Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain

Valentina Molteni,[a] Jason Greenwald,[a,b] Denise Rhodes,[c] Young Hwang,[c] Witek Kwiatkowski,[b] Frederic D. Bushman,[c] Jay S. Siegel[a] and Senyon Choe[b]*

[a]Department of Chemistry, University of California, San Diego, La Jolla, CA 92093-0358, USA, [b]Structural Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92186-5800, USA, and [c]Infectious Disease Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037, USA

Correspondence e-mail: choe@salk.edu

Received 2 November 2000
Accepted 23 January 2001

PDB References: HIV integrase catalytic domain, complex with compound (1), 1hyv; complex with compound (3), 1hyz.

Integration of the reverse-transcribed HIV cDNA into the host DNA is a required step in viral replication. The virus-encoded integrase protein catalyzes the initial DNA breaking and joining reactions that mediate cDNA integration. Here, the identification by X-ray crystallography of a small-molecule binding site on the integrase catalytic domain is reported. The small-molecule family studied consists of a core of arsenic or phosphorus surrounded by four aromatic groups. Two arsenic derivatives were visualized bound to integrase. In each case, two molecules bound at symmetry-related sites on the catalytic domain dimer interface. The first compound studied, tetraphenyl arsonium, did not inhibit integrase. However, a synthetic compound substituting a catechol for one of the phenyl rings, dihydroxyphenyltriphenylarsonium, bound to the same site and did inhibit the enzyme. Changes in the vicinity of the catalytic site were seen with the inhibitory compound only, potentially explaining its mechanism of action. Further substituting phosphonium for arsonium yielded a compound with an IC_{50} in the low micromolar range. These findings may be useful in designing new inhibitors of integrase, which is at present the only one of the three HIV enzymes for which clinically useful inhibitors are not available.

1. Introduction

The human immunodeficiency virus type 1 (HIV-1) encodes three enzymes, reverse transcriptase (RT), protease (PR) and integrase (IN). Drugs targeting RT and PR are in wide use, but owing to the development of resistant strains new drugs are needed. HIV IN is an attractive target since it is the third HIV enzyme and there are no known cellular enzymes that function similarly, raising the hope that integrase inhibitors could be relatively non-toxic (Hansen *et al.*, 1998; Bushman *et al.*, 1998; Pommier *et al.*, 1997; Coffin *et al.*, 1997).

HIV-1 IN is composed of three functional domains (Eijkelenboom *et al.*, 1997; Cai *et al.*, 1997; Yang & Steitz, 1995; Dyda *et al.*, 1994; Bujacz *et al.*, 1995). The N-terminal domain is comprised of residues 1–50 and contains a HHCC motif that binds zinc ions. The catalytic domain (comprised of amino acids 50–212 and henceforth referred to as 50–212), contains a triad of invariant acidic amino acids [the D,D(35)E motif] which is required for catalysis. These residues bind two metal atoms that carry out the chemical steps of catalysis. X-ray crystallographic analyses of the avian sarcoma virus (ASV) IN and HIV IN catalytic domains have revealed that they share very high structural homology. Both contain a

mixed α-helix β-sheet fold found in diverse polynucleotide phosphotransfer enzymes. The C-terminal domain, residues ~212–288, resembles the Src homology 3 (SH3) fold and is known to bind DNA strongly but not specifically. All three integrase domains are required for full catalytic activity, although the purified catalytic domain can carry out a permissive back reaction called disintegration (Chow *et al.*, 1992; Bushman *et al.*, 1993; Vink *et al.*, 1993). Purified recombinant proteins comprising each domain can form dimers and full-length integrase is known to act as a multimer (Engelman *et al.*, 1993; van Gent *et al.*, 1993). The recent determination of X-ray structures for two-domain fragments containing the catalytic and C-terminal domains has begun to reveal the higher order structure of integrase complexes (Yang *et al.*, 2000; Chen, Krucinski *et al.*, 2000; Chen, Yan *et al.*, 2000).

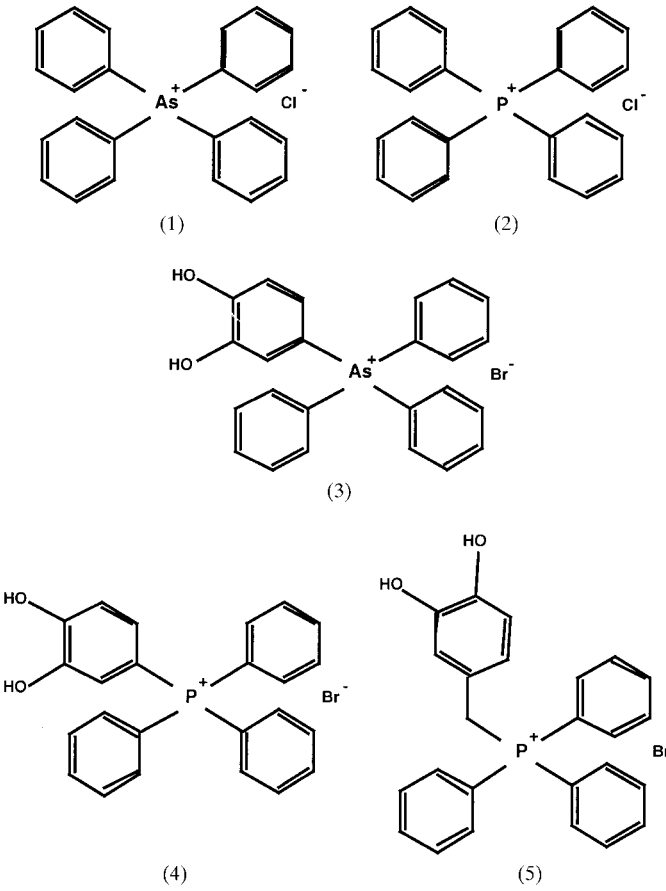
IN protein carries out the initial DNA breaking and joining reactions responsible for the attachment of HIV cDNA to host DNA. Prior to integration, two nucleotides are removed from each 3′-end in the linear cDNA precursor (terminal cleavage) (Roth *et al.*, 1989; Katzman *et al.*, 1989; Sherman & Fyfe, 1990; Craigie *et al.*, 1990). This reaction may be important to the virus by preparing a defined substrate for subsequent reaction steps (Patel & Preston, 1994; Miller *et al.*, 1997). The recessed 3′-ends are then joined to protruding 5′-ends of breaks made in the target DNA (strand transfer) (Craigie *et al.*, 1990; Bushman *et al.*, 1990; Katz *et al.*, 1990). The remaining DNA strands are then attached, probably by the action of host DNA-repair enzymes, to complete formation of an integrated provirus. The terminal cleavage and strand-transfer reactions can be modeled *in vitro* by using purified recombinant IN protein. Under simple reaction conditions, purified IN can form a covalent bond between a double-stranded DNA substrate that mimics the viral long terminal repeat (LTR) and another DNA mimicking the integration target.

