A hybrid MD-KMC algorithm for folding proteins in explicit solvent - Supplementary information†

Emanuel Karl Peter,*[a] **Joan-Emma Shea,**[a‡]

[a‡] *University of California Santa Barbara, Department of Chemistry and Biochemistry, Department of Physics, Santa Barbara, CA 93106, USA; E-mail: shea@chem.ucsb.edu*

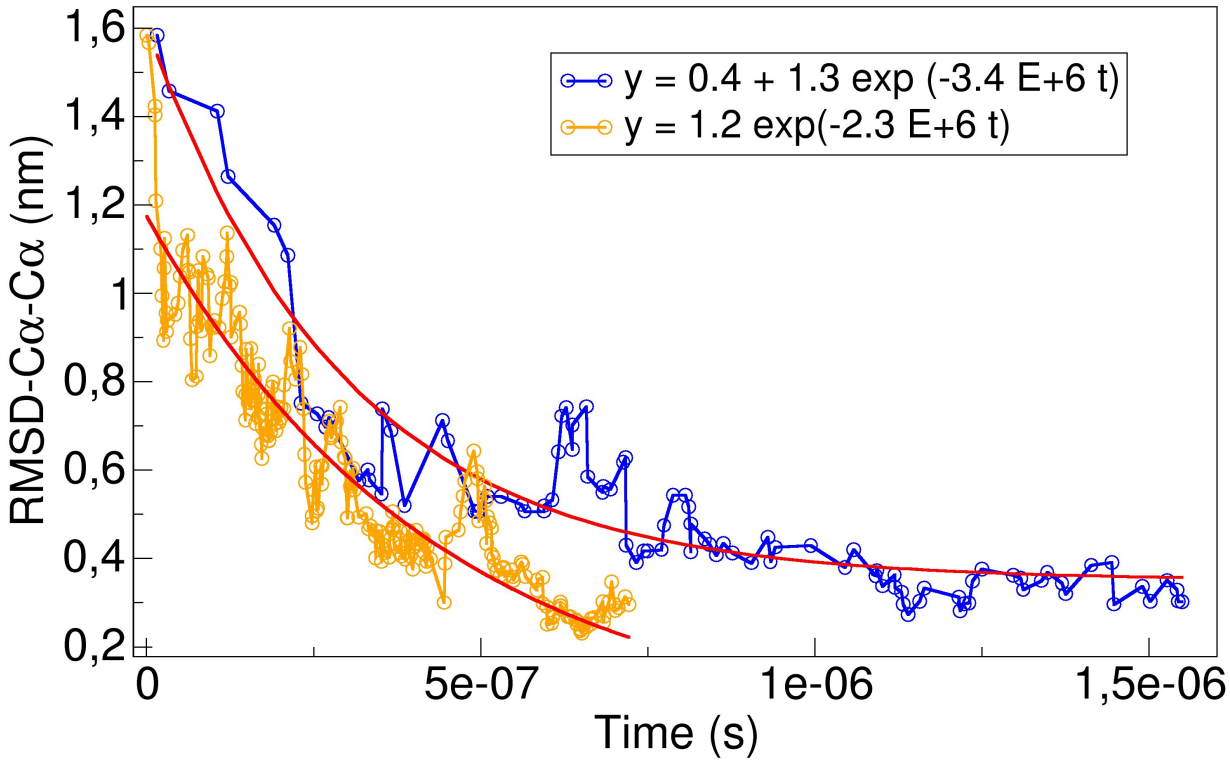

Fig. 1S Single exponential fit of the $RMSD_{C\alpha-C\alpha}$ as function of simulation time for two trajectories from $Method_a$.

