Control of reactive intermediates in enzymes and enzyme complexes

Dissertation

zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Bastian Vögeli

aus Zürich, Schweiz

Marburg/Lahn, Deutschland, 2018

Die Untersuchungen zur vorliegenden Arbeit wurden von November 2014 bis Mai 2018 unter der Betreuung von Herrn Dr. Tobias Jürgen Erb in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Abteilung „Biochemistry and Synthetic Metabolism“ durchgeführt.

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation

angenommen am: 13.07.2018

Erstgutachter: Dr. Tobias Erb

Zweitgutachter: Prof. Dr. Johann Heider

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Torsten Waldminghaus

Prof. Dr. Gert Bange

Tag der mündlichen Prüfung: 10.08.2018

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „**Control of reactive intermediates in enzymes and enzyme complexes**“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 08. Mai 2018

Bastian Vögeli

“Reality is frequently inaccurate.”

“Don't Panic.”

— Douglas Adams, The Restaurant at the End of the Universe

Contents

Summary

Enzymes are the catalysts of life. They accelerate the rate of chemical reactions that would otherwise take longer than an organism's lifetime to take just millisecond. To achieve these remarkable rate enhancements enzymes arrange into a three dimensional fold that places its amino acids in a way, which binds the transition state of the reaction better than the substrates and products of the reaction, thereby lowering the activation energy of the reaction. Enzymes are also very specific and often only catalyze one specific chemical transformation without producing side products. They are able to achieve all this under ambient temperatures and in cells that contain over 2700 different metabolites.

In this work we focus on the mechanisms enzyme use to control reactive intermediates both inside their active site and between enzymes of a metabolic pathway to avoid the formation of deleterious side products. In the first part we investigate the catalytic cycle of NAD(P)H dependent oxidoreductases. We show that the two enoyl-thioester reductases; Etr1p from *Candida tropicalis* of the MDR enzyme superfamily and InhA from *Mycobacterium tuberculosis* of the SDR enzyme superfamily form a covalent adduct between substrate and the C2 carbon of the cofactor. The observation of this reactive intermediate at the active site of enzymes from the two largest NAD(P)H dependent oxidoreductase superfamilies not only calls for a careful reconsideration of the canonical reaction mechanism of these enzymes, but also sets the basis for the development of novel tools to study, manipulate and inhibit their catalytic cycle. We demonstrate this by successfully changing the protonation specificity of Etr1p from *re*- to *si*- face. Using the molecular probe we show that a conserved threonine at the active site of Etr1p is mainly responsible for preventing the formation of a toxic side product and not for the stabilization of the wanted transition state along the reaction coordinate. This effect of destabilization of unwanted transition states, often termed ´negative catalysis´, poses a complementary mechanism of reaction control to the canonical transition state theory and is discussed in detail in this work.

In the second part of this thesis we take a look at two enzyme complexes and the strategies they use to control the transfer of a reactive intermediate from one active site to the next one. The trifunctional propionyl-CoA synthase forms a closed reaction chamber to sequester the reactive acrylyl-CoA intermediate. This reaction chamber encloses all three active sites of the enzyme fusion protein, but does not show the directionality of a conventional tunnel, and the CoA ester intermediates are not covalently attached to the enzyme but freely diffuse within the compartment. The substrate channeling mechanism of the thiolase/HMG-CoA synthase complex of archaea most closely resembles the covalent swinging arm fatty acid and polyketide synthases use to channel their intermediates. In the thiolase/HMG-CoA synthase complex the intermediate is however not covalently attached, but instead tightly bound in a shared CoA binding site, enabling the pantothenyl-arm of CoA to swing from the thiolase active site to the HMG-CoA synthase active site. The two channeling systems we describe in this work therefore represent two alternative ways of channeling CoA ester intermediates in a non-covalent fashion.

Zusammenfassung

Enzyme sind die Katalysatoren des Lebens. Sie beschleunigen die Rate chemischer Reaktionen, welche unkatalysiert länger als die Lebensspanne eines Organismus dauern würden. Enzyme erreichen diese beeindruckende Beschleunigung indem sie den Überganszustand besser als Substrat und Produkt binden, wodurch sie die Aktivierungsenergie der Gesamtreaktion senken. Zusätzlich sind Enzyme äußerst spezifisch und katalysieren normalerweise exakt eine chemische Reaktion, ohne dabei Seitenprodukte zu produzieren. All dies erreichen sie bei moderaten Temperaturen und innerhalb von Zellen, welche über 2700 unterschiedliche Metabolite enthalten.

In dieser Arbeit untersuchen wir die Mechanismen, welche Enzyme nutzen um reaktive Zwischenprodukte zu kontrollieren und damit die Bildung von schädlichen Seitenprodukten verhindern. Im ersten Teil betrachten wir den katalytischen Zyklus von NAD(P)H abhängigen Oxidoreduktasen im Detail. Wir zeigen, dass beide untersuchten Enoylthioesterreduktasen, Etr1p aus *Candida tropicalis* und der MDR Enyzmsuperfamilie und InhA aus *Mycobacterium tuberculosis* und der SDR Enzymsuperfamilie, ein kovalentes Zwischenprodukt zwischen dem Substrat und NAD(P)H bilden. Die Beobachtung, dass Enyzme aus beiden Superfamilien ein solches Zwischenprodukt herstellen, stellt nicht nur den herkömmlichen Reaktionsmechanismus in Frage, sondern setzt auch die Grundlage für die Entwicklung von neuen Methoden zur Manipulierung, Erforschung und Inhibierung solcher Oxidoreduktasen. Wir demonstrieren dies mit dem erfolgreichen Umbau der Protonierungsspezifität in Etr1p von der *re*- zur *si*-Seite. Zusätzlich nutzen wir das kovalente Zwischenprodukt um aufzuzeigen, dass ein konserviertes Threonin im aktiven Zentrum von Etr1p hauptsächlich dazu dient, die Bildung eines toxischen Seitenproduktes zu verhindern und nicht um den gewünschten Übergangszustand zu stabilisieren. Threonin destabilisiert stattdessen den Übergangszustand zum ungewünschten Seitenprodukt. Dieser Effekt der Destabilisierung wird oft ´negative Katalyse´ genannt und ergänzt die kanonische Theorie des Übergangszustandes. Die Auswirkungen dieses Effektes auf unser Verständnis der Enzymkatalyse werden in dieser Arbeit im Detail diskutiert.

Im zweiten Teil dieser Arbeit untersuchen wir zwei Enzymkaskaden und deren Mechanismen, welche den Transfer reaktiver Zwischenprodukte von einem aktiven Zentrum ans nächste ermöglichen. Die trifunktionelle Propionyl-CoA Synthase bildet eine geschlossene Reaktionskammer, um das reaktive Zwischenprodukt Acrylyl-CoA daran zu hindern ins Zytoplasma zu entweichen. Diese Reaktionskammer enthält alle drei aktiven Zentren des Fusionsproteins und ermöglicht dadurch Katalyse der gesamten Enyzmkaskade innerhalb der Kammer. Sie unterscheidet sich dabei jedoch von zielgerichteten Tunneln und anderen herkömmlichen Kanalisierungsmechanismen. Der Thiolase/HMG-CoA-Synthasekomplex von Archaeen kanalisiert das Zwischenprodukt Acetoacetyl-CoA, ähnlich wie Polyketid- und Fettsäurensynthasen, indem er die Pantothenylkomponente von CoA als Arm verwendet, um das Zwischenprodukt von einem aktiven Zentrum ans nächste weiterzureichen. Im Unterscheid zu anderen Systemen ist der Panthothenylarm hierbei jedoch nicht kovalent gebunden, sondern Teil des CoA-Esters, weshalb der Enzymkomplex eine geteilte CoA-Bindetasche besitzt. Die beiden hier beschriebenen Mechanismen stellen daher zwei neue Strategien dar, um mehrere Reaktionen nicht kovalent gebundener CoA-Ester zu kanalisieren.

Parts of this thesis have been published or are in preparation for publication:

Rosenthal, R.G.*; Vogeli, B.*; Quade, N.; Capitani, G.; Kiefer, P.; Vorholt, J.A.; Ebert, M. O.; Erb, T.J. (2015) The use of ene adducts to study and engineer enoyl-thioester reductases. *Nat Chem Biol* 11(6): 398-400

Rosenthal, R. G., B. Vögeli, T. Wagner, S. Shima and T. J. Erb (2017). A conserved threonine prevents self-intoxication of enoyl-thioester reductases. *Nat Chem Biol* 13(7): 745-749.

Vögeli, B.*; Rosenthal, R.G.*; Stoffel G.M.M.*; Wagner, T.; Kiefer, P.; Cortina, N.S.; Shima, S.; Erb, T.J. (2018) InhA, the enoyl-thioester reductase from M. tuberculosis forms a covalent reaction intermediate. *submitted*

Vögeli, B.; Gerlinger, P.D.; Geyer, K.; Benkstein, S.; Cortina, N.S.; Erb, T.J. (2018) Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis. C*ell Chem Biol* 25: 1-7

Bernhardsgrütter, I.*; Vögeli, B.*; Wagner, T.; Peter, D.M.; Cortina, N.S.; Bange, G.; Engilberge, S.; Girard, E.; Riobé, F.; Maury, O.; Shima, S.; Zarzycki, J.; Erb, T.J. (2018) Trifunctional propionyl-CoA synthase: A self-organizing, multi-catalytic 'pico'-compartment that sequesters a reactive intermediate. *submitted*

Vögeli, B; Engilberge, S.; Girard, E.; Riobe, F.; Maury, O.; Erb, T.J.; Shima, S.; Wagner, T. (2018) Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *Proc Natl Acad Sci USA* 115(13): 3380-3385.

Vögeli, B.; Erb, T.J. (2018) 'Negative' and 'positive catalysis': Complementary principles that shape the catalytic landscape of enzymes. *Submitted*

Publications that are not discussed in this thesis:

Peter, D. M.*; Vogeli, B.*; Cortina, N.S.; Erb, T.J. (2016). A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters. *Molecules* 21(4).

Krink-Koutsoubelis, N.; Loechner, A.C.; Lechner, A.; Link, H.; Denby, C.M.; Vögeli, B.; Erb, T.J.; Yuzawa, S.; Jakociunas, T.; Katz, L.; Jensen, M.K.; Sourjik, V.; Keasling, J.D. (2018). Engineered Production of Short-Chain Acyl-Coenzyme A Esters in Saccharomyces cerevisiae. *ACS Synth Biol* 7(4): 1105-1115.

1. Introduction

1.1. Enzymes

Enzymes make the chemistry of life possible by drastically increasing the rate of chemical reactions that would otherwise take too long to sustain life as we know it. They are able to accelerate rates of chemical reactions by many orders of magnitude compared to uncatalyzed reaction. In the extreme case of arginine decarboxylase the uncatalyzed reaction has a half-life of approximately the age of Earth, while the enzyme catalyzes more than 1000 reactions per second (rate enhancement of 10^{17})[1]. The mechanism of this rate acceleration was first proposed by Linus Pauling in 1946, without the knowledge of any structural information. He proposed that enzymes can accelerate rates due to better binging of the transition state than the substrate, therefore lowering the activation free energy of the reaction[2]. This key concept of lowering the activation free energy remains the main strategy of enzymes and only needs to be extended by some quantum mechanical effects that change the transmission coefficient of reactions, such as tunneling effects in reactions involving hydrogen transfers and non-equilibrium effects[3]. There are various mechanisms that enzymes can employ to lower the activation free energy (along with tight binding of the transition state). If there is a difference in charge distribution between the substrate and the transition state, electrostatic effects, such as hydrogen bonding often play an essential role in stabilizing the transition state (eg. a charged lysine lowers the activation energy of enolate intermediate formation in triosephosphate isomerase (TIM)[3-4]). Conformational stabilization of the reactant can also contribute to the lowering of the activation free energy barrier. In chorismate mutases for example, chorismate, the substrate of the reaction, is stable in an inactive conformation in solution and gets transformed into the active form upon binding to the enzyme. This preorganization of the substrate into a "near-attack conformer" has been estimated to contribute about 5 kcal/mol to lowering of the activation free energy, the more polar transition state is then further stabilized by electrostatic interactions[5-6]. The ability of enzymes to exclude water from the active site can be another effect that lowers the activation free energy. In S_N2 reactions the substrate is better solvated than the transition state, which leads to an energy penalty for its removal during the reaction. Haloalkane dehalogenases use this desolvation effect to effectively lower the activation free energy by an estimate of 6 kcal/mol compared to the reaction in water[7]. Another strategy that enzymes employ involves the enzyme free energy (often termed enzyme conformational change), which can be different for reactant and transition state binding. The binding of reactant to orotidine monophosphate decarboxylase for example induces a strain in the enzyme that is only relieved at the transition state[8]. In summary, natural selection has found a wide range of molecular mechanisms that allow enzymes to achieve remarkable rate enhancement of chemical reactions mostly by reducing the reactions active free energy.

Next to the rate enhancements, enzymes are also highly specific and are able to work in the complex metabolite mixture of the cellular cytoplasm. How do enzymes control the outcome of chemical reactions to achieve the remarkable stereo- and region-selectivity they display? How do they control highly reactive intermediates in their active site and avoid the formation of deleterious side products? These questions are discussed in Chapter 1.2.

In Chapter 1.3 the challenges enzyme cascades face in the cellular context are discussed. How do they avoid cross-talk between their intermediates and other cytosolic enzymes? How do enzymes in a cell manage to catalyze more than 1000 different reactions in parallel?

1.2. Control of reactive intermediates in Enzymes

Enzymes are often referred to as remarkably specific catalysts, meaning that they catalyze one chemical reaction with high region- and stereo-selectivity. This raises the question of how enzymes control all the reactive intermediates and transition states during a chemical transformation to guide the reaction towards the desired product. Note that there are many enzymes that use multiple substrates, but still catalyze one chemical reaction. These enzymes should still be regarded as specific, regarding their reaction mechanism, as they still catalyze the transformation of one of the substrates into exactly one corresponding region- and stereospecific product. There are however some enzymes that produce considerable amounts of unwanted side products, despite being under strong evolutionary pressure to avoid them [9]. Analyzing the challenging reactions those enzymes catalyze and the solutions nature has come up with to minimize the side reactions helps us to understand the general principles of enzymatic reaction control.

Triosephosphate isomerase, an enzyme that in terms of rate enhancement of its reaction has achieved perfection [1], forms the side product methylglyoxal (MG) approximately once every million reactions (k_{cat}/K_m for the side reaction is 14 $M^{-1}s^{-1}$). MG is produced during the reactive cycle via a phosphate elimination reaction from the reactive enediol(ate) phosphate intermediate (**Figure 1A**, (a)). Enediol(ate) phosphate intermediates are unstable in water and eliminate phosphate with estimated rate constants between $8 \cdot 10^6$ and $8 \cdot 10^8$ s^{-1}, which is at least 100 times faster than the protonation in solution [10]. This rapid decomposition of the intermediate in water highlights the need for the enzyme to control and stabilize it in the active site to avoid the fast side reaction. Knowles and coworkers showed that a loop in the enzyme closes the active site upon substrate binding and contributes critically to the stabilization of the enediol(ate) intermediate. The closing of this loop was shown by NMR studies to be the rate limiting step of the overall reaction [11]. This loop closing allows the enzyme to bind the enediol(ate) intermediate stronger than the substrate and product and locks the phosphate group in a conformation disfavoring elimination. By lowering the relative free energy of the bound intermediate, the enzyme reduces the free energies of the two adjacent transition states, accelerating the overall reaction [12]. High resolution crystal structures of the tightly binding reaction intermediate analogue phosphoglycolohydroxamate (PGH) showed that it is bound in a planar form that minimizes the π orbital overlap between the double bond and the bond to the β substituent (the phosphate), thus disfavoring β-elimination [13-14]. Lysine 12 forms direct contacts with both the phosphate and one of the hydroxyl groups of the enediol(ate) in the structure and seems to force the intermediate into the planar conformation. Mutagenesis of the lysine into a glycine lead to a 12000 fold decrease in k_{cat} of the enzyme accompanied by a 180000 fold increase of methylglyoxal formation [15]. Lysine 12 is therefore involved in stabilizing the enediol(ate) phosphate intermediate and the transition states (positive contribution) as well as in destabilizing the transition state for the β-elimination (negative contribution). In summary, the triosephosphate isomerase on one hand reduces the side reaction by lowering the activation free energies of the desired reaction and on the other hand increases the activation energies for side product formation.

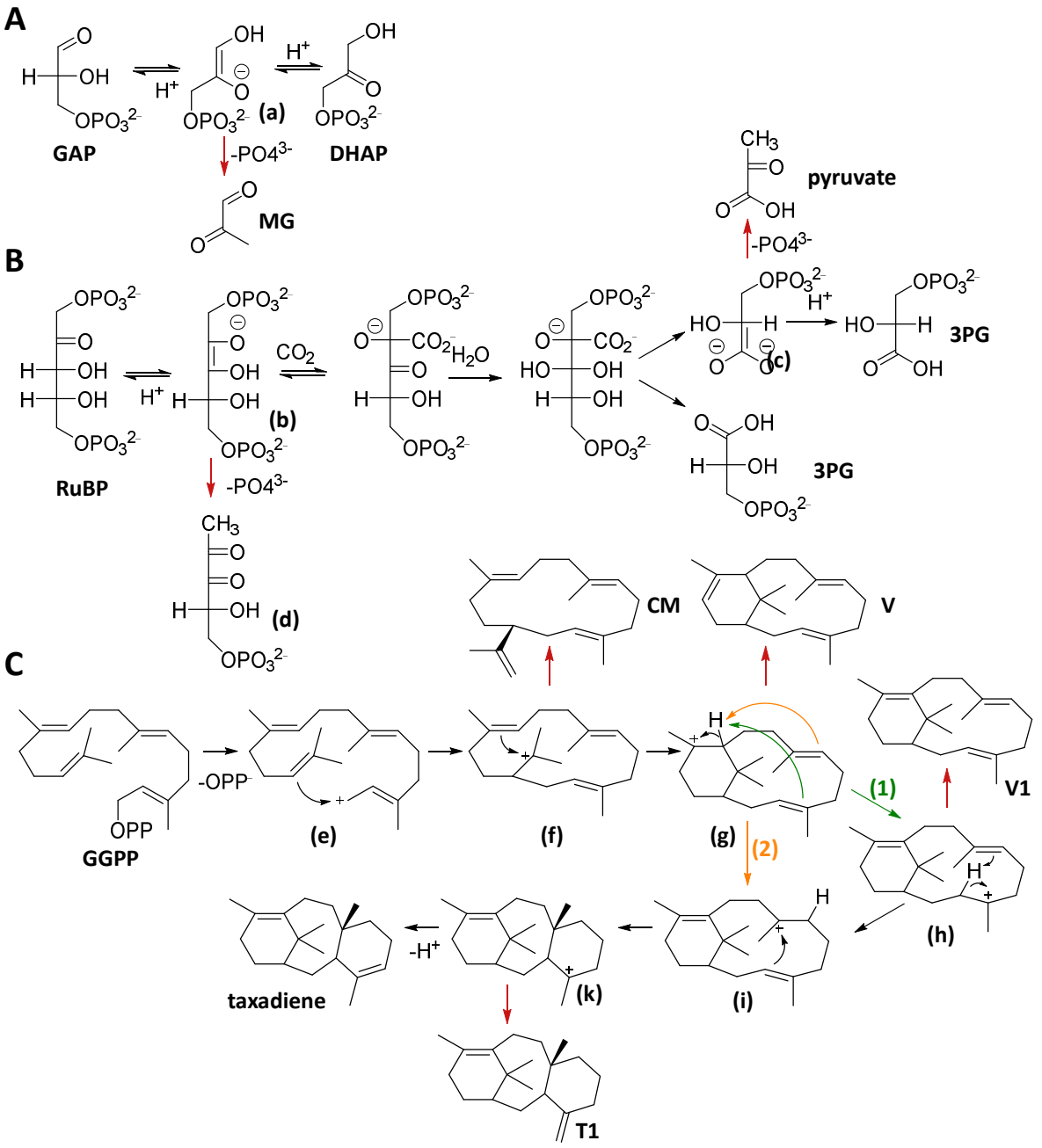


Figure 1. Reaction schemes of triosephosphate isomerase (A), ribulosebisphosphate carboxylase/oxygenase (B) and taxadiene synthase (C). Side reactions are displayed with red arrows. GAP = glyceraldehyde-3-phosphate, DHAP = dihydroxyacetone phosphate, MG = methylglyoxal, RuBP = ribulose-1,5-bisphosphate, 3PG =3-phosphoglycerate, GGPP = geranylgeranyl diphosphate, CM = cembrene A, V = verticillia-3,7,12(13)-triene, V1

Ribulosebisphosphate carboxylase/oxygenase (RuBisCO) catalyzes the carboxylation of ribulose-1,5-bisphosphate (RuBP) into two molecules of 3-phosphoglycerate (3PG). Its proposed reaction mechanism proceeds through a sequence of four intermediates. Two of them are reactive enediol(ate) phosphate intermediates analogous to TIM. In a first step a proton is abstracted from the C3 of RuBP yielding a 2,3-enediol(ate) phosphate intermediate of RuBP (b). This enolysation is readily reversible. A nucleophilic attack of the 2,3-enediol(ate) from the C2 position to CO_2 then forms a six carbon β-keto acid intermediate. The six carbon intermediate is then hydrated (it is unclear whether CO_2 addition and hydration are concerted or not [16]) and cleaved into 3PG and a *aci*-carbanion ((c), another enediol(ate) phosphate intermediate) of 3PG. The *aci*-carbanion (c) is then protonated to generate the

second 3PG [17] (**Figure 1B**). The two enediol(ate) double bonds positioned perpendicular to each other, making it difficult for the enzyme to force the phosphate group into a planar conformation with both of them to avoid elimination. Elimination from the 2,3-enediol(ate) phosphate intermediate of RuBP (b) forms a deoxypentodiulose phosphate ((d), 1-deoxy-D-glycero-2,3.pentodiulose-5-phosphate) side product, elimination from the *aci*-carbanion of 3PG (c) forms pyruvate. The formation of pyruvate physiologically only results in the cost of one ATP (could be gained in glycolysis in the transformation of 3PG to pyruvate), the deoxypentodiulose phosphate (d) on the other hand is a wasteful side product. It is therefore not surprising that all natural RuBisCOs characterized so far produce pyruvate approximately every 150 carboxylations, but none of them display any deoxypentodiulose phosphate (d) production [14]. Site-directed mutagenesis studies identified a threonine, which forms a hydrogen bonds to the C1 phosphate, to be important for preventing the elimination in the 2,3-enediol(ate) phosphate intermediate of RuBP (b). Mutagenesis of the threonine into a valine resulted in a 100 fold loss of k_{cat} for carboxylation, but was accompanied by 13% production of the deoxypentodiulose phosphate side product (d). Interestingly the pyruvate production was reduced to 18% of the WT in this variant [14]. Threonine therefore seems to stabilize the first enediol(ate) phosphate intermediate (b), but by doing so limiting the phosphate to rotate by the required 60-90° to achieve a conformation that minimizes β-elimination from the second enediol(ate) phosphate intermediate (c) [18]. RuBisCO seems therefore mechanistically unable to completely avoid both elimination reactions and was therefore evolutionarily selected to prevent the more deleterious one. Note that the second side reaction of RuBisCO, the oxidation of RuBP, is not affected by this mutation. The mechanistic discrimination between carboxylation and oxygenation during the catalytic cycle are not discussed in detail here, as the underlying factors are unclear for now (for detailed discussion see the comprehensive reviews by Tcherkez for carboxylation [17] and oxygenation [19]).

Another interesting example for the control of reactive intermediates is the family of terpene cyclases [20]. About 60% of all known natural products are terpenoids and they are all derived from relatively simple pyrophosphate carrying isoprenoid chains [21]. Even though terpene cyclases all use very similar acyclic precursors and follow a common reaction mechanism they are able to produce a huge variety of carbon backbones. The mechanism involves the formation of a reactive carbocation via the elimination of the pyrophosphate (PPi), followed by a sequence of electrophilic cycloadditions and rearrangements via further highly reactive carbocation intermediates [21]. The multitude of chemical transformations in terpene cyclases utilizing these carbocation intermediates involves region- and stereospecific ring formation, deprotonations to form double bonds, quenching of the carbocation by water to generate alcohols and stereospecific hydride, proton, methyl and methylene migrations [22]. The fact that the cyclized products are much more condensed than the substrates add to the complexity to use specific amino acid residues to stabilize formed carbocations along the reaction coordinate. It has been shown in multiple terpene cyclases that electrostatic effects stabilize mostly early stage cations and preorganization and stabilization of the substrate in a reactive conformation is used for late stage cations [23]. An interesting example for the control of reactive intermediates in terpene cyclases is the taxadiene synthase (TXS) that cyclizes geranylgeranyl diphosphate (GGPP) to taxadiene (taxa-4(5),11(12)-diene) with a yield of 93.2%, but also forms 4.7% of a taxadiene isomer (T1, taxa-4(20),11-diene) and 0.8 % of verticillia-3,7,12(13)-triene (V) and 1.1 % of verticillia-3,7,11(12)-triene (V1) (**Figure 1C**) [24]. Considering all the possible side products this is a remarkable high yield of the desired taxadiene product, which is the first committed biosynthetic step towards the clinically important tumor therapeutic Taxol [25]. Site-directed mutagenesis in combination with molecular mechanic simulations have identified the function of multiple residues in the active site of TXS that play an essential role in guiding the reactive carbocations towards the desired taxadiene product. Catalysis is initiated by Mg^{2+}-mediated GGPP elimination, forming the first reactive carbocation (e). The pyrophosphate ion is retained in a hydrophilic pocket in the active site during catalysis and

contributes to the electrostatic stabilization of multiple intermediates. The carbocation (e) is then converted into the monocyclic carbocation (f). Site-directed mutagenesis showed that Tryptophane 753 stabilizes (f) via aromatic π-π interaction, as all non-aromatic replacements resulted in inactive enzyme variants. The introduction of histidine as a proton acceptor in that position led to the predicted quenching of the carbocation and formation of the monocyclic side product (-)-(R)-cembrene A (CM) (Table 1). The carbocation (f) is then converted to (g). Valine 584 is positioned very closely to the positively charged carbocation. Mutations to methionine and leucine led to premature deprotonation and increased formation of V. In silico comparison of the WT and the valine variants showed that the carbocation moved closer to the PPi in the active site, which could be the cause for the premature deprotonation and side product formation. Precise steric positioning and stabilization of the substrate in an active conformation is therefore important to prevent this side reaction. From (g) there are two proton transfer pathways, an indirect one (1) forming first (h) and then (i) and a direct one forming (i) in one step. QM/MM calculations have shown that in the gas phase the indirect pathway (1) is favored compared to the direct one (2) (has a lower activation free energy). In the enzyme on the other hand the direct pathway (2) is favored over the indirect one (1) [9]. This suggests that TXS might have evolved to avoid the indirect pathway, which forms one more intermediate and therefore increased potential for side reactions.

TXS variant	activity	taxadiene	CM	V	V1	T1
TXS	100	93.2	N.D.	0.8	1.1	4.7
W753H	51	N.D.	100	N.D.	N.D.	N.D.
V584M	92	13.8	N.D.	83.4	1.4	N.D.
V584L	92	13.8	N.D.	83.4	1.4	N.D.

Table 1 Product distribution of interesting TXS variants, adapted from Schrepfer et al. [24]

Taken together these three examples show multiple mechanisms enzymes use to stabilize reactive intermediates and prevent the formation of undesired side products. Tight binding of the reactive intermediate in a conformation disfavoring the side reaction combined with the exclusion of water from the active site and the stabilization of the desired transition states are important mechanisms active in TIM. The case of RuBisCO additionally highlights an interesting trade-off that occurs from the necessity to prevent multiple competing side reactions. The case of terpene cyclases shows how preorganization and stabilization of the substrate in a reactive conformation and electrostatic and steric guidance as well as the choice of reaction mechanism help to prevent side reactions. All these examples also show how well enzymes are able to tune chemical reactions to achieve their remarkable region- and stereoselectivity.

1.2.1. MDR and SDR superfamily

The short- and medium-chain dehydrogenase/reductase superfamilies (SDR and MDR) are the two biggest enzyme families that use nicotinamide adenine dinucleotides (NAD(P)H) as a cofactor. They are old families that appeared early in evolution and catalyze various reduction and oxidation reactions in all domains of life. SDRs seem to have emerged earlier than MDRs, as they show greater abundance and divergence as well as a less complex fold [26]. Both families share similarities in their cofactor binding domain, the Rossmann-fold. SDRs only contain this cofactor binding domain with some extension. MDRs on the other hand typically have two domains, the cofactor binding domain and a "catalytic domain" [26]. The MDRs can then be further split up into metallo-enzymes using zinc and zinc-free families. The metallo-enzymes are believed to originate from zinc-free MDRs that have acquired zinc binding [27]. The widespread occurrence in all domains of life, their functional diversity as well as their

importance as drug targets has made MDRs and SDRs important subjects for mechanistic studies. Out of these superfamilies, the reaction mechanism of the subfamily of zinc-dependent MDRs is characterized in most detail. This is due to the early purification of both the yeast and horse liver alcohol dehydrogenase in 1935 and 1937, respectively, and the accessibility of their substrate alcohols [28-29]. Here we study the reaction mechanism of the less well characterized enoyl-thioester reductases and carboxylases, a subfamily of zinc-independent MDRs and SDRs.

1.2.2. Enoyl-thioester reductases and carboxylases

A subfamily of enoyl-thioester reductases (ETR) can be found in both the SDR and MDR superfamily. They both catalyze the NAD(P)H-dependent reduction of enoyl-thioesters into acyl thioesters involving a hydride transfer from NAD(P)H to the β-carbon and a proton from water or an active site proton donor to the α-carbon of the enoyl-thioester. There are however some differences in the active site architecture between the two families. In MDRs the pro-(4*R*) hydride of NADPH is transferred to the *re*-face of the β-carbon [30]. In SDRs on the other hand the pro-(4*S*) hydride of NADPH is transferred to the *si*-face of the β-carbon [31]. The stereochemistry of the proton donor can vary between different classes and both *re*- and *si*-face donation can be found in either. A special case are the enoyl-CoA reductase/carboxylases (ECR) that are able to use CO_2 instead of a proton as the resolving electrophile of the reaction and add it to the *re*-face of the α-carbon [30]. There are three subfamilies of ETRs in the MDR superfamily; (i) the ER domains of mammalian and insect fatty acid synthase multi-enzyme complex (called FAS I), (ii) the ER domains of bacterial and fungal polyketide synthases and (iii) the stand-alone enzymes in mitochondrial fatty acid biosynthesis (mtFAS II). Plants, apicomplexan protozoa (eg. *P. falciparum, T. gondii*) and most bacteria use SDR ETRs for fatty acid biosynthesis [32]. This difference makes the SDR ETRs very important drug targets for specific inhibition of fatty acid biosynthesis and interesting objects for mechanistic studies. Another important SDR drug target is the human 5α-reductase that reduces testosterone into dihydrotestosterone, thereby activating it. Finasteride, a mechanism-based 5α-reductase inhibitor that forms a covalent adduct with the NADPH cofactor, is used in treatment for benign prostatic hyperplasia and as a drug against male hair loss [33].

Recent reports describe the detection of a covalent adduct between the substrate crotonyl-CoA and NADPH in enoyl-CoA reductase/carboxylase (ECR) when CO_2, the resolving electrophile is omitted from the reaction [34]. Similar covalent adducts between the substrate and the dihydropyridine ring of the NAD(P)H cofactor have also been reported recently in chemical model reactions [35], in the natural product sanguinarine [36] and in the catalytic cycle of polyketide associated enoyl-CoA reductases of the MDR superfamily [37]. These observations raise the question, whether these adducts are true intermediates in the catalytic cycles of these zinc-free MDRs, which would suggest an alternative reaction mechanism for the hydride transfer from NAD(P)H to the substrate. Alternatively these adducts could also be part of a mechanism to stabilize the reactive enolate intermediate through this covalent bond and thereby preventing it to react with the wrong resolving electrophile (eg. in ECRs with a proton instead of CO_2). In chapter 2 we investigate the detailed reaction mechanism of the mitochondrial ETR from *C. tropicalis* from the MDR superfamily and show that it forms multiple covalent adducts between its substrate crotonyl-CoA and the cofactor NADPH when the resolving electrophile (a proton from an active site tyrosine) is removed. In chapter 3 we then show that the active site of Etr1p contains a conserved threonine that specifically prevents the formation of one of these adducts, which turned out to be a very competent inhibitor for the enzyme. In chapter 4 we take a detailed look at the reaction mechanism of the SDR ETR InhA from *M. tuberculosis* and show that this enzyme forms similar covalent adducts. The detection of such covalent adduct in both MDR and SDR superfamily ETRs calls for a careful reconsideration of the reaction mechanism of nicotinamide dependent oxidoreductases and in case of Etr1p highlights an interesting case of an single amino acid

residue in an active site, which prevents the highly reactive enolate intermediate from forming a toxic side product.

1.3. Reactive intermediates in enzyme complexes

The high rate enhancements under mild conditions, stereo- and region-specificity and the ability to precisely control reactive intermediates during their catalytic cycle are not the only remarkable features that make enzymes such proficient catalysts. Enzymes additionally face the challenge that all enzymes of a metabolism need to be able to work in parallel in the cytoplasm of living organisms. An *E.coli* cell for example is able to catalyze almost 2000 different reactions, contains about 1600 enzymes and about 2700 metabolites [38]. Diffusion is relatively fast compared to biochemical reactions, which leads to uniform concentration of metabolites in the cytoplasm, in particular when microbial cells are considered. This means that the enzymes need to be able to function efficiently in presence of most other metabolites of a cell. This can cause the problem of 'cross-talk' between metabolic pathways and in extreme cases cause cross-inhibition, inactivation or even irreparable damage to metabolism. Nature therefore has evolved several strategies to prevent free diffusion of problematic pathway intermediates into the bulk solvent. These strategies include i) compartmentalization and encapsulation of intermediates in membrane or protein delimited organelles (e.g., peroxisomes, bacterial microcompartments, encapsulins), ii) covalently linking of the intermediates to multi-domain enzyme complexes or carrier proteins (e.g., polyketide synthases, fatty acid synthesis type I, pyruvate dehydrogenase complex), iii) electro-statically guiding intermediates over charged protein surface from one active site to the other (e.g., dihydrofolate-thymidylate synthase, malate dehydrogenase-citrate synthase complex) and iv) formation of intramolecular tunnels between two active sites (e.g., tryptophan synthase, carbamoyl-phosphate synthetase)[39]. All these strategies ensure that the respective intermediates are not released into the bulk solvent, but directly transferred to the next, subsequent enzyme or active site; a process called substrate channeling. Besides protecting cells from toxic intermediates, substrate channeling can help to prevent the loss of unstable and volatile intermediates (e.g. loss of the indole intermediate in tryptophan synthase, which freely diffuses over the membrane), or to control the reaction sequence within a pathway (e.g. modifications of a growing polyketide chain). In the case of malate dehydrogenase the channeling of the intermediate oxaloacetate to the citrate synthase produces a low concentration microenvironment that helps to overcome the unfavorable thermodynamic equilibrium of the malate dehydrogenase reaction. This leads to an increase of the catalytic efficiency of the overall reaction. However, it is important to keep in mind that there is also a cost to benefit tradeoff for all of these mechanisms. In the case of compartmentalization the cost comes in form of the energy that is required to produce and maintain the compartment. The other three mechanisms require the stoichiometric assembly of the pathway's enzymes thus limiting the overall speed to the slowest enzyme within the reaction sequence. It follows that pathways that organize their enzymes in stoichiometric complexes or within a compartment must gain a benefit from doing so to compete with freely diffusing pathways.

1.4. Aims of this thesis

The general aim of this thesis was to gain a better understanding into the mechanisms that enzymes use to control reactive intermediate, both inside one active site and between enzymes of a metabolic pathway.

In the first part of the thesis the function of the recently discovered covalent intermediate between the substrate and the C2 carbon of the NADPH cofactor in ECR was studied in detail and the implications that this discovery has for our understanding of the reaction mechanism of NADPH dependent zinc-free MDR and SDR oxidoreductases are discussed. In Chapter 2 we show that the mitochondrial enoyl-thioester reductase Etr1p of *C. tropicalis* forms two covalent intermediates; a catalytically competent C2-adduct (analogous to the above discussed ECR) and a toxic, dead-end C4-adduct, when the resolving electrophile is removed from the reaction. We then use the competent C2-adduct as a chemical probe to access and engineer the second part of the reaction, the protonation, independent of the first part, the hydride transfer. Chapter 3 then investigates how active site residues in Etr1p prevent the formation of the toxic C4-adduct by destabilizing the transition state of the side reaction (´negative catalysis´) and shows that this feature of negative catalysis is conserved in many enzymes of the MDR superfamily. In Chapter 4 we show that the C2-adduct is also formed in InhA, an enoyl-thioester reductase of the SDR superfamily, which indicates that these adducts might be more wide spread as previously recognized. In Chapter 5 the gained insights into the reaction mechanism of ETRs and especially ECRs are then used for the biosynthesis of a large library of atypical malonyl-CoA derivatives. These derivatives are important tools for studying the function of polyketide synthases.

In the second part of the thesis we investigate two systems where reactive intermediates are transferred between two active sites of enzymes within a pathway. In Chapter 6 we analyze the trifunctional fusion protein propionyl-CoA synthase of *Erythrobacter* sp. NAP1 and show that the enzyme forms a self-assembling, multi-catalytic compartment to sequester the reactive intermediate acrylyl-CoA. In Chapter 7 we investigate the initial steps in archaeal isoprenoid biosynthesis, the mevalonate pathway. We show that archaea use an enzyme complex to couple the first reaction of the pathway, the condensation of two molecules of acetyl-CoA (catalyzed by an acetoacetyl-CoA thiolase), to the second reaction of the pathway, the condensation of acetoacetyl-CoA and acetyl-CoA to HMG-CoA (catalyzed by an HMG-CoA synthase). Channeling of acetoacetly-CoA between the active site of the thiolase and the HMG-CoA synthase allows archaea to overcome the thermodynamically unfavorable endergonic first reaction by coupling it to the exergonic second one.

1.5. References

1. Wolfenden, R.; Snider, M. J., *Accounts of chemical research* **2001,** *34* (12), 938-45.
2. Pauling, L., *Chemical and Engineering News* **1946,** *24* (10), 1375-1377.
3. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G., *Science* **2004,** *303* (5655), 186-95.
4. Cui, Q.; Karplus, M., *Journal of the American Chemical Society* **2001,** *123* (10), 2284-90.
5. Kast, P.; Grisostomi, C.; Chen, I. A.; Li, S.; Krengel, U.; Xue, Y.; Hilvert, D., *The Journal of biological chemistry* **2000,** *275* (47), 36832-8.
6. Guo, H.; Cui, Q.; Lipscomb, W. N.; Karplus, M., *Angewandte Chemie* **2003,** *42* (13), 1508-11.
7. Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J., *Journal of the American Chemical Society* **2004,** *126* (5), 1369-76.
8. Gao, J., *Curr Opin Struct Biol* **2003,** *13* (2), 184-92.
9. Freud, Y.; Ansbacher, T.; Major, D. T., *ACS catalysis* **2017,** *7* (11), 7653-7657.
10. Richard, J. P., *Biochemistry* **1991,** *30* (18), 4581-5.
11. Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E., *Journal of molecular biology* **2001,** *310* (1), 271-280.
12. Pompliano, D. L.; Peyman, A.; Knowles, J. R., *Biochemistry* **1990,** *29* (13), 3186-3194.
13. Alahuhta, M.; Wierenga, R. K., *Proteins-Structure Function and Bioinformatics* **2010,** *78* (8), 1878-1888.
14. Morell, M. K.; Paul, K.; Oshea, N. J.; Kane, H. J.; Andrews, T. J., *Journal of Biological Chemistry* **1994,** *269* (11), 8091-8098.
15. Go, M. K.; Koudelka, A.; Amyes, T. L.; Richard, J. P., *Biochemistry* **2010,** *49* (25), 5377-5389.
16. Cleland, W. W.; Andrews, T. J.; Gutteridge, S.; Hartman, F. C.; Lorimer, G. H., *Chemical reviews* **1998,** *98* (2), 549-561.
17. Tcherkez, G., *Plant Cell Environ* **2013,** *36* (9), 1586-1596.
18. Larimer, F. W.; Harpel, M. R.; Hartman, F. C., *Journal of Biological Chemistry* **1994,** *269* (24), 16984-16984.
19. Tcherkez, G., *Plant Cell Environ* **2016,** *39* (5), 983-997.
20. Firn, R., *Nature's Chemicals: The Natural Products That Shaped Our World* **2010**, 1-250.
21. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T., *ACS catalysis* **2017,** *7* (1), 812-818.
22. Major, D. T.; Freud, Y.; Weitman, M., *Curr Opin Chem Biol* **2014,** *21*, 25-33.
23. Starks, C. M.; Back, K. W.; Chappell, J.; Noel, J. P., *Science* **1997,** *277* (5333), 1815-1820.
24. Schrepfer, P.; Buettner, A.; Goerner, C.; Hertel, M.; van Rijn, J.; Wallrapp, F.; Eisenreich, W.; Sieber, V.; Kourist, R.; Bruck, T., *P Natl Acad Sci USA* **2016,** *113* (8), E958-E967.
25. Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; Mcphail, A. T., *Journal of the American Chemical Society* **1971,** *93* (9), 2325-&.
26. Jornvall, H.; Hedlund, J.; Bergman, T.; Oppermann, U.; Persson, B., *Biochemical and biophysical research communications* **2010,** *396* (1), 125-130.
27. Hedlund, J.; Jornvall, H.; Persson, B., *BMC Bioinformatics* **2010,** *11*, 534.
28. Reichel, L.; Kohle, H., *H-S Z Physiol Chem* **1935,** *236*, 158-167.
29. Negelein, E.; Wulff, H. J., *Biochem Z* **1937,** *289* (5/6), 436-437.
30. Erb, T. J.; Brecht, V.; Fuchs, G.; Muller, M.; Alber, B. E., *P Natl Acad Sci USA* **2009,** *106* (22), 8871-8876.
31. Parikh, S.; Moynihan, D. P.; Xiao, G.; Tonge, P. J., *Biochemistry* **1999,** *38* (41), 13623-34.
32. Massengo-Tiasse, R. P.; Cronan, J. E., *Cell Mol Life Sci* **2009,** *66* (9), 1507-1517.
33. Bull, H. G.; GarciaCalvo, M.; Andersson, S.; Baginsky, W. F.; Chan, H. K.; Ellsworth, D. E.; Miller, R. R.; Stearns, R. A.; Bakshi, R. K.; Rasmusson, G. H.; Tolman, R. L.; Myers, R. W.; Kozarich, J. W.; Harris, G. S., *Journal of the American Chemical Society* **1996,** *118* (10), 2359-2365.
34. Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat Chem Biol* **2014,** *10* (1), 50-U85.
35. Libby, R. D.; Mehl, R. A., *Bioorg Chem* **2012,** *40*, 57-66.

36. Sandor, R.; Slanina, J.; Midlik, A.; Sebrlova, K.; Novotna, L.; Carnecka, M.; Slaninova, I.; Taborsky, P.; Taborska, E.; Pes, O., *Phytochemistry* **2018,** *145*, 77-84.
37. Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L. C.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L., *Structure* **2015,** *23* (12), 2213-2223.
38. Keseler, I. M.; Mackie, A.; Santos-Zavaleta, A.; Billington, R.; Bonavides-Martinez, C.; Caspi, R.; Fulcher, C.; Gama-Castro, S.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Muniz-Rascado, L.; Ong, Q.; Paley, S.; Peralta-Gil, M.; Subhraveti, P.; Velazquez-Ramirez, D. A.; Weaver, D.; Collado-Vides, J.; Paulsen, I.; Karp, P. D., *Nucleic Acids Res* **2017,** *45* (D1), D543-D550.
39. Wheeldon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov, P.; Sigman, M., *Nat Chem* **2016,** *8* (4), 299-309.

CHAPTER II

The use of ene adducts to study and engineer enoyl-thioester reductases

Authors:
Raoul G. Rosenthal*, Bastian Vögeli*, Nick Quade, Guido Capitani, Patrick Kiefer, Julia A. Vorholt, Marc-Olivier Ebert, Tobias J. Erb

* These authors contributed equally to this work

Published in:
Nature Chemical Biology 11,398-400 (2015); DOI: 10.1038/nchembio.1794

Author contributions:
R.G.R., B.V. and T.J.E. conceived and performed all experiments. NMR experiments were designed with M.-O. E. and MS analyses with P.K. and J.A.V. B.V. prepared enzyme crystals, N.Q. and G.C. collected diffraction data and interpreted the results. R.G.R., B.V. and T.J.E. wrote the paper.

2. The use of ene adducts to study and engineer enoyl-thioester reductases

2.1. Abstract

An improved understanding of enzymes' catalytic proficiency and stereoselectivity would further enable applications in chemistry, biocatalysis and industrial biotechnology. We use a chemical probe to dissect individual catalytic steps of enoyl-thioester reductases (Etr), validating an active site tyrosine as the cryptic proton donor and explaining how it had eluded definitive identification. This information enabled the rational redesign of Etr, yielding mutants that create products with inverted stereochemistry at wild-type like turnover.

2.2. Introduction

Synthetic chemistry aims at achieving high regio- and stereoselectivity. Enzymes are Nature's solution to obtain both goals with unmatched catalytic efficiency. The demand for enzymes in green chemistry, biocatalysis and industrial biotechnology is reflected by the efforts that are put into creating novel enzyme activities by rational or *de novo* approaches[1-2]. Computational design is a powerful tool for designing enzyme catalysts *de novo*[3-4]. Yet, the *ab initio* design of cofactor-dependent enzyme reactions is still an unsolved problem as it involves poorly understood reactions that depend on the accurate alignment of multiple reactants and active site residues. Therefore new experimental tools are required that allow resolving catalysis beyond current limitations.

One of the evolutionary oldest and largest superfamilies of cofactor-dependent enzymes is the class of medium chain dehydrogenase reductases (MDRs) that contains ketoreductases and enoyl-thioester reductases (Etrs), which are essential in fatty acid as well as secondary metabolite biosynthesis[5]. Etrs catalyze the NAD(P)H-dependent reduction of enoyl-thioesters into acyl-thioesters. This reaction involves addition of a hydride from NAD(P)H to the β-carbon and a proton from an active site acid or water-molecule to the α-carbon of the enoyl-thioester. Several crystal structures of Etrs have been solved. Yet, the proton donor that determines reaction stereochemistry is still cryptic because no single knock-out of putative catalytic residues renders these enzymes inactive[6-11].

The obstacle to accurately study and manipulate the role of specific active site residues in Etrs was recently put in a new perspective by the suggestion[12] and experimental validation[13] that MDR family members could follow an ene-mechanism[14], which involves the formation of a transient covalent, NADPH-substrate adduct ('C2-ene adduct') for the transfer of the hydride equivalent (**Figure 1a**). Besides having mechanistic implications, the observation of such C2-ene adducts opens new options in experimental biochemistry. These adducts represent catalytically competent intermediates in which the first half of the reduction reaction *i.e.*, the transfer of a hydride equivalent has taken place, while protonation has not yet. Here, we sought to use the isolated C2-ene adduct to investigate the catalytic cycle of Etr1p, a well-studied model enzyme from yeast mitochondria. This adduct is accepted by Etr1p, albeit with a reduced k_{cat} and k_{cat}/K_m values compared to the natural substrates[13]. Despite lower catalytic parameters, we assume that the ene-adduct is close to - if not on - the reaction coordinate, as other steps might become rate limiting when using the C2-ene adduct as substrate (*e.g.* limited diffusion into the barrel shaped active site or NADPH binding related conformational changes[8]). Applying the C2-ene-adduct as molecular probe allowed us to resolve the catalytic cycle of Etr1p (and mutants) at unprecedented detail, assigning tyrosine 79 as the proton donor, and controlling the stereochemistry of reduction by rationally inverting the active site chemistry of the enzyme.

2.3. Results

2.3.1. Identification of the cryptic proton donor in Etr1p

Because several Etr structures are available, but none as ternary complex[6-8] we first solved the structure of yeast Etr1p in complex with crotonyl-CoA and $NADP^+$ (protein data base: 4WAS, **Table S1**). Inspection of the active site suggested a serine (S70), tyrosine (Y79), and threonine (T324) as possible proton donating residues (**Figure 1b**). These active site amino acids were individually mutated to yield Etr1p mutants S70V, Y79F, and T324A, which were tested for activity with the C2-ene adduct[13] of NADPH and crotonyl-CoA as molecular probe. Starting the enzyme reaction with the molecular probe, we expected only mutations that are affected in proton transfer to be severely impaired in activity (**Table S2**). Compared to Etr1p wild type (WT), Etr1p Y79F was affected by more than two orders of magnitude in conversion of the C2-ene adduct, indicating a direct role for this residue in proton donation. In contrast Etr1p S70V and T324A showed only minor losses in catalytic activities (**Table S2**), corresponding to a secondary role[15-16] of these active site residues in catalysis (*e.g.*, in transition state stabilization by hydrogen bonding).

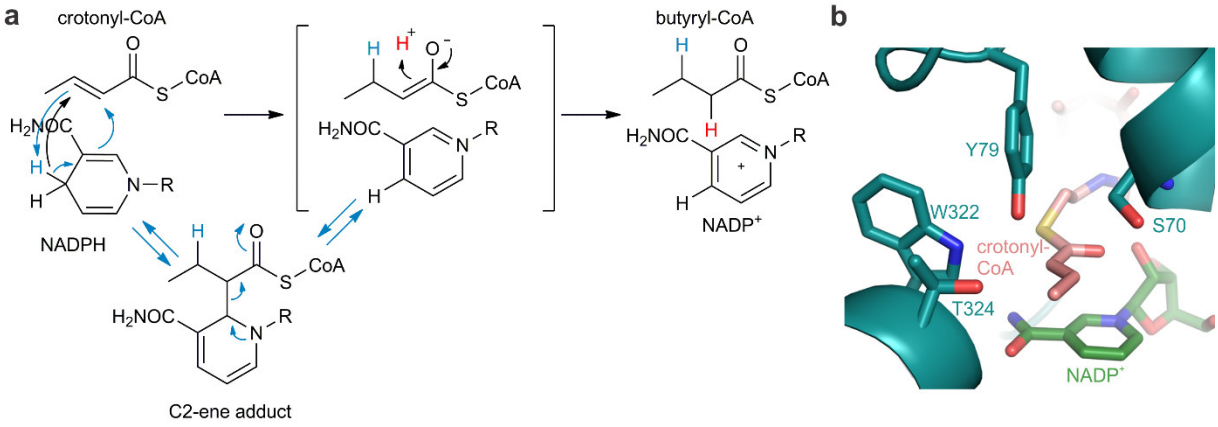


Figure 1. Proposed reaction mechanisms of enoyl-thioester reductases. (**a**) The classical mechanism for enoyl-thioester reductases (black arrows) assumes a direct hydride transfer from NAD(P)H onto the substrate to create an enolate intermediate that is subsequently protonated to yield the reaction products. The ene-mechanism[14] (blue arrows) proceeds through a distinct C2-ene intermediate that was experimentally verified recently[13]. This intermediate is further converted to the consensus enolate intermediate (in brackets) and subsequently protonated. (**b**) Active site of Etr1p WT showing potential proton donors S70, Y79 and T324.

2.3.2. A dead-end product is formed by Etr1p Y79F

In line with previous investigations using natural substrates[6, 8], but nevertheless surprising was the observation that Etr1p Y79F still catalyzed the consumption of the molecular probe even though it apparently lacked a key active site residue. However, in contrast to the WT enzyme that was able to consume the C2-ene adduct directly, as inferred from the rapid depletion of the molecular probe at λ_{max} = 370 nm, Etr1p Y79F formed a species that absorbed at λ_{max} = 328 nm, when incubated with the molecular probe (**Figure 2a, b, Figure S1**). This buildup indicated that Etr1p Y79F was not able to protonate the probe in the same fashion as the WT enzyme. The novel compound was purified from the reaction mixture and demonstrated to be a covalent C4/Cα-adduct (**1**) of NADPH and crotonyl-CoA (**Figure S2** and **S3**). In contrast to the C2-ene adduct that is a *bona fide* substrate for Etr1p WT [13], the isolated C4/Cα-ene adduct is not converted by the WT enzyme. Instead, the C4/Cα-adduct is a competitive inhibitor for NADPH in Etr1p WT with a K_i of 70 ± 45 nM and as a non-competitive inhibitor with respect to crotonyl-CoA, decaying slowly in solution with a first order constant of k = 0.02490 ±

0.00001 min^{-1} at 31 °C (**Figure S4**). In summary, these properties suggested that the C4/Cα-adduct represents a dead-end metabolite rather than a catalytic intermediate.

2.3.3. Time resolved analysis of Etr1p Y79F via NMR and stopped-flow spectroscopy

The detection of the C4/Cα-adduct in assays initiated with the molecular probe raised the question whether formation of this C4/Cα-adduct might be also of importance for the complete catalytic cycle of Etr1p Y79F. Therefore, we monitored the full reaction of the mutant enzyme by ^{1}H-NMR, starting the assay with the original substrates NADPH and crotonyl-CoA. Integration of protons specific to each compound showed a quick consumption of NADH (d, δ=5.57 ppm) and crotonyl-CoA (t, δ=1.49 ppm) and a build-up of the C2-ene adduct (d, δ=6.25 ppm). The C2-ene adduct was then further transformed into the C4/Cα-adduct (t, δ=0.42 ppm), which subsequently decomposed slowly into the final products butyryl-CoA (t, δ=2.18 ppm) and $NADP^{+}$ (d, δ= 8.71 ppm) (**Figure 2c,d**). Stopped flow measurements with Etr1p Y79F on the natural substrates showed an increase in absorption between 370-380 nm during the first 50 ms that matched the decrease in NADPH absorbance, confirming that only the C2-ene adduct is initially formed (**Figure S4**). Taken together the results show that Y79 is needed for direct conversion of the substrates to the products avoiding buildup of C4/Cα-adduct.

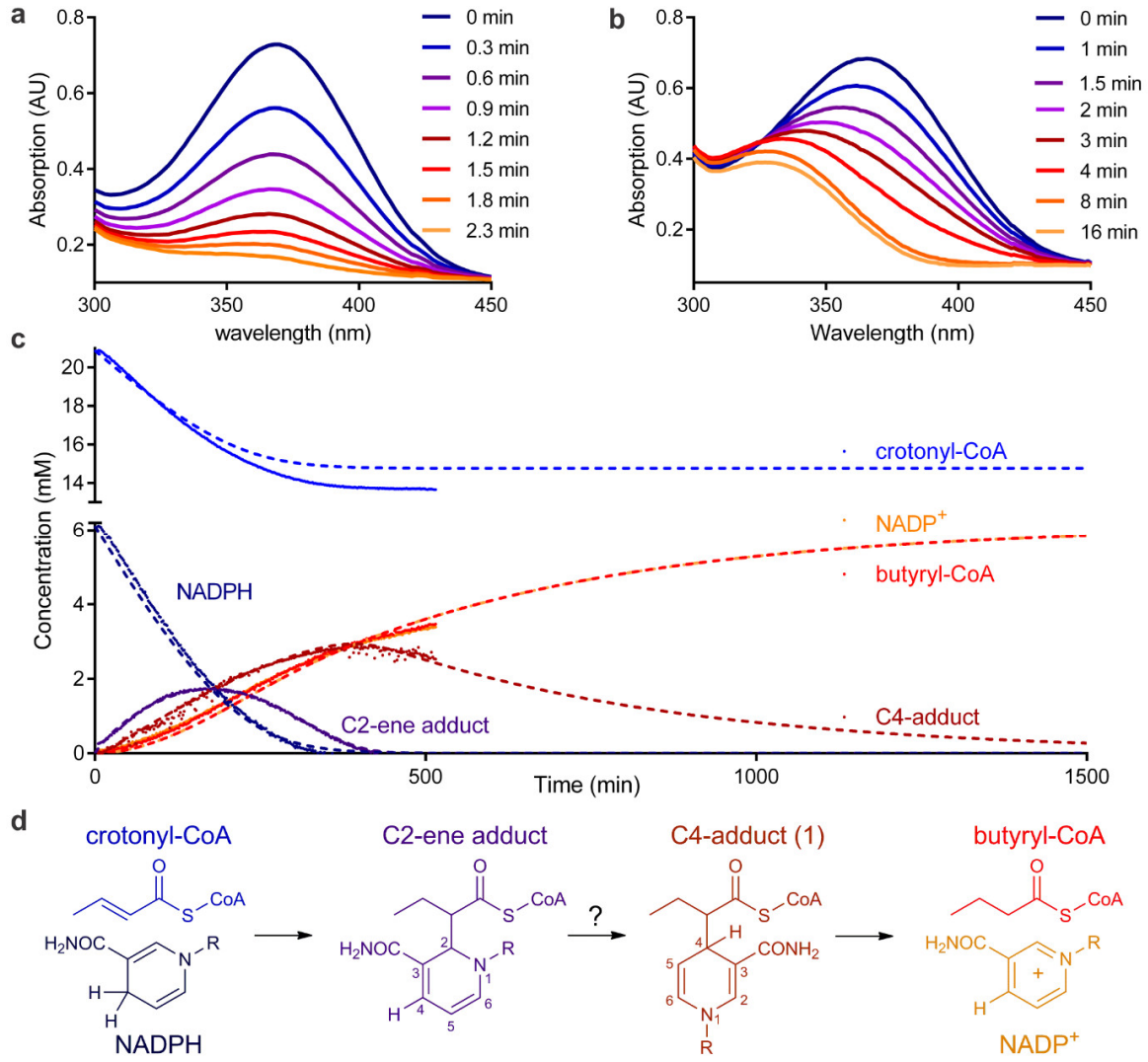


Figure 2. Identification and characterization of the proton donation deficient Etr1p Y79F mutant. (**a**) Scanning kinetic assay of Etr1p WT using the C2-ene adduct as a molecular probe. Rapid depletion of the molecular probe is observed. (**b**) Scanning kinetic assay of Etr1p Y79F with C2-ene adduct. Formation of a novel photoactive species (C4/Cα-adduct) with an absorption maximum at 328 nm is observed. (**c**) Time course assay of Etr1p Y79F. The characteristic peaks from 1H-NMR spectra were

integrated, to follow the kinetics of the substrates NADPH (dark blue) and crotonyl-CoA (light blue), the labile C2 intermediate species (purple) and the C4/Cα-adduct (dark red) as well as the products butyryl-CoA (red) and $NADP^+$ (orange). The dotted lines represent a fit of the data, collected for 500 min, modeled according to reactions (1)-(4) and simulated to 1,500 min. (**d**) The proposed reaction scheme for Etr1p Y79F with all the compounds in colors corresponding to the time course experiment.

2.3.4. Using the ene adduct to screen for new proton donors with inverted stereochemistry

With this new understanding of Y79's role, we sought to rationally control the stereochemistry of protonation in Etrs. This long-standing challenge in biocatalysis is of particular importance for Etrs that are responsible for determining chiral centers in polyketide biosynthesis[9, 17-19]. As a proof-of-principle we aimed at re-designing the last step of Etr1p catalysis by changing the stereochemistry of protonation from *re*- to *si*-face. To prevent competition for substrate protonation from the *re*-face, we started with the Y79F mutant and sought to introduce additional residues that could serve as *si*-face proton donor. Inspection of Etr1p´s active site suggested position 322 (see **Figure 1**) as good candidate for placing the new proton donor. When the corresponding double mutants were screened with the molecular probe, mutants Y79F/W322D, Y79F/W322H, and Y79F/W322K showed only unproductive conversion of the molecular probe into C4/Cα-adduct (**Figure 3a**, **Figure S5**). In contrast, Y79F/W322E was able to directly consume the C2-ene adduct like the WT enzyme suggesting this variant contained a competent *si*-face proton donor (**Figure 3b**).

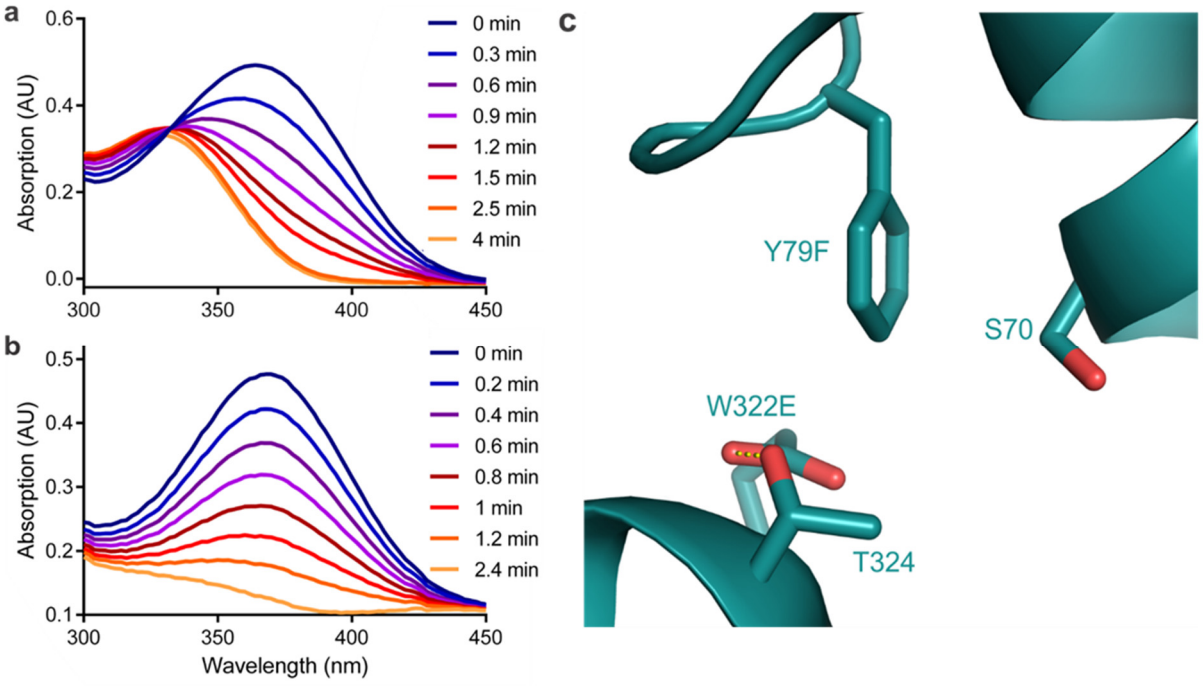


Figure 3. Screening of Etr1p mutants for *si*-face proton donors using the C2-ene adduct as a molecular probe. (**a**) Scanning kinetic assay of Etr1p Y79F W322K. Formation of C4/Cα-adduct shows that lysine does not serve as a proton donor. (**b**) Scanning kinetic assay of Etr1p Y79F W322E. The mutant behaves similar to Etr1p WT suggesting that the glutamate functions as novel proton donor. (**c**) Active site structure of Etr1p Y79F W322E site showing the mutated WT proton donor Y79 and the catalytic residue E322 with a hydrogen bond to the T324 OH-group (yellow) that positions E322 correctly for proton donation.

The successful engineering of a pro-*2S*-specific enoyl-CoA reductase was confirmed by isotopic labeling. Etr1p Y79F/W322D incorporated a deuteron with an efficiency of 97 ± 2% into the *2S*-position compared to 2 ± 2% for the WT (for Etr1p Y79F and W322E see **Table S3**). The Y79F/W322E crystal structure (protein data base: 4W99, **Table S1**) showed that E322 is positioned in proximity to the substrate by hydrogen bonding to T324 (**Figure 3c**) and oriented in a fashion that allows for a stereo-selective *si*-face proton transfer either directly or through a water molecule. The pH optimum of Etr1p Y79F/W322E was lower compared to Etr1p WT, most likely due to the lower pK_a of the engineered *si*-

glutamate compared to the *re*-tyrosine of the WT enzyme (**Figure S6**). At pH 5.0 the catalytic efficiency of Etr1p Y79F/W322E was $2.7 \pm 0.3 \cdot 10^5$ s^{-1} M^{-1}, corresponding to 3.3% of the WT (**Table S2**). Given the fact that our design was solely based on the newly obtained insights into the biochemistry of Etrs and did not involve directed evolution, the high catalytic efficiency achieved is remarkable.

2.4. Discussion

Summarizing, while our results revealed a surprising complexity of events in the Etr1p mutant reaction, they also provide the first real explanation for the enigmatic residual activity that has been reported to persist in Etr1p and other NAD(P)H-dependent reductases in the absence of the putative proton donor [6, 8-11, 20]. First, and in line with an ene-mechanism[13], these enzymes are still able to catalyze the initial transfer of a hydride equivalent, as shown by formation of the C2-ene adduct. However, because the subsequent enzymatic resolution of the C2-ene adduct through protonation is prevented, this leads to the formation of an unproductive, dead-end C4/Cα-adduct which decays non-enzymatically into butyryl-CoA and $NADP^+$. Because of overlapping spectral information of the different transient adducts formed, which cannot be resolved by standard UV-Vis assays, the chemical complexity of these (mutagenized) enzyme reactions have gone unnoticed so far. Even more seriously, the change of absorbance recorded in these widely used assays has been misinterpreted as simple Michaelis-Menten kinetics of the NAD(P)H chromophore, although it actually represents the sum of a complex series of enzymatically (and non-enzymatically) transformations. This has masked identification of a key active site residue. Our molecular probing approach represents a convenient way to distinguish between slow productive and unproductive protonation mutants. We suggest that future mutagenesis studies on Etrs and possibly other NAD(P)H-dependent enzymes should make use of a similar probing approach as chosen here to untangle the true catalytic events of these enzymes. Translation of our engineering strategy to enzymes involved in polyketide biosynthesis could lay the foundation for the (bio)synthetic production and directed manipulation of complex chemical compounds.

Lastly, our results also have implications for understanding the evolution of novel enzymatic functions. The observation that hydride equivalent transfer and protonation can be separated by transient substrate-NAD(P)H adducts, as shown here for Etr1p by molecular probing, demonstrates an implicit modularity of catalysis that might facilitate NAD(P)H-dependent enzymes to diverge in both substrate- and reaction-specificity. If novel substrates can be accommodated at the active site of enzymes that allow for formation of such adducts, reduction can take place through non-enzymatic decay (for the proposed mechanism, see **Figure S7**). Evolutionary pressure could then select for variants with a better proton donor or completely different electophiles. Notable examples for such scenarios might be the evolution of *2S*-stereoselectivity in methylenoyl-thioester reductase domains of polyketide synthases[9] and FAS I megasynthases[21] from their *2R*-specific ancestors, or the emergence of the CO_2-fixation function in the scaffold of ordinary reductases[22-23].

2.5. Methods

2.5.1. Chemicals

Butyryl-CoA and $NADP^+$ and crotonic anhydride were purchased from Sigma Aldrich AG, coenzyme A from Roche Diagnostics. Crotonyl-CoA and C2-ene adduct were synthesized and purified according to [13]. All salts and solvents were of analytical grade.

2.5.2. Cloning and mutagenesis

The gene encoding for enoyl thioester reductase (*etr1p*) was codon optimized for expression in *Methylobacterium extorquens* and synthesized by Eurofins MWG Operon (Ebersberg, Germany), see below for codon optimized sequence. An NdeI restriction site followed by an N-terminal His_{10}-tag was added in front of the start codon and an EcoRI site after the stop codon. The synthesized gene was cloned into the expression vector pT7-7 [24] yielding pTE260. The plasmids carrying desired point mutations were generated using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA) with 60 ng of template plasmid and the primers listed in **Table S4**. The resulting mutated plasmids were confirmed by sequencing (Microsynth AG, Balgach, Switzerland). Plasmids carrying the correct mutations were transformed into *E.coli* BL21(DE3) for protein expression.

DNA sequence of Etr1p WT, His10-tagged and codon optimized for expression in *Methylobacterium extorquens*:

ATGGGCCACCACCATCATCACCACCATCACCACCACTCCTCGGGCCACATCGAAGGCCGGCACATGATCACGGCCCAAGCCGTCCTCTACACGCAGCATGG
CGAGCCCAAGGACGTGCTGTTCACGCAGTCCTTCGAGATCGACGACGACAACCTGGCGCCGAACGAGGTGATCGTCAAGACCCTGGGCTCGCCGGTCAAC
CCGAGCGACATCAACCAGATCCAGGGCGTCTATCCGAGCAAGCCGGCCAAGACCACCGGGTTCGGCACCACCGAGCCCGCGGCGCCGTGCGGCAACGAG
GGCCTCTTCGAGGTGATCAAGGTCGGCTCGAACGTGAGCTCGCTCGAAGCCGGGGACTGGGTGATCCCCTCGCACGTGAACTTCGGCACCTGGCGCACGC
ACGCCCTGGGGAACGACGACGACTTCATCAAGCTCCCCAATCCCGCCCAGTCGAAGGCCAATGGCAAGCCGAACGGCCTCACCATCAACCAGGGCGCCAC
CATCTCGGTCAATCCCCTGACGGCCTACCTCATGCTCACCCACTACGTCAAGCTGACGCCGGGCAAGGACTGGTTCATCCAGAACGGCGGCACCAGCGCGG
TGGGCAAGTACGCCAGCCAGATCGGCAAGCTCCTCAACTTCAACTCGATCAGCGTGATCCGCGATCGCCCGAATCTCGATGAGGTCGTCGCCTCGCTGAAG
GAGCTGGGCGCCACCCAGGTCATCACCGAGGACCAGAACAACTCGCGGGAGTTCGGCCCGACCATCAAGGAGTGGATCAAGCAGTCGGGCGGCGAAGC
GAAGCTGGCGCTGAACTGCGTGGGCGGCAAGTCGTCCACGGGGATCGCCCGCAAGCTGAACAACAACGGGCTCATGCTCACGTATGGCGGCATGAGCTT
CCAGCCGGTCACCATCCCGACCTCCCTGTACATCTTCAAGAACTTCACCTCCGCGGGCTTCTGGGTGACCGAGCTGCTCAAGAACAACAAGGAGCTCAAGA
CCTCCACGCTGAACCAGATCATCGCCTGGTATGAGGAGGGCAAGCTCACCGATGCGAAGTCGATCGAAACCCTCTACGACGGCACCAAGCCCCTCCACGA
GCTCTACCAGGATGGCGTCGCGAACTCCAAGGATGGCAAGCAGCTCATCACGTACTGA

2.5.3. Protein expression

A single colony of *E. coli* BL21 (DE3) bearing the correct plasmid (**Table S5**) was grown over night in LB medium containing 100 μg/ml ampicillin at 37 °C. Of this overnight culture 1 ml was used to inoculate 1 l of auto-inducing medium [25] containing 100 μg/ml ampicillin. Cells were grown for 24 h at 30 °C under constant shaking before harvesting by centrifugation (10 min at 8000 x *g*, 4 °C). Harvested cells were resuspended in 15 mL of buffer A (20 mM Tris(hydroxymethyl)aminomethane (Tris-HCl), pH 7.8, 200 mM KCl) containing 0.1 mg/ml of DNase I (*Bovine* pancreas, AppliChem) and 10% glycerol. The solution was sonicated 6 times for 20 s with 40 s cooling on ice in between cycles. The lysate was centrifuged for 1 h at 40,000 x g, 4 °C. The supernatant was applied at a flow rate of 1 ml/min onto an equilibrated HisTrap FF (GE Healthcare, Little Chalfont, UK) and washed with 30 column volumes of buffer A containing 75 mM imidazole. The protein was then eluted with buffer A containing 500 mM imidazole, desalted with a PD-10 desalting column (GE Healthcare, Little Chalfont, UK) and concentrated by ultrafiltration (Amicon® Ultra-4 30 K centrifugal filters, Merk Millipore). For NMR experiments the protein was exchanged into a buffer containing 100 mM Na_2HPO_4 pH 7.9, 200 mM NaCl, stored on ice and used the same day. For crystallization the protein was stored in 20 mM TrisHCl pH 7.9, 100 mM NaCl at 4 °C and for spectrophotometric assays the protein was stored at -20 °C in 50 mM Na_2HPO_4 pH 7.9, 100 mM NaCl with 50% glycerol. Protein concentrations were determined

spectrophotometrically ($\varepsilon_{280\ nm}$ = 49.5 cm^{-1} mM^{-1}). Mutations were confirmed additionally by trypsin digestion and subsequent peptide sequencing at the functional genomics center, Zurich

2.5.4. Spectrophotometric enzyme assays

Assays were carried out on a Carry-50 UV/Vis spectrometer (Agilent) at 30 °C using quartz cuvettes (1-, 3-, or 10-mm path-length; Hellma). All assays were carried out in 100 mM Na_2HPO_4, pH 7.9. For determination of the kinetic parameters starting from NADPH and crotonyl-CoA assays contained 200 μM NADPH and were started by adding enzyme to the following end concentrations: 3.3 nM Etr1p WT, 39 nM Etr1p S70V, 11 nM Etr1p T324A or 6.8 μM Etr1p Y79F. Kinetic parameters were determined by varying the crotonyl-CoA concentration and following disappearance of NADPH at 340 nm using an absorption coefficient of $\varepsilon_{340\ nm}$ = 6.2 cm^{-1} mM^{-1} and for Etr1p Y79F $\Delta\varepsilon_{340}$ nm = 2.5 cm^{-1} mM^{-1} ($\varepsilon_{340\ nm\ NADPH}$ = 6.2 cm^{-1} mM^{-1}, $\Delta\varepsilon_{340\ nm\ C4}$ = 3.7 cm^{-1} mM^{-1} was used. Assays for determination of kinetic parameters on C2-ene adduct were started by adding varying concentrations of C2-ene adduct and following its disappearance at 370 nm using an extinction coefficient of $\varepsilon_{370\ nm}$ = 6.9 cm^{-1} mM^{-1} [13].

The screen for protonation of C2-ene adduct was done by monitoring the kinetics between 300 and 500 nm adding 0.033 μM Etr1p WT, 34 μM Etr1p Y79F, 9.3 μM Etr1p Y79F W322K, 3.4μM Etr1p Y79F W322D or 0.62 μM Etr1p Y79F W322E to start the reaction. For these assays the C2-ene adduct was dissolved directly into the assay buffer immediately before the assay was started.

For the inhibition assay C4-adduct determined using $\Delta\varepsilon_{340\ nm\ C4}$ = 3.7 cm^{-1} mM^{-1} was added to an end concentration of 5 nM to a reaction mixture containing 125 μM crotonyl-CoA and varying concentrations of NADPH or to assays containing 100 μM NADPH and varying the crotonyl-CoA concentration. The reactions were started by adding 3.3 nM Etr1p WT.

2.5.5. Purification and characterization of C4/Cα -adduct

A reaction mixture of 30 mM crotonyl-CoA, 30 mM NADPH and 0.17 mM Etr1p Y79F in 100 mM Na_2HPO_4 (pH 7.9, 100 μl) was reacted at 30 °C for 1.5 h. The mixture was separated on a Waters 2690 HPLC system on a Gemini 10 μm NX-C18 110 Å, 100 x 21.2 mm, AxiA packed column (Phenomenex) using a previously described elusion profile [13]. The C4-adduct was collected directly into liquid nitrogen, lyophilized (0.01 mbar, -70 °C) and stored at -80 °C. C4**/**Cα adduct was dissolved in 100 mM Na_2HPO_4 buffer and its absorption spectrum was measured between 200 and 400 nm (**Figure S1**). The extinction coefficients at 328 nm and 340 nm were calculated to be 4.1 ± 0.2 $cm^{-1}mM^{-1}$ and 3.7 ± 0.2 $cm^{-1}mM^{-1}$ assuming a total absorbance of 33.8 $cm^{-1}mM^{-1}$ at 260 nm ([ε_{260} nm NADPH + ε_{260} nm CoA] = [16.9 cm^{-1} mM^{-1} + 16.9 cm^{-1} mM^{-1}] [26]). For structural analysis by NMR the purified C4/Cα adduct was lyophillized three times. 2-3 mg compound were dissolved in 540 μl D_2O buffered with 60 μl > 90% deuterated 500 mM NaH_2PO_4 pD 7.9. NMR spectra were acquired at 4.4 ˚C.

2.5.6. HPLC-ESI-MS analysis

HPLC-ESI-MS analysis of the C4/Cα -adduct was performed exactly as described earlier for the C2-ene adduct [13]

2.5.7. Crystallization of Etr1p WT and Y79F/W322E

Crystallization was performed similarly to established protocols [27]. In short, first crystals were obtained by sitting drop vapor diffusion using 15 mg/mL enzyme in crystallization buffer mixed in a 1:1 ratio with reservoir solution containing 0.1 M N-(2-acetamido)-2-iminodiacetic acid (ADA)/NaOH (pH 6.3), 2 M $(NH_4)_2SO_4$, 20 mM crotonyl-CoA and 10 mM $NADP^+$ at 19 °C. Large crystals were obtained by

streak seeding 20 h after setting up crystallization plates and they reached their final size after a few weeks. For cryostabilisation a cryosolution containing 0.1 M ADA pH 6.3, 2.4 M $(NH_4)_2SO_4$ and 25% glycerol was added and the crystal immediately frozen in liquid nitrogen.

