Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion.

Francesco Vanzi, Chiara Broggio, Leonardo Sacconi, and Francesco Saverio Pavone

SUPPLEMENTARY MATERIALS

Correction of TPM measurements for missed and false events.

Our method is based on an explicit calculation of the number of events that in each trace are missed due to limited time resolution of the filter (false negatives) or created due to threshold-crossing determined by noise residual after filtering (false positives). A similar approach, limited to the missed events only, has been described by Blatz and Magleby (1). The number of missed loop and unloop events (ML and MU, respectively) and created loop and unloop events (which we call false loop and unloop events: FL and FU, respectively) depends on the extent of filtering of the data. Combining considerations on the properties of the filter, the signal-to-noise ratio of the measurements and the kinetics of the system (whose presumed biochemical scheme is shown in figure 6 and discussed in the results and discussion sections), we have obtained the following expressions for ML, MU, CL and CU:

\[
ML = N_U \cdot \frac{k_2}{2a} \left[ (a+b) + e^{(a+b)T_d \tau_L} \right] \frac{1}{(a+b)} \frac{1}{b-a} \left[ a+b + \frac{k_1}{a+b} \right] e^{-T_d \tau_L \left[ \frac{T_L}{\tau_L} + \frac{2T_d \tau_L}{\tau_L} \right]}
\]

\[
MU = N_U \cdot \frac{k_2}{2a} \left[ (a+b) + e^{(a+b)T_d \tau_L} \right] \frac{1}{(a+b)} \frac{1}{b-a} \left[ a+b + \frac{k_1}{a+b} \right] e^{-T_d \tau_L \left[ \frac{T_L}{\tau_L} + \frac{2T_d \tau_L}{\tau_L} \right]}
\]

\[
FL = N_U \cdot \frac{k_2}{2a} \cdot \lambda_{FU} \cdot \left[ (a+b) + e^{(a+b)T_d \tau_L} \right] \frac{1}{(a+b)} \frac{1}{b-a} \left[ a+b + \frac{k_1}{a+b} \right] e^{-T_d \tau_L \left[ \frac{T_L}{\tau_L} + \frac{2T_d \tau_L}{\tau_L} \right]}
\]

\[
FU = N_U \cdot \tau_L \cdot \lambda_{FU} \cdot \exp \left( -\frac{2T_d \tau_L}{\tau_L} \right)
\]

where \( T_d \) is the filter’s dead time and

\[
a = \sqrt{\frac{1}{4}(k_1 + k_{-1} + k_2)^2 - k_1 k_2}
\]

\[
b = -\frac{1}{2}(k_1 + k_{-1} + k_2)
\]

where \( k_2 = J_\text{un} k_\theta \); \( k_1 \) and \( k_{-1} \) are the rates of transition between O-OR and the other prevalent biochemical state corresponding to a TPM unloop (i.e. RO-OR at high LacI concentrations and O-O at low concentrations). So that, at LacI concentrations of 100 pM and 20 pM \( k_1 \) corresponds to \( 2k_d \) and \( k_{-1} \) to \( k_d [\text{LacI}] \), while at a concentration of 4 pM \( k_1 \) corresponds to \( 2k_d [\text{LacI}] \) and \( k_{-1} \) to \( k_d \).
\( \lambda_{FL} \) and \( \lambda_{FU} \) (the frequencies of threshold-crossing for generating false loop and unloop, respectively) were calculated based on the definition provided by Colquhoun and Sigworth (2):

\[
\lambda = 0.849 \cdot f_c \cdot \exp \left( -\frac{\phi^2}{2 \cdot \sigma_n^2} \right)
\]

(7)

where \( f_c \) is the cutoff frequency of the Gaussian filter employed, \( \phi \) is the signal amplitude (the distance between the two peaks in the \( <R(t)> \) histogram and \( \sigma_n \) is the signal standard deviation in each of the two states (\( \sigma_L \) is used for the calculation of \( \lambda_{FU} \) and \( \sigma_U \) is used for the calculation of \( \lambda_{FL} \)). The latter two parameters were measured experimentally on the TPM recordings based on the Gaussian parameters fitted to the \( <R(t)> \) distributions (see figure 2, right panels) on all the traces acquired in the presence of wt-LacI at a concentration of 100 pM. The values obtained, along with the resulting values of \( \lambda \), are reported in the Table SI.

