DISS. ETH No. 24879

De novo enzyme engineering:
Directed evolution of
computationally designed enzymes

A thesis submitted to attain the degree of

DOCTOR OF SCIENCE of ETH ZURICH
(Dr. sc. ETH Zürich)

presented by

Hans Adrian Bunzel
MSc Chemistry, Universität zu Köln
born on 23.02.1989
citizen of Germany

accepted on the recommendation of
Prof. Dr. Donald Hilvert
Prof. Dr. Helma Wennemers

2018

Et hät noch immer jot jejange!

It has always turned out great!

§3 Kölsche Jrundjesetz

§3 Constitution of Cologne

Acknowledgments

First and foremost, I would like to thank Prof. Donald Hilvert for his incredible support during the past years. It has been a privilege to work under his guidance and his broad scientific knowledge and great enthusiasm have been essential for the success of my work. Furthermore, I am particularly grateful for Don's talent to virtuously find the right tone, to either revive my mood or bring me back to earth, depending on how the experiments were going.

Before thanking all the incredible people who I was fortuned to get to know during the past years, I would like to thank my mother for always being there for me and supporting me in all of my life choices.

I would also like to thank Prof. Helma Wennemers for accepting to co-examine my thesis. I am further particularly grateful to my previous mentors, Prof. Günter Schwarz and Dr. Hajo Kries who have been great teachers during my undergraduate studies and master thesis. Furthermore, I would like to thank Prof. Peter Kast for being very helpful and supportive to overcome any experimental problems that I encountered during the course of my thesis. Also, I would like to thank the whole Hilvert Group, all present and past members, for being so incredibly awesome. The scientific and social dynamic of our group is truly unique. My time at ETH would not have been the same without the wonderful atmosphere to which everybody shared his part.

This thesis would not have been possible without the support of many talented collaborators. Dr. Hajo Kries and Joël Bloch provided me with fascinating Kemp eliminases to work with and fostered my initial attempts to study them. Special thanks goes to Prof. Stephen Boxer, Dr. Jacek Arthur Kozuch and Sam Hayes Schneider who support the research on the vibrational probes to large parts and who hosted me for a few weeks in their laboratories in Stanford. Whilst in Stanford, I was fortunate to collaborate with Prof. Jennifer Cochran, Arvind Kanaan, Dr. Jeff Galsgow on the μSCALE ultrahigh-throughput device for which I am very thankful. Also, I am grateful for the support of Prof. Dorothee Kern, Vy Nygen and Dr. Renee Otten who have been eager to support my research by deciphering the dynamics of HG3.17 using NMR spectroscopy. In addition, Prof. Kendall Houk, Dr. Gonzalo Jimenez-Oszes and Roger C. Helgeson provided beautiful Diels-Alderase designs and supported the laboratory

Danksagung

Als ehschte möch ich Prof. Donald Hilvert för sing Dirijeer en d'r vergangenen Johre danken. Et wor a Vermaach för en ze maloche, un sie jroßaatiges Wesse han d'r Usschlaach dä Arbeid bestemmp. Dorüvver erus ben ich ärch fruh för Dons Künn, emme d'r richtige Ton ze finge öm, hoorklein wie et em Labor läuf, ming Jemöt widder oplevve ze loße oder mich op de Ääd zoröck ze holle.

Bevör ich allen d'r unglaubliche Minsche danke, de ich en d'r letzten Johre kenne jelernt han, möch ich minger Moder doför danke sage, dat se emme för mich do wor un mir mi ganzes Levve lang bei allen däm Driss jeholfe hät.

Ich möch och Prof. Helma Wennemers doför danke, dat se et akzeptiert han ming Arbeid ze kontrolleere. Ärch dankbar ben ich och minge fröhere Mentore Prof. Günter Schwatz un Dr. Hajo Kries, de en mingem Bachelor un Master hervorragende Majister wohren. Prof. Pitter Kast möch ich för sing Hilfe beim Lüse vun d'r Probleme de ich en dä Arbeid em Labor hatt danke. Ich möch och d'r jesampte Hilvert Familich, allen neue un alde Mitglieder vum Trupp, doför danken, dat se esu unglaublich joldich sin. Ming Zick ahn d'r ETH wör ohne de wunderbare Minsche nit diselve jewese.

De These wör ohne der Hölp vun vielen talentierten Minsche nit müjjelich jewese. Dr. Hajo Kries un Joël Bloch han viele d'r lüsche Kemp-Eliminase, met denen ich jearbeidet han jemaach. Et wor e vermaach met üch ze arbeide. A jroße Dank jeiht ahn Prof. Stephen Boxer, Dr. Jacek Arthur Kozuch un Sam Hayes Schnieder, der der Erforschung d'r Schwingungssonde ze großen Deile unterstützt un met mir en Stanford jearbeidet han. Während ich en Stanford wor, hatt ich och dat Dusel, met Prof. Jennifer Cochran, Arvind Kanaan und Dr. Jeff Galsgow am μSCALE Ultrahochdurchsatz-Jeräd ze forsche, wofür ich ärch dankbar ben. Usserdäm ben ich dankbar för der Unterstützung vun Prof. Dorothee Kern, Vy Nygen un Dr. Renee Otten, der ming Forschung jeholfe han indem se der Bewägung vun HG3. 17 met NMR ungersökt han. Dorüvver hinaus han Prof. Kendall Houk, Dr. Gonzalo Jimenez-Oszes un Roger C. Helgeson der Diels-Alderase Designs jemaach un deren Evolution ungerstüz, wofür ich ärch dankbar ben. Zodäm möch ich Prof. Nades Hauer, för der Unterstützung bei d'r Erforschung vun promiskuitiven Aktivitäten, un Dr. Frank Biederman, för der wertvolle Enfälle för d'r Ultrahochdurchsatz-Assay, danke. Der en dä Arbeid vorjestellte Bestimmung vun

Röntgenstrukturen woot weitgehend vun PD Peer Mittl, Dr. Takhhiro Mori un Beat Blattmann ungerstüz, denen ich ärch dankbar ben.

Usserdäm hatt ich dat Dusel, viele talentierte Studente ze betreue. Angelika Eske, Daniel Joss, Ines Folger, Lukas Grenacher un Luca Machetti: Viele Dank för ör Hilf. Zodäm möch ich och Steffe Tetter, Dr. Cathleen Zeymer, Anna Weigand un Chris Fröhlich för dat Kontrolleere vun minger Arbeid danke.

Zu letz möch ich minger janze Familich danke, die während minger Zick en d'r Schweiz en d'r letzten Johre flöck jewachse es un emme för mich do wor. Der Erforschung d'r Enzyme met Don wor ohne Zweifel a Feß en d'r letzten Johre, Adrian ze treffen un a Teil d'r Egloffs un O' Gradys ze wäde, wor sicherlich a andre.

Contents

Abstract

Enzymes, Nature's catalysts, ensure that all essential metabolic processes are fast enough to sustain life. They have been refined over billions of years through natural evolution, often acquiring exceptional selectivities and rate accelerations that are unmatched by any man-made catalysts. Harnessing this power is a longstanding aim in synthetic biology. Toward that end, a combination of computational design and directed evolution has recently emerged as a powerful approach to engineer novel enzymes (Chapter 1). Novel active sites have been successfully introduced into inert protein scaffolds by computational design and subsequent optimization by directed evolution afforded enzyme-like catalysts in several cases. Despite these successes, the approach remains challenging. New designs typically display limited activity, and achieving practical rate accelerations typically requires cost and labor intensive laboratory evolution. The current challenge in *de novo* engineering is thus to improve the overall approach to reliably generate enzymes with high activities.

The computationally designed and experimentally evolved Kemp eliminase HG3.17 is one of the most efficient artificial enzymes engineered today. It catalyzes the conversion of benzisoxazoles into salicylonitriles – a mechanistically well understood, one-step model reaction for proton abstraction from carbon – with a 6×10^8-fold rate acceleration over background. To assess the origins of its remarkable activity, the enzyme was subjected to a detailed biophysical analysis (Chapter 2). First, a minimal active site composed of only the transition state and the catalytic residues was computationally modeled. The model indicates that catalysis depends on an intricate balance between positioning and electrostatics in these systems. However, the simple model that was used, which resembles those employed during computational enzyme design, is unable to fully explain catalysis in HG3.17.

Two vibrational probes were thus developed to scrutinize the contributions of active-site electrostatics to activity. The probes, based on benzotriazole transition state analogs, were tested with several single-point mutants of HG3.17 and the original design HG3. The protonation equilibrium between the probe and the catalytic base was shifted upon binding of the probe to the variants as indicated by the ratios of the peaks corresponding to the protonated and deprotonated probe in the vibrational spectra. The electrostatic environment at the enzyme active sites likely caused that perturbation as indicated by a change in peak

frequencies and supported by computational modeling. Correlation of the peak ratios and frequencies with the catalytic efficiencies of the respective variants further indicates that catalysis in HG3.17 largely relies on fine-tuned electrostatic interactions.

Biophysical characterization of HG3.17 and the original design HG3 additionally revealed pH- and temperature-dependent inactivation of the enzymes by a slow conformational change. The kinetics of that conformational change were studied by monitoring changes in tryptophan fluorescence and indicate that evolution selectively stabilized the active state of the enzyme under ambient conditions. Remodeling of conformational landscapes in this way likely increased activity of the starting catalysts by as much as 20 fold. Similar effects may have shaped the evolution of many natural enzymes.

Kemp eliminase HG3.17 achieves exceptional activity for a designed catalyst. The research presented in Chapter 3 aimed to repeat that success with another designed Kemp eliminase, 1A53-2. The latter was evolved over nine rounds of mutagenesis and screening to give 1A53-2.9, which has a catalytic proficiency ($k_{cat}/K_M/k_{uncat} = 8\times10^{10}$ M^{-1}) comparable to HG3.17 (2×10^{11} M^{-1}). Furthermore, 1A53-2.9 remained fully active up to 45 °C, allowing assessment of changes in the activation enthalpy and entropy of the reaction. The temperature dependencies revealed that evolution reduced the activation enthalpy for 1A53-2.9 compared to 1A53-2 ($\Delta\Delta H^{\ddagger} = -10.9$ kcal/mol), whereas the entropic barrier increased ($-T\Delta\Delta S^{\ddagger} = +5.4$ kcal/mol). The changes in activation enthalpy of 1A53-2.9 during evolution from a naïve design to an efficient catalyst nicely support the hypothesis that modern enzymes evolved by reducing the activation enthalpy of their primordial ancestors.

The screening step often comprises the bottleneck of directed evolution experiments. For this reason, implementation of ultrahigh-throughput assays can significantly advance enzyme engineering efforts. New fluorescence-based assays required for ultrahigh-throughput analysis of Kemp eliminase activity were therefore developed in Chapter 4. Fluorescent readouts were generated through pH changes that occur upon substrate conversion (1), a supramolecular chemosensor (2), and a novel fluorogenic substrate (3). In preliminary experiments with an ultrahigh-throughput microcapillary system, a pH-based fluorescence assay was able to discriminate cells that express Kemp eliminases on their surface from those that do not. Screening large protein libraries in this way may allow engineering of even more efficient Kemp eliminases in the future.

Chapter 5 explores the promiscuous activities of Kemp eliminase 1A53-2. Specifically, the enzyme's catalytic glutamic acid, which has an unusually high pK_a (8.2±0.2), was exploited as a Brønsted acid catalyst for the cyclization of monoterpenes such as citral or citronellal. The activity with citral at pH 6.0 was increased tenfold over five rounds of mutagenesis and screening. Further evolution of this activity may provide a stereoselective enzyme for the synthesis of precursors for commercially valuable substances such as (-)-menthol.

Another strategy for diversifying enzyme function couples the intrinsic reactivity of metal ions with the selectivities in protein pockets (Appendix). The metal-binding site of human carbonic anhydrase II was redesigned to accelerate a zinc-dependent Diels-Alder reaction between a pyridyl-based chalcone and 3-vinylindole using computational design in collaboration with the Houk group (UCLA). Nevertheless, replacement of the native zinc ion with copper was found to be crucial for Diels-Alderase activity. A screening assay involving protein immobilization and metal-ion exchange was developed and used to screen enzyme libraries. The Diels-Alderase was improved 20 fold over three rounds of directed evolution. The developed platform could be valuable for diversification of many artificial enzymes based on highly reactive metal-ion catalysis.

The research presented in this thesis contributes in several ways to *de novo* enzyme engineering. First, it shows that novel enzymes can be reproducibly engineered, while highlighting the relevance of electrostatic catalysis and conformational changes at the stage of computational design. Second, it establishes a basis for new ultrahigh-throughput assays. Third, it suggests that alternative design strategies might be profitably explored by repurposing catalytic devices present in designed and artificial enzymes. Implementation of the insights provided by this thesis may lead to many novel efficient enzymes for practical application in the future.

Zusammenfassung

Enzyme, die Katalysatoren der Natur, beschleunigen alle essentiellen Stoffwechselprozesse und ermöglichen das Leben so wie wir es kennen. Sie wurden über Milliarden von Jahren durch natürliche Evolution verfeinert und besitzen oft außergewöhnliche Effizienz und Selektivität, welche bislang von keinem künstlichen Katalysator erreicht wurden. Es ist ein langfristiges Ziel der synthetischen Biologie dieses katalytische Potential für neue Zwecke nutzbar zu machen. Eine Kombination aus computergestütztem Design und gerichteter Evolution hat sich hierbei als vielversprechend für die Entwicklung neuer Enzyme erwiesen (Kapitel 1). Neue aktive Zentren können mittels computergestütztem Design in inaktive Proteingerüste eingeführt werden. Anschließende gerichtete Evolution hat in einigen Fällen zu effizienten Katalysatoren geführt. Trotz dieser Erfolge ist das Entwickeln neuer Enzyme anspruchsvoll. Designerenzyme weisen typischerweise begrenzte Aktivität auf, und das Erreichen von signifikanten Aktivitäten erfordert häufig kosten- und arbeitsintensive gerichtete Evolution. Die aktuelle Herausforderung besteht daher darin, den gesamten Prozess zu verbessern um zuverlässig neue Enzyme mit hohen Aktivitäten zu generieren.

Einese der effizientesten künstlichen Enzyme ist Kemp Eliminase HG3.17, welche mittels computergestüztem Design und gerichteter Evolution entwickelt wurde. HG3.17 katalysiert die Umwandlung von Benzisoxazolen in Salicylonitrile - eine mechanistisch gut verstandene, einstufige Modellreaktion für die Protonenabstraktion von Kohlenstoff - mit einer 6×10^8-fachen Beschleunigung gegenüber der unkatalysierten Reaktion. Um den Ursprung dieser hohen Aktivität zu verstehen wurde das System einer detaillierten biophysikalischen Analyse unterzogen (Kapitel 2). Zuerst wurde eine computergestützte Analyse mit einem minimalen aktiven Zentrum durchgeführt, welches nur aus dem Übergangszustand des Substrats und den katalytischen Resten bestand. Die Modellierung zeigt, dass ein kompliziertes Gleichgewicht zwischen Positionierung und Elektrostatik wichtig für die Katalyse in diesem System ist. Das einfache Modell, welches in dieser Studie verwendet wurde, ist jedoch nicht in der Lage die Katalyse in HG3.17 vollständig zu erklären.

Zwei Schwingungssonden wurden daher entwickelt um die Beiträge der Elektrostatik im aktiven Zentrum zur Aktivität zu bestimmen. Diese Sonden basieren auf

Benzotriazol-Übergangszustandsanaloga und wurden mit mehreren Einzelpunktmutanten von HG3.17 und dem ursprünglichen Design HG3 getestet. Das Protonierungsgleichgewicht zwischen den Sonden und der katalytischen Base ist in den einzelnen Varianten verschoben, was durch die relativen Verhältnisse der protonierten und deprotonierten Peaks der Sonden in den Schwingungsspektren ersichtlich war. Zudem lässt die Änderung der Peakfrequenzen in Übereinstimmung mit Computermodellen vermuten, dass diese Änderungen wahrscheinlich durch die elektrostatische Umgebung im aktiven Zentrum der Enzyme verursacht wird. Die Korrelation der Peakverhältnisse und -frequenzen mit der Aktivität der jeweiligen Varianten weist darauf hin, dass die hervorragende Aktivität in HG3.17 zu grossen Teilen aus der Elektrostatik des aktiven Zentrums resultiert.

Die biophysikalische Charakterisierung von HG3.17 und dem ursprünglichen Design HG3 zeigte weiterhin eine pH- und temperaturabhängige Inaktivierung der Enzyme durch eine langsame Konformationsänderung. Die Kinetik dieser Konformationsänderung wurde untersucht, indem Veränderungen in der Tryptophanfluoreszenz beobachtet wurden. Hieraus ging hervor, dass die Evolution den aktiven Zustand des Enzyms unter Normalbedingungen selektiv stabilisierte. Durch diese Verschiebung der Konformationsgleichgewichte wurde die des Ausgangskatalysators wahrscheinlich um das 20-fache erhöht. Ähnliche Effekte könnten die Evolution vieler natürlicher Enzyme beeinflusst haben.

Kemp Eliminase HG3.17 erreicht außergewöhnliche Aktivität für ein künstliches Enzym. Die in Kapitel 3 vorgestellte Studie zielt darauf ab, diesen Erfolg mit einer anderen designten Kemp-Eliminase, 1A53-2, zu wiederholen. Letztere wurde über neun Runden Mutagenese und Screening zu einer katalytischen Leistungsfähigkeit ($k_{cat}/K_M/k_{uncat}$ = 8x10^{10} M^{-1}) verbessert, die mit HG3.17 (2x10^{11} M^{-1}) vergleichbar ist. Der evolvierte Katalysator, 1A53 -2.10, bleibt ferner bis zu einer Temperatur von 45 °C vollständig aktiv und erlaubte es daher Änderungen in der Aktivierungsenthalpie und Entropie der Reaktion zu bestimmen. Die Temperaturabhängigkeiten zeigten, dass die Evolution die Aktivierungsenthalpie für 1A53-2.9 im Vergleich zu 1A53-2 reduzierte ($\Delta\Delta H^{\ddagger}$ = -10.9 kcal/mol), während die entropische Barriere anstieg ($-T\Delta\Delta S^{\ddagger}$ = +5.4 kcal/mol). Die Änderungen der Aktivierungsenthalpie von 1A53-2.9 während der Evolution von einem naiven Design zu einem effizienten Katalysator unterstützt daher die Hypothese, dass moderne Enzyme durch die Verringerung der Aktivierungsenthalpie ihrer evolutionären Vorfahren entstanden sind.

Der Screening-Schritt ist oftmals der Engpass in der gerichteten Evolution. Aus diesem Grund kann die Implementierung eines Verfahrens mit ultrahohem Durchsatz die gerichtete Evolution von Enzymen wesentlich beschleunigen. Neue fluoreszenzbasierte Analysen, die für die Analyse von Kemp-Eliminase mit ultrahohem Durchsatz benötigt werden, wurden daher in Kapitel 4 entwickelt. Fluoreszenzsignale wurden durch (1) pH-Änderungen aufgrund des Substratumsatze, (2) einen supramolekularen Chemosensor, und (3) ein neues fluorogenes Substrat erzeugt. In ersten Experimenten mit einem Ultrahochdurchsatz-Mikrokapillarsystem war es mittels eines pH-basierten Fluoreszenzassays möglich, Zellen, die Kemp-Eliminasen auf ihrer Oberfläche exprimieren, von denen zu unterscheiden, die kein Enzym produzieren. Durch das Analysieren großer Proteinbibliotheken könnten mit solchen Systemen in Zukunft noch effizientere Kemp-Eliminasen generiert werden.

In Kapitel 5 wurden die promiskuitiven Aktivitäten der Kemp-Eliminase 1A53-2 untersucht. Insbesondere wurde die katalytische Glutaminsäure in den Enzymen, die einen ungewöhnlich hohen pKa-Wert (8.2±0.2) aufweisen, als Brønsted-Säure-Katalysatoren für die Cyclisierung von Monoterpenen wie Citral oder Citronellal genutzt. Die Aktivität mit Citral bei pH 6.0 wurde über fünf Runden der gerichteten Evolution verzehnfacht. Eine weitere Verbesserung dieser Aktivität könnte zu einem stereoselektives Enzym für die Synthese von kommerziell wertvollen Terpenen, wie beispielsweise (-)-Menthol, beisteuern.

Eine andere Strategie zur Diversifizierung von Enzymfunktion koppelt die intrinsische Reaktivität von Metallionen mit den Selektivitäten der Proteintaschen. In einem Nebenprojekt in Zusammenarbeit mit der Houk-Gruppe (UCLA) wurde die metallbindende Stelle der humanen Carboanhydrase II neu gestaltet, um eine Zink-abhängige Diels-Alder-Reaktion zwischen einem Pyridyl-substituiertem Chalkon und 3-Vinylindol mit Hilfe von Computerdesign zu beschleunigen (Appendix). Der Austausch des nativen Zinkions gegen reaktiveres Kupfer erwies sich hierbei als entscheidend für die Diels-Alderase-Aktivität. Ein Screening-Protokoll wurde entwickelt, das den Austausch von Metallionen mittels Proteinimmobilisierung ermöglicht. Die Diels-Alderase wurde über drei Runden gerichteter Evolution um das 20 fache verbessert. Die entwickelte Plattform könnte sich für die Diversifizierung künstlicher Enzyme auf der Basis hochreaktiver Metallionenkatalyse als wertvoll erweisen.

Die in dieser Arbeit vorgestellte Forschung trägt in mehrfacher Hinsicht zur Herstellung neuer Enzyme bei. Erstens wird gezeigt, dass neue Enzyme reproduzierbar entwickelt werden

können. In diesem Zusammenhang hat sich auch das Potential elektrostatischer Katalyse und konformeller Änderungen in der Phase des computergestützen Designs herauskristallisiert. Zweitens wird eine Grundlage für neuartige Analysen mit ultrahohem Durchsatz geschaffen. Drittens werden alternative Strategien zur Herstellung neuer Enzyme erkundet, indem katalytische Reste, welche sich bereits in entworfenen und künstlichen Enzymen befinden, wiederverwendet werden. Die Realisierung der Erkenntnisse dieser Arbeit könnte zu vielen neuartigen effizienten Enzymen für die praktische Anwendung führen.

1 Enzyme catalysis and *de novo* enzyme engineering

1.1 The power of enzyme catalysis

Enzymes, designed by nature over billions of years of natural evolution, have the power to accelerate their target reactions with unequaled efficiency [1]. The most proficient enzyme described to date, an alkyl sulfatase, cleaves its preferred substrate 10^{26}-fold faster than the background reaction. While enzymes have been employed by mankind for centuries [2], their unmatched activities and selectivities were only truly appreciated with the advent of industrial biocatalysis [3]. Fostered by molecular biology, biocatalysis gained increasing importance, first through rational protein engineering and later by the ability to remodel enzymes by directed evolution to suit human demands. Despite longstanding interest and intense study of these remarkable catalysts, the origins of enzymatic proficiency remain elusive. Nevertheless, enzyme design has matured over the past two decades. In a few notable cases, enzyme-like catalysts have been successfully engineered through computation and evolution [4,5]. This thesis aims at the dissection of the origins of catalysis in these biocatalysts, as well as the engineering of new efficient artificial enzymes by directed evolution.

1.1.1 Selective transition state stabilization

The first model to demystify biocatalysis was proposed by Fischer and dates back to the 19th century. Despite the lack of any structural data, he postulated a "lock and key" mechanism after observing configuration-specific enzymatic cleavage of various glucosides [6]. Through careful study of the kinetics of the responsible glycosidases, Michaelis and Menten developed a simple mechanistic model to describe enzyme catalysis [7]. Building on these kinetic studies, Alberty and Hammes noted that the apparent second-order rate constant, k_{cat}/K_M, of enzymes that achieved catalytic perfection is solely limited by diffusion of the reactants [8]. Pauling proposed that enzymes accelerate their reaction by selective transition state stabilization [9,10], a hypothesis that inspired the successful design of many transition state analogs that function as potent competitive inhibitors [11] and fostered the creation of catalytic antibodies [12]. Building on this proposal, Kurz derived a formalism, $k_{cat}/k_{uncat} = K_S/K_{TS}$, connecting rate acceleration with selective transition state binding (Figure 1, [13]).

A

B

$$\frac{v_0}{E_0} = \frac{k_{cat} \cdot [S]}{K_M + [S]}$$

C

$$\frac{k_{cat}}{k_{uncat}} = \frac{K_S}{K_{TS}}$$

D
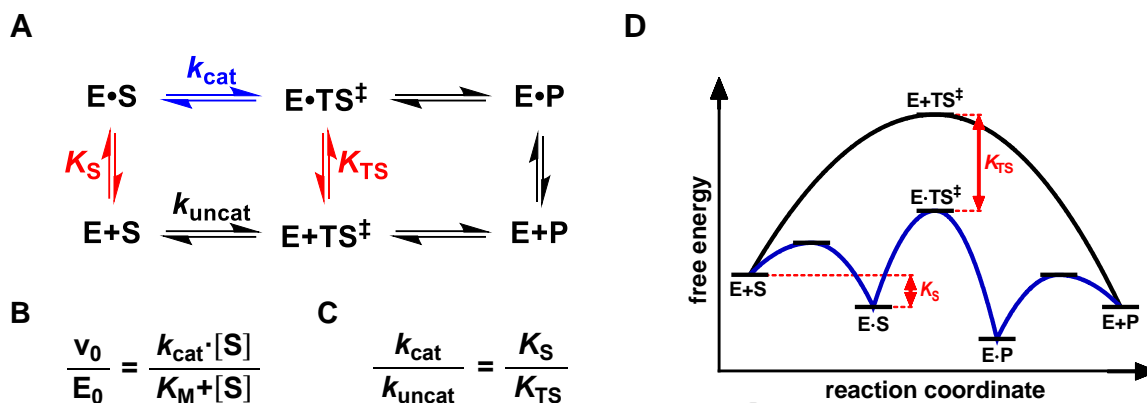
Figure 1: Enzyme catalysis. **A:** Free energy cycle involving the catalyzed (blue) and uncatalyzed (black) reaction, as well as the affinity of the enzyme to the substrate and transition state (K_S and K_{TS}, red). **B:** Michaelis-Menten kinetics derived from that cycle. **C:** The free energy cycle suggests that rate acceleration stems from selective transition state binding. **D:** Free energy diagram of the catalyzed (blue) and uncatalyzed (black) reaction. Selective stabilization of E·TS‡ over E·ES (red) leads to enzyme catalysis.

1.1.2 Chemical basis for enzyme catalysis

Various concepts have been proposed to account for the origins of enzymatic proficiency and selective transition state stabilization. These concepts include, but are not limited to, positioning and packing [11,14-16], electric fields and electrostatic catalysis [17,18], desolvation and medium effects [19-21], ground-state destabilization [9], as well as protein dynamics and conformational sampling [22-24]. These concepts are often highly debated, interpreted inconsistently, and challenging to separate from one another [25]. In order to analyze the achievements and limitations of artificial biocatalysts, the theories and controversies revolving around natural enzyme catalysis will be illustrated using the example of ketosteroid isomerase (KSI). The enzyme isomerizes 3-oxo-Δ^5 ketosteroids to their Δ^4-conjugated isomers via a dienolate intermediate with diffusion limited catalytic efficiency [26]. KSI employs an aspartate as a general base for proton shuffling and a network of H-bonding interactions to stabilize the oxyanionic intermediate and transition states (Figure 2).
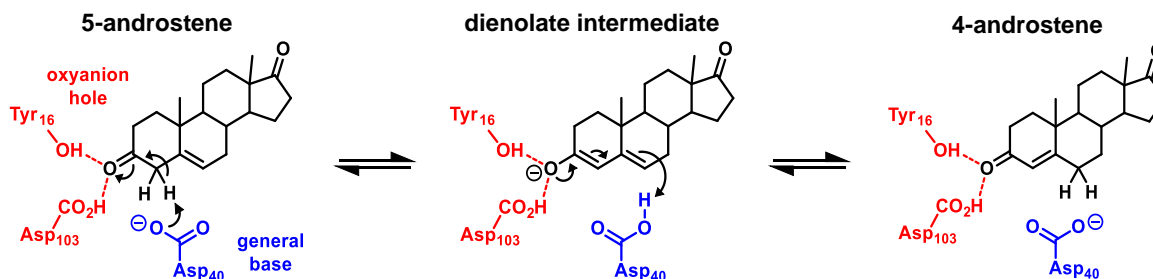
Figure 2: Mechanism of ketosteroid isomerase. The isomerization of Δ^5 ketosteroids to Δ^4 ketosteroids relies on Brønsted acid/base catalysis. A general base shuffles the proton via a dienolate intermediate. An oxyanion comprising of a complex hydrogen-bonding network stabilizes the negative charge [26].

Catalysis by positioning. Natural enzymes were refined over millions of years to precisely position their substrates and catalytic residues by tight packing and H-bonding networks to achieve geometric discrimination between the transition state and the ground state [11]. The effect of positioning on the activation energy ($\Delta G^{\ddagger}$) may be either of entropic ($\Delta S^{\ddagger}$) or enthalpic ($\Delta H^{\ddagger}$) nature. Counterintuitively, catalysis by positioning is dominated by $\Delta H^{\ddagger}$ as indicated in several model systems. For instance, increasing the steric constrain of an intramolecular ester replacement reaction primarily lead to a decrease in $\Delta H^{\ddagger}$ [27]. However, entropy and enthalpy are state functions and must be interpreted carefully since they report on countless individual microscopic contributions associated with the reactants, solvent and protein residues [28-30]. Nonetheless, enzymes are known to catalyze their reactions, with a few notable examples like the ribozyme [31], primarily by decreasing the activation enthalpy [32]. In the case of KSI, precise alignment of the general base, an oxyanion hole, and the substrate was estimated to significantly contribute to catalysis as indicated by mutagenesis studies, chemical rescue experiments, and the use of truncated substrates ([33], Figure 3).
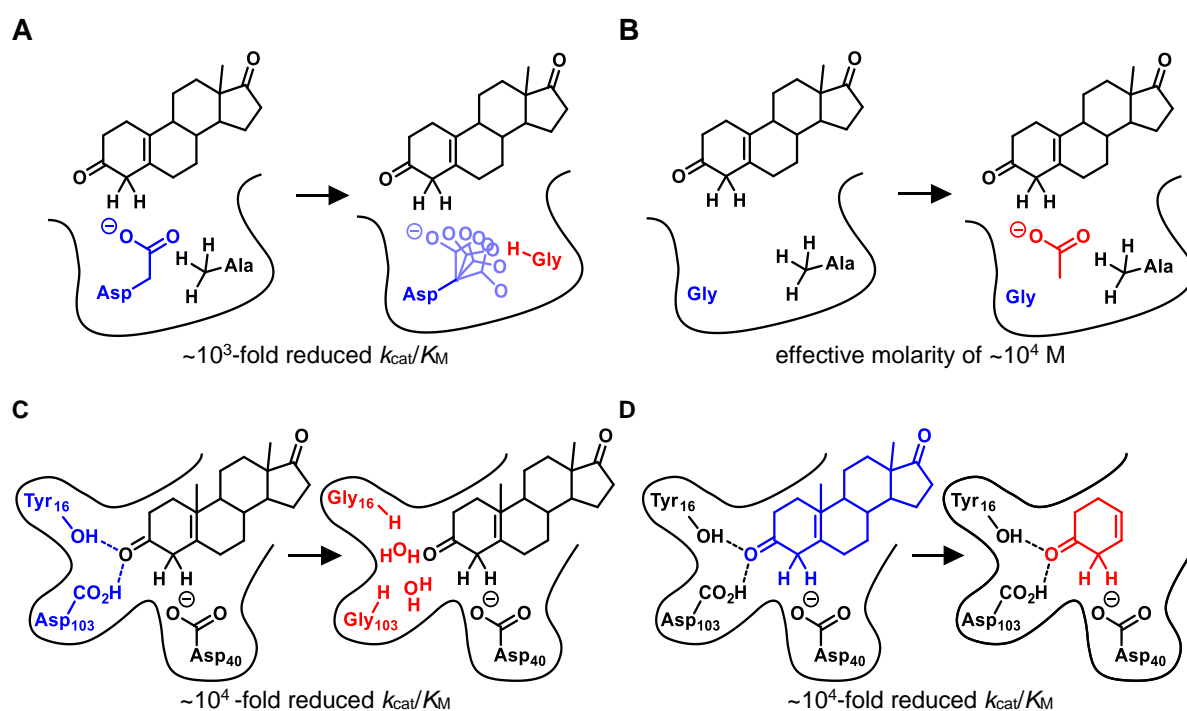
Figure 3: Catalysis by positioning. **A:** An Ala to Gly mutation (red) adjacent to the catalytic general base (blue) of KSI increases its flexibility leading to impaired positioning [34,35]. **B:** Chemical rescue with acetic acid (red) after knock-out of the general base with Gly (blue) partially restores activity [36]. **C:** Replacement of the oxyanion hole (blue) with badly positioned water (red) by mutagenesis decreases activity [37,38]. **D:** Impaired positioning of the substrate (blue) through removal of the non-reacting rings (red) impairs catalysis [39].

The precise positioning of the general base was suggested to contribute a factor of 10^3 to catalysis as indicated by double mutant cycle analyses which misplaced the general base but

conserved the overall structure of the active site [34,35]. Furthermore, chemical rescue by addition of acetic acid after active-site knock-out partially restored activity with an effective molarity for acetate of $6x10^3$ M [36]. A similar factor of $\sim10^4$-fold was observed for several mutations in which the precisely positioned hydrogen-bonding network in the oxyanion hole was replaced by water [37,38]. The influence of packing interactions on KSI's active site was probed with a truncated substrate in which non-reacting rings were removed. Although truncation leads to a $3x10^4$-fold reduction in catalytic efficiency, it was observed that k_{cat} is only decreased by a factor of 2 [39]. The observed decrease in substrate affinity thus suggests that binding interactions with the distal steroid rings provides approximately 5 kcal/mol to catalysis by KSI.

Building on the importance of reactant positioning, it was hypothesized that systems like KSI may exploit low-barrier hydrogen bonds to achieve significant rate accelerations [40-42]. Importantly, the energy of a low-barrier hydrogen bond strongly depends on precise matching of pK_A and distance, allowing selective geometric recognition of the transition state. Though a low-barrier hydrogen bond can be observed by NMR at the oxyanion hole of KSI [43], mutational studies that modulate either the pK_A [44] or the alignment of the oxyanion hole [43] revealed that the covalent character of the hydrogen bond did not yield any additional rate acceleration. In agreement with other systems, it is thus believed that low-barrier hydrogen bonds do not contribute more to catalysis than conventional hydrogen bonds [45].

Electrostatic catalysis and electric fields. Another important factor contributing to catalysis besides positioning of catalytic residues is electrostatic catalysis. Electrostatic catalysis includes the effect of all electrostatic interactions of an enzyme with the reactants to selectively stabilize the transition state. Catalysis by electrostatics, based on Markus theory for electron transfer reactions [46], suggests that enzymes are electrostatically preorganized in a way that reorganization of the environment during the reaction from ground to transition state is less energy demanding in the enzyme than in solution [17]. According to the Markus model, the activation energy ($\Delta G^{\ddagger}$) is determined by the free energy difference of the reaction (ΔG°) and the work required by the enzyme to adjust the electrostatic environment, termed the reorganization energy λ_0 (Figure 4).
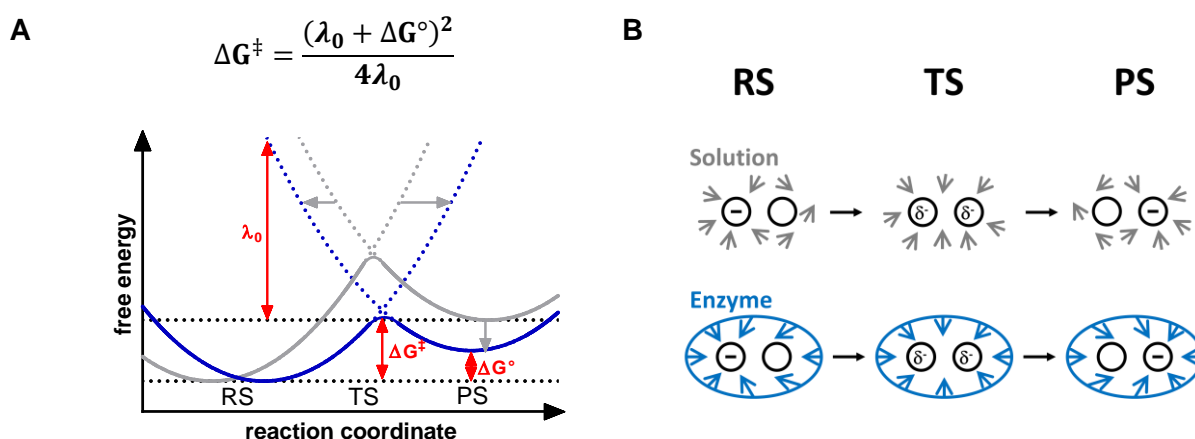
Figure 4: Electrostatic catalysis by preorganization. **(A)** Comparison of the catalyzed (blue) with the uncatalyzed (gray) reaction. The reaction proceeds from the reactant state (RS) over the transition state (TS) to the product state (PS). The activation energy ($\Delta G^{\ddagger}$) can be decreased either by reduction of the reorganization energy (λ_0) or free energy difference (ΔG°). **(B)** The reorganization energy is decreased by electrostatic preorganization of the enzyme as compared to the reference reaction in solution. Figure adapted from [47].

Importantly, computational models reliably predict rate accelerations of up to 10^7 to 10^{10} fold for electrostatic catalysis in various systems [17,48-50]. Differentiating electrostatic catalysis from catalysis by positioning is challenging, however, as electrostatic interactions with catalytic groups often critically contribute to the net electrostatic effect. The relative contribution of electrostatic catalysis [51,52] versus positioning [25,39,53,54] has thus been the subject of a vigorous ongoing debate.

While early quantification of electrostatic effects was limited to computational studies, a recently developed theoretical framework based on the Stark effect allows experimental analysis using vibrational spectroscopy. The Stark effect describes the dependence of an oscillator's vibrational frequency on its electrostatic environment. Frequency shifts of a vibrational probe can therefore be attributed to the sum of electrostatic interactions, the electric fields, at the probe. Importantly, the frequency changes observed for transition state analogs amenable to IR or Raman spectroscopy have indicated a correlation between activation energy and electric field in several systems ([18,55], Figure 5).

In the case of KSI, spectroscopic studies indicated that the total $10^{7.5}$-fold rate acceleration over the acetate-catalyzed background reaction in solution is primarily caused by a 10^5-fold acceleration due to electrostatic effects. Nevertheless, this conclusion has been questioned by others, who pointed out the challenge of separating electrostatic catalysis from individual effects like positioning [56] or desolvation of the general base [57]. Additional computational analyses [58,59] and experimental studies using non-natural amino acids [60] have since been performed that support the initial conclusion. While the ongoing debate revolves around the

magnitude of the contribution, the experimental evidences strongly indicate that electrostatic effects, including those of the overall protein scaffold, play a key role in catalysis.
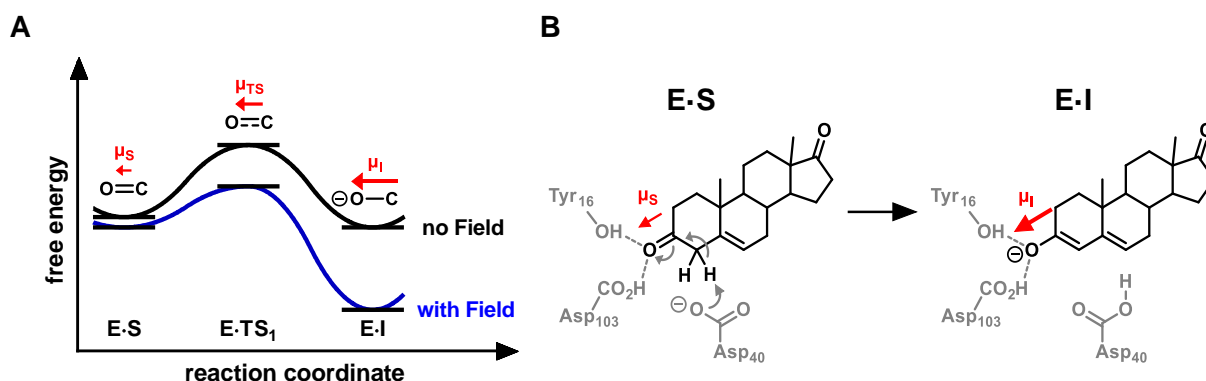
Figure 5: Electric fields and enzyme catalysis. **A:** Electric fields in enzymes accelerate reactions by stabilization of the ligand's dipole moment (blue vs. black curve). Selective stabilization can be achieved by alignment of the field to the ligand's dipole (red) which may change in intensity or direction during reaction. **B:** In KSI, the tetrahedral intermediate (**E·I**) has a larger dipole moment than the enzyme substrate complex (**E·S**). Figure adapted from [18].

Desolvation and medium effects.   Medium effects have been proposed to have a significant impact on catalysis due to the high sensitivity of many chemical reactions to the solvents in which they take place [61,62]. Biocatalytic medium effects can be ascribed to the low-dielectric environment of an enzyme's active site compared to aqueous solution [19-21]. By decreasing the dielectricity, electronic interactions are reinforced which may increase rate accelerations. Importantly, the spatially organized apolar environment of an enzyme's active site likely allows significantly larger catalytic contributions than bulk solvent.

While medium effects are actually included in electrostatic catalysis per definition, a decrease in the dielectric constant to strengthen ionic interactions and enforce electrostatic preorganization seems to be widespread in natural enzymes [63]. For instance, the active site of KSI becomes highly desolvated upon binding of a transition state analog [64]. No water is observed within 10 Å of the oxyanion hole. Computation suggested that this desolvation is key to electrostatic preorganization with an increase in bound water negatively affecting the reaction barrier [65]. The pK_a of Asp99 in the oxyanion hole appears to be shifted to ~9.5 as a consequence of desolvation [66,67] and studies using noncanonical amino acids revealed a linear correlation between the proton affinity of the oxyanion hole and activation energy [60]. The slope (ß = 1) of the experimental linear-free energy relationship between proton affinity and activation energy indicates a significant contribution of electrostatic catalysis through desolvation of the oxyanion hole in KSI. Importantly, Brønsted linear-free energy relationships,

as observed in KSI, can provide important information about catalysis by desolvation and medium effects.
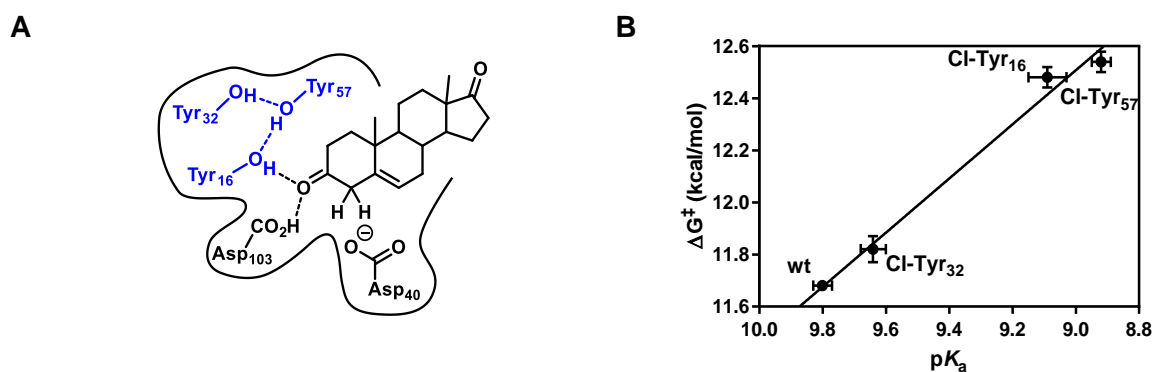
Figure 6: Desolvation and medium effects. **(A)** An intimate H-bonding network, composed of Tyr_{16}, Tyr_{32} and Tyr_{57} is involved in the oxyanion hole in KSI. **(B)** Increased electrostatic desolvation of the H-bonding network through the introduction of chloro-substituted tyrosine leads to an increase in apparent pK_a of the oxyanion hole, as well as a disfavored activation barrier. These Brønsted linear free-energy relationships can suggest medium effects arising from active site desolvation. Figure adapted from [60].

Ground state destabilization.   Similar to transition state stabilization, destabilization of the ground state can lead to lower barriers for a reaction on the protein. Ground state destabilization was believed to be a major driver of catalysis in the early years of enzymology [9], primarily based on observations in X-ray structures [68]. Nonetheless, destabilization of the enzyme-substrate complex can only result in an effect on k_{cat} (Figure 7).
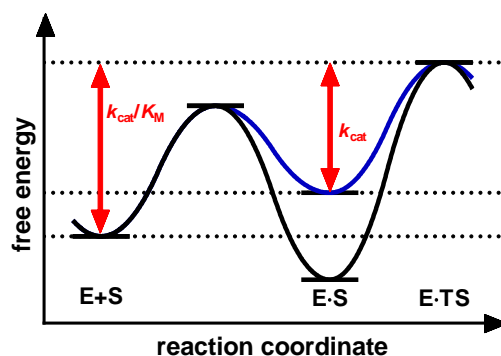
Figure 7: Ground state destabilization and enzyme catalysis. Destabilization of the ground state (ES, blue) decreases k_{cat}, but the apparent second order rate constant, k_{cat}/K_M, remains unaffected by destabilization of the ground state.

The Circe effect, proposed by Jencks, suggests that the free energy that is typically gained upon substrate binding can be partially used to destabilize the substrate's reactive groups [69]. Ground state destabilization has been observed for various systems, for instance through isotope effects [70-72]. However, the ability of enzymes to distort their substrates is likely limited [14]. As noted by Fersht, an enzyme cannot twist a double bound. But population of rare

conformers, like those observed in chorismate mutases, is not a problem [14]. Others challenged the catalytic relevance of ground state destabilization through the notion that the effect is typically accompanied by significant transition state stabilization [17]. In the case of KSI, binding studies with mutants in which the general base was knocked-out indicated that the base's negative charge destabilizes the hydrophobic substrate in the ground state, thus providing a modest contribution to catalysis [73].

Conformational sampling and catalysis by dynamics. Enzymes usually adjust their conformations during catalysis to more efficiently perform different tasks along the reaction coordinate like binding, turnover and release [74]. Importantly, networks of residues which display coupled motions can be identified in numerous systems by NMR spectroscopy [75], room-temperature X-ray crystallography [76], chemical H/D-exchange [77] as well as computation [78]. These coupled motions are often involved in catalytic turnover, may limit the overall reaction rate, and disruption of these motions typically slows down turnover [24,79]. Furthermore, statistical coupling analysis of mutations in phylogenetic models suggests coevolution of these networks which further supports the relevance of correlated movements for catalysis [80,81]. In KSI, active site closure through correlated movement of residues neighboring the catalytic general base were observed by molecular dynamics simulations (Figure 8, [65]).
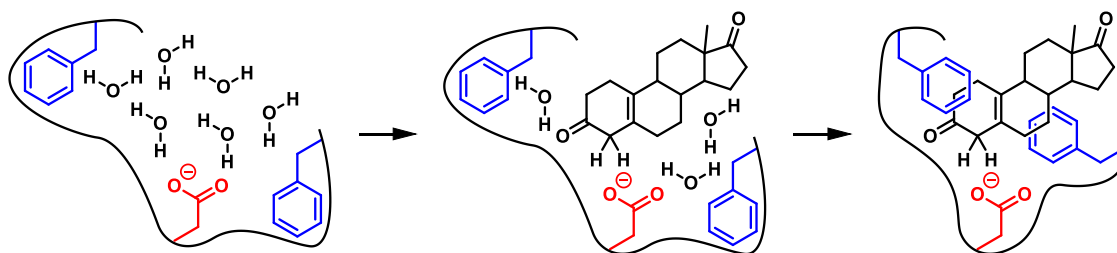
Figure 8: Conformational sampling in KSI. Binding of the substrate induces a conformational change of several first shell residues (blue). The conformational change is essential to displace water from the active site and leads to increased electrostatic preorganization. Figure adapted from [65].

A computational analysis of KSI suggested that a conformational change of the scaffold is essential to increase electrostatic preorganization for catalysis by displacing water from the active site. Thus, dynamic reorganization of the protein scaffold to selectively stabilize the transition state seems to be related to the electrostatic preorganization proposal and may be key for efficient exploitation of electrostatic catalysis [18,79].

With regard to protein motions, a catalytic role for dynamics beyond an increase in electrostatic preorganization has been postulated [82,83]. This controversial idea proposes

that either local vibrational modes (fs - ps timescale) or large-scale conformational sampling (ms - µs timescale) accelerate reactions by increased tunneling through dynamic sampling of short donor-acceptor distances [84] or by dissipating energy onto the reactant through non-equilibrium dynamics [85]. In light of debates sparked by imprecise definitions, dynamics are defined here as non-equilibrium barrier crossing effects, while catalysis describes the rate enhancement relative to the reference reaction in solution. In terms of tunneling, temperature independent [86-89] and unusually high primary kinetic isotope effects [90,91] observed for some natural enzymes support the proposed role for dynamics and suggest that enzymes dynamically sample short donor-acceptor distances, leading to more efficient tunneling. The catalytic relevance of tunneling through dynamic sampling of donor-acceptor distances has been questioned, however, since the same effects are also observed for the reference reactions in solution [92]. Even more controversially, local vibrations in the femtosecond to picosecond regime were proposed to significantly contribute to catalysis by coupling to the barrier crossing [85]. Computational models of dihydrofolate reductase, for instance, indicated that local vibrations may lead to more efficient sampling of a tunneling ready state [83]. The direct impact of such dynamics on catalysis is doubtful, however, and the observed isotope effects are likely attributable to increased preorganization of the enzyme [93,94]. Furthermore, a later analysis showed that the described modes were not coupled to the barrier crossing [95]. While it is commonly accepted that protein motions impact catalysis through an increase in electrostatic preorganization, it is unlikely that dynamic effects can directly influence the reaction rate by facilitating non-equilibrium barrier crossing.

1.2 Novel enzymes by directed evolution of *de novo* biocatalysts

The development of successful biocatalytic processes depends on the availability of suitable catalysts for the reaction of interest. *De novo* enzyme design aims to deliver made-to-order catalysts to complement the biocatalytic toolkit, particularly if natural catalysts are not available, for example for abiological reactions. Various design approaches have been explored to engineer novel enzymes based on the notion that selective transition state stabilization leads to catalysis. For example, early enzyme engineering attempts relied on the immune system to provide catalytic antibodies raised against transition-state analogs [12,96]. However, catalytic antibodies typically achieved modest efficiencies owing to the indirect selection for binding instead of catalysis and the difficulty of generating complex catalytic

devices. Also, the need for suitable and stable transition state analogs limited the success of this approach.

1.2.1   Computational enzyme design

The advent of computational design, pioneered by Baker, Mayo and Houk, allowed the *de novo* introduction of an active site for a reaction of interest into any protein scaffold [97]. During design, the ideal geometry of a theozyme, a hypothetical enzyme containing only the catalytic residues and substrate, is first determined using *ab initio* calculations. In order to introduce the theozyme into a scaffold protein, a poly-alanine pocket is generated *in silico* to provide space. The theozyme is subsequently fitted into that pocket by docking in different orientations. Finally, empty space is filled by optimizing the identities of the surrounding residues with a computational search algorithm leading to tight recognition of the transition state (Figure 9).
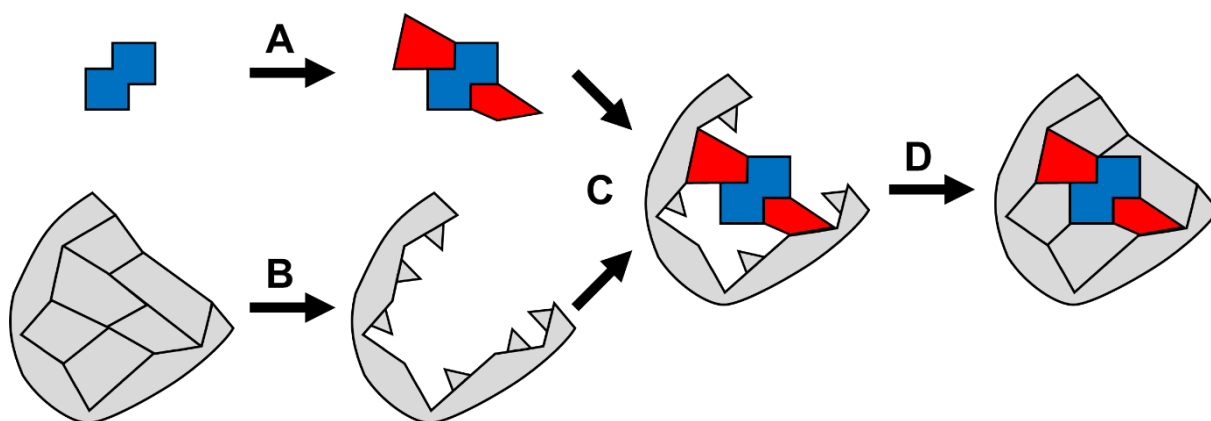
Figure 9: Computational enzyme design. **(A)** A theozyme containing only the reaction's transition state (blue) and the catalytic residues (red) can be designed using *ab initio* calculations. In order to introduce that theozyme into the scaffold (gray), **(B)** a poly-alanine pocked is formed *in silico* and **(C)** the theozyme is subsequently docked into that pocket. **(D)** Packing of the active site with additional mutations may subsequently provide active catalysts.

Design successes.   One of the first successful *de novo* computational designs was an esterase designed into the thioredoxin scaffold [98]. The designed esterase accelerated hydrolysis of *p*-nitrophenyl acetate 180-fold over the background reaction using a nucleophilic histidine on the protein surface. Later design approaches achieved rate accelerations of up to 10^5-fold over the background reaction for ester hydrolysis [98,99], aldol reactions [100-102], proton transfers [102-104], cycloadditions [105,106], and conjugated addition to enones [107] (Figure 9). While the flexibility of the design approach allowed targeting virtually any scaffold, complex reactions involving several transition states or multiple charged residues remain challenging [108,109].
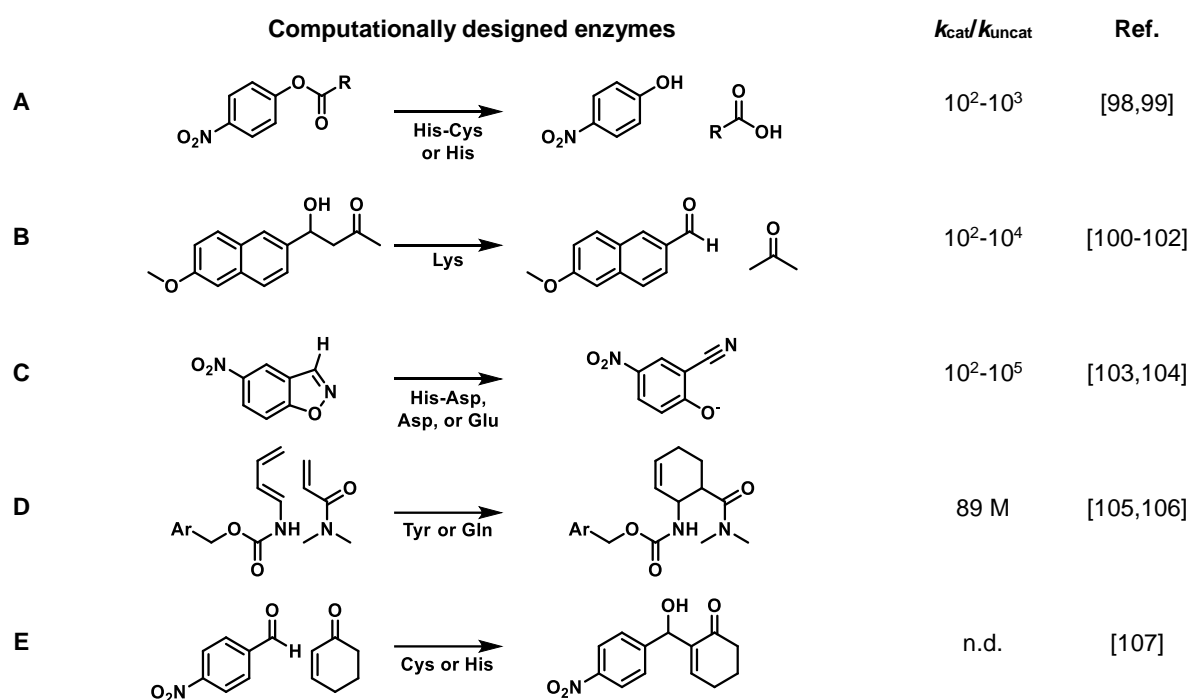
Figure 10: Rate accelerations attained by computationally designed enzymes for **A**: esterases, **B**: retro-aldolases, **C**: Kemp eliminases, **D**: Diels-Alderases and **E**: enzymes accelerating the Morita–Baylis–Hillman reaction. Catalytic residues are stated under the arrows. n.d. not determined. Figure adapted from [97].

Catalytic hypotheses. Computational design typically relies on theozymes that employed covalent catalysis, acid-base catalysis, H-bonding interactions, and packing through π–stacking interactions (Figure 4). For example, esterases designed for hydrolytic cleavage of *p*-nitrophenyl acetate employ covalent catalysis by a catalytic nucleophile (Ser, Cys or His) to attack the substrate and displace the alcohol leaving group [98,99]. Enzymes accelerating the Morita–Baylis–Hillman reaction employed covalent catalysis by an enzyme nucleophile (Cys or His) to activate the substrate enone via transient conjugate addition [107]. Similarly, retro-aldolases, which catalyze cleavage of aldol-substrates into an aldehyde and ketone, rely on covalent imine catalysis by Lys. Retro-aldolases activate the substrate's aldol-carbonyl via Schiff-base formation in a multistep reaction mechanism [100-102]. Instead of relying on covalent catalysis, the Kemp elimination [62,110], initiated by C-H proton abstraction from benzisoxazoles, has been accelerated by introduction of a general base (Glu, Asp or His) into a suitable binding pocket [111]. Finally, Diels-Alderases which catalyze [4+2]-cycloadditions have been designed that rely on activation of the diene and dienophile through H-bonding interactions and packing the two substrates in close contact with each other [105,106]. Several theozymes use additional H-bonding interactions to position the catalytic residues [99,100,104]. Furthermore, interactions that stabilize leaving groups either directly [99,103,104] or by a positioned water molecule [100] are also often included. If applicable,

aromatic residues are added to the theozymes to increase packing through π-stacking interactions with aromatic groups of the substrate [103,104,107].
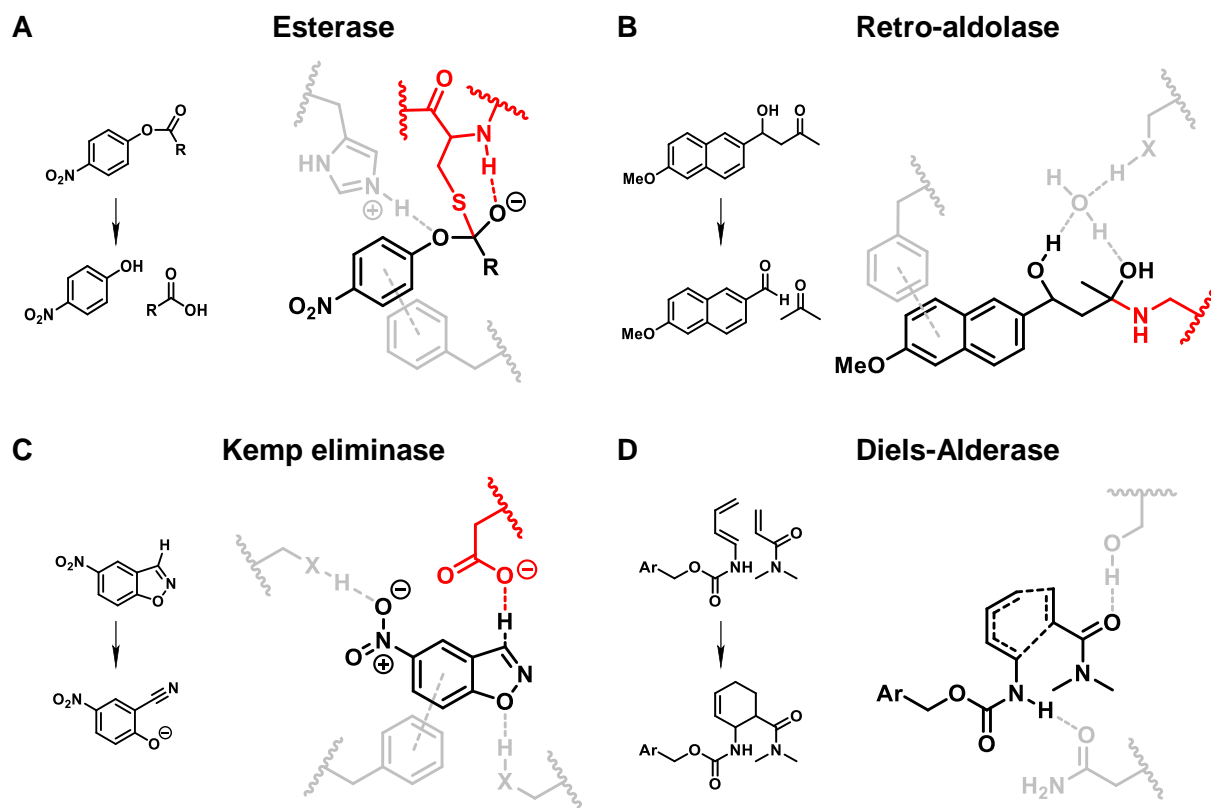
Figure 11: Theozymes employed during computational design comprise the transition states or reaction intermediates (black) and catalytic residues (red). Secondary interactions (gray) are often introduced to activate a leaving group, position catalytic residues or bind the substrate via π-stacking. **A:** Ester hydrolysis is catalyzed via nucleophilic attack to form an acyl-enzyme intermediate. **B:** Retro-aldolases have been designed to cleave aldol substrates via Schiff-base formation with the substrate carbonyl. **C:** Kemp eliminases that catalyze C-H deprotonation of benzisoxazoles have been designed using a general base. **D:** Diels-Alderases that catalyze the depicted [4+2]-cycloaddition activate their substrates through H-bonding interactions and additionally accelerate the reaction by bringing the two substrates in close contact to each other.

Design limitations. Despite the computational design of active enzymes with turnover numbers up to $3\,s^{-1}$ [103], the designs typically suffers from low success rates and limited activities [98-107]. If active, the catalytic efficiencies of designed enzymes often do not exceed those of natural enzymes that display accidental activities for the desired transformation [112-115]. This lack of efficiency can be ascribed, in part, to the inaccuracy of the design. For instance, X-ray structure analysis revealed that active sites are often not properly preorganized due to the flexibility of designed catalytic residues which may adopt multiple conformations in the apo state [103,116]. Also, the observed geometries of protein structures bound to transition state analogs usually derivate significantly from the energetic optimum determined by *ab inito* calculations [105,117]. As a consequence, computational designs often do not bind transition state analogs in the intended conformations [103,104]. Lastly,

underpacking of the active sites [101] has been shown to lead to penetration of water which impairs activity in some cases [117].

In order to improve accuracy, later design efforts aimed to introduce extended networks of polar residues and to increase the packing of the active sites. While additional H-bonding residues [103] and water molecules [116] were sometimes placed accurately, designed catalytic dyads and water networks often did not form as designed [99,100]. In one case, the designed H-bonding residues actually harmed activity by the unintended formation of a salt bridge with the catalytic general base [104]. The issue of underpacking was addressed with an additional design step in the absence of a ligand to improve the integrity of the active site in the apo state [101]. In other approaches, issues arising from solvent exposure have been solved by moving the active site to a position deeper in the scaffold protein [103] or by addition of a lid to the active site [106]. Nonetheless, significant improvement in positioning and packing with the sub-Ångstrom precision required for efficient transition-state stabilization has remained problematic, even for the most successful designs.

1.2.2   Directed evolution of computational designs

Despite impressive advancements in computational design, *de novo* engineered biocatalysts typically suffer from low proficiencies, which are orders of magnitude below those of natural enzymes [118]. Nevertheless, computationally designed biocatalysts are highly amenable to directed evolution. During directed evolution, iterative rounds of mutagenesis and screening are performed to gradually augment enzymatic activity. This simple algorithm, which efficiently mimics natural Darwinian evolution on a laboratory timescale, has been employed to optimize several computationally designed enzymes. In this way, the activities of *de novo* enzymes have been improved between two and six orders of magnitude (Table 1).

The combination of computational design and directed evolution has led to *de novo* biocatalysts with enzyme-like activity [4,5]. For instance, Kemp eliminase HG3.17 was evolved to a k_{cat} of 700 s^{-1}, which corresponds to a 6×10^8-fold rate acceleration over the uncatalyzed reaction. An even higher rate acceleration of 7×10^9 fold was achieved for the RA95 retro-aldolase after 19 rounds of evolution. Despite the significant effort invested in the optimization of these artificial biocatalysts, no catalyst has yet reached catalytic perfection. In-depth analysis of the evolved catalysts and their evolutionary trajectories is thus required to shed light on the factors that limit catalysis and were missed during design.

Table 1: Rate acceleration of selected designed and evolved enzymes					
Enzyme	Design	k_{cat}/k_{uncat}	Evolved [a]	k_{cat}/k_{uncat}	Ref.
Kemp eliminases	HG3	3×10^6	HG3.17	6×10^8	[97,103]
	KE07	2×10^4	KE07.7	1×10^6	[104,119]
	KE59	n.d.	KE59.13	8×10^6	[104,120]
	KE70	1×10^5	KE70.8	7×10^6	[104,121]
Retro-aldolases	RA95	5×10^3	RA95-8F [b]	7×10^9	[5,101,122]
Diels-Alderases	DA_20	4 M	CE20 [c]	500 M	[105,106,123]

[a] The index refers to the number of rounds of evolution. [b] RA95.5-8F was evolved over 19 rounds of directed evolution. [c] The activity of DA_20 was refined by active-site optimization and redesign of a lid domain in addition to 8 rounds of directed evolution resulting in CE20.

1.2.3 Origins of catalysis in *de novo* enzymes

Efforts to dissect the origins of catalysis in *de novo* engineered enzymes has primarily focused on base-catalyzed Kemp eliminases and retro-aldolases that employ amine catalysis. Several designs for both reactions have been subjected to directed evolution and intensive computational and experimental characterization facilitated detailed understanding of these systems. Importantly, the basal level of catalysis that is typical for *de novo* engineered systems may significantly benefit from inherently small effects that do not play an important role in natural enzymes. Identification of the key contributions to catalysis may allow engineering of more efficient enzymes in the future.

