Supporting Information: MAPLE: A Machine-Learning Force-FieldNative Platform for Automated Reaction Modeling and Enzyme Design

Details for Algorithmic Developments

Batched L-BFGS Optimization with Dynamic Batch Shrinking

For large-scale geometry optimizations involving multiple independent molecular structures, we developed a GPU-accelerated batched L-BFGS optimizer that exploits parallelism across structures while maintaining individual quasi-Newton histories for each system. This approach significantly reduces wall-clock time for high-throughput computational screening workflows.

The Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm[1] is a quasi-Newton method that approximates the inverse Hessian using a limited history of gradient and position changes. In standard L-BFGS, the search direction $\mathbf{p}_k$ at iteration k is computed via the two-loop recursion algorithm, which requires storing the most recent m pairs of vectors $\{\mathbf{s}_i, \mathbf{y}_i\}_{i=k-m}^{k-1}$, where $\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$ is the position change and $\mathbf{y}_i = \nabla f_{i+1} - \nabla f_i$ is the gradient change. Our implementation extends this to optimize B structures simultaneously on GPU with three key innovations.

Fixed-dimension tensor batching. All molecular structures are represented in a unified coordinate tensor of shape $(B, n_{\max})$, where $n_{\max} = 3N_{\max}$ is fixed at initialization based on the largest structure in the batch. Smaller structures are zero-padded, and a binary mask tensor $\mathbf{M} \in \{0,1\}^{B \times n_{\max}}$ identifies valid atomic coordinates. This uniform dimensionality enables efficient batched linear algebra operations (element-wise products, dot products, vector additions) to be executed in parallel across all structures with a single GPU kernel launch, avoiding the overhead of sequential per-structure processing.

Independent per-batch L-BFGS histories. Each structure maintains its own quasi-Newton memory consisting of m stored vector pairs and curvature information. The history for batch b is represented as $\{\mathbf{S}^{(b)}, \mathbf{Y}^{(b)}, \boldsymbol{\rho}^{(b)}\}$, where $\mathbf{S}^{(b)}, \mathbf{Y}^{(b)} \in \mathbb{R}^{m \times n_{\max}}$ store the m most recent $\mathbf{s}$ and $\mathbf{y}$ vectors, and $\boldsymbol{\rho}^{(b)} \in \mathbb{R}^m$ stores the corresponding inverse curvatures $\rho_i = 1/(\mathbf{y}_i^T \mathbf{s}_i)$. In our batched implementation, these are stored as tensors $\mathbf{S} \in \mathbb{R}^{B \times m \times n_{\max}}$, $\mathbf{Y} \in \mathbb{R}^{B \times m \times n_{\max}}$, and

$\boldsymbol{\rho} \in \mathbb{R}^{B\times m}$. The two-loop recursion is executed in parallel for all batches:

$$
\begin{aligned}
&\text{Backward pass:} \quad \mathbf{q}^{(b)} \leftarrow \nabla f_k^{(b)}, \\
&\text{for } i = k-1, \ldots, k-m: \quad \alpha_i^{(b)} = \rho_i^{(b)} (\mathbf{s}_i^{(b)} \cdot \mathbf{q}^{(b)}), \quad \mathbf{q}^{(b)} \leftarrow \mathbf{q}^{(b)} - \alpha_i^{(b)} \mathbf{y}_i^{(b)} \\
&\text{Scaling:} \quad \gamma^{(b)} = (\mathbf{y}_{k-1}^{(b)} \cdot \mathbf{s}_{k-1}^{(b)}) / (\mathbf{y}_{k-1}^{(b)} \cdot \mathbf{y}_{k-1}^{(b)}), \quad \mathbf{z}^{(b)} = \gamma^{(b)} \mathbf{q}^{(b)} \\
&\text{Forward pass: for } i = k-m, \ldots, k-1: \\
&\beta_i^{(b)} = \rho_i^{(b)} (\mathbf{y}_i^{(b)} \cdot \mathbf{z}^{(b)}), \quad \mathbf{z}^{(b)} \leftarrow \mathbf{z}^{(b)} + \mathbf{s}_i^{(b)} (\alpha_i^{(b)} - \beta_i^{(b)}) \\
&\text{Search direction:} \quad \mathbf{p}_k^{(b)} = -\mathbf{z}^{(b)}
\end{aligned}
\tag{1}
$$

All dot products and vector updates are computed using masked batched operations, ensuring that padding does not affect optimization trajectories. Each structure independently accumulates curvature information appropriate to its own potential energy surface topology.

Dynamic batch shrinking with history preservation. Structures converge at different rates depending on initial geometry quality and PES characteristics. Rather than waiting for all structures to converge, our optimizer dynamically removes converged structures from the batch at each iteration. When a subset of structures converges, the batch dimension B is reduced via index selection, and all history tensors $\{\mathbf{S}, \mathbf{Y}, \boldsymbol{\rho}\}$ are correspondingly sliced to retain only unconverged structures:

$$
B \to B' = |\{\text{unconverged structures}\}|, \quad \mathbf{S} \to \mathbf{S}[\text{survive\_idx}], \quad \text{etc.} \tag{2}
$$

This shrinking operation maintains GPU occupancy by reducing wasted computation on structures that have already reached their optima.

The step size for each structure is independently clipped to prevent large displacements:

$$
\mathbf{p}_k^{(b)} \leftarrow \mathbf{p}_k^{(b)} \cdot \min\left(1, \frac{\Delta_{\max}}{\max |\mathbf{p}_k^{(b)}|}\right) \tag{3}
$$

where $\Delta_{\max} = 0.2$ Å is the maximum allowed atomic displacement per step. The initial inverse Hessian approximation is set to $H_0^{-1} = \gamma_0^{-1}\mathbf{I}$ with curvature parameter $\gamma_0 = 70$ Eh/Å^2, and the history depth is $m = 5$ vector pairs. All optimizations used double-precision (float64)

arithmetic to ensure numerical stability in force and gradient evaluations.

Dual-Shift Partitioned Rational Function Optimization

In standard P-RFO,[2,3] the optimization problem is partitioned into two subspaces: the transition mode (uphill direction, eigenvalue $\lambda_- < 0$) where energy is maximized, and the remaining modes (downhill directions, $\lambda_+ > 0$) where energy is minimized. The step $\mathbf{s}$ is constrained by a trust radius R via a single shift parameter α that scales both subspaces simultaneously to satisfy $\|\mathbf{s}\| \leq R$.

We developed a dual-shift variant (DS-P-RFO) that introduces independent control over each subspace. The total step $\mathbf{s}$ is decomposed as:

$$\mathbf{s} = \mathbf{s}_- + \mathbf{s}_+ \tag{4}$$

where $\mathbf{s}_- \in \text{span}\{v_j\}$ represents the step in the uphill subspace (typically one-dimensional for first-order saddle points), and $\mathbf{s}_+ \in \text{span}\{v_i : i \neq j\}$ represents the step in the downhill subspace.

