Retroviral integrases and their cousins

Phoebe Rice, Robert Craigie and David R Davies

The recently determined structures of the catalytic domains of HIV integrase, avian sarcoma virus integrase and the Mu transposase are strikingly similar to each other and also exhibit significant similarity to several nucleases. All these enzymes cut polynucleotides, leaving 3′OH and 5′PO_4 groups. The integrase and transposase also possess a strand-transfer activity that splices DNA. The structural similarities among members of this superfamily of polynucleotidyl transferases suggest that they share a similar mechanism of catalysis.

Address
Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Building 5, National Institutes of Health, Bethesda, MD 20892-0540, USA

Current Opinion in Structural Biology 1996, 6:76–83

Abbreviations

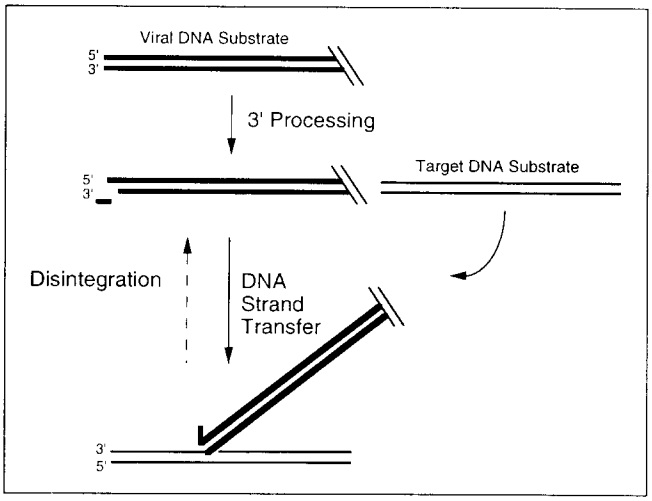
ASV	avian sarcoma virus
rms	root mean square
SH	Src homology

Introduction

Retroviruses are the causative agents of diverse diseases, including AIDS and leukemia in humans and a variety of cancers in animals. Despite the disparate consequences of retroviral infection, the steps in the replication cycles of all retroviruses are very similar [1–3]. In particular, the viral RNA genome is reverse transcribed within the host cell and the resulting double-stranded viral DNA is stably integrated into the host genome. DNA integration is required for the efficient expression of viral genes, resulting in the generation of progeny virions, and it also accounts for the ability of retroviruses to persist in infected cells. The process of DNA integration is essential for the replication of HIV, making it an attractive target in the search for inhibitors of this virus.

The first step in the retroviral integration reaction, 3′ processing, is the cleavage of two nucleotides from the 3′ ends of the viral DNA. The second step, DNA-strand transfer, is the covalent insertion of these 3′ ends into the host DNA (Fig. 1). These reactions are catalyzed by the virus-encoded integrase protein. The final steps of the integration process are the removal of the unpaired bases at the 5′ ends of the viral DNA and the repair of the single-strand gaps between the viral and host DNA. Purified integrase protein carries out both 3′ processing and DNA-strand transfer reactions *in vitro* with model DNA substrates [4–8]. By following the stereochemical course of 3′ processing and DNA-strand transfer with modified DNA substrates containing phosphorothioate linkages, it has been shown that each reaction is accompanied by an inversion of the chirality of the participating phosphorothioate [9]. This result demonstrates that both reactions proceed by a one-step transesterification mechanism that does not involve a covalent intermediate between integrase and the DNA substrate. The DNA cutting and joining reactions involved in retroviral DNA integration are essentially the same as those involved in DNA transposition [10,11]. For transposons, such as phage Mu, the enzyme that catalyzes these reactions is called a transposase.

Figure 1



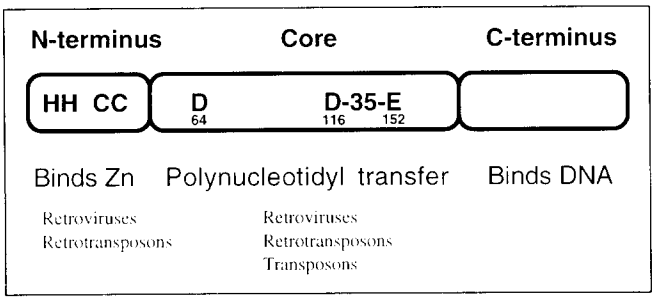
In vitro activities of retroviral integrases. The DNA substrate for 3′ processing consists of an oligonucleotide that matches one end of the viral DNA (thick black lines). Integrase cleaves two nucleotides from the 3′ end of the viral DNA substrate. Integrase also catalyzes the DNA-strand transfer reaction that inserts the cleaved viral DNA end into a target DNA (thin lines). DNA-strand transfer is a phosphoryl-transfer reaction using the 3′OH end of the viral DNA as the attacking nucleophile. The result is the covalent joining of the 3′ end of the viral DNA to the 5′ end of the target DNA at the site of insertion. Integrase can also catalyze an apparent reversal of the DNA-strand transfer reaction, known as disintegration.

Integrase

Integrase has three discrete functional domains (Fig. 2). The central core domain of HIV-1 integrase is relatively resistant to proteolytic digestion [12], suggesting that it also comprises a separate structural domain. The primary amino acid sequence of this region is highly conserved among both retroviral integrases and the transposase proteins of some mobile genetic elements. A triad of acidic residues within this domain, the D,D-35-E motif, is invariant among retroviral integrases [13–15], indicating that these residues may participate directly in catalysis. Although the core domain alone cannot catalyze 3′ processing or DNA-strand transfer, it can catalyze a

phosphoryl-transfer reaction known as disintegration [16], in which the viral DNA segment is first liberated from a branched DNA substrate that resembles the product of DNA-strand transfer and then the target DNA part of the substrate is resealed [17]. This finding [16] strengthened the notion that the core domain of integrase contains the catalytic site. Consistent with this view, mutation of any of the three acidic residues abolishes, or at least severely diminishes, all three catalytic activities in parallel [12,15,18–20].

