Supplementary Information

L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling

M. Nabuan Naufer[2], Kathryn E. Callahan[1], Pamela R. Cook[1], Cesar Perez-Gonzalez[1], Mark C. Williams[2], and Anthony V. Furano[1]

[1] The Laboratory of Molecular and Cellular Biology, NIDDK, NIH, Bethesda, MD 20892, USA.
[2] Northeastern University, Department of Physics, Boston, MA 02115, USA.

Supplementary Methods	Supplementary single molecule methods
Supplementary Figure S1	Aligned amino acid sequences of 111p, 151p and 555p
Supplementary Figure S2	Densitometry of the electropherogram in Figure 4
Supplementary Figure S3	Single molecule method to quantify fractions of ssDNA-bound ORF1p populations with slow, intermediate and fast kinetic states
Supplementary Figure S4	Representative curves for direct measurements of intermediate and fast dissociation time constants of ssDNA-bound ORF1p variants
Supplementary Figure S5	Representative return and subsequent stretch data upon different incubation times with the specified ORF1p variant, and the corresponding fits
Supplementary Table S1	Oligonucleotides used in this study
Supplementary Table S2	Fitting parameters of the normalized 680/590 FRET ratio in the NA chaperone activity assay, as a function of time
Supplementary Table S3	Fitting parameters of the extension during the constant force feedback loop as a function of time
Supplementary Table S4	Fitting parameters of ssDNA fractions bound as a function of incubation time

Supplementary Information

Supplementary Methods

Supplementary single molecule methods

A torsionally unconstrained biotinylated bacteriophage λ DNA molecule was tethered from its opposite ends between two streptavidin-coated polystyrene beads in 10 mM HEPES, 50 mM Na^+ at pH 7.5. One bead was held in an optical trap while the other was immobilized on a micropipette tip, attached to a flow cell placed on a translational piezoelectric stage. By gradually moving the fixed bead (at a pulling rate of 400 nm/s) while recording the extension and the force exerted on the single DNA molecule, the force-extension profile of a dsDNA in the absence of ORF1p was obtained.

An overstretched double-stranded DNA (dsDNA) anneals quickly with minimal hysteresis between the stretch and release curves in the absence of ORF1p. To quantify the ssDNA fractions bound by distinct ORF1p kinetic states, we first overstretched the dsDNA (Force ≈75 pN) to construct an array of single-stranded DNA (ssDNA) binding sites and then incubated it in 2 nM ORF1p variant (111p, 555p or 151p) for 360 s. After incubation we obtained the return curve of the resulting DNA-ORF1p complex (Figures 5B, E, F and Supplementary Figure S3, blue circles). A second curve is obtained by subsequently stretching the DNA-ORF1p complex immediately after the initial return (Figures 5C, D, F and Supplementary Figure S3, red circles). The experiments were conducted in 10 mM HEPES, 50 mM Na^+ buffer solution at pH 7.5 with a pulling rate of 400 nm/s and repeated for incubation times, 180, 720, 1440 and 2880 s (Figure 7A and Supplementary Figure S5).

The DNA can be considered ORF1p-saturated when the hysteresis stops increasing with protein concentration or incubation time. The resulting complex is an array of ssDNA in which all the possible binding sites of ssDNA are occupied and stabilized by ORF1p and thus significantly longer than dsDNA (Supplementary Figure S5, gold circles in top row). However, it is shorter than ssDNA (Figure 5A and Supplementary Figure S3, purple line) due to elasticity changes to the bound protein complex. In unsaturated incubation conditions the resulting complex is a combination of dsDNA and ORF1p-bound ssDNA sites in which duplex formation is inhibited. Thus the increase in the extension of the ORF1p-DNA complex in the force regime below the melting plateau compared to the extension of a dsDNA in the absence of ORF1p, measures the fraction (f) of the ORF1p-bound ssDNA. We postulated that the total fraction is bound by a combination of ORF1p populations exhibiting slow (f_{slow}), intermediate (f_{int}), and fast (f_{fast}) dissociation kinetics. We quantified these fractions with reference to the dsDNA in the absence of ORF1p (f=0) and ORF1p-saturated ssDNA (f=1) curves, which represent the two extremes. As there is no current theoretical model that provides the force-extension relationship of these protein-saturated ssDNA complexes, we fit the saturated data to a second order polynomial using least squares minimization to obtain the polynomial coefficients A, B and C and find the extension of the saturated curve ($b_{sat}(F)$) as a function of force, F :

$$b_{sat}(F) = \frac{-B + \sqrt{B^2 - 4A(C - F)}}{2A} \quad (1)$$

The polynomial coefficients [A, B, C] for 111p, 555p and 151p were found to be [878.51, -489.43, 67.81], [920.38, -543.78, 82.424] and [971.88, -607.21, 98.637] respectively (Figures 5D-F, Supplementary Figures S3, S5, gold line)

