Revisiting textbook azide-clock reactions: A "propeller-crawling" mechanism explains differences in rates

Anthony T. Bogetti,[†] Matthew C. Zwier,[‡] and Lillian T. Chong[*,†,¶]

[†]*Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

[‡]*Department of Chemistry, Drake University, Des Moines, Iowa 50311*

[¶]*Corresponding author*

E-mail: ltchong@pitt.edu

Abstract

An ongoing challenge to chemists is the analysis of pathways and kinetics for chemical reactions–including transient structures between the reactants and products that are difficult to resolve using laboratory experiments. Here, we enabled direct molecular dynamics simulations of a textbook series of chemical reactions on the hundreds of ns to µs timescale using the weighted ensemble (WE) path sampling strategy with hybrid quantum mechanical/molecular mechanical (QM/MM) models. We focused on azide-clock reactions involving addition of azide anion to each of three long-lived trityl cations in an acetonitrile-water solvent mixture. Azide additions are common click-chemistry reactions of great interest to synthetic chemists. Results reveal a two-step mechanism: (1) diffusional collision of reactants to form an ion-pair intermediate, (2) "activation", or rearrangement of the intermediate to the product. Our simulations not only yield reaction rates that are within error of experiment, but also rates for individual steps, indicating the activation step as rate-limiting for all three cations. Further, the trend

in reaction rates is due to differing extents of the azide anion "crawling" along the cation's phenyl-ring "propellers" during the activation step. Our study demonstrates the power of analyzing pathways and kinetics to gain insights on reaction mechanisms, underscoring the value of including WE and other related path sampling strategies in the modern toolbox for chemists.

Introduction

Dynamical effects on chemical reactions–involving the atomic motions and associated kinetics–have been increasingly recognized as important features of reaction mechanisms.[1] Of great interest is therefore the generation of complete atomically detailed pathways from the reactants to the products. While spectroscopy experiments can detect product formation and measure the rate constant for the overall reaction, such experiments are unable to directly provide rate constants of individual steps or atomic structures of transient states. On the other hand, high-level quantum mechanical calculations have been widely used to model the structures of transition states, metastable states (intermediates), and stable states (reactants, products). The state-of-the-art in accounting for solvent effects in such calculations is to run *ab initio* molecular dynamics (AIMD) simulations with explicit solvent molecules.[2,3] However, AIMD simulations are too computationally intensive for generating a diverse ensemble of reaction pathways with rigorous kinetics.

In principle, complete reaction pathways–including transient states that are too fleeting to be captured by experiments–can be generated at femtosecond resolution by physics-based simulations with hybrid quantum mechanical/molecular mechanics (QM/MM) models in which the reacting region of the system is modeled quantum mechanically and the remainder of the system is represented using classical molecular mechanics models. Furthermore, such simulations can complement experiment by providing direct estimates of rate constants for forming any arbitrary state of the reaction. While hybrid QM/MM molecular dynamics simulations are less computationally intensive than AIMD simulations, the generation of

unbiased reaction pathways has been a challenge, often requiring either the application of external biasing forces or modification of the free energy landscape.[4]

One promising class of approaches for enhancing the sampling of complete, unbiased reaction pathways is path sampling, particularly ones that generate continuous pathways (*e.g.* transition path sampling[5,6] and weighted ensemble sampling[7,8]). These approaches focus computing power on the functional transitions between stable states rather than the stable states themselves,[9] exploiting the fact that the transition time over the effective free energy barrier can be orders of magnitude faster than the waiting time in the initial stable state. For example, transition path sampling has been used to generate pathways for enzyme-catalyzed reactions,[10,11] and pathways along with rate constants for both chemical reactions in solution[12] and enzyme-catalyzed reactions.[13] While weighted ensemble sampling has not yet been applied to simulations of chemical reactions–until now–the strategy has been demonstrated to be orders of magnitude more efficient than conventional simulations in generating pathways and/or rate constants for complex processes that range from microseconds (e.g. binding processes of proteins[14,15] and DNA[16]) to milliseconds (e.g. protein folding)[17] to seconds (e.g. large-scale conformational switching in proteins)[18,19] and beyond (e.g. protein-ligand unbinding).[20]

Here we applied the weighted ensemble (WE) strategy to enable hybrid QM/MM simulations of textbook azide-clock reactions in a 1:2 v/v acetonitrile:water mixture of explicit solvent molecules (Figure 1A). These reactions each involve addition of azide anion to a long-lived cation and are generally assumed to occur at the diffusion limit (5×10^9 M^{-1} s^{-1})[21] such that the azide functions as a "clock" for the lifetimes of the cations.[21,22] Azide addition is commonly used in click chemistry reactions.[23] We simulated azide addition to each of three "propeller-shaped" trityl cations (Figure 1B): the unsubstituted cation (T^+), a cation with an electron-donating, methoxy substituent (4-OCH_3-T^+), and a cation with an electron-withdrawing, trifluoromethyl substituent (4-CF_3-T^+). As measured by laser flash photolysis, the rate constants for these reactions span an order of magnitude, ranging in timescale from

hundreds of ns to µs.[21] Our simulations not only yielded ensembles of reaction pathways, but also direct calculations of rate constants.

