Published in final edited form as:

Biochemistry. 2006 October 17; 45(41): 12424–12435. doi:10.1021/bi060612h.

Atomic Force Microscopy of Group II Intron Ribonucleoprotein Particle Integration into Double-Stranded DNA†

James W. Noah^{‡,⊥}, Soyeun Park[§], Jacob T. Whitt[‡], Jiri Perutka^{‡,#}, Wolfgang Frey[§], and Alan M. Lambowitz^{‡,*}

[‡]Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, and Section of Molecular Genetics and Microbiology, School of Biological Sciences, University of Texas at Austin, Austin, TX 78712–0159 USA

[§]Department of Biomedical Engineering, Texas Materials Institute, and Center for Nano and Molecular Science and Technology, University of Texas at Austin, Austin, Texas 78712–1062 USA

Abstract

The mobile *Lactococcus lactis* L1.LtrB group II intron integrates into DNA target sites by a mechanism in which the intron RNA reverse splices into one DNA strand, while the intron-encoded protein uses a C-terminal DNA endonuclease domain to cleave the opposite strand and then uses the cleaved 3' end to prime reverse transcription of the inserted intron RNA. These reactions are mediated by an RNP particle that contains the intron-encoded protein and the excised intron lariat RNA, with both the protein and base pairing of the intron RNA used to recognize DNA target sequences. Here, computational analysis indicates that *Escherichia coli* DNA target sequences that support L1.LtrB integration have greater predicted bendability than do random *Escherichia coli* genomic sequences, and atomic force microscopy shows that target DNA is bent during the reaction with L1.LtrB RNPs. Time-course and mutational analyses show that DNA bending occurs after reverse splicing and requires subsequent interactions between the intron-encoded protein and the 3'-exon, which lead to two progressively larger bend angles. Our results suggest a model in which RNPs bend the target DNA by maintaining initial contacts with the 5' exon, while engaging in subsequent 3'-exon interactions that successively position the scissile phosphate for bottom-strand cleavage at the DNA endonuclease active site and then reposition the 3' end of the cleaved bottom strand at the reverse transcriptase active site for initiation of cDNA synthesis. Our findings indicate that bendability of the DNA target site is a significant factor for L1.LtrB RNP integration.

Mobile group II introns, which are found in bacterial and organelle genomes, are retroelements that integrate site-specifically (“retrohome”) into DNA target sites at high frequency and retrotranspose to ectopic sites that resemble the normal homing site at low frequency (reviewed in refs (1–4)). The integration reactions are mediated by an RNP complex that is formed during RNA splicing and contains the intron-encoded protein (IEP)¹ and the excised intron lariat RNA (3,4). RNPs initiate intron mobility by recognizing relatively long (30–35 bp) DNA target sites, with both the IEP and base pairing of the intron RNA contributing to the recognition of DNA

[†]This work was supported by NIH grant GM37949 and Welch Foundation Grant F-1607 to A.M.L. and by support from the Welch Foundation and SPRING through the Center for Nano- and Molecular Science and Technology to W.F. J.W.N. was the recipient of NIH post-doctoral fellowship GM20780.

*To whom correspondence should be addressed: Telephone: (512)-232–3418. Fax: (512)-232–3420. E-mail: lambowitz@mail.utexas.edu.

[⊥]Current Address: Southern Research Institute, Drug Discovery Division, Emerging Infectious Disease, 2000 Ninth Avenue South, Birmingham, AL 35205.

[#]Current Address: Department of Internal Medicine, Division of Infectious Diseases, Virginia Commonwealth University, School of Medicine, 1101 E. Marshall Street, Richmond, VA 23298.

target sequences (5-8). Then, the intron RNA reverse splices into one strand of the DNA and is reverse transcribed by the IEP, yielding an intron cDNA that is integrated into the recipient genome by host DNA recombination or repair mechanisms (reviewed in refs 3,4). Their very high integration frequencies and target specificity combined with the ability to program insertion into different target sites by modifying the base-pairing sequences of the intron RNA have made it possible to develop mobile group II introns into highly efficient bacterial gene targeting vectors ("targetrons")(reviewed in ref (9)).

The *Lactococcus lactis* L1.LtrB intron, which has been studied as a model system, encodes a protein, denoted LtrA, with four conserved domains: RT, corresponding to the fingers and palm regions of retroviral RTs; X, corresponding in part to the RT thumb; DNA-binding (D), and DNA endonuclease (En) (Figure 1A; refs (1-4)). LtrA's RT and X domains function together to bind the intron RNA and stabilize its active structure for RNA splicing and reverse splicing, while the D and En domains interact with DNA target sites during intron mobility. The D domain is required for efficient reverse splicing into double-stranded DNA and is thought to contribute to DNA target site recognition, while the En domain carries out second-strand cleavage to generate the primer for reverse transcription of the inserted intron RNA (10). In addition to the predominant En-dependent integration mechanism described above, the L1.LtrB intron can also retrohome at lower frequency by using an En-independent mechanism in which a nascent strand at a DNA replication fork is used to prime reverse transcription of the inserted intron RNA (11). Additionally both En-dependent and En-independent mechanisms are used for retrotransposition of the L1.LtrB intron to ectopic sites that resemble the normal homing site (12).

A key experimental advantage of the L1.LtrB intron is the availability of an efficient *E. coli* expression system, which makes it possible to obtain large quantities of purified IEP and RNPs reconstituted with purified IEP and *in vitro*-synthesized intron RNA (13,14). The latter enabled detailed biochemical studies of DNA target site recognition. Kinetic analysis indicated that the L1.LtrB RNPs bind DNA non-specifically and then search for target sites via a facilitated-diffusion mechanism, analogous to those used by site-specific DNA-binding proteins (15). DNA target site interactions of L1.LtrB RNPs have been analyzed by mutational analysis, DNA footprinting, modification-interference, and missing-base analysis (5,7,8,16). These studies suggest that initial DNA target site recognition involves major-groove interactions between the IEP and a few specific bases in the distal 5'-exon region of the target site, including T-23, G-21, and A-20 (Figure 1B). These base interactions, bolstered by phosphate-backbone and possibly minor-groove interactions between positions -24 and -13, lead to local DNA melting, enabling intron RNA sequences EBS2, EBS1, and δ to base pair to sequences IBS2, IBS1, and δ' between DNA target site positions -12 and +3 (Figure 1B; EBS and IBS denote exon-binding site and intron-binding site, respectively). After base pairing, the intron RNA reverse splices into the top DNA strand. Bottom-strand cleavage occurs after a lag and requires additional interactions between the IEP and 3' exon, the most critical of which is recognition of T+5 in the region of the DNA target site that becomes single-stranded after DNA melting (7,8).

