Classification: Biological Sciences (Major); Biophysics and Computational Biology (Minor)

Running title: DNA unwinding mechanism of the Φ29 DNA polymerase

Active DNA unwinding dynamics during processive DNA replication

José A. Morin[a,b,1], Francisco J. Cao[c,1], José M. Lázaro[d], J. Ricardo Arias-Gonzalez[a,e], José M. Valpuesta[e], José L. Carrascosa[a,e], Margarita Salas[d,2] & Borja Ibarra[a,e,2]

[a]Nanobiosystems. Instituto Madrileño de Estudios Avanzados, IMDEA Nanoscience. Madrid 28049, Spain.

[b]Centro de Estudios Avanzados de Cuba (CEAC). Habana, 17100, Cuba.

[c]Departamento Física Atómica, Molecular y Nuclear. Universidad Complutense. Madrid, 28040, Spain.

[d]Centro de Biología Molecular 'Severo Ochoa' (C.S.I.C.-U.A.M.). Madrid, 28049, Spain.

[e]Departamento Estructura de Macromoléculas. Centro Nacional de Biotecnología (C.S.I.C.). Madrid, 28049, Spain.

[1] These authors contributed equally to this work

[2]To whom correspondence may be addressed: msalas@cmb.uam.es or borja.ibarra@imdea.org

Author contributions:

JAM and FJC contributed equally to this work. BI designed the experiments; JAM and BI performed the experiments; JAM, FJC and BI analyzed the data; JRA-G, JML, MS, JMV and JLC provided technical and biochemical tools; FJC, MS and BI wrote the paper

Keywords: DNA replication, DNA unwinding, molecular motors, optical tweezers

Number of pages: 21

Number of figures: 5

Number of tables: 1

Summary

Duplication of double-stranded DNA (dsDNA) requires a fine-tuned coordination between the DNA replication and unwinding reactions. Using optical tweezers we probed the coupling dynamics between these two activities when they are simultaneously carried out by individual Phi29 DNA polymerase molecules replicating a dsDNA hairpin. We used the wild-type and an unwinding deficient polymerase variant and found that mechanical tension applied on the DNA and the DNA sequence modulate in different ways the replication, unwinding rates and pause kinetics of each polymerase. However, incorporation of pause kinetics in a model to quantify the unwinding reaction reveals that both polymerases destabilize the fork with the same active mechanism and offers insights into the topological strategies that could be used by the Phi29 DNA polymerase and other DNA replication systems to couple unwinding and replication reactions.

Introduction

\body

Replication of dsDNA requires overcoming the energetic barrier associated with the base pair melting of its double helix. In many DNA replication systems replication and unwinding of the fork are carried out by the coordinate action of different proteins, the DNA polymerase holoenzyme and the replicative helicase, respectively (1, 2). In other systems, like the bacteriophage Phi29, these two activities are coupled within the replicative DNA polymerase (3, 4). The Phi29 DNA polymerase presents the common folding and catalytic activities characteristics of the Family B DNA polymerases (4) but in addition, it also presents an amino acid insertion in the polymerization domain, named TPR2 (Fig. S1), which confers the protein a processive strand displacement

activity (5, 6). The TPR2 insertion together with the thumb, palm and exonuclease (exo) domains forms a narrow (10 Å diameter), closed tunnel around the template strand, in comparison with the open channel described in other related DNA polymerases (5). Interestingly, this topological restriction resembles the one imposed by hexameric replicative DNA helicases at the fork junction (7, 8) and requires the dsDNA in front of the polymerase to open in order for the template to enter the active site. Therefore, the Phi29 DNA polymerase works as a hybrid *polymerase-helicase* and constitutes a good model system to understand the basic mechanistic principles of the coupling between DNA replication and unwinding reactions.

The physical mechanism by which the polymerase replicates and promotes DNA unwinding could be described as 'passive', the polymerase will not enter the duplex region until thermal fluctuations transiently open the dsDNA junction, or 'active', it destabilizes the duplex DNA near the junction shifting the equilibrium of the fork toward opening (1, 9-11). The rate of a 'passive' motor would be more sensitive than the rate of an 'active' motor to the strength of the fork ahead (11). These two behaviours are the two extremes of a continuous spectra and the polymerase is expected to present an unwinding mechanisms located between an ideally active and totally passive behaviour (11). Here we use optical tweezers and a detailed kinetic analysis to probe the coupling between DNA unwinding and replication reactions by studying the strand displacement and primer extension activities of the wild-type Phi29 DNA polymerase and a strand displacement deficient variant (12) under the same external (destabilizing) tension on the DNA and sequence context. The particular effect of fork stability on the average unwinding rate and pause kinetics of each polymerase suggests a molecular mechanism for the coupling of DNA replication and unwinding reactions on the Phi29

DNA polymerase which could be extrapolated to understand the coupling dynamics between these two reactions in other DNA replications systems.

Results

Single-molecule dsDNA replication-unwinding assay. We used optical tweezers to apply mechanical tension to the ends of the complementary strands of a single 556 bp long DNA hairpin containing a unique 3' end polymerase loading site and monitor at constant force (below 13 pN where the hairpin is stably folded) the end-to-end distance change of the DNA (Δx1 and Δx2) as the Phi29 DNA polymerase replicates the hairpin in the presence of dNTPs (Fig. 1). Initially, during the strand displacement phase, the protein moves through the hairpin using one strand as a template to incorporate the corresponding complementary nucleotide into the 3' end of the primer and at the same time displacing the complementary strand, so the distance between the beads increases (Δx1, Fig. 1). Upon reaching the end of the hairpin, no more unwinding is required and the polymerase continues replication in the primer extension mode, converting the displaced complementary single strand (ss-) to dsDNA. This reaction also changes the distance between the beads (Δx2, Fig 1) due to the different elastic properties of each DNA polymer (13). Δx1 and Δx2 were converted to the number of nucleotides incorporated as a function of time at a given force (Methods and Fig. S2).

To determine the effect of sequence on the polymerase activity the hairpin includes four GC base pair clusters with 1 to 4 repetitions of the GCC sequence, separated by a ~100 nucleotide low-GC content sequence, the GC clusters constituting barriers to DNA unwinding (Fig. S1). Therefore, this experimental setup allowed us to study and compare the particular effect of force and sequence on both the strand displacement and

primer extension reactions of the wild-type and the strand displacement deficient mutant (*sdd*) polymerases.

Different effects of fork stability on the unwinding rates of the wild-type and *sdd* mutant polymerases. The average strand displacement rate of the wild-type polymerase increased moderately with force, from V_{sd}~ 80 nt/s (3.5 pN) to V_{sd}~ 120 nt/s (13 pN), a velocity similar to its primer extension rate along the displaced complementary strand (V_{pe}~ 128 nt/s) (Fig 2A). In contrast, the average strand displacement rate of the *sdd* mutant polymerase presented a much stronger force dependence, increasing by a factor of 9, from V_{sd}~ 10 nt/s (4.2 pN) to ~ 90 nt/s (12 pN), also approaching its primer extension rate, V_{pe}~ 128 nt/s (Fig. 2C). Both polymerases present an identical primer extension rate which is independent of tension below 14 pN (Fig. 2A and Fig. 2C), indicating that template tension within this range do not affect the rate-limiting step of the replication cycle, as reported previously for these polymerases (14). Therefore, these data indicate that, during strand displacement, unwinding of the dsDNA fork is the force dependent step of the reaction, which limits moderately the wild-type and strongly the mutant polymerase advance through the hairpin (V_{sd}/V_{pe}~ 0.67 vs. 0.08, respectively).

