Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10334–10338, October 1995
Biochemistry

In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends

YEOU-CHERNG BOR, FREDERIC D. BUSHMAN, AND LESLIE E. ORGEL

The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA 92186-8500

Contributed by Leslie E. Orgel, August 4, 1995

ABSTRACT **Integration of human immunodeficiency virus type 1 cDNA into a target DNA can be strongly influenced by the conformation of the target. For example, integration *in vitro* is sometimes favored in target DNAs containing sequence-directed bends or DNA distortions caused by bound proteins. We have analyzed the effect of DNA bending by studying integration into two well-characterized protein–DNA complexes: *Escherichia coli* integration host factor (IHF) protein bound to a phage IHF site, and the DNA binding domain of human lymphoid enhancer factor (LEF) bound to a LEF site. Both of these proteins have previously been reported to bend DNA by ≈140°. Binding of IHF greatly increases the efficiency of *in vitro* integration at hotspots within the IHF site. We analyzed a series of mutants in which the IHF site was modified at the most prominent hotspot. We found that each variant still displayed enhanced integration upon IHF binding. Evidently the local sequence is not critical for formation of an IHF hotspot. LEF binding did not create preferred sites for integration. The different effects of IHF and LEF binding can be rationalized in terms of the different proposed conformations of the two protein–DNA complexes.**

During the course of normal infection by human immunodeficiency virus type 1 (HIV-1), each end of the linear viral cDNA is connected to the host genome to generate an integrated provirus. Analysis of a limited number of junctions between viral and host DNA formed *in vivo* did not reveal strongly preferred target sequences for integration (1–3). However, *in vitro* studies indicate that integration-site selection is not strictly random (4–6).

Analysis of integration reactions *in vitro* shows that proteins bound to target DNA can affect integration-site selection. Transcriptional repressor proteins bound to specific DNA sites can block integration at those sites, probably by blocking access of the integration apparatus (7, 8). Proteins that deform DNA in the course of binding (e.g., nucleosomes and *Escherichia coli* CAP) can, in addition, promote integration (9, 10).

Here we probe further the effect of target distortion on integration by examining *in vitro* integration into DNA targets bound to *E. coli* integration host factor (IHF) or the DNA binding domain of lymphoid enhancer factor (LEF). IHF protein plays a role in a number of cellular processes, including site-specific recombination, phage packaging, transposition, plasmid replication, and transcription (11). IHF often acts as an architectural element, achieving its diverse biological effects in part by bending DNA and promoting the formation of higher-order nucleoprotein complexes (12–14). LEF is a transcriptional activator protein that promotes expression of the T-cell receptor α-chain gene. LEF also bends DNA and may exert part of its effects on transcription by this mechanism (15–18).

In this study, we used purified integrase protein of HIV-1 as a source of integration activity. Short oligonucleotides mimicking one end of unintegrated HIV-1 DNA serve as model integration donor DNAs (4, 19, 20).

We find that bending of the target DNA by IHF creates preferred sites for integration *in vitro* within the region where the protein binds. A series of mutants was generated in the IHF site near the most prominent IHF-induced hotspot. Analysis of integration into these mutant sites when bound to IHF revealed that local sequence did not control the efficiency of integration, highlighting the importance of the overall conformation of the IHF–DNA complex. Bending of DNA by the LEF DNA binding domain, in contrast, does not create integration hotspots. The difference in behavior of the IHF and LEF complexes can be attributed to the different conformations they adopt.

MATERIALS AND METHODS

Oligonucleotides and Plasmids. DNA oligonucleotides were synthesized with an Applied Biosystems automatic synthesizer and purified by denaturing polyacrylamide gel electrophoresis. Oligonucleotides used as PCR primers were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase and purified using a Sephadex G-25 spin column (Boehringer Mannheim).

IHF sites used in these experiments were constructed from a series of complementary oligonucleotides, each of which contained the site of interest flanked by protruding cohesive ends matching those produced by cleavage with *Eco*RI and *Hin*dIII (see Fig. 3 for sequences). Each oligonucleotide was ligated to pUC19 cleaved with *Eco*RI and *Hin*dIII. IHF sites designated WT, IV, V, VI, and VII (see Fig. 3) were a kind gift from H. Nash and S.-W. Yang (National Institute of Mental Health, National Institutes of Health, Bethesda, MD). A plasmid containing a LEF site was constructed by ligating a synthetic oligonucleotide containing the site flanked by single-stranded 5′-GATC-3′ overhangs (the 29-mer sequence in Fig. 4) to pUC19 cleaved with *Bam*HI and *Bgl* II. The sequences of the inserts in all clones were confirmed by dideoxynucleotide sequencing. General methods have been described (21).

