Physics-based Modeling in the New Era of Enzyme Engineering

Christopher Jurich,[1,#] Qianzhen Shao,[1,#] Xinchun Ran,[1] and Zhongyue J. Yang[1-5,*]

[1]*Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States* [2]*Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States* [3]*Vanderbilt Institute of Chemical Biology, Vanderbilt University, Nashville, Tennessee 37235, United States* [4]*Data Science Institute, Vanderbilt University, Nashville, Tennessee, 37235, United States* [5]*Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, United States*

[#]*These two authors contribute equally to this work.*

ABSTRACT: Enzyme engineering techniques optimize enzymes to synthesize value-added chemicals, degrade environmental pollutants, and improve therapeutics. The field is entering a new era characterized by the increasing integration of computational strategies. While bioinformatics and artificial intelligence (AI) have been extensively applied to accelerate the screening of function-enhancing mutants, physics-based modeling methods, such as molecular mechanics and quantum mechanics, serve as essential complements in engineering objectives where setting up high-throughput screening is difficult or where a deep understanding of unknown physical principles is crucial. In this perspective, we discuss the enormous, untapped potential of physics-based modeling in guiding the next step of computational enzyme engineering. We first explore the paradigm of physics-based design principles wherein insights from natural, efficient enzymes are applied to recommend beneficial mutations *in silico.* We examine current development of high-throughput molecular modeling workflows that aid enzyme engineering campaigns through large-scale virtual applications of design principles. We then emphasize how physics-based modeling empowers AI techniques through enriching data expressiveness and interpretability. Finally, we proposed unmet challenges for the next step advancement of computational tools for enzyme engineering.

1. Introduction

Enzyme engineering concerns leveraging enzymes to suit our catalytic needs for synthesis, therapeutics, and sustainability. Industrial appetite is strong, with predicted compound annual growth rates ranging from 5 to 6% for the next decade.[1] A desirable future of enzyme engineering hinges on the creation of computational protocols capable of pinpointing functional raw enzymes and their engineered variants with quantitative accuracy, where biocatalytic development can be achieved with minimal screening efforts, as well as associated economic and environmental cost. At present, directed evolution-based (DE) protocols dominate the field and are routinely applied to create enzymes which assist in chemical production, plastic degradation, and addiction treatment.[2-4]

Despite clear and established success, the reliance of DE-based protocols on high throughput experimental screening precludes its application to some enzyme systems and engineering objectives. When side reactions are non-negligible (proteases[5]) and pure protein samples are needed for the assay, screening protocol cannot be easily constructed. In addition, it is difficult to reliably miniaturize enzymes through high-throughput deletion or truncation while maintaining their high activity.[6] The establishment of such an engineering strategy will improve therapeutic efficacy [7, 8], enhance enzyme compatibility in microfluidic devices, and reduce production costs and resource consumption. Engineering enzymes to perform optimally in non-native, extreme conditions is another critical, recurring task in industrial biosynthesis.[9] Although efficient extremophile enzymes create a clear path for reducing both cost and environmental impact, the commonplace mismatch between biological and industrial conditions (temperature, pH, etc.) makes extreme system engineering considerably less addressable with high throughput screening.

Beyond intractable engineering objectives, the prevalent use of bacterial expression systems in high throughput screening makes working with plant-based and mammalian enzymes challenging or impossible, albeit their innumerable biosynthetic functionality or reduced immunogenicity.[10] Eventually, although often viewed a strength, the choice of DE to treat catalysis as a black box process introduces the potential for evolutionary dead ends that cannot be escaped without structurally or mechanistically-derived detours[10] and prevents the quantitative prediction of enzymatic activity. When trapped in such a dead end, screening an additional 10^9 variants was unable to improve the efficiency of a human kynureninase. The logistical and theoretical limitations associated with DE and high throughput experimental screening highlight the need for complimentary techniques that overcome these limitations.

Computational sciences offer a definitive path for enzyme engineering to realize its full potential by expanding the scope of addressable enzyme properties and systems (Figure 1). Despite the growing use of bioinformatics and artificial intelligence to achieve these goals,[11, 12] physics-based molecular modeling techniques remain indispensable due to the ubiquitous insufficiency in both the quantity and quality of relational enzyme sequence-structure-function data.[13] Quantum mechanics (QM) and molecular mechanics (MM) can in theory be applied to measure experimentally-relevant functions for arguably arbitrary systems with an atom-resolved, three-dimensional structure, regardless of the enzyme's origin or preferred environment. Leveraging physics-based modeling, *de novo* enzyme design showcased the ability of first-principle approaches to create artificial enzymes that catalyze new-to-nature reactions, complimenting limitations of experimental screening. Unlike DE protocols which require a starting scaffold, *de novo* design uniquely focuses on the creation of completely artificial scaffolds whose residues contribute to catalytic efficiency by stabilizing rate limiting TS while adopting a stable fold.

Following the initial enzyme design efforts in the 1990s,[14] the seminal inside-out design protocol led to highly efficient, artificial retro-aldolases and kemp eliminases (KE),[15, 16] and further advanced approaches were developed for optimizing *de novo*-designed enzymes.[17-19] Though relying on DE protocols to optimize the designed scaffolds[15, 20] and frequently reported to hit evolutionary dead ends,[10] *de novo* enzyme design demonstrates that virtual, physics-based design can complement conventional screening-based techniques, representing a conceptual milestone for computational enzyme engineering.

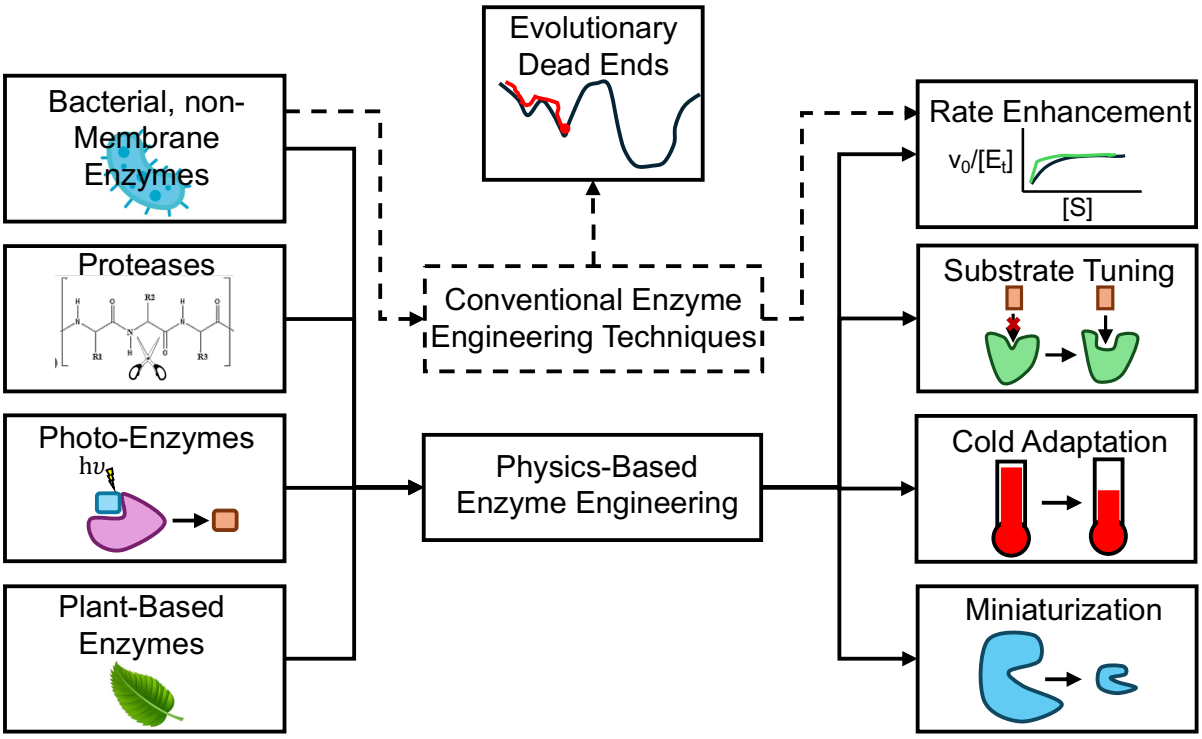

Figure 1: Physics-based computational methods as an approach to realize enzyme engineering's full potential. Conventional enzyme engineering techniques can reliably improve enzymatic efficiency for bacterial, non-membrane enzymes (top, dashed lines). A shortcoming in these workflows is the potential for campaigns to enter evolutionary dead ends (top, middle), as well as difficulties to address various systems like proteases, photo-enzymes, and plant-based

enzymes (left) or tackle other functional objectives like substrate tuning, cold adaptation, and miniaturization (right). Physics-based enzyme engineering can optimize systems accessible to conventional techniques, as well as proteases, photo-enzymes, and plant-based enzymes (left, black lines). Additionally, physics-based protocols can optimize substrate preference, cold adaptation, and miniaturization (right).

Besides enabling *de novo* enzyme design, molecular modeling has been extensively applied to elucidate an enzyme's mechanism[21] and interpret the origin of efficiency[22-24] and selectivity[25, 26] through transition state (TS) or reaction barrier calculations.[27, 28] These modeling also informs a quantitative relation or engineering principles that score and rank candidate enzyme scaffolds and mutants to achieve desired enzymatic functions. Furthermore, the resulting engineering principles inspire the design of descriptors and architectures for building machine learning (ML) models to accurately predict enzyme efficiencies, regioselectivity, substrate affinity, and other functions,[29-32] complementary to existing ML models trained solely from sequence or multiple sequence alignment (MSA) by enhancing molecular expressiveness of protein data.[13] ML methods additionally enhance physics-based modelling, performing dimension reduction on complicated MD-derived datasets and helping identify catalytically relevant modes or global conformations. Augmented with physics-based modeling, ML models may serve as an optimal pathway as the field works towards comprehensive models of catalysis.

In this perspective, we discuss the paradigm of physics-based modelling as a means for enzyme engineering to achieve the goal of optimizing arbitrary characteristics of arbitrary enzyme systems. We describe principles-based design, which aims to apply insights derived from known, efficient enzymes to predict beneficial mutants in yet unoptimized systems. Our perspective highlights the growing influence of high throughput molecular modeling workflows which stand

to accelerate the progress of principles focused enzyme engineering by increasing the number of enzymes and mutations analyzed. We also consider the fusion of ML and physics-based enzyme engineering, detailing the mutually beneficial impact of these two methods on each other. We further emphasize that ML methods will play a pivotal role in field, combining different physics-based measurements to develop a holistic view of enzyme catalysis.

2. Physics-based Modeling Elucidates New Structure-Function Relationships Underlying Enzyme Catalysis

Molecular insights of enzyme catalysis, as derived from physics-based molecular modeling, guide the identification and deployment of beneficial mutants for enhancing enzyme functions (Figure 2). Design principles are often formulated from observing and investigating known, experimentally characterized enzymes. In this sense, creating design principles allows the field to leverage molecular evolution from nature to inform our enzyme engineering efforts. Just as natural enzymes owe their catalytic efficiency to several sources, design principles survey many aspects of enzyme scaffolds. Based on classical molecular dynamics (cMD) and QM modeling, these studies may involve structural, dynamic, and electronic characteristics of enzymes including topology, enzyme electrostatics, flexibility and residue networks, or theoretical calculations with thermodynamic relevance such as activation energy, activation free energy and heat capacity. Researchers increasingly elucidated the correlation between molecular simulation-derived features and experimentally characterized kinetics and binding data.[33-35] These phenomenological models present an avenue for rapid scoring of mutation effects and improvement of catalytic functions. Due to the complexity of enzyme catalysis, there are yet undiscovered physical principles underlying enzymes' extraordinary catalytic efficiency and selectivity. The investigation of these

principles provides a direct path to further derive insights into catalytic activity as the field works towards a more holistic view of catalysis.