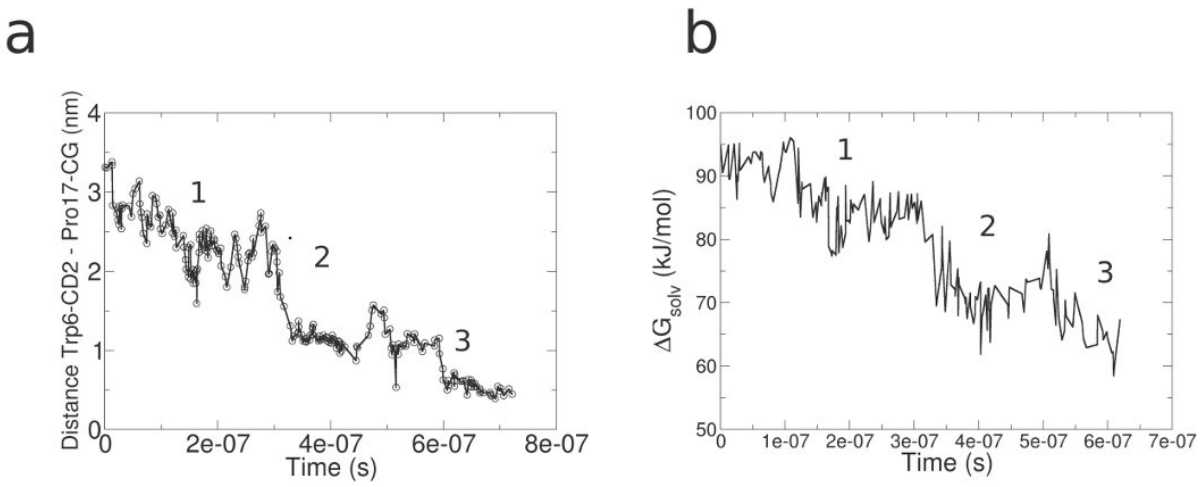

Fig. 2S Distance between atoms Trp6-CD2 and Pro17-CD as function of simulation time (a). Solvation free energy of TrpCage as function of simulation time (b). The numbers 1, 2, 3 indicate the different primary structures visited during the folding process of TrpCage. These structures are shown in the main body of the text in figure 3.

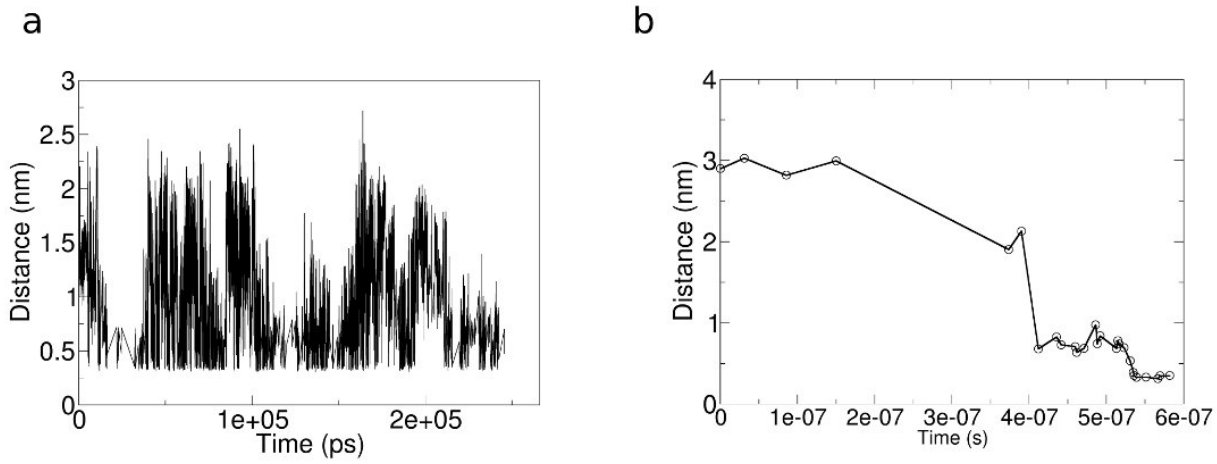

Fig. 3S Distance between Asp9-OD1 and Arg16-CZ as a measure for the stability of the Asp9-Arg16 salt-bridge in implicit solvent (a) and in explicit solvent (b).

