

Quick and Accurate Estimates of Mutation Effects on Transition-State Stabilization of Enzymes from Molecular Simulations with Restrained Transition States

Tucker Burgin, Jim Pfaendtner,* and David A. C. Beck*



Cite This: *J. Phys. Chem. B* 2022, 126, 9964–9970



Read Online

ACCESS |



Metrics & More

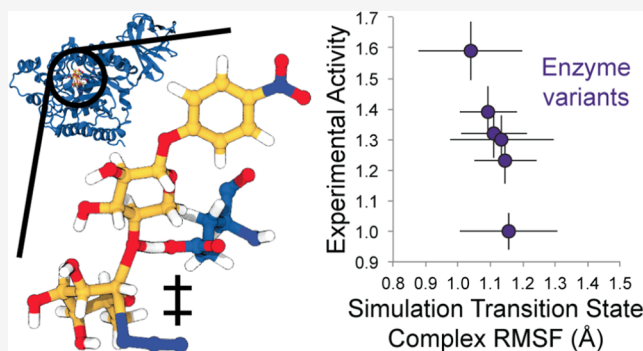


Article Recommendations



Supporting Information

ABSTRACT: Data science and machine learning are revolutionizing enzyme engineering; however, high-throughput simulations for screening large libraries of enzyme variants remain a challenge. Here, we present a novel but highly simple approach to comparing enzyme variants with fully atomistic classical molecular dynamics (MD) simulations on a tractable timescale. Our method greatly simplifies the problem by restricting sampling only to the reaction transition state, and we show that the resulting measurements of transition-state stability are well correlated with experimental activity measurements across two highly distinct enzymes, even for mutations with effects too small to resolve with free energy methods. This method will enable atomistic simulations to achieve sampling coverage for enzyme variant prescreening and machine learning model training on a scale that was previously not possible.



INTRODUCTION

Many of the most promising approaches in enzyme engineering from the past several years rely on the collection and analysis of large libraries of enzyme variants, either to directly search for improved variants as in directed evolution¹ or more recently to train machine learning methods.² This trend is not surprising given the accelerating use of data-driven science and engineering across scientific and technological disciplines, and its promise to enable the next generation of biotechnological advances is already being borne out in recent examples.^{3–5}

Methods in computational protein engineering have also begun to make use of the tools of data science, producing key breakthroughs such as AlphaFold and RoseTTAFold, neural network approaches to protein structure prediction with dramatically improved performance over predicate methods.^{6,7} However, to date, neural networks for protein engineering have been mostly trained on unlabeled sequences or experimental results, as opposed to molecular dynamics (MD) simulations. Despite ever-increasing computational power, it usually remains infeasible to perform an atomistic simulation for every one of the thousands or even millions of data points that are often necessary to train a useful neural network. This problem is further exacerbated for studies of rare events such as chemical reactions, which necessitate the use of expensive enhanced sampling methods and for especially large proteins or protein complexes.

However, MD simulations using physics-based potentials remain of high interest for certain types of problems such as

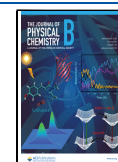
the assessment of the relative activity of variants of a given enzyme. Such data are typically collected using benchtop-directed evolution or deep mutational scanning studies; however, benchtop studies of both types are expensive, laborious, and require the development or availability of assays for the desired enzyme activity, which depending on the context is not always simple.^{8,9} Furthermore, benchtop experiments can match or correlate sequences or mutations with activity but produce no information about the local structural changes or biophysical effects underlying that relationship. This motivates the development of simulation-based methods to assess large numbers of sequence–function relationships at the molecular level directly.

