MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Dec. 2010, p. 570–588 Vol. 74, No. 4
1092-2172/10/$12.00 doi:10.1128/MMBR.00026-10

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]*

Institut Pasteur, Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, F-75015 Paris, France,[1] and CNRS, URA2171, F-75015 Paris, France[2]

* Corresponding author. Mailing address: Institut Pasteur, Unité Plasticité du Génome Bactérien, Département Génomes et Génétique, F-75015 Paris, France. Phone: 33 1 40 61 32 84. Fax: 33 1 45 68 88 34. E-mail: mazel@pasteur.fr.

INTRODUCTION

The B-helix form of DNA proposed by Watson and Crick accounts for most of the behavior of DNA in the cell. Nevertheless, it is now obvious that DNA is not 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 (158). This review focuses on DNA hairpins, i.e., DNA with intrastrand base pairing, and 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. First, in several cellular processes, DNA is single stranded (single-stranded DNA [ssDNA]), for instance, during replication on the template for lagging-strand synthesis, during DNA repair, or, more importantly, during rolling-circle replication (RCR), bacterial conjugation, natural transformation, and infection by some viruses. 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 (129): the negative supercoiling of dsDNA could provide free energy to stabilize cruciforms. Cruciforms then attracted much attention in the 1980s, when their existence was experimentally assessed *in vitro* under natural superhelical densities (127). However, 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* (29, 140). 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 (154), which is known to affect the binding of regulatory proteins for transcription, recombination, and replication (30, 59); (ii) the DNA-protein interaction can be inhibited if a hairpin overlaps a protein recognition site (70); and (iii) proteins can directly recognize and bind DNA hairpins (10, 53, 107, 110, 150).

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

DNA HAIRPIN FORMATION

Hairpin Formation from ssDNA

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

Formation of ssDNA through horizontal gene transfer. (i) Conjugation. Conjugation is the process by which one bacterium can actively transfer DNA to a neighboring cell. The mechanism of conjugation is conserved across all described systems. A protein called relaxase binds and nicks a cognate origin-of-transfer site (*oriT*). This reaction results in a covalent complex between the relaxed plasmid and the relaxase (together with accessory factors), called the relaxosome. Only the strand that is covalently bound by the relaxase is transferred to the recipient cell as ssDNA. The transferred strand (T strand) is excreted from the donor cell through the type IV secretion system, and the relaxase then directs the recircularization of the T strand in the recipient cell (for a comprehensive review, see reference 39). Two main families of conjugative elements have been described: self-transmissible plasmids and "integrative and conjugative elements" (ICEs). ICEs cannot autonomously replicate and are thus carried by chromosomes or other replicons. These elements are able to excise themselves as circular intermediates through the action of a recombinase/excisionase and are then transferred following the same conjugation mechanism. In the recipient cell, they can be integrated through homologous recombination or through the action of a site-specific recombinase (18, 77). The length of the DNA molecule that is transferred is usually the size of the whole conjugative element (usually <200 kb).

Occasionally, chromosomal DNA can be transferred. This happens when conjugative plasmids are integrated into the chromosome, with a well-known example being the plasmid F/Hfr system (105, 147). Alternatively, the conjugation functions carried by ICEs can also promote the transfer of chromosomal or plasmid DNA, as demonstrated for *Streptococcus agalactiae* (16) and for the SXT element in *Vibrio cholerae* (66). In this case, the length of the transferred strand is limited by the conjugation bridge strength and the contact time between the bacteria. Since the time of early genetic mapping of the *Escherichia coli* chromosome through Hfr conjugation by Nelson, we have learned that it takes about 100 min to transfer the whole *E. coli* chromosome (4.6 Mb) (122). Although very long DNA fragments can be transferred, the average length of the ssDNA region in the recipient cell is unknown. Indeed, the

ssDNA length and its lifetime depend on the speed of complementary-strand synthesis. The only direct data available come from microscopy experiments enabling the visualization of complementary-strand synthesis and showing that synthesis starts within 5 min after the donor and recipient cells are mixed (6). Nevertheless, the number of ssDNA replication origins is unknown in most cases. Single-stranded origins of replication have been studied in the case of rolling-circle replication, which is discussed below (see "Hairpins and Replication Origins"). The fact that specific origins of replication have evolved for the initiation of complementary-strand synthesis suggests that this process does not happen easily at random loci. This may seem at odds with the fact that the DnaG primase catalyzes the formation of primer RNAs every ~1 kb during the synthesis of the lagging strand. However, it was observed that DnaG needs to interact with the rest of the replisome (in particular, the DnaB helicase) to efficiently initiate synthesis (4, 106). Furthermore, the access of DnaG to ssDNA may be inhibited by the binding of other proteins such as the single-strand binding protein (SSB), making it hard for DnaG to prime DNA synthesis on random ssDNA sites (144). During conjugation, it is therefore unlikely that complementary-strand synthesis is initiated at numerous loci. Conjugation thus massively produces ssDNA, and conjugative plasmids are probably a place of choice for the evolution of functions where hairpins are involved. Indeed, the very process of conjugation, for instance, implies DNA secondary structures (53) (see "Hairpins and Conjugation" below).

(ii) Transformation. Bacterial competence for natural transformation is a physiological state that permits the uptake and incorporation of naked exogenous DNA. Many Gram-negative bacteria (including species of *Haemophilus*, *Neisseria*, *Helicobacter*, *Vibrio*, and *Acinetobacter*) as well as Gram-positive bacteria (including species of *Bacillus*, *Mycobacterium*, and *Streptomyces*) are capable of natural competence. In all cases, one strand of the transformed DNA is degraded, providing the energy for the transport of the complementary strand across the cytoplasmic membrane (24). Some bacteria have been shown to fragment exogenous DNA so that they take only small segments, while others can take up long DNA molecules (42). The monitoring of ssDNA fate during transformation in *Streptococcus pneumoniae* revealed that ssDNA does not subsist in the cell for more than 15 min (114). Globally, the length of the incoming DNA and the lifetime of ssDNA in the recipient cell are probably shorter than for conjugation. The entering single strand is protected from the action of nucleases essentially by the binding of SSB (26), whereas during conjugation, the relaxase is covalently bound to the T strand, effectively protecting it from exonucleases. However, for some bacteria, including *Bacillus subtilis* and *S. pneumoniae*, a protein named DprA has been found to bind the incoming ssDNA, protecting it from both endo- and exonucleases and facilitating further homologous recombination (118). All in all, during transformation, ssDNA is not long-lived in the cell; either it is quickly integrated into the chromosome through homologous recombination or it is degraded.

(iii) Phage infection. Single-stranded phages encapsidate their genome and deliver it to newly infected cells in this form. Their size is generally <10 kb, although some phages (notably filamentous phages) can accommodate longer segments of DNA simply by increasing their capsid size (61). Here again, little is known about the timing of complementary-strand synthesis and the length or availability of ssDNA in the infected cell. Nevertheless, hairpins have been found to play important roles at all steps of ssDNA phage life cycles, from the synthesis of the complementary strand (95, 155) to phage DNA encapsidation (135) (see DNA Hairpin Biological Functions below).

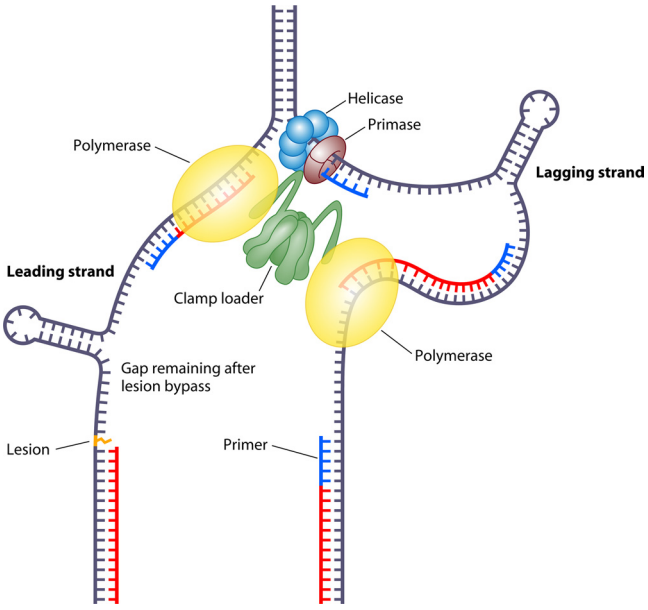

FIG. 1. Hairpin formation during replication. Hairpins can fold on the ssDNA formed by the discontinuous replication of the lagging strand or on ssDNA gaps remaining after lesion bypass.

Macromolecule synthesis and repair. (i) Transcription. RNA synthesis requires the opening of the DNA duplex. The size of the transcription bubble ranges between 12 and 25 bp (49). This small opening leaves very little room for secondary-structure formation, and transcription is thus unlikely to foster hairpin formation. On the contrary, the transcription bubble needs to unfold hairpins that it may encounter so as to enable the production of the correct transcripts by RNA polymerase (RNAP).

(ii) Replication. In contrast to transcription, DNA synthesis produces large amounts of ssDNA. First, the replication initiation step often requires the melting of a large DNA region around the origin of replication. Multiple hairpins have been found to play important roles at replication origins (20, 109) (see "Hairpins and Replication Origins" below). Second, lagging-strand replication is not continuous, and an ssDNA loop is formed to place the DNA in the correct orientation for DNA polymerase. The replication loop consists of ssDNA extruded by the helicase and of the nascent Okazaki fragment (Fig. 1). In *E. coli*, Okazaki fragments are 1 kb to 2 kb long, and the replication fork speed is about 1 kb $\cdot$ s^{-1} under optimal conditions (85). The lifetime of ssDNA should thus be on the order of a second. Evidence that inverted repeats (IRs) can fold into stable hairpins *in vivo* during replication came from the observation that large and perfect IRs are genetically unstable on plasmids in *E. coli*. Indeed, they are the cause of

mismatched alignment or slippage during replication (99, 141). In particular, deletions of IRs occur preferentially on the lagging strand (149).

Finally, a special mode of replication, called rolling-circle replication (RCR), involves the unwinding of the full lagging strand into ssDNA (82). Multiple hairpins have been found to play important roles in RCR (89, 90, 92, 123) (see Fig. 5).

(iii) DNA repair. A major source of ssDNA in the cell is through DNA repair. Double-strand breaks are processed by the RecBCD enzyme, which produces ssDNA tails through its exonuclease activity. These ssDNA tails can then be bound by RecA and may be involved in homologous-strand invasion and replication-dependent repair (86, 87, 93). Double-strand breaks can be caused by many agents, including ionizing radiation, UV light, and oxygen radicals, but in normally growing cells as well, double-strand breaks are frequently formed as a consequence of replication through imperfect DNA templates (for a comprehensive review, see reference 41).

The repair of mismatches can also produce ssDNA following a process known as methyl-directed mismatch repair (MMR) (72). The MutS protein recognizes mismatches, but it is not able to itself discriminate the correct template strand from the erroneous newly synthesized strand. This is achieved thanks to the methylation state of DNA. Immediately after replication, the DNA is hemimethylated, with the synthesized strand being transiently nonmethylated. The MutS partner, MutH, is able to find the mismatch's closest hemimethylated site (GATC in *E. coli*) and specifically cleave the newly synthesized strand. The UvrD helicase can then extrude DNA from the cleavage site to the mismatch position so that this segment can be resynthesized by polymerase III (Pol III). This leads to the production of ssDNA on the template strand, the amount of which depends on the distance between the methylation site and the mismatch and can be as much as 1 kb (17).

Finally, when replication forks encounter a lesion, the replication of the lagging and leading strands can be uncoupled in order to bypass the lesion, leaving ssDNA gaps on the damaged strand (60, 97, 126). These gaps are around 1 kb in length and can be processed by RecA-mediated recombinational repair (Fig. 1).

Single-strand DNA binding proteins. In all these processes, ssDNA in the cell is not left naked. Several proteins bind ssDNA without sequence specificity. The most important ones are the RecA and SSB proteins. SSB coats any ssDNA present in the cell and prevents intrastrand pairing, i.e., hairpin formation. The RecA protein also binds ssDNA, forming a straight nucleoproteic filament. RecA can then promote strand invasion of homologous dsDNA and catalyze recombination (86). Furthermore, SSB directs RecA binding to ssDNA (88, 132). Recent single-molecule studies have shown how tetrameric SSB can spontaneously migrate along ssDNA, melting unstable hairpins while stimulating RecA filament elongation (134).

Although ssDNA is present on many occasions in the cell, hairpin formation is strongly constrained by SSB and RecA binding. Proteins that ensure their function through hairpin binding are thus in competition with SSB and RecA for substrate availability. Hairpins that are formed need to be stable enough to resist SSB melting and coating. For instance, it was demonstrated that the RepC proteins, which initiate rolling-circle replication on plasmid pT181, can "erroneously" recognize alternative hairpin sites in the absence of SSB, but only the hairpin at the primary origin is stable enough to be recognized by RepC when SSB is present (83) [see "Double-strand DNA replication. (iii) Rolling-circle replication" below].