Systematic screening of potential inhibitors has been undertaken using mostly purified IN-based assays. From such screens several IN inhibitors have now been identified, but to date no clinically useful inhibitors have been developed and only a handful display antiviral activity (Hansen *et al.*, 1998; Pommier *et al.*, 1997; Hazuda *et al.*, 2000; King & Robinson, 1998; Reddy *et al.*, 1999). Alternatively, one can identify small molecules that bind to the integrase catalytic domain using X-ray crystallographic analysis of small molecule–integrase complexes and attempt to build up specificity and affinity groups into the compound. Where possible, we have used compounds that contain heavy atoms that have useful X-ray absorption properties (anomalous scattering) as the starting material to facilitate inhibitor identification. Here, we describe the crystal structures of 50–212 in complex with tetraphenylarsonium chloride (1) or 3,4-dihydroxyphenyltriphenylarsonium bromide (3). These compounds were found to bind specifically at the 50–212 dimer interface in the crystal. Although the mechanism of inhibition is not fully clarified, we have shown that one such chemical derivative we synthesized possesses an inhibitory activity with an IC_{50} in the low micromolar range.

2. Experimental procedures

2.1. Crystallization and structure determination

HIV integrase core domain (50–212) was purified as previously described (Jenkins *et al.*, 1995; Greenwald *et al.*, 1999). Crystals of 50–212 were obtained by vapor diffusion (1:1 mixture with the protein at 10 mg ml^{-1}) at 277 K using 15% PEG 8000, 100 m*M* cacodylic acid pH 6.5, 200 m*M* $(NH_4)_2SO_4$, 5 m*M* DTT, 5 m*M* $MnCl_2$ and 5 m*M* $MgCl_2$ as a precipitant (Greenwald *et al.*, 1999), essentially as reported by Dyda *et al.* (1994). It is known that in this crystallization condition two cysteines (Cys65 and Cys130) are covalently modified by arsenic originating from the cacodylate buffer. The cacodylation causes a local conformational alteration. The size of the crystals was increased by feeding fresh protein into a drop which contained a few large crystals. In this way, we could grow crystals that were over 0.5 mm on a side and diffracted to 2 Å on a rotating-anode X-ray source and to 1.5 Å at a synchrotron.



The compounds (1) and (3) were introduced into the crystallized protein by soaking the crystals in a solution of cryoprotectant with 2.5 m*M* compound. The cryoprotectant consisted of the same components as the crystallization solution with the addition of 30% ethylene glycol and the reduction of PEG to 10% and $(NH_4)_2SO_4$ to 100 m*M*. The crystals were soaked for 24–72 h and then flash-frozen in a nitrogen stream at 100 K. Data for the complex of 50–212 with compound (1) were collected at SSRL beamline 1-5 using the

absorption maximum at 1.04 Å in order to maximize the anomalous scattering contribution from arsenic. Data for the complex of 50–212 with compound (3) were collected using a Cu $K\alpha$ X-ray source on a rotating anode. We used the inverse-beam collection strategy to complete the F^+ and F^- anomalous pairs, to reduce systematic errors introduced by data collection and to increase the data redundancy. The data were indexed with *DENZO* and integrated with *SCALEPACK* (Otwinowski, 1993). The data were refined with *REFMAC* using PDB entry 1b9d as a starting model (Greenwald *et al.*, 1999). The X-ray data and refinement statistics are listed in Table 1.

Table 1
X-ray refinement and data statistics.

Values in parentheses refer to the outer 10% of the data. No σ cutoff was used for statistics or refinement.

	Compound (1)	Compound (3)
Space group	$P3_121$	$P3_121$
Wavelength (Å)	1.04	1.54
Resolution (Å)	50–1.7	50–2.3
Completeness (%)	94.4 (74.9)	92.7 (91.6)
R_{sym} (%)	3.1 (23.0)	4.3 (28.8)
No. of protein atoms	1137	1137
No. of water atoms	98	81
R factor (%)	22.1	23.7
R_{free}† (%)	26.3	29.1
R.m.s.d. bond distance	0.021	0.013
R.m.s.d. angle distance‡	0.036	0.034

† R_{free} was calculated with 5% of the data. ‡ *REFMAC* defines angles as 1,3 bond distances.

2.2. Identification and characterization of compounds

^{1}H and ^{13}C NMR spectra were recorded on Varian (Mercury 300 MHz/400 MHz and 500 MHz) spectrometers with tetramethylsilane as the internal standard. High-resolution mass spectra were obtained from the University of California Riverside mass-spectrometry facility in the FAB mode. Melting points were obtained on a Mel-Temp melting point apparatus and are reported uncorrected.

Analytical thin-layer chromatography was performed on aluminium baked silica gel 60 F_{254} plates from Alltech. All liquid-chromatography separations were performed using silica gel (230–425 mesh) from Fisher Scientific Company. Commercial chemicals were used as supplied. Yields refer to spectroscopically (^{1}H and ^{31}P NMR) homogeneous materials unless otherwise stated. Characterization of all new compounds was performed by ^{1}H, ^{13}C (^{1}H decoupled) and ^{31}P NMR as well as mass spectroscopy. ^{31}P NMR spectra were recorded using a solution of 85% H_3PO_4 in the appropriate solvent as an external standard.

Tetraphenylarsonium chloride (1) and tetraphenylphosphonium chloride (2) were purchased from Aldrich.

