Major and Minor Groove Contacts in Retroviral Integrase−LTR Interactions†

Tan Wang,‡ Mini Balakrishnan,§,|| and Colleen B. Jonsson*,‡

Department of Chemistry & Biochemistry and Program in Molecular Biology, New Mexico State University, Las Cruces, New Mexico 88003

Received September 2, 1998; Revised Manuscript Received January 20, 1999

ABSTRACT: The 3′-processing activities of HIV-1, HTLV-2, and M-MuLV integrases (INs) with their corresponding U5 end of the viral DNA molecule were examined to define functional group determinants of U5 terminus recognition and catalysis. Nucleotide analogues were incorporated into the U5 terminus to produce conservative modifications in the surface of the major and/or minor grooves to map the hydrogen-bonding contacts required for LTR−IN interaction. Specifically, the phylogenetically conserved CA (positions 4 and 3, respectively) and the 5′-proximal nucleotide (position 5) were replaced with base analogues in plus and/or minus strands. For each integrase, similar major and minor groove contacts were identified in the guanine and adenine of the conserved CA/GT. Overall, perturbances in the minor groove resulted in a greater decrease in 3′-processing activity than the major groove substitutions. Additionally for HIV-1 and HTLV-2 INs, we observed an increase in the 3′-processing activity with an O^4-MeThy substitution at position 3 of the minus strand. O^4-MeThy may act to destabilize Watson−Crick base pairing and in doing so provide these INs with a more favorable interaction with the adjacent scissile bond. At position 5, a substantial divergence among the three INs was noted in the functional groups required for 3′-processing activity, thereby supporting the role of this position in providing some level of substrate specificity.

The retroviral protein integrase (IN)[1] mediates the covalent insertion of the retrovirus genome into the host chromosome following reverse transcription of the viral RNA genome to a DNA intermediate (*1*, *2*). This part of the virus life cycle, integration, requires the sequential completion of three distinct enzymatic events. In the first catalytic reaction, 3′-processing, IN cleaves two to three nucleotides from each 3′ end of the linear viral DNA to expose the phylogenetically conserved CA dinucleotide. The 3′-recessed ends are then available for the second catalytic step, strand transfer. This is a direct transesterification reaction, in which the nucleophilic 3′-hydroxyl group of each adenosine ribose attacks the 5′-phosphate in the phosphodiester backbone on opposing strands of the host DNA. These two enzymatic reactions, 3′-processing and strand transfer, are catalyzed solely by the retroviral IN (*3*, *4*). The final step in integration, gap repair, requires removal of the unpaired nucleotides at the 5′ ends of the viral DNA as well as the filling in and joining of the single-strand gaps between the viral and host DNA. While host enzymes could contribute to the completion of the integrated provirus, HIV-1 IN has been shown to have an intrinsic DNA polymerase activity which may play a role in gap repair and 5′-end joining (*5*). Each of these reactions has been reconstituted in vitro with double-stranded oligonucleotide substrates which mimic the end of the viral DNA molecule, and IN. The reverse of the strand transfer reaction, disintegration, is also observed in vitro for HIV-1 (*6*), HTLV-2 (*7*), and M-MuLV (*8*) INs.

A central question in the enzymatic reaction mechanism of integration is the molecular basis for IN−substrate interaction and moreover its ability to distinguish between a nonspecific target DNA substrate and the specific viral DNA termini. Because IN binds equally well to both nonspecific and specific DNA substrates (*9*−*11*), the contribution of individual nucleotides in specific DNA recognition of the termini has been mainly inferred through catalysis of U5, U3, and disintegration substrates. Single base pair substitutions in the viral DNA termini modulate IN activity and suggest specificity determinants lie in a limited region [for example, see (*12*−*19*)]. These studies show the subterminal CA and three nucleotides upstream of the CA significantly contribute to catalysis, and, thus, may also play a role in binding of the substrate in the catalytic site. Recognition specificity for the U5 terminus maps to the nucleotide 5′ to the CA dinucleotide (position 5) in the HTLV-1, HTLV-2, and M-MuLV INs (*16*). In contrast, HIV-1 IN shows a more relaxed degree of substrate specificity for the viral DNA at this position. Differences in substrate specificity preferences

† This work was suppported entirely by Grant 02517-21-RD from the American Foundation for AIDS Research.

* To whom correspondence should be addressed at the Department of Chemistry & Biochemistry, New Mexico State University, Box 30001, Department 3C, Las Cruces, NM 88003. Phone: 505-646-3346. Fax: 505-646-2649. E-mail: cjonsson@nmsu.edu.

‡ Department of Chemistry & Biochemistry.

§ Program in Molecular Biology.

|| Current address: Department of Biochemistry and Biophysics, University of Rochester Medical Center, 601 Elmwood Ave., Box 712, Rochester, NY 14642.

[1] Abbreviations: IN, integrase; LTR, long terminal repeat; HIV-1, human immunodeficiency virus type-1; HTLV-2, human T-cell leukemia virus type 2; M-MuLV, Moloney murine leukemia virus; N^6-methyladenine, 6-MeAde; O^6-methylguanine, 6-MeGua; 7-deazaadenine, 7-deazaAde; O^4-methylthymine, O^4-MeThy; 5-methylcytosine, 5-MeCyt; 2,6-diaminopurine, 2,6-$(NH_2)_2$Pur; EDTA, (ethylenedinitrilo)-tetraacetic acid; DTT, dithiothreitol; BME, β-mercaptoethanol; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate.

10.1021/bi982124i CCC: $18.00
Published on Web 03/03/1999

among INs also extend to catalysis of disintegration substrates (*7, 20*−*23*).

While the importance of specific nucleotides is observed in each of these studies, little is known regarding the mechanism through which the specific nucleotides contact integrase. To probe the chemical basis of the IN−viral DNA interaction, in particular hydrogen bonding, isosteric nucleotide analogues were introduced into the U5 terminus at the conserved CA and at position 5. Nucleotide analogues were chosen to produce a modification in the surface of the major or minor groove. In the following work, we report on the relative importance of specific functional groups required for catalysis by HTLV-2, M-MuLV, and HIV-1 INs with their respective U5 substrates.

MATERIALS AND METHODS

Oligonucleotide Substrates. Oligonucleotide sequences corresponding to the wild-type (WT) U5 terminus of HIV-1, HTLV-2, and M-MuLV genomes were used as substrates in enzymatic assays (Figure 1A). Oligonucleotide synthesis, introduction of 2′-deoxynucleoside phosphoramidite analogues, and HPLC purification of oligonucleotides were performed at Integrated DNA Technologies. Protected 2′-deoxynucleoside phosphoramidites were purchased from Glen Research (Figure 2B). Oligonucleotides were further purified on 20% denaturing polyacrylamide gels, 5′-end-labeled with [γ-^{32}P]ATP (Dupont−NEN) and T4 DNA kinase, and hybridized to complementary strands as described previously (*16*).

