THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 269, No. 40, Issue of October 7, pp. 25031–25041, 1994
Printed in U.S.A.

The Influence of DNA and Nucleosome Structure on Integration Events Directed by HIV Integrase*

(Received for publication, May 4, 1994, and in revised form, July 27, 1994)

Dmitry Pruss, Raymond Reeves, Frederic D. Bushman‡, and Alan P. Wolffe

From the Laboratory of Molecular Embryology, NICHHD, National Institutes of Health, Bethesda, Maryland 20892 and the ‡Infectious Disease Laboratory, Salk Institute, La Jolla, California 92014

DNA copies of the human immunodeficiency virus (HIV) genome integrate nonrandomly into the chromosomal DNA of the host cell. In this report, we investigate the molecular basis of this selectivity using the virus-encoded HIV integrase to direct integration of a synthetic HIV long terminal repeat substrate into either DNA molecules of known structure or previously defined nucleosomal complexes. We find that the structure of the target greatly influences the site of integration, and, moreover, DNA curvature, flexibility, and rigidity in solution all influence the frequency of integration. Importantly, for DNA with all of these properties, the distortion of the double helix directed by association with the histone proteins promotes the integration reaction and alters the distribution of sites that are selected for integration. We suggest that both intrinsic DNA structure and the folding of DNA into chromosomal structures will exert a major influence on target site selection for integration of the viral genome.

Retroviruses and their retrotransposon relatives must integrate a DNA copy of their RNA genome into a chromosome of the host cell during their life cycle (Varmus and Swanstrom, 1985; Vink and Plasterk, 1993). The distribution of integration sites has been found to show considerable target site specificity (Shih *et al.*, 1988; Sandmeyer *et al.*, 1990). However the molecular basis of this target specificity within the chromosome is still incompletely defined.

In vivo, integration sites are often found within regions of chromatin that are also preferentially accessible to the enzyme DNase I (Rohdewohld *et al.*, 1987; Scherdin *et al.*, 1990; Vijaya *et al.*, 1986). These DNase I hypersensitive sites represent segments of chromatin where the folding of nucleosomal arrays is disrupted by DNA having unusual structural properties or by the interaction of trans-acting factors with DNA (Elgin, 1988; Gross and Garrard, 1988). These DNase I hypersensitive sites often contain histone-bound DNA interspersed with histone-free regions (Thomas and Elgin, 1988; Bresnick *et al.*, 1992; McPherson *et al.*, 1993). A well studied example concerns the integration of the yeast retrovirus-like transposons Ty1 and Ty3, which, despite their markedly different site specificities, both greatly favor integration around the DNase I hypersensitive sites associated with tRNA genes (Huibregtse *et al.*, 1987; Chalker and Sandmeyer, 1992; Ji *et al.*, 1993). Integration occurs within regions complexed either by transcription factors (Kassavetis *et al.*, 1989) or by histones (Morse *et al.*, 1992). The significance of these protein-DNA interactions at the tRNA gene in facilitating integration has not yet been resolved. Possible mechanisms include protein-protein interaction between trans-acting factors or the histones with the integration machinery or a sensitivity of integration to altered DNA conformation directed by bound proteins (Craigie, 1992).

The exact structural features of DNA recognized by the integrase machinery have not yet been determined. Varmus and co-workers (Pryciak and Varmus, 1992; Pryciak *et al.*, 1992a, 1992b) demonstrated that the incorporation of DNA into nucleosomes could enhance integration at particular DNA positions. These important studies suggested that integrase approaches DNA within the nucleosome face on, directing integration events with a 5-bp[1] stagger across the major groove and that changes in DNA structure following association with the histones might potentially directly influence the integration reaction. We have recently extended this work to demonstrate that sites of integration cluster in the most severely deformed and kinked DNA regions within the nucleosome core (Pruss *et al.* (1994), see also Richmond *et al.* (1984) and Hayes *et al.* (1991a, 1991b)). These results further emphasize an important role for histone-induced DNA structure in integration site selection. However, any role for intrinsic DNA structure in directing integration events by HIV integrase was not investigated in these studies.

In this report, we describe experiments that have directly examined the influence of both intrinsic DNA structure and that directed by association with the core histones on the integration reaction mediated by HIV integrase (Bushman and Craigie, 1991, 1992). As targets for the HIV integrase, we have made use of DNA molecules containing A/T rich sequence segments that have very different structures in solution. These DNA segments are either intrinsically straight and rigid (oligo(dA)·oligo(dT)), flexible (oligo[d(A-T)]) or curved. The organization of this DNA either when free in solution or following association with the core histones has been previously characterized in detail (Wolffe and Drew, 1989; Hayes *et al.*, 1991b). Our results demonstrate that the integration mediated by HIV integrase is sensitive to DNA structure, preferring DNA as a target that has intrinsic curvature rather than flexibility or rigidity. However, once incorporated into a nucleosome, we find that the deformation of DNA by the core histones both enhances integration and alters the distribution of sites that are selected for integration.

MATERIALS AND METHODS

Targets for Integration—Segments of bent, rigid, or flexible DNA (50 bp) cloned in about one DNA turn upstream from the T7 promoter in a T7-CAT construct were described by Wolffe and Drew (1989). DNA fragments, 271 bp in length, were excised by *Hin*dIII-*Pvu*II. A portion of the DNA was radiolabeled at the *Hin*dIII site to serve as a tracer during

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] The abbreviations used are: bp, base pair(s); HIV, human immunodeficiency virus; CAT, chloramphenicol acetyltransferase; LTR, long terminal repeat; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid; PCR, polymerase chain reaction.

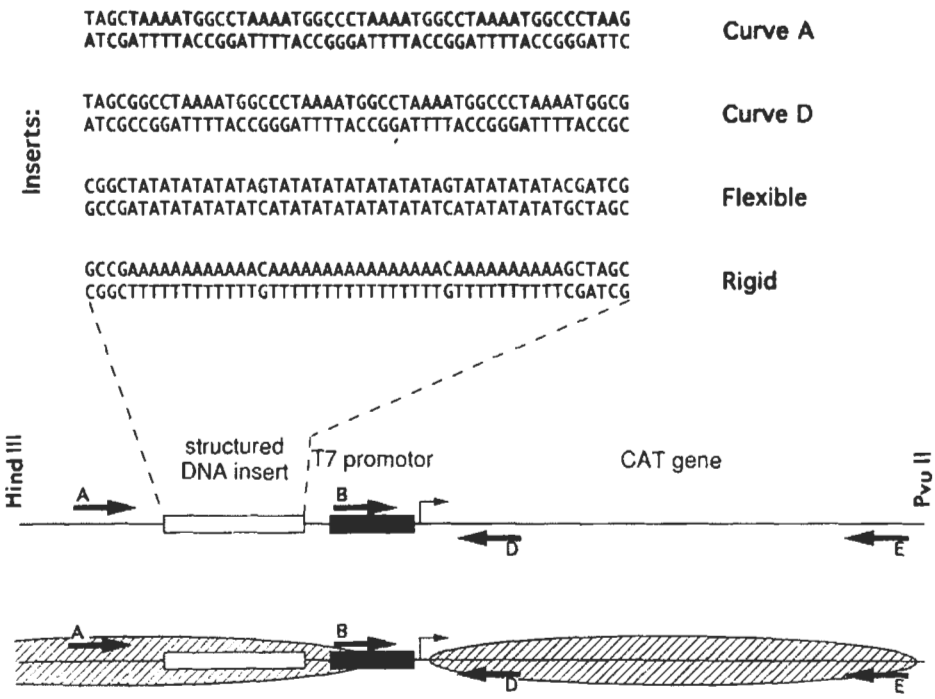

FIG. 1. **Design of the DNA fragments (Wolffe and Drew, 1989).** The four DNA 50-mers of known structure (*open box*) were inserted at a *Bam*HI site immediately upstream from the promotor (*closed box*) of the pT7CAT plasmid. *Hin*dIII-*Pvu*II fragments were probed either as naked DNA or after dinucleosome assembly (approximate nucleosome positions are shown as *hatched ovals*, see Wolffe and Drew (1989)). Also shown are the four primers used for PCR of the the products of the integration reaction. The *Hin*dIII site is 5′ to the structured DNA insert, and the *Pvu*II site is at the 3′ end of the CAT gene.

the nucleosome reconstitution. For this purpose, the plasmids were first cut with *Hin*dIII; the linearized DNA was then treated with alkaline phosphatase and radiolabeled with T4 polynucleotide kinase and [^{32}P]ATP before it was restricted with *Pvu*II, purified by electrophoresis in nondenaturing 3.5% agarose gels, and extracted either by electroelution or by QIAEX column purification (QIAGEN Inc., Chatsworth, CA). Chicken erythrocyte nucleosome core particles (to be used as a histone source for nucleosome reconstitution, see below) were prepared by the method of Libertini and Small (1980), and nucleosome core DNA was prepared by successive phenol-chloroform extractions of the core particle preparation.

