

Multisecond ligand dissociation dynamics from atomistic simulations

Steffen Wolf,^{1, a)} Benjamin Lickert,¹ Simon Bray,^{1, 2} and Gerhard Stock^{1, b)}

¹⁾*Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, 79104 Freiburg, Germany*

²⁾*Present address: Bioinformatics Group, Department of Computer Science, Albert Ludwigs University, 79110 Freiburg, Germany*

(Dated: 15 January 2020)

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 necessary for coarse-graining, we first perform dissipation-corrected targeted molecular dynamics simulations, which yield free energy and friction profiles of the molecular process under consideration. In a second step, we use these fields to perform Langevin simulations which account for the desired molecular kinetics. By introducing the concept of 'temperature boosting' of the Langevin simulations, this combination of methods allows simulation of biomolecular processes occurring on multisecond timescales and beyond. Adopting the dissociation of solvated sodium chloride as well as trypsin-benzamidine and Hsp90-inhibitor protein-ligand complexes as test problems, we are able to reproduce the rates from atomistic molecular dynamics simulation and experiments within a factor of 1.5–4 for unbinding times up to the range of milliseconds and of 1.2–10 in the range of seconds. Analysis of the friction profiles reveals that binding and unbinding dynamics are mediated by changes of the surrounding hydration shells in all investigated systems.

I. INTRODUCTION

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 better drug efficacies^{5–7}. Since these processes typically occur on timescales 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^{8–16} 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 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^{17–19}. 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 constraint²⁰. By combining a Langevin equation analysis with a cumulant expansion of Jarzynski's equality²¹, dcTMD yields both $\Delta G(x)$ and the friction field $\Gamma(x)$. The latter reflects interactions of

the system with degrees of freedom orthogonal to those which define the free energy. 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 –without further approximations– to speed up the calculations by several orders of magnitude.

II. THEORY

A. Dissipation-corrected targeted molecular dynamics

To set the stage, we briefly review the working equations of dcTMD derived in Ref. 20. TMD as developed by Schlitter²² 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)

^{a)}Electronic mail: steffen.wolf@physik.uni-freiburg.de

^{b)}Electronic mail: stock@physik.uni-freiburg.de

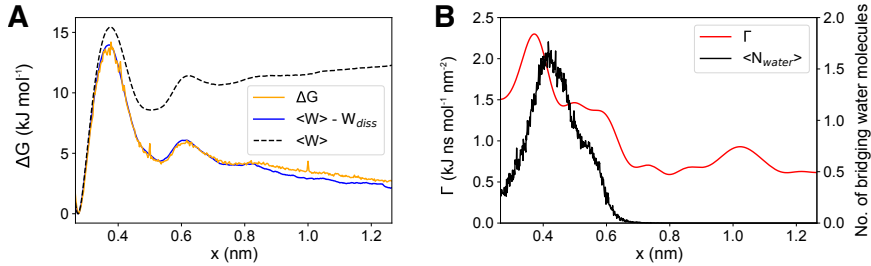


FIG. 1. Dissociation of NaCl in water. (A) Free energy profiles $\Delta G(x)$ along the interion distance x , obtained from a $1 \mu\text{s}$ long unbiased MD trajectory at 293 K (dashed green line) and $1000 \times 1 \text{ ns}$ TMD runs (full blue line). Also shown is the average work $\langle W(x) \rangle$ calculated from the TMD simulations (black). (B) Friction profile $\Gamma(x)$ obtained from TMD (red) together with the average number of water molecules (black), that connect the Na^+ and Cl^- ions in a common hydration shell.²⁸

over many TMD runs, we thus obtain the relation²⁰

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

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^{23,24}, it proves advantageous to calculate $\Gamma(x)$ directly from TMD simulations. To this end, we invoke Jarzynski's identity²¹, $e^{-\Delta G(x)/k_B T} = \langle e^{-W(x)/k_B T} \rangle$, the second-order cumulant expansion of 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 friction²⁰

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

which is readily evaluated directly from the TMD simulations.

As discussed in Ref. 20, the derivation of Langevin equation (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 change²⁵, which leaves the system virtually at equilibrium at all times. This means that the free energy (2) and the friction (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 equation (1) for $f_c = 0$. Numerical propagation of the unbiased Langevin equation then accounts for the coarse grained dynamics of the system, and yields the timescales of the considered process.

The theory developed above rests on two main assumptions. For one, we have assumed that Langevin equation (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, we have invoked a cumulant expansion to derive friction coefficient (3), which is valid under the assumption that the distribution of W within the ensemble is Gaussian. 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 an ensemble of dcTMD trajectories into classes, each corresponding to a pathway²⁶. To identify these pathways, a nonequilibrium principal component analysis of protein-ligand distances may be performed²⁶. Alternatively, path separation can be based on geometric distances between individual trajectories, making use of the NeighborNet algorithm²⁷ to cluster trajectories into classes (and thus pathways) according to these distances.

B. T -boosting

The speed-up of Langevin equation (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 Table S1), however, we still need to propagate Eq. (1) for $\gtrsim 100 \cdot 10^{15}$ 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 transition state expression

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

where ΔG^\ddagger denotes the transition state energy and $\beta_i = 1/k_B T_i$ is the inverse temperature. That is, by in-

creasing 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 .

It should be noted that the above described procedure, henceforth termed “ T -boosting,” involves no assumptions or 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. On the other hand, we wish to stress 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 , because the fields $\Delta G(x)$ and $\Gamma(x)$ obtained from MD do depend on temperature.

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 Information 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 hours 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.

III. RESULTS AND DISCUSSION

A. 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 available²⁰, making it a suitable benchmark system for our approach.

To start with, 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, i.e., the loss of di-

rect van der Waals contact and the formation of a common hydration shell. The second smaller maximum at $x \approx 0.6$ nm reflects the transition from a common to two separate hydration shells²⁸. 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. Figure 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 at $x \gtrsim 0.8$ nm. Interestingly, these features of $\Gamma(x)$ match well the changes of the average number of water molecules bridging both ions²⁸. This indicates that the increased friction in Eq. (3) is mainly caused by force fluctuations associated with the build-up of a hydration shell²⁰.

The dynamics of ion dissociation and association can be described by their mean waiting times (or the respective rates) shown in Figure 2A. For the chosen ion concentration and resulting effective simulation box size (see SI), the unbiased MD simulation at 293 K yields mean dissociation and association times of $\tau_D = 1/k_D = 120$ ps and $\tau_A = 1/k_A = 850$ ps, respectively. Using fields $\Delta G(x)$ and $\Gamma(x)$ obtained from TMD, the numerical integration of Langevin equation (1) for 1 μ s results in $\tau_D = 420$ ps and $\tau_A = 3040$ ps. That is, the Langevin predictions overestimate the true values by a factor of about four, which may be caused by various issues. For one, to be of practical use, the Langevin model was deliberately kept quite simple. It includes neither an explicit solvent coordinate to account for the complex dynamics of the solvent²⁸, nor does it account for non-Markovian memory effects as in a generalized Langevin equation¹. Moreover, we note that the calculation of the friction $\Gamma(x)$ via Eq. (3) uses constraints, which move with a pulling velocity v_c that is small compared to the transition timescale. Unlike unbiased simulations, where ions and water molecules move with comparable velocities, in TMD simulations the ions are therefore artificially held in place. This causes additional interactions with water molecules, which have the effect of increasing the effective friction³¹. This finding is supported by calculations using the data-driven Langevin approach³², which estimates friction coefficients based on unbiased MD simulations that are smaller than the ones obtained from TMD for $x \lesssim 0.7$ nm (Fig. S1). Considering the simplicity of the Langevin model and the approximate calculation of the friction coefficient by TMD, overall we are content with a factor 4 deviation of the predicted kinetics.

