

# **The L1-ORF1p coiled coil enables formation of a tightly compacted nucleic acid-bound complex that is associated with retrotransposition**

Ben A. Cashen<sup>1</sup>, M. Nabuan Naufer<sup>1</sup>, Michael Morse<sup>1</sup>, Charles E. Jones<sup>2</sup>, Mark C. Williams<sup>1†</sup>, Anthony V. Furano<sup>2</sup>

<sup>1</sup>Northeastern University, Department of Physics, Boston, MA 02115, USA

<sup>2</sup>The Laboratory of Molecular and Cellular Biology, NIDDK, NIH, Bethesda, MD, 20892, USA

<sup>†</sup>Corresponding author: ma.williams@northeastern.edu

## **Abstract**

Long interspersed nuclear element 1 (L1) parasitized most vertebrates and constitutes ~20% of the human genome. It encodes ORF1p and ORF2p which form an L1-ribonucleoprotein (RNP) with their encoding transcript that is copied into genomic DNA (retrotransposition). ORF1p binds single-stranded nucleic acid (ssNA) and exhibits NA chaperone activity. All vertebrate ORF1ps contain a coiled coil (CC) domain and we previously showed that a CC-retrotransposition null mutant prevented formation of stably bound ORF1p complexes on ssNA. Here we compared CC variants using our recently improved method that measures ORF1p binding to ssDNA at different forces. Bound proteins decrease ssDNA contour length and at low force, retrotransposition-competent ORF1ps (111p and m14p) exhibit two shortening phases: the first is rapid, coincident with ORF1p binding; the second is slower, consistent with formation of tightly compacted complexes by NA-bound ORF1p. In contrast, two retrotransposition-null CC variants (151p and m15p) did not attain the second tightly compacted state. The C-terminal half of the ORF1p trimer (not the CC) contains the residues that mediate NA-binding. Our demonstrating that the CC governs the ability of NA-bound retrotransposition-competent trimers to form tightly compacted complexes reveals the biochemical phenotype of these coiled coil mutants.

## **Introduction**

LINE-1 (L1) is a non-LTR intragenomic DNA parasite that has been evolving in mammalian genomes for ~100 Myr. It is the only autonomously active mobile genetic element in humans and constitutes approximately 20% of human DNA (1-4). L1 can also mobilize non-autonomous transposable elements, such as Alu and SVA, and as a result L1 activity has generated upwards of ~40% of the mass of many mammalian genomes (5-9). Despite their deleterious effects, L1 sequences remain active in most modern mammals, including humans, contributing to genetic diversity, and causing genetic defects and rearrangements. Additionally, L1 is subject to strong negative selection (10) and is a target of numerous host repressive mechanisms arrayed against other foreign genomic elements (11), indicating that it generally provides little benefit to its

host. The persistence of L1 activity and its evolutionary history in mammals can, in part, be understood as an ongoing arms race (12).

A full-length human L1 (~6 knt) contains a regulatory 5' untranslated region (UTR), two open reading frames (ORFs) that encode proteins required for retrotransposition (ORF1p, ORF2p) (13,14), and a 3' UTR which contains a highly conserved G-rich quadruplex-forming motif that stimulates retrotransposition (15,16). ORF1p and ORF2p bind their encoding transcript (*cis* preference) to form the L1 RNP, which mediates retrotransposition (17-20). ORF2p functions as the L1 replicase. It contains highly conserved endonuclease and reverse transcriptase domains that respectively nick host DNA, allowing a flap of the nicked strand to hybridize with the A-rich 3' end of the L1 transcript, and prime its reverse transcription to generate a DNA copy that is subsequently inserted into the genome – referred to as target site primed reverse transcription (TPRT, Fig. 1B ) (14,21-23).

ORF1p, the major component of the L1 RNP, binds NA non-specifically with high affinity and functions as a NA chaperone, *i.e.*, facilitates annealing and exchange of NA strands. It contains a 51 amino acid intrinsically disordered N-terminal domain (NTD), which harbors two highly conserved phosphorylation sites necessary for retrotransposition (24,25), followed by a 14-heptad coiled coil (26-30), which mediates trimerization of ORF1p monomers (Fig 1A). A coiled coil domain is an unusual feature of an NA chaperone and while present in ORF1 of all vertebrate L1 elements (29), and in most mammals, including humans, it is nonetheless evolutionarily labile – subject to episodic sequence changes (12). In addition, mutational analysis has shown that ORF1p activity can be quite sensitive to coiled coil substitutions (31-33).

The evolutionary lability of the coiled coil contrasts the highly conserved carboxy-terminal half of the molecule, which consists of two domains: a non-canonical RNA recognition motif (RRM)(34) that also contains two essential highly conserved phosphorylation sites, separated by an intervening intrinsically disordered loop (24,25). The RRM is followed by the carboxy-terminal domain (CTD) and the protein terminates in a 46 amino acid intrinsically disordered sequence. Several residues (*e.g.*, R261 and R262) in the carboxy terminal half have been mapped to high affinity NA binding and chaperone activity *in vitro*. These activities are only evident in the context of the trimer and their mechanistic relationship to retrotransposition is not known (13,14,18,26-28,34-40).

Studies on the interaction between purified ORF1p and NA have revealed functionalities of the protein that partially recapitulate features expected of the L1 RNP, related to both its formation and its function in TPRT (28,32,41,42). For example, as illustrated in Fig. 1B, generation of a productive cDNA primer for TPRT requires formation of a stable hybrid between the DNA flap and 3' end of the L1 transcript, and ORF1p can stabilize mismatched oligonucleotide duplexes (32). Mismatches between the target site DNA primer and the A-rich 3' terminus of the L1 transcript are likely given the degenerate nature of both the target site sequence and the 3' terminus of the A-rich L1 transcript. In addition, the protein binds mismatched duplexes with

the same affinity as it does single stranded oligonucleotides, which is 10-fold higher than to perfectly matched double stranded duplexes (28,32). These studies also showed that in addition to mediating trimerization, the coiled coil ensures the trimer-trimer interactions between nucleic acid-bound ORF1ps that support retrotransposition (Fig. 1C).

Protein cross linking studies revealed that trimer-trimer interactions between NA-bound trimers are mediated by residues in the C-terminal half of the protein (28) (also see Discussion and (30) with respect to the involvement of the coiled coil). Callahan et al also showed that trimer-trimer interactions enhance NA binding to oligonucleotides as those long enough to accommodate 2-3 trimers out compete binding by oligonucleotides that can accommodate only a single trimer (28). We had extended these observations using single molecule studies with force-melted ~50 kb  $\lambda$  phage DNA as a source of ssNA. After initial binding, ORF1p oligomerized to a far more stably bound form (32). Reducing the force on the unwound DNA allowed double helix formation, which displaced the oligomerized ORF1p, recapitulating the dissolution of the L1 RNP during TPRT (Fig. 1B). Most importantly, these studies revealed the basis of the defect in a coiled coil mutant, 151p (insert Fig. 1A), which is inactive in retrotransposition (retro<sup>-</sup>) but biochemically the same as its retrotransposition competent (retro<sup>+</sup>) counterpart ORF1p (111p) for oligonucleotide-based NA binding and chaperone activity (32). The 151p protein is unable to oligomerize to a stably bound form after it initially binds to ssNA.