Protein Purification. INs from HIV-1, HTLV-2, and M-MuLV were purified as hexahistidine-tagged fusion proteins, expressed in *Escherichia coli* BL21(DE3) cells using the T7 polymerase expression system, and purified from the insoluble fraction essentially as described previously (*8, 16*). HIV-1 IN was additionally purified from the soluble fraction, and its specific activity was comparable to HIV-1 IN isolated from the insoluble fraction. To purify HIV-1 IN from the soluble fraction, cells were grown to midlogrithmic phase in 0.5−1-L cultures, and induced for 3 h in 0.4 mM isopropyl thiogalactoside. Cells were centrifuged, resuspended in 1× binding buffer (1 M NaCl, 20 mM Tris, pH 7.9, 5 mM imidazole), and lysed by treatment with lysozyme (0.2 mg/mL) for 30 min and sonication. The lysate was centrifuged for 30 min at 30 000*g*. The soluble portion was filtered through a 0.45-mm filter and applied to a HiTrap nickel-chelating column (Pharmacia). Washing and elution were carried out in 1× binding buffer with a linear gradient (5−800 mM) of imidazole. The final preparation was dialyzed against 1 M NaCl, 20 mM HEPES, pH 7.5, 0.1 mM EDTA, 1 mM DTT, and 10% (v/v) glycerol. Protein concentrations were measured by the Bradford method (*24*) using the BioRad Micro-Assay.

Integration Assays. Results of titration experiments gave the protein concentration (0.01 μg/μL for HIV-1 IN, 0.013 μg/μL for HTLV-2 IN, and 0.023 μg/μL for M-MuLV IN) that yielded maximal activity for each IN with its respective WT blunt LTR substrate. Reactions for HIV-1 and HTLV-2 INs contained 25 mM MOPS (pH 7.2), 10 mM BME, 10% glycerol, 0.75 mM CHAPS, 7.5 mM $MnCl_2$, 1 pmol of substrate, and the appropriate IN in a final volume of 15 μL. Reaction mixtures for M-MuLV IN contained 25 mM MES (pH 6.2), 10 mM DTT, 10% ethylene glycol, 10 mM $MnCl_2$, 10 mM KCl, 1 pmol of substrate, and M-MuLV IN in a final volume of 15 μL. HTLV-2 IN reactions were incubated at 37 °C for 15 min while reactions were stopped at 60 min for M-MuLV IN because of its lower specific activity. Reactions were linear at the time points examined. In HIV-1 IN time course studies, reactions were scaled to 30 μL, and 3 μL aliquots were removed at specific time intervals. Reactions were terminated by adding an equal volume of loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.5% xylene cyanol). Reaction products were separated on 20% polyacrylamide denaturing gels, subjected to autoradiography or Phosphorimager screens (Molecular Dynamics). Products were quantitated from phosphorimaging using ImageQuant software (Molecular Dynamics). The amount of product or rate of enzymatic activity was calculated from a minimum of three separate trials for each experiment performed in triplicate. For HIV-1 reactions, the rate of 3′-processing was determined by linear regression.

RESULTS

Design of Modified U5 Substrates. Our previous cross-comparative analysis of the substrate specificity of several integrases found the 3′-processing reaction to be more dependent on the U5 terminus sequence for catalysis as compared to strand transfer and disintegration reactions (*16*). In this study, we continue our cross-comparative analysis to test the hypothesis that conserved features of viral end recognition exist among INs. The three retroviral INs explored herein, HTLV-2, M-MuLV, and HIV-1, each represent a distinct genus within the *Retroviridae* (*25*). In the context of the 3′-processing reaction, we previously noted that position 5 (Figure 1A) confers significant substrate specificity in HTLV-2 and M-MuLV reactions (*16*). While the 5 position confers substrate specificity, this position is probably recognized in conjunction with the conserved CA, which is also required for optimal catalysis. Therefore, studies herein were focused on the 3′-processing reaction to probe the specific nucleotide functional groups involved in conferring substrate specificity and recognition of the CA for HTLV-2, M-MuLV, and HIV-1 INs.

Specifically, nucleotide analogues were substituted at positions 3, 4, and 5 in the plus (E strand) and minus (A strand) strands of HTLV-2, M-MuLV, and HIV-1 U5 substrates (Figure 1A) and examined for 3′-processing activity (Figure 1B) with their respective IN. A comparison of Watson−Crick base-pairing for guanine-cytosine and adenine-thymine (Figure 2A) with the analogues (Figure 2B) indicates the effect on possible hydrogen bonding and van der Waal contacts with amino acids in the major and minor grooves of DNA. Analogue substitutions have been used extensively in the study of interaction of protein−DNA as well as enzyme−DNA interactions to explore both geometric and electronic interfaces (*26*). For example, substitution of adenine with N^6-methyladenine (6-MeAde) blocks the hydrogen-bonding capability in the major groove at W_2 with the introduction of a methyl group (Figure 2A) (*27*). Replacement of guanine with O^6-methylguanine (6-MeGua) (*28, 29*), adenine with 7-deazaadenine (7-deazaAde), thymine with O^4-methylthymine (O^4-MeThy) (*30*), and cytosine with

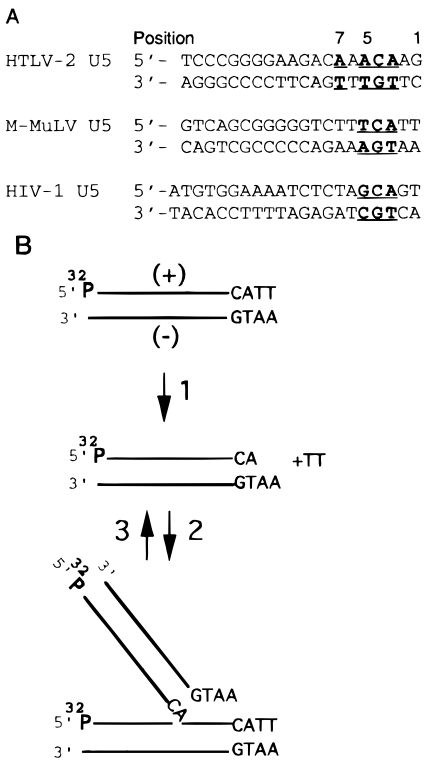

FIGURE 1: (A) U5 end sequences of HIV-1, HTLV-2, and M-MuLV substrates. The positions substituted with analogues or mutation are in boldface and underlined. The numbering system used for nucleotide identification is indicated above the sequence. The top strand in each U5 end represents the plus strand, and the bottom strand represents the minus strand. In reactions with HIV-1, position 5 in the minus strand was a Thy when Ade or analogues of Ade were employed. (B) Schematic representation of the enzymatic activities catalyzed by the retroviral IN in vitro for a single U5 end: step 1,3′-processing; step 2, strand transfer; and step 3, disintegration. Symbols: (+) plus strand; (−) minus strand.

