Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Structural studies of the catalytic core of the primate foamy virus (PFV-1) integrase

Stéphane Réty,[a]*‡ Lenka Řežábková,[a]§ Barbara Dubanchet,[a]¶ Jan Šilhán,[a]‡‡ Pierre Legrand[b] and Anita Lewit-Bentley[a]*§§

[a]Laboratoire de Biologie et Pharmacologie Appliquée, CNRS UMR8113, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan, France, and [b]Synchrotron SOLEIL, L'Orme de Merisiers, BP 48 St Aubin, 91192 Gif-sur-Yvette, France

‡ Present address: Laboratoire de Cristallographie et RMN Biologiques, CNRS UMR 8015, Faculté de Pharmacie, 4 Avenue de l'Observatoire, 75270 Paris CEDEX 06, France.
§ Present address: Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 8/2030, 128 43 Prague 2, Czech Republic.
¶ Present address: Institut Pasteur de Lille, INSERM U1011, 1 Rue du Professeur Calmette, 59000 Lille, France.
‡‡ Present address: Division of Molecular Biosciences, Imperial College London, South Kensington Campus, London SW7 2AZ, England.
§§ Present address: Unité d'Immunologie Structurale, CNRS URA 2185, Institut Pasteur, 25 Rue du Dr Roux, 75724 Paris, France.

Correspondence e-mail: stephane.rety@parisdescartes.fr, alewit@pasteur.fr

Received 8 April 2010
Accepted 14 June 2010

PDB References: catalytic core of the primate foamy virus integrase, wild type, 2x78; 3M mutant, 2x74; Mn complex, 2x6n; Mg complex, 2x6s.

Retroviral integrases are vital enzymes in the viral life cycle and thus are important targets for antiretroviral drugs. The structure of the catalytic core domain of the integrase from human foamy virus, which is related to HIV-1, has been solved. The structure of the protein is presented in two different crystal forms, each containing several molecules in the asymmetric unit, with and without the essential manganese or magnesium ion, and the structures are compared in detail. This allows regions of high structural variability to be pinpointed, as well as the effect of divalent cations on the conformation of the catalytic site.

1. Introduction

The transcription of the genetic information contained in viral RNA and the integration of the resulting DNA into host DNA is a vital step in the life cycle of all retroviruses. After its entry into the target cell, the RNA contained in the retroviral particle is reverse-transcribed into a linear viral DNA molecule (Lewinski & Bushman, 2005). Human foamy virus (PFV-1) belongs to the spumavirus genus of the retroviridae family (Delelis *et al.*, 2004). Unlike the related lentiviruses (which include HIV), the infectious genome of PFV-1 has clearly been demonstrated to be the viral DNA contained in viral particles (Linial, 1999), even though it has been suggested that the viral RNA could be reverse-transcribed early after infection (Delelis *et al.*, 2003). Retroviral infection is characterized by an obligatory integration step of the double-stranded viral DNA by the viral integrase (IN). In the first step, two nucleotides from both extremities of the linear viral DNA are removed by 3′-processing. The newly exposed 3′-OH groups of the canonical CA dinucleotides are then involved in the integration process, which is also catalysed by IN.

IN is characterized *in vitro* by three enzymatic activities: 3′-processing, strand transfer, a reaction that mimics the integration step in the infected cell, and disintegration, which has only been described in *in vitro* experiments (Delelis, Carayon, Saïb *et al.*, 2008). All retroviral INs are organized into three domains. The first domain, the N-terminal domain (NT), carries an HHCC motif which is related to zinc-finger domains and may be involved in oligomerization of the protein (Lewinski & Bushman, 2005). The second domain, the catalytic core (CC), encompasses the catalytic triad D-D35E responsible for the overall activities of the protein. The third domain, the C-terminal domain (CT), binds DNA in a nonspecific manner and is therefore mainly involved in stabilization of the complex between the protein and DNA. All integrase activities require the presence of divalent metal cations coordinated by residues of the catalytic triad (Asp64, Asp116 and Glu152 for HIV-1 IN). It is important to note that only the full-length protein is able to carry out all these activities in the presence of either Mn^{2+} or Mg^{2+}. In the case of HIV-1 IN, the protein with the NT and/or CT domains deleted is only able to perform the disintegration reaction in the presence of Mn^{2+}.

Despite years of effort, the three-dimensional structure of the entire HIV-1 IN has remained elusive. PFV-1 IN, while possessing

only 15% sequence identity to HIV-1 IN, shows very similar activity and can be inhibited by common anti-HIV-1 compounds. However, its solution properties are quite different from those of HIV-1 IN: it is highly soluble and thus much easier to study (Delelis, Carayon, Guiot *et al.*, 2008). Indeed, the structure of PFV-1 IN in complex with a DNA oligonucleotide has recently been solved (Hare *et al.*, 2010). Here, we report the crystal structure of the PFV-1 catalytic domain obtained both in the absence of divalent cations and in the presence of Mn^{2+} and Mg^{2+}.

