# JMB

Available online at www.sciencedirect.com

**ScienceDirect**

ELSEVIER

# Single-particle Image Reconstruction of a Tetramer of HIV Integrase Bound to DNA

## Gang Ren[1]†, Kui Gao*†, Frederic D. Bushman[2]* and Mark Yeager[1,3]*

[1]*The Scripps Research Institute Department of Cell Biology 10550 North Torrey Pines Road La Jolla, CA 92037, USA*

[2]*University of Pennsylvania School of Medicine Department of Microbiology 3610 Hamilton Walk Philadelphia, PA 19104-6076 USA*

[3]*Division of Cardiovascular Diseases, Scripps Clinic 10666 North Torrey Pines Road La Jolla, CA 92037, USA*

The HIV integrase enzyme (IN) catalyzes the initial DNA breaking and joining reactions that integrate viral DNA in the host chromosome. Structures for individual IN domains have been determined by X-ray crystallography and NMR spectroscopy, but the structure of the complete IN–DNA complex has remained elusive. Homogeneous complexes of IN tetramers were assembled on DNA three-way junction substrates designed to resemble integration intermediates. Electron microscopy and single-particle image analysis of these complexes yielded a three-dimensional reconstruction at ~27 Å resolution. The map of the IN–DNA complex displays four lobes of density ~50 Å in diameter. Three of the lobes form a roughly triangular base with a central channel ~20 Å in diameter. The fourth lobe is centered between two lobes and extends ~40 Å above the base. We propose that the central channel tethers the target DNA, and two of the lobes may bind the ends of the viral DNA. The asymmetry of the complex is a feature not incorporated in previous structural models and potentially provides the first view of an asymmetric reaction intermediate.

*Corresponding authors

*Keywords:* HIV; integrase; DNA-binding; electron microscopy

## Introduction

The retroviral-encoded integrase (IN) enzymes are members of a large family of recombinases that contain the D,D-35-E active site motif.[1–3] X-ray crystallography and NMR spectroscopy have been used to determine high-resolution structures of the three protein domains and two domain fragments of retroviral INs.[4–11] However, there is no structure available for the complete three-domain IN protein or the IN–DNA complex.

IN is a promising target for antiviral drugs because it is essential for HIV replication, and because there is no close counterpart in host cells.[1,12–14] Understanding the correctly assembled IN–DNA complex is crucial for structure-based design of improved integrase inhibitors. Properly assembled IN–DNA complexes respond to small-molecule inhibitors *in vitro* differently than do dissociated mixtures of free IN and DNA substrates. That is, integration reactions containing preassembled IN–DNA complexes are less prone to inhibition by nuisance compounds,[15,16] and screens using such reactions have yielded molecules that are showing success in clinical trials.[14,17] Thus, further improvement of IN inhibitors would be aided greatly by structural information on the correctly assembled HIV IN–DNA complex.

The DNA-breaking and joining reactions mediating HIV integration are illustrated by Figure 1(a).[18–24] The immediate precursor for integration is the linear viral cDNA (Figure 1(a), part 1). Before integration, two nucleotides are removed from each 3′ end by IN (Figure 1(a), part 2), a reaction that may serve to generate a homogeneous substrate for subsequent reaction steps,[25,26] and may stabilize the IN–DNA complex.[27,28] A coupled transesterification reaction mediated by IN joins the recessed 3′ ends of the viral DNA to the protruding 5′ ends in the

† G.R. and K.G. contributed equally to this work.

Present addresses: G. Ren, University of California, San Francisco, Department of Biochemistry & Biophysics, 1700 4th Street, MC 2532, Room BH301C, San Francisco, CA 94158-2330, USA; K. Gao, F.D. Bushman and M. Yeager, GenProbe, 10210 Genetic Center Drive, San Diego, CA 92121-4362, USA.

Abbreviations used: IN, integrase; BME, β-mercaptoethanol.

E-mail addresses of the corresponding authors: bushman@mail.med.upenn.edu; yeager@scripps.edu
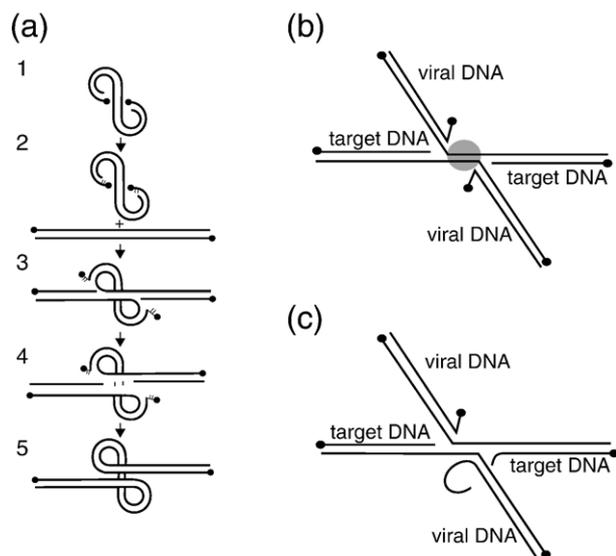
**Figure 1.** A schematic model for HIV DNA integration and the DNA substrate used in this study. (a) The DNA breaking and joining reactions involved in integration. See the text for details. (b) An integration intermediate synthesized from oligonucleotides. Two double-stranded oligonucleotides model the ends of the viral DNA. Note that the molecule is not stably paired in the grey region. (c) The stably paired DNA three-way junction substrate used to assemble the IN–DNA complexes.

target DNA (Figure 1(a), part 3).[29] The specific enzymes responsible for repair of the resulting DNA gaps at each end of the viral DNA (Figure 1(a), parts 4 and 5) are not fully clarified, but host cell gap repair enzymes are likely candidates.[30]