Waiting times on a nanosecond timescale as found for solvated NaCl are readily sampled by microsecond long Langevin trajectories at room temperature. Nonetheless,

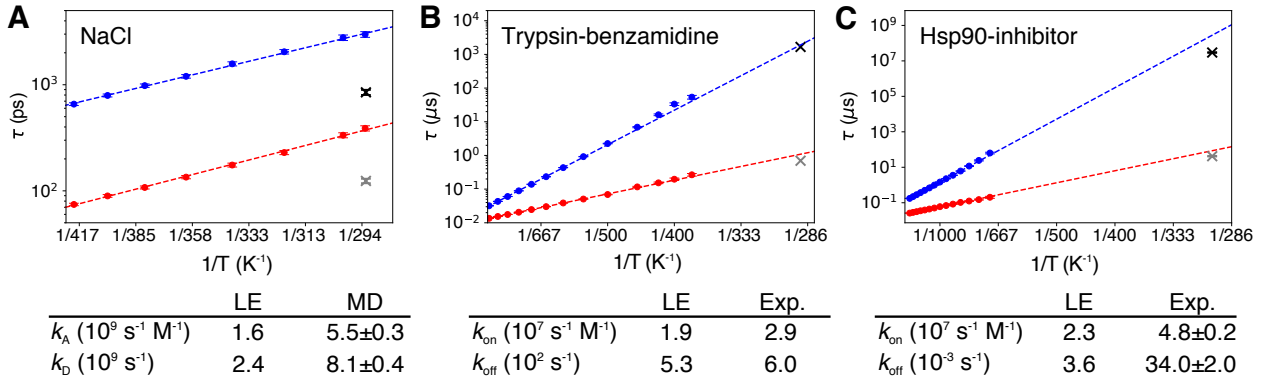


FIG. 2. Mean binding (red) and unbinding (blue) times, drawn as a function of the inverse temperature, obtained from T -boosted Langevin simulations of (A) solvated NaCl, (B) the trypsin-benzamidine complex and (C) the Hsp90 inhibitor complex. Dashed lines represent fits ($R^2 = 0.91 - 0.99$) to Eq. (4), crosses (binding in grey, unbinding in black) indicate reference results from (A) unbiased MD simulation²⁰ and (B), (C) experiment^{29,30}. Bars represent the standard error of the mean. Tables below comprise corresponding rates with M being the molarity, i.e., mol/l, and reference values. For trypsin-benzamidine and Hsp90, rate constants were fitted according to Eq. (4) at 290 K and 300 K, respectively.

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. 2). A fit to Eq. (4) yields transition state free energies ΔG^\ddagger of 13 kJ/mol and 12 kJ/mol for ion dissociation and association, respectively, which agree well with barrier heights of the free energy profile in Fig. 1A extracted from both biased and unbiased simulations. 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 (see Table S1). This indicates that high-temperature Langevin simulations can indeed be extrapolated to obtain low-temperature transition rates.

B. Trypsin-benzamidine

Let us now consider the prediction of free energies, friction profiles and kinetics in protein-ligand systems. With binding and unbinding times on a range between milliseconds and hours, these systems cannot generally be studied by sufficient sampling using unbiased simulations, owing to the limited capabilities of current computational hardware. The first system we focus on is the inhibitor benzamidine bound to trypsin^{29,33,34}, which represents a well-established model problem to test enhanced sampling techniques^{19,35-40}. The slowest dynamics in this system is found in the unbinding process, which occurs on a scale of milliseconds²⁹. To capture the kinetics of the unbinding process, so far Markov state models^{35,36}, metadynamics³⁷, Brownian dynamics³⁸ and adaptive enhanced sampling methods^{19,39,40} have been employed.

Here we combined dcTMD simulations and a subsequent nonequilibrium principal component analysis²⁶ to

identify the dominant dissociation pathways of ligands during unbinding from their host proteins (see SI Methods). Figure 3 shows TMD snapshots of the structural evolution along this pathway of benzamidine, 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. The calculated binding free energy of ~ 27 kJ/mol compares well to the experimental value of 28.0 ± 0.2 kJ/mol²⁹. In line with the findings of Tiwary et al.³⁷, 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. Being a ligand-binding system, however, 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 host hydration shell.

To calculate rates k_{on} and k_{off} describing the binding and unbinding of benzamidine from trypsin, we performed 5 ms long Langevin simulations along the dominant pathways at thirteen temperatures per system ranging from 380-900 K. As shown in Fig. 2B, the resulting rates are well fitted ($R^2 \geq 0.91$) by T -boosting expression (4). For trypsin-benzamidine, the fit yields energy barriers $\Delta G_{\text{off}}^\ddagger = 39$ kJ/mol and $\Delta G_{\text{on}}^\ddagger = 15$ kJ/mol, which agree well with the corresponding values ~ 44 and 17 kJ/mol obtained from Fig. 3B. Representing the resulting number of transitions as a function of the inverse

