

Detailed mechanism for transposition by TnpA transposase involves DNA shape rather than direct protein-DNA recognition to generate an active nucleoprotein complex

P Shing Ho

Address: Department of Biochemistry and Molecular Biology, 1870 Campus Delivery, Colorado State University, Fort Collins, CO 80523-1870, USA

Email: shing.ho@colostate.edu

F1000 Biology Reports 2009, 1:37 (doi:10.3410/BI-37)

The electronic version of this article is the complete one and can be found at: <http://F1000.com/Reports/Biology/content/1/37>

Abstract

A series of single-crystal structures determined by Barabas and colleagues provides a detailed mechanism for how the TnpA transposase from *Helicobacter pylori* recognizes, cleaves, and integrates the IS200/IS605 class of transposable elements. An interesting aspect of the mechanism is that the transposase recognizes the transposon through the unique fold-back structure adopted by the sequences of the DNA components, rather than through direct protein-DNA interactions. This is an example of indirect readout that is reminiscent of how four-stranded junctions are recognized by recombination proteins, but is also analogous to ribonucleoproteins, in that the DNA facilitates formation of an active nucleic acid-protein complex.

Introduction and context

DNA transposons are ubiquitous elements that allow genetic fragments to be integrated into host genomes and in nature are used, for example, to confer drug resistance in bacteria. They also have potential use as the delivery vehicle of new or modified DNAs in gene therapies. The mechanism of transposition is multi-step, similar in many ways to recombination-type events, and can randomly insert a transposon into the host or show some degree of specificity in integration. One particularly interesting DNA element is the IS608 transposon (a member of the IS200/IS605 class of transposable elements) found in *Helicobacter pylori* [1] that lies close to genes that are associated with gastric epithelial invasion factors. This element is associated with antibiotic resistance and is inserted 3' of specific four- or five-nucleotide sequences in the host [1] by the TnpA transposase [2]. The question is, how can this relatively small 155-amino acid enzyme affect the cleavage and rejoining steps required to integrate the associated IS608 transposon in a specific manner into the host genome? The structures implicate unique sequence-dependent

DNA fold-back conformations as the primary determinants both in recognition and in the protein conformational change that switches the transposase into the active configuration.

Major recent advances

TnpA forms an obligate dimer [3] that provides for DNA cutting and splicing at the two ends of IS608 transposable elements [2]. Thus, unlike other transposases that have two catalytic sites within the same protein, the TnpA dimer creates two sites in *trans*. This would be sufficient to deal with the mechanism for integrating the transposon if the two ends of insertion were identical and single-stranded, but they are not. The IS608 transposon includes two imperfect inverted repeats (IPs) that are nearly identical but differentially located relative to their cleavage sites [3]. The left IP is 20 nucleotides from the left end (LE) cut site, whereas the right IP is only 10 nucleotides from the right end (RE) cut site. There must be an asymmetry in the system that would not obviously come from a symmetric dimer. Furthermore, earlier crystal structures from Ronning and

colleagues [3] of TnpA alone and in complex with an isolated DNA hairpin elucidated the dimer structure, but both showed an inactive catalytic site and thus provided no additional insights into the mechanism of action. The recent paper by Barabas and colleagues [4] presents five single-crystal structures of TnpA in the presence of the left and right IP elements that extend toward the respective LEs and REs of the transposon. The key finding is that the left and right IPs and LE and RE DNA elements form specific structures that not only define their recognition by TnpA, but also induce the catalytic sites of the enzyme to adopt their active conformations. Together, the series of structures and biochemical and biophysical studies allow the authors to propose a detailed mechanism for the complete transposition cycle.

The most intriguing structures in this series are the complexes between the TnpA dimer and the IP hairpins extended with the LE or RE DNA of the transposon. The IP element forms a hairpin with a single thymine bulge. The RE DNA, however, is not a dangling end but folds back into a specific loop structure through G·C·G and T·A·A base triplets, resulting in a topology that is similar in many respects to RNA topologies [5]. This end of the DNA specifically induces a conformational change in which a helix that contains the catalytic tyrosine residue (Y127) is remodeled from the inactive conformation seen in the previous structures [3] to an active form, where Y127 is now in position to serve as a nucleophile that can attack the DNA backbone. A similar, though not identical, fold-back structure is formed by the LE IP and the associated LE element, which has a similar effect of creating an active conformation in the enzyme. Thus, the DNA itself plays a major role in its own recognition – the fold-back structures of the IP and LE or RE elements are unique and are required to induce the protein to assume its active conformation. This allows the enzyme to work only on the ‘top’ strand of the transposon.

The authors raise the question of whether the DNA sequences, in themselves, can fold into the structures seen in the LE and RE complexes [4]. If so, this would be a dramatic example of the concept known as indirect readout [6-8], in which the specificity for protein recognition of a DNA sequence is not provided by direct protein-DNA contacts in the form of, for example, pairing hydrogen bond donor and acceptor groups between protein side chains and nucleotide base pairs, but through the effects of base sequence on the conformational shape of the DNA element. This is reminiscent of most RNA-binding proteins, which recognize the tertiary structure of the fold rather than the sequence of the single-stranded polynucleotide.