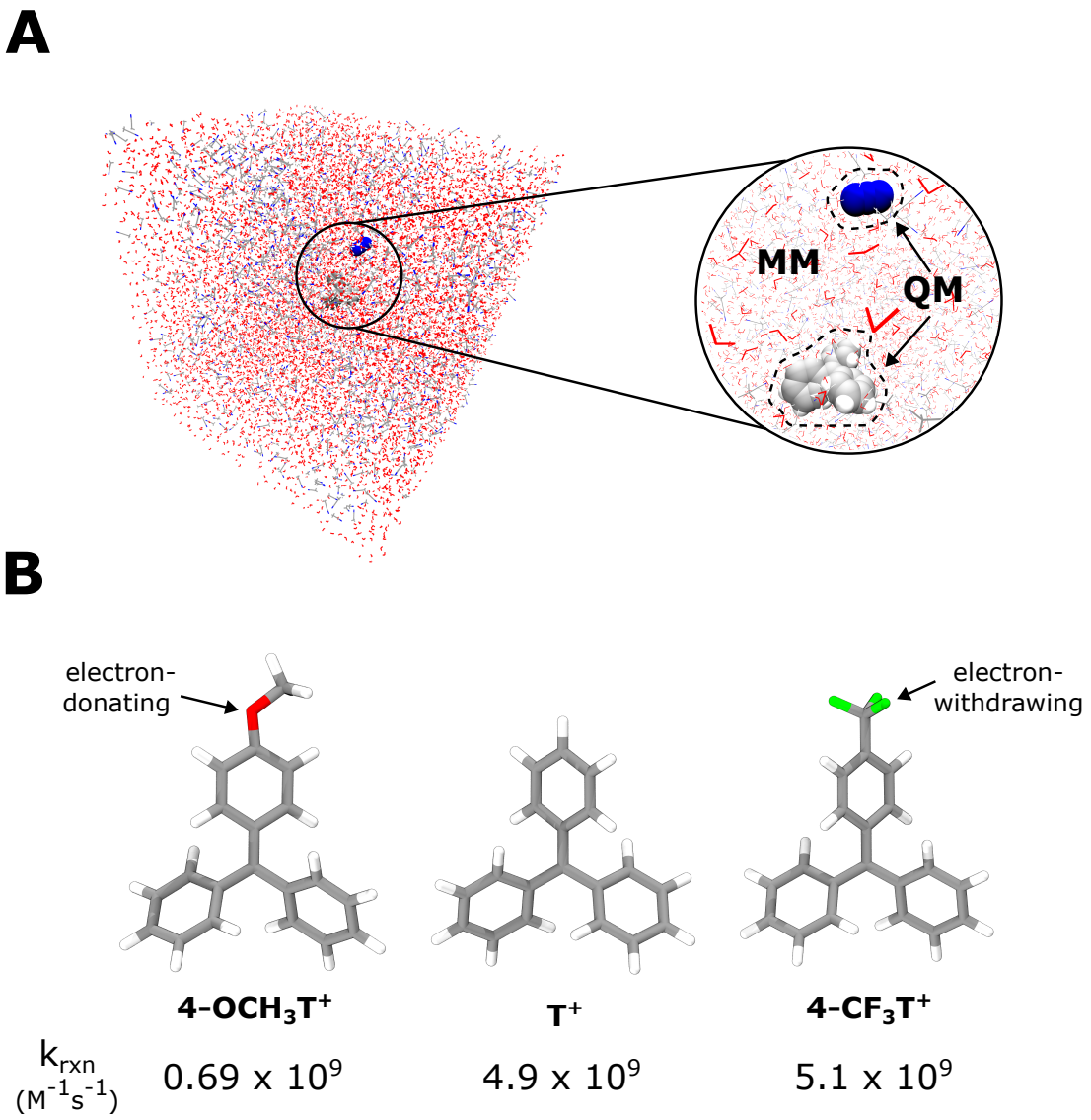

Figure 1: The azide-clock reactions in this study involved the addition of azide anion to three different trityl cations with reaction rate constants k_{rxn} that span an order of magnitude. (A) A cubic box containing the reactant solutes and explicit 1:2 v/v acetonitrile:water solute mixture, with the solutes modeled quantum mechanically and the solvent represented using classical, molecular mechanics models. (B) The three cations are shown in order of increasing reaction rate constant k_{rxn} for addition of azide anion. The 4-OCH_3-T^+ cation contains an electron-donating methoxy substituent and the 4-CF_3-T^+ cation contains an electron-withdrawing tri-fluoro substituent. The k_{rxn} values shown are published from previous laser flash photolysis experiments by others.[21]

Methods

Preparation of the hybrid QM/MM model

For each reaction simulation, we used a hybrid QM/MM model in which the QM region consisted of the reactants (azide anion and trityl cation) and the MM region consisted of the solvent molecules (1:2 v/v acetonitrile:water mixture). The QM region was modeled using the PM6-D semi-empirical method.[24] Van der Waals parameters from the GAFF force field were used for the azide and cations.[25] The MM region was presented using TIP3P water molecules[26] and acetonitrile molecules using compatible parameters.[27] Interaction energies between the QM and MM regions were treated with electrostatic embedding and long-range electrostatics with the particle mesh Ewald method.[28]