Positioning and packing.   Given the lack of precision during computational design, the local environment of the active sites often requires fine-tuning during directed evolution. Improved positioning coincided with tighter packing as well as with more extensive H-bond networks in numerous evolutionary trajectories [4,5,119,121,124,125]. During evolution of Kemp eliminase HG3.17 [4], the orientation of the substrate relative to the general base was gradually refined through better packing and positioning (Figure 12). While two ambiguous binding modes were observed in the original design, a unique conformation was selected during evolution. Evolution also introduced a potential oxyanion stabilizing H-bonding network. Computational empirical-valance bond studies indicate that the position of the key residue in this network, Gln50, was gradually refined over the course of evolution [125]. It should be noted that this interpretation was challenged by a hybrid quantum mechanics/molecular mechanics computational study in which Gln50 flipped out of the active site of HG3.17 during the

simulation of the ground state [126]. However, the calculation did not include Gln50 in the quantum mechanical region which may have biased its results.
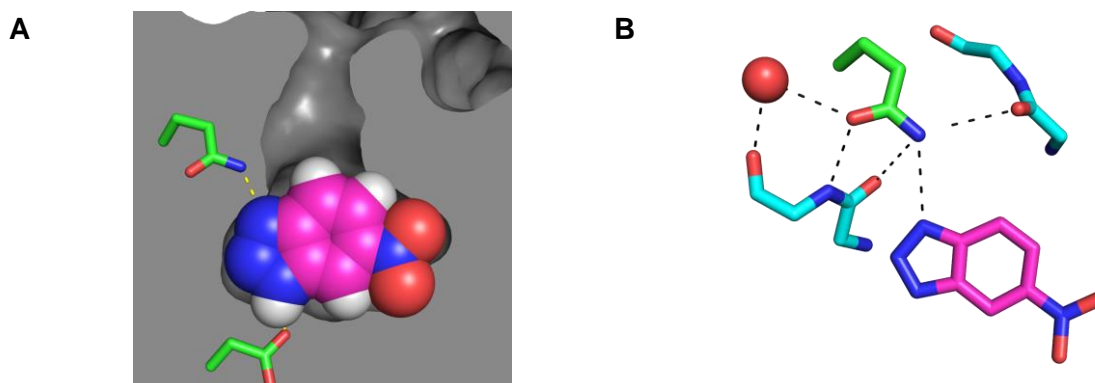
Figure 12: Improved positioning and packing during evolution of HG3.17 (PDB: 4BS0, [4]). **(A)** The inhibitor (magenta) is packed in a snug pocked and Asp127 and Gln50 (green) are nicely positioned for catalysis in HG3.17. **(B)** The evolved Gln50 (green) is positioned by an intimate H-bonding network with backbone amides (cyan) and a positioned water molecule.

Improvements in packing and H-bonding are not unique to the evolution of HG3.17. In Kemp eliminase KE07, where activity was impaired by the unintended formation of a salt bridge between a designed H-bond donor and the general base, evolution increased active site preorganization and prevented salt-bridge formation by incorporating an H-bonding residue into an extended hydrogen-bonding network [119]. Positional refinement through H-bonding networks also significantly improved activity during evolution of retro-aldolase RA95 [5], where catalysis of the designed lysine became supported by an elaborate catalytic tetrad. In an attempt to isolate the contribution of positioning in another designed retro-aldolase, RA61, activity was assessed with the normal substrate and a truncated analog lacking a naphthyl residue. Kinetic analysis indicated that positioning through packing interactions with the aromatic moiety contributed ~500-fold to the rate acceleration [124]. The relevance of positioning and packing for catalysis is also well pronounced in the engineered Diels-Alderase CE20 [105,106,123]. Occlusion of its solvent exposed active site by addition of a computationally designed lid domain, in combination with directed evolution, afforded a catalyst that efficiently packed its substrates to accelerate a [4+2]-cycloaddition. In conclusion, catalysis by positioning, which was estimated to contribute up to 10^4-fold in natural enzymes like KSI, seems to play a key role in evolved computational designs. This notion is in good agreement with the almost perfectly aligned active sites of many evolved catalysts [4,5,123].

Electrostatic catalysis and electric fields. While structural analysis of designed and evolved catalysts has provided strong evidence for catalysis by positioning, fewer studies are available regarding the impact of electrostatic catalysis. Nonetheless, a computational

analysis of the active site electric fields in Kemp eliminases KE07, KE70, and their evolutionary descendants suggested a major contribution to electrostatic catalysis from the catalytic base [127] (Figure 13).
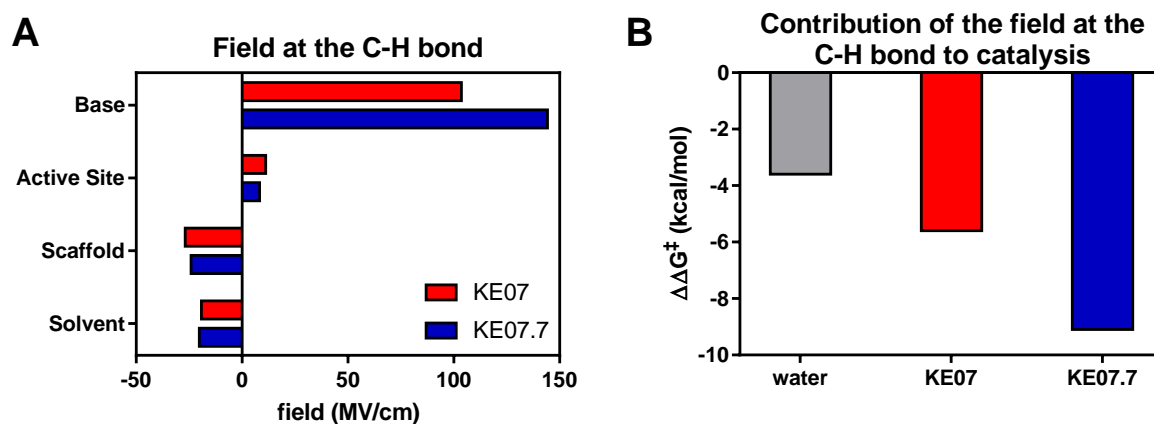
Figure 13: Electrostatic catalysis in *de novo* enzymes. **(A)** The electric field at the C-H bond in Kemp eliminase KE07 (red) is primary determined by the catalytic base, while the scaffolds and solvent field counteract catalysis. The field of the base is improved in the most evolved catalyst KE07.7 (blue) while the scaffold's field barely changed. **(B)** The effective field at the C-H bond was increased during evolution, leading to an increased rate acceleration. The impact of the solvent field on the catalyzed reaction (gray) is added as a reference. Data extracted from [127].

Remarkably, the overall electric field of the scaffold protein actually disfavored catalysis in both systems. Optimization of the enzyme's electric field on the C-H bond was furthermore almost entirely caused by changes in first shell interactions. Since the field of the protein scaffold comprises numerous interactions, each with a small contribution, optimization of the scaffold may be challenging for directed evolution. It was thus suggested that exploitation of the protein scaffold's electrostatics in the early stages of computational design may lead to significantly enhanced activities [127].

Desolvation and medium effects. Computationally designed enzymes typically targeted rather hydrophobic substrates. Collateral medium effects are thus often introduced through the design of catalytic residues in hydrophobic binding pockets. For catalysis of the Kemp elimination, Brønsted slopes of ~0.7 have been determined, indicating a sensitivity of the reaction to medium effects [110]. Analysis of a synthetic Kemp elimination catalyst further indicated that medium effects may contribute up to 10^5–fold through specific interactions in the active sites of enzymes [20]. Refinement during evolution, however, aims to generate the highest rate enhancement at the assay pH value and, as a consequence, the pK_a of catalytic residues that need to be deprotonated usually settle around five to six, reflecting a trade-off between intrinsic reactivity and the fraction of catalytically active species [4,119-121]. In the

case of Kemp eliminases employing catalytic carboxylic acids, the pK_a of the general base shifts from 4 to 6-7 which may afford a rate acceleration of up to 10^2-fold. Retro-aldolases, on the other hand, employ amine-catalysis and need to decrease the pK_a of the catalytic Lys to increase the amount of the catalytically competent deprotonated species. For example, directed evolution of RA95 gradually decreased the apparent pK_a of the catalysts from 10 to 6.2 [5,101,122]. Nonetheless, an analysis of a series of amines with varying pK_a's suggested that fine-tuning of the pK_a may increase the rate acceleration by up to tenfold at neutral pH [124].

Medium effects, which do not require a high degree of preorganization, seem to be often responsible for the promiscuous activities of natural proteins for designed reactions [112-115] by the random exposure of potentially catalytic residues in hydrophobic environments. Medium effects may thus be a key driver to achieve a basal level of activity in many computational designs. Finally, the 10 to 10^2-fold impact on catalysis due to desolvation in combination with the 10^3 to 10^4-fold gain due to positioning seems to be sufficient to explain most of the rate accelerations observed for *de novo* engineered catalysts.

Ground state destabilization. In the case of the designed KE07, computational studies suggest a significant contribution of ground-state destabilization to catalysis [128]. Ground state destabilization in the case of the designed KE07 was not achieved by imposing strain on the substrate, but by removal of an H-bonding interaction with the catalytic base. The catalytic base was thus electrostatically desolvated which is expressed by an elevated pK_a [125]. Importantly, ground state destabilization, as observed in KE07, typically represents only a minor contribution to catalysis in natural enzymes [17]. Furthermore, a recent computational analysis of HG3 and its evolutionary descendants indicates that the catalysts primarily improved through transition state stabilization during evolution [125]. However, this analysis focused primarily on the changes during evolution. It may well be that ground state destabilization in HG3, with a general base pK_A of 6.0, was already maximized in the design.

Dynamics and conformational sampling. Conformational sampling is of importance in natural systems such as KSI to facilitate several microscopic steps along the reaction pathway. For retro-aldolases, in which the catalytic cycle involves various intermediates, multistate design has been attempted with limited success [129]. The designed catalysts likely displayed low activity due to the limitations of the computational design approach which typically cannot precisely organize the active site. Also, the impact of each state of the retro-aldol reaction on

the overall rate cannot be predicted in a straightforward manner and may vary from system to system [130].

Computational design therefore typically focuses on a single active site configuration for selective stabilization of one particular state, sometimes relying on intuition with regard to which state may be most suitable [101]. Attempts to increase active site preorganization have been performed, for instance, by redesign of loops [121] and lids [123]. Nevertheless, a computational analysis of Kemp eliminases KE07 and KE70 indicates that the enzyme dynamics, expressed by the side chain entropy, were refined during evolution (Figure 14). Evolution afforded catalytic networks of correlated movements that seem to selectively stabilize the transition state. While the designed proteins show highly correlated movements in the ground state, evolution led to the emergence of correlated networks in the transition state. These observations were attributed to overdesign of the ground state relative to the transition state. Thus, correlated movements of networks may be an important design feature or an additional target for directed evolution [30].
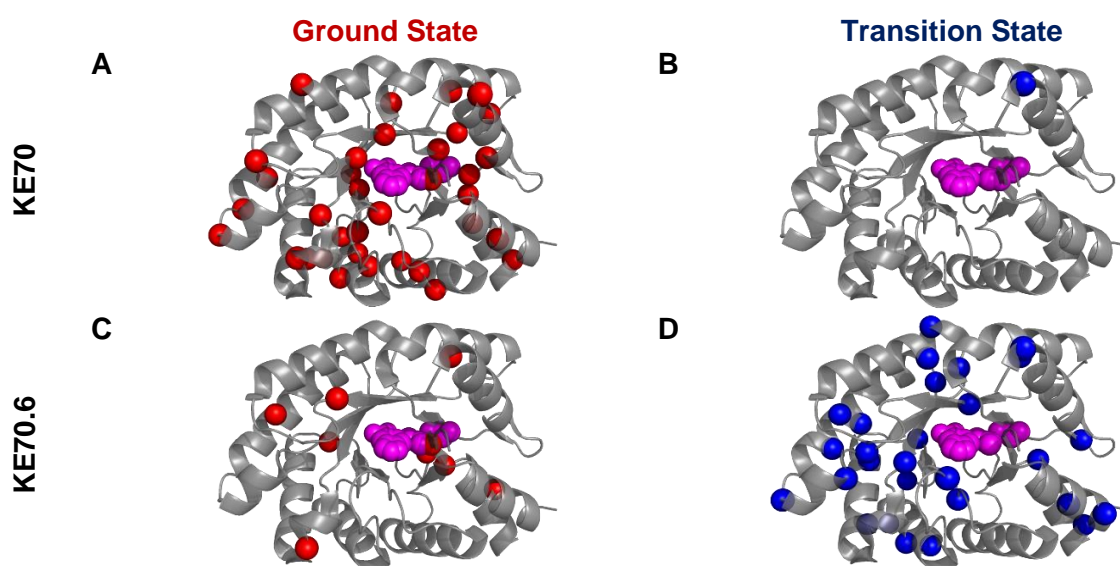
Figure 14: Correlated movements in KE70 and KE70.6. Directed evolution changed the correlated movement patterns in (**A+C**) the ground state (red) towards more correlated movement in the transition state (**B+D**) of KE07 (top) and KE07.7 (bottom). Adapted from [127].

A similar analysis of the solvent entropy in the aforementioned systems indicated that evolution also reduced the entropic cost of displacing water from the active site upon substrate binding [29]. This was either achieved by removal of low entropy waters or by increasing the net entropy of the bound water molecules. Loss of active site water was also observed during evolution in other systems, underscoring the potential relevance for catalysis [120]. Nevertheless, while entropic contributions and conformational changes may be exploited for

catalysis, a major current challenge for design in terms of dynamics is to limit the conformational diversity of the designs to achieve efficient transition state stabilization [131].

1.2.4 Future challenges

The dawn of computational enzyme design and its combination with directed evolution has enabled engineering of novel enzymes. A handful of these *de novo* catalysts have achieved enzyme-like rate accelerations [4,5]. Nevertheless, an obvious limitation of computational enzyme design is its lack of accuracy. In the light of recent successes in the field of protein design [132], however, it is evident that computational algorithms are being constantly improved. The increasing reliability of these methods has made it possible to expand from the design of simple folds [133] to complex architectures like the TIM barrel [134]. It is thus likely that technological advancements will have a similar benefit on the computational design of enzymes in the future.

Although the active sites of successfully designed enzymes often seem ideal after refinement by evolution, their apparent second order rate constants are still orders of magnitude below catalytic perfection. The theozyme approach, in its current form, cannot generate proficient enzymes from scratch. One of the future key challenges will be to find the missing links that enable reliable engineering of highly efficient enzymes as opposed to "catalytic proteins" with limited activity [118]. Analyses of designed and evolved enzymes provide hints regarding these missing links: Electrostatic preorganization should be improved, at the local level of catalytic residues through complex H-bonding networks, as well as on a global level by employing the inherent electrostatics of the enzyme's scaffold for catalysis. The reactivity of catalytic residues should furthermore be tightly controlled, for instance through intentional use of medium effects. Active-site dynamics should be harnessed during design to improve active site integrity in the apo state as well as conformational changes along the catalytic trajectory. Finally, entropy and coupled motions in the scaffold should be considered as a means to increase activity. In conclusion, a more global application of the theozyme approach may significantly improve enzyme design. However, a more detailed analysis of the individual effects is clearly needed to identify key contributions to catalysis in *de novo* enzymes.

1.3 Aims of this thesis

De novo engineered enzymes are far from catalytic perfection and the efficiencies of computationally designed enzymes are typically in the range of the promiscuous activities of natural proteins [112-115]. Nonetheless, the designs are highly amenable to directed evolution and can approach catalytic efficiencies of natural enzymes in the best cases [5,97,135]. One of the most proficient *de novo* enzymes, Kemp eliminase HG3.17, achieved an enzyme-like 6×10^8-fold rate acceleration over background after directed evolution. Nonetheless, its apparent second order rate constant (2×10^5 M^{-1}s^{-1}) is still orders of magnitude below that of natural enzymes catalyzing comparable reactions like KSI (3×10^8 M^{-1}s^{-1} [136]) or triosephosphate isomerase (4×10^8 M^{-1}s^{-1}, starting from glyceraldehyde-3-phosphate [137], Figure 15). Designed Kemp eliminases thus lack approximately three to four orders of magnitude in transition state affinity to achieve catalytic perfection.
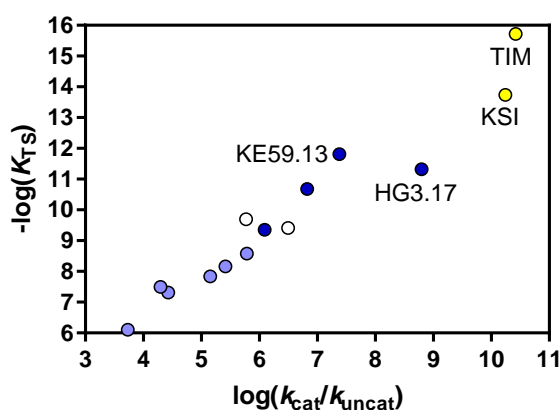
Figure 15: Natural versus designer enzymes for catalysis of proton abstraction from carbon. Designer catalysts still lack approximately 10^3-10^4-fold in transition state affinity to achieve catalytic perfection. Rate acceleration versus transition state affinities for catalytic antibodies (white), designed (light blue) and evolved (dark blue) enzymes, as well as KSI and TIM (yellow). All proteins are listed on page 114.

This thesis aims at improving the *de novo* enzyme engineering approach by examination of the origins of catalysis in artificial enzymes, generation of novel catalysts with enzyme-like activity through directed evolution, and development of novel screening methods and alternative starting points for evolution. In Chapter 2, in-depth biophysical characterization of the Kemp eliminase HG3.17 is described. In light of the proposed concepts for catalysis in natural enzymes, we focused on: (1) The impact of positioning of the general base and evolved oxyanion hole; (2) an experimental analysis on electrostatic catalysis using novel vibrational probes; and (3) the conformational landscape of HG3.17 and its relevance for catalysis (Figure 16). Based on these studies, factors that contribute to the efficiency of the catalysts were identified and suggestions formulated to improve future design endeavors.
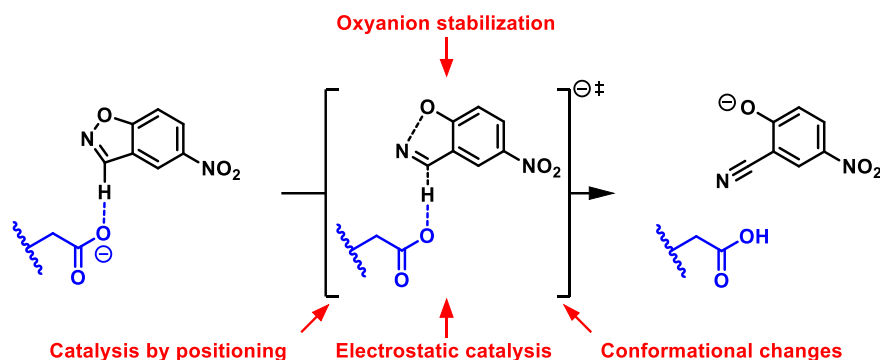
Figure 16: Catalysis by Kemp eliminase HG3.17. The enzyme uses a general base (blue) to deprotonate a benzisoxazole substrate. The role of oxyanion stabilization, positioning, electrostatics and conformational changes in HG3.17 are analyzed in Chapter 2.

HG3.17 displays exceptional activity for an artificial catalyst [5,118]. Research in Chapter 3 tests the feasibility of repeating this success through evolution of the *de novo* designed Kemp eliminase 1A53-2. Analysis of catalysis in its most evolved descendant broadens our knowledge of catalysis in these systems. Furthermore, the laboratory evolution performed in this thesis is compared to natural evolution to support a hypothesis concerning the primordial ancestors of modern enzymes.

The remainder of the thesis focuses on improving the overall engineering process. Chapter 4 attempts to capitalize on ultra-high throughput screening methods that have led to improvements beyond the reach of conventional medium-throughput methods in many cases [5,138]. Several fluorogenic systems suited for ultra-high throughput assays of Kemp eliminase activity are described. Application of such systems to the evolution of Kemp eliminases may provide even more efficient catalysts in the future. Chapter 5 describes work to broaden the catalytic scope of designed enzymes by repurposing the catalytic devices in computational designs. In this context, the promiscuous activities of the computationally designed Kemp eliminase 1A53-2 towards biotechnologically relevant synthesis of chiral cyclic monoterpenes are probed and targeted by evolution. A side project presented in the Appendix aimed to repurpose an existing metal-binding site through computational design. These experiments, performed in collaboration with the Houk group at UCLA, centered on a metal-dependent Diels-Alder reaction and had the goal of combining efficient metal-ion catalysis with the selectivities of a protein binding pocket.

By dissection of highly efficient artificial enzymes, generation of new enzyme-like catalysts, exploration of improved evolutionary methods and exploitation of existent catalysts, this thesis sheds light on catalysis in these artificial systems and will hopefully facilitate future engineering of better enzymes.

2 Dissecting catalysis in the computationally designed and evolved Kemp eliminase HG3.17

Enzyme-like catalysts have been engineered by combination of design and evolution [4,5]. However, computational design is still limited and generally requires time- and cost-intensive reengineering of the new catalysts in the laboratory to generate substantial rate accelerations [97,139]. Dissection of the engineered enzymes may help uncover problems and opportunities for both design and evolution of better catalysts in the future. In this regard, the Kemp elimination is a valuable model system comprising only a one-step C-H proton abstraction, which readily allows interpretation of experimental and computational data [62].

The Kemp eliminase with the highest rate acceleration achieved to date, HG3.17, was evolved over 17 rounds of evolution from the computationally designed enzyme HG3 [4,103]. Evolution afforded a precisely organized and compact active site with notably short catalytic interactions. Furthermore, evolution introduced an oxyanion-stabilizing residue, Gln50, which is positioned by an intimate hydrogen-bonding network. The catalyst thus has a low reorganization energy and shows significant transition state stabilization [125,126] providing an enzyme-like $6x10^8$–fold rate acceleration.

2.1 Catalysis by positioning of a general base and an oxyanion hole in an evolved Kemp eliminase

Significant contributions to catalysis in natural enzymes stem from the precise positioning of catalytic residues [11,14-16]. For instance, precise positioning of a general base has been estimated to contribute as much as 10^3 fold to the rate acceleration achieved by Δ-ketosteroid isomerase (KSI, [34,35]). Often, charges that build up in the transition state must also be stabilized via complementary catalytic groups of the enzyme. Many biocatalysts, including hydrolytic enzymes [140] and isomerases [33,38], promote reactions that involve formation of a transient oxyanion in their catalytic mechanism. These high-energy oxyanions are frequently

stabilized by oxyanion holes that provide preorganized hydrogen bonds and stabilizing electrostatic interactions. In case of KSI, replacement of the residues forming the oxyanion hole with water reduces catalytic efficiency by 10^4 fold [37,38].

Oxyanion stabilizing groups have also been introduced into small molecule catalysts [141] as well as designer enzymes [103,104] and catalytic antibodies [142] to promote the Kemp elimination (Figure 17). Computation suggested that stabilizing the incipient phenoxide with groups such as formic acid or water decreases the activation energy for the Kemp elimination by as much as 10 kcal/mol [111,128]. Hence, the theozymes for the Kemp elimination typically include an oxyanion stabilizing residue. In Kemp eliminase HG3, Ser265 was designed to stabilize the oxyanion. However, design failed to place the substrate in a unique orientation to capitalize on this interaction, as indicated by a crystal structure of the enzyme in complex with the transition state analog 6-nitrobenzotriazole **1**. During evolution, a different substrate binding mode than initially designed was selected, making Ser265 irrelevant for oxyanion stabilization. Nevertheless, a new residue, Gln50, appeared in the active site over the course of evolution. The nitrogen of its amide side chain is positioned by a hydrogen-bonding network to form a 3.10 Å interaction with N3 of **1**, which corresponds to the oxyanionic leaving group in the benzisoxazole substrate (Figure 17).
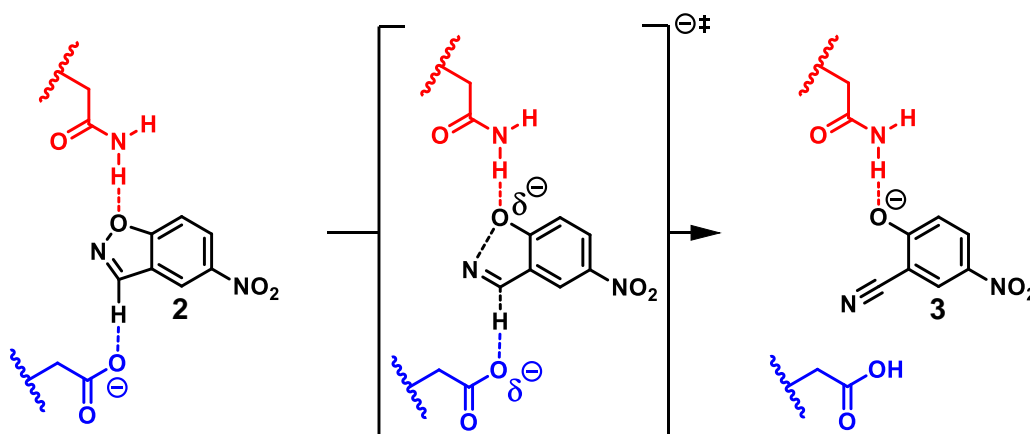
Figure 17: Catalysis of the Kemp elimination by effectively positioned catalytic resides. A catalytic base (Asp127, blue) promotes proton abstraction from the benzisoxazole-substrate (black) of Kemp eliminase HG3.17, whereas Gln50 (red), which was discovered by evolution, is positioned to stabilize the incipient oxyanion.

Despite the precise placement of Gln50 for oxyanion stabilization, site-directed mutagenesis performed by Kries revealed that Gln50Met and Gln50Phe were the next most active variants with a substitution at position 50, even though Met and Phe cannot stabilize the oxyanion by H-bonding [143]. To probe the effects of the Gln50 residue on catalysis, Kries and Bloch determined the steady-state parameters for the three most active (Met, Phe, His) and

three additional (Ser, Ala, Lys) variants with 5-nitrobenzisoxazole **2** and constructed Brønsted Plots using differently substituted substrates ([144,145], Table 2).

Table 2: Kinetic parameters for cleavage of 5-nitrobenzisoxazole **2** by Gln50 variants of HG3.17 [a]			
	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1}s^{-1})	Brønsted slope
HG3.17	470±30	2.06x10^5	-1.49±0.07
Gln50Met	320±50	5.55x10^4	-1.73±0.02
Gln50Phe	70±20	2.08x10^4	-1.72±0.02
Gln50Ser	80±20	1.39x10^4	-1.61±0.02
Gln50His	n.d.	1.33x10^4	-1.69±0.03
Gln50Ala	36±6	6.9x10^3	-1.64±0.03
Gln50Lys	n.d.	5.38x10^3	-1.64±0.11

[a] Kinetic parameters were determined at 20 °C. Adapted from [145].

Changes in the slope (ß) of the linear free-energy relationship between log(k_{cat}/K_M) and the leaving group pK_a (pK_{a_lg}) provided insights into oxyanion stabilization in Kemp eliminase HG3.17. Substrates with less acidic products show less intrinsic oxyanion stabilization and require more stabilization by the scaffold protein. The evolved HG3.17 (ß = -1.49) has a similar slope to that of catalytic antibody 34E4 (ß = −1.5, [142]) but a steeper slope compared to water (ß = -0.67). The steeper slope likely reflects the low dielectric environment at the active site which increases the energy penalty for formation of an oxyanion. As expected, the Brønsted analysis further indicates less effective oxyanion stabilization in the Gln50Phe (ß = −1.72) and Gln50Met (ß = -1.73) variants. Poorer oxyanion stabilization, however, must be compensated in these variants since the loss in net activity is comparably minor.

In order to identify these compensating effects, Kries and Bloch prepared high-resolution crystal structures (0.95-1.45 Å) of all enzyme variants in complex with **1** (Figure 18, [144,145]). The Gln50Met and Gln50Phe mutations led to a shortened distance between the ligand and the catalytic base, indicating compensation of poor oxyanion stabilization with better positioning. Importantly, the sidechains of Met50 and Phe50 packed directly against the ligand. In contrast, residue 50 points away and makes room for a water molecule in the variants with His, Ser, Ala, and Lys at position 50. Replacement by water in the Gln50His variant is particularly surprising, since modeling had suggested that His50 would be nicely poised for a hydrogen bonding interaction with the ligand.

Interestingly, the His, Ser, Ala and Lys variants all have similar Brønsted slopes of -1.61 to -1.69, intermediate between the values determined for HG3.17 (ß = -1.49) and the Gln50Met and Gln50Phe variants (ß = -1.72 and -1.73). Given the weak activity of these variants, water seems to be worse for oxyanion stabilization than the amide side chain of Gln. These observations are in agreement with those made for KSI, where penetration of water at the oxyanion hole reduced the activity compared to properly positioned H-bonding residues [37,38]. In contrast to the Gln50Met and Gln50Phe HG3.17 variants, however, replacement of the oxyanion H-bonding residues in KSI with hydrophobic residues was deleterious for activity.
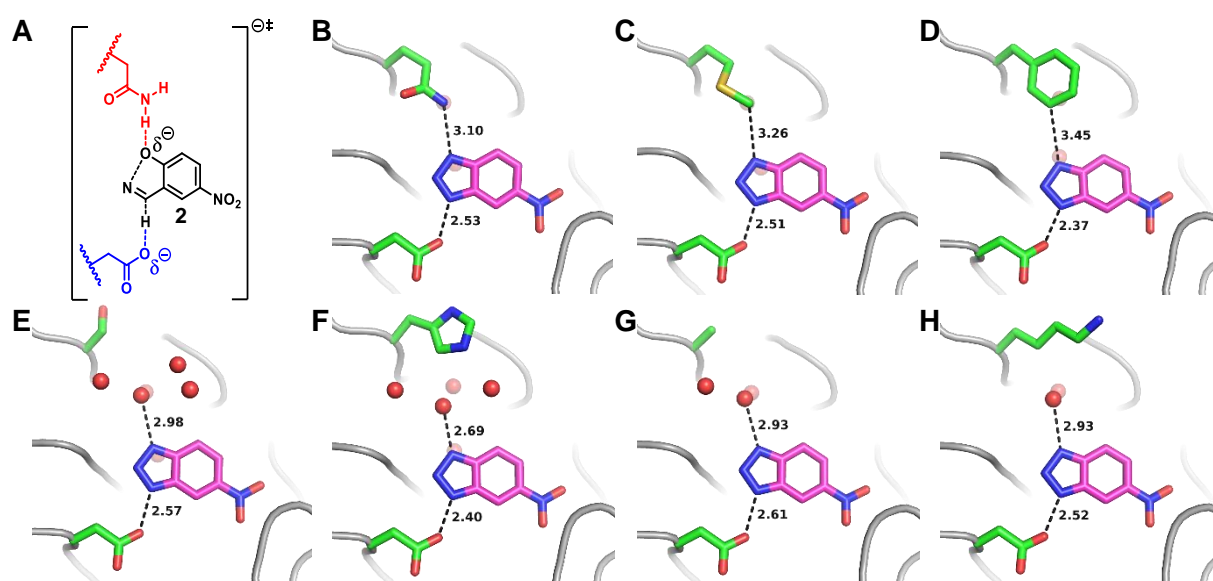
Figure 18: Structural analysis of Gln50 variants of HG3.17. Structures were solved in complex with 6-nitrobenzotriazole **1** (magenta). Residue 50 and Asp127 are shown as green sticks; the protein backbone is shown as a gray ribbon; red spheres represent water molecules. When two protein chains were present in the asymmetric unit, only chain A is shown. **(A)** Model of the transition state in HG3.17. **(B)** Active site HG3.17 (PDB code 4BS0, [4]), **(B)** Gln50Met, **(C)** Gln50Phe, **(D)** Gln50Ser, **(E)** Gln50His, **(F)** Gln50Ala, **(G)** Gln50Lys. **(C-D)** Hydrophobic Met and Phe sidechains assume a conformation similar to Gln50 although they cannot hydrogen bond with the ligand. **(E-G)** With short and polar sidechains at position 50, additional water molecules come in contact with the ligand. Adapted from [145].

In this thesis, the role of the oxyanion stabilizing residue, as well as positioning of the general base, is further scrutinized by computational analysis. In addition, primary kinetic isotope effects are presented which suggest that the documented effects on k_{cat}/K_M report on the energy barrier of the chemical reactions. In agreement with the experimental data, computational modeling indicates increased oxyanion stabilization by amides as compared to water. However, the computational predicted effects of binding geometry perturbations did not correlate with the observed activities. These results indicate a delicate interplay between positioning and oxyanion stabilization. Since computational enzyme design is typically based

on similar models as employed during this analysis, more sophisticated models will likely be necessary to design more effective catalysts.

2.1.1 Results

Substrate binding is not rate limiting. As indicated in Table 2, variants of HG3.17 display elevated K_M values (>1.5 mM). These increase the uncertainty on the extrapolated k_{cat} values due to limited substrate solubility (≤1.5 mM). Thus, analyses are typically based on k_{cat}/K_M values. To probe whether the chemical reaction or another step, such as substrate binding, is the kinetic bottleneck in the catalytic cycle of HG3.1, primary kinetic isotope effects on k_{cat} and k_{cat}/K_M were determined (Figure 19).
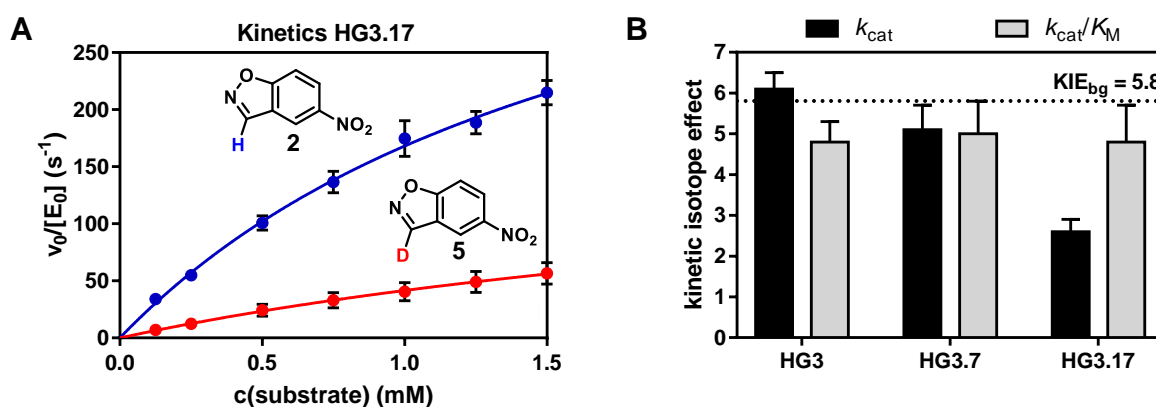
Figure 19: Primary kinetic isotope effects. **(A)** Kinetics of HG3.17 with deuterated (red) and non-deuterated (blue) substrate at 20 °C. **(B)** Isotope effects for HG3, HG3.7 and HG3.17 on k_{cat} (black) and k_{cat}/K_M (gray) indicate that product release becomes partially rate limiting in the most highly evolved HG3.17.

Primary deuterium isotope effects of 4.8±0.5 and 4.8±0.9 on k_{cat}/K_M were measured with 5-nitrobenzisoxazole **2** for HG3 and HG3.17, respectively. The isotope effect on the uncatalyzed background reaction (5.8±0.9) reproduced previously published values [62] and is only slightly larger compared to the effect on k_{cat}/K_M for HG3 and HG3.17. In contrast, the isotope effect on k_{cat} decreased from 6.1±0.4 to 2.6±0.3 during evolution of HG3 to HG3.17. For comparison, the isotope effects for catalytic antibody 34E4, which catalyzes the Kemp elimination with rate-limiting C-H proton abstraction, were 4.4 on k_{cat} and 5.8 on k_{cat}/K_M respectively. The observed isotope effects suggest that a step after the irreversible C-H/D bond cleavage becomes rate limiting in HG3.17, while the rate in HG3 is only limited by chemistry [146]. Thus, k_{cat}/K_M likely reports on changes in the activation energy barrier of the chemical reaction rather than another step like substrate binding.

Computational modeling of geometric effects. The impact of substrate positioning on catalysis in the Gln50 variants was analyzed by computation using density functional theory (DFT) in the gas phase. The transition state geometry for the reaction of acetate and 5-nitrobenzisoxazole **2** was computationally approximated by gradually increasing the scissile C-H bond distance until the isoxazole ring opened irreversibly. The transition state was further refined in a local search with the Berny algorithm, which directs geometric optimization towards local maxima on the potential energy surface [147]. A frequency calculation with the final transition state indicated the presence of a transition state with a single imaginary frequency corresponding to the reaction coordinate.

The impact of misplaced catalytic residues on the transition state energy was estimated by systematic variation of its distances, angles and dihedrals (Figure 20). The observed geometries of the inhibitor-complex in the crystal structures were subsequently compared to the computationally predicted energetic penalties. Deviation from the ideal angles for proton abstraction up to 20° led to penalties less than <2 kcal/mol. However, no correlation between the energetic penalties and experimental activities was observed. Additionally, analysis of the inhibitor-base distance revealed that the C-O distance observed in Gln50Phe (2.37 Å) is far from the calculated energy minimum at 2.66 Å. Intriguingly short inhibitor-base distances in variants such as Gln50Phe suggest that tunneling effects, which are not included in the computational model, are of increased importance for catalysis. However, this conclusion is uncertain. Given that the benzotriazole is an imperfect analog of the transition state, its active site geometry may be different from that of the true transition state [148].
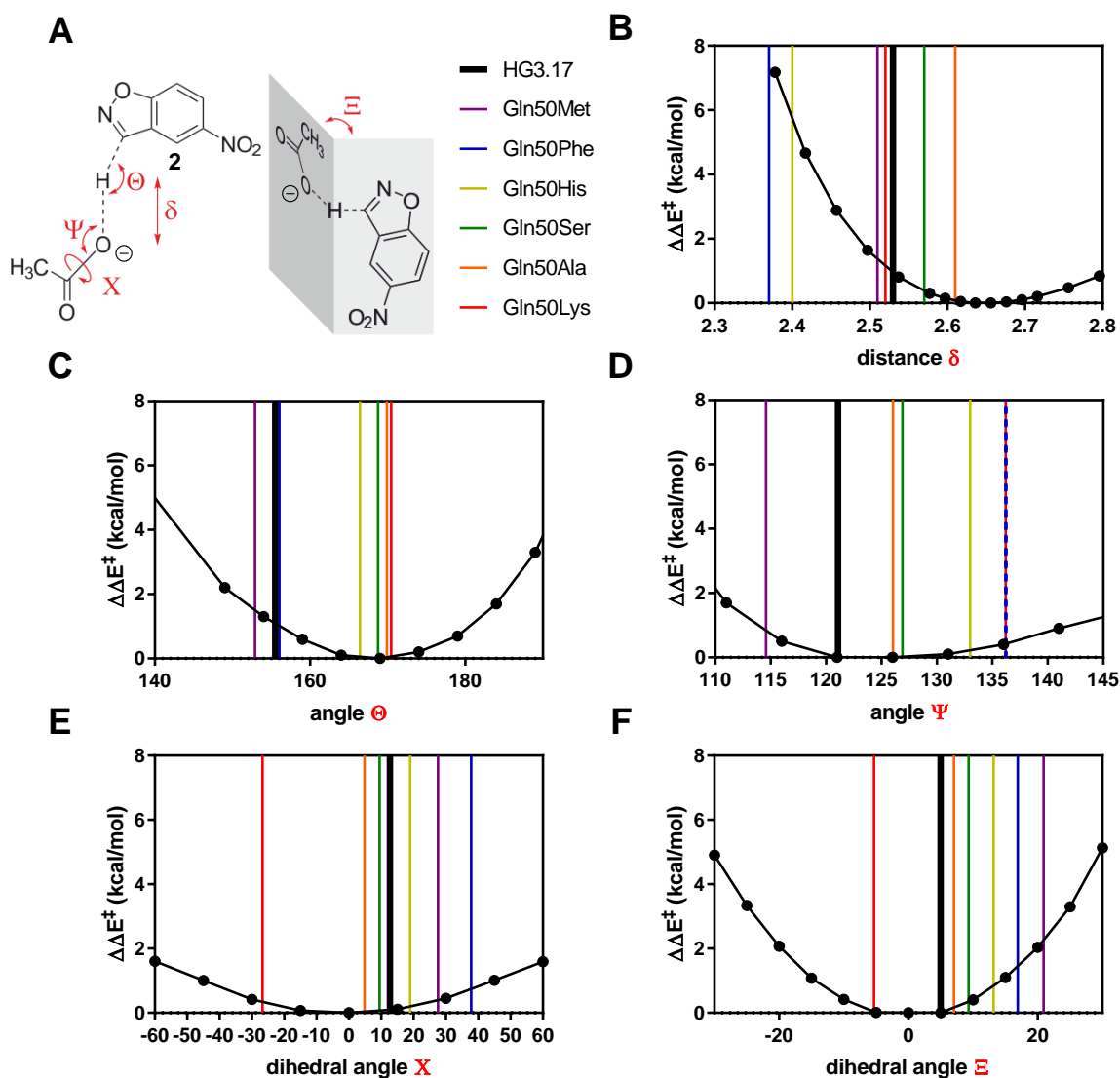
Figure 20: Impact on geometry on the activation energy. **(A)** Based on the ideal transition state geometry for the inhibitor base complex, defined distances, angles and dihedrals were modulated while fixing the other parameter and the effect on the activation energy was determined **(B-F)**. Energy changes in the observed geometries **(C-F)** never exceeded 2 kcal/mol. Only for distance δ **(B)**, large penalties of up to 7 kcal/mol were observed when comparing to structure, which indicates that the triazole inhibitor is only a poor transition state mimicry.

Computational modeling of oxyanion stabilization.   The catalytic contribution of oxyanion stabilization by protein sidechains and water was further analyzed by quantum mechanical calculations. Oxyanion-stabilizing residues, either water or acetamide, were modeled in complex with acetic acid and differently substituted benzisoxazoles. The effect on the activation energy was determined in a similar manner to the previous distance dependence analysis. The interaction of water with the phenoxide leaving group of the weakly activated substrate 5-bromobenzisoxazole decreased the activation energy by 9.6 kcal/mol, whereas acetamide lowered the barrier by 12.1 kcal/mol. The amide side chain of Gln50 thus seems to be better suited for oxyanion stabilization than water, which is present in the less stabilizing

variants with Ser, His, Ala and Lys at position 50. Importantly, oxyanion stabilization substantially decreased the activation energy even for the strongly activated 5-nitrobenzisoxazole **2** by 8.3 and 10.5 kcal/mol, respectively. Gas phase calculations thus suggest that oxyanion stabilization is even relevant for substrates with intrinsically stabilizing substituents such as the 5-nitro group.
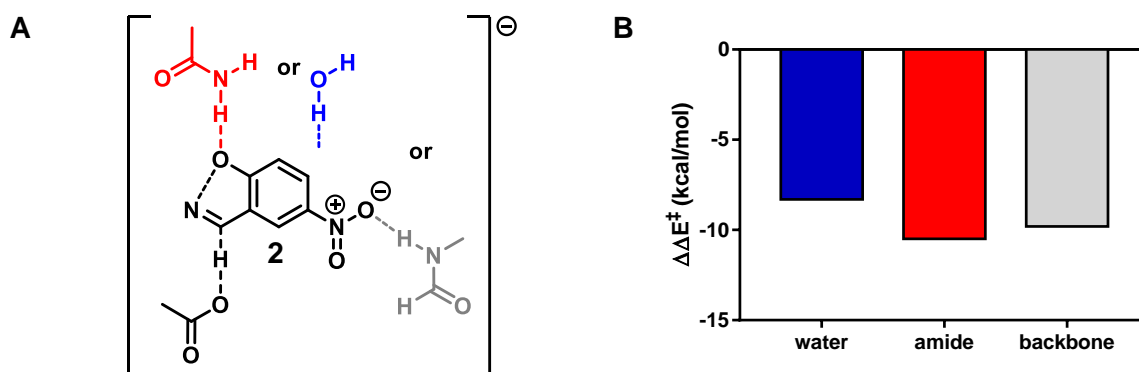
Figure 21: Impact of oxyanion stabilization on catalysis. **(A)** Activation energies were determined for the depicted substrate-base complex (black), which coordinated to either acetamide (red) or water (blue) via the phenolic oxygen. Alternatively, formamide was coordinated to the nitro group (gray). **(B)** Energy changes compared to the substrate-base complex without any ligands indicates a significant contribution of all ligands, with water performing slightly worse than the other two.

Inspection of the HG3.17 crystal structure suggests that the backbone amide of residue 237 could conceivably donate a hydrogen bond to the nitro substituent of the substrate in the transition state (Figure 22). As indicated by calculation, this interaction may stabilize the transition state by 9.8 kcal/mol, comparable to direct stabilization of the phenolic oxyanion by an H-bonding amide. H-bonding to the nitro group, which was also exploited during the design of Kemp eliminase KE59 [104], might thus be a significant contribution to catalysis. Hydrogen-bonding interactions to the backbone might furthermore contribute to the strong discrimination of HG3.17 against unsubstituted benzisoxazole (k_{cat}/K_M = 0.3 M⁻¹s⁻¹, [4]).
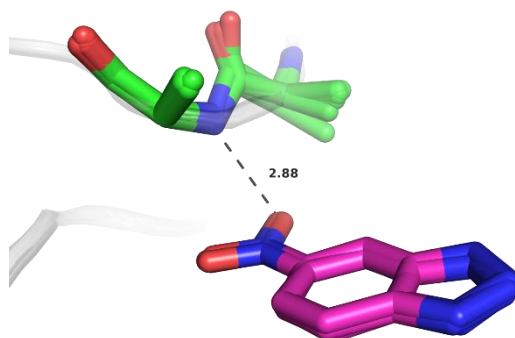
Figure 22: The hydrogen bonding interaction between the 6-nitrobenzotriazole **1** ligand (sticks in magenta) and backbone amide 237 is shown for an overlay of all HG3.17 Gln50-variants. The backbone of residues Leu236 and Met237 (including Cβ) is shown as green sticks. The N-O distance is shown for the structure of HG3.17 with black dashes.

2.1.2 Discussion

The computationally designed Kemp eliminase HG3.17 achieved a 6×10^8-fold rate acceleration over the background reaction after 17 rounds of directed evolution [4,103], approaching catalytic efficiencies of natural enzymes [118,135]. As indicated by the reduced isotope effect on k_{cat} of 2.6 ± 0.3, chemistry no longer appears to be solely rate limiting. Instead, a step after the C-H bond cleavage, likely product release, affects turnover efficiency [146]. Nonetheless, substrate binding only slightly impacts the overall rate as indicated by the significant isotope effect on k_{cat}/K_M. Hence, effects on k_{cat}/K_M can be employed to estimate the activation energy barrier of the chemical reaction.

Kries and Bloch previously performed a detailed kinetic and structural analysis of the Gln50 variants of HG3.17 [144,145]. Their analysis indicates an intricate balance between the electrostatic microenvironment and positioning of catalytic residues. Gln50Phe has the shortest, and probably most beneficial, interaction with the catalytic base, while wild-type Gln50 shows the highest degree of oxyanion stabilization. Importantly, the greasy and tightly packed active sites outperformed water-filled pockets with short (Ser, Ala) or misoriented (Lys, His) sidechains. Water thus seems to be less desirable than amide sidechains as a catalytic group, probably due to entropic penalties [149]. The computational study presented here furthermore suggests an enthalpic advantage of an amide over water for oxyanion-stabilization of up to 2 kcal/mol.

The proposed compensation of oxyanion stabilization with positioning is challenged by two Gln50 variants. The Gln50Met variant shows decreased oxyanion stabilization but has almost the same inhibitor-base distances as the wild-type Gln50. Furthermore, oxyanion stabilization by the water-filled Gln50His variant is almost as bad as that of the greasy Gln50Phe and Gln50Met variants. The two variants suggest that other effects, such as electrostatic contributions beyond those detectable by oxyanion stabilization, may be involved in catalysis. Oxyanion-stabilization via the nitro group maybe one important factor, modulations of the overall active-site electrostatics maybe another.

While characterization of HG3.17 highlights the significance of an appropriate electrostatic microenvironment and positioning of catalytic residues at the active site, a quantitative explanation is still missing. In this study, the geometric and electrostatic effects observed in the different Gln50 variants were thus translated to energetic terms using computational

methods. Changes in the activation energy of the isolated complex, comprising the transition state, general base and oxyanion stabilizing residue, were analyzed by DFT calculations. The analysis suggests that such reduced models, which are also employed during computational design, cannot rationalize catalysis in HG3.17 and its Gln50 mutants. The challenge of *de novo* designing enzyme-like catalysts may thus also be rooted in models that cannot accurately describe catalysis in actual enzyme-like catalysts. Furthermore, the highly-evolved HG3.17 is close to the targeted optimal geometry predicted by computation. Its catalytic efficiency, however, is still three orders of magnitude below the diffusion limit. More complex models will therefore be needed to accurately describe catalysis in HG3.17 and to design more efficient catalysts from scratch.

Novel experimental tools will also be required to discriminate the effects of positioning from electrostatic effects. Particularly active site electrostatics, which likely contributes to catalysis beyond stabilization of the oxyanion leaving group, cannot be experimentally analyzed. New spectroscopic probes, which have been employed to study electric fields in systems like KSI [18,150], may help to decipher catalysis in Kemp eliminases. As long as a full picture of the origins and limitations of catalysis in highly efficient novel enzymes is lacking, rational *de novo* engineering of perfect catalysts through design and evolution will remain elusive. Once the requirements for efficient catalysis have been determined, either by experiment or computation, computational design with more sophisticated models than the currently employed theozyme approach may lead to more efficient enzymes.

2.2 Electrostatic catalysis in an evolved Kemp eliminase ‡

Electrostatic catalysis plays a key role in natural enzymes [17]. While its precise contribution is still debated [56,57], experimental studies using vibrational spectroscopy [18,55,150] have reproduced the computationally predicted impact of electrostatics in several cases. Vibrational modes are inherently sensitive to the electrostatic environment, and their frequency often depends linearly on the electric field experienced by the probe. In Δ-ketosteroid isomerase (KSI), for instance, vibrational spectroscopy with the transition-state analog 19-nortestosterone indicated the presence of a 150 MV/cm field at the oxyanion hole which stabilizes a charged intermediate and transition states along the reaction coordinate [150]. It has been suggested that this extreme electric field contributes as much as 10^5 fold to the $10^{7.5}$-fold rate acceleration of the enzyme over the acetate-catalyzed background reaction.

In line with the significance of electrostatic catalysis in natural enzymes, computational studies have suggested that electrostatic preorganization increased in Kemp eliminase HG3.17 after evolution [125,126]. Likewise, computation indicated that electrostatic desolvation of the general base played an important role for catalysis in other Kemp eliminases [128,151]. However, only local electronic interactions are typically considered at the design stage. As a consequence, the inherent electric field of the protein scaffold may not be properly oriented for catalysis, as indicated by *in silico* analysis of the electric fields in Kemp eliminases KE07 and KE70 [127]. In the latter cases, the scaffolds' electrostatics appear to counteract catalysis. Since the electric field is composed of numerous small contributions, adjustment through evolution can be challenging. Harnessing the power of electric fields during design might thus improve catalysis in *de novo* enzymes considerably. Nevertheless, further studies will be required to assess the role of electrostatics in designer enzymes. Unfortunately, no assays have been reported to experimentally assess the electric fields in such systems.

Here, a new vibrational probe is presented to experimentally analyze the electric field in Kemp eliminases. The transition state analog 6-cyanobenzotriazole **4** was used to interrogate changes in the electrostatic environment around the triazole, via inductive effects at the oscillating nitrile-group (Figure 23).
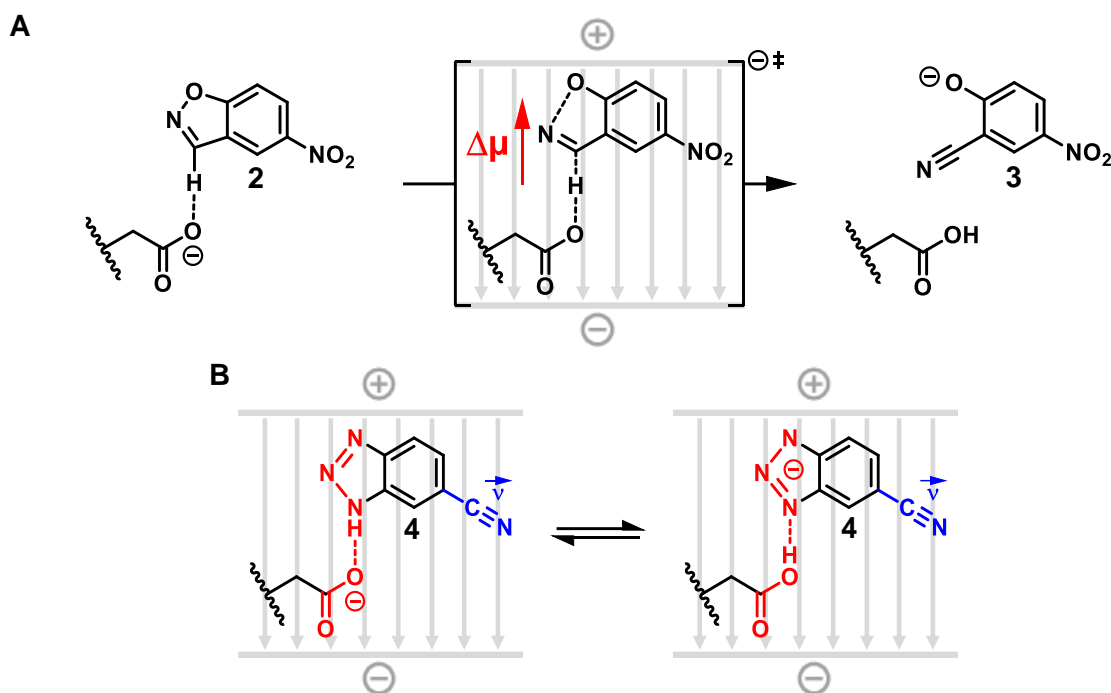
Figure 23: Electric fields in Kemp eliminases. **(A)** Kemp elimination of 5-nitrobenzisoxazole **2**. An external electric field (gray) may accelerate the reaction by stabilizing the dipole moment (red) of the transition state. **(B)** The proton between the inhibitor 6-cyanobenzotriazole **4** and the catalytic base may either reside on the base or on the triazole (red). The vibrational probe (blue) senses changes in the electric field (gray) directly via the Stark-effects and indirectly via inductive-effects from the charge distribution on the reacting groups (red). The indirect inductive effect, as well as changes in the protonation equilibrium, can be employed to study electric fields in Kemp eliminases.

The approach presented here is conceptually different from other reported approaches where the electric field is sensed directly at the oscillator [18]. Furthermore, the acidic transition state analog can exist in two different protonation states which gives rise to two independent peaks in the vibrational spectra, allowing a precise ratiometric analysis. The ratio of the two peaks, as well as their frequency differences, correlate with the catalytic efficiency of the previously discussed Gln50 variants of HG3.17. Computational analysis indicates that the nitrile probe senses remote changes in the electrostatic environment at the scissile N–H bond through inductive effects. The vibrational data further suggest that mainly electrostatic catalysis, not geometric positioning, is key to the two order of magnitude spread of activity in the Gln50 variants. Electric fields may thus be an important design feature for future engineering endeavors.

2.2.1 Results

Kinetic and spectroscopic data. The Kemp eliminase inhibitors 6-cyanobenzotriazole **4** and 6-nitrobenzotriazole **1** were analyzed by vibrational spectroscopy when bound to HG3, HG3.17 and all HG3.17 Q50X variants (Figure 24). Since the nitrile vibration is in a relatively empty part of the spectrum, the cyano-substituted inhibitor could be directly analyzed by IR spectroscopy. On the other hand, the vibration of the nitro probe overlapped with several vibrational modes of the proteins. Pre-resonance Raman spectroscopy, however, allowed selective excitation of the nitro probe. Both probes yielded qualitatively comparable results. Nonetheless, the following discussion will focus on the more accurate results for the cyano-substituted inhibitor.
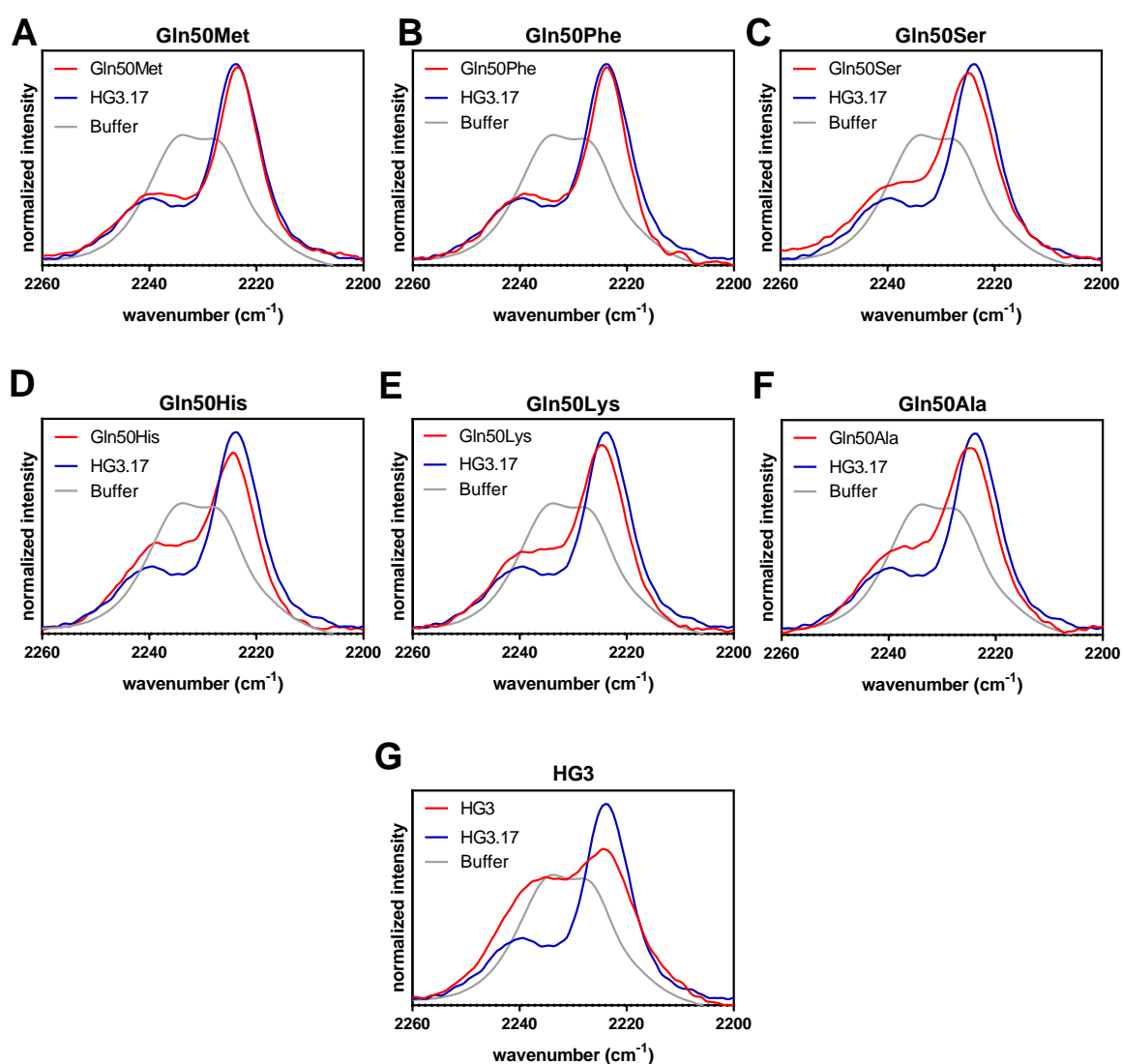
Figure 24: Background-subtracted FTIR spectra of 6-cyanobenzotriazole **4** bound to Gln50 variants. For comparison, the spectra of buffer (gray) and wild-type HG3.17 (blue) are overlaid with the respective variant (red). The ratio of the area of the two peaks, as well as their frequency difference, increases with increasing activity of the Kemp eliminases.

The vibrational spectra 6-cyanobenzotriazole **4** show two different peaks. Their frequencies (v), frequencies differences (Δv), as well as the free energy derived from the peak ratios (ΔG$_{eq}$°) all correlated the experimentally determined activation energy (ΔG‡, Figure 25 and Table 3). Samples with more than 30% free inhibitor (Q50H, HG3, empty circles in Figure 25) were excluded from all fits. One of the two species observed with IR spectroscopy is selectively stabilized with increasing catalytic efficiency. Furthermore, increasing catalytic efficiency also correlated with the absolute frequencies (v) and increased the frequency differences (Δv). The changes in v and Δv are indicative for a perturbation in the electrostatic environment of the vibrational probe. Finally, a decrease in activation energy also lead to tighter binding of the transition state analog (ΔG$_I$°) as indicated by separate K$_I$ measurements.
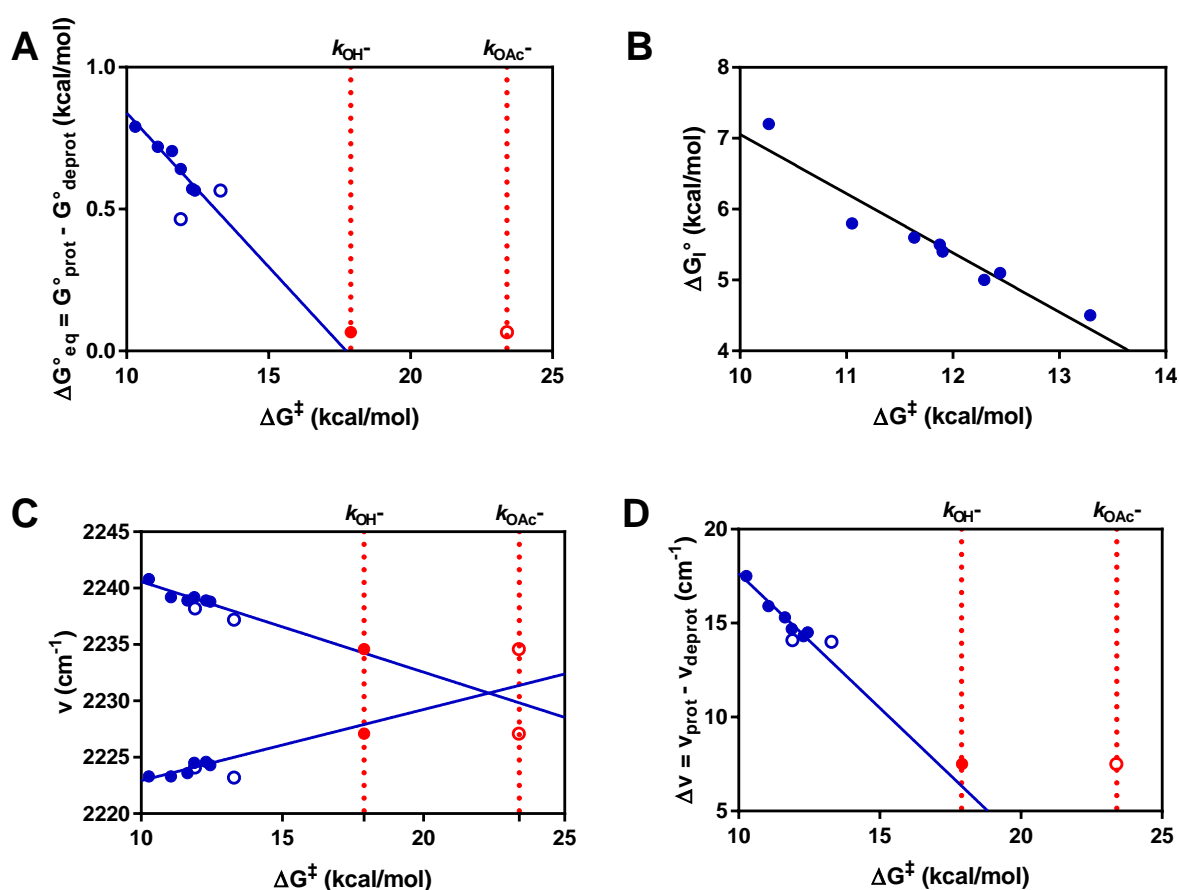
Figure 25: Spectroscopic and kinetic data for all protein variants (blue) with the cyano-probe. Samples with more than 30% free inhibitor (Q50H, HG3, empty circles) were excluded from all fits. The buffer sample is plotted at a free energy corresponding to the hydroxide- (red, full circle) or to the acetate-mediated reaction (red, empty circle) in water [62]. Free energy **(A)** of the peak equilibrium and **(B)** of inhibitor binding. **(C)** Frequencies and **(D)** frequency differences of the nitrile-vibrations.

Nitro-probe. The pre-resonance Raman spectra with bound 6-nitrobenzotriazole **1** (Figure 26) qualitatively agree with data obtained with nitrile inhibitor (Figure 25). However, the peak at lower wavenumbers has much lower intensity and the peak ratios could not be accurately quantified. Moreover, the spectrum of the free inhibitor had to be subtracted due to significant amounts of free inhibitor, giving rise to additional errors. Further analysis will hence focus on the cyano-substituted inhibitor only.
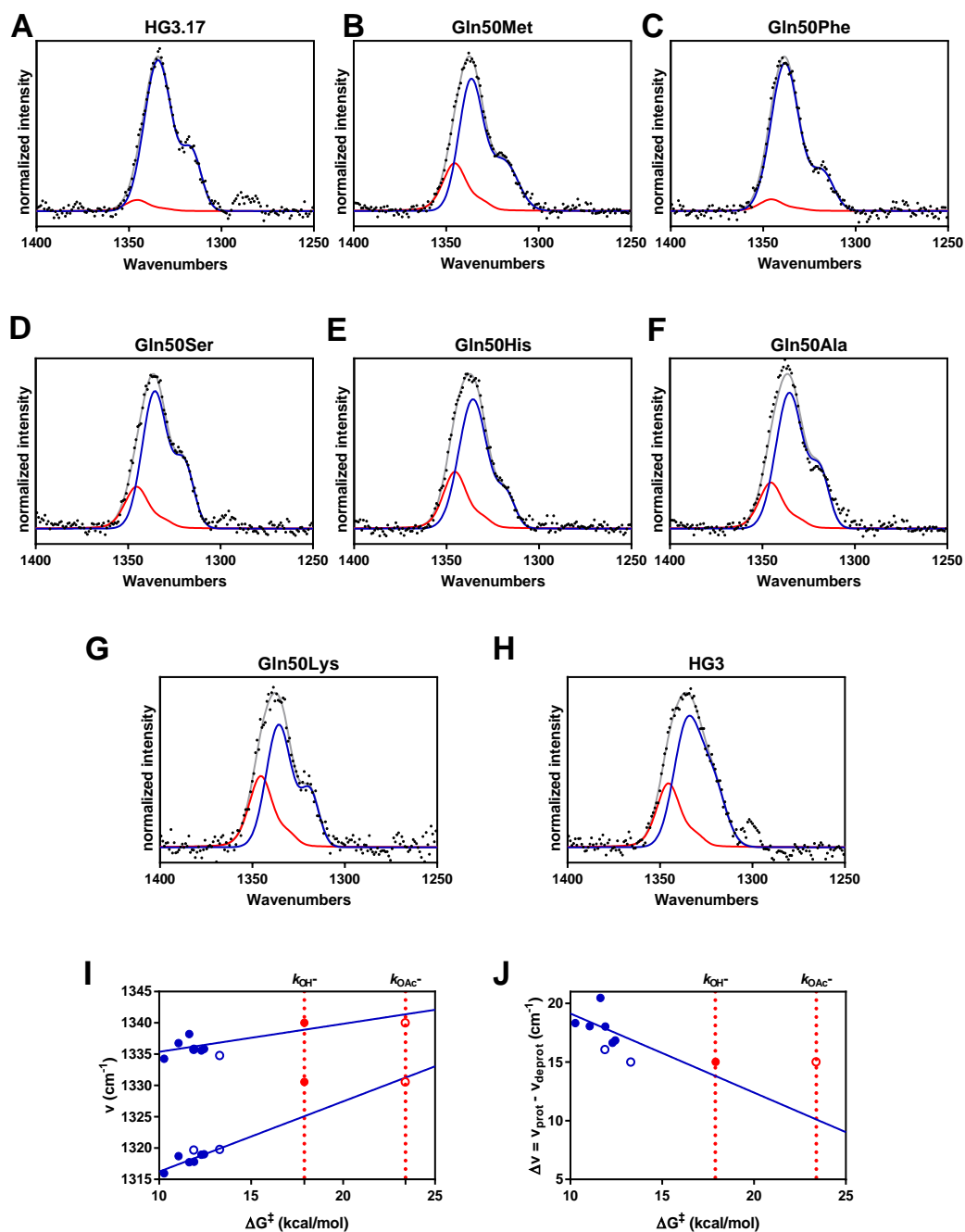
Figure 26: Pre-resonance Raman spectra of the Gln50 variants bound to 6-nitrobenzotriazole **1**. Due to the significant amount of free inhibitor, the spectrum of the free inhibitor in solution (red) was subtracted from the raw data (gray) to give the spectra of the inhibitor bound to the HG3.17 variants **(A–H)**. The absolute frequencies **(I)** as well as frequency differences **(J)** qualitatively agree with the data obtained for the nitrile probe (Figure 25).

