Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, solution properties and crystallization of SIV integrase containing a continuous core and C-terminal domain

Ying Li, Youwei Yan, Joan Zugay-Murphy, Bei Xu, James L. Cole, Marc Witmer, Peter Felock, Abigail Wolfe, Daria Hazuda, Mohinder K. Sardana, Zhongguo Chen, Lawrence C. Kuo and Vinod V. Sardana*

Department of Antiviral Research, Merck Research Laboratories, West Point, Pennsylvania 19486-0004, USA

Correspondence e-mail: vinod_sardana@merck.com

Received 12 May 1999
Accepted 12 July 1999

The C-terminal two-thirds segment of integrase derived from the simian immunodeficiency virus has been cloned, expressed in *Escherichia coli*, and purified to greater than 95% homogeneity. The protein encompasses amino-acid residues 50–293 and contains a F185H substitution to enhance solubility. In dilute solutions at concentrations below 1 mg ml^{-1}, the enzyme is predominantly dimeric. At the higher concentrations (>10 mg ml^{-1}) required to enable crystallization, the enzyme self-associates to form species with molecular weights greater than 200 kDa. Despite the apparent high aggregation in solution, the enzyme crystallizes from a 8%(*v*/*v*) polyethylene glycol (molecular weight 6000) solution in a form suitable for X-ray diffraction studies. The resulting single crystals belong to the space group $P2_12_12_1$, with unit-cell parameters $a = 79.76$, $b = 99.98$, $c = 150.2$ Å, $\alpha = \beta = \gamma = 90^\circ$ and $Z = 4$. Under X-ray irradiation generated with a rotating-anode generator, the crystals diffract to 2.8 Å resolution and allow collection of a native 3 Å resolution diffraction data set.

1. Introduction

Integration of viral cDNA into the host genome is a critical step in the life cycle of retroviruses. Integration is catalyzed by a viral enzyme called integrase in a two-step process. In the first step (3′ processing), two nucleotides are cleaved from each 3′ end of the viral DNA adjacent to a conserved CA dinucleotide. In the second step (strand transfer), 3′ viral ends are joined to 5′ phosphates of the host DNA. Subsequent trimming of the 5′ ends of viral DNA and repair of its single-stranded gaps completes the proviral integration into the host cell genome. (For recent reviews on retroviral integration, see Katz & Skalka, 1994; Vink & Plasterk, 1993.)

Mutagenesis and proteolytic studies have shown that integrase consists of three functional domains (Engelman *et al.*, 1994). The N-terminal domain contains a conserved HHCC motif which binds to zinc in a 1:1 stoichiometry. Binding of zinc to the N-terminus enhances multimerization of the full-length enzyme and increases the catalytic activity of the enzyme (Burke *et al.*, 1992; Zheng *et al.*, 1996). The structure of the N-terminal domain of HIV integrase (amino acids 1–50) has been solved by heteronuclear NMR (Cai *et al.*, 1997) revealing a dimer in solution.

The central catalytic domain contains the highly conserved DD35E triad motif (D64, D116 and E152). As in other DNA-binding proteins, these acidic residues coordinate a divalent metal ion in the resting enzyme. Substitution of one of the conserved residues in the catalytic site abolishes both the cleavage and integration activities of integrase (Bushman *et al.*, 1993; Drelich *et al.*, 1992; Engelman & Craigie, 1992; Kulkosky *et al.*, 1992; Leavitt *et al.*, 1993; van Gent *et al.*, 1992, Vink *et al.*, 1993). The crystal structures of integrase core protein from the human immunodeficiency virus (HIV-1) and the avian sarcoma virus have been determined in the absence and presence of bound metal ions and/or inhibitors with use of X-ray crystallography (Bujacz, Alexandratos *et al.*, 1996; Bujacz, Jaskolski *et al.*, 1996; Dyda *et al.*, 1994; Goldgur *et al.*, 1998; Lubkowski *et al.*, 1998). The observed structures are similar to those of Mu transposase, RuvC and RNAase H; all are proposed to catalyze phosphoryl-transfer reactions involving at least one metal ion and a DD35E catalytic triad (Dyda *et al.*, 1994; Yang & Steitz, 1995; Bujacz, Alexandratos *et al.*, 1996; Rice *et al.*, 1996).