Figure 2



The functional domains of retroviral integrases. The central core domain is highly conserved among retroviral integrases, retrotransposon integrases and some transposase proteins. It contains a triad of highly conserved acidic residues, the D,D-35-E motif, that are important for catalysis. The isolated core domain of HIV integrase catalyzes a polynucleotidyl transfer reaction termed disintegration, but both the N-terminal and C-terminal domains are required for the 3′ processing and DNA strand transfer reactions. The precise functional roles of the N- and C-terminal domains remain to be determined. The N-terminal domain is well conserved among retroviral and retrotransposon integrases, but similar sequences have not been identified in transposases. The C-terminal domain is less well conserved, even among the retroviral integrases. The C-terminal domain binds DNA non-specifically. The N-terminal domain contains a pair of highly conserved histidine and cysteine residues that resemble motifs that are frequently involved in coordinating metal ions. The numbers beneath the residues that comprise the D,D-35-E motif identify the position of these residues in HIV-1 integrase.

Although the core domain alone of HIV integrase can catalyze disintegration, the presence of the N- and C-terminal domains is required for 3′ processing and DNA-strand transfer. The functional roles of these domains have not yet been well defined. The C-terminal domain of HIV integrase binds DNA non-specifically [21–24]; this property is likely to be important for function because mutations that severely weaken DNA binding also abolish 3′ processing and DNA-strand transfer activities. However, it remains to be established whether this domain is involved in binding the viral DNA end sequences or the host sequences into which the viral DNA ends integrate. The functional role of the N-terminal domain is even less clearly defined. It contains a pair of conserved cysteine residues and a pair of conserved histidine residues that resemble motifs that coordinate metal ions, and are invariant among retroviral integrases. This domain of HIV integrase has been reported to bind zinc [16,25].

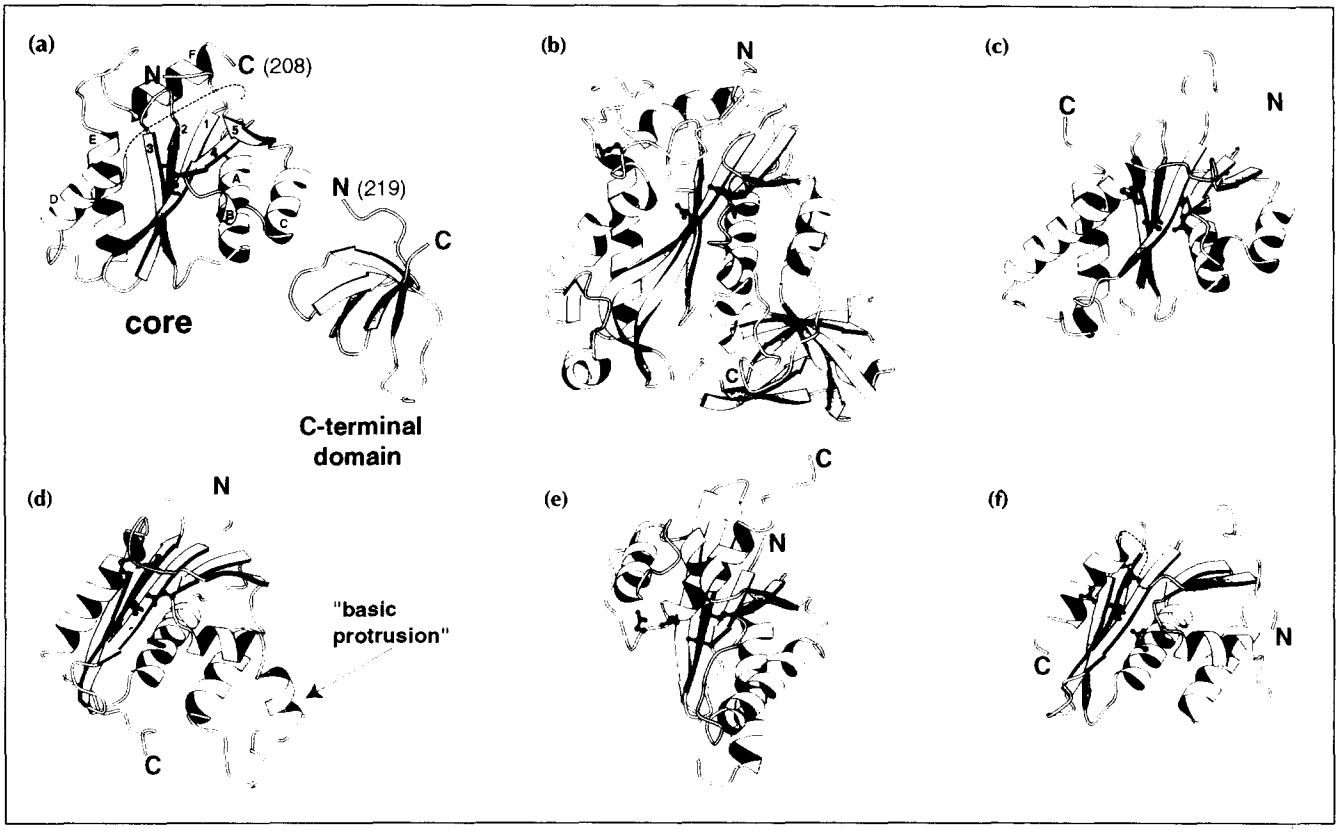
The active form of HIV-1 integrase that carries out both 3′ processing and DNA-strand transfer is multimeric, although the number of monomers in the active multimer is unknown. Evidence that multimerization of HIV-1 integrase is essential for 3′ processing and DNA-strand transfer activities comes from *in vitro* complementation experiments with integrase proteins that lack the N- or C-terminal domains [26,27]. These proteins are inactive individually, but wild-type levels of activity are restored when a pair of mutant proteins, one lacking the N terminus and the other lacking the C terminus, are both present in the reaction mixture. Similar mixing experiments, with one of the proteins carrying mutations in the active site residues, show that the functional N-terminal domain and the active catalytic domain must be located on different monomers for complementation to occur. This result implies that the N-terminal domain and the catalytic domains complement *in trans*. The avian sarcoma virus (ASV) integrase protein has also been reported to function as a multimer [28].