The complex that carries out integration *in vivo* is expected to involve a multimer of IN. Support for this idea can be inferred from the substrate symmetry, since the two viral DNA ends can be reasonably modeled with each bound by a different IN subunit in an IN multimer.[4,7–10] Additional evidence is based on the results of genetic complementation studies, in which different IN mutants were found to complement each other when present in the same complex.[31,32] Furthermore, purified IN forms multimers readily *in vitro*.[4,33,34]

A complication in studying IN–DNA complexes has been the poor solubility of the protein *in vitro*. In one approach to this problem, several studies have reported surface mutations that improved solubility and allowed 3D crystallization and X-ray structural analysis of IN domains.[7,8,35] Another strategy has been to assemble IN with DNA fragments. The use of exact mimics of integration intermediates, however, results in a molecule that is not base-paired stably (Figure 1(b)). However, DNA stabilization can be accomplished by linking the structures together as a pair of DNA three-way junctions, and such modified structures have been shown to be substrates for Rous sarcoma virus (RSV) IN.[36] Alternatively, the addition of oligonucleotides re-

sembling the viral DNA ends yielded more homogeneous and soluble RSV–IN complexes.[37]

In this study, we examined soluble HIV IN derivatives with DNA three-way junction substrates. Physical and spectroscopic analysis suggested that IN formed a tetramer bound to a single DNA substrate. Since the complexes were soluble and monodisperse, we used electron microscopy and image reconstruction to derive a 3D map at 27 Å resolution. A remarkable feature is that the triangular base of the complex encloses a central channel that we propose binds the target DNA. The structure was found to be asymmetric, a feature not previously considered in structural models for IN–DNA complexes. However, a recent functional study *in vitro* did conclude that the two viral DNA ends become integrated sequentially into target DNA in a defined order, implying the existence of an asymmetric intermediate.[28]

## Results and Discussion

### Optimizing IN–DNA complex formation

IN complexes were assembled on DNA substrates designed to resemble the product of the IN-catalyzed DNA strand transfer reaction (Figure 1(a) and (b)). Pilot studies suggested that a DNA formed by annealing five oligonucleotides as shown in Figure 1(c) yielded the optimal DNA for assembly. The horizontal parts of the DNA as drawn mimic the integration target DNA, and the two diagonal DNA duplexes match the viral DNA ends (U3 and U5). In the intermediate shown in Figure 1(b), note that each viral DNA end is joined to the target DNA on one DNA strand only. The branched DNA molecule in Figure 1(c) differs from the authentic intermediate by the DNA loop that attaches the 5′ end of the right viral DNA end to the free 3′ end of the adjacent target DNA. This prevents dissociation of the annealed oligonucleotides due to melting of the five bases of target DNA between the points of joining of the two viral DNA ends (grey circle in Figure 1(b)), which is expected to take place at physiological temperatures and is known to take place in the authentic intermediate.[38] The substrate that we used (Figure 1(c)) differs from previously reported paired DNA three way junctions[36] by having the non-biological connection in only one half of the DNA complex, a modification that was required for efficient assembly of HIV IN–DNA complexes (data not shown). The lengths of the DNA arms in the paired DNA three-way junction substrate were selected on the basis of pilot assembly experiments testing DNAs with different arm lengths (data not shown).

Several IN mutants were prepared and tested for improved complex formation and integration activity, including C56S/W131D/F139D/F185K/C280S, C56S/W131D/F139D/F185H/C280S, C56S/W131D/F139D/C280S, C56S/W131D/F139D/F185H/E246C/

C280S, and C56S/W131D/F139D/E246C/C280S. The substitutions at positions 131, 139, and 185 improved solubility,[8,35] while the substitutions of Cys56 and Cys280 were well tolerated and may have reduced formation of oxidative side-products during protein purification and storage.[33,39] To assess DNA-binding efficiency, IN mutants were further modified to contain the E246C substitution and were tested for efficient cross-linking to DNA substrates that contained tethered sulfur atoms at LTR position 7 as described.[39] The DNA substrates were end-labeled and incubated with the IN
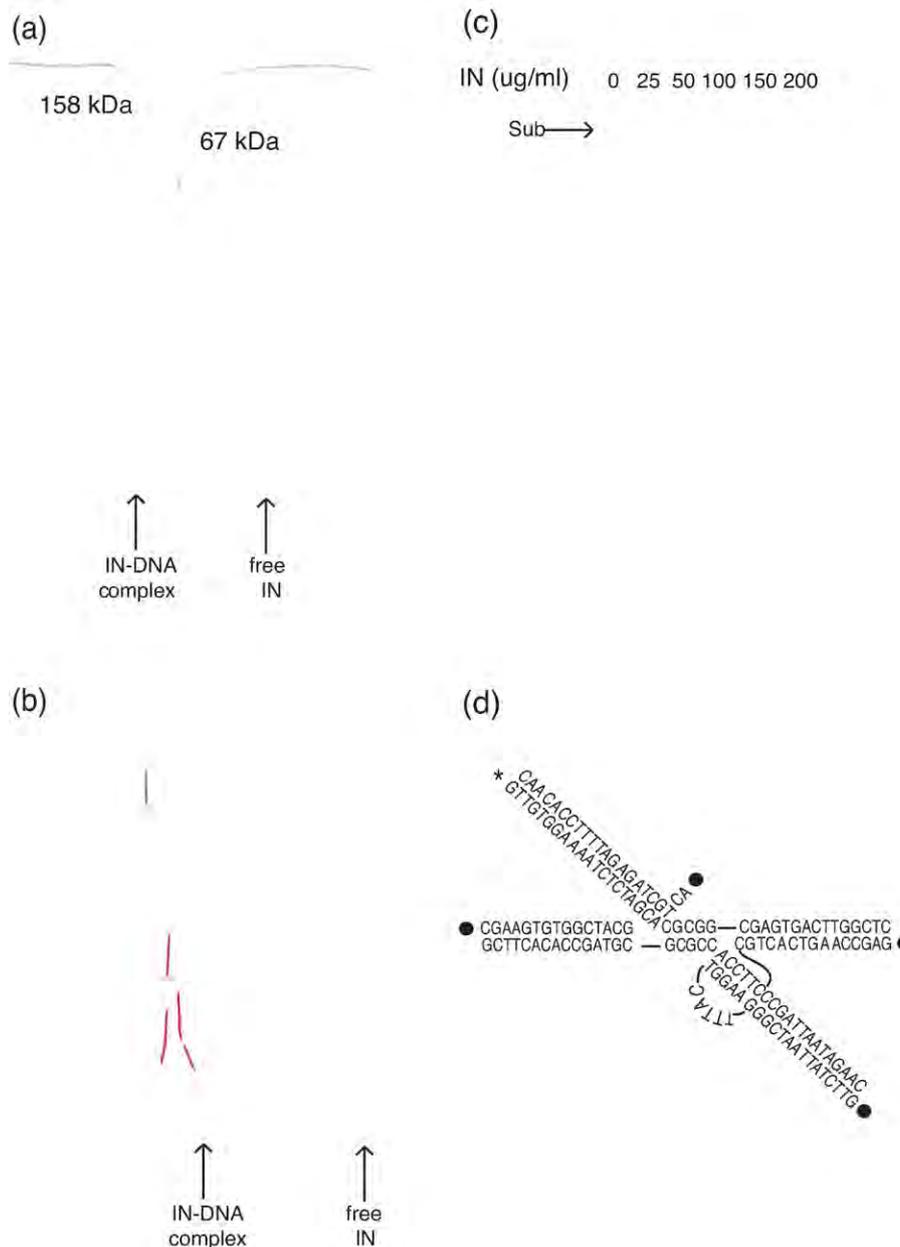