Recently, we constructed a three-dimensional homology model of the LtrA protein by threading on X-ray crystal structures of HIV-1 RT dimers (17). Docking of target DNA onto this three-dimensional model showed that the DNA target sequence is too long for the LtrA protein to bind simultaneously to its interaction sites in the distal 5'-exon and 3'-exon regions unless the DNA is strongly bent in the complex (17). These findings raised the possibility that

¹Abbreviations: AFM, atomic force microscopy; bp, base pair; EBS, exon-binding site; CS, cleavage site; D, DNA-binding domain; En domain, DNA endonuclease domain; E1, exon 1; E2, exon 2; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IBS, intron-binding site; IEP, intron-encoded protein; IS, intron-insertion site; ORF, open reading frame; nt(s), nucleotide(s); phenol-CIA (phenol-chloroform-isoamyl alcohol; 25:24:1); RNP, ribonucleoprotein; RT, reverse transcriptase; Tris, Tris(hydroxymethyl)aminomethane; WT, wild type.

bendability of the DNA target site might significantly influence the efficiency of the integration reaction.

Here, we analyzed the bendability of LI.LtrB target sites computationally and by atomic force microscopy (AFM). The former indicates that LI.LtrB DNA target sites have significantly higher predicted bendability than do random *E. coli* DNA sequences, and the latter shows that the target DNA is bent at two progressively greater angles during the integration of group II intron RNPs. Together with previous findings, our results suggest a model in which RNPs bend the target DNA by maintaining initial IEP and base-pairing interactions with the 5' exon, while engaging in subsequent interactions with the 3' exon that successively position the scissile phosphate for bottom-strand cleavage at the En active site and then reposition the 3' end of the cleaved bottom-strand at the RT active site for initiation of reverse transcription.

EXPERIMENTAL PROCEDURES

Computational Analysis of DNA Bendability

The bending propensity of DNA (bendability) was calculated by using the following algorithm based on the bendability of trinucleotides inferred from DNase I-digestion data:

$$\text{bendability} = \prod_{i=1}^4 p(a)_i$$

where $p(a)$ is the bending propensity of a trinucleotide at position i (18). Two sets of test sequences were used for group II intron-integration sites. The first set consists of 66 retrohoming sites in the *E. coli* genome used by an LI.LtrB- Δ ORF intron having randomized EBS2, EBS1, and δ sequences (19); the 66 sequences were culled from 88 total sequences by excluding sites lacking T+5, which may have been utilized via En-independent retrohoming pathways. The second set consists of 33 retrotransposition sites in the *E. coli* genome used by the wild-type LI.LtrB intron, again culled from a larger set by excluding sites lacking T+5 (12). For each test set, the average bendability was calculated for positions -30 to +10 from the intron-insertion site, using a six-nucleotide sliding window with a shift increment of one, and compared with that in control sets consisting of the same number of 45-bp sequences selected randomly from the *E. coli* genome. To construct the control sets for the retrohoming and retrotransposition sites, the *E. coli* genome was divided into 66 or 33 equal segments, respectively, and one 45-bp sequence was selected randomly from each segment by using an algorithm that randomly selects a line from a file (20).

The average bendabilities of each 6-nt segment in the test and control sets were compared by using a test statistic for the difference between two sample means, as follows:

$$z = \bar{x}_1 - \bar{x}_2 / \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

where $\bar{x}$ is an average bendability over n sequences and s is the standard deviation. The subscripts 1 and 2 denote parameters for the randomly chosen sites and group II intron-integration sites, respectively. For each test set, this comparison was repeated against 100 different control sets.

DNA Substrates

DNA substrates used for biochemical assays and AFM experiments were generated by PCR of pLHS, which contains a 70-nt sequence of ligated *ltrB* exons 1 and 2, extending from 35-nt upstream to 35-nt downstream of the intron-insertion site (positions -35 to +35), cloned between the *Bam*HI and *Eco*RI sites of pBSKS+ (Stratagene, La Jolla, CA) (13).

To generate the wild-type DNA substrate, the PCR was carried out with primers SK+113 (5'-CCCTCACTAAAGGGAACAAAAGC) and KS-121 (5'-CGACGGCCAGTGAATTGTAATAGC). The resulting 234-bp double-stranded DNA contains the 45-bp E1-E2 target sequence flanked by 5'- and 3'-sequences of 91 and 98 bp, respectively. Target site L (240 bp) contains the 45-bp L1.LtrB target sequence at one end of the DNA oligonucleotide (positions 6–50) and was generated by PCR of pLHS with primers KS-35 (5'-CTTGCAACCCACGTCGATCGTG) and SK+205 (5'-CTTCCGGCTCGTATGTTGTGTGG). Mutant DNA substrates (234 bp) contain unfavorable non-wild-type nucleotide residues at target-site positions –30 to –13, –12 to +3, +4 to +15, and –30 to +15, respectively. The non-wild-type residues are those occurring at lowest frequency at each position in the set of retrohoming sites in the *E. coli* genome obtained with an L1.LtrB intron with randomized target-site recognition sequences (Fig. 5 of ref (19)). These mutant target DNAs were generated via two PCR steps. In the first step, two separate PCR reactions were done with outside primers SK+113 or KS-121 and overlapping internal primers, containing the DNA target site mutations. The internal primers were: SK–30/–13 (5'-GTAAGATAATGGTCCTGACACATCCATAACCATATC) and KS–30/–13 (5'-TGTGTCAGGACCATTATCTTACTGCAATTATCCACTAG); SK –12/+3 (5'-TCGTGACTGAAGGTAAAAAATGATCATTTTTTAATTCTACG) and KS –12/+3 (5'-CATTTTTTACCTTCAGTCACGATCGACGTGGGTTG); SK +4/+15 (5'-CGGCAATATATGTCTACGAATCTTTATACTGG) and KS +4/+15 (5'-CGTAGACATATATTGCCGATGGTTATGGATGTGTTTCACGA); and SK –30/+15 (5'-TTTTTACCTTCAGCAGGACCATTATCTTACTGCAAGGATCCACTAG) and KS –30/+15 (5'-GGTCCTGCTGAAGGTAAAAAATGCGGCAATATATGTCTACGAATCTTTATACTGG). The products of the first PCR were gel-purified, mixed, and used in a second PCR with the outside primers to yield 234-bp DNA substrates containing the desired modifications.

To generate wild-type and mutant +4/+15 DNA substrates containing a nick at the bottom-strand cleavage site, the top strand was amplified by unidirectional PCR of pLHS or the gel-purified +4/+15 DNA substrate, respectively, with primer KS-121 (see above). After extraction with phenol-chloroform-isoamyl alcohol, 25:24:1 (phenol-CIA), the unidirectional PCR products were annealed with a ten-fold molar excess of complementary gel-purified bottom-strand oligonucleotides (129 and 105 nts) by heating to 90°C for 5 min and then slowly cooling to room temperature. The resulting 234-bp double-stranded DNAs with a nick at the bottom-strand cleavage site were purified in a 1% agarose gel.