Structural studies of HIV integrase have been impeded by the poor solubility of the protein. However, despite this obstacle, structures have recently become available for the catalytic core domain of both HIV-1 integrase [29••] and ASV integrase [30••], and the solution structure of the C-terminal domain of HIV-1 integrase has also been determined [31••,32••]. Concurrently, the structure of the catalytic domain of the Mu transposase protein was solved [33••] and was found to be remarkably similar to the integrase structure, despite very little primary sequence similarity between the proteins. Both the integrase and the transposase structures also share considerable similarity with the structures of the nucleases RNase H and RuvC (Fig. 3). Although mechanistic similarities between retroviral integrases and transposases were apparent from earlier biochemical studies [10,11], the extent of the structural similarity would not have been easily predicted given the dissimilar primary sequences. The recent structural studies indicate that retroviral integrases belong to a superfamily of polynucleotidyl transferases that includes the Mu transposase and the nucleases RNase H and RuvC. Here we discuss the similarities and differences among the members of this superfamily of polynucleotidyl transferases and relate the structures of these enzymes to their biological roles.

The HIV-1 integrase catalytic core domain

Structural studies of HIV-1 integrase have been impeded by the low solubility of the protein. Intact HIV-1 integrase is quite insoluble, making it unsuitable for crystallization. Although the core domain is somewhat better behaved, it still aggregates in solution in the absence of detergent [34]. Crystallization of the core domain was only achieved after an extensive mutagenesis investigation in which

Figure 3



Ribbon representations of the structures of the polynucleotidyl transferase superfamily members. The disordered loop between β strand 5 and helix D in the HIV integrase structure, and the corresponding loop in HIV-1 RNase H, are represented by a dotted line. The amino acid residues that have been implicated in catalysis for each superfamily member are given in parentheses for each enzyme and are shown in ball-and-stick representation in the figure. **(a)** HIV-1 integrase (Asp64, Asp116 and Glu152 [present in the disordered loop that leads into helix D]). The C-terminal end of the ordered structure is residue 208. The N terminus of the C-terminal domain structure determined by NMR is residue 219. Strands are numbered 1–5 and helices lettered sequentially. **(b)** Mu transposase (Asp269, Asp336 and Glu392). **(c)** ASV integrase (Asp64, Asp121 and Glu157). **(d)** *E. coli* RNase H (Asp10, Glu48, Asp70 [shown in black], His124 and Asn130 [shown in grey]). **(e)** *E. coli* RuvC resolvase (Asp7, Glu66, Asp138 and Asp141 [shown in black]). **(f)** HIV-1 RNase H (Asp443, Glu478, Asp499 [shown in black], Asn545 and Asp549 [shown in grey]). The structures of the HIV-1 core domain, ASV core domain and the Mu transposase core domain are strikingly similar. The main differences are the presence of longer loops between the structural elements in the case of the Mu transposase core. The integrase and transposase cores share analogous structural elements and topology with RNase H. The RuvC resolvase, while sharing considerable structural similarity with the integrases and transposases, is the least similar member of this family. A cluster of acidic residues that are important for catalysis is located in analogous positions in each of the proteins, suggesting that they share a common catalytic mechanism. Figure generated using MOLSCRIPT [65].

hydrophobic residues were systematically mutated in an attempt to obtain soluble protein [35••]. The mutation of Phe185 to lysine resulted in an active protein that was a monodisperse dimer even at concentrations of 25 mg ml^{-1}. This mutation also served to stabilize the dimer.

The structure consists of a five-stranded β sheet surrounded by six helices (Fig. 3a). Asp64 and Asp116, two of the essential carboxylate residues that make up the D,D-35-E motif, lie on adjacent β strands (1 and 4), at a point where they splay apart to form a small cleft. The third residue, Glu152, lies in a disordered region (made up of residues 141–153) between β strand 5 and helix D, and in the vicinity of the active site. However, examination of the structure indicates that this glutamate residue cannot be very far from the two aspartates. Structures of several related enzymes show evidence of flexibility in this region (see below).

The HIV-1 integrase catalytic core domain is a dimer both in solution and in the crystal (Fig. 4a). The interface seen in the crystal buries a rather large area (2600 Å^2) of previously solvent-accessible surface. In this dimer, the two active sites are 35 Å apart, which presents us with a dilemma because the insertion of a pair of viral DNA ends into the target DNA occurs at two sites that are 5 base pairs, or about 16 Å, apart. However, a similar dimeric interaction is found in the ASV integrase core structure, lending credence to the view that the dimer interface in the crystal is also the dimer interface of the protein in solution. There are a number of ways of resolving this dilemma, of which the most likely is that the active form

of the intact integrase is a tetramer, as in the case of the Mu transposase.

The ASV integrase catalytic domain

The ASV integrase has very similar catalytic properties to the HIV integrase and the structure of the catalytic domain of this enzyme has been determined at 1.7 Å resolution [30••] (Fig. 3c). Although only 24% identical in sequence, the structures of the HIV and ASV cores can be superimposed with a root mean square (rms) deviation of only 1.4 Å for 107 Cα pairs. The region containing the active site glutamate residue, which was disordered in the HIV structure, has interpretable density in the ASV structure. The N-terminal end of helix D is extended, and the side chain of Glu157 extends from it towards the other two conserved carboxylates. It is interesting that this region is also highly mobile in the ASV structure, with thermal vibration B values of 70 $Å^2$ for residues 149 and 150. In the ASV structure, the region corresponding to helix F is disordered. This probably reflects the slightly greater truncation at the C terminus of the ASV fragment rather than an intrinsic difference between the proteins.