Supplementary Information

The extensible Worm-like chain (WLC) model was used to obtain the theoretical extension $(b_{ds}(F))$ of a dsDNA molecule as a function of force (Figures 5D-F, Supplementary Figures S3, S5, grey line), where (1)

$$b_{ds}(F) = B_{ds}\left\{1 - \frac{1}{2}\left(\frac{k_B T}{F P_{ds}}\right)^{1/2} + \frac{F}{S_{ds}}\right\} \quad (2)$$

The extensible freely-jointed chain (FJC) models the extension of an ssDNA ($b_{ss}(F)$) as a function of force (Figure 5A and Supplementary Figure S3, purple line), where (2)

$$b_{ss}(F) = B_{ss}\left[\coth\left(\frac{2P_{ss}F}{K_B T}\right) - \frac{1}{2}\frac{k_B T}{P_{ss}F}\right]\left[1 + \frac{F}{S_{ss}}\right] \quad (3)$$

The typical parameters for stretch moduli (S_{ds}= 1361 pN, S_{ss}= 720 pN), persistence lengths (P_{ds}= 45 nm, P_{ss}= 0.75 nm) and contour lengths (B_{ds}= 0.34 nm/bp, B_{ss}= 0.55 nm/bp) were used.

At any given force below the melting plateau ORF1p-DNA complex is a combination of dsDNA and ORF1p-bound ssDNA fractions. Thus the extensions of the ORF1p-DNA complexes (b(F)) were modeled as a linear combination of $b_{sat}(F)$ and $b_{ds}(F)$, where

$$b(F) = (1 - f)b_{ds}(F) + f b_{sat}(F) \quad (4)$$

to find the ORF1p-bound ssDNA fraction, f. By definition, $b_{ds}(F)$ and $b_{sat}(F)$ yield the fits for which f is 0 and 1, respectively. The subsequent stretch yields the fraction (f_{slow}) of the ssDNA-bound ORF1p population that exhibits slow dissociation kinetics. We computed f_{slow} by fitting the subsequent stretch data to Supplementary equation 4 (Figures 5D, F, red line, Supplementary Figure S3, #1, and Supplementary Figure S5, bottom row). Assuming that the ORF1p-bound ssDNA fraction that emerged during the incubation remains constant throughout the return after incubation, the return curve should trace the linear combination intersecting F_0 (Figure 5E, dashed blue line, Supplementary Figure S3, #2), where F_0 is the force at which the return curve begins to approach the force regime below the melting plateau. Thus, using Supplementary equation 4, we found the linear combination intersecting F_0 to determine the total ORF1p-bound ssDNA fraction, f_T:

$$f_T = f_{slow} + f_{int} + f_{fast} \quad (5)$$

However, the discrepancy between the dashed blue line and the observed return data (Figure 5E and Supplementary Figure S3, blue circles) is due to continuous duplex formation at a rate determined by the rapidly dissociating ssDNA-bound ORF1p, which decreases the fast fraction, f_{fast} during the return. In order to account for the fast dissociating protein, we modified f_{fast} to be varied with a phenomenological force dependence ($f_{fast} \rightarrow \tilde{f}_{fast}(F)$) and modeled the return data with a varying total fraction $\tilde{f}_T(F)$;

$$\tilde{f}_T(F) = f_{slow} + f_{int} + \tilde{f}_{fast}(F) \quad , \quad (6)$$

where

$$\tilde{f}_{fast}(F) = \gamma F^3 \quad (7)$$

and γ is a constant.

