A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity

Sandra L. Martin[1,2,*], Diane Bushman[1], Fei Wang[3], Patrick Wai-Lun Li[2], Ann Walker[1], Jessica Cummiskey[2], Dan Branciforte[1] and Mark C. Williams[3,4]

[1]Department of Cell and Developmental Biology and Program in Molecular Biology, [2]Human Medical Genetics Program, University of Colorado School of Medicine, Aurora, CO 80045, [3]Department of Physics and [4]Center for Interdisciplinary Research on Complex Systems, Northeastern University, 111 Dana Research Center, Boston, MA 02115, USA

ABSTRACT

L1 is a ubiquitous interspersed repeated sequence in mammals that achieved its high copy number by autonomous retrotransposition. Individual L1 elements within a genome differ in sequence and retrotransposition activity. Retrotransposition requires two L1-encoded proteins, ORF1p and ORF2p. Chimeric elements were used to map a 15-fold difference in retrotransposition efficiency between two L1 variants from the mouse genome, T_{FC} and T_{Fspa}, to a single amino acid substitution in ORF1p, D159H. The steady-state levels of L1 RNA and protein do not differ significantly between these two elements, yet new insertions are detected earlier and at higher frequency in T_{FC}, indicating that it converts expressed L1 intermediates more effectively into new insertions. The two ORF1 proteins were purified and their nucleic acid binding and chaperone activities were examined *in vitro*. Although the RNA and DNA oligonucleotide binding affinities of these two ORF1 proteins were largely indistinguishable, D159 was significantly more effective as a nucleic acid chaperone than H159. These findings support a requirement for ORF1p nucleic acid chaperone activity at a late step during L1 retrotransposition, extend the region of ORF1p that is known to be critical for its functional interactions with nucleic acids, and enhance understanding of nucleic acid chaperone activity.

INTRODUCTION

L1 is an autonomous mammalian retrotransposon that has successfully amplified to comprise 17 and 19% of the human and mouse genomes, respectively. Most of the >600 000 copies of mouse L1 are inactive due to truncations and/or point mutations, but 3000 are estimated to be functional for further transposition (1,2). Full-length, active copies of L1 are 7 kb in length and encode two proteins necessary for retrotransposition. The ORF1 protein (ORF1p) acts as an RNA-binding protein and nucleic acid chaperone protein *in vitro* (3). The RNA-binding activity of ORF1p is necessary but not sufficient for retrotransposition (4,5), and retrotransposition efficiency depends upon nucleic acid chaperone efficacy (4). The ORF 2 protein (ORF2p) has three essential domains; two of these provide the endonuclease (EN; 6) and reverse transcriptase activities (7) required for the target-primed reverse transcription reaction (TPRT, 8) that characterizes the replication mechanism of L1 and other non-LTR retrotransposons.

Retrotransposition rates vary widely among different copies of L1. The evolution of L1 is episodic, typically characterized by one or a few distinct subtypes of L1 that dominate the dispersal process within a species and then become extinct (9). In mice, there are three subfamilies represented among the 3000 active copies of L1. These subfamilies, T_F, A and G_F, are distinguished by their distinct 5′-end sequences. Within each subfamily, individual members vary in their retrotransposition activity as much as several 100-fold, as measured by an antisense-intron (AI) reporter gene assay in cultured cells (1,2). Individual elements from the currently active subfamily of human L1 similarly exhibit different activities in the cultured cell assay. A total of 40 of 82 full-length human L1 sequences in the human genome database that contain intact ORFs were able to retrotranspose when tested in cultured cells. These active elements varied widely in their retrotransposon rates, however, with most of the total retrotransposition activity of the group (84%)

being attributable to just six individual elements. Significantly, of these six elements, the one with the greatest activity had an amino acid sequence most similar to the subfamily consensus (10).

A mouse L1 element on the X chromosome, T_{FC}, has a sequence most like the consensus of the T_F subfamily and was found to retrotranspose 15 times more efficiently than another element of the same subfamily, T_{Fspa} (4). T_{Fspa} recently inserted into the beta-glycine receptor gene, hence it is a known active mouse L1 (11). A total of 20 nt substitutions, including three that cause amino acid replacements, distinguish the two elements. The goal of this study was to define the substitution responsible for this dramatic effect on L1 retrotransposition and determine its mechanism of action. The significant substitution mapped to one of the altered amino acids in ORF1 far N-terminal to the previously described nucleic acid interaction domain of the ORF1 protein (12–14). The substitution affects a late step in retrotransposition and significantly alters the nucleic acid chaperone activity of the ORF1 protein *in vitro*. The results of this work strengthen the hypothesis that the nucleic acid chaperone activity of ORF1p is required for TPRT during L1 retrotransposition (4,15), as well as increase our understanding of the mechanism of action of nucleic acid chaperone proteins.

MATERIALS AND METHODS

Constructs

T_{FC} and T_{Fspa} constructs for the autonomous retrotransposition assay were described previously (4). Chimeric constructs that place either the two ORF1 replacement substitutions or the single ORF2 replacement substitution of T_{Fspa} into the backbone of T_{FC} were made by moving either the NheI-BstWI or the BstWI-SspI fragments, respectively, from T_{Fspa} into the homologous sites of T_{FC} (Figure 1B). The single point mutations to reciprocally alter the two ORF1 amino acids that differ between T_{FC} and T_{Fspa} were made by site-directed mutagenesis in either a T_{FC} or a T_{Fspa} subclone. The NheI-BstWI fragment containing the mutation was then used to replace the homologous fragment of the intact L1 in the retrotransposition assay vector after verification that the desired point mutation was the only change by DNA sequencing.

