The tribological behavior of C_{60} on NaCl(001) is unusual. Extremely low shear strengths of 0.05 to 0.1 MPa are found. The small dissipation energies and the reasonable cohesive energy allow C_{60} islands to move as an entity with small lateral forces. Conceivably, C_{60} islands could be used as transport devices for fabrication processes of nanometer-sized machines, whereby C_{60} islands might play the role of a transport carrier. Larger molecules (biomolecules) could be deposited on such a nanosled and then transported to a desired location.

REFERENCES AND NOTES

1. D. Tománek and M. A. Schluter, *Phys. Rev. Lett.* **67**, 2331 (1991).
2. D. W. Brenner, J. A. Harrison, C. T. White, R. J. Colton, *Thin Solid Films* **206**, 220 (1991); R. C. Mowrey, D. W. Brenner, B. I. Dunlap, J. W. Mintmire, C. T. White, *J. Phys. Chem.* **95**, 7138 (1991).
3. B. Bhushan, B. K. Gupta, G. W. Van Cleef, C. Capp, J. V. Coe, *Appl. Phys. Lett.* **62**, 3253 (1993).
4. D. M. Eigler and E. K. Schweizer, *Nature* **344**, 524 (1990); Y. W. Mo, *Phys. Rev. Lett.* **71**, 2923 (1993); Ph. Avouris and I.-W. Lyo, *Appl. Surf. Sci.* **60–61**, 426 (1992).
5. C. M. Mate, G. M. McClelland, R. Erlandsson, S. Chiang, *Phys. Rev. Lett.* **59**, 1942 (1987).
6. G. Meyer and N. M. Amer, *Appl. Phys. Lett.* **57**, 2089 (1990).
7. L. Howald *et al.*, *ibid.* **63**, 117 (1993).
8. We calibrated the piezoelectric tube scanner in the *x*, *y*, and *z* directions according to the procedure given in a study on the Si(111)7 × 7 surface by operating the multifunctional force microscope in the scanning tunneling mode [see (*7*)].
9. R. Lüthi *et al.*, *Z. Phys. B* **95**, 1 (1994).
10. B. Bhushan, J. Ruan, B. K. Gupta, *J. Phys. D* **26**, 1319 (1993); B. Bhushan, B. K. Gupta, G. W. Van Cleef, C. Capp, J. V. Coe, *Tribol. Trans.* **36**, 573 (1993).
11. M. Mate, *Wear* **168**, 17 (1993).
12. E. Meyer *et al.*, *Phys. Rev. Lett.* **69**, 1777 (1992); R. M. Overney *et al.*, *Nature* **359**, 133 (1992).
13. M. Hirano and K. Shinjo, *Phys. Rev. B* **41**, 11837 (1990).
14. G. M. McClelland, in *Adhesion and Friction*, M. Grunze and H. J. Kreuzer, Eds., vol. 17 in the Springer Series in Surface Science (Springer, Berlin, 1990), p. 81.
15. J. Belak, D. B. Boercker, I. F. Stowers, *Mater. Res. Soc. Bull.* **18**, 55 (1993); T. Gyalog *et al.*, unpublished material.
16. J. B. Sokoloff, *Phys. Rev. B* **42**, 760 (1990).
17. E. I. Altman and R. Colton, in *Atomic and Nanometer-Scale Modifications of Materials: Fundamentals and Applications*, Ph. Avouris, Ed. (vol. 239 of the NATO Advanced Study Institutes Series, Kluwer Academic, London, 1993), pp. 303–313.
18. Ph. Lambin, A. A. Lucas, J.-P. Vigneron, *Phys. Rev. B* **46**, 1794 (1992).
19. C. Pan, M. P. Sampson, Y. Chai, R. H. Hauge, J. L. Margrave, *J. Phys. Chem.* **95**, 2944 (1991).
20. Cantilevers supplied by O. Ohlsson of Nanosensors, Aidlingen, Germany.
21. The normal force was deduced with respect to the point at which the tip jumps off the surface.

22. We thank Ch. Gerber, D. Anselmetti, P. Chaudari, H.-P. Lang, V. Thommen-Geiser, and H. R. Hidber who contributed to many stimulating and clarifying discussions. We thank H. Breitenstein, A. Tonin, and R. Maffiolini for their technical help. We thank R. Hofer and D. Brodbeck for providing us with user-friendly software. We are especially grateful to J. Frommer who motivated one of us (R.L.) with donations of vitamin C. Supported by the Swiss National Science Foundation and the Kommission zur Förderung der wissenschaftlichen Forschung.

8 August 1994; accepted 14 October 1994

Crystal Structure of the Catalytic Domain of HIV-1 Integrase: Similarity to Other Polynucleotidyl Transferases

Fred Dyda,* Alison B. Hickman,* Timothy M. Jenkins, Alan Engelman, Robert Craigie, David R. Davies†

HIV integrase is the enzyme responsible for inserting the viral DNA into the host chromosome; it is essential for HIV replication. The crystal structure of the catalytically active core domain (residues 50 to 212) of HIV-1 integrase was determined at 2.5 Å resolution. The central feature of the structure is a five-stranded β sheet flanked by helical regions. The overall topology reveals that this domain of integrase belongs to a superfamily of polynucleotidyl transferases that includes ribonuclease H and the Holliday junction resolvase RuvC. The active site region is identified by the position of two of the conserved carboxylate residues essential for catalysis, which are located at similar positions in ribonuclease H. In the crystal, two molecules form a dimer with an extensive solvent-inaccessible interface of 1300 Å^2 per monomer.

Integration of HIV DNA into the host genome is an essential step in the viral replication cycle (*1*). The enzyme responsible for integration, HIV integrase, has no known functional analog in human cells and is therefore a particularly attractive target for the design of antiviral agents. The three-dimensional structures of two other essential HIV enzymes, reverse transcriptase and protease, have been determined and much progress has been made toward the structure-based design of effective inhibitors, particularly with protease (*2*).