By definition

$$f_{fast} = \tilde{f}_{fast}(F_0) = \gamma F_0^3 . \quad (8)$$

Substituting for f_{fast}, Supplementary equation 5 can be re-written as

$$f_T = f_{slow} + f_{int} + \gamma F_0^3 , \quad (9)$$

to obtain f_{int} in terms of γ, where

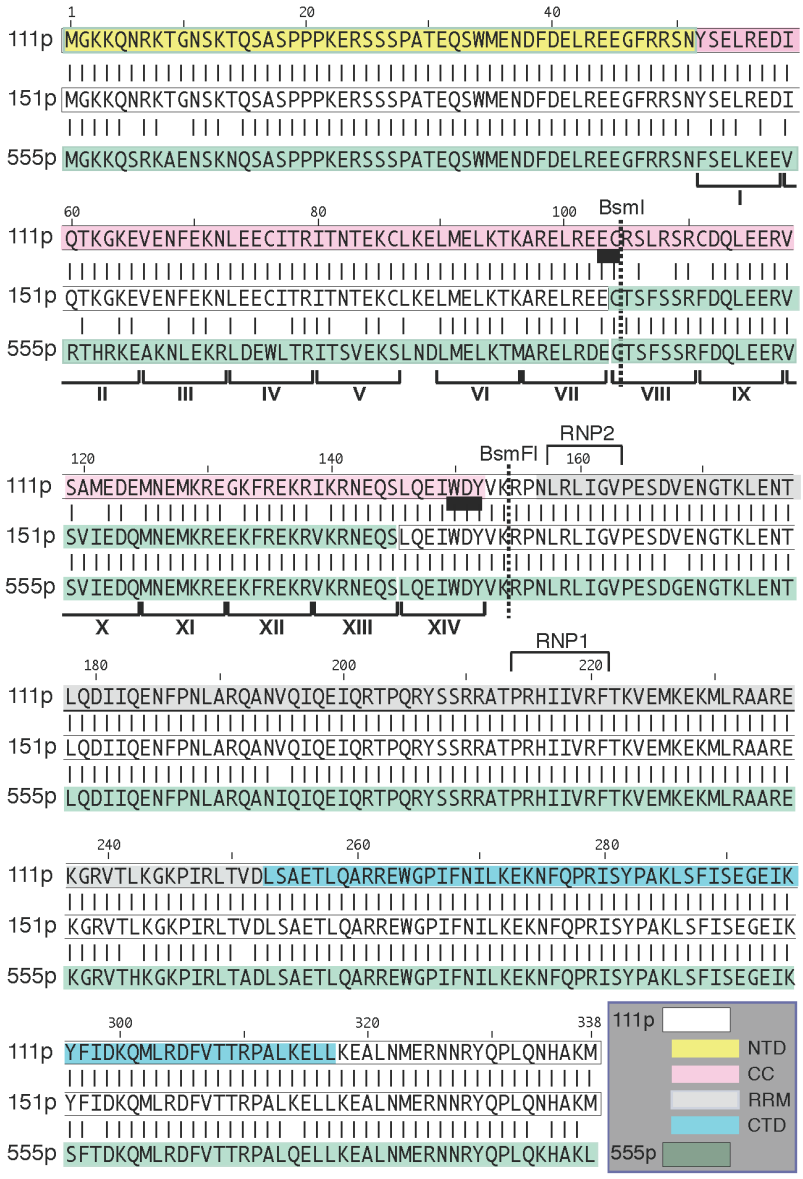
$$f_{int} = f_T - f_{slow} - \gamma F_0^3 \quad (10)$$

Substituting for f_{int} in Supplementary equation 6, we obtained $\tilde{f}_T(F)$, where

