

Research Article

Cite this article: Morse M, Cashen BA, Rouzina I, Williams MC (2025). Diverse single-stranded nucleic acid binding proteins enable both stable protection and rapid exchange required for biological function. *QRB Discovery*, 6: e1, 1–12 <https://doi.org/10.1017/qrd.2024.21>.

Received: 24 June 2024

Revised: 26 November 2024

Accepted: 27 November 2024

Keywords:


optical tweezers; protein–nucleic acid interactions; single molecule biophysics

Corresponding author:

Mark C. Williams;

Email: ma.williams@northeastern.edu

Diverse single-stranded nucleic acid binding proteins enable both stable protection and rapid exchange required for biological function

Michael Morse¹, Ben A. Cashen¹, Ioulia Rouzina² and Mark C. Williams¹ 

¹Department of Physics, Northeastern University, Boston, MA, USA and ²Department of Chemistry and Biochemistry, Ohio State University, Columbus, OH, USA

Abstract

Single-stranded nucleic acid (ssNA) binding proteins must both stably protect ssNA transiently exposed during replication and other NA transactions, and also rapidly reorganize and dissociate to allow further NA processing. How these seemingly opposing functions can coexist has been recently elucidated by optical tweezers (OT) experiments that isolate and manipulate single long ssNA molecules to measure conformation in real time. The effective length of an ssNA substrate held at fixed tension is altered upon protein binding, enabling quantification of both the structure and kinetics of protein–NA interactions. When proteins exhibit multiple binding states, however, OT measurements may produce difficult to analyze signals including non-monotonic response to free protein concentration and convolution of multiple fundamental rates. In this review we compare single-molecule experiments with three proteins of vastly different structure and origin that exhibit similar ssNA interactions. These results are consistent with a general model in which protein oligomers containing multiple binding interfaces switch conformations to adjust protein:NA stoichiometry. These characteristics allow a finite number of proteins to protect long ssNA regions by maximizing protein–ssNA contacts while also providing a pathway with reduced energetic barriers to reorganization and eventual protein displacement when these ssNA regions are diminished.

Introduction*Optical tweezers force spectroscopy*

Optical tweezers (OT) can isolate single nucleic acid (NA) molecules in suspension by tethering both ends to functionalized micron-sized beads which are held by laser trap(s). The system directly controls the end-to-end extension of the NA molecule based on the relative displacement of the two trapped beads while simultaneously measuring applied force based on the deflection of the trapping laser as the beads shift from the center of the trap. One method to measure the properties of the NA molecule is to produce a force–extension curve (FEC), in which the force response is measured over a wide range of extensions (Figure 1a), typically by increasing or decreasing the extension in fixed steps at a continuous rate starting from low or high extensions, respectively. Both DNA and RNA are well represented by polymer chain models, with double-stranded (ds) and single-stranded (ss) polymers having very different properties. While dsDNA is modeled as an extensible worm-like chain (WLC) (Baumann et al., 1997; Odijk, 2002), the end-to-end extension (X) of ssNA as a function of applied force (F) is best modeled as a freely jointed chain (Smith et al., 1996):

$$X(F) = L \left(\coth \left(\frac{2PF}{k_B T} \right) \right) \left(1 + \frac{F}{S} \right) \quad (1)$$

Both ssDNA and ssRNA have a contour length (L) of ~ 0.56 nm/nt and a persistence length (P) of ~ 0.75 nm, which reflect the length and flexibility, respectively, of the sugar–phosphate backbone (Figure 1a). The backbone can elastically stretch as well, with elongation linearly proportional to applied force, with the elastic modulus (S , ~ 600 pN for ssDNA and ssRNA) indicating the force required to double the length of the substrate. Thus, the FEC has two main regimes: at low force, the FEC gently curves upwards as the ssNA molecule is straightened, approaching its contour length, and at high force, the extension continues to increase linearly past the contour length. Note that the FJC does not account for secondary structure, which can form at low force. Structures such as large defined hairpins found in certain biological systems can persist at higher forces. OT can be used to study specific interactions between these sequences and binding proteins, but here we focus on long (>1 knt), 50% GC content, substrates without stable secondary structure (>5 pN applied tension), in order to focus on non-sequence specific interactions between ssNA and proteins. That is, we average over the behavior of many sequences

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

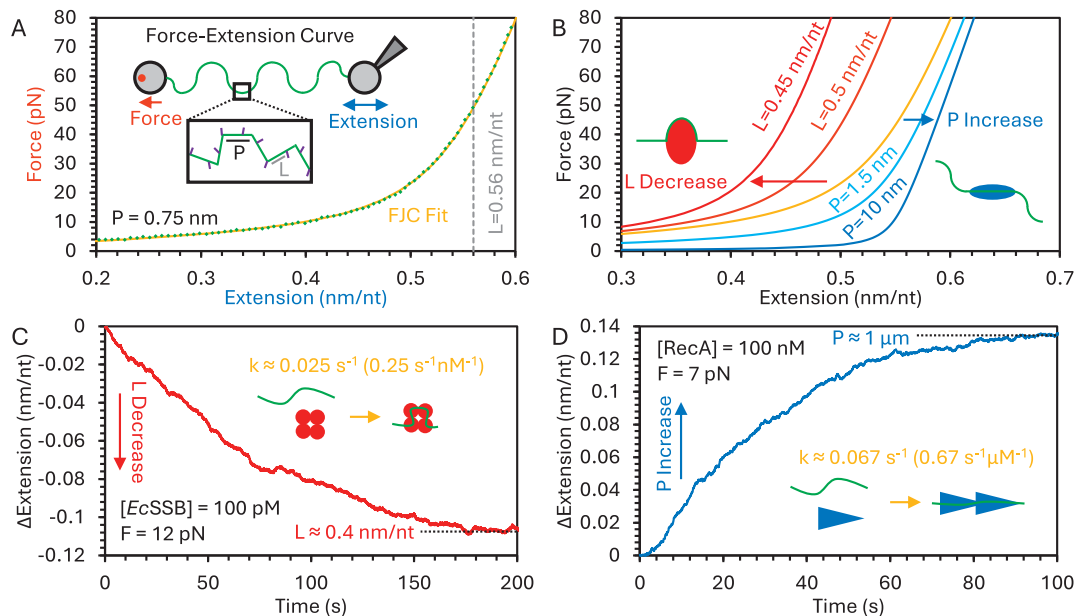


Figure 1. OT measurement of ssDNA conformation. (a) A single ssDNA molecule is tethered between two beads in an OT system to simultaneously measure end-to-end extension (blue) and applied force (red), generating an FEC (green). The data is fit by the FJC model (yellow), yielding the effective persistence and contour lengths. (b) The binding of proteins to the ssDNA can lower the contour length or increase the persistence length, resulting in a decrease or increase in extension, respectively, for a given applied force. (c) ssDNA bound by EcSSB is wrapped around the protein homotetramer, effectively shortening its contour length. (d) RecA forms long, semi-rigid filaments along ssDNA, causing an increase in persistence length. Extension changes measured for ssDNA held at fixed tension in a protein reservoir are monotonic, exponentially approaching an equilibrium value at a rate governed by the concentration-dependent bimolecular NA–protein interaction.

present on the substrate, as the proteins discussed below also must interact with a wide range of sequences to coat exposed ssNA, as opposed to interacting only with specific sequences or motifs.

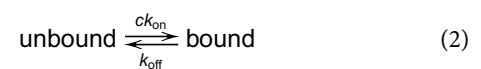
The polymer properties of the ssNA molecule determine the shape of its FEC, and if its properties change, so too will its FEC. For instance, an increase in persistence length both lowers the force required to straighten the ssNA and sharpens the transition between the low-force entropic regime and the high-force elastic regime (Figure 1b). In comparison, a reduction in contour length proportionally decreases ssNA extension at all forces, since extension scales linearly with contour length. In particular, the binding of a protein (or other small molecule) to ssNA can drastically impact its conformation. During binding, a length of ssNA adheres to a binding surface on the protein, such that its conformation is determined by the structure of the NA–protein complex. This has two primary effects. First, if the binding interface is not straight, bound ssNA must bend to adhere to the protein, which effectively lowers the substrate’s contour length (Figure 1c). This effect is more pronounced for more circuitous paths through the binding surface, such as ssNA fully wrapping around the protein. Since ssDNA extension is directly proportional to contour length (eq. (1)), a consistent decrease in ssDNA extension is observed for all forces in the FEC. Second, since the binding interface is nearly always longer than the persistence length of bare ssNA (~0.75 nm or less than 2 nt), bound protein increases the effective persistence length (Figure 2c). The new effective persistence length can be determined by the number of nt bound by each individual protein or can become even longer if multiple proteins form a filament that remains rigid over many repeating protein subunits. ssDNA extension has a non-linear response to a change in persistence length (eq. (1)), such that increased persistence length greatly increases ssDNA extension at low force but only mildly increases extension at high force when ssDNA is mostly straightened.

Proteins binding to ssDNA typically display both these effects, such that both effects can be observed in a single FEC, with increased persistence length increasing extension at low force but decreased contour length decreasing extension at high force (Cashen *et al.*, 2023; Morse *et al.*, 2019). However, applied force on the ssDNA may also impact the binding affinity and/or binding conformation of proteins, as tension inhibits wrapping.

In either case, when a ssNA substrate is incubated with a protein that has such an effect when bound, the extension change over time has the shape of exponential decay, the rate and amplitude of which depend on the fundamental nature of the protein–NA interaction as discussed below.

NA–protein binding kinetics

The simplest protein–NA interaction to characterize is one in which the protein has only one binding conformation (Figure 2a). Specifically, each bound protein has the same binding site size (occludes the same number of nts). In this case, the ssNA substrate can be modeled as an array of binding sites, where each site acts independently and can occupy one of two states, protein-free and protein-bound.



