Inhibition of Moloney Murine Leukemia Virus Integration Using Polyamides Targeting the Long-Terminal Repeat Sequences†

Fan Yang,‡ Jason M. Belitsky,§ Rodrigo A. Villanueva,‡ Peter B. Dervan,§ and Monica J. Roth*,‡

Department of Biochemistry, University of Medicine and Dentistry of New Jersey−Robert Wood Johnson Medical School, 675 Hoes Lane, Piscataway, New Jersey 08854, and Department of Chemistry, California Institute of Technology, Pasadena, California 91125

Received January 31, 2003

ABSTRACT: The retroviral integrase (IN) carries out the integration of the viral DNA into the host genome. Both IN and the DNA sequences at the viral long-terminal repeat (LTR) are required for the integration function. In this report, a series of minor groove binding hairpin polyamides targeting sequences within terminal inverted repeats of the Moloney murine leukemia virus (M-MuLV) LTR were synthesized, and their effects on integration were analyzed. Using cell-free in vitro integration assays, polyamides targeting the conserved CA dinucleotide with cognate sites closest to the terminal base pairs were effective at blocking 3′ processing but not strand transfer. Polyamides which efficiently inhibited 3′ processing and strand transfer targeted the LTR sequences through position 9. Polyamides that inhibited integration were effective at nanomolar concentrations and showed subnanomolar affinity for their cognate LTR sites. These studies highlight the role of minor groove interactions within the LTR termini for retroviral integration.

The integrase (IN)[1] catalyzes the integration of viral DNA into the host genome. In addition to the integrase, the DNA sequences present in the long-terminal repeat (LTR) at the ends of the linear viral DNA are also required for integration (for review, see refs *1* and *2*). The integration process can be divided into two steps. First, IN cleaves two nucleotides from the 3′ ends of both LTR termini, exposing the conserved 5′-CA-3′ dinucleotide. Second, the 3′ processed ends of the viral DNA are joined to the host DNA in a coordinated fashion. In vitro assays have been developed to recapitulate both steps of integration by using purified integrase protein and short oligonucleotide duplexes mimicking the viral LTR (*3*−*5*). In addition, purified integrase is able to carry out an in vitro concerted two-end integration reaction using the LTR oligonucleotide as donor and plasmid DNA as target (*6*, *7*).

The molecular details of integration remain elusive. Integration is a complex and dynamic process, which despite intense efforts (*8*−*10*) has resisted direct structural characterization. There is evidence that different IN−LTR interactions are important for the individual steps of integration (*11*, *12*). In the absence of a crystal structure, DNA binding molecules may shed light on the molecular interactions of integrase with the viral LTR. Studies with modified LTRs containing nucleotide analogues indicate that the IN does contact the LTR minor groove (*13*). DNA minor groove binders such as distamycin dimers (*14*, *15*; for review, see refs *16* and *17*) have been analyzed as integration inhibitors that target the DNA substrate. The minor groove binder netropsin has been shown to inhibit Moloney murine leukemia virus (M-MuLV) IN at micromolar concentrations (*18*). Netropsin binds to A+T-rich sequences of 4−7 base pairs in length. The terminal seven nucleotides of the M-MuLV LTR (5′-AATGAAA) are A+T-rich and contain sequences especially favorable for netropsin binding (*18* and references cited therein). In this regard, it would be especially useful to place minor groove DNA binders at precise but incrementally different positions within the 13 base pair IN recognition sequence, 5′-AATGAAAGACCCC-3′.

Hairpin polyamides containing *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) carboxamides are small molecules that bind DNA in the minor groove with high affinity in a sequence-specific manner (for review see refs *19* and *20*). On the basis of pairing rules, polyamides can target a large number of predetermined DNA sequences and bind with a 1:1 stoichiometry with its target sequence. This level of target flexibility and precision, which allows unambiguous recognition of DNA sites with single base pair resolution, is not available to classical DNA binding agents. A series of polyamides that recognize distinct LTR sites with high affinity would be molecular probes for understanding the mechanism of integration. Polyamides have been shown to block transcription factors from binding to their cognate recognition sequences, thus inhibiting transcription (*21*−*23*). Polyamides targeted to TBP and Lef-1 binding sites in the HIV-1 promoter have been shown to inhibit viral replication in human cells (*21*). Polyamides are cell-permeable com-

† This work was supported by NIH Grants CA76545 to M.J.R. and GM57148 to P.B.D. We are grateful to the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B.

* To whom correspondence should be addressed. Phone: (732) 235-5048. Fax: (732) 235-4783. E-mail: roth@waksman.rutgers.edu.

‡ UMDNJ−RWJMS.

§ California Institute of Technology.

[1] Abbreviations: Im, *N*-methylimidazole-2-carboxamide; Py, *N*-methylpyrrole-2-carboxamide; Bz, benzoyl; *β*, *β*-alanine; *γ*, *γ*-aminobutyric acid; Dp, dimethylaminopropylamine; IN, integrase; M-MuLV, Moloney murine leukemia virus.

10.1021/bi034177s CCC: $25.00
Published on Web 05/02/2003

pounds which have proven effective in human T-cells (*21*) and live *Drosophila* (*24*, *25*) without exhibiting toxicity. Recent studies have shown that the intracellular location of fluorescently labeled polyamides varies between different cell lines (*26*). If polyamides can effectively target the LTR termini, they should act as integrase inhibitors in analogy to netropsin, which would represent a complementary attack on the retroviral life cycle.

In this study, the ability of a series of polyamides (**1**−**8**), including two mismatch controls, to inhibit the in vitro integration activities of M-MuLV IN was determined. Inhibitors effective at nanomolar concentrations fell into two classes, based on the position of their cognate sites within the M-MuLV LTR. The characterization of these first generation polyamide inhibitors is presented.

EXPERIMENTAL PROCEDURES

Materials. Crude [γ-^{32}P]ATP (7000 Ci/mmol) was purchased from ICN. T4 polynucleotide kinase, T4 DNA ligase, DNase I, and restriction enzymes were obtained from New England Biolabs. Ni^{2+}-nitrilotriacetic acid−agarose was purchased from Qiagen. Target plasmid pGEM-3Zf(+) was purchased from Promega.

Polyamides. All polyamides were synthesized by solid-phase methods as previously described (*27*−*29*) and characterized by analytical HPLC and MALDI-TOF mass spectrometry: ImPyPyPyIm-γ-PyPyPyPyPy-β-Dp (**1**) [M + H] 1466.99, 1466.57 calculated for [M + H]; ImPy-β-PyIm-(*R*)-$^{H_2N}\gamma$-PyPyPyPyPy-β-Dp (**2**) [M + H] 1430.94, 1430.67 calculated for [M + H]; ImPy-β-PyIm-(*R*)-$^{H_2N}\gamma$-PyPy-β-PyPy-β-PyPy-β-Dp (**3**) [M + H] 1695.08, 1694.79 calculated for [M + H]; ImPy-β-PyIm-(*R*)-$^{H_2N}\gamma$-PyImPyPyPy-β-Dp (**4**) [M + H] 1431.48, 1431.66 calculated for [M + H]; BzImPyPy-(*R*)-$^{H_2N}\gamma$-PyPyPyPy-β-Dp (**5**) [M + H] 1233.6, 1233.57 calculated for [M + H]; BzPyPyPy-(*R*)-$^{H_2N}\gamma$-PyImPyPy-β-Dp (**6**) [M + H] 1233.54, 1233.57 calculated for [M + H]; BzPyPyPy-(*R*)-$^{H_2N}\gamma$-PyPyImPy-β-Dp (**7**) [M + H] 1233.6, 1233.57 calculated for [M + H]; ImPyPyIm-(*R*)-$^{H_2N}\gamma$-PyPyImPy-β-Dp (**8**) [M + H] 1238.73, 1238.58 calculated for [M + H].