Table 1: Nucleotide Analogues Used in This Study

nucleotide analogue	substitution	primary effect[a]
2,6-diaminopurine	adenine	minor groove at S_2
7-deazaadenine	adenine	major groove at W_1
N^6-methyladenine	adenine	major groove at W_2
5-methylcytosine	cytosine	major groove at W_1'
inosine	guanine	minor groove at S_2
O^6-methylguanine	guanine	major groove at W_2
O^4-methylthymine	thymine	major groove at W_2'

[a] Refers to position of possible interaction in minor (S) and major (W) grooves; also refer in Figure 2A for W and S position designations.

5-methylcytosine (5-MeCyt) (*27*) introduces modifications into the major groove (W_1, W_1', W_2, and W_2') (Table 1). In several of the analogues, the substitution of a hydrogen with a CH_3 group may also create a steric effect by blocking the ability of IN to interact with neighboring bases. Substitution of adenine with 2,6-diaminopurine [2,6-$(NH_2)_2$Pur] (*27*, *31*) or guanine with inosine (I) (*31*, *32*) introduces modifications into the minor groove (S_2) (Table 1). Since nucleotide analogues are not available to probe interactions in the minor groove at S_1 or S_1', we were not able to explore hydrogen-bonding interactions at this site.

Functional Group Determinants in the Interaction of HTLV-2 IN and the U5 Terminus. Previously, we reported that HTLV-2 IN shows a marked decrease in 3′-processing activity levels with substitution of adenine with thymine at position 5 in the HTLV-2 U5 substrate (*16*). Position 5 was also shown to confer substrate recognition of the M-MuLV LTR by HTLV-2 IN upon replacement of the wild-type thymine to an adenine. These results support position 5 as important for conferring substrate recognition of the U5 terminus by HTLV-2 IN and a strong requirement for adenine in this position. Similar experiments were also made with position 7; however, this position did not confer substrate recognition. Positions 5 and 7 were analyzed to probe the functional group determinants of substrate recognition; position 7 was included as an internal control.

Three base analogue substitutions were made at position 5 in the plus strand, 6-MeAde, 7-deazaAde, and 2,6-$(NH_2)_2$-Pur (Figure 1A). These analogues effect hydrogen-bonding interactions at W_2, W_1, and S_2, respectively. Neither the 6-MeAde (W_2) nor the 7-deazaAde (W_1) substrates showed any significant decrease in 3′-processing as compared to the WT substrate (Figure 3A). However, the replacement of adenine with 2,6-$(NH_2)_2$Pur, which introduced an additional amino group at S_2, substantially decreased the 3′-processing activity by 6.3-fold. The introduction of a highly polar amino group at the S_2 site adds both steric hindrance and hydrogen-bonding potential to the local environment. In the minus strand, an O^4-MeThy (W_2') substitution for thymine did not show any significant change in activity as compared to the activity noted with the WT substrate (Figure 3B). W_2' is the only site in the major groove of Thy for hydrogen bonding. Hydrogen bonding in the minor groove of the minus strand was not explored due to lack of availability of base analogues for S_1. Similar base analogue substitutions were made at position 7 in the plus strand [6-MeAde, 7-deazaAde, and 2,6-$(NH_2)_2$Pur] and in the minus strand (O^4-MeThy) (Figure 3A,B). No effect on activity was observed with any of the analogue substitutions at position 7. This result confirmed our previous observations in which we saw no effect on catalysis by HTLV-2 IN when nucleotides were substituted at this position (*16*).

Three different nucleotide analogue substitutions were made in the plus or minus strand at position 4 of the HTLV-2 U5 substrate (Figure 1A). Modification of position 4 of the plus strand with a 5-MeCyt substitution resulted in a nearly 2-fold increase in 3′-processing activity as compared to the wild-type LTR (Figure 3A). 5-MeCyt interferes with hydrogen bonding and the steric environment at W_1' by the introduction of a methyl group, and, thus, we were surprised to observe the slight increase in activity. In the minus strand, the replacement of guanine with inosine showed a 4-fold reduction in activity, while replacement of guanine with 6-MeGua showed a 5.6-fold decrease (Figure 3B). The introduction of inosine for guanine creates a deletion of the amino functional group in the minor groove at S_2. 6-MeGua introduces a methyl group in the major groove at W_2.

At position 3 (plus strand), adenine was replaced with 2,6-$(NH_2)_2$Pur, 7-deazaAde, or 6-MeAde (Figure 3A). The introduction of the 2,6-$(NH_2)_2$Pur analogue drastically reduced HTLV-2 IN 3′-processing activity to nearly undetectable levels. Substitution with 7-deazaAde did not decrease 3′-processing activity in HTLV-2 IN, while the 6-MeAde substitution showed a 3.3-fold increase in the level of activity. Substitution of O^4-MeThy for thymine in the minus strand of the HTLV-2 U5 substrate showed a 3.2-fold increase in the level of 3′-processing (Figure 3B).

