Biophysical Journal, Volume *114*

Supplemental Information

Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways

Ariane Nunes-Alves, Daniel M. Zuckerman, and Guilherme Menegon Arantes

Escape of a small molecule from inside T4 lysozyme by multiple pathways

Ariane Nunes-Alves[1], Daniel M. Zuckerman[2]* and Guilherme Menegon Arantes[1]*

[1]*Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900, São Paulo, SP, Brazil*

[2]*Department of Biomedical Engineering, School of Medicine, Oregon Health & Science University, 2730 SW Moody Avenue, 97239, Portland, OR, US*

**Corresponding authors: zuckermd@ohsu.edu, garantes@iq.usp.br*

Supporting Information

S1.1 Choice of exploratory temperature and WE parameters

High temperature exploratory simulations were performed at 400 K. This temperature was chosen intuitively to increase the rate of transitions over barriers and lead to a more even sampling of pathways with different probabilities for ligand escape. It is expected that the ideal temperature to explore unbinding pathways is system-dependent, and a systematic test of the dependence of the pathways found on different temperatures was not performed here.

A low resampling interval (τ) may increase the chances of detecting trajectories occupying unvisited bins, but it will also increase the total computational cost due to the more resources spent in trajectory analysis and bin attribution. Initial WE simulations with τ = 2, 5 and 10 ps showed that bin occupancy is reached in shorter simulation time for lower τ values. Reaching full bin occupancy in short simulation times leads to faster sampling of unbinding events. WE simulations with different τ values had similar computational cost (defined as the time to propagate the trajectories plus the time for trajectory analysis and bin attribution) to perform 10 ps of MD simulation time. Thus, τ = 2 ps was adopted for the production WE simulations here.

A high number of trajectories per bin (traj/bin, in Table S2) may result in a better occupancy of bins and also increase the chances of reaching unvisited bins, but it also increases the total number of trajectories and the computational cost of a WE simulation. Given that our simulations ran on computer nodes with 48 threads and we tried to keep one trajectory run-

ning per thread, a total of 144 trajectories (N_{tot} =3×48) was used for each resampling interval. Thus, the number of trajectories per bin was chosen as 4 (= 144/36 ≈ N_{tot}/bins) which limited the amount of correlated trajectories and the computational cost, and also allowed a reasonable bin occupancy.

The binning scheme to partition the progress coordinate was designed with the help of a rough potential of mean force (PMF) obtained from umbrella sampling simulations for benzene unbinding along the blue pathway. The binning scheme was chosen so that each bin had to transpose ∼ 0.6 kcal/mol steps of the PMF. Umbrella sampling was done with the GROMACS pull code (1), using the distance between ligand and binding site COM as a progress coordinate. Eleven (11) windows spaced at 0.2 nm bins were used, with reference distances ranging from 0.4 to 2.4 nm and a force constant of 1000 kJ/(mol nm^2). Each window run for 30 ns and the last 28.4 ns were used to generate the PMF with the weighted histogram analysis method (2).

S1.2 Description of Voronoi centers

Voronoi centers that describe each unbinding pathway were manually chosen from structures of ligand-protein complexes obtained from unbinding trajectories in the exploratory set of simulations. Each structure was aligned to the T4L L99A mutant crystal structure (PDB 1L83) using the Cα atoms of the binding site as a reference (section 2.3) and later employed in the WE simulations as Voronoi centers, to compute distances and define bins along the progress coordinate. One file for each pathway (blue, pink, cyan and orange) containing the coordinates of protein-ligand (T4L-benzene) complexes used for the definition of the Voronoi centers is included as Supporting Information.

S1.3 Reassignment of the unbinding trajectories to the correct pathways

Visual inspection of the trajectories showed that WE simulations with a progress coordinate defined from Voronoi bins for the blue pathway also sampled unbinding events for the orange pathway, and vice-versa. A post-production step was performed to properly separate unbinding events and reassign to the correct exit pathway. The unbinding event was reassigned to the pathway which contained the Voronoi center closest to the unbound state configuration. This

was possible by using 27 Voronoi centers (instead of 25 used in the WE production runs) to the blue pathway and 24 (instead of 23) centers to the orange pathway. Reassignment allowed proper separation of pathways and correct estimation of rate constants and conformational changes for each pathway.

S1.4 Identification of protein conformational transitions involved in ligand unbinding

Conformational changes allowing ligand exit from T4L were analyzed. Metastable microstates were associated to Voronoi bins with high lifetimes and the protein transitions involved in ligand progression along the unbinding pathway were found by:

- identification of the metastable microstate with the Voronoi bin with highest lifetime for each successful unbinding trajectory;
- collection of the structure in the frame immediately after the metastable state was unoccupied (transition structure);
- construction of a list of the side chains contacting benzene in the transition structure, using a distance cutoff of 0.5 nm;
- computation of the (heavy atom) RMSD between transition structure and reference structure for each side chain, after structure alignment using the binding site C_{alpha} defined in section 2.2;
- selection of 3 side chains with highest RMSD.