Reconstitution of RNPs

L1.LtrB RNPs were reconstituted with purified LtrA protein and *in vitro* synthesized L1.LtrB-ΔORF lariat RNA. Wild-type and mutant LtrA proteins were expressed in *E. coli* BL21(DE3), by using the intein-based expression plasmid pImp-1P, and purified as described (14). The LtrA mutant DD[–] has the conserved amino acid residues YADD in the RT active site changed to YAAA (13), and mutants ΔEn and ΔD/En have C-terminal truncations at amino acid residues 547 and 490, respectively, which delete the En and D + En domains, respectively (10). Protein concentrations were determined by Bradford assay (21), using LtrA protein whose concentration had been determined by using A₂₈₀ as a standard. The protein preparations were >98% pure, as judged by Coomassie blue-stained SDS-polyacrylamide gels

L1.LtrB lariat RNA was generated by self-splicing of a 971-nt L1.LtrB RNA (902-nt L1.LtrB intron flanked by 5'- and 3'-exon sequences of 32 and 37 nt, respectively) transcribed from *Bam*HI-linearized pJNΔORF with phage T7 RNA polymerase, using a MEGAScript kit (Ambion, Austin, TX) (22). Self-splicing was carried out with 2 μM L1.LtrB RNA in 1 ml of 0.5 M NH₄Cl, 50 mM MgCl₂, 40 mM Tris-HCl, pH 7.5 at 30°C. After renaturing the RNA by

heating to 50°C for 1 min and then slowly cooling to 30°C in 0.5 M NH₄Cl, 5 mM MgCl₂, 40 mM Tris-HCl, pH 7.5, self-splicing was initiated by adding the remaining MgCl₂. The reaction was incubated for 2 h at 37°C and terminated by extraction with phenol-CIA.

To reconstitute RNPs, the LtrA protein was incubated with the *in vitro* self-spliced lariat RNA in 15 ml of reaction medium containing 0.5 M NH₄Cl, 5 mM MgCl₂, 40 mM HEPES, pH 7.5. The RNA (3 nmoles in 1-ml of reaction medium) was renatured by heating to 50°C for 1 min and then slowly cooling to 30°C. After adding LtrA protein (9 nmoles), the sample was diluted to 15 ml with reaction medium and incubated for 2 h at 30°C. RNPs were then pelleted by centrifugation in a Beckman Ti50.2 rotor (Beckman-Coulter, Fullerton, CA; 40,000 rpm, ≥ 8 h at 4°C). The RNP pellet was dissolved in 300 µl of 10 mM KCl, 10 mM MgCl₂, 40 mM HEPES, pH 7.5, 50% glycerol, aliquoted, and stored at ≤ -20°C. A typical RNP preparation contained 88% lariat RNA, as judged by electrophoresis in a denaturing 4% polyacrylamide gel.

DNA Integration Assays

Reactions were carried out by incubating reconstituted RNPs (6 pmoles) with gel-purified DNA substrates (6 pmoles) in 100 µl of 10 mM KCl, 10 mM MgCl₂, 40 mM HEPES, pH 7.5 (14). The reactions were initiated by adding RNPs, incubated for 5 or 30 min at 37°C, and terminated by phenol-CIA extraction and ethanol precipitation. The products were analyzed in a denaturing 4% polyacrylamide gel, which was scanned with a PhosphorImager.

Atomic Force Microscopy

For AFM, 6 pmoles of RNPs were incubated with 6 pmoles of target DNA in 100 µl of 10 mM KCl, 10 mM MgCl₂, 40 mM HEPES, pH 7.5 for 5 or 30 min at 37°C, then placed on ice. Samples were adsorbed onto aminosilane-modified mica surfaces using a slight modification of the procedure described in ref (23). Freshly cleaved mica was exposed to the vapor of 3-aminopropyltriethoxysilane (98%; Sigma, St. Louis, MO) for 1 h, then baked for 1 h at 100°C. After diluting RNP:DNA complexes to 650 pM in 10 mM KCl, 10 mM MgCl₂, 40 mM HEPES, pH 7.5, a 200 µl aliquot was applied to the treated mica slide, incubated on ice for 20 min, washed once with 500 µl buffer, twice with 500 µl water, and air dried. AFM imaging was done in tapping mode using a MFP-3D scanning probe microscope (Asylum Research, Santa Barbara, CA). Super Sharp Silicon 225 µm tips (SSS-NCLR or SSC-NCL, Nanoworld, Neuchatel, Switzerland) with a 190 kHz frequency were used for imaging. We discerned no difference in resolution with the different tips.

DNA-bend angles were measured using Scion Image software (Scion, Frederick, MD). Only complexes that were completely visible and could be traced unambiguously were used for analysis. Selected images of single complexes were enlarged and two tangential lines were drawn along the center of the protruding DNA arms on both sides of the globular RNP, to give the measured bend angle. The angles were defined by manually determining the path of the DNA segments exiting the RNP within about 15 nm on each side (24). For naked DNA, a circular mask about the size of the RNP was placed over the center of the DNA, and the DNA-bend angle at this position was determined as described above. The angle apex was chosen as the point where the DNA path trajectories intersect. In each case, the DNA-bend angle was calculated from measurements of 100 molecules, and the values are expressed as the mean ± the standard deviation. By convention, the bend angle is defined as deviation from straight.

RESULTS

Computational Analysis of DNA Bendability

To assess the bending propensity (“bendability”) of L1.LtrB target sites, we employed an algorithm that uses a probabilistic model based on the bendability of trinucleotides deduced from DNase I-digestion data (see Experimental Procedures; ref (18)). The algorithm was used to calculate an average bendability profile for two sets of L1.LtrB insertion sites in the *E. coli* genome, one consisting of 66 retrohomologous sites used by an L1.LtrB-ΔORF intron having randomized target-site recognition sequences (19), and the other consisting of 33 retrotransposition sites used by the wild-type L1.LtrB intron (12). Both test sets were culled for larger sets of sites by excluding those lacking T+5, which may have been utilized via En-independent pathways that do not require the same IEP interactions with the 3' exon (see Introduction; (11,12)). The average bendability profiles for the test sets were then compared with those for 100 different control sets, each consisting of the same number of randomly selected 45-bp *E. coli* genomic sequences (a total of 6600 and 3300 control sequences for the retrohomologous and retrotransposition sites, respectively; see Experimental Procedures). We reasoned that a collection of target sites would provide a better picture of bendability patterns than would examination of individual target sites, including the wild-type site, in which the influence of bendability might be obscured by other features.

Figure 2 shows the bendability profiles for the retrohomologous sites (red), retrotransposition sites (blue), the wild-type L1.LtrB target site (dashed line), and 100 control sets for the retrohomologous sites (light blue). The profiles were calculated for positions −30 to +10 from the intron-insertion site, using a 6-bp sliding window and a shift increment of one. The control sets for the retrotransposition sites are similar to those for the retrohomologous sites and were omitted to simplify the plots. At a 5% significance level, both the retrohomologous and retrotransposition sites show several regions with significantly higher bendability than the control sites, albeit at somewhat different locations in the retrohomologous and retrotransposition sites (six-bp segments beginning at positions −13 to −10 and −3 in the retrohomologous sites and at positions −20 to −18, and −6 in the retrotransposition sites). A possible explanation for the different locations of the bendable regions is that the retrotransposition sites were obtained with the wild-type L1.LtrB-ΔORF intron, while the retrohomologous sites were obtained with an L1.LtrB-ΔORF intron having randomized EBS2, EBS1, and δ sequences. Consequently, the retrotransposition sites are constrained to have target sequences that base pair with a single intron (wild-type L1.LtrB), while the retrohomologous sites are not similarly constrained, providing more latitude for the location of the bendable regions. This inference is supported by the finding that the bendability profile for wild-type retrohomologous site (dashed line), which must also base pair with the wild-type L1.LtrB intron, parallels that for the retrotransposition sites. Notably, the bendability peaks for the wild-type site have significantly higher amplitudes than do those in the profiles calculated by averaging multiple retrohomologous and retrotransposition sites (see difference in scale on the right). These higher amplitudes may reflect the importance of bendability in selection of the wild-type target site. We conclude from the computational analysis that both retrohomologous and retrotransposition sites have regions of higher predicted bendability than do random *E. coli* DNA sequences. Additional computational comparisons of protein-induced deformability (25), propeller twist (26), and stacking energy (27) also showed significant differences in the retrohomologous and retrotransposition sites compared to the control sets, reinforcing the conclusion that L1.LtrB integration sites may have distinctive structural characteristics (data not shown).