There have been some attempts to develop new methodologies to address this problem. For example, the popular protein engineering suite Rosetta features tools that can be used to estimate the relative fitness of a particular protein sequence by eschewing MD simulations entirely in favor of discrete backbone torsion minimization steps and side-chain repacking.^{10,11} Though computationally efficient, methods that do not include MD simulations run the risk of missing key

Received: July 7, 2022

Revised: November 8, 2022

Published: November 22, 2022



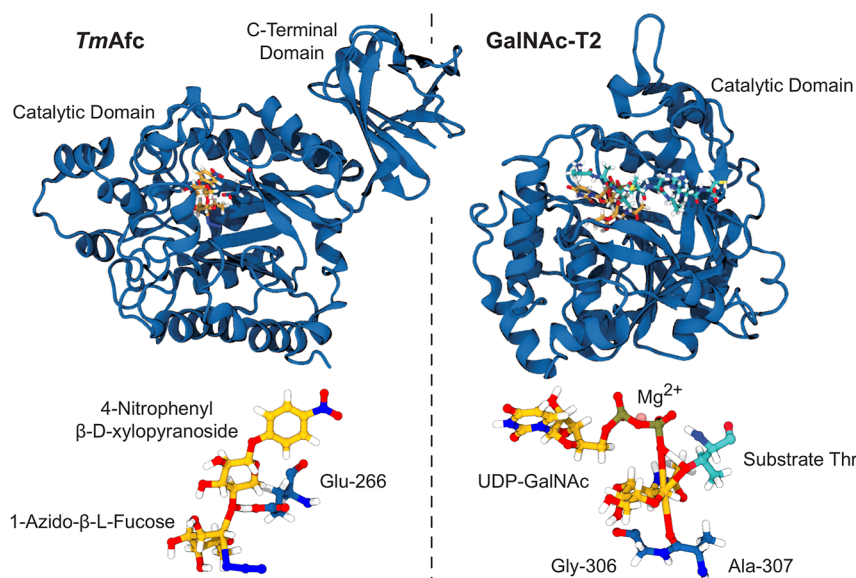


Figure 1. Models of *TmAfc* D224G (top left) and GalNAc-T2 (top right) used in this study, shown bound to their respective transition-states. At the bottom are shown the transition-state models from these enzymes enlarged, featuring stretched bonds that remain stable over the course of the simulations. Dark-blue carbon atoms correspond to the enzymes, yellow to the nonprotein substrates, and teal to the GalNAc-T2 oligopeptide substrate.

aspects of protein function not captured in backbone and side-chain reorganizations alone, and consequently, these methods feature significant error. More accurate simulation-based approaches also exist, such as CADEE, which uses full ensemble simulations combined with empirical valence bond calculations to estimate the free energy of reactions.¹² However, the change in the free energy of the reaction between enzyme variants often falls well below the sensitivity of free energy methods (in the range of 1–2 kcal/mol),¹³ making free energy approaches such as the one used in CADEE unsuited to make accurate predictions in such cases.

Here, we propose a new, simple approach to estimate the relative stabilization of a given reaction transition state in enzyme variants with full atomistic simulations using classical force fields and without any enhanced sampling scheme. Rather than directly measuring the free energy of the reaction, our method evaluates the dynamic stability of a restrained transition-state complex. It has previously been shown that the interaction energy between enzyme variants and bound transition-state analogues can be predictive of relative enzymatic activity.¹⁴ We aim to extend this work by applying it to transition states directly (not analogues) and performing full-ensemble simulations to capture long-range mutation effects.

Even for mutations far from the active site or with changes in activation free energy well below the sensitivity of free energy methods, we show that transition-state complex stability has an excellent correlation with experimental measurements of enzyme activity across two dissimilar enzymes with single-step reaction mechanisms: the engineered glycosynthase *Thermotoga maritima* α -L-fucosidase (*TmAfc*) D224G and human *N*-acetylgalactosaminyl transferase-2 (GalNAc-T2), both shown with transition-state models in Figure 1. For enzymes whose reactivity is dominated by transition-state stabilization, this method can be used to rapidly screen large libraries of variants at a modest computational expense or to generate high-quality training data for machine learning models.

METHODS

Leaving out the process of constructing a suitable molecular model of the enzyme common to all enzyme modeling approaches,¹⁵ the steps for our method are as follows:

1. Obtain or generate a model of the enzyme–transition-state complex and apply restraints to the model such that the transition state remains stable over the course of the simulation;
2. Apply the desired mutation(s) and perform classical MD simulations of the enzyme–transition-state complex; and
3. Measure the root-mean-square fluctuation of the reaction substrate and key neighboring residues over the course of the simulation(s).