Our previous single molecule method for measuring ORF1p-ssDNA interactions required force-melting of dsDNA to produce local regions of effectively single-stranded DNA. Additionally, because the degree of protein binding was measured by stretching the DNA substrate periodically at discrete times, we were unable to measure the dynamics of the protein-DNA complex continuously. Here we advanced our analyses in two ways: First, we employed a recently improved single molecule method that generates ssDNA *in situ* prior to protein binding (43-45). As such, we were not limited to forces  $\geq$  60 pN that are required to unwind dsDNA, allowing us to measure ORF1p-ssNA interactions at various (lower) tensions and in real time. Second, we compared the previously studied 111p/151p pair to a new pair of retro<sup>+</sup>/retro<sup>-</sup> coiled coil variants, m14p/m15p, which differ by a single CC residue (insert Fig. 1A). These proteins are also indistinguishable in a FRET-based NA chaperone assay (Fig. 1S). Unlike its active m14p counterpart, m15p, cannot form a stable, tightly compacted NA-ORF1p structure *in vitro* (Fig. 1C), providing another example of the biochemical phenotype of a retrotransposition-null CC mutant. Yet, seemingly paradoxically, the CC has undergone repeated evolutionary change. However, recent analysis suggested that such evolutionary lability can protect CC function from disabling mutations (12).

## Materials and methods

### Purification of ORF1p

N-terminal his-tagged ORF1ps were expressed in insect cells and purified as previously described (28). This procedure produces highly purified nuclease-free phosphorylated proteins

that exhibit high affinity NA binding and NA chaperone activity as determined by a FRET-based NA-annealing and strand exchange assay (24,28,32) and Figures S1 and S3. The ORF1 sequence was derived from L1.3 a member of the L1Pa1 (L1Hs) family (NCBI L19088).

### **Optical tweezers system for measuring ssDNA conformation**

An 8.1 knt ssDNA molecule tethered between two functionalized microspheres (anti-DIG and Streptavidin, Fig. 1C) was generated *in situ* by T7 exonuclease digestion as described previously (43-45) and held at various fixed tensions. Extension of the ssDNA was continuously altered to maintain the given force applied by the trapping laser in a binding buffer containing 50 mM Na<sup>+</sup>, 10 mM HEPES at pH 7.5. Although ORF1p trimers in the absence of NA form precipitable aggregates in 50 mM NaCl (optimal for NA binding), this does not preclude their binding to NA, or their ability to freely exchange with NA-bound trimers. However, NA-free ORF1p trimers are soluble in 0.5 M NaCl containing storage buffer (28). Therefore, concentrations of ORF1p in storage buffer were appropriately diluted into binding buffer to produce a trimer concentration of 30 nM immediately before adding to DNA. Following incubation for the indicated times, we measured the dissociation of ORF1p by replacing the protein-containing buffer with protein-free buffer. The extension of the ssDNA was controlled by a piezoelectric translational stage with 1 nm resolution, and the tension along the substrate was measured by laser deflection of the stationary optical trap (Fig. 1C). Additionally, distance between the microspheres was measured using simultaneously recorded bright-field images to calculate the absolute ssDNA extension and correct for long-term thermal drift in the system. All data were analyzed using custom scripts in MATLAB (Mathworks) with uncertainty calculated as standard error of the mean of three or more replicates.

### **ORF1p compaction at constant extension**

The ssDNA was first incubated and fully saturated with 30 nM ORF1p trimer at a tension of 30 pN. Upon reaching equilibrium, protein-containing buffer was replaced with protein-free buffer. The extension of the protein-DNA complex was subsequently lowered to, and held at ~0.2 nm/nt for 2, 5, 15 or 30 min. At this fixed extension, the force on the DNA varied between ~2-10 pN depending on the incubation time (*i.e.*, the tension increased with incubation time). The complex was then stretched at a rate of ~450 nm/s until reaching 75 pN, where it was held and monitored for 100 s while it elongated. Finally, the tension was released by reducing its extension to the initial value. The stretch and release extension values at 30 pN were normalized with respect to the extension of a bare ssDNA molecule. Uncertainties were determined by the standard error of the mean of three or more replicate curves.

## **Results**

### **Binding of ORF1p to ssDNA at low force**

The force (tension) applied to ssDNA can affect such features as its structure (*e.g.*, base stacking) and its ability to engage (conform to) protein NA-binding sites. Therefore, we

determined binding of 30 nM ORF1p trimer to ssDNA at both low force (5 pN) where the ssDNA adopts a winding/flexible conformation, and high force (30 pN) where the ssDNA is effectively straightened. Fig. 1C shows a schematic of the binding of ORF1p to an 8.1 knt ssDNA molecule. The data collected at 5 pN is shown in Fig. 2 using the ORF1p variants listed in Fig. 1A (insert – the reference sequence is L1Pa1-ORF1p, designated 111p in the text, also see Materials and Methods). Upon binding, all the variants produce an initial, rapid ( $\tau \sim 1$  s) contraction of the DNA,  $\Delta x_+^i$ , but retro<sup>+</sup> ORF1p variants (111p, m14p) then undergo a slower ( $\tau \sim 100$  s), secondary compaction,  $\Delta x_+^s$ , which is minimal for the retro<sup>-</sup> 151p and m15p proteins. We had previously shown that unlike 111p, 151p trimers cannot form stably bound, compact oligomers on ssDNA (32). Therefore, we hypothesized that m15p, which differs from m14p by a single CC mutation (R105T), and which also produces only minor secondary compaction (Fig. 2A-C), is unable to form tightly compacted structures on ssDNA. Fig. 1C shows a schematic interpretation of these data.

It is important to note that unlike 151p, which contains 4 ancestral (L1Pa5) residues in heptads 8 and 9, m15p contains only 1, R105T. Due to strong epistatic (*i.e.*, context dependent) effects of coiled coil mutations (12) it is not possible to predict or extrapolate the effect of a given coiled coil substitution on ORF1p activity to another context. Thus, m15p cannot be considered a subset of m151p but rather a distinct non-overlapping mutational state. On the other hand, the ORF1p coiled coil can be indifferent to multiple substitutions (12), *e.g.*, the ancestral L1pa5 residues, VIQEY in heptads 10 - 12, do not affect ORF1p activity.

Protein binding and subsequent oligomerization on ssDNA can be modeled by the following reaction:



where  $\Theta_0$ ,  $\Theta_b$ , and  $\Theta_{oligo}$  are the ssDNA fractions of protein in the unbound, bound (but not oligomerized), and oligomerized (tightly compacted) states, respectively.  $ck_b$ ,  $k_b$ ,  $k_{oligo}$ , and  $k_{-oligo}$  represent the characteristic transition rates between each state. In general, the resulting differential equations are analytically intractable. However, in the case of ORF1p, the following conditions allow for an approximate analytical solution to the system of equations

$$ck_b \gg k_b \quad (2)$$

$$ck_b \gg k_{oligo} + k_{-oligo} \quad (3)$$

Our data supports both conditions, as ORF1p shows high binding affinity and dissociates from the NA substrate slowly (2) and reorganization of bound protein occurs over a longer timescale than its initial binding (3). These conditions indicate that the bound state,  $\Theta_b$ , effectively reaches full occupancy, and that bound ORF1p saturates the substrate on a much shorter

timescale than subsequent oligomerization, allowing us to decouple the transitions between each state (*i.e.*, the transitions occur sequentially). This gives the following solution:

$$\Delta x(t) = \Delta x_+^i(1 - e^{-k_+^i t}) + \Delta x_+^s(1 - e^{-k_+^s t}) \quad (4)$$

where  $\Delta x_+^i$  and  $k_+^i = ck_b$  are the respective magnitude and rate of initial compaction due to ORF1p binding.  $\Delta x_+^s = \alpha(n - 1)k_{\text{oligo}}/(k_{\text{oligo}} + k_{-\text{oligo}})$  where  $\alpha$  is proportional to the magnitude of compaction per protein oligomerization event,  $n$  represents the number of proteins bound to the ssDNA, and  $k_+^s = k_{\text{oligo}} + k_{-\text{oligo}}$ .

