

UC Merced

UC Merced Electronic Theses and Dissertations

Title

Active-Site Remodeling Shapes Catalytic Efficiency Across Reactions with Distinct Mechanistic Demands

Permalink

<https://escholarship.org/uc/item/90x4r4fw>

ISBN

9798197875082

Author

Seifinoferest, Behnoush

Publication Date

2026-05-08

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

**Active-Site Remodeling Shapes Catalytic Efficiency Across Reactions
with Distinct Mechanistic Demands**

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Chemistry and Biochemistry

by

Behnoush Seifinoferest

Committee in charge:

Professor Maria Elena Zoghbi, Chair
Professor Andy LiWang
Professor Roberto Chica
Professor Michael Thompson

2026

Copyright

Behnoush Seifinoferest, 2026

All rights reserved.

The dissertation of Behnoush Seifinoferest is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

(Professor Andy LiWang)

(Professor Roberto Chica)

(Professor Michael Thompson)

(Professor Maria Elena Zoghbi, Chair)

University of California, Merced

2026

DEDICATION

To my parents, whose sacrifices shaped my path, and to my siblings,
whose love and support have always lifted me.

EPIGRAPH

"However much you read and learn, Without action, you remain ignorant."

—— Saadi Shiraazi, Bustan

TABLE OF CONTENTS

	Signature Page	iii
	Dedication	iv
	Epigraph	v
	Table of Contents	vi
	List of Figures	x
	List of Tables	xxvii
	Acknowledgements	xxix
	Vita and Publications	xxxix
	Abstract	xxxii
Chapter 1	Introduction	1
	1.1 Enzymes, Strong Biological Catalysts	1
	1.2 Structural Basis of Enzymatic Function	3
	1.3 Enzyme Dynamics Are Important for Catalytic Function	6
	1.4 Enzyme’s Flexibility Can Sometimes be Detrimental to Its Function	10
	1.4.1 The Rigidity–Flexibility Paradox	11
	1.5 Studying Protein Dynamics with X-ray Crystallography	13
	1.5.1 Crystallographic Tools for Dynamic Analysis	16
	1.5.2 Advantages of Room-Temperature X-ray Crystal- lography	22
	1.6 Engineered Enzymes as Structure–Function Relationship Models	23
	1.7 Thesis Objective	26
Chapter 2	Design of Efficient Artificial Enzymes Using Crystallographically Enhanced Conformational Sampling	28
	2.1 Statement of Contribution	28
	2.2 Abstract	29

2.3	Introduction	29
2.4	Results	33
2.4.1	Ensemble-Based Enzyme Design	33
2.4.2	Kinetic Characterization of Designs	39
2.4.3	X-ray Crystal Structures of the Highly Efficient Enzyme	43
2.4.4	Enrichment of Catalytically Competent Substates	53
2.4.5	KE70 Designs	57
2.5	Discussion	60
2.6	Methods	64
2.6.1	Ensemble Generation	64
2.6.2	Computational Enzyme Design	64
2.6.3	Computational Library Design	71
2.6.4	Design Filtering	71
2.6.5	Unconstrained MD Simulation for Filtering of HG3 Designs	73
2.6.6	Protein Expression and Purification	75
2.6.7	Steady-State Kinetics	76
2.6.8	Crystallization	77
2.6.9	X-ray Data Collection and Processing	78
2.6.10	Structure Determination	79
Chapter 3	Directed Evolution Remodels Active Sites to Stabilize Multiple Transition States in a Family of Engineered Enzymes	81
3.1	Statement of Contribution	81
3.2	Abstract	82
3.3	Introduction	82
3.4	Results	87
3.4.1	Structure Determination Using X-ray Crystallography	87
3.4.2	Evolved variants have unique active site cavity shapes	92
3.4.3	Global Rearrangements Indirectly Reshape the Active Sites of High Activity Variants	92
3.4.4	Structural Adaptations Underlying Changes in Active Site Shape	97
3.4.5	Introduction of Bulky Residues Drives Cap Domain Widening	99
3.4.6	Altered Electrostatics and H-bonding Lead to Helix Disorder in BH32.8 and BH _{MeHis} 1.0	101

3.4.7	Active Site Reshaping Allows Multiple Transition States to Be Stabilized Efficiently	103
3.4.8	Substrate-Dependent Movements in OE1.3 Resemble Evolutionary Adaptations in MBH-Catalyzing Families	110
3.4.9	The His23MeHis Substitution Does Not cause Major Structural Changes	114
3.4.10	Crystal Cryocooling Distorts Structural Measurements	114
3.5	Discussion	119
3.6	Future Directions	121
3.6.1	More Efficient Members of the BH32.n Family as a Source of Structural Insight	121
3.6.2	Obtaining Covalently Inhibited Structures for More Variants	124
3.6.3	DFT of Ester Hydrolysis Transition States and Pose Compatibility Analysis	125
3.7	Methods	125
3.7.1	Protein Expression and Purification	125
3.7.2	Crystallization	126
3.7.3	X-ray Data Collection and Processing	127
3.7.4	Structure Determination	128
3.7.5	Ensemble Generation	129
3.7.6	Theozyme placement	129
Chapter 4	Attempts to Capture Non-Equilibrium Reaction Intermediates Using Advanced Crystallographic Approaches	131
4.1	Statement of Contribution	131
4.2	Abstract	132
4.3	Introduction	132
4.4	Results	133
4.4.1	OE1.3 Shows Hinge Bending Upon Covalent Inhibition	133
4.4.2	Apo X-ray Crystallography Structure Shows High Similarity Between OE1.3 and OE1.4	138
4.4.3	Synchrotron Mix-and-Inject Serial Crystallography Can Probe the Catalytic Relevance of This Motion	140
4.4.4	Future Directions	141

4.4.5	ASB1.3 is an Engineered Branch of the BH _{MeHis} Lin- eage with Enhanced Allylic Substitution Activity . . .	144
4.4.6	Time-Resolved Soaking to Capture Enzyme-Substrate Intermediate States	145
4.4.7	Crystal Structures of Apo ASB1.3 and Substrate 1 Complex	150
4.4.8	Structural Comparison Between BH _{MeHis} 1.8 and ASB1.3 Reveals Selective Stabilization of Conformational States	151
4.4.9	ASB1.3 Has an Inherently Open Active Site	151
4.5	Future Directions	157
4.6	Methods	159
4.6.1	Protein Expression and Purification	159
4.6.2	Crystallization	160
4.6.3	X-ray Data Collection and Processing	161
4.6.4	Structure Determination	162
Chapter 5	General Discussion and Conclusions	163
5.1	Discussion	164
5.2	Conclusion	171
Bibliography		173

LIST OF FIGURES

Figure 1.1:	Reaction energy diagram illustrating the effect of enzyme catalysis on activation energy. The red curve represents the uncatalyzed reaction with a higher activation energy barrier, while the blue curve shows the enzyme-catalyzed reaction with a reduced activation energy. The transition state is stabilized in the presence of the enzyme, resulting in an increased reaction rate without altering the overall reaction free energy.	2
Figure 1.2:	Catalytic function in enzymes arises from the combined contributions of chemistry (sequence), three-dimensional structure, and specific molecular movements or dynamics.	5
Figure 1.3:	Schematic representation of an enzyme conformational ensemble, illustrating the existence of multiple interconverting structural states accessible to the protein.	8
Figure 1.4:	Structural changes accompanying the evolution of HG3 toward HG3.17. (A) Active-site interactions between the transition-state analog (6NT, yellow) and key catalytic residues across evolutionary variants. The catalytic base Asp127 is conserved in all variants, while the hydrogen-bond donor at position 50 evolves from Lys50 to His50 and subsequently to Gln50, enabling direct stabilization of the developing oxyanion in the transition state. (B) Local backbone environment underlying residue 50. The backbone of residues 83–84 (shown in white), positioned beneath residue 50, exhibits greater conformational freedom in HG3 and HG3.3b. In HG3.17, hydrogen bonding by Gln50 is associated with reduced backbone flexibility and increased rigidity at this position. (C) Conformational behavior of the loop comprising residues 87–90. This loop is more flexible in earlier variants and becomes progressively more ordered in HG3.17.	12

Figure 1.5:	X-ray crystallography workflow. (A) Experimental pipeline. Purified protein (1) is screened against a range of crystallization conditions (2) to identify optimal conditions for crystal growth. Once suitable crystals are obtained (3), the best-diffracting crystal is exposed to an X-ray beam, and diffraction data are collected (4). (B) Structure determination and model building. Diffraction images collected over a full rotation of the crystal (typically 180°) are indexed and integrated to generate a three-dimensional diffraction dataset (5). The measured structure factor amplitudes are then used in a Fourier transformation to calculate an electron density map (6). An atomic model is built and iteratively refined against this electron density to produce the final molecular model of the protein (7).	15
Figure 1.6:	Crystallographic representations of conformational heterogeneity and dynamics. (A) Atomic positional distribution modeled as a Gaussian, illustrating atomic displacement around an average position. (B) Schematic representation of order parameters, where overall motion (S^2) is decomposed into angular (S^2_{angular}) and orthogonal (S^2_{ortho}) components to account for alternative conformations. (C) Multi-conformer modeling of side-chain conformational heterogeneity within a single structure. (D) Ensemble refinement representation capturing a distribution of conformations consistent with the experimental electron density.	20
Figure 1.7:	Temperature-dependent conformational heterogeneity in Cyclophilin A (CypA). Electron density maps and modeled conformations of Phe113 are shown at 100 K, 240 K, and 310 K. At cryogenic temperature (100 K), Phe113 adopts a single dominant conformation. At intermediate temperature (240 K), partial sampling of an alternative conformation is observed, while at 310 K this alternative conformation becomes more clearly populated. These observations illustrate how increasing temperature enables access to conformational states that are less apparent under cryogenic conditions.	23

Figure 1.8: Overview of directed evolution for enzyme optimization. (A) A target gene is amplified and diversified to generate a DNA library using mutagenesis methods. (B) The library is expressed in a suitable host system. (C) Variants are screened or selected based on catalytic performance. (D) The best-performing variants are isolated for further rounds of evolution. 25

Figure 2.1: Ensemble-based computational enzyme design. (a) The procedure uses one or more theozymes—combinations of catalytic groups arranged to stabilize a TS—and a backbone ensemble representing intrinsic flexibility of the protein fold. The algorithm proceeds in two stages: placement of each theozyme onto each backbone member (bb1–3), followed by active-site repacking to identify sequences (seq1–3) that further stabilize the TS. During repacking, both residue identity and conformation are varied, and the TS is rotated and translated to minimize steric clashes. For clarity, only the theozyme containing a Gln hydrogen-bond donor is shown. Designs from individual theozyme/template combinations are selected based on energy (bold outline) and filtered using structural criteria. (b) Filtering selects designs that reproduce features associated with efficient Kemp eliminases: (1) near-ideal TS–side chain contact geometries (dashed lines), (2) high structural complementarity (low TS solvent-accessible surface area, SASA), (3) preorganized active sites with consistent rotamers in the presence (white/orange) and absence (magenta) of TS, and (4) enrichment of catalytically competent substates that maintain catalytic contacts during short MD simulations. . . 31

Figure 2.2: The Kemp elimination reaction. The Kemp elimination reaction. (a) HG-series Kemp eliminases catalyze the Kemp elimination using a catalytic motif consisting of a base (Asp) that deprotonates the 5-nitrobenzisoxazole substrate, and an H-bond donor (Gln) that stabilizes negative charge buildup on the phenolic oxygen at the transition state ($\ddagger$). This reaction yields the 4-nitro-2-cyanophenol product. (b) Structure of the 6NT transition-state analogue. 32

Figure 2.3:	Ideal geometry for catalytic base contact. Atoms defining various angles describing the hydrogen-bond interaction between the transition state and designed Asp catalytic base (θ , Ψ , χ) are shown in blue. Values in italics are optimal angles calculated for hydrogen bonding interactions between acetamide dimers.	37
Figure 2.4:	Design filtering using molecular dynamics. Following molecular dynamics, angles (θ) and distances (d) defining catalytic contacts between the 5-nitrobenzisoazole substrate and the designed catalytic residues (Asp127 and Gln50) were calculated on each snapshot. Snapshots whose θ and d values fall within the allowed ranges were considered to be catalytically competent. Catalytic competency is the percentage of snapshots in the trajectory that are catalytically competent. Only 9 variants (blue) have a catalytic competency that is greater than that of our positive control HG4 (yellow, PDB ID: 5RGF), an HG3 variant that is approximately 700-fold more catalytically efficient than HG3 [1].	40
Figure 2.5:	Steady-state kinetics of HG-series Kemp eliminases.	42
Figure 2.6:	Ensemble-based design vs directed evolution. (a) Ensemble-based design yields catalytic efficiency improvements comparable to those achieved by directed evolution with a fraction of the experimental screening effort. Full and dashed lines show ensemble-based design and directed evolution trajectories, respectively. (b) Ensemble-based design identifies a majority of the mutation sites (spheres) found by directed evolution (4/6) but mutation combinations of designed (HG649) and evolved (HG3.7) enzymes are distinct as these share only two mutations (K50Q and M84C). (c) Full and dashed lines show ensemble-based design and directed evolution trajectories, respectively. KE703b differs from KE703 by reversion of its D16/N45 residues to the H16/D45 catalytic dyad found in KE70 and its evolved variants. (d) Designed variant KE703b shares 6 of the 9 mutation sites (spheres) found by directed evolution but has only three mutations in common with evolved variant R4 8/ 5A (S20a, Y48F, A204V).	44

Figure 2.7: Crystal structures of HG-series Kemp eliminases (chain A). HG3 (PDB IDs: 5RG4, 5RGA) and HG3.7 (5RG6, 5RGC) are from Broom et al. [2]. (a) Hydrogen bonds to the 6-benzotriazole (6NT) TS analogue are shown as dashed lines with distances; angles are given in Table 2.16. The red sphere indicates a water molecule. The 2Fo–Fc map is shown in volume at 0.5 and 1.5 e Å^{−3} (light and dark blue). HG185 shows ~60% 6NT occupancy, consistent with weaker density. (b) Cutaway view showing improved active-site complementarity to 6NT relative to HG3, reflected by lower solvent-accessible surface areas (mean ± s.d.); 6NT is shown as spheres. (c) Superposition of 6NT-bound (white) with highest (magenta) and lowest (green) occupancy unbound conformers. Non-preorganized residues are marked with an asterisk. In HG3/HG3.7, W44 and M237 adopt nonproductive rotamers; in HG185, C21 adopts a single conformation overlapping the TS analogue; in HG198, HG630, and HG649, only M237 is nonproductive (47–64% occupancy). (d) Ensemble refinement shows reduced structural heterogeneity in HG198, HG630, HG649, and HG3.7 relative to HG3, and a higher fraction of conformers forming catalytic hydrogen bonds (D127, Q50) with 6NT. In HG185, lower 6NT occupancy (~60%) likely contributes to increased heterogeneity and fewer catalytic contacts. 50

Figure 2.8: Computational models vs crystal structures. Crystal structures of HG-series Kemp eliminases with bound 6NT (white) are overlaid on the corresponding design models obtained by single-state design with the average crystal structure or ensemble-based design (blue). For clarity, only the major conformer of Chain A is shown. Root-mean-square deviations between the TS in the design models and the TS analogue in corresponding crystal structures are indicated at the bottom left. The TS and TS analogue are shown at the center of each image. Side chains of all residues forming the binding pocket are shown with the exception of P45, which was omitted for clarity. The sphere shows the carbon of G83. Asterisks indicate residues that adopt side-chain rotamers varying by >20° around one or more side-chain dihedrals between the design model and crystal structure. 52

Figure 2.9:	Ensemble-based design leads to a shift in the conformational ensemble. (a) Distributions of pairwise backbone RMSDs are calculated for all pairs of structures within a single ensemble model. Comparing the distributions derived from efficient HG variants (HG198, HG630, and HG649) bound to 6NT (blue) with those derived from the HG3 ensemble model, also with bound 6NT (white), reveals that the design process leads to a narrower (more self-consistent) distribution of conformations with a lower median pairwise RMSD. This effect is most pronounced for the most active variants (HG630 and HG69). (b) Distributions of pairwise RMSD between the single conformer structure of HG3 (PDB ID: 5RGA[2]) and the individual members of various ensemble models demonstrate that the ensembles of the improved variants have evolved to include conformations that are significantly different from HG3. RMSD values are calculated over active-site residues listed in Table 2.2. Median RMSD values and Z-scores for comparison of distributions are reported in Table 2.9.	56
Figure 2.10:	Multiple sequence alignment of KE70-series Kemp eliminases. . .	58
Figure 2.11:	Steady-state kinetics of KE70-series Kemp eliminases.	62
Figure 2.12:	Multiple sequence alignment of HG-series Kemp eliminases. . . .	74
Figure 3.1:	Three enzyme families evolved from a common ancestor through directed evolution, acquiring specificity for distinct chemical reactions. (A) The crystal structure of BH32 at room temperature, highlighting the cap domain (gray), which contains the catalytic residue His23 (shown as spheres), and the core domain (orange). The schematic to the right shows the different catalytic mechanisms associated with each family. (B) From each family (comprising a range of 4–14 members), two representative enzymes were selected for structural characterization using X-ray crystallography. Mutations present in each representative are mapped onto the BH32 structure. Color coding is used throughout this study: BH32 (grey), BH32.8 (teal blue), BH _{MeHis} 1.0 (velvet), BH _{MeHis} 1.8 (magenta), OE1.3 (emerald green), and OE1.4 (evergreen).	84

Figure 3.2:	KVFinder analysis reveals how evolution reshapes active-site cavity geometry. Top) Active-site surface representations calculated for each enzyme variant illustrate distinct cavity shapes arising from evolutionary changes. Bottom) The corresponding volume and surface area measurements show that, despite similar overall sizes, each enzyme family adopts a unique pocket topology tailored to its catalytic function.	93
Figure 3.3:	Structural comparison of all variants using RMSD and correlation analysis. (A) Pairwise global RMSD values from structural alignment show that BH32.8, BH _{MeHis} 1.0, and BH _{MeHis} 1.8 exhibit the greatest divergence from the ancestral BH32, whereas OE1.3 and OE1.4 remain more structurally similar. (B–D) All-residue correlation maps comparing BH32 with BH32.8 (B), BH _{MeHis} 1.8 (C), and OE1.4 (D). Each map is generated by calculating dot products between displacement vectors of equivalent C α atoms, where blue indicates correlated motion, red indicates opposing motion, and white represents orthogonal or minimal displacement. Black dots denote disordered regions. The bars along the axes indicate domain organization, with the cap domain shown in grey and the core domain in orange.	94
Figure 3.4:	Three categories of structural changes observed in evolved enzymes relative to BH32. BH32 (grey), BH32.8 (teal), OE1.4 (forest green), BH _{MeHis} 1.8 (magenta), and BH _{MeHis} 1.0 (violet). (A) Evolved variants exhibit widening of the cap domain, with BH _{MeHis} 1.8 showing the most pronounced expansion. (B) Intermediate variants such as BH32.8 and BH _{MeHis} 1.0 display a hinge-like opening between the cap and core domains, which partially reverts in BH _{MeHis} 1.8, restoring a more ancestral geometry. (C) Helices spanning residues 180–195 and 200–210 become disordered in BH32.8 and BH _{MeHis} 1.0, and adopt new conformations in BH _{MeHis} 1.8.	96
Figure 3.5:	Comparison of C α B-factor Z-score profiles for BH32, BH _{MeHis} 1.0, and BH _{MeHis} 1.8 (top), with places of the cap and core domains and the secondary structures (from BH32) on the bottom. The vertical dashed purple line indicates the position of the catalytic (Me)His23. To avoid artificial reduction of B-factors arising from multi-conformer modeling, all values were calculated from single-conformer refinements.	97

Figure 3.6: Examples of the structural features underlying cap domain widening in evolved enzymes. A) The angle between the helices formed by residues 16–32 and that formed by residues 37–57 in the cap domain increases in the evolved enzymes, with BH_{MeHis}1.8 showing the largest change. B) The A20Y mutation between BH32 and BH32.8 illustrates how the introduction of bulkier side chains can lead to helix separation. The substitution of Ala20 with Tyr appears to push the helix formed by residues 16–32 outward to avoid steric clashes with pre-existing packing interactions. C) The S22V mutation in BH32.8 introduces a longer side chain (Val22), which likely clashes with Tyr98 on the helix formed by residues 86–99. This clash may drive the outward shift of this helix, providing a clear example of how packing disruptions can lead to structural remodeling. D) The L42W mutation between BH_{MeHis}1.0 and BH_{MeHis}1.8 helps stabilize the positioning of MeHis23, which in turn supports the repositioning of the helix formed by residues 16–32. E) The helix formed by residues 86–99 also shifts outward in the evolved enzymes, further contributing to cap domain widening. 100

Figure 3.7: Disorder and refolding of helices adjacent to the active site across evolution. (A) Structural overlay of BH32 (gray) and BH32.8 (teal) highlights the D123N mutation (inset), which disrupts a salt bridge between Asp123 and Lys184 present in BH32, leading to destabilization of the helix spanning residues 180–195. (B) Distribution of mutations in the BH_{MeHis} lineage mapped onto the BH32 backbone (shown as C α spheres), with BH_{MeHis}1.0 mutations in light purple and BH_{MeHis}1.8 mutations in magenta. The K155Y and I215R mutations are highlighted as key contributors to structural stabilization. (C) Stabilizing interactions in BH_{MeHis}1.8 following refolding: Arg215 anchors the 200–210 loop via hydrogen bonds with Asp200 and Asp210 (top), while Tyr155 forms hydrophobic interactions with Ala186, reducing flexibility in the 180–195 region (bottom). 102

Figure 3.8:	Evaluation of transition state compatibility across BH32 and evolved variants. (A) DFT-derived models of the four transition states (TS1–TS4) of the MBH reaction catalyzed by BH32.14. (B) Pose compatibility, shown as the percentage of backbone conformations capable of accommodating each transition state. (C) Representative orientations of compatible poses within the active-site backbone. (D) RMSD distributions of accepted poses, reflecting the consistency of transition state positioning across the ensemble.	111
Figure 3.9:	Structural changes induced by inhibitor binding in OE1.3 resemble evolutionary changes in BH32.8. A) The inhibitor used in the OE1.3 structure is 2-bromoacetophenone, which forms a covalent bond with the N ϵ of MeHis23. B) Structural alignment of BH32, BH32.8, OE1.3, and inhibited OE1.3 (aligned on cap domain) reveals that the inhibited OE1.3 structure shares several features with BH32.8, including cap domain widening, downward movement of helix 86–99, and a torsional shift in the core domain, although the nature of the motion differs. C) C α displacement vector correlation matrices comparing BH32 vs. BH32.8 and OE1.3 vs. inhibited OE1.3 show strikingly similar global rearrangements, suggesting that evolution in BH32.8 and inhibitor binding in OE1.3 lead to comparable structural effects. D) Like BH32.8, the inhibited OE1.3 structure shows increased backbone flexibility in the 180–195 region, as indicated by elevated C α B-factor Z-scores. As a result, this region is partly disordered and unmapable in the electron density.	112

Figure 3.10: Possible coupling between frontal loop disorder and hinge-like motions in the protein structure. (A) Structural alignment of BH32 and BH32.8, anchored on the cap domain, reveals an upward shift of helix 86–99 that propagates through adjacent helices, including 124–136, 138–142, 104–117, and 219–230. (B) Close-up views of inter-helical interactions show that packing rearrangements and a flexible linker between helices 124–136 and 138–142 enable transmission of this motion across the structure. (C) In BH_{MeHis}1.8, reduced hinge opening is observed and can be attributed to the I215R mutation, which forms stabilizing interactions with loop 200–214 and the β -sheet spanning residues 214–216. These interactions restrict the upward movement of helices 219–230, 104–117, and 138–142. (D) Despite this local restriction, the flexible linker decouples helix 124–136 from the constrained region, allowing helix 124–136 and helix 86–99 to retain similar upward shifts as observed in BH32.8 and BH_{MeHis}1.0. 115

Figure 3.11: Minimal structural impact of His23 methylation between BH32.8 and BHMeHis1.0. (A) Overlay of BH32.8 (teal) and BHMeHis1.0 (violet) shows highly similar overall structures, with no significant backbone rearrangements upon introduction of N δ -methylhistidine at position 23. The inset highlights the active site, where His23 adopts different rotamers in the two variants as a result of methylation, with BH32.8 sampling multiple conformations and BHMeHis1.0 favoring a single dominant rotamer. (B) Close-up views of the active site illustrate the local packing environment around His23. Addition of the methyl group leads to slightly tighter packing compared to the very loosely packed environment in BH32.8, although the active site remains overall relatively open. (C) C α B-factor Z-score profiles for BH32.8 (teal) and BHMeHis1.0 (violet) are largely overlapping across the sequence, indicating that methylation does not significantly alter global or local backbone dynamics. 116

Figure 3.12: Effect of temperature on structural modeling and residue dynamics. (A) RMSD between 100 K and 300 K structures (0.25–0.30 Å) indicates altered conformational sampling. (B) Order parameter profiles reveal temperature-dependent changes in the overall dynamic behavior of residues across variants. (C) Structure alignment anchored on the cap domain illustrate temperature effects on backbone conformation; in BH32, a hinge opening is observed at 300 K (dark grey) but not at 100 K (light grey). (D) The catalytic (Me)His adopts different conformations at 100 K and 300 K: in BH32.8 (teal vs green) and BH _{MeHis} 1.0 (dark vs light velvet). (E) Regions 180–195 and 200–210 are affected by cryocooling. In BH _{MeHis} 1.0, region 200–210 is resolved at 300 K (dark violet) but unmappable at 100 K (light violet); in BH _{MeHis} 1.8, this loop adopts different conformations at 100 K (light pink) and 300 K (magenta). (F) Side-chain conformations near the active site are also altered ((Me)His in yellow): in BH32.8, the alternative conformation of Trp88 is present at 300 K (teal) but absent at 100 K (green); in OE1.4, Arg50 adopts distinct conformations at 100 K (light green) and 300 K (dark green).	118
Figure 3.13: Crystallization screening of BH32.14 under a range of conditions, including standard NexTal screens, seeding, additive screening, and temperature optimization. Although crystals of moderate quality were obtained (A), they did not yield datasets suitable for structural modeling (B).	122
Figure 3.14: Evolution from BH32 to BH32.8 (and similarly BH _{MeHis} 1.0) leads to destabilization of helical regions spanning residues 180–195 and 200–210, resulting in increased conformational heterogeneity. In BH _{MeHis} 1.8, these regions become partially restabilized into flexible loop conformations. Structural characterization of BH32.14 or BH32.12 would provide a direct comparison to determine how alternative conformational states are selected and stabilized along the BH32.n evolutionary pathway.	123

Figure 4.1:	OE1.3. (A) The OE family catalyzes an ester hydrolysis reaction, which is traceable through the emission of a fluorescein group. (B) 2-bromoacetophenone acts as a covalent inhibitor and shares structural similarity with the substrate. (C) Upon reaction, the inhibitor loses the bromine atom and forms a covalent bond with the N ϵ of MeHis23. (D) Structural comparison (cryo structures from Burke et al.) shows that the bulky inhibitor induces a downward shift of the core domain (bottom domain), resulting in a hinge-like motion; the inhibited structure is shown in red and the apo structure in white. (E) The white mesh represents the shape of the active site in the absence of the inhibitor, while the red surface shows the expanded active site upon inhibition, indicating that the conformational change significantly increases the accessible volume.	135
Figure 4.2:	(A) Representative crystals of OE1.3 and OE1.4 obtained from optimized crystallization conditions derived from NexTal screening. (B) Structural alignment of OE1.3 (light green) and OE1.4 (forest green), showing a high degree of similarity (RMSD = 0.176 Å). (C) Comparison of cryogenic and room-temperature structures highlights conformational differences in the flexible linker region connecting the two domains, indicating the presence of cryocooling-induced artifacts.	139
Figure 4.3:	Schematic demonstration of the mix and inject serial crystallography experimental setting. Data is collected from a slurry of microcrystals at certain time points after being mixed with the substrate.	140
Figure 4.4:	Preparation of OE1.4 microcrystals for time-resolved crystallography experiments. (A) Optimized crystallization condition producing large single crystals of OE1.4. (B) Modification of the same condition to favor nucleation over growth, resulting in a dense population of microcrystals. (C) Schematic and experimental setup of batch crystallization [3], in which concentrated protein and reservoir solutions are rapidly mixed to promote uniform microcrystal formation. This approach yielded microcrystal densities exceeding 17,000 crystals per mL, suitable for mix-and-inject serial crystallography experiments.	142

Figure 4.5:	Evolutionary divergence from a common scaffold and distribution of mutations across the protein structure. (A) Directed evolution starting from BH32 toward the Morita–Baylis–Hillman (MBH) reaction produces BH _{MeHis} 1.0, followed by continued evolution to BH _{MeHis} 1.8. A parallel evolutionary trajectory from BH _{MeHis} 1.6 (ASB1.0) leads to ASB1.3 for allylic substitution (AS), illustrating divergence from a shared intermediate scaffold. (B) C _α positions of mutations accumulated during evolution mapped onto the BH32 scaffold. Mutations from BH _{MeHis} 1.0, BH _{MeHis} 1.6, BH _{MeHis} 1.8, and ASB1.3 are shown in different colors, illustrating how distinct evolutionary pathways redistribute mutations across the structure while maintaining a common fold.	143
Figure 4.6:	Substrates and reaction mechanism for ASB1.3-catalyzed allylic substitution. (A) Structure of substrate 1, containing a p-nitrophenol (PNP) leaving group that enables spectroscopic monitoring at 400 nm. (B) Structure of nucleophile a used as the second substrate. (C) Proposed reaction mechanism, in which the catalytic methylhistidine (E-MeHis23) forms a covalent intermediate with substrate 1, followed by nucleophilic attack by substrate a to yield the final product.	146
Figure 4.7:	Schematic representation of the experimental workflow. Crystals are first soaked in substrate 1 to generate the intermediate state, followed by introduction of substrate a. Structural data are then collected at defined time points (t1–t4) to capture progression along the reaction coordinate.	149
Figure 4.8:	Examples of ASB1.3 crystal morphology	150

Figure 4.9: Selective stabilization of distinct conformations in BH_{MeHis}1.8 and ASB1.3 from a shared dynamic ancestor. (A) Structural comparison of regions 180–195 (left) and 200–210 (right) across BH32, BH_{MeHis}1.8, and ASB1.3. In BH32, both regions adopt well-defined helical conformations. In BH_{MeHis}1.8, these regions become flexible loops, reflecting increased conformational heterogeneity. In ASB1.3, while the 180–195 region remains flexible, the 200–210 region partially regains its helical structure, indicating selective stabilization of a specific conformation. (B) The G11S mutation in ASB1.3 contributes to the reformation of the helical structure in the 200–210 region by introducing a network of stabilizing hydrogen bonds. (C) A conserved π -system at position 155 in ASB1.3 and BH_{MeHis}1.8 enables hydrophobic interaction with A186, stabilizing the 180–195 loop. In BH32, residue 155 is positioned within a helix and remains distant from A186; this interaction becomes possible only after local conformational rearrangement and increased flexibility in the intermediate BH_{MeHis}1.0 state. Together, these observations highlight how evolution selects and stabilizes distinct conformations from an initially heterogeneous structural ensemble. 152

Figure 4.10: Crystal contacts (red) in ASB1.3 apo (white) shows that chain A has been less affected, and will be used for the rest of the analysis (A) Chain A front (on the left) and back (on the right) view, (B) Chain B front (on the left) and back (on the right) view 154

Figure 4.11: Minimal structural and dynamic changes upon substrate binding in ASB1.3. (A) Structural alignment of chain A from the apo structure (cyan) and the substrate-bound structure (orange) shows a high degree of similarity between the two states. Differences are primarily localized to flexible loop regions (highlighted), with no significant rearrangement of the overall fold or active site architecture. The inset shows electron density corresponding to the bound substrate in the active site, confirming its position without requiring substantial backbone reorganization. A small helical displacement behind the active site is observed, but its presence in only one chain of the asymmetric unit suggests that it is influenced by crystal packing rather than being functionally relevant. (B) C α B-factor Z-score profiles for the apo (cyan) and substrate-bound (orange) structures largely overlap across the sequence, indicating that substrate binding does not significantly alter global or local backbone dynamics. 155

Figure 4.12: Comparison of hinge-bending conformations across related enzyme variants. ASB1.3 adopts a more open hinge-bent conformation compared to BH32.8 (blue). While OE1.3 in the inhibitor-bound state also exhibits hinge bending, its motion includes an additional twisting component that is not observed in ASB1.3. When considering the hinge-bending component alone, ASB1.3 appears to adopt a more open conformation than OE1.3 in its inhibited state, highlighting subtle but important differences in the nature of conformational changes across these systems. 156

Figure 4.13: Structural features of ASB1.3 reveal an inherently open and pre-organized active site. (A) Structural alignment of BH32 (gray), BH_{MeHis}1.8 (magenta), ASB1.3 apo (cyan), and ASB1.3 bound to substrate 1 (orange) anchored on cap domain, showing that ASB1.3 adopts a hinge-bent conformation similar to previously observed intermediates in related systems. Both apo and bound forms overlap closely, indicating that this conformation is intrinsic and not induced by substrate binding. (B) Active site cavity calculated using KVFinder for ASB1.3 apo (cyan) and substrate-bound (orange) structures. The bound state shows a more expanded cavity despite minimal global structural differences. (C) Close-up view of the active site entrance highlighting residues such as Arg50 and Lys184, which adopt slightly different conformations between apo and bound states, potentially contributing to substrate access or gating. (D) Overlay of key active site residues in apo and bound structures, showing strong structural conservation and preorganization of the catalytic environment. The ligands are shown in white. 158

Figure 5.1: A crystallographic ensemble of the low-efficiency enzyme was used as the starting point for multi-state design, leading to variants with up to a 250-fold increase in catalytic efficiency without the need for directed evolution or other labor-intensive experimental approaches. 167

Figure 5.2: Three enzyme families evolved from a common ancestor (BH32) to catalyze distinct reactions, illustrating how structural and dynamic changes reshape catalytic function. **Top panel:** Schematic representation of the divergent evolution of the BH32 scaffold into three families (BH32.n [4], BH_{MeHis}1.*n* [5], and OE1.n [6]), each adapted to a different reaction mechanism. Key catalytic residues and reaction steps are highlighted to emphasize mechanistic differences across families. **Bottom left:** Room-temperature X-ray crystal structures of representative members (BH32, BH32.8, BH_{MeHis}1.8, and OE1.4) with their active-site cavities shown as colored surfaces, demonstrating how evolution alters the size and shape of the active site. **Bottom right:** Pose-compatibility analysis of DFT-derived transition states (TS1–TS4) calculated for MBH by BH32.14 and the conformational ensembles of each enzyme. BH32.8 shows increased compatibility with multiple transition states compared to the ancestor. 168

LIST OF TABLES

Table 2.1:	Ensemble refinement results	33
Table 2.2:	Amino-acid positions optimized during computational design of HG3 variants.	35
Table 2.3:	Design filtering - HG series.	36
Table 2.4:	Angles defining the hydrogen bond between D127 and the transi- tion state or transition-state analogue for various Kemp eliminase structures	38
Table 2.5:	Kinetic Parameters of Kemp Eliminases	41
Table 2.6:	Crystallization conditions	45
Table 2.7:	Crystallographic data and refinement statistics (HG185 and HG198)	46
Table 2.8:	Crystallographic data and refinement statistics (HG630 and HG649)	48
Table 2.9:	Structural comparison of HG3 crystal structure with ensembles of HG-series variants	54
Table 2.10:	Design filtering _ KE70 series	59
Table 2.11:	Geometric constraints used to define catalytic contacts during com- putational design	65
Table 2.12:	Geometric definitions for generation of transition-state poses off the side chain of catalytic base	66
Table 2.13:	Amino-acid sequences of Kemp eliminases from HG series	67
Table 2.14:	Amino-acid positions optimized during computational design of KE70 variants	70
Table 2.15:	Amino-acid sequences of Kemp eliminases from KE70 series	72
Table 2.16:	Angles of hydrogen bonds between transition state analogue 6NT and catalytic residues in Kemp eliminase crystal structures	73
Table 3.1:	Sequences and mutations of enzyme variants.	86
Table 3.2:	Crystallization conditions for all enzyme variants. *seeding used.	88
Table 3.3:	Crystallographic data and refinement statistics for BH32, BH32.8, and BH _{MeHis} 1.0.	90
Table 3.4:	Crystallographic data and refinement statistics for BH _{MeHis} 1.8, OE1.3, and OE1.4.	91
Table 3.5:	Melting temperatures for all enzyme variants.	98
Table 3.6:	Ensemble refinement results for the room temperature structures used in theozyme placement	104
Table 3.7:	Geometric definitions for generation of transition-state poses off the side chain of catalytic residue.	104

Table 3.8:	Geometric constraints used to define catalytic contacts.	107
Table 4.1:	Crystallization conditions for OE1 variants.	136
Table 4.2:	Crystallographic data and refinement statistics for OE1.3 and OE1.4.	137
Table 4.3:	Crystallization conditions for ASB1.3 variants.	145
Table 4.4:	Crystallographic data and refinement statistics for ASB1.3 K39A .	147

ACKNOWLEDGEMENTS

I am deeply thankful to my advisor, Dr. Michael Thompson, for his invaluable mentorship, support, and guidance throughout this journey.

I would also like to sincerely thank my committee members, Dr. Roberto Chica, Dr. Andy LiWang, and Dr. Maria Elena Zoghbi, for their insight, encouragement, and thoughtful guidance throughout my PhD. I am also grateful to my former committee member, Dr. Shahar Sukenik, for his support and contributions during an important stage of this work.

My sincere thanks also go to our collaborators: Dr. Rojo Rakotoharisoa and Dr. Hang Pham from the University of Ottawa, and Dr. Anthony Green and Dr. Zachary Birch-Price from the University of Manchester, for their essential contributions to enzyme design, evolution, and analysis.

I would like to express my deepest gratitude to my parents, whose tireless strength and sacrifices gave me the freedom to choose my own path. Their resilience in the face of life's challenges has been the foundation of everything I have achieved. To my father, whose pride has been the motivation behind every step I have taken, and to my mother, from whom I learned to care deeply about humanity.

To my siblings, Bahare, Behnam, and Behnaz, for all our shared memories and for never letting me feel alone. To my extended family, Hamid, Reza, and Farzaneh, for being leaders, role models, and constant supporters, and for defining the true meaning of family. To Aven, Arteen, and Arvand, for a love unlike anything I have ever experienced.

To my professional family, my labmates: I am especially grateful to Dr. Alexander Wolff, who was both a mentor and a lasting influence. To Joshua Rodriguez for being the perfect senior, Alexander Deary for always being willing to help, Azeem Horani for being the perfect office mate, and Alec Martinez for his diligence and for making sure we never lacked anything in the lab. To my dear friends who lifted me up and never let me fall, Laura, Maryam, Luisa, Andree, Azin, Kunle, Arefeh,

and Hossein, thank you for your constant support and presence in my life. To Zahra Fayaziboroujeni, a wonderful friend and mentee, thank you for your energy, kindness, and eagerness to grow.

To my collaborators in making a dream come true, my fellow Grad Nomads officers, Camelia Haghvardian, Sneha Buhat, and Achint Sharma, thank you for your dedication and shared vision.

And last but not least, to the Thompson lab's wonderful undergraduate students, Emma Gomez and Sarah Ashrafi, who became dear friends, and Edith Gollub, Asini Taka, and Janani Pulivarthi, for their contributions to this research and to the completion of this thesis.

VITA

- 2015 B. S. in Chemistry, Kharazmi University, Tehran, Iran
- 2019-2007 M. Sc. in Organic Chemistry, University of Tehran,
Tehran, Iran
- 2026 Ph.D. in Chemistry and Biochemistry, University of California, Merced, CA, USA

PUBLICATIONS

Kianmehr, E., Seifinoferest, B. and Afaridoun, H., "Palladium-Catalyzed Regioselective Acylation of Diazines with Toluenes: A New Approach to the Synthesis of ortho-Diacylbenzenes", *European Journal of Organic Chemistry*, vol. 2020, pp.4925-4930, (2020).

Seifinoferest, B., Tanbakouchian, A., Larijani, B. and Mahdavi, M., "Ullmann-Goldberg and Buchwald-Hartwig C N Cross Couplings: Synthetic Methods to Pharmaceutically Potential N-Heterocycles", *Asian Journal of Organic Chemistry*, vol. 10, pp.1319-1344 (2021).

Farid, S.M., Seifinoferest, B., Gholamhosseyni, M., Larijani, B. and Mahdavi, M., "Modern metal-catalyzed and organocatalytic methods for synthesis of coumarin derivatives: a review.", *Organic Biomolecular Chemistry*, vol. 20, pp.4846-4883 (2022).

Rakotoharisoa, R.V., Seifinoferest, B., Zarifi, N., Miller, J., Rodriguez J.M., Thompson, M. C., Chica, R.A., "Design of efficient artificial enzymes using crystallographically-enhanced conformational sampling.", *Journal of American Chemical Society*, 146, pp. 10001-10013. (2024)

ABSTRACT OF THE DISSERTATION

Active-Site Remodeling Shapes Catalytic Efficiency Across Reactions with Distinct Mechanistic Demands

by

Behnoush Seifinoferest

Doctor of Philosophy in Chemistry and Biochemistry

University of California Merced, 2026

Professor Maria Elena Zoghbi, Chair

Understanding the relationship between enzyme structure and function is a central goal in structural biology, with important implications for enzyme design, drug discovery, and biotechnology. In this dissertation, we investigate the role of structural flexibility and dynamics in enzyme catalysis using engineered enzymes as model systems and advanced X-ray crystallography techniques.

Our results demonstrate that optimal catalytic function arises from a balance between rigidity and flexibility. In relatively simple, single-step reactions, increased rigidity can enhance catalysis by preorganizing the active site to closely match the transition state geometry, thereby reducing the need for conformational rearrangement. In contrast, more complex, multi-step reactions require enzymes to sample multiple conformational states to accommodate substrates, intermediates, and products throughout the catalytic cycle. In these cases, conformational flexibility becomes essential for efficient function. Through structural and comparative analyses of evolved enzyme variants, we show how directed evolution modulates the conformational landscape of enzymes, enabling expansion or reshaping of the active site and

selectively enriching catalytically productive states. Additionally, time-resolved and room-temperature crystallographic approaches provide insight into non-equilibrium conformations and transient states that are not accessible through conventional cryogenic methods. Together, these findings provide a unified framework for understanding how enzymes balance rigidity and flexibility to achieve efficient catalysis, and highlight the importance of dynamics in shaping enzyme function.

Chapter 1

Introduction

1.1 Enzymes, Strong Biological Catalysts

Life relies on an intricate network of chemical reactions that maintain cellular structure, generate energy, and process information. While many of these reactions are thermodynamically favorable, they are often far too slow under physiological conditions to sustain life on their own. To solve this problem, nature has evolved highly efficient biological catalysts known as enzymes [7]. These molecules can accelerate chemical reactions by many orders of magnitude while maintaining remarkable specificity [8].

For any reaction to occur, reactants must pass through a short-lived, high-energy configuration known as the **Transition State** (Figure 1.1) [9]. The energy difference between the ground state of the reactants and this transition state defines the activation energy barrier. Without a catalyst, this barrier can be too high for the reaction to proceed at biologically relevant rates. Enzymes increase reaction rates by preferentially stabilizing the transition state relative to the reactants, effectively lowering the activation energy required for product formation.

At the molecular level, the transition state represents a highly unstable con-

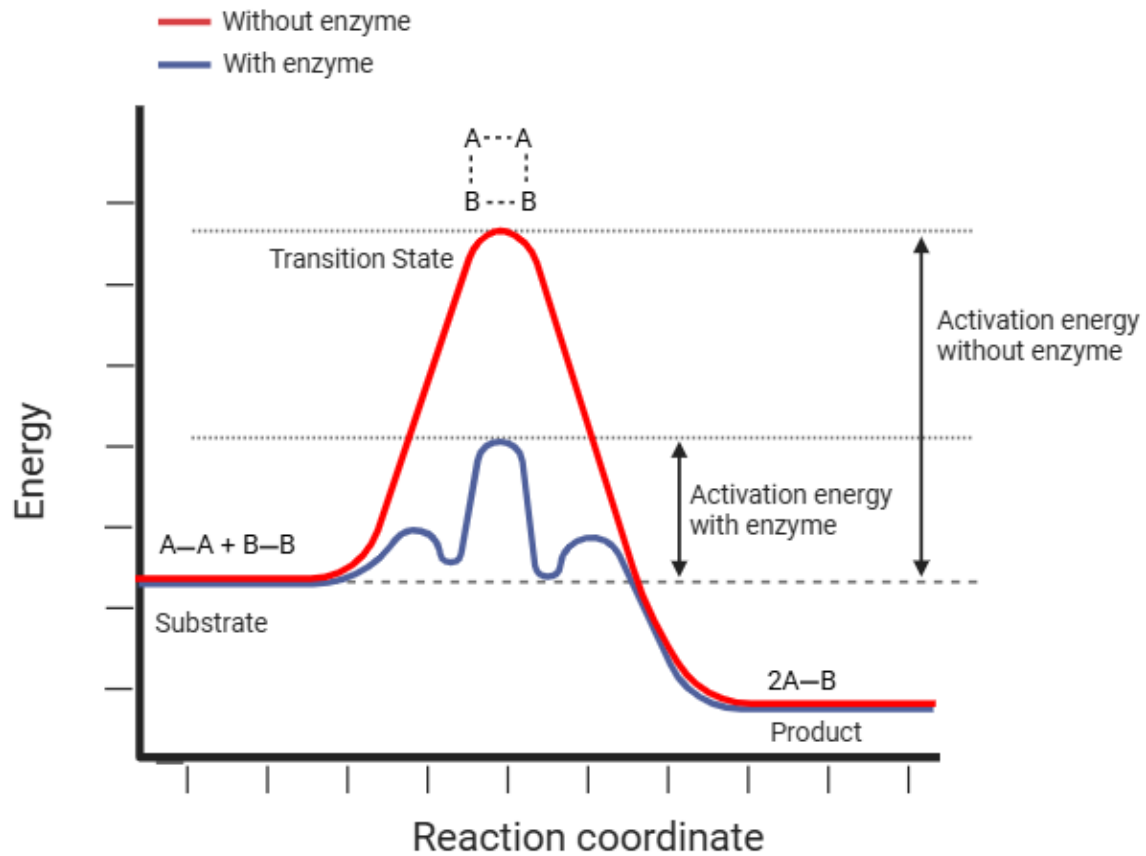


Figure 1.1: Reaction energy diagram illustrating the effect of enzyme catalysis on activation energy. The red curve represents the uncatalyzed reaction with a higher activation energy barrier, while the blue curve shows the enzyme-catalyzed reaction with a reduced activation energy. The transition state is stabilized in the presence of the enzyme, resulting in an increased reaction rate without altering the overall reaction free energy.

figuration in which bonds present in the reactants are partially broken while new bonds characteristic of the products are only partially formed. The energetic cost of reaching this state constitutes the primary rate-limiting factor for most chemical reactions. The fundamental role of an enzyme is therefore not to stabilize substrates or products, but rather to preferentially recognize and stabilize this transition state [10].

Enzymes achieve this stabilization by providing an active site that is complementary to the transition state both chemically and geometrically, enabling strong and highly specific interactions. As described by Albert L. Lehninger in his classic biochemistry textbook, enzymes function by “providing a surface that complements the transition state in stereochemistry, polarity, and charge” [10, 11]. Through this selective stabilization, enzymes can accelerate reactions that would otherwise take years to occur spontaneously, enabling them to proceed on timescales ranging from milliseconds to seconds under physiological conditions [10].

1.2 Structural Basis of Enzymatic Function

At the most fundamental level, the chemistry of the active site plays a direct and essential role in enzymatic catalysis, as the identity and reactivity of the functional groups on catalytic residues ultimately determine the chemical transformations that can occur [12, 13]. However, decades of biochemical, structural, and biophysical studies have demonstrated that efficient catalysis arises from the ability of the protein scaffold to stabilize the transition state through precise preorganization of these functional groups in space [14, 15].

One of the most straightforward examples of this principle is electrostatic stabilization, in which negatively charged or polarized transition-state atoms are stabilized by appropriately positioned hydrogen-bond donors. Short, strong hydrogen bonds have been shown to play a significant role in active-site stabilization in many enzyme

systems [16, 17]. More broadly, the active site can provide nucleophilic, acidic, and basic interactions by positioning the appropriate residues in close proximity to the reacting groups [18, 19]. In addition, hydrophobic and packing interactions contribute to substrate binding, enforce geometries favorable for reaction, and help stabilize the transition state [20]. Enzymes can also facilitate proton transfer reactions, where precise positioning of donor and acceptor groups has been shown to significantly enhance catalytic efficiency [21].

Electrostatic effects are particularly important because many biological reactions involve charge rearrangements. A chemical reaction can be accelerated by an electric field if the charge distribution (dipole moment) of the system changes upon formation of the transition state. In such cases, an electric field aligned with the transition-state dipole can stabilize it relative to the ground state and lower the activation barrier [22]. This concept highlights how the protein environment can actively contribute to catalysis beyond simple positioning of residues.

Another key factor in enzymatic catalysis is entropy [23]. A common misconception is that enzymes function primarily by bringing reacting groups into close proximity. However, this effect alone can account for only modest rate enhancements [24]. Instead, enzymes reduce the entropic cost of reaching the transition state by restricting the conformational freedom of the substrate and selecting geometries that closely resemble the transition-state configuration. Early work by Page and Jencks [25] emphasized the role of entropy in reducing the number of accessible conformations, while studies by Bruice and Pandit [26] demonstrated that constraining reacting groups into near-attack conformations can dramatically accelerate reaction rates by effectively prepaying the entropic penalty associated with transition-state formation.

Beyond defining the geometry of the active site, the overall structure establishes the spatial relationships between residues involved in substrate recognition, transition-state stabilization, and product release [27]. Even residues distant from

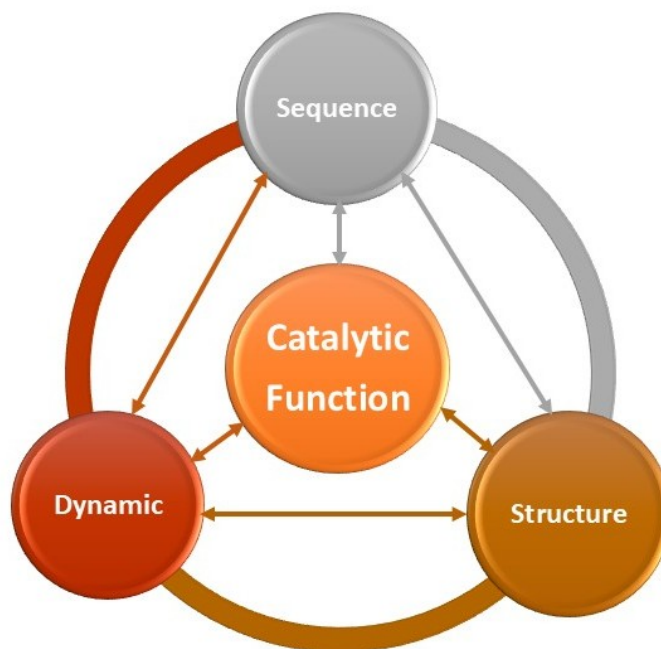


Figure 1.2: Catalytic function in enzymes arises from the combined contributions of chemistry (sequence), three-dimensional structure, and specific molecular movements or dynamics.

the active site (distal residues) can influence catalysis by shaping the overall fold, restricting or permitting conformational changes, and controlling access to the catalytic core [28, 29]. As a result, enzymatic activity reflects an intimate coupling between protein structure and chemical function, rather than a purely local chemical phenomenon.

1.3 Enzyme Dynamics Are Important for Catalytic Function

Recognizing the importance of molecular shape and structural complementarity has helped enzymologists achieve a clearer understanding of the mechanisms through which enzymes function. However, advances in structural biology have revealed a third critical factor in enzymatic catalysis: structural heterogeneity and dynamics (Figure 1.2).

In reality, enzymes are dynamic molecules that populate a hierarchy of interconverting conformations across a wide range of timescales, from femtoseconds to minutes [30, 31]. Rather than existing in a single structure, an enzyme occupies a conformational ensemble (Figure 1.3), with multiple accessible states that are close in energy and can interconvert with one another [32]. This conformational ensemble plays a central role in catalytic function. Enzymes must be able to bind substrates, carry out the chemical transformation, and release products, and these steps often require distinct structural arrangements. As a result, flexibility is necessary to allow transitions between these functional states [29, 28].

While some conformational substates may not possess sufficient catalytic capacity, enrichment of catalytically productive conformations can enhance enzymatic efficiency [2]. At the same time, conformational heterogeneity is often essential for function, as it enables access to multiple structural states required throughout the catalytic cycle.

This requirement becomes even more pronounced in enzymes that catalyze multi-step reactions. In such systems, each step along the reaction coordinate may involve a distinct transition state with unique geometric and electronic requirements. Consequently, the enzyme must adopt different conformations to stabilize each of these states [33, 34, 35]. In this view, catalysis does not arise from a single static structure, but from a dynamic distribution of conformations with varying propensities toward

chemistry supporting varying demands across the reaction coordinate.

Despite decades of work, fully understanding how these conformational motions contribute to catalysis remains challenging. In particular, the need to balance precise positioning for transition-state stabilization with sufficient flexibility to enable substrate binding, product release, and progression along the reaction coordinate makes it difficult to distinguish mechanistically relevant motions from incidental fluctuations.