HIV DNA integration occurs through a defined set of DNA cutting and joining reactions. In the first step of the integration process, 3′ processing, two nucleotides are removed from each 3′ end of the blunt-ended viral DNA made by reverse transcription. A subsequent DNA strand transfer reaction covalently joins the recessed 3′ ends of the viral DNA to the 5′ ends of the target DNA at the site of integration. Purified HIV integrase carries out these reactions in vitro with either Mg^{2+} or Mn^{2+} as a cofactor (Fig. 1). Integrase cleaves two nucleotides from the 3′ ends of DNA substrates that mimic the viral DNA ends and also inserts the processed ends into other DNA molecules that serve as targets for strand transfer (*3*). When presented with a DNA substrate that mimics the product of DNA strand transfer, integrase can catalyze an apparent reversal of the strand transfer reaction, termed disintegration (*4*). In this reaction, the viral DNA end segment of the substrate is liberated and the target DNA segment is sealed.

Several lines of evidence point to a common active site for all three catalytic activities. Three highly conserved amino acid residues in the central core domain of HIV-1 integrase, Asp^{64}, Asp^{116}, and Glu^{152}, the D,D-35-E motif, are observed in the integrase proteins of retroviruses and retrotransposons as well as in the transposase proteins of some prokaryotic transposons (*5–7*). In general, mutation of any one of these residues abolishes all enzyme activities (*6–9*). Stereochemical analysis of the 3′ processing and DNA strand transfer reactions also indicates that each occurs through a one-step transesterification mechanism (*10*). In the 3′ processing reaction, integrase activates the phosphodiester bond at the site of cleavage to nucleophilic attack by various nucleophiles (*10, 11*). The DNA strand transfer reaction may occur by a similar mechanism, with integrase playing the additional role of positioning the 3′-OH end of the viral DNA for nucleophilic attack on a phosphodiester bond in the target DNA (*10*).

While the full-length integrase protein (288 residues) is required for 3′ processing and DNA strand transfer activities (*9, 12, 13*), the central core domain can carry out the disintegration reaction and therefore contains the catalytic site for polynucleotidyl transfer (*14*). The precise roles of the NH_2- and COOH-terminal domains in the 3′ processing and DNA strand transfer reactions are not known, although the COOH-terminal domain binds to DNA nonspecifically (*12, 15, 16*).

Previous attempts to crystallize HIV-1 integrase have been obstructed by its poor solubility (*17*). In an attempt to circumvent this problem, we undertook a systematic replacement of the hydrophobic residues in the core domain of HIV-1 integrase (residues 50 to 212) (*18*). The single amino acid substitution of Lys for Phe^{185} resulted in a

Laboratory of Molecular Biology, NIDDK, NIH, Bethesda, MD 20892–0560, USA.

*These two authors contributed equally to this work.
†To whom correspondence should be addressed.

protein with considerably improved solubility and biophysical properties which was as active for disintegration as the unmutated core (*18*). The protein could be concentrated to at least 25 mg/ml, was monodisperse in solution in the absence of detergent, and was crystallized (*19*). We describe the three-dimensional structure of this catalytically active core domain of HIV-1 integrase at 2.5 Å resolution, locate the positions of the conserved acidic amino acids that have been shown to be essential for activity, and relate the structure to those of other known enzymes.

The crystal structure was determined by means of a combination of multiwavelength anomalous diffraction (MAD) and multiple isomorphous replacement with anomalous scattering (MIRAS) methods. The resulting electron density at 2.75 Å resolution was of good quality for most of the polypeptide chain (Fig. 2), allowing unambiguous chain tracing (Table 1). The current crystallographic *R* factor is 22.7% at 2.5 Å resolution with no solvent molecules added.

The structure of the core domain consists of a central five-strand β sheet and six helices (Figs. 3 and 4). Its overall topology is similar to that of ribonuclease H (RNase H) (*20*, *21*) (Fig. 5, A and B). A similar resemblance has been observed between RNase H, the Holliday junction resolvase RuvC (*22*), and the core domain of the transposase protein of bacteriophage Mu (*23*). This topological similarity has also been observed in the connection domain of HIV-1 reverse transcriptase and the adenosine triphosphatase (ATPase) domains of hexokinase, glycerol kinase, actin, and the 70-kD heat-shock protein (*24*).

The first four strands of the integrase β sheet superimpose closely on the corresponding HIV-1 RNase H sheet [root mean square deviation (rmsd) = 2.1 Å for 32 Cα atoms]. When the two β sheets are superimposed, however, it becomes apparent that α1 of integrase is displaced from the sheet by approximately 6 Å relative to the first helix of RNase H. The three-turn helix in RNase H, α2, is a one-turn helix in integrase; α3 is similar in both structures. In contrast, the orientation of α4 is quite different in the two structures; in RNase H, α4 aligns parallel to β1. After α4, there are two additional helices in the integrase structure that are not present in RNase H. These are α5, consisting of residues 171 to 186, and α6, extending from residues 196 to 208; α5 is located adjacent and parallel to β3, and α6 is located on the same side of the β sheet as α1. The orientation of α6 is approximately 90° relative to α5. Carboxyl-terminal truncation mutants of HIV-1 integrase that lack both α5 and α6 are catalytically inactive (*14*, *16*).

No interpretable density is observed for the first nine residues at the NH_2-terminus, corresponding to three amino acids that are not part of the core sequence but are remnants of the histidine tag after thombin

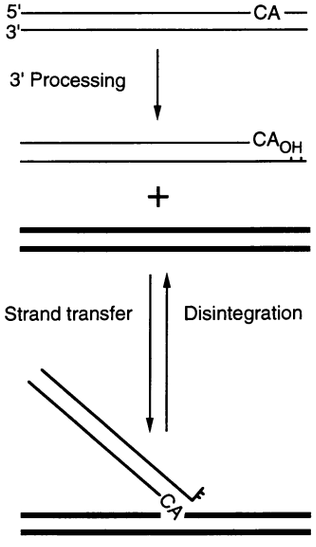

Fig. 1. The three catalytic activities of retroviral integrase. In the 3′ processing reaction, integrase cleaves viral substrate DNA (thin lines) at a specific phosphodiester bond 3′ of the conserved CA dinucleotide. In the strand transfer reaction, the recessed 3′-OH end is covalently joined to a 5′ phosphate at the site of integration in a target DNA (bold lines). Viral end sequences are separated from target DNA in the disintegration reaction. Purified HIV-1 integrase predominantly joins just a single viral DNA end to target DNA in vitro. In the complete integration reaction, integrase inserts a pair of viral DNA ends with a spacing of five base pairs between the sites of insertion on the two target DNA strands. Repair of the single strand gaps between viral and host DNA in the resulting integration intermediate, which is probably accomplished by cellular enzymes, results in a five base pair duplication of target DNA sequence flanking the integrated viral DNA.