DNase I Footprinting. Quantitative DNase I footprinting reactions were performed as previously described (*30*) using a 183 base pair 5′-^{32}P-labeled DNA fragment containing one copy of the M-MuLV LTR terminus. The labeled fragment was generated by the PCR method (*22*, *31*) using primers 5′-AGACAGGATATCAGTGGTCCA and 5′-CATGCCTTGCAAAATGGCGTT, corresponding to positions 10968−10988 and 43−63, respectively, of the M-MuLV linear provirus-containing plasmid, NCA-C (*32*), which was used as the template. Primers were prepared and purified by the Caltech Biopolymer Synthesis Facility. The sequence of the labeled fragment is listed here with the terminal 13 base pairs of the M-MuLV LTR shown in bold: 5′-^{32}P-CATGCCTTGCAAAATGGCGTTACTTAAGCTAGCTAGCTTGCCAAACCTACAGGT**GGGGTCTTTCATT**CCCCCCTTTTTCTGGAGACTAAATAAAATCTTTTATTTTATCTATGGCTCGTACTATAGGCTTCAGCTGGTGATATTGTTGAGTCAAAATAGAGCCTGGACCACTGATATCCTGTCT.

Oligonucleotides. DNA oligonucleotides used as substrates were prepared by the University of Medicine and Dentistry of New Jersey Biochemistry Department DNA Synthesis Facility and purified by electrophoresis on 20% denaturing polyacrylamide gels. Oligonucleotides used in this study are referred to by their synthesis numbers and were labeled with [γ-^{32}P]ATP by kinase reaction as previously described (*33*). Oligonucleotides 2783 (5′-GTCAGCGGGGGTCTTTCATT) and its complementary strand 2785 (5′-AATGAAAGACCCCCGCTGAC) were used for 3′ processing assay. Oligonucleotides 2784 (5′-GTCAGCGGGGGTCTTTCA) and its complementary strand 2785 were used for strand transfer and concerted two-end integration assays.

Purification of M-MuLV Integrase. Recombinant M-MuLV integrase (WT IN) containing a hexahistidine tag were expressed in *Escherichia coli* BL21(DE3) (Novagen) and purified by Ni^{2+}-nitrilotriacetate−agarose chromatography (Qiagen) as previously described (*33*).

In Vitro Assays. Strand transfer and 3′ processing reactions were performed as previously described (*33*). The reaction buffer contains 20 mM morpholineethanesulfonic acid (MES, pH 6.2), 10 mM DTT, 10 mM $MnCl_2$, 10 mM KCl, and 10% glycerol. The condition for integration into an exogenous target was the same as that of the strand transfer reactions except that 200 mM KCl and 10% DMSO (dimethyl sulfoxide) were added. The LTR oligonucleotide (2783 or 2784) was 5′ labeled by T4 polynucleotide kinase and mixed with the complementary strand at a ratio of 1:2. The oligonucleotides were annealed by heating for 3 min at 95 °C and then cooling to room temperature. Typically, one reaction mixture (30 μL) contains 1 pmol of labeled LTR, 0.6 μg of target plasmid DNA, and 20 pmol of IN protein. After the preincubation of the LTR and the polyamides, the IN protein was added and incubated on ice for 5 min and at 37 °C for 5 min. The target DNA and KCl were then added. The reactions were incubated at 37 °C for 2 h, stopped by addition of 10 mM EDTA, pH 8.0, 0.5% SDS, and 100 μg/mL proteinase K, and incubated at 37 °C for 1 h. Ten microliters of the reaction was subjected to electrophoresis on a 1% agarose gel. After gel electrophoresis, the gel was dried and exposed to Kodak X-Omat Blue XB-1 film. In all reactions, the polyamides were mixed with the LTR substrate first and incubated at room temperature for 18 h before addition of the IN protein, unless indicated otherwise. Films were scanned and quantified with a densitometric plot using the Scion Image 1.62C program (based on NIH Image).

RESULTS

Polyamide Design. The termini of the MuLV LTRs including the conserved CA::GT dinucleotide, consist of a 13 base pair (bp) inverted repeat sequence (5′-AATGAAAGACCCC). The MuLV U5 and U3 termini sequences are therefore identical. This degree of conservation of the LTR termini is not found in HIV-1 and Rous sarcoma virus (RSV) genomes. This facilitates the generation of sequence-directed polyamide inhibitors which would simultaneously recognize both ends of the viral DNA. A series of polyamides which target the M-MuLV LTR were generated (Figure 1). The pairing rules are outlined below for hairpin polyamide binding to DNA (*19*). A pyrrole opposite an imidazole (Py/Im pairing) targets a C·G bp whereas an Im/Py pair binds a G·C bp. A Py/Py pair recognizes both A·T and T·A pairs. Internal β-alanines (β) may be introduced to improve both flexibility and specificity of the polyamides. β/β, β/Py, and