Cruciform Extrusion

Mechanism of cruciform extrusion. The formation of DNA hairpins in the cell does not necessarily require the production of ssDNA. The extrusion of cruciforms occurs through the opening of the DNA double helix to allow intrastrand base pairing. Strand opening in relaxed DNA is both infrequent and transient. However, negatively supercoiled DNA molecules are much more active in the formation of cruciforms, because their topology facilitates both large- and small-scale openings of the double helix (47). Two main mechanisms for cruciform extrusion have been proposed (Fig. 2) (100). The first (type S) implies small-scale melting of the double helix at the dyad of the IR (~10 bp). This small opening allows a few bases to pair with their cognate base in the repeat. The stem can then be elongated through branch migration, which is also facilitated by negative supercoiling. The other mechanism (type C) involves the melting of a large region, which is favored by nearby AT-rich sequences. This large melting would allow hairpins to fold on both strands, leading to cruciform formation (Fig. 2). The S-type mechanism is highly dependent on the IR sequence (it is favored by the AT-rich sequence at the dyad) and works under physiological ionic conditions (143). On the other hand, C-type extrusion takes place in low-salt solutions and is highly dependent on the presence of AT-rich neighbor sequences but should theoretically be suppressed at physiological ion concentrations (120). Nevertheless, this mechanism could possibly take place in DNA regions with propensities to undergo substantial denaturation, such as replication origins.

Regulation of cruciform extrusion. Cruciforms were extensively studied in the 1980s, when techniques enabling their observation *in vitro* were developed, such as S1 sensitivity and two-dimensional (2D) electrophoresis. Although cruciform extrusion can be energetically favorable under moderate superhelical densities, the slow kinetics of cruciform extrusion raise questions as to their relevance *in vivo* (29). However, several techniques later developed led to the demonstration of cruciform formation *in vivo* under natural superhelical densities (37, 38, 70, 123). In particular, cruciforms that were tuned to fold stably at different superhelical densities have even been used to measure the natural superhelix densities of plasmids. *In vivo* cross-linking with psoralen demonstrated that the propensity of an IR to fold into a cruciform strongly depends on its sequence and context and that some IRs can exist as cruciforms at levels as high as 50% in plasmids in living *E. coli* cells (159, 160).

Nevertheless, most of the reported cruciform detection involved artificial conditions favoring hairpin extrusion: small loops, IR in AT-rich regions, perfect palindromes with AT-rich centers and GC-rich stems, topoisomerase mutants, or salt shock to increase supercoiling (141, 160, 161). Generally, IRs do not seem to fold cruciforms at significant rates under average *in vivo* supercoiling conditions. However, many factors may transiently increase local superhelical density to a critical level sufficient for cruciform extrusion (for a review, see reference

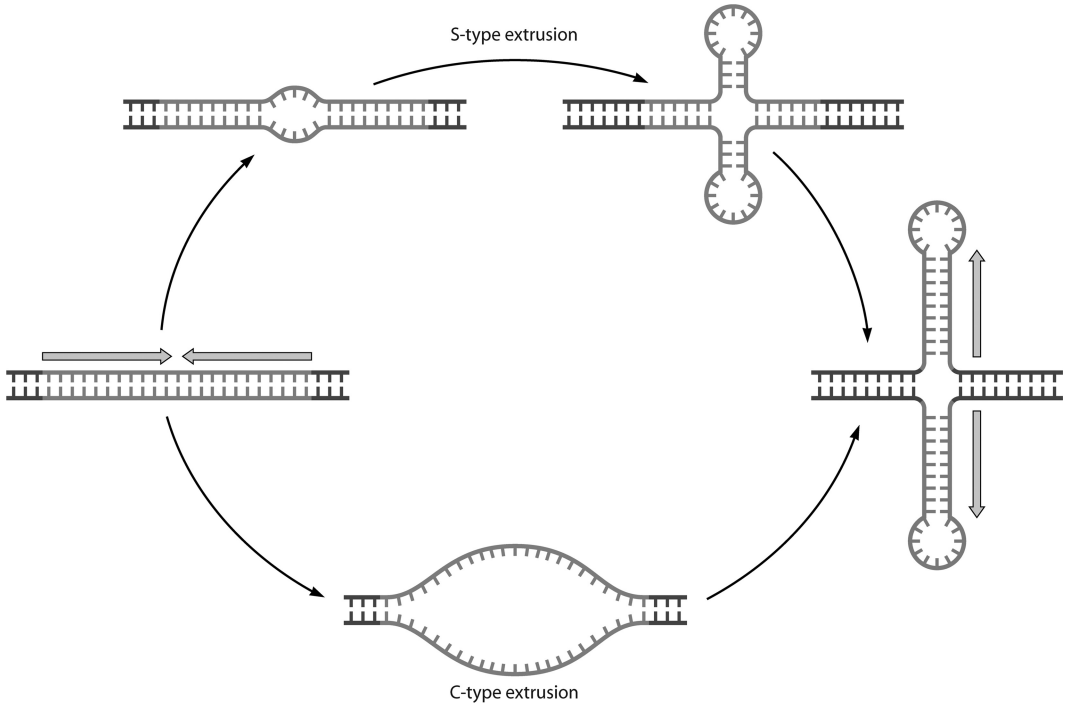

FIG. 2. Mechanisms of cruciform extrusion. In the C-type pathway, a substantial region of dsDNA is denatured, allowing the folding of the whole hairpins on both strands in one step. In the S-type pathway, a small region is denatured (~10 bp), allowing the folding of a small hairpin that can then be elongated through branch migration.

128). Biological processes such as transcription and replication may generate local and temporal domains of supercoiling on circular DNA (38, 101, 138). Indeed, during replication and transcription, enzymes alter the structure of DNA such that additional twists are added (positive supercoiling) or subtracted (negative supercoiling). Negative supercoiling favors the unwinding of the DNA double helix, which is required for the initiation of transcription and replication processes (65, 130). As transcription proceeds, DNA in front of the transcription machinery becomes positively supercoiled, and DNA behind the transcription machinery becomes negatively supercoiled. Similarly, during replication, strand separation by the helicase leads to the positive supercoiling of the duplex ahead of the fork (for a review, see reference 138).

Changes in supercoiling in response to external and/or internal stimuli could also play a significant role in the formation and stability of cruciforms. In *E. coli*, superhelicity has been shown to vary considerably during cell growth and to change under different growth conditions (9, 75). Moreover, topology analysis of reporter plasmids isolated from strains where the SOS stress response regulon is constitutively expressed revealed higher levels of negative supercoiling (108). Finally, the level of superhelicity is known to be variable between bacterial species. For instance, the average supercoiling density of a pBR322 reporter plasmid extracted from mid-log-phase cultures of *S. enterica* serovar Typhimurium is 13% lower ($\sigma = -0.060$) than that from *E. coli* ($\sigma = -0.069$) (22).

Effect of cruciform extrusion on DNA topology dynamics. The positioning of IRs within topological domains appears to be another parameter that influences cruciform extrusion. Studies involving the visualization of the cruciform on supercoiled plasmids through atomic force microscopy have shown that extrusion is favored when IRs are positioned at the apex of a plectonemic supercoil (125). Furthermore, cruciforms can exist in two distinct conformations, an X-type conformation and a planar conformation. In the X-type conformation, the cruciform arms form an acute angle, and the main DNA strand is sharply bent, whereas in the planar conformation, the arms are present at an angle of 180°C (139). It has been shown that the rest of the DNA molecule is deeply affected by the conformation adopted by the cruciform. X-type cruciforms tend to localize at the apex of the plectonemic supercoil and restrict the slithering of the molecule; i.e., they reduce the possibility of distant sites coming into contact. Environmental conditions such as salt concentration and protein binding are factors influencing the conformation choice. For instance, the RuvA protein tetramer, which binds to the Holliday junction at the base of cruciforms, forces them into a planar conformation in which the constraints upon DNA movements are relieved (139). It has thus been proposed that cruciform extrusion may act as a molecular switch that can control DNA transactions between distant sites. Such long-range contacts are known to be essential for many cellular processes, including site-specific recombination, transposition, or control of gene expression through DNA loop formation (1, 51, 102, 137).

Genetic Instability of Inverted Repeats

It was quickly noticed that long palindromes cannot be maintained *in vivo* (for a review, see reference 99) either because they are not genetically stable and will be partially mutated or deleted or because they are not viable; i.e., the mol-

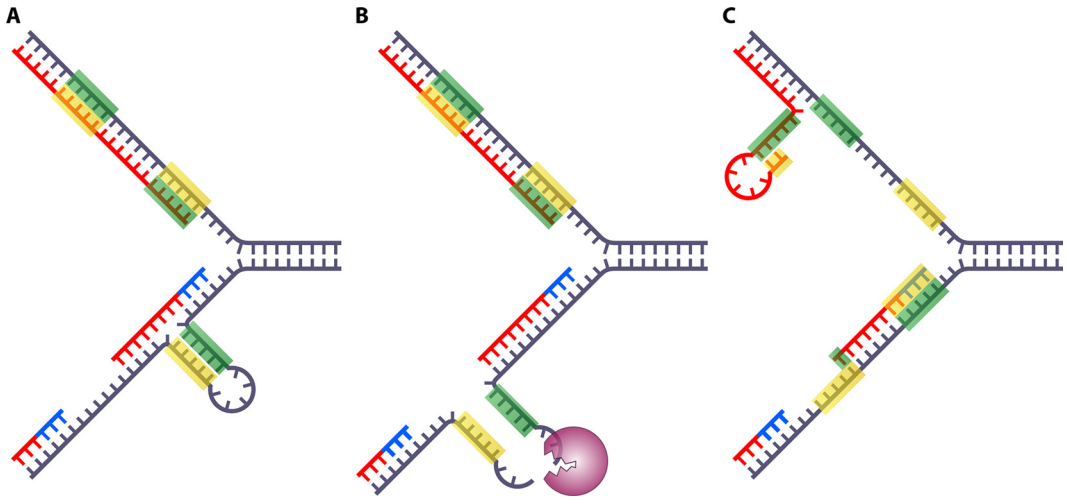

FIG. 3. Genetic instability of inverted repeats. (A) Formation of a hairpin on the template strand can lead to the deletion of the inverted repeat. (B) Hairpins can be cleaved by SbcCD, leading to double-strand breaks that can then be repaired through homologous recombination. (C) Imperfect inverted repeats can mutate toward perfection through a template switch mechanism where the first repeat becomes the template for the second repeat.

ecule carrying them cannot be replicated (28). It is assumed that instability and inviability are caused by the inability of the replication fork to process secondary structures that are too stable and by the presence of proteins destroying these structures. In particular, the SbcCD enzyme can cleave hairpins forming on ssDNA, leading to double-strand breaks that are then repaired by recombination (Fig. 3) (21, 31). This leads to constraints on the size and perfection of the inverted repeats that can be maintained *in vivo*. Typically, a size of 150 to 200 bp is a limit for IRs, although the presence of mismatches and spacers between the repeats strongly improves their maintenance. However, a mutation mechanism that tends to restore perfection to quasipalindromes during chromosomal replication was identified (43). The model proposes that during replication, the nascent DNA strand dissociates from its template strand, forming a partial hairpin loop structure. The nascent strand is then extended by DNA synthesis from the hairpin template, forming a more fully paired hairpin. IRs are thus balanced between a mechanism that tends to restore perfection and the fact that perfect IRs are not genetically stable.

DNA HAIRPIN BIOLOGICAL FUNCTIONS

Hairpins and Replication Origins

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

Priming on the single strand. The first evidence for the role of DNA hairpins in a biological function came from early studies of the primosome. The inability of DNA polymerases to initiate *de novo* replication makes the independent generation of a primer necessary (85). The primosome is a complex of proteins which carries out this priming through the *de novo* synthesis of a small RNA whose 3′ end can be used by the DNA polymerase as a starting point. The role of RNA in priming DNA replication was discovered primarily through studies of single-stranded phages, notably G4 and ϕX174 (95, 155). Single-stranded phages are delivered to the infected cells and have evolved diverse mechanisms for priming the synthesis of the complementary strand, but all the strategies described to date involve DNA hairpins.

(i) G4-type priming. In the region of replication initiation, phage G4 carries three hairpins with stems of 5 to 19 bp and loops of 4 to 8 bases. Early models invoked these structures as recognition sites for the primase DnaG (95). However, it was later shown that none of these hairpins are required for DnaG to initiate primer synthesis in the absence of SSB in *E. coli* (145). The hairpins seem, in fact, to direct the binding of SSB so that the primase recognition site 5′-CTG-3′ is exposed (144). This is likely to be the case for a large number of G4-like phages, including a3, St-1, and ϕK. This is an illustration of how hairpins can direct protein binding and structure an ssDNA region (Fig. 4).

(ii) ϕX174-type priming. Although ϕX174 is a close relative of G4, the priming mechanism leading to cDNA strand synthesis cannot be carried out by DnaG alone. The PriA protein, which is now known to play a major role in stalled replication fork restarts, was first identified as an essential component of the ϕX174 primosome (155). It catalyzes priming from a specific primosome assembly site (PAS) that can adopt a stable secondary structure (5). However, it is now clear that the main PriA substrates are not PAS sites but D loops and R loops encountered during replication, DNA repair, and recombination events. It has thus been proposed that PAS sequences have evolved to mimic the natural targets of PriA (113). A stem-loop formed on a single strand can indeed be viewed as a branched structure between a double strand and two single-strand components (a Y fork). PriA was recently shown to bind Y forks (146). This is an illustration of hairpins that have evolved to be recognized by a host protein to direct primosome assembly (Fig. 4).