**Figure 2.** Analysis of HIV IN–DNA complexes by gel-filtration and DNase I nuclease protection. (a) Spectroscopic analysis of IN–DNA complexes (0.1 mg/ml) separated by gel-filtration chromatography using Superose 12. Blue trace, IN–DNA complex measured at 254 nm; red trace, the same elution profile analyzed at 280 nm; black trace, IN alone analyzed at 280 nm; grey trace, (inverted) molecular mass markers. The greater absorbance at 254 nm *versus* 280 nm is as expected for a protein–DNA complex. The molar extinction coefficients were calculated to be 44,860 for IN (280 nm) and 1,144,100 for the DNA (260 nm). (b) Determination of the IN–DNA dissociation constant. IN–DNA complexes were diluted to different final concentrations and subjected to Superose 12 chromatography. Concentrations (mg/ml) of IN–DNA complex loaded onto the columns were: black, 0.10; red, 0.05; pink, 0.025; purple, 0.02; cyan, 0.01; and green, IN only. Peak heights were normalized to facilitate comparisons. (c) DNase I protection analysis of paired three-way junction DNA bound to IN. The paired DNA three-way junction substrate shown in (d) was labeled on the 5′ end indicated by the asterisk (*) by treatment with kinase and [γ-^{32}P]ATP. Concentrations of IN are as indicated above the gel. (d) The paired DNA three-way junction substrate used for the nuclease protection study. The asterisk indicates the position of the ^{32}P label.

derivatives, followed by electrophoresis of the reaction products. Relative affinity was assessed by monitoring the formation of complexes in the

presence of increasing concentrations of salt. This analysis indicated that IN C56S/W131D/F139D/F185H/C280S formed DNA complexes with relatively high affinity (unpublished results), so IN derivatives containing these substitutions were used in further experiments.

## Assembly and characterization of IN–DNA complexes

To assemble complexes, purified IN protein was mixed with the annealed DNA three-way junction in the presence of 1 M NaCl and 5 mM Chaps detergent. It is known that IN does not bind to DNA in high ionic strength buffers. Consequently, dialysis against buffers containing 100 mM NaCl and 5 mM Chaps allowed assembly of IN–DNA complexes.

Size-exclusion chromatography showed that the IN–DNA complexes eluted as a single peak with a mobility slightly greater than the 158 kDa marker (Figure 2(a)). The expected size of an IN tetramer bound to one molecule of the paired DNA three-way junction is 168 kDa. The ratio of absorbance at 254 nm and 280 nm was consistent with a 4:1 stoichiometric ratio of protein and DNA (Figure 2(a) and data not shown). The dissociation constant of the IN–DNA complex ($K_D = 120$ nM) was estimated by dilution of IN–DNA complexes followed by gel-filtration (Figure 2(b)).

