UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Chemistry

Mechanisms of enzyme adaptation to extreme environments

The rational design of a thermophilic chorismate mutase

Ryan Scott Wilkins

A dissertation for the degree of *Philosophiae Doctor*

February 2024

UiT The Arctic University of Norway

Ryan Scott Wilkins

Mechanisms of enzyme adaptation to extreme environments

The rational design of a thermophilic chorismate mutase

Thesis submitted in partial fulfillment for the degree of Philosophiae Doctor

Department of Chemistry
Faculty of Science and Technology

Hylleraas School of Quantum Molecular Sciences

2024

Abstract

Extremophiles, especially those living in high-temperature environments, exhibit unique enzymatic mechanisms that allow survival under conditions detrimental to most life forms. Of particular interest is thermophiles, which are able to thrive at temperatures at which psychrophiles and mesophiles unfold and cease to function. The mechanisms by which these enzymes are able to overcome this temperature extreme are not yet well understood, leading to an increasing interest in studying how they function. Thermophiles have a multitude of applications under extreme conditions in industrial processes such as pharmaceuticals, waste management, and textiles, so a more complete understanding of these molecular mechanisms can lead to the development of novel enzymes for these purposes.

This doctoral thesis focuses on exploring the molecular adaptations of the enzyme chorismate mutase (CM). Chorismate mutases are found in bacteria and plants where they catalyze the conversion of chorismate to prephenate, an important precursor in the biosynthesis of aromatic amino acids. In this thesis, I characterize a new mesophilic chorismate mutase from *B. pumilus*, and a new thermophilic chorismate mutase from an unknown organism found from bioprospect soils in Antarctica. Using the thermophile as a guide, I attempt to mutate the mesophilic CM, inducing thermophilic behavior.

The research in this thesis employs empirical valence bond (EVB) simulations to elucidate the molecular behavior of chorismate mutase. The methodology involves detailed computational modeling, to obtain accurate free energy estimates of the enzymatic reaction, emphasizing the enzyme's structural and functional adaptations to different environments. By conducting simulations across a range of temperatures, I am furthermore able to extract the enthalpic and entropic contributions to the activation free energy, providing insights into the molecular dynamics and stability mechanisms of chorismate mutase. This enables highlighting significant differences in enzyme behavior between normal and extreme environmental conditions.

The findings contribute to a deeper understanding of enzyme adaptation mechanisms in extremophiles. The study also discusses the potential of EVB simulations as a powerful tool for exploring enzyme behavior in extremophiles, setting the stage for future research in this field. The thesis concludes by underscoring the importance of computational approaches in advancing our understanding of life under extreme conditions and their practical applications in industry.

Acknowledgements

The work presented in this thesis was accomplished at the Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, Faculty of Science and Technology at the University of Tromsø from November 2019 to October 2023.

First and foremost, I would like to thank my supervisor Bjørn Olav Brandsdal for his guidance during this part of my career. Thank you for steering me along the right path, while allowing me the freedom to explore and learn from my own ideas. These skills will be invaluable to my future career in research. Thanks for the trips to Uppsala and encouraging me to attend a variety of conferences to meet and interact with other great researchers. I would also like to thank my co-supervisors, Geir Villy Isaksen and Bjarte Aarmo Lund, both for your insight and knowledge in getting my project up and running, and also for being available to bounce ideas off of, even after leaving academia. I would also like to thank Johan Åqvist for your excellent scientific insight on collaborations, as well as for the great talks about hockey whenever we would meet.

I also would like to sincerely thank my friends at both the UiT Department of Chemistry and through Hylleraas: Martin, Aleksi, Dat, Gabriel, Inga, Tonje, Sahil, Marc, Quentin. Thank you for the coffee breaks, the great beer, the trips across the border to "visit" Kilpisjärvi, the parties, the travels. These all helped keep me sane throughout the past four years, and I couldn't have done this without you all by my side. An especially big thanks goes to Bente for being my first friend when I moved here, our work trips to Uppsala and conferences, our fantastic discussions about our project, for introducing me to Buktafest, and for being my gym buddy.

Finally, I would like to thank my family for supporting your son/brother in his decision to move across the Atlantic ocean to the Arctic circle. Thank you for always being available to talk despite the distance. Thank you for always welcoming me back home during my infrequent visits. My gratitude is eternal.

Ryan Scott Wilkins
Tromsø, January 2024

Abstract

Extremophiles, especially those living in high-temperature environments, exhibit unique enzymatic mechanisms that allow survival under conditions detrimental to most life forms. Of particular interest is thermophiles, which are able to thrive at temperatures at which psychrophiles and mesophiles unfold and cease to function. The mechanisms by which these enzymes are able to overcome this temperature extreme are not yet well understood, leading to an increasing interest in studying how they function. Thermophiles have a multitude of applications under extreme conditions in industrial processes such as pharmaceuticals, waste management, and textiles, so a more complete understanding of these molecular mechanisms can lead to the development of novel enzymes for these purposes.

This doctoral thesis focuses on exploring the molecular adaptations of the enzyme chorismate mutase (CM). Chorismate mutases are found in bacteria and plants where they catalyze the conversion of chorismate to prephenate, an important precursor in the biosynthesis of aromatic amino acids. In this thesis, I characterize a new mesophilic chorismate mutase from *B. pumilus*, and a new thermophilic chorismate mutase from an unknown organism found from bioprospect soils in Antarctica. Using the thermophile as a guide, I attempt to mutate the mesophilic CM, inducing thermophilic behavior.

The research in this thesis employs empirical valence bond (EVB) simulations to elucidate the molecular behavior of chorismate mutase. The methodology involves detailed computational modeling, to obtain accurate free energy estimates of the enzymatic reaction, emphasizing the enzyme's structural and functional adaptations to different environments. By conducting simulations across a range of temperatures, I am furthermore able to extract the enthalpic and entropic contributions to the activation free energy, providing insights into the molecular dynamics and stability mechanisms of chorismate mutase. This enables highlighting significant differences in enzyme behavior between normal and extreme environmental conditions.

The findings contribute to a deeper understanding of enzyme adaptation mechanisms in extremophiles. The study also discusses the potential of EVB simulations as a powerful tool for exploring enzyme behavior in extremophiles, setting the stage for future research in this field. The thesis concludes by underscoring the importance of computational approaches in advancing our understanding of life under extreme conditions and their practical applications in industry.

List of Papers

Paper I

Wilkins, R.S. and Lund, B.A. and Isaksen, G.V. and Åqvist, J. and Brandsdal, B.O. "Accurate Computation of Thermodynamic Activation Parameters in the Chorismate Mutase Reaction from Empirical Valence Bond Simulations". In: *Journal of Chemical Theory and Computation.* Vol. 20, no. 1 (2024), pp. 451–458. DOI: 10.1021/acs.jctc.3c01105.

Paper II

Wilkins, R.S. and Lund, B.A. and Isaksen, G.V. and Åqvist, J. and Brandsdal, B.O. "Characterization of a thermophilic chorismate mutase and its utility as a template in rational enzyme design". *Preparing for publication.* Pre-print available at DOI: 10.1101/2023.04.20.537678

Paper III

Wilkins, R.S. and Skagseth, S. and Williamson, A. and Lund, B.A. and Brandsdal, B.O. "Biophysical characterization and analysis of a mesophilic chorismate mutase from B. pumilus". *Unpublished.*

Contents

List of Figures

List of Tables

List of Abbreviations

BLAST	Basic Local Alignment Search Tool
BpCM	*B. pumilus* chorismate mutase
BsCM	*B. subtilis* chorismate mutase
CM	chorismate mutase
DFT	density functional theory
ENM-NMA	elastic network model normal mode analysis
EVB	empirical valence bond
FEP	free energy peturbation
IPL	isochorismate pyruvate lyase
M1CM	chorismate mutase from unknown organism
MD	molecular dynamics
MM	molecular mechanics
MMRT	macromolecular rate theory
OPLS	optimized parameters for liquid simulation
PchB	IPL encoded by gene PchB
PS	product state
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RS	reactant state
TS	transition state
vdW	van der Waals

Chapter 1

Introduction

1.1 Enzymes: A brew-tiful past with modern applications

Even prior to the 18th century, biological catalysis was recognized as a physiological process, stemming from studies on the digestions of meat from stomach secretions. [1] However, it wasn't until 1833 that Payen and Persoz identified diastase, a "ferment" found in malted grain, was responsible for the breakdown of complex carbohydrates into simple sugars suitable for alcohol production in brewing. [2, 3] Over 40 years later Kühne coined the word *enzyme*, from the Greek "in yeast" to describe these processes. [4] Initially, following Pasteur's work on enzymes in the 1850s, it was thought that enzymes were a "vital force," being able to only operate within a living cellular environment. However, work by Buchner near the turn of the century showed that even in the absence of living cells, enzymes were still able to function. [5]

Almost 100 years after Payen and Persoz's work on diastase, the true nature of enzymes was still largely unknown. When isolating urease in 1926, Sumner found that urease crystals were proteins. However, this theory was not widely accepted until the 1930s, when further work by Northrop and Kunitz showed that digestive enzymes pepsin and trypsin were also composed of proteins when crystallized, leading to the shared Nobel prize between the three in 1946. [1, 6]. In 1965, the first high-resolution crystal structure of an enzyme was identified when Blake *et al.* published a 2 Å crystal structure of lysozyme, subsequently leading to the origins of structural biology in an effort to identify the structure and function of enzymes [7].

With the long history of enzymes in brewing, it is apparent enzymes can also aid in the production of other foodstuffs. After the role of enzymes in brewing was identified, applications in the fermentation of dairy into yogurts and cheeses were also quite apparent. However, it was not long before the applications of enzymes to other industrial processes were introduced. In the 1930s, large-scale industrial use of enzymes began in the introduction of proteases and lipases to laundry detergents, based on a 1913 patent by Röhm, who used similar enzymes in a pre-wash soaking solution. [8, 9]. In present times, enzymes are still widely used in food production and detergents, but their industrial applications have also expanded to bio-fuel production, textile production, animal feed production, pharmaceuticals, and forensics, among other applications. [10]

However, the rapid expansion of enzyme applications in industry has created the need and the want for a greater diversity in enzymes better suited to function in different environments. Early attempts to create enzymes suited for specific environments largely stem from directed evolution. This process is both costly and time-consuming, thus undesirable for industrial purposes. Thus, a better

solution was introduced: rational enzyme design. In rational enzyme design, advancements in *in silico* techniques such as molecular modelling have enabled scientists to explore site-specific mutations based on enzyme structure and kinetic calculations to better understand impacts of mutations on enzyme structure and function in a fraction of the time it takes in a traditional wet lab. [11]

1.2 A Rational Choice for Rational Enzyme Design

Due to their ability to significantly increase reaction rates, enzymes have found much utility in a wide range of industrial settings, from waste management to pharmaceutical production. However, optimal industrial settings are often quite different and extreme in comparison to the environmental conditions in which life exists. As such, placed directly into an industrial setting, many enzymes will not function optimally, if even at all. Since the introduction of enzymes into industrial processes, there has been a push to create enzymes tailored to be viable in a variety of industrial settings much different than the living environment of the host organism. This field, known as enzyme engineering, has now become a large focus of much biochemical research.

Historically, new enzymes for industrial processes were created through directed evolution. This method involves introducing random mutations into an enzyme, creating a catalogue of mutant enzymes. These new enzymes are then screened for improved performance under the desired conditions. The most promising variants are then recombined and the process is repeated until an enzyme with optimal function is created. This method mimics the natural process of evolution and can be used to create enzymes for a variety of industrial and biomedical applications. However, due to the random nature of the mutations and the time taken to express and characterize the enzymes at each iteration of mutations, directed evolution is expensive and slow, thus it is not an optimal solution for industries trying to make or save money.

Advances in computing power have made computational methods in biochemistry a powerful tool in enzyme engineering. With access to vast libraries of structural data, and physics-based molecular modelling software providing atomistic-level details of biocatalysis invisible in a wet-lab environment, the field of enzyme engineering has quickly turned to rational enzyme design. By evaluating large amounts of structural data available from computational methods such as bioinformatics and molecular modelling, we can now selectively target areas of an enzyme in which mutations are likely to affect function in a desirable way. In comparison to directed evolution, the chances of success in rational enzyme design are much higher, and using the right computational tools can often be faster than growing multiple generations of bacteria.

However, approaches to rational enzyme design are still quite primitive. As the exact mechanisms utilized by enzymes to adapt to extreme environments are highly specific to the the enzyme, many examples of rational design of extremophiles require the use of existing extremophiles as target enzymes with the desired properties. However, these target enzymes may be difficult to work with,

or not entirely suitable for the intended application. This is particularly common in the scenario of enzymes designed for applications in extreme environments, where they may be subjected to increased/decreased temperatures, increased salinity, or increased acidity/alkalinity.

1.3 Adaptation of Life to Extreme Environments

Our current view and understanding of life tends to be quite anthropocentric. However, given the vast diversity in life on Earth, and the extreme range of conditions in which it can be found, this outlook is insufficient. While we are homeothermic organisms able to produce and maintain an internal body temperature different than the external environment and have adapted as such, most organisms are poikilotherms, unable to produce or maintain an internal temperature. At the extreme ends of what we consider a comfortable temperature range exists organisms we know as *extremophiles*. Near sulfur vents on the ocean floors, where temperatures can reach in excess of 100° C, we can find thermophilic bacteria thriving. Meanwhile the Arctic waters near Tromsø, where temperatures can reach as low as 3° C, are teeming with a diverse range of psychrophilic plant, fish, and bacterial life.

The molecular mechanisms we possess in our bodies, mainly the enzymes which catalyze biochemical reactions necessary for life, are able to operate within a narrow temperature range close to our body temperature – 37° C. Outside of this temperature range, their function is severely impacted. As all of life has evolved from a common ancestor, it is logical that many organisms throughout the broad temperature ranges found on our planet require highly similar enzymes to maintain function. However, if this is the case, and the corresponding enzymes are only able to function within a narrow temperature range, what are the underlying mechanisms these enzymes have developed through evolution to be able to function and survive in their respective envrionments?

Looking at transition state theory, we observe a exponential dependence of the reaction rate on temperature. Thus, even a small decrease in temperature can be severely detrimental to an enzyme's ability to catalyze a reaction. On the other side of the temperature scale, things are also problematic as the weak interactions (e.g. hydrogen bonding and hydrophobic packing) responsible for giving a protein its structure, and therefore function, are disrupted. This results in the unfolding of an enzyme, leaving it unable to catalyze reactions.

1.4 Goals of this Thesis

Rational enzyme design is still an emerging field. While lots is known about enzyme adaptation to extreme environments, there is still a lot unknown. The main goal of this thesis was to use computational chemistry methods to design an extremophilic enzyme through rational design principles - that is, to start from a base mesophilic enzyme, employ computational studies to better understand the role the enzyme plays in catalysis, and use information acquired from our

computational methods to implement targeted mutations to allow it to function in a different environment. Paper I in this thesis explores the mechanisms of catalysis in the mesophilic chorismate mutase from *B. subtilis*. In Paper II, we characterize a chorismate mutase from *B. pumilus*, a soil bacterium, in hopes of finding a psychrophilic chorismate mutase. Finally, Paper III serves two purposes: (1) the characterization of a novel thermophilic chorismate mutase from an unknown organism, the sequence of which was bioprospected in the Antarctic dry deserts, and (2) utilization of our knowledge of our mesophilic and thermophilic chorismate mutases to create a new thermophilic enzyme.

The rest of thesis is organized as follows. Chapter 2 starts with providing the necessary foundations on enzyme kinetics and enzyme adaptation which has inspired our work, while also giving a background on what is currently known about enzyme-catalyzed chorismate mutase reactions. Chapter 3 gives an introduction to the computational methods we have used in further understanding how enzymes function. Finally, Chapter 4 provides gives some concluding remarks to the thesis, including future directions for this research.

Chapter 2

Enzyme Catalysis

Within this chapter you will find concepts relevant to understanding enzyme structure and function. It begins with an introduction to enzyme kinetics, followed by a history of attempts to understand how enzyme structure has enabled enzymes to catalyze reactions in such an efficient manner. This provides the framework necessary to understanding how enzymes have developed mechanisms enabling their function in extreme environments.

2.1 Enzyme Kinetics

Enzymes work by aiding in the conversion of a substrate (S) into products (P) necessary for an organism's function. The enzyme first obtains free substrate from surrounding media, before generating an enzyme-substrate (ES) complex. Catalysis occurs next, forming an enzyme-product (EP) complex, before the product is released by the enzyme. This process can be summarized by a general enzymatic reaction scheme:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{EP} \longrightarrow \mathrm{E} + \mathrm{P} \tag{2.1}$$

where k_1 and k_{-1} are the association and dissociation of the ES complex, respectively, and k_{cat} is the catalytic rate in the formation of the EP complex from the ES complex.

The rate-limiting step of enzyme catalysis is the conversion of the ES complex to the EP complex. We can define the rate of this step, k_{cat}, using its relation to the activation free energy from transition-state theory:

$$k_{cat} = \kappa \frac{k_B T}{h} \exp \frac{-\Delta G^{\ddagger}}{k_B T} = \kappa \frac{k_B T}{h} \exp \left(\frac{-\Delta H^{\ddagger}}{k_B T} + \frac{\Delta S^{\ddagger}}{k_B} \right) \tag{2.2}$$

Here, κ is the transmission coefficient (generally assumed to be close to unity), k_B is the Boltzmann constant, T is the system temperature, h is the Planck constant, and $\Delta G^{\ddagger}$, $\Delta H^{\ddagger}$, and $\Delta S^{\ddagger}$ are the thermodynamic activation parameters corresponding to free energy, enthalpy, and entropy, respectively.

However, there are two caveats with defining an enzyme's activity using just k_{cat}: (1) a thorough description of an enzyme's reaction rate needs a reference state to be compared to, and (2) enzyme evolution appears to prioritize overall catalytic efficiency, not just k_{cat}.

Quantifying the catalytic power of an enzyme requires some reference reaction to which the catalysis can be compared to. The most logical comparative system for enzyme catalysis is that of the uncatalyzed reaction in an aqueous environment.

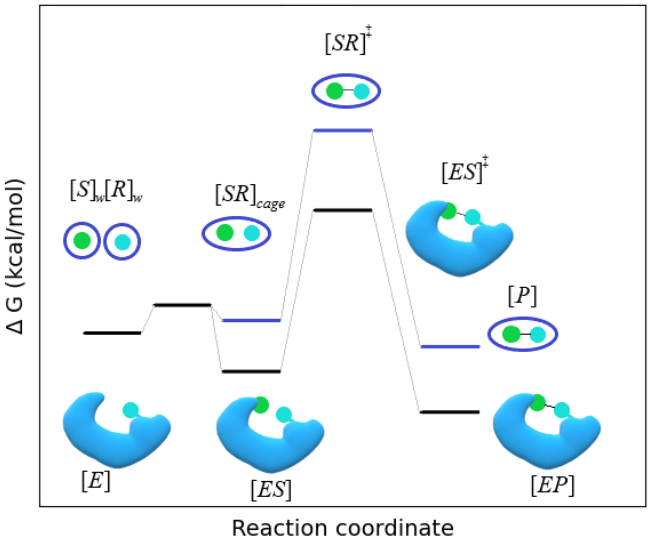


Figure 2.1: Hypothetical reaction profile showing the difference in $\Delta G^{\ddagger}$ for the formation of a product in an enzymatic reaction (light blue shape, black lines) in comparison to the reference reaction in water (blue circles, blue lines). In the reference reaction, the substrate S and the reacting group R must come within reacting distance in a single solvation cage. The formation of this transition state complex $[SR]^{\ddagger}$ in a single solvation cage leads to much of the free energy difference in comparison to the enzyme-substrate complex $[ES]^{\ddagger}$.

The catalytic efficiency of an enzyme can be defined as k_{cat}/K_m the ratio of the catalytic rate to the Michaelis constant for an enzyme. The Michaelis constant is generally defined in enzyme kinetics as the substrate concentration at which substrate is converted to product at half of the maximum reaction rate of the enzyme, but can also be defined in relation to the catalytic rate:

$$K_m = \frac{k_{-1} + k_{cat}}{k_1} \tag{2.3}$$

2.2 How do Enzymes Aid in Catalysis?

Understanding enzyme kinetics is an important aspect of understanding enzyme function, but it is not the only component. From a long history of experiments on enzymes it was evident that enzymes possess a huge catalytic power, being able to speed up reaction rates significantly with a theoretical upper-limit of $k_{cat} = 10^7 \text{sec}^{-1}$. [12] However, following the natural progression of science, this

leads to a further question: what is it about enzymes that allow them to catalyze biochemical reactions at such a high rate? With the abundance of enzymes in both research and industry, this question has both theoretical and practical implications.

The first attempts at answering this question appeared not long after enzymes were first conceptualized by Payen and Persoz. In 1894, before the molecular structure of enzymes were even known, Emil Fischer proposed the lock-and-key model of enzyme catalysis after studies on invertase showed a stereoselectivity for α-glucoside over β-glucoside. [13] True to its name, this hypothesis assumed enzyme catalysis required a unique fit between a substrate and the enzyme.

Like many early hypotheses in science, the lock-and-key model was found to be insufficient and needed to be expanded. In 1958, an analysis of early crystal structures of myoglobin and hemoglobin showed that upon entering an active site, a substrate induces conformational changes within active site to properly align a substrate for catalysis. This analysis and modification of the lock-and-key model, named the induced fit model by Koshland, further emphasized the requirement for enzyme specificity, as non-substrates would not induce the conformational changes required for catalysis. [14]

Prior to the 1970s, most work regarding the origins of enzyme catalytic power had focused on enzyme specficity and reduction in k_{cat} compared to the uncatalyzed reaction. However, in 1975 Jencks proposed the so-called Circe effect, a highly influential hypothesis attempting to explain enzyme catalysis by means of a reduction in $\Delta G^{\ddagger}$. [15] Building off of the induced fit model the Circe effect proposes that the induced alignment and restriction of substrate motion upon substrate binding has an associated entropy cost. Therefore, the entropic contributions to the activation free energy within the enzyme are significantly reduced as they are paid upon binding, lowering the activation barrier height within the enzyme.

Up until this point, most work on enzyme catalysis had been experimental in nature. However, with increasing power of computers, atomistic-scale insight into the mechanics of enzyme catalysis were within reach. In 1976, one of the first and most influential quantum mechanical simulations of enzyme active sites were performed by Warshel and Levitt. [16] In this study, they observed the charged carbonium ion transition state in a lysozyme was stabilized by interactions with Asp52. This was the first evidence of transition state stabilization contributing to a lower reaction energy barrier in enzymes. This was expanded upon two years later by Warshel [17], where he demonstrated that in bulk solvent, half of the energy gained from charge-dipole interactions is spent in rearranging the solvent molecules for optimal dipole-dipole interactions. Comparatively, in enzymes this energy cost is paid during protein folding, thus active site dipole moments are already optimally aligned for interaction with the substrate, leading to better catalysis than in bulk water.

2.3 Enzyme Adaptation to Extreme Environments

2.3.1 Temperature Adaptation of Enzymes Corresponds to a Shift in Thermodynamic Activation Parameters

As all life has evolved from a common ancestor, it is logical that many organisms require similar biomolecules to function, and that the enzymes used to create these biomolecules also have some similarity. However, despite evolving from a common ancestor, it is clear that life can currently be found in a large range of diverse environmental conditions. Thus, an important question in the understanding evolution is how the enzymes in these organisms have also evolved to adapt to such diverse conditions.

The most obvious difference in environmental conditions is temperature: some organisms are found in near-freezing conditions (psychrophiles), some can be found in hot springs and deep sea hydrothermal vents where temperatures can reach over 100°C (thermophiles), while others are found at temperatures humans and other mammals would consider comfortable (mesophiles).

With respect to Equation 2.2, this presents an obvious predicament due to the exponential dependence of k_{cat} on temperature. When considering mesophilic enzymes, lowering the temperature to near-freezing can result in a lowering of k_{cat} by 20-250 fold. [18] Conversely, when raising the temperature to optimal temperatures for thermophiles, stability of the enzyme becomes a concern as the weak intermolecular interactions responsible for a protein's structure (and therefore function) are disrupted, leading to unfolding of the enzyme and a loss in activity. Despite this problem, however, when comparing extremophiles to their mesophilic homologues, the difference in activation free energy $\Delta G^{\ddagger}$ is quite small, often found to be 1 kcal/mol or less. Thus, it appears that the likely mechanism behind enzyme temperature adaptation is a redistribution of the free energy between the enthalpic and entropic components, $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$.

When comparing homologous enzymes, there is a high degree of structural similarity and sequence conservation, even when considering extremophilic variants. Furthermore, it has been observed that the active sites of the enzymes are highly conserved, which is in line with the notion that the enzyme active sites are fine-tuned to optimize catalysis. Considering this, it appears likely that surface residues are likely responsible for tuning the enthalpy-entropy balance of psychro- and thermophilic enzymes. In short, psychrophilic enzymes display a "softer" surface than homologous thermophilic enzymes.

When looking at the dependence of the reaction rate of enzymes on temperature, it is typical that at lower temperatures, enzyme activity drops off significantly, increases to a peak as the temperature approaches the enzyme's optimal temperature T_{opt}, and decreases past T_{opt} as the enzyme denatures and is unable to catalyze the reaction. Comparing the reaction rates of homologous enzymes adapted to different temperatures we can observe a few key differences. Most obvious is the shifting of T_{opt}: warm-adapted enzymes will have a higher T_{opt} than mesophilic and psychrophilic enzymes. Less obvious, however, is the "lifting of the tail." As seen in Figure 2.2, on the lower end of T_{opt},

thermophilic enzymes will have a sharper decrease in reaction rate compared to mesophilic enzymes, which in turn also decrease in reaction rate faster than their psychrophilic counterparts.

