**Target choice**

Inspection of the public databases suggests that IS density is significantly higher in bacterial plasmids than in their host chromosomes and it seems likely that plasmids are major vectors in IS transmission (as well as in transmission of accessory traits such as resistance to antibacterials).

Indeed, plasmids, in particular those that use a rolling circle mechanism either for replication or in conjugation, appear to be preferential targets for certain TE. For example, transposon Tn7 has two modes of transposition: in one, a specific sequence within the highly conserved glmS is recognised and insertion occurs next to this essential gene; in the second, insertion occurs into replication forks directed by interactions with the b-clamp. This results in a strong orientation bias of Tn7insertions, consistent with insertion into the lagging strand of the replication fork formed during conjugative transfer. Although studies with IS are less advanced, a similar orientation bias was observed with IS903, suggesting that it too may use the b-clamp in directing insertions. It seems probable that many other IS use this type of protein–protein interaction. Other examples of sequence-specific target choice have been described. IS1, for example, shows a preference for regions rich in AT, whereas the transposon TnGBS and members of the related ISLre2family show a preference for insertion 15–17 bp upstream ofrA promoters.

There are also examples of IS (e.g. some members of the IS110,IS3and IS4families) which insert into potential secondary structures such as repeated extragenic palindromes (REP), integrons, …etc.

**Impact of ISs on genome expression**

While massive IS-mediated genomic changes leading to streamlined genomes with increased pathogenicity and virulence are important and spectacular, they do not reflect the full impact of ISs. ISs can incorporate additional genes and subsequently act as vectors for these genes . . Areas of topical relevance in this respect are the transmission of antibacterial resistance and virulence.

ISs also play more subtle roles. They can insert upstream of a gene and activate its expression This has recently received much attention due to the increase in resistance to various antibacterials.

Activation of neighbouring gene expression can occur in two principal ways: either via promoters contained entirely within the IS driving transcripts that escape into neighbouring DNA, or by the formation of hybrid promoters following insertion. Many ISs contain outward oriented -35 promoter components in their ends and insertion at the correct distance from a suitable -10 box can generate a strong promoter. This property has been noted for a very large number of ISs in a variety of bacterial strains. A preliminary survey of the literature shows that the phenomenon is associated with over 30 different ISs and has occurred in at least 17 bacterial species.

IS activity can affect efflux mechanisms resulting in increased resistance: IS1or IS10insertion can up-regulate the AcrAB-TolC pump in Salmonella enterica; IS1or IS2insertion upstream of AcrEF and IS186 insertional inactivation of the AcrAB repressor, AcrR, in Escherichia coli, all lead to increased resistance to fluoroquinolones. Insertional inactivation of specific porins can also play a significant role.

**Diversity and classification**

The primary difference between ISs is the nature of their transposases based on the type of chemistry they catalyse. These include DDE, DEDD, HUH and Ser transposases.

**ISs with DDE transposases**

Classical ISs with DDE transposases (named for a conserved amino acid triad, Asp, Asp, Glu, the active site) are small (c. 0.7–2.5 kb long) genetically compact DNA segments with one or two open reading frames (a Tpase and possibly a protein involved in regulation). They end with imperfect IRs and generate short flanking DRs on insertion. The DR length is specific for each IS type. There are clearly several dominant families among the DDE IS group. These include IS3, whose Tpases are perhaps the most closely related to the retroviral intégrase catalytic core (IN) by the spacing of the DDE triad and by the appearance of additional conserved residues. The IS3 family contains several well-defined clades delimiting subgroups. With subsequent accumulation of additional members, it has been necessary to redefine several large families, such as IS4, into a number of individual families. A similar situation has also occurred for the IS5 family whose original members are at present distributed over several families and subgroups. Some of these will certainly develop into separate families.

**ISs with DEDD transposases**

DEDD transposases (for Asp, Glu, Asp, Asp) are related to the Holiday junction resolvase, RuvC, itself related to DDE transposases. At present, only a single IS family, IS110, is known to encode this type of enzyme. The organisation of family members is quite different from that of the DDE ISs: they do not contain the typical terminal IRs of the DDE ISs and do not generate flanking target DRs on insertion. This implies that their transposition occurs using a different mechanism to the DDE ISs.

**ISs with HUH Y1 or Y2 transposases**

Two of these families encode a tyrosine (Y) Tpase. Note that these Tpases are not related to the well-characterised tyrosine site-specific recombinases such as phage integrases. Neither carries terminal IRs nor do they generate DRs on insertion. One family includes IS91, and the other includes IS200/IS605. Members of these families transpose using an entirely different mechanism to ISs with DDE transposases. These ISs carry subterminal sequences, which are able to forms hairpin secondary structures. This is particularly marked in the IS200/IS605 family elements.