2.2.1. 3,4-Dihydroxyphenyltriphenylarsonium bromide (3). A mixture of 4-bromocatechol (288.3 mg, 1.52 mmol), triphenylarsine (933.4 mg, 3.04 mmol) and palladium diacetate (68.2 mg, 0.92 mmol) in acetonitrile (1 ml) was refluxed for 3 d in a sealed tube. After cooling, the solvent was evaporated. Ether and water were added and the phases separated. The water phase was washed with more ether and evaporated. The residue was crystallized from ethanol (16.9 mg; 2% yield): m.p. >573 K. ^{1}H NMR (500 MHz, D_2O): δ = 7.70 (m, 4H), 7.58 (m, 16H). ^{13}C NMR (100 MHz, DMSO): δ = 158.31, 154.45, 134.23, 133.017, 130.85, 124.27, 122.21, 120.90, 109.21, 103.78.

2.2.2. 3,4-Dihydroxyphenyltriphenylphosphonium bromide (4). A mixture of 4-bromocatechol (200 mg, 1.06 mmol), triphenylphosphine (277 mg, 1.06 mmol) and palladium diacetate (23.7 mg, 0.1 mmol) in acetonitrile (2 ml) was refluxed for 12 h in a sealed tube. A yellow precipitate was formed. Ether and water were added and the phases separated. The water phase was washed with more ether, then heated and filtered still hot. Water was evaporated (239.28 mg, 50% yield): m.p. >573 K. ^{1}H NMR (300 MHz, D_2O): δ = 7.75–7.68 (m, 3H), 7.61–7.49 (m, 12H), 6.98–6.51 (m, 3H). ^{13}C NMR (100 MHz, DMSO): δ = 154.80, 147.03, 136.4 (d, J = 3.1 Hz), 135.8 (d, J = 10.6 Hz), 131.3 (d, J = 12.9 Hz), 130.25 (d, J = 11.3 Hz), 122.48 (d, J = 12.1 Hz), 119.96 (d, J = 90.3 Hz), 118.59 (d, J = 16 Hz), 108.44 (d, J = 96.3 Hz). ^{31}P (162 MHz, DMSO): δ = 23.04 (s). HR-MS calculated for $C_{24}H_{20}O_2P$, 371.1201; measured, 371.1213.

2.2.3. 3,4-Dihydroxybenzyltriphenylphosphonium bromide (5). To a solution of 3,4-dimethoxybenzyltriphenylphosphonium bromide (1.15 g, 2.4 mmol) in dichloromethane (8 ml) was added, at 253 K, 1.0 *M* boron tribromide in dichloromethane (17.5 ml, 17.5 mmol) under an argon atmosphere. The solution was then allowed to stir overnight at room temperature. Methanol was added dropwise until no more smoke developed and then the solvents were evaporated. The solid residue was washed with dichloromethane and dried (782.2 mg, 70% yield): m.p. >573 K. ^{1}H NMR (400 MHz, DMSO): δ = 7.91–7.88 (m, 3H), 7.76–7.71 (m, 6H), 7.66–7.60 (m, 6H), 6.56 (d, J = 8 Hz, 1H), 6.39 (d, J = 2 Hz, 1H), 6.20 (d, J = 8 Hz and J = 2 Hz, 1H), 4.94 (d, J_{H-P} = 14.8 Hz). ^{13}C NMR (100 MHz, DMSO): δ = 145.22 (d, J = 3 Hz), 145.18 (d, J = 3 Hz), 134.76 (d, J = 2.3 Hz), 133.83 (d, J = 9.1 Hz), 129.84 (d, J = 12.1 Hz), 121.84 (d, J = 6.1 Hz), 118.08 (d, J = 4.5 Hz), 118.04 (d, J = 84.2 Hz), 117.45 (d, J = 8.4), 115.49 (d, J = 3 Hz), 27.88 (d, J = 44.7 Hz). ^{31}P (162 MHz, DMSO): δ = 22.31 (s). HR-MS calculated for $C_{25}H_{22}O_2P$, 385.1357; measured, 385.1356.

2.3. Enzyme assays

Assays of purified HIV-1 IN protein were carried out essentially as described previously (Bushman & Craigie, 1991; Craigie *et al.*, 1991; Carlson *et al.*, 2000). IC_{50} values are listed in Table 2. Substrate DNA sequences are listed in Fig. 1. Oligonucleotide substrates were 5′-end labeled by treatment with γ-^{32}P ATP and kinase on FB65 (integration reactions) or FB133 (disintegration reactions). Integration reactions contained 50 m*M* NaCl, 5 m*M* $MnCl_2$, 25 m*M* HEPES pH 7.5, 1 m*M* DTT, 1 μ*M* integrase and 3 ng integration substrate per reaction. Assays in the presence of Mg^{2+} did not display

Table 2
Inhibition of HIV-1 IN.

TC, terminal cleavage; ST, strand transfer; DIS, disintegration.

	Full length IC_{50} (μM)		Core 50–212 IC_{50}
Compound	TC	SC	DIS
(1)	>240	>240	>5 mM
(2)	>200	>200	>5 mM
(3)	150	150	380 μM
(4)	13.5	13.5	200 μM
(5)	>12.5	>12.5	—

inhibitory activity (data not shown). Disintegration reactions (Chow *et al.*, 1992) contained 50 mM NaCl, 5 mM $MnCl_2$, 25 mM HEPES pH 7.5, 8.6 mM β-mercaptoethanol, 10% glycerol, 0.1 mg ml^{-1} BSA, 1 μM integrase and 2 ng disintegration substrate per reaction. Reactions were incubated for 30–45 min and the products were then analyzed on DNA-sequencing-type gels. Assays of MCV topoisomerase were analyzed using gels as described (Hwang *et al.*, 1999) or a microtiter assay (unpublished data).

3. Results

3.1. Inhibitor search strategy

We first sought to identify compounds that bind the integrase surface to provide starting points for inhibitor design. Many compounds were used in soaking crystals of the integrase catalytic domain (residues 50–212), which were then analyzed by X-ray crystallography. For these studies we used two different integrase mutants. One contained the F185K mutant, which improves solubility as previously reported (Jenkins *et al.*, 1996). The second form contained the F185K and W131E mutations. This mutant crystallizes in a different space group (Goldgur *et al.*, 1999). The first small-molecule binding site was identified with compound (1) (Fig. 2) and was only resolvable in the F185K mutant, not the F185K/W131E form. Modification of compound (1) yielded inhibitors, as is discussed below, and one (compound 3) was visualized bound to the same binding site on the protein surface.

