Kinetic modeling of molecular motors: pause model and parameter determination from single molecule experiments

José A. Morin[1], Borja Ibarra[1], Francisco J. Cao[2,*]
[1]*Instituto Madrileño de Estudios Avanzados en Nanociencia (IMDEA Nanociencia) & CNB-CSIC-IMDEA Nanociencia Associated Unit 'Unidad de Nanobiotecnología', 28049 Madrid, Spain*
[2]*Universidad Complutense de Madrid. Facultad de Ciencias Físicas. Departamento Física Atómica, Molecular y Nuclear, Av. Complutense s/n. 28040 Madrid. Spain.*
[*]*francao@fis.ucm.es*

Abstract
Single molecule manipulation experiments of molecular motors provide essential information about the rate and conformational changes of the steps of the reaction located along the manipulation coordinate. This information is not always sufficient to define a particular kinetic cycle. Recent single molecule experiments with optical tweezers showed that the DNA unwinding activity of a Phi29 DNA polymerase mutant presents a complex pause behavior, which includes short and long pauses. Here we show that different kinetic models, considering different connections between the active and the pause states can explain the experimental pause behavior. Both the two independent pause model and the two connected pause model are able to describe the pause behavior of a mutated Phi29 DNA polymerase observed in an optical tweezers single-molecule experiment. For the two independent pause model all parameters are fixed by the observed data. While for the more general two connected pause model there is a range of values of the parameters compatible with the observed data (which can be expressed in terms of two of the rates and its force dependencies). This general model includes models with indirect entry and exit to the long pause state, and also models with cycling in both directions. Additionally assuming that detailed balance is verified, which forbids cycling, reduces the ranges of the values of the parameters (which can then be expressed in terms of one rate and its force dependency). The resulting model interpolates between the independent pause model and the indirect entry and exit to long pause state model

I. INTRODUCTION

Last decades have seen amazing advances in the development of experimental techniques focused on the study of biological processes at single molecule level[1,2,3]. These, so called, single-molecule techniques allow obtaining information about the position and structure of functional biological complexes in real time. In particular, they have been widely used to explore the operation of biological molecular motors. Molecular motors are proteins that operate as transducers of chemical energy into mechanical work through conformational changes and unidirectional displacements. They perform many different tasks in the cell, which range from DNA replication to intracellular cargo transport, and even cell locomotion.

A characteristic property of the molecular motors operation is that the energy they interchange with the environment is of the same order of magnitude as that of the thermal fluctuations around them. Thus, biochemical reactions such as, DNA replication, are thermally driven processes involving the stochastic crossing of a barrier along the reaction coordinate. Therefore, at the molecular level thermal fluctuations govern the dynamics of individual proteins, which display a fast and stochastic behavior. This 'noisy' or stochastic behavior observed in single molecule data, is often an integral part of the mechanism of the reaction and contains important information over the reaction itself. Unfortunately, in single-molecule

experiments one has also to deal with thermal (and mechanical) fluctuations which affect accuracy of the measurements. These 'external' fluctuations add unwanted thermal and mechanical noise, partially masking the actual dynamics of the system under study. Thus, stochastic processes are very present in single molecule experiments with molecular motors.

In addition, the study of molecular motors using single molecule techniques is limited to monitor the time evolution of a particular spatial reaction coordinate, which is imposed by the experimental geometry. From the limited information contained in this coordinate, one has to infer the dynamics of the system, by determining the correct kinetic model and the value of its parameters. Here, based on our previous single molecule characterization of the Phi29 DNA polymerase activity, we show that this limited information can lead to ambiguities in the determination of the correct kinetic model and the values of its parameters.

II. THE PAUSE BEHAVIOUR MODULATES THE AVERAGE REPLICATION VELOCITY

The Phi29 DNA polymerase is a replicative DNA polymerase (DNAp) that has been widely used as a model system to determine the biochemical processes responsible for DNA replication[4]. Replicative DNA polymerases read the base composition of one strand of the DNA (template strand) and synthetize a new strand, by adding the corresponding complementary nucleotides in each position (primer extension mode). In addition to DNA synthesis, the Phi29 DNAp has a particular property; it is able to unwind the DNA helix as it advances along the DNA(strand displacement mode).