DNA-Integration Reactions with Wild-Type and Mutant DNA Substrates and RNPs

For AFM experiments, we constructed a series of ~240-bp wild-type (WT) and mutant DNA substrates with unfavorable, non-wild-type nucleotide residues in different regions of the

L1.LtrB target site. Figure 3 shows experiments in which the DNA substrates were tested for their ability to support DNA-integration reactions of L1.LtrB RNPs. In these experiments, L1.LtrB RNPs were reconstituted with *in vitro*-synthesized L1.LtrB- Δ ORF lariat RNA and purified LtrA protein and incubated with internally labeled DNA substrates. The products of the reaction were analyzed by electrophoresis in a denaturing 4% polyacrylamide gel, which was dried and scanned with a PhosphorImager.

Figure 3B, lanes 2 and 3, show reactions in which wild-type L1.LtrB RNPs were incubated with the WT DNA substrate for 5 and 30 min, respectively. As expected, the reactions yielded both high molecular weight products, resulting from the first and second steps of reverse splicing (referred to as partial and complete reverse splicing, respectively; see Figure 3C), and smaller products resulting from bottom-strand cleavage (5' bottom = 104 nt and 3' bottom = 130 nt). Previous time-course experiments with the wild-type intron and target site showed that the reverse-splicing products appear first, while the second-strand cleavage products appear after a lag (8). In agreement with these findings, quantifying the gel showed that after 5 min, the reverse-splicing products comprise 28% of the total radioactivity and bottom-strand cleavage products only 2%, while after 30 min, the reverse-splicing products comprise 52% and the bottom-strand cleavage products 9%.

The reactions with the remaining DNA substrates were done for 30 min. The L DNA substrate, which contains the wild-type target site near one end of the DNA, reacted as efficiently as the WT DNA substrate, with the shift of the site resulting in different sizes for the bottom-strand cleavage products (5' bottom = 196 nt; 3' bottom = 44 nt and run off the gel; Figure 3B, lane 5). The -30/+15 substrate, which has unfavorable non-wild-type target site nucleotide residues at all positions between -30 and +15, showed only a very low level of reverse splicing (< 1% that of the WT substrate; lane 4), which is likely due to inefficiently utilized cryptic sites (see ref. (7)). Mutant DNA substrates -30/-13 and -12/+3, which have non-wild-type nucleotide residues in the distal 5'-exon and IBS/ δ regions, also showed <1% of the reverse-splicing activity of the WT substrate and no detectable bottom-strand cleavage (lanes 6 and 7). These results are in agreement with previous findings showing that mutations affecting either IEP interactions with the distal 5'-exon region or intron RNA base-pairing interactions with the IBS/ δ region inhibit both reverse splicing and bottom-strand cleavage (7, 8).

Interactions between the IEP and 3' exon are not required for reverse splicing, but are required for bottom-strand cleavage (7,8). Accordingly, DNA substrate +4/+15, which has non-wild-type nucleotide residues at 3'-exon positions recognized by the IEP, supported efficient reverse splicing, but gave no detectable bottom-strand cleavage (Figure 3B; lane 8). Also in agreement with previous results (10), RNPs containing a C-terminally truncated IEP lacking the D and En domains (Δ D/En) showed no detectable reverse splicing or bottom-strand cleavage (lane 10), while RNPs lacking only the En domain (Δ En) were still capable of reverse splicing, but were unable to carry out bottom-strand cleavage (lane 9).

AFM with the WT and L DNA Substrates

Next, we used AFM to examine complexes formed between the RNPs and DNA substrates under the conditions of the DNA-integration reactions. Figure 4 shows representative complexes formed after incubating wild-type L1.LtrB RNPs with the WT and L DNA substrates for 30 min at 37°C. The samples were air-dried and imaged in tapping mode. We were unable to image in buffer due to problems with RNP aggregation and complex immobilization. Only well-separated complexes with both DNA ends visible were used for analysis.

In the AFM images, the RNPs appear as a bright dot located in the center or at the end of the WT and L DNA substrates, respectively. The shift in the position of the RNPs bound to the L substrate indicates that they specifically recognize, bind, and react with the DNA target

sequence and also that the RNP:DNA complexes are two-dimensionally fixed on the aminosilanized mica slide in their original position in the complex. Based on measurements of 100 molecules, the wild-type target DNA in complex with RNPs had a length of 85 ± 9 nm (mean $\pm$ standard deviation) compared to 81 ± 11 nm for the free DNA (see below) and a calculated length of ($3.3 \text{ \AA} \times 234 \text{ bp} =$) 77.2 nm. The similar lengths of free and bound DNA indicate that the DNA is not substantially wrapped around the RNP in the complex.

At the resolution of the AFM, the RNPs bound to the DNA substrate appear roughly circular. Based on measurement of 100 molecules, RNPs bound to the wild-type DNA substrate have a diameter of 10 ± 1 nm and a height of 0.9 ± 0.2 nm, and the corresponding measurements for RNPs bound to the L DNA substrate were 11 ± 2 nm and 1.1 ± 0.3 nm. The RNPs bound to the L DNA substrate in Figure 4 are among the larger RNPs (diameters of 11 and 12 nm and heights of 1.1 and 1.4 nm). AFM measurements of free RNPs under the same conditions gave a diameter of 11 ± 1 nm and height of 1.0 ± 0.4 nm, similar to the dimensions of the RNPs bound to the DNA (Figure 5). Although we cannot exclude damage to the RNPs during sample preparation or imaging (28), only active RNPs containing both the IEP and intron RNA can react with the DNA substrate, and RNA and protein by themselves do not bind DNA specifically (8). The relatively low height of the RNPs may reflect compaction by the tip, interaction of the RNPs with the mica substrate, dehydration, capillary forces during drying, or differences in compliance between the RNPs and mica substrate in tapping mode (29-32). It is possible that the RNPs are inherently nonglobular and flattened in one dimension, with this flattening enhanced by dehydration, similar to the situation for *E. coli* small ribosomal subunits imaged in air (ref. (32) and see Discussion).

