

Multisecond ligand dissociation dynamics from atomistic simulations

Steffen Wolf ¹✉, Benjamin Lickert ¹, Simon Bray ^{1,2} & Gerhard Stock ¹✉

Coarse-graining of fully atomistic molecular dynamics simulations is a long-standing goal in order to allow the description of processes occurring on biologically relevant timescales. For example, the prediction of pathways, rates and rate-limiting steps in protein-ligand unbinding is crucial for modern drug discovery. To achieve the enhanced sampling, we perform dissipation-corrected targeted molecular dynamics simulations, which yield free energy and friction profiles of molecular processes under consideration. Subsequently, we use these fields to perform temperature-boosted Langevin simulations which account for the desired kinetics occurring on multisecond timescales and beyond. Adopting the dissociation of solvated sodium chloride, trypsin-benzamidine and Hsp90-inhibitor protein-ligand complexes as test problems, we reproduce rates from molecular dynamics simulation and experiments within a factor of 2–20, and dissociation constants within a factor of 1–4. Analysis of friction profiles reveals that binding and unbinding dynamics are mediated by changes of the surrounding hydration shells in all investigated systems.

¹Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Hermann-Herder-Strasse 3, 79104 Freiburg, Germany. ²Present address: Bioinformatics Group, Department of Computer Science, Albert Ludwigs University, Georges-Koehler-Allee 106, 79110 Freiburg, Germany.
✉email: steffen.wolf@physik.uni-freiburg.de; stock@physik.uni-freiburg.de

Classical molecular dynamics (MD) simulations in principle allow us to describe biomolecular processes in atomistic detail¹. Prime examples include the study of protein complex formation² and protein–ligand binding and unbinding^{3,4}, which constitute key steps in biomolecular function. Apart from structural analysis, the prediction of kinetic properties has recently become of interest, since optimized ligand binding and unbinding kinetics have been linked to an improved drug efficacy^{5–9}. Since these processes typically occur on time-scales from milliseconds to hours, however, they are out of reach for unbiased all-atom MD simulations which currently reach microsecond timescales. To account for rare biomolecular processes, a number of enhanced sampling techniques^{10–18} have been proposed. These approaches all entail the application of a bias to the system in order to enforce motion along a usually one-dimensional reaction coordinate x , such as the protein–ligand distance.

While the majority of the above methods focuses on the calculation of the stationary free energy profile $\Delta G(x)$, several approaches have recently been suggested that combine enhanced sampling with a reconstruction of the dynamics of the process^{19–21}. In this vein, we recently proposed dissipation-corrected targeted MD (dcTMD), which exerts a pulling force on the system along reaction coordinate x via a moving distance constraint²². By combining a Langevin equation analysis with a cumulant expansion of Jarzynski’s equality²³, dcTMD yields both $\Delta G(x)$ and the friction field $\Gamma(x)$. Reflecting interactions with degrees of freedom orthogonal to those which define the free energy, the friction accounts for the dynamical aspects of the considered process. In this work, we go one step further and use $\Delta G(x)$ and $\Gamma(x)$ to run Langevin simulations, which describe the coarse-grained dynamics along the reaction coordinate and reveal timescales and mechanisms of the considered process. Moreover, we introduce the concept of “temperature boosting” of the Langevin equation, which allows us to speed up the calculations by several orders of magnitude in order to reach biologically relevant timescales.

Results

Dissipation-corrected targeted molecular dynamics. To set the stage, we briefly review the working equations of dcTMD derived in²². TMD as developed by Schlitter et al.²⁴ uses a constraint force f_c that results in a moving distance constraint $x = x_0 + v_c t$ with a constant velocity v_c . The main assumption underlying dcTMD is that this nonequilibrium process can be described by a memory-free Langevin equation¹,

$$m\ddot{x}(t) = -\frac{dG}{dx} - \Gamma(x)\dot{x} + \sqrt{2k_B T\Gamma(x)} \xi(t) + f_c(t), \quad (1)$$

which contains the Newtonian force $-dG/dx$, the friction force $-\Gamma(x)\dot{x}$, as well as a stochastic force with white noise $\xi(t)$, that is assumed to be of zero mean, $\langle \xi \rangle = 0$, delta-correlated, $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$, and Gaussian distributed. Since the constraint force f_c imposes a constant velocity on the system ($\dot{x} = v_c$), the total force $m\ddot{x}$ vanishes. Performing an ensemble average $\langle \dots \rangle$ of Eq. (1) over many TMD runs, we thus obtain the relation²²

$$\Delta G(x) = \langle W(x) \rangle - v_c \int_{x_0}^x \Gamma(x') dx'. \quad (2)$$

Here the first term $\langle W(x) \rangle = \int_{x_0}^x \langle f_c(x') \rangle dx'$ represents the averaged external work performed on the system, and the second term corresponds to the dissipated work $W_{\text{diss}}(x)$ of the process expressed in terms of the friction $\Gamma(x)$.

While the friction in principle can be calculated in various ways^{25,26}, it proves advantageous to invoke Jarzynski’s identity²³,

$e^{-\Delta G(x)/k_B T} = \langle e^{-W(x)/k_B T} \rangle$, which allows us to calculate $\Gamma(x)$ directly from TMD simulations. To circumvent convergence problems associated with the above exponential average²⁷, we perform a second-order cumulant expansion which gives Eq. (2) with $W_{\text{diss}}(x) = \langle \delta W^2(x) \rangle / k_B T$. Expressing work fluctuations δW in terms of the fluctuating force δf_c , we obtain for the friction²²

$$\Gamma(x) = \frac{1}{k_B T} \int_{t_0}^{t(x)} \langle \delta f_c(t) \delta f_c(t') \rangle dt', \quad (3)$$

which is readily evaluated directly from the TMD simulations.