Many of the structural states that enzymes sample are already accessible in their ground, or apo, forms. These intrinsic motions arise from the underlying energy landscape of the protein and reflect the ensemble of conformations encoded in its fold. Some of these motions may be neutral with respect to catalysis, while others may facilitate specific steps of the reaction cycle. In this sense, catalytically relevant dynamics can be an inherent property of the enzyme rather than a consequence of ligand binding.

A well-established example of intrinsic enzyme dynamics comes from the work of Dorothee Kern’s research group at The Scripps Research Institute. In studies published in *Science* in 2002 [35] and in *Nature* in 2005 [36], they investigated the prolyl cis–trans isomerase cyclophilin A (CypA) using NMR relaxation dispersion techniques. CypA catalyzes the cis–trans isomerization of prolyl peptide bonds, plays a role in HIV-1 virulence, and is a target of the immunosuppressive drug cyclosporin A (CsA) [37].

By examining CypA both in its apo form and during catalysis of the model substrate N-succinyl-Ala-Phe-Pro-Phe-p-nitroanilide, they observed that the enzyme interconverts between multiple conformational states throughout the catalytic cycle, including the free enzyme (E) and enzyme–substrate (ES) complexes with substrate in cis and trans conformations. Strikingly, the conformational exchange processes detected during catalysis were already present in the apo enzyme. Furthermore, the rates of these intrinsic motions closely matched the catalytic turnover rate. This

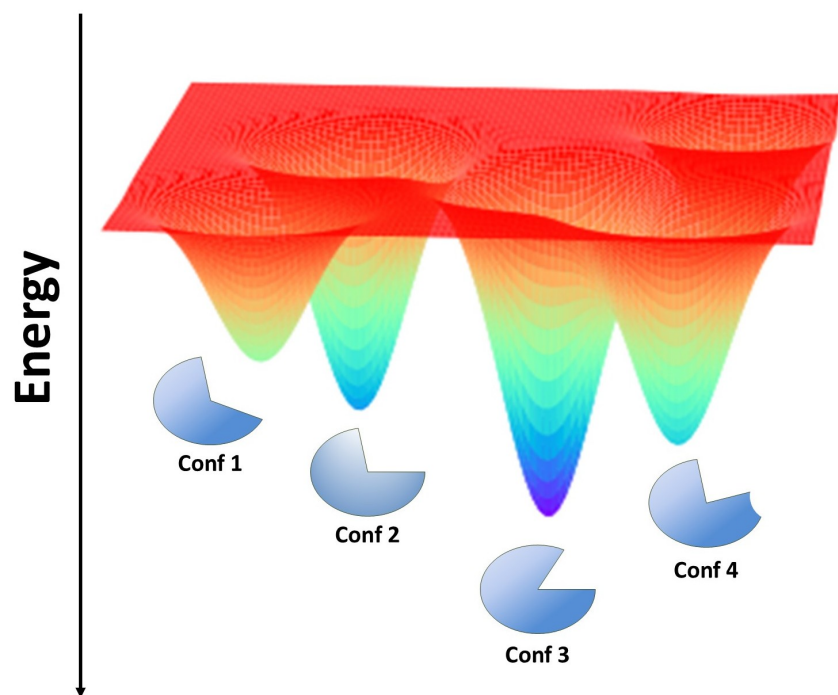


Figure 1.3: Schematic representation of an enzyme conformational ensemble, illustrating the existence of multiple interconverting structural states accessible to the protein.

correspondence suggested that the motions required for catalysis were not induced by the reaction, but instead pre-existed within the enzyme’s dynamic landscape and may even contribute to limiting the overall rate of turnover. Importantly, these dynamics were not confined to the active site, but involved a broader network of residues spanning the protein structure.

In contrast to intrinsic motions, some conformational changes arise only upon substrate binding and occur for specific catalytic purposes. Many enzymes form covalent catalytic intermediates that transiently modify active-site residues and alter the local electrostatic environment, thereby reshaping enzyme structure and conformational dynamics [33, 38, 39, 40]. Because these intermediates temporarily change bonding interactions within the active site, they can activate new regimes of motion that facilitate subsequent steps of catalysis [41]. Such movements may be triggered by disruption of an internal hydrogen bond when an active-site residue forms a new bond with the substrate [42], by steric effects introduced upon substrate binding [6], or by a combination of both [40].

An illustrative example comes from work by Mark Wilson and colleagues at the University of Nebraska, published in *PNAS* in 2019 [42]. In this study, the authors investigated the dynamics of isocyanide hydratase (ICH), a homodimeric cysteine-dependent enzyme in the DJ-1 superfamily that irreversibly hydrates isocyanides to yield N-formamides [43]. Using mix-and-inject X-ray crystallography, they captured catalytic intermediates in-crystallo and demonstrated how substrate-induced structural changes facilitate catalysis.

Previous X-ray structures of ICH had revealed mobility in a helix positioned near the catalytically important residue Cys101. In the resting enzyme, Cys101 accepts an S^- –HN hydrogen bond from the strained backbone amide of Ile152. Disruption of this hydrogen bond in an engineered C101A mutant allowed Ile152 to relax to a less strained conformation, accompanied by displacement of the adjacent helix [44, 45]. In the time-resolved experiments, formation of a thioimidate covalent in-

intermediate during catalysis produced a similar perturbation of Cys101 and triggered displacement of the same helix. Pre-steady-state kinetic analyses of wild-type and mutant ICH further suggested that these cysteine-gated conformational changes are important for enabling water access and hydrolysis of the catalytic intermediate.

1.4 Enzyme’s Flexibility Can Sometimes be Detrimental to Its Function

For optimal catalytic function, enzymes must maintain a degree of structural rigidity within their active sites to preserve the precise geometric arrangement required for chemical transformation and to support interactions necessary for transition-state stabilization [46]. Rigidity and preorganization reduce the entropic cost associated with reorganizing catalytic residues during the reaction and are key contributors to enzymatic efficiency.

Clear experimental evidence for the role of active-site rigidity was provided by Broom et al. [2], who examined X-ray crystallographic structures of HG3, an engineered Kemp eliminase, along with variants generated through directed evolution. HG3 was originally designed by Privett et al. [47] to catalyze Kemp elimination, a simple deprotonation reaction that proceeds through a single step with one intermediate [48]. Directed evolution by Blomberg et al. [49] resulted in a substantial increase in catalytic efficiency, culminating in the HG3.17 variant with an approximately 6×10^8 -fold improvement in k_{cat}/K_M . Broom et al. demonstrated that increased active-site rigidity is a major contributing factor to this enhancement.

The active site in this evolutionary series consists of a catalytic base (Asp127) that abstracts a proton from the substrate and a hydrogen-bond donor at position 50 (Lys, His, or Gln) that stabilizes the developing oxyanion in the transition state (Figure 1.4A). To characterize these interactions, members of the HG3 fam-

ily were crystallized in both the absence and presence of a transition-state analog (6-nitrobenzaldehyde, 6NT) (Figure 1.4A).

Across the evolutionary trajectory, Asp127 remains conserved, while residue 50 is mutated from lysine to histidine and subsequently to glutamine. Histidine and glutamine exhibit reduced conformational freedom compared to lysine and are capable of forming direct hydrogen bonds with the oxyanion, providing more effective stabilization than the originally designed water-mediated interaction (Figure 1.4A). Consistent with this, Broom et al. reported that while the thermal factors of Asp127 show no clear trend, residue 50 exhibits a systematic decrease in thermal factor from HG3 to HG3.17, indicating increased rigidity.

The structural basis for this rigidification is illustrated in Figures 1.4B and 1.4C. In earlier variants (HG3 and HG3.3b), the backbone of residues 83–84, located behind residue 50, does not interact with residue 50 and is able to adopt multiple conformations. In HG3.17, the glutamine side chain forms a hydrogen bond with this backbone, restricting its motion and contributing to stabilization of residue 50. A similar trend is observed in the loop comprising residues 87–90, which exhibits higher mobility in earlier variants and becomes progressively more ordered in HG3.17.

1.4.1 The Rigidity–Flexibility Paradox

For efficient catalysis, enzymes must possess sufficient rigidity to create the precise geometric and electrostatic environment required to stabilize the transition state. Accurate positioning of catalytic residues, proper orbital alignment, and optimized electrostatic interactions all depend on structural organization. At the same time, enzymes must retain enough flexibility to bind substrates, accommodate reaction intermediates, and undergo conformational changes necessary for substrate entry and product release.

These seemingly opposing requirements define a central challenge in understand-

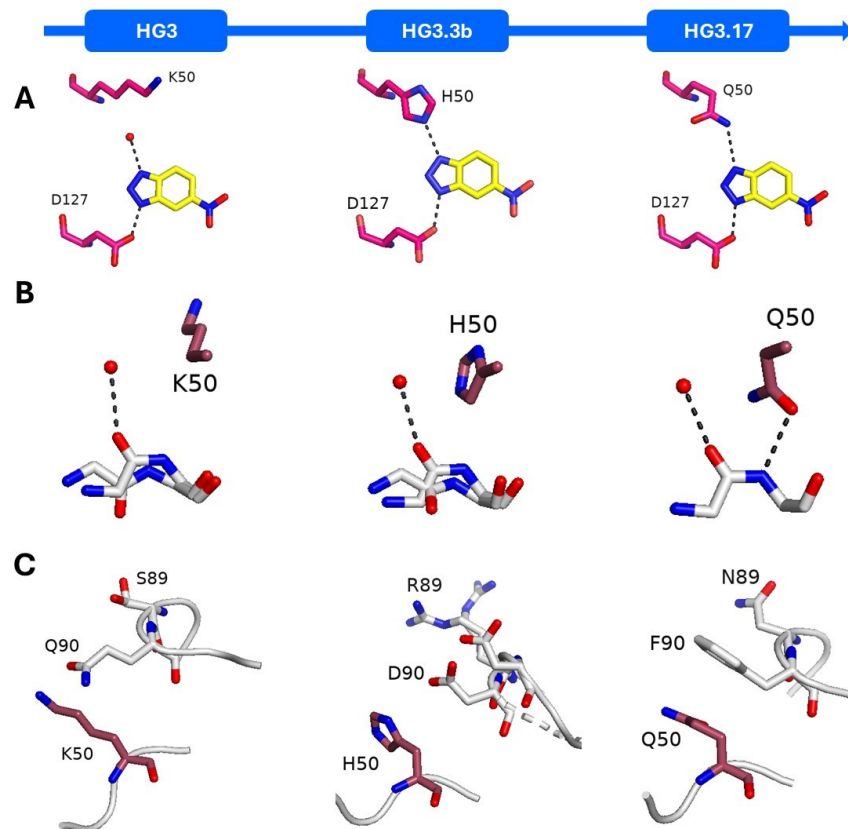


Figure 1.4: Structural changes accompanying the evolution of HG3 toward HG3.17. (A) Active-site interactions between the transition-state analog (6NT, yellow) and key catalytic residues across evolutionary variants. The catalytic base Asp127 is conserved in all variants, while the hydrogen-bond donor at position 50 evolves from Lys50 to His50 and subsequently to Gln50, enabling direct stabilization of the developing oxyanion in the transition state. (B) Local backbone environment underlying residue 50. The backbone of residues 83–84 (shown in white), positioned beneath residue 50, exhibits greater conformational freedom in HG3 and HG3.3b. In HG3.17, hydrogen bonding by Gln50 is associated with reduced backbone flexibility and increased rigidity at this position. (C) Conformational behavior of the loop comprising residues 87–90. This loop is more flexible in earlier variants and becomes progressively more ordered in HG3.17.

ing enzymatic catalysis: how are rigidity and flexibility balanced to achieve optimal function? Excessive rigidity may prevent the enzyme from accessing important and different conformational substates along the reaction coordinate, whereas excessive flexibility may reduce catalytic precision by disrupting critical interactions. Understanding how enzymes navigate this balance is therefore essential for clarifying the relationship between structure, dynamics, and catalytic efficiency.

1.5 Studying Protein Dynamics with X-ray Crystallography

Understanding protein dynamics remains far more challenging than determining single static structures. While recent advances in structure prediction and determination have been driven by large experimental datasets, characterizing conformational heterogeneity and time-dependent behavior still relies heavily on computational methods, including molecular dynamics simulations and related modeling approaches. These methods provide valuable insights into possible motions and energy landscapes, but their predictions are often difficult to validate directly due to the limited availability of high-resolution experimental data that capture dynamic states [31].

Experimentally, several techniques have been used to probe enzyme structure and dynamics, including NMR spectroscopy [31], cryo-electron microscopy [50, 51], small-angle X-ray scattering (SAXS) [52], cryo-electron tomography (CryoET) [53], vibrational spectroscopy, and molecular dynamics simulations [54]. Each of these methods provides valuable information across different timescales and resolutions. However, when atomic-level structural detail is required, X-ray crystallography remains one of the highest-resolution techniques available.

X-ray crystallography has long been a central tool in structural biology for de-

termining the three-dimensional structures of small molecules and biological macromolecules, including proteins and nucleic acids [55, 56]. The method relies on the diffraction of X-rays by a crystalline lattice composed of many ordered copies of the molecule of interest. The resulting diffraction pattern encodes information about the spatial distribution of electron density, which can be reconstructed through Fourier transformation into an electron density map. An atomic model is then built and refined against the measured diffraction data to generate a structural model [56] (Figure 1.5).

Due to its ability to provide high-resolution structural information, X-ray crystallography has become the dominant method for macromolecular structure determination. A substantial fraction of structures deposited in the Protein Data Bank (PDB) [57] have been solved using this technique, with 175,602 of 205,451 total entries in 2023 determined by crystallography [55]. This prevalence reflects both the robustness of the method and its foundational role in structural biology.

Traditionally, crystallography has been associated with static, ground-state structures, often collected under cryogenic conditions. However, the technique has evolved significantly and is now capable of capturing structural divergence and conformational heterogeneity. In essence, crystallography has a strong capacity to provide information about molecular dynamics. Each crystal contains millions of copies of the protein, which can populate multiple accessible conformational substates at equilibrium during data collection (seconds to hours), while the timescale of conformational transitions is typically in the microsecond range. As a result, the observed electron density represents an ensemble-averaged view of these conformations rather than a single static structure.

Another advantage of X-ray crystallography is its ability to probe structural responses to controlled perturbations. Changes in temperature [55, 58], controlled dehydration [59], ligand or substrate binding [42], and exposure to radiation [60] can all reveal alternative conformational states that are otherwise hidden under standard

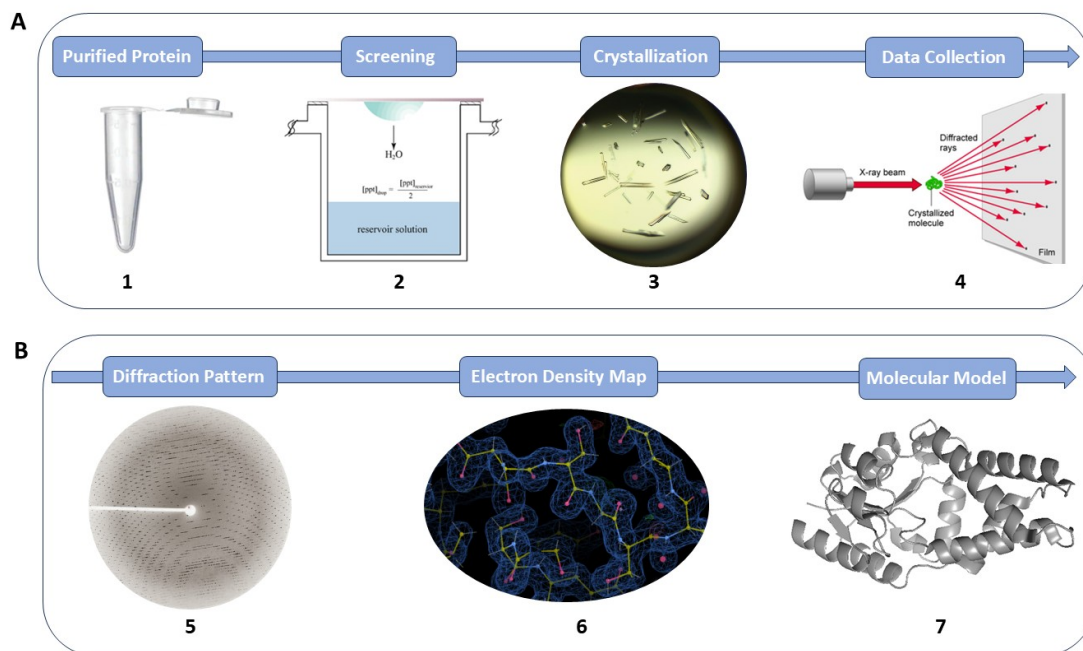


Figure 1.5: X-ray crystallography workflow. (A) Experimental pipeline. Purified protein (1) is screened against a range of crystallization conditions (2) to identify optimal conditions for crystal growth. Once suitable crystals are obtained (3), the best-diffracting crystal is exposed to an X-ray beam, and diffraction data are collected (4). (B) Structure determination and model building. Diffraction images collected over a full rotation of the crystal (typically 180°) are indexed and integrated to generate a three-dimensional diffraction dataset (5). The measured structure factor amplitudes are then used in a Fourier transformation to calculate an electron density map (6). An atomic model is built and iteratively refined against this electron density to produce the final molecular model of the protein (7).

cryogenic conditions. By examining how structures respond to these perturbations, crystallography can provide access to conformational heterogeneity and, in some cases, non-equilibrium structural states. Rather than offering only a single static snapshot, modern crystallographic approaches can sample multiple regions of the conformational landscape of an enzyme. In this way, crystallography serves not only as a structure-determination tool, but also as a powerful method for investigating the dynamic behavior underlying enzymatic function.

1.5.1 Crystallographic Tools for Dynamic Analysis

In single-crystal X-ray crystallography, information about dynamics is encoded in the electron density map and in the refinement parameters associated with each atomic model. Regions of increased atomic displacement often appear as broadened or weaker electron density.

The most commonly used metric for describing atomic mobility is the **temperature factor**, also known as the **B-factor**. The B-factor represents the extent to which electron density is distributed around an atomic position and is typically interpreted as a measure of atomic displacement [61, 62]. If B_j were purely a measure of thermal motion at atom j (assuming that the occupancies are correctly modeled), then in the simplest case of purely harmonic motion with equal magnitude in all directions—referred to as isotropic vibration— B_j is related to the magnitude of vibration as follows:

$$B_j = 8\pi^2 \langle u_j^2 \rangle = 79 \langle u_j^2 \rangle$$

[56, 63]

where $\langle u_j^2 \rangle$ represents the mean-square displacement of the atom from its equilibrium position. In practice, however, the B-factor compresses complex positional variability into a single scalar value for each atom. While useful, this simplification

assumes harmonic and isotropic motion and therefore risks oversimplifying the true nature of conformational dynamics.

In reality, atomic displacements in proteins are often anisotropic, meaning that atoms do not move equally in all directions. Instead, their motion can be directional, reflecting local packing constraints, collective domain movements, or preferred pathways of conformational change. This behavior can be described using anisotropic displacement parameters (ADPs), in which atomic motion is represented as a three-dimensional ellipsoid rather than a sphere. These ellipsoids capture both the magnitude and directionality of motion, providing a more detailed description of atomic fluctuations [64]. However, refining full anisotropic parameters for every atom requires a large number of parameters and is typically only feasible at very high resolution.

To capture anisotropic and collective motions anisotropically while maintaining a manageable number of parameters, more advanced parameterizations have been developed. One widely used approach is the **Translation–Libration–Screw (TLS)** model [65]. TLS describes rigid-body motions of groups of atoms using parameters that account for translations, rotations (librations), and coupled screw motions. Rather than refining independent anisotropic displacement parameters for every atom, TLS partitions the structure into domains and assigns a set of collective motion parameters to each group. This significantly reduces the number of refined parameters while still capturing directional and correlated motions, providing a more physically meaningful description of large-scale domain dynamics.

Although still simplified, these approaches move beyond purely isotropic descriptions of atomic displacement and provide a framework for interpreting crystallographic data in terms of structural heterogeneity and collective motion. Although B-factors and TLS parameters are useful for highlighting flexible regions of a protein, crystallographic data can reveal more detailed structural heterogeneity. In certain regions of the electron density map, clear alternative conformations can be directly

observed, affecting both backbone segments and side chains. These are not merely inferred displacements, but distinct, resolvable structural states.

Such alternative conformations are typically modeled using **multi-conformer refinement** [66]. In this approach, the protein is represented as a single structural model, while specific residues or segments are assigned multiple conformations where supported by the electron density. Each conformation is associated with an **occupancy** value, reported either as a fraction of 1 or as a percentage, representing the population of that state within the crystal. In practice, alternative conformations with occupancies above 10% can often be modeled with reasonable confidence, depending on resolution and data quality.

Multi-conformer refinement is particularly useful for capturing localized motions, especially in catalytically important residues. Examples of this type of modeling can be seen in Section 1.4 (Figure 1.4), where changes in conformational heterogeneity along the HG3 evolutionary trajectory are visualized.

Multi-conformer modeling can complicate the interpretation of B-factors. Because B-factors reflect the spread of electron density around a modeled atomic position, introducing a second conformation often redistributes that density between two discrete states. As a result, the individual B-factors of each conformer may decrease substantially compared to a single-conformer model. In such cases, simply averaging or summing the B-factors of alternative conformations does not provide an accurate estimate of the true extent of motion for that residue.

To address this limitation, a parameter referred to as an **order parameter** was introduced to better capture the overall conformational variability of a residue when alternative conformations are present [67]. This crystallographic order parameter differs from the order parameters commonly used in NMR studies. Rather than describing bond vector fluctuations derived from relaxation measurements, it is designed to account for positional heterogeneity observed in electron density while incorporating the contributions of multiple modeled conformations.

Finally, a more comprehensive and somewhat debated [68] approach for capturing conformational heterogeneity in crystallographic models is **ensemble refinement** [69]. Ensemble refinement integrates molecular dynamics (MD) simulations with crystallographic data, using the experimental diffraction data as restraints during the simulation.

Conventional MD simulations are computationally demanding and often limited to relatively short timescales and tractable system sizes. While they can provide detailed insight into local motions, many conformational changes in enzymes occur on microsecond to millisecond timescales, which remain difficult to access with standard MD approaches. As a result, sampling of relevant conformational states can be incomplete.

Ensemble refinement addresses this limitation by incorporating information encoded in the crystallographic data directly into the simulation. Because diffraction data represent an average over millions of molecules within the crystal lattice—each potentially sampling multiple conformations during data collection—the experimental observations contain implicit information about structural heterogeneity. By restraining the MD simulation with these data, ensemble refinement biases sampling toward conformations consistent with the measured diffraction intensities, improving efficiency relative to unconstrained simulations.

The output of ensemble refinement is not a single static model, but an ensemble of structures representing alternative conformational states compatible with the crystallographic data. In recent years, there has been discussion regarding the limits of this approach, including concerns about model precision and the potential for overfitting. Nevertheless, ensemble refinement has emerged as a powerful tool for visualizing conformational heterogeneity beyond what is accessible through single-structure refinement alone.

Enzymes can adopt non-equilibrium structural and dynamic states in response to external stimuli such as temperature changes and substrate binding. These

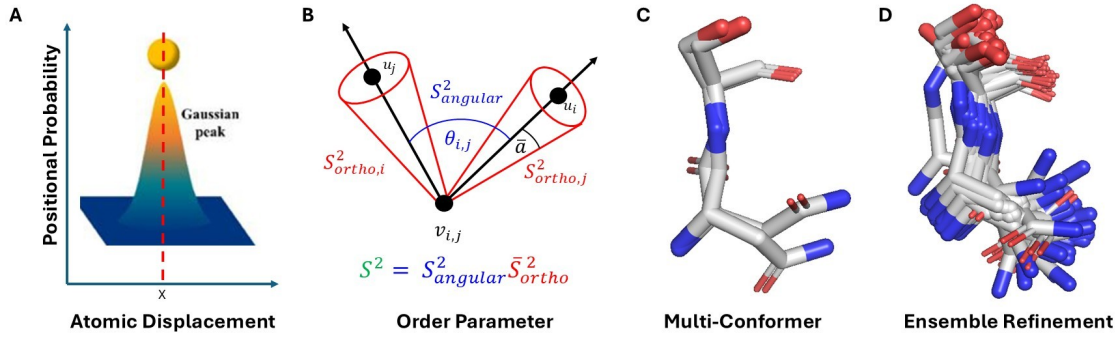


Figure 1.6: Crystallographic representations of conformational heterogeneity and dynamics. (A) Atomic positional distribution modeled as a Gaussian, illustrating atomic displacement around an average position. (B) Schematic representation of order parameters, where overall motion (S^2) is decomposed into angular ($S^2_{angular}$) and orthogonal (S^2_{ortho}) components to account for alternative conformations. (C) Multi-conformer modeling of side-chain conformational heterogeneity within a single structure. (D) Ensemble refinement representation capturing a distribution of conformations consistent with the experimental electron density.

states are often not captured in traditional crystallographic studies of apo structures, where such conformations may be sparsely populated or entirely absent. Earlier approaches attempted to mimic these non-equilibrium states using covalent inhibitors or transition-state analogs. However, advances in X-ray sources, detectors, and sample delivery methods have enabled a new class of crystallographic experiments capable of directly probing non-equilibrium protein motions in response to external perturbations [70].

Serial crystallography provides an alternative to traditional single-crystal X-ray crystallography by enabling structural analysis from large ensembles of microcrystals rather than a single stationary crystal. A key advantage of this approach is its compatibility with time-resolved experiments, in which controlled perturbations are used to initiate structural changes. Perturbations such as rapid temperature jumps [60] or substrate mixing [71, 72, 42] can drive proteins out of equilibrium, allowing their structural responses to be probed as a function of time. These methods enable direct investigation of how enzymes dynamically respond to external stimuli and progress through functionally relevant conformational states, and also enhance sensitivity to structural heterogeneity and low-population conformations, providing a more comprehensive view of the structural landscape sampled by proteins.

Mix-and-inject serial crystallography (MISC) is a time-resolved technique in which substrates or ligands are rapidly mixed with microcrystals immediately prior to X-ray exposure. By varying the delay between mixing and data collection, MISC enables structural snapshots to be captured at different stages along the reaction coordinate. This provides access to non-equilibrium states such as enzyme–substrate, enzyme–intermediate, and enzyme–product complexes, offering a powerful strategy for directly observing structural changes associated with catalysis.

MISC experiments, in particular, are well suited for studying enzyme structure–dynamics–function relationships because they provide access to otherwise hidden aspects of enzymatic behavior. By enabling the capture of substrate-induced

structural changes and transient conformational states, these experiments help reveal how specific structural and dynamic features contribute to catalytic function. Such information is often inaccessible when enzymes are studied solely in their ground states, where functionally relevant motions and intermediate conformations may remain obscured.

1.5.2 Advantages of Room-Temperature X-ray Crystallography

Early X-ray crystallographic studies faced significant limitations due to radiation damage, as protein crystals are highly sensitive to intense X-ray exposure [73]. Rapid damage often prevented complete data collection before crystal deterioration [74]. To overcome this challenge, crystals began to be cooled to cryogenic temperatures (typically around 100 K), increasing their resistance to radiation damage. This approach, known as **cryogenic X-ray crystallography**, enabled the routine collection of high-resolution datasets and played a central role in the expansion of structural biology [75, 76].

Despite these advantages, cryogenic data collection introduces important limitations, particularly in capturing protein dynamics and conformational heterogeneity [77]. Cooling to 100 K significantly reduces thermal motion and can restrict access to conformational substates that are energetically accessible at higher temperatures. As a result, higher-energy conformations that may be functionally relevant can become underpopulated or effectively inaccessible, leading to structural models that represent a narrowed portion of the protein’s conformational landscape.

This effect has been experimentally demonstrated in the literature. In one example in the work of Keedy et al. (2015) [58], crystal structures of the well-studied enzyme Cyclophilin A were determined at multiple temperatures to assess how temperature influences conformational sampling (Figure 1.7). These studies showed that

conformations not readily observed under cryogenic conditions become increasingly populated as temperature increases.

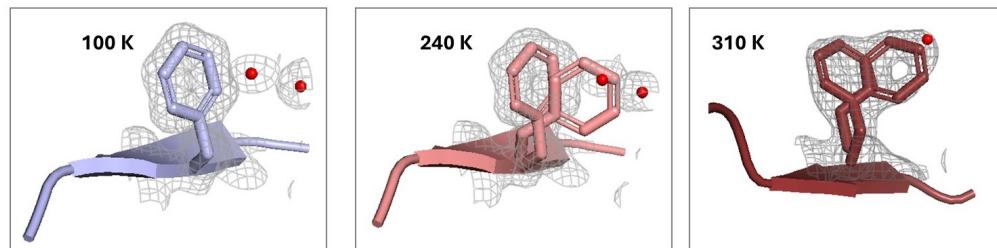


Figure 1.7: Temperature-dependent conformational heterogeneity in Cyclophilin A (CypA). Electron density maps and modeled conformations of Phe113 are shown at 100 K, 240 K, and 310 K. At cryogenic temperature (100 K), Phe113 adopts a single dominant conformation. At intermediate temperature (240 K), partial sampling of an alternative conformation is observed, while at 310 K this alternative conformation becomes more clearly populated. These observations illustrate how increasing temperature enables access to conformational states that are less apparent under cryogenic conditions.

1.6 Engineered Enzymes as Structure–Function Relationship Models

Over the past several decades, enzyme engineering has developed into an effective strategy for introducing new catalytic activities into otherwise neutral protein scaffolds, enabling the creation of enzymes capable of performing chemical transformations that lack natural biological counterparts [78, 79, 80]. While these engineered enzymes have been successful in generating new reactivities, they often exhibit significantly lower catalytic efficiencies compared to natural enzymes.

Directed evolution has emerged as a powerful approach to improve the efficiency of engineered enzymes [81, 82]. This technique mimics natural evolution in a laboratory setting and complements rational design by enabling systematic exploration of sequence space through iterative cycles of mutation and selection (Figure 1.8). Through these cycles, directed evolution can identify mutations that enhance catalytic function, both within the active site and at distal regions of the protein scaffold.

In addition to enabling the creation of novel catalysts, enzyme design strategies, including directed evolution, provide systems that are particularly well-suited for probing enzyme structure–function relationships. Because catalytic efficiency is the primary selection pressure during directed evolution, the resulting enzymes offer controlled systems for studying sequence–structure–dynamics–function relationships.

In contrast, natural enzymes have evolved over long timescales under multiple and often competing selective pressures. As a result, it can be difficult to determine whether a given structural or dynamic feature contributes directly to catalysis or instead reflects adaptations for stability, regulation, or interactions with other cellular components [83, 84]. This ambiguity is especially important when attempting to identify dynamic features that actively contribute to catalytic efficiency, as opposed to motions that are functionally incidental.

Engineered systems help mitigate this challenge. In particular, designed enzymes provide an opportunity to examine how mutations, including those distant from the active site, influence catalytic efficiency and functional behavior [28, 29].

Computationally designed enzymes typically exhibit low catalytic efficiency, but can be substantially improved through directed evolution. This process provides a clear and traceable pathway of how mutations accumulate over successive rounds and how these changes impact function [85]. As a result, these systems offer a unique framework for linking structural and dynamic features to catalytic performance.

By following this evolutionary trajectory, it becomes possible to identify which types of structural rearrangements, such as active site reshaping, changes in flex-

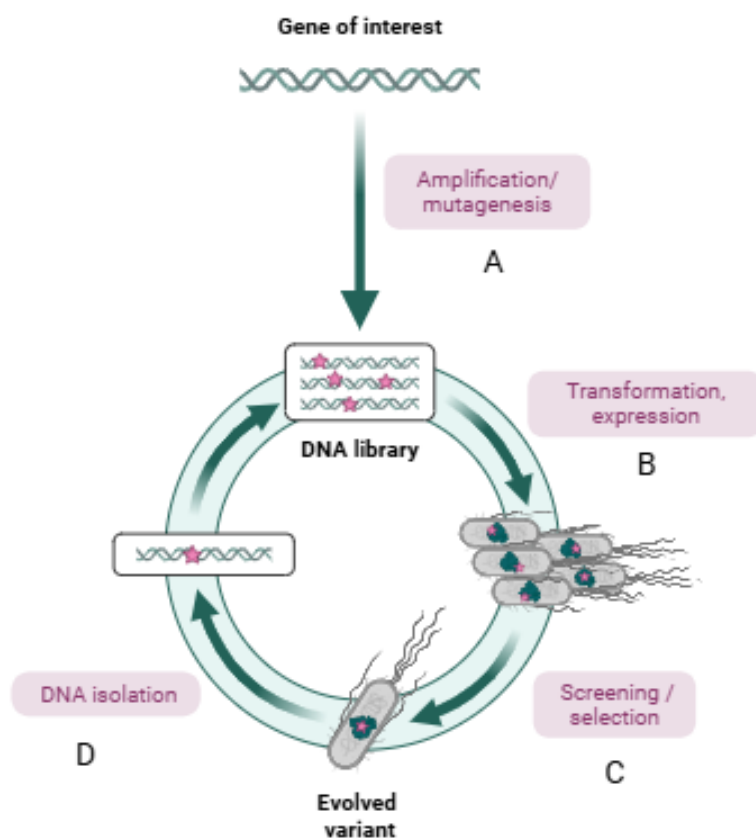


Figure 1.8: Overview of directed evolution for enzyme optimization. (A) A target gene is amplified and diversified to generate a DNA library using mutagenesis methods. (B) The library is expressed in a suitable host system. (C) Variants are screened or selected based on catalytic performance. (D) The best-performing variants are isolated for further rounds of evolution.

ibility, or distal effects, contribute to improved catalysis. In this way, engineered enzymes not only serve as model systems for studying enzyme function but also provide practical guidelines for enhancing catalytic efficiency.

1.7 Thesis Objective

The central objective of this thesis is to investigate the sequence–structure–dynamics–function relationship in enzymes using simplified, engineered systems as case studies. A primary focus is understanding how enzymes maintain a delicate balance between rigidity and flexibility: sufficient rigidity to ensure precise substrate binding and optimal geometric arrangement for chemical transformation, alongside the flexibility required to enable conformational changes associated with substrate entry, intermediate stabilization, and product release. This balance becomes especially critical for enzymes that catalyze multi-step reactions, where distinct conformational rearrangements may be required to stabilize multiple intermediates along the catalytic pathway.

Within this framework, this thesis is organized around three complementary projects. The first project focuses on the HG3 scaffold, a designed Kemp eliminase, and examines how increased active-site rigidity can be advantageous for enzymes catalyzing relatively simple reactions such as Kemp elimination. In this case, reduced structural heterogeneity is incorporated into computational design algorithms to improve catalytic performance. The second project explores how increased conformational flexibility becomes necessary as reaction complexity increases, requiring enzymes to adopt distinct dynamic features to accommodate more demanding catalytic mechanisms. This project also investigates the emergence of unique structural features in homologous enzymes in response to different catalytic demands. Finally, the third project focuses on capturing non-equilibrium enzymatic states, with the goal of directly linking structural and dynamic features to functional motions along

the reaction coordinate.

Chapter 2

Design of Efficient Artificial Enzymes Using Crystallographically Enhanced Conformational Sampling

Copyright: Rojo V. Rakotoharisoa, Behnoush Seifinoferest, Niayesh Zarifi, Jack Miller, Joshua M. Rodriguez, Michael C. Thompson, Roberto A. Chica

2.1 Statement of Contribution

R.V.R. performed the computational design experiments of HG3, and N.Z. performed computational design experiments of KE70. R.V.R. purified HG3 proteins and J.M. purified KE70 proteins. R.V.R. performed enzyme kinetics experiments on HG3 and J.M. performed enzyme kinetics experiments on KE70. B.S. crystallized proteins. J.M.R. and M.C.T. performed X-ray diffraction experiments. B.S., R.V.R., and R.A.C. performed refinements. M.C.T. designed X-ray crystallography experiments. R.A.C. and R.V.R. wrote the manuscript. N.Z., B.S., and M.C.T. edited the manuscript.

2.2 Abstract

The ability to create efficient artificial enzymes for any chemical reaction is of great interest. Here, we describe a computational design method for increasing the catalytic efficiency of de novo enzymes by several orders of magnitude without relying on directed evolution and high-throughput screening. Using structural ensembles generated from dynamics-based refinement against X-ray diffraction data collected from crystals of Kemp eliminases HG3 ($k_{\text{cat}}/K_{\text{M}}$ 125 M⁻¹ s⁻¹) and KE70 ($k_{\text{cat}}/K_{\text{M}}$ 57 M⁻¹ s⁻¹), we design from each enzyme ≤ 10 sequences predicted to catalyze this reaction more efficiently. The most active designs display $k_{\text{cat}}/K_{\text{M}}$ values improved by 100-250-fold, comparable to mutants obtained after screening thousands of variants in multiple rounds of directed evolution. Crystal structures show excellent agreement with computational models, with catalytic contacts present as designed and transition-state root-mean-square deviations of ≥ 65 Å. Our work shows how ensemble-based design can generate efficient artificial enzymes by exploiting the true conformational ensemble to design improved active sites.

2.3 Introduction

The ability to create efficient artificial enzymes for any desired chemical reaction is of great interest as it would open the door to a broad range of applications in industry and medicine. Algorithms for de novo enzyme design [86, 87], can be used to create artificial enzymes for reactions for which no naturally occurring enzyme is known. While this computational design process has been used to produce de novo enzymes for various model organic transformations [48, 88, 89, 90, 47], downstream directed evolution has been necessary to convert these enzyme prototypes into catalysts as efficient as their natural counterparts [49, 91, 92, 93, 94, 95]. However, directed evolution is time-consuming and impractical for many enzymatic reactions

of interest, for which high-throughput screening assays are unavailable. Therefore, a computational design procedure to enhance the catalytic efficiency of de novo enzymes would be beneficial as it could bypass directed evolution and accelerate the creation of efficient artificial enzymes.

Structural analyses of high-activity variants evolved from de novo enzymes revealed several mechanisms by which catalytic functions can be augmented. These mechanisms include optimization of hydrogen-bond geometries between catalytic residues and the transition state (TS)[49], enhancement of active site complementarity to the TS [49, 91, 2], improvement of active site preorganization [2], and enrichment of catalytically competent substates in the conformational ensemble[1, 96, 97]. These structural effects are challenging to design using current enzyme design protocols [48, 47, 91] because they are not sufficiently precise [49] and do not consider the dynamic nature of enzyme structures [1]. Recently, new methodologies have been developed that take advantage of the spatiotemporal averaging inherent to X-ray crystallographic data, which represent billions of crystallized molecules that are all sampling different conformations during data collection to model the conformational ensembles of proteins with high fidelity[69, 98]. New design strategies that could exploit this experimental information should facilitate the design of efficient artificial enzymes by enabling the accurate modeling of catalytically productive conformational substates.

Here, we report a computational procedure for enhancing the catalytic efficiency of de novo enzymes that exploits crystallographically derived ensemble models for increased accuracy. Starting from the noncryogenic crystal structure of de novo Kemp eliminase HG3 [47, 2], we generate a structural ensemble fitted to the diffraction data [69, 99] and use it to design mutant sequences predicted to efficiently catalyze the Kemp elimination. We filter these sequences using a variety of structural criteria aiming to reproduce molecular mechanisms known to augment enzyme catalysis and experimentally characterize 10 sequences. The most active design displays a catalytic

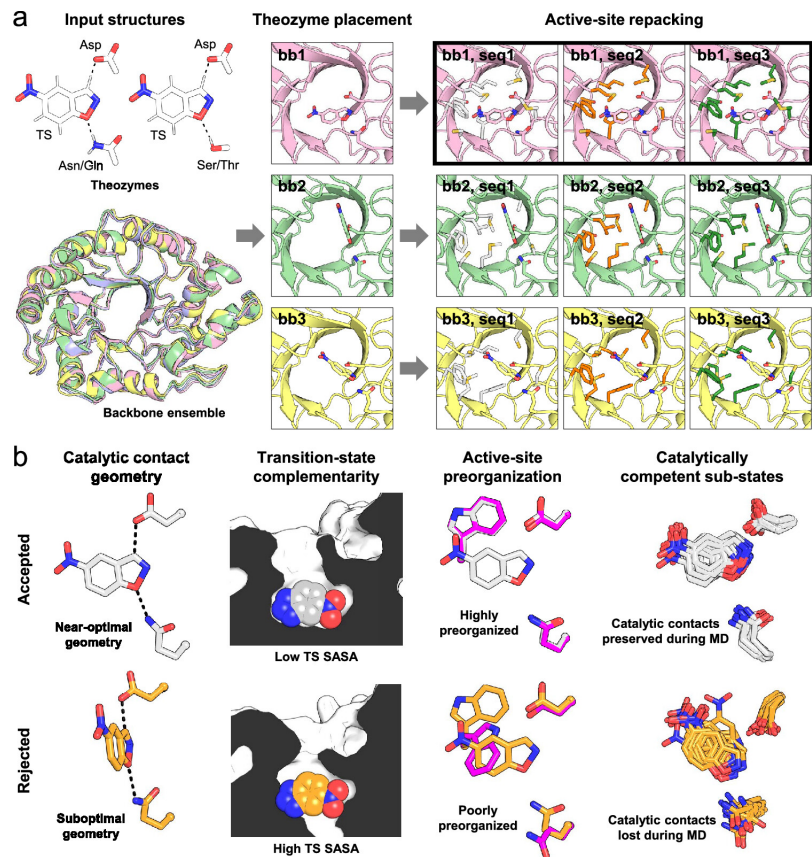


Figure 2.1: Ensemble-based computational enzyme design. (a) The procedure uses one or more theozymes—combinations of catalytic groups arranged to stabilize a TS—and a backbone ensemble representing intrinsic flexibility of the protein fold. The algorithm proceeds in two stages: placement of each theozyme onto each backbone member (bb1–3), followed by active-site repacking to identify sequences (seq1–3) that further stabilize the TS. During repacking, both residue identity and conformation are varied, and the TS is rotated and translated to minimize steric clashes. For clarity, only the theozyme containing a Gln hydrogen-bond donor is shown. Designs from individual theozyme/template combinations are selected based on energy (bold outline) and filtered using structural criteria. (b) Filtering selects designs that reproduce features associated with efficient Kemp eliminases: (1) near-ideal TS–side chain contact geometries (dashed lines), (2) high structural complementarity (low TS solvent-accessible surface area, SASA), (3) preorganized active sites with consistent rotamers in the presence (white/orange) and absence (magenta) of TS, and (4) enrichment of catalytically competent substates that maintain catalytic contacts during short MD simulations.

efficiency improved by approximately 250-fold, a result comparable to those obtained by seven rounds of directed evolution and screening of >7000 variants [49]. We solve crystal structures of several high-activity variants that show excellent agreement with the computational models and confirm that catalytic efficiency is enhanced via the intended molecular mechanisms. We also apply our approach to another de novo Kemp eliminase, KE70 [48], which yields comparable results. Our work shows how ensemble-based design can generate efficient artificial enzymes by exploiting information derived from X-ray crystallography to identify productive substates for efficient catalysis.

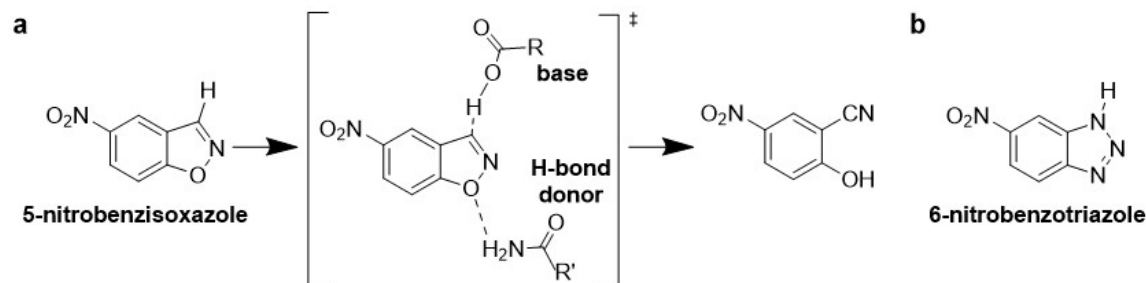


Figure 2.2: The Kemp elimination reaction. The Kemp elimination reaction. (a) HG-series Kemp eliminases catalyze the Kemp elimination using a catalytic motif consisting of a base (Asp) that deprotonates the 5-nitrobenzisoxazole substrate, and an H-bond donor (Gln) that stabilizes negative charge buildup on the phenolic oxygen at the transition state ($\ddagger$). This reaction yields the 4-nitro-2-cyanophenol product. (b) Structure of the 6NT transition-state analogue.

2.4 Results

2.4.1 Ensemble-Based Enzyme Design

Ensemble-based design can be a replacement for directed evolution, to create efficient de novo enzymes without relying on iterative rounds of mutagenesis and high-throughput screening. Here, we implemented an ensemble-based computational enzyme design pipeline that uses protein backbones generated by ensemble refinement of diffraction data [69] as templates for design (Figure 2.1).

Table 2.1: Ensemble refinement results

Structure	PDB ID	Resolution (Å)	# chains	$R_{\text{work}}/R_{\text{free}}$ before refinement	$R_{\text{work}}/R_{\text{free}}$ after refinement	Ensemble size
HG3 (+) 6NT	5RGA	1.86	2	0.1482/0.1863	0.1332/0.1814	84
HG185 (+) 6NT	8USL	1.55	2	0.1610/0.1831	0.1387/0.1740	100
HG198 (+) 6NT	8USJ	1.60	3	0.1540/0.1900	0.1402/0.1890	50
HG630 (+) 6NT	8USH	1.32	2	0.1190/0.1476	0.1316/0.1589	127
HG649 (+) 6NT	8USF	1.45	2	0.1373/0.1567	0.1302/0.1584	102
KE70 (-) 6NT	3NPV	1.48	2	0.1430/0.1700	0.1290/0.1549	84

Because the diffraction data and calculated electron density maps are derived from crystals that contain billions of molecules, which are all sampling different conformational states throughout data collection, the resulting ensemble models represent a distribution of conformations that are well populated in the ensemble at equilibrium[55]. Thus, we postulated that performing crystallographic ensemble refinement on a low activity de novo enzyme would allow sampling of conformational substates that can more accurately accommodate the TS and its interactions with catalytic residues than the single conformation modeled using traditional refinement methods, thereby facilitating the design of improved active sites. To test our hypothesis, we applied an ensemble-based design to increase the catalytic efficiency of the de novo enzyme HG3 [47] that catalyzes the Kemp elimination (Figure 2.2A), a well-established model organic transformation for benchmarking computational enzyme

design methods. We selected HG3 as the case study for three reasons.

First, X-ray diffraction data collected at noncryogenic temperatures (277 K) were available [2], which is ideal for ensemble refinement because cryocooling is known to narrow the conformational ensemble of crystallized proteins [100], reducing the number of conformations that are represented in the ensemble-averaged data. Second, these data were collected from crystals with a bound TS analogue that is resolved in the electron density, allowing inclusion of the TS in the ensemble models. Finally, HG3 was subjected to an extensive directed evolution campaign that increased catalytic efficiency by almost 3 orders of magnitude [49], allowing us to compare results with those from purely experimental approaches.

We generated an ensemble of 84 backbone templates (Table 2.1) from the crystal structure of HG3 bound with the 6-nitrobenzotriazole (6NT, Figure 2.2B) TS analogue (PDB ID: 5RGA[2]) and used these structures for placement of four theozymes containing an Asp residue as catalytic base, and either Gln, Asn, Ser, or Thr (Figure 2.1A) to stabilize the binding of the substrate and properly orient it for reaction with the base. We kept the base position fixed to that of HG3 (D127) but allowed the algorithm to place the hydrogen-bond donor at multiple positions in the active site given that no amino acid residue in HG3 forms this catalytic contact with 6NT [2]. We then repacked the active site around the most stable theozyme placed on each backbone template by optimizing the identity and side-chain conformation of residues within 4 Å of the TS and catalytic residues (Table 2.2) to maximize beneficial nonbonding interactions with the theozyme, which has been shown to contribute to increased catalytic activity [49]. After repacking, we filtered designs using a series of structural metrics aiming to reproduce mechanisms known to augment enzyme catalysis (Figure 2.1B, Table 2.3).

First, we measured angles defining the hydrogen bond between D127 and the TS for the top-scoring design obtained on each backbone template, and selected sequences from the single backbone template that gave the most stable design with an-

Table 2.2: Amino-acid positions optimized during computational design of HG3 variants.

Design objective	Designed residues^a	Allowed amino acids
Catalytic base	D127	WT
H-bond donor	A21 W44 K50 M84 W87 Q90 A125 Y170 M172 L236 M237 T265 F267	N, Q, S, T
Transition-state packing	A21 W44 M84 W87 Q90 A125 Y170 M172 L236 M237 T265 F267	A, C, F, I, L, M, V, T, WT ^b
Active-site bottleneck^c	W275	A, S, V, T, WT ^b
Floated positions^d	V16 Y17 M42 E46 N47 L79 G130 Q207 H209 S234 D239 R276	WT ^b

^a Numbering based on crystal structure of HG3 (PDB ID: 5RGA).

^b WT indicates the amino acid found at that position in the HG3 sequence (e.g., N, Q, and Y at positions 47, 90, and 170, respectively).

^c The bulky W275 residue was mutated to small amino acids to widen the active-site entrance bottleneck.

^d Amino acids at these positions were allowed to change conformation but not identity.

Table 2.3: Design filtering - HG series.

Design	Mutations from HG3	Energy (kcal/mol)	SASA (Å ²)	# Cys	# Met	Residues not preor- ganized
HG44	A21C K50Q Q90F A125V	-206.8	86.2	1	4	Cys21, Trp44, Gln50, Val125, Leu236, Asn239
HG57	A21C K50Q Q90L A125V	-205.9	86.2	1	4	Cys21, Trp44, Gln50, Val125, Leu236, Met237, Asn239, Trp275
HG185	A21C K50Q Q90F A125T W275V	-201.9	87.5	1	4	Cys21, Gln50, Met237
HG198	K50Q M84C A125T F267M	-201.6	99.8	1	4	Trp44, Gln50, Thr125, Gln207, Met237, Met267
HG259	K50Q M84C Q90L A125T F267M	-200.5	99.8	1	4	Trp44, Gln50, Thr125, Gln207, Met237, Met267
HG378	A21C K50Q Q90L A125V W275V	-198.7	87.5	1	4	Cys21, Trp44, Gln50, Val125, Leu236, Met237
HG493	K50Q Q90F A125T W275T	-197.1	87.1	0	4	Trp44, Gln50, Thr125, Leu236, Met237
HG603	K50Q Q90F A125V W275V	-195.5	87.1	0	4	Trp44, Gln50, Thr125, Leu236, Met237
HG630	K50Q M84C A125T F267M W275V	-194.9	99.8	1	4	Trp44, Asn47, Gln50, Thr125, Gln207, Met237, Met267
HG649	K50Q M84C Q90F A125T F267M W275V	-194.6	99.8	1	4	Trp44, Gln50, Thr125, Gln207, Met237, Met267

gles that were closest to the optimal geometries calculated for acetamide dimers[101] (Figure 2.3, Table 2.4), which mimic the geometry needed for efficient proton transfer.

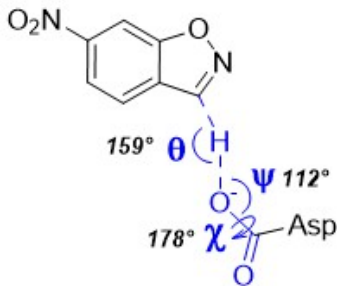


Figure 2.3: Ideal geometry for catalytic base contact. Atoms defining various angles describing the hydrogen-bond interaction between the transition state and designed Asp catalytic base (θ , Ψ , χ) are shown in blue. Values in italics are optimal angles calculated for hydrogen bonding interactions between acetamide dimers.

The selected backbone, which we name bb05, yielded sequences that contain the Q50 hydrogen-bond donor that emerged during evolution of HG3 and that is present in the highest activity variants of this enzyme family [49]. Second, we evaluated active-site complementarity to the TS by calculating the solvent-accessible surface area of this ligand in the computational models generated for 1000 sequences designed on backbone bb05. Only designs where $\geq 95\%$ of the TS surface is buried were kept. Third, we evaluated active-site preorganization by calculating the percentage of designed active-site residues that were predicted to adopt similar rotamers in the absence and presence of the TS. Sequences whose active sites were predicted to be $> 70\%$ preorganized were kept. We also eliminated sequences with calculated energies over 10 kcal mol^{-1} higher than that of an HG3 single point mutant containing the D127/Q50 catalytic dyad as well as those that contained more than four methionines or more than one cysteine in the active site to avoid excessive conformational heterogeneity and prevent formation of disulfide bonds, respectively.

Table 2.4: Angles defining the hydrogen bond between D127 and the transition state or transition-state analogue for various Kemp eliminase structures

Template	PDB ID	Angle (°)			Deviation from ideal (°)		
		ψ^{a}	θ^{a}	χ^{a}	ψ	θ	χ
Crystal structures ^b							
HG3	5RGA	134 ± 1	173 ± 2	182 ± 3	22	14	4.5
HG3.7	5RGC	125 ± 1	167 ± 1	174 ± 2	13	8	3.5
HG185	8USL	125 ± 3	163 ± 3	175 ± 9	13	4	2.5
HG198	8USJ	124 ± 2	168 ± 1	186 ± 5	12	9	8.5
HG630	8USH	125 ± 1	170 ± 1	173 ± 5	13	11	4.5
HG649	8USF	124 ± 2	169 ± 2	181 ± 9	12	10	3.5
Ensemble members ^c							
HG3 bb05	–	113.6	164.0	175.2	1.3	4.6	2.3
KE70 bb06	–	116.2	159.3	173.0	3.9	0.1	4.5

^a Average angles across all chains are reported (mean ± s.d.). Ideal values for ψ , θ and χ are 112.3, 159.4 and 177.5 degrees, respectively. Atoms defining ψ , θ and χ are shown on Figure 2.3.