For these studies, we favored, where possible, inhibitors containing atoms with useful anomalous scattering properties. This facilitates the identification of the inhibitor-binding site, particularly in cases where the occupancy of the site is incomplete. The use of anomalous scattering was crucially instrumental to resolving the bound small molecules described below.

One strategy for choosing compounds began with the identification of a negatively charged pocket on the protein surface comprising the active site. In an effort to fill this pocket, bulky positively charged compounds were soaked with crystals and analyzed by X-ray crystallography. Cationic arsenic compounds were studied, since they are both capable of anomalous scattering and are positively charged.

OLIGO	SEQUENCE	COMMENT
FB 64	5′ACTGCTAGAGATTTTCCACACGGATCCTAGGC3′	INTEGRATION SUBSTRATE
FB 65	5′GCCTAGGATCCGTGTGGAAAATCTCTAGCAGT3′	INTEGRATION SUBSTRATE
FB133	5′GAAAGCGACCGCGCC3′	DISINTEGRATION SUBSTRATE
FB134	5′GGACGCCATAGCCCCGGCGCGGTCGCTTTC3′	DISINTEGRATION SUBSTRATE
FB244	5′GTGTGGAAAATCTCTAGCAGGGGCTATGGCGTCC3′	DISINTEGRATION SUBSTRATE
FB245	5′ACTGCTAGAGATTTTCCACAC3′	DISINTEGRATION SUBSTRATE

Figure 1
Oligonucleotides used in this study.

3.2. X-ray analysis

The binding site of compound (1) was identified by locating the As atom first in anomalous difference Fourier maps. The $(2F_o - F_c)$ and $(F^+ - F^-)$ electron-density maps were calculated for this region with the phases from the model without the compound (Fig. 2*a*). Compound 1 binds at the dimer interface near the C-terminal end of the long helix (extending from residue Ala150 to Ala169) (Fig. 3). These data did not clearly resolve the four tetrahedrally oriented phenyl rings surrounding the As atom. These results indicate that these phenyl rings are either positionally flexible or the occupancy of the compound is significantly less than one per site, as reflected by their high temperature factors in the refined structure. We have not attempted to estimate the fractional occupancy based on the peak height of the electron density.

The binding specificity of (1) appears to be a consequence in part of a strong charge–charge interaction with the carbonyl O atom of Gln168, which is pointing directly towards the arsenic center. Probably as a result, the side-chain atoms of Gln168, which are disordered in the unliganded structure (Greenwald *et al.*, 1999) and thus invisible, became ordered in the presence of (1), making a direct contact with the tip of one of the phenyl rings in (1). In addition, there are two Trp residues (Trp131 and Trp 132) in this pocket that stack with one of the phenyl rings in (1).

Compound (1) has a positively charged center which is partially balanced by a counter-ion that is bound 7 Å from the As atom in compound (1). Of the two major types of anions present in the crystallization solution, Cl^- and SO_4^{2-}, Cl^- appears to be more consistent with the observed electron density.

We also collected diffraction data on complexes of integrase formed with the other chemical derivatives of (1) described below. X-ray diffraction data from a complex of 50–212 with (3) were collected at Cu $K\alpha$ wavelength to a resolution of 2.3 Å. At this wavelength, the anomalous signal is too weak to clearly locate the As atom from anomalous difference maps. However, the standard $F_o - F_c$ maps clearly located the arsenic position of (3), placing it in the same location as the As of (1) (Fig. 2*b*). This is expected since the only difference between the two compounds is the presence in (3) of a catechol group (a phenyl ring with two hydroxyl groups). The positions of the four phenyl rings were less well defined than

for compound (1). It was therefore not possible to establish the position of the catechol.

Comparison of the two complexes, however, indicates that there are significant differences between the two. In the case of compound (3), binding caused a structural change distant from its binding site (Fig. 4). There are two covalently bound As atoms in this crystal form, resulting from covalent modification of the protein by cacodylate present in the buffer (Dyda *et al.*, 1994). Both covalently bound As atoms are attached to Cys residues that lie in small hydrophobic cavities. The As atoms are bound to Cys65 and Cys130, allowing the position of these side chains to be assessed by monitoring the As signal. The binding of (3) to 50–212 causes the $C^{\alpha}-C^{\beta}$ bond in Cys65 to rotate by about 80° around the $C^{\beta}-S^{\gamma}$ bond. This altered conformation is named the 'out' form to distinguish it from the 'in' form in the unliganded structure. This rotation causes the covalently bound As atom to change position by 3 Å. The occupancy of the cacodylate is approximately 40 and 60% in the in and out positions, respectively. Analysis of ten other data sets (unsuccessful attempts to soak in other inhibitors) showed C65 in the in position only (data not shown).