2. Experimental procedures

2.1. Cloning and mutagenesis

The PFV catalytic domain (amino acids 110–309) was amplified by PCR using oligonucleotide primers 5′-CATG**CATATG**GGTCCTATTCTAAGACCAGATAGG-3′ and 5′-CCGCTCGAGCTAAGAGGAGG**CTGGAG**GGGTGGATGG-3′ and HotStart HiFi *Taq* DNA polymerase (Qiagen). The PCR product was digested by *Nde*I and *Bam*HI (NEB) and cloned in pET-15b vector (Novagen) between *Nde*I and *Xho*I restriction sites. The 3M mutant was obtained by sequentially introducing methionines using site-directed mutagenesis. Methionines were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) at positions 127 (I127M), 227 (I227M) and 253 (L253M) using the oligonucleotides 5′-CCTTTTGATAAATTCTTT**ATG**GACTATATTGGACCTTTGC-3′, 5′-GGAAAGGAAAAATAGTGAT**ATG**AAACGACTTTTAACTAAACTGC-3′ and 5′-CCTGTTGTACAACTTGCT**ATG**AACAACACCTATAGCC-3′, respectively.

2.2. Expression and purification

The plasmid encoding the integrase catalytic core domain was transformed into *Escherichia coli* strain BL21 (DE3) (Novabiogen). Protein expression was induced by adding 1 m*M* IPTG to exponentially growing cultures, which were then grown for a further 3 h at 303 K. The pellet was resuspended in 50 m*M* NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 20 m*M* imidazole (buffer *A*) supplemented with 10 mg ml^{-1} lysozyme and the suspension was sonicated. The sonicate was clarified by centrifugation and the supernatant was combined with 5 ml Ni–NTA (Qiagen) and allowed to equilibrate for 2 h at 277 K. After extensive washing with buffer *A*, the protein was eluted with 50 m*M* NaH_2PO_4 pH 7.0, 150 m*M* NaCl, 5% glycerol, 500 m*M* imidazole. The fractions containing the protein were pooled, 1 m*M* EDTA was added to the solution and the protein was purified by gel-filtration chromatography on a Superdex 75 column (Amersham Biosciences) using 50 m*M* MES pH 5.8, 150 m*M* NaCl. The final solution was adjusted to 50 m*M* NaCl by adding the appropriate volume of 50 m*M* MES pH 5.8.

2.3. Crystallization

Crystallization conditions were screened with Crystal Screen from Hampton Research using the hanging-drop vapour-diffusion technique. Bipyramidal crystals appeared in several conditions containing salts as precipitants (chloride, sulfate, acetate or phosphate), with the best conditions being 5 mg ml^{-1} protein mixed in a 1:1 ratio with 0.1 *M* sodium citrate pH 5.6 and 0.6 *M* ammonium sulfate. However, these crystals turned out to be severely multiple, making structure determination impossible. Using PEG 8000 in low-pH buffer (citrate or cacodylate), low ionic strength and after screening for additives to improve the crystal quality, the best conditions found were 14% PEG 8000, 100 m*M* sodium citrate pH 5, 10% glucose, 5% MPD mixed in a 1:1 ratio with protein at 4 mg ml^{-1}. The crystals (wt) belonged to space group $C222_1$, with unit-cell parameters a = 51.80, b = 98.04, c = 240.32 Å and three molecules per asymmetric unit. These crystals were less frequently multiple and allowed data collection from a native form.

The methionine/selenomethionine mutant (3M; see below) failed to crystallize under these conditions and a new search was undertaken. The only condition found was optimized to 1.8–2 *M* ammonium formate with 100 m*M* HEPES pH 7.5 and 10–15% glycerol. The Mn^{2+}-complexed (3M-Mn) and Mg^{2+}-complexed (3M-Mg) crystals were obtained under the same conditions with 5 m*M* $MnCl_2$ or $MgCl_2$ added to the buffer as well as to the protein, which was then incubated for several hours prior to setting up crystallization. These crystals belonged to space group $P2_12_12_1$, with unit-cell parameters a = 84.77, b = 89.23, c = 177.08 Å and six molecules per asymmetric unit.

Since the crystallization buffers contained either glucose and MPD or glycerol, the crystals were frozen directly from the crystallization drops by plunging them into liquid nitrogen.