The rates ( $k_+^i$ ) and amplitudes ( $\Delta x_+^i$ ) of initial compaction are equivalent for the four trimers, indicating that they initially bind the substrate identically at 5 pN (Table 1S). However, the magnitude of the oligomerization-driven, secondary compaction,  $\Delta x_+^s$ , is significantly reduced for the inactive variants (Fig. 2A-C). As we have no information on the oligomerization on and off rates,  $k_{\text{oligo}}$  and  $k_{-\text{oligo}}$ , we can only measure the sum of the two rates,  $k_+^s$  (which is similar for 111p, m14p, and m15p, Fig. 2D). Possible explanations for the observed differences in compaction include: (1) The degree of compaction from each oligomerization event,  $\alpha$ , differs between the active and inactive proteins but the rates,  $k_{\text{oligo}}$  and  $k_{-\text{oligo}}$ , are unchanged. (2)  $\alpha$  is unchanged, but  $k_{\text{oligo}}$  and  $k_{-\text{oligo}}$  change such that their sum remains the same, and transition into the oligomeric state,  $\Theta_{\text{oligo}}$ , is inhibited (*i.e.*, the fraction of protein in the oligomeric state at equilibrium,  $k_{\text{oligo}}/(k_{\text{oligo}} + k_{-\text{oligo}})$ , decreases). (3) These quantities,  $\alpha$ ,  $k_{\text{oligo}}$ , and  $k_{-\text{oligo}}$  are all changed. While the cause of the reduction in ORF1p-ORF1p compaction is not known, its ability to do so and thereby support retrotransposition is exquisitely sensitive to the CC sequence (12).

### Assessing inter-trimer interactions that develop at low force

We examined the properties of ORF1p-ssDNA complexes that form at low force by subjecting them to cycles of extension and release after first incubating them under protein-free conditions for 2, 5, 15 and 30 min at a low fixed extension ( $\sim 0.2$  nm/nt, Fig. 3, see Materials and Methods). After two minutes, the resulting stretch curve resembles that of a polymer with significantly reduced contour length relative to bare ssDNA (Fig. 3A, B). The average slope of the stretch curve of the protein-DNA complex increases with the duration of initial low force incubation, consistent with their reduced elasticity, likely an effect of increasing compaction due to interactions between NA-bound trimers. Additionally, following 100 s at 75 pN the release curves reveal shortening of the ORF1p-ssDNA complex concomitant with the time of the initial low force incubation. This indicates that ORF1p-ssDNA complexes that form at low force can convert to higher order compacted structures in the absence of free protein. Moreover, the shift in the extension of the release curves indicate that the compacted structures become increasingly stable with time, resisting disruption at very high force. At 75 pN the extension-time profile of the complex (Fig. 2S) shows a series of gradual increases in extension, presumably reflecting dissolution of the higher order compacted protein-DNA structures. The increases in extension show a high degree of variability, ranging from tens to several hundreds

of nanometers, indicating that the compact structures formed at low force, although quasi-stable, may be quite large and nonuniform in size.

Fig. 3C-D show that the ORF1p-ssDNA complexes formed by retro<sup>+</sup> 111p and m14p undergo significantly greater (and more stable) compaction than the retro<sup>-</sup> 151p and m15p proteins. These results recapitulate the findings on the complexes that form at 5 pN (Fig. 2) and together corroborate our earlier studies (32) that the ability of NA-bound ORF1p trimers to form stable higher order complexes is positively correlated with retrotransposition activity, and as we showed here both properties can be abolished by the single R105T CC substitution.

### Binding of ORF1p to ssDNA at high force

The binding dynamics of the ORF1 proteins at 30 pN (Fig. 4), at which the substrate tension is high enough to disfavor compaction, are dramatically different from what occurs at 5 pN (Fig. 2). Regardless of their ability to support retrotransposition, the proteins induce the same initial, rapid ( $\tau \sim 1$  s) ssDNA compaction ( $\Delta x_+^i$ ) followed by a slower ( $\tau \sim 10$  s) partial elongation, equilibrating to the same final extension ( $\Delta x_+^f$ ) less than that of bare ssDNA (Fig 4). Similar to the behavior at low force (Fig. 2), this biphasic binding signature indicates a change in the conformation of the protein-DNA complex over time. However, in contrast to what is seen at 5 pN, the oligomerization-deficient 151p binds identically to wild type, suggesting that these changes in ssDNA extension are not primarily driven by trimer-trimer interactions between NA-bound ORF1p.

Although the variants show similar biphasic binding at high force, when free unbound ORF1p is removed from the channel (Fig. 5), differences between the retro<sup>+</sup> and retro<sup>-</sup> protein complexes are readily apparent: While all exhibit an initial rapid ( $\tau \sim 10$  s) decrease in extension ( $\Delta x_-^i$ ), followed by a slower ( $\tau \sim 100$  s) elongation to a final extension that does not attain that of bare ssDNA ( $\Delta x_-^f$ ), both the rate and magnitude of these extension changes are significantly greater for the retro<sup>-</sup> 151p and m15p proteins. Specifically, re-compaction of the ssDNA during the initial dissociation phase is ~2-fold faster for the inactive trimers. Moreover, the final dissociation of 151p and m15p is both more complete (*i.e.*, the ssDNA approaches its protein-free conformation) and faster (nearly 3-fold) than the active trimers, indicating that they are less stably bound to ssDNA. This would be expected if the trimers are unable to form tightly compacted oligomers on ssDNA, supporting our conclusions from the binding experiments at low force (Fig. 2-3) and our prior studies on 151p and 111p using our previous single molecule method (32).

### Discussion

Here we extended our single molecule studies on ORF1p coiled coil variants that had shown a relation between retrotransposition and the ability of ssNA-bound ORF1p to form stably bound complexes (oligomers) (32). With our recently improved method we re-examined this interaction by probing an additional pair of retro<sup>+</sup> and retro<sup>-</sup> ORF1ps. We determined the

interaction of the ORF1ps as a function of force (tension) applied to the NA, a critical parameter that both governs formation and reveals properties of the ORF1p-NA complex. At 5 pN, retro<sup>+</sup> 111p and m14p formed tightly compacted RNPs with ssDNA that were not attainable by the retro<sup>-</sup> CC mutants 151p and m15p (Fig. 1C, Fig 2-3). These results recapitulate and extend our earlier study (32) as they include analysis of the additional pair of retro<sup>+</sup> m14p and retro<sup>-</sup> m15p CC variants, which differ by a single CC residue (Fig 1A). The four proteins were indistinguishable by an oligonucleotide-based chaperone (FRET) assay (Fig. 1S) as we had shown earlier for retro<sup>+</sup> 111p and retro<sup>-</sup> 151p (32).