We generated starting models separately for the azide anion, the three trityl cations, and solvent molecules (acetonitrile and water) by (i) constructing the molecular models separately and energy-minimizing the models in vacuum using the Avogadro software and the GAFF force field,[25] and (ii) optimizing the geometry of each model at the RI-MP2 level of theory[29] with the cc-pVTZ basis set and a cc-pVDZ/C auxiliary basis set using the Orca 4.1.2 software package (see Figure S1 for the optimized geometries).[30] For each of the three cations, we positioned the geometry optimized cation and azide conformations at a 20 Å separation distance between the central carbon of the cation and the nearest azide nitrogen along a vector perpendicular to the plane of the cation. Next, we used the PACKMOL package[31] to solvate the reactants with a pre-equilibrated cubic box containing a 1:2 v/v acetonitrile:water mixture of the corresponding geometry optimized molecules and a 20 Å buffer distance between the reactants and the edges of the box. The pre-equilibrated solvent box was previously subjected to energy minimization and 20 ps of NVT equilibration followed by 1 ns of NPT equilibration using the AMBER 18 software package.[32]

Simulation workflow

Our workflow for simulating the chemical reactions involved two stages: 1) conventional simulations of the unassociated reactants, 2) weighted ensemble (WE) simulations of the chemical reaction starting from configurations sampled in stage #1. Full details are provided below.

Conventional simulations of the reactants

The solvated starting model of each pair of unassociated reactants was energy minimized and equilibrated in two stages using a hybrid QM/MM model and position restraints on the reactants, as implemented in the AMBER 18 software.[32] In the first stage of equilibration, the solvent of each system was equilibrated for 25 ps at constant volume and temperature (20°C). In the second stage, the solvent was equilibrated for 1 ns at constant pressure (1 atm) and temperature (20°C).

Following equilibration, conventional simulations of each system were run for 6 ns at constant temperature (20°C) and pressure (1 atm) to sample different relative orientations of the unassociated azide and cation molecules, restraining the distance between the nitrogens of the azide and the central carbon of the cation to 20 Å. Temperature was maintained using a weak Langevin thermostat with a collision frequency of 0.001 ps^{-1} and pressure was maintained using the Monte Carlo barostat with a coupling constant of 1 ps^{-1}. A time step of 1 fs was used. For each chemical reaction, an ensemble of 50 unassociated reactant conformations was generated by saving conformations every 100 ps from the last 5 ns of the standard simulation. This ensemble of unassociated conformations was used to initiate WE simulations of the corresponding chemical reaction as described below.

Weighted ensemble simulations of reaction pathways

The weighted ensemble (WE) strategy involves initiating multiple weighted trajectories in parallel and iteratively applying a resampling procedure after propagating dynamics for a fixed resampling time interval τ.[7,8] Configurational space is typically divided into bins along a progress coordinate. The resampling procedure aims to provide even coverage of configurational space with a target number of N trajectories per bin by either replicating trajectories that have made transitions to less-visited regions or terminating trajectories that have not made such transitions. Importantly, trajectory weights are tracked rigorously such that *no bias is introduced into the dynamics*, enabling direct estimation of rate constants. To maintain non-equilibrium, steady-state conditions, each trajectory that reaches the target state is "recycled", terminating the trajectory and spawning off a new trajectory from the initial state with the same trajectory weight.

To generate a large ensemble of continuous pathways for each azide-clock reaction, we ran five independent WE simulations of the reaction using the open-source, highly scalable WESTPA 2.0 software package[33] according to best practices.[34] We initiated each WE simulations by running five equally weighted trajectories and applied a resampling time interval τ (WE iteration) of 0.5 ps with a target number of 5 trajectories/bin. We used a two-dimensional WE progress coordinate consisting of the minimum separation distance between any nitrogen of the azide anion and (i) the central carbon of the cation, and (ii) any carbon of the cation. This progress coordinate was binned as illustrated in Figure S2. Trajectories were recycled when the minimum separation distance between the azide and cation was $<$1.6 Å.

For each reaction, simulation convergence was assessed by examining the time-evolution of the overall reaction rate constant (k_{rxn}), averaged over all five WE simulations (Figure S3). Each simulation was run for 500 WE iterations of applying the resampling procedure. As an additional verification of convergence, we first applied the weighted ensemble steady-state (WESS) reweighting procedure[35] to each WE simulation after 500 WE iterations to

reweight trajectories for a steady state and then restarted each WE simulation with the updated trajectory weights for an additional 100 WE iterations to ensure that the average rate constant remained at a steady value. Reweighting using the WESS procedure was performed using 75% of the preceding simulation data.

Dynamics of the WE simulations were propagated using the AMBER 2018 dynamics engine.[32] Since the WE strategy requires the use of stochastic dynamics–ensuring that the dynamics of replicated trajectories diverge–we used a stochastic thermostat (i.e. a weak Langevin thermostat with a collision frequency of 0.001 ps^{-1}) to maintain a temperature of 20 °C. A Monte Carlo barostat with a coupling constant of 1 ps^{-1} was used to maintain a pressure of 1 atm.

