Mechanistic investigation of artificially designed, light regulation of naturally occurring, and characterization of ancestral $(\beta\alpha)_8$-barrel enzymes

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der Fakultät für Biologie und vorklinische Medizin der Universität Regensburg

vorgelegt von

Bernd Reisinger

aus Deggendorf

im Jahr 2013

Das Promotionsgesuch wurde eingereicht am: 26.11.2013

Die Arbeit wurde angeleitet von: Prof. Dr. Reinhard Sterner

Unterschrift:

Table of contents

Abstract

The $(\beta\alpha)_8$-barrel fold is the most frequently observed topology among enzymes. Due to the spatial separation of their activity- and stability-mediating sites, $(\beta\alpha)_8$-barrel enzymes are extremely versatile and catalyze a wide array of cellular reactions. Thus, this particular fold provides an ideal tool for modifying catalytic activity and studying enzyme evolution.

Several $(\beta\alpha)_8$-barrel enzymes are involved in the metabolism of amino acids. Along these lines, N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase (HisA) and the cyclase subunit of imidazole glycerol phosphate synthase (HisF) catalyze two consecutive steps in the biosynthesis of histidine, namely a sugar isomerization and a cycloligase/lyase reaction. By analogy with HisA, the enzyme phosphoribosyl anthranilate isomerase (TrpF) performs a chemically equivalent isomerization reaction within the biosynthesis of tryptophan. As HisA, HisF, and TrpF accommodate their phosphorylated substrates via a common phosphate binding site, an evolutionary linkage seems to exist between these three $(\beta\alpha)_8$-barrel proteins. Consequently, HisA and HisF could be engineered to bind and process the TrpF substrate phosphoribosyl anthranilate (PRA). In both cases, a single aspartate-to-valine substitution was sufficient to establish PRA isomerase activity and the combination with a second aspartate-to-valine exchange significantly improved the turnover number. Since HisA and TrpF operate through an identical acid-base mechanism, the same enzymatic mechanism could be expected for the PRA isomerase activities generated on the HisA and HisF scaffolds. However, while mutational analyses and docking studies revealed a respective general base, no appropriate general acid could be identified. Therefore, the mechanistic foundation of the artificially designed PRA isomerase activity was addressed in the first part of this work.

Initially recorded pH dependences substantiated the mechanistic differences of the naturally occurring and the artificially designed PRA isomerase activity. While TrpF wild-type exhibited the expected bell-shaped pH profile, merely an acidic limb was observed for the investigated HisA variants, suggesting that the hitherto unidentified general acid is rate-limiting for the engineered reaction. A hint on the nature of the catalytic acid was subsequently obtained from the crystal structure of the HisF double mutant with a bound product analogue. The structure suggested that within the enzyme–substrate complex the anthranilic acid moiety of PRA might donate a proton to the furanose ring oxygen of its sugar moiety. This productive geometry of PRA was compared with a non-productive binding mode in molecular dynamics simulations of

the HisA variants. In contrast to HisA wild-type, all variants clearly favor an orientation of PRA required for catalysis. Furthermore, the introduced valine residues clearly upshift the pK_a value of anthranilic acid to a catalytically useful range. Mixed quantum and molecular mechanics calculations of the HisA double mutant with bound PRA finally demonstrated that an internal proton transfer is also feasible from an energetic point of view and presumably proceeds via a bridging water molecule. In sum, the artificial PRA isomerase activity established on the HisA and HisF scaffolds appears to be partly based on substrate-assisted catalysis and thus mechanistically deviates from the PRA isomerase activity of TrpF.

The second part of this dissertation dealt with the evolution of cellular complexity. Modern organisms are highly developed molecular machineries, which rely on elaborate enzyme systems. There is considerable interest to figure out which degree of enzymatic sophistication had already been reached in the very early phases of biological evolution. Along these lines, substantial progress has been made in the field of ancestral sequence reconstruction, which links computational and evolutionary biology and enables the characterization of extinct proteins. In extreme cases, enzymes from the last universal ancestor of cellular organisms (LUCA) can be studied. The LUCA preceded the diversification into the three domains of life and existed at least 3.5 billion years ago.

Exceptional catalytic features like substrate channeling and allosteric communication are observed in the imidazole glycerol phosphate synthase bi-enzyme complex of the cyclase HisF and the glutaminase HisH. In an attempt to analyze its primal characteristics, we reconstructed a HisF enzyme from the LUCA era (LUCA-HisF). LUCA-HisF could be expressed solubly in *Escherichia coli* and purified with a high yield. The protein furthermore shows a high thermostability and a folding mechanism comparable to extant $(\beta\alpha)_8$-barrel enzymes. Accordingly, its subsequently solved crystal structure equals contemporary HisF proteins. Beyond its structural integrity, LUCA-HisF proved to be a highly active and specific enzyme. As its catalytic sophistication could only be completely assessed in combination with an interacting glutaminase, we additionally reconstructed a respective LUCA-HisH sequence. Although both proteins form a stoichiometric complex with high affinity, no catalytic activity could be determined for LUCA-HisH, probably due to uncertainties in the reconstruction process. We instead turned to a complex between LUCA-HisF and the extant HisH enzyme from *Zymomonas mobilis*. Remarkably, LUCA-HisF could both stimulate the catalytic efficiency of the interacting glutaminase and transport the produced ammonia to its active site via a molecular channel. The evolution of these elaborate features therefore must already have been completed in the LUCA era.

The artificial control of enzymatic activity has been a long-standing goal in the field of protein design. Here, light provides an ideal trigger signal, since it enables a non-invasive and spatiotemporal regulation of biological activity. However, despite various attempts, only few enzymes have been successfully regulated by light so far. Therefore, the design of a light-controllable inhibitor of the $(\beta\alpha)_8$-barrel enzyme PriA from *Mycobacterium tuberculosis* (*mt*PriA) was the aim of the third section of this thesis. Interestingly, the bisubstrate-specific isomerase *mt*PriA is able to catalyze both the HisA and TrpF reaction and displays a potential target for anti-tuberculosis drugs, since humans can synthesize neither histidine nor tryptophan.

For the construction of the potential inhibitors, two particular features of *mt*PriA could be harnessed, both of which originate from the molecular evolution from a $(\beta\alpha)_4$-half-barrel precursor: the protein exhibits a striking twofold rotational symmetry as well as two opposite phosphate binding sites. Consequently, we chose the twofold symmetric, photoswitchable 1,2-dithienylethene (DTE) as a structural core and equipped it with terminal phosphate or phosphonate anchors. The synthesized DTE compounds could reversibly be toggled between a ring-open and ring-closed form by irradiation with UV and visible light, respectively. Both isomers were thermally stable, nearly quantitatively formed and robust over various switching cycles. When tested in steady-state enzyme kinetics, the open isomers of all DTE-phosphates and DTE-phosphonates competitively inhibited the *mt*PriA activity with the inhibition constants lying in the low micromolar range. Notably, the inhibition activity was lowered up to a factor of eight upon ring-closure, where the enzymatic performance could be directly controlled during catalysis. The different binding affinities obtained upon irradiation seem to be based on a change in the conformational flexibility. Along these lines, molecular dynamics simulations of *mt*PriA with an inhibitor bound in both isomeric forms demonstrated that the ring-open isomers can readily adapt to the active site of *mt*PriA. In contrast, due to its restricted mobility, the interaction of the ring-closed form with the enzyme is energetically less favorable. Thus, the dual anchoring of photoswitchable inhibitors constitutes a viable design concept for the reversible regulation of enzymatic activity. The approach may additionally be transferred to other $(\beta\alpha)_8$-barrel proteins, as phosphate is a frequently encountered element of metabolic substrates.

Zusammenfassung

Die $(\beta\alpha)_8$-Fass-Struktur ist der am häufigsten vorkommende Faltungstyp unter Enzymen. Durch die räumliche Separation von aktivitäts- und stabilitätsvermittelnden Aminosäuren, sind $(\beta\alpha)_8$-Fass-Enzyme extrem vielseitig und katalysieren ein breites Spektrum an zellulären Reaktionen. Deswegen stellt dieser besondere Faltungstyp ein ideales Gerüst zur Modifikation katalytischer Aktivität und zur Untersuchung enzymatischer Evolutionmechanismen dar.

Gleich mehrere $(\beta\alpha)_8$-Fass-Enzyme sind am Aminosäuremetabolismus beteiligt. So katalysieren die N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazol-4-carboxamid ribonucleotid Isomerase (HisA) und die Zyklaseuntereinheit der Imidazolglycerin-phosphat Synthase (HisF) zwei aufeinanderfolgende Schritte in der Histidinbiosyn-these. HisA isomerisiert eine Ribosegruppe zu einer Aminoketose, die anschließend von HisF in einer Zykloligase/Lyase-Reaktion gespalten wird. Weiterhin existiert eine zur HisA-Reaktion äquivalente Zuckerisomerisierung innerhalb der Tryptophanbio-synthese – das zuständige Enzym wird TrpF genannt. Die $(\beta\alpha)_8$-Fass-Proteine HisA, HisF und TrpF scheinen evolutionär verknüpft zu sein, da sie ihre phosphorylierten Substrate durch eine identische Phosphatbindestelle fixieren. So konnte die Umsetzung des TrpF-Substrates Phosphoribosylanthranilat (PRA) sowohl auf dem HisA- als auch auf dem HisF-Gerüst bereits durch den Austausch eines Aspartatrestes zu Valin etabliert werden. In beiden Fällen wurde die PRA-Isomerase-Aktivität durch Kombina-tion mit dem Austausch eines weiteren Aspartatrestes zu Valin deutlich gesteigert. Da die HisA- und TrpF-Reaktion auf dem gleichen Säure-Base-Mechanismus beruhen, konnte ein identischer Reaktionsablauf auch für die auf HisA und HisF erzeugten PRA-Isomerisierungen erwartet werden. Obwohl Mutationsanalysen und Docking-Ansätze zur Identifikation einer entsprechenden katalytischen Base führten, ergaben sich mit diesen Methoden jedoch keine Hinweise auf die allgemeine Säure. Deswegen wurden die mechanistischen Grundlagen der auf den HisA- und HisF-Grundgerüsten etablierten PRA-Isomerase-Aktivität im ersten Teil dieser Arbeit eingehender analy-siert.

Dazu wurden zunächst die pH-Abhängigkeiten der neu erzeugten und der natürlich vorkommenden PRA-Isomerase-Aktivität gemessen. Dabei zeigte der TrpF-Wildtyp das erwartete glockenförmige pH-Profil, wohingegen bei den HisA-Varianten lediglich der Säureast beobachtet werden konnte. Die Varianten scheinen demnach einem Reaktionsmechanismus zu folgen, dessen Rate durch eine bis zu diesem Zeitpunkt unbekannte katalytische Säure limitiert wird. Welche Gruppe die Rolle der

allgemeinen Säure einnimmt, konnte anschließend mit Hilfe der Kristallstruktur einer HisF-Variante mit gebundenem Produktanalogon näher eingegrenzt werden. Diese Struktur legte den Schluss nahe, dass innerhalb des Enzym-Substrat-Komplexes der Anthranilsäureanteil von PRA ein Proton auf den Furanoseringsauerstoff des Zucker-anteils übertragen könnte. Eine entsprechende Geometrie des Substrates PRA wurde daraufhin in Moleküldynamiksimulationen der HisA-Varianten mit einem unproduk-tiven Bindungsmodus verglichen. Im Gegensatz zum HisA-Wildtyp, präferierten dabei alle Mutanten die katalytisch produktive Orientierung von PRA. Desweiteren verringern die eingeführten Valin-Reste die Acidität der Anthranilsäure und führen somit zu einer für die Katalyse notwendigen Erhöhung ihres pK_a-Wertes. Schließlich zeigten kombinierte quantenmechanische/molekülmechanische Berechnungen der HisA-Doppelmutante mit gebundenem PRA, dass ein Protonentransfer innerhalb des Substrates auch energetisch möglich ist, wobei dafür vermutlich ein verbrückendes Wassermolekül benötigt wird. Zusammenfassend lässt sich sagen, dass die auf den HisA- und HisF-Grundgerüsten künstlich erzeugte PRA-Isomerase-Aktivität zum Teil auf Substrat-vermittelter Katalyse beruht und sich daher von der PRA-Isomerase-Aktivität von TrpF unterscheidet.

Der zweite Teil dieser Dissertation hatte die Evolution zellulärer Komplexität zum Thema. Während moderne Organismen bekanntlich auf ausgeklügelten Enzymsys-temen basieren, existieren fast ausschließlich nur theoretische Arbeiten über die Komplexität der Spezies, die in einer sehr frühen Phase der Evolution auftraten. Jedoch wurde durch die kürzlich etablierte Rekonstruktion von Vorläufersequenzen eine Verknüpfung zwischen Bioinformatik und Evolutionsbiologie geschaffen. Mit Hilfe dieser Technik können die Sequenzen nicht mehr existierender Enzyme abgeleitet und damit ihre Eigenschaften charakterisiert werden. Im Extremfall lassen sich so Enzyme des letzten gemeinsamen Vorfahren aller zellulären Organismen (LUCA), aus dem sich vor etwa 3,5 Milliarden Jahren die heutigen Domänen des Lebens entwickelten, untersuchen.

Der Imidazolglycerinphosphat-Synthase-Komplex aus der Zyklase HisF und der Glutaminase HisH zeichnet sich durch bemerkenswerte katalytische Eigenschaften aus. So wird bei Bindung des Liganden an HisF die Glutaminase-Aktivität allosterisch aktiviert und der gebildete Ammoniak durch einen Kanal zum aktiven Zentrum von HisF transportiert. Um die ursprünglichen Eigenschaften der komplexen Synthase HisF zu charakterisieren, wurde ein HisF-Enzym aus dem LUCA-Zeitalter rekon-struiert (LUCA-HisF). Das Protein konnte löslich in *Escherichia coli* exprimiert und in hohen Ausbeuten gereinigt werden. LUCA-HisF zeigt eine hohe Thermostabilität und sowohl seine Kristallstruktur als auch sein Faltungsmechanismus gleichen rezenten HisF-Proteinen. Darüber hinaus zeigten *in vivo*-Studien und *in vitro*-Messungen, dass

LUCA-HisF ein hochaktives und spezifisches Enzym ist. Da seine katalytische Komplexität jedoch nur im Zusammenspiel mit einer interagierenden Glutaminase vollständig analysiert werden konnte, wurde zusätzlich die Sequenz eines entsprechenden LUCA-HisH-Proteins abgeleitet. Obwohl beide Proteine einen stöchiometrischen und hoch-affinen Komplex ausbilden, erwies sich LUCA-HisH als katalytisch inaktiv, was vermutlich auf einen fehlerbehafteten Rekonstruktionsprozess zurückzuführen ist. Deshalb wurde stattdessen ein Komplex zwischen LUCA-HisF und dem rezenten HisH-Enzym aus *Zymomonas mobilis* untersucht. Tatsächlich führt die Ligandenbindung an LUCA-HisF zu einer erheblichen Steigerung der Glutaminase-Aktivität, wobei der dadurch erzeugte Ammoniak durch den verbindenden Kanal zum aktiven Zentrum von HisF gelangen kann. Die Evolution dieser komplexen Eigenschaften muss deswegen bereits vor circa 3,5 Milliarden Jahren abgeschlossen gewesen sein.

Die artifizielle Regulation enzymatischer Aktivität ist ein seit Langem bestehendes Ziel auf dem Gebiet des Proteindesigns. Die Steuerung durch Licht ist besonders attraktiv, da sie eine nicht-invasive, räumliche und zeitliche Kontrolle biologischer Aktivität ermöglicht. Trotz zahlreicher Ansätze konnten jedoch bislang nur wenige Enzyme effizient durch Licht reguliert werden. Deswegen wurde im dritten Abschnitt dieser Arbeit ein lichtsteuerbarer Inhibitor des $(\beta\alpha)_8$-Fass-Enzyms PriA aus *Mycobacterium tuberculosis* (*mt*PriA) designt. Die Isomerase *mt*PriA katalysiert sowohl die HisA- als auch die TrpF-Reaktion und stellt ein potentielles Zielprotein für die Entwicklung von Antituberkulotika dar, weil Menschen weder Histidin noch Tryptophan herstellen können.

Zwei besondere Strukturmerkmale konnten für das Design potentieller *mt*PriA-Inhibitoren genutzt werden: Aufgrund seiner Evolution aus einem $(\beta\alpha)_4$-Halbfass-Vorläufer, verfügt *mt*PriA über eine auffällige zweifache Rotationssymmetrie sowie über zwei gegenüberliegende Phosphatbindestellen. Dementsprechend wurde das zweifach rotationssymmetrische, lichtschaltbare Molekül 1,2-Dithienylethen (DTE) als strukturelles Gerüst der möglichen Inhibitoren verwendet und mit terminalen Phosphat- bzw. Phosphonat-Ankern versehen. Die synthetisierten DTE-Verbindungen konnten durch Bestrahlung mit UV- bzw. sichtbarem Licht reversibel zwischen einem offenen und einem geschlossenem Isomer hin- und hergeschaltet werden. Dieser Zyklus konnte mehrfach wiederholt werden, wobei beide Formen mit hohen Ausbeuten gebildet wurden und thermisch stabil waren. Die Verbindungen wurden anschließend in *steady-state*-Enzymkinetiken charakterisiert. Dabei zeigte sich, dass die offenen Formen aller DTE-Phosphate und -Phosphonate die *mt*PriA-Aktivität kompetitiv hemmen. Aus den erhaltenen Daten konnten Inhibitionskonstanten im niedrigen mikromolaren Bereich abgeleitet werden. Bemerkenswerterweise zeigten die geschlos-

senen DTE-Formen bis zu achtfach reduzierte Inhibitionsaktivitäten. Dadurch konnte die enzymatische Umsatzgeschwindigkeit durch Bestrahlung mit UV-Licht unmittelbar während der Katalyse beeinflusst werden. Die ungleichen Bindungsaffinitäten der offenen und geschlossenen Formen basieren vorwiegend auf Unterschieden in der konformationellen Flexibilität. Dementsprechend zeigten Moleküldynamiksimulationen von *mt*PriA mit gebundenem Inhibitor, dass die offene Form sich ohne Weiteres an das aktive Zentrum anpassen kann. Im Gegensatz dazu ist das geschlossene Isomer erheblich in seiner Beweglichkeit eingeschränkt, was zu einer energetisch ungünstigeren Interaktion mit der enzymatischen Bindetasche führt. Die doppelte Verankerung von lichtschaltbaren Inhibitoren scheint daher ein erfolgreiches Designkonzept zur reversiblen Steuerung enzymatischer Aktivität darzustellen. Der hier beschriebene Ansatz kann im Prinzip auch auf weitere $(\beta\alpha)_8$-Fass-Enzyme übertragen werden, da Phosphat ein weit verbreiteter Baustein metabolischer Substrate ist.

1 General Introduction

1.1 The $(\beta\alpha)_8$-barrel: one fold with diverse functions

Enzymes are exceptional catalysts which facilitate a vast array of reactions under mild conditions [1]. In the absence of enzymes, vital reactions would last up to several million years [2] and thus by far exceed the life-span of living organisms [3]. The decarboxylation of orotidine 5`-phosphate, for example, proceeds with a half-time of 78 million years in neutral solution. However, in the presence of orotidine 5`-phosphate decarboxylase from *Saccharomyces cerevisiae* the reaction is accelerated by a factor of 10^{17}, yielding a half-time of merely 18 milliseconds [4]. The catalytic efficiencies of most enzymes lie within a similar range as they are perfectly adapted to the conditions present in cells [2, 5]. Furthermore, in order to avoid side products, enzymes perform with high specificity and excellent chemo-, regio- and stereoselectivities [6].

Except for a relative small number of catalytic RNAs, all enzymes are proteins [1]. The amino acid chain of each protein folds into a distinct three-dimensional topology which determines its stability and function. Surprisingly, although more than 41 million protein sequences are known (UniProtKB/TrEMBL, release August 2013, [7]), solely 1195 different folds have been observed to date (SCOP database, version 1.75, [8]). It has further been estimated that only around 2000 folds exist among naturally occurring proteins, some of which are more likely and hence more frequently encountered than others [9].

The most common fold among enzymes is the $(\beta\alpha)_8$- or TIM-barrel fold: it is observed in about 10 % of all structurally characterized proteins [10]. Besides the P-loop-containing nucleoside triphosphate hydrolase fold and the DNA/RNA-binding three-helical bundle fold, the $(\beta\alpha)_8$-barrel fold belongs to the most ancient protein architectures [11]. Along these lines, $(\beta\alpha)_8$-barrel enzymes are mainly involved in basic biological processes like molecular or energy metabolism [12]. Herein, they catalyze more than 60 distinct reactions and accordingly cover all enzyme classes as defined by the Enzyme Commission [13], except for ligases [12].

The catalytic versatility of the $(\beta\alpha)_8$-barrel fold is based on its unique structural assembly [10]. The canonical $(\beta\alpha)_8$-barrel is composed of eight modular $(\beta\alpha)$-units which are intrinsically linked via $\beta\alpha$-loops, whereas the individual modules are connected by $\alpha\beta$-loops (Figure 1A). In the resulting tertiary structure, the β-strands form a central β-barrel, which is surrounded by the amphiphilic α-helices (Figure 1B). Their hydrophobic residues interact with the β-strands, while the opposite, hydrophilic side is exposed to the solvent.
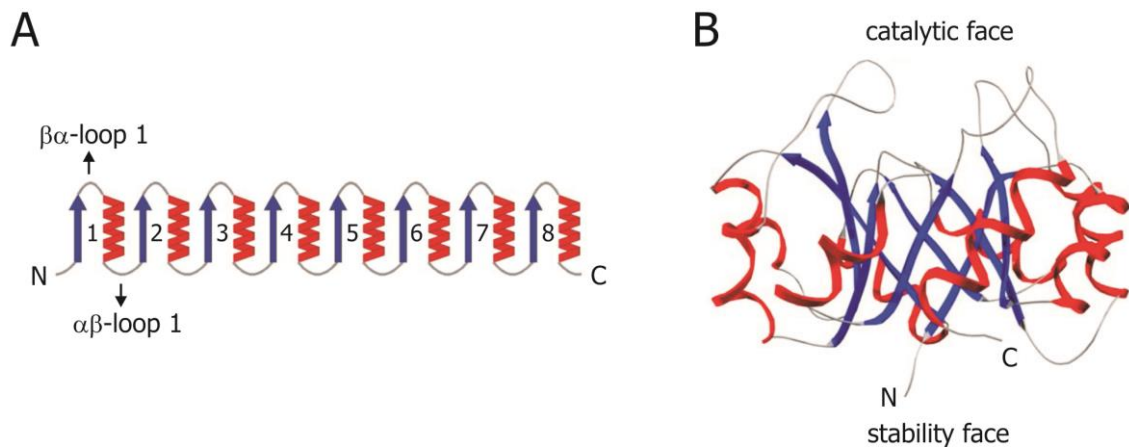
Figure 1. The $(\beta\alpha)_8$-barrel fold.
A: Schematic depiction of the right-handed $(\beta\alpha)$-super-secondary structure. B: Side view of the fold`s three-dimensional structure in ribbon representation. The barrel is separated in a catalytic and a stability face. The figure was modified from [10]. β-strands and α-helices are depicted in blue and red, respectively; the connecting loops are shown in grey.

In all known $(\beta\alpha)_8$-barrel enzymes, amino acids important for catalysis are located at the C-terminal ends of the β-strands and in the βα-loops (activity face). In contrast, the remainder of the fold, including the shorter αβ-loops, mediates the overall stability (stability face). As a consequence, the catalytic activity can in principle be modified without compromising stability, which renders the fold an ideal scaffold for studying natural enzyme evolution and designing novel catalytic activities [10, 14]. Thus, the swapping of βα-loops led to altered substrate specificities [15, 16], the introduction of few mutations enabled the catalysis of mechanistically similar reactions on either monofunctional [17] or promiscuous scaffolds [18-20] and even fairly unrelated reactions could be established [21, 22]. In addition, rational *de novo* enzyme design favored this particular fold over others, likely because catalytic and substrate-binding residues could easily be positioned in accordance with the desired transition state [23-25].

1.2 Analyzed $(\beta\alpha)_8$-barrel enzymes

The $(\beta\alpha)_8$-barrel fold is rather determined by the overall distribution of nonpolar, polar and charged residues than by details in the amino acid sequence [10]. Therefore, the sequence similarities of different $(\beta\alpha)_8$-barrel enzymes are generally low [26] with approximately 20 % identity at best [27, 28]. For this reason, it has been lively debated for more than two decades whether $(\beta\alpha)_8$-barrels result from convergent evolution to a stable protein fold or divergent evolution from a common ancestor [29, 30]. Even

extensive comparisons of sequence, structure and function could not definitely rule out one of the two scenarios [12]. However, especially $(\beta\alpha)_8$-barrel enzymes involved in tryptophan and histidine biosynthesis share some striking features strongly suggesting an evolutionary connection. For example, phosphoribosylanthranilate (PRA) isomerase (TrpF), indole-3-glycerol phosphate synthase (TrpC), and the α-subunit of tryptophan synthase (TrpA), which catalyze three consecutive steps in tryptophan biosynthesis, anchor their mono-phosphorylated substrates via a common C-terminal phosphate binding site [28]. An analogous observation was made for two successive enzymes in histidine biosynthesis, namely N´-[(5´-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and imidazole glycerol phosphate (ImGP) synthase (HisF). In order to accommodate their bi-phosphorylated substrates, HisA and HisF utilize both the above mentioned C-terminal and an additional, N-terminal phosphate binding motif [31]. Furthermore, the entire folds of both enzymes exhibit a noticeable two-fold symmetry, which prompted their evolution from a $(\beta\alpha)_4$-half-barrel ancestor via gene duplication and fusion [32]. This process was successfully mimicked in the laboratory by constructing stable $(\beta\alpha)_8$-barrel proteins from different $(\beta\alpha)_4$-barrel modules. A well-defined tertiary structure was observed for the chimeric protein HisAF, consisting of the N-terminal $(\beta\alpha)_{1-4}$-units of HisA and the C-terminal $(\beta\alpha)_{5-8}$-units of HisF [33]. After fusing two copies of the C-terminal half of HisF, further optimization of the initial construct first led to a protein termed Sym1 and finally resulted in Sym2, whose stability even surpasses the one of the natural HisF template [33-35]. Most remarkably, few mutations were sufficient to establish enzymatic activity on both the HisAF and Sym2 scaffold, corroborating the postulated evolutionary pathway [36, 37].

Interestingly, TrpF and HisA catalyze chemically equivalent isomerization reactions, even if their substrates significantly differ in size [38]. In a so-called Amadori rearrangement, the aminoaldoses PRA and ProFAR are converted to the corresponding ketoses 1-(2-carboxy-phenylamino)-1´-deoxyribulose-5´-phosphate (CdRP) and N´-[(5´-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) (Figure 2). This linkage between the biosynthetic pathways of tryptophan and histidine is strengthened by the existence of a bisubstrate-specific isomerase PriA, which was found to catalyze both the isomerization of PRA and ProFAR in some *Actinobacteria* [39, 40].
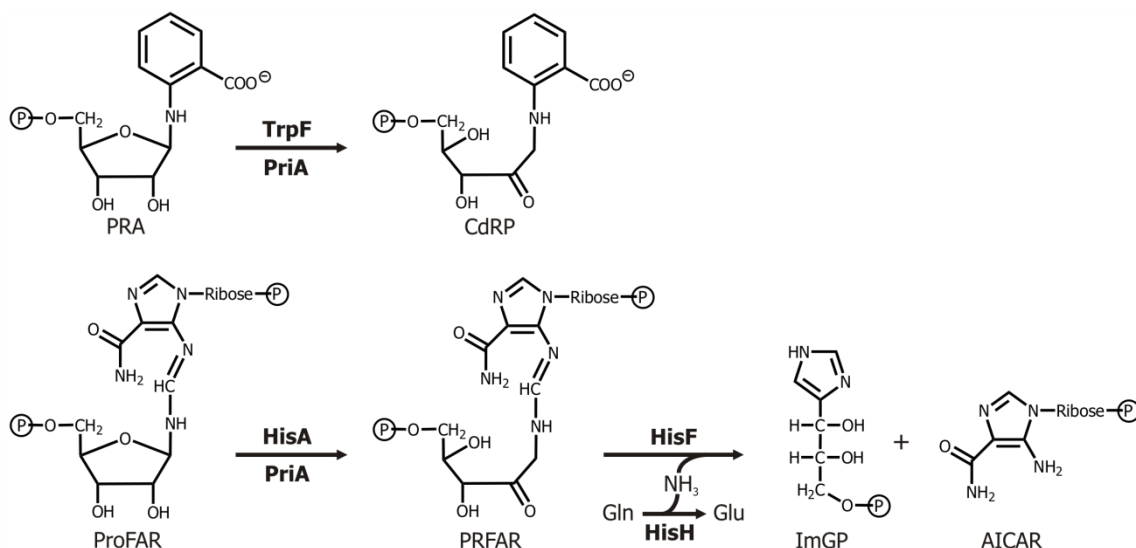
Figure 2. Reactions catalyzed by the $(\beta\alpha)_8$-barrel enzymes TrpF, HisA, PriA and HisF.
TrpF (tryptophan biosynthesis) and HisA (histidine biosynthesis) catalyze the chemically analogous Amadori rearrangements of the aminoaldoses PRA and ProFAR to the corresponding ketoses CdRP and PRFAR. The bifunctional enzyme PriA is able to catalyze both reactions. PRFAR is further metabolized to ImGP and AICAR by the ImGP synthase, which constitutes a bienzyme complex of the synthase HisF and the glutaminase HisH. While HisF performs the actual cycloligase/lyase reaction, HisH provides the required ammonia molecule via hydrolysis of glutamine (Gln) to glutamate (Glu).

In comparison to the rather simple sugar isomerizations of TrpF and HisA, the reaction catalyzed by ImGP synthase is much more sophisticated. Belonging to the family of glutamine amidotransferases, ImGP synthase consists of a heterodimeric complex of the synthase HisF and the glutaminase HisH [41]. The latter hydrolyzes glutamine to glutamate and thereby produces ammonia, which is subsequently transported to the active site of HisF through an extended molecular channel [42, 43]. HisF then mediates the nucleophilic attack of the nascent ammonia molecule to PRFAR to yield ImGP and 5-aminoimidazole-4-carboxamide ribotide (AICAR) in a unique cycloligase/lyase reaction (Figure 2) [41]. Since AICAR is further used in *de novo* purine biosynthesis, ImGP synthase constitutes a branch-point enzyme in primary metabolism [44]. Most strikingly, the enzymatic activities of HisH and HisF are strictly coupled to prevent wasteful hydrolysis of glutamine: the glutaminase activity of HisH only reaches a significant level when the substrate PRFAR is bound to HisF [41, 44, 45]. The mechanism of this allosteric stimulation seems to be a complicated, delocalized process [46-48] and still has not been completely comprehended [49].

In the course of this work, the specified enzyme systems were investigated in three different projects. In the following, the respective objectives are introduced and the obtained results are presented and discussed in three separate sections.

2 Projects

## 2.1 Project A: Mechanistic investigation of artificially generated PRA isomerase activity

2.1.1 Introduction and objective

The biosynthesis pathways of tryptophan and histidine share chemically equivalent sugar isomerization reactions, which are catalyzed by the enzymes TrpF and HisA (Figure 2). Consistently, both proteins operate via an identical acid-base mechanism [31, 50]. It therefore seemed plausible to interconvert the enzymatic activities of TrpF and HisA – an intention whose success would underpin their evolutionary relationship and the catalytic versatility of the underlying $(\beta\alpha)_8$-barrel fold [51]. Although both substrates contain a ribose-5-phosphate moiety, they significantly differ in size (Figure 2). The HisA substrate ProFAR is equipped with a second ribose-5-phosphate group and thus exhibits about twice the molecular weight of the TrpF substrate PRA. Hence, HisA was chosen as a scaffold on which PRA isomerase (PRAI) activity should be established. This goal was accomplished by a combination of random mutagenesis and *in vivo* selection [51]. Under selective pressure, HisA even acquired low PRAI activity due to spontaneous mutations [52]. Furthermore, structurally related enzymes like HisF and TrpA could readily be modified in a way that they were able to process PRA to CdRP [53, 54] and few amino acid exchanges were sufficient to generate PRAI activity on the artificial and initially inert proteins HisAF and Sym2 [36, 37]. Most remarkably, glutamine phosphoribosylpyrophosphate amidotransferase (PurF), which is unrelated to TrpF in sequence, tertiary structure and catalytic mechanism, unexpectedly showed promiscuous and evolvable PRAI activity [21].

As the active sites of HisA and HisF closely match each other in terms of sequence and structure [32], it is not surprising that analogous mutations were found to establish PRAI activity on HisA and HisF from *Thermotoga maritima* and on their chimera HisAF. Notably, the introduction of a single equivalent amino acid exchange at the C-terminal end of β-strand 5 gave rise to a weak turnover of PRA to CdRP in all three cases (D127V in HisA and HisAF, D130V in HisF) [51, 54]. It has been suggested that the removal of the aspartate side chains abolishes electrostatic repulsion with the negatively charged anthranilate moiety of PRA [36, 54]. A similar mechanism was observed for the bisubstrate-specific isomerase PriA, which effectively shields the respective aspartate residue from PRA by insertion of a positively charged arginine side chain [50]. Along these lines, the introduction of another aspartate-to-valine

mutation in βα-loops 6 (D169V in HisA, D176V in HisF and D173V in HisAF) also was sufficient to gain moderate PRAI activity, at least on the HisA and HisAF scaffold [36]. Most strikingly, the combination of both substitutions significantly enhanced the catalytic efficiencies of all three enzymes. Acting synergistically, D127V+D169V (HisA), D130V+D176V (HisAF) and D127V+D173V (HisF) increased the k_{cat}/K_M^{PRA} values 16-, 28- and 247-fold compared to the sum of the k_{cat}/K_M^{PRA} values of the single mutants (*Table 1, publication A*) [36]. While the resulting catalytic efficiency is still poor in the case of HisF, the double mutants of HisAF and HisA reached notable k_{cat}/K_M^{PRA} values of about 10^3 M^{-1}s^{-1} and 10^4 M^{-1}s^{-1}.

TrpF accomplishes the Amadori rearrangement of PRA via a general acid-base mechanism as outlined in Figure 3 [38]. The initial protonation of the furanose ring oxygen by Asp126 leads to the formation of a secondary aldimine. Acting as an electron sink, this Schiff base intermediate facilitates the deprotonation of the adjacent carbon atom by Cys7. In the final step, the generated enol form of CdRP tautomerizes to the corresponding keto form in an enzyme-independent manner [55].
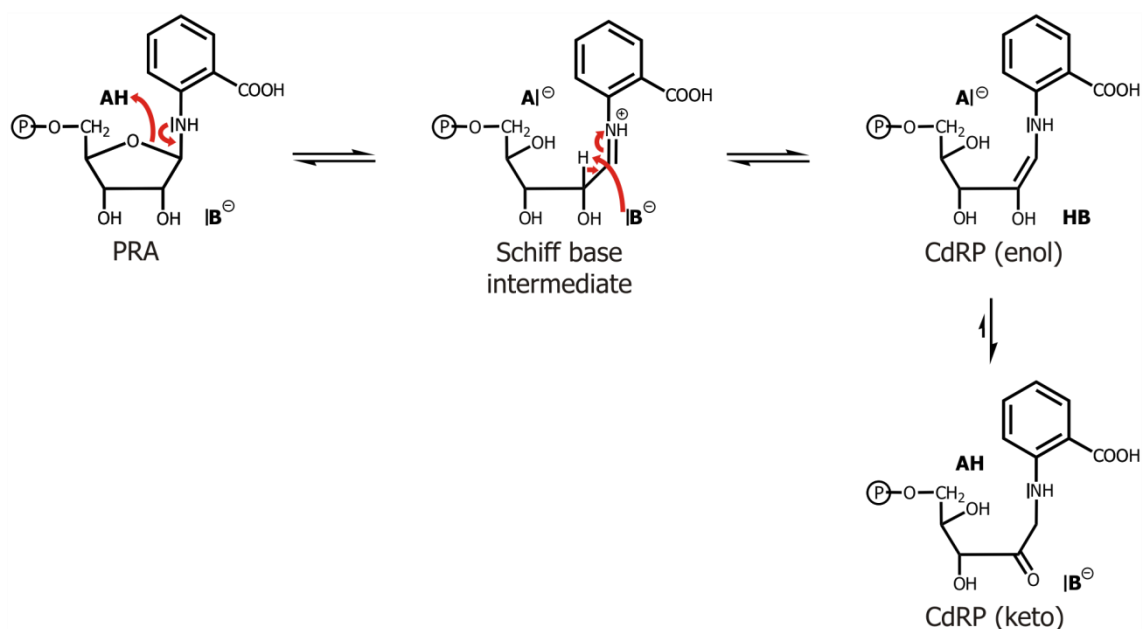
Figure 3. Mechanism of PRA isomerase TrpF.
In the initial step, the furanose ring oxygen is protonated by the general acid (Asp126). The thereby formed Schiff base intermediate acts as an electron sink and thus enables the general base (Cys7) to abstract a proton from the adjacent carbon atom. Finally, the enolamine form of CdRP spontaneously tautomerizes to the more stable keto form [38]. Modified from *Figure 2, publication A*.

Since Cys7 and Asp126 act on different parts of the sugar moiety, they are situated at opposite sides of the catalytic face, namely at the C-terminal ends of β-strands 1 and 6. As mentioned before, the isomerization of ProFAR catalyzed by HisA proceeds via

an identical reaction mechanism [38]. In addition, HisA and HisF share essential features of the active site like type and location of catalytically acting residues [32, 33]. So, it seems tempting to assume that the enzymatic mechanism observed for TrpF also holds for the PRAI activities engineered on HisA, HisF and HisAF. Accordingly, mutational studies proofed aspartate residues corresponding to Cys7 in TrpF crucial to PRAI activity (Asp8 in HisA and HisAF, Asp11 in HisF) [36, 54]. Subsequently performed docking studies revealed that these indispensable amino acids are perfectly located to act as general bases, as PRA is exclusively anchored in the C-terminal phosphate binding pocket also used by TrpF [36]. However, attempts to identify an appropriate general acid at the opposite barrel side failed. The only contemplable candidates – the aforementioned aspartate residues in β-strand 5 and βα-loop 6 – are mutated to valine in the best performing enzyme variants. To solve this apparent paradox, the artificially established PRAI activity was further investigated by pH-dependent activity measurements, crystal structure analysis and molecular dynamics (MD) as well as mixed quantum and molecular mechanics (QM/MM) calculations.

2.1.2 Summary and discussion

pH rate profiles have been shown to provide valuable insights into enzymatic reaction processes [56, 57]. Hence, to mechanistically compare naturally evolved and artificially designed PRAI activity, the turnover numbers k_{cat} of TrpF and the best performing mutant HisA_D127V+D169V were measured as a function of pH (Figure 4).
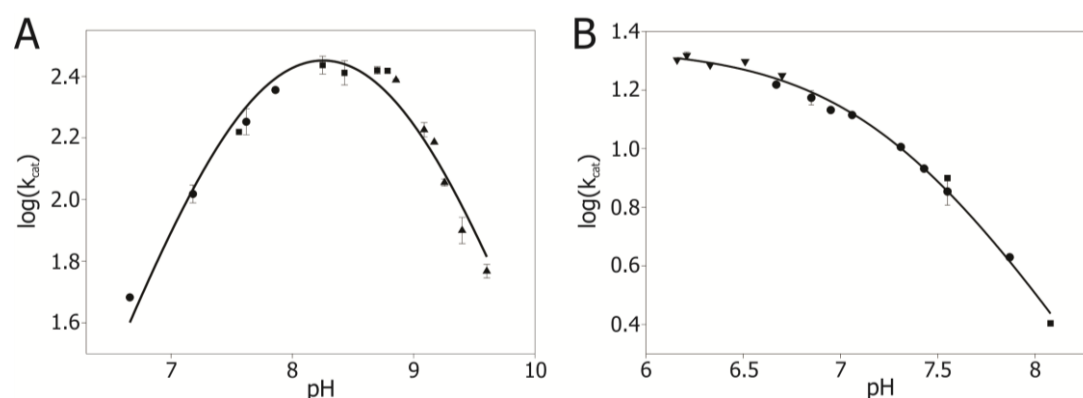

Figure 4. pH dependence of k_{cat} for TrpF (A) and HisA_D127V+D169V (B).
Turnover numbers were recorded at 15 °C using the buffer systems MES/NaOH (triangles down), HEPES/NaOH (circles), Tricin/NaOH (squares) and Glycin/NaOH (triangles up). Standard deviations, arising from at least three independent measurements, were determined across the whole pH range and are indicated, unless error bars are smaller than data symbols. The solid lines result from fitting the data to *equation 1, publication A* (panel A; TrpF) and *equation 2, publication A* (panel B; HisA_D127V+D169V), respectively. Thus, pK_a values of 7.6 ± 0.1 and 8.9 ± 0.1 for TrpF and 7.2 ± 0.1 for HisA_D127V+D169V were obtained. Modified from *Figure 3, publication A*.

As expected for a general acid-base catalysis, the k_{cat}-pH-profile of TrpF adopts a bell-shaped curve (Figure 4A). The ascending limb can be ascribed to an ionizable basic group which needs to be deprotonated to gain full activity, whereas the descending limb is distinctive of an acid which loses activity upon deprotonation with increasing pH. After fitting the data to *equation 1, publication A*, pK_a values of 7.6 ± 0.1 (base) and 8.9 ± 0.1 (acid) were obtained. While the former can easily be assigned to Cys7 [58], the latter only matches Asp126, if assuming a strong pK_a perturbation [59]. It cannot be excluded that the higher pK_a value reflects the phosphate group of PRA, whose deprotonation might cause an unfavorable change in substrate positioning and therefore reduce the overall reaction velocity. Interestingly, the k_{cat}-pH-profile of HisA_D127V+D169V merely consists of a descending limb (Figure 4), as do the curves determined for HisA_D127V and HisA_D169V (*Figure 3, publication A*). Fitting the data to *equation 2, publication A* revealed an underlying pK_a value of 7.2 ± 0.1. The exclusive presence of an acidic branch indicates that the performance of the associated general acid is rate-limiting for catalysis even at low pH values. Therefore, distinct differences seem to exist between the reaction mechanisms of TrpF and the PRAI active HisA variants, more precisely in the nature of the general acid.

Recently, crystal structures with bound ligand substantially contributed to deciphering the catalytic mechanisms of PRA isomerases TrpF and PriA [28, 60]. In both cases, the stable product analogue reduced CdRP (rCdRP), whose keto moiety has been converted to an alcohol (*Figure 2, publication A*), was used instead of the labile substrate PRA [61]. Titration studies showed that HisF_D130V+D176V and rCdRP form a stoichiometric and highly affine complex (*Figure S2, publication A*). The thermodynamic dissociation constant (K_d) of 0.2 μM is comparable to that observed for TrpF [62], and by far surpasses the rCdRP affinities of all other PRAI active variants [36, 51]. As a consequence, HisF_D130V+D176V was crystallized in the presence of rCdRP, and a structure of the complex could be resolved at 1.9 Å (Figure 5). Although HisF in principle contains two phosphate binding sites, rCdRP is solely fixed at the C-terminal one (Figure 5A). The observed binding preference is perfectly in line with a previously published model obtained by docking PRA into the apo structure of the PRAI active mutant HisA_H75Y+F111S+D127V+D169V [36]. Furthermore, the C-terminal binding mode implicates a proper positioning of Asp11 to act as the general base. Its carboxylate group is separated 4.9 Å from the respective C4 atom of rCdRP, which compares well with the distance of 3.9 Å determined between the sulfhydryl group of Cys7 and the C4 atom of rCdRP in the TrpF complex [38]. Hence, these findings substantiate mutational studies which prove not only Asp11 in HisF but also the equivalent Asp8 residues in HisA and HisAF essential for PRA isomerization [36, 54].
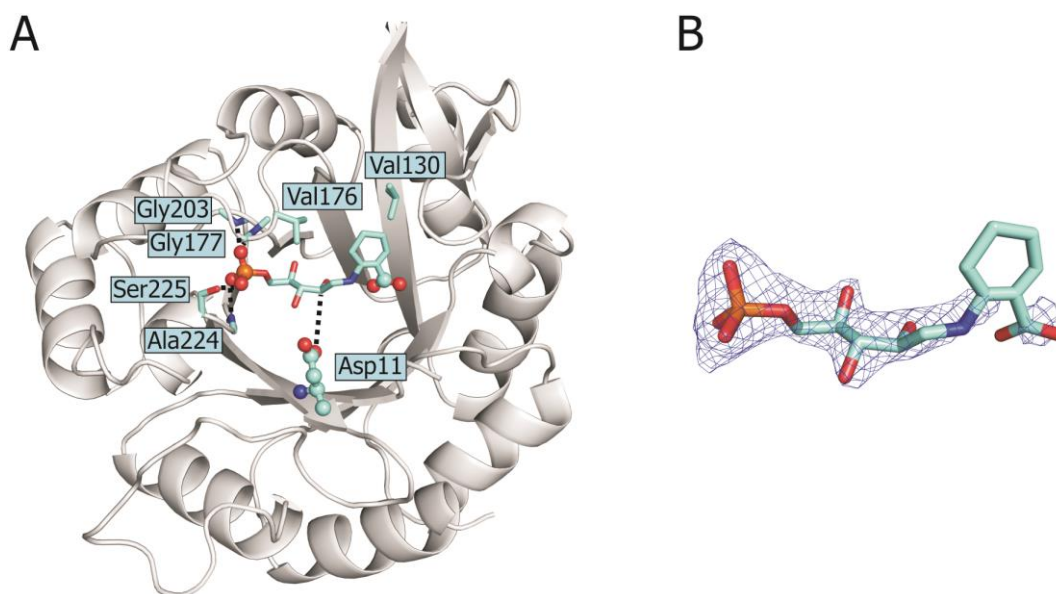
**Figure 5. Crystal structure of HisF_D130V+D176V with bound product analogue rCdRP.**
A: A ribbon representation of HisF_D130V+D176V is shown in grey. A view on the active site reveals that rCdRP is anchored in the C-terminal phosphate binding site, which is composed of the backbone atoms of Gly177, Gly203, Ala224 and the side chain of Ser225 (interactions are indicated by dashed lines). Consequently, Asp11 is properly positioned to act as a base (illustrated by a dashed line). Due to the introduced residues Val130 and Val176, the only remaining candidates for the catalytic acid are Asp11 and the phosphate and carboxylate part of rCdRP (shown as spheres). B: The omit electron density (F_o–F_c) for the ligand rCdRP is contoured at 1.5σ. PDB ID: 4ewn. Taken from *Figure 4*, *publication A*.

A corresponding general acid at the opposite barrel side, however, is looked for in vain. Instead, the crystal structure affords three alternative acid candidates: the already mentioned Asp11 as well as the phosphate and carboxylate group of the ligand (Figure 5A). If Asp11 were to play a dual role in catalysis, PRA would have to adopt a position which allows for the initial protonation of its furanose ring oxygen. In the appropriate binding mode, the anthranilate moiety of PRA points down the central β-barrel – a geometry, though, which was shown to be energetically highly unfavorable as it leads to the ejection of PRA from the active site of HisF_D130V+D176V during MD simulations (data not shown). The phosphate group of rCdRP, in turn, is tightly anchored in the phosphate binding pocket via several hydrogen bond interactions and thus seems to be rather rigid (Figure 5A). The same applies to the adjacent hydroxyl group, which corresponds to the ring-oxygen in PRA. Accordingly, a high electron density is observed for this ligand part (Figure 5B), rendering the phosphate group unlikely to be involved in catalysis. In contrast, the remaining acid candidate, which is the anthranilate moiety of the ligand, exhibits a low electron density. The accompanying high grade of conformational flexibility might facilitate an internal proton transfer to the furanose ring oxygen, which would then constitute a case of

"substrate-assisted catalysis" (SAC) [63]. SAC has so far been discovered in enzymes like GTPases, type II restriction endonucleases, serine proteases and aminoacyl-tRNA synthetases [63-66]. Indeed, SAC has been considered before to play a role in the catalytic mechanism of artificially designed PRAI activity [36, 67].

A straightforward approach to experimentally verify the hypothesis of SAC would be to remove the putative functional group of the substrate in order to impair catalysis [65]. In the present case, the carboxylate of PRA could be replaced by a non-reactive amide group. However, attempts to enzymatically synthesize phosphoribosylanthranilic amide from anthranilic amide and phosphoribosyl-α1-pyrophosphate by anthranilate phosphoribosyl transferase failed, which led to a focus on computational investigations, namely MD simulations and QM/MM calculations. MD simulations have been successfully applied to differentiate active from inactive enzyme designs [24, 68] and are a valuable tool for identifying mechanistically relevant conformations [69, 70]. Regarding biological systems, QM/MM calculations are the method of choice to probe proposed reaction mechanisms and get detailed information about the underlying activation barriers [71, 72]. Hence, a combination of both techniques is well-suited to elucidate enzyme mechanisms [69, 73] and should further substantiate the postulation that PRA is involved in its own conversion.

If the anthranilic acid moiety of PRA indeed initiates the Amadori rearrangement by protonation of the furanose ring oxygen, two requirements have to be fulfilled: the proton transfer must be geometrically feasible and the pK_a of anthranilic acid has to be in the right range to act catalytically. In order to examine both prerequisites, MD simulations were performed with the variants HisA_D127V, HisA_D169V and HisA_D127V+D169V, which feature higher catalytic activities than the respective HisF and HisAF mutants. Furthermore, wild-type HisA served as a reference. For all molecular simulations, two different binding modes of PRA were assumed (Figure 6A): in binding mode 1, the carboxylic acid group points towards the furanose ring oxygen and thus should be able to transfer a proton to the sugar moiety, at least with the assistance of a bridging water molecule. On the other hand, the anthranilic acid group is rotated through 180° in binding mode 2, resulting in a large spatial separation of the reactive substrate parts. Here, an internal proton transfer is considered to be highly unlikely. As judged by the resulting binding energies, HisA wild-type does not differentiate between mode 1 and 2 (Figure 6B). However, all PRAI active HisA variants clearly favor the productive binding mode 1 over the non-reactive binding mode 2 where the strongest preference can be observed for the single mutant HisA_D127V and for HisA_D127V+D169V. Furthermore, it had to be checked
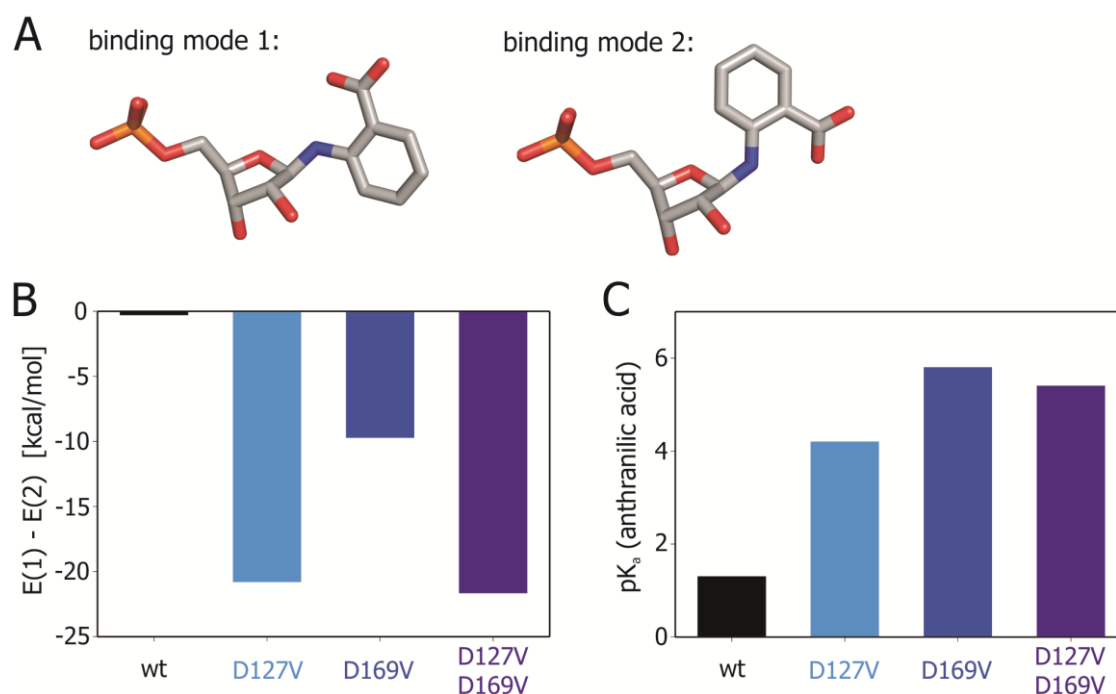
Figure 6. MD simulations of HisA, HisA_D127V, HisA_D169V and HisA_D127V+D169V with bound PRA.

A: Two binding modes of PRA with opposing orientations of the anthranilic acid moiety were considered. B: Averaged energies of binding mode 1 (E(1)) and binding mode 2 (E(2)) were subtracted and differences are shown as bars for all proteins. The more negative the energy the stronger the preference for binding mode 1. At least three independent MD simulations were performed, mean energy values were determined and subsequently averaged for each mode and protein. A t-test ensured that the mean energy values attributed to the two modes are different with statistical significance for all three HisA variants ($P \leq 0.029$). C: For the productive binding mode 1, pK_a values of anthranilic acid were determined by PROPKA [74]. PDB IDs of used structures: HisA wild-type and HisA_D169V (modeled onto the wild-type scaffold): 1qo2; HisA_D127V: 2cff; HisA_D127V+D169V: 2w79. Taken from *Figure 5*, *publication A*.

whether the pK_a values of the anthranilic acid moiety allow for repeated proton donation and acceptance. To this end, the structures obtained with PRA bound in mode 1 were used to determine the respective pK_a values by PROPKA [74]. Since anthranilic acid exhibits a pK_a of only 2.05 in aqueous solution, its acidic strength needs to be significantly lowered when bound to the protein cavities. While in the background of HisA wild-type an even lower pK_a of 1.3 is predicted, the introduced mutations seem to generate a more hydrophobic environment and therefore indeed cause the desired pK_a-shift: the pK_a values range between 4.2 and 5.8, with the largest changes observed for HisA_D169V and for HisA_D127V+D169V (Figure 6C). Therefore, also the second requirement for catalysis is met and the MD data are in line with the hypothesis of SAC. Moreover, the simulations suggest that the single point mutations D127V and D169V contribute to catalysis in a different manner: D127V mainly enables the positive discrimination of binding mode 1 (Figure 6B), whereas

D169V primarily elevates the pK$_a$ value of the anthranilic acid moiety (Figure 6C). The double mutant HisA_D127V+D169V combines both attributes and consistently displays the highest catalytic efficiency. Analogously conducted MD simulations with HisF_D130V+D176V further reinforce the obtained results, as the activating exchanges also induce a distinct preference for binding mode 1 and a marked upshift of the pK$_a$ of anthranilate (*Figure S4, publication A*). Consequently, SAC seems to be the basis for PRA isomerization on the scaffolds of both HisA and HisF.