Transition rates between these two states depend on the specific protein–NA interaction (e.g. different proteins have different binding affinities), but the rates are equivalent for each identical binding site. The rate of binding (of unbound sites becoming bound) is dependent on the concentration of free protein (c) and is denoted as $c k_{\text{on}}$, while the rate of dissociation (of bound sites becoming unbound) is concentration-independent and is denoted as k_{off} . Zooming out, the degree of protein saturation (θ) for the entire

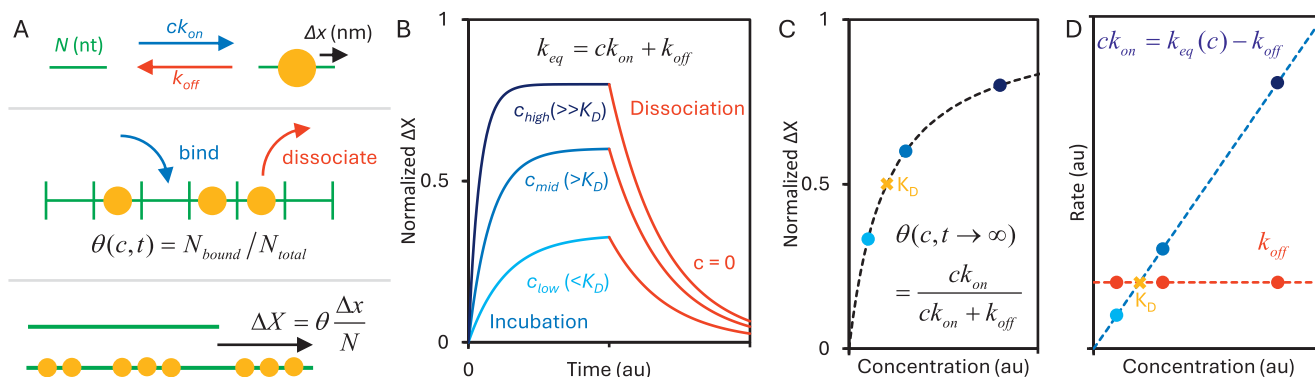


Figure 2. Simple bimolecular, reversible binding. (a) An NA substrate is modeled as an array of a fixed number of binding sites. Each site is protein-free or occupied, and the fraction of sites occupied depends on the free protein concentration and the time after the protein is introduced into the system. Each binding site is the same length (N) and acts independently. Empty sites are filled by proteins at a concentration-dependent on-rate and bound proteins leave at a fixed off-rate. The effective length of each binding site is changed by a fixed length (Δx) when protein is bound so that the length difference between an entire protein-free and protein-saturated substrate is the length change divided by the binding site size. (b) When protein is added to the sample (blue), the ssDNA extension changes at an observed rate equal to the sum of the on and off rates, and the total extension change reached at equilibrium increases with protein concentration. When free protein is removed, the ssDNA returns to its protein-free state. (c) The equilibrium ssDNA extension change as a function of protein concentration forms a binding isotherm, with the protein concentration that results in half as much change as seen at full saturation equal to K_D , by definition. (d) The rate of binding increases proportionally with protein concentration while the rate of dissociation is constant. The two rates intersect at K_D .

NA substrate can be simply calculated as the number of binding sites occupied by a protein divided by the total number of binding sites.

$$\theta(c, t) = n_{\text{bound}} / n_{\text{total}} \quad (3)$$

This value will vary with free protein concentration and over time if the system is not in equilibrium. The OT system does not directly measure protein binding, but rather NA conformation (the substrate's change in extension, ΔX). Note, that the measured NA length (X) and change in length (ΔX), are typically reported in normalized units of nm/nt, where the absolute length in nm is divided by the substrate length in nt. This allows results to be compared and interpreted consistently regardless of the specific substrate used in experiments, which is typically arbitrary and determined by (commercial) availability or technical constraints rather than a specific biological function. For example, dsDNA consistently has a normalized contour length of 0.34 nm/nt for both commercially available and widely used lambda phage DNA (48.5 kbp) and plasmid pUC19 (2.7 kbp), even though their absolute extensions vary by more than an order of magnitude.

If only one protein binding conformation is possible, then the extension change and degree of protein saturation are directly proportional.

$$\Delta X = \theta \frac{\Delta x}{N} \quad (4)$$

The maximal extension change, which is achieved when the NA is fully protein-saturated, can be experimentally determined by titrating protein concentration (at sufficiently high protein concentration, the amplitude of the extension change will asymptote, indicating saturation). If all proteins bind in the same conformation, then each binding site (size N nt) will have its extension altered by a constant value Δx .

OT systems allow for changing free protein concentration around the isolated NA molecule, either by flowing in different buffers into the sample while keeping the trap stationary or by moving the trap into a different location within the sample. This drives the system out of equilibrium, which reveals the kinetics of the protein-NA interaction. When protein-free NA is suddenly introduced to free protein

(incubation), the extension change over time takes the form of an exponential decay (Figure 2b).

$$\Delta X(t) = \Delta X_{\text{eq}} (1 - e^{-k_{\text{eq}} t}) \quad (5)$$

The rate constant of equilibration (k_{eq}) is the sum of the two fundamental rates of binding (k_{on}) and dissociation (k_{off}).

$$k_{\text{eq}} = ck_{\text{on}} + k_{\text{off}} \quad (6)$$

Thus, higher protein concentrations (c) result in the NA substrate reaching its equilibrium extension on a shorter timescale. k_{off} can be easily isolated by measuring the rate of dissociation when free protein is removed from the sample ($c = 0$). In contrast, k_{on} can be calculated from a single incubation-dissociation cycle by subtracting the dissociation rate (measured after protein is removed) from the equilibration rate (measured during protein incubation) and dividing by the protein concentration. Alternatively, the equilibration rate can be measured for several protein concentrations and then linearly fit, with the zero-concentration intercept and the concentration proportional slope yielding k_{off} and k_{on} , respectively.

Additionally, the amplitude of the extension change at equilibrium increases with protein concentration (Figure 2c), as the extension change of the NA is directly proportional to the degree of protein saturation (Eq. (4)). The fraction of binding sites occupied at equilibrium (θ_{eq}) can be written in terms of the fundamental rates of binding and dissociation, or in relation to the effective dissociation constant (K_D), which indicates the protein concentration at which the rates of binding and dissociation are equal, and equivalently, half of the binding sites are protein-occupied (Figure 2d).

$$\theta_{\text{eq}} = \frac{ck_{\text{on}}}{ck_{\text{on}} + k_{\text{off}}} = \frac{c}{c + K_D} \quad (7)$$

The degree of protein saturation can also be directly measured based on the amplitude of the extension change at equilibrium, where the extension before protein introduction indicates $\theta_{\text{eq}} = 0$ and the maximum extension change approached at the highest protein concentrations signifies $\theta_{\text{eq}} = 1$. Agreement between calculations of these fundamental parameters using either rate or amplitude measurements confirms that the system follows a simple two-

state mechanism, governed by bi-molecular binding. Thus, when agreement is not found, such as for the three different protein systems explored below, the possibility of proteins binding in more than one conformation must be explored.

Single-stranded binding proteins

Single-stranded binding proteins (SSBs) are highly abundant proteins that have been identified in all domains of life, including viruses, and eukaryotic and prokaryotic cell nuclei. All SSBs have ssDNA (or ssRNA) binding grooves that do not discriminate between the ssNA sequences but do not bind double-stranded (ds) NA regions. SSBs are generally involved in DNA (and in some viruses RNA) replication, repair, and recombination. SSBs are able to promptly engage all available ssDNA (and ssRNA) templates, thereby protecting them from degradation by nucleases, as well as eliminating NA secondary structure. SSBs are known to facilitate all NA metabolic processes, most notably, the rate of DNA replication by the polymerase complex. As an extremely important class of proteins, SSBs from many organisms have been studied extensively over the past several decades, with much information on their structure and ssDNA binding modes accumulated. Despite their similar roles in various organisms from bacteria to humans, SSBs were found to be surprisingly diverse in their structure, ssDNA binding modes, and binding cooperativity. Most importantly, despite decades of research, it remains unclear: (i) how variable amounts of ssDNA template always remain protected by the variable bulk concentrations of SSBs during DNA processing and (ii) how strongly, and often cooperatively, bound SSBs are able to promptly dissociate from ssDNA to clear the way for the rapidly moving polymerase complex as it synthesizes the complementary DNA strand with rates of ~100–1000 nt/s. Many SSBs have C-terminal unstructured or poorly structured anionic tails that compete with ssDNA for binding to their SSB binding sites, while also serving as the attachment points for multiple cellular proteins that regulate ssDNA processing. It was suggested that the binding of C-terminal tails to these regulatory factors helps to promptly dissociate SSBs from ssDNA. However, no direct evidence of such activity was provided. Moreover, SSBs are also routinely used to improve the yields of *in vitro* PCR reactions where no such cellular cofactors are provided.

Despite commonalities in function, different SSBs need not be structurally similar. In this review, we specifically examine three SSB proteins, from various sources (bacteria, bacteriophage, and retrotransposon). Despite these differences, we will show that they all exhibit similar collective behaviors when examined in single-molecule experiments. Perhaps the most well-studied such protein is the SSB of *E. coli* (*EcSSB*). *EcSSB* is a homotetramer, with each 19 kDa monomer comprising an N-terminal domain (NTD) containing an oligonucleotide binding (OB) fold (the protein's ssDNA binding site/groove), a C-terminal domain (CTD) with a conserved 9-amino acidic tip, and a poorly conserved intrinsically disordered linker (IDL) (Antony et al., 2013; Kozlov et al., 2015; Raghunathan et al., 2000; Raghunathan et al., 1997; Tan et al., 2017). The OB domain contains both the high-affinity DNA binding grooves and interfaces for interprotein interactions responsible for stable tetramerization. *EcSSB* can bind ssDNA in multiple conformations, typically identified by the total number of nt occluded, such that higher binding site size states wrap more NA substrate around the OB tetramer (Bujalowski & Lohman, 1986; Bujalowski et al., 1988; Lohman et al., 1988; Lohman & Overman, 1985; Lohman et al., 1986). Free protein concentration (or equivalently protein:nt ratio),

salt conditions, and template tension (for force spectroscopy studies) affect the occupancy of these distinct binding modes (Bujalowski & Lohman, 1986; Bujalowski et al., 1988; Kozlov et al., 2019; Lohman et al., 1988; Lohman & Overman, 1985; Lohman et al., 1986; Suksombat et al., 2015). Three wrapping states are typically observed for *EcSSB* bound to ssDNA in the absence of applied tension, with binding site sizes of 65, 56, and 35 nt. However, OT experiments observing single proteins binding to a 70 nt ssDNA substrate identified binding of as little as 17 nt by an individual tetramer under increased tension (Suksombat et al., 2015).

