Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Known knowns, known unknowns and unknown unknowns in prokaryotic transposition

Patricia Siguier[1], Edith Gourbeyre[1] and Michael Chandler[1,2]

Although the phenomenon of transposition has been known for over 60 years, its overarching importance in modifying and streamlining genomes took some time to recognize. In spite of a robust understanding of transposition of some TE, there remain a number of important TE groups with potential high genome impact and unknown transposition mechanisms and yet others, only recently identified by bioinformatics, yet to be formally confirmed as mobile. Here, we point to some areas of limited understanding concerning well established important TE groups with DDE Tpases, to address central gaps in our knowledge of characterised Tn with other types of Tpases and finally, to highlight new potentially mobile DNA species. It is not exhaustive. Examples have been chosen to provide encouragement in the continued exploration of the considerable prokaryotic mobilome especially in light of the current threat to public health posed by the spread of multiple Ab^R.

Addresses
[1] Centre National de la Recherche Scientifique (CNRS), Toulouse, France
[2] Department of Biochem., Mol. and Cell. Biol. Georgetown University Medical Center, 3900 Reservoir Rd., Washington, DC 20057-1455, USA

Corresponding author: Chandler, Michael (Mike.Chandler@ibcg.biotoul.fr)

Current Opinion in Microbiology 2017, **38**:171–180

This review comes from a themed issue on **MGE-HGT in prokaryotes**

Edited by **Andrew Lang, J Thomas Beatty, Phoebe Rice, Robin May** and **Gordon Brown**

http://dx.doi.org/10.1016/j.mib.2017.06.005

Introduction: known knowns

Understanding transposable genetic elements (TE), the processes governing their activities and their effects on the genomes they inhabit have progressed significantly since their initial identification in the late 1940s. Originally recognized as a cause of genetic instability and mutation in plants and called 'controlling elements' [1], subsequent use of bacterial models, demonstrated that their effects were due to movement of specific DNA segments from one place (donor site) to another (target site) in their host genomes catalyzed by element-specific enzymes: transposases (Tpase). The DNA segments were called transposons (Tn) [2] if they carried phenotypically detectable passenger genes such as antibiotic resistance (Ab^R), or insertion sequences (IS) if they did not [3,4].

Bioinformatic, biochemical and structural studies revealed that many TE share one particular type of Tpase structurally and catalytically related to RNaseH (see [5•]): DDE Tpases with an Asp, Asp, Glu catalytic amino acid triad. This provided a common mechanistic landscape for elements as diverse as bacterial Tn and IS, retroviruses and other eukaryotic DNA TE [6,7]. Most TE with DDE Tpases are flanked by short terminal inverted repeats (IR) that serve as transposase binding sites, and generate short direct target repeats (DR) on insertion. DDE Tpases use hydrolysis to generate a 3′OH group at the transposon ends, then catalyze the attack by that 3′OH on a new target phosphodiester bond, creating a new DNA connectivity [8,9]. Despite sharing this common catalytic core, DDE Tpases vary dramatically in size due to a wide variety of sequence-specific DNA binding domains and other accessory domains. Furthermore, differences in exactly which phosphodiester bonds are attacked can lead to large differences in the overall transposition pathway. Structural and mechanistic studies have provided an intimate understanding of transposition of several such TE (see chapters in [10•]).

Although DDE Tpases are the most widely known and arguably the best understood, other TE use different Tpase types. These include HUH enzymes (His-hydrophobe-His amino acid triad) which use a catalytic tyrosine (Y) for phosphodiester bond breakage and form a transient 5′-phosphotyrosine covalent enzyme-substrate intermediate. Structures and extensive biochemistry are also available for this large family of single-strand transposons which include the IS*200*/IS*605* and IS*91* families [11,12]. Other Tpases share properties with the serine (S) and tyrosine (Y) site-specific recombinases which also generate transitory covalent enzyme-substrate intermediates (3′-phosphotyrosine and 5′-phosophoserine respectively). Their roles in transposition have yet to be addressed in detail [10•]).

In spite of a robust understanding of transposition of some TE with DDE or HUH Tpases, there remain a number of important TE groups with potential high genome impact and unknown transposition mechanisms and yet others, only recently identified by bioinformatics, yet to be formally confirmed as mobile.

Here, we aim to point out some areas of limited understanding concerning extensively studied TE groups with

DDE Tpases, to address central gaps in our knowledge of characterised Tn with other types of Tpases and finally, to highlight new potentially mobile DNA species. This review is not exhaustive. Examples have been chosen to provide encouragement in the continued exploration of the considerable prokaryotic mobilome (the ensemble of mobile genetic elements in a genome or in a bacterial population) especially in light of the current threat to public health posed by the spread of multiple Ab^R.