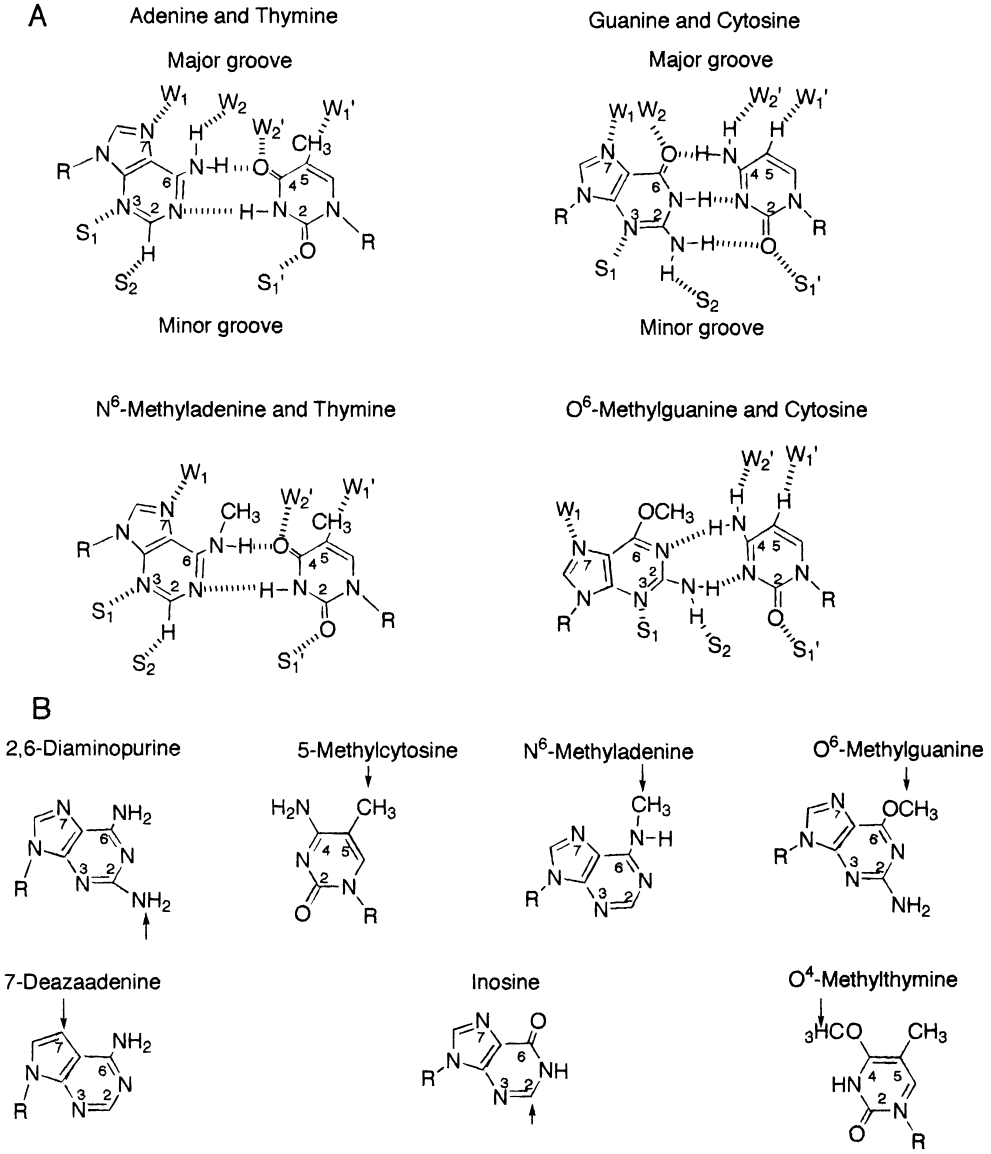

FIGURE 2: (A) Watson and Crick AT and GC base pairs show possible recognition sites in the major (W) and minor (S) grooves [Adapted from (*41*)]. Possible base-pairing schemes are shown for two of the analogues used in this study. (B) Chemical structures of the isosteric nucleotide analogues used in this work. Arrows indicate the site of modification of the nucleotide.

In summary, in the conserved CA/GT, analogues were noted to effect catalysis at adenine at S_2 in the minor groove, while for guanine both major groove (W_2) and minor groove hydrogen-bonding contacts were identified (S_2). A requirement for a hydrogen bond was not observed at W_1 in adenine. The introduction of an additional methyl group in position 3 at W_2 on the plus strand or on the minus strand at W_2', and in position 4 at W_1', increased 3′-processing activity.

Functional Group Determinants in the Interaction of M-MuLV IN and Its U5 Terminus. Previously, we observed a higher level of 3′-processing activity for M-MuLV IN with modified U5 substrates as compared to the M-MuLV wild-type substrate (*16*). These experiments suggest that the wild-type U5 terminus is not an optimal substrate for M-MuLV IN. In particular, substitution of an adenine for thymine at position 5 (5T→A substrate) increases activity substantially. The introduction of adenine for thymine at position 5 in the M-MuLV U5 substrate creates a disturbance in the possible hydrogen bonding and van der Waals contacts in both the major and minor grooves of both plus and minus strands. In the following studies, modifications of the M-MuLV U5 substrate in position 5 were based on adenine analogues substituted for thymine at position 5 to explore the contribution of the plus strand position (Figure 1A). In the minus strand, thymine was replaced with O^4-MeThy at position 5.

The wild-type LTR and the 5T→A U5 substrate were included in our study of position 5 to compare the possible loss or gain of activity during 3′-processing. As observed previously, replacement of thymine with adenine at position 5 (5T→A) increased levels of 3′-processing as compared with the M-MuLV WT substrate by 2.2-fold (Figure 4A). Substitution of position 5 with 2,6-$(NH_2)_2$Pur (S_2), 7-deazaAde (W_1), 6-MeAde (W_2), or inosine (W_2) in the plus strand and O^4-MeThy (W_2') in the minus strand all resulted in a decreased amount of activity as compared to the 5T→A substrate (Figure 4A,B). Remarkably, the levels of activity of 6-MeAde, inosine, and O^4-MeThy base analogue substrates were equivalent to the amount of activity noted with the WT substrate. Substrates with 7-deazaAde or 2,6-$(NH_2)_2$-Pur substitutions showed an additional decrease in activity, which was lower than the WT substrate. These results suggest that important hydrogen-bonding interactions with adenine