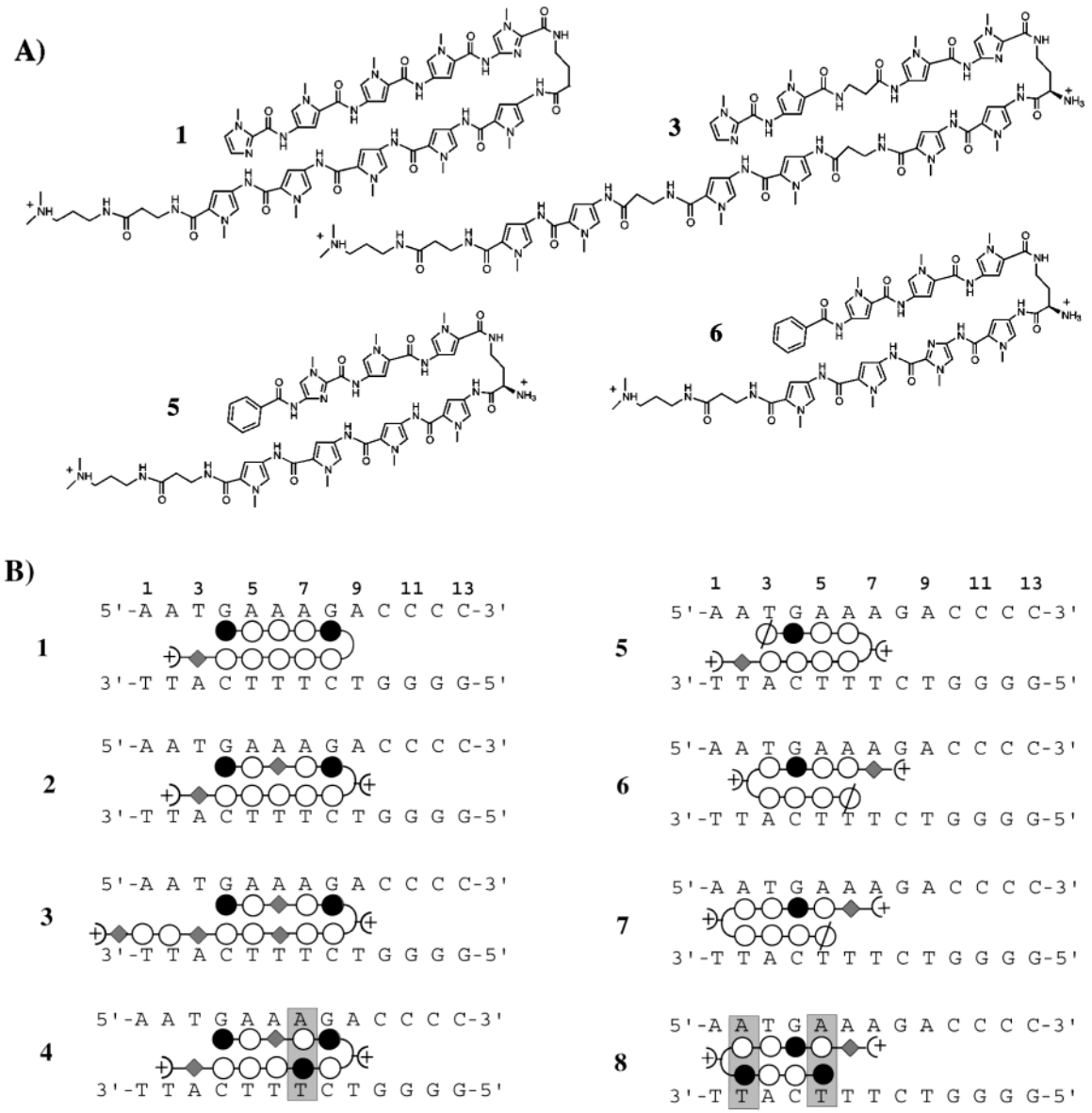


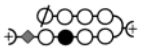
FIGURE 1: (A) Chemical structures of selected polyamides. (B) Ball and stick representations of the polyamides and their binding sites within the M-MuLV LTR. Numbering of the LTR is shown above the top entry. Mismatches are highlighted. The symbols for polyamides are as follows: open, filled, and slashed circles, Im, Py, and Bz rings, respectively; curved lines, γ-aminobutyric acid; diamonds, β-alanine; plus signs, dimethylaminopropylamine tail and amino group of (*R*)-2,4-diaminobutyric acid.

Py/β pairs recognize both A·T and T·A pairs. The "turn" and "tail" residues, γ-aminobutyric acid (γ), (*R*)-2,4-diaminobutyric acid [$(R)^{H_2N}\gamma$], unpaired β-alanine, and dimethylaminopropylamine (Dp), are also A·T and T·A specific. Recently, a benzoyl group (Bz) has been shown to be a high-affinity A·T- and T·A-specific N-terminal residue when paired with Py (*29*).

The most conserved feature of all retroviral termini is the CA/GT dinucleotide pair, invariably found precisely at the site of joining to the host DNA. Sequences internal to the CA, extending up to 15 bp from the termini, have a more limited effect on IN activity (*34*−*37*). All polyamides generated target the CA dinucleotide and vary in the extent of the surrounded sequences to which they bind (Figure 1B). Polyamides **1**−**3** were designed to bind a core sequence from positions 2−9 of the LTR. The central Py/Py pair of **1** was varied to a β/Py (**2**) or β/β pair (**3**). Polyamide **1** contains the standard γ-turn, whereas hairpins in **2** and **3** contain the charged chiral analogue, $(R)^{H_2N}\gamma$. These latter compounds are doubly charged. Polyamide **3**, which contains a C-terminal extension of β-Py-Py, is designed to bind a larger site, placing a Py ring at position 1 of the LTR. Polyamide **4** represents a single base pair mismatch for the 2−9 region. Another control compound, polyamide **8**, is a double base pair mismatch for the LTR sequence. The recent development of the A·T- and T·A-selective N-terminal Bz residue allowed for the targeting of additional sequences, which place the core of the hairpins closer to the terminus of the LTR (**5**−**7**). Polyamides **5** and **7** cover positions 1−7, and polyamide **6** covers positions 2−8. While polyamides **1**−**3** place an Im/Py pair and unpaired β-alanine against the CA dinucleotide, polyamides **5**−**7** place two central ring pairings against the CA dinucleotide, which may serve as a greater steric impediment to interaction with the integrase.

High-Affinity Binding of the Polyamides to the Viral LTR Sequence. Quantitative DNase I footprinting titration experiments (*31*) were performed to determine the equilibrium association constant (K_a) of the polyamides with the M-

Table 1: Equilibrium Association Constants for Polyamides with an Internal Copy of MuLV LTR[a]

	Polyamide	Sequence	Ka (M^{-1})
1		5'-GGGAATGAAAGACC-3'	2.2×10^9
2		5'-GGGAATGAAAGACC-3'	3.6×10^9
3		5'-GGGAATGAAAGACC-3'	2.7×10^9
4		5'-GGGAATGAAAGACC-3'	2.1×10^7
5		5'-GGGAATGAAAGACC-3'	4.0×10^{10}
6		5'-GGGAATGAAAGACC-3'	1.7×10^{10}
7		5'-GGGAATGAAAGACC-3'	3.0×10^9
8		5'-GGGAATGAAAGACC-3'	$\leq 5 \times 10^6$

[a] Values reported are the mean values obtained from three DNase I footprint titration experiments. The binding site for each polyamide is underlined. Mismatches are shown in bold. The assays were carried out at 22 °C in 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

MuLV LTR sequence (Table 1). Individual polyamides were incubated with a ^{32}P-labeled DNA fragment containing an internal copy of the LTR, followed by DNase I digestion. Note that this procedure requires an excess of polyamide, and the amount of DNA used in this procedure is on the order of 5 pM (*31*). Representative DNase I footprinting gels for polyamides **2** and **6** are shown in Figure 2. Polyamides **5**−**7**, which are constitutional isomers, varying only by the position of a single nitrogen atom within the molecule, are readily differentiated by the LTR sequence (Table 1). Among the polyamides tested, polyamide **5** has the highest equilibrium association constant ($K_a = 4 \times 10^{10}$ M^{-1}), while the double base pair mismatch **8** has the lowest binding constant ($K_a \leq 5 \times 10^6$ M^{-1}).

The similar affinity [$K_a = (1.9-3.6) \times 10^9$ M^{-1}] of polyamides **1**−**3** is of interest. Polyamide **1** appears to be particularly well matched for the LTR sequence, given its relatively high affinity, and observations that substitution with a central β/β pair and/or the positively charged $(R)^{H_2N}\gamma$-turn failed to increase its affinity (data not shown). Additionally, substitution with a central β/Py pair yielded a drop in affinity, which was recovered by substitution with $(R)^{H_2N}\gamma$ (polyamide **2**). With the internal copy of the LTR sequence used for these experiments, the tail of the extended polyamide **3** is placed against G·C base pairs, which represents a mismatch that is not present in the LTR itself.

It is important to note that, since the DNase I footprinting titrations were performed on an internal copy of the LTR sequence, the affinity constants generated are only a model for binding to the M-MuLV LTR termini. The binding of polyamides **5**−**7** was of particular interest, given the proximity of their cognate sites to the absolute end of the LTR. DNA melting experiments using oligonucleotides with terminal or internal cognate sites were performed to address this issue (see Supporting Information) (*38*). Comparisons of melting curves with and without polyamides **5**−**7** showed that the T_m shifts induced by these polyamides were similar for both sets of oligonucleotides, although slightly lower in each case (approximately 10%) for the terminal cognate site. In all cases, the T_m shifts for both sets of oligonucleotides with polyamides **5**−**7** were considerably greater than the T_m shifts induced by a mismatch polyamide. These results indicate that polyamides can bind with high affinity to the absolute ends of DNA fragments. However, since the T_m shift for the terminal cognate site was slightly lower than the internal cognate site for polyamides **5**−**7**, the affinity constants determined by DNase I footprinting for these compounds (Table 1) should be considered as upper limits for their affinity to the M-MuLV LTR termini.