At 5 pN, all ORF1ps, regardless of retrotransposition competence, attained the same initial “compacted” state,  $\Delta x_+^i$ , and at essentially the same rate,  $k_+^i$ , (Fig. 2A, Table 1S). Studies (28,32) using oligonucleotide-based binding assays also revealed no difference in binding affinity between 111p and 151p. Thus, the initial compacted state (Fig. 2) might be similar to that attained by ORF1p bound to oligonucleotides (28). NA-bound trimers were close enough ( $\sim 16\text{\AA}$ ) to be cross linked by the bifunctional cross-linking reagent EGS [ethylene glycolbis(succinimidylsuccinate)] and involved inter-trimer contacts located in the carboxy terminal half of ORF1p (28). These findings were consistent with atomic force microscopy images showing NA-bound mouse ORF1p aligned side by side on ssRNA involving contacts between the carboxy terminal half of the trimer (35).

In contrast, subsequent transition to a “tightly compacted” state ( $\Delta x_+^f$ ), was only attained by retro<sup>+</sup> 111p and m14p (Fig. 2, Table 1S). At a tension of 5 pN, the absolute extension of our bare 8.1 knt ssDNA molecule was measured at  $\sim 2.35\text{ }\mu\text{m}$  (normalized to 0.29 nm/nt), consistent with the freely jointed chain polymer model (46). The average total extension changes ( $\Delta x_+^f$ ) seen for the 111p and m14p complexes approached  $\sim 0.29\text{ nm/nt}$ , indicating that over long incubation timescales this tightly compacted state corresponds to a conformation wherein the DNA is almost fully compacted to near zero extension (Fig. 2, purple dashed line). Furthermore, these tightly compacted structures were more resistant to disruption than the complexes formed by retro<sup>-</sup> ORF1ps. Following incubation at a minimal, fixed extension ( $\sim 0.2\text{ nm/nt}$ ), the ORF1p-NA complexes were subjected to high force (75 pN) for 100 seconds (Fig. 3, Table 2S). Upon reaching 75 pN, the ssDNA elongated, likely reflecting re-organization of NA-bound ORF1p complexes. Over time, retrotransposition competent trimers remained tightly compacted but the NA complexes formed by retro<sup>-</sup> proteins reverted to nearly the extension of bare ssDNA. Thus, in addition to tight compaction, retrotransposition is positively correlated with stability of the compacted ORF1p-NA complex.

Although ORF1p binding at 30 pN produces a biphasic change in extension, it differs from what occurs at 5 pN. The length reduction attained in the first phase is about 7-fold less than at 5 pN and, rather than not changing (retro<sup>-</sup>) or undergoing further shortening (retro<sup>+</sup>), the ORF1p-ssNA complex re-elongates almost 2-fold, though to less than that of bare ssDNA, indicating that ORF1p remains bound to the DNA (cf., Tables 1S and 3S and Figs. 2 and 4). Furthermore, in contrast to the 5 pN data, the biphasic length profiles of retro<sup>+</sup> and retro<sup>-</sup> ORF1p are nearly

identical. These results likely reflect the different rigidities of ssNA at 30 and 5 pN, which would alter the binding options available to the protein. Additionally, Fig. 5 shows that the 30 pN ORF1p-ssNA complexes formed by retro<sup>-</sup> proteins are significantly more labile than those assembled by retro<sup>+</sup> ORF1p, losing protein faster and more completely than the retro<sup>+</sup> complexes in protein-free buffer (Fig. 5, Table 4S). This likely results from the weak or even lack of inter-trimer interaction between NA-bound retro<sup>-</sup> ORF1p, because the inherent affinity of retro<sup>+</sup> and retro<sup>-</sup> are the same for NA in both the single molecule assay ( $\Delta x_+^i$ , phase 1, Fig 2, Table S1) and oligonucleotide-based assays (28,32). Therefore, we felt justified in modelling the nature of the 5 pN complexes on our cross-linking evidence which showed trimer-trimer contacts between oligonucleotide-bound ORF1p. However, understanding the nature of the complexes that form at high force will require further quantification under different solution and substrate conditions (e.g., protein concentration, DNA tension, etc.) and, potentially, the development of additional analytical techniques. On the other hand, force-dependent generation of distinct ORF1p-ssNA complexes illustrates the advantages of our current single molecule method to reveal the different possible modes of NA-ORF1p interaction.

Our earlier studies indicated that the carboxy-terminal half of ORF1p can mediate the inter-trimer interactions responsible for oligomerization of NA-bound ORF1p (28,32). Before discussing these results in the current context, we address two topics relevant to trimer-trimer interaction. First, it was recently suggested that the coiled coils of different trimers could mediate their interaction (30). These conclusions were based on the interaction of coiled coils that had been solubilized to monomers in guanidinium HCl from inclusion bodies (insoluble aggregates) that had accumulated during their synthesis in *E. coli*. Fully denatured ORF1p coiled coils such as these, which also lack the NTD, are likely an *in vitro* artifact that would not exist in "nature". It is almost certain that ORF1p monomers trimerize while being synthesized on adjacent ribosomes (47), and we are not aware of any evidence showing that the coiled coils of fully formed trimers synthesized *in vivo* can unravel to mediate interactions between different trimers.

The second topic is ORF1p aggregation. This topic dates to 1996 (48) and was addressed again in 2012 using highly purified ORF1p to determine the effect of salt and nucleic acid on this process (28). These earlier studies employed chemical cross linking and gel electrophoresis and showed that ORF1p aggregates can form in either the presence or absence of NA. While NA-free ORF1p trimers are soluble in 0.5 M NaCl, they form precipitable aggregates in 50 mM NaCl (optimal for NA binding), which does not preclude their binding to NA, or their ability to freely exchange with NA-bound trimers (28,49). ORF1p aggregation has also been addressed in two recent reports respectively by Newton et al. (50) and Sil et al. (51). These studies used microscopically visible phase separated condensates (droplets) as a measure of ORF1p aggregation, droplet formation being likely mediated by the NTD, an intrinsically disordered region (IDR) (52). Both the Newton and Sil studies imply that condensate formation is intrinsic to the function of ORF1p in retrotransposition. However, there is no evidence for this assumption and the functional relevance of condensate formation by IDR-containing proteins in general has been questioned by Martin et al. (52), who stated "*The notion that the presence of an IDR means a protein has evolved to phase separate is an inaccurate inference that has*

*unfortunately been used to justify questionable lines of inquiry and questionable experimental design*”. This admonition is exemplified by Newton et al. who showed that phase separated condensates require just the N-terminal 152 amino acids (NTD + coiled coil). As this region of ORF1p does not contain the highly conserved residues in the RRM and CTD shown by mutational analysis (18,34,39) to be involved in NA binding and RNP formation, condensate formation is indifferent to these RNA binding domains that are critical to the role of ORF1p in retrotransposition.

