

Three-Dimensional Structure of the Tn5 Synaptic Complex Transposition Intermediate

Douglas R. Davies, Igor Y. Goryshin, William S. Reznikoff, Ivan Rayment*

Genomic evolution has been profoundly influenced by DNA transposition, a process whereby defined DNA segments move freely about the genome. Transposition is mediated by transposases, and similar events are catalyzed by retroviral integrases such as human immunodeficiency virus-1 (HIV-1) integrase. Understanding how these proteins interact with DNA is central to understanding the molecular basis of transposition. We report the three-dimensional structure of prokaryotic Tn5 transposase complexed with Tn5 transposon end DNA determined to 2.3 angstrom resolution. The molecular assembly is dimeric, where each double-stranded DNA molecule is bound by both protein subunits, orienting the transposon ends into the active sites. This structure provides a molecular framework for understanding many aspects of transposition, including the binding of transposon end DNA by one subunit and cleavage by a second, cleavage of two strands of DNA by a single active site via a hairpin intermediate, and strand transfer into target DNA.

DNA transposition is a process central to the evolution of genomes and is a critical component of the life cycle of retroviruses such as HIV-1. Transposable elements were first discovered in maize through a series of elegant genetic studies by Barbara McClintock (1). One class of transposable elements is a DNA sequence that has the capacity, in the presence of the transposase protein specific for its end sequences, of moving (transposing) from one site in the genome (donor DNA) to a second site (target DNA). DNA transposable elements include simple insertion sequences, transposons and some bacteriophages in eubacteria, and similar elements in archeobacteria and eukaryotes. They have the potential to remodel genomes and to facilitate the lateral transmission of genetic information such as antibiotic resistance determinants. A related class of mobile genetic elements, retroviruses (such as HIV-1) and retrotransposons containing long terminal repeats (LTRs), insert themselves into target DNA through a mechanistically similar process, although the proteins in these cases are called integrases (2-4). Despite their different names, the catalytic core domains of five different transposases and integrases exhibit remarkable similarity where each uses an active site that contains three conserved acidic residues (the DDE motif) that bind magnesium (2-7). Thus, study of the structure and function of one transposase or integrase provides a

greater understanding of all transposases and integrases.

Transposition is a complex multistep process (2-4, 8), as illustrated for Tn5 in Fig. 1. The first steps involve transposase binding to specific terminal DNA sequences and transposase-DNA oligomerization to form a synaptic complex. The catalysis of strand cleavage by transposase can only occur in the context of a synaptic complex. In all cases, catalysis involves nicking of the DNA to generate nucleophilic 3' OH groups on both strands. For some elements (for example, replicative transposable elements, such as Mu and Tn3, and retroviruses and LTR-containing retrotransposons), the 5' strands are not cut before strand transfer of the 3' ends into the target DNA. Rather, the 5' ends are resolved by a replication/recombination process or through processing of the protruding ends after transposition. For "cut and paste" elements (such as Tn5, Tn7, and Tn10), the 5' ends are also cleaved within the synaptic complex, releasing the transposable element from donor DNA. The 5' strand cleavage has been shown for Tn5 and Tn10 transposons to occur via a fascinating two-step process whereby the 3' OH generated from the initial strand cleavage step attacks the 5' strand to form a hairpin, followed by cleavage of the hairpin by attack from an activated water molecule (Fig. 1) (9, 10). The final steps of transposition involve target DNA capture, where target DNA becomes bound to the synaptic complex, followed by strand transfer, where the 3' OH groups of the transposable element perform nucleophilic attack on both strands of the target DNA. Attack of 3' OH groups on the target DNA occurs with a

staggered spacing between insertion sites that is specific for each transposable element. For Tn5 transposase, this spacing is 9 base pairs (bp).

Interestingly, the process in humans that shuffles immunoglobulin and T cell receptor genes in order to generate immune system diversity, Rag1 and Rag2 mediated V(D)J joining, also involves critical DDE residues in Rag1 (11-13) and a DNA hairpin intermediate (14). However, the orientation of hairpin formation differs in V(D)J joining. Even so, this suggests that DNA transposition is mechanistically and evolutionarily related to V(D)J joining and perhaps other cellular DNA rearrangement processes (15, 16).

Over the last 25 years, remarkable progress has been made in the molecular analysis of transposition through the use of genetics and biochemistry. Until now, struc-

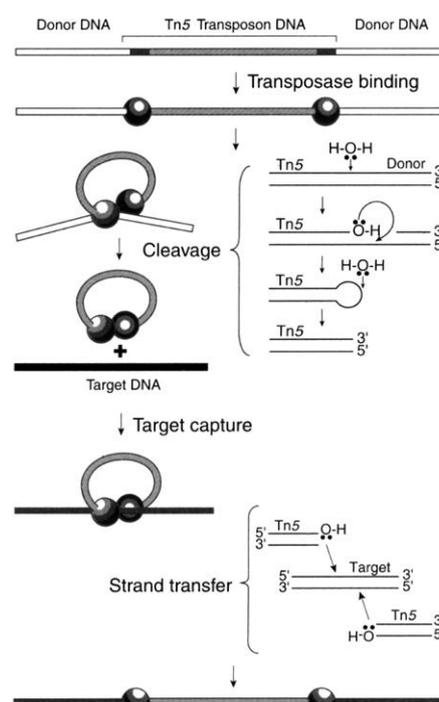


Fig. 1. Schematic diagram of the Tn5 "cut and paste" transposition mechanism (8, 23). In the first step, transposase (shaded spheres) binds to specific 19-bp recognition sequences (black bars) at the ends of the transposable element. Next, transposase dimerizes to form a catalytically active synaptic complex (29). In the first reaction, an activated water molecule performs a nucleophilic attack, hydrolyzing one strand of the DNA, which exposes a 3' OH group at the end of the transposon. This 3' OH is then activated to perform a nucleophilic attack on the opposite strand of DNA, forming a hairpin structure and excising the transposon from donor DNA (10). Next, hydrolysis of the hairpin results in blunt-ended DNA at the transposon end. Thereafter the synaptic complex binds to target DNA. In the final chemical step, activated 3' OH groups at the ends of the transposon perform nucleophilic attacks on target DNA, accomplishing strand transfer.

Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA.

*To whom correspondence may be addressed. E-mail: ivan_rayment@biochem.wisc.edu

tural studies have focused on isolated domains rather than intact proteins. This approach first yielded the three-dimensional

structures of the catalytic core domains of Mu transposase, and of the HIV-1 and avian sarcoma virus (ASV) integrases (5, 17, 18).

More recently the structures of two larger molecular fragments have been reported for Tn5 transposase (6) and rous sarcoma virus (RSV) integrase (7).