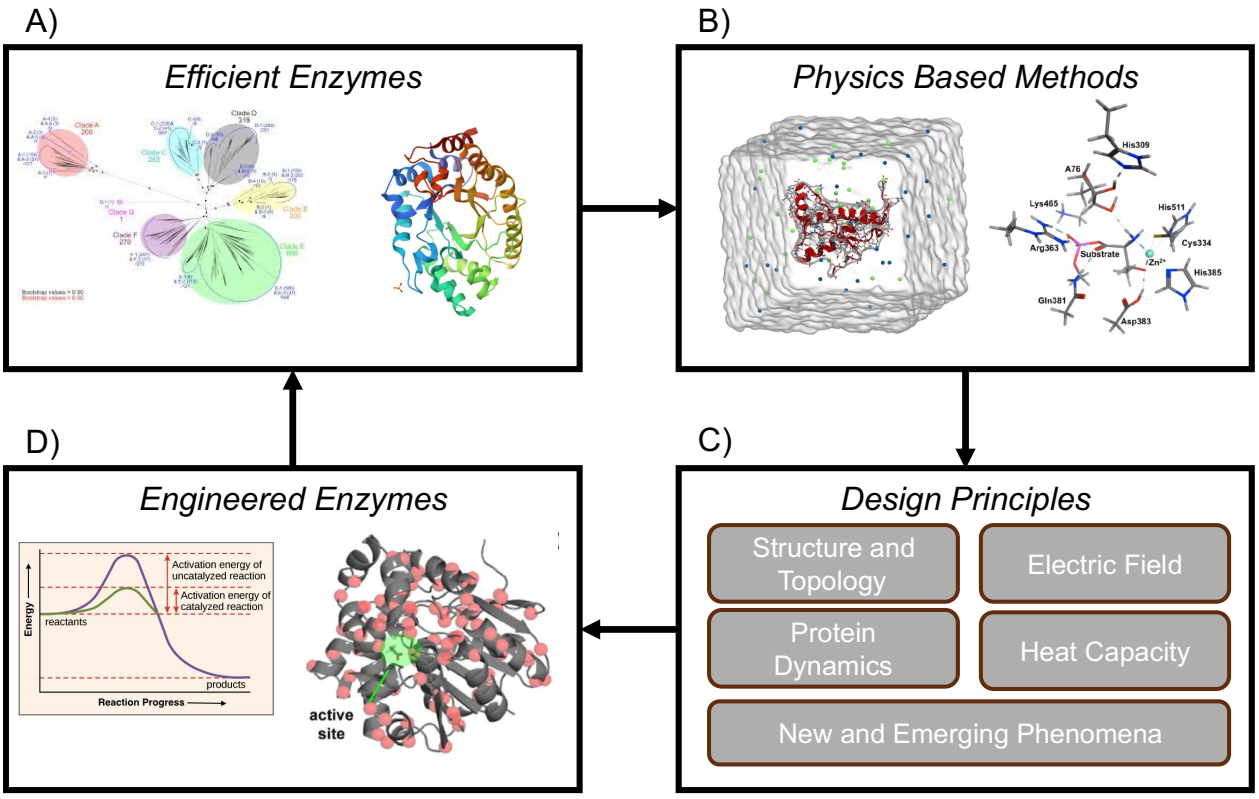

Figure 2: The lifecycle of physics-based principles. Physics-based principles are derived through observation of efficient enzymes from both natural and engineered sources (A). Individual principles and physical phenomena are identified, quantified, and better understood through physics-based computational simulations leveraging QM and MD frameworks (B). After identification in multiple systems, design principles are codified into generalized rational design rules which create definite, quantitative functional predictions (C). Design rules are applied to rank beneficial mutations to achieve a given functional objective, which can serve as another efficient enzyme (D).

2a. Structure and Topology

Structure-based enzyme engineering is uniquely convenient as beneficial mutants can be rationalized visually and AlphaFold2 has made it trivial to produce sufficiently accurate three-dimensional models for sequences of globular proteins[36]. Computational analysis of enzyme structures suggests that catalytic efficiency toward a specific substrate is improved when active sites show shape complementarity for that substrate and tunnel geometries promote rapid diffusion of associated reactants or products to and from active sites, respectively. Demonstrating the importance of active site complementarity, computational analysis shows that conserved guanine binding sites broadly drive ribozyme selectivity,[37] a single residue in catechol O-methyl transferase (COMT) positions its SAM cofactor to achieve a preferred donor-acceptor distance,[38] and that the active site residues of bacterial arylmalonate decarboxylase (AMDase) drive substrate specificity by tuning the size of a hydrophobic pocket to accommodate various substrates.[39] Topological engineering focuses on altering an enzyme's structure with these principles in mind, typically selecting mutations which favor substrate binding, or to improve tunnel accessibility. Active site mutations have been repeatedly deployed to improve substrate complementarity and enzymatic efficiency. Such mutations have enhanced an *O*-methyl transferase's ability to synthesize the pharmaceutical pinostilebene,[40] improved substrate specificity of an acyltransferase in an unfavorable aqueous environment,[41] and transformed a non-enzymatic protein to a KE.[42] Topological engineering has also been leveraged to alter substrate specificity by mutating buried residues in a channel, reducing the ability of some substrates to travel to the active site,[43] or to improve overall efficiency of a flavin dependent halogenase (FDH) by both giving better access to active sites and reducing leakage of an intermediate.[44] A clear inconvenience of topology-based design is the reliance on domain knowledge and the lack of generalizability. Rational engineering

metrics are typically specific to the system at hand and can hardly be compared to other physics-based features.

Structure-based insights can also be applied to intuitively engineering non-functional aspects of enzymes including pH preference. Assuming that surface residues play an outsized role in determining an enzyme's optimal pH, mutating these amino acids is a viable strategy for its preferred aqueous environment. By mutating only two residues, the optimal pH of vanillin dehydrogenase was shifted from 7.4 to 9.[45] This work illustrates the importance of considering a mutant's structural context and the viability of optimizing an enzyme's optimal pH. Most enzymes evolved to prefer environments closer to neutral pH's, and tolerance for less biologically common conditions opens the door for working with reactions that proceed more rapidly in a basic or acidic environment. Altering optimal pH is a largely unutilized engineering strategy for improved enzymatic efficiency.

Structure-focused enzyme engineering is inherently limited by its qualitative nature which hinders quantitative comparison between different enzyme classes. Despite the abundance of accurate, predicted structures from AlphaFold2, the inability to predict ligand locations limits the application of these principles to this breadth of data. AlphaFold3 offers some relief by predicting ligand locations, but it is not enough to simply stabilize interactions in the ground state as an enzyme must ensure that concerted interactions are in reactive conformations capable of generating products.[29] It remains to be seen how accurately this tool can predict reactive conformations. This and other shortcomings of structure-based principles are addressed partially by creating more conformations for a given structure, and new generative models are making it possible to do so easily and rapidly. The recently developed idpGAN creates structure ensembles

from sequence alone, introducing a new paradigm for structure-based design principles where insights are rapidly applied without the need for MD simulations.[46, 47]

Beyond technical approaches, applying crowd-sourcing to analyze and engineer enzyme structure remains an untapped avenue. Though low-throughput, human intuition is known to be effective and creative at optimizing various biomolecule problems as evidenced through successful efforts to game-ify protein folding, RNA folding, and drug design.[48-51] Applying Web3 block-chain technology could enhance massively parallel efforts to study and design on the basis of enzyme structure, and one could envision a public ledger that stores individual analyses and serves as an intermediary to develop consensus amongst citizen scientists.

2b. Electrostatics

Enzyme electrostatics, such as electrostatic potential and electric field (EF), mediate chemical reactivity that involves change of ionic states or charge separation. They have a quantitative connection to transition state stabilization and are convenient to calculate, making it popular in the computational enzyme engineering community. Experimental EF can be determined through vibrational Stark Shift experiments. When applied to enzymes, these methods typically require the probe molecule to have a rigid structure to enable calibrations based on MD-simulations.[52] Computationally, EF can be calculated using Coulomb's law based on atomic charges derived from MM or QM methods, and then projected onto dipole moments of reacting bonds to yield stabilization and interaction energies.[25, 53] Numerous cMD and QM/MM studies on ketosteroid isomerases, KEs, P450 enzymes, dihydrofolate reductase (DHFR), glycine N-methyltransferase (GNMTs), 20S Proteasome, and catechol-O-methyltransferase (COMT) show that enzyme EF directs substrate specificity, as scaffolds directly accommodate the electronic structure of substrates and promote the breaking and formation of bonds.[54]

Seminal work by the Head-Gordon group converted the understanding of enzyme electrostatics into a design principle. They established that individual mutations of Kemp eliminases (KE) can effectively fine-tune the magnitude of EF projected onto a catalytically relevant chemical bond, thereby enhancing KE's catalytic rate.[53] This principle was fully leveraged in the engineering of the highly efficient KE15, which saw point mutations deployed to favor bond breaking in a KE.[55] Beyond KE, observation of an electrostatic basis for efficient hydrolases motivated the inclusion of an aspartate residue which transformed the *Bacillus subtilis* esterase Bs2 into an amidase through electrostatic stabilization of the TS.[56]

Developing gold standard methods for calculating EF remains an open challenge, and providing relevant technical infrastructure stands to benefit the field at large. Researchers often project an enzyme's EF only along relevant bonds of reactant and transition state ($E = -\boldsymbol{F_{enz}} \cdot \boldsymbol{u_{bond}}$), and new methods are being developed to integrate the product of the electron density and the electric field potential of the entire reactant or transition state molecule ($E = \int \rho(\boldsymbol{r}) V_{enz}(\boldsymbol{r}) d^3\boldsymbol{r}$) to calculate the electrostatic stabilization energy.[57] Simple EF calculations treat the enzyme scaffold as a collection of point charges. Polarizable force fields like AMOEBA offer enhanced accuracy which surpasses the typical over estimation of fixed charge force fields and rivals that of QM-derived EFs.[58] However, GPU-accelerated polarizable force field calculations are only supported in Tinker-HP[59] instead of other popular MD packages such as AMBER, GROMACS, etc., presenting a need for accessibility enhancement. Rational engineering relies on deploying point mutations to improve EF strength or stabilization effects, but critically assumes that mutants retain the global fold and substrate positioning dynamics of the wild-type protein. The latter is especially pertinent as minor changes to substrate orientation may quickly abolish anticipated electrostatic gains. Analysis of residue coupling based on mutual sidechain information

presents a potential means to identifying EF-mediating residues unlikely to perturb substrate dynamics, but methods are not generalized across enzymatic systems and require further refinement.[60] Improving the accuracy of enzyme EF calculations through advances in computational methodology stands to promote the use of these design principles and presents untapped potential for growth, especially when it can be predicted which EF-mediating residues are unlikely to alter local and global conformational changes.