Histones were reconstituted onto the DNA fragments by exchange with 30–60-fold excess of chicken erythrocyte core particles (Tatchell and van Holde, 1977). The extent of reconstitution was monitored by 0.7% agarose gel electrophoresis (typically, less than 5% of mononucleosomes and negligible amounts of unbound DNA were present in the preparations). Naked DNA targets were made by mixing free DNA fragments with the equivalent excess of core particle DNA.

The Integration Reaction—The synthetic HIV LTR substrate used for the integration reaction corresponded to a 3′ processed intermediate of integration of the viral U3 end (Bushman and Craigie, 1991). It was made by annealing the oligonucleotides RB67 (30-nucleotide integrating strand, 5′-CGATAGGATCCGAGTGAATTAGCCCTTCCA) and RB50 (32-nucleotide nonintegrating strand, 5′-ACTGGAAGGGCTAATTCACTCGGATCCTATCG).

The reaction mixture for integration contained 0.5–1 μg/μl of the synthetic HIV LTR substrate, 0.1–0.5 μg of the HIV integrase stock solution, and 20 ng of either naked or nucleosomal target DNA (including 0.3–0.8 ng of the fragment DNA) in a buffer of 7.5 mM $MnCl_2$, 20 mM HEPES·NaOH, pH 7.3, 10 mM 2-mercaptoethanol, 10% glycerol, and 0.75 mM CHAPS within a final reaction volume of 24 μl. The HIV integrase was solubilized at 1 μg/μl in 1 M NaCl, 20 mM HEPES.NaOH pH 7.3, 0.1 mM EDTA, 1 mM dithiothreitol, 20% glycerol. The addition of enzyme solution or buffer solution alone leads to an elevation of ionic strength equivalent to 20 mM NaCl. The target was always the last component to be added to the reaction mixture. In the absence of target DNA, integration will occur between synthetic HIV LTR substrate molecules (self-integration). The reaction is allowed to proceed for 2 h at 37 °C before it is stopped through the addition of 25 mM EDTA and 15 μl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.4) plus 1 mM CHAPS. An aliquot (1–3 μl) was taken for PCR with an LTR-specific primer RB67 and one of four target-specific fragments; primers A and B corresponded to nucleotides 6–27 and 91–114, respectively, of the top strand, and primers D and E corresponded to nucleotides 149–126 and 237–213, respectively of the bottom strand. PCR was routinely run for 25 cycles either in the presence of 15–25 μCi of [α-^{32}P]dATP for uniform labeling or with the addition of 5′-labeled target-specific primer for end labeling. Proteins were removed by digestion with proteinase K (0.04 μg/μl) in the presence of 0.1% SDS at 37 °C for 1 h. This is followed by phenol/chloroform extraction and precipitation of DNA with ethanol.

Assignment of the Integration Sites and Quantitation—Integration products are resolved on denaturing 8% polyacrylamide sequencing gels containing 7 M urea to identify the sites of integration (Hayes *et al.*, 1991a, 1991b). The distance between the 5′ end of a target-specific primer and an integration site equals the length of a corresponding amplified product minus 30 nucleotides. A possible caveat is that mobility of an integration product may not match the mobility of a marker DNA of the same size exactly, because the sequence of integration products differs from the sequence of any available markers. The mobility difference most probably does not exceed one nucleotide, but a systematic error of this magnitude on each of the two DNA strands could lead to a 2-bp systematic error in the value of the stagger between the integration sites on the opposite strands. To overcome this problem, we synthesized a series of "pseudointegrates" with sequences exactly matching some of the integration products by ligating the LTR oligonucleotide into restriction sites of the target DNA; 50 ng/μl of a mildy *Sau*3AI-digested curve D-containing plasmid and 50 ng/μl of an oligonucleotide adaptor (made by annealing of the oligonucleotides RB67 (see above) and LM2 (5′-ACTGGAAGGGCTAATTCACTCGGATCCTATCTAG)) were ligated at 18 °C for 24 h with 0.2 units/μl T4 DNA ligase. A 100-fold dilution of the ligation reaction mixture was used for PCR as described above.

Laser scanning and quantitation of autoradiograms was performed using a Molecular Dynamics densitometer.

RESULTS

Strategy—We use purified HIV integrase (Bushman and Craigie, 1991) and a model viral DNA substrate to achieve integration of the model viral DNA into exogenous DNA targets (Fig. 1), either as naked DNA or after association with core histones (Hayes *et al.*, 1991b). Integration products are analyzed by PCR (Saiki *et al.*, 1985; Pryciak and Varmus, 1992; Kitamura *et al.*, 1992) using a combination of a specific primers for the HIV LTR and one of four primer homologous to various segments of the target DNA (Fig. 1, *A*, *B*, *D* and *E*). Integration products are resolved on a sequencing gel (Pryciak and Varmus, 1992; Pruss *et al.*, 1994). Since under our reaction conditions fewer than one integration event occurs on average per template, we are able to quantitate relative integration efficiencies within the structured DNA region relative to the same length region within the CAT gene. This is done with either naked

DNA or the nucleosomal DNA as a target.

The various A/T rich DNA sequences (Fig. 1) have been previously shown to have very different structures in solution but very similar structures within the nucleosome (Hayes *et al.*, 1991). Oligo(dA)·oligo(dT) tracts in solution are predicted to be straight and rigid (Nelson *et al.*, 1987), with a fairly narrow minor groove width. They have a helical periodicity of approximately 10.0 bp/turn (Rhodes and Klug, 1981); the base pairs have large positive propellar twists and an ordered water structure lies in the minor groove (Fratini *et al.*, 1982; Drew and Dickerson, 1981; Nelson *et al.*, 1987; Coll *et al.*, 1987). The sequence used in our experiment contains two GC base pairs. Guanine has its 2-amino group in the minor groove, which consequently interferes with the formation of an ordered water structure within the minor groove leading to its local widening (Burkhoff and Tullius, 1987; Hayes *et al.*, 1991b). In contrast to oligo(dA)·oligo(dT), oligo[d(A-T)] tracts are predicted to be conformationally flexible in solution. This flexibility is presumed to be a consequence of the instability of the TpA dinucleotide step (Drew and Dickerson, 1981; Travers and Klug, 1987). Theoretical studies (Chupina, 1985; Calladine, 1982) indicate that the TpA step is incompatible with high propellar twist, and, hence, the minor groove is forced to be wider.

We also make use of DNA similar to the naturally curved kinetoplast DNA containing short (5 bp) oligo(dA)·oligo(dT) tracts every 10–11 bp separated by GGCC/CCGG base pairs. The GpC step has a strong preference to lie with its relatively wide minor groove along the outside of a curved DNA in solution, whereas the ApA step prefers to lie with its relatively narrow minor groove along the inside of a curved DNA (Satchwell *et al.*, 1986). The GpC step is highly curved in many crystalline forms of DNA (Goodsell *et al.*, 1993), where a GGCC/CCGG sequence curves about a 20° closing of the major groove and an opening of the minor groove. Electrophoretic studies indicate that AAAAA and GGGCCC sequences curve DNA in opposite directions by the same amount (Brukner *et al.*, 1993). Thus, by having the ApA and GpC steps out of phase with each other by half of a helical turn of DNA, curvature is maximized in our constructs, with the phosphodiester backbone of GpC lying on the outside and that of ApA lying on the inside of the curve. The periodic narrowing and widening of minor groove width is easily detected by hydroxyl radical cleavage (Burkoff and Tullius, 1987, 1988; Hayes *et al.*, 1991b). Two DNA curves (curve A and curve D) are used in which the AAAA and GGCC sequences differ by 6 bp in their distance from the bacteriophage T7 RNA polymerase promoter (Wolffe and Drew, 1989). This provides a convenient means of assessing any influences of the curved region on adjacent DNA, since the flanking DNA will differ slightly for the two curves, A and D.

The DNA templates used as targets for integration are 271 bp in length bordered by *Hin*dIII and *Pvu*II restriction sites. This length of DNA is sufficient for the association of two histone octamers per molecule (Wolffe and Drew, 1989). The DNA templates contain the bacterial CAT gene, which strongly positions a histone octamer with one boundary at the edge of the gene sequence (Wolffe and Drew, 1989; Izban and Luse, 1991). The other histone octamer lies over the T7 promoter, and the 50-bp segment of structured DNA (Fig. 1). The exact position of this histone octamer is dependent upon the sequence of the DNA segment of known structure, our "test" DNA (Wolffe and Drew, 1989) (Fig. 2). It should be noted that there is no more than 110 bp of DNA available for interaction with this histone octamer, instead of canonical 146 bp/nucleosome. A number of reconstituted nucleosomes with similarly offset DNA positions has been characterized since their first discovery by Ramsay *et al.* (1984), and they appear to resemble conventional nucleosomes very closely, although a possibility of minor structural differences can not be excluded. Within the test DNA regions histone-DNA contacts begin and end at different DNA sequences (translational positioning), and different DNA sequences face toward solution or the histones (rotational positioning).