2.4. Data collection and processing

All data were collected at 173 K on synchrotron beamlines after screening on a laboratory source. Data for the wt, 3M-Mn and 3M-Mg crystals were collected on beamlines ID14-3, ID23-2 and ID14-4, respectively, at the ESRF (Grenoble, France). Anomalous data from the SeMet crystals and 3M-Mn crystals were collected on PROXIMA 1 at SOLEIL (St Aubin, France). All data were treated with *XDS* (Kabsch, 2010), followed by careful scaling with *SCALA* from the *CCP*4 program package (Collaborative Computational Project, Number 4, 1994).

2.5. Phasing and structure determination

The wild-type protein crystallized in space group $C222_1$. All attempts at molecular replacement using models of retrovirus integrase CC structures known at the time failed. A heavy-atom derivative search was unsuccessful for the wt protein, while its SeMet derivative produced systematically multiple crystals and thus could not provide a useful anomalous signal.

Finally, three residues were mutated to methionines in order to guarantee a good signal from an SeMet derivative: I127M, I227M and L253M (the 3M mutant). Data collected from SeMet-derivative crystals at the peak wavelength to 2.8 Å resolution were sufficient for automatic phasing and chain tracing using the *SOLVE*/*RESOLVE* program package (Terwilliger & Berendzen, 1999; Terwilliger, 2003) with sixfold noncrystallographic symmetry. The wt structure was solved by molecular replacement with *PHENIX* using a monomer of the 3M mutant chain trace as template. All structures were subjected to several cycles of rebuilding with *Coot* (Emsley *et al.*, 2010) and refinement with *REFMAC*5 from the *CCP*4 program package (Murshudov *et al.*, 1997).

3. Results and discussion

3.1. Structure determination

We have determined the crystal structure of the catalytic core domain of PFV-1 integrase in both the presence and the absence of the divalent cations Mg^{2+} and Mn^{2+} (Table 1). Initial attempts to solve the structure of the wt protein crystals by molecular replacement failed. The lack of success may be the consequence of a sequence identity of only 15% to lentiviral integrase CCs for which structures

were available. Indeed, the r.m.s. difference on C^{α} coordinates between HIV integrase CC (PDB code 1bis; Goldgur *et al.*, 1998) and our refined structure is 2.1 Å. Furthermore, the number of molecules per asymmetric unit and their arrangement was not clear from a preliminary data analysis. All retroviral integrase catalytic cores known to date form dimers, which led us to expect two or four molecules per asymmetric unit. However, the most likely Matthews coefficient (2.26 $Å^3$ Da^{-1}) corresponds to the presence of three molecules in the asymmetric unit, suggesting a more unusual arrangement.

In a postmortem analysis of the wt crystal data, we compared molecular replacement using our refined coordinates and the high-resolution structure of the ASV catalytic core (PDB code 1vsf; Bujacz *et al.*, 1996). The rotation function gave three peaks above background for the cognate coordinates, but none for the ASV coordinates (data not shown). With hindsight it is also possible to relate the twofold noncrystallographic axes found by the self-rotation function to the relative orientation of the three molecules in the asymmetric unit, but only two have a signal above 50% of the origin peak.

Two of the three molecules of the wt crystal (*A* and *B*) form a dimer related by noncrystallographic twofold symmetry, while the third molecule (*C*) forms a dimer across a crystallographic twofold axis (Fig. 1*a*). This type of arrangement of biological dimers has occasionally been observed before, in particular in one HIV-1 CC crystal form (Maignan *et al.*, 1998). However, in our structure one molecule (*A*) of the noncrystallographic dimer differs significantly from the other two (Fig. 2): the overall r.m.s.d. on C^{α} atoms is 0.4 Å between molecules *B* and *C* but as high as 1.9 Å between molecule *A* and the other two molecules. In comparison, the r.m.s.d. of C^{α} atoms of the six chains in the 3M mutants is only 0.4 Å. Taken together, all these factors made phasing by molecular replacement very difficult.