Cell culture and autonomous retrotransposition

Retrotransposition assays were done in 143B cells as described previously (4), using either G418 resistance or expression of eGFP as the marker of retrotransposition events. Briefly, these reporter cassettes measure retrotransposition because expression of the marker requires excision of an AI in the L1 transcript, then conversion to cDNA and insertion into the genome by TPRT before the reporter can be transcribed into an mRNA that encodes functional protein, i.e. before either growth on medium containing G418 or detection of green fluorescence, is possible. Cells were transfected the day after seeding using Lipofectamine2000 (Invitrogen, Carlsbad, CA, USA), according to the manufacturer's recommendations. Transformants were selected for 24 h in 10 μg/ml
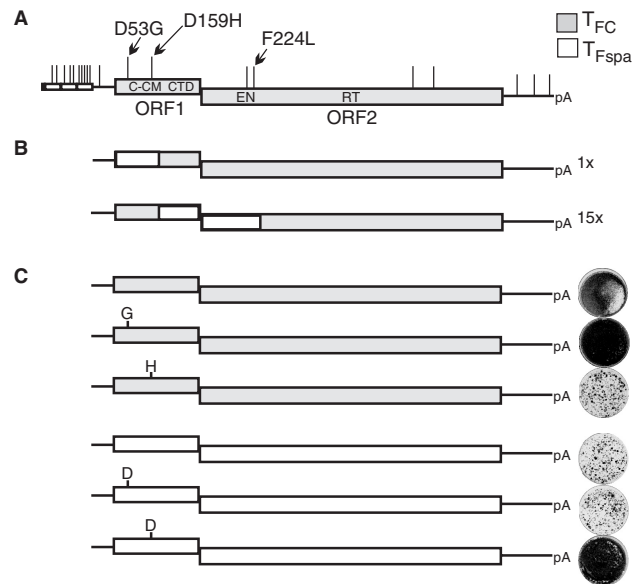
Figure 1. Locating the substitution(s) responsible for elevated retrotransposition of T_{FC} (white box) versus T_{Fspa} (grey box). (**A**) Structure of T_{FC}. From 5′ to 3′, monomers (small rectangles, with promoters), 5′ nc region (line), ORF1 with relative locations of the coiled-coil domain (C-C) and the conserved domain [comprised of the middle or M, domain and the C-terminal domain, CTD (13)], ORF2 with the endonuclease (EN) and reverse transcriptase domains (RT), the 3′ nc region (line), followed by a polyA tail (pA). ORF1 and ORF2 are not aligned because they are separated by a 40 nt intergenic region (32). (**B**) Autonomous retrotransposition assay with chimeric L1s, grey boxes indicate the locations of the restriction fragments from T_{Fspa} that were substituted into the T_{FC} backbone: NheI to BstWI with two replacement substitutions in ORF1 or BstWI to SspI with one silent and one replacement substitution in ORF2, as depicted in panel A. The fold change in retrotransposition as measured with AI-eGFP is given relative to T_{Fspa}. (**C**) Retrotransposition assay on T_{FC} (grey) and T_{Fspa} (white) elements with single amino acid substitutions. Retrotransposition causes G418^R; representative wells show stained G418-resistant cells after 10 d of selection. The uniform dark stain is due to confluent cells, whereas the spots are colonies of G418^R cells, indicative of fewer retrotransposition events.

puromycin beginning 24 h after transfection. For experiments involving the eGFP marker, fluorescence was followed daily by microscopy and quantified by flow cytometry 6 days post-transfection as described (4). For experiments involving the neo marker, cells were transfected as for eGFP, except that cells were allowed to recover for 1 day after puromycin treatment, then placed in G418 (400 μg/ml) and allowed to grow for 10 additional days before fixing and staining with crystal violet. Transfections were identical for immunoflorescence assays except they were done using cells plated on poly-lysine-coated coverslips.

Immunofluorescence microscopy of L1 ORF1p

L1-transfected 143B cells were rinsed with phosphate buffered saline (PBS) and then fixed for 20 min in 4% paraformaldehyde at various times post-transfection. The fixed cells were rinsed in PBS, then blocked and permeabilized for 2 h in PBS containing 3% BSA and 0.1% Triton X-100 (block). The coverslips were incubated overnight at 4°C with rabbit polyclonal anti-ORF1p

antibody (3.4 μg/ml) in block and then rinsed three times in block. Cells were then incubated with 1:500 Cy-5 conjugated goat anti-rabbit antibody (Jackson ImmunoResearch, West Grove, PA, USA) for 1 h at 4°C, rinsed three times in block and mounted on slides with Fluoromount G (Southern Biotech, Birmingham, AL, USA). Fluorescence was imaged and captured using a Zeiss LSM510.

RNA, DNA and protein analyses from transfected cells

Timepoints were taken from L1-transfected 143B cells by harvesting cells every 24 h post-transfection. Cells were recovered with trypsin, washed in PBS and stored frozen as cell pellets at −80°C. Pellets were resuspended in five volumes of PLB (140 mM NaCl, 200 mM Tris–HCl, pH 8.5, 2 mM MgCl$_2$) to which a 1/20 volume of 5% NP40 was added. After gentle mixing followed by 5 min on ice, the lysate was centrifuged 10 min at 2000g. The supernatant was recovered for RNA and protein analysis and the pellet was used for DNA analysis. About 10 μg/ml protease inhibitors (P8340 Protease Inhibitor Cocktail, Sigma-Aldrich, St. Louis, MO, USA) were added to aliquots collected for western blot analysis.

Retrotransposition events were detected at the DNA level by PCR amplification of the spliced reporter gene. About 200 ng of total DNA from various days post-transfection were amplified for 20 cycles with primers A7 (5′-CGTCCATGCCGAGAGTGATCCC) and A8 (5′-GCTACGT CCAGGAGCGCACCATC), followed by 35 cycles with primers A16 (5′-GCTACGTCCAGGA GCG CACCATC and A8.

ORF1p was detected by western blotting using 20 μg of transfected cell lysate (Bradford Assay, Bio-Rad, Hercules, CA, USA) and affinity purified, rabbit anti-ORF1p antibody (16) with the HRP-conjugated goat anti-rabbit antibody Plus Western Blotting Reagent Pack (GE Healthcare, Pittsburg, PA, USA). Images were captured on a Typhoon 9400 (GE Healthcare). Anti-actin antibody (sc-8432) was purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA) and used as recommended by the manufacturer to detect actin on the same blots.

L1 RNA was detected and quantified using RT–PCR. Two microgram of RNA isolated from PLB supernatants using TRIzol LS (Invitrogen) were treated with RQ1-DNase (Promega, Madison, WI, USA) according to the manufacturer's instructions. One microgram of the DNase-treated RNA was added to 20 μl reverse transcriptase reactions as recommended by the manufacturer (Reverse Transcription System, Promega). Reactions were diluted to 100 μl in nuclease-free water and 10 μl were used in 25 μl PCR reactions with primers that amplify a 265 nt region of ORF2 (172 nt downstream of the AUG) in L1 from both T$_{FC}$ and T$_{Fspa}$. The oligonucleotides were 5′-GACACTACCTCAGAATCAAAGGC TGG (forward primer) and 5′-GTGAGGCGCAATGTG TGCTTTGAGC (reverse primer). It was empirically determined that 23 or 25, cycles remained within the linear range of the assay, on days 1–3 or 4–6, respectively. PCR products were separated by electrophoresis through 2% agarose gels, and then stained with ethidum bromide. Fluorescence images were captured on the Typhoon 9400 and analysed using ImageQuant (GE Healthcare).

Assays using purified proteins

D159 (T$_{FC}$) and H159 (T$_{Fspa}$) ORF1 proteins were purified to apparent homogeneity after expression in baculovirus-infected insect cells as described (4). Briefly, insect cells were lysed 48 h post-infection and the His-tagged ORF1 proteins were purified using affinity (Ni-NTA agarose, Qiagen, Valencia, CA, USA) chromatography followed by size-exclusion (Sephacryl S300, GE Healthcare) chromatography. RNA binding, strand exchange and melting temperature (T$_m$) assays were done as described previously (4,15,17).