In comparison to the catalytic cores, less is known about the nucleic-acid binding domains and the manner in which they interact with DNA; however, structural studies of isolated DNA-binding domains have resulted in structures of the NH₂- and COOH-terminal domains of HIV-1 integrase (19, 20). A model of a Mu transposase DNA-binding subdomain with DNA has been proposed (21) and a co-crystal structure of a Tc3 transposase helix-turn-helix domain bound to DNA is available (22).

Although the previous structural studies have provided great insight into the common active site architecture of all transposase/integrase enzymes, many important questions remain unanswered. How does an intact transposase bind to the ends of its specific DNA? What is the interaction between the DNA substrate and the active site? How are divalent metal ion cofactors bound to the active site in the presence of DNA? How does the structure of the synaptic complex coordinate the cleavage of both ends of the transposable element? To address some of these questions, a structural study of full-length Tn5 transposase bound to DNA was undertaken.

The transposition reaction with wild-type Tn5 is down-regulated in vivo. Wild-type Tn5 transposase is not found to make detectable levels of synaptic complexes in vitro. A key feature of Tn5 studies has been the use of hyperactive transposase mutations to develop an efficient in vitro transposition system (23). The hyperactive in vitro system is permitting the biochemical and structural dissection of the transposition process.

We report the structure of a transposase synaptic complex determined to 2.3 Å resolution. The structure consists of intact Tn5 transposase bound to Tn5 transposon ends. This structure represents a synaptic complex at the stage following cleavage from donor DNA.

Crystallization and structure determination. The overall structure of the Tn5 transposase/DNA complex is shown in Fig. 2A. The crystallographic asymmetric unit contains one subunit of transposase and one 20-bp DNA molecule representing the transposon end. Two of each of these components are assembled within the crystal lattice to form a dimeric molecular assembly that represents a synaptic complex. A single Mn²⁺ ion is observed in each of the active sites and is coordinated by two carboxylate-containing residues and the 3' OH of the DNA. The double-stranded DNA molecules are bound to the active sites in a conformation that is consistent with the requirements of strand transfer.

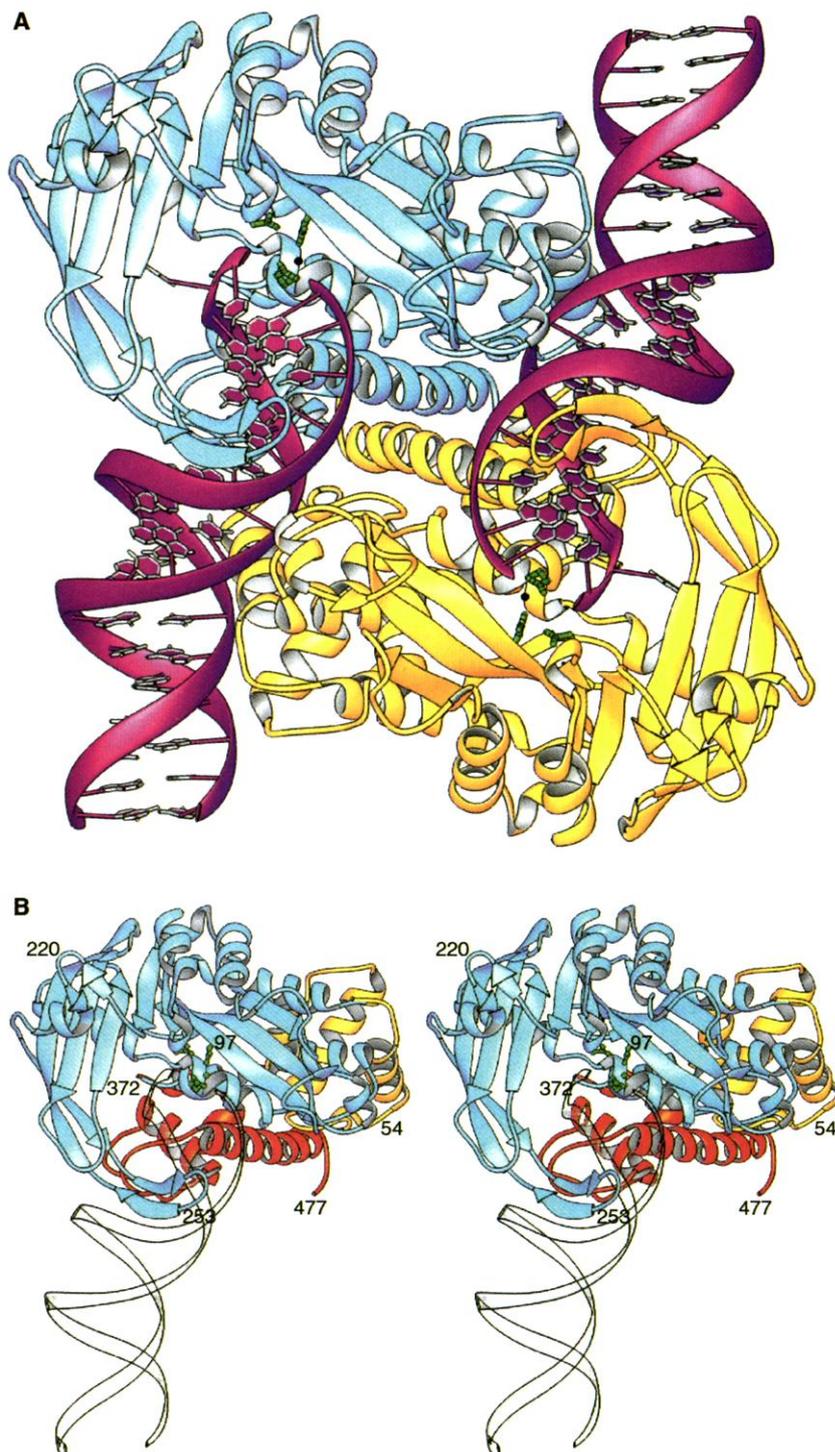


Fig. 2. The structure of the Tn5 transposase/DNA complex. (A) Ribbon representation of the transposase/DNA dimer viewed along a crystallographic twofold axis of symmetry. One protein subunit is colored yellow, the other is blue, and the two 20-bp DNA molecules are purple. The three catalytic residues are represented as green ball-and-stick structures, and the associated Mn²⁺ ion is black. (B) Stereoview of one monomer of transposase. The NH₂-terminal domain is yellow, the catalytic domain is blue, and the COOH-terminal domain is red. The active site residues Asp⁹⁷, Asp¹⁸⁸, and Glu³²⁶ and the associated Mn²⁺ ion are shown as green ball-and-stick structures. The backbone of a double-stranded DNA is represented by transparent ribbons. Figures 2, 3, 4, and 5 were generated with the program Ribbons (63).