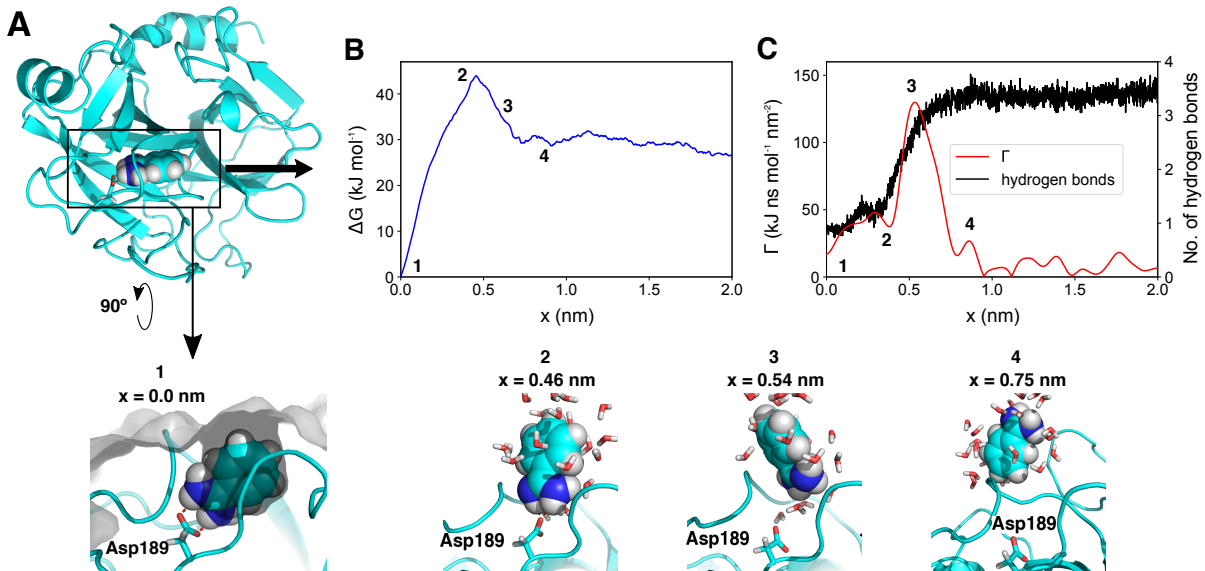


FIG. 3. Unbinding of benzamidine from trypsin. (A) TMD snapshots of the structural evolution in trypsin along the dominant dissociation pathway, showing protein surface in gray, benzamidine as van der Waals spheres, Asp189 and water molecules as sticks. Benzamidine is bound to the protein in a cleft of the protein surface via a bidental salt bridge to Asp189. dcTMD calculations of (B) free energy $\Delta G(x)$ and (C) friction $\Gamma(x)$ together with the mean number of hydrogen bonds between benzamidine and water.

temperature, we find that at 380 K only ~ 13 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 Figure 2), that is, $k_{\text{on}} = 1.9 \cdot 10^7 \text{ s}^{-1}\text{M}^{-1}$ and $k_{\text{off}} = 5.3 \cdot 10^2 \text{ s}^{-1}$, which underestimate the experimental values²⁹ $k_{\text{on}} = 2.9 \cdot 10^7 \text{ s}^{-1}\text{M}^{-1}$ and $k_{\text{off}} = 6.0 \cdot 10^2 \text{ s}^{-1}$ by a factor of 1.2–1.5. We note that the experimental value of k_{off} is given as $6.0 \pm 3.0 \cdot 10^2 \text{ M}^{-1}\text{s}^{-1}$ in some publications^{19,37}. While we could not find the source of this error estimate, it would imply a direct match of our calculations and the experiment.

As the extrapolation error due to T -boosting is negligible (see SI), 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 $\lesssim 1 k_{\text{B}}T$) require an ensemble of at least 500 simulations²⁰, 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" analysis⁴¹, for trypsin-benzamidine, we obtain an error of $\sim 2 k_{\text{B}}T$ for 150 trajectories (Fig. S2). 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 $K_D = k_{\text{off}}/k_{\text{on}}$ is known, it can be used as a further constraint on the

error of the free energy and friction fields.

C. Hsp90-inhibitor

The second protein complex of interest is the N-terminal domain of heat shock protein 90 (Hsp90) bound to a resorcinol scaffold-based inhibitor (compound **1j** in Ref. 42). This protein has recently been established as a test system for investigating the molecular effects influencing binding kinetics^{30,42–44}, and the selected inhibitor unbinds on a scale of half a minute. From the overall appearance of free energy and friction profiles (Fig. 4), this protein-ligand complex exhibits clear similarities to the case of trypsin-benzamidine. That is, the main transition barrier is also found at $x_2 \approx 0.5$ 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 $x_2 \approx 0.6$ nm with a side maximum at $x_3 \approx 0.8$ nm, which again coincides with the establishment of a hydration sphere. 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. The free energy in the unbound state of ~ 45 kJ/mol compares well to the experimental value of 40.7 ± 0.2 kJ/mol³⁰.

To calculate rates k_{on} and k_{off} describing the binding and unbinding of the resorcinol inhibitor from Hsp90, we again performed 5 ms long Langevin simulations along the dissociation pathway at fourteen different temperatures ranging from 700–1350 K. Fits result in energy barriers of $\Delta G_{\text{off}}^{\ddagger} = 71$ kJ/mol and $\Delta G_{\text{on}}^{\ddagger} = 24$ kJ/mol. The

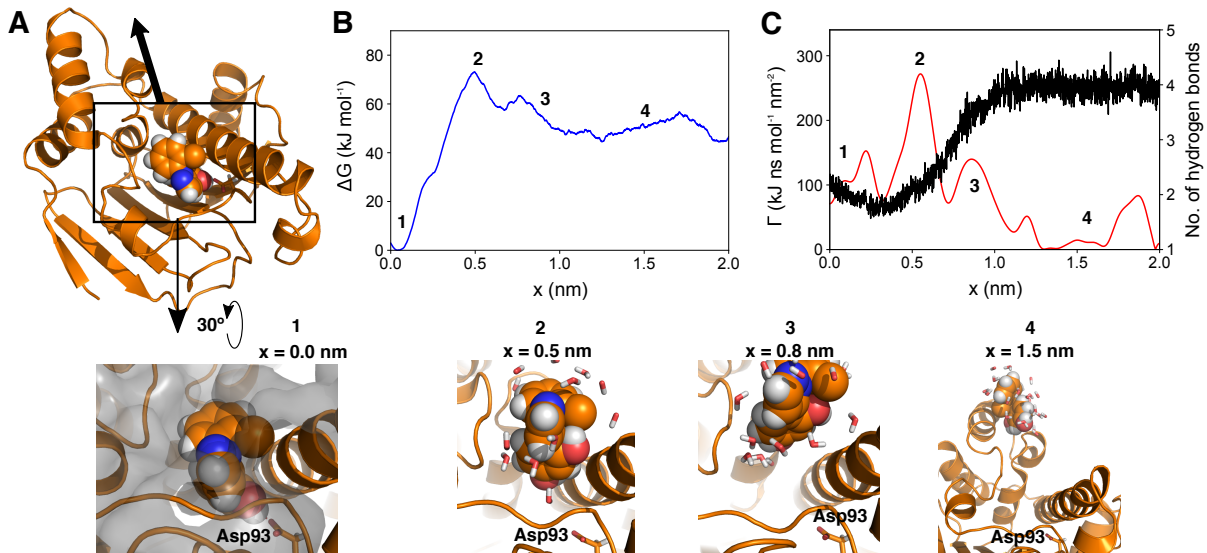


FIG. 4. Unbinding of an inhibitor from the N-terminal domain of Hsp90. (A) Structural evolution along the dissociation pathway in Hsp90, showing protein surface in gray, inhibitor as van der Waals spheres, Asp93 and water molecules as sticks. The inhibitor is bound to the protein in a cleft of the protein surface via a hydrogen bond to Asp93. dcTMD calculations of (B) free energy $\Delta G(x)$ and (C) friction $\Gamma(x)$ together with the mean number of hydrogen bonds between inhibitor and water. Color code as in Fig. 3.

agreement with the $\Delta G_{\text{off}}^{\ddagger} \approx 71$ kJ/mol and $\Delta G_{\text{on}}^{\ddagger} \approx 26$ kJ/mol from dcTMD simulation is good, as are the predictions at 300 K (see Fig. 2C), which yield $k_{\text{on}} = 2.3 \cdot 10^5$ s⁻¹M⁻¹ and $k_{\text{off}} = 3.6 \cdot 10^{-3}$ s⁻¹. The predictions again underestimate the experimental³⁰ values $k_{\text{on}} = 4.8 \pm 0.2 \cdot 10^5$ s⁻¹M⁻¹ and $k_{\text{off}} = 3.4 \pm 0.2 \cdot 10^{-2}$ s⁻¹ by a factor of 2–10, which is a fair agreement considering that we attempt to predict unbinding times on a time scale of half a minute from sub- μ s MD simulations. 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 proteins³⁰, requiring a sufficient sampling of the conformational space of the protein as a prerequisite for dcTMD pulling simulations.