Next, we measured the effect of sequence on the rates of both polymerases. For the wild-type polymerase during strand displacement, a direct comparison between the average residence time of the protein at each position in the hairpin (Methods) and the unfolding pattern of the hairpin already shows that this protein spends longer times at the four hairpin positions with the greater GC content (Fig. 2B and Fig. S3B). Quantification of this effect revealed that the unwinding rate of the more stable GC pairs is ~3 times slower than the unwinding of the AT pairs (V_{sdGC}/V_{sdAT}~ 0.36), which

in turn is close to the primer extension rate (V_{sdAT}/V_{pe}~ 0.86) (Methods and Fig. S3D). Interestingly, the *sdd* mutant polymerase during strand displacement (but not during primer extension) presents long pauses precisely located at the GC rich positions of the hairpin (Fig. 2D and Fig. S4B). Since we found that template sequence has no significant effect on the primer extension rate of the wild type and *sdd* mutant polymerases (Fig. S3C and Fig. S3D), these data indicate that unwinding of the more stable fork positions has different effects on each polymerase: 1) Slows down the wild-type polymerase translocation and 2) promotes the entrance of the *sdd* mutant polymerase to a long-lived inactive pause state. The different effect of fork stability on the unwinding reaction of each polymerase suggest differences in their unwinding mechanisms (11).

The wild-type and mutant polymerases present different pause kinetics during DNA unwinding. The data show that inactive pause states may be relevant during the unwinding reaction (at least for the *sdd* mutant). Therefore, to further investigate the unwinding mechanism of each polymerase we analyzed in detail their characteristic pause behaviour and its relation to their primer extension and strand displacement activities. We first checked that for both enzymes pauses do not correspond to the replacement of polymerase molecules at the 3' end of the primer (Fig. S4).

For the wild-type polymerase only occasional and short-lived pause events were detected during strand displacement and primer extension conditions (Methods and Fig. 3A). During both replication conditions pauses present very similar characteristics: The pause length frequency distribution (supporting information (SI) Methods) is consistent with a single step Poisson statistics (Fig. 3B). The pause frequency (pause/second) decreases exponentially with tension and increases linearly with the GC density of the DNA (Figures 3C and 3D), while the average pause length (~ 0.6 s) is tension and

sequence independent (Fig. S5). These similarities indicate that the presence of the fork does not alter much the pause kinetics characteristic of DNA replication and does not promote entering an alternative pause state, strongly suggesting that these short pause states are related to the biochemical replication cycle rather than to DNA unwinding reaction. During primer extension conditions the *sdd* mutant also presents short pauses, with very similar rates, tension and sequence dependencies than the wild-type (Fig. 4 and Table 1), suggesting an identical nature for the short inactive states during the replication cycle of both polymerases. This is consistent with ensemble experimental observations indicating that this mutant presents the primer extension properties of the wild-type polymerase (15).

However, during strand displacement the mutant's pause behaviour presents different properties (Fig. 2D). In this case, the pause length frequency distribution is better fit to a double exponential indicating that at least two types of pause events, short- (typically < 4 seconds) and long- lived (typically > 4 seconds) coexist (Fig. 4A). Analysis of the short-lived pause events revealed that they present almost identical characteristics to the pauses observed during primer extension and to the ones described for the wild-type polymerase (Fig. 4 and Table 1), suggesting that they correspond to the short-lived pause state characteristic of template replication and they are not related with unwinding. The long-lived pauses instead present quite different properties from the shorter pauses: They only occur during strand displacement and not during primer extension (Fig. 4A). Their duration is ~ 50 times longer than the short pauses (at low tensions) and more importantly presents a strong tension dependence, decreasing exponentially as the destabilization tension at the fork increases (Fig. 4B). In addition, their frequency decreases exponentially with tension (Fig. 4C) and increases linearly with the strength of the GC content of the fork (Fig. 4D). Taken together, our data

suggest that long pauses correspond to an inactive polymerase-DNA complex state induced during the unwinding of the more stable dsDNA fork positions (GC clusters) and that the external mechanical destabilization of the fork promotes the exit from this long pause state. These observations explain the inability of the *sdd* mutant to replicate *in vitro* the Phi29 dsDNA genome beyond the initial GC clusters (15).

The pause behaviour of the wild-type polymerase (both during strand displacement and primer extension) and the mutant polymerase during primer extension conditions, can be explained with a simple kinetic model in which the short-lived pauses correspond to a single off-pathway inactive state branching off the main active state of the protein during replication of the GC rich positions of the template (Short Pause 1, Fig. 3E). Instead, the simplest kinetic model that explains the mutant's pause behaviour during unwinding requires the short- and long- lived pause states to branch off independently from the main active state of the polymerase (Short Pause 1 and Long Pause 2, respectively, Fig. 4E. Alternative models were also considered, see SI Text). In these models, the tension dependent entry and exit rates to and from the pause states, $k(f)$, change with tension following (16): $k(f) = k(0) * \exp(-f * d / k_B T)$ (equation 1), where f is the applied tension, $k(0)$ the rate at zero tension (Table 1), k_B is the Boltzmann constant, T the absolute temperature and d the distance to the transition state along the mechanical tension coordinate (Table 1). We note that since the tension dependent entry rates to the short- ($k_{a1}(f)$) and long- ($k_{a2}(f)$) lived pause states are also sequence dependent (Figures 3D and 4D), we calculated the particular entrance rates to the both pause states at the GC positions, $k_{a1GC}(f)$ and $k_{a2GC}(f)$ (Table 1 and SI Text). The pause exit rate from the short-lived pauses, k_{1a}, can be directly calculated as the inverse of the average tension and sequence independent pause length (Table 1).

Inclusion of pause kinetics in a model to quantify the unwinding mechanism revealed the same active mechanism for the wild-type and mutant polymerases. Our data revealed that pauses strongly modulate the average unwinding rate of the mutant polymerase. Similarly, pauses have been shown also to be part of the unwinding mechanism of several specialized nucleic acid (NA) motors (17-20). Therefore, the kinetics of pause events must be considered for the correct interpretation of tension and sequence dependencies of the unwinding rates. We have extended the physical framework (11, 21) where the unwinding 'activeness' of the NA unwinding motor depends on the interaction energy, G_{int}, between the protein and the NA fork (Fig. 5A), to include the effect of pauses on the unwinding reaction of the Phi29 DNA polymerase. In terms of the residence time of the polymerase in each template position the generalization of the model is direct, the residence time with pauses $T(f)$ is given by $T(f) = T_a(f) + \sum_i T_i(f)$, with $T_a(f)$ the residence time without pauses (see reference 21 and SI Text), and $T_i(f)$ the residence time contribution due to pauses of type i determined by the pause kinetics. According to our kinetic models (Fig. 3E and Fig. 4E), the residence time contribution due to pause kinetics is $\sum_i T_i(f) = T_1(f) + T_2(f)$, where the additional residence times, $T_1(f)$ and $T_2(f)$, are determined by the entry and exit rates ($k_{a1}(f)$, $k_{a2}(f)$ and k_{1a}, $k_{2a}(f)$) to the short and long- lived pause states as $T_1(f) = T_a(f) \cdot \frac{k_{a1}(f)}{k_{1a}(f)}$ and $T_2(f) = T_a(f) \cdot \frac{k_{a2}(f)}{k_{2a}(f)}$ (see SI Text for sequence dependencies).