Integration. HIV-1 integrase was overexpressed in *E. coli* and purified as described (22). IHF protein was a gift of H. Nash, and LEF^{HMG88} was a gift of K. Jones (Regulatory Biology Laboratory, The Salk Institute).

The integration reactions were studied by mixing two solutions. The first (C_1) contained IHF or LEF bound to target DNA, while the second (C_2) contained integrase bound to donor DNA. The C_1 mixture was prepared by incubating 0.028 μM IHF or LEF binding sequences with appropriate concentrations of IHF or LEF^{HMG88} proteins for 10 min at 37°C in 10 μl of integration buffer [5 mM $MnCl_2$/20 mM Hepes, pH 7.0/10 mM 2-mercaptoethanol/10% (vol/vol) glycerol] to which 100 mM KCl and bovine serum albumin (2 mg/ml) had been added.

The donor DNA in the mixture C_2 (FB64/65-2; sequence in ref. 8) matched a HIV-1 DNA end. Since normal integration

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

Abbreviations: HIV-1, human immunodeficiency virus type 1; IHF, integration host factor; LEF, lymphoid enhancer factor.

in vivo is preceded by the removal of 2 nt from the 3′ end of the long terminal repeat, we used an appropriate donor DNA with two unpaired bases at its 5′ terminus. C_2 was prepared by incubating a 0.03 μM solution of this DNA (FB64/65-2) with 0.26 μg of integrase for 10 min at 37°C in integration buffer to form a stable integrase–donor complex (23).

After mixing of C_1 and C_2, the solution was incubated for 50 min at 37°C. The reaction was terminated by adding 1 μl of 0.5 M EDTA and 1 μl of 1% SDS. The products were then purified by phenol/chloroform extraction, precipitated with ethanol, and dissolved in 20 μl of 10 mM Tris·HCl, pH 8/1 mM EDTA (TE).

PCR Assay of Integration Products. Integration products were analyzed by PCR as described (7, 8, 24). Each PCR mixture contained U5-specific primer [FB66 (8)] and one of two plasmid-specific primers (forward −20 primer 1211 or reverse sequencing primer 1201; New England Biolabs). Two microliters of the integration products was added to a PCR mixture containing 10 mM Tris·HCl (pH 8.3), 50 mM KCl, 0.1 mM each dNTP, 1.5 mM $MgCl_2$, 0.1 mg of bovine serum albumin per ml, 0.5 μM primers (including 0.025 μM ^{32}P-labeled forward or reverse primer), and 1 unit of *Taq* DNA polymerase in a total vol of 20 μl. The denaturation, annealing, and extension steps were performed at 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min. After 30 cycles, the reaction mixture was incubated at 72°C for 10 min to extend incomplete products. The PCR products were resolved on an 8% denaturing polyacrylamide gel. The results were quantitated with a Molecular Dynamics PhosphorImager.

To confirm our assignments of sizes to the integration products, in representative experiments, amplification products were cleaved with *Eco*RI and analyzed by electrophoresis on sequencing gels. Comigration of the cleaved DNAs with sequencing markers of the expected sizes confirmed the reliability of sequencing ladders as markers (data not shown).

RESULTS

Experimental Plan: General Description. Two protein–DNA complexes were studied as integration targets in order to probe the effects of DNA distortion on integration complexes composed of (*i*) *E. coli* IHF protein bound to the H′ site from phage λ*attP* and (*ii*) the DNA binding domain of the human LEF protein (LEF^{HMG88}) bound to a site in the T-cell receptor α-chain enhancer. The LEF DNA binding domain is named LEF^{HMG88} because it is composed of 88 amino acids and resembles in sequence the HMG proteins (18, 25). IHF bends DNA through 140° (26); LEF^{HMG88} bends DNA through 130° (18).