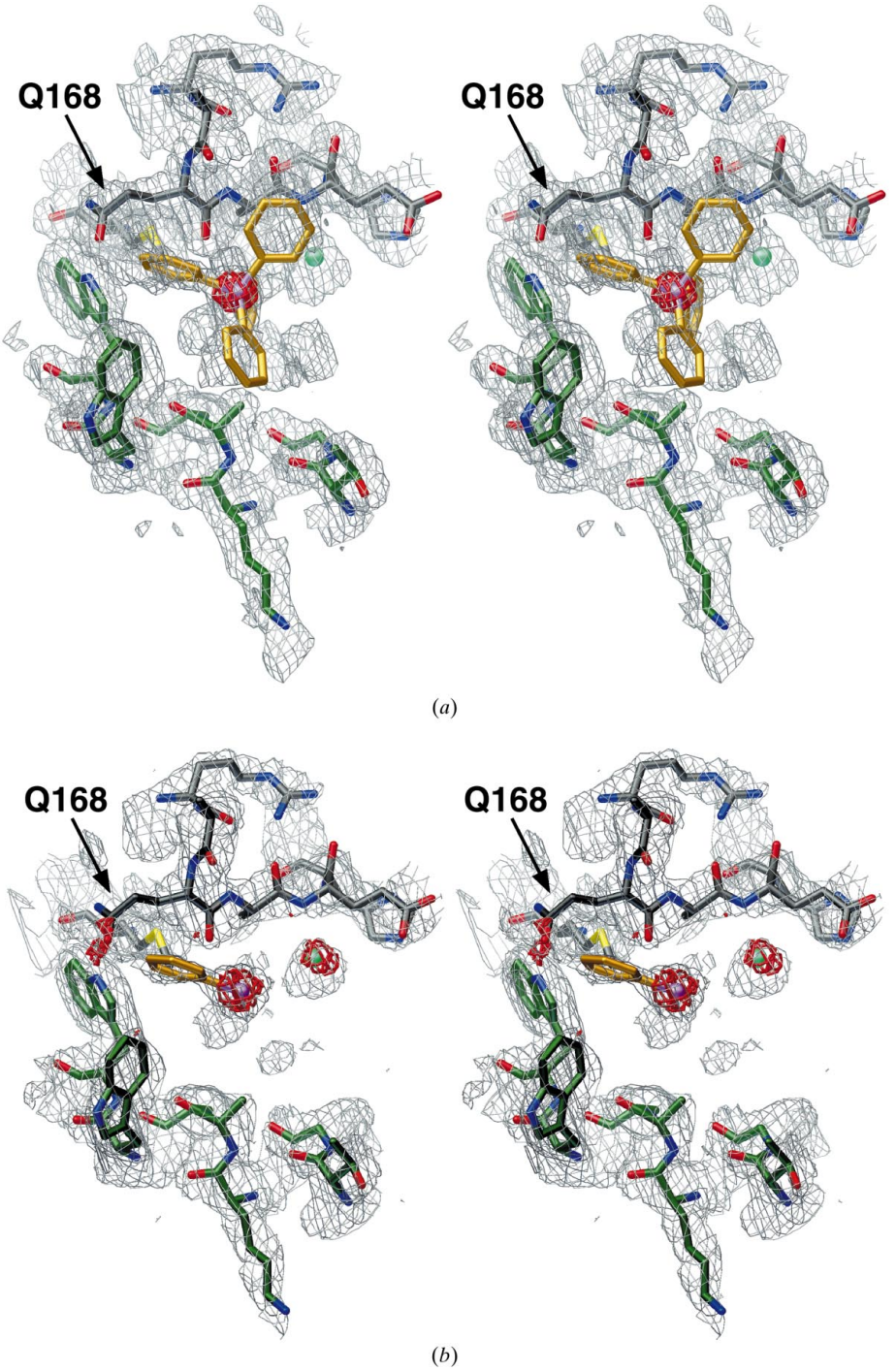


Figure 2
(*a*) Stereoview of the $2F_o - F_c$ electron density contoured at 0.6σ in gray and the $F^+ - F^-$ electron density at 8σ in red overlaid with the final model in the region of compound (1). The maps were calculated with the phases from the protein model without the compound. The bonds in the protein atoms are colored yellow in molecule 1, and green and gray in molecule 2 in the 50–212 dimer. Side chains of Gln168 are marked by arrows. (*b*) The same view as (*a*) with $2F_o - F_c$ (contoured at 1σ in gray) and $F_o - F_c$ (contoured at 4σ in red) electron-density maps showing the binding of compound (3). Only one phenyl ring is shown, as the position of the others was uncertain.

3.3. Inhibitory activity against HIV-1 integrase *in vitro*

More than 20 small molecules related to (1) were tested for their ability to inhibit the terminal cleavage (TC) and strand-transfer (ST) activities of integrase (Table 2, Fig. 5 and data not shown). Tests used model oligonucleotide substrates mimicking one end of the unintegrated viral DNA and target DNA. Products of integration reactions were separated by electrophoresis and quantitated by PhosphorImager scanner.

The IC_{50} values for the compounds in this series are summarized in Table 2. Compound (1) did not display significant inhibitory activity against integrase. Compound (2), which substitutes the As atom of (1) with phosphorus, also was not inhibitory. Compound (3) differs from (1) by the presence of a catechol ring in place of a phenyl ring. The catechol group has often been often found in compounds identified as integrase inhibitors. Compound (3) did show modest though clear activity against integrase.

Substitution of the arsenic center in compound (3) with phosphorus yielded compound (4), the most active inhibitor in the series. Compound (4) inhibited in the low micromolar range, comparing favorably to other integrase inhibitors in this class. An analog with methoxy groups in place of hydroxyl groups in (4) was completely inactive, indicating the importance of the catechol unit (data not shown). Furthermore, the precise position of the catechol group is important. Addition of a methylene unit between the catechol ring and the P atom of compound (4) yielded an inactive derivative (compound 5). This result shows that small changes in the inhibitor structure can produce large variations in the biological properties. As a control, the active compounds were tested for activity against the MCV topoisomerase enzyme (Hwang *et al.*, 1998) and found to be inactive, demonstrating specificity for integrase.