As discussed in ref. 22, the derivation of Langevin Eq. (1) assumes that the pulling speed v_c is slow compared to the timescale of the bath fluctuations, such that the effect of f_c can be considered as a slow adiabatic change²⁸. This means that the free energy Eq. (2) and the friction Eq. (3) determined by the nonequilibrium TMD simulations correspond to their equilibrium results. As a consequence, we can use $\Delta G(x)$ and $\Gamma(x)$ to describe the unbiased motion of the system via Langevin Eq. (1) for $f_c = 0$. Numerical propagation of the unbiased Langevin equation then accounts for the coarse-grained dynamics of the system. In this way, calculations of $\Delta G(x)$ and $\Gamma(x)$ as well as dynamical calculations are based on the same theoretical footing (i.e., the Langevin equation), and are therefore expected to yield a consistent estimation of the timescales of the considered process. Moreover, the exact solution of the Langevin equation allows us to directly use the computed fields $\Delta G(x)$ and $\Gamma(x)$ and thus to avoid further approximations²⁹.

The theory developed above rests on two main assumptions. For one, we have assumed that the Langevin Eq. (1) provides an appropriate description of nonequilibrium TMD simulations, and applies as well to the unbiased motion ($f_c = 0$) of the system. This means that, due to a timescale separation of slow pulling speed and fast bath fluctuations, the constraint force f_c enters this equation merely as an additive term. Secondly, to ensure rapid convergence of Jarzynski’s identity, we have invoked a cumulant expansion to derive the friction coefficient in Eq. (3), which is valid under the assumption that the distribution of the work is Gaussian within the ensemble. While this assumption may break down if the system of interest follows multiple reaction paths, we have recently shown that we can systematically perform a separation of dcTMD trajectories according to pathways by a nonequilibrium principal component analysis of protein–ligand contacts³⁰. This approach bears similarities with the work of Tiwary et al. for the construction of path collective variables³¹. Alternatively, path separation can be based on geometric distances between individual trajectories, making use of the NeighborNet algorithm³². Details on the convergence of the free energy and friction estimates, the path separation, and the choice of the pulling velocity are given in the Supplementary Methods and in Supplementary Figs. 1–4.

T-boosting. The speed-up of Langevin Eq. (1) compared to an unbiased all-atom MD simulation is due to the drastic coarse graining of the Langevin model (one instead of $3N$ degrees of freedom, N being the number of all atoms). Since the numerical integration of the Langevin equation typically requires a time step of a few femtoseconds (see Supplementary Table 1), however, we still need to propagate Eq. (1) for $\gtrsim 100 \times 10^{15}$ steps to sufficiently sample a process occurring on a timescale of seconds, which is prohibitive for standard computing resources.

As a further way to speed up calculations, we note that the temperature T enters Eq. (1) via the stochastic force, indicating that temperature is the driving force of the Langevin dynamics. That is, when we consider a process described by a transition rate

k and increase the temperature from T_1 to T_2 , the corresponding rates k_1 and k_2 are related by the Kramers-type expression²⁹

$$k_2 = k_1 e^{-\Delta G^\ddagger(\beta_2 - \beta_1)}, \quad (4)$$

where ΔG^\ddagger denotes the transition state energy and $\beta_i = 1/k_B T_i$ is the inverse temperature. Hence, by increasing the temperature we also increase the number n of observed transition events according to $n_2/n_1 = k_2/k_1$.

To exploit this relationship for dcTMD, we proceed as follows. First we employ dcTMD to calculate the Langevin fields $\Delta G(x)$ and $\Gamma(x)$ at a temperature of interest T_1 . Using these fields, we then run a Langevin simulation at some higher temperature T_2 , which results in an increased transition rate k_2 and number of events n_2 . In particular, we choose a temperature high enough to sample a sufficient number of events ($N \gtrsim 100$) for some given simulation length. In the final step, we use Eq. (4) to calculate the transition rate k_1 at the desired temperature T_1 .

As Eq. (4) arises as a consequence²⁹ of Langevin Eq. (1), the above described procedure, henceforth termed T -boosting, involves no further approximations. It exploits the fact that we calculate fields $\Delta G(x)$ and $\Gamma(x)$ at the same temperature for which we eventually want to calculate the rate. We wish to stress that this virtue represents a crucial difference to temperature accelerated MD³³. In the latter method the free energy $\Delta G(x)$ is first calculated at a high temperature and subsequently rescaled to a desired low temperature, whereupon $\Delta G(x)$ in general does change. T -boosting avoids this, because by using dcTMD we calculate $\Delta G(x)$ right away at the desired temperature. We note in passing that a Langevin simulation run at T_2 using fields obtained at T_1 in general does not reflect the coarse-grained dynamics of an MD simulation run at T_2 , but can only be used to recover k_1 from k_2 .

In practice, we perform T -boosting calculations at several temperatures T_2 in increments of 25 K to 50 K and choose the smallest T_2 such that $N \gtrsim 100$ transitions occur. In the Supporting Methods we derive an analytic expression of the extrapolation error as a function of boosting temperatures and achieved number of transitions, from which the necessary length of the individual Langevin simulations can be estimated, in order to achieve a desired extrapolation error. One-dimensional Langevin simulations require little computational effort (1 ms of simulation time at a 5 fs time step take ~ 6 h of wall-clock time on a single CPU) and are trivial to parallelize in the form of independent short runs. Hence the extrapolation error due to boosting can easily be pushed below 10% and is thus negligible in comparison to systematic errors coming from the dcTMD field estimates. As shown in Supplementary Table 1, a further increase in efficiency can be achieved if the considered dynamics is overdamped, which

is the case for both protein–ligand systems. Since overdamped dynamics neglects the inertia term $m\ddot{x}$ and therefore does not depend on the mass m , we may artificially enhance the mass in the Langevin simulations. For the protein–ligand systems, this allows us to increase the integration time step from 1 to 10 fs, i.e., a speed-up of an order of magnitude.

