JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

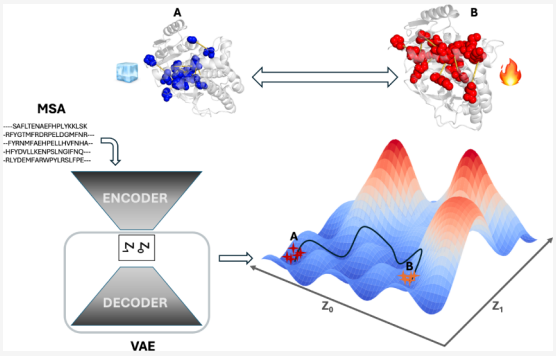
Thermal Adaptation of Cytosolic Malate Dehydrogenase Revealed by Deep Learning and Coevolutionary Analysis

Published as part of Journal of Chemical Theory and Computation special issue "Machine Learning and Statistical Mechanics: Shared Synergies for Next Generation of Chemical Theory and Computation".

Divyanshu Shukla, Jonathan Martin, Faruck Morcos, and Davit A. Potoyan*

Cite This: J. Chem. Theory Comput. 2025, 21, 3277−3287

ABSTRACT: Protein evolution has shaped enzymes that maintain stability and function across diverse thermal environments. While sequence variation, thermal stability and conformational dynamics are known to influence an enzyme's thermal adaptation, how these factors collectively govern stability and function across diverse temperatures remains unresolved. Cytosolic malate dehydrogenase (cMDH), a citric acid cycle enzyme, is an ideal model for studying these mechanisms due to its temperature-sensitive flexibility and broad presence in species from diverse thermal environments. In this study, we employ techniques inspired by deep learning and statistical mechanics to uncover how sequence variation and conformational dynamics shape patterns of cMDH's thermal adaptation. By integrating coevolutionary models with variational autoencoders (VAE), we generate a latent generative landscape (LGL) of the cMDH sequence space, enabling us to explore mutational pathways and predict fitness using direct coupling analysis (DCA). Structure predictions via AlphaFold and molecular dynamics simulations further illuminate how variations in hydrophobic interactions and conformational flexibility contribute to the thermal stability of warm- and cold-adapted cMDH orthologs. Notably, we identify the ratio of hydrophobic contacts between two regions as a predictive order parameter for thermal stability features, providing a quantitative metric for understanding cMDH dynamics across temperatures. The integrative computational framework employed in this study provides mechanistic insights into protein adaptation at both sequence and structural levels, offering unique perspectives on the evolution of thermal stability and creating avenues for the rational design of proteins with optimized thermal properties.



INTRODUCTION

Enzymes from different organisms have evolved to maintain an optimal balance of stability and functional activity under diverse thermal environments. While variations in sequence, structure and conformational dynamics influence an enzyme's thermal adaptation, the molecular rules that collectively govern stability and function across diverse temperatures remain unclear. Cytosolic malate dehydrogenase (cMDH), a well-characterized enzyme in the citric acid cycle, is an exemplary model system for investigating these mechanisms due to its temperature-sensitive conformational flexibility and widespread occurrence in species adapted to diverse thermal niches.[1−4] The homologues of cMDH display an impressive range of thermal adaptation profiles from approximately −20 °C to over 120 °C.[5,6] The information collected for cMDH enables detailed studies of specific sites within the sequence critical for ligand binding, catalysis, and subunit interactions. The conformational changes necessary for cMDH catalytic function are well understood, aiding the structure−function analyses.[2] A recent study[7] has identified key mobile regions (MRs) that undergo significant conformational changes, enabling ligand binding and catalytic activity. Dynamics of cytosolic MDH from orthologs from warm-adapted and cold-adapted species display distinct thermal responses.[5,8] For instance, cytosolic malate dehydrogenase (cMDH) from warm-adapted *Mytilus galloprovincialis* maintains activity and optimal substrate binding at higher temperatures than their cold-adapted counterparts from *Mytilus trossulus*. This is explained by strategic amino acid substitutions that dramatically enhance thermal stability.[5,9]

Numerous investigations have focused on comparing psychrophilic, thermophilic, and mesophilic proteins to pinpoint the molecular determinants of protein thermal

Received: December 31, 2024
Revised: March 6, 2025
Accepted: March 7, 2025
Published: March 13, 2025

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

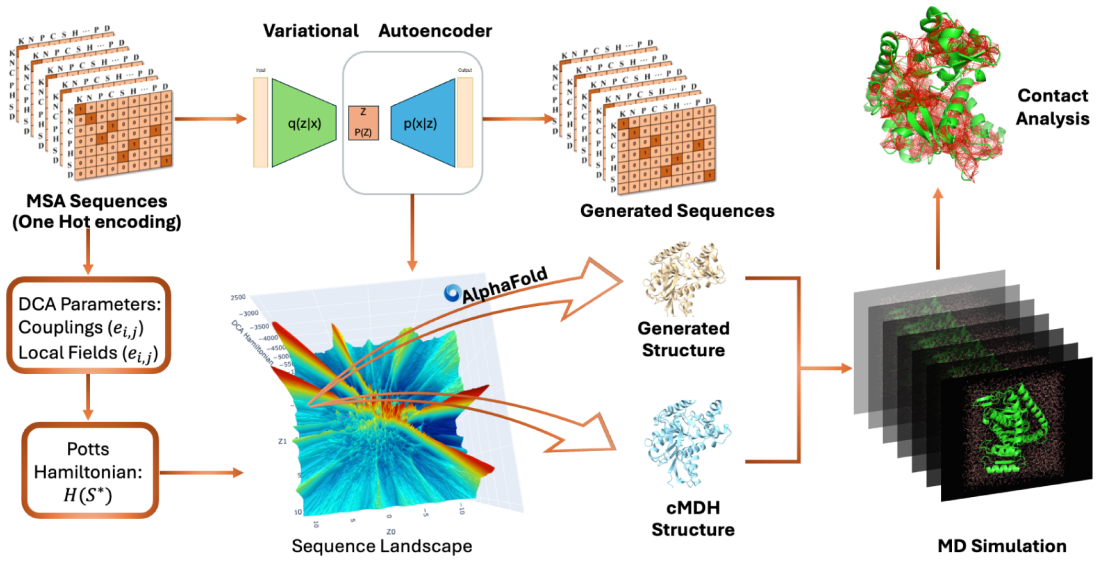


Figure 1. Workflow for cMDH landscape generation and structural analysis. The workflow begins with multiple sequence alignments (MSA) encoded using one-hot encoding. These sequences are then processed through a Variational Autoencoder (VAE) to produce a two-dimensional representation of the sequence space. Fitness is assessed using a Potts Hamiltonian, with parameters infrared from Direct Coupling Analysis (DCA) of MSA data. The combination of these features generates a comprehensive evolutionary landscape containing natural and synthetic sequences. Predicted sequences are input into AlphaFold for three-dimensional structure prediction, including the cytosolic variant of malate dehydrogenase (cMDH). Molecular dynamics simulations are subsequently used to explore the dynamic properties of the structures, with contact analysis revealing variations in hydrophobic contact networks.

stability.[10−12] Changes in noncovalent intramolecular interactions, including electrostatic interactions, hydrogen bonds, and hydrophobic contacts, have been found to correlate with varying thermal stabilities observed among homologous proteins.[13,14] Evolutionary changes, however, are known to be constrained by the fitness of an organism, which exerts selective pressures on the sequence, structure, and dynamics of proteins.[15−17] These constraints impose statistical signatures in the collection of evolutionarily related sequences, allowing structural and functional inferences from homologous sequence alignments, which is exploited in techniques collectively known as direct coupling analysis (DCA).[18−20] The inference of coevolutionary pairs in DCA has been shown to be crucial for the success of next-generation AI-based structure prediction techniques of AlphaFold and RoesttaFold, which can reconstruct 3D folds of single proteins[22−24] and protein complexes[18,20−22] from multiple sequence alignments. Co-evolutionary models have also proven to be key in inferring molecular specificity and the effects of protein mutations, thereby guiding the design of novel functional proteins such as transporters, fluorescent proteins, and enzymes.[23−25] Recent focus has shifted toward using state-of-the-art AI approaches in extending the predictive range of coevolutionary models toward functional relationships in protein families.[26,27] For instance, DeepPPI has been used to study protein interactions, demonstrating its effectiveness in identifying interaction networks.[28] Restricted Boltzmann machines (RBM) have been applied to detect motifs associated with functional roles in proteins, highlighting their utility in motif discovery.[29,30] Architectures such as variational autoencoders (VAE) and transformers have shown especially high promise in generating novel protein sequences, underlining their potential in protein design.[31] VAEs have been utilized for phylogenetic clustering and predicting the effects of protein mutations.[32]