How the CC ensures the formation of tightly compacted NA-bound ORF1p, and why it is required for retrotransposition remain open questions. As to the former, this role of the CC is highly sensitive to its sequence. Retro<sup>+</sup> m14p and retro<sup>-</sup> m15p differ by a single CC residue and we have identified a number of single CC substitutions that just as dramatically affect ORF1p activity – either abolishing it or fully restoring it (12). The phenotypic effects of CC mutations often depend on their sequence context. These are termed epistatic, and evolutionary responses to buffer negative epistatic mutations have at times governed CC evolution (12). The CC could govern the relative orientation of the carboxy-terminal half of the monomers that constitute the trimer (28). Atomic force microscopy of the mouse trimer and X-ray crystallography of the human trimer showed that this region assumes a propeller-like structure (27,35). Therefore, it is conceivable that a torque-altering coiled coil substitution could be transmitted through its length (53) and reduce the efficiency of the trimer-trimer contacts that mediate their ability to form tightly compacted NA-bound structures associated with retrotransposition.

Several possibilities could account for the correlation between retrotransposition and tightly compacted trimers. If this phenomenon also applies to the L1RNP, then in addition to protecting the parent transcript from nucleases and APOBEC3 deaminases, tight side-by-side packing of trimers could prevent formation of RNA secondary structures (35). Due to hydrogen bonds contributed by the ribose 2'OH, uncoated RNA is prone to form stable, even if short ranged, secondary structures (54,55) that could seriously impede reverse transcription. Although such possibilities seem reasonable, until we understand “tightly packed” in structural terms, they remain speculative.

## **Data Availability**

The experimental data sets are either included in the main text, supplementary material, or are available from the authors upon request.

## **Funding**

This work was supported by the National Science Foundation [MCB-1817712 to MCW]; and the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases [ZIA DK057601-25 to AVF].

## **Acknowledgements**

We thank Ioulia Rouzina for helpful discussions regarding data interpretation.

## Author contributions

BAC, MNN, MM, CEJ, MCW, and AVF designed the research and interpreted the data. MNN, MM, CEJ, and AVF contributed new reagents/analytic tools. BAC and CEJ performed research and analyzed data. BAC and AVF wrote the paper.

## References

1. Furano, A.V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog Nucleic Acid Res Mol Biol*, **64**, 255-294.
2. Goodier, J.L., Cheung, L.E. and Kazazian, H.H., Jr. (2013) Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res*, **41**, 7401-7419.
3. Kazazian Jr, H.H. and Moran, J.V. (2017) Mobile DNA in health and disease. *New England Journal of Medicine*, **377**, 361-370.
4. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.
5. Esnault, C., Maestre, J. and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*, **24**, 363-367.
6. Ostertag, E.M., Goodier, J.L., Zhang, Y. and Kazazian, H.H. (2003) SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.*, **73**, 1444-1451.
7. Wallace, N., Wagstaff, B.J., Deininger, P.L. and Roy-Engel, A.M. (2008) LINE-1 ORF1 protein enhances Alu SINE retrotransposition. *Gene*, **419**, 1-6.
8. Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A. and Batzer, M.A. (2005) SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.*, **354**, 994-1007.
9. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D. and Moran, J.V. (2001) Human L1 retrotransposition: cispreference versus trans complementation. *Molecular and cellular biology*, **21**, 1429-1439.
10. Boissinot, S., Davis, J., Entezam, A., Petrov, D. and Furano, A.V. (2006) Fitness cost of LINE-1 (L1) activity in humans. *P Natl Acad Sci USA*, **103**, 9590-9594.
11. Goodier, J.L. (2016) Restricting retrotransposons: a review. *Mob DNA*, **7**, 16.
12. Furano, A.V., Jones, C.E., Periwal, V., Callahan, K.E., Walser, J.C. and Cook, P.R. (2020) Cryptic genetic variation enhances primate L1 retrotransposon survival by enlarging the functional coiled coil sequence space of ORF1p. *PLoS Genet*, **16**, e1008991.
13. Martin, S.L. (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biol*, **7**, 706-711.
14. 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.
15. Howell, R. and Usdin, K. (1997) The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons. *Mol Biol Evol*, **14**, 144-155.
16. Sahakyan, A.B., Murat, P., Mayer, C. and Balasubramanian, S. (2017) G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nat Struct Mol Biol*, **24**, 243-247.

17. Doucet, A.J., Hulme, A.E., Sahinovic, E., Kulpa, D.A., Moldovan, J.B., Kopera, H.C., Athanikar, J.N., Hasnaoui, M., Bucheton, A., Moran, J.V. *et al.* (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet*, **6**, e1001150.
18. 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.
19. Kulpa, D.A. and Moran, J.V. (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol*, **13**, 655-660.
20. Martin, S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol*, **11**, 4804-4807.
21. Feng, Q., Moran, J.V., Kazazian, H.H., Jr. and Boeke, J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905-916.
22. 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 Chromosomal Target Site - a Mechanism for Non-Ltr Retrotransposition. *Cell*, **72**, 595-605.
23. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D. and Gabriel, A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808-1810.
24. Cook, P.R., Jones, C.E. and Furano, A.V. (2015) Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc Natl Acad Sci U S A*, **112**, 4298-4303.
25. Furano, A.V. and Cook, P.R. (2016) The challenge of ORF1p phosphorylation: Effects on L1 activity and its host. *Mob Genet Elements*, **6**, e1119927.
26. Martin, S.L., Branciforte, D., Keller, D. and Bain, D.L. (2003) Trimeric structure for an essential protein in L1 retrotransposition. *P Natl Acad Sci USA*, **100**, 13815-13820.
27. Khazina, E., Truffault, V., Buttner, R., Schmidt, S., Coles, M. and Weichenrieder, O. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat Struct Mol Biol*, **18**, 1006-1014.
28. Callahan, K.E., Hickman, A.B., Jones, C.E., Ghirlando, R. and Furano, A.V. (2012) Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. *Nucleic Acids Res*, **40**, 813-827.
29. Boissinot, S. and Sookdeo, A. (2016) The Evolution of LINE-1 in Vertebrates. *Genome Biol. Evol.*, **8**, 3485-3507.
30. Khazina, E. and Weichenrieder, O. (2018) Human LINE-1 retrotransposition requires a metastable coiled coil and a positively charged N-terminus in L1ORF1p. *Elife*, **7**.
31. Goodier, J.L., Zhang, L., Vetter, M.R. and Kazazian, H.H., Jr. (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.*, **27**, 6469-6483.
32. Naufer, M.N., Callahan, K.E., Cook, P.R., Perez-Gonzalez, C.E., Williams, M.C. and Furano, A.V. (2016) L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. *Nucleic Acids Res*, **44**, 281-293.
33. Adney, E.M., Ochmann, M.T., Sil, S., Truong, D.M., Mita, P., Wang, X., Kahler, D.J., Fenyö, D., Holt, L.J. and Boeke, J.D. (2019) Comprehensive Scanning Mutagenesis of Human Retrotransposon LINE-1 Identifies Motifs Essential for Function. *Genetics*, **213**, 1401-1414.
34. Khazina, E. and Weichenrieder, O. (2009) Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *P Natl Acad Sci USA*, **106**, 731-736.
35. Basame, S., Wai-lun Li, P., Howard, G., Branciforte, D., Keller, D. and Martin, S.L. (2006) Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *J. Mol. Biol.*, **357**, 351-357.
36. Januszyk, K., Li, P.W.L., 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.