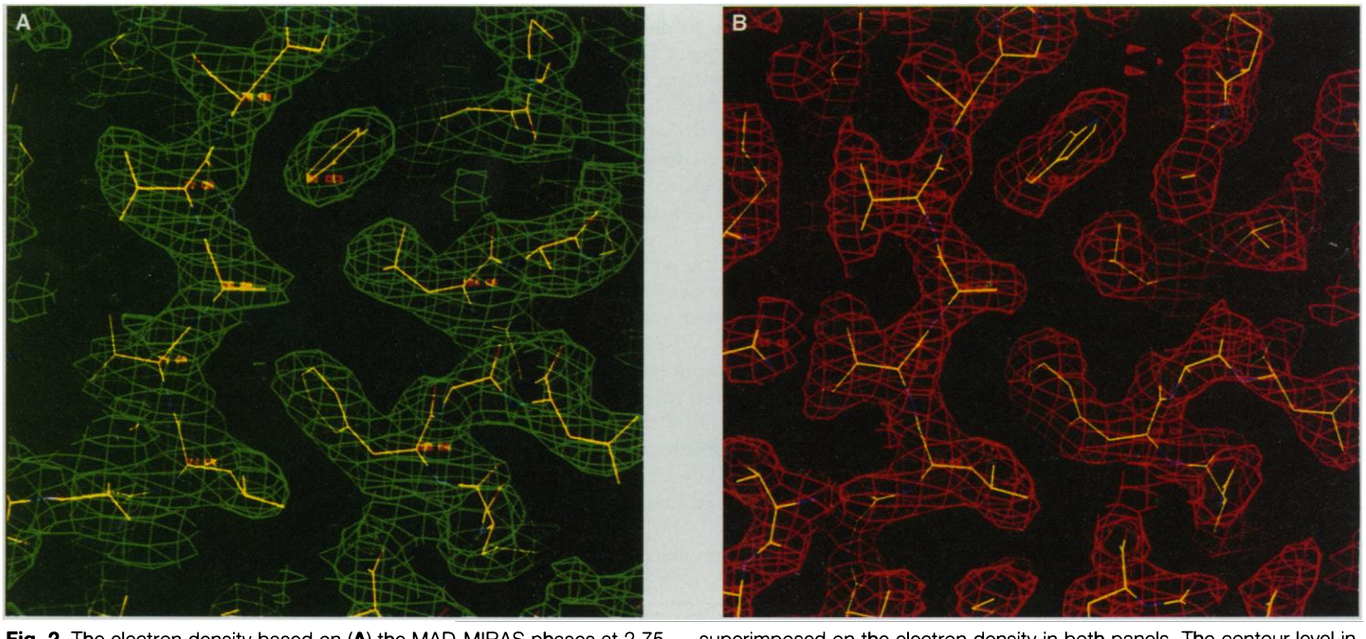

Fig. 2. The electron density based on (**A**) the MAD-MIRAS phases at 2.75 Å after solvent flattening. (**B**) Electron density ($2F_o - 2F_c$) based on phases computed from the current model at 2.5 Å resolution. The current model is superimposed on the electron density in both panels. The contour level in both is at one standard deviation.

cleavage, together with amino acids 50 to 55. There is also an internal loop region, consisting of residues 141 to 153, which is disordered. In addition, the COOH-terminal four residues are not visible.

Despite lack of sequence similarity between the core domain of HIV-1 integrase and RNase H (*25*, *26*), there is remarkable similarity in the positioning of two of the catalytic residues. Two of the conserved acidic residues of integrase that are essential for catalysis (*6*, *8*, *9*) are Asp^{64} in the middle of β1 and Asp^{116} after β4. These superimpose well on two of the catalytic residues of HIV-1 RNase H, Asp^{443} and Asp^{498}. The third integrase catalytic residue, Glu^{152}, lies in a disordered region between β5 and α4 and is therefore not part of the current interpretation, although it must be located in the general area of the two aspartates on the basis of the location of α4, which begins at residue 154. In HIV-1 RNase H, the third essential catalytic residue, Glu^{478}, is part of α1 (*20*). In the integrase structure, the displacement of α1 makes it unlikely that an active site residue is contributed by this helix. Thus, in the two structures, it appears that the location and positioning of the two aspartates in β1 and after β4 are the most highly conserved features of the active sites. A similar conclusion can be drawn from a visual inspection of the RuvC structure (*22*).

As with integrase, divalent metal ions are essential for RNase H activity (*27*) and various metals, including Mn^{2+} and Mg^{2+}, have been observed in close proximity to active site residues in both *Escherichia coli* and HIV-1 RNase H (*20*, *21*). As the crystallization medium here is free of divalent metals, none are observed at the active site. However, there is a samarium binding site close to catalytic residues Asp^{64} and Asp^{116} in one of the isomorphous derivatives used in the structure determination. Torsional movements about the side-chain bonds of these residues could place their carboxylate groups within interaction distance of the samarium.

In the crystal, there are two observed contacts between monomers. One of these, between two subunits related by a dyad axis, consists of a large solvent-excluded interface of 1300 $Å^2$ per subunit (Fig. 6). It contains a number of salt bridges and hydrogen bonds involving β3, α1, α3, α5, and α6, and can be compared with interactions between antibodies and protein antigens, where the solvent-excluded surface area is approximately 800 $Å^2$ for each molecule (*28*). The other contact in the crystal, formed between subunits related by a crystal translation, has a much less extensive interface. We see no evidence that the subunits in the dimer interact through a leucine zipper coiled coil involving residues 151 to 172, as has been suggested (*29*).