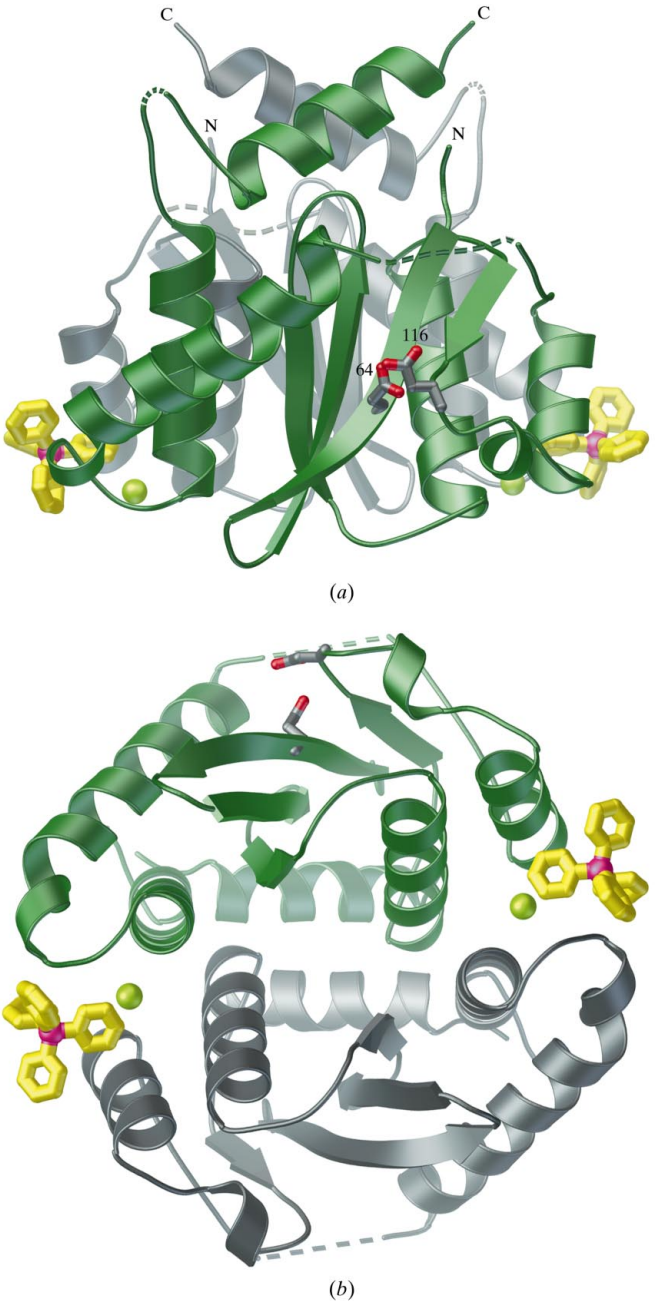


Figure 3
(*a*) Ribbon diagram of the 50–212 dimer showing the location of the binding site for compounds (1) and (3). The twofold axis of the dimer, indicated by the arrow, is nearly vertical and slightly into the page. The acidic side chains Asp64 and Asp116 are displayed on the green model as sticks. The disordered regions of the model that could not be visualized are drawn as dashed lines. Compound (1) is in yellow and its counter-ion is shown as a green ball. (*b*) The same as (*a*) as viewed from the bottom looking down the twofold axis.

3.4. Inhibition of isolated catalytic domain

Inhibition of 50–212 was also tested, since this is the form of the enzyme studied by crystallography. For tests of the inhibition of the catalytic domain, the disintegration assay was used (Chow *et al.*, 1992). The substrate for the disintegration assay contained branched DNA substrates mimicking the product of integration. Disintegration by the catalytic domain cleaved off the viral DNA end and resealed the target DNA. The results of disintegration assays testing inhibition of 50–212 are shown in Fig. 5(*b*). Substrates were end labeled on the target strand, so disintegration produces a longer DNA product (arrow on figure). Compounds (1) and (2) did not display detectable inhibitory activity. Compounds (3) and (4) did inhibit, with compound (4) displaying the lower IC_{50}. Thus the rank order of inhibitors was the same with full-length integrase and 50–212, though the IC_{50} values were higher with 50–212. This is probably because the disintegration reaction catalyzed by 50–212 has been found to be less sensitive to inhibitors than terminal cleavage and strand transfer in several previous studies (Reddy *et al.*, 1999; Farnet *et al.*, 1998).

In our X-ray studies of compounds (1) and (3), the 50–212 F185K form modified by arsenic (from the cacodylate buffer) was used because only this crystal form permitted the bound inhibitor to be identified. Previous studies have suggested that the presence of the covalently bound As atoms at Cys65 and

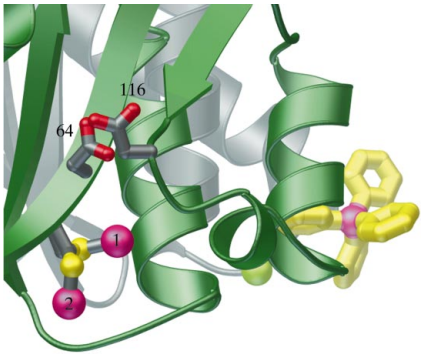


Figure 4
Changes in the catalytic domain structure distant from the compound (3) binding site. A close-up view of the two conformations for the As-modified Cys65 in the presence of compound (3). The ball attached to the sulfur represents the As atom from the cacodylate buffer. Residues 64 and 116 are two of the acidic active-site residues. Position 1 is the original position ('in') and position 2 is the predominant 'out' conformation adopted in the structure of 50–212 with compound (3) bound. Compound (3) is shown without the catechol hydroxyl groups since the locations of these groups are unknown.

Cys130 distort the positions of several enzyme side chains, raising the question of whether the structure observed was biologically relevant (Goldgur *et al.*, 1999). We therefore examined whether the cacodylate-modified form of 50–212 retained disintegration activity (Fig. 5*d*). The 50–212 F185K enzyme was pre-incubated with 100 m*M* cacodylate and 10 m*M* DTT for 30 min and the disintegration activity was assayed. Fig. 5(*d*) shows that the disintegration activity remains intact despite cacodylate treatment.

4. Discussion

4.1. Small molecules that bind the 50–212 dimer interface

Here, we describe the identification of compounds that bind at the dimer interface of the HIV-1 integrase catalytic domain. The binding pocket is ∼5 Å deep and displays an overall negative charge.

Although compound (1) occupied the dimer interface site, it was not inhibitory. In an effort to generate an inhibitor, we converted one of the phenyl rings to a catechol, a group often found in integrase inhibitors. This yielded an inhibitor (3), which displayed modest potency (IC_{50} = 150 μ*M*). X-ray diffraction analysis revealed that compound (3) bound to the same dimer interface site as compound (1). Further derivatives were then synthesized with the goal of increasing potency. The most significant improvement came with substituting the As atom with phosphorus (4), which yielded a roughly tenfold decrease in IC_{50}. Given that the active site of integrase is rather flat, this dimer interface pocket represents a potential new target for inhibitory ligands.

Why does compound (3) inhibit integrase, whereas compound (1), apparently bound to the same site, does not? Our crystallographic and enzymatic analysis cannot rule out the possibility that the binding of the compounds at the dimer interface is a crystallographic artifact and that binding to other unidentified sites on the protein or on the protein–DNA complexes could be directly responsible, although at present there is no evidence supporting this possibility. The other possible mechanism involves the changes in surface side chains seen with 50–212 bound to (3). The altered positions of the side chains near the active site in the complex of 50–212 with compound (3) but not that of 50–212 with compound (1) might interfere with enzyme activity. In support of this, Cys65 is adjacent to the catalytic residue Asp64.