Two types of protein–DNA complexes were preassembled prior to starting the integration reaction. IHF or LEF was prebound to its site in one reaction mixture, and HIV-1 integrase was assembled with a donor oligonucleotide matching one viral DNA end (23) in another. The two reaction mixtures were then mixed and incubated for 50 min at 37°C.

The frequency of integration at each site was monitored by a PCR-based assay (Fig. 1) (7, 8, 24). After the integration reactions were stopped, the products were purified and used as PCR substrates. Primers complementary to the target DNA and the viral DNA end were used for amplification. The target DNA primer was 5′ ^{32}P-labeled. The PCR products were denatured, separated by electrophoresis on a DNA sequencing-type gel, and visualized by autoradiography. Each band on the autoradiogram represented the product of integration at a unique phosphodiester bond. The band intensity was proportional to the frequency of integration at that bond. PCR primers complementary to one or another strand of the target DNA were used in separate reactions to examine integration into each strand.

Integration into the IHF–DNA Complex. The results of the PCR-based assay of integration into the IHF–DNA complex are presented in Fig. 2. In the absence of IHF protein, a radioactive band could be detected at most positions in the product DNA ladder, corresponding to integration into most phosphodiester bonds in the target DNA. The relative frequency of integration events varied over a 25-fold range. Inclusion of EDTA to chelate the divalent metal ion required during the integration reaction blocked production of the ladder of bands (Fig. 2 *A* and *B*, lanes 8; see also Fig. 4), confirming that the generation of the PCR product depended on integration.

When the IHF protein was added to the integration reaction mixture, two sites were used substantially more efficiently as integration targets (Fig. 2*A*, top strand, bp 65; Fig. 2*B*, bottom strand, bp 60). At the most prominent hotspot, nucleotide 65 on the top strand, the efficiency of integration was enhanced ≈15-fold. Integration into the hotspot on the bottom strand (nucleotide 60) was increased 3-fold.

Apart from these hotspots, integration into the IHF binding site was often diminished, probably because bound IHF blocked access of the integration machinery to the DNA. Integration into the most protected phosphodiester bonds was reduced ≈10-fold with respect to uncomplexed DNA. The DNA footprint generated by interference with integration

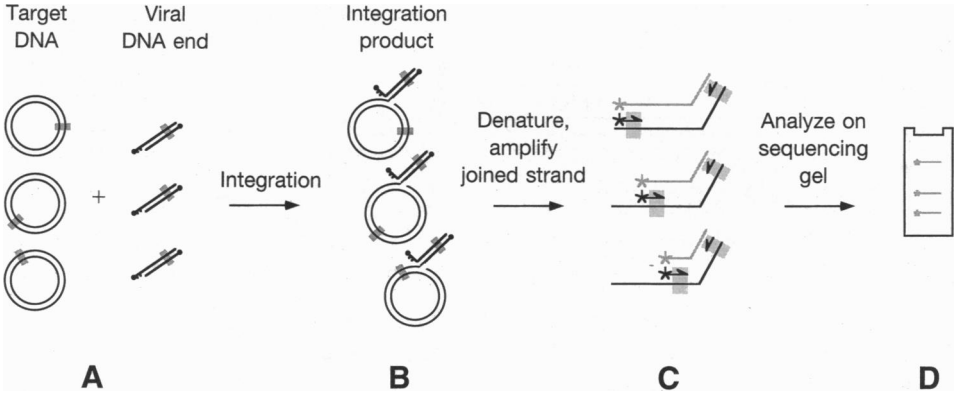

FIG. 1. Diagram illustrating the PCR assay of HIV-1 integration. During integration, viral DNAs are inserted at many sites in the target DNA. Two primers, one complementary to a target DNA and the other complementary to the viral DNA, are used for PCR. The target primer is labeled with ^{32}P on its 5′ end. PCR amplification of integration products generates a population of molecules that are then denatured and separated on a sequencing gel. The sizes of the amplification products are determined by comparison to a DNA sequencing ladder. The positions of integration sites in target DNA were deduced from the sizes of PCR products. Insertion at position N of the target DNA implies that the viral–target junction occurs between bases N and N + 1 (5′ to 3′) of the target.

