Published in final edited form as:

Nat Rev Microbiol. 2013 August ; 11(8): 525–538. doi:10.1038/nrmicro3067.

Breaking and joining single-stranded DNA: the HUH endonuclease superfamily

Michael Chandler[1], **Fernando de la Cruz**[2], **Fred Dyda**[3], **Alison B. Hickman**[3], **Gabriel Moncalian**[2], and **Bao Ton-Hoang**[1]

[1]Laboratoire de Microbiologie et Génétique Moléculaires, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5100, 118 Route de Narbonne, F31062 Toulouse Cedex, France

[2]Departamento de Biología Molecular and Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-Consejo Superior de Investigaciones Científicas-SODERCAN, 22 Calle Albert Einstein, 39011 Santander, Spain

[3]Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 5, 5 Center Drive, Bethesda, Maryland 20892, USA

Abstract

HUH endonucleases are numerous and widespread in all three domains of life. The major function of these enzymes is processing a range of mobile genetic elements by catalysing cleavage and rejoining of single-stranded DNA using an active-site Tyr residue to make a transient 5′-phosphotyrosine bond with the DNA substrate. These enzymes have a key role in rolling-circle replication of plasmids and bacteriophages, in plasmid transfer, in the replication of several eukaryotic viruses and in various types of transposition. They have also been appropriated for cellular processes such as intron homing and the processing of bacterial repeated extragenic palindromes. Here, we provide an overview of these fascinating enzymes and their functions, using well-characterized examples of Rep proteins, relaxases and transposases, and we explore the molecular mechanisms used in their diverse activities.

Although the double helix is the most iconic feature of DNA, the molecule must assume a transient single-stranded form during replication, transcription, bacterial conjugation and a variety of repair and recombination processes. Single-stranded DNA (ssDNA) also serves as the packaged genetic material in some viruses and bacteriophages. It is therefore no surprise that there is a dedicated endonuclease superfamily, the HUH endonucleases (in which U is a

Correspondence to M.C.: Michael.Chandler@ibcg.biotoul.fr.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

ISfinder: https://www-is.biotoul.fr

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

hydrophobic residue), the members of which exclusively process ssDNA using particular recognition and reaction mechanisms for site-specific ssDNA cleavage and ligation.

HUH endonucleases are numerous and widespread in all three domains of life. Two major classes within this superfamily are the Rep (replication) and relaxase, or Mob (mobilization), proteins involved in DNA processing during plasmid replication and conjugation, respectively. However, HUH endonucleases have also been identified in other processes involving DNA, such as replication of certain phages and eukaryotic viruses, and different types of transposition. These proteins have also been appropriated for cellular processes such as intron homing and processing of bacterial repetitive extragenic palindromic sequences (REP sequences). Here, we provide an overview of these fascinating enzymes, their numerous roles and functions, and the mechanisms used in their diverse activities. We explore their functional diversity using well-characterized examples of Rep proteins, relaxases and transposases.

Mechanism and overall protein architecture

The first member of the HUH superfamily, protein A (gpA) of phage φX174, was identified in early studies of phage replication[1]. Subsequent bioinformatic approaches[2,3] revealed many related proteins forming a vast superfamily, which includes members involved in the catalysis of viral and plasmid rolling-circle replication (RCR) (Rep proteins), in conjugative plasmid transfer (relaxases) and in DNA transposition (transposases)[4–9]. This familial relationship is based on several conserved protein motifs, including the HUH motif, composed of two His residues separated by a bulky hydrophobic residue, and the Y motif, containing either one or two Tyr residues separated by several amino acids. The initial classification was based largely on Rep proteins and included only a few relaxases, leading to some misclassfications in the latter group[2,3]. Indeed, the initial Y motif identified for the R388 and F plasmid relaxases was incorrect, presumably owing to the limited number of examples available at the time[9]. For convenience, here we define enzymes with only a single conserved catalytic Tyr as Y1 and those with more than one catalytic Tyr as Y2, even though some Y2 HUH endonucleases require only one of their Y motif Tyr residues for catalysis, whereas others require both.

Y1 HUH endonucleases include Rep proteins of some plasmids that replicate using ssDNA intermediates (such as pUB110 (REF 10)), Rep proteins of a wide range of eukaryotic viruses[11], most relaxases, transposases of insertion sequences with a common region (ISCRs; which are insertion sequences related to the IS*91* family)[8] and transposases of the IS*200*-IS*605* family[6,7]. Y2 enzymes[12] include φX174 gpA and the Rep proteins of other isometric-headed ssDNA and double-stranded DNA (dsDNA) phages (such as P2 (REF 13)), those of some cyanobacterial and archaeal plasmids, and those of parvoviruses (for example, adeno-associated virus (AAV)), as well as transposases ofthe IS*91* and Helitron families[4], and relaxases of the MobF family.

HUH endonucleases catalyse ssDNA breakage and joining with a unique mechanism. They use a Y motif Tyr to create a 5′-phosphotyrosine intermediate and a free 3′-OH at the cleavage site (FIG. 1a). Subsequently, the 3′-OH can be used for different tasks. The most

obvious task is to prime replication, as observed for the HUH domains in Rep proteins of single-stranded phages, RCR plasmids and conjugative relaxases. The 3′-OH group can also act as the nucleophile for strand transfer to resolve the phosphotyrosine intermediate in the termination step of RCR replication, conjugative transfer and transposition (see below).

The cleavage polarity of HUH endonucleases is inverse to that of Tyr recombinases, which make 3′-phosphoty-rosine intermediates and generate free 5′-OH groups that cannot be used as replication primers[14]. A further contrast to the cofactor-independent Tyr recombinases is that HUH enzyme activities require a divalent metal ion to facilitate cleavage by localizing and polarizing the scissile phosphodiester bond. Depending on the enzyme, Mg^{2+}, Mn^{2+} or other divalent metal ions can be used *in vitro*[15–20], although it is often presumed that Mg^{2+} or Mn^{2+} are the physiological cofactors. The His pair of the HUH motif provides two of the three ligands necessary for metal ion coordination (FIG. 1a). The location and identity of the third, invariably polar residue, which can be Glu, Asp, His or (as shown in FIG. 1a) Gln, varies across the superfamily.

Three-dimensional structures of several Rep and relaxase HUH domains have been determined with and without bound DNA (Supplementary information S1 (table) for Protein Data Bank accessions) and have been crucial to understanding the mechanisms of these enzymes. The common element of the HUH fold (FIG. 1b) is a central four- or five-stranded antiparallel β-sheet generally sandwiched between α-helices. The HUH motif is invariably located on a single β-strand, whereas the catalytic Tyr residue (or residues) is located on a nearby α-helix (FIG. 1b). The order of HUH and Y motifs varies in the primary sequence: in the relaxase group, the Y motif is upstream of the HUH motif, whereas in the Rep group it is downstream (FIG. 1b,c). This circular permutation[3,21,22] of the motifs in the primary sequence (FIG. 1c) changes the topology of the domains (FIG. 1b). Nevertheless, the three-dimensional constellation of active-site residues is virtually identical across the superfamily.

Given the diverse functions of HUH proteins, it is not surprising that other domains are often appended to the HUH domain. In many cases, these additional domains are of unknown function. However, ATP-dependent helicase, Zn-binding, primase and multimerization domains are recurring themes[4,23–28] (FIG. 1c). For example, the ssDNA substrates needed by HUH endonucleases can be generated by a dedicated DNA helicase domain carboxy-terminal to the HUH domairi[4,24,29,30] or, alternatively, by recruitment of a host helicase[25–28]. RCR uses 3′−5′ helicase activity acting on the template strand to facilitate DNA unwinding at the replication fork, whereas during conjugation, helicases (as part of the relaxase) are transported into the recipient cell and track 5′−3′ on the transported ssDNA.

DNA recognition

A feature of many HUH endonucleases is that they can recognize and bind DNA hairpin structures, and the DNA cleavage sites are located either in the hairpin or in the ssDNA located on the 5′ or 3′ side of the stem. The crucial role of hairpins has been firmly established for many processes, including plasmid conjugation, replication of plasmids and of eukaryotic viruses, and transposition[6,7,16,31–34]. In other systems, palindromic sequences that can form DNA hairpins are present near the probable HUH endonuclease cleavage

sites[35]. Such hairpins can be formed *in vivo* under a number of physiological conditions (see REF. 36).

Structural studies of HUH endonucleases revealed that small DNA hairpins can be recognized in several different ways (FIG. 2): sequence-specific recognition of the dsDNA stem, structure-specific recognition of irregularities in the stem, or sequence-specific recognition of the hairpin loop[7,15,20,22,23,33].

The hairpin-flanking DNA — in many cases, in single-stranded form — is also often important for recognition. Relaxases, for example, make extensive contacts with the bases extending between the hairpin and the cleavage site[20,22,37] (FIG. 2), and for TnpA(REP), a protein related to the IS*608* transposase (TnpA), it has been shown that nucleotides on the 5′ side of the hairpin are crucial for binding and for sequence-specific recognition[33]. Other family members[23,38] have more complex binding modes (see below).

Below, we explore the functional diversity of HUH endonucleases using well-characterized examples of various Rep proteins, relaxases and transposases.

Rep proteins and RCR

RCR (FIG. 3) was first proposed more than 40 years ago[39] as a mechanism for the replication of single-stranded phages. This process is now also well-established as a mechanism adopted by certain eukaryotic viruses and bacterial plasmids (reviewed in REFS 1,35,40).

At the origin: phage φX174.

Studies on φX174 gpA revealed several functional details that are recapitulated in other HUH domain-containing systems; thus, φX174 provides a general framework for understanding DNA processing involving HUH endonucleases.

When φX174 infects its *Escherichia coli* host, its circular positive-sense ssDNA ((+)ssDNA) genome is injected into the cell and converted by host enzymes into a duplex form resembling a supercoiled dsDNA plasmid. RCR is initiated when the Rep protein gpA recognizes the 3′ region of a short (approximately 30 bp) sequence at the origin of replication (FIG. 3a), unwinds DNA upstream and nicks the (+)strand at a specific site (called the *nic* site) upstream of the recognition sequence to form a covalent 5′-phosphotyrosine intermediate and a free 3′-OH[41–43] (FIG. 3b). Although the DNA recognition sequence and cleavage site are known, there is surprisingly no published detailed analysis of how gpA recognizes and binds the target DNA. Replication is initiated using the 3′-OH end, and the (+)strand is ‘peeled off’ (FIG. 3c,d). As replication proceeds, the recognition and *nic* sites are regenerated. After one complete cycle of replication round the circular DNA template, the replication apparatus and gpA reach the reconstituted recognition and *nic* site, and gpA recognizes this regenerated sequence (FIG. 3e) and nicks the (+)strand using the second Tyr in the catalytic Y motif[44], which becomes attached to the newly replicated DNA (FIG. 3f). The newly generated 3′-OH (at the end of the original, displaced ssDNA copy) then attacks the first phosphotyrosine linkage, generating a free

ssDNA circle (FIG. 3f). Alternating the use of the active-site Tyr residues allows gpA to spin off many ssDNA circles[45].

This process led to a model in which the Tyr residues, located on the same face of an α-helix, can be positioned alternately to attack the scissile phosphate that is localized and polarized by the HUH-bound metal ion[43,45]. Alternating the use of Tyr residues probably necessitates conformational changes, which might also be a crucial feature of the catalytic cycle of other HUH endonucleases (for example, see REFS 13,46).

RCR plasmids.