IN–DNA interactions in the assembled complexes were also characterized by DNAse I protection (Figure 2(c)). One strand of the DNA substrate was end-labeled with [32]P (asterisk in Figure 2(d)), and labeled substrates were incubated with various concentrations of IN. The entire substrate within the complex became protected from DNAseI digestion at ~150 μg/ml of IN. Thus, we infer that the IN tetramer binds the paired three-way junction so that there is steric interference with DNAseI attack over most of the DNA length.
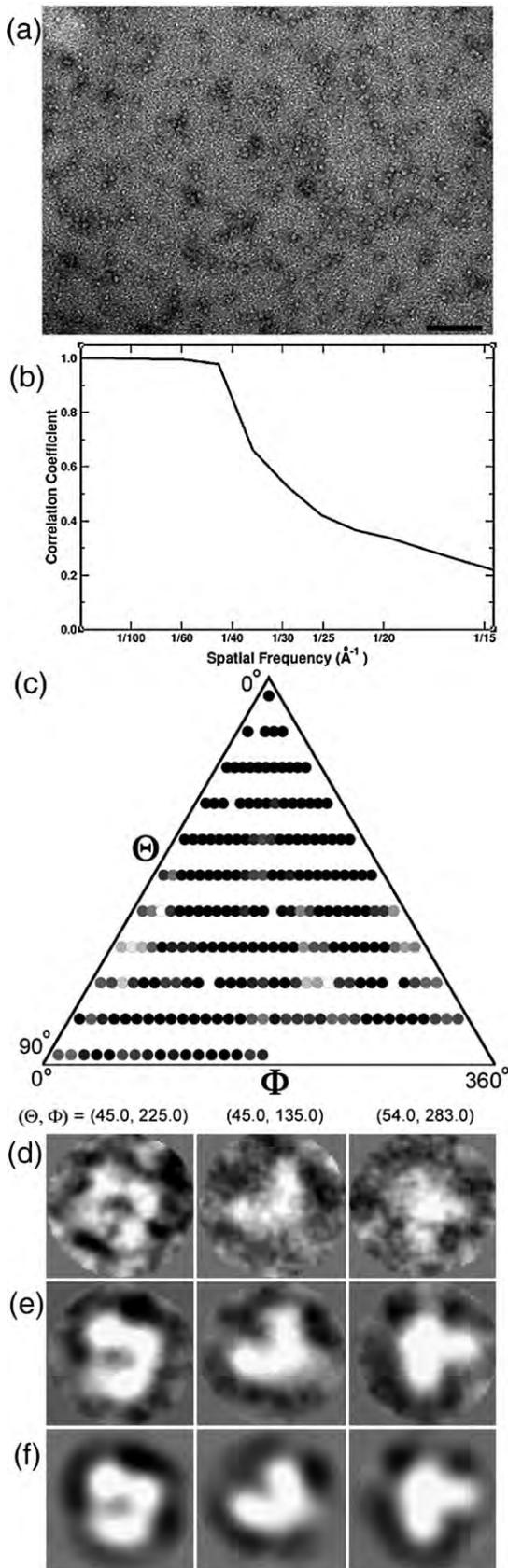


**Figure 3.** Single-particle image reconstruction of HIV IN–DNA complexes. (a) Electron micrograph of negatively stained complexes showing a homogeneous population of compact particles. The scale bar represents 500 Å. (b) Fourier shell correlation to determine the map resolution. The total data set of images was divided into two equal groups, the particles in each data set were aligned independently with projections of the final model, and two independent maps were correlated in Fourier space. The curve crosses the 50% threshold at ~27 Å resolution. (c) Histogram of the raw images in each view of the final model showing that all views of the complex are well represented and there is no preferential orientation. Images were aligned translationally (*x* and *y*) and rotationally in-plane (ω), and each particle was assigned to the group to which it correlated best. Each circle represents a projection of the final model for a given θ and φ value. (d) Three representative images and their refined θ and φ Euler angles. (e) Class averages for the particle orientations shown in (d). (f) Back projections of the final 3D reconstruction according to the same Euler angles. There is a progressive increase in the clarity of the images from (d) to (f).

## Electron microscopy and single particle image analysis

Electron micrographs of negatively stained IN–DNA complexes showed a homogeneous distribution of compact, globular particles (Figure 3(a)). Cross-correlation analysis using the EMAN software suite was used to sort individual particle images into classes of similar views that were then averaged; the data quality measures are presented in Figure 3(b) and (c). Representative individual particle images and their class averages are shown in Figure 3(d), (e) and (f), respectively. The class averages were then merged to yield a starting 3D map of the complex. Comparison of the raw images with 3D back projections of the 3D model generated an improved set of class averages and subsequently an improved 3D map. The iterative refinement was continued until there was no further statistical improvement in the 3D map. The resolution was based on the Fourier shell correlation method in which the data set was randomly divided in half and the two maps were compared in resolution shells (Figure 3(b)). Using a correlation coefficient cut-off value of 0.5, the resolution was estimated to be 27 Å.

## The 3D map of the IN–DNA displays four domains

Since we do not expect duplex DNA to be visible in a low-resolution 3D map of negatively stained particles, the observed density is interpreted as representing predominantly the 128 kDa IN tetramer only (Figure 4(a)). The map displays four lobes of density ~50 Å in diameter. Three of the lobes form a roughly triangular base with a central channel ~20 Å in diameter (Figure 4(a), top row). The fourth lobe is centered between two lobes and extends ~40 Å above the base (Figure 4(a), bottom row). We propose that the central channel tethers the target DNA, and two of the lobes may bind the ends of the viral DNA.

## Comparison to previously proposed models for the IN–DNA complex

Several models for the structure of the IN–DNA complex have been proposed on the basis of the available high-resolution structural data for IN domains, IN–DNA crosslinking data,[39–43] as well as other biochemical experiments. We generated low-resolution molecular envelopes for each model

(Figure 4(b), (c), and (d)) to allow visual comparison with our 3D map.

Model 1 (Figure 4(b)) satisfied constraints from structural and crosslinking studies, with particular emphasis on the results of disulfide-mediated cross-linking experiments.[39] Model 2 (Figure 4(c)) emphasized constraints on particle dimensions derived from fluorescence anisotropy studies.[44] Model 3 (Figure 4(d)) attempted to merge two of the two-domain IN structures, and DNA binding was modeled using the structure of a transposon–DNA complex.[10,45] A fourth model, emphasizing photo-crosslinking data, proposed that an octamer of IN was the binding moiety, but this is inconsistent with both the gel-filtration data (Figure 2(a)) and the dimensions of the reconstructed particle.[41]

A common feature of models 1, 2, and 3 is that the IN–DNA complex has 2-fold rotational symmetry (*C*2). IN dimers are bound to each viral DNA end, and these assemble as a symmetric tetramer. However, the map in Figure 4(a) does not display *C*2 symmetry. Image reconstruction from negatively stained images must be interpreted with some caution, due to possible artifacts arising during sample preparation. Nevertheless, uranyl acetate is itself a mordant, which can rapidly fix and preserve even transient biological structures.[46] Although comparable in size to the 3D reconstruction of the IN–DNA complex, it is clear that none of the previously proposed models is a close match to the map shown in Figure 4(a).