In this work, we use the previously developed latent generative landscape method (LGL),[33] which combines the predictive power of coevolutionary models with the classification and generative power of the VAE, to aid us in the generation and modification of cMDH sequences in a manner that requires no labeling information. The LGL methodology assesses the learned protein sequence manifold of the VAE with the inferred fitness score provided by DCA and produces a surprisingly complex landscape of peaks and valleys, which have already been shown to distinguish distinct functional groups of proteins and capture interesting evolutionary phenomena like selection temperature and gain-of-function mutations. In this work, through combining the LGL with AlphaFold and MD simulations, we gain insights into how evolution tunes stability-dynamics balance, enabling functional diversification and potentially guiding generative enzyme design (Figure 1).

METHODS

Generation of Multiple Sequence Alignments. For the analysis of cMDH, multiple sequence alignments (MSAs) were initially procured using the HMMSearch tool against the UniProt database, employing the GREMLIN tool developed by Baker Lab.[34,35] The MSAs were further refined by excluding sequences with more than 10% gaps, ensuring the retention of high-quality sequence data.[36] After this filtration process, a total of 2,100 sequences remained, which were compiled into a FASTA file. This curated data set was subsequently utilized to train a Variational Autoencoder (VAE), facilitating the exploration of sequence variability and potential structural predictions.[33]

VAE Model Architecture. The variational autoencoder (VAE) is designed to generate data samples $x \epsilon X$ utilizing a latent variable model defined on parameters θ with a prior $p_\theta(z)$ on latent variables z. The marginal likelihood is represented as

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

$$p_\theta(x) = \int p_\theta(x|z)dp_\theta(z) \tag{1}$$

However, the parameters θ and the latent variables z are unknown, and assessing eq 1 requires computationally difficult calculations that scale poorly with the number parameters θ. The proposed solution, as outlined in ref 37, is to approximate the posterior distribution $p_\theta(z|x)$ with another model defined on parameters ϕ:

$$q_\phi(z|x) \approx p_\theta(z|x) \tag{2}$$

The model defined on ϕ is termed the encoder, while the model defined on θ is termed the decoder. Consequently, the marginal likelihood of generating a sample x through the decoder can be expressed as

$$\log p_\theta(x) = D_{KL}(q_\phi(z|x)||p_\theta(z|x)) + \zeta(\theta, \phi, x) \tag{3}$$

Here, D_{KL} represents the Kullback−Leibler divergence, quantifying the fit between the decoder's posterior distribution on z and the encoder's posterior. In contrast, the term ζ represents the lower bound of the model's fit to the marginal distribution over z. Equation 3 can be further simplified into the evidence lower bound function (ELBO):

$$\text{ELBO} = -E_{z\sim q_\phi(z|x)}\log(p_\theta(x|z)) + D_{\text{KL}}(q_\phi(z|x)||p_\theta(z)) \tag{4}$$

This serves as the objective function minimized during training. The reconstruction error, measuring the match between encoded and generated data, constitutes the first term. In contrast, the second term evaluates the similarity between the predicted latent distribution and an assumed prior distribution. We employ the reparameterization procedure[37] by defining the encoder model to represent sequences as Gaussian parameters μ and σ, which are combined with an auxiliary noise variable ϵ as follows:

$$z = \mu + \sigma \odot \epsilon \tag{5}$$

This reparameterized z forms the code that the decoder utilizes for sequence generation, allowing us to define $p_\theta(z)$ as a Gaussian distribution, thereby providing an analytical solution to the gradient of eq 4.

Data Representation and Decoding. In our specific implementation, data is represented as a one-hot encoded vector. The array is created for a protein of length, where each row contains a 1 in a position corresponding to an amino acid identity, with the remaining row positions containing 0. A total of 23 rows are employed to encode the 20 canonical amino acids, a gap character, and additional less common amino acids. The latent variables z are decoded into a Softmax probability distribution with dimensions identical to the input. The output layer ψ:$R^{23\times L}$, with each column corresponding to 23 sequence symbols $a \in A$, is defined as

$$p(a|z)_i = \frac{\exp(\psi_{a_i}(z))}{\sum_{k\in A}\exp(\psi_{k_i}(z))} \tag{6}$$

This yields L rows with probability values summing to one in each row. The reconstruction error term in eq 4 evaluates to zero if the input and output matrices are identical, indicating that the only possible sequence at some point z is the input sequence.

Hyperparameters and Training. The model architecture comprises two layers, each of which is an encoder and decoder. For the encoder, the first layer had $2L$ hidden units, while the second layer had L hidden units, where L is the sequence length. Conversely, the decoder's first layer contained L hidden units, and the second layer had $2L$ hidden units. ReLU activation functions were employed throughout. The Adam optimizer was utilized with a learning rate of 1×10^{-4}, and L2 regularization with a penalty of 1×10^{-4} was applied to the hidden units. The training was terminated if the reconstruction loss did not improve for 50 consecutive epochs. This model, featuring two latent encoding dimensions, was trained on computing clusters equipped with NVIDIA A100 GPUs.

Landscape Generation. For scoring the generated sequences by VAE, we used the Direct Coupling Analysis (DCA) method, which models protein sequences S of length L via Boltzmann distribution.

$$P(S) = \frac{1}{Z}\exp\left(\sum_{i<j} e_{ij}(A_i, A_j) + \sum h_i(A_i)\right) \tag{7}$$

The inferred parameters e_{ij} encode pairwise couplings between MSA positions and h_i the encoding frequency of amino acids at that position. In a grid-like fashion, equally spaced coordinates of latent space are fed into the VAE decoder to generate a decoded Softmax distribution. The maximum probability sequence from this output is generated as the final sequence. The generated sequence is then given a Hamiltonian score using the parameters obtained from a Boltzmann-like DCA distribution, defined as follows:

$$H(S^*) = -\sum_{1\le i\le j\le L} e_{ij}(a_i, a_j) - \sum_{i=1}^{L} h_i(a_i) \tag{8}$$

where $S^* = a_i \cdot a_L$ and $a_i = \underset{a\in A}{\text{argmax}}\, p(a|z)_i$

Protein Structure Prediction with AlphaFold. We predict the wild-type (WT) and mutant protein structures based on their sequences using ColabFold,[38] an open-source alternative to AlphaFold2. ColabFold leverages the "Evoformer" module, which integrates multiple sequence alignments and equivariant attention architecture to accurately predict the 3D coordinates of all heavy atoms in a given protein sequence. This tool also provides the predicted local-distance difference test (pLDDT) score, which evaluates the prediction's accuracy on a scale from 0 to 100, with higher scores indicating better performance.[38,39] Using the default parameters, we imported the previously collected WT and mutant protein sequences into the ColabFold model to obtain the corresponding PDB coordinate files. This process facilitated the generation of high-confidence structural models essential for our subsequent analyses.

Hydrophobic Contacts and Network Analysis. To calculate interaction fingerprints (IFPs) from the molecular dynamics (MD) simulation and structures generated from AlphaFold, we utilized ProLIF (v1.1.0) in conjunction with RDKit (v2021.03.5) and MDAnalysis (v2.4.0), following the procedures outlined in the ProLIF documentation.[40−43] This setup enabled the extraction of detailed interaction patterns between molecular structures.