An exhaustive map of DNA sequences and their structures has shown that sequence can affect both the overall conformation and, more subtly, details of the double-helix [9]. In the latter case, there are now a sufficiently large number of structures of B-DNA in a self-consistent dataset to permit investigators to define sequences that adopt an ‘average’ conformation for this standard double-helix and sequences that deviate significantly from this average, including those of known transcriptional regulators. A recent study by Little and colleagues [10] demonstrates the concept of indirect readout at the double-helix level in convincing fashion. In this study, the authors show that the *Hinc III* endonuclease can still effect site-specific DNA cleavage even in a mutant in which all direct protein-DNA contacts have been eliminated.

The structures of the TnpA transposases by Barabas and colleagues [4], however, rely on larger effects of sequence on DNA structure, a role that is more analogous to the one that indirect readout is proposed to play in the sequence-dependent recognition of Holliday junctions during DNA recombination. Four-way DNA junctions have been shown to be stabilized explicitly by core (A/G) CC trinucleotides within perfect inverted repeat sequences [11] and this has been proposed as the means by which junction-resolving enzymes, such as T7 endonuclease I [12] and T4 endonuclease VII [13], show sequence specificity for their DNA substrates [14]. In the case of the TnpA transposase [4], the IS608 transposon ends form sequence-specific fold-back structures that are specific for the enzymes’ active form.

Future directions

The question, therefore, is one of the chicken or the egg: does the DNA fold to form a substrate that induces an active enzyme, or does the shape of the functional TnpA catalytic site induce the DNA shape? The current structures support the former scenario, but one must wait to see whether the LE and RE DNAs by themselves or, more specifically, in the context of the transposon, can fold to the conformation that is uniquely recognizable by the TnpA transposase, and it would be interesting to see whether this could influence the rate of integration. If so, this may have implications for the design of molecular systems for gene therapy applications.

Abbreviations

IP, imperfect inverted repeat; LE, left end; RE, right end.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

Studies from the author's laboratory were supported by grants from the National Institutes of Health (R01GM62957A) and the Oregon Medical Research Foundation.

References

1. Kersulyte D, Velapatiño B, Dailide G, Mukhopadhyay AK, Ito Y, Cahuayme L, Parkinson AJ, Gilman RH, Berg DE: **Transposable element ISHp608 of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity.** *J Bacteriol* 2002, **184**:992-1002.
2. Guynet C, Hickman AB, Barabas O, Dyda F, Chandler M, Ton-Hoang B: **Single-stranded transposition of IS608: in vitro reconstitution of a new transposition mechanism.** *Mol Cell* 2008, **29**:302-12.
3. Ronning DR, Guynet C, Ton-Hoang B, Perez ZN, Ghirlando R, Chandler M, Dyda F: **Active site sharing and subterminal hairpin recognition in a new class of DNA transposases.** *Mol Cell* 2005, **20**:143-54.

F1000 Factor 3.0 *Recommended*
 Evaluated by Dale Ramsden 06 Jan 2006
4. Barabas O, Ronning DR, Guynet C, Hickman AB, Ton-Hoang B, Chandler M, Dyda F: **Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection.** *Cell* 2008, **132**:208-20.
5. Hendrix DK, Brenner SE, Holbrook SR: **RNA structural motifs: building blocks of a modular macromolecule.** *Quart Rev Biophys* 2005, **38**:221-43.
6. Dickerson RE: **The DNA helix and how it is read.** *Sci Am* 1983, **249**:94-111.
7. Otwinowski Z, Schevitz RV, Zhang RG, Lawson CL, Joachimiak A, Marmorstein RQ, Luisi BF, Sigler PB: **Crystal structure of trp repressor/operator complex at atomic resolution.** *Nature* 1988, **335**:321-9.
8. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB: **DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.** *Proc Natl Acad Sci U S A* 1998, **95**:11163-8.
9. Hays FA, Teegarden A, Jones ZJR, Harms M, Raup D, Watson J, Cavaliere E, Ho PS: **How sequence defines structure: a crystallographic map of DNA structure and conformation.** *Proc Natl Acad Sci U S A* 2005, **102**:7157-62.
10. Little EJ, Babic AC, Horton NC: **Early interrogation and recognition of DNA sequence by indirect readout.** *Structure* 2008, **16**:1828-37.

F1000 Factor 6.0 *Must Read*
 Evaluated by Tom Tullius 06 Mar 2009
11. Eichman BF, Varagason JM, Mooers BHM, Ho PS: **The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions.** *Proc Natl Acad Sci U S A* 2000, **97**:3971-6.
12. Hadden JM, Declais AC, Carr SB, Lilley DM, Phillips SE: **The structural basis of Holliday junction resolution by T7 endonuclease I.** *Nature* 2007, **449**:621-4.

F1000 Factor 6.4 *Must Read*
 Evaluated by P Shing Ho 24 Sep 2007, Malcolm White 01 Nov 2007
13. Biertumpfel C, Yang W, Suck D: **Crystal structure of T4 endonuclease VII resolving a Holliday junction.** *Nature* 2007, **449**:616-20.

F1000 Factor 6.4 *Must Read*
 Evaluated by P Shing Ho 24 Sep 2007, Malcolm White 01 Nov 2007
14. Khuu P, Voth AR, Hays FA, Ho PS: **The stacked-X DNA Holliday junction and protein recognition.** *J Mol Recog* 2006, **19**:234-42.