Table 3: Kinetic data of the Gln50-variants

	k_{cat}/K_M [a] (M^{-1}s^{-1})	K_I [b] (µM)	c(Enz) [c] (mM)	bound inhibitor [d] (%)
HG3.17	2.06×10^5	5.9	2.3	98
Q50M	5.55×10^4	62	2.4	90
Q50F	2.08×10^4	83	3.5	95
Q50S	1.39×10^4	97	3.5	94
Q50H	1.33×10^4	110	1.7	71 [d]
Q50A	6.90×10^3	240	2.9	84
Q50K	5.38×10^3	199	3.5	90
HG3	1.30×10^3	560	0.6	22 [d]
background	[e]	n. A.	n. A.	n. A.

Table 3, contiued: Spectroscopic data of the Gln50-variants

	Nitrile-Probe				Nitro-Probe [f]		
	v_1 (cm^{-1})	v_2 (cm^{-1})	dv (cm^{-1})	$\Delta G_{eq}°$ (kcal/mol)	v_1 (cm^{-1})	v_2 (cm^{-1})	dv (cm^{-1})
HG3.17	2223.3	2240.8	17.5	0.79	1315.9	1334.3	18.3
Q50M	2223.3	2239.2	15.9	0.72	1318.7	1336.8	18.1
Q50F	2223.6	2238.9	15.3	0.71	1317.7	1338.2	20.5
Q50S	2224.5	2239.2	14.7	0.64	1319.7	1335.8	16.1
Q50H	2224.1	2238.2	14.1	0.47	1317.8	1335.8	18.0
Q50A	2224.6	2238.9	14.3	0.57	1318.9	1335.6	16.6
Q50K	2224.3	2238.8	14.5	0.57	1319.0	1335.8	16.8
HG3	2224.2	2237.2	14.0	0.75	1319.8	1334.8	15.0
background	2227.1	2234.6	7.5	0.07	1330.6	1345.6	15.0

[a] Catalytic efficiencies for 5-nitrobenzisoxazole **2** at 20 °C. [b] K_I for 6-cyanobenzotriazole **4** at pH 7.0 [c] Final concentration of the enzymes during the spectroscopic analysis. The inhibitor concentration was 2 mM. [d] Fraction of the bound inhibitor 6-cyanobenzotriazole **4** at the assay conditions. Q50H and HG3 were excluded from all fits due to the low levels of bound inhibitor (empty circles in Figure 25). [e] The k_{OH^-} and k_{OAc^-} for 5-nitrobenzisoxazole **2** are 0.56 M^{-1}s^{-1} and 5.77×10^{-5} M^{-1}s^{-1} respectively [62]. [f] Due to the small second peak, no peak ratio could be determined for the nitro probe.

Comparison with KSI. Remarkably, extrapolation of v, Δv, and $\Delta G_{eq}°$ of the nitrile- and nitro-probes agreed well with $\Delta G^{\ddagger}$ of the hydroxide-promoted reaction (k_{OH^-}) in water for the Kemp elimination of the respective benzisoxazole substrates [62]. This correlation indicates that changes in the electrostatic environment upon transition from bulk solvent to the protein active site significantly affect the spectroscopic and energetic properties of the inhibitor complexes. A similar trend was observed for published data on KSI (Figure 27). $\Delta G^{\ddagger}$ of the

hydroxide-mediated background reaction of KSI [152] and the vibrational frequency of its inhibitor in water also correlate well with the data collected for different KSI variants [150].
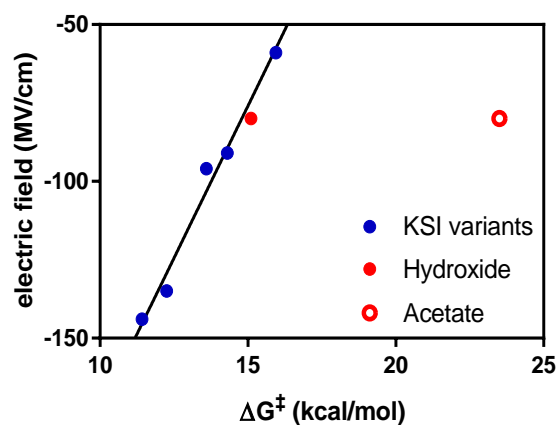
Figure 27: Electric fields in KSI variants (blue) in comparison to the hydroxide- and acetate-catalyzed background reaction in water (Data extracted from [150,152]).

Interestingly, the acetate-promoted reaction has a higher activation energy as compared to the hydroxide-promoted reaction for both systems (6.8 kcal/mol for the Kemp elimination [62] and 8.4 kcal/mol for KSI [152]). The increased activation energies may reflect a decrease in electrostatic preorganization of the acetate-promoted reaction, and can be macroscopically described by a Brønsted slope. Though the observations are remarkably similar for the KSI and Kemp eliminase variants, and may potentially be ascribed to electrostatic field effects, further studies will be required to dissect the observations for the hydroxide- and acetate-promoted reaction.

Protonation equilibrium leads to two peaks.   In order to dissect the identity of the two peaks observed in the IR measurements, the spectra of 6-cyanobenzotriazole **4** were measured at various pH values. A pH-dependent interconversion of the two peaks was observed by IR spectroscopy (pK_a=6.51±0.02, Figure 28A) and by changes in UV/Vis-absorbance (pK_a=6.49±0.03). It is therefore likely that the additional peak arises from two different protonation states of the inhibitor. Signal from free inhibitor molecules should also be negligible, since up to 98% of the molecules are bound based on the measured K_I values (Table 3).

An X-ray structure of 6-cyanobenzotriazole **4** bound to HG3.17 was determined at 2.0 Å resolution. The observed complex rules out alternative binding modes as the source of additional peaks (Figure 28B). Additionally, there are no H-bonding residues in close proximity to the nitrile substituent that might perturb the vibration and give rise to a second peak through

H-bonding [153]. Though backbone amide 273 is in close proximity to the nitrile, the geometry is unfavorable for an H-bonding interaction. In conclusion, it is likely that the two observed peaks of the inhibitor bound to the Kemp eliminase variants are caused by two different protonation states of the probe

A

B
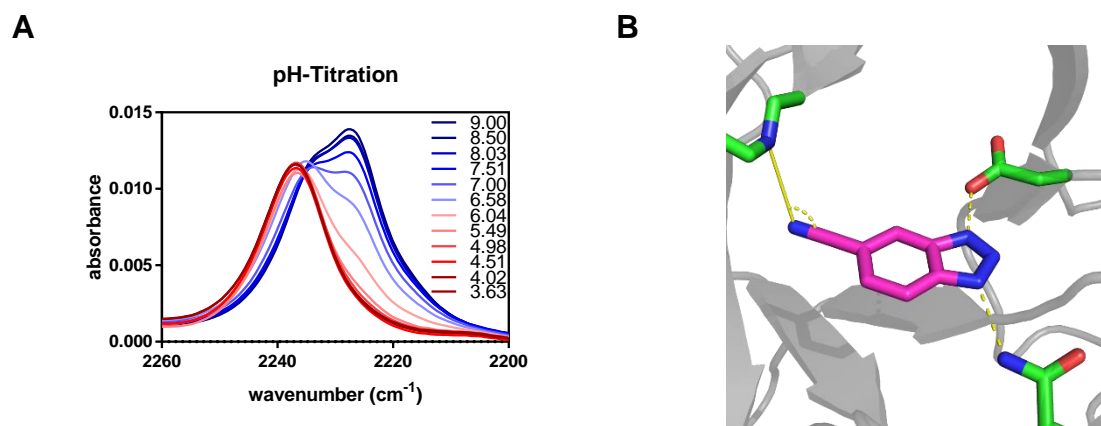
Figure 28: The inhibitor can be in two different protonation states. **(A)** IR spectra of 6-cyanobenzotriazole **4** in solution show a pH-dependent interconversion of the two peaks. **(B)** In the crystal structure of HG3.17 with 6-cyanobenzotriazole **4** (magenta) only a single binding mode was observed. N1 of the triazole forms an H-bond to the catalytic base. N3 of the triazole is intimately H-bonded with Gln50, while no H-bonding interactions are observed with the cyano substituent.

For further analysis, it will be crucial to determine which of the two ionizable nitrogens of the triazole, N1 or N3, changes its protonation state. While N3, which is on the opposite side of the base, is involved in an intimate H-bonding network in HG3.17 and likely deprotonated, N1 may be in a protonation equilibrium with the general base. Because the photometrically determined pK_a of the inhibitor (6.49±0.03) is close to the kinetic pK_a of the catalytic base (6.0, [4]), it is likely that the protonated inhibitor is in equilibrium with the protonated base at the active site, resulting in the two IR peaks observed for the enzyme-bound forms. (Figure 29).

DFT calculations were performed to dissect the origins of the spectroscopically observed effect. The computational model included only 6-cyanobenzotriazole **4** in a protonation equilibrium with acetic acid (Figure 29A). Since the deprotonated inhibitor was unstable in the gas phase, probably due to underestimation of the pK_a of acetic acid, the impact of solvation on the protonation equilibrium was first assessed by addition of a polarizable continuum model (PCM). The deprotonated state was stabilized with an increasing dielectric constant of the PCM (Figure 29B), suggesting a strong contribution from electrostatic effects to the experimentally observed peak ratios. The added dielectric essentially screens ionic interactions and stabilizes the negative charge on the acetic acid and thus shifts the equilibrium to the protonated inhibitor. Furthermore, at the dielectric constant of water (ε = 78.4), a $\Delta E_{eq}°$ of 0.25 kcal/mol

was computationally determined which is close to the experimentally observed $\Delta G_{eq}°$ at pH 7.0 of 0.22 kcal/mol. Based on these results, PCM with a dielectric constant of water was included in all subsequent calculations.
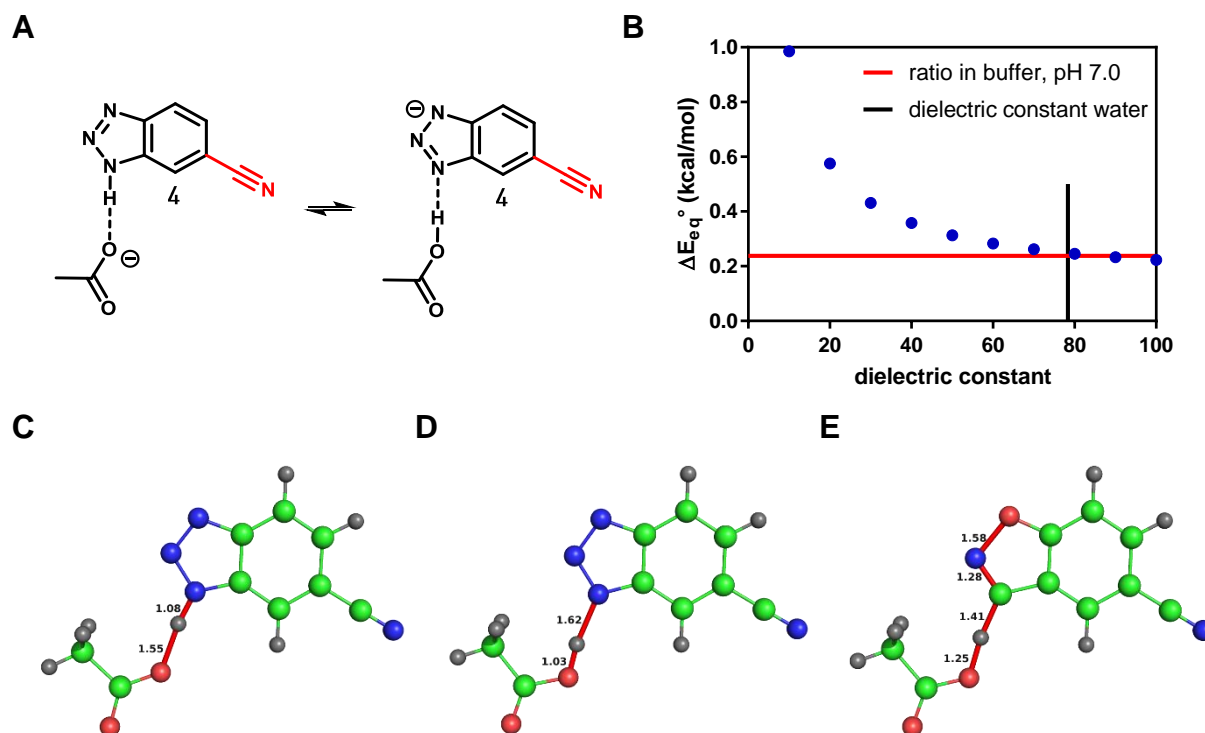
Figure 29: Computational model of the inhibitor bound at the active site. (A) Equilibrium between protonated and deprotonated inhibitor. (B) Predicted effect of the dielectric constant with PCM on the equilibrium ($\Delta E_{eq}°$). Optimized geometries of the (C) protonated inhibitor, (D) deprotonated inhibitor and (E) transition state of the Kemp elimination with PCM (ε=78.4).

Influence of electrostatics. The observed change in the ratio of protonated and deprotonated states can be described by changes in the pK_b of either the substrate or the inhibitor. These pK_b changes may be, for instance, ascribed to electrostatic effects. In order to analyze the electrostatic effect on the catalytic base, differently substituted carboxylic acids were analyzed in complex with the 6-cyanobenzotriazole **4** by computation. In these DFT calculations, the pK_b of the base was modulated by substitution while the distance to the inhibitor or substrate was fixed to 2.6 Å to exclude steric effects arising from varying equilibrium distances. Changes in the protonation equilibrium ($\Delta E_{eq}°$) and nitrile vibration (Δv) of the triazole inhibitor, as well as the activation energies ($\Delta E^{\ddagger}$) of the isoxazole-substrate, were determined after substitution of the base with mono-, di-, and tri-substituted chloro- or fluoro-acetic acid, and formic acid (Figure 30).
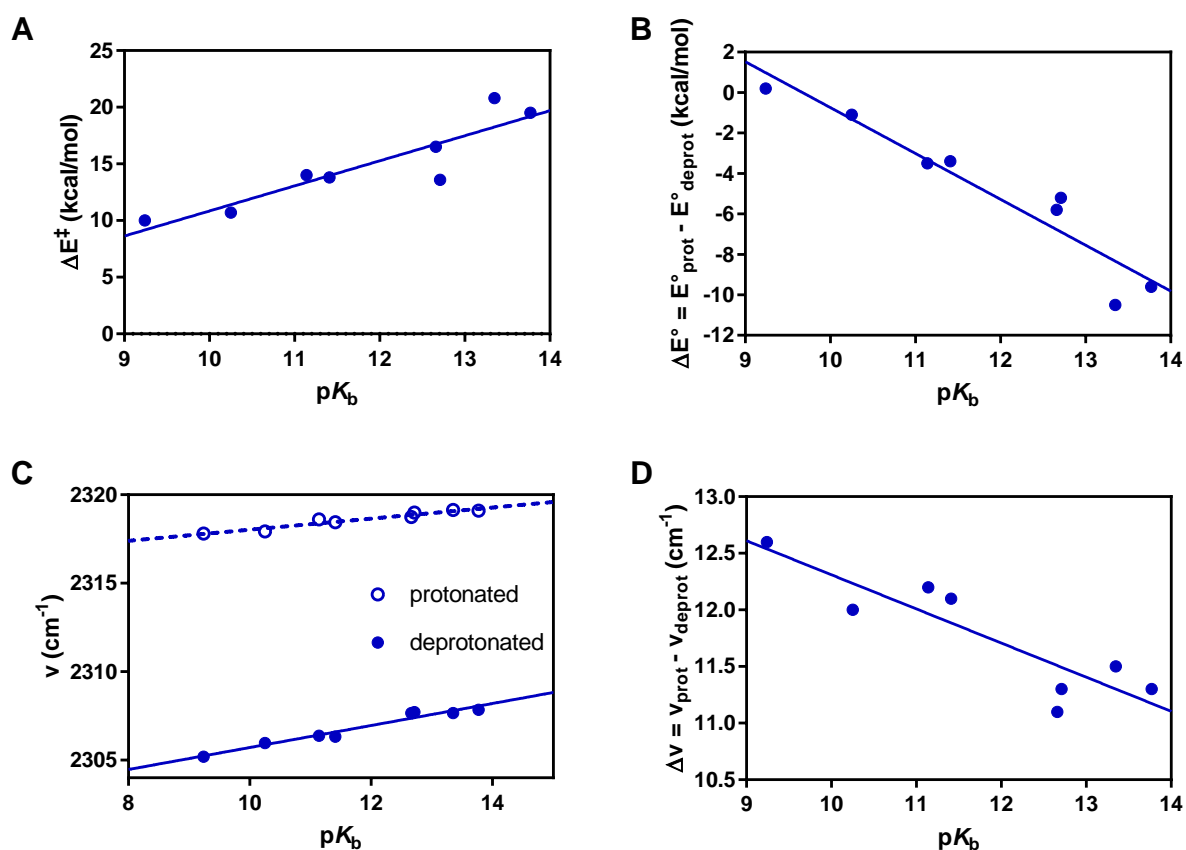
Figure 30: Effects of the base pK_b on Δv, $\Delta E_{eq}°$ and $\Delta E^{\ddagger}$. **(A)** A decrease in base pK_b leads to a decreased in $\Delta E^{\ddagger}$, stabilization of the deprotonated inhibitor **(B)**. The pK_b affects the frequencies **(C)** and leads to an increase in Δv increased **(D)**.

In agreement with literature precedent [62], the activation energy of the Kemp elimination was decreased with decreasing pK_b of the catalytic base. The decreased activation energy led to an increased frequency difference between the protonated and the deprotonated inhibitor. Also, the protonated species was selectively stabilized as base strength decreased. Finally, both the protonated and deprotonated states were stabilized with decreasing pK_b, indicating a tighter interaction with the inhibitor and leading to a decreased K_i. All analyzed effects are in good agreement with the experimental data and suggest that electrostatic effects play a key role in the observed phenomena.

In the experimentally analyzed variants of HG3.17, the Gln50 residue, which interacts with the inhibitor on the opposite site of the catalytic base, was modulated. Owing to the proximity of Gln50 to the inhibitor, it is thus more likely that mutagenesis will affect the inhibitor and not the base pK_b. While an analogous analysis by changing the substitution of the inhibitor was not performed due to the a lack of experimental pK_b values, similar trends on the protonation equilibrium, frequencies, frequency difference and activation energy [62,110] would likely be expected upon modulation of the inhibitor pK_a.

Influence of base position. The extent to which shifts in IR vibrations are caused by changes in the electric field of the enzyme or positioning of catalytic residues has been rigorously debated for KSI and other systems [18,150,154]. The effect of the position of the acetate base was therefore evaluated by analyzing the distance dependence of the base relative to the inhibitor (Figure 31). The computational analysis shows that the activation energy is increased as the distance is changed from the optimum of approximately 2.6 Å. While the data are reliable for longer distances, analysis of shorter distances was hampered by the instability of the base's anti orientation leading to a flip to the cis orientation in the ground state.
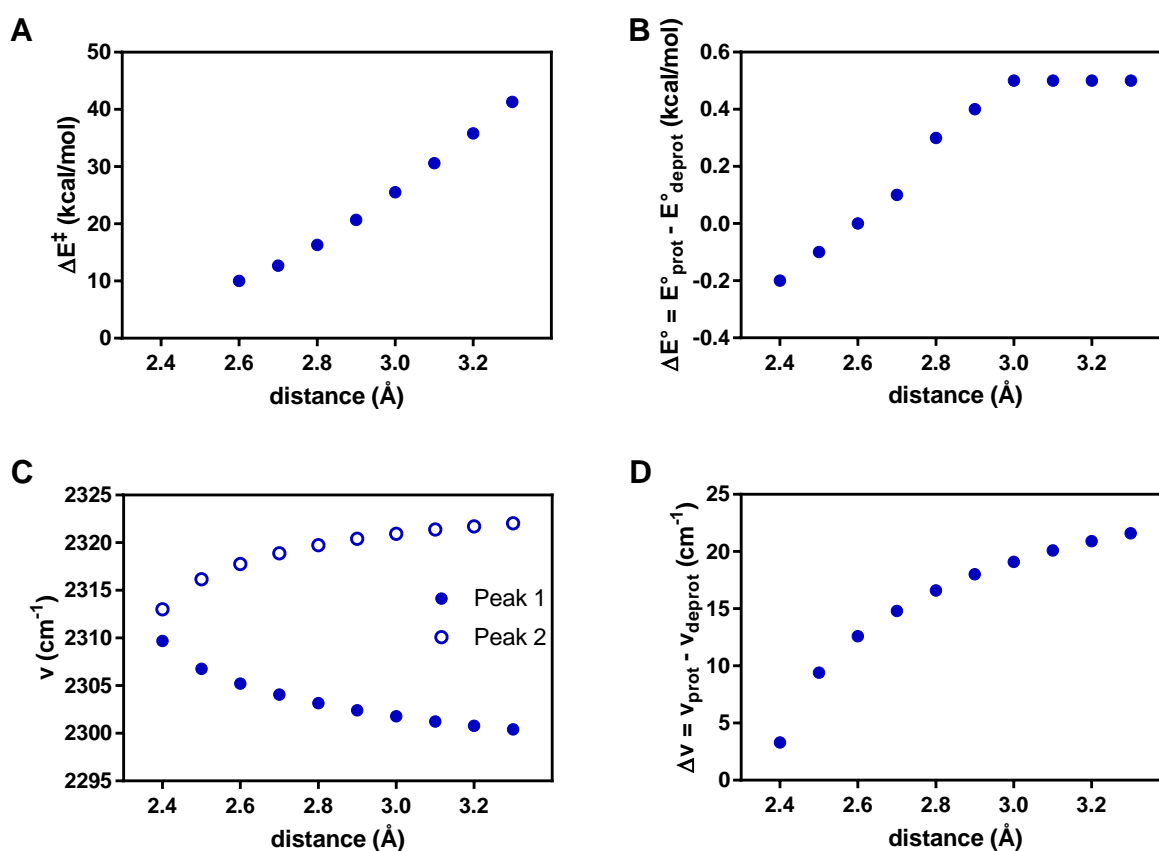
Figure 31: Effect of the ligand-base distance on the activation energy **(A)**, the equilibrium ($\Delta E_{eq}°$) between protonated and deprotonated species **(B)**, as well as the frequency (v) and frequency difference of the nitrile oscillator **(C+D)** The calculated effects show opposite trends than what is experimentally observed.

For the inhibitor complexes, it was observed that the deprotonated state was destabilized with decreasing activation energy, which contradicts the experimental observation that more efficient catalysis leads to stabilization of this state. Also, the frequency difference was decreased which does not match the experimental data. Computation thus suggests that refinement of the inhibitor / substrate – base distance cannot explain the observed effects. It

is likely that vibrational spectroscopy of the inhibitor-bound enzymes reports on true field effects rather than effects arising from optimized positioning of the base.

Molecular basis of the observed effects. The effects of distance and base substitution on the spectroscopic and thermodynamic parameters of the complexes were dissected by analyzing potential energy surfaces for a few of the aforementioned inhibitor-base complexes. On these surfaces, the N-H distance was scanned while keeping the N-O distance constant at 2.5 Å to 2.8 Å. Furthermore, energy surfaces were analyzed for a constant N-O distance of 2.6 Å for mono-, di-, and tri-fluoro substituted acetic acid (Figure 32).
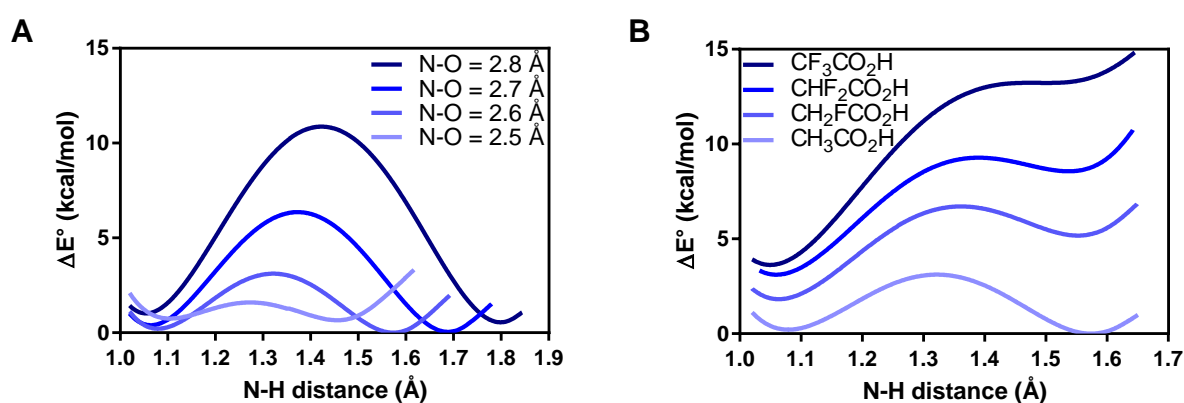
Figure 32: Potential energy surfaces of the inhibitor-base complex. The N-O distance was kept constant in all cases while the N-H distance was varied for **(A)** different initial N-O distances or **(B)** differently substituted bases.

The observed two-welled potential energy surfaces are significantly affected by changes in distance and base pK_b. The closer the base gets to the inhibitor, the closer the two minima become and the energy barrier between the two wells decreases. At some point, the base is so close to the inhibitor, that the two different minima for the proton become one, in agreement with Figure 31C. On the other hand, changing the base substitution only weakly affected the position of the wells, but strongly modulated their energies. Importantly, it was previously observed that the N-H distance predominantly determined the nitrile vibration with longer N-H distances leading to a decrease in wavenumber. A difference in the position of the minima thus translates to a change in frequency, while a change in energy leads to a different equilibrium constant for the two species. The potential energy surfaces thus demonstrate that decreasing the base distance leads to a decrease in frequency difference, contradicting the observed effect. On the other hand, modulation of the pK_b primarily leads to stabilization of the deprotonated state, which is in agreement with the reported data.

2.2.2 Discussion

This work demonstrates that the electrostatic microenvironment contributes substantially to catalysis in the *de novo* engineered enzyme HG3.17. In contrast to studies on ketosteroid isomerase [150], only qualitative data can be derived due to the indirect nature of the assay. Nonetheless, the observed electrostatic effects seem to be qualitatively sufficient to explain the 2 kcal/mol span in activation energy among the HG3.17 Gln50-variants. The effect on the peak ratios suggests selective stabilization of one protonation state over the other, which is likely caused by electronic effects as indicated by the modulated vibrational frequencies. Electrostatic catalysis also seems to provide an important contribution to catalysis in the original computational design HG3. The ambiguous binding mode observed in the crystal structure (3NYD, [103]) and low affinity for the inhibitor, however, preclude more detailed conclusions regarding the computational design.

The novel assay presented here relies on a relative analysis of two different states available to the vibrational probe. This ratiometric approach may be much more precise than other assays relying on the observation of only one species [18]. Furthermore, the presence of two species, caused by a two-welled energy potential, is key to the observed effects. DFT analysis showed that electrostatic effects reproduce the observed phenomena, while distance-related effects lead to behavior opposite than expected. The two-welled energy potential thus allows discrimination of electrostatic effects from geometry-related effects, which is challenging for other systems [150].

This study suggests that explicit consideration of enzyme electrostatics will be required to generate better catalysts by *de novo* design. Similar conclusions were drawn from other computational studies on designed enzymes which highlighted electrostatic preorganization as key to catalysis [125,128,151] and noted that the scaffold's electric field is often not properly oriented for catalysis [127]. Precise control of electrostatic effects at the local level through exact positioning of first-shell residues and at the global level through inclusion of the scaffold's electrostatics may help to improve computational design algorithms. Finally, a key characteristic of the electrostatic proposal entails selective stabilization of the transition state dipole moment compared to the ground state (Figure 5, [18]). In other words, some interactions with the transition state, those that stabilize a changing dipole moment, contribute more to catalysis than others. The conventional theozyme approach, however, focuses on stabilizing the transition state as a whole. For the design of Kemp eliminases, an

interaction that stabilizes the oxyanion would count as much as an interaction with the benzene-moiety. Expansion of the theozyme approach to selective stabilization of changing dipole moments may thus be required to design more efficient enzymes in the future.

2.3 Optimization of conformational landscapes during enzyme evolution ‡

Proteins are inherently flexible and typically undergo dynamic movements at timescales ranging from femtoseconds to minutes [155]. Sampling of conformational states on timescales slower than ns is often involved in enzyme regulation [156], adaptation [157] or promiscuity [158]. For instance, Tawfik proposed that promiscuous activity may arise if rare conformational sub-states display activity for an alternative reaction [159]. The relationship between conformational sampling and promiscuous activity is nicely exemplified with the directed evolution of a bacterial phosphotriesterase towards promiscuous arylester hydrolysis [160-163]. In that system, flexibility of a lid-forming loop is key to the catalytic mechanism and promiscuity arose through the rare opening and closing of an alternative loop [160]. Over the course of evolution, the specificity of arylesterase activity over the native activity switched by a factor of ~10^9 [161,162]. The altered specificity was enabled by complete remodeling of the conformational landscape of the enzyme [163]. Molecular dynamics and room-temperature X-ray crystallography indicated that a variant with high activity for both reactions displayed elevated conformational dynamics in both lid-forming loops. Importantly, minimization of conformational dynamics in one of the loops was further required to convert the "generalist" enzyme, capable of catalyzing both reactions, into a highly active "specialist". Thus, conformational ambiguity may lead to promiscuous activity for different reactions [158]. However, stabilizing a productive state, or an ensemble of states, seems to be typically required to convert the generalist into a highly efficient catalyst [155].

Conformational flexibility has also played a role in the directed evolution of *de novo* enzymes. A mobile loop in the designed retro-aldolase RA95 was shown to adopt several distinct conformations, allowing population of different binding pockets by the substrate over the course of evolution [5,101,122]. In another example, residues that form salt bridges with the catalytic base in two Kemp eliminases (KE07 and KE70) adopted different conformations

‡ Scientific contributors

Vy Nygen, Renee Otten, Dorothee Kern (Brandeis U) performed the NMR experiments described in the discussion.

after evolution. The change in conformation led to increased reactivity of the base [104,119,121]. During evolution of Kemp eliminase HG3, conformational ambiguity of the substrate was resolved so that only a unique binding mode is populated in the evolved catalyst [4]. Thus, evolution of *de novo* designed enzymes, like natural enzymes, often seems to lead to a decrease in conformational flexibility and stabilization of a unique productive state [30].

Here, detailed biochemical studies of HG3 and HG3.17 are presented that reveal a temperature and pH-dependent interconversion between an active and an inactive state. Kinetic analyses show that evolution selectively populated the catalytically active state of the protein at ambient temperature and pH (Figure 33). NMR analyses indicate the presence of at least two states in solution, which behave like the active and inactive states. Remodeling of conformational landscapes through evolution led to increased activity in HG3.17. Similar effects may be exploited in other catalysts that show conformational heterogeneity to increase their activity.
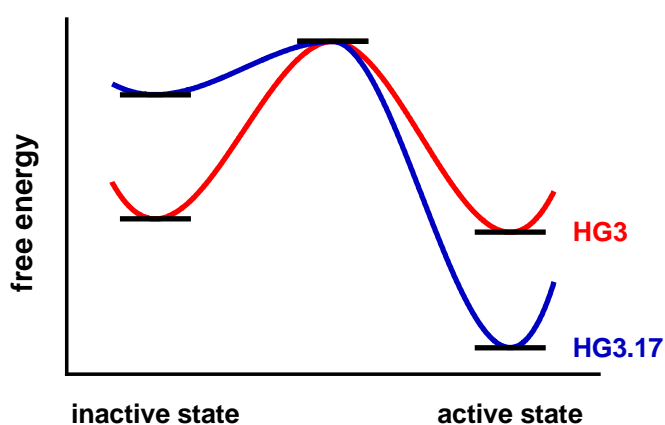
Figure 33: Evolution of conformational landscapes. HG3 (red) populates an active and an inactive state with almost equal energy. Evolution lead to selective stabilization of the active state in HG3.17 (blue).

2.3.1 Results

Activity dependence on pH and temperature.   Studying the effect of temperature on activity is of particular interest, since such analyses can provide insights into the enthalpic ($\Delta H^{\ddagger}$) and entropic ($\Delta S^{\ddagger}$) contributions to the activation energy of a reaction [164]. Natural enzymes, except the ribosome [31], are thought to accelerate their cognate reaction predominantly by reduction of $\Delta H^{\ddagger}$ [1,32,165]. Consistent with this precedent, HG3.17 and its evolutionary ancestors accelerate the Kemp elimination primarily by reducing $\Delta H^{\ddagger}$ (Figure 34).

The temperature dependence of HG3.17 is, however, complicated by an unexpected loss in activity starting at approximately ~20 °C, which limits the analyzable temperature range. As a consequence, the estimated reduction in $\Delta H^\ddagger$ is subject to large errors.

B

Thermodynamic parameters for the catalyzed and uncatalyzed Kemp elimination

	$\Delta G^\ddagger$ (kcal/mol)[b]	$\Delta H^\ddagger$ (kcal/mol)	$-T\Delta S^\ddagger$ (kcal/mol)[b]
background	22.8	15.9	7.9
HG3	13.3	9.9	3.4
HG3.7	11.4	7.5	3.9
HG3.17[a]	10.1	4.8	5.3

[a] Parameters for HG3.17 may be very error prone. [b] $\Delta G^\ddagger$ and $-T\Delta S^\ddagger$ at 25 °C.
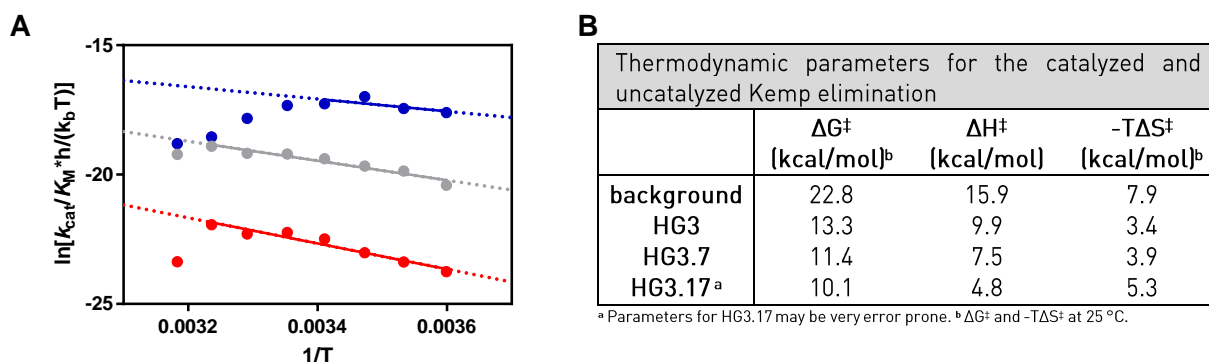
Figure 34: Temperature dependence of HG3 variants. **(A)** HG3.17 (blue) starts to lose activity around 20 °C, while HG3 (red) and HG3.7 (gray) lose activity around 35 °C. Due to the loss in activity, the Arrhenius plot of HG3.17 is only linear over a limited range. **(B)** Thermodynamic parameters for the catalyzed and uncatalyzed Kemp elimination. The parameters for HG3.17 are subject to large errors, due to the limited linear range of the Arrhenius plot.

The partially reversible loss in activity, which takes place within minutes, occurs far below the published melting point ($T_m = 53$ °C) of HG3.17 [4]. Besides this temperature-dependent activity loss, it was previously noted that HG3.17 loses activity at elevated pH values in a time-dependent fashion. The pH-dependent activity loss, which occurred with an apparent pK_a of 9.8, takes place on a similar timescale as the thermal inactivation. While it is possible that both phenomena are linked by the same process, their molecular origins are unknown.

Temperature and pH dependence may be linked to a slow conformational change. In order to test the origins of the activity loss at ambient temperature, additional denaturation analyses were performed using a thermal shift assay [166]. In this assay, a fluorogenic dye is employed that increases in fluorescence upon exposure to a hydrophobic environment like that presented by unfolding proteins. The assay was used to follow the temperature dependent denaturation of HG3.17. Two transitions were observed. The main event coincides with the T_M determined in thermal denaturation experiments monitored by circular dichroism [4]. In addition, a second process at ambient temperature was revealed, which was not observed previously (Figure 35A). This second process is characterized by a T_M of 38 °C, but starts at approximately ~20 °C, coinciding with the activity loss observed for HG3.17. Notably, this effect is less pronounced for the evolutionary ancestors of HG3.17, which matches their Arrhenius behavior. In addition, binding of the transition state analog 6-nitrobenzotriazole **1** shifts the T_M of the lower melting event to 47 °C (Figure 35B). It is hence likely that this process is connected

in some way to the active site and may be responsible for the observed loss in activity at ambient temperature.
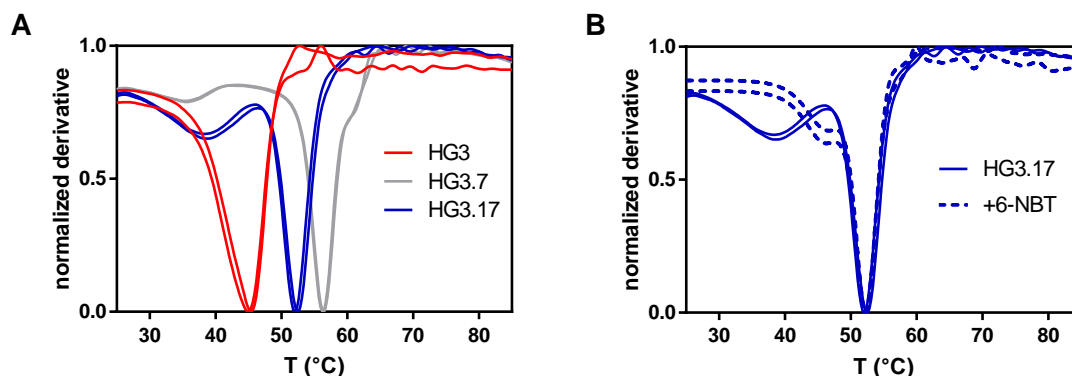
Figure 35: Thermal shift assay. **(A)** Melting curves of HG3 (red), HG3.7 (gray) and HG3.17 (blue) show a second transition in the range of 20 °C – 25 °C which was not observed by circular dichroism spectroscopy. **(B)** The first event is likely connected to the active site since binding of the inhibitor 6-nitrobenzotriazole 1 (6-NBT) increases its temperature (T_m = 38 °C to 46 °C).

In order to test whether the temperature and pH dependencies are caused by the same process, the effect of varying pH on HG3.17 was assayed in detail. Indeed, the first melting event is strongly affected by pH, with the apparent T_M decreasing with increasing pH (Figure 36A). The link between the pH- and temperature-dependent loss in activity and stability is further highlighted by the bell-shaped pH-rate profile of HG3.17 recorded at 4 °C (Figure 36B). While the lower pK_a did not change within error compared to the pH-rate profile at 20 °C, the higher pK_a increased from 7.8 at 20 °C to 9.8 at 4 °C. The lower pK_a of HG3.17 is likely associated with ionization of the catalytic base. However, no compelling explanation exists for the loss of activity at elevated pH. Due to the time scale of the process and the transition observed in the thermal shift assay, it is likely that the higher pK_a reports on a slow conformational change that is induced by either high pH or temperature.
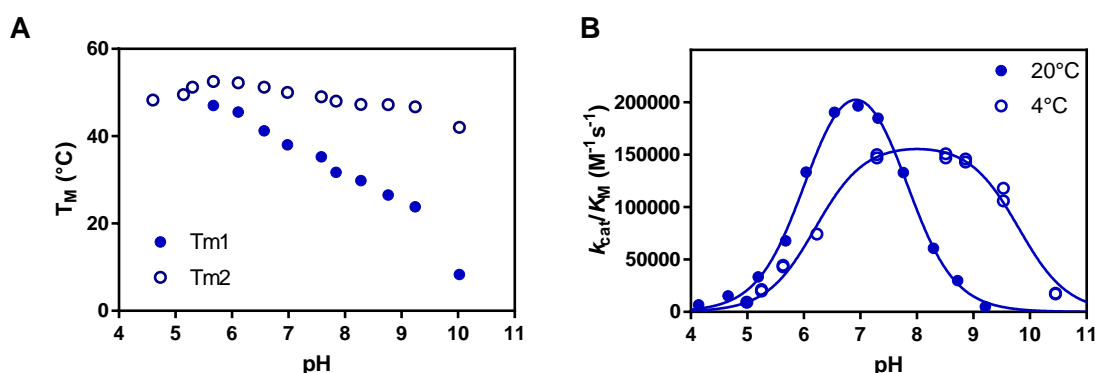
Figure 36: Effect of pH on the stability and activity of HG3.17. **(A)** pH-melt profiles of the two transitions observed in the thermal shift assay. **(B)** pH-rate profiles at 20 °C and at 4 °C indicate stabilization of the active form of the enzyme with decreasing temperature.

Proposed model for the pH-dependent conformational change. Similar effects of pH on the conformational landscape of proteins have been previously observed for other systems [167-170]. Typically, these effects are explained by a model involving equilibration between different enzyme states. In the simplest case, the model includes equilibration of the active (E_{act}) and inactive (E_{inact}) states, as well as the corresponding protonated (HE_{act}, HE_{inact}) species (Figure 37). Notably, the model separates the conformational equilibrium from pH effects. While equilibration between the active and inactive states only occurs between species with the same ionization state (k_1 and k_{-1}, k_2 and k_{-2}), pH effects are introduced via the acid dissociation constants of the active and inactive states ($K_{a,act}$, $K_{a,inact}$).
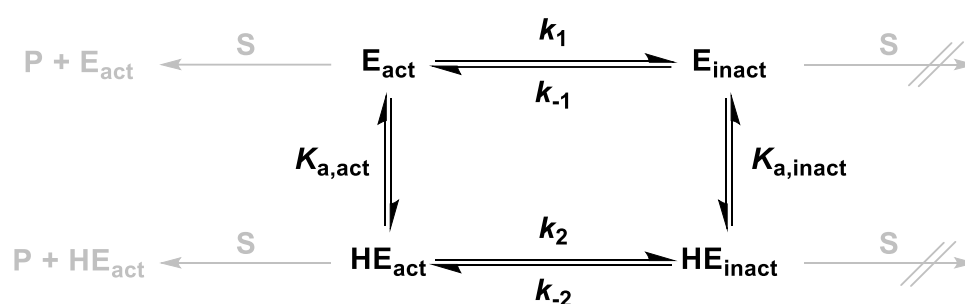
Figure 37: Proposed model of the conformational change. The enzyme equilibrates (k_1, k_{-1}, k_2, k_{-2}) between active (E_{act}, HE_{act}) and inactive (E_{inact}, HE_{inact}) conformations. The observed pH effects are caused by protonation or deprotonation of these states, which may have different proton affinities ($K_{a,act}$, $K_{a,inact}$). Only the active states convert substrate (S) into product (P).

For the proposed model, the following assumptions have to be made:

1. The enzyme equilibrates between an active and an inactive conformation;
2. The change in ionization of a key residue modulates the conformational equilibrium;
3. The change in ionization of that key residue does not directly influence enzyme activity;
4. The conformational change may result in a change in the microenvironment of the key residue which in turn may modulate its pK_a;

Protonation of the key residue should not be confused with protonation of the catalytic base, which instantaneously inactivates the enzyme. It is unlikely that protonation of the key residue directly modulates the activity (assumption 3), since the protonation equilibrium should be significantly faster than the rates of the conformational change. Based on the model, the following equation was derived for the observed rate of the conformational change (k_{obs}) [167]:

$$k_{obs} = \frac{k_1 K_{act} + k_2 [H]}{K_{act} + [H]} + \frac{k_{-1} K_{inact} + k_{-2}[H]}{K_{inact} + [H]} \tag{1}$$

For the reaction scheme shown in Figure 37, thermodynamics requires:

$$k_{-1} = k_{-2}\frac{K_{act}k_1}{K_{inact}k_2} \tag{2}$$

Substitution of k_{-1} in equation (1) leads to:

$$k_{obs} = \frac{k_1 K_{act}+k_2[H]}{K_{act}+[H]} + \frac{k_{-2}\frac{K_{act}k_1}{k_2}+k_{-2}k_{-1}[H]}{K_{inact}+[H]} \tag{3}$$

In the experiments described below, the pH-dependent conformational change was first analyzed by assaying the activity change of the enzymes at pH 5.0 following preincubation at different pH values. In order to determine the kinetics of the conformational change accurately, a more direct assay was subsequently developed in which the conformational change was monitored via tryptophan fluorescence at varying pHs after preincubation at pH 4.0 or pH 10.0.

Slow pH-dependent loss of activity. The ratio of active to inactive states was analyzed by following the change in catalytic efficiency of HG3 and HG3.17 after a pH jump. Due to the high background reaction rate at elevated pH values and the inactivation of the enzyme at low pH by protonation of the catalytic base, the conformational change was only assayed at pH 5.0 (Figure 24).

At pH 5.0, approximately 6% of the active site base is still deprotonated, and the background reaction is slow enough to allow measurements for up to 60 min with a substrate conversion below 10%. After the pH jump, HG3.17 generally relaxed more rapidly than HG3. While the conformational change in HG3.17 could be reliably analyzed, the errors for HG3 are more significant due to the slower rate of the conformational change. Nevertheless, for samples where the rate could be accurately determined, it was found that the rate was independent of the initial pH, in agreement with the proposed model. Furthermore, a lag phase was observed for samples preincubated at elevated pH, whereas the enzymes showed a burst if preincubated at low pH. By comparing the initial activity after preincubation at pH 7.0 with the initial activity after preincubation pH 4.2, it was found that HG3.17 is approximately 95% in its active state at neutral pH. For HG3, no saturation of the initial activity after preincubation at low pH values could be observed (Figure 38C). Thus, the amount of active state after preincubation at pH 7.0 cannot be deduced for HG3.
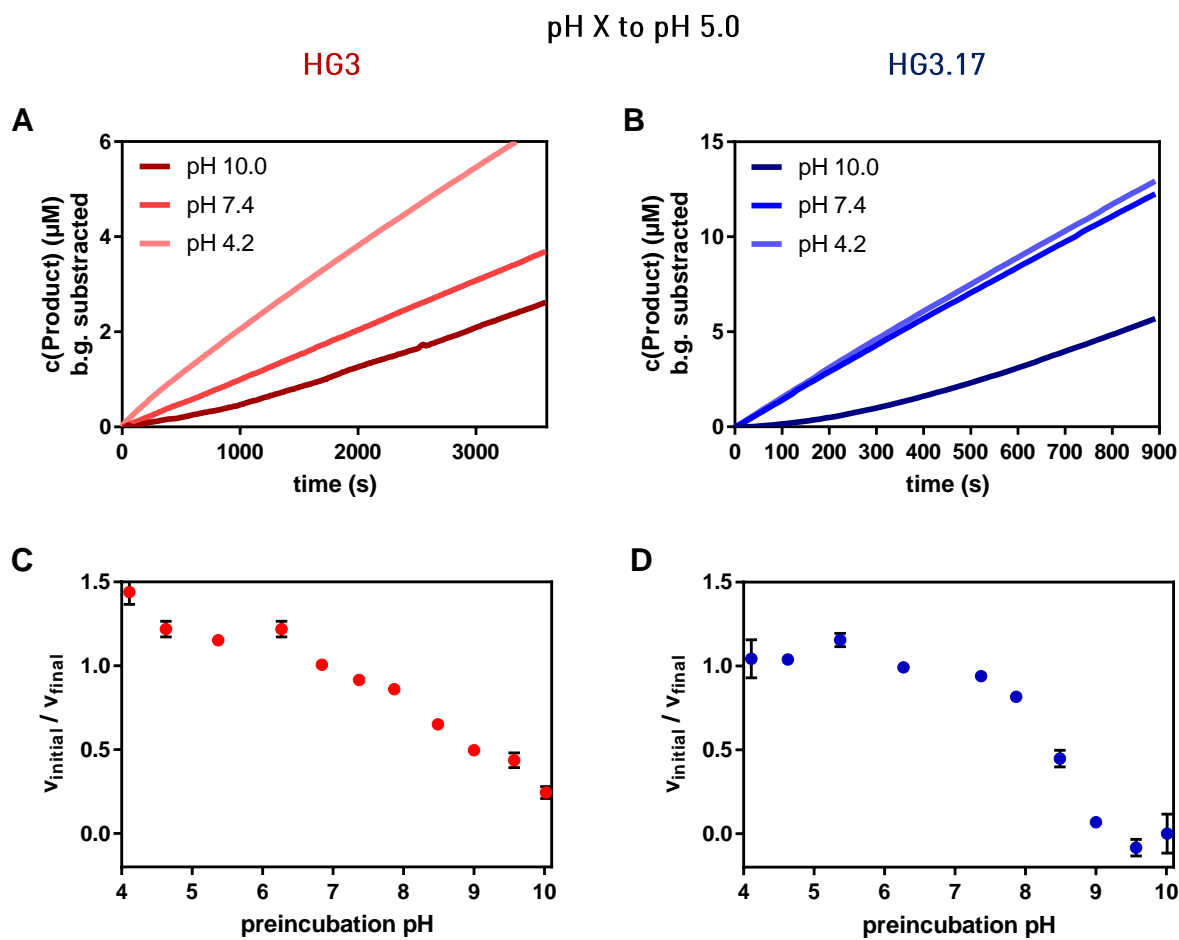
Figure 38: Change in activity of HG3 (red) or HG3.17 (blue) upon pH-jump. (A) HG3 or (B) HG3.17 were incubated at pH values ranging from 4 to 10 and the activity was assayed over time at pH 5.0. The ratio of initial velocity to the final velocity for (C) HG3 or (D) HG3.17 indicates that HG3.17 is predominantly in its active state at pH 7.0.

A pH-dependent slow conformational change. The kinetics of the conformational change could not be analyzed in detail by assaying the change in activity, since the activity assay is restricted to a small final pH range. A direct assay was therefore developed which relies on changes in tryptophan fluorescence, enabling analysis of the relaxation kinetics at any pH. The new assay is also more accurate than the activity-based assay since no background signals have to be subtracted. In order to increase the accuracy, emission was recorded ratiometrically (310 nm / 360 nm) for up to 60 min after excitation at 280 nm. The enzymes were either preincubated at pH 4.0 to assay the conformational change at alkaline pH, or at pH 10.0 to assay rates at acidic pH (Figure 39).
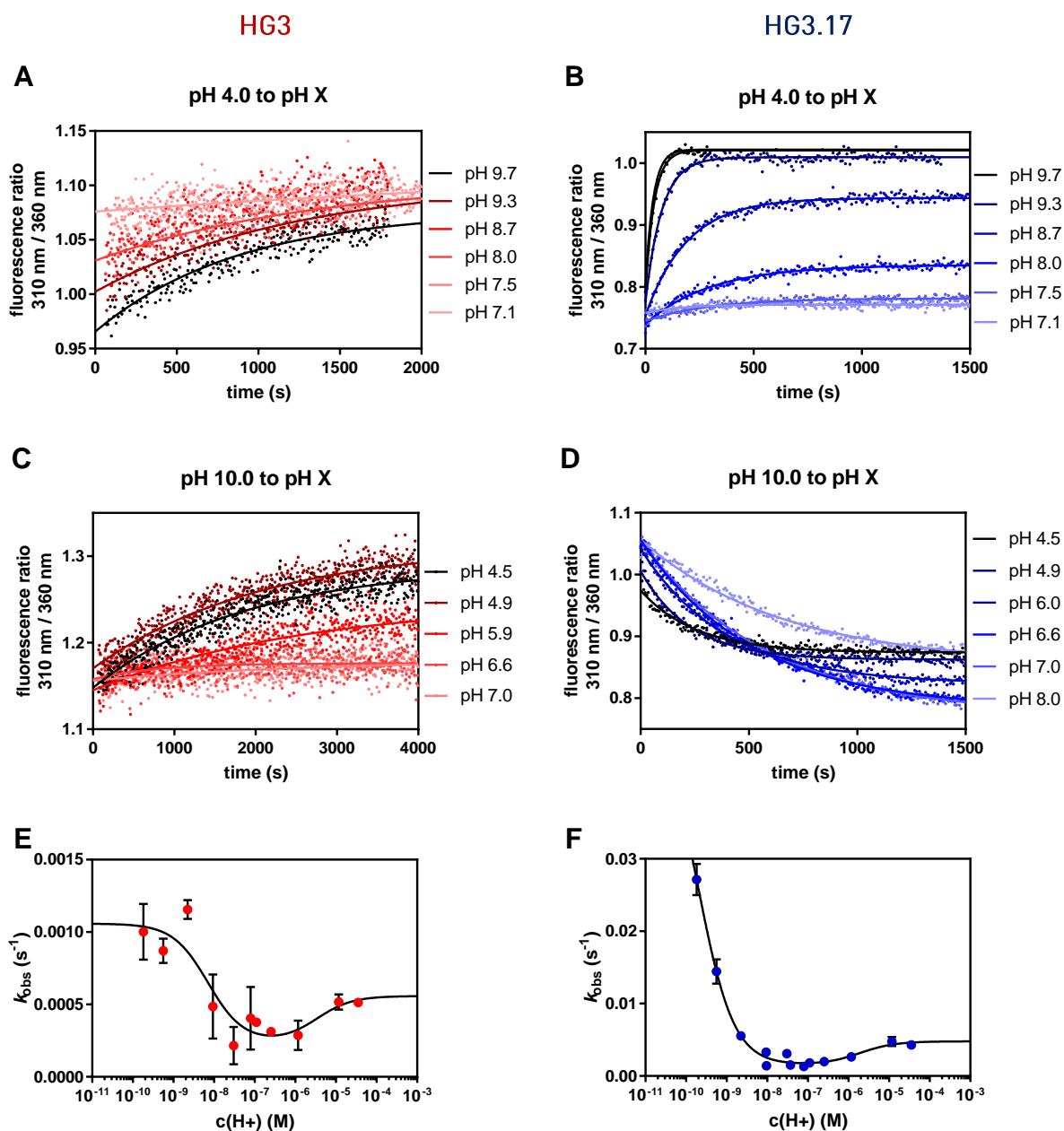
Figure 39: pH-dependence of the conformational change in HG3 (red) and HG3.17 (blue) monitored by tryptophan fluorescence. (A+B) Enzymes were incubated at pH 4.0 and rates were analyzed at the indicated alkaline pH. (C+D) Enzymes were incubated at pH 10.0 and rates were analyzed at the indicated acidic pH. (E+F) Fit of the observed rates to equation (3).

Again, HG3.17 generally relaxed faster than HG3. Global fitting of the kinetic data to the model shown in Figure 37 with *Kintek* (*KinTek Corporation*) support the conclusion that the conformational landscape was refined during evolution of HG3.17 (Figure 40). The equilibrium between the active and inactive states at pH 7.0 shifted from 4% active state for HG3 to 94% active state for HG3.17. This appears to be primarily due to the equilibrium between the protonated species, HE_{act} and HE_{inact}, which was shifted towards HE_{act} by 200 fold in the evolved variant compared to the computational design. This shift was achieved partially at the expense of a 20-fold more disfavored equilibrium for E_{act} between the deprotonated states (E_{act} and

E_{inact}). The change in equilibrium position, however, barely affects the overall ratio of the two states at pH 7.0 due to the elevated proton affinity of the active state ($K_{a,act}$) of $10^{-9.6}$. It should be noted that, though the general trends seem reliable, the rates for HG3 have rather large errors due to the long measurement times and weak signal.
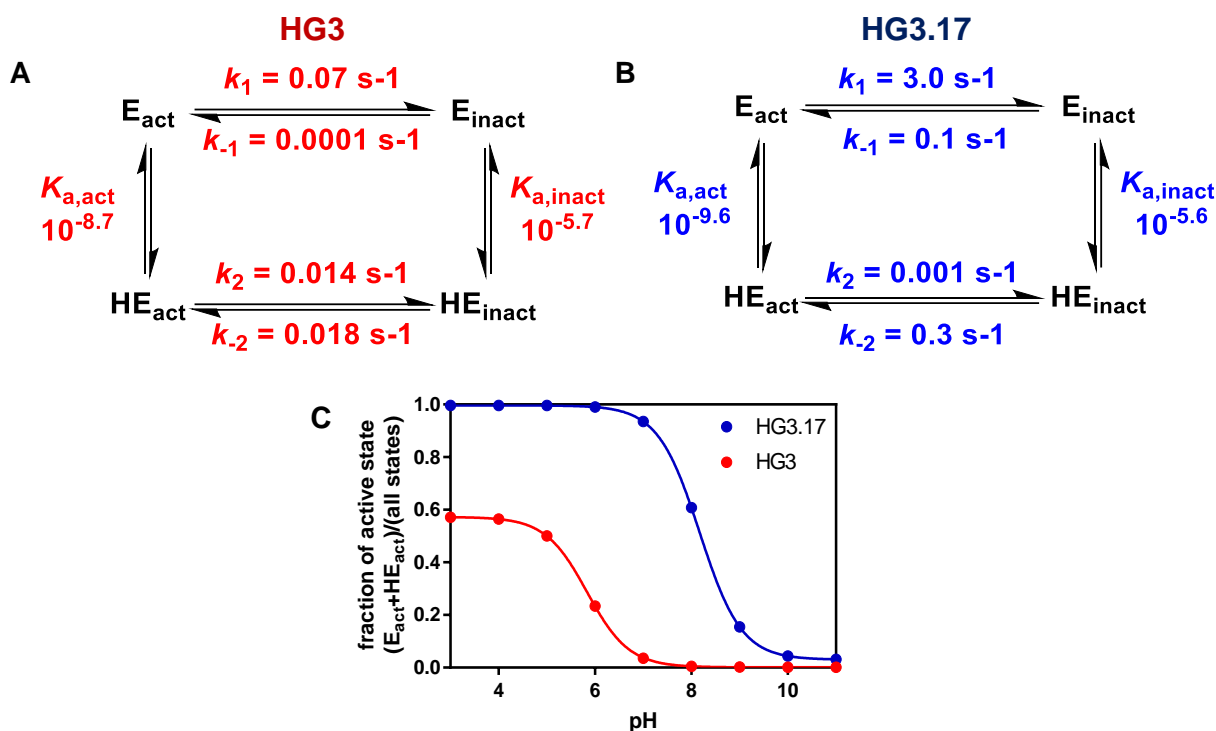
Figure 40: Conformational equilibria of **(A)** HG3 (red) and **(B)** HG3.17 (blue). The conformational equilibrium of HG3.17 is shifted towards the active state, primarily due to the 200-times more favored equilibrium of the protonated states for the active conformation. **(C)** A global fit of the date to the proposed model with *Kintek* indicates that at pH7.0, HG3.17 is almost entirely in its active state, while HG3 is predominantly in its inactive state.

Slow inhibitor binding and release of the inactive state. Due to the slow rate of conformational change of HG3.17 (k_{obs} = 0.16 min^{-1} at pH 5.9, Figure 39F), it was possible to analyze the binding of the transition state analog 6-nitrobenzotriazole **1** to either the active (pH 4.0) or inactive (pH 10.0) form. Binding was also analyzed as previously described by preincubation at pH 6.0, referred to hereafter as the native state. To eliminate pH effects on the inhibitor binding rates that are unrelated to the conformational states, the enzyme solutions were mixed in a 1:5 ratio with strongly buffered inhibitor solutions, resulting in a final pH of 6.0 in all cases. Generally, the native state behaved similar to the active state which is in good agreement with the kinetic data indicating that the native state at pH 6.0 is mostly in the active conformation. Triphasic kinetics, comprised of a fast, slow and very slow phase were observed for all samples at elevated inhibitor concentrations (Figure 41).
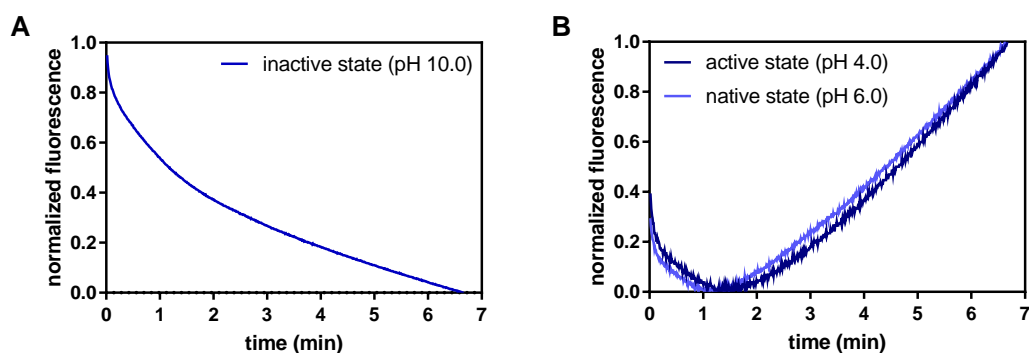
Figure 41: Binding curves of the (A) inactive and (B) active and native state at 800 µM inhibitor concentration at pH 6.0. Clear transients were only observed at high inhibitor concentrations. While the signal may be caused by nonspecific binding, the slower phases may also involve binding to the inactive state and thus require higher inhibitor concentrations.

The fast phase has rate constants in the range of $100\,\text{s}^{-1}$ for the active and $1\,\text{s}^{-1}$ for the inactive state and is likely associated with binding of 6-nitrobenzotriazole **1**. The slow phase occurs with rate constants in the range of $0.01\,\text{s}^{-1}$. No clear curvature was observed for the very slow phase. The rate constants can thus only be estimated to be below $10^{-4}\,\text{s}^{-1}$. The origins of the slow and very slow phase are less clear and will be discussed later. The binding kinetics of 6-nitrobenzotriazole **1** to the active state according to the fast phase reproduced the previously reported kinetics for the native state (Table 4 and Figure 42).
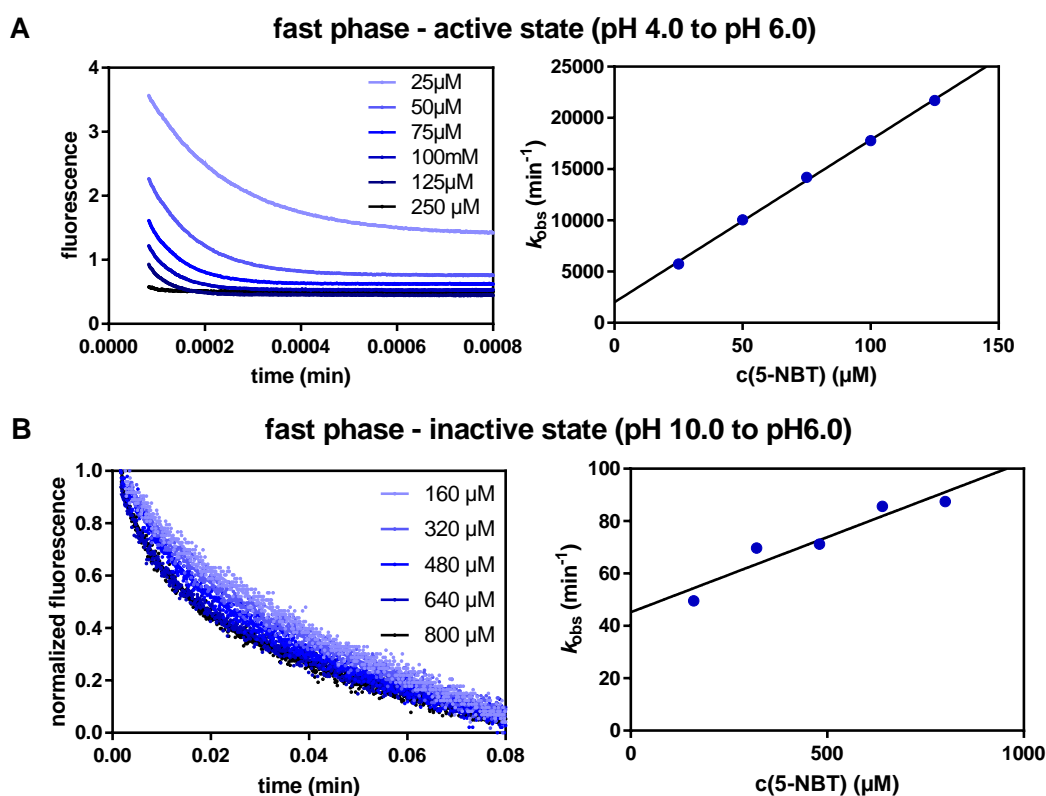
Figure 42: Inhibitor binding at pH 6.0 to the active state (preincubated at pH 4.0) and the inactive state (preincubated at pH 10.0). (A) The fast phase of the active state reproduced previous data on the native state [4]. (B) The fast phase for the inactive state was significantly slower and results in a higher K_D.

Table 4: Inhibitor binding kinetics to the different states of HG3.17				
	preincubation	k_{on} ($\mu M^{-1}s^{-1}$)	k_{off} (s^{-1})	K_D (μM)
active state	pH 4.0	2.6	34	13
native state[4]	pH 6.0	4.4	30	6
inactive state	pH 10.0	0.0011	0.76	700

Binding of the inhibitor to the inactive state was significantly slower and resulted in an 70-fold increased K_D. Thus, binding of the inhibitor to the inactive state was significantly weakened compared to the active state due to a 3,000-fold reduced k_{on}, indicating rate-limiting inhibitor binding for that state. It is thus likely that the inactive state does not accelerate the Kemp elimination because the substrate, 5-nitrobenzisoxazole 2, binds only weakly to that state.