| \( T_d \) (s) | \( T_r \) (N_r) | \( \sigma_L \) (nm) | \( \sigma_U \) (nm) | \( \phi \) (nm) | \( \lambda_{FU} \) (s\(^{-1}\)) | \( \lambda_{FL} \) (s\(^{-1}\)) | \( \tau_{FL} \) (s) | \( \tau_{FU} \) (s) |
|---|---|---|---|---|---|---|---|---|
| 1.35 | 588 (13) | 9.0 ± 0.2 | 11.9 ± 0.3 | 29.6 ± 0.4 | \( (2.9 \pm 0.2) \cdot 10^{-2} \) | \( (5.2 \pm 0.2) \cdot 10^{-2} \) | 2.2 | 1.8 |
| 2.70 | 750 (17) | 6.9 ± 0.2 | 8.9 ± 0.2 | 28.7 ± 0.4 | \( (6.5 \pm 0.9) \cdot 10^{-2} \) | \( (1.5 \pm 0.1) \cdot 10^{-2} \) | 3.7 | 3.0 |
| 4.05 | 765 (17) | 6.5 ± 0.1 | 8.0 ± 0.1 | 27.8 ± 0.5 | \( (3.0 \pm 0.6) \cdot 10^{-2} \) | \( (7.4 \pm 0.8) \cdot 10^{-3} \) | 5.1 | 4.2 |
| 5.40 | 783 (18) | 5.7 ± 0.2 | 7.0 ± 0.1 | 27.9 ± 0.4 | \( (2.0 \pm 0.5) \cdot 10^{-3} \) | \( (4.3 \pm 0.6) \cdot 10^{-3} \) | 6.4 | 5.6 |

Table SI. Signal characteristics of TPM measurements and derived frequencies of false threshold-crossing. The standard deviations (\( \sigma \)) and separation of the peaks of TPM mobility (\( \phi \)) were measured, as described in the text, on the \( <R(t)> \) distributions obtained with LacI at a concentration of 100 pM. From these values the frequencies of threshold-crossing by residual noise (\( \lambda \)) and the average duration of the false events are calculated (see text for details). \( T_r \) is the total recording time (in minutes); \( N_r \) is the number of microspheres used for the statistics.

From these expressions, we describe the number of expected loop and unloop dwells (\( N_{Lc} \) and \( N_{Uc} \)) and their average durations (\( \tau_{Lc} \) and \( \tau_{Uc} \)), starting from their true (and unknown) values (\( N_L \), \( N_U \), \( \tau_L \), and \( \tau_U \)):

\[
N_{Lc} = N_L + FL + FU - ML - MU
\]

(8)

\[
N_{Uc} = N_U + FL + FU - ML - MU
\]

(9)

\[
\tau_{Lc} = \frac{\tau_L \cdot N_L + FL \cdot \tau_{FL} - FU \cdot \tau_{FU} + MU \cdot \tau_{MU} - ML \cdot \tau_{ML}}{N_{Lc}}
\]

(10)

\[
\tau_{Uc} = \frac{1}{N_{Uc}} \left[ \frac{k_0}{2a} \left( \frac{a-b-k_0}{(a-b)^2} + \frac{a+b+k_0}{(a+b)^2} \right) + FL \cdot \tau_{FL} + FU \cdot \tau_{FU} - MU \cdot \tau_{MU} + ML \cdot \tau_{ML} \right]
\]

(11)

where

\[
\tau_{ML} = \tau_L - T_d \cdot \left( \frac{\tau_d}{\tau_L} - 1 \right)^{-1}
\]

(12)
\[ \tau_{MU} = \frac{a + b + k_1}{a + b} \left[ (T_d - \frac{1}{a + b}) e^{(a+b)\gamma_d} + \frac{1}{a + b} \right] + \frac{a - b - k_1}{b - a} \left[ (T_d - \frac{1}{b - a}) e^{(b-a)\gamma_d} + \frac{1}{b - a} \right] \]

\[ \tau_{FL} \text{ and } \tau_{FU} \text{ (the average durations of the false looped and unlooped states, respectively) were calculated with a numerical approach derived from Rice (3) and Watts (4), in any case effectively resulting in values close to } T_d. \]

We then numerically obtain the values of \( J_m \) and \( \tau_L \) that minimize, for each set of data, the following expression:

\[ \left( \frac{N_{Um} - N_{Uc}}{N_{Uc}} \right)^2 + \left( \frac{N_{Um} - N_{Uc}}{N_{Uc}} \right)^2 + \left( \frac{\tau_{Um} - \tau_{Uc}}{\tau_{Uc}} \right)^2 + \left( \frac{\tau_{Um} - \tau_{Uc}}{\tau_{Uc}} \right)^2 \]

(14)

The numerical minimization algorithm was implemented in Mathematica (v.5.0, Wolfram Research). From \( \tau_L \), the value of \( \alpha \) was calculated as \( (2k_d \tau_L)^{-1} \).