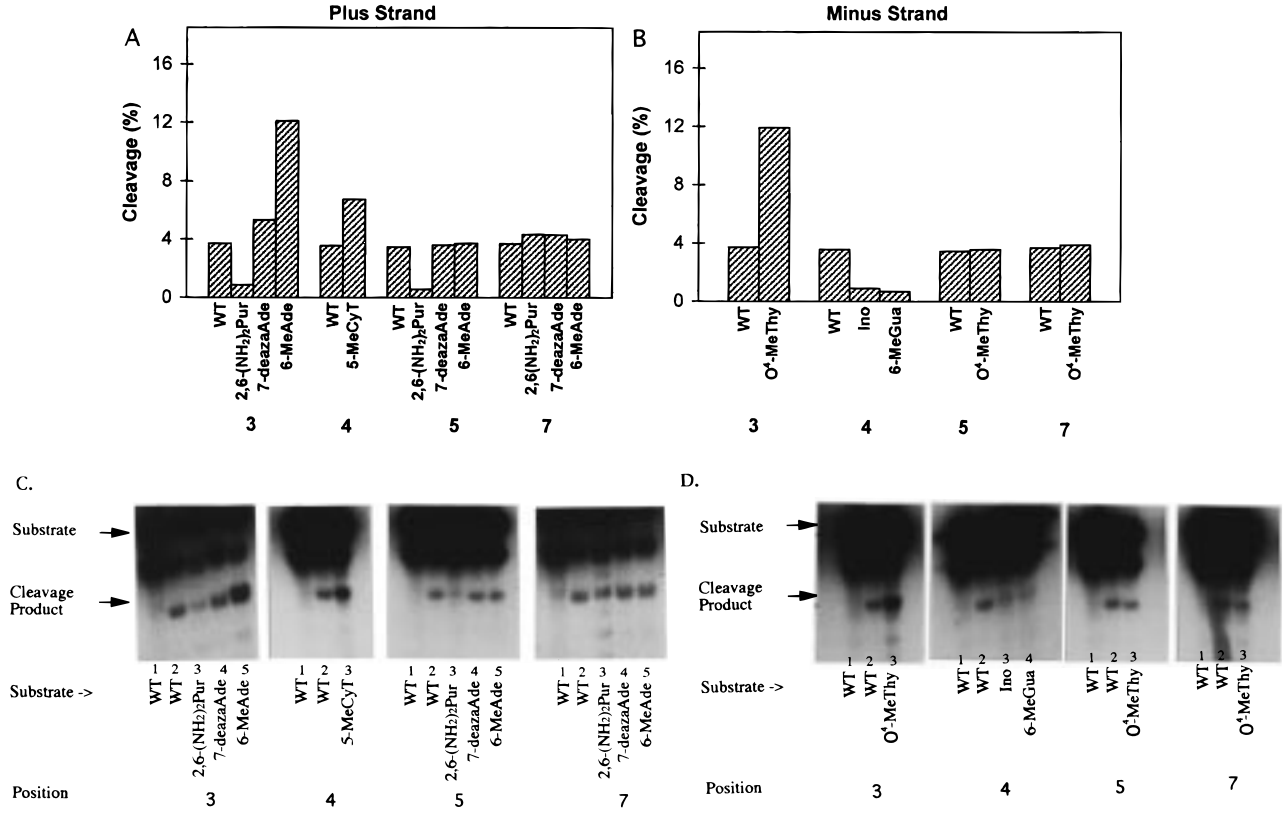

FIGURE 3: HTLV-2 IN 3′-processing activity with HTLV-2 WT and modified U5 substrates with nucleobase modifications in the plus (A) or minus (B) strands. Numbers at the bottom of each panel indicate the position of substitution in the U5 substrate. All reactions were performed in triplicate, and percent cleavage was determined from PhosphorImager analysis as described under Materials and Methods. Representative autoradiographs of HTLV-2 IN 3′-processing activity with HTLV-2 WT and modified U5 substrates with nucleobase modifications are presented for the plus (C) and minus (D) strands.

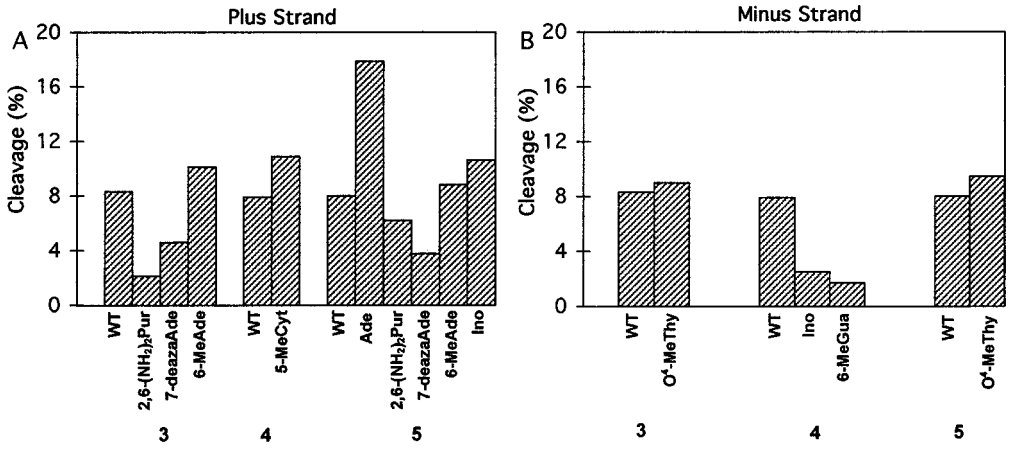

FIGURE 4: M-MuLV IN 3′-processing activity with M-MuLV WT and modified U5 substrates with nucleobase modifications in the plus (A) or minus (B) strands. Numbers at the bottom of each panel indicate the position of substitution in the U5 end. All reactions were performed in triplicate, and percent cleavage was determined from PhosphorImager analysis as described under Materials and Methods.

occur at W_1 (7-deazaAde) and W_2 (6-MeAde and inosine). At W_2, the inosine analogue substitutions allow for interaction as hydrogen bond acceptors. On the minus strand, the O^4-MeThy substrate was decreased in activity as compared to the 5T→A substrate, but was similar to the level of activity of the WT LTR. This analogue disturbs the hydrogen bonding acceptor capability of the oxygen and, thus, suggests significant contributions of this contact as well.

The conserved CA/GT was also examined for M-MuLV with five different base analogues. At position 4, the Cyt to 5-MeCyt substitution in the plus strand showed relatively no change in activity as compared to the WT substrate (Figure 4A). In contrast, at position 4 in the minus strand a substantial loss in activity occurred with substrate containing inosine (S_2) or 6-MeGua (W_2) for guanine (Figure 4B). Adenine was replaced with 2,6-$(NH_2)_2$Pur (S_2) or 7-deazaAde (W_1) at position 3 in the plus strand (Figure 4A), and thymine was replaced by O^4-MeThy (W_2') on the minus strand (Figure 4B). The 6-MeAde substitution increased only slightly, 1.23-fold, as compared to the wild-type substrate. The substrate with the 7-deazaAde incorporated into position 3 showed a 1.8-fold decrease in the level of 3′-processing. Similar to the HTLV-2 reactions, the largest effect on M-MuLV IN 3′-processing activity was with the 2,6-$(NH_2)_2$Pur substrate.

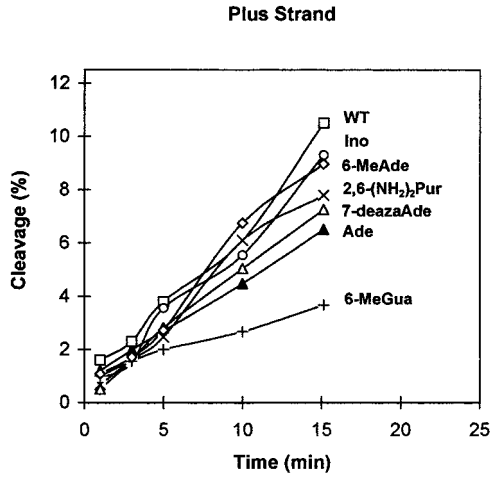

FIGURE 5: Time course of HIV-1 IN 3′-processing reaction with HIV-1 WT substrate and U5 substrates modified in position 5 in the plus (A) or minus (B) strands. Aliquots were removed at 1, 3, 5, 10, and 15 min. The *x* axis represents time in minutes, and the *y* axis represents the percent of substrate converted to product. All reactions were performed in triplicate, and percent cleavage was determined from PhosphorImager analysis as described under Materials and Methods.

This substitution decreased activity 4-fold as compared to the WT substrate. In the minus strand, the modification produced by substitution of O^4-MeThy at position 3 had no effect on activity. This observation is in stark contrast to the increase in activity obtained for HTLV-2 IN (Figure 3B).