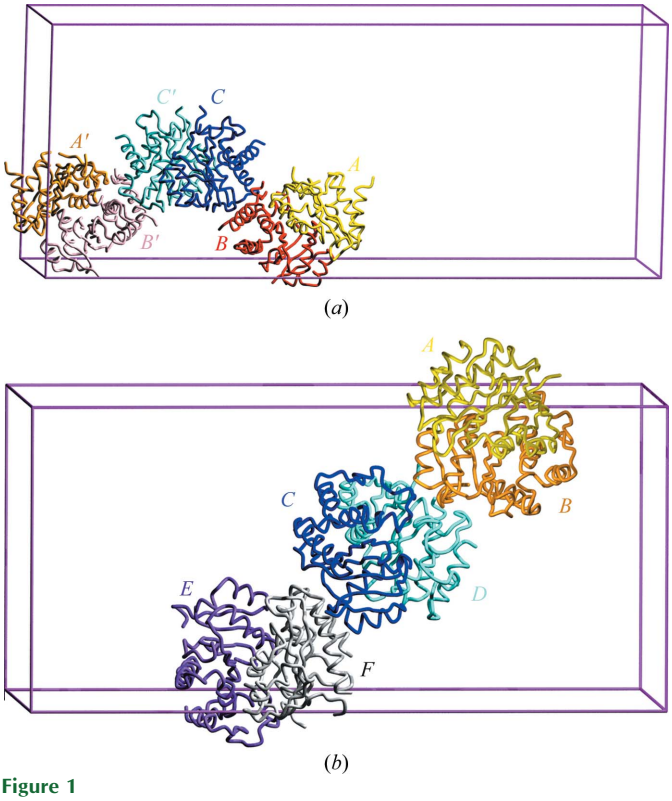

Figure 1
Arrangement of individual molecules in the crystal lattice of (*a*) the wt form ($C222_1$) and (*b*) the 3M form ($P2_12_12_1$). All figures were prepared with *CCP*4*MG* (Potterton *et al.*, 2004).

Table 1
Data, phasing and refinement statistics.

Values in parentheses are for the highest resolution shell.

(*a*) Data statistics.

Crystal	wt	3M-Se (peak)	3M-native	3M-Mn	3M-Mn (ano)	3M-Mg
Max. resolution (Å)	2.0	2.8	2.34	2.06	2.95	2.29
Wavelength (Å)	0.931	0.979	1.033	0.933	1.892	0.9835
R_{merge}	0.068 (0.499)	0.083 (0.339)	0.047 (0.356)	0.050 (0.485)	0.133 (0.587)	0.055 (0.412)
R_{anom}	n.a.	0.053	n.a.	n.a.	0.123	n.a.
$\langle I\rangle/\langle\sigma(I)\rangle$	14.2 (2.0)	26.5 (5.0)	15.1 (2.0)	14.2 (2.0)	6.2 (1.9)	15.9 (2.0)
Completeness (%)	91.2 (53.9)	99.8 (99.2)	99.0 (95.4)	93.8 (71.0)	99.0 (98.0)	97.4 (89.3)
Redundancy	6.6 (3.0)	14.6 (14.6)	3.8 (2.8)	3.5 (2.7)	3.4 (3.2)	4.5 (2.9)
Wilson *B* ($Å^2$)	29.4	41.8	51.1	32.7	64.8	44.1

(*b*) Phasing statistics (*SOLVE*).

No. of sites found	23
Resolution (Å)	30.0–2.8
Phasing power	1.652
Mean figure of merit	0.33
Mean anomalous lack-of-closure error	9.8

(*c*) Refinement statistics.

Crystal	wt	3M-native	3M-Mn	3M-Mg
Resolution (Å)	2.0	2.34	2.06	2.3
R†	0.212	0.2328	0.2257	0.2249
R_{free}	0.272	0.2937	0.2700	0.2781
Non-H atoms	4265	8882	8742	8743
Waters	364	188	316	265
Ions	n.a.	n.a.	6	6
$\langle B\rangle$ ($Å^2$)	36	57.9	45.0	48.4
R.m.s.d. bonds (°)	0.0107	0.0103	0.0078	0.0088
R.m.s.d. angles (°)	1.217	1.318	1.085	1.192
R.m.s.d. planes (°)	0.006	0.007	0.005	0.005
Ramachandran‡				
Most favoured (%)	97.7	97	97.7	97.3
Additionally allowed (%)	2.3	3	2.3	2.7

† $R = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. R_{free} is defined in Brünger (1992); 5% of data were used. ‡ The percentage of peptide bonds in the most favoured and additional allowed region of the Ramachandran plot as defined by *RAMPAGE* (Lovell *et al.*, 2002).

When all attempts at obtaining heavy-atom derivatives by soaking the wt crystals failed, we prepared an SeMet derivative (there is only one Met in the native sequence). This protein gave crystals that were systematically too multiple to provide a reliable anomalous signal. We therefore mutated a further three residues to methionines and prepared a new SeMet protein. We selected Ile127, Ile227 and Leu253, which were predicted to lie in conserved helices by sequence alignment. The mutant protein (3M) expressed well in soluble form, indicating that the mutations did not perturb the overall fold of the protein. This protein crystallized in a new unrelated crystal form which was free of twinning or multiplicity problems (Fig. 1*b*).

3.2. Crystal structure of the catalytic core of PFV-1 integrase

The predicted CC domain (Lee *et al.*, 2005) used for the crystallization trials comprises residues 109–313 of PFV-1 integrase. However, in all of the crystal forms obtained the first 8–10 residues at the N-terminus and the C-terminal 8–9 residues could not be located in electron density. The overall fold of the protein is very similar to that observed in the PFV-1 CC structure determined at the same time (Valkov *et al.*, 2009) and, as in other retroviral integrase CC domains, corresponds to the RNase fold. The pattern of secondary-structure

elements is thus the same as found in lentiviral integrase CC domains, but the PFV-1 protein is more extended at the C-terminus, covering the equivalent of the region linking the HIV-1 CC to the CT domain. The C-terminal helix (α7) is folded against the protein core as in Valkov *et al.* (2009), but in our structure there are no breaks in the loop connecting this helix, alowing us to trace the entire chain in all of the crystal forms.