To test the validity of this method, a series of numerical simulations were performed. In these simulations, the system was allowed to switch between the states shown in figure 6 with transition rates corresponding to the rate constants shown in the figure. The simulations were run at LacI concentrations of 100 pM, 20 pM, and 4 pM. In the simulations \( k_a \) and \( k_d \) were set to the same values used throughout this work. In each simulation the system starts from the O-O state; at each simulation step a double-precision floating-point random number between 0 and 1 is drawn and the value of this is compared with the probability of transition to another state accessible from the current state. If the latter probability exceeds the value of the random number, then the transition takes place, otherwise the system remains in the current state. The probability of transition to any state is calculated according to the scheme of figure 6 in the paper, using the values of \( k_a \) and \( k_d \) mentioned above, the LacI concentration selected and the values chosen for \( \alpha \) and \( J_m \) (see below) in the specific simulation running. The simulation samples the trajectory of the system on the biochemical scheme of figure 6 at the same frequency of the experimental recordings (50 Hz) and for a duration of one hour for each simulated trace. This procedure produces a trajectory of the system describing its occupancy of the states O-O, O-OR, RO-OR, and ORO as a function of time. For the purpose of simulating the TPM measurement, the O-O, O-OR and RO-OR states are mapped into the TPM unlooped state, while the ORO state is mapped into the TPM looped state. The R(t) trace is then generated from the simulated data adding to the simulated data a Gaussian noise with standard deviation which depends on whether the system is in the unlooped or looped state, so to closely reproduce the signal-to-noise ratio typical of the TPM experimental recordings. The R(t) traces are then filtered exactly as done for the experimental measurements. Figure S1 shows an example of a trace simulated choosing a concentration of LacI of 100 pM, \( \alpha=1 \), and \( J_m = 10^{-10} \) M, and filtered with a Gaussian filter with cutoff frequency of 0.066 Hz.

Different sets of simulations (each set is made of 30 simulated traces corresponding to one hour of TPM recording each) were obtained varying, independently, \( J_m \) (between \( 10^{-8} \) M and \( 10^{-11} \) M) and \( \alpha \) (between 0.1 and 10). Each set of simulated data was analyzed exactly as done for the experimental TPM recordings and the corrected values obtained for \( J_m \) and \( \alpha \) were compared with those that had been set to generate the simulated data. Comparisons between the corrected values obtained with our data analysis method and the true values of the parameters are shown in figure S2. From the figure it can be noticed that the system performs very well in a wide range of \( J_m \) and \( \alpha \) values, especially at the higher LacI.
concentrations (20 and 100 pM). For values of $J_m$ larger than about $10^{-9}$ M, the estimation of $\alpha$ is evidently not reliable. This is an obvious limitation of the TPM measurements performed with microspheres as discussed above. In fact, those large values of $J_m$ would imply transitions O-OR $\rightarrow$ORO with rates of the order of 1 s$^{-1}$ or more; in these conditions, the duration of the unlooped state is largely dominated by the permanence of the system in the RO-OR state and the rate of looping will mainly depend on the rate of exit from this state, rendering the accurate measurement of $J_m$ not achievable. However, it should be noticed that the values of $J_m$ obtained with our method still give a reliable indication of whether this is the case or not. In fact, figure S2 shows that, at [LacI] of 20 pM and 100 pM, for any of the values of $\alpha$ chosen, there is a monotonic correlation of corrected $J_m$ versus true $J_m$. Obtaining corrected values of $J_m$ of the order of $10^{-9}$ M from experimental measurements, thus, would indicate that the measurements are performed at (or below) the limits of the TPM technique. However, this is not the case for the measurements reported in this work, which are well within the range of optimal correlation between corrected and true values for both $\alpha$ and $J_m$.