IV. CONCLUSIONS

Using free energy and friction profiles from dcTMD, we have shown that T -boosted Langevin simulations yield binding and unbinding rates which are comparable to results from atomistic equilibrium MD and experiments. Adopting solvated sodium chloride, trypsin-benzamidine and Hsp90-inhibitor as test systems, the method underestimates the correct rates by a factor of 1.2–10. Considering that only sub- μ s MD runs and comparatively inexpensive Langevin simulations are required, this is comparable to the best accuracies currently achieved by enhanced sampling methods³. 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. In particular, the identification of the correct unbinding pathways of protein-ligand complexes can be challenging and requires further investigation. We have shown that friction profiles 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.

V. METHODS

Detailed information on system preparation, ligand parameterization, MD and Langevin simulations and pathway separation can be found in the SI.

A. MD simulations

All simulations employed Gromacs v2018 (Ref. 45) in a CPU/GPU hybrid implementation, using the Amber99SB force field^{46,47} and the TIP3P water model⁴⁸. For each system, 10^2 - 10^3 dcTMD calculations²⁰ 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. 20. Trypsin-benzamidine complex simulations are based on the 1.7 Å X-ray crystal structure with PDB ID 3PTB (Ref. 33). Simulation systems of the Hsp90-inhibitor complex were taken from Ref. 42.

B. Langevin simulations

Langevin simulations employed the integration scheme by Bussi and Parrinello⁴⁹, using a time step of 5 fs. As system mass m , the reduced mass of the NaCl dimer (13.88 g/mol), the trypsin-benzamidin (120.15 g/mol) and Hsp90-inhibitor (288.73 g/mol) complexes were used.

Supporting Information (SI)

Computational details on simulations, simulation system preparations, pathway separation and Langevin dynamics; waiting time distribution of NaCl dissociation and association; uncertainty for prediction of rates with T -boosting; friction estimate by data-driven Langevin equation; two Supporting Figures; one Supporting Table.

Data 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 a Jupyter notebook for Langevin simulation sampling error estimation are available at our website <https://www.moldyn.uni-freiburg.de/software/software.html>. Further data is available from the authors upon request.

Acknowledgements

We thank 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 German Research Foundation (DFG) through grant No. INST 37/935-1 FUGG.

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 nonequilibrium path separation of Hsp90 trajectories. All authors wrote the paper.

REFERENCES

- ¹H. J. C. Berendsen, *Simulating the Physical World*, Cambridge University Press, Cambridge, 2007.
- ²A. C. Pan, D. Jacobson, K. Yatsenko, D. Sritharan, T. M. Weirich, and D. E. Shaw, Atomic-level characterization of protein-protein association, *Proc. Natl. Acad. Sci. USA* **116**, 4244 (2019).
- ³N. J. Bruce, G. K. Ganotra, D. B. Kokh, S. K. Sadiq, and R. C. Wade, New approaches for computing ligand-receptor binding kinetics., *Curr. Opin. Struct. Biol.* **49**, 1 (2018).
- ⁴F. Rico, A. Russek, L. González, H. Grubmüller, and S. Scheuring, Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations, *Proc. Natl. Acad. Sci. USA* **116**, 6594 (2019).
- ⁵D. C. Swinney, Applications of Binding Kinetics to Drug Discovery, *Pharmaceut. Med.* **22**, 23 (2012).
- ⁶A. C. Pan, D. W. Borhani, R. O. Dror, and D. E. Shaw, Molecular determinants of drug-receptor binding kinetics, *Drug discovery today* **18**, 667 (2013).
- ⁷G. Klebe, The Use of Thermodynamic and Kinetic Data in Drug Discovery: Decisive Insight or Increasing the Puzzlement?, *ChemMedChem* **10**, 229 (2014).
- ⁸C. Chipot and A. Pohorille, *Free Energy Calculations*, Springer, Berlin, 2007.
- ⁹C. D. Christ, A. E. Mark, and W. F. van Gunsteren, Basic Ingredients of Free Energy Calculations: A Review, *J. Comput. Chem.* **31**, 1569 (2010).
- ¹⁰A. Mitsutake, Y. Sugita, and Y. Okamoto, Generalized-ensemble algorithms for molecular simulations of biopolymers, *Biopolymers* **60**, 96 (2001).
- ¹¹J. Kästner, Umbrella sampling, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 932 (2011).
- ¹²B. Isralewitz, M. Gao, and K. Schulten, Steered molecular dynamics and mechanical functions of proteins, *Curr. Opin. Struct. Biol.* **11**, 224 (2001).
- ¹³M. Sprik and G. Ciccotti, Free energy from constrained molecular dynamics, *J. Chem. Phys.* **109**, 7737 (1998).
- ¹⁴H. Grubmüller, Predicting slow structural transitions in macromolecular systems: Conformational flooding, *Phys. Rev. E* **52**, 2893 (1995).
- ¹⁵A. Barducci, M. Bonomi, and M. Parrinello, Metadynamics, *Comput. Mol. Sci.* **1**, 826 (2011).
- ¹⁶J. Comer, J. C. Gumbart, J. Hémin, T. Lelievre, A. Pohorille, and C. Chipot, The adaptive biasing force method: everything you always wanted to know but were afraid to ask., *J. Phys. Chem. B* **119**, 1129 (2015).
- ¹⁷P. Tiwary and M. Parrinello, From metadynamics to dynamics, *Phys. Rev. Lett.* **111**, 230602 (2013).
- ¹⁸H. Wu, F. Paul, C. Wehmeyer, and F. Noé, Multiensemble Markov models of molecular thermodynamics and kinetics, *Proc. Natl. Acad. Sci. USA* **113**, E3221 (2016).
- ¹⁹I. Teo, C. G. Mayne, K. Schulten, and T. Lelievre, Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time., *J. Chem. Theory Comput.* **12**, 2983 (2016).
- ²⁰S. Wolf and G. Stock, Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction, *J. Chem. Theory Comput.* **14**, 6175 (2018).
- ²¹C. Jarzynski, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.* **78**, 2690 (1997).
- ²²J. Schlitter, M. Engels, and P. Krüger, Targeted molecular dynamics - a new approach for searching pathways of conformational transitions, *J. Mol. Graph.* **12**, 84 (1994).
- ²³J. E. Straub, M. Borkovec, and B. J. Berne, Calculation of dynamic friction on intramolecular degrees of freedom, *J. Phys. Chem.* **91**, 4995 (1987).
- ²⁴G. Hummer, Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations, *New J. Phys.* **7**, 34 (2005).
- ²⁵J. Servantie and P. Gaspard, Methods of calculation of a friction coefficient: application to nanotubes, *Phys. Rev. Lett.* **91**, 185503 (2003).