This workflow is depicted visually in Figure 2. We will describe the methodology we used for each of these steps in turn; however, because the method is so simple, it is likely that variations on this basic framework would also be successful.

Modeling and Restraining the Transition State. Methodologies for obtaining transition-state models are outside the scope of this work, but a suitable review is available from Paul et al.¹⁶ Such methods can be broadly divided into two categories: those with predetermined collective variables (CVs) and those without. Doron et al. offer a comparison of these categories.¹⁷

In this case, we obtained transition-state models from previously published work for both models. For *TmAfc* D224G, we discovered and validated the transition-state model with transition path sampling using a workflow based on work by Bolhuis et al.¹⁸ and Mullen et al.¹⁹ that has since been published as the software tool ATESA (github.com/team-mayes/atesa). This model is based on PDB ID: 2ZXD²⁰ and did not involve any predetermined CVs. For more details regarding the method used to obtain this transition-state model, see our previous work²¹ or consult the documentation for ATESA at atesa.readthedocs.io. For our model of the GalNAc-T2, we followed the methodology laid out in ref 22 based on PDB IDs: 2FFU²³ and 4D0Z.²⁴ We excluded the

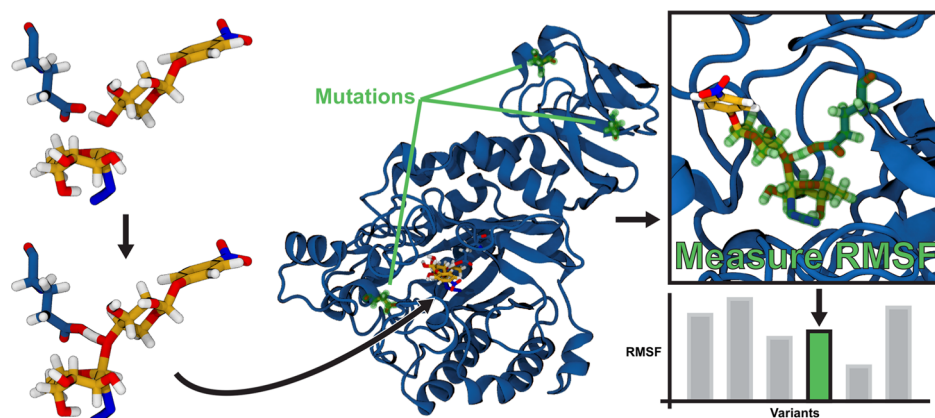


Figure 2. Process for creating and scoring a variant. Starting at left, bonds are added to the transition-state complex model and stretched bonds have their equilibrium lengths modified so that the transition-state model will be stable during classical simulation. The desired mutations are then applied (at the center, highlighted in green), and then the simulations are run. Finally, the resulting average RMSF of the transition-state complex (at right) is measured to produce a relative score for the variant to compare to those from competing variants.

noncatalytic lectin domain, which is not believed to interact directly with the substrates during the reaction.²³ For the enzyme-bound transition state, we relied on a published quantum mechanics calculation from Gómez et al. 2014,²⁵ which was obtained by scanning along predetermined CVs followed by transition-state optimization. Further details for this model are available in the [Supporting Information](#).

Models were prepared using AmberTools21.²⁶ The models were prepared with the ff19SB force field²⁷ and OPC (*TmAfc* D224G) or OPC3 (GalNAc-T2) explicit water model²⁸ using the AmberTools tleap program with the bond topology of the prereaction state. Additional details for this process are available in the [Supporting Information](#). In order to keep the transition states stable over the course of the simulations, we set the equilibrium length of all the reactive bonds, that is, bonds that either form or break over the course of the reaction, to their transition state lengths. We also applied additional 100 kcal/mol Å² restraints to the positions of the atoms involved in the reactive bonds, as well as to the ring atoms of the chiral pyranose sugars in each model (the fucose in *TmAfc* D224G and the *N*-acetylgalactosamine in GalNAc-T2) in order to stabilize their puckers. Somewhat heavier or lighter restraint weights may be acceptable as long as the transition-state structure remains stable; we observed that weights one order of magnitude larger or smaller often became unstable and crashed. The models were then energy-minimized, heated in *NVT*, and equilibrated in *NPT* in accordance with the methodology in the [Supporting Information](#).

