Supplementary Information for:

What is all this fuss about Tus?

Comparison of recent findings from biophysical and biochemical experiments

Bojk A. Berghuis[1,#], Vlad-Stefan Raducanu [2], Mohamed M. Elshenawy[2,%], Slobodan Jergic[3], Martin Depken[*,1], Nicholas E. Dixon[*,3], Samir M. Hamdan[*,2], and Nynke H. Dekker[*,1]

[1]Department of Bionanoscience, Kavli institute of Nanoscience, Delft University of Technology, van der Maasweg 9, 2629 HZ Delft, The Netherlands

[2]Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

[3]Centre for Medical and Molecular Bioscience, University of Wollongong, New South Wales 2522, Australia

[#]Present address: Department of Bioengineering, Stanford University, California 94305, United States
[%]Present address: Department of Molecular and Cell Biology, University of California Berkeley, California 94720, United States

[*]Correspondence may be addressed to: s.m.depken@tudelft.nl, nickd@uow.edu.au, samir.hamdan@kaust.edu.sa, or n.h.dekker@tudelft.nl

This supplementary information consists of:

Supplementary Discussion
- Addressing the correlation of fork velocities with stalling in magnetic tweezers experiments
- Addressing the correlation of fork velocities with stalling in flow stretch experiments
- Fitting the force-dependent locking probability

Two Supplementary Figures
Supplementary References

Supplementary Discussion

Addressing the correlation of fork velocities with stalling in magnetic tweezers experiments. Whether there is a causal relationship between fork velocities (whether enzymatic or force-induced) and the incidence of fork stalling is pivotal to model building. Such a relationship would only occur if the rates of both events directly compete with each other by being of the same order of magnitude. In such a case, a correlation between DNA unwinding rate and stalling probability should be observed in every type of single-molecule Tus–*Ter* assay, whether or not an enzyme is present at the DNA fork. In the study by Pandey *et al.* of the T7 replication fork (average rate ~100 bp/s, or an order of magnitude slower than the *E. coli* fork), a velocity dependence is not observed. This in contrast to the *E. coli* replisome, where 50% of the faster-moving forks (1700 bp/s) displace Tus without even transient stoppage, whereas 100% of the slower-moving T7 replication forks are stopped (92% permanently

stopped and 8% paused). These experiments suggest that slower moving motors are indeed more effectively stopped by Tus–*Ter*, but it remains critical to investigate whether a correlation between fork velocity and blocking probability is observed more broadly.

The experiments performed by Berghuis *et al.* represent another extreme when it comes to fork velocities (**Fig. S1**). These magnetic tweezers-induced hairpin openings have much higher (and very narrowly distributed) fork progression rates of ~20-30 kb/s, in agreement with earlier measurements [1] (depending on the hairpin length and accuracy of measurement as a result of the acquisition frequency; see Supplementary **Fig. S1** for details), and 100% of events were blocked at least temporarily during these experiments. This suggests that the rate of fork progression does not compete with the rates at which the molecular rearrangements at Tus–*Ter* take place. Another strong indication that the rates that govern Tus–*Ter* lock formation occur at quite different timescales than base pair unwinding is the observation that in the hairpin experiments, lock pre-formation produced a distribution of lock lifetimes identical to wild-type Tus–*TerB*. This indicates that, given the opportunity to form in an isolated fashion, the Tus–*Ter* lock behaves identically to a lock that has to form within the timeframe allowed by an entity approaching at 20 kb/s.

To try to understand these apparent discrepancies, we have examined both the flow-stretching and magnetic tweezers experiments more critically. First, a clear and crucial difference between the two types of experiments is that the magnetic tweezers experiments employ a mechanical force to induce DNA helix unwinding and concomitant lock formation at the non-permissive face of Tus–*Ter*. In this situation, one must consider whether the DNA geometry at the fork exactly resembles the geometry in a situation with enzyme-induced unwinding. Potentially, the mechanical strand separation might actually pull the C(6) base straight into the lock, thereby enhancing the efficiency of Tus–*Ter* lock formation. Such a 'pulling bias' would then obviously account for some of the discrepancy. Berghuis *et al*. implicitly argue that the identical blocking efficiencies observed for wild-type Tus–*Ter* and pre-formed locked Tus–*Ter* (**Fig. 4a**, 'wt' vs. 'wt 5b mm' in purple) argue against such a bias, as otherwise an increased blocking efficiency for 'wt' Tus–*Ter* would be expected. It may also be the case that a pulling bias in the hairpin experiments would obscure any inefficient (i.e., slow) base rearrangements that need to occur at the fork. In that case, such rearrangements might only be visible upon hairpin rehybridization, which would then be expected to be slower (e.g., on the order of replisome progression speeds) than the observed hairpin opening. However, no such differences are observed: the hairpin closing velocities are roughly equal to the opening velocities – ~26 kb/s for the 7 kb DNA hairpin employed (**Fig. S1c**).