Hydrophobic Contact Analysis in Static Structures. The CSV files generated from IFP calculations were further processed to evaluate hydrophobic contacts. For static structures, the total number of hydrophobic contacts was normalized by sequence length, yielding per-residue hydro-

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

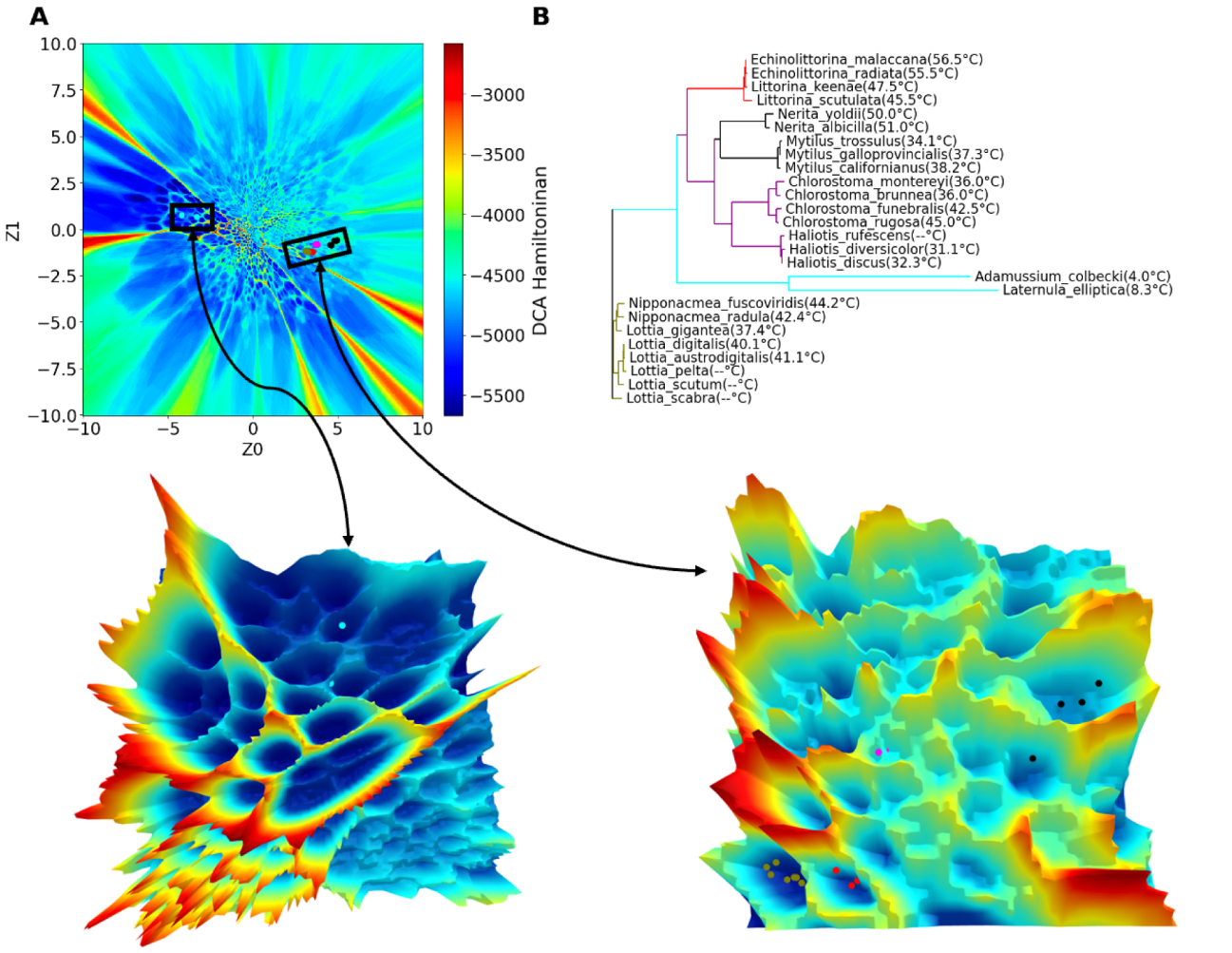


Figure 2. Clustering in the MDH evolutionary landscape. (A) An evolutionary landscape plot visualizes the distribution of Direct Coupling Analysis (DCA) scores for cytosolic malate dehydrogenase (cMDH) sequence variants generated by a Variational Autoencoder (VAE). The color gradient, ranging from blue to red, indicates DCA scores, with red representing regions of higher (unfavorable) scores. (B) The phylogenetic tree for cMDH, generated using Clustal Omega, is constructed from a distance matrix derived through pairwise sequence comparisons.

phobic contact metrics. This normalization was crucial for comparing proteins of varying lengths generated by the VAE.

Hydrophobic Contacts in MD Simulations. For the MD simulations, hydrophobic contacts were analyzed by dividing the total count by the frame length. This division helped distinguish between permanent and transient hydrophobic interactions, providing insights into the stability and behavior of the protein under simulated physiological conditions.

Molecular Dynamics Simulations. The 3D structures generated by AlphaFold2 were utilized as the initial models for the MD simulations conducted using OpenMM.[44] The amber14-all force field was applied to model the proteins, while the tip3pfb model was used for the water molecules to create the solvated systems.[45,46]

Each solvated system underwent energy minimization and equilibration in the NVT ensemble. This was followed by a production MD run in the NPT ensemble, conducted for 1.2 μs at 298 K. During the simulations, a Langevin Middle integrator was employed to maintain a constant temperature of 298 K, and OpenMM's Monte Carlo barostat ensured a constant pressure of 1 atm.[44,47]

Structural Analysis and Convergence Assessment. The root-mean-square deviation (RMSD) of the backbone atom positions and the root-mean-square fluctuation (RMSF) of individual residues were calculated using MDAnalysis.[40,42] Convergence of the simulations was assessed by comparing RMSF values at different time steps, ensuring fluctuations remained stable as the simulation progressed from shorter to longer time frames. Additionally, minimal changes in the RMSD of the backbone structure after 200 ns indicated stable structural behavior throughout the simulation duration.

RESULTS

Evolutionary Landscape of Malate Dehydrogenase and Its Connection with Structural Diversity. Analysis of the landscape conducted by [33, 48, and 49] has shown that protein sequences trained with a Variational Autoencoder (VAE) tend to get mapped into the low-dimensional manifold where functionally viable variants are grouped in distinct basins. VAE encoding also captures the most significant structural variations. We use this insight to elucidate the driving forces of thermal adaptation of cytosolic malate dehydrogenase (cMDH) by utilizing a combination of VAE and Direct Coupling Analysis (DCA) (Figure 2A).

By applying the Hamiltonian function to each sequence associated with coordinates in the 500^2-pixel grid, we construct

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

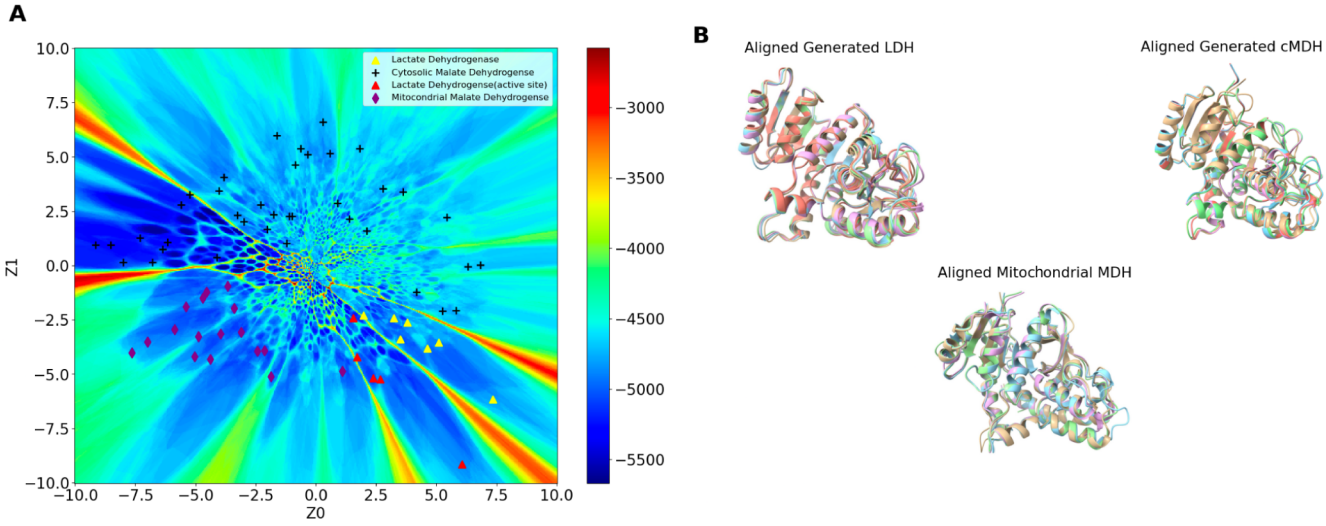


Figure 3. Generated structures from the evolutionary landscape. scatter points represent the positions of generated structures within the evolutionary landscape. (A) Scatter points indicating the positions of generated structures in the landscape, where four types of structures (cMDH with active site, cMDH without active site, lactate dehydrogenase (LDH), and mitochondrial MDH) are clustered based on their RMSD values. (B) Cytosolic malate dehydrogenase (cMDH), mitochondrial malate dehydrogenase (MDH), and lactate dehydrogenase (LDH) structures aligned with generated sequence structures, all with root-mean-square deviation (RMSD) < 2.0 Å.