At a LacI concentration of 4 pM, on the other hand, the reliability of the measurements is somewhat reduced, as also discussed in the paper on the basis of the limitations of the TPM technique itself. The simulations conducted at this lower LacI concentration (figure S2 bottom panels), in fact, show that a measured value of $10^{-10}$ M for $J_m$ does not unambiguously correspond to a true value of $10^{-10}$ M but could actually correspond also to larger true values. These simulations confirm that the estimation of $\alpha$ and $J_m$ from these measurements is not as reliable as at the higher concentrations. These results, combined with the considerations on the TPM method discussed in the paper, strongly indicate that great care needs to be taken when designing dynamic TPM measurements, so to keep the experimental system within the range of optimal validity of this single molecule experimental method and of the data filtering and correction methods that need to be applied to the experimental results. For the purpose of the present work, the LacI concentrations that are optimal for drawing quantitative conclusions from the experimental measurements are 20 pM and 100 pM, where the simulations indicate that no possible bias is taking place.

Estimation of the swelling force ($F_0$) due to the microsphere in TPM.

The one-dimensional distributions ($P_x$) measured in the TPM experiment for the positions of the microsphere centroid (in the absence of LacI, so to avoid transitions between two levels of microsphere mobility) display a characteristic non-Gaussian shape, as noticed also by Pouget et al. (5). Theoretical models previously elaborated for the physical description of the TPM system (6) provided a good prediction of some aspects of the experiment (i.e. the dependence of the average range of microsphere mobility as a function of the DNA tether length) but did not attempt an analysis of $P_x$. In fact, the description of the polymer, based, in those works, on a Gaussian model, would unavoidably lead to a Gaussian shape for $P_x$. We have implemented Metropolis-Monte Carlo simulations of the TPM system based on a Worm-like chain (WLC) description of the polymer and an additional swelling force acting radially ($F_0$), describing the effects of the microsphere. Simulations were run for a 1300bp long DNA molecule with a persistence length of 50 nm (7), and varying the value of $F_0$. At each value of $F_0$ the $\chi^2$ between simulated and experimental data was calculated (Figure S3). The figure shows that the best fit is obtained with $F_0 = 127$ fN. Confidence on the value of $F_0$ estimated with this method was evaluated following Press et al. (8): the range of $F_0$ values within a $\Delta \chi^2 = 4$ (shown by the horizontal dashed lines in the figure) from the minimum $\chi^2$ defines the interval of $F_0$ comprising the true $F_0$ value with a confidence of 95.4%. This range is shown by the vertical dashed lines in the figure and provides our measurement of $F_0 = 127 \pm 14$ fN.
Figure S1. Example of simulated data. The simulation was run at a LacI concentration of 100 pM and with $\alpha = 1.0$ and $J_m = 10^{10}$ M. The $<R(t)>$ trace (left panel) was obtained filtering the simulated data with a cutoff frequency of 0.066 Hz. The right panel shows the corresponding histogram of $<R>$ distribution. The line represents the fit of a double Gaussian to the data, as shown in figures 2 and 3 for the experimental data.
Figure S2. **Numerical validation of correction method.** Correlation of the $\alpha$ (left panels) and $J_m$ (right panels) parameters obtained with our correction method versus true values set in the data simulated for different LacI concentrations, as indicated above each set of panels (see text for description). The lines represent the perfect correlation.
Figure S3. Measurement of $F_0$. The filled circles represent the values of $\chi^2$ between simulated and experimental data for simulations run at different values of $F_0$. The dashed lines define the 95.4% confidence interval on the measured parameter (see text for details). The continuous grey line represents a third-order polynomial fit to the $\chi^2$ data, plotted to provide a means of interpolation between the points. The measured $F_0$ in the text is quoted as best fit $\pm \Delta F_0/2$. 
REFERENCES

1. Blatz, A.L. and Magleby, K.L. (1986) Correcting single channel data for missed events. Biophys. J., 49, 967-980.

2. Colquhoun, D. and Sigworth, F.J. (1983) In Sakmann, B. and Neher, E. (eds.), Single-channel recording. Plenum Press, New York, pp. 191-263.

3. Rice, S.O. (1945) Mathematical analysis of random noise. Bell. Syst. Tech. J., 24, 46-156.

4. Watts, S. (1999) Duration of radar false alarms in band-limited Gaussian noise. IEE Proc.-Radar, Sonar Navig., 146, 273-277.

5. Pouget, N., Dennis, C., Turlan, C., Grigoriev, M., Chandler, M. and Salome, L. (2004) Single-particle tracking for DNA tether length monitoring. Nucleic Acids Res., 32, e73.

6. Qian, H. and Elson, E.L. (1999) Quantitative study of polymer conformation and dynamics by single-particle tracking. Biophys J., 76, 1598-1605.

7. Smith, S.B., Cui, Y. and Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science, 271, 795-799.

8. Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (1992) Numerical recipes in C. The art of scientific computing. Second Edition ed. Cambridge University Press.