RESEARCH ARTICLES

The protein/DNA complex described here consists of a hyperactive mutant version of transposase and a 20-bp DNA duplex. The hyperactive transposase is a triple mutant (E54K/M56A/L372P) developed for in vitro transposition reactions (23). The M56A mutation was introduced to prevent expression of the inhibitor protein and has no appreciable effect on transposition activity. The mutations E54K and L372P together dramatically enhance the transposition activity of Tn5. As described below, the structural basis for the phenotype of both of these mutants can be understood in terms of the current model. The DNA in the complex contains the specific 19-bp "outside end" (OE) recognition sequence of the Tn5 transposon. The number of base pairs beyond the OE and their identity were varied to obtain high quality crystals, following a strategy outlined in previous studies (24, 25).

The structure of the transposase/DNA complex was solved by multiple isomorphous replacement with the use of two heavy-atom derivatives and selenomethionine-labeled protein (26). A representative example of the electron density for the structure is shown in Fig. 3. Data collection and refinement statistics are listed in Table 1. All 20 bp of the DNA molecule are visible in the electron density map. The current model of the transposase protein begins with residue Ala⁵ and extends to the COOH-terminal residue, Gly⁴⁷⁷. Naturally occurring transposase is a 476 amino acid residue protein, but the expression and purification method used resulted in the addition of a glycine residue at the COOH-terminus. The electron density for the

model is well defined, except for a disordered loop between Pro³⁷² and Gln³⁹¹, and the side chains of the following residues, which are disordered beyond the β -carbon: His⁷, Lys¹²⁰, Asp¹⁵⁸, Lys²⁴⁹, Lys²⁷³, Glu³⁹⁴, and Lys⁴¹⁸.

Tertiary structure. The structure of Tn5

transposase may be divided into three major folding domains (Fig. 2B). The 477-amino acid protein consists of a 70-residue, NH₂-terminal domain, a central catalytic domain of ~300 residues, and a COOH-terminal domain of ~100 residues. The secondary structure of the NH₂-terminal domain is made up

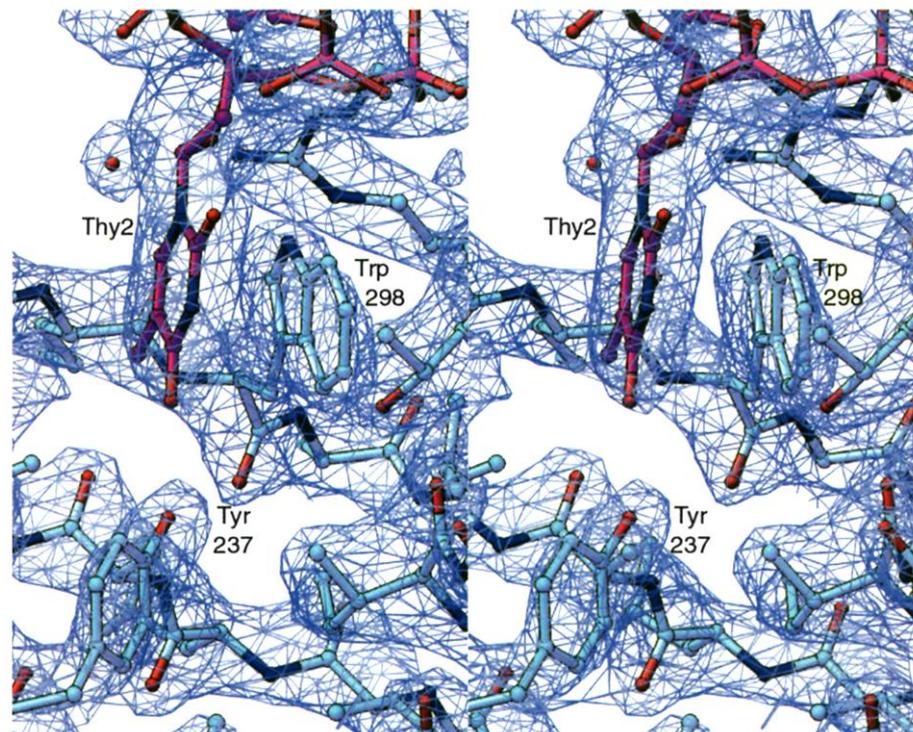


Fig. 3. Stereoview of the electron density associated with the thymine residue at position 2 of the nontransferred strand. The electron density was calculated with SIGMAA coefficients (64) and is displayed at 1.5 σ .

Table 1. Data collection and refinement statistics.

	Native	SeMet	ter(pyridine) PtCl	Uranyl acetate
<i>Data collection</i>				
λ (Å)	0.70090	0.97910	0.97630	0.72130
Space group	P6 ₃ 22	P6 ₃ 22	P6 ₃ 22	P6 ₃ 22
Unit cell dimensions a,b,c (Å)	113.7, 113.7, 228.1	114.2, 114.2, 228.9	113.6, 113.6, 228.5	113.5, 113.5, 227.5
Resolution (Å)	30–2.3	30–2.7	30–2.6	30–2.4
Mosaicity (°)	0.21	0.21	0.22	0.28
Unique reflections	39,476	25,009	27,278	34,575
Completeness (highest shell) (%)	99.6 (98.2)	99.9 (99.9)	97.3 (97.1)	99.7 (99.2)
Average I/σ^*	25.56	25.63	27.82	27.11
Redundancy	14.94	12.20	7.09	12.56
R_{sym} (%)†	5.3 (32.3)	7.9 (43.1)	6.0 (41.3)	6.0 (39.1)
Figure of merit	0.769			
<i>Refinement</i>				
Resolution range (Å)				30.0–2.3
Reflections (free)				37,502 (1,974)
$R_{\text{crystal}}^\ddagger$ (R_{free}^\ddagger)§ (percent)				22.3 (25.7)
rms deviations from ideality				
Bond lengths (Å)				0.012
Bond angles (degrees)				1.795
Planarity (trigonal) (Å)				0.002
Planarity (others) (Å)				0.010

* I/σ is the mean reflection intensity divided by the estimated error. † $R_{\text{sym}} = (\sum |I_{\text{hkl}} - \langle I \rangle|) / (\sum I_{\text{hkl}})$, where the average intensity $\langle I \rangle$ is taken over all symmetry equivalent measurements and I_{hkl} is the measured intensity for any given reflection. ‡ $R_{\text{crystal}} = \|F_o - F_c\| / F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. § R_{free} is equivalent to R_{crystal} but calculated for 5% of the reflections chosen at random and omitted from the refinement process.

for synopsis of two ends of a transposable element is a dimer of proteins as observed here. In contrast, studies of Mu transposase and the retroviral integrases have indicated that these proteins form synaptic complexes that include a tetramer of transposase/integrase proteins (33–35). Interestingly, comparison of the Tn5 transposase structure with the two-domain RSV integrase structure shows that an observed asymmetric dimer of RSV integrase occupies nearly the same volume as a Tn5 transposase monomer. Thus, it is likely that a dimer of retroviral integrase proteins is functionally equivalent to a Tn5 transposase monomer and that only one of the active sites in the integrase dimer participates in catalysis.