Recently, we described a series of single molecules experiments designed to understand the molecular mechanism used by the Phi29 DNAp to unwind the DNA and to couple the processes of DNA replication and unwinding[5,6] . To this end, using optical tweezers, we measured the combined effect of mechanical force, favoring the DNA unwinding reaction and, DNA sequence on the real time kinetics of single polymerases molecules replicating through a DNA hairpin. Figure 1 (Top) shows an schematic representation of the experimental configuration used in our studies: Initially, during the strand displacement phase, the protein moved through the DNA hairpin doing replication and DNA unwinding at the same time. These reactions modified the DNA structure, which in turn increased the distance between the beads (Δx1, Fig. 1). Upon reaching the end of the hairpin, no more DNA unwinding is required, and the polymerase continues replication in the primer extension mode, converting the displaced complementary single stranded DNA to double stranded DNA. This reaction also changes the distance between the beads (Δx2, Fig. 1) due to the different elastic properties of each DNA polymer. In addition, to get a better understanding of the unwinding mechanism we compared the activities of two Phi29 DNAp: the wild-type version and an unwinding deficient mutant (strand displacement deficient). The mutant Phi29 DNAp presents a couple of mutations which inhibit the characteristic DNA unwinding mechanism of this polymerase. Remarkably, we showed that application of increasing mechanical forces favoring the DNA unwinding reaction 'rescued' the unwinding activity of the mutant polymerase; At low forces this protein is not able to open the DNA however, at the highest forces (below 12 pN) the mutant opens the DNA with a rate similar to the wild-type polymerase.

A close inspection of the individual replication trajectories showed an interesting difference between the wild-type and the mutant activities. The activity of the wild-type polymerase is only interrupted by short pauses; while the activity of the mutant during unwinding presents short and long pauses[5,6], see Fig.1, (Bottom). The presence of these long pauses is responsible for the lower average replication velocity of the mutant polymerase during the strand displacement phase (at the lowest tensions)

A detailed analysis of the pause behavior revealed that the frequency and duration of the long pauses decrease exponentially with force. These observations indicate that mechanical force aiding the unwinding reaction decreases the entry and favors the exit rates to and from the long pause state. In addition, we found that long pauses occur with higher probabilities at DNA positions with the higher GC content (the position where the helical structure is more stable). These evidences strongly suggest that the inactive long pause state characteristic of the mutant polymerase is caused by the stability of the DNA helix in front of the protein.

The shorter pauses were found with similar frequencies during primer extension (no need of DNA unwinding) and strand displacement (DNA unwinding required) conditions, for both the wild-type and mutant polymerases, indicating they are not related to the process of DNA unwinding. Instead, short pauses are probably related to the nucleotide incorporation reaction during the replication reaction.

Here we show that different kinetic models, considering different connections between the active and pause states can explain the experimental observations. In this case, similarly to other single molecule characterizations, additional biochemical and structural information about the biological system is required to favor a particular model.

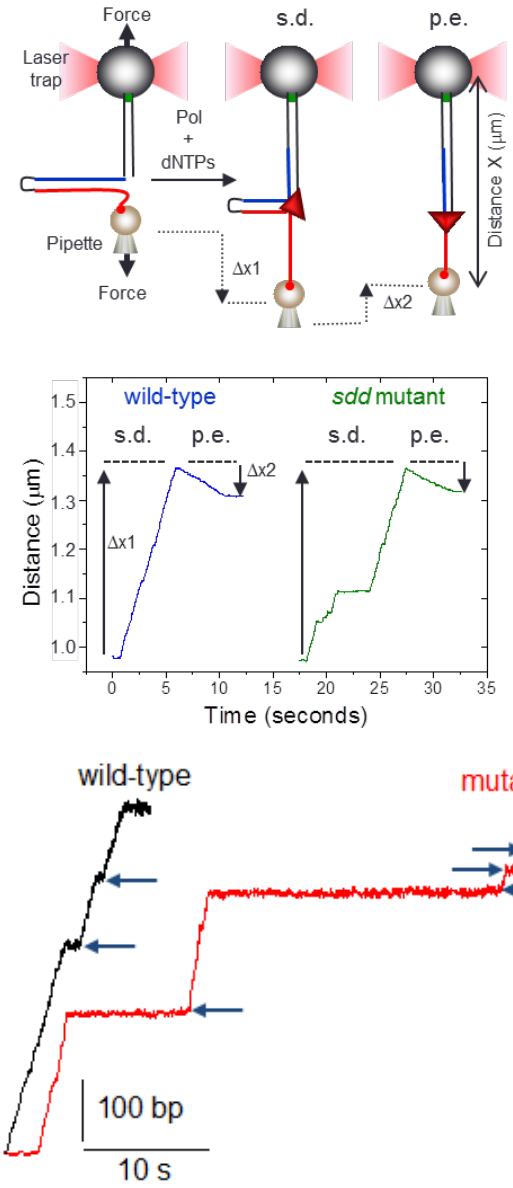

FIG. 1. **(Top)** Schematic representation of the experimental design (from Ref. 5), a single DNA hairpin was tethered to functionalized beads inside a fluidics chamber. One strand of the hairpin (blue) is attached through a dsDNA handle to a bead held in the laser trap, while the complementary strand (red) is attached to a bead on top of a mobile micropipette. In this configuration external force aids the separation of the two strands of the DNAhairpin. At a constant force the strand displacement (s.d.) and primer extension (p.e.) activities of the polymerase (red triangle) are detected as a change in distance between the beads, Δx1 and Δx2, respectively.
(Middle) (from ref. 5) Representative replication traces of the wild-type (blue) and strand displacement

