Electronic Supplementary Information for PCCP Article: Mechanistic Analysis of Light-Driven Overcrowded Alkene-Based Molecular Motors by Multiscale Molecular Simulations

Mudong Feng and Michael K. Gilson

February 16, 2021

Department of Chemistry and Biochemistry, and Skaggs School of Pharmacy and Pharmaceutical Sciences, UC San Diego, La Jolla, CA 92093

This document provides additional data and details settings of the calculations mentioned in the main text. To reproduce or adapt these calculations, see Github repository github.com/fengmudong/motor-paper, where all the input files and associated instructions can be found.

1 Additional Figures

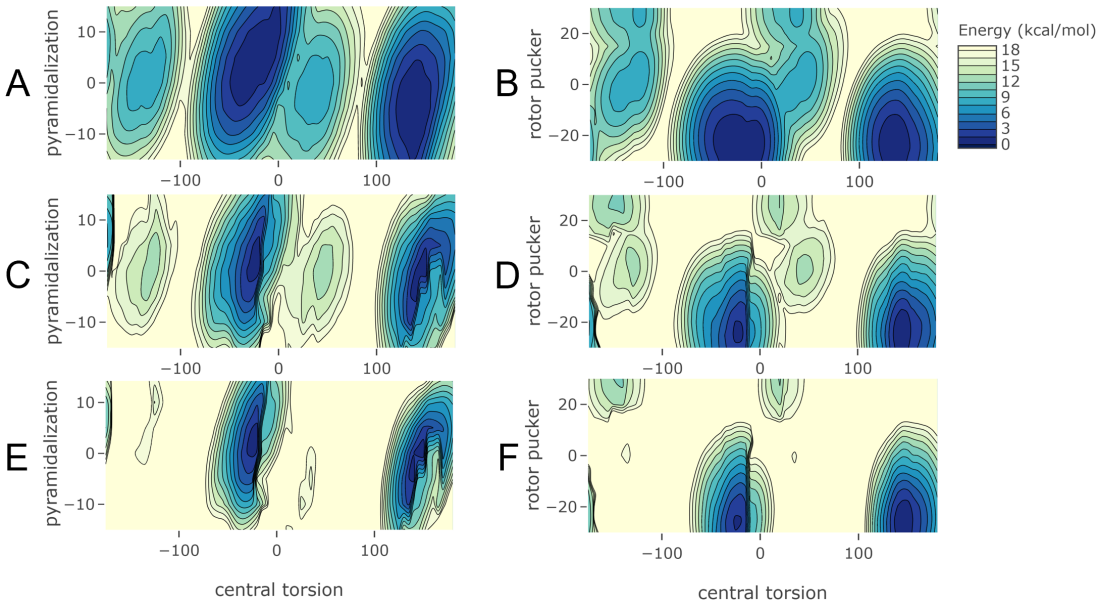

Figure S1: Potential energy surfaces of Motor N. A,B: quantum mechanical (QM) energy. C,D: energy from forcefield after fitting to QM. E,F: energy from GAFF2 forcefield, without any fitting to QM. The scanned coordinates are the central torsion angle, 18-15-14-10; the pyramidalization improper dihedral, 14-1-13-15; and the rotor pucker dihedral, 27-15-16-25.The force field energy for each grid point was evaluated at the corresponding QM conformation without further relaxation. For this larger molecule, the QM calculations used a smaller basis set, 3-21G.

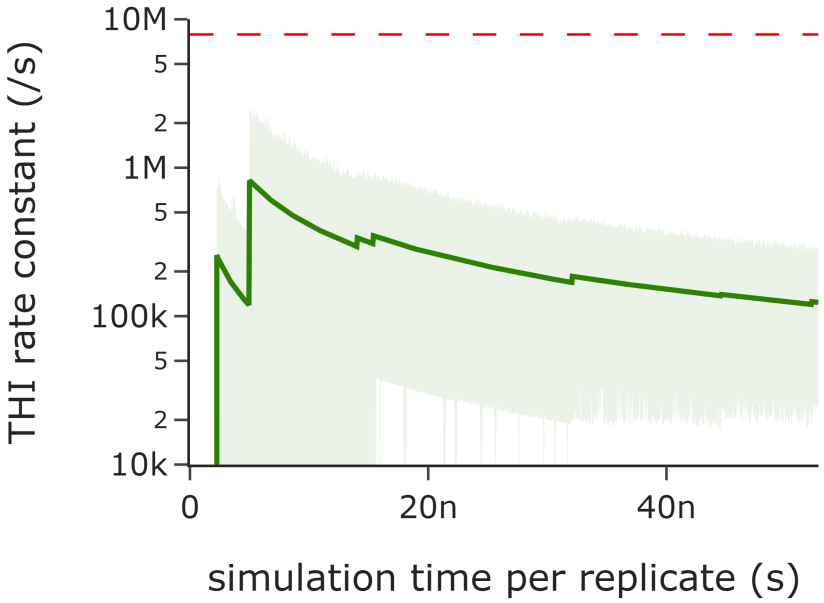

Figure S2: THI rate constant computed with WESTPA of Motor S under 3.4 kcal mol^{-1} load, as a function of simulation time per replicate, with shading to indicte 95% bootstrapping confidence interval. Dashed red line indicates experimental rate constant of the non-loaded motor at 300K, 7.9×10^6 s^{-1}, read from Arrhenius plots of Klok et al.[1].