2.5.8. Structural data collection and refinement

The crystals were measured at beamline X06SA at the Swiss light source, SLS, at a wavelength of 1Å in a cryo stream.. The data processing was carried out using XDS [28]and scaling was done with XSCALE. A set of randomly chosen reflections were set aside for the calculation of the free R factor (Rfree). To obtain initial phases molecular replacement by Phaser [29] was performed using the structure of Etr1p containing $NADP^+$ (1N9G) as a search model, finding two and three monomers in the asymmetric unit for the double mutant and the wild type complex, respectively. The structures were refined with REFMAC5 [30] and Phenix (wt complex) and completed in Coot [31]. For the double mutant, clear difference density for the mutated side chains was visible. Water molecules were added and edited manually as well as using the water find tool in Coot with a cutoff of 1 sigma within hydrogen bonding distance to protein residues. A final TLS refinement step using five TLS groups per monomer as determined by the TLS Motion Determination Server [32] was carried out with REFMAC5 for the double mutant. Both structures were refined to good geometry with more than 97% of residues in the favored region of the Ramachandran plot in both cases as well as 0.3% outliers for the double mutant structure and no outliers for the wild type structure.

2.5.9. Stereo-specific incorporation of solvent protons at Cα by Etr1p WT and mutants

HPLC-ESI-MS analysis was performed as described previously [33]. For the isotope labeling experiment the storage buffer of Etr1p WT, Etr1p Y79F , Etr1p W322E and Etr1p Y79F W322E was exchanged by three subsequent ultracentrifugation steps (Amicon® Ultra-4 30 K centrifugal filters, Merck Millipore) diluting the original buffer over 10,000 x with deuterated (98%) 50 mM Na_2DPO_4, pD 7.9. 100 µl assays containing 210 µM NADPH and 200 µM crotonyl-CoA in 100 mM NaD2PO4 buffer (pD 7.9) were started with 0.07 µM Etr1p WT, 0.1 µM Etr1p W322E, 0.5 µM Etr1p Y79F W322E or 5.4 µM Etr1p Y79F. The reactions were followed spectrophotometrically at 340 nm and run at 30 °C until complete consumption of NADPH and decay of formed C4-adduct, in case of Etr1p Y79F, occurred. butyryl-CoA was purified from the reaction mixture using the above-mentioned HPLC protocol, lyophilized (0.01 mbar, -70 °C) and dissolved in 100 µl 50 mM Na_2HPO_4 (pH 7.9). Label incorporation was checked by HPLC-ESI-MS. The samples were then incubated for 1 h at 37 °C in a reaction mixture containing 2 mM ferrocenium hexafluorophosphate (Sigma Aldrich), 2 µM FAD (Sigma Aldrich) and 2 µM H. sapiens SCAD (expressed and purified from SGC clone 2VIG, IMAGE:4842286, according to the provided protocols http://www.thesgc.org/). The reaction was quenched by adding 1 µl of 50% formic acid, centrifuged for 3 min at 16,000 x g and the supernatant analyzed by HPLC-ESI-MS.

2.5.10. NMR assays and data fitting

NMR analysis was done on a Bruker 600 MHz spectrometer at 4.4 °C. For NMR reaction kinetics of Etr1p Y79F a previously described method was used [13]. Water suppression using excitation sculpting was set on a 570 µl mixture containing 50 mM Na_2HPO_4 pH 7.9, 5% D_2O, 10 mM crotonyl-CoA and 5 mM NADPH. 30 µl of a freshly purified 8.16 mg/mL Etr1p Y79F solution was added and ^{1}H-NMR spectra recorded every 3 min for 500 min. For structural characterization of isolated C4-adduct, 5 mg of purified lyophilized C4-adduct was dissolved in 600 µl D_2O to acquire ^{1}H-NMR, COSY, HSQC, NOE and HMBC-spectra. Characteristic peaks of all reaction components were integrated and plotted against time. These data points were fitted to reactions 1-4 of which equations 1 and 2 were fitted according to enzyme catalyzed Michaelis Menten models and 3 and 4 according to uncatalyzed decay.

The uncatalyzed decay of the C2-ene adduct was described previously[13]. Fitting was performed in COPASI [34].

$$crotonyl - CoA + NADPH \xrightarrow{Etr1pY79F} C2 - ene\ adduct \quad (1)$$

$$C2 - ene\ adduct \xrightarrow{Etr1pY79F} C4 / C\alpha - adduct \quad (2)$$

$$C2 - ene\ adduct \xrightarrow{uncat.} butyryl - CoA + NADPH \quad (3)$$

$$C4 / C\alpha - adduct \xrightarrow{uncat.} butyryl - CoA + NADPH \quad (4)$$

2.5.11. Stopped flow spectroscopy

Measurements were recorded on a thermostated stopped flow unit (SX17MN, Applied Photophysics, Leatherhead, UK) and detected by a 32-channel 0.2 ms photomultiplier array (Hamamatsu Photonics, SZK, Japan). Syringe 1 contained 100 μM NADPH and 100 μM crotonyl-CoA (left out in control experiment) in 100 mM Na_2HPO_4 buffer pH 7.9 and syringe 2 contained 40 μM of Etr1p Y79F in the same buffer at 5 °C. Data was collected every 0.5 ms for 1 s and the every 5 ms for a total of 15 s covering a range between 300 and 420 nm.

2.6. References

1. Kiss, G.; Celebi-Olcum, N.; Moretti, R.; Baker, D.; Houk, K. N., *Angewandte Chemie* **2013,** *52* (22), 5700-25.
2. Baker, D., *Protein science : a publication of the Protein Society* **2010,** *19* (10), 1817-9.
3. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., *Science* **2008,** *319* (5868), 1387-1391.
4. Siegel, J. B.; Zanghellini, A.; Lovick, H. M.; Kiss, G.; Lambert, A. R.; St Clair, J. L.; Gallaher, J. L.; Hilvert, D.; Gelb, M. H.; Stoddard, B. L.; Houk, K. N.; Michael, F. E.; Baker, D., *Science* **2010,** *329* (5989), 309-13.
5. Hedlund, J.; Jornvall, H.; Persson, B., *BMC bioinformatics* **2010,** *11*, 534.
6. Chen, Z. J.; Pudas, R.; Sharma, S.; Smart, O. S.; Juffer, A. H.; Hiltunen, J. K.; Wierenga, R. K.; Haapalainen, A. M., *J Mol Biol* **2008,** *379* (4), 830-44.
7. Zheng, J. T.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T., *Nature Chemical Biology* **2012,** *8* (7), 615-621.
8. Airenne, T. T.; Torkko, J. M.; Van den Plas, S.; Sormunen, R. T.; Kastaniotis, A. J.; Wierenga, R. K.; Hiltunen, J. K., *J Mol Biol* **2003,** *327* (1), 47-59.
9. Kwan, D. H.; Leadlay, P. F., *ACS chemical biology* **2010,** *5* (9), 829-38.
10. Mesa, J.; Alsina, C.; Oppermann, U.; Pares, X.; Farres, J.; Porte, S., *Chemico-biological interactions* **2015**.
11. Wu, Y. H.; Ko, T. P.; Guo, R. T.; Hu, S. M.; Chuang, L. M.; Wang, A. H., *Structure* **2008,** *16* (11), 1714-23.
12. Libby, R. D.; Mehl, R. A., *Bioorganic chemistry* **2012,** *40* (1), 57-66.
13. Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat Chem Biol* **2014,** *10* (1), 50-5.
14. Hamilton, G. A., *Progress in Bioorganic Chemistry* **1971,** *1*, 83-157.
15. Bryan, P.; Pantoliano, M. W.; Quill, S. G.; Hsiao, H. Y.; Poulos, T., *Proc Natl Acad Sci U S A* **1986,** *83* (11), 3743-5.
16. Fersht, A. R.; Shi, J. P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M.; Winter, G., *Nature* **1985,** *314* (6008), 235-8.
17. Kwan, D. H.; Sun, Y.; Schulz, F.; Hong, H.; Popovic, B.; Sim-Stark, J. C.; Haydock, S. F.; Leadlay, P. F., *Chemistry & biology* **2008,** *15* (11), 1231-40.
18. Toogood, H. S.; Scrutton, N. S., *Current opinion in chemical biology* **2014,** *19*, 107-15.
19. Zheng, J. T.; Piasecki, S. K.; Keatinge-Clay, A. T., *ACS chemical biology* **2013,** *8* (9), 1964-1971.
20. Rafi, S.; Novichenok, P.; Kolappan, S.; Zhang, X.; Stratton, C. F.; Rawat, R.; Kisker, C.; Simmerling, C.; Tonge, P. J., *The Journal of biological chemistry* **2006,** *281* (51), 39285-93.
21. Maier, T.; Leibundgut, M.; Ban, N., *Science* **2008,** *321* (5894), 1315-22.
22. Erb, T. J.; Berg, I. A.; Brecht, V.; Muller, M.; Fuchs, G.; Alber, B. E., *Proc Natl Acad Sci U S A* **2007,** *104* (25), 10631-6.
23. Erb, T. J.; Brecht, V.; Fuchs, G.; Muller, M.; Alber, B. E., *P Natl Acad Sci USA* **2009,** *106* (22), 8871-8876.
24. Tabor, S.; Richardson, C. C., *Proc Natl Acad Sci U S A* **1985,** *82* (4), 1074-8.
25. Studier, F. W., *Protein Expres Purif* **2005,** *41* (1), 207-234.
26. Dawson, R. M. C., *Data for biochemical research*. 3rd ed.; Clarendon Press: Oxford, 1986; p xii, 580 p.
27. Torkko, J. M.; Koivuranta, K. T.; Kastaniotis, A. J.; Airenne, T. T.; Glumoff, T.; Ilves, M.; Hartig, A.; Gurvitz, A.; Hiltunen, J. K., *The Journal of biological chemistry* **2003,** *278* (42), 41213-20.
28. Kabsch, W., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 2), 125-32.
29. Mccoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., *J Appl Crystallogr* **2007,** *40*, 658-674.

30. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta Crystallogr D* **1997,** *53*, 240-255.
31. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta Crystallogr D* **2010,** *66*, 486-501.
32. Painter, J.; Merritt, E. A., *Acta Crystallogr D* **2006,** *62*, 439-450.
33. Peyraud, R.; Kiefer, P.; Christen, P.; Massou, S.; Portais, J. C.; Vorholt, J. A., *P Natl Acad Sci USA* **2009,** *106* (12), 4846-4851.
34. Hoops, S.; Sahle, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Singhal, M.; Xu, L.; Mendes, P.; Kummer, U., *Bioinformatics* **2006,** *22* (24), 3067-74.

2.7. Supplementary Information

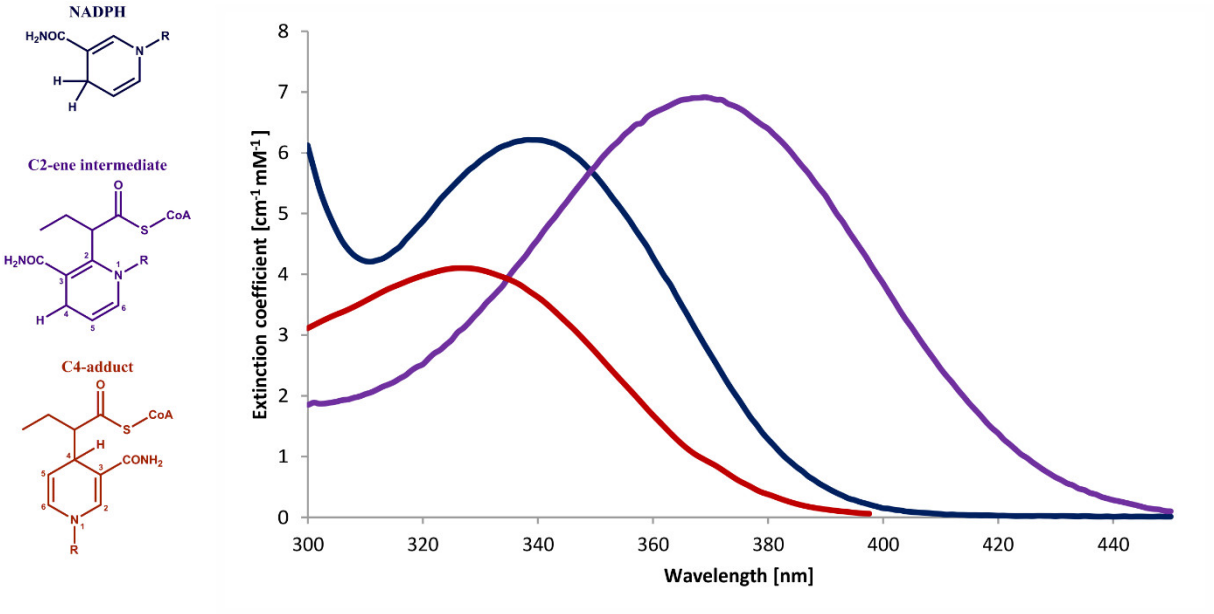


Figure S1. UV-Vis characterization of the C2-ene adduct, the C4/Cα-adduct and NADPH. Absorption spectra of NADPH, C2-ene adduct and the C4/Cα-adduct. NADPH (dark blue) has a local maximum at 340 nm, C2-ene adduct (purple) at 370 nm and C4/Cα-adduct (red) at 328 nm.

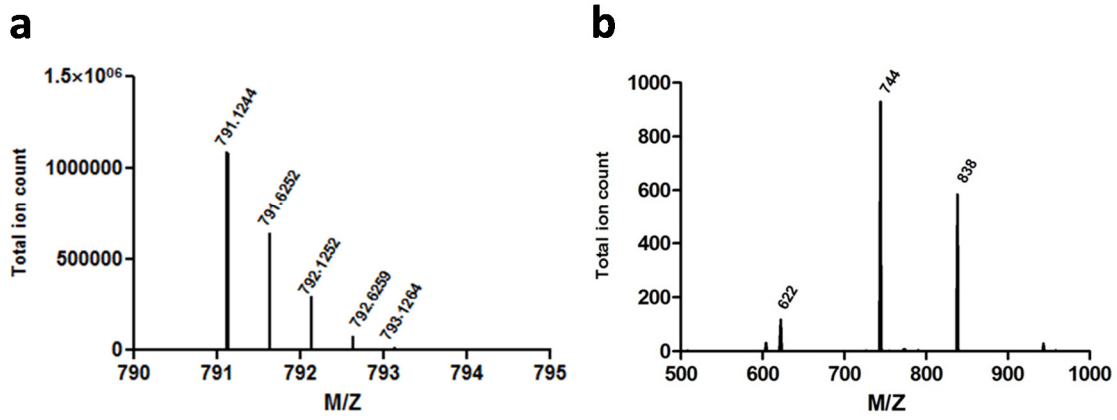
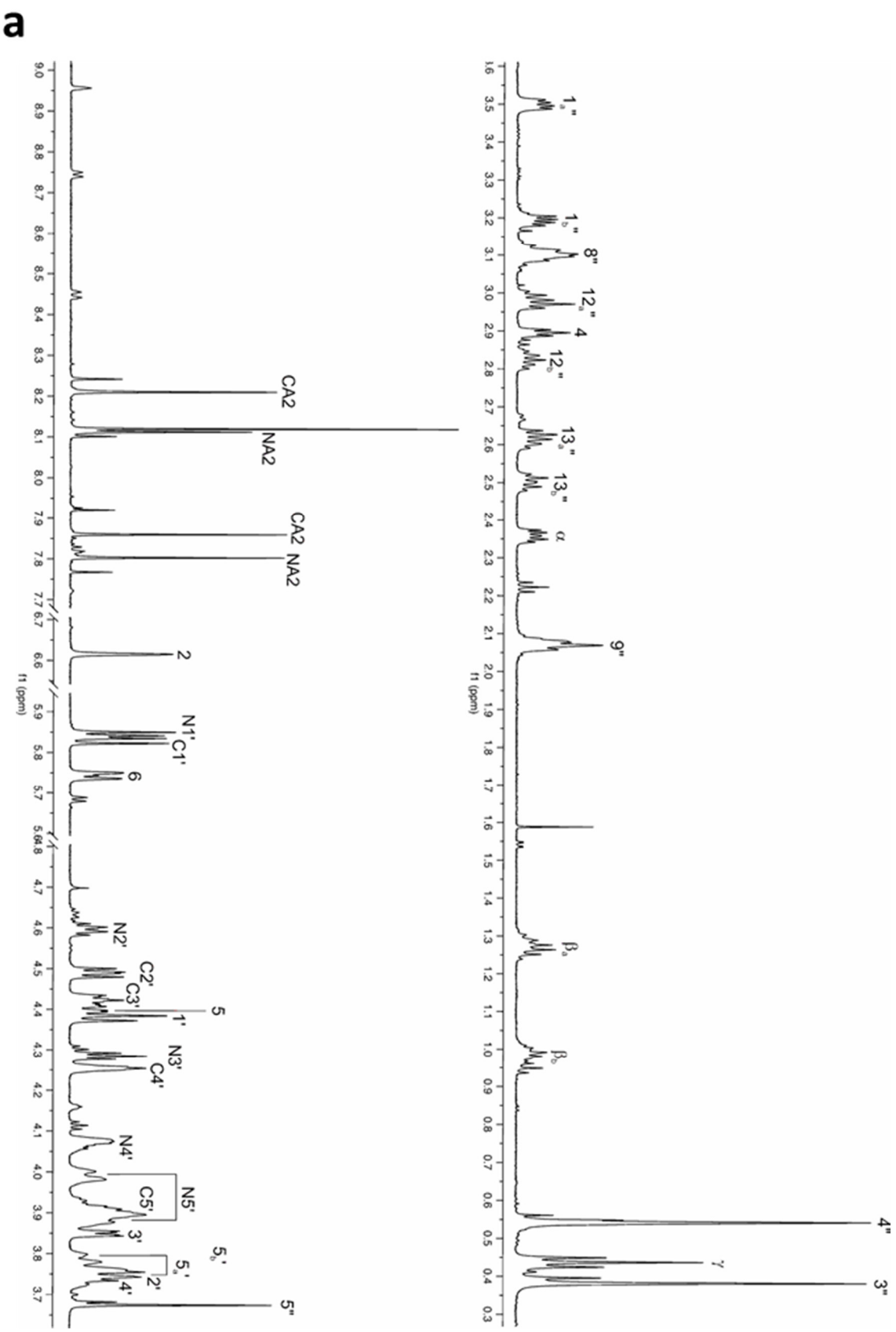
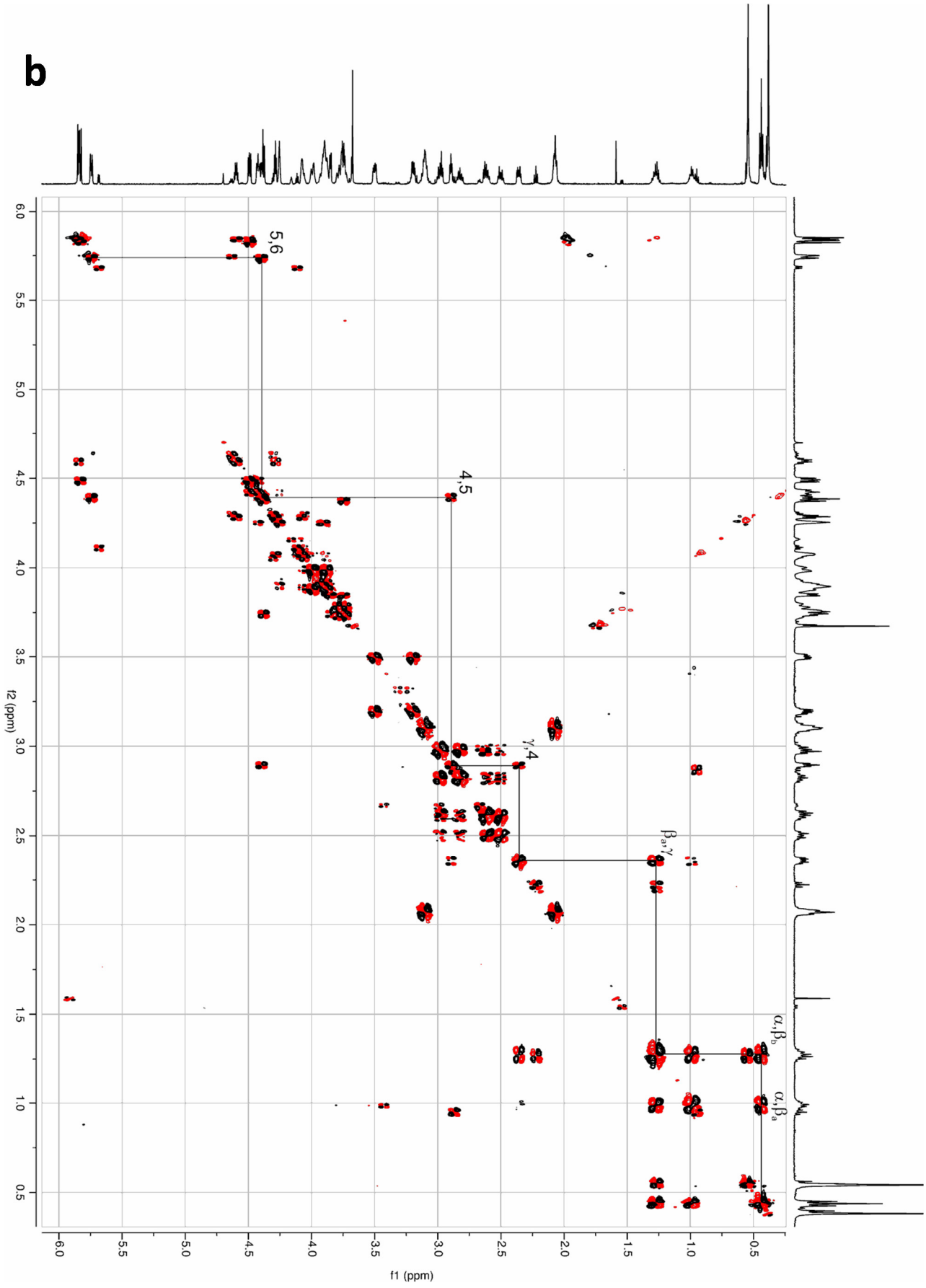


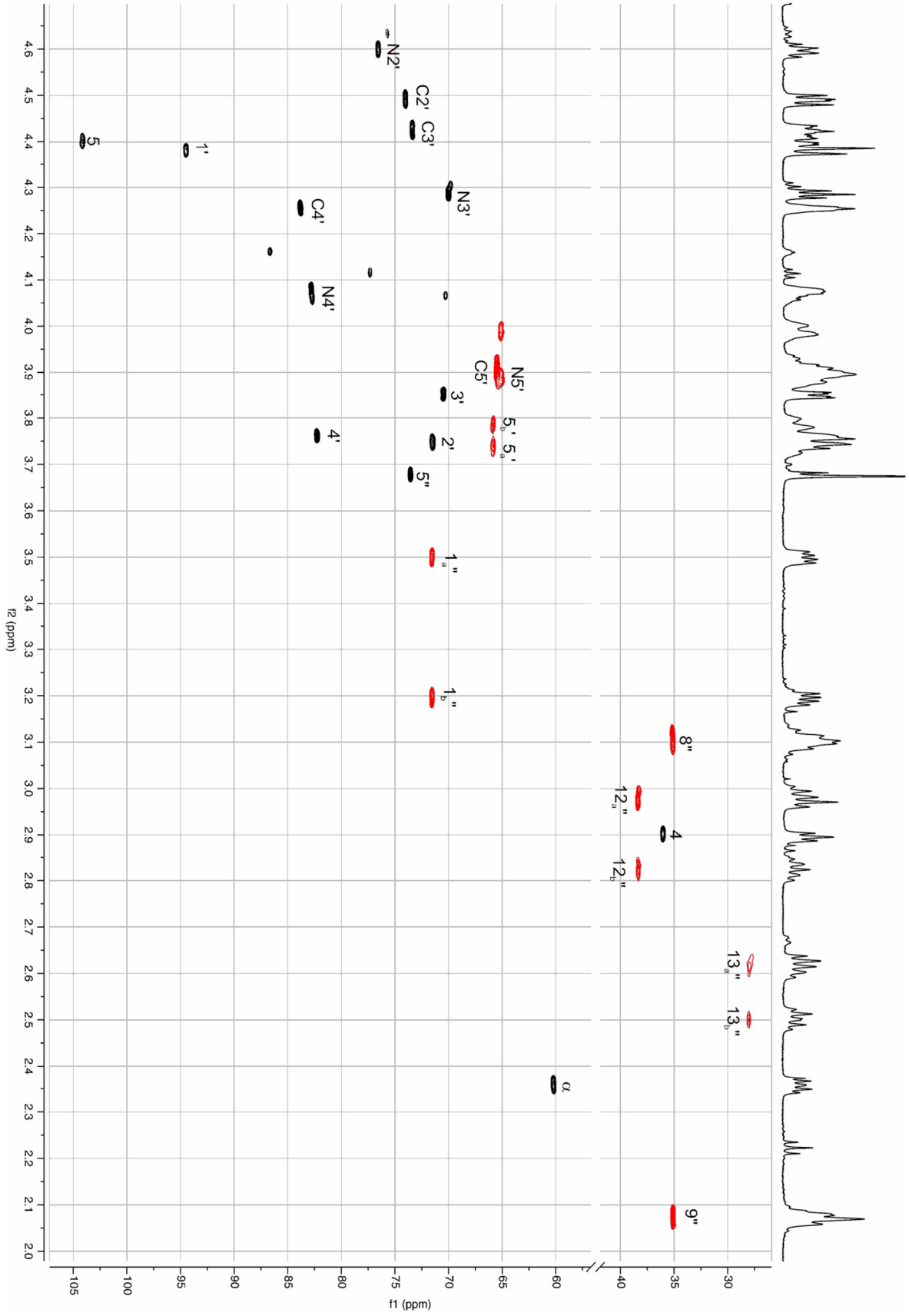
Figure S2. High-resolution MS analysis of the purified compound (C4/Cα-adduct) demonstrating its nature as a crotonyl-CoA-NADPH adduct. (**a**) HPLC-ESI-MS spectra of the purified dead end product showing the $[M+2H]^{2+}$ ion with its isotopic pattern. (**b**) MS2 spectrum of the parent ion at m/z 791.1244 showing fractioning into butyryl-CoA (838), $NADP^+$ (744) and $NADP^+$ without nicotinamide (622, loss of 122).

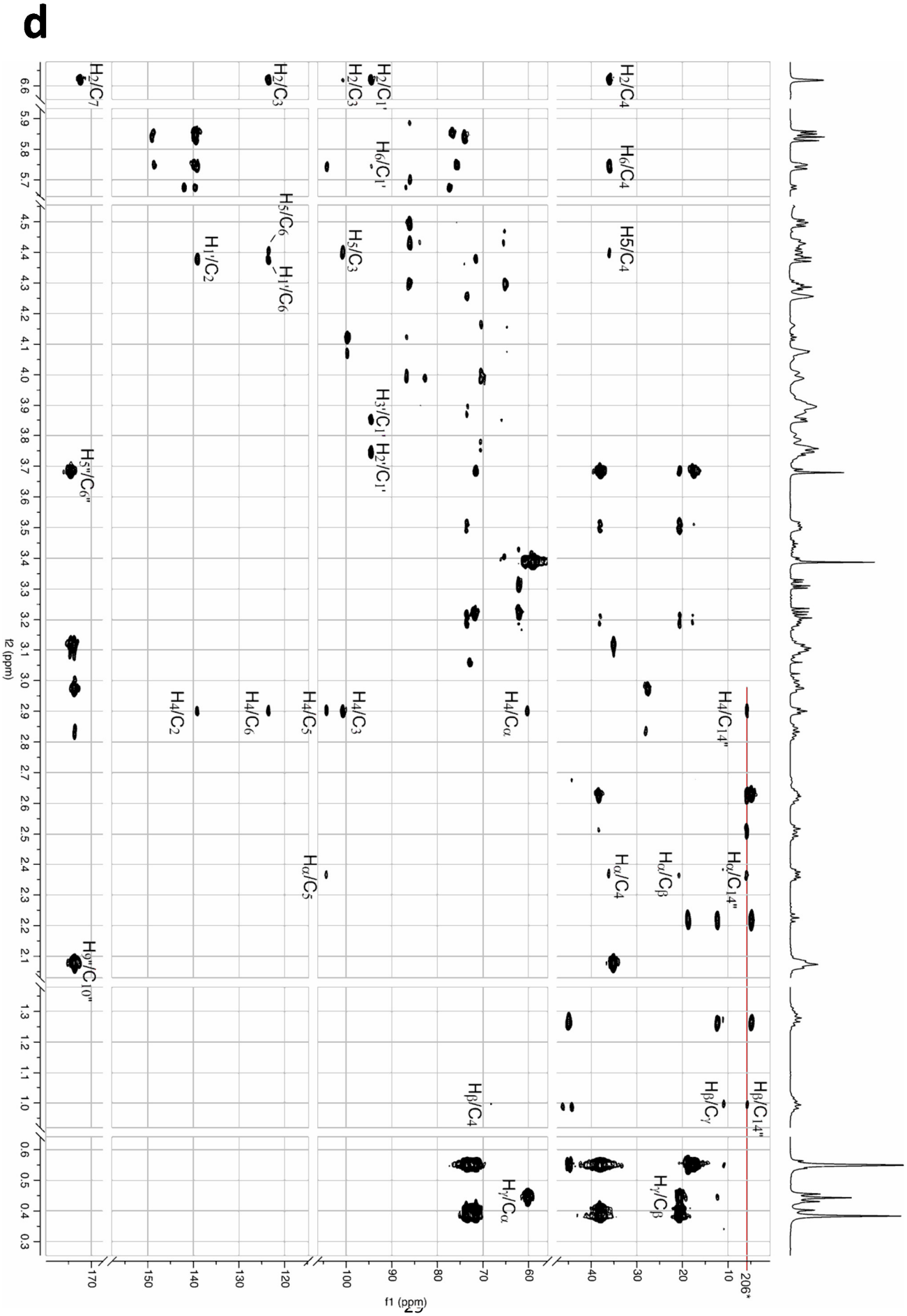
Figure S3. NMR analyses of the purified compound demonstrating its C4/Cα-adduct nature. (**a**) ^{1}H-NMR spectrum, (**b**) DQF-COSY-, (**c**) HSQC-, (**d,e**) HMBC- and (**f**) NOE- spectra of the C4/Cα-adduct. (**g**) C4/Cα-adduct structure and proton numbering. (**h**) Assignment table for the C4/Cα-adduct.

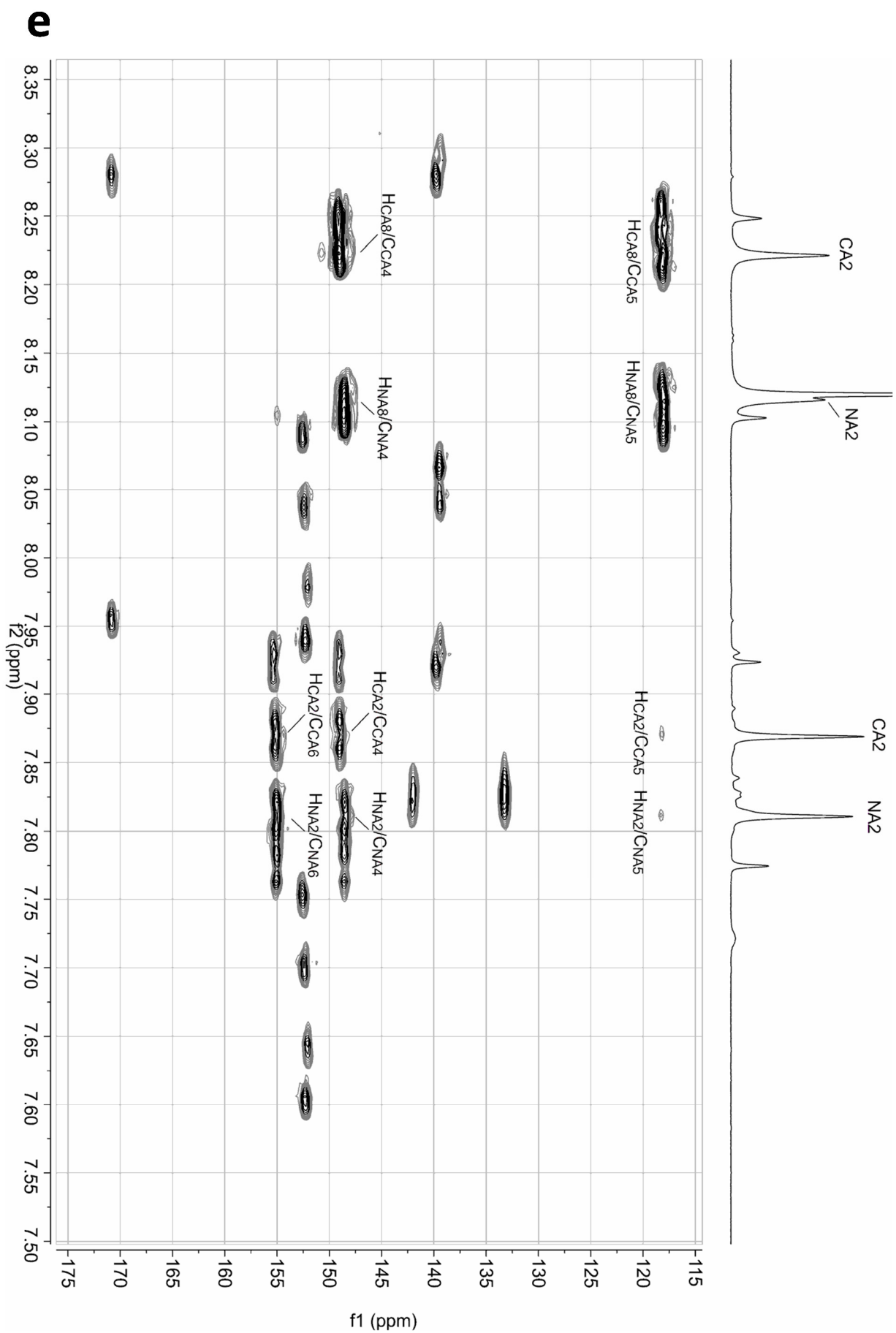


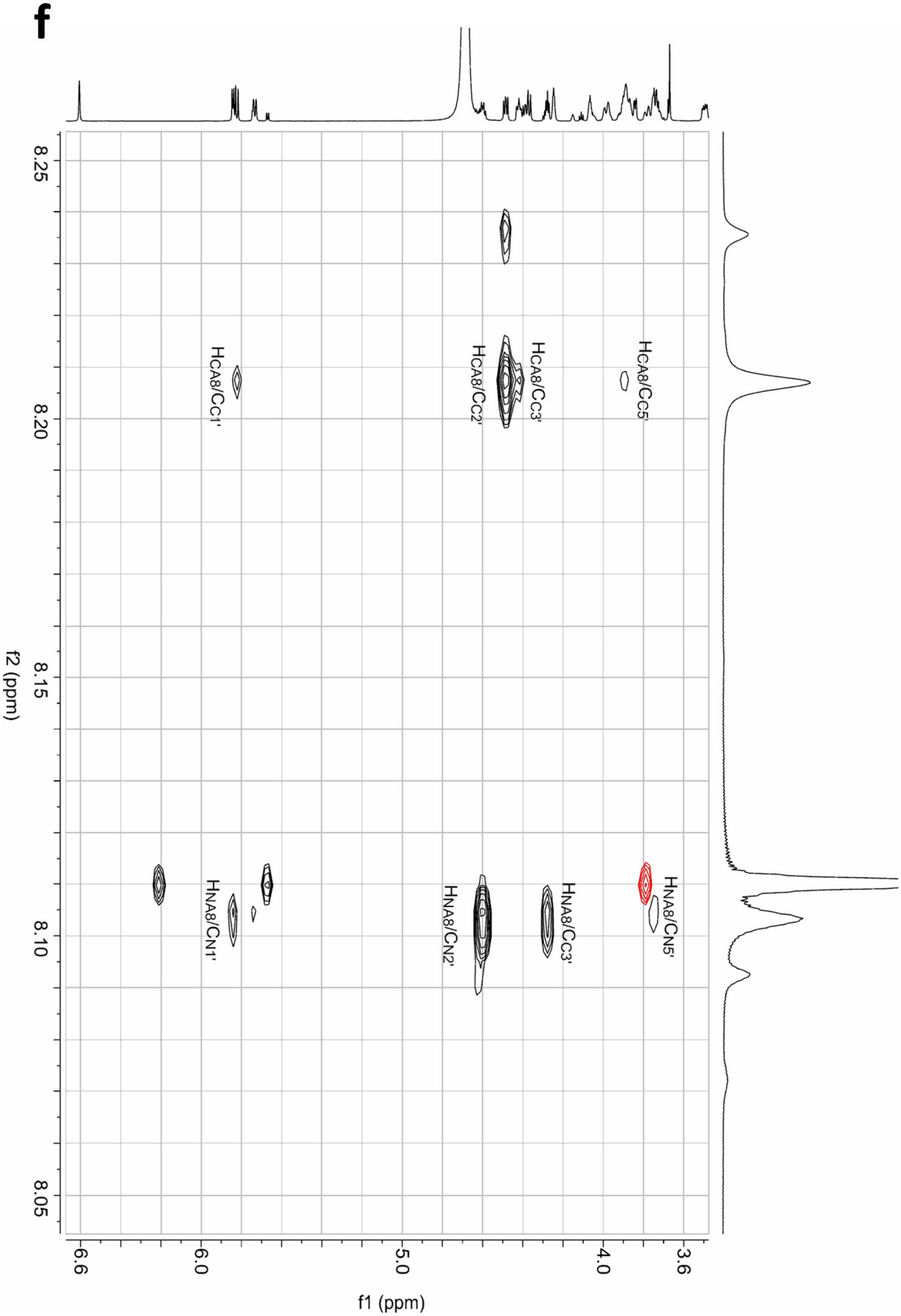

b
5,6
4,5
γ,4
βa,γ
α,βb
α,βa
f2 (ppm)
f1 (ppm)

C

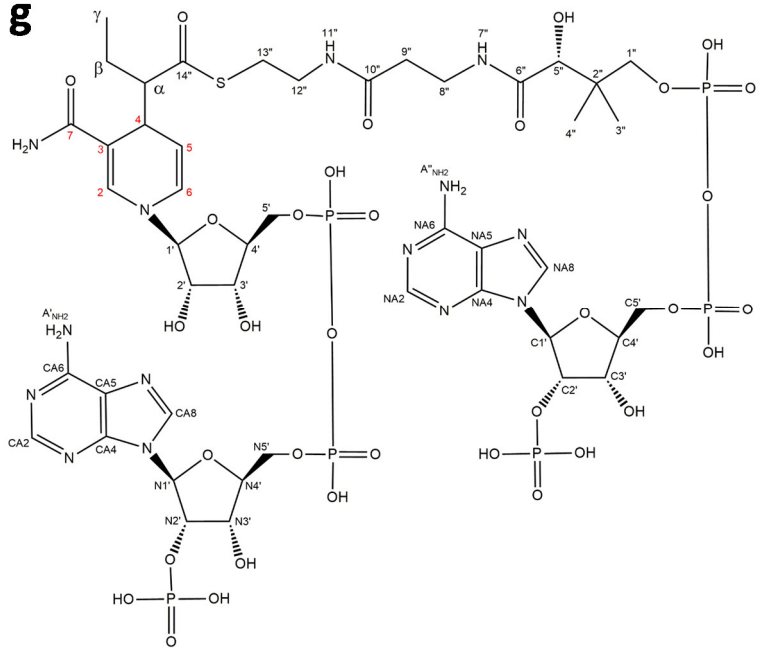

N2'
C2'
C3'
5
1'
N3'
C4'
N4'
C5'
N5'
3'
5b' 5a'
2'
4'
5"
1a"
1b"
8"
12a"
4
12b"
13a"
13b"
α
9"
f1 (ppm)
f2 (ppm)


d
H2/C7
H2/C3
H5/C6
H1'/C6
H1'/C2
H5''/C6''
H9''/C10''
H2/C4
H6/C4
H5/C4
H2/C1'
H6/C1'
H2/C3
H5/C3
H3'/C1'
H2'/C1'
H4/C5
H4/C3
Hα/C5
H4/Cα
H4/C6
H4/C2
Hβ/C4
Hγ/Cα
Hβ/C14''
Hβ/Cγ
Hγ/Cβ
H4/C14''
Hα/C14''
Hα/Cβ
Hα/C4
206*
f2 (ppm)
f1 (ppm)


e
CA2
NA2
CA2
NA2
HCA8/CCA5
HNA8/CNA5
HCA8/CCA4
HNA8/CNA4
HCA2/CCA5
HNA2/CNA5
HCA2/CCA4
HNA2/CNA4
HCA2/CCA6
HNA2/CNA6
f2 (ppm)
f1 (ppm)


f
8.25
8.20
8.15
8.10
8.05
f2 (ppm)
6.6
6.0
5.0
4.0
3.6
f1 (ppm)
HCA8/CC1'
HCA8/CC2'
HCA8/CC3'
HCA8/CC5'
HNA8/CN1'
HNA8/CN2'
HNA8/CC3'
HNA8/CN5'

g



Position	^{1}H-Shift	^{13}C-Shift	HMBC	COSY
α	2.35	60.2	6*, 11.0, 20.6, 36.0, 104.1	βa, βb,
βa	0.98	20.6	6*, 11.0, 44*, 46*, 60.2, 68*	α, βb, γ
βb	1.27	20.6	4.9*, 11.0, 12.3*, 44.9*, 60.2	α, βa, γ
γ	0.42	11.0	20.6, 60.2	βa, βb
2	6.61	139.0	36.0, 94.5, 100.7, 123.5, 172.5	-
3	-	100.7	-	-
4	2.89	36.0	6*, 20.6, 60.2, 100.7, 104.1, 123.5, 139.0, 172.5	5
5	4.41	104.1	36.0, 100.7, 123.5	4, 6
6	5.74	123.5	36.0, 94.5, 104.1, 139.0	α, 5
1"a	3.20	71.6	17.6, 20.6, 38.1, 73.5	1"b
1"b	3.49	71.6	17.6, 20.6, 38.1, 73.5	1"a
2"	-	38.1	-	-
3"	0.38	17.6	17.6, 38.1, 71.6, 73,5	-
4"	0.54	20.7	20.7, 38.1, 71.6, 73,5	-
5"	3.68	73.5	17.6, 20.7, 38.1, 71.6, 73,5, 174.6	-
6"	-	174.6	-	-
7"	-	-	-	-
8"	3.11	35.2	35.2, 173.7	9"
9"	2.08	35.2	35.2, 173.7	8"
10"	-	173.7		-
11"	-	-		-
12"a	2.82	38.4	28.0, 173.7	12"b, 13"a, 13"b
12"b	2.98	38.4	28.0, 173.7	12"a, 13"a, 13"b
13"a	2.49	28.0	38.4	12"a, 12"b, 13"b
13"b	2.62	28.0	38.4	12"a, 12"b, 13"a
14"	-	206		-
1'	4.38	94.5	71.5, 123.5, 139.0	2'

Position	^{1}H-Shift	^{13}C-Shift	HMBC	COSY
2'	3.75	71.5	94.5	1', 3'
3'	3.85	70.5	65.8, 94.5	2', 4'
4'	3.76	82.3	70.5, 94.5	3', 5'
5'a	3.73	65.8	70.5	4'
5'b	3.78	65.8	70.5	4'
C1'	5.83	85.9	74.0, 139.5, 149.1, 75.7*	2'''
C2'	4.50	74.0	85.9	1''', 3'''
C3'	4.42	73.4	65.5, 83.8	2''', 4'''
C4'	4.25	83.8	73.4	3''', 5'''
C5'	3.90	65.5	73.4, 83.8	4'''
N1'	5.85	86.0	76.6	2''''
N2'	4.60	76.6	74.0*	1'''', 3''''
N3'	4.30	70.1	86.0, 65.1	2'''', 4''''
N4'	4.07	82.7	64.8*, 99.8*	3'''', 5''''
N5a'	3.88	65.1	70.3, 73.4*	4''''
N5b'	3.99	65.1	70.3, 82.8, 86.8	
NA2	7.81	152.0	118.1, 133.2*, 142.0*, 148.5, 154.9	
NA4	-	148.5		
NA5	-	116.3		
NA6	-	155.0		
NA8	8.12	152.5	118.1, 149.0, 155.1	
CA2	7.87	139.4	118.1, 148.5	
CA4	-	149.0		
CA5	-	116.2		
CA6	-	155.2		
CA8	8.22	139.4	118.1, 149.0	

Figure S4. Inhibition assay and uncatalyzed decay of the C4/Cα-adduct. Inhibition assays for Etr1p wt catalyzed reduction of crotonyl-CoA with NADPH were measured in 100 mM Na_2HPO_4 pH 7 buffer at 31°C **A)** Inhibition assay of C4/Cα-adduct, measured at a constant crotonyl-CoA concentration of 128 μM, shows competitive inhibition towards NADPH with a K_i of 70 ± 45 nM. **B)** Inhibition assay of C4/Cα-adduct, measured at a constant NADPH concentration of 100 μM, shows non-competent inhibition towards crotonyl-CoA. **C)** Adduct decay was measured at 328 nm in 100 mM Na_2HPO_4 pH 7.9 at 31 °C. The concentration of the C4/Cα-adduct was calculated using an extinction coefficient of 4.1 ± 0.2 cm^{-1} mM^{-1}. The resulting decay constant of the C4/Cα-adduct is 0.02490 ± 0.00001 min^{-1}. Errors bars represent standard deviations of triplicates. Errors are given as 95% confidence intervals as determined from nonlinear fits of at least ten data points. **D)** Stopped flow measurements at 382 nm of enzyme assay with Etr1p Y79F (purple) show an initial buildup of C2-ene adduct (note that the C4/Cα-adduct does not show substantial absorption at 382 nm, see also Supplementary Figure 1). A control reaction without crotonyl-CoA is shown in black.

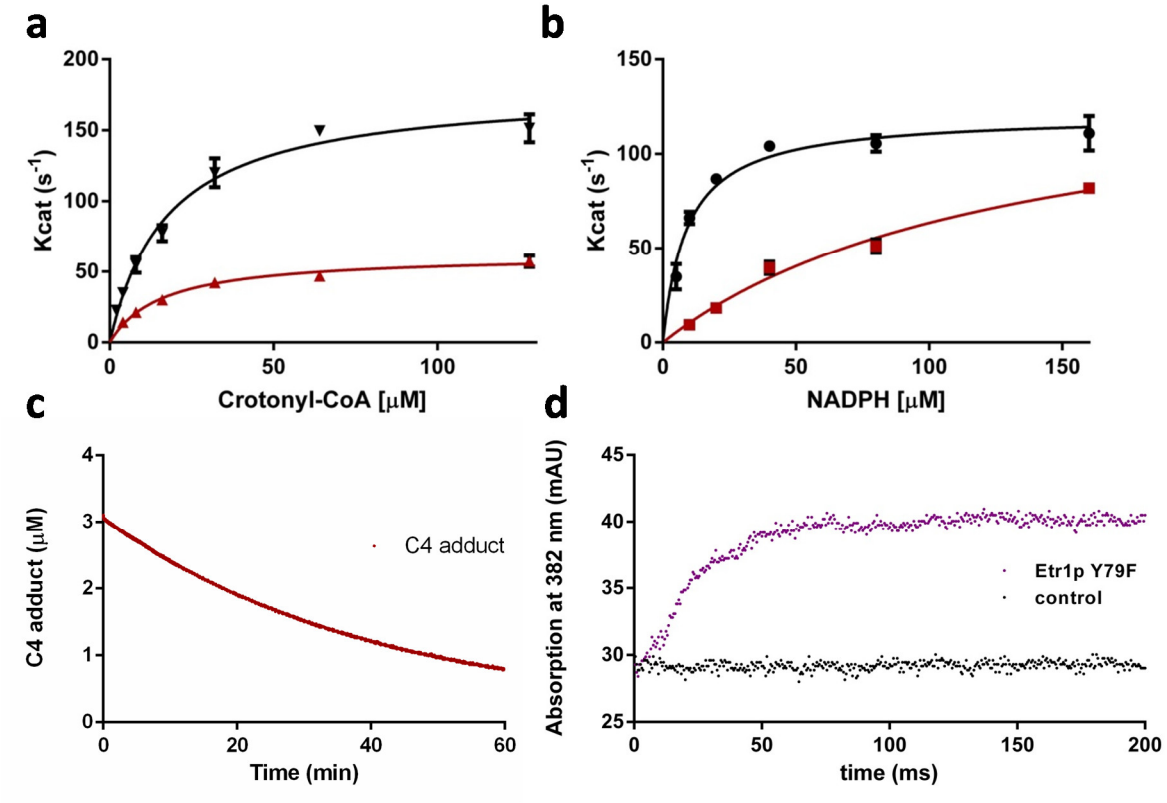


Figure S5: Scanning kinetic screening of C2-ene adduct conversion to detect *si*-face proton donor enzyme variants. All assays were performed in 100 mM Na_2HPO_4 pH 8 buffer at 30 °C. (**a**) Etr1p Y79F W322D, forms mostly C4/Cα-adduct. (**b**) Etr1p Y79F W322H forms only C4/Cα-adduct.

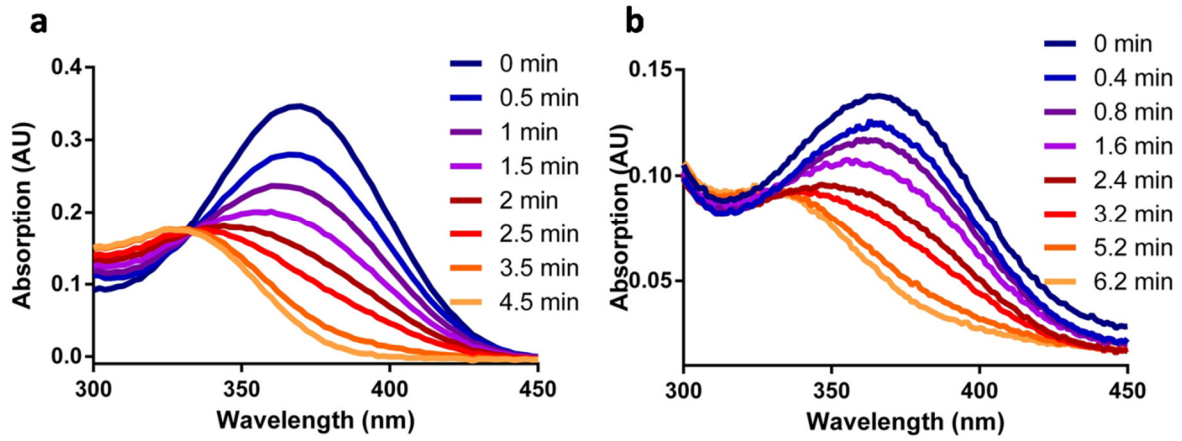


Figure S6. pH dependence of k_{cat} for Etr1p WT (black) and Etr1p Y79F W322E (red). The k_{cat} values of each mutant are fitted with a one proton-donor model (fit shown in respective color) to approximate the pK_a of the donating residue (Y79 for Etr1p WT and E322 for Etr1p Y79F W322E). The fitted values are shown below the graph. All shown data points represent Michaelis Menten fits of a minimum of 18 measurements. Error bars represent the standard deviation of those fits.

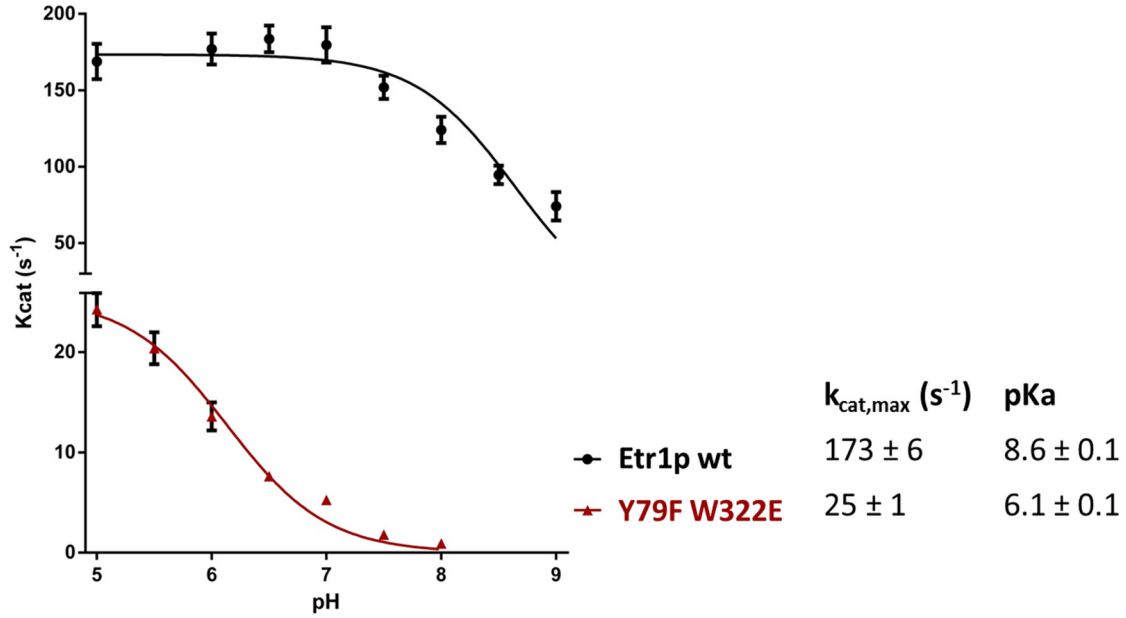
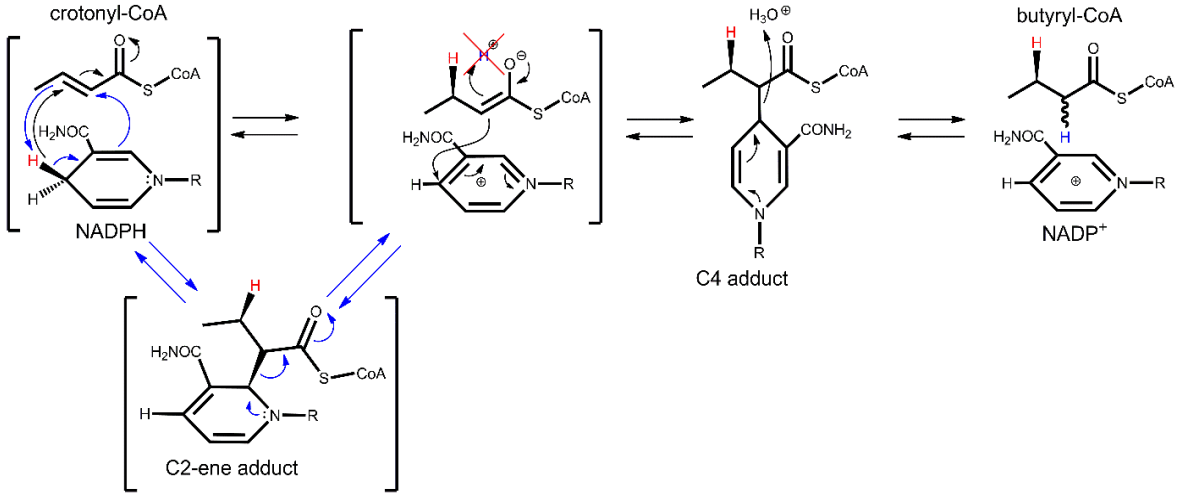


Figure S7. Proposed mechanism for formation of C4/Cα-adduct. Steps in bracket take place in the active site that lacks a proton donor (*e.g.*, Etr1p Y79F), protonation of the formed enolate is prevented and instead a Michael addition takes place between the enolate and $NADP^+$, forming the C4/Cα-adduct which can subsequently decay in solution into butyryl-CoA and $NADP^+$.



CHAPTER III

A conserved threonine prevents self-intoxication of enoyl-thioester reductases

Authors:
Raoul G. Rosenthal, Bastian Vögeli, Tristan Wagner, Seigo Shima, Tobias J. Erb

Published in:
Nature Chemical Biology 7,745-749 (2017); DOI: 10.1038/nchembio.2375

Author contributions:
R.G.R., B.V. and T.J.E. conceived and designed all experiments, with the exception of crystallography experiments, which were designed together with T.W and S.S. Enzyme kinetic assays and stopped-flow measurements were performed by R.G.R. and B.V. MS experiments were performed by B.V. and R.G.R. and analyzed together with N.S. Cortina. Crystallography experiments were performed by B.V. together with T.W. T.W. and B.V. collected the diffraction data and T.W. interpreted the results. R.G.R., B.V. and T.J.E. wrote the paper.

3. A conserved threonine prevents self-intoxication of enoyl-thioester reductases

3.1. Abstract

Enzymes are highly specific biocatalysts, yet they can perform unwanted side reactions. Here we investigated the factors that direct catalysis in the enoyl-thioester reductase Etr1p. We show that a single conserved threonine is essential in suppressing the formation of a side product that otherwise acts as a high-affinity inhibitor of the enzyme. Substitution of this threonine with isosteric valine increases side product formation by more than six orders of magnitude, while decreasing turnover frequency by only one order of magnitude. Our results show that promoting wanted reactions and suppressing unwanted side reactions are independently operating principles at the active site of Etr1p, and that the active suppression of side reactions is highly conserved in the family of medium chain dehydrogenases/reductases (MDR). Our discovery emphasizes that the active destabilization of competing transition states is an important factor during catalysis, which has implications for the understanding and the de novo-design of enzymes.

3.2. Introduction

Enzymes are known to be highly efficient and highly specific biocatalysts. Yet, many enzymes catalyze intrinsic side activities that give rise to (unwanted) side products despite sustained evolutionary pressure[1]. One prominent example is the oxygenase side reaction of ribulose-1,5-bisphosphate carboxylase/oxygenase that leads to the fixation of O_2 instead of CO_2, wasting up to 20% of the energy harvested by photosynthesis[2]. During glycolysis, the cytotoxic compound methylglyoxal is formed as side reaction of triose phosphate isomerase[3-5]. Side reactions also occur in terpene synthases that react through highly reactive intermediates, which has raised the question, how in these enzymes the reaction is guided to produce preferentially the desired product against a myriad of different reaction outcomes[6]. From these examples it becomes apparent that enzyme catalysis is a balance between positive contributions that promote the wanted reactivity along the substrate-to-product reaction coordinate, as well as negative contributions that suppresses unwanted side reactions on the same reaction coordinate[7-8]. Most research has focused on understanding the principles of positive catalysis, *i.e.*, the active site geometry that leads to stabilization of a productive transition state. Much less, however, is understood about the mechanisms that protect enzymes from performing unwanted side reactions, which have been termed 'negative catalysis'[7].

An interesting example with which to investigate these principles are nicotinamide (NAD(P)H)-dependent reductases that are among the most abundant enzymes in biology (16% of all BRENDA entries are NAD(P)H-dependent enzymes)[9]. NAD(P)H-dependent reductases serve in essential cellular processes, such as energy conservation and fatty acid metabolism[10-11]. They are important drug targets (*e.g.*, in tuberculosis treatment and hypercholesterolemia therapy[12-13], and they are increasingly used in biocatalysis and biotechnology[14-15]. For more than half a century, NAD(P)H-dependent reductases have been the prime study objects for understanding, inhibiting and engineering cofactor-dependent enzyme catalysis. Only recently, however, has it been realized that mutagenesis can perturb the catalytic cycle of NAD(P)H-dependent reductases and lead to the accumulation of catalytic intermediates and side products[9, 16].

Mitochondrial enoyl-thioester reductase Etr1p from *Candida tropicalis*[17-18] catalyzes the reduction of enoyl-CoAs during fatty acid biosynthesis (**Figure 1a, 1b**). Mutation of the active site proton donor, tyrosine 79, to phenylalanine leads to build-up of several covalent adducts between NADPH and the

enoyl-thioester substrate crotonyl-CoA[16] (**Figure 1c, 1d**). When the Y79F variant is incubated with crotonyl-CoA and NADPH a covalent adduct between the C2-carbon of NADPH and the Cα-carbon of crotonyl-CoA, a "C2-ene adduct", is formed. Note that the purified C2-ene adduct is catalytically competent for the wild type (WT) Etr1p reaction. Etr1p WT effectively converts and protonates the C2-ene adduct, indicating that this adduct lies on (or near to) the reaction coordinate (**Figure 1b, 1d**)[9]. In the Y79F variant, however, that is impaired in protonation, the C2-ene adduct is further converted into another covalent adduct between the C4 carbon of NADPH and the Cα-carbon of crotonyl-CoA, "C4-adduct"[16]. In contrast to the C2-ene adduct, the C4-adduct acts as competitive inhibitor of Etr1p WT with nanomolar affinity, indicating that it is a dead-end product of the mutated enzyme reaction (**Figure 1d**). Thus, Etr1p (and putatively other reductases) must have evolved mechanisms to funnel the reaction towards product formation while suppressing formation of the inhibitory C4-adduct.

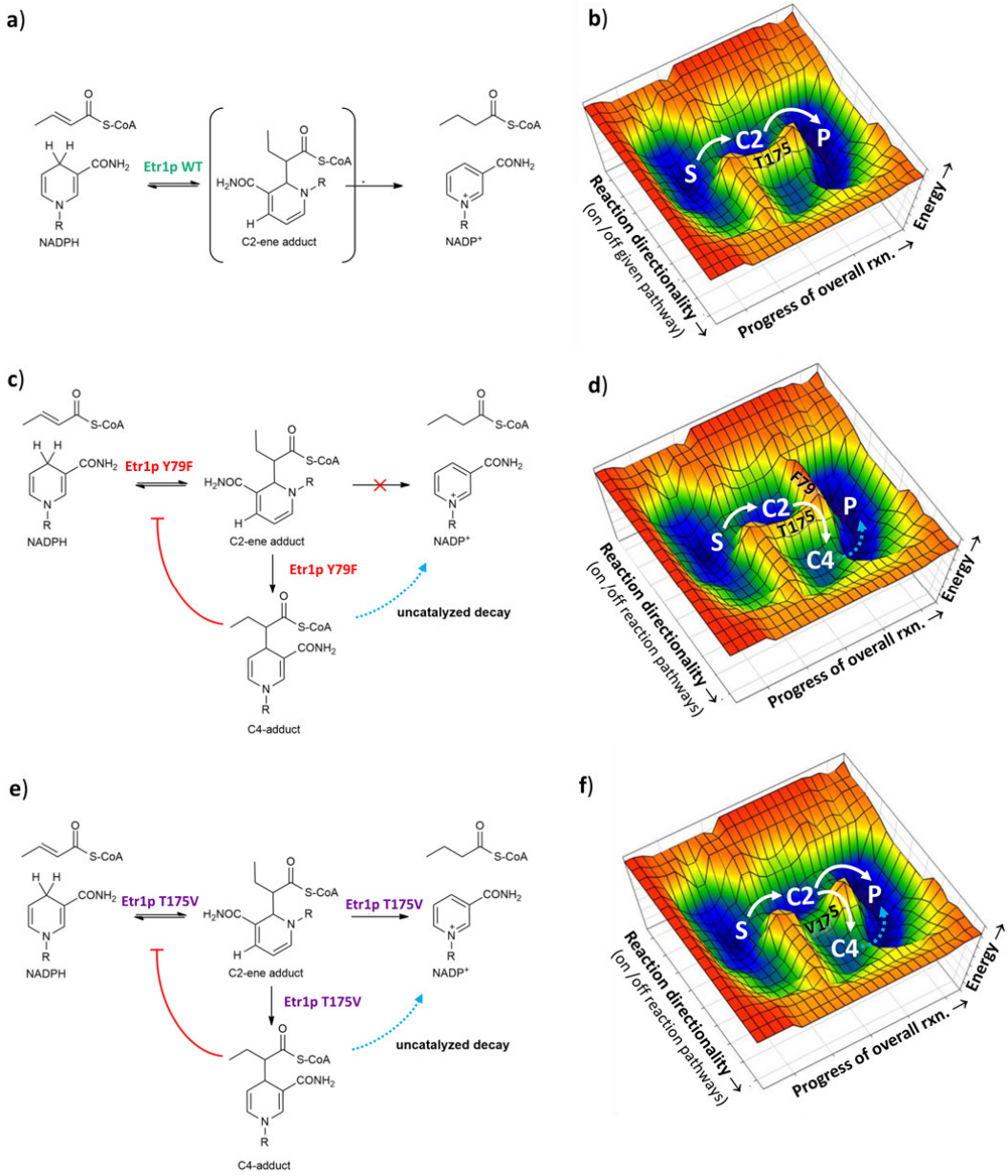


Figure 1. Reaction schemes of Etr1p WT and variants and their effect on a hypothetical catalytic landscape. a) Reaction scheme of Etr1p WT that catalyzes the NADPH-dependent conversion of crotonyl-CoA into butyryl-CoA. **b)** Hypothetical landscape of Etr1p WT catalysis. **c)** Reaction scheme of Etr1p Y79F that converts crotonyl-CoA with NADPH to the C2-ene adduct and subsequently to the C4-adduct. The C4-adduct is a strong inhibitor of Etr1p (red line) and decays slowly in solution (dotted blue lines). **d)** Hypothetical landscape of Etr1p Y79F catalysis. Mutation of the proton donor Y79F raises the energetic

barrier of the protonation step, causing formation of the C4-ene adduct. **e)** Reaction scheme of Etr1p T175V that also catalyzes formation of a C4-ene adduct from crotonyl-CoA with NADPH. **f)** Hypothetical landscape of Etr1p T175V catalysis. Mutation of T175 lowers the energetic barrier to form the C4-adduct leading to C4-adduct accumulation and self-intoxication of the enzyme even though the protonation step in the T175V is still functional. Abbreviations used: S, substrates; P, products; C2, C2-ene adduct; C4, C4-adduct.

Here, we show that the large family of MDRs feature a highly conserved threonine residue that almost exclusively serves in suppressing formation of the inhibitory C4-adduct in Etr1p. We demonstrate that the single hydroxyl group of the conserved threonine is essential to maintain a remarkably high suppression of side reactions but has only marginal effects on the catalytic rate of Etr1p (**Figure 1e, 1f**). This provides evidence that promoting catalysis and suppressing side reactions can be independently, co-existing functions at the active site of enzymes. Notably, the threonine residue is more conserved in the MDR enzyme family than residues promoting catalysis, indicating a strong evolutionary pressure on side-reaction suppression in these enzymes. Taken together, our results suggest that the active destabilization of competing transition states, commonly termed as 'negative catalysis', is fundamental to enzyme catalysis and serves as complementary force to transition state stabilization.

3.3. Results

3.3.1. A remote T175 contributes to catalysis

To understand which residues contribute to the productive resolution of the C2-ene adduct (positive catalysis) and which prevent formation of C4-adduct ('negative catalysis'), we combined site directed mutagenesis of Etr1p with kinetic isotope effect (KIE) profiling. Our previous studies showed that hydride transfer and protonation become uncoupled when the active site of Etr1p is mutated[16]. This uncoupling of NAD(P)H consumption and proton transfer makes classical spectrophotometric assays not suited to measure protonation KIEs in Etr1p, and likely in other reductases. For that reason, we established a mass spectrometry based assay that is able to measure the intramolecular KIE on protonation, similar to earlier work by Northrop and coworkers[19]. The assay quantifies the incorporation ratios of hydrogen versus deuterium from the solvent into the products, which is given as a $^{D}k_{obs}$ (refer to **3.5 Methods** for details).

Enzyme/reaction	$^{D}k_{obs}$	$^{(NADPH/NADPD)}(k_{cat})$	$^{(NADPH/NADPD)}(k_{cat}/K_m)$
WT	1.78 ± 0.08	4.3 ± 0.2	1.94 ± 0.95
S70A	1.11 ± 0.07	4.2 ± 0.2	1.1 ± 0.5
T324A	1.92 ± 0.15	3.6 ± 0.1	1.1 ± 0.3
S70A T324A	1.20 ± 0.08	n.d.	n.d.
Y79F	4.8 ± 0.8	n.d.	n.d.
T175V	4.4 ± 0.8	0.96 ± 0.04	2.0 ± 0.7
C2-ene adduct decay	4.5 ± 0.2	n.a.	n.a.
C4-adduct decay	4.4 ± 0.2	n.a.	n.a.

Table 1 Observed kinetic isotope effects in Etr1p and variants as measured by incorporation ratios of H/D from buffers with defined H/D ratios, $^{D}k_{obs}$, $^{(NADPH/NADPD)}(k_{cat})$ and $^{(NADPH/NADPD)}(k_{cat}/K_m)$ were measured in a non-competitive fashion, for assay details see **Methods 3.5.10**. All values represents averages of two independent values, based on at least four data points each. Errors given as 95% confidence intervals. Abbreviations: n.d.= not determined, n.a.= not applicable.

With this method we systematically assessed the intramolecular KIE of different Etr1p active site variants on protonation (**Table 1**). In a T175V variant, we measured a large intramolecular KIE on protonation ($^{D}k_{obs}$ = 4.8 ± 0.8) compared to the WT (1.8 ± 0.1). This result was surprising considering

that T175 is located more than 6 Å away from the Cα of crotonyl-CoA and almost 9 Å away from the Y79 hydroxyl group, the proton donor in Etr1p (**Figure S1**). The finding was even more surprising given that S70 and T324 which are in close proximity to Y79 did not show a strong KIE on protonation upon mutation (**Table 1**) raising the question how such a remote residue like T175 can affect the protonation step. At the same time, the T175V variant also appeared to be affected in hydride transfer. Compared to the WT, the apparent $^{(NADPH/NADPD)}k_{cat}$ was reduced from 4.3 ± 0.2 in the WT to 0.96 ± 0.04. Reduction of the KIE on hydride transfer to unity in the T175V variant was also remarkable, considering the close proximity of the γ-OH of T175 to the C4 position of the NADPH-nicotinamide ring. In summary, KIE profiling indicated an essential contribution of T175 to catalysis, although the exact role of this active site residue remained unclear.

3.3.2. T175 is not essential to catalysis

Spectrophotometric characterization of T175V with crotonyl-CoA and NADPH showed a consumption of several equivalents of NADPH compared to enzyme during manual mixing time. This very fast initial decrease of NADPH was followed by a slow phase with apparent kinetic parameters of k_{cat} = 0.082 ± 0.005 s^{-1} and a $K_{m\ NADPH}$ of 2.7 ± 0.8 μM, which corresponded to approximately 0.05 % of WT activity. We investigated this unusual NADPH consumption behavior in more detail by stopped flow UV-Vis spectroscopy. In the burst phase, the initial turnover of T175V was 12 ± 1 s^{-1} at 30 °C, which was 150-fold higher than under steady state conditions, corresponding to 4 % of WT activity under the same conditions (**Figure 2a**). The burst amplitude was directly proportional to the amount of enzyme (**Figure S2**) and lasted more than 3 turnovers, indicating that the biphasic behavior was not caused by slow conformational changes of Etr1p T175V or slow product release. Because the initial turnover frequency of T175V was only reduced by one order of magnitude compared to the WT, we concluded that T175 affected the energetics on the substrate-to-product reaction coordinate but was not essential to catalysis.

3.3.3. T175 controls the correct outcome of catalysis

Because results above excluded an essential role of T175 in promoting catalysis, we speculated that the residue might function in funneling intermediates along the reaction coordinate, suppressing the formation of deleterious side products, such as the inhibitory C4-adduct (**Figure 1e, 1f**). We thus devised experiments to detect whether the C4-adduct was indeed formed in the T175V variant during catalysis. We used UPLC-MS to screen for C4-adduct formation under steady state conditions. The T175V variant accumulated C4-adduct at an adduct to enzyme ratio of 1.25 ± 0.07 (**Figure 2b**). C4-adduct formation was independently confirmed by UV-Vis difference scanning (**Figure S3**). Notably, we could not detect any C4-adduct formation in the WT with either method, even not in a reaction with 40 mM NADPH, 40 mM crotonyl-CoA, and 15 μM WT (**Figure 2b**). At a detection limit of 24 nM for the C4-adduct with the UPLC-MS method (defined as detection of the target ion and isotopic pattern in 5 consecutive spectra), WT forms the C4-adduct in less than one in $1.7 \cdot 10^6$ turnovers. In other words, whereas the WT did not show detectable deviations from the substrate-to-product reaction coordinate, almost every third reaction fell off this reaction coordinate in the T175V variant.

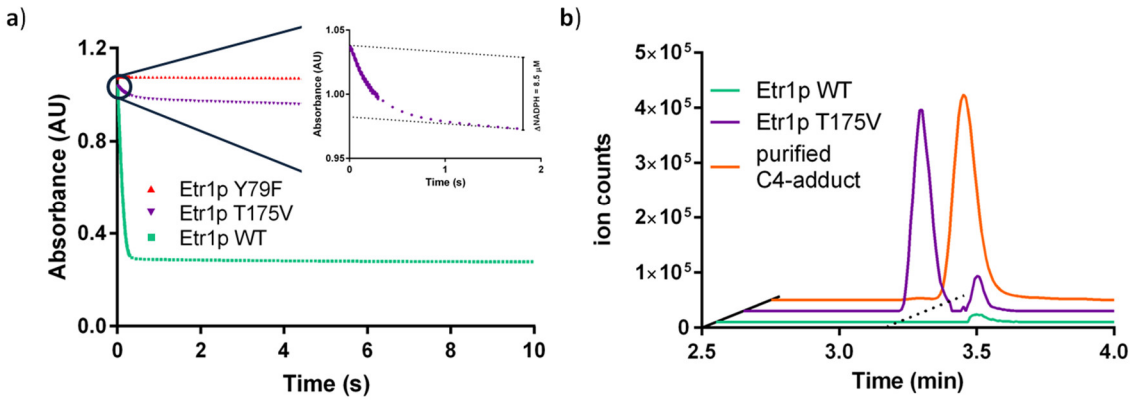


Figure 2. Detailed characterization of the reactions catalyzed by Etr1p WT and variants. The experiments show that T175V is still active but intoxicates itself rapidly by formation of the C4-adduct. **a)** Stopped flow kinetics of Etr1p WT (green), the proton donor mutant Y79F (orange) and the T175V variant (purple). T175V shows a two phasic behavior. An initial burst-phase followed by a slow linear phase (see close up). The burst phase lasted over 3 turnovers. All assays contained 2.5 μM enzyme, 160 μM NADPH and 120 μM crotonyl-CoA in 20 mM Tris-HCl buffer (pH 7.9) with 150 mM NaCl. Data shown are representatives of at least three independent experiments. **b)** UPLC-MS analysis of assays with either Etr1p WT (green) or the T175V variant (purple). The T175V variant shows accumulation of a C4-adduct (compare to authentic C4-adduct standard, orange trace), while the Etr1p WT did not accumulate detectable amounts of the C4-adduct. Data shown are representatives of at least three independent experiments.

To confirm that the observed effects were not due to changes in the secondary or tertiary structure of the variant, we investigated its folding state by circular dichroism (CD). The CD spectrum of T175V was practically superimposable with the WT and all other enzyme variants created in this study, confirming a correct folding state (**Figure S4**). To test whether cofactor binding was altered in the T175V variant, we determined the binding affinities of NADPH by isothermal titration calorimetry. T175V showed a binding constant of 1.6 ± 0.5 $\cdot 10^5$ M^{-1} for NADPH, which was comparable to the WT enzyme (3.8 ± 1.0 $\cdot 10^5$ M^{-1}), and all other variants tested albeit with slower binding kinetics. From these experiments, we concluded that the T175V variant was unaltered in its biophysical properties, and that the observed loss of catalytic control was indeed caused by an altered catalytic behavior.

3.3.4. Structural basis for the loss of catalytic control in T175V

To understand the structural base for the loss of catalytic control in T175V compared to WT we crystallized T175V. The structure confirmed a correct folding of the enzyme. The active site including the water network of apo-WT (PDB: 4W99 at 2.0 Å) and T175V (PDB: 5LB9 at 2.1 Å monoclinic crystals and 5LBX at 2.5 Å trigonal crystals) were superimposable with a root-mean-square deviation of 0.44 Å for all Cα atoms. There was one notable exception. A water molecule in direct contact with T175 was shifted by 1 Å in the T175V variant compared to the WT (**Figure S5**). We note that in contrast to earlier assumptions[20-21], this specific change in the water network did not cause a collapse of the apoenzyme's active site.

The tertiary complex of T175V occupied by $NADP^+$ and crotonyl-CoA also showed a WT-like active site geometry with a root-mean-square deviation of 0.49 Å for all Cα atoms compared to tertiary WT complex (PDB: 4WAS). The small movement of the water molecule observed in the apo-form of T175V translated to a tilt of the nicotinamide ring by 17.5° in the tertiary complex of T175V compared to the WT (**Figure 3, Figure S6** and **S7**). This tilt was observed in monoclinic and trigonal crystals of T175V (**Figure S8**), ruling out a crystal packing artifact. The tilt of the nicotinamide caused a 1 Å shift of the nicotinamide-C4 towards the Cα of crotonyl-CoA in T175V. This small difference seems to form the basis for the remarkable loss of discrimination against C4-adduct formation in the T175V variant by more than 10^6 compared to the WT.

