



HAL
open science

Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA.

Julia Agapkina, Maksim Smolov, Sophie Barbe, Evgenii Zubin, Timofei Zatsépin, Eric Deprez, Marc Le Bret, Jean-François Mouscadet, Marina Gottikh

► **To cite this version:**

Julia Agapkina, Maksim Smolov, Sophie Barbe, Evgenii Zubin, Timofei Zatsépin, et al.. Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA.. *Journal of Biological Chemistry*, 2006, 281 (17), pp.11530-40. 10.1074/jbc.M512271200 . hal-00283458

HAL Id: hal-00283458

<https://hal.science/hal-00283458>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Probing of HIV-1 Integrase/DNA Interactions Using Novel Analogs of Viral DNA^{*[S]}

Received for publication, November 15, 2005, and in revised form, February 21, 2006 Published, JBC Papers in Press, February 24, 2006, DOI 10.1074/jbc.M512271200

Julia Agapkina[‡], Maksim Smolov[‡], Sophie Barbe[§], Evgenii Zubin[‡], Timofei Zatsépin[‡], Eric Deprez[§], Marc Le Bret[§], Jean-François Mouscadet^{§1}, and Marina Gottikh[‡]

From the [‡]Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119992 Moscow, Russia and [§]LBPA, UMR 8113 CNRS, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, France

The specific activity of the human immunodeficiency virus, type 1 (HIV-1), integrase on the viral long terminal repeat requires the binding of the enzyme to certain sequences located in the U3 and U5 regions at the ends of viral DNA, but the determinants of this specific DNA-protein recognition are not yet completely understood. We synthesized DNA duplexes mimicking the U5 region and containing either 2'-modified nucleosides or 1,3-propanediol insertions and studied their interactions with HIV-1 integrase, using Mn²⁺ or Mg²⁺ ions as integrase cofactors. These DNA modifications had no strong effect on integrase binding to the substrate analogs but significantly affected 3'-end processing rate. The effects of nucleoside modifications at positions 5, 6, and especially 3 strongly depended on the cationic cofactor used. These effects were much more pronounced in the presence of Mg²⁺ than in the presence of Mn²⁺. Modifications of base pairs 7–9 affected 3'-end processing equally in the presence of both ions. Adenine from the 3rd bp is thought to form at least two hydrogen bonds with integrase that are crucial for specific DNA recognition. The complementary base, thymine, is not important for integrase activity. For other positions, our results suggest that integrase recognizes a fine structure of the sugar-phosphate backbone rather than heterocyclic bases. Integrase interactions with the unprocessed strand at positions 5–8 are more important than interactions with the processed strand for specific substrate recognition. Based on our results, we suggest a model for integrase interaction with the U5 substrate.

Following reverse transcription, a DNA copy of the human immunodeficiency virus, type 1 (HIV-1),² RNA is integrated into the genome of infected cells. Integration is a prerequisite for viral replication and is catalyzed by the viral enzyme integrase (IN). IN binds to sequences located at the end of U3 and U5 parts of long terminal repeats (LTRs) of viral DNA and catalyzes the trimming, or 3'-end processing, of the terminal dinucleotide from the 3'-ends of both strands of the DNA. IN then mediates a strand transfer reaction that inserts the viral DNA into the host DNA. During this reaction, IN must bind simultaneously to viral and target DNA. However, IN interacts with these two DNA mol-

ecules in different ways as follows: binding to host DNA does not depend directly on host DNA sequence, whereas interaction with the viral DNA is a sequence-specific process. Nevertheless, the U5 and U3 sequences recognized by IN are not exactly identical.

Strand transfer and 3'-end processing reactions may be carried out *in vitro*, using recombinant HIV IN, DNA duplexes mimicking U3 or U5 sequences of LTRs, and divalent metal ions, such as Mg²⁺ or Mn²⁺. However, the Mn²⁺- and Mg²⁺-dependent activities of IN are not equivalent, with lower specificity reported for Mn²⁺-dependent IN (1, 2). Moreover, the inhibition of HIV-1 IN by compounds such as β -diketo acids, which interact with the active site of HIV-1 IN, is also metal-dependent (3). The metal ion is coordinated within the IN catalytic core by two Asp residues (Asp-64 and Asp-116) and water molecules (4–6). These Asp residues are part of the D,D(35)E motif, a triad of invariant acidic amino acids found in all retroviral integrases (7) and ultimately required for catalytic activity (2, 8–10). Binding of the metal cofactor by the amino acids of the D,D(35)E motif causes both local and long range conformational changes and the activation of IN (5, 11). The IN catalytic core binds both the metal cofactor and the substrate. These conformational changes therefore influence the specificity of IN interactions with substrate DNA. Differences in the coordination properties of Mg²⁺ and Mn²⁺ ions may result in different conformational changes in the IN catalytic core. This may account for differences in the specificity of HIV-1 IN interactions with substrate DNA (1, 2), inhibitors (3, 12), and the nucleophiles used to carry out 3'-end processing (13). As IN is more sensitive to the structure of molecules interacting with its catalytic core in the presence of magnesium ions, it has been suggested that the catalytic core of this molecule is more rigid in the presence of magnesium ions, as reported for other magnesium-dependent enzymes (1).

IN has three domains: the N-terminal, catalytic, and C-terminal domains. The structure of each domain, and of two-domain fragments, has been determined by x-ray crystallography or NMR spectroscopy (4, 6, 14–16). However, neither the structure of the entire enzyme nor that of the IN-viral DNA complex has yet been determined. Cross-linking experiments have been used to study contacts between IN and viral DNA, and it has been found that the catalytic core (residues 50–212) and C-terminal domain (residues 213–288) are the principal areas involved in DNA binding. It has been shown that the Lys-159 residue of HIV-1 integrase may interact with the N-7 position of the third adenosine of the processed viral DNA strand (17) and that Glu-152 may also interact with this adenosine (18). Gln-148 and Tyr-143 have been shown to bind to the terminal nucleotides of the processed strands of viral DNA, and the thymidine at position 7 interacts with amino acids from the C-terminal domain (1). Glu-246 is also thought to be located near the C-2' atom of the 2'-deoxyadenosine at position 7 in the unprocessed strand (19). Nevertheless, the protein/DNA interactions through which IN forms a stable and catalytically active complex with viral DNA are not yet completely understood.

* This work was supported by the TRIOH European Project FP6 Grant 503480, the Russian Foundation for Basic Research Grants 04-04-22000 and 05-04-48743, the CNRS (PICS27), and the French National Agency for Research on AIDS (ANRS). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at <http://www.jbc.org>) contains Tables S1 and S2 and Fig. S1.

¹ To whom correspondence should be addressed: LBPA-CNRS UMR8113, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan Cedex, France. Tel.: 33-1-47-40-76-75; Fax: 33-1-47-40-76-84; E-mail: mouscadet@lbpa.ens-cachan.fr.

² The abbreviations used are: HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; IN, integrase; DTT, dithiothreitol; DAP, 2,6-diaminopurine.

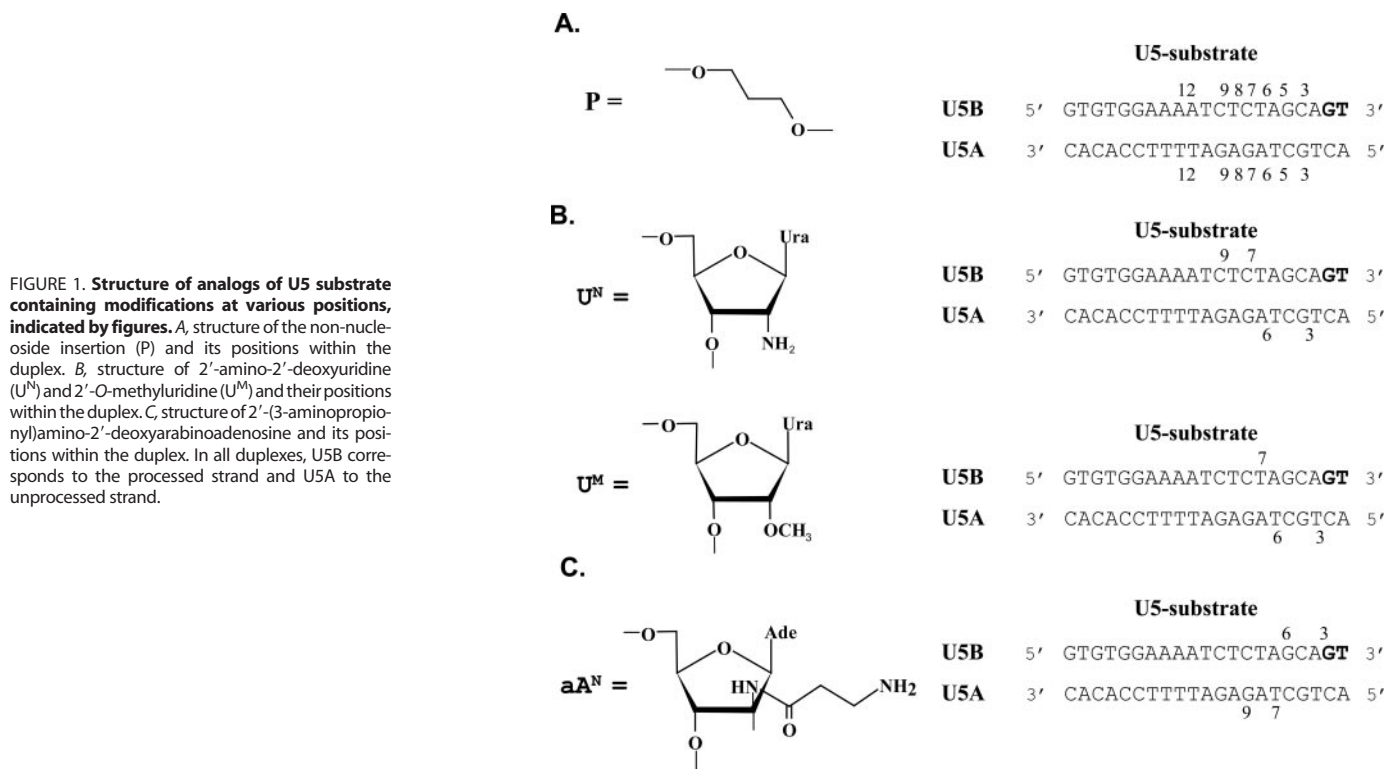


FIGURE 1. Structure of analogs of U5 substrate containing modifications at various positions, indicated by figures. *A*, structure of the non-nucleoside insertion (*P*) and its positions within the duplex. *B*, structure of 2'-amino-2'-deoxyuridine (*U^N*) and 2'-*O*-methyluridine (*U^M*) and their positions within the duplex. *C*, structure of 2'-(3-aminopropionyl)amino-2'-deoxyarabinoadenosine and its positions within the duplex. In all duplexes, U5B corresponds to the processed strand and U5A to the unprocessed strand.