DNA stretching experiments were performed using a dual beam optical tweezers instrument, described previously (18). Single bacteriophage λ DNA molecules (Roche Applied Science, Indianapolis, IN, USA) were labelled on the 5′-ends with biotin and captured between two 5 μm diameter streptavidin-coated polystyrene spheres (Bangs Labs, Fishers, IN, USA). After capturing a single DNA molecule between two spheres, other spheres and DNA were rinsed out of the optical tweezers trapping chamber, and the solution surrounding the single DNA molecule was exchanged with a solution containing a fixed protein concentration. After the volume was completely exchanged, the flow was stopped and the DNA was stretched in 100 nm steps at ∼1 step/s, and then relaxed back to the original extension at the same rate. The resulting force versus extension was then analysed to determine the amount of hysteresis between stretching and relaxation, the observed aggregation of dsDNA and ssDNA in the presence of protein, and the helix–coil transition width for a given set of solution conditions. All experiments were performed in 10 mM HEPES, 100 mM Na$^+$ solution. To quantify the effects of these proteins on the DNA helix–coil transition, we calculated the transition width as a function of protein concentration, as described previously (4). The results were fit to the McGhee–von Hippel binding isotherm:

$$\Theta = K \times n \times C \frac{(1 - \Theta)^n}{(1 - \Theta + \Theta/n)^{n-1}} \qquad \mathbf{1}$$

where Θ is the fractional change in transition width, K is the equilibrium association constant for protein binding to DNA and n is the binding site size. For simplicity we set $n = 1$ so that the equilibrium constant is the only fitting parameter.

Kinetic experiments on oligonucleotides were performed on an upgraded Biacore 3000 instrument at 25°C. The 5′-biotinylated AAAAAGTACACAGTCTAA CATCAACTCGC was annealed to either 5′-GCGAG TTGATGTTAGACTGTGTACTTTTT to make a perfectly matched, dsDNA duplex or to 5′-GCGAGTTG ACGTCAGACCGTGCACTTTTT to make the mismatched dsDNA duplex. Biotinylated oligonucleotide was captured on a CM4 chip first derivitized with NeutrAvidin biotin-binding protein (Pierce, Rockford, IL, USA) via amine coupling. dsDNA constructs

(perfectly matched and mismatched) were hybridized on the chip in running buffer (50 mM phosphate buffer, 250 mM NaCl, 0.1 mM EDTA, pH 7.6). The instrument was programmed for iterative cycles in which each kinetic cycle consisted of: (i) 300 s protein injection phase, (ii) 300 s or greater dissociation phase depending on affinity and (iii) a 120 s regeneration phase. A flow rate of 20 µl/min was maintained throughout the cycle. The concentration of proteins analysed ranged from 10 to 300 nM. The surface plasmon resonance (SPR) signal was recorded in real time every 0.5 s. Each sensorgram obtained was corrected for bulk refractive index changes by subtracting the corresponding protein injection cycle on a blank NeutrAvidin surface. The association and dissociation rate constants (k_{on} and k_{off}, respectively) for the interaction were calculated by globally fitting the data using different kinetic models available in the BIA evaluation software package with a simple 1:1 bimolecular Langmuir interaction model.

RESULTS

T_{FC}, a mouse L1 element with the sequence of the T_F subfamily consensus, retrotransposes 15-fold more frequently than T_{Fspa} in an autonomous retrotransposition assay (4). The 20 nt substitutions that distinguish these two elements are represented schematically in Figure 1A. Ten of these are in the monomers, which potentially could impact transcription and hence retrotransposition *in vivo*. However, the monomers are not present in the constructs used for the autonomous assay, rather the CMV promoter is used to drive transcription of both T_{Fspa} and T_{FC}; thus, those 10 substitutions cannot account for the difference observed between the two elements using this assay. Just three of the remaining 10 nt substitutions cause amino acid replacements: two in ORF1 and one in ORF2. In addition, there are three single nucleotide substitutions in ORF2 that are silent at the amino acid level, another in the 5′ non-coding region and three in the 3′ non-coding region.

The ORF2 replacement altered a conserved EN domain in ORF2p (6), whereas the two ORF1 replacements both lie in a relatively non-conserved coiled-coil forming region of ORF1p (13,19). All three of the replacement substitutions were rare among the 539 aligned sequences (4) that were used to derive the consensus. In ORF1, D53G was present in just five of the sequences including T_{Fspa}, D53N occurred once and D159H was present only in T_{Fspa}. In ORF2, F224L also appeared only once in T_{Fspa}, but two of the sequences had F224Y and two others had F224C. Chimeric L1 constructs were made to test the importance of the ORF2 versus the two ORF1 substitutions in the retrotransposition assay by replacing either of two restriction fragments of T_{FC} with the homologous fragments from T_{Fspa}; these two chimeric elements are both comprised of largely T_{FC}, but introduce either the single amino acid replacement (together with one nearby silent substitution) of ORF2, or the two amino acids of ORF1 from T_{Fspa} into the T_{FC} backbone. The elevated retrotransposition activity of T_{FC} unambiguously mapped to the fragment containing the two replacements in ORF1, rather than to the fragment which altered ORF2 (Figure 1B). To determine whether one or both amino acid replacements in ORF1 were important for the enhanced activity of T_{FC} and the diminished activity of T_{Fspa}, as well as to eliminate the possibility that one of the 'silent' substitutions elsewhere was critical, the two ORF1 amino acid replacements of T_{FC} were individually introduced into T_{Fspa} and the two replacements of T_{Fspa} were likewise introduced into T_{FC}. Subsequent retrotransposition assays with these altered elements identified the aspartic acid at position 159 (D159) in ORF1p as the crucial amino acid responsible for most, if not all, of the 15-fold increase in retrotransposition activity exhibited by T_{FC} (Figure 1C).

The kinetics of retrotransposition, as well as the expression of L1 RNA and ORF1p were examined in cells transfected with L1 T_{FC} or T_{Fspa} marked with the eGFP AI reporter. Transfected cells were examined daily for the presence of green cells, which indicate retrotransposition of L1 from the transfected plasmid into genomic DNA (20). Cells transfected with T_{FC} expressed eGFP at higher frequency throughout the timecourse compared with those transfected with T_{Fspa}; green cells were also initially apparent on day 3 post-transfection with T_{FC}, versus day 4 with T_{Fspa} (Figure 2A and B). Interestingly, simultaneous detection of ORF1p by immunofluorescence suggests that ORF1p accumulates similarly when expressed from the two elements (Figure 2A) because there is no evidence for a lag in expression in cells transfected with T_{Fspa} compared with T_{FC}, nor does there appear to be any difference in either the intensity or the subcellular localization of ORF1p over the timecourse examined (data not shown, but see also Figure 3). With both T_{FC} and T_{Fspa}, retrotransposition leading to eGFP expression occurred in a minor fraction of the ORF1p expressing cells.

The results of the imaging studies were corroborated and extended using biochemical techniques. The finding that transposition occurred earlier in cells expressing T_{FC} than T_{Fspa} was confirmed by PCR amplification of genomic DNA using a strategy that allowed simultaneous detection of the transfected eGFP DNA (intron present) and the retrotransposed copy (intron absent, Figure 2C). The spliced eGFP gene, indicative of successful retrotransposition, appeared earlier in genomic DNA after transfection and accumulated to higher levels in T_{FC}-transfected cells compared with those transfected with T_{Fspa}, as expected based upon the results obtained by following the appearance of green cells. Consistent with the intensity and distribution of ORF1p observed by indirect immunofluorescence, the steady-state levels of ORF1p detected by western blotting did not differ between the two elements over the timecourse examined (Figure 3). Likewise, there were no significant differences in the steady-state level of L1 RNA detected by semi-quantitative RT–PCR between T_{FC} and T_{Fspa} (Figure 4).

