

## HIV Integrase, a Brief Overview from Chemistry to Therapeutics\*

Published, JBC Papers in Press, May 9, 2001,  
DOI 10.1074/jbc.R100027200

Robert Craigie‡

From the Laboratory of Molecular Biology,  
NIDDK, National Institutes of Health,  
Bethesda, Maryland 20892

Retroviruses are a large and diverse family of RNA viruses that synthesize a DNA copy of their RNA genome after infection of the host cell. Integration of this viral DNA into host DNA is an essential step in the replication cycle of HIV<sup>1</sup> and other retroviruses (reviewed in Refs. 1–3). The integrated viral DNA is transcribed to make the RNA genome of progeny virions and the template for translation of viral proteins. Following assembly, virions bud from the cell surface and subsequently infect previously uninfected cells, thus completing the replication cycle. An infecting retrovirus introduces a large nucleoprotein complex into the cytoplasm of the host cell. This complex, which is derived from the core of the infecting virion, contains two copies of the viral RNA together with a number of viral proteins, including reverse transcriptase and integrase. Reverse transcription of the viral RNA occurs within the complex to make a double-stranded DNA copy of the viral genome, the viral DNA substrate for integration. The viral DNA remains associated with both viral and cellular proteins in a nucleoprotein complex termed the preintegration complex. One constituent of the preintegration complex is the viral integrase protein, the key player in the integration of the viral DNA into the host genome. The other components of the preintegration complex that are transported to the nucleus along with the viral DNA and integrase, and their possible functions, have not been firmly established and are not discussed here. The critical DNA cutting and joining events that integrate the viral DNA are carried out by the integrase protein itself. Here we review our current knowledge of the molecular mechanism of this reaction and discuss some of the key issues that are yet to be understood.

### The Mechanism of DNA Integration

Biochemical studies have elucidated the basic chemical mechanism of integration, even though the organization of the active complex of integrase with its DNA substrates remains to be determined. We will focus on HIV integrase, but the key properties of this enzyme appear to be shared among the entire retroviral integrase family. In the first step of the integration process, two nucleotides are removed from each 3'-end of the viral DNA, a reaction termed 3'-end processing. Cleavage occurs to the 3'-side of a CA dinucleotide that is conserved among retroviruses, retrotransposons, and many DNA transposons,

both in prokaryotes and eukaryotes. This reaction exposes the terminal 3'-hydroxyl group that is to be joined to target DNA (Fig. 1B). In the second step, DNA strand transfer, a pair of processed viral DNA ends is inserted into the target DNA (Fig. 1C). In the case of HIV, the sites of integration on the two target DNA strands are separated by 5 base pairs. Repair of this integration intermediate (Fig. 1D) results in a direct duplication of 5 base pairs flanking the integrated viral DNA (not shown). The repair step requires removal of the two unpaired nucleotides at the 5'-ends of the viral DNA, filling in the single gaps, and finally ligation. Integrase is responsible for 3'-processing and DNA strand transfer, but the latter repair steps are likely to be carried out by cellular enzymes (4, 5). There is little specificity for the sites of integration in host DNA, and insertion can occur at essentially any location. The DNA cutting and joining steps of 3'-end processing and DNA strand transfer closely parallel the reactions used by many transposons to move to new sites in the genome.

The 3'-processing and DNA strand transfer reactions can be carried out *in vitro* with purified integrase, a duplex oligonucleotide that mimics one end of the viral DNA, and a divalent metal ion. This simplified *in vitro* system has proved to be invaluable for dissecting the biochemical mechanism of DNA integration. Stereochemical analysis of the reaction pathway has demonstrated that both 3'-processing and DNA strand transfer occur by a one-step transesterification mechanism (6), the same result that was previously obtained for the corresponding reactions mediated by the related Mu transposase protein (7). In the 3'-processing reaction, water serves as the nucleophile for cleavage at the ends of the viral DNA. We may envisage that DNA strand transfer occurs by a chemically similar mechanism, except that integrase positions the 3'-hydroxyl groups at the ends of the viral DNA to simultaneously cleave the target DNA and make covalent connections between the viral and target DNA. Although 3'-processing and DNA strand transfer are very similar reactions at the chemical level, the way the active site region of integrase engages DNA substrate must differ between processing and DNA strand transfer; for the latter reaction the active sites must accommodate target DNA in addition to the viral DNA ends.

### Integrase

HIV-1 integrase is comprised of three domains (Fig. 2) based on the susceptibility of the linker regions to proteolysis (8), functional studies (8–10), and the structures of the domains (Fig. 3), which have been individually determined by x-ray crystallography or NMR.