A model of the path through which a ~65 nt ssDNA substrate accesses the binding grooves of all four OB domains has been established based on X-ray crystallographic structural data (Raghunathan et al., 1997). While the exact topologies of other binding modes have not been structurally resolved, they are generally consistent with a discrete number of the OB folds being occupied by ssDNA. Some experiments have observed evidence of ssDNA segments as short as 8 nt binding to *EcSSB*, including sedimentation of 8 nt poly dT oligos with a stoichiometry of more than 3 oligos per tetramer (Krauss et al., 1981), the addition of a poly dT ssDNA overhang to a hairpin increases protein-mediated hairpin destabilization, (Grieb et al., 2017), and AFM observation of *EcSSB* localization to 8 nt poly dT overhangs at the end of dsDNA substrates (Naufer et al., 2021).

Even when NA-bound, *EcSSB* is highly dynamic. Single-molecule FRET experiments measured a dynamic equilibrium between structural states (Roy et al., 2007), and diffusion of wrapped protein along its ssDNA substrate (Roy et al., 2009). Fluorescent imaging of *EcSSB*-ssDNA complexes has even resolved multiple sequential kinetic steps, from measurements of the concentration-dependent rate of free protein binding (Kozlov & Lohman, 2002b) and the concentration-independent rate of wrapping (Kuznetsov et al., 2006), to the slow addition of additional protein to an *EcSSB*-occupied substrate (Kunzelmann et al., 2010) and direct transfer of an *EcSSB* tetramer between two different ssDNA substrates (Kozlov & Lohman, 2002a).

The depth of research on *EcSSB* makes it a good model to compare with other less well-defined protein systems. The results we observe using optical tweezers (Naufer et al., 2021) can be directly related to previous research using different experimental systems. Since then, we have also observed significant commonalities with how both the gene 32 protein of T4 bacteriophage (T4 gp32) (Cashen et al., 2023; Cashen et al., 2024a) and the open reading frame (ORF) 1 protein of the LINE 1 retrotransposon (L1-ORF1p) (Cashen et al., 2022; Cashen et al., 2024b) interact with an ssNA substrate. As we will discuss further, the known framework in which *EcSSB* interacts with ssDNA in multiple binding conformations can be generalized to explain both our experimental data and the function of these other proteins.

T4 gp32 is a 33.5 kDa monomer comprising three domains: a central ssDNA binding core (residues 22–253), a positively charged NTD (residues 1–21), and a negatively charged CTD (residues 254–301) (Karpel, 1990). The gp32 core domain binds ssDNA (7 nt occluded site size) in a small, positively charged cleft created by a single OB-fold, conferring the protein with largely sequence-independent binding and the ability to effectively discriminate against duplexed dsDNA (Shamoo et al., 1995; Theobald et al., 2003; Wu et al., 1999). gp32 forms highly stable protein filaments on ssDNA mediated by cooperative interprotein interactions between the NTD of a nucleic acid-bound monomer and the core domain of an adjacently bound protein (Casas-Finet et al., 1992;

Lonberg et al., 1981). Previous light scattering and circular dichroism measurements suggested that NA-bound gp32 filaments wind ssDNA, resulting in a relatively stiff, helical protein-DNA structure (Kuil et al., 1990; Kuil et al., 1988; Scheerhagen et al., 1989; Scheerhagen et al., 1985a; Scheerhagen et al., 1985b; van Amerongen et al., 1990). These findings were recapitulated by recent single-molecule DNA stretching experiments, which showed that cooperatively bound gp32 simultaneously rigidifies and compacts ssDNA, characterized by its increased persistence length and reduced contour length, respectively (Cashen et al., 2023), an expected signature of a helical protein filament (Griffith & Formosa, 1985; Lee et al., 2004; Takahashi & Norden, 1994; Wu et al., 2004; Xu et al., 2017; Yang et al., 2001; Yu et al., 2004). The gp32 CTD, on the other hand, is believed to primarily help coordinate DNA replication via direct (heterotypic) interactions with other constituents of the T4 replisome (Alberts & Frey, 1970a; Alberts & Frey, 1970b; Krassa et al., 1991; Lefebvre et al., 1999; Morrical et al., 1996; Nelson et al., 2008) while also competing with the ssDNA for the central domain's binding cleft, thereby moderating the strength of the individual protein ssDNA interactions in a salt-dependent manner (Pant et al., 2005).

The exact structural details of the gp32-ssDNA complex remain incomplete. An initial X-ray crystal structure of the gp32 core (ssDNA binding) domain complexed to a dT₆ oligonucleotide revealed only weak electron density for the ssDNA lattice bound within the protein's OB-fold (Shamoo et al., 1995), suggesting that ssDNA is fairly mobile within the gp32 binding groove, allowing the protein to freely translocate (slide) along ssDNA (Jose et al., 2015a; Lee et al., 2016; Lohman & Kowalczykowski, 1981). However, the authors modeled four nucleotides of the dT₆ chain into the gp32 binding cleft, and the resulting structure suggested that at least two nucleotides were tightly bound within the protein's core. A more recent low-resolution crystal structure of gp32 in complex with the T4 Dda helicase and a dT₁₇ oligo (He et al., 2024) further defined the entire ssDNA binding surface of the gp32 monomer, as well as its interaction with Dda. Consistent with these structural studies, oligonucleotide-based binding measurements using proteolysis and DNA T_m depression methods demonstrated that at least 2–3 adjacent phosphodiester bonds are required for gp32-ssDNA binding (Wu et al., 1999). However, this study also showed an increase in gp32-ssDNA binding affinity when the oligos were increased in length from 5 to 8 nt, suggesting that the number of interactive residues within the core may be variable and dependent on substrate length.

Bulk studies of gp32 binding have often utilized relatively short ssDNA substrates, limiting measurements of gp32-ssDNA dynamics to either single noncontiguous monomers or small clusters thereof (Camel et al., 2021; Jose et al., 2015a; Jose et al., 2015b; Lee et al., 2016; Lohman & Kowalczykowski, 1981). However, the length of a typical Okazaki fragment in T4-infected *E. coli* is 1000–2000 nt (Maloy & Hughes, 2013) (i.e., can accommodate ~150–300 proteins), indicating that greater ssDNA lengths are likely needed for a complete understanding of gp32 filament structure and organizational dynamics *in vivo*. In this regard, single-molecule DNA stretching experiments are able to extend our understanding of gp32 behavior by probing its binding to long ssDNA substrates, which may accommodate >1000 proteins. Previous measurements on force-melted λ-phage DNA revealed how competing interactions of the acidic CTD for access to the protein's OB-groove regulate its salt-dependent binding to ssDNA (Pant et al., 2004; Pant et al., 2005; Pant et al., 2003; Rouzina et al., 2005). These studies also helped explain the origin of the “kinetic block” to

dsDNA helix-destabilization (melting) by full-length gp32 that was observed in thermal melting experiments.

Similar to *EcSSB*, gp32 has the seemingly paradoxical requirement to both stably bind and protect regions of ssDNA transiently exposed during replication while also ensuring their rapid release upon synthesis of the complementary strand. While gp32's high-affinity binding facilitates efficient coating of the discrete Okazaki fragments, such stable and highly cooperative binding could prevent the protein from being easily displaced from its ssDNA template. Indeed, previous stopped-flow measurements revealed that gp32 primarily dissociates from the ends of its cooperative clusters and that the rate of unbinding is too slow to account for the observed rate of DNA synthesis by T4 polymerase (Lohman, 1984a; Lohman, 1984b). Efficient protein recycling during DNA replication remains an important, open question for all SSBs.

LINE-1 (L1) is an intragenomic parasitic DNA element, comprising ~20% of the human genome, that amplifies within its host through a “copy-paste” mechanism known as retrotransposition (Furano, 2000; Goodier et al., 2013; Kazazian & Moran, 2017; Lander et al., 2001). L1 encodes two proteins, ORF1p and ORF2p, which assemble on their encoding transcript (*cis* preference) to form the L1 ribonucleoprotein (RNP), an essential intermediate of retrotransposition (Doucet et al., 2010; Howell & Usdin, 1997; Kulpa & Moran, 2005; Kulpa & Moran, 2006; Martin, 1991; Martin, 2010; Moran et al., 1996; Sahakyan et al., 2017). ORF2p provides reverse transcriptase and endonuclease activity (Feng et al., 1996; Luan et al., 1993; Mathias et al., 1991; Miller et al., 2021; Moran et al., 1996; Thawani et al., 2024). ORF1p, the major component of the L1 RNP, is a homotrimeric phosphoprotein that binds single-stranded nucleic acid (ssNA) nonspecifically with high affinity and exhibits NA chaperone activity (i.e., facilitates annealing and exchange of NA strands).

ORF1p contains a 51 amino acid intrinsically disordered NTD, which harbors two highly conserved phosphorylation sites necessary for retrotransposition (Cook et al., 2015; Furano & Cook, 2016), followed by a 14-heptad coiled-coil (CC), which mediates the trimerization of ORF1p monomers (Boissinot & Sookdeo, 2016; Callahan et al., 2012; Khazina et al., 2011; Khazina & Weichenrieder, 2018; Martin et al., 2003). The ORF1p coiled coil is evolutionarily labile (subject to rampant amino acid substitutions) (Furano et al., 2020). However, despite such variability, mutational analysis has shown that ORF1p activity is exquisitely sensitive to its CC sequence (Adney et al., 2019; Goodier et al., 2007; Naufer et al., 2016), suggesting that the persistence of L1 activity requires periodic remodeling of the ORF1p coiled coil. In contrast, the carboxy-terminal half is highly conserved and comprises two domains: a noncanonical RNA recognition motif (RRM) (Khazina & Weichenrieder, 2009), which contains two additional phosphorylation sites required for retrotransposition, and a CTD, which terminates in a 46 amino acid intrinsically disordered sequence.