Plasmid RCR closely mirrors that of phages. However, the lifestyle of viruses benefits from the synthesis of multiple genomic copies, whereas that of bacterial plasmids requires plasmid replication to be closely coordinated with host cell growth. Plasmids must therefore possess mechanisms to prevent unregulated premature reinitiation of replication.

Several models have been proposed to explain how this is achieved. For the monomeric Y1 Rep protein of pC194, the final step in replication can be accomplished by an essential active-site Glu that replaces the second Tyr. This Glu allows cleavage of the phosphotyrosine bond by a bound water molecule (FIG. 3f) and release of the enzyme. Importantly, a mutant with a second Tyr instead of a Glu, emulating the phage Rep protein sequence, generates multimeric ssDNA plasmid copies[47]. An alternative so-called suicide mechanism was proposed for other plasmid Y1 RCR systems that use a dimeric Rep, in which covalent modification of one Rep monomer generates an inactive Rep (containing one active monomer and one inactive monomer) on termination[48].

The only published example of a crystallographic structure for a Rep protein from an RCR plasmid is that of RepB from pMV158 (REF. 17). This protein has two closely spaced Y motif Tyr residues and might be considered a Y2 enzyme, although only one of the Tyr residues (Tyr99) is required for catalysis[49] and the other is not correctly placed for catalysis[17]. In addition to an amino-terminal HUH domain, RepB has a small C-terminal oligomerization domain that assembles six endonuclease domains into a ring with a small central channel, potentially providing multiple second Tyr residues for catalysis. RepB DNA recognition is complex, with two binding sites, the distal and proximal direct repeats, downstream of the hairpin that contains the cleavage site[38].

Viral Rep proteins.

Replication of eukaryotic parvoviruses provides an intriguing variation of the RCR cycle[50]. Parvoviruses such as AAV have linear ssDNA genomes and replicate by a ‘rolling hairpin’ mechanism (FIG. 4). Crucial for this are the final ~160 bases on each end of the viral ssDNA genome, which can fold into fully complementary three-way DNA junctions known as inverted terminal repeats (ITRs) (represented two-dimensionally as a T-shaped folded structure; FIG. 4a insert). As a result, replication of the 5′ end is straightforward: the 3′ ITR provides a free 3′-OH group from which leading-strand synthesis can be initiated using host cell enzymes (FIG. 4a-c). However, replication of this 3′ ITR requires a site-specific nick to generate a second 3′-OH at the terminal resolution site (*trs*) in the 3′ ITR. Cleavage is catalysed by the viral Rep (FIG. 4d,e), which possesses an N-terminal HUH domain and a

C-terminal superfamily 3 (SF3) hexameric 3′−5′ helicase domain separated by a functionally important linker region[51–54] (FIG. 1c). When the nick is introduced at the 3′ ITR, the cellular replication apparatus completes viral genome replication (FIG. 4f).

Structural characterization of the N-terminal HUH domain in AAV5 Rep revealed that several AAV5 Rep monomers recognize and bind tandem tetranucleotide repeats in the Rep-binding site (RBS) located on one branch of the 3′ ITR (FIG. 4 insert), to form a spiral curling around the RBS site[23,55]. In this case, the HUH endonuclease domain recognizes dsDNA via major- and minor-groove interactions and contacts two adjacent tetranucleotide RBS repeats. The tip of one hairpin branch (called the Rep-binding element (RBE′)) is also recognized by AAV Rep, although via a different surface of the HUH domain[23] (FIG. 2a), and this interaction is important for AAV replication[31]. Rep helicase activity is thought to generate ssDNA for *trs* hairpin formation[29] (FIG. 4e). After *trs* cleavage, AAV Rep remains covalently attached to the viral DNA through its active-site Tyr. In addition, the helicase domain directs packaging of the newly replicated genomes into capsids[56]. Although AAV Rep is a Y2 endonuclease, there is no evidence that the second Y motif is required for replication.

The HUH endonuclease domains have been structurally characterized for several other Rep proteins of eukaryotic viruses with circular ssDNA genomes. Replication of these genomes presumably proceeds using an RCR mechanism similar to that used by phages and plasmids. These Rep proteins include those of tomato yellow leaf curl virus (a geminivirus)[57], the nanovirus faba bean necrotic yellows virus and porcine circovirus[58,59]. The structure of the dimeric Rep protein from the archaeal Rudivirus *Sulfolobus islandicus* rod-shaped virus 1 has also been determined, and the biochemical activity of the enzyme has been characterized. The genome of this virus is linear and does not replicate by RCR, but replication could share some mechanistic aspects with parvovirus replication[60]. All of these viral Rep proteins have only one conserved Tyr at their active sites, raising the possibility that a second is provided by another protomer or that an alternative second nucleophile (such as water) is used, as has been proposed for copy number control in RCR plasmids (see above)[47,61–63].

Plasmid relaxases

Plasmid conjugation (FIG. 5), which was discovered more than 60 years ago[64], involves transfer of an ssDNA plasmid copy from one cell to another through a pore formed by a specialized type IV secretion system (T4SS)[65]. Conjugation is the process primarily responsible for horizontal gene transfer, for example of antibiotic resistance genes between bacterial species. Recently, a large family of non-autonomous elements called conjugative transposons or integrative conjugative elements (ICEs) has been identified. These elements are extremely wide- spread[66]. They are integrated into the host genome, excised as circular intermediates and transferred by conjugation. Depending on the ICE family, integration and excision can be catalysed either by a phage-like Tyr recombinase[67] or by a DDE family transposase[68,69]. Like conjugative plasmids, ICEs encode relaxases with both a cluster that potentially binds divalent metal ions (but not necessarily HUH ligands) and a conserved Tyr, identified first in Tn*916* (REF 70) and subsequently in many examples from clinically

important Grampositive bacteria (see REF 9). Moreover, certain ICEs have also been shown to undergo replication as part of their life cycle[68,71].

General reaction mechanism.

The biochemical signal for initiating conjugative DNA processing is unknown but is thought to be triggered by donor cell-recipient cell interaction. Conjugation begins with nicking of the circular dsDNA plasmid at the *nic* site within the plasmid origin of transfer (*oriT*); this nicking is mediated by the HUH endonuclease domain of the plasmid-encoded relaxase (FIG. 5a). This initiates 'conjugative RCR', which peels off a specific strand of the dsDNA plasmid and transfers this ssDNA to the recipient cell (FIG. 5b,c). Relaxase remains covalently attached to this ssDNA and is transported into the recipient cell, where it tracks 5′–3′ along the ssDNA while still attached to the end[72]. Relaxase catalyses recircularization of the transferred ssDNA plasmid in the recipient cell (FIG. 5d-f), where complementary-strand synthesis by a host cell polymerase converts the plasmid into its dsDNA form. Donor plasmid replication (using the non-transferred, circular strand as a template) can occur in the donor cell concomitantly with transfer, but is not essential for the transfer process[73]. Replication termination and the generation of a circular plasmid copy can occur in several ways, depending on whether the second Tyr is carried in the same relaxase molecule or in a different one, and whether the termination occurs in the donor or recipient cell (see below).

Catalytic centre.

Relaxases (FIG. 1c) have been classified into six families on the basis of their N-terminal HUH domains[9]. Four of these families, MobF, MobQ, MobP and MobV, generally include the HUH motif and a third residue for metal coordination (usually another His) along with the conserved Tyr residue (or residues), whereas the two other families, MobH and MobC, have — a different architecture. Thus, as with plasmid RCR (in which not all RCR plasmids use HUH-containing Rep proteins), conjugative plasmid transfer can have variations in the way that strand cleavage and strand transfer occur. Of the four HUH relaxase families, MobF comprises Y2 enzymes, whereas MobQ, MobP and MobV are all Y1 relaxase families. To date, crystallographic structures of the endonuclease domains of three MobF enzymes[18,22,74] and two MobQ proteins[20,75] have been determined. In addition to the HUH and Y motifs, a third motif with a conserved Asp or Glu might enhance the interaction of the His triad with the divalent cation[55,76]. There seem to be mechanistic differences even between members of one relaxase family. For example, in the MobF family member TrwC (encoded by plasmid R388), Tyr18 is necessary for the cleavage of supercoiled plasmid DNA *in vitro* and is absolutely required, together with Tyr26, for complete relaxase activity *in vivo*. This suggests that Tyr18 is involved in the initial cleavage reaction and that Tyr26 is responsible for the final strand transfer reaction[77]. By contrast, of the four candidate catalytic Tyr residues in TraI (encoded by plasmid F), only Tyr16 is essential for relaxase activity, suggesting that two TraI molecules are necessary for plasmid F transfer[78].

DNA binding.

HUH relaxases are thought to initially recognize the dsDNA form of an inverted repeat close to the *nic* site in the supercoiled plasmid[79,80] and, with the help of accessory proteins[65],

locally melt the DNA around the *nic* site (FIG. 5a insert). In the recipient cell, both arms of the inverted repeat in the transferred ssDNA will form a hairpin that is again recognized by the relaxase (FIG. 5d insert).

In the TrwC-DNA co-crystal structure[22] (FIG. 2b), the hairpin is firmly embraced by the protein through a β-ribbon conserved in all three available MobF structures, whereas in the MobQ-DNA co-crystal of DNA bound by the *Staphylococcus aureus* MobQ protein NES (FIG. 2c), the hairpin is bound by two loops[20]. Three C-terminal α-helices form the MobF relaxase 'fingers', which position the ssDNA region containing the *nic* site (immediately next to the hairpin) in the relaxase active site. These fingers are absent in MobQ proteins, and the *nic* site is instead positioned by burying an adjacent nucleotide in the protein. Conserved Arg and Ser residues have also been implicated in the stabilization of DNA-relaxase interactions and, in the case of Ser, positioning the ssDNA substrate in the active site[22,81].

For MobF relaxases, which are all Y2 proteins, one Tyr is involved in RCR initiation in the donor cell (FIG. 5a), whereas another Tyr in the same relaxase molecule catalyses termination in the recipient cell (left panels in FIG. 5d,e)[77]. The regenerated *nic* site in the transferred ssDNA is recognized by the initiating relaxase, which remains covalently linked to the 5′ ssDNA end. In all MobF relaxases, the C-terminal half contains a RecD-like translocase domain, which could track along the incoming ssDNA and reload the regenerated *nic* site onto the relaxase catalytic centre in the recipient cell (FIG. 2b). Then, a second *nic* site cleavage circularizes the transferred ssDNA but also joins the relaxase to the newly synthesized *nic* site (left panels in FIG. 5d-f). This, in principle, might lead to transfer of a second plasmid copy. It is uncertain how relaxase is removed from the ssDNA to terminate the reaction. A second model proposes that the second cleavage occurs in the donor cell and is mediated by a second relaxase, so the termination reaction will simply be the reversal of the cleavage reaction[78] (right panels in FIG. 5d-f).

Relaxases from the MobQ, MobP and MobV families, which all contain Y1 and HUH domains, have variable C-terminal domains[82,83], such as primase, helicase or DNA-binding domains. For plasmid RSF1010, the primase domain of the relaxase MobA (FIG. 1c) is required for initiation of replication in the recipient cell, but the same primase can also be produced separately without the relaxase domain to form the protein RepB′, which is also essential for replication of the plasmid[84].

HUH endonucleases as transposases

HUH endonucleases also act as transposases for members of the bacterial and archaeal IS*200*-IS*605* (REF. 6), IS*91* (REF. 85) and ISCR[8] insertion sequence families and for the eukaryotic Helitrons[4] (BOX 1). The best understood HUH transposases are those of the IS*200*-IS*605* family.

*IS*200-*IS*605 *transposons*.