37. 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.

38. Martin, S.L., Bushman, D., Wang, F., Li, P.W., Walker, A., Cummiskey, J., Branciforte, D. and Williams, M.C. (2008) A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. *Nucleic Acids Res.*, **36**, 5845-5854.

39. Martin, S.L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., 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.

40. 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.

41. Naufer, M.N., Furano, A.V. and Williams, M.C. (2019), *Seminars in cell & developmental biology*. Elsevier, Vol. 86, pp. 140-149.

42. Naufer, M.N. and Williams, M.C. (2020), *RNA Chaperones*. Springer, pp. 283-297.

43. Morse, M., Naufer, M.N., Feng, Y., Chelico, L., Rouzina, I. and Williams, M.C. (2019) HIV restriction factor APOBEC3G binds in multiple steps and conformations to search and deaminate single-stranded DNA. *Elife*, **8**, e52649.

44. Naufer, M.N., Morse, M., Moller, G.B., McIsaac, J., Rouzina, I., Beuning, P.J. and Williams, M.C. (2021) Multiprotein E. coli SSB-ssDNA complex shows both stable binding and rapid dissociation due to interprotein interactions. *Nucleic Acids Res.*, **49**, 1532-1549.

45. Naufer, M.N., Murison, D.A., Rouzina, I., Beuning, P.J. and Williams, M.C. (2017) Single-molecule mechanochemical characterization of E. coli pol III core catalytic activity. *Protein Sci.*, **26**, 1413-1426.

46. Smith, S.B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, **271**, 795-799.

47. Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., Auburger, J.J., Tans, S., Bukau, B. and Kramer, G. (2021) Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science*, **371**, 57-64.

48. Hohjoh, H. and Singer, M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630-639.

49. Callahan, K.E. (2012), Georgetown Washington DC.

50. Newton, J.C., Naik, M.T., Li, G.Y., Murphy, E.L., Fawzi, N.L., Sedivy, J.M. and Jogl, G. (2021) Phase separation of the LINE-1 ORF1 protein is mediated by the N-terminus and coiled-coil domain. *Biophys. J.*, **120**, 2181-2191.

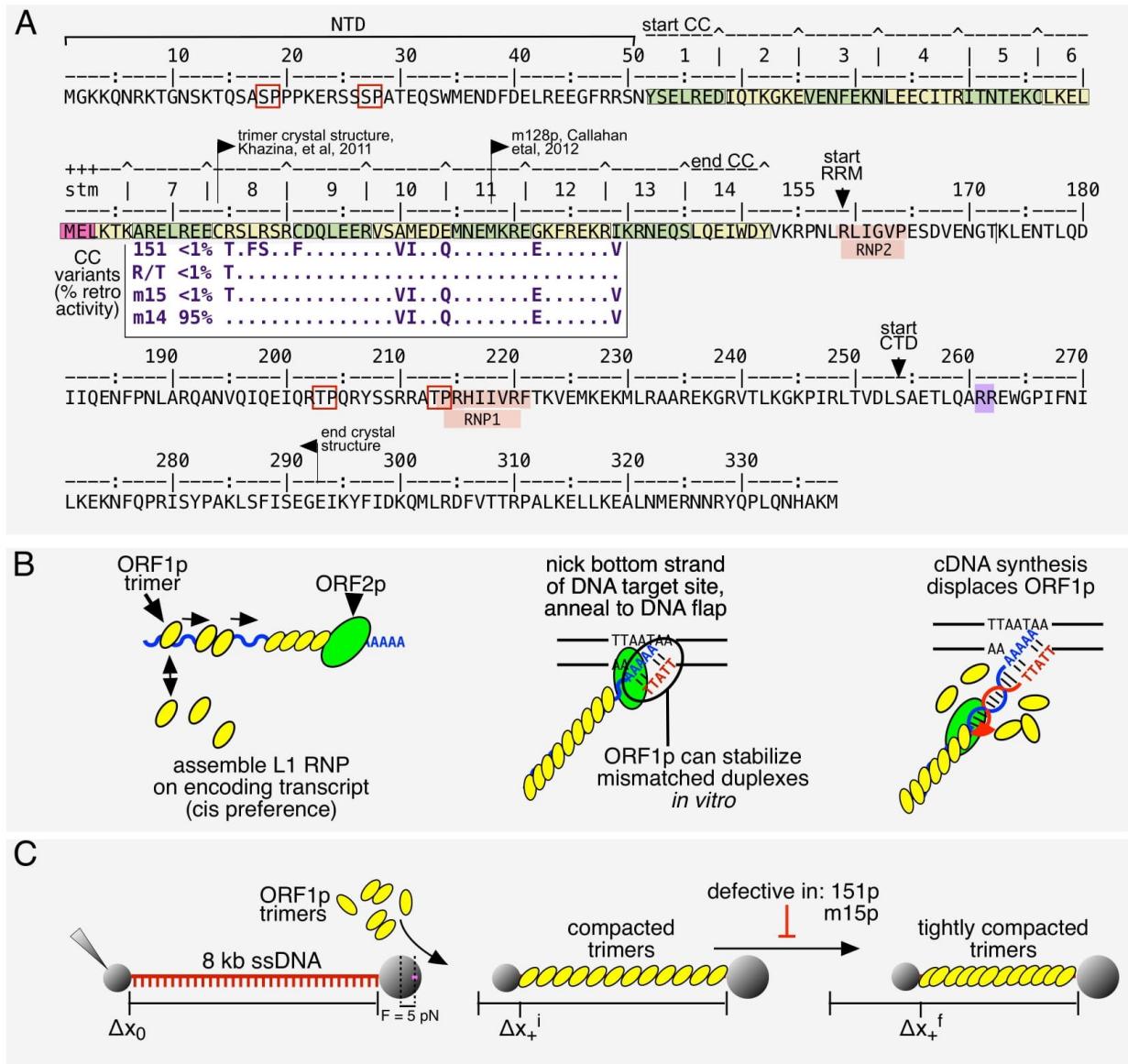
51. Sil, S., Boeke, J.D. and Holt, L.J. (2022) Condensation of LINE-1 is required for retrotransposition. *bioRxiv*, 2022.2004.2011.487880.

52. Martin, E.W. and Holehouse, A.S. (2020) Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerging Topics in Life Sciences*, **4**, 307-329.

53. Truebestein, L. and Leonard, T.A. (2016) Coiled-coils: The long and short of it. *BioEssays : news and reviews in molecular, cellular and developmental biology*, **38**, 903-916.