The origins of the slow and very slow phases of the triphasic inhibitor binding kinetics to HG3.17 are not entirely clear. While further experiments will be required to assign these phases, the acquired data suggest that the slow phase corresponds to the equilibration of the free inactive and active states, while the very slow phase reports on equilibration of the inhibitor bound forms (Figure 43).
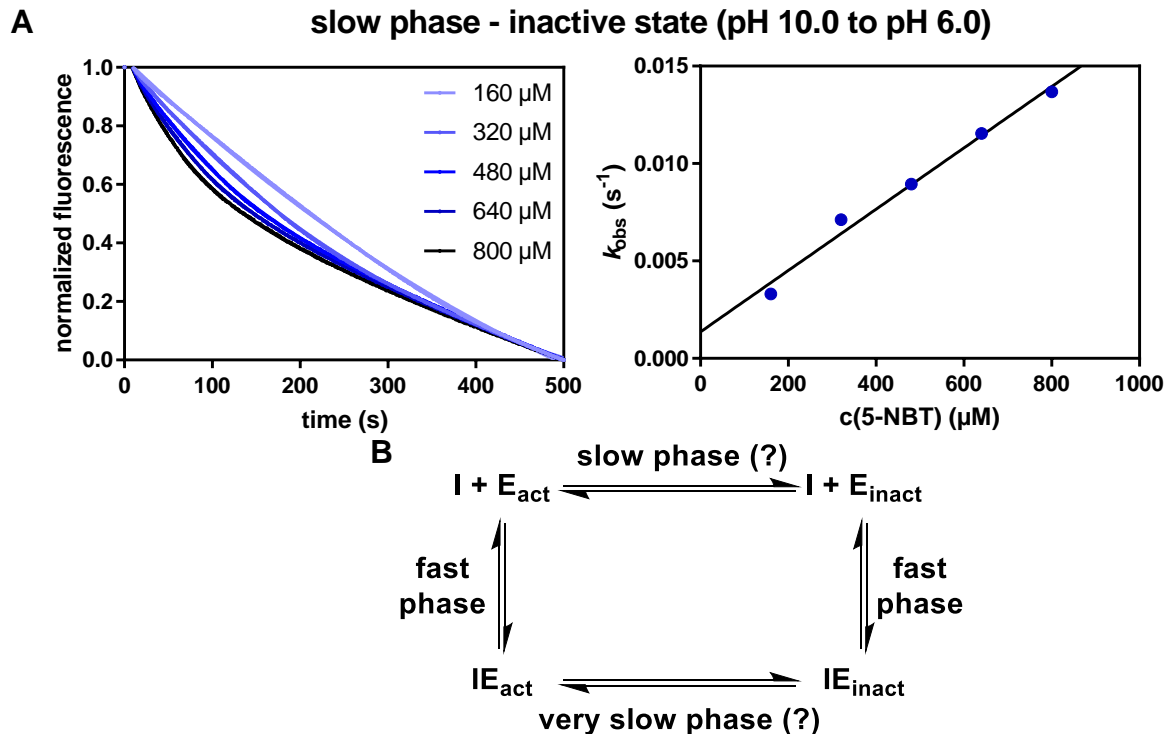
Figure 43: Proposed model for the slow and very slow inhibitor binding phases. (A) Observed slow and very slow phases starting from the inactive state. (B) The slow phase may be assigned based on the known rate of interconversion of the free enzyme. If the very slow phase corresponds to the remaining equilibrium, the model can explain the observed change in the sign of the amplitude for that phase.

The pronounced slow and very slow phases were not previously seen [4]. The high inhibitor concentrations ($\leq 800\ \mu M$ vs $\leq 125\ \mu M$), approaching the K_D of the inactive state, may have led to the new transients. Extrapolation of the rate dependence of the slow phase on the inhibitor concentration gives a k_{obs} of 0.11 min^{-1} at zero inhibitor concentration, which is close to the k_{obs} of 0.16 min^{-1} at pH 5.9 determined by pH-jump experiments. The very slow phases have opposite signs for the inactive and active state. Since inhibitor binding drastically decreases fluorescence, presumably through quenching, the observed increase for the very slow phase of the active state might indicate inhibitor release according to (4).

$$I + E_{act} \xrightleftharpoons{\text{fast binding}} IE_{act} \xrightleftharpoons{\substack{\text{very slow} \\ \text{phase (?)}}} IE_{inact} \xrightleftharpoons{\text{fast release}} I + E_{inact} \tag{4}$$
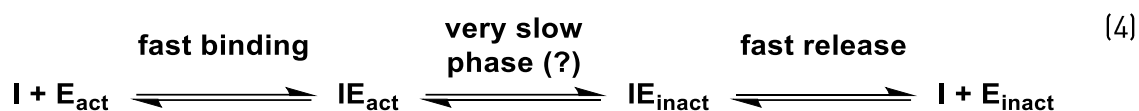
If the very slow phase corresponds to the equilibration of the inhibitor bound species (IE_{act} and IE_{inact}), very slow inhibitor release would be expected for that phase for reactions starting with the active state. Accordingly, starting from the inactive state, additional inhibitor binding may be observed after very slow equilibration of the inhibitor bound species.

2.3.2 Discussion

The conformational landscape of Kemp eliminase HG3.17 has been drastically remodeled during evolution. By refinement of the conformational equilibrium, the active state became significantly more populated in HG3.17. Inhibitor binding assays indicate that the binding of the ligand was slowed down approximately 2,000-fold in the inactive state compared to the active state of HG3.17. Thus, the inactive state binds the substrate only very weakly. By monitoring the changes in tryptophan fluorescence upon a jump in pH, it was determined that HG3 is 4% in its active state at pH 7.0, while 94% of HG3.17 is active. Importantly, transitions with similar rates were observed when activity was monitored directly. The activity-based analysis also indicated that HG3.17 is predominantly in its active state. Though the conformational equilibrium was clearly shifted during evolution, the error, particularly for HG3, is rather high. Nonetheless, the kinetic data suggest that the overall population of active states increased by as much as 20 fold through evolution.

Independent measurement by NMR spectroscopy may shed more light on the ratios of active to inactive states in HG3 and HG3.17. Preliminary HSQC-NMR experiments, performed

by Vy Nguyen, Renee Otten and Dorothee Kern (U Brandeis), were used to analyze the backbone amide peaks and confirm the observation that the proteins exist in two different conformations (Figure 44).
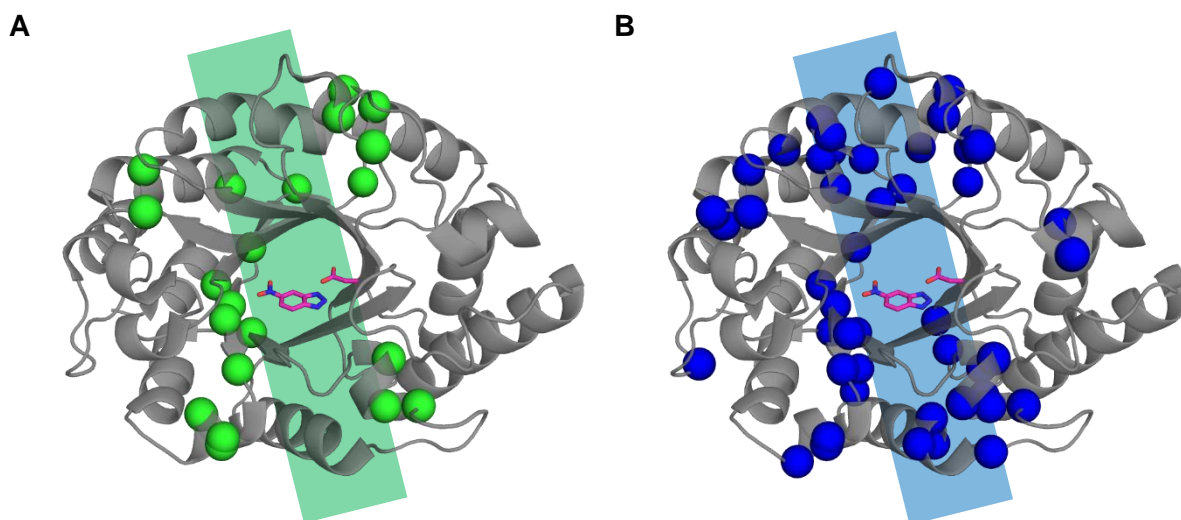
Figure 44: NMR analysis indicates that the conformational change in HG3.17 affects a cross section (colored square) of the protein. Splitting of the amide bond peaks was observed during HSQC-NMR at either **(A)** elevated pH or temperature (green) or **(B)** upon inhibitor binding (blue). The general base and inhibitor are depicted in magenta. Data from Vy Nguyen, Renee Otten and Dorothee Kern (U Brandeis)

Binding of the inhibitor or perturbation of either the pH or temperature leads to splitting of the amide peaks for a cross section of residues throughout the protein. These data thus suggest that the conformational change globally affects the protein, potentially by a twisting motion. Further atomic resolution structural data obtained by either NMR spectroscopy or room-temperature X-ray crystallography may shed further light on the precise mechanism of this conformational change.

The catalytic efficiency at 20 °C of HG3 improved by approximately 500 fold during directed evolution. Approximately 20 fold may be attributed to population of the active state through remodeling of the conformational landscape. The results from Chapter 2.2 suggest that refinement of the active site's electrostatics had a large impact on the remaining improvement. This study demonstrates the remodeling of conformational landscapes in a *de novo* designed enzyme. During evolution of HG3.17, conformational ambiguity was almost entirely eradicated. The catalyst is almost exclusively in a catalytic competent state at neutral pH and ambient temperature. While conformational uncertainty is generally not desired during design [30], stabilization of the productive conformational state during evolution may have led to significant rate acceleration in HG3.17 [4].

2.4 Materials and methods

2.4.1 Chemical Synthesis

Substrate synthesis. Substituted benzisoxazoles were synthesized as described previously from a salicaldehyde via oxime formation and PPh3/DDQ promoted cyclization [4,62,171,172]. The synthesis of the deuterated substrate 3-D-5-nitrobenzisoxazole **5** was performed as follows:

Synthesis of 2-(1,3-dithian-2-yl)phenol 6 according to [173]
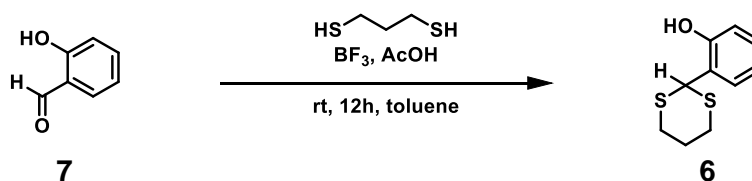
0.56 ml (4.5 mmol, 1.1 eq) Boron trifluoride ethyl etherate were added to a solution of 500 mg (4.09 mmol, 1 eq) 2-hydroxybenzaldehyde **7** and 0.45 ml (4.5 mmol, 1.1 eq) propane-1,3-dithiol in 12 ml pure acetic acid/toluene mixture (1:2) under inert conditions. The reaction mixture was stirred overnight at room temperature. After dilution with 10 ml ethylacetate and addition of 12 ml water the organic layer was washed two times with 5 ml water, three times with 5 ml sodium bicarbonate and two times with 5 ml brine. The combined organic phases were dried over anhydrous $MgSO_4$ and the solvent was removed by evaporation under reduced pressure. The raw product was used without further purification. 862.9 mg (4.06 mmol, 99%) of 2-(1,3-dithian-2-yl)phenol **6** were yielded as a colorless solid with traces of impurities.

Yield: 862.9 mg, 4.06 mmol, 99%

1H NMR: (300 MHz, CDCl3) δ (ppm) = 7.29 (dd, J = 7.9, 1.7 Hz, 1H), 7.25 – 7.18 (m, 1H), 6.91 – 6.85 (m, 2H), 6.33 (s, 1H), 5.41 (s, 1H), 3.17 – 2.83 (m, 4H), 2.27 – 1.82 (m, 2H)

EI/MS: calculated: 182.03 [M]+, found: 182.03 [M]+

Synthesis of 1-D-2-(1,3-dithian-2-yl)phenol 8 according to [173]
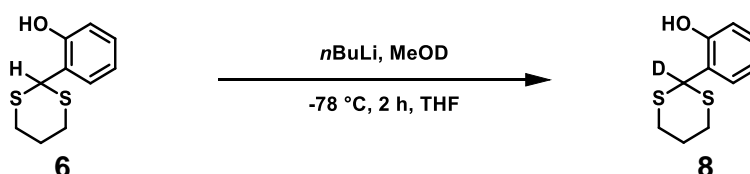
A solution of 10.0 ml (25.00 mmol, 2.4 eq) *n*-butyllithium in hexane were added to a solution of 2.21 g (10.42 mmol, 1 eq) 2-(1,3-dithian-2-yl)phenol **6** in 100 ml anhydrous THF at -78 °C for 1 h under inert conditions. After stirring for an additional hour, the reaction was warmed to 0 °C for 5 min and then cooled to -78 °C. Stirring was continued for 1 h and 2.1 ml (52.10 mmol, 5 eq) deuterated methanol were added. The reaction mixture was allowed to warm to room temperature and was stirred for 30 min. 200 ml 0.05 M hydrochloric acid was added and the solvent was removed under reduced pressure. The residue was extracted three times with 50 ml DCM/pentane (1:1) and the organic phase was washed three times with 30 ml sodium bicarbonate solution, two times with 30 ml water and one time with 30 ml brine. The organic layer was dried over MgSO₄ and solvent was removed under reduced pressure to give 2.18 g (10.23 mmol, 98%) 1-D-2-(1,3-dithian-2-yl)phenol **8**, which was used without further purification.

Yield: 2.18 g, 10.23 mmol, 98%

¹H NMR: (300 MHz, CDCl3) δ (ppm) = 7.29 (dd, J = 7.9, 1.6 Hz, 1H), 7.24-7.18 (m, 1H), 6.91 – 6.85 (m, 2H), 6.32 (s, 1H), 3.12 – 2.85 (m, 4H), 2.29   1.77 (m, 2H)

Synthesis of 1-D-2-hydroxy-benzaldehyde 9 according to [173]
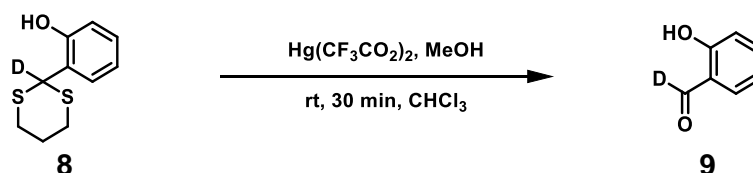
3.00 g (7.04 mmol, 1.5 eq) mercury trifluoroacetate were added to a solution of 1.00 g (4.69 mmol, 1.0 eq) 1-D-2-(1,3-dithian-2-yl)phenol **8** in chloroform/methanol (100:1) and the solution was stirred for 1h at room temperature. The reaction mixture was filtered and solvent was removed under reduced pressure. The residue was filtered over silica using ethylacetate and solvent was removed under reduced pressure to give 154 mg (1.25 mmol, 27%) of 2-hydroxy-benzaldehyde-1-d **9** as a slightly yellow liquid.

Yield: 154 mg, 1.25 mmol, 27%

¹H NMR: (300 MHz, CDCl3) δ (ppm) = 11.02 (s, 1H), 7.63 – 7.43 (m, 2H), 7.05-6.98 (m, 2H)

Synthesis of 1-D-2-hydroxy-benzaldehyde oxime 10 according to [171]
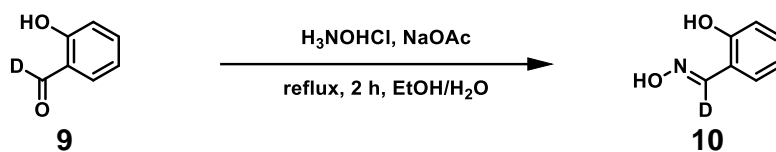
A mixture of 2.19 g (17.8 mmol, 1.0 eq) 1-D-2-hydroxy-benzaldehyde **9** and 2.47 g (35.6 mmol, 2.0 eq) hydroxylamine hydrochloride in 210 ml ethanol/water (3:1) was heated at 80 °C for 2 min. After addition of 4.84 g (35.6 mmol, 2.0 eq) sodium acetate-trihydrate in 70 ml water the reaction mixture was heated to reflux for 2 hours. The reaction mixture was allowed to cool to room temperature and 140 ml water were added before it was cooled to 0 °C for 1 h. After filtration 2.31 g (16.1 mmol, 90%) 1-D-2-hydroxy-benzaldehyde oxime **10** were obtained as a colorless solid.

Synthesis of 3-D-benzisoxazole 11 according to [171]
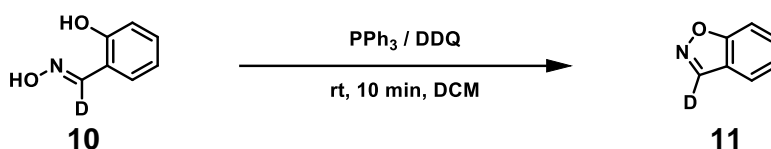
A mixture of 6.33 g (24.15 mmol, 1.5 eq) triphenylphosphine and 5.48 g (24.15 mmol, 1.5 eq) 2,3-dichloro-5,6-dicyano-1,4-benzochinon was dissolved in 80 ml DCM. The mixture was stirred for 5 min at room temperature and 2.31 g (16.1 mmol, 1.0 eq) 1-D-2-hydroxy-benzaldehyde oxime **10** was added. The solution was stirred for 10 min at rt and the solvent was evaporated under reduced pressure. The residual solid was extracted three times with 50 ml ethylacetate. The combined organic phases were dried over anhydrous $MgSO_4$ and the solvent was removed by evaporation. The crude product was purified by column chromatography (silica gel, EtOAc/CyHex = 1:2, R_f = 0.43) to yield 1.90 g (15.8 mmol, 98%) of 1-D-benzisoxazole **11** as a colorless solid.

Synthesis of 3-D-5-nitrobenzisoxazole 5 according to [174]
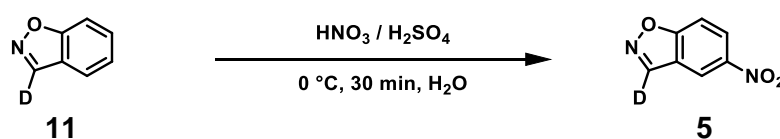
A mixture of 1.21 ml (17.4 mmol, 1.1 eq) HNO_3 and 0.98 ml (17.4 mmol, 1.1 eq) H_2SO_4 was added to a solution of 1.90 g (15.8 mmol, 1.0 eq) 1-D-benzisoxazole **11** in 8.8 ml H_2SO_4 at 0 °C over 15 min. The reaction was quenched after 15 min by addition of 50 ml ice water. The crude

product was collected as precipitate. Column chromatography (ethyl acetate/hexane 1:3, Rf = 0.52) afforded 1.54 g (9.33 mmol, 59%) of 3-D-5-nitrobenzisoxazole **5** as a colorless solid.

Yield: 1.54 g, 9.33 mmol, 59%

^{1}H NMR: (300 MHz, CDCl3) δ (ppm) = 8.84 (dd, 1H), 8.56-8.52 (dd, 1H), 7.88-7.86 (dd, 1H)
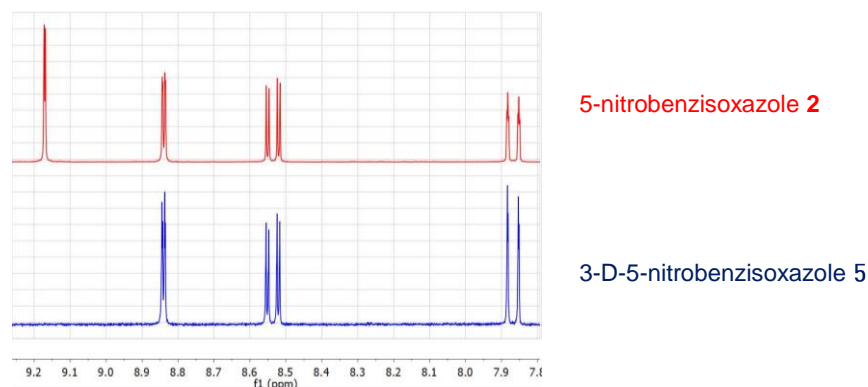
Figure 45: ^{1}H-NMR spectra of 5-nitrobenzisoxazole **2** and 3-D-5-nitrobenzisoxazole **5**. The proton at 9.17 ppm was completely replaced by deuterium.

2.4.2   Biochemical methods

Protein production and purification.   BL21-Gold(DE3) cells were transformed with the pET11 vector harboring the gene of interest and plated on LB agar plates (with 100 µg/ml ampicillin). A single colony was used to inoculate an overnight culture of LB medium containing 100 µg/ml ampicillin. 500 ml of LB medium containing 100 µg/ml ampicillin, were inoculated with the overnight culture in a ratio of 1:100 from and grown to an OD of 0.5 at 37 °C. Protein production was induced with 1 mM IPTG and cells were incubated at 18 °C overnight. The cells were harvested and the cell pellet was frozen at -80 °C or directly processed for protein purification. The cell pellet was resuspended in lysis buffer (20 ml 50 mM tris-HCl, 0.5 M NaCl, 10 mM imidazole, 100 µL protease inhibitor cocktail (*Sigma*), 1 mg/ml lysozyme, DNAseI, pH 7.4). The cell suspension was sonicated for 1 h in a sonication bath at 4 °C (*Telesonic Ultrasonics*). After centrifugation (18.000 rpm, 20 min, 4 °C). The soluble fraction was applied to 3 ml Ni-NTA slurry (*Qiagen*), washed with 10 mM imidazole before elution of the protein with 300 mM imidazole, each in 50 mM Tris-HCl, 0.5 M NaCl, pH 8. The buffer was exchanged to 20 mM sodium phosphate, 20 mM NaCl, pH 6.0 by washing three times in an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*). Proteins were further purified by anion exchange chromatography (MonoS column, *GE Healthcare*) and eluted with a salt gradient (20 mM to 1 M NaCl, 20 mM sodium phosphate, pH 6.0). Fractions absorbing at

280 nm were analyzed by SDS-PAGE (Phastsystem, *Pharmacia Biotech*) using a low molecular weight marker (*GE Helthcare*). Fractions containing the target protein were pooled and concentrated using an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*) if necessary. Protein concentrations were determined by measuring the absorbance at 280 nm using a calculated extinction coefficient [175]. The following extinction coefficients were determined for some representative enzymes: ε(HG3) = 57,410 M^{-1}cm^{-1}; ε(HG3.17) = 51,910 M^{-1}cm^{-1}. Protein masses were confirmed by mass spectrometry. For mass determination, protein samples were exchanged into 0.1% acetic acid with Vivaspin 500 centrifugal filters (*Sartorius*) and measured by ESI-MS on a Daltonics maXis ESI-Q-TOF mass spectrometer (*Bruker*). Mass spectra were deconvoluted using the MaxEnt1 software. All masses corresponded to the expected sequence lacking the N-terminal methionine.

UV-Vis kinetics. Steady-state kinetics for enzymatic conversion of 5-nitrobenzisoxazole **2** were measured spectrophotometrically at a pH value of 7.0 and 20 °C. The lower temperature of 20 °C compared to previous experiments (27 °C) was chosen because HG3.17 starts losing activity already at 25 °C (see Chapter 2.3) although thermal denaturation only occurs at much higher temperatures [4]. Product formation was monitored by following the change in absorbance at 380 nm using a Lambda 35 UV/VIS spectrometer (*PerkinElmer*). Initially, 100 µl of substrate solution in methanol (to a final concentration of 125 µM to 1.5 mM) were added to 890 µl of reaction buffer (50 mM sodium phosphate, 100 mM NaCl, pH 6.82) in a cuvette with a pathlength of 10 mm. After mixing, the final pH reached 7.00. Measurements at higher substrate concentrations were not possible because of the limited solubility of the substrate. Measurement of the background rate was performed for 30 to 40 s after mixing. Immediately afterwards, 10 µl of enzyme (5 nM) were added to the mixture. Linear curve fitting of the initial slopes and nonlinear fitting to the Michaelis-Menten equation to calculate k_{cat} and K_M were performed with the Kaleidagraph software (*Synergy Software*).

Kinetic isotope effect. Kinetic isotope effects were determined by assaying activity of Kemp eliminase variants with deuterated and non-deuterated 5-benzisoxazole as described for the UV/Vis kinetics following equation (5). Three different protein batches were analyzed twice each. Errors represent the standard deviations of six independent measurements.

$$\text{kinetic isotope effect} = \frac{k_{\text{H}}}{k_{\text{D}}} \tag{5}$$

Inhibition kinetics. Kinetics were determined in analogy to the UV-Vis kinetics at 125.8 µM 5-nitrobenzisoxazole **2** with varying concentrations of 6-cyanobenzotriazole **4**. HG3 variants (1 nM to 200 nM final concentration, depending on their respective activity) were preincubated with varying concentrations of 6-cyanobenzotriazole **4** (0.3 µM to 2500 µM final concentrations) in 50 mM phosphate, 100 mM NaCl, 10 % MeOH, 1% DMSO, pH 7 and reactions were initiated by addition of 5-nitrobenzisoxazole **2** (125.8 µM final concentration). Product formation was monitored as described previously. IC_{50} values were determined by curve fitting (Hill-Slope model with the rate at infinite inhibitor concentration set to zero) and assumed to be equal to the K_i value at c(substrate) << K_M according to the Cheng-Prusoff equation [176]:

$$K_i = \frac{IC_{50}}{\left(1 + \frac{c(\text{substrate})}{K_M}\right)} \tag{6}$$

pK_a-determination. Concentrated DMSO solutions of 5-cyanobenzotriazole **4** were diluted into aqueous buffers containing 100 mM sodium phosphate, 100 mM sodium acetate and 200 mM NaCl aqueous buffers between pH 4.0 and pH 9.0 to a final concentration of 10 µM and 1% (v/v) DMSO. The absorbance change at 285 nm was fitted to equation (7) and gave a pK_a of 6.52±0.02.

$$A = \frac{A_{max}}{(1 + 10^{pK_a - pH})} - A_{min} \tag{7}$$

pH-rate profile. The dependence of k_{cat}/K_M on pH was determined for HG3.17 at 20 °C and 4 °C. Enzymes were preincubated for at least 30 min for reactions at 20 °C and at least 2 h for reactions at 4 °C in buffer with the final pH. Acetate buffer (50 mM sodium aceteate, 100 mM NaCl) was employed for pH values from pH 4 to 5.5 and bis-Tris propane buffer (50 mM bis-Tris propane, 100 mM NaCl) from pH 6 to 9.5. The background rate without enzyme was determined for each pH value and subtracted. Due to the dependence of the apparent extinction coefficient of the product, 5-nitro-2-hydroxybenzonitrile at 380 nm, on its protonation state, the coefficient was corrected using equation (8) with a product pK_a of 3.98 and a $\Delta\varepsilon_{max}$ of 15,800 M^{-1}cm^{-1}.

$$\Delta\varepsilon_{380nm} = \frac{\Delta\varepsilon_{max}}{(1 + 10^{pK_a - pH})} \tag{8}$$

The bell-shaped pH-rate profiles of HG3.17 were fitted to equation (9) with Prism 7 (*GraphPad*) to obtain the two ionization constants, pK_{a1} and pK_{a2}.

$$k_{cat}/K_M = \frac{(k_{cat}/K_M)_{max}}{(1 + 10^{pK_{a1}-pH} + 10^{pH-pK_{a2}})} \qquad (9)$$

Temperature-dependent activity assay. The temperature-dependent activity changes of the HG3 variants were analyzed on an SX.18MV stopped-flow instrument (*Applied Photophysics*) by mixing proteins (5 nM – 30 μM, final concentration) in 95 mM sodium phosphate, 200 mM NaCl, pH 7.5 in a 1:1 ratio with 5-nitrobenzisoxazole **2** (50 μM, final concentration) in 5 mM sodium phosphate, 20% MeOH, pH 1.26 and the change in absorbance at 380 nm was measured over 1 s. $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ were determined according to equation (10) for the linear region of the Eyring-plot. Temperatures were assayed between 4 °C and 45 °C. The k_{cat}/K_M values of HG3 and HG3.7 were only included up to 35 °C for fitting. Likewise, the k_{cat}/K_M values of HG3.17 were only fitted between 4 °C and 20 °C.

$$\ln\left(\frac{\frac{k_{cat}}{K_M} \cdot h}{k_B T}\right) = -\frac{\Delta H^{\ddagger}}{R} \cdot \frac{1}{T} + \frac{\Delta S^{\ddagger}}{R} \qquad (10)$$

Melting curves. Melting curve analyses were performed with 20 μl samples containing 10 μM protein in 50 mM sodium acetate, 50 mM bis-Tris propane and 100 mM NaCl at varying pH prepared with the thermal shift assay kit (*ThermoFisher Scientific*). Melt curves between 4 °C and 99 °C were recorded with a StepOnePlus Real-Time PCR thermocycler (*ThermoFisher Scientific*) according to the user's manual. Temperature data below 20 °C are considered unreliable, because the thermocycler cannot pre-cool the sample holder. Samples therefore warmed to room temperature prior to analysis. Nonetheless, no additional transitions were observed in that range.

Activity-based pH-jump assay. The activity-based analysis of the conformational change was performed on a Lambda 35 Uv/Vis spectrometer (*PerkinElmer*). After preincubation of 10 μM HG3 or 1 μM HG3.17 at 20 °C in buffer (50 mM bis-Ttris propane, 50 mM sodium acetate and 100 mM NaCl) at pH 4.0 or 10.0, the enzymes were diluted 1:100 into the assay buffer (250 mM sodium acetate, 100 mM NaCl, pH 5.0) supplemented with 10% MeOH and a final concentration of 5-nitrobenzisoxazole **2** of 100 μM. Under these conditions, substrate

conversion was below 10% after 1 h incubation with HG3 or 15 min incubation with HG3.17. The recorded lag or burst kinetics were fitted to equation (11) with Prism 7 after subtraction of the linear background reaction [167].

$$c(\text{Product}) = v_{\text{final}} * t - (v_{\text{final}} - v_{\text{initial}}) * (1 - \exp(-k_{\text{obs}} * t)/k_{\text{obs}} \qquad (11)$$

pH-jump assay monitoring tryptophan fluorescence. The tryptophan fluorescence-based analysis of the conformational change was performed with a spectrofluorimeter (*Photon Technology International*). Upon excitation at 280 nm, the 310 nm to 360 nm emission ratio was recorded. 10 µM of HG3 or HG3.17 were preincubated at 20 °C in buffer (50 mM bis-Tris propane, 50 mM sodium acetate and 100 mM NaCl) at pH 4.0 or 10.0. Changes in fluorescence were recorded upon 1:10 dilution of the protein in buffer (50 mM bis-Tris propane, 50 mM sodium acetate and 100 mM NaCl) at varying pH values. Fluorescence was recorded alternatingly for 1 s at each wavelength. The shutter was subsequently closed for 1 s after each time point to prevent bleaching. The final pH after dilution was measured for analysis of the pH dependence.

Inhibitor binding. Stopped-flow kinetics of inhibitor binding to HG3 and HG3.17 were measured on an SX.18MV stopped-flow instrument (*Applied Photophysics*) by mixing a solution of 6 µM enzyme in buffer (50 mM bis-Tris propane, 50 mM sodium acetate, 100 mM NaCl) at pH 4.0, 6.0 or 10.0 with inhibitor solutions (30-1000 µM 6-nitrobenzotriazole **1**, 250 mM MES, 50 mM sodium phosphate, 1.2% DMSO, pH 6.0) at a 1:5 ratio. Time-dependent changes of tryptophan fluorescence upon excitation at 280 nm were measured at 20 °C for periods of 0.05 s, 5 s and 500 s. The detector voltage was kept constant for low inhibitor concentrations (25-125 µM) but was adjusted to 80% of the maximum signal for each of the higher concentrations.

Crystallography. HG3.17 E47N N300D was crystallized in the presence of the transition state analog 6-cyanobenzotriazole **4** by adding 6-cyanobenzotriazole **4** in DMSO (final concentration 1 mM) to the protein solution, with a final DMSO concentration of 0.1%. Crystals were grown at 20 °C in sitting drops using the vapor diffusion technique (3-Well Crystallization Plates, *Jena Bioscience*). A grid screen was set up with the precipitant solutions containing 0.1 M sodium acetate with a pH gradient from 5 to 6 in rows A-H and a precipitant gradient of ammonium sulfate from 1 M to 2.5 M in columns 1 to 12. Precipitant and protein solutions were mixed in a 1:1 ratio to give drop sizes of 400 nl. To enhance crystal growth, 20 nL of crystal

seeds were added. Seeds were generated by transferring initial crystals obtained under the same conditions into 100 μL ammonium sulfate solution (2 M) and the solution was vortexed for 1 min employing a Seed Bead (*Hampton Research*). All screens were performed at the Protein Crystallization Center at the University of Zurich.

X-ray diffraction data sets were collected at X06SA at the Swiss Light Source (*Paul Scherrer Institute*) using an EIGER X 16M detector. A wavelength of 1.0000 Å was used for data collection. The diffraction data were processed and scaled using the XDS program package [177]. The initial phases were determined by molecular replacement, using the structure of Kemp eliminase HG3.17 (PDB code: 4BS0, [4]) as a search model. Molecular replacement was performed with Phaser [178]. The structure was modified manually with Coot [179] and refined with PHENIX [180]. The final crystal data and intensity statistics are summarized in Table 5.

Table 5: Data collection and refinement statistics for 6-cyanobenzotriazole **4** bound to HG3.17.	
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	51.0, 57.6, 91.5
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	50.0-2.0 (2.05-2.00)*
R_{merge}	12.9 (44.5)
$I / \sigma I$	12.2 (5.6)
CC (1/2) (%)	99.4 (94.9)
Completeness (%)	99.3 (98.3)
Redundancy	6.4 (6.2)
Refinement	
Resolution (Å)	48.7-2.00
No. reflections	18644
R_{work} / R_{free}	14.5 / 19.1
No. atoms	
Protein	2277
Ligand/ion	16
Water	303
B-factors	
Protein	11.2
Ligand/ion	25.6
Water	20.2
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.841

* One crystal was used for data collection. *Values in parentheses are for highest-resolution shell.

2.4.3 Vibrational spectroscopy

FTIR spectroscopy.   FTIR spectra were recorded on a Bruker Vertex 70 spectrometer with a liquid nitrogen-cooled mercury cadmium telluride (MCT) detector. A demountable liquid cell was prepared from two CaF_2 optical windows (0.75 in diameter, 0.25 in thickness, Red Optronics), separated by two semicircular Mylar spacers (50 and 75 µm thickness), to which 20 µl of approximately of 1-10 mM protein-ligand or ligand solution in aqueous buffer (50 mM sodium phosphate, 100 mM NaCl, pH 7.0) was added (for final concentrations, see Table 3). Transmission spectra were acquired by averaging 512 scans after 10 min of purging with a nitrogen flow to remove atmospheric CO_2 and water vapor. Spectra were recorded from 4000 to 1000 cm^{-1} with 1 cm^{-1} resolution and a 6 mm aperture. The absorption spectra were calculated by taking the negative logarithm of the difference between solution (either protein-ligand or ligand) and reference (either protein or buffer) transmission spectra. All measurements were repeated in triplicate, and all frequencies determined using the OPUS software's peak picker (Bruker) as well as Voigt fitting with the Levenberg–Marquardt algorithm. For the FTIR pH titration of 5-carbonitrile benzotriazole, the same sample holder and measurement conditions were used as above. Concentrated DMSO solutions of 5-carbonitrile benzotriazole were diluted into pH-titrated 100 mM sodium phosphate 100 mM sodium acetate 200 mM NaCl aqueous buffers between pH 3.6 – 9.0 to a final concentration of 10 mM and 1% (v/v) DMSO. Fitting of the absorbance at 2227 cm^{-1} to (7) gave a pK_a of 6.51±0.02.

Pre-resonance Raman Spectroscopy.   Pre-resonance Raman spectra were recorded on a Horiba Labram HR Evolution Raman spectrometer using the 532 nm laser, a 50X objective, and a 600 grooves/mm grating. Concentrated solutions of 5-nitrobenzotriazole (approximately 1-10 mM) in complex with each protein were acquired from averaging 5 scans of 120 sec acquisition times (for final concentration, see Table 3). The sample holder was made from glass coverslips (VWR 24x50 mm micro-cover glass) affixed with a single-sided adhesive gasket (Secure-Seal spacer, eight wells, 9 mm diameter, 0.12 mm deep from Molecular Probes) as a sample reservoir. Sample volumes on the order of 2-5 µl were added to the center of the gasket to which Dow Corning vacuum grease was placed along the top-edge (non-adhesive) of the gasket to form a seal and an additional glass coverslip was placed on the grease, preventing evaporation of the sample. Difference spectra were determined by subtraction of solutions containing exclusively protein.

2.4.4   Computational Methods

Reactant and transition state structures were determined with the GAUSSIAN 09 software [181]. Calculations for Chapter 2.1 were performed as follows. Structural optimization of the transition state containing 5-nitrobenzisoxazole **2** and acetate as well as the potential energy surface scan for perturbation of that transition state were performed at the B3LYP/6-311++g(d,p) level. Transition states containing the oxyanion stabilizing residues were determined at the B3LYP/6-31+g(d,p) level. Transition states were located as maxima on the potential energy surface for the cleavage of the scissile C-H bond by performing a relaxed scan of the C-H distance. The analyzed transition states include the catalytic base represented as acetic acid, the substrate and, if stated, water or acetamide as oxyanion stabilizing groups. The given activation energies were determined as the energetic differences between the transition states and free reactants. The potential energy surface for the interaction of the base with the substrate was sampled by systematically varying the hydrogen bond geometry while keeping the rest of the transition fixed at the optimal geometry.

All calculations for Chapter 2.2 were performed at the B3LYP/6-311++g(d,p) level with a polarizable continuum model. A typical input file contained the following keywords: # opt=modredundant   B3LYP/6-311+g(d,p)   scrf=(solvent=water)   geom=connectivity int=ultrafine. Transition states were located as maxima on the potential energy surface for the cleavage of the scissile C-H bond by performing a relaxed scan of the C-H distance. Activation energies were calculated from structures including the catalytic base represented as acetic acid and the substrate 5-cyanobenzisoxazole. The activation energies were determined as the energetic differences between the transition states and relaxed ground states. The energies and vibrations of the inhibitor base complex were calculated from structures including acetic acid as catalytic base and the inhibitor 5-cyanobenzotriazole. The distance between the base O and the N1 of the inhibitor was fixed to 2.6 Å, if not otherwise stated.

3 Engineering novel enzymes by directed evolution of the computationally designed Kemp eliminase 1A53-2 ‡

De novo engineering of tailor-made biocatalysts with enzyme-like rates is a longstanding aim in synthetic biology. In nature, Darwinian evolution orchestrated the many contributions to catalysis, such as positioning of catalytic residues [11,14-16], electrostatic preorganization [17,18], and conformational dynamics [22-24] to achieve efficient catalysis. Enzymes accelerate their reactions by tight recognition of the transition state. For the reactions typically targeted during computational design, pico- to femtomolar affinities would be required to achieve catalytic perfection [182]. Catalytically perfect enzymes display apparent second-order rate constants of at least 10^8 $M^{-1}s^{-1}$ and are limited in their overall rate only by diffusion of reagents and not by the rate of the chemical reaction [8]. Triosephosphate isomerase (TIM) and Δ-ketosteroid isomerase (KSI) are two catalytically perfect catalysts that accelerate proton transfer from carbon by ten to twelve orders of magnitude [136,137]. Mimicking these exquisite catalysts by design is challenging. Nonetheless, a combination of computational design and directed evolution seems promising in that regard and has afforded highly active catalysts for several model reactions (Figure 46, [97]).
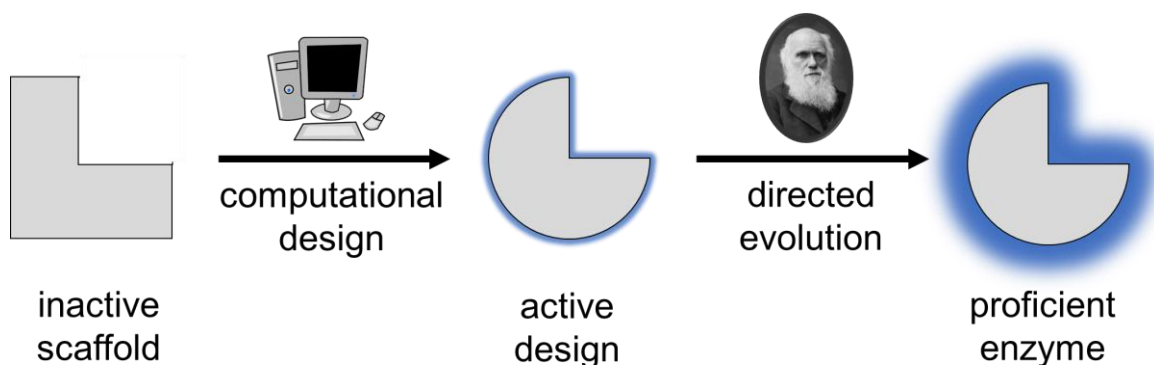
Figure 46: *De novo* enzyme engineering by computational design and directed evolution. Novel active sites have been introduced by computational design into inert protein scaffolds. Successful designs were found to be highly amenable to directed evolution and have afforded enzyme-like catalysts in some cases [4,5].

Computational enzyme design aims to achieve selective transition state stabilization in inert protein scaffolds [97,139]. In order to introduce a novel binding site for a transition state, the transition state with its catalytic residues can be computationally docked into the scaffold and tightly packed by additional mutations *in silico*. Though design has afforded active catalysts for numerous reactions, the transition state affinities typically do not exceed the low nanomolar range. Thus, these catalysts lag behind their natural counterparts by orders of magnitude in activity. Nonetheless, computational designs have proven highly amenable to directed evolution. In the best cases, catalytic efficiency has been improved by several orders of magnitude. Although *de novo* engineered catalysts are still far from catalytic perfection, comparison with all enzymes listed in the BRENDA database [135] reveals that a few engineered enzymes approach the catalytic efficiencies of "average" natural enzymes (k_{cat}/K_M ~10^5 M^{-1}s^{-1}, [4,5]). Among these *de novo* enzymes are a few Kemp eliminases [4,120]. The Kemp elimination is an important model reaction for C-H deprotonation which entails deprotonation of a benzisoxazole substrate, leading to concerted opening of the isoxazole ring (Figure 47).
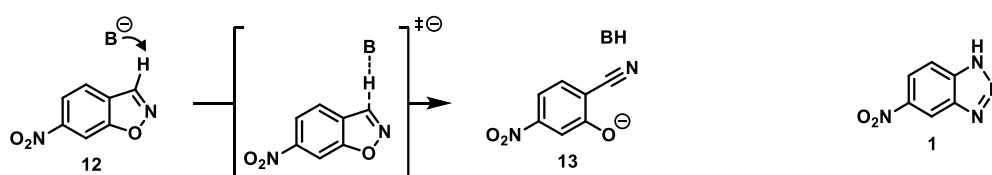
Figure 47: Mechanism of the Kemp elimination. Deprotonation of 6-nitrobenzisoxazole **12** affords salicylonitrile **13** in a one-step reaction. The transition state analog 6-nitrobenzotriazole **1** can be employed as a potent inhibitor of Kemp eliminases and as an important mechanistic probe.

Kemp eliminase HG3.17 stands out from all engineered Kemp eliminases for its exceptional rate acceleration of 6x10^8 fold over the background reaction [4]. HG3.17 was engineered over 17 rounds of directed evolution from the computationally designed protein HG3 [103]. High rate accelerations in HG3.17 were achieved through tight packing of the active site, leading to precise positioning of active site residues (Chapter 2.1), fine-tuned active site electrostatics (Chapter 2.2), and an optimized protein conformational landscape (Chapter 2.3). The efficiency of HG3.17 is unprecedented among Kemp eliminases and only matched by one other *de novo* enzyme which catalyzes retro-aldol cleavage [5]. Given the enzyme-like activity of HG3.17, one might wonder whether its evolution was just a lucky strike.

Here, another computationally designed Kemp eliminase, 1A53-2, is refined by directed evolution to test whether *de novo* engineering can reproducibly yield enzyme-like catalysts. After ten rounds of evolution, the resulting catalyst accelerates deprotonation from carbon

4×10^7-fold over the uncatalyzed reaction. pH-Rate profiles suggest that fine tuning of the pK_a through medium effects facilitated catalysis of the evolved catalysts. Furthermore, X-ray structure analysis indicates precise refinement of the positioning of active site residues. Knock-out studies underscore the importance of positioning and suggest the emergence of a catalytic dyad composed of an H-bonding interaction with the general base. In addition, Arrhenius analysis indicates that evolution significantly reduced the activation enthalpy of the catalyst but increased the entropy of activation. This finding supports the hypothesis that modern enzymes evolved from their primordial ancestors by decreasing the activation enthalpy. This research furthermore demonstrates that the *de novo* engineering successes can be repeated and emphasizes the importance of an appropriate electrostatic microenvironment and catalytic residue positioning.

3.1 Results

Starting Design. Kemp eliminase 1A53-2 [103] was previously designed by the Mayo group using the ORBIT protein design software [98] and the *Sulfolobus solfataricus* indole-3-glycerolphosphate synthase as a scaffold (Figure 48, [183]). The design revisited the same scaffold used to generate Kemp eliminase KE59, which was previously designed with the Rosetta software suite [104]. Despite targeting the same binding pocket, the catalytic base was placed at a different position by the ORBIT design algorithm, leading to a perpendicular orientation of the transition state in the two designs. Also, the actually observed orientation of the transition state analog 6-nitrobenzisoxazole **12** in the crystal structure of 1A53-2 differed from the initial design. 6-Nitrobenzisoxazole **12**, which is tightly sandwiched between two tryptophan residues, both in the design and in the experimental structure, was turned by 180° (Figure 48B). The 6-nitrobenzotriazole **1** ligand can resemble both the 5- and 6-nitro-substituted substrate depending on how it interacts with the catalytic base. As a consequence, the orientation of 6-nitrobenzotriazole **1** relative to the base is more consistent with binding of 6-nitrobenzisoxazole **12**, although 1A53-2 was designed to cleave 5-nitrobenzisoxazole **2** (Figure 48B). As opposed to the previously discussed Kemp eliminase HG3, the X-ray structure of 1A53-2 indicates only a single binding mode (Chapter 2, [103]). Nonetheless, the computationally designed 1A53-2 accepts both 5- and 6-nitrobenzisoxazole as substrates, suggesting that multiple binding modes are likely relevant in solution.
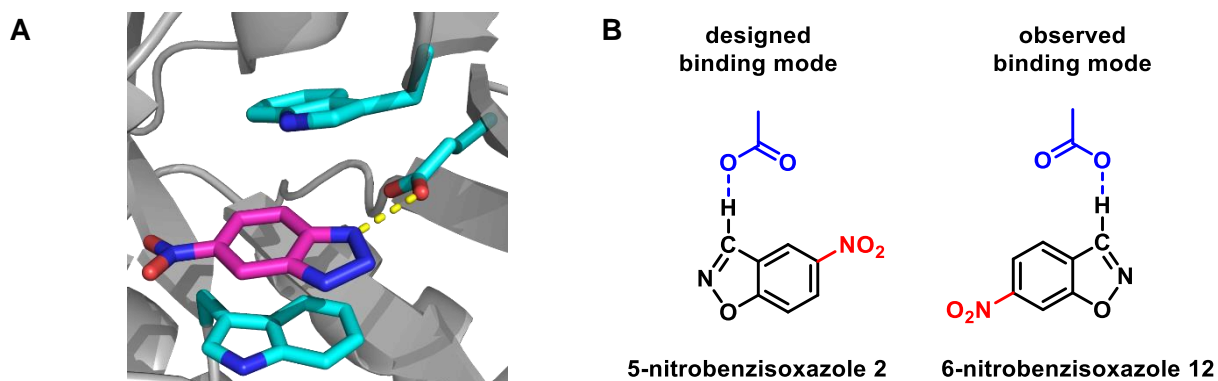
Figure 48: Designed Kemp eliminase 1A53-2. **(A)** The transition state analog 6-nitrobenzotriazole **1** (magenta) is packed between two tryptophan residues and positioned for proton abstraction by the general base (blue) of 1A53-2. **(B)** The design of 1A53-2 aimed to accelerate the Kemp elimination of 5-nitrobenzisoxazole **2**. The observed binding mode of the inhibitor in the X-ray structure, however, resembles 6-nitrobenzisoxazole **12**.

Directed Evolution. Highly proficient enzymes typically have very slow background reactions [1]. Since the acetate-promoted Kemp elimination with 6-nitrobenzisoxazole **12** is five-times slower than with 5-nitrobenzisoxazole **2** [19], it was hypothesized that 1A53-2 might be evolved to higher proficiency using the 6-nitro-substituted substrate. A pH-rate profile of 1A53-2 was determined which revealed an apparent pK_a of the general base of 8.7±0.2. Hence, the first two rounds of evolution were performed at pH 8.0, where a higher activity over the background reaction was observed in cell lysate. The pH was reduced to 7.0 in later rounds of evolution to decrease the hydroxide-catalyzed background reaction, hopefully allowing better rate accelerations after refinement of the catalyst.

Active site optimization. The way a genetic library is constructed dictates which part of the protein fitness landscape is explored during evolution. Thus, various strategies have been proposed to navigate sequence space efficiently with the limited number of variants that can be assessed during evolution [184]. Libraries are typically constructed by saturation mutagenesis, error-prone PCR or DNA shuffling. Iterative saturation mutagenesis has proved to be particularly successful for optimization of a given set of residues [185]. In this approach, active site residues are divided into small groups that are simultaneously targeted by evolution with reduced codon sets. Iterative cycling of mutagenesis through these groups of active site residues may give rapid improvements. Reduced codon sets, which encode only a fraction of all possible amino acids, are crucial for this approach, since they allow simultaneous mutagenesis of several sidechains while maintaining modest but screenable diversity.

In contrast to common iterative saturation mutagenesis, directed evolution of 1A53-2 was performed by completely randomizing each first shell residue individually by saturation

mutagenesis. Hence, each possible single-point mutant was assessed with the NNK codon set instead of relying on smaller reduced codon alphabets. Using NNK codons, all possible amino acids could be sampled at each position. Subsequent DNA shuffling of the resulting hits allowed optimization of the active site with comparably little screening effort (approximately 6,000 variants were screened in total). In this way, evolution likely afforded the best possible set of 18 active site residues that can be obtained by individual mutations (Figure 49).
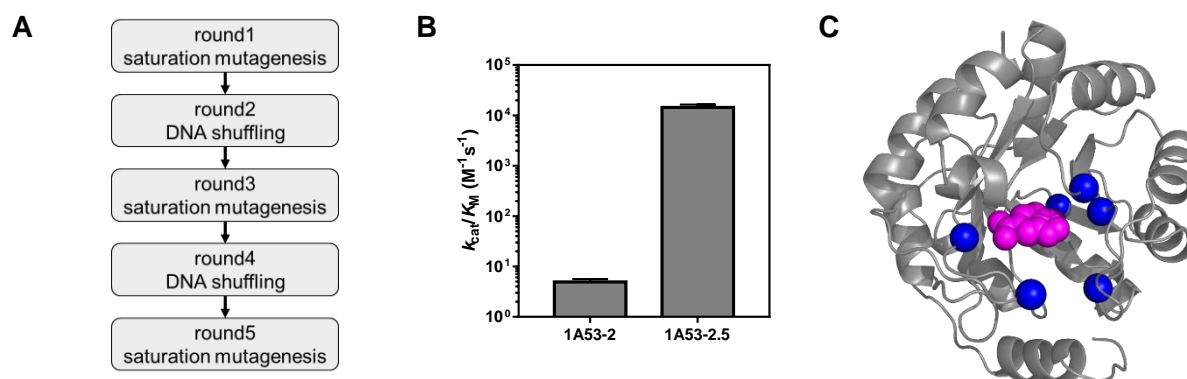
Figure 49: Active site optimization by directed evolution. **(A)** Five rounds of saturation mutagenesis and DNA shuffling were performed to optimize the active site of 1A53-2. **(B)** Evolution yielded a catalyst with a 3,000-fold improved rate acceleration. **(C)** Mutations introduced during the local optimization (blue). The inhibitor is depicted in magenta.

During evolution, mutations at six of the 18 active site residues were combined to give the best catalyst, 1A53-2.5, which has a 3,000-fold improved rate compared to 1A53-2. Two of the identified mutations (A157Y and L184F) introduced significant steric bulk at the active site. Such mutations are usually deleterious in tightly packed environments, but both increased the catalyzed rate of the reaction by approximately 10 fold. The X-ray structure of 1A53-2 suggests that these mutations likely increased activity by filling void pockets around the substrate binding site. Furthermore, evolution introduced two new residues, A157Y and A180C, which are close to the general base and may alter its reactivity and/or positioning.

Global optimization. Active site optimization involved targeting mutagenesis to specific residues. Owing to the absence of additional evident targets for mutagenesis, the protein was globally randomized during the next four rounds of evolution by error-prone PCR and DNA shuffling. Nine more mutations were introduced, increasing the rate of the Kemp eliminase another threefold (Figure 50). While the screening effort during the last four rounds (approximately 10,000 variants in total) exceeded that of the previous rounds, most of the 8,000-fold improvement of the final variant 1A53-2.9 stems from the initial active site optimization.
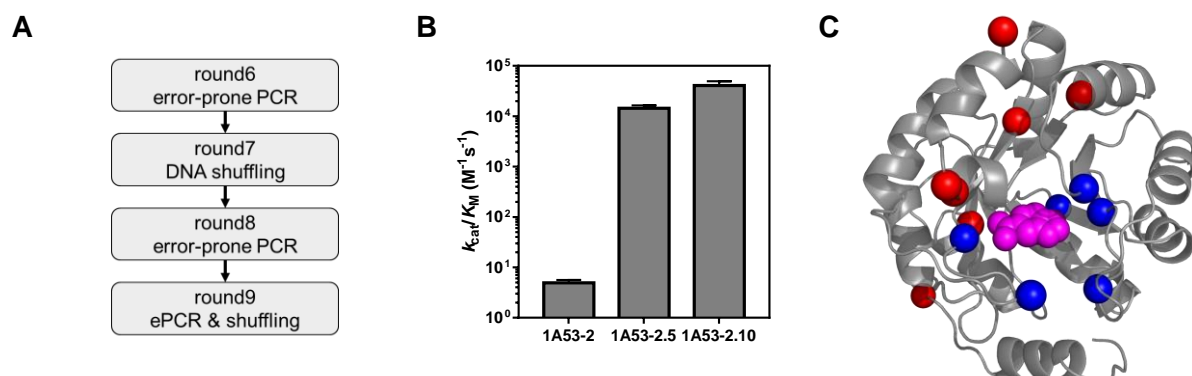
Figure 50: Global optimization by directed evolution. **(A)** Four rounds of error-prone PCR and DNA shuffling were performed to globally optimize 1A53-2.5. **(B)** Global optimization yielded an additional threefold improvement in rate acceleration compared to the round 5 variant. **(C)** Mutations introduced during the local (blue) and global (red) optimization. The inhibitor is depicted in magenta.

Biochemical characterization. The properties of the evolved catalyst were analyzed by in-depth biochemical characterization (Table 6). The round 9 catalyst, 1A53-2.9, has a k_{cat} of 24 ± 4 s^{-1} and a k_{cat}/K_M of $(4.1\pm0.9)\times10^4$ M^{-1}s^{-1} at pH 7.0. Its pK_a is decreased from 8.7 ± 0.2 for the design to 5.5 ± 0.1 for 1A53-2.9. The most evolved catalyst thus reaches its maximal activity at pH of 7.0. In agreement with literature precedent [5,119], evolution involved a tradeoff between the reactivity of the general base and the fraction of active base in the right ionization state by decreasing the pK_a of the catalytic base. Thus, intrinsic reactivity of the catalytic base was decreased to increase the fraction of deprotonated carboxylate in the resting state of the enzyme.
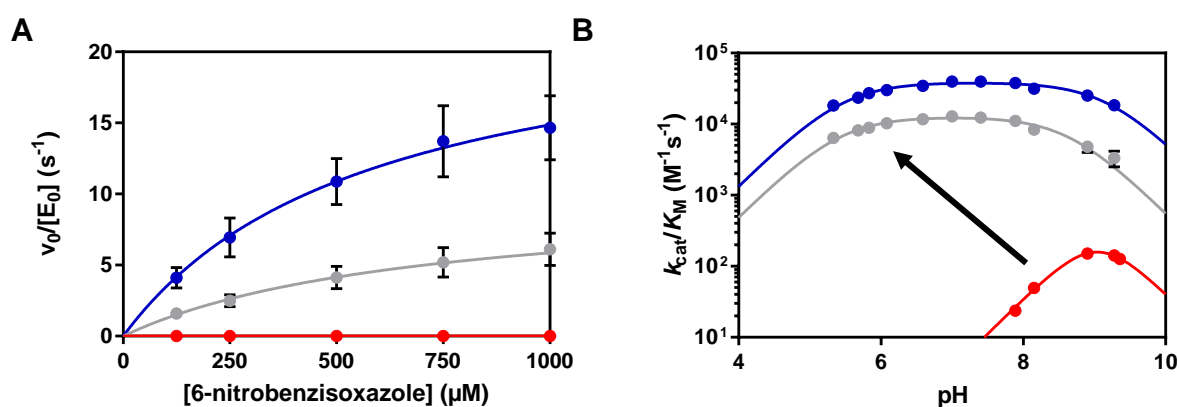
Figure 51: Kinetic characterization of 1A53-2 (red), 1A53-2.5 (gray), and 1A53-2.9 (blue). **(A)** Michaelis-Menten kinetics for the cleavage of 6-nitrobenzisoxazole **12** at pH 7.0 indicate that evolution primarily increased the k_{cat} value (T = 25 °C). **(B)** The pH-rate profile shows fine-tuning of the general base's pK_a. The evolved catalysts reach maximal activity at pH 7.0.

Table 6: Biophysical data of 1A53-2 variants			
	1A53-2	1A53-2.5	1A53-2.9
k_{cat} (s^{-1})	0.0058 ± 0.0004	10 ± 3	24 ± 4
K_M (μM)	1200 ± 200	710 ± 210	590 ± 130
k_{cat}/K_M (M^{-1}s^{-1})	4.9 ± 0.7	14000 ± 2000	40900 ± 8500
k_{cat}/k_{uncat}	(1.10 ± 0.08) × 10^4	(1.9 ± 0.6) × 10^7	(4.2 ± 0.3) × 10^7
$1/K_{TS}$ (M^{-1}) [d]	(9.4 ± 1.2) × 10^6	(2.7 ± 0.4) × 10^{10}	(7.8 ± 1.6) × 10^{10}
K_I (μM)	1.8 ± 0.2	0.46 ± 0.07	0.23 ± 0.02
p$K_{a,1}$ [c]	8.7 ± 0.2	5.4 ± 0.1	5.5 ± 0.1
p$K_{a,2}$ [c]	9.4 ± 0.2	8.6 ± 0.1	9.1 ± 0.1
T_M (°C)	74.8 ± 0.4	82.5 ± 0.9	63.5 ± 0.7
ΔH (kcal/mol)[e]	17.7 ± 0.4	7.1 ± 0.7	6.8 ± 0.7
-TΔS (kcal/mol)[e]	-1.0 ± 0.6	4.9 ± 0.7	4.4 ± 0.7
ΔG (kcal/mol)[e]	16.7 ± 0.2	11.9 ± 0.1	11.2 ± 0.1

[a] Measured at 25 °C in 50 mM phosphate, 100 mM NaCl, pH 7.0 if not otherwise stated. The estimated errors reflect the standard deviations of independent measurements. [b] A k_{uncat} of 5.24 × 10^{-7}s^{-1} was determined for the cleavage of 6-nitrobenzisoxazole. [c] p$K_{a,1}$ and p$K_{a,2}$ refer to the two apparent pK_as of bell-shaped pH-rate profile. [d] The chemical proficiency is defined as: $1/K_{TS} = k_{cat}/K_M/k_{uncat}$. [e] The thermodynamic parameters were calculated at 25 °C. The energetic parameters for the background reaction are ΔH‡ = 20.5 ± 0.4 kcal/mol, -TΔS‡ = 3.3 ± 0.3 kcal/mol, ΔG‡ = 23.8 ± 0.2 kcal/mol.

Total turnovers. 1A53-2.9 accelerates the Kemp elimination 4x10^7 fold over the background reaction. In addition, total turnover experiments with 1A53-2.5 revealed that each active site can perform more than 4,000 turnovers (Figure 52B). High total turnover numbers are highly desirable for biocatalytic applications [186].
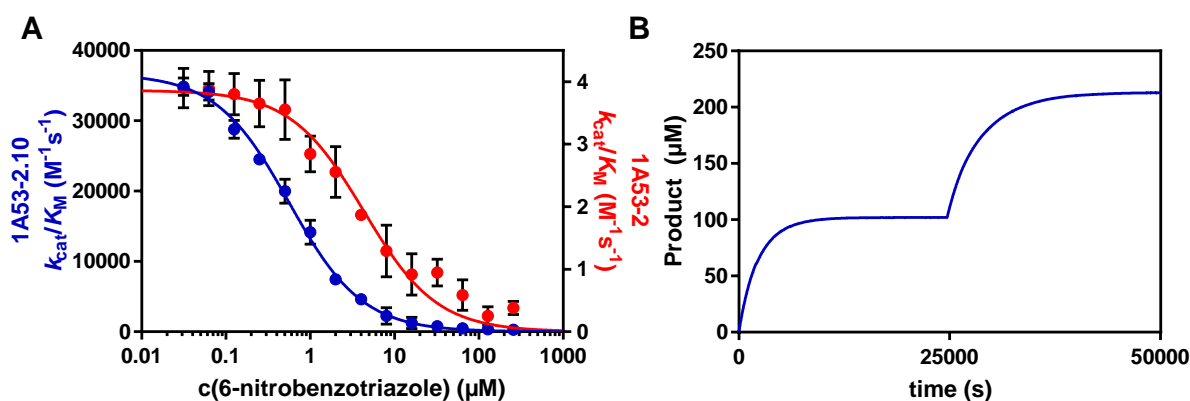
Figure 52: Kinetic characterization of 1A53-2 (red) and 1A53-2.9 (blue). **(A)** K_i measurements with the transition state analog 6-nitrobenzotriazole **1** indicate increased recognition of the transition state by the evolved variants. **(B)** Background-corrected total turnovers for 1A53-2.5 (blue, 50 nM). Two subsequent additions of 6-nitrobenzisoxazole **12** indicate that the catalyst is capable of at least 4,000 turnovers.

Since 1A53-2.9 showed slow precipitation from solution over time, 1A53-2.5 was used for the total turnover assay. The latter is less prone to precipitation and has only a five-fold lower activity. The total turnover number was determined for 1A53-2.5 by sequential addition of two batches of substrate (100 μM each) to protein (50 nM) over 14 h. A small decrease of the initial slope after the second addition of substrate by 30% may either due to precipitation of the enzyme over time or product inhibition. The significant background reaction rate at pH 7.0 ($k_{uncat} = 5.24 \times 10^{-7} s^{-1}$) further limited the assay. Thus, the total turnover number represents only a lower estimate of the catalytic potential of this enzyme.

Transition state affinity. Catalysis is equivalent to selective transition state stabilization [13]. Transition state analogs can hence be employed to examine changes in transition state affinity. K_i measurements with the 1A53-2 variants using the transition state analog 6-nitrobenzotriazole **1** show tighter binding for the evolved catalyst (0.22±0.02 μM) compared to the computational design (1.8±0.2 μM), which reflects its increased transition state recognition (Figure 52A and Table 6). 6-Nitrobenzotriazole **1**, however, is only an imperfect transition state mimic [148]. Thus, the 10-times tighter interaction with the transition state analog does not account for the entire 8,000-fold improvement in catalytic activity. As a consequence, the calculated transition state affinity (equation (12)) of the most evolved catalyst ($K_{TS} = 13$ pM) is orders of magnitude tighter than the measured K_i values.

$$\frac{1}{K_{TS}} = \frac{k_{cat}}{K_M} \cdot \frac{1}{k_{uncat}} = \text{catalytic proficiency} \tag{12}$$

Crystallography. X-ray crystallography was employed to elucidate the structural basis for tighter transition state binding. The structure of Kemp eliminase 1A53-2.5, which displays comparable activity to the round 9 variant but was more stable in solution, was solved at a resolution of 2.6 Å (Figure 53). The structure indicates that evolution afforded a tightly preorganized catalyst with improved packing and positioning. The two tryptophan residues, which form π-stacking interactions with the inhibitor, are slightly rotated allowing a tighter interaction with the ligand. The orientation of the base was furthermore refined and turned more in the plane of the inhibitor. The inhibitor-base distances, however, do not differ significantly within error for the designed (2.59 Å) and evolved variant (2.55 Å). The observed

distance may, however, be already ideal for catalysis since it is close to that of Kemp eliminase HG3.17 (2.53 Å) which achieved an only slightly higher catalytic proficiency [4].
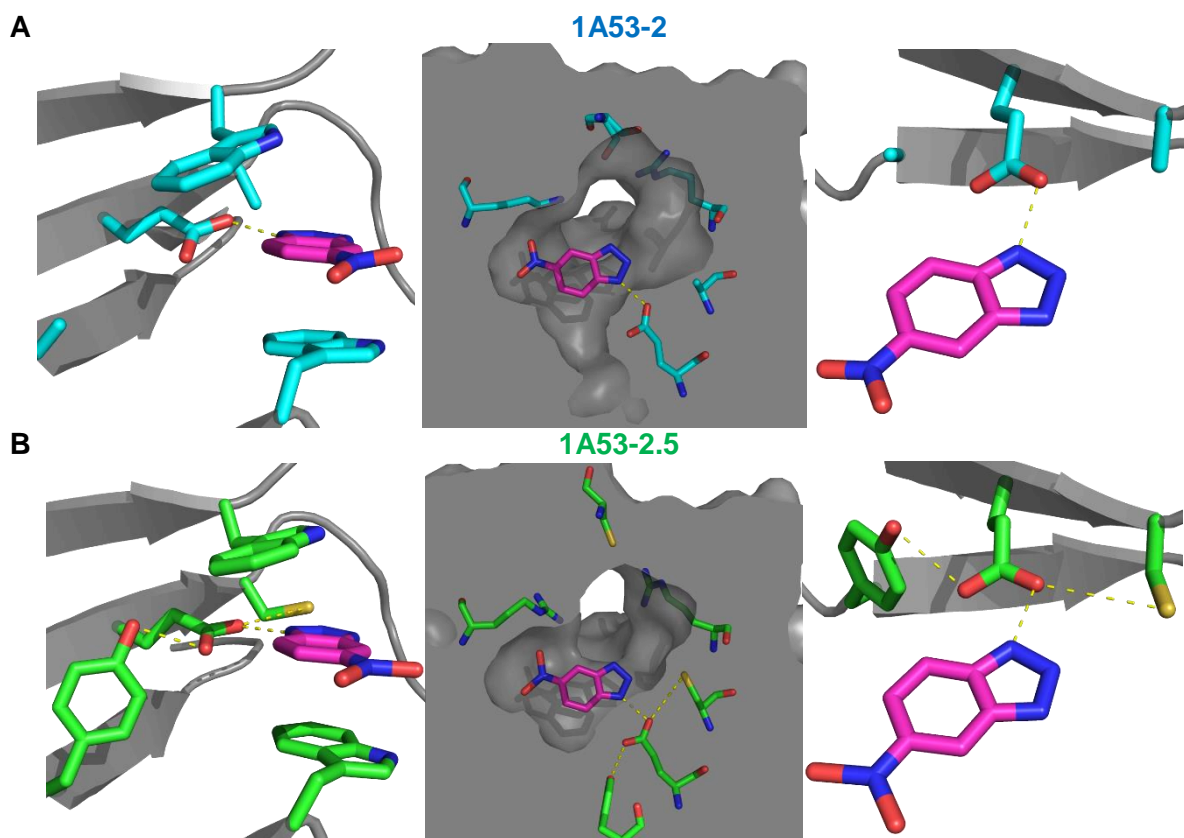
Figure 53: X-ray structure analysis of (A) 1A53-2 (blue) and (B) 1A53-2.5 (green) in complex with the 6-nitrobenzotriazole 1 (magenta) indicate refined positioning of the catalytic base as well as the tryptophan residues (left). Packing of the active site was furthermore improved by introduction of several space-filling mutations (middle). In addition, an H-bond interaction between Tyr157 and Glu178, which appeared during evolution, might contribute to positioning of the base and fine-tuning of its electrostatic environment (right). Furthermore, Cys180 could not be replaced by serine or alanine during evolution and may thus be catalytically relevant.