The ASV integrase core forms a dimer with the same general architecture as the HIV integrase dimer (Fig. 4a,b). However, the two do not exactly superpose, and if one subunit from ASV is aligned on a subunit of HIV, the corresponding second subunits show differences in the positions of the secondary structure elements of as much as 6 Å [30••]. The movement required to superimpose these second subunits would be a 13.3° rotation and a 4.8 Å translation. The dimer interface of ASV is smaller than that of HIV (766 $Å^2$ per monomer compared to 1300 $Å^2$), and Bujacz *et al.* quote data that suggest a weaker association in solution. This smaller interface is partly due to the lack of helix F, which forms part of the interface in the HIV integrase dimer. The position of the Phe185Lys mutation in HIV integrase corresponds to His198 in ASV integrase. The side chain of His198, like Lys185 in the HIV structure, forms a hydrogen bond with a main-chain carbonyl of the other subunit in the ASV integrase dimer. It may be argued, therefore, that the mutation is unlikely to have introduced any serious perturbation into the HIV integrase dimer structure.

The HIV integrase C-terminal DNA-binding domain

The isolated C-terminal domain of HIV integrase has the same non-specific DNA-binding activities as the full length integrase [21–24] and the minimal DNA-binding unit has been shown to comprise residues 220–270 [36]. The solution structure of this fragment has been determined by multidimensional heteronuclear NMR in two laboratories [31••,32••]. The monomer has five antiparallel β strands that fold into a β barrel (Fig. 3a). It shows remarkable similarity to the SH3 (Src homology 3) domain structure [37–39]. The Cα rms deviations are 1.4 Å for 41 residues of spectrin [31••] and 2.4 Å for the equivalent 45 residues of the biotin repressor BirA [38]. Although most SH3 domains are monomeric, the integrase SH3 domain is a dimer in solution, with a dissociation constant of $<10^{-6}$ M. The dimer interface is formed by predominantly hydrophobic contacts involving β strands 2, 3 and 4, with the two β sheets of the two subunits arranged antiparallel to each other. The catalytic domain dimer has a spacing of 21 Å between the two C-terminal Cαs, which may be compared with 17 Å between the two N-terminal Cαs of the C-terminal domain. The C-terminal domain can therefore be readily attached to the catalytic domains, although the relative orientations of the two domains cannot be determined because residues 213–220 are not present in either structure.

The dimer presents a saddle-shaped groove on one surface that could potentially accommodate double-stranded DNA [31••]. Consistent with this interpretation, a residue that is known to play an important role in DNA binding, Lys264, protrudes into this groove [36]. Only one other SH3 domain, the monomeric Sso7d domain [40], has been reported to be involved in DNA binding. The surfaces that interact with DNA must be quite different between the two proteins because there is extensive overlap between the surface of Sso7d that binds DNA and the surface that forms the dimer interface of the HIV C-terminal domain.

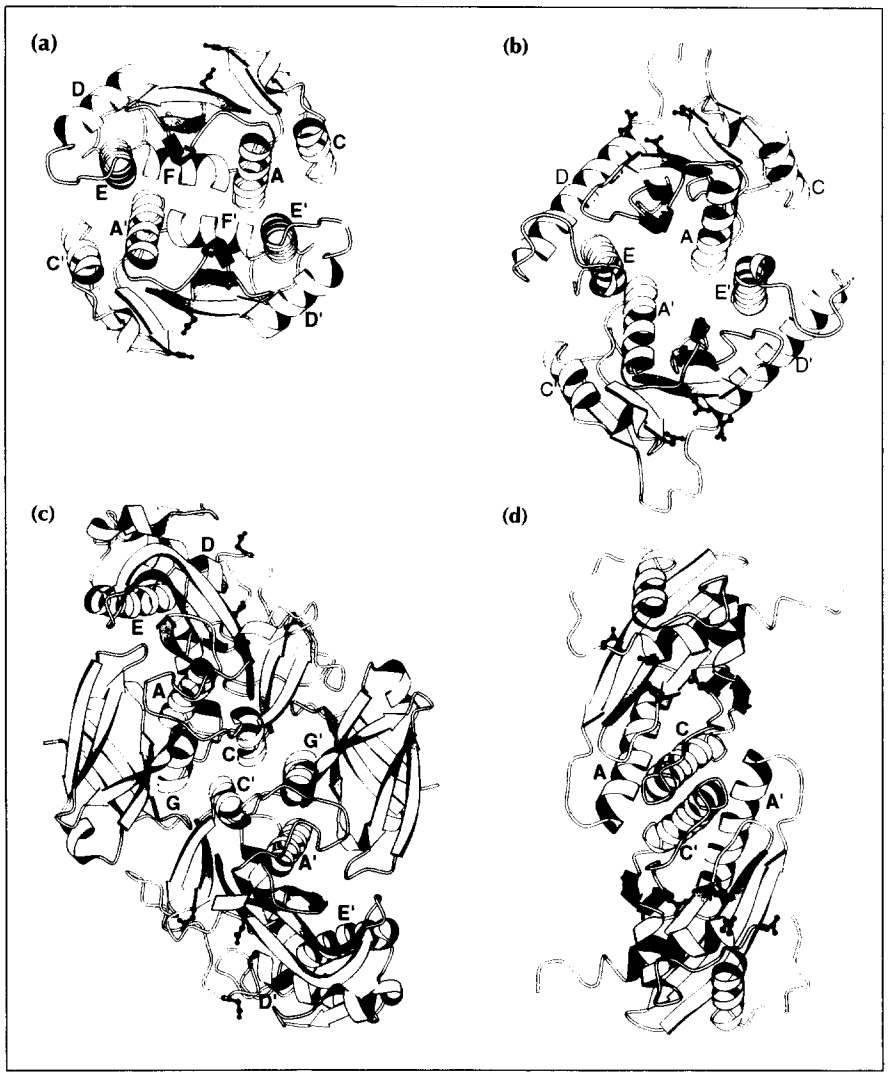
Comparison of integrase with structurally related polynucleotidyl transferases