2c. Protein Dynamics

Dynamics-inspired enzyme engineering emphasizes how conformational changes mediate catalytic activity and apply these principles to improve enzyme functions. Though convenient, a PDB-derived single enzyme geometry provides only an averaged molecular view of catalysis, which does not inform how enzyme restructuring or reorganization influences substrate positioning for barrier crossing.[61] This single geometry also fails to inform the impact of mutation on catalysis via rearranging active site residues or even dynamic allostery.[62, 63] B-factor analyses over 60,000 enzymes across 925 families show that the flexibility of active-site residues varies drastically as an enzyme evolves to adopt new functions, indicating the necessity of conformational ensemble in elucidating enzyme functions.[64] MD simulations sample enzymatic conformational ensembles and measure structural or energetic features across frames of production MD runs, generating enzyme property values, such as binding affinity, chemical selectivity, and even reaction barriers when combined with QM or QMMM methods. MD simulations offer physical insights and design hypothesis that are not accessible through the analysis of static structures.

MD analysis informs that loop mutations reduce flexibility, improving ligand binding and enzymatic activity, assisting the experimental engineering of luciferase $Anc^{HLD\text{-}RLuc}$.[65] Measuring

the evolving ratio of substrate to active site SASA through the substrate positioning index (SPI) in KE mutants provides a more nuanced view that mutations can reduce catalytic activity when the resulting active site is too loose or tight, thereby implying a Goldilocks "sweet spot" of active site size.[33] Likewise, conformational pseudo-ensembles of ketosteroid isomerase (KSI) homologs, as experimentally characterized by room temperature X-ray crystallography, reveal the presence of an optimum flexibility that allows the residue to shuffle protons between different positions as a general base with maximum catalytic efficiency.[66] Evidences from both NMR and simulation studies show that the substrate positioning and the flexibility of the enzyme mediate catalysis through affecting the conformational entropy.[67] Investigations of enzyme reactant and TS-complexes provide insight into the ways in which shifts in conformational ensembles can facilitate catalytic efficiency. Dynamics studies have shown that liver alcohol dehydrogenase (LADH) experiences ps-ns conformational changes near its transition state to facilitate hydride transfer.[68] QM/MM investigations into the TS ensembles of adenylate kinase suggest that the presence of a broad ensemble likely contributes to increased entropy of activation, suggesting that such an ensemble increases catalytic activity.[69] Mutual information analysis of side chain dynamics in directed evolution-optimized KE has illustrated emergence of evolved entropic forces which destabilize the reactant state complex.[60] Investigations into dynamics-derived ensembles of ketol-acid reductoisomerase (KARI) illustrate that specific conformational regions are associated with higher reactivity, and that mutations should be targeted to preferentially populate those regions.[70] Similar work analyzing correlated residue movement has led to the creation of the shortest path map (SPM) model, which uses MD trajectory data to identify residues that are likely instrumental to catalytically relevant conformational switches.[71] Additionally, dynamics-derived principles help identify rate-enhancing mutants. The observation that rigid catalytic residues enhance enzymatic

activity was leveraged to create the efficient HG4 KE, whose mutations promote both rigidity and active site organization for favorable catalysis.[19]

Despite storied success, logistical and theoretical limitations prevent universal application and adoption of dynamics-based principles in enzyme engineering campaigns. Sufficient conformational sampling, alongside the QM-based reaction barrier calculation, is necessary to identify catalytically relevant states. While computational costs continue to cheapen, conformational sampling remains a rate-limiting step in computational enzyme engineering efforts,[72, 73] particularly in cases where enhanced sampling or Markov-State modeling is needed. Generative models hold great promises to achieve low-cost conformational sampling directly from an input sequence, but it's application to computational enzyme engineering is likely impeded by its insensitivity to point mutations, though there have been reports of cases where AlphaFold2-generated conformal ensembles are sensitive to point mutations.[47] Another challenge for MD-guided enzyme engineering is the lack of generalizable quantitative metrics that represent the impact of protein dynamics on chemical reactivity. Like structure-based principles, MD-based design principles are often system specific. Substrate positioning index (SPI) demonstrates a volcano-like piecewise linear correlation with free energy barrier in lactonase SsoPox[35] and KE[33]. However, it remains unknown how to identify the SPI corresponding to the optimal activity *a priori*. Investigations of soybean lipoxygenase (SLO) have demonstrated the critical role that distal loop motions play in thermally activated enzymes, where the enthalpy of activation of hydrogen-deuterium exchange (HDX) energy and activation barrier energy are unexpectedly positively correlated.[74] The fundamental nature of this observation paves the way for new dynamics-inspired enzyme engineering principles that are potentially generalizable across systems.[61] Additionally, entropy has long been known critical for catalysis, but how to quantitatively factor in the role of

entropy from conformational ensemble remains an open question.[75] There is also increasing interest in predicting relative populations of reactive states capable of chemical productivity. Proximity alone is not enough for reaction procession and case studies have suggested that discriminating reactive and non-reactive geometries is a non-trivial task in need of further development.[70, 76] Treating enzymes as conformational ensembles is a fundamentally robust approach, and further refinement and application of MD-guided design principles maximize the potential of computational enzyme engineering.

2d. Heat Capacity

Considering heat capacity adds a unique dimension for engineering enzymes with a temperature-dependent behavior, such as cold-adaptation and thermostability. Negative heat capacity in a highly efficient KE is associated with stabilization of the TS state versus the ground state.[77] It has also been demonstrated that the addition of 2,2,2-trifluoro ethane induces similar effects in engineered enzymes leading to associated efficiency gains.[78] Standard Arrhenius behavior assumes that enzyme activity increases exponentially with temperature until protein degradation occurs. The incompatibility of Arrhenius behavior (non-Arrhenius behavior) seen in cold-adapted α-amylase (AHA), ancient reconstructed adenylate kinase (ANC1), and others has motivated the establishment of a heat capacity-based framework for understanding the temperature dependency of enzyme efficiency from MD simulations.[79] These simulations elucidate the emergence of inactive substrate-enzyme conformations, rather than a non-trivial activation heat capacity, to be the key factor underlying the cold adaptation of AHA. In theoretical terms, AHA keeps activation enthalpies low and activation entropies more negative by preventing conformationally competent enzyme-substrate interactions.[80] This principle was applied back to AHA, leading to the identification of mutations which shifted its thermal optima upward.[81]

Engineering enzyme temperature preference through heat capacity prediction is promising and untapped. With methods for predicting non-Arrhenius activity only being developed recently, they remain largely unapplied despite potential applications in sugar, laundry, and textile manufacturing, as well as biocatalysis, and sustainable production. Fundamental and glaring issues arise when trying to apply lessons from monodomain AHA to other industrial enzymes, such as amylases and cellulases, because the majority of these enzymes have two domains, a catalytic domain and a carbohydrate-binding module.[82, 83] Instead, it has been demonstrated that cold adaptation can be achieved via the introduction of linkers which increase domain separation index (DSI), an MD-derived descriptor which rigorously describes domain separation.[84] DSI-guided screening of linkers helped introduce a 12-fold activity increase in *Pseudomonas saccharophila* amylase (psA) at 0°C from 2.4% to 30.5%. Although this work provides a physical principle for engineering cold-adapted bidomain enzymes, its structural basis underlying the apparent non-Arrhenius activity remains to be investigated.

2e. Complex Mechanisms that Require New Engineering Principles

Until every contributor of enzyme catalysis is catalogued and codified, there is still room for the development of new physics-based design principles. At present there are an abundance of mechanisms which enable enzymes to achieve high catalytic efficiency, but have not been distilled into concise design principles. Excited state simulations have unrevealed mechanisms of photoactive enzymes.[85-87] Hydrogen tunneling is critical to the rate limiting step of soybean lipoxygenase (SLO),[61] and recent MD and QM simulations are only now beginning to unveil its mechanistic details.[88] Before that, multi-dimensional tunneling analysis has also been applied to understand how hydrogen, proton, and hydride transfer occurs in various enzyme systems, although insights have largely not been applied to further enzyme engineering efforts.[89] More

broadly, proton coupled electron transfer (PCET) serves the basis of countless highly efficient enzymatic reactions, and studies have been conducted to elucidate how mutations affect PCET.[88, 90] However, there has been limited understanding of how beneficial mutations can be predicted to gain desired functions such as activity, selectivity, etc. in an enzyme involving PCET.[91] In addition, growing interest exists in understanding the chemical reaction dynamics of enzyme catalysis on the femtosecond timescale. Ab initio path integral simulations of KSI^{D40N} discovered a quantum proton delocalization facilitated by a triad of strongly hydrogen-bonded tyrosine residues, which leads to a 10,000-fold increase in the acidity of one of the tyrosine.[92] Transition path analysis of purine nucleoside phosphorylase (PNP) highlights that a distal residue contributes a rapid promoting vibration.[93] Committor analysis of Myosin II with frozen active site residues has shown that specific global protein motions are required for ATP hydrolysis to proceed.[94] Two-dimensional QM/MM MD simulations elucidated the complex mechanism of oxidosqualene cyclase involving a nearly concerted but highly asynchronous cyclization.[95] QM/MM-based quasiclassical trajectory simulations illustrate how post-TS bifurcations determine product selectivity in SpnF-catalyzed Diels–Alder reactions, demonstrating the kinetic energy contribution of active-site hydrophobic residues in chemical activation.[21] Fundamental understanding of enzyme's chemical activation networks will suggest untapped avenue to engineer biocatalysts. Clearly, there is a diverse ecosystem of physical phenomena critical to enzyme catalysis which are not yet applied to enzyme engineering efforts. In turn, physics-based enzyme engineering is an exciting field with considerable potential to grow as computational resources continue to cheapen and phenomena like PCET, hydrogen tunneling, quantum proton delocalization, excited states, rapid enzyme motions, and post-TS bifurcations are studied in greater detail.

3. Physics-based Modeling Facilitates Discovery of New Enzymes and Variants

Automated workflows are an emerging means of fully leveraging physics-derived design principles to evolve the field of enzyme engineering. While there is a robust and validated selection of design principles, manually applying them to enzyme systems limits the field by lowering throughput-levels, reducing reproducibility, and keeping technical barriers to entry high. The low-throughput level associated with manual system preparation hurts the field by both reducing the number of enzymes that are refined and limiting the exhaustiveness of sequence search for those systems. Manual preparation of input files as well as analysis of simulations introduces innumerable failure points, translating into an escalated risk of error and reduced reproducibility when deploying mutations to a system. Computational enzyme engineering workflows typically feature multiple software packages, introducing a high technical barrier to entry and ultimately limiting the size of the community and the diversity of perspectives (Figure 3). Consequently, computational enzyme engineering workflows should aim to address these concerns.