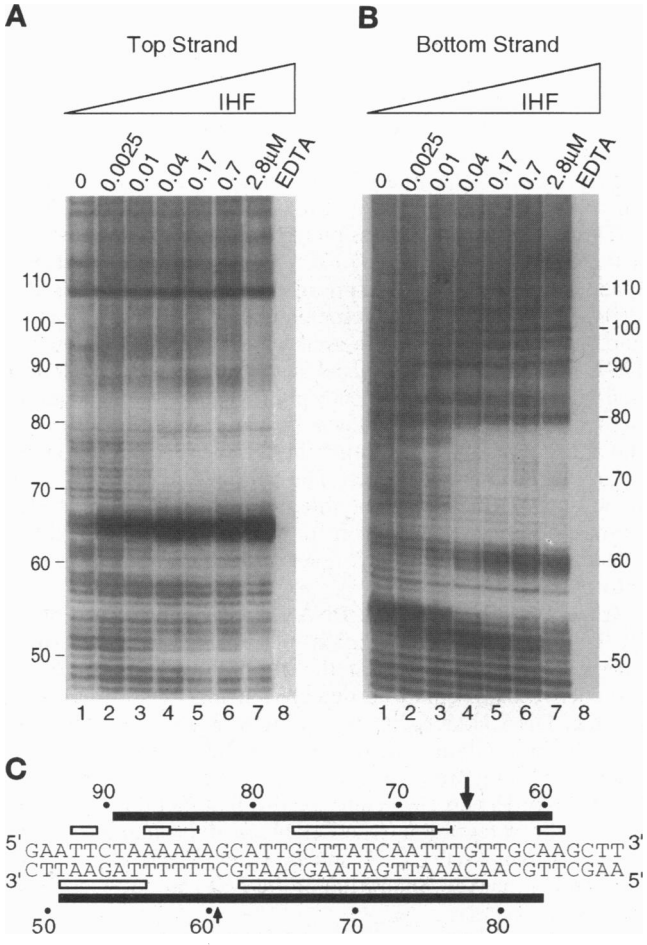

FIG. 2. PCR analysis of integration into a complex of IHF bound to the H′ DNA site. Results of amplifications using a primer annealing to the top strand (*A*) or the bottom strand (*B*) of the target DNA are shown. Amount of IHF added to each integration reaction is indicated. Sizes (bp) of amplification products, determined by comparison to a DNA sequencing ladder, are as marked. Numbers refer to the distance from the 5′ end of the target DNA primer used for amplification. (*C*) Sequence of the IHF binding site is shown. Regions protected from integration are marked by shaded boxes. Horizontal bar marks uncertainty in the margins of the footprint. Arrows indicate sites that are used more frequently as an integration target upon IHF binding. Solid boxes mark location of the IHF site as defined by protection from digestion with DNase I (27).

(Fig. 2*C*, shaded bars) corresponds quite closely to the IHF footprint seen using pancreatic DNase I as probe (Fig. 2*C*, solid bars) (27).

In some cases, effects of added IHF were seen outside the expected binding site (e.g., Fig. 2*A*, positions 50–52; data not shown), probably due to binding to additional IHF sites.

Effect of Local Sequences on an Integration Hotspot. To examine the effect of local sequence on the most prominent IHF-induced integration hotspot (residue 65; Fig. 2, top strand), several mutant sequences were prepared and the effects on integration were determined. Additional mutant IHF sites were kindly provided by S.-W. Wang and H. Nash. Three altered sites were made in which the G residue at the top-strand hotspot was changed to one of the other bases (Fig. 3, mutants I, II, and III). Mutants IV–VII contained changes elsewhere in the binding site. All mutants except mutant VII were found to bind IHF detectably as assayed by formation of a region protected from integration. However, for some mutants protection was less complete than with the wild type, particularly on the bottom strand (data not shown). The hotspot on the top strand was seen with all mutants except mutant VII, which did not bind. The stimulation of integration found upon IHF binding ranged from 7- to 20-fold. These data suggest that the local sequence at the position 65 hotspot is not of major importance in determining the efficiency of integration.

Integration into the LEF–DNA Complex. In further experiments, integration into complex of LEF^{HMG88} bound to a LEF site was studied by the methods used for the above studies of the IHF complex (Fig. 4). With increasing concentrations of LEF, a region of diminished integration became apparent on the top strand, corresponding to the previously determined LEF binding site (Fig. 4*C*) (15). Only weak footprinting over the LEF site was seen on the bottom strand. No sites of strongly enhanced integration were seen when LEF bound to the target, although some weak enhancements were seen just outside the protected area.