Although the catalytic core can catalyze the disintegration reaction, only the full-length enzyme containing the core as well as the N- and C-terminal domains is able to catalyze the cleavage and integration reactions. The C-terminal domain possesses a low degree of sequence homology among retroviruses. It has been suggested to bind DNA (Vink *et al.*, 1993; Woerner & Marcus-Sekura, 1993; Engelman *et al.*, 1994). Its three-dimensional structure in solution has been characterized using NMR (Eijkelenboom *et al.*, 1995; Lodi *et al.*, 1995), revealing a dimer with a SH3-like fold.

Despite the availability of tertiary structures for the individual domains of integrase, the quaternary juxtaposition of the domains and their interactions are unknown. Also unknown is the mode of binding of DNA to the native enzyme or to the individual domains. To elucidate the mechanism of action of integrase, a three-dimensional structure of the enzyme containing more than one domain is needed. In this paper, we describe the modified purification procedure of the SIV integrase containing the core and the C-terminal domains encompassing amino-acid residues 50–293 and a F185H mutation [SIV(50–293)F185H].[1] We find that although the enzyme exists as multimeric aggregates in solution, it is possible to obtain single crystals which diffract to better than 3 Å resolution. Our results highlight the unpredictable nature of protein crystallization, in contrast to recent suggestions that solution homogeneity may be an indicator of the intrinsic ability of a protein to crystallize (D'Arcy, 1994; Durbin & Feher, 1996; Ferre-D'Amare & Burley, 1997; George & Wilson, 1994).

2. Results and discussion

Our ability to crystallize an integrase which contains the C-terminal two-thirds of the SIV enzyme hinges on a large supply of pure protein. In the course of our search for useable single crystals for X-ray diffraction studies, more than 1.5 g quantity of various forms of integrase has been expended. Owing to the highly aggregating nature of both the HIV-1 and SIV integrase, it is also necessary to clarify the relation of protein aggregation and crystallization in order to define clearly to what extent site-specific mutagenesis is required to 'solubilize' the protein. We find that in the case of the HIV-1 core (data not shown) and the SIV integrase (reported here), the extent of protein aggregation is not a predictor of our ability to obtain single crystals suitable for structure determination.

[1] Abbreviations used: HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; HEPES, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; EDTA, ethylenediamine tetraacetic acid; MES, 2-(*N*-morpholino)ethanesulfonic acid; PAGE, polyacrylamide gel electrophoresis; PEG 6000, polyethylene glycol with molecular weight 6000; SDS, sodium dodecylsulfate; SIV(50–293)F185H, SIV integrase containing both the core and C-terminal domains from amino-acid residues 50–293 plus a site-specific mutation to histidine at Phe185.

2.1. Expression and purification

The SIV(50–293)F185H integrase has been expressed in *Escherichia coli* and purified from the soluble fraction as described in §3, with significant modifications of the published protocols for HIV-1 and SIV integrase (Du *et al.*, 1997; Jenkins *et al.*, 1996). The changes are necessary to obtain large quantities of protein crystals. In the present purification protocol, the amount of thrombin and the duration of the digestion time have been reduced to avoid generation of non-specific cleavage products. These products are not well resolved chromatographically and are difficult to remove in subsequent steps. All the dialysis steps for the purpose of buffer exchange have been replaced by simply diluting the enzyme in the required buffers in order to speed up purification and to avoid protein aggregation. Two-step SP Sepharose cation-exchange chromatography has been introduced to generate at least 95% pure integrase in buffer containing HEPES. Replacement of the SP Sepharose cation-exchange column with a one-step heparin affinity column yields an enzyme very similar in terms of purity to that obtained with the two-step SP Sepharose column based on analysis using SDS–PAGE. However, the resulting enzyme behaves very differently during crystallization. Under identical crystallization conditions, the SP Sepharose-purified enzyme forms crystals in 60–80% of the crystallization wells, compared with only 5–10% for the heparin-purified enzyme.