The IS*200*-IS*605* transposon family includes two subgroups: IS*200* and IS*605*. IS*200* group members carry only the transposase gene (*tnpA*). Those from the IS*605* group carry both

tnpA and a second ORF, *tnpB* (FIG. 6a), the function of which is unclear but is not required for transposition. The first family member, IS*200*, was identified about 30 years ago[86] in *Salmonella* spp. However, the family paradigms are the more recently identified IS*608* (in *Helicobacter pylori*[87]) and *ISDra2* (in *Deinococcus radiodurans*[88]). These insertion sequences use obligatory ssDNA intermediates for mobility. The transposons excise as an ssDNA circle with abutted left and right ends, and insert 3′ to a conserved, element-specific penta- or tetra- nucleotide target. Although this target sequence is not part of the IS, it is essential for further transposition. The transposons do not carry terminal inverted repeats but include hairpins in the left and right ends that are recognized and bound by TnpA (FIG. 6a).

The transposons excise from, and insert preferentially into, the lagging-strand template at chromosome and plasmid replication forks (FIG. 6b). This preference for the lagging strand generates an insertion bias that reflects the mode of replication (uni- or bidirectional) of the target replicon[89]. Transposition of IS*Dra2* is strongly induced following recovery of the highly radiation-resistant host from irradiation[90], and this induction is certainly due to the large amount of ssDNA that is generated during reassembly of the shattered *D. radiodurans* genome[91].

IS*200*-IS*605* family transposases are single-domain proteins containing only the essential HUH motif and a single catalytic Tyr (they are Y1 transposases) (FIG. 1c). Both IS*608* TnpA and IS*Dra2* TnpA are obligatory dimers, and the active sites of these enzymes are believed to adopt two functionally important conformations: the *trans* configuration, in which each active site is composed of the HUH motif from one monomer and the Tyr residue on an α-helix (αD) from the other monomer, and the *cis* configuration, in which both motifs in an active site are contributed by the same monomer. Only the *trans* configuration has been observed crystallographically.

Like relaxases, TnpA also binds hairpins and cleaves ssDNA at some distance. The left-end and right-end cleavage sites (which differ in sequence) are recognized not directly by the protein, but instead by particular base interactions with short ‘guide’ sequences 5′ to each hairpin. Recognition of the cleavage site requires a network of base interactions[15,46,92] that also stabilizes the nucleo- protein complex — the transpososome — within which the DNA strand cleavages and transfers are carried out[92]. Changes in the guide sequences generate predictable changes in insertion site specificity[93].

Excision and integration probably occur by alternating the catalytic site conformation between the *trans* and *cis* configurations (FIG. 6c). On cleavage of both cleavage sites by the *trans* dimer, the left end of the transposon and the chromosomal sequence immediately downstream of the right end are each attached by their 5′-phosphate to the attacking Tyr residue, leaving a 3′-OH on the break upstream of the left end and on the right end. Strand joining probably occurs by reciprocal rotation of the two Tyr-carrying a-helices, to move into the *cis* configuration, and subsequent attack of the phosphotyrosine bonds by the 3′-OH groups. This joins the left and right ends together, and also joins the two transposon-flanking sequences. The αD helices must be reset to the *trans* configuration to integrate the circular transposon elsewhere in the chromosome[92].

RCR transposons: the IS91, ISCR and Helitron families.

The earliest identified transposable elements with HUH domain transposases were members of the IS*91* family[5]. The IS*91* transposases are significantly larger than Y1 transposases (FIG. 1c), carry a Y2 motif and include an N-terminal Zn-binding motif and additional domains of as-yet-unidentified function.

Recently, a group of related elements, the ISCRs, has been described (see REF 8). ISCRs have a common region consisting of an ORF resembling the IS*91* family transposase gene but with only a single catalytic Tyr residue (FIG. 1c). ISCRs are often associated with a range of antibiotic resistance genes. In addition, eukaryotic relatives of these elements, the Helitrons, have been identified by bioinformatic approaches (BOX 1).

There is little information concerning either ISCR or Helitron transposition, but IS*91* is thought to transpose by an RCR mechanism related to RCR plasmid replication. Both Tyr residues are essential for IS*91* transposition *in vivo.* Although IS*91* ends contain two short flanking inverted repeats, it is not clear whether this is a general feature of the family. Like IS*200*-IS*605* family members, IS*91* inserts with the right inverted repeat (IR_R; also called ori-IS) 3′ to a specific tetra- nucleotide that is also required for further transposition (see REFS 94,95). Transposition is postulated to initiate at IR_r as a result of transposase-mediated cleavage, and the 3′-OH generated in the donor molecule would act as a primer for DNA replication. Transfer of the donor sequence to a target is driven by replication in the donor, during which displacement of the active transposon strand would be driven by leading-strand replication. Transfer of an ssDNA IS*91* copy into the target DNA is accomplished when IR_L (also called ter-IS) is reached and cleaved (reviewed in REF 5). Termination fails at a high frequency (1% of all insertions), and the transfer process in these cases is known as one-ended transposition, as it results in the insertion of additional flanking donor-plasmid genes 3′ to the IR_L sequence[85]. Although IR_r is essential for insertion sequence activity, IR_L is not: removal of this sequence simply increases the relative frequency of one-ended transposition. Acquisition of flanking antibiotic resistance genes by ISCRs could be a manifestation of this property.

Despite this attractive model, *in vivo* studies have identified both ssDNA and dsDNA transposon circles containing covalently joined termini without the IR_R- flanking tetranucleotide. Circle formation depends on an intact transposase Y2 motif[95], and these circles might therefore be transposition intermediates analogous to the IS*608* ssDNA circular transposition intermediates. The dsDNA circles could result from replication restart or from extension of a trapped Okazaki fragment on the excised ssDNA circle.

The relationship between transposition of IS*91* and replication of the donor and target replicons is not known.

Domestication, diversity and versatility

Domestication.

A number of studies, particularly of eukaryotes, have provided many examples of transposable elements evolving to assume alternative roles within their host[96]. There are few

bacterial examples of such ‘domestication’, but two of these examples involve HUH endonucleases closely related to IS*200*-IS*605* family transposases (BOX 2). These examples are TnpA(REP) proteins, which are TnpA-like transposases that can cleave and rejoin REP sequences *in vitro* and are thought to be responsible for the dissemination of these sequences within bacterial genomes (BOX 2); and the TnpA-like endonucleases that are encoded in some group I introns (called IStrons), which presumably play the part of homing endonucleases (BOX 2).

Diversity in the catalytic centre.

As stated above, HUH endonucleases require divalent metal ions and bind these using a triad of amino acids consisting of the HUH His pair and a third residue, which can be Glu, Asp, His or Gln. In addition to variation in the third metal ion ligand, certain viral Rep proteins with almost identical Rep folds, such as those of circoviruses and nanoviruses[58,59,97], include variations in the HUH motif itself, as do certain MobP relaxases[98]. In circoviruses and nanoviruses, the motif becomes HUQ, so bioinformatic searches for HUH endonucleases would not identify these enzymes. Indeed, there are even proteins that catalyse plasmid RCR or conjugation but do not contain an HUH motif at all[3,9,40]. For example, MobH relaxases contain an HHH motif and a hydrolase motif (both of which are potential divalent metal ion-binding clusters) and an upstream conserved Y motif, whereas MobC relaxases contain a conserved potential metal ion-binding D..E..E triad together with the Y motif[9]. The relationship of these viral proteins to HUH endonucleases cannot be defined until three-dimensional structural information becomes available. Thus, the superfamily of endonucleases that use the HUH mechanism of strand cleavage and rejoining is likely to extend significantly further than the currently defined HUH group.

Similarities with other enzymes.

HUH endonucleases are structurally similar to the replication origin-binding domains encoded by some dsDNA viruses[55,57] and also have strong similarity to the viral RNA recognition motif (RRM)[55,57,99], suggesting that there are intricate and ancient evolutionary relationships between these proteins.

Helicases and the sources of ssDNA.

A central question in the activity of HUH endonucleases is how their substrates become available in the ssDNA form. This role could be fulfilled by the associated helicase domains of many HUH endonucleases (FIG. 1c) or by the interaction of HUH proteins with cellular helicases that can unwind dsDNA. For AAV Rep, the associated SF3 helicase activity is needed for DNA cleavage, suggesting that this activity contributes to the formation of the *trs*-containing hairpin (FIG. 4a insert).

In bacterial systems, in which plasmids are super-coiled, a certain level of supercoiling can lead to hairpin extrusion without the need for helicase activity. By contrast, phage φX174 replication is absolutely dependent on the cellular Rep helicase (an SF1 3′−5′ helicase)[100], but the role of this helicase is in unwinding the phage dsDNA form during RCR, after gpA-mediated nicking.

The 5′−3′ helicase activity of relaxase is thought to promote tracking of the enzyme along the single transferred strand to position it correctly for the termination step (FIG. 5). Although relaxase is capable of binding to one arm of the dsDNA site at *oriT,* additional accessory proteins are involved in melting this region to enable hairpin formation and nicking (FIG. 5a insert). Bacterial ssDNA transposases of the IS*200*-IS*605* family rely on sources of ssDNA[89] such as the lagging-strand template at the replication fork, the ssDNA intermediates generated during DNA repair and possibly R-loops generated during intensive transcription, and these enzymes therefore do not require a dedicated helicase. However, the situation is not clear for IS*91* or, by extension, for the related ISCRs and the eukaryotic Helitrons, both of which are thought to require transposon-specific replication[5]. Helitrons encode a 5′−3′ SF1 helicase; neither IS*91* nor ISCR transposases contain helicase domains, but both have a C-terminal domain of unknown function, and it is tempting to speculate that this domain interacts with a cellular helicase.

Interchangeable functions?

As an illustration of the versatility of HUH endonucleases, members of the Rep and relaxase families can also promote intermolecular strand transfer, leading to integration. AAV2 can integrate site specifically into human chromosome 19 at a locus called *AAVS1,* which carries an RBS-like tandem repeat sequence and a nearby *trs* site[101]. Site-specific integration is Rep dependent, but the mechanism is unclear. Synapsis between the viral ITR and *AAVS1* might be facilitated, for instance, by Rep protomers assembled on *AAVS1* as hexameric or octameric rings[102,103]. In sharp contrast to Y1 transposases, integration mediated by the AAV2 transposase is not precise[104]. Most events occur hundreds of bases away from *AAVS1* in an apparently asymmetrical manner, implying that insertion is not a simple event.

Some relaxases can also site-specifically integrate the transferred strand into the genome of a recipient bacterial cell[72], provided that a second *oriT* is present in the target genome. Both the relaxase and a second TrwC domain in the 600 N-terminal residues are reauired for this activity. The reaction involves a complex sequence of events to resolve the presumed intermediates[105–107]. Moreover, relaxases can be specifically targeted to the nucleus[108], although relaxase-mediated site-specific integration of the transferred DNA in the recipient chromosome has yet to be demonstrated.

This type of site-specific integration reaction might have wide potential applications in biotechnology and biomedicine to directly transfer DNA from bacteria to eukaryotic cells, although practical implementation of this approach is clearly premature at present[109–111].

Conclusions

HUH endonucleases use the same basic catalytic mechanism to carry out a diverse set of biological functions. These enzymes all use Tyr residues as nucleophiles and form covalent 5′-phosphotyrosine bonds with the cleaved DNA strand. The phosphotyrosine bond stores energy that can be harnessed by a 3′-OH located at the other end of the Tyr-linked strand or on the end of a different ssDNA strand, allowing for either intra- or intermolecular strand transfer.