Distribution of Bend Angles at Different Stages of the Reaction

The initial AFM imaging of complexes with the WT target site, done with samples incubated for 30 min at 37°C, showed that the DNA is bent around the RNPs and that the bend angles have a distinctly bimodal distribution with peaks at $\sim 75^\circ$ and $\sim 90^\circ$. To test whether the two DNA-bend angles might reflect different stages of the reactions, we compared 100 complexes with the WT target site after 5 and 30 min of incubation (Figure 6A and B). The histograms show that the lesser bend angle predominates in the 5-min samples ($75 \pm 6^\circ$, 83%; $88 \pm 3^\circ$, 17%, panel A), while the greater bend predominates in the 30-min samples ($75 \pm 2^\circ$, 49%, $90^\circ \pm 5^\circ$, 51%; panel B). In the absence of RNPs, the same position in the WT target DNA shows a wide distribution of bend angles (Figure 6C). The distribution of bend angles in the naked wild-type DNA does appear significantly more skewed toward larger bend angles than is the case for the $-30/+15$ mutant DNA (Figure 6D), possibly reflecting that the wild-type DNA target sequence has an increased propensity for bending even in the absence of RNPs (see Discussion).

AFM Imaging of Complexes with Mutant DNA Target Sites

We were unable to detect complexes formed between wild-type L1.LtrB RNPs and mutant DNA target site $-30/-13$, which has unfavorable nucleotide residues in the distal 5'-exon region recognized by the IEP, or $-12/+3$, which has unfavorable nucleotide residues in the IBS/ δ' region recognized by base pairing of the intron RNA (data not shown). These findings agree with previous DNA-footprinting experiments, which showed that DNA substrates having unfavorable nucleotide residues in the distal 5'-exon region do not form stable complexes with the RNPs (8).

We did detect complexes with DNA substrate $+4/+15$, which has modifications in 3'-exon sequences that must be recognized by the IEP for bottom-strand cleavage (Figure 7A). Above, we showed that this DNA substrate could support reverse splicing, but not bottom-strand cleavage (Figure 3B, lane 8). Imaging of 100 complexes formed with the $+4/+15$ DNA substrate

showed RNPs bound at the correct position, but with a wide range of DNA-bend angles, as found for naked DNA (Figure 7A). These findings indicate that the initial DNA target site recognition and reverse splicing are not by themselves sufficient to induce DNA bending and that bending requires either additional IEP interactions with the 3' exon that lead to bottom-strand cleavage or bottom-strand cleavage itself.

AFM Imaging of Complexes with Mutant RNPs

Similar conclusions were reached from AFM imaging of complexes formed with mutant RNPs containing C-terminally truncated LtrA proteins. RNPs containing LtrA mutant $\Delta D/En$, which lacks both the D and En domains, gave no detectable complexes (not shown), again in agreement with DNA-footprinting experiments showing that these RNPs do not form stable complexes with the DNA substrate (8). RNPs containing LtrA mutant ΔEn were shown above to be capable of reverse splicing but not second-strand cleavage (Figure 3, lane 9). As expected, these RNPs gave stable complexes detected by AFM, and analysis of 100 such complexes again showed RNPs bound at the correct position, but with a wide distribution of bend angles (Figure 7B), reflecting either loss of required interactions with the 3' exon or bottom-strand cleavage.

AFM Imaging of Complexes with DNA Substrates Containing a Nick at the Bottom-Strand Cleavage Site

To distinguish whether DNA bending is caused by IEP interactions with the 3' exon or by bottom-strand cleavage itself, we carried out AFM imaging of RNPs bound to DNA substrates with a pre-positioned nick at the bottom-strand cleavage site. One nicked DNA substrate contained the wild-type DNA target sequence, and the other contained the +4/+15 mutations that block interaction of the IEP with the 3' exon. Figure 8 shows AFM imaging of complexes formed after a 30-min incubation of RNPs with these pre-nicked DNAs. The complexes with the nicked WT DNA substrate showed the same two bend angles ($74 \pm 2^\circ$, 50% and $89 \pm 2^\circ$, 50%) as those with the unnicked WT substrate (Figure 8A), while complexes formed with the nicked +4/+15 DNA substrate still showed a wide distribution of bend angles (Figure 8B). These experiments show that the introduction of the bottom-strand nick is not by itself sufficient to cause bending and together with the results for the ΔEn mutant with the wild-type DNA substrate (Figure 7B) indicate that bending requires interaction of the IEP with specific sequences in the 3' exon.

AFM Imaging of Complexes after Initiation of Reverse Transcription

The addition of dNTPs during the DNA-integration reaction allows LtrA to initiate target-primed reverse transcription using the cleaved bottom strand as a primer. By adding just three dNTPs (dATP, dTTP and dGTP) and omitting dCTP, reverse transcription can be made to stop opposite the first C-residue in the template RNA (17 nts), resulting in uniform complexes arrested at the same stage of the reaction. Figures 9A and B show AFM imaging of complexes formed by incubating wild-type RNPs with the WT DNA substrate in the presence of the three dNTPs for 5 and 30 min, respectively. After 5 min, the number of complexes with the two major bend angles was roughly equal ($76 \pm 3^\circ$, 48% and $90 \pm 2^\circ$, 52%) whereas after 30 min, the proportion of molecules with the greater bend angle increased dramatically ($73 \pm 2^\circ$, 16% and $90 \pm 4^\circ$, 84%). By contrast after 30-min incubation with WT DNA under the same conditions, DD^- RNPs, which have the mutation YADD $\rightarrow$ YAAA at the RT active site (13), gave predominantly complexes with the smaller bend angle ($73 \pm 2^\circ$, 46 %) and only a small number of complexes with the greater bend angle ($90 \pm 2^\circ$, 11%). Together, these results suggest that the greater bend angle is characteristic of RNPs carrying out reverse transcription. The results for DD^- RNPs may reflect that the integrity of the RT active site is required to achieve the greater bend angle, but we cannot exclude the possibility that the mutations affect the structure elsewhere in the protein.

DISCUSSION

Here, computational analysis indicates that LI.LtrB integration sites in the *E. coli* genome have regions of higher predicted bendability than do randomly selected *E. coli* DNA sequences, and AFM shows that the DNA target site is bent during the RNP integration reaction. The length of the DNA in the complex as determined by non-deconvoluted AFM measurements is 85 ± 9 nm, in good agreement with the calculated length (77.2 nm) and close to that determined by AFM measurements for free DNA substrate (81 ± 11 nm), indicating that the DNA is not wrapped around the RNPs to an appreciable extent.

By AFM imaging, the LI.LtrB RNPs bound to DNA appear roughly circular with an average diameter of ~ 10 nm and a height of ~ 1 nm. The diameter of the LI.LtrB RNPs, which consist of a 0.9-kb RNA and two molecules of the 70 kDa LtrA protein (0.45 MDa; ref (14)), is in reasonable agreement with expectations for an RNP complex based on the *E. coli* small and large ribosomal subunits. The dimensions of the *E. coli* small ribosomal subunit (0.85 MDa) determined by cryoelectron microscopy or X-ray crystallography are $10 \times 15 \times 6$ nm, and those for the large ribosomal subunit (1.7 MDa) are $24 \times 24 \times 15$ nm (33-36). The relatively low height measured for LI.LtrB RNPs may reflect compaction during air drying or imaging or compliance differences in tapping mode (see Results). *E. coli* small ribosomal subunits appear to be flattened by about 50% by AFM imaging in air versus AFM imaging in solution (32). LI.LtrB RNPs of ~ 10 nm diameter could potentially contact 30-bp of unbent DNA and > 30 -bp if the DNA is bent, in agreement with known distal contact points at positions -23 and $+10$ (8). DNase I-footprinting, which may somewhat overestimate the size of the binding site, showed protection over 40–45 bp (positions -25 to $+19$ on the top strand and -28 to $+16$ on the bottom strand (8)). The AFM imaging emphasizes the striking compaction of the 902-nt intron RNA in RNPs, which appear as a small dot on the 234-bp double-stranded DNA substrate.