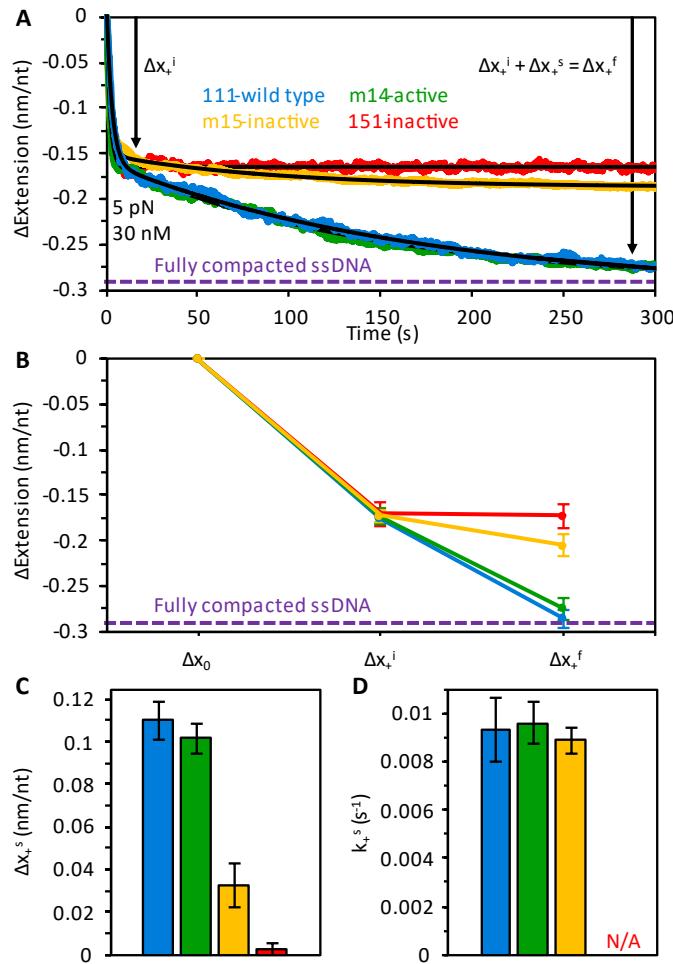
54. Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111-113.

55. Doudna, J.A. and Cate, J.H. (1997) RNA structure: crystal clear? *Curr. Opin. Struct. Biol.*, **7**, 310-316.



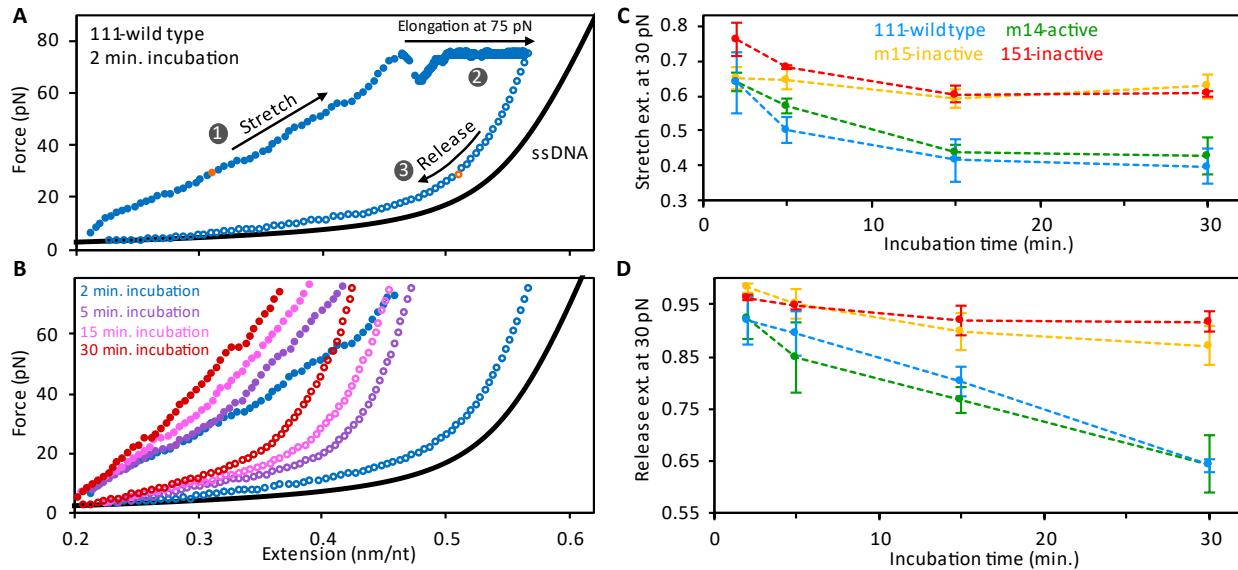
**Figure 1: ORF1p**

(A) Annotated sequence of ORF1p showing conserved phosphorylation sites (red boxes), the 14 heptads of the CC (alternating green and yellow boxes with a stammer (stm) in heptad 6), the highly conserved non-canonical RNA recognition motif (RRM), and C-terminal domain (CTD) that contains sequences (notably R261, R262) involved in NA-binding and chaperone activity. The N terminal domain (NTD) and terminal 46 amino acids of the CTD are intrinsically disordered (see text). The insert shows the relevant part of the alignment of the CC variants and their % retro(transposition) activity relative to the 111 (L1Pa1) wild type protein (adapted from Fig 1 in ref. 12). The amino acids that differentiate the coiled coil variants from 111p are their ancestral counterparts in the resuscitated L1Pa5 family (32). (B) Depiction of L1RNP assembly, involvement in, and fate during retrotransposition. (C) Depiction of an ssNA tethered between two beads and its length  $\Delta x_0$ , before and after its initial  $\Delta x_+^i$  and final  $\Delta x_+^f$  compaction.



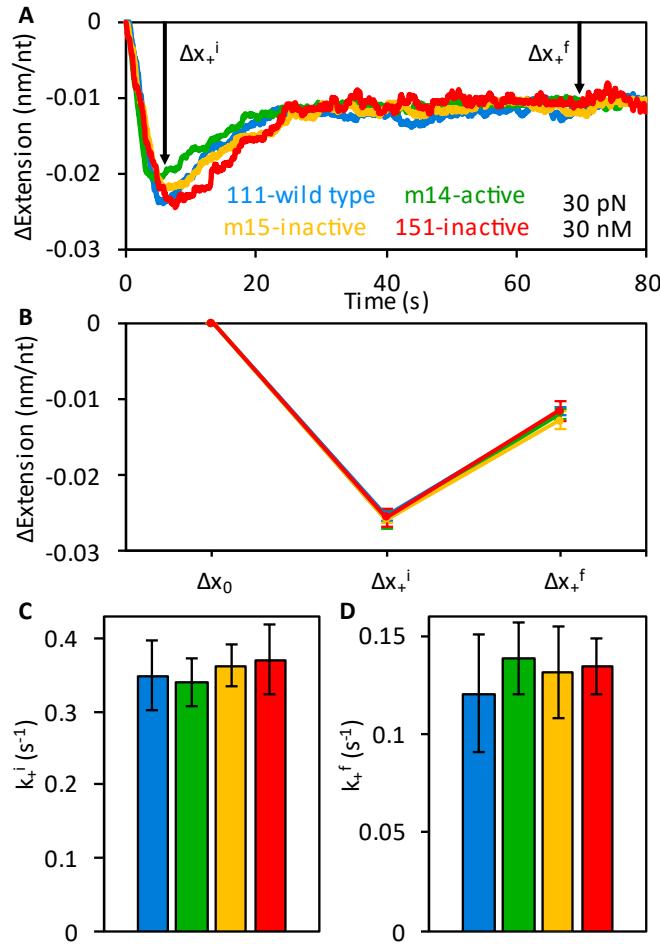
**Figure 2: Binding of wild type and ORF1p variants to ssDNA at 5 pN**

(A) When ssDNA is incubated with ORF1p at low force (5 pN), two phases of ssDNA compaction are observed: initial, rapid compaction ( $\Delta x_+^i$ ), followed by a slow, secondary compaction step ( $\Delta x_+^s$ ). The curves were fit with a two-rate decaying exponential function to extract a rate and amplitude associated with both phases of ssDNA compaction. The absolute extension of bare ssDNA at 5 pN is  $\sim 0.29$  nm/nt. The total ssDNA extension changes seen for 111p and m14 asymptote to  $\sim -0.29$  nm/nt, indicating that the DNA is almost fully compacted (purple dashed line) to near zero extension. (B) The four proteins exhibit similar initial compaction, but the magnitude of the secondary compaction is significantly reduced for complexes formed with the inactive variants. (C) The amplitudes of the secondary compaction events ( $\Delta x_+^s$ ) are plotted as bar graphs for comparison. (D) The rates of secondary compaction ( $k_+^s$ ) are similar for 111p, m14, and m15, however, we were unable to calculate  $k_+^s$  for 151p as secondary compaction was negligible.