Inhibition of the 3′ Processing. The ability of the polyamides to inhibit integration was tested in in vitro integration assays. Figure 3 schematically outlines the in vitro assays utilized in these studies. In the 3′ processing assay, a 20 base pair (bp) blunt-end oligonucleotide duplex mimicking the viral LTR end, was used as the substrate (Figure 3, top line). The strand containing the conserved CA dinucleotide was 5′ ^{32}P labeled. IN-dependent 3′ processing of the dinucleotide yields an 18 base product, which is monitored on a denaturing acrylamide gel. It is of note that the concentration of DNA substrate in these and the following integration assays is significantly greater (33 nM) than the concentration of DNA in the DNase I footprinting titrations. Figure 4 summarizes the effects of the polyamides on the 3′ processing reaction. The results of an exemplary 3′ processing assay with selective polyamides are shown in Figure 5. The eight polyamides could be divided into three classes, according to their inhibition efficiencies (Figure 4). The first class includes polyamides **1**, **3**, **5**, and **6**, all having an IC_{50} under 250 nM. The second class includes polyamides **2** and **7** with an IC_{50} between 450 and 650 nM. Polyamides **4** and **8** belong to the third class, which has an IC_{50} at or above 1400 nM. The polyamides are designed to bind with sequence specificity. The inhibitory effect of the polyamides correlates with the ability to efficiently bind the LTR sequences and is therefore not due to nonspecific interactions with the IN protein. A single mismatch within the polyamide was found to decrease the K_a value by 100-fold and showed little evidence of inhibition of 3′ processing and strand transfer. This is exemplified by polyamide **4** (Figure 5 and data not shown), with a K_a of 2.1×10^7 M^{-1} and an IC_{50} value at or greater than 1600 nM.

Inhibition of Strand Transfer. The polyamides were tested for inhibition of strand transfer. In this assay, a precleaved substrate (Figure 3, second row) was used instead of the blunt-end substrate, with integration occurring into a second double-stranded oligonucleotide (Figure 3, left arrow). Strand transfer is highly efficient using a precleaved substrate (*39*) and allows for the separation of the two steps in the analysis. Similar to 3′ processing, the polyamides can be separated into three classes on the basis of their IC_{50} for the strand transfer reactions (Figure 4). An exemplary strand transfer assay using selective polyamides is shown in Figure 6. The first class consists of **1** and **3**, with the IC_{50} under 250 nM. Moderate inhibition is found with the second class, including **2** and **5**, with IC_{50} at 450 and 1100 nM, respectively. The

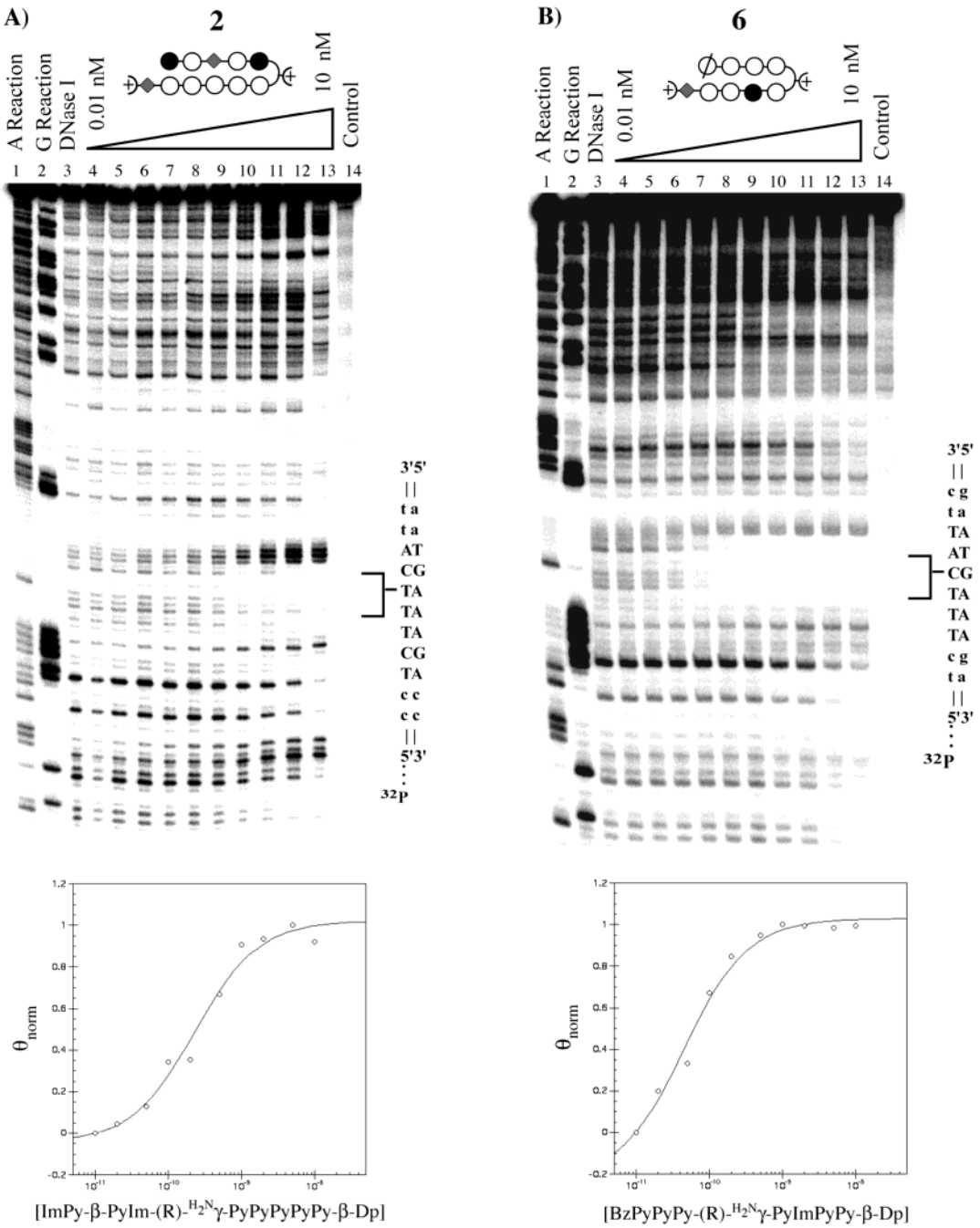


FIGURE 2: (A) Quantitative DNase I footprint titration experiment with polyamide **2** on a 5′-^{32}P-labeled 183 base pair DNA fragment containing an internal copy of the M-MuLV LTR: lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4−13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide; lane 14, intact DNA. All reactions contained 5 pM DNA, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ and were performed at 22 °C. Data were obtained for the binding site indicated at the right of the gel, 5′-aaTGAAAGAcc-3′, and are shown in the isotherm plot below. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods. The solid curve is a best-fit Langmuir binding titration isoterm obtained from a nonlinear least-squares algorithm, where $n = 1$ as previously described (*40*). (B) Quantitative DNase I footprint titration experiment with polyamide **6** on a 5′-^{32}P-labeled 183 base pair DNA fragment containing an internal copy of the M-MuLV LTR: lane 1, A reaction; lane 2, G reaction; lane 3, DNase I standard; lanes 4−13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide; lane 14, intact DNA. All reactions contained 5 pM DNA, 10 mM Tris-HCl (pH 7.0), 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ and were performed at 22 °C. Data were obtained for the binding site indicated at the right of the gel, 5′-gaATGAAAga-3′, and are shown in the isotherm plot below. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods. The solid curve is a best-fit Langmuir binding titration isotherm obtained from a nonlinear least-squares algorithm, where $n = 1$ as previously described (*40*).

third class includes **4** and **6**−**8**. The IC_{50} of the third class are at or above 1400 nM.