The key innovation is decomposing the global trust-region constraint:

$$\|\mathbf{s}\|^2 = \|\mathbf{s}_-\|^2 + \|\mathbf{s}_+\|^2 \leq R^2 \tag{5}$$

into subspace-specific constraints with adaptive allocation. We define:

$$R_-^2 = \alpha R^2, \quad R_+^2 = (1-\alpha)R^2 \tag{6}$$

where $\alpha \in [0, 1]$ is determined by the relative magnitudes of unconstrained Newton steps in each subspace. Specifically, we compute:

$$\mathbf{s}_-^{\text{unc}} = -\frac{\mathbf{g}_-}{\lambda_-}, \quad \mathbf{s}_+^{\text{unc}} = -\frac{\mathbf{g}_+}{\lambda_+} \tag{7}$$

where $\mathbf{g}_-$ and $\mathbf{g}_+$ are the gradient projections onto the uphill and downhill subspaces, respec-

tively. The allocation parameter is then:

$$\alpha = \frac{\|\mathbf{s}_-^{\text{unc}}\|^2}{\|\mathbf{s}_-^{\text{unc}}\|^2 + \|\mathbf{s}_+^{\text{unc}}\|^2} \tag{8}$$

with bounds $\alpha \in [0.05, 0.95]$ to prevent extreme allocations. This heuristic automatically dedicates larger trust-region budgets to subspaces with stronger gradient driving forces or shallower curvature.

For each subspace, we solve independently for shift parameters μ_- and μ_+ via bisection search. In the uphill subspace (maximization):

$$\sum_{i \in \mathcal{M}_-} \left(\frac{g_{-,i}}{\lambda_{-,i} - \mu_-} \right)^2 = R_-^2 \tag{9}$$

and in the downhill subspace (minimization):

$$\sum_{i \in \mathcal{M}_+} \left(\frac{g_{+,i}}{\lambda_{+,i} - \mu_+} \right)^2 = R_+^2 \tag{10}$$

where the sums are over eigenmode indices in each subspace. The step components are then:

$$s_{-,i} = -\frac{g_{-,i}}{\lambda_{-,i} - \mu_-}, \quad s_{+,i} = -\frac{g_{+,i}}{\lambda_{+,i} - \mu_+} \tag{11}$$

This approach eliminates the iterative micro-cycles required in traditional RS-P-RFO, offering improved computational efficiency while maintaining numerical stability through independent optimization of physically distinct objectives.

Dimer Method with Hessian-Vector Products

For saddle point searches without good initial guesses, we employed a modified dimer method[4] that leverages automatic differentiation to avoid numerical instability from finite-difference approximations. The dimer method is a minimum-mode following algorithm that locates first-order saddle points using only gradient information, making it particularly suitable for large systems where full Hessian calculations are prohibitive.

The dimer consists of two images separated by distance 2Δ along a unit vector $\mathbf{n}$, positioned at $\mathbf{R}_0 \pm \Delta\mathbf{n}$. The curvature along the dimer axis is approximated by:

$$\kappa = \mathbf{n}^T \mathbf{H} \mathbf{n} \approx \frac{\mathbf{F}^+ - \mathbf{F}^-}{2\Delta} \tag{12}$$

where $\mathbf{F}^{\pm}$ are forces at the two dimer endpoints. Standard implementations use finite-difference with $\Delta \approx 0.005$ Å, requiring two force evaluations per rotation step to estimate the Hessian-vector product $\mathbf{H}\mathbf{n}$.

We developed an automatic differentiation (autograd) implementation that computes $\mathbf{H}\mathbf{n}$ analytically via forward-mode differentiation:

$$\mathbf{H}\mathbf{n} = \nabla_{\mathbf{R}} \left(\nabla_{\mathbf{R}} E(\mathbf{R}) \cdot \mathbf{n} \right) \tag{13}$$

This approach eliminates the finite-difference parameter Δ and its associated numerical errors, while simultaneously providing forces $\mathbf{F}$ and energy E in a single backward pass. The implementation reduces the computational cost per iteration from two force evaluations (standard finite-difference) to one combined force + HVP evaluation.

Each dimer iteration consists of two phases:

Rotation phase: Minimize the curvature $\kappa = \mathbf{n}^T \mathbf{H} \mathbf{n}$ to align $\mathbf{n}$ with the lowest-curvature mode. The rotational force is:

$$\mathbf{F}_{\text{rot}} = \left(\mathbf{I} - \mathbf{n}\mathbf{n}^T \right) \mathbf{H}\mathbf{n} \tag{14}$$

The orientation is updated via gradient descent:

$$\mathbf{n}_{\text{new}} = \text{normalize} \left(\mathbf{n} - \alpha_{\text{rot}} \mathbf{F}_{\text{rot}} \right) \tag{15}$$

where $\alpha_{\text{rot}} = 0.5$ is the rotation step factor. Rotation converges when $\max |\mathbf{F}_{\text{rot}}| < 10^{-3}$ Eh/Å and $\text{RMS}(\mathbf{F}_{\text{rot}}) < 5 \times 10^{-4}$ Eh/Å.

Translation phase: Move along the effective force direction. The perpendicular force component is:

$$\mathbf{F}_{\perp} = \mathbf{F} - (\mathbf{F} \cdot \mathbf{n})\mathbf{n} \tag{16}$$

When $\kappa < 0$ (indicating the mode is uphill), we perform a parallel flip to accelerate escape from the convex region:

$$\mathbf{F}_{\text{trans}} = \begin{cases} \mathbf{F}_{\perp} - (\mathbf{F} \cdot \mathbf{n})\mathbf{n} & \text{if } \kappa < 0 \\ \mathbf{F}_{\perp} & \text{otherwise} \end{cases} \tag{17}$$

The position is updated with trust-region control:

$$\mathbf{R}_{\text{new}} = \mathbf{R} + \min\left(\alpha \mathbf{F}_{\text{trans}}, R_{\text{max}}\right) \tag{18}$$

where the step scaling α is adapted based on whether the step reaches the trust-radius boundary ($R_{\text{max}} = 0.15$ Å).