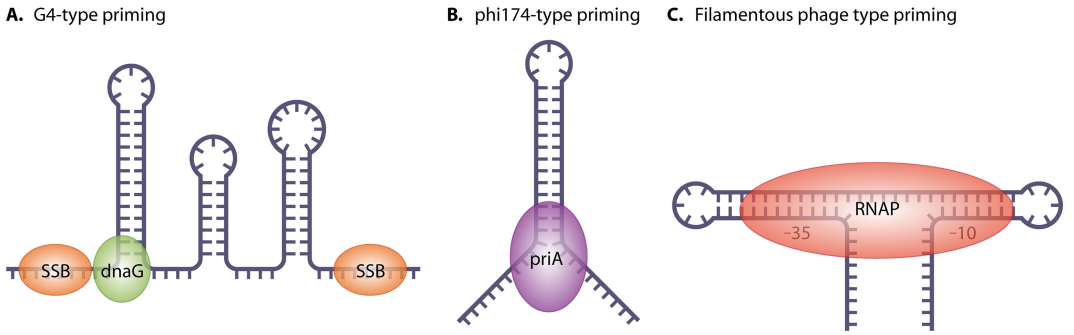

FIG. 4. Priming of replication on ssDNA hairpins. In G4-type priming, hairpins structure the region, directing the binding of SSB and allowing access to the *dnaG* primase. In ϕX174-type priming, an ssDNA hairpin forms a Y fork recognized by PriA, which directs the formation of a primosome. In filamentous-phage-type priming, a hairpin mimicking a promoter is recognized by the RNA polymerase (RNAP), which synthesizes an RNA primer for replication.

(iii) Filamentous-phage-type priming. In the case of the M13 phage and other filamentous phages (f1 and fd), the synthesis of the complementary strand is primed neither by DnaG nor by PriA but by the host RNA polymerase (RNAP) holoenzyme containing the sigma 70 subunit, which synthesizes a 20-nucleotide-(nt)-long RNA primer (63, 78). The RNAP recognizes a double-hairpin structure mimicking a promoter with −35 and −10 boxes (62) (Fig. 4). Here again, hairpins have evolved to be recognized by a host protein. Hairpins recognized by the RNAP have now been associated with several functions [see "(iii) Rolling-circle replication" and "Hairpin promoters" below].

Double-strand DNA replication. The first step in dsDNA replication is the melting of a region where the replication priming complex can load. This melting event is favored, with some exceptions, by a complex of proteins (DnaA for the chromosome or Rep for plasmids) that binds the DNA (usually at direct repeats [DnaA boxes or iterons]) and bends it (79, 84, 119). This bending promotes DNA melting but also the formation of alternative DNA structures.

A common feature of many origins of replication is the presence of inverted repeats (IRs). The extrusion of IRs as cruciforms is energetically more favorable than simple DNA melting and is thus very likely to occur, absorbing a part of the strain generated. Furthermore, when DNA melting actually occurs (which is favored by AT-rich regions present in most *ori*'s), IRs are free to fold into hairpins. There is thus ample opportunity at origins of replication for a DNA structure to arise and interact with proteins.

Hairpins have also been shown to play essential roles in primosome assembly in dsDNA replication. The generation of a primer occurs in two major ways: the opening of the DNA double helix followed by RNA priming (chromosomal, theta, and strand displacement replications) or the cleavage of one of the DNA strands to generate a 3′-OH end (RCR) (40, 82). For both mechanisms, cases where hairpins play essential roles have been described.

(i) Chromosomal and theta replication. The term "theta replication" was coined after the "theta" shape of the plasmids that carry out this type of replication (40). Similarly to chromosomal replication, it involves the melting of the parental strands, synthesis of a primer RNA, and initiation of DNA synthesis by the covalent extension of the primer RNA.

The DnaA protein or a plasmid-encoded Rep initiator protein is involved in the control of replication initiation, unwinding of the helix, and recruitment of the priming complex (for a review, see reference 119). It has been proposed that in some replication origins, a hairpin structure carrying a DnaA box folds in the region unwound by DnaA itself. This hairpin, named M13-A, is at the core of the ABC priming mechanism first described for the R6K plasmid (111). M13-A is specifically bound by DnaA, which then recruits DnaB and DnaC and finally initiates RNA priming. Putative M13-A hairpins are present in a large number of theta replicating plasmids, and this mechanism was proposed to occur at the *E. coli* origin of replication (20). However, there is to date no direct experimental evidence that this occurs *in vivo*, and the currently accepted model for the *E. coli* origin of replication does not invoke unwound DNA with hairpins.

Inverted repeats other than M13-A called single-stranded initiators (*ssi*'s) are often present at replication origins and can be involved in RNA priming. In the same way that filamentous phages prime complementary-strand synthesis, the F-plasmid origin of replication has a hairpin (*ssiD* or F*rpo*) recognized by *E. coli* RNAP, which synthesizes an RNA primer (110). Other *ssi*'s have been isolated from a variety of plasmids and shown to use ϕX174-type priming involving PriA (for a review, see reference 109).

(ii) Strand displacement replication. The best-described example of strand displacement replication is plasmid RSF1010. The plasmid-encoded RepC protein binds to iterons and unwinds the DNA in a region carrying two single-stranded initiators (*ssiA* and *ssiB*). These sequences are IRs, which fold into hairpins. The secondary structures of these hairpins and parts of their sequences have been shown to be essential for replication (116). The current model states that plasmid-encoded RepB primase specifically recognizes *ssiA* and *ssiB* and primes continuous replication from these sequences (67–69). However, it is not clear whether *ssiA* and *ssiB* fold when the region

is largely single stranded or whether they extrude as a cruciform, thanks to the action of RepC.

(iii) Rolling-circle replication. RCR is widely present among plasmids and viruses (including the filamentous phages mentioned above), with the model being plasmid pT181 (for a review, see reference 82). The plasmid-encoded Rep protein binds to the double-stranded origin of replication (*dso*) and bends the DNA, producing a strain leading to the extrusion of a hairpin carrying the Rep nicking site. This structure was among the first cruciforms probed *in vivo* (123). Rep nicks DNA in the hairpin and becomes covalently attached to the 5′ phosphate (Fig. 5). The free 3′-OH end serves as the primer for leading-strand synthesis. No synthesis occurs on the lagging strand until it is completely unwound by the helicase and released as ssDNA. The synthesis of the complementary strand is then initiated at the single-strand origin (*sso*). Four classes of *sso* have been described (*ssoA*, *ssoW*, *ssoT*, and *ssoU*). These classes have little nucleotide sequence homology but share structural features (89) necessary for their recognition by the RNAP, which primes complementary-strand synthesis (89, 91, 92).

Hairpins and Transcription

There are essentially three ways in which hairpins and cruciforms can affect transcription. (i) The extrusion of a cruciform dramatically reduces the local supercoiling of DNA. Since superhelical density is known to affect the activity of promoters, cruciform extrusions in promoter regions could reduce their activity (153). (ii) A cruciform could prevent proteins from binding to their cognate site if it overlaps the extruding sequence. (iii) RNA polymerases or transcription factors could recognize hairpins present on ssDNA or extruded from dsDNA. Since there is as yet no documented case for the first possibility, only the two other mechanisms are discussed here.

Hairpin promoters. We have discussed how the RNAP can recognize hairpin promoters to prime DNA replication (rolling-circle replication, filamentous-phage-type priming, and F-plasmid replication). The RNAP primes F-plasmid replication through the recognition of the F*rpo* hairpin, but under certain conditions, it can produce transcripts longer than the one needed for priming and express the downstream genes (110). This allows the plasmid to express the downstream genes as soon as it enters the recipient cell and before the complementary strand is synthesized.

Accordingly, transcription from a structured single-stranded promoter was suggested to occur during conjugative DNA transfer for several *oriT*-associated genes of enterobacterial conjugative plasmids, namely, *ssb*, *psiB*, and, sometimes, *ardA* (3, 76, 121). Considering that conjugation consists of ssDNA entry into the recipient cell, the products of these genes, single-strand binding, anti-SOS, and antirestriction, respectively, could be needed for maintaining the plasmid in the recipient. Indeed, the transcriptional orientation of these genes, always on the leading strand, means that the transferred strand is destined to be the transcribed strand (25). Moreover, the induction of these first loci was shown to be transfer dependent (76). The burst of activity observed shortly after the initiation of conjugation led to the proposal that this early transcription could be mediated by the presence of a secondary structure in the transferred ssDNA (3, 124) that mimics an RNA polymerase promoter recognized by the F*rpo* sigma factor (110).

Other hairpin promoters that are not involved in priming have been described. Notably, the N4 virion carries three hairpin promoters specifically recognized by the virion RNA polymerase (vRNAP) and used to direct the transcription of the phage early genes (Fig. 6). Upon infection of *E. coli*, the N4 double-stranded DNA injected into the cell is supercoiled by the host DNA gyrase, which leads to the extrusion of hairpin promoters as cruciforms (32, 33).

Promoter inhibition through cruciform extrusion. Early studies have shown how an artificial IR overlapping a promoter can regulate transcription by superhelix-induced cruciform formation (70). Although promoters usually have higher levels of activity with increasing superhelical densities, such a promoter has a lower expression level at a high superhelix density because of the extrusion of the IR as a cruciform preventing RNAP binding. It has also been shown that the N4 hairpin placed between the −10 and −35 boxes of the *rrnB* P1 promoter can repress its activity in a supercoil-dependent manner (32). DNA cruciform extrusion seems likely to be a mechanism for the regulation of genes repressed by supercoiling. However, it is not clear how common this mechanism of regulation is, since no compelling natural example has been reported. The *bgl* operon promoter, which presents a 13-bp IR, was first thought to be a natural example of such regulation (142). However, it was later shown that no cruciform is required to account for its supercoiling-dependent repression (19).

Hairpins and Conjugation

IRs are present in a majority of origins of transfer (*oriT*) (45). The best described is the origin of transfer of R388, where an IR named IR2, located 5′ to the nicking site, plays an essential role (54). Conjugation occurs as follows: DNA is nicked at *oriT* and bound covalently by the plasmid-encoded relaxase protein TrwC. The T strand is then unwound through rolling-circle replication and transferred to the recipient cell. Although the folding of IR2 into a hairpin is not required for the initial nicking of *oriT*, the recircularization of the T strand requires the folding of IR2 into a hairpin specifically recognized by the relaxase (53).

In addition to IR2, other IRs important for transfer efficiency are present in the R388 *oriT* (103), but their exact role remains to be elucidated. It is not yet known whether their sequence or structure is important. They probably help adapt *oriT* into a potentially active state through cruciform formation.

The structures of two relaxases other than TrwC have been determined by crystallography: the F-plasmid relaxase TraI (35) and the R1162 plasmid relaxase MobA (117). Although they show poor sequence homology to TrwC, the three-dimensional (3D) structures of all these relaxases are very similar. These enzymes are evolutionarily homologous and certainly have identical mechanisms of action.

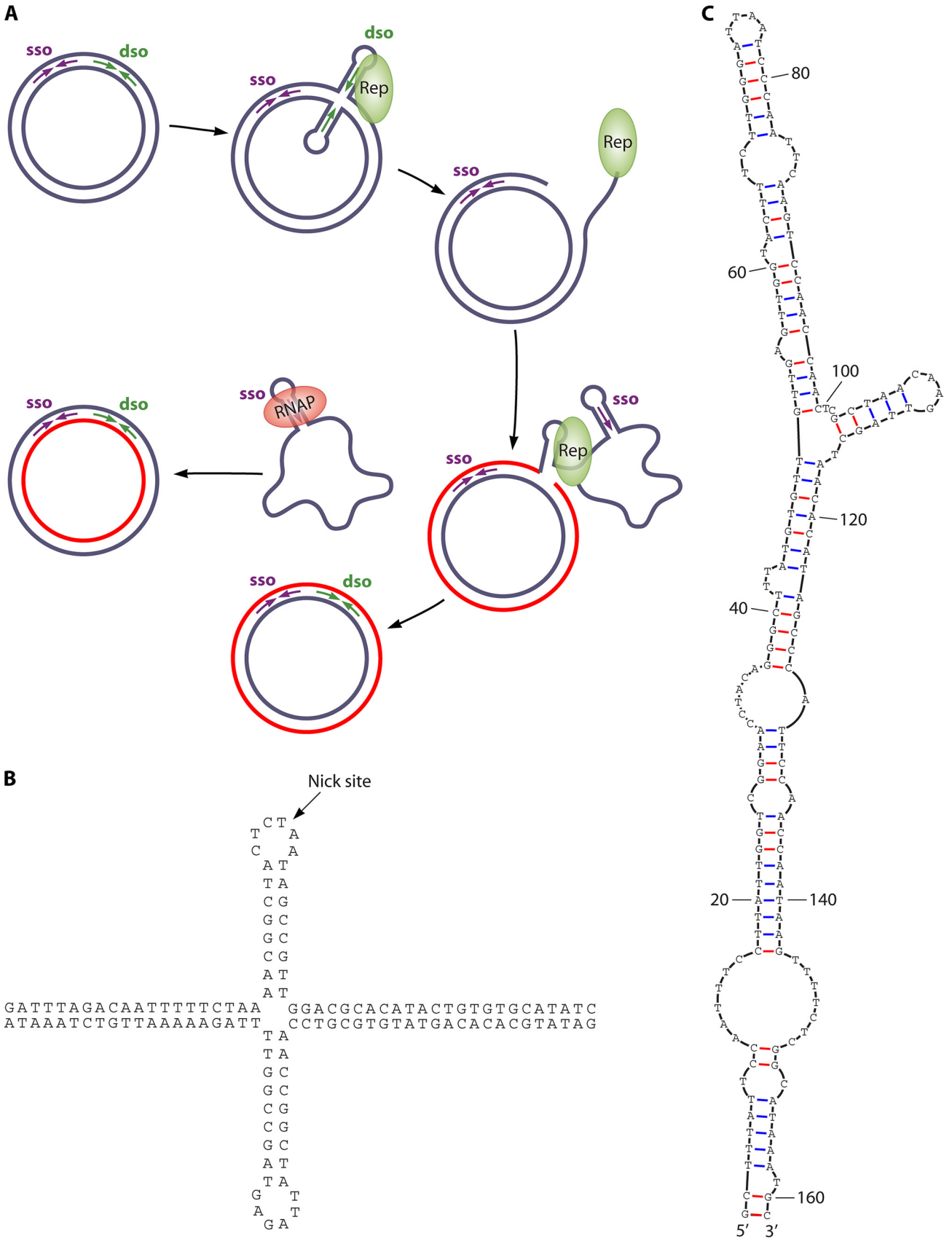

FIG. 5. Rolling-circle replication. (A) The Rep protein binds a hairpin formed by double-stranded origin (*dso*) and extruded from dbDNA as a cruciform. Rep nicks DNA and covalently binds the 5′ end, leaving a 3′ end for replication to proceed. The leading strand is replicated while the lagging strand is extruded and remains single stranded until the single-stranded origin (*sso*) is reached. The 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 use of mFOLD software.