The third and fourth nucleotides of the processed strand of the viral DNA (the so-called conserved 5'-CA-3') are important for 3'-end processing and strand transfer activities of IN (1, 20, 21). It has been shown that the region from nucleotides 4 to 9 of the viral DNA is important for efficient 3'-end processing (22–25). Modification of the nucleosides at positions 5 and 6 in the processed strand significantly decreases 3'-end processing efficiency (1, 20, 25), and it has been suggested that IN comes into contact with the heterocyclic bases of these nucleosides (26). However, the mechanism of IN interactions with the processed and unprocessed strands of the HIV DNA remains to be elucidated.

We therefore investigated the effect of the double-helix structure of viral DNA on HIV-1 integrase activity in the 3'-end processing reaction, as a means of identifying the structural features of the viral DNA determining the sequence specificity of its recognition and processing by IN. We synthesized a set of modified DNA duplexes, with sequences mimicking the U5 sequence of LTR of the viral DNA (U5 substrate). Nucleosides at various positions in the processed strand, the unprocessed strand, or both were replaced by a non-nucleoside insertion, 1,3-propanediol, or nucleosides containing modified sugar residues, 2'-amino-nucleosides and 2'-*O*-methylnucleosides (Fig. 1). Because the conserved CA sequence adenosine is considered crucial for 3'-end processing and strand transfer activities, we also prepared a number of U5 substrate analogs in which the A/T pair was replaced by unpaired nucleotides in the 3rd position and a 2,6-diaminopurine replaced the conserved adenine.

We have shown that IN probably recognizes the viral DNA by forming at least two hydrogen bonds with the adenine at position 3. The complementary base, thymine, does not interact with IN. We suggest that IN recognizes the fine structure of the sugar-phosphate backbone rather than heterocyclic bases at other positions. We also found that interactions with the unprocessed strand were more important for specific substrate recognition by IN than interactions with the processed strand. We showed that local destabilization of the A/T pair at position 3 was required for 3'-end processing, allowing such processing to occur even at internal positions of the U5 substrate within DNA. Finally, we

studied IN/DNA interactions in the presence of Mn^{2+} and Mg^{2+} ions, investigating the differences in substrate recognition specificity in the presence of these two cofactors. We found that modification of the nucleosides in the 6th, 5th, and especially the 3rd positions influenced 3'-end processing more strongly in the presence of Mg^{2+} than in the presence of Mn^{2+} ions, whereas modification of base pairs 7–9 affected 3'-end processing equally in the presence of both ions. We suggest that positions 3–6 within the substrate DNA interact with the IN catalytic core that binds the metal cofactor.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis—Oligonucleotides were synthesized using an Applied Biosystems 380B DNA synthesizer and the standard β -cyanoethyl phosphoramidite procedure. The resulting oligonucleotides were purified by electrophoresis in a 20% denaturing polyacrylamide gel. Oligonucleotides containing 2'-(3-aminopropionyl)amino-2'-deoxyarabinoadenosine (Fig. 1C) and 2'-amino-2'-deoxyuridine (Fig. 1B) were synthesized as described previously (27, 28). Oligonucleotides containing 1,3-propanediol insertions (Fig. 1A) were synthesized using the commercially available Spacer C3 (Glen Research). An oligonucleotide containing 2,6-diaminopurine was synthesized using commercially available 2-amino-dA-CE phosphoramidite (Glen Research). Structures of all modified oligonucleotides were confirmed by matrix-assisted laser desorption/ionization time of flight. Hybridization of oligonucleotides U5B and U5A (U5B being the processed strand) gave the wild-type U5 integrase substrate.

The fluorescein-labeled oligonucleotides U5B-fl (5'-fl-GTGTGGAAAATCTCTAGCAGT-3') and U5A-fl (3'-fl-CACACCTTTTGTAGATCGTCA-5') used for IN/DNA binding studies were synthesized as follows: oligonucleotide U5B, with an amino group at its 5'-end, and oligonucleotide U5A, with an amino group at its 3'-end, were synthesized with 5'-amino-modifier C6 reagent and 3'-amino-modifier C3 controlled pore glass (Glen Research), respectively. The resulting oligonucleotides were dissolved in a mixture of 60 μ l of H_2O , 20 μ l of 1 M

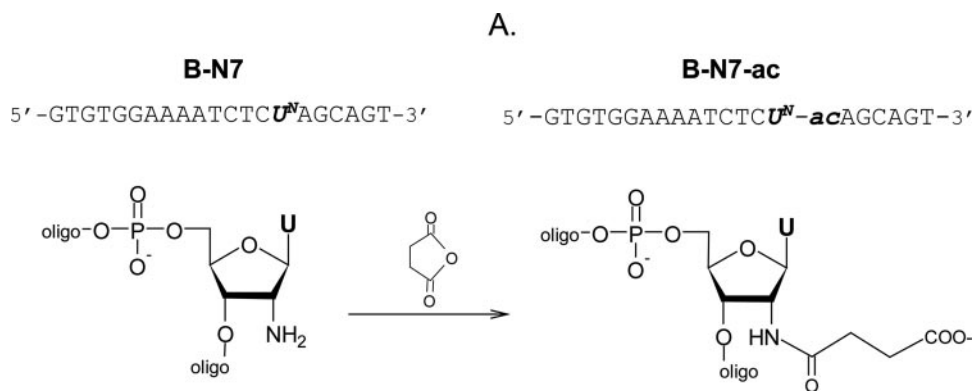
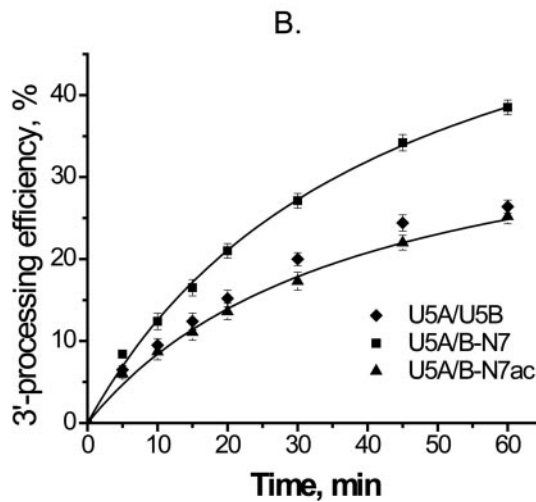


FIGURE 2. Stimulation of 3'-end processing because of the presence of the NH₂ group in the sugar 2'-position of the 2'-aminouridine at position 7. *A*, acylation scheme for the amino group of oligonucleotide B-N7. *B*, kinetic curves for the 3'-end processing of the U5 substrate, U5A/B-N7, and U5A-B-N7-ac duplexes in the presence of Mg²⁺ ions.



carbonate buffer (pH 11), and 80 μ l of dimethylformamide. We then added 1 mg of fluorescein isothiocyanate to each solution. The mixtures were incubated in the dark at room temperature for 4–12 h, and 1 ml of 2% LiClO₄ in acetone was added to precipitate the oligonucleotides. The reaction mixtures were analyzed by reversed-phase high performance liquid chromatography in ion pair mode, and the labeling efficiency was found to be greater than 90%. The conjugates were purified by electrophoresis in a 20% polyacrylamide denaturing gel, and the presence of fluorescein was confirmed by absorption spectroscopy.

Synthesis of the Oligonucleotide Containing Acylated 2'-Amino Group—A solution of 10 nmol of 2'-aminouridine containing oligonucleotide (B-N7) in 100 μ l of 1 M sodium borate buffer (pH 8.3) was supplemented with 100 μ l of a 0.1 M solution of succinic anhydride in acetonitrile, and the reaction mixture was incubated at 37 °C for 3 h (Fig. 2A). Acetonitrile was then evaporated, and the acylated oligonucleotide (B-N7-ac) was purified by gel filtration using NAP-10 columns (Amersham Biosciences). The efficiency of the amino group acylation determined by reversed-phase high performance liquid chromatography analysis in ion pair mode was about 95%. The acylated oligonucleotide (B-N7-ac) was further used for 3'-processing experiments without additional purification.

Integrase—Detergent-free recombinant integrase protein was produced in *Escherichia coli* and purified as described previously (29).