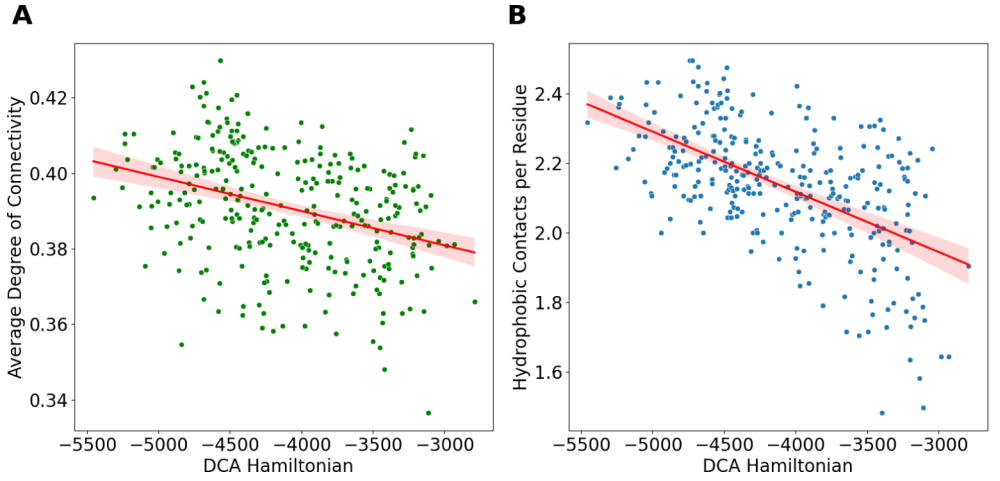


Figure 4. Relationship between DCA hamiltonian and normalized hydrophobic contacts in protein structures. (A) Scatter plot shows the correlation between DCA Hamiltonian values and the average degree of connectivity, for 400 protein structures generated from the VAE landscape. The red trend line indicates a decreasing degree of connectivity as the DCA Hamiltonian, plotted on the *x*-axis, increases. (B) Scatter plot depicts the correlation between DCA Hamiltonian values and the normalized number of hydrophobic contacts per residue across 400 protein structures generated from the VAE landscape. The red trend line indicates a decreasing number of hydrophobic contacts as the DCA Hamiltonian, plotted on the *x*-axis, increases.

the latent generative landscape (LGL). In this landscape, sequence space is organized according to the VAE latent variables and the learned distribution $q_\phi(z|x)$, with the sequence "energy" required to traverse the space defined by the Hamiltonian. This approach allows us to retrieve the VAE's intrinsic encoding of fitness by using the Hamiltonian value as a scoring metric.

The generated landscape mapped by a combination of VAE and Hamiltonian scores allows us to visualize the functional diversity within the MDH family. Characterized by distinct regions demarcated by high and low Hamiltonian scores, the landscape reveals the contributions from coevolved residues. Regions with lower Hamiltonian values are predominantly populated by extant sequences and experimental structures, which correlate with enhanced functional stability due to highly coevolved residues.[50−54] Within the evolutionary landscape (Figure 2A), species such as *Adamussium colbecki*, with a lethal temperature of 4 °C, and Laternula elliptica, with a lethal temperature of 8.3 °C, are located in distinct low Hamiltonian basins. This spatial arrangement starkly contrasts with sequences from species enduring higher lethal temperatures, ranging from 32 to 56 °C, which cluster tightly on the opposite side. To assess the pairwise sequence distances among 26 cMDH homologues, we constructed a Guide tree (Figure 2B) using Clustal Omega. Based on a distance matrix derived from pairwise sequence comparisons, this tree highlights distinct clustering patterns. Specifically, species that endure higher lethal temperatures exhibit greater sequence similarity, forming a tighter cluster within the tree. Such clustering highlights greater sequence similarity among high-

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

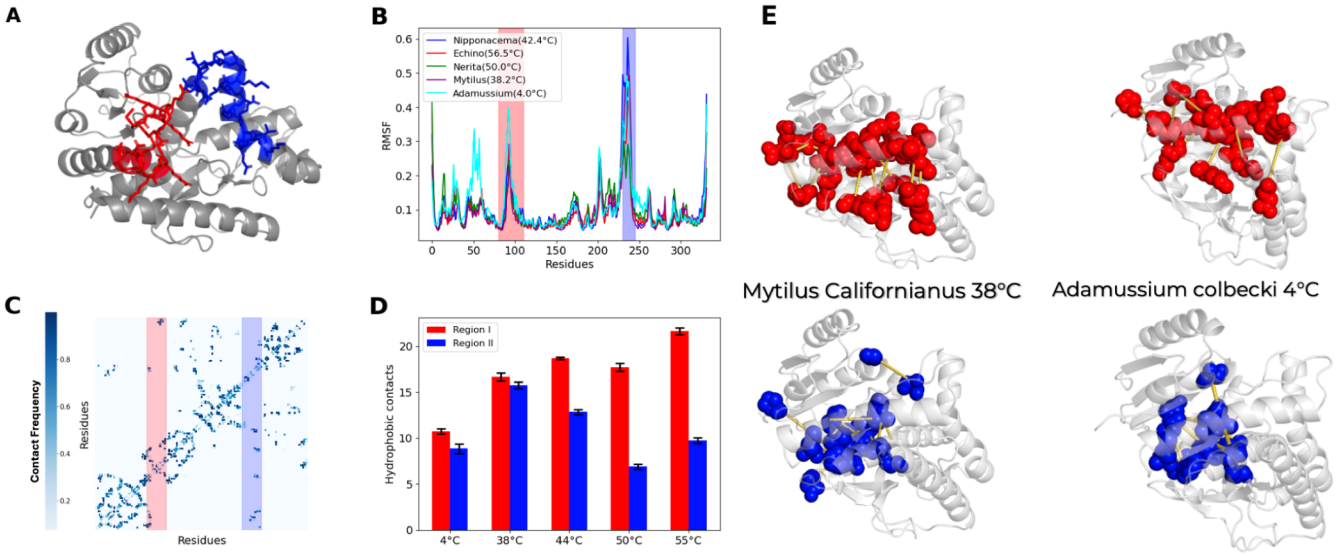


Figure 5. Molecular dynamics analysis of side chain movements (RMSF) of cMDH across various species at ambient temperature. (A) Structure of cytosolic malate dehydrogenase (cMDH) highlighting functionally active residues. The catalytic loop (residues 91−105), region 1, is shown in red, while the additional catalytic region (residues 230−245), region 2, is depicted in blue. (B) Root Mean Square Fluctuation (RMSF) profiles at 298 K for cMDH from five species: *Echinolittorina malaccana*, *Nerita yoldi*, *Mytilus californianus*, *Adamussium colbecki*, and *Nipponacmea radula*. Peaks in RMSF at residues 91−105 and 230−245, marked in red and blue, respectively, indicate regions of significant fluctuation. (C) The contact map highlights that the region containing catalytically active residues forms long-range hydrophobic contacts. (D) Permanent hydrophobic contacts in region I (residues 80−110) and region II (residues 230−260) across different cMDH families, with psychrophilic enzymes on the left and thermophilic enzymes on the right, emphasizing differences in contact dynamics related to thermal adaptation. (E) The structures illustrate permanent hydrophobic contacts in region 1 (red) and region 2 (blue) of cMDH from Mytilus californianus and Adamussium colbecki.