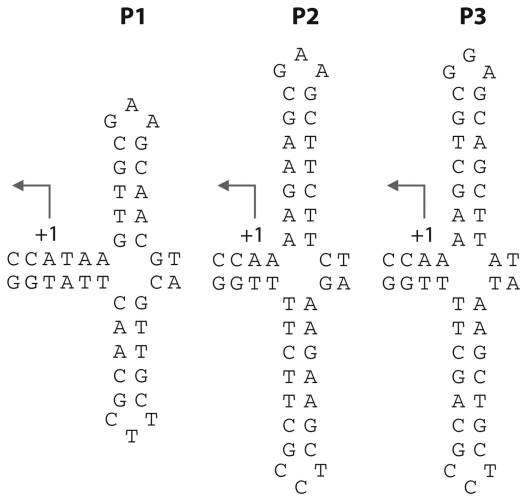

FIG. 6. N4 virion hairpin promoters. Shown are the three promoters of N4 controlling the expression of the early genes as cruciform structures.

Hairpins and Recombination

To date, there are three compelling examples of recombination systems using DNA hairpins as substrates: the CTX phage recombination site, the IS*200*/IS*605* insertion sequence (IS) family, and integron *attC* recombination sites.

The single-stranded CTX phage of *Vibrio cholerae*. CTX is a single-stranded phage involved in *V. cholerae* virulence. In the lysogenic phase, it integrates *V. cholerae* chromosome I or II at its respective *dif1* and *dif2* sites. Chromosomal *dif* sites are recombination sites recognized by the XerCD protein complex, which solves concatemers and allows proper chromosome segregation. CTX enters the infected cells as ssDNA, and the single-stranded form is integrated directly into one of the chromosomes (150). The *attP* recombination site of CTX carries a ~150-bp forked hairpin, which is homologous to *dif* sites (Fig. 7). The phage uses this hairpin to hijack the host XerCD protein complex, which catalyzes strand exchange between *attP* and the *dif* site (34).

The IS*200*/IS*605* insertion sequence family. The mechanism of transposition of the recently discovered IS*200*/IS*605* insertion sequence family greatly differs from systems already described, in particular those using DDE transposase catalysis (55). The best-studied representative of this family, IS*608*, was originally identified in *Helicobacter pylori* (81). It presents at its ends short palindromes recognized as hairpins by the TnpA transposase. "Top strands" of the two IS ends are nicked and joined together by TnpA a few base pairs away from the hairpins (19 nt upstream from the left hairpin and 10 nt downstream from the right hairpin) (10, 58). TnpA then catalyzes the formation of a single-stranded transposon circle intermediate, which is then inserted specifically into a single-stranded target. This target site is not recognized directly by TnpA but by 4 bases at the foot of the hairpin in the transposition circle (Fig. 8) (57) that undergo unconventional base pairing with the ssDNA target sequence.

The IS*91* insertion sequence. IS*91* is a member of an insertion sequence family displaying a unique mechanism of transposition. The IS*91* transposase is related to replication pro-

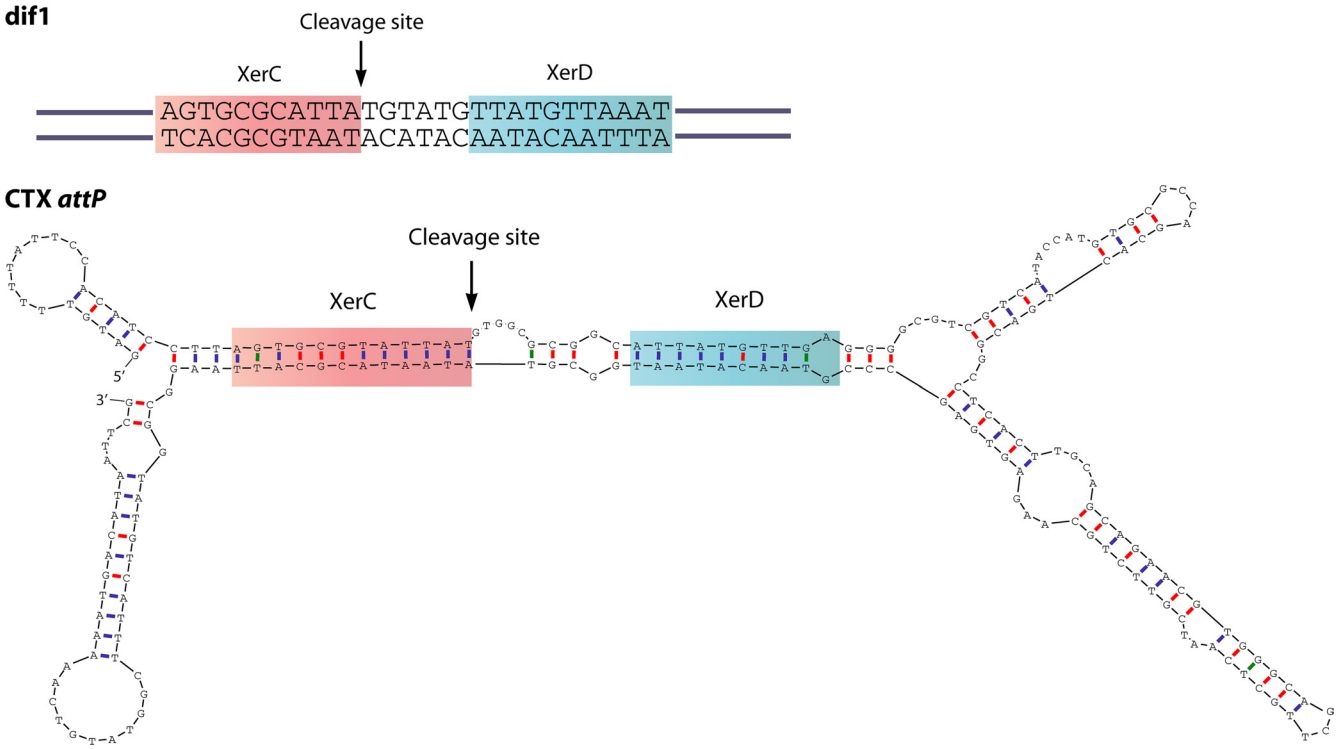

FIG. 7. The *V. cholerae* chromosome I *dif* site and the CTX phage hairpin. The CTX *attP* region folds into a forked hairpin mimicking *V. cholerae dif1*. This enables the CTX phage to use the host XerCD recombinase to catalyze its integration into the chromosome.

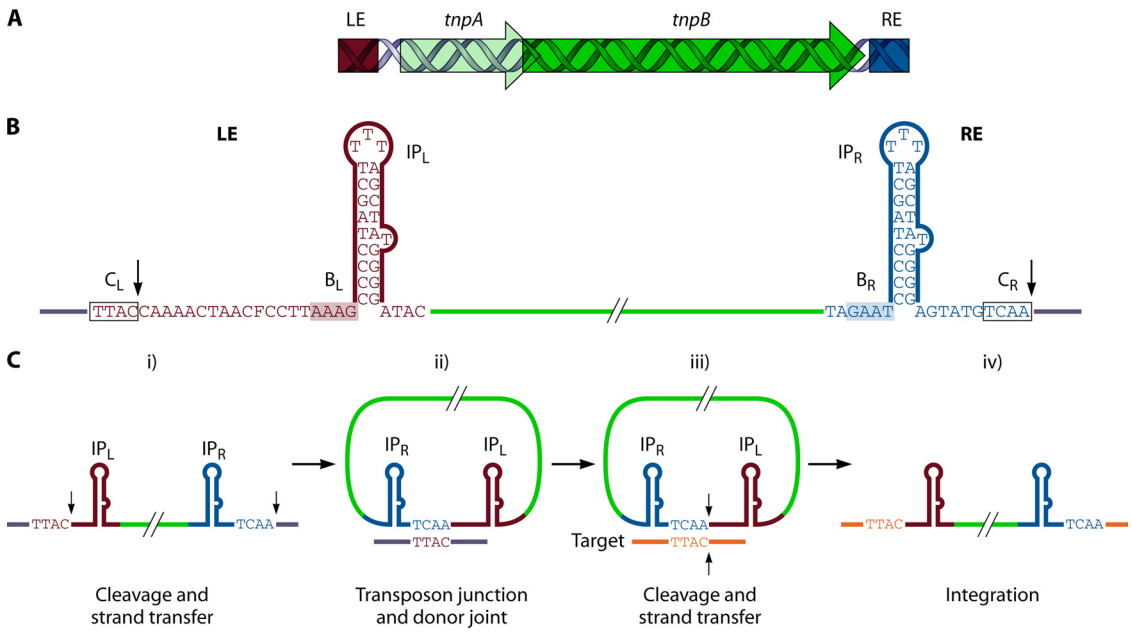

FIG. 8. Organization of IS*608* and overall transposition pathway. (A) Organization. Shown are *tnpA* and *tnpB* open reading frames (light and dark arrows, respectively) and the left end (LE) and right end (RE) (red and blue boxes, respectively).(B) Sequence of the LE and RE. Sequence and secondary structures, IP_L and IP_R, at the LE and RE of IS*608* are shown. Left and right tetranucleotide cleavage sites (C_L and C_R, respectively) are boxed in black (C_L) and underlined in blue (C_R). They are recognized by the B_L and B_R tetranucleotide boxes, respectively, through folding and unconventional base pairing. Also shown is the position of cleavage and of the formation of the 5′ phosphotyrosine TnpA-DNA intermediate (vertical arrows). (C) Transposition pathway. (i) Schematized IS*608* with IP_L and IP_R and left (TTAC) (C_L) and right (TCAA) (C_R) cleavage sites. (ii) Formation of a single-strand transposon circle intermediate with abutted left and right ends. The transposon junction (TCAA) and donor joint (TTAC) are shown. (iii) Pairing with the target (TTAC) and cleavage (vertical arrows). (iv) Inserted transposon with new left and right flanks (dotted black lines). (Reprinted from reference 57 with permission of the publisher.)

teins of RCR plasmids. IS*91* transposition involves an ssDNA intermediate generated in a rolling-circle fashion (115). Short palindromes have been identified in the regions essential for transposition just a few base pairs away from the recombination sites. Their exact functions have not been studied. Nevertheless, striking similarities between these regions, RCR plasmid *dso*, and conjugation *oriT*s suggest that these palindromes might fold into hairpins recognized by the IS*91* transposase.

Integrons. Integrons are natural recombination platforms able to stockpile, shuffle, and differentially express gene cassettes. Discovered by virtue of their importance in multiple-antibiotic resistances, they were later identified in 10% of sequenced bacterial chromosomes, where they can contain hundreds of cassettes (13). The cassettes are generally single open reading frames (ORFs) framed by *attC* recombination sites (131). When expressed, the integron integrase can recombine *attC* sites, leading to the excision of a circular cassette. Such a cassette can then be integrated at a primary recombination site named *attI*. *attC* recombination sites have been shown to be recognized and recombined by the integrase

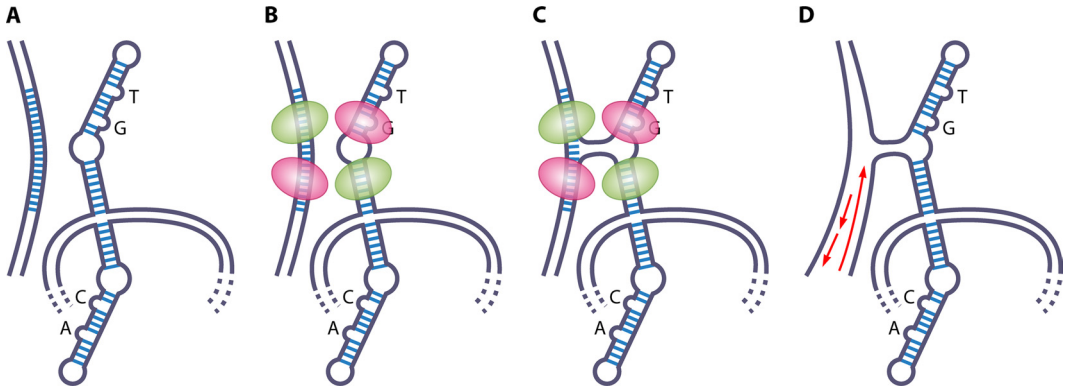

FIG. 9. Recombination between an *attC* site hairpin of an integron cassette and a double-stranded *attI* site. The first recombination steps (A to C) between the folded *attC* site and the dsDNA *attI* site are identical to classical recombination steps catalyzed by other tyrosine recombinases. (B) Four integrase monomers bind to the core sites (with the proper strand of the *attC* site being recognized through specific binding with the extrahelical G). (C) Binding to structural determinants makes the pink monomers inactive, leaving the green monomers the possibility to realize the first strand exchange. The pseudo-Holliday junction formed cannot be resolved by a second strand exchange, as occurs with classical tyrosine recombinases. (D) The current model is that replication is involved to solve the junction in a process that remains to be understood.

only as hairpins (Fig. 9) (14, 112). A surprising feature of *attC* hairpins is their huge polymorphism. Their stem length ranges from 54 to 80 bp, and their loop length ranges from 3 to 80 bp. Highly conserved mismatches known to be involved in hairpin recognition by the integrase are also present (14, 15) (see "Strand Selectivity" below).