MAPLE Functionalities

Table S1: Summary of Current and Planned Functionalities in MAPLE

Category	Method	Description
Geometry Optimization	L-BFGS[a]	Batched GPU-accelerated quasi-Newton optimizer with dynamic batch shrinking
	RFO	Rational function optimization with adaptive trust-region control
	SD / CG	Steepest descent and conjugate gradient methods
	DIIS	Direct inversion in iterative subspace acceleration
Transition State Search	CI-NEB	Climbing-image nudged elastic band with IDPP interpolation and dynamic springs
	DS-P-RFO[a]	Dual-shift partitioned RFO with independent uphill/downhill trust-region control
	Dimer[a]	Minimum-mode following with analytic Hessian–vector products via autodiff
	GSM	Growing string method with adaptive path construction and CI refinement
	AutoNEB	Automated multi-step NEB for complex reaction pathway exploration
Intrinsic Reaction Coordinate (IRC)	GS	Gonzalez–Schlegel method with mass-weighted coordinates
	HPC	Hessian predictor–corrector
	EulerPC	Euler predictor–corrector
	LQA	Local quadratic approximation
Frequency & Thermochemistry	Hessian Analysis	Analytic (autodiff) or numerical (finite-difference) mass-weighted Hessian
	Thermochemistry	ZPE, enthalpy, entropy, Gibbs free energy; RRHO and quasi-RRHO[b] treatments
PES Scan	Relaxed / Rigid	1D, 2D, and 3D scans along bond, angle, or dihedral coordinates
Single Point	Energy / Forces	Single-structure and batch trajectory energy and force evaluation
Molecular Dynamics	NVE	Velocity Verlet integrator for microcanonical ensemble
	NVT / NPT	Canonical and isothermal–isobaric ensembles with thermostats
Corrections & Solvation	DFT-D4	Dispersion correction with PyTorch automatic differentiation
	GBSA	Generalized Born surface area implicit solvation (OBC-II model)
	Explicit Solvent	Built-in library of pre-equilibrated explicit solvent environments

[a] Algorithm described in the Methods section.

Benchmark Dataset

The 100 reactions used for benchmarking were drawn by uniform random sampling (seed = 42) from the full Transition1x dataset[5] (10,073 reactions), with no additional filtering criteria applied. All cases are unimolecular rearrangements of neutral, closed-shell singlet molecules containing C, H, N, and O, spanning 43 unique molecular formulas with 6–19 atoms per molecule (mean 13.7, median 14). This composition is representative of the full Transition1x dataset, which was constructed by identifying spontaneous bond-breaking/forming events during molecular dynamics simulations of small organic molecules. A detailed breakdown of the elemental composition is provided in Table S2.

For the 100-reaction benchmark, the reactant and product structures were taken directly from the Transition1x dataset without further optimization; the same endpoint geometries were used for both DFT (ORCA) and MLFF (MAPLE) CI-NEB calculations.

Table S2: Composition of the 100-Reaction Benchmark Test Set

Elemental Composition	Count	Percentage
C, H, O	42	42%
C, H, N, O	33	33%
C, H, N	20	20%
C, H (hydrocarbons)	4	4%
C, N	1	1%
Total	100	100%

All reactions are unimolecular rearrangements of neutral, closed-shell singlet molecules. Molecular sizes range from 6 to 19 atoms (mean 13.7, median 14), with 3–6 carbon atoms per molecule. The 100 reactions span 43 unique molecular formulas and were drawn by uniform random sampling (seed = 42) from the Transition1x dataset (10,073 reactions).

Transition State Optimization

Table S3: TS Optimization Performance Starting from MLFF-Derived HEI Structures, updated to include MACE-POL-S and MACE-POL-L.

Potentials[a]	RMSD[b] (Å)	ΔIter[c]
ANI-1xnr (anal.)	0.61 ± 0.19	21.9
ANI-1xnr (num.)	0.61 ± 0.19	20.2
AIMNet2-NSE (anal.)	0.52 ± 0.14	45.1
AIMNet2-NSE (num.)	0.46 ± 0.14	26.2
UMA (num.)	$\mathbf{0.13 \pm 0.14}$	**2.8**
MACE-POL-S (anal.)	0.27 ± 0.27	6.6
MACE-POL-S (num.)	0.28 ± 0.28	4.4
MACE-POL-L (anal.)	0.18 ± 0.31	3.6
MACE-POL-L (num.)	0.20 ± 0.33	3.6

[a]Frequency calculation methods: anal. = analytic frequencies via automatic differentiation; num. = numerical frequencies via finite differences. [b]Geometric deviation from DFT-optimized TS structures when TS optimization is restarted from MLFF-derived highest-energy image (HEI) structures. [c]Iteration difference relative to DFT (MLFF − DFT).

Table S4: Decomposition of imaginary-frequency count mismatches relative to DFT reference TS calculations.

Potentials[a]	Acc.[b]	$\lvert\Delta\nu_{\min}\rvert$[c]	N_+^d	Extra $-100 < \nu < 0$[e]	Extra $\nu \leq -100$[f]	N_-^g
ANI-1xnr (anal.)	6.2%	291 ± 265	90	48 (22%)	172 (78%)	0
ANI-1xnr (num.)	6.2%	292 ± 265	90	48 (22%)	171 (78%)	0
AIMNet2-NSE (anal.)	31.2%	529 ± 600	62	25 (22%)	89 (78%)	4
AIMNet2-NSE (num.)	32.3%	521 ± 567	62	25 (22%)	90 (78%)	3
UMA (num.)	86.5%	75 ± 109	11	7 (64%)	4 (36%)	2
MACE-POL-S (anal.)	47.9%	182 ± 187	46	27 (40%)	40 (60%)	4
MACE-POL-S (num.)	47.9%	182 ± 187	46	27 (40%)	40 (60%)	4
MACE-POL-L (anal.)	85.4%	54 ± 56	13	9 (69%)	4 (31%)	1
MACE-POL-L (num.)	86.5%	55 ± 55	12	9 (69%)	4 (31%)	1

[a]Same methods as in the Hessian-quality benchmark. [b]Percentage of cases with the correct number of imaginary frequencies relative to the DFT reference. [c]Mean absolute difference in the most negative imaginary frequency, computed only for cases where both the DFT and MLFF spectra contain at least one imaginary mode; the corresponding sample size is $n = 95$ for ANI-1xnr, AIMNet2-NSE, UMA, and MACE-POL-L, and $n = 93$ for MACE-POL-S. Units are cm^{-1}. [d]Number of cases in which the MLFF predicts more imaginary frequencies than DFT. [e,f]When $N_{\mathrm{ML}} > N_{\mathrm{DFT}}$, the additional imaginary modes are defined as the shallowest $N_{\mathrm{ML}} - N_{\mathrm{DFT}}$ modes in the MLFF spectrum after sorting by $|\nu|$. Entries report the total number of such extra modes, with percentages relative to the total number of extra imaginary modes for that method. [g]Number of cases in which the MLFF predicts fewer imaginary frequencies than DFT.