The purified SIV integrase gives only a single N-terminal sequence, beginning with GSHMIHGQVN. Approximately 5% of the products are smaller (15–20 kDa) than the expected protein, as shown in Fig. 1, but give the same N-terminus sequence as the intact integrase (27.8 kDa), suggesting ∼5% C-terminus heterogeneity existed in the final integrase preparations. For the selenomethionine derivative of the enzyme, a quantitative incorporation of selenomethionine is obtained. Only one species is found, with a molecular weight of 28238.2 (theoretical molecular weight is 28239), and no methionine is seen in amino-acid analysis.

2.2. Solubility of the SIV(50–293)F185H integrase

The solubility of the SIV integrase containing amino-acid residues 50–293 and a F185H substitution is dependent on temperature and ionic strength. The protein precipitates at 277 K when its concentration is greater than 0.5 mg ml^{-1} and when the NaCl concentration is less than 0.25 *M*. However, the enzyme is soluble under the same conditions if the temperature is raised to ∼283–285 K. In buffers containing 1 *M* NaCl, the enzyme solubility is increased to ≥15 mg ml^{-1} at either 293 or 277 K (data not shown). At 0.5 *M* NaCl, the protein solubility is 10.5 mg ml^{-1} at 293 K and 7 mg ml^{-1} at 277 K. At 50 m*M* NaCl, the salt concentration at which enzymatic activity assays are generally performed, the solubility is ≥0.2 mg ml^{-1} at either temperature. Thus, for use in crystallization trials, the SIV integrase is concentrated to 25−30 mg ml^{-1} at 285 K in buffers containing 1 *M* NaCl and immediately stored at 203 K. The selenomethionine version of the enzyme is less soluble and can only be concentrated to ∼12 mg ml^{-1}.

2.3. Oligomerization of SIV integrase

The self-association of SIV integrase has been characterized using analytical ultracentrifugation. Fig. 2 shows the average molecular weight of SIV integrase as a function of protein concentration and rotor speed in buffer containing 1 *M* NaCl at 293 K. At the lowest concentration of 0.3 mg ml^{-1}, the data obtained at all rotor speeds with an average molecular weight is 52.7 kDa, a value expected for a dimer of SIV integrase (55.6 kDa). At higher protein concentrations, the molecular weights increase suggesting further aggregation. In addition, the molecular weight decreases with increasing rotor speeds such that molecular weights >200 kDa were observed at >10 mg ml^{-1} protein. In the 1:1(*v*/*v*) mixture of 1 *M* NaCl storage buffer and the reagents used to prepare the SIV integrase crystals, the molecular weight of the protein

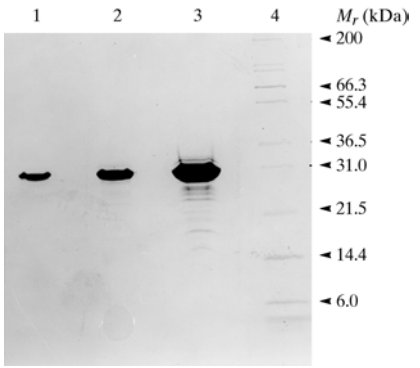

Figure 1
Reduced SDS–PAGE (16%) analysis of SIV(50–293)F185H integrase. The gel was stained with Coomassie Brilliant Blue. Lanes 1–3, purified enzyme loaded at 2, 5 and 10 µg, respectively; lane 4, molecular-weight markers.

again increases with protein concentration; however, the magnitude of the increase was less than that observed for the 1 *M* NaCl experiments (data not shown). For example, at 2 mg ml^{-1} the apparent molecular weights ranged from 69−144 in 1 *M* NaCl , but under the crystallization conditions the range was 67–83 kDa.