The catalytic core domain contains the invariant triad of acidic residues, the D,D-35-E motif (8, 11–13), comprising residues Asp<sup>64</sup>, Asp<sup>116</sup>, and Glu<sup>152</sup> in the case of HIV-1 integrase. Mutagenesis of these residues and their counterparts in related retroviral integrases abolishes or severely diminishes all catalytic activities in parallel (8, 11, 14, 15). By analogy with models of catalysis by DNA polymerases (16–18), it has been proposed that coordination of divalent metal ion to these residues plays a key role in catalysis (11). The structures of catalytic domain of HIV-1 integrase (19, 20) and the corresponding domain ASV integrase (21, 22) have been determined by x-ray crystallography. The catalytic residues Asp<sup>64</sup>, Asp<sup>116</sup>, and Glu<sup>152</sup> of HIV-1 integrase and their counterparts in the ASV structures are in close proximity, coordinate divalent metal ion,

\* This minireview will be reprinted in the 2001 Minireview Compendium, which will be available in December, 2001.

‡ To whom correspondence should be addressed: Bldg. 5, Rm. 301, 5 Center Dr. MSC 0560, NIDDK, National Institutes of Health, Bethesda, MD 20892-0560. Tel.: 301-496-4081; Fax: 301-496-0201; E-mail: bobc@helix.nih.gov.

<sup>1</sup> The abbreviations used are: HIV, human immunodeficiency virus; ASV, avian sarcoma virus.

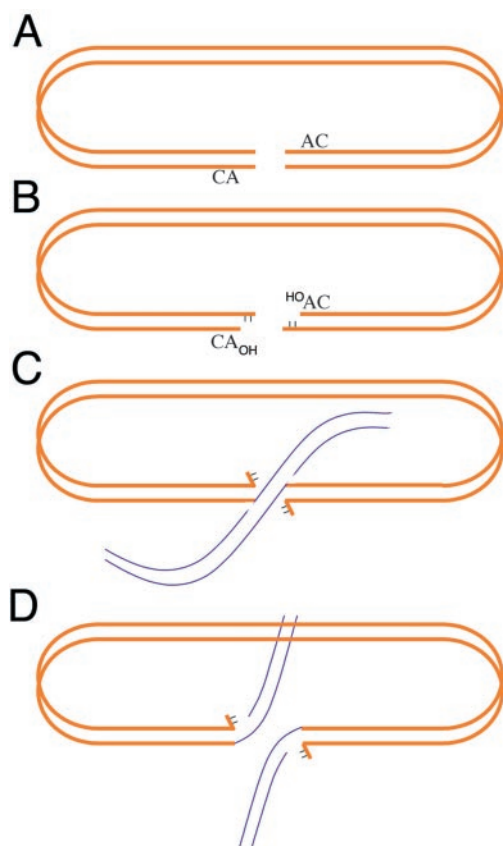


FIG. 1. DNA cutting and joining steps in retroviral integration.

A, the viral DNA (orange) made by reverse transcription is linear and blunt ended. B, in the first step of the integration process, 3'-end processing, two nucleotides are cleaved from each 3'-end of the viral DNA. C, in the next step, DNA strand transfer, the hydroxyl groups at the 3'-ends of the processed viral DNA attack a pair of phosphodiester bonds in the target DNA (blue). The spacing between the sites of attack on each target DNA strand is fixed and characteristic for each retrovirus. D, the resulting integration intermediate is redrawn to clarify the connections between viral and target DNA. Integrase is responsible for both the 3'-processing and DNA strand transfer reactions that give rise to the integration intermediate. Completion of DNA integration requires removal of the two unpaired nucleotides at the 5'-ends of the viral DNA, filling in the single strand gaps between host and viral DNA by a DNA polymerase, and finally ligation. These steps are likely to be carried out by cellular enzymes.

N-term		Catalytic core		C-term
HH	CC	D	D-35-E	
HTH fold		RNaseH fold		SH3 fold
Retroviruses		Retroviruses		Less conserved
Retrotransposons		Retrotransposons Transposons		

FIG. 2. Domain structure of retroviral integrases. Integrase is composed of three domains. The core domain contains the catalytic site. A triad of acidic residues, the D,D-35-E motif, plays a key role in catalysis. This domain is well conserved not only among retroviruses but also among many DNA transposons in both prokaryotes and eukaryotes. The N-terminal domain includes the conserved HHCC motif, which binds zinc. The C-terminal domain is less well conserved.

and define the active site. However, the residues comprising the active site region exhibit considerable flexibility, suggesting that binding of DNA substrate is required to impose the precise configuration of residues that is required for catalysis. The structures of the HIV-1 and ASV integrase core domains