$$\tilde{f}_T(F) = f_T - \gamma (F_0^3 - F^3). \quad (11)$$

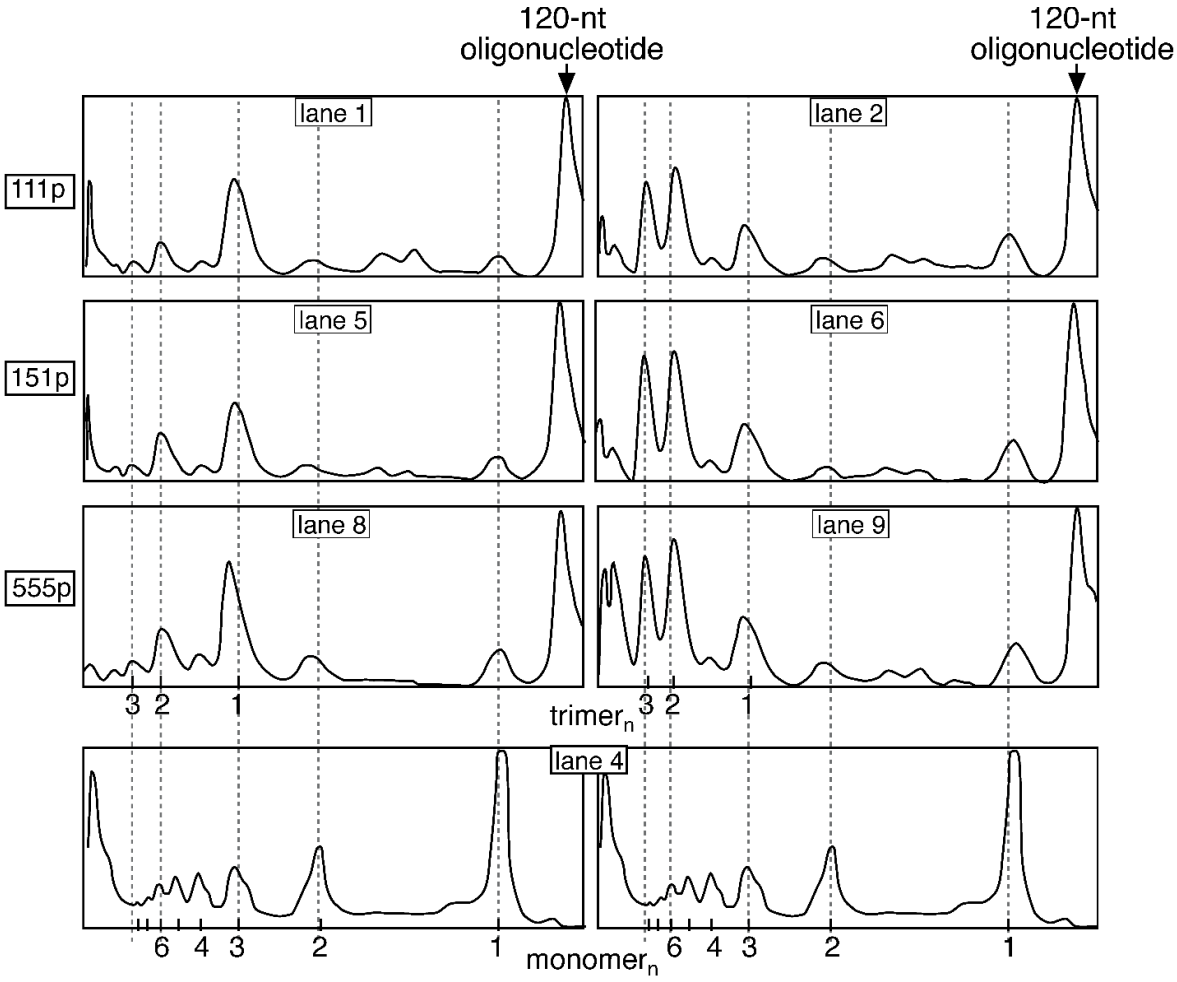
Since the values for F_0 and f_T are already found, by substituting f of Supplementary equation 4 with the expression in Supplementary equation 11 and fitting the return after incubation data (Figures 5E-F, solid blue line, Figure 7A, Supplementary Figure S3**,** #3, and Supplementary Figure S5, top row), we determined γ. We then found f_{fast} and f_{int} from Supplementary equations 8 and 10, respectively. The dashed green line in Figure 5F and Supplementary Figure S3 (#4) represents the linear combination of the combined fractions f_{int} and f_{slow}.

Supplementary Information



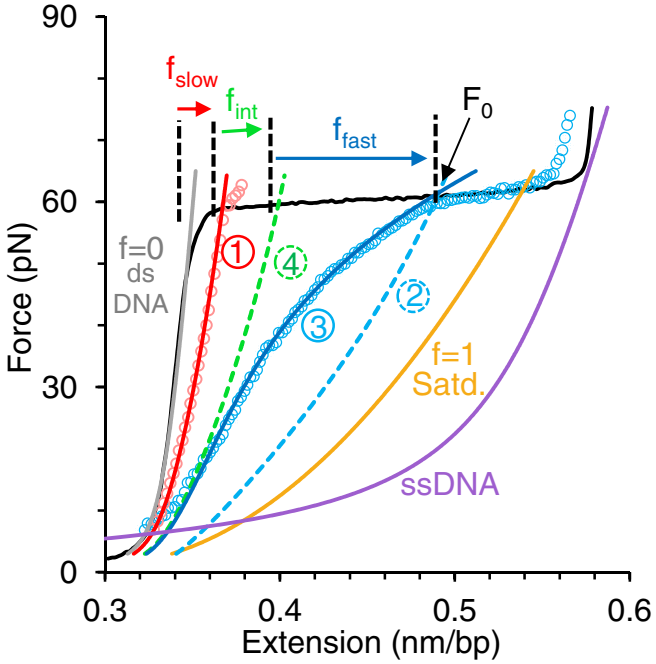
Supplementary Figure S1. Aligned amino acid sequences of 111p, 151p and 555p. The 111p sequence is shaded with the same colors as the domains depicted in Figure 1A. The amino acid sequences that correspond to ancestral 555p are shaded in green. Exact amino acid matches are indicated by vertical lines, and the highly conserved BsmI and BsmFI sites as well as the RNP2 and RNP1 domains of the RRM are indicated. The heavy black lines below the BsmI and BsmFI sites correspond to sites encoded by the recognition sequences for these enzyme. See main text for additional references.

