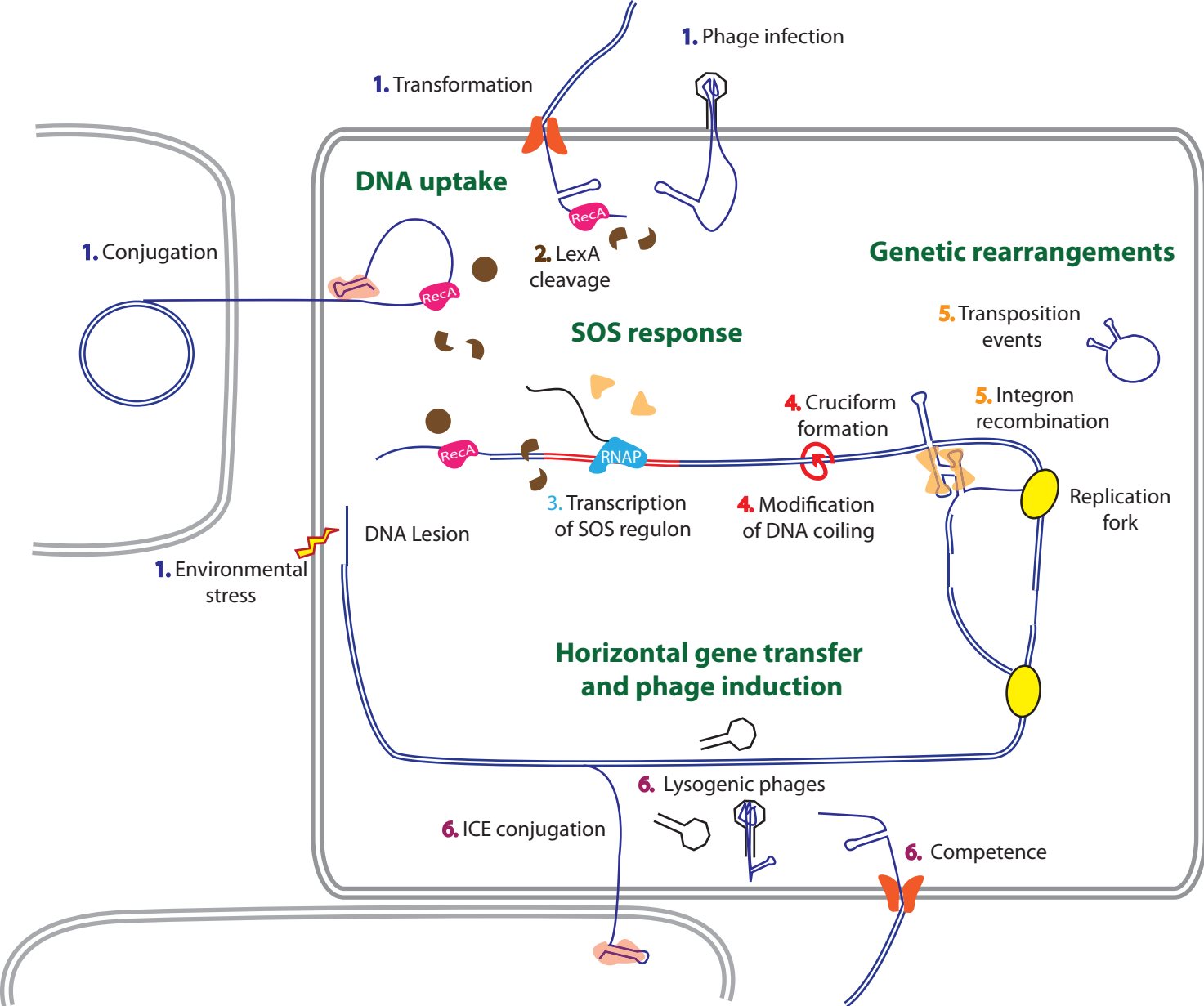


Figure 10. ssDNA: at the crossroads of horizontal gene transfer, the SOS response and genetic rearrangements. 1) Conjugation, transformation, phage infection and environmental stress lead to the production of ssDNA in the cell. 2) The RecA proteins binds ssDNA and triggers the self-cleavage of LexA (brown circles). 3) The SOS regulon is derepressed, recombinases are expressed (orange triangles), and DNA coiling is modified. 4) Increased supercoiling leads to cruciform formation. 5) Induction of IS transposition and integron recombination. 6) ICE conjugation, lysogenic phages and natural competence are induced.