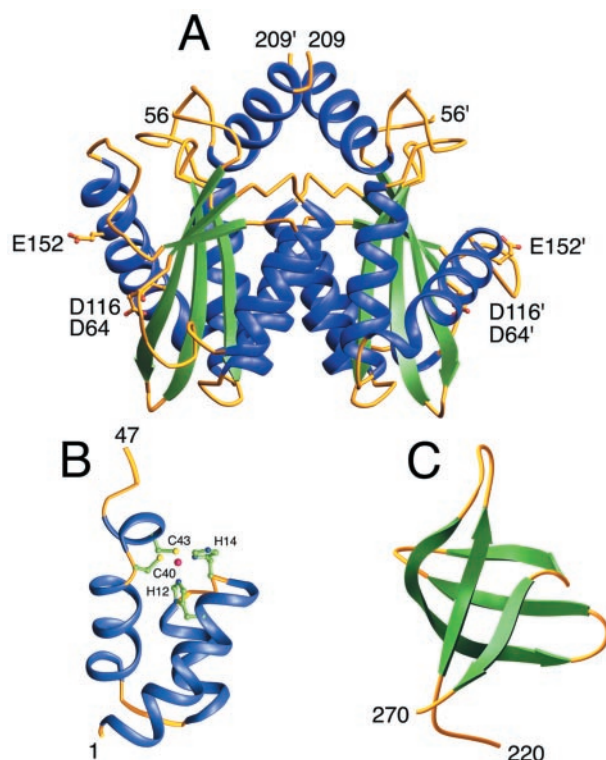


FIG. 3. Structures of the three domains of HIV-1 integrase shown as ribbon diagrams (48). A, the catalytic core domain; B, the N-terminal domain; C, the C-terminal domain. Protein Data Bank codes are 1BIS, 1WJC, and 1IHV, respectively.

are very similar to each other and to the catalytic domain of Mu transposase (23), reinforcing the parallelism of retroviral DNA integration and transposition. These structures revealed that retroviral integrases and their transposase cousins belong to a superfamily of polynucleotidyltransferases that share the same overall fold as *Escherichia coli* RNase H and have similar active sites (24, 25).

The HIV-1 catalytic domain is dimeric in solution (26) and in the crystal structures (Fig. 3A). The extensive surface area of the dimer interface suggests that it is biologically relevant. Yet, the spacing between the active sites in the nearly spherical dimer is not compatible with the spacing between the sites of insertion on the two strands of target DNA. The sites of insertion on each strand of target DNA are separated by 5 base pairs, corresponding to about 15 Å for helical B-form DNA. The functional unit of integrase should therefore have a pair of active sites separated by a similar spacing. However, in the crystal structures (Fig. 3A), the active sites in the dimer are separated by more than 30 Å measured as a straight line through the protein and by an even greater distance measured around the circumference of the dimer. Assuming that the dimer interface is maintained in the functional integrase multimer, at least a tetramer of integrase must be required for the complete integration reaction.

The N-terminal domain of HIV-1 integrase contains a conserved pair of His and Cys residues, a motif similar to the zinc-coordinating residues of zinc fingers. Although this domain does indeed bind zinc (27, 28), its structure (Fig. 3B) (29) is totally different from that of zinc fingers. It consists of a bundle of three  $\alpha$ -helices (Fig. 3C) (29, 30). It has an SH3 fold, although there is no known functional relationship with the SH3 domains of other proteins.

Although the core domain of integrase is clearly responsible for catalysis, the functional roles of the other two domains are less clear. The C-terminal domain binds DNA nonspecifically.

Because the sites of integration into target DNA are relatively nonspecific, it has been suggested that this domain may interact with target DNA. However, experiments with chimeric integrases (31, 32) assign recognition of the target site to the core domain, and cross-linking studies (33–36) suggest that the C-terminal domain interacts with a subterminal region just inside the very ends of the viral DNA end. The C-terminal domain of retroviral integrases may therefore play a similar role to that of the site-specific DNA binding domain of transposases, which also recognize a subterminal sequence at the ends of the transposon DNA. The function of the N-terminal domain of integrase is at present unknown.

### Unanswered Questions

Although the structures of all three domains of integrase have been individually determined, their spatial arrangement in the active complex with DNA substrate is unknown. Three retroviral integrase structures of the core together with the C-terminal domain have been recently reported in the absence of bound DNA (37–39). The spatial relationship between the core and C-terminal domains is different in each of these structures indicating considerable flexibility in the linkage between the two domains. It is likely that binding of viral DNA substrate imposes the proper configuration of domains for the reaction to occur. Although integrase exists as monomers, dimers, and tetramers at high ionic strength, formation of large aggregates under reaction conditions has frustrated attempts to directly determine the organization of the active unit. Efforts to crystallize integrase together with DNA are further challenged by the nonspecific nature of the interaction between integrase and the viral DNA ends.

