Computational and Structural Biotechnology Journal 19 (2021) 3765–3778

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Review

DNA replication: *In vitro* single-molecule manipulation data analysis and models

Javier Jarillo [a], Borja Ibarra [b], Francisco Javier Cao-García [b,c,*]

[a] University of Namur, Institute of Life-Earth-Environment, Namur Center for Complex Systems, Rue de Bruxelles 61, 5000 Namur, Belgium
[b] Instituto Madrileño de Estudios Avanzados en Nanociencia, IMDEA Nanociencia, C/ Faraday 9, 28049 Madrid, Spain
[c] Departamento de Estructura de la Materia, Física Térmica y Electrónica, Universidad Complutense de Madrid, Pza. de Ciencias, 1, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 15 March 2021
Received in revised form 18 June 2021
Accepted 21 June 2021
Available online 24 June 2021

Keywords:
DNA replication
DNA unwinding
DNA polymerase
Helicase
Single-molecule
Real-time kinetics

ABSTRACT

DNA replication is a key biochemical process of the cell cycle. In the last years, analysis of *in vitro* single-molecule DNA replication events has provided new information that cannot be obtained with ensembles studies. Here, we introduce crucial techniques for the proper analysis and modelling of DNA replication *in vitro* single-molecule manipulation data. Specifically, we review some of the main methods to analyze and model the real-time kinetics of the two main molecular motors of the replisome: DNA polymerase and DNA helicase. Our goal is to facilitate access to and understanding of these techniques to promote their use in the study of DNA replication at the single-molecule level. A proper analysis of single-molecule data is crucial to obtain a detailed picture of, among others, the kinetics rates, equilibrium contants and conformational changes of the system under study. The techniques presented here have been used or can be adapted to study the operation of other proteins involved in nucleic acids metabolism.

1. Introduction

DNA is the biological polymer carrying the genetic instructions for life. DNA replication (or duplication) is an essential part for biological inheritance, which ensures that upon cell division the two new daughter cells contain the same genetic information as the parent cell [10]. DNA is made up of a double helix of two complementary strands that are replicated synchronously, in a process known as semiconservative replication [63]. This process implies that each strand of the parental DNA molecule serves as a template to produce its complementary counterpart. A complex and highly dynamic protein machinery, referred to as the replisome, is in charge of robust, and accurate DNA replication needed for cell survival [6,39]. Fig. 1. The core components of prokaryotic and eukaryotic replisomes are replicative DNA polymerases, the helicase-primase and the single-stranded DNA binding proteins (SSBs). Depending on the organism, a variety of other proteins associate transiently and coordinate their activities with the core elements to carry out DNA replication [10,34]. Replicative DNA polymerases synthetize the new complementary strand of DNA by the stepwise addition of the corresponding complementary nucleotide (dNTPs) [94]. These enzymes are designed to maintain low mutation rates; they incorporate one wrong nucleotide per $10^4 - 10^5$ nucleotides polymerized. This fidelity is further enhanced by a factor of 10^2-10^3 by their associated exonuclease activities, which hydrolyze the mismatched nucleotide from the 3′ end of the hairpin [53,5]. In addition, many replicative DNA polymerases, present an intrinsic strand displacement activity (the ability to displace downstream DNA encountered during synthesis, without the help of a helicase) [15]. Replicative helicases form ring-shaped structures that encircle one (or two) of the DNA strand(s) and utilize the chemical energy of NTPs to unwind the DNA fork in coordination with the DNA polymerase [83,62,82]. Fig. 1. The SSB proteins bind with high affinity to single-stranded DNA and constitute the nucleo-protein complex upon with the other replisome components work [93,31]. Replicative DNA polymerases and helicases work as molecular motors that harness chemical and thermal energies to generate unidirectional mechanical motion. All molecular motors operate at energies comparable to those of the thermal fluctuations [50,13]. Therefore, they experience continuous agitation by random Brownian motion, which is eventually reflected in fluctuations in their real-time kinetics. Biological molecular motors have evolved to take advantage of these Brownian fluctuations, which they couple with chemical potentials (dNTP or NTPs)

* Corresponding author at: Departamento de Estructura de la Materia, Física Térmica y Electrónica, Universidad Complutense de Madrid, Pza. de Ciencias, 1, 28040 Madrid, Spain.
E-mail address: francao@ucm.es (F.J. Cao-García).

https://doi.org/10.1016/j.csbj.2021.06.032

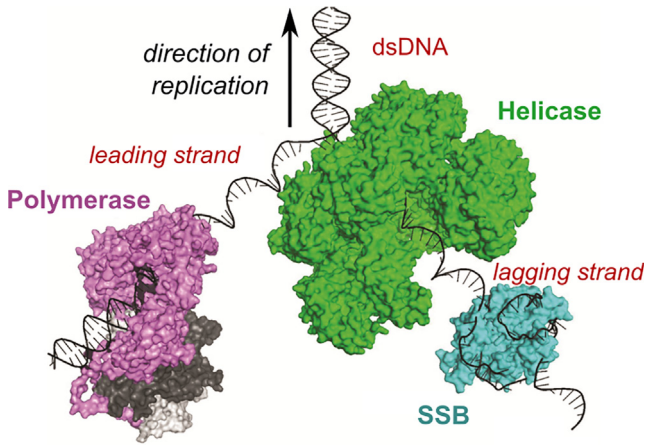

Fig. 1. Simplified view of the core components of the mitochondrial DNA replisome. Helicase opens the DNA fork separating the two strands. The leading strand, is replicated directly by the DNA polymerase, and the lagging strand, is initially bound by SSB proteins for later replication. In other replisomes, several primase subunits usually associate with the helicase to prime replication of the lagging strand, which is replicated in the opposite direction in the form of short Okazaki fragment (not shown). Adapted from [80].

to exhibit unidirectionality rather than random motion and high rates, e.g., some DNA polymerases can synthetize DNA at a rate of 500–1000 nt/s [87].

In the last two decades, the advent of *in vitro* single-molecule techniques has allowed researchers to explore, for the first time, the molecular mechanism that govern the operation of many proteins involved in DNA replication including DNA polymerases and helicases [64,99,26,68,11]. In particular, *in vitro* single-molecule manipulation techniques, such as optical and magnetic tweezers, have provided mechanistic information about the inner workings of these systems that cannot be obtained with ensemble techniques. Briefly, single-molecule detection opens the possibility to follow the activity of individual biological molecular motors, such as DNA polymerases and helicases in real-time. In this way, it is possible to detect and quantify transient features of the reaction as rare events and heterogeneous behaviour [67,30,65]. This possibility is instrumental in unveiling the complex dynamics of these biological motors. The position of the biological motor is measured indirectly through the position of a microsphere (polystyrene bead) linked to the motor or to their substrate (e.g., DNA), with a resolution of 1–10 nm. The beads are manipulated by the trapping or manipulation field (optical or magnetic), which allow to measure and apply controlled mechanical forces in the order of picoNewtons (pN), Fig. 2A and D. Thus, *in vitro* single-molecule manipulation techniques allow measuring the tiny mechanical forces generated during the course of a reaction and applying controlled mechanical forces directly on it. Note that mechanical force is a byproduct of the reaction of molecular motors. The goal of measuring and applying external forces to these systems is to determine the magnitude of the rates and free energies of the mechanical steps of their reaction. In this way, a detailed picture of the mechanical coordinate, and its relation with the chemical coordinate of the reaction (mechano-chemistry) can be obtained [50].

The spatial resolution of single-molecule manipulation techniques is limited by drift and thermal noise [35,21,56,37,81]. Thermal fluctuations affect both the instrument and the sample and provide a fundamental limit to the resolution of a given experiment. Because molecular motors operate at energy levels comparable to those of thermal motion, single-molecule manipulation data often present low signal to noise ratios [13]. In addition, in the case of DNA replication studies, the flexibility of the DNA substrate increases the noise level of the data significantly [35,14].

Low signal to noise ratio data requires special data analysis techniques, which typically accumulate evidence (locally or globally) aiming to obtain accurate and unbiased kinetic information. The optimal extraction of the kinetic information from the stochastic operation of individual DNA-based molecular motors is still an open theoretical challenge. Here, we aim to provide a perspective on some relevant techniques for the analysis (Section 2) and modelling (Section 3) of DNA replication and DNA unwinding activities obtained by *in vitro* single-molecule manipulation techniques. Data analysis techniques allow extracting the main phenomenological information from the individual trajectories. Models predict relations between the observations, which, compared with the observed relations, enable the identification of the underlying processes. Thus, models provide insight on the mechanisms and increase the future predictability of the phenomena. Here, we review fundamental models of DNA unwinding and replication (in different conditions and in the presence or absence of ligands), and also model selection criteria. We center our description on methods used to analyze and model *in vitro* single-molecule manipulation data on DNA replication. Similar/related data analysis and modelling techniques have been used to study the real-time kinetics of the activity of other nucleic acid-based molecular motors studied at the single-molecule level, such as RNA polymerases [33,32,25] and ribosomes [103,23]. Data analysis and modelling of the stochastic molecular motor trajectories have both benefited from and contributed to statistical physics developments [84]. This fruitful interaction will continue (Section 4).

2. Data analysis

Single-molecule manipulation techniques provide a way to measure properties of individual molecules (e.g., position, orientation, end-to-end extension), which can be used as reaction coordinates to follow the evolution of the molecule along a reaction pathway in real time. Using the DNA extension as a reaction coordinate, this technique allows monitoring the activity of replicative DNA polymerases as they convert single-stranded (ssDNA) to double- stranded (dsDNA) DNA, Fig. 2A [57,104,43,58,78], and replicative DNA helicases that unwind dsDNA to ssDNA, Fig. 2D [47,55,60,88,95,89,96,11,48].

For example, in optical and magnetic tweezers experiments, as those shown in Fig. 2A and D, the experiment provides the measured DNA distance X as a function of time t at the given tension f. The force extension curves of ssDNA, $x_{ss}(f)$, and of dsDNA, $x_{ds}(f)$, provides the information required to convert the changes in distances between the beads to replicated nucleotides [14]. The conversion factor depends on the experimental configuration. With the experimental configuration shown in Fig. 2A, the number of nucleotides replicated is given by $\frac{X(t)-X(t_{pe,0})}{x_{ds}(f)-x_{ss}(f)}$, as each replication step involves the convertion of one ssDNA nucleotide to its dsDNA configuration. ($t_{pe,0}$ is the initial time of primer extension DNA replication.) Whereas the number of nucleotides unwound by a single helicase using the experimental configuration shown in Fig. 2D is given by $\frac{X(t)-X(t_{u,0})}{x_{ss}(f)-x_{ds}(f)}$, as it involves the transformation of one base pair of dsDNA into one nucleotide of ssDNA along the pulling coordinate. ($t_{u,0}$ is the initial time of DNA unwinding). Other experimental configurations involve other conversions, for example, if the ssDNA is covered by SSB protein the ssDNA-SSB complex force extension curve $x_{SSB}(f)$ should be characterized and used instead of the ssDNA force extension curve $x_{ss}(f)$. See Fig. 2G [17].