State definitions

For all analysis, definitions of key states were defined as follows (see also Figure S2). The unassociated reactants state was defined as a minimum separation distance between azide and cation of >10 Å. The target product state was defined as a minimum separation distance <1.6 Å (as determined by RI-MP2 geometry optimizations of the products, see Figure S1). In addition, an ion-pair intermediate state was observed at minimum azide-cation separation distances between 5 Å to 2.25 Å.

Calculation of rate constants

Rate constants are reported as averages based on five WE simulations along with 95% credibility regions using a Bayesian bootstrapping approach.[36]

For each WE simulation, unimolecular rate constants k_{AB} (i.e., k_{-1} and k_2) for a transition from state A to state B were calculated using the Hill relation as follows:

$$k_{AB} = \frac{1}{MFPT(A \rightarrow B)} = \frac{Flux(A \rightarrow B; SS)}{p_A} [s^{-1}] \quad (1)$$

where the inverse of the mean first passage time $MFPT(A \rightarrow B)$ for the A to B non-equilibrium steady state is equal to the conditional steady-state probability flux $Flux(A \rightarrow B; SS)$ into the target state B for trajectories most recently in the initial state A. To focus on the unidirectional flux in the forward direction of the reaction, the $Flux(A \rightarrow B; SS)$ was normalized by the steady-state population p_A in state A (i.e. sum of statistical weights of trajectories most recently in state A). Both the $Flux(A \rightarrow B; SS)$ and p_A were calculated as running averages over the WE simulation.

For bimolecular rate constants (i.e., k_{rxn} and k_{-1}), the conditional flux was divided by the effective concentration of reactants C_0 to yield rate constants in units of $M^{-1}s^{-1}$.

$$bimolecular\ k_{AB} = \frac{Flux(A \rightarrow B; SS)}{p_A} \left(\frac{1}{C_0} \right) [M^{-1} s^{-1}] \tag{2}$$

Effective concentrations C_0 for each simulation system (3.70 mM, 3.75 mM and 3.69 mM for the 4-OCH_3-T^+, T^+, and 4-CF_3-T^+ cations, respectively) were calculated as $\frac{1}{N_A V}$ where N_A is Avogadro's number and V is the volume of the simulation box.

Calculation of percent productive collisions

Percent productive collisions for each reaction are reported as averages based on five WE simulations along with 95% credibility regions using a Bayesian bootstrapping approach.[36] These percentages were calculated according to the following equation.

$$\%\ productive\ collisions = \frac{Flux(reactants \rightarrow products | SS)}{Flux(reactants \rightarrow intermediate | SS)} \times 100\% \tag{3}$$

where the numerator is the steady-state flux from the reactants to the products state and the denominator is the steady-state flux from the reactants to an ion-pair contact state which is defined as a distance of <5 Å between and two atoms of the reactant molecules.

Calculation of addition-site ratios

Ratios of azide addition at various sites on the cation were calculated directly from successful reaction pathway ensembles and reported relative to the addition site with the highest amount of flux. Since each of the three para-positioned addition sites for the T^+ cation are symmetrically equivalent, the calculated ratios for those cations is the average of the total flux corresponding to addition at all three of these sites (likewise for the symmetrically-equivalent sites of the 4-OCH_3-T^+ and 4-CF_3-T^+ cations).

Clustering of pathways into distinct classes

To cluster reaction pathways into distinct classes, we applied our recently developed Linguistics Pathway Analysis of Trajectories with Hierarchical clustering (LPATH) method.[37] This method involves the three steps detailed below.

In step 1, we discretized each pathway by assigning a state label to the configuration after each resampling time interval τ of 0.5 ps. Here, the state label is the ID of the phenyl-ring carbon atom that is nearest to the nitrogen of the azide anion.

In step 2, we quantified the similarity of each pair of discretized path sequences using a modified version of the Gestalt Pattern Matching algorithm,[38] which is used in computational linguistics for comparing text strings of varying lengths in the detection of plagiarism. Using this algorithm, a distance between a pair of text strings was calculated using the following equation, which contains a correction factor in the denominator to account for pairwise pathway comparisons in which the pathway lengths are different from each other.

$$distance = 1 - \left(\frac{2 \times length\left(longest\ common\ subsequence_{AB}\right)}{(length_A + length_B) - \left(\frac{|length_A - length_B|}{2}\right)} \right) \tag{4}$$

Finally, in step 3, we clustered the discretized path sequences into distinct pathway classes using the pairwise distances and a combination of a hierarchical agglomerative clustering algorithm and the Ward linkage method, which minimizes the variance within a given

cluster.[39] Based on the resulting tree diagram (dendrogram) of clusters, we identified distinct classes of pathways by positioning a horizontal line that maximizes the distance separation between nodes in the dendrogram (Figure S4).

Results and Discussion

To enable direct simulations of complete pathways for each of the three azide-clock reactions, we applied the weighted ensemble (WE) strategy with a hybrid QM/MM model in which the reactants were treated using quantum mechanics and the acetonitrile/water explicit solvent using molecular mechanics. For each reaction, five independent WE simulations were run, generating, in aggregate, thousands of pathways (1805, 8704, and 19,220 pathways for the 4-OCH_3-T^+, T^+, and 4-CF_3-T^+ cations, respectively). Each WE simulation was completed within 4 days using 280 Intel Xeon 2.6 GHz CPU cores in parallel.