temperature cMDH orthologs as compared to their low-temperature counterparts, reinforcing the link between sequence homology and adaptability to different thermal conditions.[55]

To explore structural diversity in the evolutionary landscape, we analyzed 400 sequences, selecting 100 random sequences covering all major energy basins on the landscape (Figure 3A). Structures were predicted using AlphaFold, and their structural similarity was quantified through Root Mean Square Deviation (RMSD). We used a threshold of 2 Å RMSD to classify structures as similar to the known orthologs.[56−58] Four types of structures, clustered according to their RMSD, were identified in the landscape. Homology analysis, by using InterPro,[59] revealed two clusters corresponding to cMDH; one cluster included sequences with an active site, similarly positioned as in experimental cMDH structures, while the other lacked this active site. Additionally, one cluster was identified as mitochondrial malate dehydrogenase and another as lactate dehydrogenase. Moreover, no sequences with similar structures were found across these barrier regions, underscoring a significant demarcation in the evolutionary landscape that correlates with structural and functional divergence from cMDH.

To gain microscopic insight into the evolution of the thermal stability of malate hydrogenase, we calculated hydrophobic contacts and generated graphs using the structures generated from the VAE landscape (Figure 4). Hydrophobic contacts were calculated per residue based on atomistic distances grouped into residues. Due to the variability in sequence length, which includes gaps of insertion and deletion similar to those found in multiple sequence alignment files, the total number of hydrophobic contacts was normalized by dividing by the total number of residues. This normalization process allows for consistent comparison of hydrophobic contacts per residue across the landscape. We uncovered an inverse correlation between Hamiltonian values and hydrophobic contact per residue (Figure 4B). Since hydrophobic contacts are integral to the structural stability of proteins, particularly at higher temperatures, this trend suggests that structural stability diminishes as sequences shift toward higher DCA Hamiltonian values.[60,61] To further explore how the contacts of generated structures vary with Hamiltonian values, we constructed a graph in which C-α carbons represent the nodes and C-α to C-α contacts serve as the edges. Our analysis revealed a negative correlation between the average degree of connectivity and Hamiltonian values (Figure 4A). Given that the average degree of connectivity is linked to the thermal stability of an enzyme, this finding indicates that structures located at the peaks of the VAE landscape tend to be less thermally stable compared to those residing in the low-Hamiltonian regions.[55]

Dynamics Sheds Light on Evolving Thermal Activity in cMDH Homologues. While structures provide clues on thermal adaptation, it is, without doubt, the temperature-dependent dynamics of enzymes that report most accurately on catalytic activities.[62,63] To this end, we have conducted microsecond-long atomistic molecular dynamics simulations at room temperature and quantified enzyme dynamics via mean square fluctuation (RMSF) of contacts and Cartesian displacements. Not all of the sequences have experimentally validated structures, so to maintain consistency, we ran simulations on AlphaFold structures, which exhibit the same RMSF profile as the experimental structures.[55] We identify two regions (Figure 5B) with significant fluctuations: residues 91−105, forming the catalytic loop folding onto the catalytic site during ligand binding (highlighted in red in Figure 5A), and residues 230−245, which are involved in the catalytic process (highlighted in blue in Figure 5A).[7] Notably, *Adamussium colbecki*, adapted to

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

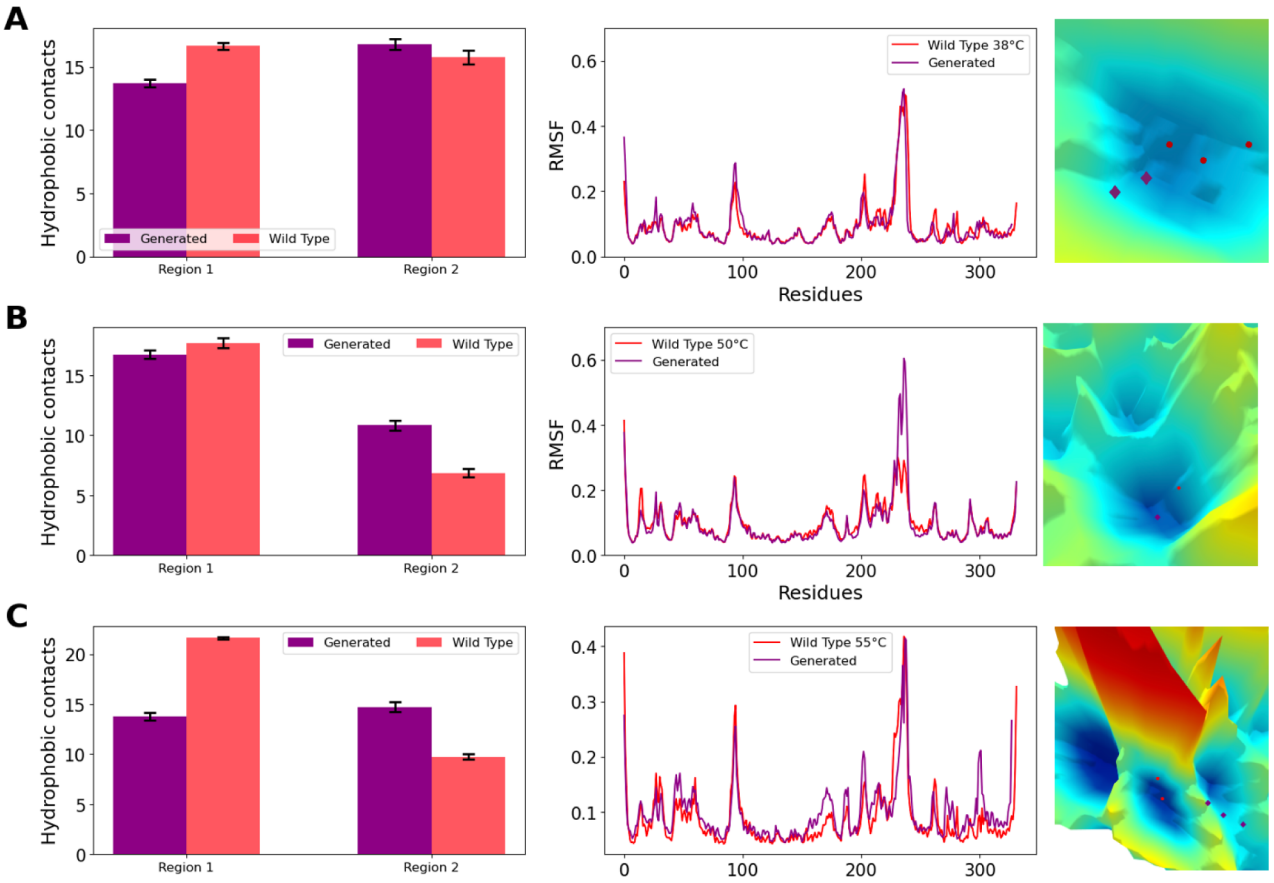


Figure 6. Comparison of changes in structural properties of generated proteins relative to their corresponding wild-type counterparts across varying Hamiltonian scores. Panels A, B, and C illustrate the differences in dynamic hydrophobic contacts (left), root-mean-square fluctuations (RMSF) across residues (center), and Hamiltonian energy differences (right) between the generated and wild-type structures. The data highlight structural and dynamic deviations influenced by Hamiltonian score variations.

colder environments, exhibits the highest RMSF, consistent with its low lethal temperature of approximately 277 K. While fluctuations in the first mobile region remained consistent across species, notable differences were observed in the dynamics of the second region as species adaptation ranged from colder to warmer environments (Figure 5B). This observation indicates that although the dynamic behavior of the catalytic area remains consistent, the distinct peaks observed in Region 2, which is also involved in catalysis, should be subject to further study. Hydrophobic interactions, salt bridges, van der Waals forces, and aromatic stacking[64−66] are key contributors to thermal stability. However, the precise mechanism by which these interactions collectively influence thermal stability remains poorly understood (Figure S1). A protein must maintain sufficient stability to preserve its structure while remaining flexible enough to carry out its biological functions near its optimal operating temperature.[67−69] The importance of balancing stability and degree of functional dynamism[70−72] has led us to investigate permanent and transient hydrophobic interactions throughout the simulation. Permanent hydrophobic contacts, defined as those present for more than 95% of the simulation time, were contrasted with transient contacts, which occur for less than 50% of the time. Interestingly, the frequency of transient hydrophobic contacts decreases in enzymes adapted to higher temperatures (Figure S1E).