ORF1p is an RNA-binding protein that forms large complexes with L1 RNA (5,12,16,17,21). It is also a nucleic acid chaperone; mutations that compromise chaperone activity block or diminish retrotransposition (4). In order to examine these activities of ORF1p, we isolated the D159 (T_{FC}) and H159 (T_{Fspa}) variant forms of ORF1p
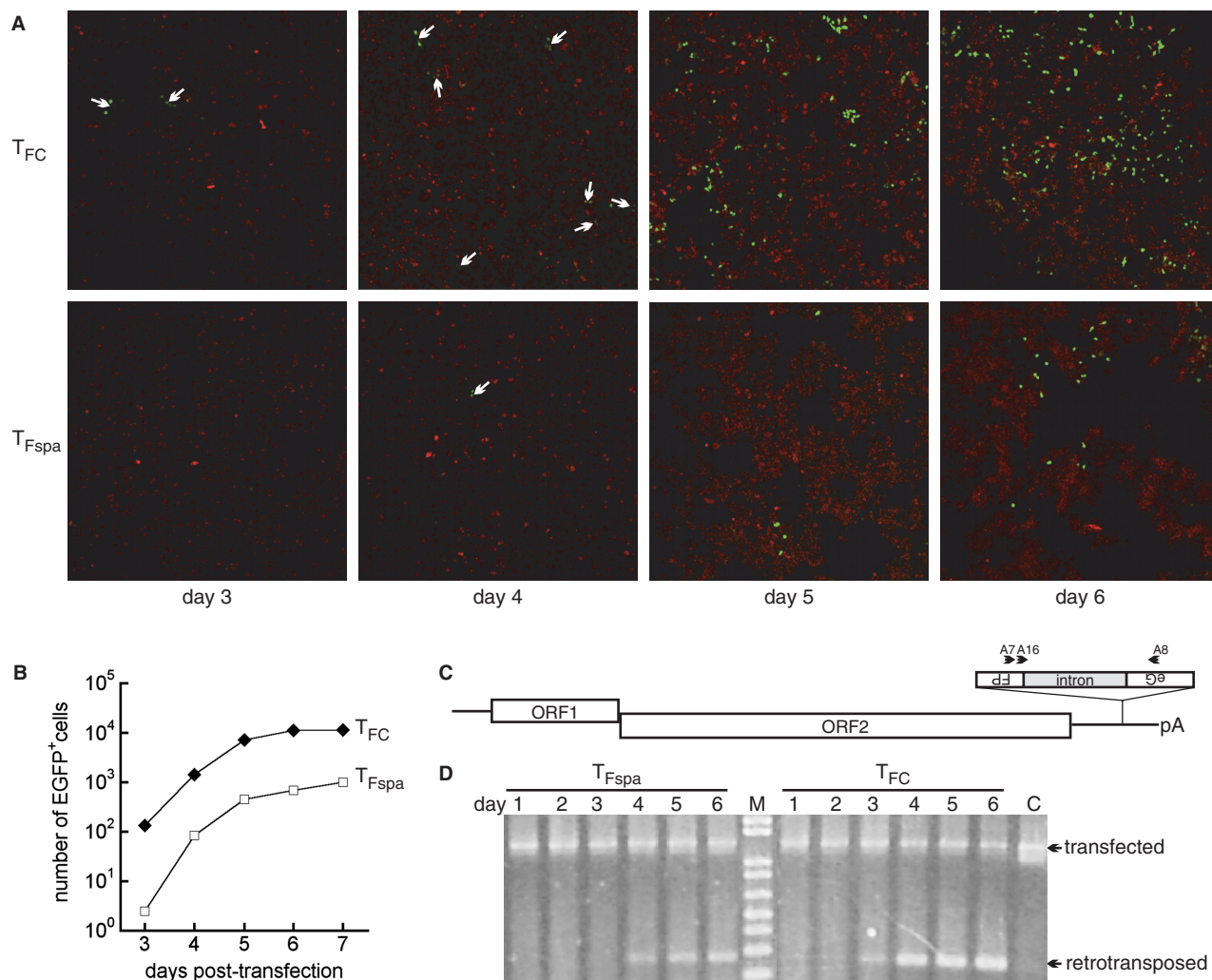
Figure 2. Timecourse of L1 retrotransposition. (**A**) Representative micrographs of 143B cells captured daily after transfection with T_{FC} or T_{Fspa}. Cells with new insertions by retrotransposition are green, cells expressing ORF1p are red, each field is 1.3 mm^2. Arrows on images for days 3 and 4 point to clusters containing at least two green cells. (**B**) Quantification of cells expressing eGFP following transfection. (**C**) Schematic of the L1 retrotransposition reporter, showing location of primers used for RT–PCR detection of unspliced and spliced eGFP DNA. The antisense eGFP gene is depicted as a white rectangle interrupted by a grey intron (in the same orientation as the L1 ORFs). Amplification across the intron results in a 1334 nt PCR product from transfected DNA, whereas after splicing and cDNA synthesis due to retrotransposition the product is 263 nt. (**D**) Ethidium bromide-stained agarose gels of PCR products obtained following a two-step PCR amplification of genomic DNA isolated from transfected cells (days 1–6 post-transfection with T_{FC} or T_{Fspa}) or the control transfected plasmid DNA (lane C). PCR was 20 cycles of amplification using the A7 and A8 primers followed by a second amplification using the primers A16 and A8. A16 spans the splice junction. The 1334 nt band amplified from transfected DNA containing the intron is readily distinguished from the 263 nt band amplified from retrotransposed DNA and thus lacking the intron. The intronless band is observed earlier and is more intense throughout the timecourse in DNA from T_{FC}-transfected cells compared with T_{Fspa}-transfected cells. The marker lane (M) shows the bands from 2000 to 200 nt from the 1 Kb Plus DNA Ladder (Invitrogen).

from baculovirus-infected insect cells. The two proteins behaved identically throughout protein purification (data not shown), eluting from size-exclusion chromatography in the identical fraction characteristic of the elongated trimer form of the protein as described previously for the T_{Fspa}, H159 variant (19). The ellipticity and T_m of these two purified proteins were also equivalent as determined by circular dichroism (−26 000 and −26 800 and 51 and 49.5°C, for H159 and D159, respectively).
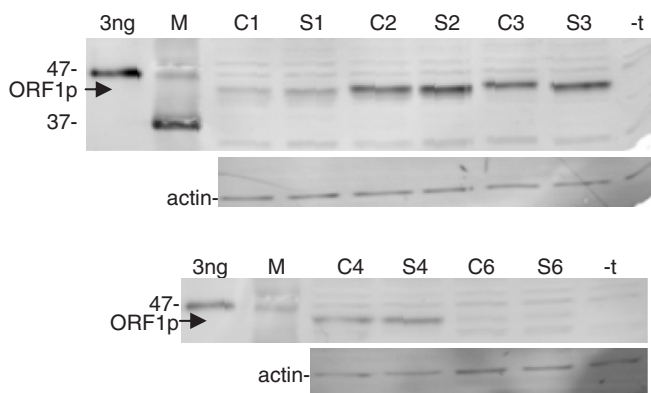
The affinity of both proteins for RNA was measured using a nitrocellulose filter-binding assay. In a side-by-side comparison containing increasing concentrations of D159 or H159 ORF1p and 25 pM of an antisense 111 nt L1 RNA in 250 mM NaCl, the apparent K_d of D159 for

RNA was 8.9 ± 1.0 nM, and H159 was 7.0 ± 0.6 nM, respectively (data not shown). This 20% change is not significant based upon multiple experiments and is unlikely to explain the 15-fold decrease in L1 retrotransposition associated with D159H because a 38% drop in apparent affinity for RNA by the ORF1p mutant, R298K, decreases retrotransposition by just 56% (4).