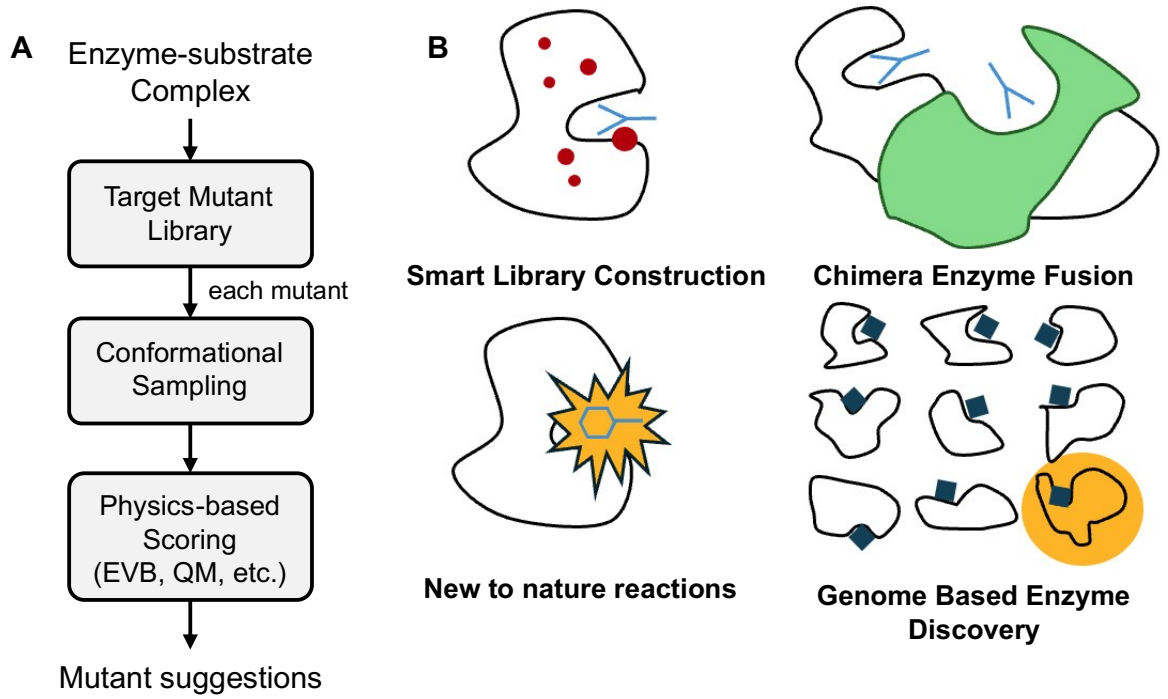

Figure 3: The role of high throughput workflows in enzyme engineering. Conventional workflows for computational enzyme engineering follow a common pattern where a mutant library is generated for a prepared enzyme-substrate complex and physics-based scoring is used to recommend mutants based on rigorous values averaged through conformational sampling (A). There are currently no computational workflows developed for optimizing non-rate enhancing objectives, leaving tasks like smart library construction, chimera-enzyme fusion, engineering new-to-nature reactions, and metagenomic enzyme discovery unaddressed (B).

Physics-based enzyme engineering workflows were pioneered with the release of CADEE in 2017.[96] Designed as a workflow to perform computer-aided directed evolution, CADEE was the first platform designed for the sole purpose of ranking and recommending individual mutants with physics-based methods (i.e., EVB). To apply CADEE to a given system, one needs to prepare the enzyme-substrate complex and calibrate the EVB force field with experimentally characterized mutants of the target enzymatic reaction. Alternatively, rigorous calculations or experimental data of the intrinsic reaction can be used. The automatic workflow then deploys mutations before running an MD simulation and ranking each mutant based on the activation energy calculated via a standard EVB free-energy perturbation/umbrella sampling (EVB-FEP/US) procedure. CADEE was initially tested on the proton transfer step of the triosephosphate isomerase (TIM) using a set of rigorous EVB parameters for this enzyme. As a pedagogical example, 128 single point mutations were computationally screened in 9.5 days with 512 CPU cores. CADEE is a clear success in the field, but there are limitations to this workflow. CADEE's performance is sensitive to EVB force field's parameterization quality, and expert input is needed when relevant experimental data is not available. This limits application of the workflow to well-studied enzymes and essentially precluding analysis of sequences folded by AlphaFold2. Reliance on implicit

knowledge gained from fitting to experimental data also introduces variability across systems, making it difficult to systematically improve the accuracy and generalizability of CADEE. The bulk of CADEE's drawbacks originate from its reliance on EVB, which could be addressed by implementing support for other techniques. Unfortunately, significant further development for CADEE has not been reported since its release, highlighting a recurring issue in the field. Software projects for enzyme engineering often have short active development lifetimes, resulting in inadvertent specialization. The field lacks a software package with robust support for numerous simulation techniques and design principles.

EnzyHTP[73] was released in 2022 and fundamentally advanced the state of the art for high throughput computational workflows. Following up CADEE, EnzyHTP was conceptualized and developed as a robust and flexible tool for general purpose enzyme engineering. Notably, EnzyHTP automates every step of enzyme engineering including preparation, mutagenesis, geometry sampling, and post-hoc analysis. This python-based package advances the state of the art by supporting arbitrary molecular modelling tasks including MD, QM, ligand docking, trajectory analysis, and more. Exposing this functionality through modular python functions uniquely enables the creation of flexible analysis and engineering workflows tailored to the system at hand. Ultimately, EnzyHTP serves as a modular breadboard from which other workflows can be built. For example,[72] an EnzyHTP-based workflow was recently developed to perform computational DE. In this workflow, *in silico* analysis allowed hundreds of mutants to be screened for both thermostability and their ability to stabilize the breaking bond in the rate-limiting TS. [43,44] Applied to the KE07, a well-characterized and optimized KE, this EnzyHTP workflow successfully identified all 4 experimentally observed rate-enhancing mutants. A total of 184 mutants were screened using a combined 18.4 μs of equilibrium MD simulations and 18,400 QM

single point calculations. Calculations were performed in 3 days using 30 GPUs and 1000 CPUs. This work highlights various pareto optimization concerns for such workflows as striking balances between computational cost and mutant ranking accuracy, smart library construction schemes, or functional scoring remain unsolved problems.

Narrow selection of targeted engineering objectives and insufficient incorporation of physics-based principles within workflows are two shortcomings in the field which both represent potential areas of growth. Protocols like CADEE, EnzyHTP, ASiteDesign[97], and CASCO[98] have demonstrated the power of high throughput workflows for the task of identifying rate-enhancing point mutations and modifying selectivity, but no analogous pipelines have been applied to solve challenges like tuning the substrate scope. This lack of functionality is problematic as altering substrate scope is a desirable task in computational enzyme engineering and especially in the context of biosynthetic enzymes with applications in industrial production and late-stage pharmaceutical functionalization. Other unaddressed functional challenges include accelerating genome-based enzyme discovery,[99] engineering new-to-nature reactions in enzymes,[100, 101] and assembling fused functional domain into chimera enzymes (Figure 3B).[102] While these objectives are addressable with existing methods created to rank point mutations, the majority of physics-based design principles remain unimplemented in a manner which could be applied to address these tasks. Given the diversity of these functional objectives, it is almost certain that many more physics-based principles will need to be codified, quantitated, and applied in high throughput workflows to see meaningful progress.

Besides expanding the scientific relevance of high-throughput enzyme modeling workflow, a technical challenge lies in software engineering. Applying coding best practices is non-trivial but frequently neglected in the initial stage of software development, leading to extensive

refactoring efforts when the inherent software architecture struggles or fails to accommodate expanded functionality. On the other hand, low code readability, over-simplified documentation, and poor co-development infrastructure are not uncommon, limiting the development and applications of the software to the developers' research group. As a result, maintenance may cease to continue as the main developer trainees move to the next career stage. Initiatives like the Molecular Sciences Software Institute have made a huge impact on raising awareness within the community about the importance of implementing software design principles in molecular modeling software, training a large group of early-stage computational chemists and biologists specializing in software development. Looking ahead, the challenge lies in how to make software engineering initiatives a routine part of training programs in traditional science departments, thereby creating a cohort of developers equipped to tackle interdisciplinary challenges like protein engineering.

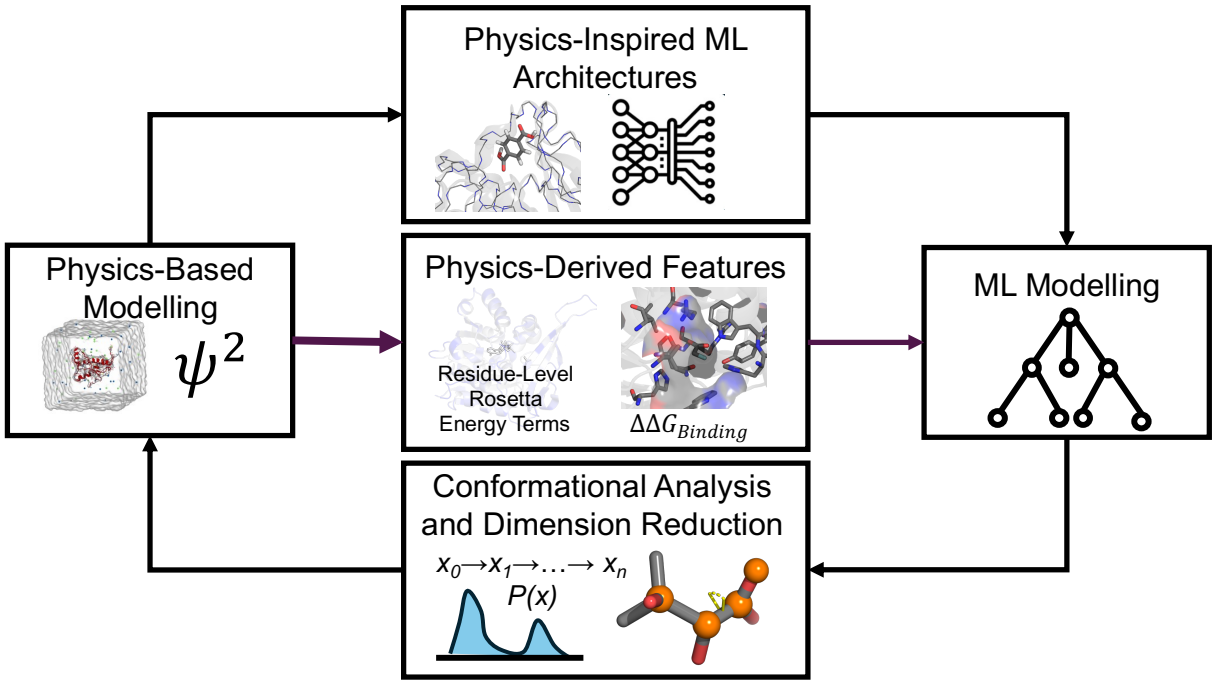

Figure 4: The symbiotic relationship of physics-based and ML modelling. Physics-based modelling (left) improves ML modelling (right) by motivating physics-inspired ML architectures

(top) and providing models with physics-derived features and descriptors (middle). ML modelling aids the advancement of physics-based understanding of catalysis through conformational analysis and dimension reduction (bottom) of large-scale dataset intractable to simplistic analytical techniques.