We found that the catalytically significant residues (90−105 and 225−245) showed an absence of permanent hydrophobic contacts. This prompted us to explore the dynamic hydrophobic interactions near the catalytic and highly flexible domain boundaries, focusing on residues 80−110 and 220−260. Notably, in addition to being catalytically active, these regions in cMDH are responsible for long-range hydrophobic interactions (Figure 5C).

We observed that cMDH homologues from warmer climates exhibit a significant increase in permanent hydrophobic contacts within region I (residues 80−110) compared to those from colder environments (Figure 5D). In contrast, mesophiles show the highest frequency of permanent hydrophobic contacts in region II (residues 220−260), surpassing both thermophiles and psychrophiles (Figure 5D). Notably, region I also displays elevated transient hydrophobic contacts, while region II, having lower frequencies of both permanent and transient contacts, likely contributes to the higher fluctuations observed in this region. Additionally, *Adamussium*, an enzyme with a lethal temperature of 4 °C, shows lower permanent contact frequencies in both regions compared to mesophilic and thermophilic enzymes (Figure 5E), suggesting that these preserved interactions play a critical role in the thermal adaptation of these enzymes.[73]

These findings highlight the crucial role of dynamic hydrophobic interactions within specific regions, potentially enhancing the enzyme's thermal adaptability by modulating the flexibility of areas critical to catalysis. The observed increase in permanent hydrophobic contacts in particular regions suggests a mechanism through which these cMDH

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

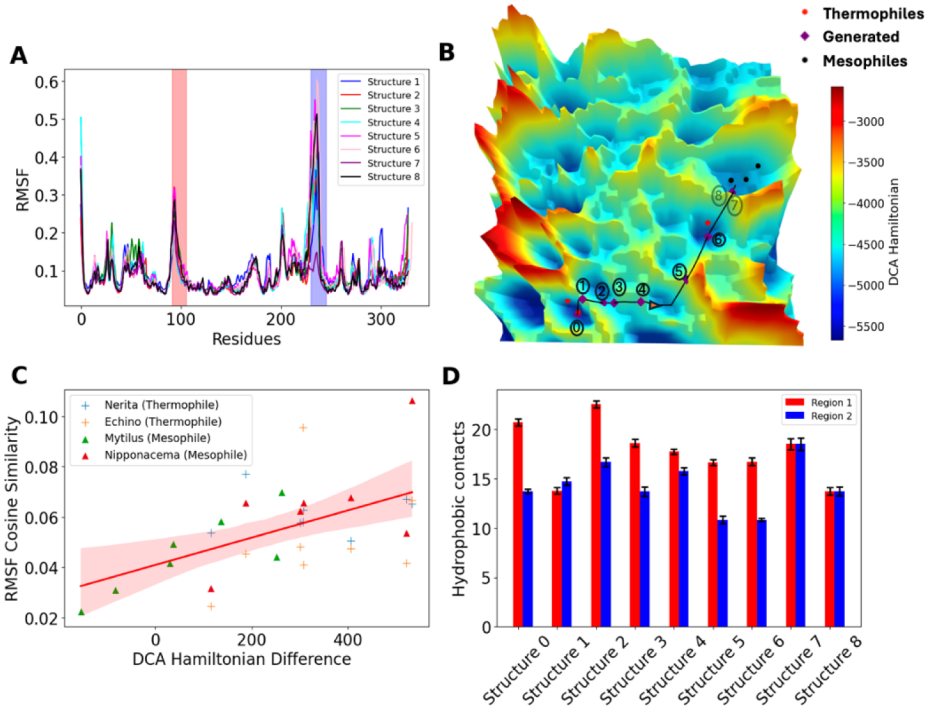


Figure 7. Change in the dynamics and thermal stability along a path connecting the cMDH thermophile to mesophile clusters. (A) Root Mean Square Fluctuation (RMSF) profiles at 298 K for eight generated structures along a trajectory connecting the *Echinolittorina, Nerita, and Mytilus* families. (B) The shortest path between families in the landscape. (C) Correlation between the cosine similarity of root-mean-square fluctuation (RMSF) profiles of the wild-type structure and those of structures generated from varying Hamiltonian values near the clustering of wild-type sequences. (D) Hydrophobic contacts in both regions, residues 80−110 and residues 220−260.

stabilize their structure under elevated temperatures, thereby compensating for increased thermal motion.

The RMSF profiles (Figure 6A−C) of the generated structures, derived from clustering in the cMDH sequence landscape, exhibit striking similarities to known cMDH structures from species such as *Mytilus californianus* and *Nerita yoldii* (Figure 6A,B). We analyzed the dynamic hydrophobic contact frequencies to investigate the basis for these similarities and the observed flexibility in catalytic regions. The RMSF profiles show that their flexibility and hydrophobic contact patterns remain consistent when the Hamiltonian difference between the wild-type and generated structures is minimal (Figure 6A). However, as the Hamiltonian difference increases (Figure 6B), we observe elevated fluctuations in Region II, while Region I remains stable. This trend is mirrored in the hydrophobic contact frequencies, where region I retains similar contacts, but Region II shows variability. As the Hamiltonian difference grows further, approaching the peak of the landscape (Figure 6C), the overall flexibility of the generated structures increases, as reflected by the rising fluctuation and altered hydrophobic contact frequencies in both regions.

The correlation of dynamic properties between the experimental and generated structures supports the validity of our computational models as accurate representations of the dynamical behavior of cMDH enzymes.[74] Additionally, this analysis highlights the importance of dynamic hydrophobic contacts, where permanent contacts correlate with thermal stability and transient contacts correlate with greater conformational flexibility and catalytic activity.

Dynamics of Sequences Sampled from Evolutionary Landscape. To explore patterns of conformational dynamics across distinct cMDH homologue clusters, we selected sequences from different parts of the generated evoltuionary landscape. We create a path evolving sequences on the landscape by starting from the Echinolittorina family (55 °C−56 °C) and extending to the Mytilus family (38 °C−42 °C), with the Nerita family (50 °C−51 °C) . We selected eight representative structures from this path. We subjected them to all-atom simulations over a microsecond at room temperature, maintaining the same conditions as those used for the cMDH enzymes. Consistent with our general observations for cMDH enzymes, the RMSF profiles for sampled structures (Figure 7A) showed similar fluctuations across most regions, except for Region 2, where each structure exhibited unique peaks. The fluctuations near the catalytic regions remained constant for all homologues, highlighting the structural conservation in these functionally important regions.

We calculated the cosine similarity of RMSF values between wild-type cMDH and generated structures from nearby coordinates to investigate the evolution of protein dynamics across the landscape. The scatter plot reveals a positive correlation between the cosine similarity score and the Hamiltonian difference between the generated and wild-type structures (Figure 7C). This trend suggests that as we move toward the peaks in the landscape, the dynamic properties deviate more from the wild-type, indicating that sequences residing in Hamiltonian pits exhibit dynamics that are more similar to the wild-type.

Among the eight generated structures, Structure 6 (Figure 7A) exhibits the highest RMSF in Region II and has the fewest permanent hydrophobic contacts. In contrast, Structure 7 (Figure 7B), which shows the lowest RMSF (Figure 7A), displays the most significant number of permanent hydrophobic contacts and one of the lowest counts of transient

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

hydrophobic contacts in Region II. Notably, all the generated structures located in nearby thermophilic pits exhibit a higher number of permanent hydrophobic contacts in Region I (Figure 7C,D) than in Region II, aligning with the hydrophobic contact patterns observed in the wild-type thermophilic proteins (Figure 5C). Similarly, two structures generated near mesophiles show comparable frequencies of permanent hydrophobic contacts, matching the contact patterns typical of wild-type mesophiles. These observations highlight the critical role of maintaining more permanent hydrophobic contacts in Region I to sustain catalytic activity at elevated temperatures.