Although HUH endonucleases are involved in a wide variety of biological processes in nature, this is not a consequence of these enzymes carrying out different catalytic reactions. Rather, it is principally a consequence of the different functional modules appended to the HUH domain or recruited as separate entities. The choice between RCR, conjugation or transposition is dictated by the topological and temporal coordination between these various alternative functional modules. Indeed there are examples of one HUH protein assuming more than one role: for example, the relaxase NicK carries out both transfer and replication functions for the transposon ICE*Bs1* (REF 71).

There remain many unanswered questions concerning this widespread class of enzymes. It is still unknown whether and how these proteins interact with the replication apparatus at the fork, and no structural data are available concerning these higher-order interactions, even in the case of φX174, the paradigm of RCR studies. In addition, ISCRs represent an important vector of multiple antibiotic resistance genes, which clearly affect public health[8], and understanding the mechanisms of acquisition and transmission of these genes is undoubtedly a principal concern. Finally, the impact of Helitrons, particularly in shaping plant genomes, makes study of their detailed mechanism an area of priority.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank C. Guynet, S. Messing and I. Molineau for discussions. This work was supported by intramural funding from the French Centre National de la Recherche Scientifique (to M.C. and B.T.H.) and the Intramural Program of the US National Institute of Diabetes and Digestive and Kidney Diseases (to A.B.H. and F.D.), and by grants from the French Agence National de Recherche (ANR-12-BSV8-0009-01; to B.T.H.), the Spanish Ministry of Science and Innovation (BIO2010-14809; to G.M.), the Spanish Ministry of Education (BFU2011-26608; to F.d.l.C.) and the European Seventh Framework Program (248919/FP7-ICT-2009-4 and 282004/ FP7-HEALTH. 2011.2.3.1-2; to F.d.l.C.).

References

1. Kornberg A & Baker TA DNA Replication 2nd edn (Freeman, 1992).
2. Ilyina TV & Koonin EV 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 (1992). [PubMed: 1630899]
3. Koonin EV & Ilyina TV Computer-assisted dissection of rolling circle DNA replication. Biosystems 30, 241–268 (1993). [PubMed: 8374079] References 2 and 3 are the first bioinformatic analyses of HUH endonucleases.
4. Kapitonov VV & Jurka J Rolling-circle transposons in eukaryotes. Proc. Natl Acad. Sci. USA 98, 8714–8719 (2001). [PubMed: 11447285]
5. Garcillan-Barcia MP, Bernales I, Mendiola MV & De la Cruz F in Mobile DNA Vol. II (eds Craig NL, Craigie R, Gellert M, & Lambowitz A) 891–904 (ASM Press, 2002).
6. Ton-Hoang B et al. Transposition of ISHp*608*, member of an unusual family of bacterial insertion sequences. EMBO J. 24, 3325–3338 (2005). [PubMed: 16163392] A description of IS608 and its behaviour.
7. Ronning DR et al. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. Mol. Cell 20, 143–154 (2005). [PubMed: 16209952]

8. Toleman MA, Bennett PM & Walsh TR ISCR elements: novel gene-capturing systems of the 21st century? Microbiol. Mol. Biol. Rev 70, 296–316 (2006). [PubMed: 16760305]
9. Garcillan-Barcia MP, Francia MV & de la Cruz F The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol. Rev 33, 657–687 (2009). [PubMed: 19396961] A discussion of the diversity of relaxases and their classification.
10. Gruss A & Ehrlich SD The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol. Rev 53, 231–241 (1989). [PubMed: 2666843]
11. Rosario K, Duffy S & Breitbart M A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch. Virol 157, 1851–1871 (2012). [PubMed: 22760663]
12. Curcio MJ & Derbyshire KM The outs and ins of transposition: from Mu to Kangaroo. Nature Rev. Mol. Cell Biol 4, 865–877 (2003). [PubMed: 14682279]
13. Odegrip R & Haggard-Ljungquist E The two active- site tyrosine residues of the A protein play nonequivalent roles during initiation of rolling circle replication of bacteriophage p2. J. Mol. Biol 308, 147–163 (2001). [PubMed: 11327759]
14. Grindley ND, Whiteson KL & Rice PA Mechanisms of site-specific recombination. Annu. Rev. Biochem 75, 567–605 (2006). [PubMed: 16756503]
15. Hickman AB et al. DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*. EMBO J. 29, 3840–3852 (2010). [PubMed: 20890269]
16. Boer R et al. Unveiling the molecular mechanism of a conjugative relaxase: the structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. J. Mol. Biol 358, 857–869 (2006). [PubMed: 16540117]
17. Boer DR et al. Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. EMBO J. 28, 1666–1678 (2009) [PubMed: 19440202] The multimeric structure of the Rep protein from plasmid pMV158.
18. Datta S, Larkin C & Schildbach J F Structural insights into single-stranded DNA binding and cleavage by F factor Tral. Structure 11, 1369–1379 (2003). [PubMed: 14604527]
19. Larkin C et al. Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. Structure 13, 1533–1544 (2005). [PubMed: 16216584]
20. Edwards JS et al. Molecular basis of antibiotic multiresistance transfer in *Staphylococcus aureus*. Proc. Natl Acad. Sci. USA 10.1073/pnas.1219701110 (2013).
21. Dyda F & Hickman AB A Mob of Reps. Structure 11, 1310–1311 (2003). [PubMed: 14604517]
22. Guasch A et al. Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nature Struct. Biol 10, 1002–1010 (2003). [PubMed: 14625590] The structure of the TrwC relaxase complexed with its DNA substrate.
23. Hickman AB, Ronning DR, Perez ZN, Kotin RM & Dyda F The nuclease domain of adeno-associated virus Rep coordinates replication initiation using two distinct DNA recognition interfaces. Mol. Cell 13, 403–414 (2004). [PubMed: 14967147] A description of the different binding modes of AAV Rep from a structural point of view.
24. Clerot D & Bernardi F DNA helicase activity is associated with the replication initiator protein Rep of tomato yellow leaf curl geminivirus. J. Virol 80, 11322–11330 (2006). [PubMed: 16943286]
25. Odegrip R, Schoen S, Haggard-Ljungquist E, Park K & Chattoraj DK The interaction of bacteriophage P2 B protein with *Escherichia coli* DnaB helicase. J. Virol 74, 4057–4063 (2000). [PubMed: 10756017]
26. Petit MA et al. PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. Mol. Microbiol 29, 261–273 (1998). [PubMed: 9701819]
27. Bruand C & Ehrlich SD UvrD-dependent replication of rolling-circle plasmids in *Escherichia coli*. Mol. Microbiol 35, 204–210 (2000). [PubMed: 10632890]
28. Chang TL et al. Biochemical characterization of the *Staphylococcus aureus* PcrA. helicase and its role in plasmid rolling circle replication. J. Biol. Chem 277, 45880–45886 (2002). [PubMed: 12244110]

29. Brister JR & Muzyczka N Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J. Virol 73, 9325–9336 (1999). [PubMed: 10516041]
30. Im DS & Muzyczka N The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell 61, 447–457 (1990). [PubMed: 2159383]
31. Brister JR & Muzyczka N Mechanism of Rep- mediated adeno-associated virus origin nicking. J. Virol 74, 7762–7771 (2000). [PubMed: 10933682]
32. Ton-Hoang B et al. Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. Nucleic Acids Res. 40, 3596–3609 (2012). [PubMed: 22199259] The identification and structure of TnpA(REP), and an analysis of the activity of this enzyme.
33. Messing SA et al. The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease. Nucleic Acids Res. 40, 9964–9979 (2012). [PubMed: 22885300]
34. Orozco BM & Hanley-Bowdoin LA DNA structure is required for geminivirus replication origin function. J. Virol 70, 148–158 (1996). [PubMed: 8523519]
35. del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M & Diaz-Orejas R Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev 62, 434–464 (1998). [PubMed: 9618448]
36. Bikard D, Loot C, Baharoglu Z & Mazel D Folded DNA in action: hairpin formation and biological functions in prokaryotes. Microbiol. Mol. Biol. Rev 74, 570–588 (2011).
37. Larkin C, Datta S, Nezami A, Dohm JA & Schildbach J F Crystallization and preliminary X-ray characterization of the relaxase domain of F factor Tral. Acta Crystallogr. D Biol. Crystallogr 59, 1514–1516 (2003). [PubMed: 12876370]
38. Ruiz-Maso JA, Lurz R, Espinosa M & del Solar G Interactions between the RepB initiator protein of plasmid pMV158 and two distant DNA regions within the origin of replication. Nucleic Acids Res. 35, 1230–1244 (2007). [PubMed: 17267412]
39. Gilbert W & Dressler D DNA replication: the rolling circle model. Cold Spring Harb. Symp. Quant. Biol 33, 473–484 (1968). [PubMed: 4891987]
40. Khan SA Plasmid rolling-circle replication: recent developments. Mol. Microbiol 37, 477–484 (2000). [PubMed: 10931341]
41. Brown DR et al. DNA structures required for ϕX174 *A*-protein-directed initiation and termination of DNA replication. Cold Spring Harb. Symp. Quant. Biol 47, 701–715 (1983). [PubMed: 6222868]
42. Brown DR, Schmidt-Glenewinkel T, Reinberg D & Hurwitz J DNA sequences which support activities of the bacteriophage φX174 gene A protein. J. Biol. Chem 258, 8402–8412 (1983). [PubMed: 6223031]
43. Roth MJ, Brown DR & Hurwitz J Analysis of bacteriophage ϕX174 gene A protein-mediated termination and reinitiation of ϕX DNA synthesis. II. Structural characterization of the covalent ϕX A protein-DNA complex. J. Biol. Chem 259, 10556–10568 (1984). [PubMed: 6236216]
44. Hanai R & Wang JC The mechanism of sequence-specific DNA cleavage and strand transfer by ϕX174 gene A* protein. J. Biol. Chem 268, 23830–23836 (1993). [PubMed: 8226920]
45. van Mansfeld AD, van Teeffelen HA, Baas PD & Jansz HS Two juxtaposed tyrosyl-OH groups participate in φX174 gene A protein catalysed cleavage and ligation of DNA. Nucleic Acids Res. 14, 4229–4238 (1986). [PubMed: 2940511] A demonstration of the importance of two Tyr residues in the φX174 Rep protein for catalysis.
46. Barabas O et al. Mechanism of IS*200*/IS*605* family DNA transposases: activation and transposon-directed target site selection. Cell 132, 208–220 (2008). [PubMed: 18243097] The structural model for single-strand transposition of IS608 and other IS200-IS605 family members.
47. Noirot-Gros MF & Ehrlich SD Change of a catalytic reaction carried out by a DNA replication protein. Science 274, 777–780 (1996). [PubMed: 8864116]
48. Novick RP Contrasting lifestyles of rolling-circle phages and plasmids. Trends Biochem. Sci 23, 434–438 (1998). [PubMed: 9852762]
49. Moscoso M, Eritja R & Espinosa M Initiation of replication of plasmid pMV158: mechanisms of DNA strand-transfer reactions mediated by the initiator RepB protein. J. Mol. Biol 268, 840–856 (1997). [PubMed: 9180376]