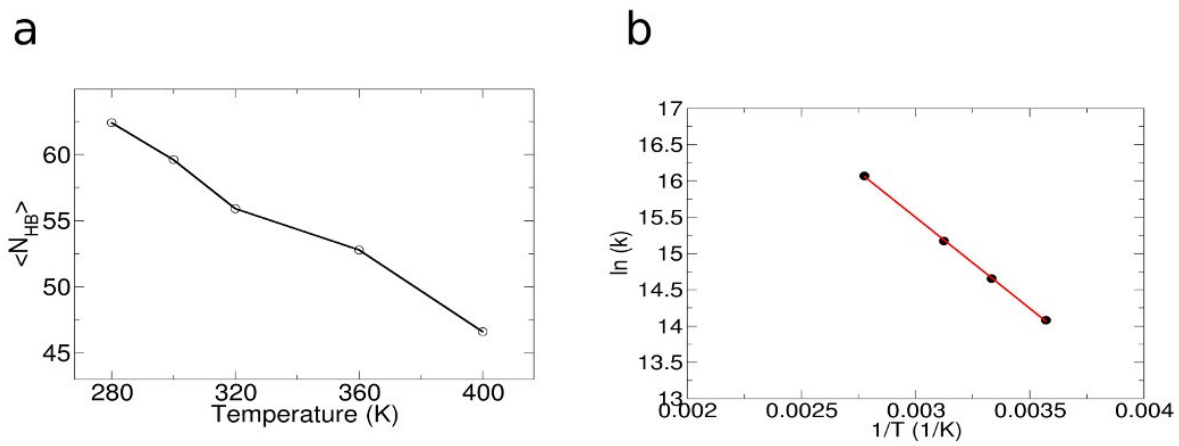

Fig. 4S Average number of H-bonds as a function of reciprocal simulation temperature (a). Folding rate constant as function of temperature (b), using $Method_a$.