DISCUSSION

We and others have found that binding of proteins to target DNA can enhance or diminish retroviral DNA integration *in vitro* in the neighborhood of the bound protein. Binding of transcriptional repressors such as Matα2 or λ repressor blocks integration (7, 8), while binding of nucleosomes or *E. coli* CAP protein, which distort DNA, enhances integration (7, 9, 10, 28). We have further characterized the influence of DNA distortion by analyzing integration into two additional protein–DNA complexes in which the DNA is thought to be distorted from straight B-form—namely, *E. coli* IHF bound to a phage λ site (13) and LEF^{HMG88} bound to a site from the T-cell receptor enhancer (16, 18). Binding of IHF to its site created a strong integration hotspot. Altering the sequence at the hotspot did

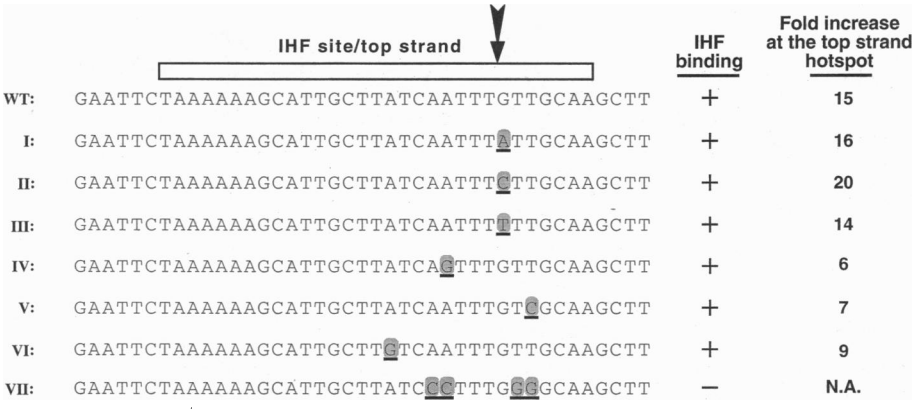

FIG. 3. Summary of hotspot formation in the IHF sites studied. Arrow marks location of the top-strand hotspot. Binding of IHF (+) indicates the presence of a half-maximal footprint on the top strand in the presence of no more than 40 nM IHF. Changes in the IHF site are shown boxed and underlined. *Eco*RI and *Hin*dIII sites used for cloning are shown at each end of the sequence.

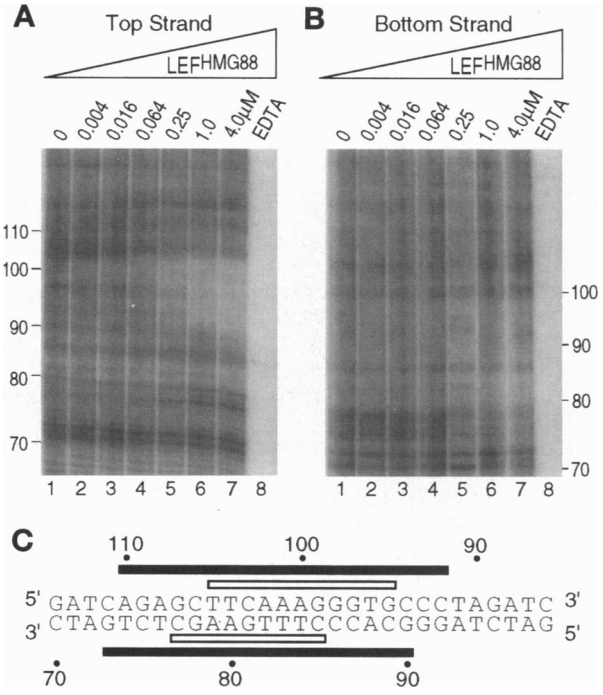

FIG. 4. (*A* and *B*) Integration into a LEF protein–DNA complex. Concentration of LEF protein in each reaction is shown. (*C*) Sequence in the region of the LEF site is shown. The extent of the integrase footprint is shown by shaded bars. Location of sites defined by DNase I footprinting are shown by solid bars. GATC sequence used for cloning is shown at each end of the sequence.

not substantially affect preferential integration at this site. However, binding of LEF to its site did not create a hotspot, indicating that DNA bending does not always result in enhanced integration.

