Supporting Information

Chen and García 10.1073/pnas.1119552109

SI Text

Number of Dissociation Simulations Required for Statistical Significance. In the preceding work the majority of the all-atom simulations were propagated until complete dissociation of the kissing-loop complex occurred—a stochastic process, which may take many hundreds of nanoseconds depending on the magnitude of the applied force. The long, indefinite simulation lengths at low forces made it necessary to estimate ahead of time how many duplicate trajectories were required for statistically significant rate constants to be extracted. A compromise must therefore be made to balance the need for adequate statistical accuracy in the measurement of the dissociation constants and the practical constraints of finite computing resources. As we have restricted our analysis to single-exponential kinetics (both for the overall dissociation process and for the detailed intermediate model of Fig. 6), we decided to examine the signal/noise ratio of simulated events drawn from a noisy, exponential decay process to assess how many dissociation simulations were needed. A random pure-death process (1) was numerically simulated with an initial population size of 5, 25, and 50 until the total population reached 0. Interevent waiting times were randomly drawn from an exponential distribution with a mean rate of 1.0/s; representative survival curves of 10 such simulations for each initial condition are shown in Fig. S1*A*. The maximum likelihood estimation (MLE) (2) method was then used to attempt to extract out the original rate constant from the individual decay trajectories assuming a pure-decay model. The resulting probability density functions (PDFs) are shown in Fig. S1*B*.

We can see from the PDFs that when there are only five observed events, the estimates range from 0.5 ± 0.5 to 1.5 ± 0.9. Although this brackets the value of the true mean of 1.0/s, these estimates suffer both from poor accuracy and poor precision, with error bars larger than the mean value itself. Therefore, we conclude that analysis of five dissociation events is only sufficient for an order-of-magnitude estimate of the true ensemble dissociation rate, and not sufficient for quantitative comparisons. With 25 events, the estimates for the rate range from 0.75 ± 0.06 to 1.25 ± 0.1; although the true error is still as large as approximately 25%, the increased precision is now sufficient for limited quantitative comparisons to be made. For example, we would be able to distinguish a process with rate 1.0 with another process that differed by as little as factor of 2. Finally, with 50 events, the estimates for the rate range from 0.9 ± 0.05 to 1.2 ± 0.04. Although clearly an improved estimate, it comes at a significantly increased computational cost when we are dealing with approximately 100 ns trajectories for each data point, each of which requires several weeks of computation on 16 cpus. Therefore, on the basis of these results we estimate that about 20 simulations would serve as the best trade-off between statistical accuracy and economical use of computer resources given the scope of the questions we wish to address. It should be noted that many of the steps in the microscopic dissociation model (Fig. 6) have recurrence rates much larger than 1, meaning that these specific transitions are in fact observed much more than 20 times when all data is pooled across 20 trajectories. Several extremely frequent transitions occur hundreds times, at which point MLE methods are no longer required as standard nonlinear least squares fitting becomes the preferred method for extracting rate constants. Conversely, transitions with recurrence rates of <1 can only be treated as order-of-magnitude estimates in the absence of addition information.

Inclusion of Auxillary Information and Censored Events for Rate Estimates. Not all transitions are as well-sampled as the highly recurrent k_2 and k_{-2} transitions, so it is important to include as much information as possible in our parameter estimation. For example, only 11 of the 20 trajectories at 100 pN experienced dissociation within the timescale simulated (400 ns); although we cannot quantify the missing nine dissociation times, we can be sure that they occur at >400 ns, and this information should be included when we estimate the overall dissociation time. The stochastic pure-death model (Eq. **3**) can compensate for these unobserved (censored) events, so long as one knows the rank order of the missing event. In this case, we evaluate Eq. **4** with n_0 set to 20 but only evaluate the product from $n = 1$ to 11, and this procedure allows us to estimate the overall lifetime accounting for the unobserved events (in this case, only a lower bound). An analogous procedure was followed in the calculation of the microscopic rates in Table 2 whenever incomplete transitions were observed due to truncation of the molecular dynamics trajectory.