The greatly improved solubility property of the core domain containing the substitution of Lys for Phe^{185} may be considered relative to the position of this residue in the structure. The side-chain amino group of the lysine of one subunit is hydrogen bonded to the main chain carbonyl oxygen of Ala^{105} of the other subunit in the dimer. This interaction occurs at the periphery of the interface between the subunits, accessible to solvent. If Lys^{185} does not induce major conformational changes, the wild-type residue Phe^{185} should also be located

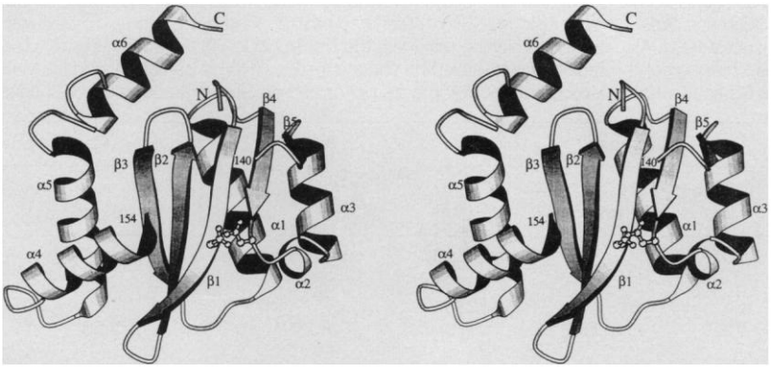

Fig. 3. Stereo diagram of a ribbon model of the core domain of HIV-1 integrase drawn with the program MOLSCRIPT v1.1 (*49*). The secondary structural elements are labeled as shown in Fig. 4. The disordered NH_2- and COOH-terminal regions are not shown. The disordered region from residues 141 to 153 is indicated by a gap in the model; residues 140 and 154 are marked. Two residues of the D,D-35-E motif, Asp^{64} and Asp^{116}, are shown as ball-and-stick models.

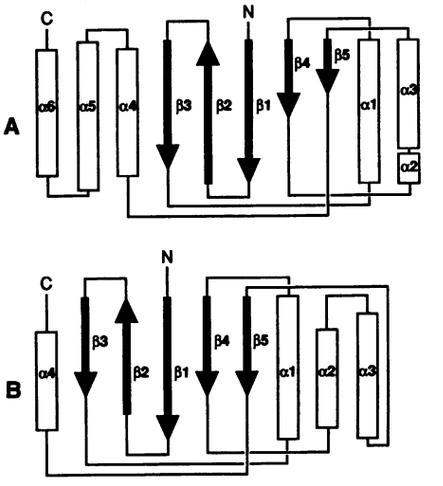

Fig. 5. Schematic diagram of the folding topologies of (**A**) the core domain of HIV-1 integrase and (**B**) RNase H from HIV-1 reverse transcriptase (*20*). Arrows are the β strands and the cylinders are α helices.

Fig. 4. Amino acid sequence of the core domain of HIV-1 integrase (amino acids 50 to 212) containing the substitution of Lys for Phe^{185} showing the secondary structure. Not shown are the first three amino acids of the crystallized protein, Gly-Ser-His, which remain after removal of the histidine tag with thrombin. Every tenth residue is numbered in the sequence. The three conserved carboxylate residues of the D,D-35-E motif are marked with asterisks. A dotted line beneath the sequence indicates disordered regions in the crystal structure.

```
50          60  *       70          80          90
MHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIEAEVIP
--------    ββββββ      ββββββ      ββββββ
                β1          β2          β3

          100         110  *      120         130
AETGQETAYFLLKLAGRWPVKTVHTDNGSNFTSTTVKAAC
          ββββββββββ      ββββββ      ββββββ
              α1          β4          α2    α3

          140         150  *      160         170
WWAGIKQEFGIPYNPQSQGVIESMNKELKKIIGQVRDQAE
          ββββββ      ββββββββββββββββββββββββ
            β5                    α4

          180         190         200         210
HLKTAVQMAVFIHNKKRKGGIGGYSAGERIVDIIATDIQTKE
          ββββββββββββββββββ      ββββββββββββββ
                    α5                  α6
```

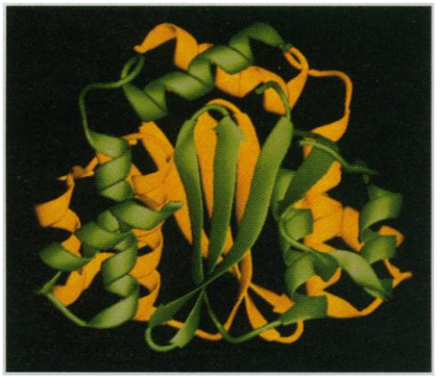

Fig. 6. A ribbon representation of the dimer observed in crystals of the core domain of HIV-1 integrase, generated with the program RIBBON v2.2 (*50*). One monomer of the dimer is shown in yellow, the other in green.

on or very close to the surface and present a hydrophobic side chain that could potentially interact with other protein molecules. Unlike lysine, phenylalanine cannot participate in hydrogen bond formation, so the unmutated core dimer will most likely be less stable than the dimer formed with the mutant protein. Both of these factors could contribute to the improved solubility properties of the mutant protein. Initial experiments suggest that the mutant protein is exclusively dimeric under conditions where the unmutated core domain exists as dimers and higher aggregates (*18*).

Although in vitro complementation experiments have established that the active form of HIV integrase for 3′ processing and DNA strand transfer is a multimer, they do not address whether the functional unit is dimeric, tetrameric, or higher order (*30*). In a dimeric model, the monomers of integrase within the dimer would each process one of the viral DNA ends. A rearrangement within the complex would then be required to introduce target DNA into the active sites and position the 3′-OH groups at the ends of the viral DNA for nucleophilic attack on a pair of phosphodiester bonds in the target DNA.

Any viable model for the authentic complex must be consistent with the observed spacing between the sites of insertion of the two viral DNA ends in target DNA. The phosphates to which the pair of viral DNA ends join on the two strands of target DNA are separated by five nucleotides, as inferred from the five base-pair duplication of host DNA that flanks integrated HIV DNA (*25*, *31*). This would require a pair of active sites separated by a spacing compatible with five nucleotides, or about 15 Å, depending on the conformation of the DNA helix. The 35 Å separation of active sites observed in the dimer is at first sight incompatible with the required spacing. It is also not possible to configure a pair of dimers so that four active sites are in close proximity. However, the interface observed in the dimer, although it could conceivably be an artifact of crystal packing, is so extensive we are persuaded it represents a functional interaction. This interpretation is compatible with several models. For example, a tetramer formed by a pair of dimers could position two active sites within the required spacing. In this model, two active sites from separate dimers would catalyze both 3′ processing and DNA strand transfer, and the other two active sites of the tetramer would not participate in catalysis.