Residues within the RRM and CTD endow ORF1p with high-affinity ssNA binding and NA chaperone activity *in vitro*. However, these properties are only evident in the context of the trimer (Basame et al., 2006; Callahan et al., 2012; Januszyn et al., 2007; Khazina et al., 2011; Khazina & Weichenrieder, 2009; Kolosha & Martin, 2003; Kulpa & Moran, 2005; Martin, 2010; Martin et al., 2003; Martin et al., 2008; Martin et al., 2005; Martin et al., 2000; Moran et al., 1996). Mutations in the RRM or CTD domains that eliminate NA chaperone activity also abolish retrotransposition, suggesting a primary role of ORF1p chaperone activity in L1 replication (Martin et al., 2005). However, the mechanistic details of this activity are not known. FRET-based assays showed that

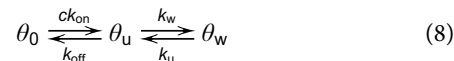
ORF1p can stabilize mismatched oligonucleotide duplexes (Callahan *et al.*, 2012), which are likely to be encountered during the hybridization of the target-site DNA and L1 transcript to generate a productive primer for cDNA synthesis by ORF2p.

Review of single-molecule experiments

Measuring multimode binding

Recently published work using similar OT single molecule techniques for three different protein systems observed strong evidence that more than one binding state must be present (Cashen *et al.*, 2023; Cashen *et al.*, 2024a; Cashen *et al.*, 2022; Cashen *et al.*, 2024b; Nauffer *et al.*, 2021). In particular, incubation experiments show a non-monotonic extension change response, both in terms of extension over time and equilibrium extension change (Figure 3a–c). Low protein concentrations (order of 1 nM) result in incubation curves that resemble simple bimolecular binding, as the ssDNA's extension decreases monotonically before approaching a highly compact equilibrium. However, while increasing protein concentration does increase the rate of initial compaction, the final equilibrium compaction is reduced as the extension begins to increase over time upon reaching a minimum value. Thus, to analyze the kinetics of the incubation process, the data must be split into two regimes (Figure 3d). First, the ssDNA extension decreases as the binding proteins wrap and compact the substrate. Eventually, the ssDNA must be sufficiently

saturated such that in order to accommodate additional protein, already bound protein must decrease its binding site size by switching to a less wrapped, and less compacted, state. As a result, both rates (initial compaction and subsequent elongation) increase with free protein concentration, but the secondary binding rate is an order of magnitude slower. This discrepancy in rate can be explained by the energetic barrier of partially unwrapping already bound protein to accommodate additional protein into the saturated complex, which is not a necessary step when the ssNA is initially protein-free. The equilibrium ssDNA extension at equilibrium can also be interpreted as a competition between a more and a less compacted binding state by extending Eq. (2) into a two-step, three-state reaction (Nauffer *et al.*, 2021).



In this scheme, the binding and wrapping of ssNA by protein are separated into two distinct steps. If the maximally and minimally compacted ssDNA extensions observed are associated with full occupancy of the wrapped (θ_w) and unwrapped state (θ_u), respectively, then the occupancy of both states can be calculated for any intermediate extension through interpolation (Figure 3e). Such analysis shows a smooth transition between the two states as a function of protein concentration that can be reproduced by simulating the three-state reaction in Eq. (8) or by approximating the interconversion between the states as a simple binding isotherm.

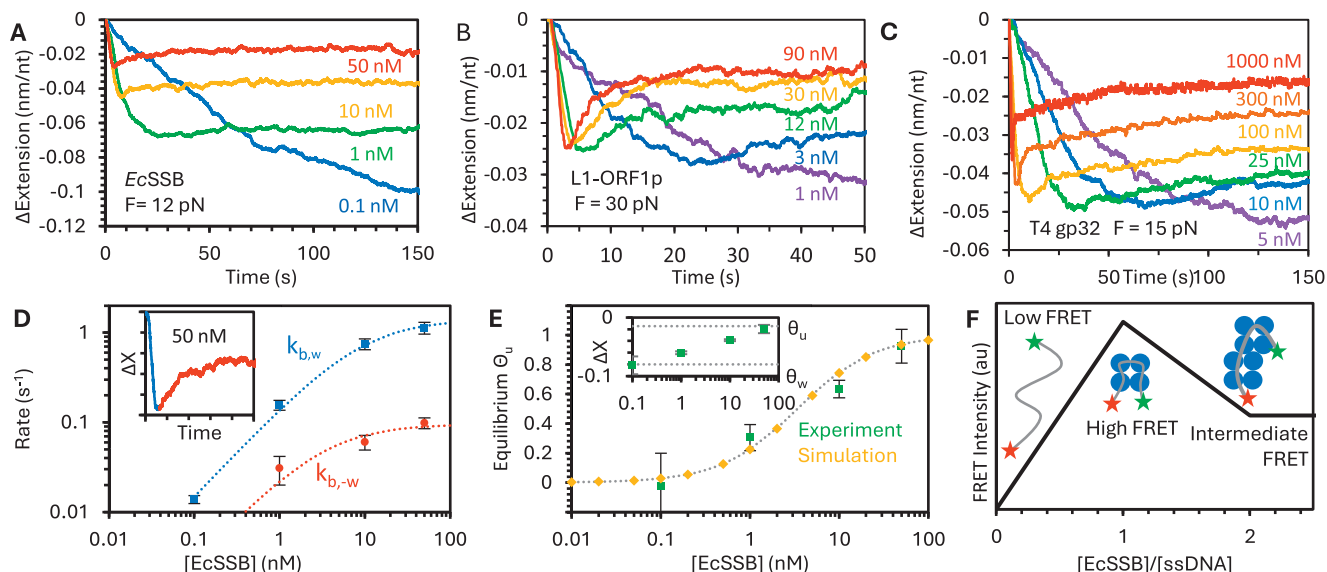


Figure 3. Non-monotonic response to protein concentration. (a) An ssDNA molecule is held at constant force (12 pN) while EcSSB is flowed into the sample, which binds and wraps the ssDNA, resulting in a decrease in extension. At low concentration (0.1 nM), the extension decreases monotonically, exponentially approaching an equilibrium value, consistent with simple concentration-dependent diffusion limited bi-molecular binding. Increasing protein concentration increases the initial rate of extension change, but ssDNA extension does not continue to decrease to the same value, and instead abruptly begins extending, equilibrating to a less compact conformation. This biphasic pattern is consistent with the protein initially wrapping the substrate to maximize binding site-NA contacts (largest binding site size), and then partially unwrapping to decrease binding site size and allow for additional protein binding. The same trend is observed for L1-ORF1p (b) and T4 gp32 (c). Note, that while the exact force used in experiments alters the absolute extension changes and kinetic rates, this trend is observed over a range of forces for each protein system. (d) The initial rate of ssDNA compaction (blue) and the secondary rate of ssDNA elongation (red) both increase with protein concentration (reflecting diffusion-limited bi-molecular binding), but the second step rate is an order of magnitude slower due to additional protein having to compete with and partially displace already bound protein. The inset shows 50 nM data from panel A with two phases marked. (e) The equilibrium extension change as a function of EcSSB concentration (inset) is converted to a measure of occupancy of the maximally and minimally compacted wrapping states. The occupancy of the unwrapped state (θ_u) increases with protein concentration, similar to a standard binding isotherm (dotted line), and can also be reproduced by numerically simulating a three-state model (yellow diamonds). (f) An additional experimental system that directly measures NA conformation, rather than protein binding itself, is FRET measurement with coupled dyes located at either end of the binding substrate (Roy *et al.*, 2007). When the pair of dyes are separated by a free 70 nt long poly dT ssDNA, little FRET intensity is observed. When the ssDNA is bound by a single EcSSB tetramer, the exact structure of the ssDNA wrapping places the labels in close proximity, resulting in high FRET intensity. Increasing the protein to ssDNA ratio to above 1:1, however, results in two tetramers simultaneously binding the substrate, each in a reduced binding site size state (35 nt), moving the dyes further from each other and decreasing FRET efficiency to an intermediate value.

The analysis shown here for *EcSSB* (Naufer et al., 2021) was also used to separate the multiple binding states and kinetic steps for L1-ORF1p (Cashen et al., 2022; Cashen et al., 2024b) and T4 gp32 (Cashen et al., 2023; Cashen et al., 2024a). One advantage of using *EcSSB* as a model system is that its ability to interconvert between different wrapping states with different binding site sizes has been well-documented using many different experimental assays. For example, one assay that also directly measures the conformation of an NA substrate (rather than directly measuring protein bound) consists of a FRET pair of dyes on both ends of a binding substrate (Roy et al., 2007)(Figure 3f). When no protein is present, the ends are effectively uncoupled, leading to low FRET intensity. For *EcSSB*, when one protein tetramer binds a 70 dT substrate, the two ends are brought together, resulting in high FRET efficiency. In contrast, if two tetramers bind the ssNA simultaneously (each occupying 35 nt), the resulting structure places the two dyes further apart. As a result, *EcSSB* titration experiments with this assay also return a non-monotonic response with concentration, similar to the OT experiments described above. In contrast, any experiment that directly measures bound protein (such as directly fluorescently labeling the protein) will instead simply measure a continuous increase in signal as free protein concentration is increased, obfuscating conformational transitions of the protein-NA complex.

Evidence of facilitated dissociation

Removing protein from the sample allows for the direct observation of protein dissociation from the ssNA substrate. One issue with working with proteins with very strong binding affinity, however, is that the binding to the substrate may be too stable to observe

measurable dissociation on experimental timescales. For *EcSSB* binding to ssDNA held at low force, removing free protein does not result in significant ssDNA extension change (Naufer et al., 2021) (Figure 4a). Subsequently increasing the *EcSSB* concentration, however, does result in a change in ssDNA extension. Compaction is reduced, and the ssDNA equilibrates to the same length as when the substrate is initially incubated with the same high protein concentration. In contrast to the equilibrium reached when the ssDNA is incubated with low *EcSSB* concentration, this less compact state is very unstable, and removing free protein at this point causes the ssDNA to quickly recompact. After reaching the same highly compact state observed during the initial incubation with low protein concentration, the system becomes stable again and no further extension change is observed. These result demonstrate that even under conditions where *EcSSB* does not fully dissociate (protein-free ssDNA is never recovered), the protein-NA complex must be able to rapidly reorganize based on the local density of protein present.