Kemp Eliminase Simulation

Protein Sequence and Structure

The amino acid sequence was extracted directly from the PDB structure (PDB ID: 5D2W) and is provided below in FASTA format:

MALAKRIIAALIVKDGRVVKGSNFENLRDSGDPVELGKFYSEIGIDELSFWDITAS
VEKRKTMLELVEKVAEQIDIPFTVGGGIHDFETASELILRGADKVEINTAAVENPSLIT
QIAQTFGSQAVVVYIAAKRVDGEFMVFTYSGKKNTGILLRDWVVEVEKRGAGEIVL
GSIDRLGTKSGYDTEMIRFVRPLTTLPIIAHGGAGKMEHFLEAFLAGADAAKANSVLH
FREIDVRELKEYLKKHGVNVRLEGLGSLEHHHHHH

Table S5: Experimental and Computational Barrier Heights for Kemp Eliminase Variants

Design	k_{cat} (s^{-1})	$\Delta G^{\ddagger}$ (kcal/mol)			Mutations[a]
		Obs.[b]	EVB[c]	MLFF[c]	
KE07	0.018 ± 0.001	20.1	19.5	17.9	WT
R6 3/7F	0.60 ± 0.070	17.8	16.1	17.8	I7D, K19E, K146T, G202R, N224D
R7 2/5B	1.20 ± 0.08	17.6	18.0	15.5	I7D, F77I, G202R, N224D
R7 1/3H	0.76 ± 0.03	17.8	17.0	17.0	I7D, V12L, F77I, H84Y, G202R, M207T, N222D, F229S

[a]Mutations relative to KE07 wild-type. [b]Observed activation barrier from experimental k_{cat} measurements. [c]Calculated activation barrier from computational method.

Table S6: Flexible Residue Definitions for Kemp Eliminase Simulations

Design	Flexible Residues[a]	Count
KE07 (WT)	9, 10, 11, 48, 49, 50, 51, 52, 80, 81, 101, 103, 128, 130, 144, 169, 171, 176, 201, 222, 224	21
R6 3/7F	9, 10, 11, 48, 49, 50, 51, 52, 80, 81, 101, 103, 128, 130, 144, 169, 171, 176, 201, 222, 224	21
R7 1/3H	9, 10, 11, 48, 49, 50, 51, 52, 80, 81, 101, 103, 128, 130, 144, 169, 171, 176, 201, 222, 224	21
R7 2/5B	9, 10, 11, 48, 49, 50, 51, 52, 80, 81, 101, 103, 128, 130, 144, 169, 171, 176, 201, 222	20

[a]Residues within 8.0 Å of ligand center were defined as flexible and allowed to optimize during simulations. All other residues were kept fixed.

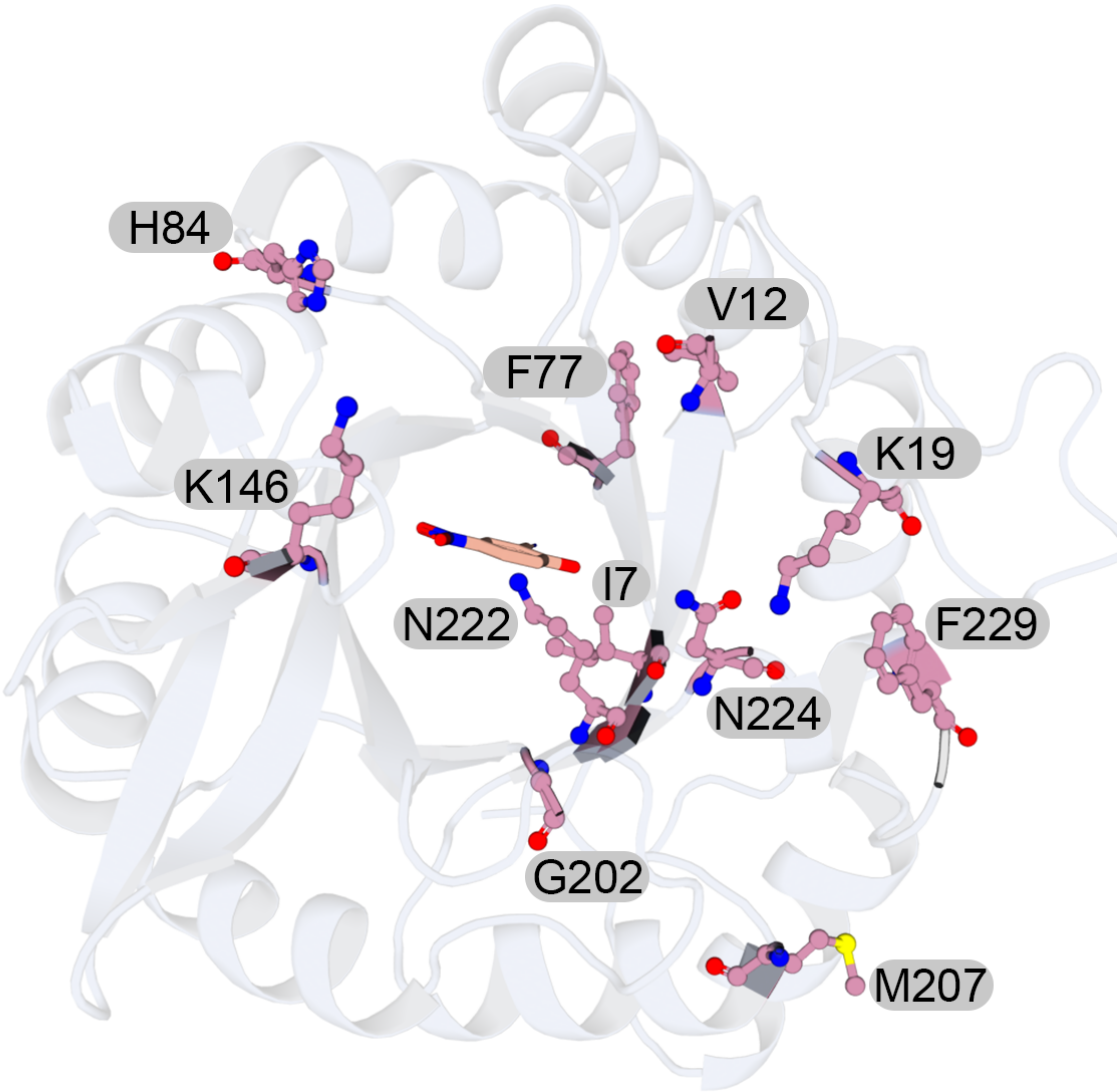

Figure S1: Overall structure of the wild-type KE07 Kemp Eliminase. Residues corresponding to mutation sites are highlighted in pink for reference, while the underlying structure remains the wild-type conformation.