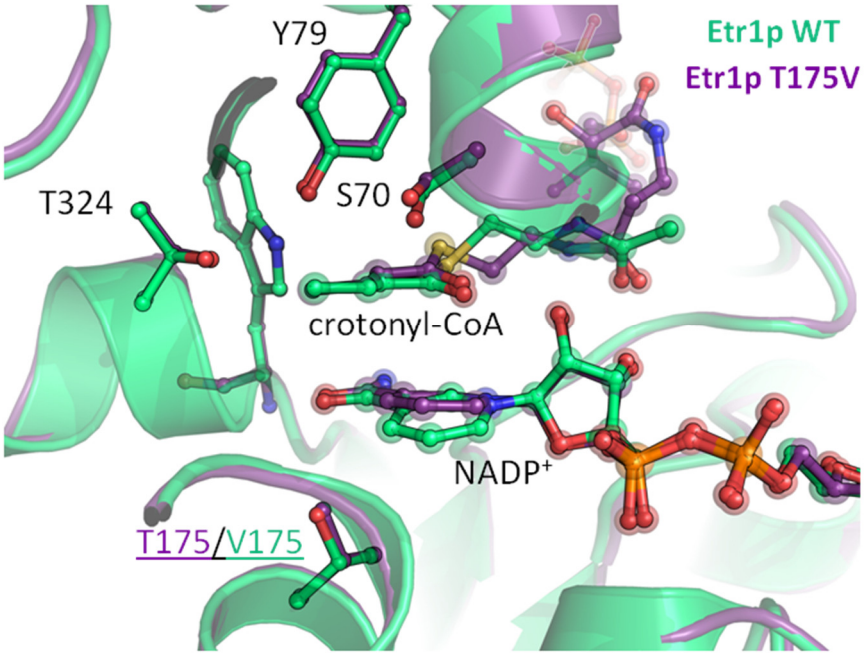


Figure 3. Crystal structure of Etr1p WT and the T175V variant. Overlay of crystal structures of Etr1p WT and the T175V variant, co-crystallized with $NADP^+$ and soaked with crotonyl-CoA. The structures superimpose with a rmsd of 0.45 Å. The only notable difference near the active site is a slight shift in the nicotinamide ring of $NADP^+$.

3.3.5. Catalytic control is highly conserved in the MDR family

Interestingly, C4-adducts are already known to be potent inhibitors of NAD(P)H-dependent enzymes. A C4-adduct of lactate and NAD^+ was reported to inhibit lactate dehydrogenase[22] , the C4-adduct of NAD^+ and 3-pentanone was described to act as inhibitor of *Drosophilla melanogaster* alcohol dehydrogenase[23], and the C4-adduct of finasteride (a steroid analog) and NADH inhibits human 5α-reductase, which is used to treat benign prostatic hyperplasia and male pattern baldness[24]. This inhibitory potential of C4-adducts postulates a strong evolutionary driving force to suppress their formation of during catalysis, which is reflected in the fact that T175V is one of the most conserved residues in the MDR superfamily (**Figure S9**).

Moreover, the high grade of conservation of T175 in the MDR superfamily suggests that suppressing off-pathways on the substrate-to-product reaction coordinate seems to be a stronger evolutionary constraint than lowering the energetics of the reaction coordinate itself. While the electrophilic donor in different members of the MDR superfamily can be ***1)*** a proton donor, such as tyrosine (Etr1p), lysine or aspartate (human FAS[25] and polyketide synthase enoyl-reductase domains[9, 26]), ***2)*** a CO_2 co-substrate (Ccr)[27-28], or ***3)*** might even lack an active site electrophile (LovC)[29], a homologue of T175 is strongly conserved in all these enzymes. This indicates that the role of T175 in preventing C4-adduct formation is essential across the entire MDR enzyme family. We tested this hypothesis by introducing an analogous T195V mutation into crotonyl-CoA carboxylase/reductase (Ccr), a CO_2-fixing member of the MDR superfamily[30]. This also resulted in the formation of the C4-adduct during catalysis (**Figure S10**).

We noted that the C4-adduct is a diffusible compound that once formed is not limited to self-intoxicate the enzyme that formed it, but could also inhibit other enzymes with overlapping substrate specificities. Such postulated cross-inhibition would have direct implications for the evolutionary pressure to avoid C4-adduct formation could act beyond the individual enzyme level. We tested this hypothesis with two different members of the MDR superfamily that catalyze completely different reactions, the enoyl-CoA reductase Etr1p, and above mentioned reductive enoyl-CoA carboxylase Ccr[30]. Etr1p WT was practically inactive in reactions mixtures of Ccr T195V with NADPH and crotonyl-

CoA indicating an effective cross inhibition between the two enzymes (**Figure 4**). This finding strongly suggests that guiding substrates effectively along the reaction coordinate by suppressing deleterious side reactions is a complementary evolutionary force to transition-state-stabilization, which is as fundamental to enzyme evolution as the latter one.

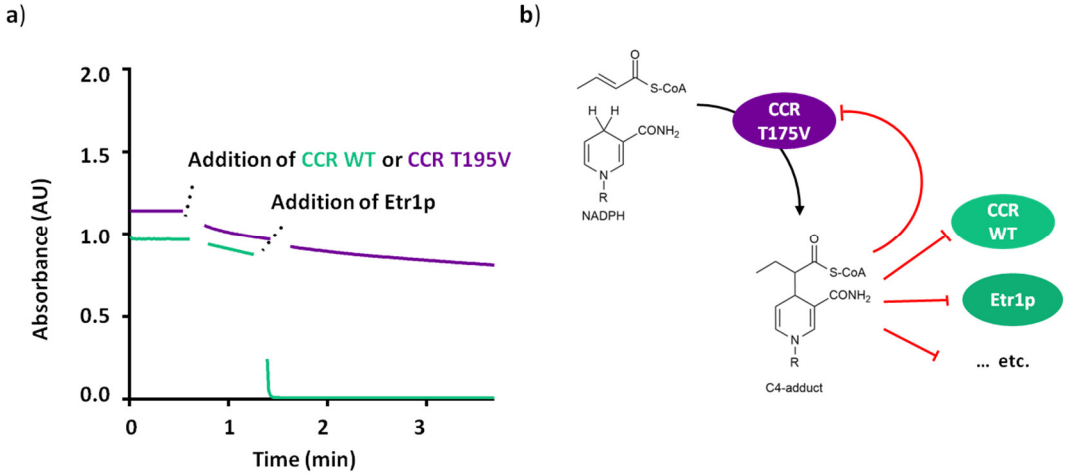


Figure 4. Cross-inhibition between Et1p WT and CCR T195V a) CCR WT (1.7 µM) was added (after 40 s) to a reaction assay containing NADPH and crotonyl-CoA that was monitored at 340 nm (green trace). After 90 s, Etr1p WT was injected into the mixture, leading to the rapid depletion of NADPH. In an equivalent assay, in which CCR T195V was used instead of CCR WT (purple trace), addition of Etr1p did not cause NADPH consumption, indicating that Etr1p was inhibited. Data shown are representatives of at least three independent experiments. **b)** Model explaining the observed cross-inhibition. CCR T195V forms the C4-adduct during catalysis, which leads to a rapid "self-intoxication" of the enzyme. The C4-adduct is also able to diffuse into other enzyme active sites, to effectively inhibit these enzymes, too. Selection for avoiding C4-adduct formation in NAD(P)H-dependent enzymes thus acts beyond individual enzyme level.

3.4. Discussion

Previous publications on enzymes from the MDR family speculated that T175 might be important for puckering of the nicotinamide ring during catalysis[20]. Here, we provide a surprising alternative explanation according to which this residue has specifically evolved to protect Etr1p and other NAD(P)H-dependent reductases from intoxication through formation of a deleterious side product of the enzymatic reaction. T175 does not promote catalysis, but serves almost exclusively in controlling the catalytic outcome of the enzyme reaction. This is reflected by the fact that the turnover frequency of T175V remains essentially unchanged, while the error rate of the reaction is increased by more than six orders of magnitude in the same variant. The reason for this dramatic loss of catalytic control apparently lies in surprisingly subtle changes in the geometry structure of the active site. Our results indicate that the hydroxyl group of T175 alone is responsible for guiding the enzyme reaction with high precision. To our knowledge there is no current example – and now simple explanation – how a single hydroxyl group could have such a strong effect on the catalytic outcome of a reaction.

On a more general note, our results demonstrate that the active site of enzymes can feature catalytic residues of distinct functionality. Residues such as Y79 that promote catalysis, and residues such as T175 that control the catalytic outcome of a given enzyme by actively suppressing unwanted side reactions. Although the existence of active site residues that guide enzyme reactions has been proposed more than 25 years ago[7], the discovery that T175 almost exclusively serves such a function is – to our knowledge – the first direct experimental evidence of this hypothesis.

Our finding that residues promoting enzyme reactions (positive catalysis) and residues directing enzyme reactions ('negative catalysis') can function independent from each other at the active site challenges the common understanding of enzyme catalysis, according to which these two traits are

inherently linked with each other. For coenzyme B12-dependent methylmalonyl-CoA mutase it was shown that an arginine is essential in suppressing an inactivating side reaction[31]. However, mutation of this arginine also affected k_{cat} by a factor of 10^4. Another example are pyridoxal phosphate (PLP) dependent enzymes, which catalyze very different reactions that all originate from a common reaction intermediate. It is generally assumed that reaction selectivity in these enzymes is maintained through promoting catalysis by stabilization of the favored transition state and/or destabilization of the ground state of the favored reaction intermediates[32-33].

In contrast, our results suggest that binding and promoting productive transition states during catalysis is not necessarily sufficient when close, competing transition states can lead to deleterious side products. It seems almost a logical consequence that in such cases side reactions need to be actively suppressed by the destabilization of competing transition states to give rise to a proficient enzyme. In this light, the function of active site residues that are conserved but not known to be essential for catalysis so far might require a closer (re-)inspection. An example are histidine 257 and phenylalanine 200 in human purine nucleoside phosphorylase. Mutation of these residues to glycine affects catalytic activity by 20- to 40-fold, but increase an N3-inosine isomerization side-reaction by at least 125- to 250-fold over the level of detection, indicating a potential role in 'negative catalysis' for these residues[34]. However, even for active site residues that are apparently essential for catalysis, it remains to be tested, whether they confer a positive catalytic function or if they prevent accumulation of inhibitory side-products like in case of Etr1p.

Taken together, our results emphasize that enzyme catalysis needs to be understood as combination of positive and negative constraints on the substrate-to-product reaction coordinate. We expect that these findings will (*i*) lead to a reinvestigation of the principles operating during enzymatic catalysis, (*ii*) clarify the role of many conserved active site residues for which no apparent function could be assigned, so far, and (*iii*) be of high relevance for rational enzyme design. Similar to the concept of 'negative design' already used in the de novo design of proteins[31], including 'negative catalysis' elements as a basic design constraint could pave the way to new efficient and precise catalysts in biology and chemistry.

3.5. Methods

3.5.1. Chemicals

$NADP^+$ and NADPH (as sodium salts) were purchased from Roth AG, crotonic anhydride were purchased from Sigma Aldrich AG, coenzyme A from Roche Diagnostics. Crotonyl-CoA and C2-ene adduct were synthesized and purified as reported earlier[8]. All salts and solvents were of analytical grade.

3.5.2. (4*R*)-[4-$^2H_1$]NADPH synthesis

(4*R*)-[4-$^2H_1$]NADPH was synthesized according to previously published protocols[28]. Synthesis products were confirmed by analytical HPLC and comparison to analytical standards of NADPH. The deuterium incorporation into (4*R*)-[4-$^2H_1$]NADPH was confirmed to be >95% by mass spectrometry. (4*R*)-[4-$^2H_1$]NADPH was dissolved in 0.5 mM NaOH (aq) after lyophilization and stored at -80 °C until it was used.

3.5.3. Cloning and mutagenesis

The plasmid containing Etr1p WT and Ccr were constructed in previous studies. Plasmids carrying desired point mutations were generated using the QuikChange® Site-Directed Mutagenesis with 60 ng of template plasmid, Phusion® DNA polymerase (Thermo Scientific™) and the primers listed in **Table S1**. The resulting mutated plasmids were confirmed by sequencing (Eurofins AG, Germany). Plasmids carrying the correct mutations were transformed into *E.coli* BL21(DE3) (Invitrogen™) for protein expression.

3.5.4. Protein expression and purification

Etr1p and variants were expressed according to previously described protocols[9] and desalted into 150 mM NaCl, 20 mM TrisHCl at pH 7.9 (at room temperature) unless otherwise stated. Ccr Wt and T195V was expressed according to previously described methods[9] and desalted into the same buffer as Etr1p and variants. Protein concentrations were determined with a Thermo Scientific™ NanoDrop 2000 and extinction coefficients calculated with the ExPASy server (PMID: 10027275).

3.5.5. Stopped flow spectroscopy

Measurements were recorded on a thermo-stated stopped flow unit (SFM-20 connected to a MOS-200, equipped with a Xe(Hg)-lamp and a TC-100/10 cuvette, Bio-Logic Science Instruments SAS, Claix, France) set to 30 °C. Syringe 1 contained 320 μM NADPH and 240 μM crotonyl-CoA in 20 mM Tris-HCl, 150 mM NaCl buffer pH 7.9 and syringe 2 contained 5.0 μM of Etr1p WT or T175V in the same buffer. Data were collected at 340 nm every 1 ms for the first 0.3 s, every 500 ms till 10 s and every 1 s up to a total of 600 s. To estimate initial turnover rates of Etr1p and its variants, the quasi-linear phase of the curve in the first 10 ms was fitted with linear regression to give a lower estimate for the initial rate constant.

3.5.6. Mass spectrometry

Kinetic isotope effects on protonation were measured with ultra-performance liquid chromatography-high resolution heated electrospray ionization mass spectrometry (UPLC HR HESI-MS) analysis a Q-Exactive Plus™ mass spectrometer (ThermoFisher Scientific, San Jose, CA, USA) connected to a Ultimate 3000 (Dionex, Sunnyvale, CA, USA) UPLC system.

3.5.7. Isothermal calorimetry

Proteins were purified as described above and concentrated to about 100 μM (±10 μM). The final concentration was determined in quadruplicate as described above and this value was used for further calculations. The ITC cell of a VP-ITC (GE Healthcare) was loaded with 220 μl of a degassed protein solution and the syringe was filled with a 1.5 mM degassed NADPH solution (the concentration was determined by the absorption at 340 nm with a Thermo Scientific™ NanoDrop 2000 using $\varepsilon_{340\ nm\ NADPH}$ = 6.2 cm^{-1} mM^{-1}) in the same buffer (150 mM NaCl, 20 mM TrisHCl at pH 7.9). The titration was spread over 20 injections with one initial dummy injection. The values association constants were determined in duplicate.

3.5.8. KIE on protonation

Reactions for the measurement of kinetic isotope effect on protonation in Etr1p WT and variants contained 100 mM Na_2HPO_4 (pH 7.9), 0.25 mM NADPH, 1.25 μM crotonyl-CoA, and a variable amount of D_2O to give the desired percentage of D_2O, between 7 and 80 %. The enzyme was equilibrated with the deuterated buffer for at least 120 min (to allow for hydrogen exchange with the buffer) before the addition of the last substrate, crotonyl-CoA, from a stock solution of 187.5 μM (to a final volume of 150 μL). Variant enzymes were added at 125 nM and the WT at 25 nM final concentration. Reactions were incubated at room temperature for 1 hour before diluting 10 times with 25 mM ammonium formate with 2% MeOH (pH 3.5) prior to analysis. Samples were analyzed with ultra-performance liquid chromatography-high resolution heated electrospray ionization mass spectrometry (UPLC HR HESI-MS) analysis on a Q-Exactive Plus™ mass spectrometer (ThermoFisher Scientific™, San Jose, CA, USA) connected to a Ultimate 3000 (Dionex™, Sunnyvale, CA, USA) UPLC system. Chromatographic separation prior to mass analysis was done with a Kinetex® 1,7 μm XB C-18 100 Å, 50 x 2.1 mm column (Phenomenex®, Torrance, CA, USA). Injection volume was 5 μL. Reaction products were separated using a mobile phase system comprised of 25 mM ammonium formate pH 8.1 (A) and methanol (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 0.55 mL min^{-1}: 0 min, 5%; 1.5 min, 5%; 6.5 min, 95%; 7.5 min, 95%; 7.7 min, 5%; 10.5 min 5%. The $[butyryl\text{-}CoA + H]^+$ mother ion was isolated (840 ± 3.5 m/z) and fragmented with argon at 25 eV. The isotopic pattern on the $[butyryl\text{-}pantothenate + H]^+$ fragment-ion (331.150 m/z) was measured to determine the deuterium incorporation content.

The kinetic isotope effect on protonation, ${}^{D}k_{obs}$, was determined by fitting the dependence of the ratio of product, butyryl-CoA, with an incorporated deuterium ($product_D$) to product with an incorporated hydrogen ($product_H$) versus the ratio of D_2O (f_{D2O}) to H_2O (1-f_{D2O}) in the buffer with the following equation:

$$\frac{Product_D}{Product_H} = {}^{D}k_{obs} \cdot \frac{f_{D2O}}{1-f_{D2O}}$$

The $product_H/product_D$ ratio was calculated by comparing the isotopologue distributions of different samples against a sample with a natural abundance isotopologue distribution as reference. The $product_D/product_H$ ratio of a sample was the calculated with the following equation from the integrated traces of the specific ions:

$$\frac{Product_D}{Product_H} = \frac{[M+1] - [M+0] \cdot (\frac{[M+1]}{[M+0]})}{[M+0]}$$

[M+0] and [M+1] represent the integrals of the [M+0] and [M+1] peaks and ([M+1]/[M+0]) reference represents the ratio of [M+0]/[M+1] in the reference sample with a natural abundance isotopologue distribution. All data was analyzed with idms_quan-0.0.9-py2-none-any.whl package

(www.emzed.ethz.ch) that was especially developed to standardize and automatize the data analysis of kinetic isotope data with EmZed32 using the equations specified above. Measurements were performed in duplicate.

3.5.9. Detection of C4-adduct

To a reaction mixture of 1 µl 500 mM Na_2HPO_4 (pH 7.9), 1 µl 200 mM NADPH and 1 µl 200 mM crotonyl-CoA; 1 µl 60 µM ($3.7 \cdot 10^{-4}$ equivalents) of Etr1p WT or T175V (in 150 mM NaCl, 20 mM TrisHCl pH 7.9) was added. The assays for Ccr WT and Ccr T195V contained 1 µl of 30 mM enzyme, 1 µl 60 µM crotonic anhydrase and additionally 1µL of 200 mM $KHCO_3$. The assay was reacted for 2 minutes on ice before being quenched with 4 µl acetonitrile and immediately injected for analysis.

The calibration curve for the C4-adduct was made with a dilution series made with pure C4-adduct isolated as previously described[16]. A stock of 12 µM of C4-adduct was made by dissolving the lyophilized C4-adduct into 100 mM Na_2HPO_4 (pH 7.9). This stock was diluted to 3.0 µM, 0.60 µM, 0.125 µM and 0.025 µM and dilutions were kept on liquid nitrogen till between diluting and injecting into the UPLC-MS. C4-adduct eluted at 3.1 min with the employed gradient (see below). The extracted ion chromatogram of m/z = 791.123 ± 0.005 (calc. mass m/z =791.1235, the mass difference of 1ppm is within the mass accuracy of the MS) was smoothed with the default algorithm (Gaussian) of MassHunter Qualitative Analysis software (Agilent Technologies Inc. Santa Clara, CA, USA).

Samples were measured on a 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies Inc. Santa Clara, CA, USA) equipped with an electrospray ionization source set to positive ionization mode through a 1290 Infinity UPLC (Agilent Technologies Inc. Santa Clara, CA, USA). Chromatographic separation was carried out on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex XB-C18, Phenomenex) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 (A) and methanol (B) with following gradient at a flow rate of 250 µl/min: 0 min 0% B; 1 min 0% B, 6 min 80% B; 8 min 80% B; 8.5 min 0 %B.

Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 l/min, 225 °C) and sheath gas (12 l/min, 350°C). MS data were acquired with a scan range of 500-1200 m/z. Targeted MS/MS was carried out on 791 m/z (z=2) with an isolation width of approximately 4 m/z and a collision energy of 20 eV with an acquisition time of 1000 ms/spectrum.

LC-MS and LC-MS/MS data were analyzed using MassHunter Qualitative Analysis software (Agilent Technologies).

3.5.10. Spectrophotometric enzyme assays

Assays were carried out on a Carry-60 UV/Vis spectrometer (Agilent Technologies Inc. Santa Clara, CA, USA) at 30 °C using quartz cuvettes (10-mm path-length; Hellma® (Germany)). All assays were carried out in 100 mM Na_2HPO_4, pH 7.9. For determination of the kinetic parameters starting from NADPH and crotonyl-CoA assays contained 200 µM crotonyl-CoA and were started by adding enzyme to the following end concentrations: 3 nM Etr1p WT, 5 nM Etr1p S70A, 5 nM Etr1p T324A, 10 nM Etr1p S70A T324A, 6.8 µM Etr1p Y79F or $1.1 \cdot 10^2$ nM Etr1p T175V. Kinetic parameters were determined by varying the NADPH or (4*R*)-[4-$^2H_1$]NADPH concentration (between 1 and 64 µM) and following disappearance of NADPH at 340 nm using an absorption coefficient of $\varepsilon_{340\,nm} = 6.2\ cm^{-1}\ mM^{-1}$ [35]. For each concentration chosen, at least triplicate measurements were performed. For Etr1p Y79F $\Delta\varepsilon_{340}$ nm = 2.5 $cm^{-1}\ mM^{-1}$ ($\varepsilon_{340\,nm\ NADPH} = 6.2\ cm^{-1}\ mM^{-1}$, $\Delta\varepsilon_{340\,nm\ C4\text{-}adduct} = 3.7\ cm^{-1}\ mM^{-1}$ was used. Assays for determination of kinetic parameters on C2-ene adduct were by directly adding lyophilized, purified C2-ene adduct and following its disappearance at 370 nm using an extinction coefficient of $\varepsilon_{370\,nm} = 6.9\ cm^{-1}\ mM^{-1}$ [9]. Kinetic

parameters with the C2-ene adduct were determined from at least 18 independent measurements at varying concentrations. The KIEs on hydride transfer were calculated from $^{H}k_{cat}/^{D}k_{cat}$ and $(^{H}k_{cat} \cdot K_{m\ D})/(^{D}K_{cat} \cdot K_{m\ H})$ ratios. For all kinetic parameters determined, the corresponding errors are given as 95% confidence intervals.

To investigate whether the C4-adducts formed in Ccr T195V could inhibit Etr1p WT, a reaction with 200 μM NADPH and 200 μM crotonyl-CoA in 100 mM Na_2HPO_4 (pH 7.9) was started by 1 volume percent of either 0.5 μM Ccr WT stock 170 μM Ccr T195V stock. The reaction was allowed to develop for 30 s before adding 1 volume percent of a 50 μM Etr1p WT stock.

3.5.11. Protein crystallization and structural analysis

The purified enzymes were crystallized in sitting drops. Etr1p was crystallized at 15 mg/mL with 5 mM NADP and 5 mM of crotonyl-CoA in the gel filtration buffer (100 mM NaCl, 20 mM TrisHCl pH 7.9). In most experiments, 1 μl of enzyme solution and 1 μl of reservoir solution were mixed in the wells of a crystallization plate (CombiClover Junior Plate, Jena Bioscience). Two different crystalline forms appeared after one week at 25°C. Cubic crystal belonging to the trigonal form appears in 2.0 M $(NH_4)_2SO_4$, 100 mM N-(2-acetamido)-2-iminodiacetic acid (ADA)/NaOH (pH 6.4). Rod shape crystal belonging to the monoclinic form appears in 1.85 M $(NH_4)_2SO4$, 100 mM ADA pH 6.4 using streak seeding to enhance crystals size. Crystals were immersed in a solution containing 30% (v/v) glycerol, 2.0 M $(NH_4)_2SO_4$, 100 mM ADA pH 6.4 prior to freezing in liquid nitrogen to prevent water-crystal formation.

Data were collected at the X10SA beamline from the Swiss Light Synchrotron (Villigen) on a Pilatus 6M (Dectris AG, Switzerland). Datasets were processed with XDS[36] and scaled with SCALA from the ccp4 suite[37]. Both crystal structures were solved by molecular replacement with Phaser from PHENIX package[38] using the PDB: 4WAS as a template. The starting models for molecular replacement were first refined with REFMAC5[39]. All models were then manually rebuilt with COOT[40] and further refined with PHENIX[38]. For both structure restraints for non-crystallographic symmetry (NCS) and a translation–liberation–screw-rotation (TLS) were applied. Final models were validated through the MolProbity server[41] (http://molprobity.biochem.duke.edu). Data collection and refinement statistics are reported in **Table S2**. Figures were generated and rendered with PyMOL (Schrödinger, LLC). The standard deviations between the structures were calculated with the Superpose program of the ccp4 suite

3.6. References

1. Khersonsky, O.; Tawfik, D. S., *Annu Rev Biochem* **2010,** *79*, 471-505.
2. Zhu, X. G.; Long, S. P.; Ort, D. R., *Curr Opin Biotech* **2008,** *19* (2), 153-159.
3. Browne, C. A.; Campbell, I. D.; Kiener, P. A.; Phillips, D. C.; Waley, S. G.; Wilson, I. A., *Journal of molecular biology* **1976,** *100* (3), 319-43.
4. Richard, J. P., *Biochemistry* **1991,** *30* (18), 4581-5.
5. Iyengar, R.; Rose, I. A., *Biochemistry* **1981,** *20* (5), 1223-9.
6. Major, D. T.; Freud, Y.; Weitman, M., *Curr Opin Chem Biol* **2014,** *21*, 25-33.
7. Retey, J., *Angew Chem Int Edit* **1990,** *29* (4), 355-361.
8. Bar-Even, A.; Milo, R.; Noor, E.; Tawfik, D. S., *Biochemistry* **2015,** *54* (32), 4969-4977.
9. Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat Chem Biol* **2014,** *10* (1), 50-U85.
10. Saraste, M., *Science* **1999,** *283* (5407), 1488-1493.
11. Harwood, J. L., *Annu Rev Plant Phys* **1988,** *39*, 101-138.
12. Quemard, A.; Sacchettini, J. C.; Dessen, A.; Vilcheze, C.; Bittman, R.; Jacobs, W. R.; Blanchard, J. S., *Biochemistry* **1995,** *34* (26), 8235-8241.
13. Stewart, M. J.; Parikh, S.; Xiao, G. P.; Tonge, P. J.; Kisker, C., *Journal of molecular biology* **1999,** *290* (4), 859-865.
14. Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B., *Nature* **2001,** *409* (6817), 258-268.
15. Mansell, D. J.; Toogood, H. S.; Waller, J.; Hughes, J. M.; Levy, C. W.; Gardiner, J. M.; Scrutton, N. S., *ACS catalysis* **2013,** *3* (3), 370-379.
16. Rosenthal, R. G.; Vogeli, B.; Quade, N.; Capitani, G.; Kiefer, P.; Vorholt, J. A.; Ebert, M. O.; Erb, T. J., *Nat Chem Biol* **2015,** *11* (6), 398-+.
17. Airenne, T. T.; Torkko, J. M.; Van den Plas, S.; Sormunen, R. T.; Kastaniotis, A. J.; Wierenga, R. K.; Hiltunen, J. K., *Journal of molecular biology* **2003,** *327* (1), 47-59.
18. Torkko, J. M.; Koivuranta, K. T.; Kastaniotis, A. J.; Airenne, T. T.; Glumoff, T.; Ilves, M.; Hartig, A.; Gurvitz, A.; Hiltunen, J. K., *Journal of Biological Chemistry* **2003,** *278* (42), 41213-41220.
19. Miwa, G. T.; Garland, W. A.; Hodshon, B. J.; Lu, A. Y.; Northrop, D. B., *The Journal of biological chemistry* **1980,** *255* (13), 6049-54.
20. Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L. C.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L., *Structure* **2015,** *23* (12), 2213-2223.
21. Almarsson, O.; Bruice, T. C., *Journal of the American Chemical Society* **1993,** *115* (6), 2125-2138.
22. Burgner, J. W.; Ray, W. J., *Biochemistry* **1984,** *23* (16), 3626-3635.
23. Benach, J.; Atrian, S.; Gonzalez-Duarte, R.; Ladenstein, R., *Journal of molecular biology* **1999,** *289* (2), 335-355.
24. Bull, H. G.; GarciaCalvo, M.; Andersson, S.; Baginsky, W. F.; Chan, H. K.; Ellsworth, D. E.; Miller, R. R.; Stearns, R. A.; Bakshi, R. K.; Rasmusson, G. H.; Tolman, R. L.; Myers, R. W.; Kozarich, J. W.; Harris, G. S., *Journal of the American Chemical Society* **1996,** *118* (10), 2359-2365.
25. Maier, T.; Leibundgut, M.; Ban, N., *Science* **2008,** *321* (5894), 1315-1322.
26. Kwan, D. H.; Leadlay, P. F., *ACS chemical biology* **2010,** *5* (9), 829-38.
27. Quade, N.; Huo, L. J.; Rachid, S.; Heinz, D. W.; Muller, R., *Nat Chem Biol* **2012,** *8* (1), 117-124.
28. Erb, T. J.; Brecht, V.; Fuchs, G.; Muller, M.; Alber, B. E., *P Natl Acad Sci USA* **2009,** *106* (22), 8871-8876.
29. Ames, B. D.; Nguyen, C.; Bruegger, J.; Smith, P.; Xu, W.; Ma, S.; Wong, E.; Wong, S.; Xie, X. K.; Li, J. W. H.; Vederas, J. C.; Tang, Y.; Tsai, S. C., *P Natl Acad Sci USA* **2012,** *109* (28), 11144-11149.
30. Erb, T. J.; Berg, I. A.; Brecht, V.; Muller, M.; Fuchs, G.; Alber, B. E., *P Natl Acad Sci USA* **2007,** *104* (25), 10631-10636.
31. Vlasie, M. D.; Banerjee, R., *Biochemistry* **2004,** *43* (26), 8410-8417.
32. Toney, M. D., *Bba-Proteins Proteom* **2011,** *1814* (11), 1407-1418.

33. Toney, M. D., *Arch Biochem Biophys* **2005,** *433* (1), 279-287.
34. Ghanem, M.; Murkin, A. S.; Schramm, V. L., *Chem Biol* **2009,** *16* (9), 971-9.
35. Dawson, R. M. C., *Data for Biochemical Research*. Clarendon Press: Oxford, 1986.
36. Kabsch, W., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 2), 125-32.
37. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., *Acta crystallographica. Section D, Biological crystallography* **2011,** *67* (Pt 4), 235-42.
38. Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Headd, J. J.; Moriarty, N. W.; Richardson, J. S.; Richardson, D. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., *Journal of applied crystallography* **2010,** *43* (Pt 4), 669-676.
39. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., *Acta crystallographica. Section D, Biological crystallography* **1997,** *53* (Pt 3), 240-55.
40. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 4), 486-501.
41. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 1), 12-21.
42. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., *Genome Res* **2004,** *14* (6), 1188-1190.

3.7. Supplementary Information

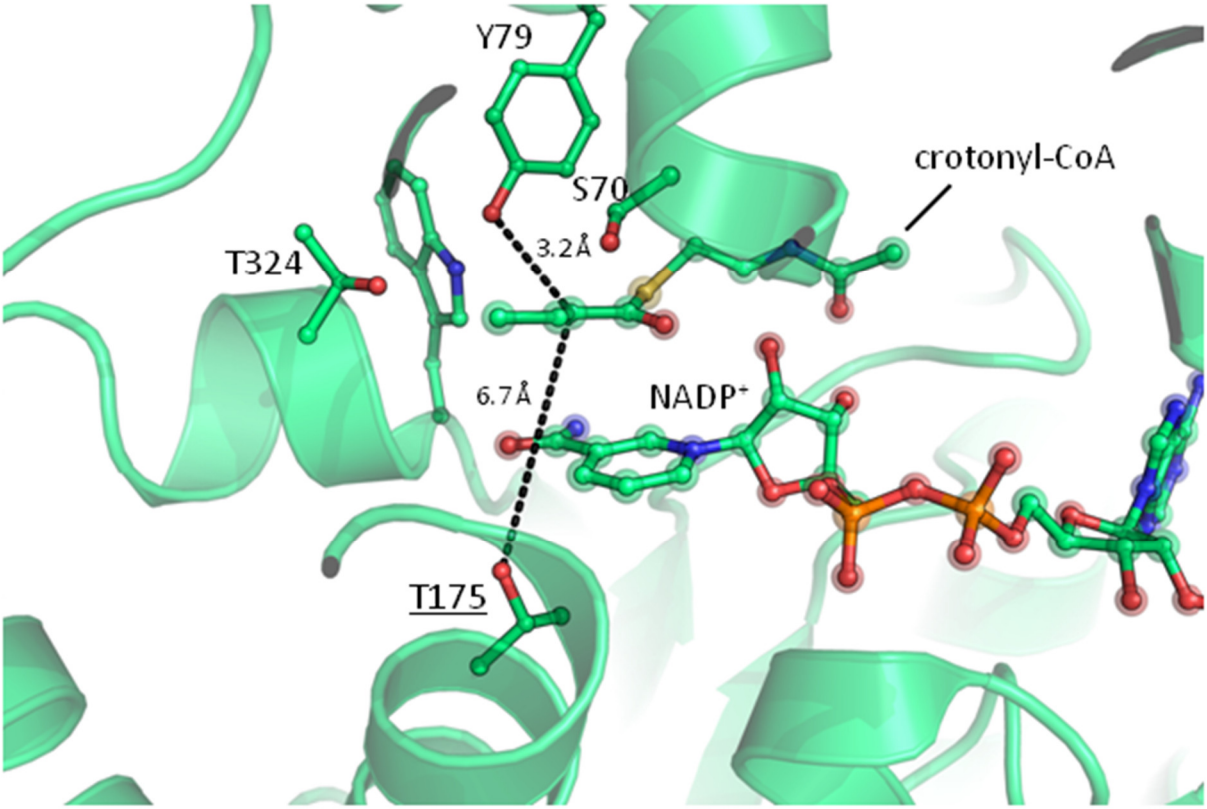


Figure S1. Active site of Etr1p. Distances from Y79 (3.2 Å) and T175 (6.7 Å) to the Cα of crotonyl-CoA, the protonation site. The distance from T175 to the Cα of crotonyl-CoA is too large to be directly involved in the protonation step.

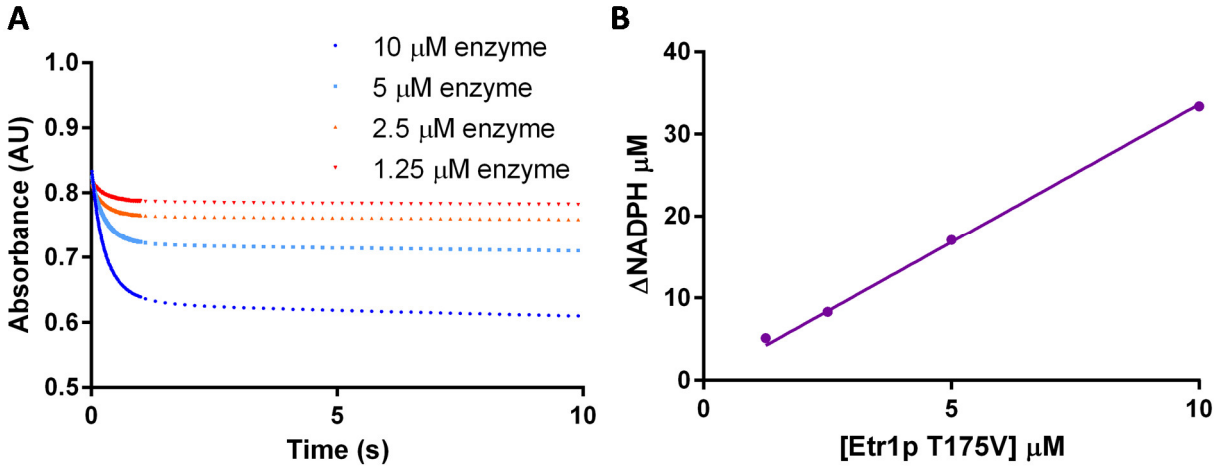


Figure S2. Linearity of the burst phase in Etr1p T175V with amount of enzyme. A) Stopped flow traces of the burst phase with the indicated amount of enzyme. **B)** The amount of NADPH that is consumed in the first 2 seconds of the reaction plotted versus the amount of enzyme. The slope of the line is 3.27± 0.07.μM NADPH per μM enzyme, indicating multiple turnovers takes place before the enzyme is inhibited.

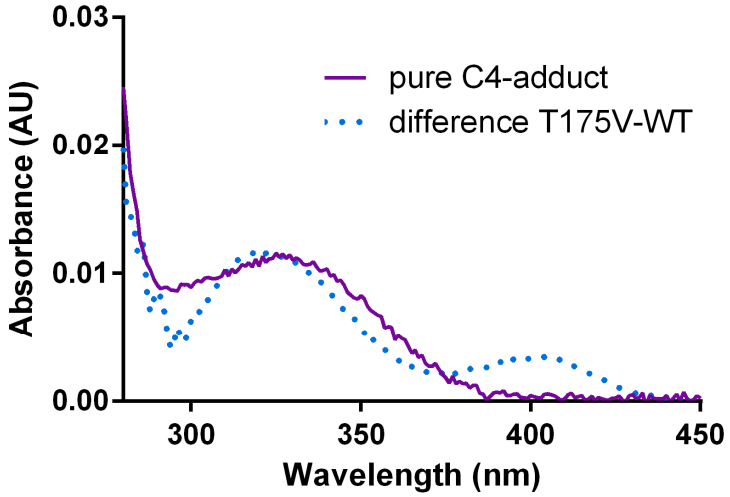


Figure S3. Difference spectroscopy demonstrates C4-adduct formation in the Etr1p T175V mutant. Spectrum of pure C4-adduct (purple) and difference spectrum (blue dotted lines) of two assays containing either 4 µM Etr1p WT or 4 µM T175V. Both assays contained 40 µM NADPH and 50 µM crotonyl-CoA in 100mM Na2HPO4 pH 7.9. Spectra were taken after both reactions depleted NADPH (within manual mixing time).

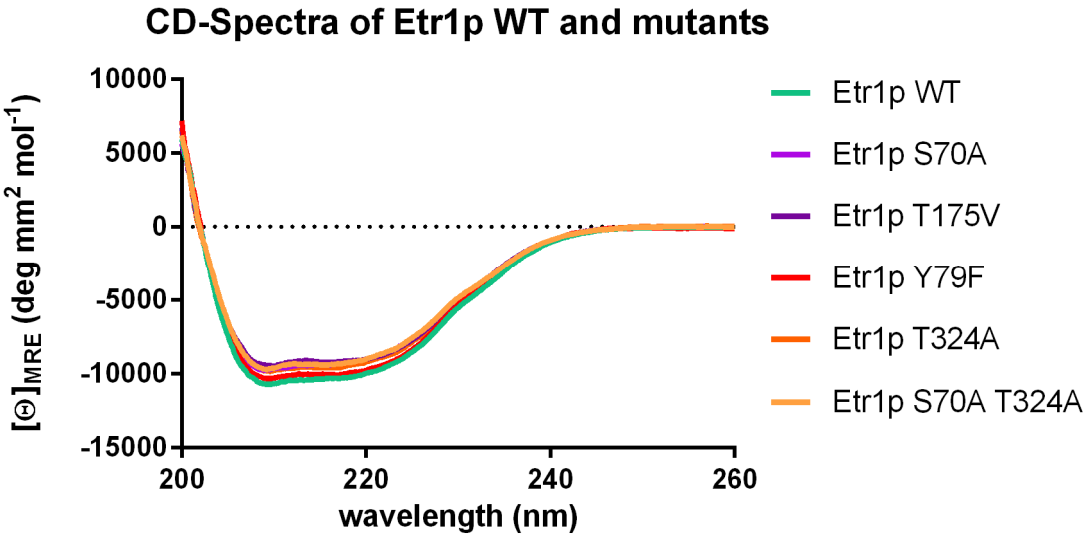


Figure S4. Circular dichroism spectra of Etr1p WT and mutants employed in this study. Each measurement was performed with 10 µM of protein and was performed in triplicate to yield the mean representative scans. Molar ellipticity was subtracted from a control buffer scan. The spectra were collected on a JASCO 815 CD spectrometer scanned from 200–260 nm in a 1 mm quartz cuvette.

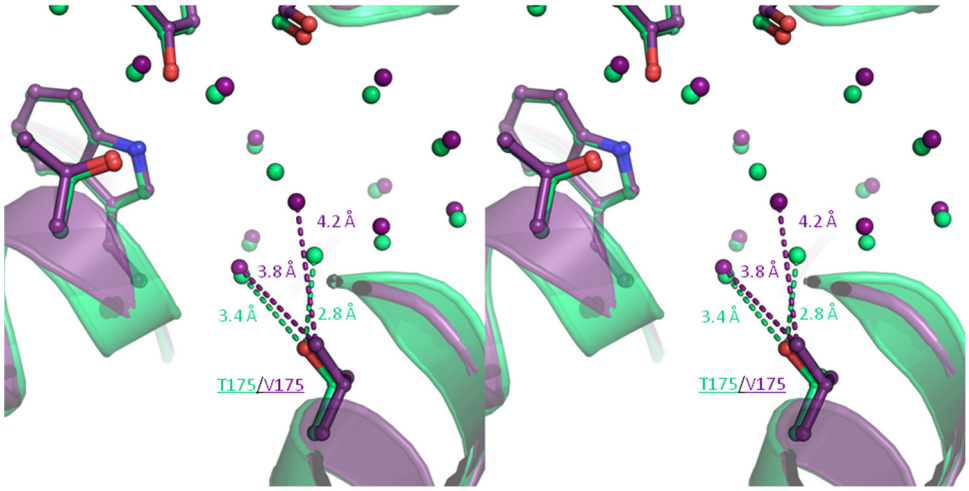


Figure S5. Stereo image of the active site of apo-Etr1p WT and T175V. Most of the water network in the active site superimposes between the WT and T175, only the crystal water in direct contact with position 175 is shifted in the mutant.

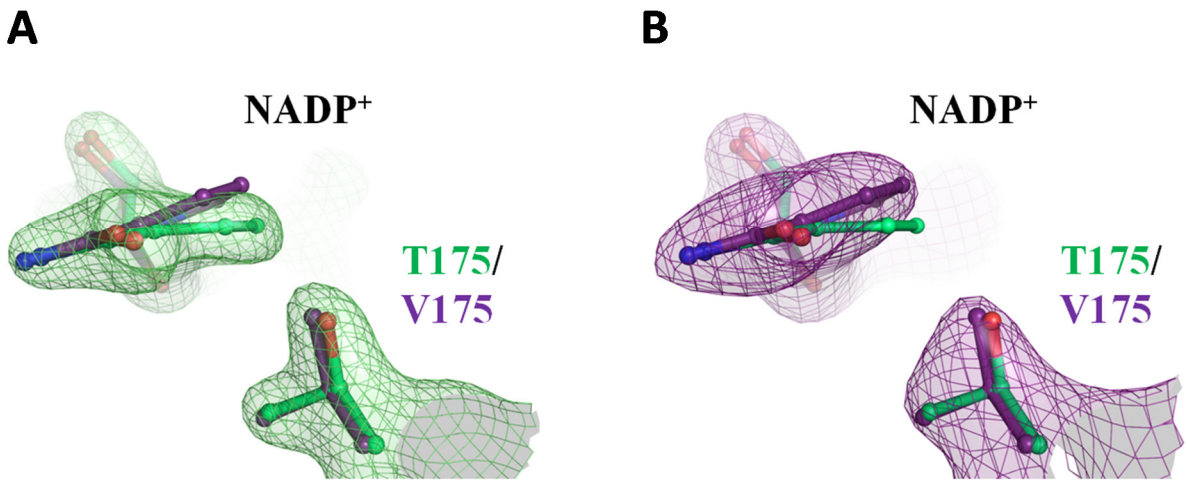


Figure S6. Superimposition of NADP⁺ and amino acid residue 175 for Etr1p WT (PDB 4W99) and the T175V variant (PDB: 5LBX) including electron densities. Overlay of NADP⁺ and the amino acid residue 175 in comparison to electron density maps. **A)** $2F_o$-F_c electron density map contoured at 1.5 σ, as well as ball and stick representation of the WT in green, with the T175V variant overlaid as ball and stick in purple. **B)** $2F_o$-F_c electron density map contoured at 1.5 σ, as well as ball and stick representation for the T175V variant in purple, with the WT overlaid only as ball and stick in green.

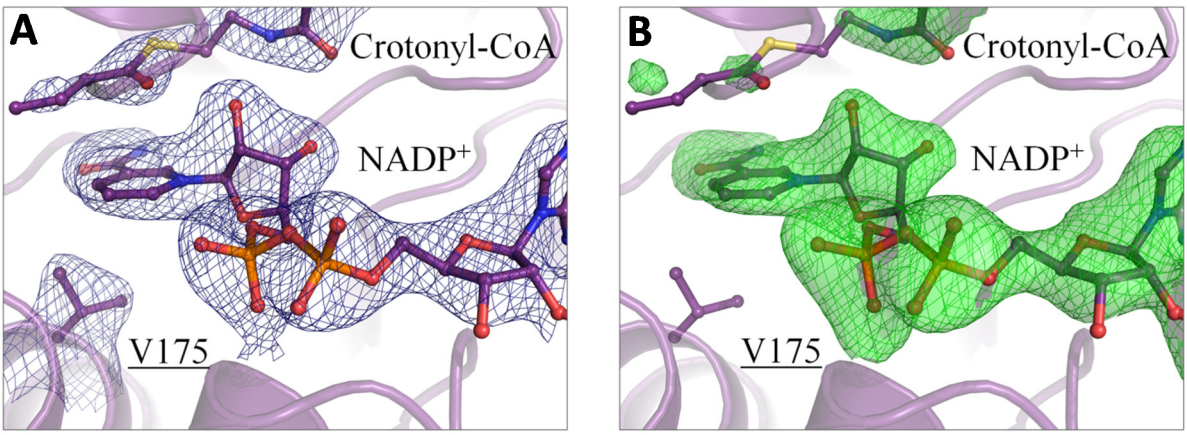


Figure S7. $2F_o$-F_c electron density maps for NADP⁺ and crotonyl-CoA bound at T175V after and before refinement. A) $2F_o$-F_c electron density map for the T175V variant with NADP⁺ and crotonyl-CoA contoured at 1.0 σ. **B)** F_o-F_c electron density map with NADP⁺ and crotonyl-CoA omitted prior to refinement. The difference map is contoured at 3.0 σ.

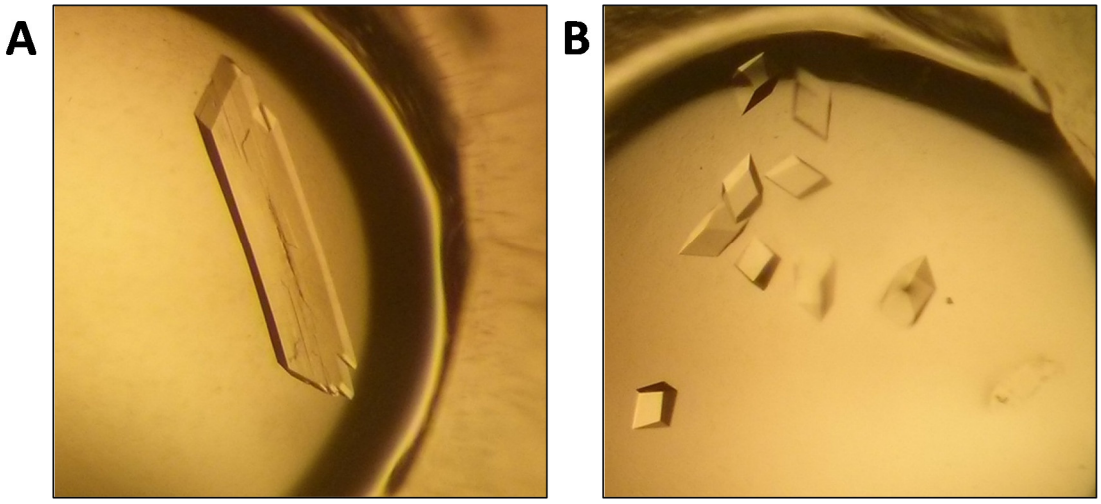


Figure S8. Picture of the **A)** trigonal and **B)** monoclinic crystalline forms (see **Table S2**).

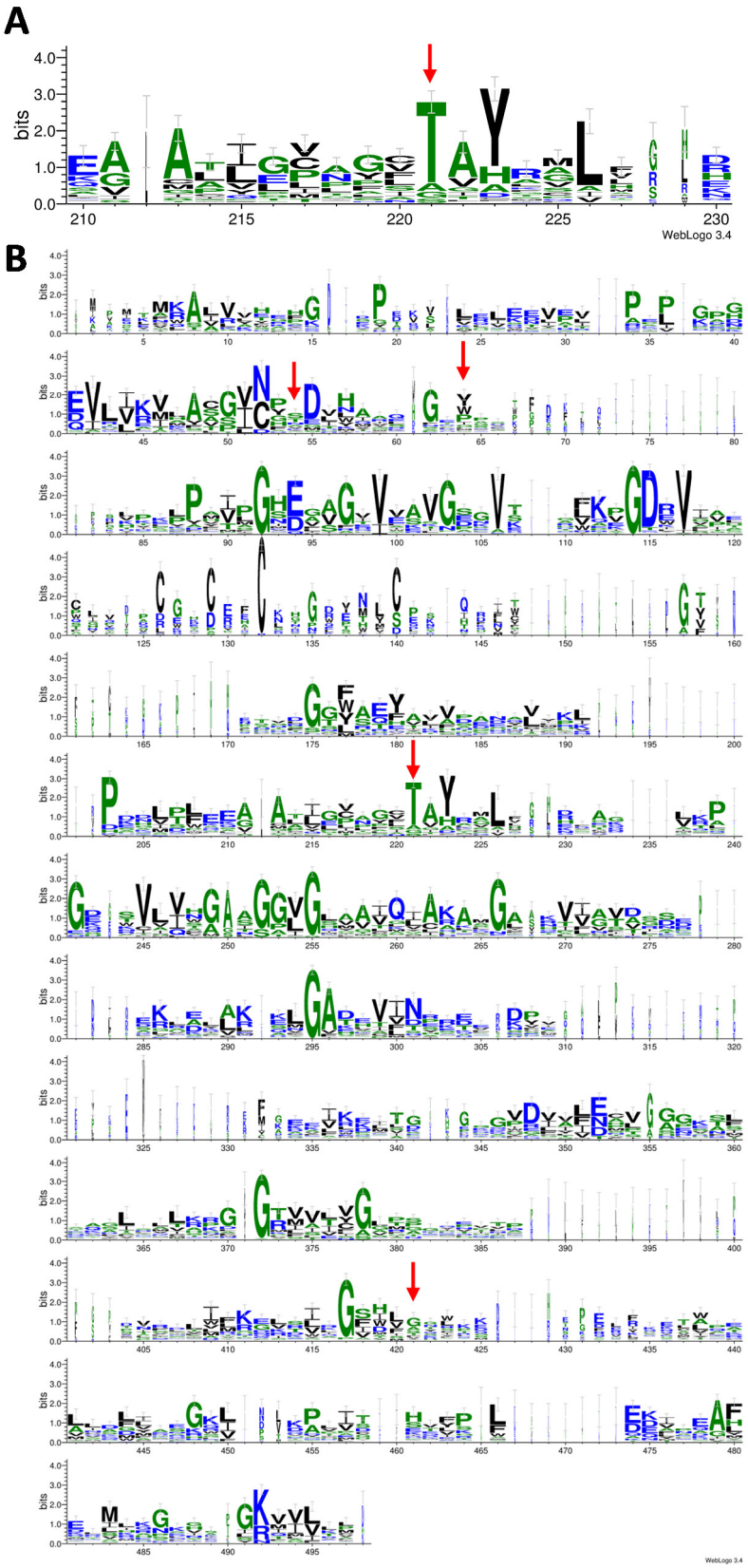


Figure S9. Weblogo analysis of conserved residues in the MDR enzyme superfamily. Sequence logo of **A)** the residues around the conserved threonine 175 (red arrow) and **B)** the complete sequence alignment. The sequence numbering is shifted due to gaps in the alignment and the presence of a histidine-tag in the crystal structure that is used for the numbering throughout the paper. S70 is at position 54, Y79 at position 64, T175 at position 221, T324 at position 421. Sequences from referenceX were aligned with MAFFT v7.273 (with iterative refinement method E-INS-i) and the sequence logo was created with the Weblogo 3.4 server [42]. Columns with many gaps or unknown residues are narrow.

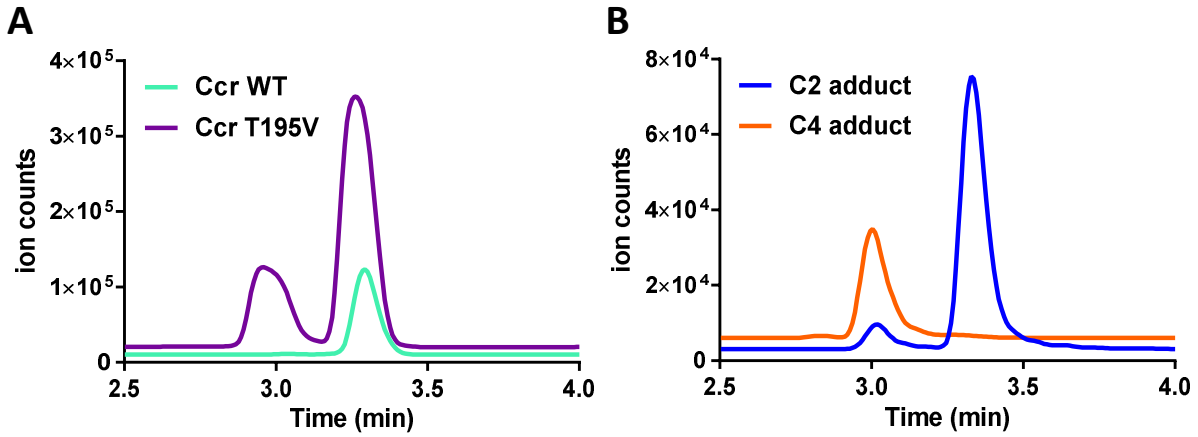


Figure S10. Detection of C4-adduct in the Ccr T195V variant and WT. A) UPLC-MS analysis of Ccr WT (green) and Ccr T195V variant (purple) containing 1 µl 30µM enzyme, 1 µl 500 mM Na_2HPO_4 (pH 7.9), 1 µl 200 mM NADPH, 1 µl 200 mM crotonyl-CoA, 1 µl 60 µM crotonic anhydrase and 1µL of 200 mM KHCO3. The T195V variant shows accumulation of a C4-adduct, while the Ccr WT did not accumulate detectable amounts of the C4-adduct. Both variants show accumulation of the catalytically competent C2-adduct. **B)** UPLC-MS analysis of authentic standards for the C2 adduct (blue) and the C4 adduct (orange). Data shown are representatives of at least three independent experiments.

Table S1. List of primers used to generate Etr1p and Ccr variants

Enzyme variant	Primer Name	Sequence
Etr1p S70A	Etr1p S70A fwd	CCGGTCAACCCGGCCGACATCAACCAG
	Etr1p S70A rev	CTGGTTGATGTCGGCCGGGTTGACCGG
Etr1p T324A	Etr1p T324A fwd	GGGCTTCTGGGTGGCCGAGCTGCTCAAG
	Etr1p T324A rev	CTTGAGCAGCTCGGCCACCCAGAAGCCC
Etr1p Y79F	Etr1p Y79F fwd	GATCCAGGGCGTCTTTCCGAGCAAGCCG
	Etr1p Y79F rev	CGGCTTGCTCGGAAAGACGCCCTGGATC
Etr1p T175V	Etr1p T175V fwd	CATCTCGGTCAATCCCCTGGTGGCCTACCTCATGCTCAC
	Etr1p T175V rev	GTGAGCATGAGGTAGGCCACCAGGGGATTGACCGAGATG
Ccr T195V	Ccr T195V fwd	GCTGACGCTCGCCGTCGCCTACCGCATGC
	Ccr T195V rev	GCATGCGGTAGGCGACGGCGAGCGTCAGC

Supplementary Table 3. Etr1p T175V data collection and refinement statistics for both, the monoclinic and trigonal crystalline forms (see **Figure S8**).

	Etr1p T175V monoclinic form	**Etr1p T175V trigonal form**
Data collection		
Wavelength (Å)	1.73913	1.73913
Space group	$P2_1$	$P3_12$
Resolution (Å)	49.47 – 2.10 (2.21 – 2.10)	47.04 – 2.50 (2.64 – 2.50)
Cell dimensions		
a, b, c (Å)	67.42, 101.86, 81.59	93.50, 93.50, 231.20
α, β, γ (°)	90, 101, 90	90, 90, 120
R_{merge} (%)[a]	5.9 (43.0)	12.4 (151.1)
R_{pim} (%)[a]	3.9 (28.6)	4.4 (55.7)
I/σ_I[a]	11.8 (2.5)	12.2 (1.5)
$CC_{1/2}$	99.6 (80.8)	99.7 (67.7)
Completeness (%)[a]	96.2 (91.5)	100.0 (100.0)
Redundancy[a]	3.3 (3.1)	8.6 (7.7)
Refinement		
Resolution (Å)	49.47 – 2.10	43.34 – 2.5
Number of reflections	60553	41370
R_{work}/R_{free}[b] (%)	15.85 / 19.70	19.16 / 21.96
Number of atoms		
Protein	5610	5586
Ligands/ions	137	209
Solvent	525	157
Average B-factors (Å^2)	40.42	53.04
Molprobity clash score, all atoms	1.93 (100th percentile)	2.8 (100th percentile)
Ramachandran plot		
Favoured regions (%)	97.94	96.69
Outlier regions (%)	0	0
rmsd[c] bond lengths (Å)	0.009	0.004
rmsd[c] bond angles (°)	1.097	0.787
PDB code	5LB9	5LBX

[a] Values relative to the highest resolution shell are within parantheses. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement. [c] msd, root mean square deviation.

CHAPTER IV

InhA, the enoyl-thioester reductase from *M.tuberculosis* forms a covalent reaction intermediate

Authors:
Bastian Vögeli*, Raoul G. Rosenthal*, Gabriele M. M. Stoffel*, Tristan Wagner, Patrick Kiefer, Niña S. Cortina, Seigo Shima, Tobias J. Erb

* These authors contributed equally to this work

Submitted to:
eLife

Author contributions:
B.V., R.G.R., G.M.M.S. and T.J.E. conceived and designed all experiments, with the exception of crystallography experiments, which were designed together with T.W and S.S. Enzyme kinetic assays and stopped-flow measurements were performed B.V. G.M.M.S. isolated the C2 ene adduct and measured NMR spectra. MS experiments were performed by B.V. and R.G.R. and analysed together with N.S. Cortina. B.V., R.G.R. and T.J.E. wrote the paper.

4. InhA, the enoyl-thioester reductase from *M. tuberculosis* forms a covalent reaction intermediate

4.1. Abstract

The enoyl thioester reductase InhA catalyzes an essential step in fatty acid biosynthesis of *Mycobacterium tuberculosis* and is a key target of anti-tuberculosis drugs. Here we use enzyme mutagenesis, NMR, stopped flow spectroscopy and LC-MS to show that a covalent reaction intermediate is formed during catalysis by InhA. The reaction intermediate is a labile (2S)-C2-ene adduct between the NADH cofactor and the CoA thioester substrate. We isolated and used the adduct as a "molecular probe" to directly access the second half reaction of the catalytic cycle of InhA (*i.e.*, the proton transfer), independent from the first half reaction (*i.e.*, the initial hydride transfer). Molecular probing allowed us to interrogate the individual contributions of the conserved active site residues tyrosine 158 and threonine 196 to either half reaction. The natural tendency of InhA to form covalent C2-ene adduct at the active site calls for a careful reconsideration of the enzyme's reaction mechanism and provides the basis for the development of novel tools to study, manipulate and inhibit the catalytic cycle of InhA.

4.2. Introduction

Tuberculosis remains one of the deadliest infectious diseases [1]. With increasing multidrug resistant strains of *M. tuberculosis*, the cause of the disease, the need for next generation treatments increases [1]. The enoyl-acyl carrier protein (enoyl-ACP) reductase InhA is a major target for the clinically relevant antibiotics isoniazid and ethionamide [2]. In the light of emerging multidrug-resistant *M. tuberculosis* strains InhA remains a prime candidate for drug design [3]. Therefore, the recent decade has seen intensive research focusing on developing new inhibitors against this enzyme [4].

InhA catalyzes the NADH-dependent reduction of enoyl-ACP in the biosynthesis of fatty and mycolic acids, which form an essential component of the membrane and cell wall of *M. tuberculosis*, respectively [5-6]. The reaction mechanism is postulated to start with a direct hydride transfer from the C4 of NADH to the β-carbon of the enoyl-ACP, followed by formation of an enolate anion, which is subsequently protonated stereospecifically to the pro-(2*R*) position of the α-carbon [3, 7]. The source of the proton however remains unknown and was suggested to originate from a solvent water due to the lack of any protic amino acid residues positioned close enough to serve as proton donor [7]. Tyrosine 158 one of the only protic groups in the active site of InhA was shown by mutagenesis studies to be involved in catalysis. When mutated to phenylalanine (Y158F) the enzyme is reported to lose about one order of magnitude in catalytic efficiency. The hydroxyl group of Tyrosine 158 was proposed to provide electrophilic stabilization of the transition state(s) by hydrogen bonding to the carbonyl of the substrate (**Figure 1**) [8]. However, the Y158F variant did not show a significantly different solvent kinetic isotope effect (^{D2O}V) in comparison to the wild type (WT) enzyme (WT: $^{D2O}V = 1.5\pm0.2$; Y158F $^{D2O}V = 1.4\pm0.2$)[7].

Here we show that the Y158F variant forms and accumulates a previously overlooked covalent C2-ene adduct between the NADH cofactor and the CoA thioester substrate during catalysis. We use the isolated adduct as "molecular probe" to demonstrate that Tyr158 is not deficient in hydride transfer, but directly affected in the protonation step. Using stopped-flow spectroscopy and HPLC-MS we show that InhA WT forms the same intermediate also in its native catalytic cycle. The detection of the C2-ene adduct in the catalytic cycle of InhA calls for a careful reconsideration of the reaction mechanism

of InhA and related NAD(P)H dependent oxidoreductases. In addition, we describe novel tools to study the catalytic cycle of InhA and potential new avenues to eventually inhibit this important drug target.

4.3. Results

4.3.1. Tyrosine 158 and Threonine 196 are involved in catalysis

Analysis of a crystal structure containing the cofactor NAD^+ and a 16C-acyl-SNAC substrate analogue suggested a water molecule bound by Threonine 196 that could be involved in catalysis and particularly in proton donation (**Figure 1**) [3]. We therefore characterized active site variants of T196 as well as the previously suggested variants of Y158 using octenoyl-CoA as a substrate for the reaction (**Table 1**). The Y158F variant retained only 1.8 % of the wild type (WT) k_{cat} with a two-fold increase of the K_m for octenoyl-CoA, which corresponds well with previous studies [7]. For an Y158S mutant, contrary to a previously report [7], we saw an even stronger decrease in k_{cat} to only 0.11% of WT activity. The T196 variants on the other hand also showed a strong effect on catalysis. T196A retained 1.0 % and the larger, isosteric T196V variant retained 0.15 % of WT k_{cat}. The K_m for octenoyl-CoA was increased by more than 5-fold in both mutants while the K_m for NADH did not change significantly compared to WT. In summary, our kinetic characterization showed that both Y158 and T196 are involved in catalysis.

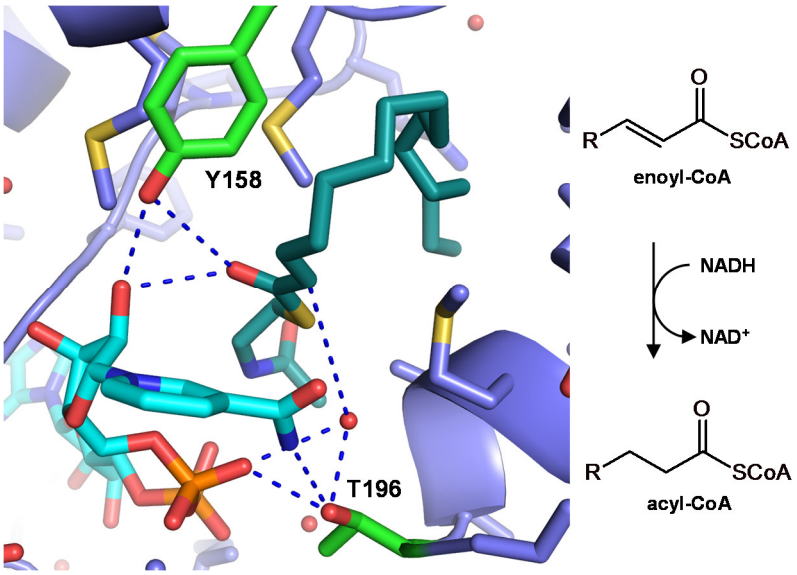


Figure 1. Active site architecture of InhA in tripartite structure containing NAD^+ (cyan) and C-16 acyl-substrate analogue (dark cyan) (PDB 1BVR[3]). Tyrosine 158 (green) is positioned for hydrogen bonding with the carbonyl of the acyl-thioester and the hydroxyl group of the NAD^+ ribose. Threonine 196 (green) is positioned below the carboxamide of NAD^+ and within hydrogen bond distance to the β-phosphate and a water molecule, which is positioned below the Cα of the substrate

4.3.2. Tyrosine 158 is essential for stereospecificity of protonation

To assess the function of Y158 and T196 in more detail we next used isotopic labeling to determine the protonation stereochemistry of InhA WT and its variants. We performed enzyme reactions in deuterated buffer and used a stereospecific oxidase Acx4 from *Arabidopsis thaliana* to quantify the incorporation pattern of deuterium at the 2 position (**Figure S1**). InhA WT incorporated the deuteron with an efficiency of 99 ± 1 % specifically into the *2R* position, which is in line with previous data [9]. The Y158F variant incorporated deuterium with an efficiency of 57 ± 2 % into the *2R* position, which indicated that this variant lost stereospecificity of proton donation. Both variants T196A and T196V showed only a partial loss in stereospecificity, incorporating the deuteron into the *2R* position with 91 ± 1 % and 77 ± 1 %, respectively (**Table S1**). These results suggested that Y158 contributes stronger to

the stereospecificity of proton donation than T196. We next aimed at determining the kinetic isotopic effects (KIE) of protonation in Y158F variant.

Table 1 Kinetic parameters of InhA WT and variants.

Enzyme	Substrate	$k_{cat,app}$ [s^{-1}]	$K_{m,app}$ [mM]
InhA WT	octenoyl-CoA	4.9 ± 0.2	0.8 ± 0.1
	NADH	3.5 ± 0.2	0.09 ± 0.01
	C2-ene adduct	3.0 ± 0.2	0.011 ± 0.002
InhA Y158F	octenoyl-CoA	0.088 ± 0.005	2.0 ± 0.3
	NADH	0.079 ± 0.001	0.01 ± 0.001
InhA Y158S	octenoyl-CoA	0.0055 ± 0.0002	0.9 ± 0.1
	NADH	0.0090 ± 0.0004	0.056 ± 0.009
InhA T196A	octenoyl-CoA	0.050 ± 0.003	7.6 ± 0.7
	NADH	0.012 ± 0.0007	0.059 ± 0.011
	C2-ene adduct	0.077 ± 0.008	0.01 ± 0.003
InhA T196V	octenoyl-CoA	0.0073 ± 0.0008	4.3 ± 1.0
	NADH	0.0051 ± 0.0003	0.11 ± 0.02
	C2-ene adduct	0.056 ± 0.003	0.022 ± 0.003

For the assays of the kinetics with octenoyl-CoA, NADH was kept constant at 400 µM, for the ones with NADH octenoyl-CoA was kept at 4 mM.

4.3.3. KIE measurements confirm that Tyrosine 158 is directly involved in protonation

Parikh *et al.* had reported almost identical KIEs on protonation for InhA WT and Y158F (WT: ^{D2O}V= 1.51; YF: ^{D2O}V = 1.39) [7]. These KIEs were measured indirectly by following the consumption of the co-substrate NADH in H_2O or D_2O, assuming that reduction and protonation are coupled. NADH consumption, however, does not necessarily remain coupled to protonation in the different enzyme variants, so that the intrinsic KIE on protonation might become masked when indirect measurements are used [10]. Therefore, we decided to apply an alternative method that would allow us to directly measure the intrinsic KIE on protonation ($^{D}k_{obs}$) [11-13]. Here, the intrinsic $^{D}k_{obs}$ on protonation is determined by running assays in buffers with different H_2O and D_2O contents and fitting the isotopic composition of the products to equation 1:

$$\frac{product_H}{product_D} = {}^{D}k_{obs} \cdot \frac{f_{H_2O}}{f_{D_2O}} \qquad \text{equation 1}^{12}$$

Using this direct method, we measured a $^{D}k_{obs}$ of 1.74 ± 0.06 for InhA WT with hexenoyl-CoA as substrate. The $^{D}k_{obs}$ decreased with increasing chain length of the substrate to 1.14 ± 0.01 for dodecenoyl-CoA (**Table 2**) indicating an increasing commitment factor with increasing chain length. Compared to WT InhA, the Y158F variant showed a significantly increased $^{D}k_{obs}$ (Table 2). The $^{D}k_{obs}$ decreased with substrate length from 3.9 ± 0.3 with hexenoyl-CoA to 2.4 ± 0.1 with dodecenoyl-CoA. The strong difference in $^{D}k_{obs}$ between WT and Y158F clearly demonstrated that the Y158 variant is directly affected in the protonation step. At the same time, the apparent differences observed in

intrinsic KIEs between the NADH consumption and the deuterium incorporation method strongly suggested that the reduction and protonation steps become (partially) uncoupled in the Y158F variant.

Table 2. $^{D}k_{obs}$ of InhA WT and Y158F

Substrate	InhA WT	InhA Y158F
hexenoyl-CoA	1.74 ± 0.14	3.9 ± 0.3
octenoyl-CoA	1.28 ± 0.04	3.4 ± 0.1
dodecenoyl-CoA	1.14 ± 0.03	2.4 ± 0.1

The $^{D}k_{obs}$ was measured by quantifying the discrimination between H and D addition at enoyl-CoA concentrations that were 20-fold below K_m to keep the commitment factor low and a NADH concentration of 10 mM, more than 100-fold over K_m, to avoid a change in the commitment factor due to a change in NADH concentration

4.3.4. Y158F variant accumulates a covalent intermediate between crotonyl-CoA and NADH

To investigate the uncoupling of the reduction and protonation steps in the Y158F variant, we characterized the enzyme's reaction in more detail. Spectrophotometric analysis of the Y158F reaction showed a transient increase of absorbance with a maximum at 375 nm, reminiscent of the formation of transient reaction intermediates that were observed recently in the enoyl-reductase Etr1p and other enzymes from the medium chain dehydrogenase superfamily (**Figure 2A**) [14-16]. This observation strongly indicated that the reaction of the Y158F variant proceeded stepwise via accumulation of an intermediate species. We next isolated the intermediate species via preparative HPLC and characterized it further by UV/Vis spectroscopy (**Figure S2A**) and high resolution mass spectrometry.

The isolated species was labile (see below) and had a mass-to-charge ratio (*m/z*) of 779.1717, which matched the expected mass of a double charged direct adduct between NADH and octenoyl-CoA (**Figure S3**). Two dimensional (COSY) NMR confirmed CoA and NADH substructures in the intermediate and identified a covalent bond between the C2 proton of the NADH nicotinamide ring and the α-proton of the octenoyl-CoA (**Figure S4**). In summary, these experiments showed that a covalent C2-ene intermediate accumulates in solution in the Y158F variant during catalysis. However, in contrast to similar C2-ene adducts previously reported for enzymes of the medium chain dehydrogenase/reductase family [14-16], the adduct from InhA Y158F is proposed to feature the opposite stereochemistry at the α-carbon (i.e. (2*S*)-C2), due to the binding geometry of the NADH cofactor and the substrate at the active site.

The isolated (2*S*)-C2-ene adduct decayed uncatalyzed in solution into the products octanoyl-CoA and NAD^+ with a decay constant of 0.186 ± 0.001 min^{-1} (**Figure S2B**). This uncatalyzed decay into the final reaction products provides an explanation for the only partial loss of activity that is observed upon mutation of Y158F. Although the Y158F is affected in the protonation step, the enzyme can still form the (2*S*)-C2-ene adduct, which then decomposes into the final products in solution. The spontaneous decay of the intermediate in solution explains the lack of stereospecificity of deuterium incorporation into octanoyl-CoA produced by Y158F. To rule out that the Y158F mutation caused major changes to the active site we co-crystallized Y158F with NAD^+. Analysis of the crystal structure confirmed that the active site geometry was not perturbed compared to WT (**Figure S5**, PDB 6EP8, **Table S2**).

4.3.5. The C2-ene adduct is also formed in the WT enzyme

The detection of the (2*S*)-C2-ene adduct in Y158F prompted the question, whether this covalent intermediate is only formed in this variant, or whether the (2*S*)-C2-ene adduct is a true intermediate in the catalytic cycle of the WT. Using stopped-flow spectroscopy on the WT enzyme, we observed an increase of absorption at 385 nm, the wavelength at which the extinction coefficients of NADH and the (2*S*)-C2-ene adduct differ most. The increase appeared in the first turnover within the first 50 ms and corresponded to approximately 4 µM (2*S*)-C2-ene adduct, which in turn corresponded to 8% of the total active sites of enzyme in the assay (**Figure 2B**). This result suggested that the (2*S*)-C2-ene adduct is formed in the catalytic cycle of the WT enzyme. Our finding was confirmed by directly injecting an enzyme assay with WT InhA during steady state catalysis into an HPLC-ESI-MS, which demonstrated existence of the (2*S*)-C2-ene adduct also under steady-state conditions (Figure 2C). In summary, these results suggest that the (2*S*)-C2-ene adduct is formed during the catalytic cycle of InhA WT and not simply as an artifact of the Y158F mutation.

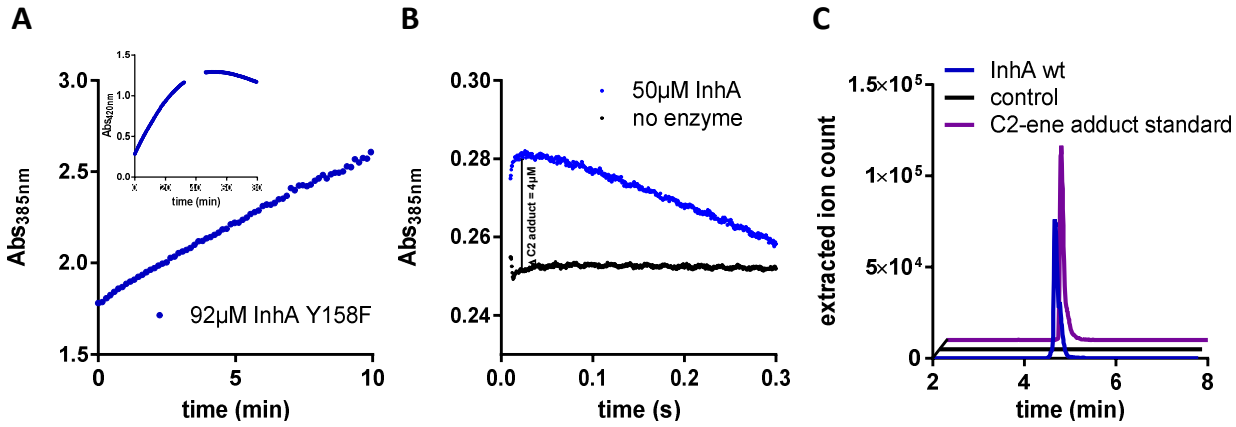


Figure 2. Detection of C2-ene adduct in InhA Y158F and WT **A)** Production of C2-ene adduct by InhA Y158F. The assay contained 15.4 mM octenoyl-CoA, 23.6 mM NADH and 92µM InhA Y158F. C2-ene adduct production was followed at 385 nm (big graph) and 420 nm (small graph). The assay was quenched when C2-ene formation plateaued after 32 min and directly injected into the HPLC for purification and further characterization **B)** Stopped flow analysis of InhA WT at 385 nm; Syringe 1 contained 100 µM InhA WT (blue line), Syringe 2 contained 4 mM octenoyl-CoA and 1 mM NADH all in 30 mM PIPES pH 6.8, 150 mM NaCl buffer. In the control Syringe 1 contained only buffer without enzyme (black line). Data shown is the average of a triplicate for each condition. **B)** LCMS analysis of InhA. The assay was directly injected after 60s incubation at room temperature during steady-state catalysis and contained 50µM InhA WT, 250mM NADH and 50mM octenoyl-CoA. In a control experiment containing 250mM NAD^+ instead of NADH, no C2-ene adduct was detected.

4.3.6. (2S)-C2-ene adduct as a "molecular probe" confirms role of Tyrosine 158 in protonation

The (2*S*)-C2-ene adduct represents an intermediary state of catalysis: In the covalent adduct the reduction has already taken place, while the protonation has not. This specific feature allows to use the (*2S*)-C2-ene adduct as a "molecular probe" to test protonation independent of the (preceeding) reduction step [14]. Indeed, purified (*2S*)-C2-ene adduct was catalytically competent and served as *bona fide* substrate for InhA WT at a k_{cat} of 3.5 ± 0.2 s^{-1} and at a K_m of 11 ± 2 µM (Table 1).