The OE DNA in the Tn5 synaptic complex structure is blunt ended, and position 1 has an exposed 3' OH that is bound to the active site. Thus, this complex corresponds to the synaptic complex that exists following donor cleavage and hairpin opening, but prior to strand transfer into target DNA. Due to the presence of a 3' OH at the end of the transferred strand, the complex can also be considered to be analogous to the preintegration complex of the retroviral integrases.

DNA conformation and protein contacts. Protein/DNA contacts are observed for all of the first 17 bp of the OE sequence, with sequence-specific protein/DNA contacts observed for 12 of these base pairs. DNA binding elements for Tn5 transposase are spread widely over nearly the entire primary sequence of the transposase protein (Fig. 4). Both transposase subunits in the dimeric synaptic complex interact with each DNA molecule.

The DNA bound to the transposase significantly differs from the conformation of standard B-DNA. The conformation of the DNA was analyzed with the program CURVES with the global parameters (36). The average helical twist of 36.86° (9.8 residues per turn) and the average rise per bp of 3.19 Å are representative of distorted B-DNA. There is a bend in the helical axis of approximately 41° at the T-A to A-T base step at positions 12 and 11 of the OE sequence. The bend in the helical axis is reflected by an increase in the roll and tilt angles, and in deviations of the major and minor groove widths and depths, compared with average B-DNA (37). Previous experiments with Tn5 have also shown the bending of DNA bound to transposase (38, 39); however, these experiments were performed with DNA that contained donor DNA sequences and indicated a bend at or near the cleavage site. Thus, the bend detected by these experiments is likely to be different from the bend observed in the crystal structure.

The nucleotide residues at positions 1 and 2 are not base-paired to one another. The phosphodiester backbone of the 5' end of the

nontransferred strand is distorted to form a hairpin-like conformation where the 5' oxygen of the nontransferred strand is ~ 3.5 Å from the 3' OH of the transferred strand (Figs. 4A and 5). The conformation of the backbone of the nontransferred strand orients the thymine residue at position 2 away from the interior of the double-stranded DNA molecule. The topic of the hairpin-like structure is discussed in more detail in the following section dealing with the active site.

The NH₂-terminal 70 residues of Tn5 transposase form an independent folding domain that binds DNA in *cis* (Fig. 4B). The NH₂-terminal domain is a compact helical bundle; however, the architecture of this domain does not match the topology of the helix-turn-helix (HTH) motif found in Cro repressor (40), HIV-1 integrase (19), or Tc3 transposase (22). Rather, the NH₂-terminal domain of Tn5 is formed from four α -helical segments, instead of the three found in the HTH domains. The fourth α -helix in Tn5 acts as the major “recognition helix” and makes several base pair-specific contacts in the major groove of the DNA from positions 7 to 13. Of particular interest is residue Lys⁵⁴ of the recognition helix, which is within hydrogen bonding distance of O4 of thymine 10 of the transferred strand. In wild-type Tn5 transposase, residue 54 is a glutamate. The E54K

mutant confers an OE-specific “hypertransposing” phenotype on Tn5. The structural basis for the E54K phenotype can be easily rationalized because the wild-type sequence would be likely to have an unfavorable charge-charge interaction between glutamate and the phosphate backbone. Conversely, the lysine mutant would increase the binding affinity for DNA through a specific contact. This interpretation is consistent with the hypothesis that suboptimal DNA binding by transposase is an important regulatory mechanism for Tn5 (31). Other *cis* binding interactions are provided by the NH₂-terminal end of helix 2, which contains three arginine residues (Arg²⁶, Arg²⁷, and Arg³⁰) that form salt bridges with phosphates in the DNA backbone. Three residues from the catalytic domain, Arg³⁴², Glu³⁴⁴, and Asn³⁴⁸, also contact DNA in *cis*.

Many of the *trans* protein/DNA interactions are provided by the interaction of a segment of anti-parallel β -sheet that protrudes from the catalytic domain between residues Ser²⁴⁰ and Lys²⁶⁰ (Fig. 4A). This extension proceeds well beyond the ends of any of the adjacent strands. In the Tn5 inhibitor protein structure, where no DNA is bound, this area of the protein is mostly disordered (6). In the presence of the OE sequence this secondary structural motif

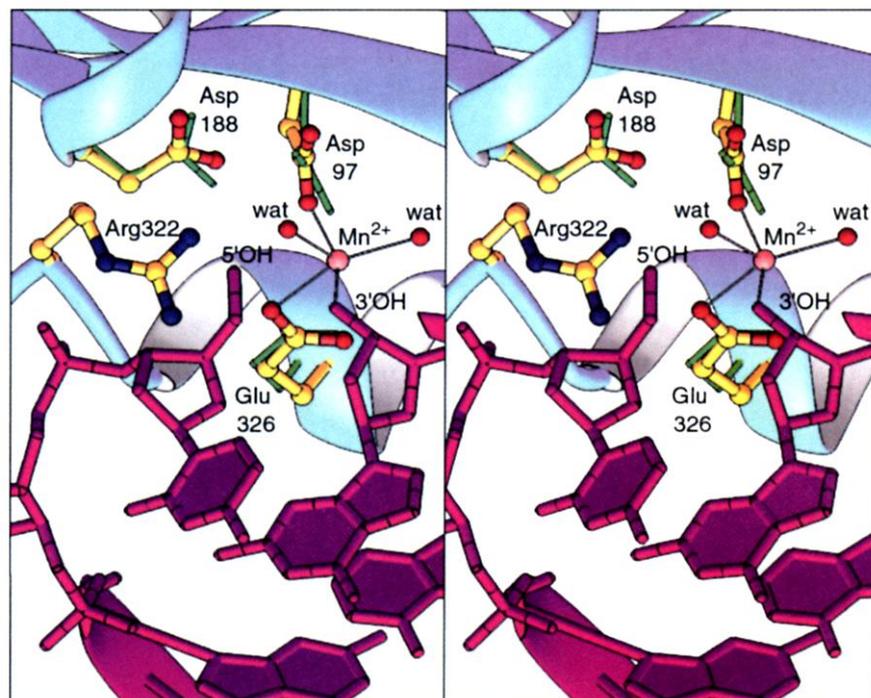


Fig. 5. Stereoview of the active site of Tn5 transposase. Transposase is shown as a blue ribbon structure with amino acid side chains as predominantly yellow ball-and-stick structures and the Mn²⁺ ion in pink. DNA is depicted in purple. Coordination of the Mn²⁺ ion with protein, DNA and water ligands is indicated by thin black lines. The active site residues and metal ions associated with an avian sarcoma virus integrase structure solved in the presence of Zn²⁺ (RCSB accession number 1VSH) are depicted in green, although the Zn²⁺ ions are not shown. The two protein models were overlapped with the program ALIGN (65, 66), where the rms deviation between active site residues is 0.55 Å for all atoms.

clamps down on the DNA close to the active site via the convex mode of major groove binding (41). Other significant *trans* protein/DNA contacts occur near the active site and their description is included in the discussion of the active site.