Functional Group Determinants in the Interaction of HIV-1 IN and Its U5 Terminus. Previous work by Sherman et al. shows the substitution of G to A at position 5 had little effect on the 3′-processing activity of HIV-1 IN (*18*). This information was useful in our analysis as more nucleotide analogues are available based on adenine. Therefore, 6-MeAde, 2,6-$(NH_2)_2$Pur, 7-deazaAde, or adenine replaced the guanine at position 5 in the plus strand of the WT HIV-1 U5 substrate (Figure 1A). In addition, the guanine at position 5 was substituted with inosine and 6-MeGua (Figure 1A). Our previous studies in substrate recognition with HIV-1 IN showed a much lower specificity as compared to HTLV-2 and M-MuLV INs. Therefore, the rate of 3′-processing activity was examined for each substitution with HIV-1 IN to obtain greater sensitivity in our analysis (Figure 5). Similar to the observations reported by others previously (*18*), a slight decrease (1.7-fold) in the rate was observed with the G to A substitution. The substrate containing the substitution of 6-MeAde had activity similar to the WT substrate. Introduction of 6-MeAde for guanine in this position will perturb both W_2 and S_2 hydrogen bonding interactions. The G to I substitution, which perturbs S_2 interactions, also showed activity comparable to the WT LTR. A moderate effect on HIV-1 IN activity was noted with the 2,6-$(NH_2)_2$Pur and 7-deazaAde modifications. The 7-deazaAde had a greater effect, but neither analogue reduced the rate of 3′-processing activity by greater than 30%. Only the substitution of 6-MeGua showed a major effect on the HIV-1 IN 3′-processing activity as compared with WT HIV-1 substrate. This substitution resulted in a 3.7-fold decrease in the rate of dinucleotide cleavage. A comparison of the effect of 6-MeAde and 6-MeGua in hydrogen bonding shows the addition of a methyl group to the carbonyl of guanine (6-MeGua) creates a very poor hydrogen bond acceptor. The unpaired electrons on the nitrogen of 6-MeAde, however, are still able to interact with hydrogen bond donors. In summary, only the major groove site, W_2, was suggested to be involved in hydrogen bonding with a hydrogen bond donor of the HIV-1 IN.

The 3′-processing activity of HIV-1 IN was followed with three different nucleotide analogue substitutions in the plus (Figure 6A) or minus strand (Figure 6B) at position 4 in the HIV-1 U5 substrate (Figure 2B). In the plus strand, substitution of cytosine with 5-MeCyt reduced the 3′-processing reaction rate 2-fold. Both substitutions in the minus strand, guanine to inosine or guanine to 6-MeGua, greatly decreased HIV-1 IN activity. Substrates with inosine had a 16-fold reduction in rate, while substrates with 6-MeGua showed a 25-fold decrease in rate. The latter nucleotide analogue substitutions suggest strong hydrogen-bonding interactions in the major groove (W_2) and minor groove (S_2) at position 4 in the minus strand.

Four analogues, 6-MeAde, 7-deazaAde, 2,6-$(NH_2)_2$Pur, and O^4-MeThy, were used to probe the major and minor grooves in position 3 (Figure 7A,B). In the plus strand, the 6-MeAde (W_2) substitution resulted in a similar rate of product as compared to the wild-type substrate. The 7-deazaAde (W_1) substrate showed a slight reduction in rate, 3-fold. Substrates containing 2,6-$(NH_2)_2$Pur (S_2) were greatly decreased in their rate of 3′-processing, showing a 10-fold decrease. Therefore, in the plus strand of position 3, major groove H-bonding contacts are suggested at W_1, although a greater impact on 3′-processing occurred by changing the minor groove hydrogen-bonding contacts at S_2 from acceptor to donor. On the minus strand at position 3, we made a single substitution of thymidine with O^4-MeThy. Unexpectedly, this substitution resulted in a 3-fold increase in activity as compared to the wild-type substrate (Figure 7B). O^4-MeThy results in blocking the hydrogen-bonding acceptor capabilities of the oxygen at position W_2′ as well as steric hindrance through the introduction of a methyl group. This result suggests the methyl group increased hydrophobic interactions or reduced base-pairing at thymidine in the conserved CA/GT to the benefit of the IN−U5 end interaction. A reduction in the normal Watson−Crick base-pairing may allow for easier accessibility of IN to the scissile bond due to an increase in the opening of this site.

DISCUSSION

Several nucleotide analogues were used to map specific functional group contacts at the U5 terminus in positions known to be critical for IN-mediated catalysis. The nucleotide analogues described here have been used extensively to explore the specificity of DNA binding of proteins such as structure-specific nuclease (*31*), restriction enzymes (*30*), and repressor proteins (*33*). The mode by which IN makes contact with the viral termini is unknown, and, therefore, we tested the hypothesis that IN interacts with the major and minor grooves. Conservative nucleotide analogue substitutions were made at three critical positions in the U5 terminus, positions 4 and 3 (the conserved CA) and position 5.

A distinct milieu of functional group determinants were observed for each IN at position 5 in the viral LTR. Figure 8 summarizes the results for each substrate. Major groove interactions were not significant in the HTLV-2 reactions at

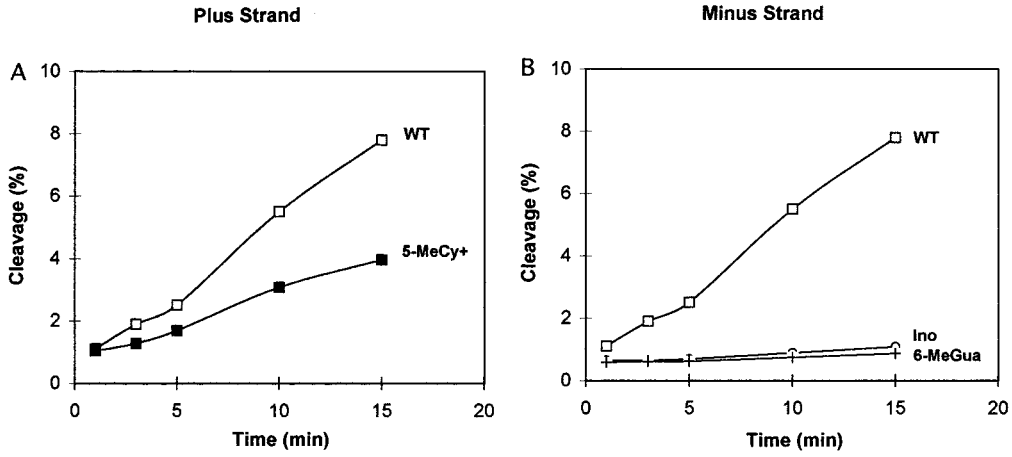

FIGURE 6: Time course of HIV-1 IN 3′-processing reaction with HIV-1 WT and modified U5 substrates in position 4 in the plus (A) or minus (B) strands. All reactions were performed in triplicate, and percent cleavage was determined from PhosphorImager analysis as described under Materials and Methods.