Ion dissociation of NaCl in water. To illustrate the above developed theoretical concepts and test the validity of the underlying approximations, we first consider sodium chloride in water as a simple yet nontrivial model system. For this system, detailed dcTMD as well as long unbiased MD simulations are available²², making it a suitable benchmark system for our approach. Fig. 1a shows the free energy profiles $\Delta G(x)$ along the interionic distance x , whose first maximum at $x \approx 0.4$ nm corresponds to the binding–unbinding transition of the two ions. The second smaller maximum at $x \approx 0.6$ nm reflects the transition from a common to two separate hydration shells³⁴. We find that results for $\Delta G(x)$ obtained from a 1 μ s long unbiased MD trajectory and from dcTMD simulations (1000×1 ns runs with $v_c = 1$ m/s) match perfectly. Since the average work $\langle W(x) \rangle$ of the nonequilibrium simulations is seen to significantly overestimate the free energy at large distances, the dissipation correction W_{diss} in Eq. (2) is obviously of importance. Fig. 1b shows the underlying friction profile $\Gamma(x)$ obtained from dcTMD, which in part deviates from the lineshape of the free energy. While we also find a maximum at $x \approx 0.4$ nm, the behavior of $\Gamma(x)$ is remarkably different for larger distances $0.5 \lesssim x \lesssim 0.7$ nm, where a region of elevated friction can be found before dropping to lower values. Interestingly, these features of $\Gamma(x)$ match well the changes of the average number of water molecules bridging both ions³⁴. This indicates that the increased friction in Eq. (3) is mainly caused by force fluctuations associated with the build-up of a hydration shell²². For $x \gtrsim 0.8$ nm, the friction is constant within our signal-to-noise resolution. The dynamics of ion dissociation and association can be described by their mean waiting times and corresponding rates shown in Fig. 2a and Table 1. For the chosen force field, ion concentration and resulting effective simulation box size, the unbiased MD simulation at 293 K yields mean dissociation and association times of $\tau_D = 1/k_D = 120$ ps and $\tau_A = 1/(k_A C) = 850$ ps, respectively, where C denotes a reference concentration (see the Supplementary Methods for details). Using fields $\Delta G(x)$ and $\Gamma(x)$ obtained from TMD, the numerical integration of Langevin Eq. (1) for 1 μ s results in $\tau_D = 420$ ps and $\tau_A = 3040$ ps. While the dissociation constants $K_D = k_D/k_A = 1.5$ M from Langevin and MD simulations match perfectly, we find that the Langevin predictions overestimate the correct rates by a factor of ~ 3.4 .

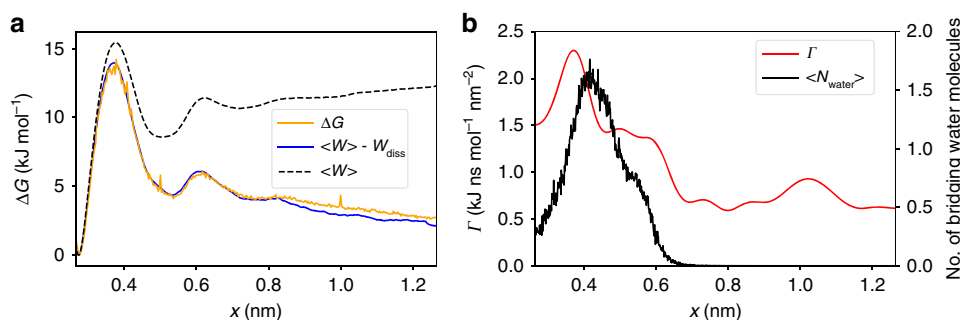


Fig. 1 Dissociation of NaCl in water. **a** Free energy profiles $\Delta G(x)$ along the interionic distance x , obtained from a 1 μ s long unbiased MD trajectory at 293 K (orange line) and 1000×1 ns TMD runs (blue line). Error bars are given in Supplementary Fig. 2. Also shown is the average work $\langle W(x) \rangle$ calculated from the TMD simulations (dashed black line). **b** Friction profile $\Gamma(x)$ (red) obtained from dcTMD after Gaussian smoothing together with the average number of water molecules (black), that connect the Na^+ and Cl^- ions in a common hydration shell³⁴.

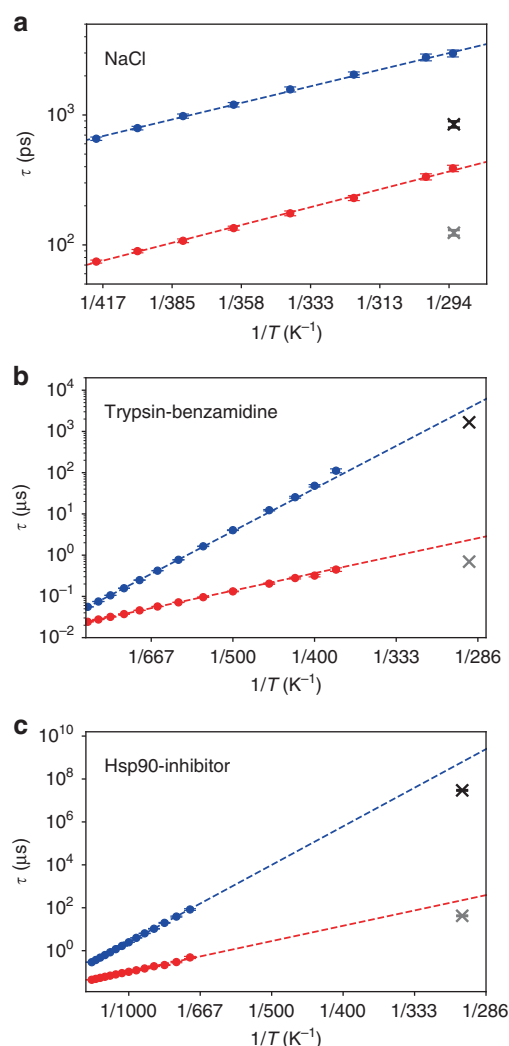


Fig. 2 Prediction of binding and unbinding times. Mean binding (red) and unbinding (blue) times, drawn as a function of the inverse temperature, obtained from *T*-boosted Langevin simulations of **a** solvated NaCl, **b** the trypsin-benzamidine complex, and **c** the Hsp90-inhibitor complex. Dashed lines represent fits ($R^2 = 0.90\text{--}0.99$) to Eq. (4), crosses (binding in grey, unbinding in black) indicate reference results from **a** unbiased MD simulation²² and **b, c** experiment^{39,48}.