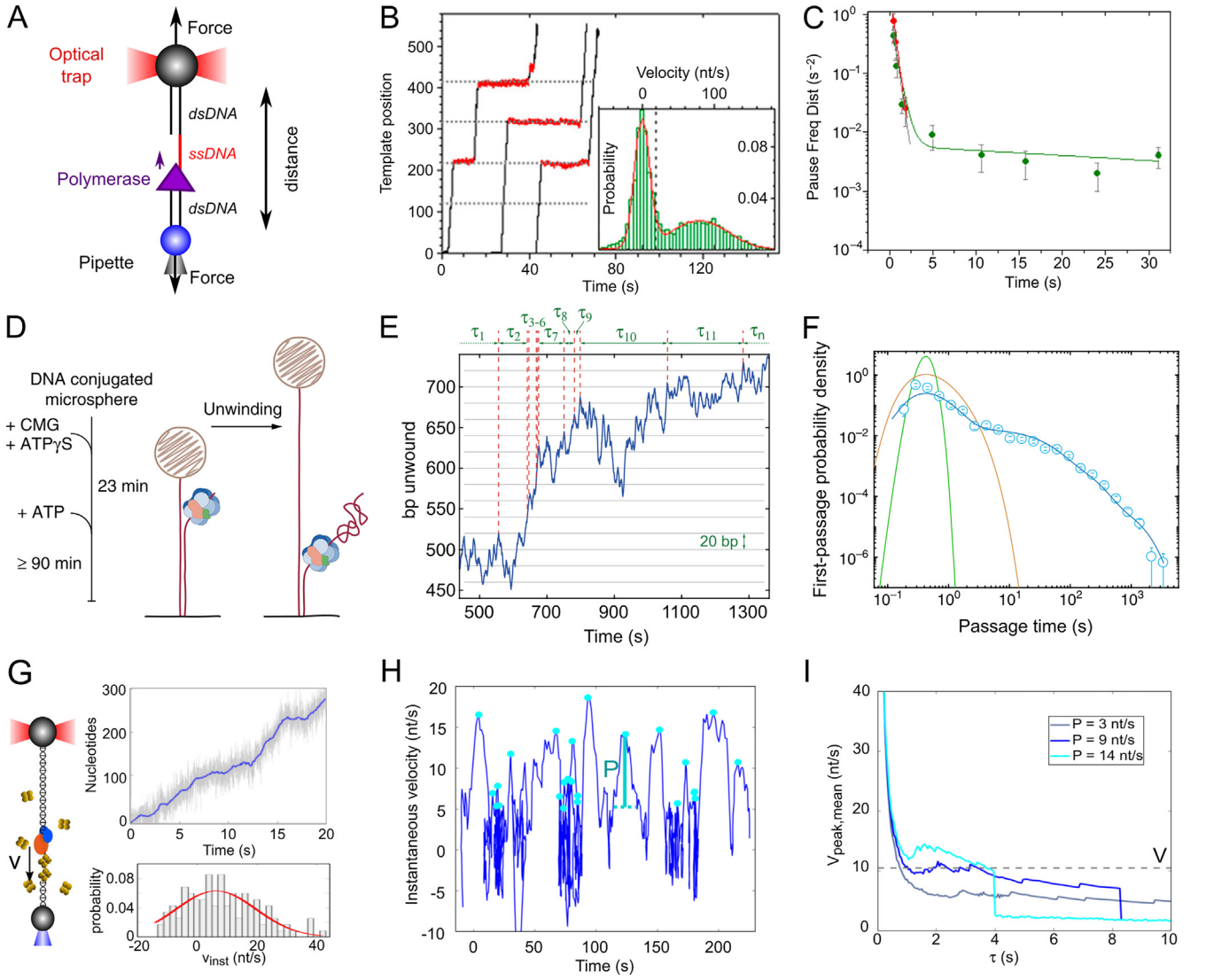

Fig. 2. A) Example of a primer extension (p.e.) DNA replication experiment with optical tweezers. The two ends of a dsDNA molecule containing a ssDNA gap in the middle are attached between two micron-sized beads: one (grey sphere) held by the optical trap (highly focalized laser, red) and the other (blue sphere) held by suction on top of a micropipette. As replication proceeds, ssDNA is converted into the more rigid dsDNA changing the distance between the beads. The change in distance is registered and later processed to obtain the polymerase trajectory (template position versus time). **B)** Representative replication traces showing transient pause events (red) intercalated with active replication events (black). The inset shows the velocity histogram. **C)** Pause length frequency distributions. Depending on the experimental conditions, the pause length frequency distribution can be compatible with a single (red line) or a double exponential distribution (green line). Other distribuitions are also possible. **D)** Diagram of a magnetic tweezers experiment measuring the DNA unwinding activity of a single replicative helicase. One of the ends of the dsDNA (bearing a helicase loading site) is attached to a glass surface and the other end to a super-paramagnetic microsphere manipulated by the magnetic field. **E)** Representative unwinding trace showing the binning in displacement and illustrating the determination of first passage times τ_i. **F)** First-passage time (FPT) distribution. Experimental data (blue circles) are fitted by a model with pauses and forward and backward stepping (blue line). Predicted FPT distribution without pauses (orange) and with only forward stepping (green) are shown for comparison. **G)** Experimental configuration to measure the replication of ssDNA covered with SSB with optical tweezers. When the replication velocity is slow, direct identification of pauses and maximum replication velocity V is not possible, neither from the traces nor from the velocity histogram. **H)** Identification of peak velocities V_{peak} with prominence greater than P (which avoids the selection of secondary peaks). Traces are averaged on a time window τ then the instantaneous velocity is represented versus time to proceed to peak velocity identification. **I)** The mean of the velocity peaks $V_{peak.mean}$ is computed for each prominence P for different time windows τ. For an intermediate value of the prominence P a clear plateau is present in the $V_{peak.mean}(\tau)$ plot, whose value gives the maximum velocity V. (Panel A in this Figure is adapted from Ref. [43]; Panels B and C are from Ref. [69]; Panels D, E and F are from Ref. [11]; Panels G, H and I are adapted from Ref. [17]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The first and most direct information we can obtain from a replication trajectory is the polymerase mean velocity V_{mean} (usually expressed in nucleotides(nt) per unit of time, e.g., nt/s) and the processivity N (the number of nucleotides replicated before detachment). This information can also be restated as the mean residence time per nucleotide $T = 1/V_{mean}$ (units of time per nucleotide). The detachment time is then given by the product $N \cdot T$, and the detachment rate as $k_{detachement} = 1/(N \cdot T)$.

Additionally, DNA replication trajectories present pauses or transient inactive states that alternate with active replication events (see Fig. 2B). However, when the polymerization rate is low, these two states cannot be easily disentangled due to the noise in the trajectory. The main source of experimental noise comes from the high flexibility of the ssDNA polymer at the low tensions relevant to study DNA unwinding and DNA replication (<10 pN). We briefly describe here the main techniques to separate the pause and active state contributions in the trajectory.

When pauses can be identified from the trajectory plot directly (as in Fig. 2B), we can obtain information on frequency of pauses and pause duration. Pause identification is performed

by direct statistical methods, as those we cite in Section 2.1. When transitions from pause to active state on a trajectory are difficult to identify, one of the pause identification methods, the velocity histogram, can still provide an estimation of the fraction of time in the pause and in the active state, as described in Section 2.2. A typical velocity histogram shows two peaks (inset Fig. 2B): one for the pause state (centered at a velocity close to zero) and the other for the active state (centered at velocity greater than zero). The later peak corresponds to the maximum velocity or velocity without pauses. However, when the signal to noise ratio decreases, the velocity histogram cannot disentangle the pause and active states contributions. In these cases, the prominence method allows to estimate the active or maximum velocity [17]. The idea of the prominence method is to identify a characteristic velocity higher than the mean velocity and relatively independent of the interval of computation of the velocity. This method is described in Section 2.4. Section 2.3 describes an alternative to the velocity histogram, the first passage time distribution, based on binning in positions and doing a histogram of first passage times. Section 2.5 comments on the Bayesian approach to data analysis, which is more model-dependent.

2.1. Direct identification of pauses

In high signal to noise ratio trajectories (as those shown in Fig. 2B) pauses can be identified using several methods, such as plateau identification techniques, step-fitting algorithms or velocity threshold algorithms [16]. Signal to noise ratio sometimes can be increased by averaging over a sliding time window [98,90,1,16]. Time averaging can help to reduce the high frequency noise at the expense of a reduction on time and position resolution.

Plateau identification techniques identify pauses directly as the constant position intervals in the trajectory (see Fig. 2B). During a pause, the next position (or a mean of next positions) is not significantly different from the mean of the previous positions [16]. This observation provides a pause identification method comparing the mean and standard deviation of the previous points with the mean and standard deviation of the next points, which allows to automatize the data processing. Additionally, step-fitting algorithms search to fit the trajectory with the optimal number of steps [52,92], when the resolution is high enough for pause identification.

Velocity threshold algorithms [42,16] identify pauses as sections of the trajectory where the local velocity is below a certain value. One effective method to select an appropriate velocity threshold is doing a velocity histogram [43,69]. For high signal to noise trajectories (as in Fig. 2B) the velocity histogram presents two clear peaks (as in the inset of Fig. 2B): one centered around zero velocity, identified as the pause state contribution; and another peak center around a nonzero velocity, identified as the active state contribution. When the system is at a pause state, the average velocity is zero (with deviations due to the noise in the trajectory that provide the width of the peak). In contrast, at the active state, there is a finite mean velocity due to the stepping of the motor. (The deviations from the mean active velocity are due to the stochastic nature of the stepping process and to the noise in the trajectory.) The valley of the velocity histogram between the two peaks provides the threshold velocity, which is then used to identify the two states along the trajectory (active and pause states). Additionally, peaks at negative velocities may appear when exonuclease events during DNA synthesis are significant [58,78].

Pause identification allows computing other magnitudes that characterize the trajectory. In particular, the fraction of time the polymerase (or helicase) is in active state is named moving probability, $MP = (\text{total time without pauses in the trajectory})/(\text{total time in the trajectory})$. We can also determine the active and passive state contribution to the mean residence time per nucleotide $T = T_a + T_p$. The active time per nucleotide is given by $T_a = T \cdot MP$, and the pause time per nucleotide by $T_p = T \cdot (1 - MP)$. We can also define the replication velocity or maximum replication velocity, $V = 1/T_a$ (or sometimes V_{max}), as the average replication velocity in the active state.

Pause identification provides a great deal of information about pause behavior and its tension dependence. Calculation of the pause length frequency distribution $\rho(t)$ reveals whether one or more characteristic pause durations are present in the trajectory (Fig. 2B, C). First, we define the points t_i of the pause duration binning. Then, we compute the pause length frequency distribution for a trajectory (or set of trajectories) as

$$\rho\left(t_i\right) = \frac{\left(\text{number of pauses of duration between } t_i \text{ and } t_{i+1}\right)}{\left(t_{i+1} - t_i\right) \cdot (\text{total time without pauses})}. \quad (1)$$

$\rho(t)dt$ provides the frequency of entering to a pause of duration between t and $t + dt$ when the polymerase is in its active state. (t_i is estimated as the mean duration of the pauses of duration between t_i and t_{i+1}.) This procedure also allows to have a non-uniform binning in the pause duration, which is required when there are two or more characteristic pause durations of different order of magnitude. For example, in Fig. 2C a smaller bin size is used for small pause durations (reflecting short pauses in the trajectories of Fig. 2B), while a larger bin size is used for large pause durations (reflecting long pauses in trajectories of Fig. 2B). This non-uniform binning allows to adequately resolve the distribution for short pauses, and to have enough counts to have a reliable value in each bin for long pauses (Fig. 2C). Pause length frequency distribution is also named as distributions of pause durations, waiting-time distributions and dwell-time distributions [54].

This pause length frequency distribution, Eq. (1), fits to the sum of one or several exponentials, depending on the number of characteristic pause durations present on the distribution [44]. For example, when two characteristic pause durations are presents (as in Fig. 2C) the pause length distribution fits to

$$\rho(t) = a_1 \lambda_1 e^{-\lambda_1 t} + a_2 \lambda_2 e^{-\lambda_2 t}, \quad (2)$$

where a_i is the pause frequency of pause i, (i.e., the frequency of entering to a pause state of type i), and $1/\lambda_i$ is the characteristic pause duration of pause i, or average pause length of pause i. These parameters of the fit are related to the enter and exit rate to (and between) pauses, as it is discussed below in Section 3.1. When the pause duration distribution is plotted in log-scale (as in Fig. 2C), the number of characteristic pause durations becomes apparent as the number of nearly straight sections in the plot (provided the distribution is well resolved and the characteristic pause durations are different enough). In cases where it is unclear the number of characteristic pause durations to use on the fit, one can resort to model comparison likelihood techniques as the AIC information measure [3,12]. Additionally, the average pause length $1/\lambda_i$ and the pause frequency a_i of the pauses can depend on the force applied to the DNA or to other relevant condition in the DNA replication experiment (as whether it is a GC or a AT bound in the fork [69]). The dependence of pause frequency and duration on force will inform about the nature of the pause state (as discussed below in Subsection 3.1).