Further evidence substantiates the absence of any significant pulling bias in the hairpin experiments. Both the Elshenawy *et al.* and Berghuis *et al.* studies find evidence for a multistep process, suggesting their different approaches probe the same mechanism. Additionally, one might also expect that a pulling bias might reduce the effect of Tus mutations not directly in contact with C(6). For example, were C(6) to be pulled into the Tus lock pocket as a result of pulling bias in the hairpin experiments, then the mutant E49K might hardly impact the efficiency of lock formation, in contrast to the substantial reduction that is experimentally observed. Furthermore, the mechanical unzipping geometry may actually approximate the fork geometry upon helicase-induced unwinding, as the physical

size of DnaB may result in a substantial wedge between the unwound parental strands. However, further examination of this geometry via structural or simulation studies remains necessary to provide conclusive insights.

Taken together, we conclude that there is no direct evidence for a significant 'pulling bias' affecting lock formation in the hairpin pulling experiments, but put forth that further investigation of the fork geometry in both experiment types will be beneficial.

Addressing the correlation of fork velocities with stalling in flow stretch experiments.
A second point of discussion in understanding the apparent discrepancies between rates of unwinding and probability of blocking in the magnetic tweezers and flow-stretching experiments concerns a further examination of the rate of unwinding. The absence of any unwinding rate dependence in the hairpin pulling experiments indicates that certain rates of barrier formation are altered by the presence of enzymes. As each enzymatic species will create its own specific chemical signature at the nonpermissive face of Tus–*Ter*, there could theoretically exist two conformationally distinct replisome populations that give rise to the populations being blocked or let through. While this scenario seems unlikely as evidenced by the ability to shift the rate dependency of Tus–*Ter* fork arrest efficiency more than 3-fold by altering the *Ter* site, we have nonetheless performed extensive statistical analysis to demonstrate that the *E. coli* replisome constitutes a single population that is stopped by Tus–*Ter* based on rate.

Our investigation starts with the histogram presented in Figure 1g in Elshenawy *et al.* (reproduced here as Supplementary **Fig. S2a**). At first sight, its shape appears both broad and somewhat skewed. While broadness characterizes many histograms constructed by single-molecule imaging techniques (including the velocity histogram for the T7 replication fork [3,4]), the degree of skew could arguably hide multiple populations. To address this, we concatenated the original two data sets, namely the Stop/Restart dataset (N = 31) and the Bypass dataset (N = 33) into a single one (N = 64) thereby omitting the rate-dependent classification with respect to the fate at the Tus–*Ter* site (Supplementary **Fig. S2b**). The empirical skew for this data set was determined to be 0.42. Such a degree of skew could arise from two effects that are distinct from multiple populations: the finite size of N and the existence of an upper or lower boundary for the dataset. In the case of the replisome velocity, the lower bound to the dataset is zero (the rate of DNA synthesis being limited to positive values), thereby inducing right skew [5]. To investigate whether the sample size influenced the degree of skew, we performed computer simulations, which indicated that sample size is a dominant factor (data not shown). For example, simulations indicate that the velocity distribution of a single replisome population with N = 64 can adopt a skew as large as 0.57 (data not shown). Given that unimodal distributions can, in general, accommodate a skew of up to 2.3 [6], the observed 0.42 positive skew in our histogram (Supplementary **Fig. S2b**) does not necessarily imply the presence of multiple replisome populations. In fact, a similar positive skew was observed for the velocity distribution of the T7 replisome, which is stopped by Tus–*Ter* in 100% of cases [7].