Although polyamides **5** and **6** were among the most effective inhibitors of 3′ processing, they showed little effect on strand transfer. Polyamides **1** and **3**, though, are highly effective at inhibition of both the 3′ processing and strand transfer reactions. This raises the interesting possibility that distinct regions of the viral termini are required for 3′ processing and strand transfer. Additionally, the absence of the TT dinucleotide in the strand transfer substrate will affect

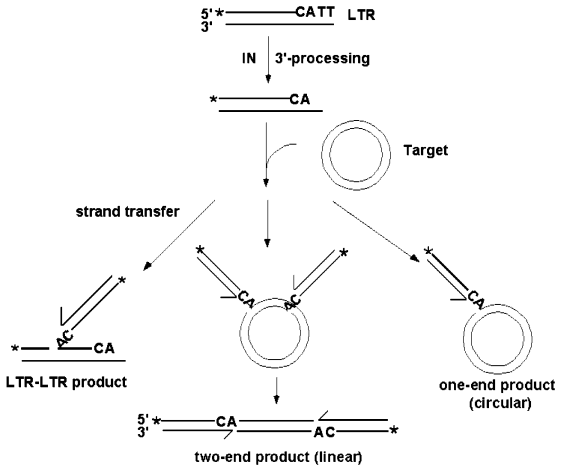


FIGURE 3: Schematic illustration of in vitro integration assays. IN first removes the TT dinucleotide from the 5′-^{32}P-labeled blunt LTR duplex, leaving a single-strand tail. Then IN integrates the cleaved strand into the target: either the LTR itself in the strand transfer assay or the plasmid DNA as an exogenous substrate. The major products of integration are as listed.

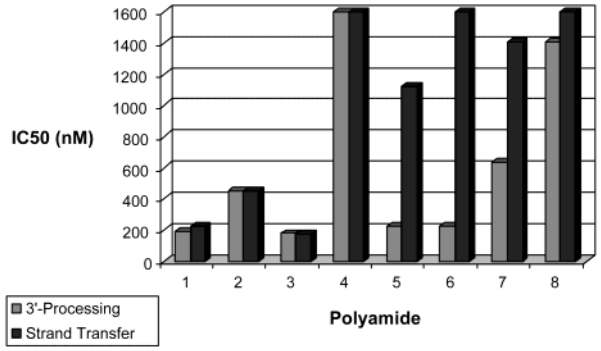


FIGURE 4: IC_{50} of polyamides on oligonucleotide-based M-MuLV integrase activities. The inhibitory effect of each polyamide (0.5−1600 nM) was titrated in either the 3′ processing (gray) or strand transfer (black) assays in the presence of 33 nM labeled LTR. Individual bands of reaction products were quantified using Scion Image 1.62C. For each assay, reaction products of the M-MuLV IN catalyzed in the absence of polyamides were defined as 100% activity. Equivalent fractions from lanes in the absence of enzyme were taken as background and subtracted from each reaction. IC_{50} values were determined as the concentration of each polyamide that inhibited IN activities by 50%.

the binding of polyamides, particularly **5**−**7**, whose cognate sites are closest to the end of the viral DNA. Each of these polyamides was less effective at inhibiting strand transfer than 3′ processing, whereas polyamides **1**−**3** showed the same efficacy in both assays (Figure 4). This further emphasizes the importance of cognate site recognition in integrase inhibition.

In the oligonucleotide-based assay, the LTR substrate functions as both the viral and target DNA. Previous analysis of binding site selection indicated that the major groove was the main determinant within the target DNA for IN docking (*40*−*44*). Therefore, it was believed that binding of the polyamides to the target DNA should not influence target site selection. Experimentally, variations in target site selection were observed. In the presence of a high concentration of polyamide **2** (640 and 1600 nM; Figure 6, lanes 15 and 16), loss of integration at the predominant sites was paralleled with enhanced production of a single product of slow electrophorectic mobility (marked by the asterisk). This product was also observed at saturating levels of polyamide **3** (Figure 6, lanes 21−23). Formation of this large product is a result of integration at one end of the target DNA. Target site selection using polyamide **5** was also altered, with preference for the faster mobility products at the highest polyamide concentrations (Figure 6, lanes 36 and 37).

Inhibition of Strand Transfer into an Exogenous Target. To separate the effects of polyamide binding to the viral versus target site, an alternative in vitro integration assay was used (Figure 3) in which a precleaved LTR oligonucleotide donor was integrated into a nonspecific circular target plasmid. The target DNA does not encode the LTR sequence, and therefore only the effect of the polyamide on the LTR is examined. Integration of the LTR into the target plasmid would result in two products: circular "lariat-like" DNA from integration of one LTR end and linear DNA from integration of two LTR ends in coordination (Figure 3, right and center arrows, respectively). The integration products into the plasmid DNA are visualized after separation on an agarose gel. Using WT M-MuLV IN, the predominant product was the lariat product (Figure 7). The general efficiency of the polyamides was as found with the oligonucleotide-based assay. Examples of polyamides in which no inhibition was observed include polyamides **4** and **6** (Figure 7, lanes 21−26 and lanes 33−38, respectively). Inhibition by polyamide **2** and **5** (Figure 7, lanes 14 and 32, respectively) was observed at the highest concentration tested (1600 nM). Polyamides **1** and **3** (Figure 7, lanes 3−8 and 15−20, respectively) were effective in inhibiting strand transfer between 250 and 640 nM. Interestingly, polyamide **3** yielded reproducibly the lowest IC_{50}, indicating recognition of the LTR sequence in the presence of excess nonspecific target DNA. Longer exposure of the gels for the two-end linear products paralleled the results for the one-end integration (data not shown). The similar profiles observed in both the two-end integration assay and the strand transfer assay indicate that the two are mechanistically the same.

DISCUSSION

In this report, a series of polyamides were designed and tested, and specific inhibitors of integration catalyzed by the M-MuLV IN were identified. Previous studies have shown that DNA binding agents can effectively inhibit retroviral integration (*14*, *15*, *18*, *45*, *46*). However, all of the compounds have very limited sequence specificity and can bind to a very large number of potential DNA sites. The polyamides developed in this report bind specifically to the minor groove of sequences required for retroviral integration, those at or flanking the conserved CA dinucleotide. Polyamides were identified with subnanomolar binding constants. The polyamides were tested in three different integration assays in cell-free experiments. The inhibition of retroviral integration observed was due to binding of the polyamides to the LTR termini. Polyamides with mismatch recognition of the M-MuLV LTR were much less efficient inhibitors. This is exemplified by polyamides **4** and **8**, which both bind to the M-MuLV LTR with K_a values of no greater than 2×10^7 M^{-1} and have IC_{50} at or greater than 1400 nM. The inhibition observed, therefore, is not due to a nonspecific binding of the polyamides to DNA or to the IN protein. A clear dependence on cognate site recognition is demonstrated by the discrepancy between the 3′ processing and strand