The magnitudes introduced up to this point constitute a complete phenomenological description of the observed replication velocities and pauses. This level of detail on the pause description is not always possible. In the next subsections, we address how to deal with more noisy trajectories where a detailed pause identification and description is not possible.

2.2. Velocity histogram

When the direct identification of pauses is not possible, a histogram of the instant velocities can still provide information on the fraction of time that the system is in pause [43]. Typical velocity histograms present two peaks, one centered around zero velocity, the pause peak, and another one centered around a nonzero velocity, the active peak. See for example the inset of Fig. 2B.

When these two peaks are well resolved, we can identify the counts of the pause states and those of the active states. In this case, individual pause identification is also possible, as each velocity count corresponds to a time in the trajectory (see previous subsection). When the two peaks cannot be well resolved, we can resource to methods that help to get a better resolved velocity histogram. One of the methods is to average the data over a sliding time window, which averages out the (high frequency) experimental noise. It is important to optimize the width of the average time window. If the average time window is too wide, we lose relevant information; while if the time window is too narrow, the experimental noise is still present. Another method is to compute the instant velocity in a larger time window, which reduces the signal to noise ratio in the velocity. Note that if the velocity time window is too wide, it will average out pause and active state velocities, while if it is too narrow, it will not change the signal to noise ratio in the velocity significantly. The average time window and the velocity time window are optimized to maximize the resolution of the two peaks in the velocity histogram.

Even upon optimization of the position of the two peaks, an overlap between them may remain, see for example the inset of Fig. 2B. In the overlap region it is unclear which counts correspond to active or to pause states. If the minimum between the two peaks is low, the two peaks are well-resolved and the location of the minimum can be used as pause identification criteria along the trajectory (as described previously in Section 2.1). If the minimum is high, the pause identification is not possible, but we can estimate the fraction of time in each state from the velocity histogram. The more elementary method to deal with this problem is to find the location of the minimum between the peaks and assign the counts below to the pause state, and the counts above to the active state. A more elaborate method is fitting the velocity histogram to a two Gaussian functions. The area under each Gaussian is proportional to the pause and active state probability, respectively. The quotient between the pause area (or counts), A_p, and the active state area (or counts), A_a, gives the value of the ratio between the pause and the active time per nucleotide $T_p/T_a = A_p/A_a$. With this information we can compute the moving probability, $MP = \frac{T_a}{T} = \frac{1}{1+T_a/T_p}$; and also, the (maximum) replication velocity $V = MP \cdot V_{mean}$.

2.3. First passage time distribution

The velocity histogram method takes equal time bins on the trajectory and gets the different displacement (in nucleotides), and therefore velocities along the trajectory. Instead, the first passage time distribution method does the binning in displacement, Fig. 2E, giving for a fixed displacement the different first passage times along the trajectory. The observed first passage time distribution is then analyzed to extract the replication rate, the number of pauses and their characteristics [22,27,11].

The theoretical first passage time (FPT) distribution for a single-step of a molecular motor stepping forward at a rate k_+ is given by an exponential distribution,

$$\rho_{FPT}(t) = k_+ e^{-k_+ t}. \tag{3}$$

It gives the probability of a single-step occurring at a time t (after the previous one). However, in practice, frequently the position noise is higher than the single-step length. Hence, the displacement binning should be large enough (higher than the typical position noise) to prevent noise dominating the first passage time distribution. For m steps binning, the FPT distribution is given by the probability that m forward steps require a time t,

$$\rho_{FPT}(t) = \frac{k_+^m}{(m-1)!} t^{m-1} e^{-k_+ t}, \tag{4}$$

which is a gamma distribution [86,27]. When the molecular motor additionally has backsteps ($k_- \neq 0$) the FPT distribution becomes wider. (In this later case, its analytic expression can be stated in terms of modified Bessel functions of the first kind [22,11]). Entrance in pause states increases the time spent to advance the displacement binning length, increasing the probability of higher first passage times. Thus, FPT distributions with large tails at high first passage times reflect the presence of one of several pauses (or backstepping). See Fig. 2D, E, F. Each interval of the FPT can be approximated by a single contribution (forward, pause), provided the characteristic times (pause, forward stepping) are well separated. This approach has allowed the identification of pause states in the operation of a DNA helicase and a RNA polymerase [27,11].

Both FPT distribution and velocity histogram methods resort to local information in the trajectory after a binning (on space or time). These methods are limited to cases where the time scales are well separated. Prominence method and Bayesian methods explained below have proven to be more effective when dealing with cases where pause identification is difficult. Their strength is that they use information on the sequence, additionally to the (averaged or binned) local information.

2.4. Prominence method

We can resort to the prominence method [17] when it is not possible to resolve the two peaks in the velocity histogram (e.g. Fig. 2G). The prominence method aims to obtain the (maximum) replication velocity V from the information contained in the trajectory. The idea is that the more prominent peaks in the velocity time series correspond to the (maximum) replication velocity V, after noise removal (Fig. 2H). Prominence is a term from topography and mountaineering, serving to identify the main peaks in a mountain ridge. The prominence of a peak is the difference between its height and the lowest closed contour line encircling it and without any higher peak inside. Here, in the velocity vs. time plot, the prominence of a velocity peak is given by the difference in height between the peak and the higher valley separating the peak from another higher peak.

The procedure involves computing the velocity time series using a time window τ, then selecting the more prominent peaks, with prominence at least P, in the velocity time series (Fig. 2H). After the mean of the peaks is computed and represented for each prominence P as a function of the time window τ. The maximum velocity appears as the height of a plateau in this plot for the appropriate prominence P. See Fig. 2I. (The adequate time window τ and the prominence P do not need to be known a priori, a wide range is taken for both and the prominence method provides the adequate range.)

The procedure is the following. First, the velocity time series is computed from the trajectory (position vs. time) using a time window of length τ, $V_{inst}(t) = (X(t+\tau) - X(t))/\tau$. Second, the peaks of prominence P or higher are selected on the velocity time series. After the mean of the peaks $V_{peak,mean}$ is computed and represented for each prominence P as a function of the time window τ. See

Fig. 2I. Finally, the plots of the mean of the velocity peaks $V_{peak,mean}$ as a function of the velocity time window τ for different prominences P, reveals that there is a plateau of $V_{peak,mean}$ which is both flatter and larger for the appropriate value of P. The height of the plateau for this optimal value of P gives a good estimate of the (maximal) replication velocity V.

Large time windows τ lead to a velocity plot with high signal to noise ratio. However, if the time window is too large it averages active replication periods and paused periods, leading to velocities below the maximum replication velocity. This is reflected in each of the plot of the mean of the velocity peaks $V_{peak,mean}$ as a function of the velocity time window τ. See Fig. 2I. At low time window τ, the plot is dominated by large velocities due to the noise in the trajectory, giving large values of $V_{peak,mean}$. For intermediate values of the time window, the influence of noise decreases, presenting a plateau at intermediate times. The plateau is clearer for the appropriate prominence P, and gives the replication velocity (on the active state) V. For large values of the time window, similar or larger than the characteristic time in the active replication state, the time window implicates averaging active and pause state periods, and $V_{peak,mean}$ goes to the mean replication velocity.

Dividing the replication velocity V obtained by the mean replication velocity of the trajectories V_{mean} we obtain the moving probability as $MP = \frac{V_{mean}}{V}$. The active time per nucleotide can be obtained as $T_a = T \cdot MP$, and the pause time per nucleotide as $T_p = T - T_a$. A complete description of the prominent method can be found in the Supplemental Information of Ref. [17].

2.5. Bayesian methods

Other class of methods to analyze the trajectories are the Bayesian methods [24]. Bayesian methods fit a model of the system and the experimental device to the observed data. The idea is to find the more probable values of the model parameters given the observed data. This is linked through Bayes theorem to the question of which are the values of the model parameters maximizing the probability that the data is observed (as in fact it was). Most of the models considered belong to the class of models known as hidden Markov models (HMM), as they assume that there is an underlying Markovian behavior of the system. A system is said to be Markovian when its next state depends only on its present state and it is independent on the previous story. (The models described in the next section belong to this class of models.)

The experimental device characteristics are also included in the model, for example, through an experimental noise parameter affecting the observed data. This noise parameter can also be fitted, and the result should be consistent with the expected experimental device accuracy.

Bayesian methods in combination with HMM have been successfully used for example for the study of gene transcription [97], after one of his early prominent uses, speech recognition [85]. They also have a wide set of potential applications on single-molecule data analysis [28,74,29]. These methods are very powerful to identify the best parameters values, and even the best model of a set of models [24], combining them with a model comparison criterion, as AIC [3,12]. They provide means to include other a priori information (obtained in previous complementary experiments). It might be argued that their drawback is that they require a concrete model (or set of models) to proceed with the analysis. However, this is also true (to a certain extent) for the more frequentist data analysis presented in previous subsections. In previous subsections we assumed a pause and an active state characterized by different velocities, and the presence of transitions between them. But in the previous subsections we do not had to assume a priori whether there was one or several pause states. In this sense the approach of the previous subsections can be considered a more model-independent analysis.

In the next section, Section 3, we present models which link the observed replication velocities and pause characteristics with the underlying processes. The models allow us to identify these underlying processes and get a deeper understanding of DNA unwinding and replication.

3. DNA replication models

Models allow the identification of the underlying processes in a phenomenon, increasing its predictability. Assuming a process, a model predicts relations and dependencies in the observations. When we fit a model to a set of experimental data, we are checking whether the assumed process is compatible with the relations and dependencies between the observations.

The fit of the model to the data can be performed with a minimization of the mean squared differences between the observed data and the predicted value of the model. This minimization sets the optimal values of the model parameters that describe this set of data. A deficient fit indicates that the model has wrong or incomplete assumptions on the possible underlying processes. Models with the same number of free parameters can be compared directly using the minimization of the mean squared differences. When comparing models with different number of free parameters, we must use model comparison criteria (based on Bayesian statistics), as the Akaike Information Criterion (AIC) [3,12]. These comparison criteria compensate that models with more parameters generally fit better leading to overfitting. It is easy to find models with n parameters that fit a set of n data, but in fact this fit would be only a reparameterization of the data. This overfitting situation does not provide information on whether the underlying process assumed on the model is compatible with the data. In practice, to prevent overfitting, we should keep the number of parameters well below the number of data points. We should also keep track of data uncertainties and the uncertainties in the fitted parameters. We should be aware that a large uncertainty in a fitted parameter is a sign of low capability of the data to determine this parameter. In other words, the model poses questions that cannot be answer with the information contained in the data. Thus, a large uncertainty in a parameter indicates that we should either go for a simpler model (with less ingredients and parameters) or to do complementary experiments (which provide more information on the process related to this parameter).

3.1. Pause modelling

The key ingredient in the modelling of pauses is the number of pause types. Pause identification was described in the previous Section 2.1. It identified the number of pause types, as the number of characteristic pause length observed, i.e., the number of exponentials with different exponents fitting the pause length frequency distribution $\rho(t)$. The number of types of pauses identified restricts the reasonable pause models, but not uniquely, as we mention below for the two-pause case.

For the cases with one pause type, Fig. 3A, the measured pause length frequency distribution fits to $\rho(t) = a_1\lambda_1 e^{-\lambda_1 t}$. For this model, the fitted parameters give the entry rate $k_{a1} = a_1$ and the exit rate $k_{1a} = \lambda_1$ from the pause state.