deficient or *sdd* mutant (green) polymerases showing the distance changes during strand displacement (s.d., Δx1) and primer extension (p.e., Δx2).
(Bottom) Representative replication traces of the wild-type (black) and mutant (red) polymerases. Arrows show the position of pause events. Y-axis represents the number of replicated base pairs (bp) and X-axis is the time (seconds).

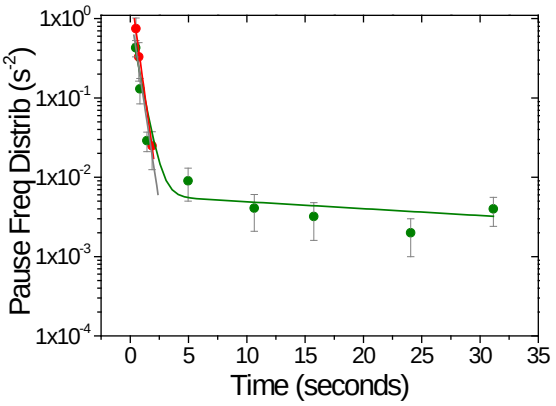

FIG. 2. Pause length frequency distribution for the mutant polymerase. During strand displacement ($F \sim 6.5pN$, green dots) a double exponential is required to fit the data (green solid line). During primer extension ($F \sim 6.5pN$, red dots) the results are compatible with a single exponential (solid red line), for comparison the wild-type distribution is shown as a grey line. (From Ref. 5)

III. KINETIC MODELS FOR THE PAUSE BEHAVIOR

During the strand displacement phase the pause length frequency distribution of the mutant polymerase presents two slopes, corresponding to two temporal scales, Fig. 2. This behavior indicates the presence of at least two different types of pauses, classified as short pauses (typically less than 4 s) and long pauses (typically greater than 4 s). In this case, the pause length frequency distribution can be fit by a two exponential distribution of the form

$$f(t) = \tilde{a}_s \cdot e^{\lambda_s \cdot t} + \tilde{a}_l \cdot e^{\lambda_l \cdot t}$$
$$= a_s \cdot (-\lambda_s) \cdot e^{\lambda_s \cdot t} + a_l \cdot (-\lambda_l) \cdot e^{\lambda_l \cdot t},$$

where $\tilde{a}_s$ and $\tilde{a}_l$ give the frequency of appearance of short and long pauses (respectively) of length t per unit of pause length, while λ_s and λ_l give the characteristic time scale of short and long pauses (respectively). Introducing a_s and a_l is convenient in order to get simpler expressions in terms of the transition rates, as shown below. The pause length frequency distribution $f(t)$ gives the probability of finding a pause of length t per unit of time without pause, and per unit of pause length (this is why it is a distribution). [Thus, $f(t)dt$ gives the quotient between the time the system is in pauses with length between t and $t + dt$ and the time without pauses.] The pause probability $P(t)$ gives instead the probability of finding the system in the pause state a time t after it has entered into pause. The relation between both functions is $f(t) = -k_{ap}\frac{dP(t)}{dt}$, with k_{ap} the rate of entrance to any pause state. Pause length frequency distribution $f(t)$ is more informative because it additionally depends on the entrance rate into pause states, which is not the case for the pause probability $P(t)$, as $P(t=0) = 1$ while $f(t=0) \neq 1$.

Different kinetic models may, in principle, explain the observed pause behavior for the mutant polymerase. Here we show that two different models can be used to fit the experimental data: the two independent pause states model, where the two pause states are independently connected to the active state, see Fig. 3, and the two connected pause states model, which allows a direct connection between the two pause states, Fig. 4. We also show below that the two connected pause model is more general and contains as particular cases: the previous independent pause model (for $k_{12} = k_{21} = 0$), and also models where one of the pause state is not directly connected to the active state (for $k_{a2} = k_{2a} = 0$ or $k_{a1} = k_{1a} = 0$).