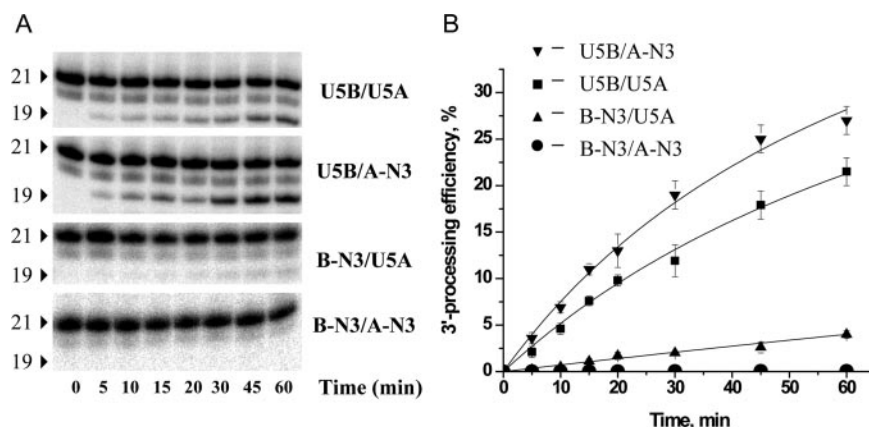
Thermal Denaturation Studies—Oligonucleotide duplexes (1 μ M in 20 mM Tris-HCl buffer (pH 7.0), 7.5 mM MgCl₂) were heated at 90 °C for 3 min and slowly cooled to 10 °C. Melting studies were carried out in Teflon-stoppered 1-cm path length quartz cells, using a Hitachi 100–20 spectrophotometer equipped with a thermal programmer. We monitored absorbance at 260 nm as the temperature was increased at a rate of 0.5 °C/min. Melting temperatures (T_m) were determined by fitting the

first derivative of absorbance with respect to $1/T$. Uncertainty in T_m was estimated to a precision of ± 0.5 °C by repeating experiments.

Preparation of 5'-Radiolabeled DNA Duplexes—For binding and processing assays, 10 pmol of the oligonucleotide U5B, or one of its modified analogs, was labeled using T4 polynucleotide kinase and 50 μ Ci of [γ -³²P]ATP (3000 Ci/mmol). T4 polynucleotide kinase was then inactivated by adding EDTA and heating at 65 °C for 5 min. The corresponding complementary unlabeled oligonucleotide, U5A, or its modified analog (10 pmol) was then added to each of the labeled oligonucleotides. The mixtures were heated at 90 °C for 3 min, and the DNA duplexes were allowed to anneal by slowly cooling to room temperature. DNA duplexes were purified using Micro Bio-Spin columns P-6 (Bio-Rad).

3'-End Processing Assay—The 3'-end processing reaction was carried out in a buffer containing 20 mM HEPES (pH 7.2), 1 mM DTT, and 7.5 mM MnCl₂ or MgCl₂ using the IN substrate U5B/U5A or an analog containing modifications in one or both strands. ³²P-Labeled DNA substrates (1.25 nM) were incubated in 20 μ l of the buffer with 100 nM integrase at 37 °C. Aliquots were taken after 5, 10, 20, 30, 60, and 120 min. The reaction was stopped by adding 80 μ l of a stop solution (7 mM EDTA, 0.3 M sodium acetate, 10 mM Tris-HCl (pH 8)). IN was extracted in phenol/chloroform, and DNA fragments were precipitated with ethanol. The products were suspended in loading buffer (80% formamide, 0.05% bromophenol blue, 0.05% xylene cyanol) and separated on 20% polyacrylamide denaturing gels. Gels were analyzed on a Storm 840TM PhosphorImager (Amersham Biosciences) and quantified with ImageQuantTM 4.1 software (Fig. 3). We fitted a nonlinear regression curve to the experimental data, and initial rates were calculated as the tangent to the kinetic curves for $t = 0$. Rate data for modified substrates were

FIGURE 3. Effect on 3'-end processing of replacing nucleosides at position 3 by 2'-aminonucleosides in the presence of Mg²⁺ ions. *A*, electrophoresis analysis of 3'-end processing products in the case of the wild-type substrate (U5B/U5A) and its analogs containing the U5B strand and U^N in the U5A strand (U5B/A-N3), aA^N in the U5B strand and the U5A strand (B-N3/U5A), and 2'-aminonucleosides in both strands (B-N3/A-N3). *B*, kinetics of 3'-end processing product accumulation for wild-type U5 substrate and its analogs containing 2'-aminonucleosides in one or both strands.



normalized to the initial rate of 3'-end processing of the natural U5 substrate. All 3'-end processing experiments were carried out at least three times; the standard deviation was between 5 and 10%.

DNA Binding Assay—³²P-Labeled duplexes (5 nM) were incubated with IN (10, 25, 50, 100, and 300 nM) in 10 μ l of a buffer containing 20 mM HEPES (pH 7.2), 1 mM DTT, 7.5 mM MgCl₂, and 5% (v/v) glycerol for 25 min at room temperature. A constant level of 20 mM NaCl was maintained in all reaction mixtures. The samples were then loaded on a 4% nondenaturing acrylamide gel containing 0.2% glycerol. Electrophoresis was carried out in a running buffer containing 50 mM Tris borate (pH 7.5) and 1 mM EDTA for 2 h at 4 °C (5.6 V/cm). A Storm 840 PhosphorImager (Amersham Biosciences) was used for analysis and quantification.

Steady-state Fluorescence Anisotropy Measurements—Steady-state fluorescence anisotropy values were recorded on a Beacon 2000 apparatus (Panvera, Madison, WI) in a thermostated cell, at 37 °C for kinetics studies or 20 °C for competition experiments. For kinetics studies, fluorescein-labeled DNA duplexes (4 nM) (wild-type substrate or modified oligonucleotides) were added to a 200- μ l buffered solution containing 20 mM HEPES (pH 7.2), 1 mM DTT, and 7.5 mM MgCl₂. IN-DNA complex formation was then triggered by adding IN (200 nM final concentration). A constant level of 20 mM NaCl was maintained in all reaction mixtures. Fluorescence anisotropy values were then recorded, and the binding curves were used to determine the pseudo first-order binding rate constant, $k_{on} = k_1 \times [IN]$ (where k_1 is the second-order binding rate constant). A single-exponential model was fitted to the data for DNA-IN complex formation.

For competition experiments, fluorescein-labeled DNA duplexes (2 nM) were first incubated with various concentrations of unlabeled DNA duplexes (either unmodified or modified oligonucleotides) in a buffered solution containing 20 mM HEPES (pH 7.2), 1 mM DTT, 20 mM NaCl, and 5 mM MgCl₂. IN (120 nM final concentration) was then added, and steady-state anisotropy values (r) were recorded after equilibrium was reached. The r value corresponding to free fluorescein-labeled DNA was subtracted from the measured r value to give Δr . Finally, Δr values were plotted against competitor concentration to determine apparent K_d (concentration of unlabeled DNA decreasing the initial Δr value by 50%).

RESULTS

This study aimed to identify the elements of viral DNA structure determining the sequence specificity of interactions with HIV-1 IN. We synthesized 21-mer DNA duplexes containing modifications inducing local distortions in the DNA double helix (27, 28, 30–32) and, therefore, DNA destabilization (supplemental Table S1). Oligonucleotides mim-

icking the HIV-1 U5 viral DNA end were synthesized with single modifications at a particular residue (Fig. 1). Modified substrate analogs contained only one modification in the processed (U5B) strand, the unprocessed (U5A) strand, or both. The native nucleosides in the U5 substrate were replaced by non-nucleoside insertions (1,3-propanediol residues), 2'-aminonucleosides, or 2'-O-methyluridine. It has been reported that the most important nucleosides for integration are those at positions 3–9 of the processed strand (17, 20, 21, 24, 26). However, nucleoside mutations distal to the LTR end (positions 10–13) have also been found to affect 3'-end processing, but only in the presence of Mg²⁺ ions (1). We assessed the importance of various nucleosides within the LTR sequence, by making modifications at positions 3, 5–9, and 12 of the U5 substrate (Fig. 1). Position 4 was not modified because the modification of this base pair has been shown to inhibit 3'-end processing (20, 21, 24, 26).

Modifications at Position 12—We investigated the influence of U5 substrate modifications on the ability of IN to mediate 3'-end processing. We found that replacement of the nucleosides at position 12 by the non-nucleoside insertion P did not affect IN activity in presence of Mg²⁺ or Mn²⁺ ions (supplemental Table S2). These nucleosides therefore do not form contacts with IN that are important for catalytic activity. By contrast, modifications at other sites on the U5 substrate significantly influenced the processing rate.

Modifications at Positions 7–9—Both non-nucleoside insertions and nucleoside modifications at position 9 decreased the 3'-end processing rate (Table 1). Similar effects were observed for modifications of the processed and unprocessed strands. Non-nucleoside insertions decreased the reaction rate more strongly than the insertion of a modified nucleoside. The 3'-end processing rate was lower if both substrate strands were modified. Therefore, nucleosides at position 9 in both strands of the DNA are involved in interactions with IN.

For positions 7 and 8, we found that all nucleoside modifications on the unprocessed strand, U5A, decreased the reaction rate (Table 1). By contrast, non-nucleoside insertions in the processed strand, U5B, had no effect on 3'-end processing rate (Table 1). Similar results were obtained if the thymidine at position 7 was replaced by 2'-O-methyluridine. Therefore, IN interacts principally with the nonprocessed strand of the U5 substrate at positions 7 and 8.