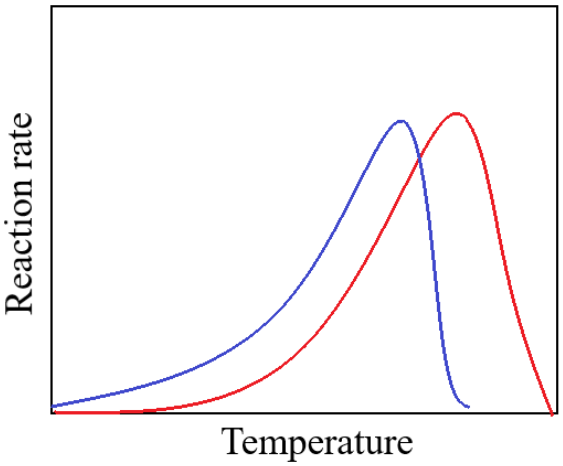


Figure 2.2: Cold-adapted enzymes exhibit a T_{opt} shifted to the left in comparison with their warm-adapted homologues, with a slower decrease in rate at lower temperatures due to the shifting of the activation enthalpy-entropy balance to favour temperature-independent entropic effects to catalysis. Furthermore, the warm-adapted enzyme is more thermostable, leading to a slower decrease in k_{cat} past T_{opt}

Arrhenius plots have long been used to look at temperature-dependent behaviour of chemical reactions, and thus are of the best methods for extracting $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$. Thus, they have proven extremely useful in the study of cold-adaptation. [19–24] From the macroscopic thermodynamic description of the Gibbs free energy:

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger}, \tag{2.4}$$

and simply dividing by T, we can see a linear behaviour in $\Delta G/T$ v.s. $1/T$, yielding a plot with slope $\Delta H^{\ddagger}$ and intercept of $\Delta S^{\ddagger}$. When comparing these plots between a cold-adapted and warm-adapted enzyme, as seen in Figure 2.3, the stronger temperature dependence and less negative entropy of the warm-adapted enzyme will yield a steeper slope and higher intercept.

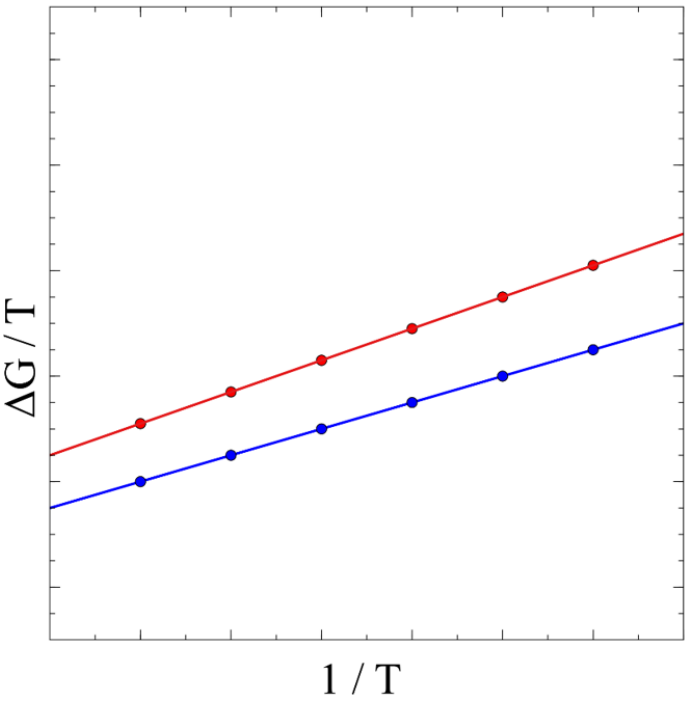


Figure 2.3: Arrhenius plots of cold- versus warm-adapted enzymes show a steeper slope and higher intercept for the warm-adapted variants in comparison to cold-adapted variants.

It is, however, important to note that this is not the only hypothesis of thermal adaptation in enzymes. In some cases of enzyme catalysis, non-linear Arrhenius behavior is observed. [25, 26] One attempt to explain this, known as macromolecular rate theory (MMRT), predicts that this curvature is the result of heat capacity. [27–29] More specifically, they propose this is caused by a $\Delta C_p^{\ddagger} < 0$, meaning the heat capacity of the rate-limiting transition state is lower than that of the reactant state. However, as pointed out by Åqvist, while there is a large $\Delta C_p^{\ddagger}$ associated with the folding of a protein due to changes in conformation, the changes in conformation during a chemical step in enzyme catalysis are relatively small, and not likely to lead to a large enough $\Delta C_p^{\ddagger}$ to influence this curvature. [21, 30]

2.3.2 Mechanisms of Enzyme Adaptation

There are a variety of proposed mechanisms for how enzymes have evolved to change the softness of the surface residues. As changes within α-helices and β-sheets can be deleterious to the structure [31, 32], most mutations affecting enzyme adaptation to extreme environments target loop regions. There are a

variety of mechanisms present across different enzymes, which can be categorized into one of five groups: (1) mutation of a polar residue into a less polar residue or vice versa, stabilizing or destabilizing the hydrogen bonding, (2) mutation of a polar residue into a charged residue or vice versa, stabilizing or destabilizing the hydrogen bonding, (3) complete changing of a loop region, completely modifying the hydrogen bonding network, (4) mutating a polar residue into a hydrophobic residue or vice versa, strengthening or disrupting hydrophobic packing at the enzyme's surface, and (5) introduction or disruption of a salt bridge. [24]

Isaksen *et al.* explored a variety of these mechanisms of temperature adaptation in a comparison between Atlantic salmon trypsin (AST) and bovine trypsin (BT). [24] The Nβ5-Nβ6 and autolysis loop in AST were found to be significantly more flexible in the cold-adapted AST in comparison to the same regions in BT. A key mutational difference between the two enzymes were found in both loops - Tyr97 and Asp150 in AST contribute to increased loop flexibility compared to the corresponding residues in BT. Both Tyr97 and Asp150 are also found to be highly conserved in a variety of other cold-adapted fish trypsins. A variety of mutations were implemented into BT, each targeting a different mechanism of increasing surface softness. A mutation of Asn97Tyr in bovine trypsin was found to increase the flexibility of the Nβ5-Nβ6 loop, while the reverse mutation in Atlantic salmon trypsin decreased loop flexibility. These two mutations are the mutation of a polar residue Asn into the less-polar Tyr, and vice versa. Remarkably, the mutated BT resulted in a shift of activation enthalpy and entropy towards a more cold-adapted trypsin, while the mutation in AST shifted activation parameters towards that of BT. Mutations of BT_{S150D} and AST_{D150S}, which completely change the loop structure and orientation, were found to have similar results. Implementation of a Ser110Lys mutation in BT disrupted hydrogen bonding reducing flexibility of the Nβ5-Nβ6 as the charged lysine prefers to interact with bulk solvent, therefore resulting in an increased loop flexibility. Finally, to reduce hydrophobic surface packing, BT was modified with a Val90Arg mutation, once again increasing loop flexibility. [24]

Mutations increasing surface hydrophobicity and their impacts on stability are a particularly interesting case. In the traditional view of globular protein folding, hydrophobic residues are generally packed within the core of the protein, while hydrophobic surface residues are largely thought to be destabilizing due to their unfavourable interactions with bulk water in the surrounding environment. [1, 33, 34] However, recent progress has challenged this long-standing notion.

Islam *et al.* found that single and double point mutations at Gly14 and 38 in bovine pancreatic trypsin inhibitor (BPTI-[5,55]) introduce small structural rearrangements around the mutation site with dramatic effects on thermal stability of enzyme. By looking at a variety of mutations, it was observed that the thermal stability is increased with the size and hydrophobicity of the mutated residue. Details from crystal structures indicate this stabilization effect is the result of a better local packing near the mutation region, further resulting in better hydration structure [35]. It is thought that increasing the density of hydrophobic packing in specific areas can shield hydrogen-bonded backbone structures from being disrupted by interactions with water. [36] These areas are

particularly sensitive to temperature, thus increasing thermal stability.

Another option to impact protein stability is the modification of salt bridges between pairs of ionic residues at the protein's surface. Like modification of hydrophobic surface packing, this methodology is also quite controversial. Early structural studies by Hendsch and Tidor show that salt bridges, especially those within the internal core of a protein, can be detrimental to protein stability. This appears to be largely due to a larger $\Delta\Delta G_{solv}$ penalty, which is reduced when the salt bridges are allowed more favourable interactions with the solvent. [37] However, work by Kumar *et al.* show this not to be the case in thermophilic dehydrogenase, with salt bridges reported to be highly stabilizing. They argue that much of this stabilizing effect is due to electrostatic contributions to $\Delta\Delta G_{prt}$, the protein neighbourhood of the salt bridge. [38] From bioinformatics analyses of enzymes, thermophiles tend to have have a higher surface charge and a greater number of surface salt bridges than their mesophilic counterparts, leading to a more favourable environment for salt bridges in thermophiles. [39] Furthermore, surface salt bridges appear to be cooperative - that is, the presence of nearby salt bridges can strengthen a salt bridge. This supports the idea that a more favourable $\Delta\Delta G_{prt}$ in the salt bridges of thermophiles can be stabilizing. [40]

Overall, it appears introduction of mutations at the surface of a protein must be done carefully. Even a single mutation can have significant effects on the stability of an enzyme. [41] While some mutations have been found to promote thermostability in some enzymes, introduction of similar mutations in different enzymes have had the opposite effect. As such, it it of great importance to not only pay attention to the target residue for mutation, but also to the surrounding area and any potential impacts a mutation may have in this region. This highlights the need for a more rational approach to enzyme design instead of the traditional directed evolution approach.

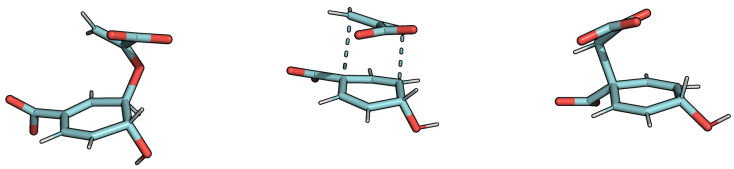
One of the underlying problems with the idea of protein surface softness being a critical feature in tuning T_{opt} is the lack of direct experimental evidence. Due to the nature of enzymes and limitations of equipment, it can be difficult to experimentally probe these loop regions. Instead, much of what is known is inferred from experiments focused on thermostability or kinetic assays. As a result, computational studies of enzymes are critically important in understanding mechanisms of adaptation used by extremophiles. [21]

Chapter 3

Chorismate Mutases

3.1 Structure and Function of Chorismate Mutases

The chorismate mutase family is an important group of enzymes found in bacteria, fungi, and higher plants, where it is a part of the Shikimate pathway. These enzymes catalyze the pericyclic rearrangement of chorismate into prephenate, which serves as a precursor for the biosynthesis of the phenyl-containing amino acids phenylalanine and tyrosine. In this reaction, the C_5-O_7 bond is broken in chorismate, while a C_1-C_9 bond is created, forming prephenate. This reaction proceeds through a chair-like transition state, as seen in Figure 3.1. Due to the simplicity of the catalysis, and the fact that this reaction is unimolecular, chorismate mutases have long served as benchmark systems in computational chemistry. However, despite their prevalence in literature, there is still much debate surrounding these biocatalysts.



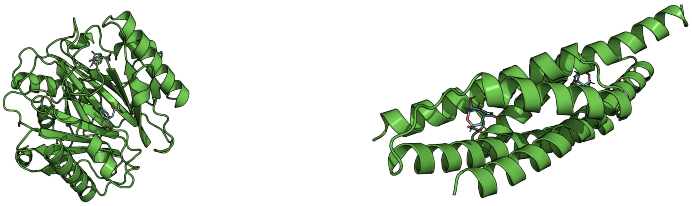
(a) (b) (c)

Figure 3.1: Chorismate mutases catalyze the Claisen rearrangement of chorismate (a) to prephenate (c) via a chair-like transition state (b).

In this thesis, three chorismate mutases and an isochorismate pyruvate lyase were studied. The first, *B. subtilis* chorismate mutase (BsCM) is extremely prevalent in literature, where it has been used as a computational benchmark in multiple QM/MM studies. This is primarily due to the abundance of experimental studies and the simplicity of the enzymatic reaction. The latter two chorismate mutases originate from *B. pumilus* (BpCM) and an unknown organism, where the sequence was obtained from bioprospecting the Antarctic dry desert (M1CM). Prior to this thesis, these latter two enzymes were uncharacterized, and as such were chosen as possible candidates for extremophilic chorismate mutases. All three enzymes belong to the AroH-class of chorismate mutases, which exhibit a

trimeric pseudo-α/β barrel structure (Figure 3.2). In AroH chorismate mutases, the active site is exposed, and lies at the interface between two monomers.

As the name implies, isochorismate pyruvate lyase (henceforth referred to by PchB, the corresponding gene in *P. aeruginosa*) catalyzes the elimination of the enolpyruvyl moeity from isochorismate, forming salicylate as a product. However, PchB has been observed to have promiscuous activity: in addition to its primary function, it exhibits chorismate mutase activity, albeit with a catalytic efficiency two orders of magnitude lower than monofunctional chorismate mutases. Structurally, PchB resembles chorismate mutases of the all-helix-bundle AroQ class (Figure 3.2), such as those from *E. coli* [42]. Due to this, and the minor chorismate mutase activity of PchB, it is hypothesized that a gene-duplication event in an AroQ-class CM is the origin of isochorismate pyruvate lyases. With the minor chorismate mutase activity of PchB, it serves as an interesting system to be compared to monofunctional chorismate mutases in the understanding of their evolutionary mechanisms.



(a) (b)

Figure 3.2: AroH (a, BsCM from PDB 1COM [43, 44]) and AroQ (b, EcCM from PDB 1ECM [42]) chorismate mutases have vastly different tertiary structures, yet are both responsible for the catalysis of chorismate to prephenate. Despite the highly different folds, they both have similar active sites, with similar residues fulfilling the same roles.

Figure 3.3: Multiple sequence alignment of AroH enzymes BsCM, BpCM, and M1CM. Residues in white are unconserved across the three enzymes.

3.1.1 Role of the Chorismate Mutase Active Site in Catalysis

Chorismate mutases are one of the few enzymes in which the enzyme-catalyzed reaction is directly comparable to the uncatalyzed reaction in solution. Both the catalyzed and uncatalyzed reactions proceed through a pericyclic rearrangement. [43, 45] The catalysis occurring within the active sites of chorismate mutases is not a true "chemical" catalysis - that is, there are no residues within the active

site contributing to proton transfer, neither are the motions of the residues significant along the reaction pathway. [46] Nevertheless, the enzyme-catalyzed reaction has a 10^6 increase in k_{cat} compared to the uncatalyzed reaction, so it is likely that the active site serves to stabilize the transition state during the reaction. [46]

Upon binding of the substrate in the active site of BsCM, the carboxylate group on the cyclohexadienyl moeity forms strong hydrogen bonds with Arg63. To initiate the conversion of chorismate to prephenate, hydrogen bonding between Glu78/Cys75 and C_4-OH hydroxyl group promotes elongation of the ether O_7 of chorismate prior to C_5-O_7 cleavage. This process produces a large negative charge on the oxygen. It is critical that this negative charge is counterbalanced, which is counterbalanced by the presence of Arg90. In mutational studies, Kleinhofer et al. replaced Arg90 with non-natural citrulline, which is structurally similar to arginine, but uncharged. This mutation results in a decrease in k_{cat} while K_M remains unchanged, showing that the presence of this positively charge hydrogen bond donor is crucial in the stabilization of the substrate leading to the formation of the chair-like transition state. Despite the structural differences compared to BsCM and other AroH-class chorismate mutases, AroQ-class enzymes such as EcCM and ScCM also exhibit similar active site features to promote C_5-O_7 cleavage and subsequent stabilization. [47]

To complete the formation of prephenate, the substrate proceeds through a chair-like transition state, where a C_1-C_9 bond must form. In order to facilitate this bond formation, the pyruvate moeity must be properly aligned with the cyclohexadienyl moiety. If this moeity is not sufficiently restrained, the activated carbon is able to form a C-C bond with many of the other carbons in the cyclohexadienyl moeity. There have been multiple key residues identified in ensuring this alignment. The carboxylate group of pyruvate . While there is a lack of direct experimental evidence directly implicating Tyr108 and Glu as important residues, assay studies on mutants disrupting the orientation of the pyruvate moeity in the transition state show a disappearing of the chorismate substrate with a decreased production of prephenate as the product in comparison to the wild-type chorismate mutase.

In *E. coli* chorismate mutase, two mutations targeting Val35 had profoundly differing outcomes on k_{cat}: a mutation of Val35Ile was found to increase k_{cat}, while a Val35Ala mutation had the opposite effect, decreasing k_{cat}. As this valine is non-polar and not directly responsible for stabilization of the transition state, it is proposed that the Val35Ile increases k_{cat} by forcing in the transition state into a more compressed geometry than the Val35Ala mutation. This is further supported by similar mutations to Val38Ile in PchB, which also results in an increase in k_{cat}. Thus, it appears that while the chorismate mutase reaction is unimolecular and the active site does not play a direct role in the conversion of chorismate to prephenate, the active site of the enzyme appears to stabilize the transition state.

Despite the prevalence of chorismate mutases in literature and how much we know about the enzymes, there is still much unknown about the reaction, namely the extent of the roles enthalpy and entropy play in the reaction. Thus,

it is important to further study these enzymes to better understand the roles activation enthalpy and entropy play in catalyzing this reaction.

Chapter 4

Computational Methods

In this chapter I will describe the computational methodology used to explore enzyme adaptation to extreme environments. It begins with an overview of statistical mechanics, laying the foundations for molecular mechanics, which further introduces molecular dynamics simulations. Next, the theory of free energy calculations is introduced, providing the framework for the final section on the empirical valence bond method and how it is utilized to probe thermodynamic activation parameters of enzyme catalysis.

4.1 Statistical Mechanics

Molecular simulation attempts to answer the question of how microscopic behaviour of a system at the atomic or molecular scale is intrinsically linked to macroscopic aggregate properties accessible experimentally. This link is fundamentally answered through the implementation of statistical mechanics, from which we are able to calculate macroscopic thermodynamic properties solely from sufficient sampling of atomic positions and momenta.

The heart of statistical mechanics is based on calculating the probability of a system being in a particular microstate when the system is in thermal equilibrium. To do this, we can introduce one of the most powerful equations in statistical mechanics, the Boltzmann distribution [48–50] (also known as the canonical ensemble, where number of particles N, volume V, and energy E are fixed quantities):

$$\mathcal{P}(s) = \frac{1}{Q} e^{-\beta \mathcal{H}(s)} \tag{4.1}$$

where Q is the partition function of the system, $\mathcal{P}(s)$ is the probability of finding the system in a microstate s, $\mathcal{H}(s)$ is the Hamiltonian of the system in microstate s, and $\beta = (k_B T)^{-1}$ is the Boltzmann constant at the system temperature. The partition function Q of the system is one of the most important concepts arising frequently in statistical mechanics. As we know that for any given system the total probability of all microstates $\mathcal{P}(s)$ must sum to unity:

$$\begin{aligned} 1 &= \sum_s \mathcal{P}(s) \\ &= \sum_s \frac{1}{Q} e^{-\beta \mathcal{H}(s)} \end{aligned} \tag{4.2}$$

The partition function is a constant independent of any particular microstate,

so we can rearrange Equation 4.2, giving:

$$Q = \sum_s e^{-\beta \mathcal{H}(s)} \tag{4.3}$$

Therefore, from Equation 4.3, we can define the partition function Q as the sum of all microstates in a system. [48]

As stated beforehand, the Boltzmann distribution is one of the most useful equations in statistical mechanics. Using it as a starting point, we can find the ensemble average of a thermodynamic observable $\langle \mathcal{O} \rangle$ for a system, as seen in Equation 4.4

$$\langle \mathcal{O} \rangle = \sum_s \mathcal{O}_s \mathcal{P}_s, \tag{4.4}$$

where $\mathcal{O}_s$ is the value of the observable in microstate s. Calculation of observable quantities, usually the internal energy E, are essential in the relation of statistical mechanics to macroscopic thermodynamics.

Starting from the calculation of the internal energy

$$\begin{aligned} \langle E \rangle &= \sum_s E_s \mathcal{P}_s \\ &= \frac{1}{Q} \sum_s E_s e^{-\beta E_s} \end{aligned} \tag{4.5}$$

it is clear we can express the sum as the fraction $-\partial Q / \partial \beta$, giving a relation between $\langle E \rangle$ and the logarithm of the partition function:

$$\begin{aligned} \langle E \rangle &= -\frac{1}{Q} \frac{\partial Q}{\partial \beta} \\ &= -\frac{\partial \ln Q}{\partial \beta} \end{aligned} \tag{4.6}$$

Substituting in $\partial \beta / \partial T$, Equation 4.6 becomes:

$$\frac{\langle E \rangle}{T^2} = -\frac{\partial (k_B \ln Q)}{\partial T} \tag{4.7}$$

Comparing this to the macroscopic thermodynamic relation:

$$\frac{E}{T^2} = \frac{\partial (A/T)}{\partial T} \tag{4.8}$$

it becomes clear that we can express the Helmholtz free energy as a function of the logarithm of the partition function:

$$A(N, V, T) = \beta^{-1} \ln Q(N, V, T) \tag{4.9}$$

This is known as the characteristic function of the ensemble. [51–53]

However, it is important to note that while the canonical ensemble is often the simplest ensemble and the first to be introduced in statistical mechanics

textbooks, it is not the only important ensemble in statistical mechanics. [48–51, 53, 54] Other important statistical mechanics ensembles exist, depending on the thermodynamic state variables used. Thus, depending on the macroscopic thermodynamic quantity of interest and the type of simulation, it may be more suitable to sample from a different ensemble. A summary of the most common statistical ensembles can be seen in Table 4.1. [51]

Ensemble	**State variables**	**Characteristic function**
Canonical	N, V, E	Helmholtz free energy (A)
Microcanonical	N, V, T	Entropy (S)
Isothermal-isobaric	N, P, T	Gibbs free energy (G)
Grand canonical	μ, V, T	Pressure-volume work (pV)

Table 4.1: Common ensembles used in molecular simulations, their associated state variables, and characteristic function.

Of course, sufficient sampling of phase states is incredibly important for calculation of macroscopic thermodynamic properties in statistical mechanics. A fundamental assumption of statistical mechanics allowing its use in molecular dynamics simulations is the ergodic hypothesis: if the Hamiltonian of the system is ergodic, over sufficiently large time scales approaching infinity, time-evolution of the system should allow for a large enough area of phase space to be sampled. Thus, we can make use of the assumption that in this limit, the time-averaged value of a thermodynamic quantity is approximately equal to the ensemble average of the quantity over all of phase space.

4.2 Molecular Mechanics

The foundation of molecular mechanics (MM) lies in the Born-Oppenheimer approximation. As the mass of a proton is three orders of magnitude larger than that of an electron, we can decouple their motion under the assumption that electron motion has negligible influence on that of the proton. This allows us to model entire atoms as balls, while atomic bonds can be modelled as springs connecting the balls.

A fundamental concept in modelling molecular energies in this ball-and-spring model of chemical systems is the use of empirical or semi-empirical energy functions, known as force fields. These force field models allow complex molecular energies to be modeled based on atomic positions $\mathbf{r}^N$. Two common force field models are OPLS and CHARMM, which share the common form based on Figure 4.1:

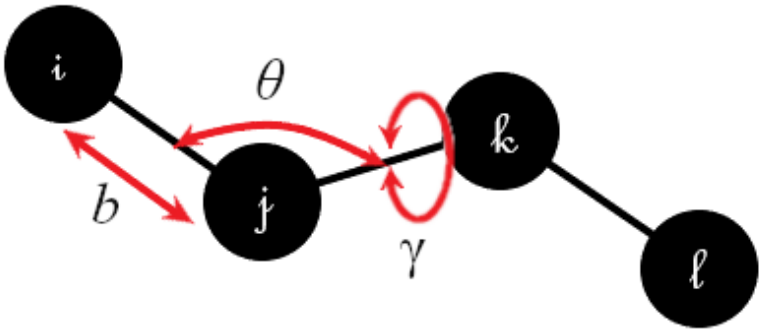


Figure 4.1: Bond stretching, bond angle bending, and torsion in four consecutive atoms i, j, k, and l. These terms form the basics for molecular mechanics force fields.

$$U(\mathbf{r}^N) = \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{imp.} \frac{1}{2} k_\varphi (\varphi - \varphi_0)^2 + \sum_{tor.} \frac{1}{2} V_n (1 - \cos(n\omega - \gamma)) + \sum_{ij} \left[4\varepsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} + \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \right] \tag{4.10}$$

The first four terms in this equation describe bonded interactions, and calculate the energies of bond stretching, angle bending, and torsion/improper twisted, where k_b, k_θ , and k_φ are the bond, angle, and improper force constants, with corresponding equilibrium values b_0, θ_0, and φ_0. The fourth bonded term describes a periodic torsion potential with. Considering four consecutive atoms i, j, k, and l, the torsional angle is the angle between the plane formed by atoms i, j, k, and atoms j, k, l. In the case where i, j, k, and l are non-consecutive (e.g. the N-terminal NH_3^+, the improper dihedral term must instead be considered.

The final two common terms in the OPLS and CHARMM force field models describe the van der Waals (vdW) and electrostatic $i+4$ non-bonded interactions. The van der Waals interactions are, in these two force fields, modelled by the Lennard-Jones 6-12 interaction, where ε_{ij} is the potential well depth at $r_{min} = 2^{1/6}\sigma_{ij}$, σ_{ij} is the distance at which the inter-particle potential $V_{ij} = 0$, and r_{ij} is the distance between the two particles. Electrostatic interactions are modelled by the Coulomb potential, where ϵ_0 is the vacuum permittivity, q_i and q_j are the charges of particle i and j, and r_{ij} is the distance between the two particles. In addition to these terms different force field models may have other terms to account for other special interactions, such as polarization and hydrogen bonding.

In a force field parameterization, Lennard-Jones parameters are given per-atom, that is they are defined using the self-interaction σ_{ii} and ε_{ii} between particle i and i. When considering the interaction between two non-identical particles i and j, we must use some form of mixing. The most common mixing rule is the simple geometric rule:

$$\sigma_{ij} = \sigma_i \sigma_j = \sigma_i^{1/2} \sigma_j^{1/2} \tag{4.11}$$

The same mixing can also be used to construct ε_{ij}.