Table 1 Predicted binding and unbinding kinetics.

NaCl	LE	MD ²²
k_A ($10^9 \text{ s}^{-1} \text{ M}^{-1}$)	1.6 ± 0.1	5.5 ± 0.3
k_D (10^9 s^{-1})	2.4 ± 0.1	8.1 ± 0.4
K_D (M)	1.5 ± 0.2	1.5 ± 0.2
Trypsin	LE	Experiment ³⁹
k_{on} ($10^6 \text{ s}^{-1} \text{ M}^{-1}$)	8.7 ± 0.5	29.0
k_{off} (10^2 s^{-1})	2.7 ± 0.4	6.0
K_D (10^{-5} M)	3.1 ± 0.6	2.1
Hsp90	LE	Experiment ⁴⁸
k_{on} ($10^4 \text{ s}^{-1} \text{ M}^{-1}$)	9.0 ± 0.8	48.0 ± 2.0
k_{off} (10^{-3} s^{-1})	1.6 ± 0.2	34.0 ± 2.0
K_D (10^{-8} M)	1.8 ± 0.3	7.1 ± 0.5

Rates resulting from fits in Fig. 2 (using units of molarity M, i.e., mol/l) with fit errors as indicated⁵⁹, and reference values from unbiased MD simulations²² and experiment^{39,48}, respectively. Dissociation constants were calculated from rate constants.

The latter may be caused by various issues. For one, to be of practical use, the Langevin model was deliberately kept quite simple. For example, it does not include an explicit solvent coordinate^{34,35}, but accounts for the complex dynamics of the solvent merely through the friction field $\Gamma(x)$. Moreover, we note that the calculation of $\Gamma(x)$ via Eq. (3) uses constraints, which have the effect of increasing the effective friction³⁶. This finding is supported by calculations using the data-driven Langevin approach^{37,38}, which estimates friction coefficients based on unbiased MD simulations that are consistently smaller than the ones obtained from dcTMD (Supplementary Fig. 5). Considering the simplicity of the Langevin model and the approximate calculation of the friction coefficient by dcTMD, overall we are content with a factor ~ 3 deviation of the predicted kinetics.

To illustrate the validity of the *T*-boosting approach suggested above, we performed a series of Langevin simulations for eight temperatures ranging from 290 to 420 K and plotted the resulting dissociation and association times as a function of the inverse temperature (Fig. 2a and Table 1). Checking the consistency of our approach, a fit to Eq. (4) yields transition state free energies ΔG^\ddagger of 13 and 12 kJ/mol for ion dissociation and association, respectively, which agree well with barrier heights of the free energy profile in Fig. 1a. Moreover, dissociation and association times obtained from the extrapolated *T*-boosted Langevin simulations ($\tau_D = 370$ ps, $\tau_A = 3050$ ps) agree excellently with the directly calculated values. This indicates that high-temperature Langevin simulations can indeed be extrapolated to obtain low-temperature transition rates.

Trypsin-benzamidine. Let us now consider the prediction of free energies, friction profiles and kinetics in protein–ligand systems. The first system we focus on is the inhibitor benzamidine bound to trypsin^{39–41}, which represents a well-established model problem to test enhanced sampling techniques^{21,31,42–45}. The slowest dynamics in this system is found in the unbinding process, which occurs on a scale of milliseconds³⁹. To capture the kinetics of the unbinding process, so far Markov state models^{42,43}, metadynamics³¹, Brownian dynamics⁴⁴ and adaptive enhanced sampling methods^{21,45} have been employed.

Here we combined dcTMD simulations and a subsequent nonequilibrium principal component analysis³⁰ to identify the dominant dissociation pathways of ligands during unbinding from their host proteins (see Supplementary Methods). Fig. 3 shows TMD snapshots of the structural evolution along this pathway, its free energy profile $\Delta G(x)$, and the associated friction $\Gamma(x)$. Starting from the bound state ($x_1 = 0$ nm), $\Delta G(x)$ exhibits a single maximum at $x_2 \approx 0.46$ nm, before it reaches the dissociated state for $x \gtrsim x_4 = 0.75$ nm. In line with the findings of Tiwary et al.³¹, the maximum of $\Delta G(x)$ reflects the rupture of the Asp189-benzamidine salt bridge, which represents the most important contact of the bound ligand. Following right after, the friction profile $\Gamma(x)$ reaches its maximum at $x_3 \approx 0.54$ nm, where the charged side chain of benzamidine becomes hydrated with water molecules. Similarly to NaCl, the friction peak coincides with the increase in the average number of hydrogen bonds between benzamidine and bulk water. The peak in friction is slightly shifted to higher x , because the ligand acts as a plug for the binding site, and first needs to be (at least partially) removed in order to allow water flowing in. As for the dissociation of NaCl in water, enhanced friction during unbinding appears to be directly linked to a rearrangement of the protein–ligand hydration shell, which is in agreement with recent results from neutron crystallography⁴¹.