When two pause types are identified, the pause length frequency distribution fits to $\rho(t) = a_1\lambda_1 e^{-\lambda_1 t} + a_2\lambda_2 e^{-\lambda_2 t}$. A possible model is the linear two pause model represented in Fig. 3C. For this model, the fitted parameters give the entry rates $k_{ai} = a_i$ and the exit rates $k_{ia} = \lambda_i$ from the pause state, with $i = 1, 2$. However, the same pause length frequency distribution is compatible to

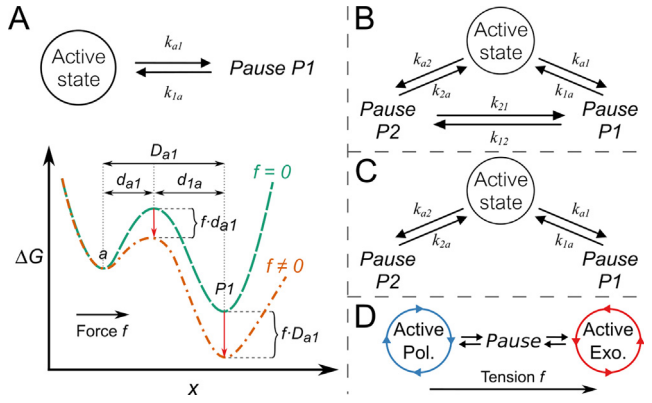

Fig. 3. Kinetic models for pause and active states of DNA polymerases. Parameters k_{ij} denotes the transition rate from state i to state j. **A)** Top: Model with a unique pause state. Bottom: Schematic representation of the effect of an external force f on the free energy landscape projected along the displacement coordinate in the direction of the force. The free energy reduction is given by the work done by the force: $f \cdot d_{a1}$ for the activation state, $f \cdot D_{a1}$ for the final state. **B)** Cyclic model with two pause states. In this model direct transitions between pause states are allowed. **C)** Linear model with two different pauses states. In this model it is not possible to go directly from one pause state to the other one, without passing through the active state. **D)** Model of polymerization-exonucleolysis transitions mediated by a pause state. DNA tension induces entrance into exonucleolysis through the pause intermediate [43,41]. (Panels A, B and C adapted from Refs. [69,73]).

the more general cyclic two pause model represented in Fig. 3B. See Refs. [44,73]. This cyclic pause model contains the linear two pause model as a particular case where the transitions between pauses have negligible rates. The general cyclic model has the drawback that has more free parameters (six transitions rates) than those provided by the pause length frequency distribution (two pause entrance rates and two characteristic duration of pauses). However, the positivity of all the six rates gives maximum and minimum values compatible with the observed fitted four parameters, as shown in Ref. [73]. Nevertheless, only if there is a biological motivation the use of a model with more parameters than those directly measure seems reasonable. In this case, the model calls for complementary experiments or information to help reduce the uncertainty in the transition rates on the cyclic case.

The characteristic entrance and exit rates can also depend on the mechanical tension applied to the systems and/or on the DNA sequence, as previously stated in Section 2.1 for the fitted parameters of the pause length frequency distribution $\rho(t)$. The force dependencies of the entry and exit rates from pause states reveal the magnitude of conformational changes d_{ij} (along the force direction) to the activation state, which governs the entry to and the exit from the pause state. See bottom of Fig. 3A. This distance, d_{ij}, can help to reveal which may be the process leading to the pause [69]. The force dependency of the entry and exit rates from a pause state is given by

$$k_{a1}(f) = k_{a1}(0) \cdot e^{\frac{f \cdot d_{a1}}{k_B T}}, \qquad k_{1a}(f) = k_{1a}(0) \cdot e^{-\frac{f d_{1a}}{k_B T}}. \tag{5}$$

The distance, d_{ij}, parameterizes the different effect of the force on the entry and exit rate. The work $f \cdot d_{ij}$ gives the magnitude of the change of the effective barrier to the activation state of the process. Its ratio with the characteristic energy of the thermal fluctuations $k_B T$ determines the magnitude of the increase or decrease of the process rates, as shown in the expressions of Eq. (5). Fitting these expressions to the observed force dependence of the rates provides the rates at zero force, $k_{ij}(0)$, and the conformational change distances d_{ij}.

The equilibrium constant of the process can be defined by

$$K_{a1}(f) = \frac{k_{a1}(f)}{k_{1a}(f)} = \frac{k_{a1}(0)}{k_{1a}(0)} \cdot e^{\frac{f(d_{a1}+d_{1a})}{k_B T}} = K_{a1}(0) \cdot e^{\frac{f D_{a1}}{k_B T}}. \tag{6}$$

This gives a complete characterization of the entry and exit from the pause, and a clue of the possible conformational changes associated to this pause through the value of $D_{a1} = d_{a1} + d_{1a}$. This simplified description of the effect of force on process rates is frequently used in biophysical studies. For a description of the simplifications involved and more exact descriptions see Refs. [51,44,101].

More advanced multistate models and their kinetics are described in Refs. [50,18,66]. Some of these models have been applied to interpret the force-velocity dependencies of replicative DNA polymerases switching between polymerization and exonucleolysis, Refs. [58,41,78], Fig. 3D. Many replicative DNA polymerases present two main active sites; the polymerization (*Pol*) and exonucleolysis (*Exo*) active sites. The *Pol* site catalyzes 5′-3′ DNA synthesis by the stepwise addition of the complementary nucleotide (dNTP) on to the terminal 3′ end of the nascent DNA strand (primer), while the Exo site hydrolyzes mismatched nucleotides from the primer strand in the 3′ to 5′ direction increasing the fidelity of the copy. The Exo site is separated by 40–60 Å [102,7] from the Pol site and only binds single-stranded DNA (ssDNA). Therefore, the primer transfer reaction implies substantial conformational changes and may involve intermediates states. A finedtuned coordination between polymerization and exonucleolysis reactions is essential for the integrity of the genome. Modeling of the pause kinetics of replicative DNA polymerases and their dependence on mechanical tension applied to the DNA template has provided insight into the Pol-Exo transfer (or proofreading mechanism) dynamics of DNA polymerases. As described below, mechanical tension applied to the DNA template decreases the polymerization rate until stalling (Fig. 4A). Interestingly, upon stalling, a further increase of tension induces processive exonuclease activity in several DNA polymerases [104,43,78], suggesting that tension can be used as a variable to study the Pol-Exo equilibrium. Modeling of the effect of tension on the moving and pause states of phages Phi29 and T7 DNA polymerases has shown that the primer transfer reaction between the two active sites is not a one step process. In the case of Phi29 DNA polymerase, the primer transfer reaction is intramolecular and implies at least two intermediates states, one of which may work as a fidelity check point [43]. In the case of T7 DNA polymerase, the primer transfer reaction is intermolecular, following DNA polymerase dissociation the primer is bound to the Exo active site of a new DNA polymerase [41] . In summary, the ability to separate transient inactive states (pauses) from active states and analyze their corresponding force dependencies has been instrumental in determining the intermediates of the proofreading reaction and to measure directly the kinetic rates, equilibrium constants, and conformational changes associated with their interconversion.

3.2. Primer extension DNA replication models

Primer extension replication is the replication of ssDNA to give dsDNA (See Fig. 2A). *In vitro* single-molecule manipulation experiments with replicative DNA polymerases have shown that the average primer extension rate presents a strong dependence on mechanical force either applied to the DNA template (Fig. 2A,D) or to the DNA polymerase directly (Fig. 4C).

When mechanical tension is applied to the DNA, the average replication rate of many DNA polymerases increases initially with tension, reaching a maximum rate at ~6 pN. Above this value of

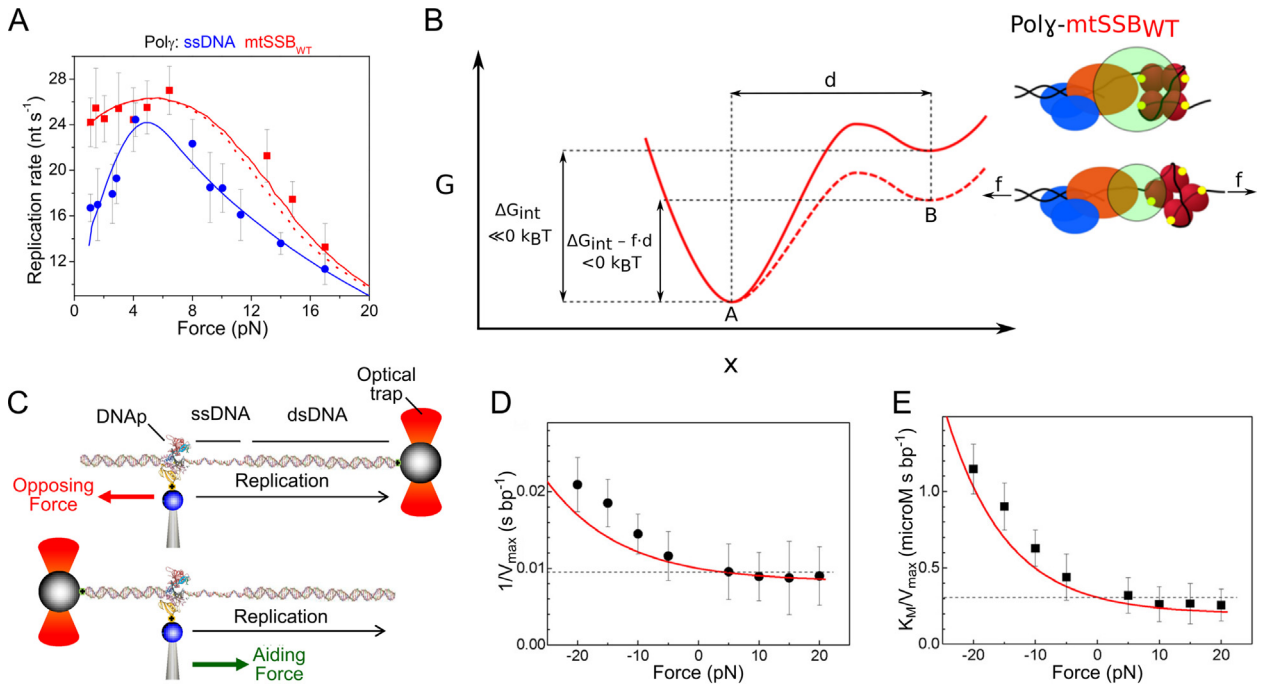

Fig. 4. Comparison between the effects of mechanical tension on the DNA (A and B) and mechanical load applied to the DNA polymerase (C-E). **A)** Effect of mechanical tension of the primer extension replication rate of the mitochondrial DNA polymerase in the absence (blue) and presence of the mitochondrial SSB (mtSSB$_{WT}$). Dots represent experimental data, and lines the best fitted theoretical models. Eq. (8) for ssDNA and Eq. (11) for SSB covered ssDNA. **B)** Comparison of polymerase-SSB coupling behavior at different tensions (f). The energy landscapes (left) between the coupled state A and the uncoupled state B, for low force (solid line) and for medium force (dashed line), show how force destabilizes the polymerase-SSB coupling. The diagrams (right) represent the polymerase-SSB coupling reduction due to force. [This decoupling effect is modeled by Eq. (10).] **C)** Diagram of a primer extension experiment applying opposing (top) or aiding (bottom) force on a DNA polymerase. **D)** Effect of load on the maximum replication rate V_{max} at saturating dNTP. **E)** Ratio of the apparent nucleotide constant and the maximum replication rate, K_M/V_{max}, (Michaelis-Menten parameters of the reaction) as a function of the force acting on the polymerase. (Panels A and B from Ref. [17], Panels C, D and E from [70]). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

tension, the replication rate decreases gradually until stalling (Fig. 4A). Originally, the so called Global Model [57,104] was proposed to explain the tension dependence as due to the activation enthalpy of converting n bases from single- to double-stranded DNA, which imposes

$$k_{pe}(f) = V(f) = k(0) \cdot \exp\left[\frac{-n \cdot f \cdot (x_{ss}(f) - x_{ds}(f))}{k_B T}\right] \quad (7)$$

where $k(0)$ is the replication rate at zero force f. The exponential argument accounts for the energy contribution involved in changing the length of n nucleotides from the ssDNA length per nucleotide $x_{ss}(f)$ to the dsDNA length per nucleotide $x_{ds}(f)$ at a DNA tension f.