We next used (*2S*)-C2-ene adduct to specifically probe the protonation step of InhA WT and different variants. WT converted the adduct to the products octanoyl-CoA and NAD^+ as judged from the uniform decrease of the adduct peak in the UV-Vis spectrum (**Figure S6**). In contrast, in the Y158F variant, the peak maximum shifted to 340 nm, which indicated formation of NADH, and thus partial resolution of the (2*S*)-C2-ene adduct back into the two substrates. From the absorbance increase, we estimated that most of the (2*S*)-C2-ene adduct was converted back to octenoyl-CoA and NADH. Thus, the Y158F variant catalyzed mainly the back reaction. This result is in line with the fact that the protonation step

– and thus product formation from the (2*S*)-C2-ene adduct – is impaired by the Y158F mutation (Figure S6).

Compared to the Y158F variant, the T196V and T196A variants were fully able to convert the (2*S*)-C2-ene adduct further into the products, indicating that these variants are not affected in catalyzing the protonation step (**Figure S6**). Notably, both T196 variants showed an increased k_{cat} and a decreased K_m for the (2*S*)-C2-ene adduct compared to either NADH or octenoyl-CoA. This clearly demonstrated that T196 is not directly involved in the protonation reaction, and suggested that T196 is rather involved in the initial reaction steps of the catalytic cycle leading to the (2*S*)-C2-ene adduct, probably the transfer of the hydride (**Table 1**).

4.4. Discussion

Understanding the mechanism of enzymes is a prerequisite to be able to manipulate their function [17]. Here we provide new insights into the catalytic cycle of the enoyl-CoA reductase InhA, the major drug target of *M. tuberculosis*. Our data suggest that InhA forms a covalent intermediate adduct between the nicotinamide cofactor and the enoyl-CoA substrate during catalysis. Pre-steady-state as well as steady-state measurements show that a significant amount of the active sites of InhA WT are occupied by the (2*S*)-C2-ene adduct. Intriguingly, similar dihydropyridine-substrate adducts were observed recently in chemical model reactions [18], the biosynthesis of the natural product sanguinarine [19], and the catalytic cycle of enoyl-CoA reductases of the medium chain dehydrogenases (MDR) superfamily [15-16]. The detection of the (2*S*)-C2-ene adduct in InhA, an enzyme from the short chain dehydrogenases (SDR) superfamily, adds to these recent observations, and suggests that the formation of such intermediates is not a singularity of one enzyme or one enzyme family, but eventually a more general catalytic principle at the active site of oxidoreductases.

How is the covalent adduct formed and why?. Based on their studies with model compounds, Libby and Mehl suggested the alignment of the reactants in an ene-like transition state (**Figure 3**) [18]. The character and position of the reactants, as well as the electronic environment (*e.g.*, the active site of the catalysts) will strongly influence the way the hydride is transferred. One (extreme) possibility is that hydride transfer proceeds in a pericyclic fashion resulting in the direct formation of the (2*S*)-C2-ene adduct. The (2*S*)-C2-ene adduct would then be resolved via the enolate to the final products in subsequent steps. Alternatively, the hydride is transferred to the Cβ of the enoyl-CoA ester according to the more canonical mechanism directly forming the enolate. The enolate can either react further to the reaction product (*e.g.*, by protonation) or re-attack the nicotinamide cofactor in a Michael type addition to form the (2*S*)-C2-ene adduct (**Figure 3**). We note that the lack of good electron acceptors in proximity to the Cα of the enolate would favor such a Michael addition, as for instance is the case in enoyl-CoA carboxylase/reductases when CO_2 is omitted [16] or Etr1p when the proton donor is mutated [14].

In available crystal structures of InhA with cofactor and product, no obvious proton donor is positioned close to the Cα [3], indicating that the formation and accumulation of the (2*S*)-C2-ene adduct might be a logical consequence of the active site geometry of InhA. Formation of the (2*S*)-C2-ene adduct could also provide an explanation for the puzzling stereochemistry of protonation, which proceeds from the same side as the hydride is transferred in InhA, although an obvious proton donor in is absent from this side [8]. Note that bond formation between the enoyl-CoA and NADH would induce a change in hybridization of the Cα of enoyl-CoA from sp^2 to sp^3, which would bring the Cα closer to the phenolic hydroxy group of Y158. This re-positioning could put Y158 into the position for a proton transfer after decomposition of the (2S)-C2-ene adduct back into the enolate and NAD^+ (**Figure S7**).

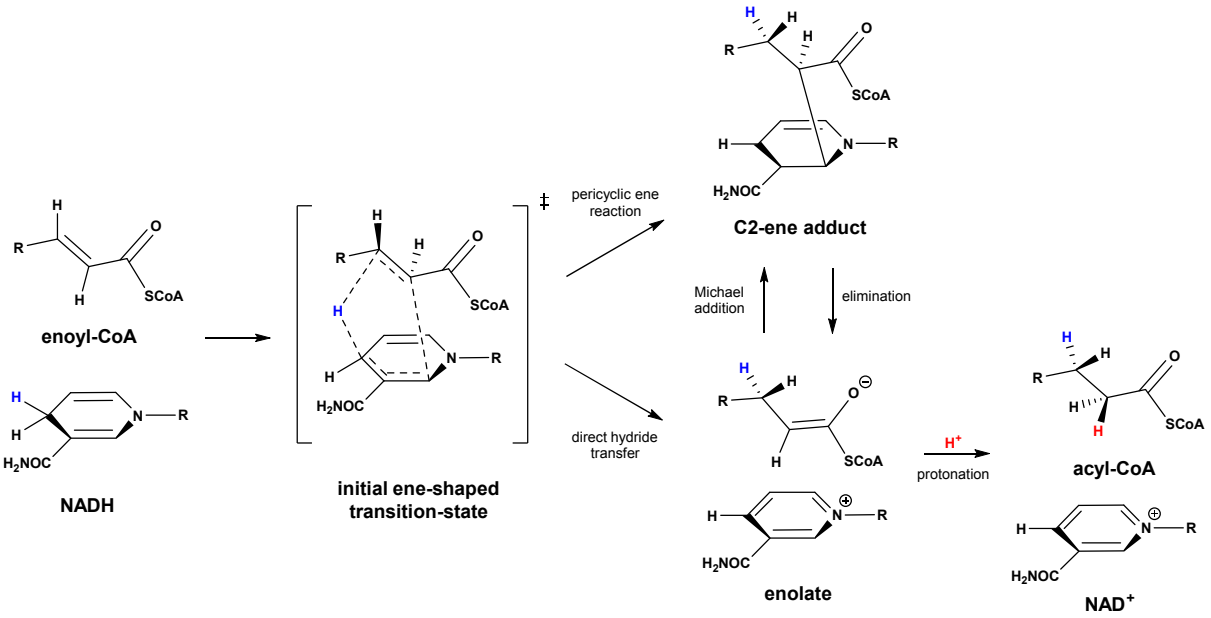


Figure 3. Proposed reaction pathways for InhA and related oxidoreductases. In a first step the substrates enoyl-CoA and NADH form an ene-shaped transition state. The observed C2-ene adduct can then be formed either directly through a pericyclic ene reaction or via a enolate formed from direct hydride transfer through a Michael-addition reaction. The C2-ene adduct can then be resolved through an elimination reaction and the enolate abstracts a proton to form the reduced acyl-CoA product.

Besides mechanistic implications, the discovery of a covalent adduct in the catalytic cycle of InhA provides the basis to prepare new tools to interrogate and eventually interfere with the mechanism of InhA. Using the (2*S*)-C2-ene intermediate as molecular probe allowed us to assess the function of individual amino acid residues at the active site of InhA. At the same time, the (2S)-C2-ene adduct could inspire the design of a new class of mechanistic inhibitors of InhA and related NAD(P)H dependent oxidoreductases. Note that compounds that are able to trap substrates at the covalent intermediate stage are expected to be high affinity binders of these enzymes. Such compounds could act in analogy to finasteride, a mechanism based inhibitor of 5α-reductase that forms a similar covalent adduct between the C4 of NADH and the prodrug within the active site of the enzyme [20]. Thus, exploiting the natural tendency of InhA to form covalent adducts with NADH might prove to be a convenient way to inhibit this key target of *M. tuberculosis*

4.5. Materials and Methods

4.5.1. Chemicals

Hexenoic- and octenoic-acid (synthesis grade) were purchased from Sigma Aldrich AG, coenzyme A and DNaseI from Roche Diagnostics, Na_2NADH (98%) was purchased from Carl Roth GmbH. All salts and solvents were of analytical grade or better. Dodecenoic acid was synthesized from decanal via Knoevenagel condensation with malonic acid according to a previously described protocol [21]. Enoyl-CoAs where synthesized using the mixed anhydride method adapted from a previously described protocol [21-22]. The unsaturated acid (156 µmol) and triethylamine (156 µmol) were dissolved in CH_2Cl_2 (3 ml) and stirred at 23°C for 30 min. The reaction was cooled to 4 °C and ethylchloroformate (156 µmol) was added. After 2 h the solvent was evaporated, the crude product dissolved in DMF (4.5 ml) and added to a stirring solution of CoA-trilithium salt (76 µmol) in 0.4 M $KHCO_3$ (4.5 ml). The reaction procedure was monitored by mixing 5 µl of reaction mixture with 35 µl of an aqueous 5,5'-dithiobis-2-nitrobenzoic acid (DTNB, Ellman's reagent) solution. Upon completion the reaction was acidified to pH = 3-4 with formic acid, diluted to 50 ml with H_2O and lyophilized. The product was resuspended in H_2O and purified by reverse phase HPLC over a Gemini 10 µm NX-C18 110 Å, 100 x 21.2 mm, AXIA packed column (Phenomenex) using a gradient from 5% to 65% (hexenoyl-CoA) and to 95% (octenoyl-, dodecenoyl-CoA) over 15 min with 25 mM ammoniumformate pH=8.1 as the aqueous phase. Fractions containing the product were pooled, lyophilized and stored at -20 °C.

4.5.2. Cloning and mutagenesis

The InhA WT gene from *Mycobacterium tuberculosis* present in plasmid pET15b (gift from Prof. John Blanchard, Albert Einstein College of Medicine, NY, USA) was used as received to express InhA WT. InhA variants were generated with the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA) using 60 ng of template plasmid and the following primer pairs:

Y158F: CGATGCCGGCCTTCAACTGGATGAC and GTCATCCAGTTGAAGGCCGGCATCG;

Y158S: CGATGCCGGCCTCCAACTGGATGAC and GTCATCCAGTTGGAGGCCGGCATCG;

T196A: CAGGCCCTATCCGGGCGCTGGCGATGAGTG and CACTCATCGCCAGCGCCCGGATAGGGCCTG;

T196V: GGCCCTATCCGGGTGCTGGCGATGAG and CTCATCGCCAGCACCCGGATAGGGCC

4.5.3. Protein expression and purification

The his-tagged proteins were expressed in *E. coli* BL21 (DE3) or *E. coli* BL21-AI™ (DE3) in terrific broth by inducing at OD_{600} = 1.5 with 0.5 mM Isopropyl-beta-D-thiogalactoside and 2.5 mM arabinose when using E. coli BL21-AI™ (DE3) at 25 °C for 12-16 hours. Harvested cells were lysed in 2.5 ml lysis buffer per gram cells: lysis buffer 500mM NaCl, 20 mM Tris HCl, 10% (v/v) glycerol pH 7.9 with 1 mM MgCl2 and DNaseI at 1 µg/ml. Cells were lysed by sonication and subsequently centrifuged at 50'000 x g for at least 30 minutes at 4°C. The clear supernatant was applied to a 1ml HisTrap FF column and washed with 24% buffer B for 30-40 column volumes, until the UV 280 nm absorbance did not decrease anymore. The protein was eluted with 100 % buffer B. Buffer A: 500mM NaCl, 20 mM Tris HCl pH 7.9, Buffer B was identical to buffer A with the addition of 250 mM imidazole. The eluted protein was desalted into 30 mM piperazine-N,N'-bis(2-ethanesulfonic acid), 150 mM NaCl, pH 6.8 within 30 minutes from eluting, as previous reports mention that the protein can precipitate in the elution buffer [7]. The enzyme was kept in this buffer at 4 °C until it was used.

4.5.4. Determining the stereochemistry of protonation

Isotopic label incorporation experiments were done analogous to a previously described method (see reaction scheme in **Figure S1**) [14]. The protein storage buffer of InhA WT and variants was exchanged using three subsequent ultracentrifugation steps (Amicon Ultra- 0.5mL Centrifugal Filters, Merck Milipore) diluting the original buffer over 10,000 x with deuterated 30 mM PIPES, 150 mM NaCl buffer pD 6.8. A 200 µL assay contained 400 µM NADH and 300 µM octenoyl-CoA in deuterated 30 mM PIPES, 150 mM NaCl buffer pD 6.8 and were started with 12.5 µM InhA WT, 22.6 µM InhA Y158F, 70.3 µM Y158S, 23.9 µM InhA T196V, 40 µM InhA T196A. The reactions were followed spectrophotometrically at 360 nm and run at 30 °C until complete consumption of NADH (ca. 1 min for WT, 3h for Y158F, Y158S and T196A) except for the assay containing T196V, which was stopped after 7h after approximately 50% of NADH was consumed. 20 µL of 50% formic acid was added to quench the reaction. The octanoyl-CoA was purified by HPLC, lyophilized and resuspended in 100 mM Tris HCl pH 8.0. Label incorporation was checked by HPLC-ESI-MS. 7.9 µM Acx4 was added to the samples and the reaction mixtures were incubated for 10 min at 30°C, quenched by adding 5% formic acid and analyzed by HPLC-ESI-MS (**Table S1**). To make sure that the oxidase Acx4 does not unspecifically scramble the signal in the α position an assay containing 300 µM octanoyl-CoA, 0.13 µM Acx4 was run in deuterated 100 mM TrisDCl pD 8.0 quenching the assay at different time points by adding 5% formic acid. Over the time course of 90 min no incorporation of a deuterium signal in either octanoyl-CoA or octenoyl-CoA could be detected.

4.5.5. KIE measurements

Reactions for the measurement of the KIEs in InhA WT and Y158F contained 30 mM PIPES buffer pH 6.8, 150 mM NaCl, 10 mM NADH, 260 µM hexenoyl-CoA, 26 µM octenoyl-CoA or 4.6 µM dodecenoyl-CoA. H_2O and D2O were added according to the desired percentage of D_2O; between 10 and 70 % (v/v). The enzymes were equilibrated with the partially deuterated buffer for at least 2 hours before the assay was started with the addition of the enoyl-CoA substrate (Enoyl-CoA substrates were dissolved as 25-fold stock solutions in 50% DMSO). Hexenoyl-CoA and octenoyl-CoA assays contained 147 nM of InhA WT or InhA Y158F, the dodecanoyl-CoA assays contained 29 nM of either enzyme. Assays with InhA WT were quenched by addition of 5% formic acid after 4 min for octenoyl-CoA and dodecanoyl-CoA and after 20 min for hexanoyl-CoA. Assays with InhA Y158F were quenched after 150 min for all three substrates.

4.5.6. C2-ene adduct purification and characterization

A mixture containing 15.5 mM octenoyl-CoA, 23.7 mM NADH and 92.5 µM InhA Y158F in 200 mM Tris HCl (pH=7, 4 ml) was reacted at 4°C until the absorption at 400 nm reached a maximum. The reaction was purified on an Agilent HPLC system using a flow rate of 25 ml min^{-1} and using a Gemini 10 µm NX-C18 110 Å, 100 × 21.2 mm, AXIA packed column (Phenomenex). The elution of the intermediate was followed at 370 nm and collected directly into liquid nitrogen. Upon lyophilization (0.01 mbar, -55°C) the compound appeared as an intense yellow powder that was stored at -80°C. For spectrophotometric analysis C2-ene adduct was dissolved in 30 mM PIPES buffer pH 6.8, 150 mM NaCl. The absorption spectrum of the C2-ene adduct (Figure S2A) and the uncatalyzed decay of C2-ene adduct at 30°C (**Figure S2B**) was determined on a Cary-60 UV/Vis spectrometer (Agilent) using quartz cuvettes (1-, 3- or 10-mm diameter; Hellma). For NMR spectroscopy the C2-ene adduct was dissolved in D_2O 25 mM Na_2DPO_4 pH=7.9 and measured at 4°C at 600 MHz (Figure S3).

4.5.7. Stopped-flow spectroscopy

Measurements were recorded on a thermo-stated stopped flow unit (SFM-20 connected to a MOS-200, equipped with a Xe(Hg)-lamp and a TC-100/10 cuvette, Bio-Logic Science Instruments SAS, Claix, France) set to 30 °C. Syringe 1 contained 1 mM NADH and 4 mM octenoyl-CoA in 30 mM PIPES, 150 mM NaCl buffer pH 6.8 and syringe 2 contained 100 µM of InhA WT in the same buffer. For the controls either the syringe 1 or syringe 2 contained only buffer. Data was collected at 385 nm every 0.5 ms for the first 4 s and then every 0.5 s up to a total of 90 s. Each assay and control was repeated at least 4 times and the resulting traces were averaged to reduce noise.

4.5.8. Protein crystallization

For protein crystallization the HisTrap purified InhA was incubated with thrombin to cleave off the His-tag (120 U thrombin/8 mg InhA) for 2h at room temperature. The solution was then passed through a HisTrap column equilibrated in 30 mM PIPES, 150 mM NaCl pH 6.8. The flow through was injected onto a Ge HiLoad 16/600 Superdex 200pg column running on 30 mM PIPES, 150 mM NaCl pH 6.8 buffer. Crystallization of poly-histidine tag cleaved InhA from *M. tuberculosis* was performed under air at 14-20°C with a protein concentration of 10 mg/ml. For all crystallization experiment, 3 to 10 mM of NADH was added to the protein solution and the protein sample was filtered through 0.2 µm filter. The 24-well crystal plate, Combiclover Junior (Jena Bioscience), was used for sitting drop and the crystallization conditions from Chollet A. *et al* was reproduced.[23] The best crystals appeared in drops of 2 µl of enzyme solution, 1 µl of the reservoir solution and 1 µl of distilled water. The reservoir solution contained 100 mM HEPES/NaOH pH 7.5, 50 mM sodium citrate pH 6.5, 7-9 % 2-methyl-2,4-pentanediol. Bi-pyramidal shaped crystals appeared typically within two to three weeks in this condition.

4.5.9. Structural analysis

The crystals were cryo-protected by soaking with 25 % glycerol (v/v) in the crystallization solution for 3-5 seconds prior freezing in liquid nitrogen. The diffraction experiments were performed at 100 K on beamline BM30A (FIP) at the ESRF (Grenoble) equipped with a CCD detector (ADSC Q315r). The data were processed with XDS [24] and scaled with SCALA from the ccp4 suite [25]. The structure was determined using the InhA wild-type from *M. tuberculosis* inhibited with the active metabolite of isoniazid (PDB code 4TRO) as a template for MOLREP. The model was manually constructed with COOT [26] and refined by PHENIX [27] including hydrogens. The final model was validated by using the MolProbity server (http://molprobity.biochem.duke.edu) [28] and deposited without hydrogens. Data collection and refinement statistics of the model are listed in Table S2. The figures were generated and rendered with PyMOL (Version 1.5, Schrödinger, LLC). For a comparison of the solved structure of InhA Y158F (PDB 6EP8) with the previously solved WT structure (PDB 1BVR [3]) see **Figure S5**.

4.5.10. Spectrophotometric enzyme assays

Assays were carried out on a Carry-4000 UV/Vis spectrometer (Agilent) at 30 °C using quartz cuvettes (1-, 3-, or 10-mm path-length; Hellma). All assays were carried out in 30 mM PIPES, 150 mM NaCl, pH 6.8 if not noted otherwise. For the determination of the kinetic parameters for NADH, octenoyl-CoA was kept constant at 4 mM; for octenoyl-CoA kinetics NADH was kept at 400 µM NADH and were measured at 340 nm. C2-ene adduct was measured at 385 nm and the concentration of C2-ene adduct was determined directly within each assay using the extinction coefficient $_{C2,385}$=7.2 $cm^{-1}mM^{-1}$ (**Figure S2A**). All Michaelis-Menten curves were determined by curve fitting using GraphPad Prism 7.02 with a minimum of 12 measuring points. Scanning traces of InhA WT and variants using C2-ene adduct as a substrate were monitored between 300 and 500 nm using 2 µM InhA WT, 1.2 µM InhA Y158F, 2 µM InhA T196V and 2 µM InhA T196A (**Figure S6**).

4.6. References

1. Zumla, A.; George, A.; Sharma, V.; Herbert, R. H.; Baroness Masham of, I.; Oxley, A.; Oliver, M., *The Lancet. Global health* **2015,** *3* (1), e10-2.
2. Hartkoorn, R. C.; Pojer, F.; Read, J. A.; Gingell, H.; Neres, J.; Horlacher, O. P.; Altmann, K. H.; Cole, S. T., *Nat Chem Biol* **2014,** *10* (2), 96-8.
3. Rozwarski, D. A.; Vilcheze, C.; Sugantino, M.; Bittman, R.; Sacchettini, J. C., *The Journal of biological chemistry* **1999,** *274* (22), 15582-9.
4. Shirude, P. S.; Ramachandran, S.; Hosagrahara, V., *Future medicinal chemistry* **2013,** *5* (5), 499-501.
5. Quemard, A.; Mazeres, S.; Sut, A.; Laneelle, G.; Lacave, C., *Biochim Biophys Acta* **1995,** *1254* (1), 98-104.
6. Quemard, A.; Sacchettini, J. C.; Dessen, A.; Vilcheze, C.; Bittman, R.; Jacobs, W. R., Jr.; Blanchard, J. S., *Biochemistry* **1995,** *34* (26), 8235-41.
7. Parikh, S.; Moynihan, D. P.; Xiao, G.; Tonge, P. J., *Biochemistry* **1999,** *38* (41), 13623-34.
8. Bell, A. F.; Stratton, C. F.; Zhang, X.; Novichenok, P.; Jaye, A. A.; Nair, P. A.; Parikh, S.; Rawat, R.; Tonge, P. J., *Journal of the American Chemical Society* **2007,** *129* (20), 6425-31.
9. Fillgrove, K. L.; Anderson, V. E., *Biochemistry* **2000,** *39* (23), 7001-11.
10. Northrop, D. B., *Annu Rev Biochem* **1981,** *50*, 103-31.
11. Rosenthal, R. G.; Vogeli, B.; Wagner, T.; Shima, S.; Erb, T. J., *Nat Chem Biol* **2017,** *13* (7), 745-749.
12. Northrop, D. B., *Biochemistry* **1975,** *14* (12), 2644-51.
13. Miwa, G. T.; Harada, N.; Lu, A. Y., *Arch Biochem Biophys* **1985,** *239* (1), 155-62.
14. Rosenthal, R. G.; Vogeli, B.; Quade, N.; Capitani, G.; Kiefer, P.; Vorholt, J. A.; Ebert, M. O.; Erb, T. J., *Nat Chem Biol* **2015,** *11* (6), 398-+.
15. Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L. C.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L., *Structure* **2015,** *23* (12), 2213-2223.
16. Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat Chem Biol* **2014,** *10* (1), 50-U85.
17. Lillelund, V. H.; Jensen, H. H.; Liang, X.; Bols, M., *Chemical reviews* **2002,** *102* (2), 515-53.
18. Libby, R. D.; Mehl, R. A., *Bioorg Chem* **2012,** *40*, 57-66.
19. Sandor, R.; Slanina, J.; Midlik, A.; Sebrlova, K.; Novotna, L.; Carnecka, M.; Slaninova, I.; Taborsky, P.; Taborska, E.; Pes, O., *Phytochemistry* **2018,** *145*, 77-84.
20. Bull, H. G.; GarciaCalvo, M.; Andersson, S.; Baginsky, W. F.; Chan, H. K.; Ellsworth, D. E.; Miller, R. R.; Stearns, R. A.; Bakshi, R. K.; Rasmusson, G. H.; Tolman, R. L.; Myers, R. W.; Kozarich, J. W.; Harris, G. S., *Journal of the American Chemical Society* **1996,** *118* (10), 2359-2365.
21. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., *Angewandte Chemie* **2015,** *54* (45), 13457-61.
22. Peter, D. M.; Vogeli, B.; Cortina, N. S.; Erb, T. J., *Molecules* **2016,** *21* (4).
23. Chollet, A.; Mourey, L.; Lherbet, C.; Delbot, A.; Julien, S.; Baltas, M.; Bernadou, J.; Pratviel, G.; Maveyraud, L.; Bernardes-Genisson, V., *Journal of structural biology* **2015,** *190* (3), 328-37.
24. Kabsch, W., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 2), 125-32.
25. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., *Acta crystallographica. Section D, Biological crystallography* **2011,** *67* (Pt 4), 235-42.
26. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 4), 486-501.
27. Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Headd, J. J.; Moriarty, N. W.; Richardson, J. S.; Richardson, D. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., *Journal of applied crystallography* **2010,** *43* (Pt 4), 669-676.

28. Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 1), 12-21.
29. Dawson, R. M. C., *Data for Biochemical Research*. Clarendon Press: Oxford, 1986.

4.7. Supplementary Information

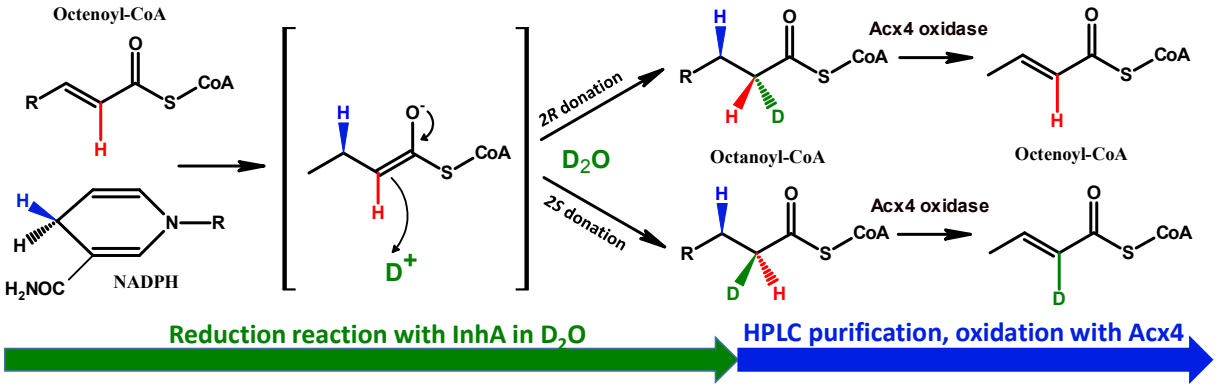


Figure S1 Scheme showing the setup for the isotopic label incorporation experiment. Octenoyl-CoA is reduced by InhA WT and variants in D_2O and a deuterium is incorporated into octanoyl-CoA in either 2R or 2S position depending on the stereospecificity of the proton doror. Octanoyl-CoA is purified via HPLC and then oxidized back to octenoyl-CoA with the stereospecific oxidase Acx4. If the proton was donated into 2R position, the resulting octenoyl-CoA will be unlabeled. In case of 2S donation the resulting octenoyl-CoA will be once deuterated.

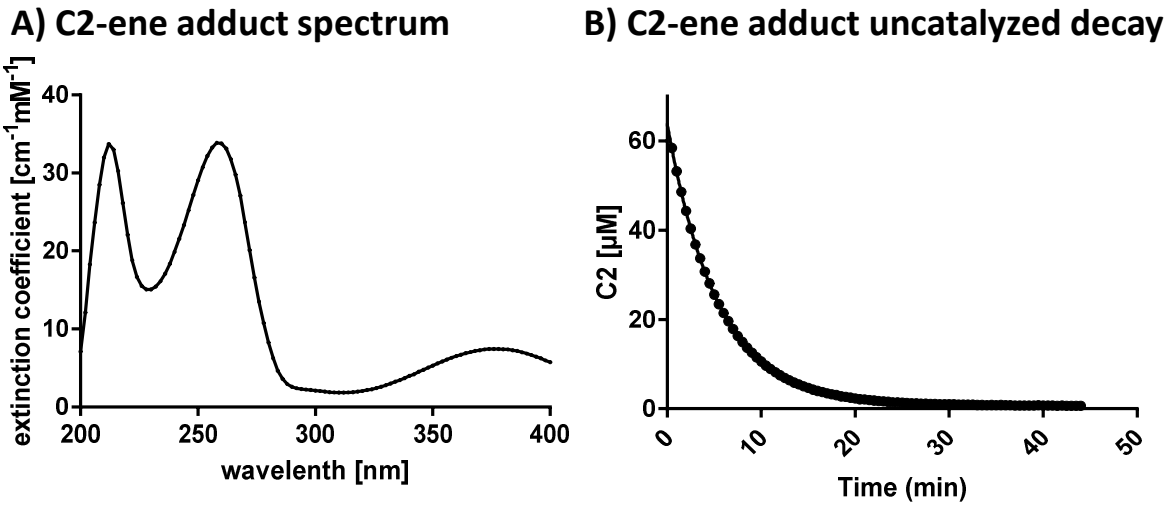


Figure S2 Spectrophotometric characterization of C2-ene adduct **A)** Absorption spectrum of C2-ene adduct. Extinction coefficients were calculated assuming the total absorbance of 33.8 cm^{-1}mM-1 at 260 nm ([ε_{260nm} NADPH + ε_{260nm} CoA] = [16.9 cm-1mM-1+ 16.9 cm-1 mM-1])[29] **B)** Uncatalyzed decay of C2-ene adduct in 30 mM PIPES buffer pH 6.8, 150 mM NaCl at 30°C was measured at 375nm, concentration was calculated using the extinction coefficient ε_{375nm}= 7.46 $cm^{-1}mM^{-1}$. Data was fitted with a first order decay, the decay rate was determined to be 0.186 ± 0.001 min^{-1} with a $t_{1/2}$ of 3.73 ± 0.01 min using three independent decay curves.

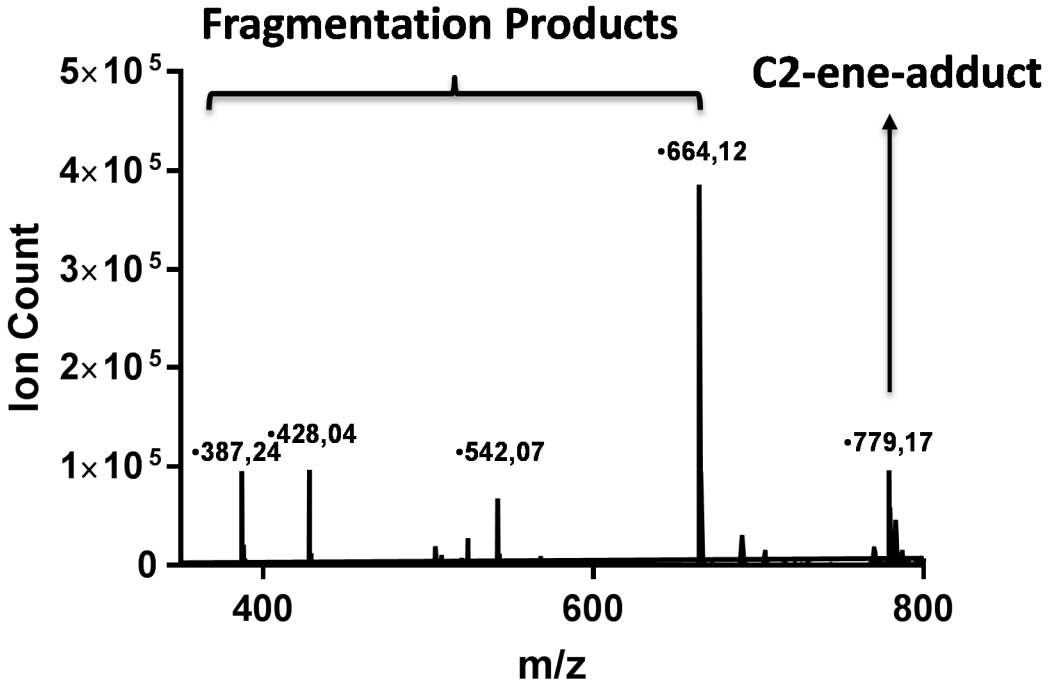
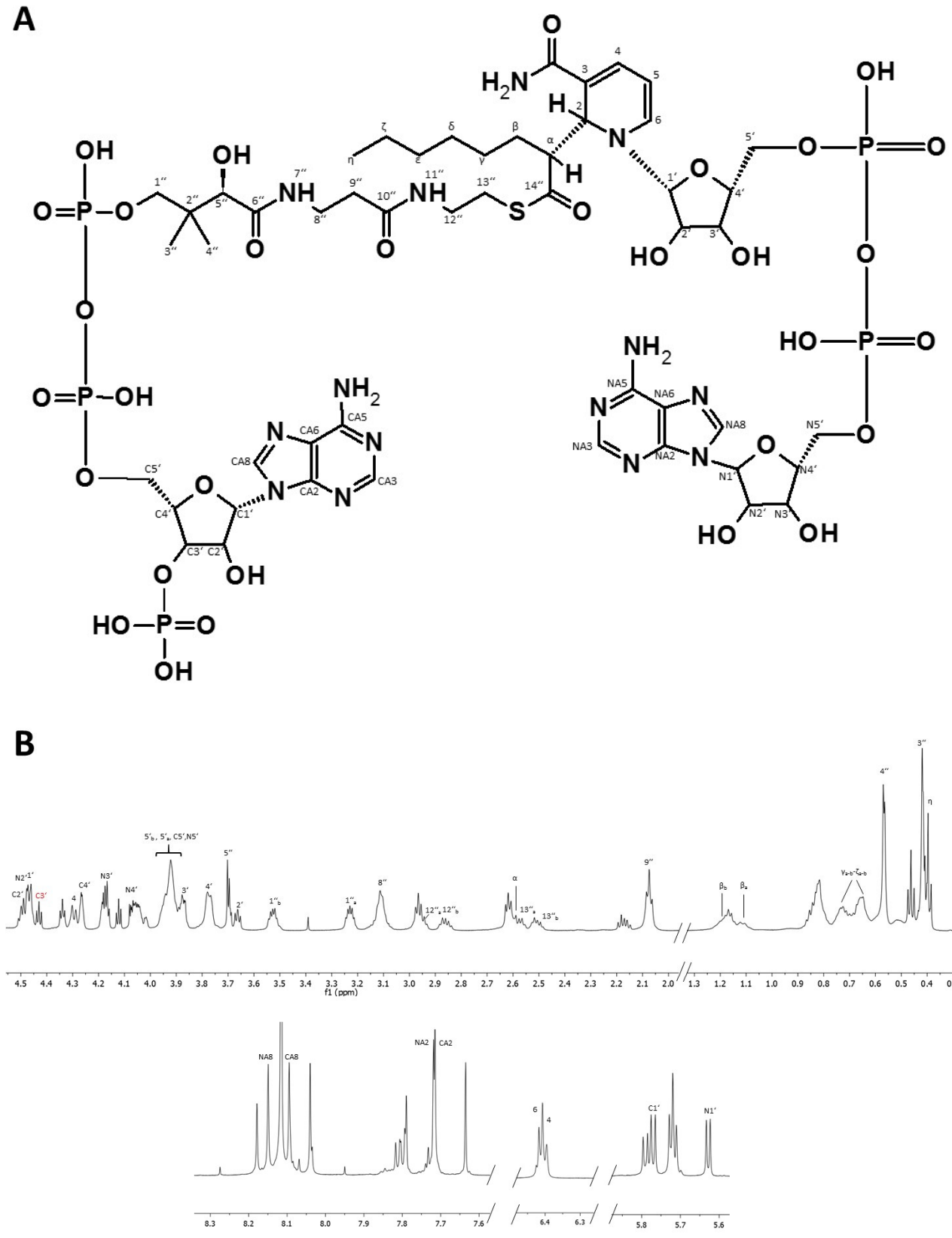
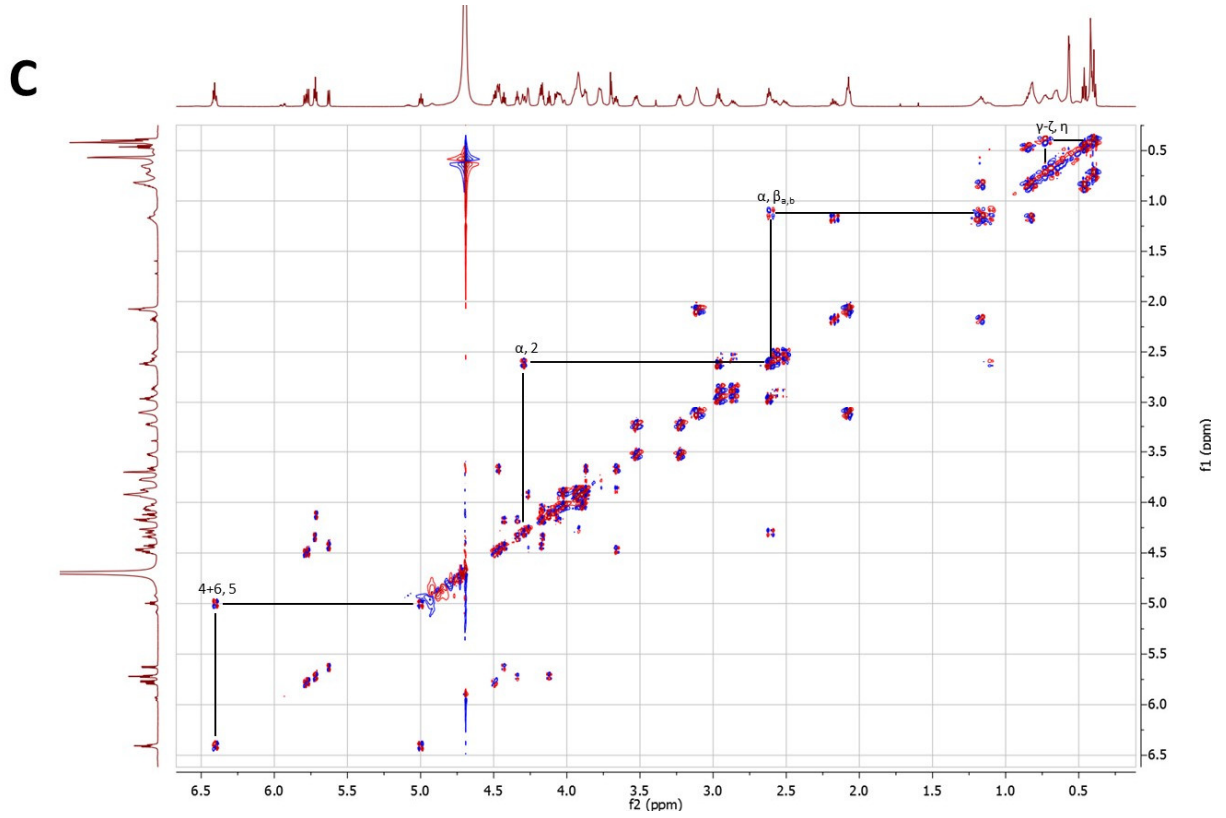
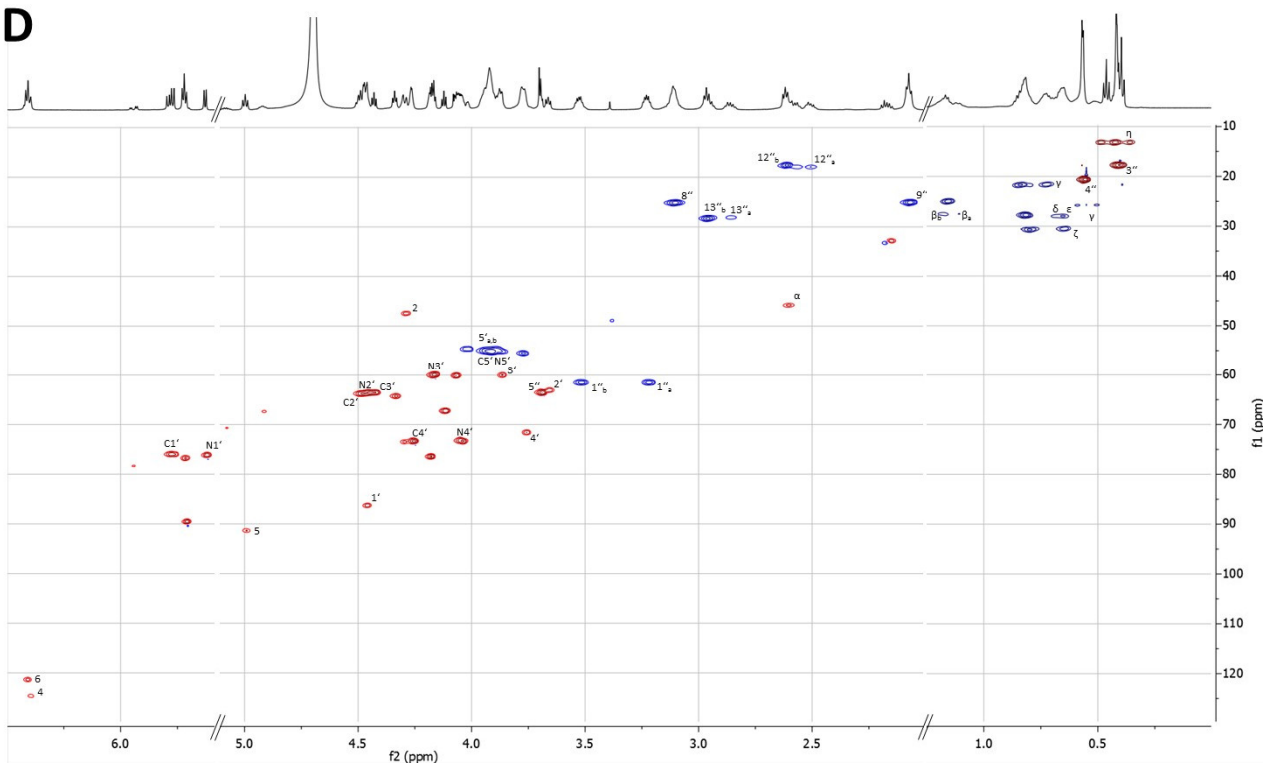


Figure S3 LC-MS analysis of C2-ene adduct. The doubly charged ion with a mass-to-charge ratio of 779.17 (m/z) corresponds to a covalent adduct of NADH and octenoyl-CoA. The ion with an (m/z) at 664 corresponds to NAD+, the one at 387 corresponds to the octenoyl-pantetheine, 428 corresponds to 3´-phospho-AMP and the 542 ion corresponds to a loss of nicotinamide of the NAD+ ion.

Figure S4 Two-dimensional NMR analysis of the C2-ene adduct recorded at 600 MHz in D_2O 25 mM Na_2DPO_4 pH=7.9 at 4.0 °C. **A)** C2-ene adduct structure and numbering **B)** ^{1}H-NMR- **C)** DQF-COSY- **D)** HSQC- spectra of the C2-ene adduct**. E)** Assignment table for the C2 ene-adduct.




C
γ-ζ, η
α, β$_{a,b}$
α, 2
4+6, 5
f2 (ppm)
f1 (ppm)


D
f2 (ppm)
f1 (ppm)

E

Position	^{1}H-shift (ppm)	^{13}C-shift (ppm)	HMBC	DQF-COSY
α	2.62, 2.16[1]	56, 43[1]	-	C_2, β_a, β_b
β_a	1.13, 1.16[1]	28, 25[1]	-	C_α
β_b	1.16, 1.16[1]	28, 25[1]	-	C_α
γ-ζ	0.65-0.75, 0.80[1]	22,28,28,31, 22,28,28,31[1]	-	C_η
η	0.40, 0.46[1]	13, 13[1]	-	$C_{\gamma\text{-}\zeta}$
2	4.3, 9.0[1]	58, 140[1]	C_α,C_4,C_5,C_6	C_α
3	-	113[2]	C_2,C_5	-
4	6.4 8.8[1]	135, 142[1]	C_2,C_5,C_6	C_5
5	5.0, 8.1[1]	101, 128[1]	C_4	C_4, C_6
6	6.4, 8.4[1]	131, 146[1]	C_2,C_4,C_5	C_5
$1''_a$	3.21	72	$C_{2''}$,$C_{3''}$,$C_{4''}$,$C_{5''}$	$C_{1''b}$
$1''_b$	3.51	72	$C_{2''}$,$C_{3''}$,$C_{4''}$,$C_{5''}$	$C_{1''a}$
2''	-	38	$C_{1''}$,$C_{3''}$,$C_{4''}$,$C_{5''}$	-
3''	0.42	18	$C_{1''a}$,$C_{1''b}$,$C_{5''}$	-
4''	0.57	21	$C_{1''a}$,$C_{1''b}$,$C_{5''}$	-
5''	3.71	73	$C_{1''a}$,$C_{1''b}$,$C_{2''}$,$C_{3''}C_{4''}$, $C_{5''}$,$C_{6''}$	-
6''	-	175	$C_{5''}$,$C_{6''}$, $C_{8''}$	-
7''	-	-	-	-
8''	3.11	35	$C_{9''}$,$C_{10''}$,$C_{12''}$	$C_{9''}$
9''	2.07	35	$C_{8''}$,$C_{10''}$	$C_{8''}$
10''	-	174	$C_{8''}$,$C_{9''}$,$C_{12''}$	-
11''	-	-	-	-
$12''_a$	2.51	28	$C_{\alpha''}$, $C_{\beta''}$,$C_{\gamma\text{-}\zeta}$, $C_{8''}$, $C_{9''}$, $C_{10''}$	$C_{12''b}$, $C_{13''a}$, $C_{13''b}$
$12''_b$	2.58	28	$C_{\alpha''}$, $C_{\beta''}$,$C_{\gamma\text{-}\zeta}$, $C_{8''}$, $C_{9''}$, $C_{10''}$	$C_{12''a}$, $C_{13''a}$, $C_{13''b}$
$13''_a$	2.95	38	$C_{14''}$	$C_{12''a}$, $C_{12''b}$, $C_{13''b}$
$13''_b$	2.86	38	$C_{14''}$	$C_{12''a}$, $C_{12''b}$, $C_{13''a}$
14''	-	205	C_α, C_β, $C_{13''a}$,$C_{13''b}$	-
1'	4.46	96	C_2, $C_{2'}$	$C_{2'}$
2'	3.66	73	$C_{1'}$	$C_{1'}$
3'	3.87	70	-	$C_{2'}$
4'	3.77	82	-	-
$5'_a$	3.91	65	$C_{3'}$	-
$5'_b$	3.93	65	$C_{3'}$	-
C1'	5.77	86	$C_{C2'}$	$C_{C2'}$
C2'	4.48	74	$C_{C1'}$,$C_{C4'}$,$C_{C5'}$	-
C3'	4.43	73	$C_{C4'}$,$C_{C5'}$	-
C4'	4.26	83	$C_{C3'}$,$C_{C5'}$	-
C5'	3.94	65	$C_{C3'}$,$C_{C4'}$	-
N1'	5.65	86	$C_{N2'}$ $C_{N3'}$ $C_{N4'}$	$C_{N2'}$
N2'	4.43	73	$C_{N1'}$	$C_{N1'}$,$C_{N3'}$
N3'	4.17	70	$C_{N1'}$	$C_{N2'}$
N4'	4.04	83	$C_{N1'}$,$C_{N2'}$,$C_{N5'}$	-
N5'	3.93	65	$C_{N3'}$	-
CA2	7.72	152	-	-
CA8	8.09	139	-	-
NA2	7.72	152	-	-
NA8	8.15	139	-	-

1 Chemical shifts in NADP+ or octanoyl-CoA

2 Chemical shift assigned from HMBC

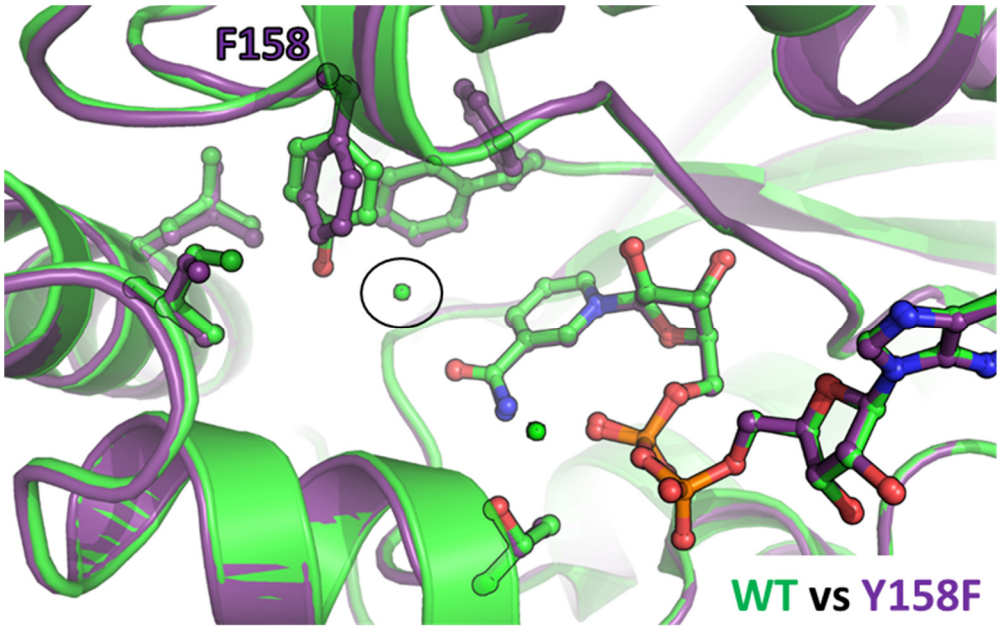


Figure S5 Crystal structure of InhA WT (green, PDB 1BVR [3]) and Y158F (purple, PDB 6EP8). The two active sites overlay very well with the phenylalanine in the Y158F variant slightly twisted compared to the WT tyrosine and the loss of a water molecule that is bound by the hydroxyl group of tyrosine in the phenylalanine variant (black circle)

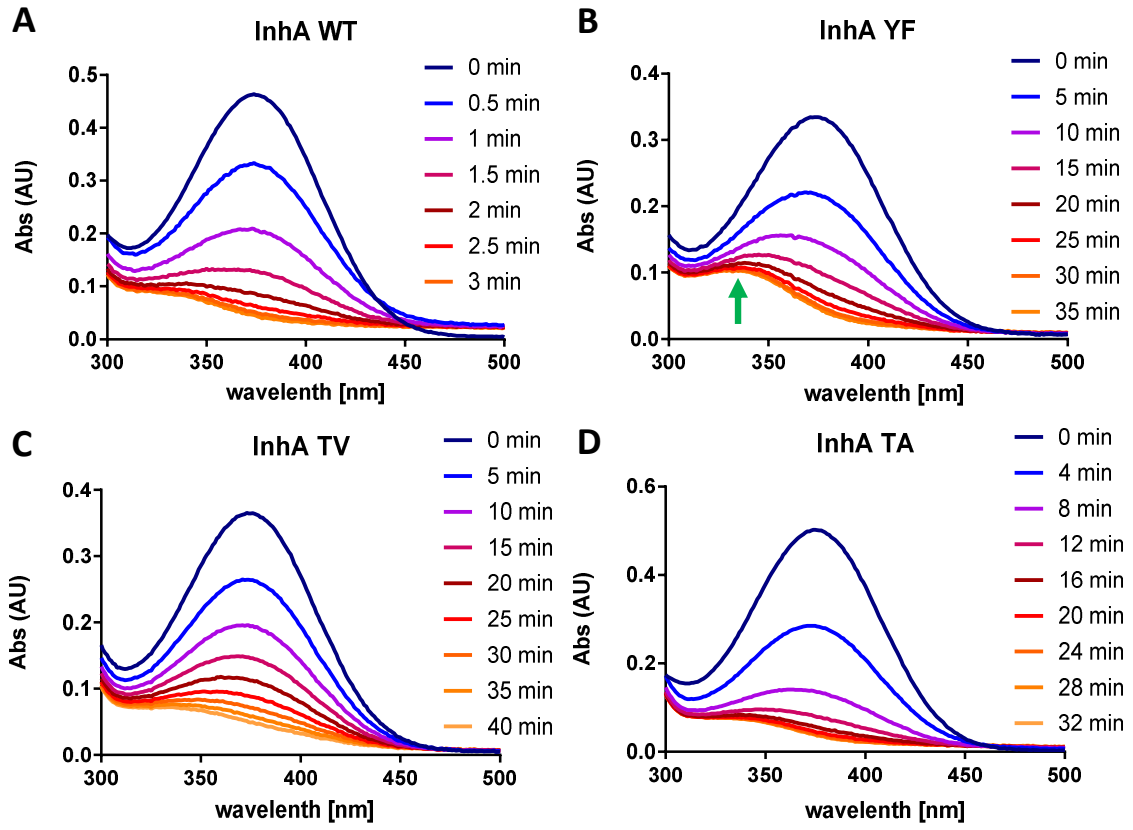


Figure S6 Scanning traces of InhA WT and variants on C2-ene adduct. Assays contained C2-ene adduct in 30mM PIPES pH 6.8, 150mM NaCl. A) 2µM InhA WT. C2-ene adduct is depleted fast into the products. B) 1.2µM InhA Y158F. C2-ene adduct competes between back reaction to NADH and octenoyl-CoA (see shift of extinction maximum towards 340nm over time, green arrow) and decay into products in solution. C) 2µM InhA T196V. C2-ene adduct is slowly consumed similar to WT, no significant amount of back reaction is observed. D) 2µM InhA T196A. Similar to T196V reaction is mainly in product direction.

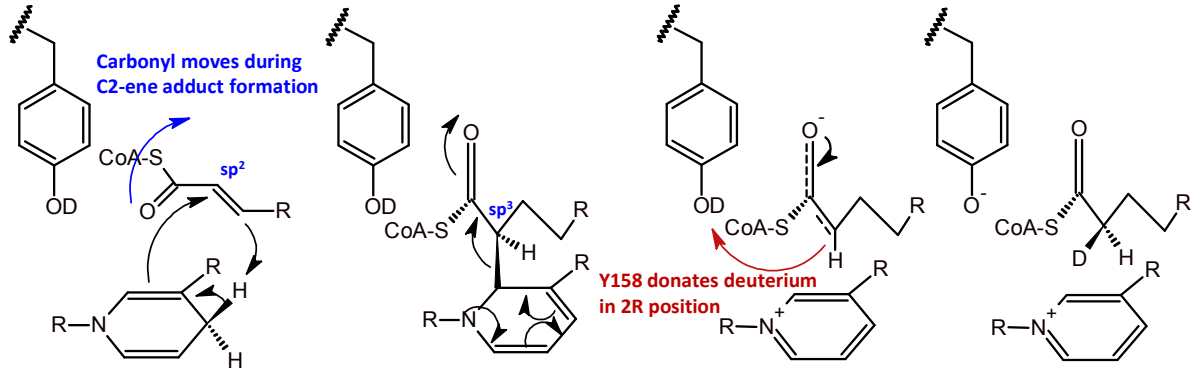


Figure S7 Possible reaction mechanism for InhA. Substrate and NADH form the C2-ene adduct via a pericyclic ene-reaction. The rehybridization (sp2 to sp3) of the Cα of the substrate causes the carbonyl to flip and expose the 2R position for protonation by tyrosine 158 upon C2-decay and enolate formation

Table S1 Label incorporation pattern for InhA WT and variants. Assay conditions are described in Supplementary chapter 4.

Enzyme variant	No label (2R)	+1 label (2S)
InhA WT	99 ± 1 %	1 ± 1 %
InhA Y158F	57 ± 2 %	43 ± 2 %
InhA Y158S	79 ± 1 %	21 ± 1 %
InhA T196A	91 ± 1 %	9 ± 1 %
InhA T196V	77 ± 1 %	23 ± 1 %

Table S2. Statistics of the crystal structure analysis.

	InhA Y158F with NADH (6EP8)
Data collection	
Wavelength (Å)	1.07244
Space group	$P6_222$
Resolution (Å)	46.61 – 1.80 (1.90 – 1.80)
Cell dimensions	
a, b, c (Å)	98.3, 98.3, 139.8
α, β, γ (°)	90.0, 90.0, 120.0
R_{merge} (%)[a]	6.5 (121.0)
R_{pim} (%)[a]	1.5 (27.1)
$CC_{1/2}$ [a]	100.0 (92.4)
I/σ_I [a]	35.6 (2.9)
Completeness (%) [a]	100.0 (100.0)
Redundancy[a]	20.5 (20.8)
Number of unique reflections [a]	37644 (5382)
Refinement	
Resolution (Å)	42.55 – 1.80
Number of reflections	37562
R_{work}/R_{free} [b] (%)	14.54 / 17.51
Number of atoms	
Protein	2086
Ligands/ions	92
Solvent	264
Mean B-value (Å^2)	37.1
Molprobity clash score, all atoms	2.06 (100th percentile)
Ramachandran plot	
Favored regions (%)	255 (95.9)
Outlier regions (%)	1 (0.38)
rmsd[c] bond lengths (Å)	0.009
rmsd[c] bond angles (°)	0.129
PDB code	6EP8

[a] Values relative to the highest resolution shell are within parentheses. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement. [c] rmsd, root mean square deviation.

CHAPTER V

Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis

Authors:
Bastian Vögeli, Patrick D. Gerlinger, Kyra Geyer, Sarah Benkstein, Niña S. Cortina, Tobias J. Erb

Published in:
Cell Chemical Biology 25, 1-7; DOI: 10.1016/j.chembiol.2018.04.009

Author contributions:
B.V. and T.J.E conceived the project. B.V., P.D.G., K.G., S.B., and T.J.E. designed and performed experiments and analyzed the data. N.S.C. performed mass spectrometry and analysis of the triketide synthesis. B.V., P.D.G., K.G. and T.J.E. wrote the manuscript with contributions from all authors.

5. Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis

5.1. Abstract

The incorporation of different extender units generates structural diversity in polyketides natural products. There is significant interest in engineering the substrate specificity of polyketide synthases (PKS) in order to change the chemical structure of polyketides. However, efforts to change extender unit selectivity in PKS are hindered by the lack of simple screening methods, as well as by the lack of easily available atypical extender units. Here, we present a chemo-biosynthetic strategy that employs biocatalytic proofreading and allows access to a large variety of extender units. In a first step saturated acids are chemically coupled to free CoA. The corresponding acyl-CoAs are further converted to their alkylmalonyl-CoAs in a 'one-pot' reaction through the combined action of an acyl-CoA oxidase and an enoyl-CoA carboxylase/reductase (ECR). We synthesized six different extender units and used them in competition *in vitro* screens to investigate active site residues conferring extender unit selectivity of PKS. Our results show the importance of an uncharacterized glutamine in extender unit selectivity and open the possibility for more comprehensive studies on extender incorporation and processing in PKS.

5.2. Introduction

Polyketides are a large class of structurally diverse natural products that exhibit a myriad of biological activities ranging from antibacterial, to antifungal, cholesterol lowering, immune system suppressing and more [1]. Although polyketides show a large structural and functional variety, they are all assembled from simple acyl- and alkylmalonyl-thioester building blocks. The building blocks are combined in successive Claisen-condensation reactions catalyzed by large multi-domain enzymes, so-called polyketide synthases (PKS). Structural diversity in the final polyketide product is achieved by a variety of different mechanisms. One important mechanism to generate structural variety in polyketides is the use of different acyl-CoA building blocks ('starter units') and their condensation with different alkymalonyl-CoA building blocks ('extender units') during biosynthesis [2]. While PKS have been recognized to accept a variety of different acyl-CoA starter units [3], the extension of the growing polyketide chain was for a long time believed to be restricted to malonyl-CoA and methylmalonyl-CoA units. However, the growing number of PKS that are able to incorporate atypical alkylmalonyl-CoA extender units, such as ethylmalonyl-CoA, benzylmalonyl-CoA or octanoylmalonyl-CoA, highlights the importance of this principle in the generation of structurally and functionally diverse polyketides [2].

Most of the atypical extender units are provided via the reductive carboxylation of unsaturated acyl-CoA precursors by enoyl-CoA carboxylase/reductases [4-6]. Genes encoding ECRs are often associated with the biosynthetic gene cluster (BGC) of the respective PKS that incorporates the atypical extender unit. Recent studies on the biochemistry of the ECRs identified active site residues that determine the substrate spectrum of these enzymes [7]. According to these studies, ECRs can be categorized into two subfamilies; the ECR-1 subfamily that possess a rather narrow substrate spectrum and the ECR-2 subfamily that display a more relaxed substrate specificity [4, 7].

The incorporation of atypical extender units through the targeted reprogramming of PKS has become an important focus in bioengineering to create polyketides that show novel properties. Most approaches in this direction focus on the manipulation of the acyl-transferase (AT) domains that are part of the PKS biosynthetic machinery. ATs are often described as the "gatekeeper" domains that decide which extender units will enter the PKS assembly line. In contrast, the downstream domains of

the PKS that perform the actual condensation and modification reactions are believed to be rather promiscuous towards unnatural α-substituents [8-10]. Several different strategies to manipulate the specificity of extender unit incorporation have been applied up to date [11-13]. Although it has become possible to incorporate atypical extender units, this was mainly achieved by lowering selectivity of ATs towards their native extender unit, which is very often accompanied by strongly reduced overall activities. An improved specificity by improving the desired selectivity in combination with reducing the native selectivity by a point mutation has to our knowledge only been achieved in one case [11]. This highlights the necessity for more comprehensive screening methods to fully understand the contributions of AT and other PKS domains towards extender unit specificity.

Another difficulty that complicates AT engineering is the limited availability of atypical extender units variants that can be used to screen for altered specificities of ATs. Commonly atypical extender units are synthesized *in situ* via enzymatic coupling of malonic acid derivatives and CoA using a promiscuous variant of the malonyl-CoA synthetase MatB from *Rhizobium trifolii* [14-15]. This synthesis route is however restricted by the limited number of commercially available malonic acid derivatives, which otherwise have to be laboriously synthesized via the acylation of Meldrum´s acid [16]. The recent description of promiscuous ECRs has opened up new routes for the chemo-enzymatic synthesis of atypical extender units at preparative scale [7]. However, promiscuous ECRs often display a significant side reactivity and simply reduce the unsaturated acyl-CoA precursors instead of carboxylation. As an example, AntE, the reductive carboxylase in the antimycin biosynthesis pathway, carboxylates only 25% of one of its natural substrates cinnamoyl-CoA, while the rest of the substrate is reduced to phenylpropionyl-CoA ('dihydrocinnamoyl-CoA')[17].

Here we increased the carboxylation yield of promiscuous ECRs by applying the principle of biocatalytic proofreading. We successfully coupled promiscuous ECRs with an acyl-CoA oxidase that is able to continuously recycle the reduction reaction side product, pushing the ECR reaction towards the carboxylation product. Based on this principle, we further developed a chemo-biosynthetic route that allows for the convenient preparation of atypical alkylmalonyl-CoA extender units from simple starting materials in 'one pot' and at high carboxylation yield. We use this route to prepare a set of six different alkylmlalonyl-CoA extender units at preparative scale and high purity and demonstrate with a recently established *in vitro* model system of the DEBS PKS [18] that these extender units can be used to systematically screen the extender unit selectivity of PKS.

5.3. Results and Discussion

5.3.1. Screening of ECRs for the production of atypical polyketide extender units from enoyl-CoAs

To first identify suitable ECRs for the preparative-scale synthesis of alkylmalonyl-CoA ester, we tested six different homologs with respect to their biosynthetic potential. Our screen included ECRs that either had been described to show relaxed substrate specificity, were part of a cluster described to produce a polyketide with atypical extender units or contained an active site motif indicating promiscuity (**Table S1**). Additionally we tested an ECR variant of *C. crescentus* ($CcrC_{PAG}$) that was recently engineered from a specific into a promiscuous enzyme through three active site mutations (C146P, I169A and F373G) [7]. The six ECRs were tested on five different enoyl-CoA thioesters of varying chain length and branching patterns; crotonyl-CoA, hexenoyl-CoA, 5-methylhexenoyl-CoA, octenoyl-CoA and cinnamoyl-CoA. We determined the carboxylation efficiency for each enzyme with each substrate by quantifying the percentage yield of carboxylated product (compared to total product formed, including reduced side product, **Figure 1A**). We also measured the kinetics for each enzyme with each substrate by following the consumption of NADPH (**Figure 1B**, **Figure S1**, **Table S2**). Based on these parameters, we identified $CcrC_{PAG}$ as most suited ECR for our biocatalytic synthesis strategy,

because the enzyme showed the highest carboxylation efficiencies as well as favorable catalytic activities for all substrates tested, with the exception of cinnamoyl-CoA. With cinnamoyl-CoA, $CcrC_{PAG}$ showed a reduced carboxylation efficiency (57%), as well as a reduced catalytic activity (< 0.15s^{-1}). Cinnamoyl-CoA posed also a challenge for other ECRs. While CinF showed higher carboxylation efficiencies (74%) compared to $CcrC_{PAG}$, the enzyme acted even slower on this substrate (k_{cat} < 0.11s^{-1}). AntE and Ndas_0488 on the other hand displayed two fold higher turnover rates with cinnamoyl-CoA compared to $CcrC_{PAG}$, albeit at reduced carboxylation efficiency.

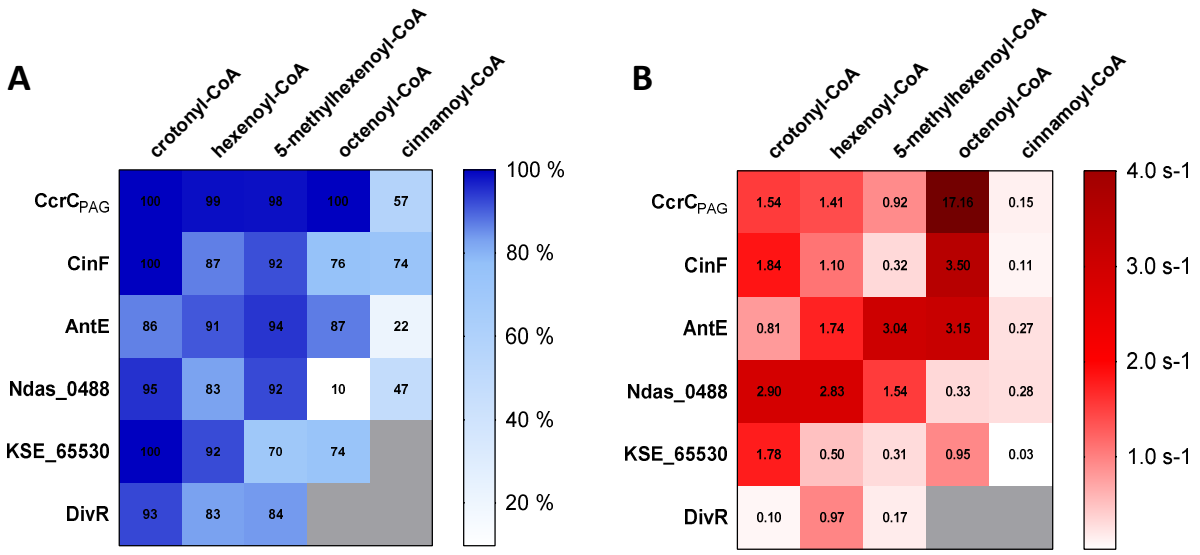


Figure 1 Characterization of ECR variants for their substrate promiscuity. A) Percentage of carboxylation product of promiscuous ECRs dependent on the substrate compared to the reduction side reaction. Assays contained 100 mM TrisHCl pH 8, 100 mM KHCO3, 10 µgmL-1 carbonic anhydrase, 150 µM NADPH, 75 µM substrate and 1.5 µg of the respective ECR, quenched with 5 % formic acid and analyzed by HPLC MS. **B)** kcat of ECRs for all the substrates tested.

5.3.2. Biocatalytic proofreading improved atypical extender unit biosynthesis

The low carboxylation yields and reduced catalytic activities of ECRs with cinnamoyl-CoA posed a problem for the preparative-scale production of benzylmalonyl-CoA. Due to the pronounced reduction side reaction of ECRs with cinnamoyl-CoA, most of the substrate is converted into phenylpropinoyl-CoA, which represents a 'dead-end' product. Thus, we sought to employ the concept of biocatalytic proofreading to our biosynthetic strategy[19]. We reasoned that an efficient regeneration of phenylpropinoyl-CoA back into the substrate cinnamoyl-CoA should increase the yield of the carboxylation reaction in the synthesis assay. Therefore we tested whether the recently described acyl-CoA oxidase Acx4 from *Arabidopsis thaliana* [20] is able to oxidize phenylpropinoyl-CoA into cinnamoyl-CoA. A detailed kinetic characterization of Acx4 showed that the enzyme is not only active with phenlypropionyl-CoA, but accepts many different acyl-CoAs, including the ones that are relevant to this study (butyanoyl-, hexanoyl, octanoyl-, and 5-methylhexanoyl-CoA **Table S3**). In the following we tested the effect of Acx4 onto the AntE-dependent carboxylation of cinnamoyl-CoA (**Figure 2A)**. While the control reaction without proofreading enzyme produced only 19 % of benzylmalonyl-CoA (**Figure 2B**), the assay containing Acx4 yielded 73 % benzylmalonyl-CoA (**Figure 2C**). The addition of catalase to remove the potentially harmful H2O2 produced by Acx4 did not alter the yield of the reaction and was therefore not used in further experiments. These results demonstrated that Acx4 can be directly added as proofreading enzyme to our biosynthetic assays and that biocatalytic proofreading by Acx4 increases carboxylation product yields by almost a factor of four. In the following we used this route to produce benzylmalonyl-CoA at the preparative scale and high purity (**Figure S3A**).

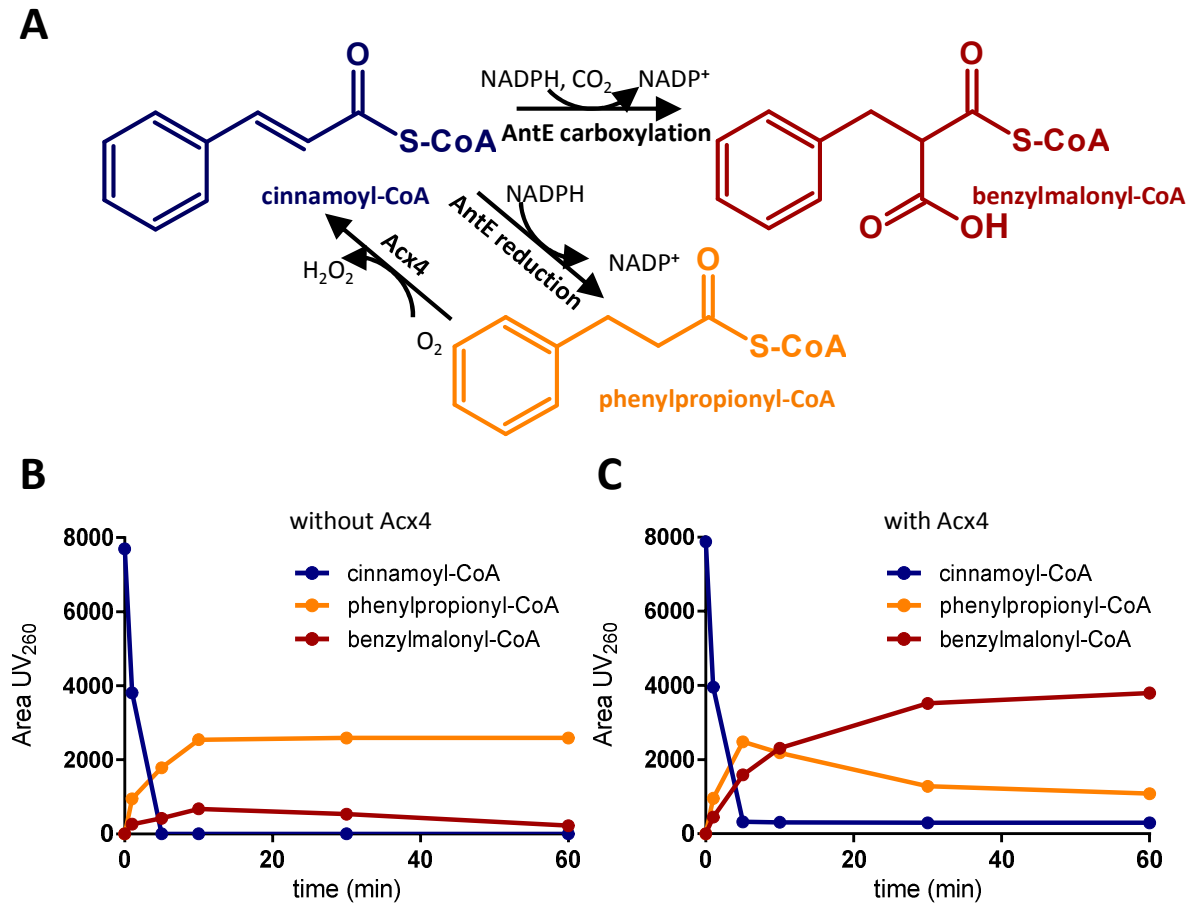


Figure 2 Cinnamoyl-CoA carboxylation using the oxidase Acx4 as a proofreading enzyme to recycle the side product dihydrocinnamoyl-CoA allows for efficient benzylmalonyl-CoA biosynthesis. A) Scheme of the benzylmalonyl-CoA synthesis reaction containing the proofreading oxidase Acx4. **B)** HPLC analysis of an assay containing 100 mM $NaPO_4$ buffer pH 8, 100 mM $KHCO_3$, 100 µM cinnamoyl-CoA, 2 mM NADPH and 5.3 µM AntE. Without proofreading the assay produces mainly the reduced side product dihydrocinnamoyl-CoA **C)** HPLC assay analogue to B containing additionally 2.6 µM of the proofreading enzyme Acx4. Using the proofreading oxidase the assay now mainly yields benzylmalonyl CoA.

5.3.3. A convenient one-pot, preparative-scale biosynthesis of atypical extender units

Acx4 was not only useful for biocatalytic proof reading, but also opened a new route to the synthesis of atypical extender units. We noticed that when promiscuous Acx4 and promiscuous $CcrC_{PAG}$ are combined, it becomes possible to synthesize alkylmalonyl-CoAs directly from the corresponding saturated acyl-CoAs. Acyl-CoAs are chemically easily accessible from free CoA and the respective saturated acids [21-22], which are commercially more available and cheaper than their desaturated counterparts. We demonstrated the new chemo-biocatalytic route in the following. First, we chemically coupled free CoA with butanoyl-tschüssikovski bruuuuudiiiiiiiiii, hexanoyl, 5-methylhexanoyl- and phenylpropinoyl-CoA in preparative scale [21]. Then, the individual synthesis assays were lyophilized, resolved in buffer (100 mM TrisHCl pH 7.5, 100 mM $KHCO_3$) to approx. 1 mM acyl-CoA, and incubated with Acx4 (2.4 µM), $CcrC_{PAG}$ (0.8 µM) as well as NADPH (10 mM) for 120 min at 30 °C (see **Figure S2**). Progress of the individual reactions was followed by HPLC MS (**Figure 3**), assays were quenched after 120 min, CoA esters were purified via HPLC, lyophilized and yields were determined (**Table S4**). Butylmalonyl-, hexylmalonyl-, and 3-methylbutylmalonyl-CoA were produced at yields of >95% from the corresponding acyl-CoA ester and around 30% from the used CoA after HPLC purification. Only the biosynthesis of benzylmalonyl-CoA did not run to full completion under the chosen conditions, even when the concentration of Acx4 and $CcrC_{PAG}$ was increased (**Figure 3E**.

Nevertheless, we could use this new chemo-biosynthetic route to prepare three alkylmalonyl-CoA extender units in the mg scale and at purity >95%, as judged by HPLC-MS (**Figure S4** and **Table S3**).