Diels-Alderase Simulation

Protein Sequence and Structure

Protein structure of MaDAs are obtained from previous studies.[6,7]

Protein sequence for MaDA-1: THEAFLECLTTRIPSNSTFTPQSIIYTPDNPSYSTILDS
TTQNPRFLSSSTRNPFAIITPLHASHIQAALYCSQKHGEQMRIRSGGDYEGLSYQSSVP
FFILDLRNLSSISIDAKSKSAWVQAGATIGELYYGIAKTSLNLSFPGGVAHTIGVGGQL
GGGGYGYSTRKYGLASDNVIDAQLIDARGRILDRKTMGEDLFWAIRGGGAGSFGIVL
AWKIRLVNTPSTVTIFEAVRSWENNTTKKFIRRYQRRASKTDKDLTIFVGFRTTSSTD
EEGNERISILTIVSATFHGSKDRLLQLVQKEFPDLGLVSEECTEMSWVRSIIHFNLFGDE
VPLEVLLNRTLNFEMKAFKLRSDYVQKPIPDDVLEKLLSKLYDEETGXGYIEFFPYGG
KMSKISESEIPFPYRAGNLYNLRYMVSWKDDGNITRTNMHLSWIKDAYDYMTPYVS
KDPRGAYLNFRDLDIGVNVNESDYDYVAKASVWGTKYFRNNFYRLVDIKTIVDPTN
FFKYEQSIPPLPPL

Protein sequence for MaDA-3: HESFLECLTTRISKSNSTSTPESIIYTKDNPSYSTILNS
TSLNPRFFPSSARYPLLIVTPLHASHVQATVHCAKKHGIQIRIRSGGDYEGLSYMSNVT
FAIVDLRNLSSIDVDVKRKSAWVQSGATLGELHYWIAKKSQNLAFPGVVGHTVGIGG
MLGAGGYGYSSRKYGLSADNILDAQLIDVRGRILNRKSMGEDLFWAIRGGGAGSFGI
VLAWKVRLVDVPSKVTVFTAIRDWDNNATKKFIHRYQRRIAKVDKDLTIIVRFLTAS
ITDEKGSKKIQISTFITATYHGSQDRLLSLMEKEFPELGLLAKECAEGAWVQSILYFNL
LTNSKSLDVLLNRTLNFEWRAFKIKSDYLKKPIPDQVLENLLVKLYEEDIGETFVEFFP
YGGKLDEISESEIPCPHRAGNLYNLRYMVLWKEGQNATAVNKHLSWIRRAYNYMTP
YVSKNPRGAFLNFRDLDIGTNPNENEINGAYNYIKQASNWGTKYFKNNFYKLIYVKT
IVDPTNFFTYEQSIPSLLPH

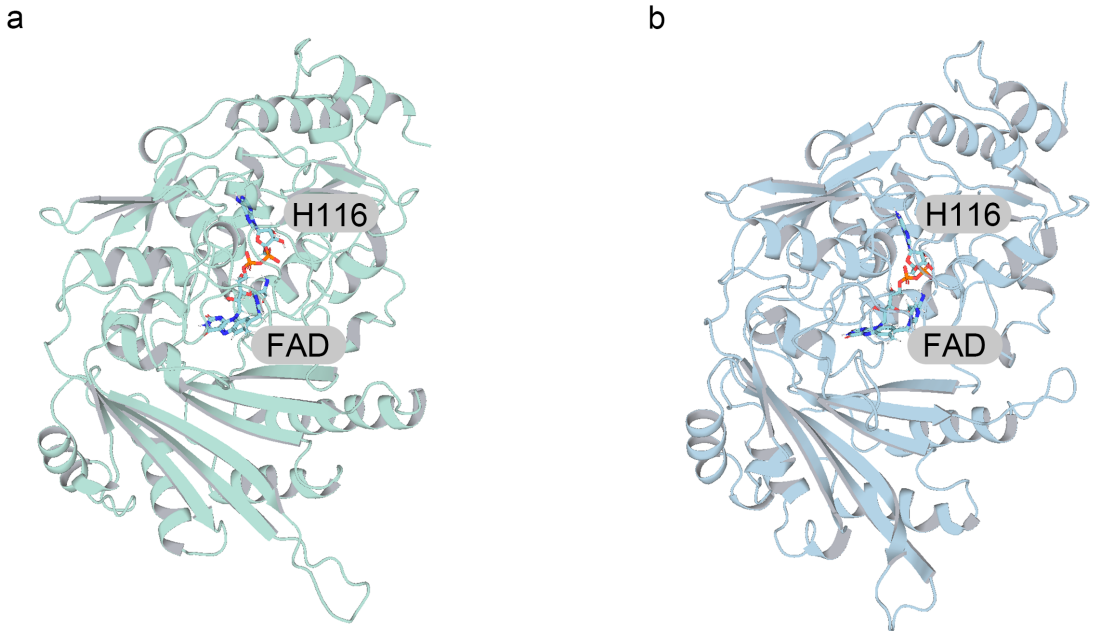

Figure S2: Overall structures of (**a**) MaDA-1 and (**b**) MaDA-3 showing the covalent linkage between residue H116 and the flavin adenine dinucleotide (FAD).

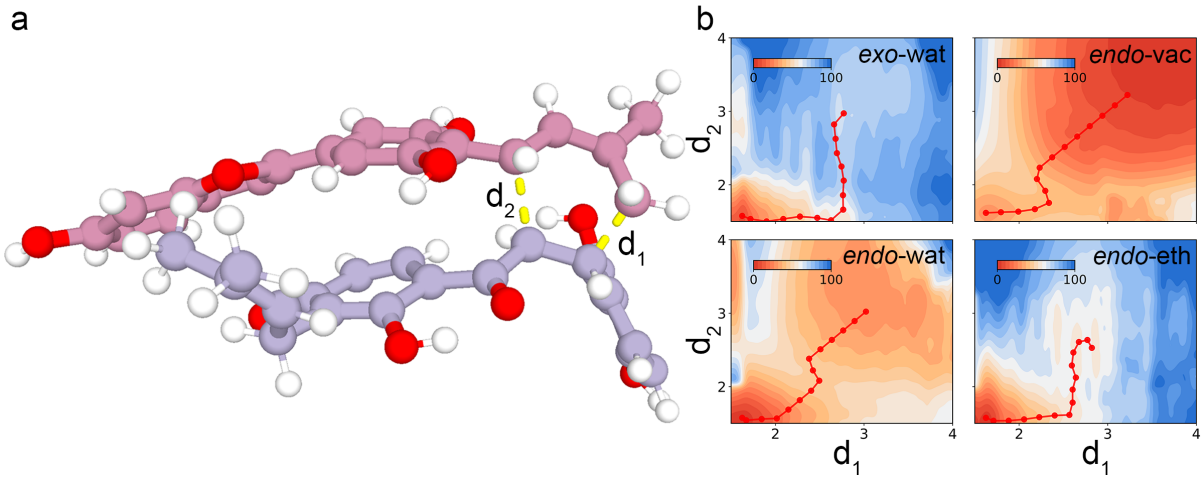

Figure S3: Two-dimensional potential energy surface (PES) analysis of the Diels-Alder reaction under different environments. (a) Definition of the two forming C–C bond distances, d_1 and d_2, used as collective variables for the PES scans, illustrated on a representative reactive geometry. (b) Two-dimensional PESs projected onto the two forming C–C bond distances d_1 and d_2 (in Å) for the competing exo and endo pathways in vacuum, water, and ether. Color contours represent relative energies (kcal mol^{-1}), and red lines denote the minimum-energy paths determined using a minimum-energy path search algorithm.