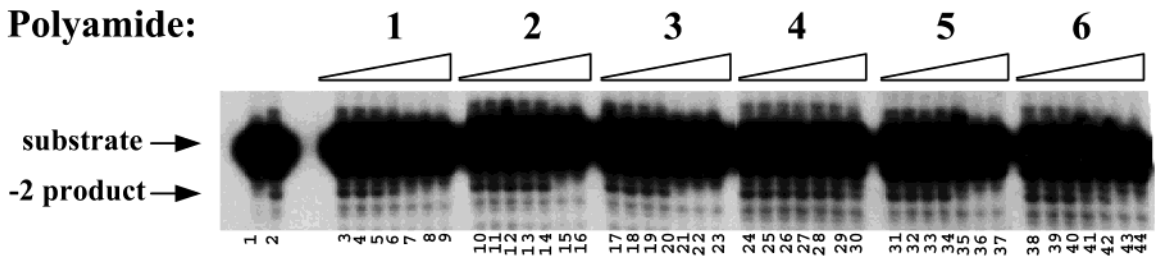


FIGURE 5: Inhibition of 3′ processing reactions by the polyamides. 3′ processing reactions were performed as described in Experimental Procedures. Exemplary assays for limited polyamides are shown. The position of the −2 3′ processing product is indicated at the left. Lane 1, no protein control; lane 2, M-MuLV IN alone; lanes 3−9, IN plus polyamide **1**; lanes 10−16, IN plus polyamide **2**; lanes 17−23, IN plus polyamide **3**; lanes 24−30, IN plus polyamide **4**; lanes 31−37, IN plus polyamide **5**; and lanes 38−44, IN plus polyamide **6**. The concentrations of the polyamide used are (left to right) 0.4, 16, 40, 100, 256, 640, and 1600 nM. Reactions contained 33 nM labeled LTR.

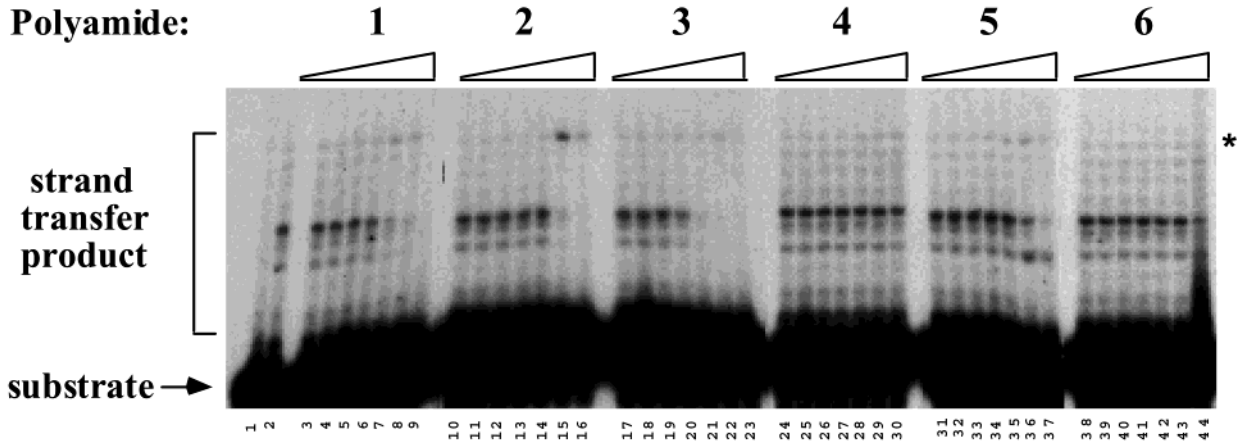


FIGURE 6: Inhibition of strand transfer reactions by the polyamides. Strand transfer assays, based on oligonucleotides substrates, were performed as described in Experimental Procedures. Strand transfer yields ^{32}P-labeled products larger than the input substrate, indicated at the left of the panel. Lane 1, no protein control; lane 2, M-MuLV IN alone; lanes 3−9, IN plus polyamide **1**; lanes 10−16, IN plus polyamide **2**; lanes 17−23, IN plus polyamide **3**; lanes 24−30, IN plus polyamide **4**; lanes 31−37, IN plus polyamide **5**; and lanes 38−44, IN plus polyamide **6**. The concentrations of the polyamide used are (left to right) 0.4, 16, 40, 100, 256, 640, and 1600 nM. Reactions contained 33 nM labeled LTR. The asterick indicates the formation of the enhanced integration product.

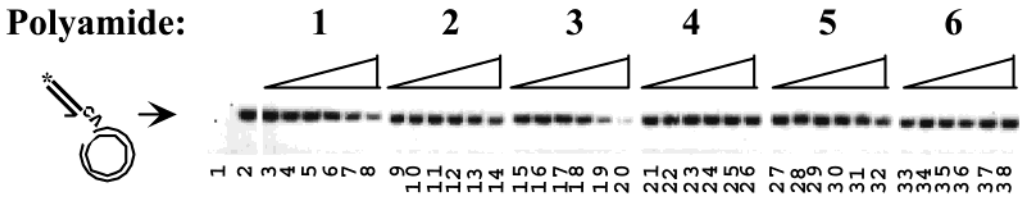


FIGURE 7: Polyamides inhibit in vitro strand transfer into an exogenous substrate. 33 nM ^{32}P-labeled LTR substrate was analyzed for integration into an unlabeled plasmid DNA. The migration on an agarose gel of the single-end lariat integration product is indicated at the left of the panel. Lane 1, no protein control; lane 2, M-MuLV IN alone; lanes 3−8, IN plus polyamide **1**; lanes 9−14, IN plus polyamide **2**; lanes 15−20, IN plus polyamide **3**; lanes 21−26, IN plus polyamide **4**; lanes 27−32, IN plus polyamide **5**; and lanes 33−38, IN plus polyamide **6**. The concentrations of the polyamide used are (left to right) 16, 40, 100, 256, 640, and 1600 nM.

transfer results for polyamides **5**−**7**. The binding site of these polyamides is affected by the removal of the terminal TT dinucleotide, and thus the cognate polyamides showed reduced strand transfer inhibition as compared to 3′ processing, where the cognate site is intact.

Despite the importance of cognate site recognition, the K_a values determined by DNase I footprinting with an internal copy of the M-MuLV LTR were not direct predictors of the efficacy of integrase inhibition. In general, all of the polyamides containing an IC_{50} lower than 640 nM had binding constants in the nanomolar range or lower. However, this correlation is not reciprocal. Polyamides with perfect sequence matches show subtle variations with the correlation of the K_a values and the effective IC_{50}. Polyamides **2** and **7** have K_a values near 3.0×10^9 M^{-1}, yet are not as effective at inhibiting 3′ processing as polyamides **1** and **3**, which have association constants in the same range. As noted, given the proximity of its binding sites to the end of the LTR, the K_a value for **7** should be considered an upper limit. This may also explain why the efficacy of polyamides **5** and **6** does not exceed that of polyamides **1** and **3**. The lower activity of **2** is more difficult to explain, given its moderately higher affinity than **1**.