It is important to note, however, that the harmonic approximation of the Lennard-Jones potential is not suitable for instances of bond breaking and forming, as real bonds are not truly harmonic in nature. This approximation is generally fine for standard molecular dynamics simulations, but may not be suitable for certain types of free energy calculations. In these cases, it is much better to use a Morse potential, given by the form:

$$V(r) = \varepsilon_{ij}(1 - e^{-a(r-r_e)})^2 \tag{4.12}$$

where ε_{ij} is the depth of the well, a is the well width, and r_e is the equilibrium bond distance. A comparison of the harmonic and Morse potentials in comparison to the "true" Lennard-Jones potential can be seen in Figure 4.2

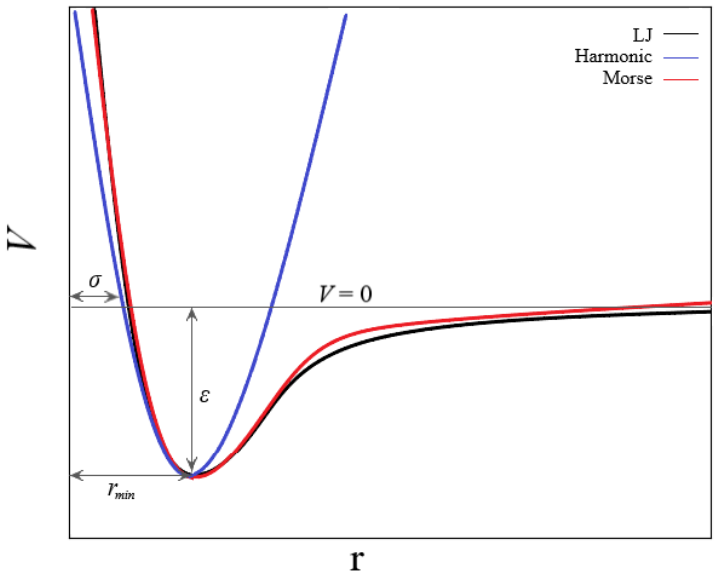


Figure 4.2: Harmonic and Morse potentials in comparison to the Lennard-Jones 6-12 potential. The harmonic approximation of bond stretching is suitable for small-magnitude bond vibrations near r_{min}, but for longer-range bond breaking and forming the Morse potential is much better suited.

4.3 Molecular Dynamics

One common simulation technique utilizing MM is molecular dynamics (MD) simulations. MD simulations treat the system of interest as a classical, many-body problem where the evolution of the system is determined by solving Newton's second law of motion:

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \tag{4.13}$$

which describes the motion of a particle of mass m_i along a coordinate x_i when acted upon by a force in that direction F_{x_i}. For systems where there is more than $N = 3$ particles, as is the case in most molecular dynamics simulations, there is no analytic solution to the system of equations, and the system must be solved numerically.

As the particle masses are known quantities, and the force acting on a particle can be calculated from the negative gradient of the potential energy function found in Equation 4.10:

$$\mathbf{F}_{\mathbf{x_i}} = -\frac{\partial U}{\partial x_i} \tag{4.14}$$

we can find easily find the acceleration of a particle i in molecular dynamics, and use it to calculate the trajectory of the particle. This is most easily accomplished using a variety of finite difference algorithms, where integration is broken into discrete time steps separated by a small time step δt.

The simplest and most commonly used finite difference algorithm for calculating the new positions of atoms in the system is the Verlet algorithm. This algorithm considers the position at a future time $\mathbf{r}_i(t+\delta t)$ and calculates it relative to the current position $\mathbf{r}_i(t)$ and the previous position, $\mathbf{r}_i(t-\delta t)$. Taylor expanding both quantities, we obtain:

$$\mathbf{r}_i(t+\delta t) = \mathbf{r}_i(t) + \delta t \mathbf{v}_i(t) + \frac{1}{2}\delta t^2 \mathbf{a}_i(t) + \mathcal{O}(t^3) \tag{4.15}$$

and

$$\mathbf{r}_i(t-\delta t) = \mathbf{r}_i(t) - \delta t \mathbf{v}_i(t) + \frac{1}{2}\delta t^2 \mathbf{a}_i(t) + \mathcal{O}(t^3) \tag{4.16}$$

Adding these two expansions, we can solve for the updated position as:

$$\mathbf{r}_i(t+\delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\delta t) + \delta t^2 \mathbf{a}_i(t) \tag{4.17}$$

At the start of the simulation, positions are known, and velocities are initialized randomly according to the Maxwell-Boltzmann distribution at the system temperature T:

$$P(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} exp \frac{1}{2}\frac{m_i v_i^2}{k_B T} \tag{4.18}$$

4.3.1 Molecular Dynamics Algorithms

In molecular dynamics simulations, it is often desirable to control some of the thermodynamic variables in order to allow sampling from a particular ensemble. While some parameters are easy to hold constant, such as energy or volume, others such as temperature or pressure, are harder to keep constant. In order to aid with this, a variety of algorithms have been developed.

As system temperature is a function of the kinetic energy of particles in the system, it makes inherent sense to control system temperature by controlling the velocities of the system. Such algorithms are known as thermostats. Among the most common thermostats used to control the temperature of a molecular dynamics simulation is the Berendsen thermostat. While it is not the earliest algorithm used for such purposes, previous attempts required modifying the Lagrangian or scaling the velocities in an unphysical manner. However, these previous algorithms contain some instabilities which require corrections after some number of steps. With the Berendsen thermostat the system is weakly coupled to a heat bath held at constant temperature and velocities are scaled at each step by a factor of:

$$\lambda = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{T_{ref}}{T(t-\Delta t/2)}\right)\right]^{1/2} \tag{4.19}$$

where T_{ref} is the desired system temperature (i.e. that of the heat bath), T is the current system temperature calculated from current particle velocities, Δt is the time step being used, and τ_T is a coupling parameter controlling how strongly the system is coupled to the heat bath.

4.4 Analysis of Structural Details of Enzymes

One of the powerful utilities of computational studies of enzymes is the ability to calculate structural quantities otherwise inaccessible from experiment. In this thesis, a variety of such methods were considered, including root-mean-square-deviation, root-mean-square fluctuations, and normal mode analysis.

Root-mean-square deviation calculations are a tool often used to compare protein similarity. It is calculated as

$$\mathbf{RMSD} = \sum_i \langle ||\delta r_i||^2 \rangle \tag{4.20}$$

where δr_i is the displacement of particle i between two structures. RMSD can either be used to compare the similarity of two related proteins, or used to calculate how much a protein deforms in time throughout a time-trajectory, e.g. from molecular simulation.

Root-mean-square fluctuations are one of the simplest and most widely used methods for comparing protein structure. [55–57] It is calculated in almost the same way as RMSD:

$$\mathbf{RMSF}_i = \langle ||\delta r_i||^2 \rangle \tag{4.21}$$

where δr_i is the displacement of particle i from frame-to-frame in a molecular dynamics simulation. [58] Whereas RMSD looks at structure similarity accross the entire protein, RMSF quantifies the amount a particle fluctuates throughout the course of the simulation. When particle i is taken to be C_α or averaged over a residue backbone, a RMSF plot can highlight areas of a protein more prone to movement. [24, 55, 56] However, RMSF profiles generally require large ensembles of a protein's conformation, such as those generacted from a molecular dynamics trajectory. Therefore, RMSF profiles can be quite expensive to compute.

A less-commonly used method for calculating structural dynamics similar to those found in RSMF calculations is an elastic network model normal mode analysis (ENM-NMA) [59, 60]. ENM-NMA is based on a classical mechanical treatment of the treating the C_α atoms of the protein backbone as ball-and-spring model. Using the C_α force field derived by Hinsen *et al.* [61], the interaction potential of a spring between residues i and j can be expressed as:

$$V = \sum_{ij} V_{ij} = \sum_{ij} \frac{1}{2}\gamma(s_{ij} - s_{ij}^0)^2 \tag{4.22}$$

where γ is a stiffness constant assumed to be equal for all pairwise interactions, s_{ij} is the instantaneous separation vector between the two atoms, and s_{ij}^0 is the equilibrium separation vector. From the the total interaction potential over the

entire protein, we can construct a Hessian from the second derivatives of the pairwise interactions: [62]

$$V_{ij} = \frac{\partial V}{\partial x_i \partial x_j} \tag{4.23}$$

The resulting matrix is of dimension $3N \times 3N$, where N is the number of residues in the protein. Eigenvectors resulting from the diagonalization of the Hessian from Equation 4.23, weighted by the masses of the corresponding residues, represent the intrinsic deformation modes of the protein structure, while the eigenvalues correspond to the squared frequency of each mode. The first 6 nodes are generally excluded from calculation, as they correspond to simple translation and rotation of the protein. [59]

In comparison to RMSF, calculation of an ENM-NMA is highly efficient. However, there are a few caveats of normal mode analyses of proteins. The first, and arguably most important, is there is much debate over how accurately the results of a NMA reflect true motions of a protein. However, low frequency modes corresponding to domain motion have been found to have good correlation to functionally-relevant protein motions. [63, 64] Secondly, in the case of comparing a single point mutation between two proteins which does not change the overall structure of the protein, normal mode analysis between the two proteins will be virtually indistinguishable. [60] As a result, RMSF is much better at predicting differences arising from single point mutations.

4.5 Free Energy Calculations

Oftentimes it is desirable to obtain energetics related to change, especially in the case of enzyme catalysis where a reactant is converted into a product. As the most common force field models used are not suited to this, and the time scale of these changes are generally longer than what is computationally feasible, the system can be perturbed along a chemical or alchemical pathway to calculate the relative free energy change between the states. Such free energy perturbations (FEP) take advantage of the ability to generate a series of trajectories along a discretized pathway from the reactant state to the product state.

Most simply, for free energy calculations we want to find the free energy change between two states A and B, which we can express as a logarithm of the isothermal-isobaric partition function:

$$\begin{aligned} \Delta G(N,P,T) &= G_B(N,P,T) - G_A(N,P,T) \\ &= -k_B T \ln Q_B(N,P,T) + k_B T \; lnQ_A(N,P,T) \\ &= -k_B T \ln \frac{Q_B(N,P,T)}{Q_A(N,P,T)} \end{aligned} \tag{4.24}$$

which corresponds to a Accurate calculation of free energies from this requires an overlap in the phase space between states A and B, with a stronger overlap providing higher accuracy. However, in the case of biomolecular simulations, especially enzyme catalysis, it is often a condition that $|E_B - E_A| \gg 1$, which implies a poor overlap in phase space.

To overcome this poor overlap and improve the accuracy of free energy calculations in enzyme catalysis, we can implement a set of coupling parameters $\lambda_n \in [0,1]$, which correspond to a set of evenly-spaced intermediate states between the reactant state ($\lambda_0 = 0$) and product state ($\lambda_n = 1$). As free energy is a thermodynamic state function it is path independent, so we can express Equation 4.24 as:

$$\begin{aligned}\Delta G &= -k_B T \ln\left(\frac{Q_B}{Q_{\lambda_{n-1}}} \times \frac{Q_{\lambda_{n-1}}}{Q_{\lambda_{n-2}}} \times \ldots \times \frac{Q_{\lambda 2}}{Q_{\lambda_1}} \times \frac{Q_{\lambda 1}}{Q_A}\right) \\ &= (G_B - G_{\lambda_{n-1}}) \times (G_{\lambda_{n-1}} - G_{\lambda_{n-2}}) \times \ldots \times (G_{\lambda_2} - G_{\lambda_1}) \times (G_{\lambda_1} - G_A) \\ &= G_B - G_A\end{aligned} \tag{4.25}$$

Implementation of a free energy perturbation is accomplished by scaling the interaction potentials of the system within each step. For each λ step, this is done as:

$$\Delta U_i(\lambda_i) = U_A(1 - \lambda_i) + U_B\lambda_i \tag{4.26}$$

That is, at any given λ step, the potential energy difference is scaled such that it experiences a linear combination of interaction potentials from states A and B. These hybrid states are unphysical in nature, however the path independence of free energy allows their use in calculations. Combining this with Equations 4.3 and 4.6, we obtain the following expression for the calculation of free energy as a sum of the free energy changes between intermediate states:

$$\Delta G = -\beta \sum_{n=0}^{n-1} \ln\langle \exp[-\beta(U_n(\lambda_{n+1}) - U_n(\lambda_n)]\rangle_n \tag{4.27}$$

4.6 Empirical Valence Bond Calculations

Methodologies in computational chemistry based upon the Born-Oppenheimer approximation are not particularly suitable for calculations dependent on the motion of electrons. Of particular relevance to this project is their inability to accurately model chemical reactions and catalysis. Thus, to explore thermodynamic properties relevant to chemical reactions, we must consider some quantum mechanical treatment of the system. For small systems, popular quantum mechanical (QM) methods such as density functional theory (DFT) are suitable for computing reaction energetics. However, when considering catalysis in biomolecules, the system size becomes prohibitively expensive for traditional methods. One particularly suitable method for these calculations is the empirical valence bond (EVB) method.

EVB is a semi-empirical methodology which can be used to calculate reaction free energy profiles for reactions in both solution and enzymes. EVB utilizes a free energy perturbation (FEP) approach to drive the reaction from the reactant state to the product state in discrete steps, passing through the transition state along the way. Along the FEP pathway, trajectories are sampled using MD.

A powerful advantage of EVB is that it is initially calibrated using reaction energetics for the uncatalyzed reaction in solution (also known as the reference reaction), and further systems involving this reaction (i.e. homologous enzymes) can utilize the same calibration parameters. This initial calibration step requires knowledge of the activation free energy and reaction free energy of the reference reaction, which can be obtained either from experiment or QM calculations. EVB draws inspiration from Marcus theory, which describes the reaction in a set of diabatic (resonance) structures, representing the reactant, transition, and product states. Mixing of these resonance states through FEP allows mapping of the adiabatic potential energy surface of the reaction, and more importantly, the transition state. For any given λ state, EVB calculates the diabatic energy profile equivalent to the Marcus parabola in electron transfer theory. Performing EVB along the perturbation pathway from reactant state to product state allows for the construction of a reaction energy profile, as seen in Figure 4.3 The energy of each state i can be obtained from the diagonal elements of the EVB Hamiltonian matrix, given by:

$$\varepsilon_i = \mathcal{H}_{ii} = \alpha^i_{gas} + U^i_{b,rr} + U^i_{nb,rr} + U^i_{nb,rs} + U^i_{ss} \tag{4.28}$$

In Equation 4.28, α^i_{gas} is the energy of diabatic state i in the gas phase with all fragments at infinite separation. $U^i_{b,rr}$ is the bonded potential energy within reacting atoms. However, in comparison to Eq.4.10 the potential energies of bonded atoms in EVB utilize a Morse potential instead of the harmonic potential seen in standard MM force fields. $U^i_{nb,rr}$, $U^i_{nb,rs}$ and U^i_{ss} are the non-bonded (i.e. vdW and electrostatic) interactions within the reacting atoms, between the reacting atoms and the surrounding protein and solvent, and within the surroundings, respectively.

The off-diagonal component of the Hamiltonian is either set to be constant, or given by the exponential function:

$$\mathcal{H}_{ij} = \mathcal{H}_{ji}\mathcal{H}_{ji} = A_{ij}e^{-\mu_{ij}r_{ij}} \tag{4.29}$$

where A_{ij} and μ_{ij} are fitting parameters. The parameters calibrated to experimental or QM calculations using simulations of the reference reaction. In most cases, μ_{ij} can be set to 0, leaving a constant $H_{ij} = A_{ij}$.

With a fully-constructed EVB Hamiltonian $\mathcal{H}_{EVB}$, which for a 2-state reaction is:

$$\mathcal{H} = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$$

we can obtain the ground-state energy E_g by solving the secular equation:

$$\mathcal{H}_{EVB}\mathbf{c}_g = E_g\mathbf{c}_g \tag{4.31}$$

where $\mathbf{c}_g$ are the corresponding eigenvectors of the system. From here, we can calculate ground state energies of the system by solving the characteristic equation:

$$|\mathcal{H} - \mathbf{I}E_g| = 0 \tag{4.32}$$

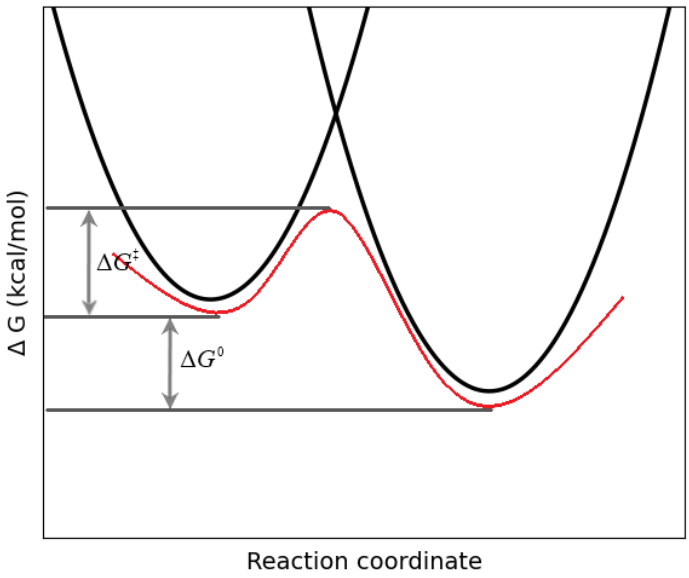


Figure 4.3: Marcus parabolas for the reactant and product states, along with reaction energy profile from tracing the minima of Marcus parabolas in a perturbation from reactant to product state.

$$\begin{vmatrix} H_{11} - E_g & H_{12} \\ H_{21} & H_{22} - E_g \end{vmatrix} = 0$$

which yeilds a ground-state energy of:

$$E_g = \frac{1}{2}(H_{11} + H_{22}) - \sqrt{(H_{11} - H_{22})^2 + 4H_{12}} \tag{4.34}$$

Obtaining the free energy changes from ground-state energies calculated from EVB can be done using the FEP-umbrella sampling method

One particularly powerful advantage of EVB is the universality of the calibrated α^i_{gas} and $\mathcal{H}_{ij}$. They are both independent of temperature and the surrounding environment, so once these parameters are fitted using the reaction free energy and activation free energy of the uncatalyzed reaction, they can be used for any system involving the same reaction. Thus, the initial calibration is performed once, and all subsequent calculations for the reaction utilize the calibrated values, even across several homologous enzymes.

4.6.1 Calculation of Thermodynamics Activation Parameters

Understanding how enzymes have evolved to adapt to extreme environments requires more than just knowledge of the activation free energy $\Delta G^{\ddagger}$ - we must also be able to decompose it into enthalpic and entropic contributions in order to understand how these parameters are affected by adaptations enzymes have made to function in differing environments. By performing EVB calculations at a series of n different temperatures, we are able to construct an Arrhenius plot of $\Delta G^{\ddagger}/T$ versus $1/T$. [20, 65] Taking a linear regression of this plot, we can extract the activation enthalpy from the slope:

$$\Delta H^{\ddagger} = \frac{\sum_{i=1}^{n}(T_i^{-1} - \langle T^{-1}\rangle)(\Delta G_i^{\ddagger} T_i^{-1} - \langle \Delta G^{\ddagger} T^{-1}\rangle)}{\sum_{i=1}^{n}(T_i^{-1} - \langle T^{-1}\rangle)^2} \tag{4.35}$$

and the activation entropy from the intercept:

$$\Delta S^{\ddagger} = \langle \Delta G^{\ddagger} T^{-1}\rangle - \Delta H^{\ddagger}\langle T^{-1}\rangle \tag{4.36}$$

Calculation of the activation enthalpy is also useful for calculating the internal energetics of the enzymatic reaction. As we know the activation enthalpy is the energy difference between $ES^{\ddagger}$ and ES in turns of heat, we can express it as:

$$\Delta H^{\ddagger} = \Delta U^{\ddagger} + P\Delta V^{\ddagger} \tag{4.37}$$

where U is the internal energy, P is pressure, and V is volume. However, the pressure-volume term in this equation is negligible, thus $\Delta H^{\ddagger}$ corresponds only to the change in internal energy of the reaction.

Using the EVB method, we can further decompose the energetics in terms of contributions from within the reacting region rr (i.e. atoms treated with the EVB method, where interaction parameters change during the simulation), between reacting region and the surrounding protein-water environment rs, and interactions within the surroundings ss:

$$\Delta H^{\ddagger} = \Delta U_{rr}^{\ddagger} + \Delta U_{rs}^{\ddagger} + \Delta U_{ss}^{\ddagger} \tag{4.38}$$

This last term, however, is difficult to calculate directly due to the large number of protein-water interactions, so convergence directly from MD simulations is impossible. Instead, it can be calculated starting by obtaining enthalpy from the Arrhenius plot, and knowledge of the other two internal energetics parameters, $\Delta U_{rr}^{\ddagger}$ and $\Delta U_{rs}^{\ddagger}$. Such a decomposition of internal energetics is particularly useful in comparing homologous enzymes adapted to different environments, as differences in $\Delta U_{rr}^{\ddagger}$, $\Delta U_{rs}^{\ddagger}$, and $\Delta U_{ss}^{\ddagger}$ can help pinpoint the origins of enthalpic changes.

Chapter 5

Concluding Remarks and Summary of the papers

Within this section, I will highlight the main results from the manuscripts following this section, with an emphasis on how they connect to the story line and goals of this thesis.

As stated in Chapter 1, the primary objectives of this thesis were to utilize computational chemistry techniques in order to gain a better understanding of enzyme adaptation to extreme environments, with an end goal of creating an extremophilic enzyme using principles of rational design. This involves starting from a mesophilic enzyme and extremophilic homolog of the enzyme, using computational analyses to gain a deeper understanding of the enzymes' catalytic function and structure, and utilizing the insights gained from these computational methods to intentionally introduce specific mutations that enable the mesophilic enzyme to operate in a different environment.

The starting point for this work was the monofunctional AroH-class of chorismate mutases. The chorismate-to-prephenate reaction catalyzed by these enzymes is relatively simple in comparison to many other enzymes, making it a prime candidate to focus on the factors affecting enzyme adaption. Furthermore, these enzymes have commonly utilized as benchmarking systems for computational methods, leading to an abundance of literature data on both the reaction thermodynamics and the structure of the enzymes, which could be used as a comparison.

In paper I, we investigate the catalytic origins of chorismate mutases by conducting EVB simulations on monofunctional BsCM and the promiscuous IPL PchB. From EVB, we were able to confirm the reaction in both enzymes is enthalpy-driven in comparison to the uncatalyzed reaction. Furthermore, from a decomposition of the internal energetics of the reaction, we can pinpoint the significant reduction in enthalpy to originate from a more favourable $\Delta U^{\ddagger}_{bonded}$. Understanding the enthalpy-entropy balance of enzymes and the origins of these effects serves as an important starting point for further studies investigating the effects of mutations of these enzymes. As such, the BsCM-catalyzed reaction studied in this paper served as a starting point for Paper II.

Paper II presents the characterization of a new chorismate mutase from *Bacillus pumilus*. This organism is an Arctic bacterium, and recent characterization of a lipase from *B. pumilus* (pLipA) in comparison with mesophilic bacterial lipases (mLipA) demonstrate behavior indication pLipA is cold-adapted. [66] As our new chorismate mutase originates from the same organism, the original hypothesis was that BpCM would be a cold-adapted chorismate mutase, which have yet to be characterized. However, within this manuscript the characterization of BpCM

with both experimental methods and the computational EVB/MM Arrhenius plot approach indicate mesophilic behaviour for this enzyme. In comparison to BsCM from Paper I, BpCM exhibits a $\Delta G^{\ddagger}$ approximately 0.5-1.0 kcal/mol higher. From our Arrhenius plots, we determined this difference to be entirely from entropic contributions to the free energy, as the activation enthalpies were identical. A structural analysis of both enzymes indicate BpCM is marginally less flexible than the BsCM; However these entropic and flexibility differences are not significant enough to result in BpCM being adapted to any particular temperature extreme. This paper also highlights an important point: as evident from the differences in behavior between BpCM and other enzymes from *B. pumilus*, if the enzyme in question is not rate-limiting in the survival of the organism, it may have not faced sufficient evolutionary pressure for adaptation to environmental extremes. Thus, even if an organism in general exhibits extremophilic behavior, not all enzymes within the organism are necessarily extremophilic. This is a particularly important point in the curation of sequence databases, especially for machine learning applications.

In Paper III, we present the synthesis and characterization of a new chorismate mutase genome sequence collected from bioprospecting in the Antarctic dry deserts. A BLAST of the sequence indicated the most similar match to be a chorismate mutase from *Thermus thermophilus*, a thermophilic chorismate mutase. Characterization of this new chorismate mutase (M1CM) using kinetic assays and EVB/MM Arrhenius plots indicate thermophilic behavior for this enzyme. While the $\Delta G^{\ddagger}$ is comparable to that of BpCM at room temperature, there is a shift of the enthalpy-entropy balance of the enzyme, with an increase in $\Delta H^{\ddagger}$ and a less-negative $\Delta S^{\ddagger}$. Furthermore, the availability of a newly-characterized mesophilic and thermophilic chorismate mutase served as an excellent starting point for understanding the structural differences which influence the differences in thermodynamic activation parameters in chorismate mutases. Using a variety of computational tools such as ENM-NMA and MD simulations to obtain flexibility profiles of the two enzymes. This allowed us to identify regions of the BpCM which were more flexible than the corresponding regions in M1CM, to serve as a starting point for mutational studies. In particular, we targeted a loop region from residues 65-70. This region was identified to be shorter and less flexible in the thermophilic M1CM. This loop is preceded by a short α-helix, and the entire helix-loop motif runs parallel to a much longer helix from residues. By implementing a Glu29Phe mutation in BpCM, we were able to increase the hydrophobic packing density of residues near the surface, leading to a lower flexibility of this loop. In M1CM, there exists a salt-bridge between Arg22 and Glu64 of the two helices, which is not found in BpCM due to a Glu64Gln substitution. Thus, we also independently implemented Gln64Glu in BpCM, also leading to a reduction in the flexibility of this loop region. Using EVB/MM, we further analyzed the effects of these mutations on the thermodynamic activation parameters of the mutant chorismate mutases. Both mutations resulted in a shifting of the enthalpy-entropy balance matching that of the thermophilic M1CM. However, this work is not yet complete. While we are confident in our approach to accurately replicate experimental application

parameters in both highly-studied and new enzymes, to supplement the findings in this paper and strengthen our findings requires further experimental testing.