Simulations reveal a two-step mechanism

Our simulations of each azide-clock reaction reveal a two-step reaction mechanism. In the first step, diffusional collision of the azide anion and trityl cation forms an ion-pair intermediate in which the anion and cation are within van der Waals contact, but not forming the target N-C bond with the central carbon of the cation. In the second step, the ion-pair intermediate "activates", rearranging to the product. For each reaction, our calculated rate constant k_{rxn} for the overall reaction is within error of that measured by laser flash-photolysis experiments, reproducing the trend in reactivity of 4-OCH_3-T^+ < T^+ < 4-CF_3-T^+ (Figure 2A). Consistent with this trend, the percentage of productive collisions (collisions of reactants that successfully reached the product state) is lowest for the least reactive 4-OCH_3-T^+ cation (0.41 ± [0.21, 0.59]%) and higher for the more reactive 4-CF_3-T^+ (16.07 ± [5.39, 34.69]%) and T^+ cations (10.22 ± [4.96, 16.25]%) (Figure 2B). The substantially larger uncertainty in the percent productive collisions for azide addition to the 4-CF_3-T^+ is due to

one of the five WE simulations yielding a particularly high percentage.

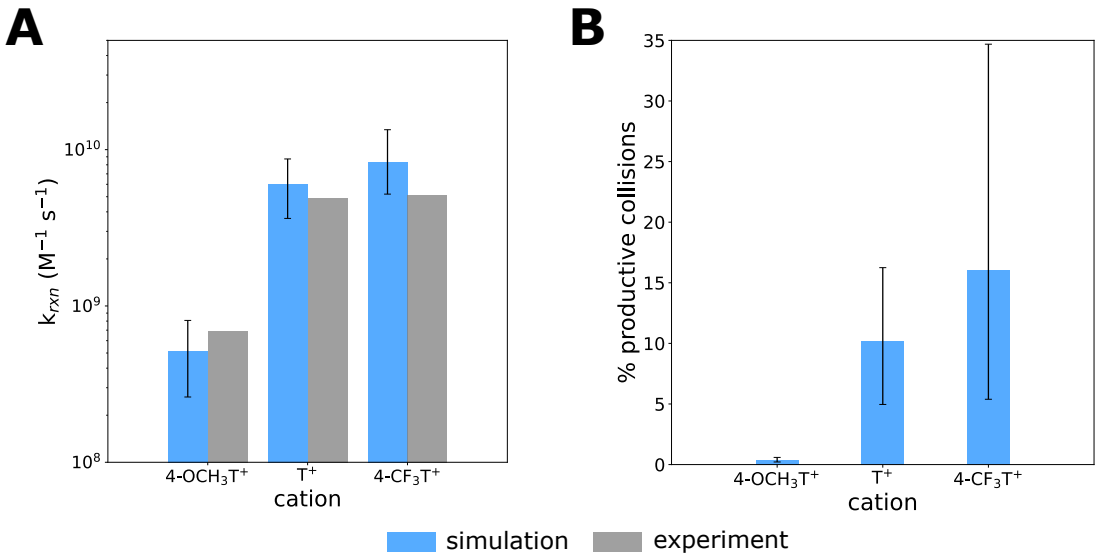

Figure 2: Direct simulations of reactions yield rate constants within error of experiment. (A) Comparison of reaction rate constants calculated from simulation (blue) to those measured by laser-flash photolysis experiments[21] (gray). (B) Percent productive collisions calculated from simulation follow the trend in reaction rate constants for azide addition to each of the three cations. Reaction rate constants and percent productive collisions are averages from five independent WE simulations with uncertainties that each represent 95% credibility regions, as estimated using a Bayesian bootstrap approach (see Methods).

Reactions are activation-limited

To determine the rate-limiting step for each reaction, we directly calculated rate constants for each individual step of the reaction (Figure 3A) from our simulations, i.e. k_1 for formation of an ion-pair intermediate, k_{-1} for the dissociation of the intermediate, and k_2 for the rearrangement of the intermediate to product (see Methods for state definitions). While the k_1 values for all three reactions are essentially identical, the corresponding k_2 values follow the trend in the overall reaction rate constant k_{rxn} (Figure 3B, Table S1) Due to the much more rapid dissociation of the ion-pair intermediate relative to rearrangement of the intermediate to the product ($k_{-1} >> k_2$), the k_2 step is rate-limiting for all three reactions. Thus, all three of the reactions in the present study are activation-limited rather than diffusion-limited. In contrast, previous fitting of data from flash photolysis experiments to a two-step mechanism