50. Tattersall P & Ward DC Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. Nature 263, 106–109 (1976). [PubMed: 967244]
51. James JA et al. Crystal structure of the SF3 helicase from adeno-associated virus type 2. Structure 11, 1025–1035 (2003). [PubMed: 12906833]
52. Smith RH & Kotin RM The Rep52 gene product of adeno-associated virus is a DNA helicase with 3′-to-5′ polarity. J. Virol 72, 4874–4881 (1998). [PubMed: 9573254]
53. Maggin JE, James JA, Chappie JS, Dyda F & Hickman AB The amino acid linker between the endonuclease and helicase domains of adeno- associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. J. Virol 86, 3337–3346 (2012). [PubMed: 22205752]
54. Zarate-Perez F et al. The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. PLoS Pathog. 8, e1002764 (2012). [PubMed: 22719256]
55. Hickman AB, Ronning DR, Kotin RM & Dyda F Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol. Cell 10, 327–337 (2002). [PubMed: 12191478]
56. King JA, Dubielzig R, Grimm D & Kleinschmidt JA DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J. 20, 3282–3291 (2001). [PubMed: 11406604]
57. Campos-Olivas R, Louis JM, Clerot D, Gronenborn B & Gronenborn AM The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. Proc. Natl Acad. Sci. USA 99, 10310–10315 (2002). [PubMed: 12130667]
58. Vega-Rocha S, Gronenborn B, Gronenborn AM & Campos-Olivas R Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circovirus structures. Biochemistry 46, 6201–6212 (2007). [PubMed: 17472345]
59. Vega-Rocha S, Byeon IJ, Gronenborn B, Gronenborn AM & Campos-Olivas R Solution structure, divalent metal and DNA binding of the endonuclease domain from the replication initiation protein from porcine circovirus 2. J. Mol. Biol 367, 473–487 (2007). [PubMed: 17275023]
60. Oke M et al. A dimeric Rep protein initiates replication of a linear archaeal virus genome: implications for the Rep mechanism and viral replication. J. Virol 85, 925–931 (2011). [PubMed: 21068244]
61. Noirot-Gros MF, Bidnenko V & Ehrlich SD Active site of the replication protein of the rolling circle plasmid pC194. EMBO J. 13, 4412–4420 (1994). [PubMed: 7925284]
62. Rasooly A & Rasooly RS How rolling circle plasmids control their copy number. Trends Microbiol. 5, 440–446 (1997). [PubMed: 9402700]
63. Laufs J et al. Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. Biochimie 77, 765–773 (1995). [PubMed: 8824773]
64. Lederberg J & Tatum EL Sex in bacteria; genetic studies, 1945–1952. Science 118, 169–175 (1953). [PubMed: 13076234]
65. de la Cruz F, Frost LS, Meyer RJ & Zechner EL Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiol. Rev 34, 18–40 (2010). [PubMed: 19919603]
66. Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F & Rocha EP The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet. 7, e1002222 (2011). [PubMed: 21876676]
67. Burrus V, Pavlovic G, Decaris B & Guedon G Conjugative transposons: the tip of the iceberg. Mol. Microbiol 46, 601–610 (2002). [PubMed: 12410819]
68. Guerillot R, Da Cunha V, Sauvage E, Bouchier C & Glaser P Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating insertion sequence transposition, plasmid replication, and conjugation for their spreading. J. Bacteriol 195, 1979–1990 (2013). [PubMed: 23435978]
69. Smyth DS & Robinson DA Integrative and sequence characteristics of a novel genetic element, *ICE6013, in Staphylococcus aureus*. J. Bacteriol 191, 5964–5975 (2009). [PubMed: 19648240]

70. Rocco JM & Churchward G The integrase of the conjugative transposon *Tn916* directs strand- and sequence-specific cleavage of the origin of conjugal transfer, *oriT,* by the endonuclease Orf20. J. Bacteriol 188, 2207–2213 (2006). [PubMed: 16513750]
71. Lee CA, Babic A & Grossman AD Autonomous plasmid-like replication of a conjugative transposon. Mol. Microbiol 75, 268–279 (2010). [PubMed: 19943900]
72. Draper O, Cesar CE, Machon C, de la Cruz F & Llosa M Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc. Natl Acad. Sci. USA 102, 16385–16390 (2005). [PubMed: 16260740]
73. Kingsman A & Willetts N The requirements for conjugal DNA synthesis in the donor strain during F *lac* transfer. J. Mol. Biol 122, 287–300 (1978). [PubMed: 151148]
74. Nash RP, Habibi S, Cheng Y, Lujan SA & Redinbo MR The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. Nucleic Acids Res. 38, 5929–5943 (2010). [PubMed: 20448025]
75. Monzingo AF, Ozburn A, Ma S, Meyer RJ & Robertus JD The structure of the minimal relaxase domain of MobA at 2.1 A resolution. J. Mol. Biol 366, 165–178 (2007). [PubMed: 17157875]
76. Larkin C, Haft RJF, Harley MJ, Traxler B & Schildbach JF Roles of active site residues and the HUH motif of the F plasmid TraI relaxase. J. Biol. Chem 282, 33707–33713 (2007). [PubMed: 17890221]
77. Gonzalez-Perez B et al. Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. EMBO J. 26, 3847–3857 (2007). [PubMed: 17660746]
78. Dostal L, Shao S & Schildbach JF Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Res. 39, 2658–2670 (2011). [PubMed: 21109533]
79. Lucas M et al. Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the *nic* site. J. Biol. Chem 285, 8918–8926 (2010). [PubMed: 20061574]
80. Kim YJ, Lin LS & Meyer RJ Two domains at the origin are required for replication and maintenance of broad-host-range plasmid R1162. J. Bacteriol 169, 5870–5872 (1987). [PubMed: 2445734]
81. Pansegrau W, Schroder W & Lanka E Concerted action of three distinct domains in the DNA cleavingjoining reaction catalyzed by relaxase (TraI) of conjugative plasmid RP4. J. Biol. Chem 269, 2782–2789 (1994). [PubMed: 8300611]
82. Meyer R Replication and conjugative mobilization of broad host-range IncQ plasmids. Plasmid 62, 57–70 (2009). [PubMed: 19465049]
83. Geibel S, Banchenko S, Engel M, Lanka E & Saenger W Structure and function of primase RepB′ encoded by broad-host-range plasmid RSF1010 that replicates exclusively in leading-strand mode. Proc. Natl Acad. Sci. USA 106, 7810–7815 (2009). [PubMed: 19416864]
84. Honda Y, Sakai H, Komano T & Bagdasarian M RepB′ is required in *trans* for the two single-strand DNA initiation signals in *oriV* of plasmid RSF1010. Gene 80, 155–159 (1989). [PubMed: 2792769]
85. Mendiola MV, Bernales I & de la Cruz F Differential roles of the transposon termini in I*S91* transposition. Proc. Natl Acad. Sci. USA 91, 1922–1926 (1994). [PubMed: 8127907]
86. Lam S & Roth JR IS200: a Salmonella-specific insertion sequence. Cell 34, 951–960 (1983). [PubMed: 6313217]
87. Kersulyte D et al. Transposable element I*SHp608* of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity. J. Bacteriol 184, 992–1002 (2002). [PubMed: 11807059]
88. Islam SM et al. Characterization and distribution of IS*8301* in the radioresistant bacterium *Deinococcus radiodurans*. Genes Genet. Syst 78, 319–327 (2003). [PubMed: 14676423]
89. Ton-Hoang B et al. Single-stranded DNA transposition is coupled to host replication. Cell 142, 398–408 (2010). [PubMed: 20691900]
90. Pasternak C et al. Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence. PLoS Genet. 6, e1000799 (2010). [PubMed: 20090938]

91. Zahradka K et al. Reassembly of shattered chromosomes in *Deinococcus radiodurans*. Nature 443, 569–573 (2006). [PubMed: 17006450]
92. He S et al. Reconstitution of a functional IS*608* single-strand transpososome: role of non-canonical base pairing. Nucleic Acids Res. 39, 8503–8512 (2011) [PubMed: 21745812]
93. Guynet C et al. Resetting the site: redirecting integration of an insertion sequence in a predictable way. Mol. Cell 34, 612–619 (2009). [PubMed: 19524540]
94. Garcillan-Barcia MP & de la Cruz F Distribution of IS*91* family insertion sequences in bacterial genomes: evolutionary implications. FEMS Microbiol. Ecol 42, 303–313 (2002). [PubMed: 19709290]
95. Pilar Garcillan-Barcia M, Bernales I, Mendiola MV & de la Cruz CF Single-stranded DNA intermediates in IS*91* rolling-circle transposition. Mol. Microbiol 39, 494–502 (2001). [PubMed: 11136468]
96. Sinzelle L, Izsvak Z & Ivics Z Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell. Mol. Life Sci 66, 1073–1093 (2009). [PubMed: 19132291]
97. Vega-Rocha S, Gronenborn AM, Gronenborn B & Campos-Olivas R ^{1}H, ^{13}C, and ^{15}N NMR assignment of the master Rep protein nuclease domain from the nanovirus FBNYV. J. Biomol. NMR 38, 169 (2007). [PubMed: 17043745]
98. Varsaki A, Lucas M, Afendra AS, Drainas C & de la Cruz F Genetic and biochemical characterization of MbeA, the relaxase involved in plasmid ColE1 conjugative mobilization. Mol. Microbiol 48, 481–493 (2003). [PubMed: 12675806]
99. Burd CG & Dreyfuss G Conserved structures and diversity of functions of RNA-binding proteins. Science 265, 615–621 (1994). [PubMed: 8036511]
100. Scott JF, Eisenberg S, Bertsch LL & Kornberg A A mechanism of duplex DNA replication revealed by enzymatic studies of phage ϕX174: catalytic strand separation in advance of replication. Proc. Natl Acad. Sci. USA 74, 193–197 (1977). [PubMed: 138139]
101. Kotin RM et al. Site-specific integration by adeno-associated virus. Proc. Natl Acad. Sci. USA 87, 2211–2215 (1990). [PubMed: 2156265]
102. Smith RH, Spano AJ & Kotin RM The Rep78 gene product of adeno-associated virus (AAV) self-associates to form a hexameric complex in the presence of AAV *ori* sequences. J. Virol 71, 4461–4471 (1997). [PubMed: 9151837]
103. Mansilla-Soto J et al. DNA structure modulates the oligomerization properties of the AAV initiator protein Rep68. PLoS Pathog. 5, e1000513 (2009). [PubMed: 19593381]
104. McCarty DM, Young SM Jr & Samulski RJ Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu. Rev. Genet 38, 819–845 (2004). [PubMed: 15568995]
105. Cesar CE, Machon C, de la Cruz F & Llosa M A new domain of conjugative relaxase TrwC responsible for efficient *oriT*-specific recombination on minimal target sequences. Mol. Microbiol 62, 984–996 (2006). [PubMed: 17038118]
106. Cesar CE & Llosa M TrwC-mediated site-specific recombination is controlled by host factors altering local DNA topology. J. Bacteriol 189, 9037–9043 (2007). [PubMed: 17921309]
107. Agundez L, Gonzalez-Prieto C, Machon C & Llosa M Site-specific integration of foreign DNA into minimal bacterial and human target sequences mediated by a conjugative relaxase. PLoS ONE 7, e31047 (2012). [PubMed: 22292089]
108. Agundez L et al. Nuclear targeting of a bacterial integrase that mediates site-specific recombination between bacterial and human target sequences. Appl. Environ. Microbiol 77, 201–210 (2011). [PubMed: 21037296]
109. Schröder G, Schuelein R, Quebatte M & Dehio C Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae*. Proc. Natl Acad. Sci. USA 108, 14643–14648 (2011). [PubMed: 21844337]
110. Fernandez-Gonzalez E et al. Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. J. Bacteriol 193, 6257–6265 (2011). [PubMed: 21908662]
111. González-Prieto C, Agúndez L, Linden RM & Llosa M HUH site-specific recombinases for targeted modification of the human genome. Trends Biotechnol. 31, 305–312 (2013). [PubMed: 23545167]