- ²⁶M. Post, S. Wolf, and G. Stock, Principal component analysis of nonequilibrium molecular dynamics simulations, *J. Chem. Phys.* **150**, 204110 (2019).
- ²⁷D. Bryant and V. Moulton, Neighbor-net: an agglomerative method for the construction of phylogenetic networks., *Mol. Biol. Evol.* **21**, 255 (2004).
- ²⁸R. G. Mullen, J.-E. Shea, and B. Peters, Transmission Coefficients, Committors, and Solvent Coordinates in Ion-Pair Dissociation, *J. Chem. Theory Comput.* **10**, 659 (2014).
- ²⁹F. Guillain and D. Thusius, 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 (1970).
- ³⁰M. Amaral, D. B. Kokh, J. Bomke, A. Wegener, H. P. Buchstaller, H. M. Eggenweiler, P. Matias, C. Sirrenberg, R. C. Wade, and M. Frech, Protein conformational flexibility modulates kinetics and thermodynamics of drug binding., *Nat. Commun.* **8**, 2276 (2017).
- ³¹J. O. Daldrop, B. G. Kowalik, and R. R. Netz, External potential modifies friction of molecular solutes in water, *Phys. Rev. X* **7**, 041065 (2017).
- ³²N. Schaudinnus, B. Lickert, M. Biswas, and G. Stock, Global Langevin model of multidimensional biomolecular dynamics, *J. Chem. Phys.* **145**, 184114 (2016).
- ³³M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors, *Acta Crystallogr. B* **39**, 480 (1983).
- ³⁴J. Schiebel, R. Gaspari, T. Wulsdorf, K. Ngo, C. Sohn, T. E. Schrader, A. Cavalli, A. Ostermann, A. Heine, and G. Klebe, Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes, *Nat. Commun.* **9**, 166 (2018).
- ³⁵I. Buch, T. Giorgino, and G. De Fabritiis, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations, *Proc. Natl. Acad. Sci. USA* **108**, 10184 (2011).
- ³⁶N. Plattner and F. Noé, Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models, *Nat. Commun.* **6**, 7653 (2015).
- ³⁷P. Tiwary, V. Limongelli, M. Salvalaglio, and M. Parrinello, Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps, *Proc. Natl. Acad. Sci. USA* **112**, E386 (2015).
- ³⁸L. W. Votapka, B. R. Jagger, and A. H. of, 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 (2017).
- ³⁹A. Dickson and S. D. Lotz, Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore, *Biophys. J.* **112**, 620 (2017).
- ⁴⁰R. M. Betz and R. O. Dror, How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding?, *J. Chem. Theory Comput.* **15**, 2053 (2019).
- ⁴¹B. Efron and C. Stein, The Jackknife Estimate of Variance, *Ann. Stat.* **9**, 586 (1981).
- ⁴²S. Wolf et al., Estimation of Protein-Ligand Unbinding Kinetics Using Non-Equilibrium Targeted Molecular Dynamics Simulations., *J. Chem. Inf. Model.* **59**, 5135 (2019).
- ⁴³D. B. Kokh et al., Estimation of Drug-Target Residence Times by τ -Random Acceleration Molecular Dynamics Simulations., *J. Chem. Theory Comput.* **14**, 3859 (2018).
- ⁴⁴D. A. Schuetz, M. Bernetti, M. Bertazzo, D. Musil, H.-M. Eggenweiler, M. Recanatini, M. Masetti, G. F. Ecker, and A. Cavalli, Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics., *J. Chem. Inf. Model.* **59**, 535 (2019).
- ⁴⁵M. J. Abraham, T. Murtola, R. Schulz, S. Pll, J. C. Smith, B. Hess, and E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* **12**, 19 (2015).
- ⁴⁶V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* **65**, 712 (2006).
- ⁴⁷R. B. Best and G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, *J. Phys. Chem. B* **113**, 9004 (2009).
- ⁴⁸W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79**, 926 (1983).
- ⁴⁹G. Bussi and M. Parrinello, Accurate sampling using Langevin dynamics, *Phys. Rev. E* **75**, 2289 (2007).

SUPPORTING INFORMATION

Supporting Methods

MD simulation details

Protein and ion interactions were described by the Amber99SB force field^{1,2}, water molecules with the TIP3P model³. Simulations were carried out using Gromacs v2018⁴ in a CPU/GPU hybrid implementation. Protein protonation states were evaluated with propka⁵. Van der Waals interactions were calculated with a cut-off of 1 nm, electrostatic interactions using the particle mesh Ewald method⁶ with a minimal real-space cut-off of 1 nm. All covalent bonds with hydrogen atoms were constrained using LINCS⁷. After an initial steepest descent minimization with positional restraints of protein and ligand heavy atoms, an initial 0.1 ns equilibration MD simulation in the NPT ensemble was performed with a 1 fs time step and positional restraints of protein and ligand heavy atoms. A temperature of 290.15 K was kept constant using the Bussi (v-rescale) thermostat⁸ (coupling time constant of 0.2 ps), the pressure was kept constant at 1 bar using the Berendsen barostat⁹ (coupling time constant of 0.5 ps), followed by a second steepest descent minimization without restraints and a short 0.1 ns equilibration MD simulation in the NPT ensemble.

dcTMD calculations¹⁰ were carried out using the PULL code implemented in Gromacs using the "constraint" option employing a SHAKE implementation¹¹. In both protein-ligand complexes, the distance constraint employed was defined as the center of mass (COM) distance between the COMs of the C_α atom of the central β-sheet of the respective protein and of the ligand heavy atoms. 400 statistically independent start points of simulations were obtained by generating different atomic velocity distributions after the 10 ns unbiased simulations, all corresponding to a temperature of 290.15 K. After a 0.1 ns preequilibration using parameters as described above with positional restraints on protein and ligand heavy atoms and a constant distance constraint of all 400 simulation systems, constant velocity calculations were carried out with $v_c = 1$ m/s covering a distance of 2 nm, switching the barostat to the Parrinello-Rahman barostat¹². The constraint pseudo-force f_c was written out each time step.

NaCl

Free energy $\Delta G(x)$ and friction profiles $\Gamma(x)$ were obtained from 1000 trajectories of previous dcTMD calculations¹⁰ at pulling velocity $v_c = 1$ m/s. For a better sampling, we continued the unbiased fully atomistic simulations described in¹⁰ and extended them to a full microsecond of simulated time. As these simulations used a cubic simulation box, binding and unbinding waiting times cannot directly be compared to the results of our Langevin simulation with "reflective" borders (see below), which represent radial dynamics. To obtain data sets that allow such a comparison, we removed all time steps with $x < 0.265$ nm and $x > 1.265$ nm from MD trajectories, and calculated mean waiting times for the resulting cut $x(t)$ trajectories.