The computational analysis showing that both retrohoming and retrotranspositions sites have regions of higher predicted bendability than do random *E. coli* DNA sequences suggests that bendability was a significant parameter in the use of these sites. Bendability does not distinguish whether the RNPs preferentially recognize DNA target sites that were already bent or whether the flexibility of the target site increases the efficiency of the reaction, by decreasing the energy required to achieve the bend. AFM comparison of bend angles at the same position in the wild-type and $-30/+15$ mutant DNA substrates in absence of RNPs showed higher bend angles for the wild-type DNA (Figure 6C and D), consistent with the possibility that the wild-type target site contains a higher proportion of molecules pre-bent at the correct positions. Assuming that the DNA on the surface represents a snapshot of the state of DNA bending in solution, these AFM results agree with recent simulations of the in-plane bending of short semiflexible polymers (37). While polymers with a ratio “ t ” of contour length to persistence length of about two can only probe bend angles up to 90° with significant probability, more flexible polymers that have reduced persistence lengths probe larger bend angles. Given that the persistence length of DNA is generally 40–45 nm (38) or about half the length of our DNA, the distribution of bend angles in the $-30/+15$ DNA show the predicted behavior for $t = 2$, while the wild-type DNA shows the behavior expected for $t > 2$ or much reduced stiffness.

Both the retrohoming and retrotransposition sites contain only a small number of nucleotide residues whose bases are recognized directly by the IEP, and with the exception of T+5, these are only moderately conserved in different sites (e.g., T-23, 41%; G-21, 79%; and A-20, 61% in the 66 retrohoming sites, and T-23, 15%; G-21, 48%; and A-20, 33% in the retrotransposition sites (12,19)). The region of the retrohoming sites recognized by base pairing of the intron RNA also has variable sequences, due to the method by which the sites were selected (i.e., with an LI.LtrB intron having randomized EBS2, EBS1, and δ sequences). Thus, it seems likely that

bendability itself is a factor in the selection of L1.LtrB target sites, beyond the nucleotide residues recognized directly by L1.LtrB RNPs.

Figure 10 summarizes our AFM results in the form of a model showing the degree of bending at different stages of the reaction. First, we find that the initial DNA target site recognition and reverse-splicing reaction do not by themselves lead to DNA bending. A mutant DNA substrate with alterations between 3'-exon positions +4 and +15 (+4/+15), and mutant RNPs that lack the En domain carry out these initial steps with relatively high efficiency, but are unable to carry out bottom-strand cleavage. In both cases, AFM imaging shows complexes with RNPs bound at the correct position, but with a wide distribution of DNA-bend angles, as found for uncomplexed DNA.

A second conclusion is that DNA bending likely results from the force generated by RNPs binding simultaneously to both the 5'- and 3'-exon regions of the DNA target site. Previous studies showed that all mutations or chemical modifications in the distal 5'-exon or IBS/ δ region of the DNA target site that affect IEP interactions or base pairing of the intron RNA inhibit not only reverse splicing, but also bottom-strand cleavage (7,8). Although we cannot exclude that some 5'-exon interactions are weakened or lost at later stages, a likely interpretation is that most if not all of the initial RNP interactions with the 5' exon are required along with 3'-exon interactions to position the IEP for bottom-strand cleavage. Our experiments with nicked DNA substrates confirm that bending results from RNP interactions with the 3' exon and not from bottom-strand cleavage itself (Figure 8). As indicated previously, three-dimensional modeling predicted that the IEP could not interact simultaneously with its target sequences in the 5'- and 3'-exons unless the DNA is bent strongly in the complex (17).

The two progressively greater bend angles seen in time-course experiments presumably reflect sequential IEP interactions that are known to occur with the 3' exon. These are recognition of T+5 and other nucleotide residues for bottom-strand cleavage, binding of the scissile phosphate between 3'-exon positions +9 and +10 at the En active site, and finally repositioning of the 3' end of the cleaved bottom strand at the RT active site. The latter is expected to require a significant conformational change in the complex. We find that the proportion of complexes containing the greater bend angle increases after initiation of reverse transcription by wild-type L1.LtrB RNPs, but decreases for complexes formed with DD⁻ RNPs, which have a mutation in the RT active site (Figure 9). X-ray crystal structures of retroviral RTs with bound template-primer substrates typically show a 40° bend in the template strand near the polymerase active site (39-41) and even greater bend angles are seen for other polymerases (*E. coli* RNA polymerase, 90° (42) and human DNA polymerase β , 90° (43)). The DNA-bend angle seen here for reverse transcribing RNPs may reflect not only an analogous bend at the RT active site, but also that the RNPs maintain contacts involved in DNA target site recognition during the initial stages of reverse transcription. At present, we can only imagine the acrobatics required to progressively unwind the highly folded intron RNA through the RT active site, a process that presumably requires disruption of the strong IEP interactions that stabilized the folded RNA structure for RNA splicing and reverse splicing, as well as IEP and intron RNA contacts with the DNA target site.

ACKNOWLEDGEMENTS

We thank Dr. Sandra Martin (University of Colorado School of Medicine) and an anonymous reviewer for comments on the manuscript and helpful suggestions.