Table S7: Flexible Residue Definitions for Diels-Alder Reaction Simulations

System	Flexible Residues[a]	Count
mada1	44, 45, 46, 47, 89, 91, 148, 149, 150, 151, 152, 153, 154, 159, 163, 164, 165, 166, 167, 168, 169, 174, 251, 263, 264, 265, 266, 269, 289, 290, 291, 326, 327, 328, 329, 330, 331, 332, 333, 342, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 382, 383, 384, 385, 387, 388, 389, 390, 391, 414, 415, 416, 417, 418, 419, 420, 421, 422, 458, 459	74
mada3[b]	42, 43, 44, 45, 46, 88, 89, 90, 91, 96, 130, 148, 149, 150, 151, 152, 153, 154, 165, 166, 167, 233, 234, 264, 265, 266, 267, 269, 287, 288, 289, 290, 325, 326, 327, 328, 329, 330, 331, 332, 333, 347, 348, 349, 350, 351, 352, 353, 354, 355, 382, 383, 384, 385, 386, 387, 388, 389, 390, 416, 417, 418, 419, 420, 421, 458, 459	67

[a]Residues within 8.0 Å of ligand center were defined as flexible and allowed to optimize during simulations. All other protein framework residues were kept fixed. [b]The same flexible residue set was used for mada3 and all its variants.

Minimum Energy Path Determination

The minimum energy path (MEP) connecting reactant and product configurations on the 2D free energy surface was determined using a Dijkstra-based algorithm combined with path smoothing and uniform resampling. The free energy surface $F(x, y)$ was represented on a regular 2D grid, with off-grid values obtained via bilinear interpolation.

The classic Dijkstra shortest-path algorithm[8] was adapted to minimize the maximum energy barrier rather than total path length. Starting from the reactant minimum, the algorithm explores neighboring grid points using 8-connectivity (including diagonal moves) and maintains for each point the lowest barrier B_{max} encountered along any path from the start. The search terminates upon reaching the product minimum, and the optimal path is reconstructed via backtracking. This approach guarantees finding the globally optimal minimum-barrier pathway on the discrete grid with $O(N \log N)$ complexity, where N is the number of grid points.

The raw Dijkstra path exhibits staircase artifacts due to grid discretization. We applied cubic B-spline smoothing ($k = 3$) to obtain a continuously differentiable path, parameterized by cumulative arc length. The smoothed path was then uniformly resampled to N_{path} points distributed with equal arc-length spacing:

$$s_i = \frac{i-1}{N_{\text{path}} - 1} \cdot L_{\text{total}}, \quad i = 1, \ldots, N_{\text{path}} \tag{19}$$

where L_{total} is the total path length. This uniform spacing ensures consistent resolution along the entire reaction coordinate. Finally, the transition state location was refined by local grid search in a small neighborhood around the maximum-energy point to accurately capture the barrier height. For all calculations, $N_{\text{path}} = 15$ images were used, providing sufficient resolution for subsequent nudged elastic band optimizations while maintaining computational efficiency.

The forward activation barrier $\Delta F^{\ddagger}_{\text{fwd}}$ and reverse barrier $\Delta F^{\ddagger}_{\text{rev}}$ were computed as the energy difference between the transition state and reactant/product minima, respectively. The reaction free energy is $\Delta F_{\text{rxn}} = F_{\text{product}} - F_{\text{reactant}}$.

DNA Photolyase Simulation

The amino acid sequence and DNA sequence were extracted directly from the PDB structure (PDB ID: 3CVU) and are provided below in FASTA format:

DNA Chain A: ACAGCGGTTGCAGGT

DNA Chain B: TACCTGCAACCGCTG

Protein sequence: MASWSHPQFEKGASTSLYKKAGLMDSQRSTLVHWFRKGLRLH
DNPALSHIFTAANAAPGRYFVRPIFILDPGILDWMQVGANRWRFLQQTLEDLDNQLR
KLNSRLFVVRGKPAEVFPRIFKSWRVEMLTFETDIEPYSVTRDAAVQKLAKAEGVRV
ETHCSHTIYNPELVIAKNLGKAPITYQKFLGIVEQLKVPKVLGVPEKLKNMPTPPKDE
VEQKDSAAYDCPTMKQLVKRPEELGPNKFPGGETEALRRMEESLKDEIWVARFEKP
NTAPNSLEPSTTVLSPYLKFGCLSARLFNQKLKEIIKRQPKHSQPPVSLIGQLMWREFY
YTVAAAEPNFDRMLGNVYCMQIPWQEHPDHLEAWTHGRTGYPFIDAIMRQLRQEG
WIHHLARHAVACFLTRGDLWISWEEGQRVFEQLLLDQDWALNAGNWMWLSASAFF
HQYFRVYSPVAFGKKTDPQGHYIRKYVPELSKYPAGCIYEPWKASLVDQRAYGCVL
GTDYPHRIVKHEVVHKENIKRMGAAYKVNREVRTGKEEESSFEEKSET

In the simulations, residues 288, 391, 394, 421, 423, 424, 425, 428, 429, 532, 8, and 9 were defined as flexible and allowed to optimize, while all DNA nucleotides except the pyrimidine(64)pyrimidone (64)-PP lesion site were kept fixed.