Supplementary Information



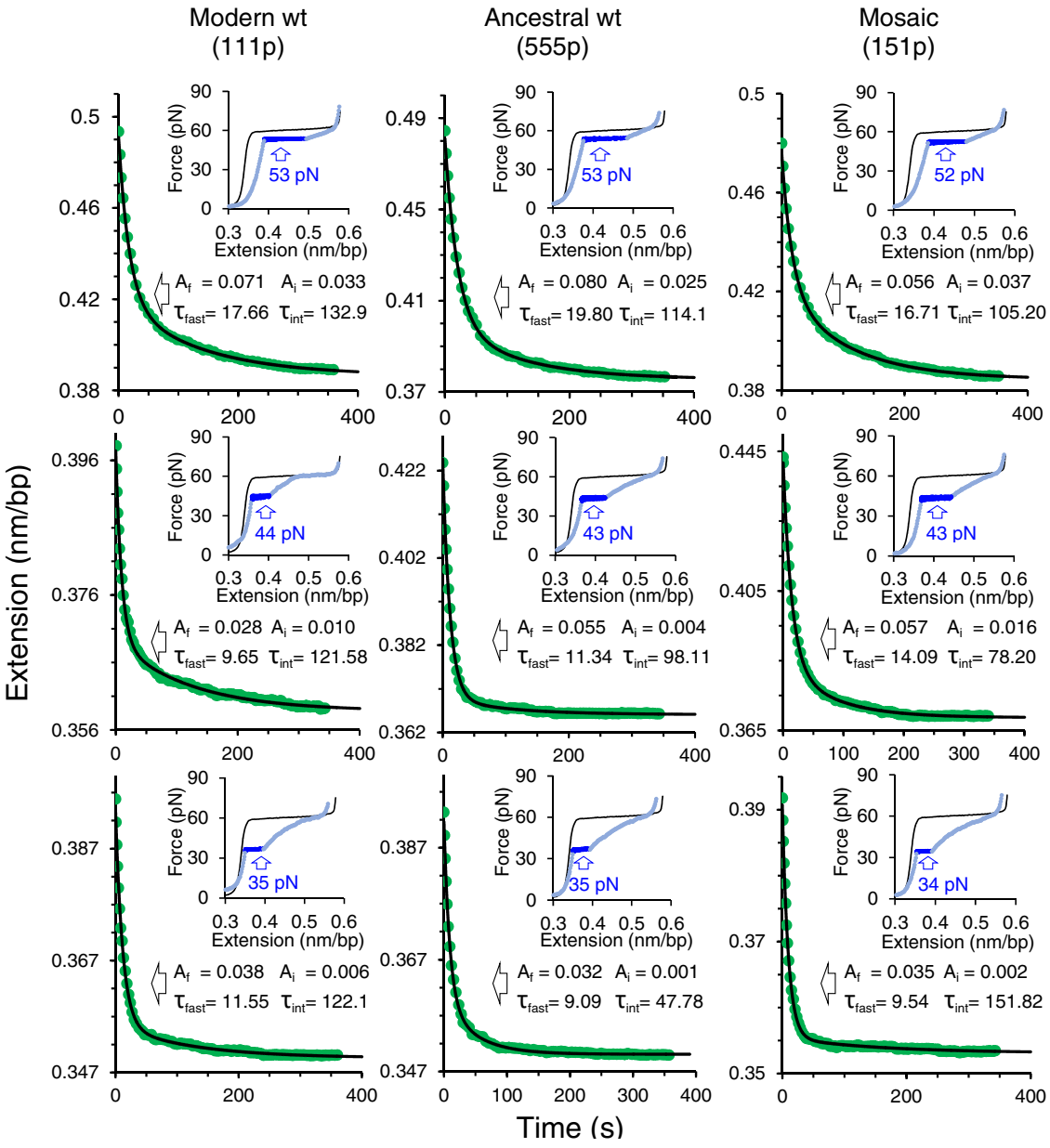
Supplementary Figure S2. Densitometry of the electropherogram in Figure 4.Traces of the indicated lanes were generated using the ImageJ gel profile tool. The band corresponding to the oligonucleotide (120-nt) is indicted with an arrow. The positions of the monomeric and trimeric ORF1p species are indicated.