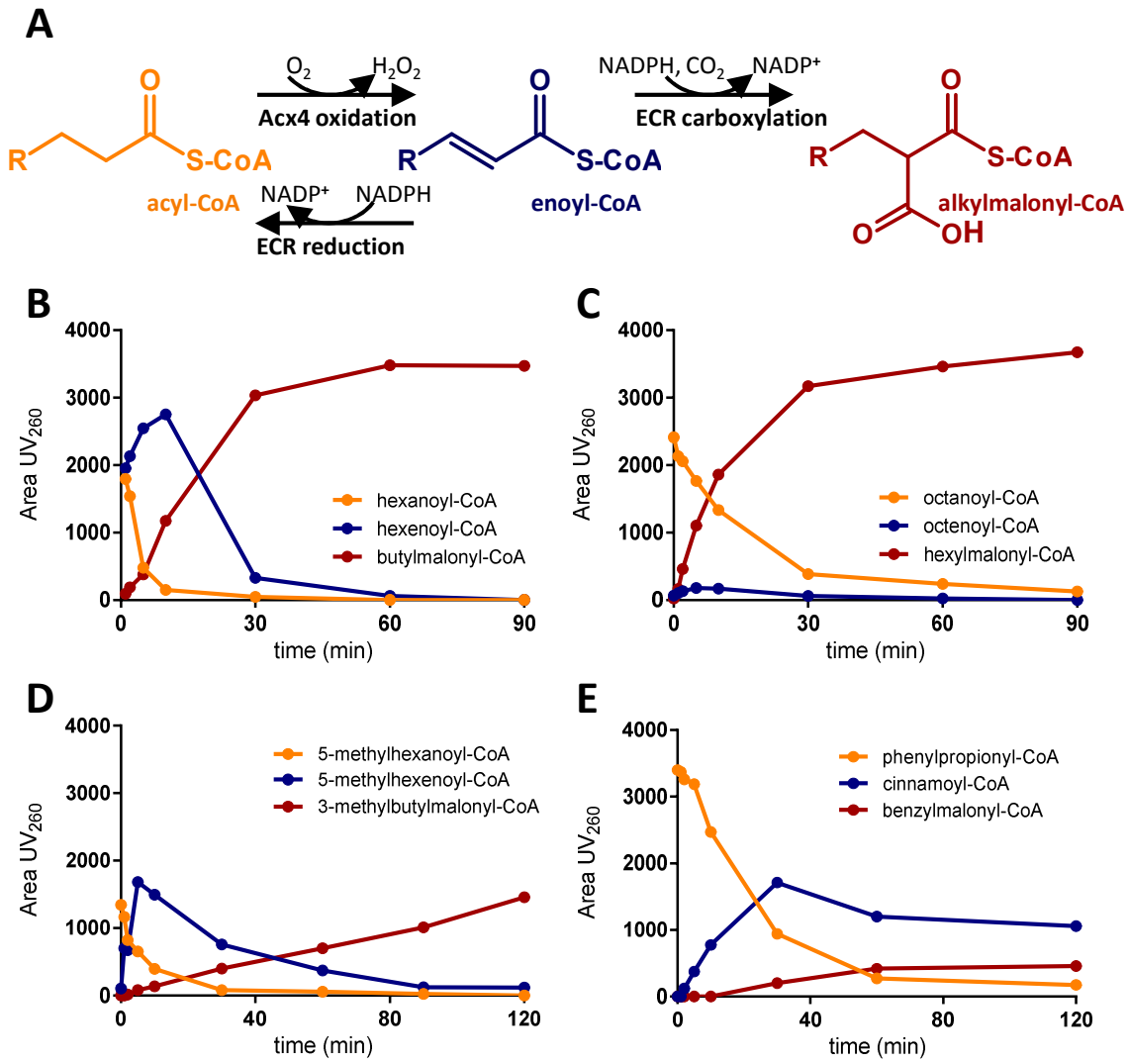


Figure 3 One pot synthesis of atypical polyketide extender units using the acyl-CoA oxidase Acx4 and the reductive carboxylase $CcrC_{PAG}$. A) General reaction scheme. Assays contained contained 100 mM TrisHCl pH 7.5, 100 mM $KHCO_3$, 10 mM NADPH, approximately 1 mM acyl-CoA from the CDI reaction mixture, 2.4 µM Acx4 and 0.8 µM $CcrC_{PAG}$ **B)** HPLC analysis of butylmalonyl-CoA synthesis. **C)** HPLC analysis of hexylmalonyl-CoA synthesis. **D)** HPLC analysis of 3-methylbutylmalonyl-CoA synthesis. **E)** HPLC analysis of benzylmalonyl-CoA synthesis using 5 fold increased $CcrC_{PAG}$ and 8 fold increased Acx4 concentrations compared to the other assays.

5.3.4. Competition assays verify selectivity-conferring active site residues of PKS

The incorporation behavior of PKS in the presence of a large variety of extender units at the same time has been investigated only sparsely so far. This prompted us to utilize our set of six diverse alkylmalonyl-CoAs in PKS competition assays. As model system, we used a modified DEBS PKS *in vitro* model system that produces triketides [18]. We first tested substrate specificity of the WT modules of the model system in a competition experiment containing 1 mM of all extender unit; the natural substrate methylmalonyl-CoA as well as the unnatural extender units malonyl-, ethylmalonyl-, butylmalonyl-, 3-methylbutylmalonyl-, and hexylmalonyl-CoA. Samples were quenched after 60 min incubation at 30 °C and triketides were analyzed by HPLC-ESI-TOF. While the WT modules showed incorporation of ethyl- and butylmalonyl-CoA, they did not accept 3-methylbutyl-, hexyl- or malonyl-CoA (**Figure 4A**), indicating that slightly larger extender units are able to enter the WT modules of the triketide model system when provided at equal concentrations as the natural extender methylmalonyl-CoA. This is well in line with earlier reports on the promiscuity of DEBS WT modules tested with individual atypical extender units [23].

Next, we tested competitive incorporation of atypical extender units into a triketide model system with altered selectivity of the AT domain of module 2. To that end we targeted Y754 that is located in the YASHS motif of the AT domain, which was previously described to confer methylmalonyl-CoA selectivity [24]. We substituted Y754 with valine to increase selectivity of the AT towards longer side chains [25]. The Y745V showed increased incorporation of atypical extender units in the competitive assay, again in line with earlier findings (**Figure 4B**). Having demonstrated that our multi-substrate competition assay can be used to screen extender unit selectivity, we decided to target a, to our knowledge, so far uncharacterized glutamine (Q624 in DEBS module 2), which is highly conserved in methyl- and ethylmalonyl-CoA incorporating ATs. In AT4 of the reveromycin PKS [26], which displays relaxed selectivity towards longer and branched alkylmalonyl-CoA extender units, this canonical glutamine is replaced with histidine. The Q624H variant showed increased incorporation of atypical extender units in the competitive assay, indicating the importance of this residue for substrate selectivity (**Figure 4C**).

Taken together our results demonstrate the potential of the newly developed synthesis route to prepare a set of atypical alkylmalonyl-CoA extender units that can be used for *in vitro* screening and site directed mutagenesis of PKS to assess their substrate selectivity. Our chemo-biosynthetic route can be easily extended for the synthesis of a large library of atypical extender units. Along with recently established *in vitro* polyketide systems, this will open the way to characterize large libraries of mutants in the AT domain, eventually leading to a detailed understanding of the function of active site residues that contribute to substrate selectivity in PKS.

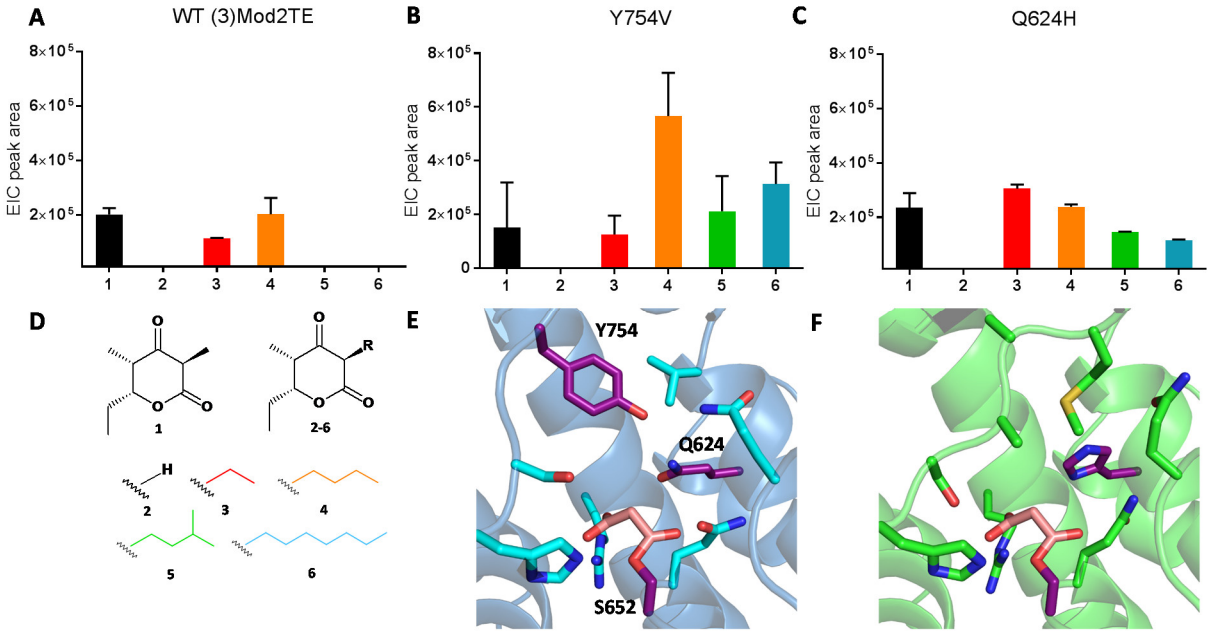


Figure 4 *In vitro* extender unit promiscuity screen for triketide biosynthesis of the DEBS polyketide synthase. The WT systems production was compared to two variants that were altered in the AT domain of module 2. Assays contained 1 mM each of the natural extender unit methylmalonyl-CoA (1) and of the atypical malonyl-CoA (2), ethylmalonyl-CoA (3), butylmalonyl-CoA (4), 3-methylbutylmalonyl-CoA (5) and hexylmalonyl-CoA (6). Assays were incubated for 60 minutes and quenched with 5 % formic acid. **A)** Triketides produced by the WT modules of the *in vitro* DEBS triketide system. **B)** Triketides produced by the Y754V variant of the AT of module 2 of the DEBS triketide system. **C)** Triketides produced by the Q624H variant of the AT of module 2. Both variants additionally incorporate 3-methylbutylmalonyl-CoA (5) and hexylmalonyl-CoA (6). **D)** Scheme of the produced triketides. **E)** Model of the AT active site of DEBS module 2 generated with SWISS-MODEL (Bienert et al. 2017) with the [KS3][AT3] didomain crystal structure (2QO3) as template (Tang et al. 2007). Highlighted in purple are Serine 652 that carries the malonyl-CoA extender unit, as well as Tyrosine 754 and Glutamine 624 that line the alkylmalonyl-CoA binding pocket and that were targeted by site directed mutagenesis in this study. **F)** Model of the promiscuous AT4 from the reveromycine PKS using the same template. Highlighted is the conserved serine that carries the malonyl-CoA extender unit, as well as the histidine which replaces the conserved glutamine of methyl- and ethylmalonyl-CoA specific ATs (i.e., Q624 of DEBS AT2).

5.4. Materials and Methods

5.4.1. Plasmid construction

The expression plasmid for CinF was a generous gift from Rolf Müller [27]. The expression plasmid for $CcrC_{PAG}$ was previously cloned into pTE-16b using NdeI and BamHI, mutations were introduced using the QuickChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA)[7]. The *antE*, *divR, Ndas_0488* and *KSE_65530* were codon optimized and synthesized by the DOE Joint Genome Intitute adding an N-terminal 10x His tag and cloned into the backbone pET16b using NcoI and XhoI (see **Table S4** for complete DNA and protein sequences).

5.4.2. ECR and Acx4 preparation

E.coli BL21 (DE3) (for Acx4 E.coli BL21 AI (DE3)) were transformed with the respective expression plasmid and plated on LB-Agar containing 100 µg/ml Ampicilin (50 µg/ml Kanamycin for CinF). Pooled colonies of overnight inoculated petri dishes were used to inoculate 1 L of TB containing the respective antibiotic. The expression culture was incubated at 37 °C until an OD_{600} of 0.8 was reached. Cultures for ECR expressions were cooled down to 20 °C and induced with 0.25 mM IPTG and incubated over night. Acx4 expression cultures were cooled to 30 °C and induced with 0.25 mM IPTG and 2 ‰ L-arabinose and incubated for four hours. Cells were harvested, resuspended in 50 ml Buffer A (50 mM TrisHCl pH 7.9, 500 mM NaCl) and lysed by sonication. Lysates were centrifuged at 42,000 g at 4 °C for

45 min. The supernatant was then loaded onto a pre-equilibrated 1 mL HisTrap column (GE Healthcare) and washed with 15 % Buffer B (50 mM TrisHCl pH 7.9, 500 mM NaCl, 500 mM imidazole) and eluted with 100 % Buffer B. The buffer of the purified proteins was exchanged into the protein storage buffer (30 % glycerol, 150 mM NaCl, 25 mM TrisHCl pH 7.9). Protein purity was checked by SDS-gel electrophoresis, concentration of proteins were determined spectroscopically at 280 nm using calculated extinction coefficients. For Acx4 the FAD concentration was determined at 450 nm and FAD was added to reach an equimolar protein to FAD ratio. Proteins were stored at -20 °C.

5.4.3. DEBS protein preparation

The expression plasmids for the *in vitro* triketide model system of DEBS were constructed according to a previously described protocol yielding the plasmids pBL12 (here termed pET28b-LD(4); the N-terminal HisTag of pBL12 was exchanged for a StrepTag), pBL13 (pET28b-(5)Mod1(2)) and pBL16 (pET28b-(3)Mod2TE) [18]. *E. coli* BAP1 was transformed with expression plasmid, 1L of TB expression culture was grown to an OD_{600} of 0.7, cooled down to 18°C, induced with 0.25 mM IPTG and incubated over night at 18°C. Cells producing module 1 and module 2 were harvested, resuspended in Buffer C (50 mM NaH_2PO_4 pH 7.6, 450 mM NaCl, 20% glycerol), lysed by sonification and centrifuged at 42,000g at 4°C for 45 min. The supernatant was incubated with pre-equilibrated Protino Ni-NTA beads (1 mL resin) for 1h. The beads were washed with 5% Buffer D (75 mM NaH_2PO_4 pH 7.6, 500 mM imidazole, 20 mM NaCl, 10% glycerol) and then eluted with 100%. The elusion was loaded onto a with Buffer E (50 mM NaH_2PO_4 pH 7.6, 10% glycerol) pre-equilibrated HiTrap Q anion exchange column and eluted in a 50 mL NaCl gradient with a final concentration of 500 nM NaCl. Cells expressing the plasmid yielding the loading domain were resupended in Buffer C, processed as described above and purified over a pre-equilibrated 1 mL StrepTrap column (GE Healthcare), using 2 mM d-Desthiobiotin (Sigma-Aldrich) for elution. Fractions were tested on an SDS-PAGE gel, pooled and concentrated with an Amicon Ultra-4 100 kDa (50 kDa for LD(4)) centrifugal filter. A final concentration of 20% glycerol was added to the concentrated proteins, their concentration was determined spectroscopically at 280 nm and they were stored at -80°C.

5.4.4. Chemical CDI coupling

CDI (1,1'-carbonyldiimidazole) coupling of carboxylic acids to CoA was done according to a previously described protocol [21]. 42 mg CDI (0.27 mmol, 4 eq.) was dissolved in 2 mL THF and the acid was added (0.31 mmol, 4.8 eq.; hexanoic acid: 38.9 µl, octanoic acid: 49.2 µl, dihydrocinnamic acid: 46.5 µl and 5-methylhexanoic acid: 43,5 µl). The reactions were stirred for 1 h at 22 °C. 50 mg CoA was dissolved in 500 µL 1 mL $KHCO_3$ and added to the reaction. The reaction was stirred for another 45 min and tested for remaining free thiols using DTNB, then flash frozen in liquid N_2 and lyophilized overnight. Samples were then dissolved in 1 mL H_2O and used for HPLC-purification or used for desaturation by Acx4 or for direct one-pot biosynthesis of the extender units.

5.4.5. Synthesis of cinnamoyl-CoA using ethylchloroformate coupling

Synthesis of cinnamoyl-CoA was done by chemically coupling cinnamic acid to CoA according to a previously described protocol[21]. 37 mg trans-cinnamic acid (0.25 mmol, 2 eq.) was dissolved in 2 mL DCM, 37 µl trimethylamine (0.26 mmol, 2.1 eq) was added and the reaction was stirred at 20°C for 30 min. The mixture was cooled on ice and 24 µl ethylchloroformate (0.25 mmol, 2 eq.) was added. The mixture was stirred on ice for 2h. The DCM was evaporated at room temperature and resuspended in 2 mL DMF. 100mg CoA (0.125 mmol, 1 eq.) was dissolved in 300µL 1 M $KHCO_3$ and added to the mixture at 20 °C. Completion of the reaction was tested using DTNB, the reaction mixture was quenched by adding 500 µL of 50 % formic acid, frozen in liquid N_2 and lyophilized overnight. The sample was dissolved in 2 mL H_2O and used for HPLC purification.

5.4.6. Synthesis of malonyl-, methylmalonal- and ethylmalonyl-CoA

Synthesis of malonyl- and methylmalonyl-CoA was done according to a previously described protocol [21]. 20 mg CoA (1 eq.), 13.2 mg malonic acid (5 eq.) or 15.2 mg methylmalonic acid (5 eq.) and 70.4 mg ATP (4 eq.) were dissolved in 5 mL of 200 mM $KHCO_3$ containing 15 mM $MgCl_2$ and 3.2 µM MatB. The reaction was incubated at 30°C and completion confirmed with DTNB. The mixture was quenched with 5% formic acid and directly used for HPLC purification. Ethylmalonyl-CoA was synthesized by chemical coupling of crotonic anhydride with CoA followed by the addition of a $CcrC_{PAG}$, NADPH and $KHCO_3$. 10 mg CoA was dissolved in 1 mL 1 M $KHCO_3$.The mixture was cooled on ice and 3.2 µl of crotonic anhydride was added and stirred on ice for 30 min. Completion of the reaction was tested using DTNB. The reaction mixture was then added to 8 mL of 250 mM TrisHCl pH 7.5 containing 30 mM NADPH, 250 mM $KHCO_3$ and 100 nM CcrCc IA/CP/FG. The reaction was incubated for 30 min at 30 °C, quenched with final concentration 5 % formic acid and directly used for HPLC purification.

5.4.7. One pot biosynthesis of extender units

The enzymatic desaturation with Acx4 as well as the reductive carboxylation was then done using this lyophilized reaction mixture without any purification in between. The lyophilized CDI reaction mixture (see above) was resuspended in 1 mL of H_2O and used directly for the biosynthesis of the extender units. A 10 mL assay for extender unit biosynthesis contained 2 mL of 0.5 M TrisHCl pH 7.5 (final concentration 100 mM), 1 mL of 1M $KHCO_3$ (100 mM), 500 µL of 200mM NADPH (10 mM), 300 µl of the dissolved CDI reaction mixture (theoretical maximum of 50 mM, resulting in an approximate final concentration of 1 mM in the assay), 500 µl of 48 µM Acx4 (2.4 µM) and 200 µl of 39 µM $CcrC_{PAG}$ (0.8 µM). The mixture was shaken at 200 rpm in a buffled Erlenmeyer flask at 30 °C for 120 min. To follow progress of the reaction, 30 µl of the reaction mixture were quenched at various time points with the addition of 3 µl of 50 % formic acid and analyzed by analytical HPLC-MS (**Figure 3**). The reaction was quenched after 120 min with the addition of 1 mL of 50 % formic acid, centrifuged for 10 min at 17000 g to remove precipitated protein and then purified via HPLC. The purified product was lyophilized, resuspended in water and the final yield of the overall reaction and purification was determined by determining the extender unit concentration via UV/Vis absorption at 260 nm (calculated as (mol product)/(mol CoA) **Table S3**).

5.4.8. HPLC purification of acyl-CoA esters

All synthesized CoA-thioesters were purified using a 1260 Infinity LC system (Agilent) using a Gemini 10 µm NX-C18 110 Å, 100 x 21.2 mm, AXOA packed column (Phenomenex). The general protocol used a flow rate of 25 ml*min^{-1} starting with 5 min of 5 % MeOH in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5 % to 40 % MeOH in 15 min, a 2 min washing step at 95 % MeOH and a re-equilibration step of 3 min at 5 % MeOH. Purified CoA-esters were lyophilized and the purity was checked using analytical HPLC-MS according to the protocol described for the carboxylation vs. reduction assays.

5.4.9. Determination of carboxylation vs. reduction ratio of ECR variants

Assays for the determination of carboxylaton vs. reduction contained 100 mM TrisHCl pH 8, 100 mM $KHCO_3$, 10 µgmL^{-1} carbonic anhydrase, 150 µM NADPH, 75 µM substrate and 1.5 µg of the respective ECR. The reaction was followed on a Cary-60 UV/Vis spectrometer and quenched upon completion with final concentration 5 % formic acid. Assays were then analyzed on a 1260 Infinity LC system (Agilent) with attached with a 6130 Quadrupole LC/MS detector (Agilent) using a Luna 3 µm C18(2) 100 Å, 150 x 2 mm (Phenomenex) column with a flow rate of 0.3 ml*min^{-1}. For malonyl-, methylmalonyl- and ethylmalonyl-CoA a gradient starting with 3 min of 5 % MeOH in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5 % to 25 % MeOH over 20 min, a 2 min washing step at 95 %

MeOH and a 3 min re-equilibration step at 5 % MeOH. All other samples were run using a gradient starting with 3 min of 5 % MeCN in 50 mM NH_4HCO_2 pH 8.2, followed by a gradient from 5 % to 35 % MeCN over 20 min, a 2 min washing step at 95 % MeCN and a 3 min re-equilibration step at 5 % CN.

5.4.10. Spectrophotometric enzyme assays

Enzyme assays were carried out on a Cary-60 UV/Vis spectrometer (Agilent) using quartz cuvettes (1- and 10 mm diameter, Hellma). Acx4 characterization was done using 10 mm cuvettes in 100 mM TrisHCl pH 8 buffer following double bond formation at 290 nm ($\Delta\varepsilon_{290}$ = 2.26 $mM^{-1}cm^{-1}$) for all substrates except for phenylpropionyl-CoA (dihydrocinnamoyl-CoA), which was measured at 308 nm ($\Delta\varepsilon_{308}$.=16.4 $mM^{-1}cm^{-1}$, **Figure S2B**). ECR characterization for all substrates was in 1-mM cuvettes following NADPH consumption at 360 nm in 100 mM TrisHCl pH8 buffer containing 100 mM $KHCO_3$ and 500µM NADPH (**Figure S4**).

5.4.11. Extender unit incorporation assays

Competition assays to test incorporation of extender units into triketides were run at 30°C in 200 mM NaH_2PO_4 with 2 mM NADPH, 4 µM epimerase, 0.2 mM propionyl-CoA, 1 mM of each extender unit and 1 µM of LD(4), (5)Mod1(2) and one of the (3)Mod2TE variants. Assays were quenched with 5% formic acid after 60 min and analyzed via HPLC-ESI-TOF on a 6550 iFunnel Q-TOF LC-MS (Agilent) using a 1.8 µm Zorbax SB-C18 column, 50 x 2.1 mm (Agilent) using H_2O (A) and MeCN (B) both containing 0.1% formic acid. The gradient condition were as follows: 0 min 5 % B; 1 min 5 % B; 6 min 95 % B,6.5 min 95 % B, 7 min 5 % B with a flow rate of 250 µl/min. Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 l/min,225 °C) and sheath gas (12 l/min, 40°C). MS data were acquired with a scan range of 100-1000 m/z. Data were analyzed using MassHunter Qualitative Analysis software (Agilent) and eMZed [28].

5.5. References

1. Chan, Y. A.; Podevels, A. M.; Kevany, B. M.; Thomas, M. G., *Natural product reports* **2009,** *26* (1), 90-114.
2. Ray, L.; Moore, B. S., *Natural product reports* **2016,** *33* (2), 150-61.
3. Moore, B. S.; Hertweck, C., *Natural product reports* **2002,** *19* (1), 70-99.
4. Wilson, M. C.; Moore, B. S., *Natural product reports* **2012,** *29* (1), 72-86.
5. Erb, T. J.; Berg, I. A.; Brecht, V.; Muller, M.; Fuchs, G.; Alber, B. E., *P Natl Acad Sci USA* **2007,** *104* (25), 10631-10636.
6. Erb, T. J.; Brecht, V.; Fuchs, G.; Muller, M.; Alber, B. E., *P Natl Acad Sci USA* **2009,** *106* (22), 8871-8876.
7. Peter, D. M.; Schada von Borzyskowski, L.; Kiefer, P.; Christen, P.; Vorholt, J. A.; Erb, T. J., *Angewandte Chemie* **2015,** *54* (45), 13457-61.
8. Koryakina, I.; McArthur, J. B.; Draelos, M. M.; Williams, G. J., *Org Biomol Chem* **2013,** *11* (27), 4449-4458.
9. Yan, J.; Hazzard, C.; Bonnett, S. A.; Reynolds, K. A., *Biochemistry* **2012,** *51* (46), 9333-41.
10. Liou, G. F.; Khosla, C., *Curr Opin Chem Biol* **2003,** *7* (2), 279-84.
11. Koryakina, I.; Kasey, C.; McArthur, J. B.; Lowell, A. N.; Chemler, J. A.; Li, S.; Hansen, D. A.; Sherman, D. H.; Williams, G. J., *ACS chemical biology* **2017,** *12* (1), 114-123.
12. Yuzawa, S.; Deng, K.; Wang, G.; Baidoo, E. E.; Northen, T. R.; Adams, P. D.; Katz, L.; Keasling, J. D., *ACS synthetic biology* **2017,** *6* (1), 139-147.
13. Tang, Y.; Tsai, S. C.; Khosla, C., *Journal of the American Chemical Society* **2003,** *125* (42), 12708-9.
14. Koryakina, I.; Williams, G. J., *Chembiochem : a European journal of chemical biology* **2011,** *12* (15), 2289-93.
15. Koryakina, I.; McArthur, J.; Randall, S.; Draelos, M. M.; Musiol, E. M.; Muddiman, D. C.; Weber, T.; Williams, G. J., *ACS chemical biology* **2013,** *8* (1), 200-8.
16. Pohl, N. L.; Hans, M.; Lee, H. Y.; Kim, Y. S.; Cane, D. E.; Khosla, C., *Journal of the American Chemical Society* **2001,** *123* (24), 5822-3.
17. Zhang, L.; Mori, T.; Zheng, Q.; Awakawa, T.; Yan, Y.; Liu, W.; Abe, I., *Angewandte Chemie* **2015,** *54* (45), 13462-5.
18. Lowry, B.; Robbins, T.; Weng, C. H.; O'Brien, R. V.; Cane, D. E.; Khosla, C., *Journal of the American Chemical Society* **2013,** *135* (45), 16809-12.
19. Sun, J.; Jeffryes, J. G.; Henry, C. S.; Bruner, S. D.; Hanson, A. D., *Metabolic engineering* **2017,** *44*, 150-159.
20. Schwander, T.; Schada von Borzyskowski, L.; Burgener, S.; Cortina, N. S.; Erb, T. J., *Science* **2016,** *354* (6314), 900-904.
21. Peter, D. M.; Vogeli, B.; Cortina, N. S.; Erb, T. J., *Molecules* **2016,** *21* (4).
22. Agarwal, V.; Diethelm, S.; Ray, L.; Garg, N.; Awakawa, T.; Dorrestein, P. C.; Moore, B. S., *Org Lett* **2015,** *17* (18), 4452-5.
23. Dunn, B. J.; Watts, K. R.; Robbins, T.; Cane, D. E.; Khosla, C., *Biochemistry* **2014,** *53* (23), 3796-806.
24. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R., *Biochemistry* **2001,** *40* (51), 15464-70.
25. Smith, L.; Hong, H.; Spencer, J. B.; Leadlay, P. F., *Chembiochem : a European journal of chemical biology* **2008,** *9* (18), 2967-75.
26. Takahashi, S.; Toyoda, A.; Sekiyama, Y.; Takagi, H.; Nogawa, T.; Uramoto, M.; Suzuki, R.; Koshino, H.; Kumano, T.; Panthee, S.; Dairi, T.; Ishikawa, J.; Ikeda, H.; Sakaki, Y.; Osada, H., *Nat Chem Biol* **2011,** *7* (7), 461-8.
27. Quade, N.; Huo, L. J.; Rachid, S.; Heinz, D. W.; Muller, R., *Nat Chem Biol* **2012,** *8* (1), 117-124.
28. Kiefer, P.; Schmitt, U.; Vorholt, J. A., *Bioinformatics* **2013,** *29* (7), 963-964.

5.6. Supplementary Information

Table S1 (related to Figure 1); Overview of the selected promiscuous ECRs

Enzyme	Phylum	Family	Organism	Promiscuity motif			Accession code	Source
$CcrC_{PAG}$	Proteobacteria	Caulobacteriaceae	Caulobacter crescentus	P	A	G	WP_010920921	Peter D. (10.1002/anie.201505282)
CinF	Actinobacteria	Streptomycetaceae	Streptomyces sp. JS360	P	A	G	WP_067247071.1	Huo L. (10.1038/nchembio.734)
AntE	Actinobacteria	Streptomycetaceae	Streptomyces sp. NRRL2288	A	V	V	AGG37751.1	synthesized DOE-JGI (this work)
Ndas_0488	Actinobacteria	Nocardiopsaceae	Nocardiopsis dassonvillei	P	V	F	WP_013151542.1	synthesized DOE-JGI (this work)
KSE_65530	Actinobacteria	Streptomycetaceae	Kitasatospora setae	T	A	G	YP_004908268.1	synthesized DOE-JGI (this work)
DivR	Actinobacteria	Streptomycetaceae	Streptomyces sp. HKI0576	A	A	V	CCA30183	synthesized DOE-JGI (this work)

Table S2 (related to Figure 1); k_{cat} and K_m of the ECR variants towards the tested substrates, Values with yellow background were fitted using substrate inhibition. ND = not determinable, errors are given as standard error.

k_{cat} [s^{-1}]

	crotonyl-CoA	hexenoyl-CoA	5-methylhexenoyl-CoA	octenoyl-CoA	cinnamoyl-CoA
$CcrC_{PAG}$	1.5 ± 0.1	1.4 ± 0.1	0.9 ± 0.1	17 ± 1	0.15 ± 0.01
CinF	1.84 ± 0.03	1.10 ± 0.04	0.32 ± 0.03	3.5 ± 0.7	0.113 ± 0.003
AntE	0.81 ± 0.04	1.7 ± 0.1	3.0 ± 0.1	3.1 ± 0.2	0.27 ± 0.01
Ndas_0488	2.9 ± 0.5	2.8 ± 0.5	1.5 ± 0.1	0.33 ± 0.01	0.28 ± 0.02
KSE_65530	1.78 ± 0.05	0.50 ± 0.02	0.6 ± 0.1	1.0 ± 0.4	0.03 ± 0.003
DivR	0.089 ± 0.003	1.0 ± 0.2	0.17 ± 0.01	ND	ND

Km [μM]

	crotonyl-CoA	hexenoyl-CoA	5-methylhexenoyl-CoA	octenoyl-CoA	cinnamoyl-CoA
$CcrC_{PAG}$	525 ± 85	74 ± 10	153 ± 45	75 ± 15	3 ± 2
CinF	154 ± 10	16 ± 2	53 ± 11	11 ±4	52 ± 6
AntE	218 ± 31	103 ± 18	125 ± 18	133 ± 18	42 ± 6
Ndas_0488	38 ± 17	58 ± 17	30 ± 5	45 ± 6	160 ± 35
KSE_65530	111 ± 8	33 ± 5	36 ± 17	100 ± 60	ND
DivR	10 ± 2	96 ± 30	56 ± 9	ND	ND

Table S3 (related to Figure 2) Kinetic characterization of the acyl-CoA oxidase Acx4

Substrate	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m [$s^{-1}M^{-1}$]
butyryl-CoA	30 ± 1.1	66 ± 9	$(4.5 \pm 0.8)\cdot10^5$
hexanoyl-CoA	66 ± 2.1	31 ± 3	$(2.2 \pm 0.3)\cdot10^6$
octanoyl-CoA	2.8 ± 0.3	33 ± 5	$(8.5 \pm 1.3)\cdot10^4$
5-methylhexanoyl-CoA	7.2 ± 0.3	32 ± 7	$(2.3 \pm 0.5)\cdot10^5$
phenylpropionyl-CoA	0.21 ± 0.01	12 ± 2	$(1.8 \pm 0.3)\cdot10^4$

Assays for butyryl-Coa, hexanoyl-CoA, octanoyl-CoA and 5-methylhexanoyl-CoA were measured at 290 nm following the double bond formation using $\Delta\varepsilon_{290}$ = 2.26 $cm^{-1}mM^{-1}$. The assay for phenylpropionyl-CoA was measured at 308 nm using $\Delta\varepsilon_{290}$ = 16.6 $cm^{-1}mM^{-1}$ as determined from the spectra of purified phenylpropionyl-CoA and cinnamoyl-CoA (**Figure S2**)

Table S4 (related to Figure 3); Yields of extender unit biosynthesis

Type of acyl-CoA	Yield for chemical coupling	Yield of one pot synthesis	Overall yield after HPLC purification
butylmalonyl-CoA	57%	99%	27%
3-methylbutylmalonyl-CoA	50%	100%	36%
hexylmalonyl-CoA	72%	93%	28%
benzylmalonyl-CoA	49%	27%	-
benzylmalonyl-CoA (one pot from cinnamoyl-CoA without proofreading)	-	8%	-
benzylmalonyl-CoA (one pot from cinnamoyl-CoA)	-	73%	37%

Table S5 (related to Figure 1); Sequences of enzymes used in this study

$CcrC_{PAG}$	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGACGATTCAGACGCTTGAGACCACCGC GCTGAAGGACCTGTACGAGATCGGTGAGATCCCGCCGGCCTTCCACGTGCCGAAAACCATGTACGCCTGGAGCATCCGCAAGGAGCGCCA CGGCAAGCCGACCCAGGCCATGCAGGTCGAGGTGGTGCCCACCTGGGAGATCGGCGAGGACGAGGTGCTGGTGCTCGTGATGGCCGCC GGCGTCAATTACAACGGGGTCTGGGCCGCTCTGGGCGAGCCGATCAGCCCGCTGGACGGCCACAAGCAGCCTTTCCACATCGCCGGGTCC GACGCCTCGGGCATCGTCTGGAAGGTCGGCGCCAAGGTGAAGCGGTGGAAGCTGGGCGACGAGGTCGTCATCCACCCGAACCAGGACG ACGGCGACGACGAGGAGTGCAACGGCGGCGACCCGATGTTCTCGTCCAGCCAGCGCGCGTGGGGCTACGAGACGCCGGACGGCAGCTTC GCCCAGTTCTGCCGGGTGCAGTCGCGCCAGCTGCTGCCGCGCCCCAAGCACCTGACCTGGGAAGAGAGCGCCTGCTACACCCTGACCCTG GCCACCGCCTACCGCATGCTGTTCGGCCACAAGCCGCATGAGCTGAAGCCCGGCCAGAACGTGCTGGTCTGGGGCGCCTCGGGCGGTCTT GGCGTCTTCGCCACCCAGCTGGCCGCCGTGGCCGGCGCCAACGCCATCGGCGTGGTGTCCTCCGAGGATAAGCGCGAGTTCGTGCTGTCG ATGGGCGCCAAGGCGGTGCTGAACCGGGGCGAGTTCAACTGCTGGGGCCAGCTGCCGAAGGTCAACGGCCCCGAGTTCAACGACTACAT GAAGGAGAGCCGCAAGTTCGGCAAAGCCATCTGGCAGATCACCGGAAACAAGGACGTCGACATGGTGTTCGAGCACCCTGGCGAGCAGA CCTTCCCGGTGTCGGTGTTCCTGGTCAAGCGCGGCGGCATGGTGGTGATCTGCGCCGGCACGACGGGCTTCAACCTGACCATGGACGCCC GCTTCCTGTGGATGCGCCAGAAGCGCGTGCAAGGGTCGCACGGCGCCAACCTGATGCAGGCCAGCGCCGCCAACCAGCTGGTCATCGACC GCCGCGTCGATCCCTGCCTGTCGGAAGTCTTCCCCTGGGACCAGATCCCGGCGGCCCACGAGAAGATGCTGGCCAACCAGCACCTGCCGG GGAACATGGCCGTGCTGGTCTGCGCCCAGCGCCCCGGCCTGCGCACCTTCGAGGAAGTGCAGGAGCTGAGCGGGGCGCCATAG
	Protein	MGHHHHHHHHHHSSGHIEGRHMTIQTLETTALKDLYEIGEIPPAFHVPKTMYAWSIRKERHGKPTQAMQVEVVPTWEIGEDEVLVLVMAAGV NYNGVWAALGEPISPLDGHKQPFHIAGSDASGIVWKVGAKVKRWKLGDEVVIHPNQDDGDDEECNGGDPMFSSSQRAWGYETPDGSFAQF CRVQSRQLLPRPKHLTWEESACYTLTLATAYRMLFGHKPHELKPGQNVLVWGASGGLGVFATQLAAVAGANAIGVVSSEDKREFVLSMGAKAV LNRGEFNCWGQLPKVNGPEFNDYMKESRKFGKAIWQITGNKDVDMVFEHPGEQTFPVSVFLVKRGGMVVICAGTTGFNLTMDARFLWMRQ KRVQGSHGANLMQASAANQLVIDRRVDPCLSEVFPWDQIPAAHEKMLANQHLPGNMAVLVCAQRPGLRTFEEVQELSGAP
AntE	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGACTAAAGACCTGTACGAACTGGGCG ACGCTCCGCCACTGGGTACTGCGCCGAAACAGATGTATGCTTCCCTTATCCGCCAGGAGCGTTATGGCCGTCCGGTTGACGCTTTCCGTACC GAAGTAGTGGATGTACCGCCGGTTGGTCCGGGTCAGGTACTGGTTAAAGTAATGGCAGCTGGCGTAAACTACAACAACGTGTGGGCGGC ACTGGGTGAGCCACTGGACGTAATTGCAGCGCGCCAGAAACAGGGCGCTACCGAAGACTTTCATATCGGCGGTTCAGACCTGTCTGGCAT TGTATGGGCGGTTGGTGACGGGGTACGTCTGAAACCGGGTGCAGAGGTTGTGGTTCTGGCTTGTCGTTGGGACGAATCTGCACAGGACAT CCGCCTGGGCGCTGACCCGGTTACCTCTTCTACTCAGCGTGTATGGGGCTATGAAGAAAACTACGGTTCTTTCGCTCAGTTTGCAGTTGTAG ACGAATATATGTGCCACCCGAAACCGCAGCGCCTGAGCTGGGCTGCGGCTTCTTGCTACATGGCCACCGCCGCGACCGCCTATCGTCAGCT GTTCGGCTGGGAACCGCACACTGTTCGTCCGGGTGACCCGGTTCTGATCTGGGGCGGTGCAGGTGGTCTGGGCTCCATCGCGATCCAGCT GGTACGTCACGTTGGTGGCATCCCGGTAGCTGTGGTGAGCTCCGAAGAACGTGGTGAATTCTGTATGCGCCTGGGCGCGAAAGGTTGGAT CGACCGTCGTGAATTCGACCACTGGGGTCGTCTGCCGGACACCACTGACGAGGAAGCTATGAAACAGTGGCTGGACGGCGCGCGCGCTTT CGGTCGTCGTTTCTGGGAAGTTCTGGGTGAACGCCGTGCGCCGCGTATCGTTCTGGAACACTCTGGTGCTGACACCATCCCGACTTCCATCT ATATGGCTGACAACGCCGGTATGGTGGTTATCTGCGGTGGCACGACTGGCTATAACGGTGACGTTGACCTGCGTTTCCTGTGGATGCGTCA AAAGCGCCTGCAGGGTAGTCACGTAGCTTCTGCACGCGAAGCACGTGAAATCACTCGCCTGATCGACCAAGGTGCAATCGACCCGTGCCT GTCCCGTACCTTCGGTTTCGAAGAGATCGGCCTGGCTCACCAGCTGATCCACGATAACCAGCACCCGTCTGGTAACATGGCTGTTCTGGTTA ACGCAACGGAGTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMTKDLYELGDAPPLGTAPKQMYASLIRQERYGRPVDAFRTEVVDVPPVGPGQVLVKVMAAGVNYNNVW AALGEPLDVIAARQKQGATEDFHIGGSDLSGIVWAVGDGVRLKPGAEVVVLACRWDESAQDIRLGADPVTSSTQRVWGYEENYGSFAQFAVV DEYMCHPKPQRLSWAAASCYMATAATAYRQLFGWEPHTVRPGDPVLIWGGAGGLGSIAIQLVRHVGGIPVAVVSSEERGEFCMRLGAKGWI DRREFDHWGRLPDTTDEEAMKQWLDGARAFGRRFWEVLGERRAPRIVLEHSGADTIPTSIYMADNAGMVVICGGTTGYNGDVDLRFLWMR QKRLQGSHVASAREAREITRLIDQGAIDPCLSRTFGFEEIGLAHQLIHDNQHPSGNMAVLVNATE

Ndas_0488	DNA	ATGGGCCATCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGCCGAAAGACCTGTACGACCTCGGCG AAGTTCCGCCGCTGGGTCACGTTCCGGCACGTATGCATGCTTTCACTCTGCGTCGCGAACGTTACGGTCAGCCGCGTGAAGCCTTCGCTCG TGAGGTAGTTCCGGTCCCGCGTCTGGAACCGGGTCACGTACTGGTTTACACCATGAGCGCGGGCATTAACTATAACAACGTATGGGCTGCA ATGGGTCACCCGGTTGACGTAATCGCAATGCGTGAAAAGGTGGGTGCGGCTGAAGAGTACCACATCGGCGGTTCTGACGGTGCCGGTGT TGTGTGGGCTGTTGGTGAAGGTGTGCGCGGTGTTGAGGTTGGTGACCACGTAATCATCGCACCGGGCCAGTGGGATGAATCCGCTGACG ACATTCGTATGGGTCGCGACCCGATCGCATCCCAGTCTATGCGTGTTTGGGGTTACGAGAACAACTTTGGCTCCTTCGGTCAGTTCACTCTG GTACGTGACATCCAGTGTCACCGCAAACCGGAAAACCTCCCGTGGGACGTTGCAGGCGGTTTCCTGGTATCTGCTGCCACCGCTTACCGTC AGCTGTTCGGTTGGGAACCGAACGTTGTACGCCCGGGTGACCCGGTTCTGATCTGGGGTGGTGCTGGTGGTCTGGGCACATCCGCAATCC AGCTGGCTCGTCAAGTTGGTGGCAAGCCGGTAGCTGTGGTTTCCACCGAAGACAAAGCTCGTATGTGCCGTGAACTGGGCGCAGTAGGCG TTATCCAGCGTACTGAATTCGACCATTGGGGTCGTGTGCCGGACGAGGGCGACACACAGGCTTACGCATCCTGGATGCGTGGTGTGCGCG CATTCGGCAAGCGCTTCTGGGAAGAACTGGGCGAGCGTCGTGCACCACGTATCGTTTTCGAACACACTGGTGCAGACACTCTGCCGACTTC TCTGTACCTGTGCGACAATGCTGGCATGGTTGTACTGTGCGGTGCAACCTCTGGTTTCCAGGCAGACGTTGATCTGCGCTTCCTGTGGATGC GTCTGAAACGCCTGCAGGGTTCCCACTTCGCATCTCCGGCTCAGTGCCGCATGGTAATCGACCTGGTAGCAGGTGGCCAGCTGGATCCGTG TGTTACTCGTATCGTGGAGTTCGACGAAATTGGTGAAGCGCATCAGCTGATCCGCGACAACGCTCAGCCGCCGGGCAACATGTCTGCTCTG GTAAACGCACGTGCAGGCCAGACTGGTCTGGACGTTTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMPKDLYDLGEVPPLGHVPARMHAFTLRRERYGQPREAFAREVVPVPRLEPGHVLVYTMSAGINYNNVWA AMGHPVDVIAMREKVGAAEEYHIGGSDGAGVVWAVGEGVRGVEVGDHVIIAPGQWDESADDIRMGRDPIASQSMRVWGYENNFGSFGQF TLVRDIQCHRKPENLPWDVAGGFLVSAATAYRQLFGWEPNVVRPGDPVLIWGGAGGLGTSAIQLARQVGGKPVAVVSTEDKARMCRELGAV GVIQRTEFDHWGRVPDEGDTQAYASWMRGVRAFGKRFWEELGERRAPRIVFEHTGADTLPTSLYLCDNAGMVVLCGATSGFQADVDLRFLW MRLKRLQGSHFASPAQCRMVIDLVAGGQLDPCVTRIVEFDEIGEAHQLIRDNAQPPGNMSALVNARAGQTGLDV
KSE_65530	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGCAGATGGACGAACTGACAGAAGCGG TTCTGCGTGACGCTCCGCCGGAGGAACTGTCTCGTCTGCCGCTGCCGAAAGACTATACCGCAGCACACCTGCTGCGCCAGGACGTAGAAAT GTTCCATGGCGTGGCTGATAAAGACGTTCGCCGCAGCCTGCATGTTGGTCGCGTACCGCTGCCGGAACTGGCTCCGGACGAAGTGCTCGT TGCTGTTATGGCGTCTTCCGTTAACTACAACACCGTATGGTCTGCAACTTTCGAGCCGGTTTCCACCTTCGACGCTCTGCGTCGTTACGCTCG CAGCGGTGGCTGGCAGGCACGCCACGATCAGCCGCACCAGGTAATCGGCTCTGACGCAGCTGGTGTTATCGTACGTACGGGCGCTGGCGT ACGCCGTTGGCAGGTAGGTGATCATGTTGCGGTATCTACCGCTGTGGTGGACGACCAGGACCCAGTTACTCACACTGACGGCATGCTGGG TGCAGATCAGAAGGCATGGGGTTACGAAACCAACTTCGGCGGCTTGGCTCACTACACTGTTGTTCGTGCGTCTCAGCTGATCGCTAAACCG CCACACCTGACCTGGGAAGAAACTGCTAGCATCCCGCTGTGCGGTGGCACAGCATACCGTATGTTAGTTTCTGAAAAAGGTGCTCGTATCA AACAGGGTGACATCGTACTTATCTGGGGTGCGTCCGGTGGTCTGGGTGCATTCGCTGTTCAGCTGGTTAAGAACGGTGGTGGTATTCCGG TAGGTGTTGTAAACTCTGAACGTAAAGCAGAACTTGTACGTCGTCTGGGCTGTGACGTAGTTATCAACCGTGAAGAAATTGGTATCGGTAA AGCACCGGAAAGCCCGCAGGAAACTGTTGAACTGGCTAAGCGCCTGGGTCGTGCCATCCGTTCTCAGGTTGGTGAAGACCCGCACATCGT ATTCGACCATGTAGGTAAAGCAACCTTTGGCATCAGCGTTATCGTTGCGCGTCGTGGTGGTACCGTTGTAACTTGCGGCTCTTCCACCGGCT ACCAGCACACCTTCGATAACCGTTACTTCTGGATGAACCTGAAGCGTATCGTAGGTTCTCATGGTATGAACCTGGGCGAGGCTGCTGAAAT GATGCGTCTGTACAAACTGGGTCTGCTGGCACCGGTTGTTTCTCGCACTTACCCACTGGCAGAAGTTGGCGAAGCGGCTCGTCTGGTTCAA AACAACCAGCACACCGGCAAAATCGGTGTGCTGTGTCTGGCAGACCAACCGGGTCTGGGTGTTACCGACCCGGCAACTCGCGCTCGTCTG GGCGAGGACTGGCTGCGTCCGCTGGCAGAGGACCGTGTTCCGGCTCTGGCTGGCTAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMQMDELTEAVLRDAPPEELSRLPLPKDYTAAHLLRQDVEMFHGVADKDVRRSLHVGRVPLPELAPDEVLV AVMASSVNYNTVWSATFEPVSTFDALRRYARSGGWQARHDQPHQVIGSDAAGVIVRTGAGVRRWQVGDHVAVSTAVVDDQDPVTHTDGM LGADQKAWGYETNFGGLAHYTVVRASQLIAKPPHLTWEETASIPLCGGTAYRMLVSEKGARIKQGDIVLIWGASGGLGAFAVQLVKNGGGIPV GVVNSERKAELVRRLGCDVVINREEIGIGKAPESPQETVELAKRLGRAIRSQVGEDPHIVFDHVGKATFGISVIVARRGGTVVTCGSSTGYQHTFD NRYFWMNLKRIVGSHGMNLGEAAEMMRLYKLGLLAPVVSRTYPLAEVGEAARLVQNNQHTGKIGVLCLADQPGLGVTDPATRARLGEDWLR PLAEDRVPALAG
DivR	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGGCGGTGGGTTCTCTGACCGATGCACCC GGCAGACGCAGTTTCTGCTGAAGAAAACGTTGGTGCAGACCTGCCGGCAGAATTCCGTGCAGTACATATCTGCGCGGAAGACGAAGCAAT GTTCTCCGACCTGTCTGATAAAGACGTACGTAAATCCCTGCGCGTTGGCCCGGTTGCTATGCCAGAACTGGCTCCGGACGAAGTTCTGATC GCAGTTATGGCATCTTCTGTTAACTACAACACCGTATGGAGCGCTCAGTTCGCACCGGTGTCTCCGTTTCGCTTCCTGGAGAAATTCGGTCG TCGTGGTGGCTGGGCTGCACGTCACGACCTGCCGCACCAGGTTCTGGGTTCTGACGCTGCTGGCGTAGTAGTGCGTACTGGTTCCGGTGTT CGCCGCTGGCGTGTAGGTGATCACGTTGTAGTTAGCGCCGCATACGTAGACGAACAGGACCCGGGTACGCACGCAGATGGTATGATCGGT GAAGACCAGCTGGCATGGGGCTACGAAACTAACTTCGGTGGCATGGCTGACTATGCGGTTGCTCGCGCAAGCCAGCTGATCCCGAAACCG CCGCACCTGACCTGGGAAGAAGCTGCATCTAACACTGCATGCGCATCTACCGCGTACCGTATGCTAGTAGGTGAGCGCGGTGCTCGTATGA AACAGGGTGACGTAGTTCTGGTATGGGGTGCAGCTGGCGGTCTCGGCTCCTACGGTGTTCAGCTGGTTCGTAACGGTGGTGGCGTACCGG TGGCAGTAGTGTCTTCCGCTCGCAAAGCTGAAGCAGTTCGTCGTCTGGGTTGTGATATTGTAGTTGACCGCGCTGAGATCGGTCTGACTGA CGATCCGACCGACGACCCGGATGAGGTAATTCGTATCGGTAAACGCCTGGGTGCAATCATCCGTGAACGTACCGGTCGCGACCCGGACAT CGTATTCGAGCACATCGGTCGTGCTACCTTCGGCGTATCTGTTTTCGTTGTACGTCGCGGTGGTGTAGTAGTTACTTGTGGTAGCTCTACCG GTTACCAGCATGTATTCGATAACCGTTATCTGTGGATGAAACTAAAACGTGTAGTAGGCTCTCACGTTGCAAACCTGCAGGAAGCGTGGGA CTGCAACCGCCTGTTTGAATTGGGTGCAATCGTACCGACTGTTAGTGCAGTATTCCCGATGGACGAGGTGGGCGAGGCTGTTCGCATCGTG CAGAACAACCGCCACATTGGCAAAATCGCGGTACTGTGTCAGGCTGACCGCCCTGGTCTGGGTGTTACCGACCCGGAACTGCGTGCTCGT GTAGGTGGCGACGATCGCCTGAACCCGCTGCGTGGGATGACTGCGGTAGGCGAAGGTGAATAA
	Protein	MGHHHHHHHHHHSSGHIEGRHMAVGSLTDAPADAVSAEENVGADLPAEFRAVHICAEDEAMFSDLSDKDVRKSLRVGPVAMPELAPDEVLIA VMASSVNYNTVWSAQFAPVSPFRFLEKFGRRGGWAARHDLPHQVLGSDAAGVVVRTGSGVRRWRVGDHVVVSAAYVDEQDPGTHADGMI GEDQLAWGYETNFGGMADYAVARASQLIPKPPHLTWEEAASNTACASTAYRMLVGERGARMKQGDVVLVWGAAGGLGSYGVQLVRNGGG VPVAVVSSARKAEAVRRLGCDIVVDRAEIGLTDDPTDDPDEVIRIGKRLGAIIRERTGRDPDIVFEHIGRATFGVSVFVVRRGGVVVTCGSSTGYQH VFDNRYLWMKLKRVVGSHVANLQEAWDCNRLFELGAIVPTVSAVFPMDEVGEAVRIVQNNRHIGKIAVLCQADRPGLGVTDPELRARVGGDD RLNPLRGMTAVGEGE
Acx4	DNA	ATGGGCCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGGCAGTGTTATCGTCAGCGGATCGTGC GTCTAACGAAAAGAAAGTGAAATCCAGCTATTTCGATCTTCCGCCTATGGAGATGTCTGTGGCCTTTCCACAAGCAACACCAGCCAGCACTT TTCCACCGTGTACGAGCGATTACTATCACTTCAACGATCTGCTCACCCCGGAAGAACAGGCGATCCGCAAGAAAGTGCGCGAATGCATGGA AAAGGAAGTAGCTCCGATCATGACAGAGTACTGGGAGAAAGCCGAATTTCCGTTTCACATTACCCCGAAACTGGGTGCAATGGGAGTGGC TGGTGGGAGCATCAAAGGCTATGGATGCCCCGGCTTGTCGATTACCGCGAATGCTATTGCCACCGCAGAGATTGCACGCGTTGATGCCAG TTGCTCAACGTTCATTCTGGTCCATTCAAGCCTGGGTATGCTGACCATCGCTTTGTGTGGCTCTGAAGCGCAGAAAGAGAAATACTTACCGT CCTTGGCGCAACTGAACACTGTAGCGTGTTGGGCCCTGACGGAACCGGATAACGGCTCTGACGCGTCGGGTCTGGGTACCACAGCAACCA AAGTTGAAGGCGGTTGGAAAATCAATGGTCAGAAACGCTGGATTGGCAATAGCACGTTTGCGGATCTGCTCATTATCTTTGCCCGTAATAC GACGACTAACCAGATTAACGGCTTTATCGTGAAGAAGGATGCTCCAGGCCTGAAAGCGACCAAGATTCCGAACAAAATCGGTCTGCGCAT GGTTCAGAATGGCGACATTCTTCTGCAGAACGTGTTCGTTCCGGATGAGGATCGCTTGCCCGGCGTCAACTCCTTCCAAGACACCAGTAAA GTGTTAGCCGTTAGCCGTGTCATGGTTGCCTGGCAACCTATCGGAATTTCGATGGGCATCTACGACATGTGTCATCGCTATCTGAAAGAAC GGAAACAGTTTGGTGCTCCTTTAGCGGCCTTCCAGCTGAATCAACAGAAACTTGTGCAGATGCTGGGGAATGTACAAGCGATGTTTCTTAT GGGCTGGCGTCTGTGCAAACTCTATGAAACGGGCCAGATGACTCCCGGACAGGCTTCGCTGGGGAAGGCATGGATTAGTTCCAAAGCCCG TGAAACCGCGTCATTAGGTCGTGAACTGCTGGGCGGCAATGGGATTCTGGCAGACTTTCTGGTCGCCAAAGCGTTCTGCGATTTGGAACC GATCTATACCTATGAGGGGACTTACGACATTAACACCCTCGTAACCGGTCGCGAAGTCACGGGTATTGCGAGTTTCAAACCGGCAACACGC AGCCGGCTGTAA

	Protein	MGHHHHHHHHHHSSGHIEGRHMAVLSSADRASNEKKVKSSYFDLPPMEMSVAFPQATPASTFPPCTSDYYHFNDLLTPEEQAIRKKVRECMEK EVAPIMTEYWEKAEFPFHITPKLGAMGVAGGSIKGYGCPGLSITANAIATAEIARVDASCSTFILVHSSLGMLTIALCGSEAQKEKYLPSLAQLNTVA CWALTEPDNGSDASGLGTTATKVEGGWKINGQKRWIGNSTFADLLIIFARNTTTNQINGFIVKKDAPGLKATKIPNKIGLRMVQNGDILLQNVFV PDEDRLPGVNSFQDTSKVLAVSRVMVAWQPIGISMGIYDMCHRYLKERKQFGAPLAAFQLNQQKLVQMLGNVQAMFLMGWRLCKLYETGQ MTPGQASLGKAWISSKARETASLGRELLGGNGILADFLVAKAFCDLEPIYTYEGTYDINTLVTGREVTGIASFKPATRSRL

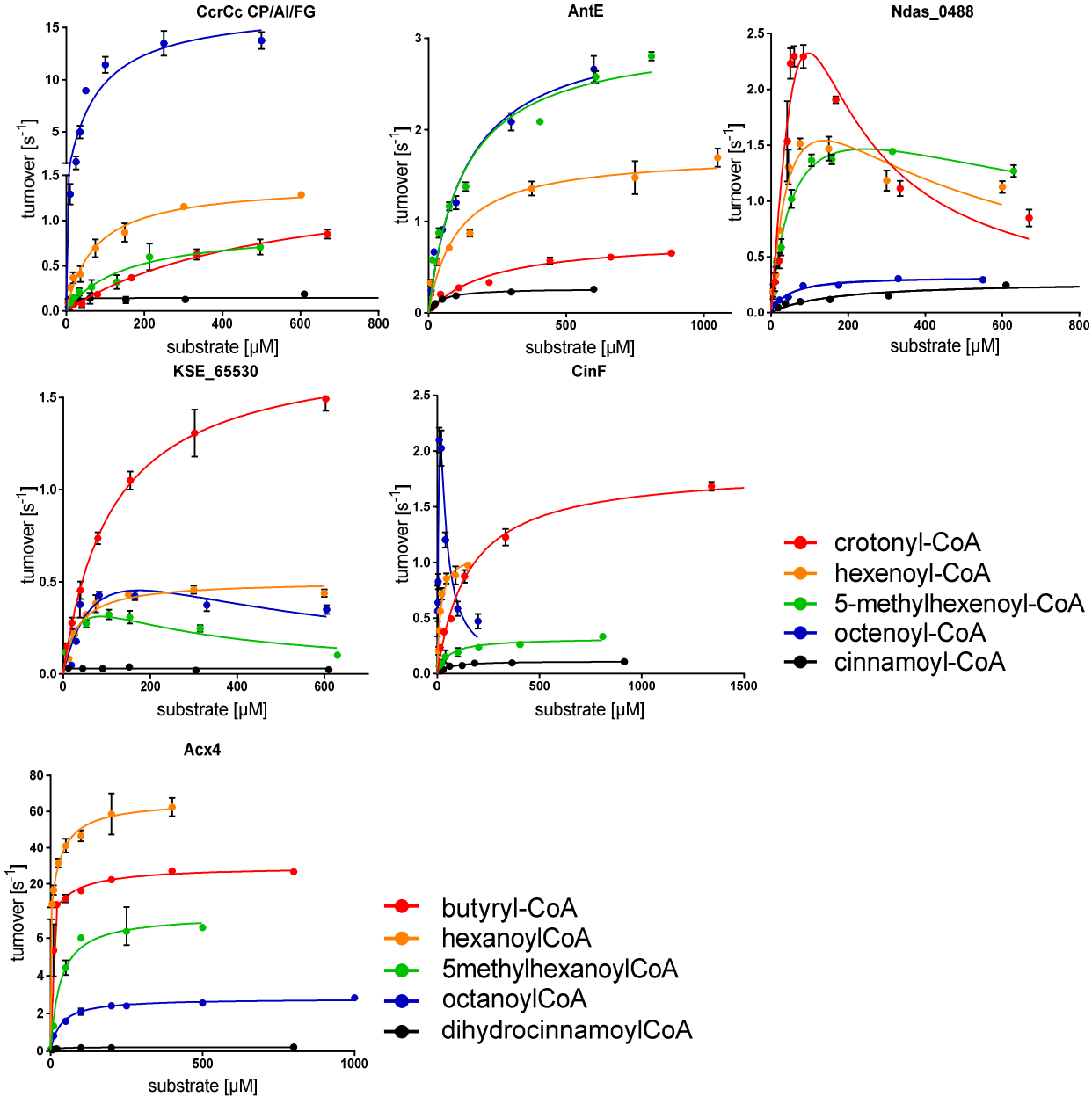


Figure S1 (related to Figure 1 and Table S2) Michaelis-Menten plots of characterized ECR variants as well as Acx4.

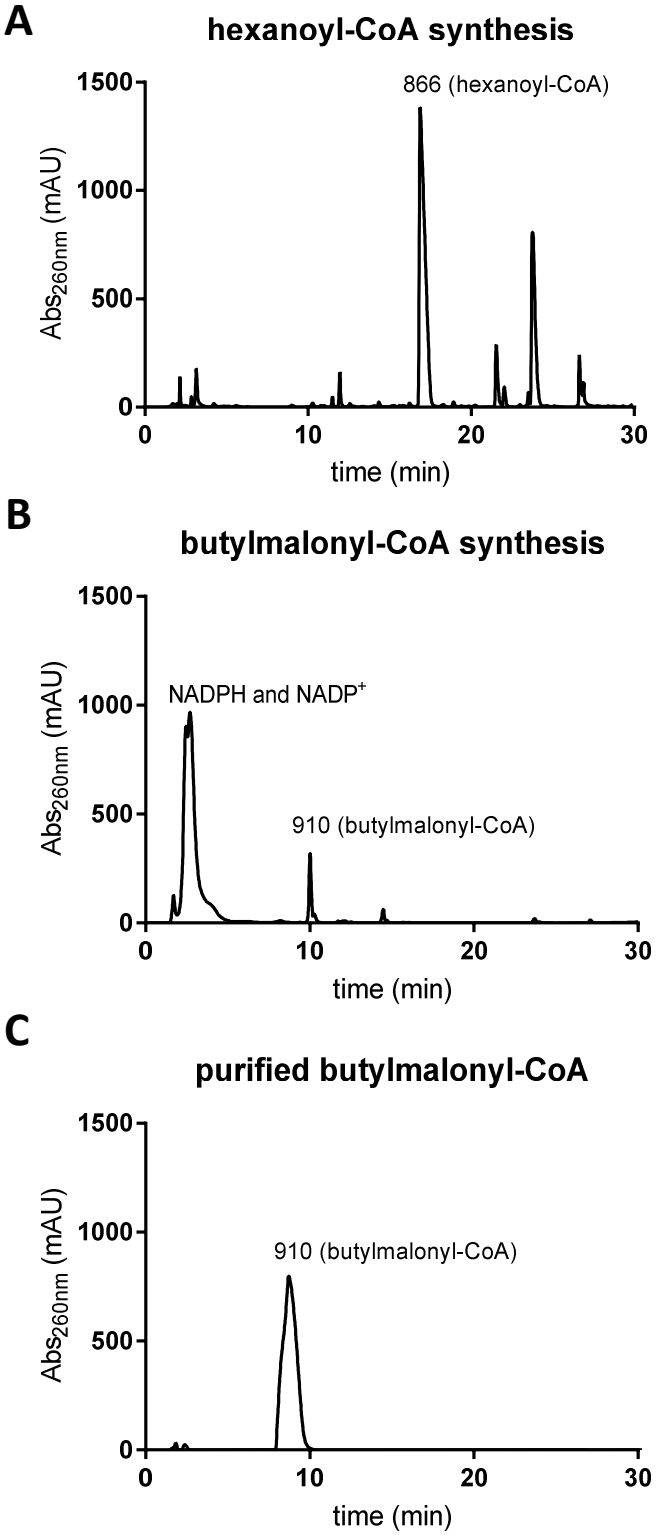


Figure S2 (related to Figure 3) HPLC traces of every step in the synthesis of butylmalonyl-CoA. A) HPLC trace after chemical coupling. Amount of hexanoyl-CoA was assessed by comparing the absorption of hexanoyl-CoA at 260nm to all other absorbing species in the mixture at this wavelength. **B)** HPLC trace after oxidation and carboxylation of hexanoyl-CoA. Purity was assessed by the absorption of butylmalonyl-CoA compaired to all other absorbing species except of the NADPH/NADP+ peak at 260 nm. **C)** HPLC trace of the purified butylmalonyl-CoA.

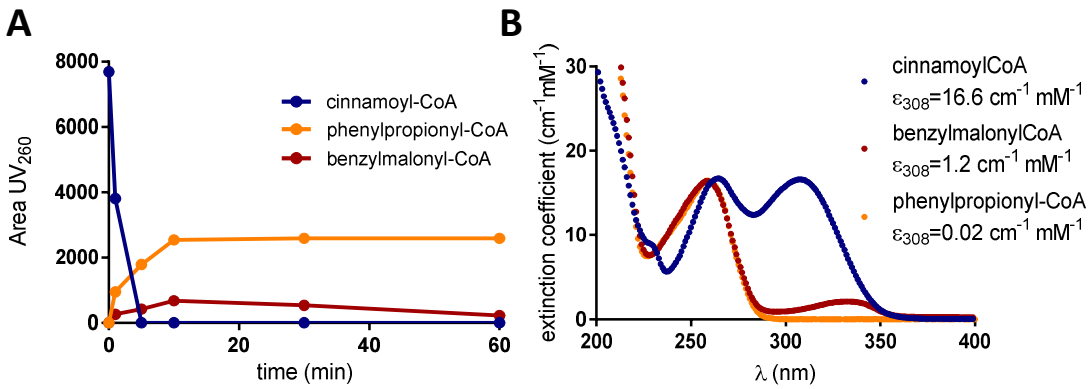


Figure S3 (related to Figure 3) Progress of up-scaled benzylmalonyl-CoA synthesis from cinnamoyl-CoA. A) The assay contained 100 mM TrisHCl pH 8, 100 mM KHCO3, 1 mM cinnamoyl-CoA, 20 mM NADPH, 40 µgmL^{-1} carbonic anhydrase, 20 Acx4 and 2.1 µM $CcrC_{PAG}$. **B)** Extinction coefficients of cinnamoyl-CoA, dihydrocinnamoyl-CoA and benzylmalonyl-CoA.

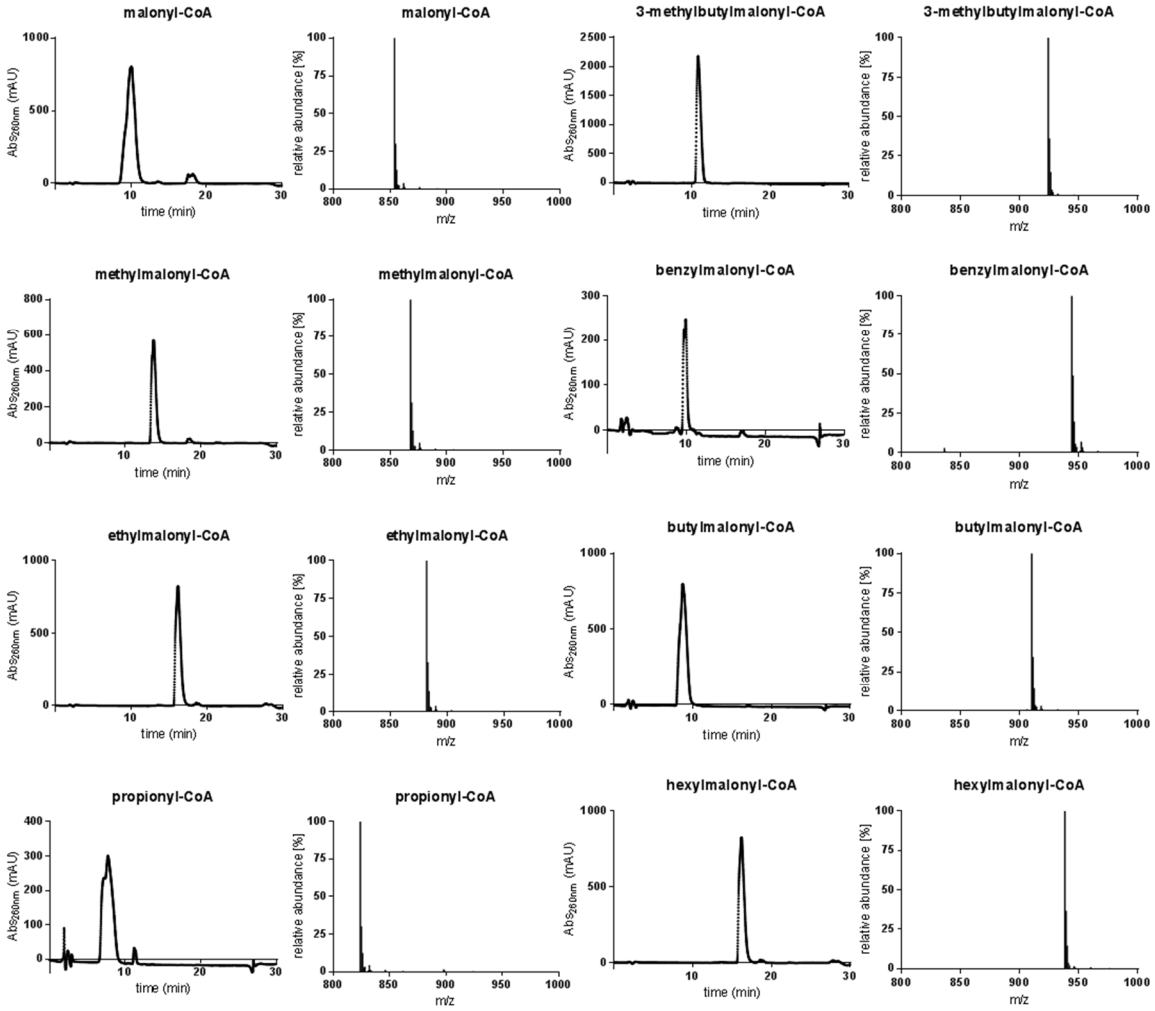


Figure S4 (related to Figure 3) Purity of all synthesized extender units assessed by HPLC-MS (left panel; HPLC trace at 260 nm, right panel; m/z of the elution peak)

CHAPTER VI

Trifunctional propionyl-CoA synthase: A self-organizing, multi-catalytic 'pico'-compartment that sequesters a reactive intermediate

Authors:
Iria Bernhardsgrütter*, Bastian Vögeli*, Tristan Wagner, Dominik M. Peter, Niña Socorro Cortina, Gert Bange, Sylvain Engilberge, Eric Girard, François Riobé, Olivier Maury, Seigo Shima, Jan Zarzycki, Tobias J. Erb

* These authors contributed equally to this work

Submitted to:
Science

Author contributions:
I.B, B.V, D.M.P, J.Z, and T.J.E conceived the project. I.B, B.V, T.W., J.Z, and T.J.E designed and performed experiments and analyzed the data. E.G, F.R, and O.M designed and prepared the phasing compound Tb-Xo4. I.B, B.V, S.E, E.G, T.W and J.Z collected X-ray data sets, T.W and J.Z solved crystal structures. E.G, G.B, and S.S oversaw crystallography and SAXS experiments, provided equipment and beam-time. N.S.C. performed mass spectrometry and analyzed the data. I.B, B.V, T.W, J.Z, and T.J.E wrote the manuscript with contributions from all other authors.

6. Trifunctional propionyl-CoA synthase: A self-organizing, multi-catalytic 'pico'-compartment that sequesters a reactive intermediate

6.1. Abstract

Cells need to cope with toxic or reactive intermediates formed during metabolism. One strategy is to sequester reactions that produce such intermediates within specialized compartments. Here we show that propionyl-CoA synthase (PCS), a dimeric ~400 kDa three-domain fusion protein and the key enzyme of the 3-hydroxypropionate bi-cycle for CO_2-fixation, forms a self-assembling, multi-catalytic compartment to sequester the reactive intermediate acrylyl-CoA. Structural analysis revealed protomer mimicking as core building principle of PCS. Kinetic analysis suggests that compartment access and catalysis are synchronized by sophisticated interdomain communication. The reaction chamber of PCS features three active sites and has a volume of only 33 nm^3. This is three orders of magnitude smaller than average bacterial microcompartments and more than two orders of magnitude smaller than other proteinaceous nanocompartments. Thus, PCS represents the smallest multi-reaction chamber described in biology to date that could inspire the engineering of a new class of dynamically regulated nanoreactors.

6.2. Introduction

Biological systems face the challenging task to efficiently catalyze hundreds to thousands of different chemical reactions in one 'pot', the cytoplasm. Diffusion is relatively fast compared to biochemical reactions, which leads to uniform concentrations of metabolites in the cytoplasm, in particular when considering the size of microbial cells [1]. This can result in 'cross-talk' between metabolic pathways and cause cross-inhibition, inactivation or even irreparable damage to metabolism, especially when the respective intermediates are instable or reactive [2]. Nature has evolved several strategies to ensure that problematic pathway intermediates are not released into the cytoplasm, but directly transferred to the next enzyme or active site. These strategies include encapsulation of intermediates in membrane or protein delimited organelles (compartmentalization), covalent linking of intermediates to multi-domain enzyme complexes or carrier proteins, electrostatic guidance of intermediates from one active site to the next, or formation of intramolecular tunnels between two active sites [3].

Here, we report on the enzyme reaction cascade from 3-hydroxypropionate to propionyl-CoA, which is the key sequence in the 3-hydroxypropionate bi-cycle for autotrophic CO_2 assimilation [4-5]. The overall sequence comprises three enzymatic steps, during which a highly reactive, toxic and unstable intermediate, acrylyl-CoA, is formed. In autotrophic *Sulfolobales* the three reactions are catalyzed by individual enzymes [6-7]. However, in several phyla (*e.g.* Proteobacteria and Chloroflexi, **Figure S1**) the three reactions are catalyzed by a fusion enzyme of about 1850 amino acids, that comprises three catalytic domains, the propionyl-CoA synthase (PCS, **Figure 1a, Figure S2**) [4]. We wondered if PCS specifically evolved as a fusion enzyme to overcome free diffusion of reactive acrylyl-CoA. Therefore, we overproduced PCS from *Erythrobacter* sp. NAP1 (GenBank accession no. EAQ29651) and studied how this multi-domain enzyme copes with its reactive intermediate, acrylyl-CoA.

6.3. Results

6.3.1. HPLC analysis of PCS reaction indicates that acrylyl-CoA is sequestered

When we followed the overall reaction of PCS, we could not detect acrylyl-CoA in the assay mixture, while 3-hydroxypropionyl-CoA and propionyl-CoA accumulated (**Figure 1b**). Only when we increased the concentration of the enzyme, acrylyl-CoA was detectable, albeit at very low levels (**Figure 1c**). The concentration of acrylyl-CoA corresponded to 1.8 ± 0.1 % of PCS monomers and stayed constant during steady state. This demonstrated that acrylyl-CoA was formed *in situ* by the dehydratase domain, but presumably stayed sequestered within PCS being quickly consumed by the reductase domain. Notably, when the reduction reaction was prevented (by omitting NADPH), acrylyl-CoA did not accumulate, while 3-hydroxypropionyl-CoA still did (**Figure 1d**). This indicated that even when interrupting the catalytic sequence, acrylyl-CoA remains sequestered within PCS.

6.3.2. The overall structure of PCS

To understand the structural basis of acrylyl-CoA sequestration, we solved the crystal structure of PCS in the presence of CoA, $NADP^+$ and an ATP analog (phosphomethylphosphonic acid adenylate ester) using a new phasing compound [8] (PDB 6EQO, **Table S1** and **Figure S3**). PCS forms a dimer (of ~400 kDa) around a central core of reductase domains (**Figure 1a**). The ligase and dehydratase domains extend to both sides, enclosing spherical compartments ('reaction chambers') that each feature three internal active sites (**Figure 2**). The active sites are interconnected through positively charged surfaces, suggesting electrostatic guidance of the CoA intermediates during catalysis. Escape of intermediates from the reaction chamber is prevented by negative charges surrounding any small openings. The inner diameter of the reaction chamber is between 3.5 – 5.5 nm with a volume of 33 nm^3. This volume is between three and six orders of magnitude smaller than that of bacterial microcompartments [9], for example carboxysomes (12×10^3 to $\sim 10 \times 10^7$ nm^3, calculated from inner diameters ranging from 30 to 600 nm) [10], two orders of magnitude smaller than of described nanocompartments, such as encapsulins (5×10^3 nm^3, calculated from inner diameter of 22 nm) [11], and even half of that of proteasomes (59 and 84 nm^3) [12]. Thus, PCS forms the smallest multi-catalytic reaction chamber observed in Nature, a 'pico'-compartment.

6.3.3. Molecular mimicry and organization of the reaction chamber

The active sites of the individual domains of PCS align well with the ones of lone-standing homologues (acyl-CoA ligase, dehydratase and reductase **Figure S4**, **S5** and **S6**). These lone-standing homologues are organized as homo-oligomers that require contributions from neighboring protomers [13-15]. In the reaction chamber of PCS that is organized as a monomer, there are no neighboring protomeric subunits that could contribute to active site formation. Instead, the domains of PCS carry extra extensions to mimic the contributions from the other subunit(s) present in the lone-standing homologues (**Figure S7**, **S8** and **S9**). As an example, the dehydratase domain of PCS features additional helices (**Figure S8**) containing two highly conserved residues (F1220 and K1223 14) that are involved in the stabilization of the CoA-ester. In case of lone-standing dehydratase homologs, which form homotrimers, these residues usually protrude into the active site of a neighboring subunit of the homo-trimeric complex. Another example is the reductase domain of PCS that carries a structural extension that mimics the part of the CoA binding site that is provided by a neighboring subunit in lone-standing enoyl-CoA reductases/carboxylase homologues, which are organized as homotetramers (**Figure S9**). Furthermore, the ligase domain of PCS carries structural extensions that are absent in any lone-standing acyl-CoA synthetase homologues (**Figure S7**), most prominently, an additional four helix bundle. This four-helix bundle appears to be unique to PCS and is exclusively found in PCS homologues (based on BLASTP analysis). This four helix bundle caps one side of the reaction chamber. Its absence would leave the enzyme compartment wide open (**Figure S12**).