The similar efficiencies of **1** and **3** are of interest. When footprinted against the internal copy of the LTR, the tail of polyamide **3** was placed over G·C base pairs, which is considered a mismatch (*47*). Under these conditions its association constant was quite close to that of polyamide **1**. These G·C pairs are not present in the 3′ processing substrate LTR, where the tail of **3** extends beyond the terminal base pair. It

is thus possible that placing the β-Dp tail of this extended-hairpin polyamide beyond the terminal base pair has energetic consequences similar to placing it against G·C base pairs. However, if **1** and **3** do have the same affinity, one might expect that **3** would be the better inhibitor of 3′ processing, due to its C-terminal extension, which places Py residues against the bases to be excised. Given that these bases are absent in the strand transfer substrate, it is also interesting that **3** maintains its activity in this assay, while the activity of **5**−**7** drops. In fact, in the presence of excess unlabeled DNA in the exogenous target assay, polyamide **3** was the most effective strand transfer inhibitor. Since the hairpin portion of **3** is internal to the excised nucleotides, it can be speculated that **3** maintains high-affinity binding through its hairpin portion. Perhaps the long C-terminal tail provides a steric impediment to strand transfer, even in the absence of cognate base-paired DNA.

It is also significant that all of the effective strand transfer inhibitors bind positions 7 through 9 of the M-MuLV LTR. This is consistent with the observations of Balakrishnan and others in the respective HTLV (human T-lymphotropic virus) and HIV-1 systems of the importance of these residues in related viral LTRs (*36*, *37*, *48*). Positions 7−9 are bound by either the tail or turn of polyamides **5**−**7** or are outside of the binding site of these polyamides. The turn and tail residues are not considered to be as effective steric blockades of the minor groove as ring pairings. Thus, reduced blockage of positions 7−9 relative to polyamides **1**−**3** correlates with reduced strand transfer inhibition, although this analysis is complicated by the expectation that polyamides **5**−**7** will have reduced affinity for the strand transfer substrate. It would be interesting to test the ability of a polyamide that binds internally to positions 7−9 to act as a strand transfer inhibitor.

The integration process itself is a dynamic process. Current models predict conformational changes accompanying cleavage of the viral DNA terminus (*12*). Differential cross-linking efficiencies to the HIV-1 LTR sequences were observed for substrates for 3′ processing, strand transfer, and disintegration (*12*). In addition, the binding of the diketo acid inhibitor L-731988 required the assembly of the IN protein with the viral end. L-731988 competes with the target DNA, indicating that the target binding site is generated after the binding of the viral LTR. These compounds selectively inhibited strand transfer, without any effect on 3′ processing (*11*). Cross-linking studies with HIV-1 IN indicate that the C-terminal domain of one IN monomer acts in trans with the catalytic core of another monomer for each viral end (*12*). This in trans assembly may involve the sequences upstream of the viral LTRs, within the region bound by polyamides **1** and **3**. Similarly, structural studies of Tn5 transposase have demonstrated that both the internal and the terminal regions of the transposon DNA are contacted by the transposase (*49*). This conformational change conceptually follows the assembly and transport of the preintegrative complexes. In vivo, 3′ processing is temporally and spatially independent of strand transfer (*50*). 3′ processing occurs in the cytoplasm for MuLV, whereas strand transfer occurs in the nucleus. Initial binding of the viral DNA requires the recognition of the linear viral terminus, including the CA dinucleotide. The requirements of protein domains for 3′ processing are stringent; for M-MuLV IN, all domains including the N-terminal HHCC region, are required (*7*, *51*, *52*).

The substrate for 3′ processing is a blunt-end double-strand DNA. Fraying or unwinding of the termini has been proposed to be a requisite step in the 3′ processing reaction (*53*, *54*). Several of the polyamides target the terminal dinucleotides. It was possible that the binding of polyamide might impede the unwinding step. However, no preferential inhibition of 3′ processing was observed for the polyamides which bound to the terminal TT sequence. Polyamides which bind to the terminal nucleotides may, however, sterically hinder subsequent assembly of a strand transfer complex. Polyamide **3** recognizes the largest binding site, from position 1 to position 9 of the LTR, including a single-stranded tail covering the terminal TT dinucleotide. As discussed, this may contribute to the overall effectiveness of inhibition by polyamide **3** in both the 3′ processing and strand transfer assays.

The variation in target site selection in the presence of polyamides is of interest. Previous studies indicated, for HIV-1 IN, that target sites were preferred within nucleosomes with wide major grooves (*41*). Altered target site selection induced by the polyamides binding to the minor groove can indicate either that minor groove contacts are important or that structure of the major groove can be altered by binding of polyamides to the minor groove. Structural evidence suggests that hairpin polyamides bind a wide minor groove, with requisite narrowing of the major groove (*55*−*58*). Therefore, it is possible that presence of the polyamides can have an indirect effect on the interaction of the preintegrative complex within the major groove. It is also expected that binding of polyamides in the minor groove will reduce the flexibility of a given stretch of DNA, reducing access to DNA conformations that could be important in the dynamic process of target site selection.

These studies represent the first generation of polyamides that inhibit MuLV integration. Four potent inhibitors were developed, two which inhibit 3′ processing (**5** and **6**) and two which inhibit both 3′ processing and strand transfer (**1** and **3**). These different classes of inhibitors could be useful tools for unraveling the complex mechanism of integration. Both sets of compounds are effective at a less than 10-fold excess over substrate DNA and less than half an equivalent of integrase protein. Next generation polyamides can incorporate hydroxypyrrole (*58*) and hydroxybenzamide (*29*) residues, which break the A·T/T·A degeneracy of the Py/Py and Bz/Py pairs, allowing for greater sequence specificity. Strategies other than increasing the binding efficiency could make the inhibitors more potent. Alterations in the structure of the polyamides could change the conformation of the LTR and/or interfere with the binding of the IN. Modified polyamides which may be able to clamp the terminal TT dinucleotide by alkylation (*59*, *60*) may be envisioned. Certainly, a crystal structure of an IN−LTR complex would aid the design of polyamide inhibitors.

Polyamides can serve as the matrix to be coupled with additional inhibitors, for example, compounds that bind the enzyme active site (*61*). Delivery of polyamides into viral particles is a complex question, as the viral RNA packaged in virions is not the double-strand DNA substrate recognized by the polyamides. Polyamide inhibitors would need to be accessible to the reverse transcription products prior to the assembly of the intasome complex (*62*, *63*). Despite this

difficulty, the modular nature of polyamide recognition and synthesis should allow for rapid and rational changes to the inhibitors in response to viral mutations. With the appearance of leukemia in two children receiving gene therapy using an MuLV-based vector (*64*), the ability to block the integration and/or expression from the MuLV LTRs is of great importance.

The HIV-1 IN and M-MuLV IN share the same mechanism of integration. It is very likely that the same strategy of designing polyamide inhibitors can be applied to the HIV-1 IN. Studies are currently underway to test polyamides with the HIV-1 IN/LTR system. Given the previously demonstrated efficacy of polyamides targeted to HIV-1 in human cells (*21*), this approach may ultimately yield potent in vivo HIV-1 IN inhibitors. With the appearance of HIV strains resistant to both the protease and the reverse transcriptase inhibitors, the need to develop potent inhibitors against new targets including the integrase is more urgent (*65*). The increased interest for inhibitors that target the integrase has provided lead compounds (*65*, *66*), although no clinically effective inhibitor is available to date.