The ability to easily identify and implement mutations to shift an enzyme's activity to favour an extreme environment is extremely promising for the field of rational enzyme design, with potential applications in both research and industry. The findings presented within this thesis contribute significantly to this goal. To further strengthen our findings, there are a variety of future directions in which this research can go.

First and foremost the mutational studies should be supplemented with experimental studies to verify the results, as previously mentioned. The EVB method provides an extremely efficient method of testing mutations computationally. Further experimental research backing EVB as a viable method for accurate calculations of enzymatic reactions can increase the output of research and application in rational enzyme design. A second avenue of future study is the application of these findings to other enzymes. While our findings have been applied only to chorismate mutase in this thesis, their potential as universal principles to guide the rational design of thermophilic enzymes would be an important asset for applications. The results of paper III also propose a very interesting question for future research. Our computational studies are limited in some aspects. While we have demonstrated the ability to accurately calculate reaction free energies in enzymes, computational studies of enzyme stability at higher temperatures are still difficult. Thus, if further extensive research can pinpoint the shifting of the enthalpy-entropy balance as the sole contributing factor to enzyme thermal stability, this promotes a new, inexpensive method to investigate thermal stability computationally. The EVB method provides a plethora of information on the dynamics of enzymatic reactions to further investigate this link.

With recent advances in machine learning [67], *de novo* protein design may overtake rational design as the preferred method of protein design in the not-too-distant future. However, for now this is limited by a lack of available sequence databases targeting extremophiles, being further complicated by the aforementioned key finding from paper II. Until these problems can be solved, improving approaches to rational enzyme design provides a path towards improving enzyme design for practical purposes.

Bibliography

(1) Nelson, D. L., Cox, M. M., *Principles of Biochemistry*; W.H. Freeman and Company: 2008.

(2) De Morveau, L., Gay-Lussac, J., Arago, F., Chevreul, M., Berthelot, M., Mascart, É., Haller, A., *Annales de chimie et de physique*; Landmarks II, scientific journals; Masson.: 1833.

(3) Tizard, W., *The Theory and Practice of Brewing Illustrated*, 1857.

(4) Kühne, W., *Verhandlungen*; v. 1, 1877.

(5) Nobelprize.org The Nobel Prize in Chemistry 1907.

(6) Nobelprize.org The Nobel Prize in Chemistry 1946.

(7) Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., Sarma, V. R. *Nature* **1965**, vol. 206, no. 4986, 757–761.

(8) Röhm, O. *German Patent* **1913**, vol. 283, 923.

(9) Olsen, H. S., Falholt, P. *Journal of Surfactants and Detergents* **1998**, vol. 1, no. 4, 555–567.

(10) *Industrial Enzymes*; Springer Netherlands: 2007.

(11) Nezhad, N. G., Rahman, R. N. Z. R. A., Normi, Y. M., Oslan, S. N., Shariff, F. M., Leow, T. C. *Applied Microbiology and Biotechnology* **2022**, vol. 106, no. 13-16, 4845–4866.

(12) Mavrovouniotis, M. L., Stephanopoulos, G., Stephanopoulos, G. *Chemical Engineering Communications* **1990**, vol. 93, no. 1, 211–236.

(13) Fischer, E. *Ber. Dtsch. Chem. Ges* **1894**, vol. 27, no. 3189, 1.

(14) Koshland, D. E. *Proceedings of the National Academy of Sciences* **1958**, vol. 44, no. 2, 98–104.

(15) Jencks, W. P. *Advances in enzymology and related areas of molecular biology* **1975**, vol. 43, 219–410.

(16) Warshel, A., Levitt, M. *Journal of Molecular Biology* **1976**, vol. 103, no. 2, 227–249.

(17) Warshel, A. *Proceedings of the National Academy of Sciences* **1978**, vol. 75, no. 11, 5250–5254.

(18) Feller, G., Gerday, C. *Nature reviews microbiology* **2003**, vol. 1, no. 3, 200–208.

(19) Mavromatis, K., Tsigos, I., Tzanodaskalaki, M., Kokkinidis, M., Bouriotis, V. *European Journal of Biochemistry* **2002**, vol. 269, no. 9, 2330–2335.

(20) Bjelic, S., Brandsdal, B. O., Åqvist, J. *Biochemistry* **2008**, vol. 47, no. 38, 10049–10057.

(21) Åqvist, J., Isaksen, G. V., Brandsdal, B. O. *Nature Reviews Chemistry* **2017**, vol. 1, no. 7.

(22) Socan, J., Isaksen, G. V., Brandsdal, B. O., Aqvist, J. *Sci Rep* **2019**, vol. 9, no. 1, 19147.

(23) Isaksen, G. V., Aqvist, J., Brandsdal, B. O. *Proc Natl Acad Sci U S A* **2016**, vol. 113, no. 28, 7822–7.

(24) Isaksen, G. V., Aqvist, J., Brandsdal, B. O. *PLoS Comput Biol* **2014**, vol. 10, no. 8, e1003813.

(25) Truhlar, D. G., Kohen, A. *Proceedings of the National Academy of Sciences* **2001**, vol. 98, no. 3, 848–851.

(26) Socan, J., Purg, M., Aqvist, J. *Nat Commun* **2020**, vol. 11, no. 1, 2644.

(27) Van der Kamp, M. W., Prentice, E. J., Kraakman, K. L., Connolly, M., Mulholland, A. J., Arcus, V. L. *Nature Communications* **2018**, vol. 9, no. 1.

(28) Bunzel, H. A., Kries, H., Marchetti, L., Zeymer, C., Mittl, P. R. E., Mulholland, A. J., Hilvert, D. *Journal of the American Chemical Society* **2019**, vol. 141, no. 30, 11745–11748.

(29) Nguyen, V., Wilson, C., Hoemberger, M., Stiller, J. B., Agafonov, R. V., Kutter, S., English, J., Theobald, D. L., Kern, D. *Science* **2017**, vol. 355, no. 6322, 289–294.

(30) Åqvist, J., van der Ent, F. *Journal of Chemical Theory and Computation* **2022**, vol. 18, no. 10, 6345–6353.

(31) Holzgräfe, C., Wallin, S. *Physical Biology* **2015**, vol. 12, 026002.

(32) Bryan, P. N., Orban, J. *Current Opinion in Structural Biology* **2010**, vol. 20, 482–488.

(33) Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., Caughey, W. S. *Biochemistry* **1993**, vol. 32, no. 1, 183–190.

(34) Saotome, T., Nakamura, S., Islam, M. M., Nakazawa, A., Dellarole, M., Arisaka, F., Kidokoro, S.-i., Kuroda, Y. *Biochemistry* **2016**, vol. 55, no. 32, 4469–4475.

(35) Islam, M. M., Kobayashi, K., Kidokoro, S.-I., Kuroda, Y. *The FEBS Journal* **2019**, vol. 286, no. 20, 4122–4134.

(36) Schmid, F. X., Perl, D., Mueller, U., Heinemann, U. *Nature Structural Biology* **2000**, vol. 7, no. 5, 380–383.

(37) Hendsch, Z. S., Tidor, B. *Protein Science* **1994**, vol. 3, no. 2, 211–226.

(38) Kumar, S., Tsai, C.-J., Ma, B., Nussinov, R. *Journal of Biomolecular Structure and Dynamics* **2000**, vol. 17, no. sup1, 79–85.

(39) Szilágyi, A., Závodszky, P. *Structure* **2000**, vol. 8, no. 5, 493–504.

(40) Olson, C. A., Spek, E. J., Shi, Z., Vologodskii, A., Kallenbach, N. R. *Proteins: Structure, Function, and Genetics* **2001**, vol. 44, no. 2, 123–132.

(41) Amelunxen, R. E., Murdock, A. L., Welker, N. E. *CRC Critical Reviews in Microbiology* **1978**, vol. 6, no. 4, 343–393.

(42) Clardy, J., Lee, A. Atomic structure of the buried catalytic pocket of Escherichia coli chorismate mutase, 1995.

(43) Chook, Y. M., Ket, H., Lipscomb, W. N. *Crystal structures of the monofunctional chorismate mutase from BaciUus subtilis and its complex with a transition state analog*; 1993, pp 8600–8603.

(44) Chook, Y., Ke, H., Lipscomb, W. The Mmonofunctional Chorismate Mutase from Bacillus subtilius: structure determination of chorismate mutase and its complexes with a transition state analog and prephenate, and implications on the mechanism of enzymatic reaction, 1994.

(45) Lee, A., Stewart, J., Clardy, J., Ganem, B. *New insight into the catalytic mechanism of chorismate mutases from structural studies*; 1995.

(46) Lyne, P. D., Mulholland, A. J., Richards, W. G. *Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction*; 1995, p 50.

(47) Zhang, X., Zhang, X., Bruice, T. C. *Biochemistry* **2005**, vol. 44, 10443–10448.

(48) Schroeder, D. V., *An Introduction to Thermal Physics*; Addison Wesley Longman: 2000.

(49) Pathria, R., Beale, P., *Statistical Mechanics (3e.)* Butterworth-Heineman: 2011.

(50) Leach, A., *Molecular Modelling: Principles and Applications*; Prentice Hall: 2001.

(51) McQuarrie, D. A., *Statistical Mechanics*; Harper Row: 1973.

(52) *An Introduction to Statistical Thermodynamics*; Hill, T. L., Ed.; Addison-Wesley: 1960.

(53) Shell, M. S., *Thermodynamics and Statistical Mechanics: An Integrated Approach*; Cambridge University Press: 2015.

(54) Frenkel, D., Smit, B., *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: 2002.

(55) Keskin, O., Jernigan, R., Bahar, I. *Biophysical Journal* **2000**, vol. 78, no. 4, 2093–2106.

(56) Papaleo, E., Riccardi, L., Villa, C., Fantucci, P., Gioia, L. D. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2006**, vol. 1764, no. 8, 1397–1406.

(57) Maguid, S., Fernández-Alberti, S., Parisi, G., Echave, J. *Journal of Molecular Evolution* **2006**, vol. 63, no. 4, 448–457.

(58) Fuglebakk, E., Echave, J., Reuter, N. *Bioinformatics* **2012**, vol. 28, no. 19, 2431–2440.

(59) Tiwari, S. P., Fuglebakk, E., Hollup, S. M., Skjærven, L., Cragnolini, T., Grindhaug, S. H., Tekle, K. M., Reuter, N. *BMC Bioinformatics* **2014**, vol. 15, no. 1.

(60) Bauer, J. A., Pavlović, J., Bauerová-Hlinková, V. *Molecules* **2019**, vol. 24, no. 18, 3293.

(61) Hinsen, K., Petrescu, A.-J., Dellerue, S., Bellissent-Funel, M.-C., Kneller, G. R. *Chemical Physics* **2000**, vol. 261, no. 1-2, 25–37.

(62) Levitt, M., Sander, C., Stern, P. S. *International Journal of Quantum Chemistry* **1983**, vol. 24, no. S10, 181–199.

(63) Krebs, W. G., Alexandrov, V., Wilson, C. A., Echols, N., Yu, H., Gerstein, M. *Proteins: Structure, Function, and Bioinformatics* **2002**, vol. 48, no. 4, 682–695.

(64) Petrone, P., Pande, V. S. *Biophysical Journal* **2006**, vol. 90, no. 5, 1583–1593.

(65) Kazemi, M., Aqvist, J. *Nat Commun* **2015**, vol. 6, 7293.

(66) Van der Ent, F., Lund, B. A., Svalberg, L., Purg, M., Chukwu, G., Widersten, M., Isaksen, G. V., Brandsdal, B. O., Åqvist, J. *Biochemistry* **2022**, vol. 61, no. 10, 933–942.

(67) Madani, A., Krause, B., Greene, E. R., Subramanian, S., Mohr, B. P., Holton, J. M., Olmos, J. L., Xiong, C., Sun, Z. Z., Socher, R., Fraser, J. S., Naik, N. *Nature Biotechnology* **2023**.

Manuscripts

Paper I

Accurate Computation of Thermodynamic Activation Parameters in the Chorismate Mutase Reaction from Empirical Valence Bond Simulations

Ryan Scott Wilkins, Bjarte Aarmo Lund, Geir Villy Isaksen, Johan Åqvist, Bjørn Olav Brandsdal

Published in *Journal of Chemical Theory and Computation*, December 2023, volume X, issue X, pp. X. DOI: 10.1021/acs.jctc.3c01105.

JCTC Journal of Chemical Theory and Computation

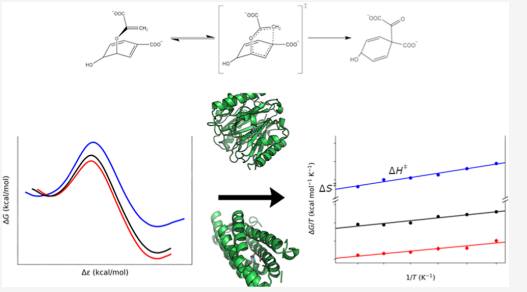
pubs.acs.org/JCTC Article

Accurate Computation of Thermodynamic Activation Parameters in the Chorismate Mutase Reaction from Empirical Valence Bond Simulations

Ryan Scott Wilkins, Bjarte Aarmo Lund, Geir Villy Isaksen, Johan Åqvist, and Bjørn Olav Brandsdal*

Cite This: https://doi.org/10.1021/acs.jctc.3c01105 Read Online

ABSTRACT: Chorismate mutase (CM) enzymes have long served as model systems for benchmarking new methods and tools in computational chemistry. Despite the enzymes' prominence in the literature, the extent of the roles that activation enthalpy and entropy play in catalyzing the conversion of chorismate to prephenate is still subject to debate. Knowledge of these parameters is a key piece in fully understanding the mechanism of chorismate mutases. Within this study, we utilize EVB/MD free energy perturbation calculations at a range of temperatures, allowing us to extract activation enthalpies and entropies from an Arrhenius plot of activation free energies of the reaction catalyzed by a monofunctional *Bacillus subtilis* CM and the promiscuous enzyme isochorismate pyruvate lyase of *Pseudomonas aeruginosa*. In comparison to the uncatalyzed reaction, our results show that both enzyme-catalyzed reactions exhibit a substantial reduction in activation enthalpy, while the effect on activation entropy is relatively minor, demonstrating that enzyme-catalyzed CM reactions are enthalpically driven. Furthermore, we observe that the monofunctional CM from *B. subtilis* more efficiently catalyzes this reaction than its promiscuous counterpart. This is supported by a structural analysis of the reaction pathway at the transition state, from which we identified key residues explaining the enthalpically driven nature of the reactions and also the difference in efficiencies between the two enzymes.



1. INTRODUCTION

A crucial component in understanding the function of enzymes is knowledge of the activation enthalpy−entropy balance and its role in catalyzing the reaction. Current studies of the significance of entropy effects in enzyme catalysis sometimes challenge the celebrated Circe effect hypothesis that ground-state destabilization of the substrate is responsible for the catalytic effect.[1] Hence, Kazemi et al. proposed that the traditional view of activation entropies being interpreted solely in terms of a loss of translational and rotational entropies upon substrate binding is oversimplified and that a complete representation should also include contributions from the reorganization of the surrounding protein and solvent.[2] Further understanding of the role of entropy in enzyme catalysis plays an important part in the optimization of catalytic rates and enzyme stability in the rational design of enzymes for industrial and scientific purposes.

A common model system used to study the entropy−enthalpy balance in enzymes is chorismate mutase (CM), which has served as a benchmark for QM/MM calculations of enzymatic reactions.[3−6] CM catalyzes the pericyclic rearrangement of chorismate to prephenate (Figure 1), which is a key step in the shikimate pathway for the biosynthesis of phenylalanine and tyrosine in bacteria, fungi, and higher plants.[7] There is some debate over the degree to which the entropy plays a role in this reaction. In the AroH-class CM from *Streptomyces aureofaciens*, the enzyme catalyzed reaction exhibits a significantly less negative entropy than the uncatalyzed reaction. This suggests that the conversion of chorismate to prephenate is driven by entropic effects.[8] This is hypothesized to be due to an electrostatic stabilization of the transition state as the average C_9-C_1 distance in the transition state decreases, an energetically costly effect in the water reaction.[9] While Hur and Bruice considered this to be an example of the formation of a so-called near-attack-complex (NAC), Warshel and co-workers argued that this conformation is the result of transition state stabilization rather than the reason for the catalytic effect of the enzyme.[3] On the other hand, in the CM reaction catalyzed by the *Bacillus subtilis* CM

Received: October 6, 2023
Revised: December 1, 2023
Accepted: December 4, 2023

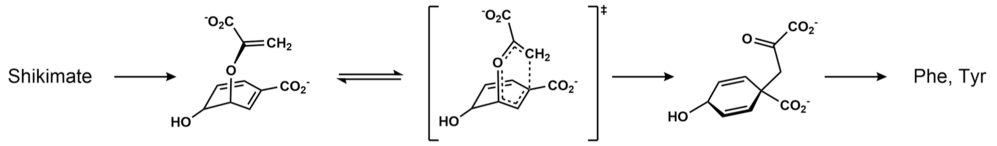


Figure 1. Claisen rearrangement of chorismate to prephenate via a chairlike transition state in CM enzymes.

(BsCM) and the promiscuous enzyme isochorismate pyruvate lyase from *Pseudomonas aeruginosa* (PchB), the activation entropy from experimental studies is comparable to that of the reaction in water, suggesting that entropy plays only a small role in the reaction.[10] In the latter enzyme, the conversion of chorismate to prephenate is secondary to the conversion of isochorismate to pyruvate and salicylate. The two reactions share an active site; however, the former reaction is approximately 2 orders of magnitude less efficient than the latter.[11] In the first reported QM/MM study on the entropic contributions of PchB, Xie et al. used DFTB2/MIO/MM to calculate thermodynamic activation parameters for the conversion of chorismate to pyruvate in PchB, finding a significantly less negative activation entropy than that observed in the uncatalyzed reaction.[12] By including a correctional term obtained from quasiharmonic analysis, they could match experimental entropy values. However, the activation free energies of the enzymatic and aqueous reactions reported by Xie et al. are also significantly lower than the experimental values, indicating that their calculated entropy from the Arrhenius plot approach is likely incorrect.

In comparison to other pyruvate lyases in the shikimate pathway, PchB is more similar to CMs of the AroQ class such as EcCM from *Escherichia coli*, both structurally and functionally.[10,13,14] The apparent origin of PchB is a gene duplication of an AroQ CM requiring a few minor mutations in the active site for efficient isopyruvate lyase activity, influencing a reduction in CM activity.[11] Mutation of the positively charged Arg90 to the non-natural neutral amino acid citrulline in BsCM decreased the catalytic rate significantly, with only a modest decrease in K_M, suggesting that electrostatic stabilization plays a key role in the CM catalytic activity.[15,16] In EcCM, Val35Ile and Val35Ala mutations result in an increase in k_{cat} of about 1.5 times and decrease in k_{cat} by about 2 times, respectively, reflecting an Ala35Ile mutation, which in EcCM introduces an overall increase in k_{cat} by about 3 times.[17] The lower enzymatic efficiency of PchB compared with that of BsCM has been suggested to be due to a less restricted effect on the substrate in the former. Thus, an equivalent Ala38Ile mutation in the active site of PchB contributes to a significant increase in k_{cat}, as it results in a better electrostatic stabilization for the transition state.[5] This suggests that enzyme efficiency is increased by favoring active sites which contribute to substrate preorganization, thereby reducing the entropic penalty paid. In both AroQ- and AroH-class CMs, preorganization of the substrate into a compressed conformer appears to be a key component in making the activation entropy less negative and therefore decreasing the activation free energy in the enzyme-catalyzed reaction. Thus, a further understanding of the phase space of the substrate conformation is crucial in elucidating the role of enthalpy and entropy in the reaction.

Understanding the roles that enthalpy and entropy play in enzymatic reactions requires an intimate knowledge of the microscopic details underlying these reactions. Due to difficulties in obtaining atomistic levels of detail in enzymatic reactions, making use of computer simulations provides an optimal route to study the details of the activation enthalpy−entropy balance of such systems. In this study, we employ semiempirical EVB/MD free energy calculations to study the CM reaction in PchB and BsCM. EVB was chosen due to its ability to accurately calculate free energies while being less computationally intensive than ab initio QM/MM methods.[18−20] Furthermore, EVB can be combined with an Arrhenius plot to extract the thermodynamic activation enthalpies and entropies. In our work, simulations are performed at six evenly spaced temperatures ranging from 288 to 313 K, allowing for calculation of the thermodynamic activation parameters from an Arrhenius plot.[2]

2. METHODS

2.1. DFT Calculations. The reaction energies for the uncatalyzed transformation of chorismate to prephenate in water were calculated with DFT using Gaussian 16.[21] Structure optimization and frequency calculations at the reactant (chorismate), transition, and product (prephenate) states were performed with the B3LYP functional[22] and the 6-31G(d,p) basis set. Dispersion effects were included in all calculations using Grimme's B3LYP-D3 method.[23,24] Intrinsic reaction coordinate calculations were performed in both directions from the TS to verify that the correct minima are connected. Electronic energies were obtained from single-point calculations on the optimized geometries (RS, TS, and PS) at the B3LYP/6-311G+(2d,2p) level of theory. The final energies reported are corrected for zero-point energy (ZPE), dispersion, and solvation effects calculated with the CPCM[25] model using water as the solvent (eps = 80). The same procedure was also performed with the M06-2X[26] functional for comparison. In contrast to B3LYP, dispersion corrections are already included in the M06-2X functional; therefore, this was not added explicitly. Additional calculations were also performed with both functionals using the SMD[27] solvation model parameterized for water.

2.2. EVB Simulations. The crystal structures of PchB (PDB entry 3REM)[28] and BsCM (PDB entry 3ZO8)[15] were used as starting models for the EVB simulations. For PchB, the chorismate was manually built in to the active site by modifying the existing salicylate and pyruvate present in the active site of 3REM.[28] The chorismate configuration for BsCM was obtained by manually modifying the transition state inhibitor present in the Arg90Cit BsCM with PDB entry 3ZP4[15] and then superimposed to the wild-type BsCM (PDB entry 3ZO8).[15] Topologies for all simulations were prepared with Q[29] and Qgui[30] where parameters were assigned in accordance with the OPLS-AA/M force field.[31] The enzyme structures were solvated in a spherical droplet of TIP3P[32] water molecules with a simulation sphere radius of 36 Å, centered at the protein center of mass and encompassing the entire protein. A 25 Å radius was used for the chorismate to

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

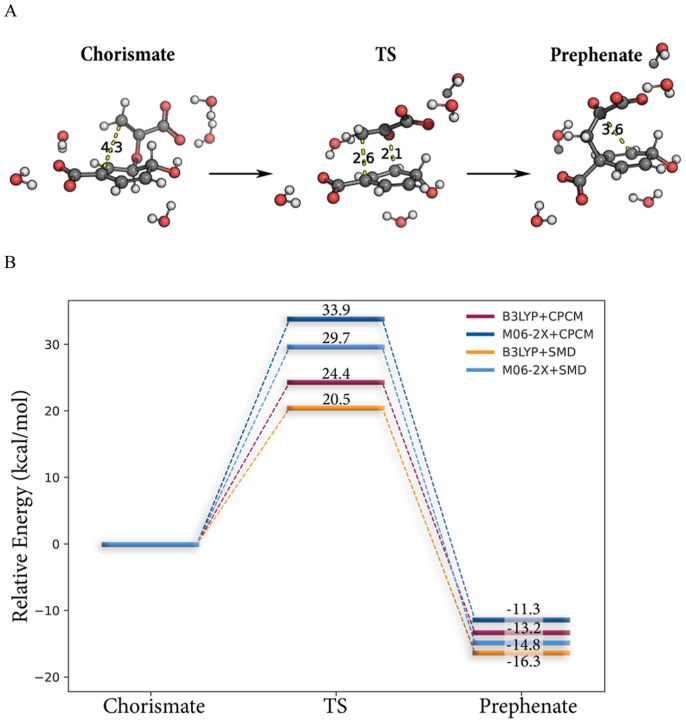


Figure 2. (A) DFT/B3LYP-optimized geometries for the conversion of the chorismate reactant state to the prephenate product state and (B) comparison of the calculated energetics with the B3LYP and M06-2X functionals including zero-point energy and solvation corrections obtained from the CPCM and SMD solvation models.

prephenate reference reaction in water. The systems were equilibrated prior to the MD/EVB simulations by gradually increasing the temperature from 1 to the final temperature (283−308 K) through a stepwise scheme (six steps). The first five steps were simulated for 10 ps, whereas the final equilibration step was simulated for 100 ps with a time step of 1 fs. The temperature was controlled by coupling the system to an external bath,[33] with a relaxation time of 10 fs for the first five steps and 100 fs for the final equilibration step and the productive phase. Long-range electrostatics were treated using the multipole expansion method (LRF)[34] and a direct 10 Å cutoff for nonbonded interactions. The reacting fragments were, however, allowed to interact with the entire system (no cutoff). Bonds and angles of solvent molecules were constrained with the SHAKE algorithm.[35] The productive simulations comprised 51 discrete free energy perturbation (FEP) steps, where each step was simulated for 100 ps, resulting in 5100 ps for each run.