Many transposases, unlike their retroviral integrase counterparts, bind specifically to the transposon ends and form discrete nucleoprotein complexes that are amenable to direct structural and functional analysis. In the case of Mu transposase, the active unit of transposase is tetramer. Within the tetramer, only two of the four active sites directly participate in the DNA cleavage and joining reactions. A dimer of transposase carries out these reactions in the case of Tn5 (40) and Tn10.<sup>2</sup> Thus, a dimer would seem to constitute the fundamental unit for this class of reactions, and the requirements for higher order multimers in some systems probably reflect differences in regulatory systems rather than a fundamental mechanism. The recent determination of the crystal structure of the Tn5 transposase dimer in complex with DNA substrate (40) provides a platform for modeling the interactions of DNA with the catalytic domain of integrase. Because the integrase catalytic domain is itself a dimer with an extensive interface, a similar architecture to that of the Tn5 complex would require a pair of dimers. Then, as in the case of Mu transposase, two of the active sites in the resulting tetramer would not directly participate in catalysis.

A puzzling phenomenon is the almost exclusive integration of a single viral DNA substrate into one strand of the DNA target in reactions with purified HIV integrase. In the cell and *in vitro* with preintegration complexes, integration is coupled so that a pair of viral DNA ends is integrated with a spacing of 5 base pairs separating the sites of insertion on the two strands of target DNA. Although inclusion of additional protein factors in the reaction has been reported to stimulate coupled integration (41, 42), single-end integration events still predominate. It appears that a yet to be understood assembly pathway, and possibly additional factors, are required to reconstitute the integration reaction with the full fidelity observed with preintegration complexes *in vitro* and *in vivo*. An alternative possi-

bility is that aggregation of purified integrase disfavors assembly of the correct complexes with two DNA ends poised for integration; such aggregation may normally be prevented by interaction with other components of the preintegration complex. Indeed, it has been shown that improving the solubility of the closely related Tn552 transposase greatly enhances the efficiency of double-end *versus* single-end strand transfer (43). Perhaps the strongest suggestion that additional host proteins may not be required is the finding that disrupted HIV-1 virions, which contain few host proteins, support efficient double-end strand transfer of exogenous viral substrate DNA (44).

### Prospects for Developing Integrase Inhibitors as Therapeutic Antiviral Agents

The development of effective inhibitors of HIV replication targeted to reverse transcriptase and protease has demonstrated the potential effectiveness of antiviral therapy for the treatment of AIDS. Drugs targeted to integrase would be a valuable complement to reverse transcriptase and protease inhibitors. However, no drugs have yet been developed that act on integrase. The bottleneck has been the absence of good lead compounds to serve as the starting point for drug development. Although many compounds have been reported to inhibit integrase, most of these lack selectivity and inhibit many other enzymes as well. A concern has also been raised that because many copies of integrase enter the cell with the infecting virus and only two integrations are required, integrase might be an intrinsically difficult enzyme to target. Recent identification of a class of compounds that unambiguously inhibits HIV replication in cell culture by targeting integrase (45) counters this argument and demonstrates the potential of integrase as an antiviral target. The structure of a related inhibitor complexed with the active site of integrase has also recently been determined (46). However, the much higher affinity of these inhibitors for integrase in the presence of viral DNA (47) and selectivity for the DNA strand transfer reaction suggest that the binding modes in the absence and presence of DNA substrate are not identical. It will likely be necessary to determine the structure of these and other inhibitors in complex with an integrase active site engaged with DNA to understand the interaction in detail and provide a platform for drug design. Nevertheless, the foundation of basic knowledge in understanding the mechanism of retroviral integration promises to bear fruit in contributing to the fight against AIDS.

*Acknowledgments*—I thank Wei Yang for help with preparation of the ribbon diagrams and Wei Yang and Kenji Adzuma for comments on the manuscript.