It should be especially noted that replacing position 7 of the U5B strand with 2'-aminouridine increased 3'-end processing rate (Table 1). This can be explained by the formation of an additional contact between the amino group of 2'-aminouridine and an amino acid inside IN. It may be an extra hydrogen bond or an electrostatic interaction that occurs if this amino group is at least partially protonated ($pK_a = 6.2$ (33, 34)). We verified this by treating an oligonucleotide containing 2'-amino-

TABLE 1

The relative initial rates of 3'-end processing of the duplexes containing modifications at positions 7–9

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand			U5A strand		Both strands	
		P	U ^N	U ^M	P	aA ^N	P	U ^N /aA ^N
7	Mg ²⁺	1.10 ± 0.09	1.81 ± 0.21	0.91 ± 0.11	0.43 ± 0.06	0.73 ± 0.09	0.69 ± 0.08	1.58 ± 0.12
	Mn ²⁺	0.83 ± 0.07	1.42 ± 0.12	0.82 ± 0.10	0.41 ± 0.05	0.57 ± 0.08	0.44 ± 0.06	1.24 ± 0.10
8	Mg ²⁺	0.92 ± 0.09	— ^a	—	0.44 ± 0.06	—	0.79 ± 0.09	—
	Mn ²⁺	0.93 ± 0.09	—	—	0.68 ± 0.08	—	0.81 ± 0.11	—
9	Mg ²⁺	0.62 ± 0.08	0.94 ± 0.08	—	0.61 ± 0.07	0.79 ± 0.08	0.53 ± 0.07	0.67 ± 0.08
	Mn ²⁺	0.59 ± 0.07	0.92 ± 0.10	—	0.62 ± 0.08	0.83 ± 0.10	0.48 ± 0.07	0.54 ± 0.07

^a — indicates 3'-end processing assay was not carried out.

TABLE 2

The relative initial rates of 3'-end processing of the duplexes containing modifications at positions 5–6

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand		U5A strand			Both strands	
		P	aA ^N	P	U ^N	U ^M	P	aA ^N /U ^N
5	Mg ²⁺	0.10 ± 0.02	— ^a	0	—	—	0	—
	Mn ²⁺	0.31 ± 0.04	—	0.11 ± 0.03	—	—	0.09 ± 0.01	—
6	Mg ²⁺	0.21 ± 0.03	0.29 ± 0.04	0	0	0.10 ± 0.02	0	0
	Mn ²⁺	0.34 ± 0.05	0.46 ± 0.07	0.18 ± 0.03	0.27 ± 0.04	0.18 ± 0.02	0.21 ± 0.04	0.19 ± 0.02

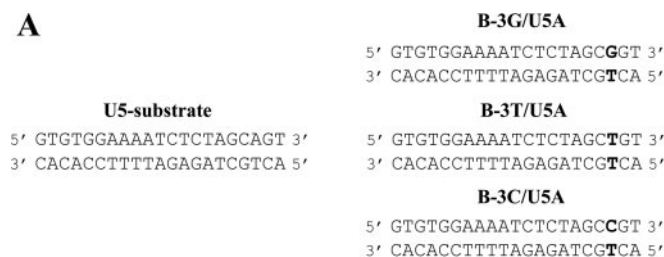
^a — indicates 3'-end processing assay was not carried out.

uridine at position 7 (B-N7) with succinic anhydride, to give an oligonucleotide containing a carboxyl group (B-N7-ac) (Fig. 2A). We found that the 3'-processing rates of the duplex with this carboxyl group (B-N7-ac/U5A) and nonmodified U5 substrate were the same (Fig. 2B). Therefore, the increase in the 3'-processing rate of the substrate containing 2'-aminonucleoside at the processed strand position 7 is because of the presence of the 2'-amino group. This processing rate increase was also observed when the seventh nucleosides in both strands were modified. However, the effect was less, probably because of the opposition of two different effects: an increase in processing rate caused by the modification of the processed strand, and a decrease caused by the modification of the non-processed strand. Finally, for all modifications at positions 7–9, an equal effect on processing rate was observed in the presence of either Mn²⁺ or Mg²⁺ ions (Table 1).

Modifications of Positions 5 and 6—By contrast with the positions 7 and 8, modifications of the nucleosides in either strand at positions 5 and 6 affected the 3'-processing rate (Table 2). However, the rate decrease was much greater when the nonprocessed strand was modified. Replacement of native nucleosides by both the non-nucleoside insertion (P) and the modified nucleosides, 2'-aminonucleosides or 2'-O-methylnucleosides, resulted in a similar effect (Table 2). Therefore, IN was sensitive to any distortion of the double helix in this DNA region and not especially affected by heterocyclic base removing. The inhibition effect of the nucleoside modification was much greater in the presence of Mg²⁺ than in the presence Mn²⁺ ions (Table 2).

Processing of Substrates Containing Modifications at Position 3—Many studies have shown that adenine at position 3 is important for the efficient processing of HIV-1 DNA (1, 17, 20, 21, 24, 26). However, some modifications of the third nucleoside do not inhibit processing (25, 35). We investigated the interaction between IN and the adenine at position 3 by studying the 3'-end processing of U5 substrate analogs in which this adenine was replaced by guanine, thymine, or cytosine (Fig. 4A). All these duplexes had noncomplementary base pairs instead of A/T at position 3, because the wild-type oligonucleotide U5A was used as the unprocessed strand in all cases. In the presence of Mg²⁺ ions, replacement of the conserved adenine by any other base decreased the proc-

A



B

duplex	U5-substrate	B-3G/U5A	B-3C/U5A	B-3T/U5A
<i>Vo rel. (Mg²⁺)</i>	1,00	0,21 ± 0,05	0,19 ± 0,07	0,29 ± 0,06
<i>Vo rel. (Mn²⁺)</i>	1,00	0,62 ± 0,07	0,51 ± 0,08	0,88 ± 0,07

FIGURE 4. Replacement of the adenine at position 3 by other heterocyclic bases. A, structures of U5 substrate analogs in which the adenine at position 3 was replaced by guanine, thymine, and cytosine. B, initial 3'-processing rates. The initial rate of U5 substrate 3'-end processing was set at 1.

essing rate by a factor of 3–4 (Fig. 4B). In the presence of Mn²⁺ ions IN cleaved the mismatched duplexes (B-3G/U5A, B-3T/U5A, and B-3C/U5A) more efficiently than in presence of Mg²⁺, and 3'-end processing rates depended on the base replacing the adenine (Fig. 4B). The processing rate of the T-containing analog was similar to that of the wild-type substrate, whereas replacement of the A by a C or a G approximately halved the processing rate.

Replacing 2'-deoxyadenosine by 2'-aminonucleoside (aA^N) at position 3 had no significant effect on processing rate in the presence of Mn²⁺ ions (Table 3). In contrast, processing was almost completely inhibited in the presence of Mg²⁺ ions (Fig. 3 and Table 3). Although 2'-aminonucleoside (aA^N) contains adenine, it is not in the same position as within the wild-type substrate. IN may therefore recognize the adenine in the modified substrate only if it has a flexible structure. The active site of IN is thought to be more rigid in the presence of Mg²⁺ than in the presence of Mn²⁺ (1). Therefore, changing the position of adenine inhibits 3'-end processing only in the presence of Mg²⁺ ions, because IN may adopt a new conformation in the presence of Mn²⁺ ions, providing greater access to the modified substrate structure.

TABLE 3**The relative initial rates of 3'-end processing of the duplexes containing modifications at the 3rd position**

The U5 substrate initial rate is considered to be equal to 1.

Position	Cofactor	Initial processing rate						
		U5B strand		U5A strand			Both strands	
		P	aA ^N	P	U ^N	U ^M	P	aA ^N /U ^N
3	Mg ²⁺ Mn ²⁺	0	0.10 ± 0.02 0.86 ± 0.08	1.31 ± 0.14 1.28 ± 0.07	1.25 ± 0.12 1.27 ± 0.09	1.80 ± 0.19 1.27 ± 0.12	0	0 2.14 ± 0.11