FIG. 3. Two independent pause states model. k_{ij} is the transition rate between the state i and the state j, where $i, j = a, 1, 2$, with a indicating the active state, 1 the pause state $P1$, and 2 the pause state $P2$.

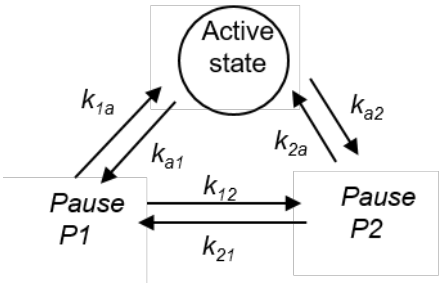

FIG. 4. Two connected pause states model. k_{ij} is the transition rate between the state i and the state j, where $i,j = a,1,2$, with a indicating the active state, 1 the pause state $P1$, and 2 the pause state $P2$.

A. TWO INDEPENDENT PAUSE STATES MODEL

The two independent pause states model assumes that the pause states are directly connected to the active state and that they are not connected between them, Fig. 3. This model leads to a two exponential pause length frequency distribution

$$f(t) = a_s \cdot (-\lambda_s) \cdot e^{\lambda_s \cdot t} + a_l \cdot (-\lambda_l) \cdot e^{\lambda_l \cdot t},$$

with

$$a_s = k_{a1},$$
$$a_l = k_{a2},$$
$$\lambda_s = -k_{1a},$$
$$\lambda_l = -k_{2a}.$$

The fits to the experimental pause length frequency distribution gives the rates at zero force $k_{ij}(0)$, and their force dependencies d_{ij} defined by

$$k_{ij}(F) = k_{ij}(0) \cdot e^{\frac{-d_{ij}F}{k_B T}},$$

are given in Tables 1 and 2. The derivation of these results and their structural and biochemical implications are discussed in detail in Ref. 5. We note that the uncertainty found for the values of these parameters is high, ranging between the 30 and 100% indicating that the values have one significant digit. Therefore, the results reported in the next sections using these values will also have only one significant digit.

	rate (s^{-1})
$k_{a1}(0)$	1.5±0.5
$k_{1a}(0)$	1.4±1.2
$k_{a2}(0)$	0.8±0.6
$k_{2a}(0)$	0.02±0.02

TABLE 1: Transition rates at zero force for the two independent states model (from Ref. 5).

	Force dependencies (nm)
d_{a1}	0.4±0.2
d_{1a}	0
d_{a2}	0.5±0.3
d_{2a}	-0.7±0.4

TABLE 2: Force dependencies of transition rates for the two independent states model (from Ref. 5). The negative value of d_{2a} indicates that the transition rate $k_{2a}(F)$ (*i.e.*, the exit rate of pause state $P2$ to the active state), increases instead of decreasing with force, (in contrast to the force dependencies measured for the other rates).

The two independent pause models is the simplest model explaining the data. However, in the next subsection we explicitly show that other kinetic models can also explain or fit the experimental data.

B. TWO CONNECTED PAUSE STATES MODEL

The double exponential form of the pause length frequency distribution is also obtained when transitions between the pause states (P1 and P2) are allowed (Ref. 7.) In this case, these transitions imply two additional rates, which in turn may also depend on force (see Fig. 4). Therefore, this model considers 4 additional parameters (when compared with the two independent pause states model).

The coefficients of the two exponential functions for the pause frequency distribution in terms of the rates of this model are (Ref. 7)

$$f(t) = a_s \cdot (-\lambda_s) \cdot e^{\lambda_s \cdot t} + a_l \cdot (-\lambda_l) \cdot e^{\lambda_l \cdot t},$$

with

$$a_s = \frac{k_{1a}k_{a1} + k_{2a}k_{a2} + (k_{a1} + k_{a2})\lambda_s}{\lambda_s - \lambda_l}$$

$$a_l = \frac{-k_{1a}k_{a1} - k_{2a}k_{a2} - (k_{a1} + k_{a2})\lambda_l}{\lambda_s - \lambda_l}$$

$$\lambda_s = \frac{1}{2}(-k_{12} - k_{1a} - k_{21} - k_{2a} - \sqrt{\Delta})$$
$$\lambda_l = \frac{1}{2}(-k_{12} - k_{1a} - k_{21} - k_{2a} + \sqrt{\Delta})$$
where $\Delta = (k_{12} + k_{1a} + k_{21} + k_{2a})^2 - 4(k_{1a}k_{21} + k_{12}k_{2a} + k_{1a}k_{2a})$.