An evolved catalytic dyad. During evolution, the two residues flanking the general base were mutated to tyrosine and cysteine. As indicated by the X-ray structure of 1A53-2.5, Tyr157 seems to be nicely positioned for an H-bonding interaction with Glu178 (Figure 53). Though Cys180 is badly positioned for H-bonding with Glu178, it may potentially affect the general base through hydrophobic interactions. Directed evolution supports this hypothesis, since Cys180 could not be replaced by serine under the screening conditions.

In order to scrutinize the impact of Tyr157 and Cys180 on catalysis, theoretical microscopic titration curves (THEMATICS, [187]) were determined for all residues in the scaffold. THEMATICS is based on the hypothesis that acid-base catalysis requires significant stabilization of both the protonated and deprotonated forms of the catalyst. A general base, such as Glu178 in 1A53-2, needs to be able to deprotonate its substrate, while the conjugated

acid also has to release that proton after the reaction is completed. In agreement with this hypothesis, it was observed that the active site residues of natural enzymes often have very shallow titration curves [188]. THEMATICS thus ranks a residue's propensity to be involved in catalysis by analysis of the curvature of titration curves for the individual protein side chains. Computational analysis revealed that the A157Y and A180C mutations increased the ranks of the corresponding positions and are thus likely to be catalytically relevant (Figure 54).
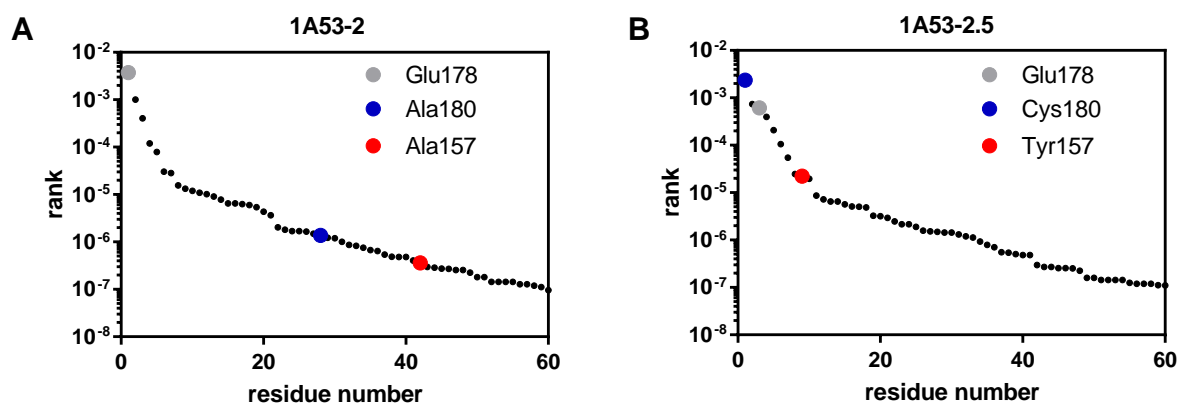
Figure 54: THEMATICS analysis of Kemp eliminase variants. The highest scoring 60 residues in **(A)** 1A53-2 and **(B)** 1A53-2.5 are listed. High scores indicate a high likelihood of catalytic relevance. The catalytic base (gray) has high scores both in the design and in the evolved variant. Residues 180 (blue) and 157 (red) increase in score indicating catalytic contribution. Particularly A180C is likely to have a high impact on catalysis according to the THEMATICS analysis.

In order to scrutinize the role of Tyr157 and Cys180 experimentally, site-directed mutagenesis was performed and the resulting variants were analyzed by Michaelis-Menten kinetics (Table 7). Only the second order rate constants are employed for a comparison of the variants, which are less uncertain than k_{cat} due to the elevated K_M (≤1.1 mM) of some of the variants and the limited solubility of 6-nitrobenzisoxazole **12** (≤1.5 mM). Knock-out of Tyr157 had the expected effect on activity, as catalytic efficiency was reduced by approximately 10 fold upon mutation to Ala or Phe. The similar effect seen for both substitutions indicates that the loss in activity is primarily due to the loss of the H-bonding interaction and not due to a decrease in packing. The effect still lags behind that in KSI, where mutagenesis, as well as chemical rescue experiments, indicate that distorted positioning of the general base leads to a 10^3-fold decreased activity [36]. KSI, however, exhibits a 300-fold higher rate acceleration than 1A53-2.9 and may thus be more sensitive to distortion of the general base.

Table 7: Michaelis-Menten parameters of the active-site knock-outs of 1A53-2.9			
variant	k_{cat} (s^{-1})	K_M (M^{-1})	k_{cat}/K_M (M^{-1}s^{-1})
1A53-2.9	24±4	590±20	(4.1±0.9)x10^4
1A53-2.9 E178Q [a]	n.d. [a]	n.d. [a]	n.d. [a]
1A53-2.9 Y157A	1.1±0.1	190±40	(6.0±1.1)x10^3
1A53-2.9 Y157F	1.3±0.4	340±110	(4.1±1.0)x10^3
1A53-2.9 C180A	39±8	1100±100	(3.6±0.7)x10^4
1A53-2.9 C180S	6.0±0.8	650±70	(9.2±1.2)x10^3

[a] Precipitation of 1A53-2.9 E178Q occurred too rapidly to analyze the protein.

Knock-out of the catalytic base (E178Q) or mutagenesis of the cysteine (C180S, C180A) primarily impaired the stability of the protein. Upon mutagenesis, the purified proteins showed accelerated precipitation from solution compared to 1A53-2.9. For E178Q, precipitation occurred too rapidly to purify the variant, and thus no activity could be determined. In contrast, the C180S and C180A variants could be characterized and C180A exhibited activity similar to that of 1A53-2.9. The decreased stability of the mutant was likely the reason why cysteine at position 180 was not replaced by alanine or another residue. Interestingly, the fivefold decrease in activity seen for the C180S variant indicates that increased H-bonding by serine is not beneficial for catalysis. The 1A53-2.9 C180S Y157F double mutant further indicated no synergy between the two mutations, highlighting the importance of a precisely positioned H-bonding interaction with Tyr157.

Arrhenius analysis. The effects of the temperature on the rate of chemical reactions can provide valuable insights regarding the enthalpic and entropic contributions to the reaction barrier. According to the linearized form of the Eyring equation (13), the slope of the temperature dependency corresponds to the activation enthalpy ($\Delta H^\ddagger$) while the activation entropy ($\Delta S^\ddagger$) can be deduced from the intercept [164]. Importantly, a decrease in $\Delta H^\ddagger$ will lead to a shallower curve and thus makes the reaction less temperature sensitive.

$$\ln\left(\frac{k_{cat}}{K_M} \cdot \frac{h}{k_B T}\right) = \frac{-\Delta H^\ddagger}{R} \cdot \frac{1}{T} + \frac{-\Delta S^\ddagger}{R} \tag{13}$$

The temperature dependence of Kemp eliminase HG3.17, discussed in Chapter 2.3, indicated that the enzyme catalyzes its reaction by a decrease of $\Delta H^\ddagger$. However, the analysis was not particularly reliably due to the limited temperature range over which HG3.17 is stable. In contrast, the temperature-dependence of k_{cat}/K_M for 1A53-2.9 can be measured over a 40 °C

range without loss of activity (Figure 55). The shallow temperature-dependence of the reaction catalyzed by the most evolved enzyme compared to the reaction catalyzed by the computational design indicates that improvements during evolution were primarily associated with a more favorable activation enthalpy ($\Delta\Delta H^{\ddagger}$ = -10.9 kcal/mol). $\Delta H^{\ddagger}$ became steadily more favorable at the expense of entropy which became increasingly less favorable during evolution ($-T\Delta\Delta S^{\ddagger}$ = +5.4 kcal/mol).

Similar behavior was noted by Wolfenden for many natural enzymes. For instance, he observed a reduction in $\Delta H^{\ddagger}$ for the enzyme-catalyzed reactions for a series of twelve randomly chosen enzymes from the literature [165]. $\Delta H^{\ddagger}$ for the catalyzed reactions ranged from 9 to 15 kcal/mol, while their uncatalyzed counterparts were in the range 22 to 47 kcal/mol. One notable exception was the ribosome, which is a true entropy trap and accelerates peptide bond formation by decreasing only $\Delta S^{\ddagger}$ [31]. Moreover, reduction of $\Delta H^{\ddagger}$ was also observed for primitive small molecule catalysts, such as pyridoxal phosphate or metal ions, which were likely utilized by primordial enzymes for catalysis. Wolfenden thus hypothesized that modern enzymes evolved from their primordial ancestors primarily by reducing $\Delta H^{\ddagger}$. That trend is well reproduced in the evolution of the designed Kemp eliminase 1A53-2.

Wolfenden furthermore notes that the cooling atmosphere of the early earth added a selective advantage for primordial enzymes that decreased $\Delta H^{\ddagger}$. Decreasing $\Delta H^{\ddagger}$ renders a catalyzed reaction less sensitive toward temperature compared to the corresponding background reaction. Enzymatic reduction of the activation enthalpy thus results in an increased rate acceleration in colder environments, which likely supported evolution of primordial enzymes. Evolution of 1A53-2, however, was performed in the absence of any temperature-associated selection pressure on $\Delta H^{\ddagger}$. The directed evolution experiment would have to be repeated at different temperatures to experimentally validate this selective advantage of $\Delta H^{\ddagger}$ as compared to $\Delta S^{\ddagger}$. Nonetheless, the laboratory evolution of 1A53-2 supports the idea that modern enzymes evolved from primitive catalysts primarily by decreasing $\Delta H^{\ddagger}$, regardless of potential benefits associated with catalyzing reactions as the early earth was cooling down.
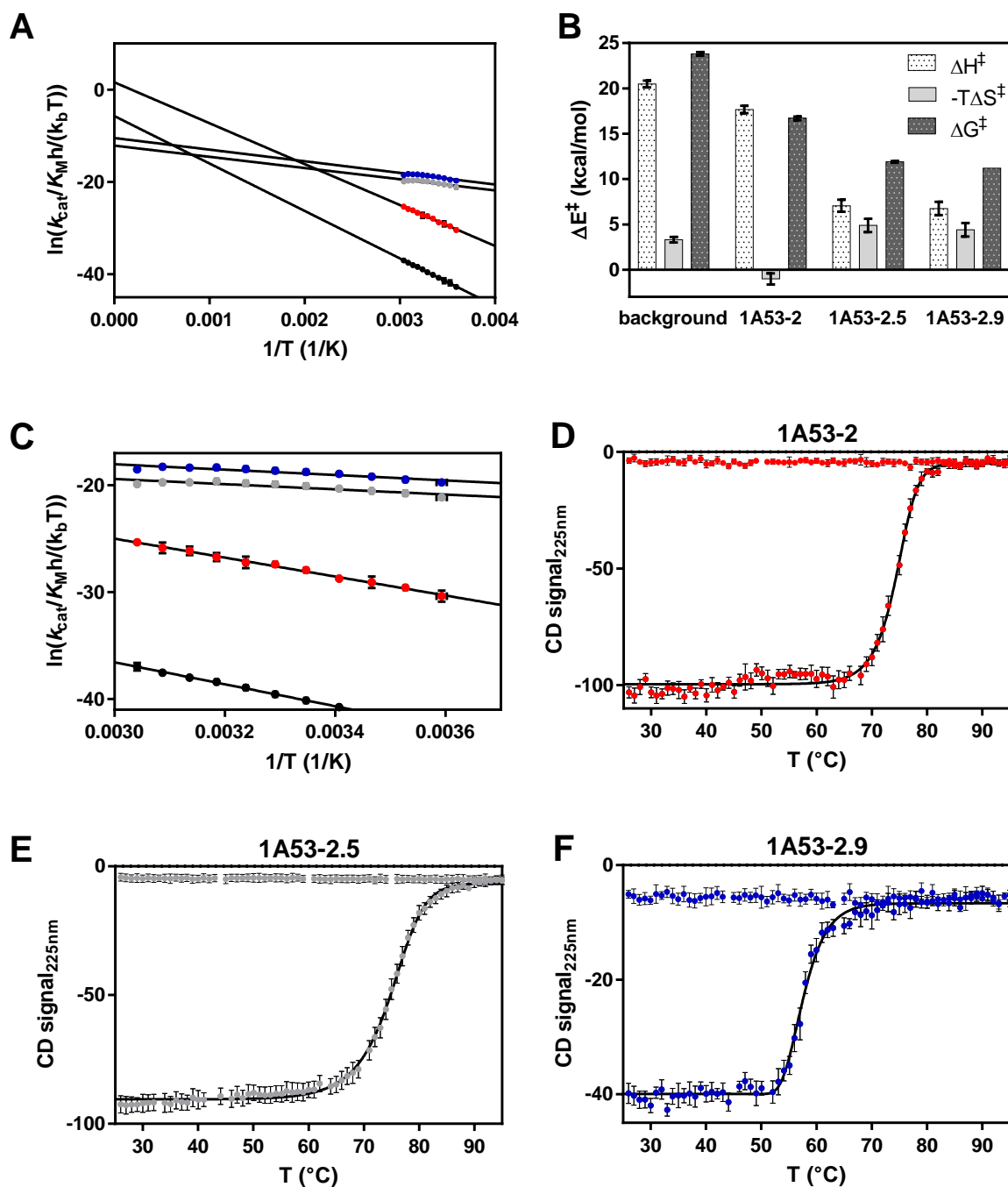
Figure 55: Temperature dependence of Kemp eliminase variants. **(A)** Arrhenius plot of 1A53-2 (red), 1A53-2.5 (gray), 1A53-2.9 (blue) and the background reaction (black). The analysis of the enzyme variants is based on their catalytic efficiency, while the reaction rate in buffer was employed for the background. **(B)** Based on the Eyring equation (13), the activation enthalpies ($\Delta H^{\ddagger}$) can be deduced from the slope of the temperature dependence, while the intercept correlates to the activation entropies ($T\Delta S^{\ddagger}$, for T = 20 °C). Evolution improved catalysis by decreases of $\Delta H^{\ddagger}$ to the expense of a slight increase in $T\Delta S^{\ddagger}$. Error bars represent the errors of the fit. **(C)** Magnified version of plot A. Circular dichroism melting curves of **(D)** 1A53-2 **(E)** 1A53-2.5 and **(F)** 1A53-2.9 indicate that all proteins are stable in the range of the temperature-dependent kinetic analysis.

3.2 Discussion

De novo engineering of perfect, diffusion-limited enzymes requires design of transition state binding pockets with pico- to femtomolar affinities for reactions such as the Kemp elimination [182]. The Kemp elimination is a valuable model reaction to tackle this design challenge. The reaction is easy to monitor experimentally and involves only a one-step C-H deprotonation event that initiates concerted isoxazole ring opening [62]. Thus, the requirements for design are much more clearly defined compared to multistep reactions [100-102,129,130].

The one-step Kemp elimination has an activation barrier of 23.1 kcal/mol for the hydroxide-promoted reaction with 6-nitrobenzisoxazole **12** and represents a significant challenge for design. The reaction is highly exothermic and particularly difficult to catalyze due to the small charge rearrangement in the relatively early transition state [62,151]. Nonetheless, Kemp eliminase 1A53-2.9, which was improved 8,000 fold by directed evolution in this study, is a highly efficient artificial enzyme and displays a 4×10^7-fold rate acceleration over the background reaction. Comparison of 1A53-2.9 to the activities of natural enzymes deposited in the BRENDA database indicates that 1A53-2.9 has acquired catalytic parameters of a typical natural enzyme [135]. In addition, 1A53-2.9 exceeds the rate-acceleration of the previously engineered Kemp eliminase KE59.13 [120] and is only 10-fold below that of HG3.17 [103]. Thus, evolution of HG3.17 was not just a lucky fluke, but computational design and directed evolution allowed reproducible generation of catalysts with enzyme-like rate accelerations (Figure 56). Nonetheless, 1A53-2.9, as well as other designed Kemp eliminases, is still several orders of magnitude less efficient than Nature's perfect catalysts, which accelerate C-H deprotonations with efficiencies that are only limited by diffusion.

Although computational design focuses on the first shell residues, most improvements in 1A53-2.9 were achieved by refinement of the active site residues. In contrast, evolution of other computationally designed enzymes required intense mutagenesis of the respective protein scaffold to achieve comparable improvements [4,5,120]. 1A53-2.5 is a very interesting variant, since most of the introduced mutations were likely assessed and a suboptimal sidechain was chosen by the computational design algorithm. Thus, the 3,000-fold improvement in the first five rounds of evolution can be directly related to design shortcomings. The lack of precision during design may have prevented the best residue being chosen for a particular site. Surprisingly, some side chains were also introduced that drastically reduced packing of the

active site. Particularly, the designed Ala153, which was replaced by tyrosine during evolution, leading to a ten-fold gain in activity, was reduced in size from leucine in the parent scaffold. This observation indicates that the evaluation of packing interactions will need to be improved in future design algorithms.
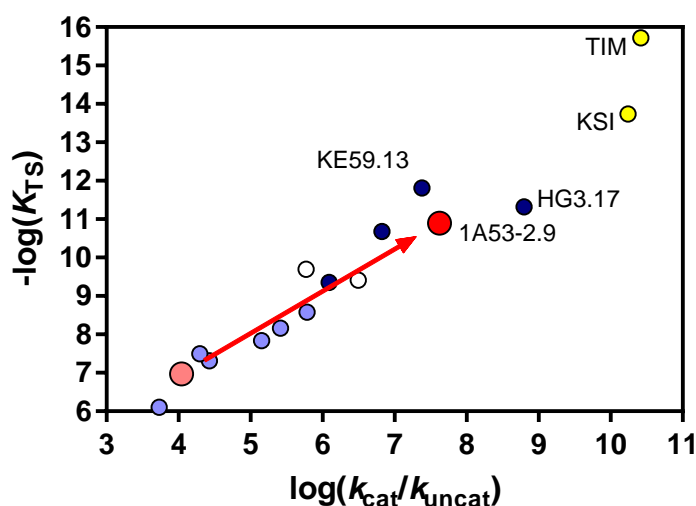
Figure 56: Transition state affinities and rate accelerations of natural and designed enzymes. The designed (light red) and evolved (red) Kemp eliminase 1A53-2 afforded a catalyst with similar catalytic proficiency as KE59.13 [120] and HG3.17 [103]. Artificial catalysts like catalytic antibodies (white), computationally designed enzymes (light blue) and evolved variants (dark blue) enzymes are shown for comparison. Nonetheless, natural enzymes like KSI and TIM (yellow) are still orders of magnitude more efficient than *de novo* engineered catalysts. The kinetic constants and respective enzymes are listed in Supplementary Table 10.

In the case of 1A53-2, active site optimization was performed with comparably little effort. Only 6,000 variants had to be screened owing to the employed evolutionary approach. The strategy was similar to iterative saturation mutagenesis, which is employed to sample the active site rapidly with a limited set of codons [185]. The approach pursued here allows rapid identification of the best combination of all possible single point mutations, whereas iterative saturation mutagenesis is limited by the range of sampled mutations. Interestingly, global optimization by error-prone PCR and DNA shuffling required intensive screening of approximately 10,000 variants and only gave a five-fold improvement.

For comparison, roughly 30,000 variants were screened in the evolution of HG3.17, resulting in a catalyst with similar catalytic proficiency and a tenfold higher rate acceleration [4]. In both cases, however, improvements in the last rounds of evolution became harder to accomplish, as the number of beneficial mutations decreased. More sophisticated screening methods may help to cover larger genetic diversities for the identification of such rare mutations ([189,190], Chapter 4). Ultrahigh-throughput methods are particularly attractive in this context, as they

allow more efficient exploration of fitness landscapes than is possible with conventional methods [5,138].

Importantly, the active site of 1A53-2.9 is still configured as the starting design. Significant rearrangements of the active site, frequently observed for other designs [4,5,155], were not seen during evolution of 1A53-2.9. Thus, the designed and evolved catalysts are directly comparable in terms of their structures. Structural analysis clearly shows refined positioning and packing during evolution. An additional driving force for evolution was increased stabilization of the general base as indicated by pH-rate profiles. The pK_a of E178 in the computational design was highly perturbed (8.7±0.2) and its decrease during evolution (5.5±0.1) provided a significant improvement in rate acceleration at neutral pH. Thus, tighter control over the pK_a of the catalytic residues is required to optimize their reactivity. Current design algorithms, however, do not explicitly analyze solvation of the catalytic residues.

Furthermore, formation of a catalytic dyad (Glu-Tyr) seems to have been highly beneficial for catalysis. Other Kemp eliminases have also successfully employed similar catalytic dyads including an His-Asp dyad in KE70 [121]. An extended H-bonding network, including a water molecule and two H-bond donating sidechains, is similarly used to position the general base of catalytic antibody 34E4 [191]. Positioning of complex catalytic devices, however, has not been reliably achieved by design. For instance, structural analysis showed that the catalytic dyads of designed esterases were misplaced in the active site [99]. Thus, future designs will likely require increased control over these networks to implement complex catalytic devices reliably during design.

Finally, the evolved catalyst 1A53-2.9 clearly accelerates the Kemp elimination by reduction of the activation enthalpy, just as most natural enzymes [165]. It was earlier proposed by Wolfenden that evolution of primordial enzymes was facilitated by decreasing $\Delta H^{\ddagger}$. With the cooling of the early earth, rate accelerations achieved by enzymes that predominantly decreased the activation enthalpy would have increased exponentially according to the Eyring equation (13). The evolution of 1A53-2 supports this proposal by demonstrating the feasibility of evolving a highly active enzyme starting from a naïve catalyst solely by reduction of $\Delta H^{\ddagger}$.

In conclusion, several of the required features for enzyme design are clearly illustrated by the evolution of 1A53-2.9. Complex catalytic moieties that are precisely positioned through H-bonding interactions seem to be required. In addition, this study provides another example

that computational design needs improved strategies for packing the binding pocket. Furthermore, design should be expanded beyond the theozyme approach and aim at gaining control over the reactivity of catalytic residues. Meanwhile, more efficient mutagenesis strategies and powerful screening methodologies are likely to accelerate evolution and yield efficient enzymes in the future. Despite these challenges, Kemp eliminase 1A53-2.9 became a truly enzyme-like catalyst after evolution, accelerating its reaction by 4×10^7-fold over the background. Furthermore, the vast reduction of its activation enthalpy during evolution supports the proposal by Wolfenden that early enzymes evolved by reducing $\Delta H^\ddagger$ as the earth was cooling down.

3.3 Materials and methods

In vitro evolution

Starting Point for *in vitro* evolution.   Kemp eliminase 1A53-2 [103] was previously designed into the thermostable indole-3-glycerolphosphate synthase scaffold from *Sulfolobus solfataricus* (PDB: 1A53, [183]). 12 mutations were introduced into the inert scaffold using the method of Lassila *et al.* [98]. The resulting enzyme served as the starting point for multiple rounds of directed evolution via iterative rounds of saturation mutagenesis, error-prone PCR and DNA shuffling, combined with microtiter plate screening.

Library construction.   Libraries were constructed by overlap-extension PCR with Phusion polymerase (*Finnzymes*) [192]. Fragments were constructed with the primer in Table 8 and the AKZ3 (TAATACGACTCACTATAGGG) and T7term (GCTAGTTATTGCTCAGCGG) primer. For a typical PCR protocol, samples were incubated 5 min at 98 °C and DNA was amplified in 32 cycles between 98 °C (30 s), 55 °C (30 s) and 72 °C (30 s). After a final elongation for 5 min at 72 °C, the samples were cooled to 4 °C. The fragments were purified by agarose gel electrophoresis. Gels usually contained 2% agarose in 0.5 g/l sodium borate ($Na_2B_4O_7$x$10H_2O$) for small fragments. Gels were stained with ez-vision one (*Amresco*). Gel electrophoresis was performed at 200 V for 20 min at a variable electrical current. DNA extractions from agarose gels were carried out using the Zymoclean Gel DNA Recovery Kit (*Zymo Research*). Fragments were assembled by PCR using a similar protocol as before. 15 cycles were performed without the T7 and T7term primer to assemble the genes and 32 cycles were performed thereafter with the outer primer for amplification. The PCR products were purified via agarose gel electrophoresis (1% agarose). The assembled genes were amplified by PCR, purified with the DNA Clean & Concentrator kit (*Zymo research*) and cloned into pET11 for screening. Plasmids were purified with the Zyppy Plasmid Miniprep Kit (*Zymo research*) from *E. coli* Xl1-Blue cells. For cloning, 2 μg of vector and 1 μg of insert were digested with NdeI and BamHI (*New England Biolabs*). Vector and insert were purified via agarose gel electrophoresis (1% agarose). Fragments were ligated using T4 DNA ligase (*New England Biolabs*) and purified with the DNA Clean & Concentrator kit. Electro-competent Xl1-Blue cells were transformed with the library [193], transferred to 10 ml LB and incubated for 1 h at 37 °C and 230 rpm. After 1 h, 0.1% of the culture was transferred onto an LB-agar plate (100 μg/ml ampicillin) for estimation of the library size. 100 μg/ml ampicillin were added to the culture and the culture was incubated overnight at 37 °C and 230 rpm. The library was ready for screening after purification with the

Zyppy Plasmid Miniprep Kit (*Zymo research*). All DNA concentrations were determined on a NanoDrop spectrophotometer (*ThermoFischer*). Gene sequences were confirmed by sequencing (*Microsynth*)

Table 8: Cloning primers for the directed evolution of 1A53-2	
Active site optimization	
Name	**Sequence 5'-3'**
1A532_K53_NNK	GCTATCATTGCCGCGTATNNKCGTAAATCCCCGTCTGGTCTGG
1A532_K53_rev	ATACGCGGCAATGATAGC
1A532_S58_NNK	CGTATAAACGTAAATCCCCGNNKGGTCTGGATGTAGAACGTGATCC
1A532_L60_NNK	CGTATAAACGTAAATCCCCGTCTGGTNNKGATGTAGAACGTGATCCGATCG
1A532_S58/L60_rev	CGGGGATTTACGTTTATACG
1A532_A147_NNK	GTTCTTACGGCATGGAACCGNNKATTGTAATCAACGACGAAAATGACC
1A532_V149_NNK	GTTCTTACGGCATGGAACCGGCGATTNNKATCAACGACGAAAATGACC
1A532_A147/V149_rev	CGGTTCCATGCCGTAAGAAC
1A532 A180_NNK	GCTCGTTTCATCGAAATTNNKAGCCGCGATCTGGAAACTCTGG
1A532_R182_NNK	GCTCGTTTCATCGAAATTGCGAGCNNKGATCTGGAAACTCTGGAAATC
1A532_A180/R182_rev	AATTTCGATGAAACGAGC
1A532_ Q211_NNK	GTTGTCAAAGTTGCGTGGNNKGGCATCTCTGAACGTAACG
1A532_ Q211_rev	CCACGCAACTTTGACAAC
L60_FP_wt	CGTATAAACGTAAATCCCCGTCTGGTCTGGATGTAGAACGTGATCCGATCG
L60_FP_T	CGTATAAACGTAAATCCCCGTCTGGTACGGATGTAGAACGTGATCCGATCG
A157_FP_wt	GTTCTTACGGCATGGAACCGGCGATTGTAATCAACGACGAAAATGACC
A157_FP_FY	GTTCTTACGGCATGGAACCGTWTATTGTAATCAACGACGAAAATGACC
A157_FP_W	GTTCTTACGGCATGGAACCGTGGATTGTAATCAACGACGAAAATGACC
A157_FP_M	GTTCTTACGGCATGGAACCGATGATTGTAATCAACGACGAAAATGACC
A180/R182_FP1	GCTCGTTTCATCGAAATTGCGAGCRKGGATCTGGAAACTCTGGAAATC
A180/R182_FP2	GCTCGTTTCATCGAAATTWSTAGCRKGGATCTGGAAACTCTGGAAATC
Q211_FP_wt	CGTTGTCAAAGTTGCGTGGCAGGGCATCTCTGAACGTAACG
Q211_FP_S	CGTTGTCAAAGTTGCGTGGTCGGGCATCTCTGAACGTAACG
Q211_FP_G	CGTTGTCAAAGTTGCGTGGGGGGGGCATCTCTGAACGTAACG
1A532_A51_NNK	GCAACATCACCGCTATCATTGCCNNKTATAAACGTAAATCCCCGTCTGG
1A532_A51_rev	GGCAATGATAGCGGTGATGTTGC
1A532_K53_NNK	GCTATCATTGCCGCGTATNNKCGTAAATCCCCGTCTGGTCTGG
1A532_S56_NNK	GCTATCATTGCCGCGTATAAACGTAAANNKCCGTCTGGTCTGGATGTAGAACG
1A532_K53_rev	ATACGCGGCAATGATAGC
1A532_S58_NNK	CGTATAAACGTAAATCCCCGNNKGGTCTGGATGTAGAACGTGATCC
1A532_S58/L60_rev	CGGGGATTTACGTTTATACG
1A532_A81_NNK	GGAACGTTACGCTGTAGGCCTGNNKATTGCGACCGAAGAAAAGTACTTTAACG
1A532_A83_NNK	GGAACGTTACGCTGTAGGCCTGGCGATTNNKACCGAAGAAAAGTACTTTAACG

Table 8, continued: Cloning primers	
1A532_A81_rev	CAGGCCTACAGCGTAACGTTCC
1A532_F89_NNK	CGATTGCGACCGAAGAAAAGTACNNKAACGGTAGCTACGAAACGCTGC
1A532_F89_rev	GTACTTTTCTTCGGTCGCAATCG
1A532_L108_NNK	GCTCTGTTAGCATTCCGATTNNKATGTGGGACTTTATCGTGAAAGAGTCC
1A532_W110_NNK	GCTCTGTTAGCATTCCGATTCTGATGNNKGACTTTATCGTGAAAGAGTCC
1A532_L108_rev	AATCGGAATGCTAACAGAGC
1A532_A131_NNK	CCTGGGTGCAGATACCGTCNNKCTGATTGTTAAAATCCTGACC
1A532_A131_rev	GACGGTATCTGCACCCAGG
1A532_L184_NNK	GCTCGTTTCATCGAAATTTGTAGCCGCGATNNKGAAACTCTGGAAATCAACAAAG
1A532_A180/R182_rev	AATTTCGATGAAACGAGC
1A532_W210_NNK	CCAACGTTGTCAAAGTTGCGNNKTCGGGCATCTCTGAACGTAACG
1A532_W210_rev	CGCAACTTTGACAACGTTGG
Global optimization	
1A532_53/58_1	GCTATCATTGCCGCGTATADGCGTAAATCCCCGKSGGGTCTGGATGTAGAACGTGATCC
1A532_53/58_2	GCTATCATTGCCGCGTATSTTCGTAAATCCCCGKSGGGTCTGGATGTAGAACGTGATCC
1A532_53/58_3	GCTATCATTGCCGCGTATCATCGTAAATCCCCGKSGGGTCTGGATGTAGAACGTGATCC
1A532_110_F	GCTCTGTTAGCATTCCGATTCTGATGTTTGACTTTATCGTGAAAGAGTCC
1A532_110_WG	GCTCTGTTAGCATTCCGATTCTGATGKGGGACTTTATCGTGAAAGAGTCC
1A532_184_L	GCTCGTTTCATCGAAATTTGTAGCCGCGATTTAGAAACTCTGGAAATCAACAAAG
1A532_184_F	GCTCGTTTCATCGAAATTTGTAGCCGCGATTTTGAAACTCTGGAAATCAACAAAG
1A532_210/211_1	CCAACGTTGTCAAAGTTGCGKSGTCTGGCATCTCTGAACGTAACG
1A532_210/211_1	CCAACGTTGTCAAAGTTGCGKSGTGCGGCATCTCTGAACGTAACG
1A532-lib7-F1	CTCTGCGTCGTCCGAGCTTCCRTGCCAGCCGTCAGCGCCCGATC
1A532-lib7-R1	GAAGCTCGGACGACGCAGAG
1A532-lib7-F2	GCATTCTGGAATTCAACAAGCRCAACATCACCGCTATCAYTGCCGCGTATAGGCGTAAATCC
1A532-lib7-R2	CTTGTTGAATTCCAGAATGC
1A532-lib7-F3	GGAAATCAACAAAGAAAACCAGMGCAAGCTGATCWCTATGATCCCGTCCAACGTTG
1A532-lib7-R3	CTGGTTTTCTTTGTTGATTTCC
1A532-lib7-F4	TTGAAGAGCTGCGTAAACYGGGTGTGAACGCCTTCGGC
1A532-lib7-R4	TTTACGCAGCTCTTCAA
1A532-lib7-F5	AAAATCAAGGAGTTTATCSTGGGCAGCATCGAGGRTCGTGGTCACCATCACCACC
1A532-lib7-R5	CCTTGATTTTCTCCGGGTTGCKCATCAGGGAGGAGCYGATGCCGAAGGCGTTCACACC
1A532-lib9-F1	CAAGCGCAACATCACCGCTAYTATTGCCGCATATAGGCG
1A532-lib9-R1	CAAGCGCAACATCACCGCTA
1A532-lib9-R1	TAGCGGTGATGTTGCGCTTG
1A532-lib9-F2	CGTCGCGCTGATTGTTAAARTCCTGACCGAACRTGAACTGGAAAGCCTGCTGG
1A532-lib9-R2	CGTCGCGCTGATTGTTAAA
1A532-lib9-R2	TTTAACAATCAGCGCGACG
1A532-lib9-F3	GGGGCGGCATCTCTGAACGTAMCGAAMTTGAAGAGCTGCGTAAACYGGGTGTGAACGCCTTCGGC
1A532-lib9-R3	GGGGCGGCATCTCTGAACGTA
1A532-lib9-R3	TACGTTCAGAGATGCCGCCCC
1A532-lib9-F4	GGAGAAAATCAAGGAGTTTATCSTGGGCAGCATCGAGGRTCGTGGTCAC
1A532-lib9-R4	GACAGCATCGAGGRTCGTGGTCACCRTCACCACCACCATTAGGGGATCCGGC
1A532-lib9-R4	GAGCCGGATCCCCTAATGGTGGTGGTGAYGGTGACCACGAYCCTCGATGCTG

Table 8, continued: Cloning primers	
Active-site knock-out	
1A10-178Q	CATCGGTGCTCGTTTCATCCAAATTTGTAGCCGCGATTTTG
1A10-178Q-rev	GATGAAACGAGCACCGATG
1A10-157A	GTTCTTACGGCATGGAACCGGCTATTGTAATCAACGACGAAAATG
1A10-157A-rev	CGGTTCCATGCCGTAAGAAC
1A10-157F	GTTCTTACGGCATGGAACCGTTTATTGTAATCAACGACGAAAATG
1A10-157F-rev	CGGTTCCATGCCGTAAGAAC
1A10-180A	GCTCGTTTCATCGAAATTGCTAGCCGCGATTTTGAAACTC
1A10-180A-rev	AATTTCGATGAAACGAGC
1A10-180S	GCTCGTTTCATCGAAATTTCTAGCCGCGATTTTGAAACTC
1A10-180S-rev	AATTTCGATGAAACGAGC

Active site optimization. During the first five rounds of directed evolution, the 20 first-shell residues were optimized individually using saturation mutagenesis with NNK-codons and subsequent shuffling. In round 1, residues in close proximity to the catalytic base E178 (residues 53, 58, 157, 159, 180, 182 and 211) were targeted individually by saturation mutagenesis and the identified hits were shuffled in round 2. In round 3, the remaining first-shell residues (residues 51, 56, 60, 81, 83, 89, 108, 110, 131, 184 and 210) were targeted individually by saturation mutagenesis and the identified hits were shuffled in round 4. In round 5, all 20 active site residues were again individually revisited with saturation mutagenesis. Cloning was performed as described above.

Global optimization. After the first-shell optimization was completed, the overall scaffold was targeted in alternating rounds of error-prone PCR and DNA shuffling (rounds 6-8). In rounds 9 and 10, shuffling and error-prone PCR were performed simultaneously in one round. The template for cloning of the error-prone PCR libraries was prepared by PCR with phusion from a pet11 plasmid harboring the hits. Randomization was performed using Mutazyme II (*Aglient*). 1 ng of the template was subjected to 32 cycles of amplification. After purification using the DNA Clean & Concentrator kit, the insert was amplified via PCR and cloning was continued as described above.

Library screening. In order to guarantee an approximately 3-fold oversampling, 100 clones were screened per saturation mutagenesis library. Alternatively, 800 to 1,000 clones were screened per error-prone PCR or DNA shuffling library. In order to ensure adequate coverage, the library sizes were restricted to either one mutation per gene on average or fewer than 1,000 different possible combinations. BL21-Gold(DE3) cells were transformed with the gene libraries and plated on LB agar plates (with 150 µg/mL ampicillin). Single colonies were used

to inoculate 600 µL LB medium (with 150 µg/mL ampicillin) in a 96-well deepwell plate. After overnight incubation at 30 °C under shaking, 200 µl culture were stored in replica plates. The remainder of the culture was induced by addition of LB (with 1 mM IPTG, 150 µg/mL ampicillin) to a final volume of 1.6 ml. The temperature was reduced to 18 °C and protein production continued for 18 h. Catalytic efficiency was assayed after appropriate dilution (1:10 to 1:1,000) into assay buffer (50 mM phosphate, 100 mM NaCl, 10% methanol, pH 7.0) containing 500 µM 6-nitrobenzisoxazole **12** at 400 nm in a plate reader (*Molecular Devices*). The most active clones were confirmed by rescreening in triplicates. The clones with the largest increase in activity relative to the preceding round were picked from the replica plate for plasmid isolation, sequencing, and further diversification. For rounds 1 and 2, cells were lysed prior to the activity assay due to the limited activity of the initial design. The cells were harvested by centrifugation, the supernatant was discarded, and the cell pellets were frozen three times at -80 °C for two hours and thawed. After thawing, the cells were resuspended in 300 µL lysis buffer (50 mM sodium phosphate, 100 mM NaCl, 1 mg/mL lysozyme, DNAseI, pH 8.0). The cells were incubated for 1 h at room temperature before clearing the lysates by centrifugation. The activity in the lysate was assayed as described above at pH 8.0.

Biophysical Characterization

Protein production and purification. BL21-Gold(DE3) cells were transformed with the pET11 vector harboring the gene of interest and plated on LB agar plates (with 100 µg/ml ampicillin). A single colony was used to inoculate an overnight culture of LB medium containing 100 µg/ml ampicillin. 500 ml of LB medium containing 100 µg/ml ampicillin, were inoculated with the overnight culture in a ratio of 1:100 from and grown to an OD of 0.5 at 37 °C. Protein production was induced with 1 mM IPTG and cells were incubated at 18 °C overnight. The cells were harvested and the cell pellet was either frozen at -80 °C or directly processed for protein purification. The cell pellet was resuspended in lysis buffer (20 ml 50 mM Tris-HCl, 0.5 M NaCl, 10 mM imidazole, 100 µL protease inhibitor cocktail (*Sigma*), 1 mg/ml lysozyme, DNAseI, pH 7.4). The cell suspension was sonicated for 1 h in a sonication bath at 4 °C (*Telesonic Ultrasonics*). After centrifugation (18.000 rpm, 20 min, 4 °C). The soluble fraction was applied to 3 ml Ni-NTA slurry (*Qiagen*), washed with 10 mM imidazole before elution of the protein with 300 mM imidazole, each in 50 mM Tris-HCl, 0.5 M NaCl, pH 8. The buffer was exchanged to 20 mM sodium phosphate, 20 mM NaCl, pH 6.0 by washing three times in an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*). Proteins were further purified by anion exchange chromatography (MonoS column, *GE Healthcare*) and eluted with

a salt gradient (20 mM to 1 M NaCl, 20 mM sodium phosphate, pH 6.0). Fractions absorbing at 280 nm were analyzed by SDS-PAGE (Phastsystem, *Pharmacia Biotech*) using a low molecular weight marker (*GE Helthcare*). Fractions containing the target protein were pooled and concentrated using an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*) if necessary. Protein concentrations were determined by measuring the absorbance at 280 nm and a calculated extinction coefficient [175]. The following extinction coefficients were determined for some representative enzymes: ε(1A53-2) = 29,910 M^{-1}cm^{-1}; ε(1A53-2.5) = 31,400 M^{-1}cm^{-1}; ε(1A53-2.9) = 31,400 M^{-1}cm^{-1}. Protein masses were confirmed by mass spectrometry. For mass determination, protein samples were exchanged into 0.1% acetic acid with Vivaspin 500 centrifugal filters (*Sartorius*) and measured by ESI-MS on a Daltonics maXis ESI-Q-TOF mass spectrometer (*Bruker*). Mass spectra were deconvoluted using the MaxEnt1 software. All masses corresponded to the expected sequence lacking the N-terminal methionine.

Substrate Synthesis.   5- and 6-Nitrobenzisoxazole were synthesized from the respective salicylaldehyde via oxime formation and PPh$_3$/DDQ promoted cyclization as previously described [4,62,171,172].

UV/Vis spectroscopic assay.   Reactions were initiated by adding enzyme (10 nM to 10 µM) to the benzisoxazole substrate (100 µM to 1.5 mM final concentration) in 50 mM sodium phosphate, 100 mM NaCl, 10% methanol (pH 7.0) at 25 °C. The pH of the buffer was measured after addition of methanol using a SenTix 81 pH electrode (*Gerber Instruments*). Product formation was monitored at the appropriate wavelength ($\varepsilon_{400\,nm,6\text{-nitrobenzisoxazle}}$ = 2,870 M^{-1}cm^{-1}; $\varepsilon_{380\,nm,5\text{-nitrobenzisoxazle}}$ = 15,800 M^{-1}cm^{-1} [19]) in a Lambda 35 UV/Vis spectrometer (*PerkinElmer*) at 25 °C. The slope before addition of protein was subtracted as background. Initial rates divided by the catalyst concentration were plotted against substrate concentration, and k_{cat} and K_M values were determined by fitting the data to the Michaelis-Menten equation:

$$\frac{v_0}{E_0} = \frac{k_{cat}\cdot[S]}{K_M + [S]} \tag{2.1}$$

The uncertainty of the catalytic parameters k_{cat} and K_M was estimated by calculating the standard deviation of multiple independent measurements performed with different batches of protein.

pH–Rate profile.   The pH dependence of k_{cat}/K_M was determined by measuring initial rates at 100 µM 6-nitrobenzisoxazole **12** in different buffers. Acetate buffer was used from pH 4 to

5.5 and Bis-Tris propane buffer from pH 6 to 9.5, both containing 100 mM NaCl and 10% methanol at 25 °C. The enzymes were preincubated in the assay buffer for at least 5 min prior to initiation of the reaction by substrate addition. Product formation was monitored as described previously. The rate of the uncatalyzed reaction was measured with an identical sample without enzyme and subtracted. The apparent extinction coefficient for 6-nitro-2-hydroxybenzonitrile at 400 nm was corrected for the ionization state of the product using the formula with $\Delta\varepsilon_{max} = 2870$ M^{-1}cm^{-1} and pK_a = 5.2.

$$\Delta\varepsilon = \frac{\Delta\varepsilon_{max}}{1 + 10^{pK_a - pH}} \tag{14}$$

The sigmoidal curves from the plot of the catalytic efficiency against the pH were fitted to the following equation for pK_a determination:

$$\left(\frac{k_{cat}}{K_M}\right) = \frac{\left(\frac{k_{cat}}{K_M}\right)_{max}}{1 + 10^{pK_a - pH}} \tag{15}$$

Total turnover determination.   50 nM enzyme was added to 990 µl buffer containing 100 µM 6-nitrobenzisoxazole **12** in 50 mM sodium phosphate, 100 mM NaCl, pH 7.0, 8% methanol at 25 °C. After 2,500 s, 10 µl 10 mM 6-nitrobenzisoxazole **12** in methanol were added. Product formation was monitored as described previously. For background correction, a similar sample without enzyme was assayed. A $k_{background}$ value was calculated from the initial velocity by dividing by the substrate concentration. For the enzyme-catalyzed reaction, the amount of substrate at each time point was calculated. The amount of product formed by the background reaction was calculated for each time point by multiplication of $k_{background}$ with the actual substrate concentration and the time-difference to the next time point. The amount of product formed in the background reaction (which added up to 2-4 µM for each substrate addition) was subtracted from the total amount of product, to determine the background corrected curve plotted in Figure 51. The total turnover number was derived by dividing the amplitude by the enzyme concentration.

K_i determination.   For K_i determination enzyme (10 nM to 10 µM) were incubated for at least 5 min in buffer containing up to 1 mM 6-nitrobenzotriazole **1** in 50 mM sodium phosphate, 100 mM NaCl, 1% DMSO (pH 7.0) at 25 °C. Reaction were initiated by addition of 6-nitrobenzisoxazole **12** in methanol to a final concentration of 100 µM and 10% methanol.

Product formation was monitored as described previously. The K_I values were determined by fitting to the following equation.

$$\left(\frac{k_{\text{cat}}}{K_{\text{M}}}\right) = \frac{\left(\frac{k_{\text{cat}}}{K_{\text{M}}}\right)_{\text{max}}}{1 + 10^{\log(K_{\text{I}} - \text{c(inhibitor)})}} \tag{16}$$

For equation (16) to be applicable, it was assumed that the inhibitor acts as a competitive inhibitor and that the substrate concentration is far below the K_M, resulting in the K_I being equal to the IC$_{50}$ according to the Cheng-Prusoff equation [176]:

$$K_{\text{I}} = \frac{\text{IC}_{50}}{1 + \dfrac{\text{c(substrate)}}{K_{\text{M}}}} \tag{17}$$

Arrhenius Analysis. The temperature dependence of k_{cat}/K_M was determined by measuring initial rates at 100 µM 6-nitrobenzisoxazole **12** in 50 mM sodium phosphate 100 mM NaCl, 10% methanol. The enzymes were preincubated in the assay buffer for at least 5 min at a temperature range between 5 °C and 45 °C prior to initiation of the reaction by substrate addition. Product formation was monitored as described previously. Data were fitted to the Eyring-equation [164], which can be derived from transition-state theory, to determine $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$.

Crystallography. Proteins were crystallized in the presence of the transition state analog 6-nitrobenzotriazole **1**, by adding 6-nitrobenzotriazole **1** in DMSO (final concentration 1 mM) to 20 mg/ml protein in 20 mM HEPES, 20 mM sodium sulfate, pH 6.0, to a final DMSO concentration of 0.1% DMSO. Crystals were grown at 20 °C in sitting drops using the vapor diffusion technique (3-Well Crystallization Plates, *Jena Bioscience*). Crystallization was performed by sitting drops. After an initial screen, the best condition (300 nl enzyme, 100 nl mother liquor (0.1 M sodium citrate, 40% v/v PEG 300, pH 4.2)) was refined by varying the pH and PEG concentration. The best crystals from this experiment (400 nl enzyme, 100 nl mother liquor (0.1 M sodium citrate, 43% v/v PEG 300, pH 5.6)) were grinded and used as seed stock. The final crystals were obtained by repeating the refinement screen with addition of 50 nl seed stock.

X-ray diffraction data sets were collected at X06SA at the Swiss Light Source (*Paul Scherrer Institute*) using an EIGER X 16M detector. A wavelength of 1.0000 Å was used for data

collection. The diffraction data were processed and scaled using the XDS program package [177]. The initial phases were determined by molecular replacement, using the structure of Kemp eliminase 1A53-2 (PDB code: 3NZ1, [103]) as a search model. Molecular replacement was performed with Phaser [178]. The structure was modified manually with Coot [179] and refined with PHENIX [180]. The final crystal data and intensity statistics are summarized in **Table 9**.

Table 9: Data collection and refinement statistics for 1A53-2.5.	
Resolution range	19.97 - 2.601 (2.693 - 2.601)
Space group	P 31 2 1
Unit cell	61.304 61.304 121.278 90 90 120
Total reflections	81317 (7640)
Unique reflections	8536 (827)
Multiplicity	9.5 (9.2)
Completeness (%)	98.90 (97.95)
Mean I/sigma(I)	3.69 (0.64)
Wilson B-factor	52.00
R-merge	0.5874 (4.301)
R-meas	0.6208 (4.553)
R-pim	0.198 (1.473)
CC1/2	0.991 (0.302)
CC*	0.998 (0.681)
Reflections used in refinement	8478 (812)
Reflections used for R-free	420 (41)
R-work	0.2475 (0.4480)
R-free	0.3208 (0.4007)
CC(work)	0.983 (0.481)
CC(free)	0.850 (0.510)
Number of non-hydrogen atoms	2054
macromolecules	2005
ligands	32
solvent	17
Protein residues	247
RMS(bonds)	0.012
RMS(angles)	1.64
Ramachandran favored (%)	90.61
Ramachandran allowed (%)	8.57
Ramachandran outliers (%)	0.82
Rotamer outliers (%)	6.85
Clashscore	9.77
Average B-factor	63.09
macromolecules	62.96
ligands	81.80
solvent	42.62

*Values in parentheses are for highest-resolution shell.

3.4 Supplementary material

3.4.1 Sequences of 1A53-2 variants

Amino acid sequences of representative 1A53-2 variants. The catalytic base is highlighted in yellow. Mutations introduced during active site optimization are colored blue. Mutations introduced during global optimization are colored red. The original naming of the variants, named on the well in which they were discovered are: 1A53-2.5-5/F7 and 1A53-2.9-5/F4.

1A53-2   catalytic base = E178

```
MPRYLKGWLK DVVQLSLRRP SFRASRQRPI ISLNERILEF NKRNITAIIA AYKRKSPSGL
DVERDPIEYS KFMERYAVGL AIATEEKYFN GSYETLRKIA SSVSIPILMW DFIVKESQID
DAYNLGADTV ALIVKILTER ELESLLEYAR SYGMEPAIVI NDENDLDIAL RIGARFIEIA
SRDLETLEIN KENQRKLISM IPSNVVKVAW QGISERNEIE ELRKLGVNAF GIGSSLMRNP
EKIKEFILGS IEGRGHHHHH H
```

1A53-2.5   1A53-2 K53R S58C A157Y A180C L184F Q211S

```
MPRYLKGWLK DVVQLSLRRP SFRASRQRPI ISLNERILEF NKRNITAIIA AYRRKSPCGL
DVERDPIEYS KFMERYAVGL AIATEEKYFN GSYETLRKIA SSVSIPILMW DFIVKESQID
DAYNLGADTV ALIVKILTER ELESLLEYAR SYGMEPYIVI NDENDLDIAL RIGARFIEIC
SRDFETLEIN KENQRKLISM IPSNVVKVAW GGISERNEIE ELRKLGVNAF GIGSSLMRNP
EKIKEFILGS IEGRGHHHHH H
```

1A53-2.9   1A53-2.5 R23H I48T I82V I219L M237L R238S L248V G253D H257R

```
MPRYLKGWLK DVVQLSLRRP SFHASRQRPI ISLNERILEF NKRNITATIA AYRRKSPCGL
DVERDPIEYS KFMERYAVGL AVATEEKYFN GSYETLRKIA SSVSIPILMW DFIVKESQID
DAYNLGADTV ALIVKILTER ELESLLEYAR SYGMEPYIVI NDENDLDIAL RIGARFIEIC
SRDFETLEIN KENQRKLISM IPSNVVKVAW GGISERNELE ELRKLGVNAF GIGSSLLSNP
EKIKEFIVGS IEDRGHRHHH H
```

3.4.2   Natural and designed enzymes catalyzing C-H proton abstractions

Kinetic parameters of selected artificial and natural enzymes catalyzing C-H deprotonation reactions are listed in Supplementary Table 10. The second-order rate constant of the hydroxide catalyzed reaction (k_{OH^-}) was used to calculate the background rate (k_{uncat}) for TIM, KSI and 1A53-2 according to (18).

$$k_{uncat} = k_{OH^-} \cdot c(OH^-)$$

(18)

For TIM, kinetic parameters are stated for the reaction starting with glyceraldehyde-3-phosphate. The background reaction rates are k_{uncat}(5-nitrobenzisoxazole **2**) = 1.12×10^{-6} s^{-1} [120], k_{uncat}(5,7-dicyanobenzisoxazole) = 8.82×10^{-7} s^{-1} [120], k_{OH^-}(6-nitrobenzisoxazole **12**) = 5.3 M^{-1}s^{-1} [62], k_{OH^-}(triosephosphate isomerization) = 0.24 M^{-1}s^{-1} [194], k_{OH^-}(steroid isomerization) = 49.2 M^{-1}s^{-1} [152].

Supplementary Table 10: Kinetic parameters of natural and designed enzymes catalyzing C-H proton abstractions.

		catalyst	k_{cat} [s^{-1}]	k_{cat}/K_M [M^{-1}s^{-1}]	k_{cat}/k_{uncat}	K_{TS} [M^{-1}]	Lit.
Kemp elimination catalysts	catalytic antibodies	4B2	3.5	2.9×10^3	3.1×10^6	2.6×10^9	[195]
		34E4	0.66	5.5×10^3	5.9×10^5	4.9×10^9	[196]
	designed enzymes	KE07	0.030	2.3×10^1	2.7×10^4	2.1×10^7	[104]
		KE15	0.022	3.5×10^1	2.0×10^4	3.1×10^7	[104]
		KE16	0.006	1.4×10^0	5.4×10^3	1.3×10^6	[104]
		KE59	0.29	1.6×10^2	2.6×10^5	1.4×10^8	[104]
		KE70	0.16	7.6×10^1	1.4×10^5	6.8×10^7	[104]
		HG3	0.68	4.3×10^2	6.1×10^5	3.8×10^8	[103]
		1A53-2	0.0058	4.9×10^1	1.1×10^4	9.4×10^6	this thesis
	evolved enzymes	KE07	1.4	2.5×10^3	1.2×10^6	2.3×10^9	[119]
		KE70	7.5	5.3×10^4	6.7×10^6	4.8×10^{10}	[121]
		KE59.13	21.2	5.7×10^5	2.4×10^7	6.5×10^{11}	[120]
		HG3.17	700	2.3×10^5	6.3×10^8	2.1×10^{11}	[4]
		1A53-2.9	24	4.1×10^4	4.2×10^7	7.8×10^{10}	this thesis
natural enzymes		TIM	2000	4.0×10^8	2.6×10^{10}	5.3×10^{15}	[137]
		KSI	86000	2.7×10^8	1.7×10^{10}	5.5×10^{13}	[136]

4 Towards novel ultra-high throughput screening systems for the directed evolution of Kemp eliminases ‡

Enzymes have found application as efficient and highly selective catalysts in manifold fields of chemistry and biology. The development of successful biocatalytic processes depends on the identification of a suitable catalyst for the reaction of interest and, if necessary, subsequent engineering of its properties [197]. Genomic databases represent a valuable source of enzymes for chemical transformations important in cellular metabolism [198]. Alternatively, viable candidates may be discovered by searching strain collections or metagenomic libraries [199]. If a natural catalyst does not exist, for example for an abiological reaction, other strategies are necessary. One possibility is to re-design an existing enzyme, capitalizing on its inherently promiscuous catalytic abilities [159]. Another is to create an artificial enzyme *de novo*, for example by computational design as described in Chapter 1 [97].

Once an appropriate starting point has been identified, it may require refinement to optimize its catalytic properties or to meet specific process needs. Directed evolution, which entails iterative rounds of genetic diversification and phenotypic screening or selection, has emerged as a particularly powerful tool for shaping protein properties in the laboratory. Specific activity, stability, substrate scope, and stereoselectivity of enzymes can be optimized using this technique [200]. Nonetheless, evolution in the laboratory is typically time-consuming and laborious and often does not match well with the pace of business [201].

In order to speed up screening campaigns, new technologies with substantially enhanced throughput have been developed ([189,190,202], Figure 57). Miniaturization of microwell-plates has been achieved, for instance, in plates consisting of up to 10^5 capillaries per standard-sized plate [203]. Further miniaturization is, for instance, achieved by microcapillary

single-cell analysis and laser extraction (µSCALE) using microchips consisting of up to 10^6 microcapillaries [204,205].
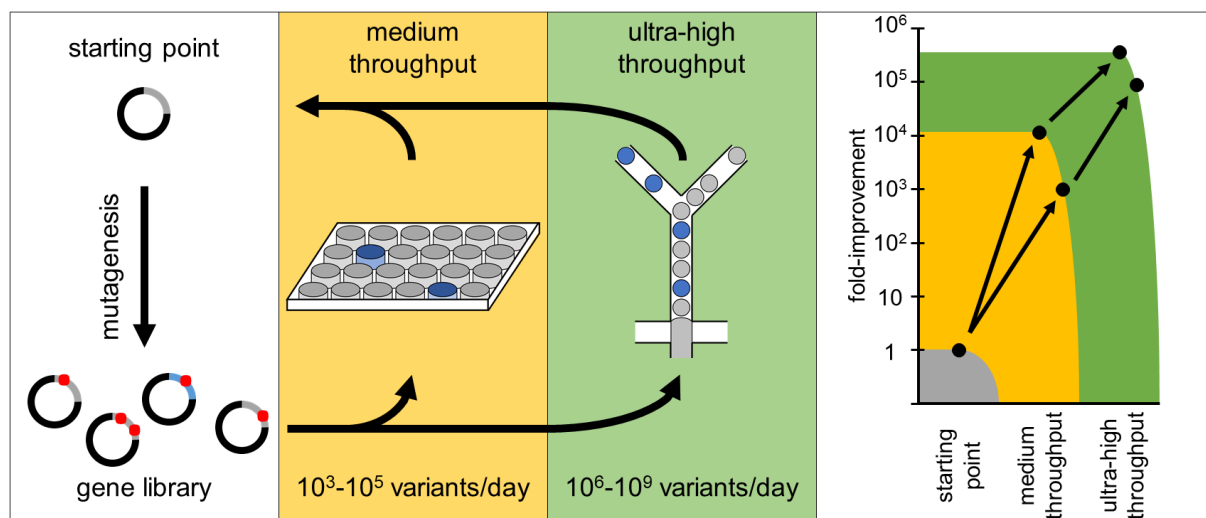
Figure 57: Medium and ultrahigh-throughput screening assays. Increased throughput, for instance by droplet-based microfluidics, enables more efficient exploration of fitness landscapes during directed evolution than is possible with conventional microtiter plate assays [5,138].

Although wells and capillaries are the conceptually simplest way to link genotype and phenotype, the assay throughput typically does not exceed ~10^6 variants. Higher numbers can be analyzed if screening is moved from a solid support to solution. For instance, intact cells can serve as sortable reaction compartments that are readily analyzed by fluorescence-activated cell sorting (FACS) [206,207]. Though efficient, FACS requires non-toxic fluorogenic substrates that either readily diffuse into the cell and become trapped upon reaction, or are tethered to the cell surface to link genotype and phenotype. Alternatively, genotype and phenotype can be linked by co-encapsulated of cells and reagents in droplet emulsions. Monodisperse droplets can be produced and sorted in large numbers (>10^9) by continuous flow on chip using droplet-based microfluidics [5,198,199,208-219].

The evolution of computationally designed enzymes by conventional microtiter plate-based assays is a cost- and time-consuming process. For instance, seventeen rounds of evolution were necessary to create Kemp eliminase HG3.17, which accelerates a proton transfer reaction by 6×10^8 fold [4]. Optimization of computationally designed Diels-Alderases [123], phosphoesterases [220] and retro-aldolases [122,221] proved similarly difficult. The activity of the computationally design retro-aldolase RA95 was increased 4,000-fold using a medium-throughput microtiter plate assay, but further gains proved elusive as beneficial mutations became increasingly rare after over thirteen rounds of evolution [122]. Subsequent droplet-based microfluidics screening of large libraries with up to 10^8 variants led to the identification

of sixteen additional substitutions that boosted activity 30 fold to give a $>10^9$-fold rate enhancement over the uncatalyzed reaction [5].

Ultrahigh-throughput screening systems are uniquely suited to aid the engineering of *de novo* enzymes by making rapid and efficient searches of sequence space possible. However, they typically require a fluorescent readout. Unfortunately, the Kemp elimination is monitored by formation of a colored product, not by fluorescence. Here, novel fluorescent-based assays for Kemp elimination that are potentially applicable to ultra-high throughput screening platforms like µSCALE and droplet-based microfluidics are presented (Figure 58). Fluorescent readouts are generated based on pH changes, recognition via supramolecular interactions or with a fluorogenic substrate (Figure 58). Efforts to employ these novel assays in an ultrahigh-throughput setting are further discussed. Once established, ultra-high throughput screening could greatly facilitate evolution of Kemp eliminases and hopefully, analogous to experiences with other systems [5,138], lead to more efficient artificial enzymes by directed evolution.
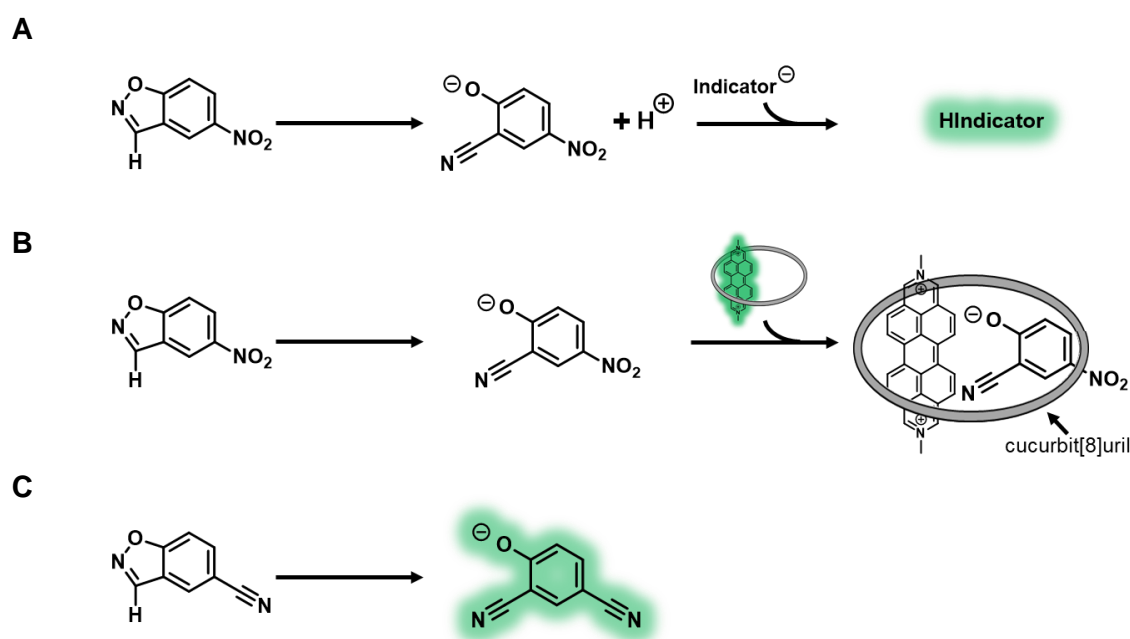
Figure 58: Strategies for ultrahigh-throughput screening of Kemp eliminases. **(A)** Formation of the acidic phenol can be detected using a fluorescent pH-indicator. **(B)** Product formation may be detected by binding to a fluorescent supramolecular host-guest system, composed of cucurbit[8]uril (gray circle) and a fluorophore such as MDPP. **(C)** Conversion of other Kemp eliminase substrates may give a direct fluorescent readout.

4.1 Results

Fluorogenic pH sensors. Reactions such as ester hydrolysis consume base which can lead to changes in pH. These changes can be monitored by addition of a pH indicator [222]. In the

Kemp elimination, deprotonation of 5-nitrobenzisoxazole **2** generates a phenolate ($pK_a = 4.1$, [19]), a process that may be also monitored in this way (Figure 59).

Here, the Kemp elimination reaction was performed in a whole-cell assay with *E. coli* cells that express Kemp eliminase HG3.17. Upon addition of the substrate 5-nitrobenzisoxazole **2**, base consumption leads to a pH-drop which was monitored by addition of a pH indicator. Fluorogenic small molecules, such as fluorescein **14** and a BODIPY-based indicator **15** (Table 11), as well as the green fluorescent protein [223] gave a fluorescent signal as the Kemp elimination of 5-nitrobenzisoxazole **2** proceeded. Similarly, non-fluorescent neutral-red can be employed to monitor pH changes upon Kemp elimination. However, fluorogenic indicators with a spectral overlap with the Kemp elimination product, like an anthracene-based indicator **16** (Table 11), gave no signal and were likely quenched by formation of the product.
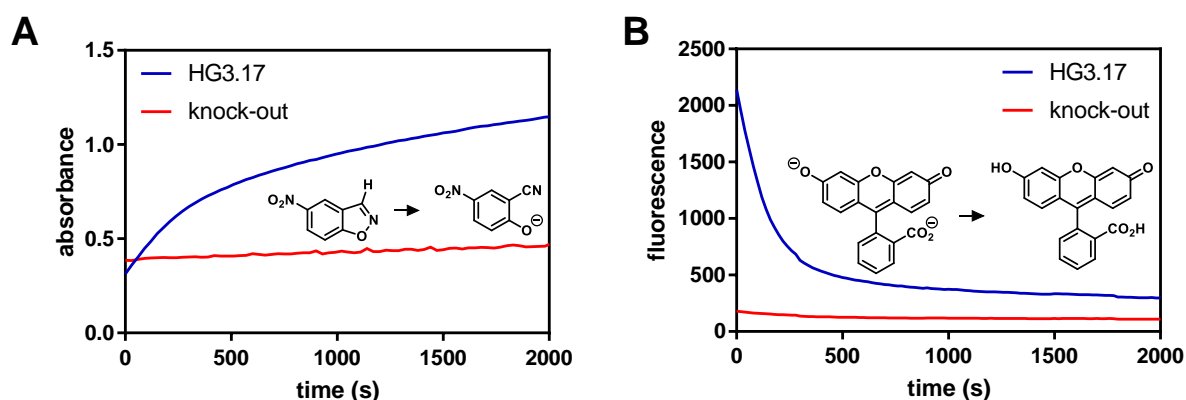
Figure 59: pH-based fluorescent readout of Kemp eliminase activity. Activity of *E. coli* cells expressing either HG3.17 (blue) or a knock-out (red). The conversion at a concentration of 300 cells/ml was monitored by following the product absorbance **(A)** or protonation of a pH indicator by base consumption **(B)**.