Our data suggest that the SIV(50–293)F185H integrase has a strong tendency to aggregate in solution. Although at protein concentrations below 1 mg ml^{-1} this protein is predominantly dimeric in both buffer systems examined, strong aggregation ensues at higher protein concentrations, producing average molecular weights >200 kDa at 10 mg ml^{-1} for the 1 *M* NaCl buffer system. In addition, the apparent molecular weights decrease with increasing rotor speed in both buffer systems. This behavior is characteristic of heterogeneous systems which are not in rapid reversible equilibrium (Laue, 1995). A number of researchers have found that proteins which exist in a homogeneous state in sub-saturated solutions tend to crystallize, whereas polydisperse aggregated systems are unlikely to crystallize (D'Arcy, 1994; Durbin & Feher, 1996; Ferre-D'Amare & Burley, 1997; George & Wilson, 1994). In the case of SIV(50–293)F185H integrase, it exists as a heterogeneous polymer at concentrations well below its solubility limit. Nevertheless, this enzyme crystallizes to yield diffracting single crystals.

2.4. Protein crystallization

The productive single crystals of the SIV(50–293)F185H integrase were obtained after three successive stages. The initial conditions for crystallization of the SIV(50–293)F185H integrase were obtained with the use of commercially available Hampton crystallization-screening kits. Although most of the crystallization solutions yielded precipitation when mixed with the protein solution, peanut-shaped pseudo-crystals with rounded edges were obtained in a Hampton solution which contained 10%(*v*/*v*) PEG 6000 and 0.1 *M* MES pH 6.0. These crystals diffracted to ∼10 Å resolution. Two such crystals are shown in Fig. 3(*a*).

Substituting the buffer in the crystallization mixture with 0.1 *M* sodium cacodylate improved the integrase crystals, giving crystals with sharp edges. Crystals grown at pH 6.5 in the presence of 0.2 *M* $MgCl_2$ and 8%(*v*/*v*) PEG 6000 are shown in Fig. 3(*b*). The new crystals diffracted to ∼4 Å resolution. However, they were very fragile. When exposed to cryo-protective agents the crystals tended to crack readily, even when the cryo-solvents were delivered to the crystals in hanging drops *via* vapor diffusion. With an addition of 0.1% of the cross-linking agent glutaraldehyde, the crystals became durable in most solutions containing a cryo-protective agent, but the diffraction power of these crystals was reduced drastically to worse than 10 Å resolution.

In the final round, crystals were improved when grown in 0.1 *M* MES pH 5.7, 8%(*v*/*v*) PEG 6000 and 1.5%(*v*/*v*) dioxane. A quantity of 0.1 *M* $MgCl_2$ was optional, but the presence of microscopic crystalline seeds prepared from crystals obtained from the MES cacodylate buffer was necessary to bring about crystal growth. As shown in Fig. 3(*c*), the resultant crystals of the selenomethionine derivative of the SIV(50–293)F185H integrase displayed sharp edges and clear surfaces; the crystals were 0.5 × 0.15 × 0.15 mm in size. A diffraction data set which contained 27263 unique reflections has been collected to 2.8 Å resolution with one single crystal. The space group determined was $P2_12_12_1$, with unit-cell parameters a = 79.76, b = 99.98, c = 150.2 Å, $\alpha = \beta = \gamma = 90^\circ$. There are four integrase molecules per asymmetric unit.

An oscillation photograph of the selenomethionine-containing SIV(50–293)F185H integrase crystal is shown in Fig. 4. Data collection at multiple wavelengths using synchrotron X-ray sources is currently under way.

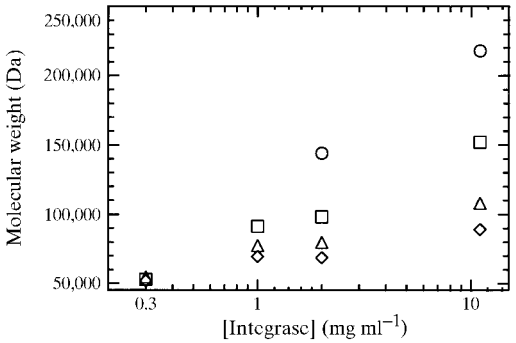

Figure 2
Average molecular weights of SIV(50–293)F185H integrase in 1 *M* NaCl buffer. The SIV(50–293)F185H integrase in a buffer containing 1 *M* NaCl, 20 m*M* Tris, 1 m*M* EDTA pH 7.5 was characterized with short-column sedimentation equilibrium at 293 K. Data were collected at rotor speeds of 12000 (circles), 16000 (squares), 20000 (triangles) and 24000 rev min^{-1} (diamonds) at the indicated protein concentrations and rotor speeds. Each channel was fitted separately to give the *Z*-average molecular weight (Laue, 1995).