These ISs are defined by the presence of a Tpase belonging to the HUH endonuclease superfamily (named for the conserved active site amino acid residues H=Histidine and U=large hydrophobic residue). There are two major HUH Tpase families: Y1 and Y2 enzymes according to whether they carry one or two Y residues involved in catalysis. One (Y1) is associated with the IS200/IS605 family. The second (Y2) is associated with the IS91 insertion sequence family, with a related and newly defined group, the ISCR and with eukaryotic helitrons. Although these enzymes use the same Y-mediated cleavage mechanism, they appear to carry out the transposition process in quite different ways.

While the IS91family is fairly homogenous, the IS200/IS605 family is divided into three major subgroups, IS200,IS1341, and those that resemble IS605. This is based on the presence or absence of two reading frames tnpA and tnpB, which can occur individually or together in several configurations: tnpA encodes the Tpase and tnpB encodes a protein with a possible role in regulation.

**ISs with serine transposases**

The third family is represented by IS607, which carries a Tpase closely related to serine recombinases such as the resolvases of Tn3 family elements. Little is known about their transposition mechanism. However, it appears likely, in view of the known activities of resolvases, that IS607transposition may involve a circular double-strand DNA intermediate.

**ISs with accessory genes**

Many ISs also include accessory genes involved in regulating their transposition. These are relatively specific for each IS family and thus also serve in definition of the family. However, in many cases, the exact role and activity of the gene product is unclear.

**IS21**

IS21family members encode a ‘helper’ gene, istB which exhibits some similarity to the DnaA replication initiator protein due to the presence of an ATP binding motif, and often appears in BLAST searches of complete genomes. The molecular details of IstB activity are not known.

**IS200/IS605and IS607**

Although IS200/IS605 and IS607 family members carry very different types of transposase, they often include a second orf, tnpB, in addition to their transposases. TnpB is not required for transposition either in vitro or in vivo. However, it has been observed to reduce ISDra2(IS200/IS605family) transposition activity both in its original host, Deinococcus radiodurans, and in E. coli. The molecular details of TnpB activity are not known. Full-length TnpB includes three domains, an N-terminal HTH, a central domain and a C-terminal zinc finger (ZF) domain. However, this TnpB configuration is quite variable and there are a large number which appear to be undergoing decay. TnpB analogues have been identified as part of IS607-like elements in the Mimi virus and other NCLDV. They have recently been identified in a variety of eukaryotic genomes sometimes associated with other TEs.

**IS66**

In addition to its DDE transposase gene, tnpC, IS66 can include two additional genes, tnpA and tnpB, whose function is as yet unknown. The three reading frames are disposed in a pattern suggesting translational coupling: tnpB is in general in translational reading frame -1 compared to tnpA and in most cases the termination codon of tnpA and the initiation codon of tnpB overlap (ATGA).

**IS91**

While the canonical IS91carries only a single orf, encoding an HUH Y2 transposase, several other family members (e.g. ISAzo26; ISCARN110; ISMno23; ISSde12; ISShvi3; ISSod25 and ISWz1) include a second orf located upstream. This orf is a Y-recombinase related to the phage integrase family whose role in transposition remains to be determined.

**Tn3family**

The Tn3 family is an extensive group of transposons which encode large (>900 aa) DDE transposases. They are included here because certain family members resemble ISs (e.g. IS1071) as they encode only the transposase.

However, the replicative transposition mechanism of this family involves formation of a cointegrate in which donor and recipient replicons are fused and separated at each junction with a directly repeated transposon copy. These structures must be ‘resolved’, by recombination between the two transposon copies, to generate the donor and target replicons each retaining a single transposon copy. This is accomplished by a ‘resolvase’, a site-specific recombinase which acts at a unique DNA sequence in the transposon, the Res site. While many Tn3 family members encode a serine recombinase, several are now known to carry a tyrosine site-specific recombinase (TnpI in Tn4655and Tn4330) resembling phage integrases. Moreover, a third group of Tn3family members include two genes, TnpS and TnpT. It is possible that, as for the bacteriophage k which uses a Y site-specific recombinase (Int) for integration together with a second protein, Xis for excision, TnpT is involved in assuring directionality in transposition. There is no evident difference in the Tpases of these two Tn3-like groups.

**ISs and relatives with passenger genes**

An increasing number of ISs carrying passenger genes are being identified. These include genes for transcription regulators, methyltransferases and antibiotic resistance. They can be located upstream, downstream or on both sides of the transposase gene. These elements, simpler than known Tns, are called transporter ISs (tISs). They can be significantly longer than typical ISs (e.g. ISCausp2, 7915 bp). Several of these ‘extended’ ISs include a significant length of DNA with no clear coding capacity (e.g. ISBse1,ISSpo3, ISSpo8).