Finally, QM/MM calculations should proof the suggested catalytic mechanism and reveal energetics along the reaction pathway [69, 71-73]. Starting up with PRA bound to HisA_D127V+D169V in mode 1, a direct proton transfer comes along with a virtually too high energy barrier of more than 50 kcal/mol. As observed for other enzymatically catalyzed substrate protonations, however, a bridging water molecule is able to mediate the transfer process [75, 76] (Figure 7). Two steps are therefore neces-
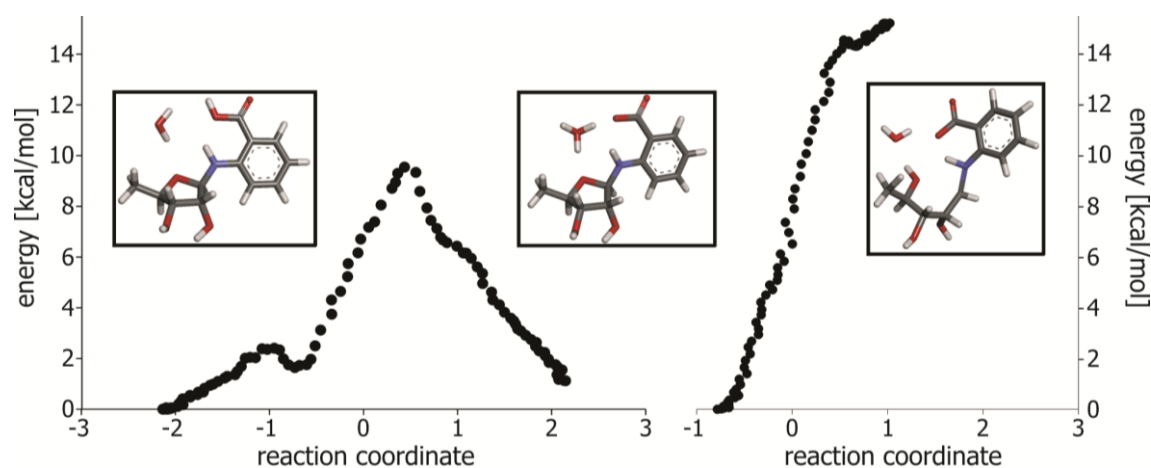
Figure 7. QM/MM calculations of HisA_D127V+D169V with PRA bound in mode 1.
Energy barriers were calculated for a water-mediated proton transfer within PRA. The reaction process includes two separate steps: beginning with PRA in binding mode 1, the carboxylic acid group protonates the bridging water molecule (left reaction coordinate). Subsequently, the generated hydronium ion is able to donate a proton to the furanose ring-oxygen (right reaction coordinate). The three involved reaction states are shown as insets. PDB ID of HisA_D127V+D169V: 2w79. Taken from *Figure 6*, *publication A*.

sary to open the sugar moiety of PRA: the anthranilic acid moiety initially donates a proton to the adjacent water molecule and thus generates a hydronium ion, which is then capable of protonating the furanose ring oxygen. The associated energy costs amount to 10 kcal/mol and 15 kcal/mol, respectively, and are thus in conformity with the energetic demands of known water-mediated proton transfers [75, 76]. Taken together, the QM/MM calculations confirm that the anthranilic acid moiety of PRA can

take over the task of the general acid from an energetic point of view and underpin the mechanism of SAC.

In order to catalyze the Amadori rearrangement of PRA, an enzyme basically requires two appropriate positioned residues acting as general acid and base, respectively [38]. However, as PRA itself brings along a carboxylic acid group which probably is capable of slipping into the role of the catalytic acid, the isomerization of PRA only demands a single base and thus follows a rather simple reaction mechanism. Accordingly, as mentioned in the introduction, this particular reaction could readily be established on various scaffolds, including the artificially designed and initially inert proteins HisAF and Sym2 [36, 37] as well as the enzyme glutamine phosphoribosylpyrophosphate amidotransferase (PurF), which is unrelated to TrpF in sequence, structure and catalytic mechanism [21]. PurF only exhibits a preexisting molecular recognition site for the phosphoribosyl moiety of PRA and an aspartate residue properly positioned to take over the part of the general base, but no obvious acid candidate could be identified. It is hence plausible to assume that the PRAI activity of PurF is also based on the mechanism of SAC. Furthermore, even the wild-type enzymes TrpF and PriA are not completely dependent on their catalytic acids. When replacing Asp126 (TrpF) and Asp175 (PriA) by asparagines, the remaining catalytic efficiency k_{cat}/K_M^{PRA} is still moderate in the case of TrpF (10^2 $M^{-1}s^{-1}$) [38] and even comparable to the best performing HisA mutant in the background of PriA (10^4 $M^{-1}s^{-1}$) [36, 60]. In contrast, the removal of the catalytic bases (C7A in TrpF and D11N in PriA) completely abolishes TrpF activity [38] and extremely reduces the catalytic efficiency of PriA to 59 $M^{-1}s^{-1}$ [60]. The residual activities of both TrpF_D126N and PriA_D175N may be ascribed to an escape route in which the anthranilate group of PRA functionally replaces the mutated catalytic acid. As this non-native reaction pathway is less demanding than the natural mechanisms of TrpF or PriA, SAC might have been a starting point or functional intermediate in the evolution of enzyme-catalyzed PRA isomerization [63].

2.2 Project B: Characterization of a HisF enzyme from the Paleoarchean era

2.2.1 Introduction and objective

The earth is approximately 4500 million years (4.5 Gyr) old and according to geologic evidence, life began about 3.5 Gyr ago in the Archean era [77]. Biological systems have been evolving ever since and have given rise to unique enzymes with specific and sophisticated catalytic features. These include complicated enzyme mechanisms with multiple reaction and conformational intermediates [78-80] and tight coordination of catalysis in multienzyme complexes due to, for example, substrate channeling and allosteric communication [81, 82]. Furthermore, large protein assemblies are observed in molecular machines such as chaperones or the proteasome, which are involved in protein folding and degradation, respectively [83, 84]. From an evolutionary point of view, it is an intriguing question, when this biological complexity developed [85]. It is widely accepted that the three domains of life – the Archaea, the Bacteria, and the Eucarya – originated from a last universal common ancestor (LUCA) [86]. Yet, the nature of this predecessor has been lively debated on [87, 88]. Based on sequence comparisons, the LUCA was claimed to be a rather simple entity- an assumption that appears intuitively plausible [86, 88, 89]. In contrast, the majority of investigations, also considering three-dimensional protein structures and coevolutionary aspects, are in favor of a complex LUCA resembling extant organisms [87, 90-95]. Direct experimental evidence for this conjecture, though, remained elusive, as molecular fossils are rare and difficult to interpret [77].

Interestingly, Pauling and Zuckerkandl suggested half a century ago that it should be possible to infer ancestral protein sequences from the amino acid compositions of their modern descendants [96]. Meanwhile, a technique called ancestral sequence reconstruction has been established and has enabled the study of extinct proteins, thus linking molecular and evolutionary biology [85, 97-99]. The underlying approach comprises three different steps and is outlined in Figure 8 [97]. Initially, a sequence set of the protein of interest is used to generate a multiple sequence alignment, which in turn provides the basis for the calculation of a phylogenetic tree. As the phylogeny reflects the evolutionary relationship of all extant proteins, every node in the tree corresponds to an individual ancestral sequence. Furthermore, branch lengths are a measure of the time lying between the connected sequences. Consequently, given the overall tree topology, an evolutionary model which takes into account the rates and probabilities of amino acid substitutions can be used to infer all ancestral sequences. Both the construction of the phylogenetic tree and the final calculation of extinct
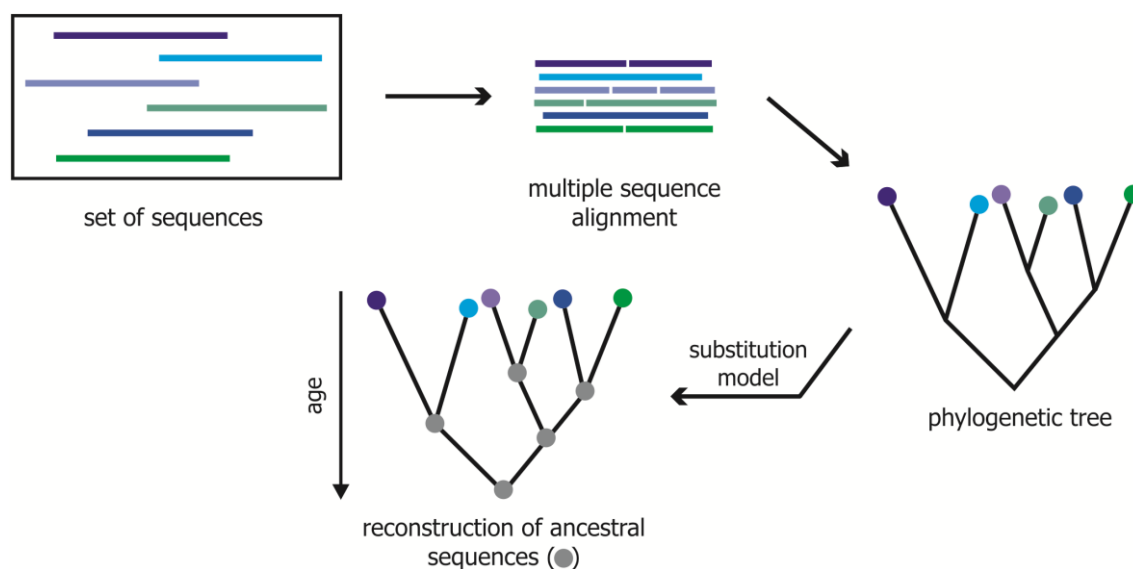
Figure 8. Strategy of ancestral sequence reconstruction.
Starting up with a selected set of sequences, a multiple sequence alignment is generated. Next, the aligned sequences are used to infer a phylogenetic tree whose topology and branch lengths define the evolutionary relationship of all extant sequences. Moreover, every node in the tree corresponds to an ancestral sequence. Finally, the predecessors can be reconstructed by applying a model for the rates and probabilities of amino acid substitution at each site.

sequences is usually based on likelihood methods, hence deducing the most probabilistic phylogeny and ancestors [100-102]. In order to derive an accurate design, the whole reconstruction process needs to be performed with care. For example, when selecting the protein sequences, it should be considered that horizontal gene transfer frequently occurred among different prokaryotes [103]. Moreover, tree nodes should exhibit a reliable statistical confidence and the overall tree topology should resemble the universal consensus tree of life [103, 104]. Regarding the underlying evolutionary model of the final reconstruction step, a variable substitution rate along lineages and the differentiation of distinct sequence sites in their substitution probabilities have proven beneficial [102, 105].

So far, various proteins have been resurrected, including the elongation factor Tu [106, 107], several visual [108-110] and fluorescent pigments [111, 112], and plenty of steroid hormone receptors [113-119]. Reconstructions also brought about functional enzymatic ancestors of alcohol dehydrogenases [120], glucosidases [121], nucleosid diphosphate kinases [122] or methyltransferases [123]. The generated biomolecules could be used to address diverse questions about the ecological features of early life [105-107, 122, 124] and revealed crucial aspects of molecular evolution. Thus, light was shed upon mechanisms of gene duplication [121], the role of epistatic mutations in the course of evolution [110, 111, 114, 118], and the development of ligand and substrate specifity [113, 115-117, 119, 123, 125]. Recently, formation of the hexameric transmem-

brane ring of the fungal V-ATPase proton pump was investigated [126]. It could be demonstrated that the composition of three paralogous protein subunits, specifically found in fungi, emerged from a simpler two-component system via gene duplication, gene loss and subsequent functional adaptation hundreds of millions of years ago. However, in order to assess the degree of complexity already present at evolutionary stages preceding the specification into the three domains of life, reconstructions have to go further back in time. Along these lines, a thioredoxin enzyme being at least 3.5 Gyr old was resurrected and the mechanism of disulfide bond reduction was examined [127]. Strikingly, the ancient enzyme acted not only by means of a simple nucleophilic bimolecular mechanism, but also a single electron transfer mechanism and a substrate binding and rearrangement reaction were observed. Thus, the complex reduction chemistry of modern thioredoxin enzymes had already been established by 3.5 Gyr ago.

It has been assumed that primary metabolic pathways like histidine biosynthesis had been assembled before the LUCA appeared [11, 128]. Consequently, an appropriate set of protein sequences should allow for the reconstruction of metabolic enzymes resembling those of the LUCA. A respective approach was recently reported for the HisF subunit of ImGP synthase [104]. As the synthase HisF forms an obligate heterodimer with the glutaminase HisH in prokaryotes [41, 129], a concatenation of both sequences was chosen as a basis for the reconstruction process to increase the phylogenetic signal. The resulting multiple sequence alignment contained 87 sequences, including the clades Crenarchaeota, Actinobacteria, Chlorobi, Cyanobacteria, Firmicutes, Proteobacteria, and Thermotogae (*Figure S1, publication B*). On this basis, a phylogenetic tree was calculated and rooted between the superkingdoms Archaea and Bacteria. The protein sequence corresponding to the root was reconstructed and will be termed LUCA-HisF hereafter. LUCA-HisF is approximately 3.5 Gyr old and thus among the most ancient predecessors inferred so far [106, 125, 127]. According to BLAST [130], LUCA-HisF differs in 55 of 250 amino acids (22 %) from its closest living relative, which is HisF from *Thermovibrio ammonificans*. Strikingly, this bacterium was isolated from a deep-sea hydrothermal vent [131], a geochemical habitat whose unique environment has repeatedly been associated with the origin of life [77, 132].

In modern HisF enzymes, binding of the substrate PRFAR triggers an allosteric signal and hence enhances the catalytic efficiency of the interacting glutaminase subunit HisH by up to several hundredfold [45, 49]. Furthermore, the ammonia produced by HisH is transported to the active site of HisF through a molecular channel spanning the whole lengths of the $(\beta\alpha)_8$-barrel fold of HisF (*Figure 1, publication B*) [42, 46, 49]. These exceptional catalytic features provide an ideal foundation for experimentally testing the grade of sophistication reached in the LUCA era. Along

these lines, the physical and catalytic properties of LUCA-HisF were investigated within the scope of this project.

2.2.2 Summary and discussion

The gene coding for LUCA-HisF was optimized for expression in *Escherichia coli*, synthesized (Life Technologies), and cloned into an expression plasmid. Production in *E. coli* yielded predominantly soluble protein, which was purified in a three-step process via ion exchange chromatography, ammonia sulfate precipitation, and size exclusion chromatography. Notably, about 30 mg of protein were obtained per liter of culture. Subsequently, LUCA-HisF was characterized by means of several biophysical methods to assure its structural and functional integrity.

Initially, LUCA-HisF was crystallized using the hanging drop vapor diffusion method and its structure was resolved at a resolution of 1.48 Å. For this purpose, a homology model of LUCA-HisF was built on the structure of HisF from *T. maritima* (*tm*HisF) (PDB ID: 1thf) and served as a template for molecular replacement [133]. As expected, LUCA-HisF adopts a canonical $(\beta\alpha)_8$-barrel fold, comparable to known structures of the HisF enzymes from *Pyrobaculum aerophilum* (PDB ID 1h5y), *Thermus thermophilus* (PDB ID 1ka9), and *T. maritima*. Thus, when aligning each of these tertiary structures with the structure of LUCA-HisF through STAMP [134], low overall root-mean-square-deviation (rmsd) values ranging from 1.14 Å to 1.43 Å were obtained. Moreover, in accordance with the postulated evolution from a half-barrel precursor via gene duplication and fusion [32], LUCA-HisF displays a clear twofold symmetry. Superposition of its N-terminal $(\beta\alpha)_{1-4}$ and C-terminal $(\beta\alpha)_{5-8}$ module resulted in a rmsd value of merely 1.68 Å. However, similar values of 1.27 Å to 1.69 Å were obtained for the corresponding analyses of the three extant HisF proteins, indicating that the main structural evolution of the $(\beta\alpha)_8$-barrel fold had been completed 3.5 Gyr ago. Among the 87 protein sequences used for reconstruction, 78 of 250 residues are less than 50 % conserved. While these variable positions are primarily located at the rim of the barrel, the functionally important activity and stability face are widely unaltered (Figure 9). Hence, LUCA-HisF contains indispensable amino acids like the catalytically acting aspartate residues 11 and 130, lying at the C-terminal ends of β-strands 1 and 5, respectively (Figure 9A) [41], and the stabilizing salt-bridge cluster at the opposite side of the barrel which forms the gate to the ammonia channel (Figure 9B) [42, 49, 135]. Moreover, the two co-crystallized phosphate ions occupy the N- and C-terminal phosphate binding site and thus mimic the terminal phosphate groups of the substrate PRFAR.
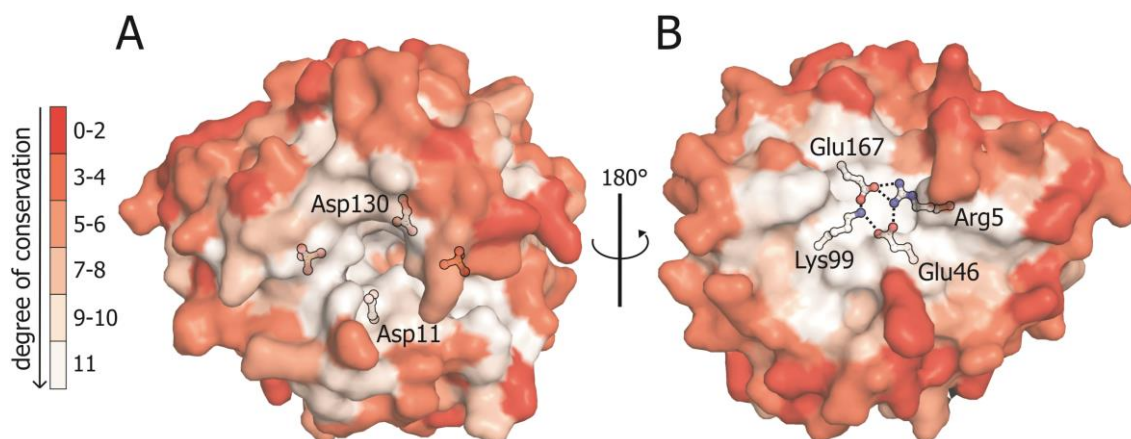
Figure 9. Crystal structure of LUCA-HisF (PDB ID: 4evz).
The surface representations of LUCA-HisF are colored according to the amino-acid conservation deduced from a multiple sequence alignment of all protein sequences used for resurrection (red: variable; white: strictly conserved). The conservation values (0-11) were determined through Jalview [136]. A: View on the activity face. Side chains of the catalytically essential residues Asp11 and Asp130 as well as bound phosphate ions are shown as sticks. B: Stability face and ammonia tunnel gate. Arg5, Glu46, Lys99 and Glu167 form a stabilizing salt bridge cluster at the bottom of the barrel, which defines the entrance to the ammonia channel (side chains are depicted as sticks; electrostatic interactions are indicated by dashed lines). Taken from *Figure 2*, *publication B*.

After closely examining the crystal structure of LUCA-HisF, its thermal and conformational stability were determined. Along these lines, the heat-induced loss of secondary structure was probed by a decrease in the far-UV circular dichroism (CD) signal, and overall thermal unfolding was monitored by differential scanning calorimetry (DSC). The combination of both methods showed that LUCA-HisF thermally unfolds in an irreversible two-step process with apparent transition midpoints of about 70 °C and 100 °C (*Figure S3A+B*, *publication B*). Although the observed transitions could not be assigned to particular structural changes, it can be concluded that LUCA-HisF features a high resistance to heat. Next, the thermodynamic stability of LUCA-HisF was analyzed by guanidinium chloride-(GdmCl)-induced equilibrium unfolding/refolding transitions (Figure 10). Starting either with folded (dissolved in buffer) or unfolded (dissolved in 6 M GdmCl) protein, the concentrations of GdmCl were adjusted between 0 and 5 M, and after equilibration, the loss or gain of tertiary structure was quantified by Trp/Tyr fluorescence. The data points for unfolding and refolding coincide, demonstrating that GdmCl-induced unfolding is reversible. Moreover, both processes are adequately described by the two-state model, enabling thermodynamic analysis to yield the Gibbs free energy of unfolding (ΔG_D), the concentration of GdmCl required to denature half of the protein ($[D]_{1/2}$), and the cooperativity of unfolding (m-value) [137]. LUCA-HisF exhibits a
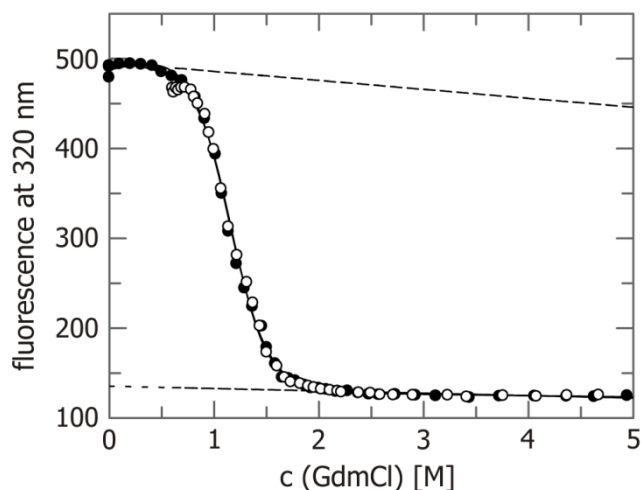
Figure 10. GdmCl-induced equilibrium unfolding/refolding transitions of LUCA-HisF.
Unfolding transitions (black circles) were started with folded protein, whereas protein used for refolding transitions (white circles) was previously unfolded in 6 M GdmCl. After adjusting to the respective GdmCl concentration and equilibration, the loss or gain of tertiary structure was monitored by protein fluorescence at 320 nm (excitation at 280 nm). Data were analyzed by the two-state model [137], where the corresponding fits are represented by continuous lines and the baselines for the pure native and unfolded state are indicated by dashed lines. The resulting thermodynamic parameters are given in *Table S2*, *publication B*. Taken from *Figure 3*, *publication B*.

moderate ΔG_D value of 18 kJ mol^{-1} and a transition midpoint $[D]_{1/2}$ at 1.2 M GdmCl and therefore is less stable than *tm*HisF [138]. However, more importantly, LUCA-HisF unfolds with a high cooperativity of 15 kJ mol^{-1} M^{-1}, which is comparable to the m-value of *tm*HisF and thus is indicative of a compact structure [138]. Moreover, subsequently performed folding kinetics demonstrate that LUCA-HisF and *tm*HisF share a common sequential folding mechanism, including the rate-limiting step (*Figures S4B+S5+S6*, *publication B*). This mechanism also holds for the artificially designed proteins Sym1 and Sym2, which should temporally precede LUCA-HisF due to their marked internal twofold rotational symmetry [34]. Hence, also the folding mechanism of the $(\beta\alpha)_8$-barrel structure had apparently already been evolved in the LUCA era.

The ultimate validation of a resurrected enzyme is the verification of its catalytic function. In the case of HisF, enzymatic activity can be measured spectrophotome-trically in the presence of saturating concentrations of externally added ammonia (*ammonia-dependent cyclase activity*) [44]. Surprisingly, the catalytic efficiency (k_{cat}/K_M^{PRFAR}) of LUCA-HisF is as high as the k_{cat}/K_M^{PRFAR} values observed for extant ImGP synthase subunits, mainly due to its low Michaelis constant K_M^{PRFAR} (Table 1).

Table 1. Steady state kinetic constants of LUCA-HisF/*zm*HisH in comparison to extant ImGP synthase pairs.

Ammonia-dependent cyclase activity

	k_{cat} [s^{-1}]	K_M^{PRFAR} [μM]	k_{cat}/K_M^{PRFAR} [M^{-1}s^{-1}]
LUCA-HisF	0.078 (± 0.003)	0.29 (± 0.04)	2.8 (± 0.3) x 10^5
*tm*HisF[a]	1.2	3.6	3.3 x 10^5
*ec*HisF[b]	5.7	21	2.7 x 10^5
*sc*HisF[c]	0.8	55	1.5 x 10^4

Glutamine-dependent cyclase activity

	k_{cat} [s^{-1}]	K_M^{PRFAR} [μM]	k_{cat}/K_M^{PRFAR} [M^{-1}s^{-1}]
LUCA-HisF/ *zm*HisH	0.058 (± 0.006)	0.36 (± 0.07)	1.6 (± 0.3) x 10^5
*tm*HisF/*tm*HisH[a]	1.1	2.0	5.5 x 10^5
*ec*HisF/*ec*HisH[b]	8.5	1.5	5.7 x 10^6
*sc*HisF/*sc*HisH[c]	5.4	5.0	1.2 x 10^6

Glutaminase activity

	k_{cat} [s^{-1}]	K_M^{Gln} [mM]	k_{cat}/K_M^{Gln} [M^{-1}s^{-1}]
LUCA-HisF/ *zm*HisH	0.21 (± 0.03)	1.9 (± 0.9)	1.2 (± 0.3) x 10^2
*tm*HisF/*tm*HisH[a]	0.1	0.8	1.3 x 10^2
*ec*HisF/*ec*HisH[b]	9.1	0.2	3.8 x 10^4
*sc*HisF/*sc*HisH[c]	6.9	1.8	3.8 x 10^3

Stimulation of glutaminase activity

	k_{cat} [s^{-1}] (without ProFAR)	k_{cat} [s^{-1}] (ProFAR saturated)	Stimulation factor[d]
LUCA-HisF/ *zm*HisH	3.85 (± 0.04) x 10^{-2}	0.483 (± 0.006)	13
*tm*HisF/*tm*HisH[a]	3.3 x 10^{-4}	0.1	303
*ec*HisF/*ec*HisH[b]	7.0 x 10^{-2}	2.7	39
*sc*HisF/*sc*HisH[c]	5.5 x 10^{-3}	6.9	1255

[a]: Data for ImGP synthase from *T. maritima* (*tm*) were taken from [49].

[b]: Data for ImGP synthase from *E. coli* (*ec*) were taken from [44].

[c]: Data for ImGP synthase from *S. cerevisiae* (*sc*) were taken from [45].

[d]: The stimulation factor is the quotient k_{cat} (ProFAR saturated)/k_{cat} (without ProFAR).

The activity of LUCA-HisF also was confirmed by *in vivo* complementation assays with an auxotrophic *E. coli* strain lacking the *his*F gene. Δ*his*F cells are not able to produce histidine and thus do not grow when plated on minimal medium without amino acids. However, after transformation with a plasmid containing the LUCA-HisF gene, visible colonies are formed within 24 hours. It is widely accepted that enzyme precursors possessed a broad substrate specificity [139]. Accordingly, recently resurrected lacta-mases process an expanded substrate repertoire and support the notion that modern enzyme specialists evolved from ancient generalists [125]. Therefore, LUCA-HisF was tested for its ability to catalyze the reactions of the sugar isomerases HisA and TrpF. As outlined in chapter 2.1, HisA precedes HisF in histidine biosynthesis and both enzymes are strongly related in terms of sequence and structure [32]. As to the PRA isomerase TrpF, a single amino acid substitution is sufficient to establish PRAI activity on HisF, suggesting an evolutionary relationship [50, 54]. Despite these similarities, both *in vivo* complementation assays and *in vitro* spectroscopical measurements demonstrated that LUCA-HisF catalyzes neither the HisA nor the TrpF reaction.

The catalytic sophistication of the synthase HisF can only completely be assessed in conjunction with an interacting glutaminase HisH. It is plausible to assume that LUCA-HisF is able to interact with extant glutaminases, since large sections of its protein interface, which is located at the stability face (*Figure 1, publication B*), are conserved (Figure 9A). Indeed, fluorescence titration revealed an interaction between LUCA-HisF and HisH from *Zymomonas mobilis* (*zm*HisH). As *zm*HisH contains a tryptophan residue at the protein interface, stepwise addition of a tryptophan-free variant of LUCA-HisF gradually shielded this particular residue from the solvent and thus induced a hypsochromic shift of the fluorescence signal [41] (Figure 11A). Analysis of the resulting titration curve showed that LUCA-HisF and *zm*HisH form a stoichiometric and affine complex, whose thermodynamic dissociation constant K_d is 113 nM.

With a glutaminase partner in hand, the synthase activity of LUCA-HisF could be measured in the presence of *zm*HisH and saturating concentrations of glutamine (*glutamine-dependent cyclase activity*). Notably, the determined k_{cat} and K_M^{PRFAR} values compare well with the catalytic parameters obtained for the ammonia-dependent reaction (Table 1), suggesting that externally added ammonia can equivalently be replaced by ammonia produced by hydrolysis of glutamine. As a consequence, the active sites of *zm*HisH and LUCA-HisF seem to be connected by an extended ammonia channel, as known from extant HisF/HisH pairs [42, 43, 49]. Moreover, in the presence of LUCA-HisF and saturating amounts of the substrate analogue ProFAR,
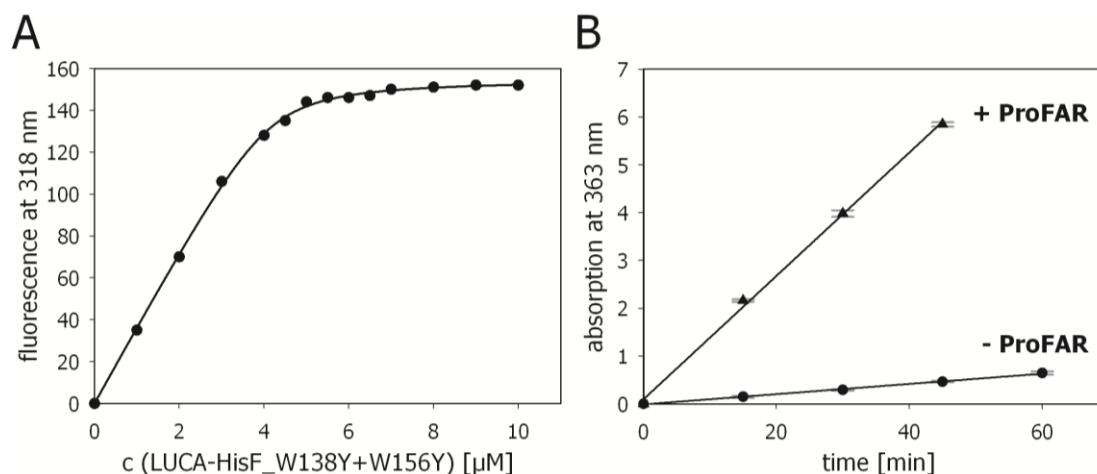
Figure 11. Interaction and glutaminase activation of LUCA-HisF and *zm*HisH.
A: The fluorescence titration curve was obtained by gradually adding a tryptophan-free variant of LUCA-HisF to 7 μM *zm*HisH and monitoring the emission change at 318 nm after excitation at 295 nm. The solid line represents a quadratic fit of the data points. Three independent titrations resulted in a K_d value of 113 ± 1 nM. B: Allosteric activation of *zm*HisH was determined in a discontinuous photometric assay at 363 nm. In identical setups, glutamine was incubated with *zm*HisH/LUCA-HisF at 25 °C both in the absence (circles) and presence (triangles) of ProFAR. Mean values and standard deviations of triplicate measurements are shown. As listed in Table 1, glutaminase activity of *zm*HisH is enhanced 13-fold in the presence of ProFAR. Taken from *Figure 4, publication B*.

*zm*HisH processes glutamine nearly as efficient as other glutaminases (Table 1). In sum, these findings clearly demonstrate that LUCA-HisF forms a functional ImGP synthase complex with *zm*HisH. However, the most remarkable characteristic of extant HisF enzymes, namely the ability to allosterically increase the glutaminase activity upon substrate binding, still had to be analyzed. In principle, glutamine hydrolysis is monitored in a coupled enzyme assay with glutamate dehydrogenase, which oxidizes glutamate with the assistance of nicotinamide adenine dinucleotide (NAD^+) or acetyl-pyridine adenine dinucleotide ($APAD^+$) and thus enables photometric detection at 340 nm and 363 nm, respectively [41, 44]. Besides the substrate PRFAR and its analogue ProFAR, however, the coenzymes NAD^+ and $APAD^+$ were also shown to stimulate the glutaminase activity [49]. On this account, the production and quantification of glutamate had to be separated in a discontinuous assay, in order to determine the extent of glutaminase activation caused by the synthase with bound ProFAR [49]. Most strikingly, the corresponding measurements revealed that glutamine hydrolysis of LUCA-HisF/*zm*HisH is 13 times more efficient in the presence than in the absence of ProFAR (Figure 11B; Table 1). The magnitude of stimulation is comparably high as the allosteric activation reported for the ImGP synthase complex of *E. coli* and definitely remarkable for a non-natural ImGP synthase pair, even if

significantly higher stimulation effects were observed for the complexes from *T. maritima* and *S. cerevisiae* (Table 1).

Although 3.5 Gyr old, LUCA-HisF constitutes an intact $(\beta\alpha)_8$-barrel protein, whose folding mechanism and overall structure is in no way inferior to modern HisF enzymes. In contrast to previously resurrected lactamases [125], LUCA-HisF does not show substrate promiscuity, but is a highly specific enzyme. Furthermore, together with the glutaminase *zm*HisH, LUCA-HisF forms a fully functional complex. Hence, LUCA-HisF exhibits all elements required for ammonia channeling and allosteric communication, indicating that these sophisticated catalytic concepts had already been evolved in the LUCA era and thus suggesting a rather complex metabolism at this early point in time. This notion is perfectly in line with the results obtained for resurrected thioredoxin enzymes [127] and also complies with the majority of theoretical investigations [87, 90-95].

The postulated existence of a functional ImGP synthase complex about 3.5 Gyr ago relies on the interplay between a resurrected synthase and an extant glutaminase. More direct evidence could be obtained, if a HisH protein from the LUCA era was used instead of *zm*HisH. As a concatenation of HisF and HisH sequences provided the basis for the reconstruction of LUCA-HisF [104], a putative HisH predecessor was inferred in an analogous manner. However, the protein was found to be completely insoluble under several expression conditions and even purification from inclusion bodies failed. The reconstruction was probably hampered by the low sequence conservation of extant HisH enzymes: In comparison to HisF proteins, twice as many amino acids are less than 50 % conserved. Even more problematically, HisH sequences significantly differ in their overall size, leading to systematic errors in the multiple sequence alignment and influences the accuracy of the final reconstruction step [140]. This problem could be circumvented by using a software called PRANK, which is superior to traditional alignment algorithms, as phylogeny is considered in the placement of gaps [140, 141]. Accordingly, PRANK was used to generate a more accurate multiple sequence alignment and to compute a protein sequence for LUCA-HisH (*Figure S2*, *publication B*). After gene synthesis (Life Technologies) and cloning into an expression plasmid, LUCA-HisH was expressed in *E. coli* cells. Although only 21 of its 226 residues are strictly conserved, LUCA-HisH was predominantly found in the soluble fraction and showed a high thermal stability. About 26 mg of the pure protein were obtained after purification via heat denaturation and Ni^{2+} affinity chromatography. The thermo-stability of LUCA-HisH was further proven by heat-induced loss of the far-UV CD signal, resulting in a single transition with a midpoint of about 79 °C (*Figure S3C*, *publication B*). Remarkably, fluorescence titration revealed that LUCA-HisH forms a

stoichiometric complex with LUCA-HisF (Figure 12) – both proteins interact with a very high affinity represented by a K_d value of merely 4 nM.
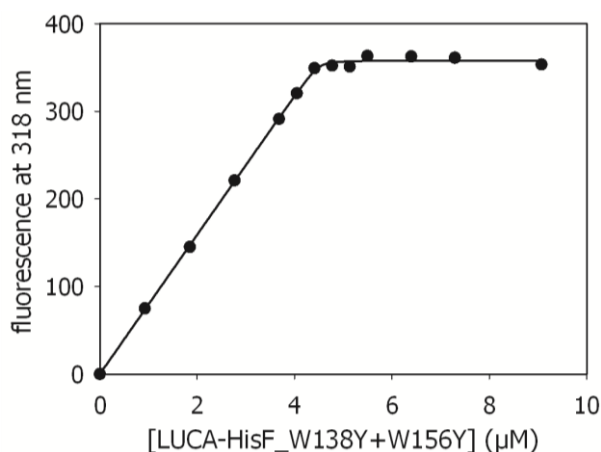
Figure 12. Interaction of LUCA-HisF and LUCA-HisH.
5 μM LUCA-HisH were titrated with LUCA-HisF_W138Y+W156Y and the change in fluorescence emission was monitored at 318 nm (excitation at 295 nm). A quadratic equation was used to fit the data represented by the solid line. Three independent titrations resulted in a K_d value of 4 ± 2 nM.

Despite this strong interaction, no enzymatic turnover could be determined when testing the LUCA-HisF/LUCA-HisH complex for the hydrolysis of glutamine and the *glutamine-dependent cyclase activity*, respectively. Thus, LUCA-HisH is inactive, which is probably to be ascribed to the indicated uncertainties in sequence reconstruction. Especially the numerous gaps observed in the alignment of HisH sequences (*Figure S2*, *publication B*) seem to negatively affect the overall accuracy of reconstruction. In contrast, the sequence lengths of the successfully reconstructed HisF and thioredoxin enzymes are very similar to the respective extant proteins. Therefore, in order to deal correctly with cases as difficult as HisH, the inference of ancestral sequences requires optimization of the placement and reconstruction of insertions and deletions.

2.3 Project C: Light regulation of PriA from *Mycobacterium tuberculosis*

2.3.1 Introduction and objective

The regulation of enzyme activity is a crucial factor for coordinating the vast array of cellular processes. Besides genetic control of enzyme synthesis, catalytic performance can directly be adjusted at the protein level through various regulatory mechanisms [142]. These include limited proteolysis as known from irreversible activation of serine proteases [143] and reversible covalent modifications like adenylation [144] and phosphorylation, which is widely used among prokaryotes and eukaryotes [145]. Furthermore, plenty of enzymes can be activated or inhibited by small metabolites. Most prominently, early catalytic steps of metabolic pathways are often allosterically inhibited by subsequently produced intermediates or by the end-products [146]. This so-called feedback inhibition has, for instance, been described for aspartate transcarbamoylase and anthranilate synthase, which catalyze the first steps of pyrimidine and tryptophan biosynthesis, respectively [147, 148]. In addition, enzymatic activity can be controlled by external stimuli like osmotic pressure, pH-changes, oxygen and light. In general, stimulus-sensitive protein domains are able to transfer the extracellular signal to effector domains such as kinases or phosphodiesterases, thus initiating a cellular response [149, 150].

Artificial control of enzyme activity would enable attractive applications for example in diagnostics and biosensing and hence has been a long-pursued aim [151, 152]. In search of a signal triggering the change in catalytic efficiency, light was proven an ideal tool, as it allows for non-invasive and spatiotemporal regulation of biological activity [153]. Three different strategies were applied to render enzymatic performance light-responsible [154]. Catalytically essential residues were caged with photolabile protecting groups, affording irreversible enzyme activation upon light irradiation [155]. Furthermore, the catalytic function of enzymes was controlled by genetic fusion with naturally occurring photoreceptors [152]. Finally, enzymatic activity was modulated in various ways by chromophores which undergo reversible photochemical transformations and therefore were termed photoswitches [151]. These chemical compounds were either randomly or site-selectively incorporated into enzymes in order to affect the active sites upon irradiation [156-158]. On the other hand, enzymes were embedded in photocontrollable environments like photoresponsive membranes or surfactants in a few cases [159, 160]. Although subjected to marked restrictions [151], these strategies indirectly control catalytic activity and thus dispense with the need for a covalent coupling step between protein and photoswitch. Along these lines, several

approaches focused on the use of photosensitive enzyme inhibitors, which were designed on the basis of known inhibitors [161-164]. Despite numerous attempts to reversibly regulate enzymatic function by light, only few successful examples have been published to date [151].

The bisubstrate-specific enzyme PriA from *Mycobacterium tuberculosis* (*mt*PriA) constitutes a branch-point enzyme in amino acid biosynthesis [39]. *mt*PriA catalyzes two chemically equivalent sugar isomerization reactions within tryptophan and histidine biosynthesis, where the aminoaldoses PRA and ProFAR are converted to the corresponding aminoketoses CdRP and PRFAR (Figure 2). Since humans can produce neither tryptophan nor histidine, *mt*PriA is a potential target for anti-tuberculosis drugs [165, 166]. Due to these unique enzymatic properties, the current project was aimed at constructing light-controllable inhibitors of *mt*PriA. After harnessing available enzyme structures for the design of potential inhibitors, the molecules were synthesized, and photochemically and functionally characterized.

2.3.2 Summary and discussion

Recently solved crystal structures revealed that *mt*PriA adopts a $(\beta\alpha)_8$-barrel fold with two special attributes [40, 60]. On the one hand, the active site comprises two opposite phosphate binding sites, which anchor the bi-phosphorylated substrate ProFAR (Figure 13A). Since high phosphate concentrations were shown to compete with substrate binding in related enzymes [36], potential inhibitors could probably be fixed through terminal phosphate groups. On the other hand, *mt*PriA exhibits an eye-catching twofold rotational symmetry (Figure 13A), which is ascribed to its evolution from a $(\beta\alpha)_4$-half-barrel precursor via gene duplication and fusion [50]. Interestingly, an artificially designed protein with threefold rotational symmetry was co-crystallized with the threefold symmetric molecule tris(hydroxymethyl)aminomethane (Tris) [167]. It therefore seems plausible to assume that compounds containing a twofold axis of symmetry are particularly suited to bind to the active site of *mt*PriA. Merely two kinds of light-controllable scaffolds feature the desired symmetry and were considered for the structural core of *mt*PriA inhibitors: the isoelectronic stilbene [168] and azobenzene switches [169], and diarylethene [170]. Both stilbene and azobenzene undergo reversible *trans*-to-*cis* isomerizations. However, the *cis* isomer of stilbene is prone to an irreversible cyclization/oxidation reaction, rendering the scaffold unsuitable for the photoregulation of enzymes. Azobenzene is the most widely used photoswitch in biological systems [151, 169], although the underlying *trans*-to-*cis* isomerization is both incomplete and thermally reversible [154]. In contrast, irradiation of 1,2-dithienyl-ethene (DTE) usually results in virtually complete photoconversion and thermally
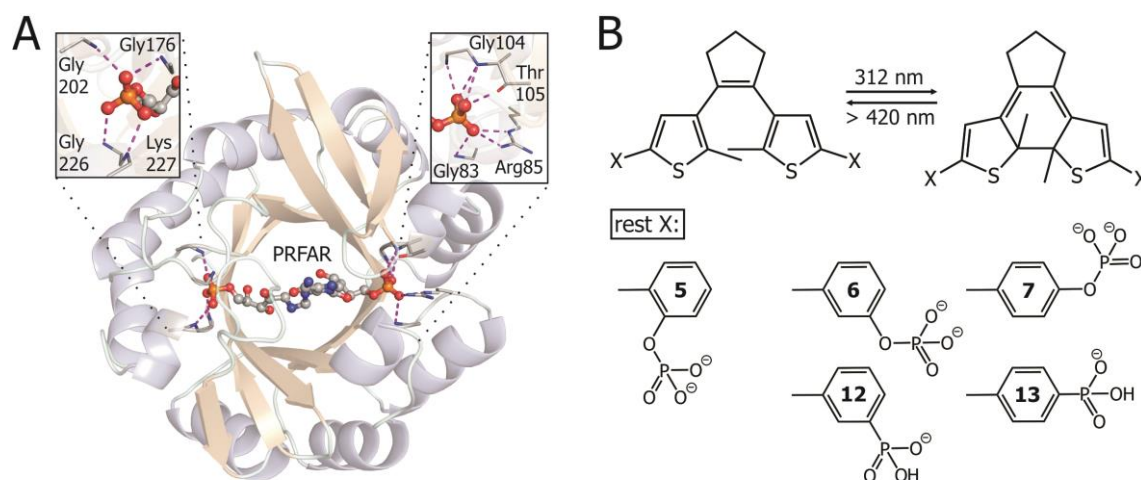
Figure 13. Structure of *mt*PriA and potential inhibitors based on the photoswitchable DTE scaffold.

A: View along the two fold symmetry axis of the $(\beta\alpha)_8$-barrel enzyme *mt*PriA, which is depicted in ribbon representation (PDB ID 3zs4 [60]). The product PRFAR is fixed by the two opposite phosphate binding sites. The coordination of the phosphate groups is enlarged in the insets, where hydrogen bonds are indicated by dashed lines (for clarity, insets are shown in different orientations). Taken from *Figure 1, publication C*. B: The DTE scaffold can reversibly be switched between a ring-open and a ring-closed isomer through irradiation with UV and visible light, respectively. Potential DTE inhibitors of *mt*PriA were equipped with terminal phosphate (compounds **5**, **6**, **7**) and phosphonate (compounds **12**, **13**) groups.

stable photoisomers [162, 170]. Therefore, DTE was chosen as a scaffold and provided with terminal phosphate and chemically similar phosphonate anchors (Figure 13B). Due to the viable synthesis through Suzuki coupling (*Scheme 1, publication C*), all potential inhibitors contain phenyl linkers between the DTE core and the terminal anchor groups (Figure 13B). In order to ensure structural variety, phosphate groups were placed in ortho-, meta-, and para-position (compounds **5**, **6**, and **7**), whereas phosphonate groups were limited to the meta- and para-position (compounds **12** and **13**). Energy minimizations of all compounds yielded distances of the phosphorous atoms between 15.8 and 19.6 Å (*Scheme 2, publication C*), which is in good accordance with the 16.9 Å observed for the *mt*PriA-PRFAR complex (PDB ID 3zs4 [60]). Only the energetically most favorable forms of compound **5** exhibited phosphate distances no longer than 12.3 Å. However, more extended conformers with higher energies were also populated. Hence, judged by their lengths, all designed molecules should be able to inhibit *mt*PriA.

The DTE architecture can be toggled between a flexible, ring-open and a rigid, ring-closed isomer through alternate irradiation with UV and visible light, respectively (Figure 13B) [170]. Thus, the switching properties of the synthesized compounds were examined to verify their photochemical integrity. As all DTE-phosphates and DTE-

phosphonates behaved virtually identical, merely the characterization of compound **6** is described in the following and depicted in Figure 14.
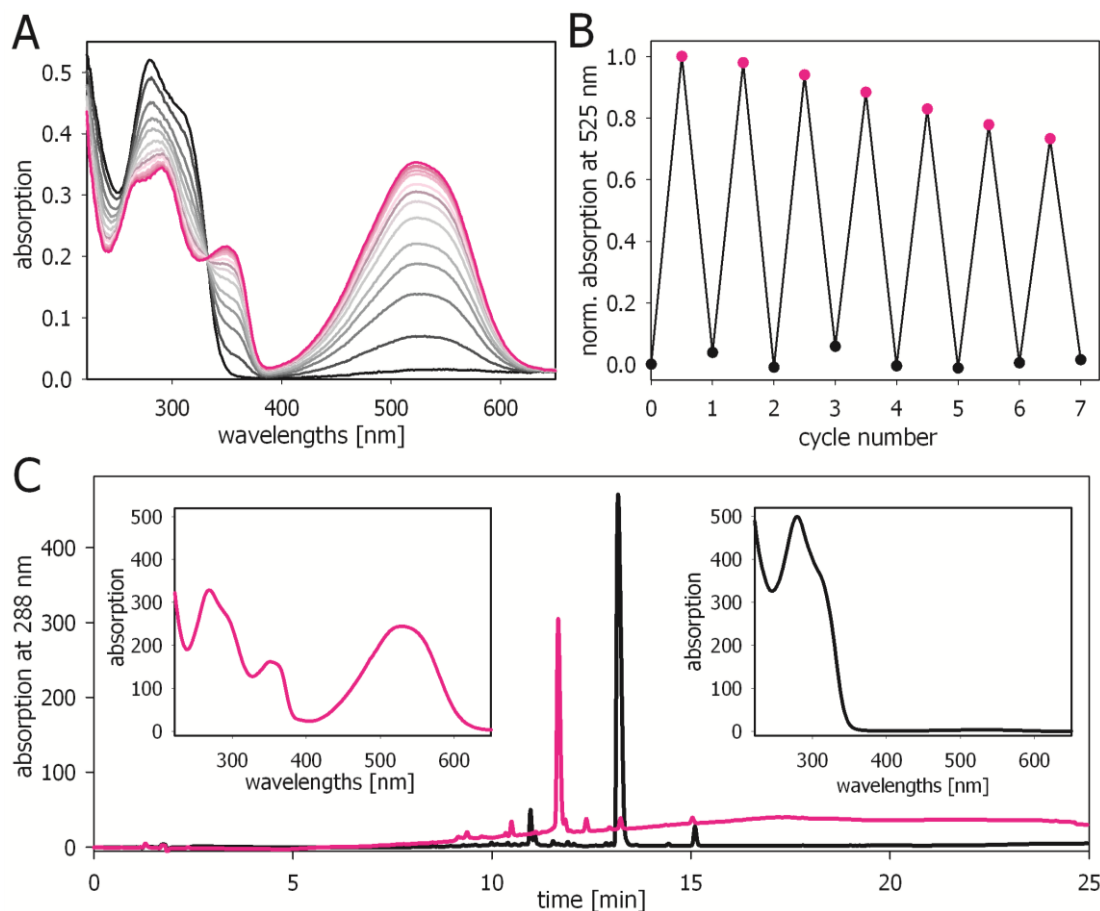
Figure 14. Representative photochemical characterization of compound 6.
A: Change in the absorption spectrum upon irradiation with 312 nm light. The open form of **6** (12.5 μM in 50 mM Tris/acetate pH 8.5) was irradiated in intervals of 2 s from 0 (black) to 30 s (pink). B: Repeated ring-closing/ring-opening cycles. DTE-phosphate **6** (12.5 μM in 50 mM Tris/acetate pH 8.5) was alternately irradiated with UV light (312 nm; 30 s) and visible light (> 420 nm; 15 min) and the absorption change was monitored at 525 nm. The open and closed isomer are depicted as black and pink dots, respectively C: Determination of the photo-stationary state. HPLC chromatograms of compound **6** (100 μM in 50 mM Tris/acetate pH 8.5) without (black) and with (pink) irradiation at 312 nm for 30 s were recorded at 288 nm and superimposed. The respective absorption spectra of the main peaks are shown as insets. According to the peak intensities, 97 % of the closed isomer were formed by irradiation.

When irradiated at 312 nm, meta-phosphate **6** immediately altered its absorption spectrum (Figure 14A). While the peak at 280 nm decreased, new maxima arose at 350 nm and especially at 525 nm, turning the initially colorless solution pink. All spectral changes were completed after 30 s. Subsequently, the original absorption spectrum could be fully restored through irradiation with visible light for 15 min. Moreover, this switching cycle could be repeated several times only leading to slight degradation

(Figure 14B). So as to determine the ratio of open and closed form at the photostationary state, DTE-phosphate **6** was analyzed by HPLC both before and after irradiation with 312 nm light for 30 s (Figure 14C). Since the isomers eluted at clearly different time points, the corresponding peak intensities could be used to determine the amount of closed isomer emerging upon irradiation. Remarkably, 97 % of compound **6** were converted to the ring-closed form. Similar results were obtained for the other photoswitches, thus confirming the magnificent photoconversion yields of the DTE scaffold [170]. Taken together, the described photochemical features are perfectly in line with previously published data on DTE-based inhibitors [162].

Subsequently, the compounds were tested for the ability to inhibit the ProFAR isomerization activity of *mt*PriA, which was monitored spectrophotometrically at 300 nm [44]. After ensuring the stability of all open and closed DTE-switches under assay conditions, steady-state kinetics were performed in the presence of different concentrations of each inhibitor candidate (saturation curves for compound **6** are shown in *Figure S4*, *publication C*). While the turnover number practically remained constant throughout all measurements, the K_M^{ProFAR} value consistently rose with increasing amounts of the respective photoswitch (Table 2). Thus, each compound inhibited *mt*PriA in a competitive fashion, proving the efficiency of the design concept. Having determined the K_M^{ProFAR} value in the absence of inhibitor, the apparent K_M^{ProFAR} values could be used to calculate the inhibitory constants K_i with *Formula S1*, *publication C* (Table 2). All deduced K_i values are in the low micromolar range and hence compare well with or even surpass the Michaelis constant for the natural substrate ProFAR. Merely considering the distances of the phosphorous atoms, the open isomers of the meta-substituted compounds **6** and **12** come closest to the 16.9 Å observed for the *mt*PriA-PRFAR complex (PDB ID 3zs4 [60]; *Scheme 2*, *publication C*) and thus should exhibit the highest inhibitory potential. Indeed, their K_i values are as low as 0.55 μM and 1.6 μM, respectively. When testing the corresponding ring-closed isomers, the inhibition activities decreased threefold in case of meta-phosphonate **12** and even eightfold for meta-phosphate **6**. In contrast, the open and closed isomer of ortho-phosphate **5** affected *mt*PriA activity to a similar extent. Although the same result was found for para-phosphate **7**, the introduction of a phosphonate moiety in para-position (compound **13**) gave rise to a threefold difference in the K_i values of the open and closed form.

Table 2. Steady-state kinetic constants for the ProFAR isomerization activity of *mt*PriA in the absence and presence of compounds 5 - 7, 12, and 13 in their open/closed forms.

Inhibitor	c(I) [μM]	k_{cat} [s⁻¹]	K_M^{ProFAR} [μM]	K_i [μM]
			open/closed	
-		0.49	8.6	
		0.49	9.2	
		0.57	8.1	
5	5.0/5.0	0.47/0.58	12.7/13.9	10.5/8.1
	10.5/10.5	0.59/0.55	20.9/21.5	7.3/7.0
	18.0/18.0	0.55/0.60	32.5/35.2	6.5/5.8
6	0.4/2.0	0.52/0.56	16.2/12.4	0.45/4.5
	0.6/3.5	0.59/0.60	18.7/15.9	0.51/4.1
	0.8/8.0	0.55/0.61	18.5/23.7	0.69/4.6
7	2.0/3.5	0.47/0.58	13.1/16.2	3.8/4.0
	3.5/7.0	0.47/0.55	17.7/24.8	3.3/3.7
	5.0/10.5	0.47/0.53	21.7/35.6	3.3/3.3
12	1.5/4.0	0.57/0.55	16.3/15.2	1.7/5.2
	2.5/10.0	0.58/0.57	22.3/27.1	1.6/4.6
	4.0/15.0	0.60/0.55	29.4/34.9	1.7/4.9
13	5.0/15.0	0.60/0.58	15.7/15.1	6.1/19.8
	10.0/15.0	0.60/0.57	20.7/15.0	7.1/20.2
	15.0/20.0	0.59/0.57	26.6/14.7	7.2/28.2

Three substrate saturation curves were recorded at different concentrations c(I) of each compound and form (open/closed). The deduced k_{cat} and K_M^{ProFAR} values were used to calculate the respective inhibition constants with *Formula S1*, *publication C*. Taken from *Table S1*, *publication C*.

According to the steady-state kinetics, the inhibitory strength of three out of five compounds depends on the photoisomeric state. When irradiated with UV light, the inhibition activity dropped in all three cases by up to eightfold. However, the activity of *mt*PriA was not directly targeted in these measurements, since the respective isomers had already been added before the reaction was initiated. Therefore, it had to be demonstrated that the enzymatic performance can also be controlled during catalysis. To this end, the isomerization of ProFAR was followed spectrophotometrically in the presence of meta-phosphate **6** in its ring-open form (Figure 15). When the maximal velocity was reached after an initial lag phase, the solution was irradiated

with UV light to generate the ring-closed isomer of **6**. As a consequence thereof, the reaction rate could be increased about threefold under the prevailing conditions.
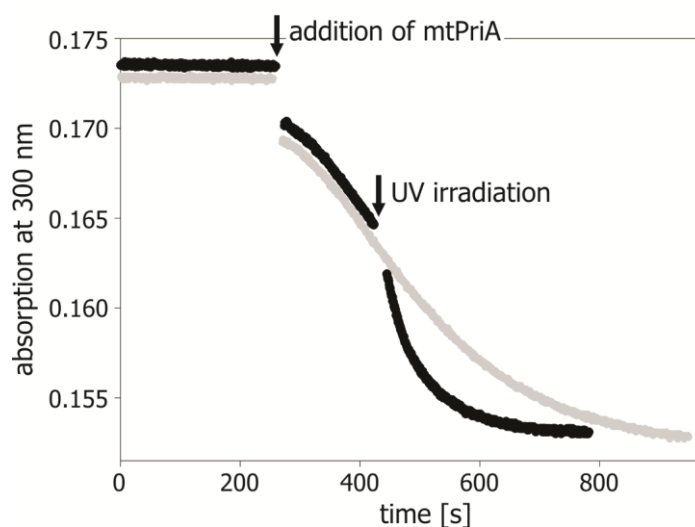
Figure 15. Remote control of *mt*PriA activity through ring-closure of compound 6.
4 µM of meta-phosphate **6** in its open form were added to a typical setup of ProFAR isomerization (5 µM ProFAR, 100 mM ammonium acetate, 0.18 µM HisF, and 0.15 µM *mt*PriA in 50 mM Tris/acetate pH 8.5 at 25 °C). The turnover of ProFAR was monitored spectrophotometrically at 300 nm. After reaching its maximal velocity, the reaction mixture was either left unchanged (grey) or irradiated with 312 nm light for 10 s (black). The accompanying baseline shift, resulting from different absorptions of the open and closed isomer at 300 nm, was corrected by means of a reference solution without enzymes. Taken from *Figure 2*, *publication C*.

