# Supplementary Material

## Free Energy Calculation and Error Estimation from Non-equilibrium Trajectories (Crooks-Gaussian-Intersection Method)

According to the Crooks Fluctuation Theorem (CFT) the work distributions calculated from the forward (P($W_f$), $\lambda$=0− >1) and reverse trajectories (P($W_r$), $\lambda$=1− >0) are related to the equilibrium free energy via

$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta G)}.$$

In the CGI analysis we approximate the forward and backward distributions by gaussians ( see fig. S1 ) which allows straightforward calculation of the intersection point and hence, the equilibrium free energy difference, via

$$\Delta G_{CGI} = \frac{\frac{W_f}{\sigma_f^2} - \frac{-W_r}{\sigma_r^2} \pm \sqrt{\frac{1}{\sigma_f \sigma_r}(W_f + W_r)^2 + 2\left(\frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2}\right) \ln \frac{\sigma_r}{\sigma_f}}}{\frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2}}$$

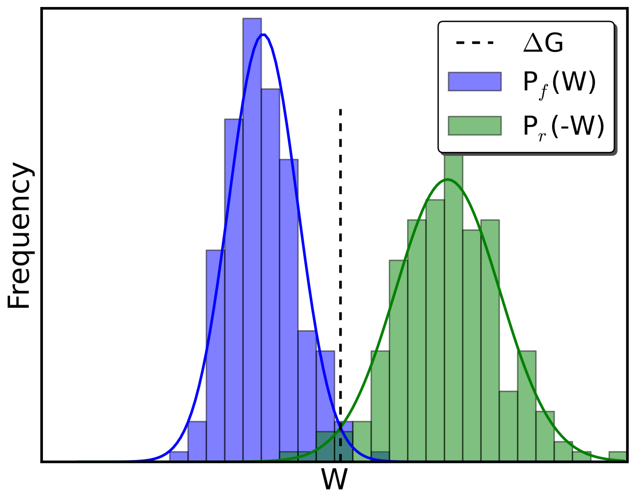


**Fig. S** 1: **Work distributions from non-equilibrium thermodynamic integration runs.** The equilibrium free energy difference is given by the intersection of the work distributions obtained from the forward and reverse non-equilibrium trajectories.

For the estimation of the statistical error we draw $N$ random numbers ( where $N$ is the number of non-equilibrium trajectories) from two gaussian distributions with mean and standard deviation obtained from the forward and reverse work distributions. For each pair of gaussians we calculate the intersection as described above, and thus obtain a distribution of free energies $P(\Delta G)$ (see fig. S2). The standard deviation of this distribution is used as an estimator of the statistical error.

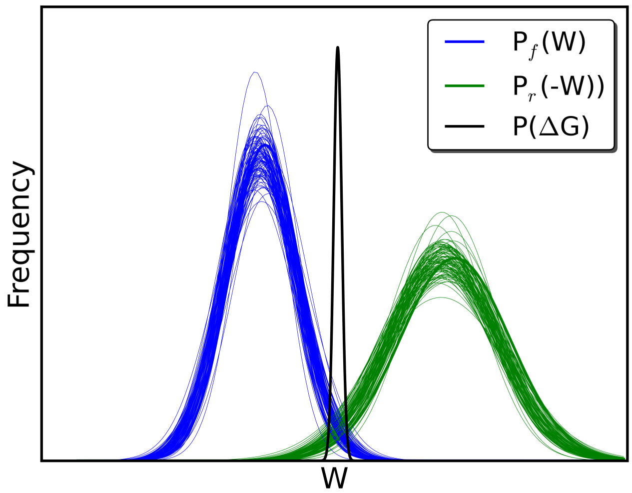


**Fig. S** 2: **Estimation of the statistical error.** Random gaussian distributions are generated for the forward and reverse work distributions. The distribution of the intersection point gives the statistical error.

## Influence of Switching Times on Histogram Overlaps

The switching time during which the system is transformed from one state to the other (the mutation) determines how far the system is driven out of the equilibrium. In the limit of infinitely slow switching the system remains in equilibrium and the dissipative work performed on the system is zero, resulting in identical histograms for the forward and the reverse work. Hence, the faster a transition is done, the more dissipative work is performed driving the histograms apart ( see fig. S3).

The optimal switching time depends on the magnitude of the perturbation and needs

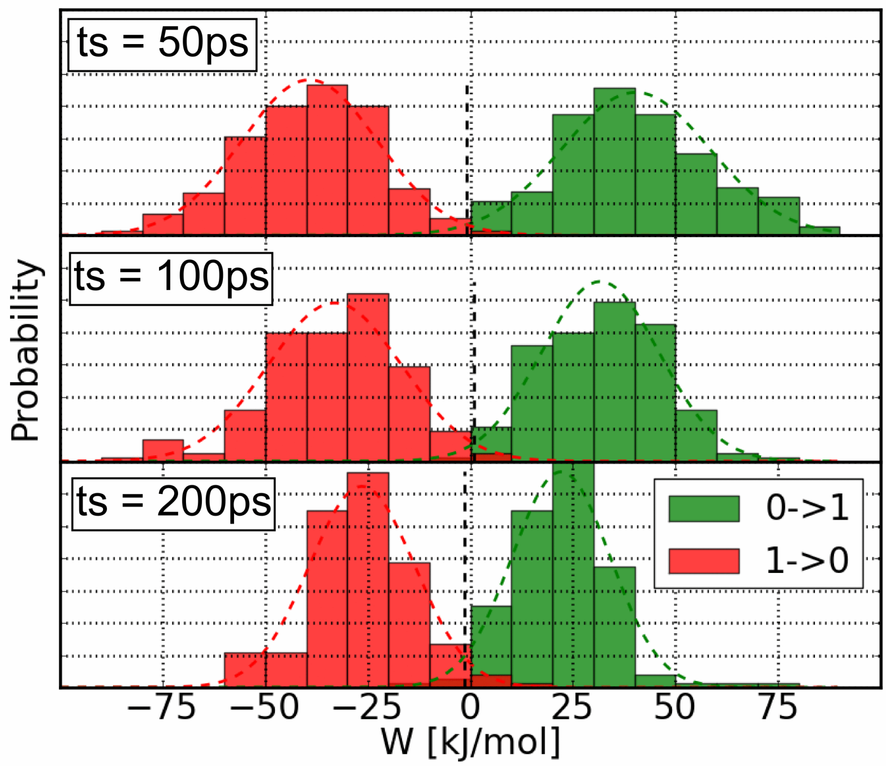


**Fig. S 3: Work histograms for different switching times.** The width of the distributions of work values depends on the switching time used for the fast-growth TI runs. At $t_s = 50\,\text{ps}$ the histograms for the forward and the backward work distributions hardly overlap, resulting in a larger statistical error.

to be adjusted for varying cases. For point mutations we found switching times from 50 to 100 ps to be sufficient, for double mutations, as for the DNA base pair exchange studied here, 100 to 200 ps are required.

## Error Estimation from Equilibrium Trajectories (MBAR method)

For the MBAR estimates of free energy differences, based on equilibrium samples of ensembles along the transformation pathway, we first applied the method of Chodera et al. (2) to estimate statistical inefficiency. Based on this estimate, sub-samples of uncorrelated data points were extracted from the full, time-correlated data sets. This sub-sample of independent data points was then used to estimate the sampling error as described by Shirts and Chodera (1).

# References


1. Shirts, M. R., and J. D. Chodera, 2008. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* 129:124105.

2. Chodera, J., W. Swope, J. Pitera, C. Seok, and K. Dill, 2007. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput* 3:26–41.