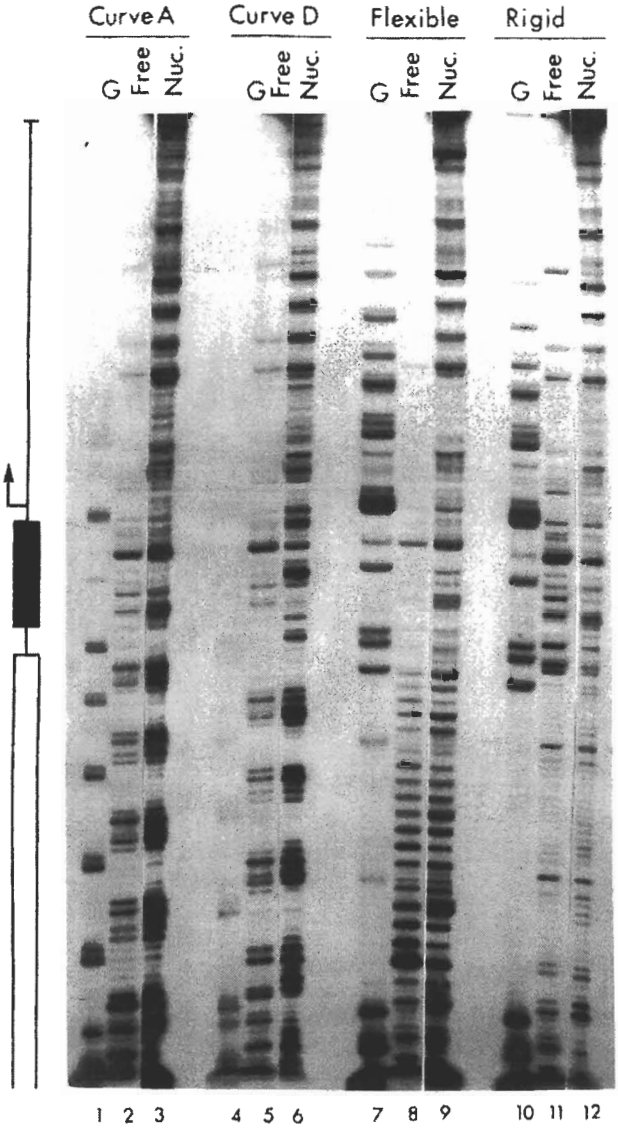

FIG. 2. **DNase I footprinting of the dinucleosomes.** The fragment was 5′-labeled at the *Hin*dIII site. DNase I cutting patterns of the fragments as dinucleosomes (*lanes 3, 6, 9,* and *12*) and of corresponding naked DNAs (*lanes 2, 5, 8,* and *11*) are shown along with G-reaction tracks generated by chemical cleavage according to Maxam and Gilbert (1977) (*lanes 1, 4, 7,* and *10*). *Open box*, DNA insert of known structure; *closed box*, T7 promoter. *Hooked arrow* indicates start site of transcription of the CAT gene.

We initially controlled for reconstitution of these DNA fragments into nucleosomal complexes identical to those that we have previously described (Wolffe and Drew, 1989; Hayes *et al.*, 1991b) by carrying out a series of nuclease digestion experiments using restriction endonucleases, micrococcal nuclease (not shown), and DNase I. DNase I footprinting reveals very similar cleavage patterns over the CAT gene region in the nucleosomal template (Fig. 2, downstream of the *hooked arrow*), but it reveals variant patterns over the segment of test DNA (Fig. 2, *open box*) and the T7 promoter (Fig. 2, *closed box*). This reflects the unique positioning, both translational and rotational, of one nucleosome over the CAT gene and the variant positioning of a second nucleosome including the structured

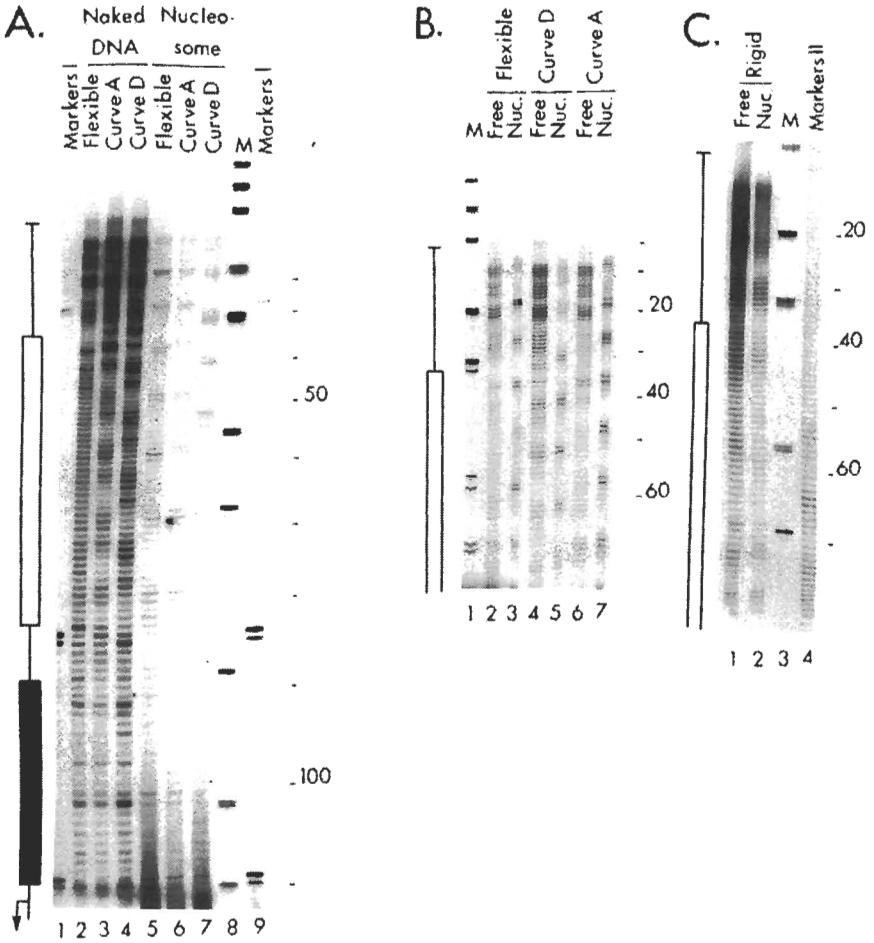

FIG. 3. **Integration footprints of the top strand of the constructs either as naked or nucleosomal targets.** Integration footprints for the curved and flexible DNA-containing constructs are shown in *panel A*, their upstream portions are shown in *panel B*, and the results using the rigid DNA-containing fragment are shown in *panel C*. *Open box*, DNA insert of known structure; *closed box*, T7 promoter. *Hooked arrow* indicates start site of transcription of the CAT gene. In this experiment primer D (Fig. 1) was used for the PCR reaction. Numbers refer to distances between the integration site and the *Hin*dIII restriction site. *M*, *Msp*I digest of pBR322; *Markers I*, pseudointegration products made by ligation of the HIV LTR oligonucleotide into *Sau*3AI-cleaved fragments of the target DNA (at positions 30, 86, and 112 from the *Hin*dIII site); *Markers II*, 5′-labeled fragments following a DNase I digest of chicken erythrocyte nuclei. In this experiment, all integration sites were mapped by uniform labeling during PCR, except for *lanes 5–7* of *panel A*, which were mapped using end-labeled target specific primer (D) (see "Materials and Methods"). Integration patterns for free DNA (*Free*) and nucleosomal DNA (*Nuc.*) are shown.