To calculate rates k_{on} and k_{off} describing the binding and unbinding of benzamidine from trypsin, we performed 10 ms long

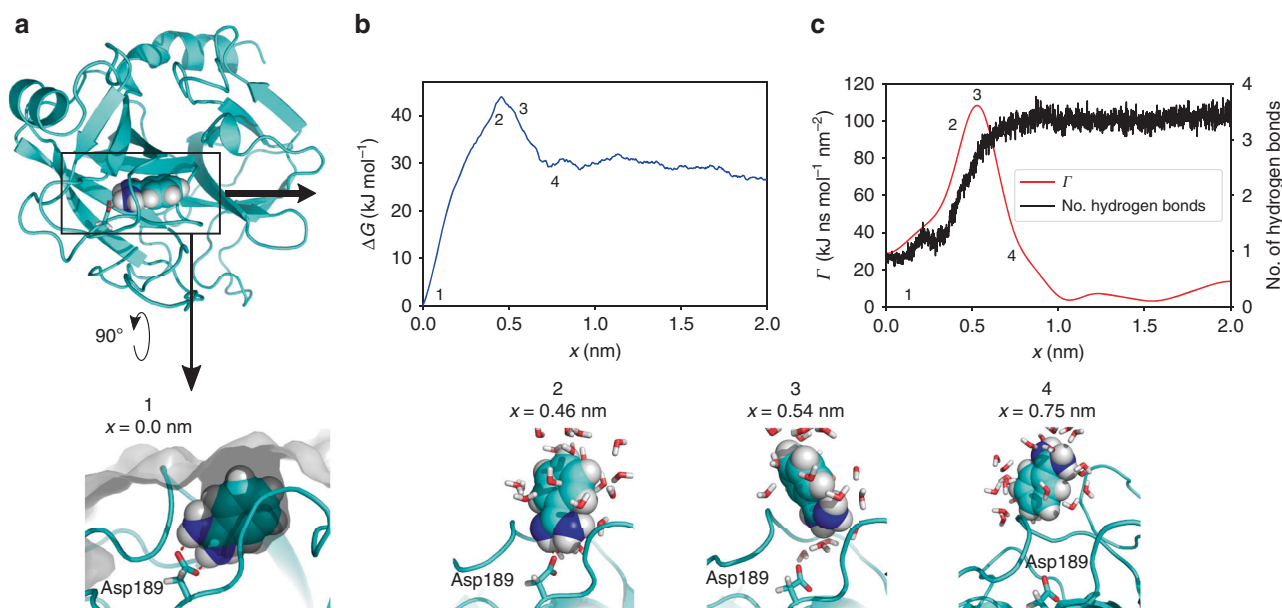


Fig. 3 Unbinding of benzamidine from trypsin. **a** TMD snapshots of the structural evolution in trypsin along the dominant dissociation pathway, showing protein surface in gray, benzamidine as van der Waals spheres, Asp189 and water molecules as sticks. Benzamidine is bound to the protein in a cleft of the protein surface via a bidentate salt bridge to Asp189. dcTMD calculations of **b** free energy $\Delta G(x)$, and **c** (Gaussian smoothed) friction $\Gamma(x)$ together with the mean number of hydrogen bonds between benzamidine and water. Highlighted are the bound state 1, transition state 2, the state with maximal friction 3 and the unbound state 4. Error bars of free energy and friction estimates are given in Supplementary Fig. 2.

Langevin simulations along the dominant pathways at thirteen temperatures ranging from 380–900 K. As shown in Fig. 2b and Table 1, the resulting rates are well fitted ($R^2 \geq 0.90$) by the T -boosting expression in Eq. (4). Representing the resulting number of transitions as a function of the inverse temperature, we find that at 380 K only ~ 9 events happen during a millisecond. That is, to obtain statistically converged rates at 290 K would require Langevin simulations at 290 K on a timescale of seconds. Using temperature boosting with Eq. (4), on the other hand, our high-temperature millisecond Langevin simulations readily yield converged transition rates at 290 K (see Fig. 2b and Table 1), that is, $k_{\text{on}} = 8.7 \times 10^6 \text{ s}^{-1} \text{ M}^{-1}$ and $k_{\text{off}} = 2.7 \times 10^2 \text{ s}^{-1}$, which underestimate the experimental values³⁹ $k_{\text{on}} = 2.9 \times 10^7 \text{ s}^{-1} \text{ M}^{-1}$ and $k_{\text{off}} = 6.0 \times 10^2 \text{ s}^{-1}$ by a factor of 2–3. Similarly, the calculated K_D overestimates the experimental result³⁹ of $K_D = 2.1 \times 10^{-5} \text{ M}$ by a factor of ~ 1.5 . As indicated by a recent review³ comparing numerous computational methods to calculate (un)binding rates of trypsin-benzamidine, our approach compares quite favorably regarding accuracy and computational effort.

As the extrapolation error due to T -boosting is negligible (see Supplementary Methods), the observed error is mainly caused by the approximate calculation of free energy and friction fields by dcTMD. In the case of NaCl, we have shown that reliable estimates of the fields (with errors $\lesssim 1 k_B T$) require an ensemble of at least 500 simulations (see ref. 22 and Supplementary Fig. 2), although the means of ΔG and Γ appear to converge already for ~ 100 trajectories. In a similar vein, by performing a Jackknife “leave-one-out” analysis⁴⁶, for trypsin-benzamidine we obtain an error of $\sim 2 k_B T$ for 150 trajectories (Supplementary Fig. 2). Interestingly, the error of the main free energy barrier is typically comparatively small, because the friction and thus variance of W increase directly after the barrier. As a consequence, the sampling error of k_{off} is small compared to that of k_{on} and the binding free energy. We note that if the experimental binding affinity K_D is known, it can be used as a further constraint on the error of the free energy and friction fields.