Table 1. Structure determination. All diffraction data were collected at 95 K, integrated with DENZO (*41*) and scaled with SCALEPACK (*41*). The crystals belong to the trigonal system, space group $P3_121$, with unit cell parameters a = 72.8, c = 66.1 (Å). The asymmetric unit contained one molecule. Data on the selenomethionine-containing crystals were collected at the Howard Hughes Medical Institute beam line (X4A) at the Brookhaven National Synchrotron Light Source with Fuji imaging plates. Images were digitized on a BAS 2000 image plate scanner. Ideal incident beam energies were determined from EXAFS scans performed in the neighborhood of the Se K absorption edge with the aim to maximize the difference in the real part of the Se scattering factor (at L2) and to maximize the imaginary part at L3. Data were collected with the inverse beam method in 2° oscillation frames with 0.3° overlap angles. Only fully recorded reflections were used. The Pb and Sm derivative data sets (and also native set 2 used in model refinement) were collected on a Raxis IIC image plate detector mounted on a RU200 rotating anode source operated at 50kV 100 mA with monochromatized CuKα radiation in 1.2° oscillation frames. Phasing calculations were performed with the PHASES package (*42*). The Se scattering factors were as described (*43*). The data set collected at wavelength L2 was chosen as native in the phasing process, the set collected at L1 was used as an isomorphous derivative, and the set collected at L3 as an anomalous scattering data set, based on the procedure of Ramakrishnan *et al.* (*44*); Pb and Sm derivatives were also included. Heavy atoms were located by Patterson and cross Fourier methods and their parameters were refined by maximum likelihood phase refinement. The final figure of merit was 0.681 at 2.75 (Å) resolution. Phases were improved by iterative solvent flattening (*45*). When the process converged, the R factor between observed and calculated structure factors obtained after map inversion was 0.212 at 2.75 (Å) resolution. The molecular model was built with O (*46*). The model was refined by several rounds of molecular dynamics and energy minimization by means of XPLOR (*47*) against the native data set 2, followed by manual rebuilding. The present model contains residues 56 to 140 and 154 to 208, and has no solvent molecules. Of the three methionine residues, the first (residue 50) is in a completely disordered region and invisible; the second (residue 154) is partially disordered. All calculations and modeling were performed on Silicon Graphics work stations.

Data set wavelength (Å)	Resolution (Å)	Detector source	Reflections						Phasing power‖	
			Total (N)	Unique (N)	Completeness* (%)	R_{sym}†	R_{Cullis}‡	R_{Kraut}§	Isomorphous	Anomalous
Se L2 0.9794 (native 1)	2.3	Fuji ×4A	74576	7079	76.3	0.050				
Se L1 0.9879	2.3	Fuji ×4A	77745	7250	78.2	0.047	0.480	0.014	2.55	
Se L3 0.9712	2.3	Fuji ×4A	81431	7303	78.5	0.059		0.022		3.01
$(CH_3)_3Pb$ acetate 1.5418	2.5	Raxis IIC RU200	50770	6789	92.0	0.040	0.713	0.116	0.74	1.13
SmAcetate 1.5418	2.8	Raxis IIC RU200	26767	5266	81.5	0.056	0.675	0.116	0.99	
Native 2 1.5418	2.5	Raxis IIC RU200	52552	6746	93.6	0.040				

Refinement	
Resolution (Å)	6.0–2.5
Atoms (N)	1059
Reflections $I > 2\sigma(I)$ (N)	5900
R factor (%)¶	22.7
R (free) (%)#	34.2
rms bond lengths (Å)	0.019
rms bond angles (°)	3.85

*Including data with $I > \sigma(I)$. †$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma \langle I \rangle$. ‡$R_{Cullis} = \Sigma \| \mid FPH_o \mid \pm \mid FP_o \| - \mid FH_c \| / \Sigma \| FPH_o \mid \pm \mid FP_o \|$ for centric reflections; §$R_{Kraut} = \Sigma \| FPH \mid - \mid FPH_c \| / \Sigma \mid FPH_o \mid$ for acentric reflections, isomorphous case; §$R_{Kraut} = \Sigma \| FPH_o^+ \mid - \mid FPH_c^+ \| + \| FPH_o^- \mid - \mid FPH_c^- \| / \Sigma (\mid FPH_o^+ \mid + \mid FPH_o^- \mid)$ for acentric reflections, anomalous case. FP is the protein, FPH is the derivative, and FH is the heavy atom structure factor, respectively. FPH^+ and FPH^- denote the Bijvoet mates. ‖The phasing power is defined as FH_c/E for the isomorphous case and $2FH_c''/E$ for the anomalous case, where E is the rms lack of closure. The R_{Cullis}, R_{Kraut}, and phasing power values are reported at 2.75 Å resolution. ¶ R factor = $\Sigma \mid FP_o - FP_c \mid / \Sigma FP_o$. #$R_{free}$ is computed on a randomly selected 10% of the data which were excluded from refinement (*48*).

There is precedence in other systems, such as Tn3 resolvase, for protomers in an active multimeric complex playing solely architectural roles (*32*). Alternatively, we cannot exclude the possibility of significant rearrangement in either the core domain or the DNA substrates in the active complex. It is possible to envision some unwinding of the host DNA, which might reflect distortions that have been observed at sites of integration in vitro (*33*). The structure and relative positioning of the missing NH_2- and COOH-terminal domains of integrase may reveal alternative interactions that bring the catalytic sites into a configuration consistent with the complete integration reaction. The active multimer of the prokaryotic MuA transposase is a tetramer, and all four active sites have been implicated in catalysis (*34*). It is also possible that the dimer interaction plays a role in the viral life cycle separate from the catalysis of integration, for example, in virus assembly.