Applying Mutations and Setting Up Individual Simulations. Mutations to the enzymes were made using PyRosetta²⁹ to pack the side chains of mutated residues (and those nearby) in a reasonable initial conformation. For each mutant, the Amber parameter/topology file was regenerated using the initial coordinates from the equilibrated model (plus the changes made by PyRosetta). Predictions of the protonation states of titratable residues were handled using PROPKA3.^{30,31} Additional details are available in the [Supporting Information](#).

We used a standard 2 fs time step for the GalNAc-T2 simulations. Before running the *TmAfc* D224G simulations, we applied hydrogen mass repartitioning (HMR)³² and set the simulation time step to 4 fs in order to demonstrate the robustness of our method to longer time steps.

Assessment of Transition-State Complex Stability from Simulations.

After the simulations, we measured root-mean-square fluctuations (RMSFs) at and around the reactive components in order to measure the relative degree to which the transition-state complex is stabilized or destabilized in each variant. Because the essential components of the reaction transition state are restrained within the active site, while the rest of the enzyme–substrate complex is free to move around them, each statistically independent frame of the simulation can be interpreted as a sample from the equilibrium distribution of the transition state–enzyme complexes. The RMSF of the transition state should therefore converge to a value interpretable as the weighted average “unfavorability” of the specific transition-state structure of interest among the equilibrium distribution of simulation states.

There is no single obvious choice for the atoms to include in the measurement of the RMSF, and the best selection for a given transition-state complex may vary. For this work, we selected the reactive components of the substrate molecules as well as the reactive residues of the enzymes and the residues immediately adjacent to them on either side. Exact definitions can be found in the [Supporting Information](#).

Simulations of *TmAfc* D224G variants were run for roughly 87 ns each, whereas the more efficient GalNAc-T2 simulations were run for roughly 115 ns each. These simulation times were achievable in roughly 6 h each on a single NVIDIA A40 GPU. For both systems, we analyzed the impact of simulation time on results, as shown in [Figure 4](#) in the [Results and Discussion](#) section.

RESULTS AND DISCUSSION

We took previously published experimental values for the relative enzymatic activity of the two enzyme systems from Agrawal et al.⁸ (*TmAfc* D224G) and Kightlinger et al.³³ (GalNAc-T2), respectively. The former paper describes a directed evolution study where the enzyme was modified directly and assayed for improved rates of product formation. By contrast, the latter paper describes the effect of modifications to the sequence of a substrate oligopeptide, leaving the enzyme itself unmodified and reducing the relative activity relative to the native substrate in most cases.

The calculated RMSF values of the transition-state complexes were well correlated with experimental values over