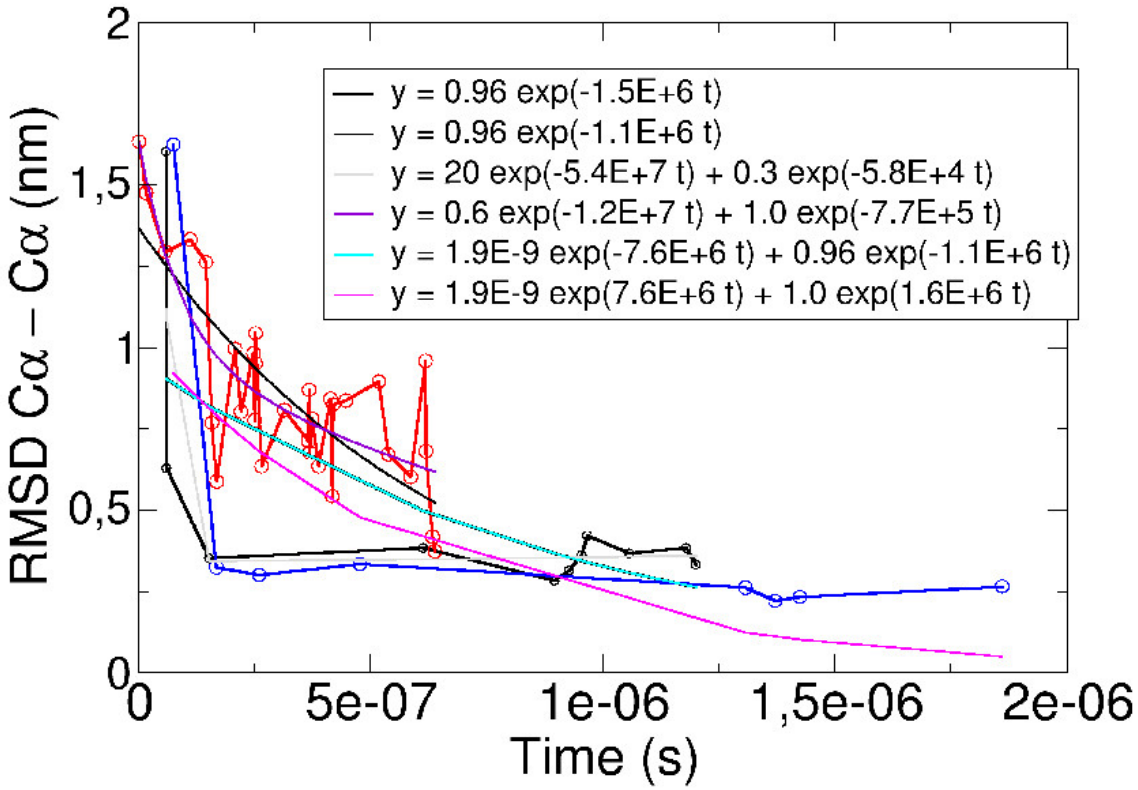

Fig. 5S Single and double exponential fit of the $RMSD_{C\alpha-C\alpha}$ as function of simulation time of 2 simulations using $Method_b$. Neither fits the trajectories well.

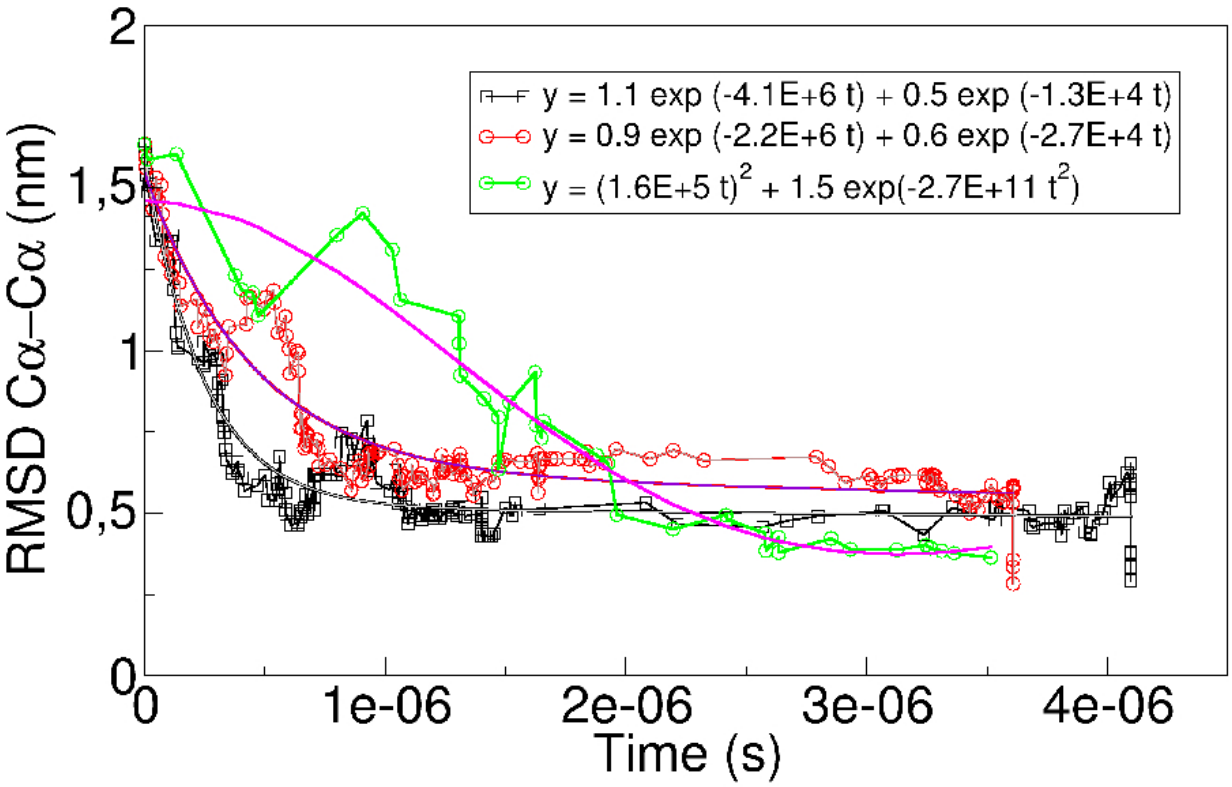

Fig. 6S Single and double exponential fits on the $RMSD_{C\alpha-C\alpha}$ as function of simulation time of 3 simulations using $Method_c$. We observe one fast component, corresponding to the collapse of the peptide and one very slow component.