The EVB free energy profiles were calculated using the FEP umbrella sampling approach described elsewhere.[36,37] These calculations utilized 51 discrete FEP windows, with a constant spacing of 0.02, between the two end-points chorismate (λ = 0) and prephenate (λ = 1). The EVB Hamiltonian was parameterized by fitting the relevant parameters ($\Delta\alpha$ and H_{12}) for the uncatalyzed transformation of chorismate to prephenate in water to reproduce the energetics obtained with DFT. That is, the activation free energy, $\Delta G^{\ddagger}$, was fitted to 24.5 kcal/mol, which is also the experimentally determined barrier,[38] whereas the reaction free energy, ΔG_0, was fitted to −12.8 kcal/mol, which is in excellent agreement with the previously estimated ΔG_0 from the observed reaction enthalpy, ΔH_0 = −13.3 kcal/mol, utilized by Warshel et al.[3] Thermodynamic activation parameters were obtained from EVB/Arrhenius plots, as explained elsewhere[36,37] based on simulations at six different temperatures evenly spaced from 288 to 313 K. A total of 100 independent simulations were carried out at each temperature point, resulting in 600 individual EVB simulations (306 ns) each for PchB, BsCM, and the uncatalyzed reaction (1800 simulations/918 ns in total).

3. RESULTS AND DISCUSSION

3.1. EVB and DFT Modeling of the Reference Reaction. Calibrating the EVB Hamiltonian requires that both activation and reaction free energies are available, from either experiments or quantum mechanical calculations. The activation free energy for the uncatalyzed transformation of chorismate to prephenate has previously been experimentally determined to be 24.5 kcal/mol,[38] but the reaction free energy, ΔG_0, is not available. Thus, to obtain a reference for ΔG_0, we carried out DFT calculations for the unimolecular conversion of chorismate to prephenate in water (Figure 2). To stabilize the charged carboxylic acid groups on the chorismate, five explicit water molecules were included. The B3LYP/6-311G +(2d,2p) model[39] with CPCM solvation[40,41] correction yields

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

a barrier of 24.4 kcal/mol, which is in excellent agreement with the corresponding experimental result.[38] Moreover, the reaction free energy was calculated to be −13.2 kcal/mol, which is very close to the experimentally observed reaction enthalpy, ΔH_0 = −13.3 kcal/mol, which was also previously used by Warshel et al. to estimate ΔG_0 for this reaction.[3]

Even though the CPCM model is often considered one of the most successful solvation models,[25,42] the more recent SMD model developed by Marenich et al.[27] is also popular, perhaps more commonly in combination with the Minnesota functionals developed by the same group. Thus, for comparison, we ran the same calculations with B3LYP and SMD, as well as M06-2X with CPCM and SMD (Figure 2B). As illustrated in Figure 2B, the resulting spread of the predicted activation energy is large. The difference between the B3LYP and M06-2X barriers is as much as 9.5 and 9.2 kcal/mol with the CPCM and SMD solvation correction, respectively. This difference is surprisingly large for the small unimolecular reaction under investigation here, demonstrating the importance of awareness of functional dependencies. However, the main aim of our DFT calculations was to obtain the reaction free energy to be used for calibrating the EVB Hamiltonian, and here, the variation is within the expected error of DFT calculations (Figure 2B). The difference between B3LYP and M06-2X reaction energies here is only 1.5 and 1.9 kcal/mol with the CPCM and SMD solvation corrections, respectively. The fact that the B3LYP calculations with CPCM reproduce both the experimentally known barrier and a reaction energy in agreement with the experimental reaction enthalpy suggests that −13.2 kcal/mol is a sufficiently reliable estimate for ΔG_0. Thus, this value was used further to calibrate the EVB reference reaction Hamiltonian. One should also note that the explicit water molecules used ensure a more proper microscopic solvation of the charges, and the orientation and the overall geometry are virtually identical when comparing the transition state obtained with B3LYP and M06-2X (Figure S1). The barrier calculated without these explicit water molecules is consistently 0.6−1.7 kcal/mol lower, depending on the functional and solvation correction, but the transition-state geometry is virtually identical compared to the calculations with explicit water (Figure S1).

While DFT methods offer a good combination of accuracy and computational cost, it should be noted that the variation observed in the activation free energy with different functionals and solvation models is a well-known problem with DFT calculations. It thus becomes difficult to trust which result should be used. A remedy to alleviate the functional dependency observed in the calculated activation free energies has been suggested, which is to use correlated ab initio electronic structure methods such as coupled cluster in multiscale modeling methods.[43] Here, Mulholland and co-workers showed that the more than 13 kcal/mol difference in the activation free energy was removed by using projector-based embedding.[43] However, for this study, the experimental barrier is known, and the thermodynamic activation parameters are not very sensitive to small changes in the reaction free energy.

EVB calculations of the thermodynamic activation parameters for the conversion of chorismate to prephenate were first performed on the uncatalyzed reference reaction in an aqueous environment. Reactions at 298 K were fit to the experimental value of the activation free energy (24.5 kcal/mol) and the reaction free energy obtained from DFT (−13.2 kcal/mol), yielding the intrinsic gas-phase energy difference between the EVB resonance structures, $\Delta\alpha$ = −94.7, and the off-diagonal coupling element, H_{12} = 86.8 kcal/mol. Using these parameters, EVB/MD calculations were performed at six other temperatures to obtain an Arrhenius plot of the solution-phase reaction. From this plot, an activation free energy of $\Delta G^{\ddagger}$ = 24.4 kcal mol^{-1} was obtained, and a linear regression was performed, yielding activation enthalpies and entropies of $\Delta H^{\ddagger}$ = 20.9 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −3.4 kcal mol^{-1} (at 298 K), respectively. These parameters are in excellent agreement with the values obtained from the experiment (Table 1).

Table 1. Thermodynamic Activation Parameters (in kcal/mol at 298 K) for the CM Reaction (Chorismate → Prephenate) in Water (Uncatalyzed) and Catalyzed by Isochorismate Pyruvate Lyase from *Pseudomonas aeruginosa* (PchB) and in CM from the *B. subtilis* (BsCM), Respectively[a]

	$\Delta G^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$
Uncatalyzed			
experimental[b]	24.5	20.7	−3.8
DFT (this study)	24.4		
DFTB2/MIO[c]	15.6 ± 0.2		−2.6
EVB (this study)	24.4 ± 0.6	20.9 ± 0.4	−3.4 ± 0.4
PchB			
experimental[d]	19.5	15.9 ± 0.2	−3.6 ± 0.2
DFTB2/MIO[c]	12.1 ± 0.1		−1.0
EVB (this study)	18.3 ± 0.7	14.0 ± 0.5	−4.3 ± 0.5
BsCM			
experimental[e]	15.4	12.7 ± 0.4	−2.7 ± 0.4
EVB (this study)	15.8 ± 12.1	12.9 ± 0.3	−2.9 ± 0.3

[a]Reported errors are given as standard error of the mean. [b]Andrews et al.[38] [c]Xie et al.[12] [d]Lamb et al.[10] [e]Kast et al.[46]

3.2. Activation Parameters of CM Enzyme Reactions.

Before constructing the Arrhenius plots for the CM enzymes, we first looked at the thermodynamic behavior at 298 K. The calculated average reaction free energy profile along the reaction coordinate for each enzyme (Figure 3A) yielded free energy barriers of $\Delta G^{\ddagger}$ = 15.8 kcal mol^{-1} for BsCM and $\Delta G^{\ddagger}$ = 18.3 kcal mol^{-1} for PchB, which show substantial catalysis compared to the uncatalyzed reference reaction activation barrier of $\Delta G^{\ddagger}$ = 24.4 kcal mol^{-1}. Activation free energies for both enzymatic reactions are also in good agreement with the experimental values for the respective enzymes.[10,46]

To obtain thermodynamic activation parameters from an Arrhenius plot (Figure 3B), EVB/MD calculations were performed at the same six temperatures used for the reference reaction, using the same $\Delta\alpha$ and H_{12} values obtained from fitting the reference reaction to activation and reaction free energies of the uncatalyzed reaction obtained from experiment[38] and DFT. From these simulations, we fitted $\Delta G^{\ddagger}/T$ vs $1/T$ using linear regression, which yielded values of $\Delta H^{\ddagger}$ = 12.9 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −2.9 kcal mol^{-1} for BsCM at 298 K. The same analysis for PchB yielded values of $\Delta H^{\ddagger}$ = 14.0 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −4.3 kcal mol^{-1} at 298 K. The values predicted using the Arrhenius plot approach for both enzymes are thus in a very reasonable agreement with the experimental values of $\Delta H^{\ddagger}$ = 12.7 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −2.7 kcal mol^{-1} for BsCM at 298 K, and $\Delta H^{\ddagger}$ = 15.9 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −3.6 kcal mol^{-1} for PchB at 298 K.

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

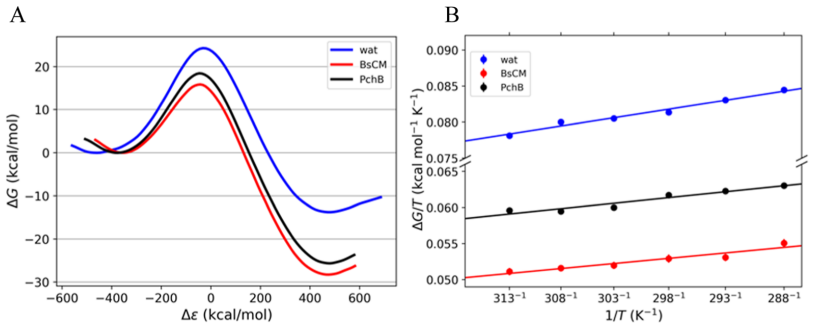


Figure 3. (A) Reaction free energy profiles obtained from EVB simulations at 298 K and (B) the corresponding Arrhenius plots in the temperature range of 288−313 K for the conversion of chorismate to prephenate in an aqueous environment (blue) and in the CM enzymes from *B. subtilis* (red) and PchB (black). Error bars are displayed but are smaller than the data point.

For the CM reaction catalyzed by PchB, an earlier computational study has been published by Xie et al.[12] They report activation parameters of $\Delta G^{\ddagger}$ = 12.1 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = −1.0 kcal mol^{-1} and claimed that since their calculated enzyme activation entropy penalty is 1.6 kcal/mol smaller than that of the uncatalyzed reaction, the enzyme-catalyzed reaction is entropically driven. However, their reported activation free energy is also significantly lower than experimental values.[10] In comparison to the enzyme-catalyzed reaction, we obtain activation parameters of $\Delta H^{\ddagger}$ = 20.9 and $T\Delta S^{\ddagger}$ = −3.4 kcal mol^{-1} K^{-1} for the reference reaction in water, in excellent agreement with experiment.[38] The enzyme-catalyzed reactions in BsCM and PchB thus show a significant reduction in the activation enthalpy, while reductions in the entropy penalty are much smaller. Hence, both from our calculations and experiment, it can be concluded that the enzyme-catalyzed reactions are enthalpy driven and not entropy driven as suggested by Xie et al.[12]

A better understanding of the difference in activation thermodynamics can be obtained from a breakdown of the energetics by analyzing the MD trajectories. Neglecting the small pressure−volume term, the activation enthalpy can be approximated by

$$\Delta H^{\ddagger}=\Delta U_{rr}^{\ddagger}+\Delta U_{rs}^{\ddagger}+\Delta U_{ss}^{\ddagger} \quad (1)$$

where the subscripts r and s denote the reacting molecule (chorismate/prephenate) and its surrounding environment, respectively. Here, the last $\Delta U_{ss}^{\ddagger}$ term involves numerically very large energies since it corresponds to the interactions within the entire protein/solvent system. However, the average $\Delta U_{rr}^{\ddagger}$ and $\Delta U_{rs}^{\ddagger}$ terms can be readily evaluated since they only involve interactions with and within the reacting molecule, and the difference is then taken between the transition and reactant states. It is also of interest here to further break down the contribution from the reacting molecule into its bonded (bonds, angles, torsions, and improper dihedrals) and nonbonded (electrostatic and van der Waals) terms

$$\Delta U_{tot,r}^{\ddagger}=\Delta U_{rr}^{\ddagger}+\Delta U_{rs}^{\ddagger}=\Delta U_{bonded}^{\ddagger}+\Delta U_{nonbond}^{\ddagger} \quad (2)$$

Such a breakdown (Table 2) clearly shows that the trend in $\Delta H^{\ddagger}$ obtained from the Arrhenius plots follows the contribution from $\Delta U_{tot,r}^{\ddagger}$. Hence, the value of $\Delta U_{tot,r}^{\ddagger}$ compared to the uncatalyzed reaction is about 10 kcal/mol lower for PchB and about 20 kcal/mol lower for BsCM. The

Table 2. Decomposition of Total Potential Activation Energy into Bonded and Nonbonded Contributions (in kcal/mol at 298 K) for the CM Reaction (Chorismate → Prephenate) in Water (Uncatalyzed) and Catalyzed by Isochorismate Pyruvate Lyase from *Pseudomonas aeruginosa* (PchB) and in CM from the *B. subtilis* (BsCM), Respectively

	$\Delta U_{tot,r}^{\ddagger}$	$\Delta U_{bonded}^{\ddagger}$	$\Delta U_{nonbond}^{\ddagger}$
uncatalyzed	111.0	110.0	1.0
PchB	101.8	96.0	5.8
BsCM	89.3	93.2	−3.9

magnitude of $\Delta U_{tot,r}^{\ddagger}$ is mainly determined by the several partially formed (single and double) bonds in the transition state, which is the reason why its main contribution comes from $\Delta U_{bonded}^{\ddagger}$. Here, one also finds that the two main contributions to the differences in $\Delta U_{bonded}^{\ddagger}$ compared to the water reaction (Table 2) originate from the Morse potential bond terms (∼10 kcal/mol) and the torsional energies (∼4 kcal/mol), while angle and improper terms have negligible contributions. Hence, both enzymes enforce a more compact transition state than in water, which results in more favorable values of $\Delta U_{bonded}^{\ddagger}$. As seen in Figure 4, the C_5-O_7 and C_1-C_9 distances are compressed in both the BsCM RS and the TS in comparison to the uncatalyzed reaction. Furthermore, the C_1 and C_9 atoms are better aligned in the BsCM RS due to differences in the $C_1-C_6-C_5-O_7$, $C_6-C_5-O_7-C_8$, and $C_5-O_7-C_8-C_9$ torsions in comparison to the uncatalyzed RS. The changes in these bond lengths and torsions are largely responsible for the more favorable $\Delta U_{bonded}^{\ddagger}$ seen in BsCM. Interestingly, one finds that the average value of $\Delta U_{nonbond}^{\ddagger}$ is actually less favorable in the promiscuous PchB enzyme than in both water and BsCM. This is likely due to poorer substrate preorganization as the CM reaction is not the primary function of this enzyme.

3.3. Structural Analysis of the Catalytic Effect of CMs.

The enthalpically driven nature of catalysis by the CM enzymes BsCM and PchB can be attributed to highly charged active sites, which serve to stabilize the transition state, as seen in Figure 5. It is notable that in comparison to PchB, the active site of BsCM has a greater degree of hydrogen and ionic bonding between the substrate and active site residues, which contribute to the lower enthalpy seen in BsCM. In particular, there are three key residues in each enzyme that contribute to

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

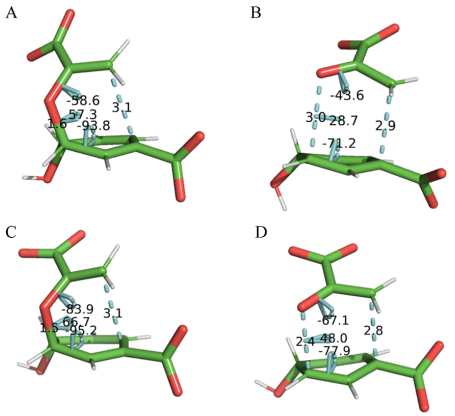


Figure 4. Geometries of chorismate displaying C_5-O_7 and C_1-C_9 distances and $C_1-C_6-C_5-O_7$, $C6-C_5-O_7-C_8$, and $C_5-O_7-C_8-C_9$ torsions for the uncatalyzed RS (A) and TS (B), as well as the BsCM RS (C) and TS (D).

this effect: Glu78, Arg90, and Tyr108 in BsCM, and Pro48, Lys41, and Arg110 in PchB. Additionally, the second carboxylate of the substrate is stabilized by Arg63 in BsCM and Arg30 in PchB.

Strong hydrogen bonding between Glu78 and C_4-OH has been proposed to cause an elongation of the C_5-O_7 bond, stabilizing the transition state in *B. subtilis*.[44] In *P. aeruginosa*, the backbone carbonyl group of Pro48 serves a similar function, as it acts as a hydrogen bond acceptor in forming a hydrogen bond with C_4-OH. However, the hydrogen bond formed in PchB is at a somewhat unfavorable angle for this interaction, thus suggesting why the enthalpic effect is greater in BsCM where the anionic oxygen is also stabilized by the amide group of Cys75. This is supported by mutagenesis studies showing that mutations of this glutamate to residues reducing this interaction result in a reduction of the catalytic rate.[45] Furthermore, the elongation of this C_5-O_7 ether bond by Glu78/Cys75 (in BsCM) or Pro48 (in PchB) results in a large negative charge on the ether oxygen. In BsCM, Arg90 provides a positive charge to counterbalance the large negative charge on this ether oxygen. Lys41 in PchB also provides a positive charge to stabilize the negative charge on this oxygen. To ensure the proper formation of the prephenate product, Arg7 and Tyr108 in BsCM form hydrogen bonds with the carboxylate group in the pyruvate moiety of the transition state. Similarly, Arg110 and Lys41 serve this function in PchB. The ionic interactions are essential in the CM reaction, as they orient the pyruvate fragment with the cyclohexadienyl moiety such that it facilitates the formation of the C_1-C_9 bond in prephenate. The differences in the active sites between the two enzymes can thus explain the decrease in the activation enthalpy seen in BsCM in comparison to that seen in PchB.

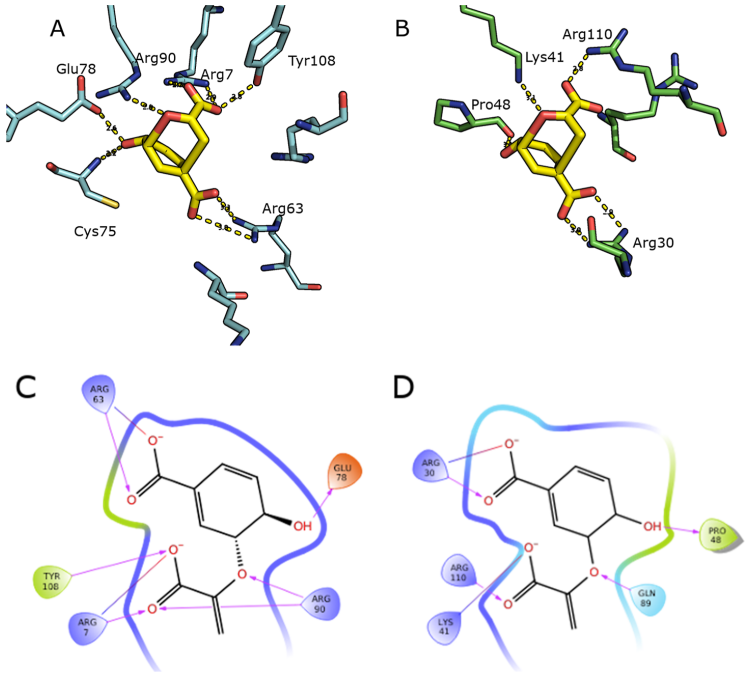


Figure 5. Active sites of CM enzymes from simulated MD trajectory snapshots of the CM reaction in BsCM and PchB showing residues of interest in catalysis. (A) Three-dimensional active site with a simulated pericyclic transition state in *B. subtilis*. (B) Three-dimensional active site with a simulated pericyclic transition state in PchB. (C) Two-dimensional ligand interaction diagram of BsCM. (D) Two-dimensional ligand interaction diagram of PchB.

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Enhanced hydrogen bonding in the active site of BsCM results in more favorable internal energetics within the reacting atoms and between the reacting atoms and their surroundings.

4. CONCLUDING REMARKS

EVB/MD simulations were used here to determine thermodynamic activation parameters for the conversion of chorismate to prephenate, in both the uncatalyzed reaction and the CM enzymes BsCM and PchB. The success of the EVB method is demonstrated by its ability to accurately match thermodynamic activation parameters obtained from experiments in all cases. These results clearly support the notion that the CM enzymes BsCM and PchB are enthalpy-driven reactions, as seen by the substantial reduction in activation enthalpies and smaller reductions in activation entropy in comparison to that observed in the uncatalyzed reaction.

Understanding the roles that entropy and enthalpy play in enzyme catalysis is crucial in understanding enzyme function as a whole. In particular, we are interested in the temperature dependence of chemical reaction rates and how enzymes have evolved to function efficiently at the freezing point of water. The present study provides knowledge of the activation parameters for BsCM and PchB that will serve as a starting point for studying the shift in the entropy−enthalpy balance for cold-adapted CM enzymes.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.3c01105.

Optimized transition state geometries and their coordinates (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Bjørn Olav Brandsdal − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway;* orcid.org/0000-0002-4681-8081; Email: bjorn-olav.brandsdal@uit.no

Authors

Ryan Scott Wilkins − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway*

Bjarte Aarmo Lund − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway;* orcid.org/0000-0001-9141-0555

Geir Villy Isaksen − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway;* orcid.org/0000-0001-7828-7652

Johan Åqvist − *Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway;* Present Address: Department of Cell and Molecular Biology, Uppsala University, Box 596, 75124 Uppsala, Sweden; orcid.org/0000-0003-2091-0610

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c01105

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support from the Norwegian Research Council through a Centre of Excellence and project grant (grant nos. 262695 and 274858), the Swedish Research Council (VR), and the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

■ REFERENCES

(1) Jencks, W. P. Binding Energy, Specificity, and Enzymic Catalysis: The Circe Effect. *Advances in Enzymology and Related Areas of Molecular Biology*; Wiley Online Library, 1975; Vol. *43*, pp 219−410.

(2) Kazemi, M.; Himo, F.; Åqvist, J. Enzyme catalysis by entropy without Circe effect. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 2406−2411.

(3) Strajbl, M.; Shurki, A.; Kato, M.; Warshel, A. Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *J. Am. Chem. Soc.* **2003**, *125*, 10228−10237.

(4) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction. *J. Am. Chem. Soc.* **1995**, *117*, 11345−11350.

(5) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J. Predicting an Improvement of Secondary Catalytic Activity of Promiscuos Isochorismate Pyruvate Lyase by Computational Design. *J. Am. Chem. Soc.* **2008**, *130*, 2894−2895.

(6) Marti, S.; Andres, J.; Moliner, V.; Silla, E.; Tunon, I.; Bertran, J.; Field, M. J. A hybrid potential reaction path and free energy study of the chorismate mutase reaction. *J. Am. Chem. Soc.* **2001**, *123*, 1709−1712.

(7) Gibson, F.; Pittard, J. Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microorganisms. *Bacteriol. Rev.* **1968**, *32*, 465−492.

(8) Galopin, C. C.; Zhang, S.; Wilson, D. B.; Ganem, B. On the mechanism of chorismate mutases: Clues from wild-type E. coli enzyme and a site-directed mutant related to yeast chorismate mutase. *Tetrahedron Lett.* **1996**, *37*, 8675−8678.

(9) Hur, S.; Bruice, T. C. Enzymes do what is expected (chalcone isomerase versus chorismate mutase). *J. Am. Chem. Soc.* **2003**, *125*, 1472−1473.

(10) Luo, Q.; Meneely, K. M.; Lamb, A. L. Entropic and enthalpic components of catalysis in the mutase and lyase activities of Pseudomonas aeruginosa PchB. *J. Am. Chem. Soc.* **2011**, *133*, 7229−7233.

(11) Gaille, C.; Kast, P.; Haas, D. Salicylate Biosynthesis in Pseudomonas aeruginosa. *J. Biol. Chem.* **2002**, *277*, 21768−21775.

(12) Xie, L.; Yang, M.; Chen, Z.-N. Understanding the entropic effect in chorismate mutase reaction catalyzed by isochorismate-pyruvate lyase fromPseudomonas aeruginosa(PchB). *Catal. Sci. Technol.* **2019**, *9*, 957−965.

(13) Gallagher, D. T.; Mayhew, M.; Holden, M. J.; Howard, A.; Kim, K. J.; Vilker, V. L. The crystal structure of chorismate lyase shows a new fold and a tightly retained product. *Proteins* **2001**, *44*, 304−311.

(14) Zaitseva, J.; Lu, J.; Olechoski, K. L.; Lamb, A. L. Two crystal structures of the isochorismate pyruvate lyase from Pseudomonas aeruginosa. *J. Biol. Chem.* **2006**, *281*, 33441−33449.

(15) Burschowsky, D.; van Eerde, A.; Okvist, M.; Kienhofer, A.; Kast, P.; Hilvert, D.; Krengel, U. Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 17516−17521.