Photoswitches like azobenzene and spiropyran significantly alter their overall geometry and polarity upon irradiation [151]. In contrast, ring-closure of the DTE scaffold only slightly affects the molecular length (*Scheme 2*, *publication C*), but mainly reduces the conformational flexibility [162]. Regarding the designed inhibitors, the open isomers possess four freely rotatable carbon-carbon bonds, which connect the thiophene heterocycles either with the central cyclopentene ring or with the terminal phenyl groups (Figure 13B). When irradiated with UV light, two flexible carbon-carbon bonds are incorporated in the newly formed cyclohexadiene ring. Furthermore, the closed isomers are completely conjugated, restricting the motility of the remaining non-cyclic carbon-carbon bonds. These conformational constraints probably hamper the adaptation to the active site of *mt*PriA and thus reduce the inhibition activities of the closed forms of compounds **6**, **12**, and **13**, even if the binding of the closed isomers is entropically favored. Furthermore, the deviating flexibilities of para-substituted compounds **7** and **13** offer a possible explanation for their different light-controllable inhibition properties (Table 2). While ring-closure of para-phosphate **7** did not alter the inhibitory activity, the K_i value of para-phosphonate **13** reduced threefold upon UV

irradiation. Although both inhibitors exhibit comparable distances between their phosphorous atoms (*Scheme 2*, *publication C*), para-phosphonate **13** lacks the versatile oxygen bridges connecting the phenyl rings with the terminal anchor groups. Consequently, the closed isomer of **13** is strongly impaired in its motility, giving rise to a distinct difference in the inhibition constants of the open and closed state.

Finally, in order to get a more detailed insight into the light-induced changes of the binding mode, MD simulations were performed with the open and closed isomer of meta-phosphate **6** bound to *mt*PriA. As expected, given the low inhibitory constants, the termini of both forms are clearly fixed in the phosphate binding sites via several hydrogen bond interactions (Figure 16A and Figure 16B). However, the molecular
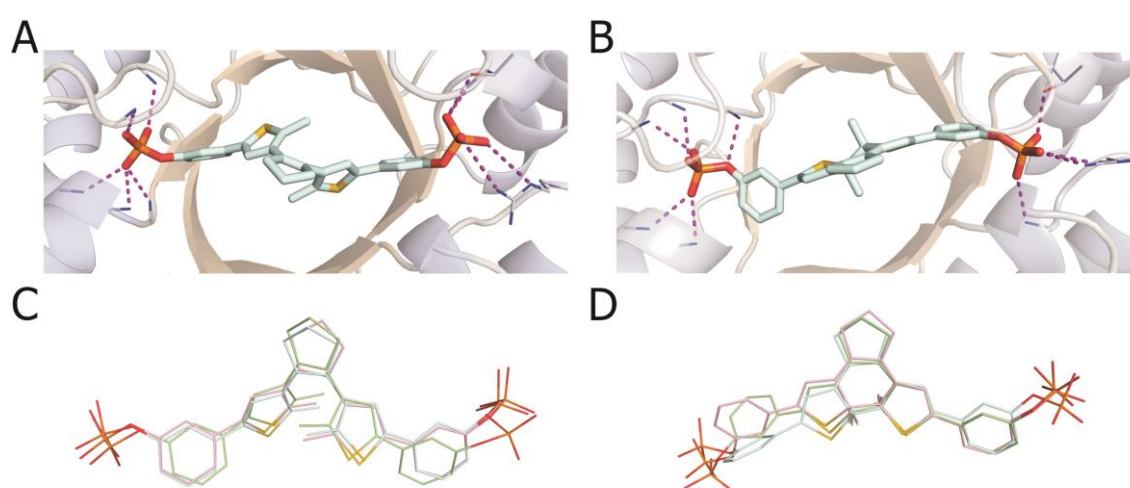
Figure 16. MD simulations of *mt*PriA with open and closed compound 6.
A (open)/B (closed): Representative ribbon diagrams are shown with views on the active sites (hydrogen bond interactions of the phosphate groups are indicated by dashed lines). C (open)/ D (closed): For each isomer, three independent calculations were performed and the respective conformers with the best binding energies are superposed. PDB ID: 3zs4. Taken from *Figure 3*, *publication C*.

cores of the inhibitors obviously vary. While the open isomer converges to widely coincident, twofold symmetric structures in three independent calculations (Figure 16C), the geometries of the closed form deviate from each other (Figure 16D). Here, one terminal phenyl ring is twisted to compensate for the structural rigidity of the molecular core and to enable an appropriate coordination of the attached phosphate group. The accompanying distortion of the twofold rotational symmetry is clearly reflected in the determined binding energies which are consistently higher than the values obtained for the open isomer (*Table S2*, *publication C*).

Recently, the design of light-inducible ion receptors and channels allowed for the spatiotemporal regulation of neuronal activity by irradiation [171, 172]. Targeting

neural systems is particularly effective, as even minor changes in the binding strength of the ligand trigger a nonlinear signal response and thus significantly affect the cellular output [173]. In contrast, light-control of enzymatic function is more pretentious and only few successful attempts have been reported [151]. Nevertheless, aiming at the metabolic branch-point enzyme *mt*PriA, the activity of DTE-based inhibitors could be reduced or enhanced by up to one order of magnitude through irradiation with UV and visible light, respectively. Remarkably, all photoisomers were nearly quantitatively formed and thermally stable, confirming the valuable photochemical features of the DTE scaffold [162]. By analogy with described two-pronged inhibitors [161, 162, 174, 175], the DTE compounds were equipped with two phosphate or phosphonate groups capable of docking at the phosphate binding sites of *mt*PriA. This dual anchoring strategy seems especially important for the effect of DTE switches, since the light-induced change in molecular flexibility only influences the inhibitory activity to a reasonable extent, if the mobility of the ligand is restricted by terminal fixation. Interestingly, other enzymes and pathways can theoretically be targeted with similar DTE switches, as phosphate is a widely distributed moiety among metabolites. Two phosphate binding sites are for instance found in pyridoxine-5′ phosphate synthase [176] and in certain aldolases [177, 178]. Hence, the presented design concept may be transferred to various enzymatic functions, enabling control of different cellular processes by light.

3   Literature

1.   Walsh, C., *Enabling the chemistry of life.* Nature, 2001. **409**(6817): p. 226-231.
2.   Wolfenden, R. and M.J. Snider, *The depth of chemical time and the power of enzymes as catalysts.* Acc Chem Res, 2001. **34**(12): p. 938-945.
3.   Radzicka, A. and R. Wolfenden, *A proficient enzyme.* Science, 1995. **267**(5194): p. 90-93.
4.   Miller, B.G. and R. Wolfenden, *Catalytic proficiency: The unusual case of OMP decarboxylase.* Annu Rev Biochem, 2002. **71**: p. 847-885.
5.   Albery, W.J. and J.R. Knowles, *Efficiency and evolution of enzyme catalysis.* Angew Chem Int Ed Engl, 1977. **16**(5): p. 285-293.
6.   Seelig, B. and J.W. Szostak, *Selection and evolution of enzymes from a partially randomized non-catalytic scaffold.* Nature, 2007. **448**(7155): p. 828-831.
7.   Apweiler, R., A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H.Z. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi, and L.S.L. Yeh, *UniProt: the Universal Protein knowledgebase.* Nucleic Acids Res, 2004. **32**: p. D115-D119.
8.   Murzin, A.G., S.E. Brenner, T. Hubbard, and C. Chothia, *Scop - a structural classification of proteins database for the investigation of sequences and structures.* J Mol Biol, 1995. **247**(4): p. 536-540.
9.   Govindarajan, S., R. Recabarren, and R.K. Goldstein, *Estimating the total number of protein folds.* Proteins: Struct, Funct, Genet, 1999. **35**(4): p. 408-414.
10. Sterner, R. and B. Höcker, *Catalytic versatility, stability, and evolution of the $(\beta\alpha)_8$-barrel enzyme fold.* Chem Rev, 2005. **105**(11): p. 4038-4055.
11. Caetano-Anolles, G., H.S. Kim, and J.E. Mittenthal, *The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture.* Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9358-9363.
12. Nagano, N., C.A. Orengo, and J.M. Thornton, *One fold with many functions: The evolutionary relationships between TIM barrel families based on their sequences, structures and functions.* J Mol Biol, 2002. **321**(5): p. 741-765.
13. Webb, E.C., *Enzyme nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology.* 1992. **Academic Press, New York**.
14. Golynskiy, M.V., J.C. Haugner, 3rd, and B. Seelig, *Highly diverse protein library based on the ubiquitous $(\beta/\alpha)_8$ enzyme fold yields well-structured proteins through in vitro folding selection.* Chembiochem, 2013. **14**(13): p. 1553-1563.
15. Campbell, E., S. Chuang, and S. Banta, *Modular exchange of substrate-binding loops alters both substrate and cofactor specificity in a member of the aldo-keto reductase superfamily.* Protein Eng Des Sel, 2013. **26**(3): p. 181-186.
16. Ma, H. and T.M. Penning, *Conversion of mammalian 3α-hydroxysteroid dehydrogenase to 20α-hydroxysteroid dehydrogenase using loop chimeras: changing specificity from androgens to progestins.* Proc Natl Acad Sci U S A, 1999. **96**(20): p. 11161-11166.
17. Vick, J.E. and J.A. Gerlt, *Evolutionary potential of $(\beta/\alpha)_8$-barrels: stepwise evolution of a "new" reaction in the enolase superfamily.* Biochemistry, 2007. **46**(50): p. 14589-14597.

18.   Meier, M.M., C. Rajendran, C. Malisi, N.G. Fox, C. Xu, S. Schlee, D.P. Barondeau, B. Höcker, R. Sterner, and F.M. Raushel, *Molecular engineering of organophosphate hydrolysis activity from a weak promiscuous lactonase template.* J Am Chem Soc, 2013. **135**(31): p. 11670-11677.

19.   Wise, E.L., W.S. Yew, J. Akana, J.A. Gerlt, and I. Rayment, *Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase suprafamily: structural basis for catalytic promiscuity in wild-type and designed mutants of 3-keto-L-gulonate 6-phosphate decarboxylase.* Biochemistry, 2005. **44**(6): p. 1816-1823.

20.   Yew, W.S., J. Akana, E.L. Wise, I. Rayment, and J.A. Gerlt, *Evolution of enzymatic activities in the orotidine 5'-monophosphate decarboxylase suprafamily: enhancing the promiscuous D-arabino-hex-3-ulose 6-phosphate synthase reaction catalyzed by 3-keto-L-gulonate 6-phosphate decarboxylase.* Biochemistry, 2005. **44**(6): p. 1807-1815.

21.   Patrick, W.M. and I. Matsumura, *A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase.* J Mol Biol, 2008. **377**(2): p. 323-336.

22.   Saab-Rincon, G., L. Olvera, M. Olvera, E. Rudino-Pinera, E. Benites, X. Soberon, and E. Morett, *Evolutionary walk between (β/α)$_8$ barrels: catalytic migration from triosephosphate isomerase to thiamin phosphate synthase.* J Mol Biol, 2012. **416**(2): p. 255-270.

23.   Jiang, L., E.A. Althoff, F.R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J.L. Gallaher, J.L. Betker, F. Tanaka, C.F. Barbas, 3rd, D. Hilvert, K.N. Houk, B.L. Stoddard, and D. Baker, *De novo computational design of retro-aldol enzymes.* Science, 2008. **319**(5868): p. 1387-1391.

24.   Privett, H.K., G. Kiss, T.M. Lee, R. Blomberg, R.A. Chica, L.M. Thomas, D. Hilvert, K.N. Houk, and S.L. Mayo, *Iterative approach to computational enzyme design.* Proc Natl Acad Sci U S A, 2012. **109**(10): p. 3790-3795.

25.   Röthlisberger, D., O. Khersonsky, A.M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J.L. Gallaher, E.A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K.N. Houk, D.S. Tawfik, and D. Baker, *Kemp elimination catalysts by computational enzyme design.* Nature, 2008. **453**(7192): p. 190-195.

26.   Copley, R.R. and P. Bork, *Homology among (βα)$_8$ barrels: implications for the evolution of metabolic pathways.* J Mol Biol, 2000. **303**(4): p. 627-641.

27.   Forsyth, W.R., O. Bilsel, Z. Gu, and C.R. Matthews, *Topology and sequence in the folding of a TIM barrel protein: global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a (βα)$_8$ barrel of unknown function from B. subtilis.* J Mol Biol, 2007. **372**(1): p. 236-253.

28.   Wilmanns, M., C.C. Hyde, D.R. Davies, K. Kirschner, and J.N. Jansonius, *Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis.* Biochemistry, 1991. **30**(38): p. 9161-9169.

29.   Farber, G.K., *An α/β-barrel full of evolutionary trouble.* Curr Opin Struct Biol, 1993. **3**(3): p. 409-412.

30.   Raine, A.R., N.S. Scrutton, and F.S. Mathews, *On the evolution of alternate core packing in eightfold β/α-barrels.* Protein Sci, 1994. **3**(10): p. 1889-1892.

31.   Henn-Sax, M., B. Höcker, M. Wilmanns, and R. Sterner, *Divergent evolution of (βα)$_8$-barrel enzymes.* Biol Chem, 2001. **382**(9): p. 1315-1320.

32. Lang, D., R. Thoma, M. Henn-Sax, R. Sterner, and M. Wilmanns, *Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion.* Science, 2000. **289**(5484): p. 1546-1550.

33. Höcker, B., J. Claren, and R. Sterner, *Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels.* Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16448-16453.

34. Carstensen, L., J.M. Sperl, M. Bocola, F. List, F.X. Schmid, and R. Sterner, *Conservation of the folding mechanism between designed primordial $(\beta\alpha)_8$-barrel proteins and their modern descendant.* J Am Chem Soc, 2012. **134**(30): p. 12786-12791.

35. Seitz, T., M. Bocola, J. Claren, and R. Sterner, *Stabillsation of a $(\beta\alpha)_8$-barrel protein designed from ldentical half barrels.* J Mol Biol, 2007. **372**(1): p. 114-129.

36. Claren, J., C. Malisi, B. Höcker, and R. Sterner, *Establishing wild-type levels of catalytic activity on natural and artificial $(\beta\alpha)_8$-barrel protein scaffolds.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3704-3709.

37. Sperl, J.M., B. Rohweder, C. Rajendran, and R. Sterner, *Establishing catalytic activity on an artificial $(\beta\alpha)_8$-barrel protein designed from identical half-barrels.* FEBS Lett, 2013. **587**(17): p. 2798-2805.

38. Henn-Sax, M., R. Thoma, S. Schmidt, M. Hennig, K. Kirschner, and R. Sterner, *Two $(\beta\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates.* Biochemistry, 2002. **41**(40): p. 12032-12042.

39. Barona-Gomez, F. and D.A. Hodgson, *Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis.* EMBO Rep, 2003. **4**(3): p. 296-300.

40. Kuper, J., C. Doenges, and M. Wilmanns, *Two-fold repeated $(\beta\alpha)_4$ half-barrels may provide a molecular tool for dual substrate specificity.* EMBO Rep, 2005. **6**(2): p. 134-139.

41. Beismann-Driemeyer, S. and R. Sterner, *Imidazole glycerol phosphate synthase from Thermotoga maritima - Quaternary structure, steady-state kinetics, and reaction mechanism of the bienzyme complex.* J Biol Chem, 2001. **276**(23): p. 20387-20396.

42. Douangamath, A., M. Walker, S. Beismann-Driemeyer, M.C. Vega-Fernandez, R. Sterner, and M. Wilmanns, *Structural evidence for ammonia tunneling across the $(\beta\alpha)_8$ barrel of the imidazole glycerol phosphate synthase bienzyme complex.* Structure, 2002. **10**(2): p. 185-193.

43. Chaudhuri, B.N., S.C. Lange, R.S. Myers, S.V. Chittur, V.J. Davisson, and J.L. Smith, *Crystal structure of imidazole glycerol phosphate synthase: a tunnel through a $(\beta/\alpha)_8$ barrel joins two active sites.* Structure, 2001. **9**(10): p. 987-997.

44. Klem, T.J. and V.J. Davisson, *Imidazole glycerol phosphate synthase: the glutamine amidotransferase in histidine biosynthesis.* Biochemistry, 1993. **32**(19): p. 5177-5186.

45. Myers, R.S., J.R. Jensen, I.L. Deras, J.L. Smith, and V.J. Davisson, *Substrate-induced changes in the ammonia channel for imidazole glycerol phosphate synthase.* Biochemistry, 2003. **42**(23): p. 7013-22.

46. Amaro, R.E., A. Sethi, R.S. Myers, V.J. Davisson, and Z.A. Luthey-Schulten, *A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase.* Biochemistry, 2007. **46**(8): p. 2156-2173.

47. Lipchock, J.M. and J.P. Loria, *Nanometer propagation of millisecond motions in V-type allostery.* Structure, 2010. **18**(12): p. 1596-1607.

48.   Rivalta, I., M.M. Sultan, N.S. Lee, G.A. Manley, J.P. Loria, and V.S. Batista, *Allosteric pathways in imidazole glycerol phosphate synthase.* Proc Natl Acad Sci U S A, 2012. **109**(22): p. E1428-E1436.

49.   List, F., M.C. Vega, A. Razeto, M.C. Häger, R. Sterner, and M. Wilmanns, *Catalysis uncoupling in a glutamine amidotransferase bienzyme by unblocking the glutaminase active site.* Chem Biol, 2012. **19**(12): p. 1589-1599.

50.   List, F., R. Sterner, and M. Wilmanns, *Related (βα)$_8$-barrel proteins in histidine and tryptophan biosynthesis: A paradigm to study enzyme evolution.* Chembiochem, 2011. **12**(10): p. 1487-1494.

51.   Jürgens, C., A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, and R. Sterner, *Directed evolution of a (βα)$_8$-barrel enzyme to catalyze related reactions in two different metabolic pathways.* Proc Natl Acad Sci U S A, 2000. **97**(18): p. 9925-9930.

52.   Näsvall, J., L. Sun, J.R. Roth, and D.I. Andersson, *Real-time evolution of new genes by innovation, amplification, and divergence.* Science, 2012. **338**(6105): p. 384-387.

53.   Evran, S., A. Telefoncu, and R. Sterner, *Directed evolution of (βα)$_8$-barrel enzymes: establishing phosphoribosylanthranilate isomerisation activity on the scaffold of the tryptophan synthase α-subunit.* Protein Eng Des Sel, 2012. **25**(6): p. 285-293.

54.   Leopoldseder, S., J. Claren, C. Jürgens, and R. Sterner, *Interconverting the catalytic activities of (βα)$_8$-barrel enzymes from different metabolic pathways: Sequence requirements and molecular analysis.* J Mol Biol, 2004. **337**(4): p. 871-879.

55.   Hommel, U., M. Eberhard, and K. Kirschner, *Phosphoribosyl anthranilate isomerase catalyzes a reversible amadori reaction.* Biochemistry, 1995. **34**(16): p. 5429-5439.

56.   McClerren, A.L., P. Zhou, Z.Q. Guan, C.R.H. Raetz, and J. Rudolph, *Kinetic analysis of the zinc-dependent deacetylase in the lipid a biosynthetic pathway.* Biochemistry, 2005. **44**(4): p. 1106-1113.

57.   Shim, J.H. and S.J. Benkovic, *Catalytic mechanism of Escherichia coli glycinamide ribonucleotide transformylase probed by site-directed mutagenesis and pH-dependent studies.* Biochemistry, 1999. **38**(31): p. 10024-10031.

58.   Grimsley, G.R., J.M. Scholtz, and C.N. Pace, *A summary of the measured pK values of the ionizable groups in folded proteins.* Protein Sci, 2009. **18**(1): p. 247-251.

59.   Harris, T.K. and G.J. Turner, *Structural basis of perturbed pK$_a$ values of catalytic groups in enzyme active sites.* IUBMB Life, 2002. **53**(2): p. 85-98.

60.   Due, A.V., J. Kuper, A. Geerlof, J.P. von Kries, and M. Wilmanns, *Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from Mycobacterium tuberculosis by active site metamorphosis.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3554-3559.

61.   Bisswanger, H., K. Kirschner, W. Cohn, V. Hager, and E. Hansson, *N-(5-Phosphoribosyl)anthranilate isomerase-indoleglycerol-phosphate synthase. 1. A substrate analogue binds to two different binding sites on the bifunctional enzyme from Escherichia coli.* Biochemistry, 1979. **18**(26): p. 5946-5953.

62.   Sterner, R., G.R. Kleemann, H. Szadkowski, A. Lustig, M. Hennig, and K. Kirschner, *Phosphoribosyl anthranilate isomerase from Thermotoga maritima is an extremely stable and active homodimer.* Protein Sci, 1996. **5**(10): p. 2000-2008.

63.   Dall'Acqua, W. and P. Carter, *Substrate-assisted catalysis: molecular basis and biological significance.* Protein Sci, 2000. **9**(1): p. 1-9.

64.   Hussain, T., V. Kamarthapu, S.P. Kruparani, M.V. Deshmukh, and R. Sankaranarayanan, *Mechanistic insights into cognate substrate discrimination*

during proofreading in translation. Proc Natl Acad Sci U S A, 2010. **107**(51): p. 22117-22121.

65. Kosloff, M. and Z. Selinger, *Substrate assisted catalysis application to G proteins.* Trends Biochem Sci, 2001. **26**(3): p. 161-166.

66. So, B.R., S. An, S. Kumar, M. Das, D.A. Turner, C.M. Hadad, and K. Musier-Forsyth, *Substrate-mediated fidelity mechanism ensures accurate decoding of proline codons.* J Biol Chem, 2011. **286**(36): p. 31810-31820.

67. Wright, H., L. Noda-Garcia, A. Ochoa-Leyva, D.A. Hodgson, V. Fulop, and F. Barona-Gomez, *The structure/function relationship of a dual-substrate (βα)₈-isomerase.* Biochem Biophys Res Commun, 2008. **365**(1): p. 16-21.

68. Kiss, G., D. Röthlisberger, D. Baker, and K.N. Houk, *Evaluation and ranking of enzyme designs.* Protein Sci, 2010. **19**(9): p. 1760-1773.

69. Kim, Y., M. Zhou, S. Moy, J. Morales, M.A. Cunningham, and A. Joachimiak, *High-resolution structure of the nitrile reductase QueF combined with molecular simulations provide insight into enzyme mechanism.* J Mol Biol, 2010. **404**(1): p. 127-137.

70. Lu, T., H.W. Tan, D. Lee, G.J. Chen, and Z.C. Jia, *New insights into the activation of Escherichia coli tyrosine kinase revealed by molecular dynamics simulation and biochemical analysis.* Biochemistry, 2009. **48**(33): p. 7986-7995.

71. Jitonnom, J., V.S. Lee, P. Nimmanpipug, H.A. Rowlands, and A.J. Mulholland, *Quantum mechanics/molecular mechanics modeling of substrate-assisted catalysis in family 18 chitinases: Conformational changes and the role of Asp142 in catalysis in ChiB.* Biochemistry, 2011. **50**(21): p. 4697-4711.

72. Senn, H.M. and W. Thiel, *QM/MM methods for biomolecular systems.* Angew Chem Int Ed Engl, 2009. **48**(7): p. 1198-1229.

73. Kong, X.Q., S.S. Ouyang, Z.J. Liang, J.Y. Lu, L. Chen, B.R. Shen, D.H. Li, M.Y. Zheng, K.K. Li, C. Luo, and H.L. Jiang, *Catalytic mechanism investigation of lysine-specific demethylase 1 (LSD1): A computational study.* PLoS ONE, 2011. **6**(9): p. e25444.

74. Li, H., A.D. Robertson, and J.H. Jensen, *Very fast empirical prediction and rationalization of protein pKₐ values.* Proteins: Struct, Funct, Bioinf, 2005. **61**(4): p. 704-721.

75. Lutz, S., I. Tubert-Brohman, Y.G. Yang, and M. Meuwly, *Water-assisted proton transfer in ferredoxin I.* J Biol Chem, 2011. **286**(27): p. 23679-23687.

76. Dourado, D.F.A.R., P.A. Fernandes, B. Mannervik, and M.J. Ramos, *Glutathione transferase: New model for glutathione activation.* Chem--Eur J, 2008. **14**(31): p. 9591-9598.

77. Nisbet, E.G. and N.H. Sleep, *The habitat and nature of early life.* Nature, 2001. **409**(6823): p. 1083-1091.

78. Bugg, T.D.H. and S. Ramaswamy, *Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations.* Curr Opin Chem Biol, 2008. **12**(2): p. 134-140.

79. Hammes, G.G., S.J. Benkovic, and S. Hammes-Schiffer, *Flexibility, diversity, and cooperativity: Pillars of enzyme catalysis.* Biochemistry, 2011. **50**(48): p. 10422-10430.

80. Henzler-Wildman, K.A., V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M.A. Wilson, G.A. Petsko, M. Karplus, C.G. Hubner, and D. Kern,

Intrinsic motions along an enzymatic reaction trajectory. Nature, 2007. **450**(7171): p. 838-844.

81. Miles, E.W., *The tryptophan synthase α2β2 complex: a model for substrate channeling, allosteric communication, and pyridoxal phosphate catalysis.* J Biol Chem, 2013. **288**(14): p. 10084-10091.

82. Nagradova, N., *Interdomain communications in bifunctional enzymes: How are different activities coordinated?* IUBMB Life, 2003. **55**(8): p. 459-466.

83. Saibil, H.R., *Chaperone machines in action.* Curr Opin Struct Biol, 2008. **18**(1): p. 35-42.

84. Groll, M., M. Bochtler, H. Brandstetter, T. Clausen, and R. Huber, *Molecular machines for protein degradation.* Chembiochem, 2005. **6**(2): p. 222-256.

85. Dean, A.M. and J.W. Thornton, *Mechanistic approaches to the study of evolution: the functional synthesis.* Nat Rev Genet, 2007. **8**(9): p. 675-688.

86. Woese, C., *The universal ancestor.* Proc Natl Acad Sci U S A, 1998. **95**(12): p. 6854-6859.

87. Kim, K.M. and G. Caetano-Anolles, *The proteomic complexity and rise of the primordial ancestor of diversified life.* BMC Evol Biol, 2011. **11**: 140.

88. Di Giulio, M., *The Last Universal Common Ancestor (LUCA) and the ancestors of Archaea and Bacteria were progenotes.* J Mol Evol, 2011. **72**(1): p. 119-126.

89. Koonin, E.V., *Comparative genomics, minimal gene-sets and the last universal common ancestor.* Nat Rev Microbiol, 2003. **1**(2): p. 127-136.

90. Tuller, T., H. Birin, U. Gophna, M. Kupiec, and E. Ruppin, *Reconstructing ancestral gene content by coevolution.* Genome Res, 2010. **20**(1): p. 122-132.

91. Ranea, J.A.G., A. Sillero, J.M. Thornton, and C.A. Orengo, *Protein superfamily evolution and the last universal common ancestor (LUCA).* J Mol Evol, 2006. **63**(4): p. 513-525.

92. Ouzounis, C.A., V. Kunin, N. Darzentas, and L. Goldovsky, *A minimal estimate for the gene content of the last universal common ancestor - exobiology from a terrestrial perspective.* Res Microbiol, 2006. **157**(1): p. 57-68.

93. Glansdorff, N., Y. Xu, and B. Labedan, *The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner.* Biol Direct, 2008. **3**: 29.

94. Doolittle, W.F., *The nature of the universal ancestor and the evolution of the proteome.* Curr Opin Struct Biol, 2000. **10**(3): p. 355-358.

95. Becerra, A., L. Delaye, S. Islas, and A. Lazcano, *The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains.* Annu Rev Ecol Evol Syst, 2007. **38**: p. 361-379.

96. Pauling, L. and E. Zuckerkandl, *Chemical paleogenetics: molecular ''restoration studies'' of extinct forms of life.* Acta Chem Scand, 1963. **17**: p. 9-16.

97. Thornton, J.W., *Resurrecting ancient genes: experimental analysis of extinct molecules.* Nat Rev Genet, 2004. **5**(5): p. 366-375.

98. Harms, M.J. and J.W. Thornton, *Analyzing protein structure and function using ancestral gene reconstruction.* Curr Opin Struct Biol, 2010. **20**(3): p. 360-366.

99. Benner, S.A., S.O. Sassi, and E.A. Gaucher, *Molecular paleoscience: systems biology from the past.* Adv Enzymol Relat Areas Mol Biol, 2007. **75**: p. 1-132.

100. Yang, Z., S. Kumar, and M. Nei, *A new method of inference of ancestral nucleotide and amino acid sequences.* Genetics, 1995. **141**(4): p. 1641-1650.

101. Hanson-Smith, V., B. Kolaczkowski, and J.W. Thornton, *Robustness of ancestral sequence reconstruction to phylogenetic uncertainty.* Mol Biol Evol, 2010. **27**(9): p. 1988-1999.

102. Blanquart, S. and N. Lartillot, *A site- and time-heterogeneous model of amino acid replacement.* Mol Biol Evol, 2008. **25**(5): p. 842-858.

103. Puigbo, P., Y.I. Wolf, and E.V. Koonin, *Search for a 'Tree of Life' in the thicket of the phylogenetic forest.* J Biol, 2009. **8**(6): p. 59.

104. Richter, M., M. Bosnali, L. Carstensen, T. Seitz, H. Durchschlag, S. Blanquart, R. Merkl, and R. Sterner, *Computational and experimental evidence for the evolution of a (βα)₈-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds.* J Mol Biol, 2010. **398**(5): p. 763-773.

105. Boussau, B., S. Blanquart, A. Necsulea, N. Lartillot, and M. Gouy, *Parallel adaptations to high temperatures in the Archaean eon.* Nature, 2008. **456**(7224): p. 942-945.

106. Gaucher, E.A., S. Govindarajan, and O.K. Ganesh, *Palaeotemperature trend for Precambrian life inferred from resurrected proteins.* Nature, 2008. **451**(7179): p. 704-707.

107. Gaucher, E.A., J.M. Thomson, M.F. Burgan, and S.A. Benner, *Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins.* Nature, 2003. **425**(6955): p. 285-288.

108. Chang, B.S.W., K. Jonsson, M.A. Kazmi, M.J. Donoghue, and T.P. Sakmar, *Recreating a functional ancestral archosaur visual pigment.* Mol Biol Evol, 2002. **19**(9): p. 1483-1489.

109. Shi, Y.S. and S. Yokoyama, *Molecular analysis of the evolutionary significance of ultraviolet vision in vertebrates.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8308-8313.

110. Yokoyama, S., H. Yang, and W.T. Starmert, *Molecular basis of spectral tuning in the red- and green-sensitive (M/LWS) pigments in vertebrates.* Genetics, 2008. **179**(4): p. 2037-2043.

111. Field, S.F. and M.V. Matz, *Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals.* Mol Biol Evol, 2010. **27**(2): p. 225-233.

112. Ugalde, J.A., B.S.W. Chang, and M.V. Matz, *Evolution of coral pigments recreated.* Science, 2004. **305**(5689): p. 1433-1433.

113. Bridgham, J.T., S.M. Carroll, and J.W. Thornton, *Evolution of hormone-receptor complexity by molecular exploitation.* Science, 2006. **312**(5770): p. 97-101.

114. Bridgham, J.T., E.A. Ortlund, and J.W. Thornton, *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution.* Nature, 2009. **461**(7263): p. 515-519.

115. Carroll, S.M., E.A. Ortlund, and J.W. Thornton, *Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor.* PLoS Genet, 2011. **7**(6): p. e1002117.

116. Eick, G.N., J.K. Colucci, M.J. Harms, E.A. Ortlund, and J.W. Thornton, *Evolution of minimal specificity and promiscuity in steroid hormone receptors.* PLoS Genet, 2012. **8**(11): p. e1003072.

117. Harms, M.J., G.N. Eick, D. Goswami, J.K. Colucci, P.R. Griffin, E.A. Ortlund, and J.W. Thornton, *Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11475-11480.

118.   Ortlund, E.A., J.T. Bridgham, M.R. Redinbo, and J.W. Thornton, *Crystal structure of an ancient protein: Evolution by conformational epistasis.* Science, 2007. **317**(5844): p. 1544-1548.

119.   Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: Ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-1717.

120.   Thomson, J.M., E.A. Gaucher, M.F. Burgan, D.W. De Kee, T. Li, J.P. Aris, and S.A. Benner, *Resurrecting ancestral alcohol dehydrogenases from yeast.* Nat Genet, 2005. **37**(6): p. 630-635.

121.   Voordeckers, K., C.A. Brown, K. Vanneste, E. van der Zande, A. Voet, S. Maere, and K.J. Verstrepen, *Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication.* PLoS Biol, 2012. **10**(12): p. e1001446.

122.   Akanuma, S., Y. Nakajima, S. Yokobori, M. Kimura, N. Nemoto, T. Mase, K. Miyazono, M. Tanokura, and A. Yamagishi, *Experimental evidence for the thermophilicity of ancestral life.* Proc Natl Acad Sci U S A, 2013. **110**(27): p. 11067-11072.

123.   Huang, R.Q., F. Hippauf, D. Rohrbeck, M. Haustein, K. Wenke, J. Feike, N. Sorrelle, B. Piechulla, and T.J. Barkman, *Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2966-2971.

124.   Hobbs, J.K., C. Shepherd, D.J. Saul, N.J. Demetras, S. Haaning, C.R. Monk, R.M. Daniel, and V.L. Arcus, *On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of Bacillus.* Mol Biol Evol, 2012. **29**(2): p. 825-835.

125.   Risso, V.A., J.A. Gavira, D.F. Mejia-Carmona, E.A. Gaucher, and J.M. Sanchez-Ruiz, *Hyperstability and substrate promiscuity in laboratory resurrections of precambrian β-Lactamases.* J Am Chem Soc, 2013. **135**(8): p. 2899-2902.

126.   Finnigan, G.C., V. Hanson-Smith, T.H. Stevens, and J.W. Thornton, *Evolution of increased complexity in a molecular machine.* Nature, 2012. **481**(7381): p. 360-364.

127.   Perez-Jimenez, R., A. Ingles-Prieto, Z.M. Zhao, I. Sanchez-Romero, J. Alegre-Cebollada, P. Kosuri, S. Garcia-Manyes, T.J. Kappock, M. Tanokura, A. Holmgren, J.M. Sanchez-Ruiz, E.A. Gaucher, and J.M. Fernandez, *Single-molecule paleoenzymology probes the chemistry of resurrected enzymes.* Nat Struct Mol Biol, 2011. **18**(5): p. 592-596.

128.   Fani, R. and M. Fondi, *Origin and evolution of metabolic pathways.* Phys Life Rev, 2009. **6**(1): p. 23-52.

129.   Klem, T.J. and V.J. Davisson, *Imidazole Glycerol Phosphate Synthase - the Glutamine Amidotransferase in Histidine Biosynthesis.* Biochemistry, 1993. **32**(19): p. 5177-5186.

130.   Boratyn, G.M., C. Camacho, P.S. Cooper, G. Coulouris, A. Fong, N. Ma, T.L. Madden, W.T. Matten, S.D. McGinnis, Y. Merezhuk, Y. Raytselis, E.W. Sayers, T. Tao, J. Ye, and I. Zaretskaya, *BLAST: a more efficient report with usability improvements.* Nucleic Acids Res, 2013. **41**(W1): p. W29-W33.

131.   Vetriani, C., M.D. Speck, S.V. Ellor, R.A. Lutz, and V. Starovoytov, *Thermovibrio ammonificans sp nov., a thermophilic, chemolithotrophic, nitrate-ammonifying bacterium from deep-sea hydrothermal vents.* Int J Syst Evol Microbiol, 2004. **54**: p. 175-181.

132. Martin, W., J. Baross, D. Kelley, and M.J. Russell, *Hydrothermal vents and the origin of life.* Nat Rev Microbiol, 2008. **6**(11): p. 805-814.

133. Sali, A. and T.L. Blundell, *Comparative protein modeling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.

134. Russell, R.B. and G.J. Barton, *Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels.* Proteins: Struct, Funct, Genet, 1992. **14**(2): p. 309-323.

135. Höcker, B., A. Lochner, T. Seitz, J. Claren, and R. Sterner, *High-resolution crystal structure of an artificial (βα)₈-barrel protein designed from identical half-barrels.* Biochemistry, 2009. **48**(6): p. 1145-1147.

136. Waterhouse, A.M., J.B. Procter, D.M.A. Martin, M. Clamp, and G.J. Barton, *Jalview Version 2-a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-1191.

137. Santoro, M.M. and D.W. Bolen, *Unfolding free-energy changes determined by the linear extrapolation method 1. Unfolding of phenylmethanesulfonyl α-chymotrypsin using different denaturants.* Biochemistry, 1988. **27**(21): p. 8063-8068.

138. Carstensen, L., G. Zoldak, F.X. Schmid, and R. Sterner, *Folding mechanism of an extremely thermostable (βα)₈-barrel enzyme: A high kinetic barrier protects the protein from denaturation.* Biochemistry, 2012. **51**(16): p. 3420-3432.

139. Jensen, R.A., *Enzyme recruitment in evolution of new function.* Annu Rev Microbiol, 1976. **30**: p. 409-425.

140. Loytynoja, A. and N. Goldman, *Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis.* Science, 2008. **320**(5883): p. 1632-1635.

141. Loytynoja, A. and N. Goldman, *An algorithm for progressive multiple alignment of sequences with insertions.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10557-10562.

142. Hammes, G.G. and C.W. Wu, *Regulation of enzyme activity.* Science, 1971. **172**(3989): p. 1205-1211.

143. Stroud, R.M., A.A. Kossiakoff, and J.L. Chambers, *Mechanisms of zymogen activation.* Annu Rev Biophys Bioeng, 1977. **6**: p. 177-193.

144. Worby, C.A., S. Mattoo, R.P. Kruger, L.B. Corbeil, A. Koller, J.C. Mendez, B. Zekarias, C. Lazar, and J.E. Dixon, *The Fic domain: Regulation of cell signaling by adenylylation.* Mol Cell, 2009. **34**(1): p. 93-103.

145. Johnson, L.N. and R.J. Lewis, *Structural basis for control by phosphorylation.* Chem Rev, 2001. **101**(8): p. 2209-2242.

146. Stadtman, E.R., *Allosteric regulation of enzyme activity.* Adv Enzymol Relat Areas Mol Biol, 1966. **28**: p. 41-154.

147. Knöchel, T., A. Ivens, G. Hester, A. Gonzalez, R. Bauerle, M. Wilmanns, K. Kirschner, and J.N. Jansonius, *The crystal structure of anthranilate synthase from Sulfolobus solfataricus: Functional implications.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9479-9484.

148. Kantrowitz, E.R., *Allostery and cooperativity in Escherichia coli aspartate transcarbamoylase.* Arch Biochem Biophys, 2012. **519**(2): p. 81-90.

149. Losi, A. and W. Gärtner, *Bacterial bilin- and flavin-binding photoreceptors.* Photochem Photobiol Sci, 2008. **7**(10): p. 1168-1178.

150. Crosson, S., S. Rajagopal, and K. Moffat, *The LOV domain family: Photoresponsive signaling modules coupled to diverse output domains.* Biochemistry, 2003. **42**(1): p. 2-10.

151. Szymanski, W., J.M. Beierle, H.A. Kistemaker, W.A. Velema, and B.L. Feringa, *Reversible photocontrol of biological systems by the incorporation of molecular photoswitches.* Chem Rev, 2013. **113**: p. 6114-6178.

152. Krauss, U., T. Drepper, and K.E. Jaeger, *Enlightened enzymes: strategies to create novel photoresponsive proteins.* Chem--Eur J, 2011. **17**(9): p. 2552-2560.

153. Drepper, T., U. Krauss, S.M.Z. Berstenhorst, J. Pietruszka, and K.E. Jaeger, *Lights on and action! Controlling microbial gene expression by light.* Applied Microbiology and Biotechnology, 2011. **90**(1): p. 23-40.

154. Brieke, C., F. Rohrbach, A. Gottschalk, G. Mayer, and A. Heckel, *Light-controlled tools.* Angew Chem Int Ed Engl, 2012. **51**(34): p. 8446-8476.

155. Deiters, A., *Principles and applications of the photochemical control of cellular processes.* Chembiochem, 2010. **11**(1): p. 47-53.

156. Weston, D.G., J. Kirkham, and D.C. Cullen, *Photo-modulation of horseradish peroxidase activity via covalent attachment of carboxylated-spiropyran dyes.* Biochim Biophys Acta, 1999. **1428**(2-3): p. 463-467.

157. Schierling, B., A.J. Noel, W. Wende, T. Hien le, E. Volkov, E. Kubareva, T. Oretskaya, M. Kokkinidis, A. Rompp, B. Spengler, and A. Pingoud, *Controlling the enzymatic activity of a restriction enzyme by light.* Proc Natl Acad Sci U S A, 2010. **107**(4): p. 1361-1366.

158. Muramatsu, S., K. Kinbara, H. Taguchi, N. Ishii, and T. Aida, *Semibiological molecular machine with an implemented "AND" logic gate for regulation of protein folding.* J Am Chem Soc, 2006. **128**(11): p. 3764-3769.

159. Willner, I., S. Rubin, R. Shatzmiller, and T. Zor, *Reversible light-stimulated activation and deactivation of α-chymotrypsin by its immobilization in photoisomerizable copolymers.* J Am Chem Soc, 1993. **115**(19): p. 8690-8694.

160. Wang, S.C. and C.T. Lee, *Enhanced enzymatic activity through photoreversible conformational changes.* Biochemistry, 2007. **46**(50): p. 14557-14566.

161. Vomasta, D., A. Innocenti, B. König, and C.T. Supuran, *Carbonic anhydrase inhibitors: Two-prong versus mono-prong inhibitors of isoforms I, II, IX, and XII exemplified by photochromic cis-1,2-alpha-dithienylethene derivatives.* Bioorg Med Chem Lett, 2009. **19**(5): p. 1283-1286.

162. Vomasta, D., C. Högner, N.R. Branda, and B. König, *Regulation of human carbonic anhydrase I (hCAI) activity by using a photochromic inhibitor.* Angew Chem Int Ed Engl, 2008. **47**(40): p. 7644-7647.

163. Herre, S., T. Schadendorf, I. Ivanov, C. Herrberger, W. Steinle, K. Rück-Braun, R. Preissner, and H. Kuhn, *Photoactivation of an inhibitor of the 12/15-lipoxygenase pathway.* Chembiochem, 2006. **7**(7): p. 1089-1095.

164. Fujita, D., M. Murai, T. Nishioka, and H. Miyoshi, *Light control of mitochondrial complex I activity by a photoresponsive inhibitor.* Biochemistry, 2006. **45**(21): p. 6581-6586.

165. Shen, H., F. Wang, Y. Zhang, Q. Huang, S. Xu, H. Hu, J. Yue, and H. Wang, *A novel inhibitor of indole-3-glycerol phosphate synthase with activity against multidrug-resistant Mycobacterium tuberculosis.* FEBS J, 2009. **276**(1): p. 144-154.

166. Mdluli, K. and M. Spigelman, *Novel targets for tuberculosis drug discovery.* Curr Opin Pharmacol, 2006. **6**(5): p. 459-467.

167. Lee, J. and M. Blaber, *Experimental support for the evolution of symmetric protein architecture from a simple peptide motif.* Proc Natl Acad Sci U S A, 2011. **108**(1): p. 126-130.

168. Waldeck, D.H., *Photoisomerization dynamics of stilbenes.* Chem Rev, 1991. **91**(3): p. 415-436.

169. Beharry, A.A. and G.A. Woolley, *Azobenzene photoswitches for biomolecules.* Chem Soc Rev, 2011. **40**(8): p. 4422-4437.

170. Irie, M., *Diarylethenes for memories and switches.* Chem Rev, 2000. **100**(5): p. 1685-1716.

171. Tochitsky, I., M.R. Banghart, A. Mourot, J.Z. Yao, B. Gaub, R.H. Kramer, and D. Trauner, *Optochemical control of genetically engineered neuronal nicotinic acetylcholine receptors.* Nat Chem, 2012. **4**(2): p. 105-111.

172. Fehrentz, T., M. Schönberger, and D. Trauner, *Optochemical genetics.* Angew Chem Int Ed Engl, 2011. **50**(51): p. 12156-12182.

173. Banghart, M.R., A. Mourot, D.L. Fortin, J.Z. Yao, R.H. Kramer, and D. Trauner, *Photochromic blockers of voltage-gated potassium channels.* Angew Chem Int Ed Engl, 2009. **48**(48): p. 9097-9101.

174. Kuil, J., L.T. van Wandelen, N.J. de Mol, and R.M. Liskamp, *Switching between low and high affinity for the Syk tandem SH2 domain by irradiation of azobenzene containing ITAM peptidomimetics.* J Pept Sci, 2009. **15**(10): p. 685-691.

175. Kuil, J., L.T. van Wandelen, N.J. de Mol, and R.M. Liskamp, *A photoswitchable ITAM peptidomimetic: synthesis and real time surface plasmon resonance (SPR) analysis of the effects of cis-trans isomerization on binding.* Bioorg Med Chem, 2008. **16**(3): p. 1393-1399.

176. Franco, M.G., B. Laber, R. Huber, and T. Clausen, *Structural basis for the function of pyridoxine 5 '-phosphate synthase.* Structure, 2001. **9**(3): p. 245-253.

177. Wagner, T., I.A. Shumilin, R. Bauerle, and R.H. Kretsinger, *Structure of 3-deoxy-d-arabino-heptulosonate-7-phosphate synthase from Escherichia coli: comparison of the Mn(2+)*2-phosphoglycolate and the Pb(2+)*2-phosphoenolpyruvate complexes and implications for catalysis.* J Mol Biol, 2000. **301**(2): p. 389-399.

178. Lorentzen, E., E. Pohl, P. Zwart, A. Stark, R.B. Russell, T. Knura, R. Hensel, and B. Siebers, *Crystal structure of an archaeal class I aldolase and the evolution of $(\beta\alpha)_8$ barrel proteins.* J Biol Chem, 2003. **278**(47): p. 47253-47260.

4   List of publications and personal contribution

A.   Reisinger, B., M. Bocola, F. List, J. Claren, C. Rajendran and R. Sterner, *A sugar isomerization reaction established on various (βa)₈-barrel scaffolds is based on substrate-assisted catalysis*. Protein Eng. Des. Sel. 2012, **25**(11), 751-760.

   DOI: 10.1093/protein/gzs080

Marco Bocola performed the MD simulations and the QM/MM calculations. Felix List partly supervised research. Jörg Claren produced and characterized variant HisAF_D127V. Chitra Rajendran collected the X-ray dataset and solved X-ray structure of HisF_D130V+D176V. All other experiments were performed by myself. The publication was written by myself and Reinhard Sterner.

B.   Reisinger, B., J. M. Sperl, A. Holinski, V. Schmid, C. Rajendran, L. Carstensen, S. Schlee, S. Blanquart, R. Merkl and R. Sterner, *Evidence for the existence of elaborate enzyme complexes in the Paleoarchean era*. J. Am. Chem. Soc. 2014, **136**(1), 122-129.

   DOI: 10.1021/ja4115677

Alexandra Holinski produced and characterized LUCA-HisH. Veronika Schmid produced and partly characterized *zm*HisH. Chitra Rajendran collected the X-ray dataset and solved X-ray structure of LUCA-HisF. Linn Carstensen performed the GdmCl-induced equilibrium unfolding/refolding transitions and folding kinetics of LUCA-HisF. Rainer Merkl and Samuel Blanquart reconstructed the protein sequences. All other experiments were performed by myself and Josef Sperl to equal parts. The publication was written by myself, Reinhard Sterner, Rainer Merkl, Josef Sperl and Sandra Schlee.

C.   Reisinger, B., N. Kuzmanovic, P. Löffler, R. Merkl, B. König and R. Sterner, *Exploiting protein symmetry to design light-controllable enzyme inhibitors*. Angew. Chem. Int. Ed. 2014, **53**(2), 595-598.

   DOI: 10.1002/anie.201307207

Natascha Kuzmanovic synthesized and partly photochemically characterized the enzyme inhibitors. Patrick Löffler performed the MD simulations. All other experiments were performed by myself. The publication was written by myself, Natascha Kuzmanovic, Reinhard Sterner, Rainer Merkl and Burkhard König.

5 Publications

5.1 Publication A

A sugar isomerization reaction established on various $(\beta\alpha)_8$-barrel scaffolds is based on substrate-assisted catalysis.

Bernd Reisinger, Marco Bocola, Felix List, Jörg Claren, Chitra Rajendran and Reinhard Sterner

A sugar isomerization reaction established on various (βα)8-barrel scaffolds is based on substrate-assisted catalysis[†]

**Bernd Reisinger[1], Marco Bocola[1,2], Felix List[1,3],
Jörg Claren[1], Chitra Rajendran[1] and Reinhard Sterner[1,4]**

[1]Institute of Biophysics and Physical Biochemistry, University of
Regensburg, Germany, [2]Department of Biotechnology (Biology VI), RWTH
Aachen University, Universitätsstrasse 31, D-93053 Germany and [3]European
Molecular Biology Laboratory (EMBL), Hamburg Unit, Hamburg, Germany

[4]To whom correspondence should be addressed. E-mail: reinhard.sterner@
biologie.uni-regensburg.de

Edited by Miroslaw Cygler

**In the course of tryptophan biosynthesis, the isomeriza-
tion of phosphoribosylanthranilate (PRA) is catalyzed by
the (βα)8-barrel enzyme TrpF. The reaction occurs via a
general acid–base mechanism with an aspartate and a
cysteine residue acting as acid and base, respectively.
PRA isomerase activity could be established on two
(βα)8-barrel enzymes involved in histidine biosynthesis,
namely HisA and HisF, and on a HisAF chimera, by
introducing two aspartate-to-valine substitutions. We
have analyzed the reaction mechanism underlying this
engineered activity by measuring its pH dependence,
solving the crystal structure of a HisF variant with bound
product analogue, and applying molecular dynamics
simulations and mixed quantum and molecular
mechanics calculations. The results suggest that PRA is
anchored by the C-terminal phosphate-binding sites of
HisA, HisF and HisAF. As a consequence, a conserved as-
partate residue, which is equivalent to Cys7 from TrpF, is
properly positioned to act as catalytic base. However, no
obvious catalytic acid corresponding to Asp126 from
TrpF could be identified in the three proteins. Instead,
this role appears to be carried out by the carboxylate
group of the anthranilate moiety of PRA. Thus, the engi-
neered PRA isomerization activity is based on a reaction
mechanism including substrate-assisted catalysis and thus
differs substantially from the naturally evolved reaction
mechanism used by TrpF.**
Keywords: (βα)8-barrel/enzyme design/histidine
biosynthesis/substrate-assisted catalysis/tryptophan
biosynthesis

Introduction

The (βα)8- or TIM-barrel is the most common and most ver-
satile fold among naturally occurring enzymes (Wierenga,
2001; Nagano *et al.*, 2002). Its structural assembly consists of
eight modular (βα) units. The β-strand and the α-helix

within the modules are connected by βα-loops, whereas suc-
cessive modules are linked by αβ-loops. In the tertiary struc-
ture, a central β-barrel is made up of the eight β-strands and
surrounded by the α-helices. The fold can be separated in a
catalytic face and a stability face: active site residues are
located at the C-terminal ends of the β-strands and in the
βα-loops, while amino acids important for stability are
mainly found in the remainder of the fold including the
αβ-loops (Sterner and Höcker, 2005). Hence, mutations can
be introduced in the active site without compromising stabil-
ity, making the (βα)8-barrel an ideal fold for enzyme design.
Along these lines, different (βα)8-barrel proteins have been
used as scaffolds to establish catalytic activities of other
enzymes (Vick and Gerlt, 2007; Jiang *et al.*, 2008) and even
reactions that do not belong to nature's repertoire
(Röthlisberger *et al.*, 2008).

Several (βα)8-barrel proteins are involved in the biosyn-
thesis of the aromatic amino acids histidine and tryptophan
(List *et al.*, 2011). As shown in Fig. 1, the enzymes
N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-car-
boxamide ribonucleotide (ProFAR) isomerase (HisA) and
imidazole glycerol phosphate (ImGP) synthase (HisF) cata-
lyze two subsequent steps in histidine biosynthesis (Lang
et al., 2000), whereas phosphoribosylanthranilate (PRA)
isomerase (TrpF) is involved in tryptophan biosynthesis.
HisA and TrpF catalyze chemically equivalent isomerization
reactions of aminoaldoses into the corresponding aminoke-
toses, their substrates, however, differ significantly in size
(Henn-Sax *et al.*, 2002). HisA and HisF contain two
phosphate-binding sites formed by βα-loops 3 and 4 and
βα-loops 7 and 8, respectively, which serve to anchor their
bi-phosphorylated substrates ProFAR and PRFAR. The
C-terminal phosphate-binding site also exists in TrpF and
serves to anchor PRA. The common fold, the shared
phosphate-binding site as well as the similarities in the cata-
lyzed reactions have allowed for the establishment of PRA
isomerization activity on the scaffolds HisA and HisF from
Thermotoga maritima (Jürgens *et al.*, 2000; Leopoldseder
et al., 2004) as well as on the chimeric scaffold HisAF,
which is composed of the four N-terminal (βα)4 modules
from HisA and the four C-terminal (βα)4 modules from HisF
(Claren *et al.*, 2009). In all cases, the introduction of a single
equivalent amino acid exchange (D127V in HisA and
HisAF, D130V in HisF) by a combination of random muta-
genesis and selection *in vivo* was sufficient to effect a weak
turnover of the aminoaldose PRA into the aminoketose
product 1-(2-carboxy-phenylamino)-1′-deoxyribulose-5′-
phosphate (CdRP).

In the TrpF wild-type enzyme the isomerization of PRA
occurs via a general acid–base mechanism (Henn-Sax *et al.*,
2002), which is outlined in Fig. 2. In the initial reaction step
the furanose ring oxygen is protonated by the acid Asp126.
Subsequently, the formed Schiff base intermediate acts as an

751
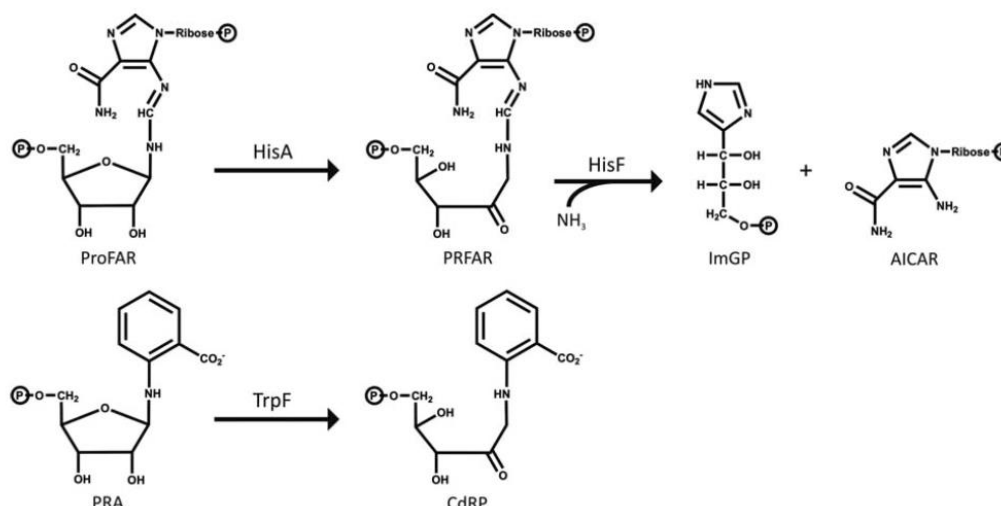
Fig. 1. Reactions catalyzed by TrpF, HisA and HisF. TrpF and HisA catalyze the Amadori rearrangements of the aminoaldoses PRA and ProFAR into the corresponding ketoses CdRP and PRFAR. The latter is converted by HisF into ImGP and AICAR. The ammonia molecule required for this reaction is provided by the glutaminase HisH (not shown).
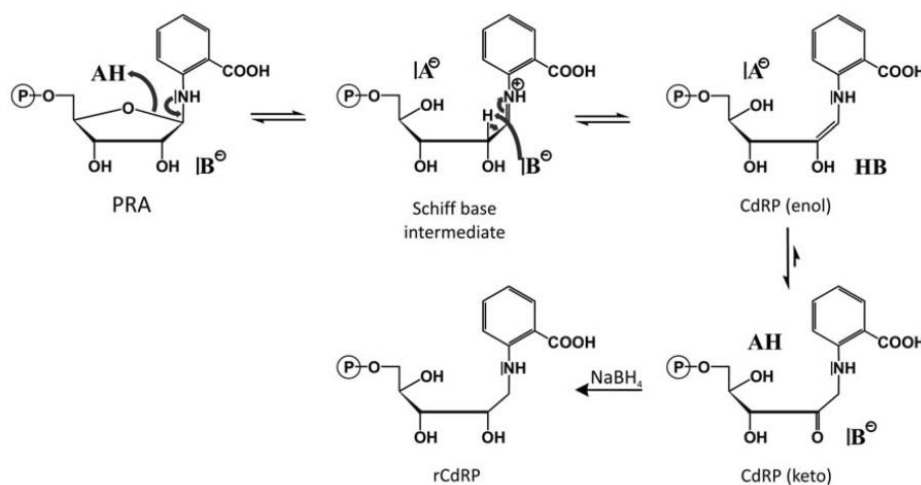
Fig. 2. Mechanism of PRA isomerization catalyzed by the TrpF enzyme from *Thermotoga maritima*. In the initial step the furanose ring oxygen of PRA is protonated by the general acid AH (Asp 126). The resulting Schiff base intermediate acts as an electron sink that enables the deprotonation of the $C2'$ atom by the general base B⁻ (Cys 7). The resulting enolamine form of CdRP tautomerizes to the corresponding keto form in an enzyme-independent manner (Henn-Sax *et al.*, 2002). Reduction of CdRP with sodium borohydride yields the stable product analogue rCdRP (Bisswanger *et al.*, 1979).

electron sink which enables the base Cys7 to abstract a proton from the carbon atom $C2'$. Finally, the enol-form of CdRP spontaneously tautomerizes to the corresponding keto-form (Hommel *et al.*, 1995). As Cys7 and Asp126 act on different parts of the sugar moiety, they are located at opposite sides of the catalytic face, namely at the C-terminal ends of β-strands 1 and 6. Assuming that the mechanism postulated for TrpF wild-type also holds for the PRA isomerization reactions established on the scaffolds of HisA, HisF and HisAF, again a catalytic acid and base are required. Previously, mutational analyses and substrate docking led to the identification of aspartate residues (Asp8 in HisA and

HisAF, Asp11 in HisF), which are located at positions equivalent to Cys7 in the TrpF wild-type enzyme and likely act as general base in the PRA isomerization reaction (Leopoldseder *et al.*, 2004; Claren *et al.*, 2009). However, no plausible acid corresponding to Asp126 of TrpF has been identified, because in HisA-D127V, HisF-D130V and HisAF-D127V potential aspartate candidates were replaced by valine, allowing for the binding of the negatively charged anthranilate moiety of PRA (Leopoldseder *et al.*, 2004; Claren *et al.*, 2009).