It has become clear that retroviral DNA integration is mechanistically closely related to the transposition of many DNA elements, including certain prokaryotic transposons (*35*). The finding that a triad of acidic amino acid residues is conserved among these elements (*5–7*), combined with the result that mutations of these residues can abolish catalytic activity (*6–9*), is consistent with their playing a key role in catalysis. The involvement of acidic residues in catalysis of polynucleotidyl transfer has been noted for many nucleases and polymerases (*20–22, 36*). In the case of the 3′-5′ exonuclease activity of *E. coli* DNA polymerase I, it has been proposed that the function of the carboxylate groups is to position two divalent metal ions, one of which activates a water molecule for nucleophilic attack on a phosphorus atom in the DNA backbone, while the second divalent metal ion stabilizes the transition state (*37*). It has been suggested that retroviral integrase may use a similar reaction mechanism (*6*). Our finding that at least two, and most probably all three, conserved residues are in close proximity in the three-dimensional structure supports the view that they are directly involved in catalysis. The relative positions of these residues are therefore consistent with a role in divalent metal ion coordination.

The structure of the core domain of HIV-1 integrase is remarkably similar to that of RNase H (*20, 21*), the Holliday junction resolving enzyme RuvC from *E. coli* (*22*), and the catalytic domain of the MuA transposase protein (*23*). The finding that the catalytic domains of MuA transposase and HIV-1 integrase share structural similarities was somewhat anticipated due to their functional similarity. Conserved acidic amino acid residues in MuA transposase have been aligned with the retroviral D,D-35-E motif, and mutation of these residues can abolish transposition in vitro (*7*). However, the functions of RNase H and RuvC are not obviously related to transposition. Together with the previously noted ATPase domains (*24*), these enzymes provide yet another example of a superfamily of proteins with low sequence similarity that are topologically related (*38*). The active sites in integrase, RNase H, RuvC, and MuA are in part formed from residues extending from analogous structural elements. Thus, these four proteins share a common structural motif that catalyzes polynucleotidyl transfer for different biological roles.

The structure of the catalytic domain of HIV integrase reported here should facilitate the development of inhibitors to this viral target. Although the search for inhibitors of integrase is at an early stage compared to the more extensively studied reverse transcriptase and protease, suitable assay systems for large-scale screening are available (*39*), and compounds that inhibit the enzyme in vitro have been identified (*40*). The structures of complexes of integrase with these compounds, and others as they become available, can be expected to elucidate their mode of action and guide the design of improved inhibitors for use in antiviral therapy.

REFERENCES AND NOTES

1. S. P. Goff, *Annu. Rev. Genet.* **26**, 527 (1992).
2. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* **256**, 1783 (1992); A. Jacobo-Molina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6320 (1993); A. Wlodawer and J. W. Erickson, *Annu. Rev. Biochem.* **62**, 543 (1993).
3. P. A. Sherman and J. A. Fyfe, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5119 (1990); F. D. Bushman and R. Craigie, *ibid.* **88**, 1339 (1991); F. D. Bushman, T. Fujiwara, R. Craigie, *Science* **249**, 1555 (1990); C. Vink, D. C. van Gent, Y. Elgersma, R. H. A. Plasterk, *J. Virol.* **65**, 4636 (1991).
4. S. A. Chow, K. A. Vincent, V. Ellison, P. O. Brown, *Science* **255**, 723 (1992).
5. O. Fayet, P. Ramond, P. Polard, M. F. Prère, M. Chandler, *Mol. Microbiol.* **4**, 1771 (1990); S.-J. Rowland and K. G. H. Dyke, *ibid.*, p. 961; P. Radström *et al.*, *J. Bacteriol.* **176**, 3257 (1994).
6. J. Kulkosky, K. S. Jones, R. A. Katz, J. P. G. Mack, A. M. Skalka, *Mol. Cell. Biol.* **12**, 2331 (1992).
7. T. A. Baker and L. Luo, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6654 (1994); E. Kremenstova, L. Luo, T. A. Baker, personal communication.
8. A. Engelman and R. Craigie, *J. Virol.* **66**, 6361 (1992); D. C. van Gent, A. A. M. Oude Groeneger, R. H. A. Plasterk, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9598 (1992); A. D. Leavitt, L. Shiue, H. E. Varmus, *J. Biol. Chem.* **268**, 2113 (1993).
9. M. Drelich, R. Wilhelm, J. Mous, *Virology* **188**, 459 (1992).
10. A. Engelman, K. Mizuuchi, R. Craigie, *Cell* **67**, 1211 (1991).
11. C. Vink, E. Yeheskiely, G. A. van der Marel, J. H. van Boom, R. H. A. Plasterk, *Nucleic Acids Res.* **19**, 6691 (1991).
12. C. Vink, A. A. M. Oude Groeneger, R. H. A. Plasterk, *ibid.* **21**, 1419 (1993).
13. M. Schauer and A. Billich, *Biochem. Biophys. Res. Commun.* **185**, 874 (1992); K. Vincent, V. Ellison, S. A. Chow, P. O. Brown, *J. Virol.* **67**, 425 (1993).
14. F. D. Bushman, A. Engelman, I. Palmer, P. Wingfield, R. Craigie, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3428 (1993).
15. A. M. Woerner and C. J. Marcus-Sekura, *Nucleic Acids Res.* **21**, 3507 (1993).
16. A. Engelman, A. B. Hickman, R. Craigie, *J. Virol.* **68**, 5911 (1994).
17. A. B. Hickman, I. Palmer, A. Engelman, R. Craigie, P. Wingfield, *J. Biol. Chem.* **269**, 29279 (1994).
18. T. M. Jenkins, A. B. Hickman, R. Ghirlando, R. Craigie, unpublished material.
19. Residues 50 to 212 of HIV-1 integrase containing the single amino acid substitution of Lys for Phe^{185}, fused with a 20 amino acid NH_2-terminal histidine tag (Novagen), were expressed in *E. coli* and purified in a one-step process by Ni affinity chromatography. The affinity tag was removed by thrombin cleavage, and thrombin was removed by adsorption to benzamidine Sepharose 6B (Pharmacia). The resulting material migrated as a single band on SDS-polyacrylamide gels and, in the presence of 5 mM dithiothreitol (DTT), migrated almost entirely as a single band on IEF gels under nondenaturing conditions. Microscale gel filtration and dynamic light scattering measurements were used to find the best protein buffer conditions for initiating crystallization trials. Crystallization conditions were determined by means of a commercial version (Hampton Research) of the sparse matrix screen [J. Jancarik and S. H. Kim, *J. Appl. Cryst.* **24**, 409 (1991)]. The best crystals were grown at 4°C by the sitting drop vapor diffusion method. Protein (20 μl) at 6 mg/ml in 0.5 M NaCl, 20 mM tris-HCl, 5 mM DTT, 1 mM EDTA, pH 7.5, was mixed with 20 μl of the well solution containing 15 percent (w/v) PEG 8000 (Fluka), 0.1 M sodium cacodylate, 0.2 M ammonium sulfate, 5 mM DTT, pH 6.5. Some precipitate formed immediately, and crystals reached final size (0.3 mm) in several weeks. For data collection, the crystals were transferred to a protein-free buffer containing 0.25 M NaCl, 10 mM tris-HCl, 5 mM DTT, 50 mM sodium cacodylate, 0.1 M ammonium sulfate, 15 percent PEG 8000, pH 6.5. This buffer also contained 15 percent glycerol as a cryoprotectant. For structure determination, a selenomethionyl version of the protein was expressed in *E. coli* strain B834(DE3) in a medium containing selenomethionine at 40 mg/l [S. Doublié and C. W. Carter Jr, in *Crystallization of Nucleic Acids and Proteins*, A. Ducruix and R. Giegé, Eds. (Oxford Univ. Press, New York, 1992), pp. 311–317]. The purified protein was crystallized in an inert atmosphere under conditions identical to those for the native protein.
20. J. F. Davies II, Z. Hostomska, Z. Hostomsky, S. R. Jordan, D. A. Matthews, *Science* **252**, 88 (1991).
21. W. Yang, W. A. Hendrickson, R. J. Crouch, W. Satow, *ibid.* **249**, 1398 (1990); K. Katayanagi *et al.*, *Nature* **347**, 306 (1990); K. Katayanagi *et al.*, *J. Mol. Biol.* **223**, 1029 (1992); K. Katayanagi, M. Okumura, K. Morikawa, *Proteins* **17**, 337 (1993).
22. M. Ariyoshi *et al.*, *Cell* **78**, 1063 (1994).
23. P. A. Rice and K. Mizuuchi, personal communication.
24. P. J. Artymiuk, H. M. Grindley, K. Kumar, D. W. Rice, P. Willet, *FEBS Letters* **324**, 15 (1993).
25. M. A. Muesing *et al.*, *Nature* **313**, 450 (1985).
26. L. Ratner *et al.*, *ibid.* **313**, 277 (1985); S. Wain-Hobson, P. Sonigo, O. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985).
27. R. J. Crouch and M. L. Dirksen, in *Nucleases*, S. M. Linn and R. J. Roberts, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 211–241; R. J. Crouch, *New Biol.* **2**, 771 (1990); M. C. Starnes and Y. Cheng, *J. Biol. Chem.* **264**, 7073 (1989).
28. D. R. Davies, E. A. Padlan, S. Sheriff, *Annu. Rev. Biochem.* **59**, 439 (1990).
29. T.-H. Lin and D. P. Grandgenett, *Protein Eng.* **4**, 435 (1991).
30. A. Engelman, F. D. Bushman, R. Craigie, *EMBO J.* **12**, 3269 (1993); D. C. van Gent, C. Vink, A. A. M. Oude Groeneger, R. H. A. Plasterk, *ibid.* **12**, 3261 (1993).
31. K. A. Vincent, D. York-Higgins, M. Quiroga, P. O. Brown, *Nucleic Acids Res.* **18**, 6045 (1990); C. Vink *et al.*, *J. Virol.* **64**, 5626 (1990).
32. D. Sherratt, in *Mobile DNA*, D. E. Berg and M. M.