The nucleic acid chaperone activity of ORF1p can be assessed by determining the T_m of a mismatched dsDNA oligonucleotide in the presence of protein (4). About 30 nM H159 ORF1p shifts the T_m of a 29 nt DS DNA oligonucleotide with four non-contiguous mismatches from 42°C to 22°C. At the same concentration of protein, D159 ORF1p has a significantly different effect, with only a

small fraction of the mismatched double-stranded oligo-nucleotide being converted to single-stranded form (Figure 5A). This effect occurs over a broad concentration range of protein (compare B and C in Figure 5), and suggests that H159 ORF1p interacts more strongly with single-stranded DNA than D159 ORF1p as the duplex transiently or fully melts.
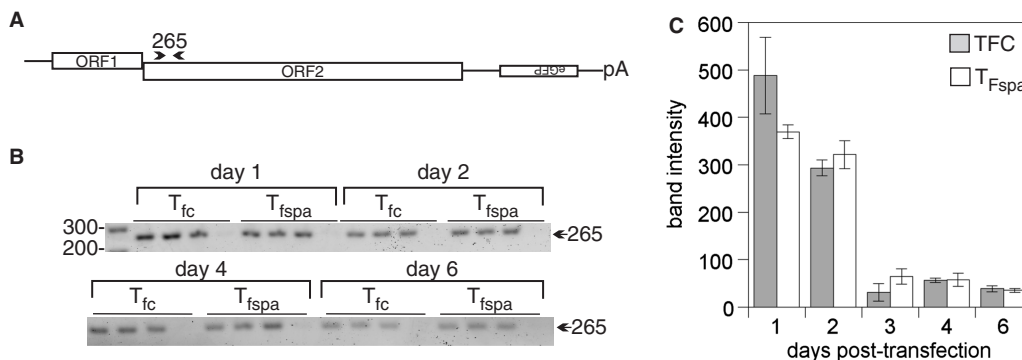
This difference in the interaction of ORF1p with a mismatched DNA oligonucleotide was explored in more detail using SPR. The interactions of both ORF1p proteins with ssDNA and with dsDNA containing either perfect or imperfect heteroduplex were examined for comparison. The interactions of the two proteins with the ssDNA oligonucleotide, as well as with the perfectly matched dsDNA oligonucleotide were similar. In contrast,

a relatively large difference was observed between these two proteins in their interaction with the imperfect double-strand duplex of the same length; D159 ORF1p displays rapid kinetics of association and dissociation with the mismatched duplex whereas H159 ORF1p dissociates 10 times more slowly (Table 1).

Single-molecule analysis of DNA stretching is a sensitive assay for nucleic acid chaperone activity (see 18, and references therein). Typical DNA stretching experiments in the absence of protein and in the presence of 15 nM D159 or H159 ORF1p are shown in Figure 6A. The stretching curves (solid lines) for DNA in the presence of both proteins show significant changes in the shape of the helix–coil transition. In the absence of protein, at a force of about 60 pN, the DNA stretching force increases very little as the DNA extension is increased by a factor of about 1.7. This plateau represents a cooperative DNA helix–coil transition, and the length at which the force begins to increase dramatically at the end of the transition (indicated by an arrow) demonstrates almost complete conversion of the DNA from double- to single-stranded form. The width of the transition, which is smaller for a more cooperative transition, is only about 4 pN in the absence of protein. In contrast, the transition width (or the force change from the beginning to the end of the plateau) is much higher in the presence of both proteins, but D159 ORF1p shows a much greater increase in the transition width relative to that observed for H159 ORF1p at this protein concentration. Finally, both proteins induce a change in the extension at which the ssDNA stretching force increases at the end of the transition, shown by the arrows. This change in ssDNA extension represents ssDNA aggregation, in which protein-induced effects make ssDNA attracted to itself, thus effectively decreasing the ssDNA length at a given force. The magnitude of the ssDNA aggregation is clearly much greater in the case of H159 ORF1p. To quantify the effects of these

Figure 3. Timecourse of ORF1p expression. Representative western blot of proteins recovered from untransfected cells (-t) or after transfection with T_{FC} (C) or T_{Fspa} (S) L1 constructs at days 1–4 and 6 following transfection. Lanes marked 3 ng contain baculovirus-expressed ORF1p fusion protein (47 kDa). Sizes on the left are in kDa, the position of native ORF1p is indicated by the arrow. The actin signal from the same lanes on these blots is shown below to verify approximately equal loading.

Figure 4. Timecourse of L1 RNA expression. 143B cells were transfected with either T_{FC} or T_{Fspa} containing the eGFP AI reporter gene in the 3′-UTR, and then harvested on the indicated days following transfection for isolation of RNA. (**A**) Schematic of L1RNA and location of primers to detect L1RNA; these primers amplify a 265 nt region near the 5′-end of ORF2. (**B**) Fluorescence image of ethidium bromide-stained gels after RT–PCR; each set of four lanes contains the PCR product from the same DNAse-treated RNA RT–PCR in triplicate plus a control reaction without reverse transcriptase. Twenty-three cycles of PCR were performed on cDNA from days 1 and 2, and 25 for days 4 and 5, to remain in the linear range of the assay. (**C**) Steady-state levels of L1 RNA as determined by semi-quantitative RT–PCR using primers within ORF2. PCR bands were quantified from fluorograms and are plotted as mean ±SD from triplicates; the actual above background intensity values for days 4 and 6 were divided by four to correct for the cycle number difference between these 2 days and days 1–3. This experiment is representative of the results obtained in three separate transfection experiments. The steady-state levels of L1 RNA in transfected cells were not significantly different between T_{FC} (grey bars) and T_{Fspa} (white bars) on any day across the timecourse (Student's *t*-test, $P > 0.05$).
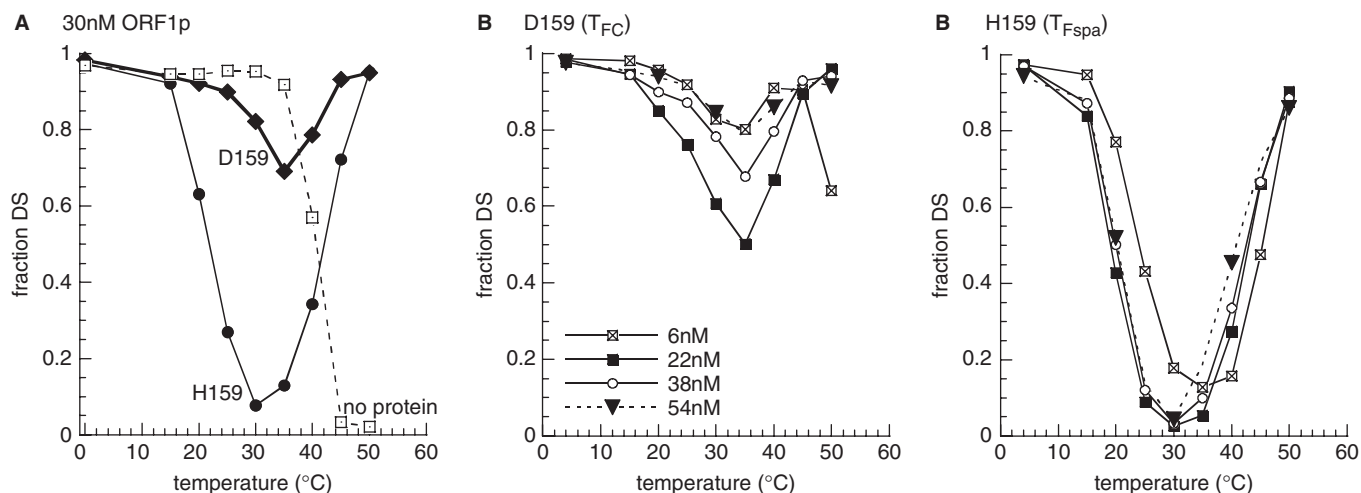
Figure 5. T_m assay for nucleic acid chaperone activity of ORF1p. Addition of ORF1p alters the T_m of a mismatched 29 nt dsDNA oligonucleotide as measured by conversion of the double-stranded form to single-stranded form. (**A**) ORF1p from T_{FC} (D159) and T_{Fspa} (H159) at the same concentration (trimer) compared with naked DNA. (**B**) T_m of the mismatched oligonucleotide in the presence of increasing concentrations of D159 ORF1p or (**C**) H159 ORF1p. ORF1p (trimer) concentrations are identical to those in (**B**). The temperature range needed to fully denature the DNA in the absence of protein is shown; both proteins begin to denature at 35°C, which enhances their charge neutralization capabilities and DNA reannealing (4).