Reference structures were collected either 100 ps or 10 ps before the transition structure in the same trajectory.

Analysis revealed that movement of helices C, D, F, H and J could also be involved in ligand unbinding. This was quantified by:

- identification of the metastable microstate and collection of transition structure for each successful unbinding trajectory;

- check if one of the side chains of helices C (residues 69-81), D (residues 83-90) F (residues 107-114), H (residues 126-135) or J (residues 143-155) was contacting benzene in the transition structure, using a cutoff of 0.5 nm;
- check the helix displacement.

The criteria for helix displacement was based on average atom-pair distance distributions shown in figure S2. Helix C was considered displaced when the average distance was higher than 1.01 nm, helix D when higher than 1.02 nm, helix F when higher than 1.25 nm and helices H and J when higher than 0.67 nm.

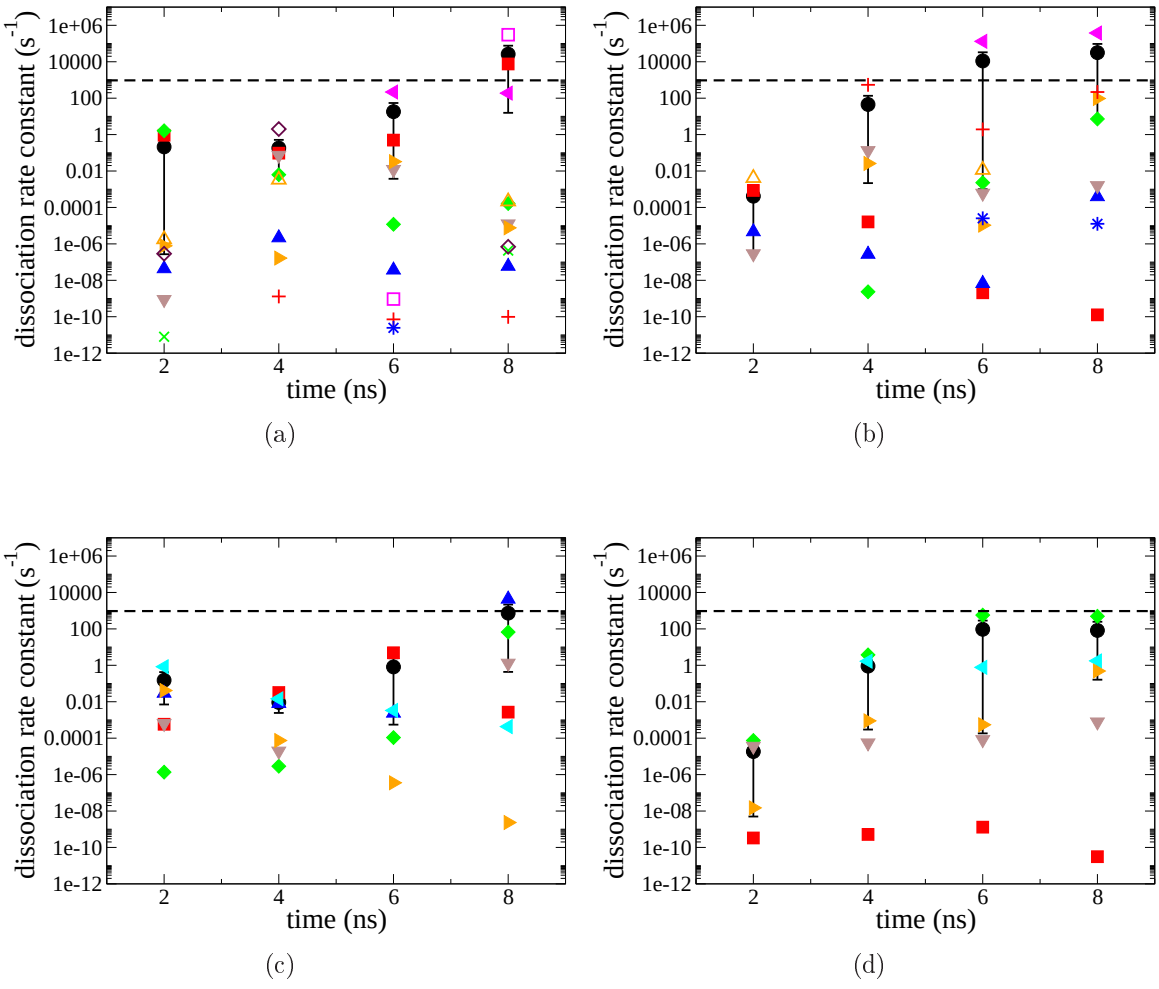


Figure S1: Average and individual rate constant estimated for each pathway of benzene unbinding from T4L. (a) Blue pathway, (b) orange pathway, (c) pink pathway and (d) cyan pathway. Black circles indicate averages, remaining symbols indicate individual estimates from one WE simulation. Bars indicate a 90% confidence interval from bootstrapping for the averages, which are dominated by larger values. Dotted lines indicate the experimental rate constant.

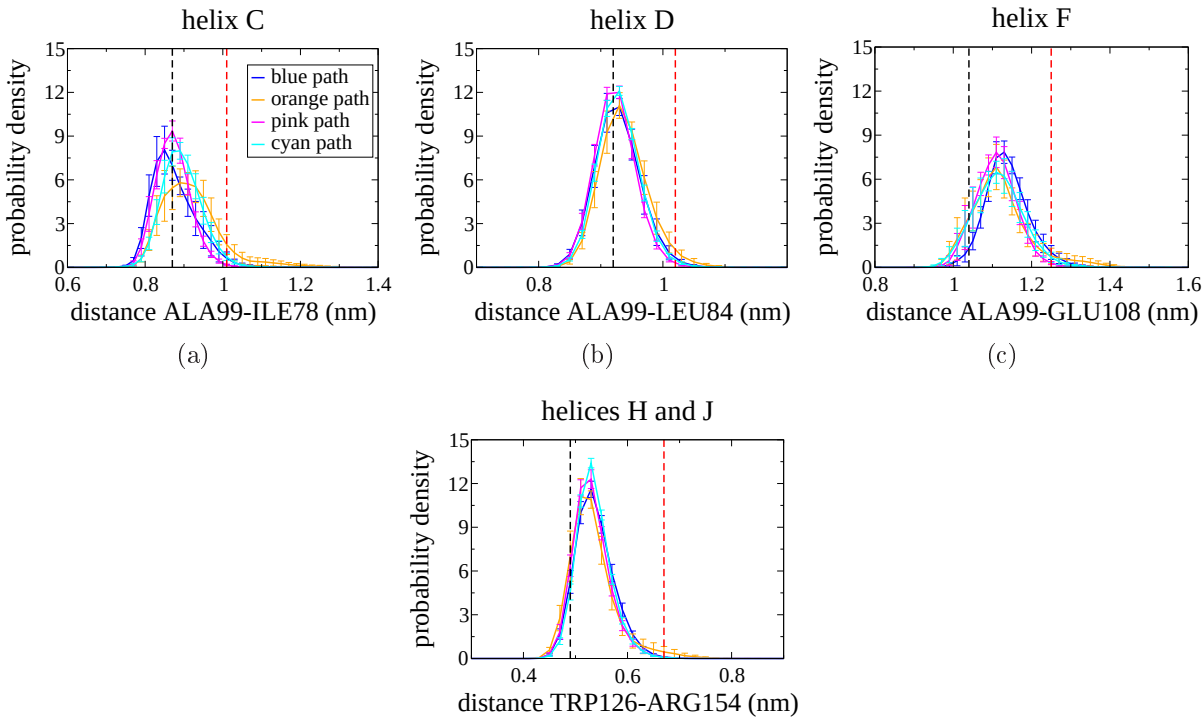
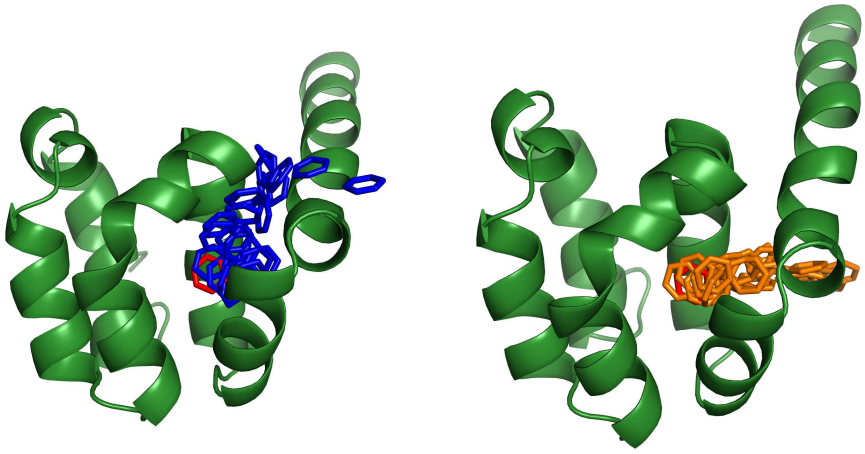
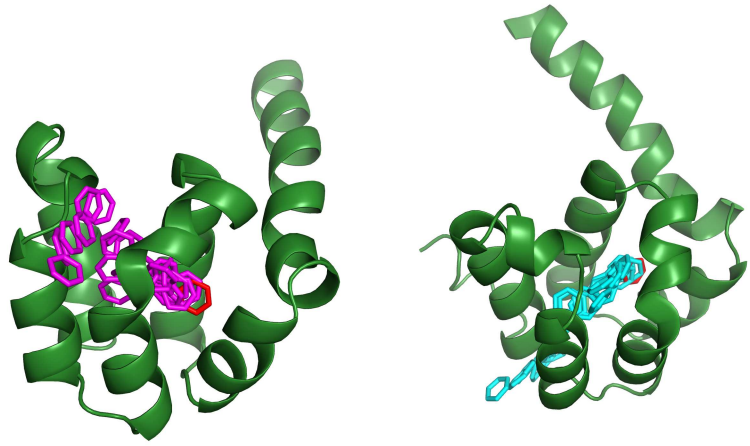