Other Hairpin DNAs: Phage Packaging and Retrons, etc.

Single-stranded phage packaging. The single-stranded filamentous phages (f1, fd, M13, and Ike) contain IRs that can fold into hairpins. We have described above the hairpins involved in complementary-strand synthesis, but the largest hairpin identified on these genomes is the packaging signal (PS) recognized in the translocation of ssDNA into the virion capsid. This hairpin is probably recognized by the phage transmembrane protein pI and determines the orientation of DNA within the particle (135). Both the structure and sequence determinants of the PS hairpin are required for its function (136).

Retrons. Retrons are DNA sequences found in the genomes of a wide variety of bacteria (96). They code for a reverse transcriptase similar to that produced by retroviruses and other types of retroelements. They are responsible for the synthesis of an unusual satellite DNA called msDNA (multicopy single-stranded DNA). msDNA is a complex of DNA, RNA, and probably protein. It is composed of a small single-stranded DNA linked to a small single-stranded RNA molecule folded together into a secondary structure. msDNA is produced in many hundreds of copies per cell (96). Whether msDNAs are selfish elements or play a role in the cell remains to be discovered.

HAIRPIN FORMATION: CRUCIFORM EXTRUSION VERSUS SINGLE-STRANDED HAIRPIN

Under what conditions do DNA hairpins fold? Do they extrude from the double helix as cruciforms, or do they fold from ssDNA during replication, repair, or horizontal gene transfer? Both the single-stranded phage hairpins and the *sso* of RCR plasmids obviously fold from ssDNA. On the other hand, there is consistent evidence that the N4 hairpin promoters and the hairpin of the RCR plasmid *dso* fold as cruciforms (32, 123). However, there are only a few cases of successful cruciform detection of natural IRs *in vivo*. Indeed, most reported *in vivo* cruciform detections involved artificial conditions favoring hairpin extrusion: small loops, IRs in AT-rich regions, perfect palindromes with AT-rich centers and GC-rich stems, topoisomerase mutants, or salt shock to increase supercoiling (141, 160, 161).

Ton-Hoang and colleagues have recently uncovered how the transposition of IS*200*/IS*605* family members is coupled with replication (148). Those authors were able to show that the excision of IS*608* is greatly stimulated when the recombinogenic "top strand" is on the lagging strand template. Furthermore, in their experiments, integration events occurred exclusively on the lagging-strand template, in agreement with *in silico* data showing that the orientation of IS*200*/IS*605* family members in their respective host genomes is strongly skewed in this direction. Interestingly, this integration preference could be abolished in the case of the transposition of another member of the family, IS*Dra2* in *Deinococcus radiodurans*, when cells were subjected to gamma irradiation (148). This treatment induces a repair pathway resulting in massive amounts of ssDNA with no strand bias. This observation is consistent with transposition events occurring on ssDNA generated during DNA repair. Such events might account for the few cases where IS*200*/IS*605* family members are found to be integrated on the leading strand of the replication fork, and cruciforms are probably not involved in the transposition of these elements.

We recently investigated the conditions that can lead to integron *attC* site folding (104). These recombination sites are extremely good candidates for the study of hairpin formation *in vivo*. Recombination events can occur only with folded *attC* sites and can be detected at very low frequencies. Furthermore, only the bottom strand of the *attC* site is recognized by the integrase (14, 46). This enables the distinguishing of recombination events occurring with hairpins formed during replication on the template for lagging-strand synthesis from events occurring with hairpins extruding as cruciforms or during other processes such as repair. Apparently, *attC* hairpins fold much more frequently during replication on the lagging-strand template than through other processes. However, it was noted that the recombinogenic strand of *attC* sites is always found on the leading strand template in natural chromosomal integrons (104). Recombination in chromosomal integrons can therefore happen only with sites folded as cruciforms or during DNA repair. This contrasts with the IS*200*/IS*605* family elements, which are almost always oriented so that the recombinogenic strand is on the lagging-strand template, where it can take advantage of the ssDNA produced between the Okazaki fragments to recombine. The pathways and conditions in which these two systems recombine are thus likely to be different. It is important to note that *attC* sites are imperfect IRs with at least two extrahelical bases, a bulge of 4 to 5 bp, and a spacer sequence between the IRs (the loop of the hairpin, called the variable terminal sequence) of up to 80 bp. Such imperfections are known to hinder cruciform formation, and the extrusion of imperfect IRs has been detected *in vitro* only for very AT-rich IRs (12). Nevertheless, the transformation of nonreplicative plasmids carrying *attC* sites into cells where they could be maintained only after a recombination event enabled us to show that *attC* sites can extrude cruciforms at low frequencies ($<10^{-3}$). Most surprisingly, *attC* sites with large spacer sequences (80 bp) between the repeats were also able to fold cruciform structures. Integron cassettes are particularly AT rich (112), which could favor *attC* site extrusion following a C-type mechanism. The biological relevance of cruciform extrusion in natural integrons remains to be properly investigated, together with the role of DNA repair, which might be the main mechanism by which ssDNA production could lead to integron recombination.

In summary, large perfect IRs can presumably fold into cruciforms but are genetically unstable because of their propensity to hinder replication and be cleaved by SbcCD. Small perfect (or almost-perfect) IRs can fold into cruciforms only when their sequence and context allow it. The N4 promoters and plasmid pT181 origin of replication are examples of such IRs with biological functions. Imperfect IRs are genetically

more stable regardless of their size but fold into cruciforms only rarely. They could still be involved in biological functions that take place at low frequencies, such as integron recombination. Alternatively, imperfect IRs present in topologically constrained regions such as replication origins could also fold into cruciforms, which might be the case for the M13-A hairpin and for the *ssi* present in some origins of replication. Note that these hairpins are specifically bound by cognate proteins that could stabilize cruciforms.

PROTEIN/HAIRPIN RECOGNITION

Mimicry: Subverting the Host Proteins

Some of the hairpins described in the literature have evolved to mimic the "natural" target of the proteins with which they interact. The PAS sequences of single-stranded phages mimic Y forks that are recognized by PriA. The *sso* of RCR plasmids, the F*rpo* hairpin, and the filamentous-phage priming hairpins all mimic promoters recognized by the host RNAP. The M13-A hairpin mimics a natural *dnaA* box, and the CTX *attP* recombination site mimics the *V. cholerae dif* sites recognized by XerCD.

There is a noteworthy difference between hairpins like the CTX *attP* site, where mimicry is clear-cut, and the variety of hairpins recognized by RNAP. The latter indeed display an impressive diversity of structures and sequences. Although elements of the *ssoA* class present a large hairpin with near-consensus −35/−10 boxes (92), other *sso* classes, like *ssoU*, present much more complex structures with several hairpins and −35/−10 boxes that are harder to recognize (89). Another structural variation is that used by the filamentous phages. Here, a double hairpin acts as the recognition site, with the −35 box on one stem-loop and the −10 box on the other (62). The fact that they are all recognized by RNAP suggests a poor specificity of RNAP binding to hairpin DNA. The few common features of all these sequences are the widespread presence of mismatches in the hairpins and the fact that they do not work as promoters in the dsDNA form but bind RNAP very strongly when single stranded (in some cases even more strongly than strong double-stranded promoters [62]). These observations are consistent with the fact that sigma A and sigma 70 of *B. subtilis* and *E. coli*, respectively, bind strongly to ssDNA-containing promoter −10 sequences (73). The mismatches that often span the −10 box could be there to ease access for RNAP and increase hairpin-promoter activity. A high level of activity might be required by single-stranded molecules, which need to synthesize their complementary strand promptly before triggering the SOS response of the host, as was observed for phages defective in complementary-strand synthesis (64).

In all these cases, the mimicry of dsDNA is not perfect: to different extents, mismatches are present in the hairpins. These mismatches are probably, in some cases, necessary for the maintenance of long IRs *in vivo*, as discussed above, but do they have a role in and an impact upon hairpin recognition? CTX might be the only mimicry case in which imperfection has a clear function: mismatches are essential for the irreversibility of single-stranded phage integration (150).

Protein Recognition of Hairpin Features

Other systems have evolved proteins recognizing special features of hairpin DNA. This is the case for the integron integrase IntI, for the IS*200*/IS*605* family transposase TnpA, for mobilizable plasmid relaxases (TrwC, etc.), for N4 virion RNAP, and probably for the strand displacement replication protein RepB. The features that make a hairpin structurally different from dsDNA are essentially (i) the bottom of the stem, which can be either a Y fork or a Holliday junction depending on whether the hairpin forms on ssDNA or as a cruciform; (ii) the loop, which is single stranded; and (iii) extrahelical bases and bulges produced by mismatches between the IRs.

The crystal structure of the interaction between IntI, N4 vRNAP, TnpA, TrwC, and their cognate hairpins has been obtained (52, 54, 107, 133). All four highlight different mechanisms of recognition. IntI binds as a dimer to the stem of the hairpin and specifically recognizes two extrahelical bases. A central bulge in the stem also seems to be important for the formation of a recombination synapse involving four IntI monomers. N4 vRNAP presents a base-specific interaction with the single-stranded loop of the hairpin and fits the stem structure through interactions with the phosphate-and-sugar backbone. TnpA binds the stem primarily through contact with the phosphate backbone but also shows a base-specific interaction with the bases of the loop and, importantly, with an extrahelical T in the middle of the stem. Finally, the TrwC interaction is somewhat different from the others, since it binds not only to the hairpin structure but also to the ssDNA 3′ to the stem-loop, where the nicking site is present. The binding to the ssDNA part is base specific, whereas the interaction with the hairpin occurs essentially through contact with the DNA backbone (54).

Strand Selectivity

Whether it be during phage complementary-strand synthesis, at the *sso* of RCR plasmids, or during conjugation, only one DNA strand is available. In these cases, the question of strand selectivity is not physiologically relevant. However, when both DNA strands are free to fold into hairpins, the erroneous recognition of one strand over the other may be problematic. Indeed, an inverted repeat, once folded, generates the same hairpins on the top and bottom strands, except for the loop and eventual bulges and extrahelical bases. Still, in all the processes in which a protein recognizes hairpin features, strand selectivity has been observed: the protein recognizes one strand and not the other. In light of the hairpin-protein interactions described above, it is easy to understand how proteins discriminate between the two strands. They all show base-specific interactions with bases either in the loop, at the single-stranded base of the stem, or with extrahelical bases. Any of these interactions can account for strand selectivity. Some of these systems appear to have good reason to process one strand and not the other. The N4 virion needs to initiate transcription in the right direction. Recombination of the wrong strand for integron cassettes would lead to their integration in the wrong direction, where they could not be transcribed. Finally, if a different strand of IS*608* is recognized at each end of the

transposon, this would lead to the junction of the top strand with the bottom strand, a configuration that cannot be processed further and that is likely to be lethal. Therefore, one strand had to be chosen, and the other had to be strongly discriminated against.

EVOLUTION OF HAIRPINS WITH BIOLOGICAL FUNCTIONS

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

Single-Stranded DNA, Stress, and Horizontal Transfer

We have seen that hairpin formation in the cell is most likely to occur in the presence of ssDNA in the cell. Intracellular single-stranded DNA triggers the SOS response (Fig. 10). ssDNA is the substrate for RecA polymerization. The formation of a RecA nucleofilament on ssDNA stimulates the self-cleavage of the general repressor LexA, leading to its inactivation. Promoters from the SOS regulon, controlling mostly DNA repair, recombination, and mutagenic polymerases, are than derepressed (Fig. 10).

SOS is thus induced when an abnormal amount of ssDNA is present in the cell. The formation of hairpins from ssDNA is thus likely to occur in a context where the SOS response is activated. The induction of the SOS response is often synonymous with stress. This happens, for example, when the cell tries to replicate damaged DNA, causing replication forks to stall (152). Another source of ssDNA comes from DNA intake by horizontal gene transfer and phage infection. For instance, the conjugative transfer of R plasmids, conjugative plasmids carrying multiple resistances, has been shown to induce the SOS stress response in the recipient cell, except when an anti-SOS factor is encoded by the plasmid (*psiB*, mentioned above in "Hairpins and Transcription") (7, 8). Interestingly, the expression of these anti-SOS genes is under the control of ssDNA promoters, i.e., of hairpin substrates.

Furthermore, in the case of integrons, the expression of the integrase (*intI*) has recently been shown to be controlled by SOS (56). Some antibiotics are known to induce the SOS response in Gram-negative and Gram-positive bacteria (80). These antibiotics, such as quinolones, trimethoprim, and beta-lactams, were tested and found to be inducers of the expression of the *intI* promoter. This is certainly a way for integrons to "know" when potential substrates are present in the cell and to recombine them. Indeed, the induction of SOS during the conjugative transfer of R plasmids results in the induction of the integrase, allowing genome rearrangements in the recipient bacterium (8). Furthermore, integrons are often found on conjugative plasmids and may well take advantage of the single-stranded transfer to acquire cassettes and spread horizontally. Similarly, for IS*200*/IS*605* family members, specific integration into the ssDNA substrate has been proposed as a mechanism for targeting mobile elements and ensuring interbacterial spread (58). Gamma irradiation has been shown to increase the frequency of transposition of IS*Dra2*, a member of this family (148). It is also known that SOS induces the transposition of other classes of insertion sequences such as IS*10* (44) and, possibly, of Tn*1*, Tn*5*, and Tn*10* (2).