Table 1. Kinetics of interaction of ORF1p with short DNA oligonucleotides

	ssDNA binding			dsDNA perfect binding			dsDNA mismatch binding		
	k_a	k_d	K_D	k_a	k_d	K_D	k_a	k_d	K_D
H159	4.7E+04	6.0E−04	1.3E−08	3.8E+04	8.1E−01	2.1E−05	9.2E+03	8.2E−03	8.9E−07
D159	3.7E+04	4.3E−04	1.2E−08	2.1E+04	7.9E−01	3.7E−05	1.5E+04	8.2E−02	5.3E−06

Measurements obtained by SPR with the indicated 29 nt oligonucleotide immobilized via a 5′ biotin, protein added at 10, 20 and 40 nM. Units are: k_a, 1/Ms; k_d, 1/s; κ_D, M. Measurement error is ~10%.
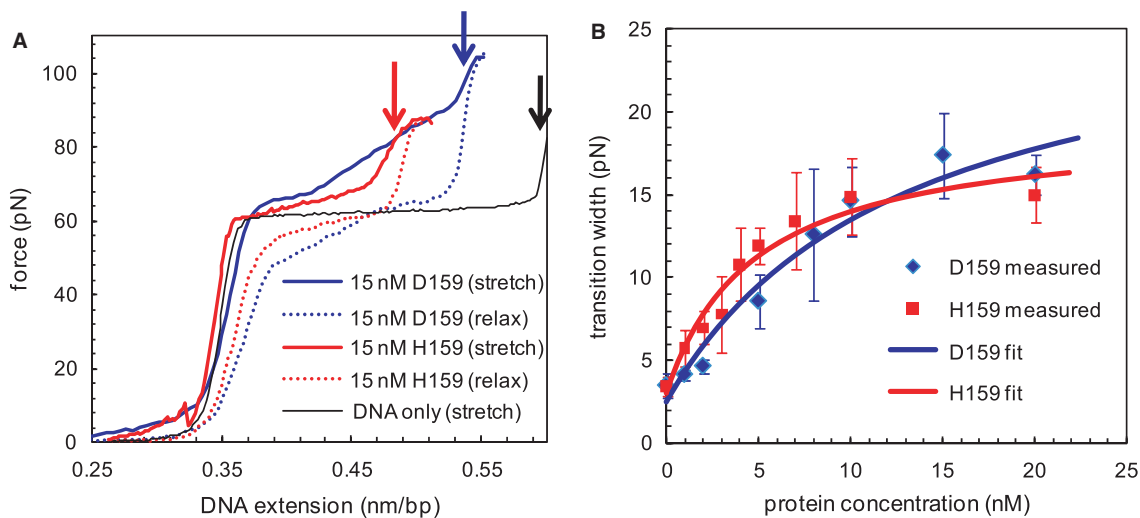
proteins on the DNA helix–coil transition, we calculated the transition width as a function of protein concentration, as described previously (4). The results are shown as data points in Figure 6B, along with lines that represent fits to the McGhee–von Hippel binding isotherm (Methods section). A saturated transition width of 27.5 ± 1.2 pN and an equilibrium association constant of $0.78 \pm 0.08 \times 10^8 M^{-1}$ $(K_D = 1.28 \pm 0.13 \times 10^{-8} M)$ was calculated from the fit to the data obtained using D159 ORF1p, compared with a saturated transition width of 19.2 ± 0.8 pN and an equilibrium association constant of $2.1 \pm 0.2 \times 10^8 M^{-1}$ $(K_D = 0.48 \pm 0.05 \times 10^{-8} M)$ from the data obtained in the presence of H159 ORF1p. These equilibrium binding constants for ssDNA determined from DNA stretching experiments agree well with those presented in Table 1 from bulk measurements.

DISCUSSION

The human and mouse genomes contain hundreds or thousands of intact and therefore potentially retrotransposition-competent L1s, respectively. The activities of individual L1 elements vary widely when tested in a cultured cell assay (1,2,10). Analogous variations in retrotransposition activity are also detected on an evolutionary timescale

through phylogenetic analysis of genomic sequences (22–24); thus this property of individual variation in the ability to produce progeny is likely intrinsic to L1 and not an artifact of the cultured cell assay. T_{Fspa}, the first active mouse L1 described, was isolated because it had retrotransposed into the glycine receptor beta subunit and disrupted the expression of this critical neurotransmitter receptor gene (11,25). T_{FC}, on the other hand, was predicted to be an active element based upon its near identity to the consensus sequence of all of the T_F elements in the mouse genome that are closely related to T_{Fspa} (4). Although both elements are active, T_{FC} retrotransposes 15-fold more effectively than T_{Fspa} in the assay used here. The goal of this investigation was to determine the molecular basis for this difference in order to gain insight into the mechanism and control of L1 retrotransposition.