^b Crystal structures contain the 6NT transition-state analogue.

^c Ensemble members contain the transition state.

Importantly, we did not eliminate entirely these amino acids from our designed active sites nor explicitly specify which positions should contain the methionines or cysteine that were preserved. This filtering process narrowed the original list of 1000 designs obtained from backbone bb05 down to 55 designs. A final filtering step consisting of a 20 ns MD simulation was used to identify designs that were enriched in catalytically competent conformational substates, which we define as snapshots that make the designed catalytic contacts (Figure 2.4). This final filtering resulted in nine sequences containing four or five mutations (Table 2.5), which we experimentally characterized.

2.4.2 Kinetic Characterization of Designs

All nine designs catalyzed the Kemp elimination (Figure 2.5), with all but two having catalytic efficiencies greater than that of HG3 (Table 2.5). The most active of these variants, HG630, had a catalytic efficiency of $15,000 \pm 500 \text{ M}^{-1} \text{ s}^{-1}$, a value that is 120-fold higher than that of the starting template HG3. Furthermore, the three most active designs (HG198, HG259, and HG630) were more catalytically efficient than the HG3.3b variant obtained after three rounds of directed evolution and screening of > 3000 mutants [49].

These designs all contain the M84C mutation that was previously shown to increase catalytic efficiency by 1 order of magnitude when combined with K50Q to yield the HG3-K50Q/M84C double mutant [1] (Table 2.5). Comparison of the sequences of HG44 and HG57, which differ by a single point mutation at position 90, suggested that the introduction of the Q90F mutation into HG630 would further enhance its catalytic efficiency as it doubles $k_{\text{cat}}/K_{\text{M}}$ when introduced into HG57 to yield HG44. Introduction of this mutation into HG630 doubled its catalytic efficiency and resulted in a hexamutant (HG649) that is 256-fold more catalytically efficient than HG3 ($k_{\text{cat}}/K_{\text{M}} = 32,000 \pm 3000 \text{ M}^{-1} \text{ s}^{-1}$) and more active than HG3.7,

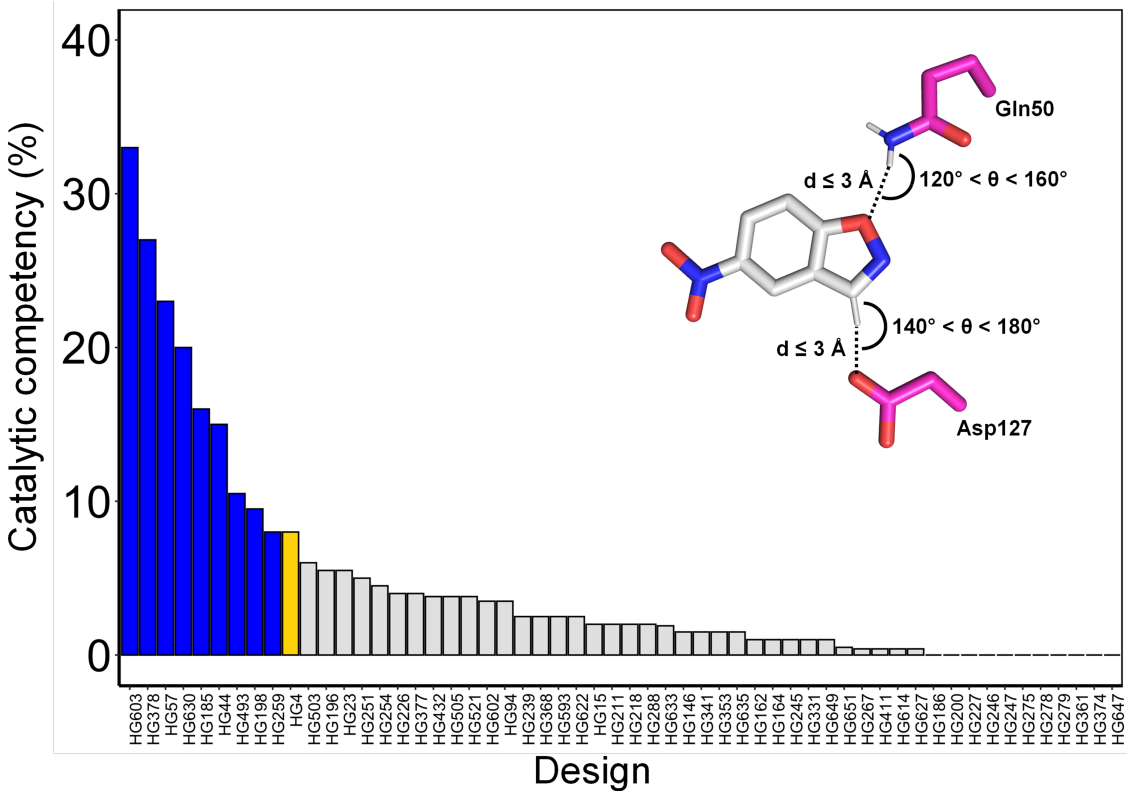


Figure 2.4: Design filtering using molecular dynamics. Following molecular dynamics, angles (θ) and distances (d) defining catalytic contacts between the 5-nitrobenzisoxazole substrate and the designed catalytic residues (Asp127 and Gln50) were calculated on each snapshot. Snapshots whose θ and d values fall within the allowed ranges were considered to be catalytically competent. Catalytic competency is the percentage of snapshots in the trajectory that are catalytically competent. Only 9 variants (blue) have a catalytic competency that is greater than that of our positive control HG4 (yellow, PDB ID: 5RGF), an HG3 variant that is approximately 700-fold more catalytically efficient than HG3 [1].

Table 2.5: Kinetic Parameters of Kemp Eliminases

enzyme	mutations ^a	# mutations	k_{cat} (s ⁻¹) ^b	K_M (mM) ^b	k_{cat}/K_M (M ⁻¹ s ⁻¹) ^b	relative k_{cat}/K_M
HG series – controls						
HG3	–	0	N.D.	N.D.	125 ± 8 (146 ± 6) ^c	1
HG3.3b	V61 K50H M84C S89R Q90D A125N	6	N.D.	N.D.	3400 ± 700 (2200 ± 100) ^c	27
HG3-K50Q/M84C	K50Q M84C	2	N.D.	N.D.	4300 ± 300	34
HG3.7	V61 Q37 K50Q M84C S89R Q90H A125N	7	N.D.	N.D.	24,000 ± 3000 (27,000 ± 2000) ^c	192
HG series – designs						
HG378	A21C K50Q Q90L A125V W275V	5	N.D.	N.D.	70 ± 5	0.6
HG57	A21C K50Q Q90L A125V	4	N.D.	N.D.	120 ± 8	1
HG44	A21C K50Q Q90F A125V	4	N.D.	N.D.	250 ± 10	2
HG603	K50Q Q90F A125V W275V	4	N.D.	N.D.	800 ± 10	6.4
HG493	K50Q Q90F A125T W275T	4	N.D.	N.D.	2100 ± 100	17
HG185	A21C K50Q Q90F A125T W275V	5	N.D.	N.D.	2300 ± 100	18
HG198	K50Q M84C A125T F267M	4	N.D.	N.D.	8000 ± 300	64
HG259	K50Q M84C Q90L A125T F267M	5	N.D.	N.D.	13,000 ± 400	104
HG630	K50Q M84C A125T F267M W275V	5	N.D.	N.D.	15,000 ± 500	120
<i>without MD filtering^d</i>						
HG649	K50Q M84C Q90F A125T F267M W275V	6	N.D.	N.D.	32,000 ± 3000	256
KE70 series – controls						
KE70	–	0	N.D.	N.D.	57 ± 3 (78 ± 14) ^c	1
KE70-S20a/A240a ^e	S20a A240a	2	N.D.	N.D.	150 ± 7	2.6
R2 7/12F	D23G Y48F D212E H251Y	4	0.50 ± 0.01	1.00 ± 0.06	500 ± 30 (1330 ± 13) ^c	9
R4 8/5A ^f	S20a K29N Y48F W72C H166Y N197N A204V G240a	8	3.0 ± 0.1	1.0 ± 0.1	3000 ± 300 (4380 ± 350) ^c	53
KE70 series – designs						
KE701 ^g	H16D S20a D45Q Y48F W72C G101S S138Q H166N V168A A204V A240a	11	0.040 ± 0.001	0.10 ± 0.02	400 ± 80	7
KE703 ^g	H16D S20a D45N Y48F W72V G101M S138Q H166N V168A A204V A240a	11	0.50 ± 0.01	0.20 ± 0.02	2500 ± 300	44
KE703b ^g	S20a Y48F W72V G101M S138Q H166N V168A A204V A240a	9	N.D.	N.D.	5500 ± 100	96

^a Mutations are relative to the starting template (HG3 or KE70). ^b Data represent the average of six to 12 individual replicate measurements from two to four independent protein batches. N.D. indicates that individual parameters K_M and k_{cat} could not be determined accurately because saturation was not possible at the maximum substrate concentration tested (2 mM, Figures 2.5 and 2.11), which is the substrate’s solubility limit. In these cases, catalytic efficiencies (k_{cat}/K_M) were calculated from the slope of the linear portion ($[S] \leq K_M$) of the Michaelis-Menten model ($v_0 = (k_{\text{cat}}/K_M)[E_0][S]$). Errors of regression fitting, which represent the absolute measure of the typical distance that each data point falls from the regression line, are provided. ^c Values in parentheses are from Broom et al.[2] (HG series), or Röthlisberger et al.[48] and Khersonsky et al.[92] (KE70 series). ^d HG649 is one of the 55 designs that passed all filtering criteria prior to the final MD step. ^e These enzymes contain two insertions, S20a and A240a/G240a. ^f The KE702 design was inactive.

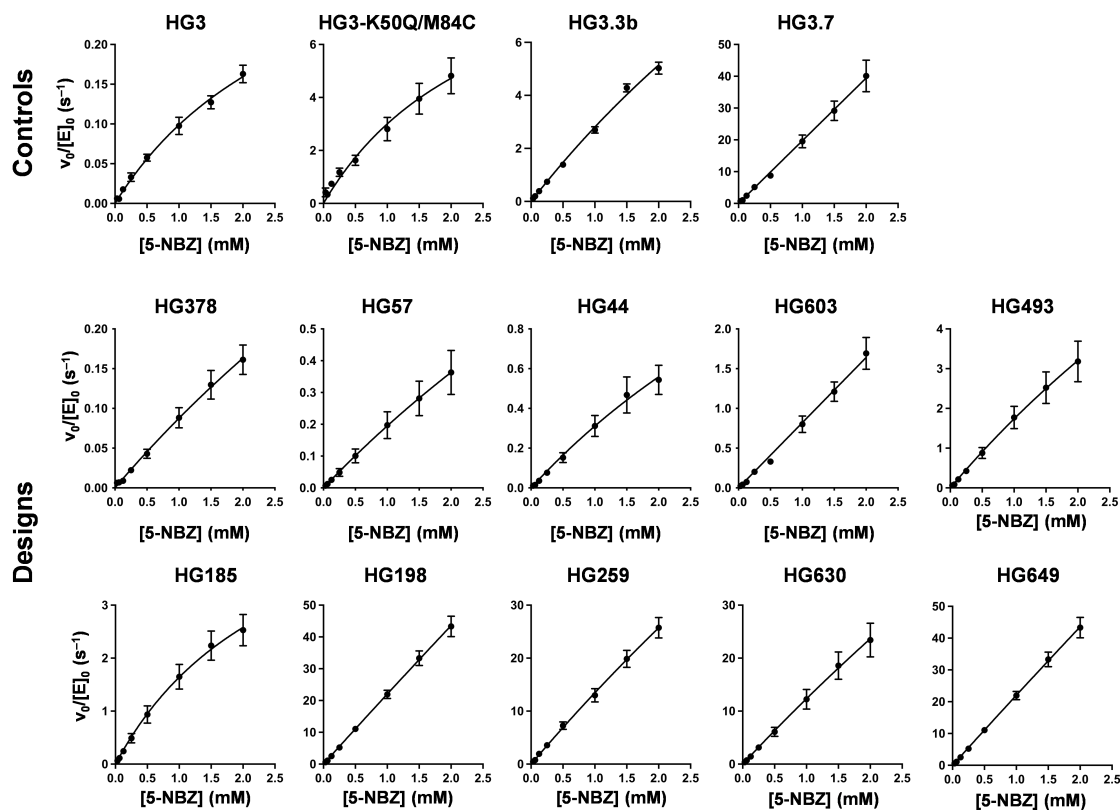


Figure 2.5: Steady-state kinetics of HG-series Kemp eliminases. Michaelis–Menten plots of normalized initial rates as a function of 5-nitrobenzisoxazole (5-NBZ) concentrations are shown. Data represent the average of six or nine individual replicate measurements from two or three independent protein batches, with error bars indicating the SEM. Saturation was not achieved for any enzyme at the substrate’s solubility limit (2 mM). Therefore, only k_{cat}/K_M values are reported on Table 2.5, and these values were calculated using linear regression of rates measured at the three or four lowest substrate concentrations.

a variant found after seven rounds of directed evolution involving the screening of >7000 mutants [49] (Table 2.5, Figure 2.6a).

HG649 shares a majority of its mutation sites with HG3.7, but the combination of specific amino acid substitutions is distinct, with only two mutations in common (Figure 2.6b). These results demonstrate how ensemble-based enzyme design produces similar activity increases as directed evolution but without extensive experimental screening. Interestingly, HG649 was one of the 55 designs that passed all filtering criteria except for the final MD analysis, which led to its elimination due to its low predicted catalytic competency (Figure 2.4). This false negative, along with the two false positives (Table 2.5, HG378, and HG57), indicates that improvements are required to make the MD filtering more robust, for example, by running multiple replicas and/or longer simulations, and suggests that other eliminated sequences may also display high catalytic efficiency.

2.4.3 X-ray Crystal Structures of the Highly Efficient Enzyme

To evaluate whether the catalytic efficiency of our designs was increased via the predicted molecular mechanisms (Figure 2.1b), we solved room-temperature (280 K) X-ray crystal structures of four of the most active designs (HG185, HG198, HG630, and HG649) in the presence and absence of the 6NT TS analogue. We determined structures at room temperature for direct comparison with the HG3 noncryogenic structure used to generate the ensemble and because this method can reveal conformational heterogeneity in protein structures that would not be visible at cryogenic temperatures, providing insights into the conformational ensemble that is sampled by a protein [55, 102]. Crystals of HG185, HG630, and HG649 were obtained in the absence of 6NT and soaked into solutions containing this molecule prior to X-ray data collection to obtain the bound structures (Table 2.6). These three enzyme variants all crystallize in space group P21 with two molecules in the asymmetric

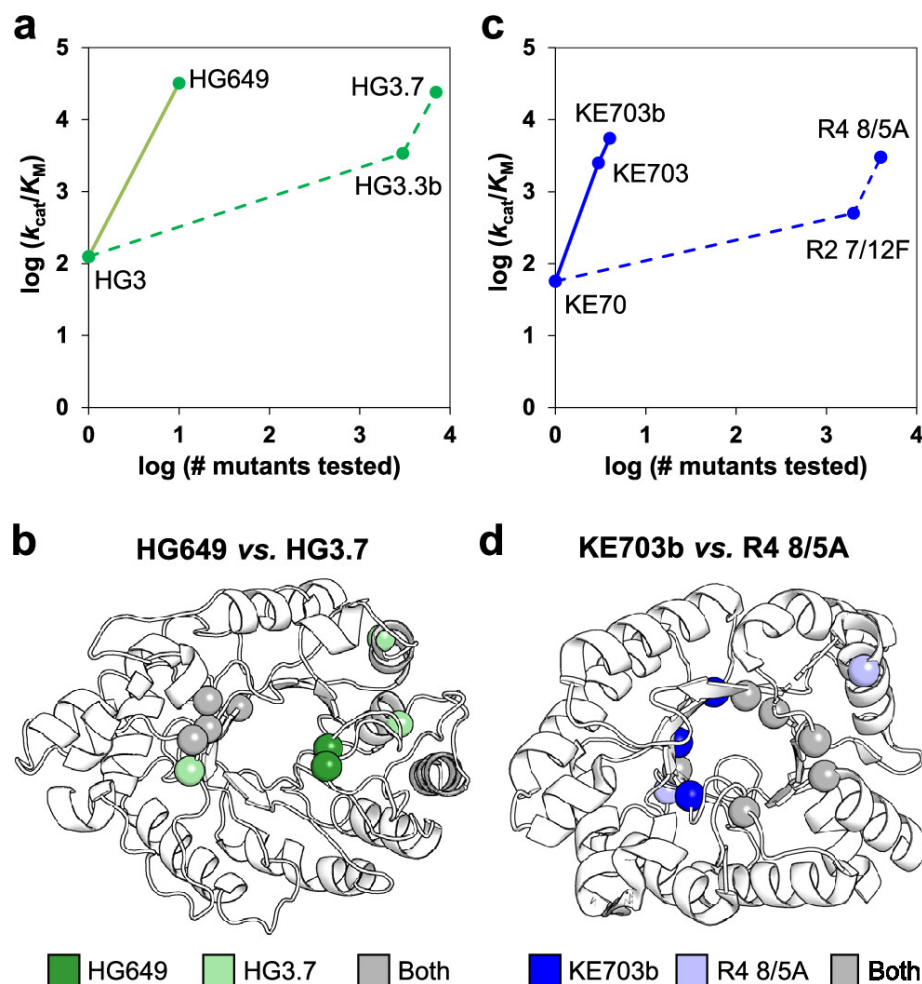


Figure 2.6: Ensemble-based design vs directed evolution. (a) Ensemble-based design yields catalytic efficiency improvements comparable to those achieved by directed evolution with a fraction of the experimental screening effort. Full and dashed lines show ensemble-based design and directed evolution trajectories, respectively. (b) Ensemble-based design identifies a majority of the mutation sites (spheres) found by directed evolution (4/6) but mutation combinations of designed (HG649) and evolved (HG3.7) enzymes are distinct as these share only two mutations (K50Q and M84C). (c) Full and dashed lines show ensemble-based design and directed evolution trajectories, respectively. KE703b differs from KE703 by reversion of its D16/N45 residues to the H16/D45 catalytic dyad found in KE70 and its evolved variants. (d) Designed variant KE703b shares 6 of the 9 mutation sites (spheres) found by directed evolution but has only three mutations in common with evolved variant R4 8/ 5A (S20a, Y48F, A204V).

unit, roughly matching the crystal form reported for HG3 [2]. On the other hand, HG198 could only be crystallized in the presence of 6NT, and its unbound structure was derived from serially soaking these crystals in solutions lacking the TS analogue prior to data collection.

Table 2.6: Crystallization conditions

Enzyme	6NT	Method ^a	Buffer	pH	Protein(mg mL ⁻¹)	Precipitant #1	Precipitant #2
HG185	(-)	Crystallization	100 mM sodium acetate	4.6	12.8	1.6 M ammonium sulfate	–
	(+)	Soaking	100 mM sodium acetate	4.6	12.8	1.6 M ammonium sulfate	–
HG198	(-)	Soaking	100 mM sodium acetate	4.6	6.0	1.8 M ammonium sulfate	–
	(+)	Co-crystallization	100 mM sodium acetate	4.6	6.0	1.6 M ammonium sulfate	–
HG630	(-)	Crystallization	100 mM sodium acetate	4.8	10.6	1.4 M ammonium sulfate	–
	(+)	Soaking	100 mM sodium acetate	4.8	10.6	1.4 M ammonium sulfate	–
HG649	(-)	Crystallization	–	n.d.	12.5	0.3 M ammonium sulfate	30% PEG-4000
	(+)	Soaking	–	n.d.	12.5	0.3 M ammonium sulfate	30% PEG-4000

^a Crystals of ligand-bound HG185, HG630 and HG649 were obtained by soaking ligand-free crystals in 1 L drops of mother liquor containing 5 mM 6NT and 5% DMSO in a single step. Crystals of ligand-free HG198 were obtained by sequentially soaking crystals obtained by co-crystallization with 6NT in three new drops of 1 L mother liquor containing no ligand (30 min incubation per drop).

In the presence of 6NT, HG198 crystallizes in low symmetry space group P1 with three molecules in the asymmetric unit, but removing the TS analogue via serial soaking transforms the crystals to space group P21 with a single molecule in the asymmetric unit. This unusual transformation of the crystal system was consistent over multiple crystal specimens. All crystals diffracted at resolutions of 1.32-1.90 Å, which enabled us to create high-quality structural models (Table 2.7 and Table 2.8).

Table 2.7: Crystallographic data and refinement statistics (HG185 and HG198)

	HG185	HG185	HG198	HG198
6NT	-	+	-	+
PDB ID	8USK	8USL	8USI	8USJ
Data collection				
Temperature (K)	280	280	280	280
Resolution (Å)	1.60	1.55	1.90	1.60
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Cell params.				
a b c (Å)	51.12	51.11	45.68	46.65
	77.44	77.49	58.72	51.03
	93.75	93.85	51.30	86.30
α β γ (°)	90.00	90.00	90.00	87.01
	105.55	105.51	95.36	81.12
	90.00	90.00	90.00	87.01
Chains per asym. unit	2	2	1	3
R _{merge}	0.071	0.065	0.169	0.123
(Highest shell)	(0.542)	(0.470)	(0.517)	(0.547)
CC _{1/2}	0.992	0.994	0.955	0.981
(Highest shell)	(0.474)	(0.531)	(0.493)	(0.466)
I/ σ I	10.3	11.1	6.0	7.7
(Highest shell)	(1.4)	(1.9)	(1.7)	(0.4)
Completeness (%)	96.4	99.7	99.4	91.8
(Highest shell)	(94.1)	(99.9)	(91.8)	(94.3)
Multiplicity	3.4	3.4	1.8	1.8
(Highest shell)	(3.5)	(3.3)	(3.1)	(1.8)
Wilson B-factor (Å ²)	14.880	14.290	9.580	10.130

Continued on next page

Table 2.7 (continued)

	HG185	HG185	HG198	HG198
6NT	-	+	-	+
Refinement				
# unique reflections	89399	101929	21360	97130
(Highest shell)	(4304)	(5042)	(1030)	(4660)
R _{work} /R _{free}	0.160/0.183	0.161/0.181	0.158/0.199	0.150/0.190
No. atoms				
Protein	5068	5116	2408	7489
Ligand	0	24	0	105
Water	543	362	294	820
Averaged B-factors (Å²)				
Protein	19.74	19.04	14.02	14.54
Ligand	—	32.92	—	10.84
Water	35.18	29.88	27.68	30.53
RMSD				
bond lengths (Å)	0.005	0.007	0.003	0.007
bond angles (°)	0.834	0.848	0.607	0.724
MolProbity statistics				
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Ramachandran allowed (%)	1.67	1.57	1.66	1.56
Ramachandran favored (%)	98.33	98.49	98.34	98.44
Rotamer outliers (%)	0.00	0.00	0.00	0.00
MolProbity clashscore	4.15	1.67	1.67	1.73

* Highest resolution shell is shown in parentheses.

Table 2.8: Crystallographic data and refinement statistics (HG630 and HG649)

	HG630	HG630	HG649	HG649
	-	+	-	+
PDB ID	8USG	8USH	8USE	8USF
Data collection				
Temperature (K)	280	280	280	280
Resolution (Å)	1.37	1.32	1.55	1.45
Space group	P1	P1	P2 ₁	P2 ₁
Cell params.				
a b c (Å)	51.12	51.02	51.13	51.07
	77.51	77.51	77.47	77.47
	93.88	93.90	93.78	93.81
$\alpha \beta \gamma$ (°)	90.00	90.00	90.00	90.00
	105.42	105.27	105.39	105.28
	90.00	90.00	90.00	90.00
Chains per asym. unit	2	2	2	2
R _{merge}	0.062	0.052	0.070	0.059
(Highest shell)	(0.443)	(0.585)	(0.513)	(0.552)
CC _{1/2}	0.990	0.995	0.991	0.996
(Highest shell)	(0.551)	(0.477)	(0.505)	(0.452)
I/ σ I	8.2	9.4	7.9	8.7
(Highest shell)	(1.9)	(1.5)	(1.7)	(1.5)
Completeness (%)	99.4	99.7	99.8	99.7
(Highest shell)	(99.4)	(99.7)	(99.4)	(99.7)
Multiplicity	3.2	3.3	3.3	3.2
(Highest shell)	(2.7)	(3.1)	(3.1)	(3.1)
Wilson B-factor (Å ²)	14.910	13.640	13.180	12.510

Continued on next page

Table 2.8 (continued)

	HG630	HG630	HG649	HG649
	-	+	-	+
Refinement				
# unique reflections	147144	164672	105181	124369
(Highest shell)	(6954)	(8177)	(9411)	(6172)
R _{work} /R _{free}	0.127/0.147	0.110/0.147	0.117/0.137	0.117/0.133
No. atoms				
Protein	5166	5168	5188	5260
Ligand	0	24	0	105
Water	557	743	748	743
Averaged B-factors (Å²)				
Protein	19.37	17.87	17.97	17.50
Ligand	–	17.12	–	17.10
Water	36.59	39.27	37.66	35.87
RMSD				
bond lengths (Å)	0.007	0.007	0.006	0.007
bond angles (°)	0.710	0.683	0.835	0.683
MolProbity statistics				
Ramachandran outliers (%)	0.00	0.00	0.00	0.00
Ramachandran allowed (%)	1.51	1.34	1.84	1.17
Ramachandran favored (%)	98.49	98.66	98.16	98.83
Rotamer outliers (%)	0.00	0.00	0.00	0.00
MolProbity clashscore	2.52	2.32	3.57	3.03

* Highest resolution shell is shown in parentheses.

All designs bound 6NT in the same catalytically productive pose (Figure 2.7a) as that observed in HG3, in excellent agreement with the computational models gen-

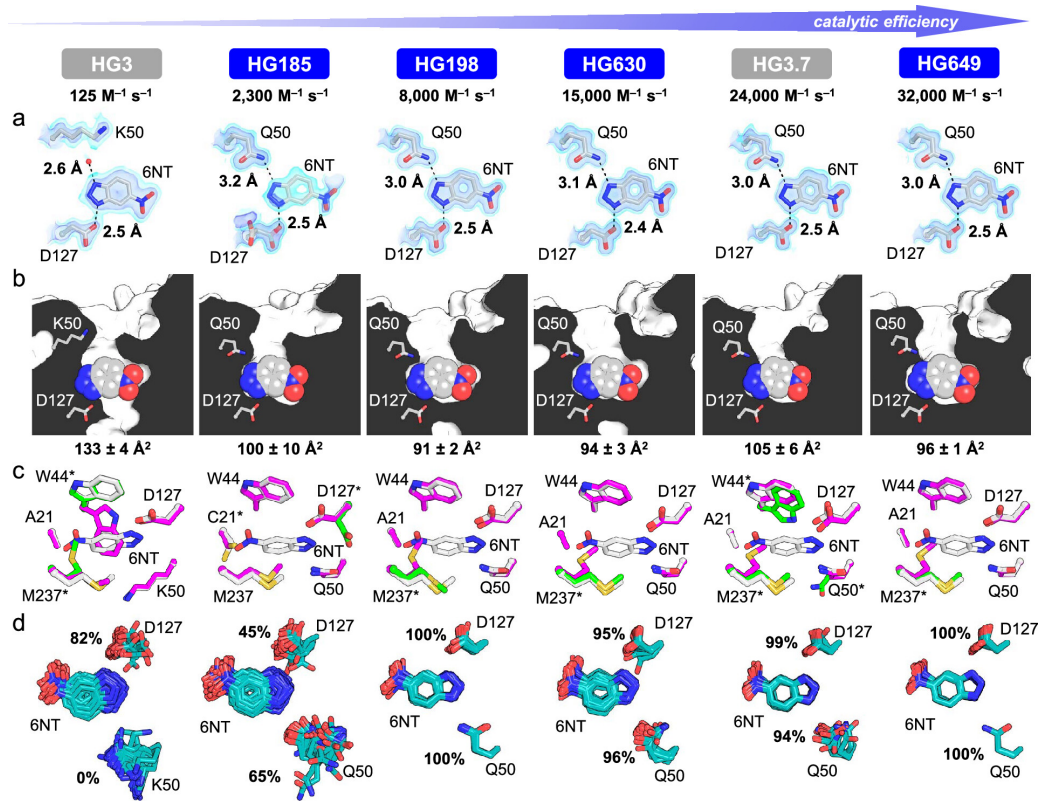


Figure 2.7: Crystal structures of HG-series Kemp eliminases (chain A). HG3 (PDB IDs: 5RG4, 5RGA) and HG3.7 (5RG6, 5RGC) are from Broom et al. [2]. (a) Hydrogen bonds to the 6-benzotriazole (6NT) TS analogue are shown as dashed lines with distances; angles are given in Table 2.16. The red sphere indicates a water molecule. The 2Fo-Fc map is shown in volume at 0.5 and 1.5 $e \text{\AA}^{-3}$ (light and dark blue). HG185 shows ~60% 6NT occupancy, consistent with weaker density. (b) Cutaway view showing improved active-site complementarity to 6NT relative to HG3, reflected by lower solvent-accessible surface areas (mean $\pm$ s.d.); 6NT is shown as spheres. (c) Superposition of 6NT-bound (white) with highest (magenta) and lowest (green) occupancy unbound conformers. Non-preorganized residues are marked with an asterisk. In HG3/HG3.7, W44 and M237 adopt nonproductive rotamers; in HG185, C21 adopts a single conformation overlapping the TS analogue; in HG198, HG630, and HG649, only M237 is nonproductive (47–64% occupancy). (d) Ensemble refinement shows reduced structural heterogeneity in HG198, HG630, HG649, and HG3.7 relative to HG3, and a higher fraction of conformers forming catalytic hydrogen bonds (D127, Q50) with 6NT. In HG185, lower 6NT occupancy (~60%) likely contributes to increased heterogeneity and fewer catalytic contacts.

erated by ensemble-based design but in stark contrast with those generated using the average crystal structure (Figure 2.8), highlighting the increased accuracy afforded by the use of a crystallographically derived ensemble [2].

In this pose, the acidic N-H bond of 6NT that mimics the cleavable C-H bond of the substrate is located within hydrogen-bonding distance to the carboxylate oxygen of D127, while the basic nitrogen atom corresponding to the phenolic oxygen of the TS forms a hydrogen bond with the N ϵ atom of Q50, similar to what was observed for evolved variant HG3.7 [2]. By contrast, HG3 does not form this second hydrogen bond because its K50 residue does not adopt a suitable side-chain conformation, and instead, it is a water molecule that hydrogen bonds to the basic nitrogen atom of 6NT. In all cases, angles defining the hydrogen bond between D127 and the TS analogue were improved or within error compared to HG3 (Table 2.4). In contrast to all other enzymes, the 6NT TS analogue of HG185 was refined to an occupancy of approximately 60%, indicating that its crystal comprises a mixture of enzyme molecules in their bound and unbound forms. Given that the soaking procedure we used to generate crystals of HG185 bound to 6NT was identical to the one used to soak isomorphous crystals of HG630 and HG649 (Table 2.6), which yielded 6NT bound at 100% occupancy, it is likely that the lower occupancy of this ligand in HG185 reflects its weaker binding in the active site, in agreement with the enzyme’s 6.5_14-fold lower catalytic efficiency compared to HG630 and HG649 (Table 2.5). The lower occupancy of 6NT in the HG185 structure is reflected by the weaker electron density in the active site and the presence of an alternate D127 rotamer that does not hydrogen bond with the TS analogue (Figure 2.7a). Regardless, the active site of all designed variants is more complementary to the TS than that of HG3, as indicated by the lower solvent-accessible surface area for 6NT when bound in the active site of these variants (Figure 2.7b).

Next, we evaluated active-site preorganization, which is an important contributor to enzyme catalysis [103], by comparing structures in the presence and absence of

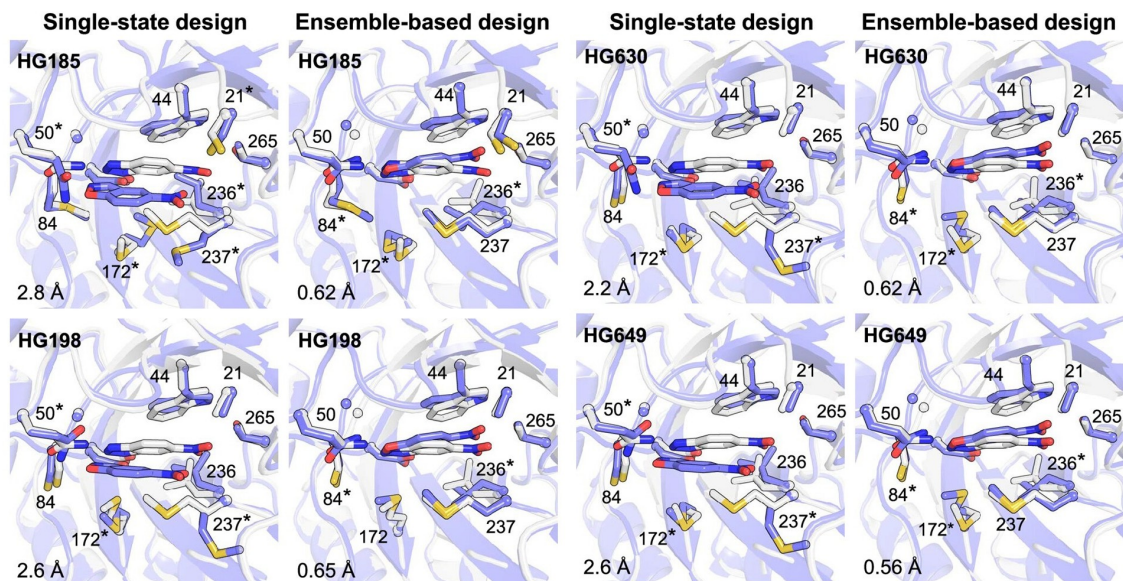


Figure 2.8: Computational models vs crystal structures. Crystal structures of HG-series Kemp eliminases with bound 6NT (white) are overlaid on the corresponding design models obtained by single-state design with the average crystal structure or ensemble-based design (blue). For clarity, only the major conformer of Chain A is shown. Root-mean-square deviations between the TS in the design models and the TS analogue in corresponding crystal structures are indicated at the bottom left. The TS and TS analogue are shown at the center of each image. Side chains of all residues forming the binding pocket are shown with the exception of P45, which was omitted for clarity. The sphere shows the carbon of G83. Asterisks indicate residues that adopt side-chain rotamers varying by $>20^\circ$ around one or more side-chain dihedrals between the design model and crystal structure.

bound 6NT. In HG3, the unbound form is not preorganized for catalysis as both W44 and M237 adopt conformations that would prevent productive binding of 6NT in the major or minor conformations, respectively (Figure 2.7C). This is also the case for evolved variant HG3.7 [2]. HG185 is also poorly preorganized in the unbound form because C21 exclusively adopts a rotamer that would clash with 6NT. Furthermore, the D127 catalytic residue of HG185 adopts a low occupancy (7 and 14% for chains A and B, respectively), catalytically nonproductive conformation in the unbound form that cannot interact favorably with 6NT, likely contributing to its inferior activity compared with the other variants. By contrast, active sites of higher activity variants HG198, HG630, and HG649 are correctly preorganized for catalysis in a substantial portion of the molecules in the crystal, with only M237 adopting a nonproductive conformation at 47-64% occupancy (Figure 2.7c). Overall, crystal structures confirmed that the activity of designed HG variants was increased by several of the mechanisms known to enhance enzyme catalysis and that we aimed to reproduce using our filtering criteria.

2.4.4 Enrichment of Catalytically Competent Substates

The ability of an enzyme to frequently sample productive substates in its conformational ensemble is another important feature of efficient catalysis [1, 97, 104, 105]. To evaluate this property for our designs, we performed ensemble refinement of their 6NT-bound structures and evaluated the proportion of ensemble members that make the designed catalytic contacts with the TS analogue.

We observed that in HG3, the hydrogen bond between the catalytic base and 6NT is formed in 82% of ensemble members and no interaction is observed between the ligand and K50 (Figure 2.7d). In HG185, interactions between 6NT and catalytic residues D127 and Q50 are present in 45 and 65% of ensemble members, respectively, likely due to the lower occupancy of 6NT. In the case of the most active variants

Table 2.9: Structural comparison of HG3 crystal structure with ensembles of HG-series variants

Ensemble ^b	Median pairwise backbone RMSD		Z-score
	from reference structure		between
	(Å) ^a		distributions
	Self	HG3 xtal ^d	HG3 xtal ^d
HG3	0.22	0.17	–
HG198	0.17	0.23	–10.6
HG630	0.12	0.22	–9.2
HG649	0.09	0.19	–4.9

^a Median pairwise backbone RMSDs were calculated for designed active-site residues listed on Table 2.2.

^b Ensembles were generated by ensemble refinement of 6NT-bound crystal structures listed on Table 2.1.

^c Median pairwise backbone RMSD between every pair of backbones from the same ensemble.

^d HG3 crystal structure bound with 6NT (PDB ID: 5RGA, Chain A).

HG198, HG630, and HG649, catalytic contacts to both D127 and Q50 are presented in $\geq 95\%$ of ensemble members, and their ensembles are less structurally diverse than those of HG3 or HG185 (Figures 2.7D and 2.9A, Table 2.9). The higher percentage of hydrogen bonds between 6NT and catalytic residues, and narrower conformational ensembles of high-activity HG variants whose structures contain TS analogues at 100% occupancy, confirm a population of catalytically competent substates higher than in the ensemble of HG3.

Next, we quantified the extent to which ensemble-based design was able to shift the conformational ensemble observed for HG3 by analyzing the crystallographic ensemble models of the 6NT-bound forms of HG3 and the efficient HG variants that bound this TS analogue with 100% occupancy (HG198, HG630, and HG649). First, we calculated the pairwise backbone root-mean-square deviations (RMSDs) for all pairs of structures within a single ensemble model and plotted their distributions (Figure 2.9A). Comparing the more active variants to HG3 demonstrated that all of these improved enzymes have a narrower (more self-consistent) ensemble, whose members have a lower median pairwise RMSD (Table 2.9), which decreases as catalytic efficiency increases (e.g., 0.22, 0.17, 0.12, and 0.09 Å for HG3, HG198, HG630, and HG649, respectively). Next, we wanted to ask whether the ensembles of the more active variants are converging on the most highly populated conformations in the HG3 ensemble or around a different subset of conformations that might provide a better active-site structure. To assess this, we calculated the pairwise RMSD between the single conformer structure of HG3 bound to 6NT (PDB ID: 5RGA[2]) and the individual members of various ensemble models (HG198, HG630, HG649, and HG3). We observed that the single conformer structure of HG3 has a lower median RMSD relative to members of the HG3 ensemble model (0.17 Å) than it does relative to structures present in the ensemble models corresponding to higher activity variants (0.19-0.23 Å, Figure 2.9B, Table 2.9). Taken together, these RMSD calculations suggest that our design procedure narrows the ensembles of the more

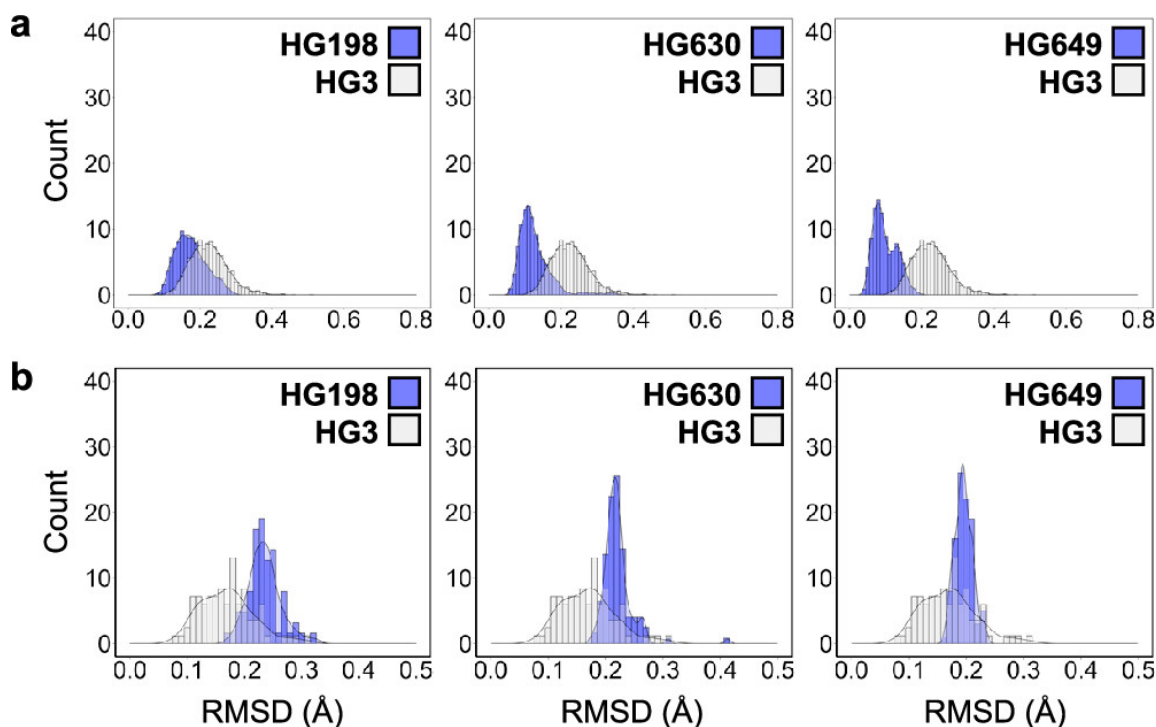


Figure 2.9: Ensemble-based design leads to a shift in the conformational ensemble. (a) Distributions of pairwise backbone RMSDs are calculated for all pairs of structures within a single ensemble model. Comparing the distributions derived from efficient HG variants (HG198, HG630, and HG649) bound to 6NT (blue) with those derived from the HG3 ensemble model, also with bound 6NT (white), reveals that the design process leads to a narrower (more self-consistent) distribution of conformations with a lower median pairwise RMSD. This effect is most pronounced for the most active variants (HG630 and HG69). (b) Distributions of pairwise RMSD between the single conformer structure of HG3 (PDB ID: 5RGA[2]) and the individual members of various ensemble models demonstrate that the ensembles of the improved variants have evolved to include conformations that are significantly different from HG3. RMSD values are calculated over active-site residues listed in Table 2.2. Median RMSD values and Z-scores for comparison of distributions are reported in Table 2.9.

active variants around conformations that are substantially different from those that are most frequently populated by HG3.

2.4.5 KE70 Designs

To further test our ensemble-based enzyme design method, we applied it to another Kemp eliminase that was designed on a different protein scaffold. This enzyme, KE70, contains a His/Asp dyad that acts as catalytic base and a Ser hydrogen-bond donor to stabilize productive substrate binding [48]. To allow direct comparison to our HG enzymes, we designed variants using a theozyme comprising an Asp residue as a catalytic base and a Gln as a hydrogen-bond donor. Starting from the available unbound cryogenic crystal structure of KE70 (PDB ID: 3NPV[92]), we generated an 84-member ensemble (Table 2.1) and used it to design variants predicted to efficiently catalyze the Kemp elimination using a similar protocol as that used for HG enzymes but without the final filtering step involving MD (Methods section). Three variants were selected for kinetic analysis, and these contain between 11 and 13 mutations (Table 2.10), including the D16 and Q138 catalytic motif as well as two insertions in active-site loops (S20a and A240a) that we added after design.

These loop insertions were manually introduced into KE70 during its directed evolution [92], and we added them to allow for direct comparison with evolved KE70 variants that contain these insertions, which increase the catalytic efficiency by 2.6-fold when introduced on their own into KE70 (Table 2.5). Two of the three tested designs were active (Figure 2.11). The most active variant, KE703, displays a $k_{\text{cat}}/K_{\text{M}}$ value of $2500 \pm 300 \text{ M}^{-1} \text{ s}^{-1}$, which is 44-fold higher than KE70 and approximately 17-fold higher than that of a KE70 variant containing the two loop insertions (Table 2.5).

To compare with evolved KE70 variants that contain the His/Asp dyad as a catalytic base, we introduced the D16H and N45D mutations into KE703, generating

```

1 .....10.....20.....30.....40.....50.....60.....70
KE70 1 MTDLKASSLRALKLMHLAT-ANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
KE70-s20a/A240a 1 MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
R2_7/12F 1 MTDLKASSLRALKLMHLAT-ANCDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
R4_8/5A 1 MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
KE701 1 MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
KE703 1 MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR
KE703b 1 MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTDDAIFYPRFIPIARKTLKEQGTPEIR

71 .....80.....90.....100.....110.....120.....130.....140
KE70 70 IWTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
KE70-s20a/A240a 71 IWTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
R2_7/12F 70 IWTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
R4_8/5A 71 ICTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
KE701 71 ICTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
KE703 71 IWTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI
KE703b 71 IWTSTNFPHGNDIDIALAETRAAIAYGADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVI

141 .....150.....160.....170.....180.....190.....200.....210
KE70 140 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
KE70-s20a/A240a 141 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
R2_7/12F 140 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
R4_8/5A 141 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
KE701 141 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
KE703 141 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA
KE703b 141 IETGELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDMGVEKTVGFIPVGGVRTA

211 .....220.....230.....240.....250.....260.....270.
KE70 210 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE70-s20a/A240a 211 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
R2_7/12F 210 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
R4_8/5A 211 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE701 211 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE703 211 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE703b 211 EDAQKYLAIADFLFGADWADARHYAFGAS-SLLASLLKALGHGDGKSASSYGSLEHHHHHH

```

Figure 2.10: Multiple sequence alignment of KE70-series Kemp eliminases.

Table 2.10: Design filtering _ KE70 series

Design	Mutations from KE70	Energy (kcal mol ⁻¹)	SASA (Å ²) ^a	# Cys	# Met	# Preorganized residues ^b
KE701	H16D S20a D45Q Y48F W72C G101S	-29.4	101.0	1	0	12
	S138Q H166N V168A A204V A240a					
KE702	H16D S20a D45Q Y48F R70K W72Y G101M	-17.9	41.4	0	2	12
	S138Q H166N V168A I202V A204M A240a					
KE703	H16D S20a D45N Y48F W72V G101M	-17.8	101.6	0	1	12
	S138Q H166N V168A A204V A240a					

^a Solvent-accessible surface area of the transition state in the design model.
^b Preorganized residues are those predicted to adopt the same rotamer in the presence and absence of transition state.
^c Numbering based on sequence shown on Table 2.15 and Figure 2.10.

the KE703b variant. This variant has a $k_{\text{cat}}/K_{\text{M}}$ value of $5500 \pm 100 \text{ M}^{-1} \text{ s}^{-1}$, which is approximately 100-fold higher than KE70 and 2-fold higher than that of the evolved variant R4 8/5A that was isolated after 4 rounds of directed evolution and screening of approximately 4000 clones [92](Figure 2.6C). KE703b shares a majority of its mutation sites with R4 8/5A, but its specific amino-acid substitutions are distinct, with only three mutations in common (Figure 2.6D). These results confirm the ability of ensemble-based design to generate artificial enzymes that are orders of magnitude more active than the original de novo enzymes at a fraction of the experimental effort required by directed evolution. Furthermore, our results show that ensemble-based design does not require an ensemble generated from a room-temperature crystal structure with a bound TS analogue to be successful, which is advantageous given that such structures are less common than unbound ones solved at a cryogenic temperature.

2.5 Discussion

In this work, we used ensemble-based computational design to increase the catalytic efficiency of the HG3 and KE70 de novo Kemp eliminases by 2 orders of magnitude, yielding artificial enzymes with $k_{\text{cat}}/K_{\text{M}}$ values that are only 4- or 10-fold lower than those of the most active variants evolved from these enzyme prototypes (e.g., $k_{\text{cat}}/K_{\text{M}}$ of 126,000 and 53,100 $\text{M}^{-1} \text{s}^{-1}$ for HG3.17 [2] and KE70 R8 12/12B10 [92], respectively). To achieve these results, we tested a number of variants that are several orders of magnitude lower than those previously used to achieve comparable improvements by directed evolution [49, 92]. The success of our ensemble-based design relied on exploiting the spatiotemporally averaged information contained in X-ray crystallographic data to overcome the backbone sampling challenge that is responsible for much inaccuracy in enzyme design [49]. While others have previously attempted to address the backbone sampling issue by using ensembles generated via Monte Carlo-based backrub sampling and incorporating loop remodeling during the enzyme design process, activity improvements were modest (<5-fold) [92]. These results, in combination with our previous observation that use of an ensemble generated by unconstrained MD does not allow accurate recapitulation of catalytic contact and active-site configurations for an efficient artificial enzyme, [2] suggest that to obtain large activity increases, it is not sufficient to introduce backbone flexibility in the design process. Instead, the correct backbone conformation must be sampled. Crystallographic ensemble refinement of the initial de novo enzyme satisfies this stringent requirement in that the resulting model represents the true conformational ensemble sampled by the enzyme, which should include productive substates that are better suited for stabilization of the TS than the average crystal structure. Generating the ensemble from a structure with a bound TS analogue likely helps to enrich such productive substates and could partially explain why the KE70 ensemble, which was derived from an unbound structure, led to a lower overall activity improvement. Al-

though the use of an ensemble derived from enzymes in their bound and unbound forms led to activity improvements, one important disadvantage of our approach is that it requires a crystal structure of the original de novo enzyme to generate the ensemble. Since de novo enzymes are typically generated from small, monomeric protein scaffolds whose crystal structures have been determined to high resolution, we do not anticipate that structure determination will create a substantial bottleneck in most cases.

Our results support the idea that to design efficient artificial enzymes, it is essential to consider residual conformational heterogeneity present in the designs and whether it leads to the population of catalytically unproductive conformations. In other words, dynamic aspects of enzyme structure, such as active-site preorganization and enrichment of catalytically competent substates, must also be considered. This is supported by the fact that the most active designs described here, HG630 and HG649, were ranked lower than less active ones such as HG44 and HG57 based on their predicted ability to stabilize the TS (as reflected by computed energies, Table 2.3). Recently, researchers have proposed that enzyme design should be viewed as a population shift problem, where relative stabilities of conformational substates are redistributed by the introduction of mutations to stabilize productive substates. [1, 32] Previous work on the HG series of Kemp eliminases by our team [2] and others [1] demonstrated that directed evolution improves activity through essentially the same mechanism, with higher activity variants having reduced flexibility and greater active-site preorganization, both of which favor the population of catalytically active substates. The similarity is further highlighted by the fact that our ensemble-based design approach and directed evolution identify activity-enhancing mutations at many of the same positions.

Moving forward, we expect that ensemble-based design will find application in the creation of de novo enzymes for other organic transformations. However, for complex reactions involving multiple steps and transient species, additional developments will

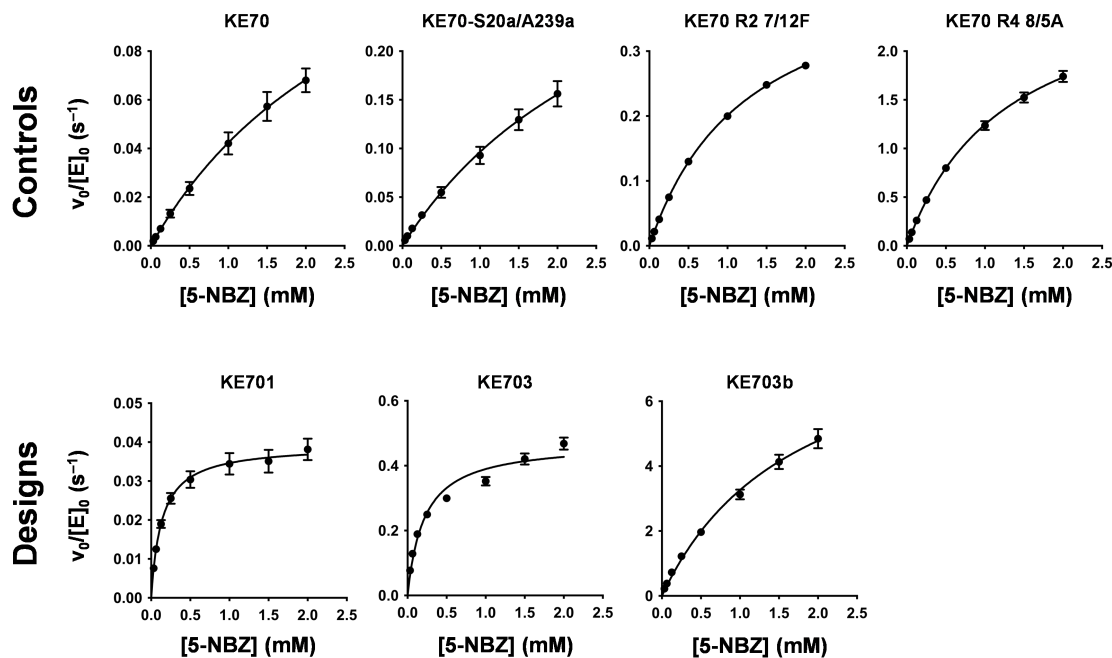


Figure 2.11: **Steady-state kinetics of KE70-series Kemp eliminases.** Michaelis–Menten plots of normalized initial rates as a function of 5-nitrobenzisoxazole (5-NBZ) concentrations are shown. Data represent the average of six or nine individual replicate measurements from two or three independent protein batches, with error bars indicating the SEM. Saturation was not achieved for some enzymes at the substrate’s solubility limit (2 mM). In these cases, only k_{cat}/K_M values are reported on Table 2.5, and these values were calculated using linear regression of rates measured at the three or four lowest substrate concentrations.

be required to make de novo enzyme design robust. For example, improvement of scoring functions used to evaluate sequences is necessary to accurately model forces that stabilize TSs in enzyme active sites, and implementation of multistate design approaches [106] for fine-tuning conformational transitions will be required to control progression through the catalytic cycle. We anticipate that ensemble-based design will be extended for achieving the latter objective, as it could facilitate the identification of sequences that stabilize multiple TSs by using multiple backbone templates from the ensemble. Additionally, we envision using crystallographic ensemble models to further enhance the population of catalytically competent conformations by identifying sequences that can stabilize productive substates while destabilizing non-productive ones. The ability to remodel an enzyme conformational landscape [107] in this way, using negative design strategies to widen the energy gap between productive and nonproductive conformations, may open the door to the design of more efficient artificial enzymes than previously possible. Beyond the scope of enzyme design, we expect that a similar strategy, which takes advantage of crystallographically derived ensemble models, could find utility in a broad range of protein design applications. For example, directed evolution following computational design of protein-protein interactions has been shown to enhance affinity by eliminating conformations from the ensemble that are nonproductive for binding [102]. Additionally, conformational heterogeneity is detrimental to the functions of computationally designed fluorescent proteins [108] and ligand-binding proteins. [109] Each of these observations suggests a potential application of our proposed approach.