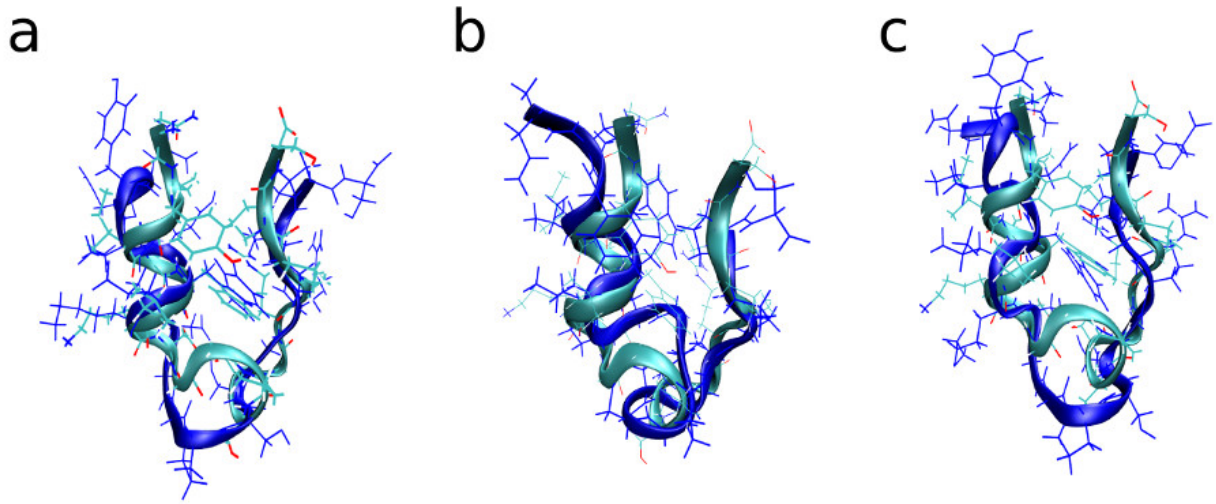

Fig. 7S Overlay of final structures with NMR-solution structure of Neidigh *et al.* (PDB: 1L2Y) (cyan). $Method_a$ (a), $Method_b$ (b) and $Method_c$ (c) (blue).