4. Symbiotic Fusion of Physics-Based Modelling and ML

ML has been extensively used to guide enzyme engineering, but the lack of robust datasets combining structure, sequence, and function has hampered model development. Physics-based modelling is uniquely poised to address this limitation and enhance the expressivity and performance of ML models by providing an abundance of microscopic, information dense descriptors. Beyond feature generation, physics-based techniques enable model interpretability by establishing concrete links between catalysis and enzyme-substrate interactions in the ground state, reactive conformational state, and even transition state. As an example, incorporating structure-based features increases prediction performance in ML models. In the case of predicting the impact of mutations on relative improvement of the activity in bovine enterokinase (EKB), the introduction of MD-derived conformational descriptors improved Pearson correlation, root mean square error (RMSE), and mean average error (MAE) versus a sequence-only model.[32] In addition, structural features enhance stereoselectivity predictions. Explicit encoding of active site enzyme-ligand interactions propelled EnzyKR to outperform the general-purpose k_{cat} predictor (DLKcat) when predicting favored enantiomers for hydrolases.[31] Considering the lack of explicit substrate-enzyme interaction information in one-dimensional protein sequence and substrate SMILES string, the notion that structural features boost ML performance may seem obvious. The practical challenge, however, lies in the acquisition of quality enzyme-substrate complexes. Unlike sequence databases that are numerous, robust, and accessible, integrated sequence-structure

libraries are rare. The development of integrated sequence-structure-function database, such as IntEnzyDB[103] and OpenEnzymeDB,[104] emerges as a solution, but the size of curated quality data still largely lags behind the community needs. Moreover, despite advances of reactive docking algorithms,[105, 106] a comprehensive database of large-scale catalytically relevant pre-reaction complexes across diverse enzymes and substrates remains undeveloped. Equally urgent is the establishment of enzymology databases with physics-based modeling-derived descriptors. An indicative example is the BioFragment Database, a repository of QM-derived protein interaction energies that provides a template for how researchers can create generalized values which can be readily adapted as features for training ML models.[107] A similar database containing QM and MD descriptors has also been developed for antimicrobials, and given the widespread use of these techniques in the enzyme engineering community, developing analogous resources for enzyme structures would both accelerate and improve ML efforts.[108] Molecular modeling-inspired sequence embedding tailored to enzyme engineering tasks is likewise under-developed, and serves as another avenue for improving the quality of ML models. There are also deficiencies in existing data as fitness-related values remain the most reported in large-scale dataset, such as ProteinGym.[109] Kinetic parameters like k_{cat} and K_m are important for training biocatalysis-oriented models, but they disperse in scientific literatures with a wide variety of reporting formats and units. Training GPT agents with biochemistry knowledge-guided prompt engineering serves as a promising solution. A longer-term strategy would be to build a community-level Web3-based infrastructure that provide token incentives for experimentalists to contribute their quality kinetic data to the community.

Supplying ML models with multiple feature classes improve performance today and paves the way for comprehensive catalytic understanding and the design of electrostatically optimized

sequences. Individual physics-based principles are useful for ranking mutations in a specific case but are hard to combine and generalize across families of enzymes. While binding affinity and TS barrier height both impact catalysis, the lack of quantitative models describing their coupling makes it difficult to predict the impact of mutations which improve one quantity and impair the other. This problem is complicated further when catalytically relevant conformational descriptors are unitless. Contemporary ML models address this problem by combining multiple disparate features in predictive efforts and additionally use this wealth of information to their benefit. By using docking scores, QM-derived charges, and other physics-based metrics, a recently developed classifier was able to predict substrate promiscuity of bacterial nitrilases against a test library with high accuracy.[30] When residue-level Rosetta energy terms and sequence identity were combined in a structure-aware protein graph convolutional network (PGCN), protease specificity was predicted with a minimum of 86.62% accuracy and greater than 90% in many cases.[110] Beyond improving model accuracy today, such studies pave the way for ML models to elucidate the interplay of competing physics-based phenomena on catalytic efficiency. Supplying models with information about substrate binding, product binding, and reaction-related features presents an opportunity to develop a holistic understanding of enzyme catalysis. Moreover, the success of the PGCN encoding and EnzyKR demonstrate the importance of representing spatial information in ML models and highlights how the physical origin of catalysis can inspire better model design and performance through the explicit modelling of systems in catalytically competent conformations. Looking to ML design, the introduction of ProteinMPNN has made it possible to generate expressible, soluble, and stable *de novo* enzyme scaffolds.[111] Despite this success, ProteinMPNN's focus on the scaffold stability and disregard for properties that are more catalytic relevant such as side-chain conformations, dynamics, electric field, etc. means that many enzymes require further

rounds of catalytic engineering before potential use. Developing similar models which consider protein EF presents a direct method to design enzymes which are biased towards higher initial efficiency.

Descending from broad views of catalytic activity to the transition state, ML models play key roles in addressing long standing problems associated with the generation and classification of reactive states. Generating accurate TS geometries inside an enzyme remains a major challenge, despite their importance in calculating barrier height. Recent work has demonstrated that equivariant diffusion models can generate highly accurate gas phase TS-geometries from structures of reactants and products alone.[112] Considerable further effort will be needed to extend this technology to account for interactions with both active site residues and solvent molecules, but using ML algorithms to generate accurate TS-geometries is an elegant means to make barrier calculations possible for more systems. Distilling catalytic meaning from high-dimensional MD simulations is another major challenge aided by ML integration. Connecting stochastically sampled enzyme geometries with activity is often hampered by the tendency for conformational shifts to occur. Markov chain Monte Carlo (MCMC) models solve this problem on a global scale by organizing MD snapshots into probability-weighted states which combine with physics-based calculations to show high agreement with experimentally measured values.[113] Looking to individual molecular coordinates, the high number of distances, angles, and dihedrals associated with even small ligand systems makes manual analysis potentially spurious and counterproductive. In the case of ketol-acid reductoisomerase (KARI), ML models analyzed substrate turnover events and identified measurables strongly associated with reactivity.[29] This application of ML is transformative and potentially not replicable through manual analysis. Further application of this technique to systems beyond KARI stands to grow knowledge of intrinsic reactivity and paves the

way for the distillation of a more generalized understanding of how ligand geometry impacts reactivity.

5. Challenges in Computational Tool Development for Enzyme Engineering

The advancement of physics-based enzyme engineering will depend on methodological and technical improvements. Enzyme engineering's reliance on computational tools provides both technical limitations at present and the potential for rapid gains as first-principles software packages see continued improvement. Computational cost is a primary bottleneck for most techniques in the field. MD and QM/MM are critical for conformational sampling and activation barrier height calculations, respectively, but both techniques can take days to calculate values meaningful for enzyme engineering efforts. Addressing the challenge demands advancements in both computing hardware and algorithms. On the hardware front, quantum computing presents a promising engine to drive the next-generation electronic structure simulations, with early applications demonstrating its potential in the structure prediction of proteins[114]. The hybrid use of quantum computers alongside existing classical processing units could lead to significant advances in modeling the rare events within enzyme systems, though the realization of a true quantum advantage remains uncertain. On the algorithmic side, the development of artificial intelligence (AI) to accelerate high-accuracy energy calculations and sampling is a thriving direction. Machine learning potentials have shown promise in facilitating QM/MM simulations, particularly for evaluating chemical barriers[115]. Furthermore, generative models are increasingly being used to map path-dependent free energy changes by leveraging information from end-point states, such as in targeted free energy perturbation studies[116].

Many tools have been developed, but there is no consensus on how to compare and titrate computational performance for enzyme engineering. In contrast to traditional computational

chemistry tasks, such as thermochemistry predictions with well-established benchmark sets, computational enzyme engineering faces a constantly evolving target that depends on the intrinsic mutational landscape of the selected enzyme. Some model systems like KE have become de facto benchmarks, but the practice of drawing conclusions from a single enzyme is biased and prone to create analyses which hyper fixate on artifacts or aberrations of a specific enzyme. AlphaFold2 and its origins in the Critical Assessment of Structure Prediction (CASP) program offer a potential path to more sound methodology in the enzyme engineering. CASP uses a blind process in which researchers' predictions are validated against unreleased protein structures and contain protein targets across different families. Such an approach creates a maximally robust validation environment and inspires the future creation of an initiative perhaps called Critical Assessment of Enzyme Functional Prediction. Notably, ProteinGym[109] offers a large dataset obtained from deep mutational screening (DMS) and clinical observation, mapping the protein sequence to the DMS score. While the dataset is valuable for AI model developments, ProteinGym does not contain information such as substrates, reaction mechanism, enzyme-substrate complex structures, or more physically meaningful kinetic properties such as k_{cat} or K_m, which are all critical for assessing a physics-based enzyme engineering tool. A gold standard of benchmarking in the field of enzyme engineering should comprise of enzyme data with a diverse range of wild-type sequences, mutations, reactions, reaction mechanisms, substrate types, experimental hit rates, as well as evaluation metrics and algorithms which critically assess the performance of software packages in a blind, unbiased manner.

Although the abilities of directed evolution and high-throughput screening are impressive, we only consider this approach as an intermediate step towards developing methods that can address any engineering objectives across any enzyme systems. Physics-based modeling plays an

essential role in advancing the next generation of enzyme engineering methods due to its unique ability to directly predict experimental observables from first principles, elucidate molecular mechanisms, and identify key molecular descriptors as design principles. This approach is crucial for ultimately unlocking the full potential of enzyme engineering, leading us into a new era of enzyme innovation and discovery.

AUTHOR INFORMATION

Corresponding Author

*Email: zhongyue.yang@vanderbilt.edu phone: 615-343-9849

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

This research was supported by the startup grant from Vanderbilt University. C. Jurich is supported by Vanderbilt Chemistry-Biology Interface Training Grant (T32GM149371 and T32GM065086). Z. J. Yang is supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM146982 and the Rosetta Commons Seed grant.

References

(1) Yadav, V.; Biswas, S.; Goyal, A. Enzymes of Industrial Significance and Their Applications. In *Industrial Microbiology and Biotechnology: An Insight into Current Trends*, Verma, P. Ed.; Springer Nature Singapore, 2024; pp 277-307.

(2) Liao, C.; Seebeck, F. P. S-adenosylhomocysteine as a methyl transfer catalyst in biocatalytic methylation reactions. *Nature Catalysis* **2019**, *2* (8), 696-701. DOI: 10.1038/s41929-019-0300-0.

(3) Bell, E. L.; Smithson, R.; Kilbride, S.; Foster, J.; Hardy, F. J.; Ramachandran, S.; Tedstone, A. A.; Haigh, S. J.; Garforth, A. A.; Day, P. J. R.; et al. Directed evolution of an efficient and thermostable PET depolymerase. *Nature Catalysis* **2022**, *5* (8), 673-681. DOI: 10.1038/s41929-022-00821-3.

(4) Rossi, L.; Pierigè, F.; Agostini, M.; Bigini, N.; Termopoli, V.; Cai, Y.; Zheng, F.; Zhan, C. G.; Landry, D. W.; Magnani, M. Efficient Cocaine Degradation by Cocaine Esterase-Loaded Red Blood Cells. *Front Physiol* **2020**, *11*, 573492. DOI: 10.3389/fphys.2020.573492 From NLM.