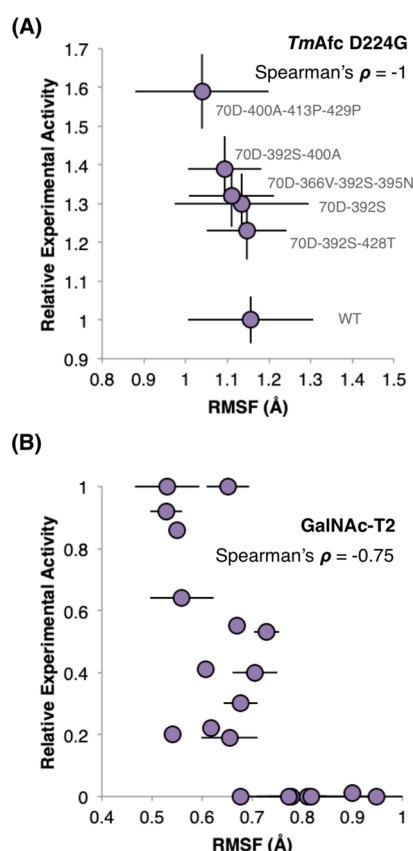


Figure 3. Correlations between experimentally determined relative enzyme activities and fluctuation of transition-state complexes (RMSF) from simulations for (A) *TmAfc* D224G and (B) GalNAc-T2. Error bars for experimental values in (A) are approximations of the standard deviation reported in Agrawal et al.⁸ Error bars for RMSF values in both plots represent standard deviations across independent simulations of each variant [nine simulations each in (A) and three each in (B)]. No error values were available for the experimental values for (B), which are taken from Kightlinger et al.³³ Data labels in (A) indicate the sequence of mutations characterizing each variant, where WT indicates that no further mutations beyond D224G were made to the canonical *TmAfc* sequence taken from PDB ID: 1HL9.³⁴ “Experimental activity” values correspond to hydrolysis rescue specific activity in the presence of aqueous azide (a proxy for glycosynthase activity) for *TmAfc* D224G⁸ and substrate specificity for GalNAc-T2.³³ A version of panel (B) with labels indicating the identity of each point is available in the Supporting Information.

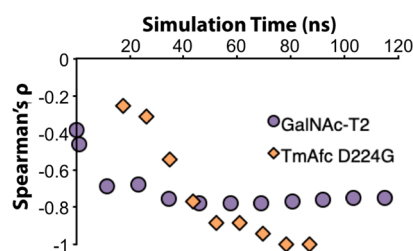


Figure 4. Relationship between simulation time and Spearman's ρ for both models.

the range of available data for both enzymes, as shown in Figure 3. The negative correlation is entirely expected: an increase in RMSF indicates that the transition-state complex is less stable, so the enzyme is expected to be less active and vice versa. Although transition-state theory predicts that the

energetic stability of the transition-state complex should correlate with reactivity, to our knowledge, this is the first time that the dynamic fluctuations of the transition-state complex itself have been shown to directly correlate as well. This should not necessarily be interpreted as evidence for (or against) the “dynamic” model of enzyme function, as discussed by Kamerlin and Warshel;³⁵ instead, we interpret the “dynamic fluctuations” of the restrained transition-state model as a simplified proxy for measurement of the free energy of the transition state. In other words, we contend that the same molecular mechanisms that affect the dynamic stability of the restrained transition state complex in our models may also have the effect of energetically stabilizing the transition state complex in the actual reaction context.

The differences in free energies of activation associated with these mutations can be quite small. For example, the difference in activation energies between relative activity values of 1 and 1.39 in the *TmAfc* D224G study is only about 0.2 kcal/mol,⁸ well below the sensitivities of free energy methods broadly, which are generally on the order of 1–2 kcal/mol.¹³ The question of why our method achieves superior sensitivity is worth considering. Although free energy methods are highly diverse, typically they require converged sampling of multiple different states along a predefined order parameter or collective variable, and always, they can only measure relative free energy between states. Furthermore, all degrees of freedom other than the one(s) being sampled along are left unrestrained, and sampling along these dimensions must be converged. By contrast, our method sidesteps these issues by focusing on only a single state (the transition state) and measuring an exact (not relative) value that can be directly observed during MD (the transition-state complex RMSF), which eliminates some sources of error (while introducing others, discussed below). Because we have no need to measure free energy directly, our method is also able to feature more restraints across the entire substrate structure, which in turn results in faster convergence.

Besides the small differences in free energy, another complication for comparing the *TmAfc* D224G variants reported in Agrawal et al.⁸ is that all of the differences between them are far from the active site in the C-terminal domain (Figure 1). The purpose of this domain is not well understood, and there is no obvious reason we can see that mutations to this region should arise so frequently during directed evolution; nevertheless, our approach was successful in distinguishing the allosteric effects of these distal mutations on the dynamics of the transition-state complex. This result underscores the importance of capturing the long-range effects of mutations in making successful predictions of relative fitness, a task for which classical MD simulations are well suited.

Although the shape of the relationship between RMSF and enzyme activity in Figure 3 appears roughly linear over the range of values tested here (with the exception of the wide range of values with approximately zero activity in Figure 3B since further transition-state destabilization cannot reduce activity below zero), physically, this cannot continue forever: RMSF can only become so low. Hypothetical enzymes with exceptionally stable transition-state complexes would have to experience diminishing returns on decreases in RMSF as their activity further improves, which would be reflected in an upward inflection (positive second derivative) of the relationship between them when plotted as in Figure 3. Therefore, the sensitivity of our method is expected to diminish as the enzyme

becomes highly optimized. However, we suggest that enzymes approaching this limit are likely to be more easily further improved by alternate mechanisms (e.g., changes to substrate binding, product release, etc.) instead of by further transition-state stabilization, in which case the method we present here would not be an appropriate tool to measure the relevant effects.