2.6 Methods

2.6.1 Ensemble Generation

Ensembles of backbone templates were generated using the ensemble refinement protocol [69] implemented in the PHENIX macromolecular structure determination software (v1.18.2-3874) [110]. In ensemble refinement, local atomic fluctuations are sampled by using MD simulations accelerated and restrained by electron density to produce ensemble models fitted to diffraction data. Briefly, input crystal structures were edited to remove low-occupancy conformers and assign an occupancy of 1.0 to the remaining conformer. Following the addition of riding hydrogens, parallel ensemble refinement simulations were performed using various combinations of the parameters pTLS (0.6, 0.8, 0.9, 1.0), τx (0.5, 1.5, 2.0), and wX-ray (2.5, 5.0, 10.0), where pTLS describes the percentage of atoms included in a translation-libration-screw (TLS) model use to remove the effects of global disorder, τx is the simulation time step, and wX-ray is the coupled temperature bath offset, which controls the extent to which the electron density contributes to the simulation force field such that the simulation runs at a target temperature of 300 K. The ensemble generated from each crystal structure that displayed the lowest Rfree value was selected for design and analysis (Table 2.1).

2.6.2 Computational Enzyme Design

All calculations were performed with the Triad protein design software [111] (Protabit, Pasadena, CA, USA) using a Monte Carlo with simulated annealing search algorithm for rotamer optimization. Individual backbone templates from each structural ensemble were prepared for computational design using the addH.py application within Triad, which optimized and renamed hydrogen atoms, assigned protonation state to His residues, flipped His/Asn/Gln side chains if necessary, and rotated hy-

Table 2.11: Geometric constraints used to define catalytic contacts during computational design

Contact	Residue	Type	Atom				Min ^b	Max ^b
			1 ^a	2 ^a	3 ^a	4 ^a		
Catalytic base	Asp	Distance	OD1 or OD2	H3			1.0 (1.0)	1.6 (1.6)
		Angle	CG	OD1 or OD2	H3		109 (109)	131 (131)
		Angle	OD1 or OD2	H3	C3		159 (159)	180 (180)
		Torsion	CB	CG	OD1 or OD2	H3	-21 (-21)	21 (21)
H-bond donor	Gln	Distance	1HE2 or 2HE2	O1			1.2 (1.2)	2.3 (2.3)
		Angle	NE2	1HE2 or 2HE2	O1		145 (145)	157 (157)
		Angle	1HE2 or 2HE2	O1	N2		120 (120)	140 (140)
		Torsion	1HE2 or 2HE2	O1	N2	C3	160 (160)	200 (200)
	Asn	Distance	1HD2 or 2HD2	O1			1.2 (1.2)	2.3 (2.3)
		Angle	ND2	1HD2 or 2HD2	O1		145 (145)	157 (157)
		Angle	1HD2 or 2HD2	O1	N2		120 (120)	140 (140)
		Torsion	1HD2 or 2HD2	O1	N2	C3	160 (160)	200 (200)
	Thr	Distance	HG1	O1			1.2 (1.2)	2.3 (2.3)
		Angle	OG1	HG1	O1		145 (145)	157 (157)
		Angle	HG1	O1	N2		120 (120)	140 (140)
		Torsion	HG1	O1	N2	C3	160 (160)	200 (200)
	Ser	Distance	HG	O1			1.2 (1.2)	2.3 (2.3)
		Angle	OG	HG	O1		145 (145)	157 (157)
		Angle	HG	O1	N2		120 (120)	140 (140)
		Torsion	HG	O1	N2	C3	160 (160)	200 (200)
Pi-stacking ^c	Phe	Distance between aromatic planes	CE1 CD2	C4 C9			3.5 (3.5)	4 (4)
		Angle between aromatic planes	CE1 CR2	C5 N2			0 (120)	40 (180)

^a Atoms in bold are from the transition state. All other atoms are from the catalytic residues.

^b Distance measurements given in Å, all others are in degrees. Values are for theozyme placement, while those in parentheses are for active-site repacking.

^c This catalytic contact was used for KE70 designs only.

droxyl, sulfhydryl, methyl, and amino groups in Ser/Thr, Cys, Met, and Lys, respectively.

Table 2.12: Geometric definitions for generation of transition-state poses off the side chain of catalytic base

Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Values ^b
Asp	Distance	OD1 or OD2	H3			1.0, 1.2, 1.5
	Angle	CG	OD1 or OD2	H3		112, 117, 122
	Angle	OD1 or OD2	H3	C3		159, 164, 169, 174, 179
	Torsion	CB	CG	OD1 or OD2	H3	0, 5, 10, 170, 175, 180
	Torsion	CG	OD1 or OD2	H3	C3	170, 175, 180, 185, 190
	Torsion	OD1 or OD2	H3	C3	N2	0, 5, 170, 175, 180

^a Atoms in bold are from the transition state. All other atoms are from the catalytic residue.

^b Distance measurements given in Å, all others are in degrees.

The Kemp elimination TS was built using parameters described by Privett and co-workers. [47] For HG3 designs, D127 was used as the catalytic base, and Gln/Asn/Thr/Ser was tested as hydrogen-bond donors at various positions (Table 2.2). For KE70 designs, D16 served as the catalytic base (corresponding to position 17 in the KE70 crystal structure, PDB ID: 3NPV), Q138 as hydrogen bond donor, and F48 was utilized for the introduction of a pi-stacking interaction with the TS (Table 2.14). For both HG3 and KE70 designs, all other positions in the active site were mutated to Gly during theozyme placement to avoid steric clashes with the designed catalytic residues. The 2002 Dunbrack backbone-independent rotamer library [112] with expansions of ± 1 standard deviation around $\chi 1$ and $\chi 2$ was used to provide side-chain conformations. A library of TS poses was generated in the active site by targeted ligand placement [86] using the contact geometries listed in Table 2.12. TS pose energies were calculated using a modified version of the Phoenix protein design energy

function [111] consisting of a Lennard-Jones 12-6 van der Waals term from the Dreiding II force field [113] with atomic radii scaled by 0.9, a direction-dependent hydrogen bond term with a well depth of 8.0 kcal mol⁻¹ and an equilibrium donor-acceptor distance of 2.8 Å,[114] and an electrostatic energy term modeled using Coulomb's law with a distance-dependent dielectric of 10. During the energy calculation step, TS-side-chain interaction energies were biased to favor interactions that satisfy contact geometries (Table 2.11) by applying an energy benefit of -100 kcal mol⁻¹ as described by Lassila et al.[86].

Table 2.13: Amino-acid sequences of Kemp eliminases from HG series

Enzyme	# Mutations from HG3	Sequences
HG44	4	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIIQADFGMVEENSMQNDAT PSQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITL TMTRYKGGKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKL IYMDYLDDASAYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILA SAGTPEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFD GNFNPKAYNAIVQDLQQSGIERGHHHHH
HG57	4	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIIQADFGMVEENSMQNDAT PSQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLT MTRYKGGKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIY MDYLDDASAYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASA GTPEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGN FNPKAYNAIVQDLQQSGIERGHHHHH
HG185	5	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIIQADFGMVEENSMQNDAT PSQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITL TMTRYKGGKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKL IYMDYLDDASAYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILA SAGTPEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFD GNFNPKAYNAIVQDLQQSGIERGHHHHH
HG198	4	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIIQADFGMVEENSMQNDAT PSQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLT MTRYKGGKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIY

Continued on next page

Table 2.13 (continued)

Enzyme	# Mutations from HG3	Sequences
		MDYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASA GTPEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGN FNP KAYNAIVQDLQQSGIERGHHHHH
HG259	5	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIHQADFGMVEENSMQNDATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYK GKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYM DYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGT PEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNP KAYNAIVQDLQQSGIERGHHHHH
HG378	5	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIHQADFGMVEENSMQNDATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYK GKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYM DYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGT PEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNP KAYNAIVQDLQQSGIERGHHHHH
HG493	4	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIHQADFGMVEENSMQNDATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYK GKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYM DYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGT PEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNP PKAYNAIVQDLQQSGIERGHHHHH
HG603	4	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIHQADFGMVEENSMQNDATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYK GKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYM DYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGT PEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNP KAYNAIVQDLQQSGIERGHHHHH
HG630	5	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAIHQADFGMVEENSMQNDATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYK GKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYM DYLDDASASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGT PEVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNP

Continued on next page

Table 2.13 (continued)

Enzyme	# Mutations from HG3	Sequences
		KAYNAIVQDLQQSGIERGHHHHH
HG649	6	MAEAQSDVQLIKARGKYVFGVCTDQNRLTKGNAAHQADFGMVEENSMQNATP SQGNENFAGADYLVNWAQNGGKLGGMQLVWSHFLPSWSSDILTNVMKNHITLTM TRYKGGKIRVDVVGFEANEDGSLRQTVELNVIGEDYIPIAFQTRARADPNAKLIYMD YLDDSASYRKTQAIVNRVKQKRAGVRIDGGSQHTLSAGQAGVQLAPLILASAGTP EVSILLVDAGASPTDVVNVACLYVQSCVGITVGVADPDSVRASSTTPLFDGNFNPK AYNAIVQDLQQSGIERGHHHHH

Next, active-site repacking calculations were performed on the theozyme-containing templates. In these calculations, the TS was translated ± 0.4 Å in each Cartesian coordinate in 0.2 Å increments and rotated 10° about all three axes (origin at TS geometric center) in 5° increments for a total combinatorial rotation/translation search space of 15,625 possible poses. The identities of catalytic residues were fixed and allowed to sample all rotamers of that amino acid type. Residues that were converted to Gly in the theozyme placement step were allowed to sample rotamers of selected amino acids (Tables 2.2 and 2.14) to form a mostly hydrophobic active site. Side-chain-TS interaction energies were biased to favor contacts satisfying geometries as done during the theozyme placement step (Table 2.11). Rotamer optimization was carried out using the search algorithm, rotamer library, and energy function described above.

After active-site repacking, we calculated energy differences between each repacked structure and the corresponding structure in the absence of the TS where all designed residues (including catalytic residues) were converted to Gly. This energy subtraction was done to allow a comparison of energies for sequences across multiple backbone templates. We also measured the angles defining the hydrogen-bond contact between the catalytic base and the TS of the top-scoring sequence obtained from each backbone template. Sequences from the HG3 and KE70 backbone template that gave

Table 2.14: Amino-acid positions optimized during computational design of KE70 variants

Design objective	Designed residues ^a	Allowed amino acids
Catalytic base	H17 ^b	D
H-bond donor	S138	Q
Pi-stacking	Y48	F
Removal of original catalytic dyad ^c	D45	N, Q
Transition-state packing ^d	A19 W72 G101 A103 I140	A, F, I, L,
	H166 V168 S170 I202 A204	M, N, V, Y
Surface residue ^e	R70	K, R

^a Numbering based on crystal structure of KE70 (PDB ID: 3NPV). For comparison with sequences of designed variants, see Figure 2.10.

^b Position 17 in the 3NPV structure corresponded to position 16 in the sequence of our designed variants. See sequence alignment on Figure 2.10.

^c KE70 contains the H17/D45 catalytic dyad. The D45N and D45Q mutations were introduced to make this residue compatible with the newly designed catalytic base (D16).

^d C and S were also allowed at positions W72 and G101 instead of A.

^e This surface residue is located next to D45.

the lowest energy top-ranked sequence with angles between its catalytic base and the TS within 3 or 3.5% error of the optimal geometries calculated for acetamide dimers [101] (Figure 2.3, Table 2.4), respectively, were selected for library design and cleaning, as described later.

2.6.3 Computational Library Design

with the CLEARSS algorithm [115] using the energies obtained after active-site repacking on the single template selected as described earlier. For a specific library size configuration, which is the specific number of amino acids at each position in the protein, the highest probability set of amino acids at each position (based on the sum of their Boltzmann weights) were included in the library. A combinatorial library of approximately 1000 sequences derived from the selected HG3 or KE70 ensemble member was generated. To ensure that the lowest energy rotameric configuration was obtained for each library sequence, we used the `cleanSequences.py` app in Triad to “clean” their structures by optimizing rotamers on the input backbone template using the active-site repacking protocol described above, but without modification of the catalytic residue rotameric configuration. To compare energies of “cleaned” sequences, we calculated the energy difference between each optimized structure and the corresponding structure in the absence of the TS where all designed residues (including catalytic residues) were converted to Gly. These energies are reported throughout the figures and text.

2.6.4 Design Filtering

To decrease the number of sequences to test experimentally, “cleaned” structures were filtered by using geometrical criteria. We kept sequences with (i) a TS solvent-accessible surface area lower than 100 or 102 Å² for HG3 and KE70 designs, respectively, and (ii) $\geq 70\%$ of all designed active-site residues predicted to be preorganized,

Table 2.15: Amino-acid sequences of Kemp eliminases from KE70 series

Enzyme	# Mutations from KE70	Sequences
KE70_S20a A240a	2	MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTTDAF IYIYPRIFIARKTLKEQGTPEIRIWTSTNFPHGNDIDIALAETRAAIAY GADGVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLSVIIETG ELKDEALIRKASEISIKAGADHIVTSTGKVAVGATPESARIMMEVIRDM GVEKTVGFIPAGGVRTAEDAQKYLAIADELFGADWADARHYAFGASA SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE701	11	MTDLKASSLRALKLMDLATSANDDDTDEKVIALCHQAKTPVGTTQAI FIYPRFIPIARKTLKEQGTPEIRICTSTNFPHGNDIDIALAETRAAIAY GADSVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLQVIIETG ELKDEALIRKASEISIKAGADNIATSTGKVAVGATPESARIMMEVIRDM GVEKTVGFIPVGGVRTAEDAQKYLAIADELFGADWADARHYAFGASA SLLASLLKALGHGDGKSASSYGSLEHHHHHH
KE701	11	MTDLKASSLRALKLMDLATSANDDDTDEKVIALCHQAKTPVGTTQAIF IYPRFIPIARKTLKEQGTPEIRICTSTNFPHGNDIDIALAETRAAIAYGA DSVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLQVIIETGELK DEALIRKASEISIKAGADNIATSTGKVAVGATPESARIMMEVIRDMGVE KTVGFIPVGGVRTAEDAQKYLAIADELFGADWADARHYAFGASASLLA SLLKALGHGDGKSASSYGSLEHHHHHH
KE702	13	MTDLKASSLRALKLMDLATSANDDDTDEKVIALCHQAKTPVGTTNAIF IYPRFIPIARKTLKEQGTPEIKIYTSTNFPHGNDIDIALAETRAAIAYGA DMVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLQVIIETGELK DEALIRKASEISIKAGADNIATSTGKVAVGATPESARIMMEVIRDMGVEK TVGFVPMGGVRTAEDAQKYLAIADELFGADWADARHYAFGASASLLAS LLKALGHGDGKSASSYGSLEHHHHHH
KE703	11	MTDLKASSLRALKLMDLATSANDDDTDEKVIALCHQAKTPVGTTNAIF IYPRFIPIARKTLKEQGTPEIRIVTSTNFPHGNDIDIALAETRAAIAYGA DMVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLQVIIETGELK DEALIRKASEISIKAGADNIATSTGKVAVGATPESARIMMEVIRDMGVEK TVGFIPVGGVRTAEDAQKYLAIADELFGADWADARHYAFGASASLLASL LKALGHGDGKSASSYGSLEHHHHHH
KE703b	9	MTDLKASSLRALKLMHLATSANDDDTDEKVIALCHQAKTPVGTTDAIFI YPRFIPIARKTLKEQGTPEIRIVTSTNFPHGNDIDIALAETRAAIAYGAD MVAVVFPYRALMAGNEQVGFDLVKACKEACAAANVLLQVIIETGELKD EALIRKASEISIKAGADNIATSTGKVAVGATPESARIMMEVIRDMGVEKT VGFIPVGGVRTAEDAQKYLAIADELFGADWADARHYAFGASASLLASLL KALGHGDGKSASSYGSLEHHHHHH

Table 2.16: Angles of hydrogen bonds between transition state analogue 6NT and catalytic residues in Kemp eliminase crystal structures

Enzyme	PDB ID	Angle (°) ^a	
		D127	Q50
HG3	5RGA	171.8	— ^b
HG44	8USL	165.0	143.1
HG198	8USJ	168.5	150.3
HG630	8USH	170.6	150.7
HG637	8RGC	166.0	151.2
HG649	8USF	166.8	152.5

^a All angles are for Chain A.

^b HG3 does not contain the Q50 residue. In this enzyme, the hydrogen bond donor is a water molecule.

which was determined as those that adopted the same rotamer in the absence and presence of TS. Furthermore, sequences that contained ≥ 5 Met residues or ≥ 2 Cys residues were discarded to reduce conformational heterogeneity of the active site and prevent formation of disulfide bonds, respectively. A final filtering step consisting of short MD simulations was performed for HG3 designs, as described later.

2.6.5 Unconstrained MD Simulation for Filtering of HG3 Designs

All simulations were performed using the Amber 2020 software (<http://ambermd.org/>) with the AMBER19SB force field.[116] Longrange electrostatics (>10 Å) were modeled using the particle mesh Ewald method,[117] and a time step of 2 fs was used for the production phase. Parameters for the 5-nitrobenzisoxazole substrate were generated using the Antechamber [118] package.

“Cleaned” computational models generated by Triad for the 55 sequences that passed the filtering procedure described earlier were prepared for MD. The TS structure was converted to that of the substrate, hydrogen atoms were added using Reduce,[119] and each substrate-bound protein molecule was placed in a dodecahe-

```

1 .....10.....20.....30.....40.....50.....60.....70
HG3 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG44 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG57 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG185 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG198 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG259 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG378 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG493 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG603 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG630 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY
HG649 1 MAEAAQSVVDQLIKARGKVYFGVATDQNRLTTGKNAAI IQADFGMVWPENSMQWDATEPSQGNFNFAGADY

71 .....80.....90.....100.....110.....120.....130.....140
HG3 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG44 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG57 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG185 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG198 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG259 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG378 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG493 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG603 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG630 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS
HG649 71 LVNWAQQNGKLIGGGMLVWHSLPSWVSSITDKNTLTNVMKNHITTLTRYKGKIRVWDVVGEAFNEDGS

141 .....150.....160.....170.....180.....190.....200.....210
HG3 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG44 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG57 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG185 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG198 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG259 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG378 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG493 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG603 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG630 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT
HG649 141 LRQTVFLNVIGEDYIPIAFQTARAADPNAKLYIMDYNLDSASYPKTQAI VNRVKQWRAAGVPIDGIGSQT

211 .....220.....230.....240.....250.....260.....270.....280
HG3 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG44 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG57 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG185 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG198 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG259 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG378 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG493 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG603 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG630 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS
HG649 211 HLSAGQGAGVLQALPLLASAGTPEVSIIMLDVAGASPTDYVNVVNACLVQSCVGITVFGVADPDSWRAS

281 .....290.....300.....310.....
HG3 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG44 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG57 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG185 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG198 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG259 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG378 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG493 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG603 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG630 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH
HG649 281 TTPLLFDGNFNPKPAYNAIVQDLQQGSIEGRGHHHHHH

```

Figure 2.12: Multiple sequence alignment of HG-series Kemp eliminases.

dral box with periodic boundary conditions where the distance between the protein surface and the box edges was set to 10 Å. After explicit TIP3P water molecules [120] were added, charges on protein atoms were neutralized with Na⁺ and Cl⁻ counterions at a concentration of 0.15 M. The structures were then energy minimized with the steepest descent method to a target maximum force of 1000 kJ mol⁻¹ nm⁻¹. Before equilibration, the system was heated to a target temperature of 300 K for 100 ps. The system was then equilibrated under an NVT ensemble for 100 ps at a temperature of 300 K using a Nose-Hoover thermostat,[121] while applying position restraints for heavy protein atoms. A second equilibration step under an NPT ensemble was performed for 100 ps with constant pressure and temperature of 1 bar and 300 K, respectively, using the Berendsen barostat.[122] Following the removal of position restraints, 20 ns production runs under Parrinello-Rahman pressure coupling [123] were initiated from the final snapshot of the NPT equilibration. After MD, designs were evaluated and ranked according to their catalytic competency, which is the percentage of snapshots in which catalytic contacts between substrate and catalytic side chains were maintained, as calculated using distance and angle metrics listed in Figure 2.4.

2.6.6 Protein Expression and Purification

Codon-optimized (*E. coli*) and his-tagged (C-terminus) genes for Kemp eliminases (Tables 2.13-2.15, Figures 2.12 and 2.10) cloned into the pET-29b(+) vector via NdeI and XhoI were obtained from Twist Bioscience. Enzymes were expressed in *E. coli* BL21-Gold (DE3) cells (Agilent) using lysogeny broth (LB) supplemented with 100 g mL⁻¹ kanamycin. Cultures were grown at 37 °C with shaking to an optical density at 600 nm of 0.3-0.7, and protein expression was initiated with 1 mM isopropyl- β -D-1-thiogalactopyranoside. Following incubation at 16 °C for 16 h with shaking (250 rpm), cells were harvested by centrifugation, resuspended in 8 mL of lysis buffer (5

mM imidazole in 100 mM potassium phosphate buffer pH 8.0) supplemented with 1.25 g mL⁻¹ of lyophilized lysozyme and 1 U mL⁻¹ of benzonase nuclease (Merck Millipore), and lysed with an EmulsiFlex-B15 cell disruptor (Avestin). Proteins were purified by immobilized metal affinity chromatography using NiNTA agarose (Qia-gen) pre-equilibrated with lysis buffer in individual Econo-Pac gravity-flow columns (Bio-Rad). Columns were washed twice, first with 10 mM imidazole in 100 mM potassium phosphate buffer (pH 8.0), and then with the same buffer containing 20 mM imidazole. Bound proteins were eluted with 250 mM imidazole in 100 mM potassium phosphate buffer (pH 8.0) and exchanged into 100 mM sodium phosphate buffer (pH 7.0) supplemented with 100 mM sodium chloride using Econo-Pac 10DG desalting preppacked gravityflow columns (Bio-Rad). Proteins used for crystallization were further purified by gel filtration in 50 mM sodium citrate buffer (pH 5.5) and 150 mM sodium chloride using an ENrich SEC 650 size-exclusion chromatography column (Bio-Rad). Purified samples were concentrated using Amicon Ultracel-3K centrifugal filter units (EMD Millipore) and quantified by measuring the absorbance of the denatured protein in 6 M guanidinium chloride at 280 nm and applying Beer-Lambert's law using calculated extinction coefficients obtained from the ExPASy ProtParam tool (<https://web.expasy.org/protparam>).

2.6.7 Steady-State Kinetics

All assays were carried out at 27 °C in 100 mM sodium phosphate buffer (pH 7.0) supplemented with 100 mM sodium chloride. Triplicate 200 μ L reactions with varying concentrations of freshly prepared 5-nitrobenzisoxazole (Aablocks) dissolved in methanol (10% final concentration, pH of reaction mixture adjusted to 7.0 after addition of methanol-solubilized substrate) were initiated by the addition of 0.02-5.0 or 0.03-16.0 μ M for HG or KE70 variants, respectively. Product formation was monitored spectrophotometrically at 380 nm ($\epsilon = 15,800 \text{ M}^{-1} \text{ cm}^{-1}$) in individual

96-well plates (Greiner Bio-One) wells using a Biotek Synergy H1 multimode plate reader (Agilent). Path lengths for each well were calculated ratiometrically using the difference in absorbance of 100 mM sodium phosphate buffer (pH 7.0) supplemented with 100 mM sodium chloride and 10% methanol at 900 and 975 nm (27 °C). Linear phases of the kinetic traces were used to measure the initial reaction rates. For cases where saturation was not possible at the maximum substrate concentration tested (2 mM), data were fitted to the linear portion of the Michaelis-Menten model ($v_0 = (k_{\text{cat}}/K_M)[E_0][S]$), and k_{cat}/K_M was deduced from the slope. In all other cases, the data were fitted to the Michaelis-Menten equation to calculate individual k_{cat} and K_M parameters.

2.6.8 Crystallization

Enzyme variants were prepared in 50 mM sodium citrate buffer (pH 5.5) at the concentrations listed in Table 2.6. For samples that were cocrystallized with the TS analogue, a 100 mM stock solution of 6-nitrobenzotriazole (6NT, AstaTech) was prepared in dimethyl sulfoxide (DMSO) and diluted 20-fold in the enzyme solutions for a final concentration of 5 mM (5% DMSO). For each enzyme variant, we carried out initial crystallization trials in 15-well hanging-drop format using EasyXtal crystallization plates (Qiagen) and a crystallization screen that was designed to explore the chemical space around the crystallization conditions reported by Broom et al.[2] Crystallization drops were prepared by mixing 1 μL of protein solution with 1 μL of the mother liquor and sealing the drop inside a reservoir containing an additional 500 μL of the mother liquor solution. This procedure was successful for HG185, HG198, and HG630; however, crystallization of HG649 required the identification of a new crystallization condition using sparse matrix screening. We used commercial screening kits (NeXtal) to prepare crystallization drops in 96-well sitting drop plates by mixing 1 μL of the protein solution with 1 μL of the mother liquor and

sealing the drop inside a reservoir containing an additional 100 μL of the mother liquor solution. After identifying a suitable mother liquor solution from screening experiments, large crystals suitable for X-ray diffraction experiments were obtained in a 15-well hanging-drop format as described for the other HG variants. The specific growth conditions that yielded the crystals of each HG variant used for X-ray data collection are provided in Table 2.6. HG185, HG630, and HG649 crystallized in the absence of the 6NT TS analogue, and the resulting crystals were isomorphous with the HG3 crystal form reported in the PDB (PDB ID: 5RGA and 5RG4). For these variants, crystals with bound 6NT were obtained by soaking the unbound crystals in mother liquor containing 10 mM 6NT for at least 30 min prior to data collection. On the other hand, HG198 could not be crystallized without 6NT present, and the crystals that were obtained belonged to space group $P1$ with three polypeptides in the asymmetric unit. To prepare crystals of HG198 without 6NT bound, we serially soaked 6NT-bound crystals in three drops of crystallization mother liquor without 6NT. Each sequential soaking step was carried out for at least 30 min. After the removal of 6NT by soaking, the $P1$ crystals undergo a transition to space group $P2_1$, with a smaller unit cell and only one molecule per asymmetric unit. Although we could not determine the cause of this unusual change in crystal symmetry upon soaking, we observed that it occurred with a high reproducibility. We measured 10 individual crystals of 6NT-bound HG198, which all belonged to space group $P1$ and were roughly isomorphous. Additionally, we measured four individual crystals of HG198 without 6NT, and we found that all of them demonstrated the transition to monoclinic symmetry.

2.6.9 X-ray Data Collection and Processing

Prior to X-ray data collection, crystals were mounted in polyimide loops and sealed by using a MicroRT tubing kit (MiTeGen). Single-crystal X-ray diffraction

data were collected on beamline 8.3.1 at the Advanced Light Source. The beamline was equipped with a Pilatus3 S 6 M detector and was operated at a photon energy of 11,111 eV. Crystals were maintained at 280 K throughout the course of data collection. Each dataset was collected using a total X-ray dose of 50 kGy or less and covered a 180° wedge of reciprocal space. Multiple datasets were collected for each enzyme variant either from different crystals, or if their size permitted, from non-overlapping volumes of larger crystals, in order to identify optimally diffracting samples. X-ray data were processed with the Xia2 0.5.492 program (10.1107/S0021889809045701),[124] which performed indexing, integration, scaling, and merging with DIALS.[125, 126] For each dataset, a resolution cutoff was taken where the CC1/2 and $\langle I/\sigma I \rangle$ values for the merged intensities fell to approximately 0.5 and 1.0, respectively. Information regarding data collection and processing is presented in Table 2.7 and 2.8. The reduced diffraction data were analyzed with phenix.xtriage to check for crystal pathologies, and no complications were identified.

2.6.10 Structure Determination

We obtained initial phase information for calculation of electron density maps by molecular replacement using the program Phaser[127], as implemented in v1.13.2998 of the PHENIX suite.[128] Several different models of HG variants, including crystal structures and design models, were used for molecular replacement. For all datasets except that of HG630 (+) 6NT, we performed iterative steps of manual model rebuilding followed by refinement of atomic positions, atomic displacement parameters, and occupancies using a TLS model, a riding hydrogen model, and automatic weight optimization. For HG630 (+) 6NT, refinement was done as described earlier but without the usage of a TLS model. All model building was performed using Coot 0.8.9.236[129] and refinement steps were performed with phenix.refine within the PHENIX suite (v1.13-2998).[128],[130] Restraints for 6NT were generated us-

ing phenix.elbow,[131] starting from coordinates available in the Protein Data Bank (PDB ligand ID: 6NT). Further information regarding model building and refinement, as well as PDB accession codes for the final models, is presented in Tables 2.7 and 2.8.

Chapter 3

Directed Evolution Remodels Active Sites to Stabilize Multiple Transition States in a Family of Engineered Enzymes

Copyright: Behnoush Seifinoferest, Ngoc Thu Hang Pham, Edith Gollub, Asini Taka, Janani Pulivarthi, Roberto A. Chica, Anthony Green, Michael C. Thompson.

3.1 Statement of Contribution

M.C.T., R.A.C., and A.G. conceived the project. B.S., J.P., E.G., and A.T. purified and crystallized proteins. M.C.T. and B.S. designed and performed X-ray diffraction experiments. B.S. conducted refinements and computational analysis. B.S. and N.T.H.P. performed computational analysis, B.S., N.T.H.P., and M.C.T. wrote the manuscript.

3.2 Abstract

An enzyme’s ability to stabilize the transition state is a primary determinant of its catalytic efficiency. In multi-step reactions, this requirement becomes more complex, as the enzyme must accommodate multiple transition states while maintaining sufficient preorganization to minimize the entropic cost of catalysis. How enzymes achieve this balance between structural flexibility and preorganization remains an open question. Here, we leverage a set of engineered enzymes evolved from a common ancestor through three parallel directed evolution trajectories, each selected for a distinct catalytic function, to investigate how enzymes adapt to stabilize multiple transition states. Our results show that evolution reshapes the active site through a combination of structural rearrangements, including disruption and reformation of stabilizing interactions, changes in secondary structure, domain movements, and expansion of the active-site cavity. These changes collectively alter the conformational landscape, enriching substates that are favorable for catalysis. Computational analysis further reveals that evolution does not optimize the enzyme for perfect stabilization of a single transition state. Instead, it favors active-site geometries that can accommodate multiple transition states with moderate efficiency. Together, these findings provide insight into how evolution balances rigidity and flexibility to enhance catalytic efficiency in enzymes that operate along complex reaction coordinates.

3.3 Introduction

Enzymes, like every other catalyst, mainly function through stabilizing the transition state and lowering the activation energy [132]. The ways in which enzymes stabilize transition states have been a subject of intense investigation for decades [133, 134]. A variety of effects have been invoked, including stabilization of transition state charges and dipoles by electrostatic [135] and hydrogen-bonding interac-

tions [136], and entropy reduction mechanisms [23] whereby the binding affinity of the substrates effectively “pre-pays” the entropic cost of precisely orienting the reactants in the transition state geometry. Most studies directed toward understanding how enzymes stabilize transition states have focused on reactions that involve only a single step [137, 138], or on just a single step in a multi-step reaction [134, 139]. Considerably less work has been done to understand how a single enzyme active site can stabilize multiple transition states in reactions that involve multiple chemical steps [140, 141].

To address this knowledge gap, we exploited a family of engineered enzymes [142] that catalyze multi-step chemical reactions as a simplified and experimentally tractable model system. These artificial enzymes create a useful model system because the evolutionary history is known, making it possible to study variants with known evolutionary relationships and different catalytic efficiencies, providing a framework for understanding how changes in structure correlate with functional changes. Additionally, unlike natural enzymes, the laboratory evolutionary process [81] involves only a single selective pressure, which is the improvement of catalytic activity, so all the accumulated mutations optimize this one specific property. The selected enzyme family includes three distinct lineages, each developed by laboratory-directed evolution, from the computationally designed ancestor BH32 [90] (Figure 3.1A, left). BH32 was originally designed by Bjelic et al.(2013) to catalyze the Morita-Baylis-Hillman (MBH) reaction [90, 143, 144], an important carbon-carbon bond-forming reaction that allows functionalization of an α,β -unsaturated cyclic ketone at the C α position using an electrophilic carbon, such as an aldehyde (Figure 3.1A, top right). The BH32.n family [4] (Figure 3.1A, blue box) was evolved to enhance catalytic efficiency for the MBH reaction, and after fourteen rounds of directed evolution, the optimized variant BH32.14 exhibited a 27-fold increase in efficiency compared to the ancestor. The second family, BH_{MeHis}1.n [5] (Figure 3.1A, middle right), branches from the BH32.n lineage. Specifically, BH32.8 was modified to in-

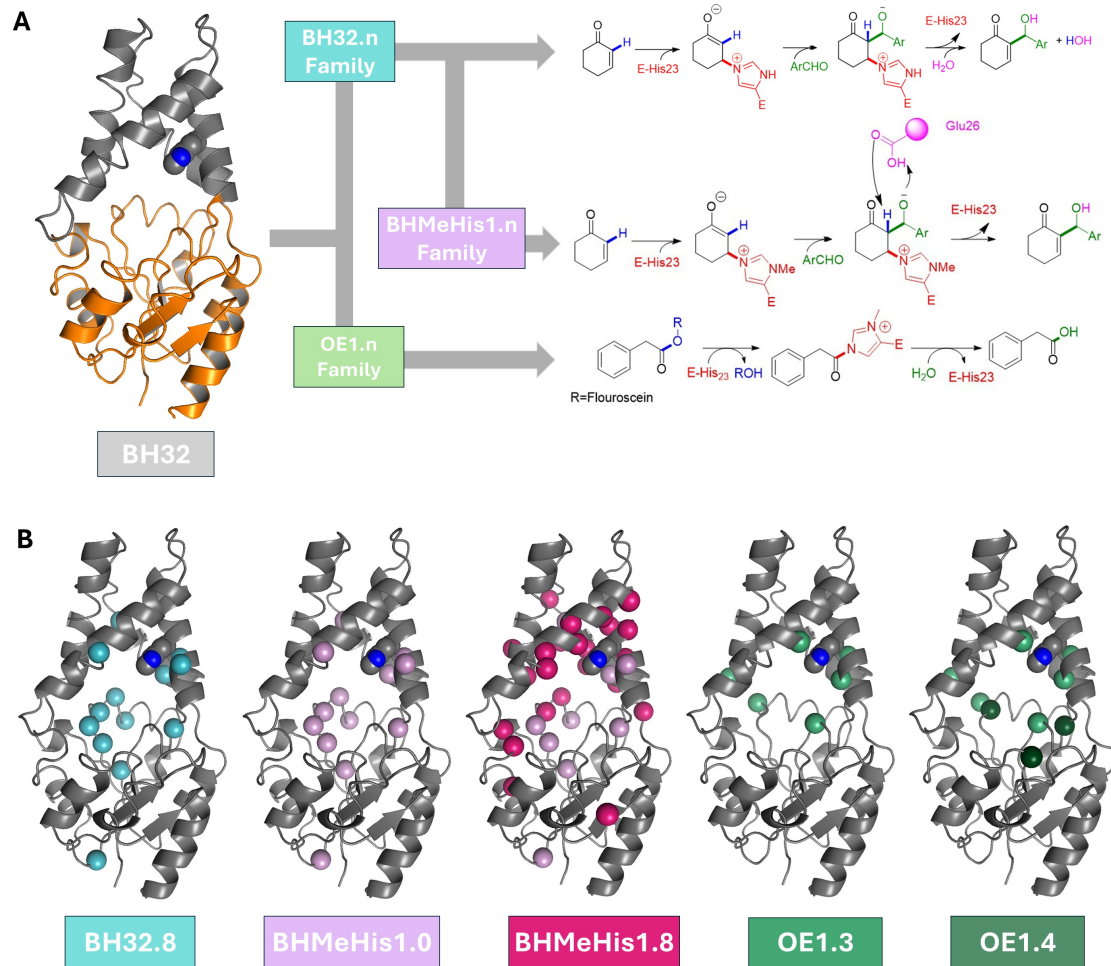


Figure 3.1: Three enzyme families evolved from a common ancestor through directed evolution, acquiring specificity for distinct chemical reactions. (A) The crystal structure of BH32 at room temperature, highlighting the cap domain (gray), which contains the catalytic residue His23 (shown as spheres), and the core domain (orange). The schematic to the right shows the different catalytic mechanisms associated with each family. (B) From each family (comprising a range of 4–14 members), two representative enzymes were selected for structural characterization using X-ray crystallography. Mutations present in each representative are mapped onto the BH32 structure. Color coding is used throughout this study: BH32 (grey), BH32.8 (teal blue), BH_{MeHis}1.0 (velvet), BH_{MeHis}1.8 (magenta), OE1.3 (emerald green), and OE1.4 (evergreen).

corporate an artificial amino acid, N δ -methylhistidine [145] in place of the catalytic residue Histidine23 using genetic code expansion [146, 147] to yield BH_{MeHis}1.0, the founding member of the second lineage. Eight rounds of directed evolution resulted in BH_{MeHis}1.8, which exhibited a 440-fold increase in catalytic efficiency for the MBH reaction compared to BH_{MeHis}1.0.

A separate evolutionary trajectory, also incorporating the His23MeHis substitution, was pursued to develop catalytic activity for an ester hydrolysis reaction, creating the OE1.n family [6] (Figure 3.1A, bottom right). Five rounds of directed evolution resulted in OE1.4, which demonstrated a 200-fold increase in efficiency compared to OE1. The ester hydrolysis reaction proceeds through a nucleophilic attack by the N ϵ of MeHis23, forming an acyl-enzyme intermediate that is subsequently resolved by a water molecule. Notably, the substitution of His23 with MeHis23 prevents the formation of a stable intermediate due to deprotonation of the imidazole ring in the acyl-enzyme state [6].

These three lineages, evolved from a common ancestor, share a conserved protein fold and a common catalytic residue (Figure 3.1 and Table 3.1), while catalyzing distinct chemical reactions or similar reactions with distinct mechanisms. This makes them an ideal system for studying how structural and dynamic features emerge to enable reaction specificity and improve catalytic efficiency.

Here, we show, using room-temperature X-ray crystallography [55] and computational modeling, that for enzymes that catalyze the MBH reaction, directed evolution serves to widen and reshape the active site so that all four transition states can be stabilized within the same active site. Specifically, we selected two representative enzymes from each evolutionary lineage with differing (low and high) activities (Figure 3.1B). From the BH32.n family, we selected BH32 and BH32.8, which is also the branching point for the BH_{MeHis}1.n family. From this second lineage, we studied BH_{MeHis}1.0, which differs from BH32.8 only by a single chemical modification (His23MeHis), as well as BH_{MeHis}1.8, the most evolved enzyme that catalyzes the

Table 3.1: Sequences and mutations of enzyme variants.

Protein	Sequence	Mutations (from BH32)
BH32	MIRAVFFDSLGLTNSVEGAAKSHLKIMEEVLGDYPLNPKTL LDEYEKLREAFSNIYAGKPYRPLRDILEEVMRKLAKEYGFK YPENFWEISLRMSQRYGELYPEVEVLKSLKGKYWGMI TDSDEQAMAFDLALGKDLFDSITTSEEAGFFKPHPRIFE LALKAGVKGEAEVYVGDNPVKDCGGSKNLGMTSLLDDR KGERERFWDKCDFIVSDLREVIKIVDELNGQGSLEHHHH H	–
BH32.8	MIRAVFFDSLGLTNSVEGAYKVHLKIMEEVLGDYPLNPKTLL DEYEKLREAFSNIYAGKPYRPLRDILEEVMRKLAKEYGFKN YPDNLWEISLRMAGRYGELYPEVEVLKSLKGKYWGMI NENFATATDLALGIKDLFDSITTSEEAGFFKPHYPRIE LKKAGVGKEKAVYVGDNPVKDAGGSKNLGMTSILLDRKG EKREFWDKADFIVSDLREVIKINGQGSLEHHHHHH	N14I, A20Y, S22V, T49A, F87L, S95A, M120V, T122L, D123N, S124R, Q128P, M130T, E174K
BH _{MeHis} 1.0	MIRAVFFDSLGLTISVEGAYKV(MHIS)LKIMEEVLGDYPLNP KTLLDEYEKLREAFSNIYAGKPYRPLRDILEEVMRKLAEK YGFKYPENLWEISLRMAQRYGELYPEVEVLKSLKGKYW GIVLNRDTEPATAFLDALGIKDLFDSITTSEEAGGFFKPHPRI FELALKKAGVKGEKAVYVGDNPVKDAGGSKNLGMTSILLD RKGEREFWDKADFIVSDLREVIKIVDELNGQGSLEHHHH HH	<i>His23MeHis</i> , N14I, A20Y, S22V, T49A, F87L, S95A, M120V, T122L, D123N, S124R, Q128P, M130T, E174K
BH _{MeHis} 1.8	MIRAVFFDSLGLTISVEGAYKI(MHIS)VKEMEEVLGDYPLNP KTLTWDEDKLFKFEAFSNIYAGKPYRPARFILEEVMRKLKGY FGFKYPGNLQEIGVRMARRYGGLYPEVEVLKSLKGKYW GVILNWDTENATAFLDALGIKDLFDSITTSEEAGFAYPHPRI RILELALKKAGVKGEKAVYVGDNPVKDAGGSKNLGMTSI LLDRKGERREFWDKADFVRSDDLREVIKIVDELNGQGSLEH HHHH	<i>His23MeHis</i> , N14I, A20Y, V22I, L24V, I26E, K39R, L42W, Y45F, E46D, A49F, R50K, L64A, D66F, E85G, F87L, W88Q, S91G, L92V, S95A, Q96R, E100G, M120V, T122L, D123N, R124W, P128N, A129S, M130T, F154A, K155V, F161L, E174K, I215R
OE1.3	MIRAVFFDSPGTLNSVEGHAKM(MHIS)LKIMEEVLGDYPLN PKTLLDEYENKLTREAFSNIYAGKPYRGLRDIEEVMRKLA KYGFKYPENFWEISLRMSQRYGELYPEVEVLKSLKGKY HVGMITDSGTEQAMAFDLALGIKDLFDSITTSEEAGFFKPH PRIFELALKKAGVKGEAEVYVGDNPVKDCGGSKNLGMTSI LLDRKGERERFWDKCDFIVSDLREVIKIVDELNGQGSLEH HHHH	<i>His23MeHis</i> , L10P, A19H, S22M, E46N, P63G, C125G
OE1.4	MIRAVFFDSPGTLQSVEGHAKM(MHIS)LKIMEEVLGDYPLN PKTLLDEYNLKTREAFSNIYAGKPYRGLRDIEEVMRKLA KYGFKYPENFWEISLRMSQRYGELYPEVEVLKSLKGKY HVGMITDLGTEQAMAFDLALGIKDLFDSITTSEEAGFFKPH PRIFELALKKAGVKGEAEVYVGFNPVKDCGGSKNLGMTSI LLDRKGERERFWDKCDFIVSDLREVIKIVDELNGQGSLEH HHHH	<i>His23MeHis</i> , L10P, N14Q, A19H, S22M, E48N, P63G, S124L, C125G, D180F

MBH reaction. Finally, from the OE1.n family, we selected OE1.4, the most evolved variant, and OE1.3.

Our results suggest that evolution introduces structural changes, including disruption and reformation of stabilizing interactions, widening of the active site cavity through shifts in helix positioning, hinge-bending motions that increase the angle between the cap and core domains, and localized disorder and conformational rearrangements. Together, these changes reshape the active site. Computational theozyme placement of DFT-derived transition states for the MBH reaction catalyzed by BH32.14 shows that these structural adaptations improve the complementarity of BH32.8 with all four transition states. In contrast, BH32 effectively stabilizes the first transition state but shows substantially weaker complementarity for the remaining three. The ability to better accommodate multiple transition states is therefore directly associated with the increased catalytic efficiency observed for BH32.8 relative to BH32, emphasizing the importance of stabilizing multiple transition states in multi-step reactions.

3.4 Results

3.4.1 Structure Determination Using X-ray Crystallography

To initiate our analysis of structural changes induced by directed evolution, we determined high-resolution, room temperature (300 K) crystal structures of BH32, BH32.8, BH_{MeHis}1.0, BH_{MeHis}1.8, OE1.3, and OE1.4, (crystallization condition are available in Table 3.2) to understand which features are tuned to improve catalytic efficiency. We also determined structures at cryogenic temperature (100 K), but since cryocooling is known to reduce the volumes of protein molecules [148] and distort their conformational ensembles [77], our analysis focused on the room temperature structures. All enzymes crystallized in similar unit cells corresponding to space group

Table 3.2: Crystallization conditions for all enzyme variants. *seeding used.

Protein	Temp.	Buffer	pH	Protein	Precipitant #1	Precipitant #2
	(K)			(mg mL ⁻¹)		
BH32	100	Tris	8.5	22.92	26% w/v PEG 4000	0.1 M Sodium Acetate
	300	Tris	8.0	10	26% w/v PEG 4000	0.1 M Sodium Acetate
		Tris	8.5	10	30% w/v PEG 4000	0.2 M Sodium Acetate
BH32.8	100	–	–	20	20% w/v PEG 3350	0.1 M Sodium Citrate
	300	–	–	20	18% w/v PEG 3350	0.1 M Sodium Citrate
BH _{MeHis} 1.0	100	–	–	22	18% w/v PEG 3350	0.1 M Sodium Citrate
	300	–	–	22	18% w/v PEG 3350	0.1 M Sodium Citrate
BH _{MeHis} 1.8	100	–	–	15	18% w/v PEG 4000	18% w/v Isopropanol
	300	–	–	15	14% w/v PEG 4000	20% w/v Isopropanol
OE1.3	100	BisTris	5.5	5.1	18% w/v PEG 3350	0.2 M Ammonium Acetate
	300	HEPES	7.5	5.1	16% w/v PEG 3350	0.2 M Sodium Nitrate
OE1.4	100	BisTris	5.5	8.8	16% w/v PEG 3350	0.2 M Ammonium Acetate
	300	BisTris	5.5	8.8	16% w/v PEG 3350	0.2 M Ammonium Acetate

$P2_1$, with the exception of BH_{MeHis}1.8, which crystallized in the higher symmetry space group $P2_12_12_1$. The resulting datasets were collected at resolutions ranging from 1.55 to 2.10 and refined to optimal R-factors (Tables 3.3 and 3.4).

Table 3.3: Crystallographic data and refinement statistics for BH32, BH32.8, and BH_{MeHis}1.0.

	BH32		BH32.8		BH _{MeHis} 1.0	
Temp. (K)	100	300	100	300	100	300
Data Collection						
Resolution (Å)	1.85	1.95	1.90	1.95	2.10	2.00
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Unit cell						
<i>a</i> (Å)	34.39	34.49	36.52	36.95	36.53	36.96
<i>b</i> (Å)	69.85	70.77	69.23	71.13	68.78	70.56
<i>c</i> (Å)	52.60	54.01	45.48	46.68	45.57	46.33
β (°)	106.4	103.5	99.56	98.89	99.41	99.15
Chains / asym. unit	1	1	1	1	1	1
Completeness (%)	97.7	94.3	98.4	99.9	99.9	98.3
Multiplicity	6.6	3.6	6.8	3.4	3.3	3.4
I/ σ I	12.4 (0.8)	12.2 (1.2)	7.7 (0.3)	12.3 (1.0)	9.9 (1.6)	7.9 (0.8)
R _{pim}	0.039 (0.121)	0.044 (0.702)	0.050 (0.505)	0.041 (0.660)	0.068 (0.512)	0.064 (0.642)
CC _{1/2}	0.998 (0.955)	0.995 (0.491)	0.997 (0.632)	0.998 (0.471)	0.995 (0.604)	0.996 (0.471)
Wilson B-factor (Å ²)	24.96	35.17	32.18	35.54	28.20	33.96
Unique reflections	19921	17468	17314	17393	12943	15679
Refinement						
R _{work} /R _{free}	0.19/0.23	0.14/0.19	0.19/0.24	0.16/0.21	0.19/0.25	0.17/0.24
No. atoms						
Protein	4158	4384	3719	4396	3634	3986
Ligand	34	1	11	40	39	29
Water	91	51	31	45	40	39
RMSD						
Bond lengths (Å)	0.012	0.012	0.015	0.013	0.013	0.014
Bond angles (°)	1.160	1.093	1.295	1.175	1.188	1.166
MolProbity statistics						
Ramachandran favored (%)	97.37	97.37	97.63	96.82	97.54	98.62
Rotamer outliers (%)	4.44	3.81	7.54	8.55	5.15	7.08
Clashscore	5.70	6.81	10.16	11.03	5.71	6.71

Values in parentheses correspond to the highest-resolution shell.

Table 3.4: Crystallographic data and refinement statistics for BH_{MeHis}1.8, OE1.3, and OE1.4.

	BH _{MeHis} 1.8		OE1.3		OE1.4	
Temp. (K)	100	300	100	300	100	300
Data Collection						
Resolution (Å)	1.87	1.49	1.60	1.80	1.60	1.90
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Unit cell						
<i>a</i> (Å)	32.56	33.23	34.23	34.40	33.98	34.45
<i>b</i> (Å)	46.93	47.67	72.46	71.73	72.16	72.09
<i>c</i> (Å)	146.72	150.01	52.48	52.68	52.11	52.72
β (°)	90	90	105.57	105.19	105.11	105.24
Chains / asym. unit	1	1	1	1	1	1
Completeness (%)	99.1	96.4	99.7	99.9	99.8	99.7
Multiplicity	12.5	3.4	6.5	3.4	6.3	3.4
I/ σ I	36.9 (7.3)	12.5 (0.9)	16.2 (0.9)	12.4 (1.0)	14.6 (0.5)	9.4 (0.9)
R _{pim}	0.027 (0.062)	0.038 (0.972)	0.021 (0.293)	0.039 (0.635)	0.031 (0.187)	0.051 (0.737)
CC _{1/2}	0.995 (0.984)	0.997 (0.340)	1.000 (0.864)	0.998 (0.573)	0.998 (0.909)	0.995 (0.518)
Wilson B-factor (Å ²)	19.20	22.23	20.69	29.86	23.29	33.25
Unique reflections	19293	38517	32489	22907	31995	193636
Refinement						
R _{work} /R _{free}	0.15/0.19	0.16/0.18	0.17/0.21	0.16/0.19	0.18/0.21	0.16/0.20
No. atoms						
Protein	4146	4388	4602	4376	4508	4138
Ligand	120	64	92	80	110	48
Water	702	360	690	250	665	205
RMSD						
Bond lengths (Å)	0.010	0.011	0.014	0.012	0.005	0.012
Bond angles (°)	1.061	1.038	1.232	1.071	0.622	1.060
MolProbity statistics						
Ramachandran favored (%)	98.10	97.54	99.02	99.87	96.90	98.45
Rotamer outliers (%)	2.41	3.62	3.48	3.05	2.89	2.31
Clashscore	6.10	7.25	5.90	7.80	6.45	7.10

Values in parentheses correspond to the highest-resolution shell.

3.4.2 Evolved variants have unique active site cavity shapes

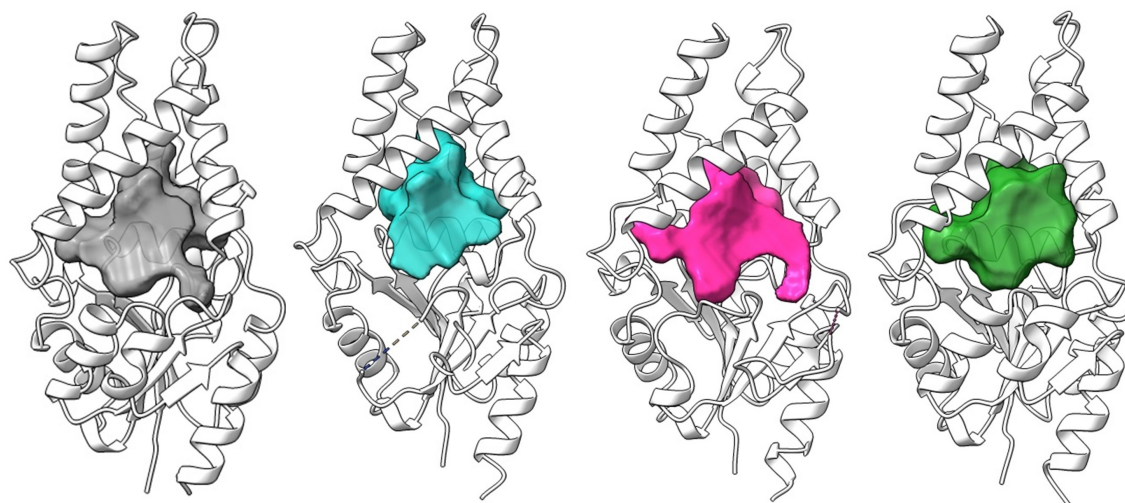
Active-site cavity calculations using KVFinder [149, 150] show that evolution reshapes the active site in the evolved variants relative to BH32. As shown in Figure 3.2A, each enzyme family exhibits distinct pocket shapes and topologies. Consistent with this, the mutations accumulated during evolution, many of which are located around the active site (Figure 3.1B), appear to drive these changes in the active-site shape.