The active site. Transposases and integrases share a common catalytic triad of acidic residues; the DDE motif. The components of the active site are thought to form a binding site for magnesium that is required for catalysis. Mutational studies in Tn10 transposase (42) and Tn5 transposase (43) have shown that changing any one of these acidic residues to alanine abolishes the catalytic activity. Furthermore, experiments with active site cysteine mutants in Tn10 (44) have resulted in enzymes with an altered preference for manganese over magnesium. Together these experiments suggest that these residues are involved in divalent metal ion binding. Structural studies of the catalytic core domains of the retroviral integrases of HIV-1, ASV, and RSV, as well as those of Mu transposase and Tn5 transposase, show a common spatial arrangement of carboxylates that can potentially bind two metal ions: one located between the two aspartate residues and one between the first aspartate and the glutamate. The three residues are too far apart to form a single binding site. Several crystal structures of integrase core domains have been solved in the presence of metal ions. Typically, these structures show the binding of Mg²⁺ or Mn²⁺ only at the site located between the two aspartate residues (45, 46); however, crystals of the ASV integrase core soaked in Zn²⁺ and Cd²⁺ exhibit metal binding at both potential metal-binding sites (47).

In the Tn5 transposase/DNA complex, there is one Mn²⁺ ion in the active site. The Mn²⁺ ion in the active site is coordinated by Asp⁹⁷ and Glu³²⁶ and by the 3' OH of the transferred strand of DNA. The electron density map does not show binding of a second metal ion between Asp⁹⁷ and Asp¹⁸⁸. However, the conformations of all three residues bear a similarity to the conformations of the side chains of the ASV/Zn²⁺ structure. The corresponding residues in ASV integrase overlap with those of Tn5 transposase with an rms deviation of 0.55 Å for all atoms (Fig. 5). The Mn²⁺ ion in the Tn5 structure is within 0.5 Å of the Zn²⁺ ion in the ASV structure. The observed Mn²⁺ ion is in position to activate the 3' OH of the transferred strand of DNA for nucleophilic attack on target DNA.

It has been suggested that the transposase/integrase enzymes may employ a catalytic mechanism that uses two divalent metal ions in a way similar to that seen in the DNA polymerase enzymes (48). In the DNA polymerases, the active site contains two aspartic acid residues that coordinate two Mg²⁺ ions. One of the Mg²⁺ ions activates a 3' OH in

order to perform a nucleophilic attack on the α -phosphate of nucleotide triphosphate. The second metal ion coordinates the β - and γ -phosphates and serves to orient the nucleotide triphosphate in the active site and then to coordinate the pyrophosphate leaving group. It appears unlikely that the two metal hypothesis can be transferred directly to the transposase in light of the current structure. Indeed, the role of the conserved residue Asp¹⁸⁸ of the DDE motif cannot be defined by the present structure.

The conformation of DNA in the active site of Tn5 transposase is consistent with the hairpin mechanism for the cleavage of two strands of DNA by a single active site (9, 10). The hairpin hypothesis predicts that the nontransferred strand of DNA is cleaved by attack from the 3' OH of the nicked transferred strand to form an intermediate hairpin structure. The main structural requirement for the hairpin mechanism is that the scissile phosphodiester bond of the nontransferred strand must be close enough to the 3' OH of the transferred strand for transesterification to occur. The DNA used in crystallization lacks 5' phosphate groups. Nonetheless, in the Tn5 transposase structure, the 5' OH of the nontransferred strand is only 3.5 Å away from the activated 3' OH of the transferred strand (Figs. 4A and 5). Indeed, a phosphate moiety could be readily accommodated between the 3' and 5' hydroxyls with only a small rotation in the conformational angles of the 5' nucleotide.

The close approach of the 5' and 3' ends of the DNA appears to be facilitated by several protein/DNA contacts near the active site which form a binding pocket that stabilizes a bend in the backbone of the nontransferred strand. A thymine residue second from the 5' end of the nontransferred strand is rotated away from the interior of the DNA molecule such that the thymine base is held in a hydrophobic binding pocket and forms stacking interactions with Trp²⁹⁸ (Figs. 3 and 4A). This phenomenon of "base flipping" has been observed in protein/DNA co-crystal structures and has been proposed to be a common feature of enzymes that perform chemistry on DNA (49). The 5' phosphate group of thymine 2 is held in place by specific contacts with NH1 of Arg²¹⁰, OH of Tyr³¹⁹, and Ne and NH2 of Arg³²². Interestingly, Tyr³¹⁹ and Arg³²² in Tn5 transposase are part of a conserved sequence motif of Y(2)R(3)E(6)K in the IS4 family of transposases known as the YREK signature (50). For Tn5, the entire signature sequence consists of Tyr³¹⁹, Arg³²², Glu³²⁶, and Lys³³³, where Glu³²⁶ is a catalytic residue of the DDE motif. Mutational studies in Tn5 (43) and Tn10 (42) have shown that changes of the Tyr and Arg residues in the YREK signature have a deleterious effect on overall catalytic activity, with the Arg residue being far more important for catal-

ysis. These results emphasize the importance of stabilizing the hairpin conformation of the DNA for the mechanism of "cut and paste" transposases. Interestingly, the retroviral integrases and Mu transposase, which have not been shown to catalyze transposition via a hairpin intermediate, do not show sequence conservation of the Tyr and Arg residues and lack a binding pocket for a nucleotide base observed in Tn5 transposase.

Although Tn5 transposase and retroviral integrases differ in the number of cleavage steps in transposition, there are many similarities between the active sites of these enzymes. Besides the conserved cluster of acidic residues in the DDE motif, there is a pair of conserved basic residues near the active site. In the Tn5 transposase/DNA structure, the NH₂ group of Lys³³⁰ makes a specific contact on the minor groove side of the double helix with a cytosine residue and Lys³³³ is in contact with a phosphate group on the same strand of DNA. Structural and sequence alignment between Tn5 transposase and HIV-1 integrase show that these Lys residues are equivalent to Lys¹⁵⁶ and Lys¹⁵⁹ in HIV-1 integrase. Earlier crosslinking studies with HIV-1 integrase have shown that these two lysine residues are in contact with DNA (51). This comparison indicates a high probability that the binding of the transferred strand of DNA to the active site of Tn5 transposase may be very similar to the mode of binding of viral DNA to the active site of integrase. Recently, promising new drug leads were obtained by screening a library of chemical compounds against preformed complexes of HIV-1 integrase and viral DNA, suggesting that a synaptic complex of HIV-1 integrase may be a unique drug target (52). Because of the similarity of the catalytic active sites and overall catalytic core architecture common to all transposase/integrase enzymes, the Tn5 transposase synaptic complex structure may prove to be an important first step toward modeling the integrase/DNA interaction for the purpose of structure-based design of inhibitors of HIV-1 integrase as potential antiviral therapies for the treatment of AIDS.