(5) Ferrall-Fairbanks, M. C.; Kieslich, C. A.; Platt, M. O. Reassessing enzyme kinetics: Considering protease-as-substrate interactions in proteolytic networks. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (6), 3307-3318. DOI: doi:10.1073/pnas.1912207117.
(6) Emond, S.; Petek, M.; Kay, E. J.; Heames, B.; Devenish, S. R. A.; Tokuriki, N.; Hollfelder, F. Accessing unexplored regions of sequence space in directed enzyme evolution via insertion/deletion mutagenesis. *Nat. Commun.* **2020**, *11* (1), 3469. DOI: 10.1038/s41467-020-17061-3.
(7) Shams, A.; Higgins, S. A.; Fellmann, C.; Laughlin, T. G.; Oakes, B. L.; Lew, R.; Kim, S.; Lukarska, M.; Arnold, M.; Staahl, B. T.; et al. Comprehensive deletion landscape of CRISPR-Cas9 identifies minimal RNA-guided DNA-binding modules. *Nat. Commun.* **2021**, *12* (1), 5664. DOI: 10.1038/s41467-021-25992-8 From NLM.
(8) Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **2020**, *578* (7794), 229-236. DOI: 10.1038/s41586-020-1978-5.
(9) Wu, H.; Chen, Q.; Zhang, W.; Mu, W. Overview of strategies for developing high thermostability industrial enzymes: Discovery, mechanism, modification and challenges. *Crit. Rev. Food Sci. Nutr.* **2023**, *63* (14), 2057-2073. DOI: 10.1080/10408398.2021.1970508.
(10) Blazeck, J.; Karamitros, C. S.; Ford, K.; Somody, C.; Qerqez, A.; Murray, K.; Burkholder, N. T.; Marshall, N.; Sivakumar, A.; Lu, W. C.; et al. Bypassing evolutionary dead ends and switching the rate-limiting step of a human immunotherapeutic enzyme. *Nat Catal* **2022**, *5* (10), 952-967. DOI: 10.1038/s41929-022-00856-6 From NLM.
(11) Menke, M. J.; Ao, Y.-F.; Bornscheuer, U. T. Practical Machine Learning-Assisted Design Protocol for Protein Engineering: Transaminase Engineering for the Conversion of Bulky Substrates. *ACS Catalysis* **2024**, *14* (9), 6462-6469. DOI: 10.1021/acscatal.4c00987.
(12) Yu, T.; Boob, A. G.; Volk, M. J.; Liu, X.; Cui, H.; Zhao, H. Machine learning-enabled retrobiosynthesis of molecules. *Nature Catalysis* **2023**, *6* (2), 137-151. DOI: 10.1038/s41929-022-00909-w.
(13) Mazurenko, S.; Prokop, Z.; Damborsky, J. Machine Learning in Enzyme Engineering. *ACS Catalysis* **2020**, *10* (2), 1210-1223. DOI: 10.1021/acscatal.9b04321.
(14) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. Protein Design: A Hierarchic Approach. *Science* **1995**, *270* (5238), 935-941. DOI: doi:10.1126/science.270.5238.935.
(15) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; Dechancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; et al. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453* (7192), 190-195. DOI: 10.1038/nature06879.
(16) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; et al. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319* (5868), 1387-1391. DOI: doi:10.1126/science.1152692.
(17) Privett, H. K.; Kiss, G.; Lee, T. M.; Blomberg, R.; Chica, R. A.; Thomas, L. M.; Hilvert, D.; Houk, K. N.; Mayo, S. L. Iterative approach to computational enzyme design. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (10), 3790-3795. DOI: doi:10.1073/pnas.1118082108.
(18) Garrabou, X.; Beck, T.; Hilvert, D. A Promiscuous De Novo Retro-Aldolase Catalyzes Asymmetric Michael Additions via Schiff Base Intermediates. *Angew. Chem. Int. Ed.* **2015**, *54* (19), 5609-5612. DOI: https://doi.org/10.1002/anie.201500217.

(19) Broom, A.; Rakotoharisoa, R. V.; Thompson, M. C.; Zarifi, N.; Nguyen, E.; Mukhametzhanov, N.; Liu, L.; Fraser, J. S.; Chica, R. A. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nature Communications* **2020**, *11* (1), 4808. DOI: 10.1038/s41467-020-18619-x.
(20) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., 3rd; et al. De novo computational design of retro-aldol enzymes. *Science* **2008**, *319* (5868), 1387-1391. DOI: 10.1126/science.1152692 From NLM.
(21) Yang, Z.; Yang, S.; Yu, P.; Li, Y.; Doubleday, C.; Park, J.; Patel, A.; Jeon, B.-s.; Russell, W. K.; Liu, H.-w.; et al. Influence of water and enzyme SpnF on the dynamics and energetics of the ambimodal [6+4]/[4+2] cycloaddition. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (5), E848-E855. DOI: doi:10.1073/pnas.1719368115.
(22) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106* (8), 3210-3235. DOI: 10.1021/cr0503106.
(23) Jiménez-Osés, G.; Osuna, S.; Gao, X.; Sawaya, M. R.; Gilson, L.; Collier, S. J.; Huisman, G. W.; Yeates, T. O.; Tang, Y.; Houk, K. N. The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat. Chem. Biol.* **2014**, *10* (6), 431-436. DOI: 10.1038/nchembio.1503.
(24) Yang, L.; Zhang, K.; Xu, M.; Xie, Y.; Meng, X.; Wang, H.; Wei, D. Mechanism-Guided Computational Design of ω-Transaminase by Reprograming of High-Energy-Barrier Steps. *Angew. Chem. Int. Ed.* **2022**, *61* (52). DOI: 10.1002/anie.202212555.
(25) Yang, Z.; Liu, F.; Steeves, A. H.; Kulik, H. J. Quantum Mechanical Description of Electrostatics Provides a Unified Picture of Catalytic Action Across Methyltransferases. *The Journal of Physical Chemistry Letters* **2019**, *10* (13), 3779-3787. DOI: 10.1021/acs.jpclett.9b01555.
(26) Fu, Y.; Chen, H.; Fu, W.; Garcia-Borràs, M.; Yang, Y.; Liu, P. Engineered P450 Atom-Transfer Radical Cyclases are Bifunctional Biocatalysts: Reaction Mechanism and Origin of Enantioselectivity. *J. Am. Chem. Soc.* **2022**, *144* (29), 13344-13355. DOI: 10.1021/jacs.2c04937.
(27) Pan, Y.; Gao, D.; Yang, W.; Cho, H.; Yang, G.; Tai, H.-H.; Zhan, C.-G. Computational redesign of human butyrylcholinesterase for anticocaine medication. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (46), 16656-16661. DOI: doi:10.1073/pnas.0507332102.
(28) García-Meseguer, R.; Ortí, E.; Tuñón, I.; Ruiz-Pernía, J. J.; Aragó, J. Insights into the Enhancement of the Poly(ethylene terephthalate) Degradation by FAST-PETase from Computational Modeling. *J. Am. Chem. Soc.* **2023**, *145* (35), 19243-19255. DOI: 10.1021/jacs.3c04427.
(29) Karvelis, E.; Swanson, C.; Tidor, B. Substrate Turnover Dynamics Guide Ketol-Acid Reductoisomerase Redesign for Increased Specific Activity. *ACS Catalysis* **2024**, *14* (14), 10491-10509. DOI: 10.1021/acscatal.4c01446.
(30) Mou, Z.; Eakes, J.; Cooper, C. J.; Foster, C. M.; Standaert, R. F.; Podar, M.; Doktycz, M. J.; Parks, J. M. Machine learning-based prediction of enzyme substrate scope: Application to bacterial nitrilases. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89* (3), 336-347. DOI: https://doi.org/10.1002/prot.26019.
(31) Ran, X.; Jiang, Y.; Shao, Q.; Yang, Z. J. EnzyKR: a chirality-aware deep learning model for predicting the outcomes of the hydrolase-catalyzed kinetic resolution. *Chemical Science* **2023**, *14* (43), 12073-12082, 10.1039/D3SC02752J. DOI: 10.1039/D3SC02752J.
(32) Venanzi, N. A. E.; Basciu, A.; Vargiu, A. V.; Kiparissides, A.; Dalby, P. A.; Dikicioglu, D. Machine Learning Integrating Protein Structure, Sequence, and Dynamics to Predict the Enzyme Activity

of Bovine Enterokinase Variants. *J. Chem. Inf. Model.* **2024**, *64* (7), 2681-2694. DOI: 10.1021/acs.jcim.3c00999.
(33) Jiang, Y.; Ding, N.; Shao, Q.; Stull, S. L.; Cheng, Z.; Yang, Z. J. Substrate Positioning Dynamics Involves a Non-Electrostatic Component to Mediate Catalysis. *The Journal of Physical Chemistry Letters* **2023**, *14* (50), 11480-11489. DOI: 10.1021/acs.jpclett.3c02444.
(34) Kari, J.; Schaller, K.; Molina, G. A.; Borch, K.; Westh, P. The Sabatier principle as a tool for discovery and engineering of industrial enzymes. *Curr. Opin. Biotechnol.* **2022**, *78*, 102843. DOI: https://doi.org/10.1016/j.copbio.2022.102843.
(35) Jiang, Y.; Yan, B.; Chen, Y.; Juarez, R. J.; Yang, Z. J. Molecular Dynamics-Derived Descriptor Informs the Impact of Mutation on the Catalytic Turnover Number in Lactonase Across Substrates. *J. Chem. Phy. B* **2022**, *126* (13), 2486-2495. DOI: 10.1021/acs.jpcb.2c00142.
(36) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2.
(37) Gaines, C. S.; Piccirilli, J. A.; York, D. M. The L-platform/L-scaffold framework: a blueprint for RNA-cleaving nucleic acid enzyme design. *RNA* **2020**, *26* (2), 111-125, Article. DOI: 10.1261/rna.071894.119.
(38) Zhang, J.; Kulik, H. J.; Martinez, T. J.; Klinman, J. P. Mediation of donor–acceptor distance in an enzymatic methyl transfer reaction. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (26), 7954-7959. DOI: doi:10.1073/pnas.1506792112.
(39) Biler, M.; Crean, R. M.; Schweiger, A. K.; Kourist, R.; Kamerlin, S. C. L. Ground-State Destabilization by Active-Site Hydrophobicity Controls the Selectivity of a Cofactor-Free Decarboxylase. *J. Am. Chem. Soc.* **2020**, *142* (47), 20216-20231. DOI: 10.1021/jacs.0c10701.
(40) Herrera, D. P.; Chánique, A. M.; Martínez-Márquez, A.; Bru-Martínez, R.; Kourist, R.; Parra, L. P.; Schüller, A. Rational Design of Resveratrol O-methyltransferase for the Production of Pinostilbene. *International Journal of Molecular Sciences* **2021**, *22* (9), 4345.
(41) Staudt, A.; Terholsen, H.; Kaur, J.; Müller, H.; Godehard, S. P.; Itabaiana, I., Jr.; Leal, I. C. R.; Bornscheuer, U. T. Rational Design for Enhanced Acyltransferase Activity in Water Catalyzed by the Pyrobaculum calidifontis VA1 Esterase. *Microorganisms* **2021**, *9* (8). DOI: 10.3390/microorganisms9081790 From NLM.
(42) Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; et al. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **2015**, *137* (47), 14905-14911. DOI: 10.1021/jacs.5b07812.
(43) Kaushik, S.; Marques, S. M.; Khirsariya, P.; Paruch, K.; Libichova, L.; Brezovsky, J.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Impact of the access tunnel engineering on catalysis is strictly ligand-specific. *The FEBS Journal* **2018**, *285* (8), 1456-1476. DOI: https://doi.org/10.1111/febs.14418.
(44) Prakinee, K.; Phintha, A.; Visitsatthawong, S.; Lawan, N.; Sucharitakul, J.; Kantiwiriyawanitch, C.; Damborsky, J.; Chitnumsub, P.; Van Pée, K.-H.; Chaiyen, P. Mechanism-guided tunnel engineering to increase the efficiency of a flavin-dependent halogenase. *Nature Catalysis* **2022**, *5* (6), 534-544. DOI: 10.1038/s41929-022-00800-8.
(45) Xu, L.; Nawaz, M. Z.; Khalid, H. R.; Waqar Ul, H.; Alghamdi, H. A.; Sun, J.; Zhu, D. Modulating the pH profile of vanillin dehydrogenase enzyme from extremophile Bacillus ligniniphilus L1