At least some incomplete ISs presently identified in sequenced genomes may be of this type because the second IS end would occur at an unexpectedly distant position and would not necessarily have been identified.

tISs are generally present in low copy number. Many occur only in single copy in a given genome raising the question of whether they are active. Moreover, more than one closely related but nonidentical derivative can be found in a single genome (e.g. ISSpo3,and ISSpo8in Silicibacter pomeroyi DSS-3 and ISNwi4and ISNha5in Nitrobacter winograskiNb-255). Others are present in more than one copy. ISPre3, an IS66 relative from Pseudomonas resinovorans plasmid pCAR1 includes a hypothetical protein and is present as 2 copies with different insertion sites as judged by their typical 8 bp DRs.

The mechanisms involved in acquisition of additional genes to generate tISs are at present unclear. They do not appear to carry programmed recombination systems as do some members of the Tn3family. One possibility is that tISs are derived by deletion from ancestral compound transposons. These are composed of two ISs flanking any DNA segment either in direct or inverted orientation. The flanking ISs are able to mobilize the intervening DNA segment. They were among the first types of transposon described and include the models Tn5 (flanking IS50) and Tn10 (flanking IS10) as well as Tn9 (flanking IS1).

Some early observations concerning Tn5and Tn10suggest that the flanking ISs can undergo mutation rendering them less autonomous [for example mutations within one IS which inactivate its transposase]. Furthermore, studies on IS101from the pSC101 plasmid clearly indicated that transposition can occur using one established IS end and a second surrogate end located at some distance from the IS. If, as in this case, the intervening DNA includes a passenger gene, this creates a novel transposon.

It has also been observed that other ISs such as IS911 can use surrogate ends during transposition. Moreover, isolated individual IS ends are often observed in sequenced genomes and could provide a source of surrogate ends.

**The Tn3transposon family derivatives**

Tn3family members are quite variable: several examples lack passenger genes and therefore do not fall into the strict definition of a transposon, while others lack both passenger and resolvase genes (e.g. ISVsa19,ISShfr9, ISBusp1,IS1071...) and therefore closely resemble ISs.

Many family members carry a number of passenger genes. These can represent entire operons, notably mercury resistance, or individual genes involved in antibiotic resistance, breakdown of halogenated aromatics or virulence. They often carry integron recombination platforms enabling them to incorporate additional resistance genes by recruiting integron cassettes.

**IS-related ICEs**

Other structures which obscure the definition of an IS have been identified among various TEs. For example, ICEs (integrative conjugative elements) are found integrated into the host genome but can excise and transfer from cell to cell. Their insertion and excision are generally catalysed by enzymes related to site-specific recombinases, whereas their transfer depends on a second set of proteins, which includes a ‘relaxase’, often a single-strand endonuclease of the HUH superfamily. However, ISSag10, a tIS member of the IS1595family from Streptococcus agalactiae which includes an O-lincosamide nucleotidyltransferase gene, encodes a DDE transposase and undergoes cell-to cell transfer when complemented with an autonomous ICE, Tn916. In this case, a cryptic origin of transfer is located within the 3’end of the resistance gene. These nonautonomous ICEs have been called IMEs [integrative mobilisable elements] or cis-mobilisable elements [CIMEs].

More recent studies have identified a new ICE family, transposon of Group B Streptococcus (TnGBS), in which the enzyme catalysing their integration and excision belongs to another DDE-group Tpase. Indeed, this has led to the identification of an entirely new family of classic ISs carrying DDE Tpases, the ISLre2 family). In addition to the TnGBS family, other ICEs have been identified which include a DDE Tpase closely related to that of the IS30family. It seems likely that examples of ICEs with other IS family Tpases are awaiting identification. Moreover, in addition to a variety of transfer functions, certain ICEs carry plasmid-related replication genes important in ensuring sufficient stability of the transposition intermediates to enable their subsequent integration. Indeed, early examples of ICEs were initially thought to be resistance plasmids and assigned an incompatibility group, incJ. These are maintained as an integrated copy in the host chromosome but can nevertheless give rise to circular copies. This is yet another example of the increasingly indistinct frontiers betweenphage, plasmids and transposons

**IS91-related ISCRs**

These MGE include a putative transposase of the HUH family similar to that of the IS91 family, the so-called ‘common region’ (CR). Although it has yet to be demonstrated that such structures indeed transpose, ISCRs are associated with multiple flanking antibiotic resistance genes. It is thought that these genes are transmitted during the rolling circle type of transposition mechanism postulated to occur in IS91 transposition. This involves an initiation event at one IS end, polarised transfer of the IS strand into a target molecule and termination at the second end. Flanking gene acquisition is thought to occur when the termination mechanism fails and rolling circle transposition extends into neighbouring DNA where it may encounter a second surrogate end. This type of mobile element may prove to play an important role in the assembly and transmission of multiple antibiotic resistance.