## Implications for structure and function of IN–DNA complexes

The image reconstruction reported here suggests that the IN–DNA complex has a triangular base with a central channel, which resembles a variety of other DNA-binding proteins that wrap around their substrates, including PCNA,[47] topoisomerases,[48–50] and polymerases.[51–53] For the case of IN, the simplest interpretation is that the channel tethers the target DNA, and the lobes of density may anchor the viral DNA ends during catalysis. Circumferential binding around the target DNA may serve to exclude solvent from the IN active site during the strand transfer reaction, thereby favoring use of the viral DNA 3′ end instead of water as a nucleophile.

The unexpected asymmetry in the complex is intriguing. It is not excluded that the asymmetry results from the averaging of images of particles with different conformations. In this case, the map

**Figure 4.** (a) The 3D surface-shaded view of the IN–DNA complex at 27 Å resolution. The complex is formed by four lobes of density, three of which form a roughly triangular base (top row), and a fourth that is centered between two domains and projects from the base (bottom row). Three models have been proposed for the IN–DNA complex and are displayed as ribbon diagrams: (b) model 1;[39] (c) model 2;[44] and (d) model 3.[10] Density maps of the models shown in (b), (c) and (d) were computed at the same resolution as the experimental map. Although comparable in size, the shape of the experimental map is clearly different from the models. Compared with the view in the top row, the views in rows 2 and 3 are rotations of 45° and 90° towards the viewer, the views in rows 4 and 5 are rotations of 45° and 90° away from the viewer, and the view in row 6 is a 90° anti-clockwise rotation about the vertical axis. The scale bar represents 50 Å. The resolution of the available EM structure did not permit *de novo* modeling of the IN–DNA complex.
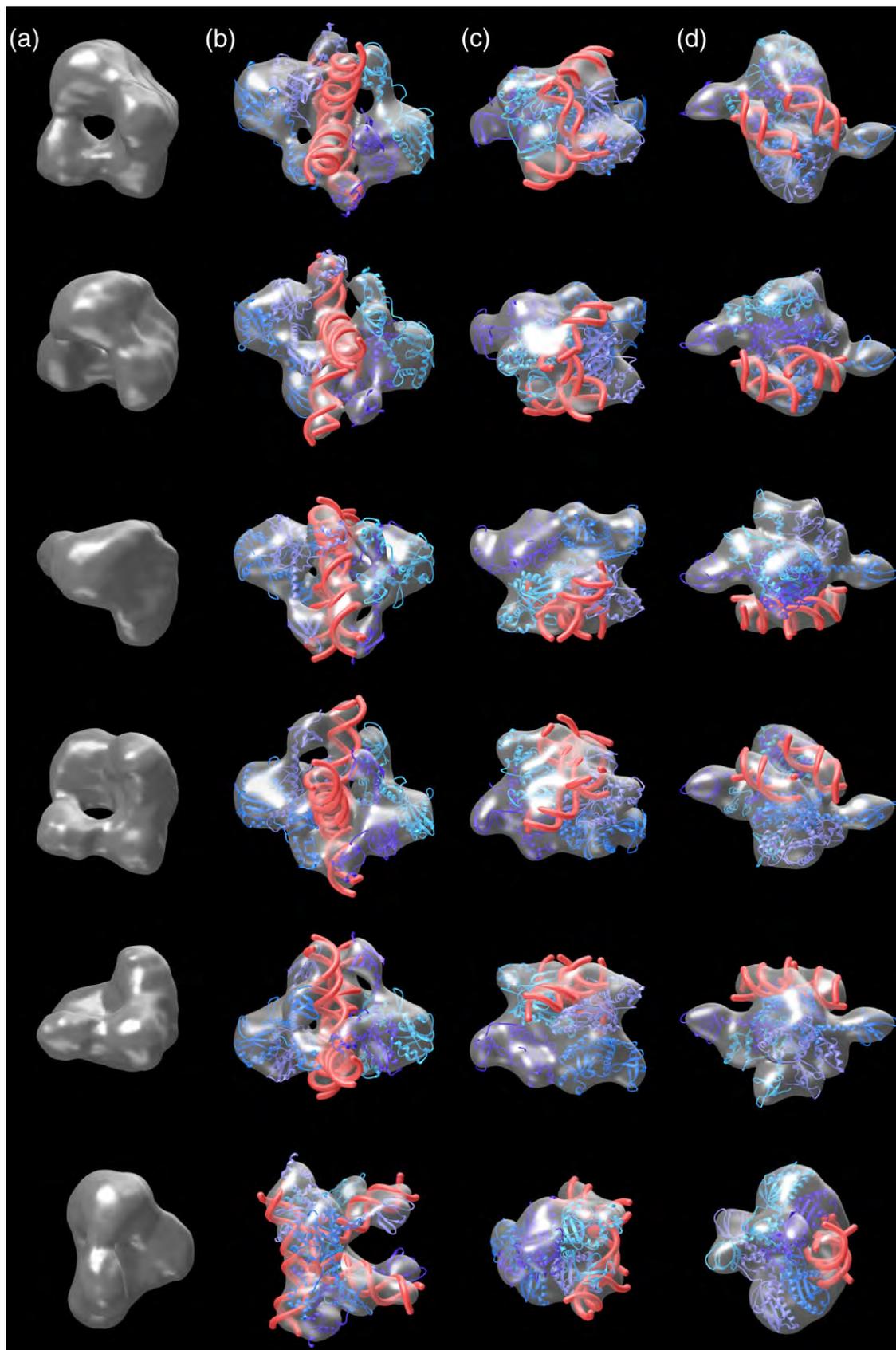
**Figure 4** (*legend on previous page*)