We next validated the use of a single Gaussian fit to describe the behavior of the rate of DNA synthesis by a single replisome population. The empirical cumulative distribution function (CDF) of the rate of the replisome fits well with a single Gaussian distribution (R^2 = 0.982)

(Supplementary **Fig. S2c**). Furthermore, use of the Bayesian information criterion (BIC) showed that use of a single Gaussian achieves the lowest BIC, supporting the existence of a single replisome population (Supplementary **Fig. S2d**). Furthermore, to strengthen the observed correlation between rate of the replisome and its fate at the Tus–*Ter* site, we performed overlap integral analysis, denoted as the *j* factor that corresponds to the percentage of the overlap between two distributions, i.e., of the Stopped/Restart and Bypassed distributions. This is opposed to the analysis performed in Elshenawy *et al.* that relied on the correlation coefficient. Both analyses have led to the same outcome, that there is a correlation between rate of DNA synthesis and fate of the replisome at Tus–*Ter* (Supplementary **Fig. S2e**). Bypass and Stop/Restart events are therefore separated only by prior knowledge of the type of event, in agreement with the BIC-based finding that the combined distribution is best described by a single Gaussian (Supplementary **Fig. S2d**). Finally, to eliminate the apparent effect of skew in the rate distribution histogram on the rate-dependent conclusion, we performed a Box-Cox transformation on the concatenated data set without prior knowledge of the fate of the event at Tus–*Ter*. This effectively reduced the skew to ~zero. The inverse of the Box-Cox transformation gave a mean that is very similar to the mean of the untransformed initial data set (Supplementary **Fig. S2f**). This clearly demonstrates that skew does not influence rate determination and that single Gaussian model fitting of the rate is appropriate. It also validates the cause-and-effect interpretation between the rate of the replisome and its fate at Tus–*Ter*.

Fitting the force-dependent locking probability.

For the fit to the force dependence of the Tus full lock probability as shown in **Fig. 4f**, we assume that the probability of full locking is a result of a direct competition between two rates: the rate of entering the locked state k_l and the rate of not going into the locked state $k_\mathrm{u} = 1 - k_\mathrm{l}$, which makes the probability of locking:

$$P_\mathrm{l} = \frac{k_\mathrm{l}}{k_\mathrm{l} + k_\mathrm{u}} \qquad (1)$$

From Kramer's theory of reaction kinetics we have an expression for the rates k_l and k_u:

$$k_\mathrm{l} = k_\mathrm{l}^0 \mathrm{e}^{(E_\mathrm{l}^\ddagger - \delta_\mathrm{l} F)\cdot\beta} \qquad (2)$$

$$k_\mathrm{u} = k_\mathrm{u}^0 \mathrm{e}^{(E_\mathrm{u}^\ddagger - \delta_\mathrm{u} F)\cdot\beta} \qquad (3)$$

where the k^0's are the pre-exponential factors at zero force, $E^\ddagger$ the energy difference between the respective states and the transition state, δ the distance to the transition state, F the applied force and $\beta = [k_\mathrm{B}T]^{-1}$. Defining a reaction rate at zero force $K^0 = k^0 \mathrm{e}^{E^\ddagger\cdot\beta}$, and combining eqns. 2 and 3, we obtain:

$$P_{\mathrm{l}} = \frac{K_{\mathrm{l}}^{0}\mathrm{e}^{-\delta_{\mathrm{l}}\cdot F\cdot\beta}}{K_{\mathrm{l}}^{0}\mathrm{e}^{-\delta_{\mathrm{l}}\cdot F\cdot\beta} + K_{\mathrm{u}}^{0}\mathrm{e}^{-\delta_{\mathrm{u}}\cdot F\cdot\beta}} \tag{4}$$

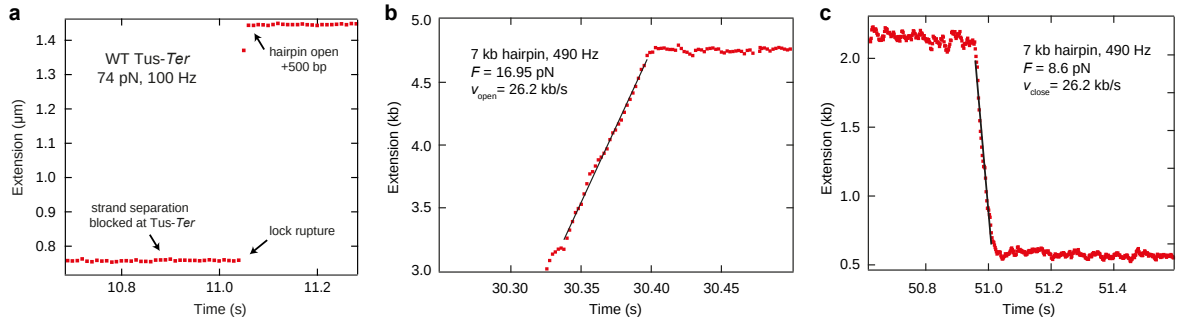
$$= \frac{1}{1 + \frac{K_{\mathrm{u}}^{0}}{K_{\mathrm{l}}^{0}}\mathrm{e}^{F(\delta_{\mathrm{u}}-\delta_{\mathrm{l}})\cdot\beta}} \tag{5}$$