**Transposases and mechanism**

Transposition requires a set of DNA cleavages at the ends of a TE and a set of strand transfer reactions which move these ends into a suitable target DNA molecule catalysed by the TE-encoded Tpase. Transposases are often multidomain proteins. In addition to catalysis, they must also recognise specific DNA sequences at the IS ends and engage with the DNA target to form multimeric nucleoprotein assemblies, transpososomes. These nucleoprotein assemblies provide a precise architecture within which the chemical steps of transposition are carried out. They are composed of two or more Tpase monomers and, in some cases, accessory proteins such as DNA-architectural proteins (for example IHF, HU and HNS; or possibly protein chaperones [e.g. GroEL in the case of IS1]. They are dynamic and undergo conformational changes to coordinate DNA cleavages and strand transfers and ensure that, once started, transposition goes to completion.

**DDE transposases**

Many of the presently identified Tpases are members of the DDE family. These are structurally and catalytically related to RNaseH and other nucleic acid processing enzymes and are said to have an RNaseH fold. The highly conserved DDE triad serves to coordinate one or two divalent cations such as Mg2+ which in turn assist polarisation of the phosphate group belonging to the target phosphodiester bond, facilitating cleavage. DDE enzymes use hydroxyl groups as nucleophiles for cleavage and for strand transfer: H2O for initial cleavage and a 3’OH, generated by cleavage of the TE ends, for the strand transfer reaction. The DDE triad is often followed by a basic amino acid residue, generally K or R, 7 amino acids downstream and on the same face of the a-helix that also carries the final conserved E residue. Differences in the DDE spacing as well as the presence or absence of specific submotifs are used in distinguishing different groups and families. The earliest DDE motifs from transposases, those of the IS3 family, were identified by their similarities with retroviral integrases (IN) and, like IN, have a D35E spacing. Other DDE-group enzymes exhibit a spacing of 33–35 residues between the conserved D and E.

However, the DDE transposases of several families carry relatively long distinctive insertion domains of either a-helical orb-strand. Certain IS groups appear to carry variants of the conserved DDE motif with N or H residues replacing the E. The role of these alternative residues in catalysis has yet to be tested. Some groups, for example the IS1595family, include several of these variant motifs and the fact that these IS exist in several copies suggests that they are active.

DDE enzymes catalyse cleavage of only one DNA strand generally generating a 3’OH at the IS end. This is known as the transferred strand because the 3’OH is used as a nucleophile in the integration step to attack the target phophodiester bond and complete strand transfer. However, transposition of these ISs occurs via double strand DNA intermediates and therefore requires processing of the second strand (called the non transferred strand) to liberate the IS from flanking DNA sequences in the donor molecule. This can occur in several different ways and is also IS family-specific and serves to reinforce groupings derived from bioinformatics comparison by providing a mechanistic coherence. Thus, for ISs of the IS4family (IS50,IS10), the initial 3’OH is used to attack the opposite strand forming a transient hairpin bridge at the IS end. This is then cleaved using H2O as a nucleophile to liberate the IS and regenerate the 3’OH on the transferred strand ready for strand transfer and integration. This is known as a cut-and-paste mechanism.

Other ISs such as the IS630family, related to the eukaryote mariner/Tc superfamily, employ a mechanism in which the initial H2O-catalysed cleavage of the non-transferred strand occurs with a small offset of 2 bases into the IS prior to cleavage of the transferred strand as do their eukaryotic cousins.

An additional mechanism adopted by certain elements with DDE Tpases is cointegrate formation. Here, the transposon inserts into a target replicon in a process accompanied by TE replication. This results in fusion of the donor and target replicons with a directly repeated TE copy at each junction known as a Shapiro intermediate. Tn3 and IS6 family members transpose using this pathway as does bacteriophage Mu and its relatives.

However, by far the most common mechanism is the so-called copy–paste mechanism, which generates a transient double-strand circular DNA intermediate. This has been adopted by a significant number of IS families including IS3,IS30,IS110,IS256, ISLre2and possibly others. IS1uses this as one of several transposition pathways. For IS3family members, circle formation occurs in an asymmetric manner. One IS end is cleaved to generate the characteristic 3’OH of the transferred strand. This then serves to attack several nucleotides exterior to the second end to generate a single-strand bridge leaving a free 3’OH on the IS flank. The 3’OH can act as a replication primer.

IS replication would regenerate an intact copy reconstituting the donor plasmid and produce a double-strand circular DNA intermediate. Due to low basal Tpase levels, this initial step may occur in a stochastic manner. However, formation of the circular intermediate results in the assembly of a transient strong promoter composed of a -35 promoter element in the right IS end oriented outwards and a -10 promoter element in the left end oriented inwards. This promoter serves to drive transposase synthesis and consequent integration and disassembly of the promoter. Thus, the circular intermediate once generated is committed to terminate transposition.