If the ssDNA tension is increased, binding is further destabilized such that full protein dissociation can be observed. That is, eventually the ssDNA returns to its original protein-free conformation (Figure 4b). However, even under these conditions, the ssDNA-*EcSSB* complex first rapidly recompacts when free protein is removed before slowly elongating as the rest of the protein dissociates. The same effect is again observed for L1-ORF1p (Cashen et al., 2022; Cashen et al., 2024b) (Figure 4c) and T4 gp32 as well (Cashen et al., 2023; Cashen et al., 2024a) (Figure 4d).

Similar to the incubation experiments, the dissociation data must be split into two distinct regimes. The rapid reorganization that results from protein re-wrapping the ssDNA when excess

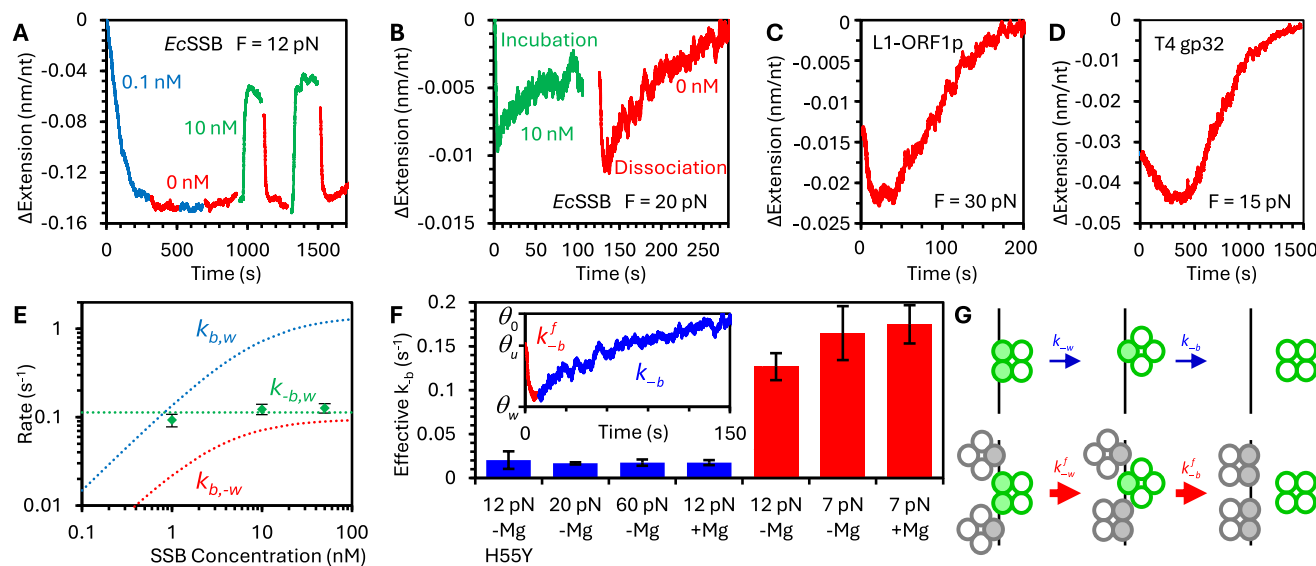


Figure 4. Facilitated dissociation of over-saturated protein. (a) After ssDNA has been incubated with protein, the free protein concentration is suddenly changed to measure the re-equilibration of the complex. At low forces, *EcSSB* binding is very stable, and little dissociation is observed on the ~100 s timescale. However, subsequently, increasing protein concentration reduces the decrease in extension, similar to when the ssDNA is initially incubated with high protein concentration (Figure 3a). When free protein is removed after oversaturation, the ssDNA rapidly recompacts, consistent with facilitated dissociation of excess protein to allow remaining bound proteins to return to their most wrapped, stably bound state. (b) At higher forces, *EcSSB* wrapping is destabilized, allowing measurable protein dissociation, as the ssDNA returns to its protein-free state. However, the initial dissociation of excess (oversaturating) protein occurs much faster than the final full dissociation, consistent with a mechanism of facilitated dissociation. Similar trends are observed for L1-ORF1p (c) and T4 gp32 (d). (e) The rate of *EcSSB* dissociation that results in further wrapping of the ssDNA is constant with respect to protein concentration during incubation. Compared to the rates of binding, the dissociation rate crosses the initial binding rate at <1 nM but only approaches the secondary binding rate at >10 nM, consistent with the ssDNA becoming saturated at low protein concentration but requiring much higher concentrations to oversaturate. (f) While full protein dissociation is enhanced by increased force, salt, or certain protein mutations, this rate remains an order of magnitude slower than the fast initial rate of facilitated protein dissociation. (g) Full protein dissociation (top) is slow due to the energetic barrier of removing binding energy between the ssDNA substrate and the binding domains of *EcSSB*. For oversaturated substrates (bottom), binding contacts released during dissociation are replaced by other bound proteins, removing this energy barrier and facilitating dissociation.

protein is removed occurs at a constant rate regardless of the initial EcSSB concentration during incubation (Figure 4e). For EcSSB, when compared to the two rates of binding observed during incubation, this initial dissociation rate intersects the initial binding rate at <1 nM, consistent with low protein concentration able to fully saturate the ssDNA substrate. In contrast, the secondary binding rate asymptotes to the initial dissociation rate at a high concentration (>10 nM), consistent with the high EcSSB concentration needed for the less wrapped state to become dominant. The secondary dissociation rate, where the ssDNA extends back to its original protein-free length, is much slower and is only detectable under conditions that augment full dissociation, such as destabilizing binding using excess applied force or high salt buffer that screens electrostatic binding interactions (Figure 4f). The order of magnitude difference between these two dissociation rates indicates that some biophysical process must be stimulating protein dissociation from this oversaturated state. This can be explained by the presence of multiple binding modes with different effective binding site sizes. If the ssDNA is protein oversaturated (bound protein is in its lower binding site size state to accommodate excess protein), then when a single protein dissociates, neighboring proteins can switch to the higher binding site size state, effectively absorbing the released NA substrate (Figure 4g). This process, in which the ssDNA substrate released by a dissociating protein is reabsorbed by another protein, is referred to as facilitated dissociation (Erbaş & Marko, 2019). In contrast, when protein dissociates from an undersaturated substrate, protein-free NA is left behind. Thus, full dissociation has an additional energy barrier due to the loss of protein-NA binding energy, while facilitated dissociation is closer to an isoenergetic process in which the loss of the final ssDNA contact before a protein is released from the substrate is replaced by the analogous interaction with a neighboring protein.

Protein structure and function

The ssDNA interactions exhibited by all three proteins discussed here can be related to the common features of their complexes with

ssNA, despite clear differences in structure and multimerization (Figure 5a). EcSSB and L1-ORF1p naturally form homotetramers and homotrimers, respectively. As a result, each homo-oligomer intrinsically has multiple binding domains. If each binding domain can bind ssNA semi-independently, with the substrate winding around the protein oligomer in different conformations to access the binding grooves of a discrete number of subunits, then the proteins inherently have the ability to alter their effective binding site size.

T4 gp32, conversely, is primarily monomeric in solution. However, interactions between the NTD and the core domains of neighboring proteins bound to an ssDNA substrate confer the protein with cooperative binding, resulting in the formation of long protein clusters on the substrate. Again, this results in many binding interfaces present on each homo-oligomeric filament, which must allow for the modulation of protein:ssDNA stoichiometry. Our results are consistent with the ssDNA substrate helically winding around the protein filament while remaining highly dynamic due to the ability to partially unwind to accommodate additional protein at high protein:ssDNA stoichiometries (Cashen *et al.*, 2023). Moreover, we found that critical protein oversaturation resulted in filament unwinding such that the cooperative interprotein interactions largely vanished, enabling rapid protein displacement from across the entire ssDNA substrate, relieving oversaturation.

However, the structural differences between these proteins confer different biophysical interactions with the ssNA that are measurable using OT. First, L1-ORF1p differs from the other proteins in that also exhibits an RNA packaging function mediated by inter-trimer interactions, in which the protein must stably bind and compact a copy of the L1 sequence. Correspondingly, when an L1-ORF1p saturated ssDNA is held at low force (≤ 5 pN), the substrate continues to compact over time (Figure 5b). Additionally, when left for increasingly long incubation times, high-force stretching of the ssDNA reveals permanent compaction. When WT L1-ORF1p is replaced with an inactive (retrotransposition-deficient) mosaic of modern and ancestral strains of L1, this secondary compaction ability is lost (Cashen *et al.*, 2022). As T4 gp32