Note that the pause length frequency distribution $f(t)$ used here is related to the pause probability $P(t)$ described in Ref. 7, through $f(t) = -(k_{a1} + k_{a2})\, dP(t)/dt$. This implies that the a_i coefficients described here can be expressed in terms of the coefficients $\hat{a}_{i,Jackson}$ of the probability function used in Ref. 7, through $a_i = (k_{a1} + k_{a2}) \cdot \hat{a}_{i,Jackson}$, while the pause length parameters λ_i have the same expressions. We recall that the pause length frequency distribution is more informative, as it contains the entrance rate into pause states, which is not contained in the pause probability as $P(t=0) = \hat{a}_{s,Jackson} + \hat{a}_{l,Jackson} = 1$, while $f(t=0) \neq 1$.

The previous expressions for the observed quantities a_s, a_l, λ_s and λ_l imply that giving two of the rates, the other four rates can be determined

$$k_{a1} = \frac{-a_l k_{2a} - a_s k_{2a} - a_l \lambda_l - a_s \lambda_s}{k_{1a} - k_{2a}},$$
$$k_{21} = \frac{(k_{2a} + \lambda_l)(k_{2a} + \lambda_s)}{k_{1a} - k_{2a}},$$
$$k_{a2} = \frac{a_l k_{1a} + a_s k_{1a} + a_l \lambda_l + a_s \lambda_s}{k_{1a} - k_{2a}},$$
$$k_{12} = \frac{-(k_{1a} + \lambda_l)(k_{1a} + \lambda_s)}{k_{1a} - k_{2a}}.$$

All the other rates can be expressed in terms of the rates k_{1a} and k_{2a}, and the values of the rates of the observed quantities a_s, a_l, λ_s and λ_l.

Using these expressions, the positiveness of all the rates ($k_{ij} \geq 0$) implies

$$\frac{-a_l \lambda_l - a_s \lambda_s}{a_l + a_s} < k_{1a} < -\lambda_s \; and$$
$$0 < k_{2a} < -\lambda_l,$$

which in our case for zero force leads to

$$0.9 s^{-1} < k_{1a}(0) < 1.4 s^{-1} \text{ and}$$
$$0 < k_{2a}(0) < 0.02 s^{-1}.$$

These ranges of values of $k_{1a}(0)$ and $k_{2a}(0)$ can be used to directly derive the ranges of values for all the other rates at zero force, shown in Table 3. It would also be valid to interchange the values of $k_{1a}(0)$ and $k_{2a}(0)$, however this will exchange the other rates too, finally giving the same result.

	Minimum (s^{-1})	Maximum (s^{-1})
$k_{a1}(0)$	1.5	2.3
$k_{1a}(0)$	0.9	1.4
$k_{21}(0)$	0	0.03
$k_{12}(0)$	0	0.5
$k_{a2}(0)$	0	0.8
$k_{2a}(0)$	0	0.02

TABLE 3: Extreme (*i.e.*, minimum and maximum) values of the rates at zero force. (Note that these values are obtained from the most probable values in Table 1, considering uncertainties in these values would increase the range, even double it for some of them.)

The solution $k_{1a}(0) = 1.4\, s^{-1}$ and $k_{2a}(0) = 0.02\, s^{-1}$ recovers the unconnected two pauses model of previous subsection. Other possible solutions include:

1. Indirect entry to and exit from pause P2, which yields the fastest interconversion rates between P1 and P2 pause states (Table 4).
2. Cycling in $a \rightarrow P1 \rightarrow P2$ direction, as well as in the opposite direction, see Table 4. Cycling violates detailed balance since implies a net driving Gibbs energy ΔG_c for the cycle, given by

$$\frac{k_{a1}}{k_{1a}} \frac{k_{12}}{k_{21}} \frac{k_{2a}}{k_{a2}} = e^{\frac{-\Delta G_c}{k_B T}}.$$

ΔG_c may arise from the nucleotide hydrolysis reaction, which would supply as much as 30 K_BT per nucleotide hydrolysis. This energy could, in principle, promote cycling in a particular direction. However, the large conformational changes (or force dependencies) associated with the entry/exit to/from the pause states indicate that pauses correspond to off-pathway, conformationally inactive polymerase-DNA complexes, branching off the main active state of the protein during replication. In other words, pauses are not integrated within the nucleotide

incorporation cycle, arguing against the possibility of a nucleotide-fueled cycling.

It is interesting to note that all three extreme models of Table 4 involve an exit rate from pause state 2 of the order of $0.02\ s^{-1}$, which is one order of magnitude smaller than the other rates, and it is what leads to the observed long pauses.