A particularly informative comparison is between BH32.8 and BH_{MeHis}1.8, both of which catalyze the Morita–Baylis–Hillman (MBH) reaction on the same substrate, yet display markedly different binding pocket architectures. Analysis of active-site volume and surface area (Figure 3.2, bottom) indicates that these quantities do not differ substantially; however, each family has remodeled the overall pocket shape to better accommodate its specific catalytic requirements.

3.4.3 Global Rearrangements Indirectly Reshape the Active Sites of High Activity Variants

As an initial step toward comparative structural analysis, we performed pairwise alignments of structures. To quantify structural divergence across the evolved enzymes, we calculate pairwise RMSDs between their structures (Figure 3.3A), providing a global overview of structural differences across all variants. This analysis revealed that the OE1.n family members (OE1.3 and OE1.4) retained the highest similarity to BH32, suggesting minimal deviation from the ancestral fold. In contrast, BH_{MeHis}1.8 exhibited the greatest divergence between the highly evolved variants, indicative of substantial structural adaptation. Notably, variants BH32.8 (branch point of BH_{MeHis}1.n family) and BH_{MeHis}1.0 also showed moderate divergence.

Although global RMSD comparisons are useful, they do not capture more subtle, localized differences in structural behavior. To address this, we employed all-residue



Protein	Active site Volume ($\text{\AA}^3$)	Active site Area ($\text{\AA}^2$)
BH32	1,295.14	852.97
BH32.8	1,119.74	679.65
BHMeHis1.8	1,165.75	708.12
OE1.4	1,304.21	804.85

Figure 3.2: KVFinder analysis reveals how evolution reshapes active-site cavity geometry. **Top)** Active-site surface representations calculated for each enzyme variant illustrate distinct cavity shapes arising from evolutionary changes. **Bottom)** The corresponding volume and surface area measurements show that, despite similar overall sizes, each enzyme family adopts a unique pocket topology tailored to its catalytic function.

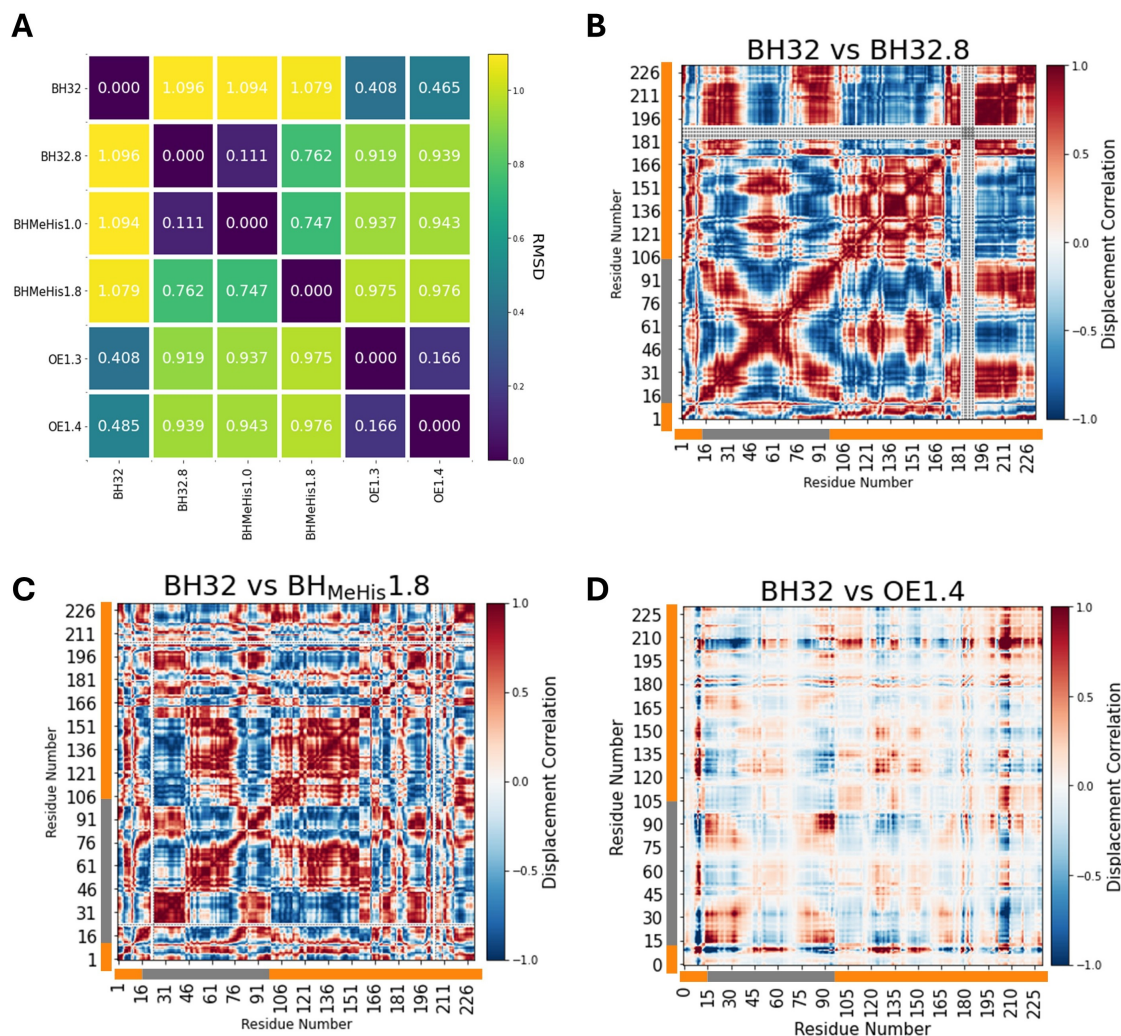


Figure 3.3: Structural comparison of all variants using RMSD and correlation analysis. (A) Pairwise global RMSD values from structural alignment show that BH32.8, BH_{MeHis}1.0, and BH_{MeHis}1.8 exhibit the greatest divergence from the ancestral BH32, whereas OE1.3 and OE1.4 remain more structurally similar. (B–D) All-residue correlation maps comparing BH32 with BH32.8 (B), BH_{MeHis}1.8 (C), and OE1.4 (D). Each map is generated by calculating dot products between displacement vectors of equivalent C α atoms, where blue indicates correlated motion, red indicates opposing motion, and white represents orthogonal or minimal displacement. Black dots denote disordered regions. The bars along the axes indicate domain organization, with the cap domain shown in grey and the core domain in orange.

C α displacement correlation maps to analyze pairwise shifts of residues in the high-efficiency variants relative to their counterparts in BH32. These maps serve as structural “fingerprints,” enabling direct comparison of collective motions across enzyme variants.

Correlation maps were generated by defining displacement vectors for each C α atom between aligned structures and calculating dot products between all pairs of vectors. Positive dot products (blue) indicate correlated changes in atomic position in similar directions, negative values (red), in opposite directions, and values near zero (white) reflect either orthogonal shifts or minimal displacement. Disordered regions are indicated by black dots.

Figures 3.3B–D show the correlation maps comparing BH32 with the three efficient variants from each family: BH32.8, BH_{MeHis}1.8, and OE1.4. These maps reveal that the largest global deviations from BH32 occur in BH32.8. While BH_{MeHis}1.8 retains many of the structural changes it has inherited from BH32.8, the correlations indicate less pronounced shifts in this variant compared to BH32.8. In contrast, OE1.4 exhibits the least global deviation from BH32, suggesting less dramatic propagation of structural changes.

Importantly, while mutations within the active site directly affect local geometry in a predictable manner, the correlation maps highlight the broader, indirect effects of distal mutations on global protein motions. The observed patterns suggest that many of these global movements originate in BH32.8 and are subsequently retained with lower intensity in the BH_{MeHis} lineage, while remaining largely absent or attenuated in the OE family.

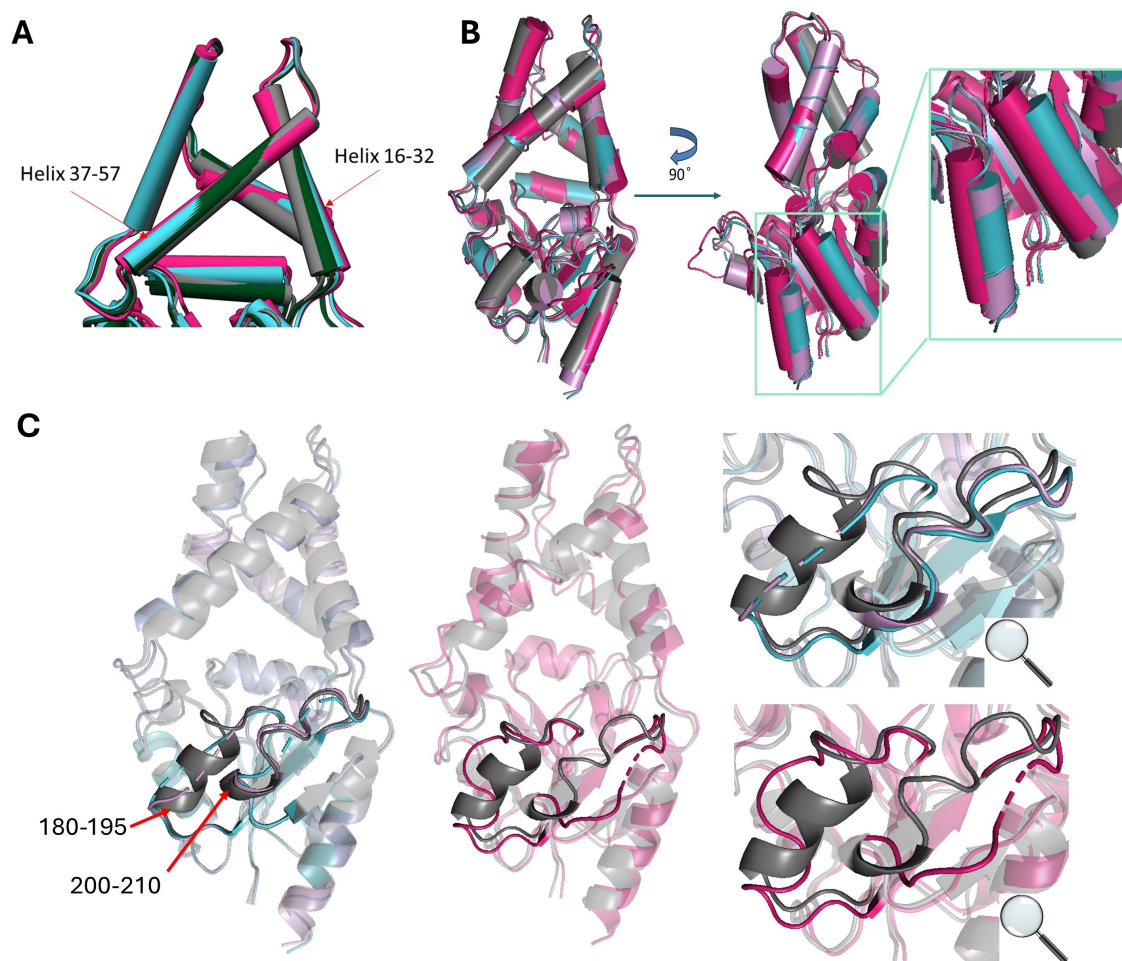


Figure 3.4: Three categories of structural changes observed in evolved enzymes relative to BH32. BH32 (grey), BH32.8 (teal), OE1.4 (forest green), BH_{MeHis}1.8 (magenta), and BH_{MeHis}1.0 (violet). (A) Evolved variants exhibit widening of the cap domain, with BH_{MeHis}1.8 showing the most pronounced expansion. (B) Intermediate variants such as BH32.8 and BH_{MeHis}1.0 display a hinge-like opening between the cap and core domains, which partially reverts in BH_{MeHis}1.8, restoring a more ancestral geometry. (C) Helices spanning residues 180–195 and 200–210 become disordered in BH32.8 and BH_{MeHis}1.0, and adopt new conformations in BH_{MeHis}1.8.

3.4.4 Structural Adaptations Underlying Changes in Active Site Shape

A broad comparison of the evolved enzyme structures with the ancestral BH32 highlights three major structural adaptations present in high activity variants (Figure 3.4). First, nearly all evolved variants exhibit a widening of the cap domain, primarily due to an outward shift of the helix containing the catalytic residue (residues 16–32) relative to the adjacent helix (residues 37–57) (Figure 3.4A).

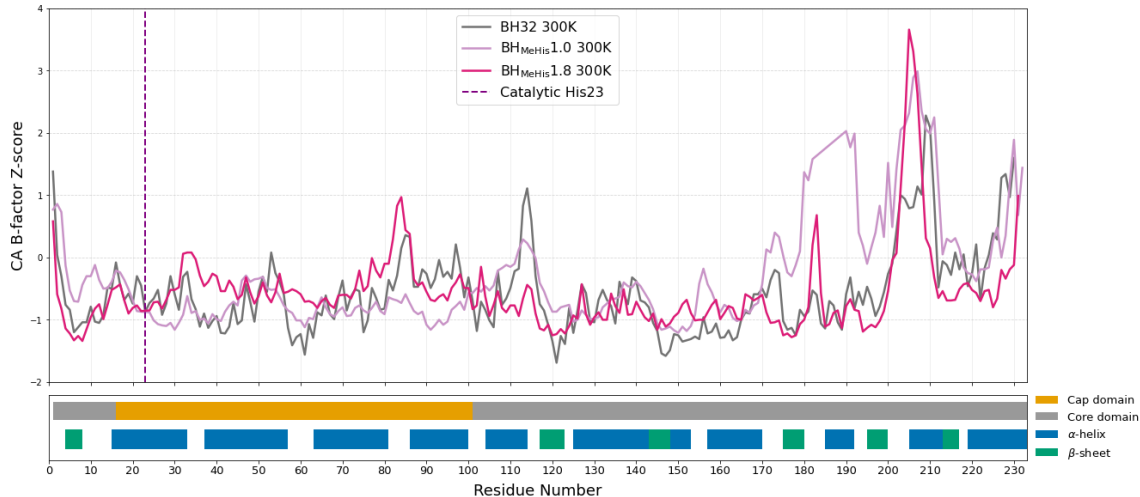


Figure 3.5: Comparison of C α B-factor Z-score profiles for BH32, BH_{MeHis}1.0, and BH_{MeHis}1.8 (top), with places of the cap and core domains and the secondary structures (from BH32) on the bottom. The vertical dashed purple line indicates the position of the catalytic (Me)His23. To avoid artificial reduction of B-factors arising from multi-conformer modeling, all values were calculated from single-conformer refinements.

Second, BH32.8 and BH_{MeHis}1.0 display a hinge-like opening between the cap and core domains, with increased inter-domain angle (Figure 3.4B). This hinge movement appears to be partially or fully reversed in the more evolved members, with

BH_{MeHis}1.8 retaining a mild opening. Hints of this hinge-bending movement, specifically in BH32.8 compared to BH32, can be observed in the C α displacement correlation maps previously discussed (Figure 3.3B), where the cap and the core domain are shifted away from one another.

Third, the helices formed by residues 180–195 and by residues 200–210 show increased crystallographic disorder in BH32.8 and BH_{MeHis}1.0, suggesting they become more flexible in these variants, and later adopt distinct new conformations in BH_{MeHis}1.8 as evolution progresses (Figure 3.4C, 3.4D, and 3.5).

The observation that highly active variants have increased flexibility is supported by thermal unfolding experiments performed using circular dichroism spectroscopy (Table 3.5), which show that in BH32.n and BH_{MeHis}1.n lineages, thermodynamic stability is anti-correlated with activity.

Table 3.5: Melting temperatures for all enzyme variants.

Enzyme	Melting Temperature (°C)
BH32	110.00
BH32.8	76.11
BH32.12	67.01
BH _{MeHis} 1.0	77.48
BH _{MeHis} 1.8	68.25
OE1.3	110.00
OE1.4	110.01

Together, these three categories of structural adaptations act to reshape the active sites of enzymes that efficiently catalyze the MBH reaction. In contrast, highly active members of the OE1 lineage retain structural features most similar to BH32, showing no/little significant changes in these regions.

3.4.5 Introduction of Bulky Residues Drives Cap Domain Widening

As shown in Figure 3.4A, all evolved enzymes exhibit a widening of the cap domain. Specifically, the helix formed by residues 37–57, positioned on the opposite side of the active site cavity, which contains the catalytic residue, has shifted outward. Angle measurements between the helix axis of these two helices using the UCSF Chimera [151] axis define tool confirm this observation: both intermediate and highly evolved enzymes display a wider cap domain compared to BH32, with BH_{MeHis}1.8 showing the largest angle and hence the widest cap domain (Figures 3.6A). In contrast, OE family members exhibit the smallest changes in cap domain geometry relative to BH32.

To investigate the structural basis for this widening, we analyzed BH32.8 as a representative case. To mention some examples, as shown in Figure 3.6B, the A20Y mutation appears to contribute to the separation between helices formed by residues 16–32 and residues 37–57. In BH32, the smaller Ala20 allows Lys39 to pack tightly in front of it on the catalytic helix (Figure 3.6B, left). Upon mutation to Tyr, however, the bulky side chain of Tyr20 causes a steric clash with the original rotamer of Lys39, forcing the helices apart and resulting in a new conformation for Lys39 (Figure 3.6B, right). This serves as a clear example of how the introduction of a bulkier residue can contribute to the opening of the cap domain through local structural rearrangement.

Similar cases can be spotted in other enzymes as well. In BH_{MeHis}1.8, the L42W mutation, which was also highlighted in the study by Hutton et al.[145] for its role in positioning and improving the packing of MeHis23, appears to influence the positioning of the helix formed by residues 16–32 as well (Figure 3.6D). In addition to the outward movement of helices formed by residues 16–32 and residues 37–57, we also observed a shift in helix formed by residues 86–99, which contributes further to cap domain widening (Figure 3.6E). This movement is particularly pronounced in

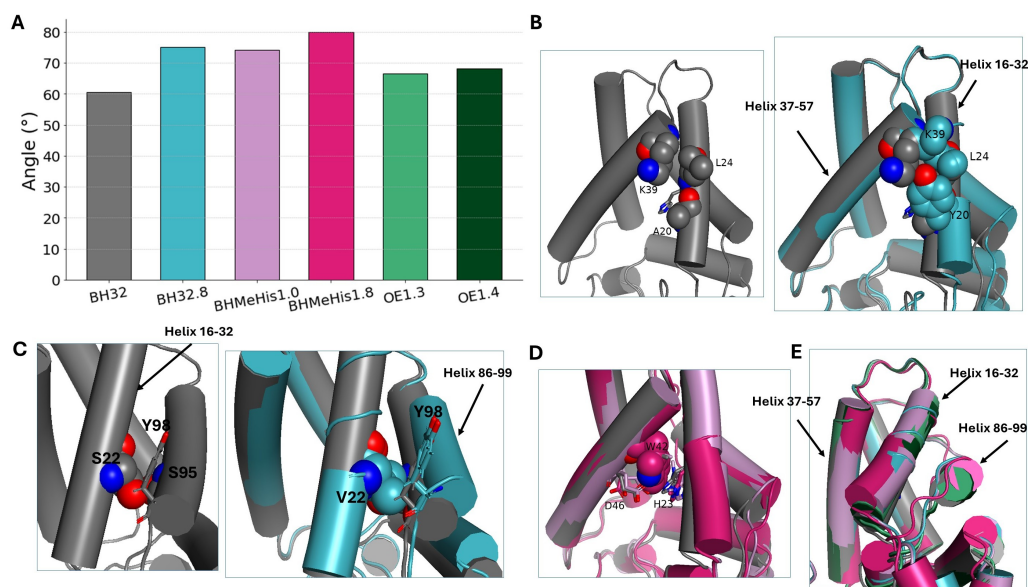


Figure 3.6: Examples of the structural features underlying cap domain widening in evolved enzymes. A) The angle between the helices formed by residues 16–32 and that formed by residues 37–57 in the cap domain increases in the evolved enzymes, with BH_{MeHis}1.8 showing the largest change. B) The A20Y mutation between BH32 and BH32.8 illustrates how the introduction of bulkier side chains can lead to helix separation. The substitution of Ala20 with Tyr appears to push the helix formed by residues 16–32 outward to avoid steric clashes with pre-existing packing interactions. C) The S22V mutation in BH32.8 introduces a longer side chain (Val22), which likely clashes with Tyr98 on the helix formed by residues 86–99. This clash may drive the outward shift of this helix, providing a clear example of how packing disruptions can lead to structural remodeling. D) The L42W mutation between BH_{MeHis}1.0 and BH_{MeHis}1.8 helps stabilize the positioning of MeHis23, which in turn supports the repositioning of the helix formed by residues 16–32. E) The helix formed by residues 86–99 also shifts outward in the evolved enzymes, further contributing to cap domain widening.

BH32.8, BH_{MeHis}1.0, and BH_{MeHis}1.8, but is also present, to a lesser extent, in OE1.3 and OE1.4. In an example from BH32.8, we identified the S22V mutation as one of the likely drivers of this shift (Figure 3.6C). The longer side chain of Val22, positioned between the helix formed by residues 16–32 and that formed by residues 86–99, would clash with Tyr98 if the helix formed by residues 86–99 remained in its original location. To avoid this steric conflict, the helix formed by residues 86–99 appears to shift outward. This provides another clear example of how evolutionary mutations can reshape the cap domain architecture through local packing constraints.

3.4.6 Altered Electrostatics and H-bonding Lead to Helix Disorder in BH32.8 and BH_{MeHis}1.0

In addition to cap domain widening and hinge opening, another key structural adaptation observed in the evolved enzymes is the disordering and subsequent refolding of two helices adjacent to the active site, spanning residues 180–195 and 200–210. In both BH32.8 and BH_{MeHis}1.0, the 180–195 region becomes unresolved in the electron density map, indicating increased local disorder in this segment. At the same time, the 200–210 region remains visible but exhibits elevated flexibility, as reflected by increased C α B-factor Z-scores in these regions (Figure 3.5).

At the molecular level, this destabilization can be traced to the D123N mutation between BH32 and BH32.8. In the ancestral BH32 structure, the side-chain carboxylate of Asp123, located on the loop above the 180–195 helix, forms a stabilizing salt bridge with the side chain of Lys184 on this helix (Figure 3.7A). This interaction serves as a structural anchor that helps maintain the helical conformation. Substitution of Asp123 with Asn removes the negative charge required for this interaction, thereby disrupting the salt bridge and destabilizing the helix. Consistent with this loss of stabilization, Lys184 no longer shows well-defined electron density in BH32.8 (and BH_{MeHis}1.0), reflecting increased conformational disorder in this region.

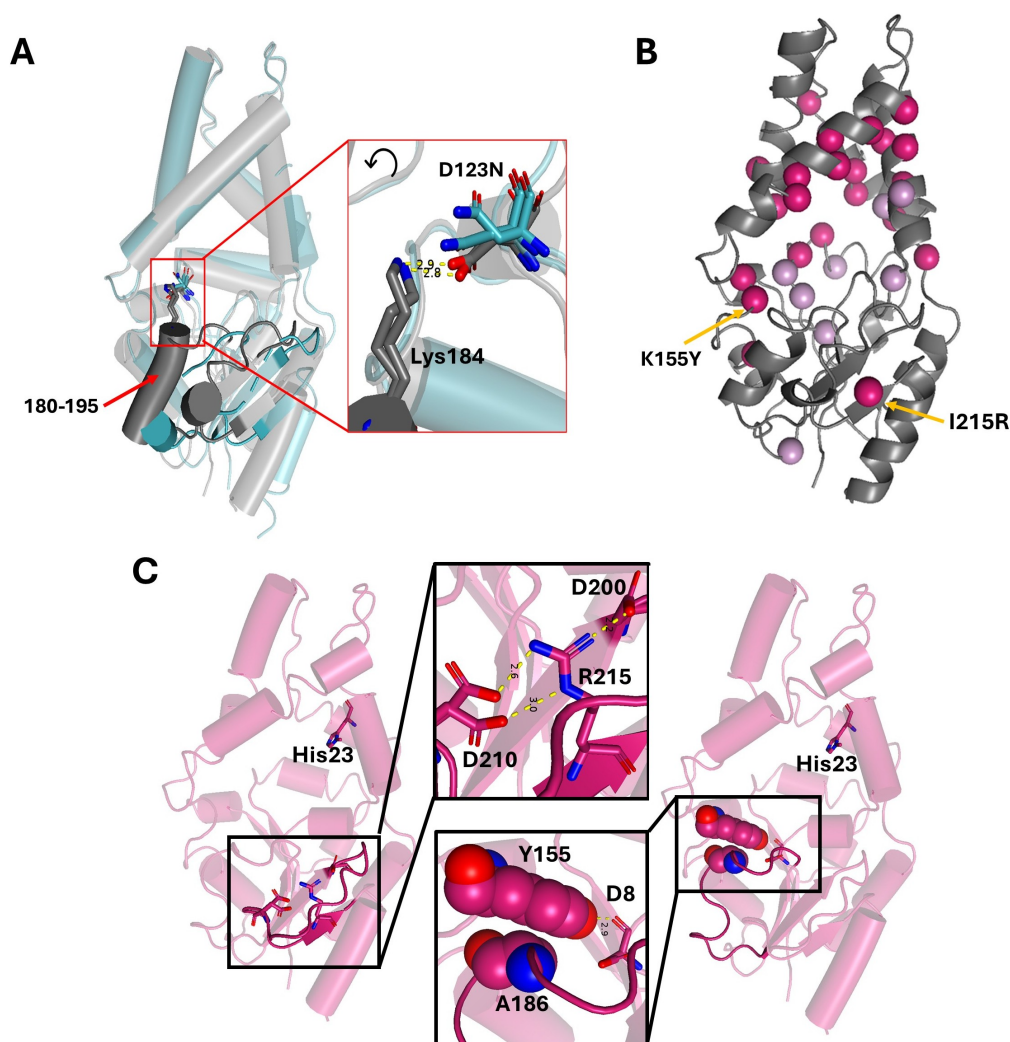


Figure 3.7: Disorder and refolding of helices adjacent to the active site across evolution. (A) Structural overlay of BH32 (gray) and BH32.8 (teal) highlights the D123N mutation (inset), which disrupts a salt bridge between Asp123 and Lys184 present in BH32, leading to destabilization of the helix spanning residues 180–195. (B) Distribution of mutations in the BH_{MeHis} lineage mapped onto the BH32 backbone (shown as C α spheres), with BH_{MeHis} 1.0 mutations in light purple and BH_{MeHis} 1.8 mutations in magenta. The K155Y and I215R mutations are highlighted as key contributors to structural stabilization. (C) Stabilizing interactions in BH_{MeHis} 1.8 following refolding: Arg215 anchors the 200–210 loop via hydrogen bonds with Asp200 and Asp210 (top), while Tyr155 forms hydrophobic interactions with Ala186, reducing flexibility in the 180–195 region (bottom).

Remarkably, this locally disordered state observed in BH32.8 and BH_{MeHis}1.0 is resolved in the more evolved variant BH_{MeHis}1.8. In this enzyme, both regions (180–195 and 200–214) become reordered into structured loops, but with elevated flexibility relative to the rest of the protein (Figure 3.7C and Figure 3.4C, right). This structural reorganization appears to be stabilized by two key mutations. First, the I215R mutation, located on the β -sheet spanning residues 214–216 directly beneath the 200–210 region, introduces an arginine residue that forms hydrogen bonds with Asp200 and Asp210, effectively anchoring the loop at both ends (Figure 3.7C, left). Second, the K155Y mutation, located on a loop positioned above the 180–195 region, introduces an aromatic tyrosine side chain that engages in hydrophobic interactions with Ala186 within the disordered segment (Figure 3.7C, right).

Together, these interactions provide compensatory stabilization that restores structural order in these regions. This is reflected in the Bfactor Zscore profiles for BH_{MeHis}1.8 (Figure 3.5), which show reduced flexibility compared to earlier variants in residues 180–195 and in residues flanking the 200–210 loop.

3.4.7 Active Site Reshaping Allows Multiple Transition States to Be Stabilized Efficiently

To better understand how changes in active-site shape relate to catalytic function, we used Density Functional Theory [152]-derived models from Crawshaw et al. [4], representing the four transition states of the Morita–Baylis–Hillman (MBH) reaction catalyzed by the BH32.14 (Figure 3.8A). These transition state structures were used to evaluate pose compatibility within the backbone conformational ensembles of the ancestral BH32 and the three high-efficiency variants.

These calculations were performed using ensemble-refined models derived from room-temperature X-ray crystallography data (Table 3.6), allowing us to explicitly account for conformational heterogeneity. For each enzyme, every transition state

Table 3.6: Ensemble refinement results for the room temperature structures used in theozyme placement

Structure	Resolution (Å)	$R_{\text{work}}/R_{\text{free}}$ before refinement	$R_{\text{work}}/R_{\text{free}}$ after refinement	Ensemble size	pTLS	τ_x	W_{xray}
BH32	1.95	0.1400/0.1883	0.1407/0.1865	50	0.25	0.50	2.5
BH32.8	1.95	0.1659/0.2244	0.1523/0.1991	50	0.25	0.50	5.0
BH_{MeHis}1.8	1.55	0.1471/0.1740	0.1339/0.1635	100	0.25	0.50	7.5
OE1.4	1.90	0.1472/0.1929	0.1380/0.1877	100	0.25	0.50	5.0

was systematically placed within the active-site backbone geometry of each ensemble member. During this process, the positions of key catalytic residues of BH32.14, namely His23 and Arg124, were constrained according to their geometries in the DFT models, while sequence-specific interactions and full electrostatic contributions were not explicitly considered. This approach isolates the ability of the backbone architecture to support the catalytic arrangement required for optimum MBH transition state stabilization. The geometries and constraints used for this calculations can be found in Table 3.7 and Table 3.8.

Table 3.7: Geometric definitions for generation of transition-state poses off the side chain of catalytic residue.

TS	Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Value ^b
TS1	HID	Distance	NE2	C5			1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5
		Angle	CE1	NE2	C5		117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127
		Angle	NE2	C5	C6		112.8, 113.8, 114.8, 115.8, 116.8, 117.8, 118.8, 119.8, 120.8, 121.8, 122.8

Continued on next page

TS	Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Value ^b
		Torsion	ND1	CE1	NE2	C5	-172.7, -171.7, -170.7, -169.7, -168.7, -167.7, -166.7, -165.7, -164.7, -163.7, -162.7
		Torsion	CE1	NE2	C5	C6	-8.4, -7.4, -6.4, -5.4, -4.4, -3.4, -2.4, -1.4, -0.4, 0.6, 1.6
		Torsion	NE2	C5	C6	C1	-111.5, -110.5, -109.5, -108.5, -107.5, -106.5, -105.5, -104.5, -103.5, -102.5, -101.5
TS2	HID	Distance	NE2	C5			1.3, 1.4, 1.5
		Angle	CE1	NE2	C5		121.9, 122.9, 123.9, 124.9, 125.6, 126.6, 127.6, 128.6, 129.6, 130.6, 131.9
		Angle	NE2	C5	C6		105.4, 106.4, 107.4, 108.4, 109.4, 110.4, 111.4, 112.4, 113.4, 114.4, 115.4
		Torsion	ND1	CE1	NE2	C5	172.6, 173.6, 174.6, 175.6, 176.6, 177.6, 178.6, 179.6, 180.6, 181.6, 182.6
		Torsion	CE1	NE2	C5	C6	34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44
		Torsion	NE2	C5	C6	C1	-139.9, -138.9, -137.9, -136.9, -135.9, -134.9, -133.9, -132.9, -131.9, -130.9, -129.9
TS3	HID	Distance	NE2	C5			1.3, 1.4, 1.5
		Angle	CE1	NE2	C5		122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132
		Angle	NE2	C5	C6		106.7, 107.7, 108.7, 109.7, 110.7, 111.7, 112.7, 113.7, 114.7, 115.7, 116.7
		Torsion	ND1	CE1	NE2	C5	161.6, 162.6, 163.6, 164.6, 165.6, 166.6, 167.6, 168.6, 169.6, 170.6, 171.6
		Torsion	CE1	NE2	C5	C6	29.9, 30.9, 31.9, 32.9, 33.9, 34.9, 35.9, 36.9, 37.9, 38.9, 39.9
		Torsion	NE2	C5	C6	C1	-150.4, -149.4, -148.4, -147.4, -146.4, -145.4, -144.4, -143.4, -142.4, -141.4, -140.4
TS4	HID	Distance	NE2	C5			1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5
		Angle	CE1	NE2	C5		117.5, 118.5, 119.5, 120.5, 121.5, 122.5, 123.5, 124.5, 125.5, 126.5, 127.5

Continued on next page

TS	Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Value ^b
		Angle	NE2	C5	C6		98.4, 99.4, 100.4, 101.4, 102.4, 103.4, 104.4, 105.4, 106.4, 107.4, 108.4
		Torsion	ND1	CE1	NE2	C5	173.3, 174.3, 175.3, 176.3, 177.3, 178.3, 179.3, 180.3, 181.3, 182.3, 183.3
		Torsion	CE1	NE2	C5	C6	44.5, 45.5, 46.5, 47.5, 48.5, 49.5, 50.5, 51.5, 52.5, 53.5, 54.5
		Torsion	NE2	C5	C6	C1	-100.8, -99.8, -98.8, -97.8, -96.8, -95.8, -94.8, -93.8, -92.8, -91.8, -90.8

^aAtoms in bold are from the transition state; all other atoms are from the catalytic residue.

^bDistances are in Å; all other values are in degrees.

Table 3.8: Geometric constraints used to define catalytic contacts.

TS	Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Min ^b	Max ^b
TS1	HID	Distance	NE2	C5			1.5	2.5
		Angle	CE1	NE2	C5		117	127
		Angle	NE2	C5	C6		112.8	122.8
		Torsion	ND1	CE1	NE2	C5	-172.7	-162.7
		Torsion	CE1	NE2	C5	C6	-8.4	1.6
		Torsion	NE2	C5	C6	C1	-111.5	-101.5
	ARG	Distance	NH1	O1			2.6	3.0
		Distance	NH2	O1			2.6	3.0
		Distance	NH2	O7			2.8	3.2
		Angle	O1	NH1	NH2		55.5	75.1
		Angle	O7	O1	NH1		77.2	97.2
		Torsion	O1	NH1	NH2	CZ	161.1	181.1
		Torsion	O7	O1	NH1	NH2	4.4	24.4
		Torsion	C1	O7	O1	NH1	-179.1	-159.7
TS2	HID	Distance	NE2	C5			1.3	1.5
		Angle	CE1	NE2	C5		121.9	131.9
		Angle	NE2	C5	C6		105.4	115.4
		Torsion	ND1	CE1	NE2	C5	172.6	182.6
		Torsion	CE1	NE2	C5	C6	34	44
		Torsion	NE2	C5	C6	C1	-139.9	-129.9
	ARG	Distance	NH1	O1			2.5	2.9
		Distance	NH2	O1			3.2	3.6
		Distance	NH2	O7			2.5	2.9
		Angle	O1	NH1	NH2		74	94
		Angle	O7	O1	NH1		64.9	84.9
		Torsion	O1	NH1	NH2	CZ	124	144
		Torsion	O7	O1	NH1	NH2	32.5	52.5
		Torsion	C1	O7	O1	NH1	158.1	178.1
TS3	HID	Distance	NE2	C5			1.3	1.5
		Angle	CE1	NE2	C5		122	132

Continued on next page

TS	Residue	Type	Atom 1 ^a	Atom 2 ^a	Atom 3 ^a	Atom 4 ^a	Min ^b	Max ^b
	ARG	Angle	NE2	C5	C6		106.7	116.7
		Torsion	ND1	CE1	NE2	C5	161.6	171.6
		Torsion	CE1	NE2	C5	C6	29.9	39.9
		Torsion	NE2	C5	C6	C1	-150.4	-140.4
		Distance	NH1	O1			2.5	2.9
		Distance	NH2	O1			2.6	3.0
		Distance	NH2	O7			3.0	3.4
		Angle	O1	NH1	NH2		68.1	88.1
		Angle	O7	O1	NH1		73.2	93.2
		Torsion	O1	NH1	NH2	CZ	130.6	150.6
		Torsion	O7	O1	NH1	NH2	33.5	53.5
		Torsion	C1	O7	O1	NH1	123.4	153.4
TS4	HID	Distance	NE2	C5			1.8	2.5
		Angle	CE1	NE2	C5		117.5	127.5
		Angle	NE2	C5	C6		98.4	108.4
		Torsion	ND1	CE1	NE2	C5	173.3	183.3
		Torsion	CE1	NE2	C5	C6	44.5	54.5
	ARG	Torsion	NE2	C5	C6	C1	-100.8	-90.8
		Distance	NH1	O1			2.6	3.0
		Distance	NH2	O1			2.6	3.0
		Angle	O1	NH1	NH2		54.9	74.9
		Angle	O7	O1	NH1		121.4	141.4
		Torsion	O1	NH1	NH2	CZ	163.5	183.5
		Torsion	O7	O1	NH1	NH2	20.6	40.6
		Torsion	C1	O7	O1	NH1	-180.1	-160.1

^aAtoms in bold are from the transition state. All other atoms are from the catalytic residue.

^bDistance measurements given in Å; all others are in degrees.

For each placement, the backbone conformation was evaluated to determine whether it could accommodate the transition state without steric clashes and while maintaining the required geometric relationships between the transition state and the

catalytic residues. The fraction of ensemble members that successfully supported a given transition state was quantified as the “pose percentage,” representing the proportion of conformations compatible with stabilizing that transition state. These values therefore reflect how well the conformational landscape of each enzyme is preorganized, at the structural level, to support different steps along the reaction coordinate. The resulting pose percentages for all four transition states are shown in Figure 3.8B.

As shown in this analysis, the ancestral BH32 enzyme, which was originally designed to stabilize the first transition state, is highly enriched in conformations compatible with TS1. However, it shows very limited compatibility with the later transition states, with essentially no conformations capable of stabilizing TS4. In contrast, evolution appears to reshape the backbone conformational landscape and the geometry of the active site in a way that broadens this compatibility. In BH32.8, although the pose percentage for TS1 is reduced relative to BH32, there is a clear increase in compatibility across all four transition states, indicating that the enzyme has evolved a more balanced ability to support multiple steps along the reaction pathway.

Interestingly, although it might be expected that variants from other lineages, particularly OE1.4, would show little to no pose compatibility with these transition states, the calculations indicate that OE1.4 does exhibit measurable pose percentages (Figure 3.8B). However, this apparent compatibility does not necessarily translate into functionally relevant stabilization.

To further examine this, the orientations of the accepted poses were analyzed (Figure 3.8C). While BH32.8 and BH_{MeHis}1.8 show relatively consistent positioning of the transition states across ensemble members, OE1.4 displays a much broader distribution of orientations. This suggests that, although the backbone can geometrically accommodate the transition state, it does not enforce a well-defined or catalytically productive arrangement.

This observation is quantified in Figure 3.8D, where the RMSD between accepted poses is reported for each transition state. In OE1.4, the significantly higher RMSD values indicate substantial variability in transition state positioning across the ensemble. In contrast, BH32.8 and BH_{MeHis}1.8 show lower RMSD values, reflecting more consistent and convergent positioning of the transition states, which would facilitate passage along the reaction coordinate.

Together, these results demonstrate that evolution enhances the ability of MBH-catalyzing enzymes to stabilize multiple transition states within a single active site. In BH32.8 and BH_{MeHis}1.8, the reshaping of the backbone and active-site geometry leads to both increased compatibility across all transition states and more consistent positioning of those states within the ensemble. This reflects a shift from a highly specialized active site, optimized primarily for a single transition state in BH32, to a more adaptable catalytic environment capable of supporting multiple steps along the reaction coordinate.

3.4.8 Substrate-Dependent Movements in OE1.3 Resemble Evolutionary Adaptations in MBH-Catalyzing Families

Although members of the OE family exhibit the fewest structural adaptations in our evolutionary analysis, OE1.3 was shown to undergo significant domain movement upon binding to the covalent inhibitor 2-bromoacetophenone (Figure 3.9A), as previously reported by Burke et al [6]. This irreversible inhibitor was designed to mimic the enzyme–acyl intermediate. We aligned the published structure of OE1.3 bound to the inhibitor (PDB ID: 6Q7R) with our room-temperature structure of OE1.3, as well as with BH32 and BH32.8 (all aligned on cap domain). Strikingly, the domain repositioning observed in the inhibited OE1.3 structure closely resembles the structural changes that arose through the evolution of BH32 to yield higher activity variants of MBH enzymes.

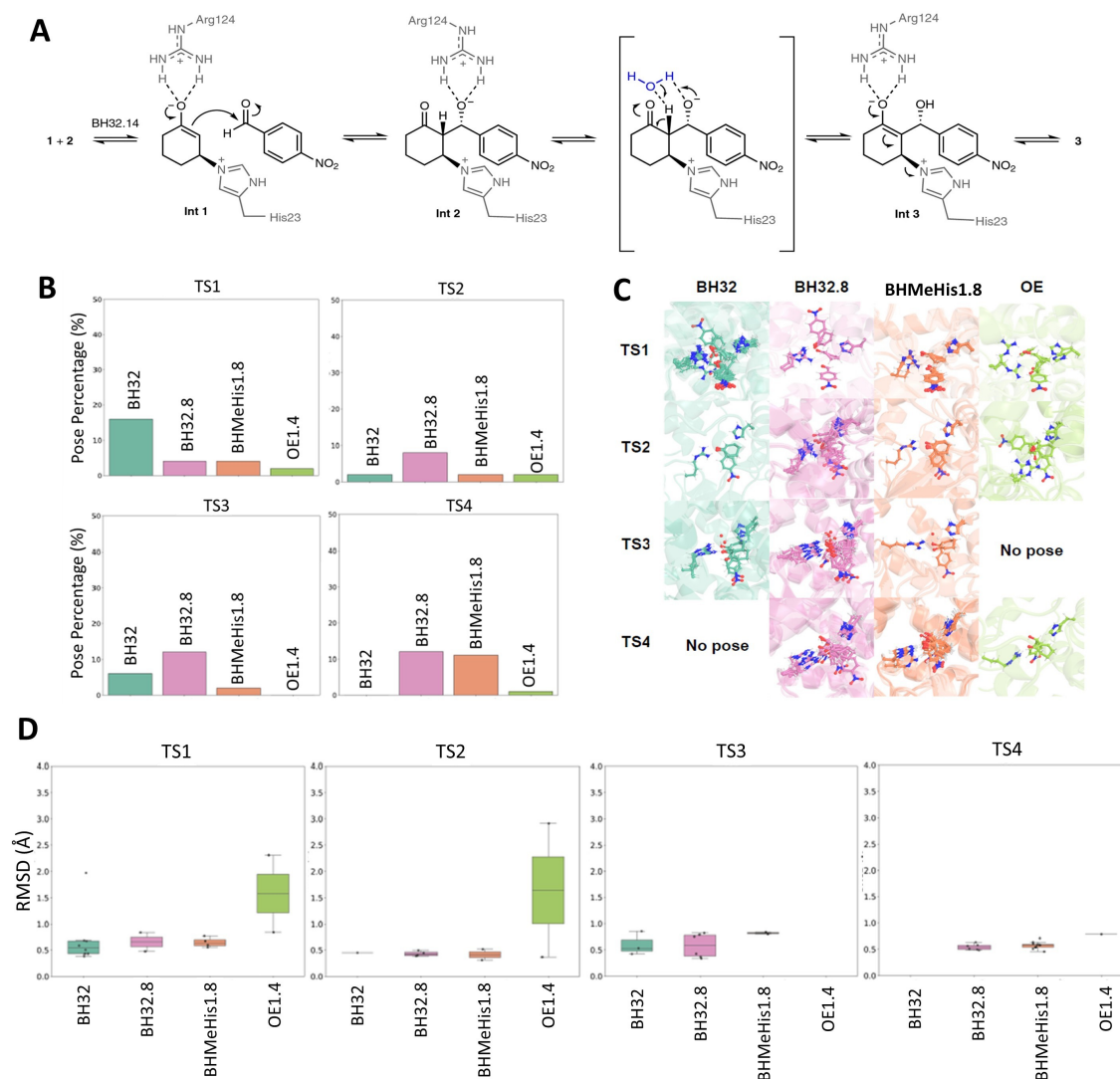


Figure 3.8: Evaluation of transition state compatibility across BH32 and evolved variants. (A) DFT-derived models of the four transition states (TS1–TS4) of the MBH reaction catalyzed by BH32.14. (B) Pose compatibility, shown as the percentage of backbone conformations capable of accommodating each transition state. (C) Representative orientations of compatible poses within the active-site backbone. (D) RMSD distributions of accepted poses, reflecting the consistency of transition state positioning across the ensemble.

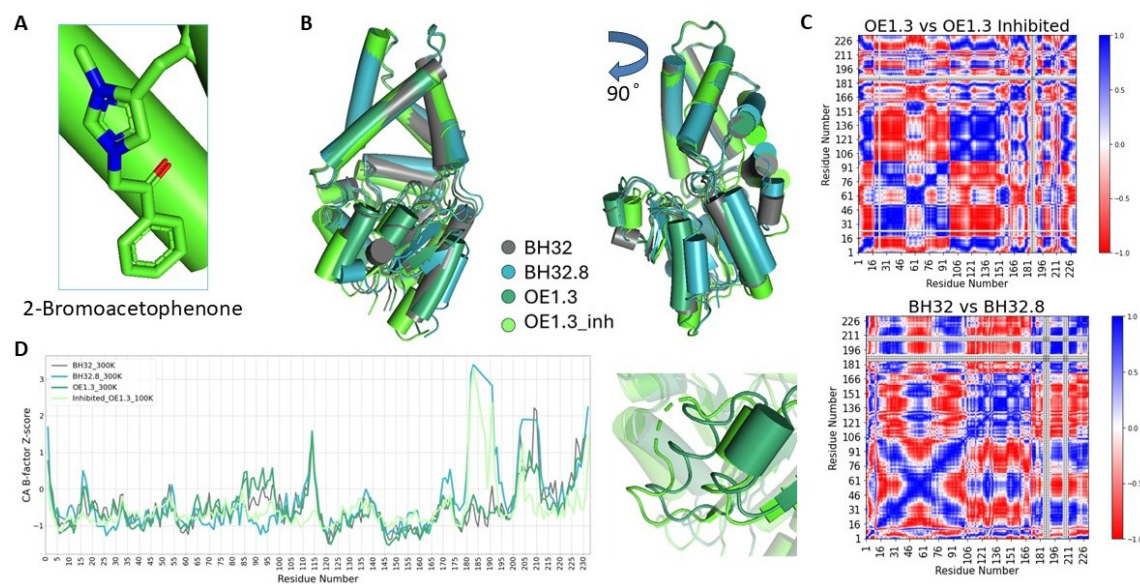


Figure 3.9: Structural changes induced by inhibitor binding in OE1.3 resemble evolutionary changes in BH32.8. A) The inhibitor used in the OE1.3 structure is 2-bromoacetophenone, which forms a covalent bond with the N ϵ of MeHis23. B) Structural alignment of BH32, BH32.8, OE1.3, and inhibited OE1.3 (aligned on cap domain) reveals that the inhibited OE1.3 structure shares several features with BH32.8, including cap domain widening, downward movement of helix 86–99, and a torsional shift in the core domain, although the nature of the motion differs. C) C α displacement vector correlation matrices comparing BH32 vs. BH32.8 and OE1.3 vs. inhibited OE1.3 show strikingly similar global rearrangements, suggesting that evolution in BH32.8 and inhibitor binding in OE1.3 lead to comparable structural effects. D) Like BH32.8, the inhibited OE1.3 structure shows increased backbone flexibility in the 180–195 region, as indicated by elevated C α B-factor Z-scores. As a result, this region is partly disordered and unmappable in the electron density.

As shown in Figure 3.9C, $C\alpha$ displacement vector correlation matrices for apo OE1.3 vs. its inhibited form and BH32 vs. BH32.8 reveal a strong similarity between the effect of inhibitor binding in OE1.3 and the evolutionary changes observed between BH32 and BH32.8. However, structural alignment of BH32, BH32.8, OE1.3, and inhibited OE1.3 (Figure 3.9B, aligned on the cap domain) reveals key differences in these motions. While BH32.8 primarily exhibits a hinge-like opening, the inhibited OE1.3 structure undergoes a twisting motion centered in the core domain, while the cap domain remains similarly widened in both. Notably, the inter-domain angle in inhibited OE1.3 remains close to that of BH32 and uninhibited OE1.3 (92.5°), suggesting that overall domain geometry is conserved even as the internal architecture shifts. Importantly, Burke et al. reported that the ability of the two domains to move relative to one another first emerged between OE1.2 and OE1.3, facilitated by a Leu10Pro mutation in the loop connecting the cap and core domains [6]. This points to an alternative evolutionary solution: instead of permanently expanding the binding pocket, OE1.3 may have evolved conformational flexibility, allowing the structure to dynamically open in response to the formation of the acyl-enzyme intermediate. Similar “substrate-assisted dynamics” have been observed for other enzymes that catalyze bi-substrate reactions [42]. However, it remains to be determined whether members of the BH32.n and BH_{MeHis}1.n family also dynamically open when bound to the substrate.

Further support for this mechanism comes from the $C\alpha$ B-factor Z-score analysis (Figure 3.9D), which highlights a highly dynamic region spanning residues 180–195. This segment is disordered in both inhibited OE1.3 and BH32.8, but remains ordered in uninhibited OE1.3. It’s another clear example of how inhibition can mimic the kinds of structural changes that appear during evolution. Overall, these results suggest a connection between hinge-bending motion and disorder in the regions close to the active site entrance, as both occur concurrently in response to evolutionary mutations and upon inhibition of OE1.3. Although the current evidence does not allow

a definitive causal conclusion, disorder in residues 180-195 may function analogously to a loosening of a clip that connects the cap and core domains, thereby enabling the hinge-bending motion required for changes in active-site shape (Figure 3.10).

3.4.9 The His23MeHis Substitution Does Not cause Major Structural Changes

Studying the BH32.8/BH_{MeHis}1.0 pair allowed us to isolate and examine the direct structural impact of the N δ -methyl group on the catalytic histidine. Figure 3.11 shows the overall similarity between BH32.8 and BH_{MeHis}1.0, with structural alignment of their X-ray crystallography models yielding an RMSD of 0.111 .

The packing of the catalytic (Me)His differs slightly between the two variants, as the introduction of the additional methyl group alters both its orientation and local packing environment. However, comparison of the C α B-factor Z-scores indicates no significant differences, suggesting that this modification does not substantially affect global backbone dynamics.

3.4.10 Crystal Cryocooling Distorts Structural Measurements

Although cryocooling is routinely used to mitigate radiation damage, it can alter conformational sampling and obscure functionally relevant structural features. To evaluate the impact of cryogenic data collection on structural interpretation, we compared models obtained at 100 K and 300 K for each enzyme variant. Representative examples of these effects are shown in Figure 3.12.