Howe, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 163–184.
33. D. Pruss, F. D. Bushman, A. P. Wolffe, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5913 (1994); H.-P. Müller and H. E. Varmus, *EMBO J.* **13**, 4704 (1994).
34. T. A. Baker, M. Mizuuchi, H. Savilahti, K. Mizuuchi, *Cell* **74**, 723 (1993); T. A. Baker, E. Kremenstova, L. Luo, *Genes Dev.* **8**, 2416 (1994).
35. K. Mizuuchi, *Annu. Rev. Biochem.* **61**, 1011 (1992).
36. C. M. Joyce and T. A. Steitz, *ibid.* **63**, 777 (1994); S. M. Linn, R. S. Lloyd, R. J. Roberts, Eds., *Nucleases* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1993); J. F. Davies II, R. J. Almassy, Z. Hostomska, R. A. Ferre, Z. Hostomsky, *Cell* **76**, 1123 (1994); M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, J. Kraut, *Science* **264**, 1930 (1994).
37. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988); L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991).
38. A. G. Murzin and C. Chothia, *Curr. Opinion Struct. Biol.* **2**, 895 (1992).
39. R. Craigie, K. Mizuuchi, F. D. Bushman, A. Engelman, *Nucleic Acids Res.* **19**, 2729 (1991); D. J. Hazuda, J. C. Hastings, A. L. Wolfe, E. A. Emini, *ibid.* **22**, 1121 (1994); C. Vink, M. Banks, R. Bethell, R. H. A. Plasterk, *ibid.*, p. 2176.
40. M. Cushman and P. Sherman, *Biochem. Biophys. Res. Commun.* **185**, 85 (1992); S. Carteau, J. F. Mouscadet, H. Goulaouic, F. Subra, C. Auclair, *ibid.* **192**, 1409 (1993); M. R. Fesen, K. W. Kohn, F. Leteurtre, Y. Pommier, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2399 (1993); A. Mazumder, D. Cooney, R. Agbaria, A. Gupta, Y. Pommier *ibid.* **91**, 5771 (1994).
41. Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Bailey, Eds. (Science and Engineering Research Council, Warrington, United Kingdom, 1993), pp. 56–62.
42. W. Furey and S. Swaminathan, *Phases—A program package for the processing and analysis of diffraction data for macromolecules*. Poster, American Crystallographic Association meeting, 1990.
43. H. Wu, J. W. Lustbader, Y. Liu, R. E. Canfield, W. A. Hendrickson, *Structure* **2**, 545 (1994).
44. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet, *Nature* **262**, 219 (1993).
45. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
46. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Cryst.* **A47**, 110 (1991).
47. A. T. Brünger, *X-PLOR Version 3.1. A system for X-ray crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
48. A. T. Brünger, *Nature* **355**, 472 (1992).
49. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
50. M. Carson, *ibid.*, p. 958.
51. We thank P. Rice and K. Mizuuchi for advice; C. Ogata and P. Rice for assistance in collecting the MAD data at NSLS, Brookhaven; R. Ghirlando for determining the dynamic light scattering properties of the protein; and S. Landry for directing our attention to the similarity between RNase H and ATPase domains. The structural study reported here was based on a long-term effort on the part of a number of people. In particular, we thank F. Bushman, P. Wingfield, and I. Palmer for developing protein purification protocols and supplying a number of integrase derivatives in the early stages of the project; P. Sun for initial crystallization attempts; and M. Carmichael, S. Hosseini, and R. Madabhushi for technical assistance. Supported in part by the NIH Intramural AIDS Targeted Antiviral Program. Coordinates of the structure at the present stage of refinement have been deposited in the Protein Data Bank under accession number 1ITG/T5588.