(16) Kienhofer, A.; Kast, P.; Hilvert, D. Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor. *J. Am. Chem. Soc.* **2003**, *125*, 3206−3207.
(17) Lassila, J. K.; Keeffe, J. R.; Kast, P.; Mayo, S. L. Exhaustive mutagenesis of six secondary active-site residues in Escherichia coli chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis. *Biochemistry* **2007**, *46*, 6883−6891.
(18) Kamerlin, S. C.; Warshel, A. The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discuss.* **2010**, *145*, 71−106.
(19) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(20) Åqvist, J.; Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* **1993**, *93*, 2523−2544.
(21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, D. F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*: Wallingford, CT, 2016.
(22) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(23) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(24) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(25) Takano, Y.; Houk, K. N. Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 70−77.
(26) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(27) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(28) Olucha, J.; Ouellette, A. N.; Luo, Q.; Lamb, A. L. pH Dependence of catalysis by Pseudomonas aeruginosa isochorismate-pyruvate lyase: implications for transition state stabilization and the role of lysine 42. *Biochemistry* **2011**, *50*, 7198−7207.
(29) Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q. Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graphics Modell.* **1998**, *16*, 213−225.
(30) Isaksen, G. V.; Andberg, T. A.; Åqvist, J.; Brandsdal, B. O. Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems. *J. Mol. Graphics Modell.* **2015**, *60*, 15−23.
(31) Robertson, M. J.; Qian, Y.; Robinson, M. C.; Tirado-Rives, J.; Jorgensen, W. L. Development and Testing of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.* **2019**, *15*, 2734−2742.
(32) 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.
(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(34) Lee, F. S.; Warshel, A. A Local Reaction Field Method for Fast Evaluation of Long-Range Electrostatic Interactions in Molecular Simulations. *J. Chem. Phys.* **1992**, *97*, 3100−3107.
(35) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(36) Åqvist, J.; Isaksen, G. V.; Brandsdal, B. O. Computation of enzyme cold adaptation. *Nat. Rev. Chem* **2017**, *1*, 0051.
(37) Isaksen, G. V.; Åqvist, J.; Brandsdal, B. O. Enzyme surface rigidity tunes the temperature dependence of catalytic rates. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 7822−7827.
(38) Andrews, P. R.; Smith, G. D.; Young, I. G. Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate. *Biochemistry* **1973**, *12*, 3492−3498.
(39) Kazemi, M.; Sheng, X.; Himo, F. Origins of Enantiopreference of Mycobacterium smegmatis Acyl Transferase: A Computational Analysis. *Chemistry* **2019**, *25*, 11945−11954.
(40) Andzelm, J.; Kölmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995**, *103*, 9312−9320.
(41) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(42) Rayne, S.; Forest, K. Accuracy of computational solvation free energies for neutral and ionic compounds: Dependence on level of theory and solvent model. *Nat. Prec.* **2010**, DOI: 10.1038/npre.2010.4864.1.
(43) Ranaghan, K. E.; Shchepanovska, D.; Bennie, S. J.; Lawan, N.; Macrae, S. J.; Zurek, J.; Manby, F. R.; Mulholland, A. J. Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution. *J. Chem. Inf. Model.* **2019**, *59*, 2063−2078.
(44) Worthington, S. E.; Roitberg, A. E.; Krauss, M. An MD/QM Study of the Chorismate Mutase-Catalyzed Claisen Rearrangement Reaction. *J. Phys. Chem. B* **2001**, *105*, 7087−7095.
(45) Cload, S. T.; Liu, D. R.; Pastor, R. M.; Schultz, P. G. Mutagenesis Study of Active Site Residues in Chorismate Mutase from Bacillus subtilis. *J. Am. Chem. Soc.* **1996**, *118*, 1787−1788.
(46) Kast, P.; Asif-Ullah, M.; Hilvert, D. Is chorismate mutase a prototypic entropy trap? - Activation parameters for the Bacillus subtilis enzyme. *Tetrahedron Lett.* **1996**, *37*, 2691−2694.

https://doi.org/10.1021/acs.jctc.3c01105
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Paper II

Biophysical characterization and analysis of a mesophilic chorismate mutase from B. pumilus

Ryan Scott Wilkins, Bjarte Aarmo Lund, Geir Villy Isaksen, Johan Åqvist, Bjørn Olav Brandsdal

Unpublished, to be submitted soon. Pre-print available at DOI: https://doi.org/10.1101/2023.04.20.537678.

II

Biophysical characterization and analysis of a mesophilic chorismate mutase from *B. pumilus*

Ryan Scott Wilkins[a], *Bjarte Aarmo Lund* [a], *Geir Villy Isaksen*[a,1], *Johan Åqvist*[a,b] *and Bjørn Olav Brandsdal*[a,1]

[a]Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway

[b]Department of Cell and Molecular Biology, Uppsala University, Box 596, 75124 Uppsala, Sweden

Abstract

Chorismate mutases have extensively been used as computational benchmarking systems for enzyme catalysis, yet the roles entropy and enthalpy play in catalysis are still not fully understood. Thus, it is important to better understand these enzymes for potential research or industrial applications. Here, we report the first crystal structure and kinetic characterization of a chorismate mutase from *Bacillus pumilus*. This enzyme exhibits a high degree of similarity to a known mesophilic chorismate mutase from *Bacillus subtilis*. Using this crystal structure, we further employ EVB/MD simulations to construct Arrhenius plots, allowing us to extract thermodynamic activation parameters. Overall, this study provides new insights into the structural and functional features of the *B. pumilus* chorismate mutase and highlights its potential as a valuable enzyme for biocatalytic and biotechnological applications.

Introduction

In order to better understand the mechanisms behind enzyme catalysis, it is important to know the role the activation enthalpy-entropy balance plays in the reaction. The current hypothesis regarding entropic contributions to enzyme activity proposed by Åqvist *et al.* is that the once-

popular Circe effect is an oversimplification[1] – activation entropies should not solely include contributions from a loss of translational and rotational entropies upon substrate binding, but also need to account for entropic contributions from the reorganization of the surrounding enzyme and solvent. A common model system used in understanding the balance of these activation parameters in computational chemistry is the chorismate mutase enzyme.[2-5] Chorismate mutase enzymes are found in bacteria, fungi, and higher plants, where they play a role in the Shikimate pathway. The enzyme is responsible for the catalysis of the pericyclic rearrangement of chorismate to prephenate, which proceeds through a chair-like transition state. This serves as a precursor for the biosynthesis of benzyl-containing amino acids phenylalanine and tyrosine.[6] Further understanding of the roles activation enthalpy and entropy play in enzyme catalysis serves as groundwork for fine-tuning catalytic rates and enzyme stability for the rational design of enzymes to be used in industrial and scientific contexts.

Chorismate mutase enzymes are a common model system to study entropic and enthalpic contributions to enzymatic reactions and are prevalent as a benchmark system for QM/MM calculations. Despite this, there is still debate over how important entropic effects are in the chorismate mutase reaction. Early studies by Gorisch indicate that the chorismate mutase reaction in *S. aurofaciens* is primarily entropy driven. [7] However later studies on *E. coli* and *B. subtilis* show a significant entropic contribution to the free energy but they are still enthalpically driven reactions. Hur and Bruice originally consider an electrostatic stabilization of the transition state as the C_1-C_9 distance decreases the formation of a near-attack conformer (NAC), contributing to the catalytic effect of the enzyme. [8] On the other hand, Warshel *et al.* later argue that this confirmation is likely a result of electrostatic stabilization, rather than the formation of a NAC. [2]

Preorganization of the substrate into a compressed conformer appears to be a key component underlying the enthalpy-entropy balance in chorismate mutase enzymes, thus understanding the mechanisms leading to this conformer plays a vital role in understanding the role enthalpy and entropy play in the reaction. There has been some work done in identifying key active site residues contributing to the formation of this

compressed conformer. In BsCM, substitution of the positively charged Arg90 for a neutral Citrulline results in a significant reduction of k_{cat}, while only moderately affecting K_m, supporting electrostatic stabilization as a key mechanism in the catalytic reaction. [9, 10] In the promiscuous enzyme PchB, a lower k_{cat} in comparison to monofunctional BsCM has been suggested to be the result of a less restrictive active site. This is supported by an Ala38Ile mutation in *E. coli* CM, [11] increasing electrostatic stabilization and resulting in a significant increase in k_{cat}. Additionally, Val35Ile and Val35Ala mutations in EcCM result in an increase in k_{cat} of 1.5 times and decrease in k_{cat} by about 2 times, respectively. This suggests that chorismate mutase activity can be increased by favoring more restrictive active sites, contributing to higher substrate preorganization and reducing the entropic penalty paid in the reaction. [4]

In this study, we present a new crystal structure for *B. pumilus* chorismate mutase (uniprot A8FEK3, BpCM) obtained from X-ray diffraction crystallography along with thermodynamic activation parameters obtained from constructing an Arrhenius plot from UV-Vis spectroscopy enzyme kinetics experiments. As a full understanding of the roles entropy and enthalpy play in BpCM requires intimate knowledge of atomistic-level details, we further employ computer calculations to elucidate the microscopic details of the enzyme-catalyzed reaction. For this, we construct Arrhenius plots using the empirical valence bond method (EVB) in tandem with molecular dynamics (MD) simulations to obtain free energy calculations of this reaction. Compared to *ab initio* QM/MM methods, EVB maintains a high level of accuracy while remaining less computationally intensive. [12-14] Combining this with molecular dynamics allows us to obtain trajectories of the enzyme and substrate during the reaction. To construct the Arrhenius plot from simulation data, we performed simulations at 7 evenly spaced temperatures ranging from 283-313K on a microsecond-timescale.

Methods

Protein production and expression

The gene encoding the *Bacillus pumilus* AroH-type Chorismate mutase (uniprot A8FEK3, BpCM) with a dual C-terminal stop codon (TAATAA) was codon optimized for expression in *E. coli*, synthesized and subcloned into pET22b by

GenScript Biotech (Netherlands) B.V. The plasmid was transformed into *E. coli* BL21 (DE3) Star (Invitrogen) using standard heat shock protocol. Protein was produced in shake flasks using ZYP-5052 auto induction media with 50 µg/mL Ampicillin. Cells were grown at 37°C for 4-5 hours, before the temperature was reduced to 20 °C for overnight expression. Cells were harvested by centrifugation, 6000 rpm for 30 minutes. Cells were disrupted by sonication, for 15 minutes active at 40 % amplitude, 3/6 s pulses on/off, 12 °C maximum temperature. The cell lysate was clarified by centrifugation, 20000 rpm for 45 minutes. The clarified cell lysate was loaded on a HisTrap FF 1 mL column equilibrated with 50 mM HEPES pH 7.5 with 500 mM NaCl. Contaminating proteins were eluted by a stepwise elution of 5 and 10 % buffer B containing 500 mM Imidazole and 500 mM NaCl pH 7.5. The protein was eluted by 75 % buffer B. For polishing the concentrated fractions were loaded on a Superdex 75 26/600 column equilibrated with 50 mM potassium phosphate pH 7.5. The main peak was concentrated to 1 mM concentration (as determined by UV280 measurements with the calculated extinction coefficient) and flash-frozen in liquid nitrogen for storage at -20°C. SEC-MALS was performed with a Shimadzu UFLC system with a Wyatt dawn 8+ detector using Shodex HPLC lb804/lb805 columns with 0.1M $NaNO_3$ with 20 mg/L NaN_3 as running buffer at 50 °C using a protein concentration of 0.65 mg/mL.

Crystallization

From a set of four commercial crystallization screens from Molecular Dimensions multiple hits were identified for BpCM when screened at 1 mM. For optimization simple gradients of PEG-concentrations were used and diffraction-quality crystals were obtained with 25 % PEG 1500 and 0.1 M MES pH 6.5. Diffraction data was collected using the MX14.2 beamline at the BESSY II electron storage ring at the Helmholtz-Zentrum Berlin für Materialien und Energie Data was processed in XDS.
Molecular replacement was performed with Phaser [15] using the trimer of *Bacillus subtilis* chorismate mutase 3ZO8 [9] as the search model. The solution was improved with Autobuild [16] before several cycles of semi-manual rebuilding in coot [17] and automatic refinement in phenix.refine. [18]

Chorismic acid assay

A 10.5 mM stock solution was used to make dilutions from 500 µM to 16 µM. The concentration of BpCM was kept at 10 nM for the experiments. The concentration of chorismic acid was followed at 274 nm, and concentrations were calculated using an extinction coefficient of 2630 $mol^{-1}cm^{-1}$. Measurements were done with 1 cm UV-quartz cuvettes in a qChanger6 Peltier-cooled thermostat and sample-changer with a Cary60 UV/VIS spectrophotometer. The velocities from triplicated measurements were fitted against the substrate concentrations with non-linear regression using GraphPad PRISM 6.07 for Windows (GraphPad Software, La Jolla California USA).

Differential scanning calorimetry

Using 50 mM potassium phosphate pH 7.5 as the buffer and reference the temperature range of 20-80°C was scanned at a rate of 1°C/min.

EVB Simulations

The X-ray crystal structure obtained in this study was used as the starting structure for our EVB simulations. The chorismate substrate configuration was obtained by manually superimposing the BpCM crystal structure with the structure of Arg90Cit BsCM with PDB entry 3ZP4. [9] Topologies for the system were generated using the EVB/MM program *Q*[19] interfaced with *Qgui*, [20] using the OPLS-AA/M force field to parameterize the enzyme. [21] A 38 Å spherical boundary centered on the protein center-of-mass was defined around the enzyme, which was filled with TIP3P[22] water molecules to solvate the system. Before the production runs of EVB/MD simulations, the systems were equilibrated by gradually increasing the temperature in 6 steps from 1 K up until the final system temperature. Equilibration was performed in 1 fs timesteps, with the first 5 equilibrations steps being ran for 10 ps, and the final step being performed for 100 ps. System temperature was controlled via coupling to an external heat bath[23] with a relaxation time of 10 fs for the first 5 equilibration steps and 100 fs for the final equilibration step and the production runs. A multipole expansion method (LRF)[24] was used to calculate long-range electrostatics, and nonbonded interactions were calculated using a 10 Å cut-off. However, reacting fragments in the catalytic center were free to interact with the entire system. The SHAKE[25] algorithm was

used to constrain bonds and angles of solvent molecules.

The FEP umbrella sampling[26, 27] approach was used to calculate EVB free energy profiles. Calculations were divided into 51 discrete FEP windows with a spacing of 0.02 between chorismate ($\lambda = 0$) and prephenate ($\lambda = 1$). The EVB Hamiltonian was parameterized using the uncatalyzed conversion of chorismate to prephenate in our previous work. To obtain thermodynamic activation parameters for the enzyme, the EVB/Arrhenius[29, 30] approach was carried out at 7 evenly spaced temperatures from 283-313 K. At each temperature, 100 independent replicas were carried out for a total of 357 ns of simulated trajectories.

Molecular Dynamics Simulations

The initial solvated structure for molecular dynamics (MD) simulations was the same structure used in the setup of EVB simulations. Equilibration, electrostatics, non-bonded interactions, and force fields were also identical to those used in the EVB simulations. However, the active site was treated entirely with molecular mechanics. For production runs, 20 independent runs at 3.5 ns of simulation time per run were performed, for a total of 70 ns of simulation time.

Results and Discussion

Protein characterization

BpCM was expressed with reasonable yields of 5-10 mg per liter of culture and appeared to be stable. From differential scanning calorimetry a single transition was observed with a T_m of 65.5°C compared to 58.3°C for BsCM. No refolding was observed after cooling. SEC-MALS confirmed that BpCM is a trimer in solution.

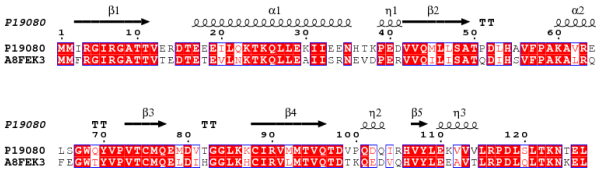


Figure 1: Sequence alignment of BsCM (P19080) and BpCM (A8FEK3) shows an identity of 69%. Furthermore, the similarity between the sequences is 93%.

Protein structure

The overall structure of BpCM is nearly identical to BsCM, with a RMSD value of 0.28 Å over 717 atoms in the monomers. The RMSD value is surprisingly low as the sequence identity is only 69%, which is on the order of two randomly selected structures of the same protein with 100% identity. However, the trimeric quaternary structure of these chorismate mutases may enforce a certain structure in order to maintain the binding interfaces, forcing any mutations to be compliant within the structural constraints. Calculating the electrostatics using the APBS[31] approach indicates that the BpCM chorismate binding site is more positive than its BsCM homolog (Figure 1). Sitemap[32, 33] calculations of the three binding pockets of the trimers also show that the ratio of hydrogen bond donors to acceptors are higher for the BpCM. Meanwhile BpCM has a lower calculated pI than BsCM (5.33 compared to 5.6) and fewer positively charged amino acids (14 compared to 16) showing that this is a local effect.

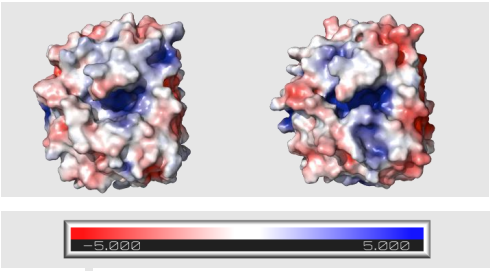


Figure 2: Electrostatics of BpCM and BsCM trimers indicate that the chorismate binding pocket is more positively charged in BpCM than in BsCM. Units given in kT/e

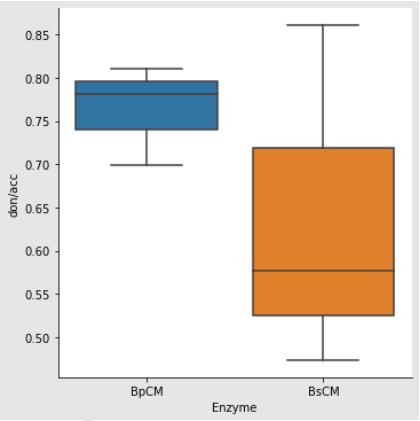


Figure 3: Comparison of binding pockets using SiteMap shows that the donor/acceptor ratio is higher for BpCM than for BsCM

Activation Parameters of B. pumilus Chorismate Mutase

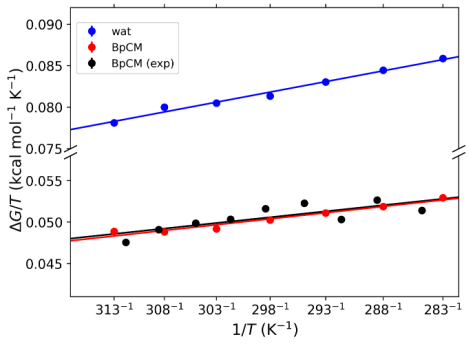


Figure 4: Arrhenius plots from 283-313 K comparing spectrophotometry (black) and EVB results for BpCM to EVB results of the reaction in an aqueous environment (blue).

Using the same intrinsic gas phase energy gap $\Delta\alpha$ and off-diagonal coupling H_{12} calibrated in our previous work on BsCM, we performed EVB simulations on BpCM at 7 different temperatures ranging from 283-313 K, from which an Arrhenius plot was constructed. Fitting $\Delta G^{\ddagger}/T$ vs. $1/T$ using a linear regression, we obtained thermodynamic activation parameters of $\Delta H^{\ddagger}$ = 12.9 ± 0.5 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = -2.1 ± 0.5 kcal mol^{-1}. An Arrhenius plot and similar regression of experimental results, obtained by fitting ln(k_{cat}) versus temperature values obtained from UV-Vis spectroscopy in a chorismic acid assay, yielded activation parameters of $\Delta H^{\ddagger}$ = 13.1 ± 0. kcal mol^{-1} and $T\Delta S^{\ddagger}$ = -2.1 ± 0.5 kcal mol^{-1}. The thermodynamic activation parameters predicted from the Arrhenius plot approach applied to both simulation and experimental results are in excellent agreement with each other.

Table 1: Thermodynamic activation parameters for the chorismate mutase (CM) reaction (chorismate → prephenate) in water (uncatalyzed) and catalyzed by chorismate mutase enzymes from *B. pumilus* (BpCM) and *B. subtilis* (BsCM).

		$\Delta\mathbf{G}^{\ddagger}$ (kcal/mol)	$\Delta\mathbf{H}^{\ddagger}$ (kcal/mol)	$\mathbf{T}\Delta\mathbf{S}^{\ddagger}$ (cal mol^{-1} K^{-1})
Uncatalyzed				
	Experimental[a]	24.5	20.7	-3.8
	EVB[c]	24.4 ± 0.6	20.9 ± 0.4	-3.4 ± 0.4
BpCM				
	Experimental	15.2	13.1	-2.1
	EVB	15.0 ± 0.7	12.9 ± 0.5	-2.1± 0.5
BsCM				
	Experimental[b]	15.4	12.7 ± 0.4	-2.7 ± 0.4
	EVB[c]	15.8 ± 1.2	12.9 ± 0.8	-2.9 ± 0.3

[a]Andrews et al.[28] [b] Kast et al. [34] [c]Wilkins et al.

Compared to the related chorismate mutase enzyme from *B. subtilis* (BsCM), which shares a 69% sequence identity and 93% similarity to BpCM, the thermodynamic activation parameters also indicate that like BsCM, BpCM is likely a mesophilic enzyme, despite other enzymes from *B. pumilus* exhibiting psychrophilic behavior.[35,36] The activation free energy for BpCM is less than 1 kcal mol^{-1} lower than reported values for BsCM, both experimentally and computationally. The activation enthalpies of both enzymes are the same within reported errors, thus the difference in $\Delta G^{\ddagger}$ is likely a result of BpCM having a lower $T\Delta S^{\ddagger}$ of about 0.6 kcal mol^{-1} at 298 K.

Structural Analysis of the Catalytic Effect of B. pumilus Chorismate Mutase

To further understand the mechanisms contributing to the catalytic effect in BpCM, we compare it to the structure to of BsCM, which has been well documented. According to our analysis of the thermodynamic activation parameters of the two enzymes,

the differences between the two appears to mostly due to entropic effects, as the activation enthalpies of the two enzymes are remarkably similar.

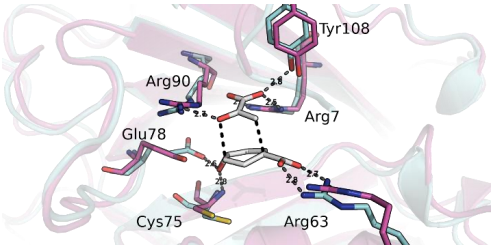


Figure 4: Active site comparison of BsCM (cyan) and BpCM (pink) with the transition state complex (white) obtained from the trajectory of MD simulations showing the highly conserved residues of interest.

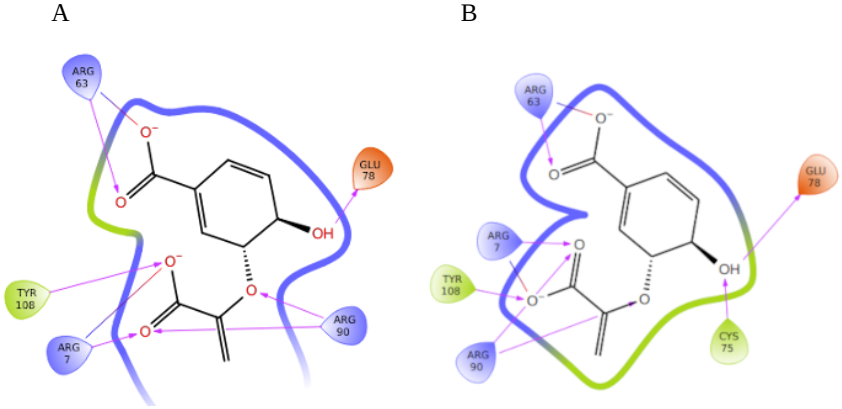


Figure 5: Ligand interaction diagrams of (A) BsCM and (B) BpCM showing a high degree of similarity in the active sites of the two enzymes.

An analysis of the active sites using both a 3D snapshot of the active site during the transition state (Figure 4) and a 2D ligand interaction diagram (Figure 5) generated in Maestro (Schrödinger Release 2019-3) [37] indicates a highly similar active site between the two enzymes, with identical hydrogen bonding between active site residues in both enzymes and the substrate, with the exception of the presence of an additional

hydrogen bond in the case of BpCM, between Cys75 and the hydroxyl group attached to the C_4 atom on the substrate. However, an analysis of trajectories generated from the molecular dynamics portion of our EVB simulations indicates a distance of 3.2 ± 0.2 Å between the main chain amide proton in Cys75 of BsCM, compared to a distance of 2.9 ± 0.2 Å with the corresponding Cys75 in BpCM (Figure 6). This is indicative that there is also hydrogen bonding present between this residue and the substrate in BsCM, however it lies outside of Maestro's distance cut-off. Hydrogen bonding between these residues and the C_4-OH are responsible for an elongation of the C_5-O_7 bond in chorismate, leading to a stabilization effect in the transition state of the substrate.[38] Mutagenesis studies where Glu78 was mutated to residues disrupting this hydrogen bonding indicate a reduction in the catalytic rate of BsCM, supporting the role for these residues in the catalytic effect of the enzymes.

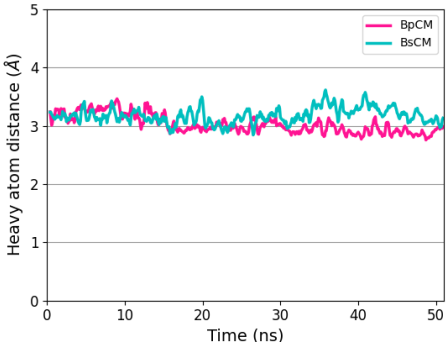


Figure 6: Heavy atom distances between the substrate hydroxyl group on C_4 and amide protons of Cys75 (BpCM in pink, BsCM in cyan) throughout the course of the chorismate mutase reaction.

Furthermore, a large negative charge is generated on the ether oxygen in the transition state by the elongation of the C_5-O_7 ether bond. In BpCM, the proximity of Arg90 likely provides a positive charge to counterbalance this effect. This corresponds directly to the Arg90 in BsCM fulfilling the same function due to the high active site similarity. Thus, the presence of these arginines in the active site is likely to help provide a positive charge stabilize a negatively charged region in transition state.

Formation of the C_1-C_9 bond is crucial in the formation of prephenate. To form this bond, the pyruvate moiety in the transition state must maintain a proper orientation with

respect to the cyclohexadienyl moiety. Mutagenesis studies on a related *E. coli* CM show a 10^4 reduction in k_{cat}/K_m when residues orienting the pyruvate moiety were mutated to prevent proper alignment during the transition state. [39] To maintain the proper orientation, Arg7 and Tyr108 in BpCM form strong ionic and hydrogen bonds with the pyruvate moiety in the transition state. Similarly, in BsCM these bonds to the pyruvate moiety are made with the same Arg7 and Tyr108.

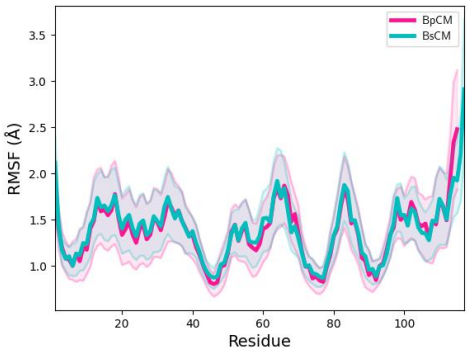


Figure 7: Root mean square fluctuations per residue of each monomer averaged over the N-C_α-C-O backbone atoms in BpCM (pink) versus BsCM (cyan) with error bands shown in respective colours.