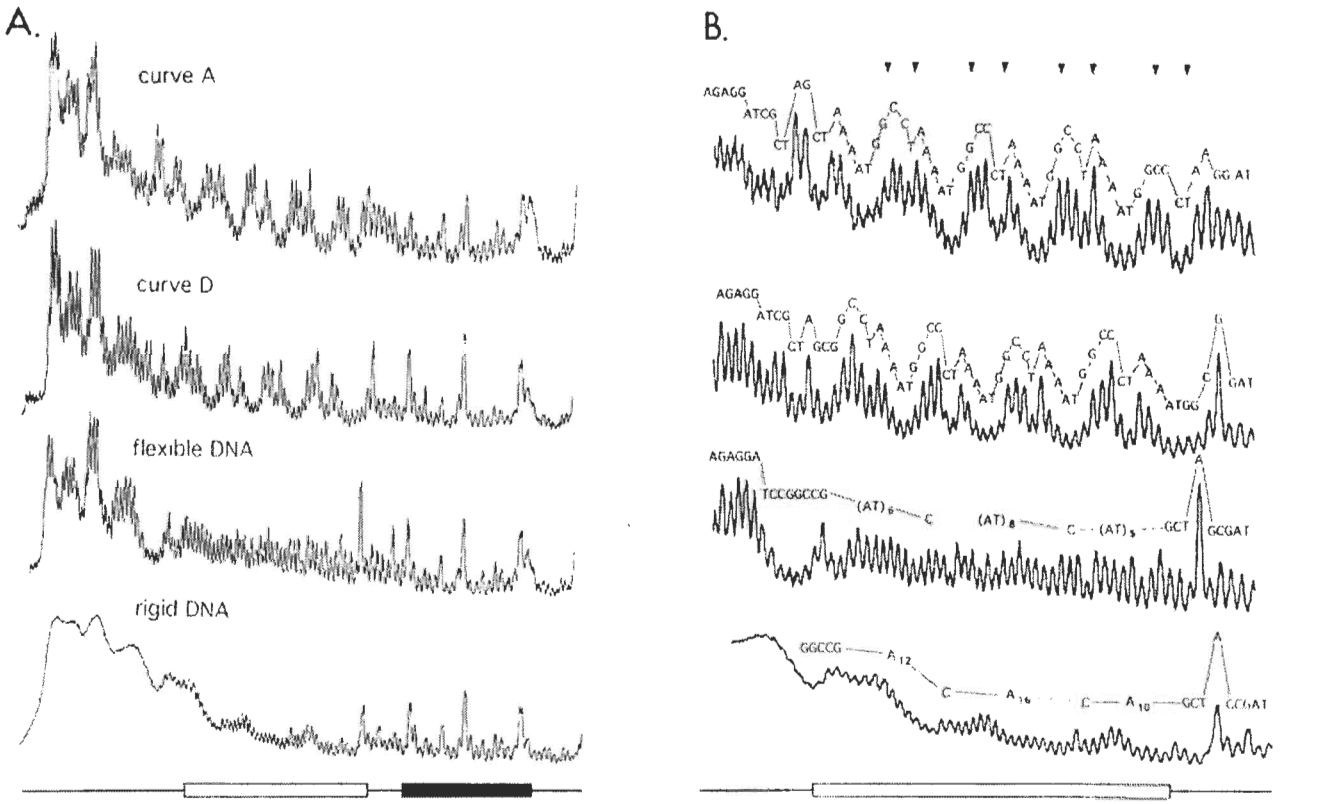

FIG. 4. **Integration profiles of the top strands of the constructs (in naked DNA).** Superimposed densitometer scans of autoradiographs of three sequencing gels are shown in *panel A*, and the blown-up partions encompassing the structured DNA inserts are shown in *panel B* along with the DNA sequences. *Open box*, DNA insert of known structure; *closed box*, T7 promotor.

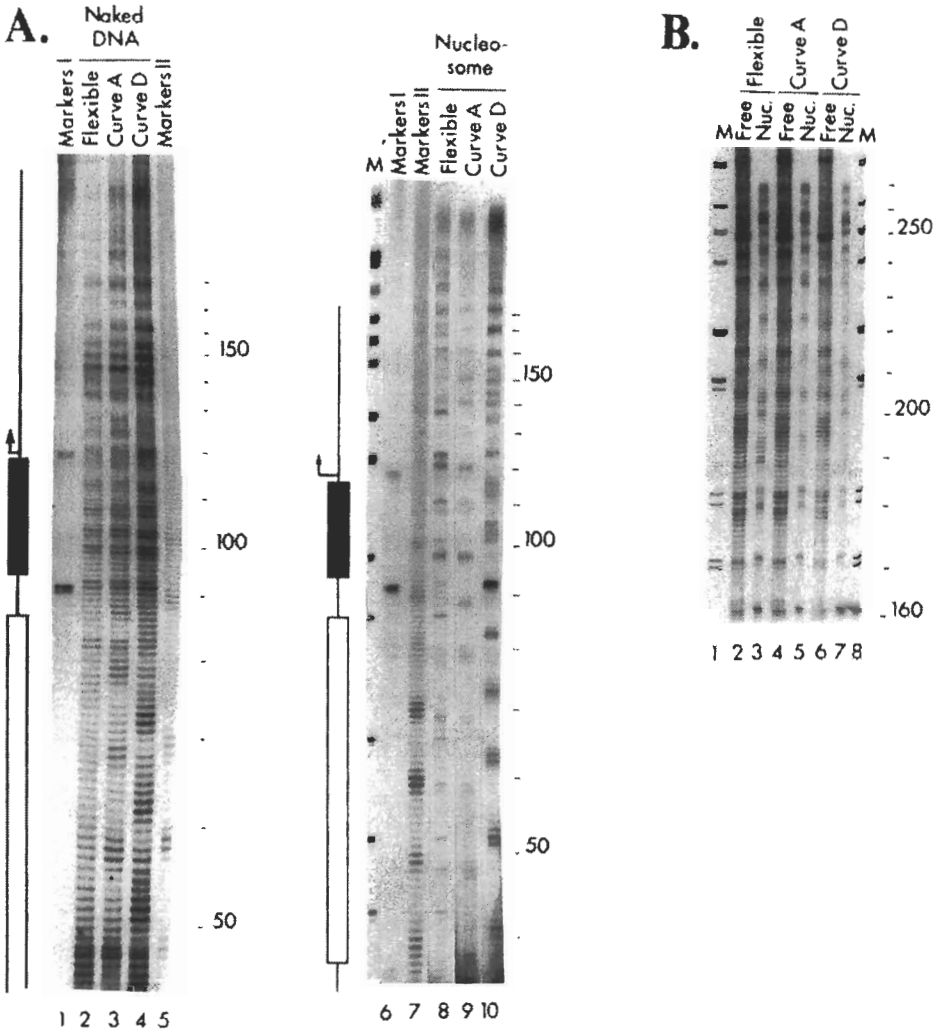

FIG. 5. **Integration footprints of the bottom strand of the DNA constructs.** *A*, the whole of the DNA construct is shown; *B*, the DNA sequence common to all of the constructs including the CAT gene is shown. Designations are the same as in Fig. 1. *Markers I* correspond to "integration events" at positions 35, 91, and 117 on this strand. The primers used were A (*panel A*) and B (*panel B*). In this experiment, all integration events were mapped by uniform labeling during PCR except for *lanes 2–4* of *panel A*, where end-labeled primer A was used. Integration patterns for free DNA (*Free*) and nucleosomal DNA (*Nuc.*) are shown.

DNA segment (Wolffe and Drew, 1989).

DNA Structural Influences on HIV Integrase—We find that each of the test DNA regions in solution (Fig. 3, open box) are utilized as templates for integration (Fig. 3*A*, *lanes 2–4*; Fig. 3*B*, *lanes 2*, *4*, and *6*; and Fig. 3*C*, *lane 1*). However, the frequency of integration over these sequences is not uniform (Fig. 4, *open box*). The curved DNA shows strong modulations in integration frequency, whereas flexible and rigid DNA show more uniform patterns of integration. The T7 promoter in solution serves as a convenient internal control, giving a similar pattern of integration for each fragment (Fig. 4*A*, *black box*). Within the DNA curves, two sequence elements are clearly favored for integration, the GGCC sequences in which the minor groove will lie on the outside of the DNA curve and the TpA step immediately adjacent to the GGCC sequence (Fig. 4*B*, *arrows* above *curve A*, see Fig. 11*A*). The bifurcated peaks of integration are distributed over 6 bp. The AAAA/TTTT sequences show a progressive reduction in integration from 5′ to 3′, and the ApT dinucleotide step is the site at which minimum integration occurs. Only minor differences in integration at the TpA and ApT steps are seen in the flexible DNA, indicating that the frequency of integration also depends on the flanking DNA sequence, although the extended oligo(dA)·oligo(dT) sequence shows much reduced integration (Fig. 4*B*).

The separation between peaks of integration in the curved DNA is approximately 10–11 bp, reflecting the reiteration of the sequence. Mapping of integration sites on the opposite strand (Fig. 5*A*, *lanes 2–4*) also reveals a 10–11-bp separation of peaks. The peaks of integration on this strand are within the GGCC sequence with a shoulder over the TA step (Fig. 6*A*, see Fig. 11*A*). A 1–2-bp stagger exists between integration events on opposite strands of naked curved DNA (Fig. 6*B*); this contrasts with the 5-bp stagger between strands previously observed on nucleosomal targets (Pryciak and Varmus, 1992; Pruss *et al.*, 1994) (see "Discussion").

Outside of the structured DNA region, selection of integration sites is the same over the identical DNA sequences (Figs. 3–5). This result implies that different DNA structures in one part of the DNA molecule do not transmit any influence on integrase activity. This is in contrast to their influence on RNA polymerase activity (see Collis *et al.* (1989)). Normalizing the efficiency of integration in the structured DNA regions relative to integration in a comparable length of the CAT gene reveals that curved DNA is favored as a substrate for integrase relative to flexible or rigid DNA (Fig. 7) (see "Discussion").