Trypsin-benzamidine

Benzamidine parameters were obtained using Antechamber¹³ and Acypype¹⁴ with atomic parameters derived from GAFF parameters¹⁵. Atomic charges were obtained as RESP charges¹⁶ based on QM calculations at the HF/6-31G* level using Orca¹⁷ and Multiwfn¹⁸. Trypsin (PDB ID 3PTB¹⁹) was placed into a dodecahedral box with dimensions of 7.5 x 7.5 x 5.3 nm³ side length and solvated with 8971 water molecules. 16 sodium and 25 chloride ions were added to yield a charge neutral box with a salt concentration of 0.1 M²⁰. After the initial equilibration, we added an additional 10.0 ns unbiased MD simulation to yield a converged protein structure.

Hsp90-inhibitor

Parameters of the resorcinol inhibitor were taken from Ref.²¹: here, inhibitor parameters were generated using Antechamber¹³ and Acypype¹⁴ with atomic parameters derived from GAFF parameters¹⁵ and AM1-BCC atomic charges^{22,23}. Solvated simulation boxes of the Hsp90-inhibitor complex were taken from²¹ (compound **1j**), which in turn are based on the 2.5 Å X-ray crystal structure with PDB ID 6FCJ²⁴.

Pathway separation

In the case of trypsin-benzamidine, pathway separation was performed by employing nonequilibrium principal component analysis²⁵ using a protein-ligand distance covariance matrix²⁶. Trajectories were projected onto the first two principal components and sorted according to pathways by visual inspection. The dcTMD correction was then carried out separately for such bundles of trajectories. Performing 200 pulling simulations, we found 84 trajectories to constitute the major unbinding pathway, for which free energy and friction profiles were converged, and whose free energy difference between bound and unbound state matches well the value known from experiment (see Fig. S2).

For the Hsp90-inhibitor complex, we employed a path separation based on geometric distances between individual trajectories²⁷. After aligning trajectories with the protein's C_α atoms as fit reference, we calculated the matrix of means over time of the root mean square distance between individual trajectories. We then applied the NeighborNet algorithm²⁸ to the matrix to cluster trajectories according to distances. From the considered 400 trajectories, the cluster that gave a free energy difference that was in best agreement with experiment (see Fig. S2) was taken by 93 single trajectories.

Langevin simulations

Langevin simulations used the integration scheme by Bussi and Parrinello⁸. Simulations were carried out with integrator step sizes $\Delta t = 1, 2, 5$ and 10 fs at a constant temperature of 300 K, or at 5 fs integrator step size and temperatures between 293.15 K to 420 K for NaCl, 380 K to 900 K for trypsin-benzamidine and 700 K to 1350 K for the Hsp90-inhibitor complex. See Table S1 for the convergence of binding and unbinding times of NaCl with respect to the time step. Simulations were run for 1 μ s of simulated time for NaCl and 5 milliseconds for protein-ligand systems. System coordinates were written out each 1 ps for NaCl and each 1 ns for protein-ligand systems.

The gradient of the potential of mean force was approximated as

$$\frac{dG(x)}{dx} \approx \frac{[\Delta G(x + \Delta x) - \Delta G(x)] + [\Delta G(x) - \Delta G(x - \Delta x)]}{2\Delta x} \quad (\text{S05})$$

Input free energy and friction profiles obtained from dcTMD were smoothed with a Gauss filter ($\sigma = 10$ and 40 ps, respectively). As friction fields in some cases still exhibited negative friction values after smoothing, we used the absolute values $|\Gamma(x)|$ as input for simulations. For x , we used a resolution of 1 pm. For compensation of data borders, we employed "fully reflective" boundary conditions: If the system jumped over a boundary x_{\max} at any time step by a distance a , it was put back to $x = x_{\max} - a$, and its velocity sign reversed.

Mean waiting times were calculated by defining geometric cores²⁹: For NaCl, the free energy surface was separated into the bound state $x < 0.31$ nm and unbound state $x > 0.43$. For trypsin-benzamidine, we used a bound state $x < 0.3$ nm and unbound state $x > 0.6$ nm, while for the Hsp90-inhibitor complex, we applied cores of $x < 0.3$ nm and $x > 0.9$ nm. As the native units of k_{on} are $s^{-1} M^{-1}$, all accounting rates were scaled by the "concentration" of ion or protein-ligand pairs in an effective volume of $\frac{4}{3}\pi x_{\text{end}}^3$, amounting to a molarity of 0.2 M for NaCl Langevin simulations and 50 mM for protein-ligand Langevin simulations.

Data evaluation

Minimal distance evaluation was performed using the MDanalysis Python library³⁰, nonequilibrium principal component analysis was carried out using the fastpca program³¹. Data evaluation was carried out using a Jupyter notebook⁷ employing the numpy³³, scipy³⁴ and astropy³⁵ Python libraries. Graphs were plotted using the matplotlib³⁶ Python library, molecular structures were displayed via PyMOL³⁷.

Uncertainty for prediction of rates with temperature boosting

To estimate the extrapolation error of T -boosting, we assume that the waiting time τ of an unbinding or binding process is exponentially distributed,

$$P(\tau) = \frac{1}{\langle \tau \rangle} e^{-\frac{\tau}{\langle \tau \rangle}}, \quad (\text{S06})$$

where $\langle \tau \rangle$ is a function of temperature T . Hence the expectation $\bar{\tau}$ is given by its mean $\langle \tau \rangle$ plus/minus the error of the mean

$$\bar{\tau}(T) = \langle \tau(T) \rangle \pm \frac{\langle \tau(T) \rangle}{\sqrt{N(T)}}, \quad (\text{S07})$$

where $N(T)$ denotes the number of simulated transitions. By changing to the dimensionless rate $k = t_0/\langle \tau \rangle$ (with t_0 being some timescale, e.g., ns) and employing Gaussian error propagation to lowest order³⁸, we obtain accordingly

$$\ln(\bar{k}(T)) = \ln(k(T)) \pm \frac{1}{\sqrt{N(T)}}. \quad (\text{S08})$$

Due to the rate expression $k \propto e^{-\Delta G/k_B T}$ with transition state energy ΔG , $\ln(k)$ depends linearly on $1/T$, i.e.,

$$\ln(k(T)) = \frac{a}{T} + b. \quad (\text{S09})$$

Linear regression theory³⁸ yields estimates for a and b as well as uncertainties

$$\sigma_b = \sqrt{\frac{\sum_i \frac{N(T_i)}{T_i^2}}{\Delta}} \quad (\text{S10})$$

and

$$\sigma_a = \sqrt{\frac{\sum_i N(T_i)}{\Delta}} \quad (\text{S11})$$

with

$$\Delta = \sum_i N(T_i) \sum_i \frac{N(T_i)}{T_i^2} - \left(\sum_i \frac{N(T_i)}{T_i} \right)^2, \quad (\text{S12})$$

where T_i denotes a discrete set of temperatures at which simulations are performed. Employing error propagation, we estimate the uncertainty of $\ln(k)$ at $T_{\text{ref}} = 300$ K as

$$\sigma_{\ln(k(T_{\text{ref}}))} = \sqrt{\left(\frac{\sigma_a}{T_{\text{ref}}} \right)^2 + \sigma_b^2}. \quad (\text{S13})$$