would be a composite and the variable features would be smeared. It is attractive to consider a model in which the structure reflects an authentic asymmetric intermediate in the integration reaction. It is possible, and perhaps likely, that strand transfer at each of the two viral DNA ends is not simultaneous but is instead sequential. If so, the integration reaction would proceed through a series of asymmetric reaction intermediates to accomplish the sequential integration of the two DNA ends and, indeed, a recent biochemical study suggested that this is the case.[28] The DNA substrate in the complex is asymmetric in primary sequence and contains the stabilizing DNA loop at only one of the two junctions between the viral DNA and the target DNA. Thus, the asymmetric DNA may have stabilized a previously unappreciated asymmetric reaction intermediate.

## Materials and Methods

### Assembly of IN–DNA complexes

The plasmid expressing IN C56S/W131D/F139D/F185H/C280S from a phage T7 promoter was constructed by replacing segments of the IN-coding region of a synthetic IN gene containing a His$_6$ affinity tag[54] with synthetic oligonucleotides.[39] The modified IN protein was expressed in *Escherichia coli* strain BL21/DE3 by the addition of IPTG. A cell pellet was resuspended in 20 mM Tris (pH 7.9), 0.2 M NaCl and lysed by sonication in the presence of 2 mg/ml of lysozyme. The suspension was adjusted to 1 M NaCl, 5 mM β-mercaptoethanol (BME), 10 mM Chaps, 5 mM imidazole, and Protease Inhibitor Cocktail 1 (Calbiochem) and then clarified by centrifugation for 30 min at 15,000 rpm in a JA20 rotor. Aggregates were removed by passage through a 0.45 μm pore size filter and purified by binding and elution on Ni-NTA agarose (Qaigen) as described.[39] A Centriprep YM-30 concentrator was used to exchange the elution buffer for thrombin buffer (20 mM Tris (pH 7.9), 1.0 M NaCl, 5 mM BME, 10 mM Chaps) and to remove imidazole. The His$_6$ tag was removed by overnight incubation with thrombin (1.5 units/ml of IN solution) at 4 °C. The cleaved His$_6$ tag was removed by repeated passage of the IN solution over the Ni-NTA agarose. Thrombin was removed by chromatography using benzamidine Sepharose 6B. IN was concentrated using a Centriprep YM-30 and dialyzed overnight at 0 °C against sample buffer (10 mM Hepes (pH 7.5), 1 M NaCl, 10 mM BME, 10 μM ZnSO$_4$, 10 mM Chaps). IN activity was assayed using end-labeled DNA substrates as described.[39]

The paired DNA three-way junction substrate was prepared by annealing 80 nmol of each of five oligonucleotides of sequence:

U3Bb 5′ CAAGTCACTGCTTTTACTGGAAGGGCTAATTA 3′;
U3Tb 5′ TAATTAGCCCTTCCACCGCGCGTAGCCACAC 3′;
U5B1b 5′ pACTGCTAGAGATTTTCC 3′;
U5B2b 5′ GTGTGGCTACG 3′;
U5Tb 5′ GGAAAATCTCAGCACGCGGGCAGTGACTTG 3′.

Oligonucleotides were mixed, heated to 95 °C and cooled to 4 °C over 45 min.

To prepare IN–DNA complexes, 80 nmol of DNA substrate was mixed with 320 nmol of purified IN in 10 mM Tris–HCl (pH 8.0), 1 M NaCl, 5 mM Chaps, 10 mM DTT, 1 mM EDTA. The mixture was dialyzed against 5 mM Hepes (pH 7.3), 5 mM DTT, 5 mM Chaps, 100 mM NaCl, 10 μM ZnSO$_4$. For gel-filtration analysis, complexes were diluted in running buffer (20 mM Hepes (pH 7.5), 80 mM NaCl, 10 mM Chaps, and 10 mM BME) to 0.1 mg/ml, or as indicated in Figure 2, and then separated in running buffer by Superose 12 chromatography.

### Electron microscopy and image analysis

IN–DNA complexes prepared as described above were diluted to 0.35 mg/ml using 20 mM Hepes (pH 7.6), 10 mM Chaps, 80 mM NaCl, 10 mM DTT. Samples (~4 μl) were incubated for 1 min at room temperature on carbon-coated Maxtaform, 300-mesh Cu/Rh grids (Ted Pella, Inc., Redding, CA) rendered hydrophilic by glow-discharge in the presence of amylamine. Excess solution was removed by blotting, and the sample was stained for 30 s with 2% (w/v) uranyl acetate. Images were recorded on Kodak SO163 film using a CM100 electron microscope (FEI/Philips) at a magnification of 52,000×(±1%) and an underfocus of ~2.5 μm. Negatives were digitized on a Zeiss SCAI flat-bed scanning densitometer (ZI/Zeiss) with a step size of 7 μm, followed by twofold pixel-averaging, which resulted in a pixel size of 2.69 Å on the object scale.