Integration within the Nucleosome—The assembly of the 271-bp DNA fragments into dinucleosomes leads to major changes in integration pattern. These changes occur both within and outside of the test DNA regions and reflect both the translational positioning of the histone octamer with respect to DNA sequence and the rotational positioning of DNA with respect to the surface of the histone octamer (Wolffe and Drew, 1989). For the test DNA segments, the resulting integration patterns within the nucleosome are still influenced by the primary DNA sequence; however, general features emerge. Peaks of integration are separated by 10–11 bp, dependent on the rotational positioning of DNA on the surface of the histone octamer, which is itself determined by DNA secondary struc-

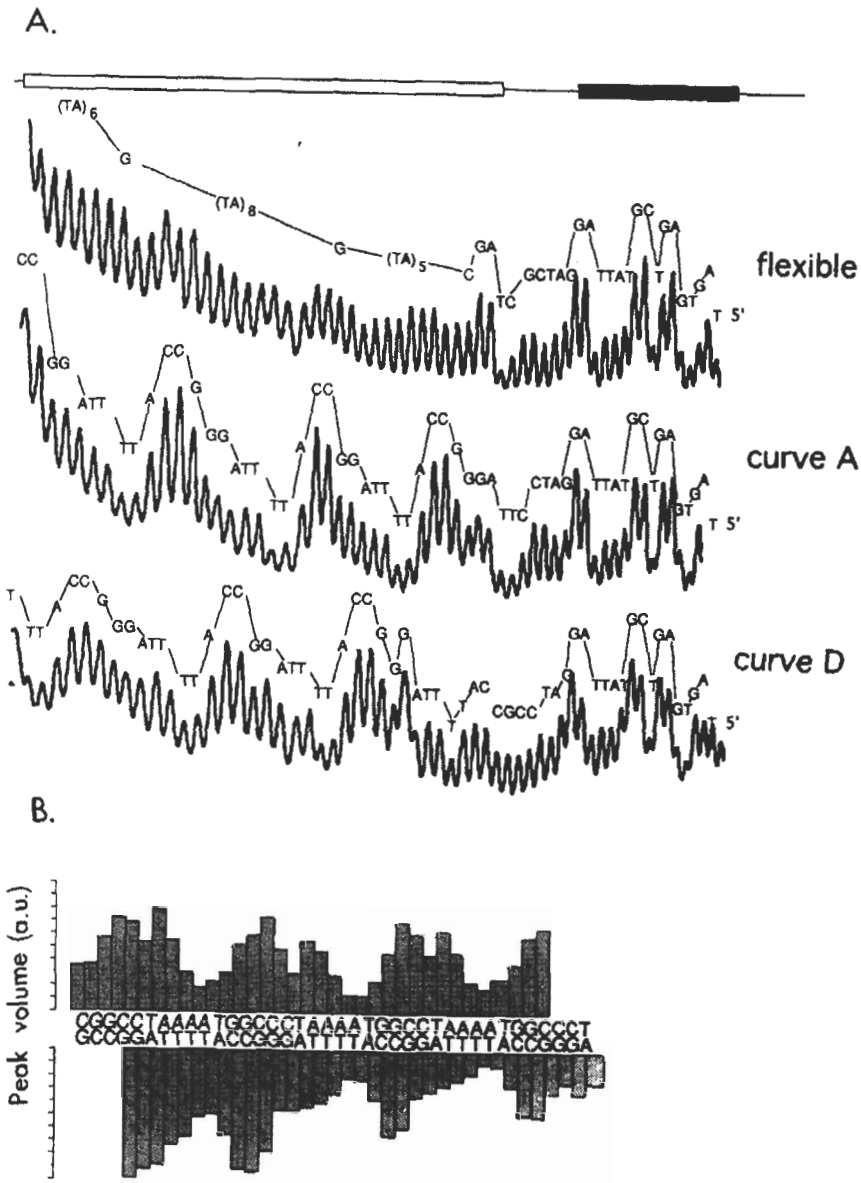

FIG. 6. **Integration profiles of the constructs (in naked DNA).** *A*, superimposed densitometer scans of autoradiographs of two sequencing gels are shown along with the DNA sequences (bottom strand). *Open box*, DNA insert of known structure; *closed box*, T7 promotor. *Hooked arrow* indicates start site of transcription of the CAT gene. *B*, yields of integration into phosphodiester bonds of the two strands of the naked curve D DNA insert. Densitometer scans of the curve D integration products (see Figs. 6*A* and 4*B*) are integrated and plotted.

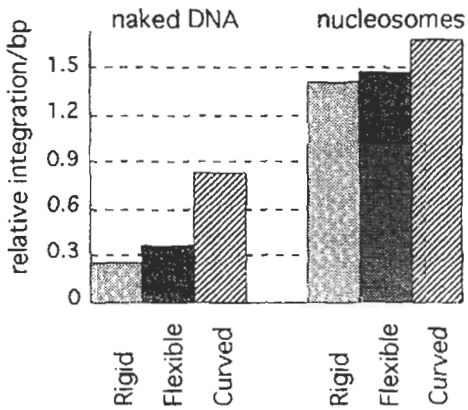

FIG. 7. **Overall efficiency of the integration into various DNA segments of known structure before and after the nucleosome assembly.** Total peak volume (arbitrary units) of the insert of DNA of known structure (top strand between positions 47 and 87) divided by the total peak volume of the CAT gene (115–169) (see Fig. 10).

ture. The integrase is only able to gain access to DNA, where it is exposed toward solution and is not in direct contact with the surface of the histone octamer. This occurs once per helical turn of DNA within the nucleosome (Figs. 3 and 8). For the curved DNAs, the positions of peaks of integration resemble those of naked DNA (Fig. 3, compare naked, *lanes 3* and *4*; with nucleosomal, *lanes 6* and *7*). However the distribution of peak integration sites is much narrower (compare scans of Fig. 4*B* with those of Fig. 8, see Fig. 11). Integration at the GGCC sequence remains high, but that at the TpA step is now suppressed. This result is consistent with the wrapping of DNA around the nucleosome that occurs with the minor groove of the GGCC sequence exposed toward solution as predicted (Satchwell *et al.*, 1986). The TpA step is close to the histone surface and, hence, might be expected to be inaccessible to the integrase enzyme. The similarities between the cleavage pattern of naked DNA in the curves and that in the nucleosome further substantiate the prediction that the naked DNA used in these experiments is intrinsically curved in a similar manner to that found in the nucleosome (Wolffe and Drew, 1989). Moreover the integrase enzyme might be expected to approach DNA preferentially from the outside of the curve. In this conformation DNA bends away from the enzyme.

In contrast to these similarities, the flexible oligo[d(A-T)] tract shows a major change in integration pattern following wrapping around the histone octamer (Fig. 3*A*, compare *lanes 2* and *5*; and Fig. 3*B*, compare *lanes 2* and *3*). The uniform pat-

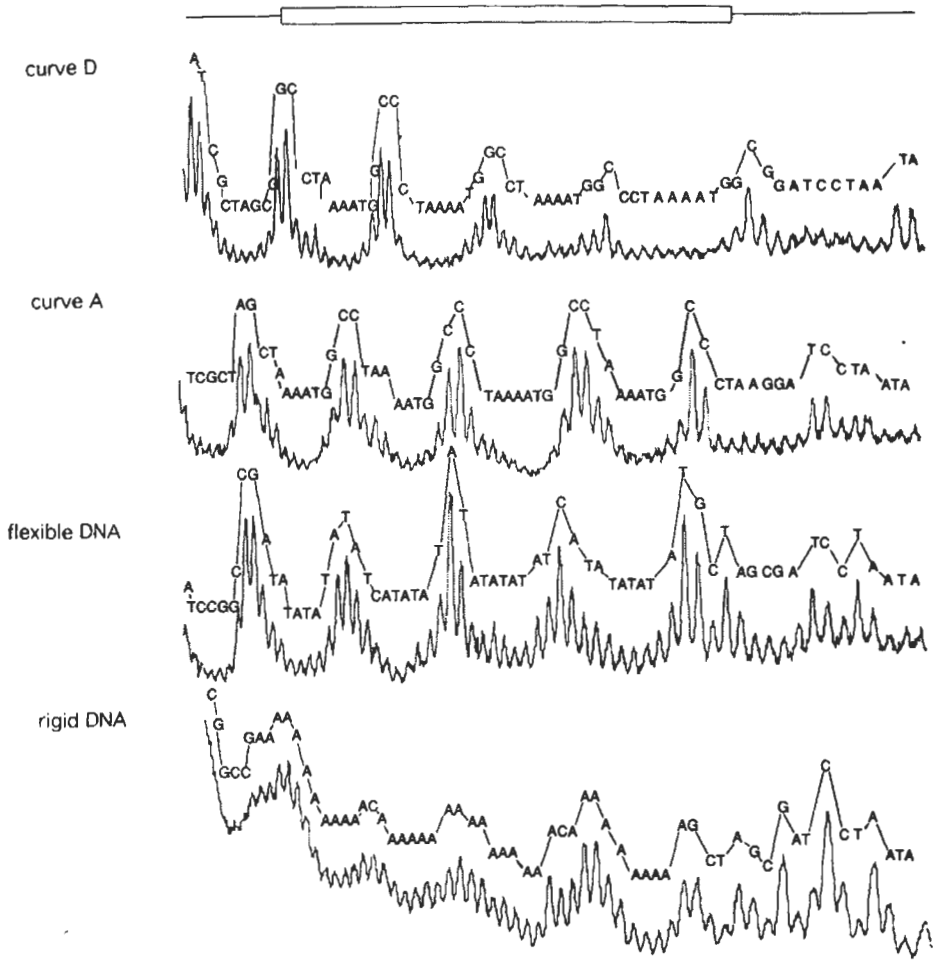

FIG. 8. **Integration profiles of the top strands of the various DNA segments of known structure after nucleosome assembly.** Superimposed densitometer scans with the DNA sequences indicated. *Open box*, structured DNA insert.

tern of integration seen with naked flexible DNA is changed to one of strongly modulated peaks and valleys within the nucleosome. These modulations are very similar to those seen with curve A following association with the histones. This similarity between the organization of flexible DNA and curve A DNA within the nucleosome has been previously observed after hydroxyl radical cleavage (Hayes *et al.*, 1991b) or DNase I cleavage (Fig. 2). These changes in integration site selection from those observed within the naked flexible DNA reflect its distortion as a consequence of wrapping around the histones.