This model, Eq. (7), explains well the replication rate decay with tension, f, for tension values above $4-6$ pN, with $n = 1$. This value of n indicates that only one template base is converted from ssDNA to dsDNA, which is in accordance with strong evidence from structural, bulk and single-molecule experiments [75,76,4,41]. However, Eq. (7) can only explain the entire force–velocity plot (including data at tension below $4-6$ pN) with a value of $n > 1$ [57,104]. Since only one nucleotide is added per polymerase step (n = 1), this model implies that $n-1$ bases have to be reverted to the ss geometry after the activation state. These results are not supported by previous structural and bulk kinetic studies. Alternative models that only involve the two neighboring DNA segment have been proposed (Local Model and its variations: Restricted-Cone Local Model and Minimalist Two Segment Model) [36,4,79]. These models explained well the tension dependence of the polymerization rate of some DNA polymerases, considering $n = 1$. However, the models relied on several assumptions about the DNA-polymerase interactions and the nature of the rate-limiting step, which need further experimental validation (see Supplementary Information of Ref. [17] for a more detailed discussion).

Recently, Ref. [17] proposed that the initial increase of the DNA replication rate with tension is caused by the mechanical disruption of the self-binding energies (secondary structure) of the ssDNA template [9]. Template secondary structures are known to hinder or slow down the advance of DNA polymerases [49,38,46]. This leads to the expression

$$k_{pe}(f) = V(f) = k(0) \cdot \exp\left\{\frac{-[n \cdot f \cdot (x_{ss}(f) - x_{ds}(f)) + \Delta G_{sec} \cdot (1 - \varphi_{sec}(f))]}{k_B T}\right\}, \quad (8)$$

where $k(0)$ would be the replication rate at zero force in the absence of secondary structure. The first term in the exponential argument accounts for the energy contribution involved in changing the length of n nucleotides (for DNA replication $n = 1$ nt) from the ssDNA length per nucleotide $x_{ss}(f)$ to the dsDNA length per nucleotide $x_{ds}(f)$ at a DNA tension f. The first term of the exponential dominates the decay of the replication rate at higher forces (Fig. 4A blue points and line.) The second term accounts for the contribution of the secondary structure, which slows down the replication at low forces, as shown in Fig. 4A. $(1 - \varphi_{sec}(f))$ gives the fraction of ssDNA template bases forming secondary structure and decreases for increasing force (Ref. [9] describes how to experimentally determine $\varphi_{sec}(f)$ from ssDNA force-extension curves). Each of the bases forming secondary structure imposes an average effective energy barrier of ΔG_{sec} to the advance of the polymerase. Fig. 4A (blue line) shows the fit of Eq. (8) to the force dependence replication rate of the mitochondrial DNA polymerase, Polγ. Future experiment would clarify whether this model, Eq. (8), gives good fits for other replicative DNA polymerases.

In vitro single molecule manipulation experiments showed that, in contrast to the effect of tension on the DNA, application of mechanical load opposing the direction of movement directly on the DNA polymerase decreases the average DNA replication rate monotonically (Fig. 4C,D) [70]. Mechanical load interferes with the translocation step of the polymerase, which becomes rapidly the rate-limiting step of the reaction upon of application of force, explaining the marked effect of force on the replication velocity [50]. Modeling of the combined effects of load and dNTP concentration on the maximum replication rate (at saturating dNTP concentration) and apparent nucleotide binding constant of the reaction, (V_{max} and K_M, respectively, Fig. 4D,E) provided a detailed picture of the coupling between the mechanical and chemical steps of the nucleotide incorporation reaction [70].

The model that explained well the data considered that mechanical translocation is independent on chemistry and therefore, the only force dependent rates of the reaction were the forward and backward translocation rates. According to this model, the kinetic expressions for the Michaelis-Menten parameters V_{max}, and K_M can be expressed as the sum of a force independent and a force dependent term

$$\frac{1}{V_{max}(F)} = a + b \cdot e^{F \cdot d_b/(k_B T)}, \quad \frac{K_M(F)}{V_{max}(F)} = r + s \cdot e^{F \cdot d_s/(k_B T)}, \tag{9}$$

where the coefficients *a, b, r,* and *s* are related to the rates of several steps of the nucleotide incorporation cycle such as *i.e.*, the catalytic rate (a), the dNTP binding rate (r), and the forward/backward translocation rates (b, r, and s). On the other hand, the d_b and d_s are the characteristic distances from the pre- and post- translocation positions to the transition state. Fits of the data with this model yielded the values of several of the main rates and force dependencies of the nucleotide incorporation cycle. In summary, modeling of the data revealed that chemical catalysis and mechanical translocation are not directly coupled. Instead, upon chemical catalysis, mechanical translocation of the enzyme occurs by thermal diffusion. This diffusion is biased towards the post-translocation state by binding of the next complementary nucleotide (dNTP) to the polymerization active site [70].

3.2.1. Effects of DNA ligands on primer extension replication

The presence of ligands bound to ssDNA, as the single stranded DNA binding proteins (SSB), favor DNA replication by suppressing the formation of secondary structure, but at the same time, they could also be a barrier to the access of the polymerase to the ssDNA template. However, some polymerase-SSB pairs are found to interact in such a way that the barrier imposed by the SSB for the ssDNA replication is negligible [17]. This collaborative interaction is force sensitive and is inhibited by different force values depending on the polymerase-SSB pair.

The probability to find the polymerase-SSB pair forming the collaborative pair can be parameterized as a transition between two states (collaborative and non-collaborative state),

$$P_{int}(f) = \frac{1}{1+\exp\left(\frac{\Delta G_{int}+f\cdot d}{k_B T}\right)}, \tag{10}$$

where ΔG_{int} is the coupling energy between the pair and d is the characteristic length of the conformational change that inhibits the formation of the polymerase-SSB collaborative pair.

Thus, in this case, the primer extension replication rate is given by

$$V(f) = k(0) \cdot exp\left[\frac{-\delta f \cdot (x_{SSB}(f) - x_{ds}(f))}{k_B T}\right] \cdot \left\{P_{int}(f) + (1 - P_{int}(f)) \cdot \exp\left[\frac{-n \cdot \Delta G_{SSB}(f)}{k_B T}\right]\right\}. \tag{11}$$

The free parameter n, representing the mean number of nucleotides to release from SSB per step, and the two free parameters of $P_{int}(f)$, ΔG_{int} and d, are fixed by fits to the experimental data on ssDNA replication in the presence of SSB. $k(0)$ is the replication rate at zero force in the absence of secondary structure previously obtained from the fit of Eq. (8) to the experimental data on ssDNA replication in the absence of SSB. $x_{SSB}(f)$ is the length of the ssDNA in the presence of SSB per nucleotide at tension f, and it is determined previously by force-extension experiments. $\Delta G_{SSB}(f)$ is the Gibbs energy to release a nucleotide from SSB at tension f, it is determined by the comparison of the integrals above the ssDNA-SSB and the naked ssDNA force extension curves [$x_{SSB}(f)$ and $x_{ss}(f)$, respectively]. A study of different polymerase-SSB pairs [17] found similar values of d, pointing to a similar conformational change, but different pairing energies, ΔG_{int}, indicating different stability of the collaboration state. Fig. 4A (red line) shows the fit of Eq. (11) to the force dependence replication rate of the polymerase Polγ replicating a ssDNA covered by mtSSB.

Overall, modelling of the effect of mechanical tension on the DNA replication rate in the presence of ligands (SSBs) revealed that elimination of template secondary structure by SSB binding promoted the maximum replication rate of DNA polymerases. However, for this stimulation to occur functional interactions between DNA polymerase and the SSB are required. These interactions, i.e., electrostatic repulsion, decrease the energy barrier of ssDNA unwrapping from the SSB and facilitate its release from the template without compromising the replication rate of the DNA polymerase.

3.3. Strand displacement DNA replication and DNA unwinding models

Some replicative DNA polymerases carry out strand displacement DNA synthesis, which is the ability to displace downstream dsDNA encountered during replication (Fig. 5A, B). Current strand displacement replication models mainly describe how the stability of the dsDNA fork ahead of the polymerase slow down the maximum replication rate of the enzyme (in the absence of fork) [8,47]. The main idea is that replication cannot proceed at tension f if the next base pair is closed, which, at tension f, happens with probability $P_0(l, m = 0, f)$, where l is the template position of the polymerase and m the number of base pairs of the fork opened ahead. Averaging over the complete template gives

$$\frac{V_{sd}(f)}{V_{pe}(f)} = 1 - \frac{1}{L}\sum_{l=0}^{L-1} P_0(l, m = 0, f)\,. \tag{12}$$

This probability is given by a balance of the Gibbs energy contributions involved in the fork opening

$$P_0(l, m, f) = \frac{\exp\left[-\frac{\Delta G(l,m,f)}{k_B T}\right]}{Z(l,f)}, \tag{13}$$

with $Z(l,f) = \sum_{m=0}^{L-l} \exp\left[-\frac{\Delta G(l,m,f)}{k_B T}\right]$. The Gibbs energy required to open m base pairs ahead of position l, at force f, is given by

$$\Delta G(l, m, f) = \sum_{i=l+1+m}^{L} \Delta G_{bp}(i) - 2m \int_0^f x_{ss}\left(f'\right) df' + [M - \min(m, M)] \cdot \Delta G_d. \tag{14}$$

The first term accounts for the stability of the base pairs ahead, $\Delta G_{bp}(i) \sim 2 - 3\ k_B T$ [105]. The second term accounts for the tension destabilization contribution, which is computed from the ssDNA elasticity $x_{ss}(f)$. The third term accounts for the interaction energy between the polymerase and the dsDNA fork. The polymerase is assumed to destabilize the M closer base pairs ahead by an amount ΔG_d each. These are the two free parameters in the model. ΔG_d parameterizes the activity of the polymerase (Fig. 5C) [8,60,59], and M, which parameterizes the range of fork destabilization. Large values of M should be interpreted with caution. They might be induced by the simplifying assumption in the model that the destabilization energy is the same for all the M next base pairs.