There is also additional information at branch points where one path is sampled much more than the other (i.e., low recurrence in Table 3). Although direct estimation of the infrequently sampled transition is unreliable due to insufficient statistics (recurrence $\leq$1), there is additional information contained in the recurrence value of the more frequently sampled transition. The estimate can therefore be improved using standard properties of Markovian processes (3).

If our Markovian process is characterized by a transition rate matrix Q, as follows:

$$Q = \begin{bmatrix} -\mu_1 & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & -\mu_2 & \lambda_{ij} & \lambda_{2n} \\ \vdots & \lambda_{ji} & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \dots & -\mu_n \end{bmatrix},$$

then the total transition rate out of state i is characterized by the diagonal element:

$$\mu_i = \sum_{j \neq i} \lambda_{ij}.$$

Given that the system is initially at state i, the probability of a particular transition $i \rightarrow j$ is

$$p_{ij} = \frac{\lambda_{ij}}{\mu_i}.$$

If there are only two competing transitions, and one of the transitions is an irreversible one (i.e., recurrence = 1), then the ratio of the recurrences is the same as the ratio of the probabilities and also of the individual transition rates. We can therefore improve our estimate of the poorly sampled transition by enforcing the recurrence ratio to the highly sampled transition with its correspondingly smaller error bars. Rates denoted by an asterisk (*) in Table 2 were estimated in this fashion.

1. Renshaw E (1991) in *Modelling Biological Populations in Space and Time* (Cambridge University Press, Cambridge, UK), pp 28–36.
2. Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos Trans R Soc A* 222:309–68.
3. Nelson BL (1995) *Stochastic Modeling Analysis and Simulation* (McGraw-Hill, New York).