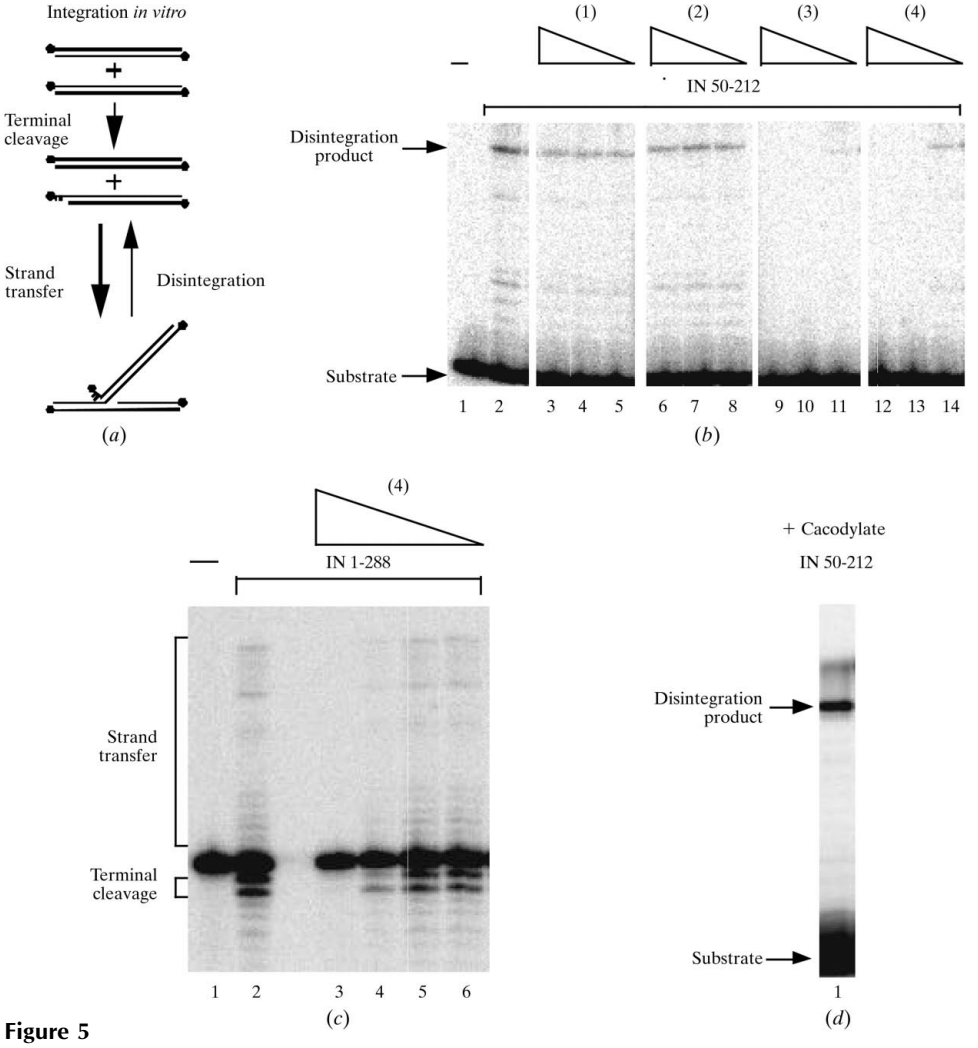


Figure 5
Inhibition of HIV-1 integrase *in vitro* by arsenic-containing compounds. Diagram of activities of purified integrase *in vitro*. The lines represent an oligonucleotide DNA matching one end of the unintegrated HIV cDNA. DNA 5′-ends are indicated by balls. In the presence of purified integrase protein, two nucleotides are removed from the 3′-end matching the cDNA end (terminal cleavage) and cleaved molecules are then joined to breaks made in a target DNA (strand transfer). (*b*) Inhibitory activities of compounds (1)–(4) against the HIV-1 integrase catalytic domain (50–212). The disintegration assay was used to measure the activity of the catalytic domain. The gel mobilities of the substrates and products are as indicated. Compounds tested are listed above the gel. Lane 1 contained substrate only; all other lanes contained 50–212. Lanes 2, no added inhibitor; lane 3, 2.5 m*M* (1), lane 4, 1.25 m*M* (1); lane 5, 0.6 m*M* (1); lane 6, 2.5 m*M* (2); lane 7, 1.25 m*M* (2); lane 8, 0.6 m*M* (2); lane 9: 2.5 m*M* (3); lane 10, 1.25 m*M* (3); lane 11, 0.6 m*M* (3); lane 12, 800 u*M* (4); lane 13, 400 u*M* (4); lane 14, 200 u*M* (4). (*c*) Autoradiogram illustrating inhibition of full-length integrase by compound (4). Lane 1, substrate only. Lane 2, substrate incubated with integrase. Lanes 3–6 contained standard reactions with varying amounts of compound (4). Lane 3, 200 μm; lane 4, 50 μm; lane 5, 17.5 μm; lane 6, 3 μm. The mobilities of the terminal cleavage and strand-transfer products are indicated by the brackets. (*d*) Integrase 50–212 modified by incubation with cacodylate and DTT retains disintegration activity. Integrase 50–212 was incubated with 100 m*M* cacodylate and 10 m*M* DTT for 30 min at room temperature, then diluted 1:10 into a standard disintegration reaction (final concentration 1 μ*M* integrase 50–212).

Other mechanisms are also conceivable. For example, the catechol moiety might bind a metal atom, a known property of this group, thereby forming a larger 'knob' on the protein surface that could sterically interfere with activity. A previous study led to the suggestion of such a mechanism for catechol-containing integrase inhibitors based on other data (Farnet *et al.*, 1998). Two observations suggest the compounds (3) and (4) may inhibit assembly of integration complexes. First, the catechol derivatives did not inhibit reactions with pre-assembled preintegration complexes isolated from infected cells (data not shown). Second, we find that the IC_{50} is increased twofold to fourfold if IN is preassembled with DNA substrates prior to adding (3) or (4) (data not shown).

4.2. Exploiting anomalous scattering in inhibitor development

Compounds that contained atoms that exhibit an anomalous scattering signal facilitated the localization of the binding site on 50–212. The results emphasize the usefulness of anomalous scatterers in screening for bound ligands (see Hendrickson, 1991 for detailed discussion). This is particularly useful in early stages of study in which weak signals arising from partial occupancy are often encountered. In our experiments, the presence of anomalous signal was crucial to locate the As center of the compounds. The inclusion of an anomalous scatterers such as As, Se or Br can help to locate the compound with less ambiguity in initial crystallographic screening of inhibitors. We are attempting to exploit this approach systematically by assembling a library of compounds containing anomalous scattering atoms to facilitate the identification of small-molecule binding sites in the course of crystallographic analyses.