Hsp90-inhibitor. The second investigated protein complex is the N-terminal domain of heat shock protein 90 (Hsp90) bound to a resorcinol scaffold-based inhibitor (**1j** in ref. 47). This protein has recently been established as a test system for investigating the molecular effects influencing binding kinetics^{47–50}, and the selected inhibitor unbinds on a scale of half a minute. From the overall appearance of free energy and friction profiles (Fig. 4), we observe clear similarities to the case of trypsin-benzamidine. That is, the main transition barrier is also found at $x_2 \approx 0.5 \text{ nm}$, which stems from the ligand pushing between two helices at this point in order to escape the binding site. Moreover, the friction peaks at $x_2 \approx 0.5 \text{ nm}$, as well, but with an additional shoulder at $x_3 \approx 0.8 \text{ nm}$, which again coincides with changes of the ligand’s hydration shell. The unbound state is reached after $x \gtrsim 1.0 \text{ nm}$. We note that the ligand is again bound to the protein via a hydrogen bond to an aspartate (Asp93) and at a position that is open to the bulk water.

To calculate rates k_{on} and k_{off} , we again performed 5 ms long Langevin simulations along the dissociation pathway at fourteen different temperatures ranging from 700–1350 K. Rate prediction (see Fig. 2c and Table 1) yields $k_{\text{on}} = 9.0 \times 10^4 \text{ s}^{-1} \text{ M}^{-1}$ and $k_{\text{off}} = 1.6 \times 10^{-3} \text{ s}^{-1}$, and underestimates the experimental⁴⁸ values $k_{\text{on}} = 4.8 \pm 0.2 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$ and $k_{\text{off}} = 3.4 \pm 0.2 \times 10^{-2} \text{ s}^{-1}$ by a factor of 5–20. The resulting value for $K_D = 1.8 \times 10^{-8} \text{ M}$ underestimates the experimental value⁴⁸ $7.1 \times 10^{-8} \text{ M}$ by a factor of ~ 4 . Considering that we attempt to predict unbinding times on a time scale of half a minute from sub- μs MD simulations, and that a factor 20 corresponds to a free energy difference of about $3 k_B T$ (i.e., 15 % of the barrier height in Hsp90), we find this agreement remarkable for a first principles approach which implies many uncertainties of the physical model⁵¹. We attribute the larger deviation in comparison to trypsin to issues with the sampling of the correct unbinding pathways: especially unbinding rates in the range of minutes to hours fall into the same timescale as slow conformational dynamics of host proteins⁴⁸, requiring a sufficient sampling of the conformational space of the protein as a prerequisite for dcTMD pulling simulations.

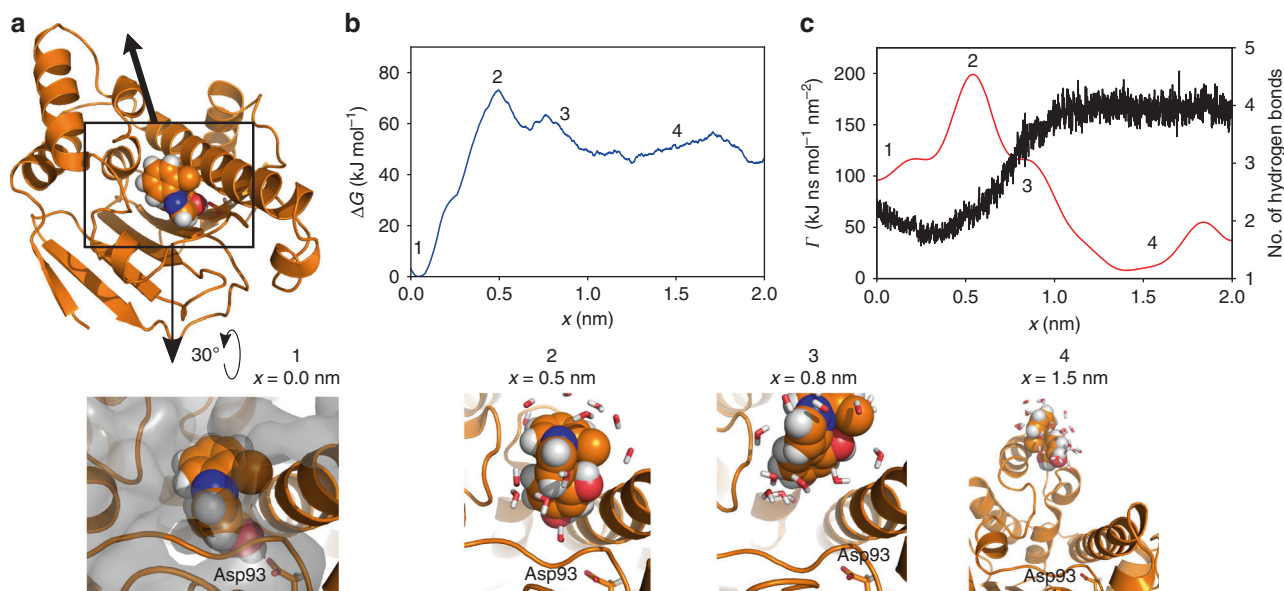


Fig. 4 Unbinding of an inhibitor from the N-terminal domain of Hsp90. **a** Structural evolution along the dissociation pathway in Hsp90, showing protein surface in gray, inhibitor as van der Waals spheres, Asp93 and water molecules as sticks. The inhibitor is bound to the protein in a cleft of the protein surface via a hydrogen bond to Asp93. dcTMD calculations of **b** free energy $\Delta G(x)$ and **c** (Gaussian-smoothed) friction $\Gamma(x)$ together with the mean number of hydrogen bonds between inhibitor and water. Highlighted are the bound state 1, transition state and state with maximal friction 2, an additional state with increased friction 3 and the unbound state 4. Error bars of free energy and friction estimates are given in Supplementary Fig. 2. Fluctuations of $\Gamma(x)$ for $x \geq 1$ nm are due to noise. Color code as in Fig. 3.