The structure of the HIV integrase core domain revealed that it belongs to a structurally related superfamily of polynucleotidyl transferases [29••,41•] of which the earliest structure elucidated was that of RNase H, but which also includes RuvC resolvase, and the core domain of bacteriophage Mu transposase (Fig. 3). This group of enzymes also seems to be more distantly related to both the kinase family and the 3′-5′ exonuclease domain of the Klenow fragment [42]. These enzymes have five β strands and helices equivalent to A, B, and C of the structures shown in Figure 3, with the same topological arrangement. In many cases, however, there are large insertions between these structural elements and, although they all have one active site residue near the C terminus of β strand 1, the location of the other active site residues is quite variable.

The Mu transposase catalytic core

Despite multiple insertions and a mere 15% sequence identity, the structure of the core domain of the phage Mu transposase [33••], or MuA protein, is strikingly similar to that of the retroviral integrases (Fig. 3). The MuA core is approximately twice the size of the integrase core and can be divided into two subdomains: the first corresponds to the integrase core and can be superimposed on the HIV-1 structure with an rms distance of 1.7 Å over 114 Cαs, whereas the second subdomain is a six-stranded β barrel that may be involved in non-specific DNA binding.

Figure 4



Ribbon representations of the structures of the polynucleotidyl transferase superfamily members, showing the major intersubunit interactions in the crystal structures. Helices (A through F) are labelled to correspond with those of integrase. Residues that have been implicated in catalysis are shown in ball-and-stick representation (see Fig. 3). **(a)** HIV integrase core. **(b)** ASV integrase core. **(c)** Mu transposase core. **(d)** RuvC resolvase core. Figure prepared using MOLSCRIPT [65].

The two active site aspartate residues are found in the same location as those of the integrases and the third active site residue, Glu392, is located on a turn just before helix D [43,44]. It is visible in the electron density map, but, unlike its ASV counterpart, it extends away from the active site. This probably reflects the fact that Mu transposition is more tightly regulated than retroviral integration, and the protein is normally inactive until it has formed a transpososome (a tetramer of MuA bound to the two ends of Mu DNA) [45]. Unlike the two integrase fragments discussed above, no chemical activity could be found for the isolated MuA core. A conformational change in the loop, triggered by protein–protein or protein–DNA interactions with structures not present in this crystal, probably brings Glu392 into proximity with the rest of the active site. This cannot be modeled by simply extending helix D backward into the loop region in a manner analogous to that at the active site of ASV integrase, as MuA contains an extra two residues between this glutamate and the parts of helix D that can be aligned.

The presence of a large patch of positive electrostatic potential on the surface of the barrel suggests that it is responsible for the non-specific DNA-binding activity of this fragment [46]. It is interesting that although both this subdomain and the integrase C-terminal domain are located in the primary sequence directly after the polynucleotidyl transferase domain, are probably used for non-specific DNA binding, and are antiparallel β barrel structures, they appear to be topologically unrelated to one another. It is not clear whether this is an extreme case of divergent evolution or merely coincidence. Among transposases, the primary sequence conservation in this region is so poor that it is unclear how conserved this structural motif will prove to be. However, the contacts between the two subdomains of the MuA core are clearly

Mu specific, as they are formed largely by two loops extending from the active site subdomain that are longer than those present in most related enzymes (Fig. 3b).

Intact MuA is a monomer in solution in the absence of DNA. Different protein–protein contacts are seen in two different crystal forms of the core domain, but the same general region of surface is buried at the interface. The dimeric pairing involving the largest amount of buried surface area (2060 Å^2) is shown in Figure 4c. A dimer interaction such as that formed by the integrase cores could not be modelled because of steric conflicts with helix G, which would lie between the two halves of such a dimer.

RNase H

There have been two crystal structure determinations of *Escherichia coli* RNase H [47–49]. The structure of RNase H from *Thermus thermophilus* has also been determined at 2.8 Å resolution [50]. In the case of HIV-1, RNase H is a domain of the p66 chain of the reverse transcriptase. The structure of HIV-1 RNase H has been determined both as an isolated molecule [51], where it is essentially inactive, and as part of the intact reverse transcriptase [52–55].

The *E. coli* and *T. thermophilus* RNase H structures are very similar, sharing 52% sequence identity and an rms deviation of 0.95 Å for 109 Cα atoms within the secondary structure elements. A positively charged loop in the *E. coli* RNase H, called the basic protrusion (Fig. 3d), is thought to be involved in nucleic acid binding [56] and this region is also present in the *T. thermophilus* enzyme. Although two of the essential carboxylates in both enzymes lie on β strands 1 and 4, as in the case of the integrases and Mu transposase, the third is found on helix A, and thus approaches the active site from the opposite side of the β sheet. Direct evidence for the involvement of these carboxylates in binding divalent metal ions comes from the difference map calculated from diffraction data obtained from metal-free crystals and crystals that had been soaked in 45 mM Mn^{2+} [51]. Two tightly bound Mn^{2+} ions were found in close proximity to these residues.