2 Cartesian Coordinates of Selected Structures

Files MS1.pdb, MS2.pdb, MS3.pdb, and MS4.pdb each contains 10 sample conformations of Motor S metastable substates MS1, MS2, MS3, MS4, respectively. File stable.pdb contains a DFT optimized structure of the stable form of Motor S. For more technical purposes or for reproducing calculations in this study, more Cartesian structures in various formats can be found in the Github repository.

3 Trajectory Videos

Four trajectories videos are mentioned in the main text to animate the motor rotation process, and are available online. To make thermal fluctuations less distracting, the trajectory videos were smoothed in VMD [2] by averaging over a 20 frame moving window. We did not observe any noteworthy solvent motion, so solvent molecules were stripped out. To facilitate visualization, center-of-mass translation and rotation were removed and the stator atoms were aligned across snapshots to make the stator appear steady. Note that in mndo-amber.mp4, the surface-hopping trajectory at the beginning has one frame per 1fs, whereas the classical trajectory that follows has one frame per 800fs, and therefore appears less smooth.

4 Settings Related to Surface-hopping simulations

To describe the photoexcitation leading to central double bond rotation, we chose a minimal active space consisting of the two π bonded orbitals on the two central carbon atoms. Electron correlation was computed to the CISD level. During MD, the active space was tracked by an overlap criterion so that the molecular orbitals retain their character, to a threshold of at least 70 percent overlap with previous active space. The time step of nuclear motion was 0.5fs for ground state sampling simulations, or 0.1fs for most surface-hopping simulations. One trajectory frame was saved to disk every 10 steps. For motor N, which has more atoms and slower calculations, we increased the surface-hopping timestep to 0.2fs. We found the 0.2fs simulation results of Motor N to be similar to 0.1fs results of Motor N.

5 Settings Related to Plain Classical MD Simulations

The short-ranged PME cutoff was 8 Å, the time step was 2fs, the SHAKE algorithm was used to constrain bonds involving hydrogen, and a snapshot was written to disk every 400 steps. The Monte Carlo barostat was used, with a compressibility of comp=103 (in unit $1 \times 10^{-6}\,\mathrm{bar}^{-1}$), based on measured compressibility of dichloromethane.

6 Settings Related to WESTPA Simulations

The weighted ensemble MD simulations used the same parameters used in above plain MD simulations. We found that whether or not SHAKE was used does not affect the computed rate constant. After each WESTPA iteration of 50ps, replicates were examined for cloning or merging, for the target of having 20 replicates in each bin. For Motor S, the bin boundaries, in degrees, were -180, -80, -70, -60, -50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 60, 80, 100, 120, 140, 160, 180 for the first progress coordinate, and 0, 20, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 360 for the second. For Motor N, we used slightly different bin boundaries in the first coordinate: -180, -60, -50, -40, -30, -20, -10, 0, 10, 20, 30, 40, 60, 80, 100, 120, 140, 180.

7 Settings related to MSM

The high dimensional trajectories of atomic positions were first condensed to time series of all the dihedrals in the solute motor's topology file, each associated with a forcefield term. The time series were calculated at intervals of 0.8ps. This interval was also used as the lag time, i.e., the time resolution of the MSM. We also tested lag times up to 80 ps, and found that longer lag times did not significantly change the computed rate constants. Then the number of dimensions were further reduced to 8 for Motor S, and 19 for Motor N, by using time-lagged independent component analysis (TICA) on the time series. The number of dimensions was chosed as the minimal number that still preserves high VAMP2 score. Next, the continuous time series were discretized to 64 conformational clusters using k-means clustering. Then the clusters were classified by Perron cluster analysis (PCCA) [3] into four metastable substates. We used sharp classification instead of fuzzy probabilistic classification, so that pairs of substates that do not interconvert directly receive transition counts of exactly zero rather than small numbers. Then the MSM, as a 4x4 transition matrix, was computed from the time series of interconversion among these four substates. The Chapman-Kolmogorov test [3] was conducted which validates the MSM. Our workflow is available in a more detailed and easily reproducible format, as Jupyter notebooks in our Github repository.

8 Settings Related to Umbrella Sampling Simulations

MD simulation parameters are the same as in plain MD. All simulation windows start with the same AMBER restart file used in plain MD; in the restart file, the motor is at metastable form and the system has been equilibrated for plain MD. The spring constant of the window restraints were determined by trials; we found that a too large spring constant may crash the simulation, while a too small spring constant may fail to pull the reaction coordinate near the window center. The first 100 frames of each window were excluded from analysis to eliminate pulling transients.

References

[1] M. Klok, N. Boyle, M. T. Pryce, A. Meetsma, W. R. Browne, and B. L. Feringa, "MHz Unidirectional Rotation of Molecular Rotary Motors," *Journal of the American Chemical Society*, vol. 130, pp. 10484–10485, Aug. 2008.

[2] W. Humphrey, A. Dalke, and K. Schulten, "VMD Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[3] C. Wehmeyer, M. K. Scherer, T. Hempel, B. E. Husic, S. Olsson, and F. Noé, "Introduction to Markov state modeling with the PyEMMA software [Article v1.0]," *Living Journal of Computational Molecular Science*, vol. 1, no. 1, p. 5965, 2019.