In order to elucidate the mechanism of PRA isomerization by the HisA, HisF and HisAF scaffolds, both the hitherto

752

known and the newly generated variants were characterized by pH-dependent activity measurements, crystal structure analysis and molecular dynamics (MD) and mixed quantum and molecular mechanics (QM/MM) calculations. The results suggest that PRA is complexed by the C-terminal phosphate-binding site of these scaffolds and confirm the assignment of the catalytic bases. Moreover, our data suggest that the carboxylic acid of the anthranilate moiety of PRA initiates the Amadori rearrangement by protonating the furanose ring oxygen, in line with 'substrate-assisted catalysis' contributing to the straightforward conversion of PRA to CdRP by $(\beta\alpha)_8$-barrel proteins other than TrpF.

Materials and methods

Site-directed mutagenesis

All point mutations were introduced with the megaprimer method (Sarkar and Sommer, 1990) using oligonucleotides carrying modified codons for amino acid exchanges (underlined) or restriction sites for cloning (boldface type). The plasmid pTNA-*malE-his*AF (Claren *et al.*, 2009) was used as template for generating *his*AF_D127V. The 5′-megaprimer was produced in an initial polymerase chain reaction (PCR) using the oligonucleotides 5′-CCG TAT **CAT ATG** CTC GTT GTC CCG GCG AT-3′ (5′ *his*A_*Nde*I) and 5′-ACT CTT TTT GCA ACT ATC GCC ACG AC-3′ (3′ *his*F_D130V) and combined with 5′-ATA **GCG GCC GCG** GGA GCA TAT CTC TTC ATC AC -3′ (3′ *his*A_*Not*I) in the final PCR to amplify the entire mutant gene. The gene was cloned into the expression plasmid pET21a, using the introduced terminal restriction sites for *Nde*I and *Not*I. The plasmids pDS56/RBSII_*his*A and pET11c_*his*F (Thoma *et al.*, 1999) served as templates for generating *his*A_D169V and *his*F_D176V, respectively. Megaprimers were produced in an initial PCR using the oligonucleotides 5′-CAC ACG GAG ATC GAA AAA GTT GGC ACT CT-3′ (5′ *his*A_D169V) together with 5′-TGC **CAA GCT T**TA GCG AGC ATA TCT-3′ (3′ *his*A_*Hind*III), and 5′- AGT ATC GAC AGA GTC GGC ACA AAA TCG GGT TAC-3′ (5′ *his*F_D176V) together with 5′-GTG GTG **CAA GCT T**CA CAA CCC CTC CAG TCT CAC-3′ (3′ *his*F_*Hind*III). In the final PCR the resulting 3′-megaprimers were used together with the 5′-primers 5′ *his*A_*Nde*I and 5′- AGC **CAT ATG** CTC GCT AAA AGA ATA ATC GCG-3′ (5′ *his*F_*Nde*I) to amplify the entire mutant genes. The genes were cloned into the expression plasmid pET28a, using the introduced terminal restriction sites for *Nde*I and *Hind*III. In an analogous manner the mutants *his*A_D127V + D169V and *his*F_D130V + D176V were generated using pET28a_*his*A_D169V and pET28a_*his*F_D176V as templates. Megaprimers were produced using 5′ *his*A_*Nde*I together with the oligonucleotide 5′-AC TCT TCC ACC TCG AGT GAC CAG ACT GAA CAC-3′ (3′ *his*A_D127V), and 5′ *his*F_*Nde*I together with the oligonucleotide 3′ *his*F_D130V. In the final PCR these megaprimers were used together with 3′ *his*A_*Hind*III and 3′ *his*F_*Hind*III, respectively, to amplify the entire mutant genes. The genes were cloned into pET28a. Furthermore, *this*A_D127V was subcloned into pET28a using pTNA_*his*A_D127V (Leopoldseder *et al.*, 2004) as a template and the above mentioned terminal *his*A-primers. Finally, *trp*F was subcloned into pET21a using pQE60_ *trp*F

(Sterner *et al.*, 1996) and the oligonuleotides 5′-AGC **CAT ATG** GTC AGA GTG AAA ATC TGC-3′ (5′ *trp*F_*Nde*I) and 5′- GTG GTG **CTC GAG** CAA CCC CTT TGC ATT TTT-3′(3′ *trp*F_*Xho*I). All gene constructs were entirely sequenced to exclude inadvertent mutations.

Heterologous expression and purification of recombinant proteins

HisAF_D127V was expressed and purified as previously described (Claren *et al.*, 2009). The various pET28a_*his*A and pET28a_*his*F constructs were used to transform competent *Escherichia coli* T7 Express cells (New England Biolabs) containing the pRARE helper plasmid (Claren *et al.*, 2009). Following the inoculation of 2 l LB medium containing 150 μg/ml kanamycin, the cells were grown at 37°C until an OD$_{600}$ of 0.6 was reached. After cooling down to 30°C, gene expression was induced by adding 0.5 mM IPTG, and growth was continued overnight. Subsequently, cells were harvested by centrifugation (Sorvall/RC5B, GS3 rotor, 20 min, 4000 rpm, 4°C) and washed with 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 10 mM imidazole. The washed cells were pelletized again (Eppendorf Centrifuge 5810R, 35 min, 4000 rpm, 4°C), suspended and lysed by sonification (Branson Sonifier W-250D; 2 min, 2 s intervals, 70&per; pulse, 4°C). The successive centrifugation step (Sorvall/RC5B, SS34 rotor, 30 min, 14 000 rpm, 4°C) was used to separate the insoluble from the soluble fraction of the cell extract, from which the recombinant HisA and HisF variant proteins were purified. All variants showed very high thermal stability, which allowed for the removal of most host proteins by heat denaturation (70°C, 15 min) followed by centrifugation (Sorvall/RC5B, SS34 rotor, 45 min, 14 000 rpm, 4°C). The recombinant proteins were then further purified by Ni^{2+} affinity chromatography using their N-terminal hexa-histidine tags. For this purpose, the supernatants of the heat steps were loaded onto a HisTrapFF crude column (5 ml; GE Healthcare), which had been equilibrated with 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 10 mM imidazole. After washing the column with the equilibration buffer, bound proteins were eluted by applying a linear gradient of 10–500 mM imidazole. The HisA and HisF variants eluted between 50 and 150 mM imidazole and were >95% pure, as judged by sodium dodecyl sulphate-polyacrylamide gel electrophoresis. The yields were between 10 and 30 mg of protein per l of cell suspension. Finally, the proteins were dialyzed twice against 5 l of 50 mM Tris/HCl pH 7.5.

The plasmid pET21a_*trp*F was used to produce recombinant TrpF with the same protocol as described above for the HisA and HisF variants, except for the buffer system. Because TrpF tends to aggregate in Tris/HCl buffer, 10 mM potassium phosphate pH 7.5 was used for purification, which was exchanged for 50 mM Tris/HCl pH 7.5 with NAP5 columns (GE Healthcare) immediately before functional characterization. Yield and purity of recombinant TrpF were comparable to the HisA and HisF variants.

After subcloning *his*A_D127V + D169V into the pQE70 expression plasmid using the *Sph*I and *Hind*III restriction sites, the recombinant HisA variant was also produced in *E. coli* strain W3110 *trp*EA2, which contains the helper plasmid pDMI,1 (Jürgens *et al.*, 2000). W3110 *trp*EA2 is lacking on its chromosome the entire *trp* operon including *trp*F (Schneider *et al.*, 1981). Protein purification was

performed as described above, resulting in a similar yield and purity of HisA_D127V + D169V as obtained with the other variants.

Steady-state enzyme kinetics and ligand titration

The conversion of the TrpF substrate PRA into the product CdRP by the HisA and HisF variants was followed at 25°C under steady-state conditions by a fluorimetric assay and analyzed as described (Hommel et al., 1995; Leopoldseder et al., 2004; Claren et al., 2009). PRA was produced in situ from anthranilate and a 30-fold molar excess of phosphoribosyl-α1-pyrophosphate (PRPP) with the help of 1 μM of anthranilate phosphoribosyl transferase from yeast (yTrpD). Additionally, 2.5-μM indole-3-glycerol phosphate synthase from Thermotoga maritima (TrpC), which catalyzes the transformation of CdRP to IGP, was added to prevent product inhibition. The PRAI activities of HisA_D127V + D169V purified from either W3110 trpEA2 or T7 Express cells were identical, excluding the manipulation of the results by a contamination with TrpF from the E. coli host cells. The binding of the product analogue rCdRP to HisF_D130V + D176V was monitored by fluorescence energy transfer and analyzed as described (Bisswanger et al., 1979; Henn-Sax et al., 2002).

pH dependence of catalytic activity

In order to determine the PRAI activities of HisA_D127V, HisA_D169V, HisA_D127V + D169V and wild-type TrpF at various pH values, the enzymatic assay setup described above was modified. To assure during the individual measurements, a constant concentration of the unstable substrate PRA, which readily decomposes into anthranilate and ribose-5-phosphate, especially at acidic conditions (Kirschner et al., 1987), the assay was performed at 15°C (instead of 25°C) using 3 μM (instead of 1 μM) yTrpD and a 75-fold (instead of a 30-fold) molar excess of PRPP over anthranilate. Under these conditions the k_{cat} value of the HisA variants was determined from roughly pH 6 to 8 using the buffer systems MES/NaOH, HEPES/NaOH and Tricin/NaOH. The k_{cat} value of TrpF was determined up to pH 9.5 by using the additional buffer system Glycin/NaOH. To assure the stability of the characterized enzymes at acidic and basic conditions, HisA_D127V + D169V and TrpF were pre-incubated at the lowest and highest applied pH value for 15 min, and subsequently the activities were measured at pH 7.5. Both enzymes were equally active with and without pre-incubation. To ensure that the used substrate concentrations (200 μM PRA for the HisA variants, 10–100 μM PRA for wild-type TrpF) were saturating at each pH value, individual reactions were repeated with higher concentrations of substrate, and no change in the rates was observed. In the case of TrpF entire progress curves were recorded and analyzed with the program COSY using the integrated form of the Michaelis-Menten equation taking product inhibition into account (Eberhard, 1990; Hommel et al., 1995). The latter was necessary because TrpC is unable to prevent product inhibition by CdRP at pH values of >8.5.

The data from the k_{cat} versus pH profile for TrpF were fit to equation (1), whereas those for HisA_D169V and HisA_D127V + D169V were fit to equation (2) (Shim and Benkovic, 1999), where c is the maximum activity, H is the

754

hydrogen ion concentration, and K_{a1} and K_{a2} are the acid dissociation constants:

$$\log(k_{cat}) = \log\left[\frac{c}{(1 + H/K_{a1} + K_{a2}/H)}\right] \quad (1)$$

$$\log(k_{cat}) = \log\left[\frac{c}{(1 + K_{a2}/H)}\right] \quad (2)$$

Crystallisation, data collection and refinement

Initial crystallization trials of HisF_D130V + D176V in complex with rCdRP were carried out at 292 K using the sitting-drop vapor-diffusion method in 96-well plates (Greiner) at the EMBL Hamburg High-throughput Crystallization Facility (Mueller-Dieckmann, 2006). Crystals were obtained under various conditions. Further optimization of these conditions was performed manually in 24-well plates (Qiagen) using the hanging-drop vapor-diffusion method at 292 K. In the final condition drops contained 1 μL of 1.8 mM sodium chloride and 0.1 M sodium acetate at pH 5.0 mixed with 1 μL of protein (14 mg/ml), to which 1 mM rCdRP was added. Equilibration was performed against 500 μL of reservoir buffer. After an additional soaking step, in which crystals were put into reservoir buffer containing 2.7 mM rCdRP for about 1 min, the crystals were flash frozen in liquid nitrogen, and data of single crystals were collected at the synchrotron beamline ID29 (ESRF) at 100 K. The data were processed using XDS (Kabsch, 1993), and the data quality assessment was done using phenix.xtriage (Adams et al., 2002). Molecular replacement was performed with MOLREP within the CCP4i suite (Potterton et al., 2004) using the coordinates of HisF from Thermotoga maritima as a search model (PDB ID code 1thf). Initial refinement was performed using REFMAC (Murshudov et al., 1997). The model was further improved in several refinement rounds using automated restrained refinement with the program PHENIX (Adams et al., 2002) and interactive modeling with Coot (Emsley and Cowtan, 2004). The ligand rCdRP was identified in the first electron density map based on molecular replacement phases and was included in the model using the program Ligandfit embedded in PHENIX (Adams et al., 2002). The data collection and refinement statistics are summarized in Supplementary Table SI. The final model was analyzed using the program MolProbity (Davis et al., 2007).

MD simulations and QM/MM calculations

MD simulations were performed by using the AMBER10 suite of programs (Case et al., 2005) and employing force fields Amber03 for proteins (Duan et al., 2003) and GAFF for ligands (Wang et al., 2004). The simulations were based on the crystal structures of HisF_D130V + D176V (PDB ID code 4ewn), HisA wild type (PDB ID code 1qo2), HisA_D127V (PDB ID code 2cff) and HisA_H75Y + F111S + D127V + D169V (PDB ID code 2w79). The latter variant will be termed HisA_D127V + D169V hereafter, as the mutations H75Y and F111S are remote from the active site and thus do not affect the calculations. The structures were initially protonated using PROPKA 1.0 (Li et al., 2005), the acid moiety of the substrate PRA was treated as glutamate for pK_a estimation. The protein/PRA complexes

were solvated in a cubic box filled with Tip3P (Tan *et al.*, 2007) molecules. The substrate PRA was parameterized according to the standard GAFF procedure, assigning partial charges by using a HF/631G* RESP (Cieplak *et al.*, 1995) procedure. A stepwise minimization/equilibration procedure was applied to all structures with and without bound PRA, minimizing first the solvent and then the entire system for 2500 steps using periodic boundaries and a cutoff of 10 Å. During heating from 0 to 300 K over 20 ps, the protein backbone and the ligand were restrained by a force 10 kcal/molA^2, before an additional heating of the unrestrained system over 100 ps was performed. All bonds to hydrogen atoms were restrained using SHAKE (Miyamoto and Kollman, 1992) during minimization and MD simulations. Productive MD simulations under NPT-ensemble conditions (Martyna *et al.*, 1999) were performed at 300 K and 1 bar pressure over 3–10 ns using Langevin temperature control (ntt = 3) and isotropic position scaling to maintain the pressure (NTP = 1) using a relaxation time of 2 ps (TAUP = 2.0). The simulation of different PRA-binding modes in the binding pocket of all protein variants was at least performed in triplicates for 3 ns. The meaning of ntt, NTP and TAUP is given in the AMBER 11 manual.

QM/MM simulations of PRA-ring opening reactions in pure solvent and in the active site were performed using QMMM SCCDFTB Hamiltonian (Seabra *et al.*, 2007) for the substrate atoms excluding the phosphate group and the adjacent water molecules or active-site residues within 6 Å from the substrate. The energy along the ring opening reaction coordinate was scanned by Adaptively Biased MD (Babin *et al.*, 2008) using a linear combination of distances of the formed and broken bonds for all reaction pathways. The Jarzynski averaged potential of mean force was calculated to estimate the free energy of the respective reaction. Since proton transfer reactions were analyzed, QMSHAKE was turned off.

Results

Establishing PRA isomerase activity on the (βα)$_8$-barrel scaffolds of HisA, HisF and HisAF

PRA isomerization (TrpF) activity was previously established on the (βα)$_8$-barrel scaffolds of HisA and HisF by introducing an aspartate-to-valine exchange at the C-terminal end of β-strand 5 (D127V in HisA, D130V in HisF) (Jürgens *et al.*, 2000; Leopoldseder *et al.*, 2004). Analogously, the expressed and purified mutant HisAF_D127V exhibited moderate PRAI activity (Table I). Moreover, PRAI activity in HisAF was also generated by an aspartate-to-valine exchange at the C-terminal end of β-strand 6 (D173V) (Claren *et al.*, 2009). Remarkably, the catalytic efficiency (k_{cat}/K_M^{PRA}) of variant HisAF_D127V + D173V, which contains both activating exchanges, was found to be 247-fold higher than the sum of the k_{cat}/K_M^{PRA} values of HisAF_D127V and HisAF_D173V (Table I) (Claren *et al.*, 2009). We tested whether the two equivalent aspartate-to-valine exchanges promote PRA isomerization in a more-than-additive manner also in the backgrounds of HisA and HisF. For this purpose, in addition to the known HisA_D127V and HisF_D130V variants, HisA_D169V and HisF_D176V as well as HisA_D127V + D169V and HisF_D130V + D176V were produced and characterized by steady-state kinetics. While HisA_D169V

Table I. Steady-state kinetic constants for the PRA isomerization activities of HisA, HisF and HisAF variants in comparison to wild-type TrpF from *Thermotoga maritima*

Protein	k_{cat}, min^{-1}	K_M^{PRA}, μM	k_{cat}/K_M^{PRA}, M^{-1} s^{-1}	Synergy factor[e]
TrpF wild-type[a]	222	0.28	1.3×10^7	
HisA variants				
HisA_D127V[b]	0.52	74	1.2×10^2	
HisA_D169V	1.7	36	7.8×10^2	
HisA_D127V + D169V	31.3	36	1.5×10^4	16
HisF variants				
HisF_D130V[b]	0.011	74	2.5	
HisF_D176V	ND	ND	ND	
HisF_D130V + D176V	0.073	18	69	28
HisAF variants				
HisAF_D127V	0.013	94[d]	2.3	
HisAF_D173V[c]	0.017	53	5.4	
HisAF_D127V + D173V[c]	1.9	19	1.9×10^3	247

Experimental conditions: 50 mM Hepes, pH 7.5, 4 mM MgCl$_2$, 4 mM EDTA, 2 mM DTT, 25°C.
The k_{cat} and K_M^{PRA} values for the HisA and HisF variants are the mean deduced from two PRA saturation curves, which were recorded with different protein preparations and are shown in Supplementary Fig. S1.
ND, no detectable activity in the presence of 50 μM protein and 150 μM PRA.
[a]Data are taken from Sterner *et al.* (1996).
[b]Data are taken from Leopoldseder *et al.* (2004).
[c]Data are taken from Claren *et al.* (2009).
[d]K_M^{PRA} value was determined in presence of >0.8 mM orthophosphate.
[e]Fold increase of k_{cat}/K_M^{PRA} of the double mutant compared to the sum of the k_{cat}/K_M^{PRA} values of the corresponding single mutants.

showed PRA isomerization activity, for HisF_D176V no measurable conversion of PRA into CdRP was observed. Interestingly, in accordance with the results obtained for HisAF, the k_{cat}/K_M^{PRA} values of the variants HisA_D127V + D169V and HisF_D130V + D176V were 16-fold and 28-fold higher, respectively, than the sum of the k_{cat}/K_M^{PRA} values of the variants containing the single exchanges (Table I). The comparison of the individual k_{cat} and K_M^{PRA} values shows that this effect can be mainly ascribed to the disproportionally high increase of the turnover numbers that is caused by combining the two substitutions (Table I; Supplementary Fig. S1).

Catalytic activities of wild-type TrpF and the HisA variants as a function of pH

To obtain information about the mechanisms of PRA isomerization, the catalytic activities of wild-type TrpF and the HisA variants, which performed better than the respective HisF and HisAF variants (Table I), were measured as a function of pH. In TrpF, Cys7 and Asp126 act as catalytic base and acid, respectively (Fig. 2) (Henn-Sax *et al.*, 2002). In accordance with this notion, the k_{cat} value of TrpF shows a bell-shaped dependency on pH (Fig. 3A). The fit of equation (1) (see Materials and methods section) to the data yielded pK$_a$ values of 7.6 ± 0.1 and 8.9 ± 0.1 for the ascending (base) and descending (acid) limbs, respectively. While the former can readily be assigned to Cys7, the latter might be too high for Asp126, even when assuming a strong pK$_a$ perturbation of its carboxylic acid group by neighboring residues (Harris and Turner, 2002). An alternative explanation for the decrease of activity at high pH could be the release of a proton from the phosphate group of PRA, which might be

755
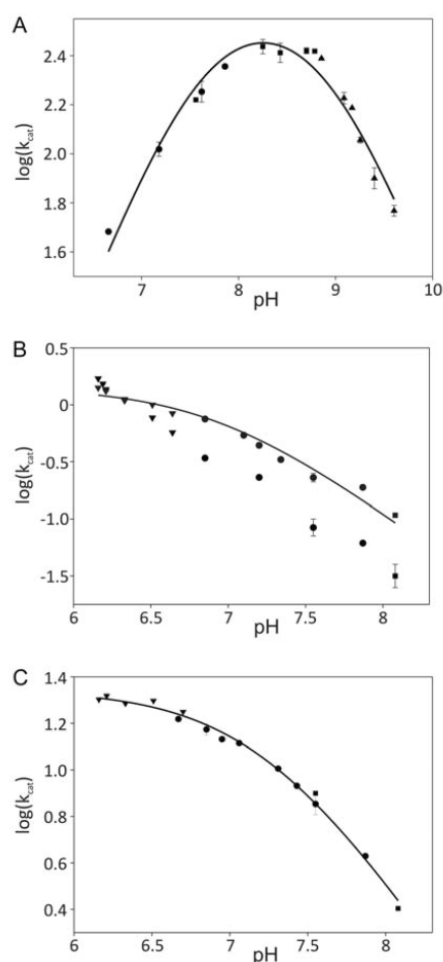
Fig. 3. Dependence of the turnover number (k_{cat}) on pH. Data recorded at 15°C are shown for wild-type TrpF (**A**), HisA_D127V (**B**), HisA_D169V (B, fitted curve) and HisA_D127V + D169V (**C**). MES/NaOH (triangles down), HEPES/NaOH (circles), Tricin/NaOH (squares) and Glycin/NaOH (triangles up) were used as buffer systems. Buffer overlap was ensured in case of TrpF and HisA_D127V + D169V. Standard deviations arising from at least three independent measurements are indicated, unless error bars are smaller than the symbols for the data points. Solid lines in panel A show the results of fitting the data for wild-type TrpF to equation 1, yielding two pK_a values of 7.6 ± 0.1 and 8.9 ± 0.1; solid lines in panels B and C show the results of fitting the data for HisA_D169V and HisA_D127V + D169V to equation 2, yielding pK_a values of 6.9 ± 0.1 and 7.2 ± 0.1, respectively.

rate limiting for catalysis. However, in this case the pK_a value for the second deprotonation step, which is 7.20 in free orthophosphate, would also have to be strongly perturbed when PRA is bound to TrpF (Czekster *et al.*, 2009). In contrast to TrpF wild-type, the k_{cat} versus pH profile of the HisA variants merely consist of a descending limb, indicating that the protonated form of a single acidic group catalyzes the rate-limiting step within the entire analyzed pH range from 6.15 to 8.25 (Fig. 3B and C). Unfortunately, kinetics could not be measured at pH values <6.15, due to the rapid degradation of the substrate PRA at acidic conditions (Kirschner

756

et al., 1987). The k_{cat} values of HisA_D169V and HisA_D127V + D169V reached a plateau at pH values of <6.5. The data could be fit with equation (2) (see Materials and methods section), yielding apparent pK_a values for the catalytic acid of 6.9 ± 0.1 and 7.2 ± 0.1, respectively. However, the molecular interpretation of the apparent pK_a value of HisA_D169V is difficult as the slope of the linear part of the curve deviates from unity. The activity of HisA_D127V continuously increased down to the lowest analyzed pH value, which indicates that the pK_a value of its acid must be <6. The virtual absence of an ascending limb for all HisA variants suggests that the acid is rate limiting for catalysis even at low pH where only minute concentrations of the active base are populated. Taken together, the pH dependences of enzymatic activity reveal distinct differences between wild-type TrpF and the analyzed HisA variants, which may be caused by modifications of the catalytic mechanisms of PRA isomerization and the groups acting as acid and base.

Crystal structure analysis of HisF_D130V+D176V with bound rCdRP

The previously solved crystal structure of TrpF with bound product analogue rCdRP provided insights into the reaction mechanism of the wild-type enzyme and allowed for the assignment of Cys7 and Asp126 as catalytic base and acid, respectively (Henn-Sax *et al.*, 2002). We anticipated that crystal structures of our catalytically active variants with bound rCdRP could help to understand how PRA isomerization is catalyzed by the HisA, HisF and HisAF variants. Titration studies showed that HisF_D130V + D176V binds rCdRP with a 1:1 stoichiometry and a thermodynamic dissociation constant (K_D) of only 0.2 μM (Supplementary Fig. S2), which corresponds to an at least 15-fold higher protein-ligand affinity than observed with all other tested variants (data not shown; Jürgens *et al.*, 2000; Claren *et al.*, 2009). Based on this finding HisF_D130V + D176V was crystallized in the presence of rCdRP, and the structure of the complex was solved with a resolution of 1.9 Å (Fig. 4, PDB ID code 4ewn). Although rCdRP contains a linearized sugar moiety and is therefore less compact than PRA (Fig. 2) the obtained structure provides insights into substrate binding and catalysis. HisF_D130V + D176V contains two phosphate-binding sites, which are formed by βα-loops 3 and 4 and βα-loops 7 and 8, respectively. Remarkably, the phosphate moiety of rCdRP is exclusively bound to the C-terminal-binding site, where it interacts with the backbone atoms of Gly177, Gly203 and Ala224, as well as with the hydroxyl group of Ser225 (Fig. 4A). As a consequence of this binding geometry, the carboxylate group of Asp11 at the C-terminal end of β-strand 1 is located in a distance of 4.9 Å from the C2 atom of rCdRP, which is similar to the distance between the sulfhydryl group of Cys7 and the C2 atom of rCdRP in the TrpF complex (Henn-Sax *et al.*, 2002). This finding suggests that Asp11 in HisF_D130V + D176V fulfills the same role as Cys7 in TrpF, namely acting as the catalytic base that abstracts a proton from the Schiff's base intermediate (Fig. 2). In accordance with this conclusion, previous modeling and site-directed mutagenesis studies indicated that the equivalent residue Asp8 acts as catalytic base upon PRA isomerization by the variant HisA_H75Y + F111S + D127V + D169V (Claren *et al.*, 2009). However,
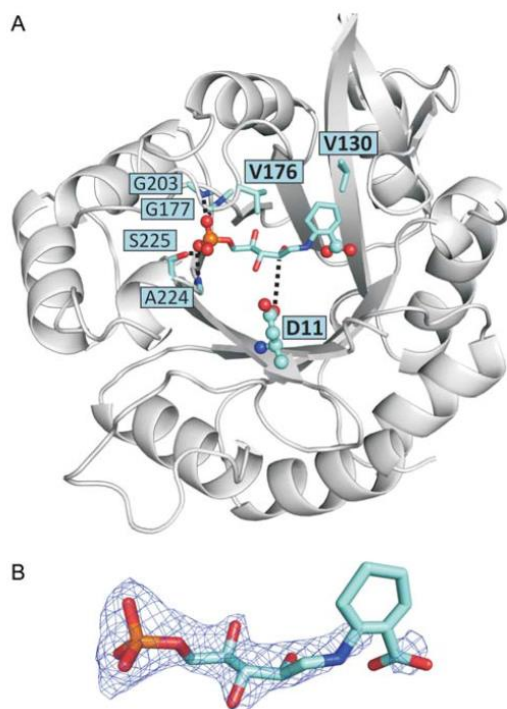
Fig. 4. Crystal structure of HisF_D130V + D176V with bound product analog rCdRP. (**A**) Ribbon diagram with relevant residues highlighted in cyan. The rCdRP molecule is bound to the C-terminal phosphate binding site via hydrogen bonds to backbone atoms of G177, G203, A224, and to the side chain of S225. These hydrogen bonds are indicated by dotted lines. The distance of 4.9 Å between the carboxylate group of the putative catalytic base D11 and carbon atom C2′ of rCdRP is also indicated by a dotted line. Candidates for the catalytic acid (D11, phosphate and carboxylate group of ligand) are depicted as spheres. (**B**) The omit electron density (F_o-F_c) for the ligand rCdRP is contoured at 1.5σ.

the HisF_D130V + D176V:rCdRP crystal structure does not afford an obvious candidate that could act as catalytic acid. In wild-type TrpF residue Asp126 at the C-terminal end of β-strand 6 protonates the furanose ring oxygen (Fig. 2; Henn-Sax et al., 2002) but in the HisA, HisF and HisAF variants all aspartate residues at the C-terminal ends of β-strands 5 and 6 are replaced by valine, probably to allow for the binding of PRA (Leopoldseder et al., 2004; Claren et al., 2009). Three alternative candidates for the catalytic acid come to mind: Asp11, as well as the phosphate and the anthranilate moieties of the substrate PRA. If Asp11 were to play a dual role by acting not only as catalytic base but also as catalytic acid, PRA would have to rotate by 180°C with respect to the orientation of rCdRP observed in the crystal structure, in order to allow for the protonation of the furanose ring oxygen. However, this additional binding mode, in which the anthranilate moiety points down the central β-barrel, was shown to be energetically highly unfavorable as it leads to the ejection of the ligand from the protein's binding pocket during MD simulations (data not shown). The protonation of the furanose ring oxygen by the phosphate moiety of PRA would require that its hydrogen bonds with the protein are disintegrated during catalysis. However, the

high electron density observed for the phosphate part of rCdRP suggests that it is tightly anchored and rigid (Fig. 4B). The remaining candidate is the anthranilic acid moiety of PRA, which was already previously considered as catalytic acid (Wright et al., 2008; Claren et al., 2009). The electron density of the anthranilate ring is rather low (Fig. 4B) and additional density is found around it (data not shown) suggesting a considerable grade of flexibility. Therefore, it might be feasible for the anthranilic acid moiety to transfer a proton to the furanose ring oxygen, which would constitute a case of 'substrate-assisted catalysis' (Dall'Acqua and Carter, 2000).

MD simulations and QM/MM calculations

The hypothesis that the anthranilic acid moiety of PRA acts as catalytic acid was further evaluated by MD simulations and QM/MM calculations, which have been shown to contribute to the elucidation of enzyme mechanisms (Alexandrova et al., 2008; Lu et al., 2009; Kim et al., 2010; Kiss et al., 2010; Jitonnom et al., 2011; Kong et al., 2011). As with the pH-dependency measurements, the MD simulations focused on the most active HisA variants with the wild-type HisA protein serving as a reference. In order to allow for substrate-assisted catalysis to occur, (i) the protonation of the furanose ring oxygen of PRA by the carboxylic acid group of its anthranilic acid moiety must be geometrically feasible and (ii) the pK_a value of the carboxylic acid group needs to be upshifted by the protein environment. Therefore, as a first step, two different binding modes for the flexible anthranilic acid moiety of PRA were considered (Fig. 5A). In binding mode 1 the carboxylic acid group points toward the sugar ring and thus should be able to act catalytically, with the interaction probably mediated by a water molecule. In binding mode 2 the anthranilic acid moiety is rotated through approximately 180° resulting in a large spatial separation of its carboxylic acid group and the furanose oxygen. While HisA wild-type shows almost identical binding energies for both anthranilic acid conformations, the HisA variants show a clear preference for the productive binding mode 1 (Fig. 5B). In a second step, the pK_a values for anthranilic acid orientated in binding mode 1 were calculated using PROPKA (Li et al., 2005). For wild-type HisA, the predicted pK_a of 1.3 is even lower than the pK_a of 2.05 measured for anthranilic acid in aqueous solution (Rappopory, 1984). Significantly higher pK_a values between 4.2 and 5.8 were calculated for the anthranilate moiety of PRA when bound to the HisA variants (Fig. 5C), probably due to the removal of the negative charges and the generation of a more hydrophobic environment brought about by the aspartate-to-valine exchanges. Although these pK_a values are somewhat lower than the pK_a values of the catalytic acid of HisA_D169V and HisA_D127V + D169V as deduced from pH-dependency measurements (Fig. 3B and C), the data are in line with the hypothesis that the anthranilic acid moiety protonates the ring oxygen of PRA.

The outcome of the MD simulations suggests that the two exchanges D127V and D169V contribute to PRA isomerization by the HisA variants in different manner: D127V is mainly responsible for the positive discrimination of binding mode 1 (Fig. 5B), whereas D169V appears to be more important for elevating the pK_a value of anthranilic acid (Fig. 5B). This notion is illustrated in representative

757

B.Reisinger *et al.*
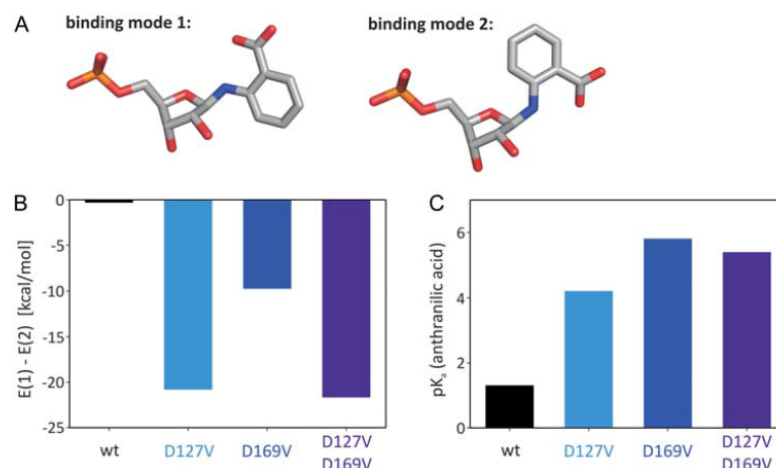
Fig. 5. Molecular dynamics simulations of wild-type (wt) HisA and HisA variants containing the indicated exchanges. (**A**) Structure of the substrate PRA with two different anthranilic acid binding modes. (**B**) Resulting energy differences E(1)−E(2) of simulations with binding modes 1 and 2. For each mode, at least three independent energy values were determined. A low energy value indicates a strong preference for binding mode 1 versus 2. A t-test showed that the energy values attributed to the two modes are different with statistical significance for all three variants ($P \leq 0.029$). (**C**) pK$_a$ values of the anthranilic acid in binding mode 1 as determined by PROPKA (Li *et al.*, 2005). PDB ID codes of used structures: HisA wild-type and HisA_D169V (modeled onto the wild-type scaffold): 1qo2; HisA_D127V: 2cff; HisA_D127V + D169V: 2w79.
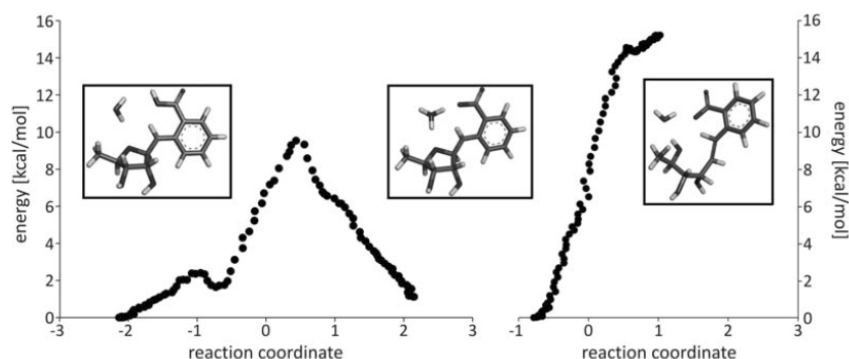
Fig. 6. QM/MM calculations of HisA_D127V + D169V with PRA bound in binding mode 1. With PRA in binding mode 1 (see Fig. 6A), a substrate-internal proton transfer includes two steps: protonation of the water molecule by the substrate's carboxylic acid group (left reaction coordinate), and the subsequent protonation of the furanose ring oxygen (right reaction coordinate). Reaction states are shown as insets. PDB ID code for HisA_D127V + D169V: 2w79.

structural models, which substantiate that the D127V exchange is important for a high affinity of PRA in binding mode 1 and that the D169V exchange increases the hydrophobicity around the carboxylate of anthranilic acid (Supplementary Fig. S3). These findings explain why HisA_D127V + D169V exhibits a significantly higher catalytic activity than HisA_D127V and HisA_D169V, although the calculations are not precise enough to account for the synergistic effect of the two exchanges in a quantitative manner (Table I). Additional MD simulations performed with HisF_D130V + D176V confirm that the activating exchanges come along with a preference for binding mode 1 and an upshift of the pK$_a$ of anthranilate, supporting the view that substrate-assisted catalysis is the basis for PRA binding and isomerization also in the background of the HisF scaffold (Supplementary Fig. S4).

Finally, QM/MM calculations of HisA_D127V + D169V with bound PRA were performed (Alexandrova *et al.*, 2008; Kiss *et al.*, 2010; Jitonnom *et al.*, 2011). The protonation of the furanose ring oxygen was calculated without and with the assistance of a water molecule positioned between the anthranilic acid moiety and the sugar ring. Since direct protonation is associated with a high energy barrier (>50 kcal/mol), we focused on proton transfer via a bridging water molecule. With PRA in binding mode 1 the initial generation of the hydronium ion requires the crossing of an energy barrier of 10 kcal/mol, and the subsequent proton transfer leading to ring-opening costs 15 kcal/mol (Fig. 6). These energies are consistent with other water-mediated proton transfers (Dourado *et al.*, 2008; Lutz *et al.*, 2011) and thus the reaction pathway is feasible from an energetic point of view. However, when starting off with PRA in its

758

deprotonated form or in binding mode 2, the activation energies in QM/MM calculations get virtually too high (40 and 50 kcal/mol) (Supplementary Fig. S5).

Discussion

In organisms producing the essential amino acid tryptophan, the Amadori rearrangement of the aminoaldose PRA into the corresponding ketose CdRP is catalyzed by the $(\beta\alpha)_8$-barrel enzyme TrpF. The underlying acid–base mechanism is supported by two amino acids, which are positioned on opposite sides of the central β-barrel: Cys7 and Asp126 (Henn-Sax et al., 2002). Interestingly, PRA isomerization activity can readily be established on the two $(\beta\alpha)_8$-barrel enzymes HisA and HisF, both of which are involved in histidine biosynthesis, as well as on the artificial chimera HisAF. In each case the substitutions of two aspartate residues by valines promote catalysis in a synergistic manner. As TrpF and HisA catalyze chemically equivalent reactions and HisF is related to HisA in terms of structure and sequence (Lang et al., 2000), it is plausible to assume a TrpF-like reaction mechanism for PRA isomerization by the HisA, HisF and HisAF variants. However, while in all three cases the presumable base was found to be an aspartate residue that is located at the C-terminal end of β-strand 1 and superposes well with Cys7 of TrpF (Leopoldseder et al., 2004; Claren et al., 2009), all potential acid candidates are replaced by valine in the best performing variants. In addition, studies of pH dependences revealed the action of the catalytic acid being rate-limiting within the HisA variants. Hence, distinct differences seem to exist between the reaction mechanisms of TrpF wild-type and the variants, more precisely in the nature of the general acid. The crystal structure of HisF_D130V + D176V with bound product analog rCdRP confirmed an earlier hypothesis that the ligand is anchored at the C-terminal phosphate-binding site (Claren et al., 2009) and suggested functional groups that could act as catalytic acid. The most promising candidate turned out to be the anthranilic acid moiety of the substrate PRA whose carboxylic acid could initiate the isomerization reaction by protonating the furanose ring oxygen.

A straightforward experimental approach to test this hypothesis would be to perform the reaction with an alternate substrate in which the carboxylate of anthranilic acid is replaced by a non-reactive group, for example the amide (Kosloff and Selinger, 2001). However, our efforts to enzymatically synthesize phosphoribosyl-anthranilic amide from anthranilic amide and PRPP by anthranilate phosphoribosyl transferase failed. Therefore, we turned to MD simulations to further substantiate our hypothesis. The results showed that in contrast to the wild-type HisA protein, the variant HisA_D127V has a clear preference for a binding geometry of PRA that allows for a proton transfer reaction to occur. Moreover, the calculations suggest that the pK_a value of the carboxylate group of anthranilate is significantly up-shifted when bound to HisA_D169V, an important prerequisite for its ability to act as catalyst at physiological pH conditions. Both fundamental features are combined in HisA_D127V + D169V, which is consequently far more active than the variants that contain only a single exchange. QM/MM calculations performed with HisA_D127V + D169V demonstrated that the participation of the carboxylic acid in the reaction

mechanism indeed is also feasible from an energetic point of view. Taken together, our results suggest that a non-natural reaction mechanism involving substrate-assisted catalysis is used by the HisA, HisF and HisAF variants for the isomerization of PRA.

Several experimental observations underpin the proposed non-natural reaction mechanism. First of all, the ease and frequency with which PRA isomerase activity has been established on several protein scaffolds is striking. This applies not only to the naturally occurring $(\beta\alpha)_8$-barrel enzymes HisA and HisF, but also to the artificial, initially inactive HisAF protein (Jürgens et al., 2000; Leopoldseder et al., 2004; Claren et al., 2009). In addition, we recently established PRA isomerization activity on the scaffold of the α-subunit of tryptophan synthase, another $(\beta\alpha)_8$-barrel enzyme from tryptophan biosynthesis (Evran et al., 2012). Most notably, glutamine phosphoribosylpyrophosphate amidotransferase (PurF) unexpectedly showed promiscuous and evolvable PRA isomerase activity (Patrick and Matsumura, 2008). PurF is unrelated to TrpF in sequence, tertiary structure and catalytic mechanism. However, it contains a preexisting molecular recognition site for phosphoribosylated substrates, which can be recruited for the binding of PRA. While again an aspartate residue is well positioned to act as catalytic base, a potential acid candidate is less obvious (Patrick and Matsumura, 2008). It therefore seems tempting to assume that also PRA activity of PurF is based on the substrate's anthranilic acid moiety acting as catalytic acid.

Moreover, mutational studies performed with the TrpF enzyme are in accordance with the assumption of substrate-assisted catalysis. While no residual activity could be measured after removing the catalytic base by the C7A exchange, the replacement of the general acid by the D126N exchange still allowed for a moderate catalytic efficiency of about $10^2 \, M^{-1} \, s^{-1}$ (Henn-Sax et al., 2002). Similar results were obtained with the bisubstrate-specific isomerase PriA from *Mycobacterium tuberculosis*, which is capable of catalyzing both the HisA and TrpF reaction via two aspartate residues (Asp11 acting as base, and Asp175 acting as acid) (Due et al., 2011). Remarkably, the k_{cat}/K_M^{PRA} value of PriA_D175N remains as high as $10^4 \, M^{-1} \, s^{-1}$, whereas the k_{cat}/K_M^{PRA} for PriA_D11N is only $50 \, M^{-1} \, s^{-1}$. The extant activities of both TrpF_D126N and PriA_D175N may arise from an escape route, in which the substrate PRA takes over the task of the catalytic acid.

Up to now the phenomenon of substrate-assisted catalysis has been discovered in naturally occurring enzymes like GTPases, type II restriction endonucleases or aminoacyl-tRNA synthetases (Dall'Acqua and Carter, 2000; Hussain et al., 2010; Kosloff and Selinger, 2001; So et al., 2011), as well as in engineered serin proteases or GTPases (Dall'Acqua and Carter, 2000; Kosloff and Selinger, 2001). In extension of these hydrolytic reactions, we have shown here that engineered sugar isomerization is presumably facilitated by the participation of the carboxylic acid group of the substrate. Nevertheless, this non-native reaction mechanism affords catalytic efficiencies that are lower compared with the wild-type enzyme by two orders of magnitude at best (Claren et al., 2009). Hence, the mechanism of substrate-assisted catalysis might have been a starting point or functional intermediate in the course of the evolution of enzyme-catalyzed PRA isomerization (Dall'Acqua and Carter, 2000).

759

66

B.Reisinger *et al.*

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

Funding

References

Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K. and Terwilliger,T.C. (2002) *Acta Crystallogr. D.*, **58**, 1948–1954.

Alexandrova,A.N., Röthlisberger,D., Baker,D. and Jorgensen,W.L. (2008) *J. Am. Chem. Soc.*, **130**, 15907–15915.

Babin,V., Roland,C. and Sagui,C. (2008) *J. Chem. Phys.*, **128**, 134101.

Bisswanger,H., Kirschner,K., Cohn,W., Hager,V. and Hansson,E. (1979) *Biochemistry*, **18**, 5946–5953.

Case,D.A., Cheatham,T.E., Darden,T., Gohlke,H., Luo,R., Merz,K.M., Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) *J. Comput. Chem.*, **26**, 1668–1688.

Cieplak,P., Cornell,W.D., Bayly,C. and Kollman,P.A. (1995) *J. Comput. Chem.*, **16**, 1357–7377.

Claren,J., Malisi,C., Höcker,B. and Sterner,R. (2009) *Proc. Natl. Acad. Sci. USA.*, **106**, 3704–3709.

Czekster,C.M., Neto,B.A.D., Lapis,A.A.M., Dupont,J., Santos,D.S. and Basso,L.A. (2009) *Arch. Biochem. Biophys.*, **486**, 19–26.

Dall'Acqua,W. and Carter,P. (2000) *Protein Sci.*, **9**, 1–9.

Davis,I.W., Leaver-Fay,A., Chen,V.B., *et al.* (2007) *Nucleic Acids Res.*, **35**, W375–383.

Dourado,D.F.A.R., Fernandes,P.A., Mannervik,B. and Ramos,M.J. (2008) *Chemistry*, **14**, 9591–9598.

Duan,Y., Wu,C., Chowdhury,S., *et al.* (2003) *J. Comput. Chem.*, **24**, 1999–2012.

Due,A.V., Kuper,J., Geerlof,A., von Kries,J.P. and Wilmanns,M. (2011) *Proc. Natl. Acad. Sci. USA*, **108**, 3554–3559.

Eberhard,M. (1990) *CABIOS*, **6**, 213–221.

Emsley,P. and Cowtan,K. (2004) *Acta Crystallogr. D*, **60**, 2126–2132.

Evran,S., Telefoncu,A. and Sterner,R. (2012) *Protein Eng. Des. Sel.*, **25**, 285–293.

Harris,T.K. and Turner,G.J. (2002) *Iubmb Life*, **53**, 85–98.

Henn-Sax,M., Thoma,R., Schmidt,S., Hennig,M., Kirschner,K. and Sterner,R. (2002) *Biochemistry*, **41**, 12032–12042.

Hommel,U., Eberhard,M. and Kirschner,K. (1995) *Biochemistry*, **34**, 5429–5439.

Hussain,T., Kamarthapu,V., Kruparani,S.P., Deshmukh,M.V. and Sankaranarayanan,R. (2010) *Proc. Natl. Acad. Sci. USA*, **107**, 22117–22121.

Jiang,L., Althoff,E.A., Clemente,F.R., *et al.* (2008) *Science*, **319**, 1387–1391.

Jitonnom,J., Lee,V.S., Nimmanpipug,P., Rowlands,H.A. and Mulholland,A.J. (2011) *Biochemistry*, **50**, 4697–4711.

Jürgens,C., Strom,A., Wegener,D., Hettwer,S., Wilmanns,M. and Sterner,R. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 9925–9930.

Kabsch,W. (1993) *J. Appl. Cryst.*, **26**, 795–800.

Kim,Y., Zhou,M., Moy,S., Morales,J., Cunningham,M.A. and Joachimiak,A. (2010) *J. Mol. Biol.*, **404**, 127–137.

Kirschner,K., Szadkowski,H., Jardetzky,T.S. and Hager,V. (1987) *Methods Enzymol.*, **142**, 386–397.

Kiss,G., Röthlisberger,D., Baker,D. and Houk,K.N. (2010) *Protein Sci.*, **19**, 1760–1773.

Kong,X., Ouyang,S., Liang,Z., Lu,J., Chen,L., Shen,B., Li,D., Zheng,M., Li,K.K., Luo,C., *et al.* (2011) *PLoS One*, **6**, e25444.

Kosloff,M. and Selinger,Z. (2001) *Trends Biochem. Sci.*, **26**, 161–166.

Lang,D., Thoma,R., Henn-Sax,M., Sterner,R. and Wilmanns,M. (2000) *Science*, **289**, 1546–1550.

Leopoldseder,S., Claren,J., Jürgens,C. and Sterner,R. (2004) *J. Mol. Biol.*, **337**, 871–879.

Li,H., Robertson,A.D. and Jensen,J.H. (2005) *Proteins, Struct. Funct. Bioinform.*, **61**, 704–721.

List,F., Sterner,R. and Wilmanns,M. (2011) *ChemBioChem*, **12**, 1487–1494.

Lu,T., Tan,H.W., Lee,D., Chen,G.J. and Jia,Z.C. (2009) *Biochemistry*, **48**, 7986–7995.

Lutz,S., Tubert-Brohman,I., Yang,Y.G. and Meuwly,M. (2011) *J. Biol. Chem.*, **286**, 23679–23687.

Martyna,G.J., Hughes,A. and Tuckerman,M.E. (1999) *J. Chem. Phys.*, **110**, 3275–3290.

Miyamoto,S. and Kollman,P.A. (1992) *J. Comput. Chem.*, **13**, 952–962.

Mueller-Dieckmann,J. (2006) *Acta Crystallogr. D.*, **62**, 1446–1452.

Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) *Acta Crystallogr. D.*, **53**, 240–255.

Nagano,N., Orengo,C.A. and Thornton,J.M. (2002) *J. Mol. Biol.*, **321**, 741–765.

Patrick,W.M. and Matsumura,I. (2008) *J. Mol. Biol.*, **377**, 323–336.

Potterton,L., McNicholas,S., Krissinel,E., Gruber,J., Cowtan,K., Emsley,P., Murshudov,G.N., Cohen,S., Perrakis,A. and Noble,M. (2004) *Acta Crystallogr. D*, **60**, 2288–2294.

Rappopory,Z. (1984) *CRC Handbook of Tables for Organic Compound Identification*. Cleveland, OH: CRC Press.

Röthlisberger,D., Khersonsky,O., Wollacott,A.M., *et al.* (2008) *Nature*, **453**, 190–195.

Sarkar,G. and Sommer,S.S. (1990) *BioTechniques*, **8**, 404–407.

Schneider,W.P., Nichols,B.P. and Yanofsky,C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2169–2173.

Seabra,G.D., Walker,R.C., Elstner,M., Case,D.A. and Roitberg,A.E. (2007) *J. Phys. Chem. A*, **111**, 5655–5664.

Shim,J.H. and Benkovic,S.J. (1999) *Biochemistry*, **38**, 10024–10031.

So,B.R., An,S., Kumar,S., Das,M., Turner,D.A., Hadad,C.M. and Musier-Forsyth,K. (2011) *J. Biol. Chem.*, **286**, 31810–31820.

Sterner,R. and Höcker,B. (2005) *Chem. Rev.*, **105**, 4038–8055.

Sterner,R., Kleemann,G.R., Szadkowski,H., Lustig,A., Hennig,M. and Kirschner,K. (1996) *Protein Sci.*, **5**, 2000–2008.

Tan,C., Tan,Y.H. and Luo,R. (2007) *J. Phys. Chem. B*, **111**, 12263–12274.

Thoma,R., Obmolova,G., Lang,D.A., Schwander,M., Jeno,P., Sterner,R. and Wilmanns,M. (1999) *FEBS Lett.*, **454**, 1–6.

Vick,J.E. and Gerlt,J.A. (2007) *Biochemistry*, **46**, 14589–94597.

Wang,J.M., Wolf,R.M., Caldwell,J.W., Kollman,P.A. and Case,D.A. (2004) *J. Comput. Chem.*, **25**, 1157–1174.

Wierenga,R.K. (2001) *FEBS Lett.*, **492**, 193–198.

Wright,H., Noda-Garcia,L., Ochoa-Leyva,A., Hodgson,D.A., Fulop,V. and Barona-Gomez,F. (2008) *Biochem. Biophys. Res. Commun.*, **365**, 16–21.

Supplementary Material

A sugar isomerization reaction established on various (βα)₈-barrel scaffolds is based on substrate-assisted catalysis

Bernd Reisinger[1], Marco Bocola[1,2], Felix List[1,3], Jörg Claren[1], Chitra Rajendran[1], Reinhard Sterner[1,4]

[1]Institute of Biophysics and Physical Biochemistry, University of Regensburg, Germany

[2]Department of Biotechnology (Biology VI), RWTH Aachen University, Germany

[3]European Molecular Biology Laboratory (EMBL), Hamburg Unit, Hamburg, Germany.

[4]To whom correspondence should be addressed: Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany. Phone: +49 941 943 3015. Fax: +49 941 943 2813.