The *E. coli* and HIV enzymes are similar, with a Cα rms of 1.23 Å for all elements of secondary structure (reviewed by Hughes *et al.* in [57]). A major difference is that the basic protrusion in the *E. coli* enzyme, comprising helix C and the twelve residues that follow it, is missing in the HIV-1 enzyme (Fig. 3d,f). The role of this protrusion is presumably filled by the proximal parts of the reverse transcriptase in the case of the HIV RNase H [57]. Consistent with this interpretation, the introduction of the basic protrusion of *E. coli* RNase H (residues 76–102) into the equivalent region of the HIV enzyme yielded a hybrid protein that was highly active in the presence of Mn^{2+} [58].

The loop between β strand 5 and helix D of HIV-1 RNase H appears to be flexible; it is disordered in the HIV enzyme as an isolated domain [57] and is only visible when it is part of the intact reverse transcriptase. It is noteworthy that this loop exhibits flexibility in several members of the polynucleotidyl transferase family discussed here. It is disordered in HIV integrase and HIV RNase H, mobile in ASV integrase, and in the Mu transposase it must adopt an alternative conformation to assemble the complete active site. Mutating residues within the corresponding loop of *E. coli* RNase H does not abolish catalytic activity, but does have large effects on K_{cat} and K_M [59].

RuvC resolvase

The structure of RuvC resolvase [60••], a Holliday junction specific endonuclease from *E. coli*, exhibits a topology very similar to that of integrase and RNase H, although it is the most distantly related member of this family (Fig. 4). Comparison with *E. coli* RNase H shows that part of the common structural motif (β1–4 and helix A) contains 35 Cαs that superimpose with an rms deviation of 2.1 Å. Unlike RNase H, however, the RuvC protein forms a dimer in the active form [61]. The major intersubunit contact in the dimer is between the B helices of each monomer (consisting of residues 76–93, corresponding to helix C of the integrase cores), which lie roughly parallel to the dimer axis. The dimer interface contains both hydrophobic and polar residues, and places the two active sites 30 Å apart. This dimer is quite different from that formed by the integrase cores.

The catalytic center lies at the bottom of a cleft that can be nicely fitted onto a DNA duplex. It contains a cluster of carboxylic acid residues (Asp7, Glu66, Asp138 and Asp141) that have been shown to be necessary for RuvC resolvase activity *in vitro* [62]. Two of these residues lie on β strands 1 and 4, the same as in all the structures discussed above. The other two lie near the N terminus of a helix that passes across the face of the β sheet. The direction of this helix is opposite to that of helix D in the other structures. Soaking the crystals in 5 mM $MnCl_2$ resulted in a single new peak in the electron density map in contact with the carboxylate oxygens of Asp7 and Asp141, as would be expected if manganese is coordinated by these residues.

Conclusions

The catalytic domains of HIV-1 integrase, ASV integrase, and Mu transposase belong to the structurally related superfamily of polynucleotidyl transferases, which also includes the nucleases RNase H and RuvC. Among these structures, the integrases and the transposase are the most similar and RuvC resolvase is the most divergent. All these enzymes cut polynucleotides leaving 3′OH and 5′PO_4 groups, and the integrases and the transposase share a strand-transfer activity that splices DNA. In all cases, a divalent metal ion cofactor is required for catalytic activity. These proteins have remarkably similar active sites that contain three or four catalytic carboxylate residues. In all five structures, two of these residues are located in

identical positions in the secondary structure. In contrast, the third carboxylate residue is contributed by different secondary structural elements. In RNase H it is on helix A, in HIV-1 integrase it is on a disordered loop leading to helix D, in ASV it is on a mobile region near the N-terminal end of helix D, and in RuvC resolvase it is on the fifth helix (Fig. 3). In all these structures, the active-site carboxylates either adopt a very similar configuration in the tertiary structure (e.g. the three carboxylates in RNase H can be superimposed onto those of RuvC resolvase with an rms deviation of 2.1 Å [12]), or could adopt such a configuration without any major structural perturbation of the protein. It is likely that all these enzymes share a common mechanism of chemical catalysis similar to that proposed for the 3′-5′ exonuclease activity of *E. coli* DNA polymerase I [63,64]. Although integrases, transposases and nucleases at first sight may seem to catalyze dissimilar reactions, the major difference need only be the ability to position the 3′OH end of a DNA molecule as the attacking nucleophile in the case of integrases and transposases [10]. The proteins we have discussed appear to have similar active sites that have been recruited to serve different biological functions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Goff SP: **Genetics of retroviral integration.** *Annu Rev Genet* 1992, **26**:527–544.
2. Whitcomb JM, Hughes SH: **Retroviral reverse transcription and integration: progress and problems.** *Annu Rev Cell Biol* 1992, **8**:275–306.
3. Katz RA, Skalka AM: **The retroviral enzymes.** *Annu Rev Biochem* 1994, **63**:133–173.
4. Craigie R, Fujiwara T, Bushman F: **The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*.** *Cell* 1990, **62**:829–837.
5. Bushman FD, Fujiwara T, Craigie R: **Retroviral DNA integration directed by HIV integration protein *in vitro*.** *Science* 1990, **249**:1555–1558.
6. Katz RA, Merkel G, Kulkosky J, Leis J, Skalka AM: **The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*.** *Cell* 1990, **63**:87–95.
7. Sherman PA, Fyfe JA: **Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity.** *Proc Natl Acad Sci USA* 1990, **87**:5119–5123.
8. Bushman FD, Craigie R: **Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA.** *Proc Natl Acad Sci USA* 1991, **88**:1339–1343.
9. Engelman A, Mizuuchi K, Craigie R: **HIV1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.** *Cell* 1991, **67**:1211–1221.
10. Mizuuchi K: **Polynucleotidyl transfer reactions in transpositional DNA recombination.** *J Biol Chem* 1992, **267**:21273–21276.
11. Mizuuchi K: **Transpositional recombination: mechanistic insights from studies of mu and other elements.** *Annu Rev Biochem* 1992, **61**:1011–1051.
12. Engelman A, Craigie R: **Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*.** *J Virol* 1992, **66**:6361–6369.
13. Rowland SJ, Dyke KG: **Tn552, a novel transposable element from *Staphylococcus aureus*.** *Mol Microbiol* 1990, **4**:961–975.
14. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM: **Retroviral integrase domains: DNA binding and the recognition of LTR sequences.** *Nucleic Acids Res* 1991, **19**:851–860.
15. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM: **Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases.** *Mol Cell Biol* 1992, **12**:2331–2338.
16. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R: **Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding.** *Proc Natl Acad Sci USA* 1993, **90**:3428–3432.
17. Chow SA, Vincent KA, Ellison V, Brown PO: **Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus.** *Science* 1992, **255**:723–726.
18. Drelich M, Wilhelm R, Mous J: **Identification of amino acid residues critical for endonuclease and integration activities of HIV1 IN protein *in vitro*.** *Virology* 1992, **188**:459–468.
19. Leavitt AD, Rose RB, Varmus HE: **Both substrate and target oligonucleotide sequences affect *in vitro* integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*.** *J Virol* 1992, **66**:2359–2368.
20. Van Gent DC, Groeneger AA, Plasterk RH: **Mutational analysis of the integrase protein of human immunodeficiency virus type 2.** *Proc Natl Acad Sci USA* 1992, **89**:9598–9602.
21. Schauer M, Billich A: **The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding.** *Biochem Biophys Res Commun* 1992, **185**:874–880.
22. Vink C, Oude Groeneger AM, Plasterk RH: **Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein.** *Nucleic Acids Res* 1993, **21**:1419–1425.
23. Woerner AM, Marcussekura CJ: **Characterization of a DNA binding domain in the C-terminus of HIV1 integrase by deletion mutagenesis.** *Nucleic Acids Res* 1993, **21**:3507–3511.
24. Engelman A, Hickman AB, Craigie R: **The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding.** *J Virol* 1994, **68**:5911–5917.
25. Burke CJ, Sanyal G, Bruner MW, Ryan JA, Lafemina RL, Robbins HL, Zeft AS, Middaugh CR, Cordingley MG: **Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase.** *J Biol Chem* 1992, **267**:9639–9644.
26. Engelman A, Bushman FD, Craigie R: **Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex.** *EMBO J* 1993, **12**:3269–3275.
27. Van Gent DC, Vink C, Groeneger AA, Plasterk RH: **Complementation between HIV integrase proteins mutated in different domains.** *EMBO J* 1993, **12**:3261–3267.
28. Jones KS, Coleman J, Merkel GW, Laue TM, Skalka AM: **Retroviral integrase functions as a multimer and can turn over catalytically.** *J Biol Chem* 1992, **267**:16037–16040.
29. •• Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR: **Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases.** *Science* 1994, **266**:1981–1986.