Rates (s^{-1})	Indirect entry and exit to pause 2.	$a \to 1 \to 2$ cycling	$a \to 2 \to 1$ cycling
$k_{a1}(0)$	2.3	2.3	1.5
$k_{1a}(0)$	0.9	0.9	1.4
$k_{21}(0)$	0.03	0	0.02
$k_{12}(0)$	0.5	0.5	0
$k_{a2}(0)$	0	0	0.8
$k_{2a}(0)$	0	0.02	0

TABLE 4: Three of the four combinations of the extreme values of $k_{1a}(0)$ and $k_{2a}(0)$, the combination not shown in this table just recovers the unconnected two pause model of the previous section. (Extreme values compatible with the experimental data were shown in Table 3.)

The experimental constraints for the force dependencies of the rates can be obtained repeating the previous computations with the rates at a force of $10pN$ (*i.e.*, in the other extreme of the interval of measured forces). In this case, the same relations hold yielding

$$1.1s^{-1} < k_{1a}(10pN) < 1.4s^{-1}$$
$$0 < k_{2a}(10pN) < 0.11s^{-1},$$

Table 5 shows the range values obtained for the other rates of the two connected pause state mode.

	Minimum (s^{-1})	Maximum (s^{-1})
$k_{a1}(10pN)$	0.4	0.6
$k_{1a}(10pN)$	1.1	1.4
$k_{21}(10pN)$	0	0.14
$k_{12}(10pN)$	0	0.3
$k_{a2}(10pN)$	0	0.15
$k_{2a}(10pN)$	0	0.11

TABLE 5: Minimum and maximum values of the rates at a force of $10pN$, obtained using the most probable values in Table 1 and Table 2.

The force dependencies of the rates, $k_{ij}(F) = k_{ij}(0) \cdot e^{\frac{-d_{ij}F}{k_BT}}$, are given by d_{ij}, which can be obtained as

$$d_{ij} = \frac{k_BT}{10pN} \ln \frac{k_{ij}(0)}{k_{ij}(10pN)}.$$

If no further assumptions are considered, the values of d_{a1} and d_{1a} are bounded between 0.4 and 0.7 and d_{1a} between -0.17 and 0.11, respectively; while all the other force dependencies d_{ij} of the model are not bounded by the experimental results.

Detailed balance

Detailed balance implies no net cycling, and therefore

$$\frac{k_{a1}}{k_{1a}} \frac{k_{12}}{k_{21}} \frac{k_{2a}}{k_{a2}} = 1$$

Thus, assuming detailed balance the number of degrees of freedom is reduced by two (one rate and its force dependency), allowing us to express the values of the rates compatible with experimental data in terms of the values of one of the rates and its force dependency, we choose k_{1a}. As the detailed balance relation implies

$$k_{2a} = \frac{k_{a2}k_{21}}{k_{a1}k_{12}} k_{1a}.$$

The dependencies of the rates at zero force with the value of $k_{1a}(0)$ are shown in Fig. 5. The highest value of $k_{1a}(0)$ recovers the independent pause model (*i.e.*, $k_{12} = 0$ and $k_{21} = 0$), while its lowest value would promote a model where entrance to and exit from the second pause state (P2) takes place only through the first pause state (P1) (*i.e.*, $k_{a2} = 0$ and $k_{2a} = 0$).

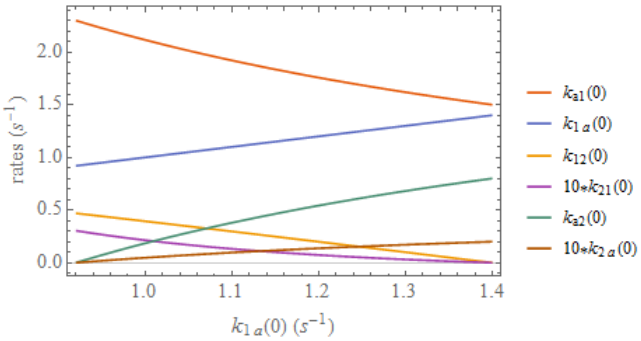

FIG. 5: Rates at zero force as a function of $k_{1a}(0)$, under detailed balance conditions.

The corresponding Gibbs energies of the individual transitions are given by

$$\frac{k_{ij}}{k_{ji}} = e^{\frac{-\Delta G_{ij}}{k_B T}},$$

implying

$$\Delta G_{ij} = -k_B T \cdot \ln\left(\frac{k_{ij}}{k_{ji}}\right) = -k_B T \cdot \ln\left(\frac{k_{ij}(0)}{k_{ji}(0)}\right) + \left(d_{ij} - d_{ji}\right) \cdot F.$$

The Gibbs energies at zero force $\Delta G_{ij}(0)$ are shown in Fig. 6.