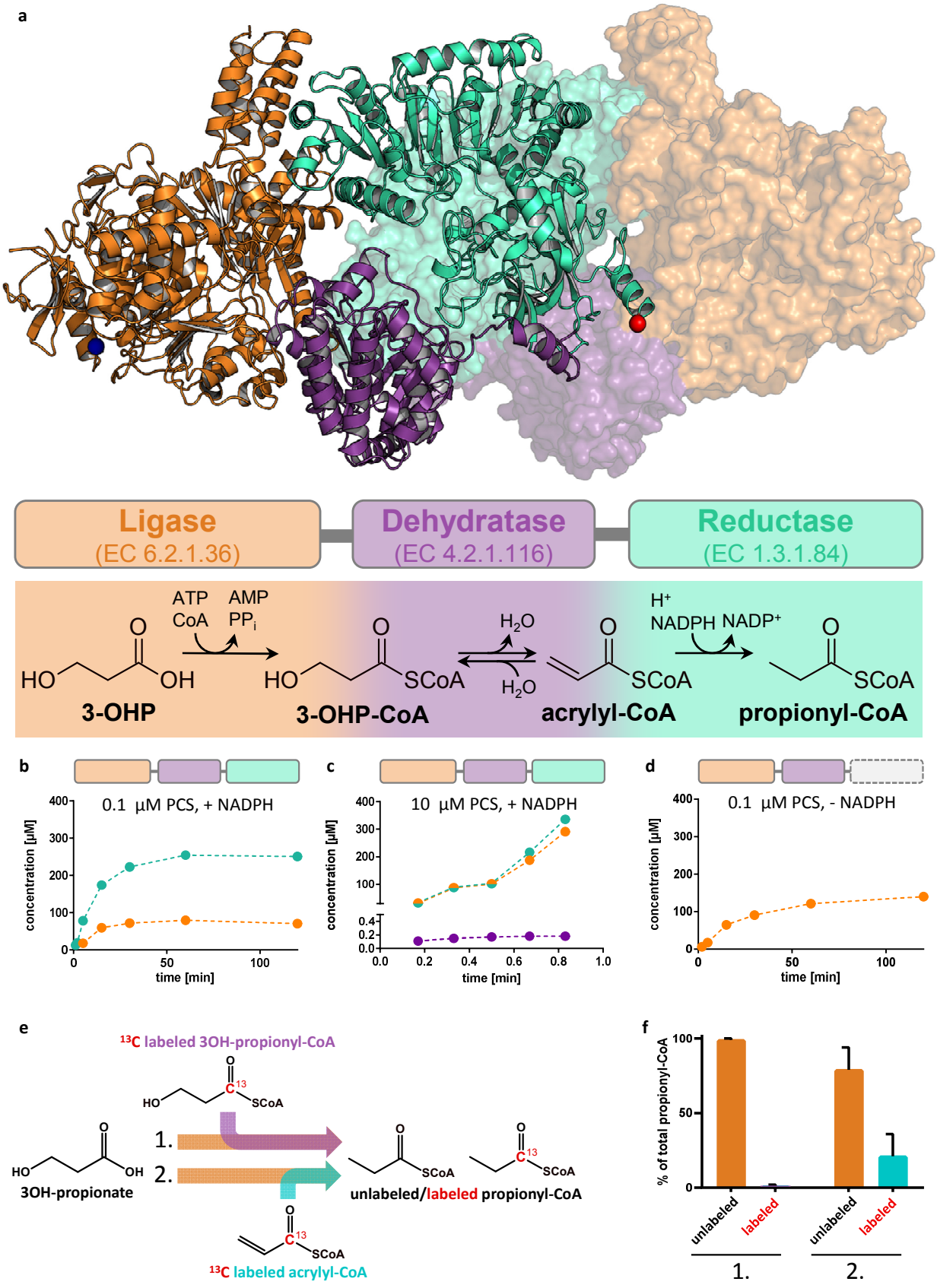


Figure 1. Trifunctional PCS: Structure and catalytic behavior. a, Dimeric structure of PCS from Erythrobacter sp. NAP1 (PDB 6EQO). One protomer is depicted in cartoon and one in surface representation. The multi-domain organization is highlighted by different colors: orange, ligase domain; purple, dehydratase domain; cyan, reductase domain; blue sphere, N-terminus; red sphere, C-

terminus. Schematic arrangement of the three domains and their individual reactions are shown using the same color code. **b,** Time course of the overall reaction with 0.1 µM PCS, 800 µM CoA, 500 µM 3-hydroxypropionate, 800 µM ATP and 300 µM NADPH. Production of the 3-hydroxypropionyl-CoA intermediate (orange) and the final product propionyl-CoA (cyan) was observed. In contrast no free acrylyl-CoA was detectable. **c,** Time course of the reaction containing 10 µM PCS, 5 mM CoA, 5 mM 3-hydroxypropionate, 5 mM ATP and 5 mM NADPH. At these high enzyme concentrations acrylyl-CoA (purple) was detected at 0.18 µM during steady-state corresponding to 1.8% occupancy of reductase active sites. 3-Hydroxypropionyl-CoA und propionyl-CoA accumulate over time. **d,** as in ab, but without NADPH. Again, formation of 3-hydroxypropionyl-CoA was observed, but not of free acrylyl-CoA. **e,** Isotopic labeling competition experiment containing unlabeled 3-hydroxypropionate and either 13C-labeled 3-hydroxypropionyl-CoA (experiment 1) or acrylyl-CoA (experiment 2). The reaction was started by the addition of PCS. Products were analyzed by LC-MS (see Table S4 for detailed assay conditions). f, Results of the isotopic labeling competition experiment. Only 0.8 ± 0.4 % of propionyl-CoA was produced from exogenous 13C-labeled 3-hydroxypropionyl-CoA during steady state (experiment 1). Approximately every fifth propionyl-CoA (21 ± 15%) was formed from exogenous 13C-labeled acrylyl-CoA during steady-state (experiment 2). b – d, data of a representative single experiment. **f,** data mean ± s.d. (n=3).

6.3.4. Biochemical characterization of PCS

How is the sequence of reactions orchestrated within the compartment? The three enzyme reactions of PCS can be measured individually, when the appropriate substrates and cofactors are provided, demonstrating that all active sites are in principle accessible [4]. We determined the kinetic parameters for the overall reaction of PCS as well as for each catalytic domain. While the ligase and the dehydratase domain had apparent turnover frequencies (k_{cat}) comparable to the overall reaction of PCS, the k_{cat} of the reductase domain was almost 30-fold higher (**Figure S10**, **Table S2** and **S3**). This suggests that acrylyl-CoA is immediately consumed upon its formation *in situ*. To study whether externally provided intermediates can access the reaction chamber of PCS during steady-state, we performed an isotopic labeling competition experiment [16-17]. When starting from 3-hydroxypropionate, PCS preferentially catalyzed the overall reaction. Externally added 3-hydroxypropionyl-CoA or acrylyl-CoA were used only to a small extent by PCS under these conditions (**Figure 1e**, **1f** and **Table S4**). In other words: despite the high catalytic efficiency of the reductase domain, almost no externally added acrylyl-CoA was reduced by PCS in the steady-state. Note that the formed ^{13}C labelled propionyl-CoA (21 ± 15 %) corresponds with the amount that the reductase is able toform by incorporation of external acrylyl-CoA within the first 0.6 s and thus before steady-state of the overall reaction is reached. These results demonstrated that catalysis in PCS is consecutive, and that internally produced 3-hydroxypropionyl-CoA and acrylyl-CoA are channeled within the enzyme.

6.3.5. Interdomain communication in PCS

Apparently, external acrylyl-CoA is prevented from entering the reaction chamber of PCS indicating that the enzyme assumes a 'closed state' during catalysis. Therefore, we wondered, if any of the substrates (CoA, ATP, 3-hydroxypropionate) or products (AMP) would restrict access of exogenous acrylyl-CoA to the reductase domain during steady-state. Indeed, CoA had a strong effect on both the K_M and the k_{cat} of acrylyl-CoA consumption by the reductase domain. To measure the influence of CoA on the reduction reaction, we used the E1027Q variant that is unable to hydrate acrylyl-CoA and added different concentrations of CoA. At low concentrations (0.3 mM), CoA acted as competitive inhibitor of the reduction reaction (**Figure S11**), suggesting that free CoA directly competes with acrylyl-CoA at the active site of the reductase domain. At higher concentrations (3.0 mM), free CoA additionally acted as non-competitive inhibitor, indicating that a second CoA binding site outside of the reductase domain influenced the reduction reaction. We speculated that this binding site was located in the ligase domain. We substituted a conserved lysine (K783) in the ligase domain that directly interacts with CoA by methionine in the E1027Q variant (**Figure S12**). The interaction between K783 and CoA is suggested to stabilize the closed conformation of PCS. The K783M mutation rendered the ligase as well as the reductase domains still functional. Notably the non-competitive inhibition of CoA on the reductase domain was completely removed in this variant (**Figure S11**). Thus, we concluded that the ligase domain and the reductase domain act not independently from each other and that the K783 plays an important role in communication between the two domains. The hypothesis that the domains in

PCS act in a synchronized fashion was additionally supported by the observation that NADPH (the co-substrate of the third reaction) in turn had a strong effect on the kinetic parameters of the first reaction. The ligase reaction alone had an apparent Michaelis-Menten constant for CoA (KM,CoA) of 6 ± 3 mM. However, when we added NADPH to the reaction mixture, KM,CoA decreased by a factor of 30 to 0.22 ± 0.05 mM (**Figure S11**, **Table S2**). To test whether the effect of NADPH onto the ligase reaction was independent from an active reductase domain, we repeated the same assay with a dehydratase mutant of PCS (E1027Q). In this enzyme variant, the reductase domain is still functional, but is not provided with substrate because of the inactive dehydratase domain. In the E1027Q variant KM,CoA dropped even below that of wt PCS to 0.019 ± 0.003 mM (Figure S11, Table S2). These results clearly confirmed that the individual domains of PCS are coupled with each other and that interdomain communication in PCS works in both directions (**Figure S10**, **Table S2**).

6.3.6. Opening and closing of the reaction chamber

The kinetic characterization demonstrates a functional coupling of the last reaction step in PCS to the first one and *vice versa*. Apparently, PCS undergoes synchronized conformational changes during catalysis allowing substrates and products to enter and leave the reaction chamber. The gatekeeper to the reaction chamber of PCS is presumably the ligase domain. Stand-alone CoA ligases undergo significant conformational changes between 'open' and 'closed' states during catalysis [13, 18]. When we superposed a 'closed-state' *Salmonella enterica* ligase (PDB 2P2F) [13] with the PCS ligase domain, the structures aligned almost perfectly (rmsd of 0.932 Å over 441 Cα-atoms) with the exception of a four-helix bundle extension only present in PCS (**Figure S7**). The N-terminus of an 'open-state' ligase of *Saccharomyces cerevisiae* (PDB 1RY2) [18] still aligned well to the PCS ligase domain (rmsd of 0.955 Å over 374 Cα-atoms), whereas the C-terminal domain (~130 residues) was rotated outwards. Modelling a corresponding conformational change onto PCS results in the exposure of a 'hole' that provides access to the interior of an 'open state' compartment (**Figure S12**). Upon binding of CoA and formation of 3-hydroxypropionyl-CoA by the ligase domain, the reaction chamber becomes shut, keeping the intermediates enclosed.

We performed limited proteolysis on PCS with trypsin in the absence and presence of different substrates and products to directly test for conformational changes (**Figure S13**). When incubated in the presence of CoA, which restricts access to the active site of the reductase (see above), PCS was fragmented within 90 min. In contrast, the simultaneous addition of 3-hydroxypropionate, $NADP^+$ and ATP protected PCS from total proteolysis (**Figure S14**). This confirmed that the enzyme assumed different conformations depending on the presence of different substrates or products, orchestrating and synchronizing catalysis. SAXS analyses of PCS in the presence of different substrates support this conclusion (**Table S5**).

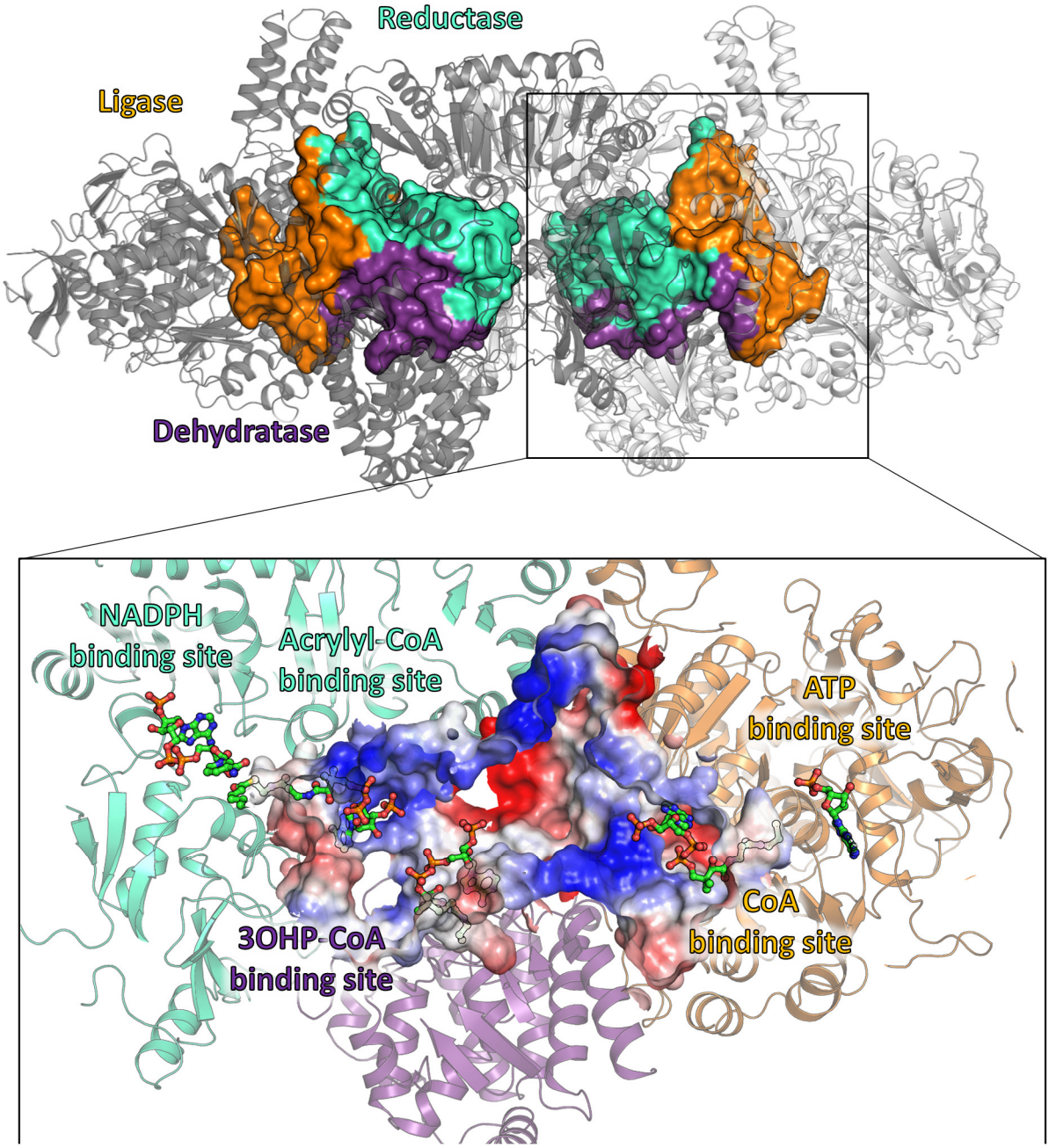


Figure 2. Multi-catalytic reaction chamber of PCS. Volume filling representation[19] of the reaction chambers enclosed by PCS. The central catalytic reaction chamber of each protomer is formed through the contribution of all three domains. Orange, contribution of the ligase domain; purple, contribution of the dehydratase domain; cyan, contribution of the reductase domain. The close up shows a cross section through the reaction chamber. Electrostatic charge distribution is shown as a gradient from red – negatively charged to blue – positively charged. The three active sites are well connected within the reaction chamber. Large positively charged patches may guide the CoA-ester intermediates or help retaining them inside during catalysis. Negative charges around the small openings may also prevent leakage of the negatively charged CoA-derivatives. The PCS structure co-crystalized with CoA (no density), an ATP analog and $NADP^+$ is depicted in cartoon showing the ligase domain in orange, the dehydratase domain in purple and the reductase domain in cyan. CoA binding sites have been modelled based on the superposition of the structures of lone-standing CoA ligase (PDB 2P2F) [13], dehydratase (PDB 5JBX) [14] and reductase (PDB 4A0S) [15] onto PCS (compare Figure S4, S5 and S6). Distances between the active sites have been determined by measuring the distance between sulfur atoms of modelled CoA moieties to be: ligase - dehydratase 42.5 Å, dehydratase – reductase 33.7 Å, ligase – reductase 63.5 Å.

6.3.7. The catalytic cycle of PCS

Taken all our results together we propose the following catalytic mechanism (**Figure 3**). First the ligase forms the 3-hydroxypropionyl-AMP from ATP and 3-hydroxypropionate in the open conformation. Binding of CoA then closes the reaction chamber. NADPH facilitates the binding of CoA, which is consistent with the lowered $K_{M,CoA}$ observed upon NADPH addition and would ensure that the enzyme contains all necessary cofactors to catalyze the complete reaction sequence upon closing the chamber. The formed 3-hydroxypropionyl-CoA is subsequently dehydrated to acrylyl-CoA within the closed reaction chamber, isolated from the external environment. The final reduction reaction would trigger the re-opening of the reaction chamber for product release and prepare the enzyme for the next catalytic cycle. This mechanism fits well with the observation that the k_{cat} of the ligase reaction alone drops significantly in presence of NADPH, indicating that NADPH stabilizes the closed conformation and limits the catalytic rate of the ligase domain. The conformational changes are supported by limited proteolysis experiments, which indicate that the enzyme features at least two different states depending on the substrates available (**Figure S13** and **S14**). Notably, the peptides that appear more prominent in the sample of closed PCS are corresponding to the flexible cap in the ligase domain that is exposed in the closed conformation but hidden within the PCS structure in the open conformation. In summary the enzyme displays a synchronized reaction mechanism to sequester and channel the toxic acrylyl-CoA between active sites.

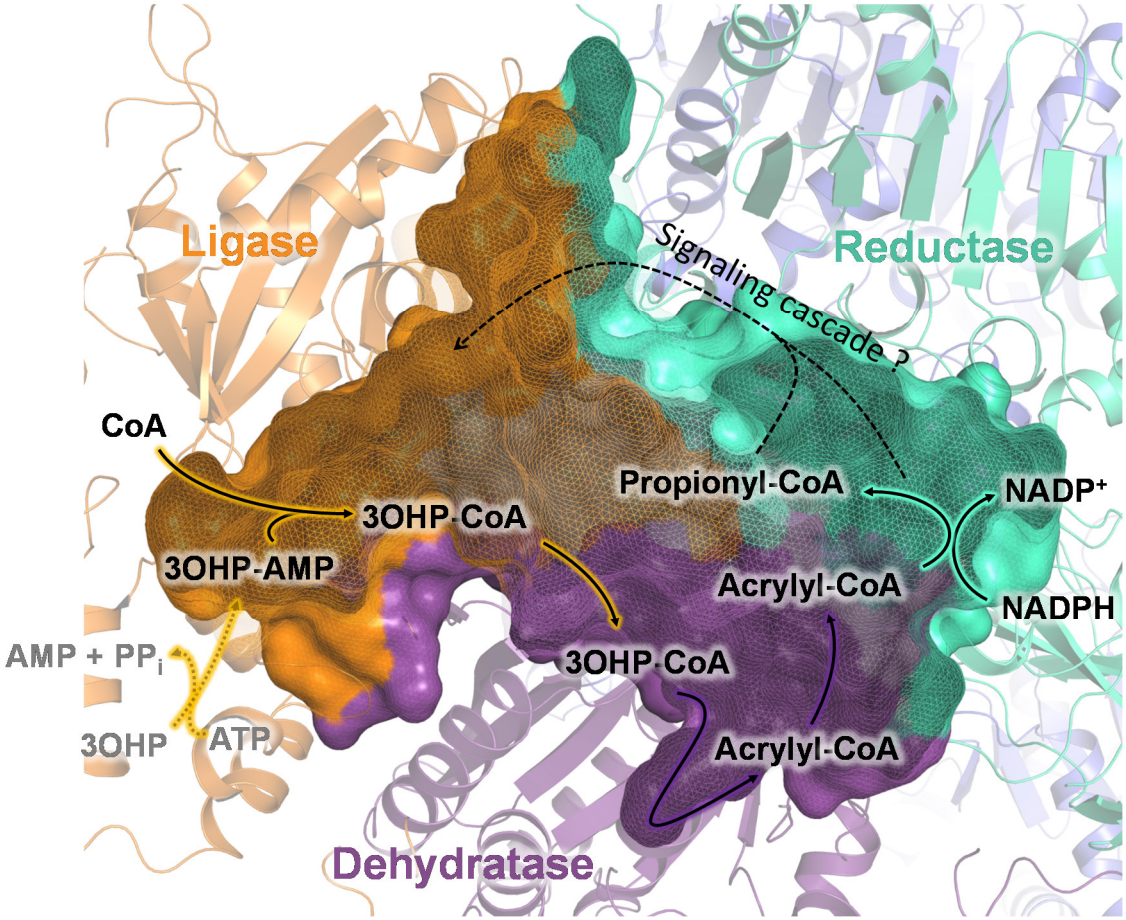


Figure 3. Proposed catalytic cycle of PCS. In the open conformation 3-hydroxypropionate (3OHP) and ATP are converted to 3-hydroxypropionyl-AMP (3OHP-AMP) through the ligase domain (orange). The binding of CoA induces closing of the enzyme and the formation of 3-hydroxypropionyl-CoA (3OHP-CoA). 3OHP-CoA is released into the reaction chamber, where it is converted by the dehydratase domain (purple) to acrylyl-CoA. Acrylyl-CoA then enters the active site of the reductase domain (cyan). The reduction of acrylyl-CoA to propionyl-CoA causes the re-opening of the reaction chamber. Propionyl-CoA is released, which leaves PCS ready for the next catalytic cycle.

6.4. Discussion

In summary, PCS, a key enzyme in the 3-hydroxypropionate bi-cycle for CO_2 fixation, is not a simple fusion protein of three individual enzymes, but represents a sophisticated structural arrangement of three domains enclosing a reaction chamber, which connects all three active sites. This novel strategy of substrate channeling combines the mechanisms of compartmentalization with the allosteric gating in enzyme channels. However, key features distinct PCS from either of the two substrate channeling mechanisms. The volume of the PCS reaction chamber is several orders of magnitude smaller than those of known bacterial micro- or nanocompartments. Usually, such compartments consist of self-assembling shell proteins that encapsulate enzymes [20-23]. Compared to these, PCS is to our knowledge the first self-organizing compartment in central carbon metabolism where just one polypeptide fulfills the structural role of forming the reaction chamber, as well as the catalytic role of driving the multi-reaction sequence. The orchestration of the three-reaction sequence in PCS is guaranteed through interdomain communication (**Figure 3**) and conformational changes, similar as described in other channeling complexes[24-26]. Conformational change of the PCS ligase domain controls the only entry and exit site to the three active sites buried within the 33 nm^3 reaction chamber. Other channeling enzymes in contrast, connect two active sites through a narrow channel where the conformational change of single residues suffice to gate the separate entry and exit sites. Channeling in other CoA-ester metabolic enzymes requires covalent attachment of the CoA-ester or a shared binding site of the CoA moiety between two active sites[27-29]. Thus, the channeling mechanism of PCS is not only a novelty in CoA-ester metabolic enzymes but also generally represents an unprecedented principle for directing intermediates along a defined multi-reaction sequence.

PCS is able to catalyze three very fundamental chemical reactions in CoA-ester biochemistry and retain intermediates within its reaction chamber. As such, the enzyme employs an intriguingly 'simple' principle to catalyze a consecutive reaction sequence within a controlled environment. This notably comes at minimal biosynthetic cost, because no additional proteins are required. The natural example of a minimal self-assembling nanoreactor that is dynamically regulated could serve as a model for the engineering of spatially and temporally controlled reaction sequences [30-35], especially such that proceed via toxic, reactive or unstable intermediates.

6.5. Materials and Methods

6.5.1. Chemicals

Chemicals were obtained from Sigma-Aldrich (Munich, Germany) and CARL ROTH GmbH (Karlsruhe, Germany). 3-hydroxypropionate was bought from TCI Deutschland GmbH (Eschborn, Germany). Coenzyme A was purchased from Roche Diagnostics. 1-^{13}C-propionate sodium salt was purchased from Cambridge Isotope Laboratories Inc. (Tewksbury, USA). Biochemicals and materials for cloning and expression were obtained from Thermo Fisher Scientific (St. Leon-Rot, Germany), New England Biolabs GmbH (Frankfurt am Main, Germany) and Macherey-Nagel GmbH (Düren, Germany). Carbonic anhydrase was bought from MP Biomedicals (Illkirch, France). Primers or synthesized genes were obtained from Eurofins MWG GmbH (Ebersberg, Germany) or the DOE Joint Genome Institute (California, USA), respectively. Materials and equipments for protein purification were obtained from GE Healthcare (Freiburg, Germany), Bio Rad (Munich, Germany) or Merck Millipore GmbH (Schwalbach, Germany).

6.5.2. Synthesis of 3-hydroxypropionyl-CoA, 13C-acrylyl-CoA and 13C-3-hydroxypropionyl-CoA

For the synthesis of unlabeled 3-hydroxypropionyl-CoA a previously described method using carbonyldiimidazole coupling of the precursor acid with coenzyme A was used [36]. Unlabeled acrylyl-CoA was synthesized using a previously described mixed anhydride coupling [36]. ^{13}C-acrylyl-CoA and ^{13}C-3-hydroxypropionyl-CoA were synthesized in two steps from ^{13}C-propionate. In the first step the CDI coupling method was adapted for the synthesis of ^{13}C-propionyl-CoA by a protonation step. ^{13}C-propionate (0.156 mmol, 4.8 eq.) was dissolved into 1 mL THF containing pTsOH (0.156 mmol, 4.8 eq.) for 15 min, the mixture was centrifuged and CDI (0.130 mmol, 4 eq.) added to the supernatant. The mixture was stirred at RT for 1 h. CoA (0.0325 mmol, 1 eq.) dissolved in 250 µL 0.5M $NaHCO_3$ was added and stirred for 1h. The mixture was lyophilized, HPLC purified and again lyophilized. ^{13}C-acrylyl-CoA was synthesized enzymatically using the acyl-CoA oxidase Acx4 from *Arabidopsis thaliana* [37]. A 1 mL assay contained 100 µL 1 M $KHPO_4$ 200 µL 30 mM ^{13}C-propiony-CoA and 600 µL 1mg/mL Acx4. The reaction was quenched after 1 h by adding 20 µL 50% formic acid and directly injected into the HPLC-MS for purification using a previously described purification protocol [36]. In case of ^{13}C-3-hydroxypropionyl-CoA the assay contained additionally 50 µL of the dehydratase PhaJ from *Pseudomonas aeruginosa* for direct hydration of the *in situ* generated acrlyl-CoA.

6.5.3. Bacterial strains and growth conditions

E. coli DH5α (Thermo Scientific™) strains were used for cloning and grown in LB medium [38]. For protein expression *E. coli* BL21-AI™ (Invitrogen) or Arctic-Express (DE3) RIL (Agilent Technologies) were grown in TB medium[39]. Incubation temperature was 37°C. Antibiotics for selection purposes were used accordingly: 100 µg/ml ampicillin, 15 µg/ml gentamycin.

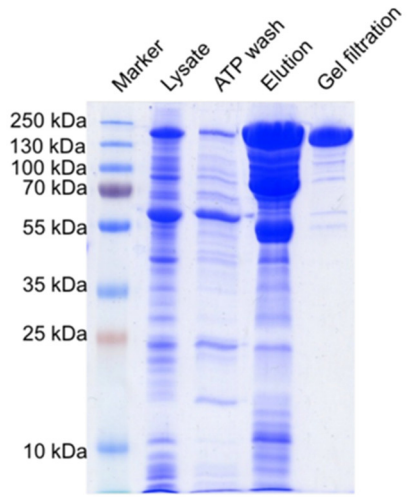
Cloning. All *in silico* cloning was performed with Clone Manager 9 (Scientific & Educational Software). For purification, preparation, cloning, transformation and amplification of DNA, standard protocols were used [40]. Plasmid isolation and PCR product purification was performed with kits from Macherey Nagel (Düren, Germany) according to the manufacturer's protocols.

The PCS gene with an N-terminal 10x His tag was synthesized by the DOE Joint Genome Institue. The construct was cloned into the expression backbone pET-16b by restriction cloning, resulting in the plasmid pTE1005. Point mutants were generated by QuickChange® Site-Directed mutagenesis (Stratagene, La Jolla, USA). Following primers were used: forward primer (5'-CGT TTC GGT CAA CCA CAA ATC AAT CTT CGC-3') and reverse primer (5'-GCG AAG ATT GAT TTG TGG TTG ACC GAA ACG-3') for the E1027Q variant; forward primer (5'-CGG AAA TTT TTG GCA CAG CGC TGT GCA ATG CTT ATG AG-3') and reverse primer (5'-CTC ATA AGC ATT GCA CAG CGC TGT GCC AAA AAT TTC CG-3') for the H1769A variant; forward primer (5'-CCT CAC AGC CAG

ATG GGT GTA ACT CC-3') and reverse primer (5'-GGA GTT ACA CCC ATC TGG CTG TGA GG-3') for the K783M variant.

6.5.4. Expression and purification of PCS

PCS was expressed from the plasmid pTE1005 using *E.coli* ArcticExpress (DE3) RIL as expression host. The cells were transformed with the expression plasmid and plated on LB agar containing selective antibiotic and grown overnight. The colonies were used to inoculate 1 L TB medium. The expression culture was incubated at 37°C while shaking at 110 rpm until an OD_{600} of 0.7 – 0.9 was reached. The *E.coli* ArcticExpress (DE3) RIL culture was cooled down to 14°C before induction. Expression was induced by adding 0.25 mM IPTG. The culture was incubated for 16-20 h. The cells were harvested by centrifugation at 5000 *x g* for 10 min. The pellet was stored at -20°C, optionally. Cells were resuspended in a 1:3 ratio (w/w) in Buffer A (50 mM Tris-HCl pH 7.9, 500 mM NaCl) containing SIGMAFAST™ protease inhibitor (Sigma-Aldrich, Munich, Germany) and lysed by ultrasonication. The lysate was cleared by ultracentrifugation at 50'000 *x g* for 45 min at 4°C followed by filtration through a 0.45 µm syringe filter. The lysate was loaded onto a 1 mL His-Trap (GE Healthcare). Unspecifically bound proteins were washed off with 15 mL of 5 % Buffer B (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 500 mM imidazole). To wash away the *E.coli* ArcticExpress (DE3) RIL Cpn60 chaperone, an additional wash step was performed with 15 mL removal buffer (50 mM Tris-HCl pH 7.5, 50 mM KCl, 20 mM $MgCl_2$, 5 mM ATP). PCS was eluted with 100% buffer B and applied to a pre-equilibrated HiLoad 16/60 200 pg superdex (GE Life Science) column (150 mM NaCl, 20 mM Tris HCl pH 7.9). The purity of the PCS was tested by SDS-PAGE.



Purification of PCS using a HisTrap and Gelfiltration column

6.5.5. Kinetic characterization of PCS

Spectrophotometric assays were set up to measure the activity of PCS. The assays were performed in 10 mm quartz cuvettes (Hellma Analytics) on a Cary-60 UV/Vis spectrometer (Agilent Technologies Inc. Santa Clara, CA, USA). The assay temperature was set to 30 °C. The parameters for the CoA ligase domain alone were measured using a coupling assay via myokinase (purified from ASKA JW1375), pyruvate kinase and lactate dehydrogenase (SigmaAldrich P02694). To probe the influence of acrylyl-CoA on the ligase reaction, the assay was repeated using the PCS E1027Q variant deficient in the enoyl-CoA hydratase reaction to avoid back reaction of acrylyl-CoA to 3-hydroxypropionyl-CoA. In this assay PCS and acrylyl-CoA were added to the reaction and incubated for 5 min at 30°C before starting the reaction with the addition of CoA. The effect of NADPH on the ligase reaction was also tested with the PCS E1027Q variant to avoid overall reaction. The reaction catalyzed by the dehydratase domain was assayed using the PCS H1769A variant that is deficient in the reductase reaction. Acrylyl-CoA formation was coupled to its reduction by a stand-alone reductase (Etr1p

from *Saccharomyces cerevisiae*). The PCS reductase reaction was measured using the E1027Q variant that is deficient in the enoyl-CoA hydratase reaction to avoid the back reaction of acrylyl-CoA to 3-hydroxypropionyl-CoA. The assay was repeated in presence of different concentrations of free CoA. PCS and CoA were added to the reaction mixture and incubated for 10 min at 30°C before starting the reaction with the addition of acrylyl-CoA.

All reactions were measured by following the consumption of NADPH or NADH (ligase coupling assay) at 340 nm ($\varepsilon_{NAD(P)H}$ = 6.22 mM^{-1} cm^{-1}) or at 365 nm (ε_{NADH} = 3.4 mM^{-1} cm^{-1}). The detailed conditions for all assays can be found in **Table S2**.

6.5.6. Crystallization of PCS

All crystallization was performed at 18°C using the sitting drop method in 96-well 2-drop MRC Crystallization Plates in polystyrene (Molecular Dimensions, Suffolk, UK). Crystallization drops (1.4 – 2 µL) contained PCS at 10 mg/ml premixed with 2mM of CoA, $NADP^+$ and phosphomethylphosphonic acid adenylate ester each mixed with reservoir solution in a 1:1 ratio. First thin needle-shaped crystals appeared after several weeks in 100 mM BisTris pH 6.5, 200 mM NaAc, 25 % (w/v) polyethylene glycol (PEG) 3350 supplemented with 3 % (w/v) trimethylamine N-oxide dihydrate as additive (condition 1). These crystals had a *C*2 symmetry and the best resolution obtained was 2.7 Å, which was used for the final structure model. Increasing the additive trimethylamine N-oxide dihydrate to 6 % still lead to crystal formation (condition 2) but exhibited strong twinning. Crystals could also be reproduced in the same condition replacing the additive with 100 mM D-(-)-fructose (condition 3) or 4 % (v/v) tert-butanol (condition 4). Crystals of the space group $P\ 2_1\ 2_1\ 2$ were obtained in the same condition supplemented with 2 % (w/v) benzamidine hydrochloride as additive (condition 5). Phasing was achieved by soaking crystals of condition 3 in 100 mM Xo4 [8] for 4 minutes. All crystals were cryo-protected with the respective crystallization solution supplemented with 20 – 30 % ethylene glycol.

6.5.7. X-ray crystallography analysis

Numerous heavy atom derivatives have been tested attempting to solve the structure of PCS such as: Potassium tetrachloroplatinate(II), organic mercury derivatives, 5-amino-2,4,6-triiodoisophthalic acid and lanthanide phasing compounds (NatX-ray SAS, Saint Martin d'Hères, France). These derivatives resulted in either very low occupancy of the heavy atoms or a significant decrease in diffraction. A recently developed lanthanide complex, Xo4, containing a terbium ion [8] gave the best results with high anomalous signal not interfering with diffraction quality for short time soakings. We solved PCS in the *C*2 crystalline form using the single-wavelength anomalous scattering method. Datasets were collected at beamline Proxima-2A of the SOLEIL synchrotron (Paris, France) at the Tb LIII absorption edge (wavelength of 1.649165 Å) on two different crystals (condition 3, see above) soaked in Tb-Xo4. Merging provided a 3.45 Å resolution data set with high redundancy facilitating the location of 18 terbium sites using the SHELX [41] software and PHASER [42]. After multiple cycles of phasing, electron density modification, and secondary structure building using AUTOSOL from the PHENIX package [43], the electron density quality was sufficient to build a model with Buccaneer from the CCP4 package [42]. The initial model was then used as template for molecular replacement with a dataset of a native crystal (condition 1, see above) using PHASER-MR. The native dataset was collected (wavelength of 0.97625 Å) at beamline ID29 of the ESRF (Grenoble, France). Manual extension of the model was done using COOT [44]. Several rounds of manual and automatic refinements were performed using COOT and PHENIX-Refine. The model (PDB 6EQO) structure was refined with Ramachandran statistics of 94.93 % favored, 4.82 % allowed, and 0.25 % outliers.

6.5.8. Limited proteolysis

PCS at 0.5 mg/mL was forced into a supposedly closed or open state by adding 3 mM CoA or a combination of 3.4 mM ATP, 2 mM 3-hydroxypropionate and 2 mM $NADP^+$, respectively. A zero time sample of 10 µL was taken. Trypsin (Promega, diluted in 25 mM Tris-HCl pH 7.5, 10 mM $CaCl_2$) was added in a protein:protease ratio of 200:1. Samples of 10 µL were taken at different time points. The sample was quenched with 10 µL 4x SDS buffer and heated at 90°C for 10 min. Samples were applied onto an SDS-PAGE gel. For peptide quantification, the limited proteolysis was repeated as described. However, the samples were quenched by adding PMSF protease inhibitor (dissolved in 2-propanol) to a final concentration of 1 mM. The propylation was performed overnight in the dark (at RT) in 100 mM HEPES buffer pH 7.5, 400 mM sodium cyanoborohydride and 5 % (v/v) acetone [45]. The open state PCS samples were treated with unlabeled acetone while the closed state PCS samples were incubated with D6-labeled acetone. The reaction was stopped with 0.07 % TFA. Samples were concentrated and dried in a Speedvac. The open and closed state samples of each time point were combined and purified over a C18 membrane (cut from Empore™ SPE disks). Peptides were eluted with 0.1 % trifluoroacetic acid (TFA) in 50 % acetonitrile (ACN). Samples were dried in the Speedvac. Peptides were resuspended in 50 µl 0.1 % TFA. 1 µl of the peptide sample was mixed with 1 µl solution of 3 g/L alpha-Cyano-4-hydroxycinnamic acid in 80 % ACN (v/v) containing 0.3 % TFA onto a MALDI plate. The dried spots were measured automatically for MS and MSMS in a MALDI TOF/TOF analyzer (Applied Biosystems/MDS Sciex, Framingham, MA, USA) and the 4800 Series Explorer Software.

6.5.9. Time course assays

The time course assays with 0.1 µM PCS contained 0.8 mM CoA, 0.5 mM 3-hydroxypropionate, 0.8 mM ATP, 4 mM $MgCl_2$, 40 mM KCl, 50 mM $KHCO_3$ and 0.3 mM NADPH, if stated, in 100 mM potassium phosphate buffer pH 8. At specific time points 20 µL of the assay were quenched with 20 µL of 50 % formic acid. The time course assay with 10 µM PCS contained 5 mM CoA, 5 mM 3-hydroxypropionate, 5 mM ATP, 6 mM $MgCl_2$, 60 mM KCl and 2 mM NADPH in 100 mM Tris-HCl pH 7.8. At specific time points 10 µL of the assay were quenched with 10 µL acetonitrile and 10 % formic acid. Standard curves (1 µM to 500 µM) for quantification for 3-hydroxypropionyl-CoA, acrylyl-CoA and propionyl-CoA were prepared in the corresponding buffer conditions. The samples were centrifuged at 17´000 × *g* and frozen in liquid nitrogen. Samples were immediately thawed before application to hrLC-MS.

6.5.10. Isotopic labeling competition experiment

The competition contained 3 mM CoA, 2 mM 3-hydroxypropionate, 200 µM NADPH, 5 mM ATP, 7.5 mM Mg_2Cl, 60 mM KCl, 100 mM $KHCO_3$ and 100 mM Tris-HCl pH 7.8. For the competition either 100 µM ^{13}C-3-hydroxypropionyl-CoA or 100 µM ^{13}C-acrylyl-CoA was added. The assay was started with 2 µL of 1.28 mg/mL PCS wt and the reaction monitored photospectrometrically at 340 nm using a Cary-60 UV/Vis spectrometer (Agilent Technologies Inc. Santa Clara, CA, USA) at 30°C using quartz cuvettes (10-mm path-length; Hellma® (Germany)). The assay was quenched after a ΔAbs of 0.36 that corresponds to a turnover of 60 µM. The isotopic pattern of the produced propionyl-CoA was analyzed by hrLC-MS.

6.5.11. High resolution LC-MS (hrLC-MS)

3-hydroxypropionyl-Coa, acrylyl-CoA and propionyl-CoA were analyzed using an Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode through a 1290 Infinity UPLC (Agilent Technologies Inc. Santa Clara, CA, USA). Compounds were separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex XB-C18, Phenomenex, Aschaffenburg, Germany) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 (A) and methanol (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250

µl/min: 0 min 0% B; 1 min 0% B, 3 min 2.5% B; 9 min 23% B; 14 min 80 %B; 16 min 80%; 17 min 0 % B; 18 min 0 % B.

Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 l/min, 225 °C) and sheath gas (12 l/min, 400°C). The TOF was calibrated using an ESI-L Low Concentration Tuning Mix (Agilent Technologies Inc. Santa Clara, CA, USA) before measurement (residuals less than 2 ppm for five reference ions). MS data were acquired with a scan range of 750-1200 *m/z*.

CoA-thioesters were additionally detected by UV absorbance at 260 nm using a diode array detector (1290 Infinity II, Agilent Technologies Inc. Santa Clara, CA, USA)

LC-MS data were analyzed using MassHunter Qualitative Analysis software (Agilent).

6.5.12. SAXS analysis

PCS was freshly purified as described above two days before SAXS analysis was performed. The protein was stored on ice until measurements. Gel filtration buffer for dilutions and blank measurements was treated equally. SAXS data were recorded at the European Synchrotron Radiation Facility (Grenoble, France) on beamline BM29. The protein was up-concentrated at the beamline. If stated, cofactors were added to the concentrated protein at following concentrations: 3-hydroxypropionate, 2mM; CoA, 3 mM; ATP, 3.4 mM. Two-fold dilution series (4 mg/mL to 0.125 mg/ml) were prepared by dilution with gel filtration buffer containing the corresponding cofactors. The different dilutions were measured to investigate sample quality. Sample storage and measurement temperature was set to 20°C. The ESRF BM29 online software was employed for primary data reduction. PrimusQt (version 4.8.1 [46]) was used for data analysis.

6.6. References

1. Wheeldon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov, P.; Sigman, M., *Nat. Chem.* **2016,** *8* (4), 299-309.
2. Linster, C. L.; Van Schaftingen, E.; Hanson, A. D., *Nat. Chem. Biol.* **2013,** *9* (2), 72-80.
3. Wheeldon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov, P.; Sigman, M., *Nature chemistry* **2016,** *8* (4), 299.
4. Alber, B. E.; Fuchs, G., *J. Biol. Chem.* **2002,** *277* (14), 12137-12143.
5. Zarzycki, J.; Brecht, V.; Müller, M.; Fuchs, G., *Proc. Natl. Acad. Sci. USA* **2009,** *106* (50), 21317-22.
6. Teufel, R.; Kung, J. W.; Kockelkorn, D.; Alber, B. E.; Fuchs, G., *J. Bacteriol.* **2009,** *191* (14), 4572-4581.
7. Berg, I. A.; Kockelkorn, D.; Buckel, W.; Fuchs, G., *Science* **2007,** *318* (5857), 1782-1786.
8. Engilberge, S.; Riobé, F.; Di Pietro, S.; Lassalle, L.; Coquelle, N.; Arnaud, C.-A.; Pitrat, D.; Mulatier, J.-C.; Madern, D.; Breyton, C., *Chem. Sci.* **2017,** *8* (9), 5909-5917.
9. Chowdhury, C.; Sinha, S.; Chun, S.; Yeates, T. O.; Bobik, T. A., *Microbiol. Mol. Biol. Rev.* **2014,** *78* (3), 438-468.
10. Sutter, M.; Greber, B.; Aussignargues, C.; Kerfeld, C. A., *Science* **2017,** *356* (6344), 1293-1297.
11. Sutter, M.; Boehringer, D.; Gutmann, S.; Günther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., *Nat. Struct. Mol. Biol.* **2008,** *15* (9), 939-947.
12. Jung, T.; Grune, T., *Redox Biol.* **2013,** *1* (1), 178-182.
13. Reger, A. S.; Carney, J. M.; Gulick, A. M., *Biochemistry* **2007,** *46* (22), 6536-6546.
14. Bock, T.; Reichelt, J.; Müller, R.; Blankenfeldt, W., *Chembiochem* **2016,** *17* (17), 1658-1664.
15. Quade, N.; Huo, L.; Rachid, S.; Heinz, D. W.; Müller, R., *Nat. Chem. Biol.* **2012,** *8* (1), 117-124.
16. Spivey, H. O.; Ovadi, J., *Methods* **1999,** *19* (2), 306-21.
17. Lyle, S.; Ozeran, J. D.; Stanczak, J.; Westley, J.; Schwartz, N. B., *Biochemistry* **1994,** *33* (22), 6822-6827.
18. Jogl, G.; Tong, L., *Biochemistry* **2004,** *43* (6), 1425-1431.
19. Ho, B. K.; Gruswitz, F., *BMC Struct. Biol.* **2008,** *8* (1), 49.
20. Tanaka, S.; Sawaya, M. R.; Yeates, T. O., *Science* **2010,** *327* (5961), 81-84.
21. Giessen, T. W.; Silver, P. A., *Nat. Microbiol.* **2017,** *2* (6), 17029.
22. Fan, C. G.; Cheng, S. Q.; Sinha, S.; Bobik, T. A., *PNAS* **2012,** *109* (37), 14995-15000.
23. Kerfeld, C. A.; Sawaya, M. R.; Tanaka, S.; Nguyen, C. V.; Phillips, M.; Beeby, M.; Yeates, T. O., *Science* **2005,** *309* (5736), 936-938.
24. Mouilleron, S.; Badet-Denisot, M.-A.; Golinelli-Pimpaneau, B., *J. Biol. Chem.* **2006,** *281* (7), 4404-4412.
25. Singh, H.; Arentson, B. W.; Becker, D. F.; Tanner, J. J., *PNAS* **2014,** *111* (9), 3389-3394.
26. Smith, N. E.; Vrielink, A.; Attwood, P. V.; Corry, B., *Biophys. J.* **2012,** *102* (4), 868-877.
27. Ishikawa, M.; Tsuchiya, D.; Oyama, T.; Tsunaka, Y.; Morikawa, K., *EMBO J.* **2004,** *23* (14), 2745-2754.
28. Smith, S.; Tsai, S.-C., *Nat Prod Rep* **2007,** *24* (5), 1041-1072.
29. Vögeli, B.; Engilberge, S.; Girard, E.; Riobé, F.; Maury, O.; Erb, T. J.; Shima, S.; Wagner, T., *PNAS* **2018**, 201718649.
30. Bale, J. B.; Gonen, S.; Liu, Y. X.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D., *Science* **2016,** *353* (6297), 389-394.
31. Aussignargues, C.; Pandelia, M. E.; Sutter, M.; Plegaria, J. S.; Zarzycki, J.; Turmo, A.; Huang, J. C.; Ducat, D. C.; Hegg, E. L.; Gibney, B. R.; Kerfeld, C. A., *J. Am. Chem. Soc.* **2016,** *138* (16), 5262-5270.
32. Giessen, T. W.; Silver, P. A., *Chembiochem* **2016,** *17* (20), 1931-1935.
33. Azuma, Y.; Zschoche, R.; Tinzl, M.; Hilvert, D., *Angew. Chem.* **2016,** *55* (4), 1531-4.
34. Burton, A. J.; Thomson, A. R.; Dawson, W. M.; Brady, R. L.; Woolfson, D. N., *Nat. Chem.* **2016,** *8* (9), 837-844.
35. Brasch, M.; Putri, R. M.; de Ruiter, M. V.; Luque, D.; Koay, M. S. T.; Caston, J. R.; Cornelissen, J. J. L. M., *J. Am. Chem. Soc.* **2017,** *139* (4), 1512-1519.
36. Peter, D. M.; Vögeli, B.; Cortina, N. S.; Erb, T. J., *Molecules* **2016,** *21* (4), 517.
37. Schwander, T.; von Borzyskowski, L. S.; Burgener, S.; Cortina, N. S.; Erb, T. J., *Science* **2016,** *354* (6314), 900-904.
38. Bertani, G., *J. Bacteriol.* **1951,** *62* (3), 293.

39. Tartof, K.; Hobbs, C., *Focus* **1987,** *9* (2), 12.
40. Sambrook, J.; Russel, D., *Molecular cloning: a laboratory manual*. 3 ed.; Cold Spring Laboratory Press, New York: 2001.
41. Sheldrick, G. M., *Acta Crystallogr. A* **2008,** *64* (1), 112-122.
42. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A., *Acta Crystallogr. D* **2011,** *67* (4), 235-242.
43. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W., *Acta Crystallogr. D* **2010,** *66* (2), 213-221.
44. Emsley, P.; Cowtan, K., *Acta Crystallogr. D* **2004,** *60* (12), 2126-2132.
45. Eichacker, L.; Granvogl, B.; Gruber, P. Method for quantitative comparison of two or more proteins. EP1947461, 2008.
46. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H.; Svergun, D. I., *J. Appl. Crystallogr.* **2003,** *36* (5), 1277-1282.
47. Le, S. Q.; Gascuel, O., *Mol. Biol. Evol.* **2008,** *25* (7), 1307-1320.
48. Sudhir Kumar, G. S., and Koichiro Tamura, *Mol. Biol. Evol.* **2015**.
49. Felsenstein, J., *Evolution* **1985,** *39* (4), 783-791.

6.7. Supplementary Information

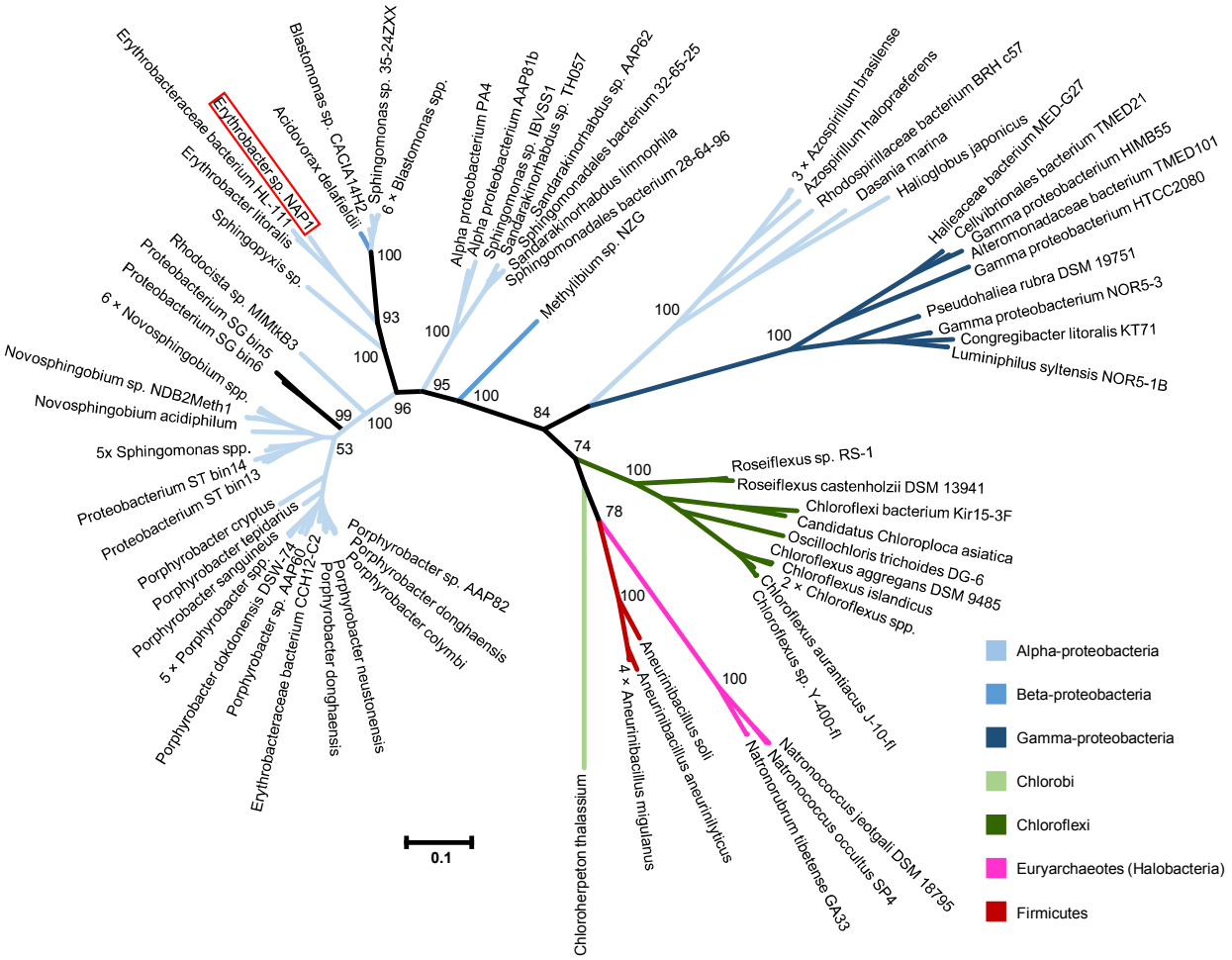


Figure S1. Maximum likelihood phylogenetic tree of PCS homologues. The evolutionary history was inferred based on the Le-Gascuel model [47]. The scale of the branch lengths is measured in number of substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 [48]. Bootstrap values [49] for confidence limits are given at important nodes. The position of the PCS homologue of *Eryhrobacter sp*. NAP1 is highlighted with a red box.

Figure S2. PCS domain distribution. Sequence alignment of *Erythrobacter* sp. NAP1 PCS (PDB 6EQO) with its closest lone-standing homologues; the acetyl-CoA ligase of *Salmonella enterica* (PDB 2P2F) [13], the HMG-CoA dehydratase of *Myxococcus xanthus* (PDB 5JBX) [14] and the enoyl-CoA carboxylase/reductase of *Streptomyces* sp. JS360 (PDB 4A0S) [15]. The secondary structure of PCS is represented above the sequence and colored according to their domain contribution; orange - ligase domain, purple – dehydratase domain, cyan - reductase domain.

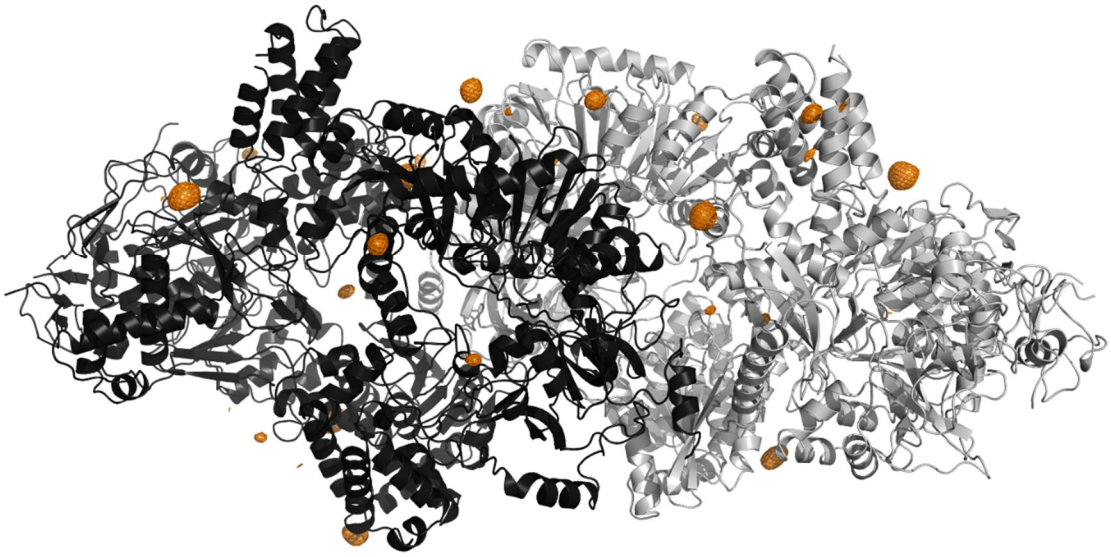


Figure S3. Anomalous map contoured at 6 σ indicating the positions (in orange) of terbium contained in the new phasing compound Tb-Xo4 [8].

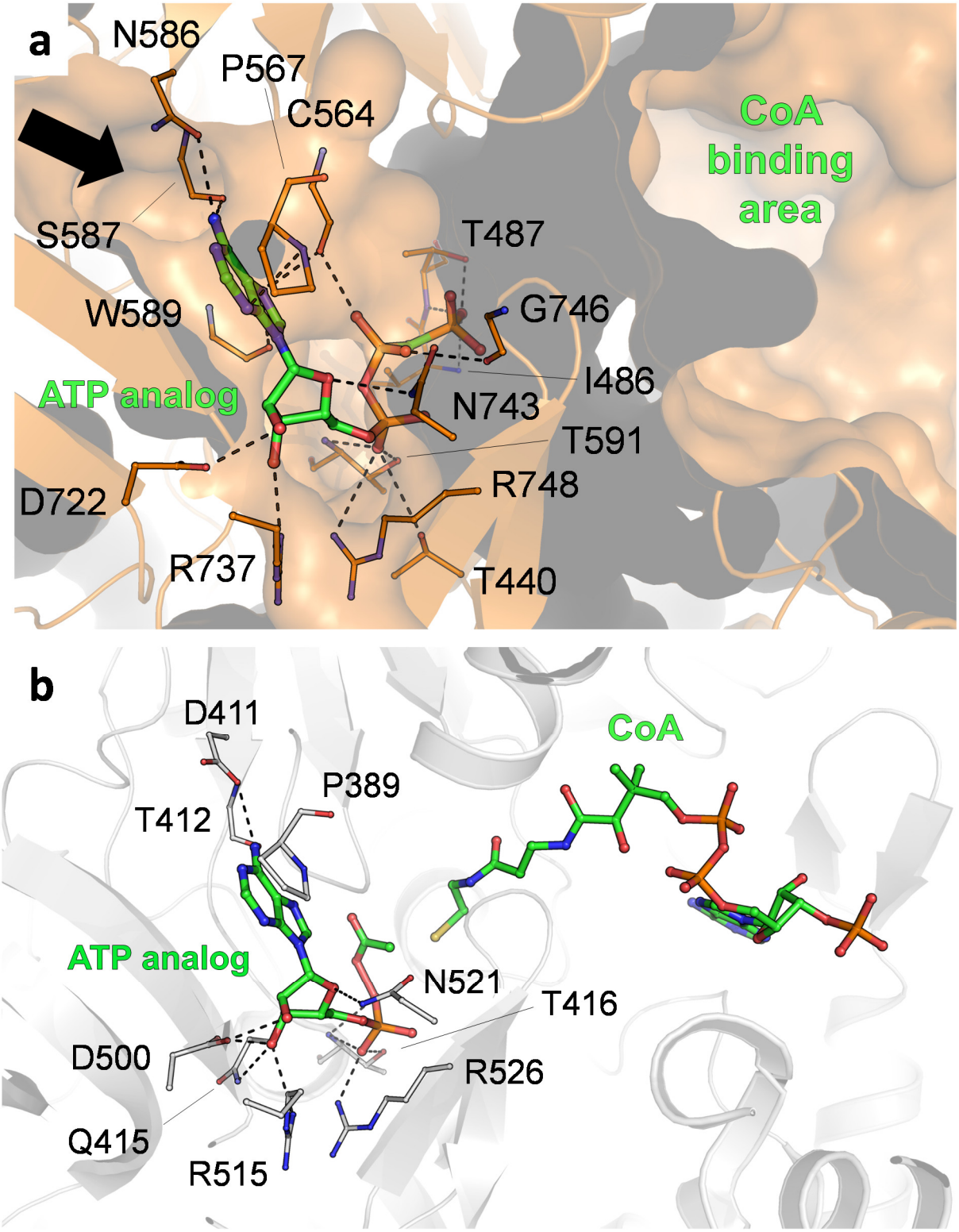


Figure S4. Close-ups of the active sites of the ligase domain of PCS and a close homologue. a, Active site of the PCS ligase domain co-crystalized with an ATP analog (green) and with the predicted CoA binding site indicated (green label). The ATP entry site is indicated by a black arrow. **b,** Active site of a lone-standing CoA ligase of *Salmonella enterica* (PDB 2P2F) [13] co-crystalized with an ATP analog and CoA (green).

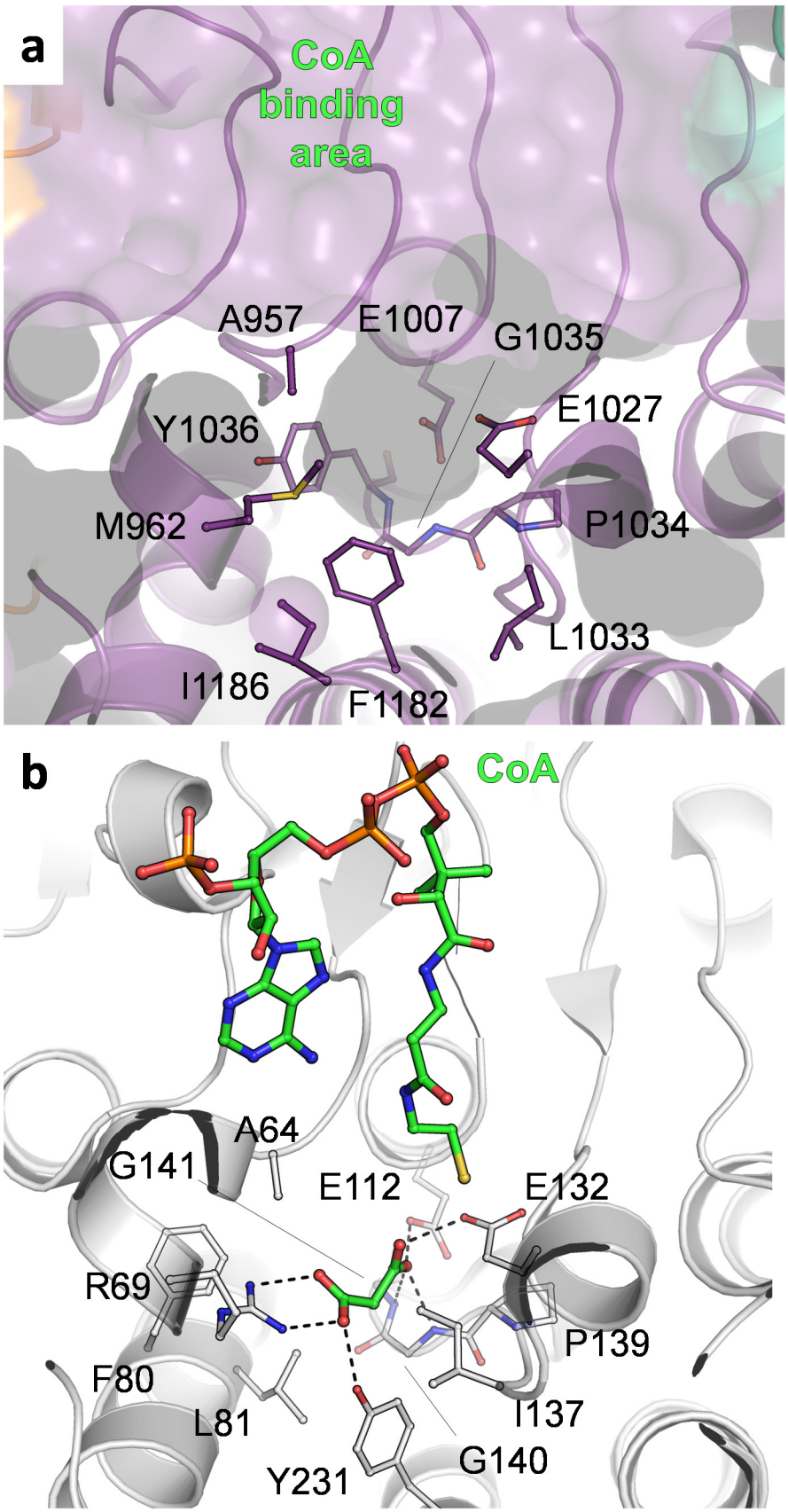


Figure S5. Close-ups of the active sites of the dehydratase domain of PCS and a close homologue. a, Active site of the PCS dehydratase domain with the essential active site glutamate E1027. **b,** Active site of a lone-standing dehydratase of *Myxococcus xanthus* co-crystalized with CoA and malonic acid (green) (PDB 5JBX) [14] with the essential active site glutamate E132.

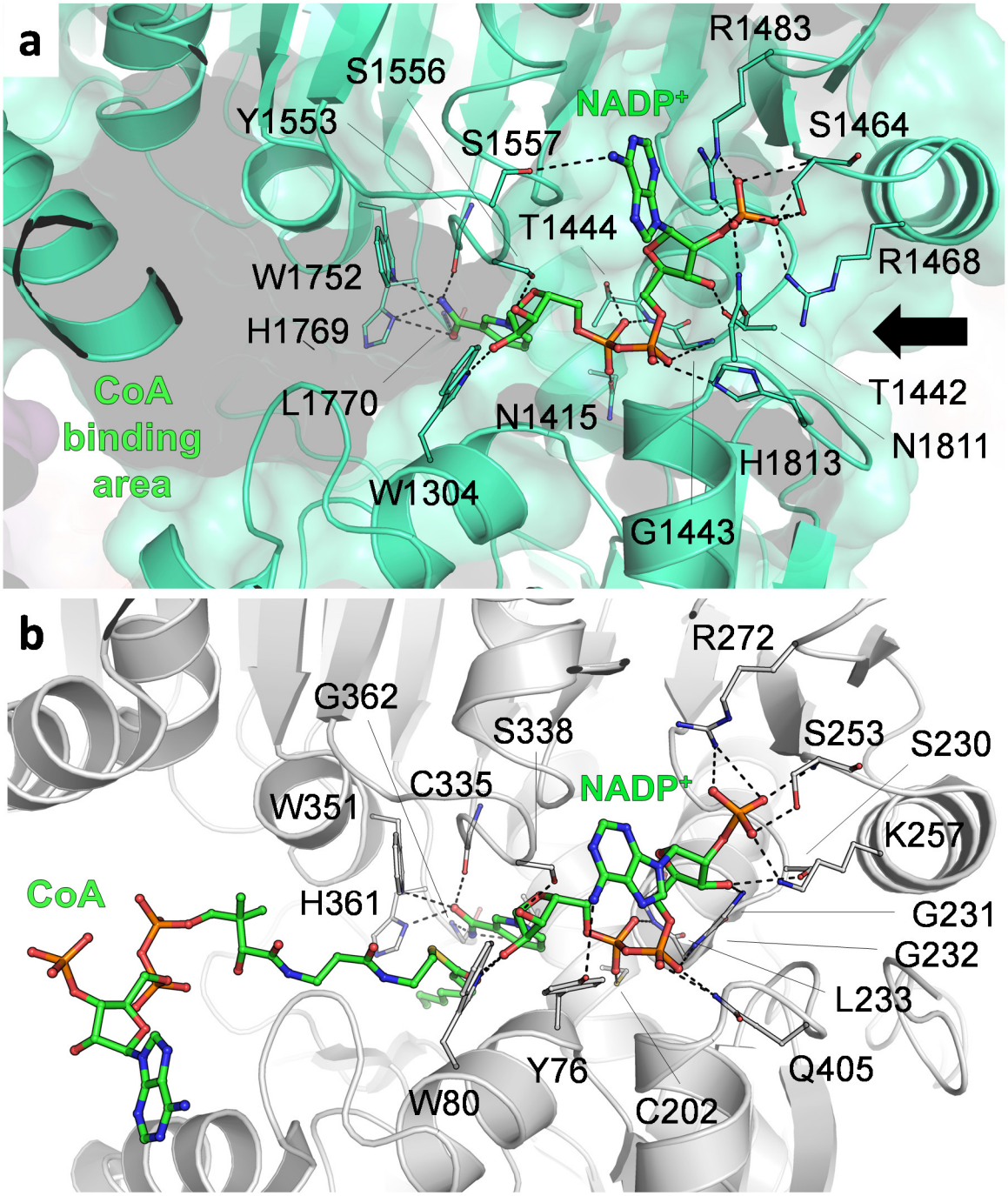


Figure S6. Close-ups of the active sites of the reductase domain of PCS and a close homologue. a, Active site of the PCS reductase domain with essential active site histidine H1769 co-crystalized with $NADP^+$ (green). The $NADP^+$ entry site is indicated by a black arrow. **b,** Active site of a lone-standing enoyl-CoA carboxylase/reductase of *Streptomyces* sp. JS360 co-crystalized with octenoyl-CoA and $NADP^+$ (green) (PDB 4A0S) [15] with the essential active site histidine H361.

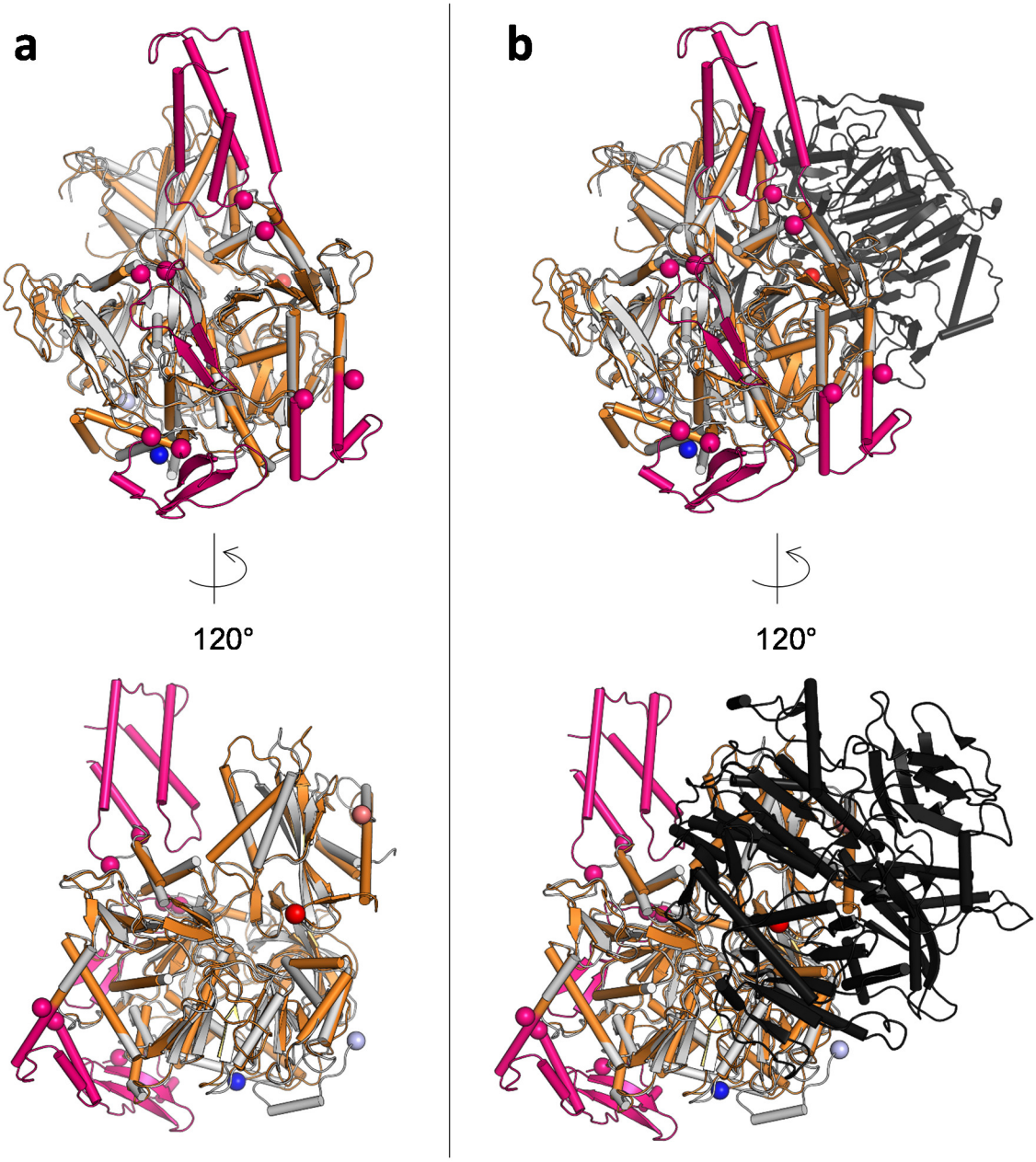


Figure S7. Superposition of the ligase domain with a close homologue, the acetyl-CoA synthetase of *Salmonella enterica* (PDB 2P2F) [13]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. a, Superposition of a PCS ligase domain (orange) with the monomer of the homologue (gray, rmsd of 0.930 Å over 447 Cα-atoms). Additional structural elements of the PCS monomer are shown in pink. b, Superposition of a PCS ligase domain with the dimer of the homologue (gray and black). The peripheral extensions of the PCS ligase domain are not aligning with the second chain of the dimeric homologue. The largest additional structural element is a four helix bundle on top of the structure.

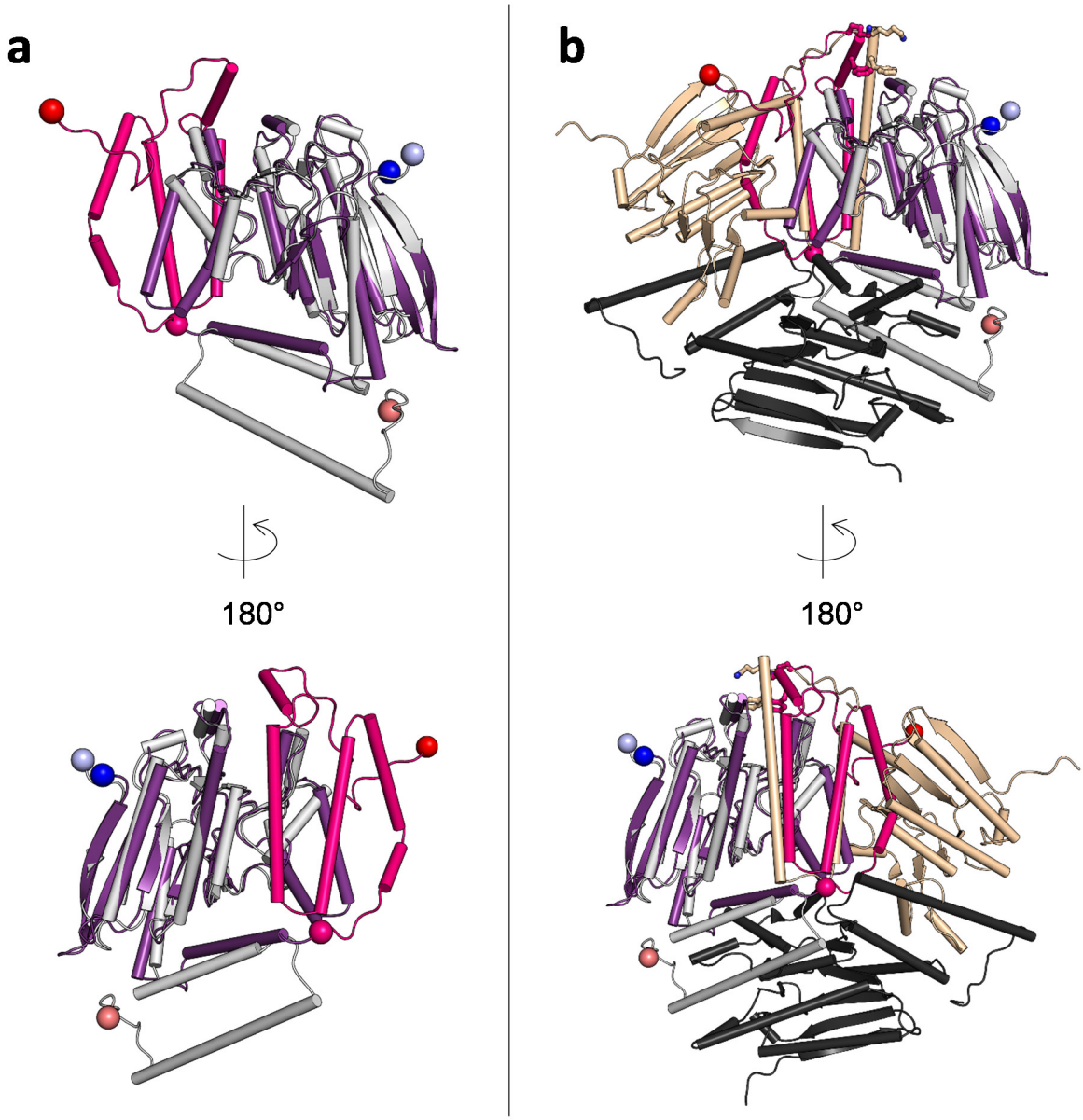


Figure S8. Superposition of the PCS dehydratase domain with a close homologue, the HMG-CoA dehydratase of *Myxococcus xanthus* (PDB 5JBX) [14]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. a, Superposition of a PCS dehydratase domain (purple) with the monomer of the homologue (gray, rmsd of 1.085 Å over 147 Cα-atoms). Additional structural elements of PCS dehydratase domain are shown in pink. b, Superposition of a PCS dehydratase domain with the trimeric homologue (gray, sand and black). The additional secondary structure elements in the PCS dehydratase domain partly mimic the neighboring chain of the trimeric homologue.