Rigid oligo(dA)·oligo(dT) DNA can also be distorted around the histone octamer on nucleosome formation, although such distortion is energetically unfavorable (Hayes *et al.*, 1991b). Such distortion is not easily revealed by DNase I (Fig. 2, compare *lanes 11* and *12*) but is readily observed with hydroxyl radical cleavage (Hayes *et al.*, 1991b). This difference in cleavage efficiency is presumably due to the sequence selectivity of DNase I cleavage, since the enzyme does not favor rigid oligo(dA)·oligo(dT) as a substrate (Drew, 1984). Similar considerations may explain the weak and diffuse modulation of integration observed over this structured DNA region in the nucleosome (Fig. 3*C*, compare *lanes 1* and *2*; Fig. 8). This can be most clearly seen when the rigid DNA segment with a low frequency of integration is flanked by regions of normal DNA in the nucleosome exhibiting much higher integration frequencies (Fig. 9, *lanes 3* and *9*). It is possible that integrase so disfavors oligo(dA)·oligo(dT) as a target (Fig. 7) that even on wrapping around the histone octamer, integration is severely reduced. A convenient control in these experiments is the integration observed into the adjacent nucleosome that includes the CAT gene, which shows very similar integration patterns for all four DNA fragments within the nucleosome (Figs. 5*B* and 9). Normalizing the integration occurring in the DNA of known structure relative to that in the CAT gene reveals that compared with naked DNA, DNA within the nucleosome is favored for integration. However, even within the nucleosome, oligo(dA)·oligo(dT) tracts are underutilized (Fig. 7). Nevertheless, the segment of DNA containing the rigid sequence is surprisingly a better target than CAT DNA in nucleosomes (Fig. 7). A potential explanation for the enhanced integration frequency of rigid DNA in the nucleosome comes from an observation that the nucleosome formation over oligo (dA)·oligo (dT) tracts is relatively energetically unfavorable (Hayes *et al.*, 1991b). One can argue that these less stably bound histone octamers partially dissociate during integration. To check nucleosome stability in the presence of a large excess of the enzyme, we used agarose gel electrophoresis (not shown). We had to incubate nucleosomes at higher salt concentrations in these assays (0.15 M instead of ~0.04 M) since, at lower ionic strength, integrase readily forms a partially insoluble agregate. With at least three times higher enzyme concentration than we used for the integration reactions, we observed what seemed to be a very low level of histone displacement (not shown). We assume that at lower salt and enzyme concentrations, the nucleosomes are more stable, yet we cannot exclude the possibility of limited octamer dissociation.

Finally, we examined the stagger between integration sites on opposite DNA strands within the nucleosome for curve D compared with the CAT DNA sequence. We found that for curve D, there is a 2–3-bp stagger, whereas for the CAT gene, there is 5–6-bp stagger (Fig. 10). Both of the values are similar to those obtained using naked DNA (Fig. 5, data not shown). Thus, local

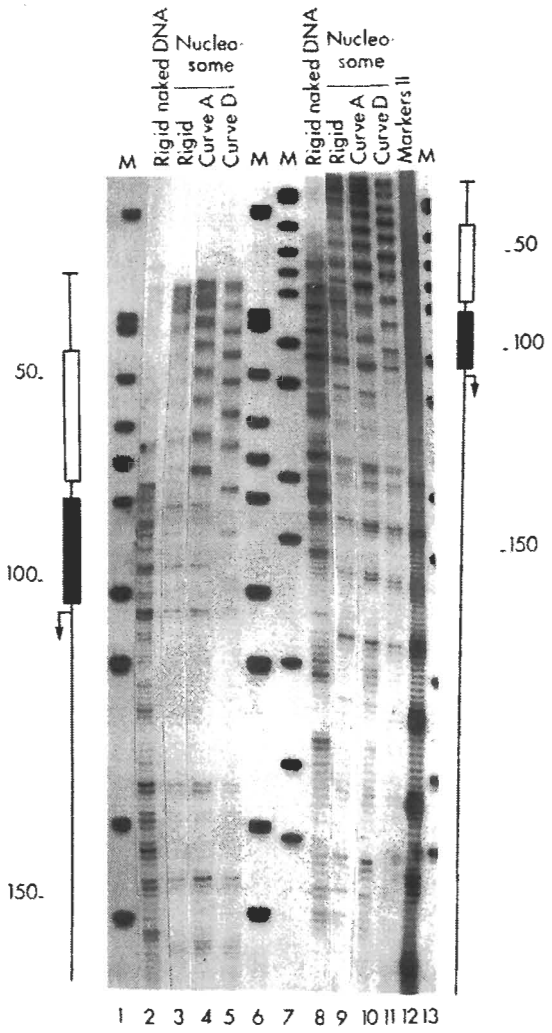

FIG. 9. **Integration profiles of the top strand of the constructs in naked and nucleosomal targets.** Designations are the same as on Fig. 1. Primer E was used for PCR with uniform labeling in order to make possible simultaneous resolution of the integration products in the insert and CAT gene area. Right- and left-hand sections of the gel were loaded with the same samples (left-hand side is a long electrophoretic run and right-hand side a short run of the same samples). *Open box*, DNA insert of known structure. *Closed box*, T7 promoter. *Hooked arrow* indicates start site of transcription of the CAT gene.

DNA sequence can apparently exert a strong influence on the interaction of HIV integrase with the phosphodiester backbone of DNA (see "Discussion").

DISCUSSION

Our experiments were designed to examine what features of free DNA and of DNA within chromatin might promote the selective integration of retroviral genomes (Rohdewohld *et al.*, 1987; Scherdin *et al.*, 1990; Shih *et al.*, 1988; Sandmeyer *et al.*, 1990). All of the unusual test DNA sequences used as targets in our experiments are found *in vivo*: oligo(dA)·oligo(dT) tracts (Valenzuela *et al.*, 1977), oligo[d(A-T)] (Brutlag, 1980), and curved DNA (Marini *et al.*, 1982). In no case do these DNA sequences preclude nucleosome formation (Kunkel and Martinson, 1981; Losa *et al.*, 1990); moreover, the conformation of these structured test DNAs becomes very similar within the nucleosome (Hayes *et al.*, 1991b). The unusual DNA sequences have highly variant properties in solution, they have different helical periods, different minor groove widths, and different flexibilities (see "Strategy"). We find that these properties influence the action of HIV integrase.

Although all of the DNA sequences can be used as targets, curved DNA is favored (Fig. 7). Certain other large protein complexes may favor curved DNA as a binding site, *e.g.* topoisomerase II (Hsieh and Griffith, 1988) and RNA polymerase (Lamond and Travers, 1983; Bossi and Smith, 1984; Gourse *et al.*, 1986; Almouyal and Buc, 1987). This preference is believed to occur from a decrease in the energy required to distort DNA around the protein (Travers and Klug, 1987). However, the preference of integrase for curved DNA appears to differ from that of these other complexes. The sites of integration cluster around where the minor groove of DNA is facing outward, away from the direction of curvature (Figs. 4, 5, and 8). This is not the direction anticipated if integrase binds inside the curve of DNA. Moreover, very similar sequence preferences for integration are seen when DNA is wrapped around the core histones and, hence, is inaccessible for wrapping around the integrase enzyme (Fig. 11). We suggest that HIV integrase actually favors association with DNA that has a similar conformation to that found when it is wrapped around the core histones, *i.e.* that is bent away from the enzyme even when that DNA is naked in solution. The increased frequency of integration into curved DNA is consistent with this hypothesis (Fig. 7).

We find that flexible DNA is favored for integration relative to rigid DNA (Fig. 7). This may reflect the necessity for the integrase to distort DNA or unstack base pairs as part of the reaction mechanism (see Vink and Plasterk (1993) and Pruss *et al.* (1994)). The TpA step is among the most unstable within the double helix (Drew and Dickerson, 1981; Travers and Klug, 1987) and, consequently, might be most easily distorted (Travers and Schwabe, 1993), thereby potentially favoring integration. Even within the curved DNA, where the TpA step lies inside of the curve, this dinucleotide provides a site of enhanced integration efficiency (Figs. 4*B* and 11).