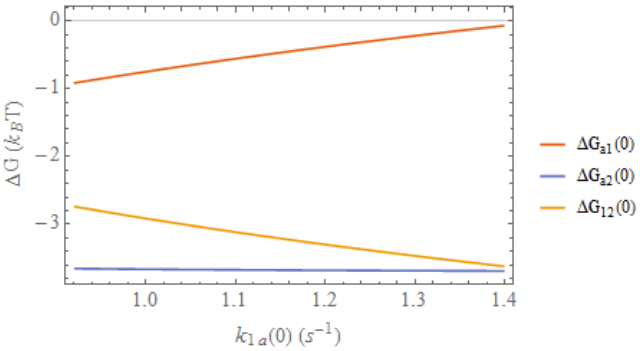

FIG. 6: Gibbs free energies at zero force as a function of $k_{1a}(0)$, under detailed balance conditions.

Analogously, the dependency of the rates at $10pN$ can be shown as a function of $k_{1a}(10pN)$, see Fig. 7.

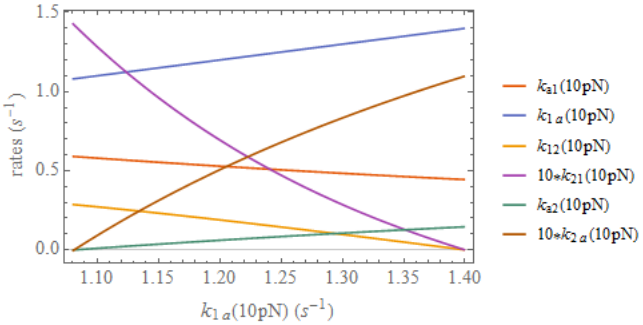

FIG. 7: Rates at a force of $10pN$ as a function of $k_{1a}(10pN)$, if detailed balance is assumed.

In the two connected pause states model the value of d_{1a} can range between -0.17 and 0.11, while in the unconnected model it was approximately zero.

However, when d_{1a} is assumed to be zero, k_{1a} is force independent, and we can obtain as a function of a single parameter $k_{1a}(0) = k_{1a}(10pN)$: the force dependencies of all the other rates (Fig. 8), and the force dependencies of the Gibbs energy changes associated to the transitions (Fig. 9).

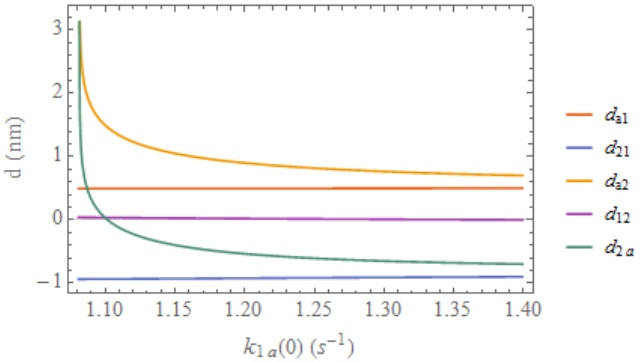

FIG. 8: Force dependencies, d_{ij}, as a function of $k_{1a}(0)$, under detailed balance and $d_{1a} = 0$ conditions.

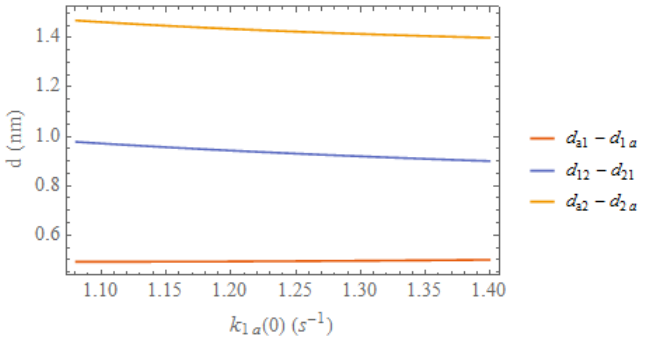

FIG. 9: Force dependency of the Gibbs energy of each transition as a function of $k_{1a}(0)$, under detailed balance and $d_{1a} = 0$ conditions.

IV. CONCLUSIONS

Single molecule experiments provide the unique capability of following the dynamics of individual molecular motors along a precise mechanical coordinate. Time trajectories contain valuable information about the number of intermediate states, while the residence times in these states indicate transition rates between them. This information ultimately dictates the average behavior of a given property, such as the average replication velocity of the Phi29 DNA polymerase.