E-mail: reinhard.sterner@biologie.uni-regensburg.de

1

Supplementary Figures
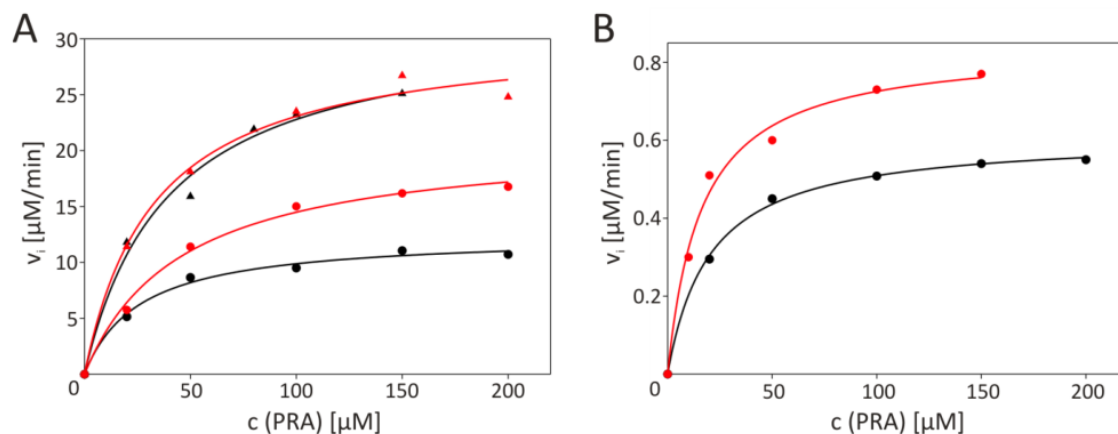
Fig. S1. Steady-state enzyme kinetics of HisA and HisF variants. In each case, PRA saturation curves of two different protein preparations were monitored (black and red). (A) HisA_D169V (circles; 10 µM protein) and HisA_D127V+D169V (triangles, 1 µM protein). (B) HisF_D130V+D176V (circles, 10 µM protein). The measurements were performed in 50 mM Hepes, pH 7.5, 4 mM MgCl$_2$, 4 mM EDTA, 2 mM DTT at 25 °C. The solid lines result from a hyperbolic fit of the data points, which yielded values for k_{cat} and K_M^{PRA}. The averaged values for the two saturation curves of each variant are shown in **Table 1**.
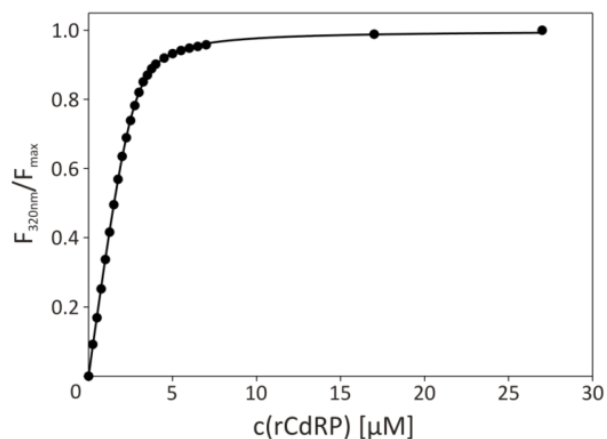
Fig. S2. Titration of HisF_D130V+D176V with rCdRP. The product analogue rCdRP was added to 5 µM HisF_D130V+D176V in 50 mM Tris/HCl, pH 7.5, 25 °C and fluorescence emission at 320 nm was determined. A quadratic fit of the obtained data points resulted in a K_D^{rCdRP} value of 0.18 ± 0.01 µM. The complex stoichiometry is approximately 1 molecule of rCdRP per enzyme molecule.

2
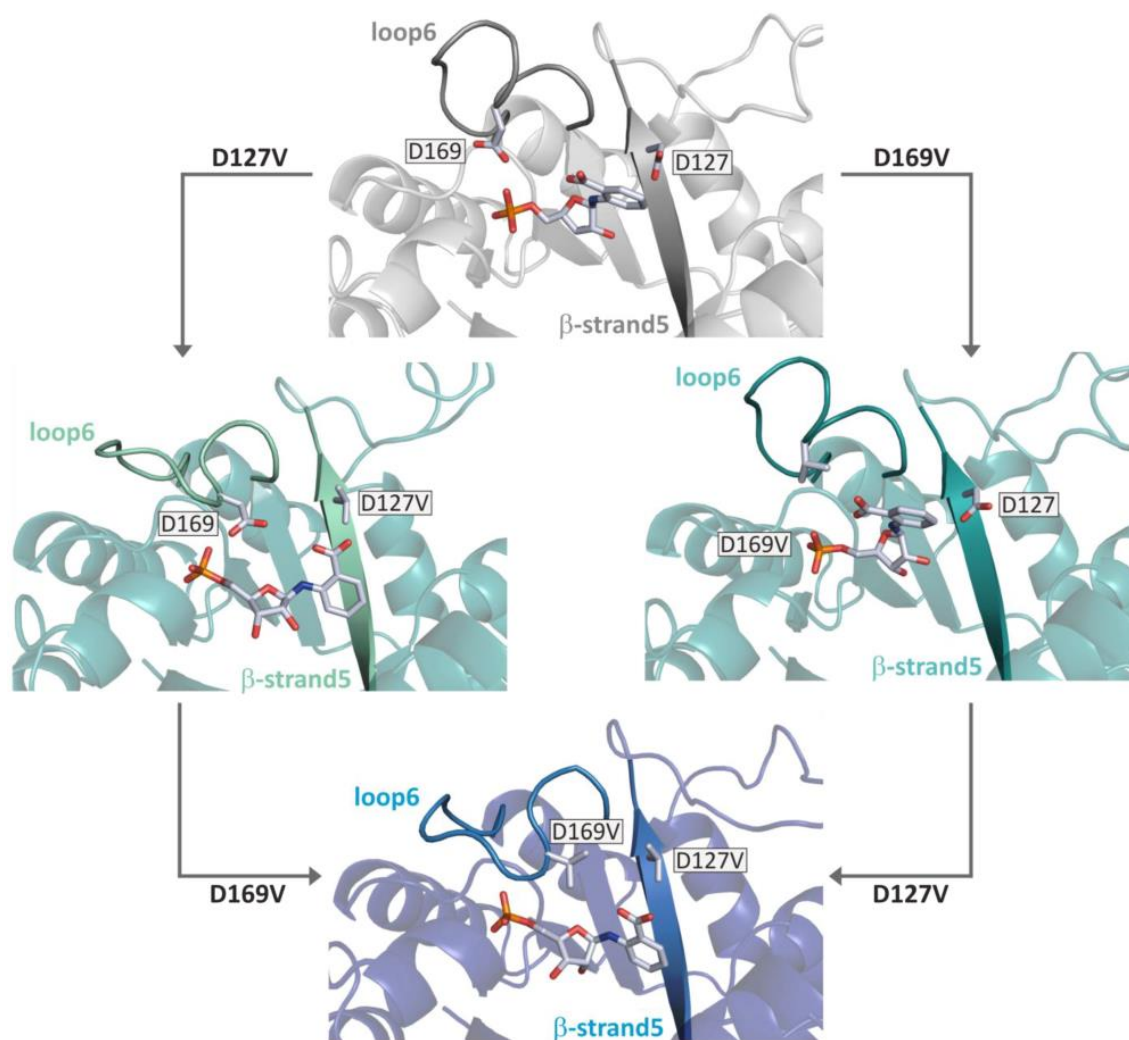
Fig. S3. Representative MD simulation structures of wild-type HisA, HisA_D127V, HisA_D169V, and HisA_D127V+D169V with PRA bound in mode 1. In wild-type HisA PRA is weakly bound with its carboxylic acid group pointing towards Asp169 and thus being surrounded by a rather hydrophilic environment (grey, upper panel). The D169V exchange does not affect the binding of PRA, but the proximity of the carboxylic acid group is now hydrophobic (dark green, right panel). On the other hand, as a consequence of the D127V exchange βα-loop 6 undergoes a distinct conformational rearrangement, which forces PRA deeper into the hydrophobic binding pocket and thus presumably increases the affinity between substrate and protein (light green, left panel). The combination of D127V and D169V finally results in both, a strongly bound PRA and a carboxylic acid group being surrounded by two hydrophobic valine residues (blue, lower panel). PDB ID codes of used structures: wild-type HisA and HisA_D169V (modeled onto the wild-type scaffold): 1qo2; HisA_D127V: 2cff; HisA_D127V+D169V: 2w79.

3
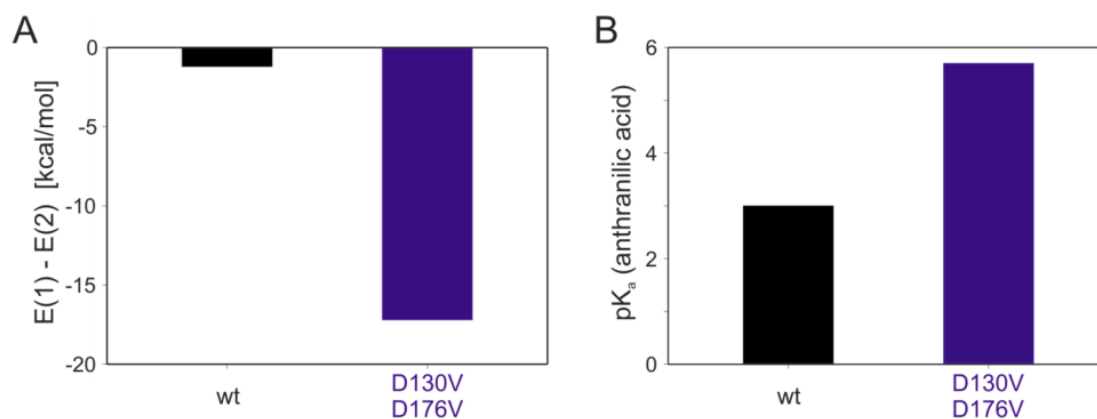
Fig. S4. Molecular Dynamics simulations of wild-type (wt) HisF and HisF_D130V+D176V. (A) Resulting energy differences E(1)-E(2) of simulations with binding modes 1 and 2 (see Fig. 6A). For each mode, at least three independent energy values were determined. A low energy value indicates a strong preference for binding mode 1 versus 2. A t-test showed that the energy values attributed to the two modes are different with statistical significance in case of HisF_D130V+D176V (p = 0.045). (B) pK$_a$-values of the anthranilic acid in binding mode 1 as determined by PROPKA (Li *et al.*, 2005). PDB ID codes of used structures: wild-type HisF wild-type: 1thf; HisF_D130V+D176V: 4ewn.
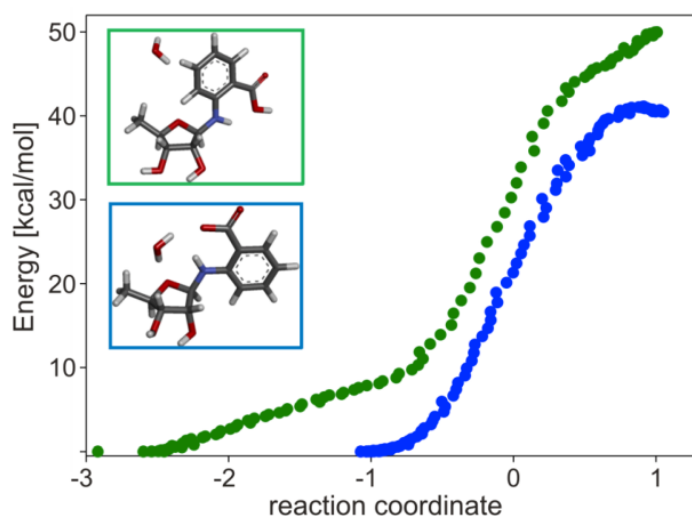
Fig. S5. QM/MM calculations of HisA_D127V+D169V with PRA bound in non-productive modes. With PRA in binding mode 2 (green inset; see Fig. 6A) or deprotonated (blue inset), protonation of the furanose ring oxygen via a water molecule is highly energetically unfavorable (around 40 and 50 kcal/mol, respectively). PDB ID code for HisA_D127V+D169V: 2w79.

4

Supplementary Table

Table S1. Crystal structure determination of HisF_D130V+D176V with bound rCdRP: data collection and refinement statistics.

Data collection	
Wavelength, (Å)	0.976
Space group	F222
Unit cell dimensions	
a, b, c, (Å)	57.65, 112.52, 163.50
α, β, γ, (°)	90, 90, 90
Resolution, (Å)	46.3-1.9 (2.0-1.9)
Total reflections	197996 (22031)
Unique reflections	20802 (2755)
Mosaicity, (°)	0.157
Completeness, (%)	98.4 (90.9)
R_{merge},(%)	5.0 (74.2)
I/σ(I)	21.8 (2.6)
Refinement statistics	
Resolution, (Å)	46.3-1.9
R_{cryst} / R_{free} (%)	20.4 / 23.7
Average B-value, ($Å^2$)	51.9
Number of atoms	
Protein	1871
Ligand	23
Water	66
r.m.s.d. of bond length, (Å)	0.007
r.m.s.d. of angle, (°)	1.079
Model quality (Molprobity)	
Residues in favored region, (%)	97.9
Residues in additional allowed region, (%)	1.7
Residues in outlier region, (%)	0.4

Supplementary Reference

Li, H., Robertson, A.D. and Jensen, J.H. (2005) *Proteins: Struct. Funct. Bioinform.,* **61**, 704-721.

6

5.2   Publication B

74

Evidence for the existence of elaborate enzyme complexes in the Paleoarchean era.

Bernd Reisinger, Josef Sperl, Alexandra Holinski, Veronika Schmid, Chitra Rajendran, Linn Carstensen, Sandra Schlee, Samuel Blanquart, Rainer Merkl and Reinhard Sterner

J. Am. Chem. Soc. 2014, **136**(1), 122-129

J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Evidence for the Existence of Elaborate Enzyme Complexes in the Paleoarchean Era

Bernd Reisinger,[†] Josef Sperl,[†] Alexandra Holinski,[†] Veronika Schmid,[†] Chitra Rajendran,[†] Linn Carstensen,[†] Sandra Schlee,[†] Samuel Blanquart,[‡] Rainer Merkl,*[,†] and Reinhard Sterner*[,†]

[†]Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, Germany

[‡]Equipe Bonsai, Institut National de Recherche en Informatique et en Automatique, INRIA Lille Nord Europe, 40 avenue Halley, 59650 Villeneuve d'Ascq, France

🅢 *Supporting Information*

ABSTRACT: Due to the lack of macromolecular fossils, the enzymatic repertoire of extinct species has remained largely unknown to date. In an attempt to solve this problem, we have characterized a cyclase subunit (HisF) of the imidazole glycerol phosphate synthase (ImGP-S), which was reconstructed from the era of the last universal common ancestor of cellular organisms (LUCA). As observed for contemporary HisF proteins, the crystal structure of LUCA-HisF adopts the $(\beta\alpha)_8$-barrel architecture, one of the most ancient folds. Moreover, LUCA-HisF (i) resembles extant HisF proteins with regard to internal 2-fold symmetry, active site residues, and a stabilizing salt bridge cluster, (ii) is thermostable and shows a folding mechanism similar to that of contemporary $(\beta\alpha)_8$-barrel enzymes, (iii) displays high catalytic activity, and (iv) forms a stable and functional complex with the glutaminase subunit (HisH) of an extant ImGP-S. Furthermore, we show that LUCA-HisF binds to a reconstructed LUCA-HisH protein with high affinity. Our findings suggest that the evolution of highly efficient enzymes and enzyme complexes has already been completed in the LUCA era, which means that sophisticated catalytic concepts such as substrate channeling and allosteric communication existed already 3.5 billion years ago.
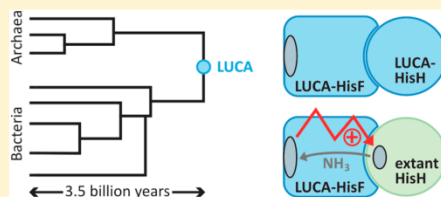
■ INTRODUCTION

Modern enzyme complexes are elaborate molecular machineries that have been optimized in the course of evolution for the efficient and specific processing of their substrates. One prominent example is the imidazole glycerol phosphate synthase (ImGP-S), a bienzyme complex which belongs to the family of glutamine amidotransferases[1] and constitutes a branch point connecting amino acid and nucleotide biosynthesis. ImGP-S consists of the cyclase subunit HisF and the glutaminase subunit HisH. HisF binds the substrate N'-[(5′phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) and performs a cycloligase/lyase reaction that generates imidazole glycerol phosphate (ImGP) and 5-aminoimidazole-4-carboxamide ribotide (AICAR), which are further used in histidine and de novo purine biosynthesis, respectively[2] (Figure 1). The ammonia molecule required for this transformation is produced by the glutaminase subunit HisH and transported to the active site of HisF through an extended molecular channel. This channeling hampers diffusion of ammonia into bulk solvent and thus presumably prevents its protonation to the nonproductive ammonium ion. Another specific feature of the HisF/HisH complex is the tight coordination of the two enzymatic activities: Binding of PRFAR (or its analogue N'-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide, ProFAR) to HisF results in an allosteric signal that leads to a several-hundred-fold stimulation of the

glutaminase activity of HisH.[3−5] This property precludes the hydrolysis of glutamine by HisH in the absence of an acceptor substrate at the active site of HisF.[6]

We were interested to find out whether the characteristics of modern HisF enzymes were already present in those species that colonized Earth in a very early phase of biological evolution. A straightforward answer to this question is difficult due to the lack of macromolecular fossils. However, computational techniques of amino acid sequence reconstruction[7,8] make it possible to travel back in time and to study extinct proteins.[9−16] In extreme cases, these algorithms enable us to study enzymes from the last universal common ancestor of cellular organisms (LUCA), which preceded the diversification of life and existed in the Paleoarchean era, i.e., at least 3.5 billion years ago.[17]

Along these lines, we have previously computationally reconstructed the amino acid sequence of HisF from the LUCA era (LUCA-HisF).[18] To this end, a set of 87 extant HisF and HisH proteins from the seven phylogenetic clades Crenarchaeota, Actinobacteria, Chlorobi, Cyanobacteria, Firmicutes, Proteobacteria, and Thermotogae has been used to determine a phylogenetic tree t_{HisF_HisH} based on the CAT model[19] (Supporting Information, Figure S1). After having rooted this tree between the superkingdoms Archaea and
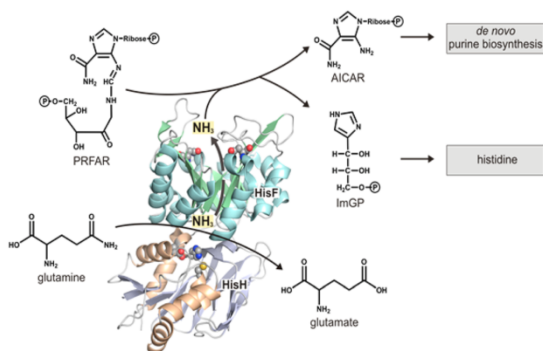
Figure 1. Reaction catalyzed by the heterodimeric ImGP synthase complex. The synthase subunit HisF catalyzes the reaction of N'-[(5′phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) with ammonia (NH_3) to imidazole glycerol phosphate (ImGP) and 5-aminoimidazole-4-carboxamide ribotide (AICAR). ImGP is further utilized in the synthesis of histidine, whereas AICAR is an intermediate in de novo purine biosynthesis, rendering HisF a branch-point enzyme of amino acid and nucleotide biosynthesis. The ammonia molecule required for the HisF reaction is produced by the glutaminase subunit HisH (catalytic triad residues are depicted as spheres) and subsequently channeled to the active site of HisF (catalytic aspartate residues are depicted as spheres). In the absence of HisH, HisF can also use external ammonia that is added as ammonia salt. HisF adopts the $(\beta\alpha)_8$-barrel fold, an ubiquitous and catalytically versatile protein architecture,[21] which is considered one of the three most ancient protein folds.[22] HisH adopts the α/β hydrolase fold.[23]

Bacteria, we reconstructed a predecessor of HisF from Bacteria and Crenarchaeota as described.[19,20] Thus, although the precise lineage of the three superkingdoms is still under debate,[24] LUCA-HisF is among the oldest so far reconstructed proteins, if not the oldest hitherto calculated predecessor.[16,25,26] Among the 87 descendants of LUCA-HisF used for reconstruction, 78 of the 250 residues are less than 50% conserved, whereas 49 residues are strongly conserved. Accordingly, LUCA-HisF differs in 55 amino acids (22%) from the closest BLAST match,[27] which is HisF from *Thermovibrio ammonificans*. The nucleotide and amino acid sequences of LUCA-HisF are given in the Supporting Information.

We have now produced LUCA-HisF in *Escherichia coli*, and analyzed its crystal structure, conformational stability, folding mechanism, and catalytic activity. The observed molecular characteristics of LUCA-HisF turned out to be similar to contemporary HisF proteins. Moreover, LUCA-HisF activates an extant HisH protein and thus comprises all elements required for allosteric interaction. Finally, we have also reconstructed and produced a LUCA-HisH protein and could show that it binds to LUCA-HisF with high affinity. Taken together, our results suggest that the protein inventory of the LUCA already contained elaborate enzyme complexes.

■ METHODS

Cloning, Expression, and Purification of LUCA-HisF. The gene coding for LUCA-HisF was optimized for its expression in *E. coli*, synthesized (GeneArt), and cloned into the vector pET24a(+) (Stratagene) using the terminal restriction sites for *Nde*I and *Xho*I. Since the addition of a C-terminal hexahistidine tag to LUCA-HisF might influence its interaction with HisH proteins, a stop codon was integrated at the end of the gene. The gene was expressed in *E. coli* T7-

Express cells (New England Biolabs) transformed with pET24a(+)-LUCA-*his*F. To this end, 4 L of Luria broth (LB) medium supplemented with 75 μg/mL kanamycin were inoculated with a preculture and incubated at 37 °C. After an OD_{600} of 0.6 was reached, the temperature was lowered to 30 °C. Expression was induced by adding 0.5 mM IPTG, and growth was continued overnight. Cells were harvested by centrifugation (Sorvall/RC5B, GS3, 15 min, 4000 rpm, 4 °C), washed with 50 mM potassium phosphate, pH 7.5, and centrifuged again. The cells were suspended in the same buffer, lysed by sonication (Branson Sonifier W-250D, 2 × 2 min in 15 s intervals, 45% pulse, 0 °C), and centrifuged again (Sorvall/RC5B, SS34, 30 min, 13.000 rpm, 4 °C) to separate the soluble from the insoluble fraction of the cell extract. In a first step, the soluble supernatant was subjected to ion exchange chromatography using a MonoQ column (HR 16/10, 20 mL, Pharmacia), which had been equilibrated with 50 mM potassium phosphate, pH 7.5. The column was washed with equilibration buffer, and bound LUCA-HisF was eluted by applying a linear gradient of 0−1.5 M NaCl. Protein-containing fractions were pooled, dialyzed against 50 mM potassium phosphate, pH 7.5, and subjected to ammonia sulfate precipitation. After 80% saturation with ammonia sulfate, precipitated protein was centrifuged (Sorvall/RC5B, SS34, 30 min, 13.000 rpm, 4 °C), dissolved in 50 mM potassium phosphate, pH 7.5, and 300 mM potassium chloride, and finally purified via size exclusion chromatography. For this purpose a Superdex200 column (HiLoad 26/60, 320 mL, GE Healthcare) was operated with 50 mM potassium phosphate, pH 7.5, and 300 mM potassium chloride at 4 °C. Fractions with pure protein were pooled and dialyzed against 50 mM Tris·HCl, pH 7.5. According to SDS-PAGE (12.5% acrylamide), LUCA-HisF was more than 95% pure. About 30 mg of protein was obtained per liter of culture.

In order to determine the binding properties of LUCA-HisF to HisH proteins via fluorescence titration, all tryptophan residues of LUCA-HisF were replaced by tyrosines. Hence, LUCA_*his*F_W138Y+W156Y was generated via overlap extension polymerase chain reaction (PCR)[28] using pET24a(+)-LUCA-*his*F as a template (see the Supporting Information for oligonucleotide sequences), and subsequently cloned into pET24a(+) via the terminal restriction sites for *Nde*I and *Xho*I. Expression and purification were performed as described for LUCA-HisF, yielding a comparable amount and purity of LUCA-HisF_W138Y+W156Y.

Cloning, Expression, and Purification of zmHisH. Genomic DNA of *Zymomonas mobilis* (DSM424) was ordered from the Leibniz Institute DSMZ. In order to remove the internal restriction site for *Nde*I, the zm*his*H gene was amplified by overlap extension PCR[28] (see the Supporting Information for oligonucleotide sequences) and cloned into pET24a(+) using the terminal restriction sites for *Nde*I and *Xho*I. After transformation of *E. coli* strain BL21(DE3) (Stratagene), expression was carried out at 30 °C overnight in 4 L of LB medium, supplemented with 75 μg/mL kanamycin. Protein purification was performed as described for LUCA-HisF including ion exchange chromatography using 50 mM Tris·HCl, pH 9, as buffer, ammonia sulfate precipitation, size exclusion chromatography, and final dialysis against 50 mM Tris·HCl, pH 7.5. According to SDS-PAGE (12.5% acrylamide), zmHisH was more than 95% pure. About 8 mg of protein was obtained per liter of culture.

Sequence Reconstruction, Cloning, Expression and Purification of LUCA-HisH. As for LUCA-HisF, the reconstruction of LUCA-HisH was based on the tree t_{HisF_HisH} (Supporting Information, Figure S1), which is close to an accepted organism phylogeny. In comparison to the multiple sequence alignment (MSA) of extant HisF sequences, the 87 extant HisH sequences exhibit a significantly higher variability. In fact, 140 of 226 residues are less than 50% conserved. Furthermore, the MSA ($HisH_{ext}$) contains several gaps. Recently, it has been shown that a novel algorithm for the phylogeny-aware gap placement named PRANK[29] improves MSA quality. This is why we used PRANK with the option −showanc to deduce LUCA-HisH from the MSA $HisH_{ext}$ under the control of t_{HisF_HisH} (Supporting Information, Figure S2). The nucleotide and amino acid sequences of LUCA-HisH are given in the Supporting Information. The protein

shares 123 of 226 residues (54%) with the closest BLAST[27] match, which is HisH from *Syntrophothermus lipocalidus*.

The gene coding for LUCA-HisH was optimized for its expression in *E. coli*, synthesized (GeneArt), and cloned into the vector pET24a(+) (Stratagene) using the terminal restriction sites for *Nde*I and *Xho*I. (The gene encodes a C-terminal hexahistidine tag; see the Supporting Information.) Subsequently, pET24a(+)-LUCA-*his*H was used to transform *E. coli* strain BL21-Gold (DE3) (Stratagene). Protein expression, harvesting of cells, and cell lysis were performed as described for LUCA-HisF. As LUCA-HisH showed a high thermal stability, most of the host proteins could be removed by heat denaturation (70 °C, 15 min) followed by centrifugation (Sorvall/RC5B, SS34, 30 min, 13.000 rpm, 4 °C). For further purification, the supernatant of the heat step was loaded onto a HisTrapFF crude column (5 mL; GE Healthcare), which had been equilibrated with 50 mM potassium phosphate, pH 7.5, 300 mM potassium chloride, and 10 mM imidazole. After washing with equilibration buffer, the bound protein was eluted by applying a linear gradient of 10−375 mM imidazole. Fractions with pure protein were pooled, and LUCA-HisH was dialyzed against 10 mM potassium phosphate, pH 7.5. As judged by SDS-PAGE, the protein was more than 95% pure. About 26 mg of LUCA-HisH was obtained per liter of culture.

Crystallization, Data Collection, and Refinement of LUCA-HisF. Crystallization trials were carried out using the PEG/Ion screen (Hampton Research). The hanging drop vapor diffusion method was performed in 96-well plates (Greiner) at 291 K. Drops contained 300 nL of the respective reservoir buffer mixed with 300 nL of LUCA-HisF (13.9 mg/mL) in 10 mM potassium phosphate, pH 7.5. In each well equilibration was performed against 100 μL of reservoir buffer. Crystals were obtained with 0.2 M sodium phosphate monobasic monohydrate, pH 4.7, and 20 wt %/vol PEG 3350. After flash freezing in liquid nitrogen, data of single crystals were collected at the synchrotron beamline PX2 (SLS) at 100 K. Data were processed using XDS,[30] and the data quality assessment was done using phenix.xt-riage.[31] Molecular replacement was performed with MOLREP within the CCP4i suite.[32] A homology model of LUCA-HisF with HisF from *Thermotoga maritima* (*tm*HisF) (PDB ID 1THF) was built with MODELLER[33] and served as a search model. Initial refinement was performed using REFMAC.[34] The model was further improved in several refinement rounds using automated restrained refinement with the program PHENIX[31] and interactive modeling with Coot.[35] The data collection and refinement statistics are summarized in Table S1 in the Supporting Information. The final model was analyzed using the program MolProbity.[36]

Analysis of the Thermal Stability of LUCA-HisF and LUCA-HisH. Differential scanning calorimetry (DSC) was performed with LUCA-HisF in 50 mM potassium phosphate, pH 7.5, by heating the sample in a CSC 5100 Nano differential scanning calorimeter with a scan rate of 1 °C min⁻¹. The DSC data were analyzed with the program CpCalc (version 2.1; Calorimetry Sciences Corp., 1995) to determine apparent melting temperatures (T_M^{app} DSC). Thermal denaturation traces of LUCA-HisF and LUCA-HisH in 50 mM potassium phosphate, pH 7.5, were monitored with a JASCO J-815 circular dichroism (CD) spectrometer in a 0.1 cm cuvette by following the loss of ellipticity at 220 nm. Unfolding was induced by raising the temperature in 1 °C increments at a ramp rate of 1 °C min⁻¹ with a Peltier-effect temperature controller. The midpoint temperatures of the unfolding transitions (T_M^{app} CD) were determined. Data are shown in Figure S3 in the Supporting Information. The irreversibility of the denaturation traces precluded the thermodynamic analyses of the DSC and CD unfolding measurements.

Equilibrium Unfolding/Refolding Transitions and Formation of a Burst-Phase Intermediate by LUCA-HisF. The thermodynamic stability of LUCA-HisF was determined by guanidinium chloride (GdmCl) induced equilibrium unfolding transitions. The loss of tertiary structure was probed by protein fluorescence; the loss of secondary structure was probed by far-UV CD. Samples with 2 μM protein were prepared in 50 mM Tris·HCl buffer (pH 7.5) containing different concentrations of GdmCl. GdmCl (ultrapure) was purchased from MP Biomedicals (Illkirch, France), and its concentration was

determined by the refractive index of the solution.[37] To reach equilibrium, LUCA-HisF was preincubated at the indicated concentration of GdmCl for 24 h at 25 °C.

The fluorescence emission signal at 320 nm (bandwidth 5 nm) after excitation at 280 nm (bandwidth 3 nm) was monitored with a JASCO Model FP-6500 spectrofluorimeter. The equilibrium unfolding transition of LUCA-HisF obtained by monitoring fluorescence is shown in Figure S4A in the Supporting Information in comparison to *tm*HisF and its artificially designed precursors Sym1 and Sym2, which were constructed by duplication and fusion of the C-terminal half-barrel HisF-C followed by the optimization of the initial construct.[38−40] The transitions were analyzed according to the two-state equilibrium model, assuming a linear dependency of the free-energy of unfolding on the GdmCl concentration.[41] The obtained values for ΔG_D, m, and $[D]_{1/2}$ are listed in Table S2 in the Supporting Information.

The far-UV circular dichroism (CD) signal at 225 nm was monitored using a JASCO Model J815 CD spectrophotometer (path length 5 mm; bandwidth 1 nm). The equilibrium unfolding/refolding transitions of LUCA-HisF obtained by monitoring the far-UV CD signal are shown in Figure S4B in the Supporting Information.

Kinetics of refolding of LUCA-HisF in Figure S4B in the Supporting Information were obtained by following the far-UV CD signal for 200 s in manual mixing experiments at various concentrations of GdmCl and extrapolating the exponential curve to zero time. The observed amplitude was plotted as a function of the GdmCl concentration and is shown in Figure S4B in the Supporting Information. It was significantly lower than the amplitude observed in the refolding equilibrium transitions, indicating that the major part of the CD change occurred within the dead time of the manual mixing experiment. This is interpreted with the formation of a compact burst-phase refolding intermediate with a high content of secondary structure.

Fluorescence Titration of *zm*HisH and LUCA-HisH with LUCA-HisF. Fluorescence titration was used to determine the binding stoichiometry and affinity of the LUCA-HisF/*zm*HisH and LUCA-HisF/LUCA-HisH complexes, as in the course of complex formation a tryptophan residue lying at the HisH interface is shielded from the solvent.[5] Hence, when titrating either 7 μM *zm*HisH or 5 μM LUCA-HisH in 50 mM potassium phosphate, pH 7.5, with LUCA-HisF_W138Y+W156Y, the emission maxima shifted from 345 to 325 nm and from 345 to 329 nm, respectively (excitation at 295 nm). The decreases in fluorescence emission at 318 nm were plotted against the added amounts of LUCA-HisF_W138Y+W156Y, and the resulting curves were analyzed with a quadratic fit. Both titrations were performed in triplicate.

Analysis of Enzymatic Activity in Vitro and in Vivo. In vitro enzymatic activities were determined by steady-state kinetics. The ammonia- and glutamine-dependent conversions of PRFAR into ImGP and AICAR (HisF reaction) were measured spectrophoto-metrically at 300 nm as previously described.[5] At 25 °C, entire progress curves at four different PRFAR concentrations were recorded either in 50 mM Tris-acetate, pH 8.5, in the presence of 100 mM ammonium acetate (*ammonia-dependent cyclase reaction*) or in 50 mM Tris-acetate, pH 8.0, in the presence of 15 mM glutamine and 2 μM *zm*HisH (*glutamine-dependent cyclase reaction*). In both cases an excess of HisA from *T. maritima* was added in order to synthesize PRFAR in situ from ProFAR,[44] and 0.5 μM LUCA-HisF was used to initiate the measurements. Data were analyzed using the integrated form of the Michaelis−Menten equation using the program COSY[45] to obtain k_{cat} and K_M^{PRFAR}. In case of the LUCA-HisF/LUCA-HisH complex (10 μM), no *glutamine-dependent cyclase activity* could be determined in the presence of 10 mM glutamine and 100 μM ProFAR. The glutaminase activity of *zm*HisH (1 μM) in complex with liganded LUCA-HisF (2 μM; 40 μM ProFAR) was measured in a coupled enzymatic assay as previously described.[5] At 25 °C, produced glutamate was oxidized by a molar excess of glutamate dehydrogenase (Roche) in 50 mM Tricine hydroxide, pH 8.0. Thus, the reduction of the coenzyme NAD⁺ to NADH could be monitored spectrophotometrically at 340 nm. Three glutamine saturation curves were recorded and fitted with the
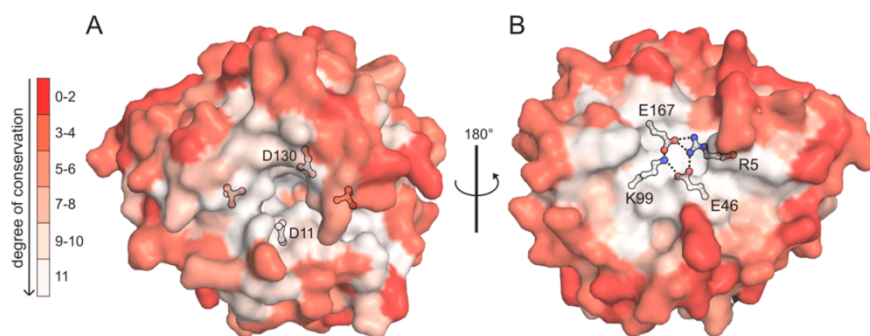
Figure 2. Crystal structure of LUCA-HisF. The surface is color coded according to residue conservation deduced from the MSA used for reconstruction. Conservation values [0−11] were determined by means of Jalview;[42] strictly conserved residues are white. (A) Catalytic face of HisF and view along the ammonia channel. The catalytically important aspartate residues D11 and D130 as well as two bound phosphate ions, which mimic the phosphate moieties of the substrate PRFAR, are shown as sticks. (B) Stability face and ammonia tunnel gate at the bottom of the β-barrel. The salt bridge cluster between the residues R5, E46, K99, and E167 (depicted as sticks; electrostatic interactions indicated by dashed lines) defines the entrance to the ammonia channel.[43]

Michaelis−Menten equation to obtain k_{cat} and K_M^{Gln}. In an identical setup, no glutaminase activity (12 mM glutamine) could be detected for LUCA-HisH (20 μM) in complex with ligand-bound LUCA-HisF (20 μM; 200 μM ProFAR). The extent to which LUCA-HisF liganded with ProFAR activates zmHisH had to be determined in a discontinuous assay, since NAD⁺ also exhibits a stimulating effect on glutaminase activity.[6] To this end, 10 mM glutamine was incubated at 25 °C with 0.5 μM zmHisH and 5 μM LUCA-HisF either in the absence of ProFAR or in the presence of 40 μM ProFAR. Aliquots of 150 μL of the reaction mixture were collected after 15, 30, 45, and 60 (only in the absence of ProFAR) min and spun through a 10 kDa filter (Roth) to remove the enzymes. The V_{max} value was calculated from the linear increase of glutamate production with time, which was determined with the help of 1 mg/mL glutamate dehydrogenase and 0.7 mM APAD⁺ (Sigma) (the reaction mixture was diluted 1:7.5, and absorption was measured at 363 nm). All measurements were performed in triplicates.

ProFAR to PRFAR isomerization activity (HisA reaction) was measured with the enzymatic assay described for the ammonia-dependent HisF reaction, however, in the presence of an excess of HisF.[46] PRA to CdRP isomerization activity (TrpF reaction) was followed at 25 °C by a fluorimetric assay (excitation at 350 nm, emission at 400 nm).[47,48] The substrate PRA was generated in situ by 1 μM yeast anthranilate phosphoribosyl transferase from anthranilate and PRPP, which was provided in a 30-fold molar excess. Moreover, 2.5 μM indole-3-glycerol phosphate synthase from *T. maritima* was added to prevent product inhibition.

To test for enzymatic activity in vivo, the gene coding for LUCA-HisF was subcloned into the pTNA vector, which allows for constitutive expression in *E. coli*.[49] The resulting pTNA-LUCA-*his*F plasmid was used to transform cells of auxotrophic Δ*his*F, Δ*his*A, or Δ*trp*F *E. coli* strains.[50,51] These strains lack the *his*F, *his*A, or *trp*F gene on their chromosome and are, therefore, unable to grow on medium without histidine or tryptophan, respectively. Growth experiments and controls were performed as described.[50]

RESULTS AND DISCUSSION

Structure Determination of LUCA-HisF. The gene coding for LUCA-HisF was synthesized, cloned into a plasmid, and expressed in *E. coli*. The LUCA-HisF protein was predominantly found in the soluble fraction of the host cell extract, and purified in a three-step process using ion exchange chromatography, ammonia sulfate precipitation, and size exclusion chromatography. Purified LUCA-HisF was crystallized and its three-dimensional structure was determined at 1.48

Å resolution by molecular replacement based on the structure of *tm*HisF[52] (Figure 2, Table S1 in the Supporting Information). LUCA-HisF (PDB ID 4EVZ) adopts the conserved $(\beta\alpha)_8$-barrel structure observed in the three extant HisF proteins from *Pyrobaculum aerophilum* (PDB ID 1H5Y), *Thermus thermophilus* (PDB ID 1KA9), and *T. maritima* (PDB ID 1THF), for which crystal structures have been previously determined. The superposition of LUCA-HisF with each of these structures by means of STAMP[53] resulted in an overall root-mean-square deviation (rmsd) ranging from 1.14 to 1.43 Å. In agreement with the postulated evolution of the $(\beta\alpha)_8$-barrel fold from a $(\beta\alpha)_4$-half-barrel,[54,55] LUCA-HisF displays a clear 2-fold symmetry: The superposition of its N-terminal $[(\beta\alpha)_{1-4}]$ and C-terminal $[(\beta\alpha)_{5-8}]$ halves yielded an rmsd of 1.68 Å, which is similar to the corresponding values for the three extant HisF proteins (1.27 Å for 1H5Y, 1.52 Å for 1KA9, and 1.69 Å for 1THF). Consistent with the internal symmetry, the two catalytically important aspartate residues[5] are found on opposite sides of the active site at the C-terminal ends of β-strand 1 and β-strand 5. Likewise, the two cocrystallized phosphate groups, which represent the two phosphate groups of the substrate PRFAR (Figure 1), are anchored by the C-terminal ends of β-strands 3 and 4, and β-strands 7 and 8, respectively (Figure 2A). Moreover, a stabilizing salt-bridge cluster at the N-terminal end of the β-barrel, which contains four charged and invariant residues that form the gate to the cyclase ammonia channel,[6,38,43] is also present in LUCA-HisF (Figure 2B).

Stability and Folding Mechanism of LUCA-HisF. The thermal stability of LUCA-HisF was determined by differential scanning calorimetry (DSC), which monitors overall unfolding, and the heat-induced decrease of the far-UV CD signal, which indicates the loss of secondary structure. The combination of both methods showed that thermal unfolding of LUCA-HisF is a two-step process with apparent transition midpoints of about 70 and 100 °C (Supporting Information, Figure S3A,B). These results characterize LUCA-HisF as an enzyme with a high resistance to heat. Interestingly, even higher denaturation temperatures were previously observed for enzymes from the common ancestors of Bacteria, Archaea, and Archaea/ Eukaryota.[16] These findings and our results are interesting in the light of rRNA and protein sequence analyses which have

provided independent support for the increase of thermotolerance from the LUCA to the ancestors of Bacteria and Archaea/Eukaryota.[56] Furthermore, the conformational stability of LUCA-HisF was analyzed by GdmCl-induced equilibrium unfolding and refolding transitions. The loss or gain of tertiary structure was probed by protein (Tyr/Trp) fluorescence. The equilibrium unfolding and refolding curves superpose well, which proves the reversibility of unfolding (Figure 3).
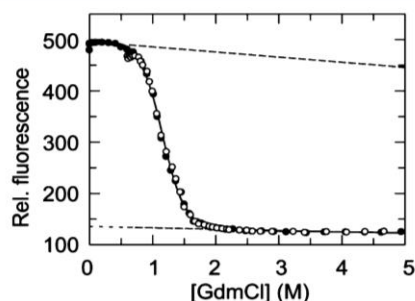
Figure 3. GdmCl-induced equilibrium unfolding/refolding transitions of LUCA-HisF. The transitions were followed by Trp/Tyr fluorescence (excitation at 280 nm; emission at 320 nm) in 50 mM Tris·HCl buffer, pH 7.5. Closed symbols represent the unfolding experiment, started with folded protein, and open symbols represent the refolding experiment, started with protein that was previously unfolded in 6.0 M GdmCl. The continuous line represents a fit to the unfolding transition on the basis of the two-state model. The dashed lines indicate the baselines for the pure N and U states. The thermodynamic parameters deduced from the analysis are given in the text and listed in Table S2 in the Supporting Information.

Moreover, the transitions are adequately described by the two-state model,[41] indicating that no significant amounts of stable equilibrium intermediates are populated. The analysis yielded an m-value of ~15 kJ mol^{-1} M^{-1}, a transition midpoint ($[D]_{1/2}$) at 1.2 M GdmCl, and a free energy of unfolding in the absence of denaturant (ΔG_D) of 18 kJ mol^{-1} (Supporting Information, Table S2). LUCA-HisF has a lower ΔG_D but a comparably high m-value as tmHisF and its artificially designed precursors Sym1 and Sym2,[39,40] indicating that it is comparably compact as these proteins but less stable (Supporting Information, Figure S4A, Table S2). Folding and unfolding

kinetics followed by Tyr/Trp fluorescence showed that the reduced stability of LUCA-HisF is due to strongly increased unfolding rates of LUCA-HisF in comparison to T. maritima HisF (tmHisF) (Supporting Information, Figure S5A). The comparison of the refolding kinetics of LUCA-HisF, tmHisF, Sym1, and Sym2 followed by fluorescence and far-UV CD (Supporting Information) showed that all four proteins share a common sequential folding mechanism including a nonproductive burst-phase intermediate (Supporting Information, Figure S4B) and two productive intermediates (Supporting Information, Figure S5). The rate-limiting step that synchronizes folding is conserved (Supporting Information, Figure S6).

Catalytic Activity of LUCA-HisF. The enzymatic activity of LUCA-HisF was measured in vitro using steady-state kinetics. The analysis of PRFAR to ImGP/AICAR progress curves obtained in the presence of saturating concentrations of externally added ammonia (ammonia-dependent cyclase activity) yielded a catalytic efficiency (k_{cat}/K_M^{PRFAR}) of 2.8 × 10^5 M^{-1} s^{-1}, which is similar to the catalytic efficiency of 3.3 × 10^5 M^{-1} s^{-1} that was obtained for tmHisF (Table 1). As ancient enzymes have been proposed to be less specific (more promiscuous) than their modern descendants,[57] we tested LUCA-HisF for its ability to catalyze related metabolic reactions. The homologous enzyme HisA, which precedes HisF in the histidine biosynthesis pathway, catalyzes the Amadori rearrangement of ProFAR to PRFAR. HisA shares with HisF the overall $(\beta\alpha)_8$-barrel fold as well as the location of the two symmetry-related catalytic aspartate residues and phosphate binding sites.[52] Phosphoribosyl anthranilate (PRA) isomerase (TrpF) catalyzes an Amadori rearrangement in tryptophan biosynthesis analogous to HisA in histidine biosynthesis.[49] Remarkably, a single amino acid exchange in the HisA and HisF proteins from T. maritima leads to TrpF activity, suggesting that these three phosphate-binding $(\beta\alpha)_8$-barrel proteins have evolved from a common precursor.[48,58] We examined LUCA-HisF for the isomerization activity toward ProFAR and PRA. However, no substrate turnover could be detected, even in the presence of 50 μM protein. These findings were complemented by assessing catalytic activity in vivo using metabolic selection. For this purpose, a plasmid harboring the LUCA-HisF gene was used to transform auxotrophic E. coli strains lacking either the intrinsic hisF, hisA, or trpF gene. When plated on minimal medium without histidine or tryptophan, the ΔhisF cells formed visible colonies within 24 h, whereas the ΔhisA and ΔtrpF cells did not

Table 1. Steady-State Kinetic Constants of the ImGP Synthase Pairs LUCA-HisF/zmHisH and tmHisF/tmHisH

ammonia-dependent cyclase activity[a]	k_{cat}, s^{-1}	K_M^{PRFAR}, μM	k_{cat}/K_M^{PRFAR}, M^{-1} s^{-1}
LUCA-HisF	0.078 (±0.003)	0.29 (±0.04)	2.8 (±0.3) × 10^5
tmHisF[b]	1.2	3.6	3.3 × 10^5
glutamine-dependent cyclase activity[c]	k_{cat}, s^{-1}	K_M^{PRFAR}, μM	k_{cat}/K_M^{PRFAR}, M^{-1} s^{-1}
LUCA-HisF/zmHisH	0.058 (±0.006)	0.36 (±0.07)	1.6 (±0.3) × 10^5
tmHisF/tmHisH[b]	1.1	2.0	5.5 × 10^5
glutaminase activity[d]	k_{cat}, s^{-1}	K_M^{Gln}, mM	k_{cat}/K_M^{Gln}, M^{-1} s^{-1}
LUCA-HisF/zmHisH	0.21 (±0.03)	1.9 (±0.9)	1.2 (±0.3) × 10^2
tmHisF/tmHisH[b]	0.1	0.8	1.3 × 10^2
stimulation of glutaminase activity[d]	k_{cat}, s^{-1} (without ProFAR)	k_{cat}, s^{-1} (ProFAR satd)	stimulation factor[e]
LUCA-HisF/zmHisH	3.85 (±0.04) × 10^{-2}	0.483 (±0.006)	13
tmHisF/tmHisH[b]	3.3 × 10^{-4}	0.1	303

[a]Reaction conditions: 50 mM Tris-acetate buffer, pH 8.5, at 25 °C. [b]Data taken from ref 6. [c]Reaction conditions: 50 mM Tris-acetate buffer, pH 8.0, at 25 °C. [d]Reaction conditions: 50 mM Tricine hydroxide, pH 8.0, at 25 °C. [e]The stimulation factor is the quotient k_{cat}(ProFAR saturated)/k_{cat}(without ProFAR).

grow within 1 week. Taken together, these results suggest that LUCA-HisF is a monofunctional enzyme.

Formation of LUCA-HisF/HisH Complexes. In order to test whether LUCA-HisF contains all structural elements required for complex formation, substrate channeling, and allosteric communication, we assayed its functional interaction with the extant *zm*HisH enzyme from *Zymomonas mobilis*. For this purpose, *zm*HisH was produced in *E. coli* and purified. The binding of *zm*HisH to LUCA-HisF was analyzed via fluorescence titration,[5] which showed that the two proteins form a stoichiometric complex with a thermodynamic dissociation constant (K_D) of 113 nM (Figure 4A).

The steady-state kinetic constants k_{cat} and K_M^{PRFAR} of LUCA-HisF in the presence of *zm*HisH and saturating concentrations of glutamine (*glutamine-dependent cyclase activity*) compare well with the above-reported *ammonia-dependent cyclase activity* (Table 1). This outcome confirms the functionality of the LUCA-HisF/*zm*HisH complex, as ammonia produced at the active site of HisH by means of glutamine hydrolysis is used as efficiently by LUCA-HisF as externally added ammonia. Moreover, this finding suggests that ammonia is transported from HisH to the active site of the synthase through a molecular channel formed by the central β-barrel of LUCA-HisF, as observed for extant HisF enzymes.[43,59] Furthermore, glutamine hydrolysis by *zm*HisH in the presence of LUCA-HisF and saturating concentrations of the substrate analogue ProFAR (*glutaminase activity*) is as efficient as glutaminase activity of HisH from *T. maritima* (*tm*HisH) in complex with ProFAR-liganded *tm*HisF[4] (Table 1). The comparison of the *zm*HisH activity in the presence and absence of ProFAR indicates a 13-fold stimulation by the HisF-ligand in this non-native complex (Figure 4B), which is 23-fold lower than the stimulating effect of ProFAR in the native *tm*HisF/*tm*HisH complex (Table 1).

Following the characterization of LUCA-HisF, we also reconstructed the amino acid sequence of the corresponding glutaminase LUCA-HisH. Again, we used the tree t_{HisF_HisH}, but opted for the phylogeny-aware gap placement of PRANK[29] to deduce LUCA-HisH from the MSA $HisH_{ext}$, which contains several insertions and deletions. The gene coding for LUCA-HisH was synthesized, cloned into a plasmid, and expressed in *E. coli*. The produced protein was soluble and could be purified by a combination of heat denaturation and Ni^{2+} affinity chromatography. As observed for LUCA-HisF, LUCA-HisH exhibits a high thermotolerance. Unfolding followed by CD resulted in a single transition with a midpoint of about 79 °C (Supporting Information, Figure S3C). Complex formation between LUCA-HisH and LUCA-HisF was probed by fluorescence titration.[5] Both proteins interacted stoichiometrically with very high affinity as demonstrated by a K_D value of 4 nM (Figure 4C). However, when testing the LUCA-HisF/LUCA-HisH complex for *glutamine-dependent cyclase activity* or LUCA-HisH for the hydrolysis of glutamine in the presence of LUCA-HisF and saturating concentrations of ProFAR, no enzymatic turnover could be determined. Thus, unlike LUCA-HisF, LUCA-HisH is catalytically inactive. As outlined in the following, uncertainties in the reconstruction process are probably responsible for this finding.

The evolutionary models underlying reconstruction consider each residue position independently of all other positions. Thus, the reliability of a given reconstruction is limited not by sequence length, but by the composition of the MSA and the topology of the deduced phylogenetic tree. In the case of
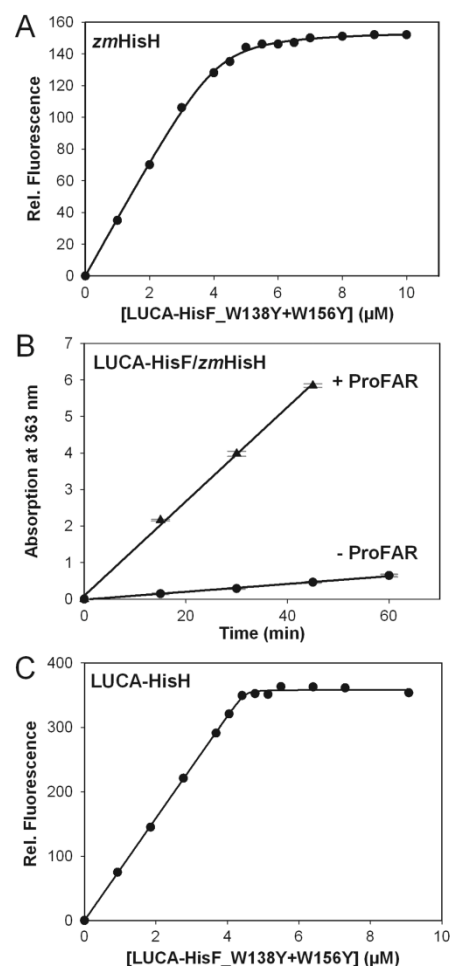
Figure 4. Fluorescence titration curve of *zm*HisH and LUCA-HisH with LUCA-HisF and activation of *zm*HisH by LUCA-HisF. (A) LUCA-HisF_W138Y+W156Y was added to 7 μM *zm*HisH in 50 mM potassium phosphate, pH 7.5, and 25 °C. Fluorescence emission at 318 nm was determined following excitation at 295 nm. Quadratic fits of the obtained data points resulted in a K_D value of 113 ($\pm$1) nM. The stoichiometry is slightly deviating from a 1:1 complex, indicating that a small fraction of *zm*HisH is not active. (B) Glutaminase activity of the LUCA-HisF/*zm*HisH complex in the absence (circles) and presence (triangles) of ProFAR was tested in a discontinuous assay (see the Supporting Information for detailed information). Mean values and standard deviations of triplicate measurements are shown. Glutamine (10 mM) was incubated with 0.5 μM *zm*HisH and 5 μM LUCA-HisF at 25 °C in both cases. Glutaminase activity is enhanced 13-fold in the presence of ProFAR (see Table 1). (C) Titration of 5 μM LUCA-HisH with LUCA-HisF_W138Y+W156Y was performed and analyzed analogous to (A), yielding a stoichiometric complex with a K_D value of 4 ($\pm$2) nM.

LUCA-HisF, 49 out of 250 residues are strictly conserved; among them are the two active site aspartate residues[5] and amino acids contributing to the central ammonia channel[38,43] (Figure 2). Furthermore, the four central nodes of t_{HisF_HisH} (Supporting Information, Figure S1) possess posterior probabilities of $\geq$0.88. Taken together, these features suggest

that tree topology and choice of the most likely residues for the corresponding predecessors and LUCA-HisF is largely unambiguous. In contrast, in the case of LUCA-HisH, only 21 out of 226 residues are strictly conserved, which makes the reconstruction much more prone to uncertainties.

Even more than a certain conservation of amino acid sequence composition, conservation of sequence length is an important prerequisite for a valid reconstruction. Along these lines, the sequence lengths of extant and reconstructed thioredoxins, which is the only other example for a fully functional enzyme from the LUCA era,[16] are very similar. In contrast, MSA $HisH_{ext}$ contains several gaps. Nonetheless, the phylogeny-aware gap placement by means of PRANK did allow us to reconstruct a stable LUCA-HisH protein with a fully functional protein–protein interface, albeit lacking enzyme activity. Obviously, t_{HisF_HisH} was sufficiently informative to reconstruct ancestral residues at positions whose role did not change during evolution such as the catalytic triad[3] and residues involved in binding of the substrate glutamine, as deduced from the structure of tmHisH.[6] In contrast, reconstruction seems to have failed at residue positions that underwent frequent changes during evolution due to insertions and deletions. It has to be shown that highly articulated phylogenetic trees will enable us to reconstruct the correct series of indels and to further improve reconstruction for such difficult cases.