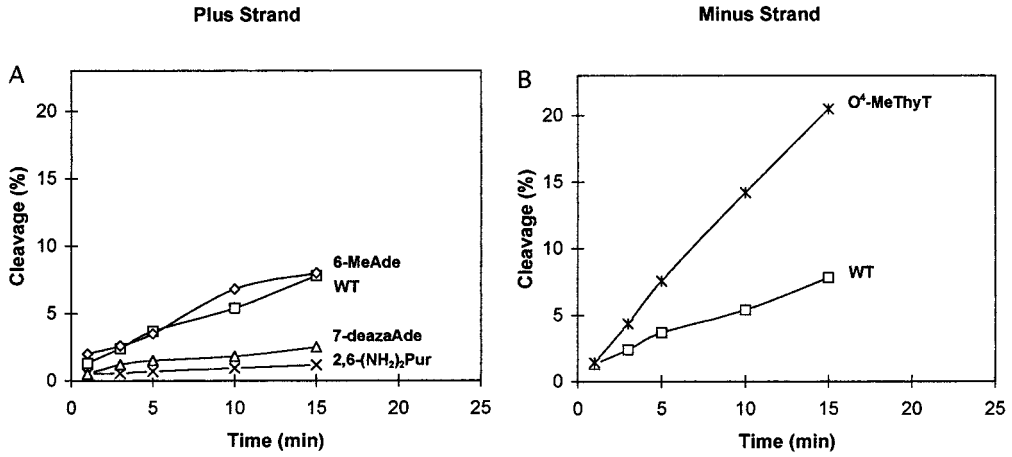

FIGURE 7: Time course of HIV-1 IN 3′-processing reaction with HIV-1 WT and modified U5 substrates in position 3 in the plus (A) or minus (B) strands. All reactions were performed in triplicate, and percent cleavage was determined from PhosphorImager analysis as described under Materials and Methods.

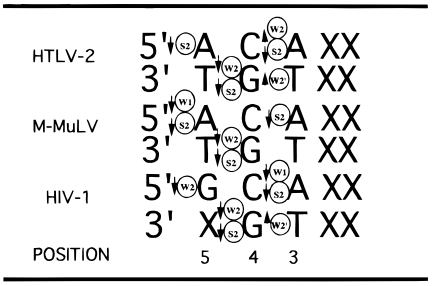

FIGURE 8: Summary of the critical sites of IN−U5 terminus interactions in major (W) and minor (S) grooves. Arrows indicate whether an increase or decrease in 3′-processing was observed.

position 5. In contrast, HIV-1 and M-MuLV IN reactions required hydrogen-bonding interactions within the major groove for optimal enzymatic activity. However, in the case of M-MuLV, our findings are based on adenine at position 5. In HTLV-2 IN−U5 terminus interactions, the introduction of an amino group into the minor groove (S_2) decreased 3′-processing activity. The apparent lack of hydrogen-bonding interactions was unexpected given our previous observations of the strong requirement of HTLV-2 IN for adenine in this position. The minor groove only permits discrimination between AT and GC base pairs. Previously, we noted that substitution of position 5 with Thy greatly reduces 3′-processing activity. These combined observations suggest that the local DNA conformation at this position confers substrate recognition. We knew from previous work that the native U5 substrate is not optimal for catalysis and the M-MuLV 5T→A substrate supports a higher integration activity of the M-MuLV IN in vitro (*16*). Thus, we anticipated that replacing the adenine residue with nucleotide analogues may decrease 3′-processing activity. As expected, all the modifications based on adenine in position 5 in major (W_1, W_2) and minor grooves (S_1) of the M-MuLV U5 terminus resulted in decreased M-MuLV IN activity. The adenine substitution may permit hydrogen bonding in the major groove with neighboring amino acids. However, it is difficult to determine the influence of local conformation at this site within the context of these studies. HIV-1 IN tolerated most modifications in the major and minor groove in the HIV-1 U5 terminus. Only the 6-MeGua substitution was observed to have a dramatic decrease on catalysis. Comparison of the activity displayed by 6-MeAde and 6-MeGua reveals the importance of the carbonyl oxygen of guanine as a hydrogen acceptor in the U5 end−IN interactions. Similar steric effects are introduced by 6-MeAde and 6-MeGua in the major groove, and thus cannot be considered to play a role in the observed decreased activity. Further, no effect was noted with the introduction of an amino group at

S_2. The flexibility of position 5—with regard to the number of substitutions which can be made without effecting activity—supports our previous finding of the relaxed substrate specificity displayed by HIV-1 IN as compared to HTLV-1, HTLV-2, and M-MuLV INs (*16*).

Base analogue modifications of the conserved CA/ GT, particularly at the guanine and adenine residues, revealed a common mechanism in the interaction of the three INs. In all three, the modifications in the minor groove were more critical to activity than those in the major groove. Substitution of guanine with 6-MeGua or inosine, which modifies the major groove (W_2) or minor groove (S_2), respectively, decreased 3′-processing activity. On the minus strand, however, we did not note any significant effect in a cytosine to 5-MeCyt substitution in position 4, which modifies the major groove at W_1. This modification resulted in a moderate decrease in the 3′-processing activity of HIV-1 IN, and a slight increase with HTLV-2 and M-MuLV INs.

Base analogue substitutions at position 3 in the plus strand with 6-methyl-A (W_2) in HIV-1, HTLV-2, and M-MuLV U5 termini also showed an effect on the activity of the three INs. Interestingly, the HTLV-2 IN reactions showed an enhancement in 3′-processing with the introduction of a methyl group at W_2, while HIV-1 and M-MuLV INs showed a decrease in activity. A drastic decrease in 3′-processing activity was observed with the adenine to 2,6-$(NH_2)_2$Pur (S_2) substitution in the three U5 termini at position 3 with all three INs. With the O^4-MeThy substitution at position 3 on the minus strand, we observed an increase in the 3′-processing activity of HIV-1 and HTLV-2 INs, but not with M-MuLV IN. The O^4-MeThy, which creates more of a hydrophobic surface in the major groove, may stabilize the IN−U5 terminus complex by enhancement of van der Waal contacts between the protein and the viral DNA. Alternatively, the methyl group may act to destabilize Watson−Crick base-pairing and in doing so create a more favorable interaction with the adjacent scissile bond. Destabilization of the Watson−Crick base-pairing in the three terminal positions by introduction of mismatched base pairs has been previously shown to enhance end processing (*34*−*36*). Therefore, it is possible that the introduction of an amino group into the minor groove at position 3 may prevent distortion or bending of the DNA and interfere with the creation of a requisite conformation of the DNA required for optimal interaction of IN for additional contacts with the viral DNA end. It has been suggested that Lys 159 in HIV-1 IN may interact with the N7 position (W_1) of the conserved deoxyadenosine adjacent to the scissile phosphodiester bond of the viral DNA (*37*). This residue is conserved among retroviral INs. In our analysis, the substitution of 7-deazaAde in the HIV-1 U5 terminus showed a 3-fold decrease in 3′-processing, which is consistent with this hypothesis. M-MuLV IN was also effected by this substitution; however, HTLV-2 IN was not. This suggests a different orientation of the HTLV-2 U5 terminus in the catalytic site with regard to its Lys.