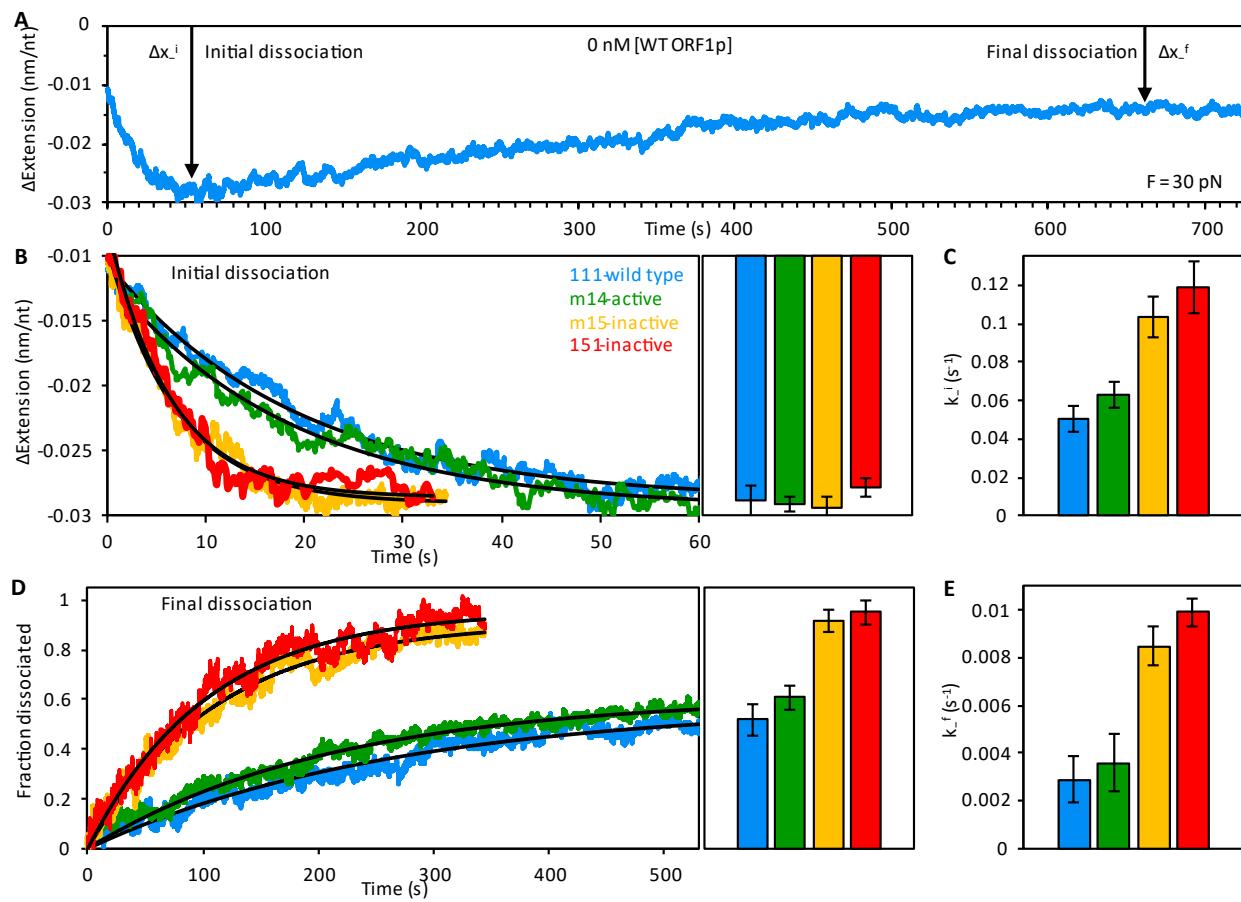
**Figure 3: Dynamics of ORF1p-ssDNA compaction at low, fixed extension**

(A) ssDNA previously incubated for 100 seconds at 30 pN with 30 nM 111p was held at a minimal, fixed extension ( $\sim 0.2$  nm/nt) in protein-free buffer for 2 minutes. The protein-ssDNA complex was then stretched until reaching a tension of 75 pN (1), where it was held for 100 seconds (2). Tension on the strand was then released by reducing its extension to the initial value (3). (B) Incubation of the complexes formed by 111p in protein-free buffer at low extension (0.2 nm/nt) was repeated for 5, 15, or 30 minutes. The ssDNA extension for both the initial stretch (closed circles) and subsequent release (open circles) is inversely proportional to the initial incubation time. (C) Extension of the pre-formed ssDNA-ORF1p complex at 30 pN during stretching (normalized to the length of protein-free ssDNA) shows that the compaction of the active protein-DNA complexes (111p and m14) is greater than that of the inactive proteins (151p and m15). (D) Similarly, during release, the reduction of extension at 30 pN is greater for the active proteins than the inactive variants.



**Figure 4: Initial binding phases of active (111p, m14) and inactive (151p, m15) trimers at 30 pN**

(A, B) The extension changes,  $\Delta x_+^i$  and  $\Delta x_+^f$ , of ssDNA (absolute extension of bare ssDNA at 30 pN is  $\sim 0.54$  nm/nt) during incubation with wild type ORF1p and the three coiled coil variants (m14, m15, and 151p) show identical biphasic binding behavior, indicating that the proteins initially bind ssDNA in a similar manner at 30 pN. The initial rates of ssDNA compaction ( $k_r^i$ , C) and subsequent elongation ( $k_r^f$ , D) due to ORF1p binding are equivalent for the four trimers.



**Figure 5: Dissociation phases of active (111p, m14) and inactive (151p, m15) trimers at 30 pN**

(A) Representative 111p dissociation curve showing two phases of dissociation at 30 pN: an initial re-compaction ( $\Delta x_i$ ) followed by ssDNA elongation ( $\Delta x_f$ ). (B) All ORF1p trimers eventually compact the DNA to the same extent during the initial dissociation phase. Bar graphs show the average amplitude of compaction from multiple ( $n \geq 3$ ) experiments with each variant. However, re-compaction of the ssDNA occurs approximately twice as fast with the inactive variants as it does with the active proteins (C). Final dissociation of the inactive ORF1 proteins is both more complete (D) and faster (E) than the active trimers. In contrast, the ssDNA binding dynamics of all three ORF1p variants (m14, m15, and 151p) are identical to those of the wild type at 30 pN (shown in Fig. 4).