Several experimental studies indicate that DNA distortion may promote the formation of integration hotspots (7, 9, 10). Integration *in vitro* occurs on the major groove side of DNA (7, 9) and widened major grooves in the nucleosome have been found to be integration hotspots (9). However, other forms of DNA distortion may also promote integration. Integration hotspots in the CAP–DNA complex occur at major grooves adjacent to a widened minor groove (10). Integration is also enhanced at A-tract DNA sequences containing static bends. These sequences display enhanced reactivity with hydroxyl radicals, which is believed to be due to distortion, possibly from widening of the minor groove (29, 30). Integration is enhanced at these same sequences. It is interesting that the integration hotspots in the IHF–DNA complex also fall in regions of enhanced hydroxyl radical sensitivity (14). This is consistent with the idea that the hotspots found here also lie in or near sites of DNA distortion.

Mutations in the IHF site in the region of the IHF-induced hotspot on the top strand did not significantly influence the efficiency of integration. Evidently, the structural constraints imposed by wrapping the DNA around IHF protein dominated constraints imposed by local DNA sequence. A parallel results has been reported in comparisons of integration efficiency in DNA sequences bound to nucleosomes or free in solution. In this study, each naked DNA displayed a characteristic pattern of integration hotspots and cold spots, but the differences were less pronounced when the DNA was wrapped on nucleosomes (28).

The IHF protein–DNA complex has been modeled based on the known structure of the related HU protein of *Bacillus stearothermophilus* (31) and biochemical studies (14). IHF protein is a heterodimer that binds to a 30-bp nonsymmetrical site. IHF recognizes the DNA site primarily through contacts

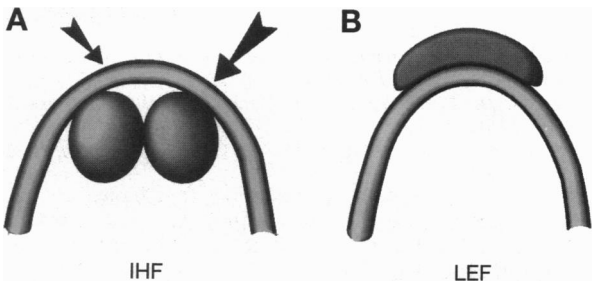

FIG. 5. Schematic diagrams of the IHF–H′ complex (*A*) and the LEF–DNA complex (*B*). Each protein is shown stippled. DNA is shown as the bent cylinder. Arrows mark approximate location of integration hotspots in the IHF site.

to the base pair edges on the minor groove side of the helix (14, 32).

The hotspots for integration lie on the outside of the proposed IHF–DNA complex, within DNA sequences that are expected to be exposed to solvent (Fig. 5*A*). The asymmetry in the placement and magnitude of the hotspots likely reflects the asymmetry in the IHF–DNA complex. Many sites on the inner face of the curved DNA are protected from integration, as expected if the IHF protein lies on the inner surface (32).

In the study of integration into a complex of LEF^{HMG88} bound to DNA, we found that LEF binding did not result in the formation of significant hotspots for integration. This was surprising, since LEF^{HMG88} bends DNA to almost the same extent as does IHF (13, 18, 26), and both bind specifically to base-pair edges on the minor groove side of DNA (14, 16). Binding of LEF^{HMG88} to its recognition site resulted primarily in inhibition of HIV-1 integration in the region of the binding site.

A model for the LEF–DNA complex provides one possible explanation for this difference. LEF is a member of the HMG/SRY family of DNA binding proteins. Models based on structural and genetic data (33, 34) place proteins of this class on the outside of the protein-induced bend (Fig. 5*B*), unlike IHF, which binds to the inside of the bend (14) (Fig. 5*A*). LEF-induced bending may thus be unfavorable for integration because LEF protein itself lies over the region of the greatest distortion in the DNA.

Note added in proof. A recent NMR study by Wright and co-workers (35) confirms that the DNA in the LEF–DNA complex is bent and that the protein is on the outside of the bend.