Global structural comparisons reveal RMSD values of approximately 0.25 Å across all enzymes, reaching up to 0.30 Å in OE1.3 and OE1.4 (Figure 3.12A). While such values might appear small, in this system they correspond to meaningful structural differences, particularly in inter-domain orientations. As illustrated in Figure 3.12C, these shifts can alter the relative angle between the cap and core do-

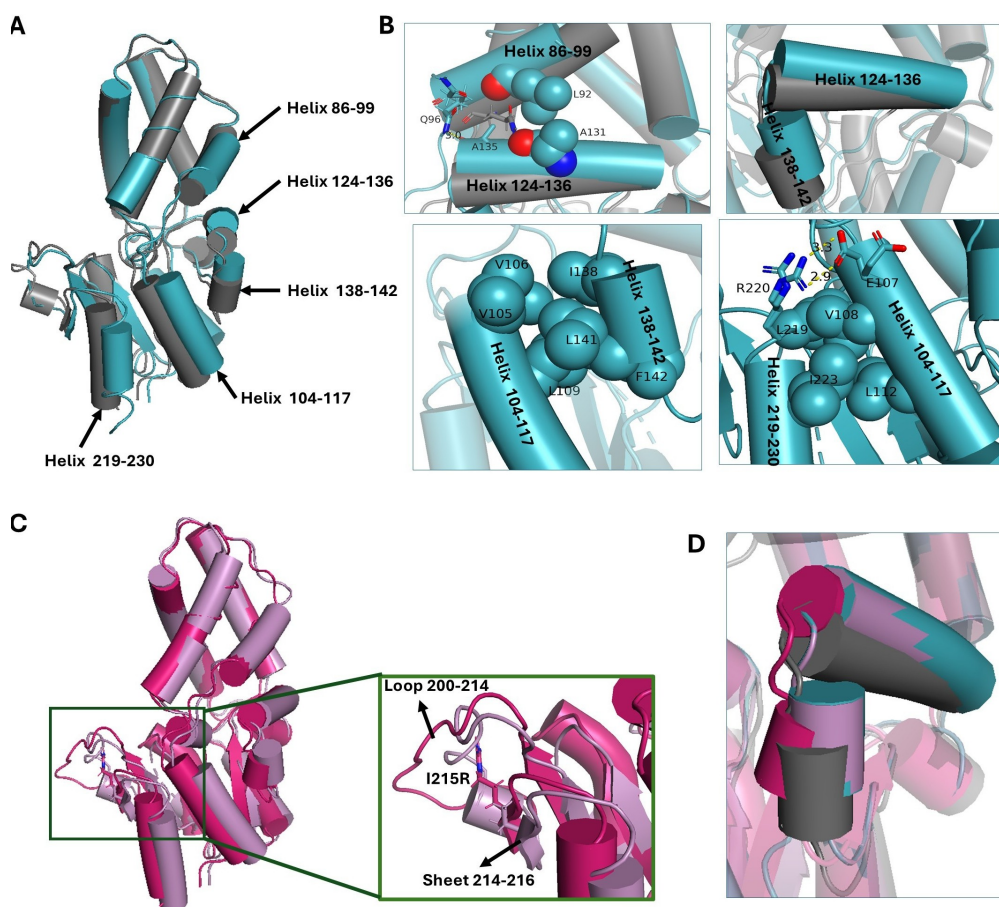


Figure 3.10: Possible coupling between frontal loop disorder and hinge-like motions in the protein structure. (A) Structural alignment of BH32 and BH32.8, anchored on the cap domain, reveals an upward shift of helix 86–99 that propagates through adjacent helices, including 124–136, 138–142, 104–117, and 219–230. (B) Close-up views of inter-helical interactions show that packing rearrangements and a flexible linker between helices 124–136 and 138–142 enable transmission of this motion across the structure. (C) In BH_{MeHis}1.8, reduced hinge opening is observed and can be attributed to the I215R mutation, which forms stabilizing interactions with loop 200–214 and the β -sheet spanning residues 214–216. These interactions restrict the upward movement of helices 219–230, 104–117, and 138–142. (D) Despite this local restriction, the flexible linker decouples helix 124–136 from the constrained region, allowing helix 124–136 and helix 86–99 to retain similar upward shifts as observed in BH32.8 and BH_{MeHis}1.0.

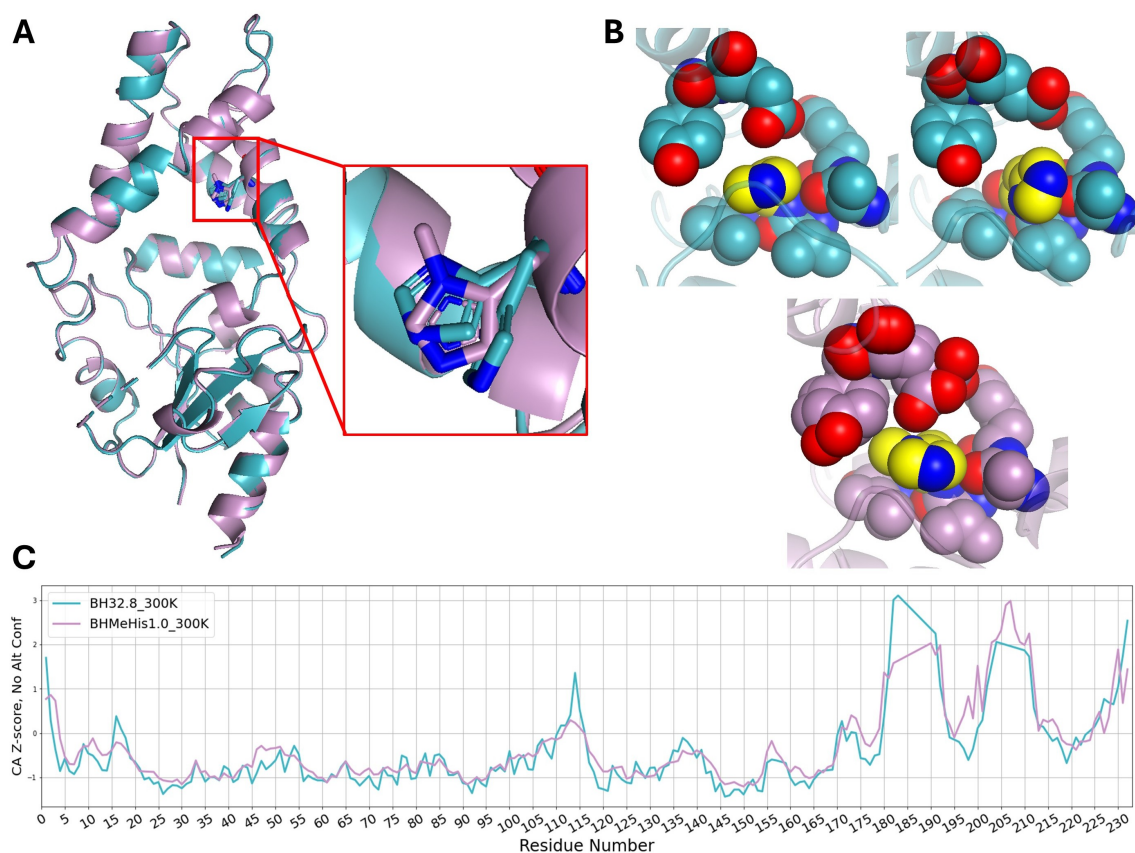


Figure 3.11: Minimal structural impact of His23 methylation between BH32.8 and BHMeHis1.0. (A) Overlay of BH32.8 (teal) and BHMeHis1.0 (violet) shows highly similar overall structures, with no significant backbone rearrangements upon introduction of N δ -methylhistidine at position 23. The inset highlights the active site, where His23 adopts different rotamers in the two variants as a result of methylation, with BH32.8 sampling multiple conformations and BHMeHis1.0 favoring a single dominant rotamer. (B) Close-up views of the active site illustrate the local packing environment around His23. Addition of the methyl group leads to slightly tighter packing compared to the very loosely packed environment in BH32.8, although the active site remains overall relatively open. (C) C α B-factor Z-score profiles for BH32.8 (teal) and BHMeHis1.0 (violet) are largely overlapping across the sequence, indicating that methylation does not significantly alter global or local backbone dynamics.

mains, potentially leading to misleading conclusions when comparing conformational changes across evolutionary variants.

Analysis of overall residue dynamics further highlights the impact of temperature. Order parameter profiles show differences in the overall dynamic behavior of the cryocooled and room-temperature structures (Figure 3.12B). Direct structural comparisons provide further insight into these effects. In BH32, a hinge-like opening is observed at 300 K (Figure 3.12C, dark grey) that is absent in the cryogenic structure (Figure 3.12C, light grey), indicating that cryocooling restricts access to this conformational state. The previously discussed regions 180–195 and 200–210 are also affected by cryocooling. In BH_{MeHis}1.0, segments that are well-defined at room temperature become disordered or unmappable at 100 K (Figure 3.12E, left). In BH_{MeHis}1.8, the conformation of the loop spanning residues 200–210 also differs between the two temperatures (Figure 3.12E, right). Panel D in Figure 3.12 shows that the rotamer of (Me)His23 is also affected by cryocooling. In the top panel, the teal rotamer corresponds to the 300 K structure, while the green rotamer corresponds to the cryo-cooled structure. Similarly, in BH_{MeHis}1.0, the violet structure represents 300 K and the lighter violet structure represents 100 K, and their conformations show significant differences.

Additional local effects are also evident. Cryocooling influences side-chain conformations within the active site, including alternative rotameric states of catalytic residues and residues close to the active site (Figure 3.12F). In BH32.8, the alternative conformation of Trp88, a key residue that directly affects backside access to the active site, cannot be mapped at cryogenic temperature (Figure 3.12F, top). In OE1.4, Arg50, which is located directly in front of the catalytic histidine, also adopts different rotamers at cryogenic temperature (Figure 3.12F, bottom).

Together, these results demonstrate that cryo-cooling can reshape both global and local features of the conformational ensemble. These effects can obscure functionally relevant conformations and complicate comparisons across enzyme variants,

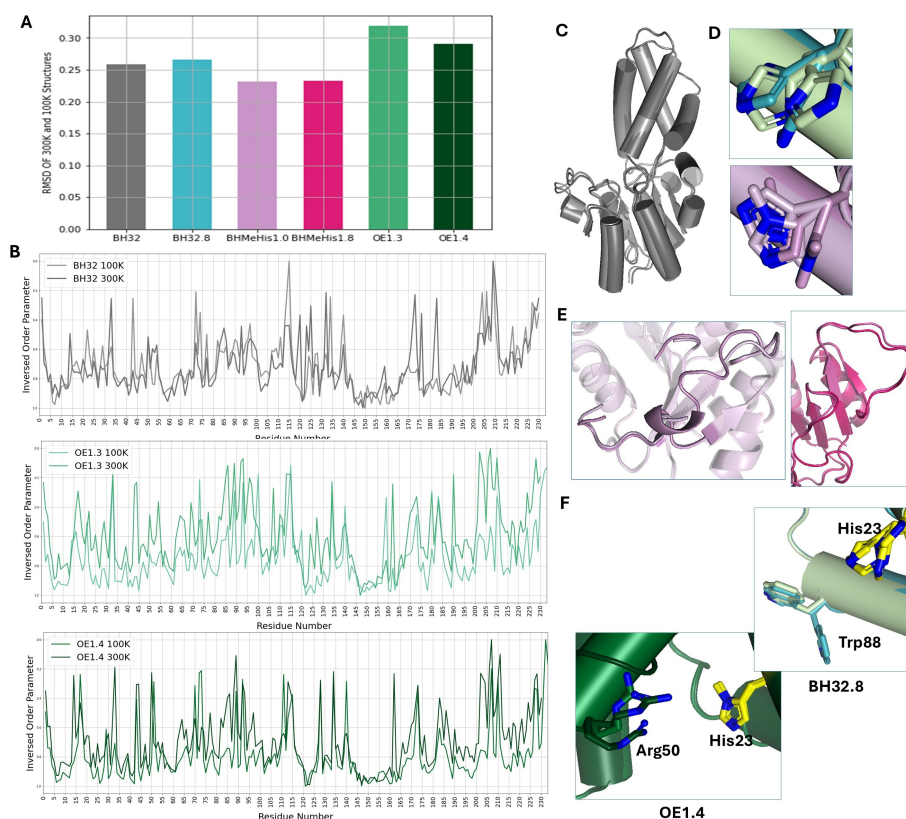


Figure 3.12: Effect of temperature on structural modeling and residue dynamics. (A) RMSD between 100 K and 300 K structures (0.25–0.30 Å) indicates altered conformational sampling. (B) Order parameter profiles reveal temperature-dependent changes in the overall dynamic behavior of residues across variants. (C) Structure alignment anchored on the cap domain illustrate temperature effects on backbone conformation; in BH32, a hinge opening is observed at 300 K (dark grey) but not at 100 K (light grey). (D) The catalytic (Me)His adopts different conformations at 100 K and 300 K: in BH32.8 (teal vs green) and BH_{MeHis}1.0 (dark vs light velvet). (E) Regions 180–195 and 200–210 are affected by cryocooling. In BH_{MeHis}1.0, region 200–210 is resolved at 300 K (dark violet) but unmappable at 100 K (light violet); in BH_{MeHis}1.8, this loop adopts different conformations at 100 K (light pink) and 300 K (magenta). (F) Side-chain conformations near the active site are also altered ((Me)His in yellow): in BH32.8, the alternative conformation of Trp88 is present at 300 K (teal) but absent at 100 K (green); in OE1.4, Arg50 adopts distinct conformations at 100 K (light green) and 300 K (dark green).

highlighting the importance of room-temperature crystallography for accurately capturing the structural basis of enzyme function.

3.5 Discussion

In this study, a cross-family comparison of enzymes sharing a common ancestor provides insight into how structural features evolve in response to functional demands. Enzyme families evolved to catalyze the Morita–Baylis–Hillman (MBH) reaction (BH32.n and BH_{MeHis}1.n) exhibit convergent structural adaptations, including active-site opening, increased flexibility of surface loops adjacent to the active site, and a pronounced clamshell-like hinge motion. However, the extent and functional implications of these changes depend on the specific orientation of the catalytic residue and the geometry of the reaction intermediates, which differ slightly between the BH32.n and BH_{MeHis}1.n lineages. In contrast, these features are largely absent, or only weakly present (e.g., modest cap-domain widening), in the OE1.3 and OE1.4 enzymes that catalyze ester hydrolysis. Notably, OE1.3 undergoes substantial structural rearrangements upon covalent inhibition, resembling those observed in the MBH-catalyzing families in their unbound states. This suggests that, rather than maintaining a constitutively open active site, the OE1 lineage may rely on substrate-induced conformational changes.

The combined effect of the structural adaptations observed in the MBH lineages is the formation of distinct active-site cavity shapes (Figure 3.2). Pose compatibility analysis indicates that these reshaped cavities are optimized to accommodate multiple transition states along the reaction coordinate (Figure 3.8). While the ancestral BH32 enzyme is highly compatible with the first transition state (consistent with its original design) it shows poor compatibility with subsequent states and lacks conformations capable of stabilizing TS4. In contrast, evolved variants such as BH32.8 display increased compatibility across all four transition states. Importantly, this

evolution does not lead to optimal stabilization of any single transition state, but rather to a more balanced ability to support multiple states.

These observations highlight a key principle in enzyme catalysis. For reactions in which a single step is rate-limiting, optimization of one transition state may be sufficient. However, for multi-step reactions with several kinetically significant steps, catalytic efficiency requires the ability to stabilize multiple transition states. In this context, the structural features observed in the MBH-catalyzing families, including the active-site widening, increased flexibility, and coupled domain motions, collectively enable a more versatile catalytic environment capable of supporting the full reaction pathway.

It is noteworthy that although both the MBH and ester hydrolysis reactions proceed through multiple steps, they differ significantly in substrate requirements. The MBH reaction involves a bulky second substrate, likely necessitating a wider and more accessible active site, whereas ester hydrolysis involves a much smaller second substrate, a water molecule, reducing this requirement. For us, due to the lack of DFT-derived transition states for ester hydrolysis, it was difficult to directly assess the compatibility of OE1.3 with its reaction coordinate. Furthermore, the structures analyzed here are primarily apo or inhibitor-bound and may not fully represent the substrate-bound or intermediate states. As such, the apparent absence of large-scale structural adaptations in the OE family, nor the helix shift in inhibited OE1.3, does not preclude the possibility of substrate-induced conformational changes, nor does it imply that the MBH-catalyzing families lack access to similar dynamic mechanisms.

Finally, since the present analysis is based on equilibrium structures, it does not capture non-equilibrium enzyme–substrate or enzyme–intermediate states along the reaction coordinate. Time-resolved or mixing experiments (discussed in Chapter 4) that directly probe these transient states will therefore be essential to fully understand how these enzymes stabilize different transition states during catalysis.

3.6 Future Directions

3.6.1 More Efficient Members of the BH32.n Family as a Source of Structural Insight

As mentioned before, the most catalytically efficient member in the BH32.n family is BH32.14 [4]. However, even though extensive crystallization and data collection efforts were carried out for this enzyme, which was initially targeted for comparison with BH_{MeHis}1.8 and OE1.4, it was not successful. While these efforts produced crystals under multiple conditions (Figure 3.13A), the resulting datasets did not reach a quality sufficient for reliable model building.

Despite this, the crystallization behavior of BH32.14 provides useful guidance for future optimization and highlights that obtaining structures of later-stage variants in this family remains an achievable and important goal. While structural and dynamic trends have already been established across the available variants, inclusion of higher-efficiency members such as BH32.12 or BH32.14 would complete the evolutionary trajectory and allow for a more continuous mapping of how conformational landscapes evolve with catalytic improvement.

Importantly, BH32.8 represents a key branching point in this system, as it is the latest common ancestor to both the BH32.n and BH_{MeHis}1.*n* families. This makes it a particularly informative reference state, where multiple conformations may coexist and serve as potential starting points for divergent evolutionary paths. Access to structures of more advanced BH32.n variants would therefore provide the opportunity to directly compare how evolution selects and stabilizes distinct conformational states from a shared ensemble.

In particular, the region adjacent to the active site that becomes disordered in intermediate variants is of significant interest. The disorder likely provides a pool of conformations with similar energetic accessibility, which can be optimized by further

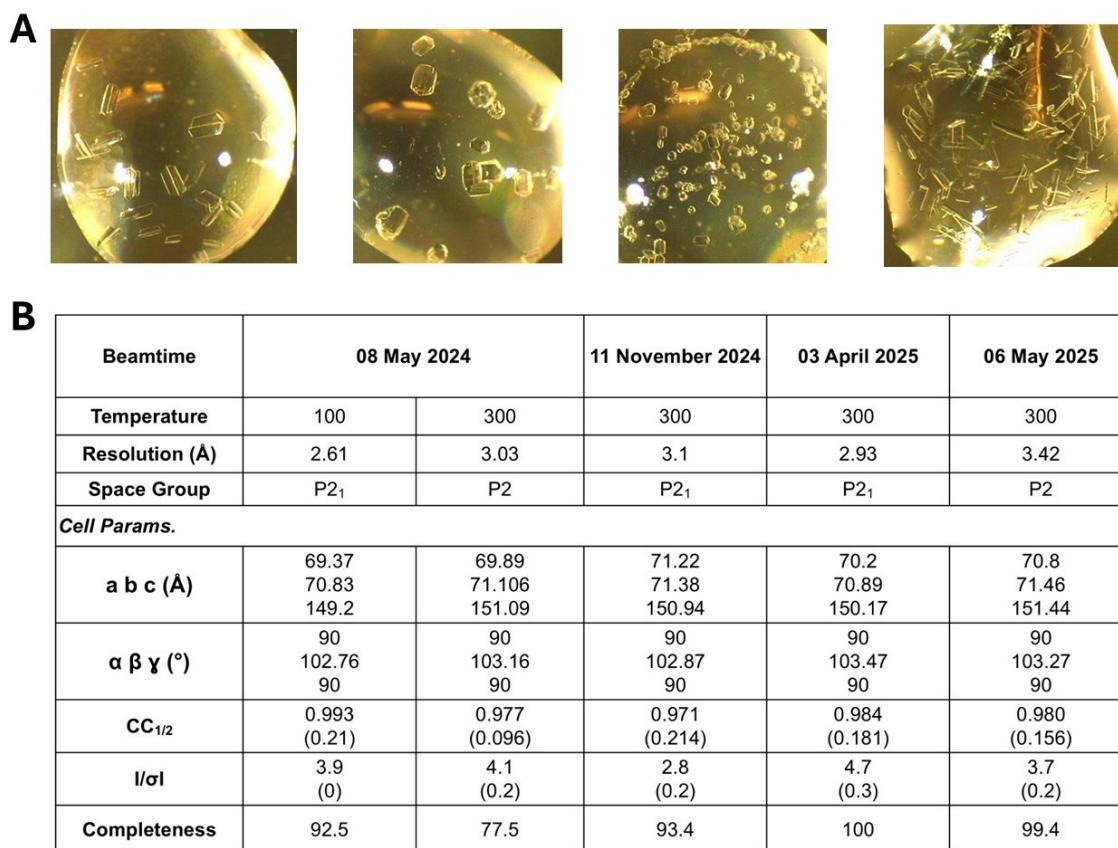


Figure 3.13: Crystallization screening of BH32.14 under a range of conditions, including standard NexTal screens, seeding, additive screening, and temperature optimization. Although crystals of moderate quality were obtained (A), they did not yield datasets suitable for structural modeling (B).

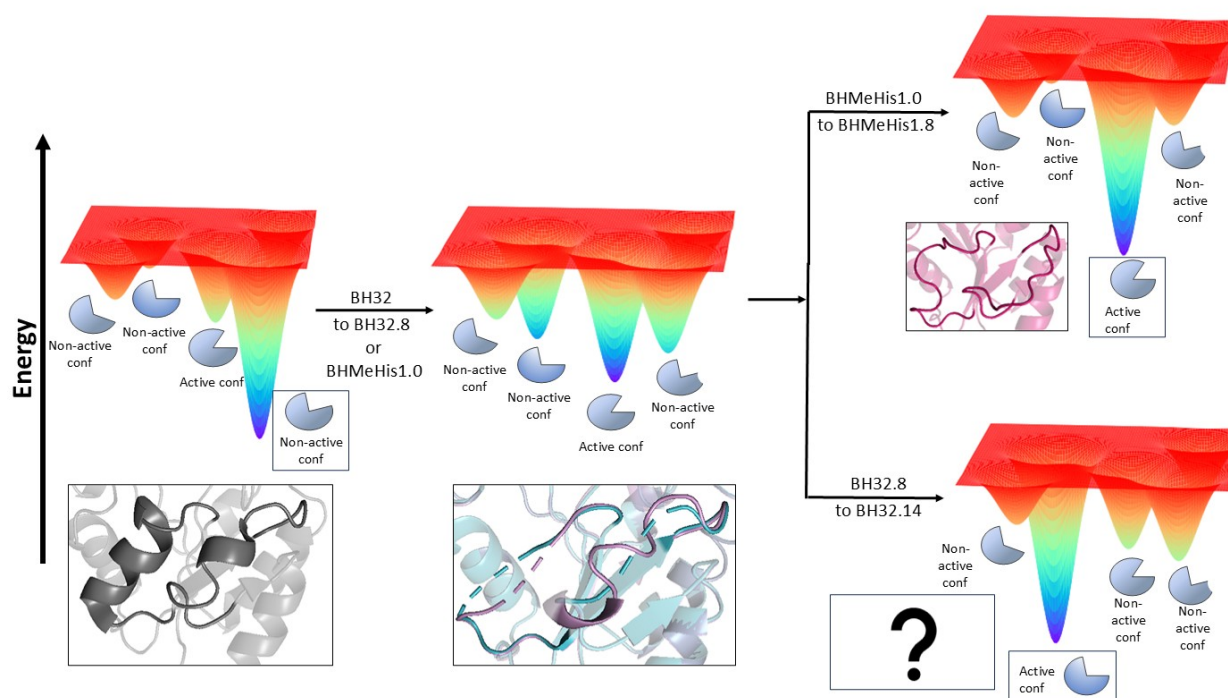


Figure 3.14: Evolution from BH32 to BH32.8 (and similarly BH_{MeHis}1.0) leads to destabilization of helical regions spanning residues 180–195 and 200–210, resulting in increased conformational heterogeneity. In BH_{MeHis}1.8, these regions become partially restabilized into flexible loop conformations. Structural characterization of BH32.14 or BH32.12 would provide a direct comparison to determine how alternative conformational states are selected and stabilized along the BH32.n evolutionary pathway.

rounds of directed evolution. Comparing these regions across BH32.n and BH_{MeHis}1.n families would allow us to determine whether evolution converges on similar structural solutions or diverges to stabilize distinct functional conformations (Figure 3.14).

Future efforts will therefore focus on optimizing crystallization conditions and data collection strategies to obtain high-quality room-temperature datasets for BH32.12 and BH32.14. Achieving this will enable a more complete comparison across evolutionary trajectories and provide deeper insight into how conformational heterogeneity is selectively reshaped to support catalytic function.

3.6.2 Obtaining Covalently Inhibited Structures for More Variants

Expanding the set of covalently inhibited structures across enzyme families would provide a more complete view of how active-site geometry and dynamics respond to ligand binding. In particular, obtaining room-temperature inhibited structures of OE1.3 and OE1.4 would help determine whether the structural rearrangements observed in OE1.3 upon inhibition in cryo temperature are a general feature of the OE1 lineage without the artifacts of cryocooling. Similarly, acquiring inhibited structures for additional members of the BH32.n and BH_{MeHis}1.n families would allow direct comparison of how active-site reshaping influences ligand accommodation across evolutionary trajectories.

Additionally, testing covalent inhibitors with varying degrees of steric bulk would provide insight into the functional consequences of active-site widening in the MBH-catalyzing families. Such studies would enable systematic evaluation of how increased cavity size and flexibility impact the ability of these enzymes to accommodate larger substrates or intermediates. Together, these experiments would help establish whether the observed structural adaptations translate into enhanced tolerance for steric diversity and further clarify the relationship between active-site architecture

and catalytic function.

3.6.3 DFT of Ester Hydrolysis Transition States and Pose Compatibility Analysis

In parallel with obtaining additional inhibited structures, performing DFT calculations for the transition states of the ester hydrolysis reaction would provide valuable insight into the functional relevance of the observed OE1.3 and OE1.4 apo and inhibited structures. These transition states could be used to evaluate pose compatibility with the inhibited OE1.3 structure, enabling direct comparison between the active-site geometry and the requirements of the catalytic mechanism.

Such analysis would help determine whether the inhibited conformation represents a catalytically relevant state by assessing its ability to accommodate the transition states of the reaction. In this way, combining structural and computational approaches would provide a more complete understanding of how active-site architecture relates to catalytic function in the OE family.

3.7 Methods

3.7.1 Protein Expression and Purification

For the expression of BH32, BH32.8, and BH32.14, chemically competent *E. coli* DH5 α cells were transformed with pBbE8k_BH32-His, pBbE8k_BH32.8-His, and pBbE8k_BH32.14-Strep, respectively. Single colonies of freshly transformed cells were cultured for 18 hours in 25 mL LB medium supplemented with 35 g/mL kanamycin at 37 °C. Starter cultures (10 mL) were used to inoculate 1 L 2 $\times$ YT medium containing 35 g/mL kanamycin. Cultures were grown at 37 °C, 200 rpm until they reached an OD600 of 0.7, at which point L-arabinose (3.33 mM final concentration) was added to induce protein expression. For the expression of OE1.3,

OE1.4, BH_{MeHis}1.0, and BH_{MeHis}1.8, *E. coli* DH5 α cells containing pEVO_PylRSM_{MeHis}/tRNACUA were transformed with the appropriate pBbE8k construct. The expression protocol followed that of BH32, with the addition of chloramphenicol (25 g/mL). To incorporate Me-His, 10 mM Me-His was added to the culture before induction. Following induction, cultures were incubated for 20 hours at 25 °C, and cells were harvested by centrifugation (4,500g for 40 minutes). Pelleted cells were resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, pH 7.5, 20 mM imidazole) and lysed by sonication. Cell lysates were cleared by centrifugation (4,500g for 30 minutes), and supernatants were subjected to affinity chromatography using Ni Sepharose High Performance columns (Cytiva). For His-tagged variants, proteins were eluted using 50 mM HEPES, 300 mM NaCl, pH 7.5, 250 mM imidazole via a HisTrap HP column (Cytiva) and analyzed by SDS-PAGE. For Strep-tagged variants, pelleted cells were resuspended in Buffer NP (50 mM NaHPO, 300 mM NaCl, pH 8.0) and lysed by sonication. Cleared lysates were applied to a StrepTrapXT preppacked chromatography column (Qiagen), and purified protein was eluted using 50 mM NaHPO, 300 mM NaCl, 50 mM biotin, pH 8.0, followed by SDS-PAGE analysis. For further purification, BH32.14 was dialyzed against 50 mM HEPES, 300 mM NaCl, pH 8.0 and purified by size-exclusion chromatography using a Superdex 75 column. The remaining proteins were purified by size-exclusion chromatography on a Superdex 75 column in 50 mM HEPES, 300 mM NaCl, pH 7.5. Protein concentrations were determined by UV absorption at 280 nm.

3.7.2 Crystallization

Enzyme variants were prepared at the concentrations listed in Table 3.2. Initial crystallization trials were conducted for each enzyme variant using 96-well NeXtal crystallization screens (NeXtal Biotechnologies) and further optimized using a 15-well hanging drop format with EasyXtal crystallization plates (NeXtal). Crystallization

drops were prepared by mixing 1 μL of protein solution with 1 μL of mother liquor, followed by sealing the drop inside a reservoir containing an additional 500 μL of the mother liquor solution. The mother liquor solutions and the specific growth conditions that yielded the crystals used for X-ray data collection are provided in Table 3.2. In some cases, a microseeding protocol was required to obtain high-quality crystals. Microseeds were prepared by vortexing crystals in their mother liquor in the presence of 0.5 mm glass beads, followed by a 40x dilution into the mother liquor solutions used for crystallization drop formation.

3.7.3 X-ray Data Collection and Processing

Prior to X-ray data collection, crystals were mounted in polyimide loops and sealed using a MicroRT tubing kit (MiTeGen). Single-crystal X-ray diffraction data were collected on beamline 8.3.1 at the Advanced Light Source. The beamline was equipped with a Pilatus3 S 6M detector and was operated at a photon energy of 11111 eV. Also, the Stanford Synchrotron Radiation Lightsource beamline 12-1, equipped with an Eiger X 16M detector and was operated at a photon energy of 11111 eV, was used. Crystals were maintained at 100K and 300K throughout the course of data collection. Each data set was collected using a total X-ray dose of 200 kGy or less and covered a 180° wedge of reciprocal space. Multiple data sets were collected for each enzyme variant either from different crystals or, if their size permitted, from unique regions of larger crystals. X-ray data were processed with the Xia2 0.5.492 program [153], which performed indexing, integration, and scaling with the 20180126 version of XDS and XSCALE [154], followed by merging with Pointless as distributed in CCP4 7.0.05333 [155]. For each variant, multiple individual data sets were merged to obtain the final set of reduced intensities, and the resolution cutoff was taken where the CC1/2 and $\langle I/I \rangle$ values for the merged intensities fell to 0.5 and 0.8, respectively. We determined which individual data

sets should be combined by evaluating the overall effects of adding or removing individual data sets on the CC1/2 and I/ in the high-resolution bins of the merged data set. Information regarding data collection and processing is presented in Table 3.3 and 3.4. The reduced diffraction data were analyzed with phenix.xtriage (http://www.ccp4.ac.uk/newsletters/newsletter43/articles/PHZ_RWG_K_PDA.pdf) to check for crystal pathologies, and no complications were identified.

3.7.4 Structure Determination

Initial phase information for the calculation of electron density maps was obtained by molecular replacement using Phaser[127], as implemented in v1.13.2998 of the PHENIX suite[128]. The cryogenic structure of BH32 alkylated with the mechanistic inhibitor 2-bromoacetophenone (PDB ID: 6Q7N) [4] was used as the molecular replacement search model for all enzymes. Each enzyme contained one copy of the molecule in the crystallographic asymmetric unit. Iterative steps of manual model rebuilding were performed, followed by refinement of atomic positions, atomic displacement parameters, and occupancies. Refinements included the use of a translation-libration-screw (TLS) model, a riding hydrogen model, and automatic weight optimization. Model building was conducted using Coot 0.8.9.2[129], while refinement was performed with phenix.refine within the PHENIX suite (v1.13-2998)[130, 128]. Further details regarding model building, and refinement for the final models are provided in Tables 3.3 and 3.4. Time-averaged ensemble refinements were generated for all enzymes using phenix.ensemble_refinement was implemented in PHENIX v1.15.2-3472 [69]. For B-factor Z-score comparison, alternative conformations were removed from the final models using phenix.pdbtools, and they were then refined again using Phenix.refine. Hydrogen atoms were subsequently added using phenix.ready_set. This procedure yielded ensembles containing between 20 and 100 templates for all enzymes. Order parameters were calculated using the qfit-3.0

suite [156].

3.7.5 Ensemble Generation

Ensembles of backbone templates were generated using the ensemble refinement protocol [69] implemented in the PHENIX macromolecular structure determination software (v1.18.2-3874) [110]. In ensemble refinement, local atomic fluctuations are sampled by using MD simulations accelerated and restrained by electron density to produce ensemble models fitted to diffraction data. Following the addition of riding hydrogens, ensemble refinement simulations were performed using various combinations of the parameters pTLS (0.6, 0.8, 0.9, 1.0), τx (0.5, 1.5, 2.0), and wX-ray (2.5, 5.0, 10.0), where pTLS describes the percentage of atoms included in a translation-libration-screw (TLS) model use to remove the effects of global disorder, τx is the simulation time step, and wX-ray is the coupled temperature bath offset, which controls the extent to which the electron density contributes to the simulation force field such that the simulation runs at a target temperature of 300 K. The ensemble generated from each crystal structure that displayed the lowest Rfree value was selected for design and analysis (Table 3.6).

3.7.6 Theozyme placement

All transition states (TS) were modeled using the Triad protein design software [111] (Protabit, Pasadena, CA, USA). After extracting heavy-atom coordinates from each backbone structure, the `addH.py` application was used to add hydrogen atoms and format atom names according to the Triad convention prior to calculation. All TS coordinates for the Morita–Baylis–Hillman (MBH) reaction were provided by the Anthony Green group [4].

All active-site residues, except the catalytic residues HID23 and Arg124 (positions 10, 14, 19, 22, 26, 53, 64, 68, 91, 92, 94, 95, 122, 123, 128, 132), were mutated to

Gly. TSs were placed into the active site based on contact geometries with HID23 [86] (Table 3.7). Side-chain conformations of HID23 and Arg124 were sampled using the 2002 Dunbrack backbone-independent rotamer library, with expansions of ± 1 standard deviation around χ_1 and χ_2 [112].

Rotamer optimization was performed using a Monte Carlo simulated annealing algorithm. TS pose energies were calculated using a modified version of the Phenix energy function [111], consisting of: (i) a Lennard–Jones 12–6 van der Waals term from the Dreiding II force field [113] with atomic radii scaled by 0.9; (ii) a direction-dependent hydrogen bond term with a well depth of 8.0 kcal mol⁻¹ and an equilibrium donor–acceptor distance of 2.8 Å [114]; and (iii) an electrostatic term modeled using Coulomb’s law with a distance-dependent dielectric of 10.

For non-covalent states (TS1 and TS4), a standard force field distance cutoff of 35.0 Å was used to capture long-range non-bonded interactions. For covalent intermediates (TS2 and TS3), the cutoff was reduced to 1.0 Å to suppress short-range steric repulsion (van der Waals clashes) at covalent bonding distances. A penalty of 1000 kcal mol⁻¹ was applied when TS–side-chain interactions failed to satisfy catalytic contact constraints (Table 3.8). TS poses with energies greater than 500 kcal mol⁻¹ were discarded.

The pose percentage for each structural ensemble was calculated as the ratio of structures yielding valid TS poses to the total number of input structures. To assess accuracy, each TS pose was superimposed onto the DFT model by pair-fitting TS atoms. The root-mean-square deviation (RMSD) was then calculated for the guanidinium group of Arg124 and the imidazole group of HID23.

Chapter 4

Attempts to Capture Non-Equilibrium Reaction Intermediates Using Advanced Crystallographic Approaches

Copyright: Behnoush Seifinoferest, Zak Birch-Price, Anthony Green, Roberto A. Chica, Michael C. Thompson.

4.1 Statement of Contribution

M.C.T., R.A.C., and A.G. conceived the project, Z.B. conducted directed evolution experiments, B.S. purified and crystallized proteins, M.C.T., and B.S. designed and performed X-ray diffraction experiments, and B.S. conducted refinements and computational analysis. B.S. wrote the chapter, M.C.T. edited the chapter.

4.2 Abstract

Time-resolved X-ray crystallography provides a powerful approach for capturing enzyme–intermediate complexes that exist transiently outside of equilibrium and are typically inaccessible to conventional single-crystal X-ray crystallography. Here, we describe a series of efforts aimed at capturing non-equilibrium intermediates using approaches such as mix-and-inject serial crystallography and slow-reacting engineered enzyme systems. These approaches aim to provide insight into the structural and dynamic features that govern enzymatic function by enabling direct observation of transient states along the reaction coordinate.

4.3 Introduction

X-ray crystallography has proven to be a highly effective method for capturing protein structures at atomic resolution [56]. However, conventional single-crystal structures provide only a static snapshot and cannot directly report on how enzymes behave along the course of an enzymatic reaction. To address this limitation, various approaches such as site-directed mutagenesis [157, 158], transition-state analogues [159, 160], and inhibitors [161] have been used to mimic enzyme–substrate or enzyme–transition state complexes and infer aspects of catalytic mechanisms. While valuable, these approaches rely on approximations and do not capture the actual progression of the reaction.

Crystallography, in principle, offers the opportunity to directly observe non-equilibrium intermediates such as enzyme–substrate, enzyme –intermediate, and enzyme–product complexes. Mixing-based crystallographic experiments [71, 162, 42] have therefore emerged as powerful approaches for probing reactions across different time scales. At one extreme, mix-and-inject experiments using either XFELs [163, 164] or synchrotron sources [165, 166] enable access to ultrafast processes on

picosecond to millisecond timescales.

In this chapter, we explore two complementary strategies aimed at capturing such non-equilibrium intermediates. In the first, we take initial steps toward developing an experimental design for synchrotron-based mix-and-inject serial crystallography on OE1.4. This effort is motivated by the large-scale helix shift observed in OE1.3, a lower-efficiency ancestor of OE1.4, upon covalent inhibition [5]. This conformational change suggests the possibility of functionally relevant structural transitions that may be captured through time-resolved approaches.

In the second strategy, we pursue a simplified approach that does not require specialized mix-and-inject instrumentation. Using a family of enzymes evolved in parallel from BH_{MeHis}1.6 [5] to catalyze an allylic substitution reaction (ASB1.3), we take advantage of the intrinsic kinetics of the system to capture non-equilibrium states using single-crystal X-ray crystallography. In this system, the enzyme forms a stable complex with substrate 1 in the absence of substrate 2, effectively trapping the reaction at an early stage. Upon addition of substrate 2, the reaction proceeds slowly, creating a time window in which intermediate states can be captured. This approach enables structural characterization of non-equilibrium states along the reaction coordinate using conventional crystallographic methods.

4.4 Results

4.4.1 OE1.3 Shows Hinge Bending Upon Covalent Inhibition

OE1.3 is an engineered enzyme evolved through directed evolution to catalyze an ester hydrolysis reaction (Figure 4.1A) [5]. This enzyme was developed by diversifying BH32 [4], a computationally designed enzyme originally created to catalyze the Morita–Baylis–Hillman reaction (see Chapter 3). In OE1.3, catalysis is mediated by a non-canonical amino acid, ϵ N-methylhistidine, introduced at position 23, which

plays a central role in the reaction mechanism.

As reported by Hutton et al., OE1.3 undergoes a pronounced conformational change upon covalent inhibition by 2-bromoacetophenone (Figure 4.1B and 4.1C) [5]. This inhibitor closely resembles the ester substrate in size and geometry and reacts with the catalytic MeHis23 residue. Upon binding, the bromine atom is displaced and a covalent bond is formed with the N ϵ of the methylhistidine, generating a stable enzyme-inhibitor complex. This interaction was characterized using cryogenic X-ray crystallography by Hutton et. al.

Structural comparison between the apo and inhibited forms reveals that binding of this bulky ligand induces a large-scale hinge-bending motion, in which the core domain is displaced relative to the rest of the protein (Figure 4.1D). This motion appears to be a direct consequence of steric pressure exerted by the inhibitor within the active site. In the apo structure, the active site is relatively compact (Figure 4.1E), suggesting that it is not inherently optimized to accommodate the substrate in a static configuration. Instead, the enzyme appears to rely on conformational flexibility to transiently expand the active site upon ligand binding.

This observation is particularly notable from an evolutionary perspective. Rather than enlarging the active site permanently through mutation, evolution has introduced a dynamic solution, in which domain motion enables the enzyme to access catalytically relevant conformations only when needed. This capability emerges following the L10P mutation between OE1.2 and OE1.3, which has been associated with increased catalytic efficiency [5]. This suggests that the ability to undergo conformational rearrangement is functionally important and may play a role during the catalytic cycle.

To investigate whether this hinge-bending motion reflects a catalytically relevant transition along the reaction coordinate, we sought to explore time-resolved crystallographic approaches. In particular, capturing intermediate states would allow direct observation of whether similar domain movements occur during substrate turnover,

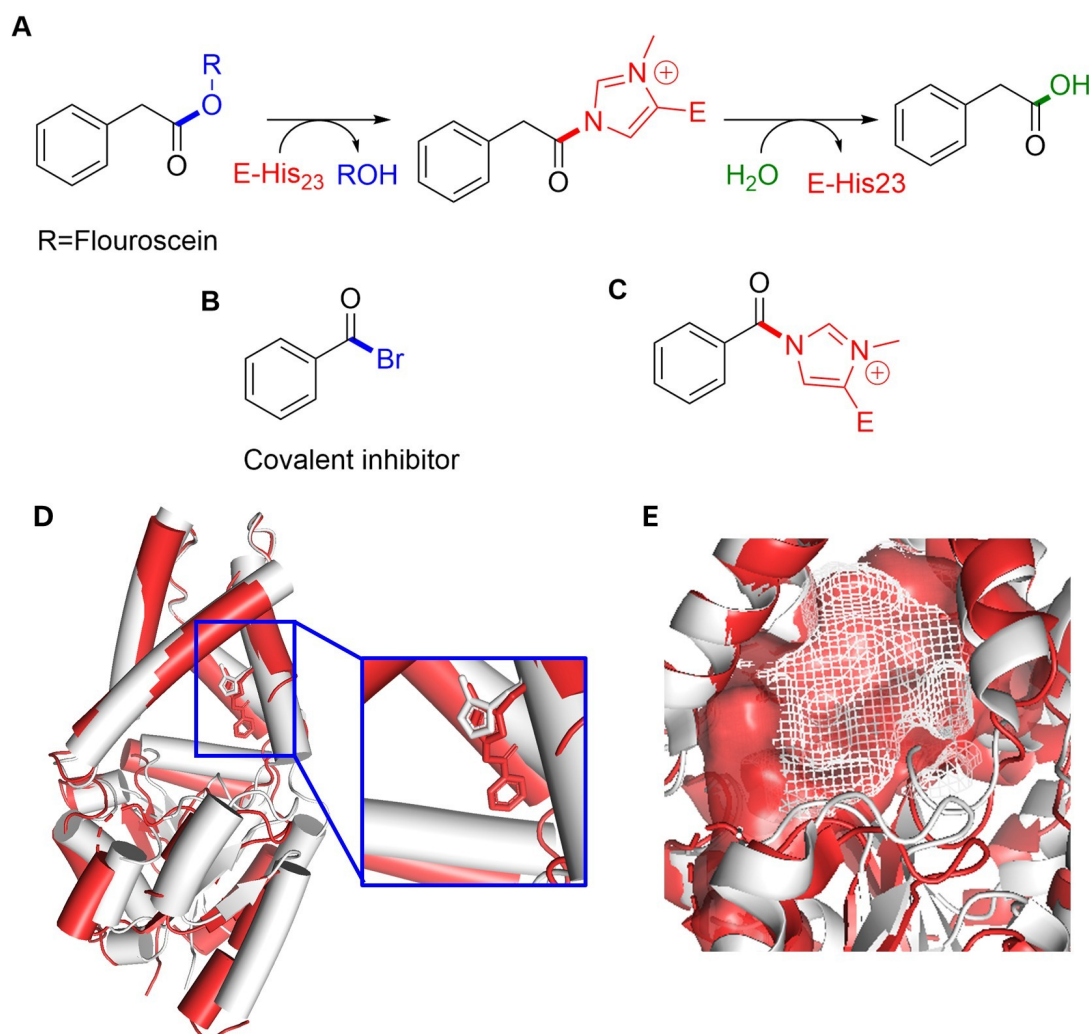


Figure 4.1: OE1.3. (A) The OE family catalyzes an ester hydrolysis reaction, which is traceable through the emission of a fluorescein group. (B) 2-bromoacetophenone acts as a covalent inhibitor and shares structural similarity with the substrate. (C) Upon reaction, the inhibitor loses the bromine atom and forms a covalent bond with the N ϵ of MeHis23. (D) Structural comparison (cryo structures from Burke et al.) shows that the bulky inhibitor induces a downward shift of the core domain (bottom domain), resulting in a hinge-like motion; the inhibited structure is shown in red and the apo structure in white. (E) The white mesh represents the shape of the active site in the absence of the inhibitor, while the red surface shows the expanded active site upon inhibition, indicating that the conformational change significantly increases the accessible volume.

rather than only upon inhibitor binding.

Table 4.1: Crystallization conditions for OE1 variants.

Condition	Temp. (K)	Buffer	pH	Protein (mg mL ⁻¹)	Precipitant #1	Precipitant #2
OE1.3	100	BisTris	5.5	5.1	18% w/v PEG 3350	0.2 M ammonium acetate
	300	HEPES	7.5	5.1	16% w/v PEG 3350	0.2 M sodium nitrate
OE1.4	100	BisTris	5.5	8.8	16% w/v PEG 3350	0.2 M ammonium acetate
	300	BisTris	5.5	8.8	16% w/v PEG 3350	0.2 M ammonium acetate

As an initial step, both cryogenic and room-temperature X-ray crystallography datasets were obtained for OE1.3 and OE1.4. OE1.4, being the more efficient variant, has not yet been structurally characterized in the presence of the inhibitor. For this reason, OE1.4 was selected as the primary target for subsequent mix-and-inject serial crystallography experiments, with the goal of probing potential substrate-induced conformational changes under non-equilibrium conditions.

Table 4.2: Crystallographic data and refinement statistics for OE1.3 and OE1.4.

Temp. (K)	OE1.3		OE1.4	
	100	300	100	300
Data Collection				
Resolution (Å)	1.60	1.80	1.60	1.90
Space group	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Unit cell				
<i>a</i> (Å)	34.23	34.40	33.98	34.45
<i>b</i> (Å)	72.46	71.73	72.16	72.09
<i>c</i> (Å)	52.48	52.68	52.11	52.72
β (°)	105.57	105.19	105.11	105.24
Chains / asym. unit	1	1	1	1
Completeness (%)	99.7	99.9	99.8	99.7
Multiplicity	6.5	3.4	6.3	3.4
I/ σ I	16.2 (0.9)	12.4 (1.0)	14.6 (0.5)	9.4 (0.9)
R _{pim}	0.021 (0.293)	0.039 (0.635)	0.031 (0.187)	0.051 (0.737)
CC _{1/2}	1.000 (0.864)	0.998 (0.573)	0.998 (0.909)	0.995 (0.518)
Wilson B-factor (Å ²)	20.69	29.86	23.29	33.25
Unique reflections	32489	22907	31995	193636
Refinement				
R _{work} /R _{free}	0.17/0.21	0.16/0.19	0.18/0.21	0.16/0.20
No. atoms				
Protein	4602	4376	4508	4138
Ligand	92	80	110	48
Water	690	250	665	205
RMSD				
Bond lengths (Å)	0.014	0.012	0.005	0.012
Bond angles (°)	1.232	1.071	0.622	1.060
MolProbity statistics				
Ramachandran favored (%)	99.02	99.87	96.90	98.45
Rotamer outliers (%)	3.48	3.05	2.89	2.31
Clashscore	5.90	7.80	6.45	7.10

Values in parentheses correspond to the highest-resolution shell.

4.4.2 Apo X-ray Crystallography Structure Shows High Similarity Between OE1.3 and OE1.4

Initial crystallization screening was performed using approximately 800 conditions from commercial NexTal screens. Multiple hits were identified for OE1.3, which were subsequently optimized to obtain crystals suitable for diffraction. The optimized conditions obtained for OE1.3 were also directly applicable to OE1.4, producing crystals of similar morphology and quality (Figure 4.2A). This is consistent with the high similarity between the two variants and allowed us to use the same crystallization strategy for both systems. Details of the crystallization conditions are summarized in Table 4.1.

Using these conditions, diffraction-quality crystals were obtained for both OE1.3 and OE1.4, and datasets were collected at both cryogenic (100 K) and room temperature (300 K). Structures were solved and aligned to assess their structural similarity. As shown in Figure 4.2B, OE1.3 (light green) and OE1.4 (forest green) are highly similar, with an RMSD of 0.176 Å. This confirms that OE1.4 provides a structurally comparable system to OE1.3, while also being a more efficient variant, making it a suitable candidate for further Mix-and-Inject experiments. Data collection and refinement statistics are provided in Table 4.2.

While the overall structures remain highly similar, comparison between cryogenic and room-temperature datasets reveals localized differences that are most pronounced in the flexible linker region connecting the two domains (Figure 4.2C). This observation suggests that the cryogenic structures, such as those reported by Hutton et al., may be influenced by cryocooling artifacts and may not fully capture the native conformational behavior of the enzyme. Notably, this flexible region coincides with the location of the L10P mutation introduced between OE1.2 and OE1.3, which has been associated with the emergence of the hinge-bending motion observed in this enzyme series. The fact that this same region shows the largest deviation upon

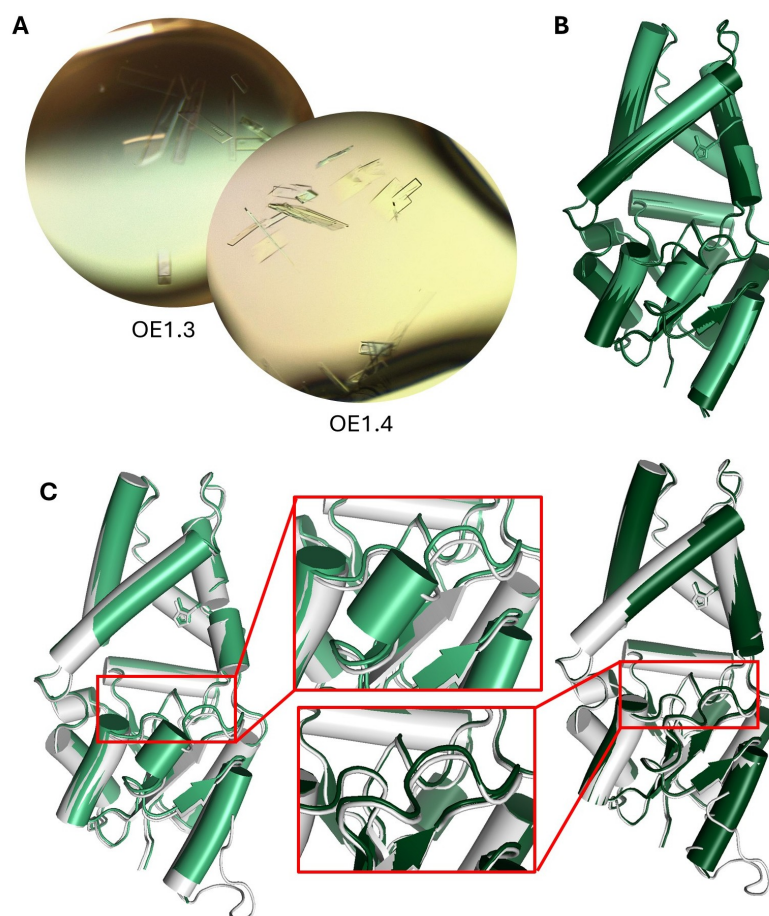


Figure 4.2: (A) Representative crystals of OE1.3 and OE1.4 obtained from optimized crystallization conditions derived from NexTal screening. (B) Structural alignment of OE1.3 (light green) and OE1.4 (forest green), showing a high degree of similarity ($\text{RMSD} = 0.176 \text{ \AA}$). (C) Comparison of cryogenic and room-temperature structures highlights conformational differences in the flexible linker region connecting the two domains, indicating the presence of cryocooling-induced artifacts.

cryocooling further supports the idea that the mutation has introduced increased conformational flexibility. This flexibility is therefore likely to be intrinsic to the system and may underlie the enzyme's ability to undergo domain movement during catalysis.

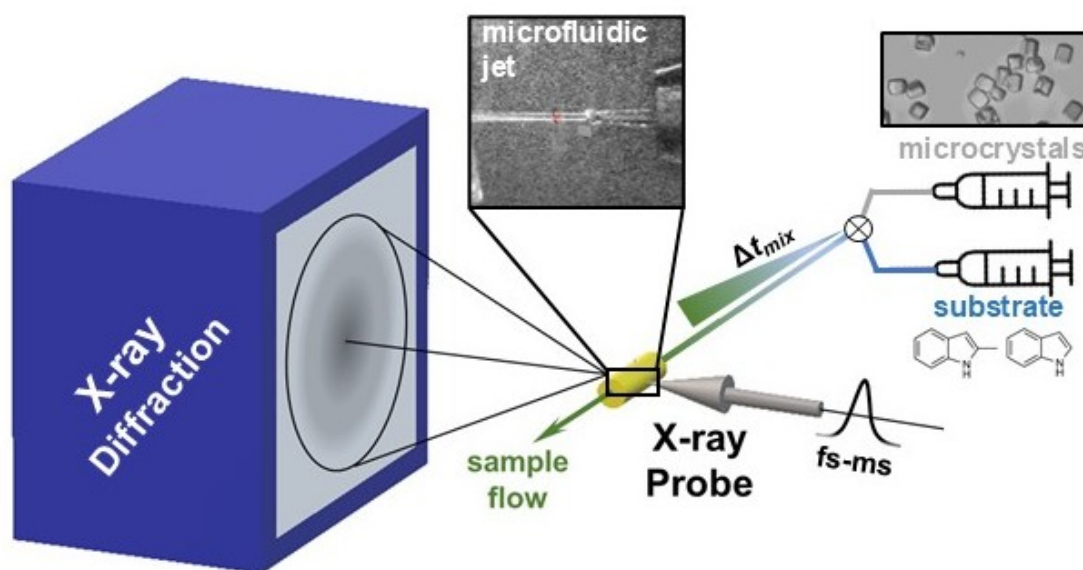


Figure 4.3: Schematic demonstration of the mix and inject serial crystallography experimental setting. Data is collected from a slurry of microcrystals at certain time points after being mixed with the substrate.