We used our model to quantify the DNA unwinding activity of both polymerases when pauses are included and excluded from the average strand displacement rate. We considered that the polymerase can only move forward when there is ssDNA template available in front of it ($m \geq 1$, Fig. 5A). Importantly, in the case of the DNA

polymerase, since replication is known to occur in discrete steps of 1 nucleotide at a time, δ= 1 nt (22) and backtracking is not expected (SI Text) and not observed with our current resolution, only two variables determine the fits; the interaction energy between polymerase and the fork, ΔG_{int}, and the range of this interaction, M (Fig. 5A and SI Text). We found that the best fit of the model to the wild-type polymerase tension dependent strand displacement rate (with and without pauses) was obtained with an interaction potential increase per base of ΔG_{int}= 2 k_BT and M= 2 (Fig. 2A and Fig. S6). In addition, the same set of parameters also correctly predicted the sequence dependent residence time (and velocity) of the polymerase during unwinding of the hairpin supporting the validity of the model (Fig. 2B and Fig. S3D). Therefore, the tension and sequence dependent unwinding rate of the wild-type Phi29 DNA polymerase (with and without pauses) support an active DNA unwinding model where the polymerase destabilizes at least 2 bases at the junction facilitating in this way the fork opening.

The strong tension and sequence dependency of the strand displacement rate of the *sdd* mutant may suggest this protein uses a different or less active unwinding mechanism than the wild-type. However, when pause kinetics are consistently considered, the values ΔG_{int} = 2 k_BT and M= 2 nicely predict the average strand displacement rate of the mutant (Fig. 2C and Fig. S6). Therefore, our data strongly support that both enzymes use the same active DNA unwinding mechanism, but the mutant's mechanism fails during unwinding of the more stable fork positions, which promotes entering into the long-lived inactive pause state. The identical values and tension dependencies of the average unwinding rates without pauses found for both polymerases further support this hypothesis (Fig. 2A and Fig. 2C).

Discussion

Our data show that the rate limiting step of the polymerization cycle of the Phi29 DNA polymerase is not affected by tension (below 14 pN) or sequence. However, replication along GC rich template sequences promotes the entrance to a short-lived pause state (Short Pause 1), which is more frequent (~2 to 4-fold) and presents a stronger tension dependence (~2-fold) during primer extension than during strand displacement conditions (k_{a1GC} and d_{a1GC}, Table 1). These observations suggest that translocation along the DNA template could be interrupted by the particular arrangements of the GC rich positions of the template (i.e. GG base stacking), which would occur with less probability and involve fewer bases during strand displacement due to the limited number of bases opened in front of the polymerase. Mechanical extension of the template could prevent base stacking avoiding in this way the entrance to the short-lived pause state. In fact, other factors known to affect the stability of DNA, as betaine and increasing temperatures, also decrease the pause frequency at GC rich sequences during primer extension conditions for Phi29 and other DNA polymerases, supporting this hypothesis (19, 23, 24).

In contrast, the unwinding of the fork limits the forward translocation of both, the wild-type and the *sdd* mutant polymerases, although in different ways: The more stable fork positions slow down the wild-type but pause the advance of the mutant polymerase. Despite these differences inclusion of pause kinetics in a model to quantify the polymerase unwinding mechanism revealed that both polymerases actively destabilize the 2 nearest bp of the fork (M= 2) with a destabilization energy of ΔG_{int}~ 2 k_BT per bp (Fig. 2). The extension of the Betterton and Jülicher (11) model we present in this work points out the importance to consider pause kinetics and can be applied to the quantification of the unwinding mechanism of other NA unwinding motors in which pauses are known to be relevant or are expected to occur during their operation (17-20).

To rationalize the above observations we propose a model for the mechanism of coupling of DNA unwinding and replication. The active destabilization of the fork promoted by the Phi29 DNA polymerase is likely due to the particular polymerase-DNA interactions at the fork junction since, in contrast with other polymerases, the fingers domain seems not to be involved in the DNA unwinding reaction (25-27) and in addition, there are no evidences for the interaction of the displaced strand with the polymerase during unwinding or for different translocation modes between the primer extension and strand displacement activities (6, 22). Structural studies of replicative polymerases during primer extension, including the Phi29 DNA polymerase, showed that during replication the template strand is sharply bent (~90^{o}) inside the polymerase (Fig. 5B) (22). In the case of Phi29 DNA polymerase, this sharp template bending would be maintained during strand displacement activity by the stable thread of the template strand (2 nucleotides) through the narrow tunnel formed by the TPR2 and the exo domains (22). At the same time, the small dimensions of the tunnel (10 Å) would promote the steric exclusion of the complementary strand, probably forcing its bending near the tunnel entrance. Due to the elastic properties of ssDNA (13) bending of the two DNA strands may incur in generation of mechanical stress at the fork junction promoting its destabilization during strand displacement conditions (Fig. 5B).

Our data show that the active unwinding mechanism of the *sdd* mutant fails (*Long* Pause 2) at the more stable GC base pairs of the hairpin, which will be closed with higher probability (28). In fact, we found that the tension dependence of the entry rate to the *Long* Pause 2 state at the GC rich positions, d_{a2GC}~ 0. 5 nm (Table 1), can be explained when considering that entrance occurs when the next base-pair is closed (m= 0) (SI Text). The value of d_{p2GC} is consistent with the interpretation that tension would prevent entry to long pauses by promoting the opening of the next base-pair of the fork,

since the average distance between ssDNA nucleotides under our experimental conditions is 0.25 nm/nt (Fig. S2C), 0.5 nm/ 0.25 nm/nt = 2 nt=1 base pair. Mutations D12AD66A at the exo domain are thought to disturb the specific arrangement of the TPR2 and the exo domains (29), which in turn would affect the stability of the template tunnel. According to our model, this defect would impede the mutant to hold the mechanical stress at a closed fork junction, which in turn would induce loosening of the correct interactions with the template and would favour the entrance to the long-lived inactive pause state (Fig. 5B). The fork pressure will prevent the functional binding of the template back to the polymerization domain, stopping the polymerase advance. According to our hypothesis the mutant polymerase would exit from the *Long* Pause 2 inactive state only when the fork is opened ($m > 0$) and the polymerase has the chance to recover the control on the template strand. Indeed, we checked that the tension dependence of the exit rate from the *Long* Pause 2 state, $d_{2a} \sim -0.7$ nm, can be explained when considering $m > 0$, in other words, when the fork ahead of the polymerase is opened (SI Text). The value of d_{2a} indicates that the unwinding of the first ~2 bp of the fork (0.7 nm/ 0.25 nm/nt) is required to rescue the polymerase from the *Long* Pause 2 inactive state (Fig. 5B, the negative sign of d_{2a} simply reflects that k_{2a} increases with growing tension, see Eq. 1).

The Phi29 DNA polymerase presents a more efficient strand displacement activity than other replicative DNA polymerases like the bacteriophage T4, T7 or the prokaryotic Pol III (30, 31). Due to the open template-tunnel characteristic of these polymerases (32, 33), it is possible that they cannot keep the sharp bending of the template against a closed fork junction impeding further polymerization (as observed in our experiments with the *sdd* mutant). Therefore, to replicate through dsDNA these polymerases should work in coordination with their respective helicases, which by steric exclusion of the

leading strand (34, 35) would increase the probability of the fork to be opened. The steric exclusion of the leading strand by the helicase plus its sharp bending, induced by the polymerase, may account for the efficient fork destabilization during DNA replication. In the case of the Phi29 DNA polymerase the TPR2-exo tunnel, which closely wraps around the template, allows the protein to maintain both, the bending of the template and the steric exclusion of the complementary strand against a closed fork promoting its active destabilization. In fact, the fork junction destabilization energy of the Phi29 DNA polymerase ($\Delta G_{int} \sim 2\ k_BT$) is apparently higher than the one reported for hexameric DNA helicases, $\Delta G_{int} \sim$ 0.05-1.6 k_BT (20, 21, 36, 37) and polymerases, $\Delta G_{int} \sim$ 1.4 k_BT (19), studied in isolation. These features plus the energy provided by the dNTPs binding and hydrolysis during replication would allow the Phi29 DNA polymerase to work as a hybrid polymerase-helicase and efficiently couple DNA replication and unwinding activities within the same polypeptide.