3. Materials and methods

3.1. Expression of SIV(50–293)F185H integrase

A plasmid encoding amino acids 50–293 of the SIV integrase and a 21-residue N-terminal tag was cloned into the *Nde*I/*Bam*HI sites of the pET15B vector (Novagen) and expressed under control of the T7 promoter. The N-terminal tag contained a thrombin cleavage site and a string of six histidines. The phenylalanine side chain of residue 185 was substituted with a histidine in order to enhance solubility (Dyda *et al.*, 1994; Jenkins *et al.*, 1996). The plasmid was transformed into *E. coli* BL21DE3 cells (Stratagene). Cells were cultured at 310 K to an optical density of 0.8–0.9 at 600 nm in LB broth which contained 50 µg ml^{-1} ampicillin. Induction of plasmid expression took place with 1 m*M* isopropyl-D(−)-thiogalactopyranoside, and the cells were allowed to grow for an additional 2 h prior to storage at 203 K until use.

3.2. Purification of SIV(50—293)F185H integrase

Packed cells from a 30 l culture were suspended at 277 K in 100 ml of lysis buffer (25 m*M* HEPES pH 7.5, 5 m*M* β-mercaptoethanol, 5 m*M* imidazole, 10% glycerol, 1 *M* NaCl) and treated for 15–20 min with 0.02 mg ml^{-1} DNAase (Type IIS: bovine pancreas, Sigma) in 10 m*M* $MgCl_2$. Phenylmethyl sulfonyl flouride was added to the cell suspension to a final concentration of 2 m*M* and the cells were immediately disrupted with two passes through a microfluidizer (Model 110-S) at 6 × 10^5 Pa pressure. The cell lysate was centrifuged at 10000 r min^{-1} for 30 min. The pellet was homogenized in 60 ml of lysis buffer and centrifuged at 10000 r min^{-1} for 30 min. Two 5 ml His-trap columns (Pharmacia Biotech) were pre-charged with $NiSO_4$ and equilibrated with buffer *A* (25 m*M* HEPES pH 7.5, 5 m*M* β-mercaptoethanol, 20 m*M* imidazole, 10% glycerol, 0.5 *M* NaCl). The two supernatants were combined and the imidazole concentration was increased to 20 m*M*. The pool was loaded at 3.5 ml min^{-1} onto the His-trap column. The column was washed with lysis buffer containing 20 m*M* imidazole for at least 20 column volumes, followed by washing with buffer *A* containing 80 m*M* imidazole for 10 column volumes. The enzyme was eluted from the column with a 80−500 m*M* imidazole gradient. Integrase was eluted at 300 m*M* imidazole. The pooled fractions were diluted with a solution containing 25 m*M* HEPES at pH 7.5 and 5 m*M* β-mercaptoethanol to final