112. Pritham EJ & Feschotte C Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. Proc. Natl Acad. Sci. USA 104, 1895–1900 (2007). [PubMed: 17261799]

113. Kapitonov VV & Jurka J Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet. 23, 521–529 (2007). [PubMed: 17850916]

114. Feschotte C & Wessler SR Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes. Proc. Natl Acad. Sci. USA 98, 8923–8924 (2001). [PubMed: 11481459]

115. Higgins CF, Ames GF, Barnes WM, Clement JM & Hofnung M A novel intercistronic regulatory element of prokaryotic operons. Nature 298, 760–762 (1982). [PubMed: 7110312]

116. Bachellier S, Clément J-M & Hofnung M Short palindromic repetitive DNA elements in enterobacteria: a survey. Res. Microbiol 150, 627–639 (1999). [PubMed: 10673002]

117. Nunvar J, Huckova T & Licha I Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes. BMC Genomics 11,44 (2010). [PubMed: 20085626]

118. Bertels F & Rainey P B. Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria. PLoS Genet. 7, e1002132 (2011). [PubMed: 21698139]

119. Braun V et al. A chimeric ribozyme in *Clostridium difficile* combines features of group I introns and insertion elements. Mol. Microbiol 36, 1447–1459 (2000). [PubMed: 10931294]

Box 1 | The Helitron superfamily

All the HUH transposases (in which U represents a hydrophobic residue) described in the main text are specific to bacteria and archaea and thus catalyse the movement of DNA elements in bacterial and archaeal genomes. However, Helitrons are a family of mobile genetic elements that are present in eukaryotic genomes. Helitrons are found in high numbers in *Arabidopsis* thaliana, *Oryza sativa,Zea maize, Caenorhabditis elegans, Drosophila melanogaster*[112] and Myotis lucifugus[113]. Bioinformatic analysis suggests that these elements are mobilized by rolling-circle replication-based transposition, similarly to certain bacterial and archaeal elements[112]. It has been proposed that Helitrons gave rise to geminiviruses[114], although the helicase domains of Helitrons differ significantly from the superfamily 3 (SF3) helicases that are generally associated with geminiviruses.

Helitrons contain a particularly long ORF, *RepHel,* composed of a number of exons (sometimes more than ten)[4,112] encoding a protein with both HUH and Y2 motifs, together with putative amino-terminal Zn-finger and carboxy-terminal SF1-like helicase domains (FIG. 1c). Moreover, in many Helitrons, RepHel exhibits additional apurinic endonuclease and/or cysteine protease domains[112]. These elements can also include a separate protein that binds single-stranded DNA. Helitrons have fairly well conserved ends (generally 5′-TC and often CTAG-3′) and short 16–20 -bp subterminal hairpins at the right end, but no inverted repeats, and are also often flanked by 5′-T and A-3′.

Box 2 | Domestication of IS*200*-IS*605* family transposases

In bacteria, two groups of mobile genetic elements, the repetitive extragenic palindromic sequences (REP sequences; not to be confused with the Rep (replication) HUH proteins) and the IStrons, use transposases that are closely related to the transposase (TnpA) proteins of the insertion sequence IS*200*-IS*605* family.

REP sequences

REP sequences, which were identified 30 years ago[115], are inverted DNA repeats that can form hairpin structures and are almost exclusively located in intergenic regions of bacterial genomes[115]. REP sequences are often present in high copy numbers (there are ~590 copies in *Escherichia coli* K12 and >1,600 copies in *Stenotrophomonas* maltophilia) and grouped into pairs in structures called bacterial interspersed mosaic elements (BIMEs). BIMEs are composed of a REP sequence and an inverted REP (iREP) sequence[116], and might have several key roles in aspects of host physiology, including organization of the genome structure, regulation of gene expression, and genome plasticity (see REF 32).

Recently, a group of REP-associated proteins called TnpA(REP) proteins[32] (also known as RAYTS[117]) was identified and found to be closely related to IS*200*-IS*605* family transposases. *tnpA*(*REP*) is generally present in a single copy located between flanking REP sequences. TnpA(REP) catalyses REP sequence cleavage and joining *in vitro*[32], but is less specific than IS*200*-IS*605* family transposases.

TnpA(REP) is structurally closely related to the TnpA proteins of IS*608* and IS*Dra2* but, unlike these dimeric TnpA proteins, is a monomer[33] (FIG. 2). The structure has been determined for TnpA(REP) bound to a REP hairpin extended on its 5′ side by 4 bases (analogous to the guide sequence found in IS*608*), and in this structure the single-stranded DNA ‘guide sequence’ is extensively contacted by the protein. As is the case for IS*Dra2* TnpA, hairpin recognition by TnpA(REP) requires a mismatch in the stem, and the 5′ guide sequence is important for cleavage[33]. In spite of the evident catalytic properties of TnpA(REP) and the high prevalence of REP sequences, only scant information is available about the origin and behaviour of REP sequences or the role of TnpA(REP) *in vivo.* Recently, bioinformatic approaches detected a BIME excision event in the *Pseudomonas fluorescence* GW25 genome, suggesting that BIMEs are in fact mobile elements[118], but there is no direct evidence for the involvement of TnpA(REP) in this process. As TnpA(REP)-mediated cleavage and rejoining requires only a single REP sequence (the iREP sequence is refractory to TnpA(REP) activity[32]), it is intriguing that REP sequences are grouped in BIMEs. This paired configuration might be linked to genome replication because the iREP sequence on the lagging strand would become a REP sequence on the leading strand. The mechanism by which REP sequences invade and propagate within genomes remains to be addressed.

Group I introns

Other known TnpA-like proteins occur in certain group I bacterial introns called IStrons. These elements are rich in secondary structure and encode a TnpA-like potential homing

endonuclease[119] similar to TnpA of IS*Dra2*. All of the IStron copies analysed so far have been found to be inserted 3′ to the pentanucleotide TTGAT, which is also the target sequence used by IS*Dra2*. TTGAT is complementary to the intron internal guide sequence and, at the RNA level, is presumably required in the splicing reaction. Little is known about IStron behaviour.

Repetitive extragenic palindromic sequences

Abundant non-coding repeats that are found in bacterial genomes and form DNA hairpin structures that can have regulatory functions.

Author Manuscript

Author Manuscript

Author Manuscript

Rolling-circle replication

Unidirectional replication of circular DNA molecules (such as plasmids and phage genomes) in which a single-stranded product is ‘peeled’ from a circular DNA template.

Insertion sequences with a common region

Insertion sequences that contain a common ORF which has similarity to the transposase ORF encoded in insertion sequence IS*91*. These elements seem to be able to sequester additional neighbouring genes during their transposition.

Author Manuscript

Author Manuscript

Author Manuscript

Insertion sequences

Short transposable DNA segments that include one or more genes involved in their own mobility.

Author Manuscript

Author Manuscript

Author Manuscript

Helitron

A type of eukaryotic transposon that is thought to transpose using a rolling-circle replication mechanism similar to that of insertion sequence *IS91*.

Helicase

An enzyme that unwinds and separates double-stranded DNA. Helicases are classified into six major superfamilies (SF1-SF6) on the basis of the motifs and consensus sequences shared by the molecules and their activities (for example, 5′−3′ or 3′−5′ directionality).

Type IV secretion system

A multiprotein apparatus that is used by bacteria to transport both DNA and proteins across the bacterial cell envelope.

Integrative and conjugate element

A mobile genetic element that has the transfer properties of a conjugative plasmid but that is generally unable to replicate autonomously and possesses dedicated integration system to allow the element to be maintained by integration into the host chromosome.

Author Manuscript

Author Manuscript

Author Manuscript

DDE family transposase

A transposase that contains a characteristic DDE amino acid motif.

R-loops

Regions of DNA in which the two strands are separated by a short RNA segment that forms a complementary RNA-DNA hybrid with one of the DNA strands.

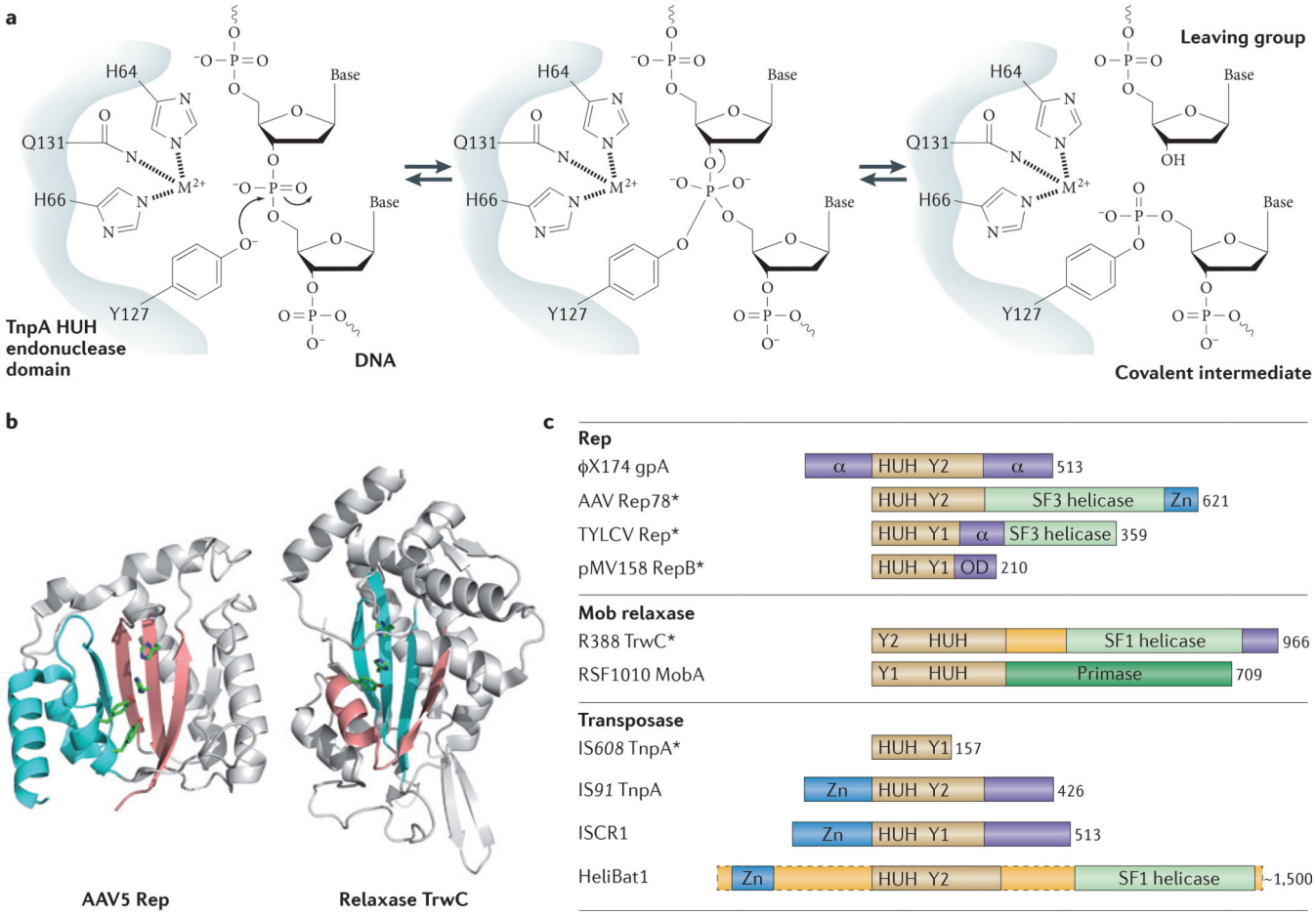

Figure 1 |. Reaction mechanism and organization of selected HUH proteins.