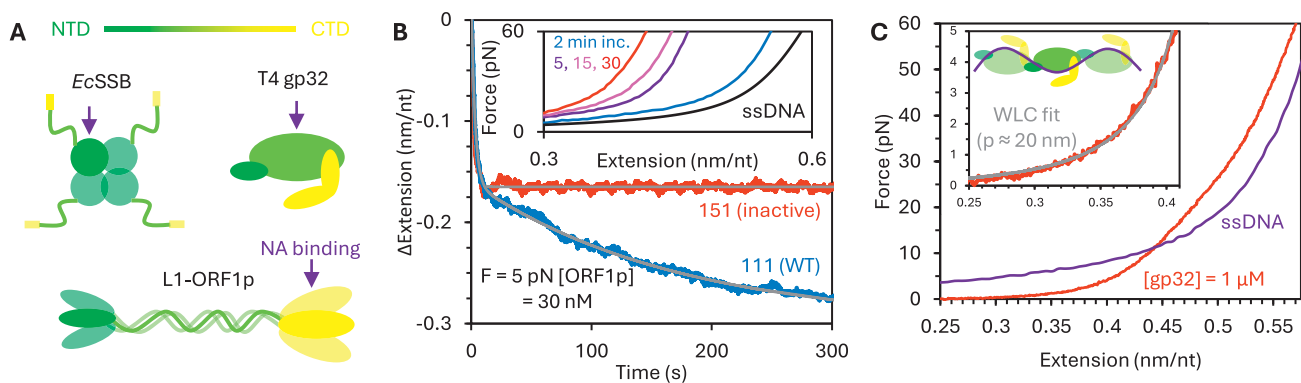


Figure 5. Comparison of protein structures. (a) EcSSB is a homotetramer mediated by its oligonucleotide binding domains (NTD, green). Each subunit has an intrinsically disordered tail (green-yellow) ending with an acidic tip at the C-terminus (yellow). L1-ORF1p forms trimers through its central coiled coil domain, with NA binding activity located in the RRM-CTD. T4 gp32 is a monomer in solution, but its NTD binds to the core domain of a neighboring protein to enable cooperative binding and oligomerization of NA-bound proteins. The T4 gp32 CTD competes with NA for access to the core binding domain and modulates binding in a salt-dependent manner. Purple arrows indicate the NA binding domain of each protein. (b) At very low forces, when ssDNA is not straightened, L1-ORF1p exhibits a secondary compaction phase that further reduces ssDNA extension after protein saturation. Stretching the ssDNA–protein complex after increasing incubation times (inset) confirms a stable, non-reversible shortening of the substrate. Certain protein variants that are deficient in retrotransposition activity lack this secondary compaction function. As this tight compaction is effectively irreversible, it would seemingly interfere with the protein’s SSB-like function during polymerization. However, tight NA compaction by L1-ORF1p is likely critical to its role in RNA packaging, a function not exhibited by the other studied proteins. (c) T4 gp32 forms long protein filaments on ssDNA, greatly reducing the force needed to straighten the substrate (or equivalently, increasing ssDNA extension at low force). Fitting the FEC of protein-saturated ssDNA to the WLC model returns an effective persistence length of ~ 20 nm, much longer than the length of a single protein, indicating that the interprotein interface is semi-rigid and preserves the relative orientation of neighboring proteins in the filament.

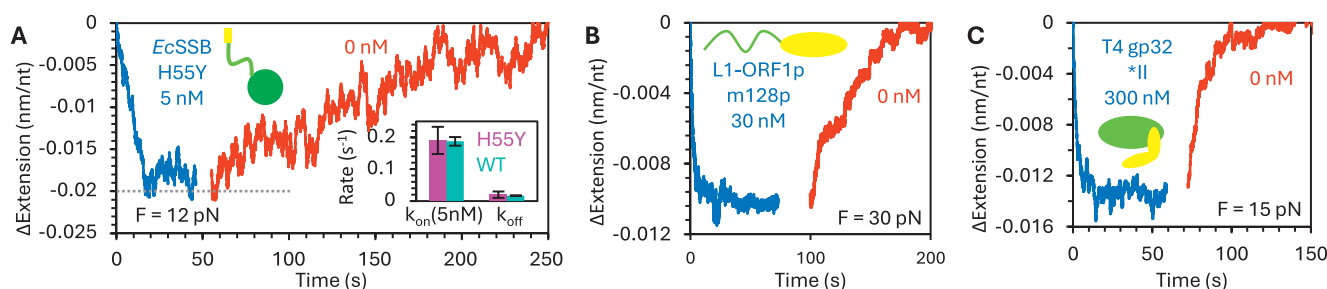


Figure 6. Monomeric, non-cooperative protein variants. (a) ssDNA is incubated with H55Y mutant *EcSSB*, which does not form protein tetramers. ssDNA extension decreases monotonically during incubation and increases monotonically when free protein is removed, returning to its initial protein-free state. The amplitude of compaction at saturation (gray dotted line) matches the extension of ssDNA over-saturated with WT protein and the rates of binding and dissociation match the rates of initial binding and full dissociation of the WT protein (inset). Monomeric protein variants m128p (truncated at 128th residue) L1-ORF1p (b) and *II (NTD removed) T4 gp32 (c) also display similar two-state, bimolecular, non-cooperative, reversible binding.

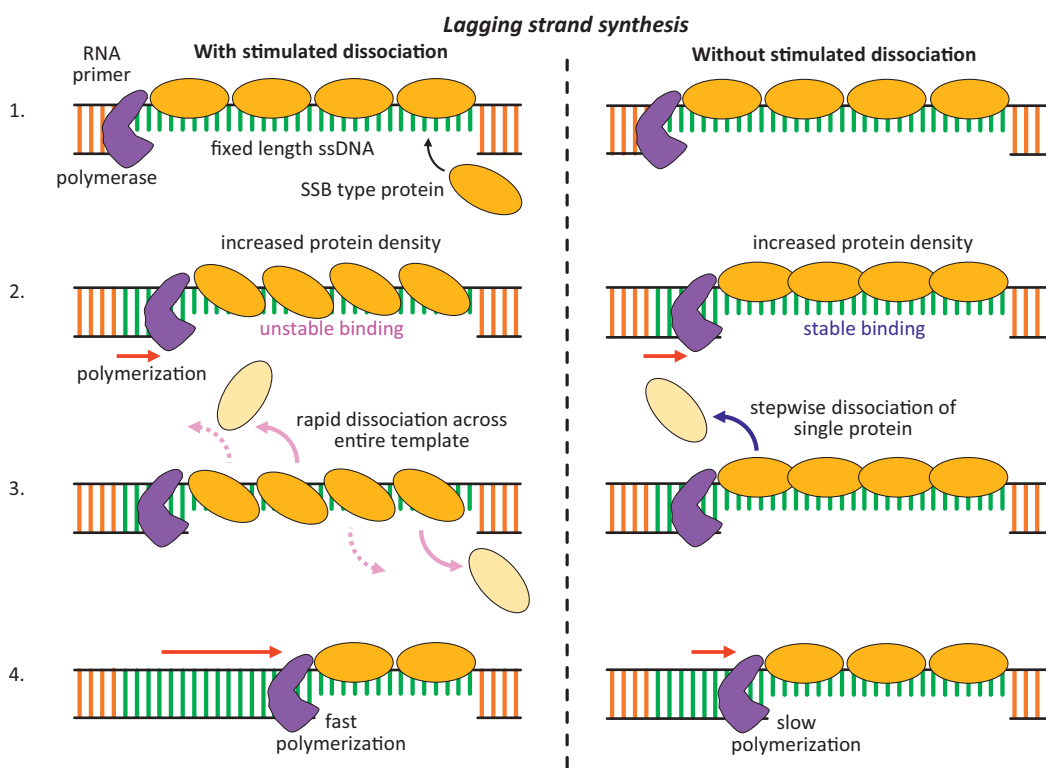


Figure 7. Biological impact of unstable, oversaturated protein binding state. A protein with high-affinity ssNA binding that is expressed in sufficient quantity will quickly saturate any transiently exposed ssNA regions that form during processes such as Okazaki fragments production during DNA replication. As a polymerase proceeds along the template strand, the ssNA region is reduced in length, effectively increasing protein density. If the proteins are able to switch to a lower binding site size and reduced binding free energy conformation (left side), facilitated dissociation is enabled in which any protein on the lattice can dissociate while neighboring proteins compensate for the loss of interaction free energy by absorbing any additional nucleotides released by switching back to a larger binding site size state. In contrast, if the protein remains in its most stable state (right side), in which the binding site size is maximized by utilizing the entire binding surface on each protein, then protein dissociation is inhibited. Instead, some other mechanism, such as direct interactions between the polymerase and the adjacent protein, must remove fully bound proteins in a stepwise, sequential manner, such that polymerization is rate limited by protein dissociation.

forms long, continuous protein filaments rather than discrete tetramers or trimers, the length scale of interprotein interaction is greatly increased relative to that observed for L1 ORF1p or *EcSSB*. In addition to decreasing the contour length of the protein-saturated ssDNA substrate, a large (~30-fold) increase in persistence length is also observed (Figure 5c). ssDNA bound with T4 gp32 has an effective persistence length of approximately 20 nm, which is much longer than the length scale of the protein itself (Cashen et al., 2023). Thus, the orientation of neighboring proteins in the filament must

be preserved, resulting in the protein–ssNA complex behaving like a semirigid structure.

Our hypothesis that the dynamic multistate binding of these proteins is related to their ability to multimerize is supported by comparison experiments with non-multimerizing protein variants that exhibit neither oversaturated protein binding nor facilitated dissociation. Instead, simple single-state binding is recovered (Figure 6). First, a point mutation in *EcSSB*'s OB domain (H55Y) prevents the formation of homotetramers in

solution but does not inhibit the binding of the domain to ssDNA (Figure 6a) (Naufer et al., 2021). As a result, the binding of this monomeric mutant results in minimal ssDNA compaction, similar in amplitude to oversaturating WT protein conditions, where wrapping is destabilized to accommodate additional protein binding. Additionally, the binding is completely reversible, and the protein immediately begins dissociating when free protein is removed, extending the ssDNA back to its protein-free conformation. However, the rates of initial binding and full dissociation are still consistent with the rates observed for WT protein. Similarly, truncation of the L1-ORF1p protein at the 128th residue (m128p), removing the NTD and 10.5 heptads of the 14-heptad coiled coil, prevents trimer formation while leaving the binding domains in the RRM-CTD intact (Cashen et al., 2024b). Again, compaction is reduced and single-phased, and binding is reversible (Figure 6b). Finally, truncation of T4 gp32 to remove the entire NTD domain responsible for cooperative binding has the same effect (Figure 6c) (Cashen et al., 2023). Besides removing multistate binding, the large increase in persistence length is no longer observed, but is instead consistent with the length scale of a single binding site, indicating that neighboring proteins are no longer associated with a continuous filament, but instead act independently of one another.