3.3. Comparison of the wt and 3M mutant structures

A small but important change in the angle between helices α4 and α5 is observed in the 3M mutant form compared with the wild type, which probably accounts for the change in space group ($C222_1$ to $P2_12_12_1$) as well as that in crystal packing. In the 3M crystal forms there are six molecules in the asymmetric unit, forming three dimers, which all show subtle differences in conformation. The catalytic loop connecting β5 and α4 is often disordered in retroviral integrase CC crystal structures, with residue Glu221, which lies at the N-terminal end of helix α4, being poorly defined. We could only trace the complete course of this loop in the *A* molecules of divalent-ion-containing crystals. Indeed, we observed a wide range of stability (*i.e.* electron-density quality) of this loop in the 21 independent molecules in our four crystal forms (Fig. 3). In the wt crystal structure the loop is the most disordered, while in several of the molecules in the divalent-cation-bound crystals the gaps are much smaller.

We introduced three methionines in order to solve the structure by SeMet MAD (I127M, I227M and L253M). Leucine-to-methionine substitution is considered to be a safe mutation with minor structural changes (Bordo & Argos, 1991), while isoleucine-to-methionine substitution can have a destabilizing effect (Gassner *et al.*, 2003; Ohmura *et al.*, 2001). This is partly a consequence of the change in hydrophobicity and the greater flexibility of the methionine side chain. Methionines 227 and 253 are in helices α4 and α5, respectively, with their side chains pointing towards each other. While the position and orientation of helix α5 is very similar in the wt and 3M crystals, helix α4 is shifted to allow the packing of the slightly larger volume of the Met side chain compared with the Ile and Leu side chains (Fig. 4). Upon superposition of the two structures with alignment about the catalytic centre, there is a very subtle displacement of helix α5 (about 0.5 Å at the C^α atom of residue 227). Helix α4, on the other hand, is shifted by an entire turn at residue 253 in the mutant compared with the wt structure (3.8 Å between C^α atoms). Helix α6 is very similar in both crystal forms, but the large loop connecting it to helix α7, as well as the C-terminal helix itself, are again shifted (C^α distance of 1.8 Å at residue 300). Interestingly, helix α4 in our wt structure (as well as in the structure published by Valkov *et al.*, 2009) follows a somewhat different path compared with the structures of the CC from ASV and HIV-1, with an angle of 20.6° between them. In spite of this, helix α5 superposes well in the PFV-1 and ASV/HIV CC structures.

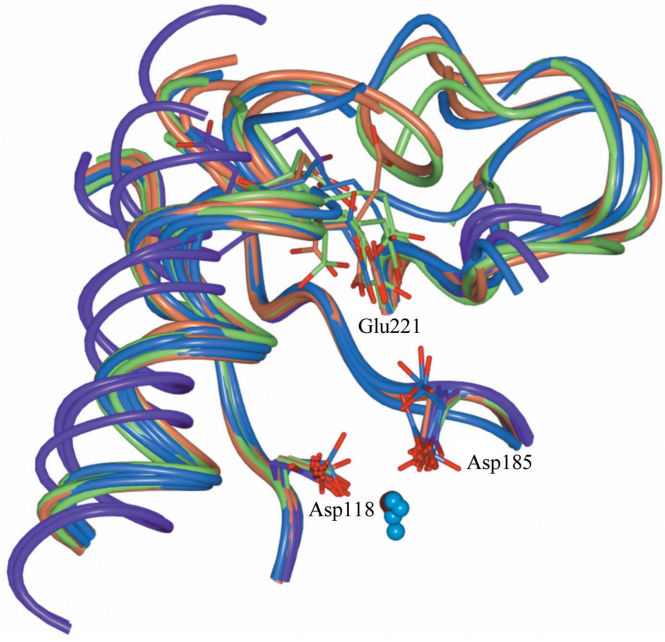

Figure 3
Loop connecting β5 to α4. A superposition of C^α traces for all 21 independent molecules: three wt molecules in purple and six copies of each 3M molecule: no ion, blue; Mg, coral; Mn, green. Catalytic residues Asp118, Asp185 and Glu221 are highlighted (note the double conformations for Asp185 of the 3M mutant in the absence of divalent cations). All Mn^{2+} ions (chocolate) superpose upon each other, while the positions of the Mg^{2+} ions (blue) are more divergent.