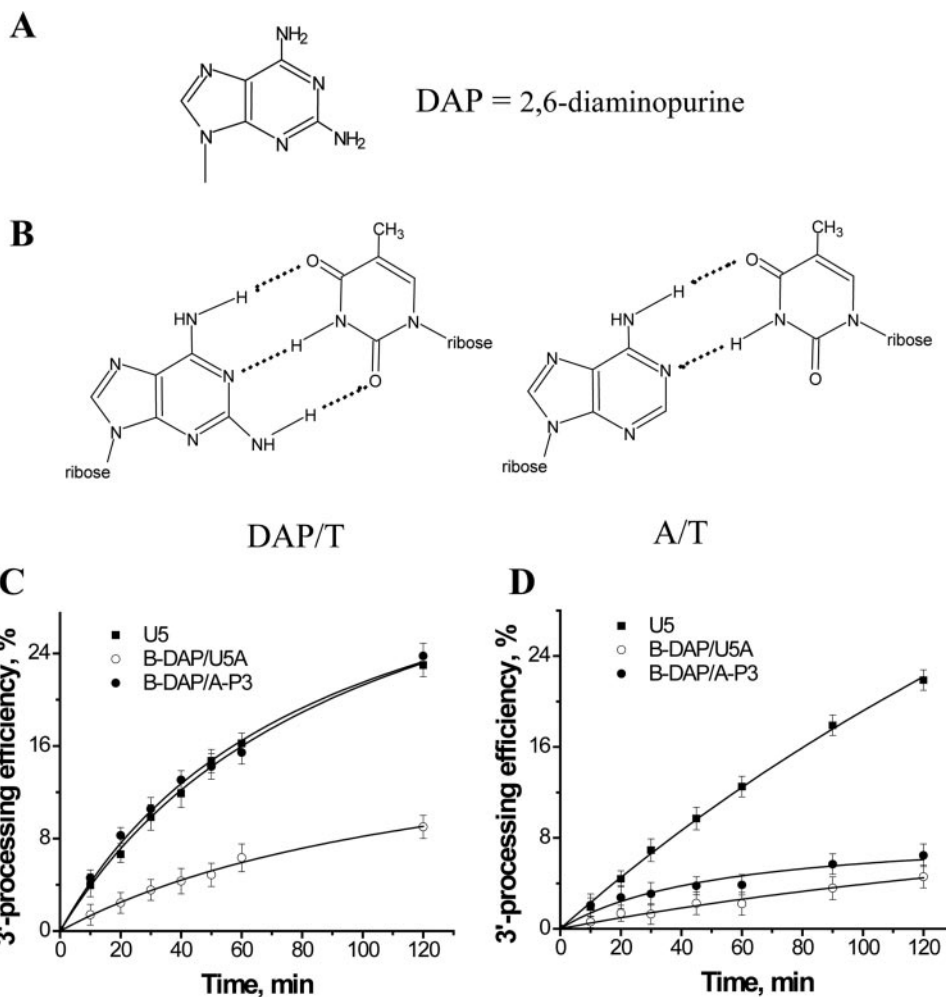


FIGURE 5. Effect on 3'-end processing of replacing the adenine at position 3 by DAP. *A*, structure of DAP. *B*, formation of Watson-Crick base pairs between thymine and adenine (A/T) and between thymine and 2,6-diaminopurine (DAP/T). Kinetics of 3'-end processing product accumulation in the case of U5 substrate and its analogs containing 2,6-diaminopurine in the U5B strand and the wild-type U5A strand (B-DAP/U5A) or the U5A strand with a non-nucleotide insertion at position 3 (B-DAP/A-P3), in the presence of Mn²⁺ ions (*C*) and Mg²⁺ ions (*D*) is shown.

Replacement of the 2'-deoxyadenosine at position 3 by 1,3-propanediol (P), resulting in the elimination of all functional groups of the adenine, completely inhibited 3'-end processing, regardless of the cofactor used (Table 3).

Replacement of the thymidine at position 3 in the unprocessed strand by 1,3-propanediol (P) or any modified nucleoside (U^N or U^M) increased the 3'-end processing rate (Table 3 and Fig. 3), regardless of the cofactor used. All these modifications at position 3 destabilized the DNA double helix (supplemental Table S1). This destabilization may stimulate 3'-end processing, consistent with previous data demonstrating an increase in 3'-processing activity because of destabilization of the 3rd bp (26, 35, 36).

The duplex with non-nucleoside insertions in both strands was not processed, regardless of the cofactor used (Table 3). In the presence of Mg²⁺ ions, no 3'-processing was also observed for the duplex containing 2'-aminonucleosides in both strands, because of the inhibitory effect of the processed strand modification (Fig. 3 and Table 3). As replacing

the native A by aA^N did not affect 3'-end processing in the presence of Mn²⁺ ions, we observed an increase in reaction rate for the substrate analog containing two 2'-aminonucleosides in presence of this ion (Table 3). This effect may be due to the modification of both strands, further destabilizing the duplex.

As destabilization of the A/T base pair at position 3 stimulated 3'-end processing, we hypothesized that the stabilization of this base pair would decrease the processing rate. We tested this hypothesis by studying 3'-end processing of U5 substrate analogs containing 2,6-diaminopurine (DAP) instead of the adenine at position 3 of the U5B strand (Fig. 5*A*). The DAP base stabilizes the DNA structure by forming three hydrogen bonds with the complementary thymine, instead of two bonds for the normal Watson-Crick pair A/T (Fig. 5*B*) (37). Processing was strongly inhibited in the presence of either metallic cofactor following the incorporation of DAP into the U5B strand (duplex B-DAP/U5A) (Fig. 5, *C* and *D*). However, the 3'-end processing rate was increased by elimination of the hydrogen bonding between DAP and T because of

New HIV DNA Analogs for Studying Interactions with Integrase

replacement of the T residue by 1,3-propanediol (duplex B-DAP/A-P3). Processing was as rapid as for the wild-type substrate in the presence of Mn^{2+} ions (Fig. 5C). A smaller increase in processing rate was observed in presence of Mg^{2+} ions (Fig. 5D).

We also studied interactions between IN and two 26-mer duplexes, because we assumed that A/T pair destabilization would enable IN to mediate cleavage of the U5 sequence in a more extended DNA duplex (Fig. 6). One duplex contained the sequence of the U5 substrate with 5 additional base pairs (26B/A). In the second duplex (26B/Amut), we replaced the thymine preceding the conserved adenine by an adenine, resulting in the noncomplementary pair A/A (Fig. 6A). The 3'-end proc-

essing efficiency of the 26B/A duplex was very low in the presence of Mn^{2+} ions (Fig. 6B), and no processing was observed in the presence of Mg^{2+} ions (Fig. 6C). In the case of the duplex containing the noncomplementary pair (26B/Amut), the 3'-processing efficiency increased in the presence of Mn^{2+} ions (Fig. 6B). Moreover, IN was able to cleave the mutated substrate even in presence of Mg^{2+} ions (Fig. 6C). Thus, A/T pair destabilization is essential for DNA cleavage by IN.

Integrase Binding to Modified Substrate Analogs—DNA substrate modification may influence the overall enzymatic reaction, either by affecting the substrate binding step or by influencing the catalytic activity of the enzyme. We therefore carried out electrophoretic mobility shift assays and steady-state fluorescence anisotropy experiments to test the effect of U5 substrate modifications on DNA binding by IN. We showed previously that the anisotropy parameter (r) is particularly useful for studies of IN/DNA interactions using fluorescein-labeled DNA, with monitoring of the difference in r value between the IN-free and the IN-bound DNA (38). We studied IN binding to substrate analogs containing the modifications with the most pronounced effect (inhibition or stimulation) on 3'-end processing rate (Table 4). All experiments were carried out in the presence of Mg^{2+} ions.

We first studied the binding isotherms for DNA/IN interactions, for the modified U5 substrate analogs, by means of electrophoretic mobility shift assays (supplemental Fig. S1). IN bound equally well to all the modified duplexes tested. The binding curves for the modified substrates were similar to that for the control wild-type U5 substrate (Fig. 7). Thus, substrate modifications with major effects on 3'-end processing seemed to have no effect on substrate-IN affinity.

We then investigated IN/DNA interactions by steady-state fluorescence anisotropy (Table 4). The fluorescein-labeled wild-type DNA substrate was incubated in the presence of various concentrations of unlabeled modified DNA duplexes (Table 4) before the addition of IN. We calculated apparent K_d values from the competition curves for the modified DNA duplexes (Table 4). Consistent with the results of electrophoretic mobility shift assays, the apparent affinities of IN for the various modified duplexes were comparable with the one for wild-type substrate. A slight affinity increase was obtained for substrates containing the abasic site, which display no activity under magnesium conditions (U5B/A-P5 and B-P3/U5A $K'_d \approx 8.5$ nM, as compared with U5B/U5A $K'_d \approx 21$ nM). Moreover, other DNA substrates (U5B/A-N6 and B-N3/U5A) that display no 3'-processing activity have an apparent affinity for IN comparable with the wild-type substrate. Finally, processing of the modified duplex B-N7/U5A, which is characterized by the weakest affinity ($K'_d \approx 36$ nM), was stimulated (Table 1). Therefore, no

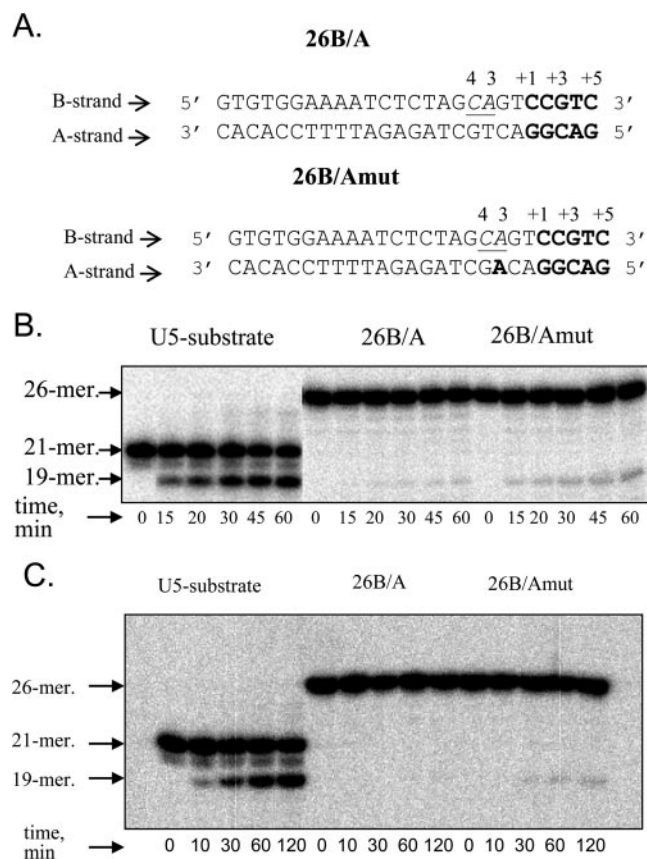


FIGURE 6. 3'-End processing of substrate analogs containing additional base pairs. A, structure of substrate analogs containing additional base pairs. Analysis of the processing products in the presence of Mn^{2+} ions (B) and Mg^{2+} ions (C) is shown.