An analysis of residue fluctuations in both enzymes was calculated from a total 35ns of molecular dynamics trajectories over 10 independent runs. Due to the high sequence identity and similarity between the two enzymes, the RMSF profiles are remarkably similar, with only slight deviations corresponding to residues found mostly in surface loop regions. An exception can be found in residue 45, found in the middle of the β-sheets forming the internal core of the enzymes, in which a polar Met45 in BsCM is found in pocket formed otherwise by hydrophobic residues. In contrast, the corresponding residue in BpCM is Ile45, which is likely to be more stable in this hydrophobic pocket.

To further understand the internal energetics of the reaction, we can perform a breakdown into bonded and nonbonded contributions (Table 2), as seen in our previous work (cite JCTC paper here when available). The total internal energy of the reacting region in BsCM is slightly more favorable than that of the BpCM, likely due to slightly less favorable nonbonded interactions in BpCM despite slightly more favorable bonded interactions. Analyzing the contributions to $\Delta U^{\ddagger}_{bonded}$ shows more favorable torsional and improper contributions to the internal energy, while contributions from bond and angle interactions are approximately same in both BsCM and BpCM. A similar

breakdown of the nonbonded interactions shows that while van der Waals contributions to the nonbonded internal energy are also similar between the two enzymes, the active site of BsCM provides a slightly more favorable electrostatic environment.

Table 2: Total potential activation energy within the reacting region and surroundings (in kcal/mol at 298 K) for the chorismate mutase (CM) reaction (chorismate → prephenate) from *B. pumilus* (BpCM) and from *B. subtilis* (BsCM), respectively.

	$\Delta \boldsymbol{U}^{\ddagger}_{\text{tot,r}}$	$\Delta \boldsymbol{U}^{\ddagger}_{\text{bonded}}$	$\Delta \boldsymbol{U}^{\ddagger}_{\text{nonbond}}$
BpCM	91.2	91.5	-0.3
BsCM	89.3	93.2	-3.9

*Subscripts rr, rs, and ss denote interactions between atoms in the EVB region, interactions between this region and surroundings, and internal interactions within the surroundings, respectively.

Concluding remarks

We have presented a new crystal structure and accompanying characterization of thermodynamic activation parameters for a new chorismate mutase enzyme from *Bacillus pumilus*. The sequence and structure show high similarity to the previously characterized *B. subtilis* chorismate mutase. Thermodynamic activation parameters obtained from experiment and molecular simulation, along with an accompanying structural analysis, show enthalpic contributions to the reaction pathway are identical, but there exists a small difference in activation entropy between BpCM and BsCM. Understanding the thermodynamic activation parameters and the underlying structural details is a crucial component leading to rational enzyme design. Understanding the way in which residues in the enzyme contribute to enthalpic and entropic effects in the chorismate mutase reaction establishes BpCM as a possible starting point in the design of psychrophilic or thermophilic chorismate mutase enzymes.

ACKNOWLEDGMENT

Support from the Norwegian Research Council through a Centre of Excellence and project grant (Grant Nos. 262695 and 274858), the Swedish Research Council (VR) and the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

References

[1] Kazemi, M., Himo, F., and Aqvist, J. (2016) Enzyme catalysis by entropy without Circe effect, *Proc Natl Acad Sci U S A 113*, 2406-2411.

[2] Strajbl, M., Shurki, A., Kato, M., and Warshel, A. (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization, *J Am Chem Soc 125*, 10228-10237.

[3] Lyne, P. D., Mulholland, A. J., and Richards, W. G. (1995) Insights into Chorismate Mutase Catalysis from a Combined QM/MM Simulation of the Enzyme Reaction, *Journal of the American Chemical Society 117*, 11345-11350.

[4] Marti, S., Andres, J., Moliner, V., Silla, E., Tunon, I., and Bertran, J. (2008) Predicting an improvement of secondary catalytic activity of promiscuous isochorismate pyruvate lyase by computational design, *J Am Chem Soc 130*, 2894-2895.

[5] Marti, S., Andres, J., Moliner, V., Silla, E., Tunon, I., Bertran, J., and Field, M. J. (2001) A hybrid potential reaction path and free energy study of the chorismate mutase reaction, *J Am Chem Soc 123*, 1709-1712.

[6] Gibson, F., and Pittard, J. (1968) Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microorganisms, *Bacteriol Rev 32*, 465-492.

[7] Görisch, H. (1978) A new test for chorismate mutase activity, *Analytical Biochemistry 86*, 764-768.

[8] Hur, S., and Bruice, T. C. (2003) Enzymes do what is expected (chalcone isomerase versus chorismate mutase), *J Am Chem Soc 125*, 1472-1473.

[9] Burschowsky, D., van Eerde, A., Okvist, M., Kienhofer, A., Kast, P., Hilvert, D., and Krengel, U. (2014) Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis, *Proc Natl Acad Sci U S A 111*, 17516-17521.

[10] Kienhofer, A., Kast, P., and Hilvert, D. (2003) Selective stabilization of the chorismate mutase transition state by a positively charged hydrogen bond donor, *J Am Chem Soc 125*, 3206-3207.

[11] Lassila, J. K., Keeffe, J. R., Kast, P., and Mayo, S. L. (2007) Exhaustive mutagenesis of six secondary active-site residues in Escherichia coli chorismate mutase shows the importance of hydrophobic side chains and a helix N-capping position for stability and catalysis, *Biochemistry 46*, 6883-6891.

[12] Åqvist, J., and Warshel, A. (1993) Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical

approaches, *Chemical Reviews 93*, 2523-2544.
[13] Kamerlin, S. C., and Warshel, A. (2010) The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions, *Faraday Discuss 145*, 71-106.
[14] Warshel, A., and Weiss, R. M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes, *Journal of the American Chemical Society 102*, 6218-6226.
[15] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.
[16] Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D Biol Crystallogr 64*, 61-69.
[17] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.
[18] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr 68*, 352-367.
[19] Marelius, J., Kolmodin, K., Feierberg, I., and Åqvist, J. (1998) Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, *Journal of Molecular Graphics and Modelling 16*, 213-225.
[20] Isaksen, G. V., Andberg, T. A., Aqvist, J., and Brandsdal, B. O. (2015) Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems, *J Mol Graph Model 60*, 15-23.
[21] Robertson, M. J., Qian, Y., Robinson, M. C., Tirado-Rives, J., and Jorgensen, W. L. (2019) Development and Testing of the OPLS-AA/M Force Field for RNA, *J Chem Theory Comput 15*, 2734-2742.
[22] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics 79*, 926-935.
[23] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics 81*, 3684-3690.
[24] Lee, F. S., and Warshel, A. (1992) A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations, *The Journal of Chemical Physics 97*, 3100-3107.
[25] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.

[26] Warshel, A. (1991) Computer modelling of chemical reactions in enzymes and solutions.

[27] Hwang, J. K., King, G., Creighton, S., and Warshel, A. (1988) Simulation of free energy relationships and dynamics of SN2 reactions in aqueous solution, *Journal of the American Chemical Society 110*, 5297-5311.

[28] Andrews, P. R., Smith, G. D., and Young, I. G. (1973) Transition-state stabilization and enzymic catalysis. Kinetic and molecular orbital studies of the rearrangement of chorismate to prephenate, *Biochemistry 12*, 3492-3498.

[29] Isaksen, G. V., Aqvist, J., and Brandsdal, B. O. (2014) Protein surface softness is the origin of enzyme cold-adaptation of trypsin, *PLoS Comput Biol 10*, e1003813.

[30] Kazemi, M., and Aqvist, J. (2015) Chemical reaction mechanisms in solution from brute force computational Arrhenius plots, *Nat Commun 6*, 7293.

[31] Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G. W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018) Improvements to the APBS biomolecular solvation software suite, *Protein Sci 27*, 112-128.

[32] Halgren, T. (2007) New method for fast and accurate binding-site identification and analysis, *Chem Biol Drug Des 69*, 146-148.

[33] Halgren, T. A. (2009) Identifying and characterizing binding sites and assessing druggability, *J Chem Inf Model 49*, 377-389.

[34] Kast, P., Asif-Ullah, M., and Hilvert, D. (1996) Is chorismate mutase a prototypic entropy trap? - Activation parameters for the Bacillus subtilis enzyme, *Tetrahedron Letters 37*, 2691-2694.

[35] Ranaghan, K. E., Shchepanovska, D., Bennie, S. J., Lawan, N., Macrae, S. J., Zurek, J., Manby, F. R., and Mulholland, A. J. (2019) Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution, *J Chem Inf Model 59*, 2063-2078.

[36] van der Ent, F., Lund, B. A., Svalberg, L., Purg, M., Chukwu, G., Widersten, M., Isaksen, G. V., Brandsdal, B. O., and Aqvist, J. (2022) Structure and Mechanism of a Cold-Adapted Bacterial Lipase, *Biochemistry 61*, 933-942.

[37] Schrödinger Release 2019-3: Maestro, Schrödinger, LLC.

[38] Worthington, S. E., Roitberg, A. E., and Krauss, M. (2001) An MD/QM Study of the Chorismate Mutase-Catalyzed Claisen Rearrangement Reaction, *The Journal of Physical Chemistry B 105*, 7087-7095.

[39] Liu, D. R., Cload, S. T., Pastor, R. M., and Schultz, P. G. (1996) Analysis of Active Site Residues in Escherichia coli Chorismate Mutase by Site-Directed Mutagenesis, *Journal of the American Chemical Society 118*, 1789-1790.

Paper III

Characterization of a thermophilic chorismate mutase and its utility as a template in rational enzyme design

Ryan Scott Wilkins, Susann Skagseth, Adele Williamson, Bjarte Aarmo Lund, Johan Åqvist, Bjørn Olav Brandsdal

Unpublished.

III

Characterization of a thermophilic chorismate mutase and its utility as a template in rational enzyme design

Ryan Scott Wilkins[a], *Susann Skagseth*[a], *Adele Williamson*[b], *Bjarte Aarmo Lund*[a] *and Bjørn Olav Brandsdal*[a,1]

[a]Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway

[b]School of Science, University of Waikato, Hamilton 3240, New Zealand

Abstract

Enzymes serve important roles in both research and industry. However, enzymes are fine-tuned to operate within a narrow range environmental conditions suitable to their host organism's natural environment. The demands placed on enzymes for applications in research and industry can often be outside of this narrow range of suitable conditions. Traditionally, the design of enzymes tailored to survive within laboratory or industrial conditions has been accomplished through random mutations in a process known as directed evolution. With increases in computing power, a better understanding of the structure and function of enzymes has enabled the selective mutation of enzymes. This rational design of enzymes is much more efficient than the traditional approach of randomized mutations. Within this manuscript, we aim to accomplish two goals: 1) the characterization of a new thermophilic chorismate mutase, and 2) the rational design of a thermophilic chorismate mutase, using the mesophilic *B. pumilus* chorismate mutase as a starting point and our new chorismate mutase as a template, through the combination of empirical valence bond simulations and analysis of structural details. The study utilizes a computational approach to compare the thermodynamic activation parameters of the mesophilic and thermophilic chorismate mutase and investigate the impact targeted mutations have in tuning these parameters. The results of the simulations and analysis provide insights into the key structural features that contribute to the stability of the enzyme in higher temperature ranges and inform the design of an optimized variant with improved thermophilic properties.

Introduction

Enzymes play crucial parts in both scientific research and industry due to their ability to increase reaction rates by several orders of magnitude. [1-4] However, enzymes are optimized to function within a limited range of conditions that are appropriate for their source organism. The controlled conditions found in research and industry are often quite different and demanding in comparison to the optimal environment for an enzyme. Furthermore, for more specific applications corresponding environmental conditions in nature have made bioprospecting for new enzymes difficult. [2] As such, a demand has been created to design novel enzymes tailored to function optimally within an applied setting. [3, 5, 6]

Traditional approaches to this problem involve directed evolution of enzymes. This involves introducing random mutations into a target gene, expressing the mutated genes in cells or organisms, and then selecting those with desired improved characteristics under the desired conditions, such as increased stability or specificity. With a selection of improved enzymes, the mutations are shuffled and repeated multiple times to generate new generations of the protein with improved properties. At each iteration, the best performing enzymes are selected until a threshold for optimal function is reached. [7-9] However, this process requires multiple iterations of implementing mutations, expressing the desired protein, and characterizing activity within the desired environment, so it is both time-consuming and expensive.

An increase in the power of computers has further led to an increase in computational methods in biochemistry which have provided new insights into enzyme structure and function previously unavailable in a wet lab. This has granted the ability for a different approach to enzyme design, where detailed structural information or bioinformatical information is utilized to pinpoint specific regions likely to impact function in a desirable manner. [6, 10, 11] This approach, known as rational enzyme design, is much more efficient and cost effective than the previously used directed evolution, as it removes much of the trial-and-error from the process. [5, 12]

However, there is still much controversy over the differences between mesophilic and extremophilic enzymes, leading to difficulties in rational enzyme design. In particular, without a template extremophile to be compared to showing the subtle

differences between a particular mesophilic-extremophilic homolog pair, it can be difficult to know the exact mutations to implement and where to implement them within an enzyme. [13] Therefore, to improve the process of rational enzyme design to a point where template enzymes are not required, it is imperative to gain a further understanding of the driving factors behind enzyme adaptation to extreme environments.

In the consideration of psychrophilic enzymes, the exponential dependence of the reaction rate on temperature is a large concern., as lowering the temperature can have detrimental effects on enzyme activity. [14-19] However, at the other temperature extreme, stability of the enzyme becomes the biggest concern. By increasing the temperature, weak interactions responsible for an enzyme's structure are weakened or destroyed, leading to degradation of the enzyme and subsequent loss of function. [20-23]

In order to cope with adaptation to extreme temperatures, enzymes have evolved numerous mechanisms to overcome the challenges presented by their respective environments. Looking at the active sites of temperature-adapted enzyme homologs, a high degree of conserved residues can be seen. This is likely due to enzyme active sites being highly selected for activity. [16] Thus, it is likely that surface resides located outside of the active site are largely responsible for altering an enzyme's ability to function at different temperatures. [16, 17, 21]

There are a variety of adaptive methods thermophiles have evolved to increase enzyme stability in hot environments. Bioinformatics-based analyses have shown an increase in hydrophobic packing near the surfaces, ion pairs forming salt bridges at near the surface, and shorter surface loops with lower flexibility are present in thermophilic enzymes. [20, 21] Structural analyses comparing mesophilic and thermophilic glutamate dehydrogenases and proteases show an increase in the density of surface ion pairs able to form salt bridges in thermophiles. [21, 24, 25]

There is evidence that these structural differences in protein surface flexibility correlate with the thermodynamics of the enzyme-catalyzed reaction. From computational studies on cold-adapted Atlantic salmon and mesophilic bovine trypsin, it was observed that while the activation free energies $\Delta G^{\ddagger}$ were similar between the two enzymes, the increased flexibility of the cold-adapted trypsin was accompanied by a redistribution of the

activation enthalpy $\Delta H^{\ddagger}$ and entropy terms $\Delta S^{\ddagger}$ comprising the free energy. Specifically, the cold-adapted trypsin exhibited a lower enthalpy and a less-negative entropy, thus reducing the temperature-dependence of the reaction. [16]

In this manuscript, we demonstrate the ability to rationally design mutate enzymes to fine-tune thermodynamic activation parameters of a mesophilic chorismate mutase from *B. pumilus*, resulting in thermophilic variants of BpCM. Using structural details and an analysis of thermodynamic activation parameters from M1CM, a novel thermophilic chorismate mutase also characterized within this paper, we were able to identify a loop region of BpCM to selectively mutate in an effort to increase thermal activity. An analysis of the thermodynamic activation parameters of the new variants indicates a shift of the activation enthalpy-entropy balance away from mesophilic activity towards increased thermal activity. This is further supported by a structural analysis of the new variants, which display a less-flexible loop, more closely matching the same loop region in M1CM.

Methods

Identification of Chorismate mutase from Antarctic metagenome

To identify novel chorismate mutase enzymes, metagenomic datasets from Miers, Marshall and Garwood valleys Antarctic Dry valleys were searched using a combination of HMM profiles and sequence similarity networks. [26] Protein coding sequences in FASTA format were downloaded from JGI's IMG/MER web portal [27] and searched with the pfam profile CM_1 (PF07736) using hmmsearch. [28] Sequences above the inclusion threshold (E-value <1) were used to construct a sequence similarity network via the EFI-EST server . [29, 30] The BLAST score cutoff threshold was set to E=1, sequences length cutoffs of >110 and <140 amino acids were used to exclude partial or misassembled metagenome and the BpCM sequence was included for reference. Sequences sharing one or more first-neighbor or second-neighbor edges with BpCM were aligned to check for completeness (presence of start and stop codon).

Protein production and purification

The full-length gene encoding the Chorismate Mutase from a metagenome found in the Antarctic was codon-optimized for expression in *Escherichia coli* and delivered by GenScript Biotech (Leiden, Netherlands) in a pET-22b(+) vector. The chorismic acid was supplied by Sigma-Aldrich.

The pET-22b(+)-M1CM was transformed into NiCo21(DE3) chemically competent *E. coli* cells (New England BioLabs, Ipswich, MA, USA) using a standard heat shock protocol. Cells were grown in shaking flasks in ZYP-5052 auto-induction media with 100 µg/ml Ampicillin after inoculated with a 3 ml over-night culture of LB media with 100 µg/ml Ampicillin. The culture was grown at 37 °C until a log phase (OD_{600} of 0,6 – 1,0) was reached and the temperature was adjusted to 16 °C for over-night expression. Cells were harvested by 20 minutes centrifugation at 6000 rpm and 4 °C. The cell pellets were resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl and 10 % Glycerol) before sonicated for 30 minutes with pulses of 1/1 seconds with active cooling. The cell extract was clarified by centrifugation at 14500 rpm for 45 minutes, and the supernatant collected. M1CM was affinity purified using a 5 ml HisTrap HP column (GE Healthcare) in buffer A washed with 5 % buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 10 % Glycerol, 375 mM Imidazole), before elution with 5 to 100 % buffer B. Fraction peaks were investigated using sodium dodecyl sulfate (SDS)-12% polyacrylamide gel electrophoresis (PAGE) (Bio-Rad). Fractions containing M1CM was dialyzed overnight in buffer C (50 mM HEPES pH 7.5, 10 % Glycerol) before anion exchange purification using a 5 ml HiTrap Q column (GE Healthcare) in buffer C. The M1CM was eluted with a gradient of buffer D (50 mM HEPES pH 7.5, 10 % Glycerol, 1 M NaCl). The elution peaks were investigated using SDS-PAGE gel before loading on a HiLoad Superdex 75 16/260 column in PBS buffer. M1CM was concentrated to 16 mg/ml and flash-frozen with liquid nitrogen for storage at -80 °C.

Thermal stability measurements using Thermal Shift Assay

The thermal stability of M1CM was found using a fluorescence-based thermal stability assay in a MJ minicycler (Bio-Rad) across a temperature gradient of 10 to 85 °C with a heating rate of 1 °C per minute. The assay volume used was 20 µl which included 20 µM enzyme, 5.5 X SYPRO Orange solution (Sigma-Aldrich), and buffer with 50 mM HEPES pH 7.0. The excitation and emission wavelengths of SYPRO orange are 470 nm and 570 nm. All experiments were performed in triplicates, and the melting temperatures were determined as the inflection point of the melting transition found from the first derivative.

Differential Scanning Calorimetry

To verify the results obtained from the Thermal stability measurements in the Thermal Shift Assay, Differential Scanning Calorimetry (DSC) was performed. The M1CM was dialyzes overnight in PBS buffer using a Slide-A-Lyzer Dialysis Cassette 30K MWCO (Thermo Fisher Scientific). The sample were then diluted to 1 mg/ml. PBS buffer were used as reference cell solvent. Prior to analyses, the reference solution and the M1CM sample were degassed using a Microcal Thermovac 2 (GE Healthcare). DSC data were obtained on CSC Nano-Differential Scanning Calorimeter III (N-DSC III) equipment with twin cells and operated on differential mode at a rate of 1 °C/min from 15-95 °C and a constant pressure of 3 atm. Melting points were determined using the NanoAnalyze 3.6 Software (TA Instruments, New Casttle, Delaware, USA).

Chorismic acid assay

A 10 mM stock solution of chorismic acid was used to make dilutions from 500 µM to 15 µM. The concentration of M1CM was kept at 10 nM for the experiments. The concentration of chorismic acid was followed at 274 nm, and concentrations were calculated using the extinction coefficient of 2630 $mol^{-1}cm^{-1}$. Measurements were done with 1 cm UV-quartz cuvettes in a sample-changer with a Cary60 UV/VIS spectrophotometer and a qChanger6 Peltier-cooled thermostat. The measurements were performed in triplicates and the velocities were fitted against the substrate concentrations with non-linear regression using GraphPad Prism 9.2.0 for Windows (GraphPad Software, La Jolla, California, USA).

Crystallization

From a set of eight commercial crystallization screens from Molecular Dimensions multiple hits were identified for M1CM when screened at 10 mg/ml. The screens were set up using MRC 96-well plates (Molecular Dimensions) and a Phoenix DT crystallization robot (Rigaku) and the sitting drop method at room temperature. Crystals from the Ligand-Friendly screen were obtained with 0.1 M SPG pH 8.0 and 30 % PEG 1000. Crystals appeared after 4 months. Diffraction data was collected at the ID30a remote from the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The data set were integrated, scaled and truncated using XDS, [31] POINTLESS, [32] and AIMLESS. [33] Molecular replacement was performed with PHASER[34] using the previously published trimer of *Thermus thermophilus* chorismate mutase (PDB: 4NWO) [35] as the search model. The structure was refined in several cycles using PHENIX [36, 37] and molecular modelled in WinCoot [38] according to the 2Fo-Fc and Fo-Fc map used to obtain the final structure.

Empirical valence bond simulations

In this study, we used the X-ray structure of M1CM superimposed with chorismate bound in BpCM from our previous work, which was used as the starting point for further simulations. For our simulations, we used the EVB/MM program Q [39] interfaced with Qgui[40] to generate topologies for the systems of interest, which were parameterized using the OPLS-AA/M force field. [41] The system was solvated within a 38 Å spherical boundary centered on the protein center-of-mass and filled with TIP3P water molecules [42] to simulate the effect of the solvent. To equilibrate the system before production runs, we gradually increased the temperature in six steps from 1 K up until the final system temperature, which was maintained using the Berendsen thermostat. [43] The SHAKE algorithm [44] was used to constrain bonds and angles of solvent molecules.

To calculate the free energy profiles from EVB simulations, we used the EVB/FEP umbrella sampling approach, dividing the calculations into 51 discrete windows between the chorismate and prephenate states. We parameterized the EVB Hamiltonian by fitting the uncatalyzed conversion of chorismate to prephenate in water to an activation free energy of $\Delta G^{\ddagger}$ = 24.5 kcal/mol and ΔG^{0} = -12.8 kcal/mol obtained from DFT calculations.

To obtain thermodynamic activation parameters for the enzyme, we employed the EVB/Arrhenius approach at 9 evenly spaced temperatures from 293-333 K for thermophilic enzymes and. At each temperature, 100 independent replicas were carried out for a total of 459 ns of simulated trajectories. We controlled the system temperature by coupling it to an external heat bath using a relaxation time of 10 fs for the first 5 equilibration steps and 100 fs for the final equilibration step and the production runs. Long-range electrostatics were calculated using a multipole expansion method (LRF), and nonbonded interactions were calculated using a 10 Å cut-off. However, reacting fragments in the catalytic center were free to interact with the entire system.

Results and Discussion

Characterization of a novel thermophilic chorismate mutase

The hmmsearch tool identified 3,376 hits to the CM_1 profile, 1356 of which were within the length constraints and were used to build the SSN. There were two were first-neighbor and 900 second-neighbor sequences to BpCM with edge values of >46% identity. Of these, both contigs containing the first-neighbor sequences annotated as *Anaerolinea thermophila* UNI-1 and the sequence Ga0136635_100017581 (hereafter M1CM) was selected on account of it being full-length and found within a cluster of shikimate-pathway biosynthetic enzymes.

The previously published trimer of *Thermus thermophilus* chorismate mutase (TtCM, PDB: 4NWO) [35] was used as the search model in molecular replacement as a blast search showed this to have the closest identity to the Antarctic chorismate mutase with a sequence identity of 60.3 %. The trimeric structure of M1CM is highly similar to both the search model TtCM and BpCM, with RMSD values of 0.32 Å and 1.56 Å, respectively.

The thermal unfolding curve of M1CM in Figure 1 was used to determine a melting temperature of T_m = 79.8 °C (Figure 1). This is higher than what was expected for a cold-adapted enzyme and indicates that the chorismate mutase from the Antarctic is not cold-adapted. In comparison, mesophilic chorismate mutases from *B. pumilus* and *B. subtilis* have T_m = 65.5 °C and T_m = 58.3 °C, respectively.

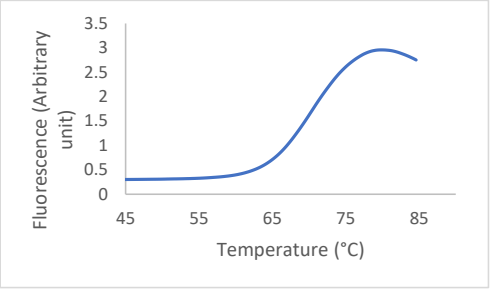


Figure 1: The thermal unfolding curve from ThermoFluor experiment. The fluorescence signal displays M1CM with an observed melting temperature (T_m) of 79.8 °C

An Arrhenius plot (Figure 2) was constructed by performing EVB simulations on M1CM at 9 different temperatures ranging from 293-333 K. Thermodynamic activation parameters $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ were obtained by fitting $\Delta G^{\ddagger}/T$ vs. $1/T$ using a linear regression. This resulted in calculated values of $\Delta H^{\ddagger}$ = 14.4 ± 0.5 kcal mol^{-1} and

$T\Delta S^{\ddagger}$ = -0.5 ± 0.5 kcal mol^{-1}. A similar analysis of converting ln(k_{cat}) values, as obtained from UV/VIS spectroscopy in a chorismic acid assay, to $\Delta G^{\ddagger}$ and constructing the corresponding $\Delta G^{\ddagger}/T$ vs. 1/T regression yielded activation parameters of $\Delta H^{\ddagger}$ = 15.0 kcal mol^{-1} and $T\Delta S^{\ddagger}$ = 0.2 kcal mol^{-1}. From both simulation and experiment, we see a shifting of the enthalpy-entropy balance opposite to psychrophilic enzyme behavior, resulting in a higher activation enthalpy and a less-negative activation entropy, as seen in Table 1. In combination with the high experimental melting point, it is likely that M1CM is a thermophilic or hyperthermophilic chorismate mutase.