Biological function of variable conformation binding

The role of these proteins' multistate binding may be related to their need to perform the seemingly opposed functions of stably binding ssNA for protection while remaining dynamic enough to reorganize as NA processing proceeds. All three proteins bind lengths of ssNA that must be polymerized into ds form, ssDNA Okazaki fragments formed during lagging strand synthesis for EcSSB and T4 gp32, and during replication of the L1 RNA transcript for L1 ORF1p. When first formed, the ssNA region has a discrete length and the limited pool of binding proteins must be sufficient to fully saturate this length, and all other ssNA regions present at a given time (Figure 7). Under these conditions, it is beneficial that ssNA binding proteins can occupy and occlude as many nts as possible. For example, by fully wrapping around all available domains, 65 nt ssNA (~35 nm of linear length) can be fully occluded by a single EcSSB tetramer. However, for polymerization to proceed, these occluded nt must eventually be accessed by the polymerase enzyme. As the ssNA binding proteins do not naturally dissociate on a short timescale (as required by the protection function), the proteins must be removed by an active process. One candidate is a specific interaction between the polymerase and the binding proteins. Such interactions would be limited, however, to a single protein at the ds-ss junction and would have to proceed in a stepwise fashion. The ability of such strongly bound proteins to be removed one at a time in sufficiently rapid sequence as to allow efficient polymerization would appear difficult. However, the presence of multiple binding states alleviates this bottleneck. As polymerization proceeds, bound proteins can switch to a lower binding site state, giving the polymerase access to additional nts. Additionally, since these states have reduced contact with the ssNA substrate, binding is weakened, and dissociation is enabled for all proteins across the ssNA region. Finally, this mechanism also allows for facilitated dissociation, as observed in the OT experiments, where the dissociation of proteins from an oversaturated substrate is an order of magnitude faster than dissociation that leaves behind bare ssNA. Thus, the ssNA binding proteins can promptly reorganize and dissociate in front of the advancing polymerase so as to not delay DNA synthesis.

Conclusions

Our OT experiments shed new light on the mechanism of prompt protein dissociation in front of a moving DNA polymerase. We observed that ssNA oversaturation (crowding) with any of the three studied ssNA binding proteins leads to rapid non-cooperative dissociation of excess protein from along the ssNA template. This mechanism is enabled by the structure of the proteins and the oligomers they form, in which multiple binding interfaces are present, allowing proteins to interconvert between stable and unstable conformations with distinct dissociation kinetics. Competition between these binding interfaces, in which the NA substrate released from one site can be reabsorbed by another empty site, enables rapid protein reorganization and facilitated dissociation.

Open peer review. To view the open peer review materials for this article, please visit <http://doi.org/10.1017/qrd.2024.21>.

Data availability statement. All data available in previously published manuscripts and upon reasonable request from M.C.W.

Author contribution. M.M., B.A.C., I.R. wrote the original manuscript. M.M., B.A.C., I.R., and M.C.W. reviewed and edited the manuscript.

Financial support. This work was supported by the National Science Foundation [MCB-1817712 to M.C.W.]

Competing interest. The authors declare no conflicts of interest.

References

- Adney EM, et al. (2019) Comprehensive scanning mutagenesis of human Retrotransposon LINE-1 identifies motifs essential for function. *Genetics* 213(4), 1401–1414
- Alberts BM and Frey L (1970a) T4 bacteriophage gene 32: A structural protein in the replication and recombination of DNA. *Nature* 227(5265), 1313–1318
- Alberts BM and Frey L (1970b) T4 bacteriophage gene 32: A structural protein in the replication and recombination of DNA. *Nature* 227(5265), 1313–1318
- Antony E, et al. (2013) Multiple C-terminal tails within a single SSB homotetramer coordinate DNA replication and repair. *Journal of Molecular Biology* 425(23), 4802–4819
- Basame S, et al. (2006) Spatial assembly and RNA binding stoichiometry of a LINE-1 protein essential for retrotransposition. *Journal of Molecular Biology* 357(2), 351–357
- Baumann CG, et al. (1997) Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* 94(12), 6185–6190
- Boissinot S and Sookdeo A (2016) The evolution of LINE-1 in vertebrates. *Genome Biology and Evolution* 8(12), 3485–3507
- Bujalowski W and Lohman TM (1986) *Escherichia coli* single-strand binding protein forms multiple, distinct complexes with single-stranded DNA. *Biochemistry* 25(24), 7799–7802
- Bujalowski W, et al. (1988) Binding mode transitions of *Escherichia coli* single strand binding protein-single-stranded DNA complexes. Cation, anion, pH, and binding density effects. *Journal of Biological Chemistry* 263(10), 4629–4640
- Callahan KE, et al. (2012) Polymerization and nucleic acid-binding properties of human L1 ORF1 protein. *Nucleic Acids Research* 40(2), 813–827
- Camel BR, et al. (2021) Mapping DNA conformations and interactions within the binding cleft of bacteriophage T4 single-stranded DNA binding protein (gp32) at single nucleotide resolution. *Nucleic Acids Research* 49(2), 916–927
- Casas-Finet JR, et al. (1992) Structural basis for the nucleic acid binding cooperativity of bacteriophage T4 gene 32 protein: The (Lys/Arg)₃(Ser/Thr)₂ (LAST) motif. *Proceedings of the National Academy of Sciences of the United States of America* 89(3), 1050–1054
- Cashen BA, et al. (2023) Dynamic structure of T4 gene 32 protein filaments facilitates rapid noncooperative protein dissociation. *Nucleic Acids Research* 51(16), 8587–8605

- Cashen BA, et al. (2024a) C-terminal domain of T4 gene 32 protein enables rapid filament reorganization and dissociation. *Journal of Molecular Biology* **436**(9), 168544
- Cashen BA, et al. (2022) The L1-ORF1p coiled coil enables formation of a tightly compacted nucleic acid-bound complex that is associated with retrotransposition. *Nucleic Acids Research* **50**(15), 8690–8699
- Cashen BA, et al. (2024b) L1-ORF1p nucleoprotein can rapidly assume distinct conformations and simultaneously bind more than one nucleic acid. *Nucleic Acids Research*, **52**(22), 14013–14029
- Cook PR, et al. (2015) Phosphorylation of ORF1p is required for L1 retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America* **112**(14), 4298–4303
- Doucet AJ, et al. (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genetics* **6**(10), e1001150
- Erbas A and Marko JF (2019) How do DNA-bound proteins leave their binding sites? The role of facilitated dissociation. *Current Opinion in Chemical Biology* **53**, 118–124
- Feng Q, et al. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**(5), 905–916
- Furano AV (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Progress in Nucleic Acid Research and Molecular Biology* **64**, 255–294
- Furano AV and Cook PR (2016) The challenge of ORF1p phosphorylation: Effects on L1 activity and its host. *Mobile Genetic Elements* **6**(1), e1119927
- Furano AV, et al. (2020) Cryptic genetic variation enhances primate L1 retrotransposon survival by enlarging the functional coiled coil sequence space of ORF1p. *PLoS Genetics* **16**(8), e1008991
- Goodier JL, et al. (2013) Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Research* **41**(15), 7401–7419
- Goodier JL, et al. (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Molecular and Cellular Biology* **27**(18), 6469–6483
- Grieb MS, et al. (2017) Dynamic stepwise opening of integron *attC* DNA hairpins by SSB prevents toxicity and ensures functionality. *Nucleic Acids Research* **45**(18), 10555–10563
- Griffith J and Formosa T (1985) The Uvsx protein of bacteriophage-T4 arranges single-stranded and double-stranded DNA into similar helical nucleoprotein filaments. *Journal of Biological Chemistry* **260**(7), 4484–4491
- He X, et al. (2024) Structural and functional insights into the interaction between the bacteriophage T4 DNA processing proteins gp32 and Dda. *Nucleic Acids Research* **52**(20), 12748–12762
- Howell R and Usdin K (1997) The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3' end of L1 retrotransposons. *Molecular Biology and Evolution* **14**(2), 144–155
- Januszyk K, et al. (2007) Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *Journal of Biological Chemistry* **282**(34), 24893–24904
- Jose D, et al. (2015a) Mapping the interactions of the single-stranded DNA binding protein of bacteriophage T4 (gp32) with DNA lattices at single nucleotide resolution: Polynucleotide binding and cooperativity. *Nucleic Acids Research* **43**(19), 9291–9305
- Jose D, et al. (2015b) Mapping the interactions of the single-stranded DNA binding protein of bacteriophage T4 (gp32) with DNA lattices at single nucleotide resolution: gp32 monomer binding. *Nucleic Acids Research* **43**(19), 9276–9290
- Karpel, R. L. (1990). T4 bacteriophage gene 32 protein. *The Biology of Non-specific DNA Protein Interactions*. CRC Press, Inc., Boca Raton, Fla, 103–130
- Kazazian HH and Moran JV (2017) Mobile DNA in health and disease. *New England Journal of Medicine* **377**(4), 361–370
- Khazina E, et al. (2011) Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nature Structural & Molecular Biology* **18**(9), 1006–1014
- Khazina E and Weichenrieder O (2009) Non-LTR retrotransposons encode non-canonical RRM domains in their first open reading frame. *Proceedings of the National Academy of Sciences of the United States of America* **106**(3), 731–736
- 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**, e34960
- Kolosha VO and Martin SL (2003) High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *Journal of Biological Chemistry* **278**(10), 8112–8117
- Kozlov AG and Lohman TM (2002a) Kinetic mechanism of direct transfer of SSB tetramers between single-stranded DNA molecules. *Biochemistry* **41**(39), 11611–11627
- Kozlov AG and Lohman TM (2002b) Stopped-flow studies of the kinetics of single-stranded DNA binding and wrapping around the *Escherichia coli* SSB tetramer. *Biochemistry* **41**(19), 6032–6044
- Kozlov AG, et al. (2019) Regulation of nearest-neighbor cooperative binding of *E. coli* SSB protein to DNA. *Biophysical Journal* **117**(11), 2120–2140
- Kozlov AG, et al. (2015) Intrinsically disordered C-terminal tails of *E. coli* single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA. *Journal of Molecular Biology* **427**(4), 763–774
- Krassa KB, et al. (1991) Protein-protein interactions with the acidic COOH terminus of the single-stranded DNA-binding protein of the bacteriophage T4. *Proceedings of the National Academy of Sciences, USA* **88**(9), 4010–4014
- Krauss G, et al. (1981) *Escherichia coli* single-strand deoxyribonucleic acid binding protein: Stability, specificity, and kinetics of complexes with oligonucleotides and deoxyribonucleic acid. *Biochemistry* **20**(18), 5346–5352
- Kuil ME, et al. (1990) A refined calculation of the solution dimensions of the complex between gene 32 protein and single stranded DNA based on estimates of the bending persistence length. *Journal of Biomolecular Structure and Dynamics* **7**(4), 943–957
- Kuil ME, et al. (1988) The internal dynamics of gene 32 protein-DNA complexes studied by quasi-elastic light scattering. *Biophysical Chemistry* **32**(2–3), 211–227
- Kulpa DA and Moran JV (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human Molecular Genetics* **14**(21), 3237–3248
- Kulpa DA and Moran JV (2006) Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nature Structural & Molecular Biology* **13**(7), 655–660
- Kunzelmann S, et al. (2010) Mechanism of interaction between single-stranded DNA binding protein and DNA. *Biochemistry* **49**(5), 843–852
- Kuznetsov SV, et al. (2006) Microsecond dynamics of protein-DNA interactions: Direct observation of the wrapping/unwrapping kinetics of single-stranded DNA around the *E. coli* SSB tetramer. *Journal of Molecular Biology* **359**(1), 55–65
- Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**(6822), 860–921
- Lee MH, et al. (2004) Self-polymerization of archaeal RadA protein into long and fine helical filaments. *Biochemical and Biophysical Research Communications* **323**(3), 845–851
- Lee W, et al. (2016) Single-molecule FRET studies of the cooperative and non-cooperative binding kinetics of the bacteriophage T4 single-stranded DNA binding protein (gp32) to ssDNA lattices at replication fork junctions. *Nucleic Acids Research* **44**(22), 10691–10710
- Lefebvre SD, et al. (1999) Simultaneous interactions of bacteriophage T4 DNA replication proteins gp59 and gp32 with single-stranded (ss) DNA. Co-modulation of ssDNA binding activities in a DNA helicase assembly intermediate. *Journal of Biological Chemistry* **274**(32), 22830–22838
- Lohman TM (1984a) Kinetics and mechanism of dissociation of cooperatively bound T4 gene 32 protein-single-stranded nucleic acid complexes. 1. Irreversible dissociation induced by sodium chloride concentration jumps. *Biochemistry* **23**(20), 4656–4665
- Lohman TM (1984b) Kinetics and mechanism of dissociation of cooperatively bound T4 gene 32 protein-single-stranded nucleic acid complexes. 2. Changes in mechanism as a function of sodium chloride concentration and other solution variables. *Biochemistry* **23**(20), 4665–4675
- Lohman TM, et al. (1988) Interactions of the *E. coli* single strand binding (SSB) protein with ss nucleic acids. Binding mode transitions and equilibrium binding studies. *Biochemical Pharmacology* **37**(9), 1781–1782