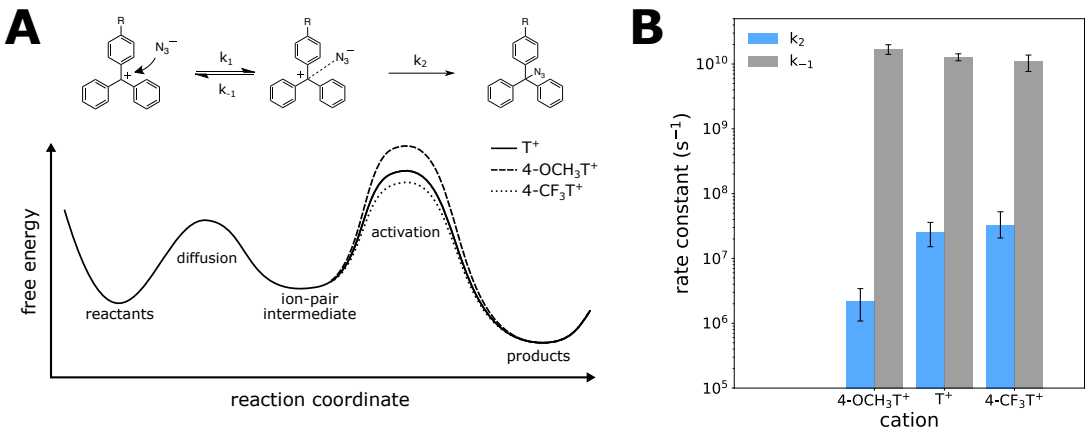

Figure 3: Evidence of activation-controlled reactions. (A) Illustration of the azide clock reaction as a two-step reaction involving the formation of an ion-pair intermediate and a rate-limiting rearrangement of this intermediate to the product. (B) Direct calculations of rate constants for the dissociation of the ion-pair intermediate (k_{-1}) and rearrangement of the intermediate to the product (k_2) reveal that for all three cations, $k_{-1} >> k_2$, which indicates that the k_2 step is rate-limiting (activation-controlled). The trend in k_2 values explains the trend in the overall reaction rate constant k_{rxn} whereas k_1 values are the same for all three reactions.

suggested that the diffusion step to be rate-limiting for the less reactive cations (e.g., 4-OCH_3-T^+, T^+) and the activation step to be rate-limiting for the more reactive cations (e.g., 4-CF_3-T^+).[21] However, these experiments monitored only the lifetime of the cation reactant, underscoring the value of using simulations to directly calculate rate constants for individual steps of a chemical reaction.

The activation step involves "propeller crawling"

A hallmark of WE simulations is not only the direct calculation of rates, but also ensembles of continuous, unbiased pathways between the initial and target states (here, the reactants and products, respectively). While each WE simulation consists of many pathways with shared history (common trajectory segments), we were able to obtain two distinct classes of pathways for each reaction and their corresponding probabilities by clustering the pathways from all five WE simulations of that reaction based on the sequence of configurations visited (see Methods).

Our pathway analysis reveals that the activation step involves the azide anion "crawling" among the cation's three propellers (phenyl rings) on its journey to the cation's central carbon. Further, the trend in the rate constants k_2 for the activation step of 4-OCH_3-T^+ < T^+ < 4-CF_3-T^+ (Figure 3B) is due to the 4-OCH_3-T^+ reaction involving a greater range of "propeller crawling" relative to the T^+ and 4-CF_3-T^+ reactions, i.e. the azide anion is more likely to contact the ortho- (X) and para-positioned (S) carbons of the cation's propellers before reaching the target central carbon (T) of the cation to form the target N-C bond (Figure 4). This greater probability is likely due to the reduced partial positive charge of the 4-OCH_3-T^+ cation's central carbon relative to T^+ and 4-CF_3-T^+. Movies of this propeller-crawling mechanism are provided for the most probable pathway of each reaction (Movies S1-S3).

A closer examination of the propeller-crawling mechanism reveals interesting features for each reaction. For the unsubstituted cation, the most probable pathway class involves a greater range of propeller-crawling compared to the minor pathway class with a higher probability of the azide anion first contacting an ortho-positioned carbon of a phenyl ring to form the ion-pair intermediate followed by contacts with both the ortho and para positions of phenyl rings while crawling to the central carbon of the cation. For the 4-OCH_3-T^+ cation, the minor pathway class (class 2) involves a high probability of the azide anion first contacting the carbon that is directly bound to the methoxy substituent; this carbon is more positively charged compared to the other carbons in the phenyl ring due to resonance effects of the electron-donating methoxy substituent. For the 4-CF_3-T^+ reaction, azide addition does not occur at the para carbon, which is directly bound to the partial negatively-charged carbon of the trifluoro substituent.

Detection of azide addition at propeller sites.

Our reaction simulations captured azide addition at not just the target central carbon of each cation, but also at the 2- or 4- positions of the cation's phenyl-ring "propellers" (Figure