Not only does the SOS response promote genetic rearrangements, it also induces horizontal gene transfer. It is known, for instance, that stress can induce competence in some bacteria (27) (Fig. 10). Another effect of SOS induction is the derepression of genes involved in the single-stranded transfer of integrating conjugative elements (ICEs), such as SXT from *V. cholerae*, which is a ~100-kb ICE that transfers and integrates the recipient bacterium's genome, conferring resistance to several antibiotics (11). Moreover, different ICEs are able to combine and create their own diversity in a RecA-dependent manner (i.e., using homologous recombination, which is also induced by SOS) (50, 156). As for R plasmids, SXT transfer was observed to induce SOS in *V. cholerae*. Finally, some lysogenic phages are also known to induce their lytic phase under stressful conditions (48). One might thus see the use of ssDNA by integrons and other recombination systems as a mechanism for evolving: diversity is generated under stressful conditions.

Origins of Folded DNA Binding Proteins

While in many examples described above, one can see that hairpins evolved to subvert the host machinery, in other instances, proteins evolved to specifically and sometimes exclusively recognize hairpin structures. This is the case for the RCR Rep proteins, the relaxases of conjugative elements, the transposase of IS*608*, the integron integrases, and phage N4 vRNAP. Where do these proteins come from, and what pushed them to recognize ssDNA rather than dsDNA?

RCR Rep proteins, relaxases, and IS*608* transposase. Interestingly, the IS*608* transposase as well as conjugative relaxases have been found to be structurally similar to RCR Rep proteins (133). All of these proteins have in common the use of a tyrosine residue to covalently bind DNA. The Rep proteins belong to a vast superfamily spanning eubacteria, archaea, and eukaryotes (74). The superfamily is characterized by two sequence motifs: an HUH motif (histidine-hydrophobic residue-histidine), which coordinates an Mg^{2+} ion and is required for nicking, and a YxxxY motif, where the tyrosines (Y) bind the DNA covalently, with one of the tyrosines being optional (35, 98, 133). All these proteins thus probably have a common ancestor ancient enough to account for the diversity of their functions and their spread among the kingdoms of life. The ability to bind hairpin DNA might have been an important feature in early stages of life when single-stranded DNA might have been more widely present. In this instance, the relaxases of conjugative plasmids obviously need to recognize ssDNA

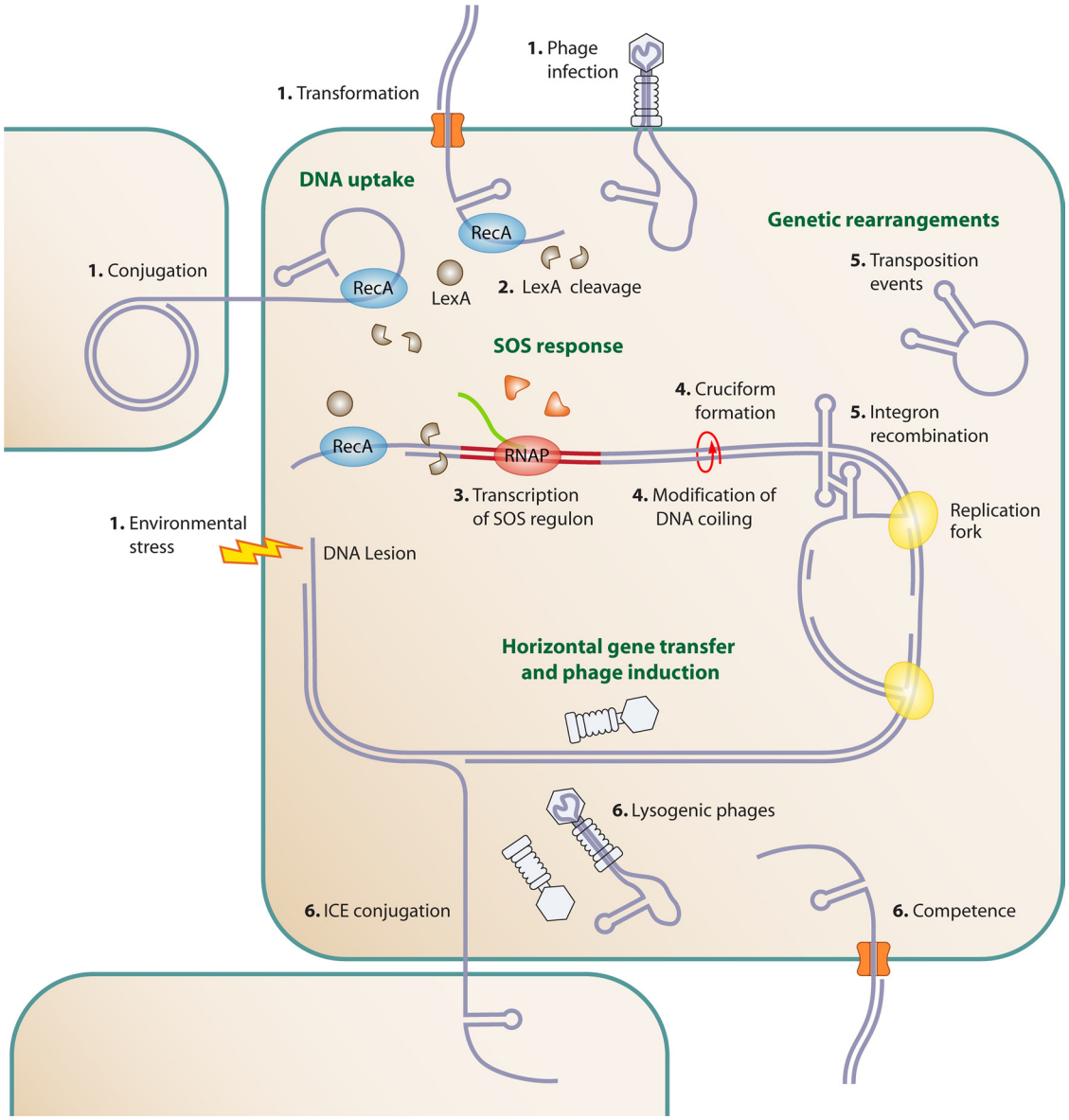

FIG. 10. ssDNA, at the crossroads of horizontal gene transfer, the SOS response, and genetic rearrangements. (1) Conjugation, transformation, phage infection, and environmental stress lead to the production of ssDNA in the cell. (2) The RecA proteins bind ssDNA and trigger the self-cleavage of LexA (brown 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.

features to process the ssDNA in the recipient cell. The recombination of ssDNA by the IS*608* transposase is probably a way to target mobile elements and to ensure their spread. Finally, the reason why RCR plasmid Rep proteins would recognize hairpins rather than the more stable dsDNA is probably that origins of replications need to be strongly negatively coiled to unwind the double helix, and under these conditions, hairpins can be the most stable conformation of DNA.

Integron integrases. Integron integrases (IntI) are also tyrosine recombinases covalently binding DNA. However, they are not related to the Rep protein superfamily. The closest relatives of integron integrases are the XerCD proteins. However, IntI proteins carry an additional domain compared to XerCD. This domain is involved in the binding of the extrahelical bases of the *attC* hairpins that are essential for strand selectivity (15, 107). It would be tempting to speculate that integrons diverged from a single-stranded CTX-like phage that already used XerCD to recombine hairpin DNA. This special feature of ssDNA recombination would then have been selected to form an evolving recombination platform, thanks to its ability to sense both stressful conditions and the occurrence of horizontal gene transfer.

N4 vRNAP. N4 vRNAP is an evolutionarily highly divergent member of the T7 family of RNAPs (36). N4 vRNAP and T7 RNAP recognize their promoter with similar domains and motifs. However, N4 vRNAP recognizes a hairpin, whereas T7

RNAP recognizes dsDNA. The difference lies in the domain interacting with the hairpin loop. It displays substantial architectural complexity and base-specific interactions for N4 vRNAP, whereas the same domain in its counterpart fits just an AT-rich DNA sequence without base recognition (23). The reason why the N4 phage has evolved to transcribe several genes only from cruciform promoters is unclear. It is likely a way for the virion to sense the coiling state of DNA in the cell, which is known to be modified during the cell cycle and is particularly negative during the SOS stress response (108).

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 were suggested to play a major role in genome translocation (94) and replication initiation (157).

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. However, 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, and RNAP) in three different ways. Another example of the opportunistic use of host machinery is the CTX phage that 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 involves 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 IS*608* or integrons. To close the loop, the SOS response can trigger more horizontal transfer, notably through the activation of natural transformation, ICE conjugation, and lysogenic phages.

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

ACKNOWLEDGMENTS

We acknowledge Marie-Eve Val and Adam Wilkins for helpful discussions.

This study was carried out with financial assistance from the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-URA 2171), the French National Research Agency (ANR-08-MIE-016), the European Union (NoE EuroPathoGenomics LSHB-CT-2005-512061), and the Fondation pour la Recherche Médicale (program Équipe FRM 2007).

REFERENCES

1. **Adhya, S.** 1989. Multipartite genetic control elements: communication by DNA loop. Annu. Rev. Genet. **23:**227–250.
2. **Aleshkin, G. I., K. V. Kadzhaev, and A. P. Markov.** 1998. High and low UV-dose responses in SOS-induction of the precise excision of transposons Tn1, Tn5 and Tn10 in Escherichia coli. Mutat. Res. **401:**179–191.
3. **Althorpe, N. J., P. M. Chilley, A. T. Thomas, W. J. Brammar, and B. M. Wilkins.** 1999. Transient transcriptional activation of the IncI1 plasmid anti-restriction gene (ardA) and SOS inhibition gene (psiB) early in conjugating recipient bacteria. Mol. Microbiol. **31:**133–142.
4. **Arai, K., and A. Kornberg.** 1979. A general priming system employing only dnaB protein and primase for DNA replication. Proc. Natl. Acad. Sci. U. S. A. **76:**4308–4312.
5. **Arai, K., and A. Kornberg.** 1981. Unique primed start of phage phi X174 DNA replication and mobility of the primosome in a direction opposite chain synthesis. Proc. Natl. Acad. Sci. U. S. A. **78:**69–73.
6. **Babic, A., A. B. Lindner, M. Vulic, E. J. Stewart, and M. Radman.** 2008. Direct visualization of horizontal gene transfer. Science **319:**1533–1536.
7. **Bagdasarian, M., A. Bailone, M. M. Bagdasarian, P. A. Manning, R. Lurz, K. N. Timmis, and R. Devoret.** 1986. An inhibitor of SOS induction, specified by a plasmid locus in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. **83:**5723–5726.
8. **Baharoglu, Z., D. Bikard, and D. Mazel.** 2010. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. PLoS Genet. **6:**e1001165.
9. **Balke, V. L., and J. D. Gralla.** 1987. Changes in the linking number of supercoiled DNA accompany growth transitions in Escherichia coli. J. Bacteriol. **169:**4499–4506.
10. **Barabas, O., D. R. Ronning, C. Guynet, A. B. Hickman, B. Ton-Hoang, M. Chandler, and F. Dyda.** 2008. Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. Cell **132:**208–220.
11. **Beaber, J. W., B. Hochhut, and M. K. Waldor.** 2004. SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature **427:**72–74.
12. **Benham, C. J., A. G. Savitt, and W. R. Bauer.** 2002. Extrusion of an imperfect palindrome to a cruciform in superhelical DNA: complete determination of energetics using a statistical mechanical model. J. Mol. Biol. **316:**563–581.
13. **Boucher, Y., J. E. Koenig, H. W. Stokes, and M. Labbate.** 2007. Integrons: mobilizable platforms that promote genetic diversity in bacteria. Trends Microbiol. **15:**301–309.
14. **Bouvier, M., G. Demarre, and D. Mazel.** 2005. Integron cassette insertion: a recombination process involving a folded single strand substrate. EMBO J. **24:**4356–4367.
15. **Bouvier, M., M. Ducos-Galand, C. Loot, D. Bikard, and D. Mazel.** 2009. Structural features of single-stranded integron cassette attC sites and their role in strand selection. PLoS Genet. **5:**e1000632.
16. **Brochet, M., C. Rusniok, E. Couve, S. Dramsi, C. Poyart, P. Trieu-Cuot, F. Kunst, and P. Glaser.** 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. Proc. Natl. Acad. Sci. U. S. A. **105:**15961–15966.
17. **Bruni, R., D. Martin, and J. Jiricny.** 1988. d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch. Nucleic Acids Res. **16:**4875–4890.
18. **Burrus, V., and M. K. Waldor.** 2004. Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. **155:**376–386.