Methods

Optical tweezers experiments. A custom made 515 bp DNA fragment containing four GCC bp clusters of increasing lengths (3, 6, 9 and 12) separated by 97 bp of a low GC content sequence was purchased from Genscript Corp. The final substrate presents a unique 3' end loading site for the DNA polymerase and a 556 bp long hairpin (Fig. S1). We used a dual-beam optical tweezers instrument (38) to manipulate individual DNA hairpins. Generation and purification of the wild-type and *sdd* polymerases was described previously (12, 39). Phi29 DNA polymerase (2, 20, 200 nM) was flowed into the chamber diluted in the reaction buffer (50 mM Tris-HCl (pH 7.5), 20 mM ammonium sulphate, 10 mM $MgCl_2$, 7 mM 2-Mercaptoethanol, 4% Glycerol (w/v), 0.025% Tween20 (w/v), 0.1 μg/ml BSA) containing 50 μm dNTPs. Data presented in

the main text were taken at a polymerase concentration of 20 nM. Data were collected at 60 Hz at 23 +/- 2 ^{o}C in the 'constant force feedback' mode, in which the distance between the beads was adjusted to maintain a constant tension in the DNA.

Data Analysis. The number of nucleotides incorporated as a function of time during strand displacement was obtained by dividing the observed distance change ($\Delta x1$) by the change in extension at a given tension accompanying the generation of one new base pair and one single-stranded nucleotide (Fig. S2). The number of replicated nucleotides during primer extension was calculated as described earlier (14). The average rates (with and without pauses) were determined by a line fit to the traces showing the number of replicated nucleotides versus time. The final rate at each tension was obtained by averaging over all of the traces at similar tension values (± 1.6 pN).

The residence time (defined as the inverse of the instantaneous replication rate) was obtained from a linear fit of the number of replicated nucleotides (in each replication condition) over a sliding time window of 0.7 s (50 data points) for all velocities at different tensions. Individual residence time traces within the same tension range were aligned to calculate the average residence time at each DNA position at different tensions (Fig. S3). A moving window of 10 bp was used to obtain the velocity as a function of the GC percentage (Fig. S3).

Pause events were identified following the method described previously (14) with a resolution of 0.4-0.8 s (Figures 2D and 3A). The accurate alignment of independent extension vs time records allowed us to compute the average pause frequency and duration at each sequence position within a window of ~20 bp.

This work was supported by grants BFU2008-02328/BMC and CSD2007-00010 (to JLC), BFU2010-15703/BMC (to JMV), BFU2008-00215/BMC and CSD2007-00015 (to MS), S2009MAT-1507 (to JLC, JMV and MS), FIS2010-17440 and GR35/10-A-920GR35/10-A-911 (to FJC).

References

1. von Hippel PH & Delagoutte E (2001) A General Model for Nucleic Acid Helicases and Their "Coupling" within Macromolecular Machines. *Cell* 104:177-190.
2. Stano NM, *et al.* (2005) DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. *Nature* 435:370-373.
3. Blanco L, *et al.* (1989) Highly efficient DNA synthesis by the phage phi29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* 264:8935-8940.
4. Salas M (1991) Protein-Priming of DNA Replication. *Annu Rev Biochem* 60:39-71.
5.Rodriguez I, *et al.* (2005) A specific subdomain in Phi29 DNA polymerase confers both processivity and strand-displacement capacity. *Proc Natl Acad Sci USA* 102:6407-6412.
6. Kamtekar S, *et al.* (2004) Insights into Strand Displacement and Processivity from the Crystal Structure of the Protein-Primed DNA Polymerase of Bacteriophage Phi29. *Mol Cell* 16:609-618.
7. Patel SS & Picha KM (2000) Structure and Function of Hexameric Helicases. *Annu Rev Biochem* 69:651-697.
8. Enemark EJ & Joshua-Tor L (2006) Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* 442:270-275.
9. Lohman TM & Bjornson KP (1996) Mechanisms of Helicase-Catalyzed DNA Unwinding. *Annu Rev Biochem* 65:169-214.
10. Soultanas P & Wigley DB (2001) Unwinding the 'Gordian knot' of helicase action. *TIBS* 26:47-54.
11. Betterton MD & Julicher F (2005) Opening of nucleic-acid double strands by helicases: Active versus passive opening. *Phys Rev E* 71:011904.
12. Soengas MS EJ, Lázaro JM, Bernad A, Blasco M A, Salas M , and L Blanco (1992) Site-directed mutagenesis at the Exo III motif of Phi29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J.* 11:4227-4237.
13. Smith SB, Cui Y, & Bustamante C (1996) Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* 271:795-799.
14. Ibarra B, *et al.* (2009) Proofreading dynamics of a processive DNA polymerase. *EMBO J* 28:2794-2802.
15. Esteban JA, Soengas MS, Salas M, & Blanco L (1994) 3'-->5' exonuclease active site of phi 29 DNA polymerase. Evidence favouring a metal ion-assisted reaction mechanism. *J Biol Chem* 269:31946-31954.
16. Bustamante C, Chemla YR, Forde NR, & Izhaky D (2004) Mechanical Processes in Biochemistry. *Annu Rev Biochem* 73:705-748.

17. Cheng W, Dumont S, Tinoco I, & Bustamante C (2007) NS3 helicase actively separates RNA strands and senses sequence barriers ahead of the opening fork. *Proc Natl Acad Sci USA* 104:13954-13959.
18. Sun B, *et al.* (2008) Impediment of E. coli UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. *EMBO J* 27:3279-3287.
19. Kim S, Schroeder CM, & Xie XS (2009) Single-Molecule Study of DNA Polymerization Activity of HIV-1 Reverse Transcriptase on DNA Templates. *J Mol Biol* 395:995-1006.
20. Ribeck N, Kaplan DL, Bruck I, & Saleh OA (2010) DnaB Helicase Activity Is Modulated by DNA Geometry and Force. *Biophysical J* 99:2170-2179.
21. Johnson DS, Bai L, Smith BY, Patel SS, & Wang MD (2007) Single-Molecule Studies Reveal Dynamics of DNA Unwinding by the Ring-Shaped T7 Helicase. *Cell* 129:1299-1309.
22. Berman AJ, *et al.* (2007) Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. *EMBO J* 26:3494-3505.
23. Mytelka DS & Chamberlin MJ (1996) Analysis and Suppression of DNA Polymerase Pauses Associated with a Trinucleotide Consensus. *Nuc Acids Res* 24:2774-2781.
24. Schwartz JJ & Quake SR (2009) Single molecule measurement of the 'speed limit' of a DNA polymerase. *Proc Natl Acad Sci USA* 106:20294-20299.
25. Fisher TS, Darden T, & Prasad VR (2003) Substitutions at Phe61 in the β3-β4 Hairpin of HIV-1 Reverse Transcriptase Reveal a Role for the Fingers Subdomain in Strand Displacement DNA Synthesis. *J Mol Biol* 325:443-459.
26. Yin YW & Steitz TA (2004) The Structural Mechanism of Translocation and Helicase Activity in T7 RNA Polymerase. *Cell* 116:393-404.
27. Singh K, Srivastava A, Patel SS, & Modak MJ (2007) Participation of the Fingers Subdomain of Escherichia coli DNA Polymerase I in the Strand Displacement Synthesis of DNA. *J Biol Chem* 282:10594-10604.
28. Leroy JL, Kochoyan M, Huynh-Dinh T, & Guéron M (1988) Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton. *J Mol Biol* 200:223-238.
29. Pérez-Arnaiz P, Lázaro JM, Salas M, & de Vega M (2009) Functional Importance of Bacteriophage Phi29 DNA Polymerase Residue Tyr148 in Primer-terminus Stabilisation at the 3'-5' Exonuclease Active Site. *J Mol Biol* 391:797-807.
30. Canceill D, Viguera E, & Ehrlich SD (1999) Replication Slippage of Different DNA Polymerases Is Inversely Related to Their Strand Displacement Efficiency. *J Biol Chem* 274:27481-27490.
31. Yuan Q & McHenry CS (2009) Strand Displacement by DNA Polymerase III Occurs through a Link to Single-stranded DNA-binding Protein Coating the Lagging Strand Template. *J Biol Chem* 284:31672-31679.
32. Doublie S, Tabor S, Long AM, Richardson CC, & Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2Å resolution. *Nature* 391:251-258.
33. Franklin MC, Wang J, & Steitz TA (2001) Structure of the Replicating Complex of a Polα Family DNA Polymerase. *Cell* 105:657-667.
34. Hacker KJ & Johnson KA (1997) A Hexameric Helicase Encircles One DNA Strand and Excludes the Other during DNA Unwinding. *Biochemistry* 36:14080-14087.