Discussion

Using free energy and friction profiles obtained from dcTMD, we have shown that T -boosted Langevin simulations yield binding and unbinding rates which are well comparable to results from atomistic equilibrium MD and experiments. That is, rates are underestimated by an order of magnitude or less which, in comparison to other methods that have been applied to the trypsin-benzamidine and Hsp90 complexes (see refs. 3,52 for recent reviews), is within the top accuracy currently achievable. At the same time, the few other methods that aim at predicting absolute rates (such as Markov state models^{42,43} and infrequent metadynamics^{31,53}) require substantial more MD simulation time, while dcTMD only requires sub- μ s MD runs, that is, at least an order of magnitude less computational time. As the extrapolation error due to T -boosting is negligible, the error is mainly caused by the approximate calculation of free energy and friction fields by dcTMD. We have shown that friction profiles, which correspond to the dynamical aspect of ligand binding and unbinding, may yield additional insight into molecular mechanisms of unbinding processes, which are not reflected in the free energies. Although the three investigated molecular systems differ significantly, in all cases friction was found to be governed by the dynamics of solvation shells.

Methods

MD simulations. All simulations employed Gromacs v2018 (ref. 54) in a CPU/GPU hybrid implementation, using the Amber99SB* force field^{55,56} and the TIP3P water model⁵⁷. For each system, 10^2 – 10^3 dcTMD calculations²² at pulling velocity $v_c = 1$ m/s were performed to calculate free energy $\Delta G(x)$ and friction $\Gamma(x)$. For the NaCl-water system, dcTMD as well as unbiased MD simulations were taken from ref. 22. Trypsin-benzamidine complex simulations are based on the 1.7 Å X-ray crystal structure with PDB ID 3PTB⁴⁰. Simulation systems of the Hsp90-inhibitor complex were taken from ref. 47. Detailed information on system preparation, ligand parameterization, MD simulations and pathway separation can be found in the Supplementary Methods.

Langevin simulations. Langevin simulations employed the integration scheme by Bussi and Parrinello⁵⁸. Details on the performance of this method with respect to the employed integration time step and system mass can be found in the Supplementary Methods.

Data availability

Simulation data on NaCl, Trypsin-benzamidine, and Hsp90-inhibitor is available from the authors upon request.

Code availability

Python scripts for dcTMD calculations, the fastpca program package for nonequilibrium principal component analysis, the data-driven Langevin package, the Langevin simulation code, and Jupyter notebooks for T -boosting analysis and sampling error estimation in Langevin simulations are available at our website www.moldyn.uni-freiburg.de.

Received: 5 February 2020; Accepted: 12 May 2020;

Published online: 10 June 2020

References

- Berendsen, H. J. C. *Simulating the Physical World* (Cambridge University Press, Cambridge, 2007).
- Pan, A. C. et al. Atomic-level characterization of protein-protein association. *Proc. Natl Acad. Sci. USA* **116**, 4244–4249 (2019).
- Bruce, N. J., Ganotra, G. K., Kokh, D. B., Sadiq, S. K. & Wade, R. C. New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* **49**, 1–10 (2018).
- Rico, F., Russek, A., González, L., Grubmüller, H. & Scheuring, S. Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations. *Proc. Natl Acad. Sci. USA* **116**, 6594–6601 (2019).
- Copeland, R. A., Pompliano, D. L. & Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* **5**, 730–739 (2006).
- Swinney, D. C. Applications of Binding Kinetics to Drug Discovery. *Pharm. Med.* **22**, 23–34 (2012).