Conformational differences. Comparison of the Tn5/DNA complex reported here with the earlier truncated Tn5 transposase structure (6) reveals major conformational differences in the protein. Both the truncated Tn5 protein structure and the transposase/DNA structure show very similar protein/protein interactions associated with the COOH-terminal domain across the molecular twofold axis. This interaction is dominated by symmetry-related α -helices that are in contact from Ser⁴⁵⁸ to Met⁴⁷⁰. A single point mutation in this interface, A466D, abolishes the inhibitory activity of the inhibitor protein and also catalytic activity of the transposase (53, 54), illustrating the importance of the COOH-terminal domains. At the time the structure of

is truncated (53), and binding of monomeric transposase to DNA is not observed unless the COOH-terminus is truncated (56). By disordering the helix connecting the catalytic and COOH-terminal domains, the L372P mutant may facilitate the conformational change required to remove inhibitory interactions between the NH₂- and COOH-terminal domains and allow the formation of a synaptic complex.

Implications for target DNA binding. As discussed above, the observed Tn5 transposase/DNA dimer is consistent with the strand transfer requirements of Tn5. Assuming that no major conformational changes occur between formation of the complex seen here and the beginning of strand transfer, the structure of the synaptic complex should allow for binding of target DNA and should have spacing between the two activated 3' OH groups consistent with the geometry of strand transfer. Examination of the synaptic complex structure clearly shows a potential binding site for target DNA. The two catalytic active sites and the activated 3' OH groups lie at the bottom of a continuous extended trench that is lined with several basic amino acid side chains: Arg¹⁰⁴, Arg¹⁵⁰, Lys¹⁶⁰, Lys¹⁶⁴, Arg¹⁸⁹, His¹⁹⁴, Arg²¹⁰, Lys²¹², His²¹³, Arg²¹⁵, Lys²¹⁶, Lys²⁴⁹, Arg²⁵⁰, Lys²⁵⁴, Lys²⁹⁷, and Lys³³³. Modeling with a segment of B-form DNA indicates that the trench is wide enough to accommodate a DNA duplex, if one allows for some movement of the loops that form the walls of this structure.

Tn5 catalyzes transfer of the 3' OH groups at each end of the transposable element into opposite strands of target DNA with a spacing of 9 bp between the two insertion sites. For standard B-DNA, the required spacing between scissile phosphodiester bonds would be approximately 35 Å. The spacing between 3' OH groups in the current model of the synaptic complex is nearly 41 Å. This discrepancy could be resolved either by conformational changes of the complex upon target DNA binding and/or by the observed preference of Tn5 transposase for negatively supercoiled target (23, 57). Supercoiled DNA has a decreased twist angle leading to an increased spacing between scissile bonds.

Conclusions. The Tn5 transposase/DNA structure provides a framework for understanding the molecular basis of Tn5 transposition: cleavage of DNA in *trans*, cleavage of the non-transferred strand of DNA via a hairpin intermediate, and strand transfer into target DNA with a 9-bp stagger between insertion sites. Comparison with the retroviral integrases suggests many similarities between the architecture of the active site and the binding of the transferred strand of DNA, which may prove useful in the modeling of retroviral preintegration complexes. The structure also provides a structural basis for the phenotype of the hyperactive mutations of Tn5. Although the Tn5 transposase/

DNA structure represents a significant step forward in understanding the structural basis of transposition, many questions remain. The current structure represents only one intermediate in a process involving four distinct reactions. A full understanding of the molecular basis of Tn5 transposase requires a knowledge of the structures of transposase in the absence of DNA, in the presence of donor DNA, of the hairpin intermediate, and of the strand transfer intermediate. Studies to achieve these goals are in progress.