The first structure of a retroviral integrase catalytic core domain. Described at 2.5 Å, the structure showed an unexpected similarity to RNase H and RuvC resolvase. See also [30••].

30. •• Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A: **High resolution structure of the catalytic domain of avian sarcoma virus integrase.** *J Mol Biol* 1995, in press.

The 1.7 Å structure of a second retroviral catalytic core showing a highly conserved structure despite only 24% sequence identity to the HIV core. A region containing one of the catalytic residues that is disordered in the in the HIV structure is visible in this structure. See also [29••].

31. •• Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM: **Solution structure of the DNA binding domain of HIV-1 integrase.** *Biochemistry* 1995, **34**:9826–9833.

The NMR structure of the C-terminal domain of HIV integrase, demonstrating similarity to SH3 domains. A model for DNA binding is presented. Novel use of dimeric SH3 structure to bind DNA.

32. •• Eijkelenboom APAM, Puras Lutzke RA, Boelems R, Plasterk RHA, Kaptein R, Hard K: **The DNA-binding domain of HIV-1 integrase has an SH3-like fold.** *Nature Struct Biol* 1995, **2**:807–810.
The NMR structure of the C-terminal domain of HIV integrase, which demonstrates similarity to SH3 domains. Novel use of dimeric SH3 structure to bind DNA.

33. •• Rice P, Mizuuchi K: **Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration.** *Cell* 1995, **82**:209–220.
The 2.4 Å structure of the catalytic domain of the Mu transposase (Mu A protein). Despite almost no obvious similarity in the primary sequence, it exhibits remarkable structural similarity to the catalytic domains of HIV and ASV integrase. The main differences are longer loops between the structural elements.

34. Hickman AB, Palmer I, Engelman A, Craigie R, Wingfield P: **Biophysical and enzymatic properties of the catalytic domain of HIV-1 integrase.** *J Biol Chem* 1994, **269**:29279–29287.

35. •• Jenkins TM, Hickman AB, Dyda F, Ghirlando R, Davies DR, Craigie R: **Catalytic domain of human immunodeficiency virus type-1 integrase: identification of a soluble mutant by systematic replacement of hydrophobic residues.** *Proc Natl Acad Sci USA* 1995, **92**:6057–6061.
A novel approach to resolve desperate situations in crystallization. All the hydrophobic residues in HIV-1 integrase were mutated, either singly or in small groups, to lysine or alanine respectively. The presence of just one mutation dramatically improved the solubility of the protein.

36. Puras Lutzke RA, Vink C, Plasterk RH: **Characterization of the minimal DNA-binding domain of the HIV integrase protein.** *Nucleic Acids Res* 1994, **22**:4125–4131.

37. Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M: **Crystal structure of a Src-homology 3 (SH3) domain.** *Nature* 1992, **359**:851–855.

38. Wilson KP, Shewchuk LM, Brennan RG, Otsuka AJ, Matthews BW: ***Escherichia coli* biotin holoenzyme synthetase/bio repressor crystal structure delineates the biotin and DNA-binding domains.** *Proc Natl Acad Sci USA* 1992, **89**:9257–9261.