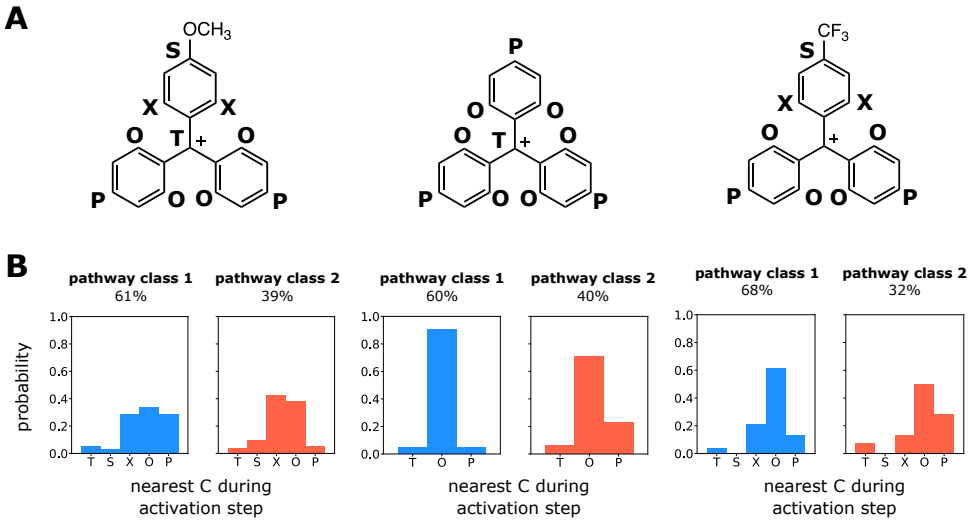

Figure 4: Distinct classes of pathways reveal differences in mechanism among the three reactions. (A) Symmetry-adapted labels for carbons of the phenyl rings that are contacted by the azide anion during its crawling to the target central carbon (T) of the cation: S is the carbon bonded to the substituent of a substituted phenyl ring, X is the ortho-positioned carbon that is bonded to the substituent, O is the ortho-positioned carbon of an unsubstituted phenyl ring, and P is the para-positioned carbon of an unsubstituted ring. (B) Probability distributions as a function of carbon atoms contacted by the azide anion at any time during the activation step, as considered every 0.25 ps. The azide displays more even crawling across all carbons considered in the 4-OCH_3-T^+ cation.

S5). These propeller sites of addition would be expected based on the delocalization of positive charge in resonance models of the cations; however, the corresponding addition species result in nonaromatic rings and would therefore be too transient for detection in the previously reported flash photolysis experiments.[21] These species likely rearrange to the expected product (involving azide addition to the cation's central carbon) to preserve aromaticity of the phenyl rings. While such rearrangements are beyond the target states of our simulations, their occurrence would further contribute to the activation step being rate-limiting for the reactions.

Outlook for simulating chemical reactions

We have demonstrated the power of the WE strategy in enabling the direct molecular simulation of pathways and rate constants for chemical reactions using hybrid QM/MM models.

Such simulations would be feasible for chemical reactions on the µs timescale with < 50 reacting atoms in the QM region using a semi-empirical level of theory. The reactions can be either in solution or enzyme-catalyzed. The future looks bright for simulating chemical reactions with more complex reactants and/or higher levels of QM theory, including the use of GPU-accelerated dynamics engines that can greatly extend the timescales of hybrid QM/MM simulations (NAMD)[40] and deep-learning potentials approaching the "gold standard" of coupled-cluster accuracy at the cost of a classical force field (ANI-1ccx).[41]

Conclusions

We report direct molecular dynamics simulations of azide-clock reactions in explicit solvent involving the addition of an azide anion to each of three different cations, 4-OCH_3-T^+, T^+, and 4-CF_3-T^+. These simulations were enabled by applying the WE path sampling strategy with hybrid QM/MM model. Our simulations generated thousands of continuous pathways for each reaction, yielding reaction rate constants that are within error of experiment. Results revealed that each reaction involves a two-step mechanism in which the first step involves diffusional collision of reactants to form an ion-pair intermediate and the second step involves "activation", or rearrangement of the intermediate to the product. In contrast to previous assumptions, all three reactions are activation-controlled rather than diffusion-controlled. Based on our simulated ensemble of reaction pathways, the slower activation step of the 4-OCH_3-T^+ reaction relative to the T^+ and 4-CF_3-T^+ reactions is due to the ability of the azide to "crawl" along a greater range of the cation's three propellers (phenyl rings). Our work not only provides the most detailed views to date of azide-clock reactions, but also presents a rare-events sampling method that enables simulation of pathways and kinetics for many chemical reactions, either in solution or enzyme-catalyzed.

Acknowledgements

Computational resources were provided by NSF XSEDE allocation TG-MCB100109 to L.T.C. and the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the H2P cluster, which is supported by NSF award number OAC-2117681. We would like to thank Adrian Elcock (U Iowa), Geoff Hutchison, Dean Tantillo (UC Davis), and Kennie Merz (MSU) for helpful discussions.

References

(1) Tantillo, D. J. Dynamic Effects on Organic Reactivity—Pathways to (and from) Discomfort. *J of Physical Organic Chem* **2021**, *34*, e4202.

(2) Feng, Z.; Tantillo, D. J. Dynamic Effects on Migratory Aptitudes in Carbocation Reactions. *J. Am. Chem. Soc.* **2021**, *143*, 1088–1097.

(3) Fu, Y.; Bernasconi, L.; Liu, P. Ab Initio Molecular Dynamics Simulations of the S_N1/S_N2 Mechanistic Continuum in Glycosylation Reactions. *J. Am. Chem. Soc.* **2021**, *143*, 1577–1589.

(4) Van Der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708–2728.

(5) Van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A Novel Path Sampling Method for the Calculation of Rate Constants. *The Journal of Chemical Physics* **2003**, *118*, 7762–7774.

(6) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition Path Sampling and the Calculation of Rate Constants. *The Journal of Chemical Physics* **1998**, *108*, 1964–1977.

(7) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophys. J.* **1996**, *70*, 97–110.

(8) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.

(9) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-Sampling Strategies for Simulating Rare Events in Biomolecular Systems. *Current Opinion in Structural Biology* **2017**, *43*, 88–94.