CONCLUSIONS

Taken together, LUCA-HisF, which presumably existed about 3.5 billion years ago, is similar to extant HisF proteins with respect to structure, stability, folding, and activity. Since similar results were obtained for predecessors of thioredoxin,[16] experimental evidence accumulates for the existence of highly effective enzymes in the LUCA era. In addition, LUCA-HisF forms a stable complex with LUCA-HisH and a functional enzyme complex with the extant glutaminase zmHisH. It is therefore plausible to assume that the evolution of the ImGP-S complex, including ammonia channeling and allosteric communication, had been completed in the LUCA era. Thus, our experimental findings are in line with the hypothesis that the LUCA had already a rather diverse metabolism, which was as sophisticated as the metabolisms of its archaeal and bacterial successors are.[60]

ASSOCIATED CONTENT

Ⓢ Supporting Information

Nucleotide and amino acid sequences of LUCA-HisF and LUCA-HisH. Oligonucleotides used for the construction of LUCA_hisF_W138Y+W156Y and for the genomic amplification of zmHisH. Comparison of the folding mechanism of LUCA-HisF, *T. maritima* HisF, and its artificial precursors Sym1 and Sym2. Crystal structure determination of LUCA-HisF (PDB ID 4EVZ): data collection and refinement statistics (Table S1). Thermodynamic unfolding parameters of LUCA-HisF, *T. maritima* HisF, Sym1, and Sym2 (Table S2). Phylogenetic tree t_{HisF_HisH} used for the reconstruction of LUCA-HisF and LUCA-HisH (Figure S1). PRANK output for the reconstruction of LUCA-HisH (Figure S2). Thermal denaturation of LUCA-HisF and LUCA-HisH (Figure S3). Equilibrium unfolding transitions of LUCA-HisF, *T. maritima* HisF, Sym1, and Sym2, and formation of a burst-phase intermediate by LUCA-HisF (Figure S4). Apparent rate constants (λ) and amplitudes of refolding and unfolding kinetics of LUCA-HisF (Figure S5). Unifying folding mechanism for LUCA-HisF, *T. maritima* HisF, and Sym1 and Sym2 (Figure S6). References for Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
Rainer.Merkl@ur.de; Reinhard.Sterner@ur.de

Author Contributions
B.R. and J.S. contributed equally. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Zalkin, H. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1993**, *66*, 203.
(2) Chittur, S. V.; Chen, Y.; Davisson, V. J. *Protein Expression Purif.* **2000**, *18*, 366.
(3) Myers, R. S.; Jensen, J. R.; Deras, I. L.; Smith, J. L.; Davisson, V. J. *Biochemistry* **2003**, *42*, 7013.
(4) Klem, T. J.; Davisson, V. J. *Biochemistry* **1993**, *32*, 5177.
(5) Beismann-Driemeyer, S.; Sterner, R. *J. Biol. Chem.* **2001**, *276*, 20387.
(6) List, F.; Vega, M. C.; Razeto, A.; Haeger, M. C.; Sterner, R.; Wilmanns, M. *Chem. Biol.* **2012**, *19*, 1589.
(7) Hanson-Smith, V.; Kolaczkowski, B.; Thornton, J. W. *Mol. Biol. Evol.* **2010**, *27*, 1988.
(8) Harms, M. J.; Thornton, J. W. *Curr. Opin. Struct. Biol.* **2010**, *20*, 360.
(9) Benner, S. A.; Sassi, S. O.; Gaucher, E. A. *Adv. Enzymol. Relat. Areas Mol. Biol.* **2007**, *75*, 1.
(10) Thomson, J. M.; Gaucher, E. A.; Burgan, M. F.; De Kee, D. W.; Li, T.; Aris, J. P.; Benner, S. A. *Nat. Genet.* **2005**, *37*, 630.
(11) Chang, B. S.; Jonsson, K.; Kazmi, M. A.; Donoghue, M. J.; Sakmar, T. P. *Mol. Biol. Evol.* **2002**, *19*, 1483.
(12) Malcolm, B. A.; Wilson, K. P.; Matthews, B. W.; Kirsch, J. F.; Wilson, A. C. *Nature* **1990**, *345*, 86.
(13) Stackhouse, J.; Presnell, S. R.; McGeehan, G. M.; Nambiar, K. P.; Benner, S. A. *FEBS Lett.* **1990**, *262*, 104.
(14) Jermann, T. M.; Opitz, J. G.; Stackhouse, J.; Benner, S. A. *Nature* **1995**, *374*, 57.
(15) Finnigan, G. C.; Hanson-Smith, V.; Stevens, T. H.; Thornton, J. W. *Nature* **2012**, *481*, 360.
(16) Perez-Jimenez, R.; Inglés-Prieto, A.; Zhao, Z. M.; Sanchez-Romero, I.; Alegre-Cebollada, J.; Kosuri, P.; Garcia-Manyes, S.; Kappock, T. J.; Tanokura, M.; Holmgren, A.; Sanchez-Ruiz, J. M.; Gaucher, E. A.; Fernandez, J. M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 592.
(17) Nisbet, E. G.; Sleep, N. H. *Nature* **2001**, *409*, 1083.
(18) Richter, M.; Bosnali, M.; Carstensen, L.; Seitz, T.; Durchschlag, H.; Blanquart, S.; Merkl, R.; Sterner, R. *J. Mol. Biol.* **2010**, *398*, 763.
(19) Lartillot, N.; Philippe, H. *Mol. Biol. Evol.* **2004**, *21*, 1095.
(20) Blanquart, S.; Lartillot, N. *Mol. Biol. Evol.* **2008**, *25*, 842.
(21) Wierenga, R. K. *FEBS Lett.* **2001**, *492*, 193.
(22) Caetano-Anollés, G.; Kim, H. S.; Mittenthal, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9358.
(23) Ollis, D. L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J.; Sussman, J. L.; Verschueren, K. H. G.; Goldman, A. *Protein Eng.* **1992**, *5*, 197.
(24) Gribaldo, S.; Poole, A. M.; Daubin, V.; Forterre, P.; Brochier-Armanet, C. *Nat. Rev. Microbiol.* **2010**, *8*, 743.

(25) Gaucher, E. A.; Govindarajan, S.; Ganesh, O. K. *Nature* **2008**, *451*, 704.

(26) Risso, V. A.; Gavira, J. A.; Mejia-Carmona, D. F.; Gaucher, E. A.; Sanchez-Ruiz, J. M. *J. Am. Chem. Soc.* **2013**, *135*, 2899.

(27) Boratyn, G. M.; Camacho, C.; Cooper, P. S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T. L.; Matten, W. T.; McGinnis, S. D.; Merezhuk, Y.; Raytselis, Y.; Sayers, E. W.; Tao, T.; Ye, J.; Zaretskaya, I. *Nucleic Acids Res.* **2013**, *41*, W29.

(28) Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. *Gene* **1989**, *77*, 51.

(29) Löytynoja, A.; Goldman, N. *Science* **2008**, *320*, 1632.

(30) Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125.

(31) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1948.

(32) Potterton, L.; McNicholas, S.; Krissinel, E.; Gruber, J.; Cowtan, K.; Emsley, P.; Murshudov, G. N.; Cohen, S.; Perrakis, A.; Noble, M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2288.

(33) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779.

(34) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240.

(35) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126.

(36) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res.* **2007**, *35*, W375.

(37) Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.

(38) Höcker, B.; Lochner, A.; Seitz, T.; Claren, J.; Sterner, R. *Biochemistry* **2009**, *48*, 1145.

(39) Carstensen, L.; Sperl, J. M.; Bocola, M.; List, F.; Schmid, F. X.; Sterner, R. *J. Am. Chem. Soc.* **2012**, *134*, 12786.

(40) Carstensen, L.; Zoldak, G.; Schmid, F. X.; Sterner, R. *Biochemistry* **2012**, *51*, 3420.

(41) Santoro, M. M.; Bolen, D. W. *Biochemistry* **1988**, *27*, 8063.

(42) Waterhouse, A. M.; Procter, J. B.; Martin, D. M.; Clamp, M.; Barton, G. J. *Bioinformatics* **2009**, *25*, 1189.

(43) Douangamath, A.; Walker, M.; Beismann-Driemeyer, S.; Vega-Fernandez, M. C.; Sterner, R.; Wilmanns, M. *Structure* **2002**, *10*, 185.

(44) Thoma, R.; Obmolova, G.; Lang, D. A.; Schwander, M.; Jeno, P.; Sterner, R.; Wilmanns, M. *FEBS Lett.* **1999**, *454*, 1.

(45) Eberhard, M. *Comput. Appl. Biosci.* **1990**, *6*, 213.

(46) Jürgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmanns, M.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9925.

(47) Hommel, U.; Eberhard, M.; Kirschner, K. *Biochemistry* **1995**, *34*, 5429.

(48) Leopoldseder, S.; Claren, J.; Jürgens, C.; Sterner, R. *J. Mol. Biol.* **2004**, *337*, 871.

(49) Henn-Sax, M.; Thoma, R.; Schmidt, S.; Hennig, M.; Kirschner, K.; Sterner, R. *Biochemistry* **2002**, *41*, 12032.

(50) Claren, J.; Malisi, C.; Höcker, B.; Sterner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3704.

(51) Sterner, R.; Dahm, A.; Darimont, B.; Ivens, A.; Liebl, W.; Kirschner, K. *EMBO J.* **1995**, *14*, 4395.

(52) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. *Science* **2000**, *289*, 1546.

(53) Russell, R. B.; Barton, G. J. *Proteins* **1992**, *14*, 309.

(54) Höcker, B.; Beismann-Driemeyer, S.; Hettwer, S.; Lustig, A.; Sterner, R. *Nat. Struct. Biol.* **2001**, *8*, 32.

(55) Sterner, R.; Höcker, B. *Chem. Rev.* **2005**, *105*, 4038.

(56) Boussau, B.; Blanquart, S.; Necsulea, A.; Lartillot, N.; Gouy, M. *Nature* **2008**, *456*, 942.

(57) Jensen, R. A. *Annu. Rev. Microbiol.* **1976**, *30*, 409.

(58) List, F.; Sterner, R.; Wilmanns, M. *ChemBioChem* **2011**, *12*, 1487.

(59) Chaudhuri, B. N.; Lange, S. C.; Myers, R. S.; Chittur, S. V.; Davisson, V. J.; Smith, J. L. *Structure* **2001**, *9*, 987.

(60) Glansdorff, N.; Xu, Y.; Labedan, B. *Biol. Direct* **2008**, *3*, 29.

Supporting Information for

Evidence for the existence of elaborate enzyme complexes in the Paleoarchean era

Bernd Reisinger[1], Josef Sperl[1], Alexandra Holinski[1], Veronika Schmid[1], Chitra Rajendran[1], Linn Carstensen[1], Sandra Schlee[1], Samuel Blanquart[2], Rainer Merkl*,[1], Reinhard Sterner*,[1]

[1]Institute of Biophysics and Physical Biochemistry, University of Regensburg, Universitätsstraße 31, D-93053 Regensburg, Germany

[2]Equipe Bonsai, Institut National de Recherche en Informatique et en Automatique, INRIA Lille Nord Europe, 40 avenue Halley, 59650 Villeneuve d'Ascq, France

- S1 -

Table of Contents

Nucleotide and amino acid sequences of LUCA-HisF and LUCA-HisH

Nucleotide sequence of LUCA-HisF (including restriction sites for *Nde*I and *Xho*I in boldface):

CATATGCTGGCAAAACGTATTATTCCGTGCCTGGATGTTAAAGATGGTCGTGTTGTTAAA
GGCGTGAATTTTGAAAATCTGCGTGATGCCGGTGATCCGGTTGAACTGGCAGCACGTTAT
GATGAAGAAGGCGCTGACGAACTGGTGTTTCTGGATATTACCGCAAGCCATGAAGGTCGC
GAAACCATGCTGGAAGTTGTTGAACGTACCGCAGAACAGGTTTTTATTCCGCTGACCGTT
GGTGGTGGTATTCGTAGCGTTGAAGATGCAAGCCGTCTGCTGCGTGCCGGTGCAGATAAA
GTTAGCATTAATACCGCAGCCGTGAAAAATCCGGAACTGATTACCGAAGCAGCAGAAGA
ATTTGGTAGCCAGGCAGTTGTTGTTGCAATTGATGCAAAACGTGTGGGTGGTGGTTGGGA
AGTTTTTACCCATGGTGGTCGTAAACCGACCGGTCTGGATGCAGTTGAATGGGCACGTAA
AGTTGTGGAACTGGGTGCCGGTGAAATTCTGCTGACCAGCATGGATCGTGATGGCACCAA
AGCAGGTTATGATCTGGAACTGACCCGTGCAGTTAGCGAAGCAGTTAGCGTTCCGGTTAT
TGCAAGCGGTGGTGCGGGTGAACTGGAACATTTTGCAGAAGTTTTTGAACTGGAAGGTGC
CGATGCAGCACTGGCAGCAAGCATTTTTCATTTTGGCGAAATTACCATTCGCGAAGTTAA
AGCATATCTGCGTGAACGTGGTATTGAAGTTCGTTAA**CTCGAG**

Amino acid sequence of LUCA-HisF; catalytically essential residues (D11 and D130) [1] and amino acids defining the entrance to the ammonia channel entrance (R5, E46, K99 and E167) [2] are underlined.

MLAK<u>R</u>IIPCL<u>D</u>VKDGRVVKGVNFENLRDAGDPVELAARYDEEGAD<u>E</u>LVFLDITASHEGRETM
LEVVERTAEQVFIPLTVGGGIRSVEDASRLLRAGAD<u>K</u>VSINTAAVKNPELITEAAEEFGSQAVV
VAI<u>D</u>AKRVGGGWEVFTHGGRKPTGLDAVEWARKVVELGAG<u>E</u>ILLTSMDRDGTKAGYDLEL
TRAVSEAVSVPVIASGGAGELEHFAEVFELEGADAALAASIFHFGEITIREVKAYLRERGIEVR

Nucleotide sequence of LUCA-HisH (including restriction sites for *Nde*I and *Xho*I in boldface):

CATATGAGCAAAACCATGCGTGTGGCCATTATTGATTATGGTATGGGTAATCTGCGTAGC
GTTAGCAAAGCACTGGAACGTGTTGGTGCCGAAGTTGTTGTTACCAATGATCCGGAAGAA
CTGAAAGAAGCAGACGCACTGATTCTGCCTGGTGTGGGTGCATTTGATGAAGCAATGAAA
AATCTGCGTTCACGTGGTCTGGTTGAAGTGATTAAAGAAGTGGATAAATTCAACCGTGTG
GCGAAAGGTAAACCGCTGCTGGGTATTTGTCTGGGTATGCAGCTGCTGTTTGAAAGCAGC
GAAGAAGGTGGCACCACCAAAGGTCTGGGCATTATTCCGGGTCGTGTTAACGTATTCGT
AGCGAAGCAGCAGAAGGTGATAATCTGAAAATTCCGCACATGGGTTGGAATCAGGTTAA
TGTTGTTCGTGAAAGCCCTCTGCTGGAAGGTATTCCGGAAGGTAGCTATTTCTATTTTGTG
CACAGCTACTATGTTGTGTATCCGACCAATGAAGAACATATTGTTGCAACCACCGAATAC
TATGGCCAGAAATACACCGCAGCAGTTGCACGTGGTAACATTTTTGGCACCCAGTTTCAT
CCGGAAAAAAGCGGTAAAGCAGGTCTGAAACTGCTGAAAAACTTTCTGGAATGGGTGGA
ACGTGAAAATAATGCACGTCGC**CTCGAG**

Amino acid sequence of LUCA-HisH; putative catalytic triad residues (C89, H199 and E201) [2] and substrate binding residues (G52, Q93, E101 and Y162) [3] are underlined).

MSKTMRVAIIDYGMGNLRSVSKALERVGAEVVVTNDPEELKEADALILPGV<u>G</u>AFDEAMKNL
RSRGLVEVIKEVDKFNRVAKGKPLLGI<u>C</u>LGM<u>Q</u>LLFESSE<u>E</u>GGTTKGLGIIPGRVERIRSEAAEG
DNLKIPHMGWNQVNVVRESPLLEGIPEGSYFYFVHS<u>Y</u>YVVYPTNEEHIVATTEYYGQKYTAA
VARGNIFGTQF<u>H</u>P<u>E</u>KSGKAGLKLLKNFLEWVERENNARRLEHHHHHH

Oligonucleotides used for the construction of LUCA_*his*F_W138Y+W156Y and for the genomic amplification of *zm*HisH

Construction of LUCA_*his*F_W138Y+W156Y (restriction sites are shown in boldface):

5`-LUCA_*his*F_*Nde*I:    5`-GTTGT**CATATG**CTGGCAAAACGTATTATTCCGTGCC-3`

3`- LUCA_*his*F_*Xho*I:    5`-GTGGTG**CTCGAG**TTAACGAACTTCAATACCACGTTCACGC-3`

5`- LUCA_*his*F_W156Y:

5`-GGTGGTCGTAAACCGACCGGTCTGGATGCAGTTGAATACGCACGTAAAGTTGTG-3`

3`- LUCA_*his*F_W138Y:

5`-ACCGGTCGGTTTACGACCACCATGGGTAAAAACTTCGTAACCACCACCCACACG-3`

Genomic amplification of *zm*HisH (restriction sites are shown in boldface):

5`-*zm*HisH_*Nde*I:        5`-GTTTGT**CATATG**AAAAATCAGGCATCTTC-3`

3`-*zm*HisH_*Xho*I:        5`-ATGTGC**CTCGAG**TTAGGGATTCCAGTCCA-3`

5`-*zm*HisH_T393C:        5`-TGTAAGGTGCCTCACATGGGATGGAACCAG-3`

3`-*zm*HisH_T393C:        5`-CTGGTTCCATCCCATGTGAGGCACCTTACA-3`

Comparison of the folding mechanisms of LUCA-HisF, *T. maritima* HisF, and its artificial precursors Sym1 and Sym2

The observed rate constants of unfolding and refolding at 25°C are shown in Figure S5A as function of GdmCl concentration in the form of a chevron plot. The refolding kinetics of HisF from *T. maritima* and its putative evolutionary precursors Sym1 and Sym2 are remarkably similar to each other and to LUCA-HisF [4,5]. In all cases a burst-phase intermediate I_{BP} is rapidly formed within the dead time of the stopped-flow apparatus (Figure S4B). It contains a significant amount of secondary structure but is most likely off-pathway [4,5]. This reaction is followed by a very fast folding reaction with a time constant ($\tau = 1/\lambda$) of 0.27 s, detected for LUCA-HisF and Sym2. Probably, the intermediate I' being formed in this reaction, is energetically unfavorable for *tm*HisF and Sym1 [4]. On the contrary, a subsequent fast folding reaction with a τ-value of about 3 s is only observed for *tm*HisF, Sym1, and Sym2 (Figure S5A). Probably, the intermediate I being formed in this reaction, is energetically unfavorable for LUCA-HisF. In the slow rate-limiting reaction leading to the native state N the refolding limb of LUCA-HisF is shifted to lower denaturant concentrations, resulting in decreased τ values of refolding compared to *tm*HisF, Sym1, and Sym2.

The reciprocal change of the amplitudes of the fast 3s-phase and the slow 20s-phase of *tm*HisF folding with increasing concentration of GdmCl was previously shown to reflect the formation of a GdmCl-sensitive on-pathway folding intermediate I and its subsequent transformation to the native state N in a sequential folding mechanism [5]. Such a sequential folding pathway has also been confirmed for Sym1 and Sym2 [4]. For LUCA-HisF, the amplitudes of the consecutive refolding phases also change in a reciprocal manner with denaturant concentration (Figure S5B). This result suggests that the folding of LUCA-HisF also occurs sequentially via intermediates. Taken together, our findings suggest a common folding mechanism for LUCA-HisF, *tm*HisF, Sym1 and Sym2 (Figure S6).

LUCA-HisF and the other characterized proteins differ strongly in the rate of unfolding (Figure S5A). LUCA-HisF and Sym1 denature relatively fast even at moderate GdmCl concentrations, which allowed for the determination of the complete unfolding limb of their chevron diagrams. In contrast, unfolding of *tm*HisF is roughly three orders of magnitude slower, and its rate could not be determined below 4 M GdmCl. The unfolding of Sym2 is further decelerated about 1000-fold. These strong differences in the unfolding rates account to a large extent for the observed differences in the thermodynamic stabilities between the four proteins (Table S2).

- S6 -

Table S1. Crystal structure determination of LUCA-HisF (PDB ID 4EVZ): Data collection and refinement statistics.

Data collection	
Wavelength, (Å)	1.0
Space group	P 1 21 1
Unit cell dimensions	
a, b, c, (Å)	43.07, 90.74, 57.37
α, β, γ, (°)	90.00, 100.25, 90.00
Resolution, (Å)	47.93-1.46 (1.50-1.46)
Total reflections	214865 (14064)[1]
Unique reflections	69543 (3819)[1]
Mosaicity, (°)	0.064
Completeness, (%)	92.86 (50.95)[1]
R_{merge}	0.05 (0.14)[1]
$I/\sigma(I)$	12.37 (3.59)[1]
Refinement statistics	
Resolution, (Å)	47.93-1.46
R_{cryst} / R_{free} (%)	15.1 / 17.6
Average B-value, (Å^2)	12.4
Number of atoms	
Protein	3845
Ligand	20
Water	667
RMSD of bond length, (Å)	0.010
RMSD of angle, (°)	1.294
Model quality (Molprobity)	
Residues in favored region, (%)	98.6
Residues in additional allowed region, (%)	1.4

[1] The numbers in parentheses indicate the data collection parameters for the resolution between (1.50-1.46) Å.

Table S2. Thermodynamic unfolding parameters of LUCA-HisF, *T. maritima* HisF, Sym1, and Sym2.

	ΔG_D (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})	[D]$_{1/2}$ (M)
LUCA-HisF	17.8 ± 0.4	15.2 ± 0.4	1.2
*tm*HisF	53.5 ± 1.7	19.1 ± 0.6	2.8
Sym1	42.8 ± 0.2	20.2 ± 0.4	2.1
Sym2	62.2 ± 3.9	19.2 ± 1.2	3.2

The Gibbs free energy of denaturation (ΔG_D), the cooperativity (m), and the denaturant concentration required to unfold 50 % of the protein ([D]$_{1/2}$) were obtained by analyzing the fluorescence-detected unfolding transitions (see Figure S4A) with the two-state model. The data for *T. maritima* HisF, Sym1 and Sym2 were taken from references 4 and 5, respectively.
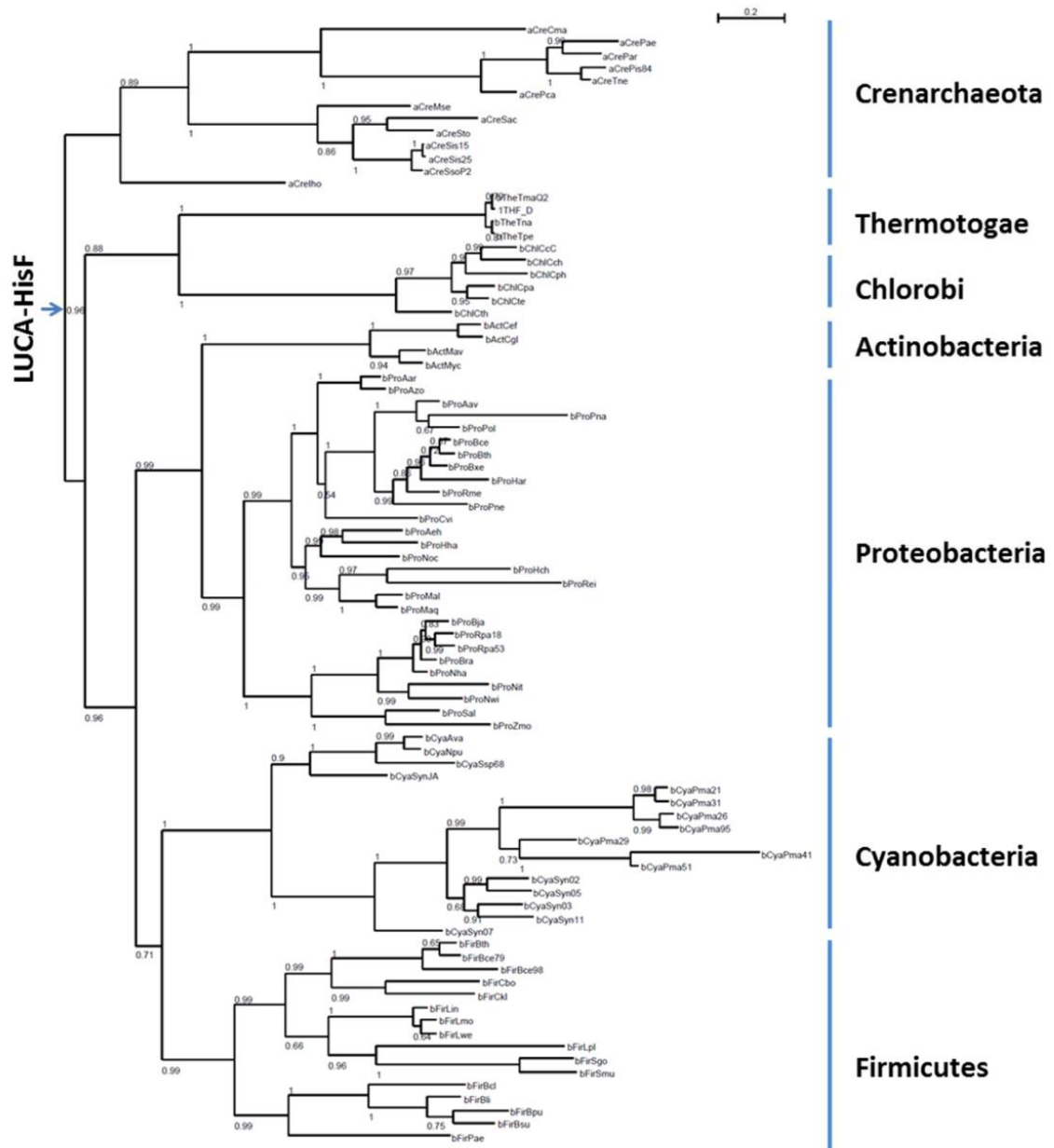
Figure S1. Phylogenetic tree *t_HisF_HisH* **used for the reconstruction of LUCA-HisF and LUCA-HisH.**

87 extant HisF sequences from seven archeal and bacterial clades were used to determine a phylogenetic tree and to deduce ancestral sequences. The tree was rooted to determine the position of LUCA-HisF. Numbers give *posterior* probabilities. For details and abbreviation of species names see ref. 6.

Figure S2. PRANK output for the reconstruction of LUCA-HisH

```
>LUCA-HisH
--------MS---KTMRVAIIDYGMGNL----RSVSKA---LERVG---A----------EV-VVT--NDPE
ELKEA-DALILPGVGAFDEAMKNLRSR-------GLVEVI------KEVDKFN---------RVAKG--KPL
LGICLGMQLLFESSE-----E-GGTT--------KGLGIIPGRVER------I--------R-SEAAEG--D
-NLKIPHMGWNQVNVV-----RE-S-PLL---------EGIPEGS-Y--------------FYFVHSY--YV
VYPT------N---EEHIVAT-TEY-YGQKYTAAV---ARGNIFGTQFHPEKSGKAGLKLLKNFLEW-VERE
NNAR----------R--------------
>aCreCma
------MRVR---GVVKVGVVNYGVGNV----GSIFNA---LRRIE---A----------EP-LLI--NNTA
DLRSV-DAVILPGVGSFNSAMVNLSRL-------T-------------DELN---------RVRGS--SPI
LGICLGLQLMFKGSD-----E-GEL---------RGLGWYDGWVNR------I--------R-G-------
--PRVPHIGWDYVKLS-----GD-C-NL-----------GV-ARG-Y--------------YYFMHSY--AV
VNPV------N---EPPFVGF-TK--YGDSTILSVLCDEGNATYGTQFHPEKSGKLGLSILNGLVNL-----
--AR----------R--------------
>bActCef
--------MT---KTV--ALLDYGSGNL----RSAQRA---LEHVG---A----------EV-IVT--SDPD
ICTNA-DGLLVPGVGAFDACMKGLRGV-------FGHRII------GT-------------RLAGG--RPV
MGICVGMQILFDEGI-----E-HGIQTRGCGEW-------SGRVER------L--------Q-AR-------
---ILPHMGWNTV-EK-QPGS-----EMF---------AGLSEDE-R--------------YYFVHTY--GV
R-DW-TLVTDDLTTPPL-VTW-AVH-ENDRFVAAV---ENGALWATQFHPEKSGDAGAQLLRN---W-----
-INH----------I--------------
>bChlCcC
--------M------V--FIADYGAGNL----RSVLKA---FEFLG---I----------KA-IVS--NDPR
KMAGY-RKVLLPGVGAFGQAMQSLEAL-------GFVSAL------LE-------------HVDKG--GHL
LGICLGMQLLLSESE-----E-MGTH--------KGLNLVPGKVKH------F--------V-SS----SD-
---KIPQIGWNAVDFS-----KQ-S-DLF---------RNVADHS-F--------------FYFVHSY--YC
E-TE------S---VEAVAAT-TLF-AGQNFCSAI---EKNGIFAVQFHPEKSADAGLKVLANFAEL-----
----------------------------
>bCyaAva
------MPV------V--AVIDYEMGNL----HSVCKG---LEKAG---A----------TP-IIT--HSHQ
ELTKA-DAVILPGVGAFDPAVQSLRSR-------DLEQPI------KD--------------TIASG--KPF
LGICLGLQILFESSA-----E--GTQ--------PGLGIIKGKVRR------F--------I-SEP-----
-GITIPHMGWNQLELT-----QP-K-SILW--------EHLPPQP-W--------------VYFVHSY--YV
D-PV------E---PQVRAAT-VTH-GTQTVTAAI---AHENLMAVQFHPEKSSNIGLQILSNFVSQ-V---
--------------R-------EKIAA---
>bFirBce79
--------M------I--AIIDYGMGNI----RSVEQA---LKYIG---A----------EY-IVT--SDKK
EILRS-DGVILPGVGAFPKAMDVLEER-------DLVCVL------KE--------------VCDIG--KPL
LGICLGMQLLFERSE-----E-LKDC--------SGLGLLPGEIRK------L-----------KV------
-SYKIPHMGWNELRKE-----RE-F-PLW---------NGLVDGS-F--------------VYYVHSY--YA
D--C------P---DEIVCGV-SDY-GMQ-VPGFV---AKGNVFGAQFHPEKSGEIGMQILKNFQGV-VE--
------AWKSSQLSI--------------
--------MS----DV--AIIDYGMGNL----RSVAKA---IEHV----APGK-------RV-AVT--SDPA
VVAAA-ARVVFPGQGAMPDCMRELDLR-------GLREVV------KT--------------AAAS--KPF
LGICIGQQMLFEHSE-----E--G------N--VPGLGILPGGVVRFPDAKMV--------A-AD------G
SRLKVPHMGWNEVCQ-----RKP-H-PMW---------EGIPDNE-R--------------FYFVHSY--FV
A-PA------D---AAL-VAA-ESD-YGTRFTSAV---ARANIFAVQFHPEKSAQAGLKMLANFISW-----
-AP--------------------------
>bProAci
--------MNVEAKTV--AVVDYGMGNL----RSVSQA---VQAA----AEGSGW-----TV-VVT--SRPE
DVRAA-QRVVLPGQGAMPDCMRELRES-------GLQESV------LE--------------AAAS--KPL
FGVCVGMQMLLDHSA-----E--G------D--TPGLGLIPGDVVRFELAGRL--------Q-PD------G
SRYKVPQMGWNRVRQM-PHGGAV-H-PVW---------AGIPDEN-Y--------------FYFVHSF--HA
L-PR------D---AAH-TVG-ETD-YGGRFASAV---ARDNIFATQFHPEKSSGHGLALYRNFLHW-----
-NP--------------------------
>bTheTna
------MR------I--GIISVGPGNIMNLYRGVKRASENFEDVS---I----------EL-VESPQDDL-
----Y-DLLFIPGVGHFGEGMRRLREN-------GLVEFI------KK--------------HVEDG--KYV
VGVCLGMQLLFEESE-----E-APGV--------KGLSLIEGNVVK------L--------K-SR-------
---RLPHMGWNE-------------VIF---------KGTFPNG-Y--------------YYFVHTYRAVC
K------------EEHVLGT-TEY-DGEIFPSAV---RKGRILGFQFHPEKSSKIGRKLLEKVIEC-SLSR
R--------------------------
```

Representative sequences belonging to *HisH*$_{ext}$ are shown to illustrate the phylogeny. Be aware of gap placement introduced by PRANK [7]. For reconstruction, PRANK v.130410 was used with the options `prank -uselogs -twice -showall -showanc -protein -support -d=` *HisH*$_{ext}$ `-t=` t_{HisF_HisH}. LUCA-HisH is the reconstructed sequence. For abbreviations of species names see ref. 6.
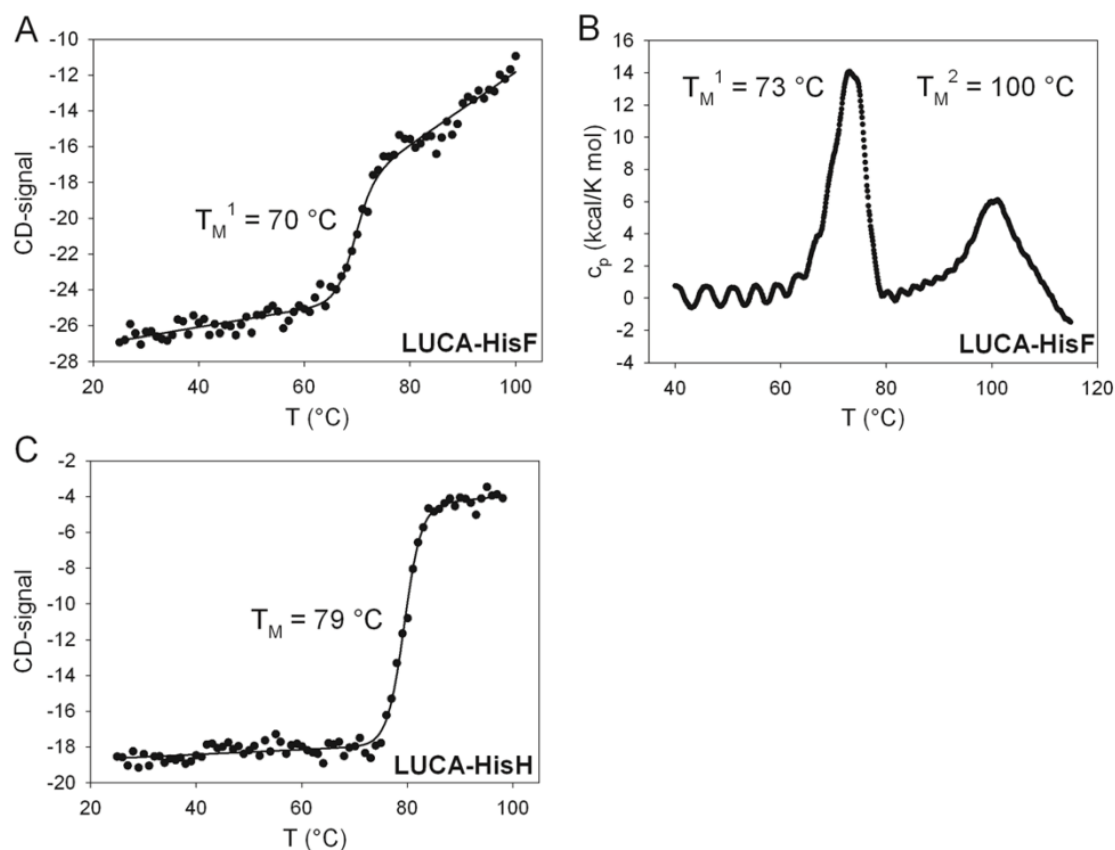
Figure S3. Thermal denaturation of LUCA-HisF and LUCA-HisH.

(A) Thermal denaturation of LUCA-HisF as followed by CD. The unfolding of 10 µM protein in 50 mM potassium phosphate, pH 7.5 was followed at 220 nm. The solid line is shown to guide the eye. The first transition midpoint lies at 70 C, whereas the second transition midpoint could not be determined due to the incompleteness of unfolding. (B) Thermal denaturation of LUCA-HisF as followed by DSC. Unfolding of 35 µM LUCA-HisF in 50 mM potassium phosphate, pH 7.5 resulted in transition midpoints at 73 °C and 100 °C. (C) Thermal denaturation of LUCA-HisH as followed by CD. The unfolding of 9 µM protein in 50 mM potassium phosphate, pH 7.5 was followed at 220 nm. The data were fit with the two-state model (solid line), which yielded a transition midpoint at 79 °C.
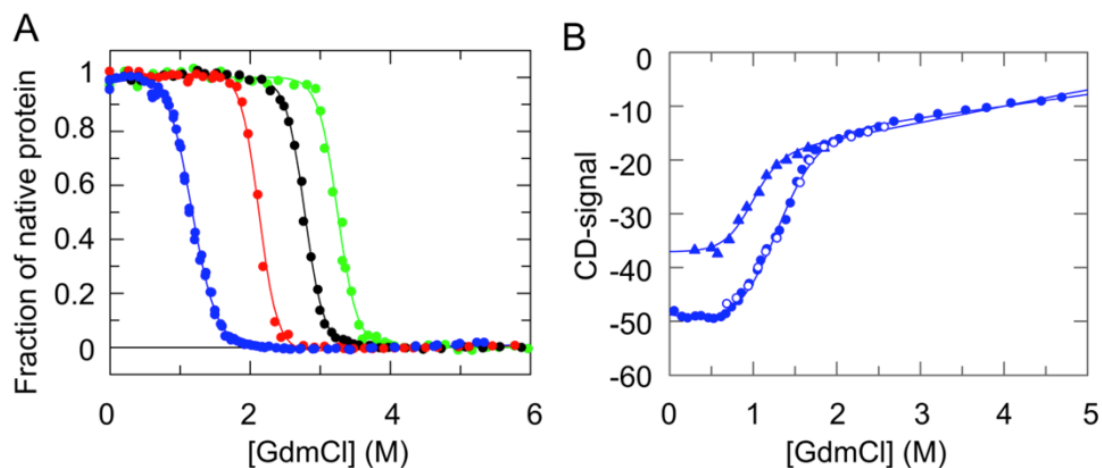
Figure S4. Equilibrium unfolding transitions of LUCA-HisF, *T. maritima* HisF, Sym1, and Sym2, and formation of a burst-phase intermediate by LUCA-HisF.

(A) GdmCl-induced unfolding of LUCA-HisF (blue), *tm*HisF (black), Sym1 (red), and Sym2 (green) was followed in 50 mM Tris/HCl buffer, pH 7.5, at 25 °C by monitoring Tyr/Trp fluorescence (excitation: 280 nm; emission: 320 nm). The continuous lines represent the fit of the two-state model to the normalized unfolding transitions, yielding the thermodynamic parameters listed in Table S2. The data for *T. maritima* HisF, Sym1 and Sym2 were taken from references 4 and 5, respectively. (B) Equilibrium unfolding/refolding transitions of LUCA-HisF in comparison with the unfolding transition of its burst-phase intermediates (I_{BP}). GdmCl-induced unfolding/refolding was followed in 50 mM Tris/HCl buffer, pH 7.5 at 25 °C by the far-CD signal at 225 nm. Closed circles represent unfolding experiments, started with folded protein and open circles represent refolding experiments, started with protein that was previously unfolded in 6.0 M GdmCl. The unfolding transition of I_{BP} (filled triangles) was calculated by plotting the initial CD signal upon refolding, which was obtained by extrapolation of the kinetics to zero time, as a function of the corresponding GdmCl concentration.
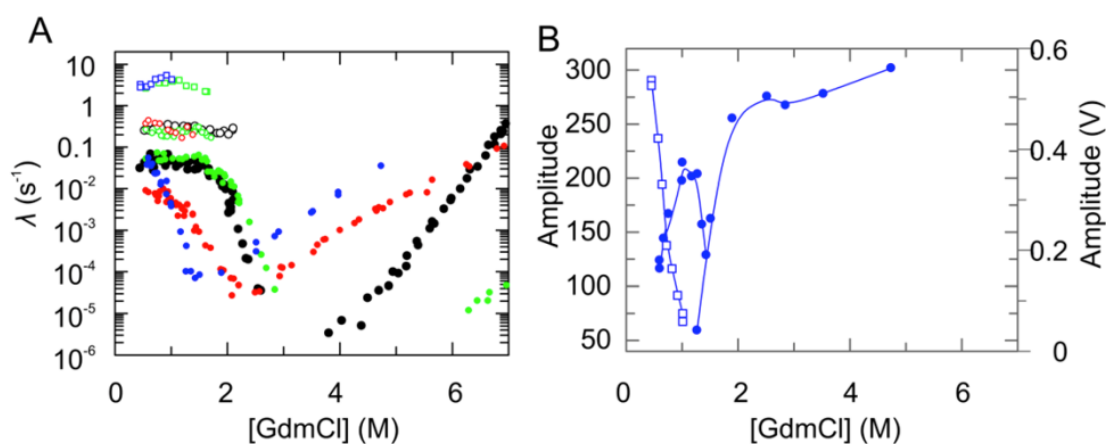
Figure S5. Apparent rate constants (λ) and amplitudes of refolding and unfolding kinetics of LUCA-HisF.

(A) Chevron diagram showing the apparent rate constantsfor the refolding/unfolding reactions of LUCA-HisF (blue), *tm*HisF (black), Sym1 (red) and Sym2 (green). Reactions were followed in 50 mM Tris/HCl buffer, pH 7.5, at 25 °C by stopped-flow or manual mixing monitoring Trp/Tyr fluorescence (excitation: 280 nm; emission: 320 nm). The dependence on GdmCl concentration of the apparent rate constant (λ) is shown for the slow unfolding phases leading to U and the slow refolding phases leading to N (filled circles), for the fast refolding phase leading to I (open circles), and for the very rapid refolding phase leading to I' (open squares). The data for *T. maritima* HisF, Sym1 and Sym2 were taken from references 4 and 5, respectively.

(B) Amplitudes of the very rapid (open squares) and slow (closed circles) refolding phases and the unfolding phase (closed circles) of LUCA-HisF.
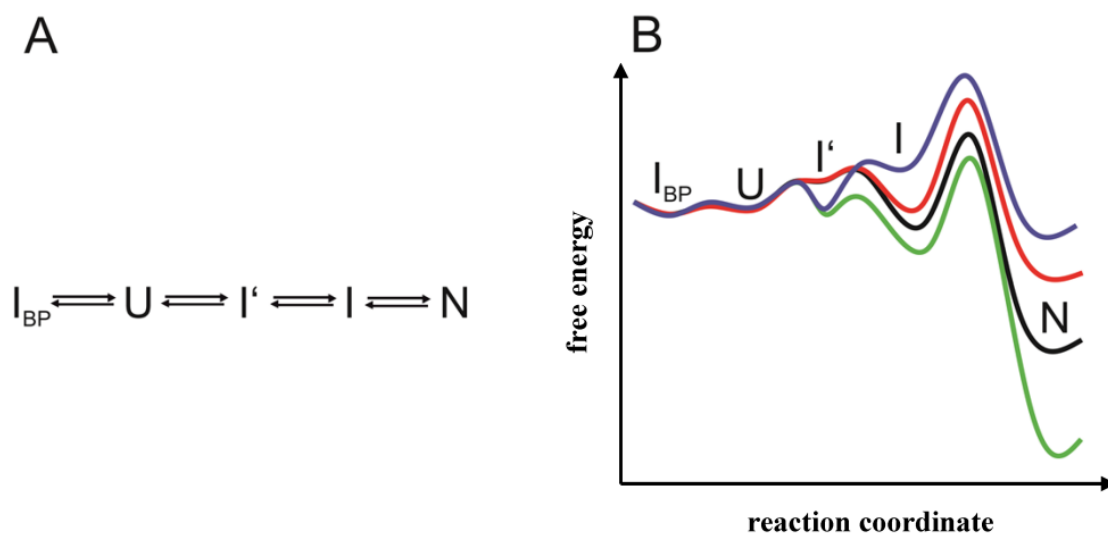
Figure S6. Unifying folding mechanism for LUCA-HisF, *T. maritima* HisF, and Sym1 and Sym2.

(A) An off-pathway equilibrium between the unfolded state (U) and the burst-phase intermediate (I_{BP}) precedes the formation of the on-pathway intermediates I' and I. For *tm*HisF and Sym1, I' is assumed to be a high energy intermediate. For LUCA-HisF, I is assumed to be a high energy intermediate. (B) Energy diagram for the folding of LUCA-HisF (blue), and *tm*HisF (black), Sym1 (red), and Sym2 (green). The different heights of the energy barriers indicate different folding/unfolding rates but are not at scale. The data for *T. maritima* HisF, Sym1 and Sym2 were taken from references 4 and 5, respectively.

References for Supporting Information

(1) Beismann-Driemeyer, S.; Sterner, R. *Journal of Biological Chemistry* **2001**, *276*, 20387.

(2) Douangamath, A.; Walker, M.; Beismann-Driemeyer, S.; Vega-Fernandez, M. C.; Sterner, R.; Wilmanns, M. *Structure* **2002**, *10*, 185.

(3) List, F.; Vega, M. C.; Razeto, A.; Hager, M. C.; Sterner, R.; Wilmanns, M. *Chemistry and Biology* **2012**, *19*, 1589.

(4) Carstensen, L.; Sperl, J. M.; Bocola, M.; List, F.; Schmid, F. X.; Sterner, R. *Journal of the American Chemical Society* **2012**, *134*, 12786.

(5) Carstensen, L.; Zoldak, G.; Schmid, F. X.; Sterner, R. *Biochemistry* **2012**, *51*, 3420.

(6) Richter, M.; Bosnali, M.; Carstensen, L.; Seitz, T.; Durchschlag, H.; Blanquart, S.; Merkl, R.; Sterner, R. *Journal of Molecular Biology* **2010**, *398*, 763.

(7) Löytynoja, A.; Goldman, N. *Science* **2008**, *320*, 1632.

5.3   Publication C

Exploiting protein symmetry to design light-controllable enzyme inhibitors.

Bernd Reisinger, Natascha Kuzmanovic, Patrick Löffler, Rainer Merkl, Burkhard König and Reinhard Sterner

Light-Regulated Enzyme Activity

Exploiting Protein Symmetry To Design Light-Controllable Enzyme Inhibitors**

Bernd Reisinger, Natascha Kuzmanovic, Patrick Löffler, Rainer Merkl, Burkhard König, and Reinhard Sterner**

Abstract: *The activity of the metabolic branch-point enzyme PriA from Mycobacterium tuberculosis (mtPriA) can be controlled reversibly by light. Two-pronged inhibitors based on the dithienylethene scaffold were designed utilizing mtPriA's natural rotational symmetry. Switching from the flexible, ring-open to the rigid, ring-closed isomer reduces inhibition activity by one order of magnitude.*

The artificial control of biological processes by light is a rapidly emerging area of protein design.[1] Three basic strategies for the light regulation of biomolecules have been reported: key positions have been functionalized with photolabile protecting groups,[2] naturally occurring photoreceptors have been reprogrammed,[3] and designed molecules that can be reversibly switched by light (photoswitches) have been used to direct protein or cellular function.[4] With respect to the latter, substantial progress has been made in the regulation of neuronal activity by designing light-inducible ligands for ion channels and receptors.[5] As the molecular recognition of specific ligand parts leads to a nonlinear signal response in neural systems, even small changes in the binding efficacy upon light irradiation significantly influence the cellular output.[6] In contrast, for the reversible control of enzymatic activities, the switching of a photoresponsive group must substantially affect the

enzyme's active site. This task can be fulfilled either by the covalent incorporation of a molecular photoswitch near the catalytic center[4,7] or by the design of a noncovalently bound, light-controlled inhibitor.[4,8]

We aimed to design a light-controlled inhibitor for phosphoribosyl isomerase A from *Mycobacterium tuberculosis* (*mt*PriA). *mt*PriA is a branch-point enzyme in amino acid biosynthesis as it catalyzes two chemically equivalent sugar isomerization reactions in tryptophan and histidine biosynthesis.[9] In the latter, the aminoaldose *N'*-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) is converted to the corresponding
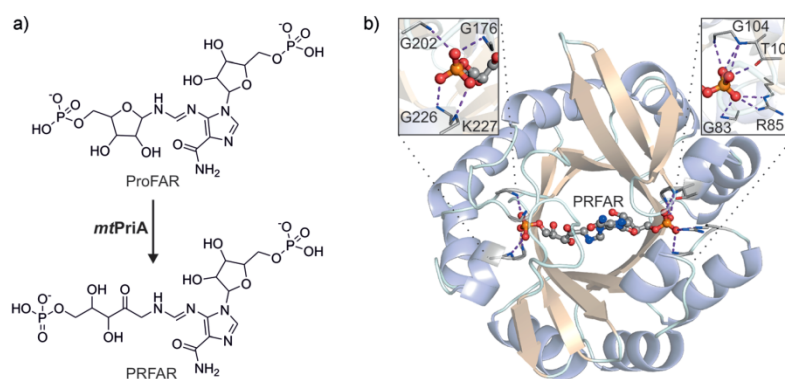
Figure 1. Reaction and structure of *mt*PriA. a) *mt*PriA catalyzes the conversion of ProFAR to PRFAR in the histidine biosynthesis. b) Ribbon representation of the $(\beta\alpha)_8$-barrel structure of *mt*PriA with bound product PRFAR (PDB ID: 3ZS4[12a]). The view is along the twofold symmetry axis of the protein. PRFAR is anchored by two opposite phosphate binding sites, which are enlarged in the insets. Hydrogen bonds are indicated by dashed lines.

[*] B. Reisinger,[+] P. Löffler, Prof. Dr. R. Merkl, Prof. Dr. R. Sterner
Institut für Biophysik und physikalische Biochemie
Universität Regensburg, 93040 Regensburg (Germany)
E-mail: Reinhard.Sterner@ur.de

N. Kuzmanovic,[+] Prof. Dr. B. König
Institut für Organische Chemie
Universität Regensburg, 93040 Regensburg (Germany)
E-mail: Burkhard.Koenig@ur.de

[+] These authors contributed equally to this work.

aminoketose *N'*-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) (Figure 1a). Since humans can synthesize neither histidine nor tryptophan, *mt*PriA is a potential target for anti-tuberculosis drugs.[10] Structurally, *mt*PriA belongs to the class of $(\beta\alpha)_8$-barrels, which is a frequently encountered and highly versatile fold among enzymes.[11] The protein exhibits a clear twofold symmetry (Figure 1b),[12] which indicates its evolution from a $(\beta\alpha)_4$-half-barrel precursor.[13] Consequently, two phosphate binding sites are found opposite each other to fix the substrate ProFAR and the product PRFAR (Figure 1b). We thus reasoned that a C_2-symmetric photoswitch with terminal phosphate anchors would be an excellent foundation for building a light-controllable inhibitor of *mt*PriA.

Wiley Online Library

Two types of organic photochromic systems possess the desired twofold rotational symmetry: stilbene[14] and azobenzene switches[15] and the diarylethene scaffold.[16] Although azobenzene derivatives have been widely used in biological systems, they suffer from incomplete photoconversion and thermal reversibility.[1] In contrast, photoresponsive compounds based on 1,2-dithienylethene (DTE) generally feature photoconversions of over 90% with both photoisomers being thermally stable.[8b,16] Hence, we opted for DTE as a core and provided it with different phosphate and phosphonate anchors (Scheme 1). Starting with the bis-chlorodithienylethene **1**, Suzuki coupling yielded either the aromatic hydroxides **2–4** or the aromatic bromides **8** and **9**. The former were subsequently converted to *ortho-*, *meta-* and *para-*phosphates **5–7**, while the latter were used to synthesize the phosphonic acid esters **10** and **11**, which were finally hydrolyzed to afford *meta-* and *para-*phosphonates **12** and **13**.

The DTE molecular structure can be toggled reversibly between a ring-open and a ring-closed photoisomer (Table 1), which significantly alters its overall conformational flexibility.[8b] Energy minimizations of the open and closed forms of all potential inhibitors confirmed distances between the phosphorus atoms of 15.8 to 19.6 Å (Table 1), which is in good accordance with the 16.9 Å observed for the corresponding atoms in the *mt*PriA structure with co-crystallized PRFAR (PDB ID: 3ZS4[12a]; see the Supporting Information for details). Only the open and closed isomers of *ortho-*phosphate **5** exhibit rather short P–P distances of 12.3 Å and 11.3 Å, respectively, in their energetically most favorable geometries; in addition, less populated, more extended conformers can be observed.

When ring-open forms of compounds **5–7**, **12**, and **13** are irradiated with UV light (312 nm), the absorption band at 280 nm immediately decreases. Simultaneously, new absorption maxima at 350 nm and 525 nm formed, turning the initially colorless solutions pink (Figure S1). In each case, the spectral changes are complete after 30 s of irradiation and the corresponding photostationary states consist of between 93% and 97% of the closed isomers, as judged by HPLC analyses (Figure S2). The open forms can be recovered by irradiation with visible light (>420 nm) and all switches are robust over several ring-closing/ring-opening cycles (Figure S3).

The activity of *mt*PriA can be measured spectrophotometrically at 300 nm in a coupled enzyme assay (Scheme S1).[17] As all synthesized compounds were stable under the assay conditions, their inhibitory effect could be investigated in steady-state enzyme kinetics. For this purpose, substrate saturation curves were monitored in the presence of different concentrations of compounds **5–7**, **12**, and **13** in their open and closed forms (curves are shown for compound **6** in Figure S4). Indeed, all investigated DTE phosphates and
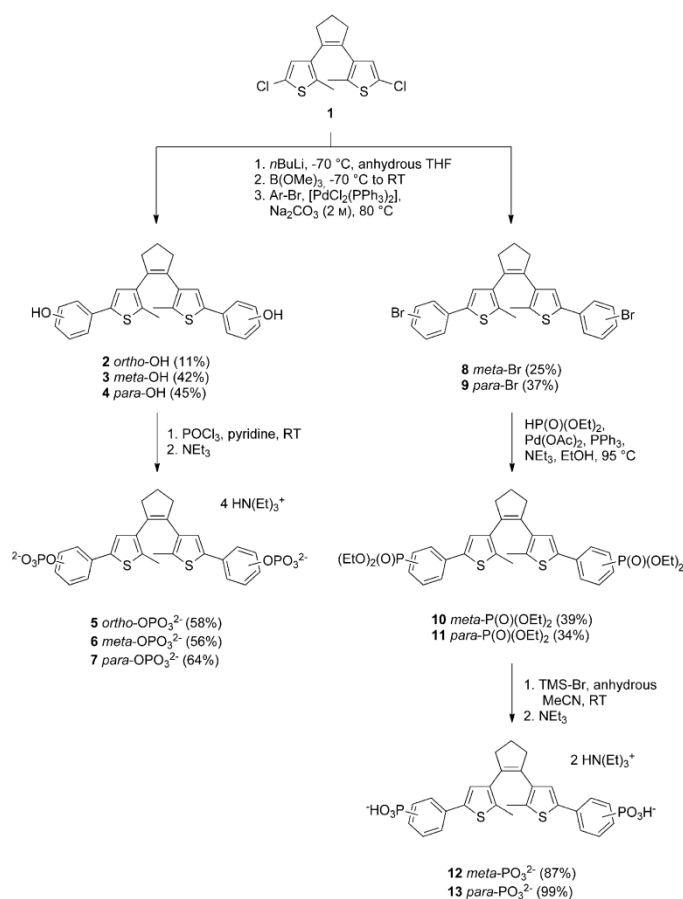
Scheme 1. Synthesis of DTE phosphates and DTE phosphonates.

Table 1: Photochemical switching and corresponding calculated P–P distances in DTE phosphates and DTE phosphonates **5–7**, **12**, and **13**.

Compound	d(P–P) [Å]	
	open	closed
5	12.3	11.3
6	16.7	16.2
7	19.6	18.4
12	16.9	15.8
13	18.9	18.1

DTE phosphonates are able to inhibit the *mt*PriA reaction in both isomeric forms, thus proving the viability of the design concept. As expected for competitive inhibition, the turnover numbers k_{cat} were identical in the presence and absence of

inhibitor (Table S1). The observed increase of the Michaelis constants caused by the inhibitors (Table S1) was used to calculate the inhibition constants K_i (Equation S1), which are given in Table 2.

Table 2: Inhibition constants K_i for compounds **5–7**, **12**, and **13** in their ring-open and ring-closed forms.

Inhibitor	K_i [μM][a]	
	ring-open	ring-closed
5	8.1 ± 2.1	7.0 ± 1.2
6	0.55 ± 0.12	4.4 ± 0.3
7	3.5 ± 0.3	3.7 ± 0.4
12	1.6 ± 0.1	4.9 ± 0.3
13	6.8 ± 0.6	22.7 ± 4.7

[a] The values of K_i were obtained from the data shown in Table S1 using Equation S1.

The determined inhibition constants are in the low micromolar range (Table 2) and therefore comparable or even better than the enzyme's K_M value for the natural substrate ProFAR ($K_M^{ProFAR} = 8.6$ μM; Table S1). In agreement with the appropriate distance between its phosphorus atoms (Table 1), the open isomer of phosphate **6** exhibits the highest binding affinity ($K_i = 0.55$ μM). However, when **6** is switched to its rigid, closed form, the inhibition activity is lowered by roughly one order of magnitude ($K_i = 4.4$ μM). A similar trend is observed for the corresponding phosphonate **12**, whose binding affinity in the open form ($K_i = 1.6$ μM) decreases about threefold upon ring-closure ($K_i = 4.9$ μM). In contrast, the inhibitory effects of phosphates **5** and **7** are nearly identical in their ring-open and ring-closed forms (Table 2). Interestingly, the introduction of a phosphonate moiety in *para* position (compound **13**) causes a threefold difference in the inhibition activity of the open ($K_i = 6.8$ μM) and closed photoisomer ($K_i = 22.7$ μM). The removal of the oxygen bridge between the terminal anchor and the switchable core further reduces the overall flexibility, which seems to predominantly affect the already rigid, closed isomer.