4.4.3 Synchrotron Mix-and-Inject Serial Crystallography Can Probe the Catalytic Relevance of This Motion

Mix-and-inject serial X-ray crystallography [167] (Figure 4.3) presents a powerful approach to directly probe the role of this conformational transition during catalysis. By capturing non-equilibrium states such as enzyme–substrate, enzyme–intermediate, and enzyme–product complexes, these experiments can reveal whether the observed

hinge-bending motion is coupled to specific steps along the reaction coordinate.

To assess the feasibility of such experiments, batch microcrystallization [3] of OE1.4 was performed under optimized conditions, producing a high density of microcrystals suitable for time-resolved studies. These results demonstrate that OE1.4 can be reliably prepared in a form compatible with mix-and-inject experiments.

This establishes a strong experimental foundation for future work, where time-resolved datasets can be used to directly test whether the observed domain motion is populated during catalysis. Such experiments will provide deeper insight into whether substrate-induced conformational changes are an evolved and functionally relevant component of catalysis in the OE1 family.

4.4.4 Future Directions

The next step in this work is to perform synchrotron-based mix-and-inject serial crystallography experiments on OE1.4 to capture time-resolved structural snapshots along the reaction coordinate. By collecting datasets at defined time points after mixing the substrate with the microcrystals, it will be possible to observe the progression from enzyme–substrate to intermediate and product states under non-equilibrium conditions.

These datasets can then be modeled and compared to determine whether the hinge-bending motion observed upon covalent inhibition is also populated during catalysis. Establishing this connection would provide direct evidence for the functional role of this conformational transition and offer deeper insight into how structural dynamics contribute to catalytic efficiency in the OE1 enzyme family.

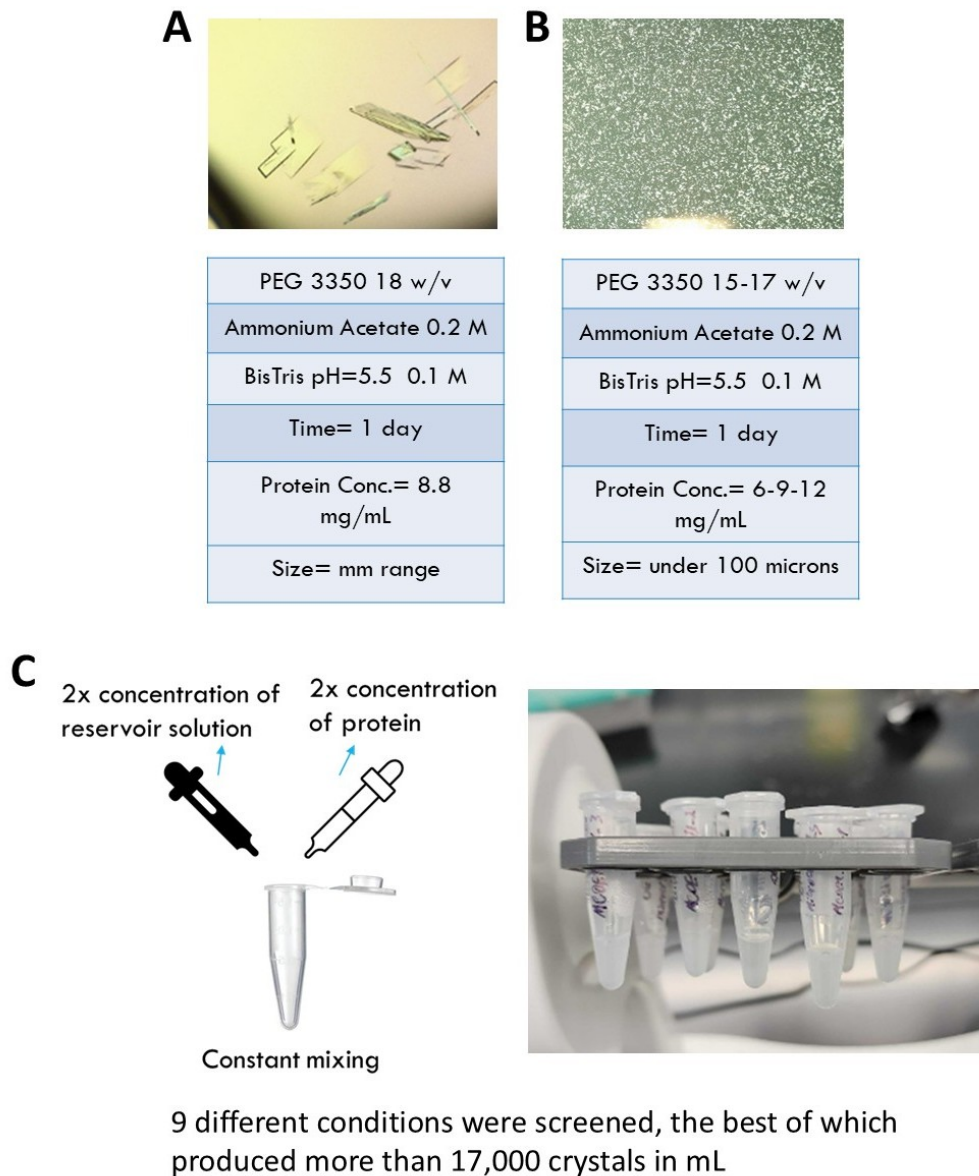


Figure 4.4: Preparation of OE1.4 microcrystals for time-resolved crystallography experiments. (A) Optimized crystallization condition producing large single crystals of OE1.4. (B) Modification of the same condition to favor nucleation over growth, resulting in a dense population of microcrystals. (C) Schematic and experimental setup of batch crystallization [3], in which concentrated protein and reservoir solutions are rapidly mixed to promote uniform microcrystal formation. This approach yielded microcrystal densities exceeding 17,000 crystals per mL, suitable for mix-and-inject serial crystallography experiments.

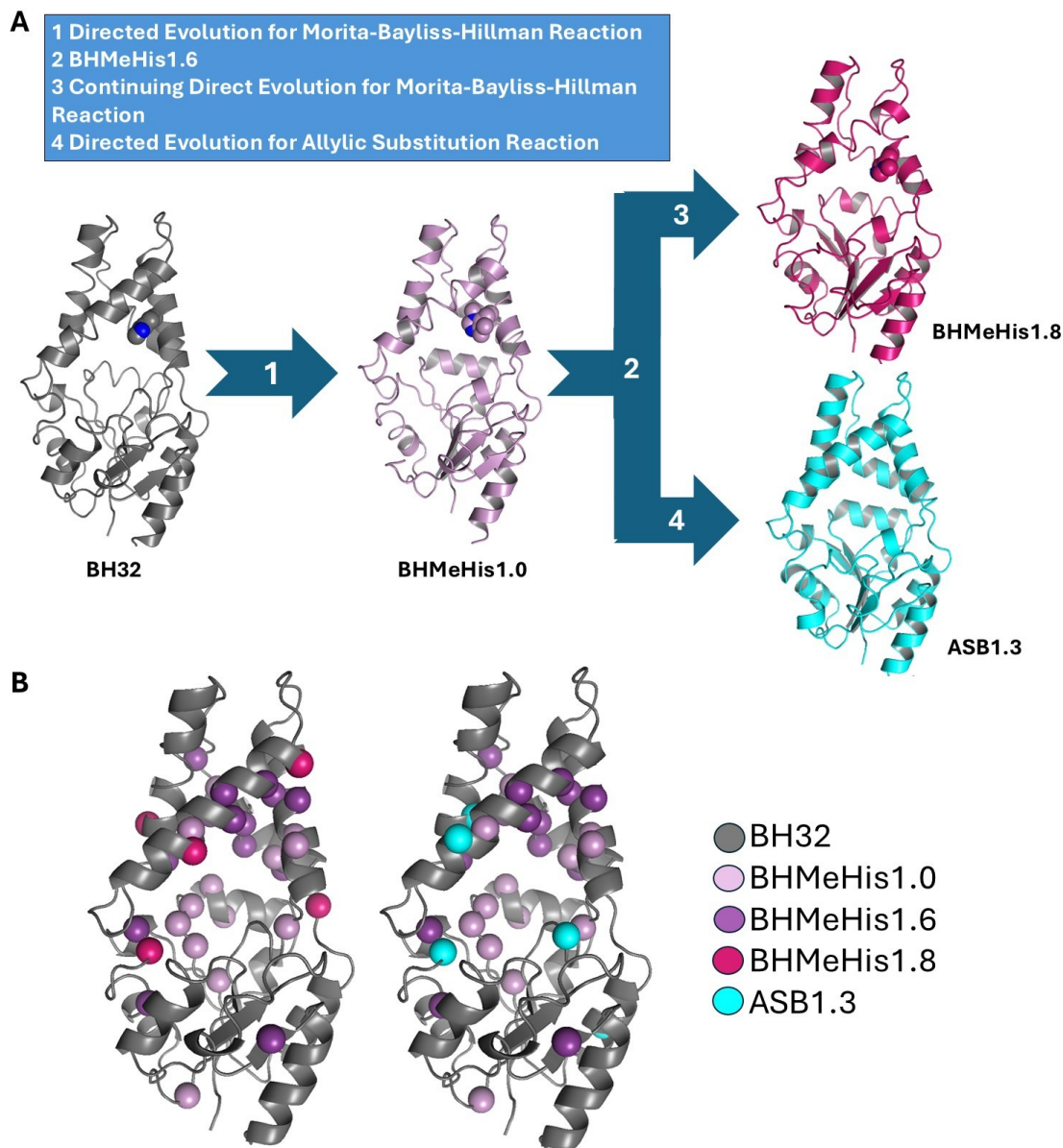


Figure 4.5: Evolutionary divergence from a common scaffold and distribution of mutations across the protein structure. (A) Directed evolution starting from BH32 toward the Morita–Baylis–Hillman (MBH) reaction produces BH_{MeHis}1.0, followed by continued evolution to BH_{MeHis}1.8. A parallel evolutionary trajectory from BH_{MeHis}1.6 (ASB1.0) leads to ASB1.3 for allylic substitution (AS), illustrating divergence from a shared intermediate scaffold. (B) C_α positions of mutations accumulated during evolution mapped onto the BH32 scaffold. Mutations from BH_{MeHis}1.0, BH_{MeHis}1.6, BH_{MeHis}1.8, and ASB1.3 are shown in different colors, illustrating how distinct evolutionary pathways redistribute mutations across the structure while maintaining a common fold.

4.4.5 ASB1.3 is an Engineered Branch of the BH_{MeHis} Lineage with Enhanced Allylic Substitution Activity

To investigate structural changes associated with catalytic progression in descendants of BH32 in a simpler system, we focused on ASB1.3, an evolved variant from the BH_{MeHis} enzyme lineage. This enzyme proceeds through its reaction very slowly, enabling capture of non-equilibrium intermediates using traditional single-crystal X-ray crystallography and soaking experiments.

ASB1.3 originates from BH_{MeHis}1.6, a precursor to BH_{MeHis}1.8 in the evolutionary trajectory (Chapter 3, Figure 4.5). While the BH_{MeHis} lineage was evolved to enhance Morita–Baylis–Hillman (MBH) activity, ASB1.3 represents a distinct evolutionary branch selected for improved allylic substitution activity.

Directed evolution of BH_{MeHis}1.6 (referred to as ASB1.0 in allylic substitution studies) introduced a set of mutations that resulted in ASB1.3, with enhanced catalytic efficiency for allylic substitution reactions. As shown in Figure 4.5B, these mutations are distributed across the protein scaffold and differ from those accumulated along the pathway leading to BH_{MeHis}1.8, highlighting the divergence between MBH-optimized and allylic substitution-optimized evolutionary trajectories.

The allylic substitution reaction catalyzed by ASB1.3 (Figure 4.6C) begins with nucleophilic attack by the catalytic methylhistidine (MeHis) on the β -position of an α, β -unsaturated ketone substrate. This step forms a covalent intermediate and induces delocalization of the π -system, facilitating release of the p-nitrophenol (PNP) leaving group. The release of PNP generates a characteristic absorbance at 400 nm, enabling convenient monitoring of the reaction.

In the subsequent step, the second substrate (nucleophile a) attacks the allylic position, restoring the conjugated system and yielding the final product. This two-step mechanism provides a clear framework for capturing intermediate states during catalysis.

Table 4.3: Crystallization conditions for ASB1.3 variants.

Condition	Temp.(K)	Buffer	pH	Protein (mg mL ⁻¹)	Precipitant
Apo	300	MMT	6.0	20	27% w/v PEG 1500
Substrate 1	300	MMT	5.5	20	27% w/v PEG 1500

For in crystallo experiments, substrate 1 (Figure 4.6A) and nucleophile a (Figure 4.6B) were selected to balance reactivity and temporal resolution, enabling controlled progression of the reaction within the crystal.

4.4.6 Time-Resolved Soaking to Capture Enzyme–Substrate Intermediate States

To probe structural changes along the catalytic pathway, we employ a time-resolved soaking strategy using ASB1.3, which operates on a sufficiently slow timescale to allow capture of intermediate states using conventional single-crystal X-ray crystallography. This approach enables the observation of structural snapshots along the reaction coordinate without the need for rapid mixing or serial crystallography methods [42, 72].

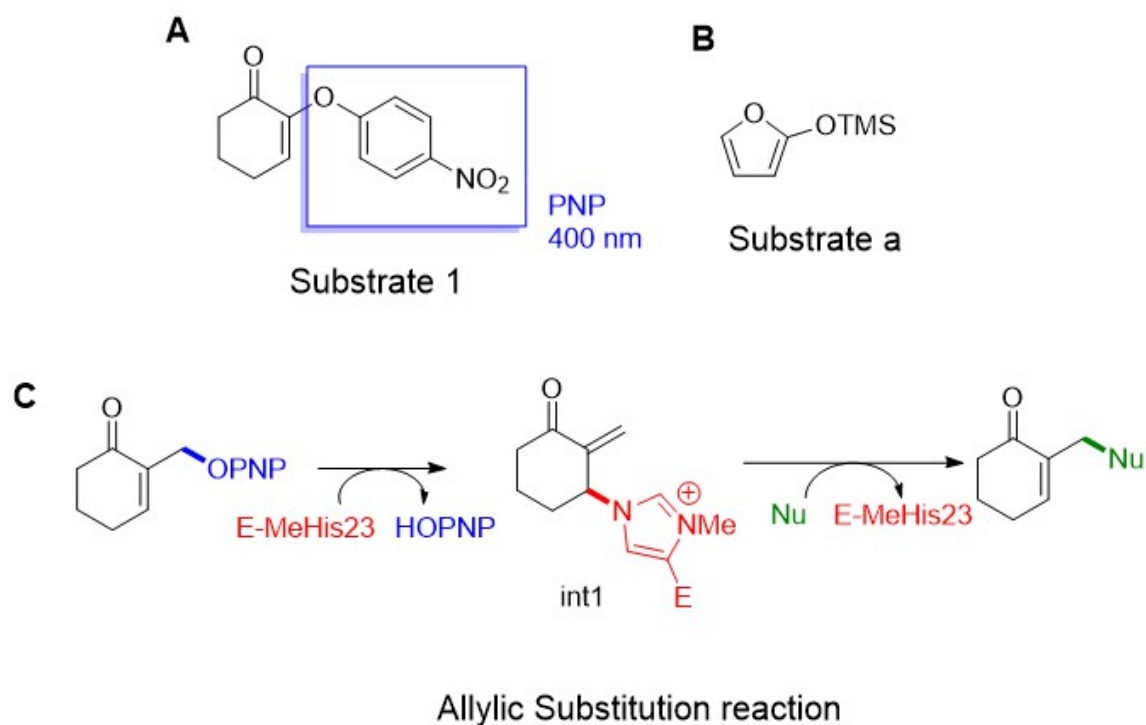


Figure 4.6: Substrates and reaction mechanism for ASB1.3-catalyzed allylic substitution. (A) Structure of substrate 1, containing a p-nitrophenol (PNP) leaving group that enables spectroscopic monitoring at 400 nm. (B) Structure of nucleophile a used as the second substrate. (C) Proposed reaction mechanism, in which the catalytic methylhistidine (E-MeHis23) forms a covalent intermediate with substrate 1, followed by nucleophilic attack by substrate a to yield the final product.

Table 4.4: Crystallographic data and refinement statistics for ASB1.3 K39A

	Apo	Bound
Data collection		
Temperature (K)	300	300
Resolution (Å)	2.20	2.40
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)		
<i>a, b, c</i>	37.42, 97.75, 124.05	37.48, 97.90, 124.11
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Chains per asym. unit	2	2
R _{pim}	0.035 (0.188)	0.029 (0.330)
CC _{1/2}	0.998 (0.960)	0.999 (0.902)
I/ σ I	11.7 (1.0)	13.5 (1.0)
Completeness (%)	100.0	100.0
Multiplicity	6.3	6.3
Wilson B-factor (Å ²)	42.49	53.51
Unique reflections	23975	18671
Refinement		
R _{work} /R _{free}	0.1657 / 0.2317	0.1753 / 0.2463
No. atoms		
Protein	3834	3809
Ligand	22	44
Water	58	21
Avg. B-factors (Å²)		
Protein	62.26	81.58

Continued on next page

Table 4.4 (continued)

	Apo	Bound
Ligand	74.88	79.99
Water	52.75	62.78
RMSD		
Bond lengths (Å)	0.009	0.010
Bond angles (°)	0.99	1.14
MolProbity statistics		
Ramachandran (favored %)	96.83	95.08
Rotamer outliers (%)	8.42	12.25
Clashscore	4.65	8.13

Values in parentheses correspond to the highest-resolution shell.

In the absence of a suitable nucleophilic coupling partner, the covalent intermediate formed after reaction with substrate 1 is expected to persist without progressing to product formation. This behavior has been previously observed for BH_{MeHis}1.6, which corresponds to ASB1.0 in the allylic substitution system, in recent unpublished work from the collaboration of Dr. Anthony Green and Dr. Zak Birch-Price. Importantly, ASB1.3 is expected to exhibit similar behavior, suggesting that this intermediate-trapping strategy is applicable across this branch of the enzyme lineage.

Based on this framework, the experimental design involves first soaking enzyme crystals (macrocrystals, in the 100 μ m range) in substrate 1 to accumulate the covalent intermediate within the crystal. Single-crystal X-ray diffraction data can then be collected to capture this trapped state. In a second step, the nucleophilic substrate (substrate a) is introduced, and additional datasets are collected at defined time intervals following mixing (Figure 4.7).

This approach allows progression along the reaction coordinate to be followed structurally over time, with each dataset representing a distinct time point. Data collection can be performed either at different positions within a single crystal or across multiple crystals prepared under identical conditions, enabling flexibility in experimental implementation.

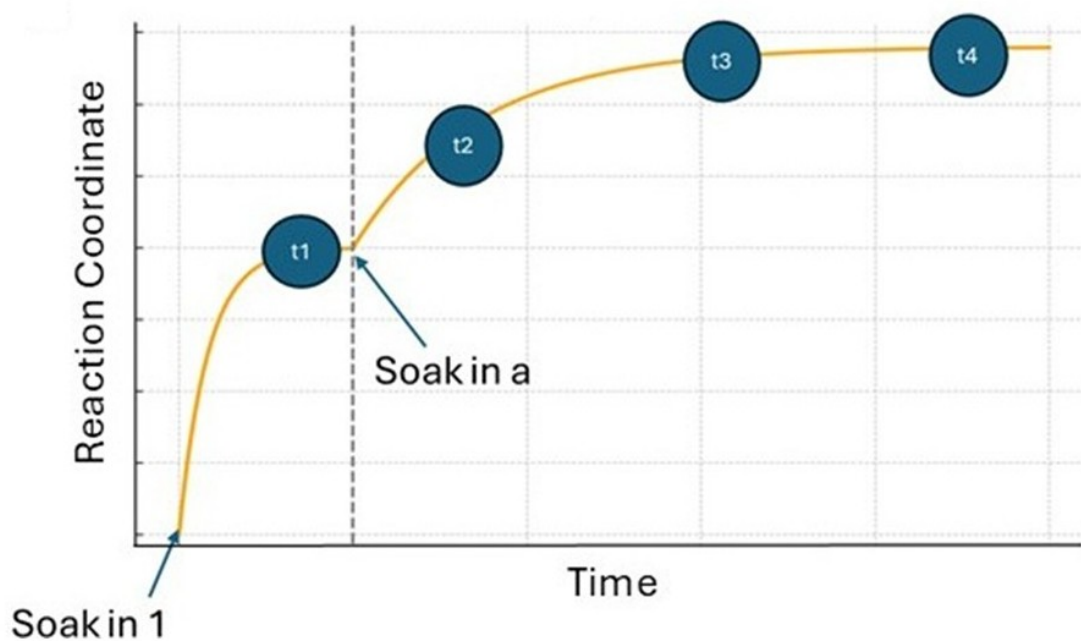


Figure 4.7: Schematic representation of the experimental workflow. Crystals are first soaked in substrate 1 to generate the intermediate state, followed by introduction of substrate a. Structural data are then collected at defined time points (t_1 – t_4) to capture progression along the reaction coordinate.

4.4.7 Crystal Structures of Apo ASB1.3 and Substrate 1 Complex

In the first stage of this study, the X-ray crystallographic structure of the apo form of ASB1.3 K39A was determined at 300 K and used as the reference state for subsequent structural analysis. The K39A mutation was introduced to improve crystallization behavior and was verified to have a negligible effect on catalytic function. The apo structure was solved to a resolution of 2.2 Å.

Crystals of ASB1.3 suitable for diffraction were obtained under the conditions listed in Table 4.3, with representative crystal morphology shown in Figure 4.8. Using these conditions, a second dataset was obtained for crystals soaked with substrate 1, resulting in a structure solved to 2.4 Å resolution. The apo structure serves as the

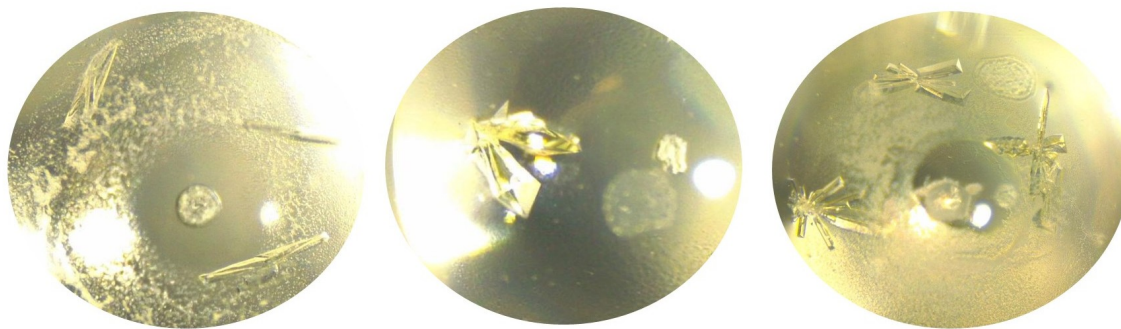


Figure 4.8: Examples of ASB1.3 crystal morphology

baseline for comparison with the substrate-bound state, enabling direct analysis of ligand-induced structural changes within the same enzyme scaffold. Together, these datasets establish the foundation for subsequent time-resolved experiments aimed at capturing intermediate states along the reaction coordinate.

4.4.8 Structural Comparison Between BH_{MeHis}1.8 and ASB1.3 Reveals Selective Stabilization of Conformational States

In BH_{MeHis}1.0, the shared ancestor of BH_{MeHis}1.8 and ASB1.3 (Figure 4.5), the regions spanning residues 180–195 and 200–215 are disordered or highly dynamic, providing a set of energetically accessible conformational substates. These regions therefore represent a flexible structural landscape from which evolution can selectively stabilize specific conformations.

X-ray crystallographic structures show that BH_{MeHis}1.8 and ASB1.3 adopt distinct structural solutions from this shared ensemble. As shown in Figure 4.9A, these regions, which are helical in BH32, become flexible loops in BH_{MeHis}1.8. In contrast, in ASB1.3, the region spanning residues 200–210 partially regains its helical conformation, while the 180–195 region remains flexible.

One contributing factor to this behavior may be the G11S mutation in ASB1.3, which introduces a network of hydrogen bonds that stabilizes the helical conformation in the 200–210 region (Figure 4.9B). In addition, ASB1.3 retains a π -system at position 155, similar to BH_{MeHis}1.8, which appears to stabilize the 180–195 loop through hydrophobic interactions with A186 (Figure 4.9C).

In BH32, residue 155 is positioned within a helix and remains distant from A186. However, through local conformational changes and increased disorder in BH_{MeHis}1.0, these residues become proximal, enabling new stabilizing interactions. These observations suggest that evolution selectively stabilizes distinct conformations from an initially heterogeneous ensemble, leading to different structural solutions that support catalytic function in each variant.

4.4.9 ASB1.3 Has an Inherently Open Active Site

The first observation from the X-ray structures of ASB1.3, in both the apo form and in complex with substrate 1, is the minimal structural difference between the

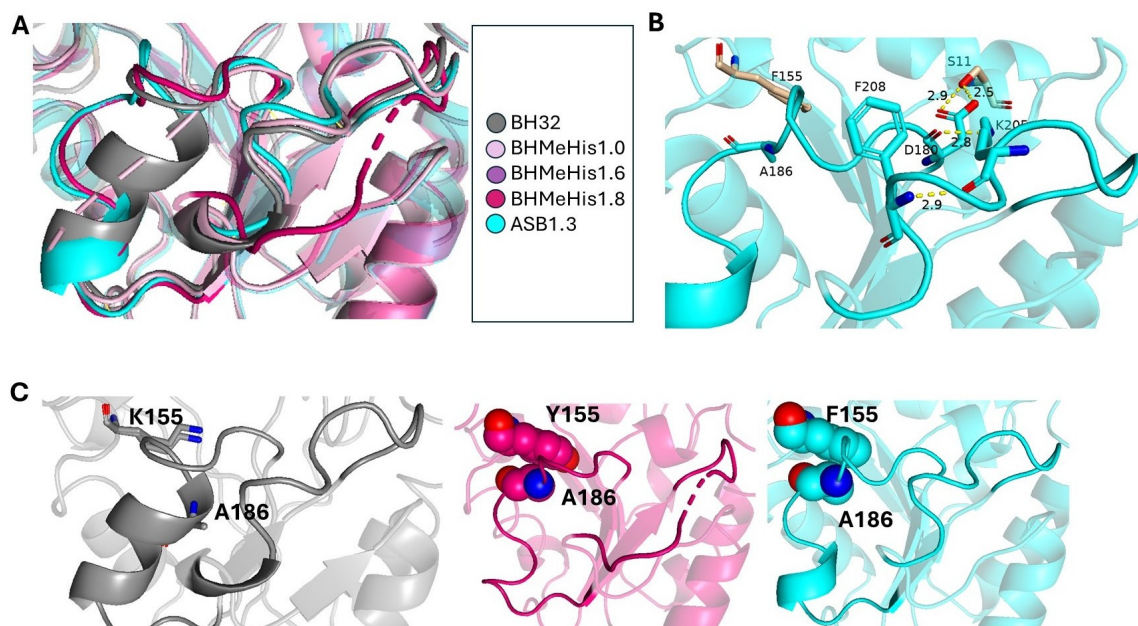


Figure 4.9: Selective stabilization of distinct conformations in BH_{MeHis}1.8 and ASB1.3 from a shared dynamic ancestor. (A) Structural comparison of regions 180–195 (left) and 200–210 (right) across BH32, BH_{MeHis}1.8, and ASB1.3. In BH32, both regions adopt well-defined helical conformations. In BH_{MeHis}1.8, these regions become flexible loops, reflecting increased conformational heterogeneity. In ASB1.3, while the 180–195 region remains flexible, the 200–210 region partially regains its helical structure, indicating selective stabilization of a specific conformation. (B) The G11S mutation in ASB1.3 contributes to the reformation of the helical structure in the 200–210 region by introducing a network of stabilizing hydrogen bonds. (C) A conserved π -system at position 155 in ASB1.3 and BH_{MeHis}1.8 enables hydrophobic interaction with A186, stabilizing the 180–195 loop. In BH32, residue 155 is positioned within a helix and remains distant from A186; this interaction becomes possible only after local conformational rearrangement and increased flexibility in the intermediate BH_{MeHis}1.0 state. Together, these observations highlight how evolution selects and stabilizes distinct conformations from an initially heterogeneous structural ensemble.

two states (Figure 4.11). The observed differences are largely limited to flexible loop regions, with no significant rearrangement of the overall fold or active site architecture. This indicates that binding of the first substrate does not induce major conformational changes, suggesting that the active site is already in a conformation compatible with ligand binding. In other words, ASB1.3 appears to be preorganized, with an inherently accessible active site that does not require substantial structural adaptation upon substrate entry.

Both the apo and substrate-bound structures crystallize in the $P2_12_12_1$ space group with two chains in the asymmetric unit. Although these chains arise from non-crystallographic symmetry, chains A and B within each structure display noticeable differences in conformation and apparent flexibility, likely influenced by crystal packing effects. However, when comparing equivalent chains between the apo and substrate-bound states, the overall structural differences remain minimal.

Alignment of chain A from the apo structure (cyan) with chain A from the substrate-bound structure (orange) further highlights this similarity (Figure 4.11). Only small local deviations are observed, primarily in flexible loop regions and a minor helical shift behind the active site. The presence of this helical movement in only one chain of the asymmetric unit suggests that it may be influenced by crystal packing rather than representing a functionally relevant conformational change.

Consistent with these structural observations, comparative analysis of B-factor Z-scores between the apo and bound structures does not reveal significant differences, indicating that the overall dynamic behavior of the enzyme remains largely unchanged upon binding of the first substrate.

To further investigate this behavior, we focused on chain A, which appears to be less affected by crystal packing interactions (Figure 4.10), and compared it with BH32 and BH_{MeHis}1.8. Alignment of these structures based on the cap domain reveals that ASB1.3, in both apo and substrate-bound forms, adopts a hinge-bent conformation similar to that observed in BH32.8 (Chapter 3) (Figure 4.13A) and in inhibitor-

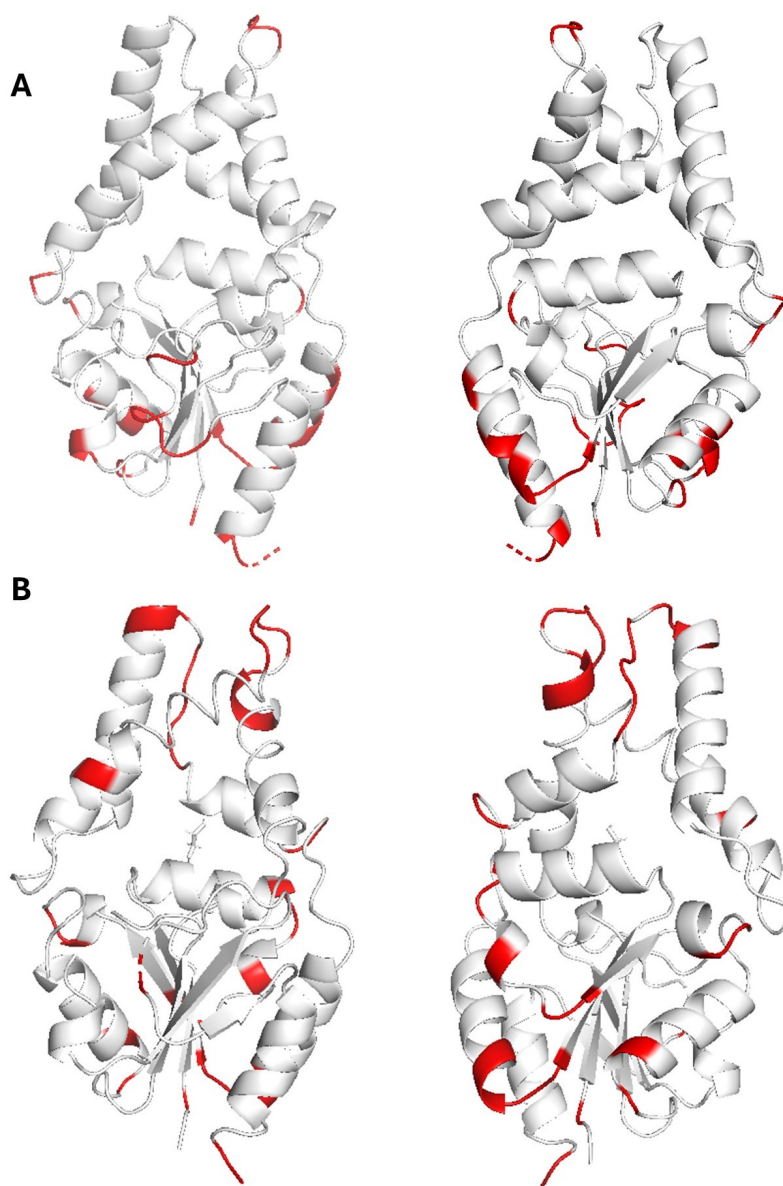


Figure 4.10: Crystal contacts (red) in ASB1.3 apo (white) shows that chain A has been less affected, and will be used for the rest of the analysis (A) Chain A front (on the left) and back (on the right) view, (B) Chain B front (on the left) and back (on the right) view

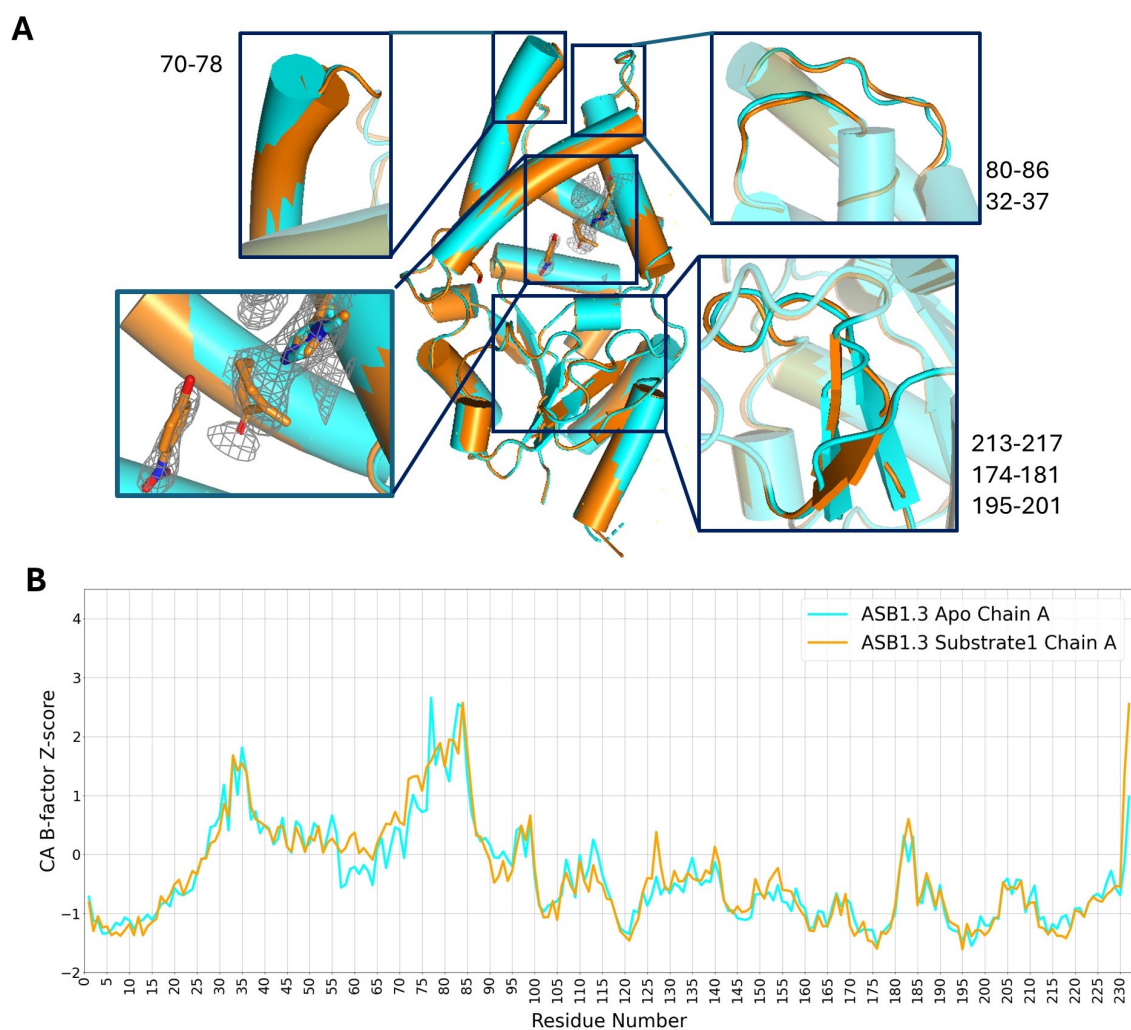


Figure 4.11: Minimal structural and dynamic changes upon substrate binding in ASB1.3. (A) Structural alignment of chain A from the apo structure (cyan) and the substrate-bound structure (orange) shows a high degree of similarity between the two states. Differences are primarily localized to flexible loop regions (highlighted), with no significant rearrangement of the overall fold or active site architecture. The inset shows electron density corresponding to the bound substrate in the active site, confirming its position without requiring substantial backbone reorganization. A small helical displacement behind the active site is observed, but its presence in only one chain of the asymmetric unit suggests that it is influenced by crystal packing rather than being functionally relevant. (B) C α B-factor Z-score profiles for the apo (cyan) and substrate-bound (orange) structures largely overlap across the sequence, indicating that substrate binding does not significantly alter global or local backbone dynamics.

bound OE1.3. Notably, this hinge opening appears to be slightly more pronounced than that observed in BH32.8 and OE1.3 inhibited (Figure 4.12). Importantly, this conformational state is present in both apo and ligand-bound structures, indicating that it is an intrinsic feature of ASB1.3 rather than a response to substrate binding.

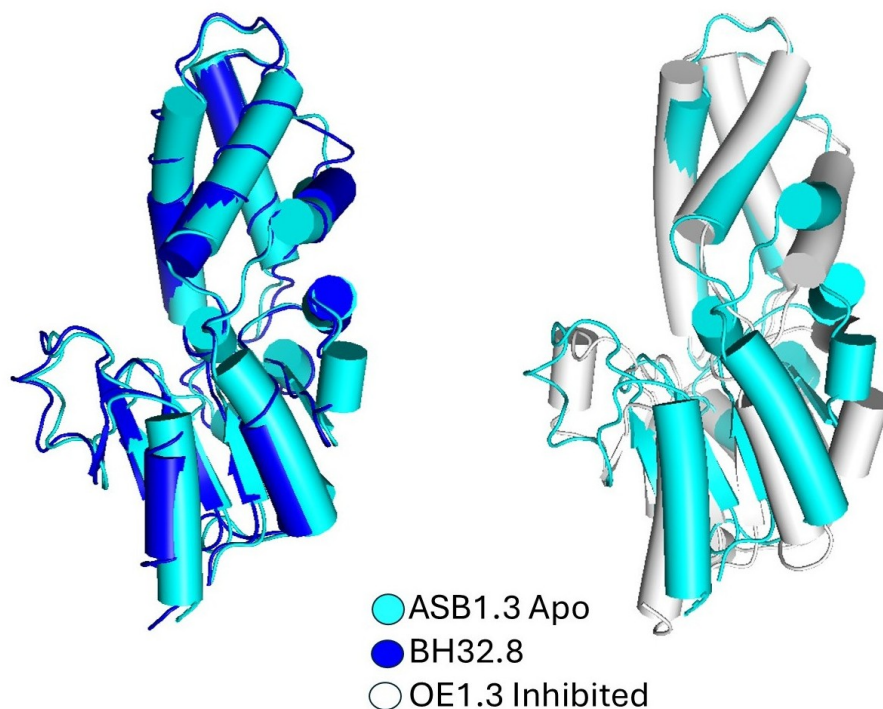


Figure 4.12: Comparison of hinge-bending conformations across related enzyme variants. ASB1.3 adopts a more open hinge-bent conformation compared to BH32.8 (blue). While OE1.3 in the inhibitor-bound state also exhibits hinge bending, its motion includes an additional twisting component that is not observed in ASB1.3. When considering the hinge-bending component alone, ASB1.3 appears to adopt a more open conformation than OE1.3 in its inhibited state, highlighting subtle but important differences in the nature of conformational changes across these systems.

A second observation comes from analysis of the active site volume using KVFinder [149, 150]. The calculated cavity for the substrate-bound structure appears more ex-

panded than that of the apo structure (Figure 4.13B), despite the absence of any obvious large-scale structural rearrangement. This suggests that relatively small, localized changes are sufficient to modulate the shape and accessibility of the active site. In particular, residues located at the entrance of the active site, such as Lys184 and Arg50, adopt slightly different conformations in the bound state, appearing to move into more occluding positions (Figure 4.13C). This raises the possibility that these residues may play a role in gating substrate entry or stabilizing transient states during binding.

However, comparison of the core active site residues between the apo and bound structures shows a high degree of structural conservation and alignment (Figure 4.13D), indicating strong preorganization of the catalytic environment. Substrate binding occurs through a covalent interaction between the N ϵ of the catalytic residue and the C β of the electrophilic substrate, while the released p-nitrophenol (PNP) group is stabilized through a π - π interaction with Trp124.

Together, these observations support a model in which ASB1.3 combines an inherently open and accessible active site with a preorganized catalytic core, allowing substrate binding to occur with minimal structural rearrangement while still maintaining precise positioning of key functional groups required for catalysis.

4.5 Future Directions

The time-resolved crystallographic study of ASB1.3 represents an ongoing effort to capture non-equilibrium states along an enzymatic reaction coordinate. Initial results demonstrate that this system is well-suited for mixing-based approaches due to its relatively slow reaction kinetics, which allow intermediate states to accumulate within the crystal.

However, the current dataset is limited in temporal resolution, and additional time points following substrate mixing are needed to more fully characterize the

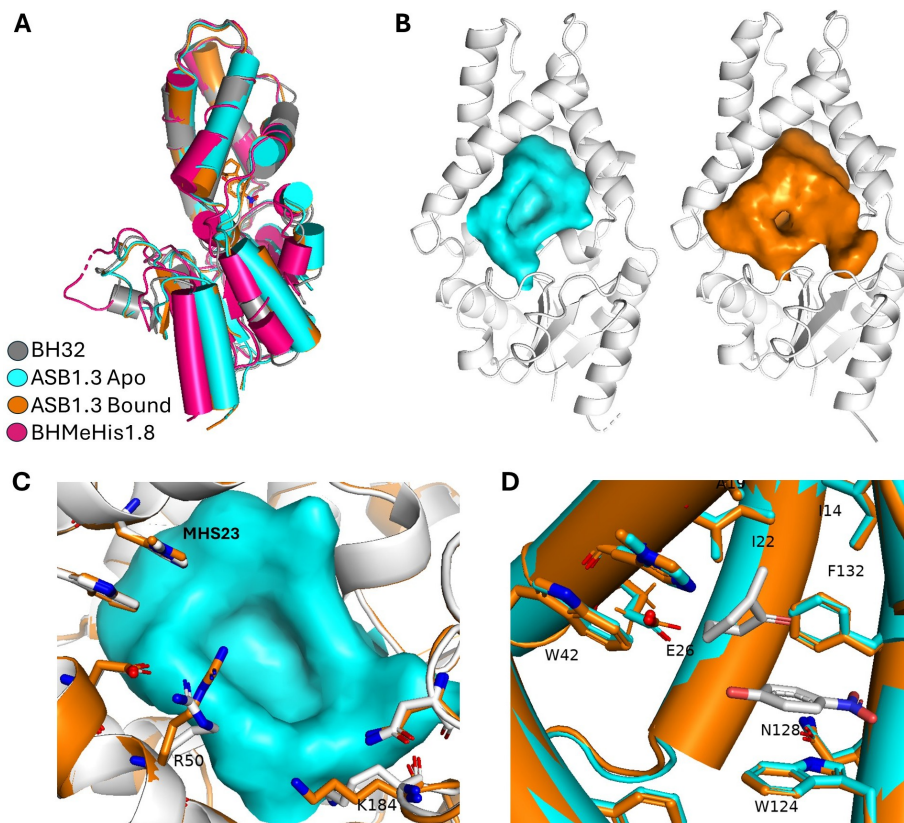


Figure 4.13: Structural features of ASB1.3 reveal an inherently open and preorganized active site. (A) Structural alignment of BH32 (gray), BH_{Me}His1.8 (magenta), ASB1.3 apo (cyan), and ASB1.3 bound to substrate 1 (orange) anchored on cap domain, showing that ASB1.3 adopts a hinge-bent conformation similar to previously observed intermediates in related systems. Both apo and bound forms overlap closely, indicating that this conformation is intrinsic and not induced by substrate binding. (B) Active site cavity calculated using KVFinder for ASB1.3 apo (cyan) and substrate-bound (orange) structures. The bound state shows a more expanded cavity despite minimal global structural differences. (C) Close-up view of the active site entrance highlighting residues such as Arg50 and Lys184, which adopt slightly different conformations between apo and bound states, potentially contributing to substrate access or gating. (D) Overlay of key active site residues in apo and bound structures, showing strong structural conservation and preorganization of the catalytic environment. The ligands are shown in white.

progression of the reaction. In particular, obtaining structures at intermediate time scales between the existing measurements, as well as extending to longer time points, would provide a more continuous view of the structural changes occurring during catalysis.

Furthermore, while initial soaking experiments have provided insight into substrate binding, obtaining structures with the second substrate fully incorporated remains essential for understanding the complete catalytic mechanism. These structures would enable direct visualization of how the active site accommodates both substrates simultaneously and how the enzyme reorganizes during the chemical step.

Future work will therefore focus on expanding the time-resolved dataset, improving ligand occupancy, and refining structural models to better capture partial states and conformational heterogeneity. Together, these efforts aim to provide a more detailed and mechanistic understanding of how enzyme structure and dynamics evolve along the reaction coordinate.

4.6 Methods

4.6.1 Protein Expression and Purification

For the expression of OE1.3, OE1.4 and ASB1.3 K39A containing methylhistidine (Me-His), *E. coli* DH5 α cells containing pEVOL_PylRSMe-His/tRNA_{CUA} were transformed with the appropriate pBbE8k_ASB1.3 K39A-His construct. Single colonies of freshly transformed cells were cultured for 18 hours in 25 mL LB medium supplemented with 35 μ g/mL kanamycin and 25 μ g/mL chloramphenicol at 37 °C. Starter cultures (10 mL) were used to inoculate 1 L 2 $\times$ YT medium containing 35 μ g/mL kanamycin and 25 μ g/mL chloramphenicol. Cultures were grown at 37 °C, 200 rpm until they reached an OD₆₀₀ of $\sim$ 0.7. To incorporate Me-His, 10 mM Me-His was added to the culture prior to induction, and protein expression was

induced by the addition of L-arabinose (3.33 mM final concentration). Following induction, cultures were incubated for 20 hours at 25 °C, and cells were harvested by centrifugation (4,500g for 40 minutes).

Pelleted cells were resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, pH 7.5, 20 mM imidazole) and lysed by sonication. Cell lysates were cleared by centrifugation (22,000g for 30 minutes), and supernatants were subjected to affinity chromatography using Ni Sepharose High Performance columns (Cytiva) where the His-tagged ASB1.3 K39A protein was eluted using 50 mM HEPES, 300 mM NaCl, pH 7.5, 250 mM imidazole via a HisTrap HP column (Cytiva) and analyzed by SDS-PAGE. For further purification, the protein was subjected to size-exclusion chromatography on a Superdex 75 column in 50 mM HEPES, 300 mM NaCl, pH 7.5. Protein concentrations were determined by UV absorption at 280 nm.

4.6.2 Crystallization

All variants were prepared at the concentration listed in Table 4.1 and 4.3. Initial crystallization trials were conducted using 96-well NeXtal crystallization screens (NeXtal Biotechnologies) and further optimized using a 15-well hanging drop format with EasyXtal crystallization plates (NeXtal). Crystallization drops were prepared by mixing 1 μ L of protein solution with 1 μ L of mother liquor, followed by sealing the drop inside a reservoir containing an additional 500 μ L of the mother liquor solution. The mother liquor solutions and the specific growth conditions that yielded the crystals used for X-ray data collection are provided in Table 4.1 and Table 4.3.

Ligand soaking experiments were performed using pre-formed ASB1.3 K39A crystals grown in the mother liquor conditions listed in Table 4.3. The same mother liquor solutions were used for preparation of soaking solutions.

Stock solutions of substrate 1 was prepared in acetonitrile (MeCN) at a concentration of 140 mM, and mixed into the soaking solutions by dilution to a final

concentration of approximately 7 mM in the mother liquor. For substrate 1, crystals were soaked for 24 hours prior to data collection.

Microcrystals were prepared using a batch crystallization approach [3] by combining 2× mother liquor with 2× protein solution, followed by rapid mixing to promote nucleation over crystal growth.

4.6.3 X-ray Data Collection and Processing

Prior to X-ray data collection, crystals were mounted in polyimide loops and sealed using a MicroRT tubing kit (MiTeGen). Single-crystal X-ray diffraction data were collected at beamline 8.3.1 of the Advanced Light Source, equipped with a Pilatus3 S 6M detector and operated at a photon energy of 11,111 eV.

Crystals were maintained at 300 K throughout data collection. Each data set was collected using a total X-ray dose of 200 kGy or less and covered a 180° wedge of reciprocal space. Multiple data sets were collected for each enzyme variant, either from different crystals or, when crystal size permitted, from distinct regions of larger crystals.

X-ray diffraction data were processed using the `xia2` (v0.5.492) pipeline [153], which performed indexing, integration, and scaling with `XDS` (v20180126) and `XS-CALC` [154], followed by merging with `pointless` as implemented in `CCP4` (v7.0.05333). For each variant, multiple individual data sets were merged to obtain the final set of reduced intensities. Resolution cutoffs were determined based on $CC_{1/2}$ and $\langle I/\sigma(I) \rangle$ values of approximately 0.5 and 0.8, respectively, in the highest-resolution shell.

The reduced diffraction data were further analyzed using `phenix.xtriage` to assess potential crystal pathologies, and no significant issues were identified. Data collection and processing statistics are summarized in Tables 4.2 and 4.4.

4.6.4 Structure Determination

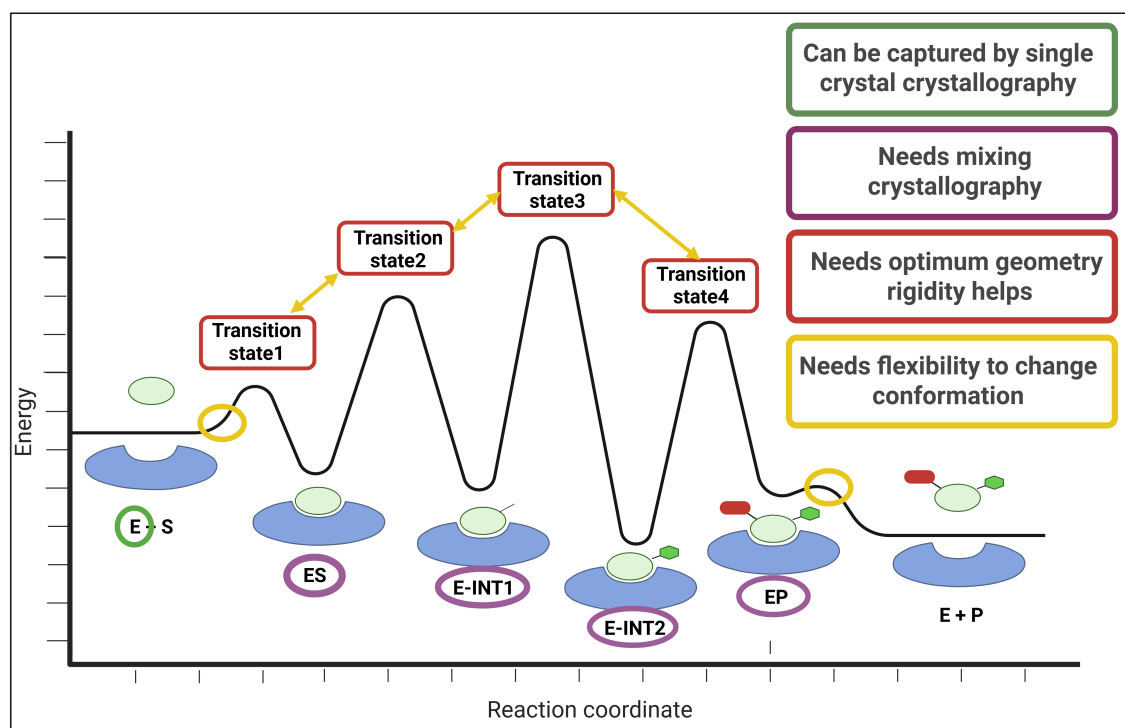
Initial phase information for the calculation of electron density maps was obtained by molecular replacement using Phaser [127], as implemented in the PHENIX suite (v1.13-2998). The 300K structure of BH_{MeHis}1.8 (refer to Chapter 2) was used as the search model for all ASB1.3 K39A data sets, and each structure contained 2 molecules in the crystallographic asymmetric unit. The cryogenic structure of BH32 alkylated with the mechanistic inhibitor 2-bromoacetophenone (PDB ID: 6Q7N) [6] was used as the search model for OE1.3 and OE1.4 data sets. Each structure contained 1 molecules in the crystallographic asymmetric unit.

Iterative cycles of manual model rebuilding and refinement were performed, including refinement of atomic positions, atomic displacement parameters, and occupancies. For consistency across data sets, R_{free} flags for all structures were copied from the ASB1.3 K39A apo data set. Refinement protocols included the use of a translation–libration–screw (TLS) model, a riding hydrogen model, and automatic weight optimization. Model building was carried out in Coot (v0.8.9.2) [129], and refinement was performed using phenix.refine within the PHENIX suite (v1.13-2998) [128, 130].