- Lohman TM and Kowalczykowski SC** (1981) Kinetics and mechanism of the association of the bacteriophage T4 gene 32 (helix destabilizing) protein with single-stranded nucleic acids. Evidence for protein translocation. *Journal of Molecular Biology* **152**(1), 67–109
- Lohman TM and Overman LB** (1985) Two binding modes in *Escherichia coli* single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration. *Journal of Biological Chemistry* **260**(6), 3594–3603
- Lohman TM**, et al. (1986) Salt-dependent changes in the DNA binding cooperativity of *Escherichia coli* single strand binding protein. *Journal of Molecular Biology* **187**(4), 603–615
- Lonberg N**, et al. (1981) Interactions of bacteriophage T4-coded gene 32 protein with nucleic acids. III. Binding properties of two specific proteolytic digestion products of the protein (G32P**I* and G32P**III*). *Journal of Molecular Biology* **145**(1), 123–138
- Luan DD**, et al. (1993) Reverse transcription of R2bm Rna is primed by a Nick at the chromosomal target site - a mechanism for non-Ltr Retrotransposition. *Cell* **72**(4), 595–605
- Maloy S and Hughes K** (2013) *Brenner's Encyclopedia of Genetics*. Academic Press, Cambridge, MA.
- Martin SL** (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and Cellular Biology* **11**(9), 4804–4807
- Martin SL** (2010) Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1. *RNA Biology* **7**(6), 706–711
- Martin SL**, et al. (2003) Trimeric structure for an essential protein in L1 retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America* **100**(24), 13815–13820
- Martin SL**, et al. (2008) A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. *Nucleic Acids Research* **36**(18), 5845–5854
- Martin SL**, et al. (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *Journal of Molecular Biology* **348**(3), 549–561
- Martin SL**, et al. (2000) Deletion analysis defines distinct functional domains for protein–protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *Journal of Molecular Biology* **304**(1), 11–20
- Mathias SL**, et al. (1991) Reverse transcriptase encoded by a human transposable element. *Science* **254**(5039), 1808–1810
- Miller I**, et al. (2021) Structural dissection of sequence recognition and catalytic mechanism of human LINE-1 endonuclease. *Nucleic Acids Research* **49**(19), 11350–11366
- Moran JV**, et al. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* **87**(5), 917–927
- Morrill SW**, et al. (1996) The gene 59 protein of bacteriophage T4. Characterization of protein–protein interactions with gene 32 protein, the T4 single-stranded DNA binding protein. *Journal of Biological Chemistry* **271**(33), 20198–20207
- Morse M**, et al. (2019) HIV restriction factor APOBEC3G binds in multiple steps and conformations to search and deaminate single-stranded DNA. *eLife* **8**, e52649
- Naufer MN**, et al. (2016) L1 retrotransposition requires rapid ORF1p oligomerization, a novel coiled coil-dependent property conserved despite extensive remodeling. *Nucleic Acids Research* **44**(1), 281–293
- Naufer MN**, et al. (2021) Multiprotein SSB ssDNA complex shows both stable binding and rapid dissociation due to interprotein interactions. *Nucleic Acids Research* **49**(3), 1532–1549
- Nelson SW**, et al. (2008) RNA primer handoff in bacteriophage T4 DNA replication: The role of single-stranded DNA-binding protein and polymerase accessory proteins. *Journal of Biological Chemistry* **283**(33), 22838–22846
- Odijk T** (2002) Stiff chains and filaments under tension. *Macromolecules* **28**(20), 7016–7018
- Pant K**, et al. (2004) Mechanical measurement of single-molecule binding rates: Kinetics of DNA helix-destabilization by T4 gene 32 protein. *Journal of Molecular Biology* **336**(4), 851–870
- Pant K**, et al. (2005) Salt dependent binding of T4 gene 32 protein to single and double-stranded DNA: Single molecule force spectroscopy measurements. *Journal of Molecular Biology* **349**(2), 317–330
- Pant K**, et al. (2003) Kinetic regulation of single DNA molecule denaturation by T4 gene 32 protein structural domains. *Journal of Molecular Biology* **327**(3), 571–578
- Raghunathan S**, et al. (2000) Structure of the DNA binding domain of *E. Coli* SSB bound to ssDNA. *Nature Structural Biology* **7**(8), 648–652
- Raghunathan S**, et al. (1997) Crystal structure of the homo-tetrameric DNA binding domain of *Escherichia coli* single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-Å resolution. *Proceedings of the National Academy of Sciences of the United States of America* **94**(13), 6652–6657
- Rouzina I**, et al. (2005) Theory of electrostatically regulated binding of T4 gene 32 protein to single- and double-stranded DNA. *Biophysical Journal* **89**(3), 1941–1956
- Roy R**, et al. (2007) Dynamic structural rearrangements between DNA binding modes of *E. Coli* SSB protein. *Journal of Molecular Biology* **369**(5), 1244–1257
- Roy R**, et al. (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature* **461**(7267), 1092–1097
- Sahakyan AB**, et al. (2017) G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nature Structural & Molecular Biology*, **24**(3), 243–247.
- Scheerhagen MA**, et al. (1989) A model for the complex between the helix destabilizing protein GP32 of bacteriophage T4 and single-stranded DNA. *Journal of Biomolecular Structure and Dynamics* **6**(4), 701–706
- Scheerhagen MA**, et al. (1985a) Hydrodynamic studies of a DNA-protein complex. Dimensions of the complex of single-stranded 145 base DNA with gene 32 protein of phage T4 deduced from quasi-elastic light scattering. *FEBS Letters* **184**(2), 221–225
- Scheerhagen MA**, et al. (1985b) Hydrodynamic studies of a DNA-protein complex. Elongation of single stranded nucleic acids upon complexation with the gene 32 protein of phage T4 deduced from electric field-induced birefringence experiments. *FEBS Letters* **179**(2), 221–224
- Shamoo Y**, et al. (1995) Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. *Nature* **376**(6538), 362–366
- Smith SB**, et al. (1996) Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**(5250), 795–799
- Suksombat S**, et al. (2015) Structural dynamics of *E. Coli* single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *eLife* **4**
- Takahashi M and Norden B** (1994) Structure of RecA–DNA complex and mechanism of DNA strand exchange reaction in homologous recombination. *Advances in Biophysics* **30**, 1–35
- Tan HY**, et al. (2017) The intrinsically disordered linker of *E. Coli* SSB is critical for the release from single-stranded DNA. *Protein Science* **26**(4), 700–717
- Thawani A**, et al. (2024) Template and target-site recognition by human LINE-1 in retrotransposition. *Nature* **626**(7997), 186–193
- Theobald DL**, et al. (2003) Nucleic acid recognition by OB-fold proteins. *Annual Review of Biophysics and Biomolecular Structure* **32**(1), 115–133
- van Amerongen H**, et al. (1990) Structure calculations for single-stranded DNA complexed with the single-stranded DNA binding protein GP32 of bacteriophage T4: A remarkable DNA structure. *Biochemistry* **29**(23), 5619–5625
- Wu M**, et al. (1999) Details of the nucleic acid binding site of T4 gene 32 protein revealed by proteolysis and DNA tm depression methods. *Journal of Molecular Biology* **286**(4), 1107–1121
- Wu Y**, et al. (2004) Crystal structure of archaeal recombinase RADA: A snapshot of its extended conformation. *Molecular Cell* **15**(3), 423–435
- Xu J**, et al. (2017) Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange. *Nature Structural & Molecular Biology* **24**(1), 40–46
- Yang SX**, et al. (2001) Comparison of bacteriophage T4 UvsX and human Rad51 filaments suggests that RecA-like polymers may have evolved independently. *Journal of Molecular Biology* **312**(5), 999–1009
- Yu X**, et al. (2004) What is the structure of the RecA–DNA filament? *Current Protein and Peptide Science* **5**(2), 73–79