35. Ahnert P & Patel SS (1997) Asymmetric Interactions of Hexameric Bacteriophage T7 DNA Helicase with the 5'- and 3'-Tails of the Forked DNA Substrate. *J Biol Chem* 272:32267-32273.
36. Donmez I, Rajagopal V, Jeong Y-J, & Patel SS (2007) Nucleic Acid Unwinding by Hepatitis C Virus and Bacteriophage T7 Helicases Is Sensitive to Base Pair Stability. *J Biol Chem* 282:21116-21123.
37. Lionnet T, Spiering MM, Benkovic SJ, Bensimon D, & Croquette V (2007) Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. *Proc Natl Acad Sci USA* 104:19790-19795.
38. Smith SB, Cui Y, Bustamante C (2003) Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol* 361:134-162.
39. Lazaro JM BL, Salas M (1995) Purification of bacteriophage Phi29 DNA polymerase. *Methods Enzymol* 262:42-49.

Figure 1. Overview of experimental assay. (A) Schematic representation of the experimental design. 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 (Methods). At a constant tension the strand displacement and primer extension activities of the polymerase (red triangle) are detected as a change in distance between the beads, $\Delta x1$ and $\Delta x2$, respectively. (B) Representative replication traces of the wild-type (blue) and *sdd* mutant (green) polymerases showing in detail the distance changes during strand displacement (s.d., $\Delta x1$) and primer extension (p.e., $\Delta x2$) activities ($f \sim$ 11 pN). See Fig. S3 for additional experimental traces at different tensions.

Figure 2. Tension and sequence dependencies of the wild-type and *sdd* mutant rates. Only activities that replicate the full hairpin length were considered (Figure S4A). (A) The tension dependency of the average strand displacement rate (V_{sd}) of the wild-type polymerase with and without pauses (full and empty blue circles, respectively) is well explained by our active model (solid blue line). (B) During strand displacement the

wild-type polymerase spends longer times (blue, *f*~ 4 pN, N= 15) at the more stable hairpin positions (dotted lines). This behaviour is well predicted by the proposed model (orange). For comparison the experimental force-unzipping curve of the hairpin is shown in grey. (C) The different tension dependencies of the average strand displacement rate (V_{sd}) of the mutant with and without pauses (full and empty green circles, respectively) can be explained with the same active model when pause kinetics are included in (solid green line) or excluded from (dashed green line) the model. (D) Representative position vs. time traces of the mutant polymerase during strand displacement. Identified pause events are shown in red, long pauses are located at the GC rich positions of the hairpin (grey lines). Insert shows the *sdd* mutant velocity distribution during strand displacement conditions (see Fig. S5C). (A) and (C) For both polymerases full and empty red circles represent the average primer extension rates (V_{pe}) with and without pauses. Solid red lines represent the sequence independent velocity value used in the model (128 nt/s). Error bars represent the standard error.

Figure 3. Pause kinetics of the wild-type polymerase. (A) Representative position vs. time traces of the wild-type polymerase during strand displacement. Identified pause events are shown in red, grey lines show GC rich positions of the hairpin. Insert shows the wild-type velocity distribution during strand displacement conditions (see Fig. S5A). (B), (C) and (D) red dots represent primer extension and blue dots strand displacement conditions. (B) The pause length frequency distributions (s^{-2}) can be fitted with a single exponential (*f*~ 5pN). Primer extension, red line (R^2= 0.89, N= 22), strand displacement, blue line (R^2= 0.87, N= 38). (C) The pause frequency (s^{-1}) decreases exponentially with tension (solid lines represent fit to data with equation 1) and (D) increases linearly with the GC content of the DNA (*f*~ 5 pN). Primer extension, red line

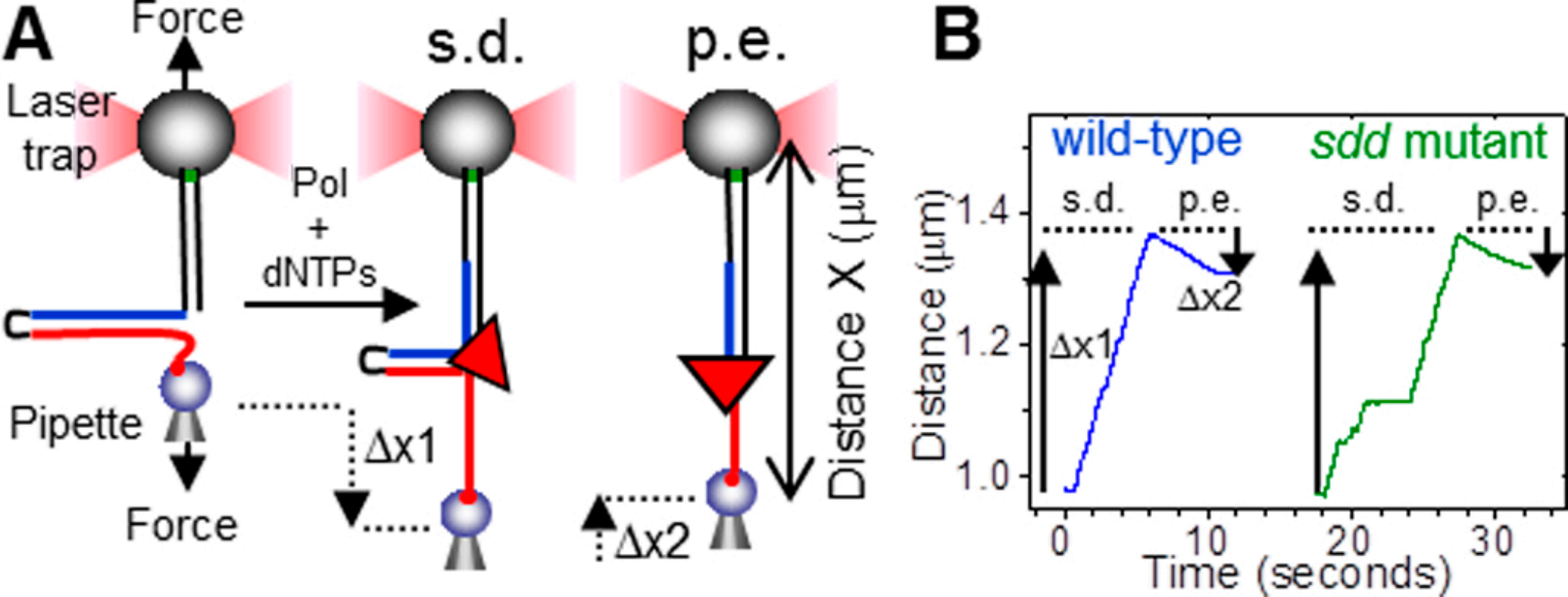
(R^2= 0.81, N= 38), strand displacement, blue line (R^2= 0.78, N= 22). (E) Proposed kinetic model that explains the behaviour of short-lived pauses (Short Pause 1). See Table 1 for rate values and the associated tension dependence, d_{a1}. For (B) and (D) error bars calculation is described in SI Text. For (C) error bars represent standard deviation.