REFERENCES

1. Bonen L, Vogel J. The ins and outs of group II introns. Trends Genet 2001;17:322–331. [PubMed: 11377794]

2. Lehmann K, Schmidt U. Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol* 2003;38:249–303. [PubMed: 12870716]
3. Lambowitz AM, Zimmerly S. Mobile group II introns. *Annu. Rev. Genet* 2004;38:1–35. [PubMed: 15568970]
4. Lambowitz, AM.; Pyle, AM. Group II Introns: Ribozymes that Splice RNA and Invade DNA. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. *The RNA World*. 3rd ed.. Cold Spring Harbor Laboratory Press; New York: 2006. p. 469-505.
5. Guo H, Karberg M, Long M, Jones JP 3rd, Sullenger B, Lambowitz AM. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* 2000;289:452–457. [PubMed: 10903206]
6. Guo H, Zimmerly S, Perlman PS, Lambowitz AM. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J* 1997;16:6835–6848. [PubMed: 9362497]
7. Mohr G, Smith D, Belfort M, Lambowitz AM. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev* 2000;14:559–573. [PubMed: 10716944]
8. Singh NN, Lambowitz AM. Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol* 2001;309:361–386. [PubMed: 11371159]
9. Lambowitz, AM.; Mohr, G.; Zimmerly, S. Group II Intron Homing Endonucleases: Ribonucleoprotein Complexes with Programmable Target Specificity. In: Belfort, M.; Derbyshire, V.; Stoddard, B.; Wood, D., editors. *Homing Endonucleases and Inteins*. Springer-Verlag; Heidelberg: 2005. p. 121-145.
10. San Filippo J, Lambowitz AM. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol* 2002;324:933–951. [PubMed: 12470950]
11. Zhong J, Lambowitz AM. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J* 2003;22:4555–4565. [PubMed: 12941706]
12. Coros CJ, Landthaler M, Piazza CL, Beauregard A, Esposito D, Perutka J, Lambowitz AM, Belfort M. Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol. Microbiol* 2005;56:509–524. [PubMed: 15813740]
13. Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, Cavanagh S, Dunny GM, Belfort M, Lambowitz AM. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev* 1997;11:2910–2924. [PubMed: 9353259]
14. Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry* 1999;38:9069–9083. [PubMed: 10413481]
15. Aizawa Y, Xiang Q, Lambowitz AM, Pyle AM. The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell* 2003;11:795–805. [PubMed: 12667460]
16. Perutka J, Wang W, Goerlitz D, Lambowitz AM. Use of computer-designed group II introns to disrupt *Escherichia coli* DEXH/D-box protein and DNA helicase genes. *J. Mol. Biol* 2004;336:421–439. [PubMed: 14757055]
17. Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M, Lambowitz AM. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* 2005;11:14–28. [PubMed: 15574519]
18. Brukner I, Sanchez R, Suck D, Pongor S. Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *EMBO J* 1995;14:1812–1818. [PubMed: 7737131]
19. Zhong J, Karberg M, Lambowitz AM. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res* 2003;31:1656–1664. [PubMed: 12626707]
20. Wall, L.; Orwant, J.; Christiansen, T. *Programming Perl*. 3rd ed.. O'Reilly and Associates, Inc.; CA: 2000. p. 6
21. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* 1976;72:248–254. [PubMed: 942051]

22. Noah JW, Lambowitz AM. Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* 2003;42:12466–12480. [PubMed: 14580192]
23. Balagurumoorthy P, Lindsay SM, Harrington RE. Atomic force microscopy reveals kinks in the p53 response element DNA. *Biophys. Chem* 2002;101–102:611–623.
24. Janicijevic A, Sugawara K, Shimizu Y, Hanaoka F, Wijgers N, Djurica M, Hoeijmakers JH, Wyman C. DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair (Amst.)* 2003;2:325–336. [PubMed: 12547395]
25. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci., U.S.A* 1998;95:11163–11168. [PubMed: 9736707]
26. el Hassan MA, Calladine CR. Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol* 1996;259:95–103. [PubMed: 8648652]
27. Ornstein RL, Rein R, Breen DL, Macelroy RD. An optimized potential function for the calculation of nucleic acid interaction energies: base stacking. *Biopolymers* 1978;17:2341–2360.
28. Giro A, Bergia A, Zuccheri G, Bink HH, Pleij CW, Samori B. Single molecule studies of RNA secondary structure: AFM of TYMV viral RNA. *Microsc. Res. Tech* 2004;65:235–245. [PubMed: 15630685]
29. Yang J, Mou J, Yuan JY, Shao Z. The effect of deformation on the lateral resolution of atomic force microscopy. *J. Microsc* 1996;182(Pt 2):106–113. [PubMed: 8683559]
30. Spatz JP, Sheiko S, Moller M, Winkler RG, Reineker P, Marti O. Tapping scanning force microscopy in air -- theory and experiment. *Langmuir* 1997;13:4699–4703.
31. Skabkin MA, Kiselyova OI, Chernov KG, Sorokin AV, Dubrovin EV, Yaminsky IV, Vasiliev VD, Ovchinnikov LP. Structural organization of mRNA complexes with major core mRNP protein YB-1. *Nucleic Acids Res* 2004;32:5621–5635. [PubMed: 15494450]
32. Matsuura T, Tanaka H, Matsumoto T, Kawai T. Atomic force microscopic observation of *Escherichia coli* ribosomes in solution. *Biosci. Biotechnol. Biochem* 2006;70:300–302. [PubMed: 16428856]
33. Lata KR, Agrawal RK, Penczek P, Grassucci R, Zhu J, Frank J. Three-dimensional reconstruction of the *Escherichia coli* 30 S ribosomal subunit in ice. *J. Mol. Biol* 1996;262:43–52. [PubMed: 8809178]
34. Verschoor A, Warner JR, Srivastava S, Grassucci RA, Frank J. Three-dimensional structure of the yeast ribosome. *Nucleic Acids Res* 1998;26:655–661. [PubMed: 9421530]
35. Tocilj A, Schlunzen F, Janell D, Gluhmann M, Hansen HA, Harms J, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. The small ribosomal subunit from *Thermus thermophilus* at 4.5 Å resolution: pattern fittings and the identification of a functional site. *Proc. Natl. Acad. Sci., U.S.A* 1999;96:14252–14257. [PubMed: 10588692]
36. Ban N, Freeborn B, Nissen P, Penczek P, Grassucci RA, Sweet R, Frank J, Moore PB, Steitz TA. A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* 1998;93:1105–1115. [PubMed: 9657144]
37. Lattanzani G, Munk T, Frey E. Transverse fluctuations of grafted polymers. *Physical Reviews E* 2004;69:021801.
38. Wang MD, Yin H, Landick R, Gelles J, Block SM. Stretching DNA with optical tweezers. *Biophys. J* 1997;72:1335–1346. [PubMed: 9138579]
39. Ding J, Das K, Hsiou Y, Sarafianos SG, Clark AD Jr. Jacobo-Molina A, Tantillo C, Hughes SH, Arnold E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol* 1998;284:1095–1111. [PubMed: 9837729]
40. Cote ML, Yohannan SJ, Georgiadis MM. Use of an N-terminal fragment from Moloney murine leukemia virus reverse transcriptase to facilitate crystallization and analysis of a pseudo-16-mer DNA molecule containing G-A mispairs. *Acta Crystallogr. D. Biol. Crystallogr* 2000;56(Pt 9):1120–1131. [PubMed: 10957631]
41. Sarafianos SG, Das K, Tantillo C, Clark AD Jr. Ding J, Whitcomb JM, Boyer PL, Hughes SH, Arnold E. Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J* 2001;20:1449–1461. [PubMed: 11250910]

42. Rees WA, Keller RW, Vesenka JP, Yang G, Bustamante C. Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science* 1993;260:1646–1649. [PubMed: 8503010]
43. Maitra M, Gudzelak A Jr. Li SX, Matsumoto Y, Eckert KA, Jager J, Sweasy JB. Threonine 79 is a hinge residue that governs the fidelity of DNA polymerase beta by helping to position the DNA within the active site. *J. Biol. Chem* 2002;277:35550–35560. [PubMed: 12121998]

FIGURE 1.

The LtrA protein encoded by the *L. lactis* L1.LtrB group II intron, and DNA target site interactions of L1.LtrB RNPs. (A) LtrA protein. The protein is shown as a rectangle with RT, X, D, and En domains delineated and amino acid positions at the domain boundaries indicated below. The RT domain contains conserved RT sequence motifs RT1–7, along with an upstream region (RT-0) characteristic of non-LTR-retroelement RTs ((17) and refs therein). (B) DNA target site interactions for the L1.LtrB RNPs. The 45-bp target DNA sequence is shown in complex with the RNPs containing the LtrA protein (gray) and intron RNA (stem-loops containing EBS2 and EBS1/δ sequences). The intron-insertion site (IS) in the top DNA strand and the DNA endonuclease cleavage site (CS) in the bottom strand are indicated by arrowheads. Critical target site nucleotide residues recognized by the LtrA protein (8) are highlighted in gray.