DISCUSSION

Despite decades of research on cytosolic malate dehydrogenase (cMDH), detailed molecular level understanding of how these enzymes adapt to varying thermal environments to catalyze the conversion of malate to oxaloacetate are still missing. In this study, we employed a combination of machine learning,[37] global sequence modeling,[34,35] and molecular simulation techniques[44] to generate an evolutionary landscape for cMDH, providing a novel framework for understanding thermal adaptation at the molecular level.

A key result of our work is the demonstration that the latent generative landscape (LGL),[33] constructed using variational autoencoders (VAE) and informed by coevolutionary models, enables predictive insights into how different regions of sequence space correlate with thermal enzymatic properties (e.g., being a thermophile, mesophile, and psychrophile) (Figure 6). By exploring this landscape, we not only identify potential pathways and critical sequence changes that influence thermal adaptation but also lay the foundation for future studies focused on engineering specific sequence changes to achieve desired thermal stability and dynamical properties. This predictive power positions the LGL as a valuable tool for protein engineering and rational design.

Additionally, our structural and dynamic analyses revealed that the ratio of hydrophobic contacts between two key regions of the enzyme, region 1 and region 2, serves as a predictive order parameter for thermal stability. This ratio quantitatively reflects the dynamic interplay of hydrophobic interactions that stabilize cMDH under varying thermal conditions. Higher ratios were associated with enhanced thermal stability, underscoring the critical role of hydrophobic interactions in the enzyme's dynamics and adaptation.

The dynamic analysis of generated sequences, elucidated through microsecond-long atomistic simulations, closely mirrored the behavior of experimentally characterized cMDH structures.[55] These simulations revealed that variability in dynamics across species adapted to different temperature ranges is primarily driven by changes in the number of dynamic hydrophobic interactions near mobile regions critical for catalysis, influenced by specific mutations. In conclusion, this study highlights the importance of the LGL's predictive power for understanding sequence-thermal property relationships and the hydrophobic contact ratio as an order parameter for enzyme dynamics. Together, these findings provide significant insights into the molecular mechanisms of thermal adaptation and open new avenues for the rational design of enzymes with tailored thermal properties, contributing to advancements in biotechnology and protein engineering.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c01774.

Frequencies of key molecular interactions per trajectory frame across marine mollusk species; transient hydrophobic contacts in regions I and II; root mean square deviation (RMSD) profiles for cytosolic malate dehydrogenase (cMDH) from five species adapted to different temperatures (PDF)

AUTHOR INFORMATION

Corresponding Author

Davit A. Potoyan − *Department of Chemistry and Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa 50011, United States; Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, United States;* orcid.org/0000-0002-5860-1699; Email: potoyan@iastate.edu

Authors

Divyanshu Shukla − *Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa 50011, United States*

Jonathan Martin − *Department of Biological Sciences, UT Dallas, Richardson, TX 75080, United States;* orcid.org/0000-0003-0946-3864

Faruck Morcos − *Department of Biological Sciences, UT Dallas, Richardson, TX 75080, United States; Departments of Bioengineering and Physics and Center for Systems Biology, UT Dallas, Richardson, TX 75080, United States;* orcid.org/0000-0001-6208-1561

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c01774

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by funds from the National Institute of General Medical Sciences with grant nos. R35GM138243 awarded to DAP and R35GM133631 to FM. We also acknowledge support from the NIH National Institute for Allergy and Infectious Diseases with grant no. R01AI178692 to FM and JM. FM acknowledges support from the National Science Foundation (grant no. MCB-1943442).

REFERENCES

(1) Nava, G.; Laclette, J. P.; Bobes, R.; Carrero, J. C.; Reyes-Vivas, H.; Enriquez-Flores, S.; et al. Cloning, sequencing and functional expression of cytosolic malate dehydrogenase from Taenia solium: Purification and characterization of the recombinant enzyme. *Exp Parasitol.* **2011**, *128*, 217−224.

(2) Minárik, P.; Tomásková, N.; Kollárová, M.; Antalík, M. Malate dehydrogenases–structure and function. *Gen. Physiol. Biophys.* **2002**, *21* (3), 257−265.

(3) Goto, Y.; Fink, A. L. Conformational states of beta-lactamase: molten-globule states at acidic and alkaline pH with high salt. *Biochemistry* **1989**, *28*, 945−952.

(4) Yueh, A. Y.; Chung, C. S.; Lai, Y. K. Purification and molecular properties of malate dehydrogenase from the marine diatom Nitzschia alba. *Biochem. J.* **1989**, *258*, 221−228.

(5) Chao, Y.-C.; Merritt, M.; Schaefferkoetter, D.; Evans, T. G. High-throughput quantification of protein structural change reveals potential mechanisms of temperature adaptation in Mytilus mussels. *BMC Evol. Biol.* **2020**, *20* (1), 28.
(6) Pucci, F.; Rooman, M. Physical and molecular bases of protein thermal stability and cold adaptation. *Curr. Opin. Struct. Biol.* **2017**, *42*, 117−128.
(7) Dong, Y.-W.; Liao, M.-L.; Meng, X.-L.; Somero, G. N. Structural flexibility and protein adaptation to temperature: Molecular dynamics analysis of malate dehydrogenases of marine molluscs. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 1274−1279.
(8) Dong, M. A Minireview on Temperature Dependent Protein Conformational Sampling. *Protein J.* **2021**, *40*, 545−553.
(9) Zhang, J.; Liu, S.; Chen, M.; Chu, H.; Wang, M.; Wang, Z.; et al. Unsupervisedly prompting AlphaFold2 for accurate few-shot protein structure prediction. *J. Chem. Theory Comput.* **2023**, *19*, 8460−8471.
(10) Siddiqui, K. S.; Cavicchioli, R. Cold-adapted enzymes. *Annu. Rev. Biochem.* **2006**, *75*, 403−433.
(11) Szilágyi, A.; Závodszky, P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **2000**, *8*, 493−504.
(12) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1−43.
(13) Somero, G. N. Proteins and temperature. *Annu. Rev. Physiol.* **1995**, *57*, 43−68.
(14) Feller, G.; Gerday, C. Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **2003**, *1*, 200−208.
(15) DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **2005**, *6*, 678−687.
(16) Tokuriki, N.; Tawfik, D. S. Protein dynamism and evolvability. *Science* **2009**, *324*, 203−207.
(17) Bloom, J. D.; Arnold, F. H. In the light of directed evolution: Pathways of adaptive protein evolution. *Proc. Int. Acad. Sci.* **2009**, *106*, 9995−10000.
(18) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. Identification of direct residue contacts in protein−protein interaction by message passing. *Proc. Int. Acad. Sci.* **2009**, *106*, 67−72.
(19) Sułkowska, J. I.; Morcos, F.; Weigt, M.; Hwa, T.; Onuchic, J. N. Genomics-aided structure prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10340−10345.
(20) Morcos, F.; Hwa, T.; Onuchic, J. N.; Weigt, M. Direct coupling analysis for protein contact prediction. *Methods Mol. Biol.* **2014**, *1137*, 55−70.
(21) Ovchinnikov, S.; Kamisetty, H.; Baker, D. Robust and accurate prediction of residue−residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, *3*, No. e02030.
(22) Michel, M.; Hayat, S.; Skwark, M. J.; Sander, C.; Marks, D. S.; Elofsson, A. PconsFold: improved contact predictions improve protein models. *Bioinformatics* **2014**, *30*, i482−8.
(23) Morcos, F.; Pagnani, A.; Lunt, B.; Bertolino, A.; Marks, D. S.; Sander, C.; Zecchina, R.; Onuchic, J. N.; Hwa, T.; Weigt, M.; et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (49), No. E1293−301.
(24) Hopf, T. A.; Schärfe, C. P. I.; Jpglm, R.; Green, A. G.; Kohlbacher, O.; Sander, C.; et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **2014**, *3*, No. e03430.
(25) Ekeberg, M.; Hartonen, T.; Aurell, E. Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences. *J. Comput. Phys.* **2014**, *276*, 341−356.
(26) Brookes, D. H.; Park, H.; Listgarten, J. Conditioning by adaptive sampling for robust design. *arXiv*, **2019**, .
(27) Linder, J.; Bogard, N.; Rosenberg, A. B.; Seelig, G. A generative neural network for maximizing fitness and diversity of synthetic DNA and protein sequences. *Cell Syst.* **2020**, *11*, 49−62.e16.
(28) Du, X.; Sun, S.; Hu, C.; Yao, Y.; Yan, Y.; Zhang, Y. DeepPPI: Boosting Prediction of Protein−Protein Interactions with Deep Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 1499−1510.
(29) Zou, X.; Wang, G.; Yu, G. Protein Function Prediction Using Deep Restricted Boltzmann Machines. *BioMed. Res. Int.* **2017**, *2017*, 1729301.
(30) Christensen, P. M.; Martin, J.; Uppuluri, A.; Joyce, L. R.; Wei, Y.; Guan, Z.; et al. Lipid discovery enabled by sequence statistics and machine learning. *Elife* **2024**, *13*, RP94929.
(31) Ingraham, J.; Garg, V. K.; Barzilay, R.; Jaakkola, T.Generative Models For Graph-Based Protein Design*Advances in neural information processing systems*NIPS201915794−15805
(32) Riesselman, A. J.; Ingraham, J. B.; Marks, D. S. Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* **2018**, *15*, 816−822.
(33) Ziegler, C.; Martin, J.; Sinner, C.; Morcos, F. Latent generative landscapes as maps of functional diversity in protein sequence space. *Nat. Commun.* **2023**, *14* (1), 2222.
(34) Finn, R. D.; Clements, J.; Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29−37.
(35) Eddy, S. R. Accelerated Profile HMM Searches. *PloS Comput. Biol.* **2011**, *7*, No. e1002195.
(36) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; et al. Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7* (1), 539.
(37) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv*, .
(38) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nat. Methods* **2022**, *19*, 679−682.
(39) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(40) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(41) Bouysset, C.; Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. *J. Cheminf.* **2021**, *13* (1), 72.
(42) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.et al.MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations*Proceedings Of The Python In Science Conference. SciPy*Los Alamos National Laboratory201698−105
(43) Landrum, G. *RDKit: Open-Source Cheminformatics Software*, https://www.rdkit.org/.
(44) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PloS Comput. Biol.* **2017**, *13*, No. e1005659.
(45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(46) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett* **2014**, *5*, 1885−1891.
(47) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys* **1994**, *101*, 4177−4189.
(48) Tian, H.; Jiang, X.; Trozzi, F.; Xiao, S.; Larson, E. C.; Tao, P. Explore Protein Conformational Space With Variational Autoencoder. *Front. Mol. Biosci.* **2021**, *8*, 781635.
(49) Detlefsen, N. S.; Hauberg, S.; Boomsma, W. Learning meaningful representations of protein sequences. *Nat. Commun.* **2022**, *13* (1), 1914.