Image processing was performed with the EMAN software suite;[55] 2196 particles were manually selected and extracted as 100×100 pixel images. The absorbance histograms for the pixels in each image were scaled to the mean and standard deviation for all images. The contrast transfer function (CTF) parameters for each micrograph were determined using the routine *ctfit* from the computed Fourier transform of the carbon film of each micrograph, and phase corrections were applied to each particle image, and the particles were then centered using *cenalignint*. To minimize the influence of surrounding noise, the 100×100 pixel images were masked at 64×64 pixels. Reference free class averages of the particles were then generated using *startnrclasses* with about 60 particles in each of the 25 classes. The starting 3D model for reference-based alignment was generated by a cross common lines approach (program *startany*), using the 12 class averages with the highest signal-to-noise ratio. The class averages were first low-pass filtered to 20 Å$^{-1}$, and five rounds of iteration were performed to determine the Euler angles for each group of class averages. Projections of the 3D starting model were computed at 9° intervals. The program *refine* was used to determine the $x,y$ origin and the Euler angles for each particle by cross-correlation with the 216 projections of the starting model. Particle images with the same Euler angles were averaged, and the distribution of correlation coefficient values was determined. The images with correlation coefficients that deviated by ≥0.8σ were rejected. The final set of 216 class averages was used to generate a new 3D model. For the next round of refinement, the new model was smoothed using *three.1a.mrc*. After 20 cycles of refinement, the process was halted because the Fourier shell correlation with the previous model did not yield any substantial differences within the resolution cutoff. The final 3D map was generated from 1783 particles without applying any symmetry. To estimate the resolution of the final map, these 1783 particles were divided randomly into two groups, and two 3D maps were correlated in Fourier

space. The resolution was defined using a Fourier shell cut-off value of 0.5. The final 3D map was visualized by the use of Chimera and Vis5d software‡. A protein partial specific volume of 0.81 Da/Å$^3$ was used to set the isosurface threshold that corresponded to the molecular volume.

## Acknowledgements

## References

1. Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997). *Retroviruses.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Bushman, F. D. (2001). *Lateral DNA Transfer: Mechanisms and Consequences.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. (2002). *Mobile DNA II.* ASM Press, Washington, DC.
4. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science*, **266**, 1981–1986.
5. Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Struct. Biol.* **4**, 567–577.
6. Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T. *et al.* (1997). The solution structure of the amino-terminal HHCC domain of HIV-2 integrase; a three-helix bundle stabilized by zinc. *Curr. Biol.* **1**, 739–746.
7. Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. (2000). Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296**, 535–548.
8. Chen, J. C.-H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D. & Stroud, R. M. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl Acad. Sci. USA*, **97**, 8233–8238.
9. Chen, Z., Yan, Y., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B. *et al.* (2000). X-ray structure of simian immunodeficiency virus integrase containing the core and C-terminal domain (residues 50-293) – an initial glance of the viral DNA binding platform. *J. Mol. Biol.* **296**, 521–533.
10. Wang, J. Y., Ling, H., Yang, W. & Craigie, R. (2001). Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **20**, 7333–7343.
11. Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, R. A. & Skalka, A. M. (1997). Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.* **272**, 18161–18168.
12. Pommier, Y., Pilon, A. A., Bajaj, K., Mazumder, A. & Neamati, N. (1997). HIV-1 integrase as a target for antiviral drugs. *Antiviral Chem. Chemother.* **8**, 463–483.
13. Hansen, M. S. T., Carteau, S., Hoffmann, C., Li, L. & Bushman, F. (1998). Retroviral cDNA integration: mechanism, applications and inhibition. In *Genetic Engineering. Principles and Methods* (Setlow, J. K., ed) vol. 20, pp. 41–62 Plenum Press, New York.
14. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A. *et al.* (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science*, **287**, 646–650.
15. Farnet, C., Lipford, R., Wang, B. & Bushman, F. D. (1996). Differential inhibition of HIV-1 preintegration complexes and purified integrase protein by small molecules. *Proc. Natl Acad. Sci. USA*, **93**, 9742–9747.
16. Hazuda, D. J., Felock, P. J., Hastings, J. C., Pramanik, B. & Wolfe, A. L. (1997). Differential divalent cation requirements uncouple the assembly and catalytic reactions of human immunodeficiency virus type 1 integrase. *J. Virol.* **71**, 7005–7011.
17. Hazuda, D. J., Anthony, N. J., Gomez, R. P., Jolly, S. M., Wai, J. S., Zhuang, L. *et al.* (2004). A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc. Natl Acad. Sci. USA*, **101**, 11233–11238.
18. Bushman, F. D., Fujiwara, T. & Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science*, **249**, 1555–1558.
19. Bushman, F. D. & Craigie, R. (1990). Sequence requirements for integration of Moloney murine leukemia virus DNA *in vitro*. *J. Virol.* **64**, 5645–5648.
20. Katzman, M., Katz, R. A., Skalka, A. M. & Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J. Virol.* **63**, 5319–5327.
21. Sherman, P. A. & Fyfe, J. A. (1990). Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.
22. Craigie, R., Fujiwara, T. & Bushman, F. (1990). The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell*, **62**, 829–837.
23. Bushman, F. D. & Craigie, R. (1991). Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. *Proc. Natl Acad. Sci. USA*, **88**, 1339–1343.
24. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell*, **63**, 87–95.
25. Miller, M. D., Farnet, C. M. & Bushman, F. D. (1997).

‡ http://www.ssec.wisc.edu/~billh/vis5d.html

Human immunodeficiency virus type 1 preintegra-
tion complexes: studies of organization and composi-
tion. *J. Virol.* **71**, 5382–5390.

26. Patel, P. H. & Preston, B. D. (1994). Marked infidelity
of human immunodeficiency virus type 1 reverse
transcriptase at RNA and DNA template ends. *Proc.
Natl Acad. Sci. USA*, **91**, 549–553.

27. Ellison, V. & Brown, P. O. (1994). A stable complex
between integrase and viral DNA ends mediates
human immunodeficiency virus integration *in vitro*.
*Proc. Natl Acad. Sci. USA*, **91**, 7316–7320.

28. Li, M., Mizuuchi, M., Burke, T. R. J. & Craigie, R.
(2006). Retroviral DNA integration: reaction pathway
and critical intermediates. *EMBO J.* **25**, 1295–1304.

29. Engelman, A., Mizuuchi, K. & Craigie, R. (1991). HIV-
1 DNA integration: mechanism of viral DNA cleavage
and DNA strand transfer. *Cell*, **67**, 1211–1221.