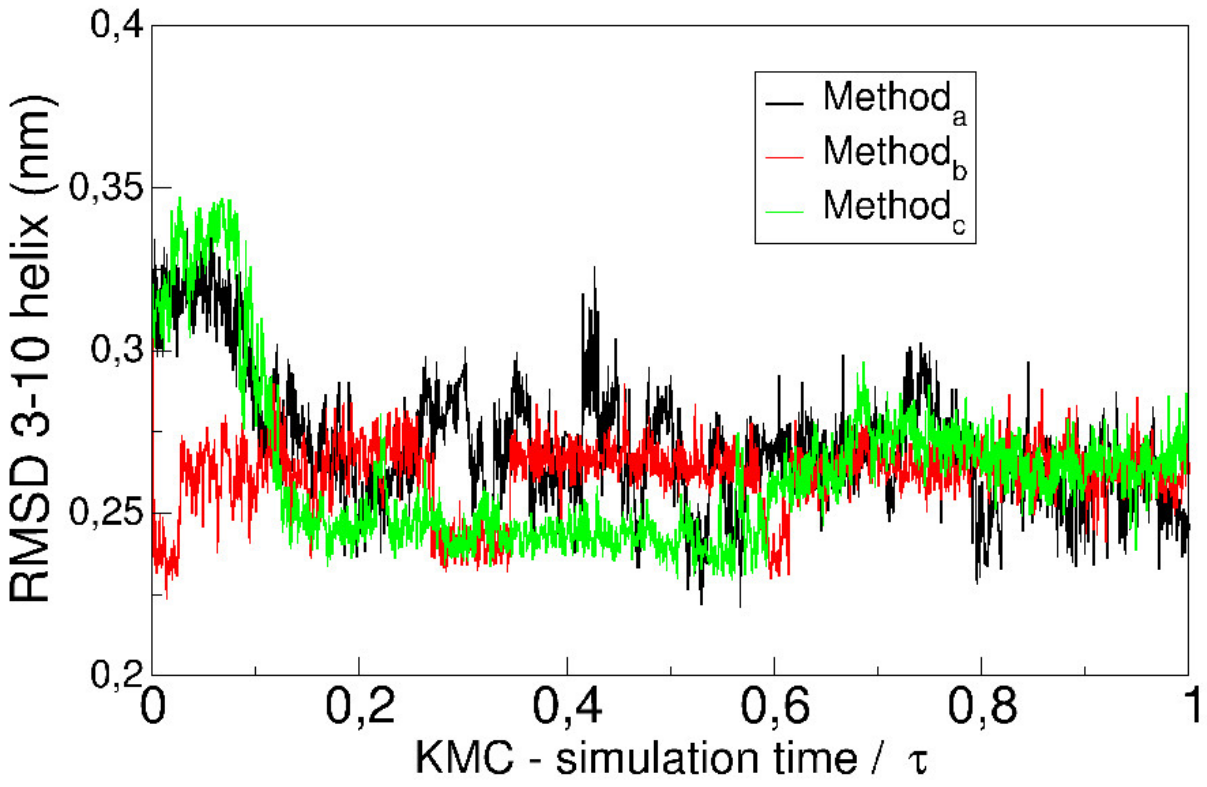

Fig. 8S $RMSD_{3-10}$ helix to the native structure of TrpCage from simulations with the three implemented methods (a-c) as function of normalized simulation time.

1 Determination of friction coefficient

For the determination of the friction coefficient γ, we implemented an algorithm, which follows a sequence of virtual events generated randomly and updates the time according to the Bortz-Kalos Lebowitz algorithm. We chose an appropriate range for the activation energies and number of aminoacids. Within this algorithm, we changed the value of the friction coefficient γ, in order to determine *a priori* a good estimate for γ. From different runs with different values for γ ranging from $1 ps^{-1}$ to $40000 ps^{-1}$, as shown in Figure 9S, where we show the total simulation time as function of KMC-steps. We found that the update in time converges to at a total time of approximately 1 μs at values of γ larger than 10000 ps^{-1}. However, we chose a value of 4000 ps^{-1}, since we aimed at simulating in a brownian regime, were the protein is mostly decoupled from the solvent. The algorithm consists of following steps :

- Select a number of aminoacids N, which can perform N_2 events, by solving following equation :

$$N = N_{max}\xi \ , \ \xi \in]0,1] \ ; \tag{1}$$

  where $N_{max} = 4$ and ξ is a equally distributed random number.

- Select a activation energy $\Delta G_{ij}^{\dagger}$ for each event j and each aminoacid i, by :

$$\Delta G_{ij}^{\dagger} = \xi \Delta G_{range} + G_1 \ , \ \xi \in]0,1]; \tag{2}$$

  where $\Delta G_{range} = 1000 \ kJ/mol$ and $G_1 = 100 \ kJ/mol$.

- Select the frequency factor ω_{ij} for each event j and each aminoacid i, by :

$$\omega_{ij} = \xi \Delta \omega_{range} + \omega_1 \ , \ \xi \in]0,1]; \tag{3}$$

  where $\Delta \omega_{range} = 10 \times 10^{-3} \ ps^{-1}$ and $\omega_1 = 1 \times 10^{-3} \ ps^{-1}$.

- Continue with

$$\nu_{ij} = \left[\left(\frac{\gamma^2}{4} + \omega_b(ij)^2 \right)^{\frac{1}{2}} - \frac{\gamma}{2} \right] \frac{\omega_a(ij)}{2\pi\omega_b(ij)} \ ; \tag{4}$$

  where γ is the friction coefficient, and $\omega_j(i) \ j = a,b$ are the squared angular frequencies inside each metastable minimum along the process i.