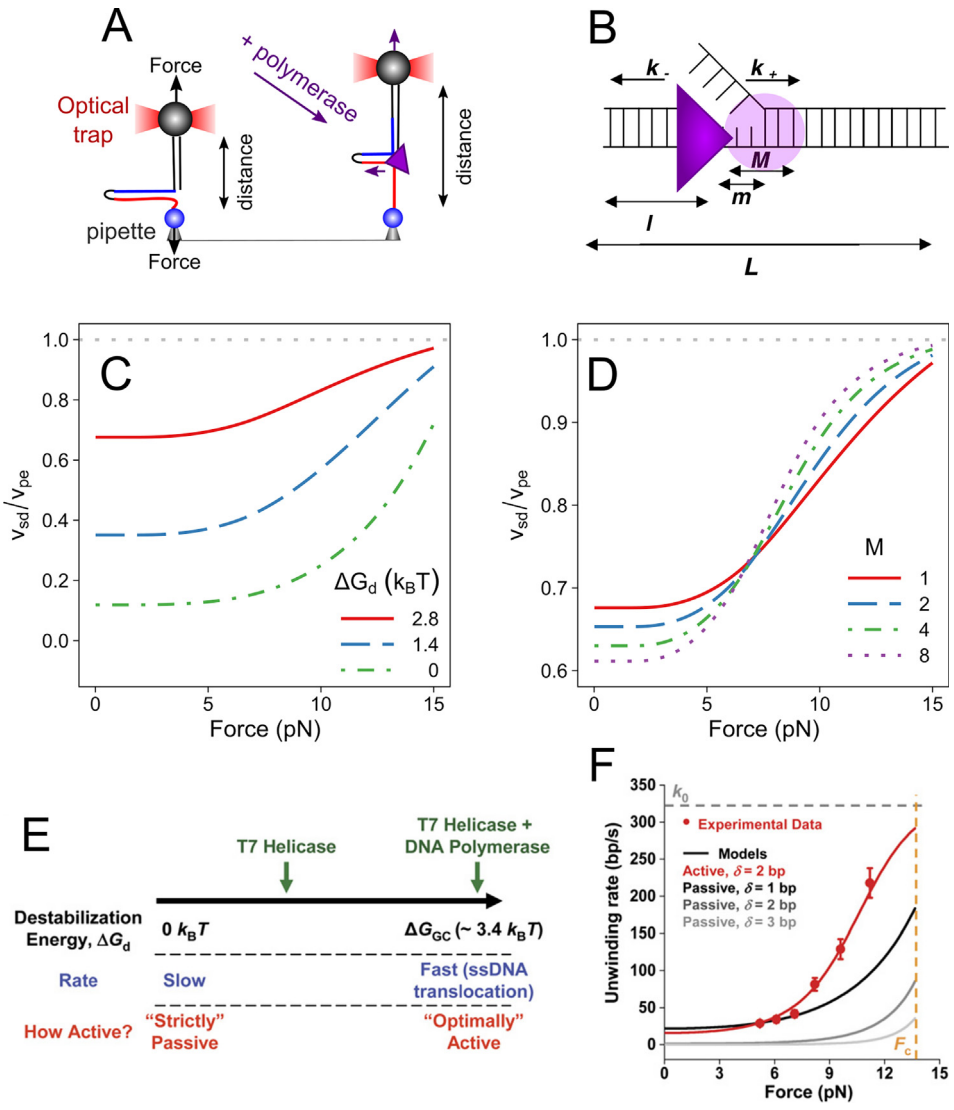

Fig. 5. **A)** Diagram of a strand displacement DNA replication experiment with optical tweezers. (Left) The two ends of a DNA hairpin are attached between two micron-sized beads (grey and blue spheres) one held by the optical trap and the other held by suction on top of a micropipette. Double parallel lines represent double stranded DNA (dsDNA). (Right) During strand displacement conditions (s.d.), the DNA polymerase (purple triangle) opens the DNA fork, replicates one strand (blue line), and displaces the other (red line). **B)** Scheme for the polymerase (purple triangle) dynamics during strand displacement DNA synthesis. *L* denotes the length in nucleotides of the DNA template, *l* the number of nucleotides replicated, *m* the number of base pairs opened between the polymerase and the DNA fork, while *M* stands from the number of base pairs that are destabilized by the polymerase (purple circle). All variables used in the strand displacement replication model are described in Section 3.3. (Adapted from [69].) **C)** DNA polymerases with high fork destabilization energies, ΔG_d, would present s.d. rates V_{sd} similar to those found during primer extension V_{pe} (red line). On the contrary, DNA polymerases with low ΔG_d present lower V_{sd}/V_{pe} ratios with stronger force dependencies (green dashed dotted line). (For all lines $M = 1$.) **D)** Variation of the force dependent V_{sd}/V_{pe} ratio with the interaction range *M*. Higher M values yield stronger force dependencies. Different values of M can fit the same set of data with different interaction intensities ΔG_d. (Values of the lines in this panel are: $\Delta G_d = 2.8\ k_BT$ for $M = 1$, $\Delta G_d = 2.0\ k_BT$ for $M = 2$, $\Delta G_d = 1.6\ k_BT$ for $M = 4$, $\Delta G_d = 1.5\ k_BT$ for $M = 8$.) **E)** Helicases are classified as active or passive according to their ability to destabilize the fork, parameterized by the interaction intensity ΔG_d. They are optimally active when ΔG_d is of the order of the higher base pair binding energy ΔG_{GC}. The coordinate operation of a polymerase and a helicase can increase the effective interaction intensity ΔG_d. **F)** Helicase with steps δ larger than one require the simultaneous opening of δ base pairs, implying a stronger tension dependence. (Active $\Delta G_d = 1.2\ k_BT$, $M = 6$; passive $\Delta G_d = 0$.) (Panels E and F are from [47].) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Assuming smaller destabilization energy for the more distant base pairs seems reasonable but it opens the questions of how fast is this decrease). See Fig. 5C, D, and E for further insight on the implications of the different values of the parameters ΔG_d and *M* to the V_{sd}/V_{pe} replication ratio.

Note that some effects that are detrimental for the primer extension replication might not be present in the strand displacement replication. For example, the formation of secondary structure is prevented by the presence of the fork, and we expect this effect to be absent in the $V_{pe}(f)$ used to compute the strand displacement replication velocity $V_{sd}(f)$ in Eq. (12). This description of the Betterton and Julicher model adapted for DNA polymerases, Eqs. (12)-(14), considers that replication occurs one nucleotide at a time and assumes that there is no significant back-stepping, $k_- \neq 0$, (i.e., due to exonucleolysis), Fig. 5B. How to include these additional effects (and others) is discussed on Refs. [47,60].

Fits of the force and sequence dependencies of the strand displacement rates of T4 and Phi29 DNA polymerases with this model revealed the interaction energy of each polymerase with the fork, $\Delta G_d = 1.6 - 2\ K_BT$ respectively [59,69]. Interpretation of the single-molecule data together with biochemical and structural information on polymerase-DNA complexes suggested that the ability of DNA polymerases to unwind DNA during replication depends on two competing processes: On the one hand, binding

and bending of the template strand by the DNA polymerase generates mechanical stress at the fork junction, which forces the separation of the dsDNA strands. On the other hand, the complementarity between the template and the displaced strands generates a regression pressure on the enzyme that competes for template binding, which prevents further polymerization and shifts the equilibrium towards the exonuclease conformation.

Similarly, single-molecule manipulation experiments have shown that the DNA unwinding rate of replicative DNA helicases is strongly affected by the stability of the fork [47,55,60]. The effect of fork stability on the unwinding rate can be explained using the Betterton and Julicher model explained above. In fact, this model was originally developed to quantify the effect of fork stability on the unwinding rate of helicases [8,47]. The same schemes as in Fig. 5A and B can be used for helicase, just accounting that the helicase opens the fork but does not convert ssDNA into dsDNA. Thus, the ratio between the DNA unwinding and translocation rates is obtained just replacing V_{sd}/V_{pe} by $V_{unwinding}/V_{translocation}$ in Eq. (12).

In fact this model was originally developed for helicases [8,47]. Quantification of the DNA destabilization energy by helicases using this model is not straightforward because of the uncertainty of helicase step size and the significant probability of backsliding events. Helicases can have large backstepping $k_{-}\neq 0$. The effect of varying the step size, δ, is shown in Fig. 5F. Increasing the step size, δ, leads to stronger force dependence (Fig. 5F), as also does an increase of the interaction range, M (Fig. 5D) (See Ref. [60] for a more detailed discussion). In any case, the strong dependency of the average unwinding rate of replicative helicases on DNA sequence and mechanical tension (fork stability), suggest that, when working in isolation, these enzymes present weak DNA destabilization energies. Interestingly, helicases may need the assistance of other partner proteins at the fork [59].

Ligands attached to dsDNA or to ssDNA can act as inhibitors or activators of strand displacement DNA replication or DNA unwinding. On the inhibitory side, ligands might represent a barrier if attached to dsDNA stabilizing it and preventing the fork opening. In this case, their effects must be accounted in the base pair stability term, ΔG_{bp}, in the model presented in Subsection 3.3. Ligands might also act as activators of the strand displacement replication, lowering the effective base pair binding energy ΔG_{bp}. Ligands attached to the lagging ssDNA (as SSBs) can help fork opening, or simply inhibit rezipping, effectively increasing the active opening of the fork ΔG_d. This inhibitory or activatory role of the ligand can be force modulated through a force dependent transition, as in primer extension, Eq. (10).

4. Conclusions

Dynamic biological processes such as DNA replication and DNA unwinding are inherently stochastic processes. Single-molecule manipulation and detection techniques allow researchers following the progress of an individual molecule and measuring the instantaneous rates and their fluctuations (such as pauses). These fluctuations can provide detailed information about the underlying kinetic and mechano-chemical cycle that governs the behavior of the motor under study. Extracting this information from experiments requires the ability to identify the pause states and their proper quantification.

The data analysis methods presented here approach the analysis of individual replication trajectories obtained with single-molecule manipulation methods differently. The direct pause identification approach is based on local analysis of the trajectory comparing the positions in a time interval to the positions in the following time interval to identify steps and plateaus. Instead, the velocity histogram introduces a more global approach, making the histogram of local velocities, but for the whole trajectory. In this method, the resolution of the two velocity peaks of the whole trajectory allows us to optimize the local average time windows (which filters high frequency noise) and the local velocity time window. Similar comments apply to the first passage time distribution analysis, which takes instead displacement windows. The prominence method combines the local selection of the velocity peaks by its prominence, with the invariance properties of the velocity peak mean along the whole trajectory. These methods combining local and global approaches extract reliable information from noisy trajectories. The key is performing a proper accumulation of small local evidences. Therefore, their philosophy is similar to the Bayesian methods, but from a more phenomenological or model independent approach (to a certain extent). We think there is still room for improvements to trayectory data analysis with new methods using phenomenological approaches. For example, new methods may arise from the mathematical study of the combination of quite general molecular motor models with the Bayesian approach. These mathematical developments could propose improved estimators to extract the relevant biological magnitudes from the trajectories (as the maximum replication velocity).

Models provide hypothesis of possible mechanisms for the DNA replication process, and the effects of tension, coordination with helicases, or the role of ligands as SSBs. The fit of the experimental data to the models allows us to check these hypotheses and see whether observations are compatible with the proposed mechanism. New models allow to explore new potential mechanisms or further details of the processes.

Further progress in the theory of binding of ligands to long polymers is required to complete the understanding of the observed multimode SSB binding to DNA [45,72,77] and mutual interaction between SSB binding and DNA replication [72,17]. Although the basis for the computation of the equilibrium coverage of ligands bound to a long polymer has been stablished by Mc Ghee and von Hippel, Ref. [61], there is still the need to develop a complete theory for the mechanics, kinetics, and thermodynamics of these systems. Mechanical models and simple kinetic and thermodynamic models have already been developed for one and two modes of binding [45]. However, recent results show that accurate description of the binding process (at medium or high coverages) requires a detailed count of the binding possibilities [100].

The analysis and modeling techniques described in this review have proven to be useful for the interpretation of the activities of replisome components when working in isolation (or in pairs) and importantly, have set the stage for the analysis of replication traces of fully reconstituted replisomes in the future. Analysis and modeling of single-molecule manipulation data will evolve hand by hand with the development of new methods that enable increased resolution, multiplexing or access to measure multiple variables of the system at the same time [40,20,2]. The methods described in this document have been used or could be adapted to study RNA polymerases [1,33,32,19,25,91] and other molecular motors [98,90].

Author contributions

FJCG wrote the first draft. JJ prepared the figures. BI wrote the first part of the introduction and made general improvements

and suggestions. All authors made relevant suggestions to improve the paper text and figures.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the European Regional Development Fund (ERDF) and by the Spanish Ministry of Economy and Competitiveness [BFU2015-63714-R and PGC2018-099341-B-I00 to B.I. and FIS2015-67765-R and RTI2018-095802-B-I00 to F.J.C.G.].