$$= \frac{1}{1 + \mathrm{e}^{(\ln\frac{K_{\mathrm{u}}^{0}}{K_{\mathrm{l}}^{0}}\frac{1}{\delta_{\mathrm{u}}-\delta_{\mathrm{l}}} - F)(\delta_{\mathrm{u}}-\delta_{\mathrm{l}})\cdot\beta}} \tag{6}$$

$$= \frac{1}{1 + \mathrm{e}^{(F_{\mathrm{eq}}-F)\Delta x\cdot\beta}} \tag{7}$$

where $F = \ln K_{\mathrm{u}}^{0}/K_{\mathrm{l}}^{0}\,(\delta_{\mathrm{u}} - \delta_{\mathrm{l}})^{-1}$ and $\Delta x = \delta_{\mathrm{u}} - \delta_{\mathrm{l}}$. Berghuis *et al.* extract the full lock probability by fitting a 3-state kinetic model to the data, making the full lock probability a result of two preceding exponential processes. This would add to another term in P_{l}, but since the trend is coarse-grained, we will not be able to fit out the additional parameters. This is why we denote the distance to the transition state Δx as the *effective* distance to the transition state.

Supplementary Figures



Supplementary Figure S1. DNA hairpin opening and closing velocities in magnetic tweezers experiments. (**a**) A typical experiment of Berghuis *et al.*, where the 1.1 kb hairpin extension is tracked at high force (74 pN) with 100 Hz acquisition frequency. Tus–*Ter* lock rupture allows for further opening of an additional 500 base pairs of dsDNA. The instantaneous velocity, characteristic for systems at low Reynolds numbers, is too high to measure accurately, as there is only 1 datapoint (maximum) between lock rupture and hairpin opening. (**b**) Using a 7 kb hairpin and an acquisition frequency of 490 Hz, the opening velocity can be more accurately measured (data in red, fit in black). (**c**) The closing velocity is very similar to the opening velocity, indicating that in the MT experiments, the ssDNA bases have time to rearrange and form Watson-Crick base pairing at this rate (data in red, fit in black). Note that the estimated velocities in Berghuis' experiments are higher (Supplementary Figure S1d in Berghuis *et al.* [2]); this might be a result of the known effect of the increased drag the 7-fold longer DNA hairpin experiences here [1].