13 October 1994; accepted 15 November 1994

Splicing of the *rolA* Transcript of *Agrobacterium rhizogenes* in *Arabidopsis*

Armando Magrelli, Kerstin Langenkemper, Christoph Dehio,* Jeff Schell, Angelo Spena†

The *rolA* gene encoded on the Ri plasmid A4 of *Agrobacterium rhizogenes* is one of the transferred (T_L-DNA) genes involved in the pathogenesis of hairy-root disease in plants. The function of the 100–amino acid protein product of *rolA* is unknown, although its expression causes physiological and developmental alterations in transgenic plants. The *rolA* gene of *A. rhizogenes* contains an intron in its untranslated leader region that has features typical of plant pre–messenger RNA introns. Transcription and splicing of the *rolA* pre–messenger RNA occur in the plant cell.

The *rolA* gene from the Ri plasmid A4 of *A. rhizogenes* is one of the T_L-DNA genes transferred from the bacterium to the plant, and it is involved in the pathogenesis of hairy-root disease (*1*). Although transferred DNA (T-DNA)–encoded genes of *A. tumefaciens* can be transcribed and translated in bacterial extracts (*2*), we are not aware of data reporting bacterial transcription of T_L-DNA–encoded genes of *A. rhizogenes*. Expression takes place in transformed plant cells, and the *rolA* gene by itself causes plant developmental alterations, including dwarfism (due to reduced growth and internode distance) and wrinkled leaves (due to reduced growth of the midrib and of vascular tissue) (*3*).

Conflicting results concerning the initiation of transcription of the *rolA* gene [mapped by primer extension to position −29 from the ATG initiation codon (*4*) or to position −100 (*5*)] led us to analyze the structure of the transcript in more detail. We did reverse transcription–polymerase chain reaction (RT-PCR) with polyadenylated [poly(A)$^+$] RNA extracted from *Arabidopsis thaliana* plants transgenic for the *rolA* gene (line 23) (*5*), using as a 5′ primer an oligonucleotide spanning nucleotides −100 to −82 (primer a in Fig. 1) and as 3′ primer an oligonucleotide spanning nucleotides +300 to +282 of the *rolA* gene (primer b in Fig. 1). The polyadenylation site is at position +530 (*5*). The PCR produced fragments of two lengths, differing by 76 nucleotides (Fig. 1). DNA sequence analysis indicated that, in *Arabidopsis*, *rolA* transcripts have either an untranslated leader region (ULR) of at least 100 bases (I in Fig. 1) or a ULR of 24 bases (II in Fig. 1). The shorter class of transcripts is identical to the long one except for the deletion of 76 bases in the ULR of the *rolA* mRNA. The deletion starts with the sequence GT at position −76 and ends with AG at position −3. The dinucleotides GT (at the 5′ end) and AG (at the 3′ end) are known to delimit eukaryotic introns and to be invariable parts of the splice sites (*6*). Thus, the two classes of RT-PCR products represent the unspliced and spliced mRNA of the *rolA* gene. The reported discrepancy could be partially explained by the fact that mapping experiments have used poly(A)$^+$ RNA extracted from tobacco (*4*) and *Arabidopsis* (*5*) plants. The efficiency of *rolA* pre–messenger RNA (pre-mRNA) splicing differs in these two plant species (*7*).

To further characterize the spliced *rolA* mRNA, 12 independent RT-PCR products, corresponding to spliced mRNAs, were cloned and sequenced. Ten clones had a ULR of 27 bases (III in Fig. 1B), whereas in the remaining two clones, the ULR was 24 bases long (II in Fig. 1B). This result defines two classes of spliced transcripts that were generated by the use of alternative 5′ splice sites: the GT at position −73 for class I and the GT at position −76 for class II. Consequently, splicing of *rolA* pre-mRNA removes an intron of 73 to 76 nucleotides, which is in agreement with a minimum length of 70 to 73 nucleotides reported for efficient splicing of introns in plants (*8*). Furthermore, the *rolA* intron has a 71% AT content and the AG 3′ splice site is preceded by a T-rich region, features considered typical of plant pre-mRNA introns (*9*). The two 5′ exon-intron junctions show homology to the plant consensus sequence ($G_{72}G_{100}T_{99}A_{70}A_{55}G_{65}T_{49}$) (*6*).

Many mutant alleles are caused by mutations that interfere with RNA splicing (*10*). Thus, to confirm the molecular data, *rolA* alleles from seven independent null mutants isolated by ethylmethane sulfonate mutagenesis of an *Arabidopsis* line transgenic for the *rolA* gene (line 23) (*5*) were cloned by PCR (5′ primer: AACGCTTCAATACGGTGAG; 3′ primer: AATACGCACGTGGCTGGCGGTCTT) and sequenced. Four out of seven were single point mutations leading either to amino acid substitutions [Arg at position 37 to Trp in mutant line 23-4(1); Pro at position 40 to Ser in mutant line 23-1; Pro at position 40 to Leu in mutant line 23-6] or to change

Max-Plank-Institut für Züchtungsforschung, Carl-von-Linné Weg 10, 50829 Cologne, Germany.

*Present address: Institut Pasteur, 28 rue du Dr Roux, 75724 Paris, Cedex 15, France.
†To whom correspondence should be addressed.