NaCl and imidazole concentrations of 0.25 *M* and 0.15 *M*, respectively. Thrombin (Sigma) digestion was performed for 30 min at ambient temperature at a concentration of 10 NIH units of thrombin per milligram of integrase. The resultant insoluble material was removed with centrifugation at 3000*g* for 10 min. The supernatant was loaded onto a benzamidine 6B Sepharose column (Pharmacia-Biotech) previously equilibrated with buffer *B* (25 m*M* HEPES pH 7.5, 5 m*M* β-mercaptoethanol, 5% glycerol, 0.25 *M* NaCl) to bind thrombin. The flow-through which contained integrase was loaded onto the His-trap column to remove histidine-tag-containing (undigested) integrase. The flowthrough from the His-trap column was diluted sixfold with buffer *C* (50 m*M* MES pH 5.5, 1 m*M* EDTA, 5 m*M* β-mercaptoethanol, 10% glycerol) and loaded onto a SP Sepharose high-performance column (XK16/20, Pharmacia Biotech) at a flow rate of 3.5 ml min^{-1}. The enzyme was eluted with a 0.15–1.0 *M* NaCl gradient. Eluent fractions were analyzed with SDS–PAGE and Coomassie Blue stain. Fractions which contained >95% purified integrase were pooled and buffer-exchanged on a SP Sepharose Hi-trap (10 ml) column into a buffer solution containing 25 m*M* HEPES (pH 7.5), 1 m*M* EDTA, 5 m*M* β-mercaptoethanol and 1 *M* NaCl and concentrated to 25–30 mg ml^{-1} with Centriprep-10 (Amicon). N-terminal sequence analysis was carried out with Edman degradation on an Applied Biosystem model 470 A gas-phase sequencer. Protein concentration was determined by quantitative amino-acid analysis. The enzyme was stored at 203 K until use.

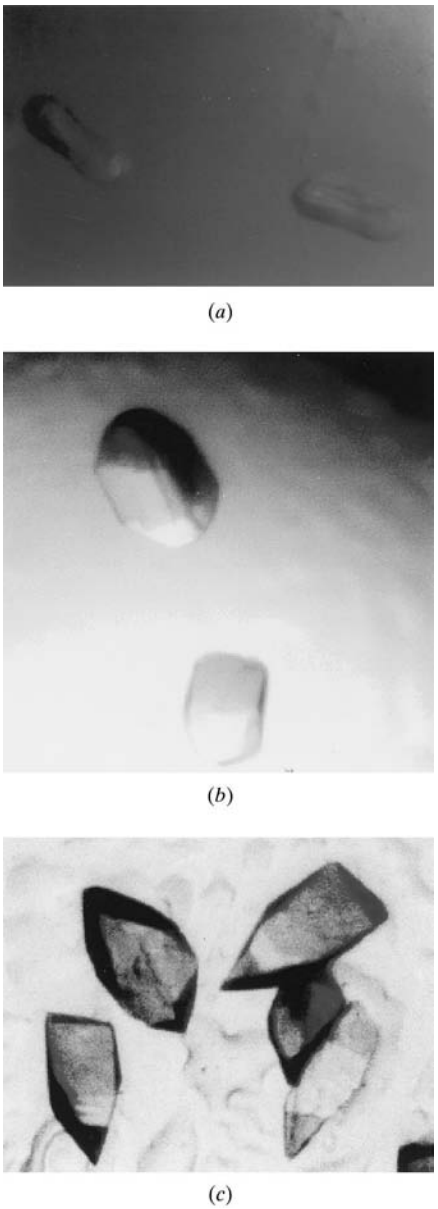

Figure 3
Crystals of the SIV(50–293)F185H integrase grown under three different conditions. All crystals were prepared in hanging drops with a 1:1(*v*/*v*) mix of the protein solution at 10–15 mg ml^{-1} (see §3) and a 1 ml reservoir of mother liquor. (*a*) Initial crystals obtained with a mother liquor which contained 10%(*v*/*v*) PEG 6000 and 0.1 *M* MES pH 6.0; (*b*) crystals of the selenomethionine enzyme derivative obtained with mother liquor containing 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* $MgCl_2$ and 8%(*v*/*v*) PEG 6000; (*c*) final single crystals obtained with use of micro-seedings prepared from crystals of the selenomethionine enzyme derivative obtained in (*b*) and a mother liquor of 0.1 *M* MES pH 5.7, 8%(*v*/*v*) PEG 6000 and 1.5%(*v*/*v*) dioxane. A quantity of 0.1 *M* $MgCl_2$ is optional.