The increased activity of T_{FC} was mapped to aspartate 159 in ORF1p (Figure 1), which affects a step that follows the accumulation of L1 intermediates (RNA and ORF1p) but precedes the successful insertion of a new cDNA copy of L1 into genomic DNA (Figures 2–4). Thus, the L1 expression products, RNA and ORF1p, are more effectively converted to new insertion events in T_{FC} compared with T_{Fspa}. RNA binding and nucleic acid chaperone activities are the only two essential retrotransposition

Figure 6. Effect of ORF1p on DNA as measured by DNA stretching. (**A**) Effect of T_{FC} (D159) and T_{Fspa} (H159) ORF1 proteins on the helix–coil transition of λ DNA. Stretching (continuous lines) and relaxing (dotted lines) curves for λ DNA, either without protein (black) with 15 nM D159 (blue), or H159 (red) ORF1p from T_{FC} and T_{Fspa}, respectively. Arrows indicate ssDNA contour length. The average maximum stretching force was the same for both forms of ORF1p. (**B**) Binding titrations for ORF1p T_{Fc} and T_{Fspa} variant binding to DNA as measured by DNA stretching. Scattered data points represent measurements and continuous curves are fits to the data using the McGhee–von Hippel isotherm (as shown in Equation 1 in Methods section). For fits to D159 ORF1p titration data (blue), we obtain a saturated transition width of 27.5 ± 1.2 pN and an equilibrium association constant of $0.78 \pm 0.08 \times 10^8 M^{-1}$. For H159 (red), we obtain a saturated transition width of 19.2 ± 0.8 pN and an equilibrium association constant of $2.1 \pm 0.2 \times 10^8 M^{-1}$.

functions presently attributable to ORF1p (4,5). This study revealed that H159 ORF1p has a diminished nucleic acid chaperone activity compared with D159 ORF1p (Figures 5 and 6), although its affinities for RNA, a short ssDNA oligonucleotide, a perfectly base paired short dsDNA oligonucleotide and a long dsDNA were not affected (data not shown, Table 1, Figure 6). R297K and RR297:298KK substitutions in ORF1p disrupt nucleic acid chaperone activity without affecting RNA binding, but RR297:298AA causes a significant drop in RNA affinity and also destroys nucleic acid chaperone activity (4). Hence, no mutations in ORF1p are known to significantly reduce RNA binding without disrupting nucleic acid chaperone activity; such mutations would be useful for understanding the relationship between the RNA binding and nucleic acid chaperone activities of ORF1p.

The location of the amino acid substitution responsible for the improved nucleic acid chaperone function of ORF1p and the elevated retrotransposition activity of T_{FC} was surprising based upon our understanding of the structure and function of L1 ORF1p prior to this work. Primary sequence analysis had revealed two domains in ORF1p: (i) an N-terminal, coiled-coil domain which is highly divergent and perhaps not homologous among mammalian L1s and (ii) a C-terminal conserved domain that is shared among all mammalian L1s and some non-LTR retrotransposons in fish (3, and references therein). Results of several studies indicate that the coiled-coil region is necessary and sufficient for the formation of highly stable ORF1p homotrimers, and the basic region of the conserved domain is likewise both necessary and sufficient for high-affinity interactions between ORF1p and RNA (12,14,19). The crucial residue for high retrotransposition activity identified here, D159, lies within the

coiled-coil domain, but near its C-terminus (13); this is the first report of residues that are critical for functional protein–nucleic acid interactions in this region of ORF1p. Based upon the transfer of ^{32}P from RNA to protein, a polypeptide with residues 244–371 of ORF1p binds RNA, but one with residues 1–251 does not (14). In addition, four consecutive alanine substitutions for REGK beginning at residue 235 in ORF1p from human L1 (homologous to 271 in mouse ORF1p) alter interaction of ORF1p with RNA (5). Thus, D159 in mouse L1 ORF1 is at least 85 amino acids N-terminal to the closest residue previously shown to be involved in nucleic acid interactions. Although it is possible that these regions are adjacent to one another in the presently unknown 3D structure of ORF1p, our data strongly suggest the presence of a heretofore unrecognized site on the protein for nucleic acid interactions.

The binding constant calculated from DNA stretching was similar for D159 and H159 ORF1 proteins, in agreement with the SPR analysis results shown in Table 1. In addition, the amount of hysteresis observed between the stretching and relaxation curves is also similar for these two variants. There are two primary differences between the DNA stretching results for D159 ORF1p and those obtained for H159 ORF1p. First, the amount of ssDNA aggregation is much greater for H159 relative to that observed in the presence of D159. Second, the maximum helix–coil transition width as predicted from fits to the binding titrations is greater for D159 ORF1p by about 40%. An increase in helix–coil transition width is positively correlated with nucleic acid chaperone activity (26–28), however, this increase in transition width with protein binding involves the combination of several effects related to chaperone activity that are not easily separated (29).

Recent studies demonstrated that the primary determinants of efficient nucleic acid chaperone activity are the capability to induce nucleic acid attraction, the ability to partially but not completely destabilize the DNA helix, and the ability to rapidly switch between ssDNA- and dsDNA-binding modes (30,31). Therefore, the chaperone activity of a specific protein is the result of a balance between competing effects. For example, the ability to induce DNA attraction, or aggregation of DNA, facilitates nucleic acid rearrangements by bringing complementary strands together. Conversely, this property also tends to stabilize the DNA helix and inhibit the mobility of the DNA–protein complex, thereby inhibiting rearrangements of nucleic acid secondary structure. HIV-1 nucleocapsid protein (NC) is a well-studied example of a protein that has optimized these various competing effects (31). Subtle changes in the architecture of the zinc fingers in NC destroy this delicate balance, resulting in an inefficient nucleic acid chaperone that is defective in retroviral replication (30).

The two ORF1p proteins studied here exhibit all of the characteristics of a nucleic acid chaperone, but to subtly different extents. They both aggregate DNA, bind preferentially to ssDNA and therefore stabilize the DNA helix, and do not strongly inhibit annealing of long DNA strands. These general nucleic acid chaperone characteristics were previously demonstrated using DNA stretching measurements for the T_{Fspa} ORF1p (4), and are also apparent in the results of DNA stretching experiments with T_{FC} ORF1p (Figure 6A). More subtle features of the stretching experiments distinguish the two ORF1 proteins, however; the results show significantly lower aggregation for the T_{FC} ORF1p, suggesting that DNA complexes with this protein will have increased mobility and therefore increased nucleic acid chaperone activity. The hypothesis that lower aggregation of ssDNA results in increased mobility of protein–DNA complexes is supported by the SPR results, which provide evidence of more rapid binding kinetics for the T_{FC} ORF1p with a mismatched dsDNA oligonucleotide compared with the ORF1p from T_{Fspa}. It is also consistent with much stronger effect in melting the mismatched dsDNA oligonucleotide in the gel-based T_m experiments. Thus, the biophysical data presented here are fully consistent with the observed stronger nucleic acid chaperone capabilities of D159 compared with H159 ORF1p.

The results of earlier studies established that mutations in the C-terminus of T_{Fspa} ORF1p greatly inhibited nucleic acid chaperone activity, which in turn abolished L1 retrotransposition (4). There, the primary effect of the mutations on DNA stretching was to induce such strong ssDNA and dsDNA aggregation that the DNA could not be melted by force. Those earlier results illustrated that aggregation that is too strong inhibits chaperone activity. This hypothesis is supported by other studies in which similar ssDNA aggregation effects were observed for DNA stretching in the presence of HIV-1 Gag, a nucleic acid packaging protein (29), as well as in the presence of HIV-1 NC variants that were identified as poor nucleic acid chaperones (30). Increased DNA aggregation that results in decreased chaperone activity could be due directly to changes in a DNA-binding region of the protein, as is likely the case with mutations in the zinc finger regions of HIV-1 NC (30). It is possible that a tighter interaction with nucleic acids by H159 is a simple reflection of the more basic nature of histidine than glutamate. If D159H does directly alter a DNA binding site in ORF1p, it is not likely to be the previously known site because interactions of ORF1p with nucleic acids have consistently mapped to the distant C-terminal third of the protein (5,12–14). The known, C-terminal binding site is apparently unaltered by the D159H substitution, based upon the indistinguishable affinities of D159 and H159 ORF1p for RNA and ssDNA.