(10) Basner, J. E.; Schwartz, S. D. How Enzyme Dynamics Helps Catalyze a Reaction in Atomic Detail: A Transition Path Sampling Study. **2005**, *127*, 13822–13831.

(11) Crehuet, R.; Field, M. J. A Transition Path Sampling Study of the Reaction Catalyzed by the Enzyme Chorismate Mutase. **2007**, *111*, 5708–5718.

(12) Moqadam, M.; Lervik, A.; Riccardi, E.; Venkatraman, V.; Alsberg, B. K.; Van Erp, T. S. Local Initiation Conditions for Water Autoionization. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*.

(13) Bonk, B. M.; Weis, J. W.; Tidor, B. Machine Learning Identifies Chemical Characteristics That Promote Enzyme Catalysis. *J. Am. Chem. Soc.* **2019**, *141*, 4108–4118.

(14) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chem. Sci.* **2019**, *10*, 2360–2372.

(15) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein-Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered P53 Peptide. *J. Phys. Chem. Lett.* **2016**, *7*, 3440–3445.

(16) Brossard, E. E.; Corcelli, S. A. Molecular Mechanism of Ligand Binding to the Minor Groove of DNA. *J. Phys. Chem. Lett.* **2023**, *14*, 4583–4590.

(17) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J. Am. Chem. Soc.* **2019**, *141*, 6519–6526.

(18) Sztain, T. et al. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.* **2021**, *13*, 963–968.

(19) Bogetti, X.; Bogetti, A.; Casto, J.; Rule, G.; Chong, L.; Saxena, S. Direct Observation of Negative Cooperativity in a Detoxification Enzyme at the Atomic Level by Electron Paramagnetic Resonance Spectroscopy and Simulation. *Protein Science* **2023**, *32*, e4770.

(20) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. *Biophysical Journal* **2021**, *120*, 158–167.

(21) McClelland, R. A.; Kanagasabapathy, V. M.; Banait, N. S.; Steenken, S. Reactivities of Triarylmethyl and Diarylmethyl Cations with Azide Ion Investigated by Laser Flash Photolysis. Diffusion-controlled Reactions. *J. Am. Chem. Soc.* **1991**, *113*, 1009–1014.

(22) McClelland, R. A.; Kanagasabapathy, V. M.; Banait, N. S.; Steenken, S. Flash-Photolysis Generation and Reactivities of Triarylmethyl and Diarylmethyl Cations in Aqueous Solutions. *J. Am. Chem. Soc.* **1989**, *111*, 3966–3972.

(23) Bertozzi, C. R. A Decade of Bioorthogonal Chemistry. *Acc. Chem. Res.* **2011**, *44*, 651–653.

(24) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes

Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749–1760.

(25) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25*, 1157–1174.

(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(27) Nikitin, A. M.; Lyubartsev, A. P. New Six-site Acetonitrile Model for Simulations of Liquid Acetonitrile and Its Aqueous Mixtures. *J Comput Chem* **2007**, *28*, 2020–2026.

(28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(29) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of Approximate Integrals in Ab Initio Theory. An Application in MP2 Energy Calculations. *Chemical Physics Letters* **1993**, *208*, 359–363.

(30) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *The Journal of Chemical Physics* **2020**, *152*, 224108.

(31) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. P PACKMOL : A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J Comput Chem* **2009**, *30*, 2157–2164.

(32) Case, D. et al. *AMBER 2018*; 2018.

(33) Russo, J. D. et al. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble Simulations and Analysis of Longer-Timescale Applications. *J. Chem. Theory Comput.* **2022**, *18*, 638–649.

(34) Bogetti, A. T.; Leung, J. M. G.; Russo, J. D.; Zhang, S.; Thompson, J. P.; Saglam, A. S.; Ray, D.; Abraham, R. C.; Faeder, J. R.; Andricioaei, I.; Adelman, J. L.; Zwier, M. C.; LeBard, D. N.; Zuckerman, D. M.; Chong, L. T. A Suite of Advanced Tutorials for the WESTPA 2.0 Rare-Events Sampling Software [Article v2.0]. *LiveCoMS* **2022**, *5*.

(35) Bhatt, D.; Zhang, B. W.; Zuckerman, D. M. Steady-State Simulations Using Weighted Ensemble Path Sampling. *J. Chem. Phys.* **2010**, *133*, 014110.

(36) Mostofian, B.; Zuckerman, D. M. Statistical Uncertainty Analysis for Small-Sample, High Log-Variance Data: Cautions for Bootstrapping and Bayesian Bootstrapping. *J. Chem. Theory Comput.* **2019**, *15*, 3499–3509.

(37) Bogetti, A. T.; Leung, J. M. G.; Chong, L. T. LPATH: A Semiautomated Python Tool for Clustering Molecular Pathways. *Journal of Chemical Information and Modeling* **2023**, *63*, 7610–7616.

(38) Ratcliff, J. W.; Metzener, D. E. Pattern Matching: The Gestalt Approach. *Dr. Dobb's Journal* **1988**,

(39) Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963**, *58*, 236–244.

(40) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. NAMD Goes Quantum: An Integrative Suite for Hybrid Simulations. *Nat Methods* **2018**, *15*, 351–354.

(41) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat Commun* **2019**, *10*, 2903.