Supplementary Information



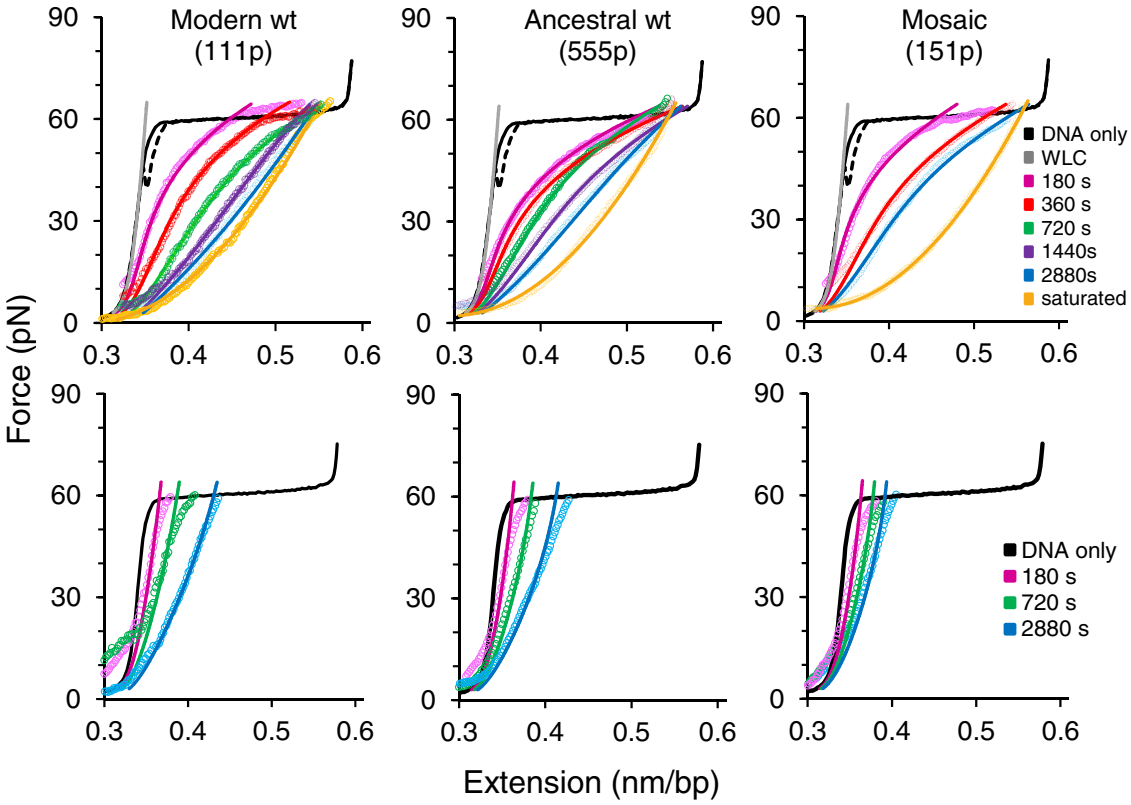
Supplementary Figure S3. Single molecule method to quantify fractions of ssDNA-bound ORF1p populations with slow, intermediate and fast kinetic states. Grey and gold solid lines are theoretical fits for dsDNA (extensible WLC, Supplementary equation 2) and ORF1p (111p)-saturated ssDNA (Supplementary equation 1), respectively. Purple line represent the theoretical ssDNA curve obtained by the extensible FJC model (Supplementary equation 3). Blue and red open circles are representative data of return after incubating in 2 nM 111p for 360 s and the subsequent stretch, respectively. Data modeling follows the numbered order as described in the Supplementary Methods to find f_{slow}, f_{int} and f_{fast}. 1. Subsequent stretch data are fit as a linear combination of dsDNA and ORF1p-saturated ssDNA (Supplementary equation 4) to find f_{slow} (red line) 2. Total ORF1p-bound ssDNA fraction (f_T, Supplementary equation 5) is found by the linear combination intersecting F_0 (dashed blue line), where F_0 is the force at which the return curve begins to approach the force regime below the melting plateau. 3. The return data are modeled (solid blue line) accounting for the rapidly dissociating ORF1p during the return to find f_{fast} and f_{int} (Supplementary equations 6-11). 4. Dashed green line is the linear combination of the combined fractions f_{slow} and f_{int}.

Supplementary Information



Supplementary Figure S4. Representative curves for direct measurements of intermediate and fast dissociation time constants of ssDNA-bound ORF1p variants (Figure 6). Insets are the DNA stretch (black line) and return after incubating for 360 s (light blue circles) with the specified ORF1p variant. The DNA-ORF1p complex was stopped and maintained at the noted force (constant force feedback loop) during the return while recording the extension as a function of time (dark blue line). Green circles represent the recorded extension as a function of time during the constant force feedback loop, which yields an exponential decay (black line) with specified intermediate (τ_{int}) and fast (τ_{fast}) dissociation time constants (in seconds). A_i and A_f are the corresponding amplitudes (in nm/bp) of the exponentials (Supplementary Table S3).