References and Notes

1. B. McClintock, *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13 (1951).
2. K. Mizuuchi, *Annu. Rev. Biochem.* **61**, 1011 (1992).
3. ———, *Genes Cells* **2**, 1 (1997).
4. N. L. Craig, *Science* **270**, 253 (1995).
5. G. Bujacz *et al.*, *J. Mol. Biol.* **253**, 333 (1995).
6. D. R. Davies, L. M. Braam, W. S. Reznikoff, I. Rayment, *J. Biol. Chem.* **274**, 11904 (1999).
7. Z. N. Yang, T. C. Mueser, F. D. Bushman, C. C. Hyde, *J. Mol. Biol.* **296**, 535 (2000).
8. W. S. Reznikoff *et al.*, *Biochem. Biophys. Res. Commun.* **266**, 729 (1999).
9. A. K. Kennedy, A. Guhathakurta, N. Kleckner, D. B. Haniford, *Cell* **95**, 125 (1998).
10. A. Bhasin, I. Y. Goryshin, W. S. Reznikoff, *J. Biol. Chem.* **274**, 37021 (1999).
11. M. A. Landree, J. A. Wibbenmeyer, D. B. Roth, *Genes Dev.* **13**, 3059 (1999).
12. D. R. Kim, Y. Dai, C. L. Mundy, W. Yang, M. A. Oettinger, *Genes Dev.* **13**, 3070 (1999).
13. S. D. Fugmann, I. J. Villey, L. M. Ptaszek, D. G. Schatz, *Mol. Cell* **5**, 97 (2000).
14. J. F. McBlane *et al.*, *Cell* **83**, 387 (1995).
15. A. Agrawal, Q. M. Eastman, D. G. Schatz, *Nature* **394**, 744 (1998).
16. D. C. van Gent, K. Mizuuchi, M. Gellert, *Science* **271**, 1592 (1996).
17. P. Rice and K. Mizuuchi, *Cell* **82**, 209 (1995).
18. F. Dyda *et al.*, *Science* **266**, 1981 (1994).
19. M. Cai *et al.*, *Nature Struct. Biol.* **4**, 567 (1997).
20. A. P. Eijkelenboom *et al.*, *Proteins* **36**, 556 (1999).
21. S. Schumacher *et al.*, *EMBO J.* **16**, 7532 (1997).
22. G. van Pouderooyen, R. F. Ketting, A. Perrakis, R. H. Plasterk, T. K. Sixma, *EMBO J.* **16**, 6044 (1997).
23. I. Y. Goryshin and W. S. Reznikoff, *J. Biol. Chem.* **273**, 7367 (1998).
24. S. R. Jordan, T. V. Whitcombe, J. M. Berg, C. O. Pabo, *Science* **230**, 1383 (1985).
25. Tn5 transposase E54K/M56A/L372P was purified from the IMPACT intein expression system (New England Biolabs, Beverly, MA), as described (10), with the exception that sample buffer containing 100 mM hydroxylamine was used as the cleavage reagent. The purified protein consists of the 476-amino acid residue transposase triple mutant with an additional glycine residue at the COOH-terminus. DNA was purchased from Integrated DNA Technologies (Coralville, IA) on a 1- μ mol scale and was purified by the manufacturer using HPLC. Oligonucleotides were annealed and combined with purified Tn5 transposase at a molar ratio of protein:DNA of 1:1.5, and dialyzed against 0.3 M KCl, 20 mM HEPES (pH 7.7), and 2 mM EDTA. Dialyzed protein/DNA complex was concentrated to ~10 mg/ml for use in crystallization experiments. The length and sequence of the DNA were screened to produce crystals with optimal diffraction. Only one sequence gave crystals that diffracted beyond 3.5 Å resolution on a laboratory area detector. The DNA in the present structure consists of the OE sequence with an additional G-C base pair at the end distal from the scissile phosphodiester bonds. The sequence of the strand of DNA with the 3' OH coordinated to the active site is 5'-GACTTGTGTATAAGAGTCAG-3'. Transposase/DNA crystals were grown by micro batch at room temperature. Typically, 10 μ l of complex was combined with 9 to 11 μ l of a mother liquor containing 30% (w/v) PEG 400, 100 mM MES (pH 6.0), and 60 mM (NH₄)₂SO₄. Crystals appeared spontaneously and reached sizes in excess of 0.3 mm by 0.3 mm by 0.2 mm in 10 to 30 days. Crystals belong to the space group P6₃22, with unit cell parameters of a = b = 113.65 Å and c = 228.08 Å for the native crystal measured at -160°C. For preparation of heavy-atom derivatives and for data collection at cryogenic temperatures, crystals were transferred into a synthetic mother liquor/cryoprotectant solution containing 35% PEG 400, 40 mM (NH₄)₂SO₄, 0.4 M KCl, 50 mM MES (pH 6.0), and 10 mM MnCl₂. The addition of MnCl₂ to crystals improved the diffraction limit of native crystals on a laboratory rotating anode x-ray source from ~3.4 Å to better than 2.9 Å.
26. Selenomethionine-labeled protein was prepared by growing the expression strain of *Escherichia coli* in liquid minimal media with selenomethionine replacing methionine. Labeled protein was purified and crystallized in the same manner as the native transposase, except for the inclusion of 2 mM tris (2-carboxyethyl)-phosphine hydrochloride. A ter(pyridine)PtCl derivative containing two heavy atom sites was produced by soaking crystals in a 0.5 mM solution for 12 hours, and a uranyl acetate derivative containing three heavy-atom sites was produced from a 36-hour soak at 5 mM. Diffraction data were collected at the Structural Biology Center beamline at the Advanced Photon Source in Argonne, IL (Table 1). Data were integrated, scanned and reduced with the HKL software (58). Phases from multiple isomorphous replacement were calculated and improved by density modification using CNS, version 0.5 (59). The model was built into the resulting electron density map with the program TURBO-FRODO (60). Thereafter, the model was improved through cycles of manual model building and refinement by simulated annealing with cross-validated maximum likelihood (61, 62) with CNS, version 1.0.
27. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia, *J. Mol. Biol.* **247**, 536 (1995).
28. I. Y. Goryshin, data not shown.
29. A. Bhasin, I. Y. Goryshin, W. S. Reznikoff, *J. Mol. Biol.*, in press.
30. M. D. Weinreich, L. Mahnke-Braam, W. S. Reznikoff, *J. Mol. Biol.* **241**, 166 (1993).
31. M. Zhou and W. S. Reznikoff, *J. Mol. Biol.* **271**, 362 (1997).
32. T. A. Naumann and W. S. Reznikoff, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
33. T. A. Baker, M. Mizuuchi, H. Savilahti, K. Mizuuchi, *Cell* **74**, 723 (1993).
34. D. C. van Gent, C. Vink, A. A. Groeneger, R. H. Plasterk, *EMBO J.* **12**, 3261 (1993).
35. V. Ellison, J. Gerton, K. A. Vincent, P. O. Brown, *J. Biol. Chem.* **270**, 3320 (1995).
36. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 655 (1989).
37. E. Stofer and R. Lavery, *Biopolymers* **34**, 337 (1994).
38. D. York and W. S. Reznikoff, *Nucleic Acids Res.* **25**, 2153 (1997).
39. R. A. Jilk, D. York, W. S. Reznikoff, *J. Bacteriol.* **178**, 1671 (1996).
40. W. F. Anderson, Y. Takeda, H. Echols, B. W. Matthews, *J. Mol. Biol.* **130**, 507 (1979).
41. M. Tateno *et al.*, *Biopolymers* **44**, 335 (1997).
42. S. Bolland and N. Kleckner, *Cell* **84**, 223 (1996).
43. T. Naumann and W. S. Reznikoff, in preparation.
44. J. S. Allingham, P. A. Pribil, D. B. Haniford, *J. Mol. Biol.* **289**, 1195 (1999).
45. G. Bujacz *et al.*, *Structure* **14**, 89 (1996).
46. Y. Goldgur *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9150 (1998).
47. G. Bujacz *et al.*, *J. Biol. Chem.* **272**, 18161 (1997).
48. T. A. Steitz, S. J. Smerdon, J. Jäger, C. M. Joyce, *Science* **266**, 2022 (1994).
49. R. J. Roberts and X. Cheng, *Annu. Rev. Biochem.* **67**, 181 (1998).
50. R. Reznikoff, B. Hallet, J. Delcour, J. Mahillon, *Mol. Microbiol.* **9**, 1283 (1993).
51. T. M. Jenkins, D. Esposito, A. Engelman, R. Craigie, *EMBO J.* **16**, 6849 (1997).
52. D. J. Hazuda *et al.*, *Science* **287**, 646 (2000).
53. L. M. Braam, I. Y. Goryshin, W. S. Reznikoff, *J. Biol. Chem.* **274**, 86 (1999).