It is important to recognize that because this approach only involves the simulation of an approximation of the transition state and not of the reactant or product states, it cannot be used to make direct inferences about relative changes in activation energy or overall reaction energy. The RMSF values computed here only relate to the efficacy of the enzyme in stabilizing the transition-state complex itself. Therefore, even in the limit of an ideal model and perfect sampling, our method is not robustly guaranteed to correlate with enzyme activity in all cases, and particularly not in enzymes where the rate of reaction is significantly influenced by nonreactive steps in the catalytic cycle, such as product release. We interpret the fact that this method nevertheless produces high-quality results as evidence that, at least in the tested examples, mutation effects on enzyme activity are dominated by changes in transition-state complex stabilization rather than in reactant-state destabilization. Nevertheless, this method must be validated with experimental results for each individual system, and accuracy could possibly be improved in some cases by repeating simulations in the reactant state as well and evaluating the relative change in stabilization between the two states. The fact that this method may not be uniformly applicable to all reactions of potential interest should not be interpreted as diminishing its potential for rapid screening of systems where it is appropriate.

Because this approach is designed to assess the relative stability of a specific transition-state complex, it cannot accurately estimate the effects of mutations that alter the transition-state geometry significantly relative to the enzyme variant for which the transition-state model was obtained. We believe that such mutations will be rare even among beneficial mutations, supported by the fact previous efforts to computationally engineer enzymes based on their affinity for a particular transition-state geometry have been successful.^{14,36} This concern could be addressed at some additional computational expense by performing committor analysis simulations on variants to ensure that the given transition state is still valid in each case.¹⁸

The relationship between simulation time and Spearman's rank correlation coefficient for each system is shown in Figure 4. Even very short simulation times are sufficient to achieve strong performance for the GalNAc-T2 model, whereas correlation for the *TmAfc* D224G model is much slower to converge. The most likely explanation for this difference is that the mutations made to the GalNAc-T2 model are immediately adjacent to the active site, whereas the mutations in *TmAfc* D224G are mostly in a different protein domain entirely. Simulations need to be run for long enough for the allosteric effects of the most distal mutations of interest to reach equilibrium, and this timescale must be determined for each enzyme of interest on a case-by-case basis. Although Spearman's rank correlation for *TmAfc* D224G reaches a more negative (i.e., more strongly correlated) value compared to GalNAc-T2, we caution against the interpretation that our method was more suited to *TmAfc* D224G. Instead, this is

likely an effect of the relatively lower number of *TmAfc* D224G variants available for testing compared to GalNAc-T2.

CONCLUSIONS

We present a novel method for assessing the relative effects of mutations on the transition-state stabilization of enzymes by direct measurement of the fluctuation of a restrained transition-state model within the active site. This method produces highly satisfactory results when compared to experimental reactivity data from two dissimilar enzyme systems, despite the associated changes in activation free energy falling below the magnitude that can be observed with simulation free energy methods in some cases. Combined with case-by-case model validation, we propose that this approach could be used for the prescreening of mutations and training of machine learning models to guide directed evolution with improved accuracy and reduced computational expense compared to existing methods. In combination with ultra-high-performance computing resources such as Anton, our method could be used to screen massive enzyme variant libraries with higher accuracy than has heretofore been possible.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.2c04802>.

Methodology used in generating models of the enzyme system and variants and further details on the exact simulation methods used (PDF)

AUTHOR INFORMATION

Corresponding Authors

Jim Pfaendtner – Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0001-6727-2957; Email: jpfaendt@uw.edu

David A. C. Beck – Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; Email: dacb@uw.edu

Author

Tucker Burgin – Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States; orcid.org/0000-0002-3714-3537

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.2c04802>

Author Contributions

T.B. was responsible for project ideation, model building, data collection and analysis, and primary writing of the manuscript. D.A.C.B. and J.P. provided critical guidance and feedback at all stages in addition to editing the manuscript. All authors have given approval for the final version of the manuscript.