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287

(50) Dimas, R. P.; Jiang, X.-L.; Alberto de la Paz, J.; Morcos, F.; Chan, C. T. Y. Engineering repressors with coevolutionary cues facilitates toggle switches with a master reset. *Nucleic Acids Res.* **2019**, *47*, 5449−5463.
(51) Russ, W. P.; Figliuzzi, M.; Stocker, C.; Barrat-Charlaix, P.; Socolich, M.; Kast, P.; et al. An evolution-based model for designing chorismate mutase enzymes. *Science* **2020**, *369*, 440−445.
(52) Cheng, R. R.; Nordesjö, O.; Hayes, R. L.; Levine, H.; Flores, S. C.; Onuchic, J. N.; et al. Connecting the Sequence-Space of Bacterial Signaling Proteins to Phenotypes Using Coevolutionary Landscapes. *Mol. Biol. Evol.* **2016**, *33*, 3054−3064.
(53) Figliuzzi, M.; Jacquier, H.; Schug, A.; Tenaillon, O.; Weigt, M. Coevolutionary Landscape Inference and the Context-Dependence of Mutations in Beta-Lactamase TEM-1. *Mol. Biol. Evol.* **2016**, *33*, 268−280.
(54) Bisardi, M.; Rodriguez-Rivas, J.; Zamponi, F.; Weigt, M. Modeling Sequence-Space Exploration and Emergence of Epistatic Signals in Protein Evolution. *Mol. Biol. Evol.* **2022**, *39*, msab321.
(55) Liao, M.-L.; Somero, G. N.; Dong, Y.-W. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 679−688.
(56) Xin, F.; Radivojac, P. Post-translational modifications induce significant yet not extreme changes to protein structure. *Bioinformatics* **2012**, *28*, 2905−2913.
(57) Krissinel, E. On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* **2007**, *23*, 717−723.
(58) Lazar, T.; Guharoy, M.; Vranken, W.; Rauscher, S.; Wodak, S. J.; Tompa, P. Distance-Based Metrics for Comparing Conformational Ensembles of Intrinsically Disordered Proteins. *Biophys. J.* **2020**, *118*, 2952−2965.
(59) Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B. L.; Salazar, G. A.; et al. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51*, D418−D427.
(60) Baldwin, R. L. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 8069−8072.
(61) Widom, B.; Bhimalapuram, P.; Koga, K. The hydrophobic effect. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3085−3093.
(62) Thomas, T. M.; Scopes, K. R. The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. *Biochem. J.* **1998**, *330* (Pt 3), 1087−1095.
(63) Daniel, R. M.; Danson, M. J. A new understanding of how temperature affects the catalytic activity of enzymes. *Trends Biochem. Sci.* **2010**, *35*, 584−591.
(64) Lee, C.-W.; Wang, H.-J.; Hwang, J.-K.; Tseng, C.-P. Protein thermal stability enhancement by designing salt bridges: a combined computational and experimental study. *PLoS One* **2014**, *9*, No. e112751.
(65) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133−7155.
(66) Anderson, D. E.; Hurley, J. H.; Nicholson, H.; Baase, W. A.; Matthews, B. W. Hydrophobic core repacking and aromatic-aromatic interaction in the thermostable mutant of T4 lysozyme Ser 117–&#x003E. *Phe. Protein Sci.* **1993**, *2*, 1285−1290.
(67) Kim, S. Y.; Hwang, K. Y.; Kim, S. H.; Sung, H. C.; Han, Y. S.; Cho, Y. Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile Aquaspirillium arcticum. *J. Biol. Chem.* **1999**, *274*, 11761−11767.
(68) Fitter, J.; Herrmann, R.; Dencher, N. A.; Blume, A.; Hauss, T. Activity and stability of a thermostable alpha-amylase compared to its mesophilic homologue: mechanisms of thermal adaptation. *Biochemistry* **2001**, *40*, 10723−10731.
(69) Folch, B.; Dehouck, Y.; Rooman, M. Thermo- and mesostabilizing protein interactions identified by temperature-dependent statistical potentials. *Biophys. J.* **2010**, *98*, 667−677.
(70) Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys. Chem.* **2010**, *148*, 1−15.
(71) Karshikoff, A.; Nilsson, L.; Ladenstein, R. Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *Febs J.* **2015**, *282*, 3899−3917.
(72) Burns, D.; Venditti, V.; Potoyan, D. A. Temperature-sensitive contact modes allosterically gate TRPV3. *bioRxivorg*, **2023**, .
(73) Noivirt-Brik, O.; Horovitz, A.; Unger, R. Trade-off between positive and negative design of protein stability: from lattice models to real proteins. *PloS Comput. Biol.* **2009**, *5*, No. e1000592.
(74) Somero, G. N. Temperature Adaptation of Enzymes: Biological Optimization Through Structure-Function Compromises. *Annu. Rev. Ecol Syst.* **1978**, *9*, 1−29.

https://doi.org/10.1021/acs.jctc.4c01774
J. Chem. Theory Comput. 2025, 21, 3277−3287