Fluorogenic supramolecular chemosensor. An alternative fluorescence readout for product formation was based on a fluorescent supramolecular chemosensor [224]. The employed chemosensor relied on supramolecular interactions of a fluorophore with a cucurbit[8]uril host. Association of a small molecule, like the reaction's substrate or product, to the supramolecular host modulates its fluorescence. The reaction can thus be monitored through changes in the host-guest equilibria. Kemp elimination of 5-nitrobenzisoxazole **2** was monitored with two different fluorescent dyes complexed with cucurbit[8]uril (Table 11). The fluorescence was assayed with the chemosensor alone, after addition of HG3.17 and 5-nitrobenzisoxazle **2** and over the course of the reaction (Figure 60). The resulting fluorescence correlated with product formation. Unfortunately, the dynamic range of the

fluorescence assay was limited to a ~2-fold signal change. As a consequence, this approach is not suitable for ultrahigh-throughput screening.
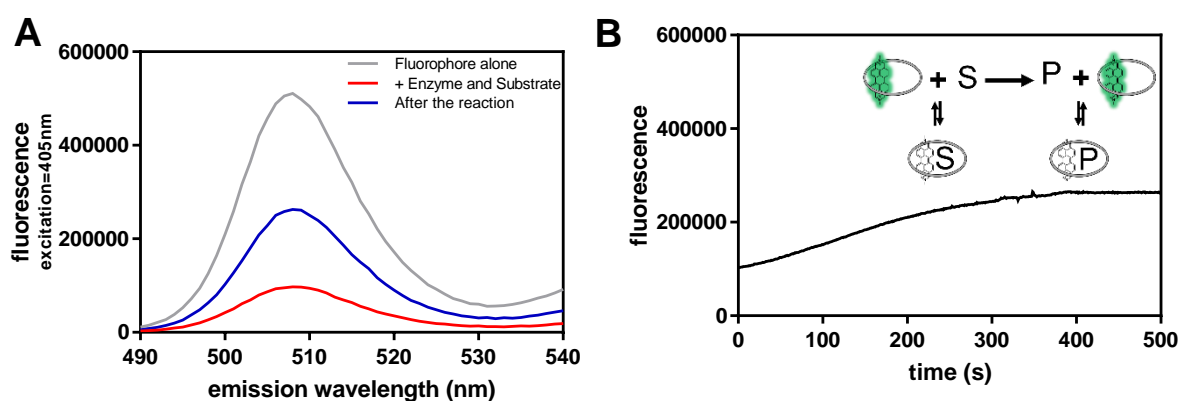
Figure 60: Supramolecular fluorescent readout of Kemp eliminase activity. **(A)** The supramolecular fluorescent sensor (gray) decreases in fluorescent upon substrate addition (red) and is increased after the reaction (blue). **(B)** Time course of the fluorescence change which is caused by changes in the host-guest equilibria of the fluorescent dye (green), cucurbit[8]uril (gray circle), and the Kemp elimination substrate (S) and product (P).

Fluorescence readout via a fluorogenic substrate. Another assay was established relying on the intrinsic fluorescence of product derivatives. Though fluorogenic coumarin-derived substrate for the Kemp elimination were reported earlier [225], their steric bulk is unlikely to fit in the active sites of designed Kemp eliminases. Here, differently substituted benzisoxazoles were screened for their fluorogenic properties. It was discovered that Kemp elimination of 5-cyanobenzisoxazole **17** yielded a readily fluorescent product (Figure 61). The reaction could thus be monitored by following either the product absorbance or fluorescence. Importantly, the substrate is accepted by Kemp eliminase HG3.17 and shows comparable activity to 5-nitrobenzisoxazole **2**. However, the excitation wavelength of the product (331 nm) will limit its applicability for some ultrahigh-throughput systems which do not have lasers that can excite dyes at such low wavelengths. Nevertheless, a 20-fold increase in fluorescence upon Kemp elimination was observed, which is likely sufficient for an ultra-high screening system.
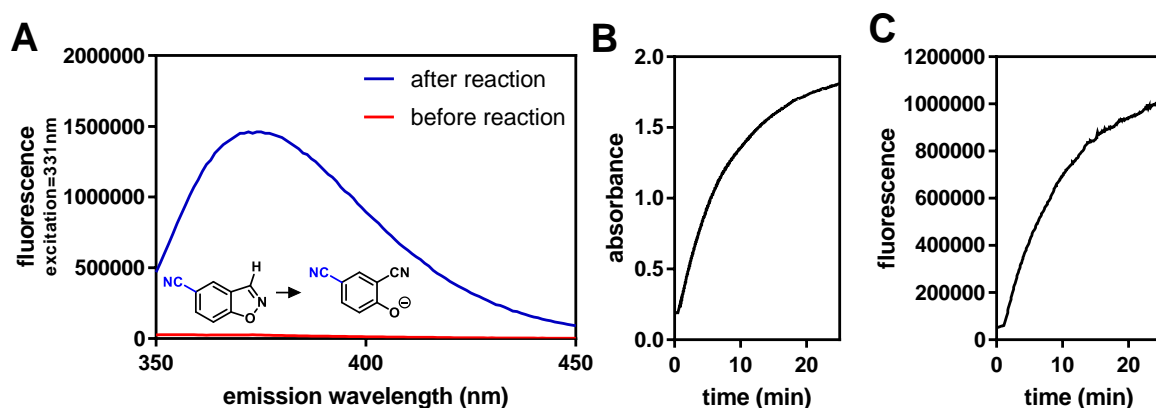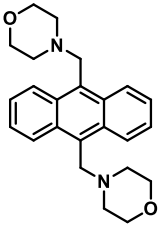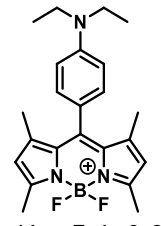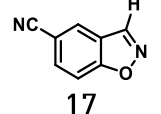
Figure 61: Substrate-based fluorescent readout of Kemp eliminase activity. **(A)** Kemp elimination of 5-cyanobenzisoxazole gives a fluorescence signal. The reaction can thus be monitored by **(B)** absorbance or **(C)** fluorescence.

Table 11: Properties of fluorescence detection systems					
Name	Excitation wavelength	Emission wavelength	Dynamic range	Direction [a]	Lit.
pH-based systems					
anthracene-based indicator 16 pK_a = 5.4±0.2	373 nm	425 nm	~10-fold	turn-on	[226]
BODIPY-based indicator 15 pK_a = 5.4±0.2	468 nm	507 nm	n.d.	turn-on	[227]
fluorescein 14 pK_a = 6.4	485 nm	515 nm	n.d.	turn-off	
fluorescence supramolecular chemosensor					
MDAP [b]	405 nm	424 nm	2-fold	turn-off	[224]
MDPP [b]	424 nm	508 nm	2-fold	turn-on	[224]
substrate-based system					
17	331 nm	375 nm	20-fold	turn-on	

[a] Direction indicates, whether the fluorescent signal will increase (turn-on) or decrease (turn-off) upon Kemp elimination. [b] The respective dyes were complexed by mixing with cucurbit[8]uril as described in [224].

Ultra-high throughput screening. Having established three fluorescence-based screening assays for Kemp eliminase activity, the suitability of these assays for ultrahigh-throughput screening was next investigated. Early attempts aimed at monitoring the Kemp elimination of colonies on agar-plates using neutral-red for visualization of the pH-drop, according to esterase assays performed by Bornscheuer [222]. Attempts to monitor the activity of colonies, however, failed due to the high background reaction on plates, as well as bleeding of the signal from colonies. Furthermore, the reaction could be monitored with indicators such as the cell impermeable fluorescein [228] suggests that the pH is decreased outside the cell. Hence, the substrate likely penetrates the cell, is converted to product and the product is detected outside the cells after release. Screening for the reaction with whole cells in a FACS setting, relying on localization of the signal inside the cell, was thus not possible.

Localization of the signal to the cells was subsequently attempted using droplet-based microfluidics. Cells and reagents can be compartmentalized in droplet emulsions using microfluidics leading to a robust link between activity and signal. In order to implement the pH-based assay, two different droplet populations, with and without the 5-nitrobenzisoxazole **2** substrate, were formed. Both populations additionally contained Kemp eliminase HG3.17 and the pH-indicator fluorescein. The buffer strength for the pH-based assay needed to be decreased to allow a pH change upon reaction. Unfortunately, the pH equilibrated between the weakly-buffered droplets within seconds after mixing of the two populations and the active and inactive droplets became indistinguishable. Alternative attempts to use the fluorescent supramolecular chemosensor in droplets led to rapid clogging of the microfluidics chip, possibly due to precipitation of the host-guest system with either the fluorinated oil or the surfactant. Lastly, experiments with the fluorogenic substrate did not give any signal with the microfluidics station available at the Hilvert lab. The lowest excitation wavelength of the setup was 375 nm and not sufficiently low to excite the fluorescent product (λ_{ex} = 331 nm).

As an alternative to droplet-based microfluidics, implementation of the pH-based assay with µSCALE was attempted. Due to the physical separation of the individual capillaries, it was anticipated that µSCALE would give a stable signal for pH-based readouts. In order to load the µSCALE chip with the reaction mixture, a suspension of yeast cells that display HG3.17 on their surface was mixed with a solution containing the substrate 5-nitrobenzisoxazole **2** and fluorescein (Figure 62). The fluorescein dye decreases in fluorescence when protonated, thus darker wells indicate Kemp eliminase activity. While yeast cells which did not display Kemp eliminase HG3.17 gave a homogeneous population, wells with decreased fluorescence were

observed with yeast cells displaying HG3.17 on their surface. However, the µSCALE setup suffers from rapid bleaching. The issue of bleaching can be partially overcome by dimming the incoming light. Nevertheless, some pH indicators, like the BODIPY-based dye **15**, are bleached too rapidly to be employed in this setting. Future experiments, including an active-site knock-out of HG3.17 as negative control and more homogeneous loading of the chip, may be able to build on these preliminary results.
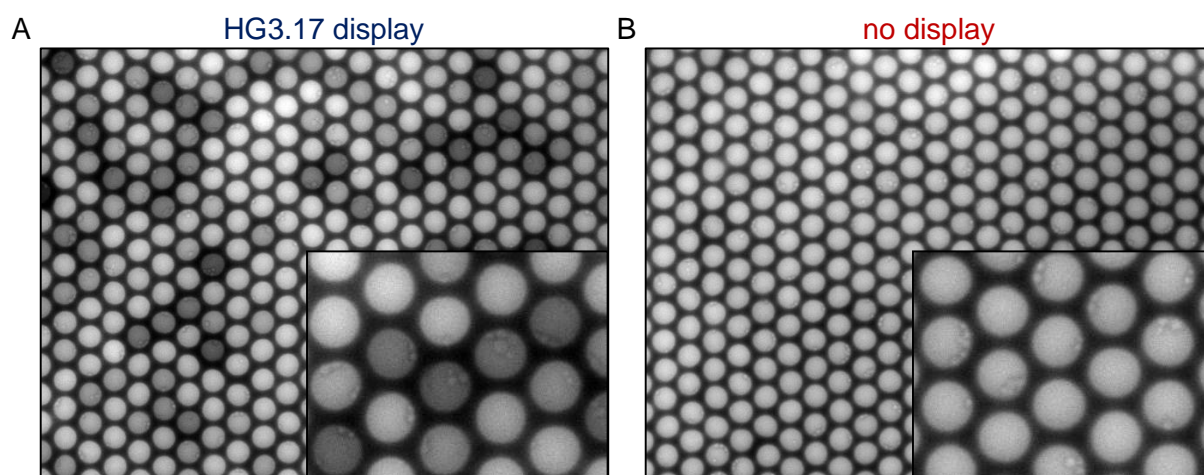
Figure 62: Kemp eliminase assay in the µSCALE. Yeast cells are visible (inset, blobs) in the capillaries of the microchip (black grid). A decrease in fluorescence indicates protonation of fluorescein and thus Kemp eliminase activity. Kemp eliminase activity was detected for induced cells **(A)**, while non-induced cells **(B)** did not show any activity. Only 50% of the induced cells displayed HG3.17. Furthermore, the wells were not uniformly loaded, which caused variations in the fluorescence signal.

4.2 Discussion

Enzyme engineering by directed evolution relies on efficient methods to efficiently sample sequence space. The success of such endeavors can profit from ultrahigh-throughput screening methods that allow coverage of larger DNA libraries. Implementation of such approaches has led to improvements beyond those achieved in microtiter-plate screening systems in several cases [5,138]. For example, the efficiency of the *de novo* engineered retro-aldolase RA95.5-8 could be improved by an additional 30-fold using an ultrahigh-throughput droplet-based microfluidics screen, after evolution with a microtiter-plate screen had stalled. Further methods development will be required to make emerging ultrahigh-throughput technologies such as droplet-based microfluidics or µSCALE accessible to a broader range of reactions.

Most ultrahigh-throughput setups rely on a fluorescence readout. A potential ultrahigh-throughput screening assay for Kemp eliminase activity thus requires linking product formation with a change in fluorescence. Several assays for the reaction have been developed making it possible to monitor product formation by detecting changes in pH, by exploiting supramolecular interactions, or by direct monitoring of a fluorescent product. The pH- and substrate-based assays gave good signals with a high dynamic range. In contrast, the fluorescence supramolecular chemosensor seems to be poorly suited for an ultrahigh-throughput screen due to its limited 2-fold dynamic range.

The substrate-based screening with 5-cyanobenzisoxazole **17** for Kemp eliminase activity using the droplet-based microfluidics system of the Hilvert lab was not possible due to the product's low excitation wavelength of 331 nm. Nevertheless, this approach is still attractive due to the direct readout and turn-on fluorescence. Derivatization of the substrate may yield more red-shifted products that can be analyzed with conventional systems, such as the microfluidics system with a minimum excitation wavelength of 375 nm. Engineering of a droplet-based screening system was consequently attempted with the pH-based assay. The rapid exchange of pH between droplets, however, prevents assaying changes in pH in individual compartments. Since the expected drop in pH is small, it would be necessary to slow down the pH exchange by several orders of magnitude. Unfortunately, pH equilibration between droplets is not well understood and cannot be suppressed in a straightforward manner.

Fortunately, initial μSCALE experiments demonstrate the feasibility of pH assays in an ultra-high throughput fashion. Compared to the droplets, the capillaries of the μSCALE setup are separated by a true physical barrier. No equilibration of the fluorescence signal between individual capillaries was observed. However, bleaching represents a significant practical hurdle for μSCALE. The setup is based on confocal microscopy and thus has elevated light intensity in the focal area. Monitoring conversion with fluorescein was possible but required rapid analysis of the μSCALE chip. Furthermore, fluorescein decreased in fluorescence upon Kemp elimination. For a robust screening assay, a dye with turn-on fluorescence, like the BODIPY-based indicator **15**, would be more desirable. An increase in signal is easier to monitor experimentally and also less prone to artifacts, such as bleaching. Unfortunately, the BODIPY-based dye **15** rapidly bleached and could not be used. Another dye that combines photostability, turn-on readout, and no spectral overlap with the Kemp elimination product would be desirable for a successful μSCALE screening setup.

This work establishes the foundation for ultrahigh-throughput screening assays for Kemp eliminase activity. Though further methods development is required, a viable ultrahigh-throughput screening assay will, hopefully, facilitate the directed evolution of Kemp eliminases, leading to better *de novo* engineered biocatalysts.

4.3 Materials and methods

Chemical synthesis

10-(4-(Diethylamino)phenyl)-5,5-difluoro-1,3,7,9-tetramethyl-5H-dipyrrolo [1,2-c:2',1'-f][1,3,2]diazaborinin-4-ium-5-uide 15 according to [227].
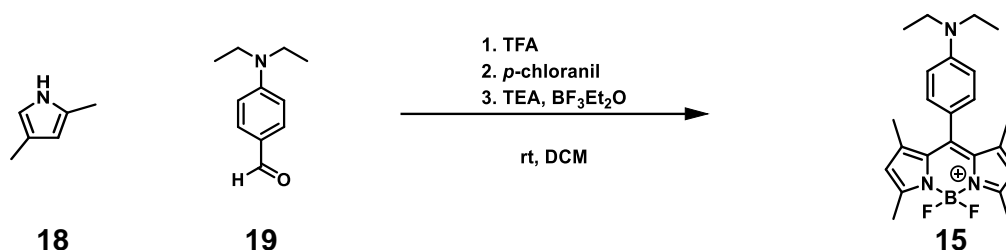
2,4-Dimethylpyrrole **18** (158 mg, 1.6 mmol, 2 eq), 4-(diethylamino)benzaldehyde **19** (147 mg, 0.8 mmol, 1 eq) and trifluoroacetic acid (0.5 ml) were added to dichloromethane (150 ml), and the suspension was stirred overnight at room temperature. p-Chloranil (204 mg, 0.8 mmol, 1 eq) was added and the solution was stirred for 30 minutes. Triethylamine (1.6 ml) and boron trifluoride in Et_2O (1.6 ml) were added and the solution was stirred for one hour. The organic phase was washed with water (30 ml) and brine (30 ml), dried over sodium sulfate and the solvent was removed under reduced pressure. Product **15** was obtained as a red solid after flash column chromatography on silica gel eluting with hexane / dichloromethane / triethylamine (100:100:1, R_f = 0.37).

Yield: 100 mg (0.3 mmol, 19%).

1H-NMR: (300 MHz, CDCl3) δ [ppm] = δ 7.01 (d, 2H, J = 8.7 Hz), 6.73 (d, 2H, J = 8.7 Hz), 5.66 (s, 2H), 3.39 (q, 4H, J = 7.1 Hz), 2.54 (s, 6H), 1.51 (s, 6H), 1.19 (t, 6H, J = 7.1 Hz)

9,10-Bis(morpholinomethyl)anthracene 16 according to [226].

20   **16**
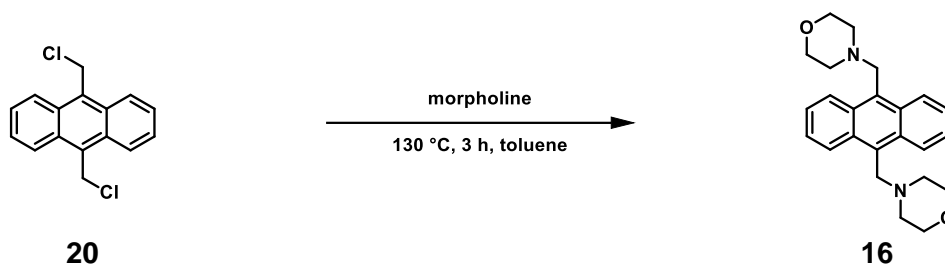
9,10-Bis(chloromethyl)anthracene **20** (0.2 g, 0.72 mmol, 1.0 eq) and morpholine (250 µl, 2.88 mmol, 4.0 eq) were added to toluene (2 ml) and the suspension was heated to 130 °C for 3 hours. The suspension was cooled to room temperature. The precipitate was dissolved by addition of 1 M HCl. The organic phase was separated. Product **16** was precipitated by addition of 10 M NaOH. The solid was removed by filtration and the product **16** was obtained as a yellow solid purified by flash column chromatography on silica gel eluting with hexane/ethyl acetate/triethylamine (100:50:1, R_f = 0.37).

Yield: 0.059 g, (0.16 mmol, 22%).

¹H-NMR: (300 MHz, CDCl₃) δ [ppm] = 8.47 (4H, q, J = 3.7 Hz), 7.46 (4H, q, J = 3.4 Hz), 4.39 (4H, s), 3.57 (2H, t, J = 4.7 Hz), 2.56 (2H, t, J = 4.6 Hz)

5-Cyanobenzisoxazole. This compound was prepared as previously described from 3-formyl-4-hydroxybenzonitrile via oxime formation and PPh₃/DDQ-promoted cyclization [4,62,171,172].

Protein production and purification. BL21-Gold(DE3) cells were transformed with the pET11 vector harboring HG3.17 and plated on LB agar plates (with 100 µg/ml ampicillin). A single colony was used to inoculate an overnight culture of LB medium (with 100 µg/ml ampicillin). 500 ml of LB medium (with 100 µg/ml ampicillin) were inoculated in a ratio of 1:100 from the overnight culture and grown to an OD of 0.5 at 37 °C. Protein production was induced with 1 mM IPTG and cells were incubated overnight at 18 °C. The cells were harvested and the cell pellet was frozen at -80 °C or directly processed further for protein purification. The cell pellet was resuspended in lysis buffer (20 ml 50 mM tris-HCl, 0.5 M NaCl, 10 mM imidazole, 100 µL protease inhibitor cocktail (*Sigma*), 1 mg/ml lysozyme, DNAseI, pH 7.4). The cell suspension was sonicated for 1 h in a sonication bath at 4 °C (*Telesonic Ultrasonics*). After centrifugation (18.000 rpm, 20 min, 4 °C). The soluble fraction was applied to 3 ml Ni-NTA slurry (*Qiagen*), washed with 10 mM imidazole before elution with 300 mM imidazole, each in

50 mM tris-HCl, 0.5 M NaCl, pH 8. The buffer was exchanged to 20 mM sodium phosphate, 20 mM NaCl, pH 6.0 by washing with an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*) three times. Proteins were then further purified by anion exchange chromatography (MonoS column, *GE Healthcare*) and eluted with a salt gradient (20 mM to 1 M NaCl, 20 mM sodium phosphate, pH 6.0). Fractions absorbing at 280 nm were analyzed with SDS-PAGE (Phastsystem, *Pharmacia Biotech*) using the low molecular weight marker (*GE Helthcare*). Fractions containing the target protein were pooled and concentrated using an Amicon Ultra-15 unit with a 10 kDa molecular cut-off (*Millipore*), if necessary. Protein concentrations were determined by measuring the absorbance at 280 nm using a calculated extinction coefficient (ε(HG3.17) = 51,910 M^{-1}cm^{-1}, [175]). Protein masses were checked by mass spectrometry. For mass determination, protein samples were exchanged into 0.1% acetic acid with Vivaspin 500 centrifugal filters (*Sartorius*) and measured by ESI-MS on a Daltonics maXis ESI-Q-TOF mass spectrometer (*Bruker*). Mass spectra were deconvoluted using the MaxEnt1 software. All masses corresponded to the expected sequence lacking the N-terminal methionine.

pH-based fluorescence assay. Fluorescence and absorbance spectra were measured on a Varioscan plate reader (*ThermoFisher*). The pK_a of the indicators as well as their spectroscopic properties were determined by pH titration in 50 mM sodium phosphate, 50 mM sodium acetate and 100 mM NaCl, pH 4.0 to 10.0. Time curves were analyzed at the wavelength listed in Table 11 and were performed with 1 mM 5-nitrobenzisoxazole **2**, 10 µM indicator, 160 mM sodium chloride, 0.1% DMSO, 10% methanol, pH 7.0. *E. coli* BL21-Gold(DE3) cells expressing either HG3.17 or an active-site knock-out (HG3.17 D127N) were freshly transformed and picked from LB-amp agar plates and used to inoculate 5 mL LB medium containing 100 µg/mL ampicillin (LB-amp). Cells were grown overnight at 30 °C and 230 rpm in a rotary shaker, and protein expression was induced by 1 to 10 dilution into 5 LB medium containing 100 µg/mL ampicillin and 250 µM IPTG. After 16 h incubation at 20 °C and 230 rpm, cells were diluted directly in assay buffer to the appropriate concentration.

Fluorescent chemosensor assay for Kemp eliminase activity. Fluorescence spectra and time curves were recorded on a fluorometer (*Photon Technology International*). 500 mM MDPP/cucurbit[8]uril or 5 µM MDPA/cucurbit[8]uril, prepared as described in [224], were dissolved in 440 µl 20 mM sodium phosphate, 20 mM sodium chloride, pH 7.0. Spectra and time curves were recorded after addition of 50 µl 1 mM 5-nitrobenzisoxazole **2** in methanol and 10 µl 70 nM HG3.17.

Substrate-based fluorescence assay. Fluorescence spectra, as well as absorbance and fluorescence time curves, were recorded on a Varioscan platereader (*ThermoFisher*). Time curves for the conversion of 120 µM fluorogenic 5-cyanobenzisoxazole **17** were recorded after addition of 8 nM HG3.17 in 50 mM sodium phosphate, 100 mM sodium chloride, 10% methanol, pH 7.0. Absorbance changes were monitored at 380 nm and fluorescence was monitored at 375 nm upon excitation at 331 nm.

Droplet-based microfluidics. The microfluidics station was set up as previously described using a second-generation integrated chip [5] for all experiments. The samples were prepared in the same way as for the respective bulk assays. In addition, two different droplet populations were produced, either with or without the substrate in the aqueous phase. Furthermore, 2 nM or 10 nM fluorescent kiton red were added to the aqueous phase as barcodes to allow discrimination of the two populations. The kiton red was detected at 561 nm upon excitation at 609 nm. The two different droplet populations were prepared in parallel on the chip by using the junctions of the channels behind the droplet nozzle for droplet productions. For the supramolecular assay rapid clogging of the chip prevented any analysis. The pH-based assay was performed for fluorescein. Fluorescein was excited at 488 nm and fluorescence was detected at 520 nm. The fluorescence of the two populations equilibrated rapidly after mixing. For the product-based assay, excitation at the lowest possible wavelength of 375 nm and detection at 448 nm did not give any signal.

Yeast surface display of HG3.17. HG3.17 was cloned into pCTRB [229] using the NdeI and XhoI restriction enzymes, Phu polymerases and T4 ligase from *New England Biolabs*. The gene coding for HG3.17 was amplified from a pet11 plasmid by PCR using the T7 terminator reverse primer (AAG ACC CGT TTA GAG GCC CCA A) and the T7 promotor forward primer (GAT CCC GCG AAA TTA ATA CGA CTC ACT ATA GG). Vector and Insert were sequentially digested with NdeI overnight at 37 °C and XhoI for 2 h at 37 °C. After ligation with T4 DNA ligase, Xl1blue cells were transformed with the DNA by electroporation [193]. EBY100 yeast cells were transformed with the pCTRB plasmid harboring HG3.17 using the Frozen-EZ Yeast Transformation II Kit (*Zymogene*), streaked out on SD-CAA agar-plates and incubated for 2-3 days at 30 °C. Cells were picked and grown in 2 ml SD-CAA overnight at 30 °C and 230 rpm. A fresh culture of 2 ml SD-CAA was inoculated with the initial culture and cells were incubated for 7 h at 30 °C and 230 rpm. Cells were induced by addition of SG-CAA and were incubated overnight at 18 °C and 230 rpm. The cells were pelleted by centrifugation, resuspended in PMB (20 µl), and incubated with monoclonal mouse anti-c-myc antibody 9E10 (*Roche*, 250 ng/µl).

After incubation at 4 °C for 30 min, the cells were pelleted by centrifugation, resuspended in PMB (20 µl), and labelled with goat anti-mouse IgG-FITC antibody (*Sigma*, 50 ng/µl) at 4 °C for another 30 min. After labelling, the cells were pelleted by centrifugation and washed with PMB (3 x 180 µl). Labelled cells were resuspended in PMB and analyzed on a LSRFortessa (*BD*) at the Flow Cytometry Core Facility of ETH Zurich. Data was analyzed using the FlowJo software (*LLC*).

µSCALE analysis of HG3.17. The µSCALE was set up as previously described [204]. Briefly, yeast cells displaying HG3.17 were diluted into 1 mM 5-nitrobenzisoxazole **2**, 10 µM fluorescein, 160 mM sodium chloride, 0.1% DMSO, 10% methanol, pH 7.0 and immediately loaded onto a µSCLAE chip. The chip was covered with an agarose-slab to prevent evaporation. The chip was analyzed for fluorescence intensity with 525/45 nm excitation and 565 nm long-pass emission filters. An additional dimmer which eliminates 90% of the incoming light was applied to prevent photobleaching. Nonetheless, photobleaching was observed after a partially reversible, rapid decay of fluorescence upon irradiation. In order to keep the irradiation homogenous, images were typically obtained by rastering over the µSCLAE chip.

5 Exploiting the promiscuity of designed Kemp eliminases for industrially relevant transformations ‡

Promiscuous enzyme activity often provides a useful starting point for the engineering of new biocatalysts [3]. Minor side activities of natural enzymes have been exploited and refined by directed evolution for several transformations for practical application. In a landmark paper, researchers from Merck and Codexis evolved a transaminase for the enantioselective synthesis of the antidiabetic drug sitagliptin [230]. In this case, successful directed evolution relied on a substrate walk to increase the size of the binding pocket for the target substrate. Optimization for each intermediate substrate offered novel substrate promiscuity for even larger ligands. Another level of promiscuity was exploited by Arnold, who, instead of relying on substrate ambiguity, utilized the chemical promiscuity of natural heme-proteins to catalyze a manifold of abiological chemical transformations [231]. For instance, starting from a catalytically promiscuous cytochrome P450, exchange of the axial heme ligand and subsequent directed evolution lead to efficient and selective catalysts for non-natural carbene and nitrene transfer reactions.

Catalytic promiscuity may have various mechanistic origins [159]. First, enzymes often accept alternative substrates that undergo the same reaction mechanism but have different substituents [230,232,233]. As discussed in Chapter 2.3, promiscuity may also be enhanced by conformational diversity [158,234,235]. For instance, a bacterial phosphotriesterase acquired promiscuous aryl esterase activity by altering the dynamics of lid-forming loops that cover the active site [160]. Different sub-sites within the same active site may also give rise to new activities [236-238]. In this context, the promiscuous transformation may employ a different set of catalytic residues than the natural reaction. In addition, alternative cofactors [239], including exchange of metal cofactors [240,241] or differently oriented water molecules [242], can yield novel reactivities. Finally, varying protonation states of catalytic residues may also alter enzyme reactivity [243,244]. For example, the catalytic N-terminal Pro1 in trans-3-

---

chloroacrylic acid dehalogenase (pK_a ~9.2) acts as a general acid whereas in the related 4-oxalo-crotonate tautomerase Pro1 (pK_a~6.4) Pro1 serves as a general base. Based on the pK_a of the tautomerase, a small fraction of the catalytic residue is present in its protonated form and the catalyst thus displays promiscuous dehalogenase activity [245].

Another example of promiscuous Brønsted acid catalysis was reported by Hauer, who diverted the substrate promiscuity of squalene-hopene cyclase towards industrially relevant monoterpenes [246-248]. Squalene-hopene cyclase utilizes a catalytic Asp with a highly perturbed pK_a for protonation of a C=C double bond in their substrates. Protonation of the linear triterpene squalene, which is precisely folded at the enzyme's active site, induces a cascade cyclization reaction leading to the pentacyclic hopene or hopenol [246]. However, Hauer discovered that squalene-hopene cyclases also accept smaller mono- and sesquiterpenes with good selectivity after introduction of bulky residues to shrink the active site (Figure 63, [247,248]). Further diversification led to catalysts that produce industrially relevant cyclic monoterpenes, such as (-)-isopulegol **23**, with good enantiomeric excess [249,250].
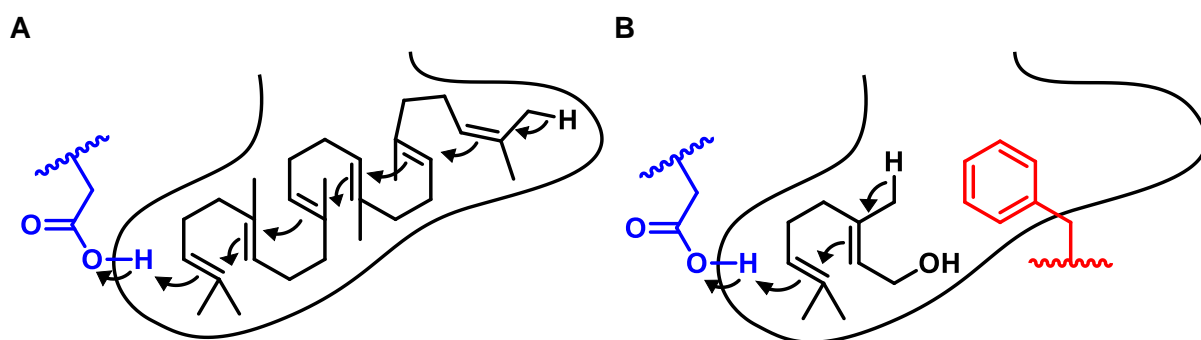
Figure 63: Substrate promiscuity of squalene-hopene cyclases. **(A)** Protonation of a C-C double bound of the wild-type substrate squalene (black) by the general acid (blue) induces a cascade cyclization reaction. **(B)** Introduction of bulky residues, like G600F (red), shrinks the binding pocket and increases cyclization activity for monoterpenes such as geraniol (black).

In 1976, Jensen formulated the hypothesis that primordial enzymes were often promiscuous, a property that facilitated the gradual evolution of highly specific natural enzymes for various reactions from common ancestors [251]. His hypothesis has since been backed up by the discovery of countless enzyme families and superfamilies [252]. Furthermore, reconstruction of common ancestors from phylogenetic relationships [253] often [254], but not always [255], yields promiscuous catalysts. In this context, computationally designed enzymes may also be regarded as primordial catalysts due to their naïve activity. Catalytic promiscuity is, for instance, well documented in retro-aldolase RA95 and its

evolutionary descendants. Utilizing a key catalytic intermediate, the Schiff-base formed by the substrate's ketone and a Lys side chain, a diverse range of chemistries have been explored. Promiscuous activity was found for Knoevenagel condensations with ketones and aldehydes [256], Michael additions to α,β-unsaturated ketones [257,258] and Henry reactions of nitroalkanes [259]. In line with Jensen's hypothesis [251], the evolutionary intermediate RA95.5-8 turned out to be far more promiscuous than its highly specialized evolutionary descendant RA95.5-8F.
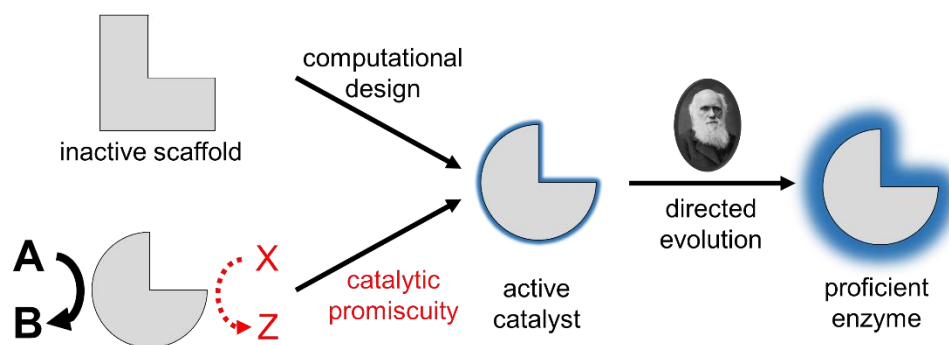
Figure 64: Catalytic promiscuity (red) can provide alternative starting points for the engineering of novel, proficient enzymes via directed evolution.

Catalytic promiscuity offers a valuable alternative origin for the evolution of novel enzymes (Figure 64). Owing to the primordial nature of computational designs, promiscuous activities of designed catalysts seem particularly promising. Here, the promiscuous activity of the designed Kemp eliminase 1A53-2 is explored [103]. The elevated pK_a of the general base suggested that promiscuous acid-catalyzed cyclization of monoterpenes, analogous to the aforementioned squalene-hopene cyclase [246-248] might be possible. Indeed, catalytic promiscuity for the reaction with various monoterpenes was observed. The evolvability of the promiscuous activity for the cyclization of citral, leading to an industrially relevant (-)-menthol precursor, was investigated. (-)-Menthol, a popular additive in the flavor and fragrance industry [260,261], is produced on a 20,000-ton scale per year [262]. Evolution of artificial enzymes capable of stereoselective cyclization of one of its synthetic precursors, would significantly shorten the lengthy industrial synthesis of this commodity chemical (Figure 69, [263-265]).

5.1 Results

Computationally designed Kemp eliminase 1A53-2 utilizes Glu178 to promote the C-H deprotonation of benzisoxazoles [103]. Kinetic characterization revealed that 1A53-2 acquires full activity only above pH 9, suggesting an unusually high pK_a for the catalytic carboxylic acid of 8.7±0.2 (Figure 65). Furthermore, the hydrophobic active site, with two Trp residues packing the benzisoxazole substrate, seems nicely suited to accommodate monoterpenes similar to those accepted by squalene-hopene cyclases [246-248]. Thus, it was hypothesized that the general base might function as a general acid at lower pH values.
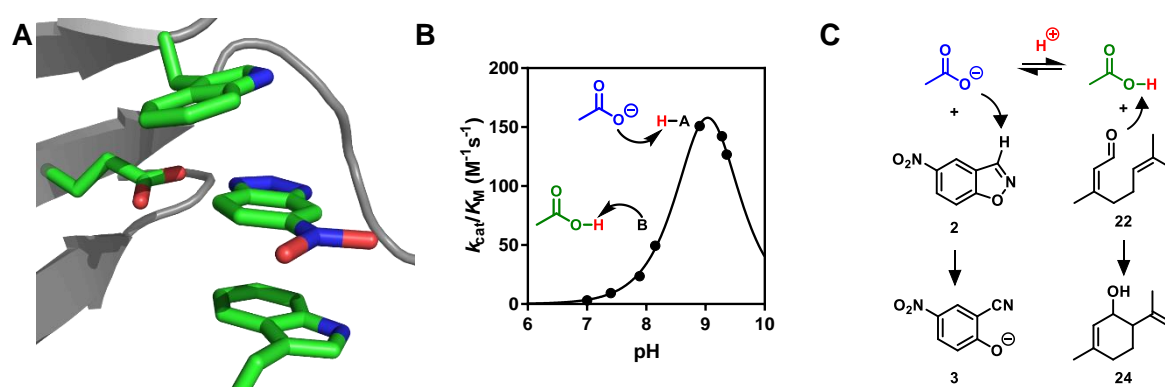
Figure 65: Promiscuous catalysis in Kemp eliminase 1A53-2 (PDB: 3NZ1 [103]). **(A)** The transition state analog (6-nitrobenzotriazole **1**) is sandwiched between two tryptophan residues, positioning it for a tight interaction with the catalytic base Glu178. **(B)** The loss of activity of 1A53-2 at low pH suggests protonation of the general base with a pK_a of 8.2±0.2. **(C)** The active site hydrophobicity and elevated base pK_a seem well suited for promiscuous acid-catalyzed cyclization of monoterpenes, as observed for squalene-hopene cyclases [248].

Promiscuous activity. The cyclization activity of 1A53-2, 1A53-2.5 and 1A53-2.9 (Chapter 3), as well as HG3, HG3.7 and HG3.17 [4,103] was tested with *rac*-citronellal **21** and citral **22**. As expected, only 1A53-2 displayed detectable activity at acidic pH (6.0), likely owing to its perturbed pK_a (Figure 66). The reaction product of *rac*-citronellal **21** was identified as isopulegol **23** by comparison to a commercial standard. The reaction of citral **22**, composed of the (*E*)- and (*Z*)-isomers geranial and neral, leads to only one observable product. In agreement with literature precedent [266], this product was assumed to be the corresponding cyclization product **24**. 1A53-2 also accepted two alternative monoterpene substrates of squalene-hopene cyclases [248]. The enzyme accelerated reactions of geraniol and 6,7-epoxygeraniol which led to formation of several unidentified products observed by liquid chromatography.
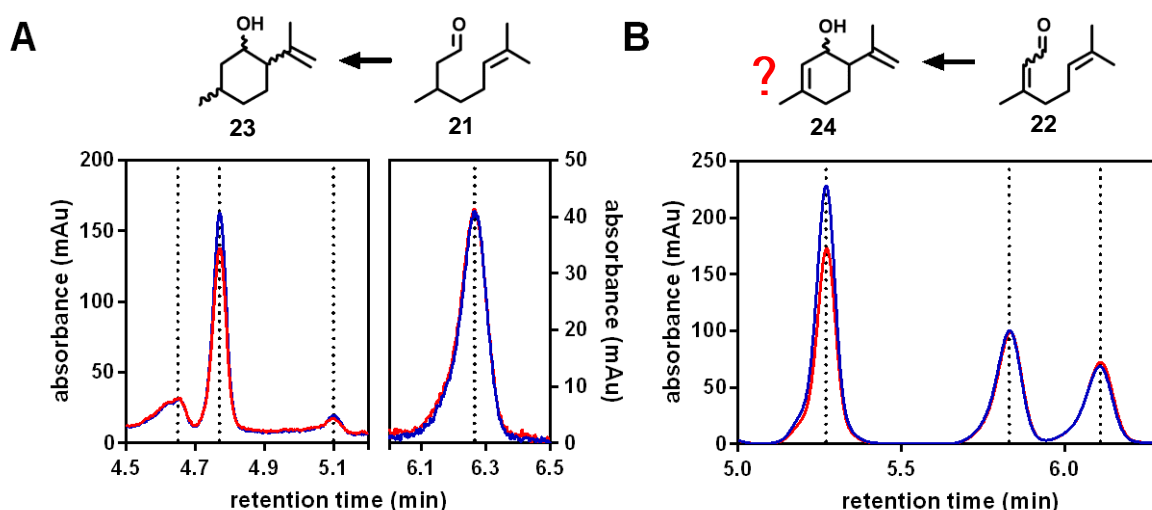
Figure 66: Cyclization activity of Kemp eliminase 1A53-2 (blue) and an active site knock-out 1A53-2 E178Q (red). **(A)** Isopulegol **23** is formed by cyclization of *rac*-citronellal **21**. The product peaks were identified by comparison with a commercial standard. Only three of the four possible diastereomers are observed. The forth diastereomer is only formed in low amounts. **(B)** The reaction of citral **22**, composed of the (*E*)- and (*Z*)-isomers geranial and neral, leads to only one observable product. In agreement to literature precedent [266], this product may be the depicted cyclization product **24**. **(A+B)** Loss of substrate, potentially through unspecific degradation or evaporation, was observed for both reactions. The effects on product formation are thus more pronounced than on substrate consumption.

Directed evolution. Kemp eliminase 1A53-2 was submitted to directed evolution to increase its cyclization activity for citral **22**. Evolution was performed with a whole cell assay in 96-deepwell plates. *E. coli* cells expressing the cyclase variants were incubated with citral **22** at pH 6.0 and conversion was analyzed by liquid chromatography after extraction with butanol. Owing to the limited efficiency of 1A53-2, the observed activity was barely above background in the screens, leading to several false positive hits. The activity of the hits was thus always confirmed with purified protein after each round of directed evolution. Round 1 of directed evolution entailed saturation mutagenesis of residue A157, which gave a 10-fold improvement in Kemp eliminase activity upon mutation to large residues like tryptophan or phenylalanine (Chapter 3). In the case of the cyclase activity, however, only small residues like leucine and glycine were accepted and cysteine gave the best improvement of approximately two fold. In round 2, saturation mutagenesis of individual residues was performed, targeting the other active site residues that had beneficial effects in the evolution of 1A53-2 for Kemp elimination. Again, different mutations were identified compared to the Kemp elimination, reflecting the different steric and chemical requirements of the cyclization reaction. The hits from round 2 were shuffled in round 3. Finally, during round 4, all 21 active site residues were again targeted by saturation mutagenesis and the hits were shuffled in round 5. The mutations

of the best variant from round 5 (Cyc5), round 3 (Cyc3) as well as the starting computational design (Cyc0) and the knockout lacking the catalytic acid (CycX) are listed in Figure 67.

Cyc0: 1A53-2
CycX: Cyc0 E178Q
Cyc3: Cyc0 K53E D111N A157L L184S Q211R
Cyc5: Cyc0 A51G K53E S58W A83C R97H D111N A157L A180S L184S Q211V

```
MPRYLKGWLK DVVQLSLRRP SFRASRQRPI ISLNERILEF NKRNITAIIA GYERKSPWGL DVERDPIEYS
KFMERYAVGL AICTEEKYFN GSYETLHKIA SSVSIPILMW NFIVKESQID DAYNLGADTV ALIVKILTER
ELESLLEYAR SYGMEPLIVI NDENDLDIAL RIGARFIEIS SRDSETLEIN KENQRKLISM IPSNVVKVAW
VGISERNEIE ELRKLGVNAF GIGSSLMRNP EKIKEFILGS IEGRGHHHHH H
```

Figure 67: Sequences and mutations of different cyclase variants. Cyc0 is equivalent to the computational design 1A53-2. CycX is the knockout of the catalytic acid. Mutations introduced up to round three (Cyc3, blue) and round 5 (Cyc5, red) are highlighted. The given sequence is that of the highest evolved Cyc5.

Characterization of purified protein.   Cyclase variants Cyc0, Cyc3, Cyc5 and CycX were produced, purified and characterized (Figure 68). At a citral **22** concentration of 2 mM, Cyc3 had an approximately 4-fold improved cyclization activity compared to Cyc0, while Cyc5 was 12-fold improved. Noteworthy, the second substrate peak with a retention time of 5.4 min was always consumed slightly faster than the first peak at 5.1 min. This observation was most pronounced for Cyc5 and may indicate that this peak corresponds to the (*Z*)-isomer of citral, which can undergo direct cyclization, as opposed to (*E*)-citral which requires isomerization for cyclization [266].
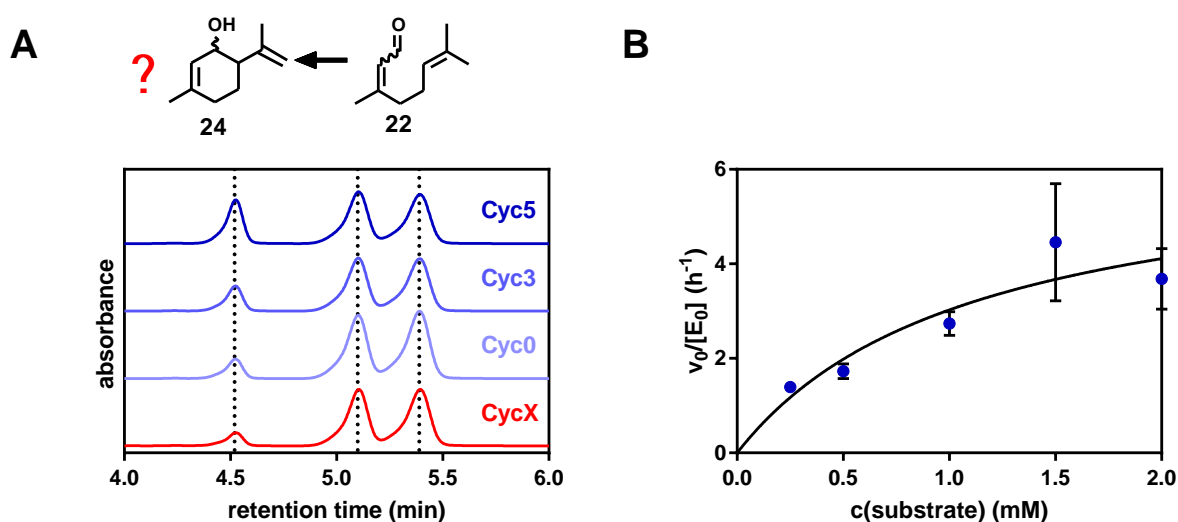
Figure 68: Directed evolution of cyclases. **(A)** Round 5 (Cyc5, dark blue), round3 (Cyc3, blue), the computational design (Cyc0, light blue), and a knockout lacking the catalytic acid (Cyc0 E178Q = CycX, red) were analyzed. An approximately 12-fold improvement was achieved during evolution. **(B)** Michaelis-Menten kinetics revealed turnover number of approximately $6\pm2\,h^{-1}$ based on substrate consumption.

Michaelis-Menten kinetics revealed a turnover number of Cyc5 of 6 ± 2 h^{-1} and a catalytic efficiency of 5 ± 2 M^{-1}s^{-1}, based on consumption of citral **22**. These values were determined using a substrate standard curve to determine concentration, but are subject to relatively large errors because of nonspecific loss of substrate over time. Also, the results varied slightly from analysis to analysis. While the results of each experiment were consistent, variations of the background reaction and overall rate by a factor of up to two fold were observed resulting in large standard deviations. In addition, enzyme production occasionally yielded inactive variants. The reproducibility issues indicate the contribution of unknown factors to the reaction. Careful study of shaking speed, reaction volume, and buffer composition, on the background reaction may help to solve these issues. Furthermore, testing some reactions with the squalene-hopene cyclase as a positive control may be insightful [247].

5.2 Discussion

(-)-Menthol is a fragrance and food additive with a worldwide demand of 20,000 tons per year [262]. Its fresh test and cooling effect is appreciated in a manifold of products ranging from cigarettes [260] to toothpastes and chewing gums [261]. The industrial synthesis of (-)-menthol relies on the synthesis of the non-natural (*R*)-isomer of citronellal **21**, e.g. by asymmetric hydrogenation and subsequent cyclization using a Lewis acid [263-265]. The resulting mixture of cyclized stereoisomers is typically separated by cost-intensive distillation. (-)-Menthol can be subsequently produced by hydrogenation (Figure 69).

Biocatalysis has the potential to shorten the lengthy industrial synthesis of (-)-menthol by exploiting the exquisite selectivity of enzymes. Squalene-hopene cyclases have already been evolved to accept (*R*)-citronellal **21** for enantiospecific formation of the desired diastereomer [247]. An enzyme which stereoselectivly cyclizes the achiral citral **22** would offer a valuable alternative synthetic route. Furthermore, starting from achiral citral **22** for the synthesis of (-)-menthol would circumvent for stereoselective synthesis of (*R*)-citronellal **21**. Subsequent stereospecific hydrogenation of the product can easily be achieved using the resulting alcohol as a steering group [267]. Squalene-hopene cyclases, however, do not accept achiral citral **22** as a substrate [247].
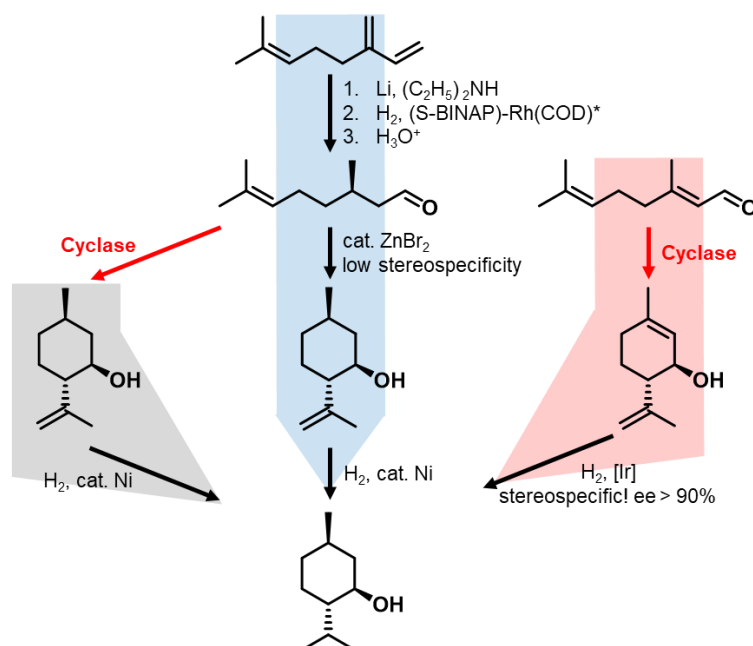
Figure 69: Chemical versus biochemical routes to (-)-menthol. One industrial route to (-)-menthol (cyan) starts from the hydrocarbon myrcene which is converted to (*R*)-citronellal **21** via asymmetric hydrogenation. Subsequent cyclization and hydrogenation affords (-)-menthol. Biocatalysis may alternatively be employed for stereospecific cyclization of (*R*)-citronellal **21** (gray) or stereoselective cyclization of the achiral citral **22** (red).

Here, the promiscuous activity of the computationally designed Kemp eliminase 1A53-2 towards cyclization of citral **22** was investigated. Promiscuity was only observed in the designed 1A53-2, which was enabled by the perturbed pK_a of the catalytic carboxylate. Other Kemp eliminases with less perturbed general bases did not display any activity for the reaction. Furthermore, the high hydrophobicity of the Kemp eliminase active site, which resembles the active site of squalene-hopene cyclases [247], was likely key to the promiscuous activity. Since promiscuous activities are generally low compared to the main activity, they need to be optimized, for instance through directed evolution, to generate practical catalysts [159]. The evolvability of the 1A53-2 catalyst was demonstrated by five rounds of evolution with the substrate citral **22**. The low observed activity during the screen led to several false-positives, which was likely hindering evolution. Thus, the resulting improvements are rather modest. Further evolution will, hopefully, give a catalyst with the potential to significantly shorten the industrial biosynthesis of (-)-menthol ([263-265], Figure 69).

Additional experiments are still required to characterize the enzyme kinetics, verify the identity of the reaction product, and demonstrate further conversion of the cyclized product to (-)-menthol. Nonetheless, this study illustrates the promiscuous potential of naïve, *de novo* enzyme designs.

5.3 Materials and methods

Protein production and purification. BL21-Gold(DE3) cells were transformed with the pET11 vector harboring the gene of interest and plated on LB agar plates (100 μg/ml ampicillin). A single colony was used to inoculate an overnight culture of LB medium (100 μg/ml ampicillin). 500 ml of LB medium (100 μg/ml ampicillin) were inoculated in a ratio of 1:100 from the overnight culture and grown to an OD of 0.5 at 37 °C. Protein production was induced with 1 mM IPTG and cells were incubated overnight at 18 °C. The cells were harvested and the cell pellet was either frozen at -80 °C or directly processed for protein purification. The cell pellet was resuspended in lysis buffer (20 ml 50 mM Tris-HCl, 0.5 M NaCl, 10 mM imidazole, 100 μL protease inhibitor cocktail (*Sigma*), 1 mg/ml lysozyme, DNAseI, pH 7.4). The cell suspension was sonicated for 1 h in a sonication bath at 4 °C (*Telesonic Ultrasonics*). After centrifugation (18.000 rpm, 20 min, 4 °C), the soluble fraction was applied to 3 ml Ni-NTA slurry (*Qiagen*), washed with 10 mM imidazole before elution with 300 mM imidazole, each in 50 mM Tris-HCl, 0.5 M NaCl, pH 8. The buffer was exchanged to 50 mM sodium phosphate, 100 mM NaCl, pH 6.0 using a PD10 buffer exchange column (*GE Healthcare*). Protein concentrations were determined by measuring the absorbance at 280 nm and a calculated extinction coefficient [175]. The following extinction coefficients were determined for some representative enzymes: $\varepsilon(Cyc0) = \varepsilon(Cyc0) = \varepsilon(Cyc3) = 29,910\ M^{-1}cm^{-1}$; $\varepsilon(Cyc5) = 35,410\ M^{-1}cm^{-1}$; Protein masses were confirmed by mass spectrometry. For mass determination, protein samples were exchanged into 0.1% acetic acid with Vivaspin 500 centrifugal filters (*Sartorius*) and measured by ESI-MS on a Daltonics maXis ESI-Q-TOF mass spectrometer (*Bruker*). Mass spectra were deconvoluted using the MaxEnt1 software. All masses corresponded to the expected sequence lacking the N-terminal methionine.

Activity assay. Substrate stocks were prepared in DMSO (200 mM). 20 μl of the DMSO stock solution was added to 980 μl cyclase (10 μM) in 50 mM sodium phosphate, 100 mM NaCl, pH 6.0. Reactions with 6,7-epoxygeraniol, rac-citronellal **21** and citral **22** were incubated for 20 h at 30 °C and reactions with geraniol were incubated 20 h at 37 °C. Reactions in buffer and with the 1A53-2 E178Q and HG3.17 D127N variants (knockout of the Brønsted acid) served as negative controls. After the conversion, the reaction mixture was extracted with 500 μl butanol by shaking for 10 min at 1,000 rpm to allow extraction. The phases were separated by centrifugation (1 min, 16,000 rpm) and the organic phase was analyzed with an Ultimate3000 liquid chromatography system (*Dionex*). Reagents were resolved on a C-18 X-Bridge column

(*Waters*) with a linear 50% ACN/H_2O gradient containing 0.1% TFA. The column was washed in 95% ACN prior to injection of the samples at 5% ACN. The standard for the cyclization product of *rac*-citronellal **21** (*rac*-isopulegol **23**) is commercially available. The main product from the reaction with citral **22** was potentially the corresponding cyclization product, based on literature precedent [266]. For Michaelis-Menten kinetics, the substrate concentration was adjusted while keeping the DMSO concentration constant. The reported kinetics are based on the consumption of the substrate and are corrected for background.

Library construction. Libraries were constructed by overlap-extension PCR with Phusion polymerase (*Finnzymes*) [192]. Fragments were constructed with the primers in the Table 12 and the AKZ3 (TAATACGACTCACTATAGGG) and T7term (GCTAGTTATTGCTCAGCGG) primer. For a typical PCR protocol, samples were incubated 5 min at 98 °C and DNA was amplified in 32 cycles between 98 °C (30 s), 55 °C (30 s) and 72 °C (30 s). After a final elongation for 5 min at 72 °C, the samples were stored at 4 °C. The fragments were purified by agarose gel electrophoresis. Gels usually contained 2% agarose in 0.5 g/l sodium borate ($Na_2B_4O_7$x10H_2O) for small fragments. Gels were stained with ez-vision one (*Amresco*). Gel electrophoresis was performed at 200 V for 20 min at a variable electrical current. DNA extractions from agarose gels were carried out using the Zymoclean Gel DNA Recovery Kit (*Zymo Research*). Fragments were assembled by PCR using a similar protocol as before. 15 cycles were performed without the T7 and T7term primer to assemble the genes and 32 cycles were performed thereafter with the outer primer for amplification. The PCR products were purified via agarose gel electrophoresis (1% agarose). The assembled genes were amplified by PCR, purified with the DNA Clean & Concentrator kit (*Zymo research*) and cloned into pET11. Plasmids were purified with the Zyppy Plasmid Miniprep Kit (*Zymo research*) from *E. coli* Xl1-Blue cells. For cloning, 2 μg of vector and 1 μg of insert were digested with NdeI and BamHI (*New England Biolabs*). Vector and insert were purified via agarose gel electrophoresis (1% agarose). Fragments were ligated using T4 DNA ligase (*New England Biolabs*) and purified with the DNA Clean & Concentrator kit. Electro-competent Xl1-Blue cells were transformed with the library [193], transferred to 10 ml LB and incubated for 1 h at 37 °C and 230 rpm. After 1 h, 0.1% of the culture was transferred onto an LB agar plate (100 μg/ml ampicillin) for estimation of the library size. 100 μg/ml ampicillin were added to the culture and the culture was incubated overnight at 37 °C and 230 rpm. The library was ready for screening after purification with the Zyppy Plasmid Miniprep Kit (*Zymo research*). All DNA concentrations were determined on a

NanoDrop spectrophotometer (*ThermoFischer*). Gene sequences were confirmed by sequencing (*Microsynth*).

Table 12: Cloning primers for the directed evolution of cyclases	
Name	Sequence 5'-3'
Library 1	
1A532_A147_NNK	GTTCTTACGGCATGGAACCGNNKATTGTAATCAACGACGAAAATGACC
1A532_A147/V149_REV	CGGTTCCATGCCGTAAGAAC
Library 2	
1A532_K53_NNK	GCTATCATTGCCGCGTATNNKCGTAAATCCCCGTCTGGTCTGG
1A532_K53_REV	ATACGCGGCAATGATAGC
1A532_S58_NNK	CGTATAAACGTAAATCCCCGNNKGGTCTGGATGTAGAACGTGATCC
1A532_S58/L60_REV	CGGGGATTTACGTTTATACG
1A532 A180_NNK	GCTCGTTTCATCGAAATTNNKAGCCGCGATCTGGAAACTCTGG
1A532_A180/R182_REV	AATTTCGATGAAACGAGC
1A532_L184_NNK	GCTCGTTTCATCGAAATTTGTAGCCGCGATNNKGAAACTCTGGAAATCAACAAAG
Library 3	
CYC3-1F	CCGCTATCATTGCCGCGTATRAACGTAAATCCCCGKYTGGTCTGGATGTAGAACGTG
CYC3-1R	ATACGCGGCAATGATAGCGG
CYC3-2F	GCATTCCGATTCTGATGTGGRATTTTATCGTGAAAGAGTCCC
CYC3-2R	CCACATCAGAATCGGAATGC
CYC3-3F1	GCTCGTTTCATCGAAATTGCTAGCCGCGATTYAGAAACTCTGGAAATCAACAAAG
CYC3-3F2	GCTCGTTTCATCGAAATTTDTAGCCGCGATTYAGAAACTCTGGAAATCAACAAAG
CYC3-3R	AATTTCGATGAAACGAGC
CYC3-4F	CCAACGTTGTCAAAGTTGCGYGGCRGGGCATCTCTGAACGTAACG
CYC3-4R	CGCAACTTTGACAACGTTGG
Library 4	
CYC4-NNK1	CATCACCGCTATCATTGCCNNKTATGAACGTAAATCCCCGTC
CYC4-NNK2	CATCACCGCTATCATTGCCGCGTATNNKCGTAAATCCCCGTCTGGTCTG
CYC4-REV12	GGCAATGATAGCGGTGATG
CYC4-NNK3	CATTGCCGCGTATAAACGTAAANNKCCGTCTGGTCTGGATGTAGAAC
CYC4-NNK4	CATTGCCGCGTATAAACGTAAATCCCCGNNKGGTCTGGATGTAGAACGTG
CYC4-REV34	TTTACGTTTATACGCGGCAATG
CYC4-NNK5	GGAACGTTACGCTGTAGGCCTGNNKATTGCGACCGAAGAAAAGTAC
CYC4-NNK6	GGAACGTTACGCTGTAGGCCTGGCGATTNNKACCGAAGAAAAGTACTTTAACG
CYC4-REV56	CATTGCCGCGTATAAACGTAAA
CYC4-NNK7	GATTGCGACCGAAGAAAAGTACNNKAACGGTAGCTACGAAACGC
CYC4-REV7	GTACTTTTCTTCGGTCGCAATC
CYC4-NNK8	GCTCTGTTAGCATTCCGATTNNKATGTGGAATTTTATCGTGAAAG
CYC4-REV8	AATCGGAATGCTAACAGAGC
CYC4-NNK9	CCTGGGTGCAGATACCGTCNNKCTGATTGTTAAAATCCTGAC
CYC4-REV9	GACGGTATCTGCACCCAGG
CYC4-NNK10	GTTCTTACGGCATGGAACCGNNKATTGTAATCAACGACGAAAATG
CYC4-NNK11	GTTCTTACGGCATGGAACCGTTGATTNNKATCAACGACGAAAATGACC
CYC4-REV101	CGGTTCCATGCCGTAAGAAC
CYC4-NNK12	GTGCTCGTTTCATCGAAATTNNKAGCCGCGATTCAGAAACTCTGG
CYC4-NNK13	GTGCTCGTTTCATCGAAATTGCTAGCNNKGATTCAGAAACTCTGGAAATC
CYC4-NNK14	GTGCTCGTTTCATCGAAATTGCTAGCCGCGATNNKGAAACTCTGGAAATCAAC
CYC4-REV1234	AATTTCGATGAAACGAGCAC
CYC4-NNK15	CGTTGTCAAAGTTGCGTGGNNKGGCATCTCTGAACGTAACG
CYC4-REV15	CCACGCAACTTTGACAACG
Library 5	
CYC5-F1G-1	CATCACCGCTATCATTGCCGSGTATRAGCGTAAAATGCCGTBGGGTCTGGATGTAGAACGTG
CYC5-F1G-2	CATCACCGCTATCATTGCCGSGTATRAGCGTAAATCTCCGTBGGGTCTGGATGTAGAACGTG
CYC4-REV12	GGCAATGATAGCGGTGATG
CYC4-F2G	CGGTAGCTACGAAACGCTGCRTAAAATCGCTAGCTCTGTTAGC
CYC4-REV9G	CAGCGTTTCGTAGCTACCG
CYC5-F3-1	GTGCTCGTTTCATCGAAATTKCTAGCCGCGATTCAGAAACTCTGGAAATC
CYC5-F3-2	GTGCTCGTTTCATCGAAATTKCTAGCGTTGATTCAGAAACTCTGGAAATC
CYC4-REV1234	AATTTCGATGAAACGAGCAC

Library screening. In order to guarantee an approximately 3-fold oversampling, 100 clones were screened per saturation mutagenesis library. Alternatively, 800 to 1,000 clones were screened per DNA shuffling library. In order to ensure adequate coverage, the library size was restricted to fewer than 1,000 possible combinations. Calcium competent BL21-Gold(DE3) cells [193] were transformed with the gene libraries and plated on LB agar plates (100 μg/mL ampicillin). Single colonies were used to inoculate 600 μL LB medium (100 μg/mL ampicillin) in a 96-well deepwell plate. After overnight incubation at 30 °C, 200 μl culture were stored in replica plates. The remainder of the culture was induced by addition of LB (with 1 mM IPTG, 150 μg/mL ampicillin) to a final volume of 1.6 ml. The temperature was reduced to 18 °C and protein production continued for 18 h. The induced cells were harvested by centrifugation (4,000 rpm, 10 min, 4 °C) and resuspended in 800 μl assay buffer (50 mM sodium phosphate, 100 mM NaCl, 2% DMSO, pH 6.0) containing 1 mg/ml lysozyme, a spatula tip of DNAseI, and 2 mM citral **22**. Plates were covered with an aluminum seal and incubated for 20 h at 30 °C with shaking at 1,000 rpm. 400 μl *n*-butanol was added and the plates were shaken for 10 min at 1,000 rpm to allow extraction. The phases were separated by centrifugation (10 min, 4,000 rpm, 4 °C). The organic phase was transferred into a fresh 96-well PCR plate. The plate was covered with a silicon seal and analyzed via liquid chromatography as described before.

6   Perspective – The coming of age of *de novo* enzyme design

In 2016, Baker published a review on the recent advances in the design of novel protein structures [132] in which he highlighted "the coming of age of *de novo* protein design". Numerous examples of *de novo* proteins that were successfully designed by computation without any sequence homology to natural scaffolds were presented. Just ten to twenty years ago, design of an atomically accurate protein like the *de novo* TIM barrel [134] was thought to be too challenging.