Funding

This work was supported by National Science Foundation grant number 1934292. This work used the Extreme Science and Engineering Discovery Environment (XSEDE),³⁷ which is supported by National Science Foundation grant number ACI-1548562.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the University of Washington eScience Institute for useful feedback and support. They also thank Professor Jeffrey Gray for bringing the example case of GalNAc-T2 to their attention.

■ ABBREVIATIONS

MD, molecular dynamics; TmAfc, *Thermotoga maritima* α -L-fucosidase; GalNAc-T2, N-acetylgalactosaminyl transferase-2; CV, collective variable; QM, quantum mechanics; HMR, hydrogen mass repartitioning; RMSF, root-mean-square fluctuation

■ REFERENCES

- (1) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem., Int. Ed.* **2018**, *57*, 4143–4148.
- (2) Mazurenko, S.; Prokop, Z.; Damborsky, J. Machine Learning in Enzyme Engineering. *ACS Catal.* **2020**, *10*, 1210–1223.
- (3) Wu, Z.; Kan, S. B.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine Learning-Assisted Directed Protein Evolution with Combinatorial Libraries. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 8852–8858.
- (4) Bedbrook, C. N.; Yang, K. K.; Robinson, J. E.; Mackey, E. D.; Gradinaru, V.; Arnold, F. H. Machine Learning-Guided Channelrhodopsin Engineering Enables Minimally Invasive Optogenetics. *Nat. Methods* **2019**, *16*, 1176–1184.
- (5) Singh, N.; Malik, S.; Gupta, A.; Srivastava, K. R. Revolutionizing Enzyme Engineering through Artificial Intelligence and Machine Learning. *Emerg. Top. Life Sci.* **2021**, *5*, 113–125.
- (6) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
- (7) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871–876.
- (8) Agrawal, A.; Bandi, C. K.; Burgin, T.; Woo, Y.; Mayes, H. B.; Chundawat, S. P. S. Click-Chemistry-Based Free Azide versus Azido Sugar Detection Enables Rapid in Vivo Screening of Glycosynthase Activity. *ACS Chem. Biol.* **2021**, *16*, 2490–2501.
- (9) Mays, Z. J. S.; Mohan, K.; Trivedi, V. D.; Chappell, T. C.; Nair, N. U. Directed Evolution of *Anabaena Variabilis* Phenylalanine Ammonia-Lyase (PAL) Identifies Mutants with Enhanced Activities. *Chem. Commun.* **2020**, *56*, S255–S258.
- (10) Barlow, K. A.; O'Conchúir, S.; Thompson, S.; Suresh, P.; Lucas, J. E.; Heinonen, M.; Kortemme, T. Flex DdG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J. Phys. Chem. B* **2018**, *122*, 5389–5399.
- (11) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS Cent. Sci.* **2018**, *4*, 1708–1718.
- (12) Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCr* **2017**, *4*, 50–64.
- (13) Mey, A. S. J. S.; Allen, B. K.; Bruce Macdonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2020**, *2*, 18378.
- (14) Grisewood, M. J.; Gifford, N. P.; Pantazes, R. J.; Li, Y.; Cirino, P. C.; Janik, M. J.; Maranas, C. D. OptZyme: Computational Enzyme Redesign Using Transition State Analogues. *PLoS One* **2013**, *8*, No. e75358.
- (15) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J. Comp. Mol. Sci.* **2019**, *1*, 5957.
- (16) Paul, S.; Nair, N. N.; Vashisth, H. Phase Space and Collective Variable Based Simulation Methods for Studies of Rare Events. *Mol. Simul.* **2019**, *45*, 1273–1284.
- (17) Doron, D.; Kohen, A.; Nam, K.; Major, D. T. How Accurate Are Transition States from Simulations of Enzymatic Reactions? *J. Chem. Theory Comput.* **2014**, *10*, 1863–1871.
- (18) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (19) Mullen, R. G.; Shea, J. E.; Peters, B. Easy Transition Path Sampling Methods: Flexible-Length Aimless Shooting and Permutation Shooting. *J. Chem. Theory Comput.* **2015**, *11*, 2421–2428.
- (20) Wu, H.-J.; Ho, C.-W.; Ko, T.-P.; Popat, S. D.; Lin, C.-H.; Wang, A. H.-J. Structural Basis of α -Fucosidase Inhibition by Iminocyclitols with K_i Values in the Micro- to Picomolar Range. *Angew. Chem., Int. Ed.* **2010**, *49*, 337–340.
- (21) Burgin, T.; Mayes, H. B. Mechanism of Oligosaccharide Synthesis via a Mutant GH29 Fucosidase. *Chem. React. Eng.* **2019**, *4*, 402–409.
- (22) Mahajan, S. P.; Srinivasan, Y.; Labonte, J. W.; DeLisa, M. P.; Gray, J. J. Structural Basis for Peptide Substrate Specificities of Glycosyltransferase GalNAc-T2. *ACS Catal.* **2021**, *11*, 2977–2991.
- (23) Fritz, T. A.; Raman, J.; Tabak, L. A. Dynamic Association between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide α -N-Acetylgalactosaminyltransferase-2. *J. Biol. Chem.* **2006**, *281*, 8613–8619.
- (24) Lira-Navarrete, E.; Iglesias-Fernández, J.; Zandberg, W. F.; Compañón, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R. Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the PolypeptideN-Acetylgalactosaminyltransferase 2. *Angew. Chem., Int. Ed.* **2014**, *53*, 8206–8210.
- (25) Gómez, H.; Rojas, R.; Patel, D.; Tabak, L. A.; Lluch, J. M.; Masgrau, L. A Computational and Experimental Study of O-Glycosylation. Catalysis by Human UDP-GalNAc Polypeptide:GalNAc Transferase-T2. *Org. Biomol. Chem.* **2014**, *12*, 2645.
- (26) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., 3rd; Cisneros, G. A.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber 2021*; University of California: San Francisco, 2021.
- (27) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguet, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; Simmerling, C. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.* **2020**, *16*, 528–552.
- (28) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863–3871.