This yields for the desired relative uncertainty of the average waitingtime

$$\bar{\tau}(T_{\text{ref}}) = \langle \tau(T_{\text{ref}}) \rangle \pm \langle \tau(T_{\text{ref}}) \rangle \cdot \sigma_{\ln(k(T_{\text{ref}}))}. \quad (\text{S14})$$

To illustrate the typical magnitude of this uncertainty, we assume that we perform 10 Langevin simulations of length t_{LE} at different temperatures T_i ($i = 0, \dots, 9$)

$$T_i = T_0 + i \left(25 \frac{T_0}{T_{\text{ref}}} \right) \text{K}. \quad (\text{S15})$$

The first three temperatures (T_0 , T_1 and T_2) are chosen such that we observe $\approx 10^2$ transitions during the simulation time t_{LE} . Similarly, we assume to observe 10^3 transitions at T_3 , T_4 and T_5 , 10^4 transitions at T_6 , T_7 and T_8 and 10^5 transitions at T_9 . Using the case of $T_0 = T_{\text{ref}} = 300$ K, since we assumed 10^2 transitions at $T_0 = T_{\text{ref}}$ during the Langevin simulation time t_{LE} , the observed rate at 300 K is $k = 10^2/t_{\text{LE}}$. Choosing $t_{\text{LE}} = 5$ ms, the corresponding rate is $k(300\text{K}) = 1/50\mu\text{s}$, and we obtain for the error of the rate $\sigma_{\ln(k(T_{\text{ref}}))} = 7.7\%$. Alternatively, when we assume that we need to choose $T_0 \approx 450$ K in order to achieve 10^2 transitions, employing the boosting relation (4) in the main text and $t_{\text{LE}} = 5$ ms, the observed rate at 300 K is $k(300\text{K}) = 0.063 \text{ ms}^{-1}$ with an error of 10.6%.

Considering our Langevin simulations of trypsin described in the main text, where we used T -boosting at 13 temperatures from 380–900 K, the error at 300 K is estimated to be $\sigma_{\ln(k(T_{\text{ref}}))} = 3.3\%$. For Hsp90, the system with the highest considered barrier, we obtain $\sigma_{\ln(k(T_{\text{ref}}))} = 11.0\%$ at 300 K using Langevin simulations at 14 temperatures from 700–1350 K. As the overestimation of friction factors due to usage of constraints (see main text) results in a underestimation of rates by a factor of ~ 4 , and as the error of free energy profiles enters Eq. [4] in the exponent, the extrapolation error due to T -boosting can easily be made negligible in all practical cases.

SUPPORTING FIGURES

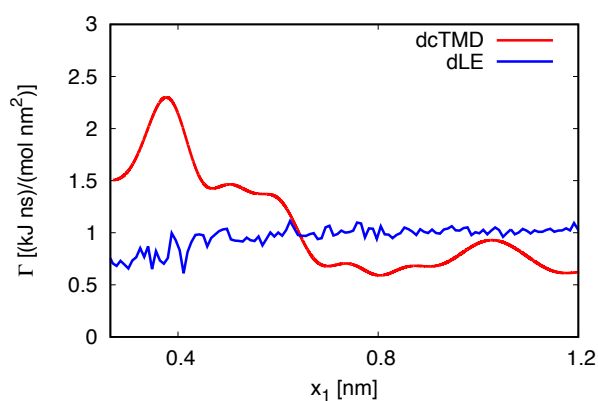


FIG. S5. Friction calculated from data-driven LE (dLE)³⁹ in comparison to friction profiles from dcTMD.

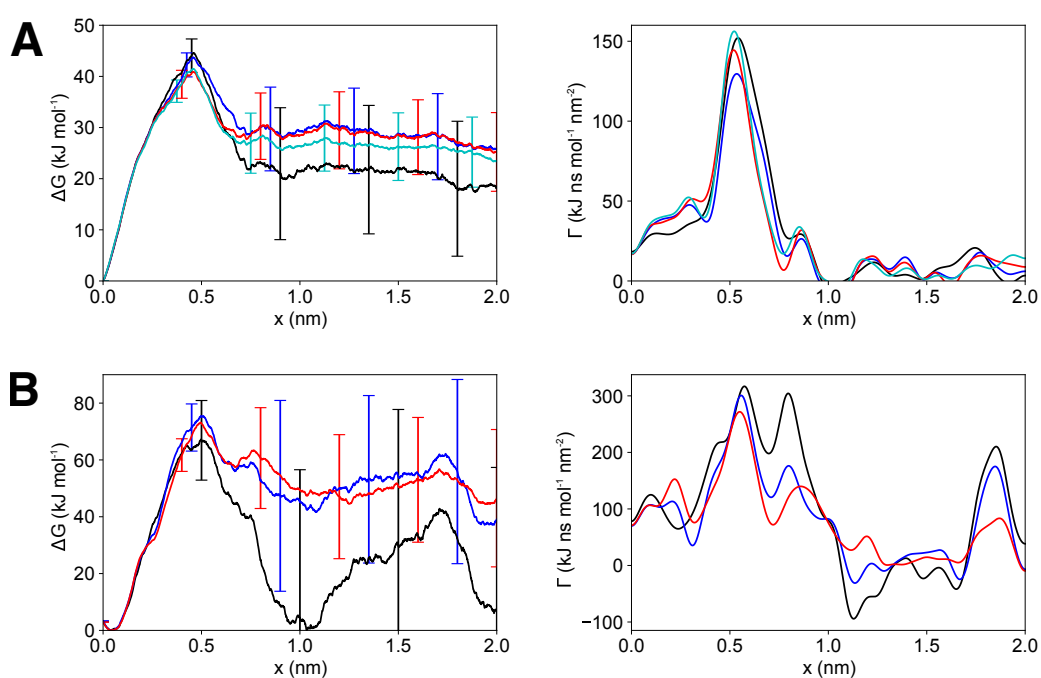


FIG. S6. Jackknife (“leave-one-out”) analysis⁴⁰ of free energy profiles and friction for (A) trypsin-benzamidine and (B) Hsp90-inhibitor complex. Error bars denote the Jackknife standard error obtained for various numbers of TMD trajectories (“samples”). Color code in (A): 52 samples in black, 84 in blue, 117 in red, 148 in cyan. Color code in (B): 30 in black, 50 in blue, 93 in red. The calculated binding free energies of ~ 27 kJ/mol (Trypsin) and ~ 45 kJ/mol (Hsp90) compare well to the experimental values of 28.0 ± 0.2 kJ/mol²⁰ and 40.7 ± 0.2 kJ/mol⁴¹, respectively.

SUPPORTING TABLES

TABLE S1. Convergence of the Bussi-Parrinello Langevin equation integration scheme⁸ with respect to the integration time step Δt . Shown are dissociation and association times and corresponding number of transitions obtained for NaCl at $T = 293$ K, as well as results from fully atomistic MD simulations (single trajectory of 1 μ s, 1 fs step size). Errors denote the standard error of the mean. A time step of 5 fs appears as the longest time step that results in suitable dissociation and association times.