References

[1] Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM. Direct observation of base-pair stepping by RNA polymerase. Nature 2005;438 (7067):460–5. https://doi.org/10.1038/nature04268.
[2] Agarwal R, Duderstadt KE. Multiplex flow magnetic tweezers reveal rare enzymatic events with single molecule precision. Nat Commun 2020;11 (1):4714. https://doi.org/10.1038/s41467-020-18456-y.
[3] Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control 1974;19(6):716–23. https://doi.org/10.1109/TAC.1974.1100705.
[4] Andricioaei I, Goel A, Herschbach D, Karplus M. Dependence of DNA polymerase replication rate on external forces: a model based on molecular dynamics simulations. Biophys J 2004;87(3):1478–97. https://doi.org/10.1529/biophysj.103.039313.
[5] Bębenek A, Ziuzia-Graczyk I. Fidelity of DNA replication—a matter of proofreading. Curr Genet 2018;64(5):985–96. https://doi.org/10.1007/s00294-018-0820-1.
[6] Benkovic SJ, Valentine AM, Salinas F. Replisome-Mediated DNA Replication. Annu Rev Biochem 2001;70(1):181–208. https://doi.org/10.1146/annurev.biochem.70.1.181.
[7] Berman AJ, Kamtekar S, Goodman JL, Lázaro JM, de Vega M, Blanco L, et al. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. EMBO J 2007;26 (14):3494–505. https://doi.org/10.1038/sj.emboj.7601780.
[8] Betterton M, Jülicher F. Opening of nucleic-acid double strands by helicases: Active versus passive opening. Phys Rev E 2005;71(1):. https://doi.org/10.1103/PhysRevE.72.029906029906.
[9] Bosco A, Camunas-Soler J, Ritort F. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. Nucleic Acids Res 2014;42(3):2064–74. https://doi.org/10.1093/nar/gkt1089.
[10] Burgers PMJ, Kunkel TA. Eukaryotic DNA Replication Fork. Annu Rev Biochem 2017;86(1):417–38. https://doi.org/10.1146/annurev-biochem-061516-044709.
[11] Burnham DR, Kose HB, Hoyle RB, Yardimci H. The mechanism of DNA unwinding by the eukaryotic replicative helicase. Nat Commun 2019;10 (1):2159. https://doi.org/10.1038/s41467-019-09896-2.
[12] Burnham KP, Anderson DR. Model Selection and Inference. Model Select Infer 1998. https://doi.org/10.1007/978-1-4757-2917-7. Springer, New York, New York, NY.
[13] Bustamante C, Keller D, Oster G. The Physics of Molecular Motors. Acc Chem Res 2001;34(6):412–20. https://doi.org/10.1021/ar0001719.
[14] Camunas-Soler J, Ribezzi-Crivellari M, Ritort F. Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy. Annu Rev Biophys 2016;45 (1):65–84. https://doi.org/10.1146/annurev-biophys-062215-011158.
[15] Canceill D, Viguera E, Ehrlich SD. Replication Slippage of Different DNA Polymerases Is Inversely Related to Their Strand Displacement Efficiency. J Biol Chem 1999;274(39):27481–90. https://doi.org/10.1074/jbc.274.39.27481.
[16] Carter BC, Vershinin M, Gross SP. A Comparison of Step-Detection Methods: How Well Can You Do?. Biophys J 2008;94(1):306–19. https://doi.org/10.1529/biophysj.107.110601.
[17] Cerrón F, De Lorenzo S, Lemishko KM, Ciesielski GL, Kaguni LS, Cao FJ, et al. Replicative DNA polymerases promote active displacement of SSB proteins during lagging strand synthesis. Nucleic Acids Res 2019;47(11):5723–34. https://doi.org/10.1093/nar/gkz249.
[18] Chemla YR, Moffitt JR, Bustamante C. Exact Solutions for Kinetic Models of Macromolecular Dynamics †. J Phys Chem B 2008;112(19):6025–44. https://doi.org/10.1021/jp076153r.
[19] Cheng W, Arunajadai SG, Moffitt JR, Tinoco I, Bustamante C. Single-Base Pair Unwinding and Asynchronous RNA Release by the Hepatitis C Virus NS3 Helicase. Science 2011;333(6050):1746–9. https://doi.org/10.1126/science.1206023.
[20] Chuang C-Y, Zammit M, Whitmore ML, Comstock MJ. Combined High-Resolution Optical Tweezers and Multicolor Single-Molecule Fluorescence with an Automated Single-Molecule Assembly Line. J Phys Chem A 2019;123 (44):9612–20. https://doi.org/10.1021/acs.jpca.9b08282.
[21] Czerwinski F, Richardson AC, Oddershede LB. Quantifying noise in optical tweezers by allan variance. Opt Express 2009;17(15):13255–69. https://doi.org/10.1364/oe.17.013255.
[22] Depken M, Galburt EA, Grill SW. The Origin of Short Transcriptional Pauses. Biophys J 2009;96(6):2189–93. https://doi.org/10.1016/j.bpj.2008.12.3918.
[23] Desai VP, Frank F, Lee A, Righini M, Lancaster L, Noller HF, et al. Co-temporal Force and Fluorescence Measurements Reveal a Ribosomal Gear Shift Mechanism of Translation Regulation by Structured mRNAs. Mol Cell 2019;75(5):1007–1019.e5. https://doi.org/10.1016/j.molcel.2019.07.024.
[24] Douglas J, Kingston R, Drummond AJ. Bayesian inference and comparison of stochastic transcription elongation models. PLoS Comput Biol 2020;16 (2):1–21. https://doi.org/10.1371/journal.pcbi.1006717.
[25] Dulin D, Arnold JJ, van Laar T, Oh H-S, Lee C, Perkins AL, et al. Signatures of Nucleotide Analog Incorporation by an RNA-Dependent RNA Polymerase Revealed Using High-Throughput Magnetic Tweezers. Cell Reports 2017;21 (4):1063–76. https://doi.org/10.1016/j.celrep.2017.10.005.
[26] Dulin D, Lipfert J, Moolman MC, Dekker NH. Studying genomic processes at the single-molecule level: introducing the tools and applications. Nat Rev Genet 2013;14(1):9–22. https://doi.org/10.1038/nrg3316.
[27] Dulin D, Vilfan ID, Berghuis BA, Hage S, Bamford DH, Poranen MM, et al. Elongation-Competent Pauses Govern the Fidelity of a Viral RNA-Dependent RNA Polymerase. Cell Reports 2015;10(6):983–92. https://doi.org/10.1016/j.celrep.2015.01.031.
[28] Eddy SR. What is a hidden Markov model?. Nat Biotechnol 2004;22 (10):1315–6. https://doi.org/10.1038/nbt1004-1315.
[29] El Beheiry M, Türkcan S, Richly MU, Triller A, Alexandrou A, Dahan M, et al. A Primer on the Bayesian Approach to High-Density Single-Molecule Trajectories Analysis. Biophys J 2016;110(6):1209–15. https://doi.org/10.1016/j.bpj.2016.01.018.
[30] Elting MW, Spudich JA. Future Challenges in Single-Molecule Fluorescence and Laser Trap Approaches to Studies of Molecular Motors. Dev Cell 2012;23 (6):1084–91. https://doi.org/10.1016/j.devcel.2012.10.002.
[31] Flynn RL, Zou L. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. Crit Rev Biochem Mol Biol 2010;45 (4):266–75. https://doi.org/10.3109/10409238.2010.488216.
[32] Galburt EA, Grill SW, Bustamante C. Single molecule transcription elongation. Methods 2009;48(4):323–32. https://doi.org/10.1016/j.ymeth.2009.04.021.
[33] Galburt EA, Grill SW, Wiedmann A, Lubkowska L, Choy J, Nogales E, et al. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature 2007;446(7137):820–3. https://doi.org/10.1038/nature05701.
[34] Gao Y, Cui Y, Fox T, Lin S, Wang H, de Val N, et al. Structures and operating principles of the replisome. Science 2019;363(6429):eaav7003. https://doi.org/10.1126/science.aav7003.
[35] Gittes F, Schmidt CF. Signals and noise in micromechanical measurements. Methods Cell Biol 1998;55(55):129–56. https://doi.org/10.1016/s0091-679x(08)60406-9.
[36] Goel A, Frank-Kamenetskii MD, Ellenberger T, Herschbach D. Tuning DNA "strings": modulating the rate of DNA replication with mechanical tension. PNAS 2001;98(15):8485–9. https://doi.org/10.1073/pnas.151261198.
[37] Gollnick B. Optical and Magnetic Tweezers for Applications in Single-Molecule Biophysics and Nanotechnology. Universidad Autónoma de Madrid; 2014.
[38] Hacker KJ, Alberts BM. The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. J Biol Chem 1994;269(39):24221–8. https://doi.org/10.1016/S0021-9258(19)51071-7.
[39] Hamdan SM, Richardson CC. Motors, Switches, and Contacts in the Replisome. Annu Rev Biochem 2009;78(1):205–43. https://doi.org/10.1146/annurev.biochem.78.072407.103248.
[40] Heller I, Sitters G, Broekmans OD, Farge G, Menges C, Wende W, et al. STED nanoscopy combined with optical tweezers reveals protein dynamics on densely covered DNA. Nat Methods 2013;10(9):910–6. https://doi.org/10.1038/nmeth.2599.
[41] Hoekstra TP, Depken M, Lin S-N, Cabanas-Danés J, Gross P, Dame RT, et al. Switching between Exonucleolysis and Replication by T7 DNA Polymerase Ensures High Fidelity. Biophys J 2017;112(4):575–83. https://doi.org/10.1016/j.bpj.2016.12.044.
[42] Hua W, Young EC, Fleming ML, Gelles J. Coupling of kinesin steps to ATP hydrolysis. Nature 1997;388(6640):390–3. https://doi.org/10.1038/41118.
[43] Ibarra B, Chemla YR, Plyasunov S, Smith SB, Lázaro JM, Salas M, et al. Proofreading dynamics of a processive DNA polymerase. EMBO J 2009;28 (18):2794–802. https://doi.org/10.1038/emboj.2009.219.
[44] Jackson MB. Molecular and Cellular Biophysics. Cambridge: Cambridge University Press; 2006.
[45] Jarillo J, Morín JA, Beltrán-Heredia E, Villaluenga JPG, Ibarra B, Cao FJ. Mechanics, thermodynamics, and kinetics of ligand binding to biopolymers. (M. S. Kellermayer, ed.). PLOS ONE 2017;12(4):. https://doi.org/10.1371/journal.pone.0174830e0174830.