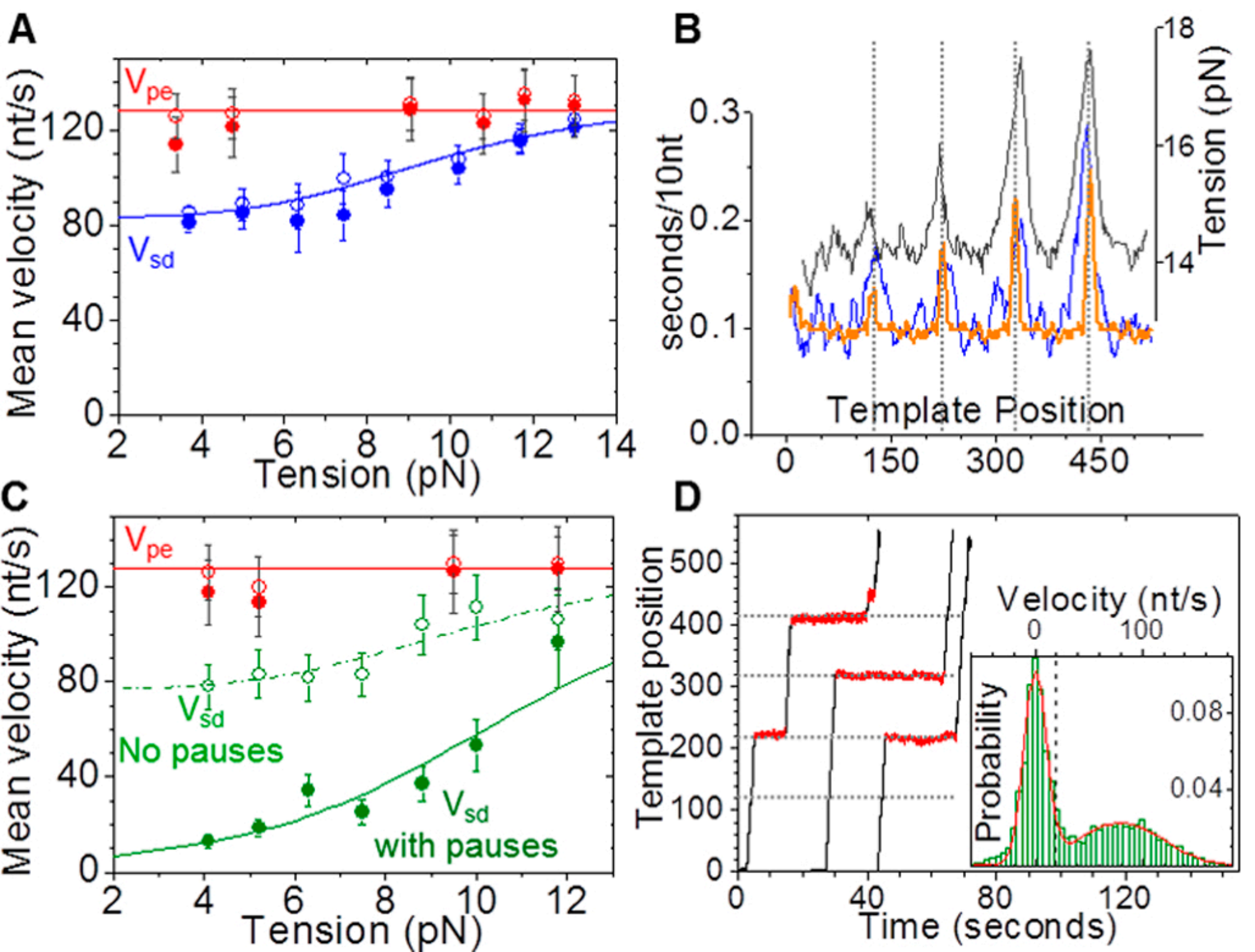
Figure 4. Pause kinetics of the *sdd* mutant polymerase. (A) The pause length frequency distribution (s^{-2}) during primer extension (*f*~ 6.5 pN, red dots) is compatible with a single exponential (solid red line R^2= 0.97, N= 25). For comparison the wild-type distribution from Fig. 3B is shown as a grey line. During strand displacement (*f*~ 6.5 pN, green dots) a double exponential is required to fit the data (green solid line, R^2= 0.93, N= 41). For (B), (C) and (D), green dots represent short pauses during strand displacement; red dots, short pauses primer extension; black dots, long pauses strand displacement. (B) The average length (seconds) of short pauses is tension independent while the average length of long pauses decreases exponentially with tension (black line fit to the data with equation 1). (C) For both pause types the pause frequency (s^{-1}) decreases exponentially with tension (solid lines represent the fit to the data with equation 1) and (D) increases linearly with the GC content of the DNA (solid lines represent linear fits, R^2=0.98, 0.99, 0.92). (E) Proposed kinetic model that explains the mutant pause behaviour during strand displacement conditions. See Table 1 for rates and conformational changes values. For (A) and (D) error bars calculation is described in SI Text. For (B) and (C) error bars represent standard deviation.