De novo enzyme engineering, just like protein design, has made huge improvements over the last decade. The development and refinement of computational methods were game changing in a field that was previously limited to semi-rational and random experimental approaches for exploring sequence space. Designing enzymatic activity, however, is an inherently more challenging task than the design of stable protein folds. Placement of catalytic residues with sub-Ångstrom precision [11,14-16], modeling of fine-tuned electrostatics [17,18], and consideration of delicate conformational motions [22-24] are still extremely difficult. The currently employed theozyme approach, which attempts to uniformly stabilize a static transition state complex, may be too simplistic for reliable implementation of such features [97,139]. Nonetheless, several novel enzymes have been successfully engineered [97,139] and future enzyme design endeavors may, like *de novo* protein design, benefit from recent improvements in computational methodologies [132]. Meanwhile, directed evolution has been applied with remarkable success to optimize a few computational designs. In some cases, catalysts have been generated that achieve enzyme-like rate accelerations [4,5].

Engineered enzymes will likely only be as good as the underlying models employed for their design. Broadening the knowledge on *de novo* engineered catalysts by in-depth biophysical characterization may help scrutinize the limitations and opportunities for computational enzyme design. In this context, the highly evolved Kemp eliminase HG3.17 [4,103] was subjected to detailed structural and mechanistic studies. Our results suggested that a delicate interplay between positioning and electrostatics were important for evolution of its high activity [144,145]. However, computational modeling of the active site indicated that the simple models underlying the theozyme approach were unable to explain catalysis in different Gln50

variants of HG3.17 or rationalize the activity of the most evolved catalyst. More complex models that explicitly include electrostatic effects, which are certainly important in natural enzymes [17,18], will be required for future progress.

Owing to the lack of appropriate tools to assess electrostatic effects in Kemp eliminases, vibrational probes were developed to investigate electrostatic catalysis in HG3.17 experimentally. Spectroscopic analysis of the probes bound to different variants of HG3.17 indicated a correlation between rate acceleration and the observed perturbation of the acid-base equilibrium of the inhibitor and the aspartate base. Computational analysis further suggested that this perturbation is caused by a change in the apparent electric field at the active site. Thus, electrostatic catalysis seems to provide a significant contribution to the 6×10^8-fold rate acceleration of HG3.17, even though the theozyme approach does not explicitly consider such effects (Figure 70).
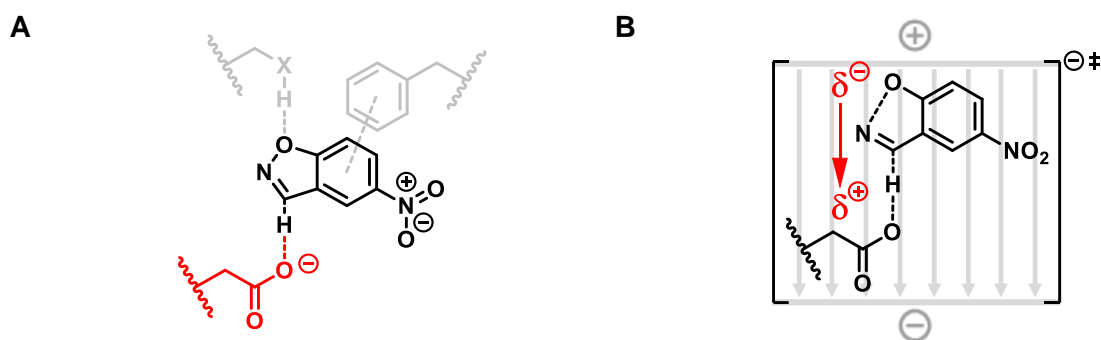
Figure 70: Expanding the theozyme approach. **(A)** The classical theozyme approach relies on stabilizing the transition state complex, composed of the ligand (black), primary catalytic residues (red) and additional interactions (gray) into a protein scaffold. **(B)** Electrostatic effects, such as of the overall electric field (gray) onto the transition states dipole moment (red) are not explicitly included during design. Given the importance of electrostatic catalysis, computational design will likely benefit from including dipole stabilization in addition to stabilization of to the overall complex.

Computational design typically relies on stabilization of the transition state as a whole, with each interaction to the ligand weighted the same during design. However, electrostatic stabilization of transition state dipole moment is likely a key driver for catalysis and could be an important consideration for future designs. In fact, computational algorithms have been developed by others to assess protein electrostatics on a practical timescale [65]. Application of such tools might improve the quality of first generation catalysts. Furthermore, computational studies on other Kemp eliminases have suggested that the electric field of the protein scaffold is often misaligned for catalysis, likely because it is not considered during design [127]. By aligning the transition state dipoles of novel active sites to the scaffold's

natural electric field at an early stage of design, catalysis might successfully harness electrostatic effects from the protein scaffold.

Improvements achieved during HG3.17's evolution resulted partially from remodeling of its conformational landscape. All HG3 variants exist in two conformations, an active and an inactive state. The pH dependencies of the conformational interconversion revealed that evolution selectively stabilized the active conformation of HG3.17 at ambient pH. Preliminary NMR studies from the Kern lab (Brandeis University) indicate that the conformational changes affect a cross section of the whole catalyst. Once the molecular details of the conformational change are dissected, directed evolution targeted to the affected regions may allow further increases in efficiency by preventing losses in activity at elevated pH or temperature.

Conformational motions may be far more important in designed enzymes than originally anticipated. Directed evolution often remodels the active sites of designer enzymes, allowing side chains and substrates to adopt new conformations or even populate other binding pockets [155]. Such conformational heterogeneity, observed in numerous static crystal structures, may be subject to dynamic interconversions that can likely be targeted by evolution to improve catalysis, as observed for HG3.17. Remodeling of protein motions and conformational landscapes may thus allow reengineering of the activities of many other systems.

Kemp eliminase HG3.17 is unusually efficient [119-121], and its catalysis is only matched by one other *de novo* enzyme that catalyzes retro-aldol cleavage [5]. The engineering of a new Kemp eliminase, 1A53-2, complements these previously described catalysts, demonstrating that enzyme-like rate accelerations can be repeatedly achieved through design and evolution. Optimization of evolutionary naïve computational designs to enzyme-like catalysts in the laboratory may well be comparable to the natural evolution of modern enzymes from primordial catalysts. It has been proposed that modern enzymes evolved from their primordial ancestors by gradually decreasing $\Delta H^\ddagger$ [268], with the ribosome as the only notable exception [31]. The laboratory evolution of 1A53-2 supports this hypothesis. The activation enthalpy in the most evolved catalyst was decreased relative to the computational design ($\Delta\Delta H^\ddagger$ = -10.9 kcal/mol), which was partially offset by a decrease in the activation entropy ($-T\Delta\Delta S^\ddagger$ = +5.4 kcal/mol, Figure 71). Decreasing of the activation energy in this way would have automatically enhanced the rate acceleration over the corresponding background reaction for primordial enzymes as the early earth was cooling down [268]. Laboratory evolution of 1A53-2 at different temperatures, which may be performed comparably rapidly with the developed

active site optimization approach, may shed light on the temperature-dependent selection pressure on $\Delta H^\ddagger$ and $\Delta S^\ddagger$. Divergent evolution of enzymes that either decrease $\Delta H^\ddagger$ or increase $\Delta S^\ddagger$ would provide valuable insights into the molecular requirements for optimization of these parameters.
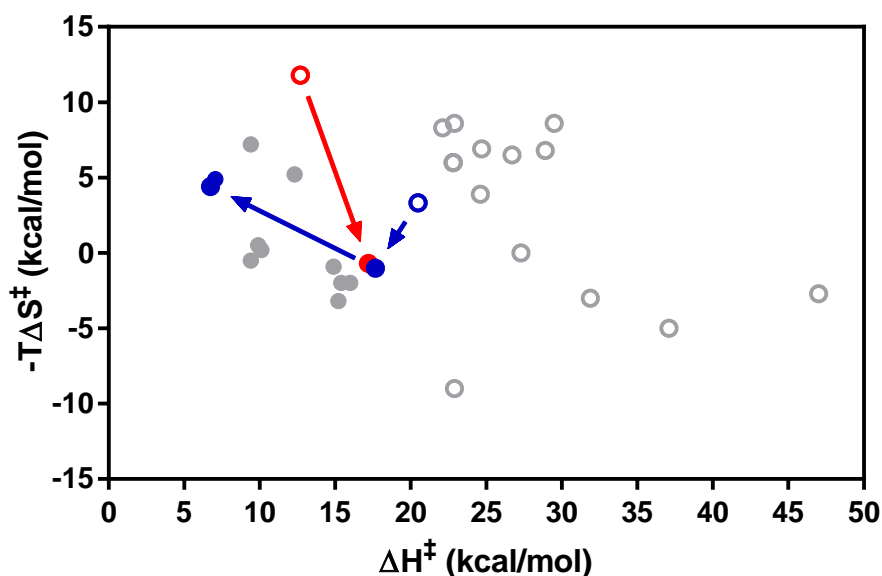
Figure 71: Activation enthalpies ($\Delta H^\ddagger$) and entropies ($-T\Delta S^\ddagger$) for enzymatic reactions. Various natural enzymes (gray, listed in [32]) are compared to variants of Kemp eliminase 1A53-2 (blue). Filled circles represent enzymatic reactions, while empty circles represent the respective background reactions. Remarkably, natural enzymes, as well as the computationally designed 1A53-2 and its evolutionary descendants achieve catalysis by lowering $\Delta H^\ddagger$. In case of 1A53-2.9, lowering $\Delta H^\ddagger$ occurs even at the expense of $\Delta S^\ddagger$. The only exception is the ribosome (red), which is a true entropy-trap and achieves catalysis only by lowering $-T\Delta S^\ddagger$. Adapted from [32].

Kinetic and structural characterization of 1A53-2 and its evolutionary descendants have highlighted several requirements for efficient catalysis. The importance of steric and electrostatic effects is particularly evident in the evolved variants. Structural comparison of 1A53-2.5 with the computational design was straightforward since the designed and evolved catalysts have similarly organized active sites, unlike many other evolved *de novo* enzymes [4,5]. The structural analysis strongly suggested that improved packing interactions, refined positioning, and fine-tuning of base reactivity through secondary interactions enhanced catalysis over the course of evolution. Stabilization of the deprotonated glutamate base through H-bonding interactions, as indicated by a decrease in the apparent pK_a (8.2±0.2 to 5.8±0.1), likely contributed significantly to the activity of 1A53-2.9. The identified effects seem to be common to many other systems where similar phenomena have been observed. Future design endeavors may thus benefit from improving packing interactions and positioning. Reliable implementation of more complex catalytic devices is also clearly needed. In this

regard, tuning the reactivity of the catalytic residues through secondary interactions will be essential. Realization of these may provide increased activity of designed catalysts in the future.

Efforts to expand the scope of *de novo* enzyme engineering can benefit substantially from development of novel ultra-high throughput screening systems [5,189,190]. Increased sampling capacity allows coverage of a larger fraction of the fitness landscape and facilitates the identification of beneficial mutations which become increasingly rare during evolution [189,190]. Since ultrahigh-throughput screening typically relies on fluorogenic assays, we established several assays with a fluorescent readout for the Kemp elimination. These assays rely either on pH changes, a supramolecular chemosensor, or a fluorogenic substrate. Initial methods development demonstrated the viability of the pH-based assay in ultrahigh-throughput μSCALE screening. The substrate-based approach is also likely to be viable in such assays if the absorbance of the benzisoxazole derivative can be shifted to a slightly higher excitation wavelength. Based on the assays reported in this thesis, novel methods to accelerate evolution of Kemp eliminases in an ultrahigh-throughput setting may become accessible in the future. These assays would likely support the evolution of more efficient Kemp eliminases, as shown for other systems [5]. They might also allow rapid extrapolation of divergent evolutionary trajectories, for example at different temperatures or pH values. It would further be possible to explore different starting designs. Valuable insights into catalysis and evolution in these artificial systems can be expected from such experiments [211].

In principle, designed enzymes can be employed for manifold synthetically relevant transformations using the inherent reactivities of the designed catalytic devices [256-259]. Kemp eliminase 1A53-2 relied on a carboxylic acid to promote proton abstraction from its benzisoxazole substrates. The general base, however, was placed in a hydrophobic environment during design, perturbing its pK_a. The activity of this group as a general acid at neutral pH for the cyclization of monoterpenes, such as citronellal **21** or citral **22**, opens up exciting new opportunities. These cyclized monoterpenes are of high industrial relevance and often required in optical purity for further derivatization [249,250]. The evolvability of the original design for the promiscuous activity was demonstrated. Although only modest improvements were achieved, further evolution may lead to an enzymatic process for the synthesis of (-)-menthol starting from achiral citral **22**. The primordial nature of naïve computational designs is thus a potentially attractive source for new reactivities. Exploitation

of existing catalytic devices also offers a valuable alternative to computational design for the future engineering of novel enzymes.

With the increasing precision of computational methods, *de novo* enzyme engineering may become less limited by the currently available tools and more by fundamental questions regarding the origins of efficient enzyme catalysis. With the coming of age of enzyme design, the key question will less likely be how a catalyst can be designed, but more what features need to be introduced to generate a proficient catalyst. Dissection of designed catalysts, as described in this thesis, should foster future design endeavors. More efficient computational and evolutionary methods, as well as diversification of existing catalysts, have enormous potential to enhance the overall engineering pipeline. Based on the insights gained and extrapolated from past achievements, *de novo* enzyme engineering is likely destined to provide ever more efficient catalysts in the future.

7 Appendix: Towards design and evolution of metal-dependent Diels-Alderases ‡

Metalloenzymes catalyze some of the most challenging transformations in Nature. Metal cofactors facilitate reactions such as water oxidation [269] and nitrogen fixation [270]. The inherent reactivity of metal ions often leads to high efficiency in these enzymes. In an attempt to harness such reactivities, many artificial metalloenzymes have been engineered to productively combine the selectivities and turnover numbers of protein catalysts with the catalytic performance of metal-ions [271]. Likewise, computational enzyme design may benefit from the incorporation of these reactive cofactors [272], adding a powerful and diverse source of high rate accelerations. Aiming for new biocatalysts with improved activities, novel metal binding sites have been introduced into scaffold proteins by exploiting covalent, supramolecular or dative interactions, as well as by substitution of metals at existing binding sites (Figure 72).
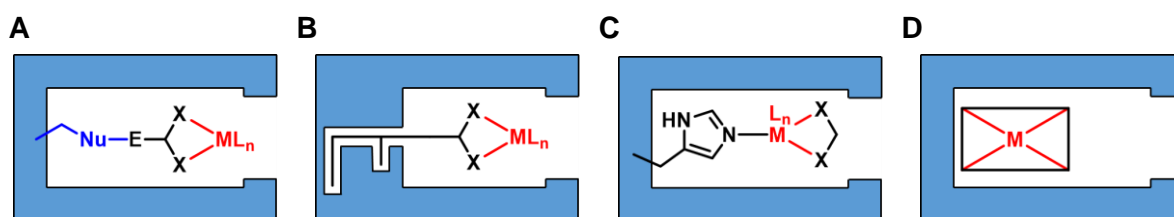
Figure 72: Anchoring strategies for the engineering of artificial metalloproteins. Binding of a metal (red) in an enzyme scaffold (blue) can be achieved via **(A)** covalent, **(B)** supramolecular, **(C)** dative anchoring, as well as by **(D)** metal substitution.

Covalent anchoring relies on any of the widespread bioconjugation techniques that have been developed in recent years [273]. For instance, decoration of a metal complex with an electrophilic moiety, such as a maleimide, allows direct nucleophilic coupling to a cysteine residue of the protein scaffold [274,275]. Supramolecular anchoring can be achieved by tight binding of specific ligands to natural enzymes. Attachment of the metal complex to a ligand such as biotin has been used to dock the catalysts to a suitable interaction partner such as streptavidin to give an artificial metalloenzyme. Many different reactions have been catalyzed

---

‡ **Scientific contributors**

Gonzalo Jimenez-Oses, and Kendall N. Houk performed the computational design. Roger C. Helgeson synthesized the substrates.

in this way [276,277]. Dative anchoring, reminiscent of most natural metalloenzymes, employs amino acid residues that are naturally able to coordinate metal ligands. Importantly, incorporation of non-natural amino acids with side chains such as bipyridinyl groups gained increasing importance for the dative anchoring of metal cofactors [239]. Finally, exchange of the metal cofactor at existent binding sites has led to several novel reactivities in enzymes such as carbonic anhydrase [278] and carboxypeptidase A [279].

A reaction which is particularly susceptible to metal-ion catalysis is the Diels-Alder reaction. Diels-Alder [4+2] cycloadditions of a diene and a dienophile are important in chemical synthesis and often yield chiral products upon C-C bond formation. Several Diels-Alderases have been rationally designed by introduction of artificial metal chelating residues in protein [280-282] or DNA [283] scaffolds (Figure 73). The resulting catalysts activate their dienophile, typically a pyridyl-substituted α,β-unsaturated ketone, through coordination of the carbonyl and the pyridine-nitrogen by the metal. Subsequent cycloaddition with dienes such as cyclopentadiene often proceeds with good yields and enantiomeric excesses of up to 97%. Furthermore, the reaction with the activated Michael systems has been diverted to a manifold of other transformations like nucleophilic additions [272] or cyclopropanations [271].
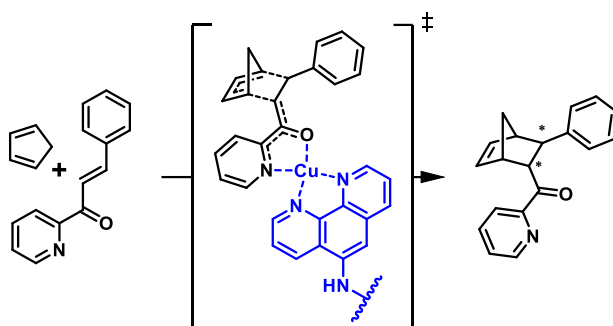
Figure 73: Rationally engineered metal-dependent Diels-Alderases. Diels-Alderase were engineered by introduction of phenanthroline-based artificial amino acids (blue) into a protein scaffold. Coordination of the dienophile to copper afforded a Diels-Alderases with an enantiomeric access of up to 97% [282].

The introduction of novel metal binding sites in protein scaffolds has led to notable enzymatic activities [271]. Computational design approaches, however, typically separate the challenge of introducing a new metal binding site and the design of a transition state stabilizing pocket. Likewise, one of the first computationally designed metalloenzymes, a phosphatase [220], utilized the natural zinc cofactor of its scaffold to hydrolyze a model organophosphate. Despite that early success, metal-ion catalysis is still challenging for computational design. Metal cofactors are dificult to accurately describe using the scoring functions of conventional design algorithms [284] and typically require complex quantum mechanical calculations to be

accurately modeled [285]. Nevertheless, given the proficiency of metal-ion catalysis, it is likely that computational design will benefit from utilizing metals as inherently reactive prosthetic groups to accelerate diverse reactions.

Here, the design and evolution of an artificial metal-dependent Diels-Alderase was investigated in collaboration with the Houk group (UCLA). The designs exploit the natural zinc-binding site of human carbonic anhydrase II (hCAII) to accelerate the Diels-Alder reaction the dienophile (*E*)-4-(3-oxo-3-(pyridin-2-yl)prop-1-en-1-yl)-benzoic acid **25** and diene 6-hydroxy-3-vinylindol **26**. As previously shown, the dienophile **25** can be activated for cycloaddition by coordination to the metal-cofactor (Figure 73). To address some of the problems with the standard design algorithm, extensive molecular dynamics simulations were used to vet the designs prior to experimental testing. However, none of the designs displayed any Diels-Alderase activity. As described in this chapter, experimental troubleshooting by reversion of designed mutations, testing of other reactions, and examination of different metal ions ultimately yielded an active Diels-Alderases. The evolvability of the Diels-Alderases was demonstrated in several rounds of directed evolution with a screening assay allowing metal-exchange via enzyme immobilization. The results provide a platform for further diversification by directed evolution and indicate the potential of reactive metal-ion cofactors for the engineering of novel enzymes.

7.1 Results

Computational design. Diels-Alderases were designed using the Rosetta software suite based on human carbonic anhydrase II (hCAII) by Dr. Gonzalo Jimenez-Oses and Prof. Kendall Houk from the University of California, Los Angeles (Figure 74). The designs aimed to accelerate the Diels-Alder reaction between dienophile **25** and diene 6-hydroxy-3-vinylindol **26** by coordination of the dienophile **25** to the zinc ion bound at the active site of the scaffold. The ideal geometry of the transition state in complex with the metal cofactor and three coordinating imidazole ligands was determined by density functional theory calculations. A binding site for the complex was introduced into carbonic anhydrase II using the Rosetta Enzdes algorithm. The designs were ranked and filtered by the Rosetta software suite. The most promising 400 designs were subjected to 40 ns molecular dynamics simulations with explicit water. Of these, 92 designs were identified after additional filtering for geometric

constraints. These designs are listed in the supplementary materials (see Supplementary Table 16). Initial tests of the background reaction by the Houk group revealed that 6-hydroxy-3-vinylindole **26** was unstable. All activity assays were thus performed with the diene 3-vinylindole **27** instead.
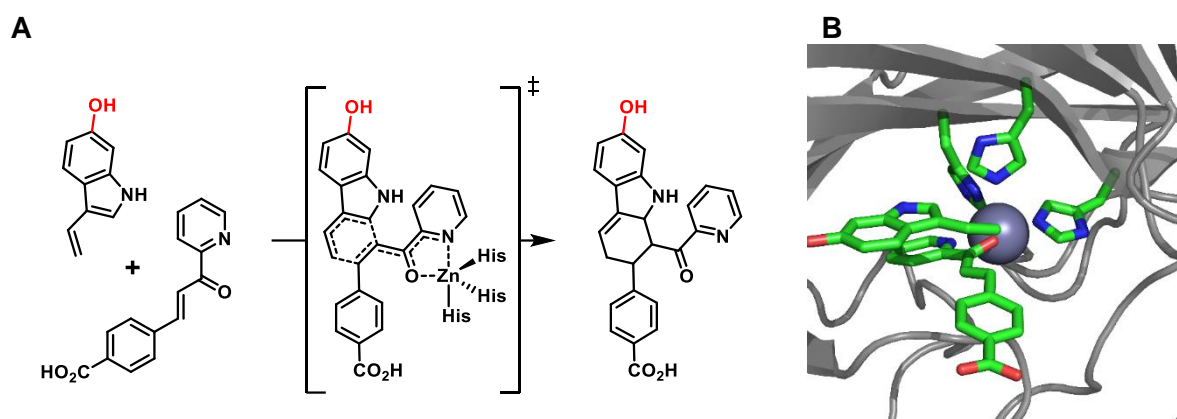
Figure 74: Computational Diels-Alderase design. **(A)** The designs targeted the [4+2]-cycloaddition of the depicted diene **26** and dienophile **25**. 6-Hydroxy-3-vinylindole **26** was experimentally found to be unstable and replaced by 3-vinylindole **27** for all experimental analyses. **(B)** Active site of the computational design 3-10 based on human carbonic anhydrase II. Coordination to the zinc (gray sphere) cofactor at the native (His)$_3$-binding site was employed to activate the dienophile **25**.

Buffer effects on the background Diels–Alder reaction. In order to detect low levels of Diels-Alderase activity, the assay conditions had to be refined. Activity was measured by monitoring the depletion of the dienophile **25** by the decrease in the absorbance maximum at 320 nm. All assays were performed at pH 8.0, because the solubility of the dienophile **25** was very low below pH 8.0. Even so, it had to be dissolved in DMSO to give linear absorbance upon addition to the reaction mix. If the dienophile **25** was dissolved in methanol or acetonitrile, the first 30 min of the time course were typically not linear. BSA also had to be added, to avoid undesired acceleration of the Diels-alder reaction by free metal ions. BSA naturally acts as a plasma carrier and interacts non-specifically with metal ions and other small molecules [286]. At a concentration of 1 mg/ml BSA, 5 µM of free Cu^{2+}, which displays high Diels-Alderase activity in solution, was entirely quenched and did not accelerate the reaction above background.

Testing the designs. Due to the similarity of the top 92 Diels-Alderase designs, only five were chosen for experimental characterization. These five designs were picked based on their net ranking and their distinct patterns of mutations (Table 13). In two of these designs, the metal coordination sphere was altered by replacing H94 with either aspartate or glutamate. The His-tagged designs were loaded with zinc *in vivo* during protein production. After

purification via a Ni-NTA column, the activity of 20 µM enzyme was tested with dienophile **25** and 3-vinylindole **27**. Neither the designs nor the wild-type accelerated the Diels-Alder reaction at enzyme concentrations up to 20 µM, as indicated by similar depletion of the dienophile **25** using UV/Vis-spectroscopy. In addition, none of the designs catalyzed the promiscuous esterase activity of the parent carbonic anhydrase scaffold with *p*-nitrophenyl-acetate which was also monitored by UV/Vis-spectroscopy. Furthermore, the previously reported strong interaction of wild-type hCAII with the competitive inhibitor acetazolamide was strongly diminished. While the wild-type showed a significant shift in melting temperature, the melting curves of the designs were unaffected by the inhibitor. Only two designs, 3-10 and 25-3, retained some affinity for the inhibitor, as indicated by a shift in melting temperature.

Table 13: Mutations of the experimentally tested Diels-Alderase designs [a]

Design	4	5	94	121	134	140	142	197	199	RANK
hCAII	H	W	H	V	V	L	V	L	T	
3-10	R	L	H	D	E	L	T	M	S	1
29-4	R	W	H	D	Q	L	T	M	S	17
40-10	H	W	H	D	E	H	T	M	S	22
10-5	R	L	E	D	E	H	T	M	S	40
25-3	R	L	D	I	E	L	V	M	T	63

[a] Mutated residues compared to hCAII are highlighted in blue

Activity rescue. The mutations introduced by design significantly remodeled the active site as indicated by the loss of both esterase activity and acetazolamide binding. While some changes in phenotype are expected, complete loss of activity may, however, indicate that some of the mutations impaired the integrity of the active site. Attempts to rescue the promiscuous esterase activity, as well as restore inhibitor binding, were performed by reversion of some of the mutations. Though unlikely, it was anticipated that restoration of the active site's integrity might also allow Diels-Alder activity. Individual as well as double reversion (DR) of mutations in the 3-10 and 25-3 designs were tested (Figure 35). Designs 3-10 and 25-3 were chosen since they displayed some residual binding of acetazolamide. In both cases, the V134E mutation, which was introduced to interact with the hydroxyl group of 6-hydroxy-3-vinyl-indole **26**, was reverted. In addition, the V121D mutation in 3-10, as well as the mutation of the zinc-coordinating histidine (H94D) in 25-3, were reverted.
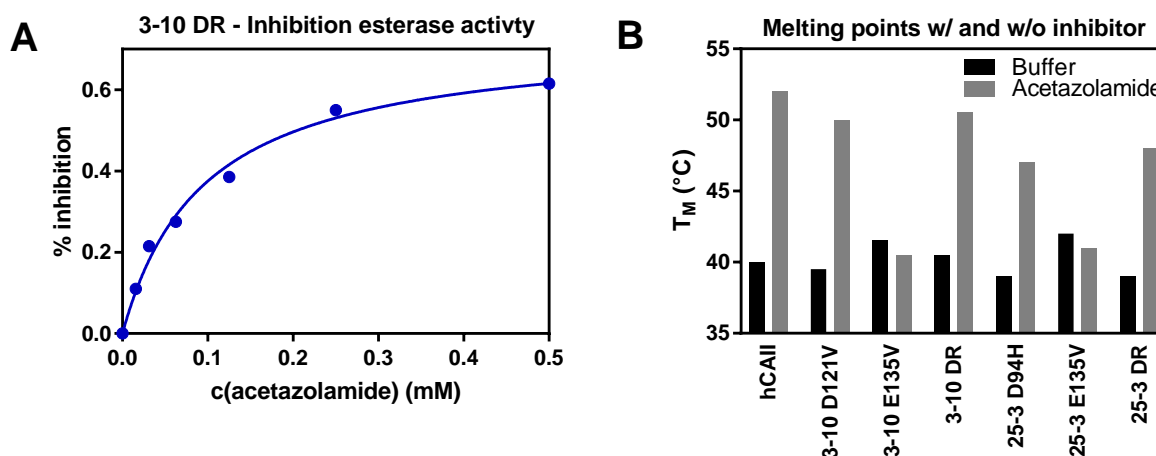
Figure 75: Activities of the Diels-Alderase designs. **(A)** Designs 3-10 DR and 25-3 DR displayed esterase activity, which could be inhibited with the carbonic anhydrase inhibitor acetazolamide. **(B)** Acetazolamide binding was further monitored by a shift in the melting temperature of some of the Diels-Alderase variants.

The esterase activity of the respective single and double reversions (DR) was tested with *p*-nitrophenyl-acetate. In addition, binding of the inhibitor acetazolamide was tested either by thermal-shift assay or inhibition of the esterase reaction. Reversion of the E134V mutation increased the protein yield without changing the phenotype. In contrast, reversion of the D121V mutation in 3-10, as well as D94H in 25-3, led to significantly stronger inhibitor binding as indicated by melting curve analysis. Furthermore, inhibition of esterase activity with acetazolamide was observed after these reversions. Binding of the metal-cofactor was verified by inductively coupled plasma mass spectrometry. Purified wild type hCAII and the E134V variants of 3-10 and 25-3 were analyzed. The original designs were not analyzed due to their reduced protein yield. It was found that the variants had 10% (3-10 E134V) and 36% (25-3 E134V) less zinc bound than the wild-type hCAII. Despite the presence of the zinc ion and restoration of the wild-type phenotype, none of the variants displayed any Diels-Alderase activity.

Owing to the absence of any Diels-Alderase activity, detection of dienophile **25** binding was probed independently. It was hypothesized that binding of the dienophile **25** to the metal leads to activation for other reactions such as Michael additions of nucleophiles such as ethyl cyanoacetate. Indeed, the double reversions 3-10 DR as well as 25-3 DR accelerated this Michael reaction, as indicated by accelerated depletion of the substrate absorbance. No activity was observed for the wild-type hCAII. Furthermore, the Michael reaction promoted by the design could be inhibited by addition of acetazolamide. Thus, the designs 3-10 DR and 25-3 DR likely accelerate the reaction by coordination of the dienophile **25** to the active site zinc (Figure 76).
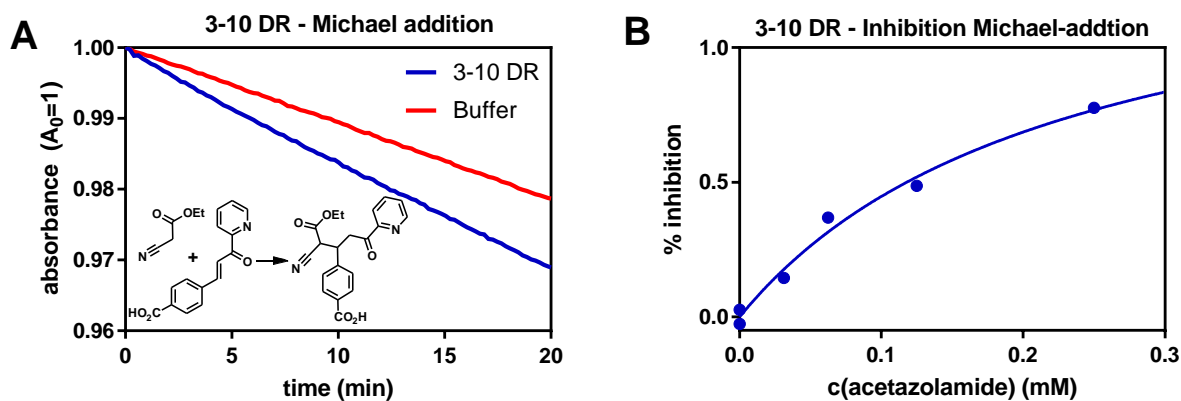
Figure 76: Binding of dienophile **25** to the Diels-Alderase designs. **(A)** Designs 3-10 DR and 25-3 DR displayed acceleration of a Michael reaction between the dienophile **25** and ethyl cyanoacetate **(B)** The Michael addition could be inhibited with acetazolamide, indicating that the designs accelerate the reaction by coordination of the zinc cofactor to dienophile **25**.

Metal exchange leads to Diels-Alderase activity. Zinc, which is the native cofactor of hCAII, reportedly possesses comparably low Diels-Alderase activity compared to copper in solution [280-282]. It was thus proposed that exchange of the metal-cofactor to copper could induce Diels-Alderase activity. The picomolar affinity of wild-type hCAII for zinc required removal of the metal cofactor by overnight incubation with 2,6-pyridinedicarboxylic acid [287]. After removal of the chelating agent through buffer exchange, copper nitrate was added. Removal of excess copper likely afforded designs with copper bound to the active-site. As predicted, the metal-exchanged 25-3 DR and 3-10 DR displayed modest activity for the Diels-Alder reaction with 8.5 µM enzyme (Figure 77). The highest Diels-Alderase activity was observed for the copper-loaded 25-3 DR which was dubbed DA0. Diels-Alderase activity, which was initially detected spectrophotometrically, was further confirmed by HPLC.
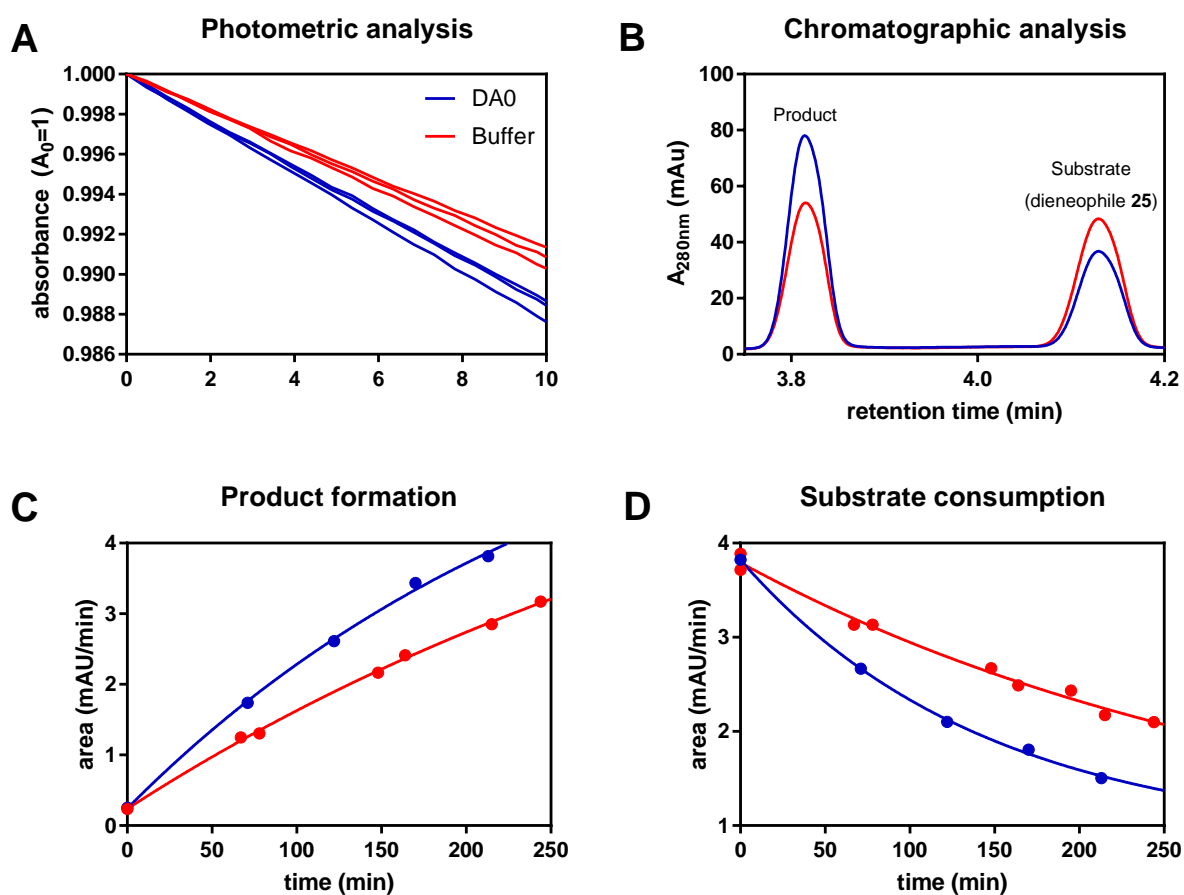
Figure 77: Diels-Alderase activity of DA0 (blue) and background reaction in buffer (red). **(A)** Absorbance traces observed for three different samples. For better comparison, the initial absorbance, which showed slight scattering, was set to one. **(B)** UPLC trace showing the substrate and product of the reaction between diene **27** and dienophile **25**. Increased product formation **(C)** and substrate consumption **(D)** were observed over time.

Directed evolution. After discovery of modest Diels-Alderase activity of the computational designs, the most active design (25-3 DR = DA0) was subjected to directed evolution. Since the copper ion could not be loaded *in vivo*, the metal had to be exchanged from zinc after protein production. Due to the challenges associated with metal-ion exchange in a microtiter plate-based assay, the first round of evolution was performed by screening of purified enzyme variants. Individual variants were constructed, produced, purified and assayed for Diels-Alderase activity. The first round of evolution targeted active site histidines His94, His107 and His119, which were each mutated to cysteine, aspartate or glutamate. DA0 H107E displayed the highest degree of activity, approximately 4-fold compared to DA0, and was dubbed DA1. Importantly, His107 is not one of the zinc-coordinating histidines, but part of an extended H-bonding network involved in positioning of His119 [288].

Development of a novel screening assay. In order to facilitate screening of larger libraries, a novel assay was developed, which involves immobilization of the Diels-Alderase

variants in microtiter plates (Figure 78). Immobilization of the proteins allowed exchange of the metal cofactor by incubation with the chelating agent 2,6-pyridinedicarboxylic acid and subsequent washing with copper.
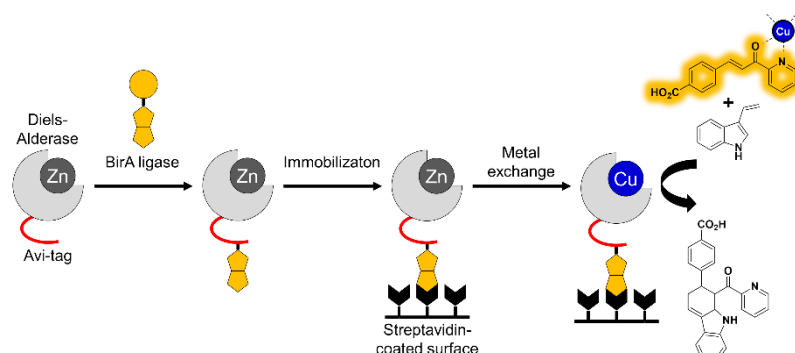
Figure 78: Metal-exchange screening assay. The zinc-containing Diels-Alderase variants (gray) are expressed with a biotinylation tag (Avi tag, red [289]). Biotin (yellow) is attached to the tag *in vivo* by the coexpressed BirA ligase. The biotinylated enzymes are immobilized on streptavidin-coated microtiter plates after cell lysis. Metal exchange is subsequently performed by incubation first with a chelating agent and then with copper (blue). Finally, Diels-Alderase activity can be monitored by following the depletion of the dienophile **25**.

Immobilization was achieved via a biotinylation tag (Avi tag) [289] that was attached to the N-terminus of the Diels-Alderase variants. The Avi tag was biotinylated *in vivo* by coexpression of a ligase and addition of biotin to the growth medium. After cell lysis, the lysate was loaded onto microtiter plates coated with streptavidin. The metal ion of the immobilized proteins was subsequently removed by overnight incubation with the chelating agent 2,6-pyridinedicarboxylic acid. Subsequently, the excess chelating agent was removed by washing with buffer, and copper was added. Removal of the excess copper by washing with buffer afforded active Diels-Alderase variants bound to the microtiter plates. According to the manufacturer, approximately 100 pmol biotin can be immobilized per well, resulting in a maximal final concentration of approximately 0.5 μM enzyme. The activity of DA1 could thus be monitored, although the acceleration over background was only (1.6±0.3) fold over the background reaction. Nonetheless, the microtiter plate screen could be employed for further evolution.

To further evolve DA1, saturation mutagenesis libraries individual targeting residues 65, 67, 70, and 121 were screened with the new assay. The low activity of the Diels-Alderase variants lead to high errors in the screen. Thus, several of the isolated hits had wild-type sequence and only nine out of 15 sequenced variants contained mutations. Two additional variants were false positives with deletions in their genes. The best six variants from the screen were recloned, expressed and purified. Only half of the hits yielded soluble protein and

just one variant, DA1 A65M, had improved activity compared to DA1. It was dubbed DA2. To avoid recloning in subsequent rounds of evolution, a vector was constructed which had both the biotinylation tag and the His tag at the N-terminus. In the third round of evolution, saturation mutagenesis libraries were prepared that targeted residues 198-202 individually using NNK libraries. The most active variant from that round of evolution was DA2 P210M (dubbed DA3).

25-3: hCAII H4R W5L H94D V121I V134E L197M
DA0: hCAII H4R W5L D94H V121I E134V L197M
DA1: hCAII H4R W5L D94H V121I H107E E134V L197M
DA2: hCAII H4R W5L A65M D94H V121I H107E E134V L197M
DA3: hCAII H4R W5L A65M D94H V121I H107E E134V L197M P201M

```
MSHRLGYGKH NGPEHWHKDF PIAKGERQSP VDIDTHTAKY DPSLKPLSVS YDQATSLRIL NNGHMFNVEF
DDSQDKAVLK GGPLDGTYRL IQFHFHWGSL DGQGSEETVD KKKYAAELHL IHWNTKYGDF GKAVQQPDGL
AVLGIFLKVG SAKPGLQKVV DVLDSIKTKG KSADFTNFDP RGLLPESLDY WTYPGSMTTP MLLECVTWIV
LKEPISVSSE QVLKFRKLNF NGEGEPEELM VDNWRPAQPL KNRQIKASFK LEHHHHHH
```

Figure 79: Diels-Alderase sequences. Mutations introduced during design (red) or the evolution (blue) are stated relative to hCAII. The mutations are highlighted in the sequence of DA3.

Characterization of DA0, DA1, DA2 and DA3. Single-point kinetic measurements with 1 mM diene **27** and 40 µM dienophile **25** were performed to estimate the catalytic efficiency of DA0, DA1, DA2 and DA3. The determined catalytic efficiencies are thus only a rough approximation of the true activity. For DA3, an apparent third order rate constant of approximately 2,000 $M^{-1}M^{-1}s^{-1}$ was determined (Figure 80), which corresponds to a 20-fold improvement over DA0. Its metal binding selectivity was also modified (Figure 80C). As indicated by melting curve analysis, the hCAII protein scaffold tightly binds zinc and shows no interaction with copper. The affinity for zinc is drastically decreased in DA3 but the affinity for copper seems to be slightly increased as compared to hCAII. Finally, the stereoselectivity of the reaction was analyzed via chiral liquid chromatography (Figure 80D). In agreement with the four different possible stereoisomeric Diels-Alderase products, four different peaks were observed by separation on a chiral column. Diels-Alderase DA1 shows some stereoselectivity as indicated by a slightly increased yield of the two products.
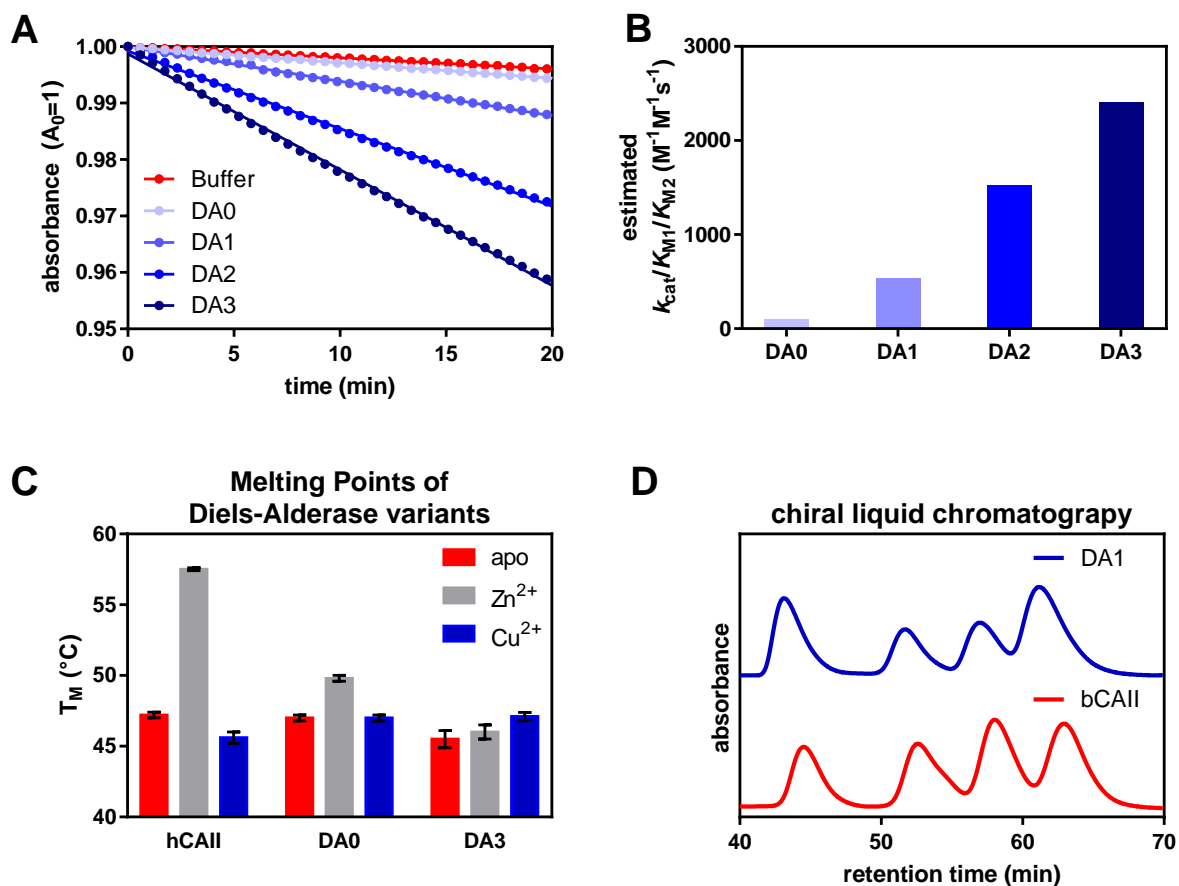
Figure 80: Evolution of Diels Alderase DA3. **(A)** Diels-Alderase variants (gray to blue) are gradually improved during evolution to higher activity compared to the background reaction (red). For better comparison, the initial absorbance, which showed slight scattering, is set to 1. **(B)** Estimated catalytic efficiencies derived from that single-point measurement. **(C)** Melting curves indicate that the affinity towards Zn^{2+} is decreased during evolution compared to the scaffold protein hCAII. Nonetheless, DA3 shows slightly increased stabilization when Cu^{2+} is bound as compared to hCAII, which is destabilized by Cu^{2+}. **(D)** Chiral liquid chromatography of the methylated reaction products from DA1 and the background reaction with the inactive bovine carbonic anhydrase II (inactive bovine isoform of hCAII). DA1 shows some selectivity for 2 of the 4 possible reactions products.

7.2 Discussion

De novo enzyme engineering will likely benefit from the increasing precision of computational algorithms [132], the development of more powerful screening methods (Chapter 4 and [189,190]), and growing knowledge about enzyme catalysis (Chapter 2.1-2.4 and [11,18,23]). In addition, the combination of inherently reactive groups, such as metal-cofactors, with a suitable protein scaffold may allow engineering of even better enzymes in the future. Thus, computational design and directed evolution of metal-dependent Diels-Alderases was pursued.

While it was assumed that these peaks correspond to the four different product enantiomers, recent unpublished X-ray data from S. Basler indicate that different products then expected are formed [290]. Specifically, the peaks observed in chiral liquid chromatography correspond to the endo and exo hetero Diels-Alder products resulting from cycloaddition of the unsaturated ketone as diene and the vinyl group of 3-vinylindole **27** dienophile. Further analysis will be required to verify the identity of the first two peaks. Nonetheless, observation of stereoselectivity is a strong indication that the observed rate acceleration stems from enzymatic catalysis and not from free metal in solution.

The activity of the designed Diels-Alderases critically depends on the reactive copper ion and is not observed if the less reactive zinc is bound. Copper thus seems to be absolutely essential for the activity of DA3, which may also be reflected in this variant's increased affinity for copper over zinc. Further problems with the designs likely arose due to the exchange of the unstable 6-hydroxy-3-vinylindole **26** with the more stable and less electron-rich 3-vinylindole **27**. Particularly the V134E mutation introduced to interact with the hydroxyl group impaired scaffold stability, but may even hamper catalysis with 3-vinylindole **27**.

The estimated apparent third order rate constant of DA3 ($2x10^3\,M^{-1}M^{-1}s^{-1}$) can be used to compare its activity with other Diels-Alderases. For example, this value is similar to that exhibited by the computationally designed Diels-Alderase CE20 ($5x10^2\,M^{-1}M^{-1}s^{-1}$), after intensive evolution. The uncatalyzed background reaction of CE20 [123] is, however, 4,000-fold slower than the uncatalyzed reaction of diene **27** and dienophile **25** [290]. The chemical proficiency, $(k_{cat}/K_{M,diene}/K_{M,dienophile})/k_{uncat}$, of CE20 ($9x10^7\,M^{-1}$) is thus orders of magnitude higher than that of DA3 ($9x10^4\,M^{-1}$). Noteworthy, another Diels-Alderase design by Gonzalo Jimenez-Oses and Kendall Houk has recently been evolved in the Hilvert group by Sabine Studer and Sophie Basler [290]. The design targeted the same reaction as DA3 but was based on a computationally designed zinc-binding protein called MID1 [291]. This Diels-Alderase design, which is 97 amino acids long, achieved an apparent third order rate constant of $6x10^8\,M^{-1}M^{-1}s^{-1}$ for the target Diels-Alderase reaction after nine rounds of directed evolution using its zinc ion as a Lewis acid. In contrast to the MID1-based design, in which 20% of the residues in the protein were exchanged during evolution, modification of DA3 has only just begun. The newly designed screening assay via immobilization of the Diels-Alderase variants provides a powerful tool for further evolution.

7.3 Materials and methods

Substrate synthesis. 3-Vinylindol **26** and (*E*)-4-(3-oxo-3-(pyridin-2-yl)prop-1-en-1-yl)-benzoic acid **25** were kindly supplied by Roger C. Helgeson who prepared the substrates following published procedures [292,293].

Protein production and purification. BL21-Gold(DE3) cells were transformed with the pET11 vector harboring the gene of interest and plated on LB agar plates (with 100 µg/ml ampicillin). A single colony was used to inoculate an overnight culture of LB medium (with 100 µg/ml ampicillin). 500 ml of LB medium (with 50 µM $ZnSO_4$ and 100 µg/ml ampicillin) were inoculated in a ratio of 1:100 from the overnight culture and grown to an OD of 0.5 at 37 °C. Protein production was induced with 1 mM IPTG, 500 µM $ZnSO_4$ were added, and cells were incubated overnight at 18 °C. The cells were harvested and the cell pellet was frozen at -80 °C or directly processed for protein purification. The cell pellet was resuspended in lysis buffer (20 ml 50 mM Tris-HCl, 0.5 M NaCl, 10 mM imidazole, 100 µL protease inhibitor cocktail (*Sigma*), 1 mg/ml lysozyme, DNAseI, pH 7.4). The cell suspension was sonicated for 1 h in a sonication bath at 4 °C (*Telesonic Ultrasonics*). After centrifugation (18.000 rpm, 20 min, 4 °C). The soluble fraction was applied to 3 ml Ni-NTA slurry (*Qiagen*), washed with 10 mM imidazole before elution with 300 mM imidazole, each in 50 mM tris-HCl, 0.5 M NaCl, pH 8. The buffer was exchanged to 100 mM of the chelating agent 2,6-pyridinedicarboxylic acid in 50 mM MOPS, 100 mM NaCl, pH 7.0 with a PD10 buffer exchange column (*GE Healthcare*) and the proteins were incubated for 20 h at 15 °C. A series of buffer exchanges were subsequently performed using PD10 columns. First, the buffer was exchanged to 50 mM MOPS, 100 mM NaCl, pH 7.0, then to the same buffer with 100 µM $Cu(NO_3)_2$ and the excess of copper was again removed by two additional buffer exchanges to 50 mM MOPS, 100 mM NaCl, pH 8.0. Proteins were analyzed with SDS-PAGE (Phastsystem, *Pharmacia Biotech*) using a low molecular weight marker (*GE Helthcare*). Protein concentrations were determined by measuring the absorbance at 280 nm and a calculated extinction coefficient [175]. An extinction coefficients of 31,400 $M^{-1}cm^{-1}$ was calculated for all variants discussed in this Chapter. Protein masses were checked by mass spectrometry. For mass determination, protein samples were exchanged into 0.1% acetic acid with Vivaspin 500 centrifugal filters (*Sartorius*) and measured by ESI-MS on a Daltonics maXis ESI-Q-TOF mass spectrometer (*Bruker*). Mass spectra were deconvoluted using the MaxEnt1 software. All masses corresponded to the expected sequence

lacking the N-terminal methionine. Zinc binding to hCAII, 3-10 E134V and 25-3 E134V was probed by inductively coupled plasma mass spectrometry by the Günther Group (ETHZ).

Activity assay.   Activity was assayed spectrophotometrically with a Lambda 35 UV/VIS spectrometer (*PerkinElmer*) by following substrate depletion at 320 nm in a cuvette with a 0.5 cm path length. The enzymes (20 µM for the assay in Figure 77, 8.5 µM for the assay in Figure 80) were added to 1 mM diene **27** and 40 µM dienophile **25** in 20 mM MOPS, 100 mM NaCl, 1 mg/ml BSA, 5% DMSO, pH 8.0 at 20 °C. Alternatively, activity was assayed with an Ultimate 3000 liquid chromatography system (*Dionex*). The reaction, under the same conditions as described earlier, was quenched by addition of *n*-butanol at a one-to-one ratio. The organic phase was separated by centrifugation (10,000 rpm, 10 min, 4 °C) and injected onto a C-18 X-Bridge column (*Waters*) with a linear 50% ACN/H_2O gradient with 0.1% TFA. The column was washed in 95% ACN prior to injection of the organic phase at 5% ACN.

Chiral high-performance liquid chromatography.   The Diels-Alder reactions were set up in 3 ml with 60 µM DA1 or bovine carbonic anhydrase as a negative control, 250 µM dienophile **25**, and 1 mM diene **27** in 20 mM MOPS, 100 mM NaCl, 1 mg/ml BSA, 5% DMSO, pH 8.0 at 20 °C. The reaction was run to completion as indicated by spectrophotometrical analysis. The reaction mixture was extracted three times with 5 ml diethyl ether, 0.1% TFA and the solvent was removed under reduced pressure. The remainder was dissolved in 1 ml methanol and cooled to 0 °C. For methylation of the Diels-Alder products, 10 µl trimethylsilyldiazomethane (2.0 M in diethyl ether) were added to the reaction and stirred for 30 min at 0 °C. The reaction was subsequently quenched with 30 µl acetic acid. The solvent was removed under reduced pressure and the residual was purified over a small silica column in a Pasteur pipette (ethyl acetate / hexane 4:1, R_f = 0.5). The solvent was removed under reduced pressure and the residual was dissolved in 300 µl 10% isopropanol in hexane. The supernatant was filtered through a Minisart SRP 4 filter (*Sartorius Stedim Biotech*) directly into an HPLC vial. 50 µl were injected into a normal-phase high performance liquid chromatography system (*Waters*) and separated on an OD-H Chiracel column (*Chiral Technologies*) with an isocratic gradient of 10% isopropanol in hexane (0.05 ml/min and 0.45 ml/min, respectively).

Library construction.   Libraries were constructed by overlap-extension PCR with Phusion polymerase (*Finnzymes*) [192]. Fragments were constructed with the primers in Table 14 and the AKZ3 (TAATACGACTCACTATAGGG) and T7term (GCTAGTTATTGCTCAGCGG) primer. For a typical PCR protocol, samples were incubated 5 min at 98 °C and DNA was amplified in 32

cycles between 98 °C (30 s), 55 °C (30 s) and 72 °C (30 s). After a final elongation for 5 min at 72 °C, the samples were cooled to 4 °C. The fragments were purified by agarose gel electrophoresis. Gels usually contained 2% agarose in 0.5 g/l sodium borate ($Na_2B_4O_7x10H_2O$) for small fragments. Gels were stained with ez-vision one (*Amresco*). Gel electrophoresis was performed at 200 V for 20 min at a variable electrical current. DNA extractions from agarose gels were carried out using the Zymoclean Gel DNA Recovery Kit (*Zymo Research*). Fragments were assembled by PCR using a similar protocol as before. 15 cycles were performed without the T7 and T7term primer to assemble the genes and 32 cycles were performed thereafter with the outer primer for amplification. The PCR products were again purified via agarose gel electrophoresis (1% agarose). The Genes were again amplified by PCR and purified with the DNA Clean & Concentrator kit (*Zymo research*). Genes were subsequently cloned into pET11 for protein production or pAvi or pAvi2 (Table 15) for screening. Plasmids were purified with the Zyppy Plasmid Miniprep Kit (*Zymo research*) from *E. coli* Xl1-Blue cells. For cloning, 2 µg of vector and 1 µg of insert were restricted with NdeI and XhoI (*New England Biolabs*). Vector and insert were purified via agarose gel electrophoresis (1% agarose). Fragments were ligated using T4 DNA ligase (*New England Biolabs*) and purified with the DNA Clean & Concentrator kit. Electro competent Xl1-Blue cells were transformed with the library [193], transferred to 10 ml LB and incubated for 1 h at 37 °C and 230 rpm. After 1 h, 0.1% of the culture were transferred onto an LB agar plate (100 µg/ml ampicillin) for estimation of the library size. 100 µg/ml ampicillin were added to the culture and the culture was incubated overnight at 37 °C and 230 rpm. The library was ready for screening after purification with the Zyppy Plasmid Miniprep Kit (*Zymo research*). All DNA concentrations were determined on a NanoDrop spectrophotometer (*ThermoFischer*). Gene sequences were confirmed by sequencing (*Microsynth*).

Table 14: Cloning Primers for the directed evolution of Diels-Alderase variants

Name	Sequence 5'-3'
Reversion of designed mutations	
hCAII-94H	CCTATCGTCTGATCCAGTTCCACTTTCATTGGGGTAGCTTAGATGG
hCAII-94r	GAACTGGATCAGACGATAGG
hCAII-121V	CGCCGCGGAATTACACTTAGTTCATTGGAACACTAAGTATGG
hCAII-121r	TAAGTGTAATTCCGCGGCG
hCAII-134V	GTATGGCGATTTTGGCAAAGCAGTGCAGCAGCCTGATGGCCTGGC
hCAII-134r	TGCTTTGCCAAAATCGCCATAC
hCAII-142V	CAGCAGCCTGATGGCCTGGCCGTTCTGGGCATCTTTTTAAAAGTGG
hCAII-142r	GGCCAGGCCATCAGGCTGCTG
Mutation of His to Cys/Asp/Glu	
HCAII-94C	CCTATCGTCTGATCCAGTTCTGCTTTCATTGGGGTAGCTTAGATGGACAGG
HCAII-94D	CCTATCGTCTGATCCAGTTCGATTTTCATTGGGGTAGCTTAGATGGACAGG
HCAII-94E	CCTATCGTCTGATCCAGTTCGAGTTTCATTGGGGTAGCTTAGATGGACAGG
HCAII-96C	CCTATCGTCTGATCCAGTTCCACTTTTGTTGGGGTAGCTTAGATGGACAGG
HCAII-96D	CCTATCGTCTGATCCAGTTCCACTTTGATTGGGGTAGCTTAGATGGACAGG
HCAII-96E	CCTATCGTCTGATCCAGTTCCACTTTGAATGGGGTAGCTTAGATGGACAGG
HCAII-94REV	GAACTGGATCAGACGATAGG
HCAII-107C	GCTTAGATGGACAGGGCTCGGAATGTACCGTTGACAAAAAGAAATACG
HCAII-107D	GCTTAGATGGACAGGGCTCGGAAGATACCGTTGACAAAAAGAAATACG
HCAII-107E	GCTTAGATGGACAGGGCTCGGAAGAAACCGTTGACAAAAAGAAATACG
HCAII-107REV	TTCCGAGCCCTGTCCATCTAAGC
HCAII-119C	GAAATACGCCGCGGAATTATGCTTAGATCATTGGAACACTAAGTG
HCAII-119D	GAAATACGCCGCGGAATTAGATTTAGATCATTGGAACACTAAGTG
HCAII-119E	GAAATACGCCGCGGAATTAGAATTAGATCATTGGAACACTAAGTG
HCAII-119REV	TAATTCCGCGGCGTATTTC
Avi Tag	
HCAII-AVITAG	TAAGCTCGAGAACCTGTATTTTCAGGGCCTGAATGACATTGAAGCGCAG
HCAII-AVITAG_REV	GCTGACTAGTCATTATTCATGCCACTCGATTTTCTGCGCTTCAATGTCATTC
Avi2 Tag	
HCAII-AVITAG2	TAAGCTCGAGCATCATCACCATCACCATGGCCTGAATGACATTGAAGCGCAG
HCAII-AVITAG2_REV	GCTGACTAGTCATTATTCATGCCACTCGATTTTCTGCGCTTCAATGTCATTC
Library 2	
HCAII-65NNK	CGTATCCTGAATAACGGACATNNKTTTAACGTCGAATTCGATGATTCGC
HCAII-67NNK	CGTATCCTGAATAACGGACATGCGTTTNNKGTCGAATTCGATGATTCGC
HCAII-65R	ATGTCCGTTATTCAGGATACG
HCAII-92NNK	GACGGTACCTATCGTCTGATCNNKTTCCACTTTCATTGGGGTAGC
HCAII-92R	GATCAGACGATAGGTACCGTC
HCAII-121NNK	CGCCGCGGAATTACACTTANNKCATTGGAACACTAAGTATGG
HCAII-121R	TAAGTGTAATTCCGCGGCG
HCAII-142NNK	CAGCAGCCTGATGGCCTGGCCNNKCTGGGCATCTTTTTAAAAGTGG
HCAII-142R	GGCCAGGCCATCAGGCTGCTG
Library 3	
HCAII-198NNK	CTGGACCTACCCAGGTAGTNNKACGACCCCGCCGCTGTTAGAATG
HCAII-199NNK	CTGGACCTACCCAGGTAGTATGNNKACCCCGCCGCTGTTAGAATG
HCAII-200NNK	CTGGACCTACCCAGGTAGTATGACGNNKCCGCCGCTGTTAGAATGCGTG
HCAII-201NNK	CTGGACCTACCCAGGTAGTATGACGACCNNKCCGCTGTTAGAATGCGTGAC
HCAII-202NNK	CTGGACCTACCCAGGTAGTATGACGACCCCGNNKCTGTTAGAATGCGTGACATG
HCAII-198R	CTGGACCTACCCAGGTAGT

Library screening. In order to guarantee an approximately 3-fold oversampling, 100 clones were screened per saturation mutagenesis library. Alternatively, 800 to 1,000 clones were screened per DNA shuffling library. In order to ensure adequate coverage, the library seize was restricted to fewer than 1,000 possible combinations. Calcium competent AVB100 *E. coli* K12 strain (*Avidity*) cells [193] were transformed with the gene libraries and plated on LB agar plates (100 µg/mL ampicillin). The AVB100 *E. coli* K12 strain has the BirA the biotin ligase gene stably integrated into its chromosome. Single colonies were used to inoculate 600 µL LB medium (100 µg/mL ampicillin, 50 µM ZnSO₄) in a 96-well deepwell plate. After overnight incubation at 30 °C, 200 µl culture were stored in replica plates. The remainder of the culture was induced by addition of LB (with 1 mM IPTG, 150 µg/mL ampicillin, 500 µM ZnSO₄, 50 µM biotin and 0.4% arabinose) to a final volume of 1.6 ml. The temperature was reduced to 18 °C and protein production continued for 18 h. All following steps were performed at 4 °C, if not otherwise stated. The induced cells were harvested by centrifugation (4,000 rpm, 10 min, 4 °C) and resuspended in 300 µl of 1 mg/ml lysozyme, 50 mM MOPS, 100 mM NaCl, pH 7.0. The wells of a streptavidin coated plate (*Pierce™ Streptavidin Coated High Capacity, Thermo Scientific*) were washed 3-times with 50 mM MOPS, 100 mM NaCl, pH 7.0 and 200 µl of the cell suspension per well were added. After incubation for 1 h at 4 °C, the wells were washed 3-times with 50 mM MOPS, 100 mM NaCl, pH 7.0. The buffer in each well was exchanged to 100 mM of the chelating agent 2,6-pyridinedicarboxylic acid in 20 mM MOPS, 100 mM NaCl, pH 7.0. The plate was incubated overnight at 15 °C. The wells were washed three times with 20 mM MOPS, 100 mM NaCl, pH 7.0. After incubation for 5 min in 100 µM $Cu(NO_3)_2$ in 20 mM MOPS, 100 mM NaCl, pH 7.0, the wells were washed again 3-times with 20 mM MOPS, 20 mM NaCl, pH 8.0. Activity was assayed after addition of 200 µl diene **27** (1 mM) and dienophile **25** (40 µM) in 20 mM MOPS, 20 mM NaCl, 5% DMSO, 1mg/ml BSA, pH 8.0 at 340 nm and room temperature using a SpectraMax platereader (*Molecular Devices*). The clones with the largest increase in activity relative to the preceding round were picked from the replica plate for plasmid isolation, sequencing, and further diversification.

Construction of Diels-Alderase vectors. Diels-Alderase genes were cloned into variants of the pKTCTET-0 vector [294]. The 3' end of the multiple-cloning site, which encodes a His₆ tag and three stop codons between the restriction sequences for XhoI and SpeI in pKTCTET-0 was modified in the pAvi and pAvi2 vectors. Modification of the vector was performed with the primer listed in the Table 14 (Avi tag and Avi2 tag, respectively) analogous to the protocol for library cloning. The insert was prepared by amplifying the corresponding forward and reverse

primers without insert via PCR, and cloning was performed via the XhoI and SpeI restriction sites. pAvi has a TEV cleavage site, an Avi tag and two stop codons, and pAvi2 carries a His$_6$ tag, the Avi tag and two stop codons between XhoI and SpeI (Table 15).

Table 15: Multiple cloning sites of Diels-Alderase vectors [a]
pKTCTET-0:
5'- NdeI – Protein of interest (POI) – XhoI – His$_6$ tag – Stop$_3$ – SpeI – 3'   5'- CAT ATG – POI – CTC GAG CAC CAC CAC CAC CAC CAC TAA TAA TGA CTA GTC –3'
pAvi:
5'- NdeI – Protein of interest (POI) – XhoI – TEV site – Avi tag – Stop$_2$ – SpeI – 3'   5'- CAT ATG – POI – CTC GAG AAC CTG TAT TTT CAG GGC CTG AAT GAC ATT GAA GCG CAG AAA ATC GAG TGG CAT GAA TAA TGA CTA GTC –3'
pAvi2:
5'- NdeI – Protein of interest (POI) – XhoI – His$_6$ tag – Avi tag – Stop$_2$ – SpeI – 3'   5'- CAT ATG – POI – CTC GAG CAT CAT CAC CAT CAC CAT GGC CTG AAT GAC ATT GAA GCG CAG AAA ATC GAG TGG CAT GAA TAA TGA CTA GTC –3'

[a] Multiple cloning sites of Diels-Alderase vectors pKTCT-0, pAvi and pAvi2. The sequences of the multiple cloning sites are colored as follows: Restriction sites (cyan), protein of interest (POI, gray), His$_6$ tag (pink), Avi tag (yellow) and stop-codons (red).