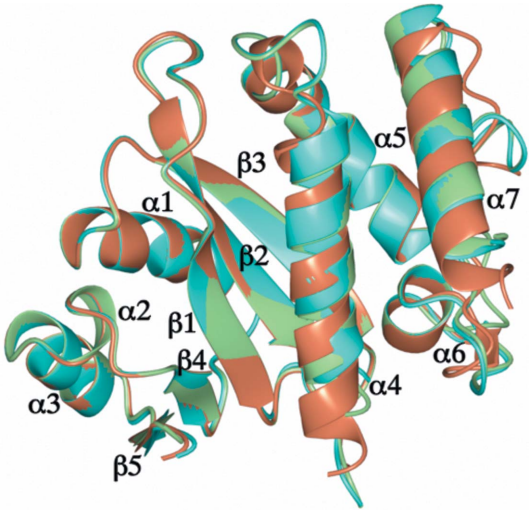

Figure 2
Superposition of the three independent molecules in the wt crystal asymmetric unit: molecule *A* in blue, molecule *B* in green and molecule *C* in coral.

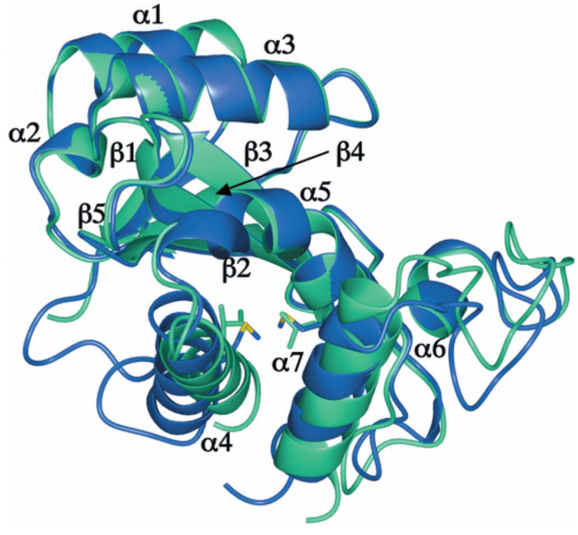

Figure 4
Superposition of the wt (green) and 3M mutant (blue) structures. The side chains of residues 227 and 253 are highlighted.

3.4. Catalytic site: divalent cations

We were not able to introduce divalent cations into the wt crystal form by either cocrystallization or soaking experiments. The conformation of the catalytic site reflects the absence of a divalent cation: in molecules *B* and *C* the Glu121 side chains point away from the catalytic site, while in molecule *A* its density is very poorly defined.

However, we were able to obtain crystals of the 3M mutant protein in the presence of Mg^{2+} or Mn^{2+}, as well as without any ion, under otherwise identical crystallization conditions. This is the first time that Mn^{2+} cations have been introduced into an IN CC structure by cocrystallization rather than by soaking a preformed crystal. The presence and type of divalent ion has a significant effect on the stability of the catalytic site as well as the crystal quality. One divalent cation site is visible in the catalytic site of all molecules in both the Mg^{2+} and Mn^{2+} crystals, but the electron density and coordinating water molecules are different (Fig. 5 shows examples for both ions in molecule *E*). The Mn^{2+}-ion positions were verified by an anomalous difference map using data collected at the Mn absorption-edge wavelength and the anomalous peak height again differs for individual molecules (Table 2).

Upon superposition of the catalytic loops for all molecules in all mutant crystal forms (Fig. 3), as well as from the quality of the electron-density maps, it is obvious that the Mn^{2+} sites are the most stable and well defined. The positions of the Mn^{2+} ion in all six molecules overlap well and the coordination spheres are complete for four of them. On the other hand, the Mg^{2+} ion could not be positioned with confidence in molecules *C* and *D*. Indeed, in the case of Mg^{2+} only molecule *A* has a completely ordered ion site. The active site of PFV-1 IN is analogous to other retroviral integrases and is formed by the catalytic triad D-D35E motif around a divalent cation. The loop connecting catalytic residues Asp118 and Glu221 is very flexible and is highly disordered in most crystal structures of retroviral integrase catalytic cores. The presence of divalent cations

Table 2
Characteristics of the Mn^{2+} and Mg^{2+} ions in all molecules of the 3M crystal forms (see text).