a | The HUH endonuclease domain (in which U is a hydrophobic residue), together with the divalent metal ion (M^{2+}) ligand, nicks single-stranded DNA (ssDNA) to form a covalent intermediate and release the cleaved leaving group. **b** | Both the Rep (replication) domain of adeno-associated virus 5 (AAV5) Rep (Protein Data Bank (PDB) accession 1M55) and the relaxase domain of plasmid R388 TrwC (PDB accession 1OMH) have a five-stranded β-sheet with a βαββ core that harbours the HUH motif on the central strand. Owing to the different domain organization in Rep proteins and relaxases, the position of this core differs (βαββαβαβ for Rep proteins and βαβαβαββ for relaxases). The more amino-terminal residues in the primary sequence are shown in pink; those more carboxy-terminal are in blue. **c** | The organization of representative HUH domain-containing proteins is shown; they contain HUH, helicase, oligomerization (OD) and proposed Zn-binding (not necessarily structurally related) domains. The length of each protein is indicated in numbers of amino acids, and those proteins for which HUH domain structures are available are indicated with an asterisk; the HUH motif data are taken from REF. 3 (motif 2) and REF. 9 (motif III); the Y motif data are taken from REF. 3 (motif 3) and REF. 9 (motif I). The assigned domain organizations are taken from phage ϕXl74 protein A (gpA)[17], AAV Rep78 (REF. 102), tomato yellow leaf curl virus (TYLCV) Rep[57], plasmid pMV158 RepB[17], plasmid R388 TrwC[22], plasmid RSF1010 MobA (mobilization protein A)[75], transposases from the

insertion sequences IS608 (REF. 7), IS*91* (REF. 7) and insertion sequence with a common region 1 (ISCR1) (S. Messing, A.B.H. and F.D., unpublished observations), and HeliBat1 (a consensus sequence from a bioinformatic prediction)[113]. Image in part **a** is modified, with permission, from REF. 7 © (2005) Cell Press.

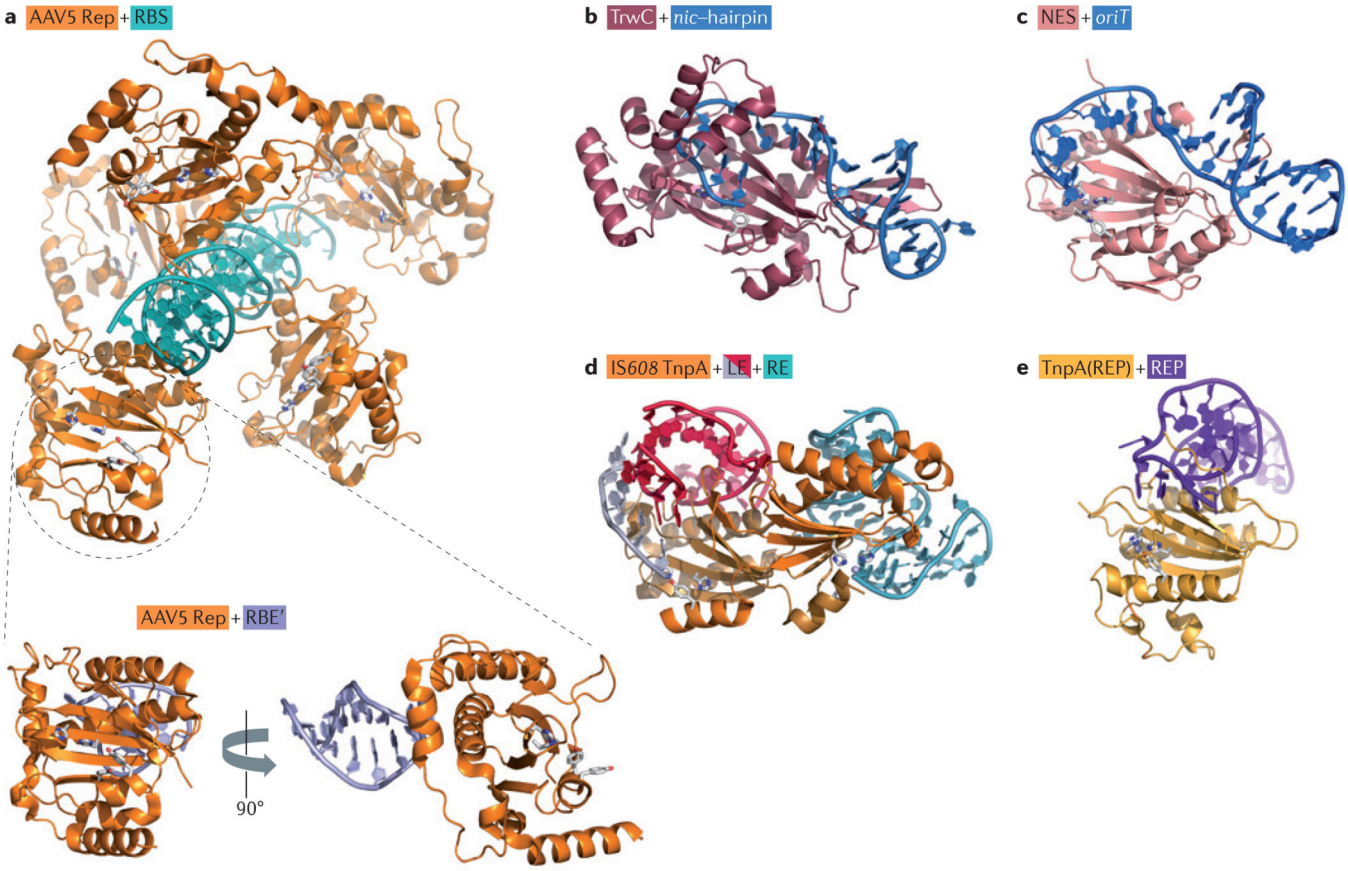

Figure 2 |. Structures of various HUH enzymes with their substrate DNAs.

a | The interaction of adeno-associated virus (AAV) Rep (replication) protein with the inverted terminal repeat hairpin in the DNA; the top panel shows the interaction with the 20 bp Rep-binding site (RBS) (Protein Data Bank (PDB) accession 1RZ9), and the lower panel shows the interaction with the Rep-binding element (RBE') (PDB accession 1UUT). **b** | The interaction between the plasmid R388 relaxase, TrwC, and a 25-base oligonucleotide containing both the plasmid *nic* site and the recognition hairpin (PDB accession 2CDM). **c** | The interaction of nicking enzyme in *Staphyloeoeeus* (NES) with the origin of transfer (*oriT*), represented by an oligonucleotide of 30 bases (PDB accession 4HT4). **d** | A model of the interaction between insertion sequence IS*6*08 transposase (TnpA) and both the transposon left end (LE) (PDB accession 2VJV) and right end (RE) (PDB accession 2VHG). **e** | The interaction of TnpA(REP) with a repetitive extragenic palindromic (REP) sequence (PDB accession 4ER8).

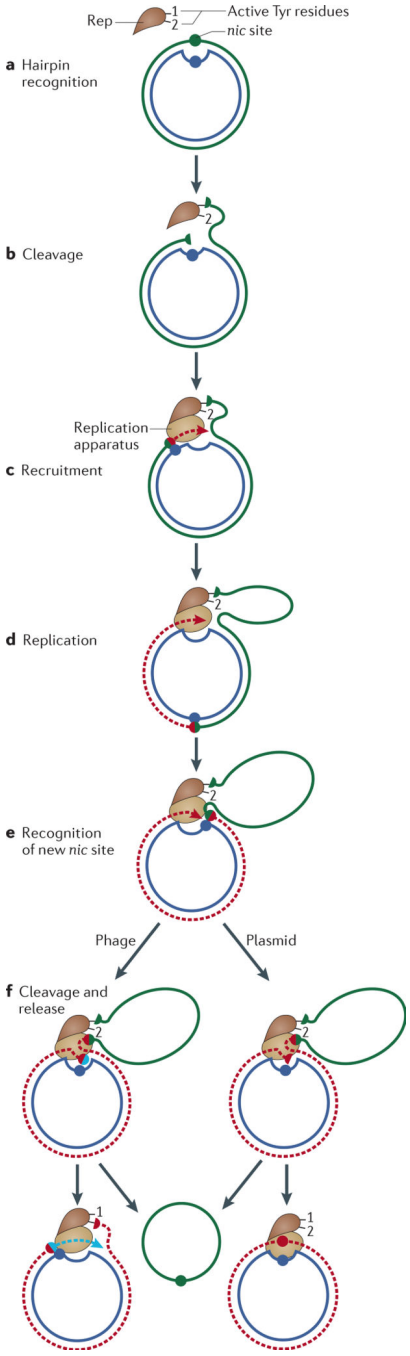

Figure 3 |. Rolling-circle replication.

a | The Rep (replication) protein contains two active-site Tyr residues, 1 and 2, and recognizes circular double-stranded DNA (positive-sense strand in green, negative-sense strand in blue) at the the 3′ recognition sequence. **b** | Rep cleaves the DNA at the *nic* site and forms a phosphotyrosine bond with Tyr residue 1. **c** | The replication apparatus is recruited. **d** | Replication is initiated and reconstitutes the *nic* site. **e** | When the entire circle of DNA has been replicated, Rep and the replication apparatus reach the reconstituted *nic* site. **f** | In the case of phage replication, the *nic* site is cleaved by the Rep active-site Tyr residue 2 to form a

new phosphotyrosine. The resulting 3′-OH attacks the original phosphotyrosine bond (with Tyr residue 1) to release a single-stranded circular phage copy. Replication of the Rep-bound circular DNA then continues round the circle again. In the case of copy number-regulated plasmid replication, no second phosphotyrosine bond is formed, and the plasmid regains its initial circular dsDNA form possibly through the use of a second nucleophile (such as water) instead of a Tyr[47]. Failure to terminate at this stage would lead to multimeric single-stranded products.