- **Determination of rate r_i for process path i**, according to :

$$r_{ij} = \nu_{ij} \exp\left(\frac{\Delta G_{ij}^{\dagger}}{RT} \right) \ ; \tag{5}$$

  where $R = k_B N_A$, where k_B is Boltzmann's constant, N_A is Avogadro's number and T the absolute temperature.

  The cumulative sum R_j is calculated according to :

$$R_j = \sum_{i=1}^{N} \sum_{j=1}^{N_2} r_{ij} \; ; \tag{6}$$

until $j = N$, where N is the total number of processes. Using a random number $\xi \in]0,1]$, one process j is chosen, solving following equation :

$$R_{j-1} < \xi R_N \leq R_j \tag{7}$$

- **Execution of event j**, update of time, using

$$\Delta t = -\frac{\ln \xi}{R_N} \; , \tag{8}$$

with a newly generated random number $\xi \in]0,1]$.

- **Continuation with 1.**

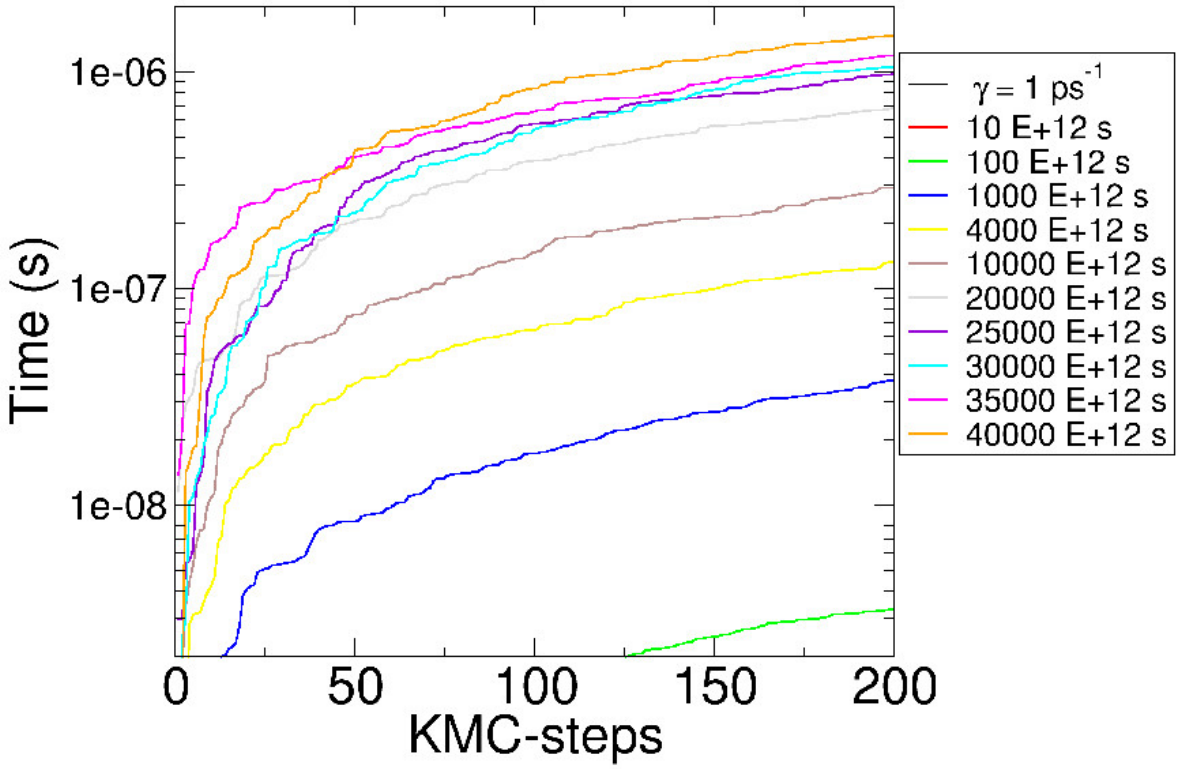

Fig. 9S Simulation time as function of KMC-steps from KMC-sampling at different γ-values ranging from 1 ps^{-1} to 40000 ps^{-1}.