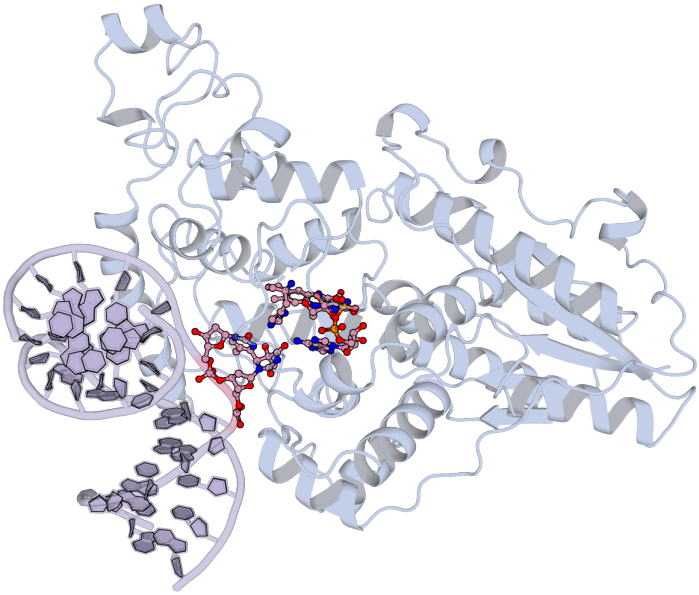

Figure S4: Overall structures of DNA Photolyase with DNA double strands.

Multi-Step Reaction Pathway Construction

The three-step repair mechanism (PCET, hydroxyl transfer, and C–C bond cleavage) was modeled as sequential elementary reactions, each constructed and optimized individually. The general workflow for each step is as follows:

(i) Endpoint definition. Reactant and product structures for each elementary step were defined based on the established mechanism from QM/MM studies[9]. For the first step, the globally optimized enzyme-substrate complex served as the reactant; for subsequent steps, the optimized product of the preceding step was used. (ii) CI-NEB pathway optimization. CI-NEB calculations were performed using 10 intermediate images with the settings described above. When the optimization did not converge within the maximum number of iterations, the last set of images was extracted and used as the starting point for continued optimization with reset L-BFGS history. (iii) Pathway inspection and refinement. After each CI-NEB run, the converged pathway was visually inspected. If individual images deviated significantly from the expected minimum energy path, for example, exhibiting unphysical bond lengths, steric clashes, or discontinuous geometric changes relative to neighboring images, these problematic images were identified and removed. The remaining images were then used to reinitialize a new CI-NEB calculation with interpolated replacements. This iterative refinement was repeated until a smooth, physically reasonable pathway was obtained. (iv) Energy profile assembly. The converged pathways from all three steps were concatenated to produce the overall reaction energy profile, with the endpoint energies of consecutive steps matched to ensure continuity.

This workflow requires manual intervention and chemical judgment at several stages, reflecting the current state of the art for modeling complex multi-step enzymatic mechanisms. Automating multi-step reaction discovery and pathway validation is an active development direction for MAPLE.

Enzymatic Systems Calculation Performance

Table S8: Wall-clock time for a single L-BFGS optimization step on the three enzymatic systems using the UMA model. All calculations were performed on a single NVIDIA A100 80 GB GPU with an AMD EPYC 7742 CPU.

System	Num. of Atoms	GPU Memory (MB)	Wall Time Per Round (s)
Kemp Eliminase (R7-1)	3,936	15,309	145.0
Diels–Alderase (MaDA1)	8,351	18,696	170.8
DNA Photolyase	9,240	18,943	175.8

Computational Details

Climbing-image Nudged Elastic Band Calculation

Ab initio climbing-image nudged elastic band (CI-NEB) calculations were performed using the ORCA-6.0 program package. All NEB calculations were carried out at the ωB97M-V/def2-TZVPD level of theory.[10,11] The reactant and product structures were provided as Cartesian coordinates, and the reaction pathway was discretized into 10 intermediate images connecting the reactant and product minima. Initial NEB pathways were generated by linear interpolation of Cartesian coordinates optimized by the Image-Dependent Pair Potential (IDPP) method[12] to produce physically reasonable geometries, particularly for reactions involving significant bond rearrangements. Geometry optimizations along the reaction pathway were performed using the CI-NEB algorithm to locate the minimum-energy path and corresponding transition-state region. All calculations employed a closed-shell singlet electronic configuration. All optimizations were performed using the default convergence criteria implemented in ORCA.

MAPLE CI-NEB calculations were performed using a fixed number of intermediate images (10, excluding the end points) and optimized with a limited-memory BFGS (L-BFGS) algorithm. Dynamic spring constants were employed, with values ranging from 0.03 to 0.3 Hartree Å^{-2} and a decay factor of 0.5, to improve path resolution near the saddle-point region. Initial reaction pathways were generated using the image-dependent pair potential (IDPP) method. NEB optimizations were converged based on projected forces, with thresholds of 9.5×10^{-3} and 5.0×10^{-3} Hartree Å^{-1} for the maximum and root-mean-square projected forces, respectively. A maximum of 500 optimization steps was allowed, using an L-BFGS history size of 20 and an initial step length of 2×10^{-2} Å. When CI-NEB refinement was applied, stringent force criteria were used together with a reduced initial step size to accurately locate the highest-energy image.

Frequency Analysis

Frequency analyses were performed in ORCA using numerical differentiation, with all other settings kept at their default values.

In MAPLE, harmonic vibrational frequencies were computed from the Cartesian Hessian expressed in Hartree Å^{-2}. Prior to diagonalization, the six external degrees of freedom corresponding to overall translations and rotations were removed by projection in mass-weighted coordinates. The projected Hessian was constructed as

$$\mathbf{H}_{\text{proj}} = \mathbf{PHP}, \qquad \mathbf{P} = \mathbf{I} - \mathbf{QQ}^{\mathsf{T}}, \tag{20}$$

where $\mathbf{Q}$ denotes the orthonormalized mass-weighted translationrotation basis.

The projected Hessian was then mass-weighted and diagonalized to obtain eigenvalues $\{\lambda_k\}$, which were converted to harmonic vibrational frequencies according to

$$\tilde{\nu}_k = \text{sgn}(\lambda_k)\sqrt{|\lambda_k|}\, C, \tag{21}$$

with $C = 2721.1383$ converting eigenvalues in (Hartree Å^{-2})/amu to frequencies in cm^{-1}. Normal modes were normalized in the mass-weighted metric following the ORCA convention.

The quality of Hessians was further evaluated by comparing the normal modes associated with the lowest imaginary frequency against DFT references. Two complementary metrics were used: the angular deviation and the mass-weighted mode overlap.

The mass-weighted overlap between an MLFF normal mode $\mathbf{Q}_{\text{ML}}$ and the corresponding DFT mode $\mathbf{Q}_{\text{DFT}}$ was defined as

$$\text{Overlap} = \frac{\left|\mathbf{Q}_{\text{DFT}}^{\mathsf{T}}\mathbf{M}\mathbf{Q}_{\text{ML}}\right|}{\sqrt{\left(\mathbf{Q}_{\text{DFT}}^{\mathsf{T}}\mathbf{M}\mathbf{Q}_{\text{DFT}}\right)\left(\mathbf{Q}_{\text{ML}}^{\mathsf{T}}\mathbf{M}\mathbf{Q}_{\text{ML}}\right)}}, \tag{22}$$

where $\mathbf{M}$ is the diagonal mass matrix and Q denotes the eigenvector of the mass-weighted Hessian matrix corresponding to a vibrational normal mode. The absolute value in the numerator is taken to eliminate the arbitrary sign of the normal-mode eigenvectors, ensuring that the overlap reflects physical similarity between modes.