4.3. Three binding sites for integrase inhibitors

This study reports the third inhibitor-binding site on the surface of integrase proteins identified to date: a binding site at the catalytic domain dimer interface. Previously, Lubkowski *et al.* (1998) have reported a structure of the avian sarcoma virus integrase catalytic domain complexed with an inhibitor bound to a site near the catalytic pocket. Goldgur *et al.* (1999) reported another structure of the HIV catalytic domain containing a small molecule bound to a part of the active site. Some caution is warranted in interpreting data from all three complexes, since the inhibitors in each case make contacts to multiple integrase monomers in the crystal lattice and these potentially non-biological contacts may influence the conformation observed (Striffer *et al.*, 2000). However, these data taken together suggest that it may be possible to develop inhibitors of integrase that bind to multiple sites. Potentially, the HIV virus would need independent mutations in the integrase coding region to evade inhibition by molecules binding to each site. This raises the exciting hope that it may be possible to develop multiple combination therapy of integrase to add to the antiviral armamentarium.

We thank people in the groups of A. McCammon, J. Briggs and J. Faulkner for numerouse discussions and insight. This work was supported by a NIH AIDS-related structural biology program grant (GM56553) and AI34786 (FDB). FDB is a scholar of the Leukemia Society of America. JG is supported by a Markey fellowship and a Howard Hughes Medical Institute predoctoral fellowship.

References

Bujacz, G., Jaskolski, M., Alexandratos, J., Wlodawer, A., Merkel, G., Katz, R. A. & Skalka, A. M. (1995). *J. Mol. Biol.* **253**, 333–346.
Bushman, F. D. & Craigie, R. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 1339–1343.
Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P. & Craigie, R. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.
Bushman, F. D., Fujiwara, T. & Craigie, R. (1990). *Science*, **249**, 1555–1558.
Bushman, F. D., Landau, N. R. & Emini, E. A. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 11041–11042.
Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). *Nature Struct. Biol.* **4**, 567–577.
Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., Briggs, J. M. & McCammon, J. A. (2000). *J. Med. Chem.* **43**, 2100–2114.
Chen, J. C.-H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D. & Stroud, R. M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8233–8238.
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. (2000). *J. Mol. Biol.* **296**, 521–533.
Chow, S. A., Vincent, K. A., Ellison, V. & Brown, P. O. (1992). *Science*, **255**, 723–726.
Coffin, J. M., Hughes, S. H. & Varmus, H. E. (1997). *Retroviruses.* Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
Craigie, R., Fujiwara, T. & Bushman, F. (1990). *Cell*, **62**, 829–837.
Craigie, R., Mizuuchi, K., Bushman, F. D. & Engelman, A. (1991). *Nucleic Acids Res.* **19**, 2729–2734.
Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). *Science*, **266**, 1981–1986.
Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Vos, A., Doreleijers, J. F., Hard, K., Tullius, T., Plasterk, R. H. A., Kaptein, R. & Boelens, R. (1997). *Curr. Biol.* **1**, 739–746.
Engelman, A., Bushman, F. D. & Craigie, R. (1993). *EMBO J.* **12**, 3269–3275.
Farnet, C. M., Wang, B., Hansen, M., Lipford, J. R., Zalkow, L., Robinson, W. R. J., Siegel, J. & Bushman, F. D. (1998). *Antimicrob. Agents Chemother.* **42**, 2245–2253.
Gent, D. C. van, Vink, C., Oude Groeneger, A. A. M. & Plasterk, R. H. A. (1993). *EMBO J.* **12**, 3261–3267.
Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. & Davies, D. R. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13040–13043.
Greenwald, J., Le, V., Butler, S. L., Bushman, F. D. & Choe, S. (1999). *Biochemistry*, **38**, 8892–8898.
Hansen, M. S. T., Carteau, S., Hoffmann, C., Li, L. & Bushman, F. (1998). *Genetic Engineering, Principles and Methods*, edited by J. K. Setlow, Vol. 20, pp. 41–62. New York/London: Plenum Press.
Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C. & Miller, M. D. (2000). *Science*, **287**, 646–650.
Hendrickson, W. A. (1991). *Science*, **254**, 51–254.
Hwang, Y., Rowley, D., Rhodes, D., Gertsch, J., Fenical, W. & Bushman, F. D. (1999). *Mol. Pharmacol.* **55**, 1049–1053.
Hwang, Y., Wang, B. & Bushman, F. D. (1998). *J. Virol.* **72**, 3401–3406.

Jenkins, T. M., Engelman, A., Ghirlando, R. & Craigie, R. (1996). *J. Biol. Chem.* **271**, 7712–7718.
Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6057–6061.
Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990). *Cell*, **63**, 87–95.
Katzman, M., Katz, R. A., Skalka, A. M. & Leis, J. (1989). *J. Virol.* **63**, 5319–5327.
King, P. J., & Robinson, W. E. J. (1998). *J. Virol.* **72**, 8420–8424.
Lubkowski, J., Yang, F., Alexandratos, J., Wlodawer, A., Zhao, H., Burke, T. R., Neamati, N., Pommier, Y., Merkel, G. & Skalka, A. M. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 4831–4836.
Miller, M. D., Farnet, C. M. & Bushman, F. D. (1997). *J. Virol.* **71**, 5382–5390.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Patel, P. H. & Preston, B. D. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 549–553.
Pommier, Y., Pilon, A. A., Bajaj, K., Mazumder, A. & Neamati, N. (1997). *Antiviral Chem. Chemother.* **8**, 463–483.
Reddy, M. V. R., Rao, M. R., Rhodes, D., Hansen, M., Rubins, K., Bushman, F., Venkateswarlu, Y. & Faulkner, D. J. (1999). *J. Med. Chem.* **42**, 1901–1907.
Roth, M. J., Schwartzberg, P. L. & Goff, S. P. (1989). *Cell*, **58**, 47–54.
Sherman, P. A. & Fyfe, J. A. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.
Striffer, C. A., Hi, H. & McCammon, J. A. (2000). *J. Am. Chem. Soc.* **122**, 6136–6137.
Vink, C., Oude Groeneger, A. M. & Plasterk, R. H. A. (1993). *Nucleic Acids Res.* **21**, 1419–1425.
Yang, W. & Steitz, T. A. (1995). *Curr. Biol.* **3**, 131–134.
Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. (2000). *J. Mol. Biol.* **296**, 535–548.