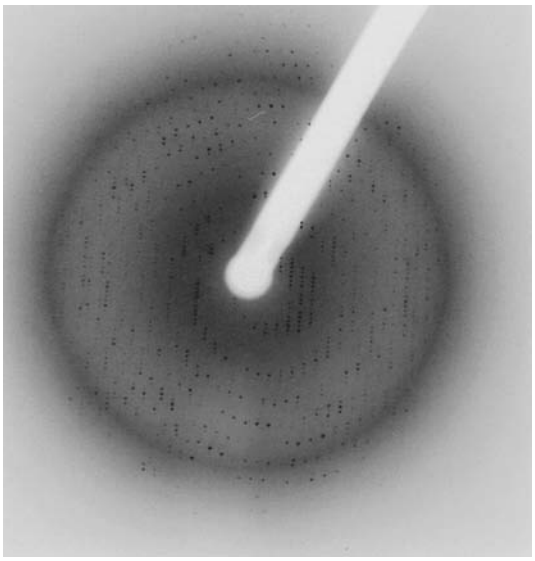

Figure 4
An X-ray oscillation photograph of a SIV(50–293)F185H integrase crystal. Diffraction pictures were recorded at 2° oscillation intervals with use of a MSC double-mirror focusing system and irradiation generated with a Rigaku RU 200 rotating-anode generator operating at 50 kV and 100 mA.

The selenomethionine substitution was carried out as described (Pryor & Leiting, 1997) using the *E. coli* methionine auxotropic strain B843(DE3) from Novagen in a minimal culture broth. Purification was conducted in a manner similar to that performed for the native enzyme, with the exception that the amount of reducing agent was doubled in the final cation-exchange chromatography step. The extent of selenomethionine substitution was determined with electrospray ionization mass spectrometry and amino-acid analysis.

3.3. Sedimentation equilibrium

The SIV(50–293)F185H integrase was buffer-exchanged using BioGel P-6 spin gel-filtration columns (Biorad). Solubility measurements were performed by exchanging integrase at 15 mg ml^{-1} from a buffer at pH 7.5 containing 1 *M* NaCl, 20 m*M* Tris, 1 m*M* EDTA into buffers with variable concentrations of NaCl. Following the exchange, samples were incubated at either 277 or 293 K for 30 min, followed by centrifugation at 16000*g* for 10 min. The concentration of soluble integrase in the supernatant was measured by the absorbance at 280 nm using a calculated molar extinction coefficient of 39000 M^{-1} cm^{-1}. A partial specific volume of 0.735 ml g^{-1} was calculated from the amino-acid sequence of the enzyme.

For sedimentation-equilibrium measurements, 14 µl samples with a 5 µl layer of FC-43 were loaded into eight-channel short-column (0.75 mm) centerpieces (Laue, 1995). For the highest protein concentration (11 mg ml^{-1}), 10 µl samples were loaded into two-channel 3 mm path-length cells. The short-column centerpieces were purchased from the National Analytical Ultracentrifugation Facility, Storrs, CT. Sedimentation measurements were performed at rotor speeds of 12000, 16000, 20000 and 24000 rev min^{-1} in a Beckman XL-I analytical ultracentrifuge at 293 K using interference optics. Sedimentation equilibrium was generally achieved within 6 h at each speed. Data at each protein concentration and rotor speed were fitted to a single ideal species model using the program *IGOR Pro* (Wavemetrics).

3.4. Protein crystallization

Single crystals were prepared by mixing 5 µl of 10–15 mg ml^{-1} SIV(50–293)F185H integrase with an equal volume of mother liquor containing 0.1 *M* MES pH 5.7, 8%(*v*/*v*) PEG 6000 and 1.5%(*v*/*v*) dioxane. The mixture was seeded with microscopic crystals of the native protein prepared in sodium cacodylate buffer, as described in §2, and was equilibrated at room temperature as a hanging drop against 1 ml of mother liquor. Small crystals were formed in 1−2 d and grew to 0.5 × 0.15 × 0.15 mm within another several days.