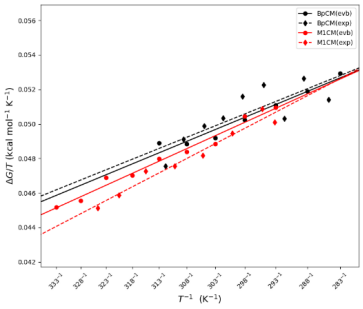


Figure 2: Arrhenius plots comparing M1CM (red) to BpCM (black) show thermophilic behavior for the M1CM.

Structural comparisons of a mesophilic and thermophilic chorismate mutase

As seen in Figure 3. the active sites of both BpCM and M1CM are highly conserved except for Arg63Lys. This arginine in BpCM forms hydrogen bonds with Cx-COO^{-}, which helps to orient the cyclohexadienyl moiety of the transition state. However, the substituted lysine in M1CM is also able to sufficiently fulfill this role, so it is unlikely that any differences in thermodynamic activation parameters are due to differences in the active sites between the mesophile and thermophile. This is in line with the popular hypothesis that active sites are highly optimized for activity, and any differences in thermodynamic activation parameters arise from differences in flexibility within surface residues on the enzyme. [10, 16-18, 45]

To further explore this hypothesis, we conducted a comparative elastic network model normal mode analysis (ENM-NMA). [46] From Figure 4, we can see a few areas of interest where the mesophilic BpCM is more flexible than the thermophilic M1CM.

A B

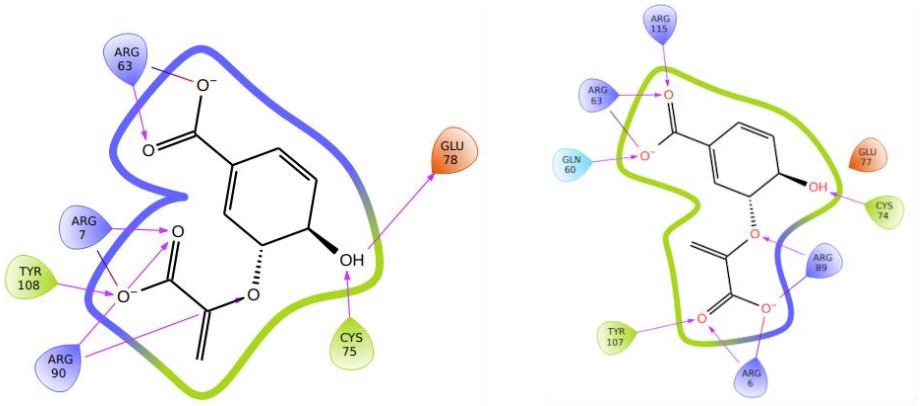
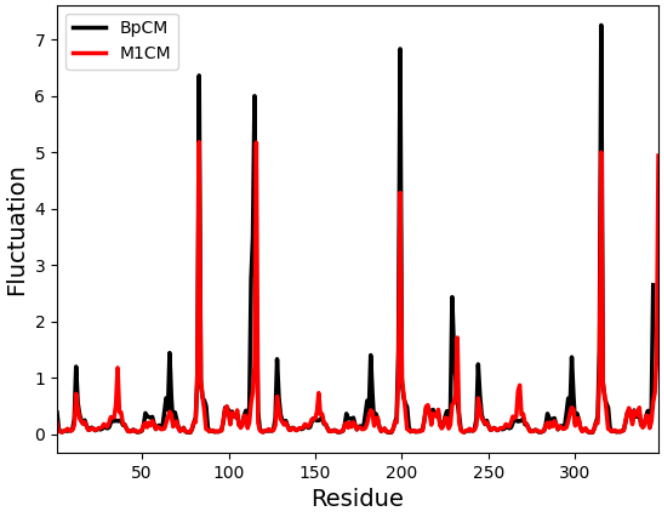


Figure 3: Ligand interaction diagrams of (A) BpCM and (B) M1CM show that important interacting residues within the active site are highly conserved between the two enzymes.

The area of highest disparity is the loop region from residues 83-87. However, this region closely borders Arg90, which is incredibly important for the formation of the transition state and is such not a suitable region for mutations. [47] Two other notable regions are the loop regions from residues 10-15 and from residues 65-70. Performing a normal mode analysis also allows for the construction of contact mapping, which can help predict correlated motions between residues. As seen in Figure 4b. the contact map resulting from our ENM-NMA shows a strong correlated motion between the two helices in the proximity of this loop, the shorter of which directly precedes the loop in residues 65-70.

A



B C

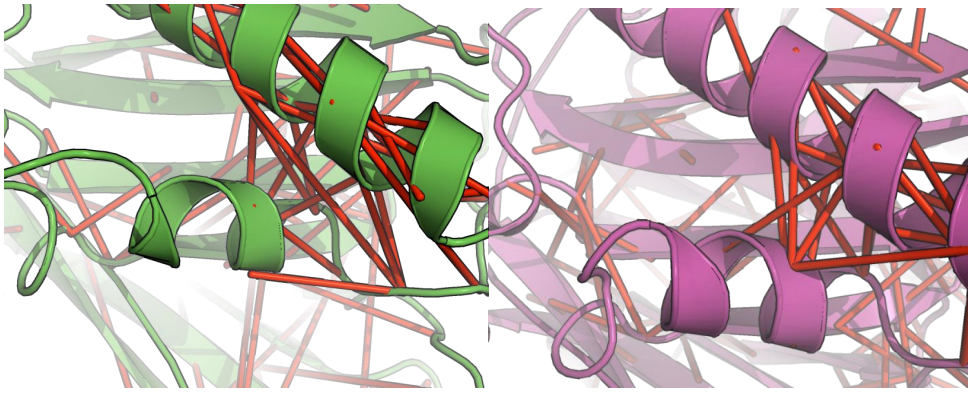


Figure 4: Comparative ENM-NMA showing fluctuations in BpCM and M1CM (A) and correlation mapping generated from ENM-NMA showing correlated motions between Calpha atoms in the two helices in BpCM (B) and M1CM (C).

From a structural analysis of this region in both BpCM and M1CM, as seen in Figure 5., there are three notable differences between the two enzymes. Firstly, in the thermophilic M1CM, the loop region in question is shorter in comparison to that of BpCM, corresponding to a deletion of Glu66 and Gly67 from BpCM. Secondly, there is an increase in hydrophobic packing between this loop and the longer helix in the thermophile. Finally, in the thermophilic chorismate mutase variant there exists a salt bridge between the two helices. These three observations are in agreement with predictions made from bioinformatical analyses comparing thermophilic and mesophilic enzymes. [20-23]

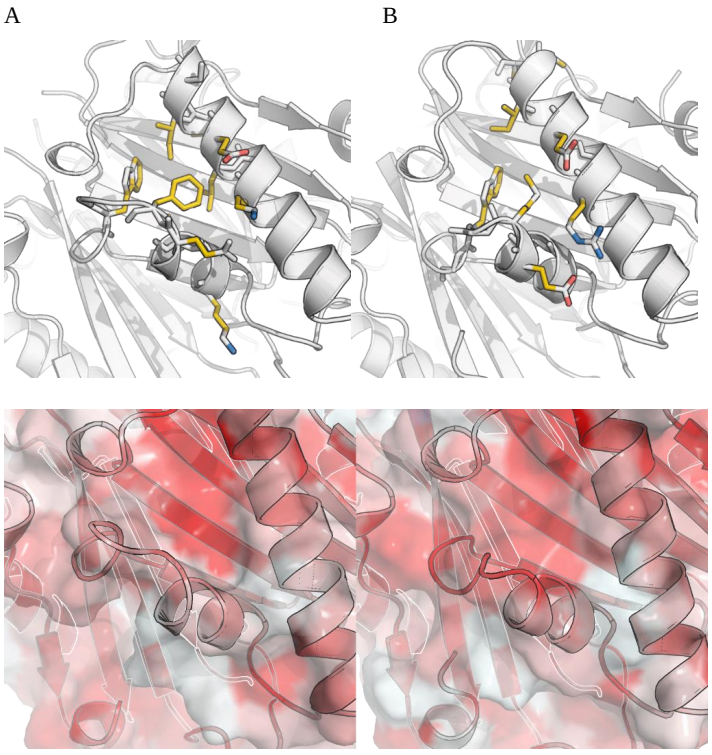


Figure 5: The flexible loop from residues 65-70 in mesophilic BpCM (A,C) with reduced mobility in thermophilic M1CM (B,D) This region is shown and contrasted in the YRB (A,B) [50] and Eisenberg (C,D) color schemes for hydrophobicity.[51]

Rational design of a thermophilic chorismate mutase

To produce a thermophilic chorismate mutase variant starting from the mesophilic BpCM, it is most logical to stabilize surface loops in which there is more flexibility in the mesophilic variant in comparison to its thermophilic counterpart. There is often no one mutation that will lead to better performance of extremophiles – instead, a variety of mutations in the same region may produce similar results via differing mechanisms. Comparing the loop of interest, located at approximately residues 65-70, between BpCM and M1CM, we observe three key differences – (1) an increase in hydrophobic packing in the region between the loop and two helices in the thermophilic M1CM, (2) a salt bridge between the two helices in M1CM, which does not exist in BpCM, and (3) a longer loop in the BpCM in comparison to M1CM, corresponding to the deletion of two residues. For this study, we focus only on point mutations, so we predict (1) and (2) to be potential targets for mutations in BpCM, either by increasing the hydrophobic packing or introducing a new salt bridge.

To increase the hydrophobic packing between the loop and helices in BpCM, we introduced a mutation of Glu29Phe. Contrary to other studies which generally target non-conserved residues, [16] Glu29 is conserved between BpCM and M1CM. In a more traditional approach, we also introduced mutations into a non-conserved region between the two enzymes through Gln64Glu. This mutation introduces a salt bridge between the mutated glutamate in the shorter helix and Arg24 in the longer helix, which also exists in the thermophilic M1CM. Both mutations resulted in an increased activation enthalpy and decreased activation entropy, matching the parameters of the thermophilic wild-type M1CM, as seen from the Arrhenius plots in Figure 6. and corresponding activation parameter breakdown in Table 1.

Remarkably, this shift in enthalpy-entropy balance was reflected in the flexibility of the mutated chorismate mutases. As seen in a comparison of RMSF plots (Figure 7), Both mutations of Glu29Phe and Gln64Glu resulted in a reduction of the flexibility of the wild-type BpCM, matching more similarly to the RMSF profile of the thermophilic M1CM. It is also notable that this effect was observed throughout the almost the entire enzyme.

Table 1: Thermodynamic activation parameters (in kcal/mol at 298 K) for the chorismate mutase (CM) reaction (chorismate to prephenate) in water (uncatalyzed) and catalyzed by chorismate mutases M1CM and from *B. pumilus* (BpCM), respectively. Reported errors are given as standard error of the mean.

		$\Delta G^{\ddagger}$	$\Delta H^{\ddagger}$	$T\Delta S^{\ddagger}$
Uncatalyzed				
	Experimental[a]	24.5	20.7	–3.8
	EVB	24.4 ± 0.6	20.9 ± 0.4	–3.4 ± 0.4
BpCM$_{wt}$				
	Experimental[c]	15.2	13.1 ± 0.4	–2.1 ± 0.4
	EVB	15.0 ± 0.7	12.9 ± 0.5	–2.1 ± 0.5
BpCM$_{E29F}$				
	EVB	15.1 ± 0.8	14.4 ± 0.8	–0.7 ± 0.5
BpCM$_{Q64E}$				
	EVB	15.1 ± 0.3	14.6 ± 0.2	–0.5 ± 0.2
M1CM				
	Experimental	14.8	15.0	0.2
	EVB	14.9 ± 0.8	14.5 ± 0.6	–0.5 ± 0.6
M1CM$_{E28F}$				
	EVB	16.1 ± 0.7	12.2 ± 0.5	-3.9 ± 0.5
M1CM$_{E64Q}$				
	EVB	16.0 ± 1.0	12.6 ± 0.7	-3.4 ± 0.8

A B

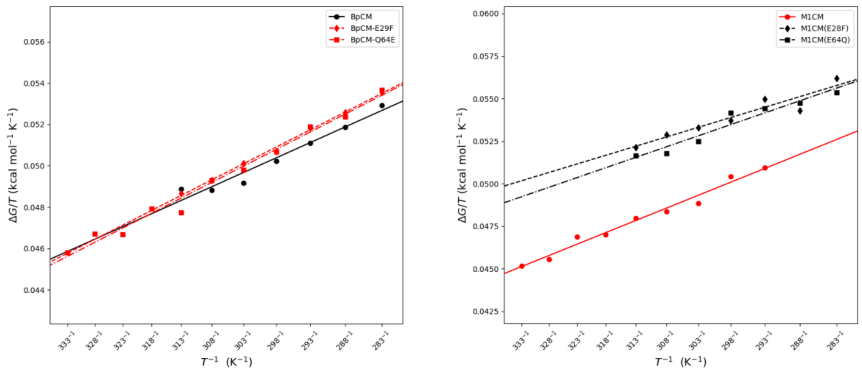


Figure 6: Arrhenius plots of the effects of mutations on BpCM (A) show a steeper slope and lower intercept in comparison to the wild-type, indicative of improved thermophilic activity. However, corresponding mutations in M1CM (B) have significant effects on the overall $\Delta G^{\ddagger}$ and are harder to quantify.

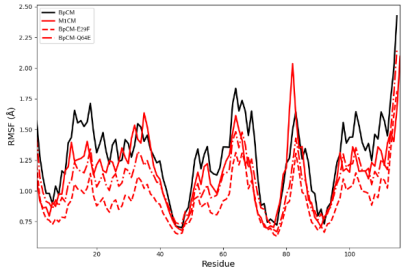


Figure 7: Comparison of RMSF profiles between BpCM (wild-type, E29F, and Q64E) and M1CM indicate a reduced flexibility in BpCM mutants, matching closely with the reduced flexibility of wild-type M1CM.

However, implementing mutations in M1CM proved to be more problematic. Disrupting the salt bridge with a mutation of Glu64Gln in M1CM, comparable to a reversal of the Gln64Glu mutation implemented in BpCM, did shift the thermophile's activation parameters more towards that of the mesophilic wild-type. However, doing so also increased the activation free energy up by approximately 1 kcal/mol, resulting overall in a less efficient reaction. This is in contract to findings by Isaksen *et al.* who found that implementing a set of reverse mutations between Atlantic salmon and bovine trypsin had minimal effects on $\Delta G^{\ddagger}$. [16]

As Glu29/Glu28 is conserved in both BpCM and M1CM, it was not possible to reverse this mutation. Instead, an identical mutation of Glu28Phe was implemented with the aim of further increasing the hydrophobicity in this region, possibly increasing thermostability. However, this mutation resulted in both an increase in $\Delta G^{\ddagger}$ by about 1 kcal/mol, and a shifting of the activation parameters towards those expected from a mesophile.

While implementing mutations into mesophilic BpCM had the expected effect of creating a mutant enzyme with thermodynamic activation parameters comparable to that of a thermophilic homolog, mutations intended to increase thermostability or increase flexibility in the thermophilic M1CM were less successful. This may be due to a multitude of reasons. For one, in BpCM the mutation of Glu29Phe creates a ladder cluster of aromatic rings with Phe65 and Trp68, which promotes stability. [48] In M1CM, however, the corresponding residue to Phe65 is a Met64. While sulfur-aromatic interactions can also have stabilizing effects, [49] this methionine is buried deeper within the hydrophobic pocket than the corresponding Phe65, which may influence in its ability to form this interaction in M1CM. Secondly, thermophilic enzymes may already be highly selected for increased thermostability, and even single point mutations can result in detrimental effects on and activity. However, thermostability is difficult to quantify using computational methods, so further biophysical characterization should be performed to determine if the implemented mutations affect both thermal stability and the thermodynamic activation parameters of the enzymes.

Conclusion

In this study, we have characterized a previously uncharacterized thermophilic chorismate mutase M1CM from a metagenomic sequence obtained in the Antarctic dry deserts using a combination of experimental and computational approaches. Our analysis of this enzyme has provided insights into its thermostability and activity at elevated temperatures, strongly indicating thermophilic behavior. Furthermore, we have utilized structural and trajectory information from our EVB simulations to introduce mutations into the mesophilic chorismate mutase from *B. pumilus*, resulting in the creation of new thermophilic variants that exhibit improved activity at higher temperatures. However, these mutations have yet to be tested

experimentally, so it is unclear if the shifting of the enthalpy-entropy balance in the activation free energy can entirely compensate for improved thermostability.

The discovery of this previously uncharacterized thermophilic chorismate mutase and the successful rational design of a novel variant with improved properties has significant implications for rational enzyme design. Our findings demonstrate the potential of using computational methods, such as the empirical valence bond method, to guide the rational design of enzymes for specific applications. Moreover, the insights gained from this study may inform the development of new computational tools and strategies for the rational design of enzymes with desired thermal activity.

References

[1] Whitaker, J. R. (1990) New and future uses of enzymes in food processing, *Food Biotechnology 4*, 669-697.

[2] Schmid, A., Hollmann, F., Park, J. B., and Buhler, B. (2002) The use of enzymes in the chemical industry in Europe, *Curr Opin Biotechnol 13*, 359-366.

[3] Cavicchioli, R., Charlton, T., Ertan, H., Mohd Omar, S., Siddiqui, K. S., and Williams, T. J. (2011) Biotechnological uses of enzymes from psychrophiles, *Microb Biotechnol 4*, 449-460.

[4] Radzicka, A., and Wolfenden, R. (1995) A proficient enzyme, *Science 267*, 90-93.

[5] Chica, R. A., Doucet, N., and Pelletier, J. N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, *Curr Opin Biotechnol 16*, 378-384.

[6] Steiner, K., and Schwab, H. (2012) Recent advances in rational approaches for enzyme engineering, *Comput Struct Biotechnol J 2*, e201209010.

[7] Arnold, F. H., and Volkov, A. A. (1999) Directed evolution of biocatalysts, *Curr Opin Chem Biol 3*, 54-59.

[8] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat Rev Genet 16*, 379-394.

[9] Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S. T., and Zhao, H. (2021) Directed Evolution: Methodologies and Applications, *Chem Rev 121*, 12384-12444.

[10] Socan, J., Isaksen, G. V., Brandsdal, B. O., and Aqvist, J. (2019) Towards Rational Computational Engineering of Psychrophilic Enzymes, *Sci Rep 9*, 19147.

[11] Teze, D., Zhao, J., Wiemann, M., Kazi, Z. G. A., Lupo, R., Zeuner, B., Vuillemin, M., Ronne, M. E., Carlstrom, G., Duus, J. O.,

Sanejouand, Y. H., O'Donohue, M. J., Nordberg Karlsson, E., Faure, R., Stalbrand, H., and Svensson, B. (2021) Rational Enzyme Design without Structural Knowledge: A Sequence-Based Approach for Efficient Generation of Transglycosylases, *Chemistry 27*, 10323-10334.

[12] Cherry, J. R., and Fidantsef, A. L. (2003) Directed evolution of industrial enzymes: an update, *Curr Opin Biotechnol 14*, 438-443.

[13] Renugopalakrishnan, V., Garduno-Juarez, R., Narasimhan, G., Verma, C. S., Wei, X., and Li, P. (2005) Rational design of thermally stable proteins: relevance to bionanotechnology, *J Nanosci Nanotechnol 5*, 1759-1767.

[14] Feller, G., and Gerday, C. (2003) Psychrophilic enzymes: hot topics in cold adaptation, *Nat Rev Microbiol 1*, 200-208.

[15] Siddiqui, K. S., and Cavicchioli, R. (2006) Cold-adapted enzymes, *Annu Rev Biochem 75*, 403-433.

[16] Isaksen, G. V., Aqvist, J., and Brandsdal, B. O. (2014) Protein surface softness is the origin of enzyme cold-adaptation of trypsin, *PLoS Comput Biol 10*, e1003813.

[17] Isaksen, G. V., Aqvist, J., and Brandsdal, B. O. (2016) Enzyme surface rigidity tunes the temperature dependence of catalytic rates, *Proc Natl Acad Sci U S A 113*, 7822-7827.

[18] Åqvist, J., Isaksen, G. V., and Brandsdal, B. O. (2017) Computation of enzyme cold adaptation, *Nature Reviews Chemistry 1*.

[19] Aqvist, J. (2017) Cold Adaptation of Triosephosphate Isomerase, *Biochemistry 56*, 4169-4176.

[20] Jaenicke, R., and Bohm, G. (1998) The stability of proteins in extreme environments, *Curr Opin Struct Biol 8*, 738-748.

[21] Kumar, S., and Nussinov, R. (2001) How do thermophilic proteins deal with heat?, *Cell Mol Life Sci 58*, 1216-1233.

[22] Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability, *Microbiol Mol Biol Rev 65*, 1-43.

[23] Berezovsky, I. N., and Shakhnovich, E. I. (2005) Physics and evolution of thermophilic adaptation, *Proc Natl Acad Sci U S A 102*, 12742-12747.

[24] Kumar, S., Tsai, C. J., Ma, B., and Nussinov, R. (2000) Contribution of salt bridges toward protein thermostability, *J Biomol Struct Dyn 17 Suppl 1*, 79-85.

[25] Mitra, D., and Das Mohapatra, P. K. (2021) Discovery of Novel Cyclic Salt Bridge in Thermophilic Bacterial Protease and

Study of its Sequence and Structure, *Appl Biochem Biotechnol 193*, 1688-1700.

[26] Williamson, A., Rzoska-Smith, E., Stelzer, R., Monteiro, M. R., and Cary, S. C. DNA Repair Enzymes of the Antarctic Dry Valley Metagenome, *Frontiers in Microbiology 14*, 1101.

[27] Chen, I. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., Hajek, P., Ritter, S., Varghese, N., Seshadri, R., Roux, S., Woyke, T., Eloe-Fadrosh, E. A., Ivanova, N. N., and Kyrpides, N. C. (2021) The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities, *Nucleic Acids Res 49*, D751-D763.

[28] Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010) Hidden Markov model speed heuristic and iterative HMM search procedure, *BMC Bioinformatics 11*, 431.

[29] Zallot, R., Oberg, N., and Gerlt, J. A. (2019) The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways, *Biochemistry 58*, 4169-4182.

[30] Oberg, N., Zallot, R., and Gerlt, J. A. (2023) EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools, *Journal of Molecular Biology*.

[31] Kabsch, W. (2010) Xds, *Acta Crystallogr D Biol Crystallogr 66*, 125-132.

[32] Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics, *Acta Crystallogr D Biol Crystallogr 67*, 282-292.

[33] Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr 62*, 72-82.

[34] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

[35] King, N. P., Bale, J. B., Sheffler, W., McNamara, D. E., Gonen, S., Gonen, T., Yeates, T. O., and Baker, D. (2014) Accurate design of co-assembling multi-component protein nanomaterials, *Nature 510*, 103-108.

[36] Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D Biol Crystallogr 64*, 61-69.

[37] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr 68*, 352-367.

[38] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

[39] Marelius, J., Kolmodin, K., Feierberg, I., and Åqvist, J. (1998) Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, *Journal of Molecular Graphics and Modelling 16*, 213-225.

[40] Isaksen, G. V., Andberg, T. A., Aqvist, J., and Brandsdal, B. O. (2015) Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems, *J Mol Graph Model 60*, 15-23.

[41] Robertson, M. J., Qian, Y., Robinson, M. C., Tirado-Rives, J., and Jorgensen, W. L. (2019) Development and Testing of the OPLS-AA/M Force Field for RNA, *J Chem Theory Comput 15*, 2734-2742.

[42] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics 79*, 926-935.

[43] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics 81*, 3684-3690.

[44] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.

[45] van der Ent, F., Lund, B. A., Svalberg, L., Purg, M., Chukwu, G., Widersten, M., Isaksen, G. V., Brandsdal, B. O., and Aqvist, J. (2022) Structure and Mechanism of a Cold-Adapted Bacterial Lipase, *Biochemistry 61*, 933-942.

[46] Tiwari, S. P., Fuglebakk, E., Hollup, S. M., Skjaerven, L., Cragnolini, T., Grindhaug, S. H., Tekle, K. M., and Reuter, N. (2014) WEBnm@ v2.0: Web server and services for comparing protein flexibility, *BMC Bioinformatics 15*, 427.

[47] Burschowsky, D., van Eerde, A., Okvist, M., Kienhofer, A., Kast, P., Hilvert, D., and

Krengel, U. (2014) Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis, *Proc Natl Acad Sci U S A 111*, 17516-17521.

[48] Lanzarotti, E., Biekofsky, R. R., Estrin, D. A., Marti, M. A., and Turjanski, A. G. (2011) Aromatic-aromatic interactions in proteins: beyond the dimer, *J Chem Inf Model 51*, 1623-1633.

[49] Valley, C. C., Cembran, A., Perlmutter, J. D., Lewis, A. K., Labello, N. P., Gao, J., and Sachs, J. N. (2012) The methionine-aromatic motif plays a unique role in stabilizing protein structure, *J Biol Chem 287*, 34979-34991.

[50] Hagemans, D., van Belzen, I. A., Moran Luengo, T., and Rudiger, S. G. (2015) A script to highlight hydrophobicity and charge on protein surfaces, *Front Mol Biosci 2*, 56.

[51] Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot, *Journal of Molecular Biology 179*, 125-142.