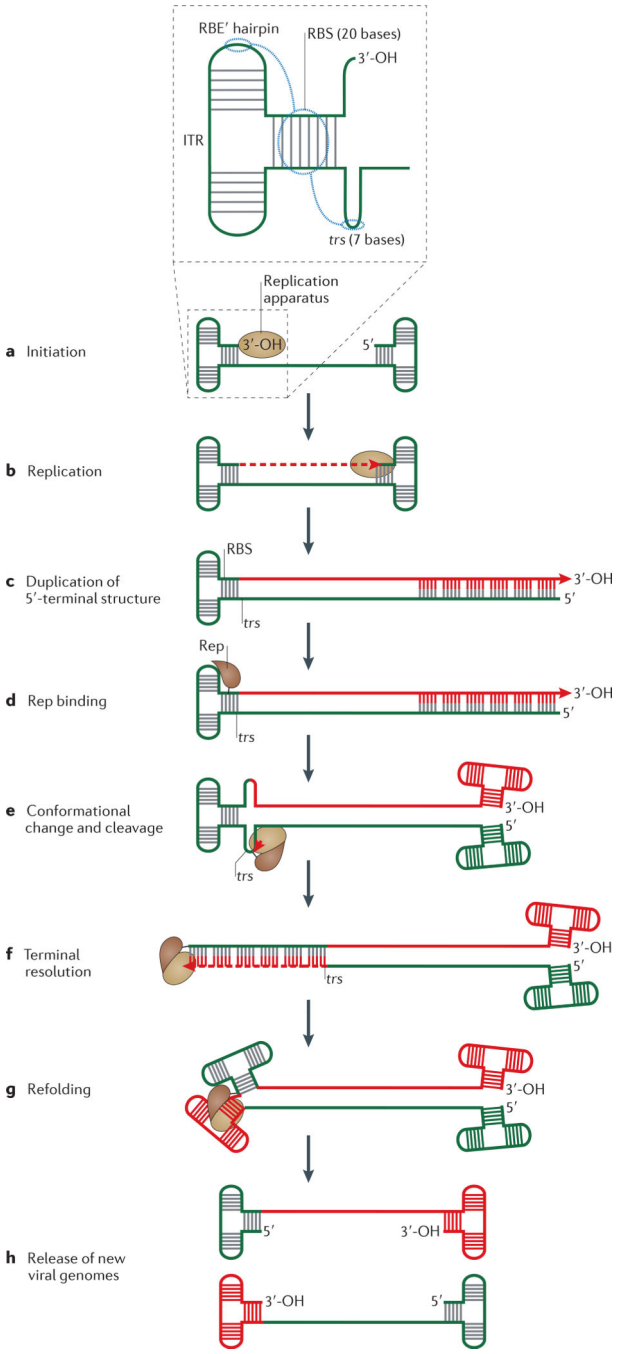

Figure 4 |. Rolling hairpin replication of adeno-associated virus.

For simplicity, secondary structures at the ends of the single-stranded adeno-associated virus (AAV) genome (called inverted terminal repeats (ITRs)) are shown as T junctions with complementary base pairing. Parental DNA is green and newly synthesized DNA is red. **a** | The insert shows a schematic organization of the ITR, and blue circles indicate ITR regions that interact with the protein Rep (replication). Replication is initiated using the host replication machinery and the 3′-OH of the 3′ ITR as a primer. **b**,**c** | Replication continues to the end of the genome, duplicating the 5′-terminal ITR structure. **d** | For AAV5, five Rep

molecules bind to the 20-base $(GCTC)_5$ Rep-binding site (RBS) in the 3′ ITR and contact the Repbinding element (RBE′) hairpin tip. **e** | This binding is proposed to provoke a conformational change in the DNA at the palindromic terminal resolution site (*trs*), and to induce Rep-mediated cleavage at *trs* to generate a phosphotyrosine bond between Rep and the DNA. **f** | Recruitment of the host replication machinery then allows replication of the 3′-terminal structure (a step called terminal resolution). **g** | Refolding of the ends generates structures resembling those that are present before replication initiation. **h** | The end result is a fully replicated viral genome.

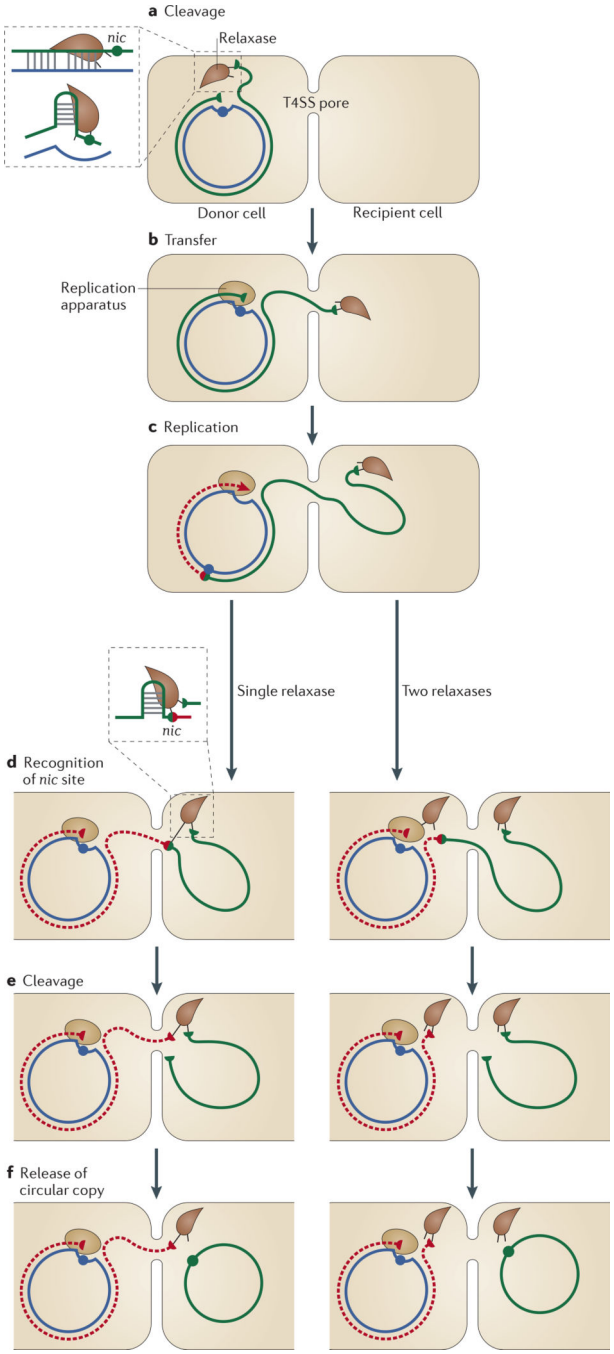

Figure 5 |. Rolling-circle replication-mediated conjugation.

a | Conjugative transfer occurs through a type IV secretion system (T4SS) pore between the donor and recipient cells. Conjugation is initiated by the relaxase recognizing one side of a double-stranded inverted repeat structure (insert) proximal to a 3′ *nic* site. Nucleophilic attack of the *nic* site generates a phosphotyrosine bond between the cleaved strand of the *nic* site and a Tyr residue in the relaxase. **b** | The linked single-stranded DNA (ssDNA) is transferred with the relaxase into the recipient cell. **c** | Replication is initiated in the donor cell using the host replication apparatus and regenerates the *nic* site. **d-f** | Replication

continues until the entire circle of DNA has been replicated. The *nic* site that was reconstituted at the onset of replication is then cleaved by a relaxase (insert) to generate a circular ssDNA copy of the genome in the recipient cell. This cleavage can be mediated by the relaxase bound to the original nicked end, in the recipient cell, or by a second relaxase, in the donor cell. Replication in the donor cell is generally concomitant with transfer but is not essential. Replication of the transferred single strand is carried out by the replication apparatus of the recipient cell.

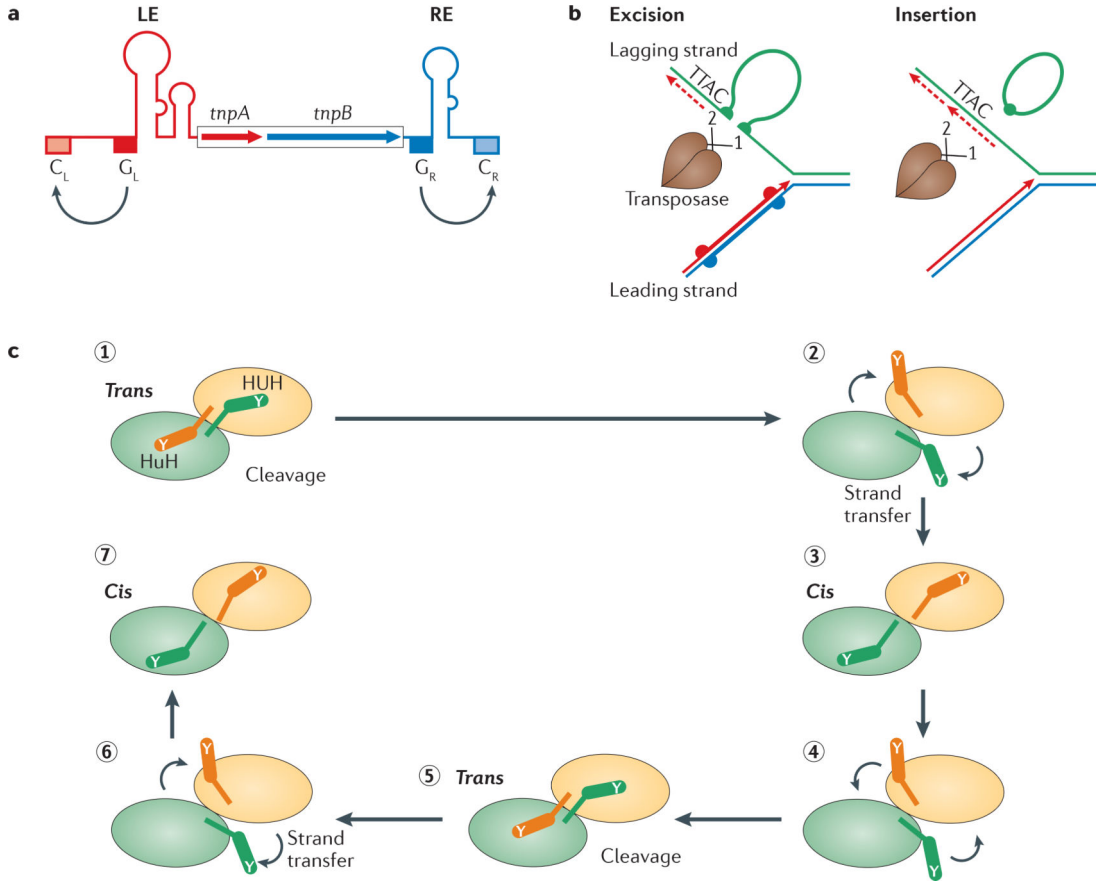

Figure 6 |. Single-strand transposons.

a | The genetic organization of the insertion sequence IS*608*, a member of the IS*200*-IS*605* family. The ORFs *tnpA* and *tnpB* encode transposase (TnpA; a Y1 transposase) and a protein of unknown function, TnpB. The left end (LE) and right end (RE) regions have potential secondary structures, as indicated schematically. The left and right guide sequences (G_L and G_R, respectively) recognize and interact with the left and right cleavage sites (C_L and C_R, respectively). **b** | IS*200*-IS*605* transposons preferentially excise from and integrate into the lagging-strand template of the replication fork. The dashed red line represents an Okazaki fragment, and the solid red line represents the newly synthesized leading strand. The dimeric TnpA (black sticks indicating catalytic Tyr residues) catalyses excision of the single-stranded insertion sequence, resulting in a deletion in the lagging-stand template. The single-stranded insertion sequence then attacks the lagging-strand template in a target molecule and inserts 3′ to the TTAC target tetranucleotide. **c** | Each monomer of the IS*608* TnpA dimer undergoes a series of *cis-trans* conformational changes during the transposition cycle, displacing the HUH motif (in which U is a hydrophobic residue) in the main body of the protein and the catalytic Tyr residue (Y) situated on an α-helix (αD). For binding and cleavage of the left and right ends of the insertion sequence, the active sites adopt a *trans* configuration, in which the HUH motif of an active site is contributed by one TnpA monomer and the Tyr residue is contributed by the other (step 1). Following DNA cleavage, the two αD-helices rotate to the *cis* configuration (step 2). In this configuration, the αD-helices then catalyse the formation of a circular single-stranded transposon and join the

cleaved ends of the donor sequence (step 3). The configuration is then reset to the *trans* form to end excision (step 4). Integration starts with the active sites, in a *trans* configuration, cleaving the single-stranded transposon and the target DNA (step 5). The two αD-helices rotate for strand transfer (step 6), and the active sites return to a *cis* configuration to integrate the transposon (step 7).