39. Yu H, Rosen MK, Shin TB, Seideldugan C, Brugge JS, Schreiber SL: **Solution structure of the SH3 domain of Src and identification of its ligand-binding site.** *Science* 1992, **258**:1665–1668.

40. Baumann H, Knapp S, Karshikoff A, Ladenstein R, Hard T: **DNA-binding surface of the Sso7d protein from *Sulfolobus solfataricus*.** *J Mol Biol* 1995, **247**:840–846.

41. • Yang W, Steitz TA: **Recombining the structures of HIV integrase, RuvC and RNase H.** *Structure* 1995, **3**:131–134.
A very nice review of the similarity among structures of the polynucleotidyl transferase superfamily.

42. Artymiuk PJ, Grindley HM, Kumar K, Rice DW, Willett P: **Three-dimensional structural resemblance between the ribonuclease H and connection domains of HIV reverse transcriptase and the ATPase fold revealed using graph theoretical techniques.** *FEBS Lett* 1993, **324**:15–21.

43. Baker TA, Luo L: **Identification of residues in the Mu transposase essential for catalysis.** *Proc Natl Acad Sci USA* 1994, **91**:6654–6658.

44. Kim K, Namgoong SY, Jayaram M, Harshey RM: **Step-arrest mutants of phage Mu transposase. Implications in DNA–protein assembly, Mu end cleavage, and strand transfer.** *J Biol Chem* 1995, **270**:1472–1479.

45. Baker TA, Mizuuchi K: **DNA-promoted assembly of the active tetramer of the Mu transposase.** *Genes Dev* 1992, **6**:2221–2232.

46. Nakayama C, Teplow DB, Harshey RM: **Structural domains in phage Mu transposase: identification of the site-specific DNA-binding domain.** *Proc Natl Acad Sci USA* 1987, **84**:1809–1813.

47. Yang W, Hendrickson WA, Crouch RJ, Satow Y: **Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein.** *Science* 1990, **249**:1398–1405.

48. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Ikehara M, Matsuzaki T, Morikawa K: **Three-dimensional structure of ribonuclease H from *E. coli*.** *Nature* 1990, **347**:306–309.

49. Katayanagi K, Miyagawa M, Matsushima M, Ishikawa M, Kanaya S, Nakamura H, Ikehara M, Matsuzaki T, Morikawa K: **Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution.** *J Mol Biol* 1992, **223**:1029–1052.

50. Ishikawa K, Okumura M, Katayanagi K, Kimura S, Kanaya S, Nakamura H, Morikawa K: **Crystal structure of ribonuclease H from *Thermus thermophilus* HB8 refined at 2.8 Å resolution.** *J Mol Biol* 1993, **230**:529–542.

51. Davies JF, Hostomska Z, Hostomsky Z, Jordan SR, Matthews DA: **Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase.** *Science* 1991, **252**:88–95.

52. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA: **Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor.** *Science* 1992, **256**:1783–1790.

53. Jacobomolina A, Ding J, Nanni RG, Clark AD Jr, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P: **Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA.** *Proc Natl Acad Sci USA* 1993, **90**:6320–6324.

54. Ren J, Esnouf R, Garman E, Somers D, Ross C, Kirby I, Keeling J, Darby G, Jones Y, Stuart D: **High resolution structures of HIV-1 RT from four RT–inhibitor complexes.** *Nature Struct Biol* 1995, **2**:293–302.

55. Rodgers DW, Gamblin SJ, Harris BA, Ray S, Culp JS, Hellmig B, Woolf DJ, Debouck C, Harrison SC: **The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1.** *Proc Natl Acad Sci USA* 1995, **92**:1222–1226.

56. Kanaya S, Katsudanakai C, Ikehara M: **Importance of the positive charge cluster in *Escherichia coli* ribonuclease HI for the effective binding of the substrate.** *J Biol Chem* 1991, **266**:11621–11627.

57. Hughes SH, Arnold E, Hostomsky Z: **RNase H of retroviral reverse transcriptases.** In *Ribonucleases.* Edited by Toulme JJ, Crouch RJ. Paris: INSERM Editions; 1996:in press.

58. Stahl SJ, Kaufman JD, Vikictopic S, Crouch RJ, Wingfield PT: **Construction of an enzymatically active ribonuclease H domain of human immunodeficiency virus type 1 reverse transcriptase.** *Protein Eng* 1994, **7**:1103–1108.

59. Kanaya S, Kohara A, Miura Y, Sekiguchi A, Iwai S, Inoue H, Ohtsuka E, Ikehara M: **Identification of the amino acid residues involved in an active site of *Escherichia coli* ribonuclease H by site-directed mutagenesis.** *J Biol Chem* 1990, **265**:4615–4621.

60. •• Ariyoshi M, Vassylyev DG, Iwasaki H, Nakamura H, Shinagawa H, Morikawa K: **Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E. coli*.** *Cell* 1994, **78**:1063–1072.
Structure of the Holliday junction resolving enzyme RuvC. This structure is very similar to that of RNase H.

61. Iwasaki H, Takahagi M, Shiba T, Nakata A, Shinagawa H: ***Escherichia coli* RuvC protein is an endonuclease that resolves the Holliday structure.** *EMBO J* 1991, **10**:4381–4389.

62. Saito S, Iwasaki H, Ariyoshi M, Morikawa K, Shinagawa H: **Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase.** *Proc Natl Acad Sci USA* 1995, **92**:7470–7474.

63. Freemont PS, Friedman JM, Beese LS, Sanderson MR, Steitz TA: **Cocrystal structure of an editing complex of Klenow fragment with DNA.** *Proc Natl Acad Sci USA* 1988, **85**:8924–8928.

64. Beese LS, Steitz TA: **Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism.** *EMBO J* 1991, **10**:25–33.

65. Kraulis PJ: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.** *J Appl Crystallogr* 1995, **24**:946–950.