The performance of *mt*PriA can also be controlled by irradiation with light during catalysis. When the reaction was started with compound **6** in its strongly inhibiting, open form, the reaction rate could be enhanced by about threefold upon switching to the less active, ring-closed isomer (Figure 2).

As pointed out before, the mechanistic principle behind the different binding affinities of the photoisomers is based on a change in conformational flexibility.[8b] Due to free rotation around the C−C bonds joining the thiophene heterocycles to the central cyclopentene ring and the terminal phenyl groups, DTE phosphate **6** is able to adopt various geometries in its ring-open form. On the other hand, the closed isomer is completely conjugated and thus far more restricted in its mobility. To gain insight into the binding modes of the inhibitors, we performed molecular dynamics (MD) simulations of the open and closed isomers of compound **6** bound to *mt*PriA. Although both forms are clearly fixed at the phosphate binding sites (Figure 3a,c), obvious differences can be observed in their structural cores. While the open isomer converges to similar, C_2-symmetric conformers in
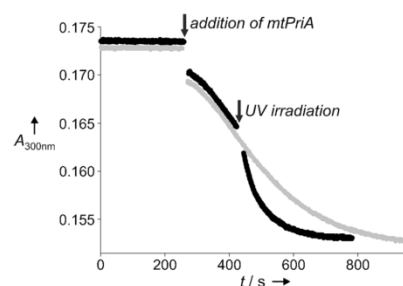
Figure 2. Change in *mt*PriA activity upon ring-closure of compound **6**. The turnover of 5 μM ProFAR was followed photometrically at 300 nm in the presence of 4 μM **6** in its open form under typical assay conditions (25 °C, 50 mM Tris/acetate pH 8.5, 100 mM ammonium acetate, 0.18 μM HisF, and 0.15 μM *mt*PriA). After having reached its maximum rate, the reaction mixture was either left in the spectrophotometer (gray) or removed and irradiated with 312 nm light for 10 s (black). A reference solution without enzymes was used to correct the baseline shift resulting from different absorption values of the open and closed isomer at 300 nm (see Figure S1).
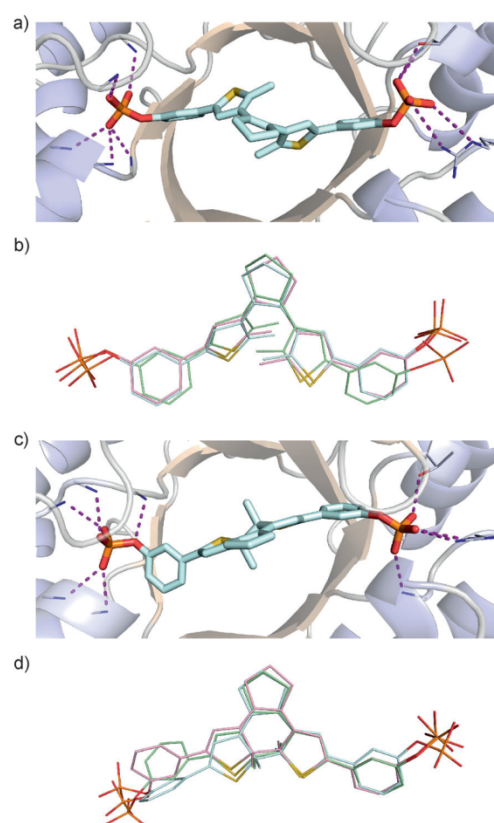
Figure 3. MD simulations of *mt*PriA and bound *meta*-phosphate **6**. For each isomer, three independent calculations were performed and representative enzyme structures for the open (a) and closed (c) form are shown. A superposition of the energetically most favorable conformer of each simulation is depicted in for the open conformation (b) and the closed conformation (d).

three independent calculations (Figure 3b), more diverse binding modes are found for the ring-closed isomer (Figure 3d). Here, one terminal phenyl ring is twisted and adopts various geometries to facilitate proper coordination of the terminal phosphate group. The measured difference in inhibition activity is also reflected in the binding energies determined during the simulations, which consistently show that interaction with the open form is energetically more favorable (Table S2). Taken together, the higher flexibility of the open isomer allows for better adaptation to the enzyme's active site and apparently overcompensates for the loss in entropy upon binding.

In summary, we have demonstrated that natural protein symmetry can be advantageously utilized to design light-controllable enzyme inhibitors. The two-pronged DTE switches can be toggled reversibly between a high- and low-affinity form, where both photoisomers are nearly quantitatively formed and thermally stable. Hence, the enzyme's performance can alternately be enhanced and reduced by irradiation with UV and visible light, respectively. The viability of a dual-anchored DTE inhibitor has been shown before[8b] and the design concept can in principle be transferred to functionally quite different enzyme systems. Phosphate is a frequently encountered element of metabolic substrates, and various other $(\beta\alpha)_8$-barrel enzymes such as pyridoxine 5′-phosphate synthase[18] and aldolases[19] possess two phosphate binding sites. Thus, for these enzymes, inhibitors may be designed by a similar approach, which would allow for the photocontrol of several metabolic processes independently in a spatiotemporal fashion.

[1] C. Brieke, F. Rohrbach, A. Gottschalk, G. Mayer, A. Heckel, *Angew. Chem.* **2012**, *124*, 8572–8604; *Angew. Chem. Int. Ed.* **2012**, *51*, 8446–8476.

[2] A. Deiters, *ChemBioChem* **2010**, *11*, 47–53.

[3] a) U. Krauss, T. Drepper, K. E. Jaeger, *Chem. Eur. J.* **2011**, *17*, 2552–2560; b) O. Yizhar, L. E. Fenno, T. J. Davidson, M. Mogri, K. Deisseroth, *Neuron* **2011**, *71*, 9–34; c) F. Zhang, J. Vierock, O. Yizhar, L. E. Fenno, S. Tsunoda, A. Kianianmomeni, M. Prigge, A. Berndt, J. Cushman, J. Polle et al., *Cell* **2011**, *147*, 1446–1457.

[4] W. Szymanski, J. M. Beierle, H. A. Kistemaker, W. A. Velema, B. L. Feringa, *Chem. Rev.* **2013**, *113*, 6114–6178.

[5] a) T. Fehrentz, M. Schönberger, D. Trauner, *Angew. Chem.* **2011**, *123*, 12362–12390; *Angew. Chem. Int. Ed.* **2011**, *50*, 12156–12182; b) I. Tochitsky, M. R. Banghart, A. Mourot, J. Z. Yao, B. Gaub, R. H. Kramer, D. Trauner, *Nat. Chem.* **2012**, *4*, 105–111.

[6] M. R. Banghart, A. Mourot, D. L. Fortin, J. Z. Yao, R. H. Kramer, D. Trauner, *Angew. Chem.* **2009**, *121*, 9261–9265; *Angew. Chem. Int. Ed.* **2009**, *48*, 9097–9101.

[7] S. Muramatsu, K. Kinbara, H. Taguchi, N. Ishii, T. Aida, *J. Am. Chem. Soc.* **2006**, *128*, 3764–3769.

[8] a) S. Herre, T. Schadendorf, I. Ivanov, C. Herrberger, W. Steinle, K. Rück-Braun, R. Preissner, H. Kuhn, *ChemBioChem* **2006**, *7*, 1089–1095; b) D. Vomasta, C. Högner, N. R. Branda, B. König, *Angew. Chem.* **2008**, *120*, 7756–7759; *Angew. Chem. Int. Ed.* **2008**, *47*, 7644–7647; c) D. Vomasta, A. Innocenti, B. König, C. T. Supuran, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 1283–1286.

[9] F. Barona-Gomez, D. A. Hodgson, *EMBO Rep.* **2003**, *4*, 296–300.

[10] a) K. Mdluli, M. Spigelman, *Curr. Opin. Pharmacol.* **2006**, *6*, 459–467; b) H. Shen, F. Wang, Y. Zhang, Q. Huang, S. Xu, H. Hu, J. Yue, H. Wang, *FEBS J.* **2009**, *276*, 144–154.

[11] a) B. Höcker, C. Jürgens, M. Wilmanns, R. Sterner, *Curr. Opin. Biotechnol.* **2001**, *12*, 376–381; b) R. K. Wierenga, *FEBS Lett.* **2001**, *492*, 193–198.

[12] a) A. V. Due, J. Kuper, A. Geerlof, J. P. von Kries, M. Wilmanns, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3554–3559; b) J. Kuper, C. Doenges, M. Wilmanns, *EMBO Rep.* **2005**, *6*, 134–139.

[13] F. List, R. Sterner, M. Wilmanns, *ChemBioChem* **2011**, *12*, 1487–1494.

[14] D. H. Waldeck, *Chem. Rev.* **1991**, *91*, 415–436.

[15] A. A. Beharry, G. A. Woolley, *Chem. Soc. Rev.* **2011**, *40*, 4422–4437.

[16] M. Irie, *Chem. Rev.* **2000**, *100*, 1685–1716.

[17] T. J. Klem, V. J. Davisson, *Biochemistry* **1993**, *32*, 5177–5186.

[18] M. G. Franco, B. Laber, R. Huber, T. Clausen, *Structure* **2001**, *9*, 245–253.

[19] a) E. Lorentzen, E. Pohl, P. Zwart, A. Stark, R. B. Russell, T. Knura, R. Hensel, B. Siebers, *J. Biol. Chem.* **2003**, *278*, 47253–47260; b) T. Wagner, I. A. Shumilin, R. Bauerle, R. H. Kretsinger, *J. Mol. Biol.* **2000**, *301*, 389–399.

Supporting Information

Table of Contents

Experimental Material and Methods

Synthesis and characterization of new compounds

General. Commercial reagents and starting materials were purchased from Acros Organics, Alpha-Aesar or Sigma Aldrich and used without further purification. Solvents were used in p.a. quality and dried according to common procedures, if necessary. Flash column chromatography was performed on a Biotage Isolera One automated flash purification system with UV/Vis detector using Sigma Aldrich MN silica gel 60 M (40-63 μm, 230-400 grain diameter) for normal phase or pre-packed Biotage SNAP cartridges (KP-C18-HS) for reversed phase chromatography. Reaction monitoring via TLC was performed on alumina plates coated with silica gel (Merck silica gel 60 F_{254}, 0.2 mm). NMR spectra were recorded on a Bruker Avance 400 (^{1}H 400.13 MHz, ^{13}C 100.61 MHz, ^{31}P 161.96 MHz, T = 300 K) instrument. The spectra are referenced against the NMR-solvent, chemical shifts δ are reported in ppm and coupling constants J are given in Hz. Resonance multiplicity is abbreviated as: s (singlet), d (doublet), t (triplet), q (quadruplet), m (multiplet) and b (broad). Carbon NMR signals are reported using DEPT 135 spectra with (+) for primary/tertiary, (−) for secondary and (q) for quaternary carbons. Mass spectra were recorded on Finnigan MAT95 (EI-MS), Agilent Q-TOF 6540 UHD (ESI-MS, APCI-MS), Finnigan MAT SSQ 710 A (EI-MS, CI-MS) or ThermoQuest Finnigan TSQ 7000 (ES-MS, APCI-MS) spectrometer. UV/Vis absorption spectroscopy was performed using a Varian Cary BIO 50 UV/Vis/NIR spectrometer.

Photochemistry. Standard hand-held lamps (Herolab, 312 nm, 6 W) were used for visualizing TLC plates and to carry out the ring-closure reactions at 312 nm. The ring-opening reactions were performed with the light of a 200 W tungsten light bulb which was passed through a 420 nm cut-off filter to eliminate higher energy light. The power of the light is given based on the specifications supplied by the company when the lamps were purchased. A light detector was not used to measure the intensity during the irradiation experiments.

General procedure A for Suzuki coupling. 1,2-Bis(5-chloro-2-methylthiophen-3-yl)cyclopent-1-ene **1**[1] (1.0 eq) was dissolved in dry THF under nitrogen atmosphere. After cooling to -78 °C, *n*-butyl lithium (2.2 eq) was added dropwise via syringe and the resulting purple solution was stirred in the cold for 45 min. The reaction mixture was treated with trimethyl borate (4.0 eq) and the yellow solution was stirred at -78 °C for 30 min. Subsequently, the cold bath was removed and the reaction was stirred at r.t. for further 60 min to form the boronic ester intermediate. After quenching with aqueous sodium carbonate (2 M) the two phase system was degassed for 15 min by nitrogen purge. Then the appropriate aryl bromide (2.2 or 4.0 eq) and Pd(PPh$_3$)$_2$Cl$_2$ (10 mol%) were added and the reaction mixture was heated to reflux overnight. After cooling to r.t., it was diluted with EtOAc and water and the phases were separated. The aqueous phase was extracted with EtOAc twice, the combined organic phases were dried over sodium sulfate and the solvents removed under reduced

pressure. Purification of the crude product was performed by automated flash column chromatography.

General procedure B for phosphate ester formation. A crimp top vial was equipped with the desired bisphenol DTE (1.0 eq) and a stirring bar and capped. It was set under nitrogen atmosphere and pyridine (20 eq) in dry CH_2Cl_2 (1 ml) was added by syringe. After cooling to 0 °C a solution of phosphorous oxychloride (10 eq) in dry THF (1 ml) was added dropwise via syringe over 5 min. The purple reaction mixture was stirred for 2 h at r.t. under nitrogen atmosphere and subsequently quenched with acetone/water (1:1, 2 ml). Volatiles were removed *in vacuo* and the crude product was purified by automated reversed phase flash column chromatography.

Synthesis of 2,2'-(4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))diphenol (2). Compound **2** was prepared from 1,2-bis(5-chloro-2-methylthiophen-3-yl)cyclopent-1-ene **1** (1.50 g, 4.55 mmol), *n*-butyl lithium (1.3 M in hexane, 7.68 ml, 10.0 mmol), trimethyl borate (2.03 ml, 18.2 mmol), 2-bromophenol (1.16 ml, 10.0 mmol) and Pd(PPh₃)₂Cl₂ (320 mg, 0.46 mmol) following the general procedure **A** for Suzuki coupling with 50 ml dry THF and 20 ml aq. Na_2CO_3 (2 M). Automated flash column chromatography (PE/EtOAc, 15-25% EtOAc) and subsequent reversed phase automated flash column chromatography (MeOH/water with 0.05% TFA, 3-100% MeOH) afforded 223 mg (0.50 mmol, 11%) of **2** as purple oil. **¹H-NMR: δ_H(400 MHz, CDCl₃):** 2.10 (2H, dt, J = 7.4, 7.5 Hz, cyclopentene-CH_2), 2.16 (6H, s, 2 CH_3), 2.84 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 6.85 (2H, s, 2 thiophene-H), 6.89 − 6.97 (4H, m, 4 phenyl-H), 7.15 − 7.22 (2H, m, phenyl-H), 7.30 (2H, dd, J = 7.6, 1.6 Hz, phenyl-H); **¹³C-NMR: δ_C(101 MHz, CDCl₃):** 14.1 (+), 23.2 (−), 37.8 (−), 116.3 (+), 120.8 (+), 121.1 (q), 127.6 (+), 129.2 (+), 129.8 (+), 134.5 (q), 135.4 (q), 135.6 (q), 136.6 (q), 152.4 (q); **ESI-MS:** m/z (%): 443.1 (100) [M-H]⁻; calcd. for $C_{27}H_{24}O_2S_2$ 443.1145; found 443.1147.

Synthesis of 3,3'-(4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))diphenol (3). Compound **3** was prepared from 1,2-bis(5-chloro-2-methylthiophen-3-yl)cyclopent-1-ene **1** (1.84 g, 5.59 mmol), *n*-butyl lithium (1.6 M in hexane, 7.68 ml, 12.3 mmol), trimethyl borate (2.49 ml, 22.4 mmol), 3-bromophenol (2.13 g, 12.3 mmol) and Pd(PPh₃)₂Cl₂ (393 mg, 0.56 mmol) following the general procedure **A** for Suzuki coupling with 100 ml dry THF and 20 ml aq. Na_2CO_3 (2 M). Automated flash column chromatography (PE/EtOAc, 15-30% EtOAc) and subsequent reversed phase automated flash column chromatography (MeOH/water with 0.05% TFA, 30-100% MeOH) afforded 1.04 g (2.35 mmol, 42%) of **3** as red oil. **¹H-NMR: δ_H(400 MHz, CDCl₃):** 2.00 (6H, s, 2 CH_3), 2.08 (2H, dt, J = 7.4 Hz, cyclopentene-CH_2), 2.83 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 3.51 (2H, bs, 2 phenyl-OH), 6.71 (2H, dd, J = 1.9 Hz, 8.0 Hz, 2 phenyl-H), 6.95 − 6.99 (2H, m, 2 phenyl-H), 7.00 (2H, s, 2 thiophene-H), 7.07 (2H, J = 7.9 Hz, 2 phenyl-H), 7.19 (2H, t, J = 7.9 Hz, 2 phenyl-H); **¹³C-NMR: δ_C(101 MHz, CDCl₃):** 14.4 (+), 23.0 (−), 38.4 (−), 112.2 (+), 114.1 (+), 117.9 (+), 124.4 (+), 130.1 (+), 134.7 (q), 134.8 (q), 136.1 (q), 136.7 (q), 139.2 (q), 156.0 (q); **EI-MS:** m/z (%) 444.1 (100) [M]⁺, 429.0 (35) [M-CH₃]⁺; calcd. for $C_{27}H_{24}O_2S_2$ 443.1145; found 443.1156.

Synthesis of 4,4'-(4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))diphenol (4). Compound **4** was prepared from 1,2-bis(5-chloro-2-methylthiophen-3-yl)cyclopent-1-ene **1** (209 mg, 0.64 mmol), n-butyl lithium (1.6 M in hexane, 0.88 ml, 1.41 mmol), trimethyl borate (0.29 ml, 2.56 mmol), 4-bromophenol (439 mg, 2.54 mmol) and Pd(PPh$_3$)$_2$Cl$_2$ (42 mg, 0.06 mmol) following the general procedure **A** for Suzuki coupling with 50 ml dry THF and 10 ml aq. Na$_2$CO$_3$ (2 M). Automated flash column chromatography (PE/EtOAc, 15-25% EtOAc), subsequent reversed phase automated flash column chromatography (MeOH/water with 0.05% TFA, 30-100% MeOH) and lyophilization afforded 127 mg (0.29 mmol, 45%) of **4** as red oil. ^{1}**H-NMR:** δ_H**(400 MHz, CDCl$_3$):** 1.98 (6H, s, 2 CH_3), 2.10 (2H, p, J = 7.7 Hz, cyclopentene-CH_2), 2.83 (4H, t, J = 7.5 Hz, 2 cyclopentene-CH_2), 4.88 (2H, bs, 2 phenyl-OH), 6.80 (4H, d, J = 8.7 Hz, 4 phenyl-H), 6.90 (2H, s, 2 thiophene-H), 7.37 (4H, d, J = 8.7 Hz, 4 phenyl-H); ^{13}**C-NMR:** δ_C**(101 MHz, CDCl$_3$):** 14.4 (+), 23.0 (−), 38.5 (−), 115.7 (+), 123.1 (+), 126.8 (+), 127.7 (q), 133.5 (q), 134.6 (q), 136.6 (q), 139.4 (q), 154.8 (q); **EI-MS:** m/z (%) 444.0 (100) [M]$^+$, 429.0 (35) [M-CH$_3$]$^+$; calcd. for C$_{27}$H$_{24}$O$_2$S$_2$ 443.1145; found 443.1148.

Synthesis of (4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(2,1-phenylene) bis(phosphate) (5). Compound **5** was prepared from **2** (65 mg, 0.15 mmol), pyridine (0.24 ml, 2.92 mmol) and phosphorous oxychloride (0.13 ml, 1.46 mmol) following the general procedure **B** for phosphate ester formation. Automated reversed phase flash column chromatography (MeOH/water with 0.5% NEt$_3$, 20-100% MeOH) and subsequent lyophilization afforded 66 mg (0.09 mmol, 58%) of **5** as purple foam. ^{1}**H-NMR:** δ_H**(400 MHz, MeOD):** 1.21 (18H, t, J = 7.3 Hz, 2 H$^+$N(CH$_2$CH_3)$_3$), 1.99 (6H, s, 2 thiophene-CH_3), 2.04 − 2.13 (2H, m, cyclopentene-CH_2), 2.87 (4H, t, J = 7.5 Hz, 2 cyclopentene-CH_2), 3.03 (12H, q, J = 7.3 Hz, 2 H$^+$N(CH_2CH$_3$)$_3$), 6.99 (2H, t, J = 7.6 Hz, 2 phenyl-H), 7.10 − 7.18 (2H, m, 2 phenyl-H), 7.40 (2H, s, thiophene-H), 7.51 (2H, d, J = 7.8 Hz, 2 phenyl-H), 7.63 (2H, d, J = 8.3 Hz, 2 phenyl-H); ^{13}**C-NMR:** δ_C**(101 MHz, MeOD):** 9.1 (+), 14.4 (+), 24.0 (−), 39.6 (−), 47.5 (−), 121.4 (+), 123.8 (+), 126.6 (q), 128.2 (+), 128.4 (+), 128.7 (+), 135.8 (q), 136.1 (q), 136.5 (q), 137.4 (q), 150.8 (q); ^{31}**P-NMR:** δ_P**(162 MHz, MeOD):** -3.40 (s); **ESI-MS:** m/z (%) 806.0 (100) [(M$^+$+3H)+2(N(Et)$_3$)]$^-$, 603.0 (44) [M$^+$+3H]$^-$; calcd. for C$_{27}$H$_{25}$O$_8$P$_2$S$_2$ 603.0472; found 603.0477.

Synthesis of (4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(3,1-phenylene) bis(phosphate) (6). Compound **6** was prepared from **3** (26 mg, 0.06 mmol), pyridine (100 µL, 1.18 mmol) and phosphorous oxychloride (50 µL, 0.56 mmol) following the general procedure **B** for phosphate ester formation. Automated reversed phase flash column chromatography (MeOH/water with 0.5% NEt$_3$, 20-100% MeOH) and subsequent lyophilization afforded 33 mg (0.03 mmol, 56%) of **6** as purple foam. ^{1}**H-NMR:** δ_H**(400 MHz, MeOD):** 1.21 (36H, t, J = 7.3 Hz, 4 H$^+$N(CH$_2$CH_3)$_3$), 1.96 (6H, s, 2 thiophene-CH_3), 2.11 (2H, dt, J = 7.5, 7.6 Hz, cyclopentene-CH_2), 2.86 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 2.97 (24H, q, J = 7.3 Hz, 4 H$^+$N(CH_2CH$_3$)$_3$), 7.11 − 7.13 (4H, m, 4 phenyl-H),

7.15 (2H, s, 2 thiophene-*H*), 7.17 − 7.23 (2H, m, 2 phenyl-*H*), 7.50 (2H, s, 2 phenyl-*H*); **^{13}C-NMR: δ$_C$(101 MHz, MeOD):** 9.5 (+), 14.5 (+), 24.1 (−), 39.5 (−), 47.1 (−), 118.1 (+), 120.0 (+), 120.3 (+), 125.3 (+), 130.5 (+), 135.3 (q), 136.1 (q), 136.7 (q), 138.2 (q), 141.1 (q), 156.4 (q); **^{31}P-NMR: δ$_P$(162 MHz, MeOD):** -0.78 (s). **ESI-MS:** m/z (%) 806.0 (100) [(M^{4-}+3H)+2(N(Et)$_3$)]$^-$, 603.0 (21) [M^{4-}+3H]$^-$; calcd. for $C_{27}H_{25}O_8P_2S_2$ 603.0472; found 603.0478.

Synthesis of (4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(4,1-phenylene) bis(phosphate) (7). Compound **7** was prepared from **4** (50 mg, 0.11 mmol), pyridine (191 μL, 2.36 mmol) and phosphorous oxychloride (103 μL, 1.12 mmol) following the general procedure **B** for phosphate ester formation. Automated reversed phase flash column chromatography (MeOH/water with 0.5% NEt$_3$, 20-100% MeOH) and subsequent lyophilization afforded 73 mg (0.07 mmol, 64%) of **7** as purple oil. **^{1}H-NMR: δ$_H$(400 MHz, MeOD):** 1.26 (36H, t, *J* = 7.3 Hz, 4 H$^+$N(CH$_2$C*H$_3$*)$_3$), 1.99 (6H, s, 2 thiophene-C*H$_3$*), 2.08 (2H, dt, *J* = 7.5, 7.6 Hz, cyclopentene-C*H$_2$*), 2.84 (4H, t, *J* = 7.4 Hz, 2 cyclopentene-C*H$_2$*), 3.09 (24H, q, *J* = 7.3 Hz, 4 H$^+$N(C*H$_2$*CH$_3$)$_3$), 6.97 (2H, s, 2 thiophene-*H*), 7.20 (4H, d, *J* = 8.7 Hz, 4 phenyl-*H*), 7.40 (4H, d, *J* = 8.6 Hz, 4 phenyl-*H*); **^{13}C-NMR: δ$_C$(101 MHz, MeOD):** 9.2 (+), 14.5 (+), 24.1 (−), 39.3 (−), 47.4 (−), 121.8 (+), 124.6 (+), 127.1 (+), 130.6 (q), 134.8 (q), 136.2 (q), 138.1 (q), 140.9 (q), 154.5 (q); **^{31}P-NMR: δ$_P$(162 MHz, MeOD):** -2.26 (s). **ESI-MS:** m/z (%) 806.0 (100) [(M^{4-}+3H)+2(N(Et)$_3$)]$^-$, 603.0 (15) [M^{4-}+3H]$^-$; calcd. for $C_{27}H_{25}O_8P_2S_2$ 603.0472; found 603.0465.

Synthesis of tetraethyl((4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(3,1-phenylene)) bis(phosphonate) (10). A crimp top vial was loaded with compound **8**[2] (277 mg, 0.49 mmol), palladium acetate (11 mg, 0.05 mmol), triphenylphosphine (39 mg, 0.15 mmol) and a stirring bar and capped. It was set under nitrogen atmosphere at which point diethyl phosphite (0.15 ml, 1.18 mmol), triethylamine (0.20 ml, 1.47 mmol) and ethanol (4 ml) were added via syringes and the purple solution was refluxed for 48 h. After cooling to r.t. the reaction mixture was diluted with ethanol (5 ml), transferred to a round bottom flask and the volatiles were removed at the rotary evaporator. The brown residue was purified by automated flash column chromatography (PE/EtOAc, 40-95% EtOAc) yielding **10** (128 mg, 0.19 mmol, 39%) as purple oil. **^{1}H-NMR: δ$_H$(400 MHz, CDCl$_3$):** 1.29 (12H, t, *J* = 7.1 Hz, 2 P(O)(OCH$_2$C*H$_3$*)$_2$), 1.96 (6H, s, 2 thiophene-C*H$_3$*), 1.99 − 2.13 (2H, m, cyclopentene-C*H$_2$*), 2.82 (4H, t, *J* = 7.4 Hz, 2 cyclopentene-C*H$_2$*), 3.91 − 4.24 (8H, m, 2 P(O)(OC*H$_2$*CH$_3$)$_2$), 7.10 (2H, s, 2 thiophene-*H*), 7.39 (4H, td, *J* = 7.7, 4.4 Hz, 2 phenyl-*H*), 7.54 − 7.71 (4H, m, 4 phenyl-H), 7.93 (2H, dt, *J* = 14.1, 1.4 Hz, 2 phenyl-*H*); **^{13}C-NMR: δ$_C$(101 MHz, CDCl$_3$):** 14.5 (+), 16.3 (+), 23.0 (−), 38.6 (−), 62.2 (−), 124.8 (+), 128.2 (+), 128.3 (q), 129.0 (+), 129.1 (+), 130.0 (+), 134.7 (q), 134.9 (q), 135.4 (q), 136.9 (q), 138.4 (q); **^{31}P-NMR: δ$_C$(162 MHz, CDCl$_3$):** 19.0 (s); **ESI-MS:** m/z (%) 685.2 (100) [M+H]$^+$, 1391.4 (65) [2M+Na]$^+$; calcd. for $C_{35}H_{42}O_6P_2S_2$ 685.1971; found 685.1976.

S5

Synthesis of tetraethyl ((4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(4,1-phenylene)) bis(phosphonate) (11). A crimp top vial was loaded with compound **9**[3] (248 mg, 0.44 mmol), palladium acetate (10 mg, 0.04 mmol), triphenylphosphine (34 mg, 0.14 mmol) and a stirring bar and capped. It was set under nitrogen atmosphere at which point diethyl phosphite (0.13 ml, 1.04 mmol), triethylamine (0.18 ml, 1.31 mmol) and ethanol (4 ml) were added via syringes and the purple solution was refluxed for 48 h. After cooling to r.t. the reaction mixture was diluted with ethanol (5 ml), transferred to a round bottom flask and the volatiles were removed at the rotary evaporator. The brown residue was purified by automated flash column chromatography (PE/EtOAc, 95% EtOAc) yielding **11** (102 mg, 0.15 mmol, 34%) as purple oil. **^{1}H-NMR: δ_H(400 MHz, CDCl$_3$):** 1.31 (12H, t, J = 7.1 Hz, 2 P(O)(OCH$_2$CH_3)$_2$), 1.99 (6H, s, 2 thiophene-CH_3), 2.02 − 2.14 (2H, m, cyclopentene-CH_2), 2.83 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 3.92 − 4.29 (8H, m, 2 P(O)(OCH_2CH$_3$)$_2$), 7.11 (2H, s, 2 thiophene-H), 7.55 (4H, dd, J = 8.3, 3.7 Hz, 4 phenyl-H), 7.75 (4H, dd, J = 12.9, 8.3 Hz, 4 phenyl-H); **^{13}C-NMR: δ_C(101 MHz, CDCl$_3$):** 14.5 (+), 16.3 (+), 16.4 (+), 23.0 (−), 38.5 (−), 62.1 (−), 124.9 (+), 125.1 (+), 125.4 (+), 127.2 (q), 132.4 (+), 132.5 (+), 134.8 (q), 136.2 (q), 137.0 (q), 138.2 (q), 138.4 (q); **^{31}P-NMR: δ_C(162 MHz, CDCl$_3$):** 19.3 (s); **ESI-MS:** m/z (%) 685.2 (100) [M+H]$^+$, 1391.4 (44) [2M+Na]$^+$; calcd. for C$_{35}$H$_{42}$O$_6$P$_2$S$_2$ 685.1971; found 685.1977.

Synthesis of ((4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(3,1-phenylene))-diphosphonic acid (12). Compound **10** (118 mg, 0.17 mmol) was dissolved in dry acetonitrile (15 ml) and treated with trimethylsilyl bromide (1.14 ml, 8.62 mmol). The reaction mixture was stirred at r.t. overnight. Volatiles were removed at the rotary evaporator. After purification by automated reversed phase flash column chromatography (MeOH/water with 0.5% NEt$_3$, 3-100% MeOH) and lyophilization compound **12** (114 mg, 0.15 mmol, 87%) was obtained as dark purple oil. **^{1}H-NMR: δ_H(400 MHz, MeOD):** 1.23 (18H, t, J = 7.3 Hz, 2 H$^+$N(CH$_2$CH_3)$_3$), 1.98 (6H, s, 2 thiophene-CH_3), 2.05 − 2.16 (2H, m, cyclopentene-CH_2), 2.86 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 3.07 (12H, q, J = 7.3 Hz, 2 H$^+$N(CH_2CH$_3$)$_3$), 7.16 (2H, s, thiophene-H), 7.33 (2H, td, J = 7.7, 3.5 Hz, 2 phenyl-H), 7.44 − 7.59 (2H, m, 2 phenyl-H), 7.67 (2H, dd, J = 12.1, 7.5 Hz, 2 phenyl-H), 8.01 (2H, d, J = 13.1 Hz, 2 phenyl-H); **^{13}C-NMR: δ_C(101 MHz, MeOD):** 9.2 (+), 14.6 (+), 24.1 (−), 39.5 (−), 47.4 (−), 125.4 (+), 127.4 (+), 128.7 (+), 129.5 (+), 130.6 (+), 135.2 (q), 135.7 (q), 136.2 (q), 138.3 (q), 139.2 (q), 141.0 (q); **^{31}P-NMR: δ_P(162 MHz, MeOD):** 12.5 (s); **ESI-MS:** m/z (%) 571.1 (100) [M^4+3H]$^-$; calcd. for C$_{27}$H$_{25}$O$_6$P$_2$S$_2$ 571.0573; found 571.0570.

Synthesis of ((4,4'-(cyclopent-1-ene-1,2-diyl)bis(5-methylthiophene-4,2-diyl))bis(4,1-phenylene))-diphosphonic acid (13). Compound **11** (82 mg, 0.12 mmol) was dissolved in dry acetonitrile (5 ml) and treated with trimethylsilyl bromide (0.79 ml, 6.00 mmol). The reaction mixture was stirred at r.t. overnight. Volatiles were removed at the rotary evaporator. After purification by automated reversed phase flash column chromatography (MeOH/water with 0.5% NEt$_3$, 3-100% MeOH) and

lyophilization compound **13** (76 mg, 0.10 mmol, 82%) was obtained as dark purple oil. **^{1}H-NMR:** **δ_H(400 MHz, MeOD):** 1.25 (18H, t, J = 7.3 Hz, 2 H$^+$N(CH$_2$CH_3)$_3$), 1.99 (6H, s, 2 thiophene-CH_3), 2.09 (2H, dt, J = 7.5, 7.6 Hz, cyclopentene-CH_2), 2.85 (4H, t, J = 7.4 Hz, 2 cyclopentene-CH_2), 3.09 (12H, q, J = 7.3 Hz, 2 H$^+$N(CH_2CH$_3$)$_3$), 7.17 (2H, s, thiophene-H), 7.50 (4H, dd, J = 8.4, 2.8 Hz, 4 phenyl-H), 7.73 (4H, dd, J = 12.1, 8.3 Hz, 4 phenyl-H); **^{13}C-NMR:** **δ_C(101 MHz, MeOD):** 9.2 (+), 14.7 (+), 24.1 (−), 39.4 (−), 47.4 (−), 125.4 (+), 125.8 (q), 132.6 (+), 136.2 (q), 136.7 (q), 136.9 (q), 138.4 (q), 138.5 (q), 140.7 (q); **^{31}P-NMR:** **δ_P(162 MHz, MeOD):** 12.9 (s); **ESI-MS:** m/z (%) 571.1 (100) [M^{4+}+3H]$^-$; calcd. for C$_{27}$H$_{25}$O$_6$P$_2$S$_2$ 571.0573; found 571.0579.

Photochromism of DTE-phosphates and DTE-phosphonates

Photochemical syntheses of ring-closed isomers. Solutions of compounds **5-7, 12** and **13** (12.5 µM) in 50 mM Tris/acetate pH 8.5 were irradiated for 30 s with a 312 nm lamp yielding pink solutions containing the ring-closed isomers. The changes in the UV/Vis absorption spectra are representatively shown for compound **6** in Figure S1. The photostationary states contain between 93 and 97% of the ring-closed isomers according to HPLC analyses. The respective HPLC chromatograms of compound **6** in its open and closed form are depicted in Figure S2.

Photochemical cycling studies. In order to test the robustness of the photochromic systems, photochemical cycling studies were recorded for all switches (Figure S3). In each case, a solution of 12.5 µM inhibitor in 50 mM Tris/acetate pH 8.5 was alternately irradiated with 312 nm light for 30 s and with visible light for 15 min over various cycles and the absorption change at 525 nm was monitored.

Cloning, heterologous expression in Escherichia coli, and purification of mtPriA

The gene coding for *mt*PriA was amplified by PCR using pETM-11-*mt*PriA[4] as a template and cloned into the vector pET24a(+) (Stratagene) via the terminal restriction sites for *Nde*I and *Xho*I. *mt*PriA was heterologously expressed in *E. coli* BL21-Gold cells (Stratagene) transformed with pET(24a)(+)-mt*pri*A. To this end, five liters of Luria Broth (LB) medium supplemented with 75 µg/ml kanamycin were inoculated with a pre-culture and incubated at 37 °C. After an OD$_{600}$ of 0.6 was reached, the temperature was lowered to 20 °C. Expression was induced by adding 0.5 mM IPTG, and growth was continued overnight. Cells were harvested by centrifugation (Sorvall/RC5B, GS3, 15 min, 4000 rpm, 4 °C) and suspended in 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 10 mM imidazole. After lysing by sonication (Branson Sonifier W-250D, 2 x 2 min in 15 sec intervals, 45% pulse, 0 °C), cells were centrifuged again (Sorvall/RC5B, SS34, 30 min, 13.000 rpm, 4 °C) to separate the soluble from the insoluble fraction of the cell extract. As a first purification step, *mt*PriA was subjected to a Ni^{2+}

affinity chromatography using its C-terminal hexa-histidine tag. To this end, the soluble supernatant was loaded onto a HisTrapFF crude column (5 ml; GE Healthcare), which had been equilibrated with 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 10 mM imidazole. After washing the column with the equilibration buffer, a linear gradient of 10-500 mM imidazole was applied and *mt*PriA eluted between 40 and 100 mM imidazole. Main fractions were pooled and further purified via size exclusion chromatography. For this purpose, a Superdex200 column (HiLoad 26/60, 320 ml, GE Healthcare) was equilibrated and operated with 50 mM Tris/HCl pH 8.0, 200 mM sodium chloride, 2 mM dithiothreitol at 4 °C. Fractions with pure protein were pooled and dialyzed twice against 50 mM Tris/HCl pH 8.0. According to SDS-PAGE (12.5% acrylamide), *mt*PriA was more than 95% pure. About 1 mg of protein was obtained per liter of culture.

Steady-state enzyme kinetics of mtPriA

Steady-state enzyme kinetics of the *mt*PriA reaction were measured spectrophotometrically at 300 nm as described.[5] In a coupled enzyme assay the substrate ProFAR was initially converted to PRFAR, which was further processed by the enzyme HisF from *Thermotoga maritima* (*tm*HisF) to yield imidazole glycerol phosphate (ImGP) and aminoimidazole carboxamide ribonucleotide (AICAR) (Scheme S1). As ProFAR and PRFAR exhibit identical absorption spectra, the second step is indispensible in order to monitor the reaction progress at 300 nm ($\Delta\varepsilon_{300}$ (ProFAR-AICAR) = 5.637 mM^{-1}cm^{-1}). It furthermore prevents product inhibition of the *mt*PriA reaction.[6]
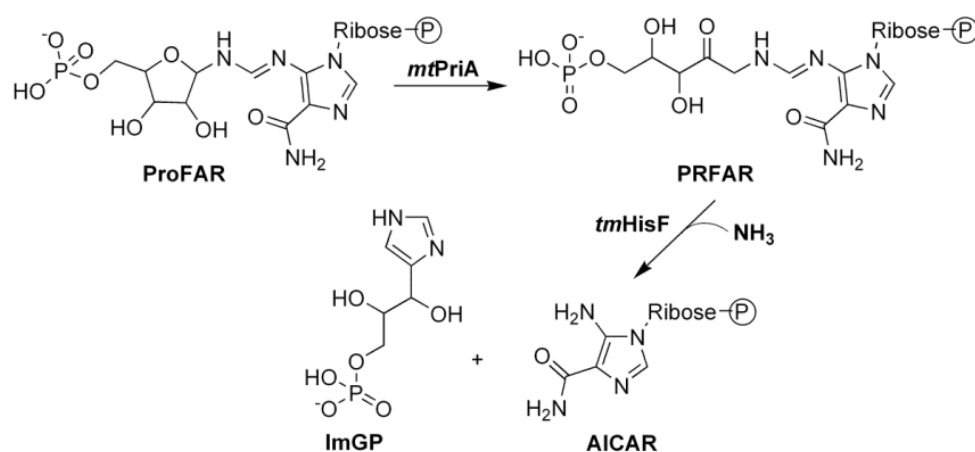
Scheme S1. Coupled enzyme assay for photometric determination of *mt*PriA activity.

The steady-state catalytic parameters of *mt*PriA in the absence of inhibitor were determined in triplicate at 25 °C in 50 mM Tris/acetate pH 8.5 in the presence of 100 mM ammonium acetate, 2-73 µM ProFAR, and 0.2 µM *tm*HisF (see Table S1). Reactions were initiated with 0.2 µM *mt*PriA,

when the baseline signal was constant. To determine the inhibition constants of compounds **5-7**, **12** and **13** in their open and closed forms, three saturation curves were measured in presence of different inhibitor concentrations in each case. Reaction mixtures contained 100 mM ammonium acetate, 3-190 μM ProFAR, 0.4-20 μM inhibitor (closed inhibitor forms were synthesized photochemically prior to addition to the reaction mixture), and 0.04-0.2 μM *tm*HisF in 50 mM Tris/acetate pH 8.5. Substrate turnover was initiated with 0.04-0.2 μM *mt*PriA. The molar concentration of *mt*PriA and *tm*HisF did not exceed a tenth of the respective inhibitor concentration. Additionally, to rule out artifacts, the photostability of inhibitor **6** was representatively investigated under assay conditions: 15 μM of **6** in both isomeric forms were irradiated for 30 minutes at 300 nm in the spectrophotometer and no change in the absorption spectra was observed. Moreover, it was assured that the performance of *mt*PriA is rate-limiting in each kinetic setup, as doubling of its concentration resulted in a doubled initial velocity. Finally, the inhibition constant K_i of each DTE compound could be calculated from Formula S1, where K_m is the Michaelis constant of *mt*PriA and K_m^{app} is the apparent K_m value determined in the presence of a distinct inhibitor concentration $c(I)$. All kinetic parameters are given in Table S1 and steady-state enzyme kinetics of **6** in its open and closed form are illustrated in Figure S4.

$$K_i = \frac{K_m \cdot c(I)}{K_m^{app} - K_m}$$

Formula S1. Determination of inhibition constant K_i.

Molecular Modeling

Energy minimization of the open and closed isomers of compounds 5-7, 12 and 13.

The conformer distribution of compounds **5**, **6**, **7**, **12** and **13** in their open and closed isomers was calculated by Wavefunction Spartan'10[7] using molecular mechanics in water (MMFFaq) in the ground state with 20.000 conformers examined, a total charge of -2 for phosphonates or -4 for phosphates, singlet multiplicity, subject to symmetry and global calculations. Sorted by energy minimum, the five best results for each calculation were averaged and are shown in Table S2.

Molecular Dynamics (MD) simulations of *mt*PriA with compound 6 bound in its open and closed form.

Modeling and MD simulations were performed with YASARA Structure Version 13.4.21 employing the YAMBER3 force field.[8] Simulations were run at 298 K under periodic boundary conditions and with explicit water, using a multiple time step of 1 fs for intramolecular and 2 fs for intermolecular forces. Lennard-Jones forces and long-range electrostatic interactions were treated with a 7.86 Å cutoff, the latter were calculated using the Particle Mesh Ewald method.[9] Temperature was adjusted

using a Berendsen thermostat based on the time-averaged temperature and simulations were carried out at constant pressure. The parameterization of the open and closed forms of compound **6** was performed using the AM1BCC protocol[10] that assigns atomic charges by applying additive bond charge corrections (BCCs) to semi-empirical AM1 atomic charge calculations. Based on the crystal structure of *mt*PriA (PDB ID 3ZS4), the original ligand PRFAR was removed and manually replaced by compound **6** in the open and closed form in order to fit the phosphate binding pockets. The simulation cell was defined as 5 Å larger than the protein along each axis (cell dimensions 52 × 47 × 44 Å^3), filled with water to a density of 0.997 g/ml, and counterions were added to a final concentration of 0.9% NaCl. The protonation states of protein side chains were assigned according to Krieger *et al.*[11] In order to remove conformational stress, two phases of equilibration were conducted: After a 100 ps equilibration with frozen protein coordinates, the whole system was equilibrated for 1 ns. The two equilibrated models of the open and closed form were subsequently used for the six following (three for each conformer) production MD simulations. To reassign initial atom velocities and thus seed independent calculations, the temperature was slightly changed (+/-0.0001 K) for the respective second and third simulation runs of the open/closed form. Trajectories were sampled at intervals of 100 ps for a total of 10 ns for each model. Each snapshot was energy minimized as follows: After removing conformational stress by a steepest descent minimization, the procedure continued by simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e. the energy improved by less than 0.05 kJ/mol per atom during 200 steps. Binding energies were obtained for each energy minimized snapshot using YASARA's integrated binding energy function that computes the energetic difference of the ligand at bound state and at infinite distance from its binding site. Representative enzyme models (shown in Figure 3) for each simulation are based on the energy minimized snapshots with the best binding energy. The ligand binding energies and standard deviations (see Table S2) were calculated by using the full production trajectory.

Supplementary Figures
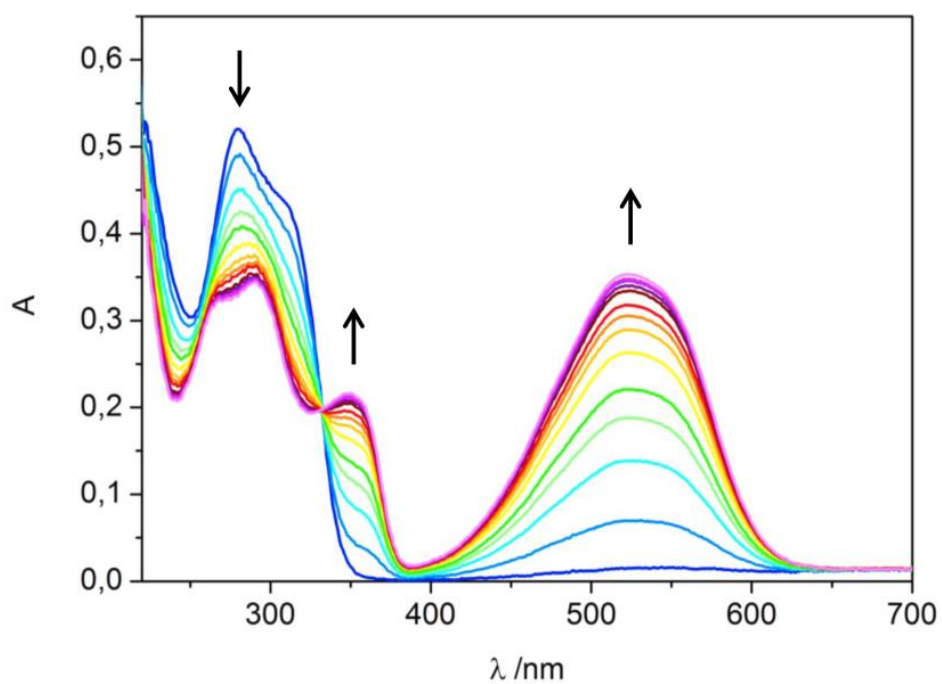
Figure S1. UV/Vis absorption spectra evolution of compound **6** upon irradiation with 312 nm light. Arrows indicate the changes of the absorption maxima of 12.5 μM **6** in 50 mM Tris/acetate pH 8.5 with irradiation intervals of 2 s from 0 to 30 s, which is illustrated by the color change from blue (open isomer) to purple (closed isomer).

S11
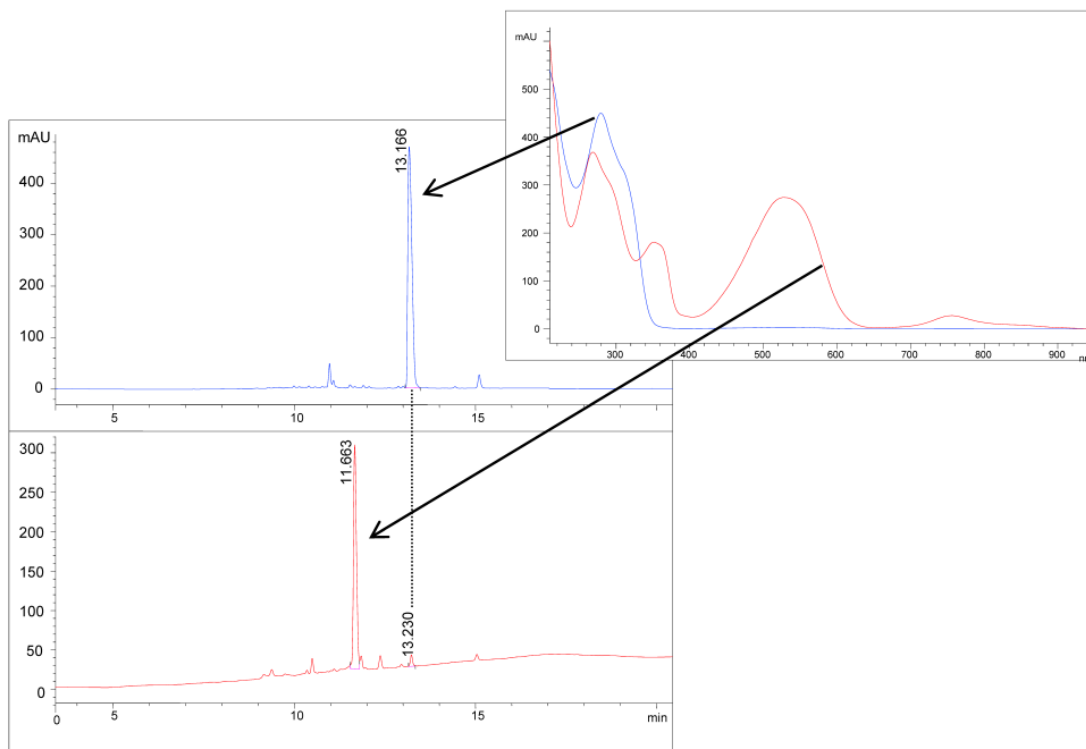
Figure S2. Representative HPLC chromatograms for the determination of the photostationary state of compound **6** (100 μM in 50 mM Tris/acetate pH 8.5). The UV/Vis signal was detected at 288 nm. The chromatograms before (upper) and after (lower) irradiation with 312 nm light for 30 s indicate that 97% of the open isomer (blue curve) were converted to the closed isomer (red curve). The absorption spectra of both forms are shown in the right panel.

S12
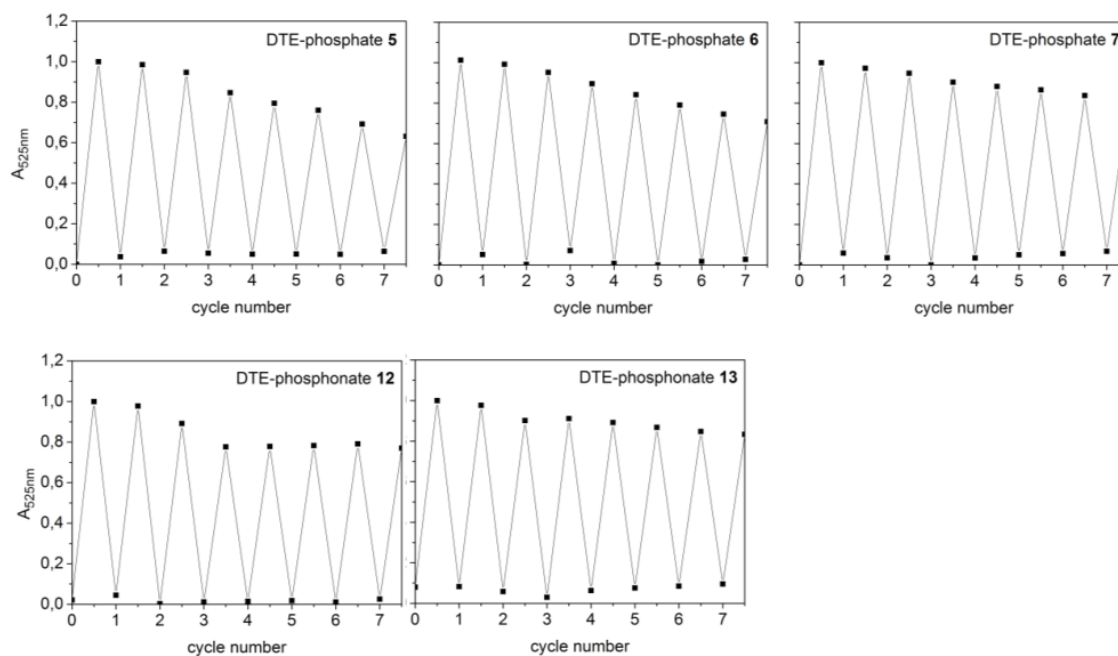
Figure S3. Cycle performance of compounds **5**-**7**, **12** and **13**. Changes in absorption at 525 nm were measured during an alternated irradiation of each inhibitor solution (12.5 µM in 50 mM Tris/acetate pH 8.5) with 312 nm light for 30 s and greater than 420 nm light for 15 min.

S13
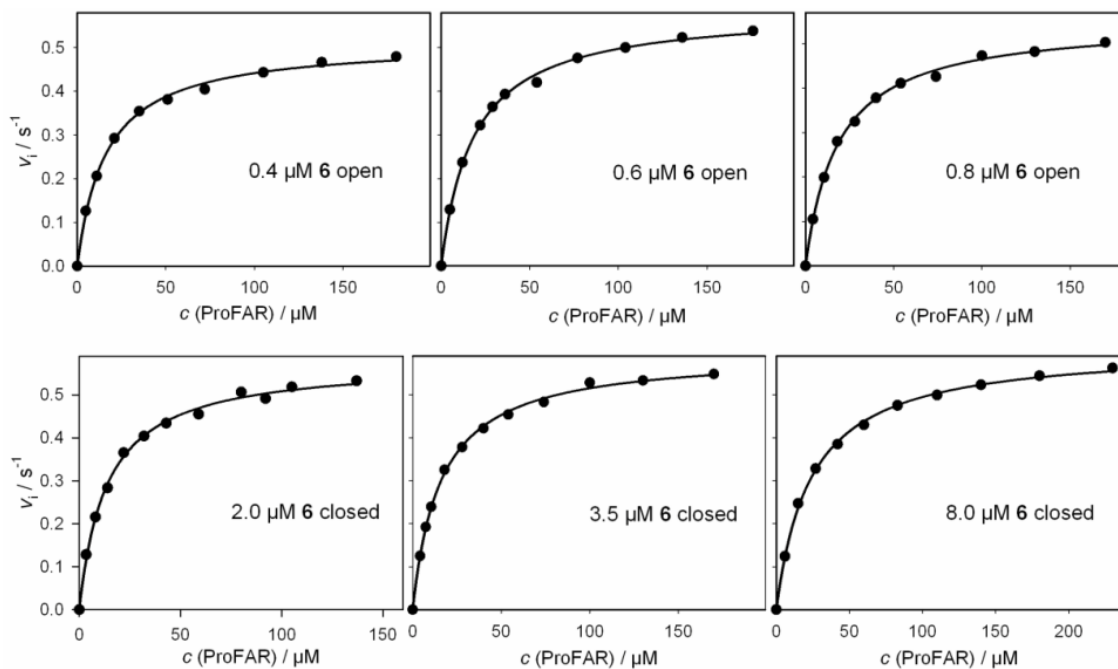
Figure S4. Substrate saturation curves of *mt*PriA in the presence of inhibitor **6** in its open (upper curves) and closed (lower curves) form. See Table S1 for deduced steady-state kinetic parameters.

Supplementary Tables

Table S1. Steady-state kinetic constants for the ProFAR isomerization activity of *mt*PriA in absence and presence of compounds **5-7**, **12** and **13** in their ring-open and ring-closed forms.

Inhibitor	$c(I)$ / µM	k_{cat} / s^{-1}	K_m / µM	K_i / µM
			open / closed	
-		0.49	8.6	
		0.49	9.2	
		0.57	8.1	
5	5.0 / 5.0	0.47 / 0.58	12.7 / 13.9	10.5 / 8.1
	10.5 / 10.5	0.59 / 0.55	20.9 / 21.5	7.3 / 7.0
	18.0 / 18.0	0.55 / 0.60	32.5 / 35.2	6.5 / 5.8
6	0.4 / 2.0	0.52 / 0.56	16.2 / 12.4	0.45 / 4.5
	0.6 / 3.5	0.59 / 0.60	18.7 / 15.9	0.51 / 4.1
	0.8 / 8.0	0.55 / 0.61	18.5 / 23.7	0.69 / 4.6
7	2.0 / 3.5	0.47 / 0.58	13.1 / 16.2	3.8 / 4.0
	3.5 / 7.0	0.47 / 0.55	17.7 / 24.8	3.3 / 3.7
	5.0 / 10.5	0.47 / 0.53	21.7 / 35.6	3.3 / 3.3
12	1.5 / 4.0	0.57 / 0.55	16.3 / 15.2	1.7 / 5.2
	2.5 / 10.0	0.58 / 0.57	22.3 / 27.1	1.6 / 4.6
	4.0 / 15.0	0.60 / 0.55	29.4 / 34.9	1.7 / 4.9
13	5.0 / 15.0	0.60 / 0.58	15.7 / 15.1	6.1 / 19.8
	10.0 / 15.0	0.60 / 0.57	20.7 / 15.0	7.1 / 20.2
	15.0 / 20.0	0.59 / 0.57	26.6 / 14.7	7.2 / 28.2

Three substrate saturation curves were recorded at different inhibitor concentrations $c(I)$ for each form (open/closed). The deduced k_{cat} and K_m values are shown. The shown K_i values were calculated with Formula S1 from the increase of the apparent K_m value with respect to the K_m value in absence of inhibitor. The mean values and standard deviations of K_i for each inhibitor are given in Table 1.