Known unknowns: holes in our knowledge of mechanism in traditional DDE transposons

Tn3 family

The important and diverse Tn*3* family plays key roles in sequestering and transmitting many different passenger gene types from entire operons (e.g. Hg^R) to individual genes involved in Ab^R or virulence (e.g. [13]). They often carry integron recombination platforms and recruit integron cassettes with additional resistance genes [14]. In spite of their prevalence, especially in bacterial plasmids, their exact transposition mechanism is only just being unravelled at the molecular level. Tn*3*, among the earliest transposons identified [2], uses a replicative mechanism fusing donor and target replicons (cointegrates) with a directly repeated Tn copy at each of the two-donor/target junctions (Figure 1a). Members have characteristic long conserved IR and generate 5 bp DR on insertion. They carry dedicated site-specific recombination systems to drive exchange between short specific DNA sequences (*res*) on each Tn copy and resolve the cointegrate to complete transposition products ([15•] for review). A number of different resolution systems are associated with different family members and several are extremely well characterised [10•].

Studying Tn*3* family transposition has proved problematic due to their particularly long Tpases, TnpA (950-1025 aa), their replicative transposition mechanism and a phenomenon called immunity where a resident Tn copy inhibits insertion of a second copy in its vicinity. Immunity is also observed in two well understood transposition paradigms, phage Mu and Tn7. These encode a second essential protein with DNA dependent ATPase activity in addition to the Tpase whereas Tn*3* family members do not (chapters in [10•]).

There has been some recent progress in understanding Tn*3* transposition using Tn*4430* as a model [16,17]. Robust *in vitro* IR strand cleavage and transfer have been described using a TnpA mutant with reduced immunity. Wildtype TnpA appears less active suggesting that the 'immunity' mutations in some way 'unlock' TnpA. Moreover it appears that that single Tpase molecule can simultaneously bind two DNA ends.

Many important questions remain unanswered including: the nature of immunity, the structure and organization of the transpososome (a complex of the transposase and transposon ends), how replication enzymes are recruited, and whether there is interaction with extant host replication forks.

IS*6* family

Another important DDE IS family, IS*6*, is intimately involved in acquisition rearrangement and transmission of Ab^R genes. Family members IS*257* (aka IS*431*) and IS*26* play particularly active roles in plasmid plasticity (examples [16,17]), and occasionally of pathogenicity islands through Hfr formation [18].

These IS generally are ~800 bp long with Tpase identity levels of 40–94%, short related (14–20 bp) terminal IR and generally 8 bp DR [19]. Members invariably form compound transposons with directly repeated copies of the IS element flank intervening genes (Figure 1c). This arrangement is the same as that of the cointegrate intermediates of Tn*3*-like replicative transposition (Figure 1a) [20–24]. Unlike Tn*3*-related elements, no specific resolvase has been identified and cointegrate resolution (Figure 1a) presumably occurs *via recA*-dependent recombination between the two IS. Indeed RecA is required for cointegrate resolution [20,24]. This can only be achieved if compound IS*6*-based transposons carry directly repeated flanking IS copies.

Recent *in vivo* results suggest that IS*6* transposition may be more complex than originally thought [25]. IS*26* intermolecular cointegrate formation increased significantly in a Tpase-dependent, recA-independent reaction if the target replicon also contained an IS*26* copy (Figure 1b). The resulting cointegrates did not contain an additional IS*26* copy as would be expected if replicative transposition had taken place (Figure 1a,b), suggesting that the phenomenon results from conservative recombination. The observed cointegrate is structurally equivalent to the recombination product between an IS*26* copy in donor and target plasmids. At least one copy of an intact Tpase gene was required.

Another interesting aspect of IS*26* transposition, with striking implications for dissemination of other genes, is its ability to generate so called 'transposable units' (TUs) from compound transposons [26,27]. These are DNA circles that carry one copy of neighboring or passenger DNA together with a single IS*26* copy and can integrate efficiently into a target DNA carrying an IS*26* copy, recreating the original compound transposon (Figure 1c).

The biochemical mechanisms of these rearrangements have not been investigated in detail. However, it is likely that they arise when the Tpase synapses the ends of two different IS copies — either copies found in different replicons for cointegrate formation, or within the same compound transposon for TU formation. Note that the