19. **Caramel, A., and K. Schnetz.** 1998. Lac and lambda repressors relieve silencing of the Escherichia coli bgl promoter. Activation by alteration of a repressing nucleoprotein complex. J. Mol. Biol. **284:**875–883.
20. **Carr, K. M., and J. M. Kaguni.** 2002. Escherichia coli DnaA protein loads a single DnaB helicase at a DnaA box hairpin. J. Biol. Chem. **277:**39815–39822.
21. **Chalker, A. F., D. R. Leach, and R. G. Lloyd.** 1988. Escherichia coli sbcC mutants permit stable propagation of DNA replicons containing a long palindrome. Gene **71:**201–205.
22. **Champion, K., and N. P. Higgins.** 2007. Growth rate toxicity phenotypes and homeostatic supercoil control differentiate Escherichia coli from Salmonella enterica serovar Typhimurium. J. Bacteriol. **189:**5839–5849.
23. **Cheetham, G. M., and T. A. Steitz.** 1999. Structure of a transcribing T7 RNA polymerase initiation complex. Science **286:**2305–2309.
24. **Chen, I., P. J. Christie, and D. Dubnau.** 2005. The ins and outs of DNA transfer in bacteria. Science **310:**1456–1460.
25. **Chilley, P. M., and B. M. Wilkins.** 1995. Distribution of the ardA family of antirestriction genes on conjugative plasmids. Microbiology (Reading) **141:**2157–2164.
26. **Claverys, J. P., B. Martin, and P. Polard.** 2009. The genetic transformation machinery: composition, localization, and mechanism. FEMS Microbiol. Rev. **33:**643–656.
27. **Claverys, J. P., M. Prudhomme, and B. Martin.** 2006. Induction of competence regulons as a general response to stress in Gram-positive bacteria. Annu. Rev. Microbiol. **60:**451–475.
28. **Collins, J., G. Volckaert, and P. Nevers.** 1982. Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of recA-independent deletion events in Escherichia coli. Gene **19:**139–146.
29. **Courey, A. J., and J. C. Wang.** 1983. Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell **33:**817–829.
30. **Cozzarelli, N. R., and J. C. Wang.** 1990. DNA topology and its biological effects, vol. 20. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. **Cromie, G. A., C. B. Millar, K. H. Schmidt, and D. R. Leach.** 2000. Palindromes as substrates for multiple pathways of recombination in Escherichia coli. Genetics **154:**513–522.
32. **Dai, X., M. B. Greizerstein, K. Nadas-Chinni, and L. B. Rothman-Denes.** 1997. Supercoil-induced extrusion of a regulatory DNA hairpin. Proc. Natl. Acad. Sci. U. S. A. **94:**2174–2179.
33. **Dai, X., and L. B. Rothman-Denes.** 1998. Sequence and DNA structural determinants of N4 virion RNA polymerase-promoter recognition. Genes Dev. **12:**2782–2790.
34. **Das, B., J. Bischerour, M.-E. Val, and F.-X. Barre.** 2010. Molecular keys of the tropism of integration of the cholera toxin phage. Proc. Natl. Acad. Sci. U. S. A. **107:**4377–4382.
35. **Datta, S., C. Larkin, and J. F. Schildbach.** 2003. Structural insights into single-stranded DNA binding and cleavage by F factor TraI. Structure **11:**1369–1379.
36. **Davydova, E. K., I. Kaganman, K. M. Kazmierczak, and L. B. Rothman-Denes.** 2009. Identification of bacteriophage N4 virion RNA polymerase-nucleic acid interactions in transcription complexes. J. Biol. Chem. **284:**1962–1970.
37. **Dayn, A., S. Malkhosyan, D. Duzhy, V. Lyamichev, Y. Panchenko, and S. Mirkin.** 1991. Formation of (dA-dT)n cruciforms in Escherichia coli cells under different environmental conditions. J. Bacteriol. **173:**2658–2664.
38. **Dayn, A., S. Malkhosyan, and S. M. Mirkin.** 1992. Transcriptionally driven cruciform formation in vivo. Nucleic Acids Res. **20:**5991–5997.
39. **de la Cruz, F., L. S. Frost, R. J. Meyer, and E. L. Zechner.** 2010. Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiol. Rev. **34:**18–40.
40. **del Solar, G., R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas.** 1998. Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. **62:**434–464.
41. **Dillingham, M. S., and S. C. Kowalczykowski.** 2008. RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol. Mol. Biol. Rev. **72:**642–671.
42. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu. Rev. Microbiol. **53:**217–244.
43. **Dutra, B. E., and S. T. Lovett.** 2006. cis and trans-acting effects on a mutational hotspot involving a replication template switch. J. Mol. Biol. **356:**300–311.
44. **Eichenbaum, Z., and Z. Livneh.** 1998. UV light induces IS10 transposition in Escherichia coli. Genetics **149:**1173–1181.
45. **Francia, M. V., A. Varsaki, M. P. Garcillán-Barcia, A. Latorre, C. Drainas, and F. de la Cruz.** 2004. A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol. Rev. **28:**79–100.
46. **Francia, M. V., J. C. Zabala, F. de la Cruz, and J. M. Garcia-Lobo.** 1999. The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. J. Bacteriol. **181:**6844–6849.
47. **Furlong, J. C., K. M. Sullivan, A. I. Murchie, G. W. Gough, and D. M. Lilley.** 1989. Localized chemical hyperreactivity in supercoiled DNA: evidence for base unpairing in sequences that induce low-salt cruciform extrusion. Biochemistry **28:**2009–2017.
48. **Galkin, V. E., X. Yu, J. Bielnicki, D. Ndjonka, C. E. Bell, and E. H. Egelman.** 2009. Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. J. Mol. Biol. **385:**779–787.
49. **Gamper, H. B., and J. E. Hearst.** 1982. A topological model for transcription based on unwinding angle analysis of E. coli RNA polymerase binary, initiation and ternary complexes. Cell **29:**81–90.
50. **Garriss, G., M. K. Waldor, and V. Burrus.** 2009. Mobile antibiotic resistance encoding elements promote their own diversity. PLoS Genet. **5:**e1000775.
51. **Gellert, M., and H. Nash.** 1987. Communication between segments of DNA during site-specific recombination. Nature **325:**401–404.
52. **Gleghorn, M. L., E. K. Davydova, L. B. Rothman-Denes, and K. S. Murakami.** 2008. Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. Mol. Cell **32:**707–717.
53. **Gonzalez-Perez, B., M. Lucas, L. A. Cooke, J. S. Vyle, F. de la Cruz, and G. Moncalián.** 2007. Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. EMBO J. **26:**3847–3857.
54. **Guasch, A., M. Lucas, G. Moncalian, M. Cabezas, R. Perez-Luque, F. X. Gomis-Ruth, F. de la Cruz, and M. Coll.** 2003. Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat. Struct. Biol. **10:**1002–1010.
55. **Gueguen, E., P. Rousseau, G. Duval-Valentin, and M. Chandler.** 2005. The transpososome: control of transposition at the level of catalysis. Trends Microbiol. **13:**543–549.
56. **Guerin, E., G. Cambray, N. Sanchez-Alberola, S. Campoy, I. Erill, S. Da Re, B. Gonzalez-Zorn, J. Barbé, M. C. Ploy, and D. Mazel.** 2009. The SOS response controls integron recombination. Science **324:**1034.
57. **Guynet, C., A. Achard, B. T. Hoang, O. Barabas, A. B. Hickman, F. Dyda, and M. Chandler.** 2009. Resetting the site: redirecting integration of an insertion sequence in a predictable way. Mol. Cell **34:**612–619.
58. **Guynet, C., A. B. Hickman, O. Barabas, F. Dyda, M. Chandler, and B. Ton-Hoang.** 2008. In vitro reconstitution of a single-stranded transposition mechanism of IS608. Mol. Cell **29:**302–312.
59. **Hatfield, G. W., and C. J. Benham.** 2002. DNA topology-mediated control of global gene expression in Escherichia coli. Annu. Rev. Genet. **36:**175–203.
60. **Heller, R. C., and K. J. Marians.** 2006. Replication fork reactivation downstream of a blocked nascent leading strand. Nature **439:**557–562.
61. **Herrmann, R., K. Neugebauer, H. Zentgraf, and H. Schaller.** 1978. Transposition of a DNA sequence determining kanamycin resistance into the single-stranded genome of bacteriophage fd. Mol. Gen. Genet. **159:**171–178.
62. **Higashitani, A., N. Higashitani, and K. Horiuchi.** 1997. Minus-strand origin of filamentous phage versus transcriptional promoters in recognition of RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. **94:**2909–2914.
63. **Higashitani, N., A. Higashitani, Z. W. Guan, and K. Horiuchi.** 1996. Recognition mechanisms of the minus-strand origin of phage f1 by Escherichia coli RNA polymerase. Genes Cells **1:**829–841.
64. **Higashitani, N., A. Higashitani, A. Roth, and K. Horiuchi.** 1992. SOS induction in Escherichia coli by infection with mutant filamentous phage that are defective in initiation of complementary-strand DNA synthesis. J. Bacteriol. **174:**1612–1618.
65. **Hirose, S., and K. Matsumoto.** 2005. Possible roles of DNA supercoiling in transcription, p. 138–143. *In* T. Ohyama (ed.), DNA conformation and transcription, vol. XII. Springer Sciences, New York, NY.
66. **Hochhut, B., J. Marrero, and M. K. Waldor.** 2000. Mobilization of plasmids and chromosomal DNA mediated by the SXT element, a constin found in Vibrio cholerae O139. J. Bacteriol. **182:**2043–2047.
67. **Honda, Y., H. Sakai, H. Hiasa, K. Tanaka, T. Komano, and M. Bagdasarian.** 1991. Functional division and reconstruction of a plasmid replication origin: molecular dissection of the oriV of the broad-host-range plasmid RSF1010. Proc. Natl. Acad. Sci. U. S. A. **88:**179–183.
68. **Honda, Y., H. Sakai, and T. Komano.** 1988. Two single-strand DNA initiation signals located in the oriV region of plasmid RSF1010. Gene **68:**221–228.
69. **Honda, Y., H. Sakai, T. Komano, and M. Bagdasarian.** 1989. RepB′ is required in trans for the two single-strand DNA initiation signals in oriV of plasmid RSF1010. Gene **80:**155–159.
70. **Horwitz, M. S., and L. A. Loeb.** 1988. An E. coli promoter that regulates transcription by DNA superhelix-induced cruciform extrusion. Science **241:**703–705.
71. Reference deleted.
72. **Hsieh, P.** 2001. Molecular mechanisms of DNA mismatch repair. Mutat. Res. **486:**71–87.
73. **Huang, X., F. J. Lopez de Saro, and J. D. Helmann.** 1997. Sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with −10 region single-stranded DNA. Nucleic Acids Res. **25:**2603–2609.
74. **Ilyina, T. V., and E. V. Koonin.** 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse

replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. **20:**3279–3285.
75. **Jaworski, A., N. P. Higgins, R. D. Wells, and W. Zacharias.** 1991. Topoisomerase mutants and physiological conditions control supercoiling and Z-DNA formation in vivo. J. Biol. Chem. **266:**2576–2581.
76. **Jones, A. L., P. T. Barth, and B. M. Wilkins.** 1992. Zygotic induction of plasmid ssb and psiB genes following conjugative transfer of Incl1 plasmid Collb-P9. Mol. Microbiol. **6:**605–613.
77. **Juhas, M., D. W. Crook, and D. W. Hood.** 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. Cell. Microbiol. **10:**2377–2386.
78. **Kaguni, J. M., and A. Kornberg.** 1982. The rho subunit of RNA polymerase holoenzyme confers specificity in priming M13 viral DNA replication. J. Biol. Chem. **257:**5437–5443.
79. **Katayama, T., S. Ozaki, K. Keyamura, and K. Fujimitsu.** 2010. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. Nat. Rev. Microbiol. **8:**163–170.
80. **Kelley, W. L.** 2006. Lex marks the spot: the virulent side of SOS and a closer look at the LexA regulon. Mol. Microbiol. **62:**1228–1238.
81. **Kersulyte, D., B. Velapatino, G. Dailide, A. K. Mukhopadhyay, Y. Ito, L. Cahuayme, A. J. Parkinson, R. H. Gilman, and D. E. Berg.** 2002. Transposable element ISHp608 of Helicobacter pylori: nonrandom geographic distribution, functional organization, and insertion specificity. J. Bacteriol. **184:**992–1002.
82. **Khan, S. A.** 2005. Plasmid rolling-circle replication: highlights of two decades of research. Plasmid **53:**126–136.
83. **Koepsel, R. R., and S. A. Khan.** 1987. Cleavage of single-stranded DNA by plasmid pT181-encoded RepC protein. Nucleic Acids Res. **15:**4085–4097.
84. **Konieczny, I.** 2003. Strategies for helicase recruitment and loading in bacteria. EMBO Rep. **4:**37–41.
85. **Kornberg, A., and T. A. Baker.** 1992. DNA replication. W. H. Freeman & Co., New York, NY.
86. **Kowalczykowski, S. C.** 1994. In vitro reconstitution of homologous recombination reactions. Experientia **50:**204–215.
87. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer.** 1994. Biochemistry of homologous recombination in Escherichia coli. Microbiol. Rev. **58:**401–465.
88. **Kowalczykowski, S. C., and R. A. Krupp.** 1987. Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. J. Mol. Biol. **193:**97–113.
89. **Kramer, M. G., M. Espinosa, T. K. Misra, and S. A. Khan.** 1999. Characterization of a single-strand origin, ssoU, required for broad host range replication of rolling-circle plasmids. Mol. Microbiol. **33:**466–475.
90. **Kramer, M. G., M. Espinosa, T. K. Misra, and S. A. Khan.** 1998. Lagging strand replication of rolling-circle plasmids: specific recognition of the ssoA-type origins in different Gram-positive bacteria. Proc. Natl. Acad. Sci. U. S. A. **95:**10505–10510.
91. **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1998. Lagging-strand replication from the ssoA origin of plasmid pMV158 in Streptococcus pneumoniae: in vivo and in vitro influences of mutations in two conserved ssoA regions. J. Bacteriol. **180:**83–89.
92. **Kramer, M. G., S. A. Khan, and M. Espinosa.** 1997. Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. EMBO J. **16:**5784–5795.
93. **Kreuzer, K. N.** 2005. Interplay between DNA replication and recombination in prokaryotes. Annu. Rev. Microbiol. **59:**43–67.
94. **Kurahashi, H., H. Inagaki, T. Ohye, H. Kogo, T. Kato, and B. S. Emanuel.** 2006. Chromosomal translocations mediated by palindromic DNA. Cell Cycle **5:**1297–1303.
95. **Lambert, P. F., D. A. Waring, R. D. Wells, and W. S. Reznikoff.** 1986. DNA requirements at the bacteriophage G4 origin of complementary-strand DNA synthesis. J. Virol. **58:**450–454.
96. **Lampson, B. C., M. Inouye, and S. Inouye.** 2005. Retrons, msDNA, and the bacterial genome. Cytogenet. Genome Res. **110:**491–499.
97. **Langston, L. D., and M. O'Donnell.** 2006. DNA replication: keep moving and don't mind the gap. Mol. Cell **23:**155–160.
98. **Larkin, C., R. J. Haft, M. J. Harley, B. Traxler, and J. F. Schildbach.** 2007. Roles of active site residues and the HUH motif of the F plasmid TraI relaxase. J. Biol. Chem. **282:**33707–33713.
99. **Leach, D. R.** 1994. Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. Bioessays **16:**893–900.
100. **Lilley, D. M.** 1985. The kinetic properties of cruciform extrusion are determined by DNA base-sequence. Nucleic Acids Res. **13:**1443–1465.
101. **Liu, L. F., and J. C. Wang.** 1987. Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. **84:**7024–7027.
102. **Liu, Y., V. Bondarenko, A. Ninfa, and V. M. Studitsky.** 2001. DNA supercoiling allows enhancer action over a large distance. Proc. Natl. Acad. Sci. U. S. A. **98:**14883–14888.
103. **Llosa, M., S. Bolland, and F. de la Cruz.** 1991. Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. Mol. Gen. Genet. **226:**473–483.
104. **Loot, C., D. Bikard, A. Rachlin, and D. Mazel.** 2010. Cellular pathways controlling integron cassette site folding. EMBO J. **29:**2623–2634.
105. **Low, K. B.** 1972. Escherichia coli K-12 F-prime factors, old and new. Bacteriol. Rev. **36:**587–607.
106. **Lu, Y. B., P. V. Ratnakar, B. K. Mohanty, and D. Bastia.** 1996. Direct physical interaction between DnaG primase and DnaB helicase of Escherichia coli is necessary for optimal synthesis of primer RNA. Proc. Natl. Acad. Sci. U. S. A. **93:**12902–12907.
107. **MacDonald, D., G. Demarre, M. Bouvier, D. Mazel, and D. N. Gopaul.** 2006. Structural basis for broad DNA specificity in integron recombination. Nature **440:**1157–1162.
108. **Majchrzak, M., R. P. Bowater, P. Staczek, and P. Parniewski.** 2006. SOS repair and DNA supercoiling influence the genetic stability of DNA triplet repeats in Escherichia coli. J. Mol. Biol. **364:**612–624.
109. **Masai, H., and K. Arai.** 1996. DnaA- and PriA-dependent primosomes: two distinct replication complexes for replication of Escherichia coli chromosome. Front. Biosci. **1:**d48–d58.
110. **Masai, H., and K. Arai.** 1997. Frpo: a novel single-stranded DNA promoter for transcription and for primer RNA synthesis of DNA replication. Cell **89:**897–907.
111. **Masai, H., N. Nomura, and K. Arai.** 1990. The ABC-primosome. A novel priming system employing dnaA, dnaB, dnaC, and primase on a hairpin containing a dnaA box sequence. J. Biol. Chem. **265:**15134–15144.
112. **Mazel, D.** 2006. Integrons: agents of bacterial evolution. Nat. Rev. Microbiol. **4:**608–620.
113. **McGlynn, P., A. A. Al-Deib, J. Liu, K. J. Marians, and R. G. Lloyd.** 1997. The DNA replication protein PriA and the recombination protein RecG bind D-loops. J. Mol. Biol. **270:**212–221.
114. **Mejean, V., and J. P. Claverys.** 1984. Use of a cloned DNA fragment to analyze the fate of donor DNA in transformation of Streptococcus pneumoniae. J. Bacteriol. **158:**1175–1178.
115. **Mendiola, M. V., I. Bernales, and F. de la Cruz.** 1994. Differential roles of the transposon termini in IS91 transposition. Proc. Natl. Acad. Sci. U. S. A. **91:**1922–1926.
116. **Miao, D. M., Y. Honda, K. Tanaka, A. Higashi, T. Nakamura, Y. Taguchi, H. Sakai, T. Komano, and M. Bagdasarian.** 1993. A base-paired hairpin structure essential for the functional priming signal for DNA replication of the broad host range plasmid RSF1010. Nucleic Acids Res. **21:**4900–4903.
117. **Monzingo, A. F., A. Ozburn, S. Xia, R. J. Meyer, and J. D. Robertus.** 2007. The structure of the minimal relaxase domain of MobA at 2.1 A resolution. J. Mol. Biol. **366:**165–178.
118. **Mortier-Barriere, I., M. Velten, P. Dupaigne, N. Mirouze, O. Pietrement, S. McGovern, G. Fichant, B. Martin, P. Noirot, E. Le Cam, P. Polard, and J. P. Claverys.** 2007. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell **130:**824–836.
119. **Mott, M. L., and J. M. Berger.** 2007. DNA replication initiation: mechanisms and regulation in bacteria. Nat. Rev. Microbiol. **5:**343–354.
120. **Murchie, A. I., and D. M. Lilley.** 1987. The mechanism of cruciform formation in supercoiled DNA: initial opening of central basepairs in salt-dependent extrusion. Nucleic Acids Res. **15:**9641–9654.
121. **Nasim, M. T., I. C. Eperon, B. M. Wilkins, and W. J. Brammar.** 2004. The activity of a single-stranded promoter of plasmid ColIb-P9 depends on its secondary structure. Mol. Microbiol. **53:**405–417.
122. **Nelson, T. C.** 1951. Kinetics of genetic recombination in Escherichia coli. Genetics **36:**162–175.
123. **Noirot, P., J. Bargonetti, and R. P. Novick.** 1990. Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. Proc. Natl. Acad. Sci. U. S. A. **87:**8560–8564.
124. **Nomura, N., H. Masai, M. Inuzuka, C. Miyazaki, E. Ohtsubo, T. Itoh, S. Sasamoto, M. Matsui, R. Ishizaki, and K. Arai.** 1991. Identification of eleven single-strand initiation sequences (ssi) for priming of DNA replication in the F, R6K, R100 and ColE2 plasmids. Gene **108:**15–22.
125. **Oussatcheva, E. A., J. Pavlicek, O. F. Sankey, R. R. Sinden, Y. L. Lyubchenko, and V. N. Potaman.** 2004. Influence of global DNA topology on cruciform formation in supercoiled DNA. J. Mol. Biol. **338:**735–743.
126. **Pages, V., and R. P. Fuchs.** 2003. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo. Science **300:**1300–1303.
127. **Panayotatos, N., and R. D. Wells.** 1981. Cruciform structures in supercoiled DNA. Nature **289:**466–470.
128. **Pearson, C. E., H. Zorbas, G. B. Price, and M. Zannis-Hadjopoulos.** 1996. Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. J. Cell. Biochem. **63:**1–22.
129. **Platt, J. R.** 1955. Possible separation of intertwined nucleic acid chains by transfer-twist. Proc. Natl. Acad. Sci. U. S. A. **41:**181–183.
130. **Pruss, G. J., and K. Drlica.** 1989. DNA supercoiling and prokaryotic transcription. Cell **56:**521–523.
131. **Recchia, G. D., and R. M. Hall.** 1995. Gene cassettes: a new class of mobile element. Microbiology **141:**3015–3027.

132. **Reddy, M. S., M. B. Vaze, K. Madhusudan, and K. Muniyappa.** 2000. Binding of SSB and RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA polymerization on single-stranded DNA. Biochemistry **39:**14250–14262.
133. **Ronning, D. R., C. Guynet, B. Ton-Hoang, Z. N. Perez, R. Ghirlando, M. Chandler, and F. Dyda.** 2005. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. Mol. Cell **20:**143–154.
134. **Roy, R., A. G. Kozlov, T. M. Lohman, and T. Ha.** 2009. SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. Nature **461:**1092–1097.
135. **Russel, M., N. A. Linderoth, and A. Sali.** 1997. Filamentous phage assembly: variation on a protein export theme. Gene **192:**23–32.
136. **Russel, M., and P. Model.** 1989. Genetic analysis of the filamentous bacteriophage packaging signal and of the proteins that interact with it. J. Virol. **63:**3284–3295.
137. **Schleif, R.** 2000. Regulation of the L-arabinose operon of Escherichia coli. Trends Genet. **16:**559–565.
138. **Schvartzman, J. B., and A. Stasiak.** 2004. A topological view of the replicon. EMBO Rep. **5:**256–261.
139. **Shlyakhtenko, L. S., P. Hsieh, M. Grigoriev, V. N. Potaman, R. R. Sinden, and Y. L. Lyubchenko.** 2000. A cruciform structural transition provides a molecular switch for chromosome structure and dynamics. J. Mol. Biol. **296:**1169–1173.
140. **Sinden, R. R., S. S. Broyles, and D. E. Pettijohn.** 1983. Perfect palindromic lac operator DNA sequence exists as a stable cruciform structure in supercoiled DNA in vitro but not in vivo. Proc. Natl. Acad. Sci. U. S. A. **80:** 1797–1801.
141. **Sinden, R. R., G. X. Zheng, R. G. Brankamp, and K. N. Allen.** 1991. On the deletion of inverted repeated DNA in Escherichia coli: effects of length, thermal stability, and cruciform formation in vivo. Genetics **129:**991–1005.
142. **Singh, J., M. Mukerji, and S. Mahadevan.** 1995. Transcriptional activation of the Escherichia coli bgl operon: negative regulation by DNA structural elements near the promoter. Mol. Microbiol. **17:**1085–1092.
143. **Sullivan, K. M., and D. M. Lilley.** 1987. Influence of cation size and charge on the extrusion of a salt-dependent cruciform. J. Mol. Biol. **193:**397–404.
144. **Sun, W., and G. N. Godson.** 1998. Structure of the Escherichia coli primase/single-strand DNA-binding protein/phage G4oric complex required for primer RNA synthesis. J. Mol. Biol. **276:**689–703.
145. **Swart, J. R., and M. A. Griep.** 1993. Primase from Escherichia coli primes single-stranded templates in the absence of single-stranded DNA-binding protein or other auxiliary proteins. Template sequence requirements based on the bacteriophage G4 complementary strand origin and Okazaki fragment. J. Biol. Chem. **268:**12970–12976.
146. **Tanaka, T., T. Mizukoshi, K. Sasaki, D. Kohda, and H. Masai.** 2007. Escherichia coli PriA protein, two modes of DNA binding and activation of ATP hydrolysis. J. Biol. Chem. **282:**19917–19927.
147. **Tatum, E. L., and J. Lederberg.** 1947. Gene recombination in the bacterium Escherichia coli. J. Bacteriol. **53:**673–684.
148. **Ton-Hoang, B., C. Pasternak, P. Siguier, C. Guynet, A. B. Hickman, F. Dyda, S. Sommer, and M. Chandler.** 2010. Single-stranded DNA transposition is coupled to host replication. Cell **142:**398–408.
149. **Trinh, T. Q., and R. R. Sinden.** 1991. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli. Nature **352:**544–547.
150. **Val, M.-E., M. Bouvier, J. Campos, D. Sherratt, F. Cornet, D. Mazel, and F.-X. Barre.** 2005. The single-stranded genome of phage CTX is the form used for integration into the genome of Vibrio cholerae. Mol. Cell **19:**559–566.
151. Reference deleted.
152. **Walker, G. C.** 1996. The SOS response of Escherichia coli, p. 1400–1416. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 1. ASM Press, Washington, DC.
153. **Wang, J. C., and A. S. Lynch.** 1993. Transcription and DNA supercoiling. Curr. Opin. Genet. Dev. **3:**764–768.
154. **White, J. H., and W. R. Bauer.** 1987. Superhelical DNA with local substructures. A generalization of the topological constraint in terms of the intersection number and the ladder-like correspondence surface. J. Mol. Biol. **195:**205–213.
155. **Wickner, S., and J. Hurwitz.** 1975. Association of phi X174 DNA-dependent ATPase activity with an Escherichia coli protein, replication factor Y, required for in vitro synthesis of phi X174 DNA. Proc. Natl. Acad. Sci. U. S. A. **72:**3342–3346.
156. **Wozniak, R. A., D. E. Fouts, M. Spagnoletti, M. M. Colombo, D. Ceccarelli, G. Garriss, C. Dery, V. Burrus, and M. K. Waldor.** 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. PLoS Genet. **5:**e1000786.
157. **Zannis-Hadjopoulos, M., W. Yahyaoui, and M. Callejo.** 2008. 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. Trends Biochem. Sci. **33:**44–50.
158. **Zhao, J., A. Bacolla, G. Wang, and K. M. Vasquez.** 2010. Non-B DNA structure-induced genetic instability and evolution. Cell. Mol. Life Sci. **67:**43–62.
159. **Zheng, G., D. W. Ussery, and R. R. Sinden.** 1991. Estimation of superhelical density in vivo from analysis of the level of cruciforms existing in living cells. J. Mol. Biol. **221:**122–129.
160. **Zheng, G. X., T. Kochel, R. W. Hoepfner, S. E. Timmons, and R. R. Sinden.** 1991. Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells. J. Mol. Biol. **221:**107–122.
161. **Zheng, G. X., and R. R. Sinden.** 1988. Effect of base composition at the center of inverted repeated DNA sequences on cruciform transitions in DNA. J. Biol. Chem. **263:**5356–5361.