In a previous work, we analyzed the pause behavior of the wild type and of a strand displacement deficient mutant of the Phi29 DNA polymerase. Pauses have a fundamental role in modulating the average replication rate. The pause length frequency distribution during the DNA unwinding activity of the mutant polymerase presented two time scales. This indicates that, at least, two types of pause events coexist for this case, and required the introduction of a two pause states model. Here, the different interconnection schemes between these two states have been explored.

A simple model considering two independent pause states branching off the active state (Fig. 3) was shown to be consistent with single molecule and biochemical observations (Ref. 5). In this model, all the parameters can be determined from the experimental data. According to this model the short pause state is caused by local arrangements of the template (i.e. base stacking), which can be unraveled by mechanical extension of the template (d_{a1}), while the long pause state is caused by the closed conformation of the DNA fork, which is destabilized (d_{a2}, d_{2a}) by tension favoring the unwinding reactions. Here we show that an alternative model considering a direct connection between these pause states is also compatible with the results obtained from the single molecule experiments. In this case, the values of the parameters are not fully determined by the experimental data. However, they are all constrained by the observations. This model presents a two dimensional set of solutions. Sets of parameters compatible with experiments include: 1) an indirect entry/exit to/from pause state and 2) cycling around the active and inactive (pause) states in both directions.

Cycling between the active and pause states requires an external source of energy, as nucleotide hydrolysis, which as discussed before is highly improbable. Therefore, if no cycling is present, detailed balance is verified, given an additional relation valid for all forces, which allows to decrease by two the number of free parameters. Experimental data indicates that the exit rate from pause $P1$ to the active state, k_{1a}, can only depend mildly on force, in particular $|d_{1a}| \leq 0.17$. If we assume $d_{1a} = 0$ we are left with only one free parameter. This leads to a one parameter set of solutions that interpolates between two models, from the unconnected two pause states model to the indirect entry and exit to pause P2 model. We have also shown that the Gibbs energies of the transitions are of the order of a few k_BT, and while their values at zero force change along the set of solutions, their force dependencies are quite constant.

Alternative models considering direct connections between the pause states imply that the transition between these states involve local conformational changes within the polymerase DNA complex of ~1 nm (d_{12}-d_{21}=1 nm). Further single molecule, structural and biochemical information are required to correctly interpret these putative conformational changes and, to

validate one of the possible kinetic models of DNA replication along the DNA fork.

These results show that models with more parameters give sets of solutions. However, these solutions can be analyzed, as shown here, to get insights into the rates and the energies involved in the transitions in the various possible cases. The methods shown here can also be used to study the dynamics of other molecular processes governing the operation of molecular motors, such us the mechano-chemical steps governing the translocation mechanism of the DNA polymerases (Ref. 8).

ACKNOWLEDGEMENTS

This work was supported by grants FIS2010-17440 and to F.J.C., and BFU2012-31825 to B.I. from Ministerio de Ciencia e Innovación (Spain), and GR35/14-920911to F.J.C. from Universidad Complutense de Madrid and Banco Santander (Spain).

[1] K. Visscher, M. J. Schnitzer, and S. M. Block, *Single kinesin molecules studied with a molecular force clamp*, Nature **400**, 184 (1999).

[2] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante, *Recent Advances in optical tweezers*, Annu. Rev. Biochem. **77**, 205 (2008).

[3] C. Bustamante, W. Cheng, Y. X. Mejia, and Y. X. Meija, *Revisiting the central dogma one molecule at a time*, Cell **144**, 480 (2011).

[4] M. Salas, *Protein-priming of DNA replication*, Annu. Rev. Biochem. **60**, 39–71(1991).

[5] J. A. Morin, F. J. Cao, J. M. Lázaro, J. R. Arias-Gonzalez, J. M. Valpuesta, J. L. Carrascosa, M. Salas, B. Ibarra, *Active DNA unwinding dynamics during processive DNA replication*, Proc. Natl. Acad. Sci. U. S. A. **109**, 8115 (2012).

[6] J. A. Morin, F. J. Cao, J. M. Valpuesta, J. L. Carrascosa, M. Salas, and B. Ibarra, *Manipulation of single polymerase-DNA complexes: A mechanical view of DNA unwinding during replication*, Cell Cycle **11**, 2967 (2012).

[7] M. B. Jackson, *Molecular and Cellular Biophysics*, Cambridge University Press (2006).

[8] J. A. Morin, F. J. Cao, J. M. Lázaro, J. R. Arias-Gonzalez, J. M. Valpuesta, J. L. Carrascosa, M. Salas, B. Ibarra, *Mechano-chemical kinetics of DNA replication: identification of the translocation step of a replicative DNA polymerase*, Nucleic Acids Research **43**, 3643-3652 (2015).