The corresponding angular deviation was computed as

$$\theta = \arccos(\text{Overlap}), \tag{23}$$

where θ is reported in degrees.

Transition State Optimization

Transition-state optimizations in ORCA were performed starting from the converged CI-NEB highest-energy image. Numerical Hessians were first evaluated at the ωB97M-V/def2-TZVPD level of theory and the numerical Hessian was read into ORCA as the initial Hessian guess. Subsequent transition-state optimizations were performed using ORCAs default settings.

Transition-state optimizations in MAPLE were carried out using a trust-regionbased partitioned rational function optimization (P-RFO) algorithm. A dual-shift formulation with mode following was employed to explicitly locate first-order saddle points, where the reaction coordinate was tracked by maximizing overlap between successive eigenmodes of the mass-weighted Hessian. Hessians were obtained directly from the underlying MLFF calculators and used to construct the quadratic model at each optimization step. The trust-region radii were dynamically adjusted based on the agreement between predicted and actual energy changes. The optimization was terminated when both force- and displacement-based convergence criteria were satisfied, with thresholds defined on the maximum and root-mean-square values. Unless otherwise stated, all remaining algorithmic parameters were kept at their default MAPLE settings.

Explicit Solvent Potential Energy Surface Scan

Reactant structures were taken from previous reports[6] and initially optimized in MAPLE using the L-BFGS algorithm with default convergence criteria. Starting from the optimized geometry, the two forming CC bond distances were manually elongated to 4.0 Å and subsequently constrained during geometry optimization to generate the initial scanning structure. This structure was directly used for gas-phase potential energy surface (PES) scans.

To best account for the solvent effect, explicit solvent models were employed. Solvent boxes were constructed using the GAFF[13] with ABCG2 charges.[14] Each system was first subjected to energy minimization followed by a 1 ns NVT molecular dynamics simulation to ob-

tain equilibrated solvent configurations. All solvent molecules within a $35 \times 35 \times 35$ Å^3 cube were extracted to form an equilibrium solvent box for the solvent. To accelerate the solvation process, MAPLE provides a library of nearly one hundred such pre-equilibrated solvents representing diverse environments.

Explicit solvent models were built using the MAPLE solvation builder, where solvent molecules were placed within a 10 Å radius from the geometric center of the reactant. Solvent molecules located beyond 8 Å were kept fixed to mimic the bulk environment, while the reactant and solvent molecules within 8 Å were allowed to relax. Solvent molecules exhibiting atomic clashes within 1.5 Å of the reactant were removed to eliminate any unrealistic forces that could compromise physically reasonable configurations. Constrained geometry optimizations were then performed to obtain the initial solvated structures in water and diethyl ether.

Two-dimensional PES scans were carried out by simultaneously varying the two CC bond distances from 4.0 to 1.5 Å with a step size of 0.05 Å, resulting in a total of $51 \times 51 = 2601$ sampling points.

Protein Structure Preparation

The raw PDB file was first processed using the `pdb4amber` module in AmberTools[15] to repair common structural issues, including missing atoms, nonstandard residue naming, and to add hydrogen atoms consistent with standard protonation states. This procedure also ensured a standardized and simulation-ready PDB format.

Based on the repaired wild-type structure, site-specific residue mutations were introduced using the `tleap` module in AmberTools, resulting in mutant variants. All mutations were applied while preserving the overall protein backbone conformation.

The corresponding small-molecule ligand present in the experimental protein–ligand complex was then incorporated into each system according to the crystallographic binding mode. The full protein–ligand complexes were subsequently imported into the MAPLE framework, where geometry optimizations of the entire systems were performed using the UMA machine-learning force field. Optimizations were carried out under superloose convergence criteria to efficiently relax the structures while maintaining the experimentally resolved binding geometry.

The optimized protein–ligand complexes obtained from these procedures were used as the initial structures for all subsequent catalytic reaction calculations.

References

(1) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **1989**, *45*, 503–528.

(2) Baker, J. An algorithm for the location of transition states. *Journal of Computational Chemistry* **1986**, *7*, 385–395.

(3) Besalú, E.; Bofill, J. M. On the automatic restricted-step rational-function-optimization method. *Theoretical Chemistry Accounts* **1998**, *100*, 265–274.

(4) Henkelman, G.; Jónsson, H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics* **1999**, *111*, 7010–7022.

(5) Schreiner, M.; Bhowmik, A.; Vegge, T.; Busk, J.; Winther, O. Transition1x—a dataset for building generalizable reactive machine learning potentials. *Scientific Data* **2022**, *9*, 779.

(6) Gao, L. et al. FADdependent enzymecatalysed intermolecular [4+2] cycloaddition in natural product biosynthesis. *Nat. Chem.* **2020**, *12*, 620–628.

(7) Gao, L.; Zou, Y.; Liu, X.; Yang, J.; Du, X.; Wang, J.; Yu, X.; Fan, J.; Jiang, M.; Li, Y.; Houk, K. N.; Lei, X. Enzymatic control of endo and exostereoselective DielsAlder reactions with broad substrate scope. *Nat. Catal.* **2021**, *4*, 1059–1069.

(8) Ensing, B.; Laio, A.; Parrinello, M.; Klein, M. L. A Recipe for the Computation of the Free Energy Barrier and the Lowest Free Energy Path of Concerted Reactions. *J. Phys. Chem. B* **2005**, *109*, 6676–6687.

(9) Faraji, S.; Groenhof, G.; Dreuw, A. Combined QM/MM Investigation on the Light-Driven Electron-Induced Repair of the (6–4) Thymine Dimer Catalyzed by DNA Photolyase. *J. Phys. Chem. B* **2013**, *117*, 10071–10079.

(10) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of Chemical Physics* **2016**, *144*, 214110.

(11) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297–3305.

(12) Smidstrup, S.; Pedersen, A.; Stokbro, K.; Jónsson, H. Improved Initial Guess for Minimum Energy Path Calculations. *J. Chem. Phys.* **2014**, *140*, 214106.

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

(14) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. ABCG2: A Milestone Charge Model for Accurate Solvation Free Energy Calculation. *J. Chem. Theory Comput.* **2025**, *21*, 3032–3043.

(15) Case, D. A. et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63*, 6183–6191.