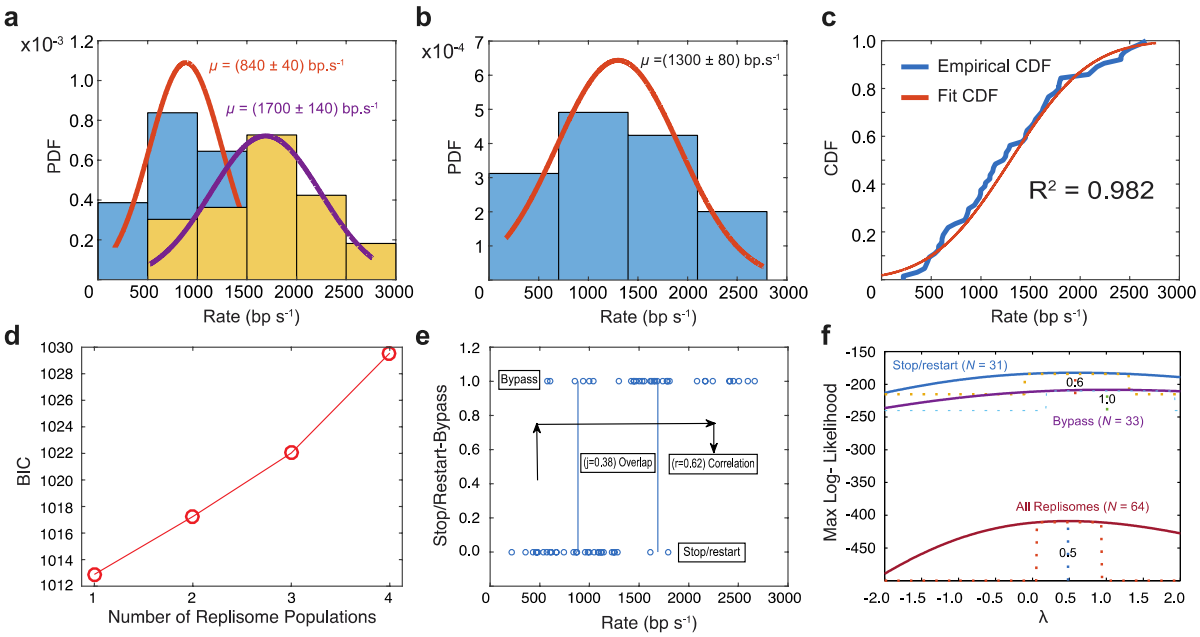


Figure S2. Statistics of *E. coli* replisomes in the presence of Tus–*Ter*. (**a**) Distribution of rates of Stop/Restart (N = 31) and Bypass (N = 33) replisomes. The mean rate of Bypass replisomes (1700 bp/s) is two-fold higher than the mean rate of Stop/Restart replisomes (840 bp/s). The two distributions were constructed by prior knowledge and separation of the type of events. There is a visible overlap region between the two distributions. (**b**) Rates from all replisomes (Stop/Restart and Bypass) are combined into one set of data points. This histogram contains all rates (N = 64), regardless of the type of event. The histogram is fit with a single Gaussian distribution with the mean of 1300 bp/s. (**c**) Plot of the empirical cumulative distribution function (CDF) of the rate distribution from all replisomes (N = 64) and the CDF of the fit single Gaussian distribution, showing an excellent fit (R^2 = 0.982). (**d**) Gaussian mixture model (GMM) with variable number of Gaussian distributions was used to fit the histogram of rates from all replisomes (N = 64). The Bayesian information criterion (BIC) was used as an indicator of the most accurate model. A single Gaussian was found to achieve the lowest BIC value and therefore the replisome was found to be a single population from the perspective of rates. (**e**) The correlation between the type of event and the rates was found to be described by a Pearson correlation coefficient r = 0.62, consistent with Elshenawy *et al*. The overlap integral of the two distributions (j factor) (Stop/Restart and Bypass) was independently calculated and found to be 0.38. (**f**) Data sets of Stop/Restart (N = 31), Bypass (N = 33) and all replisomes (N = 64) were transformed by one-parameter Box-Cox transformation, where the value of the parameter is chosen to be the one with which the transformation minimizes the skewness. All three transformations are convergent for values of the parameter between 0.5 and 1, demonstrating that skewness can be removed without influencing the rate.

Supplementary References

1. Essevaz-Roulet, B., Bockelmann, U. & Heslot, F. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA* **94**, 11935–11940 (1997).
2. Berghuis, B.A. *et al.* Strand separation establishes a sustained lock at the Tus–*Ter* replication fork barrier. *Nat. Chem. Biol.* **11**, 579–585 (2015).
3. Hamdan, S.M., Loparo, J.J., Takahashi, M., Richardson, C.C. & van Oijen, A.M. (2009). Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* **457**, 336–339 (2009).
4. Lee, J.-B., Hite, R.K., Hamdan, S.M., Xie, X.S., Richardson, C.C. & van Oijen, A.M., 2006. DNA primase acts as a molecular brake in DNA replication. *Nature* **439**, 621–624 (2006).
5. Benedict, L.H. & Gould, R.D. Towards better uncertainty estimates for turbulence statistics. *Exp. Fluids* **22**, 129–136 (1996).
6. Johnson, N.L. & Rogers, C.A. The moment problem for unimodal distributions. *Ann. Math. Stat.* **22**, 433–439 (1951).
7. Pandey, M. *et al.* Two mechanisms coordinate replication termination by the *Escherichia coli* Tus–*Ter* complex. *Nucleic Acids Res.* **43**, 5924–5935 (2015).