We thank Verna Stitt for the figures and Sylvia Bailey for manuscript preparation. F.D.B. is a Special Fellow of the Leukemia Society of America. We thank Howard Nash, S.-W. Yang, and K. Jones for materials and K. Jones, H. Nash, and members of the Salk Institute Infectious Diseases Laboratory for comments on the manuscript. This work was supported by Grants AI 29850-04 and AI 34786 from the National Institutes of Health and Grant R94-SAL-104 from the University of California Universitywide AIDS Research Program.

1. Vincent, K. A., York-Higgins, D., Quiroga, M. & Brown, P. O. (1990) *Nucleic Acids Res.* **18,** 6045–6047.
2. Vink, C., Groenink, M., Elgersma, Y., Fouchier, R. A. M., Tersmette, M., Plasterk, R. H. A. (1990) *J. Virol.* **64,** 5626–5627.
3. Stevens, S. W. & Griffith, J. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5557–5561.
4. Bushman, F. D. & Craigie, R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1339–1343.
5. Leavitt, A. D., Rose, R. B. & Varmus, H. E. (1992) *J. Virol.* **66,** 2359–2368.
6. Vink, C., van Gent, D. C., Elgersma, Y. & Plasterk, R. H. A. (1991) *J. Virol.* **65,** 4636–4644.
7. Pryciak, P. M. & Varmus, H. E. (1992) *Cell* **69,** 769–780.
8. Bushman, F. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9233–9237.

9. Pruss, D., Bushman, F. D. & Wolffe, A. P. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5913–5917.
10. Müller, H.-P. & Varmus, H. E. (1994) *EMBO J.* **13,** 4704–4714.
11. Nash, H. A. (1990) *Trends Biochem. Sci.* **15,** 222–227.
12. de Vargas, L. M., Kim, S. & Landy, A. (1989) *Science* **244,** 1457–1461.
13. Robertson, C. A. & Nash, H. A. (1989) *J. Biol. Chem.* **263,** 3554–3557.
14. Yang, C.-C. & Nash, H. A. (1989) *Cell* **57,** 869–880.
15. Waterman, M. L. & Jones, K. A. (1990) *New Biol.* **2,** 621–636.
16. Giese, K., Cox, J. & Grosschedl, R. (1992) *Cell* **69,** 185–195.
17. Giese, K., Kingsley, C., Kirshner, J. R. & Grosschedl, R. (1995) *Genes Dev.* **9,** 995–1008.
18. Sheriden, P. L., Sheline, C. T., Cannon, K., Yoz, M. L., Pazin, M. J., Kadonaga, J. T. & Jones, K. A. (1995) *Genes Dev.*, in press.
19. Craigie, R., Fujiwara, T. & Bushman, F. (1990) *Cell* **62,** 829–837.
20. Katz, R. A., Merkel, G., Kulkosky, J., Leis, J. & Skalka, A. M. (1990) *Cell* **63,** 87–95.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
22. Sherman, P. A. & Fyfe, J. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 5119–5123.
23. Ellison, V. & Brown, P. O. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 7316–7320.
24. Kitamura, Y., Lee, Y. M. & Coffin, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5532–5536.
25. Carlsson, P., Waterman, M. L. & Jones, K. A. (1993) *Genes Dev.* **7,** 2418–2430.
26. Thompson, J. F. & Landy, A. (1988) *Nucleic Acids Res.* **16,** 9687–9705.
27. Craig, N. L. & Nash, H. A. (1984) *Cell* **39,** 707–716.
28. Pruss, D., Reeves, R., Bushman, F. D. & Wolffe, A. P. (1994) *J. Biol. Chem.* **269,** 25031–25041.
29. Burkhoff, A. M. & Tullius, T. D. (1988) *Nature (London)* **331,** 455–457.
30. Nagaich, A., Bhattacharyya, D., Brahmachari, S. K. & Bansal, M. (1994) *J. Biol. Chem.* **269,** 7824–7833.
31. Tanaka, I., Appelt, K., Dijk, J., White, S. W. & Wilson, K. S. (1984) *Nature (London)* **310,** 376–381.
32. Yang, S. W. & Nash, H. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12183–12187.
33. King, C.-Y. & Weiss, M. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11990–11994.
34. Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995) *Cell* **81,** 705–714.
35. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995) *Nature (London)* **376,** 791–795.