SUPPORTING INFORMATION AVAILABLE

One figure and one table detailing the DNA melting experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Brown, P. O. (1997) Integration, in *Retroviruses* (Coffin, J. M., Hughes, S. H., and Varmus, H. E., Eds.) pp 161−203, Cold Spring Harbor Laboratory Press, Plainview, NY.
2. Hindmarsh, P., and Leis, J. (1999) *Microbiol. Mol. Biol. Rev. 63*, 836−843.
3. Katzman, M., Katz, R. A., Skalka, A. M., and Leis, J. (1989) *J. Virol. 63*, 5319−5327.
4. Craigie, R., Fujiwara, T., and Bushman, F. (1990) *Cell 62*, 829−837.
5. Sherman, P. A., and Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5119−5123.
6. Singh, I. R., Crowley, R. A., and Brown, P. O. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 1304−1309.
7. Yang, F., and Roth, M. J. (2001) *J. Virol. 75*, 9561−9570.
8. Craigie, R. (2001) *J. Biol. Chem. 276*, 23213−23216.
9. Hyde, C. C., Bushman, F. D., Mueser, T. C., and Yang, Z.-N. (1999) *J. Mol. Biol. 296*, 535−538.
10. Wang, J.-Y., Ling, H., Yang, W., and Craigie, R. (2001) *EMBO J. 20*, 7333−7343.
11. Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J., Anthony, N., Egbertson, M., Melamed, J. Y., Young, S., Hamill, T., Cole, J. L., and Hazuda, D. J. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 11244−11249.
12. Gao, K., Butler, S. L., and Bushman, F. (2001) *EMBO J. 20*, 3585−3576.
13. Wang, T., Balakrishnan, M., and Jonsson, C. B. (1999) *Biochemistry 38*, 3624−3632.
14. Neamati, N., Mazumder, A., Sunder, S., Owen, J. M., Tandon, M., Lown, J. W., and Pommier, Y. (1998) *Mol. Pharmacol. 54*, 280−290.
15. Ryabinin, V. A., Sinyakov, A. N., de Soultrait, V. R., Caumont, A., Parissi, V., Zakharova, O. D., Vasyutina, E. L., Yurchenko, E., Bayandin, R., Litvak, S., Tarrago-Litvak, L., and Nevinsky, G. A. (2000) *Eur. J. Med. Chem. 35*, 989−1000.
16. Pommier, Y., and Neamati, N. (1999) *Adv. Virus Res. 52*, 427−458.
17. Pommier, Y., Marchand, C., and Neamati, N. (2000) *Antiviral Res. 47*, 139−148.
18. Carteau, S., Mouscadet, J. F., Goulaouic, H., Subra, F., and Auclair, C. (1994) *Biochem. Pharmacol. 47*, 1821−1826.
19. Dervan, P. B., and Burli, R. W. (1999) *Curr. Opin. Chem. Biol. 3*, 688−693.
20. Dervan, P. B. (2001) *Bioorg. Med. Chem. 9*, 2215−2235.
21. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 12890−12895.
22. Chiang, S. Y., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B., and Beerman, T. A. (2000) *J. Biol. Chem. 275*, 24246−24254.
23. Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expression 9*, 77−91.
24. Janssen, S., Cuvier, O., Muller, M., and Laemmli, U. (2000) *Mol. Cell 6*, 1013−1024.
25. Janssen, S., Durussel, T., and Laemmli, U. (2000) *Mol. Cell 6*, 999−1011.
26. Belitsky, J., Leslie, S., Arora, P., Beerman, T., and Dervan, P. (2002) *Bioorg. Med. Chem. 10*, 3313−3318.
27. Baird, E. E., and Dervan, P. B. (1996) *J. Am. Chem. Soc. 118*, 6141−6146.
28. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc. 120*, 1382−1391.
29. Ellervik, U., Wang, C. C. C., and Dervan, P. B. (2000) *J. Am. Chem. Soc. 122*, 9354−9360.
30. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature 382*, 559−561.
31. Trauger, J. W., and Dervan, P. B. (2001) *Methods Enzymol. 340*, 450−466.
32. Felkner, R. H., and Roth, M. J. (1992) *J. Virol. 66*, 4258−4264.
33. Jonsson, C. B., Donzella, G. A., and Roth, M. J. (1993) *J. Biol. Chem. 268*, 1462−1469.
34. LaFemina, R. L., Callahan, P. L., and Cordingley, M. G. (1991) *J. Virol. 65*, 5624−5630.
35. Vink, C., van Gent, D. C., Elgersma, Y., and Plasterk, R. H. (1991) *J. Virol. 65*, 4636−4644.
36. Balakrishnan, M., and Jonsson, C. B. (1997) *J. Virol. 71*, 1025−1035.
37. Esposito, D., and Craigie, R. (1998) *EMBO J. 17*, 5832−5843.
38. Pilch, D. S., Poklar, N., Baird, E. E., Dervan, P. B., and Breslauer, K. J. (1999) *Biochemistry 38*, 2143−2151.
39. Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1339−1343.
40. Pruss, D., Reeves, R., Bushman, F. D., and Wolffe, A. P. (1994) *J. Biol. Chem. 269*, 25031−25041.
41. Pruss, D., Bushman, F. D., and Wolffe, A. P. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 5913−5917.
42. Pryciak, P. M., Sil, A., and Varmus, H. E. (1992) *EMBO J. 11*, 291−303.
43. Pryciak, P. M., and Varmus, H. E. (1992) *Cell 69*, 769−780.
44. Pryciak, P. M., Muller, H. P., and Varmus, H. E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9237−9241.
45. Carteau, S., Mouscadet, J. F., Goulaouic, H., Subra, F., and Auclair, C. (1993) *Biochem. Biophys. Res. Commun. 192*, 1409−1414.
46. Fesen, M. R., Kohn, K. W., Leteurtre, F., and Pommier, Y. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2399−2403.
47. Swalley, S. E., Beard, E. E., and Dervan, P. B. (1999) *J. Am. Chem. Soc. 121*, 1113−1120.
48. Katzman, M., and Katz, R. A. (1999) *Adv. Virus Res. 52*, 371−395.
49. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000) *Science 289*, 77−85.
50. Roe, T., Chow, S. A., and Brown, P. O. (1997) *J. Virol. 71*, 1334−1340.
51. Jonsson, C. B., Donzella, G. A., Gaucan, E., Smith, C. M., and Roth, M. J. (1996) *J. Virol. 70*, 4585−4597.
52. Jonsson, C. B., and Roth, M. J. (1993) *J. Virol. 67*, 5562−5571.
53. Scottoline, B. P., Chow, S., Ellison, V., and Brown, P. O. (1997) *Genes Dev. 11*, 371−382.
54. Katz, R. A., DiCandeloro, P., Kukolj, G., and Skalka, A. M. (2001) *J. Biol. Chem. 276*, 34213−34220.
55. Clairac, R. P. L. d., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *J. Am. Chem. Soc. 119*, 7909−7916.
56. Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Nat. Struct. Biol. 5*, 104−109.
57. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) *J. Mol. Biol. 295*, 557−567.
58. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Science 282*, 111−115.
59. Chang, A. Y., and Dervan, P. B. (2000) *J. Am. Chem. Soc. 122*, 4856−4864.

60. Wurtz, N. R., and Dervan, P. B. (2000) *Chem. Biol. 7*, 153−161.
61. Erlanson, D. A., Braisted, A. C., Raphael, D. R., Randal, M., Stroud, R. M., Gordon, E. M., and Wells, J. A. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 9367−9372.
62. Wei, S. Q., Mizuuchi, K., and Craigie, R. (1997) *EMBO J. 16*, 7511−7520.
63. Chen, H., Wei, S. Q., and Engelman, A. (1999) *J. Biol. Chem. 274*, 17358−17364.
64. Marshall, E. (2003) *Science 299*, 320.
65. Miller, M. D., and Hazuda, D. (2001) *Curr. Opin. Microbiol. 4*, 535−539.
66. Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000) *Science 287*, 646−650.

BI034177S