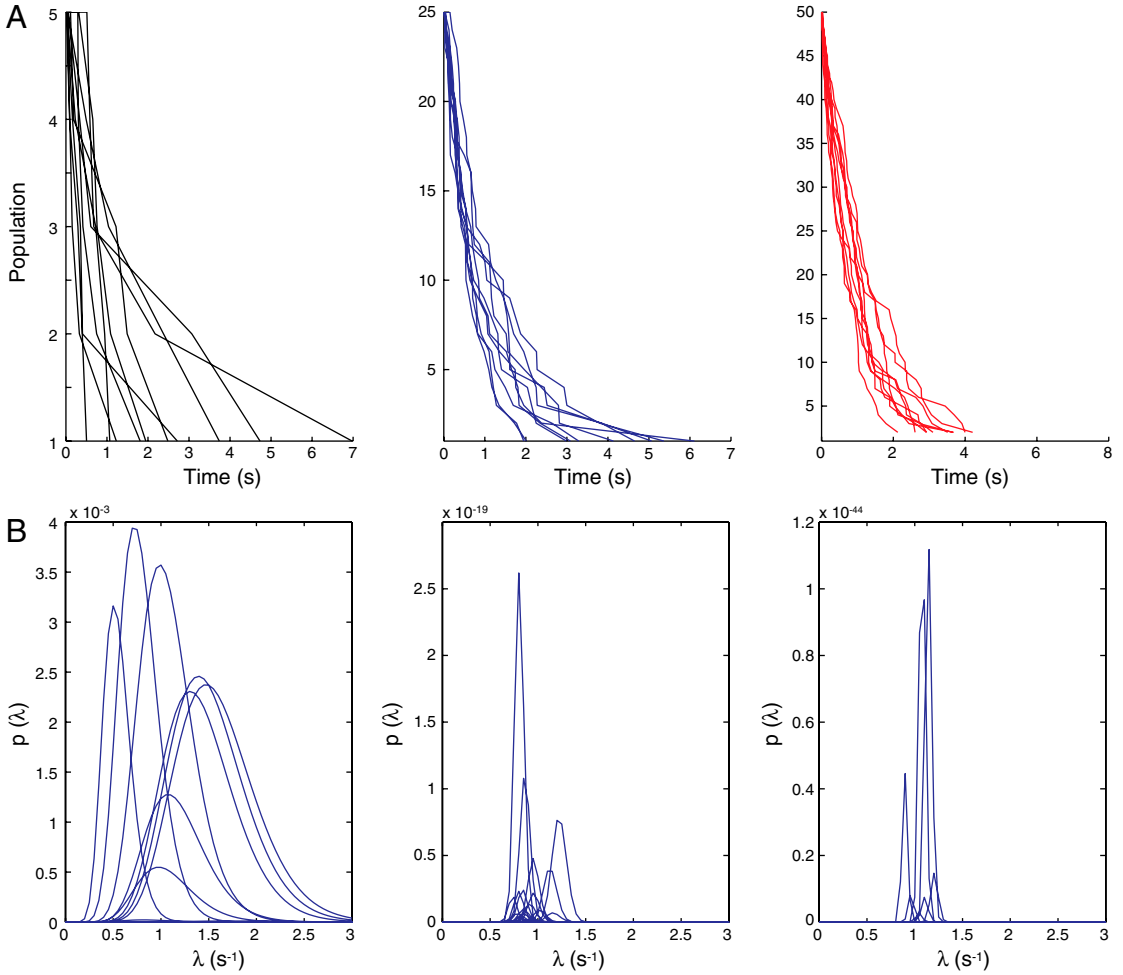
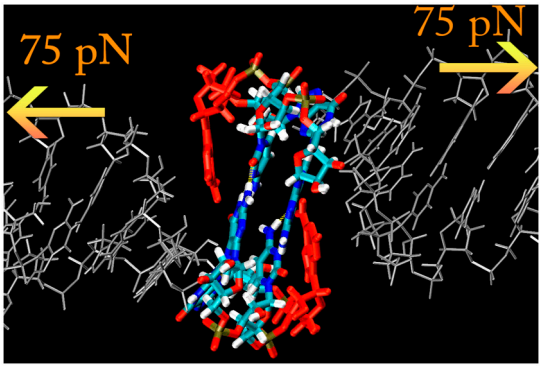
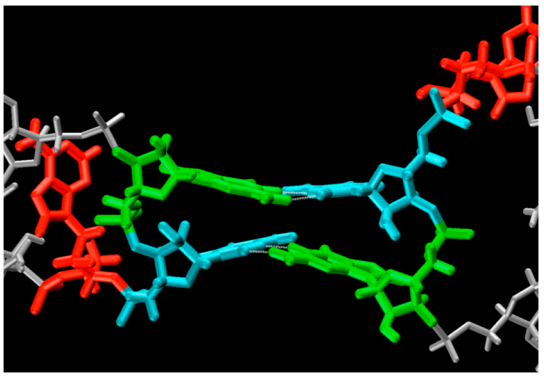


Fig. S1. (*A*) Survival curves of simulated stochastic decay events with initial populations of 5, 25, and 50. Ten trajectories are simulated at each population size to assess the expected signal/noise ratio. The mean decay rate is 1.0/s, and interevent times were randomly drawn from an exponential distribution. (*B*) Probability density functions resulting from maximum likelihood regression of the decay trajectories from *A* using a decay-only (i.e., pure-death) stochastic model. It can be seen that the noise overwhelms the signal with only 5 events, with 25 events a reasonable confidence interval can be ascertained, and that there are diminishing returns when scaling up to 50 events.



Movie S1. This montage of molecular dynamics trajectories shows the initial conformational rearrangement from coaxially stacked to parallel bonds of the MMLV 2-bp complex. It then shows the dissociation pathway for the stack sandwich form of the complex where both flanking stacks are intact at 150 pN constant force. Key nucleotides at the loop-loop interface are color coded: *G*s are green, *C*s are blue, and flanking *A*s are red. Because of file size considerations and long waiting times in-between events, several uninteresting intervening stretches of the trajectory have been omitted as indicated by the simulation time index in the lower right corner. Intermediates are labeled according to the scheme in the summary to this article.

Movie S1 (MPG)



Movie S2. This animation is a molecular dynamics trajectory for the dissociation of the MMLV 2-bp complex with no flanking stacks (open faced) at 150 pN constant force. Key nucleotides at the loop-loop interface are color coded: *G*s are green, *C*s are blue, and flanking *A*s are red. This entire trajectory lasts less than 5 ns.

Movie S2 (MPG)