Table S2. Ligand binding energies derived from MD simulations of *mt*PriA and compound **6** in its open and closed form.[a]

	Ligand binding energy (open) [kJ/mol]	Ligand binding energy (closed) [kJ/mol]
run 1	-2171 ± 55	-2062 ± 64
run 2	-2162 ± 59	-1919 ± 71
run 3	-2122 ± 77	-2049 ± 60

[a] Three independent simulations were performed in each case; see Experimental Methods for details.

Supplementary NMR spectra

^{1}H-NMR for compound **2**:
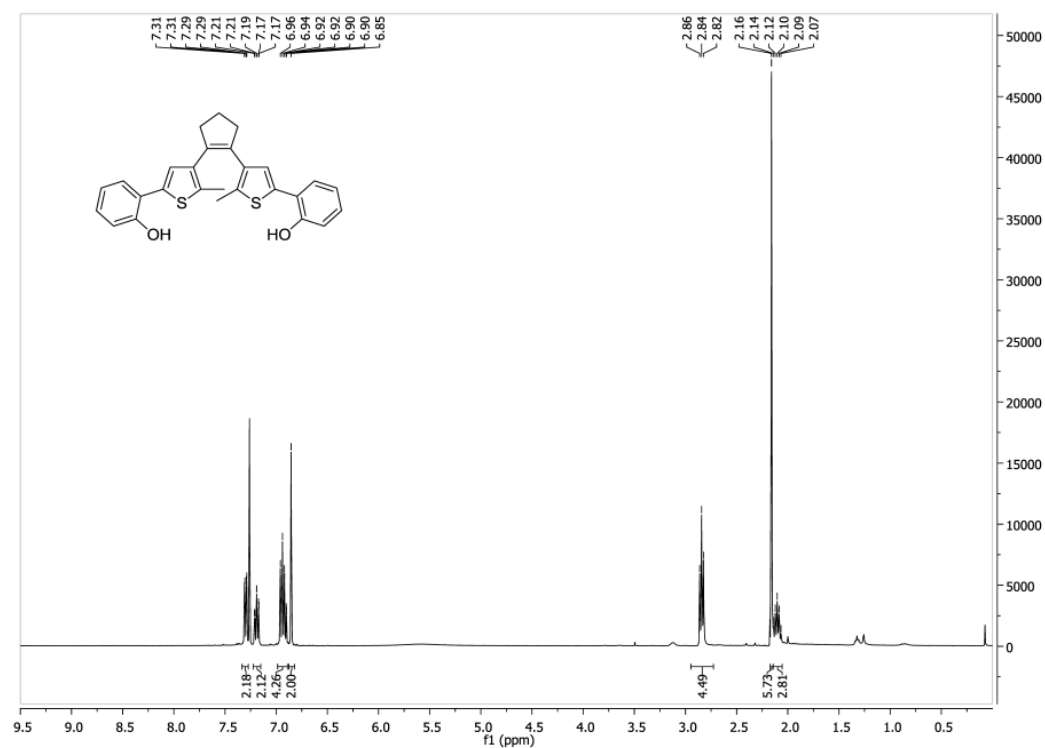
^{13}C-NMR for compound **2**:
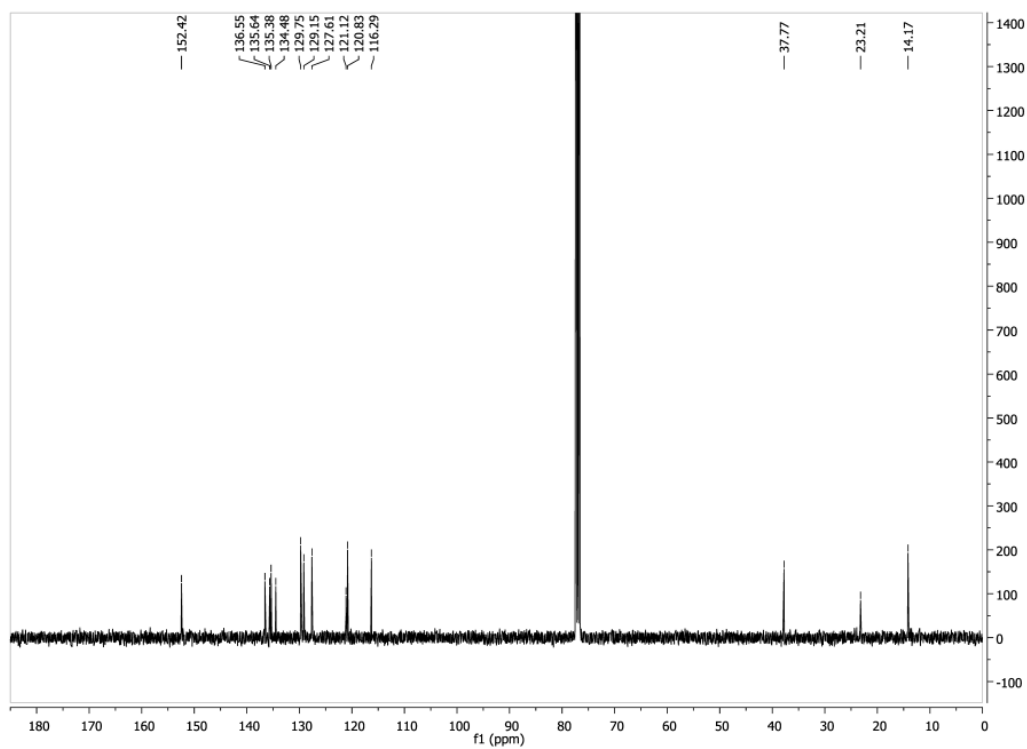
¹H-NMR for compound **3**:
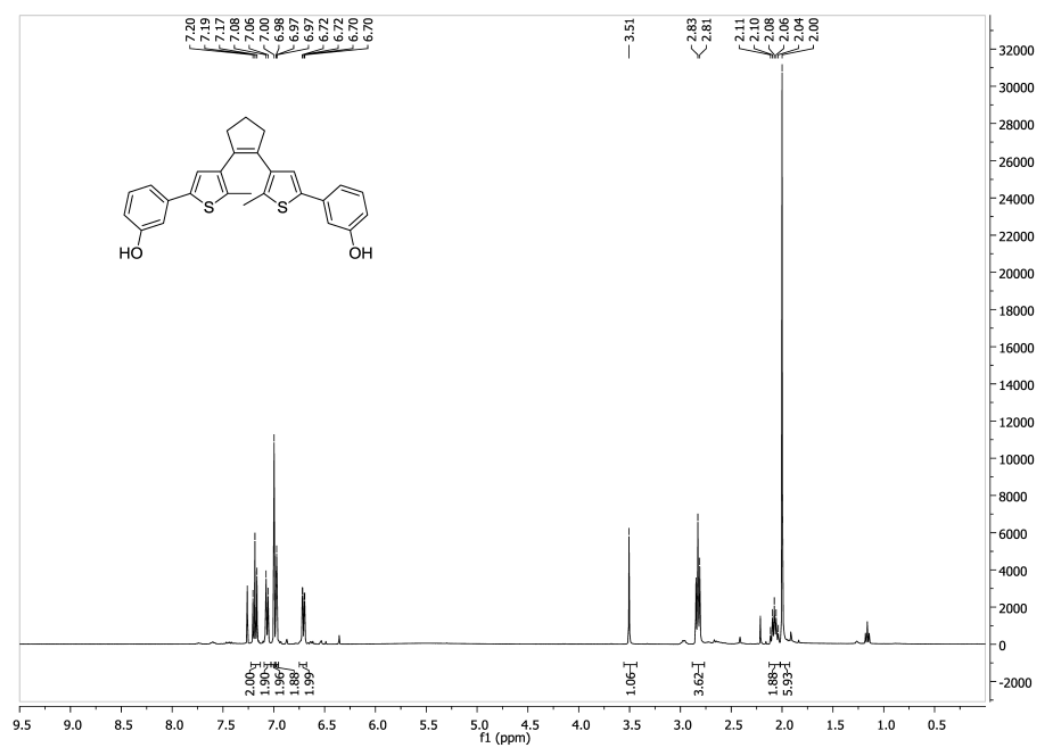
¹³C-NMR for compound **3**:

S17
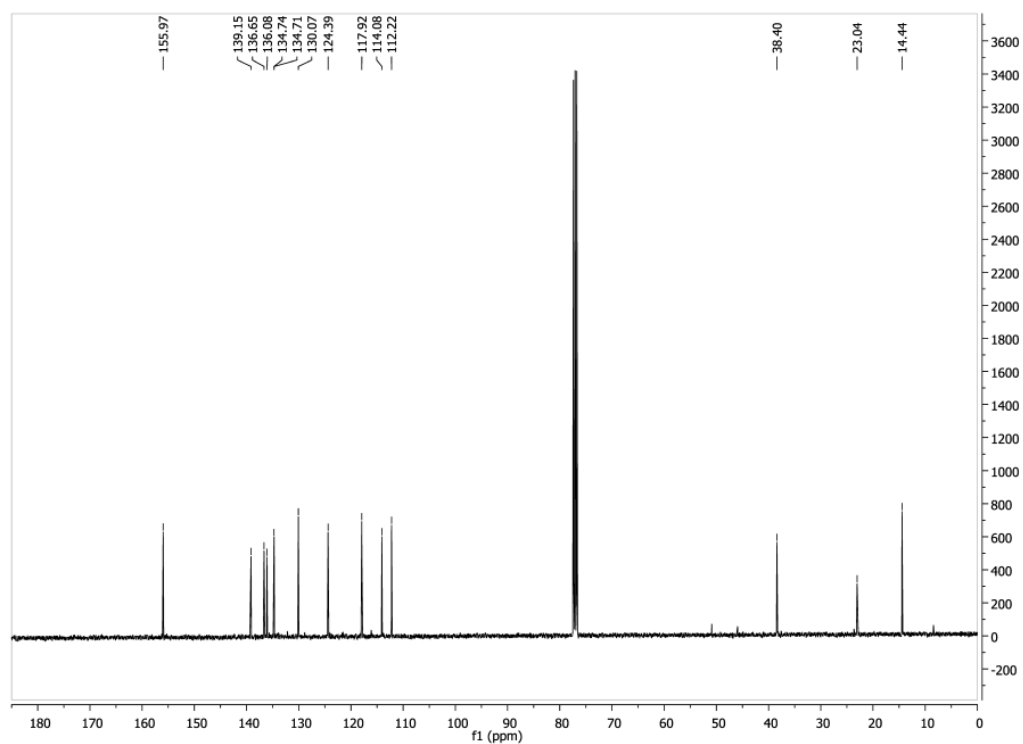
^{1}H-NMR for compound **4**:
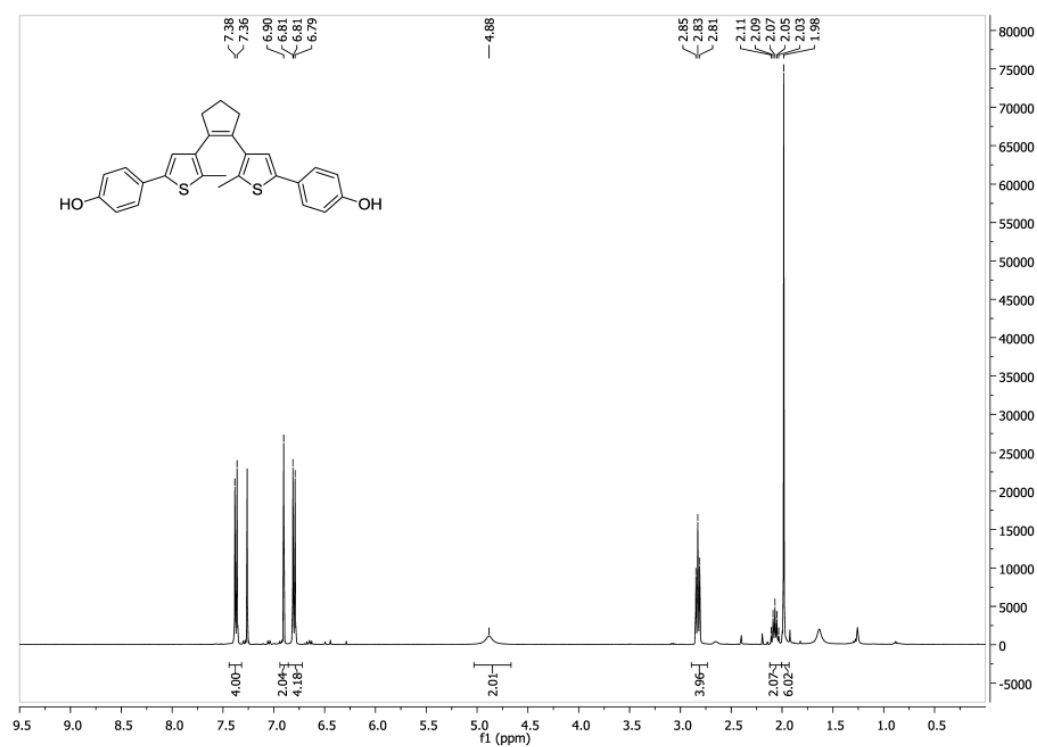
^{13}C-NMR for compound **4**:

S18
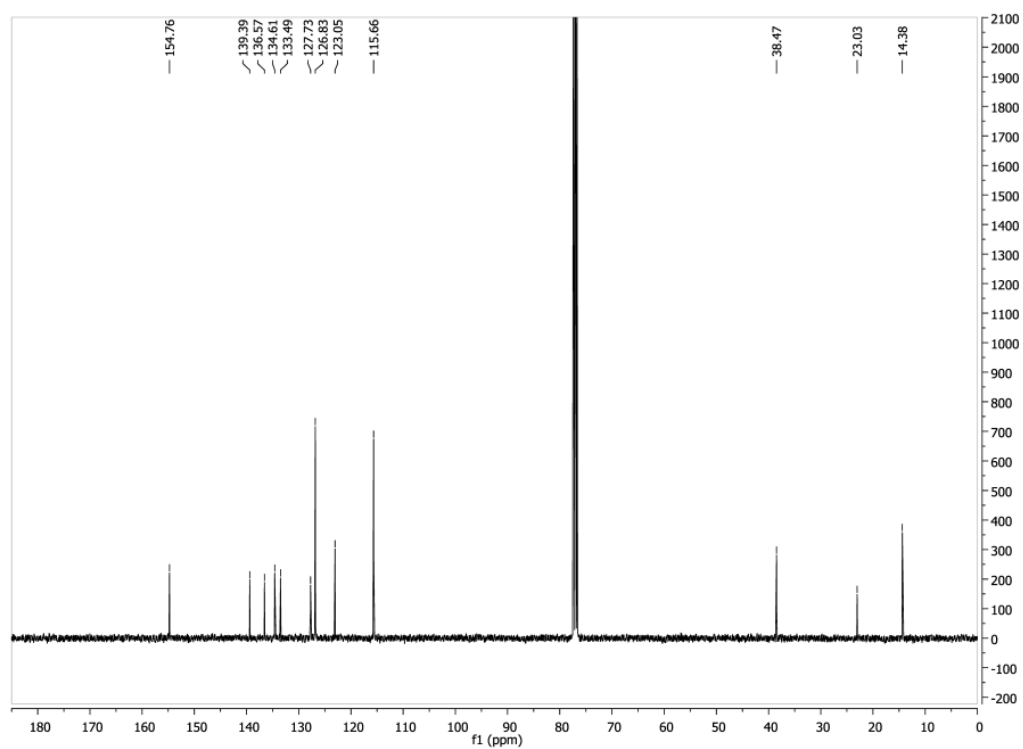
¹H-NMR for compound **5**:
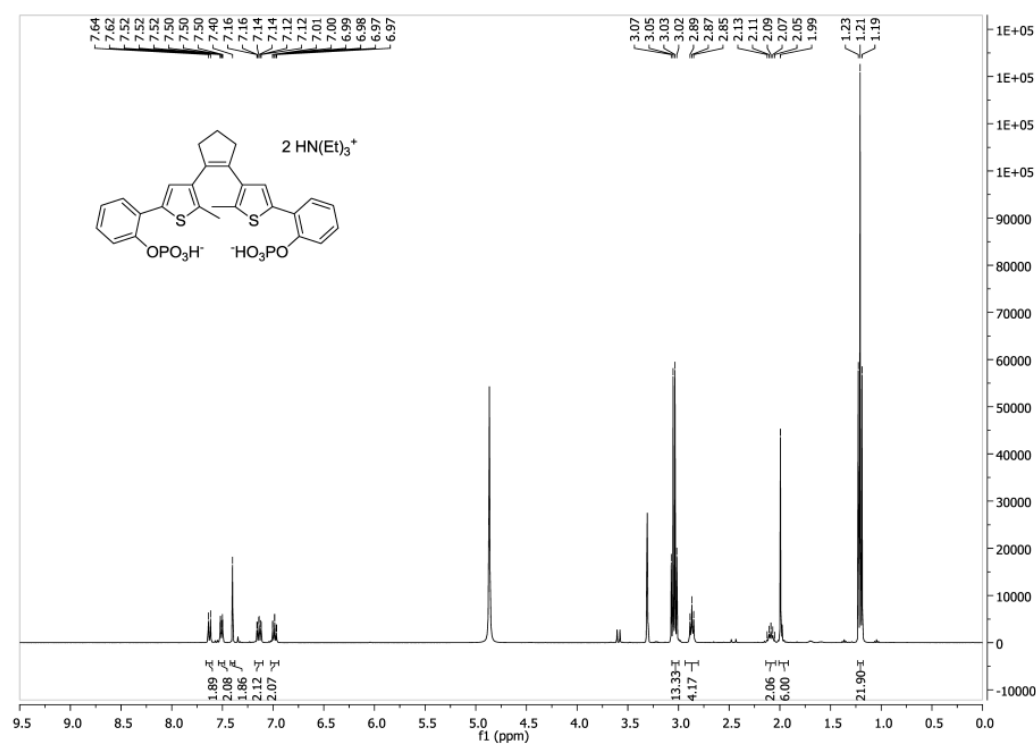
¹³C-NMR for compound **5**:
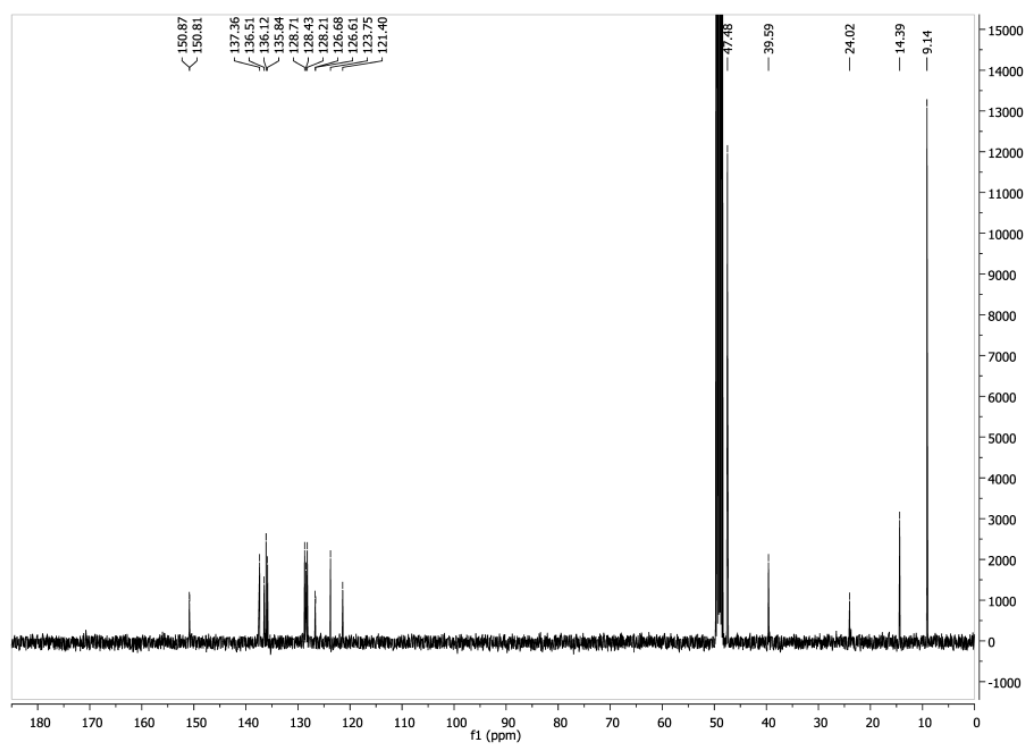
^{1}H-NMR for compound **6**:
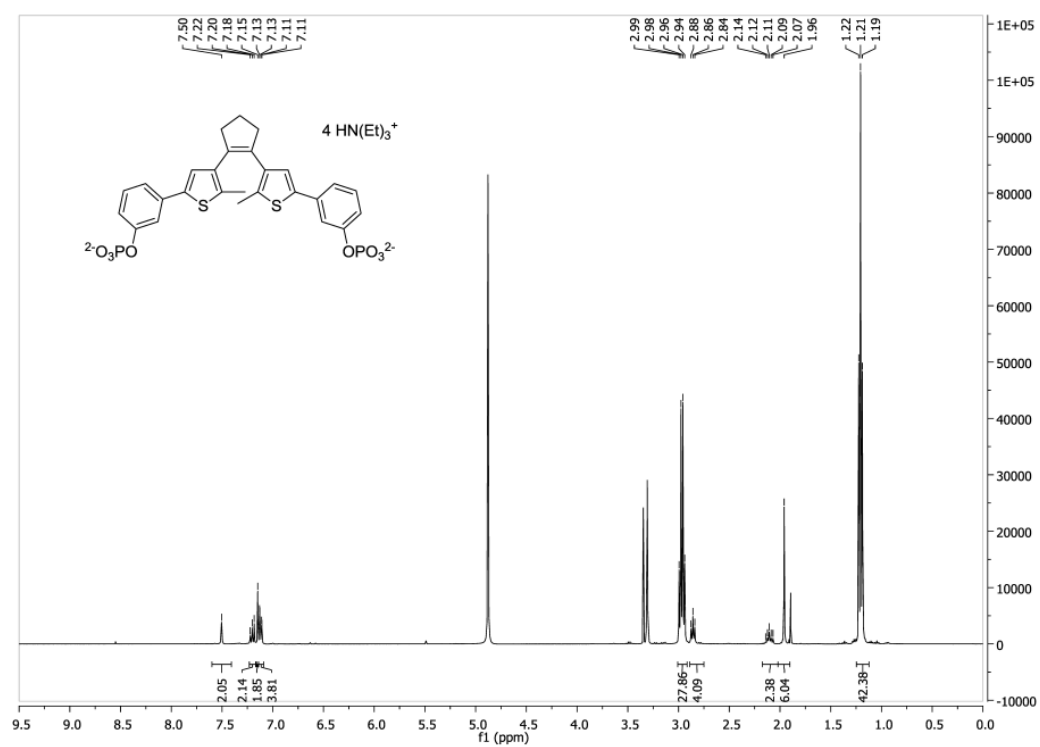
^{13}C-NMR for compound **6**:

S20
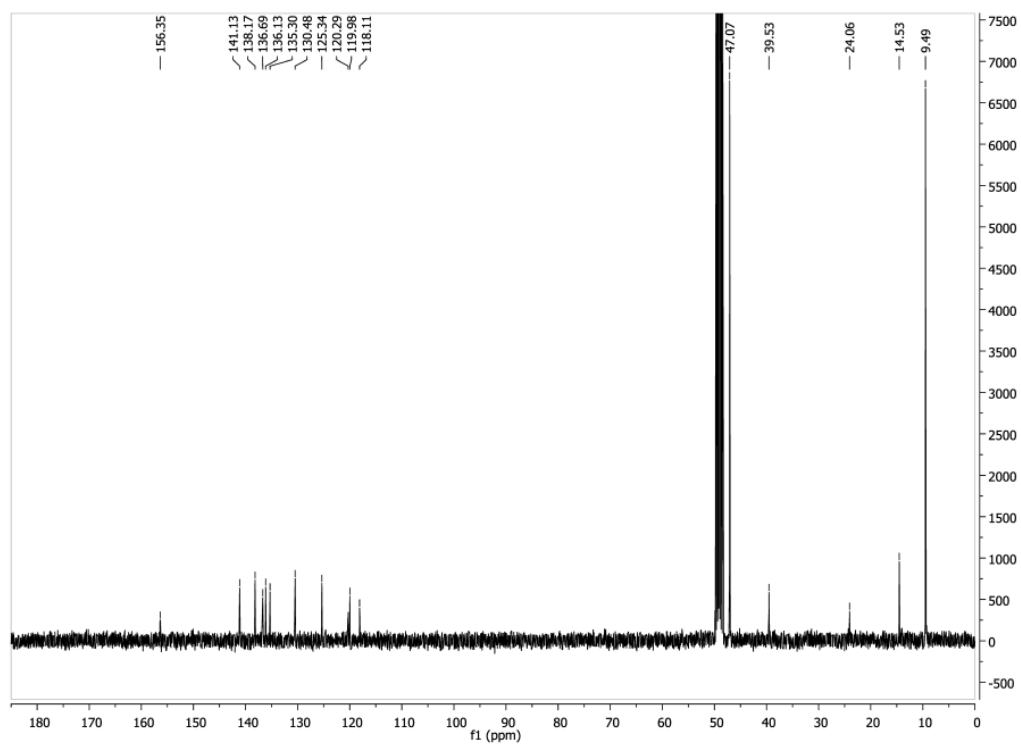
¹H-NMR for compound **7**:
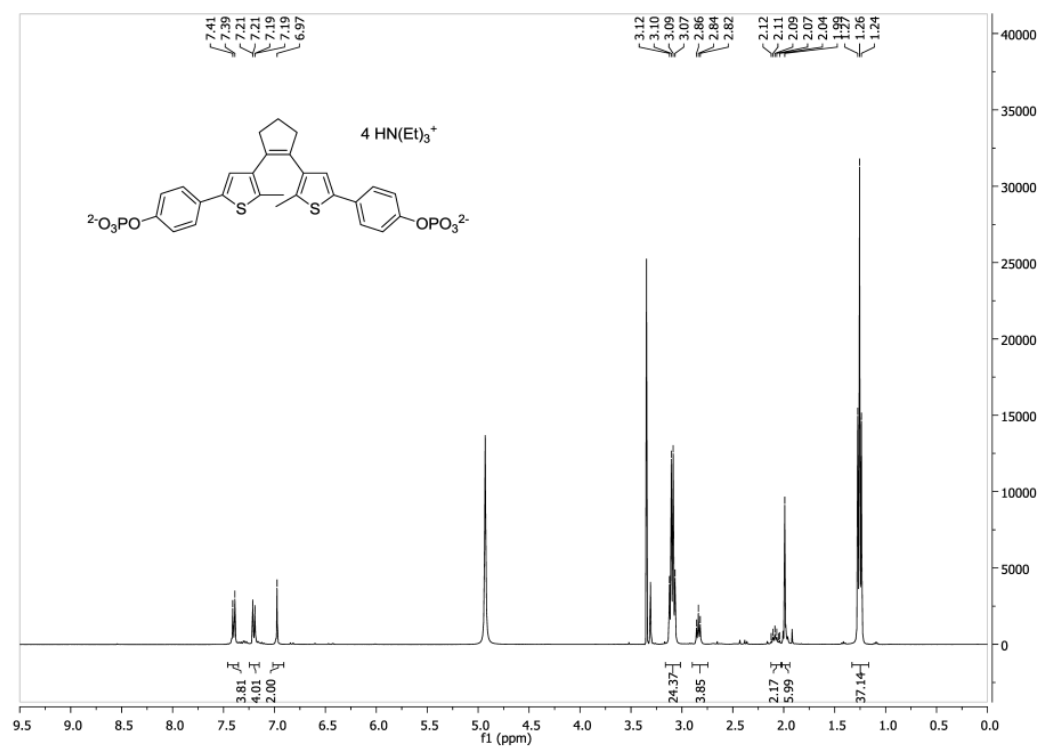
¹³C-NMR for compound **7**:
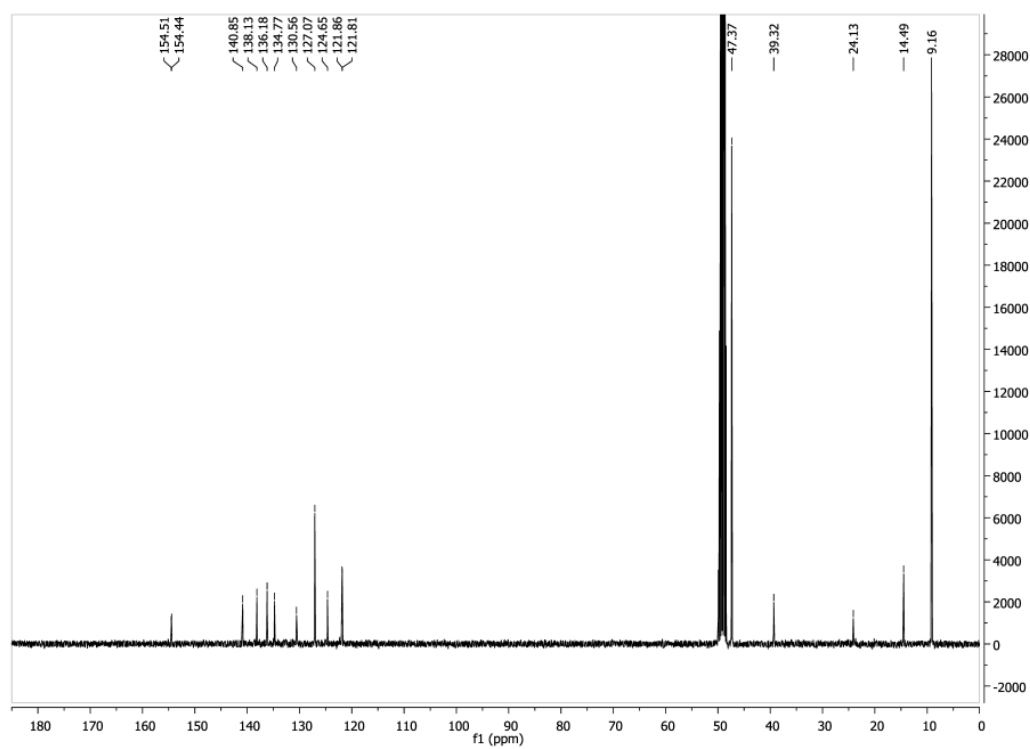
S21

¹H-NMR for compound **10**:
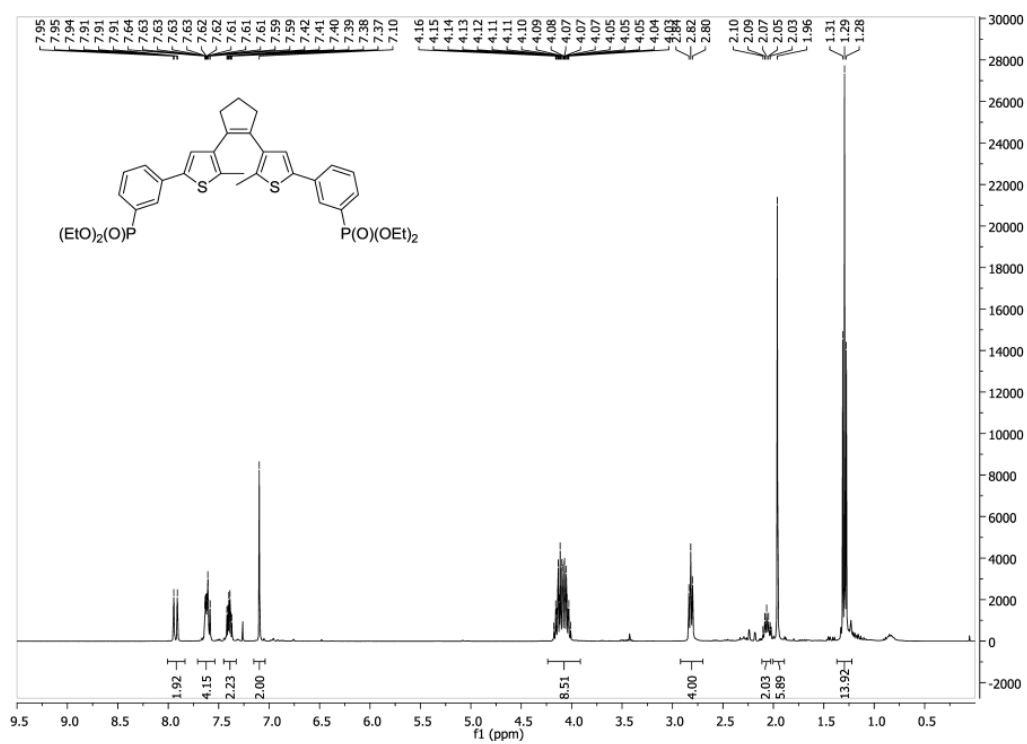
¹³C-NMR for compound **10**:
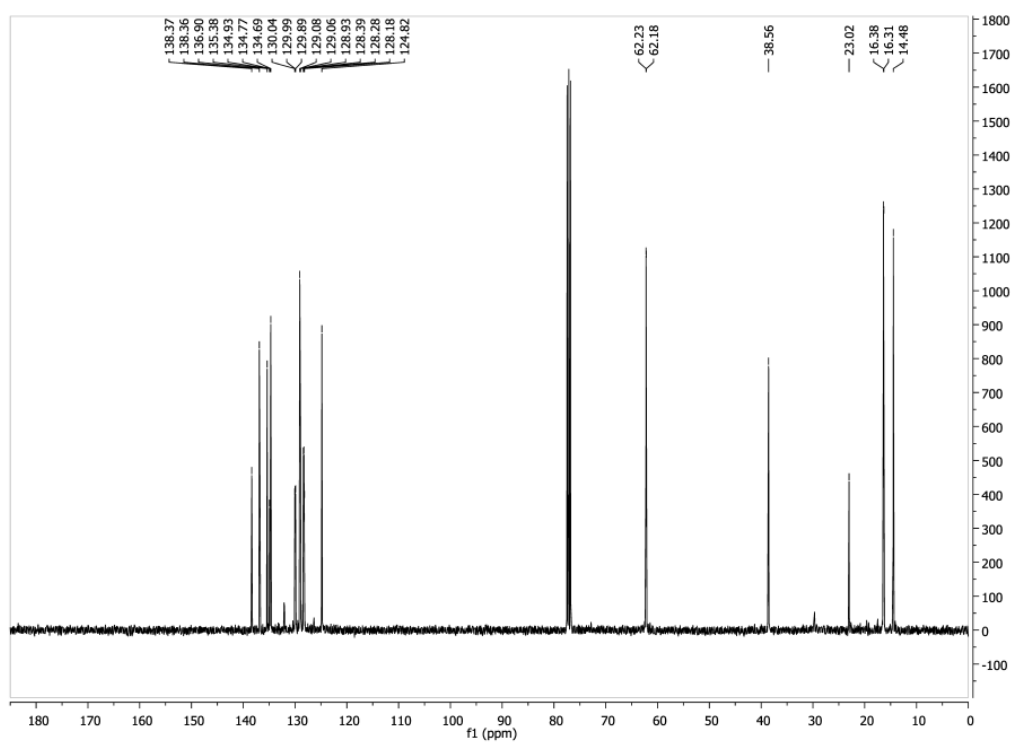
^{1}H-NMR for compound 11:

^{13}C-NMR for compound 11:

S23
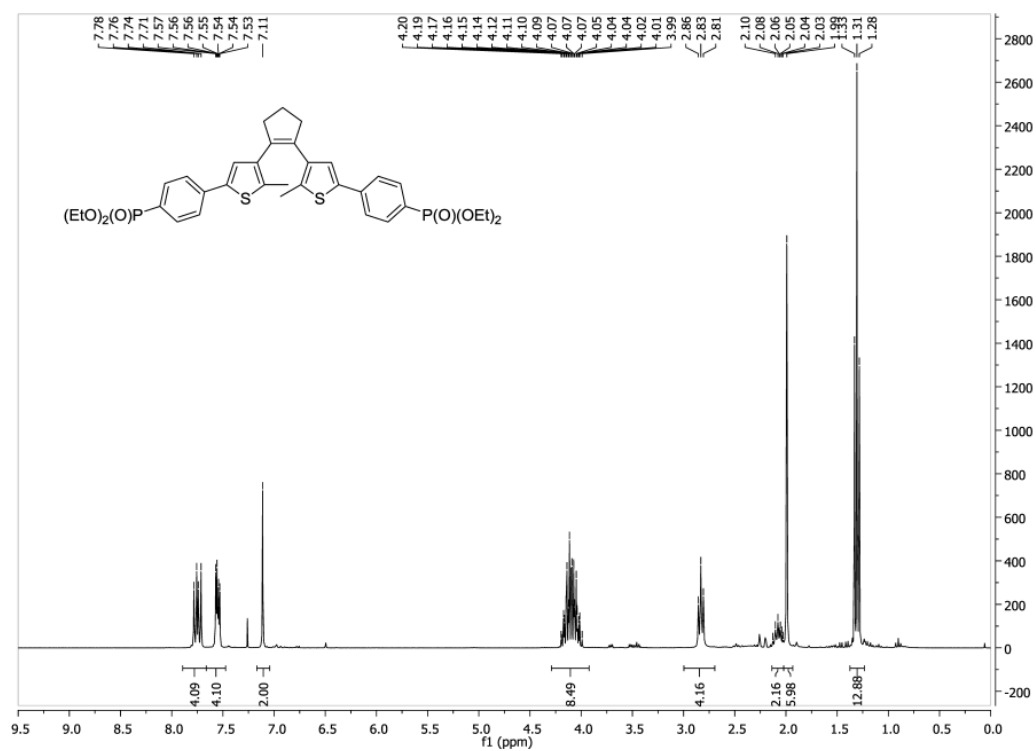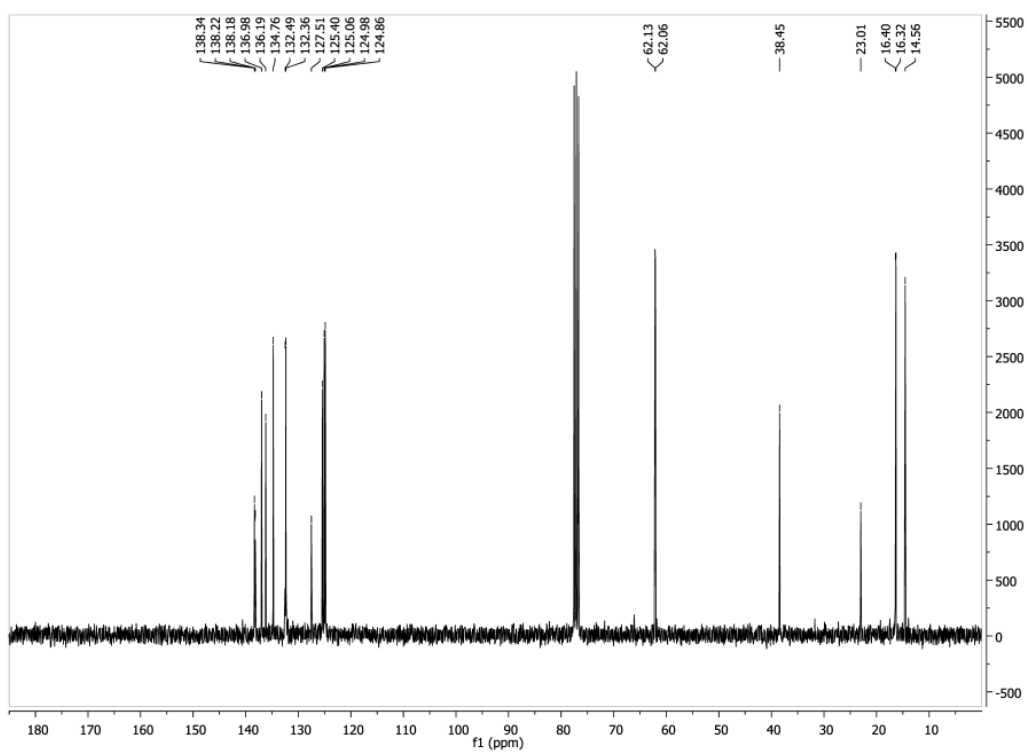
¹H-NMR for compound **12**:
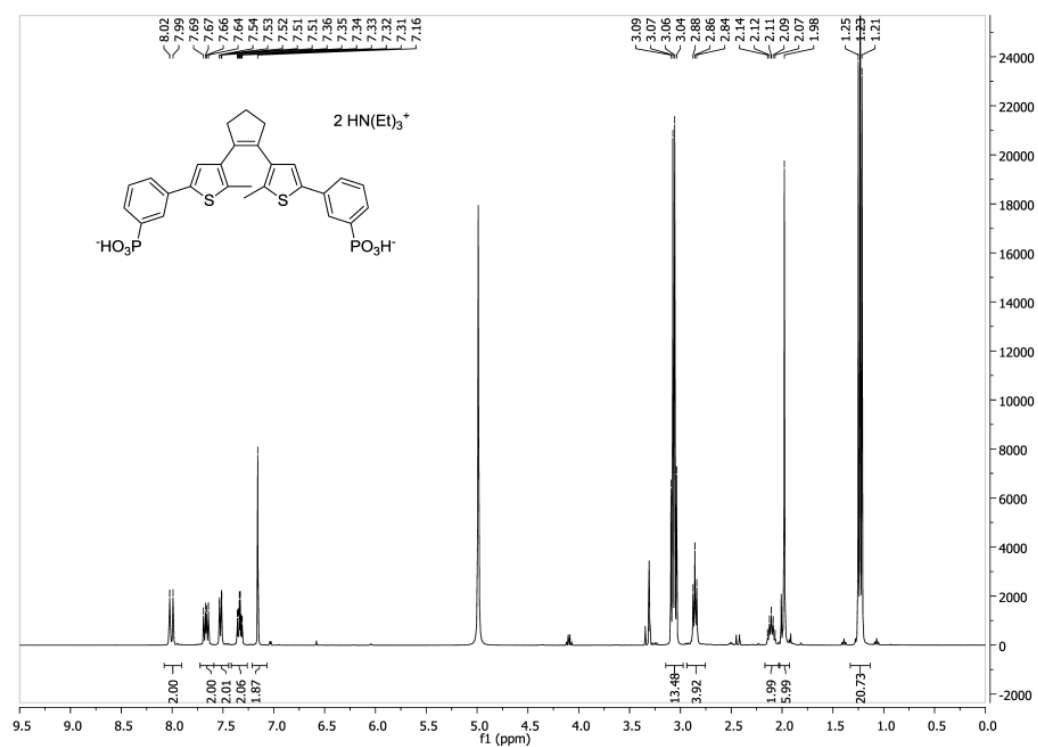
¹³C-NMR for compound **12**:
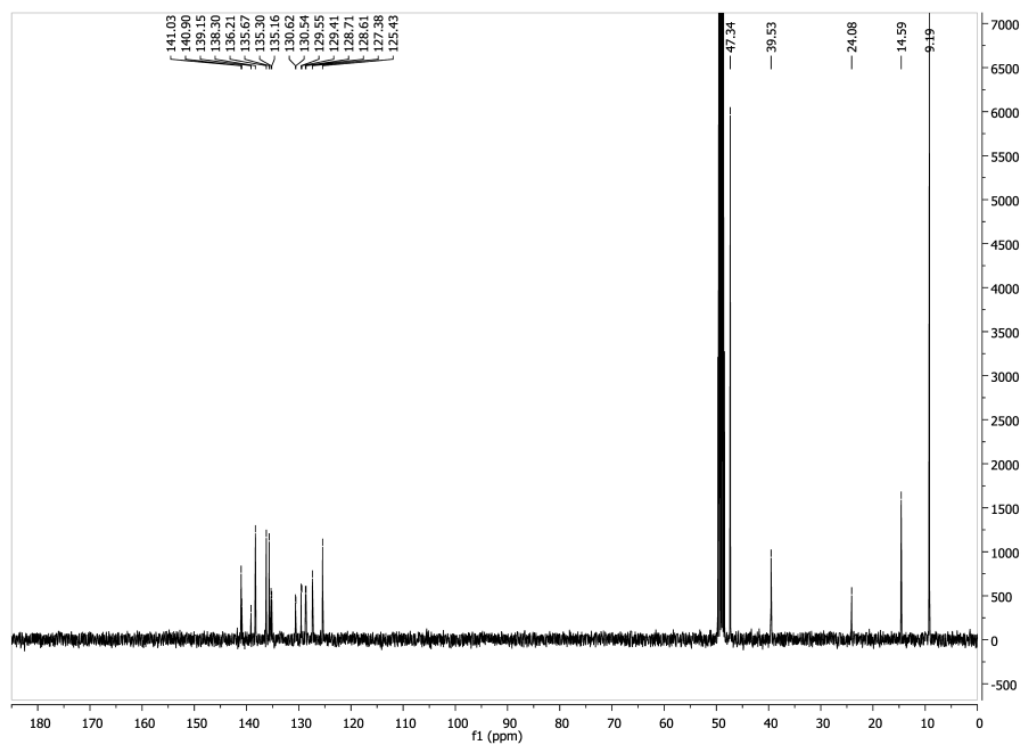
¹H-NMR for compound **13**:
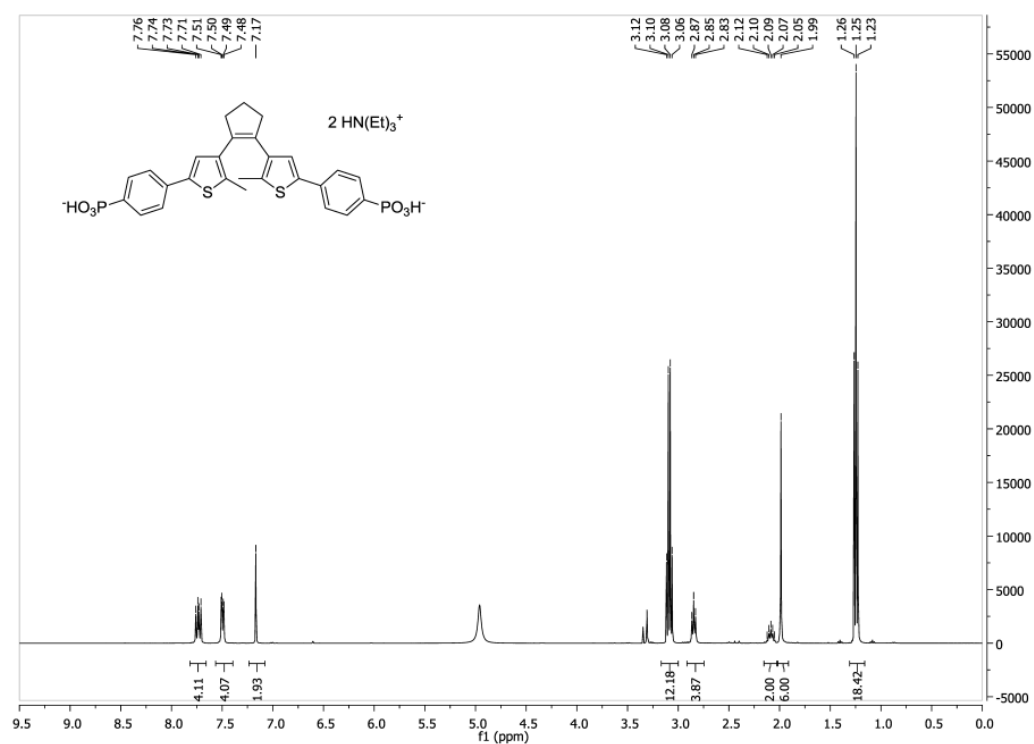
¹³C-NMR for compound **13**:
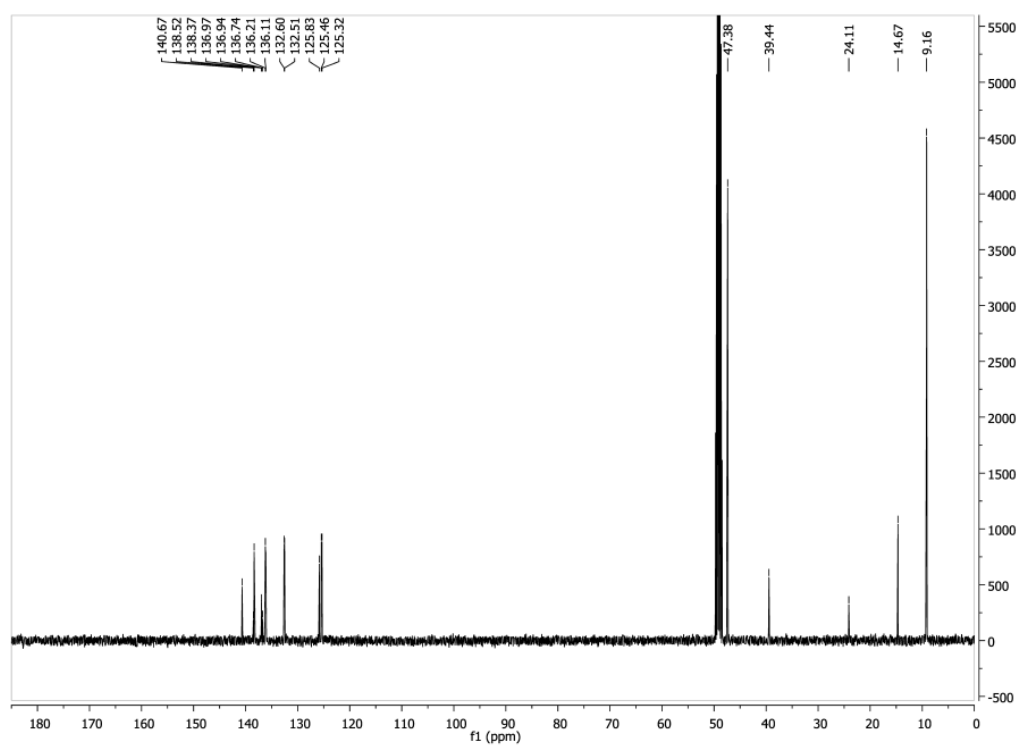
Supplementary References

[1]   L. N. Lucas, J. J. D. de Jong, J. H. van Esch, R. M. Kellogg, B. L. Feringa, *Eur. J. Org. Chem.* **2003**, 155-166.

[2]   T. Kudernac, J. J. De Jong, J. Van Esch, B. L. Feringa, D. Dulic, S. J. Van Der Molen, B. J. Van Wees, *Mol. Cryst. Liq. Cryst.* **2005**, *430*, 205-210.

[3]   M. Akazawa, K. Uchida, J. J. D. de Jong, J. Areephong, M. Stuart, G. Caroli, W. R. Browneb, B. L. Feringa, *Org. Biomol. Chem.* **2008**, *6*, 1544-1547.

[4]   A. V. Due, J. Kuper, A. Geerlof, J. P. von Kries, M. Wilmanns, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3554-3559.

[5]   T. J. Klem, V. J. Davisson, *Biochemistry* **1993**, *32*, 5177-5186.

[6]   C. Jürgens, A. Strom, D. Wegener, S. Hettwer, M. Wilmanns, R. Sterner, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9925-9930.

[7]   Spartan '10 v.11.11.10, *Wavefunction Inc.*, 18401 Von Karman Ave., Suite 18370, Irvine, CA 92612.

[8]   E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein, G. Vriend, *Proteins* **2004**, *57*, 678-683.

[9]   U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577-8593.

[10]   A. Jakalian, D. B. Jack, C. I. Bayly, *J. Comput. Chem.* **2002**, *23*, 1623-1641.

[11]   E. Krieger, J. E. Nielsen, C. A. Spronk, G. Vriend, *J. Mol. Graphics Modell.* **2006**, *25*, 481-486.

Abbreviated Main Reference

[3]   c) F. Zhang, J. Vierock, O. Yizhar, L. E. Fenno, S. Tsunoda, A. Kianianmomeni, M. Prigge, A. Berndt, J. Cushman, J. Polle, J. Magnuson, P. Hegemann, K. Deisseroth, *Cell* **2011**, *147*, 1446-1457.

S27

6 List of figures

7   List of tables

8   List of acronyms and abbreviations

Å	Ångström (10^{-10} m)
AICAR	5-aminoimidazole-4-carboxamide ribotide
$APAD^+$	acetylpyridine adenine dinucleotide
CD	circular dichroism
CdRP	1-(2-carboxy-phenylamino)-1´-deoxyribulose-5´-phosphate
$[D]_{1/2}$	transition midpoint of denaturant-induced unfolding
DSC	differential scanning calorimetry
DTE	1,2-dithienylethene
E. coli	*Escherichia coli*
ec	Enzyme from *Escherichia coli*
GdmCl	guanidinium chloride
Gyr	gigayear (10^9 years)
HisA	N´-[(5´-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide isomerase
HisAF	chimera consisting of the N-terminal $(\beta\alpha)_{1-4}$-units of HisA and the C-terminal $(\beta\alpha)_{5-8}$-units of HisF
*his*F	gene coding for HisF
HisF	synthase subunit of imidazole glycerol phosphate synthase
HisH	glutaminase subunit of imidazole glycerol phosphate synthase
ImGP	imidazole glycerol phosphate
k_{cat}	turnover number
k_{cat}/K_M	catalytic efficiency parameter
K_d	dissociation constant
K_M	Michaelis constant; equivalent to the substrate concentration at half-maximum rate
K_i	inhibition constant
LUCA	last universal common ancestor
LUCA-HisF	reconstructed HisF protein from the era of the last universal common ancestor
MD	molecular dynamics
mt	Enzyme from *Mycobacterium tuberculosis*
m-value	parameter defining the cooperativity in denaturant-induced unfolding
n	nano (10^{-9})
NAD^+	nicotinamide adenine dinucleotide

PDB ID	4-character unique identifier of every structure entry in the Protein Data Bank
PRA	phosphoribosylanthranilate
PRAI	phosphoribosylanthranilate isomerase
ProFAR	N´-[(5´-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide
PRFAR	N´-[(5´-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide
PurF	glutamine phosphoribosylpyrophosphate amidotransferase
QM/MM	quantum mechanics/molecular mechanics
rmsd	root mean square deviation
rCdRP	reduced 1-(2-carboxy-phenylamino)-1´-deoxyribulose-5´-phosphate
S. cerevisiae	*Saccharomyces cerevisiae*
SAC	substrate-assisted catalysis
sc	Enzyme from *Saccharomyces cerevisiae*
SCOP	structural classification of proteins
T. maritima	*Thermotoga maritima*
tm	Enzyme from *Thermotoga maritima*
Tris	tris(hydroxymethyl)aminomethane
TrpA	α-subunit of tryptophan synthase
TrpC	indole-3-glycerol phosphate synthase
TrpF	phosphoribosylanthranilate isomerase
zm	Enzyme from *Zymomonas mobilis*
ΔG_D	Gibbs free-energy of unfolding
ΔhisF	*Escherichia coli* strain with deleted *his*F gene
μ	micro (10^{-6})

9 Acknowledgement

I am deeply grateful to all current and former members of the Sterner research group. I definitely enjoyed the excellent working atmosphere including mutual support and valuable discussions. I also appreciated their humor and entertainment as well as plenty of exciting table soccer and dart matches.

I cordially thank my parents and friends for their unlimited encouragement and support. Particularly, I am grateful to Maria Pirzer, who proofread this work.

Dear Sabine: your contribution to this dissertation and to far more important things in life is hard to put into words. The last years with you were marvelous and incredibly precious. Your charming way and constant support are invaluable companions.