Rigid DNA such as the oligo(dA)·oligo(dT) tracts is particularly difficult to distort (Hayes *et al.*, 1991b); this inflexibility may account for the low levels of integration observed (Figs. 4*B* and 7). Incorporation of DNA into a nucleosome enhances the integration process for all the DNA sequences examined (Pryciak and Varmus, 1992; Pruss *et al.*, 1994). However this is especially true for rigid DNA (Fig. 7) and less so for DNA that already has intrinsic curvature. Our results suggest that it is the conformational change of DNA directed by association with the histones that promotes the integration reaction. DNA within the nucleosome is distorted into an 80-bp circle (Richmond *et al.*, 1984). At certain points in this circle, DNA is more severely bent (Hogan *et al.*, 1987); it is at these sites that integration events preferentially occur (Pruss *et al.*, 1994). These effects are not observed to a significant level in the defined sequence nucleosomes used in the experiments reported here, presumably because sequence specific effects predominate. For example, compare the DNase I cleavage patterns of mixed sequence nucleosomes with those containing defined sequences (Hayes *et al.*, 1991a). Local sequence-specific effects also predominate for DNase I cleavage.

An additional constraint on the DNA shape is imposed by the interaction between the two nucleosomes in the integration buffer, where high concentrations of bivalent cations promote condensation of the linker DNA (Yao *et al.*, 1990; 1991). The nucleosomes formed over both of our specimens of straight test DNA as well as over curve A all adopt about the same rotational position with respect to DNA as the CAT nucleosome. A 10-bp ladder of the integration hotspots continuing across the junction of the two nucleosomes is particularly well seen at Fig. 9 (*lanes 9* and *10*). This is consistent with the idea that the DNA adopts a continuous superhelical path around the stack of the two adjacent histone octamers. In the case of curve D, the

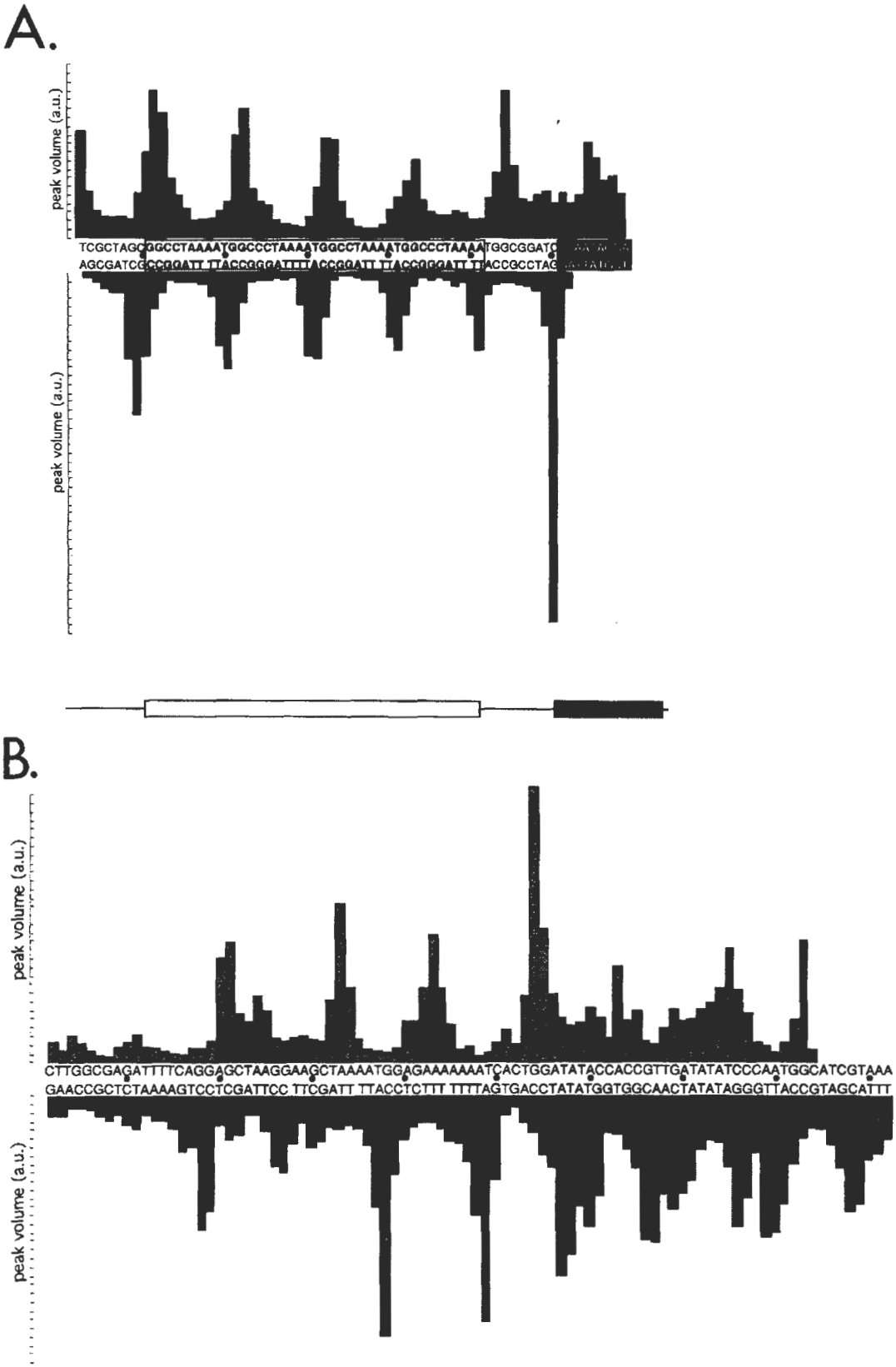

FIG. 10. **Yields of integration (arbitrary units, *a.u.*) into phosphodiester bonds of the two DNA strands of the dinucleosome.** *A*, curve D insert and promotor area; *B*, CAT gene area. The top strand is *above* the bottom strand. *Open box*, DNA insert of known structure; *closed box*, T7 promoter.

preexisting test DNA curvature precludes it from forming a nucleosome with the rotational setting matching the one of the CAT nucleosome. Instead, the two nucleosomes have an approximately 4-bp difference in their rotational phases (Wolffe and Drew, 1989; Fig. 9, lane 11). If DNA within curve D-containing dinucleosomes were to adopt a shape (or superhelical path) similar to the three other DNAs, a major distortion of DNA between the nucleosomes would have to occur. Most strik-

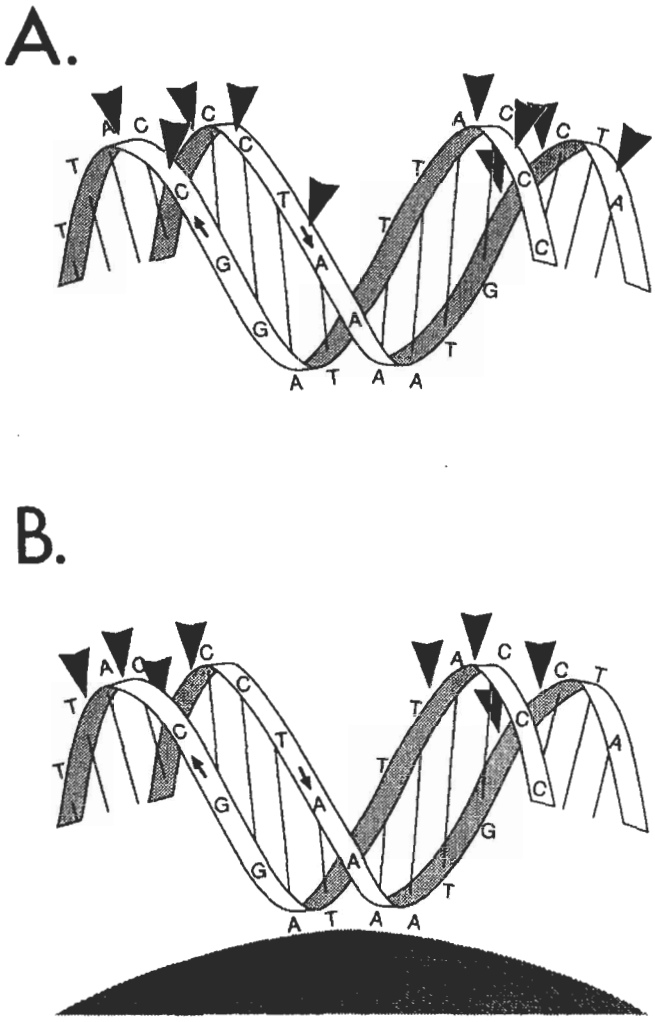

FIG. 11. **Model for integration into curved DNA, free in solution (*A*) or in the nucleosome (*B*).** Integration events are indicated by *arrowheads*, the path of DNA from 5′ to 3′ is indicated by the *arrows*, and the histone octamer is indicated by the *gray shape*.

ingly, it is within this segment of the curve D fragment where we observe the strongest nucleosome-induced integration site of all sites of all 4 constructs (Fig. 5, *lane 8*; Fig. 11*A*). We interpret this enhanced integration as an indication of the dominant effect of nucleosome-nucleosome interaction on the structure of linker DNA, which, in case of markedly different rotational positions of the adjacent nucleosomes, results in significant distortion of the DNA at the junction of the two nucleosomes.