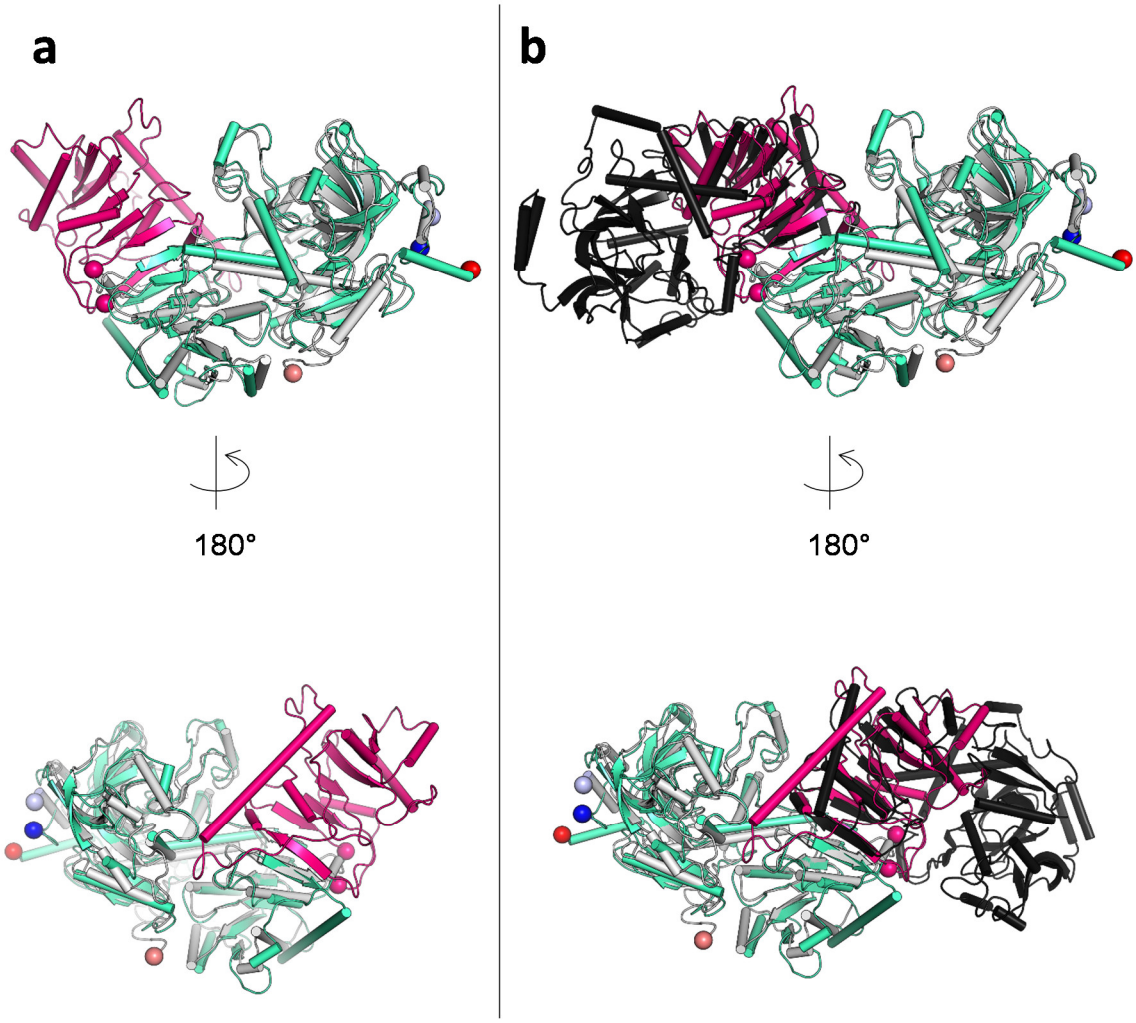
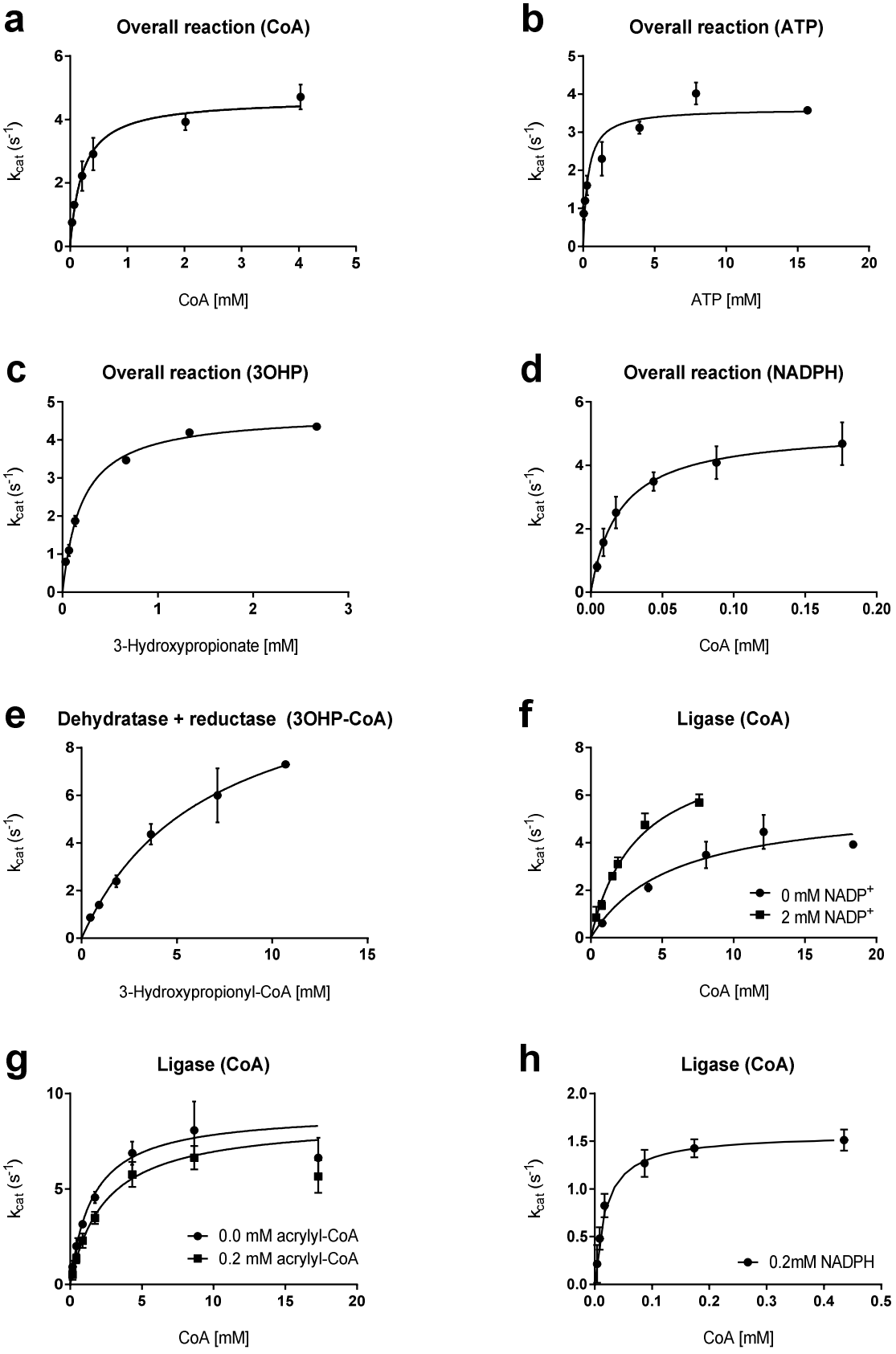


Figure S9. Superposition of the PCS reductase domain with a close homologue, the enoyl-CoA carboxylase/reductase of *Streptomyces* sp. JS360 (PDB 4A0S) [15]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. a, Superposition of a PCS reductase domain (cyan) with the monomer of the homologue (gray, rmsd of 1.530 Å over 291 Cα-atoms). Additional structural elements of PCS reductase domain are shown in pink. b, Superposition of a PCS reductase domain with the dimer of the homologue (gray and black). The additional secondary structure elements in the PCS reductase domain partly mimic the neighboring chain of the homologue dimer.


a
Overall reaction (CoA)
kcat (s-1)
CoA [mM]
b
Overall reaction (ATP)
kcat (s-1)
ATP [mM]
c
Overall reaction (3OHP)
kcat (s-1)
3-Hydroxypropionate [mM]
d
Overall reaction (NADPH)
kcat (s-1)
CoA [mM]
e
Dehydratase + reductase (3OHP-CoA)
kcat (s-1)
3-Hydroxypropionyl-CoA [mM]
f
Ligase (CoA)
kcat (s-1)
0 mM NADP+
2 mM NADP+
CoA [mM]
g
Ligase (CoA)
kcat (s-1)
0.0 mM acrylyl-CoA
0.2 mM acrylyl-CoA
CoA [mM]
h
Ligase (CoA)
kcat (s-1)
0.2mM NADPH
CoA [mM]

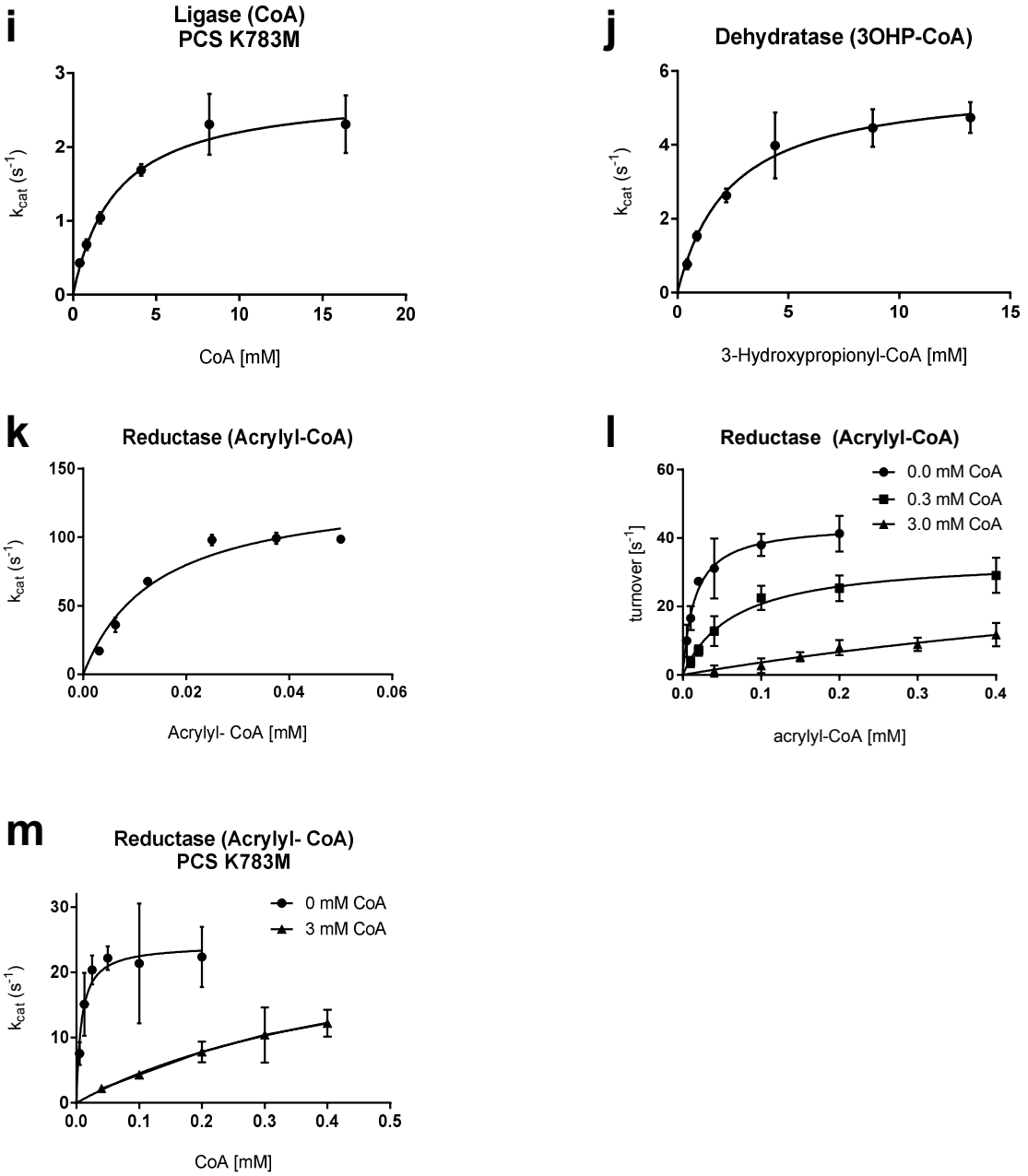
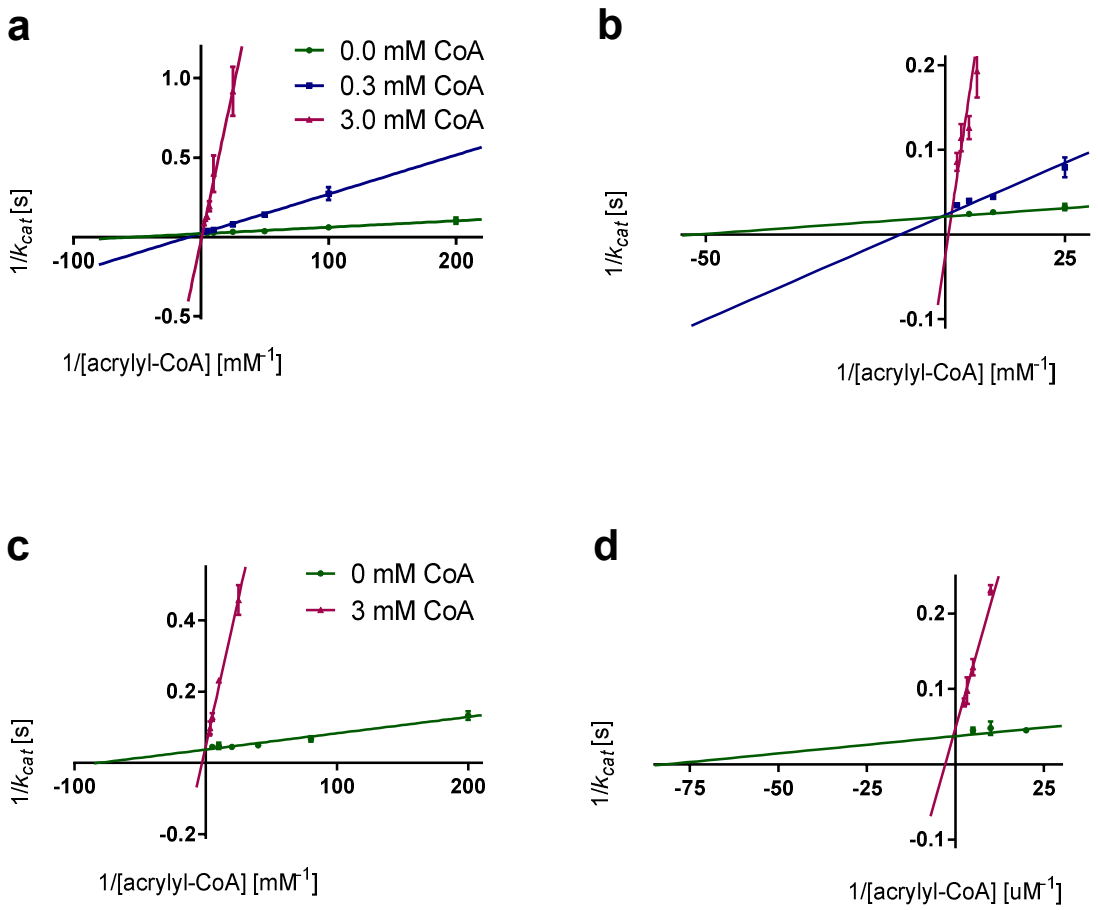


Figure S10. Michaelis-Menten plots of the kinetic characterizations of the reactions of PCS from *Erythrobacter* sp. NAP1. All points were measured in triplicates and the 95% confidence intervals are indicated. For kinetic parameters see **Table S2**, for assay compositions see **Table S3. a,** Overall reaction of PCS dependent on CoA concentration. The curve was fitted excluding values above 5 mM CoA, which resulted in substrate inhibition. **b,** Overall reaction of PCS in dependence of ATP concentration. Substrate inhibition was observed at 10 mM ATP. **c,** Overall reaction of PCS in dependence of 3-hydroxypropionate concentration. **d,** Overall reaction of PCS in dependence of NADPH concentration. **e,** The coupled dehydratase and reductase reaction depending on the 3-hydroxypropionyl-CoA concentration. **f,** The ligase reaction alone depending on the CoA concentration in presence or absence of $NADP^+$. **g,** The ligase reaction alone using the PCS E1027Q variant deficient in the dehydratase reaction to test the effect of acrylyl-CoA. Substrate inhibition can be observed above 10 mM ATP. The curve was fitted including those values. **h,** The ligase reaction alone using the PCS E1027Q variant deficient in the dehydratase reaction to test the effect of NADPH. **i,** The ligase reaction alone using the PCS E1027Q K783M variant depending on the CoA concentration. **j,** The dehydratase reaction alone using the H1769A variant deficient in the reductase activity to allow coupling to the stand-alone reductase Etr1p (from *Saccharomyces cerevisiae*) to monitor acrylyl-CoA formation. **k,** The reductase reaction alone depending on the acrylyl-CoA concentration. **l,** The reductase reaction alone using the PCS E1027Q variant depending on the acrylyl-CoA concentration in presence of different CoA concentrations. **m,** The reductase reaction alone using the PCS E1027Q K783M variant depending on the acrylyl-CoA concentration in presence of different CoA concentrations.



Figure

S11. Lineweaver-Burk representation of the effects of CoA on the reductase reaction of PCS. All points were measured in triplicates and the 95% confidence intervals are indicated. **a,** The reductase reaction alone catalyzed by the PCS E1027Q variant depending on the acrylyl-CoA concentration in presence of three different CoA concentrations. **b,** Detailed view of the data of a,. **c,** The reductase reaction alone catalyzed by the PCS E1027Q K783M variant depending on the acrylyl-CoA concentration in presence of two different CoA concentrations. **d,** Detailed view of the data of c,. While PCS E1027Q is also non-competitively inhibited by 3 mM CoA (changes in apparent KM, CoA and kcat) PCS E1027Q K783M variant is only inhibited competitively (change in apparent KM, CoA).

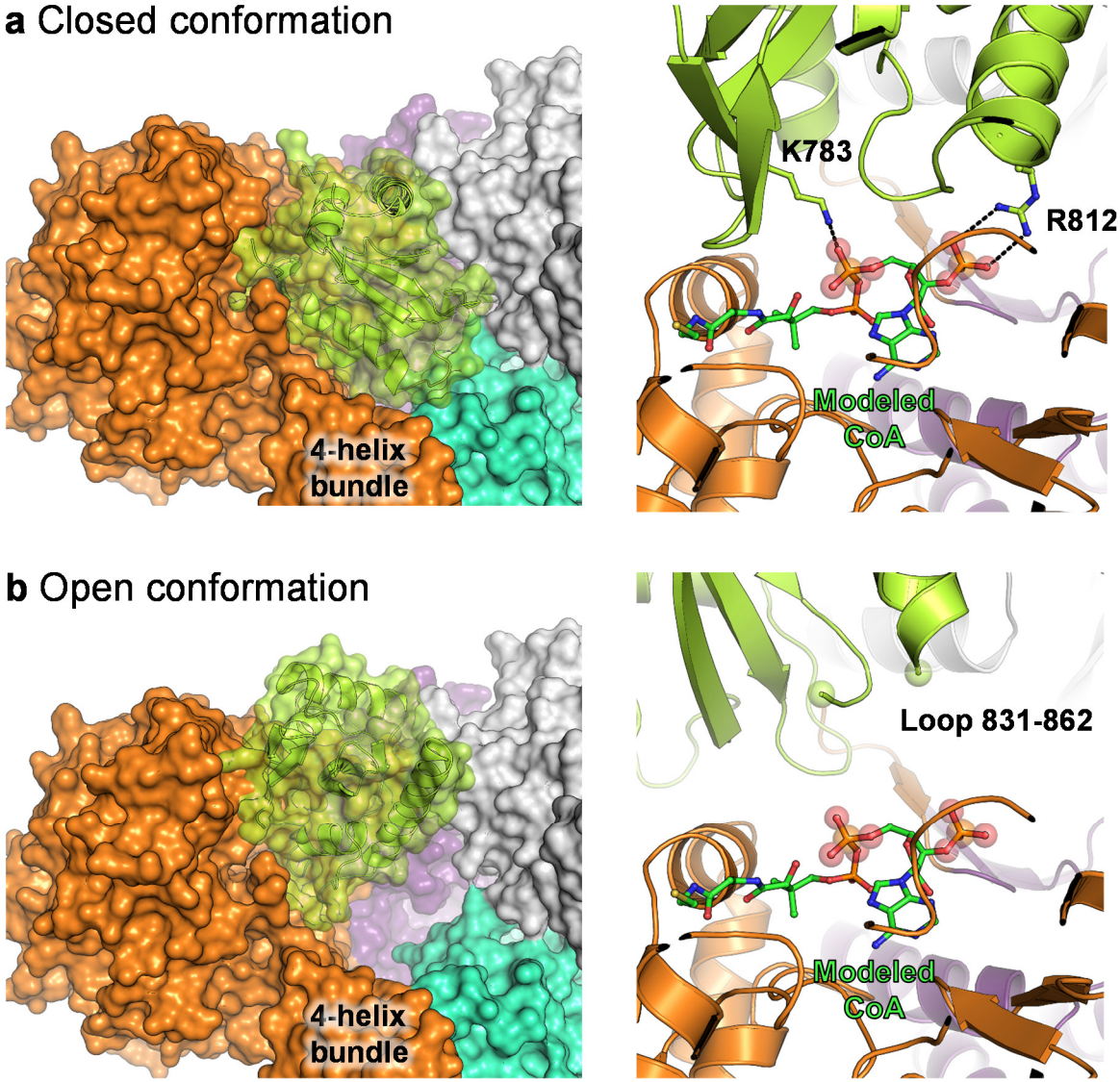


Figure S12. Hypothetic conformational change of PCS upon CoA binding. The left panel depicts the effect of the plausible conformational change on the overall structure by surface representation, while the right panel shows a close up of the ligase domain active site in the different conformations with modeled CoA. The modeled open conformation suggests how PCS allows substrates to enter the reaction chamber. Our observation of CoA preventing access of intermediates to the reaction chamber can be explained by this hypothesis. Note, that the PCS-specific four-helix bundle is required to enclose the enzyme compartment.

a, closed conformation of PCS as it was obtained from the crystal structure in this study. The closed conformation might be stabilized by interactions of K783 and R812 of the flexible cap domain with bound CoA. **b,** open conformation of PCS was modeled by replacing the flexible cap subdomain of PCS with the one from the open-state ligase of *Saccharomyces cerevisiae* (PDB 1RY2) [18]. The size of the opening is roughly 13 Å and 24 Å in the two dimensions. A loop corresponding to PCS residues 831-862 had to be omitted from the model because of clashes. The green spheres mark its insertion points. Orange – non-flexible part of ligase domain, light green – flexible cap of ligase domain (defined by comparing PDB 2P2F and 1RY2 [2, 5]), purple – dehydratase domain, cyan and grey – reductase domains from both protomers.

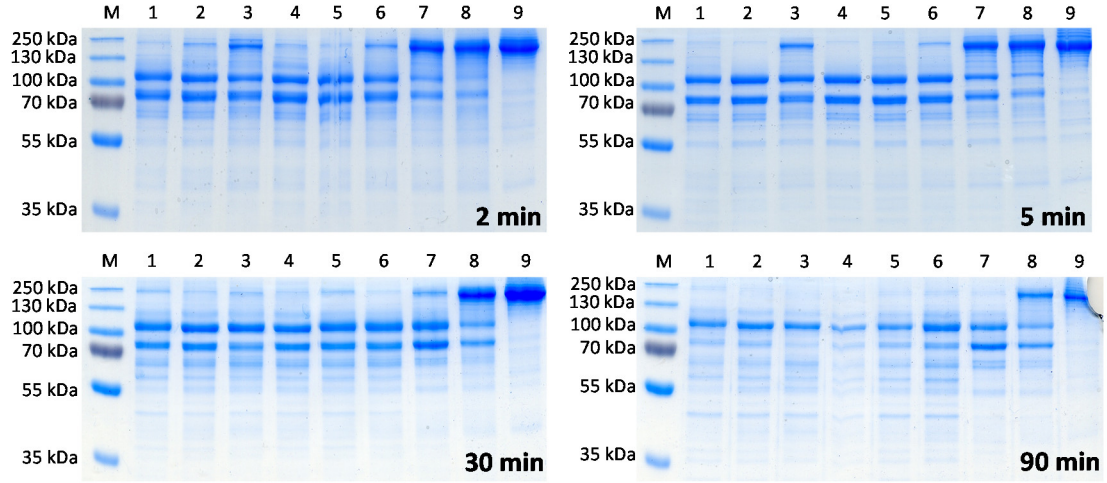


Figure S13. Limited proteolysis on PCS with different substrates/products. Limited proteolysis of PCS with trypsin (200:1 protein:protease ratio) analysed by SDS-PAGE. 1, PCS (0.5mg/ml) apo; 2, with CoA; 3, with CoA, ATP, $NADP^+$; 4, with an ATP analog, CoA, 3-hydroxypropionate; 5, with ATP; 6, with ATP, $NADP^+$; 7, with ATP, 3-hydroxypropionate; 8, with ATP, 3-hydroxypropionate, $NADP^+$; 9, PCS apo without trypsin. CoA at 3 mM, ATP and analog at 3.4 mM, 3-hydroxypropionate at 2mM, $NADP^+$ at 2mM.

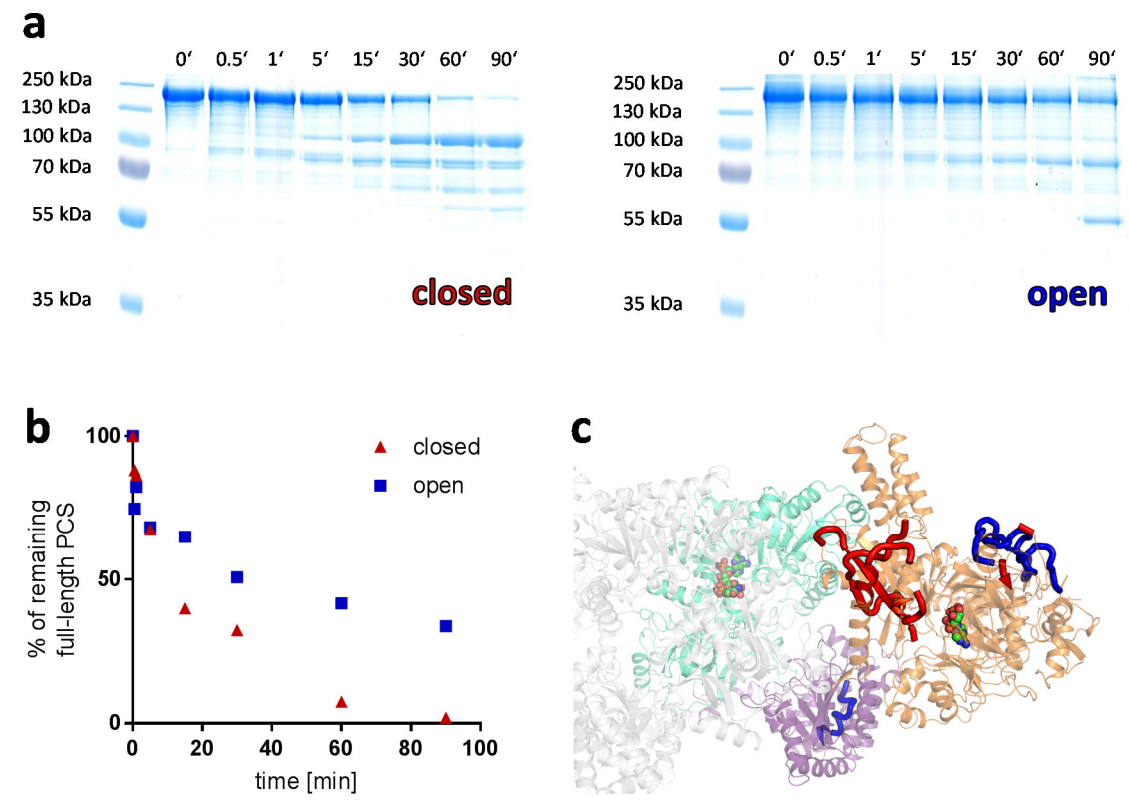


Figure S14. Limited Proteolysis of PCS in open and closed conformation. a, limited proteolysis of PCS with trypsin (200:1 protein:protease ratio) analysed by SDS-PAGE. The closed conformation (left) was induced by addition of 3 mM CoA, while the open conformation was stabilized by 2 mM 3HP, 3.4 mM ATP and 2 mM $NADP^+$ (right). **b,** percent of remaining full-length PCS over time was quantified from the SDS-PAGE gel. Samples on both gels derive from the same experiment and gels were processed in parallel. **c,** peptides from the closed and open conformation appearing after the first 30 seconds of the limited proteolysis were quantitatively compared. Peptides in blue are slightly less prominent in the sample of closed PCS (ratio of peptide in closed sample to open sample: 0.73 for peptide 866-890, 0.82 for peptide 714-724, 0.84 for peptide 1053-1065, 0.90 for peptide 121-131) while peptides in red are overrepresented in the sample of closed PCS (ratio of peptide in closed sample to open sample: 1.38 for peptide 664-673, 1.71 for peptide 866-890, 2.05 for peptide 805-818, 2.3 for peptide 837-855). Data of a representative single experiment.

Table S1. X-ray diffraction data collection and model refinement statistics

	PCS with $NADP^+$, phosphomethylphosphonic acid adenylate ester (PDB: 6EQO)	PCS with $NADP^+$, phosphomethylphosphonic acid adenylate ester, Tb-Xo4 (SAD data)
Data collection		
Space group	*C* 1 2 1	*C* 1 2 1
Cell dimensions		
a, *b*, *c* (Å)	383.41, 86.74, 133.96	389.68, 88.02, 134.91
α, β, γ (°)	90.00, 108.89, 90.00	90.00, 108.53, 90.00
Resolution (Å)	46.18 - 2.70 (2.75 - 2.70)*	49.66 - 3.45 (3.54 – 3.45)†
R_{merge}	0.212 (0.886)	0.497 (3.374)
$I / \sigma I$	4.2 (1.7)	7.5 (1.3)
$CC_{1/2}$ (%)	95.8 (18.9)	98.2 (30.5)
Completeness (%)	97.9 (94.2)	99.7 (99.0) / ano. 99.7 (98.8)
Redundancy	3.1 (2.7)	20.2 (16.9) / ano. 10.2 (8.5)
Refinement		*Not fully refined*
Resolution (Å)	46.18 - 2.70 (2.75 - 2.70)	*NA*
No. unique reflections	111918 (5260)	*NA*
R_{work} / R_{free}	0.191 / 0.229	*NA*
No. atoms	28218	*NA*
Protein	27636	*NA*
Ligands	150	*NA*
Water	432	*NA*
B-factors		
Protein	52.99	*NA*
Ligands	52.86	*NA*
Water	42.07	*NA*
R.m.s. deviations		
Bond lengths (Å)	0.006	*NA*
Bond angles (°)	0.765	*NA*

*The structure was determined from a single crystal. †Phasing was achieved with a merged dataset from two crystals. Values in parentheses are for highest-resolution shell. *NA*, not applicable.

Table S2. Kinetic parameters of the reactions of propionyl-CoA synthase of *Erythrobacter* sp. NAP1**.**

reaction	PCS variant	substrate	app. K_M in mM	app. k_{cat} in s^{-1}
overall	wt	CoA	0.22 ± 0.05	4.7 ± 0.3
	wt	ATP	0.34 ± 0.14	3.6 ± 0.3
	wt	3OHP	0.20 ± 0.02	4.71 ± 0.12
	wt	NADPH	0.020 ± 0.003	5.1 ± 0.2
dehydratase + reductase	wt	3OHP-CoA	6.6 ± 1.1	11.7 ± 1.0
ligase alone	wt	CoA	6 ± 3	5.7 ± 1.2
	wt	CoA (+ 2 mM NADP+)	3.3 ± 0.5	8.3 ± 0.6
	E1027Q	CoA	1.6 ± 0.4	9.1 ± 0.7
	E1027Q	CoA (+ 0.2 mM NADPH)	0.019 ± 0.003	1.63 ± 0.06
	E1027Q	CoA (+ 0.2 mM acrylyl-CoA)	2.4 ± 0.3	8.6 ± 0.5
	E1027Q K783M	CoA	2.6 ± 0.2	2.77 ± 0.08
dehydratase alone	H1769A	3OHP-CoA	2.4 ± 0.5	5.7 ± 0.4
reductase alone	E1027Q	acrylyl-CoA	0.014 ± 0.005	137 ± 17
	E1027Q	acrylyl-CoA (+ 0.0 mM CoA)	0.015 ± 0.003	44 ± 3
	E1027Q	acrylyl-CoA (+ 0.3 mM CoA)	0.070 ± 0.016	34 ± 3
	E1027Q	acrylyl-CoA (+ 3.0 mM CoA)	NA	NA
	E1027Q K783M	acrylyl-CoA	0.008 ± 0.003	24 ± 2
	E1027Q K783M	acrylyl-CoA (+ 3.0 mM CoA)	0.5 ± 0.4	29 ± 12

Data are shown as the mean ± 95% confidence intervals, as determined from a non-linear fit of at least 18 data points. Michaelis-Menten graphs of the original data are shown in Figure S11. NA, not applicable. For detailed reaction compositions, see Table S3.

Table S3. Composition of enzyme assays to determine kinetic parameters of PCS.

	overall (CoA)	overall (ATP)	overall (3OHP)	overall (NADPH)	dehydratase + reductase (3OHP-CoA)	ligase (CoA)	ligase + $NADP^+$ (CoA)	ligase +/- NADPH (CoA)	ligase +/- acrylyl-CoA (CoA)	dehydratase (3OHP-CoA)	reductase (acrylyl-CoA)	reductase +/- CoA (acrylyl-CoA)
P_i buffer pH 8	100	100	100	100	100	100	100	100	100	100	100	100
$KHCO_3$	50	50	50	50	50						50	50
3OHP	2.3	1.3	X	2.1		2.3	2.3	2.1	2.1			
CoA	X	2	2	1.7		X	X	X	X			X
ATP	5	X	5	3		5	3	3	3			
3OHP-CoA					X					X		
acrylyl-CoA									0.2 / 0		X	X
NADPH	0.3	0.3	0.3	X	0.3			0.2 / 0			0.3	0.3
KCl	40	40	40	40	40	40	40	40	40			
$MgCl_2$	4	4	4	10	4	10	10	10	10			
PEP						1	1		1			
NADH						0.3	0.3		0.3			
NADP+							2					
PCS [nM]	32	32	32	47	32	X[1]	17.3	104.0[2]/25.2[2]	22.3[2]	34.2[3]	1.8[2]	X[4]
CA [nM]	67	67	67	67	67						2	2
Etr1p [µM]										0.5		
PK/LDH [U]						~4.8/6.9	~2.4/3.5	~2.4/3.5	~2.4/3.5			
Myo [µM]						4.6	2.4	2.9	2.9			

All concentrations given in mM if not stated otherwise. Abbreviations: 3OHP, 3-hydroxypropionate; 3OHP-CoA, 3-hydroxypropionyl-CoA; CA, carbonic anhydrase (Simga-Aldrich C3934); PEP, phosphoenolpyruvate; Myo, myokinase; PK/LDH, mixture of pyruvate kinase and lactate dehydrogenase (Sigma-Aldrich P0294)

[1] 14 nM of wild type PCS or 68 nM of PCS E1027Q K783M.

[2] PCS E1027Q mutant that is deficient in the dehydratase reaction was used.

[3] PCS H1769A mutant that is deficient in the reductase reaction was used.

[4] PCS concentration was different depending on CoA concentration in assay. Assay with PCS E1027Q: without CoA, 5 nM PCS; with 0.3 mM CoA, 9 nM PCS; with 3 mM CoA, 17 nM PCS. Assay with PCS E1027Q K783M: without CoA, 12 nM PCS; with 3 mM CoA, 22 nM PCS.

CHAPTER VII

Archaeal acetoacetyl-CoA-thiolase/HMG-CoA-synthase complex channels the intermediate via a fused CoA-binding site

Authors:
Bastian Vögeli, Sylvain Engilberge, Eric Girard, François Riobé, Olivier Maury, Tobias J. Erb, Seigo Shima, Tristan Wagner

Published in:
PNAS 115, 3380-3385 (2018); DOI: 10.1073/pnas.1718649115

Author contributions:
B.V. and T.W. designed research; T.W. performed cultivation, enzyme purification, crystallization. T.W., S.E. and E.G. collected X-ray data; T.W. solved the structure. S.E. refined and deposited the structure. B.V. performed preparation of HMGCR and activity assays; S.E., E.G., F.R. and O.M. prepared Tb-Xo4 for phasing. B.V., T.J.E., S.S. and T.W. analyzed the data and wrote the paper with contribution by all other authors.

7. Archaeal acetoacetyl-CoA-thiolase/HMG-CoA-synthase complex channels the intermediate via a fused CoA-binding site

7.1. Abstract

Many reactions within a cell are thermodynamically unfavorable. To efficiently run some of those endergonic reactions, nature evolved intermediate-channeling enzyme complexes, in which the products of the first endergonic reactions are immediately consumed by the second exergonic reactions. Based on this concept, we studied how archaea overcome the unfavorable first reaction of isoprenoid biosynthesis; the condensation of two molecules of acetyl-CoA to acetoacetyl-CoA catalyzed by acetoacetyl-CoA thiolases (thiolase). We natively isolated an enzyme complex comprising the thiolase and HMG-CoA synthase (HMGCS) from a fast-growing methanogenic archaeon *Methanothermococcus thermolithotrophicus*. HMGCS catalyzes the second reaction in the mevalonate pathway, the exergonic condensation of acetoacetyl-CoA and acetyl-CoA to 3-hydroxy-3-methylglutaryl-CoA. The 380-kDa crystal structure revealed that both enzymes are held together by a third protein with so far unknown function (DUF35 protein). The active-site clefts of thiolase and HMGCS form a fused CoA-binding site, which allows for efficient coupling of the endergonic thiolase reaction with the exergonic HMGCS reaction. The tripartite complex is found in almost all archaeal genomes and in some bacterial ones. In addition, the DUF35 proteins are also important for polyhydroxyalkanoate (PHA) biosynthesis most probably by functioning as a scaffold protein that connects thiolase with 3-ketoacyl-CoA reductase. This natural and highly conserved enzyme complex offers great potential to improve isoprenoid and PHA biosynthesis in biotechnologically relevant organisms.

7.2. Introduction

Biological systems have evolved to efficiently catalyze a myriad of enzymatic reactions in metabolism. Among these reactions, some exhibit extremely unfavorable thermodynamics. One prominent example, found in the tricarboxylic acid cycle (TCA), is the oxidation of malate to oxaloacetate that is catalyzed by malate dehydrogenase (MDH). Under physiological conditions, the equilibrium constant for this reaction in the forward direction is K_{eq} = (2.86 ± 0.12) ×10^{-5} [1]; estimated ΔG°' = +30.3 kJ/mol using eQuilibrator [2]. If considered as an isolated reaction alone, this unfavorable step would drastically reduce the flux through the TCA cycle. However, recent *in vivo* chemical cross-linking experiments have shown that MDH forms a weakly associated complex with the next enzyme in the TCA cycle, citrate synthase (CS) [3-4]. The oxaloacetate produced by MDH is electrostatically guided from the active site of the MDH to the active site of the CS, without freely diffusing into the bulk solvent [5-6]. The direct substrate channeling from one active site to the other leads to an increased consumption of oxaloacetate by CS in a low-concentration microenvironment and allows high flux through the MDH/CS complex despite unfavorable thermodynamics of the first reaction [5]. This coupling mechanism balances the first endergonic reaction with the second and drives the high rate of citrate formation required [4-5].

Another example of a reaction with unfavorable thermodynamics in central carbon metabolism is the non-decarboxylative Claisen condensation of two acetyl-CoA molecules into acetoacetyl-CoA catalyzed by acetoacetyl-CoA thiolases (thiolase); K_{eq} = (1.1 ± 0.2) ×10^{-5} [7]; estimated ΔG°' = +26.1 kJ/mol using eQuilibrator [2]. Thiolases operate in all domains of life and are a part of several important metabolic pathways. There are two types of thiolases [8]. Degradative thiolases are involved in the β-oxidation of fatty acids, where they catalyze the thermodynamically favorable reaction; thiolytic cleavage of a wide range of 3-ketoacyl-CoAs into shorter acyl-CoAs and acetyl-CoA. Biosynthetic thiolases operate in

polyhydroxyalkanoate (PHA) and the mevalonate biosynthesis pathways, where they catalyze the thermodynamically unfavorable condensation of two molecules of acetyl-CoA to acetoacetyl-CoA. Both types of thiolases have significant sequence identity and show a similar protein fold [8]. In the mevalonate pathway, acetoacetyl-CoA is condensed with a third acetyl-CoA molecule to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by HMG-CoA synthase (HMGCS) and then reduced to mevalonate by HMG-CoA reductase (HMGCR) using NADPH as reductant (**Figure 1A**). The exergonic HMGCS and HMGCR reactions make the total mevalonate biosynthetic reaction thermodynamically favorable.

The mevalonate pathway is found in eukaryotes, archaea as well as some gram positive bacteria, where it forms the starting compound of isoprenoid biosynthesis [9]. Isoprenoids are one of the largest groups of natural products and display a wide variety of biological functions. They serve as hormones, photosynthetic pigments, quinones, plant defense compounds and as a component of membranes [10]. Particularly, archaea require a lot of isoprenoids as building blocks of the membrane lipid-bilayer. This raises the question how archaea realize high flux through the mevalonate pathway despite the unfavorable thermodynamics of the thiolase reaction.

Here, we show that the thiolase and HMGCS of the methanogenic archaeon *Methanothermococcus thermolithotrophicus* form a 380-kDa enzyme-complex in combination with a small scaffold protein (Pfam family: DUF35), which connects the two enzymes. The crystal structure of the native complex revealed a unique, shared CoA-binding site formed by both the thiolase and HMGCS subunit. Structural and biochemical data indicates that the complex uses this shared binding site to couple the endergonic non-decarboxylative Claisen condensation to the exergonic HMG-CoA forming reaction. This complex is conserved in most of archaea and can be found in many bacteria.

7.3. Results

7.3.1. A tripartite complex found by native purification

We initially identified the presence of the thiolase/HMGCS complex from *M. thermolithotrophicus*, which is a fast growing methanogen with a doubling time of about 30 minutes in a minimum mineral medium at 65 °C [11-12], by the shotgun purification of its proteome (see Supporting Information, **Figure 1B**). After identification of the multi-component enzyme, we purified the thiolase/HMGCS complex from the cell lysate by following mevalonate producing activity using a coupled enzyme assay containing HMGCR. In the purification steps, the total activity increased to 4.4 fold after the first two purification steps, which indicated that there are some inhibitory or regulatory components for the reactions in the lysate that were lost during these purification steps. At the end of the five-step purification, the specific activity was increased by 1000-fold (**Table 1**). Four protein bands were visible on SDS-PAGE in the final fraction (**Figure 1B**, **Figure S1**).

Multiple crystalline forms were observed in one crystallization drop. We dissolved single crystals and tested the enzyme activity (**Figure 1B**). A type of crystals exhibited the HMG-CoA-forming enzyme activity from acetyl-CoA and contained three proteins as shown by SDS-PAGE; the ratio of the band intensities was approximately 1:1:1. By mass spectrometry sequencing, these proteins were annotated as the 41.9-kDa thiolase, the 37.5-kDa HMGCS, and a 15-kDa DUF35 family-protein. The function of the DUF35 family proteins was at that point unknown [13]. Complexing of the DUF35 protein with thiolase and HMGCS, however, was in line with the observation that DUF35 proteins are typically encoded within the same transcriptional unit. Other crystals in the crystallization drop, which did not exhibit the HMG-CoA-forming activity, contained a protein with a molecular mass of 27.9 kDa present

in the SDS-PAGE. This protein was identified by mass spectrometry sequencing as thiazole synthase, which is not functionally related to the mevalonate pathway.

Table 1. Target purification of the thiolase/HMGCS complex.

Purification step	Protein (mg)	Total activity (U)	Specific activity (U/mg)	Purification (-fold)
Soluble fraction	4000	31	0.008	1
DEAE Sepharose	2300	99	0.043	5.4
Q Sepharose	1100	130	0.12	15
Hydroxyapatite	310	59	0.19	24
Source 15 Phe	4.3	43	10	1300

The activity was determined using HMGCR coupled assay. The assays contained 2-mM acetyl-CoA, 2-µM HMGCR, 450-µM NADPH in 100-mM HEPES-NaOH pH 7.5.

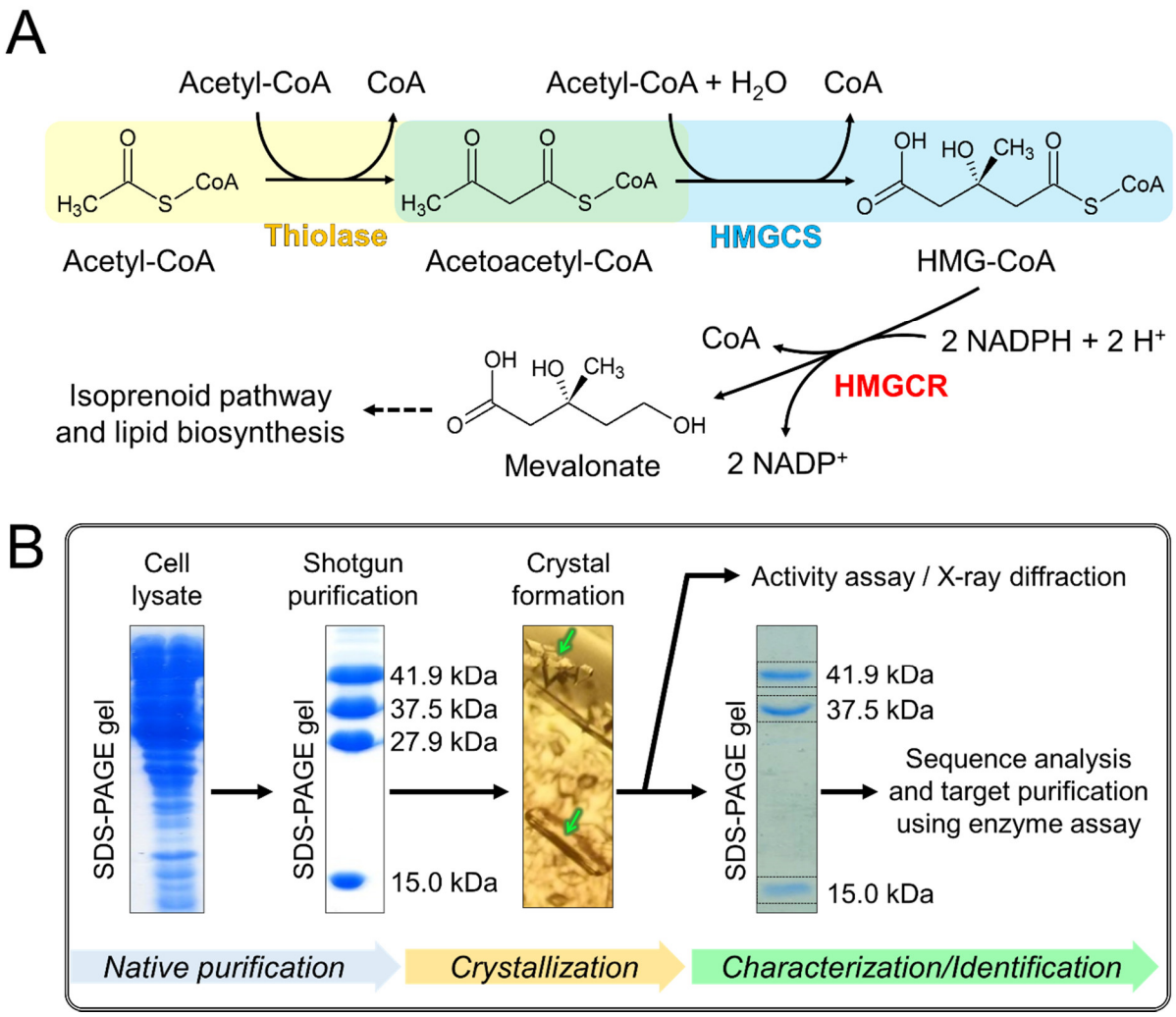


Figure 1. Mevalonate pathway and purification strategy. **(A)** Reaction scheme of the three enzymes involved in the mevalonate pathway. **(B)** Workflow applied in this work to identify and analyze the thiolase/HMGCS complex. Proteins in the cell extract was isolated by shotgun purification (see SI). Two different crystalline forms appeared (green arrows) from a fraction. Crystals were washed and dissolved. The protein bands on SDS-PAGE of the dissolved crystals were used for mass-spectrometry identification. Based on the protein-sequence analysis, the function of the crystallized proteins was predicted and the thiolase/HMGCS activity were measured by the HMGCR-coupled assay. The thiolase/HMGCS complex was purified

again from the cell extract using the enzyme assay (target-purification) and the purified enzyme complex was enzymologically characterized.

7.3.2. Architecture of the thiolase/HMGCS complex

The crystals containing thiolase, HMGCS and the DUF35 protein were used for X-ray diffraction experiments. We solved the structure of this enzyme complex using the newly developed terbium-containing phasing-reagent, Tb-Xo4 [14]. Diffraction data of the Tb-Xo4 soaked crystals were collected at the terbium L_{III} absorption edge and the 16 terbium sites with high occupancy were used to determine the structure by means of anomalous phasing (Materials and Methods, **Figure S2**). The complex contained two dimers of thiolase, two dimers of HMGCS and four monomers of the DUF35 family protein (**Figure 2A**). The structure of the thiolase dimer in the complex showed the highest similarity to the lone-standing Scp2 type-II thiolases from eukaryotic parasites (i.e. *Leishmania mexicana*, PDB: 3ZBG, root mean square deviation (rmsd) of 1.0 Å), which are folded as a five-layered αβαβα catalytic domain (**Figure S3A**) [15-17]. The HMGCS dimer from *M. thermolithotrophicus* has a very similar fold as the HMGCS from eukaryotes (i.e. *Homo sapiens*, PDB: 2P8U, rmsd 1.8 Å) and gram-positive bacteria, which adopt a five-layered αβαβα catalytic fold (**Figure S3B**) [18-19]. However, the HMGCS from *M. thermolithotrophicus* is reduced in size and lacks a C-terminal extension of about 50-100 amino acids compared to HMGCS from eukaryotes and gram positive bacteria.

The third component of the complex, the DUF35 protein, has high similarity with a homolog from *Sulfolobus solfataricus*, whose structure has been solved by a structural genomics initiative (PDB: 3IRB, rmsd 2.9 Å) (**Figure S3C** and **S3D**) [13]. A role in lipid and polyketide antibiotic biosynthesis was proposed for this homologue [13]. The DUF35-encoding genes are also found in operons containing sterol-carrier protein X related thiolases [20-22]. In the crystal structure of the *S. solfataricus* DUF35 homolog, the N-terminal helix is bound to the hydrophobic core of the protein and also bound to another DUF35 monomer [13]. This helix was speculated to serve as a scaffold for binding other proteins. In contrast, in the DUF35 protein from the thiolase/HMGCS complex, the hydrophobic core of the DUF35 protein binds to the thiolase dimer and its N-terminal segment interacts with one thiolase subunit (**Figure S3C**). A Zn(II) ion is coordinated in its rubredoxin domain and the rubredoxin binds to the second thiolase subunit. Furthermore, the core of the DUF35 protein displays multiple contacts with the thiolase and the HMGCS dimers with a total interface of 2019 Å² and 1640 Å² respectively (**Figure S4**). Compared to the DUF35 interaction, thiolase and HMGCS dimers perform only direct contacts with an area of interaction below 300 Å². The interface area between thiolase-DUF35-HMGCS is governed by electrostatic ring-like interactions in the periphery and some hydrophobic contacts close to the center, which are largely conserved throughout methanogens (**Figure S5**). Thus, the small DUF35 protein functions as a scaffold to connect the thiolase and HMGCS dimers.

7.3.3. Thiolase and HMGCS share a single CoA binding site

We investigated the substrate-binding sites of the thiolase/HMGCS complex by soaking with CoA, acetyl-CoA and HMG-CoA. We obtained one dataset at 2.95 Å resolution (see **Table S1**) from a soak with 100 mM acetyl-CoA for 1.5 min at 18 °C (see Materials and Methods). In this structure, an extra electron density appeared at the interface of the two enzyme-subunits and was modeled as an acetyl-CoA molecule (**Figure S6**). The adenine moiety of the acetyl-CoA molecule was bound to the thiolase part and the phosphate groups were fixed by three salt-bridges to the HMGCS part. The adenine- and phosphate-binding sites are located at the equivalent positions of the acetyl-CoA binding sites of the lone-standing thiolase and HMGCS, respectively (**Figure S7**). This finding indicated that thiolase and HMGCS in the enzyme complex, have only one common acyl-CoA-binding site, which is shared by both enzymes (**Figure 2B**). While the adenine interaction is well resolved in our structure, there is only partial electron density visible for the pantetheinyl-arm reaching into the thiolase active site. This is

probably due to the flexibility of the pantetheinyl-arm that can swing between the thiolase and HMGCS active sites. By using non-crystallographic symmetry on the omit map, we were able to detect peaks of electron density close to the two known catalytic cysteines in the thiolase (Cys85) and HMGCS (Cys114) (**Figure S8**). Since these two residues could be temporally acetylated in the catalytic reactions [15], we interpreted these densities as partially acetylated cysteine-sulfurs. The fused acetyl-CoA binding-site suggested that the two enzymatic reactions are directly coupled via the common CoA-binding site. In this hypothetical mechanism, the intermediate acetoacetyl-CoA swings from the thiolase active site to the HMGCS active site, where the intermediate is directly converted to HMG-CoA without releasing it into the bulk solvent.

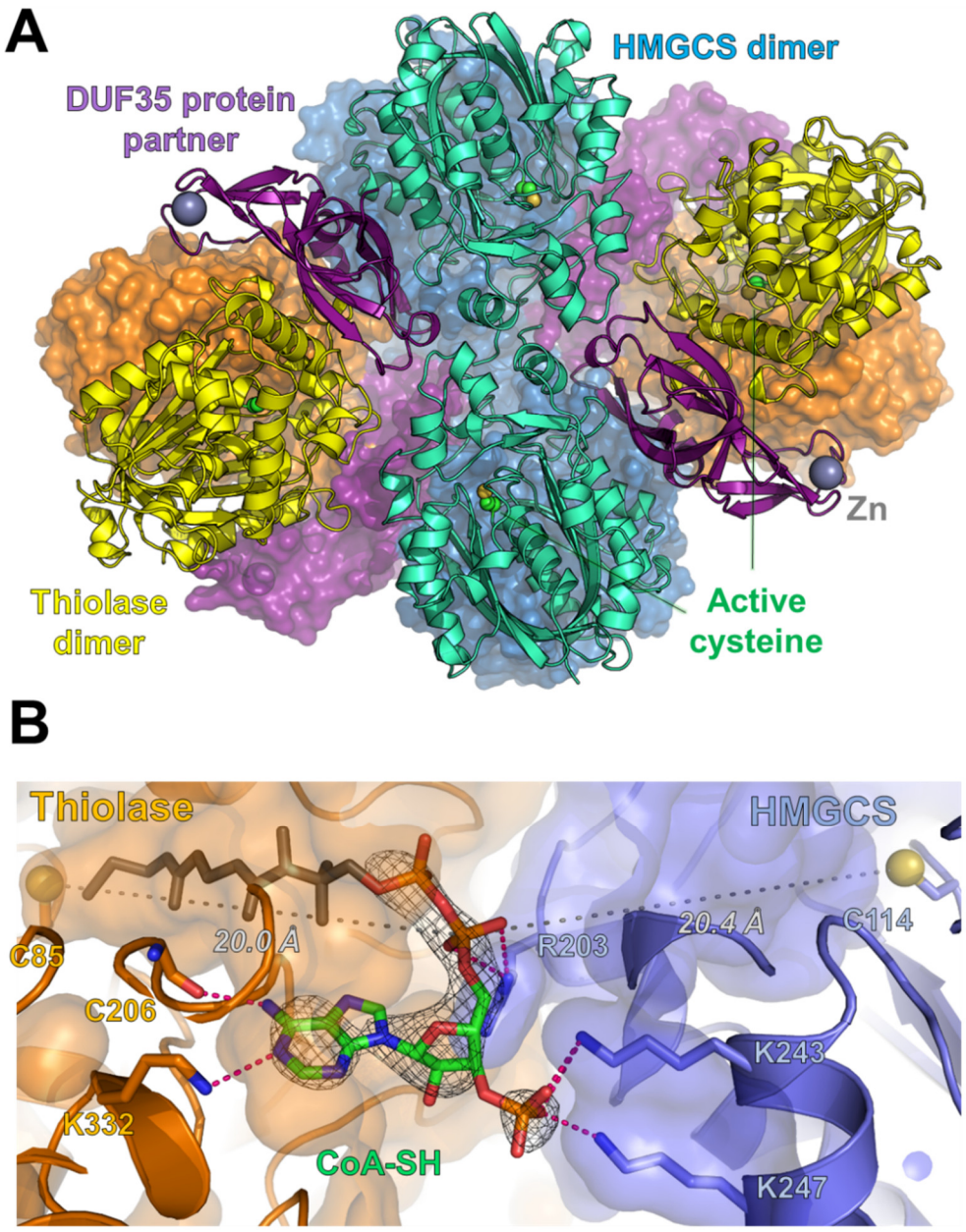


Figure 2. Thiolase and HMGCS are assembled in an active complex via the DUF35 protein, which provides a bridging platform for the organization of a shared CoA-binding site. (**A**) Quaternary representation of the thiolase/HMGCS complex in cartoon and surface models. Zn(II) ion and reactive cysteines are shown in ball and stick models. Color code highlights the different protein: yellow/orange for thiolase dimers, cyan/dark blue for the HMGCS dimers and purple for the DUF35 protein. (**B**) Binding site of acyl-CoA made at the interface of both enzymes. CoA (adenosine moiety in green) and CoA-binding residues are shown with stick model. The active-site cysteines are shown in yellow ball and sticks. The $2F_o$-F_c electron density from the omit map for CoA-SH is contoured at 1.0 σ.

7.3.4. Kinetic assays indicate channeling

The enzyme assay using HPLC-MS showed that the thiolase/HMGCS complex alone is able to convert acetyl-CoA to HMG-CoA (**Figure 3A**). Addition of an excess amount of HMGCR to the assay increased neither the reaction rate nor the stoichiometry of the final products (**Figure 3B**). These findings indicated that the coupled reactions of thiolase and HMGCS are thermodynamically favorable, which allows the production of HMG-CoA from acetyl-CoA. To determine the kinetic parameters of the thiolase/HMGCS complex, we used a photometrical assay with an excess amount of HMGCR. In this assay, oxidation of NADPH was followed. The thiolase/HMGCS complex showed a specific activity of 2.9 ± 0.2 U/mg considering the purity of the complex (60% judged by SDS-PAGE) and a $K_{m,app}$ for acetyl-CoA of 210 µM at 65 °C and at pH of 7.5 (**Figure 3C**).

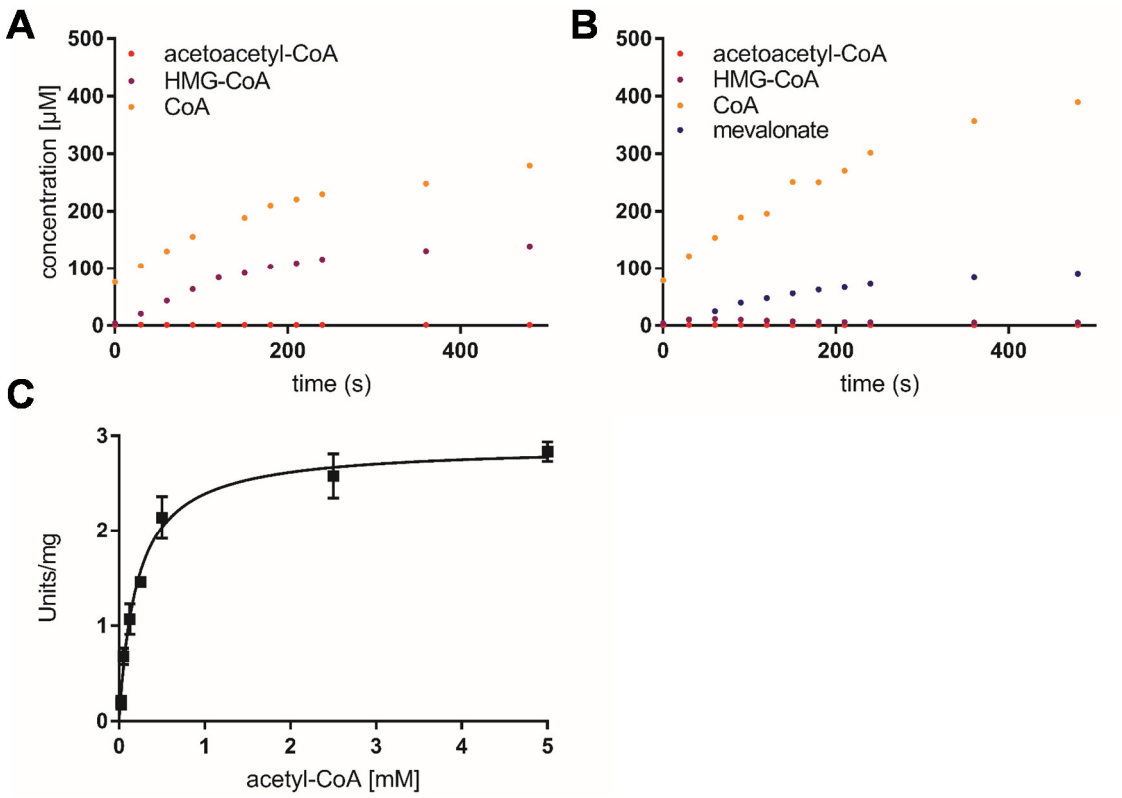


Figure 3. Activity assay of the thiolase/HMGCS complex. (**A**) Production of acetoacetyl-CoA, HMG-CoA and CoA catalyzed by the thiolase/HMGCS complex at 60 °C determined by the LC/MS analysis. The assays contained 100 mM HEPES-KOH pH 7.5, 0.1 mg/mL thiolase/HMGCS complex and 2 mM acetyl-CoA. (**B**) Production of acetoacetyl-CoA, HMG-CoA, CoA and mevalonate catalyzed by the thiolase/HMGCS complex in the presence of HMGCR at 60 °C determined by the LC/MS analysis. The assay contained 100 mM HEPES-KOH pH 7.5, 0.1 mg/mL thiolase/HMGCS complex, 200 µM NADPH and 1.35 µM HMGCR. (**C**) Effect of substrate (acetyl-CoA) concentration on the reaction rate catalyzed by the thiolase/HMGCS complex. The activity was measured with the HMGCR coupled assay by following the oxidation of NADPH at 340 nm. Standard errors of triplicate samples are shown.

In the next step, we tested whether acetoacetyl-CoA is released into the bulk solvent during the catalysis. We quenched assays during the steady state of the reactions that either contained only the thiolase/HMGCS complex or the complex and HMGCR. Using HPLC-MS analysis we could not detect any acetoacetyl-CoA above the detection limit of 100 nM in any of the assays, indicating that acetoacetyl-CoA concentrations were below 100 nM during steady state catalysis of the thiolase/HMGCS complex. It is unlikely that such low concentrations of acetoacetyl-CoA in solution can account for the measured turnover rates of 2.9 U/mg ($k_{cat,app}$ = 4.6 s^{-1}) of the complex if we compare it to the kinetic parameters of the archaeal HMGCS of *H. volcanii* (k_{cat} = 4.6 s^{-1}, $K_{m,acetoacetyl\text{-}CoA}$ = 1.4 µM [23]).

The data therefore indicates that acetoacetyl-CoA was consumed directly by the HMGCS without release of the intermediate from the active site of the thiolase into the bulk solvent.

We next investigated whether external acetoacetyl-CoA can enter the active site of the HMGCS by adding external acetoacetyl-CoA to the thiolase/HMGCS complex. The kinetic data shows that addition of external acetoacetyl-CoA slows down the overall turnover of the reaction, which suggests that external acetoacetyl-CoA is able to diffuse into the active site and acts as a competitive inhibitor by binding to the acetyl-CoA binding site (**Figure S9**). Such an inhibitory effect of acetoacetyl-CoA was shown for several lone-standing HMGCS [24-25]. Furthermore, we tested the utilization of acetoacetyl-CoA by a stable isotope-labeling experiment using the HMGCR-coupled assay. We started the reaction of the complex in the presence of 1 mM [1,2-$^{13}C_2$] acetyl-CoA and 1 mM unlabeled acetoacetyl-CoA (**Figure S10**). Mass spectrometric analysis of the HMGCR-coupled reaction products showed that the thiolase/HMGCS complex preferentially used the unlabeled acetoacetyl-CoA (**Table S2**). This finding confirmed that external acetoacetyl-CoA was diffusible into the active site, which suggested that the acetoacetyl-CoA intermediate formed from acetyl-CoA is not trapped in an internal compartment.

To assess the effect of complex formation on activity, we heterologously expressed the three proteins individually in *E. coli*. Unfortunately we were unable to purify the thiolase, as it was always found in insoluble inclusion bodies. The purified HMGCS did not show any activity towards acetoacetyl-CoA (concentrations of up to 10 mM acetoacetyl-CoA were tested) in coupled assays with HMGCR suggesting that the contribution of the thiolase to the binding site is required for activity.

7.3.5. DUF35 protein is a universal scaffold in archaea for lipid and PHA biosynthesis

We found that almost all archaea encode the three proteins of the thiolase/HMGCS complex; the thiolase, the HMGCS and the DUF35 protein (**Figure S11**, **Table S3** and **S4**), which indicated that the tripartite complex is essential in archaea for lipid biosynthesis. In some archaea, the DUF35 protein is directly fused to the C-terminus of HMGCS (i.e. in *Halobacteria*). The only exception is the recently discovered group of Lokiarchaea, which neither encodes a lone-standing DUF35 protein nor a DUF35-HMGCS fusion protein. However, the HMGCS of Lokiarchaea contains a C-terminal extension with a similar size as the DUF35 protein but a unique sequence. This C-terminal extension might fulfil a similar role as the DUF35 protein in the thiolase/HMGCS complex.

The DUF35 proteins are also found in the two PHA biosynthetic gene clusters of *Haloferax mediterranei* that is known to produce branched and unbranched polyalkanoates [26]. Hou et al. showed that the DUF35 proteins are essential for PHA biosynthesis. A double knock out of both DUF35 proteins completely abolished PHA production in *H. mediterranei*. This suggests that in PHA biosynthesis, the DUF35 protein also acts as a scaffold that brings together the thiolase with a 3-ketoacyl-CoA reductase, which couples the endergonic thiolase reaction to the exergonic reduction of acetoacetyl-CoA to 3-hydroxybutyryl-CoA.

Notably, the DUF35 scaffolding proteins are also found throughout the genomes of many bacteria (**Table S4**). The bacterial DUF35 homologues are co-located with a thiolase and in many cases either a HMGCS or a 3-ketoacyl-CoA reductase homolog, which suggests that these DUF35 proteins also organize tripartite mevalonate and/or PHA biosynthetic complexes in bacteria although no such enzyme complexes are reported so far.

In eukaryotes that also synthesize isoprenoids via the mevalonate pathway, no DUF35 homologues are found. Interestingly, however, eukaryotic and bacterial lone-standing HMGCS contain a C-terminal hairpin (e.g., 487-508 for the mitochondrial human HMGCS, PDB: 2WYA [19]) that superimposes well with the C-terminal part of the DUF35 scaffolding protein (i.e., residues 100-129 of the DUF35) from

the archaeal tripartite complex (**Figure S12**). These eukaryotic HMGCS have been characterized to act as lone-standing enzymes, however, the structural homology of the C-terminus suggests that they might complex with other protein in the cells.

7.4. Discussion

The identification of this natural channeling complex highlights the importance of native purification as a method to discover physiological enzyme complexes. The isolation of the native thiolase/HMGCS complex enabled us to find the third partner of these two enzymes, the DUF35 protein family, whose functions have been enigmatic. The DUF35 protein organizes the spatial co-localization of the thiolase and HMGCS dimers to form the thiolase/HMGCS complex. A shared acyl-CoA binding site is formed between the thiolase and the HMGCS. This shared binding site could enable acetoacetyl-CoA produced by the thiolase to swing directly from the thiolase into the HMGCS active site. This couples the endergonic first reaction with the exergonic second one to increase overall biosynthetic flux. The proposed reaction scheme is shown in **Figure 4**. Both active-site cysteines have to be acetylated first in order to ensure formation of acetoacetyl-CoA and HMG-CoA. Desynchronized reaction order will lead to unreactive enzyme substrate/intermediate complexes, as seen in the inhibitory effect of added bulk-solvent acetoacetyl-CoA (see above). To perform such a sequential ordered reaction, the thiolase and HMGCS subunits have to communicate with each other through conformational changes.

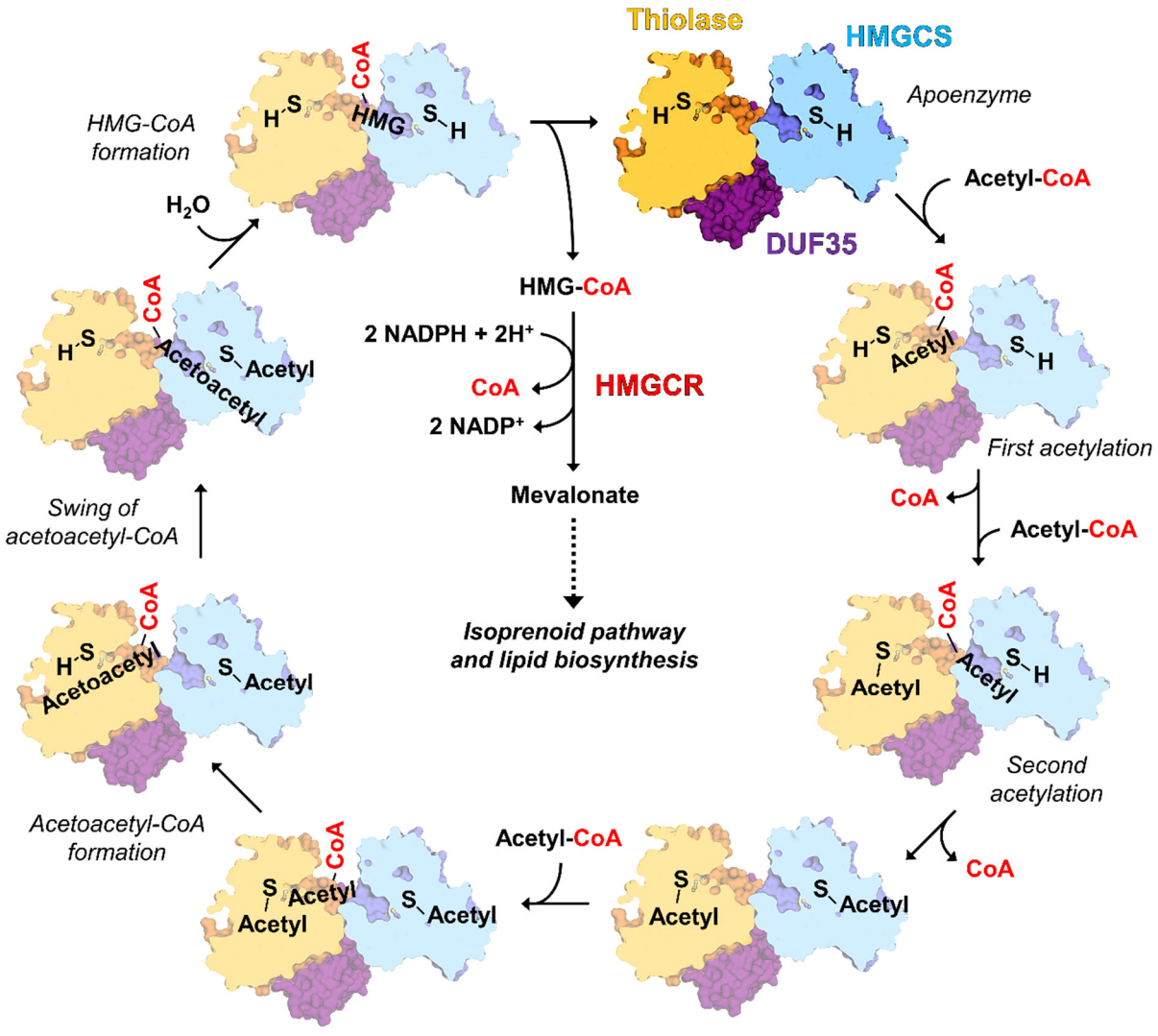


Figure 4. Proposed reaction sequence of the thiolase/HMGCS complex. The active-site Cys85 of thiolase (in yellow) and Cys114 of HMGCS (in blue) are shown as S-(H). In the initial two steps, both specific cysteine residues of thiolase and the

HMGCS are acetylated. In the next step, thiolase forms acetoacetyl-CoA using the third molecule of acetyl-CoA. The acetoacetyl-S-pantetheinyl-arm of acetoacetyl-CoA swings to the acetylated HMGCS active site and is converted to the final product HMG-CoA. The released HMG-CoA is further processed by HMGCR to form mevalonate through the mevalonate pathway.

The high total kinetic rate of the thiolase/HMGCS complex (2.9 ± 0.2 U/mg) indicated that the highly endergonic thiolase reaction was not only thermodynamically overcome but also kinetically accelerated by the channeling system in this enzyme complex, most probably by means of the shared CoA-binding site. A similar shared CoA-binding site has been reported for the bacterial multi-enzyme complex that catalyzes the last three sequential reactions in the fatty acid β-oxidation cycle [27]. In this complex, 2-enoyl-CoA hydratase, 3-ketoacyl-CoA thiolase and 3-hydroxyacyl-CoA dehydrogenase share the same CoA-binding site and use the long-spanning pantetheinyl arm to swing the intermediate from one active site to the next ones.

This mode of action of the thiolase/HMGCS complex resembles that of polyketide and fatty acid synthases that also channel catalytic intermediates from one active site to the other [28]. However, in contrast to the latter complexes, the intermediates of the thiolase/HMGCS are not covalently bound. It should be noted that polyketide synthases and the fatty acid synthases use decarboxylative Claisen condensation for the C-C bond formation. Instead of condensing two molecules of acetyl-CoA, they condense acetyl-CoA with malonyl-CoA releasing CO_2 during the reaction. This release of CO_2 serves as the driving force to catalyze the condensation reaction. The malonyl-CoA is synthesized from acetyl-CoA via ATP dependent carboxylation by acetyl-CoA carboxylase (ACC). The condensation of two molecules of acetyl-CoA by polyketide synthases and fatty acid synthases therefore uses one molecule of ATP. The archaeal tripartite complex on the other hand saves one molecule of ATP by coupling the endergonic condensation of two molecules of acetyl-CoA directly to the exergonic formation of HMG-CoA.

Although the presented tripartite complex is universal to most archaea, it does not appear to be essential for acetyl-CoA condensation reactions in some bacteria and in particular eukaryotes, which completely lack a DUF35 scaffolding protein. However, in bacteria and eukaryotes the demand for isoprenoids is generally lower compared to archaea, which use isoprenoids as a major component of membrane lipids. It remains to be seen whether in bacteria and eukarya lone-standing enzymes are sufficient to sustain isoprenoid production, or whether other mechanisms might be operating that help to overcome the unfavorable condensation reaction (e.g. high acetyl-CoA pools, compartmentalization).

Dueber et al. found that channeling by artificial connection of *E. coli* thiolase AtoB with the HMGCS and HMGCR from *S. cerevisiae* increased the production of mevalonate over lone-standing enzymes in *E. coli* by about 77 fold [29]. Our discovery of a naturally evolved channeling strategy in archaea harbors great potential to further improve the heterologous production of mevalonate in biotechnologically relevant strains.

7.5. Materials and Methods

7.5.1. Cultivation condition of *Methanothermococcus thermolithotrophicus*

M. thermolithotrophicus was grown at 65 °C in 1.5 liters of medium in the 2-liter fermenters. The growth medium is similar to that described by Belay *et al.* [30], but contained in addition 1 µM $NaHSeO_3$ per liter of the medium [31], 25 mM piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) and 25 mM 2-(N-morpholino)ethanesulfonic acid (MES). The inoculum was cultivated in anoxic bottles (50 ml) containing 5 ml of medium, whose gas phase included 20% CO_2/80% H_2 with +0.6 bar. *M. thermolithotrophicus* cells needs 1-2 weeks to adapt to the growth medium containing PIPES and MES in the preculture. In the fermenter cultures, the medium was gassed with 80% H_2/20% CO_2/0.01% H_2S with a flow of 1 liter per min and stirred at 1,000 rpm. The 1.5-l cultivation started with inoculation of 120−150-ml preculture. During the exponential phase, the doubling time was 25−30 min. When the optical density at 600 nm (OD_{600}) reached 2−3, the cells were harvested by centrifugation (6,000 × g) for 30 min under anoxic conditions at 8 °C, yielding approximately 2.5 g cells (wet mass) per liter. The cell pellets were frozen and stored at −75 °C before enzyme purification.