(29) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyR. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26*, 689–691.

(30) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of PKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284.

(31) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical PKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(32) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(33) Kightlinger, W.; Lin, L.; Rosztoczy, M.; Li, W.; DeLisa, M. P.; Mrksich, M.; Jewett, M. C. Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat. Chem. Biol.* **2018**, *14*, 627–635.

(34) Sulzenbacher, G.; Bignon, C.; Nishimura, T.; Tarling, C. A.; Withers, S. G.; Henrissat, B.; Bourne, Y. Crystal Structure of *Thermotoga maritima* α -L-Fucosidase. *J. Biol. Chem.* **2004**, *279*, 13119–13128.

(35) Kamerlin, S. C. L.; Warshel, A. At the Dawn of the 21st Century: Is Dynamics the Missing Link for Understanding Enzyme Catalysis? *Proteins* **2010**, *78*, 1339–1375.

(36) Khersonsky, O.; Röthlisberger, D.; Wollacott, A. M.; Murphy, P.; Dym, S.; Albeck, G.; Kiss, K. N.; Houk, D.; Baker, D. S.; Tawfik, D. S. Optimization of the In-Silico-Designed Kemp Eliminase KE70 by Computational Design and Directed Evolution. *J. Mol. Biol.* **2011**, *407*, 391–412.

(37) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, R. J.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16*, 62–74.

Recommended by ACS

Generating Protein Folding Trajectories Using Contact-Map-Driven Directed Walks

Ziad Fakhoury, Scott Habershon, *et al.*

MARCH 30, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Protein Dynamics and Enzymatic Catalysis

Steven D. Schwartz.

MARCH 21, 2023

THE JOURNAL OF PHYSICAL CHEMISTRY B

READ 

Accurate Prediction of Enzyme Thermostabilization with Rosetta Using AlphaFold Ensembles

Francesca Peccati, Gonzalo Jiménez-Osés, *et al.*

JANUARY 16, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Effective Molecular Dynamics from Neural Network-Based Structure Prediction Models

Alexander Jussupow and Ville R. I. Kaila

MARCH 24, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Get More Suggestions >