Further details regarding model building and refinement, for the final models are provided in Table 4.4 and Table 4.2.

Chapter 5

General Discussion and Conclusions



Chapter 2 Ensemble refinement reveals and enriches catalytically favorable conformations, enabling efficient enzyme design without extensive directed evolution

Chapter 3 Comparison of homologous enzymes evolved from a common ancestor reveals structural and dynamic features that govern functional selectivity and catalytic efficiency

Chapter 4 Simplified mixing experiments with X-ray crystallography and slow engineered enzymes enable capture of non-equilibrium intermediate states

5.1 Discussion

Enzymes rely heavily on their three-dimensional structure and dynamic behavior for function. While extensive work in structural biology has established how enzyme structure complements its chemistry to enable catalysis, the role of dynam-

ics introduces an additional layer of complexity. On one hand, enzymes require a degree of rigidity to maintain the precise geometry of the active site and to stabilize the transition state through preorganization, thereby lowering the energetic cost of reorganization. On the other hand, several steps in the catalytic cycle necessitate conformational flexibility, allowing the enzyme to move between different substates. For example, in multi-step reactions, distinct transition states may require different structural arrangements, and the enzyme must be able to access these conformations as the reaction progresses. This raises a fundamental question: how do enzymes balance these seemingly opposing requirements of rigidity and flexibility during catalysis?

In this thesis, we take advantage of advanced X-ray crystallography approaches, including room-temperature crystallography, mixing strategies, and ensemble refinement, together with engineered enzyme systems to investigate this balance in simplified and controlled settings. Natural enzymes have evolved over millions of years to optimize catalytic function and adapt to their environments; as a result, their structures and dynamics reflect a combination of features that are both related and unrelated to catalysis, making it difficult to directly attribute specific properties to catalytic activity.

In this context, directed evolution [81, 1, 82] provides a useful framework, not as the primary focus of the study, but as a tool to isolate and better understand structural and dynamic features that contribute to catalysis. Engineered enzymes offer an additional advantage beyond their relative simplicity. Their evolutionary trajectory begins with low catalytic efficiency and improves through directed evolution, with selection at each stage based specifically on catalytic performance. As a result, they provide a streamlined system in which structural and dynamic features can be more directly linked to function, reducing the ambiguity often present in natural enzymes.

This thesis addresses this question through three complementary approaches, each presented in a separate chapter.

First, we consider the role of conformational heterogeneity in enzyme catalysis in simple single-step kemp eliminases. While structural flexibility can be essential for function, excessive heterogeneity can be detrimental, as it disrupts the precise geometry required for optimal positioning of active-site residues. In Chapter 2, we show that incorporating crystallographic ensemble refinement into computational enzyme design enables the identification of functionally relevant backbone conformations that are already present within a low-activity enzyme. By selectively enriching these conformations through a new round of *in silico* design, catalytic efficiency can be improved without relying on directed evolution.

This approach was applied to HG3, a low-efficiency artificial Kemp eliminase, and resulted in variants that surpassed the activity of enzymes obtained through directed evolution, with the most active variant reaching a 250-fold increase in k_{cat}/K_M . Structural analysis of these improved variants showed a reduction in conformational heterogeneity and an increase in active-site preorganization, consistent with enhanced catalytic performance.

Building on this, we next examine how enzymes evolve to meet the demands of multi-step catalysis, where efficient turnover requires stabilization of multiple transition states within a single active site. In this context, flexibility is not simply permissive, but functionally necessary, as the enzyme must adopt distinct conformations to accommodate the geometric and electronic requirements of each step along the reaction coordinate.

To investigate this, we take advantage of three enzyme families that have evolved in parallel from a common ancestor to catalyze distinct multi-step organic reactions. Because these enzymes are highly homologous and share the same structural scaffold, they provide a unique system to directly compare how different catalytic requirements shape structure and dynamics within a common framework. This enables us to dissect how evolution tunes the conformational landscape to support multiple transition states rather than a single optimized geometry.

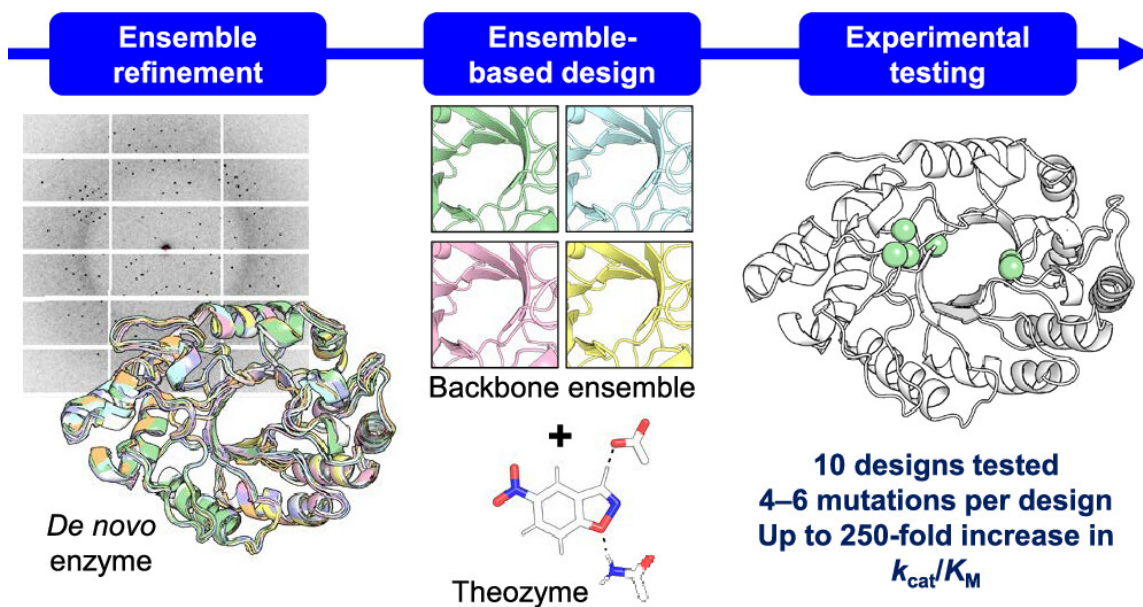


Figure 5.1: A crystallographic ensemble of the low-efficiency enzyme was used as the starting point for multi-state design, leading to variants with up to a 250-fold increase in catalytic efficiency without the need for directed evolution or other labor-intensive experimental approaches.

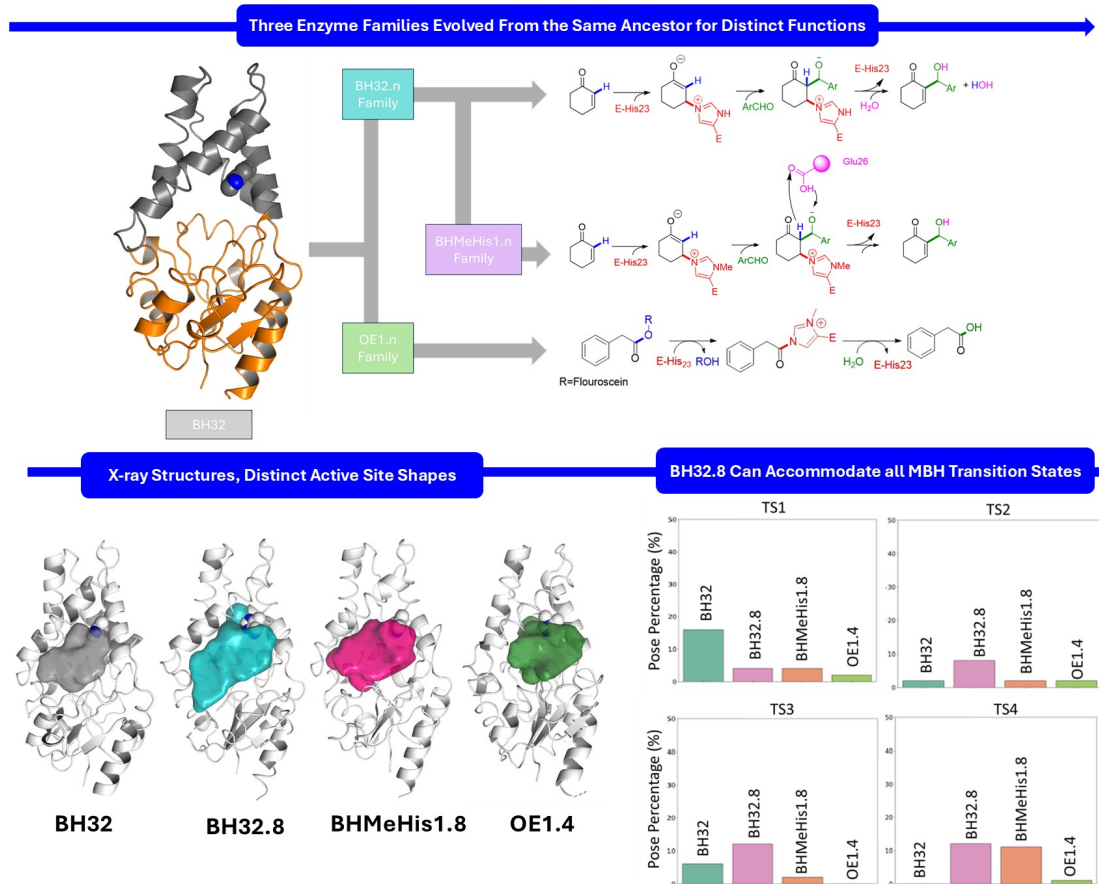


Figure 5.2: Three enzyme families evolved from a common ancestor (BH32) to catalyze distinct reactions, illustrating how structural and dynamic changes reshape catalytic function. **Top panel:** Schematic representation of the divergent evolution of the BH32 scaffold into three families (BH32.n [4], BH_{MeHis}1.n [5], and OE1.n [6]), each adapted to a different reaction mechanism. Key catalytic residues and reaction steps are highlighted to emphasize mechanistic differences across families. **Bottom left:** Room-temperature X-ray crystal structures of representative members (BH32, BH32.8, BH_{MeHis}1.8, and OE1.4) with their active-site cavities shown as colored surfaces, demonstrating how evolution alters the size and shape of the active site. **Bottom right:** Pose-compatibility analysis of DFT-derived transition states (TS1–TS4) calculated for MBH by BH32.14 and the conformational ensembles of each enzyme. BH32.8 shows increased compatibility with multiple transition states compared to the ancestor.

Room-temperature X-ray crystallographic structures were obtained for two members of each family and compared both to the ancestral enzyme and across families. Active-site shape and volume calculations, together with pose-compatibility analysis using conformational ensembles and DFT-derived transition states for the Morita–Baylis–Hillman reaction, reveal that differences in catalytic function are accompanied by systematic changes in active-site geometry.

In particular, BH32.8 exhibits an active site that accommodates a broader range of transition states compared to the ancestral BH32, consistent with its improved catalytic efficiency. Rather than optimizing for a single transition state, evolution reshapes the active site to support multiple states with moderate efficiency.

These functional changes arise from coordinated structural and dynamic adaptations, including clamshell-like opening between the cap (grey) and core (orange) domains, increased separation of cap-domain helices, and localized disorder and conformational rearrangements. Together, these features reshape the active site and modulate its flexibility, enabling the enzyme to traverse multiple states along the reaction coordinate.

Overall, this chapter provides insight into how enzymes balance structural organization and conformational dynamics to meet the demands of multi-step catalysis. By using directed evolution as a controlled framework, we establish a direct link between active-site remodeling, conformational heterogeneity, and the ability to stabilize multiple transition states within a single catalytic system.

Along the reaction coordinate, enzymes must pass through a series of short-lived, non-equilibrium states that can differ substantially from those observed in static apo structures. Capturing these states is therefore essential for understanding how structural dynamics contribute to catalytic function. In Chapter 2, active-site cavity and pose-compatibility analyses were performed on equilibrium conformational ensembles, which may not fully represent the transient, functionally relevant states populated during catalysis.

For example, the inhibitor-induced conformational change observed in OE1.3 may reflect a substrate-induced motion along the reaction coordinate, or it may be specific to the inhibited state. Similarly, other family members, including the BH32.n and BH_{MeHis}1.*n* lineages, may access analogous conformational transitions during catalysis that are not captured in equilibrium structures. Resolving this ambiguity requires direct observation of structures along the reaction coordinate rather than relying solely on apo or inhibitor-bound models.

In Chapter 4, we therefore focus on experimental strategies to access these non-equilibrium states. Time-resolved crystallographic approaches, such as mix-and-inject serial crystallography, provide a powerful means of capturing enzyme–substrate and enzyme–intermediate complexes by initiating reactions within microcrystals and monitoring structural changes as a function of time. To enable such experiments, we prepared OE1.4 as a candidate system by obtaining cryogenic and room-temperature structures and developing conditions for microcrystallization. The successful generation of high-density microcrystal suspensions establishes the feasibility of directly testing whether the observed conformational transitions are populated during catalysis.

In parallel, we explored a simplified approach that leverages slow reaction kinetics in engineered enzyme systems. Using ASB1.3, we took advantage of the intrinsic timescale of the reaction to perform controlled mixing and soaking experiments in single crystals. This approach enables access to non-equilibrium states without the need for rapid-mixing instrumentation, providing a practical alternative for systems with sufficiently slow turnover.

Together, these ongoing efforts provide a path toward addressing the remaining gaps in our understanding of how structural and dynamic adaptations support catalytic function in these enzyme families. By directly probing transient states along the reaction coordinate, these approaches will enable evaluation of whether the structural features identified in equilibrium structures correspond to functionally relevant

conformations, and how they contribute to catalytic efficiency.

5.2 Conclusion

Taken together, the results presented in this thesis define how enzymes balance structural preorganization and conformational adaptability to achieve catalytic efficiency across reactions with different mechanistic demands. In Chapter 2, we show that catalytic efficiency in single-step reactions can be improved by reducing conformational heterogeneity and enriching preorganized, catalytically competent conformations. This demonstrates that, when a single transition state dominates, precise active-site geometry and rigidity are sufficient to enhance catalysis.

In contrast, Chapter 3 shows that this strategy is not sufficient for multi-step reactions. Instead of optimizing for a single transition state, evolution reshapes the active site to accommodate multiple transition states along the reaction coordinate. This is achieved through coordinated structural and dynamic changes, including active-site expansion, domain motions, and localized rearrangements that broaden the accessible conformational landscape. Importantly, computational analysis reveals that evolved enzymes do not perfectly stabilize any single transition state, but instead adopt geometries that support multiple states with moderate efficiency. At the same time, experimental observations, including reduced thermal stability, indicate that this increased catalytic versatility is accompanied by increased flexibility, highlighting a trade-off between stability and functional adaptability.

Finally, Chapter 4 addresses a key limitation of equilibrium structural analysis by emphasizing that the relevant conformations for catalysis may not be fully captured in apo or inhibitor-bound states. The observation of large conformational changes, such as the helix shift in OE1.3, raises the possibility that additional functionally relevant states are accessed during catalysis. By developing both time-resolved crystallographic approaches and simplified mixing strategies, this work establishes a frame-

work for directly probing non-equilibrium states along the reaction coordinate.

Together, these results provide a more detailed picture of how enzymes achieve efficient catalysis. Rather than operating through a single optimized structure, enzymes balance preorganization and flexibility in a manner that depends on the complexity of the reaction. For single-step reactions, efficiency is achieved through rigidification and precise positioning, whereas for multi-step reactions, it arises from reshaping the active site to enable stabilization of multiple transition states. These findings highlight that catalytic efficiency is governed not only by static structure, but by how the conformational landscape is tuned to meet the demands of the reaction coordinate. More broadly, this work demonstrates that combining engineered systems with advanced structural methods provides a powerful approach for dissecting the relationship between enzyme structure, dynamics, and function.

Bibliography

- [1] Renee Otten, Ricardo AP Pádua, H Adrian Bunzel, Vy Nguyen, Warintra Pitsawong, MacKenzie Patterson, Shuo Sui, Sarah L Perry, Aina E Cohen, Donald Hilvert, et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science*, 370(6523):1442–1446, 2020.
- [2] Aron Broom, Rojo V Rakotoharisoa, Michael C Thompson, Niayesh Zarifi, Erin Nguyen, Nurzhan Mukhametzhanov, Lin Liu, James S Fraser, and Roberto A Chica. Ensemble-based enzyme design can recapitulate the effects of laboratory directed evolution in silico. *Nature communications*, 11(1):4808, 2020.
- [3] Brian P Mahon, Justin J Kurian, Carrie L Lomelino, Ian R Smith, Lilien Socorro, Antonette Bennett, Alex M Hendon, Paul Chipman, Daniel A Savin, Mavis Agbandje-McKenna, et al. Microbatch mixing: “shaken not stirred”, a method for macromolecular microcrystal production for serial crystallography. *Crystal Growth & Design*, 16(11):6214–6221, 2016.
- [4] Rebecca Crawshaw, Amy E Crossley, Linus Johannissen, Ashleigh J Burke, Sam Hay, Colin Levy, David Baker, Sarah L Lovelock, and Anthony P Green. Engineering an efficient and enantioselective enzyme for the morita–baylis–hillman reaction. *Nature chemistry*, 14(3):313–320, 2022.
- [5] Amy E Hutton, Jake Foster, Rebecca Crawshaw, Florence J Hardy, Linus O Johannissen, Thomas M Lister, Emilie F Gérard, Zachary Birch-Price, Richard Obexer, Sam Hay, et al. A non-canonical nucleophile unlocks a new mechanistic pathway in a designed enzyme. *Nature Communications*, 15(1):1956, 2024.
- [6] Ashleigh J Burke, Sarah L Lovelock, Amina Frese, Rebecca Crawshaw, Mary Ortmayer, Mark Dunstan, Colin Levy, and Anthony P Green. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature*, 570(7760):219–223, 2019.

- [7] Peter K Robinson. Enzymes: principles and biotechnological applications. *Essays in biochemistry*, 59:1, 2015.
- [8] Robert A Copeland. *Enzymes: a practical introduction to structure, mechanism, and data analysis*. John Wiley & Sons, 2023.
- [9] Donald G Truhlar, Bruce C Garrett, and Stephen J Klippenstein. Current status of transition-state theory. *The Journal of physical chemistry*, 100(31):12771–12800, 1996.
- [10] Albert L Lehninger. *Lehninger Principles of Biochemistry: David L. Nelson, Michael M. Cox*. Recording for the Blind & Dyslexic New York, 2004.
- [11] Joseph Kraut. How do enzymes work? *Science*, 242(4878):533–540, 1988.
- [12] ET Kaiser and DS Lawrence. Chemical mutation of enzyme active sites. *Science*, 226(4674):505–511, 1984.
- [13] Gail J Bartlett, Craig T Porter, Neera Borkakoti, and Janet M Thornton. Analysis of catalytic residues in enzyme active sites. *Journal of molecular biology*, 324(1):105–121, 2002.
- [14] Scott C-H Pegg, Shoshana D Brown, Sunil Ojha, Jennifer Seffernick, Elaine C Meng, John H Morris, Patricia J Chang, Conrad C Huang, Thomas E Ferrin, and Patricia C Babbitt. Leveraging enzyme structure- function relationships for functional inference and experimental design: the structure- function linkage database. *Biochemistry*, 45(8):2545–2555, 2006.
- [15] DE Koshland Jr. Correlation of structure and function in enzyme action: Theoretical and experimental tools are leading to correlations between enzyme structure and function. *Science*, 142(3599):1533–1541, 1963.
- [16] W Wallace Cleland, Perry A Frey, and John A Gerlt. The low barrier hydrogen bond in enzymatic catalysis. *Journal of Biological Chemistry*, 273(40):25529–25532, 1998.
- [17] Perry A Frey, Sean A Whitt, and John B Tobin. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science*, 264(5167):1927–1930, 1994.
- [18] Alan Fersht. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. Macmillan, 1999.

- [19] Jon D Robertus, Joseph Kraut, Richard A Alden, and Jens J Birktoft. Subtilisin. stereochemical mechanism involving transition-state stabilization. *Biochemistry*, 11(23):4293–4303, 1972.
- [20] Dudley H Williams. Enzyme catalysis from improved packing in their transition-state structures. *Current opinion in chemical biology*, 14(5):666–670, 2010.
- [21] Florian P Seebeck and Donald Hilvert. Positional ordering of reacting groups contributes significantly to the efficiency of proton transfer at an antibody active site. *Journal of the American Chemical Society*, 127(4):1307–1312, 2005.
- [22] Stephen D Fried and Steven G Boxer. Electric fields and enzyme catalysis. *Annual review of biochemistry*, 86:387–415, 2017.
- [23] Johan Åqvist, Masoud Kazemi, Geir Villy Isaksen, and Bjørn Olav Brandsdal. Entropy and enzyme catalysis. *Accounts of chemical research*, 50(2):199–207, 2017.
- [24] L Mario Amzel. Loss of translational entropy in binding, folding, and catalysis. *Proteins: Structure, Function, and Bioinformatics*, 28(2):144–149, 1997.
- [25] Michael I Page and William P Jencks. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proceedings of the National Academy of Sciences*, 68(8):1678–1683, 1971.
- [26] Thomas C Bruice and Upendra K Pandit. The effect of geminal substitution ring size and rotamer distribution on the intramolecular nucleophilic catalysis of the hydrolysis of monophenyl esters of dibasic acids and the solvolysis of the intermediate anhydrides. *Journal of the American Chemical Society*, 82(22):5858–5865, 1960.
- [27] Vern L Schramm. Enzymatic transition states and transition state analog design. *Annual review of biochemistry*, 67(1):693–720, 1998.
- [28] Serena E Hunt, Cindy Klaus, Aqza E John, Niayesh Zarifi, Alec Martinez, Ferran Feixas, Marc Garcia-Borràs, Michael C Thompson, and Roberto A Chica. Distal mutations in a designed retro-aldolase alter loop dynamics to shift and accelerate the rate-limiting step. *Journal of the American Chemical Society*, 147(34):30723–30736, 2025.

- [29] Niayesh Zarifi, Pooja Asthana, Hiva Doustmohammadi, Cindy Klaus, Janet Sanchez, Serena E Hunt, Rojo V Rakotoharisoa, Silvia Osuna, James S Fraser, and Roberto A Chica. Distal mutations enhance catalysis in designed enzymes by facilitating substrate binding and product release. *bioRxiv*, 2025.
- [30] Robert Callender and R Brian Dyer. The dynamical nature of enzymatic catalysis. *Accounts of chemical research*, 48(2):407–413, 2015.
- [31] Hannah K Wayment-Steele, Gina El Nesr, Ramith Hettiarachchi, Adedolapo M Ojoawo, Hasindu Kariyawasam, Sergey Ovchinnikov, and Dorothee Kern. Learning millisecond protein dynamics from what is missing in nmr spectra. *BioRxiv*, pages 2025–03, 2025.
- [32] Silvia Osuna. The challenge of predicting distal active site mutations in computational enzyme design. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(3):e1502, 2021.
- [33] Gordon G Hammes. Multiple conformational changes in enzyme catalysis. *Biochemistry*, 41(26):8221–8228, 2002.
- [34] Mireia Garcia-Viloca, Jiali Gao, Martin Karplus, and Donald G Truhlar. How enzymes work: analysis by modern rate theory and computer simulations. *Science*, 303(5655):186–195, 2004.
- [35] Elan Zohar Eisenmesser, Daryl A Bosco, Mikael Akke, and Dorothee Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–1523, 2002.
- [36] Elan Z Eisenmesser, Oscar Millet, Wladimir Labeikovsky, Dmitry M Korzhnev, Magnus Wolf-Watz, Daryl A Bosco, Jack J Skalicky, Lewis E Kay, and Dorothee Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005.
- [37] Stephen P Goff. Genetic control of retrovirus susceptibility in mammalian cells. *Annu. Rev. Genet.*, 38(1):61–85, 2004.
- [38] Sachin Kale, Gözde Ulas, Jaeyoung Song, Gary W Brudvig, William Furey, and Frank Jordan. Efficient coupling of catalysis and dynamics in the e1 component of escherichia coli pyruvate dehydrogenase multienzyme complex. *Proceedings of the National Academy of Sciences*, 105(4):1158–1163, 2008.

- [39] Pedram Mehrabi, Eike C Schulz, Raison Dsouza, Henrike M Müller-Werkmeister, Friedjof Tellkamp, RJ Dwayne Miller, and Emil F Pai. Time-resolved crystallography reveals allosteric communication aligned with molecular breathing. *Science*, 365(6458):1167–1170, 2019.
- [40] Kyle M Stiers, Abigail C Graham, Jian-She Zhu, David L Jakeman, Jay C Nix, and Lesa J Beamer. Structural and dynamical description of the enzymatic reaction of a phosphohexomutase. *Structural Dynamics*, 6(2), 2019.
- [41] Jack B Greisman, Kevin M Dalton, Dennis E Brookner, Margaret A Klureza, Candice J Sheehan, In-Sik Kim, Robert W Henning, Silvia Russi, and Doeke R Hekstra. Perturbative diffraction methods resolve a conformational switch that facilitates a two-step enzymatic mechanism. *Proceedings of the National Academy of Sciences*, 121(9):e2313192121, 2024.
- [42] Medhanjali Dasgupta, Dominik Budday, Saulo HP De Oliveira, Peter Madzelan, Darya Marchany-Rivera, Javier Seravalli, Brandon Hayes, Raymond G Sierra, Sébastien Boutet, Mark S Hunter, et al. Mix-and-inject xfel crystallography reveals gated conformational dynamics during enzyme catalysis. *Proceedings of the National Academy of Sciences*, 116(51):25634–25640, 2019.
- [43] Masahiko Goda, Yoshiteru Hashimoto, Sakayu Shimizu, and Michihiko Kobayashi. Discovery of a novel enzyme, isonitrile hydratase, involved in nitrogen-carbon triple bond cleavage. *Journal of Biological Chemistry*, 276(26):23480–23485, 2001.
- [44] Mahadevan Lakshminarasimhan, Peter Madzelan, Ruth Nan, Nicole M Milkovic, and Mark A Wilson. Evolution of new enzymatic function by structural modulation of cysteine reactivity in pseudomonas fluorescens isocyanide hydratase. *Journal of Biological Chemistry*, 285(38):29651–29661, 2010.
- [45] Andrew E Brereton and P Andrew Karplus. Native proteins trap high-energy transit conformations. *Science Advances*, 1(9):e1501188, 2015.
- [46] Krithika Ramakrishnan, Rachel L Johnson, Samuel D Winter, Harley L Worthy, Christopher Thomas, Diana C Humer, Oliver Spadiut, Sarah H Hindson, Stephen Wells, Andrew H Barratt, et al. Glycosylation increases active site rigidity leading to improved enzyme stability and turnover. *The FEBS Journal*, 290(15):3812–3827, 2023.

- [47] Heidi K Privett, Gert Kiss, Toni M Lee, Rebecca Blomberg, Roberto A Chica, Leonard M Thomas, Donald Hilvert, Kendall N Houk, and Stephen L Mayo. Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences*, 109(10):3790–3795, 2012.
- [48] Daniela Röthlisberger, Olga Khersonsky, Andrew M Wollacott, Lin Jiang, Jason DeChancie, Jamie Betker, Jasmine L Gallaher, Eric A Althoff, Alexandre Zanghellini, Orly Dym, et al. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–195, 2008.
- [49] Rebecca Blomberg, Hajo Kries, Daniel M Pinkas, Peer RE Mittl, Markus G Grütter, Heidi K Privett, Stephen L Mayo, and Donald Hilvert. Precision is essential for efficient catalysis in an evolved kemp eliminase. *Nature*, 503(7476):418–421, 2013.
- [50] Roman Zadorozhnyi, Angela M Gronenborn, and Tatyana Polenova. Integrative approaches for characterizing protein dynamics: Nmr, cryoem, and computer simulations. *Current opinion in structural biology*, 84:102736, 2024.
- [51] Shigeyuki Matsumoto, Shoichi Ishida, Mitsugu Araki, Takayuki Kato, Kei Terayama, and Yasushi Okuno. Extraction of protein dynamics information from cryo-em maps using deep learning. *Nature Machine Intelligence*, 3(2):153–160, 2021.
- [52] Dina Schneidman-Duhovny and Michal Hammel. Modeling structure and dynamics of protein complexes with saxs profiles. In *Protein Complex Assembly: Methods and Protocols*, pages 449–473. Springer, 2018.
- [53] Lei Zhang and Gang Ren. Ipet and fetr: experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure. *PloS one*, 7(1):e30249, 2012.
- [54] Guy G Dodson, David P Lane, and Chandra S Verma. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *The EMBO Reports*, 9(2):144–150, 2008.
- [55] Michael C Thompson. Combining temperature perturbations with x-ray crystallography to study dynamic macromolecules: A thorough discussion of experimental methods. *Methods in Enzymology*, 688:255–305, 2023.
- [56] Gale Rhodes. *Crystallography made crystal clear: a guide for users of macromolecular models*. Elsevier, 2010.

- [57] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- [58] Daniel A Keedy, Lillian R Kenner, Matthew Warkentin, Rahel A Woldeyes, Jesse B Hopkins, Michael C Thompson, Aaron S Brewster, Andrew H Van Benschoten, Elizabeth L Baxter, Monarin Uervirojnangkoorn, et al. Mapping the conformational landscape of a dynamic enzyme by multitemperature and xfel crystallography. *Elife*, 4:e07574, 2015.
- [59] Ki Hyun Nam. Dehydration-induced space group transition triggers conformational changes in protein structure. *Crystals*, 15(8):674, 2025.
- [60] Michael C Thompson, Benjamin A Barad, Alexander M Wolff, Hyun Sun Cho, Friedrich Schotte, Daniel MC Schwarz, Philip Anfinrud, and James S Fraser. Temperature-jump solution x-ray scattering reveals distinct motions in a dynamic enzyme. *Nature chemistry*, 11(11):1058–1066, 2019.
- [61] Cristiano Malica and Andrea Dal Corso. Temperature-dependent atomic b factor: An ab initio calculation. *Foundations of Crystallography*, 75(4):624–632, 2019.
- [62] Hyuntae Na, Konrad Hinsén, and Guang Song. The amounts of thermal vibrations and static disorder in protein x-ray crystallographic b-factors. *Proteins: Structure, Function, and Bioinformatics*, 89(11):1442–1457, 2021.
- [63] Georg Mlynek, Kristina Djinović-Carugo, and Oliviero Carugo. B-factor rescaling for protein crystal structure analyses. *Crystals*, 14(5):443, 2024.
- [64] Lei Zhou and Qinglian Liu. Aligning experimental and theoretical anisotropic b-factors: water models, normal-mode analysis methods, and metrics. *The Journal of Physical Chemistry B*, 118(15):4069–4079, 2014.
- [65] MD Winn, MN Isupov, and Garib N Murshudov. Use of tls parameters to model anisotropic displacements in macromolecular refinement. *Biological Crystallography*, 57(1):122–133, 2001.
- [66] Siyuan Du, Stephanie A Wankowicz, Filip Yabukarski, Tzanko Doukov, Daniel Herschlag, and James S Fraser. Refinement of multiconformer ensemble models from multi-temperature x-ray diffraction data. In *Methods in enzymology*, volume 688, pages 223–254. Elsevier, 2023.

- [67] R Bryn Fenwick, Henry Van den Bedem, James S Fraser, and Peter E Wright. Integrated description of protein dynamics from room-temperature x-ray crystallography and nmr. *Proceedings of the National Academy of Sciences*, 111(4):E445–E454, 2014.
- [68] Mehagan S Hopkins, Thomas C Terwilliger, Pavel V Afonine, Helen M Ginn, and James M Holton. The untangle challenge for accurate ensemble models. *bioRxiv*, pages 2026–02, 2026.
- [69] B Tom Burnley, Pavel V Afonine, Paul D Adams, and Piet Gros. Modelling dynamics in protein crystal structures by ensemble refinement. *elife*, 1:e00311, 2012.
- [70] Jose M Martin-Garcia, Chelsie E Conrad, Jesse Coe, Shatabdi Roy-Chowdhury, and Petra Fromme. Serial femtosecond crystallography: A revolution in structural biology. *Archives of biochemistry and biophysics*, 602:32–47, 2016.
- [71] Jason R Stagno, Juraj Knoska, Henry N Chapman, and Yun-Xing Wang. Mix-and-inject serial femtosecond crystallography to capture rna riboswitch intermediates. In *RNA Structure and Dynamics*, pages 243–249. Springer, 2022.
- [72] KA Zielinski and L Pollack. Developing mix-and-inject serial crystallography at synchrotrons. *Acta Cryst*, 79:C550, 2023.
- [73] Elspeth F Garman and Martin Weik. Radiation damage in macromolecular crystallography. *Protein Crystallography: Methods and Protocols*, pages 467–489, 2017.
- [74] Elspeth F Garman and Robin Leslie Owen. Cryocooling and radiation damage in macromolecular crystallography. *Biological Crystallography*, 62(1):32–47, 2006.
- [75] David J Haas. The early history of cryo-cooling for macromolecular crystallography. *IUCrJ*, 7(2):148–157, 2020.
- [76] Douglas H Juers and Brian W Matthews. Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization. *Quarterly reviews of biophysics*, 37(2):105–119, 2004.
- [77] James S Fraser, Henry Van Den Bedem, Avi J Samelson, P Therese Lang, James M Holton, Nathaniel Echols, and Tom Alber. Accessing protein conformational ensembles using room-temperature x-ray crystallography. *Proceedings of the National Academy of Sciences*, 108(39):16247–16252, 2011.

- [78] Alexandre Zanghellini. De novo computational enzyme design. *Current opinion in biotechnology*, 29:132–138, 2014.
- [79] Enrique Marcos and Daniel-Adriano Silva. Essentials of de novo protein design: Methods and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(6):e1374, 2018.
- [80] Po-Ssu Huang, Scott E Boyken, and David Baker. The coming of age of de novo protein design. *Nature*, 537(7620):320–327, 2016.
- [81] Olga Kuchner and Frances H Arnold. Directed evolution of enzyme catalysts. *Trends in biotechnology*, 15(12):523–530, 1997.
- [82] Nicholas J Turner. Directed evolution of enzymes for applied biocatalysis. *TRENDS in Biotechnology*, 21(11):474–478, 2003.
- [83] Katherine A Henzler-Wildman, Vu Thai, Ming Lei, Maria Ott, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A Wilson, Gregory A Petsko, Martin Karplus, et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171):838–844, 2007.
- [84] Adelajda Hadzipasic, Christopher Wilson, Vy Nguyen, Nadja Kern, Chan-sik Kim, Warintra Pitsawong, Janice Villali, Yuejiao Zheng, and Dorothee Kern. Ancient origins of allosteric activation in a ser-thr kinase. *Science*, 367(6480):912–917, 2020.
- [85] Wai Shun Mak and Justin B Siegel. Computational enzyme design: transitioning from catalytic proteins to enzymes. *Current opinion in structural biology*, 27:87–94, 2014.
- [86] Jonathan Kyle Lassila, Heidi K Privett, Benjamin D Allen, and Stephen L Mayo. Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences*, 103(45):16710–16715, 2006.
- [87] Alexandre Zanghellini, Lin Jiang, Andrew M Wollacott, Gong Cheng, Jens Meiler, Eric A Althoff, Daniela Röthlisberger, and David Baker. New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, 15(12):2785–2794, 2006.

- [88] Lin Jiang, Eric A Althoff, Fernando R Clemente, Lindsey Doyle, Daniela Rothlisberger, Alexandre Zanghellini, Jasmine L Gallaher, Jamie L Betker, Fujie Tanaka, Carlos F Barbas III, et al. De novo computational design of retroaldol enzymes. *science*, 319(5868):1387–1391, 2008.
- [89] Justin B Siegel, Alexandre Zanghellini, Helena M Lovick, Gert Kiss, Abigail R Lambert, Jennifer L St. Clair, Jasmine L Gallaher, Donald Hilvert, Michael H Gelb, Barry L Stoddard, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329(5989):309–313, 2010.
- [90] Sinisa Bjelic, Lucas G Nivon, Nihan Celebi-Olcum, Gert Kiss, Carolyn F Rosewall, Helena M Lovick, Erica L Ingalls, Jasmine Lynn Gallaher, Jayaraman Seetharaman, Scott Lew, et al. Computational design of enone-binding proteins with catalytic activity for the morita–baylis–hillman reaction. *ACS chemical biology*, 8(4):749–757, 2013.
- [91] Andy Hsien-Wei Yeh, Christoffer Norn, Yakov Kipnis, Doug Tischer, Samuel J Pellock, Declan Evans, Pengchen Ma, Gyu Rie Lee, Jason Z Zhang, Ivan Anishchenko, et al. De novo design of luciferases using deep learning. *Nature*, 614(7949):774–780, 2023.
- [92] Olga Khersonsky, Daniela Röthlisberger, Andrew M Wollacott, Paul Murphy, Orly Dym, Shira Albeck, Gert Kiss, KN Houk, David Baker, and Dan S Tawfik. Optimization of the in-silico-designed kemp eliminase ke70 by computational design and directed evolution. *Journal of molecular biology*, 407(3):391–412, 2011.
- [93] Olga Khersonsky, Gert Kiss, Daniela Röthlisberger, Orly Dym, Shira Albeck, Kendall N Houk, David Baker, and Dan S Tawfik. Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed kemp eliminase ke59. *Proceedings of the National Academy of Sciences*, 109(26):10358–10363, 2012.
- [94] Lars Giger, Sami Caner, Richard Obexer, Peter Kast, David Baker, Nenad Ban, and Donald Hilvert. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nature chemical biology*, 9(8):494–498, 2013.
- [95] Richard Obexer, Alexei Godina, Xavier Garrabou, Peer RE Mittl, David Baker, Andrew D Griffiths, and Donald Hilvert. Emergence of a catalytic tetrad during

- evolution of a highly active artificial aldolase. *Nature chemistry*, 9(1):50–56, 2017.
- [96] Eleanor Campbell, Miriam Kaltenbach, Galen J Correy, Paul D Carr, Benjamin T Porebski, Emma K Livingstone, Livnat Afriat-Jurnou, Ashley M Buckle, Martin Weik, Florian Hollfelder, et al. The role of protein dynamics in the evolution of new enzyme function. *Nature chemical biology*, 12(11):944–950, 2016.
 - [97] H Adrian Bunzel, JL Ross Anderson, Donald Hilvert, Vickery L Arcus, Marc W Van Der Kamp, and Adrian J Mulholland. Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nature Chemistry*, 13(10):1017–1022, 2021.
 - [98] Daniel A Keedy, James S Fraser, and Henry van den Bedem. Exposing hidden alternative backbone conformations in x-ray crystallography using qfit. *PLoS computational biology*, 11(10):e1004507, 2015.
 - [99] Elena J Levin, Dmitry A Kondrashov, Gary E Wesenberg, and George N Phillips. Ensemble refinement of protein crystal structures: validation and application. *Structure*, 15(9):1040–1052, 2007.
 - [100] Daniel A Keedy, Henry Van Den Bedem, David A Sivak, Gregory A Petsko, Dagmar Ringe, Mark A Wilson, and James S Fraser. Crystal cryocooling distorts conformational heterogeneity in a model michaelis complex of dhfr. *Structure*, 22(6):899–910, 2014.
 - [101] Alexandre V Morozov, Tanja Kortemme, Kiril Tsemekhman, and David Baker. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proceedings of the National Academy of Sciences*, 101(18):6946–6951, 2004.
 - [102] Justin T Biel, Michael C Thompson, Christian N Cunningham, Jacob E Corn, and James S Fraser. Flexibility and design: conformational heterogeneity along the evolutionary trajectory of a redesigned ubiquitin. *Structure*, 25(5):739–749, 2017.
 - [103] Garima Jindal and Arie Warshel. Misunderstanding the preorganization concept can lead to confusions about the origin of enzyme catalysis. *Proteins: Structure, Function, and Bioinformatics*, 85(12):2157–2161, 2017.

- [104] Miguel A Maria-Solano, Eila Serrano-Hervás, Adrian Romero-Rivera, Javier Iglesias-Fernández, and Sílvia Osuna. Role of conformational dynamics in the evolution of novel enzyme function. *Chemical Communications*, 54(50):6622–6634, 2018.
- [105] Filip Yabukarski, Tzanko Doukov, Margaux M Pinney, Justin T Biel, James S Fraser, and Daniel Herschlag. Ensemble-function relationships to dissect mechanisms of enzyme catalysis. *Science Advances*, 8(41):eabn7738, 2022.
- [106] James A Davey and Roberto A Chica. Multistate approaches in computational protein design. *Protein Science*, 21(9):1241–1252, 2012.
- [107] Antony D St-Jacques, Joshua M Rodriguez, Matthew G Eason, Scott M Foster, Safwat T Khan, Adam M Damry, Natalie K Goto, Michael C Thompson, and Roberto A Chica. Computational remodeling of an enzyme conformational landscape for altered substrate selectivity. *Nature Communications*, 14(1):6058, 2023.
- [108] Sandrine Legault, Derek P Fraser-Halberg, Ralph L McAnelly, Matthew G Eason, Michael C Thompson, and Roberto A Chica. Generation of bright monomeric red fluorescent proteins via computational design of enhanced chromophore packing. *Chemical Science*, 13(5):1408–1418, 2022.
- [109] Christine E Tinberg, Sagar D Khare, Jiayi Dou, Lindsey Doyle, Jorgen W Nelson, Alberto Schena, Wojciech Jankowski, Charalampos G Kalodimos, Kai Johnsson, Barry L Stoddard, et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature*, 501(7466):212–216, 2013.
- [110] Dorothee Liebschner, Pavel V Afonine, Matthew L Baker, Gábor Bunkóczi, Vincent B Chen, Tristan I Croll, Bradley Hintze, L-W Hung, Swati Jain, Airle J McCoy, et al. Macromolecular structure determination using x-rays, neutrons and electrons: recent developments in phenix. *Biological Crystallography*, 75(10):861–877, 2019.
- [111] Frederick S Lee, Amos G Anderson, and Barry D Olafson. Benchmarking triadab using targets from the second antibody modeling assessment. *Protein Engineering, Design and Selection*, 36:gza013, 2023.
- [112] Roland L Dunbrack Jr. Rotamer libraries in the 21st century. *Current opinion in structural biology*, 12(4):431–440, 2002.

- [113] Stephen L Mayo, Barry D Olafson, and William A Goddard. Dreiding: a generic force field for molecular simulations. *Journal of Physical chemistry*, 94(26):8897–8909, 1990.
- [114] Bassil I Dahiyat and Stephen L Mayo. Probing the role of packing specificity in protein design. *Proceedings of the National Academy of Sciences*, 94(19):10172–10177, 1997.
- [115] Benjamin D Allen, Alex Nisthal, and Stephen L Mayo. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proceedings of the National Academy of Sciences*, 107(46):19838–19843, 2010.
- [116] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.
- [117] Tom Darden, Darrin York, Lee Pedersen, et al. Particle mesh ewald: An $n \log(n)$ method for ewald sums in large systems. *Journal of chemical physics*, 98:10089–10089, 1993.
- [118] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics and modelling*, 25(2):247–260, 2006.
- [119] J Michael Word, Simon C Lovell, Jane S Richardson, and David C Richardson. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of molecular biology*, 285(4):1735–1747, 1999.
- [120] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.
- [121] Denis J Evans and BL Holian. The nose–hoover thermostat. $A)'=(A \text{ } Is)$, 1:18, 1985.
- [122] Herman JC Berendsen, JPM van Postma, Wilfred F Van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

- [123] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.
- [124] Graeme Winter, Carina MC Lobley, and Stephen M Prince. Decision making in xia2. *Biological Crystallography*, 69(7):1260–1273, 2013.
- [125] Graeme Winter, David G Waterman, James M Parkhurst, Aaron S Brewster, Richard J Gildea, Markus Gerstel, Luis Fuentes-Montero, Melanie Vollmar, Tara Michels-Clark, Iris D Young, et al. Dials: implementation and evaluation of a new integration package. *Biological Crystallography*, 74(2):85–97, 2018.
- [126] Philip Evans. Scaling and assessment of data quality. *Biological crystallography*, 62(1):72–82, 2006.
- [127] Airlie J McCoy, Ralf W Grosse-Kunstleve, Paul D Adams, Martyn D Winn, Laurent C Storoni, and Randy J Read. Phaser crystallographic software. *Applied Crystallography*, 40(4):658–674, 2007.
- [128] Paul D Adams, Pavel V Afonine, Gábor Bunkóczi, Vincent B Chen, Ian W Davis, Nathaniel Echols, Jeffrey J Headd, L-W Hung, Gary J Kapral, Ralf W Grosse-Kunstleve, et al. Phenix: a comprehensive python-based system for macromolecular structure solution. *Biological crystallography*, 66(2):213–221, 2010.
- [129] Paul Emsley, Bernhard Lohkamp, William G Scott, and Kevin Cowtan. Features and development of coot. *Biological crystallography*, 66(4):486–501, 2010.
- [130] Pavel V Afonine, Ralf W Grosse-Kunstleve, Nathaniel Echols, Jeffrey J Headd, Nigel W Moriarty, Marat Mustyakimov, Thomas C Terwilliger, Alexandre Urzhumtsev, Peter H Zwart, and Paul D Adams. Towards automated crystallographic structure refinement with phenix. refine. *Biological crystallography*, 68(4):352–367, 2012.
- [131] Nigel W Moriarty, Ralf W Grosse-Kunstleve, and Paul D Adams. electronic ligand builder and optimization workbench (elbow): a tool for ligand coordinate and restraint generation. *Biological Crystallography*, 65(10):1074–1080, 2009.
- [132] Jon R Lorsch. Practical steady-state enzyme kinetics. *Methods in enzymology*, 536:3–15, 2014.

- [133] Pratul K Agarwal. A biophysical perspective on enzyme catalysis. *Biochemistry*, 58(6):438–449, 2018.
- [134] Gabriel E Jara, Francesco Pontiggia, Renee Otten, Roman V Agafonov, Marcelo A Martí, and Dorothee Kern. Wide transition-state ensemble as key component for enzyme catalysis. *Elife*, 12:RP93099, 2025.
- [135] Sophie E Jackson and Alan R Fersht. Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin bpn’. *Biochemistry*, 32(50):13909–13916, 1993.
- [136] KB Schowen, H-H Limbach, GS Denisov, and RL Schowen. Hydrogen bonds and proton transfer in general-catalytic transition-state stabilization in enzyme catalysis. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1458(1):43–62, 2000.
- [137] Daniel Burschowsky, André van Eerde, Mats Ökvist, Alexander Kienhöfer, Peter Kast, Donald Hilvert, and Ute Krengel. Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis. *Proceedings of the National Academy of Sciences*, 111(49):17516–17521, 2014.
- [138] Kirill Zinovjev and Iñaki Tuñón. Quantifying the limits of transition state theory in enzymatic catalysis. *Proceedings of the National Academy of Sciences*, 114(47):12390–12395, 2017.
- [139] Steven D Schwartz and Vern L Schramm. Enzymatic transition states and dynamic motion in barrier crossing. *Nature chemical biology*, 5(8):551–558, 2009.
- [140] Guanhua Hou and Qiang Cui. Stabilization of different types of transition states in a single enzyme active site: Qm/mm analysis of enzymes in the alkaline phosphatase superfamily. *Journal of the American Chemical Society*, 135(28):10457–10469, 2013.
- [141] Adam JT Smith, Roger Muller, Miguel D Toscano, Peter Kast, Homme W Hellinga, Donald Hilvert, and KN Houk. Structural reorganization and preorganization in enzyme active sites: comparisons of experimental and theoretically ideal active site geometries in the multistep serine esterase reaction cycle. *Journal of the American Chemical Society*, 130(46):15361–15373, 2008.

- [142] Anshula Sharma, Gaganjot Gupta, Tawseef Ahmad, Sheikh Mansoor, and Baljinder Kaur. Enzyme engineering: current trends and future perspectives. *Food Reviews International*, 37(2):121–154, 2021.
- [143] Min Shi, Mei-Xin Zhao, Yin Wei, and Feijun Wang. *Chemistry of the Morita-Baylis-Hillman Reaction*. Royal Society of Chemistry, 2011.
- [144] Deevi Basavaiah, Bhavanam Sekhara Reddy, and Satpal Singh Badsara. Recent contributions from the baylis- hillman reaction to organic chemistry. *Chemical reviews*, 110(9):5447–5674, 2010.
- [145] Amy E Hutton, Jake Foster, James EJ Sanders, Christopher J Taylor, Stefan A Hoffmann, Yizhi Cai, Sarah L Lovelock, and Anthony P Green. An efficient pyrrolysyl-trna synthetase for economical production of mehis-containing enzymes. *Faraday Discussions*, 252:295–305, 2024.
- [146] Mia A Shandell, Zhongping Tan, and Virginia W Cornish. Genetic code expansion: a brief history and perspective. *Biochemistry*, 60(46):3455–3469, 2021.
- [147] Ana Crnković, Tateki Suzuki, Dieter Söll, and Noah M Reynolds. Pyrrolysyl-trna synthetase, an aminoacyl-trna synthetase for genetic code expansion. *Croatica Chemica Acta*, 89(2):163–174, 2016.
- [148] Douglas H Juers and Brian W Matthews. Reversible lattice repacking illustrates the temperature dependence of macromolecular interactions. *Journal of molecular biology*, 311(4):851–862, 2001.
- [149] Joao VS Guerra, Helder V Ribeiro-Filho, José GC Pereira, and Paulo S Lopes-de Oliveira. Kvfinder-web: a web-based application for detecting and characterizing biomolecular cavities. *Nucleic acids research*, 51(W1):W289–W297, 2023.
- [150] João Victor da Silva Guerra, Helder Veras Ribeiro Filho, Leandro Oliveira Bortot, Rodrigo Vargas Honorato, Jose Geraldo de Carvalho Pereira, and Paulo Sergio Lopes-de Oliveira. Parkvfinder: A thread-level parallel approach in biomolecular cavity detection. *SoftwareX*, 12:100606, 2020.
- [151] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

- [152] Maylis Orio, Dimitrios A Pantazis, and Frank Neese. Density functional theory. *Photosynthesis research*, 102(2):443–453, 2009.
- [153] Graeme Winter. xia2: an expert system for macromolecular crystallography data reduction. *Applied Crystallography*, 43(1):186–190, 2010.
- [154] Guangyan Zhang, Guiyong Wu, Yu Lu, Jie Wu, and Weimin Zheng. Xscale: online x-code raid-6 scaling using lightweight data reorganization. *IEEE Transactions on Parallel and Distributed Systems*, 27(12):3687–3700, 2016.
- [155] Jon Agirre, Mihaela Atanasova, Haroldas Bagdonas, Charles B Ballard, Arnaud Baslé, James Beilsten-Edmands, Rafael J Borges, David G Brown, J Javier Burgos-Mármol, John M Berrisford, et al. The ccp4 suite: integrative software for macromolecular crystallography. *Biological Crystallography*, 79(6):449–461, 2023.
- [156] Blake T Riley, Stephanie A Wankowicz, Saulo HP de Oliveira, Gydo CP van Zundert, Daniel W Hogan, James S Fraser, Daniel A Keedy, and Henry van den Bedem. qfit 3: Protein and ligand multiconformer modeling for x-ray crystallographic and single-particle cryo-em density maps. *Protein science*, 30(1):270–285, 2021.
- [157] CR Wagner and SJ Benkovic. Site directed mutagenesis: a tool for enzyme mechanism dissection. *Trends in biotechnology*, 8:263–270, 1990.
- [158] Forest H Andrews and Michael J McLeish. Using site-saturation mutagenesis to explore mechanism and substrate specificity in thiamin diphosphate-dependent enzymes. *The FEBS Journal*, 280(24):6395–6411, 2013.
- [159] Vern L Schramm. Transition states and transition state analogue interactions with enzymes. *Accounts of Chemical Research*, 48(4):1032–1039, 2015.
- [160] Craig J Markin, Daniel A Mokhtari, Siyuan Du, Tzanko Doukov, Fanny Sundén, Jordan A Cook, Polly M Fordyce, and Daniel Herschlag. Decoupling of catalysis and transition state analog binding from mutations throughout a phosphatase revealed by high-throughput enzymology. *Proceedings of the National Academy of Sciences*, 120(29):e2219074120, 2023.
- [161] Jun Hiratake. Enzyme inhibitors as chemical tools to study enzyme catalysis: rational design, synthesis, and applications. *The Chemical Record*, 5(4):209–228, 2005.

- [162] George D Calvey, Andrea M Katz, Kara A Zielinski, Boris Dzikovski, and Lois Pollack. Characterizing enzyme reactions in microcrystals for effective mix-and-inject experiments using x-ray free-electron lasers. *Analytical chemistry*, 92(20):13864–13870, 2020.
- [163] Haiguang Liu and Weontae Lee. The xfel protein crystallography: developments and perspectives, 2019.
- [164] JR Stagno, Y Liu, YR Bhandari, CE Conrad, S Panja, Mamata Swain, L Fan, Gerald Nelson, C Li, DR Wendel, et al. Structures of riboswitch rna reaction states by mix-and-inject xfel serial crystallography. *Nature*, 541(7636):242–246, 2017.
- [165] Philip Willmott. *An introduction to synchrotron radiation: techniques and applications*. John Wiley & Sons, 2019.
- [166] Ki Hyun Nam. Guide to serial synchrotron crystallography. *Current Research in Structural Biology*, 7:100131, 2024.
- [167] Jose L Olmos Jr, Suraj Pandey, Jose M Martin-Garcia, George Calvey, Andrea Katz, Juraj Knoska, Christopher Kupitz, Mark S Hunter, Mengning Liang, Dominik Oberthuer, et al. Enzyme intermediates captured “on the fly” by mix-and-inject serial crystallography. *BMC biology*, 16(1):59, 2018.