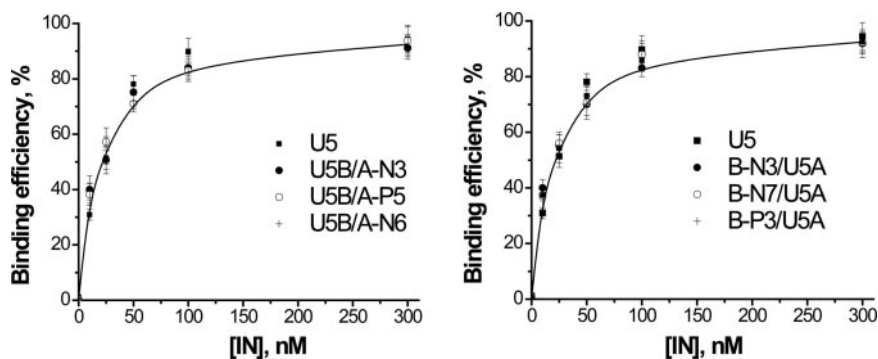
TABLE 4
Study of the IN binding with U5 substrate and its modified analogs

The type of the strand (A or B) modification is noted by N for 2'-aminonucleosides and P for 1,3-propanediol insertion; position of the modified nucleoside in the strand is noted by a corresponding figure.

DNA duplex	Sequence	K'_d	k_{on}^a
		nM	min^{-1}
U5 substrate	5'-GTGTGGAAAATCTCTAGCAGT-3' U5B strand	21.0 ± 3.5	0.74 ± 0.04
U5B/U5A	3'-CACACCTTTTAGAGATCGTCA-5' U5A strand		
U5B/A-N3	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	26 ± 3	0.89 ± 0.04
U5B/A-P5	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATPCTCA	8.6 ± 0.4	1.07 ± 0.05
U5B/A-N6	GTGTGGAAAATCTCTAGCAGT CACACCTTTTAGAGATCGTCA	20 ± 2	1.07 ± 0.05
B-P3/U5A	GTGTGGAAAATCTCTAGCPGT CACACCTTTTAGAGATCGTCA	8.2 ± 0.8	1.00 ± 0.05
B-N3/U5A	GTGTGGAAAATCTCTAGCAA ^N GT CACACCTTTTAGAGATCGTCA	20 ± 3	0.71 ± 0.04
B-N7/U5A	GTGTGGAAAATCTCTU ^N AGCAGT CACACCTTTTAGAGATCGTCA	36 ± 5	1.13 ± 0.06

^a k_{on} and K'_d (apparent K_d) values were obtained from steady-state anisotropy measurements as indicated under "Experimental Procedures."

FIGURE 7. Binding of modified DNA substrates to IN (5 nM duplex, 10–300 nM integrase) in the presence of Mg^{2+} ions. Curves corresponding to the accumulation of IN-bound substrates were calculated using electrophoretic mobility shift assays data (supplemental Fig. S1). Structures of the modified substrates containing 2'-amino-nucleotides (N) or 1,3-propanediol insertion (P) are shown in the Table 4.



correlation between the apparent affinity and the specific activity was detected.

As we showed previously that IN-DNA complexes form fairly slowly (38, 39), this step might be differentially affected by DNA modifications, with effects on subsequent IN activity. We therefore investigated the effects of substrate modifications on the kinetics of DNA binding to IN (Table 4). Fluorescein was attached to the 5'-end of the U5B strand of the duplexes when the U5A strand was modified and was attached to the 3'-end of the U5A strand when the U5B strand was modified. The pseudo first-order rate constants (k_{on}) were calculated for the various modified DNA duplexes (Table 4). Again, no significant difference was observed between modified and unmodified substrates, and no correlation between the effect of a given modification on DNA-binding rate and 3'-end processing rate was observed. Thus, the catalytic step is more sensitive to DNA modifications than the binding step. IN is known to have the intrinsic ability to bind target DNA in a nonspecific manner. This nonspecific binding may lead to underestimation, at least *in vitro*, of specific binding, minimizing the structural effects of DNA modifications on IN/DNA recognition. In the 3'-end processing assay, a smaller population of IN-DNA complexes is observed, the subpopulation corresponding to the correct and specific positioning of IN on its DNA substrate. Structural effects of DNA on this subpopulation may be more pronounced and easier to detect in activity assays than in DNA binding assays.

DISCUSSION

In the absence of crystallographic or NMR structures for IN-DNA complexes, most of what is known about interactions between IN and its substrates is based on functional studies using modified DNA analogs. We tried to identify the elements of viral DNA structure important for IN binding and 3'-end processing. We first used a novel type of U5 substrate analog containing sugar-phosphate backbone modifications at various positions in one or both strands (Fig. 1). We investigated two different types of DNA modification: removal of the nucleoside by replacement with a 1,3-propanediol residue (P) or modification of the sugar conformation, changing the furanose 2'-position. It has been shown that the replacement of thymidines in the U5 substrate by 2'-deoxyuridines has no major influence on 3'-end processing (25), indicating that the methyl groups of the thymine residues in the U5 substrate are not involved in interactions with IN. The effect on the 3'-end processing rate of replacing T residues by 2'-modified uridines as demonstrated here therefore results from the unusual furanose structure rather than the elimination of methyl groups from heterocyclic bases.

It is important that all our modifications introduced a local distortion in the DNA double helix structure, decreasing T_m (supplemental Table S1). The instability caused by 2'-amino modification probably results from heterocyclic base destacking (27), causing base displacement in the double helix structure. In the case of U^M , the

destabilization may result from its C-3'-endo sugar puckering typical of the A-form geometry of RNA (32, 40). However, 2'-modified nucleosides retain their heterocyclic bases and may maintain at least some contacts with IN, whereas the incorporation of 1,3-propanediol results in the complete elimination of all functional groups that might interact with IN.

Most of the substrate positions involved in the specific activity of IN are located upstream from the 9th bp (20, 22, 25). However, nucleotides at positions 10–13 have also been shown to be important (1). We found that modifications at position 12 did not alter IN activity. We therefore focused on positions 3–9. We found that their modifications altered the 3'-end processing rate. As the conflicting results reported in previous publications may arise from the differential use of Mn^{2+} or Mg^{2+} ions, we also compared, for each modification, the 3'-end processing rate in the presence of these two cations.

We found that the catalytic step was sensitive to modifications to DNA-substrate structure but that the modifications had no effect on substrate binding to IN. In our hands, it was not possible to distinguish between specific and nonspecific binding of IN to DNA. The K_d value derived from binding to the U5 sequence was very close to the one obtained with nonspecific double-stranded DNA (25 nM; data not shown). Under our enzymatic conditions ($E > S$), IN activity was correlated to the amount of IN-DNA complexes. It was therefore crucial to verify whether the substrate modifications did not alter the quantity of complexes formed *in vitro*. No modification of IN binding was observed, thus indicating that the alteration of the catalysis was related to a post-binding event, most likely a modification of the DNA conformation, which prevented the cleavage to occur.

Integrase Interacts with the DNA Sugar-Phosphate Backbone at Positions 5–9—We identified three different sites, between nucleosides 5 and 9 inclusive, where modification affected 3'-end processing rate in different ways.

All modifications at position 9 affected processing rate, but this effect was moderate in all cases (Table 1). The strongest effect was obtained by replacing the base pair by 1,3-propanediol (P). However, despite the P insertion causing the largest DNA distortion (the T_m shift was 8.3 °C) because of the disruption of both hydrogen bonding and heterocyclic base stacking, the processing rate was no more than halved. This suggests that IN interacts with the sugar-phosphate backbone, without making specific contacts with the bases. Moreover, single modifications on either strand affected IN activity equally, suggesting that this interaction likely involves both strands.

At positions 7 and 8, the total removal of heterocyclic bases from the processed strand had no significant effect on 3'-end processing rate (Table 1, modification P), ruling out the possibility of specific contact with heterocyclic bases. By contrast, similar modifications of the unprocessed strand decreased IN activity (Table 1, modification P) thus allowing specific interactions with IN for heterocyclic bases from this strand. It has been

New HIV DNA Analogs for Studying Interactions with Integrase

shown, however, that substitutions at positions 7 and 8 with other complementary base pairs, which do not disturb the DNA double helix structure but change the pattern of putative bases contacts with IN, do not affect 3'-end processing (1). Thus, the inhibitory effect of P insertion results from the loss of IN contacts with the sugar-phosphate backbone because of its local distortion rather than a loss of contacts with heterocyclic bases.

Any modification of the fifth and sixth native nucleosides, either by P insertion or by replacement with 2'-modified nucleosides, led to a large decrease in the 3'-end processing rate (Table 2). Similar levels of inhibition were observed with all substrate modifications at position 6, suggesting that IN recognizes the conformation of the DNA backbone rather than nucleoside bases. Previous studies have shown that replacing the guanine at position 5 with *O*⁶-methylguanine decreases substrate cleavage (26). It has been suggested that IN interacts with this heterocyclic base via the major groove of the DNA. However, the 5th and 6th bp are inverted in the U5 and U3 extremities (5'-A(G/C)T-3' in U5 and 5'-T(C/G)A-3' in U3), although both sequences are substrates of the enzyme. This observation is not consistent with the direct recognition of the bases at these positions by IN. Moreover, the effect of *O*⁶-methylguanine may also result from the DNA structure distortion, because guanine *O*⁶-methylation modifies the conformation of the phosphodiester backbone at the modification site (41). It therefore seems more likely that IN recognizes a DNA fine structure generated by the G/C and A/T pairs at positions 5 and 6, respectively. Deformation of the double helix by either base replacement or nucleoside modification prevents DNA recognition, thereby strongly inhibiting 3'-end processing (Table 2).

We conclude that IN recognition involves only interactions with the DNA sugar-phosphate backbone at positions 5–9. Both nucleosides of the 9th bp participate in these interactions. For positions 5–8, modifications of the unprocessed strand decrease 3'-end processing rate more strongly than modifications of the processed strand (Tables 1 and 2), indicating that active IN-DNA complex formation primarily requires interactions with the unprocessed strand. Nevertheless, at positions 5 and 6, contact between IN and both strands is important for specific DNA recognition (Table 2).