Ion	*B* factor (Å^2)	Anomalous peak (σ)	No. of H_2O ligands
Mn1	31.2	10.6	4
Mn2	55.2	10.4	1
Mn3	60.4	8.0	4
Mn4	85.5	4.5	0
Mn5	44.4	9.5	4
Mn6	41.9	8.8	4
Mg1	46.0	—	4
Mg2	67.6	—	1
Mg3	77.2	—	1
Mg4	80.7	—	1
Mg5	52.9	—	3
Mg6	56.2	—	3

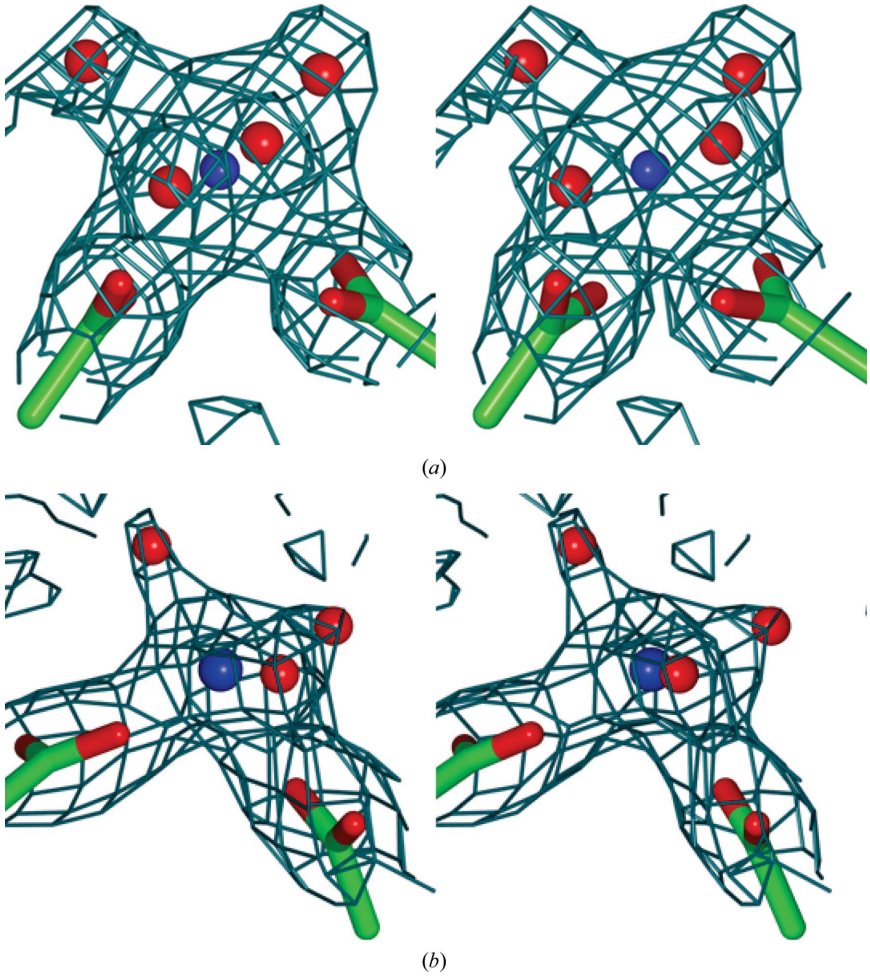

Figure 5
Detail of the catalytic ion. (*a*) Mn^{2+} and (*b*) Mg^{2+} with their ligands in molecule *E*. Weighted F_{obs} maps were contoured at 1.2σ. Metal ions are shown in blue, water molecules in red and the side chains of Asp118 and Asp185 are depicted as cylinders.

stabilizes the catalytic loop conformation, with Mn^{2+} being more effective than Mg^{2+}. Interestingly, the PFV-1 integrase showed a preference for Mn^{2+} in the endonuclease and integration reactions (Liu *et al.*, 1977; Pahl & Flügel, 1993). The effect of these two ions has been compared for the disintegration reaction (O. Delelis, personal communication): while Mg^{2+} showed a behaviour similar to HIV integrase (with activity peaking at 10 m*M* Mg^{2+} and then decreasing; Delelis, Carayon, Saïb *et al.*, 2008), in the presence of Mn^{2+} ions the disintegration activity is much higher and does not saturate even at a tenfold higher ion concentration.

The recent structure of the entire PFV-1 integrase (Hare *et al.*, 2010) was obtained in the presence of a DNA oligonucleotide that mimics the viral DNA. The complex is composed of two protein dimers but only one DNA duplex since the target oligonucleotide is not present. The integrase dimer that is not engaged in a nucleic acid complex is less well ordered and binds only one divalent ion in its catalytic site. We can make an interesting observation by superimposing the catalytic site of our CC structures upon those of the two dimers in the complex. For the DNA-free dimer our wt structure superposes very well, while the mutant shows the displacement of helix α4 mentioned above. The catalytic loops, nevertheless, follow a similar path. Only one divalent ion is present in the DNA-free dimer, similar to that of our 3M Mn^{2+} structure. On the other hand, when DNA is engaged by the integrase dimer a second divalent ion is bound that is absent in our structures and the entire catalytic loop changes position significantly from its position in the free CC.