through computational guided site-directed mutagenesis. *Int. J. Biol. Macromol.* **2024**, *263* (Pt 1), 130359. DOI: 10.1016/j.ijbiomac.2024.130359 From NLM.
(46) Janson, G.; Valdes-Garcia, G.; Heo, L.; Feig, M. Direct generation of protein conformational ensembles via machine learning. *Nat. Commun.* **2023**, *14* (1), 774. DOI: 10.1038/s41467-023-36443-x.
(47) Brown, B. P.; Stein, R. A.; Meiler, J.; McHaourab, H. S. Approximating Projections of Conformational Boltzmann Distributions with AlphaFold2 Predictions: Opportunities and Limitations. *J. Chem. Theory Comput.* **2024**, *20* (3), 1434-1447. DOI: 10.1021/acs.jctc.3c01081.
(48) Kleffner, R.; Flatten, J.; Leaver-Fay, A.; Baker, D.; Siegel, J. B.; Khatib, F.; Cooper, S. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* **2017**, *33* (17), 2765-2767. DOI: 10.1093/bioinformatics/btx283 (acccessed 7/24/2024).
(49) Barber-Zucker, S.; Mindel, V.; Garcia-Ruiz, E.; Weinstein, J. J.; Alcalde, M.; Fleishman, S. J. Stable and Functionally Diverse Versatile Peroxidases Designed Directly from Sequences. *J. Am. Chem. Soc.* **2022**, *144* (8), 3564-3571. DOI: 10.1021/jacs.1c12433.
(50) Lee, J.; Kladwang, W.; Lee, M.; Cantu, D.; Azizyan, M.; Kim, H.; Limpaecher, A.; Gaikwad, S.; Yoon, S.; Treuille, A.; et al. RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (6), 2122-2127. DOI: doi:10.1073/pnas.1313039111.
(51) Boby, M. L.; Fearon, D.; Ferla, M.; Filep, M.; Koekemoer, L.; Robinson, M. C.; Consortium‡, T. C. M.; Chodera, J. D.; Lee, A. A.; London, N.; et al. Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. *Science* **2023**, *382* (6671), eabo7201. DOI: doi:10.1126/science.abo7201.
(52) Kozuch, J.; Schneider, S. H.; Zheng, C.; Ji, Z.; Bradshaw, R. T.; Boxer, S. G. Testing the Limitations of MD-Based Local Electric Fields Using the Vibrational Stark Effect in Solution: Penicillin G as a Test Case. *J. Chem. Phy. B* **2021**, *125* (17), 4415-4427. DOI: 10.1021/acs.jpcb.1c00578.
(53) Bhowmick, A.; Sharma, S. C.; Head-Gordon, T. The Importance of the Scaffold for de Novo Enzymes: A Case Study with Kemp Eliminase. *J. Am. Chem. Soc.* **2017**, *139* (16), 5793-5800. DOI: 10.1021/jacs.6b12265.
(54) Ruiz-Pernía, J. J.; Świderek, K.; Bertran, J.; Moliner, V.; Tuñón, I. Electrostatics as a Guiding Principle in Understanding and Designing Enzymes. *J. Chem. Theory Comput.* **2024**, *20* (5), 1783-1795. DOI: 10.1021/acs.jctc.3c01395.
(55) Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T. Computational Optimization of Electric Fields for Improving Catalysis of a Designed Kemp Eliminase. *ACS Catalysis* **2018**, *8* (1), 219-227. DOI: 10.1021/acscatal.7b03151.
(56) Galmés, M. À.; Nödling, A. R.; He, K.; Luk, L. Y. P.; Świderek, K.; Moliner, V. Computational design of an amidase by combining the best electrostatic features of two promiscuous hydrolases. *Chemical Science* **2022**, *13* (17), 4779-4787, 10.1039/D2SC00778A. DOI: 10.1039/D2SC00778A.
(57) Yan, S.; Ji, X.; Peng, W.; Wang, B. Evaluating the Transition State Stabilization/Destabilization Effects of the Electric Fields from Scaffold Residues by a QM/MM Approach. *J. Chem. Phy. B* **2023**, *127* (19), 4245-4253. DOI: 10.1021/acs.jpcb.3c01054.
(58) Bradshaw, R. T.; Dziedzic, J.; Skylaris, C.-K.; Essex, J. W. The Role of Electrostatics in Enzymes: Do Biomolecular Force Fields Reflect Protein Electric Fields? *J. Chem. Inf. Model.* **2020**, *60* (6), 3131-3144. DOI: 10.1021/acs.jcim.0c00217.

(59) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; et al. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory Comput.* **2021**, *17* (4), 2034-2053. DOI: 10.1021/acs.jctc.0c01164.
(60) Bhowmick, A.; Sharma, S. C.; Honma, H.; Head-Gordon, T. The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70. *Physical Chemistry Chemical Physics* **2016**, *18* (28), 19386-19396. DOI: 10.1039/c6cp03622h.
(61) Zaragoza, J. P. T.; Offenbacher, A. R.; Hu, S.; Gee, C. L.; Firestein, Z. M.; Minnetian, N.; Deng, Z.; Fan, F.; Iavarone, A. T.; Klinman, J. P. Temporal and spatial resolution of distal protein motions that activate hydrogen tunneling in soybean lipoxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (10), e2211630120. DOI: doi:10.1073/pnas.2211630120.
(62) Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Computational Molecular Science* **2021**, *11* (3), e1502. DOI: https://doi.org/10.1002/wcms.1502.
(63) Corbella, M.; Pinto, G. P.; Kamerlin, S. C. L. Loop dynamics and the evolution of enzyme activity. *Nature Reviews Chemistry* **2023**, *7* (8), 536-547. DOI: 10.1038/s41570-023-00495-w.
(64) Riziotis, I. G.; Ribeiro, A. J. M.; Borkakoti, N.; Thornton, J. M. Conformational Variation in Enzyme Catalysis: A Structural Study on Catalytic Residues. *J. Mol. Biol.* **2022**, *434* (7), 167517. DOI: 10.1016/j.jmb.2022.167517.
(65) Schenkmayerova, A.; Pinto, G. P.; Toul, M.; Marek, M.; Hernychova, L.; Planas-Iglesias, J.; Daniel Liskova, V.; Pluskal, D.; Vasina, M.; Emond, S.; et al. Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun.* **2021**, *12* (1), 3616. DOI: 10.1038/s41467-021-23450-z.
(66) Yabukarski, F.; Biel, J. T.; Pinney, M. M.; Doukov, T.; Powers, A. S.; Fraser, J. S.; Herschlag, D. Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray–derived conformational ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (52), 33204-33215. DOI: 10.1073/pnas.2011350117.
(67) Wankowicz, S.; Fraser, J. Making sense of chaos: uncovering the mechanisms of conformational entropy. *Chemrxiv* **2023**, 10.26434/chemrxiv-22023-26439b26435k26437. DOI: 10.26434/chemrxiv-2023-9b5k7.
(68) Billeter, S. R.; Webb, S. P.; Agarwal, P. K.; Iordanov, T.; Hammes-Schiffer, S. Hydride Transfer in Liver Alcohol Dehydrogenase: Quantum Dynamics, Kinetic Isotope Effects, and Role of Enzyme Motion. *J. Am. Chem. Soc.* **2001**, *123* (45), 11262-11272. DOI: 10.1021/ja011384b.
(69) Jara, G. E.; Pontiggia, F.; Otten, R.; Agafonov, R. V.; Martí, M. A.; Kern, D. Wide Transition-State Ensemble as Key Component for Enzyme Catalysis. Cold Spring Harbor Laboratory: 2023.
(70) Bonk, B. M.; Weis, J. W.; Tidor, B. Machine Learning Identifies Chemical Characteristics That Promote Enzyme Catalysis. *J. Am. Chem. Soc.* **2019**, *141* (9), 4108-4118. DOI: 10.1021/jacs.8b13879.
(71) Duran, C.; Casadevall, G.; Osuna, S. Harnessing Conformational Dynamics in Enzyme Catalysis to achieve Nature-like catalytic efficiencies: The Shortest Path Map tool for computational enzyme design. *Faraday Discuss.* **2024**, 10.1039/D3FD00156C. DOI: 10.1039/D3FD00156C.
(72) Shao, Q.; Jiang, Y.; Yang, Z. J. EnzyHTP Computational Directed Evolution with Adaptive Resource Allocation. *J. Chem. Inf. Model.* **2023**, *63* (17), 5650-5659. DOI: 10.1021/acs.jcim.3c00618.