Supplementary Information



Supplementary Figure S5. Top row: Representative return curves upon different incubation times with the specified ORF1p variants. Measured data are in open circles and the solid lines are the corresponding fits (Supplementary Figure S1, #3). Bottom row: Representative subsequent stretches after the return upon chosen incubation times are in open circles) and the solid lines are the corresponding fits (Supplementary Figure S1, #1).

Supplementary Information

Supplementary Table S1. Oligonucleotides used in this study L is the oligonucleotide length in bases and the name is as discussed in the text.

#	name	L	sequence
1	d120_c	120	AGCGAGTTGATGTTAGACTGTGTACTTTTTGCGAGTTGATGTTAGACTGTGTACTTTTTAAGCGAGTTGATGTTAGACTGTGTACTTTTTGCGAGTTGATGTTAGACTGTGTACTTTTTA
2	d29	29	AAAAAGTACACAGTCTAACATCAACTCGC
3	d29_c-mm	29	GCGAGTTGAcGTcAGACcGTGcACTTTTT
4	d29_c	29	GCGAGTTGATGTTAGACTGTGTACTTTTT
5	21r_dna	21	ACTGCTAGAGATTTTCCACAT
6	cy3_21r_mm_dna	21	ACTGCcAGAGAcTTcCCACAT
7	cy5_21f_dna	21	ATGTGGAAAATCTCTAGCAGT
8	F-BamHI-111	30	CGCGGATCCGCAATGGGGAAAAAACAGAAC
9	R-Flag-EcorI-111	49	GCCGGAATTCCTACTTGTCGTCGTCGTCCTTATAATCCATTTTGGCATG
10	F-BamHI-555	30	CGCGGATCCGCAATGGGGAAAAAACAGAGC
11	R-Flag-EcorI-555	49	GCCGGAATTCCTACTTGTCGTCGTCGTCCTTATAATCCAATTTGGCATG
12	histev_t	69	GTACGCCATTATGGTGCATCATCATCATCATCATGCTAGCATCGAGAACCTGTACTTCCAGGGCATCGC
13	histev_b	69	GGCCGCGATGCCCTGGAAGTACAGGTTCTCGATGCTAGCATGATGATGATGATGATGCACCATAATGGC

Supplementary Information

Supplementary Table S2. Fitting parameters of the normalized 680/590 FRET ratio in the NA chaperone activity assay, as a function of time. Annealing and exchange phases of NA chaperone activity reactions (Figure 3) are fit to single and double exponential functions in time (t), respectively. These reactions were carried out in duplicate and fits and uncertainties for each measurement were obtained using the minimization of χ^2 method. Weighted means and uncertainties in the fitting parameters for the two measurements are shown in the table. The annealing phase is started by incubating mismatched oligonucleotides cy3_21r_mm_dna and cy5_21f_dna (Supplementary Table S1, #6-7), at t=0. The exchange phase is initiated at t = t*=180 s by addition of a 10 fold excess of the perfect complement, 21r_dna (Supplementary Table S1, #5) of cy5_21f_dna (Supplementary Table S1, #7).

	Annealing phase		Exchange phase			
Fit	$y_{an} = a_0(1-e^{-k_0 t})$		$y_{ex}(t \geq t^*) = a_1 e^{-k_1(t-t^*)} + a_2 e^{-k_2(t-t^*)} + (a_0 - a_1 - a_2)$			
ORF1p variant	a_0	k_0 (s^{-1})	a_1	k_1 (s^{-1})	a_2	k_2 (s^{-1})
Modern wt (111p)	0.95 ± 0.01	0.11 ± 0.01	0.388 ± 0.004	0.42 ± 0.06	0.273 ± 0.004	0.014 ± 0.002
Ancestral wt (555p)	0.98 ± 0.01	0.18 ± 0.02	0.383 ± 0.004	0.42 ± 0.07	0.341 ± 0.004	0.031 ± 0.002
Mosaic (151p)	0.98 ± 0.01	0.08 ± 0.01	0.479 ± 0.004	0.42 ± 0.05	0.255 ± 0.004	0.027 ± 0.002

Supplementary Information

Supplementary Table S3. Fitting parameters of the extension during the constant force feedback loop as a function of time. A double exponential b(t) with characteristic time constants τ_i and τ_f well describes the temporal extension at the reported stopped forces (F_{st}) revealing the existence of fast and intermediately dissociating ssDNA-bound ORF1p populations (f_{fast} and f_{int}). The exponential function converges to an extension (b_0) larger than that of an extension of a dsDNA ($b_{ds}(F)$, Supplementary equation 2) at corresponding forces, due to the ssDNA-bound ORF1p population (f_{slow}) exhibiting negligible dissociation during the timescales of this experiment (see also Figure 6 and Supplementary Figure S4). Standard errors are calculated using at least three measurements.