30. Yoder, K. & Bushman, F. D. (2000). Repair of gaps in
retroviral DNA integration intermediates. *J. Virol.* **74**,
11191–11200.

31. Engelman, A., Bushman, F. D. & Craigie, R. (1993).
Identification of discrete functional domains of HIV-1
integrase and their organization within an active
multimeric complex. *EMBO J.* **12**, 3269–3275.

32. van Gent, D. C., Vink, C., Oude Groeneger, A. A. M. &
Plasterk, R. H. A. (1993). Complementation between
HIV integrase proteins mutated in different domains.
*EMBO J.* **12**, 3261–3267.

33. Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R.
(1996). A soluble active mutant of HIV-1 integrase: in-
volvement of both the core and carboxyl-terminal do-
mains in multimerization. *J. Biol. Chem.* **271**, 7712–7718.

34. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M. &
Skalka, A. M. (1992). Retroviral integrase functions as
a multimer and can turn over catalytically. *J. Biol.
Chem.* **267**, 16037–16040.

35. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R.,
Davies, D. R. & Craigie, R. (1995). Catalytic domain of
human immunodeficiency virus type 1 integrase: Identi-
fication of a soluble mutant by systematic replacement of
hydrophobic residues. *Proc. Natl Acad. Sci. USA*, **92**,
6057–6061.

36. Johnson, E. P. & Bushman, F. D. (2001). Paired DNA
three-way junctions as scaffolds for assembling
integrase complexes. *Virology*, **286**, 304–316.

37. Bao, K. K., Wang, H., Miller, J. K., Erie, D. A., Skalka,
A. M. & Wong, I. (2002). Functional oligomeric state
of avian sarcoma virus integrase. *J. Biol. Chem.* **278**,
1323–1327.

38. Chow, S. A. & Brown, P. O. (1994). Juxtaposition of
two viral DNA Ends in a bimolecular disintegration
reaction mediated by multimers of human immuno-
deficiency virus type 1 or murine leukemia virus
integrase. *J. Virol.* **68**, 7869–7878.

39. Gao, K., Butler, S. L. & Bushman, F. D. (2001). Human
immunodeficiency virus type 1 integrase: arrange-
ment of protein domains in active cDNA complexes.
*EMBO J.* **20**, 3565–3576.

40. Heuer, T. S. & Brown, P. O. (1997). Mapping features

of HIV-1 integrase near selected sites on viral and
target DNA molecules in an active enzyme-DNA
complex by photo-cross-linking. *Biochemistry*, **36**,
10655–10665.

41. Heuer, T. S. & Brown, P. O. (1998). Photo-cross-linking
studies suggest a model for the architecture of an
active human immunodeficiency virus type-1 inte-
grase-DNA complex. *Biochemistry*, **37**, 6667–6678.

42. Jenkins, T. M., Esposito, D., Engelman, A. & Craigie,
R. (1997). Critical contacts between HIV-1 integrase
and viral DNA identified by structure-based analysis
and photo-crosslinking. *EMBO J.* **16**, 6849–6859.

43. Esposito, D. & Craigie, R. (1998). Sequence specificity of
viral end DNA binding by HIV-1 integrase reveals critical
regions for proteIN–DNA interaction. *EMBO J.* **17**,
5832–5843.

44. Podtelezhnikov, A. A., Gao, K., Bushman, F. D. &
McCammon, A. (2003). Modelling HIV-1 integrase
complexes based on their hydrodynamic structure.
*Biopolymers*, **68**, 110–120.

45. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S. &
Rayment, I. (2000). Three-dimensional structure of the
Tn5 synaptic complex transposition intermediate.
*Science*, **289**, 77–85.

46. Zhao, F. Q. & Craig, R. (2003). Capturing time-
resolved changes in molecular structure by negative
staining. *J. Struct. Biol.* **141**, 43–52.

47. Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. &
Kuriyan, J. (1994). Crystal structure of the eukaryotic
DNA polymerase processivity factor PCNA. *Cell*, **79**,
1233–1243.

48. Berger, J. M., Gamblin, S. J., Harrison, S. C. & Wang,
J. C. (1996). Structure and mechanism of DNA
topoisomerase II. *Nature*, **379**, 225–232.

49. Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J.
& Hol, W. G. J. (1998). Crystal structures of human
topoisomerase i in covalent and noncovalent com-
plexes with DNA. *Science*, **279**, 1504–1513.

50. Perry, K., Hwang, Y., Bushman, F. D. & Van Duyne, G.
(2006). Structural basis for specificity in the poxvirus
topoisomerase. *Mol. Cell*, **23**, 343–354.

51. Ollis, D. L., Kline, C. & Steitz, T. A. (1985). Domain of
*E. coli* DNA polymerase I showing sequence homol-
ogy to T7 DNA polymerase. *Nature*, **313**, 818–819.

52. Arnold, E., Jacobo-Molina, A., Nanni, R. G., Wil-
liams, R. L., Lu, X., Ding, J. *et al.* (1992). Structure of
HIV-1 reverse transcriptase/DNA complex at 7 A
resolution showing active site locations. *Nature*, **357**,
85–89.

53. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A.
& Steitz, T. A. (1992). Crystal structure at 3.5 A
resolution of HIV reverse transcriptase complexed
with an inhibitor. *Science*, **256**, 1783–1790.

54. Gerton, J. L. & Brown, P. O. (1997). The core domain of
HIV-1 integrase recognizes key features of its DNA
substrates. *J. Biol. Chem.* **272**, 25809–25815.

55. Ludtke, S. J., Baldwin, P. R. & Chiu, W. (1999). EMAN:
semiautomated software for high-resolution single-
particle reconstructions. *J. Struct. Biol.* **128**, 82–97.

*Edited by J. Karn*