7.4 Supplementary Material

Supplementary Table 16: Top 92 Diels-Alderase Designs*											
Design	**4**	**5**	**92**	**94**	**121**	**130**	**134**	**140**	**142**	**197**	**199**	**RANK**
hCAII	H	W	Q	H	V	F	V	L	V	L	T	
round4b__3-10	R	L	Q	H	D	F	E	L	T	M	S	1
round4d__33-12	H	W	Q	H	D	F	E	L	T	I	S	2
round4b__18-17	R	A	Q	H	D	F	D	H	T	M	S	3
round4d__15-4	H	W	Q	H	T	F	S	L	T	I	T	4
round4a__22-8	R	W	Q	H	D	F	S	L	T	M	S	5
round4d__31-9	H	W	Q	H	D	F	D	H	T	I	S	6
round4a__35-14	R	W	Q	H	T	F	E	L	T	M	S	7
round4c__43-13	R	W	A	H	D	S	S	L	T	I	S	8
round4d__20-14	H	W	Q	H	D	F	E	L	T	I	T	9
round444b__30-7	R	L	Q	D	I	F	E	L	V	M	S	10
round4c__27-19	R	W	M	H	D	S	S	L	T	A	S	11
round44c__32-12	R	W	M	D	D	S	S	H	V	M	T	12
round4d__14-18	H	W	Q	H	D	F	E	V	T	M	S	13
round44a__22-10	R	W	Q	D	D	F	E	L	T	M	T	14
round4b__14-7	R	A	Q	H	D	F	E	H	T	M	S	15
round44a__36-4	R	W	Q	D	D	F	N	H	V	M	T	16
round4a__29-4	R	W	Q	H	D	F	E	H	T	M	S	17
round44a__11-7	R	W	Q	E	D	F	E	L	T	M	T	18
round4a__25-10	R	W	Q	H	D	F	S	H	T	M	S	19
round4d__50-4	H	W	Q	H	T	F	E	L	T	I	T	20
round4b__2-6	R	A	Q	H	D	F	E	L	V	M	S	21
round4d__40-10	H	W	Q	H	D	F	Q	L	T	M	S	22
round44c__15-5	R	W	M	D	D	S	E	L	T	I	S	23
round4c__14-7	R	W	M	H	D	S	Q	L	T	M	S	24
round44d__34-13	H	W	Q	E	D	F	E	L	T	M	S	25
round4d__24-15	H	W	Q	H	T	F	E	V	T	I	T	26
round4b__13-11	R	A	Q	H	D	F	E	L	T	M	S	27
round4c__32-18	R	W	M	H	D	S	E	L	T	A	S	28
round44a__50-20	R	W	Q	D	D	F	S	H	V	M	T	29
round4d__47-8	H	W	Q	H	D	F	E	Y	T	M	S	30
round4c__35-16	R	W	M	H	D	S	E	H	T	A	S	31
round4c__44-2	R	W	M	H	D	S	Q	H	T	M	S	32
round4a__16-2	R	W	Q	H	D	F	E	M	T	M	S	33
round4c__11-8	R	W	M	H	D	S	E	H	T	M	S	34
round44d__18-17	H	W	Q	D	D	F	E	L	V	M	S	35
round44b__42-20	R	A	Q	D	T	F	E	H	V	M	T	36
round4c__36-9	R	W	M	H	D	S	Q	L	T	A	S	37
round44d__30-4	H	W	Q	D	D	F	S	H	V	M	T	38
round4c__48-12	R	W	V	H	D	S	E	L	T	I	S	39
round44b__10-5	R	L	Q	E	D	F	E	H	T	M	S	40
round4c__5-11	R	W	M	H	T	S	E	L	V	I	S	41
round44b__19-8	R	L	Q	D	D	F	E	H	T	M	S	42
round44c__12-17	R	W	M	D	D	S	E	L	T	M	T	43
round4a__22-16	R	W	Q	H	D	F	E	Y	T	M	S	44
round44c__47-2	R	W	M	D	D	S	S	L	T	A	S	45
round4c__42-14	R	W	M	H	T	S	S	L	V	I	S	46

| Supplementary Table 16, continued: Top 92 Diels-Alderase Designs* | | | | | | | | | | | |
Design	4	5	92	94	121	130	134	140	142	197	199	RANK
hCAII	H	W	Q	H	V	F	V	L	V	L	T	
round4c__15-12	R	W	A	H	D	S	E	L	T	I	S	47
round44c__12-3	R	W	M	D	T	S	E	L	T	I	S	48
round444c__11-15	R	W	V	D	V	S	E	L	V	I	S	49
round44b__11-19	R	A	Q	D	T	F	E	H	T	M	S	50
round44c__20-10	R	W	M	D	D	S	E	H	V	M	S	51
round44a__41-4	R	W	Q	D	D	F	E	L	V	M	T	52
round44d__39-20	H	W	Q	D	T	F	E	L	V	M	T	53
round444c__46-17	R	W	M	D	V	S	E	H	V	M	S	54
round444c__34-10	R	W	V	D	V	S	S	L	V	I	S	55
round444c__22-15	R	W	M	D	I	S	Q	L	V	M	S	56
round44c__28-16	R	W	M	D	D	S	Q	L	V	M	T	57
round44c__37-14	R	W	M	D	T	S	E	L	T	M	T	58
round444a__25-14	R	W	Q	D	V	F	E	L	V	M	S	59
round44a__48-17	R	W	Q	D	T	F	E	L	T	M	S	60
round44d__10-8	H	W	Q	D	T	F	E	L	T	I	T	61
round44a__24-19	R	W	Q	D	T	F	E	H	T	M	T	62
round444b__25-3	R	L	Q	D	I	F	E	L	V	M	T	63
round444c__19-17	R	W	M	D	I	S	E	H	V	M	T	64
round444c__5-18	R	W	M	D	I	S	E	H	V	M	S	65
round44a__16-20	R	W	Q	D	T	F	E	L	V	M	S	66
round444c__43-11	R	W	M	D	V	S	Q	L	V	M	S	67
round4c__1-20	R	W	M	H	D	S	D	H	T	M	S	68
round444c__4-17	R	W	M	D	V	S	S	L	V	A	S	69
round4c__1-15	R	W	M	H	D	S	E	L	V	M	S	70
round44b__9-18	R	L	Q	D	T	F	E	H	T	M	T	71
round444d__8-16	H	W	Q	D	V	F	E	L	V	M	T	72
round444c__40-3	R	W	M	D	I	S	E	L	V	M	S	73
round44b__11-13	R	A	Q	D	D	F	E	H	T	M	S	74
round4c__50-2	R	W	V	H	D	S	S	L	T	I	S	75
round44c__49-1	R	W	V	D	T	S	E	L	T	I	T	76
round44c__23-2	R	W	M	D	T	S	E	L	T	A	S	77
round444c__41-19	R	W	V	D	V	S	E	L	V	M	T	78
round444c__34-3	R	W	M	D	I	S	S	L	V	M	S	79
round44c__17-8	R	W	M	D	V	S	E	L	V	M	S	80
round44c__38-12	R	W	M	D	D	S	E	L	V	A	S	81
round444c__29-9	R	W	V	D	V	S	S	H	V	I	S	82
round4a__49-11	R	W	Q	H	D	F	Q	H	T	M	S	83
round44a__34-14	R	W	Q	D	T	F	E	L	T	I	T	84
round4b__1-9	R	L	Q	H	D	F	E	H	T	M	S	85
round44a__2-17	R	W	Q	D	D	F	N	H	T	M	T	86
round44d__9-3	H	W	Q	D	T	F	E	H	T	M	T	87
round44b__4-9	R	A	Q	D	D	F	E	H	V	M	T	88
round44c__24-12	R	W	M	D	D	S	E	H	V	A	S	89
round444b__24-19	R	L	Q	D	I	F	E	H	V	M	T	90
round444c__37-17	R	W	M	D	V	S	E	L	V	I	T	91
round4b__32-12	R	A	Q	H	D	F	S	H	T	M	S	92

*For simplicity, design rounds (for instance: round4a__) are omitted in the main text. Tested designs are highlighted in light blue.

Bibliography

1. Wolfenden R, Snider MJ: **The depth of chemical time and the power of enzymes as catalysts.** *Acc Chem Res* 2001, **34**:938-945.

2. Sicard D, Legras J-L: **Bread, beer and wine: Yeast domestication in the** *Saccharomyces sensu stricto* **complex.** *C R Biol* 2011, **334**:229-236.

3. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K: **Engineering the third wave of biocatalysis.** *Nature* 2012, **485**:185-194.

4. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grutter MG, Privett HK, Mayo SL, Hilvert D: **Precision is essential for efficient catalysis in an evolved Kemp eliminase.** *Nature* 2013, **503**:418-421.

5. Obexer R, Godina A, Garrabou X, Mittl PR, Baker D, Griffiths AD, Hilvert D: **Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase.** *Nat Chem* 2017, **9**:50-56.

6. Fischer E: **Einfluss der Configuration auf die Wirkung der Enzyme.** *Ber Dtsch Chem Ges* 1894, **27**:2985-2993.

7. Michaelis L, Menten ML: **Die Kinetik der Invertinwirkung.** *Biochem* 1913, **Z**:333-369.

8. Alberty RA, Hammes GG: **Application of the theory of diffusion-controlled reactions to enzyme kinetics.** *J Phys Chem* 1958, **62**:154-159.

9. Pauling L: **Molecular architecture and biological reactions.** *Chem Eng News* 1946, **24**:1375-1377.

10. Pauling L: **Chemical achievement and hope for the future.** *Am Sci* 1948, **36**:51-58.

11. Amyes TL, Richard JP: **Specificity in transition state binding: The Pauling model revisited.** *Biochemistry* 2013, **52**:10.1021/bi301491r.

12. Hilvert D: **Critical analysis of antibody catalysis.** *Annu Rev Biochem* 2000, **69**:751-793.

13. Kurz JL: **Transition state characterization for catalyzed reactions.** *J Am Chem Soc* 1963, **85**:987-991.

14. Fersht A: **Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding.** Edited by. New York: W.H. Freeman; 1999.

15. Robertus JD, Kraut J, Alden RA, Birktoft JJ: **Subtilisin: A stereochemical mechanism involving transition-state stabilization.** *Biochemistry* 1972, **11**:4293-4303.

16. Williams DH: **Enzyme catalysis from improved packing in their transition-state structures.** *Curr Opin Chem Biol* 2010, **14**:666-670.

17. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MH: **Electrostatic basis for enzyme catalysis.** *Chem Rev* 2006, **106**:3210-3235.

18. Fried SD, Boxer SG: **Electric fields and enzyme catalysis.** *Annu Rev Biochem* 2017, **86**:387-415.

19. Hu Y, Houk KN, Kikuchi K, Hotta K, Hilvert D: **Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles.** *J Am Chem Soc* 2004, **126**:8197-8205.

20. Hollfelder F, Kirby AJ, Tawfik DS: **On the magnitude and specificity of medium effects in enzyme-like catalysts for proton transfer.** *J Org Chem* 2001, **66**:5866-5874.

21. O'Leary MH, Piazza GJ: **Medium effects in enzyme-catalyzed decarboxylations.** *Biochemistry* 1981, **20**:2743-2748.

22. Kohen A: **Role of dynamics in enzyme catalysis: substantial versus semantic controversies.** *Acc Chem Res* 2015, **48**:466-473.

23. Hanoian P, Liu CT, Hammes-Schiffer S, Benkovic S: **Perspectives on electrostatics and conformational motions in enzyme catalysis.** *Acc Chem Res* 2015, **48**:482-489.

24. Hammes-Schiffer S: **Catalytic efficiency of enzymes: A theoretical analysis.** *Biochemistry* 2013, **52**:2012-2020.

25. Kraut DA, Carroll KS, Herschlag D: **Challenges in enzyme mechanism and energetics.** *Annu Rev Biochem* 2003, **72**:517-571.

26. Pollack RM: **Enzymatic mechanisms for catalysis of enolization: Ketosteroid isomerase.** *Bioorg Chem* 2004, **32**:341-353.

27. Lightstone FC, Bruice TC: **Enthalpy and entropy in ring closure reactions.** *Bioorg Chem* 1998, **26**:193-199.

28. Åqvist J, Kazemi M, Isaksen GV, Brandsdal BO: **Entropy and enzyme catalysis.** *Acc Chem Res* 2017, **50**:199-207.

29. Belsare S, Pattni V, Heyden M, Head-Gordon T: **Solvent entropy contributions to catalytic activity in designed and optimized Kemp eliminases.** *J Phys Chem B* 2017.

30. Bhowmick A, Sharma SC, Honma H, Head-Gordon T: **The role of side chain entropy and mutual information for improving the de novo design of Kemp eliminases KE07 and KE70.** *Phys Chem Chem Phys* 2016, **18**:19386-19396.

31. Sievers A, Beringer M, Rodnina MV, Wolfenden R: **The ribosome as an entropy trap.** *Proc Natl Acad Sci USA* 2004, **101**:7897-7901.

32. Wolfenden R, Snider M, Ridgway C, Miller B: **The temperature dependence of enzyme rate enhancements.** *J Am Chem Soc* 1999, **121**:7419-7420.

33. Herschlag D, Natarajan A: **Fundamental challenges in mechanistic enzymology: progress toward understanding the rate enhancements of enzymes.** *Biochemistry* 2013, **52**:2050-2067.

34. Schwans JP, Sunden F, Lassila JK, Gonzalez A, Tsai Y, Herschlag D: **Use of anion-aromatic interactions to position the general base in the ketosteroid isomerase active site.** *Proc Natl Acad Sci USA* 2013, **110**:11308-11313.

35. Schwans JP, Hanoian P, Lengerich BJ, Sunden F, Gonzalez A, Tsai Y, Hammes-Schiffer S, Herschlag D: **Experimental and computational mutagenesis to investigate the positioning of a general base within an enzyme active site.** *Biochemistry* 2014, **53**:2541-2555.

36. Lamba V, Yabukarski F, Pinney M, Herschlag D: **Evaluation of the catalytic contribution from a positioned general base in ketosteroid isomerase.** *J Am Chem Soc* 2016, **138**:9902-9909.

37. Schwans JP, Sunden F, Gonzalez A, Tsai Y, Herschlag D: Correction to "Evaluating the catalytic contribution from the oxyanion hole in ketosteroid isomerase". *J Am Chem Soc* 2016, **138**:7801-7802.

38. Schwans JP, Sunden F, Gonzalez A, Tsai Y, Herschlag D: Evaluating the catalytic contribution from the oxyanion hole in ketosteroid isomerase. *J Am Chem Soc* 2011, **133**:20052-20055.

39. Schwans JP, Kraut DA, Herschlag D: Determining the catalytic role of remote substrate binding interactions in ketosteroid isomerase. *Proc Natl Acad Sci USA* 2009, **106**:14271-14275.

40. Frey PA, Whitt SA, Tobin JB: A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* 1994, **264**:1927-1930.

41. Cleland WW, Frey PA, Gerlt JA: The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 1998, **273**:25529-25532.

42. Gerlt JA, Gassman PG: Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters. *Biochemistry* 1993, **32**:11943-11952.

43. Jang DS, Choi G, Cha HJ, Shin S, Hong BH, Lee HJ, Lee HC, Choi KY: Contribution of a low-barrier hydrogen bond to catalysis is not significant in ketosteroid isomerase. *Mol Cells* 2015, **38**:409-415.

44. Natarajan A, Schwans JP, Herschlag D: Using unnatural amino acids to probe the energetics of oxyanion hole hydrogen bonds in the ketosteroid isomerase active site. *J Am Chem Soc* 2014, **136**:7643-7654.

45. Kirby AJ: Efficiency of proton transfer catalysis in models and enzymes. *Acc Chem Res* 1997, **30**:290-296.

46. Marcus RA: On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer. I. *J Chem Phys* 1956, **24**:966-978.

47. Fuxreiter M, Mones L: The role of reorganization energy in rational enzyme design. *Curr Opin Chem Biol* 2014, **21**:34-41.

48. Feierberg I, Åqvist J: The catalytic power of ketosteroid isomerase investigated by computer simulation. *Biochemistry* 2002, **41**:15728-15735.

49. Szefczyk B, Claeyssens F, Mulholland AJ, Sokalski WA: Quantum chemical analysis of reaction paths in chorismate mutase: Conformational effects and electrostatic stabilization. *Int J Quantum Chem* 2007, **107**:2274-2285.

50. Fuxreiter M, Warshel A: Origin of the catalytic power of acetylcholinesterase: Computer simulation studies. *J Am Chem Soc* 1998, **120**:183-194.

51. Warshel A, Sharma PK, Chu ZT, Aqvist J: Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* 2007, **46**:1466-1476.

52. Kamerlin SC, Sharma PK, Chu ZT, Warshel A: Ketosteroid isomerase provides further support for the idea that enzymes work by electrostatic preorganization. *Proc Natl Acad Sci USA* 2010, **107**:4075-4080.

53. Sigala PA, Kraut DA, Caaveiro JMM, Pybus B, Ruben EA, Ringe D, Petsko GA, Herschlag D: **Testing geometrical discrimination within an enzyme active site: Constrained hydrogen bonding in the ketosteroid isomerase oxyanion hole.** *J Am Chem Soc* 2008, **130**:13696-13708.

54. Kraut DA, Sigala PA, Fenn TD, Herschlag D: **Dissecting the paradoxical effects of hydrogen bond mutations in the ketosteroid isomerase oxyanion hole.** *Proc Natl Acad Sci USA* 2010, **107**:1960-1965.

55. Schneider SH, Boxer SG: **Vibrational stark effects of carbonyl probes applied to reinterpret IR and Raman data for enzyme inhibitors in terms of electric fields at the active site.** *J Phys Chem B* 2016, **120**:9672-9684.

56. Natarajan A, Yabukarski F, Lamba V, Schwans JP, Sunden F, Herschlag D: **BIOPHYSICS. Comment on "Extreme electric fields power catalysis in the active site of ketosteroid isomerase".** *Science* 2015, **349**:936.

57. Chen D, Savidge T: **BIOPHYSICS. Comment on "Extreme electric fields power catalysis in the active site of ketosteroid isomerase".** *Science* 2015, **349**:936.

58. Beker W, van der Kamp MW, Mulholland AJ, Sokalski WA: **Rapid estimation of catalytic efficiency by cumulative atomic multipole moments: Application to ketosteroid isomerase mutants.** *J Chem Theory Comput* 2017, **13**:945-955.

59. Wang L, Fried SD, Markland TE: **Proton network flexibility enables robustness and large electric fields in the ketosteroid isomerase active site.** *J Phys Chem B* 2017, **121**:9807-9815.

60. Wu Y, Boxer SG: **A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase.** *J Am Chem Soc* 2016, **138**:11890-11895.

61. Reichardt C: **Solvents and solvent effects in organic chemistry, 3rd Ed.** *Wiley-VCH, Weinheim* 2003.

62. Casey ML, Kemp DS, Paul KG, Cox DD: **Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles.** *J Org Chem* 1973, **38**:2294-2301.

63. Richard JP: **A paradigm for enzyme-catalyzed proton transfer at carbon: Triosephosphate isomerase.** *Biochemistry* 2012, **51**:2652-2661.

64. Kim SW, Cha SS, Cho HS, Kim JS, Ha NC, Cho MJ, Joo S, Kim KK, Choi KY, Oh BH: **High-resolution crystal structures of delta5-3-ketosteroid isomerase with and without a reaction intermediate analogue.** *Biochemistry* 1997, **36**:14030-14036.

65. van der Kamp MW, Chaudret R, Mulholland AJ: **QM/MM modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis.** *Febs j* 2013, **280**:3120-3131.

66. Li YK, Kuliopulos A, Mildvan AS, Talalay P: **Environments and mechanistic roles of the tyrosine residues of Δ^5-3-ketosteroid isomerase.** *Biochemistry* 1993, **32**:1816-1824.

67. Weintraub H, Alfsen A, Baulieu EE: **Étude de la $\Delta 5 \rightarrow 4$ 3-oxostéroïde isomérase.** *Eur J Biochem* 1970, **12**:217-221.

68. Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR: **Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution.** *Nature* 1965, **206**:757-761.

69. Jencks WP: **Binding energy, specificity, and enzymic catalysis: The circe effect.** In *Advances in Enzymology and Related Areas of Molecular Biology.* Edited by: John Wiley & Sons, Inc.; 2006:219-410.

70. Anderson VE: **Quantifying energetic contributions to ground state destabilization.** *Archives of Biochemistry and Biophysics* 2005, **433**:27-33.

71. Griswold WR, Castro JN, Fisher AJ, Toney MD: **Ground-state electronic destabilization via hyperconjugation in aspartate aminotransferase.** *J Am Chem Soc* 2012, **134**:8436-8438.

72. Zhang Y, Schramm VL: **Ground-state destabilization in orotate phosphoribosyltransferases by binding isotope effects.** *Biochemistry* 2011, **50**:4813-4818.

73. Ruben EA, Schwans JP, Sonnett M, Natarajan A, Gonzalez A, Tsai Y, Herschlag D: **Ground state destabilization from a positioned general base in the ketosteroid isomerase active site.** *Biochemistry* 2013, **52**:1074-1081.

74. Gutteridge A, Thornton J: **Conformational change in substrate binding, catalysis and product release: an open and shut case?** *FEBS Letters* 2004, **567**:67-73.

75. Lisi GP, Loria JP: **Using NMR spectroscopy to elucidate the role of molecular motions in enzyme function.** *Prog Nucl Magn Reson Spectrosc* 2016, **92-93**:1-17.

76. Fenwick RB, van den Bedem H, Fraser JS, Wright PE: **Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR.** *Proc Natl Acad Sci USA* 2014, **111**:E445-E454.

77. Offenbacher AR, Hu S, Poss EM, Carr CAM, Scouras AD, Prigozhin DM, Iavarone AT, Palla A, Alber T, Fraser JS, Klinman JP: **Hydrogen–deuterium exchange of lipoxygenase uncovers a relationship between distal, solvent exposed protein motions and the thermal activation barrier for catalytic proton-coupled electron tunneling.** *ACS Cent Sci* 2017, **3**:570-579.

78. Carvalho ATP, Barrozo A, Doron D, Kilshtain AV, Major DT, Kamerlin SCL: **Challenges in computational studies of enzyme structure, function and dynamics.** *Journal of Molecular Graphics and Modelling* 2014, **54**:62-79.

79. Nashine VC, Hammes-Schiffer S, Benkovic SJ: **Coupled motions in enzyme catalysis.** *Curr Opin Chem Biol* 2010, **14**:644-651.

80. Rivoire O, Reynolds KA, Ranganathan R: **Evolution-based functional decomposition of proteins.** *PLoS Comput Biol* 2016, **12**:e1004817.

81. Lockless SW, Ranganathan R: **Evolutionarily conserved pathways of energetic connectivity in protein families.** *Science* 1999, **286**:295-299.

82. Hay S, Scrutton NS: **Good vibrations in enzyme-catalysed reactions.** *Nat Chem* 2012, **4**:161-168.

83. Boekelheide N, Salomon-Ferrer R, Miller TF, 3rd: **Dynamics and dissipation in enzyme catalysis.** *Proc Natl Acad Sci USA* 2011, **108**:16159-16163.

84. Nagel ZD, Klinman JP: **A 21st century revisionist's view at a turning point in enzymology.** *Nat Chem Biol* 2009, **5**:543-550.

85. Schwartz SD, Schramm VL: **Enzymatic transition states and dynamic motion in barrier crossing.** *Nat Chem Biol* 2009, **5**:551-558.

86. Nagel ZD, Klinman JP: **Tunneling and dynamics in enzymatic hydride transfer.** *Chem Rev* 2006, **106**:3095-3118.

87. Nagel ZD, Klinman JP: **Update 1 of: Tunneling and dynamics in enzymatic hydride transfer.** *Chem Rev* 2010, **110**:Pr41-67.

88. Knapp MJ, Rickert K, Klinman JP: **Temperature-dependent isotope effects in soybean lipoxygenase-1: Correlating hydrogen tunneling with protein dynamics.** *J Am Chem Soc* 2002, **124**:3865-3874.

89. Stojkovic V, Perissinotti LL, Willmer D, Benkovic SJ, Kohen A: **Effects of the donor-acceptor distance and dynamics on hydride tunneling in the dihydrofolate reductase catalyzed reaction.** *J Am Chem Soc* 2012, **134**:1738-1745.

90. McCusker KP, Klinman JP: **An active-site phenylalanine directs substrate binding and C-H cleavage in the alpha-ketoglutarate-dependent dioxygenase TauD.** *J Am Chem Soc* 2010, **132**:5114-5120.

91. Glickman MH, Wiseman JS, Klinman JP: **Extremely large isotope effects in the soybean lipoxygenase-linoleic acid reaction.** *J Am Chem Soc* 1994, **116**:793-794.

92. Kamerlin SC, Warshel A: **At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?** *Proteins* 2010, **78**:1339-1375.

93. Liu H, Warshel A: **Origin of the temperature dependence of isotope effects in enzymatic reactions: the case of dihydrofolate reductase.** *J Phys Chem B* 2007, **111**:7852-7861.

94. Liu H, Warshel A: **The catalytic effect of dihydrofolate reductase and its mutants is determined by reorganization energies.** *Biochemistry* 2007, **46**:6011-6025.

95. Dametto M, Antoniou D, Schwartz SD: **Barrier crossing in dihydrofolate reductasedoes not involve a rate-promoting vibration.** *Mol Phys* 2012, **110**:531-536.

96. Pollack SJ, Jacobs JW, Schultz PG: **Selective chemical catalysis by an antibody.** *Science* 1986, **234**:1570-1573.

97. Kries H, Blomberg R, Hilvert D: **De novo enzymes by computational design.** *Curr Opin Chem Biol* 2013, **17**:221-228.

98. Bolon DN, Mayo SL: **Enzyme-like proteins by computational design.** *Proc Natl Acad Sci USA* 2001, **98**:14274-14279.

99. Richter F, Blomberg R, Khare SD, Kiss G, Kuzin AP, Smith AJ, Gallaher J, Pianowski Z, Helgeson RC, Grjasnow A, et al.: **Computational design of catalytic dyads and oxyanion holes for ester hydrolysis.** *J Am Chem Soc* 2012, **134**:16197-16206.

100. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF, 3rd, et al.: **De novo computational design of retro-aldol enzymes.** *Science* 2008, **319**:1387-1391.

101. Althoff EA, Wang L, Jiang L, Giger L, Lassila JK, Wang Z, Smith M, Hari S, Kast P, Herschlag D, et al.: **Robust design and optimization of retroaldol enzymes.** *Protein Sci* 2012, **21**:717-726.

102. Bjelic S, Kipnis Y, Wang L, Pianowski Z, Vorobiev S, Su M, Seetharaman J, Xiao R, Kornhaber G, Hunt JF, et al.: **Exploration of alternate catalytic mechanisms and optimization strategies for retroaldolase design.** *J Mol Biol* 2014, **426**:256-271.

103. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL: **Iterative approach to computational enzyme design.** *Proc Natl Acad Sci USA* 2012, **109**:3790-3795.

104. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al.: **Kemp elimination catalysts by computational enzyme design.** *Nature* 2008, **453**:190-195.

105. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, et al.: **Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction.** *Science* 2010, **329**:309-313.

106. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, Players F, Stoddard BL, Popovic Z, Baker D: **Increased Diels-Alderase activity through backbone remodeling guided by Foldit players.** *Nat Biotechnol* 2012, **30**:190-192.

107. Bjelic S, Nivón LG, Çelebi-Ölçüm N, Kiss G, Rosewall CF, Lovick HM, Ingalls EL, Gallaher JL, Seetharaman J, Lew S, et al.: **Computational design of enone-binding proteins with catalytic activity for the Morita–Baylis–Hillman reaction.** *ACS Chem Biol* 2013, **8**:749-757.

108. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D: **De novo enzyme design using Rosetta3.** *PLoS One* 2011, **6**:e19230.

109. Hongdilokkul N: **Biochemical and biophysical basis of enzyme evolvability.** *PhD thesis* 2014.

110. Kemp DS, Casey ML: **Physical organic chemistry of benzisoxazoles. II. Linearity of the Broensted free energy relation for the base-catalyzed decomposition of benzisoxazoles.** *J Am Chem Soc* 1973, **95**:6670-6680.

111. Na J, Houk KN, Hilvert D: **Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production.** *J Am Chem Soc* 1996, **118**:6462-6471.

112. Lamba V, Sanchez E, Fanning LR, Howe K, Alvarez MA, Herschlag D, Forconi M: **Kemp eliminase activity of ketosteroid isomerase.** *Biochemistry* 2017, **56**:582-591.

113. Miao Y, Metzner R, Asano Y: **Kemp elimination catalyzed by naturally occurring aldoxime dehydratases.** *Chembiochem* 2017, **18**:451-454.

114. Li A, Wang B, Ilie A, Dubey KD, Bange G, Korendovych IV, Shaik S, Reetz MT: **A redox-mediated Kemp eliminase.** *Nat Commun* 2017, **8**:14876.

115. Hollfelder F, Kirby AJ, Tawfik DS: **Off-the-shelf proteins that rival tailor-made antibodies as catalysts.** *Nature* 1996, **383**:60-62.

116. Wang L, Althoff EA, Bolduc J, Jiang L, Moody J, Lassila JK, Giger L, Hilvert D, Stoddard B, Baker D: **Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of de novo enzyme design.** *J Mol Biol* 2012, **415**:615-625.

117. Kiss G, Rothlisberger D, Baker D, Houk KN: **Evaluation and ranking of enzyme designs.** *Protein Sci* 2010, **19**:1760-1773.

118. Mak WS, Siegel JB: **Computational enzyme design: Transitioning from catalytic proteins to enzymes.** *Curr Op Struc Biol* 2014, **27**:87-94.

119. Khersonsky O, Rothlisberger D, Dym O, Albeck S, Jackson CJ, Baker D, Tawfik DS: Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J Mol Biol* 2010, **396**:1025-1042.

120. Khersonsky O, Kiss G, Rothlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS: Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci USA* 2012, **109**:10358-10363.

121. Khersonsky O, Rothlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk KN, Baker D, Tawfik DS: Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. *J Mol Biol* 2011, **407**:391-412.

122. Giger L, Caner S, Obexer R, Kast P, Baker D, Ban N, Hilvert D: Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat Chem Biol* 2013, **9**:494-498.

123. Preiswerk N, Beck T, Schulz JD, Milovnik P, Mayer C, Siegel JB, Baker D, Hilvert D: Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase. *Proc Natl Acad Sci USA* 2014, **111**:8013-8018.

124. Lassila JK, Baker D, Herschlag D: Origins of catalysis by computationally designed retroaldolase enzymes. *Proc Natl Acad Sci USA* 2010, **107**:4937-4942.

125. Jindal G, Ramachandran B, Bora RP, Warshel A: Exploring the development of ground-state destabilization and transition-state stabilization in two directed evolution paths of Kemp eliminases. *ACS Catal* 2017, **7**:3301-3305.

126. Swiderek K, Tunon I, Moliner V, Bertran J: Revealing the origin of the efficiency of the de novo designed Kemp eliminase HG-3.17 by comparison with the former developed HG-3. *Chemistry* 2017, **23**:7582-7589.

127. Bhowmick A, Sharma SC, Head-Gordon T: The importance of the scaffold for de novo enzymes: A case study with Kemp eliminase. *J Am Chem Soc* 2017, **139**:5793-5800.

128. Frushicheva MP, Cao J, Chu ZT, Warshel A: Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *J Chem Educ* 2010, **107**:16869-16874.

129. Davey JA, Chica RA: Multistate approaches in computational protein design. *Protein Science* 2012, **21**:1241-1252.

130. Zeymer C, Zschoche R, Hilvert D: Optimization of enzyme mechanism along the evolutionary trajectory of a computationally designed (retro-)aldolase. *J Am Chem Soc* 2017, **139**:12541-12549.

131. Lassila JK: Conformational diversity and computational enzyme design. *Curr Op Chem Biol* 2010, **14**:676-682.

132. Huang PS, Boyken SE, Baker D: The coming of age of de novo protein design. *Nature* 2016, **537**:320-327.

133. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003, **302**:1364-1368.

134. Huang PS, Feldmeier K, Parmeggiani F, Velasco DAF, Hocker B, Baker D: De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* 2016, **12**:29-34.

135. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R: **The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters.** *Biochemistry* 2011, **50**:4402-4410.

136. Hawkinson DC, Eames TC, Pollack RM: **Energetics of 3-oxo-delta 5-steroid isomerase: source of the catalytic power of the enzyme.** *Biochemistry* 1991, **30**:10849-10858.

137. Knowles JR, Albery WJ: **Perfection in enzyme catalysis: The energetics of triosephosphate isomerase.** *Acc Chem Res* 1977, **10**:105-111.

138. Gupta RD, Goldsmith M, Ashani Y, Simo Y, Mullokandov G, Bar H, Ben-David M, Leader H, Margalit R, Silman I, et al.: **Directed evolution of hydrolases for prevention of G-type nerve agent intoxication.** *Nat Chem Biol* 2011, **7**:120-125.

139. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN: **Computational enzyme design.** *Angew Chem Int Ed* 2013, **52**:5700-5725.

140. Bryan P, Pantoliano MW, Quill SG, Hsiao HY, Poulos T: **Site-directed mutagenesis and the role of the oxyanion hole in subtilisin.** *Proc Natl Acad Sci USA* 1986, **83**:3743-3745.

141. Taylor MS, Jacobsen EN: **Asymmetric catalysis by chiral hydrogen-bond donors.** *Angew Chem Int Ed* 2006, **45**:1520-1543.

142. Debler EW, Müller R, Hilvert D, Wilson IA: **An aspartate and a water molecule mediate efficient acid-base catalysis in a tailored antibody pocket.** *Proc Natl Acad Sci USA* 2009, **106**:18539-18544.

143. Kries H: **Tailor-made biocatalysts by enzyme design, redesign, and directed evolution.** *PhD thesis* 2014.

144. Bloch SJ: **Dissecting an oxyanion-hole in a computatioanlly derived enzyme.** *Mater Thesis* 2014.

145. Kries H, Bloch JS, Bunzel HA, Pinkas DM, Mittl PRE, Hilvert D: **Contribution of oxyanion stabilization to Kemp eliminase proficiency.** *manuscript in preperation.*

146. Northrop DB: **On the meaning of K_m and V/K in enzyme kinetics.** *J Chem Educ* 1998, **75**:1153.

147. Peng C, Bernhard Schlegel H: **Combining synchronous transit and quasi-newton methods to find transition states.** *Israel Journal of Chemistry* 1993, **33**:449-454.

148. Mader MM, Bartlett PA: **Binding energy and catalysis: The implications for transition-state analogs and catalytic antibodies.** *Chem Rev* 1997, **97**:1281-1302.

149. Dunitz JD: **The entropic cost of bound water in crystals and biomolecules.** *Science* 1994, **264**:670.

150. Fried SD, Bagchi S, Boxer SG: **Extreme electric fields power catalysis in the active site of ketosteroid isomerase.** *Science* 2014, **346**:1510-1514.

151. Frushicheva MP, Cao J, Warshel A: **Challenges and advances in validating enzyme design proposals: The case of Kemp eliminase catalysis.** *Biochemistry* 2011, **50**:3849-3858.

152. Houck WJ, Pollack RM: **Activation enthalpies and entropies for the microscopic rate constants of acetate-catalyzed isomerization of 5-androstene-3,17-dione.** *J Am Chem Soc* 2003, **125**:10206-10212.

153. Fafarman AT, Sigala PA, Herschlag D, Boxer SG: **Decomposition of vibrational shifts of nitriles into electrostatic and hydrogen-bonding effects.** *J Am Chem Soc* 2010, **132**:12811-12813.

154. Fried SD, Boxer SG: **BIOPHYSICS. Response to Comments on "Extreme electric fields power catalysis in the active site of ketosteroid isomerase".** *Science* 2015, **349**:936.

155. Campbell EC, Correy GJ, Mabbitt PD, Buckle AM, Tokiriki N, Jackson CJ: **Laboratory evolution of protein conformational dynamics.** *Curr Opin Struct Biol* 2017, **50**:49-57.

156. Goodey NM, Benkovic SJ: **Allosteric regulation and catalysis emerge via a common route.** *Nat Chem Biol* 2008, **4**:474-482.

157. Nguyen V, Wilson C, Hoemberger M, Stiller JB, Agafonov RV, Kutter S, English J, Theobald DL, Kern D: **Evolutionary drivers of thermoadaptation in enzyme catalysis.** *Science* 2017, **355**:289-294.

158. Tokuriki N, Tawfik DS: **Protein dynamism and evolvability.** *Science* 2009, **324**:203-207.

159. Khersonsky O, Tawfik DS: **Enzyme promiscuity: A mechanistic and evolutionary perspective.** *Annu Rev Biochem* 2010, **79**:471-505.

160. Jackson CJ, Foo J-L, Tokuriki N, Afriat L, Carr PD, Kim H-K, Schenk G, Tawfik DS, Ollis DL: **Conformational sampling, catalysis, and evolution of the bacterial phosphotriesterase.** *Proc Natl Acad Sci USA* 2009, **106**:21631-21636.

161. Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS: **Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme.** *Nat Commun* 2012, **3**:1257.

162. Kaltenbach M, Jackson CJ, Campbell EC, Hollfelder F, Tokuriki N: **Reverse evolution leads to genotypic incompatibility despite functional and active site convergence.** *Elife* 2015, **4**.

163. Campbell E, Kaltenbach M, Correy GJ, Carr PD, Porebski BT, Livingstone EK, Afriat-Jurnou L, Buckle AM, Weik M, Hollfelder F, et al.: **The role of protein dynamics in the evolution of new enzyme function.** *Nat Chem Biol* 2016, **12**:944-950.

164. Evans MG, Polanyi M: **Some applications of the transition state method to the calculation of reaction velocities, especially in solution.** *Trans Faraday Soc* 1935, **31**:875-894.

165. Wolfenden R: **Massive thermal acceleration of the emergence of primordial chemistry, the incidence of spontaneous mutation, and the evolution of enzymes.** *J Biol Chem* 2014, **289**:30198-30204.

166. Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas AF, Gilmanshin RI: **Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe.** *Biopolymers* 1991, **31**:119-128.

167. Gontero B, Meunier JC, Buc J, Ricard J: **The 'slow' pH-induced conformational transition of chloroplast fructose 1,6-bisphosphatase and the control of the Calvin cycle.** *Eur J Biochem* 1984, **145**:485-488.

168. Ricard J, Noat G, Nari J: **pH-induced co-operative effects in hysteretic enzymes.** *Eur J Biochem* 1984, **145**:311-317.

169. Valero E, García-Carmona F: **pH-induced kinetic co-operativity of a thylakoid-bound polyphenol oxidase.** *Biochem J* 1992, **286**:623-626.

170. Nari J, Noat G, Ricard J: **pH-induced co-operative effects in hysteretic enzymes. 2. pH-induced co-operative effects in a cell-wall beta-glucosyltransferase.** *Eur J Biochem* 1984, **145**:319-322.

171. Iranpoor N, Firouzabadi H, Nowrouzi N: **A novel method for the highly efficient synthesis of 1,2-benzisoxazoles under neutral conditions using the Ph3P/DDQ system.** *Tetrahedron Lett* 2006, **47**:8247-8250.

172. Manetsch R, Zheng L, Reymond MT, Woggon W-D, Reymond J-L: **A catalytic antibody against a tocopherol cyclase inhibitor.** *Chem Eur J* 2004, **10**:2487-2506.

173. Moore JL, Silvestri AP, de Alaniz JR, DiRocco DA, Rovis T: **Mechanistic investigation of the enantioselective intramolecular Stetter reaction: Proton transfer is the first irreversible step.** *Org Lett* 2011, **13**:1742-1745.

174. Hollfelder F, Kirby AJ, Tawfik DS, Kikuchi K, Hilvert D: **Characterization of proton-transfer catalysis by serum albumins.** *J Am Chem Soc* 2000, **122**:1022-1029.

175. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A: **Protein identification and analysis tools on the ExPASy server.** *John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press* 2005:571-607

176. Cheng Y, Prusoff WH: **Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction.** *Biochem Pharmacol* 1973, **22**:3099-3108.

177. Kabsch W: **XDS.** *Acta Crystallogr D* 2010, **66**:125-132.

178. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software.** *J Appl Crystallogr* 2007, **40**:658-674.

179. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics.** *Acta Crystallogr D* 2004, **60**:2126-2132.

180. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD: **Towards automated crystallographic structure refinement with phenix.refine.** *Acta Crystallogr D* 2012, **68**:352-367.

181. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al.: **Gaussian 09, Revision B.01.** Edited by. Wallingford CT; 2009.

182. Miller BG, Wolfenden R: **Catalytic proficiency: The unusual case of OMP decarboxylase.** *Annu Rev Biochem* 2002, **71**:847-885.

183. Hennig M, Darimont BD, Jansonius JN, Kirschner K: **The catalytic mechanism of indole-3-glycerol phosphate synthase: crystal structures of complexes of the enzyme from Sulfolobus solfataricus with substrate analogue, substrate, and product.** *J Mol Biol* 2002, **319**:757-766.

184. Reetz MT, Kahakeaw D, Lohmer R: **Addressing the numbers problem in directed evolution.** *ChemBioChem* 2008, **9**:1797-1804.

185. Acevedo-Rocha CG, Hoebenreich S, Reetz MT: **Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution.** *Methods Mol Biol* 2014, **1179**:103-128.

186. Rogers TA, Bommarius AS: **Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes.** *Chemical engineering science* 2010, **65**:2118-2124.

187. Somarowthu S, Ondrechen MJ: **POOL server: Machine learning application for functional site prediction in proteins.** *Bioinformatics* 2012, **28**:2078-2079.

188. Somarowthu S, Brodkin HR, D'Aquino JA, Ringe D, Ondrechen MJ, Beuning PJ: **A tale of two isomerases: Compact versus extended active sites in ketosteroid isomerase and phosphoglucose isomerase.** *Biochemistry* 2011, **50**:9283-9295.

189. Mair P, Gielen F, Hollfelder F: **Exploring sequence space in search of functional enzymes using microfluidic droplets.** *Curr Opin Chem Biol* 2017, **37**:137-144.

190. Longwell CK, Labanieh L, Cochran JR: **High-throughput screening technologies for enzyme engineering.** *Curr Opin Biotechnol* 2017, **48**:196-202.

191. Debler EW, Ito S, Seebeck FP, Heine A, Hilvert D, Wilson IA: **Structural origins of efficient proton abstraction from carbon by a catalytic antibody.** *Proc Natl Acad Sci USA* 2005, **102**:4984-4989.

192. Higuchi R, Krummel B, Saiki RK: **A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions.** *Nucleic Acids Res* 1988, **16**:7351-7367.

193. Sambrook J, Russell DW: **Molecular Cloning: A Laboratory Manual. .** *Quart Rev Biol* 2001, **76**:348-349.

194. Richard JP: **Acid-base catalysis of the elimination and isomerization reactions of triose phosphates.** *J Am Chem Soc* 1984, **106**:4926-4936.

195. Genre-Grandpierre A, Tellier C, Loirat M-J, Blanchard D, Hodgson DRW, Hollfelder F, Kirby AJ: **Catalysis of the Kemp elimination by antibodies elicited against a cationic hapten.** *Bioorg Med Chem Lett* 1997, **7**:2497-2502.

196. Thorn SN, Daniels RG, Auditor M-TM, Hilvert D: **Large rate accelerations in antibody catalysis by strategic use of haptenic charge.** *Nature* 1995, **373**:228.

197. Turner NJ: **Directed evolution drives the next generation of biocatalysts.** *Nat Chem Biol* 2009, **5**:567-573.

198. Colin PY, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvonen M, Morgavi DP, Janssen DB, Hollfelder F: **Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics.** *Nat Commun* 2015, **6**:10008.

199. Najah M, Calbrix R, Mahendra-Wijaya IP, Beneyton T, Griffiths AD, Drevelle A: **Droplet-based microfluidics platform for ultra-high-throughput bioprospecting of cellulolytic microorganisms.** *Chem Biol* 2014, **21**:1722-1732.

200. Jackel C, Hilvert D: **Biocatalysts by evolution.** *Curr Opin Biotechnol* 2010, **21**:753-759.

201. Truppo MD: **Biocatalysis in the pharmaceutical industry: The need for speed.** *ACS Med Chem Lett* 2017, **8**:476-480.

202. Packer MS, Liu DR: **Methods for the directed evolution of proteins.** *Nat Rev Genet* 2015, **16**:379-394.

203. Lafferty M, Dycaico MJ: **GigaMatrix: A novel ultrahigh throughput protein optimization and discovery platform.** *Methods Enzymol* 2004, **388**:119-134.

204. Chen B, Lim S, Kannan A, Alford SC, Sunden F, Herschlag D, Dimov IK, Baer TM, Cochran JR: **High-throughput analysis and protein engineering using microcapillary arrays.** *Nat Chem Biol* 2016, **12**:76-81.

205. Lim S, Chen B, Kariolis MS, Dimov IK, Baer TM, Cochran JR: **Engineering high affinity protein–protein interactions using a high-throughput microcapillary array platform.** *ACS Chem Biol* 2017, **12**:336-341.

206. Yang G, Withers SG: **Ultrahigh-throughput FACS-based screening for directed enzyme evolution.** *ChemBioChem* 2009, **10**:2704-2715.

207. Varadarajan N, Cantor JR, Georgiou G, Iverson BL: **Construction and flow cytometric screening of targeted enzyme libraries.** *Nat Protoc* 2009, **4**:893-901.

208. Agresti JJ, Antipov E, Abate AR, Ahn K, Rowat AC, Baret JC, Marquez M, Klibanov AM, Griffiths AD, Weitz DA: **Ultrahigh-throughput screening in drop-based microfluidics for directed evolution.** *Proc Natl Acad Sci USA* 2010, **107**:4004-4009.

209. Kintses B, Hein C, Mohamed MF, Fischlechner M, Courtois F, Laine C, Hollfelder F: **Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution.** *Chem Biol* 2012, **19**:1001-1009.

210. Granieri L, Baret JC, Griffiths AD, Merten CA: **High-throughput screening of enzymes by retroviral display using droplet-based microfluidics.** *Chem Biol* 2010, **17**:229-235.

211. Obexer R, Pott M, Zeymer C, Griffiths AD, Hilvert D: **Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting.** *Protein Eng Des Sel* 2016, **29**:355-366.

212. Beneyton T, Coldren F, Baret JC, Griffiths AD, Taly V: **CotA laccase: High-throughput manipulation and analysis of recombinant enzyme libraries expressed in E. coli using droplet-based microfluidics.** *Analyst* 2014, **139**:3314-3323.

213. Ng EX, Miller MA, Jing T, Chen CH: **Single cell multiplexed assay for proteolytic activity using droplet microfluidics.** *Biosens Bioelectron* 2016, **81**:408-414.

214. Price AK, MacConnell AB, Paegel BM: **hnuSABR: Photochemical dose-response bead screening in droplets.** *Anal Chem* 2016, **88**:2904-2911.

215. Larsen AC, Dunn MR, Hatch A, Sau SP, Youngbull C, Chaput JC: **A general strategy for expanding polymerase function by droplet microfluidics.** *Nat Commun* 2016, **7**:11235.

216. Ryckelynck M, Baudrey S, Rick C, Marin A, Coldren F, Westhof E, Griffiths AD: **Using droplet-based microfluidics to improve the catalytic properties of RNA under multiple-turnover conditions.** *RNA* 2015, **21**:458-469.

217. Beneyton T, Wijaya IP, Postros P, Najah M, Leblond P, Couvent A, Mayot E, Griffiths AD, Drevelle A: **High-throughput screening of filamentous fungi using nanoliter-range droplet-based microfluidics.** *Sci Rep* 2016, **6**:27223.

218. Ostafe R, Prodanovic R, Lloyd Ung W, Weitz DA, Fischer R: **A high-throughput cellulase screening system based on droplet microfluidics.** *Biomicrofluidics* 2014, **8**:041102.

219. Gielen F, Hours R, Emond S, Fischlechner M, Schell U, Hollfelder F: **Ultrahigh-throughput-directed enzyme evolution by absorbance-activated droplet sorting (AADS).** *Proc Natl Acad Sci USA* 2016, **113**:E7383-e7389.

220. Khare SD, Kipnis Y, Greisen P, Jr., Takeuchi R, Ashani Y, Goldsmith M, Song Y, Gallaher JL, Silman I, Leader H, et al.: **Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis.** *Nat Chem Biol* 2012, **8**:294-300.

221. Obexer R, Studer S, Giger L, Pinkas DM, Grütter MG, Baker D, Hilvert D: **Active site plasticity of a computationally designed retro-aldolase enzyme.** *ChemCatChem* 2014, **6**:1043-1050.

222. Bornscheuer UT, Altenbuchner J, Meyer HH: **Directed evolution of an esterase: screening of enzyme libraries based on pH-indicators and a growth assay.** *Bioorg Med Chem* 1999, **7**:2169-2173.

223. Scharnagl C, Raupp-Kossmann RA: **Solution pK_a Values of the Green Fluorescent Protein Chromophore from Hybrid Quantum-Classical Calculations.** *J Phys Chem B* 2004, **108**:477-489.

224. Biedermann F, Hathazi D, Nau WM: **Associative chemosensing by fluorescent macrocycle-dye complexes - a versatile enzyme assay platform beyond indicator displacement.** *Chem Commun (Camb)* 2015, **51**:4977-4980.

225. Sakamoto S, Komatsu T, Ueno T, Hanaoka K, Urano Y: **Fluorescence detection of serum albumin with a turnover-based sensor utilizing Kemp elimination reaction.** *Bioorg Med Chem Lett* 2017, **27**:3464-3467.

226. Beckett AH, Walker J: **Steric interactions in substituted cyclohexadienes—II: Meso-substituted dihydroanthracenes: steric effects in the reactions of cis- and trans-isomers.** *Tetrahedron* 1963, **19**:545-556.

227. Ying L-Q, Branchaud BP: **Selective labeling and monitoring pH changes of lysosomes in living cells with fluorogenic pH sensors.** *Bioorg Med Chem Lett* 2011, **21**:3546-3549.

228. Aeschbacher M, Reinhardt CA, Zbinden G: **A rapid cell membrane permeability test using fluorescent dyes and flow cytometry.** *Cell Biol Toxicol* 1986, **2**:247-255.

229. Blomberg RMB: **Optimizing artifical enzymes with directed evolution.** *PhD thesis* 2011.

230. Savile CK, Janey JM, Mundorff EC, Moore JC, Tam S, Jarvis WR, Colbeck JC, Krebber A, Fleitz FJ, Brands J, et al.: **Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture.** *Science* 2010, **329**:305-309.

231. Brandenberg OF, Fasan R, Arnold FH: **Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions.** *Curr Opin Biotechnol* 2017, **47**:102-111.

232. Lu X, Li L, Wu R, Feng X, Li Z, Yang H, Wang C, Guo H, Galkin A, Herzberg O, et al.: **Kinetic Analysis of *Pseudomonas aeruginosa* Arginine Deiminase Mutants and Alternate Substrates Provides Insight into Structural Determinants of Function.** *Biochemistry* 2006, **45**:1162-1172.

233. Milburn CC, Lamble HJ, Theodossis A, Bull SD, Hough DW, Danson MJ, Taylor GL: **The structural basis of substrate promiscuity in glucose dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*.** *J Biol Chem* 2006, **281**:14796-14804.

234. James LC, Tawfik DS: **Conformational diversity and protein evolution - a 60-year-old hypothesis revisited.** *Trends Biochem Sci* 2003, **28**:361-368.

235. Meier S, Ozbek S: A biological cosmos of parallel universes: Does protein structural plasticity facilitate evolution? *Bioessays* 2007, **29**:1095-1104.

236. Blum M-M, Timperley CM, Williams GR, Thiermann H, Worek F: Inhibitory potency against human acetylcholinesterase and enzymatic hydrolysis of fluorogenic nerve agent mimics by human paraoxonase 1 and squid diisopropyl fluorophosphatase. *Biochemistry* 2008, **47**:5216-5224.

237. Khersonsky O, Tawfik DS: The histidine 115-histidine 134 dyad mediates the lactonase activity of mammalian serum paraoxonases. *J Biol Chem* 2006, **281**:7649-7656.

238. Yeung DT, Lenz DE, Cerasoli DM: Analysis of active-site amino-acid residues of human serum paraoxonase using competitive substrates. *Febs j* 2005, **272**:2225-2230.

239. Kaiser ET, Lawrence DS: Chemical mutation of enzyme active sites. *Science* 1984, **226**:505-511.

240. Bakker M, Rantwijk Fv, Sheldon RA: Metal substitution in thermolysin: Catalytic properties of tungstate thermolysin in sulfoxidation with H2O2. *Canadian Journal of Chemistry* 2002, **80**:622-625.

241. da Silva GFZ, Ming L-J: Catechol Oxidase Activity of Di-Cu2+-Substituted Aminopeptidase from Streptomyces griseus. *J Am Chem Soc* 2005, **127**:16380-16381.

242. Kourist R, Bartsch S, Fransson L, Hult K, Bornscheuer UT: Understanding promiscuous amidase activity of an esterase from *Bacillus subtilis*. *ChemBioChem* 2008, **9**:67-69.

243. Wang SC, Johnson WH, Jr., Whitman CP: The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3E-haloacrylates: implications for the evolution of new enzymatic activities. *J Am Chem Soc* 2003, **125**:14282-14283.

244. Wang SC, Person MD, Johnson WH, Whitman CP: Reactions of trans-3-chloroacrylic acid dehalogenase with acetylene substrates: Consequences of and evidence for a hydration reaction. *Biochemistry* 2003, **42**:8762-8773.

245. Poelarends GJ, Serrano H, Johnson WH, Hoffman DW, Whitman CP: The hydratase activity of malonate semialdehyde decarboxylase: Mechanistic and evolutionary implications. *J Am Chem Soc* 2004, **126**:15658-15659.

246. Wendt KU, Poralla K, Schulz GE: Structure and function of a squalene cyclase. *Science* 1997, **277**:1811-1815.

247. Syren PO, Henche S, Eichler A, Nestl BM, Hauer B: Squalene-hopene cyclases-evolution, dynamics and catalytic scope. *Curr Opin Struct Biol* 2016, **41**:73-82.

248. Hammer SC, Marjanovic A, Dominicus JM, Nestl BM, Hauer B: Squalene hopene cyclases are protonases for stereoselective Brønsted acid catalysis. *Nat Chem Biol* 2015, **11**:121-126.

249. Merlini V, Luparia M, Porta A, Zanoni G, Vidari G: Biomimetic cyclization of geraniol derivatives, a useful tool in the total synthesis of bioactive monocyclic terpenoids. *Nat Prod Commun* 2011, **6**:465-476.

250. Lenardão EJ, Botteselle GV, de Azambuja F, Perin G, Jacob RG: Citronellal as key compound in organic synthesis. *Tetrahedron* 2007, **63**:6671-6712.

251. Jensen RA: **Enzyme recruitment in evolution of new function.** *Annu Rev Microbiol* 1976, **30**:409-425.

252. Glasner ME, Gerlt JA, Babbitt PC: **Evolution of enzyme superfamilies.** *Curr Opin Chem Biol* 2006, **10**:492-497.

253. Merkl R, Sterner R: **Reconstruction of ancestral enzymes.** *Perspectives in Science* 2016, **9**:17-23.

254. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM: **Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-lactamases.** *J Am Chem Soc* 2013, **135**:2899-2902.

255. Reisinger B, Sperl J, Holinski A, Schmid V, Rajendran C, Carstensen L, Schlee S, Blanquart S, Merkl R, Sterner R: **Evidence for the existence of elaborate enzyme complexes in the pleoarchean era.** *J Am Chem Soc* 2014, **136**:122-129.

256. Garrabou X, Wicky BI, Hilvert D: **Fast Knoevenagel condensations catalyzed by an artificial Schiff-base-forming enzyme.** *J Am Chem Soc* 2016, **138**:6972-6974.

257. Garrabou X, Verez R, Hilvert D: **Enantiocomplementary synthesis of γ-nitroketones using designed and evolved carboligases.** *J Am Chem Soc* 2017, **139**:103-106.

258. Garrabou X, Beck T, Hilvert D: **A promiscuous de novo retro-aldolase catalyzes asymmetric michael additions via schiff base intermediates.** *Angew Chem Int Ed* 2015, **54**:5609-5612.

259. Garrabou X, Macdonald DS, Hilvert D: **Chemoselective Henry condensations catalyzed by artificial carboligases.** *Chemistry* 2017, **23**:6001-6003.

260. Wickham RJ: **How menthol alters tobacco-smoking behavior: A biological perspective.** *The Yale Journal of Biology and Medicine* 2015, **88**:279-287.

261. Kamatou GP, Vermaak I, Viljoen AM, Lawrence BM: **Menthol: A simple monoterpene with remarkable biological properties.** *Phytochemistry* 2013, **96**:15-25.

262. Schäfer B: **Menthol: Minze versus Tagasako-Prozess.** *Chemie in unserer Zeit* 2013, **47**:174-182.

263. Heydrich G, Gralla G, Rauls M, Schmidt-Leithoff J, Ebel K, Krause W, Oehlenschläger S, Jäkel C, Friedrich M, Bergner EJ, et al.: **(-)-Menthol synthesis by BASF.** *BASF SE, Ludwigshafen, DE* 2010, **US20100249467 A1**.

264. Friedrich M, Ebel K, Götz N: **(-)-menthol synthesis by BASF.** *BASF SE, Ludwigshafen, DE* 2009, **US7550633**.

265. Iwata T, Okeda Y, Hori Y: **(-)-menthol synthesis by Takasago.** *Takasago International Corporation, Tokyo, JP* 2004, **US6774269**.

266. Baines DA, Jones RA, Webb TC, Campion-Smith IH: **The chemistry of terpenes—I: The effect of hydrogen ion concentration and oxygen upon the acid catalysed cyclization of citral.** *Tetrahedron* 1970, **26**:4901-4913.

267. Crabtree RH, Davis MW: **Directing effects in homogeneous hydrogenation with [Ir(cod)(PCy3)(py)]PF6.** *J Org Chem* 1986, **51**:2655-2661.

268. Stockbridge RB, Lewis CA, Jr., Yuan Y, Wolfenden R: **Impact of temperature on the time required for the establishment of primordial biochemistry, and for the evolution of enzymes.** *Proc Natl Acad Sci USA* 2010, **107**:22102-22105.

269. Loll B, Kern J, Saenger W, Zouni A, Biesiadka J: **Towards complete cofactor arrangement in the 3.0Å resolution structure of photosystem II.** *Nature* 2005, **438**:1040.

270. Burgess BK, Lowe DJ: **Mechanism of molybdenum nitrogenase.** *Chem Rev* 1996, **96**:2983-3012.

271. Schwizer F, Okamoto Y, Heinisch T, Gu Y, Pellizzoni MM, Lebrun V, Reuter R, Köhler V, Lewis JC, Ward TR: **Artificial metalloenzymes: Reaction scope and optimization strategies.** *Chem Rev* 2017.

272. Drienovska I, Alonso-Cotchico L, Vidossich P, Lledos A, Marechal JD, Roelfes G: **Design of an enantioselective artificial metallo-hydratase enzyme containing an unnatural metal-binding amino acid.** *Chem Sci* 2017, **8**:7228-7235.

273. Hermanson G: **Bioconjugate Techniques.** *Academic Press* 3rd Edition.

274. Reetz MT, Rentzsch M, Pletsch A, Taglieber A, Hollmann F, Mondiere RJ, Dickmann N, Hocker B, Cerrone S, Haeger MC, et al.: **A robust protein host for anchoring chelating ligands and organocatalysts.** *Chembiochem* 2008, **9**:552-564.

275. Reetz MT, Peyralans JJ, Maichele A, Fu Y, Maywald M: **Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein.** *Chem Commun (Camb)* 2006:4318-4320.

276. Mallin H, Hestericova M, Reuter R, Ward TR: **Library design and screening protocol for artificial metalloenzymes based on the biotin-streptavidin technology.** *Nat Protoc* 2016, **11**:835-852.

277. Thomas CM, Ward TR: **Artificial metalloenzymes: Proteins as hosts for enantioselective catalysis.** *Chem Soc Rev* 2005, **34**:337-346.

278. Okrasa K, Kazlauskas RJ: **Manganese-substituted carbonic anhydrase as a new peroxidase.** *Chemistry* 2006, **12**:1587-1596.

279. Yamamura K, Kaiser ET: **Studies on the oxidase activity of copper(II) carboxypeptidase A.** *J Chem Soc Chem Commun* 1976:830-831.

280. Bos J, Roelfes G: **Artificial metalloenzymes for enantioselective catalysis.** *Curr Opin Chem Biol* 2014, **19**:135-143.

281. Coquiere D, Bos J, Beld J, Roelfes G: **Enantioselective artificial metalloenzymes based on a bovine pancreatic polypeptide scaffold.** *Angew Chem Int Ed* 2009, **48**:5159-5162.

282. Bos J, Fusetti F, Driessen AJ, Roelfes G: **Enantioselective artificial metalloenzymes by creation of a novel active site at the protein dimer interface.** *Angew Chem Int Ed* 2012, **51**:7472-7475.

283. Rioz-Martinez A, Roelfes G: **DNA-based hybrid catalysis.** *Curr Opin Chem Biol* 2015, **25**:80-87.

284. Voigt CA, Gordon DB, Mayo SL: **Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design.** *J Mol Biol* 2000, **299**:789-803.

285. Nechay MR, Valdez CE, Alexandrova AN: **Computational treatment of metalloproteins.** *J Phys Chem B* 2015, **119**:5945-5956.

286. Topală T, Bodoki A, Oprean L, Oprean R: **Bovine serum albumin interactions with metal complexes.** *Clujul Medical* 2014, **87**:215-219.

287. Hakansson K, Carlsson M, Svensson LA, Liljas A: **Structure of native and apo carbonic anhydrase II and structure of some of its anion-ligand complexes.** *J Mol Biol* 1992, **227**:1192-1204.

288. Tu C, Couton JM, Van Heeke G, Richards NG, Silverman DN: **Kinetic analysis of a mutant (His107-->Tyr) responsible for human carbonic anhydrase II deficiency syndrome.** *J Biol Chem* 1993, **268**:4775-4779.

289. Graslund S, Savitsky P, Muller-Knapp S: **In vivo biotinylation of antigens in** *E. coli*. *Methods Mol Biol* 2017, **1586**:337-344.

290. Studer S, Basler S, Bunzel HA, A. C, Jimenez-Oses G, Helgeson RC, Houk KN, Hilvert D: **Design and evolution of a metallo Diels-Alderase based on an artificial helical bundle.** *unpublished data.*

291. Salgado EN, Ambroggio XI, Brodin JD, Lewis RA, Kuhlman B, Tezcan FA: **Metal templated design of protein interfaces.** *Proc Natl Acad Sci USA* 2010, **107**:1827-1832.

292. Zhong S, Nieger M, Bihlmeier A, Shi M, Brase S: **Asymmetric organocatalytic synthesis of 4,6-bis(1H-indole-3-yl)-piperidine-2 carboxylates.** *Organic & Biomolecular Chemistry* 2014, **12**:3265-3270.

293. St. John SE, Jensen KC, Kang S, Chen Y, Calamini B, Mesecar AD, Lipton MA: **Design, synthesis, biological and structural evaluation of functionalized resveratrol analogues as inhibitors of quinone reductase 2.** *Bioorganic & Medicinal Chemistry* 2013, **21**:6022-6037.

294. Roderer K, Neuenschwander M, Codoni G, Sasso S, Gamper M, Kast P: **Functional mapping of protein-protein interactions in an enzyme complex by directed evolution.** *PLoS ONE* 2014, **9**:e116234.

Curriculum Vitae

Personal Data

Name	Hans Adrian Bunzel
Date of Birth	23.02.1989
Place of Birth	Remscheid, Germany
Citizenship	Germany

Education

PhD Studies
2013 – 2017

ETH Zurich, Switzerland

Degree:	Doctor of Science in Organic Chemistry
Supervisor:	Prof. Dr. Donald Hilvert
Thesis title:	De novo enzyme engineering: Directed evolution of computationally design enzymes

Undergraduate Studies
2011 - 2013

Chemistry studies at the University of Cologne, Germany

Degree:	M.Sc. in Chemistry
Supervisors:	Prof. Dr. Günter Schwarz
	Prof. Dr. Donald Hilvert
Thesis title:	Directed evolution of a highly efficient Kemp eliminase

2008 - 2011

Degree:	B.Sc. in Chemistry
Supervisors:	Prof. Dr. Günter Schwarz
Thesis title:	Identification of splice-specific interaction partners of gephyrin

Secondary School
1999 - 2008

Gymnasium Wermelskirchen, Germany

Degree:	Abitur (general qualification for university entrance)

Teaching Experiences

2013 – 2017	Supervision of undergraduate bachelor and masters level research projects in biochemistry and chemistry at ETH Zurich
2013 – 2017	Teaching assistant for biological chemistry courses at ETH Zurich

Publications

Bunzel HA, Garrabou X, Pott M, Hilvert D: **Speeding up enzyme discovery and engineering with ultrahigh-throughput methods.** *Curr Op Chem Biol*, **2018**, 149.

Selected Oral Presentations

Bunzel HA, Kries H, Hilvert D: **Characterization of a Kemp eliminase reveals strategies for evolution.** *Bridging Chemistry and Biology Workshop* 2017 Bad Boll, Germany.

Bunzel HA, Jiminez-Oses G, Houk KN, Hilvert D: **De novo design of a metallo Diels-Alderase.** *Bridging Chemistry and Biology Workshop* 2015, Schwarzenberg, Austria.

Bunzel HA, Kries H, Folger I, Joss D, Mittl PRE, Hilvert D: **Novel enzymes by computational design and catalytic promiscuity.** *Bridging Chemistry and Biology Workshop* 2017, Schwarzenberg, Austria.

Selected Poster Presentations

Bunzel HA, Kries H, Mittl PRE, Hilvert D: **Evolution of a computationally designed enzyme.** *4th International Conference on Novel Enzymes* 2014, Gent, Brussels.

Bunzel HA, Jiminez-Oses G, Houk KN, Hilvert D: **De novo engineering of a metallo Diels-Alderase by design and evolution.** *GRC "Metals in Biology"* 2016, Ventura, USA.

Bunzel HA, Kries H, Folger I, Joss D, Mittl PRE, Hilvert D: **Novel enzymes by computational design and catalytic promiscuity.** *Benzon Symposium* 2017, Copenhagen, Denmark.