Δt (fs)	dissociation time (ps)	no. of transitions	association time (ps)	no. of transitions
0.1	406 \pm 8	273 \pm 4	3,248 \pm 66	273 \pm 4
0.2	404 \pm 8	279 \pm 4	3,174 \pm 63	279 \pm 4
0.5	411 \pm 8	278 \pm 5	3,166 \pm 65	279 \pm 5
1.0	428 \pm 9	277 \pm 4	3,157 \pm 65	278 \pm 4
2.0	403 \pm 8	279 \pm 5	3,162 \pm 63	279 \pm 5
5.0	420 \pm 9	288 \pm 4	3,035 \pm 61	288 \pm 4
10.0	345 \pm 6	335 \pm 6	2,631 \pm 47	335 \pm 6
MD	124 \pm 6	1,023	848 \pm 47	1,024

SUPPORTING REFERENCES

- ¹V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* **65**, 712 (2006).
- ²R. B. Best and G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, *J. Phys. Chem. B* **113**, 9004 (2009).
- ³W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79**, 926 (1983).
- ⁴M. J. Abraham, T. Murtola, R. Schulz, S. Pll, J. C. Smith, B. Hess, and E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* **12**, 19 (2015).
- ⁵M. H. Olsson, C. R. S ndergaard, M. Rostkowski, and J. H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions, *J. Chem. Theory Comput.* **7**, 525 (2011).
- ⁶T. Darden, D. York, and L. Petersen, Particle mesh Ewald: An N log(N) method for Ewald sums in large systems, *J. Chem. Phys.* **98**, 10089 (1993).
- ⁷B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations, *J. Comp. Chem.* **18**, 1463 (1997).
- ⁸G. Bussi and M. Parrinello, Accurate sampling using Langevin dynamics, *Phys. Rev. E* **75**, 2289 (2007).
- ⁹H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* **81**, 3684 (1984).
- ¹⁰S. Wolf and G. Stock, Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction, *J. Chem. Theory Comput.* **14**, 6175 (2018).
- ¹¹J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical-integration of cartesian equations of motions of a system with constraints-molecular dynamics of n-alkanes, *J. Comput. Phys.* **23**, 327 (1977).
- ¹²M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *J. Appl. Phys.* **52**, 7182 (1981).
- ¹³J. Wang and R. Br uschweiler, 2D entropy of discrete molecular ensembles, *J. Chem. Theory Comput.* **2**, 18 (2006).
- ¹⁴A. W. Sousa da Silva and W. F. Vranken, ACPYPE - AnteChamber PYthon Parser interface, *BMC Res. Notes* **5**, 367 (2012).
- ¹⁵J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* **25**, 1157 (2004).
- ¹⁶C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The resp model, *J. Phys. Chem.* **97**, 10269 (1993).
- ¹⁷F. Neese, The ORCA program system, *WIREs Comput. Mol. Sci.* **2**, 73 (2012).
- ¹⁸T. Lu and F. Chen, Multiwfn: A multifunctional wavefunction analyzer, *J. Comput. Chem.* **33**, 580 (2012).
- ¹⁹M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors, *Acta Crystallogr. B* **39**, 480 (1983).
- ²⁰F. Guillain and D. Thusius, 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 (1970).
- ²¹S. Wolf et al., Estimation of Protein-Ligand Unbinding Kinetics Using Non-Equilibrium Targeted Molecular Dynamics Simulations., *J. Chem. Inf. Model.* **59**, 5135 (2019).
- ²²A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method, *J. Comput. Chem.* **21**, 132 (2000).
- ²³A. Jakalian, D. B. Jack, and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model - II. Parameterization and validation., *J. Comput. Chem.* **23**, 1623 (2002).

- ²⁴J. Güldenhaupt, M. Amaral, C. Kötting, J. Schartner, D. Musil, M. Frech, and K. Gerwert, Ligand-Induced Conformational Changes in HSP90 Monitored Time Resolved and Label Free-Towards a Conformational Activity Screening for Drug Discovery., *Angew. Chem. Int. Ed.* **57**, 9955 (2018).
- ²⁵M. Post, S. Wolf, and G. Stock, Principal component analysis of nonequilibrium molecular dynamics simulations, *J. Chem. Phys.* **150**, 204110 (2019).
- ²⁶M. Ernst, F. Sittel, and G. Stock, Contact- and distance-based principal component analysis of protein dynamics, *J. Chem. Phys.* **143**, 244114 (2015).
- ²⁷S. Bray, *Approaches to analyzing protein-ligand dissociation with targeted molecular dynamics*, Master Thesis, 2018.
- ²⁸D. Bryant and V. Moulton, Neighbor-net: an agglomerative method for the construction of phylogenetic networks., *Mol. Biol. Evol.* **21**, 255 (2004).
- ²⁹D. Nagel, A. Weber, B. Lickert, and G. Stock, Dynamical coring of Markov state models, *J. Chem. Phys.* **150**, 094111 (2019).
- ³⁰N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, Mdanalysis: A toolkit for the analysis of molecular dynamics simulations, *J. Comput. Chem.* **32**, 2319 (2011).
- ³¹F. Sittel, A. Jain, and G. Stock, Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates, *J. Chem. Phys.* **141**, 014111 (2014).
- ³²T. Kluyver et al., Jupyter Notebooks-a publishing format for reproducible computational workflows., in *Positioning and Power in Academic Publishing*, edited by F. Loizides and B. Schmidt, pages 87–90, IOP Press, 2016.
- ³³S. van der Walt, S. C. Colbert, and G. Varoquaux, The NumPy Array: A Structure for Efficient Numerical Computation, *Comput. Sci. Eng.* **13**, 22 (2011).
- ³⁴T. E. Oliphant, Python for Scientific Computing, *Comput. Sci. Eng.* **9**, 10 (2007).
- ³⁵A. M. Price-Whelan, B. M. Sipcz, H. M. Gnther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, and et al., The astropy project: Building an open-science project and status of the v2.0 core package, *Astron. J.* **156**, 123 (2018).
- ³⁶J. D. Hunter, Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* **9**, 90 (2007).
- ³⁷Schrödinger, LLC, The PyMOL Molecular Graphics System, Version 1.8, (2010).
- ³⁸I. Hughes and T. Hase, *Measurements and Their Uncertainties*, A Practical Guide to Modern Error Analysis, Oxford University Press, 2010.
- ³⁹N. Schaudinnus, B. Lickert, M. Biswas, and G. Stock, Global Langevin model of multidimensional biomolecular dynamics, *J. Chem. Phys.* **145**, 184114 (2016).
- ⁴⁰B. Efron and C. Stein, The Jackknife Estimate of Variance, *Ann. Stat.* **9**, 586 (1981).
- ⁴¹M. Amaral, D. B. Kokh, J. Bomke, A. Wegener, H. P. Buchstaller, H. M. Eggenweiler, P. Matias, C. Sirrenberg, R. C. Wade, and M. Frech, Protein conformational flexibility modulates kinetics and thermodynamics of drug binding., *Nat. Commun.* **8**, 2276 (2017).