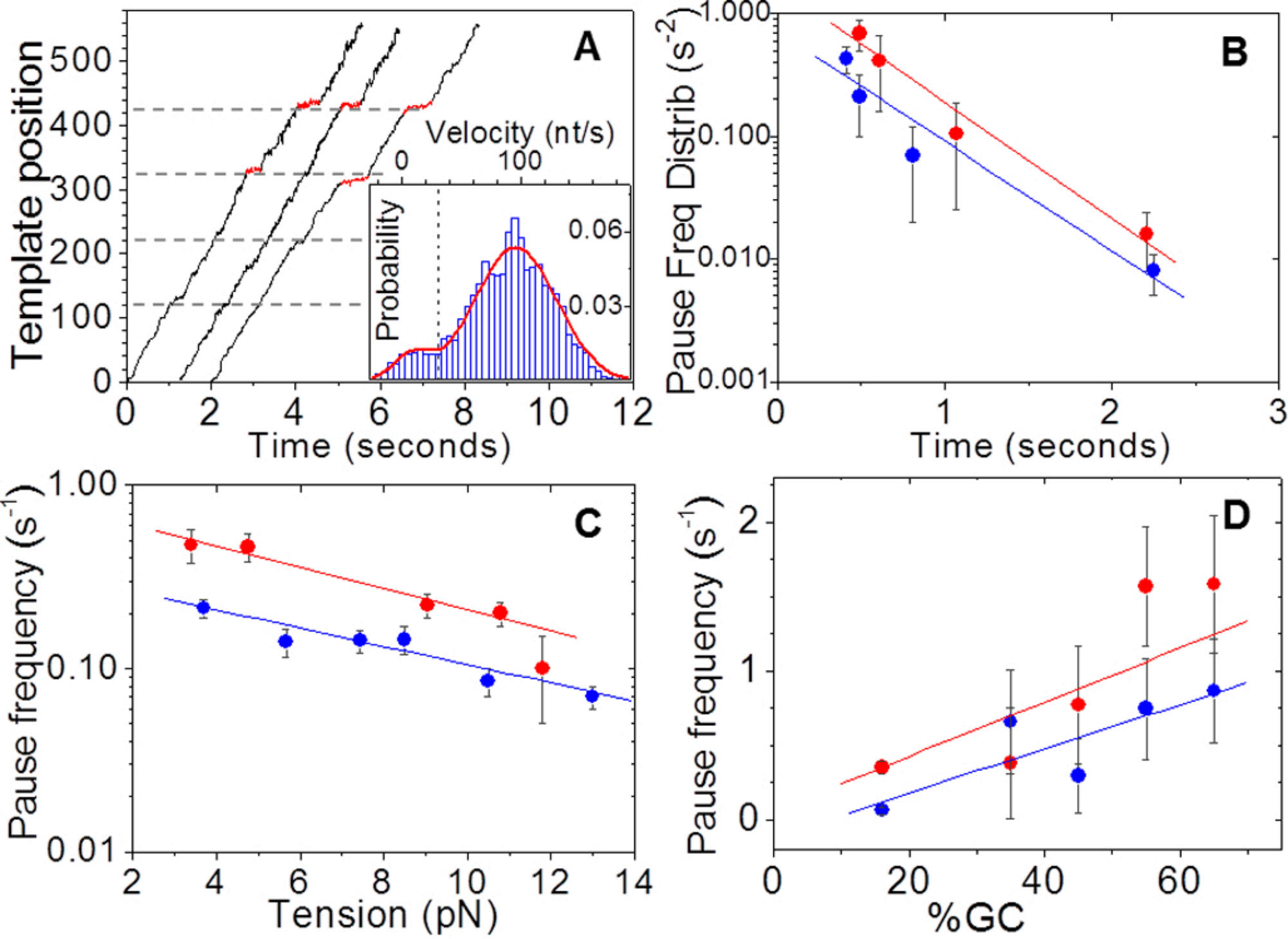
Figure 5. DNA unwinding model. (A) Diagram showing notation used for modelling the polymerase movement during strand displacement conditions. *M* (red circle) defines the range of interaction between the polymerase and the junction; *m* is the number of

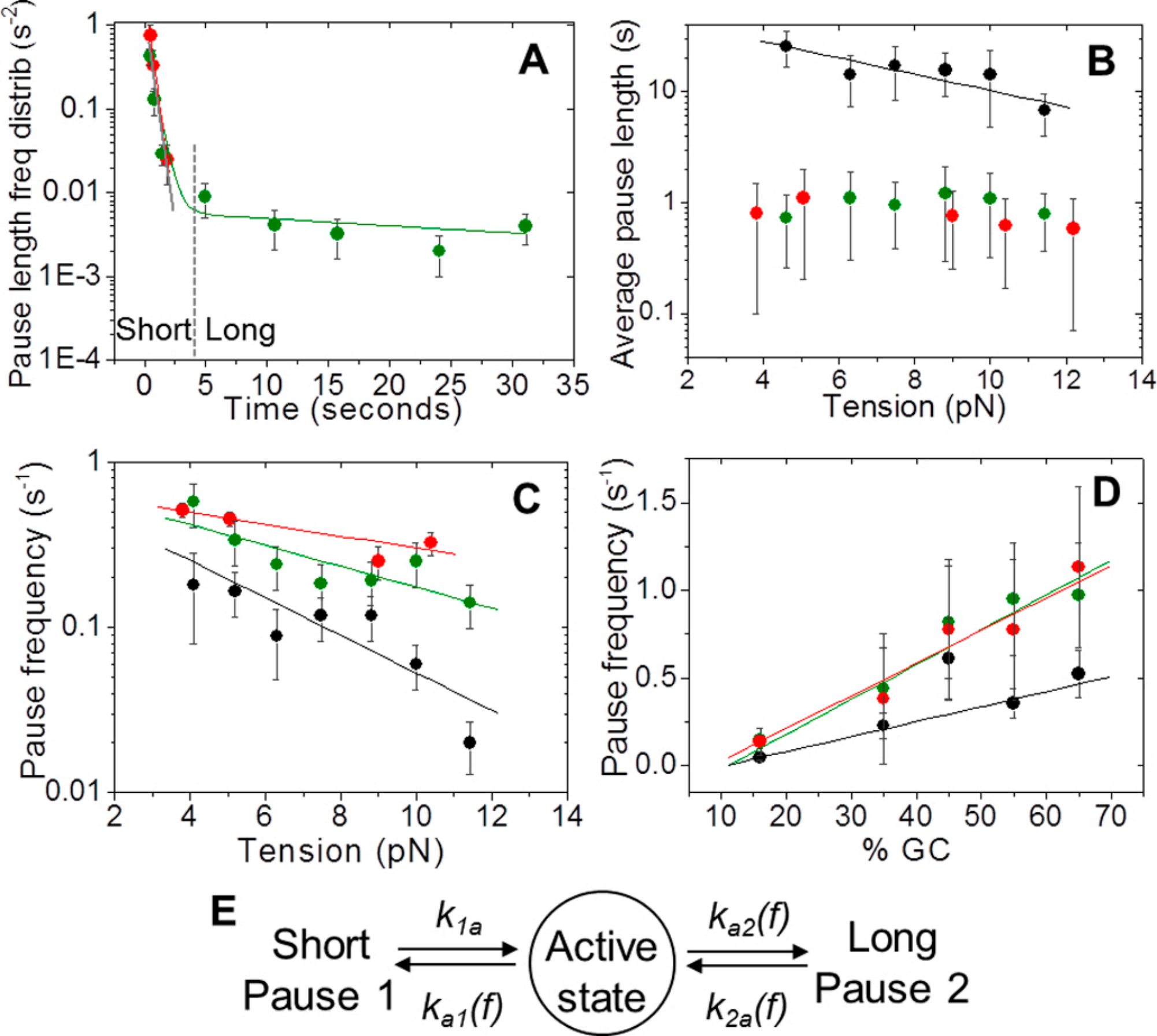
ssDNA template nucleotides between the polymerase and the junction; *l* is the total number of replicated nucleotides and *L* is the full length of the hairpin. (B) Proposed DNA unwinding mechanism. Figures show the schematic representation of the Phi29 DNA polymerase with the DNA substrate during strand displacement (6). The protein is diagrammed in two levels. The upper level contains the exo (green), the TPR2 (blue) and thumb (orange) domains. The rest of the protein is shown as a grey circle. 1) and 2) Template bending at the TPR2-exo tunnel and steric exclusion of the complementary strand may generate mechanical stress at the fork junction promoting the unwinding of the first 2 bp (blue) of the fork. dNTP binding and hydrolysis fuelled the polymerase forward movement. 3) For the *sdd* mutant, mechanical stress at the junction during unwinding of the more stable fork positions (where $m = 0$) could lead to destabilization of the template-tunnel interactions, favouring the entrance to the Long Pause 2 state. Fork destabilization by external tension (*f*) favours DNA unwinding, preventing the entrance to and promoting the exit from the inactive Long Pause 2 state. In 2) and 3) the initial position of the fork is shown in light grey.