Destabilization of the A/T Pair at Position 3 Is Necessary for 3'-End Processing—Modification of the thymidine at position 3 in the unprocessed strand, by P insertion or 2'-modified nucleosides (U^N and U^M), increased 3'-end processing rate (Table 3) probably because of local destabilization of the A/T base pair, as suggested previously (36). The processing stimulation following the thymidine replacement likely points to an absence of specific contact between IN and this base. Alternatively, one could consider that mispairing may result in a loss of a specific contact otherwise required for the unpairing step. However it must be noted that according to Ref. 42, binding of the avian homolog of HIV integrase to its DNA substrate enhances the accessibility and the subsequent modification of the third thymidine by KMnO₄. Thus, although integrase disrupts the A/T base pair, it does not protect the heterocyclic base from the modification, thereby ruling out a close contact between the base and the enzyme. The stimulation caused by A/T base pair destabilization is also consistent with the inhibition elicited by replacing the adenine at position 3 with DAP, a base known to stabilize the double helix as a result of the formation of three hydrogen bonds (Fig. 5B) (37). Accordingly, the disruption of these hydrogen bonds by replacing thymidine in the DAP/T base pair by 1,3-propanediol (duplex B-DAP/A-P3) led to an increase in 3'-end processing rate (Fig. 5, C and D).

The negative effect of DAP substitution on 3'-processing has been reported before, and it has been suggested that DAP disrupts a specific contact between IN and the C-2 position of the adenine at position 3

(26). However, such a contact is not consistent with the efficient cleavage of the mismatched substrate B-3T/U5A (Fig. 4) or the stimulation observed if the thymidine complementary to DAP is replaced by 1,3-propanediol (Fig. 5C). We therefore suggest that DAP prevents the local A/T pair destabilization that is essential for efficient processing. The necessity of this destabilization also found for avian sarcoma virus integrase (42) may account for the strict conservation of the CA/TG sequence in integrase substrates, as intrinsic conformational mobility is known to favor DNA flexibility (43–47). Such a local destabilization allows IN activity, even at an internal position, as seen in the internal processing of a 26-mer duplex (26B/A). As predicted, processing efficiency increased considerably when the A/T base pair was replaced by a noncomplementary A/A pair (duplex 26B/Amut) (Fig. 6).

Specific and Metal-dependent Interactions with the Adenosine at Position 3 in the Processed Strand—It has been reported that replacement of the A/T base pair at position 3 by other complementary pairs (1, 35) or replacement of the adenine in this position with *N*⁷-deaza-adenine (26) decreases 3'-end processing efficiency, suggesting that specific contacts with this adenine are required. However, other studies have shown that replacement of the A/T pair by either noncomplementary base pairs (35) or an abasic site (AP-site) (25) does not affect 3'-end processing efficiency. Our study reconciles these contradictory data by demonstrating the important role played by the metal cofactor. Indeed, we observed that the replacement of 2'-deoxyadenosine by P insertion completely inhibited 3'-processing in the presence of both Mg²⁺ and Mn²⁺ (Table 3). By contrast, the modification of this residue by 2'-aminonucleoside, aA^N, yielded a metal-dependent effect; 3'-end processing was completely inhibited in the presence of Mg²⁺ ions, whereas no inhibition was observed in the presence of Mn²⁺ (Table 3). It has been suggested that Mg²⁺ coordination makes the active site of IN more rigid, whereas Mn²⁺ coordination does not (1, 4). The loss of flexibility caused by Mg²⁺ coordination may prevent IN from establishing specific contacts with the adenine within aA^N, as the sugar modification changes the location of the adenine. As a result, 3'-processing of the aA^N-containing substrate is inhibited in the presence of Mg²⁺ but unaffected in the presence of Mn²⁺, which maintains structural flexibility of IN.

We investigated possible specific contacts between IN and the adenine at position 3 by studying the processing of substrates in which this adenosine was replaced by other nucleosides (Fig. 4). Adenine, which is simultaneously involved in Watson-Crick base pairing, can form hydrogen bonds with IN via its N-7 atom (proton acceptor) and the *exo*-cyclic amino group (proton donor) in the major groove and via the N-3 atom (proton acceptor) in the minor groove. A specific contact between IN and the N-7 atom has been suggested, based on the decrease in processing efficiency by a factor of 3 caused by replacing the adenine at position 3 with *N*⁷-deaza-adenine, which cannot maintain this contact (26). The low processing rate observed for the C-containing duplex (B-3C/U5A) (Fig. 4B), for which a contact similar to that formed by N-7 is impossible, confirmed the necessity of this contact. However, the strong decrease in processing rate for the B-3G/U5A duplex (Fig. 4B), despite the presence of the same proton acceptors in guanine and adenine (N-3 and N-7 atoms), indicates that the NH₂ group of adenine is also involved in an interaction with IN. The suggestion that two contacts are formed between the adenine and IN is supported by the finding that the T-containing duplex (B-3T/U5A) was processed almost as efficiently as the wild-type substrate (Fig. 4B). Indeed, thymine, which is not involved in Watson-Crick base pairing in this duplex, may form two hydrogen bonds with IN, involving O-4 as the proton acceptor (similar to N-7 of adenine) and the NH group as the proton donor. As the geometry of the bonds formed by thymine is not identical to that of the bonds

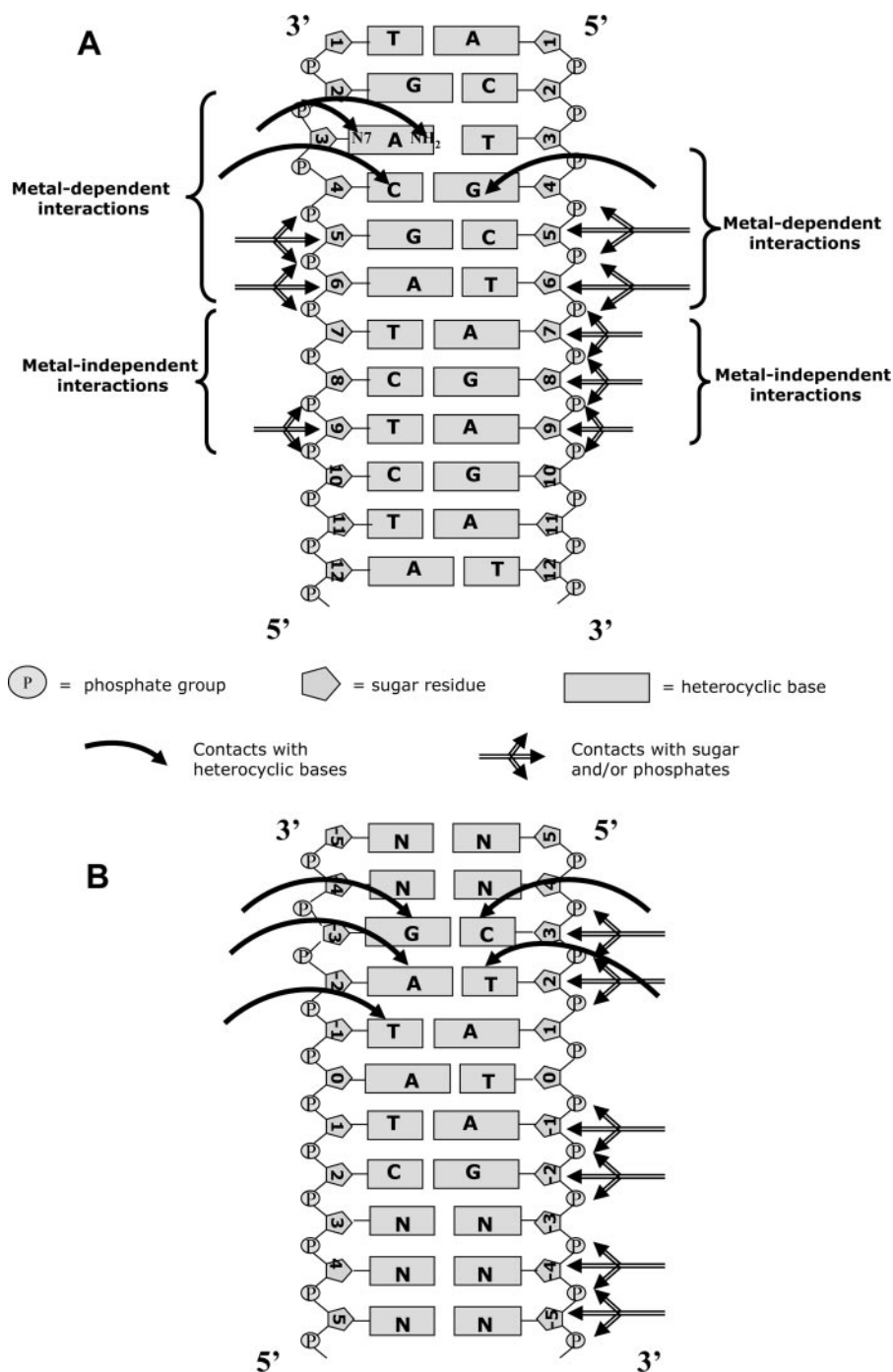


FIGURE 8. Scheme of contacts formed between HIV-1 IN and the U5 substrate proposed in this study (A) and EcoRV restriction endonuclease and its DNA substrate described in Ref. 51 (B). Contacts of IN with heterocyclic bases at position 4 of the U5 substrate have been described elsewhere (1, 20, 21). The size of the arrows corresponds to the importance of contacts for 3'-end processing rate.

formed by adenine, these bonds can be generated only if the structure of IN structure is flexible enough to accept T-containing duplexes as a substrate, *i.e.* in presence of Mn^{2+} ions. Clearly, these contacts cannot be formed if the thymine participates in base pairing, and as result, the 3'-end processing of a substrate analog containing the T/A pair instead of A/T is inefficient (35).