Alternatively, increased DNA aggregation that inhibits chaperone activity could also occur with amino acid substitutions that increase or alter homotypic protein–protein interactions, as is likely the explanation for the increased aggregation that was observed for HIV-1 Gag and the Gag cleavage product NCp9. For these two proteins, the DNA binding site remained intact, yet aggregation was increased and nucleic acid chaperone activity decreased compared with NCp7 (29). Given the location of D159H in the coiled-coil domain (13) where it is likely exposed to solvent, it is plausible that this residue could be involved in further protein–protein interactions between trimers. The differences observed between D159 and H159 in the assays involving short oligonucleotides are not easily explained by such protein–protein interactions, however, because each trimer occupies 50 nt (12). Hence, the 29 nt oligos used for T_m and SPR studies would only bind a single ORF1p trimer and would not be expected to interact with more than one trimer at a time.

A full understanding of the interactions of ORF1p with single and double-stranded nucleic acids, and the relationship between its high-affinity RNA-binding function and the nucleic acid chaperone function necessarily awaits a high resolution structure of the protein with and without its various nucleic acid ligands. Nevertheless, the results presented here illustrate the importance of maintaining a delicate balance between strong DNA binding and DNA–protein complex mobility for efficient nucleic acid chaperone activity. This work also lends further support to the conclusion that the nucleic acid chaperone activity of ORF1p plays an essential role during L1 retrotransposition; this role likely occurs late in the process, consistent with a function in facilitating the strand exchanges that are required to initiate TPRT or melting secondary structure in the RNA template for reverse transcription (15).

REFERENCES

1. DeBerardinis,R.J., Goodier,J.L., Ostertag,E.M. and Kazazian,H.H. Jr. (1998) Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.*, **20**, 288–290.
2. Goodier,J.L., Ostertag,E.M., Du,K. and Kazazian,H.H.J. (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.*, **11**, 1677–1685.
3. Martin,S.L. (2006) The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. *J. Biomed. Biotechnol.*, **2006**, 45621–45626.
4. Martin,S.L., Cruceanu,M., Branciforte,D., Li,P.W.-L., Kwok,S.C., Hodges,R.S. and Williams,M.C. (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.*, **348**, 549–561.
5. Kulpa,D.A. and Moran,J.V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.*, **14**, 3237–3248.
6. Feng,Q., Moran,J.V., Kazazian,H.H.J. and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
7. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H. Jr. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
8. Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chomosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.
9. Furano,A.V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog. Nucleic Acid Res. Mol. Biol.*, **64**, 255–294.
10. Brouha,B., Schustak,J., Badge,R.M., Lutz-Prigge,S., Farley,A.H., Moran,J.V. and Kazazian,H.H.J. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA*, **100**, 5280–5285.
11. Kingsmore,S.F., Giros,B., Suh,D., Bieniarz,M., Caron,M.G. and Seldin,M.F. (1994) Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nat. Genet.*, **7**, 136–142.
12. Basame,S., Li,P.W.-1., Howard,G., Branciforte,D., Keller,D. and Martin,S.L. (2006) Spatial assembly and RNA binding stoichiometry of a protein essential for LINE-1 retrotransposition. *J. Mol. Biol.*, **357**, 351–357.
13. Januszyk,K., Li,P.W., Villareal,V., Branciforte,D., Wu,H., Xie,Y., Feigon,J., Loo,J.A., Martin,S.L. and Clubb,R.T. (2007) Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *J. Biol. Chem.*, **282**, 24893–24904.
14. Martin,S.L., Li,J. and Weisz,J.A. (2000) Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *J. Mol. Biol.*, **304**, 11–20.
15. Martin,S.L. and Bushman,F.D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell Biol.*, **21**, 467–475.
16. Martin,S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell Biol.*, **11**, 4804–4807.
17. Kolosha,V.O. and Martin,S.L. (2003) High affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J. Biol. Chem.*, **278**, 8112–8117.
18. McCauley,M.J. and Williams,M.C. (2007) Mechanisms of DNA binding determined in optical tweezers experiments. *Biopolymers*, **85**, 154–168.
19. Martin,S.L., Branciforte,D., Keller,D. and Bain,D.L. (2003) Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl Acad. Sci. USA*, **100**, 13815–13820.
20. Ostertag,E.M., Prack,E.T.L., DeBerardinis,R.J., Moran,J.B. and Kazazian,H.H. Jr. (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.*, **28**, 1418–1423.
21. Hohjoh,H. and Singer,M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630–639.
22. Furano,A.V., Duvernell,D.D. and Boissinot,S. (2004) L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet.*, **20**, 9–14.
23. Boissinot,S., Chevret,P. and Furano,A.V. (2000) L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.*, **17**, 915–928.
24. Khan,H., Smit,A. and Boissinot,S. (2006) Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.*, **16**, 78–87.
25. Mullhard,C., Fischer,M., Gass,P., Simon-Chazottes,D., Guenet,J.-L. and Becker,C.-M. (1994) The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. *Neuron*, **13**, 1003–1015.
26. Williams,M.C., Gorelick,R.J. and Musier-Forsyth,K. (2002) Specific zinc-finger architecture required for HIV-1 nucleocapsid protein's nucleic acid chaperone function. *Proc. Natl Acad. Sci. USA*, **99**, 8614–8619.
27. Williams,M.C., Rouzina,I. and Bloomfield,V.A. (2002) Thermodynamics of DNA interactions from single molecule stretching experiments. *Acc. Chem. Res.*, **35**, 159–166.
28. Williams,M.C., Rouzina,I., Wenner,J.R., Gorelick,R.J., Musier-Forsyth,K. and Bloomfield,V.A. (2001) Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proc. Natl Acad. Sci. USA*, **98**, 6121–6126.
29. Cruceanu,M., Urbaneja,M.A., Hixson,C.V., Johnson,D.G., Datta,S.A., Fivash,M.J., Stephen,A.G., Fisher,R.J., Gorelick,R.J., Casas-Finet,J.R. *et al.* (2006) Nucleic acid binding and chaperone properties of HIV-1 Gag and nucleocapsid proteins. *Nucleic Acids Res.*, **34**, 593–605.
30. Cruceanu,M., Gorelick,R.J., Musier-Forsyth,K., Rouzina,I. and Williams,M.C. (2006) Rapid Kinetics of Protein-Nucleic Acid Interaction is a Major Component of HIV-1 Nucleocapsid Protein's Nucleic Acid Chaperone Function. *J. Mol. Biol.*, **363**, 867–877.
31. Levin,J.G., Guo,J., Rouzina,I. and Musier-Forsyth,K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism. *Prog. Nucleic Acid Res. Mol. Biol.*, **80**, 217–286.
32. Li,P.W.-L., Li,J., Timmerman,S.L., Krushel,L.A. and Martin,S.L. (2006) The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition. *Nucleic Acids Res.*, **34**, 853–864.