### REFERENCES

- Brown, P. O. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., eds) pp. 161–203, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Asante-Appiah, E., and Skalka, A. M. (1997) *Antiviral Res.* **36**, 139–156
- Hindmarsh, P., and Leis, J. (1999) *Microbiol. Mol. Biol. Rev.* **63**, 836–843
- Brin, E., Yi, J. Z., Skalka, A. M., and Leis, J. (2000) *J. Biol. Chem.* **275**, 39287–39295
- Yoder, K. E., and Bushman, F. D. (2000) *J. of Virol.* **74**, 11191–11200
- Engelman, A., Mizuuchi, K., and Craigie, R. (1991) *Cell* **67**, 1211–1221
- Mizuuchi, K., and Adzuma, K. (1991) *Cell* **66**, 129–140
- Engelman, A., and Craigie, R. (1992) *J. Virol.* **66**, 6361–6369
- Engelman, A., Bushman, F. D., and Craigie, R. (1993) *EMBO J.* **12**, 3269–3275
- van Gent, D. C., Vink, C., Groeneger, A. A., and Plasterk, R. H. (1993) *EMBO J.* **12**, 3261–3267
- Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338
- Polard, P., and Chandler, M. (1995) *Mol. Microbiol.* **15**, 13–23
- Rowland, S. J., and Dyke, K. G. (1990) *Mol. Microbiol.* **4**, 961–975
- Leavitt, A. D., Shiue, L., and Varmus, H. E. (1993) *J. Biol. Chem.* **268**, 2113–2119
- van Gent, D. C., Groeneger, A. A., and Plasterk, R. H. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9598–9602
- Beese, L. S., and Steitz, T. A. (1991) *EMBO J.* **10**, 25–33
- Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8924–8928

<sup>2</sup> D. Haniford, personal communication.

18. Steitz, T. A., and Steitz, J. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6498–6502
19. Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) *Science* **266**, 1981–1986
20. Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R., and Davies, D. R. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9150–9154
21. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1995) *J. Mol. Biol.* **253**, 333–346
22. Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A., and Skalka, A. M. (1996) *Structure* **4**, 89–96
23. Rice, P., and Mizuuchi, K. (1995) *Cell* **82**, 209–220
24. Yang, W., and Steitz, T. A. (1995) *Structure* **3**, 131–134
25. Rice, P., Craigie, R., and Davies, D. R. (1996) *Curr. Opin. Struct. Biol.* **6**, 76–83
26. Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlardo, R., Davies, D. R., and Craigie, R. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6057–6061
27. Burke, C. J., Sanyal, G., Bruner, M. W., Ryan, J. A., LaFemina, R. L., Robbins, H. L., Zeff, A. S., Middaugh, C. R., and Cordingley, M. G. (1992) *J. Biol. Chem.* **267**, 9639–9644
28. Zheng, R., Jenkins, T. M., and Craigie, R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13659–13664
29. Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M., and Gronenborn, A. M. (1995) *Biochemistry* **34**, 9826–9833
30. Eijkelenboom, A. P., Lutzke, R. A., Boelens, R., Plasterk, R. H., Kaptein, R., and Hard, K. (1995) *Nat. Struct. Biol.* **2**, 807–810
31. Katzman, M., and Sudol, M. (1995) *J. Virol.* **69**, 5687–5696
32. Shibagaki, Y., and Chow, S. A. (1997) *J. Biol. Chem.* **272**, 8361–8369
33. Esposito, D., and Craigie, R. (1998) *EMBO J.* **17**, 5832–5843
34. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16**, 6849–6859
35. Heuer, T. S., and Brown, P. O. (1997) *Biochemistry* **36**, 10655–10665
36. Heuer, T. S., and Brown, P. O. (1998) *Biochemistry* **37**, 6667–6678
37. Chen, Z. G., Yan, Y. W., Munshi, S., Li, Y., Zugay-Murphy, J., Xu, B., Witmer, M., Felock, P., Wolfe, A., Sardana, V., Emini, E. A., Hazuda, D., and Kuo, L. C. (2000) *J. Mol. Biol.* **296**, 521–533
38. Chen, J. C. H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8233–8238
39. Yang, Z. N., Mueser, T. C., Bushman, F. D., and Hyde, C. C. (2000) *J. Mol. Biol.* **296**, 535–548
40. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000) *Science* **289**, 77–85
41. Carteau, S., Gorelick, R. J., and Bushman, F. D. (1999) *J. Virol.* **73**, 6670–6679
42. Hindmarsh, P., and Leis, J. (1999) *Adv. Virus Res.* **52**, 397–410
43. Leschziner, A. E., Griffin, T. J., and Grindley, N. D. F. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7345–7350
44. Goodarzi, G., Im, G. J., Brackmann, K., and Grandgenett, D. (1995) *J. Virol.* **69**, 6090–6097
45. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000) *Science* **287**, 646–650
46. Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H., and Davies, D. R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13040–13043
47. Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J. Y., Young, S., Hamill, T., Cole, J. L., and Hazuda, D. J. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11244–11249
48. Carson, M. (1987) *J. Mol. Graph.* **5**, 103–106