Fit / ORF1p variant	$b(t) = A_f e^{-t/\tau_{fast}} + A_i e^{-t/\tau_{int}} + b_0$						
	F_{st} (pN)	A_f (nm/bp) × 10^{-3}	τ_{fast} (s)	A_i (nm/bp) × 10^{-3}	τ_{int} (s)	b_0 (nm/bp)	$b_{ds}(F)$ (nm/bp)
Modern wt (111p)	53.4 ± 0.1	67 ± 9	16.0 ± 0.7	106 ± 35	142.6 ± 35.7	0.379 ± 0.003	0.348
	44.3 ± 0.4	27 ± 1	9.6 ± 0.2	12 ± 1	92.4 ± 15.4	0.358 ± 0.001	0.345
	34.5 ± 0.8	21 ± 8	9.3 ± 1.2	8 ± 1	85.3 ± 26.7	0.357 ± 0.097	0.341
Ancestral wt (555p)	52.7 ± 0.4	71 ± 5	19.8 ± 0.4	28 ± 6	104.5 ± 5.3	0.377 ± 0.002	0.348
	43.7 ± 0.3	51 ± 6	12.4 ± 1.1	6 ± 1	120.8 ± 17.5	0.363 ± 0.003	0.345
	35.1 ± 0.4	30 ± 3	9.2 ± 2.3	5 ± 2	88.6 ± 42.8	0.369 ± 0.012	0.342
Mosaic (151p)	52.1 ± 0.2	44 ± 7	16.6 ± 0.8	42 ± 8	105.5 ± 4.0	0.393 ± 0.004	0.347
	42.5 ± 0.4	45 ± 4	12.6 ± 1.6	19 ± 3	81.8 ± 10	0.372 ± 0.005	0.344
	33.9 ± 0.1	35 ± 6	12.2 ± 3.3	12 ± 4	121.3 ± 36	0.352 ± 0.001	0.341

Supplementary Information

Supplementary Table S4. Fitting parameters of ssDNA fractions bound as a function of incubation time, $\tilde{t}$. Corresponding exponential fitting functions and their parameters are reported for each ORF1p variant. The characteristic decay time constant of f_{fast} is defined as the oligomerization time constant (T_{oligo}) and shown in **bold** text. Fits and uncertainties are obtained using minimization of χ^2 method (see Figure 7B).

	Slow		Intermediate		Fast			Total	
Fit	$f_{slow}(\tilde{t}) = \theta_s(1-e^{-\tilde{t}/T_s})$		$f_{int}(\tilde{t}) = \theta_i(1-e^{-\tilde{t}/T_i})$		$f_{fast}(\tilde{t}) = \theta_f(e^{-\tilde{t}/T_f} - e^{-\tilde{t}/T_{oligo}})$			$f_T(\tilde{t}) = \theta_T(1-e^{-\tilde{t}/T_T})$	
ORF1p variant	θ_s	T_s (s)	θ_i	T_s (s)	θ_f	T_f (s)	$\mathbf{T_{oligo}}$ **(s)**	θ_T	T_T (s)
Modern wt (111p)	0.410 ± 0.005	885 ± 78	0.58 ± 0.02	632 ± 87	14 ± 1	384 ± 2	**356 ± 2**	0.99 ± 0.1	219 ± 24
Ancestral wt (555p)	0.342 ± 0.004	1010 ± 37	0.57 ± 0.04	1610 ± 175	0.94 ± 0.03	101 ± 2	**1620 ± 124**	0.98 ± 0.01	96 ± 14
Mosaic (151p)	0.24 ± 0.01	967 ± 9	0.31 ± 0.01	614 ± 6	0.57 ± 0.01	78 ± 6	**18000 ± 8190**	1.00 ± 0.01	205 ± 20

Supplementary References

1. Odijk, T. (1995) Stiff chains and filaments under tension. *Macromolecules*, **28**, 7016-7018.
2. Smith, S.B., Finzi, L. and Bustamante, C. (1992) Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, **258**, 1122-1126.