54. M. Steiniger-White and W. S. Reznikoff, *J. Biol. Chem.*, in press.
55. T. W. Wiegand and W. S. Reznikoff, *J. Bacteriol.* **174**, 1229 (1992).
56. D. York and W. S. Reznikoff, *Nucleic Acids Res.* **24**, 3790 (1996).
57. R. R. Isberg and M. Syvanen, *J. Biol. Chem.* **260**, 3645 (1985).
58. Z. Otwinowski and W. Minor, *Methods Enzymol.* **276**, 307 (1997).
59. A. T. Brunger *et al.*, *Acta Crystallogr. D* **54**, 905 (1998).
60. A. Roussel and C. Cambilau, in *Silicon Graphics Geometry Partners Directory* (Silicon Graphics, Mountain View, CA, 1991), vol. 86.
61. N. S. Pannu and R. J. Read, *Acta Crystallogr. A* **52**, 659 (1996).
62. P. D. Adams, N. S. Pannu, R. J. Read, A. T. Brünger, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5018 (1997).
63. M. Carson, *Methods Enzymol.* **277**, 493 (1997).
64. R. J. Read, *Acta Crystallogr. A* **42**, 140 (1986).
65. G. H. Cohen, *J. Mol. Biol.* **190**, 593 (1986).
66. ———, *J. Appl. Crystallogr.* **30**, 1160 (1997).
67. We thank L. Mahnke for providing the first samples of purified transposase, J. B. Thoden for help in collecting the x-ray data, and T. Naumann, A. Bhasin, M. Steiniger-White and R. Saecker for helpful discussions. We gratefully acknowledge the help of N. Duke, F. Rotella, and A. Joachimak at the Structural Biology Center Beamline, Argonne National Laboratory.

in collecting the data. This research was supported in part by NIH grants AR35186 (I.R.) and GM50692 (W.S.R.). W.S.R. is a recipient of a Vilas Associates Award and is the Evelyn Mercer Professor of Biochemistry and Molecular Biology. D.R.D. was supported by the NIH Biotechnology Training Grant. The use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Energy Research, Contract W-31-109-Eng-38. The PDB accession number for the coordinates and structure factors is 1F3I for the transposase/DNA complex.

26 April 2000; accepted 5 June 2000

fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size

Anne Frary,^{1*} T. Clint Nesbitt,^{1*} Amy Frary,^{1†}
Silvana Grandillo,^{1‡} Esther van der Knaap,¹ Bin Cong,¹
Jiping Liu,¹ Jaroslaw Meller,² Ron Elber,² Kevin B. Alpert,¹
Steven D. Tanksley^{1§}

Domestication of many plants has correlated with dramatic increases in fruit size. In tomato, one quantitative trait locus (QTL), *fw2.2*, was responsible for a large step in this process. When transformed into large-fruited cultivars, a cosmid derived from the *fw2.2* region of a small-fruited wild species reduced fruit size by the predicted amount and had the gene action expected for *fw2.2*. The cause of the QTL effect is a single gene, *ORFX*, that is expressed early in floral development, controls carpel cell number, and has a sequence suggesting structural similarity to the human oncogene *c-H-ras* p21. Alterations in fruit size, imparted by *fw2.2* alleles, are most likely due to changes in regulation rather than in the sequence and structure of the encoded protein.

In natural populations, most phenotypic variation is continuous and is effected by alleles at multiple loci. Although this quantitative variation fuels evolutionary change and has been exploited in the domestication and genetic improvement of plants and animals, the identification and isolation of the genes underlying this variation have been difficult.

Conspicuous and important quantitative traits in plant agriculture are associated with domestication (1). Dramatic, relatively rapid evolution of fruit size has accompanied the domestication of virtually all fruit-bearing crop species (2). For example, the progenitor of the domesticated tomato (*Lycopersicon esculen-*

tum) most likely had fruit less than 1 cm in diameter and only a few grams in weight (3). Such fruit was large enough to contain hundreds of seeds and yet small enough to be dispersed by small rodents or birds. In contrast, modern tomatoes can weigh as much as 1000 grams and can exceed 15 cm in diameter (Fig. 1A). Tomato fruit size is quantitatively controlled [for example, (4)]; however, the molecular basis of this transition has been unknown.

Most of the loci involved in the evolution and domestication of tomato from small berries to large fruit have been genetically mapped (5, 6). One of these QTLs, *fw2.2*, changes fruit weight by up to 30% and appears to have been responsible for a key transition during domestication: All wild *Lycopersicon* species examined thus far contain small-fruit alleles at this locus, whereas modern cultivars have large-fruit alleles (7). By applying a map-based approach, we have cloned and sequenced a 19-kb segment of DNA containing this QTL and have identified the gene responsible for the QTL effect.

Genetic complementation with *fw2.2*. A yeast artificial chromosome (YAC) containing *fw2.2* was isolated (8) and used to screen a

cDNA library (constructed from the small-fruited genotype, *L. pennellii* LA716). About 100 positive cDNA clones were identified that represent four unique transcripts (cDNA27, cDNA38, cDNA44, and cDNA70) that were derived from genes in the *fw2.2* YAC contig. A high-resolution map was created of the four transcripts on 3472 F₂ individuals derived from a cross between two nearly isogenic lines (NILs) differing for alleles at *fw2.2* (Fig. 2A) (8). The four cDNAs were then used to screen a cosmid library of *L. pennellii* genomic DNA (9). Four positive, nonoverlapping cosmids (cos50, cos62, cos69, and cos84) were identified, one corresponding to each unique transcript. These four cosmid clones were assembled into a physical contig of the *fw2.2* region (10) (Fig. 2B) and were used for genetic complementation analysis in transgenic plants.

The constructs (11) were transformed into two tomato cultivars, Mogeor (fresh market-type) and TA496 (processing-type) (12). Both tomato lines carry the partially recessive large-fruit allele of *fw2.2*. Because *fw2.2* is a QTL and the *L. pennellii* allele is only partially dominant, the primary transformants (R0), which are hemizygous for the transgene, were self-pollinated to obtain segregating R1 progeny. In plants containing the transgene (13), a statistically significant reduction in fruit weight indicated that the plants were carrying the small-fruit allele of *fw2.2* and that complementation had been achieved. This result was only observed in the R1 progeny of primary transformants fw71 and fw107, both of which carried cos50 (Fig. 1B and Table 1) (14). That the two complementing transformation events are independent and in different tomato lines (TA496 and Mogeor) indicates that the cos50 transgene functions similarly in different genetic backgrounds and genomic locations. Thus, the progeny of plants fw71 and fw107 show that *fw2.2* is contained within cos50.

Most QTL alleles are not fully dominant or recessive (5). The small-fruit *L. pennellii* allele for *fw2.2* is semidominant to the large-fruit *L. esculentum* allele (7). R2 progeny of fw71 were used to calculate the gene action [*d/a* = dominance deviation/additivity; calculated as described in (5)] of cos50 in the transgenic plants.

¹Department of Plant Breeding and Department of Plant Biology, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA. ²Department of Computer Science, Cornell University, Ithaca, NY 14853, USA.

*These authors contributed equally to this work.

†Present address: Department of Biological Sciences, Clapp Laboratory, Mount Holyoke College, South Hadley, MA 01075, USA.

‡Present address: Research Institute for Vegetable and Ornamental Plant Breeding, IMOF-CNR, Via Università 133, 80055 Portici, Italy.

§To whom correspondence should be addressed.