(73) Shao, Q.; Jiang, Y.; Yang, Z. J. EnzyHTP: A High-Throughput Computational Platform for Enzyme Modeling. *J. Chem. Inf. Model.* **2022**, *62* (3), 647-655. DOI: 10.1021/acs.jcim.1c01424.
(74) Offenbacher, A. R.; Hu, S.; Poss, E. M.; Carr, C. A. M.; Scouras, A. D.; Prigozhin, D. M.; Iavarone, A. T.; Palla, A.; Alber, T.; Fraser, J. S.; et al. Hydrogen–Deuterium Exchange of Lipoxygenase Uncovers a Relationship between Distal, Solvent Exposed Protein Motions and the Thermal Activation Barrier for Catalytic Proton-Coupled Electron Tunneling. *ACS Cent. Sci.* **2017**, *3* (6), 570-579. DOI: 10.1021/acscentsci.7b00142.
(75) Jara, G. E.; Pontiggia, F.; Otten, R.; Agafonov, R. V.; Martí, M. A.; Kern, D. Wide Transition-State Ensemble as Key Component for Enzyme Catalysis. eLife Sciences Publications, Ltd: 2024.
(76) Otten, R.; Pádua, R. A. P.; Bunzel, H. A.; Nguyen, V.; Pitsawong, W.; Patterson, M.; Sui, S.; Perry, S. L.; Cohen, A. E.; Hilvert, D.; et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **2020**, *370* (6523), 1442-1446. DOI: doi:10.1126/science.abd3623.
(77) Bunzel, H. A.; Anderson, J. L. R.; Hilvert, D.; Arcus, V. L.; van der Kamp, M. W.; Mulholland, A. J. Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nature Chemistry* **2021**, *13* (10), 1017-1022. DOI: 10.1038/s41557-021-00763-6.
(78) Hindson, S. A.; Bunzel, H. A.; Frank, B.; Svistunenko, D. A.; Williams, C.; van der Kamp, M. W.; Mulholland, A. J.; Pudney, C. R.; Anderson, J. L. R. Rigidifying a De Novo Enzyme Increases Activity and Induces a Negative Activation Heat Capacity. *ACS Catalysis* **2021**, *11* (18), 11532-11541. DOI: 10.1021/acscatal.1c01776.
(79) Åqvist, J.; Sočan, J.; Purg, M. Hidden Conformational States and Strange Temperature Optima in Enzyme Catalysis. *Biochemistry* **2020**, *59* (40), 3844-3855. DOI: 10.1021/acs.biochem.0c00705.
(80) Sočan, J.; Purg, M.; Åqvist, J. Computer simulations explain the anomalous temperature optimum in a cold-adapted enzyme. *Nat. Commun.* **2020**, *11* (1), 2644. DOI: 10.1038/s41467-020-16341-2.
(81) van der Ent, F.; Skagseth, S.; Lund, B. A.; Sočan, J.; Griese, J. J.; Brandsdal, B. O.; Åqvist, J. Computational design of the temperature optimum of an enzyme reaction. *Science Advances* **2023**, *9* (26), eadi0963. DOI: doi:10.1126/sciadv.adi0963.
(82) Souza, P. M. d. Application of microbial α-amylase in industry-A review. *Braz. J. Microbiol.* **2010**, *41*, 850-861.
(83) Apic, G.; Gough, J.; Teichmann, S. A. Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* **2001**, *310* (2), 311-325.
(84) Ding, N.; Jiang, Y.; Ge, R.; Ran, X.; Yang, Z. J. Enhance Cold Adaptation of Bidomain Amylases via High-throughput Computational Engineering. *ChemRxiv* **2024**, 10.26434/chemrxiv-22024-rstbz.
(85) Liang, R.; Yu, J. K.; Meisner, J.; Liu, F.; Martinez, T. J. Electrostatic Control of Photoisomerization in Channelrhodopsin 2. *J. Am. Chem. Soc.* **2021**, *143* (14), 5425-5437. DOI: 10.1021/jacs.1c00058.
(86) Carceller, J.; Jayee, B.; Page, C.; Oblinsky, D.; Chintala, N.; Mondragón-Solórzano, G.; Cao, J.; Alassad, Z.; Zhang, Z.; White, N.; et al. Engineering a Photoenzyme to Use Red Light. **2024**, 10.26434/chemrxiv-22024-cjs26435j. DOI: 10.26434/chemrxiv-2024-cjs5j.
(87) Groenhof, G.; Bouxin-Cademartory, M.; Hess, B.; De Visser, S. P.; Berendsen, H. J. C.; Olivucci, M.; Mark, A. E.; Robb, M. A. Photoactivation of the Photoactive Yellow Protein: Why Photon

Absorption Triggers a Trans-to-Cis Isomerization of the Chromophore in the Protein. *J. Am. Chem. Soc.* **2004**, *126* (13), 4228-4233. DOI: 10.1021/ja039557f.
(88) Li, P.; Soudackov, A. V.; Hammes-Schiffer, S. Fundamental Insights into Proton-Coupled Electron Transfer in Soybean Lipoxygenase from Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *J. Am. Chem. Soc.* **2018**, *140* (8), 3068-3076. DOI: 10.1021/jacs.7b13642.
(89) Masgrau, L.; Truhlar, D. G. The Importance of Ensemble Averaging in Enzyme Kinetics. *Acc. Chem. Res.* **2015**, *48* (2), 431-438. DOI: 10.1021/ar500319e.
(90) Zhong, J.; Reinhardt, C. R.; Hammes-Schiffer, S. Role of Water in Proton-Coupled Electron Transfer between Tyrosine and Cysteine in Ribonucleotide Reductase. *J. Am. Chem. Soc.* **2022**, *144* (16), 7208-7214. DOI: 10.1021/jacs.1c13455.
(91) Warburton, R. E.; Soudackov, A. V.; Hammes-Schiffer, S. Theoretical Modeling of Electrochemical Proton-Coupled Electron Transfer. *Chem. Rev.* **2022**, *122* (12), 10599-10650. DOI: 10.1021/acs.chemrev.1c00929.
(92) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111* (52), 18454-18459. DOI: 10.1073/pnas.1417923111.
(93) Frost, C. F.; Balasubramani, S. G.; Antoniou, D.; Schwartz, S. D. Connecting Conformational Motions to Rapid Dynamics in Human Purine Nucleoside Phosphorylase. *J. Chem. Phy. B* **2023**, *127* (1), 144-150. DOI: 10.1021/acs.jpcb.2c07243.
(94) Baldo, A. P.; Tardiff, J. C.; Schwartz, S. D. Mechanochemical function of myosin II: investigation into the recovery stroke and ATP hydrolysis. *J. Chem. Phy. B* **2020**, *124* (45), 10014-10023.
(95) Chen, N.; Wang, S.; Smentek, L.; Hess, B. A.; Wu, R. Biosynthetic Mechanism of Lanosterol: Cyclization. *Angew. Chem. Int. Ed.* **2015**, *54* (30), 8693-8696. DOI: 10.1002/anie.201501986.
(96) Amrein, B. A.; Steffen-Munsberg, F.; Szeler, I.; Purg, M.; Kulkarni, Y.; Kamerlin, S. C. L. CADEE: Computer-Aided Directed Evolution of Enzymes. *IUCrJ* **2017**, *4* (1), 50-64. DOI: doi:10.1107/S2052252516018017.
(97) Roda, S.; Terholsen, H.; Meyer, J. R. H.; Cañellas-Solé, A.; Guallar, V.; Bornscheuer, U.; Kazemi, M. AsiteDesign: a Semirational Algorithm for an Automated Enzyme Design. *The Journal of Physical Chemistry B* **2023**, *127* (12), 2661-2670. DOI: 10.1021/acs.jpcb.2c07091.
(98) Wijma, H. J.; Floor, R. J.; Bjelic, S.; Marrink, S. J.; Baker, D.; Janssen, D. B. Enantioselective Enzymes by Computational Design and In Silico Screening. *Angew. Chem. Int. Ed.* **2015**, *54* (12), 3726-3730. DOI: 10.1002/anie.201411415.
(99) Hon, J.; Borko, S.; Stourac, J.; Prokop, Z.; Zendulka, J.; Bednar, D.; Martinek, T.; Damborsky, J. EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities. *Nucleic Acids Res.* **2020**, *48* (W1), W104-W109. DOI: 10.1093/nar/gkaa372.
(100) Zhou, Q.; Chin, M.; Fu, Y.; Liu, P.; Yang, Y. Stereodivergent atom-transfer radical cyclization by engineered cytochromes P450. *Science* **2021**, *374* (6575), 1612-1616. DOI: doi:10.1126/science.abk1603.
(101) Chen, K.; Arnold, F. H. Engineering new catalytic activities in enzymes. *Nature Catalysis* **2020**, *3* (3), 203-213. DOI: 10.1038/s41929-019-0385-5.

(102) García-Paz, F. D. M.; Del Moral, S.; Morales-Arrieta, S.; Ayala, M.; Treviño-Quintanilla, L. G.; Olvera-Carranza, C. Multidomain chimeric enzymes as a promising alternative for biocatalysts improvement: a minireview. *Mol. Biol. Rep.* **2024**, *51* (1). DOI: 10.1007/s11033-024-09332-9.
(103) Yan, B.; Ran, X.; Gollu, A.; Cheng, Z.; Zhou, X.; Chen, Y.; Yang, Z. J. IntEnzyDB: an Integrated Structure–Kinetics Enzymology Database. *Journal of chemical information and modeling* **2022**, *62* (22), 5841-5848.
(104) Yuan, L.; Ran, X.; Yang, Z.; Zhao, H. OpenEnzymeDB. **2024**.
(105) Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. EnzyDock: Protein–Ligand Docking of Multiple Reactive States along a Reaction Coordinate in Enzymes. *J. Chem. Theory Comput.* **2019**, *15* (9), 5116-5134. DOI: 10.1021/acs.jctc.9b00366.
(106) O'Brien, T. E.; Bertolani, S. J.; Zhang, Y.; Siegel, J. B.; Tantillo, D. J. Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases—bornyl diphosphate synthase as a representative case. *ACS catalysis* **2018**, *8* (4), 3322-3330.
(107) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Merz, K. M., Jr.; Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147* (16), 161727. DOI: 10.1063/1.5001028 From NLM.
(108) Gervasoni, S.; Malloci, G.; Bosin, A.; Vargiu, A. V.; Zgurskaya, H. I.; Ruggerone, P. AB-DB: Force-Field parameters, MD trajectories, QM-based data, and Descriptors of Antimicrobials. *Scientific Data* **2022**, *9* (1), 148. DOI: 10.1038/s41597-022-01261-1.
(109) Notin, P.; Kollasch, A.; Ritter, D.; Van Niekerk, L.; Paul, S.; Spinner, H.; Rollins, N.; Shaw, A.; Orenbuch, R.; Weitzman, R. Proteingym: Large-scale benchmarks for protein fitness prediction and design. *Advances in Neural Information Processing Systems* **2024**, *36*.
(110) Lu, C.; Lubin, J. H.; Sarma, V. V.; Stentz, S. Z.; Wang, G.; Wang, S.; Khare, S. D. Prediction and design of protease enzyme specificity using a structure-aware graph convolutional network. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120* (39), e2303590120.
(111) Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I.; Courbet, A.; de Haas, R. J.; Bethel, N. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **2022**, *378* (6615), 49-56.
(112) Duan, C.; Du, Y.; Jia, H.; Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nature Computational Science* **2023**, *3* (12), 1045-1055. DOI: 10.1038/s43588-023-00563-7.
(113) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Computational Molecular Science* **2018**, *8* (1), e1343. DOI: https://doi.org/10.1002/wcms.1343.
(114) Doga, H.; Raubenolt, B.; Cumbo, F.; Joshi, J.; Difilippo, F. P.; Qin, J.; Blankenberg, D.; Shehab, O. A Perspective on Protein Structure Prediction Using Quantum Computers. *J. Chem. Theory Comput.* **2024**, *20* (9), 3359-3378. DOI: 10.1021/acs.jctc.4c00067.
(115) Pan, X.; Yang, J.; Van, R.; Epifanovsky, E.; Ho, J.; Huang, J.; Pu, J.; Mei, Y.; Nam, K.; Shao, Y. Machine-Learning-Assisted Free Energy Simulation of Solution-Phase and Enzyme Reactions. *J. Chem. Theory Comput.* **2021**, *17* (9), 5745-5758. DOI: 10.1021/acs.jctc.1c00565.
(116) Olehnovics, E.; Liu, Y. M.; Mehio, N.; Sheikh, A. Y.; Shirts, M. R.; Salvalaglio, M. Assessing the Accuracy and Efficiency of Free Energy Differences Obtained from Reweighted Flow-

Based Probabilistic Generative Models. *J. Chem. Theory Comput.* **2024**, *20* (14), 5913-5922. DOI: 10.1021/acs.jctc.4c00520.