3.5. C-terminal helix

The PFV-1 integrase catalytic core domain (CC) definition was proposed by Engelman & Craigie (1992) from a multiple sequence alignment of all retroviral integrase sequences then available as comprising residues 106–257. Nevertheless, a more recent paper by Lee *et al.* (2005) defined the domain as terminating at Pro309, a delimitation used by Valkov *et al.* (2009) as well as in our work.

However, the CC crystal structures, and more recently the intasome structure (Hare *et al.*, 2010), confirm that the true CC domain terminates after helix α5, as first suggested by Engelman and Craigie in 1992. Thus, the following loop–helix α6–loop–helix α7 should correspond to the linker between the catalytic and C-terminal domains. We attempted to prepare a CC domain that would more closely correspond to the CC domains of HIV and ASV integrases, but all constructs gave rise to insoluble protein. A construct extending the present C-terminal limit by an extra 30 residues did not express well either. Indeed, when trying to grow crystals of the two-domain PFV-1 CC and CT construct, we obtained some very poor crystals of the CC domain up to helix α7 after several months, presumably after slow proteolytic cleavage. It thus appears that the sequence that we used does correspond to a stable folding unit. In all molecules of both crystal forms the C-terminal helix (α7) lies against the protein core despite very different packing environments for the different molecules. This is clearly an energetically most favourable configuration for this region and could explain the expression behaviour of the different constructs.

In the crystal structure of the entire PFV integrase complex, the dimer with its cognate viral DNA shows that helix α7 must move away from the catalytic core to make room for the DNA, which it contacts with its C-terminal end. However, in the dimer lacking DNA this helix as well as the rest of the CT domain could not be traced in the electron density. This again suggests that the linker between the CC domain and helix α7 is extremely flexible, presumably in order to adjust for the engagement of the cellular DNA.

The authors wish to acknowledge the financial support of the CNRS and of the EU (TrioH European Project Grant FP6 503880 and ERASMUS exchange programme between the Ecole Normale Supérieure de Cachan and the Faculty of Natural Sciences, Charles University, 2006–2008). We acknowledge the European Synchrotron Radiation Facility and Synchrotron SOLEIL for provision of synchrotron-radiation facilities and the staff of beamlines ID14-3, ID23-2 and ID14-4 at ESRF and PROXIMA 1 at SOLEIL for assistance. We are grateful to Dr G. A. Bentley and members of the Institut Pasteur Block Allocation Group for access to ESRF and for their help with numerous data collections. We wish to thank Olivier Delelis (LBPA) for advice and many stimulating discussions.

References

Bordo, D. & Argos, P. (1991). *J. Mol. Biol.* **217**, 721–729.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996). *Structure*, **4**, 89–96.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Delelis, O., Carayon, K., Guiot, E., Leh, H., Tauc, P., Brochon, J. C., Mouscadet, J. F. & Deprez, E. (2008). *J. Biol. Chem.* **283**, 27838–27849.
Delelis, O., Carayon, K., Saïb, A., Deprez, E. & Mouscadet, J. F. (2008). *Retrovirology*, **5**, 114–127.
Delelis, O., Lehmann-Che, J. & Saïb, A. (2004). *Curr. Opin. Microbiol.* **7**, 400–406.
Delelis, O., Saïb, A. & Sonigo, P. (2003). *J. Virol.* **77**, 8141–8146.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Engelman, A. & Craigie, R. (1992). *J. Virol.* **66**, 6361–6369.
Gassner, N. C., Baase, W. A., Mooers, B. H., Busam, R. D., Weaver, L. H, Lindstrom, J. D., Quillin, M. L. & Matthews, B. W. (2003). *Biophys. Chem.* **100**, 325–340.
Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.
Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. (2010). *Nature (London)*, **464**, 232–236.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Lee, H. S., Kang, S. Y. & Shin, C. G. (2005). *Mol. Cells*, **19**, 246–255.
Lewinski, M. K. & Bushman, F. D. (2005). *Adv. Genet.* **55**, 147–181.
Linial, M. L. (1999). *J. Virol.* **73**, 1747–1755.
Liu, W. T., Natori, T., Chang, K. S. & Wu, A. M. (1977). *Arch. Virol.* **55**, 187–200.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2002). *Proteins*, **50**, 437–450.
Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clément-Mella, C. & Mikol, V. (1998). *J. Mol. Biol.* **282**, 359–368.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Ohmura, T., Ueda, T., Hashimoto, Y. & Imoto, T. (2001). *Protein Eng.* **14**, 421–425.
Pahl, A. & Flügel, R. M. (1993). *J. Virol.* **67**, 5426–5434.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* D**60**, 2288–2294.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Valkov, E., Gupta, S. S., Hare, S., Helander, A., Roversi, P., McClure, M. & Cherepanov, P. (2009). *Nucleic Acids Res.* **37**, 243–255.