3.5. X-ray diffraction

A native data set was collected at 107 K with monochromated Cu *Kα* X-ray irradiation generated with a Rigaku RU-200 rotating-anode generator operating at 50 kV and 100 mA and focused with a double-mirror focus system. Data reduction and determination of space group and unit-cell constants were performed using the *DENZO* program package. The diffraction data was scaled with the *SCALEPACK* program (Otwinowski, 1993).

Note added in proof: The X-ray structure of the SIV integrase containing a continuous core and C-terminal domain has now been solved and refined at 3 Å resolution, with an *R* value of 0.2 and an $R_{\rm free}$ value of 0.36 (Chen *et al.*, 1999).

We thank Mr Collin Lellis for mass spectrometry, Ms Theresa Wood for N-terminal amino-acid sequence analysis and Ms Mai Tang for amino-acid analysis.

References

Bujacz, G., Alexandratos, J., Qing, Z. L., Clement-Mella, C. & Wlodawer, A. (1996). *FEBS Lett.* **398**, 175–178.

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1996). *Structure*, **4**, 89–96.

Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zeft, A. S., Middaugh, C. R. & Cordingley, M. G. (1992). *J. Biol. Chem.* **267**, 9639–9644.

Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). *Nature Struct. Biol.* **4**, 567–577.

Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emini, E. A., Hazuda, D. & Kuo, L. C. (1999). Submitted.

D'Arcy, A. (1994). *Acta Cryst.* D**50**, 469–471.

Drelich, M., Wilhelm, R. & Mous, J. (1992). *Virology*, **188**, 459–468.

Du, Z., Ilyinskii, P. O., Lally, K., Desrosiers, R. C. & Engelman, A. (1997). *J. Virol.* **71**, 8124–8132.

Durbin, S. D. & Feher, G. (1996). *Annu. Rev. Phys. Chem.* **47**, 171–204.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). *Science*, **266**, 1981–1986.

Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R. & Hard, K. (1995). *Nature Struct. Biol.* **2**, 807–810.

Engelman, A. & Craigie, R. (1992). *J. Virol.* **66**, 6361–6369.

Engelman, A., Hickman, A. B. & Craigie, R. (1994). *J. Virol.* **68**, 5911–5917.

Ferre-D'Amare, A. R. & Burley, S. K. (1997). *Methods Enzymol.* **276**, 157–166.

Gent, D. C. van, Oude Groeneger, A. A. M. & Plasterk, R. H. A. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 9598–9602.

George, A. & Wilson, W. W. (1994). *Acta Cryst.* D**50**, 361–365.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.

Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. (1996). *J. Biol. Chem.* **271**, 7712–7718.

Katz, R. A. & Skalka, A. M. (1994). *Annu. Rev. Biochem.* **63**, 133–173.

Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P. & Skalka, A. M. (1992). *Mol. Cell Biol.* **12**, 2331–2338.

Laue, T. M. (1995). *Methods Enzymol.* **259**, 27–452.

Leavitt, A. D., Shiue, L. & Varmus, H. E. (1993). *J. Biol. Chem.* **268**, 2113–2119.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1995). *Biochemistry*, **34**, 9826–9833.

Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T. R. Jr, Neamati, N., Pommier, Y., Merkel, G. & Skalka, A. M. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4831–4836.

Pryor, K. D. & Leiting, K. (1997). *Protein Expr. Purif.* **10**, 309–319.

Rice, P., Craigie, R. & Davies, D. R. (1996). *Curr. Opin. Struct. Biol.* **6**, 76–83.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. N. Sawyer, N. W. Isaacs & S. Bailey, pp. 55–62. Warrington: Daresbury Laboratory.

Vink, C., Oude Groeneger, A. M. & Plasterk, R. H. (1993). *Nucleic Acids Res.* **21**, 1419–1425.

Vink, C. & Plasterk, R. H. (1993). *Trends Genet.* **9**, 433–438.

Woerner, A. M. & Marcus-Sekura, C. J. (1993). *Nucleic Acids Res.* **21**, 3507–3511.

Yang, W. & Steitz, T. A. (1995). *Structure*, **3**, 131–134.

Zheng, R., Jenkins, T. M. & Craigie, R. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 13659–13664.