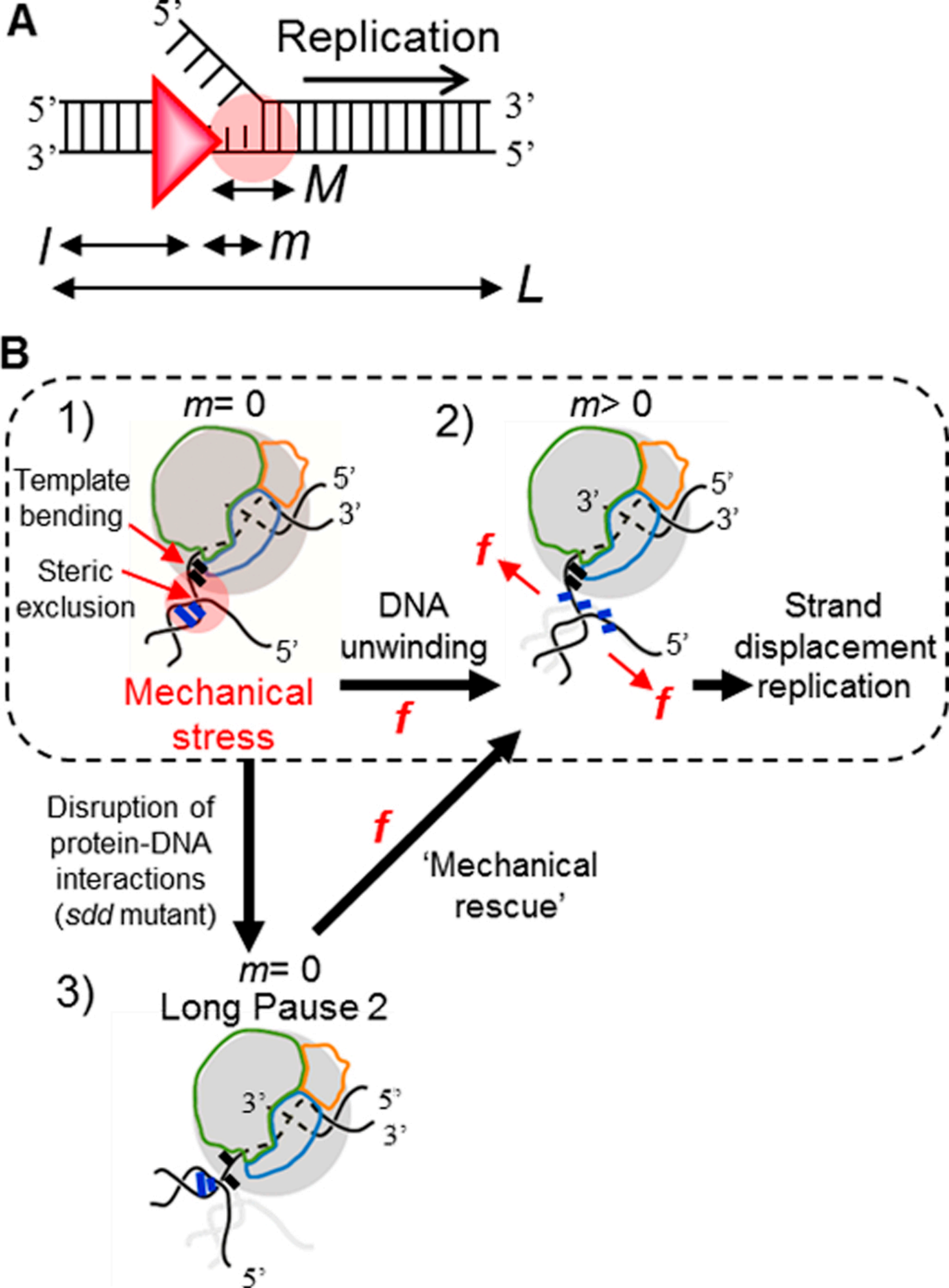
T**able 1.** Rates and conformational changes during primer extension (p.e.) and strand displacement (s.d.) conditions in the absence of tension. Rates (*k*) are in s^{-1}, Arrhenius tension dependences (*d*) in nm and error represents standard error. Similar values were obtained from fits to pause length frequency distributions (Table S1).


A
Force
Laser trap
Pol + dNTPs
Pipette
Force
Δx1
s.d.
p.e.
Δx2
Distance X (μm)
B
wild-type
sdd mutant
s.d.
p.e.
s.d.
p.e.
Δx1
Δx2
Distance (μm)
1.0
1.2
1.4
0
10
20
30
Time (seconds)


A
Mean velocity (nt/s)
Tension (pN)
V_{pe}
V_{sd}
B
seconds/10nt
Tension (pN)
Template Position
C
Mean velocity (nt/s)
Tension (pN)
V_{pe}
V_{sd}
No pauses
V_{sd}
with pauses
D
Template position
Time (seconds)
Velocity (nt/s)
Probability


A
Template position
Time (seconds)
Velocity (nt/s)
Probability
B
Pause Freq Distrib (s-2)
Time (seconds)
C
Pause frequency (s-1)
Tension (pN)
D
Pause frequency (s-1)
%GC
E
Short
Pause 1
k1a
ka1(f)
Active
state


A
Pause length freq distrib (s^{-2})
1
0.1
0.01
1E-3
1E-4
Short
Long
0
5
10
15
20
25
30
35
Time (seconds)
B
Average pause length (s)
10
1
0.1
2
4
6
8
10
12
14
Tension (pN)
C
Pause frequency (s^{-1})
1
0.1
0.01
2
4
6
8
10
12
14
Tension (pN)
D
Pause frequency (s^{-1})
1.5
1.0
0.5
0.0
10
20
30
40
50
60
70
% GC
E
Short
Pause 1
k_{1a}
$k_{a1}(f)$
Active
state
$k_{a2}(f)$
$k_{2a}(f)$
Long
Pause 2


A
5'
Replication
5'
3'
3'
5'
M
l
m
L
B
1)
m= 0
Template bending
Steric exclusion
5'
3'
5'
Mechanical stress
DNA unwinding
f
2)
m> 0
3'
5'
3'
f
5'
f
Strand displacement replication
Disruption of protein-DNA interactions (sdd mutant)
f
'Mechanical rescue'
3)
m= 0
Long Pause 2
3'
5'
3'
5'

	wild-type p.e.	wild-type s.d.	*sdd* p.e.	*sdd* s.d.
$k_{a1}(0)$	0.8 ± 0.12	0.35 ± 0.05	0.7 ± 0.3	0.7 ± 0.3
$k_{a1GC}(0)$	3.2 ± 0.11	0.72 ± 0.11	3.0 ± 1.3	1.5 ± 0.6
d_{a1}	0.5 ± 0.08	0.45 ± 0.08	0.6 ± 0.2	0.5 ± 0.2
d_{a1GC}	0.53 ± 0.18	0.28 ± 0.18	0.6 ± 0.2	0.4 ± 0.2
k_{1a}	1.5 ± 0.9	1.5 ± 1.4	1.4 ± 1.1	1.4 ± 1.2
$k_{a2}(0)$	n.a.	n.a.	n.a.	0.4 ± 0.3
$k_{a2GC}(0)$	n.a.	n.a.	n.a.	0.8 ± 0.6
d_{a2}	n.a.	n.a.	n.a.	0.7 ± 0.3
d_{a2GC}	n.a.	n.a.	n.a.	0.5 ± 0.3
$k_{2a}(0)$	n.a.	n.a.	n.a.	0.02 ± 0.02
d_{2a}	n.a.	n.a.	n.a.	-0.7 ± 0.4

Table 1