7. Pan, A. C., Borhani, D. W., Dror, R. O. & Shaw, D. E. Molecular determinants of drug–receptor binding kinetics. *Drug Discov. Today* **18**, 667–673 (2013).
8. Klebe, G. The use of thermodynamic and kinetic data in drug discovery: decisive insight or increasing the puzzlement? *ChemMedChem* **10**, 229–231 (2014).
9. Copeland, R. A. The drug–target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discov.* **15**, 87–95 (2016).
10. Chipot, C. & Pohorille, A. *Free Energy Calculations* (Springer, Berlin, 2007).
11. Christ, C. D., Mark, A. E. & van Gunsteren, W. F. Basic ingredients of free energy calculations: a review. *J. Comput. Chem.* **31**, 1569–1582 (2010).
12. Mitsutake, A., Sugita, Y. & Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **60**, 96 (2001).
13. Torrie, G. M. & Valleau, J. P. Non-physical sampling distributions in Monte-Carlo free-energy estimation—umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
14. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224–230 (2001).
15. Sprik, M. & Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **109**, 7737–7744 (1998).
16. Grubmüller, H. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E* **52**, 2893–2906 (1995).
17. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Comput. Mol. Sci.* **1**, 826–843 (2011).
18. Comer, J. et al. The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* **119**, 1129–1151 (2015).
19. Tiwary, P. & Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **111**, 230602 (2013).
20. Wu, H., Paul, F., Wehmeyer, C. & Noé, F. Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl Acad. Sci. USA* **113**, E3221–E3230 (2016).
21. Teo, I., Mayne, C. G., Schulten, K. & Lelievre, T. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *J. Chem. Theory Comput.* **12**, 2983–2989 (2016).
22. Wolf, S. & Stock, G. Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction. *J. Chem. Theory Comput.* **14**, 6175–6182 (2018).
23. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
24. Schlitter, J., Engels, M. & Krüger, P. Targeted molecular dynamics - a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* **12**, 84–89 (1994).
25. Straub, J. E., Borkovec, M. & Berne, B. J. Calculation of dynamic friction on intramolecular degrees of freedom. *J. Phys. Chem.* **91**, 4995–4998 (1987).
26. Hummer, G. Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J. Phys.* **7**, 34 (2005).
27. Vaikuntanathan, S. & Jarzynski, C. Escorted free energy simulations: Improving convergence by reducing dissipation. *Phys. Rev. Lett.* **100**, 190601 (2008).
28. Servantie, J. & Gaspard, P. Methods of calculation of a friction coefficient: application to nanotubes. *Phys. Rev. Lett.* **91**, 185503 (2003).
29. Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251 (1990).
30. Post, M., Wolf, S. & Stock, G. Principal component analysis of nonequilibrium molecular dynamics simulations. *J. Chem. Phys.* **150**, 204110 (2019).
31. Tiwary, P., Limongelli, V., Salvalaglio, M. & Parrinello, M. Kinetics of protein–ligand unbinding: predicting pathways, rates, and rate-limiting steps. *Proc. Natl Acad. Sci. USA* **112**, E386–E391 (2015).
32. Bryant, D. & Moulton, V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265 (2004).
33. Sørensen, M. R. & Voter, A. F. Temperature-accelerated dynamics for simulation of infrequent events. *J. Comp. Phys.* **112**, 9599–9606 (2000).
34. Mullen, R. G., Shea, J.-E. & Peters, B. Transmission coefficients, committers, and solvent coordinates in ion-pair dissociation. *J. Chem. Theory Comput.* **10**, 659–667 (2014).
35. Geissler, P. L., Dellago, C. & Chandler, D. Kinetic pathways of ion pair dissociation in water. *J. Phys. Chem. B* **103**, 3706–3710 (1999).
36. Daldrop, J. O., Kowalik, B. G. & Netz, R. R. External potential modifies friction of molecular solutes in water. *Phys. Rev. X* **7**, 041065 (2017).
37. Hegger, R. & Stock, G. Multidimensional Langevin modeling of biomolecular dynamics. *J. Chem. Phys.* **130**, 034106 (2009).
38. Schaudinnus, N., Lickert, B., Biswas, M. & Stock, G. Global Langevin model of multidimensional biomolecular dynamics. *J. Chem. Phys.* **145**, 184114 (2016).
39. Guillain, F. & Thusius, D. Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J. Am. Chem. Soc.* **92**, 5534–5536 (1970).
40. Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. B* **39**, 480–490 (1983).
41. Schiebel, J. et al. Intriguing role of water in protein–ligand binding studied by neutron crystallography on trypsin complexes. *Nat. Commun.* **9**, 166 (2018).
42. Buch, I., Giorgino, T. & De Fabritiis, G. Complete reconstruction of an enzyme–inhibitor binding process by molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **108**, 10184–10189 (2011).
43. Plattner, N. & Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6**, 7653 (2015).
44. Votapka, L. W., Jagger, B. R., Heyneman, A. & Amaro, R. E. SEEKR: simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin–benzamidine binding. *J. Phys. Chem. B* **121**, 3597–3606 (2017).
45. Betz, R. M. & Dror, R. O. How effectively can adaptive sampling methods capture spontaneous ligand binding? *J. Chem. Theory Comput.* **15**, 2053–2063 (2019).
46. Efron, B. & Stein, C. The Jackknife estimate of variance. *Ann. Stat.* **9**, 586–596 (1981).
47. Wolf, S. et al. Estimation of protein–ligand unbinding kinetics using non-equilibrium targeted molecular dynamics simulations. *J. Chem. Inf. Model.* **59**, 5135–5147 (2019).
48. Amaral, M. et al. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat. Commun.* **8**, 2276 (2017).
49. Kokh, D. B. et al. Estimation of drug–target residence times by τ -random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* **14**, 3859–3869 (2018).
50. Schuetz, D. A. et al. Predicting residence time and drug unbinding pathway through scaled molecular dynamics. *J. Chem. Inf. Model.* **59**, 535–549 (2019).
51. Capelli, R. et al. On the accuracy of molecular simulation-based predictions of koff values: a Metadynamics study. Preprint at <https://www.biorxiv.org/content/10.1101/2020.03.30.015396v1> (2020).
52. Nunes-Alves, A., Kokh, D. B. & Wade, R. C. Recent progress in molecular simulation methods for drug binding kinetics. Preprint at <https://arxiv.org/abs/2002.08983> (2020).
53. Casanovas, R., Limongelli, V., Tiwary, P., Carloni, P. & Parrinello, M. Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* **139**, 4780–4788 (2017).
54. Abraham, M. J. et al. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
55. Hornak, V. et al. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).
56. Best, R. B. & Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* **113**, 9004–9015 (2009).
57. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
58. Bussi, G. & Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **75**, 2289–22897 (2007).
59. Hughes, I. & Hase, T. *Measurements and Their Uncertainties. A Practical Guide to Modern Error Analysis* (Oxford University Press, 2010).

Acknowledgements

We thank Peter Hamm and Matthias Post for numerous instructive and helpful discussions. The authors acknowledge support by the Deutsche Forschungsgemeinschaft (Sto 247/11), by the bwUniCluster computing initiative, the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the Deutsche Forschungsgemeinschaft through grant No. INST 37/935-1 FUGG, and the Freiburg Institute for Advanced Studies (FRIAS) of the Albert-Ludwigs-University Freiburg. The article processing charge was partially funded by the Albert-Ludwigs-University Freiburg in the funding programme Open Access Publishing.

Author contributions

S.W. and G.S. designed and supervised research. S.W. performed TMD and Langevin simulations and nonequilibrium path separation of Trypsin trajectories. B.L. performed dLE analysis and implemented Langevin simulations. S.B. performed the nonequilibrium path separation of Hsp90 trajectories. All authors wrote the paper.

Competing Interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16655-1>.

Correspondence and requests for materials should be addressed to S.W. or G.S.

Peer review information *Nature Communications* thanks Max Bonomi, Pratyush Tiwary and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020