Metal-dependent sensitivity of IN to substrate structure was also observed when positions 5 and 6 were modified (Table 2), although this sensitivity was less marked than for modifications at position 3. As the cation present primarily affects the conformation of the core domain (4–6), metal-dependent sensitivity indicates possible interactions with this domain. We therefore suggest that nucleosides at positions 3–6 within the DNA substrate interact with the catalytic core. By

contrast, the C-terminal domain is not directly involved in metal ion binding, and the structure of this domain therefore depends less on the metal present than does that of the catalytic core. Consequently, modifications of nucleosides interacting with the C-terminal domain should affect 3'-end processing regardless of the cofactor used. This was indeed found to be the case for nucleosides at positions 7–9 (Tables 1), consistent with known interaction of the C-terminal domain with the thymidine at position 7 (1).

Summary of IN/DNA Interactions—The interactions between IN and DNA substrate identified in this study are summarized in Fig. 8A. This scheme provides new insight into mechanistic similarities between IN and restriction endonucleases. Retroviral integrases and type II restriction enzymes have been reported to share conserved structural features

(48). Thus, in the archetypal EcoRV/DNA interaction model, the protein contacts the base just upstream from the scissile bond but does not contact the complementary base in the opposite strand (Fig. 8B). Similarly, IN interacts with the adenine at position 3, but not its complementary thymine (Fig. 8A). Position -2 in the EcoRV recognition sequence, which is similar position 4 in the HIV-1 DNA, displays the highest level of contact within the EcoRV-DNA complex, with contacts established on both strands. It has been suggested that heterocyclic bases from the 4th bp interact with IN (1, 20, 21). Moreover, in addition to base-specific contacts, numerous contacts of EcoRV with the DNA backbone allow recognition of the cognate site through a sequence-dependent backbone conformation (49, 50). Regarding HIV-1 IN, it was found that specific protein mutations altering ionic interactions or hydrogen bonding result in a change of DNA recognition specificity when the mutated amino acids are located close to DNA positions 5–11 (51). These data support our conclusion that IN recognizes the fine conformation of DNA backbone at positions 5–9 rather than nucleoside bases.

Finally, EcoRV endonuclease initially binds DNA nonspecifically, subsequently forming a specific complex as a result of conformational changes in both the DNA and the enzyme (49). The DNA is locally distorted from a regular B form to a highly strained kinked conformation, leading to unwinding of the DNA, unstacking of the two central base pairs of the recognition site, and bending of the DNA (50). For IN, this bending and unstacking may be favored by the conformational mobility of the CA/TG fragment (43–44) and a loss of regular base stacking (45).

Acknowledgment—We thank Françoise Simon for technical assistance.

REFERENCES

- Esposito, D., and Craigie, R. (1998) *EMBO J.* **17**, 5832–5843
- Engelman, A., and Craigie, R. (1995) *J. Virol.* **69**, 5908–5911
- Marchand, C., Johnson, A. A., Karki, R. G., Pais, G. C., Zhang, X., Cowansage, K., Patel, T. A., Nicklaus, M. C., Burke, T. R., Jr., and Pommier, Y. (2003) *Mol. Pharmacol.* **64**, 600–609
- 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
- Asante-Appiah, E., Seeholzer, S. H., and Skalka, A. M. (1998) *J. Biol. Chem.* **273**, 35078–35087
- Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., and Mikol, V. (1998) *J. Mol. Biol.* **282**, 359–368
- Polard, P., and Chandler, M. (1995) *Mol. Microbiol.* **15**, 13–23
- Kulkosky, J., Jones, K. S., Katz, R. A., Mack, J. P., and Skalka, A. M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338
- Drelich, M., Wilhelm, R., and Mous, J. (1992) *Virology* **188**, 459–468
- Gao, K., Wong, S., and Bushman, F. (2004) *J. Virol.* **78**, 6715–6722
- Asante-Appiah, E., and Skalka, A. M. (1997) *J. Biol. Chem.* **272**, 16196–16205
- Neamati, N., Lin, Z., Karki, R. G., Orr, A., Cowansage, K., Strumberg, D., Pais, G. C., Voigt, J. H., Nicklaus, M. C., Winslow, H. E., Zhao, H., Turpin, J. A., Yi, J., Skalka, A. M., Burke, T. R., Jr., and Pommier, Y. (2002) *J. Med. Chem.* **45**, 5661–5670
- Katzman, M., and Sudol, M. (1996) *J. Virol.* **70**, 2598–2604
- Chen, J. C., Krucinski, J., Miercke, L. J., 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
- Esposito, D., and Craigie, R. (1999) *Adv. Virus Res.* **52**, 319–333
- Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J.* **20**, 7333–7343
- Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J.* **16**, 6849–6859
- Gerton, J. L., Ohgi, S., Olsen, M., DeRisi, J., and Brown, P. O. (1998) *J. Virol.* **72**, 5046–5055
- Gao, K., Butler, S. L., and Bushman, F. (2001) *EMBO J.* **20**, 3565–3576
- LaFemina, R. L., Callahan, P. L., and Cordingley, M. G. (1991) *J. Virol.* **65**, 5624–5630
- Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992) *J. Virol.* **66**, 3593–3601
- Katzman, M., and Sudol, M. (1996) *J. Virol.* **70**, 9069–9073
- Mazumder, A., Neamati, N., Pilon, A. A., Sunder, S., and Pommier, Y. (1996) *J. Biol. Chem.* **271**, 27330–27338
- van den Ent, F. M. I., Vink, C., and Plasterk, R. H. A. (1994) *J. Virol.* **68**, 7825–7832
- Mazumder, A., and Pommier, Y. (1995) *Nucleic Acids Res.* **23**, 2865–2871
- Wang, T., Balakrishnan, M., and Jonsson, C. B. (1999) *Biochemistry* **38**, 3624–3632
- Zubin, E. M., Antsyovich, S. I., Oretskaya, T. S., Romanova, E. A., Volkov, E. M., Tashlitsky, V. N., Dolinnaya, N. G., and Shabarova, Z. A. (1998) *Nucleosides Nucleotides* **17**, 425–440
- Kuznetsova, L. G., Volkov, E. M., Romanova, E. A., Tashlitsky, V. N., Oretskaya, T. S., and Shabarova, Z. A. (1991) *Bioorg. Khim.* **17**, 1289–1291
- Leh, H., Brodin, P., Bischerour, J., Deprez, E., Tauc, P., Brochon, J. C., LeCam, E., Coulaud, D., Auclair, C., and Mouscadet, J. F. (2000) *Biochemistry* **39**, 9285–9294
- Gelfand, C. A., Plum, G. E., Grollman, A. P., Johnson, F., and Breslauer, K. J. (1998) *Biochemistry* **37**, 7321–7327
- Hendrix, C., Devreese, B., Rozenki, J., Van Aerschot, A., De Bruyn, A., Van Seeumen, J., and Herdewijn, P. (1995) *Nucleic Acids Res.* **23**, 51–57
- Lubini, P., Zurcher, W., and Egli, M. (1994) *Chem. Biol.* **1**, 39–45
- Aurup, H., Tuschl, T., Benseler, F., Ludwig, J., and Eckstein, F. (1994) *Nucleic Acids Res.* **22**, 20–24
- Kan, L.-S., Bauman, R., Bhan, P., and Miller, P. S. (1993) *J. Chin. Chem. Soc.* **40**, 469–473
- Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) *Genes Dev.* **11**, 371–382
- Agapkina, J., Smolov, M., Zubin, E., Mouscadet, J.-F., and Gottikh, M. (2004) *Eur. J. Biochem.* **271**, 205–211
- Bailey, C., and Waring, M. J. (1998) *Nucleic Acids Res.* **26**, 4309–4314
- Deprez, E., Barbe, S., Kolaski, M., Leh, H., Zouhiri, F., Auclair, C., Brochon, J.-C., Le Bret, M., and Mouscadet, J.-F. (2004) *Mol. Pharmacol.* **65**, 85–98
- Brodin, P., Pinskaya, M., Buckle, M., Parsch, U., Romanova, E., Engels, J., Gottikh, M., and Mouscadet, J.-F. (2002) *Biochemistry* **41**, 1529–1538
- Venkateswarlu, D., Lind, K. E., Mohan, V., Manoharan, M., and Ferguson, D. M. (1999) *Nucleic Acids Res.* **27**, 2189–2195
- Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) *Biochemistry* **25**, 1027–1036
- Katz, R. A., DiCandeloro, P., Kukulj, G., and Skalka, A. M. (2001) *J. Biol. Chem.* **276**, 34213–34220
- Dickerson, R. E. (1998) *Nucleic Acids Res.* **26**, 1906–1926
- Packer, M. J., Dauncey, M. P., and Hunter, C. A. (2000) *J. Mol. Biol.* **295**, 71–83
- Renisio, J. G., Cosquer, S., Cherrak, I., Antri, S., Mauffret, O., and Femandjian, S. (2005) *Nucleic Acids Res.* **33**, 1970–1981
- Dickerson, R. E., and Chiu, T. K. (1997) *Biopolymers* **44**, 361–403
- Jones, S., van Heyningen, P., Berman, H. M., and Thornton, J. M. (1999) *J. Mol. Biol.* **287**, 877–896
- Venclovas, C., and Siksny, V. (1995) *Nat. Struct. Biol.* **2**, 838–841
- Pingoud, A., and Jeltsch, A. (2001) *Nucleic Acids Res.* **29**, 3705–3727
- Pingoud, A., and Jeltsch, A. (1997) *Eur. J. Biochem.* **246**, 1–22
- Chen, A., Weber, I. T., Harrison, R., and Leis, J. (2006) *J. Biol. Chem.* **281**, 4173–4182