[46] Johnson A, O'Donnell M. Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem 2005;74:283–315. https://doi.org/10.1146/annurev.biochem.73.011303.073859.
[47] Johnson DS, Bai L, Smith BY, Patel SS, Wang MD. Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. Cell 2007;129(7):1299–309. https://doi.org/10.1016/j.cell.2007.04.038.
[48] Joo S, Chung BH, Lee M, Ha TH. Ring-shaped replicative helicase encircles double-stranded DNA during unwinding. Nucleic Acids Res 2019;47 (21):11344–54. https://doi.org/10.1093/nar/gkz893.
[49] Kaguni LS, Clayton DA. Template-directed pausing in in vitro DNA synthesis by DNA polymerase a from Drosophila melanogaster embryos. Proc Natl Acad Sci 1982;79(4):983–7.
[50] Keller D, Bustamante C. The Mechanochemistry of Molecular Motors. Biophys J 2000;78(2):541–56. https://doi.org/10.1016/S0006-3495(00)76615-X.
[51] Keller D, Swigon D, Bustamante C. Relating single-molecule measurements to thermodynamics. Biophys J 2003;84(2 Pt 1):733–8. https://doi.org/10.1016/S0006-3495(03)74892-9.
[52] Kerssemakers JWJ, Munteanu EL, Laan L, Noetzel TL, Janson ME, Dogterom M. Assembly dynamics of microtubules at molecular resolution. (Supplementary Methods: Step-fitting algorithm.). Nature 2006;442(7103):709–12.
[53] Kunkel TA, Bebenek K. DNA Replication Fidelity. Annu Rev Biochem 2000;69 (1):497–529. https://doi.org/10.1146/annurev.biochem.69.1.497.
[54] Liao JC, Spudich JA, Parker D, Delp SL. Extending the absorbing boundary method to fit dwell-time distributions of molecular motors with complex kinetic pathways. PNAS 2007;104(9):3171–6. https://doi.org/10.1073/pnas.0611519104.
[55] Lionnet T, Spiering MM, Benkovic SJ, Bensimon D, Croquette V. Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. PNAS 2007;104(50):19790–5. https://doi.org/10.1073/pnas.0709793104.
[56] Lipfert J, Hao X, Dekker NH. Quantitative Modeling and Optimization of Magnetic Tweezers. Biophys J 2009;96(12):5040–9. https://doi.org/10.1016/j.bpj.2009.03.055.
[57] Maier B, Bensimon D, Croquette V. Replication by a single DNA polymerase of a stretched single-stranded DNA. Proc Natl Acad Sci 2000;97(22):12002–7. https://doi.org/10.1073/pnas.97.22.12002.
[58] Manosas M, Spiering MM, Ding F, Bensimon D, Allemand J-F, Benkovic SJ, et al. Mechanism of strand displacement synthesis by DNA replicative polymerases. Nucleic Acids Res 2012;40(13):6174–86. https://doi.org/10.1093/nar/gks253.
[59] Manosas M, Spiering MM, Ding F, Croquette V, Benkovic SJ. Collaborative coupling between polymerase and helicase for leading-strand synthesis. Nucleic Acids Res 2012;40(13):6187–98. https://doi.org/10.1093/nar/gks254.
[60] Manosas M, Xi XG, Bensimon D, Croquette V. Active and passive mechanisms of helicases. Nucleic Acids Res 2010;38(16):5518–26. https://doi.org/10.1093/nar/gkq273.
[61] McGhee JD, von Hippel PH. Theoretical aspects of DNA-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. J Mol Biol 1974;86(2):469–89. https://doi.org/10.1016/0022-2836(74)90031-x.
[62] Medagli B, Onesti S. Structure and Mechanism of Hexameric Helicases. In: Spies M, editor. DNA Helicases and DNA Motor Proteins. New York, NY: Springer-Verlag; 2013. p. 75–95. https://doi.org/10.1007/978-1-4614-5037-5_4.
[63] Meselson M, Stahl FW. The replication of DNA in Escherichia coli. Proc Natl Acad Sci 1958;44(7):671–82. https://doi.org/10.1073/pnas.44.7.671.
[64] Michaelis J, Muschielok A, Andrecka J, Kügel W, Moffitt JR. DNA based molecular motors. Phys Life Rev 2009;6(4):250–66. https://doi.org/10.1016/j.plrev.2009.09.001.
[65] Miller H, Zhou Z, Shepherd J, Wollman AJM, Leake MC. Single-molecule techniques in biophysics: a review of the progress in methods and applications. Rep Prog Phys 2018;81(2):. https://doi.org/10.1088/1361-6633/aa8a02024601.
[66] Moffitt JR, Chemla YR, Bustamante C. Methods in Statistical Kinetics. In: Walter NG, editor. Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods. cop: Academic Press; 2010. p. 221–57. https://doi.org/10.1016/S0076-6879(10)75010-2.
[67] Moffitt JR, Chemla YR, Smith SB, Bustamante C. Recent Advances in Optical Tweezers. Annu Rev Biochem 2008;77(1):205–28. https://doi.org/10.1146/annurev.biochem.77.043007.090225.
[68] Monachino E, Spenkelink LM, van Oijen AM. Watching cellular machinery in action, one molecule at a time. J Cell Biol 2017;216(1):41–51. https://doi.org/10.1083/jcb.201610025.
[69] Morin JA, Cao FJ, Lazaro JM, Arias-Gonzalez JR, Valpuesta JM, Carrascosa JL, et al. Active DNA unwinding dynamics during processive DNA replication. Proc Natl Acad Sci 2012;109(21):8115–20. https://doi.org/10.1073/pnas.1204759109.
[70] Morin JA, Cao FJ, Lázaro JM, Arias-Gonzalez JR, Valpuesta JM, Carrascosa JL, et al. Mechano-chemical kinetics of DNA replication: identification of the translocation step of a replicative DNA polymerase. Nucleic Acids Res 2015;43(7):3643–52. https://doi.org/10.1093/nar/gkv204.
[71] Morin JA, Cao FJ, Valpuesta JM, Carrascosa JL, Salas M, Ibarra B. Manipulation of single polymerase-DNA complexes: A mechanical view of DNA unwinding during replication. Cell Cycle 2012;11(16):2967–8. https://doi.org/10.4161/cc.21389.
[72] Morin JA, Cerrón F, Jarillo J, Beltran-Heredia E, Ciesielski GL, Arias-Gonzalez JR, et al. DNA synthesis determines the binding mode of the human mitochondrial single-stranded DNA-binding protein. Nucleic Acids Res 2017;45(12):7237–48. https://doi.org/10.1093/nar/gkx395.
[73] Morin JA, Ibarra B, Cao FJ. Kinetic modeling of molecular motors: pause model and parameter determination from single-molecule experiments. J Stat Mech: Theory Exp 2016;2016(5):. https://doi.org/10.1088/1742-5468/2016/05/054031054031.
[74] Müllner FE, Syed S, Selvin PR, Sigworth FJ. Improved hidden Markov models for molecular motors, part 1: Basic theory. Biophys J 2010;99(11):3684–95. https://doi.org/10.1016/j.bpj.2010.09.067.
[75] Myers TW, Romano LJ. Mechanism of stimulation of T7 DNA polymerase by Escherichia coli single-stranded DNA binding protein (SSB). J Biol Chem 1988;263(32):17006–15. https://doi.org/10.1016/s0021-9258(18)37490-8.
[76] Nakai H, Richardson CC. The effect of the T7 and Escherichia coli DNA-binding proteins at the replication fork of bacteriophage T7. J Biol Chem 1988;263 (20):9831–9. https://doi.org/10.1016/s0021-9258(19)81592-2.
[77] Naufer, M. N., M. Morse, G. B. Möller, J. McIsaac, I. Rouzina, P. J. Beuning, and M. C. Williams. 2021. Multiprotein E. coli SSB-ssDNA complex shows both stable binding and rapid dissociation due to interprotein interactions. Nucleic acids research 49(3):1532–1549. 10.1093/nar/gkaa1267.
[78] Naufer MN, Murison DA, Rouzina I, Beuning PJ, Williams MC. Single-molecule mechanochemical characterization of E. coli pol III core catalytic activity. Protein Sci 2017;26(7):1413–26. https://doi.org/10.1002/pro.3152.
[79] Nong EX, DeVience SJ, Herschbach D. Minimalist model for force-dependent DNA replication. Biophys J 2012;102(4):810–8. https://doi.org/10.1016/j.bpj.2012.01.020.
[80] Oliveira MT, de Bovi Pontes C, Ciesielski GL. Roles of the mitochondrial replisome in mitochondrial DNA deletion formation. Genet Mol Biol 2020;13 (1):. https://doi.org/10.1590/1678-4685-GMB-2019-0069e20190069.
[81] Ostrofet E, Papini FS, Dulin D. Correction-free force calibration for magnetic tweezers experiments. Sci Rep 2018;8(1):15920. https://doi.org/10.1038/s41598-018-34360-4.
[82] Pandey M, Patel SS. Helicase and polymerase move together close to the fork junction and copy DNA in one-nucleotide steps. Cell Rep 2014;6(6):1129–38. https://doi.org/10.1016/j.celrep.2014.02.025.
[83] Patel SS, Picha KM. Structure and Function of Hexameric Helicases. Annu Rev Biochem 2000;69(1):651–97. https://doi.org/10.1146/annurev.biochem.69.1.651.
[84] Qian H, Kou SC. Statistics and related topics in single-molecule biophysics. Annu Rev Stat Appl 2014;1:465–92. https://doi.org/10.1146/annurev-statistics-022513-115535.
[85] Rabiner LR. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc IEEE 1989;77(2):257–86. https://doi.org/10.1109/5.18626.
[86] Redner S. A Guide to First-Passage Processes. Cambridge University Press; 2001.
[87] Reyes-Lamothe R, Possoz C, Danilova O, Sherratt DJ. Independent Positioning and Action of Escherichia coli Replisomes in Live Cells. Cell 2008;133 (1):90–102. https://doi.org/10.1016/j.cell.2008.01.044.
[88] Ribeck N, Kaplan DL, Bruck I, Saleh OA. DnaB helicase activity is modulated by DNA geometry and force. Biophys J 2010;99(7):2170–9. https://doi.org/10.1016/j.bpj.2010.07.039.
[89] Ribeck N, Saleh OA. DNA unwinding by ring-shaped T4 helicase gp41 is hindered by tension on the occluded strand. PLoS ONE 2013;8(11):. https://doi.org/10.1371/journal.pone.0079237e79237.
[90] Rief M, Rock RS, Mehta AD, Mooseker MS, Cheney RE, Spudich JA. Myosin-V stepping kinetics: A molecular model for processivity. Proc Natl Acad Sci 2000;97(17):9482–6. https://doi.org/10.1073/pnas.97.17.9482.
[91] Righini M, Lee A, Cañari-Chumpitaz C, Lionberger T, Gabizon R, Coello Y, et al. Full molecular trajectories of RNA polymerase at single base-pair resolution. Proc Natl Acad Sci 2018;115(6):1286–91. https://doi.org/10.1073/pnas.1719906115.
[92] Schwartz JJ, Quake SR. Single molecule measurement of the "speed limit" of DNA polymerase. Proc Natl Acad Sci 2009;106(48):20294–9. https://doi.org/10.1073/pnas.0907404106.
[93] Shereda RD, Kozlov AG, Lohman TM, Cox MM, Keck JL. SSB as an Organizer/Mobilizer of Genome Maintenance Complexes. Crit Rev Biochem Mol Biol 2008;43(5):289–318. https://doi.org/10.1080/10409230802341296.
[94] Steitz TA. DNA Polymerases: Structural Diversity and Common Mechanisms. J Biol Chem 1999;274(25):17395–8. https://doi.org/10.1074/jbc.274.25.17395.
[95] Sun B, Johnson DS, Patel G, Smith BY, Pandey M, Patel SS, et al. ATP-induced helicase slippage reveals highly coordinated subunits. Nature 2011;478 (7367):132–5. https://doi.org/10.1038/nature10409.
[96] Sun B, Wang MD. Single-molecule perspectives on helicase mechanisms and functions. Crit Rev Biochem Mol Biol 2016;51(1):15–25. https://doi.org/10.3109/10409238.2015.1102195.
[97] Suter, D. M., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. Science (New York, N.Y.) 2011; 332(6028):472–4. 10.1126/science.1198817.
[98] Svoboda K, Schmidt CF, Schnapp BJ, Block SM. Direct observation of kinesin stepping by optical trapping interferometry. Nature 1993;365(6448):721–7. https://doi.org/10.1038/365721a0.
[99] Van Oijen AM, Loparo JJ. Single-molecule studies of the replisome. Annu Rev Biophys 2010;39(1):429–48. https://doi.org/10.1146/annurev.biophys.093008.131327.

[100] Villaluenga JPG, Vidal J, Cao-García FJ. Noncooperative thermodynamics and kinetic models of ligand binding to polymers: Connecting McGhee–von Hippel model with the Tonks gas model. Phys Rev E 2020;102(1):. https://doi.org/10.1103/PhysRevE.102.012407012407.
[101] Walcott S. The load dependence of rate constants. J Chem Phys 2008;128 (21):1–10. https://doi.org/10.1063/1.2920475.
[102] Wang J, Sattar AKMA, Wang CC, Karam JD, Konigsberg WH, Steitz TA. Crystal Structure of a pol α Family Replication DNA Polymerase from Bacteriophage RB69. Cell 1997;89(7):1087–99. https://doi.org/10.1016/S0092-8674(00)80296-2.
[103] Wen J-D, Lancaster L, Hodges C, Zeri A-C, Yoshimura SH, Noller HF, et al. Following translation by single ribosomes one codon at a time. Nature 2008;452(7187):598–603. https://doi.org/10.1038/nature06716.
[104] Wuite GJL, Smith SB, Young M, Keller D, Bustamante C. Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. Nature 2000;404(6773):103–6. https://doi.org/10.1038/35003614.
[105] Yakovchuk P, Protozanova E, Frank-Kamenetskii MD. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res 2006;34(2):564–74. https://doi.org/10.1093/nar/gkj454.