Even when DNA molecules having distinct conformations in solution are constrained into a very similar structure (Hayes *et al.*, 1991b) local DNA sequence can influence integrase activity (Fig. 8). This influence of sequence is especially prominent within rigid DNA, which is relatively refractory to integrase action under all of our experimental conditions.

The presence of a ~2 bp (or alternatively a ~8 bp) stagger between integration events on opposite strands of the double helix for curved DNA, either when naked in solution or in association with the histones, is unexpected (Figs. 10 and 11). This is because previous work both *in vivo* and *in vitro* has indicated that a 5-bp stagger normally occurs between opposite strands within chromosomal substrates (Pryciak and Varmus, 1992; Pruss *et al.*, 1994). However, Pryciak and Varmus (1992) studied the coupled integration of two HIV LTRs mediated by the HIV viral nucleoprotein complex *in vivo*, not the integration of a single HIV LTR mediated by HIV integrase *in vitro* (this work). These studies demonstrated that the two ends of the viral DNA were most commonly inserted into target DNA with a separation of 5 bp from one another. This result is consistent with the viral nucleoprotein complex approaching the target DNA from the major groove, with the two active sites of the nucleoprotein complex interacting with the two opposite DNA strands of the target DNA at a single site across the major groove (Pryciak and Varmus, 1992). In our *in vitro* experiments, we make use of purified HIV integrase, which is an oligomeric protein with two active sites/complex (see Vink and Plasterk, 1993; Craigie, 1992). Using purified integrase, the integration of a single model HIV LTR substrate into target DNA proceeds in an uncoupled fashion, resulting in a single integration event on a single DNA strand at any one time. Thus there are potentially fewer constraints using the purified integrase for selection of integration sites than with the HIV viral nucleoprotein complex (Figs. 10 and 11). Future experiments will examine the sites of contact within the major and minor grooves of DNA between the integrase enzyme, specific target DNAs, and other defined nucleoprotein complexes. This information will be important in further elucidation of the molecular mechanisms facilitating the integration of the HIV genome into the host chromosome.

Acknowledgments—We thank Thuy Vo for manuscript preparation.

REFERENCES

Amouyal, M., and Buc, H. (1987) *J. Mol. Biol.* **195,** 795–808
Bossi, L., and Smith, D. M. (1984) *Cell* **39,** 643–652
Bresnick, E. H., Bustin, M., Marsaud, V., Richard-Foy, H., and Hager, G. L. (1992) *Nucleic Acids Res.* **20,** 273–278
Brukner, I., Dlakic, M., Savic, A., Susic, S., Pongor, S., and Suck, D. (1993) *Nucleic Acids Res.* **21,** 1025–1029
Brutlag, D. (1980) *Annu. Rev. Genet.* **14,** 121–144
Burkhoff, A. M., and Tullius, T. D. (1987) *Cell* **48,** 935–943
Burkhoff, A. M., and Tullius, T. D. (1988) *Nature* **331,** 445–457
Bushman, F. D., and Craigie, R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1339–1343
Bushman, F. D., and Craigie, R. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3458–3462
Calladine, C. R. (1982) *J. Mol. Biol.* **161,** 343–352
Chalker, D. L., and Sandmeyer, S. B. (1992) *Genes & Dev.* **6,** 117–128
Chuprina, V. P. (1985) *FEBS Lett.* **186,** 98–102
Coll, M., Frederick, C. A., Wang, A. H. J., and Rich, A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8385–8389
Collis, C. M., Molloy, P. L., Both, G. W., and Drew, H. R. (1989) *Nucleic Acids Res.* **17,** 9447–9468
Craigie, R. (1992) *Trends Genet.* **8,** 187–190
Drew, H. R. (1984) *J. Mol. Biol.* **176,** 535–557
Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol.* **151,** 535–556
Elgin, S. C. R. (1988) *J. Biol. Chem.* **263,** 19259–19262
Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) *J. Biol. Chem.* **257,** 14686–14707
Goodsell, D. S., Kopka, M. L., Cascio, D., and Dickerson, R. E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2930–2934
Gourse, R. L., DeBoer, H. A., and Nomura, M. (1986) *Cell* **44,** 197–205
Gross, D. S., and Garrard, W. T. (1988) *Annu. Rev. Biochem.* **57,** 159–197
Hayes, J. J., Clark, D. J., and Wolffe, A. P. (1991a) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 6829–6833
Hayes, J. J., Bashkin, J., Tullius, T. D., and Wolffe, A. P. (1991b) *Biochemistry* **30,** 8434–8440
Hogan, M. E., Rooney, T. F., and Austin, R. H. (1987) *Nature* **328,** 554–557
Hsieh, C. H., and Griffith, J. D. (1988) *Cell* **52,** 535–544
Huibregtse, J. M., Evans, C. F., and Engelke, D. R. (1987) *Mol. Cell. Biol.* **7,** 3212–3220
Izban, M. G., and Luse, D. S. (1991) *Genes & Dev.* **5,** 683–696
Ji, H., Moore, D. P., Blomberg, M. A., Braiterman, L. T., Voytas, D. F., Natsoulis, G., and Boeke, J. D. (1993) *Cell* **73,** 1007–1018
Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H., and Geiduschek, E. P. (1989) *Mol. Cell. Biol.* **9,** 2551–2566
Kitamura, Y., Lee, Y. M. H., and Coffin, J. M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 5532–5536
Kunkel, G. R., and Martinson, H. G. (1981) *Nucleic Acids Res.* **9,** 6869–6888
Lamond, A. I., and Travers, A. A. (1983) *Nature* **305,** 248–250
Libertini, L. J., and Small, E. W. (1980) *Nucleic Acids Res.* **8,** 3517–3534
Losa, R., Omari, S., and Thoma, F. (1990) *Nucleic Acids Res.* **18,** 3495–3502
Marini, J. C., Levene, S. D., Crothers, D. M., and Englund, P. T. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 7664–7668
Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 560–564
McPherson, C. E., Shim, E. Y., Friedman, D. S., and Zaret, K. S. (1993) *Cell* **75,** 387–398

Morse, R. H., Roth, S. Y., and Simpson, R. T. (1992) *Mol. Cell. Biol.* **12,** 4015–4025
Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987) *Nature* **330,** 221–226
Pruss, D., Bushman, F. D., and Wolffe, A. P. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5913–5917
Pryciak, P. M., and Varmus, H. E. (1992) *Cell* **69,** 769–780
Pryciak, P. M., Sil, A., and Varmus, H. E. (1992a) *EMBO J.* **11,** 291–303
Pryciak, P. M., Muller, H. P., and Varmus, H. E. (1992b) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 9237–9241
Ramsay, N., Felsenfeld, G., Rushton, B. M., and McGhee, J. D. (1984) *EMBO J.* **3,** 2605–2611
Rhodes, D., and Klug, A. (1981) *Nature* **292,** 378–380
Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984) *Nature* **311,** 532–537
Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987) *J. Virol.* **61,** 336–343
Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) *Science* **230,** 1350–1354
Sandmeyer, S. B., Hansen, L. J., and Chalker, D. L. (1990) *Annu. Rev. Genet.* **24,** 1499–1515
Satchwell, S. C., Drew, H. R., and Travers, A. A. (1986) *J. Mol. Biol.* **191,** 659–675
Scherdin, U., Rhodes, K., and Breindl, M. (1990) *J. Virol.* **64,** 907–912
Shih, C. C., Stoye, J. P., and Coffin, J. M. (1988) *Cell* **53,** 531–537
Tatchell, K., and van Holde, K. E. (1977) *Biochemistry* **16,** 5293–5303
Thomas, G. H., and Elgin, S. C. R. (1988) *EMBO J.* **7,** 7291–7301
Travers, A. A., and Klug, A. (1987) *Philos. Trans. R. Soc. London B Biol. Sci.* **317,** 537–561
Travers, A. A., and Schwabe, J. W. R. (1993) *Curr. Biol.* **3,** 898–900
Valenzuela, P., Bell, G. I., Masiarz, F. R., DeGennaro, J. L., and Rutter, W. J. (1977) *Nature* **267,** 641–643
Varmus, H. E., and Swanstrom, R. (1985) in *RNA Tumor Viruses* (Weiss *et al.*, ed) Vol. 1 and 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Vijaya, S., Stefan, D. L., and Robinson, H. L. (1986) *J. Virol.* **60,** 683–692
Vink, C., and Plasterk, R. H. A. (1993) *Trends Genet.* **9,** 433–437
Wolffe, A. P., and Drew, H. R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 9817–9821
Yao, J., Lowary, P. D., and Widom, J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 7603–7607
Yao, J., Lowary, P. T., and Widom, J. (1991) *Biochemistry* **30,** 8408–8414