FIGURE 2.

Bendability profiles of Ll.LtrB group II intron-integration sites and randomly selected *E. coli* DNA sequences. The profiles show average bendability versus nucleotide position calculated for sets of 66 previously identified Ll.LtrB retrohoming sites (red line) and 33 Ll.LtrB retrotransposition sites (blue line). The bendability profiles were compared to those for 100 control sets consisting of the same number of randomly selected *E. coli* DNA sequences (6600 and 3300 individual sequences, respectively). The profiles for the 100 retrohoming control sets are shown as light blue lines. The profiles for the 100 retrotransposition control sets are similar to those for the retrohoming control sets and are not shown in order to simplify the figure. Bendability profiles were calculated over a 6-bp sliding window for positions -30 to +10 from the intron-insertion site by using an algorithm that employs a probabilistic model based on DNase I-susceptibility of different trinucleotides, as described in Experimental Procedures. Dots mark the nucleotide position at the 5' end of six-bp segments that are significantly more bendable in the retrohoming or retrotransposition sites than in their respective control sets, as determined by a statistical test for the difference between two sample means using a 5% significance level. The calculated bendability profile for the wild-type Ll.LtrB intron-insertion site (dashed line) is shown for comparison (note difference in scale on the right).

FIGURE 3.

DNA-integration assays with wild-type and mutant DNA substrates and RNPs. (A) Schematic of DNA substrates used in biochemical and AFM experiments. Substrates are ~240-bp double-stranded DNAs generated by PCR. The wild-type (WT) *Ll.LtrB* DNA target site (positions -30 to +15 from the intron-insertion site) is shown as a gray rectangle, with black shading showing regions changed in mutant DNA substrates. In DNA substrate L, the *Ll.LtrB* target site is moved near the end of the DNA oligonucleotide. Arrowheads above and below indicate the intron-insertion site (IS) and bottom-strand cleavage site (CS), respectively. (B) DNA-integration assays. Wild-type or mutant *Ll.LtrB* RNPs were incubated with internally labeled DNA substrates for 5 min (lane 2) or 30 min (lanes 3–10) at 37°C, as described in Experimental Procedures, and the products were analyzed in a denaturing 4% polyacrylamide gel, which was dried and scanned with a PhosphorImager. Lane 1 shows the WT DNA target site incubated in reaction medium for 30 min at 37°C without RNPs. Positions of products resulting from partial and complete reverse splicing and bottom-strand cleavage are indicated to the left of the gel, and the predicted size of each product (nts) is indicated to the right. (C) Diagram showing products of full- and partial reverse splicing and bottom-strand cleavage reactions with the WT and L DNA substrates. Sizes of DNA segments and the *Ll.LtrB*- Δ ORF intron RNA (not drawn to scale) are indicated.

FIGURE 4.

AFM showing RNPs reacted with DNA substrates. (A) and (B) show representative images of wild-type Ll.LtrB RNPs bound to DNA substrates in which the Ll.LtrB target sequence is at the center (WT) or one end (L) of the DNA oligonucleotide, respectively (see Figure 3). Images were taken in tapping mode over a single μm^2 .

FIGURE 5.

AFM imaging of L1.LtrB RNPs. The figure shows a representative field of RNPs (left panel) imaged as in Figure 4, after incubation in 10 mM KCl, 10 mM MgCl₂, 40 mM HEPES, pH 7.5 for 5 min at 37° C. The center and right panels show diameter and height measurements for 100 imaged RNPs.

FIGURE 6.

AFM imaging of complexes formed with wild-type LI.LtrB RNPs and DNA substrate for different times. The figure shows partial field and expanded views of (A) wild-type (WT) DNA incubated with LI.LtrB RNPs for 5 min at 37°C, highlighting a complex with a 78° DNA-bend angle; (B) wild-type DNA incubated with LI.LtrB RNPs for 30 min at 37°C, highlighting a complex with a 90° DNA-bend angle; (C and D) wild-type and mutant -30/+15 DNA substrates in the absence of RNPs. The histograms beneath the images show the distribution of DNA-bend angles in 100 complexes or DNA molecules for each panel. Gray shaded bell-curves show the bimodal distribution of DNA-bend angles.

FIGURE 7.

AFM imaging of complexes formed with mutant DNA target sites and RNPs that affect 3'-exon interactions and bottom-strand cleavage. (A) Wild-type L1.LtrB RNPs incubated with DNA substrate +4/+15, which has disfavored nucleotide-residues at 3'-exon positions +4 to +15 from the intron-insertion site; (B) wild-type DNA substrate incubated with RNPs containing a mutant LtrA protein lacking the C-terminal En domain (Δ En). The incubations were for 30 min at 37°C. The histograms below the images show the distribution of DNA-bend angles in 100 complexes for each condition.

FIGURE 8.

AFM imaging of complexes formed by wild-type Ll.LtrB RNPs with DNA substrates containing a nick at the bottom-strand cleavage site. (A) and (B) Wild-type Ll.LtrB RNPs incubated with wild-type (WT) and mutant +4/+15 DNA substrates, respectively, containing a pre-positioned nick at the bottom-strand cleavage site. The incubations were for 30 min at 37°C. The histograms below the images show the distribution of DNA-bend angles in 100 complexes for each nicked DNA substrate.

FIGURE 9.

AFM imaging of complexes after initiation of reverse transcription. (A) and (B) show wild-type L1.LtrB RNPs incubated with wild-type DNA substrate at 37°C in the presence of 2 μM each dATP, dTTP, and dCTP for 5 and 30 min, respectively. The highlighted molecules illustrate the predominant bend angle at each time, 78° at 5 min and 90° at 30 min. (C) shows mutant DD⁻ RNPs, with the mutation YADD→YAAA at the RT active site, incubated with wild-type DNA substrate in the presence of 2 μM each dATP, dTTP, and dCTP for 30 min at 37°C. The histograms beneath the images show the distribution of DNA-bend angles in 100 complexes for each condition.

FIGURE 10.

Model showing how RNP integration may cause DNA bending. (A) RNPs bind DNA non-specifically and search for target sites. (B) Initial DNA target site recognition involves IEP interactions with the distal 5'-exon region and base pairing of the intron RNA's EBS2, EBS1, and δ sequences to the IBS2, IBS1, and δ' sequences between positions -12 to +3 of the DNA target site. These initial DNA target site interactions and reverse splicing of the intron RNA into the intron-insertion site (IS) in the top DNA strand do not lead to detectable DNA bending. (C) Subsequent IEP interactions with the 3' exon lead to an initial DNA-bend angle of 75° and bring the scissile phosphate at the bottom-strand cleavage site (CS) to the En active site. (D)

Shift of the 3' end of the cleaved bottom strand from the En to RT active site for initiation of reverse transcription leads to a greater bend angle of 90°.