In summary, conserved as well as unique modes of functional group interaction were observed among the three INs. We speculate that these conserved sites reflect a common mode of interaction with the U5 viral terminus in the active site of IN. Our analysis using base analogues to probe the functional group determinants in the U5 end suggests very few hydrogen-bonding interactions in the major groove are essential for substrate recognition or substrate processing. Further, the major interactions appear limited to three positions in the plus and minus strands. This agrees with previous ethylation interference experiments which defined critical contacts with HIV-1 IN and its termini in positions 3, 4, and 5 (*38*). Future kinetic analysis of the critical functional groups defined herein will ascertain their contribution to the role in binding or catalysis by IN.

In light of our studies, we were led to consider that the mode of U5 end recognition constitutes an indirect readout of the B-DNA by IN rather than a direct readout. The direct readout mechanism has been demonstrated for a number of sequence-specific protein−DNA complexes and is characterized as the interaction of a protein with specific functional groups in the nucleobases (*39*). In the case of indirect readout, the functional groups present in the nucleobases provide the overall geometry of the recognition site, and the major direct contacts are formed from interactions of the phosphodiester residues of the DNA and amino acid side chains (*40*). The combined observations of our laboratory and others suggest a model whereby IN may initially recognize the viral and host DNA through an indirect readout mechanism. However, in the case of the viral end, continuation to 3′-processing may require that the appropriate functional groups are available to form a transition state via specific hydrogen-bonding of the nucleobases. The ability to form a transition state for catalysis would be absent in a nonviral substrate. Entry into the transition state may require significant distortion of the viral DNA ends at base pairs 1, 2, and 3 (*36*). Thus, additional recognition of the U5 substrate may occur during catalysis. In studies with disintegration substrates, numerous changes in the viral DNA nucleotides can be made with little impact on catalysis of the substrate (*21*−*23*). This also suggests that IN may rely on a particular geometrical shape for recognition and, therefore, the involvement of an indirect readout mechanism.

Over the past few years, integrase has become a highly sought target for directing antivirals to treat individuals with AIDS. Our investigations will hopefully provide insight for those concerned with designing effective inhibitors of the retroviral integrase. Given that IN interacts with both major and minor grooves of the U5 end, competitive suicide inhibitor analogues of the conserved CA or GT may be useful in blocking IN activity.

ACKNOWLEDGMENT

We thank Marcus Bader for technical assistance and Jeffrey Arterburn for critical reading of the manuscript.

REFERENCES

1. Katz, R. A., and Skalka, A. M. (1994) *Annu. Rev. Biochem. 63*, 133−173.
2. Whitcomb, J. M., and Hughes, S. H. (1992) *Annu. Rev. Cell Biol. 8*, 275−306.
3. Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell 62*, 829−837.
4. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990) *Cell 63*, 87−95.
5. Acel, A., Udashkin, B. E., Wainberg, M. A., and Faust, E. A. (1998) *J. Virol. 72*, 2062−2071.
6. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) *Science 255*, 723−726.

7. Balakrishnan, M., Zastrow, D., and Jonsson, C. B. (1996) *Virology 219*, 77−86.
8. Jonsson, C. B., Donzella, G. A., and Roth, M. J. (1993) *J. Biol. Chem. 268*, 1462−1469.
9. Pemberton, I. K., Buckle, M., and Buc, H. (1996) *J. Biol. Chem. 271*, 1498−1506.
10. van Gent, D. C., Elgersma, Y., Bolk, M. W. J., Vink, C., and Plasterk, R. H. A. (1991) *Nucleic Acids Res. 19*, 3821−3827.
11. Khan, E., Mack, J. P. G., Katz, R. A., Kulkosky, J., and Skalka, A. M. (1991) *Nucleic Acids Res. 19*, 851−860.
12. Leavitt, A. D., Rose, R. B., and Varmus, H. E. (1992) *J. Virol. 66*, 2359−2368.
13. Reicin, A. S., Kalpana, G., Paik, S., Marmon, S., and Goff, S. (1995) *J. Virol. 69*, 5904−5907.
14. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol. 63*, 5319−5327.
15. Bushman, F. D., and Craigie, R. (1990) *J. Virol. 64*, 5645−5648.
16. Balakrishnan, M., and Jonsson, C. B. (1997) *J. Virol. 71*, 1025−1035.
17. Colicelli, J., and Goff, S. P. (1988) *J. Mol. Biol. 199*, 47−59.
18. Sherman, P. A., Dickson, M. L., and Fyfe, J. A. (1992) *J. Virol. 66*, 3593−3601.
19. Roth, M. J., Schwartzberg, P. L., and Goff, S. P. (1989) *Cell 58*, 47−54.
20. Gerton, J. L., and Brown, P. O. (1997) *J. Biol. Chem. 272*, 25809−25815.
21. Chow, S., and Brown, P. O. (1994) *J. Virol. 68*, 3896−3907.
22. Donzella, G. A., Jonsson, C. B., and Roth, M. J. (1993) *J. Virol. 67*, 7077−7087.
23. Jonsson, C. B., and Roth, M. J. (1993) *J. Virol. 67*, 5562−5571.
24. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
25. Vogt, P. K. (1997) in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., Eds.) pp 1−26, Cold Spring Harbor Press, Plainview, NY.
26. Aiken, C. R., and Gumport, R. I. (1991) *Methods Enzymol. 208*, 433−457.
27. Brennan, C. A., van Cleve, M. D., and Gumport, R. I. (1986) *J. Biol. Chem. 261*, 7270−7278.
28. Spratt, T. E., and Levy, D. E. (1997) *Nucleic Acids Res. 25*, 3354−3361.
29. Voigt, J. M., and Topal, M. D. (1990) *Biochemistry 29*, 1632−1637.
30. Newman, P. C., Nwosu, V. U., Williams, D. M., Cosstick, R., Seela, F., and Connolly, B. (1990) *Biochemistry 29*, 9891−9901.
31. Wang, S., Cosstick, R., Gardner, J. F., and Gumport, R. I. (1995) *Biochemistry 34*, 13082−13090.
32. Duggan, L. J., Hill, T. M., Wu, S., Garrison, K., Zhang, X., and Gottlieb, P. A. (1995) *J. Biol. Chem. 270*, 28049−28054.
33. Mazzarelli, J. M., Rajur, S. B., Iadarola, P. L., and McLaughlin, L. W. (1992) *Biochemistry 31*, 5925−5936.
34. van den Ent, F. M. I., Vink, C., and Plasterk, R. H. A. (1994) *J. Virol. 68*, 7825−7832.
35. Mazumder, A., and Pommier, Y. (1995) *Nucleic Acids Res. 23*, 2865−2871.
36. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) *Genes Dev. 11*, 371−382.
37. Jenkins, T. M., Esposito, D., Engelman, A., and Craigie, R. (1997) *EMBO J. 16*, 6849−6859.
38. Bushman, F. D., and Craigie, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3458−3462.
39. Steitz, T. A. (1990) *Q. Rev. Biophys. 23*, 205−280.
40. Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature 335*, 321−329.
41. Branden, C., and Tooze, J. (1991) *Introduction to Protein Structure*, Garland Publishing, Inc., New York.

BI982124I