Figure S2: Average distance distributions from 4 to 8 ns of six WE simulations for each pathway (color coded as shown in the insert). Bars represent the standard error in the averages. Black and red lines represent the distance in the crystal structure and the distance criteria for helix displacement, respectively.



(a) (b)



(c) (d)

Figure S3: Benzene positions in the structures used as Voronoi centers to sample the blue (a), orange (b), pink (c) and cyan (d) pathways. Benzene position in the crystal structure is shown in red sticks and was also used as a Voronoi center in each path. Only the T4L C-terminal domain is shown.

Table S1: Definition of additional progress coordinates for benzene unbinding. Two atom-pair distances were selected for each pathway. The bin boundaries (in nm) and the Voronoi bins where the progress coordinate was added are also shown. Voronoi bins are numbered in order of increasing distance from the binding site.

pathway	**distance**	**bin boundaries**	**Voronoi bins**
blue	CZ:Y88 - CA:A99	0.70, 0.83, 0.95	1, 2
	CA:L84 - CA:A99	0.90, 0.97, 1.05	5, 6
orange	CZ:Y88 - CA:A99	0.70, 0.83, 0.95	1, 2
	CB:Y88 - CB:I78	0.50, 0.58, 0.65	7, 8
pink	CB:V111 - CA:A99	0.80, 0.85, 0.90	2, 3
	CE:M102 - CA:A146	0.70, 0.77, 0.85	7, 11
cyan	CG:L133 - CA:S117	0.62, 0.72, 0.80	8, 9
	CD:R154 - CA:W126	0.60, 0.70, 0.80	19, 20

Table S2: Parameters used to run WE simulations for different progress coordinates (PC). WE sim.: number of WE simulations run, traj/bin: number of trajectories per bin, τ: resampling frequency (in ps), iter: number of resampling steps of one WE simulation, max length: maximum length of one WE simulation (in ns, calculated as τ*iter), bins: amount of bins of the progress coordinate calculated as (N-M)+M*J when Voronoi bins are used as progress coordinate, where N is the total number of Voronoi bins and M=4 is the number of Voronoi bins which were partitioned by an additional progress coordinate composed of J=4 bins, as shown in table S1, aggreg. time: aggregate simulation time of the progress coordinate in ns, calculated as (WE sim.)*(traj/bin)*(max length)*bins.

PC	**WE sim.**	**traj/bin**	τ	**iter**	**max length**	**bins**	**aggreg. time**
dist	4	5	10	150	1.5	12	360
RMSD	5	5	10	150	1.5	17	637.5
V_{blue}	6	4	2	4000	8	37	7104
V_{orange}	6	4	2	4000	8	35	6720
V_{pink}	6	4	2	4000	8	36	6912
V_{cyan}	6	4	2	4000	8	38	7296

Bin boundaries for distance PC (in nm): 0.30, 0.35, 0.38, 0.41, 0.50, 0.55, 0.60, 0.65, 0.80, 1.60, 1.68. Boundaries for RMSD PC (nm): 0.15, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70, 0.80, 0.90, 1.00, 1.10, 1.50, 1.90.

Supporting References

1. Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. 2013, GROMACS 4.5: a high–throughput and highly parallel open source molecular simulation toolkit. Bioinformatics, 29, 845–854.

2. Roux, B. 1995, The calculation of the potential of mean force using computer simulations. Comp. Phys. Comm., 91, 275–282.