7.5.2. Shotgun purification of the native proteins

To determine the structure of the enzymes, which are able to be only functionally produced in the native microorganism, we purified proteins from the cell extract. The cells were disrupted and the proteins were fractionated by multiple column chromatography (i.e. DEAE Sepharose, Q Sepharose, Source 15 Phe, gel filtration). When several proteins were dominated in SDS-PAGE, the samples were subjected for crystallization. When protein crystals were obtained, the crystals were picked up from the crystallization drops, washed three times with reservoir solutions, dissolved in water and analyzed by SDS-PAGE. The protein bands obtained in the SDS-PAGE gel were sequenced with matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry as described previously [32]. After examination of the importance of the proteins, the crystals obtained were structurally analyzed by X-ray crystallography. When the function of the protein was predictable, the enzymatic activity of the dissolved crystals was confirmed. To obtain the target enzyme for further studies of their catalytic properties, the target was purified from the cell extract again using the enzyme activity assay to improve purification.

7.5.3. HMG-CoA reductase purification and characterization

The gene of HMG-CoA reductase (HMGCR) amplified by PCR from *M. thermolithotrophicus* DNA was cloned into the pET28b expression vector using the forward primer GCGGACATATGATGGAAAAAGGAAATAACGAAATAC and the reverse primer GTCCGAATTCTTATCTGCCTAATTGGGCATG, which included *Nde*I and *Eco*RI restriction sites, respectively. HMGCR was produced in *Escherichia coli* BL21-AI in Terrific Broth (TB) medium containing 50 µg/ml kanamycin. The overnight preculture (5 ml) was inoculated into 1-liter TB medium and cultivated at 37 °C until the OD_{600} reached 0.8. The culture was cooled to 30 °C and the gene expression was induced with 0.25 mM IPTG and 0.02% L-arabinose (final concentrations in the medium) and further cultivated for 12 h. Cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl pH 8.0 containing 500 mM NaCl and 5 mM β-mercaptoethanol (buffer A). Cells were disrupted by sonication and then the lysate was centrifuged at 40,000 × g for 60 min at 4°C. Supernatant was applied onto the HisTrap FF (5 ml) column (GE Healthcare, Freiburg, Germany) equilibrated with buffer A with a flow rate of 1 ml/min and washed with 30-column volumes of buffer A containing 75 mM imidazole. The protein was then eluted with buffer A containing 500 mM imidazole. The protein was desalted with a HiTrap Desalting column (GE Healthcare) into 20 mM Tris-HCl pH 8.0, 10 mM DTT, 150 mM NaCl,

20% glycerol and stored at −20 °C. Protein concentration was determined spectrophotometrically (ε_{280nm} = 23.8 cm^{-1} mM^{-1}). The enzyme activity was measured at 60 °C by following NADPH consumption at 340 nm using a Carry-60 UV-Vis spectrophotometer (Agilent) using 10mm quartz cuvettes (Hellma, Mülheim, Germany). The assay mixture contained 100-mM Tris-HCl pH 8.0, 12.6-nM HMGCR, 250-µM NADPH and various amounts of HMG-CoA. Its kinetic parameters were characterized; apparent k_{cat} = 22 ± 1 s^{-1} and apparent K_m for HMG-CoA = 28 ± 4 µM.

7.5.4. Native purification of the thiolase/HMGCS complex

The subsequent steps were performed under anoxic conditions (95%N_2/5%H_2) at 18 °C. Routinely 35-45 g of wet cell pellets are used for purification. Cells were thawed at room temperature. Lysis buffer (50-mM MOPS/NaOH, pH 7.0, 10-mM $MgCl_2$ and 2-mM dithiothreitol) was added at a ratio of 10 ml per g pellet to lyse the cells by osmotic shock. Lysis is performed at 18 °C under gentle stirring with an addition of 2.0 U/ml benzonase. The lysate was centrifuged twice at 10,000 ×g for 60 min at 4−10 °C. After the cell lysis, enzymes were purified without any freezing steps. The soluble fraction was applied on a DEAE Sepharose fast-flow column (75 ml) equilibrated with 50 mM Tricine-NaOH pH 8.0 containing 2-mM dithiothreitol (DTT) (buffer B). The column was washed at a flow rate of 4.5 ml/min for two column volumes. Elution was performed with an increasing gradient of NaCl from 50 mM to 350 mM in five column volumes. The thiolase/HMGCS complex eluted in the 130-240 mM NaCl fractions. The fractions were pooled and diluted with an equal volume of buffer B and the sample was loaded on a Q-Sepharose fast-flow column (55 ml) (GE Healthcare), which was pre-equilibrated in buffer B. The column was washed with 200-mM NaCl, and the proteins were eluted at a flow rate of 3.0 ml/min with an increasing gradient of NaCl from 200 mM to 400 mM in three column volumes. The thiolase/HMGCS complex typically eluted at 235–370-mM NaCl. Fractions containing the complex were diluted with an equal volume of 25-mM sodium phosphate buffer pH 7.6, containing 2-mM DTT (buffer C) and passed through a Ceramic Hydroxyapatite Type I column (40 ml) (Macroprep; Bio-Rad; München, Germany) equilibrated with buffer C. The column was washed with five-column volumes of buffer C and the enzyme complex was eluted with an increasing gradient of sodium phosphate buffer pH 7.6 from 25 to 500 mM with four-column volumes. The complex eluted at 220−490 mM of sodium phosphate buffer pH 7.6. Pooled fractions were subsequently diluted with three volume of 25-mM Tris-HCl pH 7.6, containing 2.0-M $(NH_4)_2SO_4$ and 2-mM DTT (buffer D), and loaded on a Source 15 Phe column (15 ml) that was equilibrated in the same buffer. After washing the column with one-column volume of buffer D, the enzymes were eluted at a flow rate of 1 ml/min with a decreasing gradient of $(NH_4)_2SO_4$ from 2.0 M to 0 M in a fourteen-column volumes. The thiolase/HMGCS complex eluted at 1.06–0.92 M $(NH_4)_2SO_4$. The fractions were pooled and concentrated by passing through a 50-kDa cut-off filter (Merck Millipore, Darmstadt, Germany). Finally, the concentrated sample was loaded on a 10/300 Superose 6 column (GE Healthcare) equilibrated with 25-mM Tris-HCl pH 7.6, containing 10% glycerol and 2-mM DTT and eluted at a flow rate of 0.4 ml/min. The thiolase/HMGCS complex eluted at 11.8−14.4 ml with a peak at 12.9 ml.

7.5.5. Enzymatic activity assay of the thiolase/HMGCS complex

The assay coupled with HMGCR was carried out with a Cary-4000 UV-Vis spectrometer (Agilent) at 60°C by following NADPH consumption at 340 nm using 10 mm quartz cuvettes (Hellma, Mülheim, Germany). To measure the activity of the fractions obtained in the purification steps, the assay solution contained 100 mM HEPES-KOH pH 7.5, 2 mM acetyl-CoA, 450 µM NADPH and 2 µM HMGCR. Kinetic parameters of the purified thiolase/HMGCS complex were determined at 60 °C using the assay solution containing 500 mM potassium phosphate buffer pH 7.5, 400 µM NADPH, 2 µM HMGCR, 4.5 µg/mL thiolase/HMGCS complex and varying concentrations of acetyl-CoA. The LC-MS experiments were done with the Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization

source set to positive ionization mode. Compounds were separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex XB-C18, Phenomenex) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 and methanol. Assays contained 100 mM HEPES/KOH pH 7.5, 0.1 mg/mL thiolase/HMGCS complex, 2 mM acetyl-CoA and in the coupled assay additionally 200 µM NADPH and 1.35 µM HMGCR. Samples were quenched with 5% formic acid (final concentration) and centrifuged for 5 min at 17,000 × g to remove precipitated proteins and then the supernatant was analyzed by LC-MS. For this assay, the thiolase/HMGCS complex stored for about 2 weeks at −80°C was used. The activity of the stored preparation was 15% of that of the fresh sample. Compounds in the assay was quantified using external standard curves prepared in the same buffer system (**Figure S13**).

7.5.6. Crystallization of the thiolase/HMGCS complex from *M. thermolithotrophicus*

Crystallization was done under air at 18−20 °C. The best-diffracting crystals were obtained using the sitting drop method (in a 24-well junior clover plate from Jena Bioscience). The crystallization reservoir contained 100 mM Tris-HCl pH 8.0, 25−28% v/v pentaerythritol ethoxylate (15/4 EO/OH, average molecular mass ≈ 797 Da) and 50 mM $MgCl_2$. The crystallization drop contained 1−2 µl of the purified complex (50−60 mg/ml) and 1 µl of the mother liquor. The best diffracted crystal was obtained in the presence of 10 mM Tb-Xo4 (14) (**Table S1**). For phasing, the crystals obtained in the crystallization solution in the presence of 10 mM Tb-Xo4 were soaked with 100 mM Tb-Xo4. Substrate-soaking experiment was done using crystals, which were obtained in the absence of Tb-Xo4. The best data were obtained when the crystals were soaked for 1.5 min in the mother liquor supplemented with 100 mM acetyl-CoA. The thiolase/HMGCS crystals were also soaked with several concentrations of CoA, acetyl-CoA and HMG-CoA; however, the majority of the crystals lost diffraction power, which probably came from conformational changes of the enzymes upon substrate-binding.

7.5.7. Structural analysis

A first intensive screening using crystals soaked in heavy atoms was performed on the BM30A beamline at ESRF. Soaking of the crystals in the mother liquor supplemented with 100 mM Tb-Xo4 [14] for 6 min gave the best result. The heavy atom derivative exhibited a very strong anomalous signal without altering the diffraction quality. After a fluorescent scan, a dataset was collected at the Tb L_{III} edge on Proxima 2A beamline (Synchrotron SOLEIL). Diffraction frames were integrated using XDS [33] and the integrated intensities were scaled using the CCP4 programs SCALA [34]. The structure was solved *de novo* by the single-wavelength anomalous diffraction method. Firstly, we localized the 39 terbium sites using SHELX software [35] and phased with Autosol from the PHENIX package [36]. The model building was completed with Buccaneer from the CCP4 package [37]. Molecular replacements were done with PHASER using this first model to solve the structure of the native and the acyl-CoA thiolase/HMGCS complex. Models were completed and improved in COOT [38] prior to refine with BUSTER [39] or PHENIX (33). These models were then optimized through iterative rounds of refinement and model building. Automated non-crystallography symmetry (NCS) and translation-liberation-screw-rotation (TLS) provided in the BUSTER and PHENIX program suits were applied. The final refinement rounds were carried out with the hydrogens in riding position and were omitted for the final deposition. Refinement statistics are summarized in **Table S1**. The figures were generated with PyMOL (Version 1.5, Schrödinger, LLC). The structural homologs search was done with the DALI server [40]. Contact areas between the DUF35 protein and the two-other enzyme were analyzed with the PISA server [41].

7.6. References

1. Guynn, R. W.; Gelberg, H. J.; Veech, R. L., *The Journal of biological chemistry* **1973,** *248* (20), 6957-65.
2. Flamholz, A.; Noor, E.; Bar-Even, A.; Milo, R., *Nucleic Acids Res* **2012,** *40* (Database issue), D770-5.
3. Wu, F.; Minteer, S., *Angewandte Chemie-International Edition* **2015,** *54* (6), 1851-1854.
4. Lindbladh, C.; Rault, M.; Hagglund, C.; Small, W. C.; Mosbach, K.; Bulow, L.; Evans, C.; Srere, P. A., *Biochemistry* **1994,** *33* (39), 11692-11698.
5. Wheeldon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov, P.; Sigman, M., *Nat Chem* **2016,** *8* (4), 299-309.
6. Shatalin, K.; Lebreton, S.; Rault-Leonardon, M.; Vélot, C.; Srere, P. A., *Biochemistry* **1999,** *38* (3), 881-889.
7. Lan, E. I.; Liao, J. C., *Proc Natl Acad Sci U S A* **2012,** *109* (16), 6018-23.
8. Heath, R. J.; Rock, C. O., *Natural product reports* **2002,** *19* (5), 581-96.
9. Miziorko, H. M., *Arch Biochem Biophys* **2011,** *505* (2), 131-43.
10. Lange, B. M.; Rujan, T.; Martin, W.; Croteau, R., *Proc Natl Acad Sci U S A* **2000,** *97* (24), 13172-7.
11. Wagner, T.; Wegner, C. E.; Kahnt, J.; Ermler, U.; Shima, S., *J Bacteriol* **2017**.
12. Huber, H.; Thomm, M.; König, H.; Thies, G.; Stetter, K. O., *Arch Microbiol* **1982,** *132* (1), 47-50.
13. Krishna, S. S.; Aravind, L.; Bakolitsa, C.; Caruthers, J.; Carlton, D.; Miller, M. D.; Abdubek, P.; Astakhova, T.; Axelrod, H. L.; Chiu, H. J.; Clayton, T.; Deller, M. C.; Duan, L.; Feuerhelm, J.; Grant, J. C.; Han, G. W.; Jaroszewski, L.; Jin, K. K.; Klock, H. E.; Knuth, M. W.; Kumar, A.; Marciano, D.; McMullan, D.; Morse, A. T.; Nigoghossian, E.; Okach, L.; Reyes, R.; Rife, C. L.; van den Bedem, H.; Weekes, D.; Xu, Q.; Hodgson, K. O.; Wooley, J.; Elsliger, M. A.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wilson, I. A., *Acta crystallographica. Section F, Structural biology and crystallization communications* **2010,** *66* (Pt 10), 1160-6.
14. Engilberge, S.; Riobé, F.; Di Pietro, S.; Lassalle, L.; Coquelle, N.; Arnaud, C.-A.; Pitrat, D.; Mulatier, J.-C.; Madern, D.; Breyton, C.; Maury, O.; Girard, E., *Chem Sci* **2017,** *8* (9), 5909-5917.
15. Modis, Y.; Wierenga, R. K., *Structure* **1999,** *7* (10), 1279-90.
16. Harijan, R. K.; Kiema, T. R.; Karjalainen, M. P.; Janardan, N.; Murthy, M. R.; Weiss, M. S.; Michels, P. A.; Wierenga, R. K., *The Biochemical journal* **2013,** *455* (1), 119-30.
17. Harijan, R. K.; Mazet, M.; Kiema, T. R.; Bouyssou, G.; Alexson, S. E.; Bergmann, U.; Moreau, P.; Michels, P. A.; Bringaud, F.; Wierenga, R. K., *Proteins* **2016,** *84* (8), 1075-96.
18. Campobasso, N.; Patel, M.; Wilding, I. E.; Kallender, H.; Rosenberg, M.; Gwynn, M. N., *The Journal of biological chemistry* **2004,** *279* (43), 44883-8.
19. Shafqat, N.; Turnbull, A.; Zschocke, J.; Oppermann, U.; Yue, W. W., *Journal of molecular biology* **2010,** *398* (4), 497-506.
20. Kuhner, S.; Wohlbrand, L.; Fritz, I.; Wruck, W.; Hultschig, C.; Hufnagel, P.; Kube, M.; Reinhardt, R.; Rabus, R., *J Bacteriol* **2005,** *187* (4), 1493-503.
21. Leuthner, B.; Heider, J., *Journal of Bacteriology* **2000,** *182* (2), 272-277.
22. Murdoch, R. W.; Hay, A. G., *Microbiol-Sgm* **2013,** *159*, 621-632.
23. VanNice, J. C.; Skaff, D. A.; Wyckoff, G. J.; Miziorko, H. M., *J Bacteriol* **2013,** *195* (17), 3854-62.
24. Steussy, C. N.; Vartia, A. A.; Burgner, J. W., 2nd; Sutherlin, A.; Rodwell, V. W.; Stauffacher, C. V., *Biochemistry* **2005,** *44* (43), 14256-67.
25. Lowe, D. M.; Tubbs, P. K., *The Biochemical journal* **1985,** *227* (2), 591-9.
26. Hou, J.; Feng, B.; Han, J.; Liu, H.; Zhao, D.; Zhou, J.; Xiang, H., *Applied and environmental microbiology* **2013,** *79* (17), 5104-11.
27. Ishikawa, M.; Tsuchiya, D.; Oyama, T.; Tsunaka, Y.; Morikawa, K., *The EMBO journal* **2004,** *23* (14), 2745-54.
28. Smith, S.; Tsai, S. C., *Natural product reports* **2007,** *24* (5), 1041-72.

29. Dueber, J. E.; Wu, G. C.; Malmirchegini, G. R.; Moon, T. S.; Petzold, C. J.; Ullal, A. V.; Prather, K. L.; Keasling, J. D., *Nature biotechnology* **2009,** *27* (8), 753-9.
30. Belay, N.; Sparling, R.; Daniels, L., *Applied and environmental microbiology* **1986,** *52* (5), 1080-5.
31. Wagner, T.; Koch, J.; Ermler, U.; Shima, S., *Science* **2017,** *357* (6352), 699-703.
32. Wagner, T.; Kahnt, J.; Ermler, U.; Shima, S., *Angewandte Chemie* **2016,** *55* (36), 10630-10633.
33. Kabsch, W., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 2), 125-32.
34. Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S., *Acta crystallographica. Section D, Biological crystallography* **2011,** *67* (Pt 4), 235-42.
35. Sheldrick, G. M., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 4), 479-85.
36. Afonine, P. V.; Grosse-Kunstleve, R. W.; Chen, V. B.; Headd, J. J.; Moriarty, N. W.; Richardson, J. S.; Richardson, D. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., *Journal of applied crystallography* **2010,** *43* (Pt 4), 669-676.
37. Cowtan, K., *Acta crystallographica. Section D, Biological crystallography* **2006,** *62* (Pt 9), 1002-11.
38. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., *Acta crystallographica. Section D, Biological crystallography* **2010,** *66* (Pt 4), 486-501.
39. Bricogne, G.; Blanc, E.; Brandl, M.; Flensburg, C.; Keller, P.; Paciorek, W.; Roversi, P.; Sharff, A.; Smart, O.; Vonrhein, C.; Womack, T. *BUSTER version 2.10.3*, Global Phasing Ltd.: Cambridge, United Kingdom, 2016.
40. Holm, L.; Laakso, L. M., *Nucleic Acids Res* **2016,** *44* (W1), W351-5.
41. Krissinel, E.; Henrick, K., *Journal of molecular biology* **2007,** *372* (3), 774-97.
42. Robert, X.; Gouet, P., *Nucleic Acids Res* **2014,** *42* (Web Server issue), W320-4.

7.7. Supplementary Information

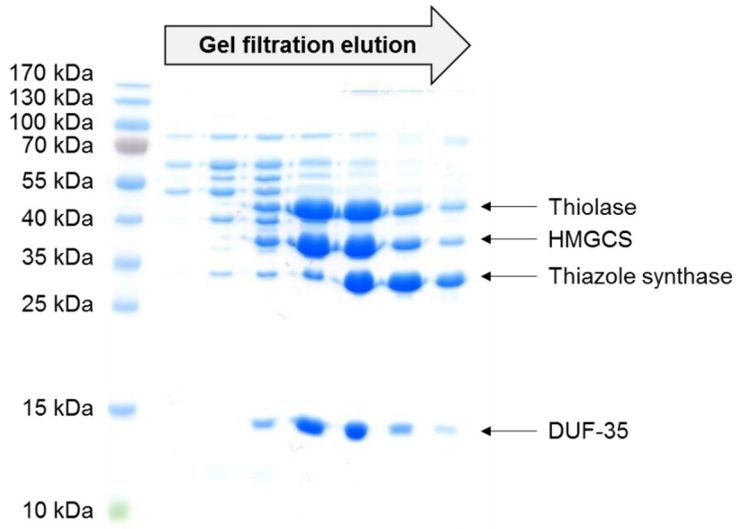


Figure S1. Gel filtration profile of the thiolase/HMGCS complex. The major contaminant thiazole synthase is not co-eluting with the thiolase/HMGCS complex.

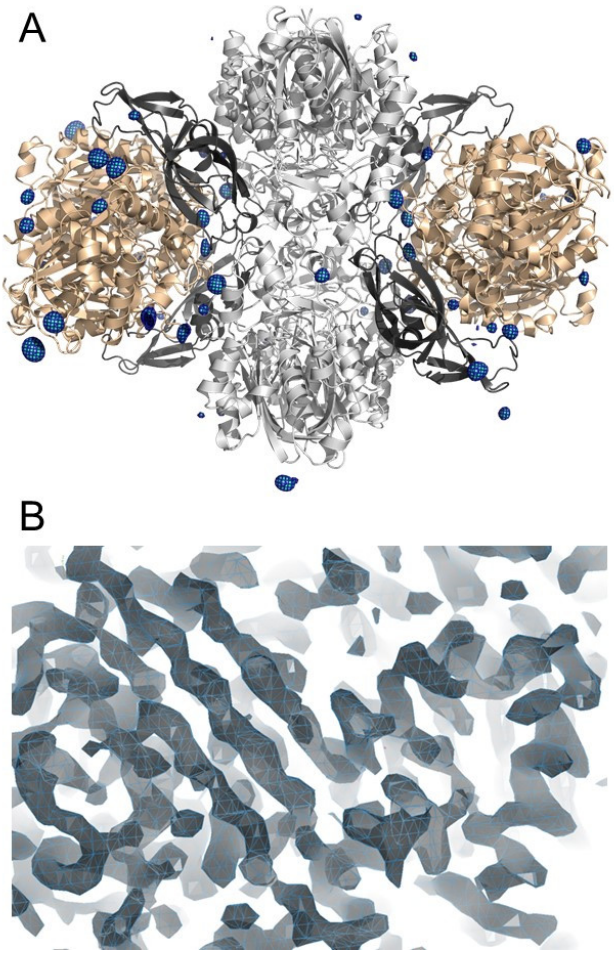


Figure S2. The terbium-binding sites of the thiolase/HMGCS complex and experimental map. (**A**) Terbium localization in the quaternary structure of the thiolase/HMGCS complex obtained from the experimental data of the Tb-Xo4 derivative. The anomalous map contoured at 6.0 σ (blue mesh and cyan surface) indicates the localization of a total of 39 sites. (**B**) A close-up view of an experimental density map contoured at 1.5 σ after solvent flattening and electron density modification.

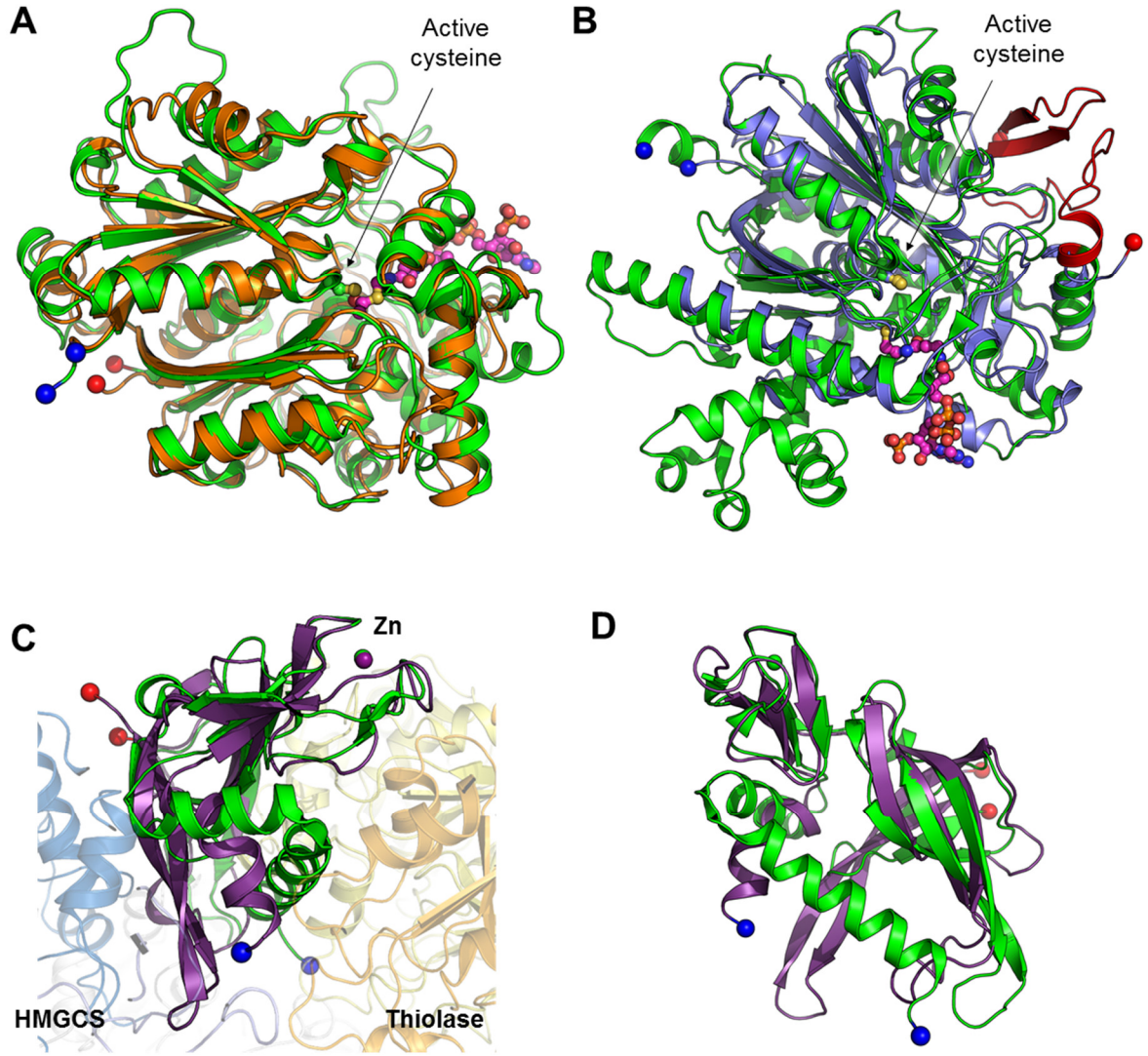


Figure S3. Superposition of the three independent-subunits of the thiolase/HMGCS complex with their closest structural homologs. The N- and C-termini of each models are highlighted in blue and red spheres, respectively. To avoid confusion, the representation contains only one protomer. (**A**) Superposition of the cartoon models of thiolase from *M. thermolithotrophicus* (orange) and the thiolase Scp2 from *Leishmania mexicana* (green, PDB: 3ZBG). The acetyl-CoA, shown in pink comes from the superposition of Scp2 C123A mutant from *L. mexicana* (PDB: 5LOT). (**B**) Superposition of the cartoon models of HMGCS from *M. thermolithotrophicus* (blue) and HMGCS1 from *Homo sapiens* (green, PDB: 2P8U) in complex with CoA. The C-terminal extension contained in lone-standing HMGCS is highlighted in red. (**C**) Superposition of the cartoon models of the DUF35 protein of the thiolase/HMGCS complex from *M. thermolithotrophicus* (purple) and the DUF35 family protein from *Sulfolobus solfataricus* (green, PDB: 3IRB). The thiolase (orange) and HMGCS subunits (blue) from *M. thermolithotrophicus* are shown in transparent cartoon models. (**D**) Same view than C turn at 90°, without the thiolase and HMGCS from *M. thermolithotrophicus*.

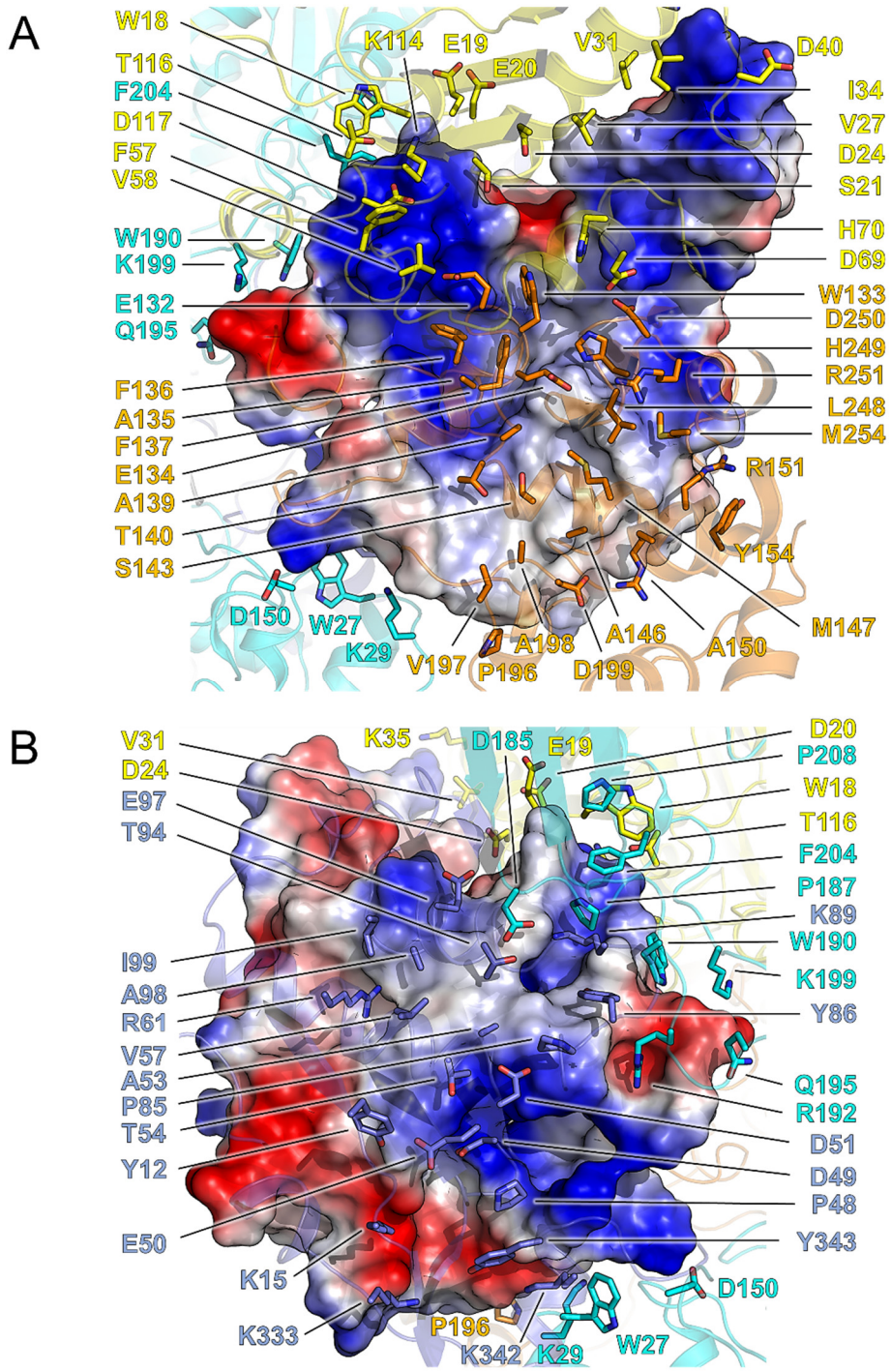
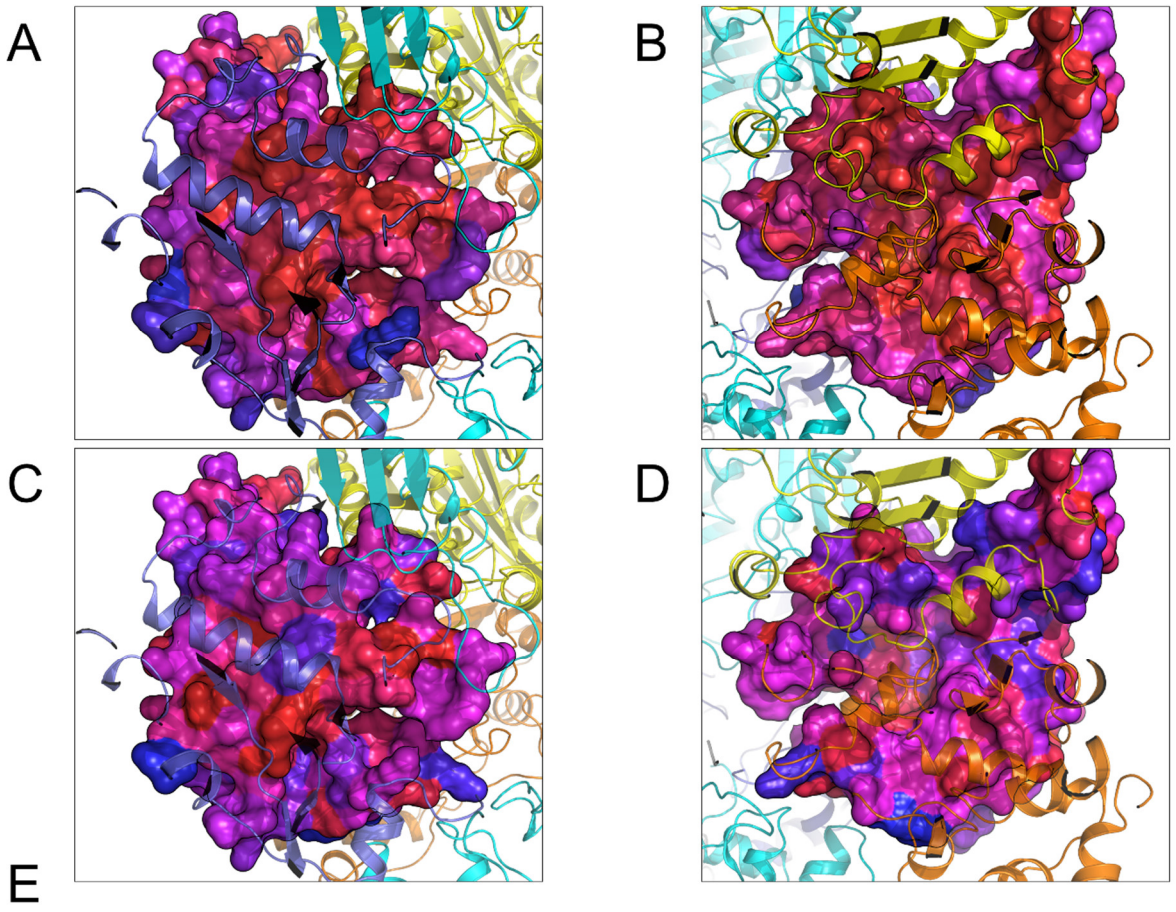


Figure S4. Binding interface area of the DUF35 protein and the enzyme partners in the thiolase/HMGCS complex from *M. thermolithotrophicus*. The electrostatic feature of the surface of DUF35 protein is represented using a color codes with charge gradients from acidic (in red) to basic (in blue). The amino-acid residues of each thiolase protomer at the interface are colored in orange and yellow. The amino-acid residues of each HMGCS protomer at the interface are colored in cyan and blue. (**A**) View from thiolase. (**B**) View from the HMGCS.




```
M_thermolithotrophicus_WP_018154495.1
                                                α1        α2        β1        β2                  β3
                                                          10        20        30        40
M_thermolithotrophicus_WP_018154495.1     .............MVVRSWRHMKERYNLIGTRCKTCGKVYFFSRTVCPDCRRKGELEEFQ
Methanocaldococcus_jannaschii_WP_064496977.1 .............MVVRSWRHIKERYCLIGVRCKNCGTVYFPSREICPKCRRKTEFEEIK
Ignicoccus_hospitalis_KIN4/I_ABU81561.1   ...........MEISPAAVWREKPYRYRLEVRRCPSCKRVFREEVTHCPRCGVET..VRER
Haloferax_volcanii_WP_004040874.1         ITGGEPAGGGAYVSVPTWKRTIPQRHRLVAGECTECGALSFPPEGACADCGSLAGYDDVT
Pyrococcus_furiosus_COM1_AFN03769.1       ........MGKPMQVSRYWRHFREKYRLIGGKCE.NGHLLFPKRPICPICGSR.NIEEFQ
Candidatus_Korarchaeum_cryptofilum_WP_012309820.1 .......MWAYWLSVPAHWRELPARYRLEGGLCRSCGHTMLPKGEVCPICGSK.DIVTKQ
Nitrosopumilus_maritimus_WP_012216013.1   .....MFSRVCKMSIEEQVIENAKNGKLLTHKCTECGYLHLSTAYFCREGGSK.GFEDVL
Sulfolobus_acidocaldarius_WP_011278197.1  ...........MKISPSTSWRLKDTLYRLALGKCKKCGNIKFPYQKICEVCGSL.DVEKIF
Metallosphaera_sedula_WP_012021590.1      ...........MRILPPKIWRNKDTHYSLVASVCEKCGHVNFPARTVCSNCGSF.RVTTKR

M_thermolithotrophicus_WP_018154495.1
                                                β4                  η1                  β5                  β6
                                                50        60                  70        80
M_thermolithotrophicus_WP_018154495.1     LSGKGKIYTYSIVYAP..PKEFNK....LTPYVIAIVELEE...........GPKVTAQ.
Methanocaldococcus_jannaschii_WP_064496977.1 LSGKGKVYTYSVVHVA..PKDFEK....QAPYVIAIIELEE...........GARITGQI
Ignicoccus_hospitalis_KIN4/I_ABU81561.1   LPWRGKLLAWTKICQV..PEGFED....QAPLYVGLVELEN...........GERVVARL
Haloferax_volcanii_WP_004040874.1         LPGTGTVEAATAIGQGGAPPEFVEQQQRAGSFPVAVVAFDGPGGEGGETDERTVSAPAQV
Pyrococcus_furiosus_COM1_AFN03769.1       FSGKGKVITWTIVRNP..PSGFEY....YKPYPLALVQLEE...........GPVVLAQL
Candidatus_Korarchaeum_cryptofilum_WP_012309820.1 LPRRGKIISYSVIWNA..PRGYEY....YTPYIVALIELED...........GTRLMSQL
Nitrosopumilus_maritimus_WP_012216013.1   LDGIGKIATYTIITVP..PAGFEK....YTPYAFVVLEVDNSDLRISGF.MAGIATPA..
Sulfolobus_acidocaldarius_WP_011278197.1  SKGTGSLVDYTILYQP..RDGFEK....QSPMIVGLVRLDE...........GVNIVAPI
Metallosphaera_sedula_WP_012021590.1      LSGKGVLVSYTVSHQV..RTGHEK....ESPVYYGLIRLDE...........GVELVAPL

M_thermolithotrophicus_WP_018154495.1
                                                β7                  β8
                                          90        100       110       120       130
M_thermolithotrophicus_WP_018154495.1     VDCDINKISIGIPVEAAFRRIKEDGKDGIISYGYKFVPITE....
Methanocaldococcus_jannaschii_WP_064496977.1 VDCKPEDVYIGMQVEAVFRRIKEDGDDGVITYGYKFKPIEN....
Ignicoccus_hospitalis_KIN4/I_ABU81561.1   TDVM.EEPKEGSEVQAVLRKVRADGETGLIEYALFFRVIPK....
Haloferax_volcanii_WP_004040874.1         VHSPAGTPAIGDRVEAVPRRIYE..QEGVVRYGFKVVPTEARRE.
Pyrococcus_furiosus_COM1_AFN03769.1       TDVEPEEIKEGMEVEMVTRKIREFDEDGLILYGYKFRPIIKSE..
Candidatus_Korarchaeum_cryptofilum_WP_012309820.1 TDVEPSQVTEGMEVEMVLRKVRVSGESGIIAYAYKFRPLLKK...
Nitrosopumilus_maritimus_WP_012216013.1   .DLPVG.....TPAKITGFDERGIIIEKQ.....
Sulfolobus_acidocaldarius_WP_011278197.1  TDAEPSKLKEGVRVEAVFRRMRNDSYNGLIQYGIKFRVVLDAGNN
Metallosphaera_sedula_WP_012021590.1      VDVDE.EPREGQRVEATIRRLRVDSSNGLIEYGTKFRLVSDGQN.
```

Figure S5. DUF35 sequence conservation. (**A-D**) CONSURF representation of the sequence conservation, using the same view than Figure S4. The conservation is made by a color gradient: from blue, not conserved, to red perfectly conserved. (**A, B**) Conservation among 160 methanogens DUF35. (**C, D**) Conservation among the archaea from the table S3. (**E**) Sequence alignment among the archaea species of table S3, using Espript server [42]. For *Pyrococcus furiosus* one of the highly conserved cysteine part of the Zn(II) finger seems absent.

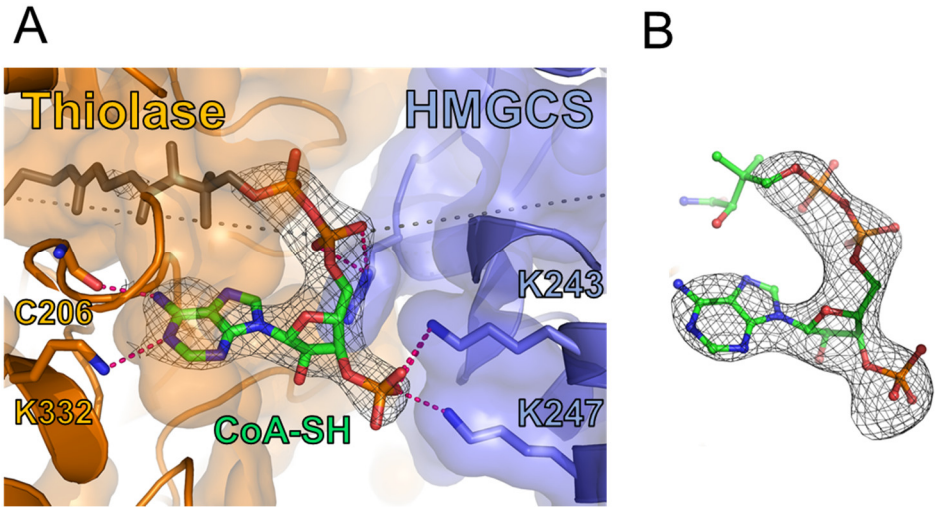


Figure S6. CoA-SH presence at the interface of the thiolase / HMGCS complex. (**A**) Same representation shown in Figure 2B with an electron density map after refinement of the model including CoA-SH. The $2F_o$-Fc electron density map is contoured at 1.0 σ. (**B**) Non-crystallography symmetry applied to the omit map for the CoA-SH. This map corresponds to the signal of the $2F_o$-F_c after averaging of the 4 binding sites in the asymmetric unit and is contoured at 1.0 σ.

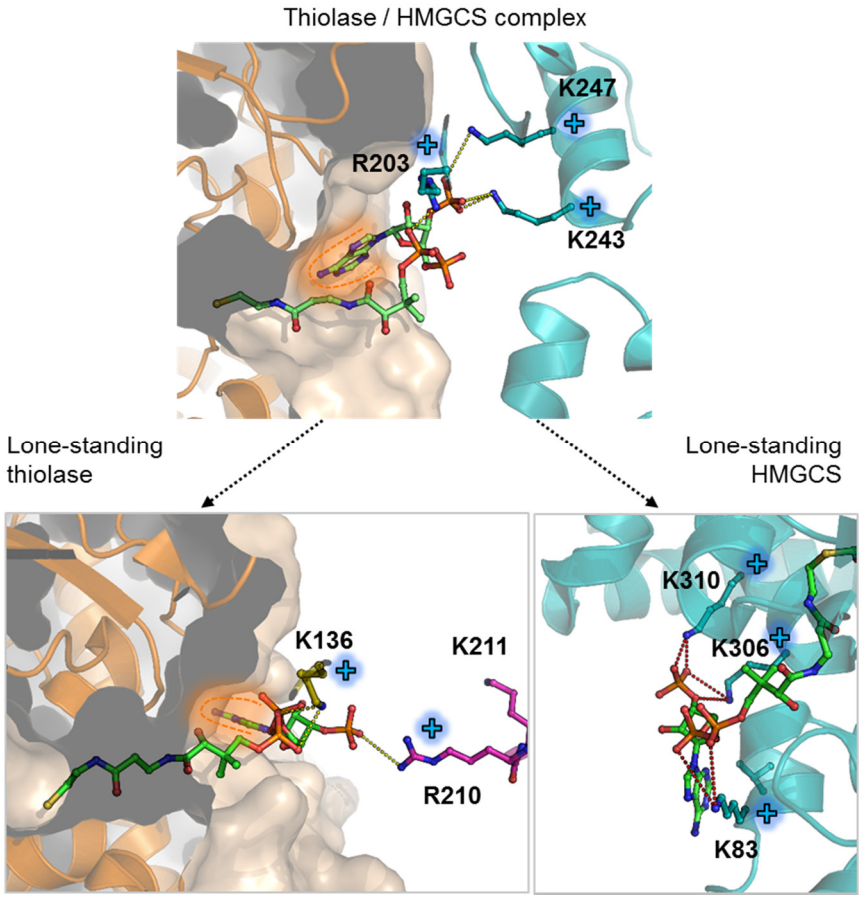


Figure S7. Comparison of the fused-CoA binding site in the thiolase/HMGCS complex from *M. thermolithotrophicus* and the CoA-binding site of the lone-standing thiolase and HMGCS. Top picture represents the CoA-binding site made of the thiolase (orange), and the HMGCS (blue) in the thiolase/HMGCS complex. The two pictures at the bottom represent the CoA-binding site structures of human cytosolic lone-standing thiolase (PDB: 1WL4) (bottom, left) and lone-standing HMGCS from human mitochondria (PDB: 2WYA) (bottom, right). The angle of the three structures are unified by superposition of the lone-standing ones to the thiolase/HMGCS complex. The CoA moiety (carbons in green) and the amino-acid residues interacting with the phosphate groups of CoA (lysine and arginine) are shown in sticks. The positively charged residues forming the electrostatic clamp are highlighted with a "+" symbol, and the cleft for binding of the adenine moiety of CoA is highlighted by a red dashed line.

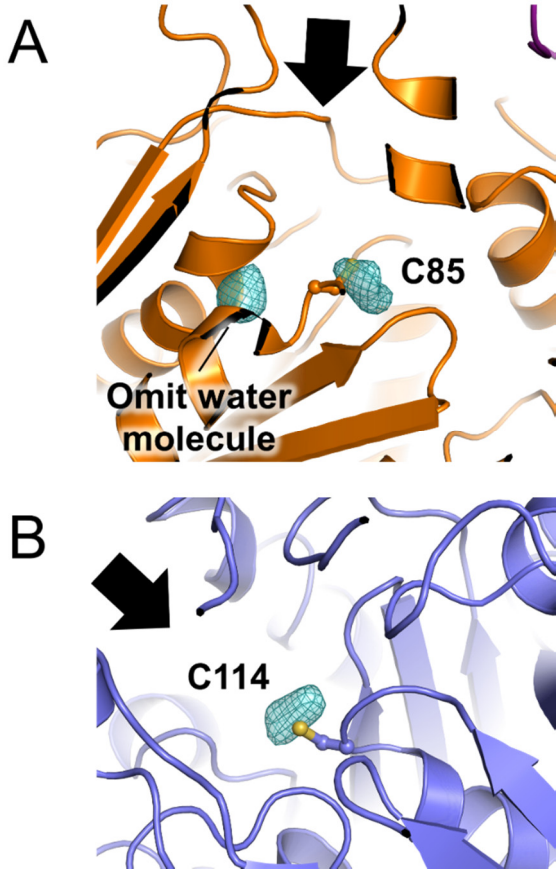


Figure S8. The electron-density map calculated with non-crystallography symmetry (NCS) suggested a partial acetylation of the active cysteines (C85 in the thiolase subunit and C114 in the HMGCS subunit). Since the resolution is borderline to differentiate between an acetylation and an oxidation we cannot exclude the possibility that these extra electron densities come from a partial cysteine oxidation. The NCS map was calculated by the COOT software from the F_o-F_c map and is contoured at 6.5 σ in (**A**) the thiolase (in orange) and (**B**) HMGCS subunits (in blue). In panel *A*, close to the Cys85, a water molecule was omitted during the refinement as a control. The thick black arrows indicate the entrance of the tunnel of the active site.

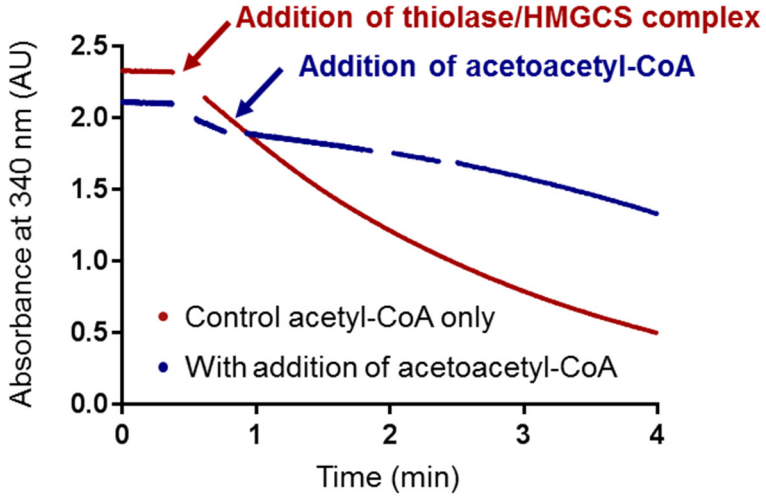


Figure S9. Inhibition of the thiolase/HMGCS complex by external acetoacetyl-CoA. The assay contained 1.6-µM HMGCR, 400-µM NADPH, 1-mM acetyl-CoA in 500-mM potassium phosphate buffer pH 7.5. The reaction was started by adding 45 µg/ml of the thiolase/HMGCS complex (final concentration) at the time indicated with red and blue arrows. Progress of the reaction was monitored spectrophotometrically at 340 nm. In the test, acetoacetyl-CoA (1mM final concentration) was added to the assay at the time indicated by an arrow (blue curve). In a control experiment, acetoacetyl-CoA was not added (red curve). Gaps of the curves come from the injections and mixing events during experiments.

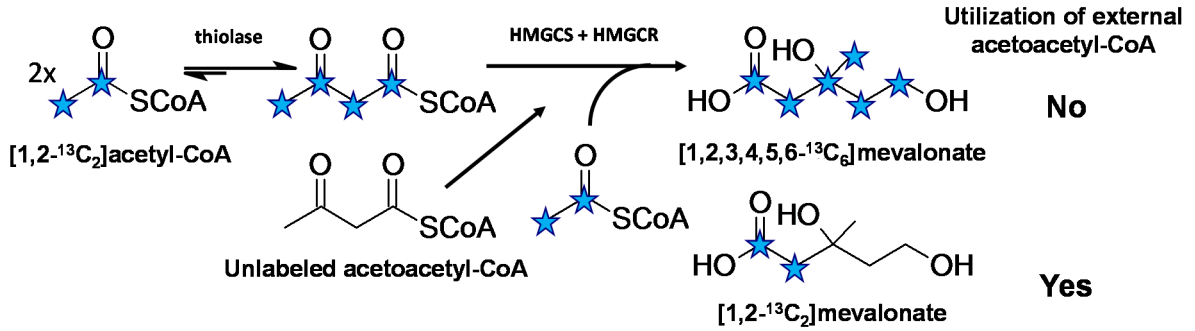


Figure S10. Predicted labeling-patterns of mevalonate produced from [1,2-$^{13}C_2$]acetyl-CoA and non-labeled acetoacetyl-CoA in the reactions catalyzed by the thiolase/HMGCS complex and HMGCR. The assay contained 1.6 µM HMGCR, 400 µM NADPH, 1 mM [1,2-$^{13}C_2$]acetyl-CoA and 1 mM unlabeled acetoacetyl-CoA in 500 mM potassium phosphate buffer pH 7.5. The reaction was started by adding 45 µg/ml of the thiolase/HMGCS complex (final concentration). Progress of the reaction was monitored spectrophotometrically at 340 nm. When 50% of NADPH was consumed, the reaction was quenched with 5% formic acid (final concentration) and centrifuged for 5 min at 17,000 × g to remove precipitated proteins and then the supernatant was analyzed by LC-MS (See Method section in the main text). If external unlabeled-acetoacetyl-CoA was not used by the thiolase/HMGCS complex, only [1,2,3,4,5,6-^{13}C6]mevalonate is produced. If external unlabeled-acetoacetyl-CoA was used by the thiolase/HMGCS complex, [1,2-$^{13}C_2$]mevalonate is additionally produced. In the latter case, the external acetoacetyl-CoA could be transformed to unlabeled acetyl-CoA by the thiolase reverse-reaction and then the unlabeled acetyl-CoA could then be condensed with the unlabeled-acetoacetyl-CoA, which yield unlabeled mevalonate.

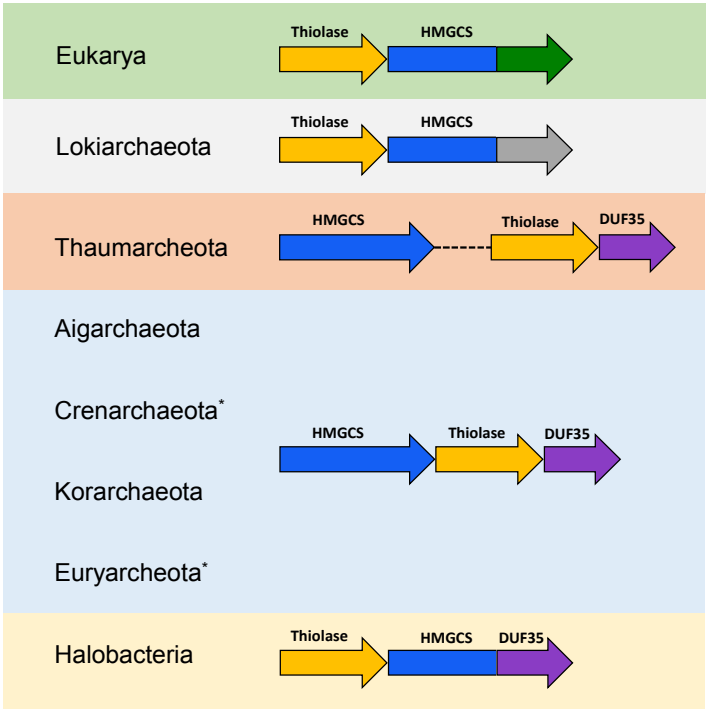


Figure S11. The thiolase/HMGCS complex operon is widespread among archaea. All three proteins, thiolase (yellow), HMGCS (blue) and the DUF35 (purple) are conserved in archaea with only one exception, Lokiarchaeota. HMGCS from Lokiarchaeota (gray background) and Eukarya (green background) contain a C-terminal extension (highlighted in green and gray, respectively) which are not conserved each other. The genes encoding the three proteins are organized in a conserved operon in most archaea (light blue background) with the exception of Thaumarchaeota and *Halobacteria*. In Thaumarchaeota (light orange background) HMGCS is encoded in a different region of the chromosome. In *Halobacteria* (yellow background) the DUF35 protein is fused to the C-terminus of HMGCS. *In *Caldivirga* and *Staphylothermus* in Crenarcheota and *Methanobacterium* in Euryarchaeota, the three genes encoding HMGCS, thiolase and the DUF35 proteins are separated in the genome.

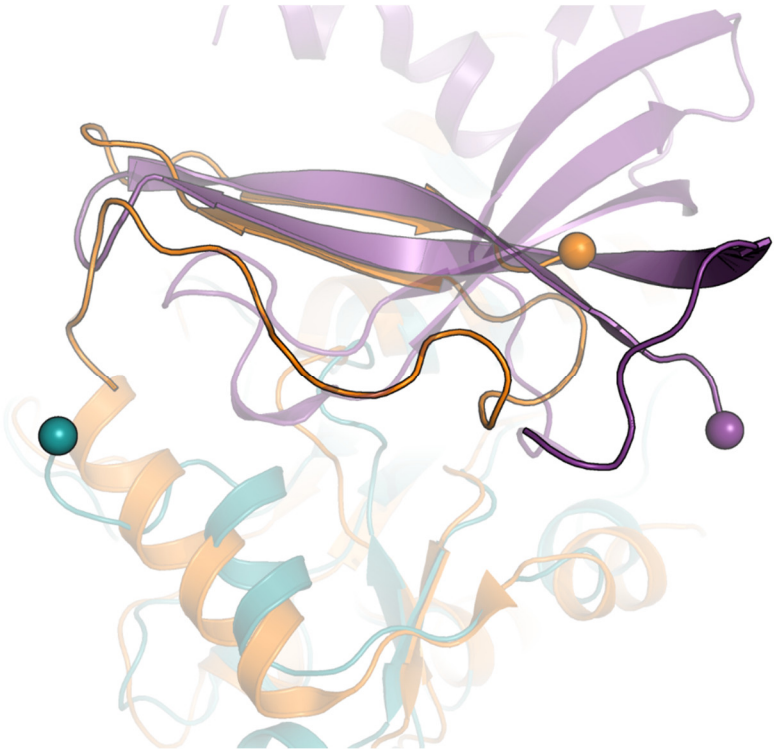


Figure S12. Structural similarity of the C-terminal motifs of the lone-standing HMGCS and the DUF35 protein of the thiolase/HMGCS complex. The human HMGCS1 (PDB: 2P8U) in orange is superposed to the archaeal HMGCS from *M. thermolithotrophicus* shown in blue. The DUF35 of the thiolase/HMGCS complex is colored in purple. The C-terminal extension of the human HMGCS (residues 434-470) and the DUF35 (residues 94-129) are highlighted in non-transparent compared to the rest of the sequence. The spheres indicate the C-terminal amino acid.

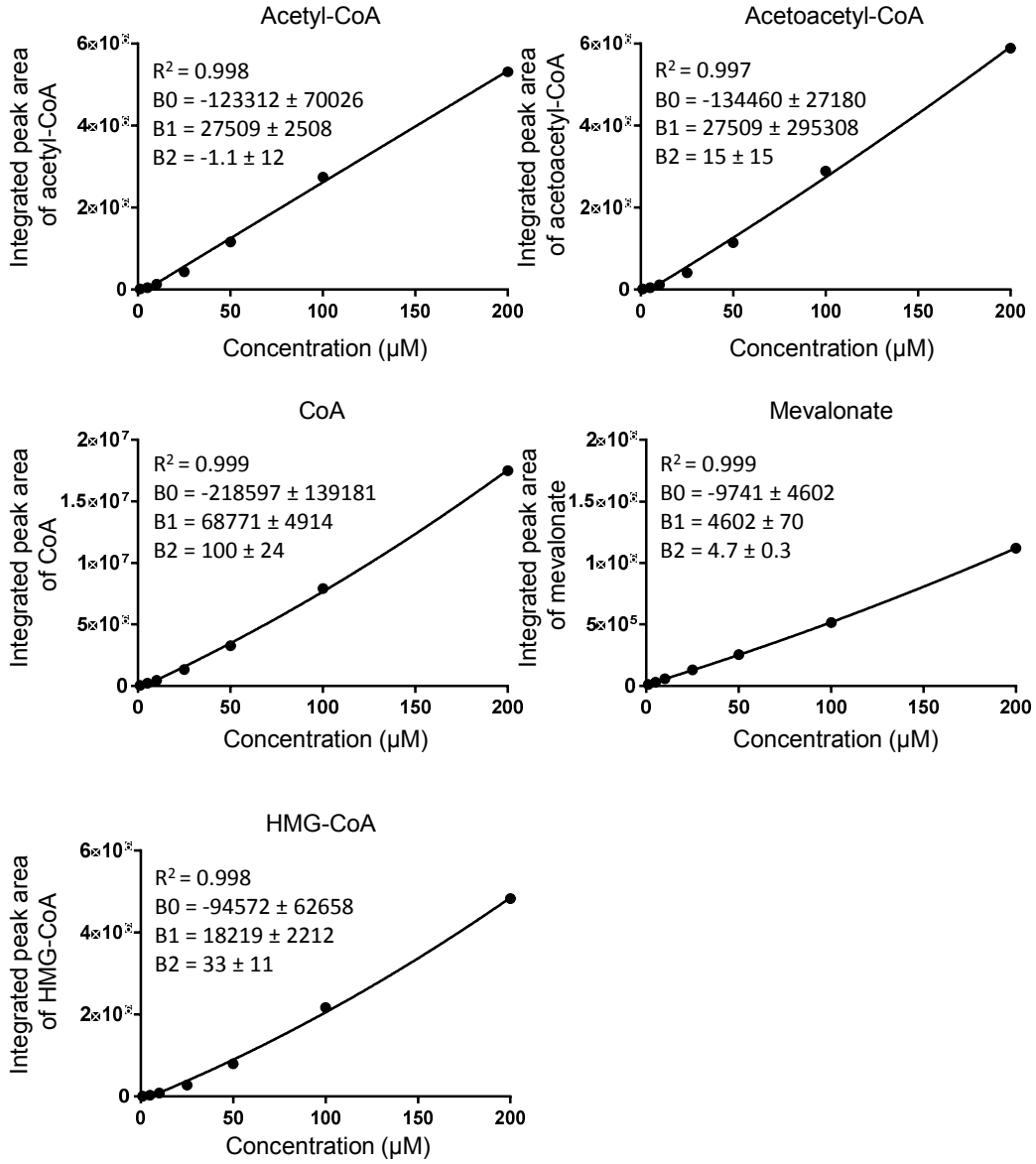


Figure S13. Standard curves used for HPLC-MS quantification of assay substrates, intermediates and products. Standard curves were fitted with a quadratic regression for all compounds.

Table S1. X-ray analysis statistics.

	Thiolase/HMGCS complex with Tb-Xo4 (SAD)	Thiolase/HMGCS complex native	Thiolase/HMGCS soaked with acetyl-CoA
Data collection			
Synchrotron	SOLEIL, Proxima 2A	ESRF, ID29	ESRF, ID23-1
Wavelength (Å)	1.64851	1.38546	1.64861
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution (Å)	49.20 – 3.29 (3.47 – 3.29)	48.90 – 2.75 (2.90 – 2.75)	49.30 – 2.95 (3.11 – 2.95)
Cell dimensions			
a, b, c (Å)	107.4, 145.1, 230.9	106.6, 144.0, 230.6	107.6, 145.2, 230.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
R_{merge} (%)[a]	11.4 (63.6)	27.9 (124.9)	14.3 (202.4)
R_{pim} (%)[a]	4.6 (27.2)	7.9 (35.4)	4.0 (57.7)
$CC_{1/2}$ (%)[a]	99.8 (91.9)	99.3 (32.4)	99.9 (48.1)
I/σ_I[a]	18.8 (4.4)	8.6 (2.1)	14.9 (1.3)
Completeness (%)[a]	99.7 (98.4)	100.0 (100.0)	99.4 (98.8)
Redundancy[a]	13.4 (12.3)	13.4 (13.3)	13.3 (13.1)
Number of unique reflections[a]	55597 (7919)	92965 (13426)	76363 (10942)
Refinement			
Resolution (Å)		48.39 – 2.75	49.29 – 2.95
Number of reflections		92874	76299
R_{work}/R_{free}[b] (%)		19.0 / 22.5	18.2 / 23.0
Number of atoms			
Protein		26314	26334
Ligands/ions		146	172
Solvent		471	61
Mean B-value (Å^2)		52.62	106.71
Molprobity clash score, all atoms		2.82 (100th percentile)	4.68 (100th percentile)
Ramachandran plot			
Favoured regions (%)		97.01	97.07
Outlier regions (%)		0.23	0.29
rmsd[c] bond lengths (Å)		0.008	0.005
rmsd[c] bond angles (°)		0.825	0.58
PDB code		6ET9	6ESQ

[a] Values relative to the highest resolution shell are within parentheses. [b] R_{free} was calculated as the R_{work} for 5% of the reflections that were not included in the refinement. [c] rmsd, root mean square deviation.

Table S3. Examples of HMGCS complexes found in archaea. Representative organisms from each phyla were picked and searched for the DUF35 protein in the Integrated Microbial Genomes system (IMG). Subsequently they were checked for the HMGCS and thiolase (abbreviated Thio). GenBank accession code is given for all the proteins, if available, otherwise the locus tag from JGI IMG is given.

Phylum	Class	Organism	Thiolase	DUF35	HMGCS	Cluster architecture
Euryarchaeota	**Halobacteria**	*Haloferax volcanii* DS2	WP_004040873.1	WP_004040874.1	WP_004040874.1	Thio-HMGCS DUF35 (HMGCS and DUF35 are fused)
			YP_003534895	YP_003534894		Thio-DUF35
	Methanococci	*Methanocaldococcus jannaschii* DSM 2661	NP_248557	NP_248560	NP_248554	TR-HMGCS-Thio-HP-HP-DUF35 (TR=transcript. regulator; HP=hypothet. protein)
	Thermococci	*Pyrococcus furiosus* COM1	AFN03768.1	AFN03769.1	AFN03767.1	HMGCS-Thio-DUF35
Korarchaeota		*Korarchaeum cryptofilum* OPF8	YP_001737605	YP_001737606	YP_001737604	HMGCS-Thio-DUF35
Thaumarchaeota		*Nitrosopumilus maritimus* SCM1	WP_012216014.1	WP_012216013.1		Thio-DUF35
Crenarchaeota	**Thermoprotei**	*Sulfolobus acidocaldarius* DG1	B6A19_RS06605	B6A19_RS06600	B6A19_RS06610	HMGCS-Thio-DUF35
		Metallosphaera sedula ARS120-1	Ga0100896_112014	Ga0100896_112015	Ga0100896_112013	HMGCS-Thio-DUF35-HMGCR
			Ga0100896_111636	Ga0100896_111637		Thio-DUF35
			Ga0100943_11682	Ga0100943_11683		Thio-DUF35-phAC-CoA ligase-3OHacyl-CoADH
		Ignicoccus hospitalis KIN4	Igni_0377	Igni_0378	Igni_0376	DUF84-HMGCS-Thio-DUF35

Table S4. Examples of the predicted thiolase/HMGCS complexes found in bacteria. Completed bacterial genomes from the Integrated Microbial Genomes system (IMG) were searched for the DUF35 protein. Subsequently the presence of HMGCS, thiolase (abbreviated as Thio) and HMGCR were analyzed. GenBank accession code is given for all the proteins, if available, otherwise the locus tag from JGI IMG is given.

Phylum	Organism	Thiolase	DUF35	HMGCS	HMGR	Cluster architecture
Chloroflexi	*Roseiflexus sp.* RS-1	YP_001274810	YP_001274809	YP_001274811	YP_001274812	HMGR-HMGCS-Thio-DUF35
	Anaerolinea thermophila UNI-1	YP_004175506	YP_004175505	YP_004175507	YP_004175508	HMGR-HMGCS-Thio-DUF35
	Chloroflexus aurantiacus J-10-fl	YP_001633876	YP_001633874	YP_001633877	YP_001633878	HMGR-HMGCS-Thio-duf-DUF35
	Herpetosiphon aurantiacus DSM 785	ABX06504.1	ABX06505.1	ABX06503.1	not in same cluster	HMGCS-Thio-DUF35
Actinobacteria	*Streptomyces albus* DSM 41398	Ga0077996_113376/ Ga0077996_113374	Ga0077996_113375	Ga0077996_113375	not in same cluster	Thio-HMGCS DUF35-Thio (HMGCS and DUF35 fused)
Firmicutes	*Streptococcus mutans* NN2025	YP_003485695	YP_003485694	YP_003485002	YP_003485003	Thio- DUF35 ……HMGCS-HMGR
	Desulfosporosinus orientis DSM765	AET68488.1	AET68487.1	AET68487.1	not in same cluster	HMGCS DUF35-Thio (HMGCS and DUF35 fused)
	Desulfonomaculum kuznetsovii DSM6115	AEG13966.1	AEG13968.1	AEG13967.1	not in same cluster	Thio-HMGCS- DUF35
	Clostridium autoethanogenum DSM 10061	AGY77788.1	AGY77789.1	AGY77787.1	not in same cluster	HMGCS-Thio- DUF35
	Carboxydothermus hydrogenoformans Z-2901	ABB16153.1	ABB15728.1	ABB15916.1	not in same cluster	Thio-HMGCS- DUF35
Proteobacteria	*Desulfococcus oleovorans* Hdx3	ABW67355.1	ABW67354.1	ABW67354.1	not in same cluster	HMGCSDUF35SDR-Thio (HMGS fused to DUF35 and SDR)

8. Conclusions and General Discussion

Enzymes are able to catalyze a myriad of chemical reactions with high regio- and stereo-specificity under very mild reaction conditions and in the context of the whole cellular metabolism (more than 2700 different metabolites). To achieve this remarkable specificity enzymes have evolved various strategies to i) control reactive intermediates within an enzymes active site by both lowering the activation energies along the desired reaction pathway (´positive catalysis´) and by increasing the activation energies of side reactions (´negative catalysis´) and to ii) control the transfer of reactive intermediates between enzyme active sites (´substrate channeling´).

In the first part of this thesis we looked at NAD(P)H dependent enoyl-thioester reductases from both the short-chain reductase/dehydrogenase (SDR) and the medium-chain reductase/dehydrogenase (MDR) superfamilies. We were able to dissect the catalytic cycle of these enzymes in two distinct steps, the hydride transfer and the proton transfer. We identified a covalent C2-adduct as an intermediate that is part of the reaction coordinate or lays at least very close to it. The detailed implications of this detected intermediate for our understanding of the reaction mechanism of enoyl-thioester reductases and it´s use as a molecular probe are discussed in chapter 8.1 below. We were also able to identify the function of multiple residues in the active site of these reductases and showed that some of them mainly destabilize transition states of unwanted side reactions (i.e. have a negative catalytic function). The concept of such negative catalysis as a key principle in enzyme catalysis and the control of reactive intermediates is discussed in detail in chapter 8.2.

In the second part of this thesis we then have a close look at two examples of substrate channeling between enzyme active sites of a metabolic pathway. The general strategies nature has evolved to control the flux of intermediates between enzymes are discussed in chapter 8.3. In the outlook (chapter 8.4) we then outline possible ways to build artificial substrate channeling.

8.1. C2-adducts in enoyl-thioester reductases

The detection of covalent C2-adducts between the nicotinamide cofactor and the enoyl-thioester substrate in zink-free enzymes from the both the MDR and SDR superfamily as well as in the biosynthesis of the natural product sanguinarine[1] and in chemical model reactions using modified dihydropyridine substrates[2] suggests that such intermediates are a more general phenomenon than so far assumed and may even be a general catalytic principle at the active site of oxidoreductases. Interestingly, there is also a report of a transient intermediate formed during the catalytic cycle of the zinc dependent horse liver alcohol dehydrogenase with the model substrate 3-hydroxy-4-nitrobenzaldehyde, which shows an absorption maximum around 370 nm [3], which is similar to the C2-adducts we described.

It is not yet clear through what mechanism the C2 adduct is formed and whether it is on the reaction coordinate (RC) or just energetically very close to it. One possibility is that the hydride transfer from NAD(P)H to the β-carbon of the substrate proceeds through a pericyclic transition state following an ene-mechanism as proposed over 50 years ago by Hamilton[4]. This would put the C2 adduct on the RC and challenge the canonical mechanism of direct hydride transfer used to describe NAD(P)H dependent reduction reactions of oxidoreductases. The formed C2 adduct would then decay to a reactive enolate, which can attack the resolving electrophile (a proton in reductases or CO_2 in reductive carboxylases), either in a concerted mechanism or in successive steps.

Alternatively, the hydride is transferred in a more canonical mechanism to the β-carbon forming the reactive enolate directly. The enolate could then either attack the resolving electrophile or instead re-attack the nicotinamide ring in a Michael type addition reaction to form the C2 adduct. In this case the

C2 adduct is just a catalytically competent side product formed during the reaction, which can successively be transformed into the products. Note that the lack of good electron acceptors in proximity to the Cα of the enolate would favor such a Michael addition, as for instance is the case in ECRs when CO_2 is omitted [5], Etr1p when the proton donor is mutated (Chapter 2), or in InhA, where no protic group is positioned well for protonation (Chapter 4). The enzymes additionally still need to actively prevent the formation of the toxic C4 side product (Chapter 3). In reductive carboxylases the C2 adduct could serve an additional purpose of storing the reactive enolate until CO_2 is positioned correctly for carboxylation, thus preventing the reactive enolate from prematurely reacting with a proton instead. Further experiments are needed to decipher the precise reaction mechanism and the role of the C2 adduct during catalysis (see Outlook).

Independent of the fundamental question of how the hydride is transferred, the C2 adduct has some interesting applications. Because it is a competent substrate for all enoyl-thioester reductases tested so far and represents an intermediate state in catalysis, where the hydride transfer has already occurred, it is a great tool to study the second part of the catalytic cycle of these enzymes (whether it's the protonation or the carboxylation). We demonstrated this by successfully changing the protonation specificity of the enoyl-thioester reductase Etr1p from *re*- to *si*- face (Chapter 2). It additionally allowed us to assign the function of individual amino acids in Etr1p as well as in InhA. Due to its close proximity to the catalytic transition state it also poses an interesting starting point for the design of mechanistic inhibitors. The fact that the C4 adduct observed in the Y79F variant of Etr1p is a nano molar inhibitor for the wild type enzyme highlights this potential. The design of a substrate analogue that shifts the formation of the C2 adduct towards the C4 adduct could therefore already act as such an inhibitor. Introduction of functional groups that stabilize the C2 or C4 adduct and prevent its uncatalyzed decay in solution could additionally help in the design of such inhibitors. Finasteride, a mechanistic inhibitor of 5α-reductase used to treat benign prostatic hyperplasia and hair loss in men, uses both of these principles and has been shown to form a covalent intermediate with NADPH. It is a structural analogue to testosterone, the natural substrate of 5α-reductase, but the enoyl-double bond is shifted (most likely positioning the enolate better for Michael addition to the cofactor) and an amide group is introduced in the molecule (which stabilizes the intermediate and prevents it from fast decay). It therefore represents a blue print for the development of further NADPH oxidoreductase inhibitors that are interesting drug targets (for a list of potential candidates see Chen et al. [6]).

The detailed understanding of enoyl-thioester reductases and reductive carboxylases then allowed us to use them for the up scaled synthesis of a large library of malonyl-CoA derivatives, which can be used as atypical extender units in polyketide synthases (Chapter 5). Interestingly our approach included a metabolic proofreading step. Another strategy nature often uses to deal with side reaction of enzymes that can occur when the mechanisms to control reaction selectivity at an enzymes active site are unable to prevent the formation of side products. These mechanisms to control reactive intermediates are discussed in detail in the following chapter.

8.2 'Negative' and 'positive catalysis': Complementary principles that shape the catalytic landscape of enzymes

Authors:

Bastian Vögeli[1] and Tobias J. Erb[1,2]

Submitted to:
Current Opinion in Chemical Biology

Author contributions:
B.V. and T.J.E. wrote the paper.

8.2. 'Negative' and 'positive catalysis': Complementary principles that shape the catalytic landscape of enzymes

8.2.1. Abstract

Our understanding of enzyme catalysis is dominated by transition state theory. According to this concept, an enzymatic reaction is guided along a desired reaction coordinate through the stabilization of favorable transition state. But how much is the outcome of an enzyme reaction controlled by the destabilization of unwanted transition states? Here, we revive and critically review the hypothesis that the active site of enzymes also features elements of 'negative catalysis'. We provide examples that show that enzyme catalysis can be achieved by the combined action of positive and negative constraints at the active site of an enzyme. This integrated view of enzyme catalysis has direct consequences for our studies on the catalytic landscape of enzymes, as well as current efforts in enzyme engineering and the *de novo*-design of enzymes.

8.2.2. Introduction

Enzymes are remarkable catalysts that can achieve rate accelerations of up to nineteen orders of magnitude at high regio-, stereo- and reaction specificity compared to the uncatalyzed reaction [7]. The concept of enzyme catalysis as formulated by Pauling states that enzymes accelerate reaction rates by binding transition states better than substrates thereby lowering the activation energy of the reaction [8]. Pauling's original concept was expanded and further refined, by including additional factors that also contribute to the lowering of activation energies, such as ground state destabilization, conformational substrate stabilization and enzyme preorganization [9-12]. Besides catalyzing a given reaction enzymes are very efficient at preventing alternative reaction outcomes, which would be favored in solution or in the gas phase over the desired reaction. This enables enzymes to catalyze chemically 'difficult' or 'improbable' reactions involving highly reactive intermediates [13]. The exquisite control of intermediates over the course of a reaction is a key feature of enzyme catalysis.

8.2.3. The catalytic landscape of enzymes is multi-dimensional and shaped by positive and negative contributions

The progress of enzyme reactions is typically visualized in two-dimensional plots, the free energy along the reaction coordinate (RC). In these graphs, transition states are depicted as energy maxima along the reaction path (Figure 1B). Although used very often to describe enzyme reactions, the two-dimensional representation of catalysis is rather misleading. Transition states actually represent local energy minima along the chosen RC when compared to the non-catalyzed reaction or alternative reaction outcomes (*i.e.*, side reactions). Conceptually it is thus more correct (and more helpful) to think of enzyme catalysis in a multi-dimensional landscape, where transition states represent saddle-points along different possible reaction outcomes (Figure 1C).

How do enzymes control their catalytic landscape to form the correct reaction product? In principle the outcome of a reaction can be determined through two different mechanisms: Enzymes could guide a reaction along a minimum energy path by lowering the energy of productive transition states that lead to formation of the wanted product, as formulated by Pauling (Figure 1D). Alternatively, enzymes could increase the energy of competing transition states that would lead to the formation of alternative reaction products (Figure 1E). Even though the final product of the reaction would be the same, it is a conceptual, as well as a mechanistic difference whether a preferred reaction is promoted, or whether competing side reactions are suppressed by an enzyme during catalysis.

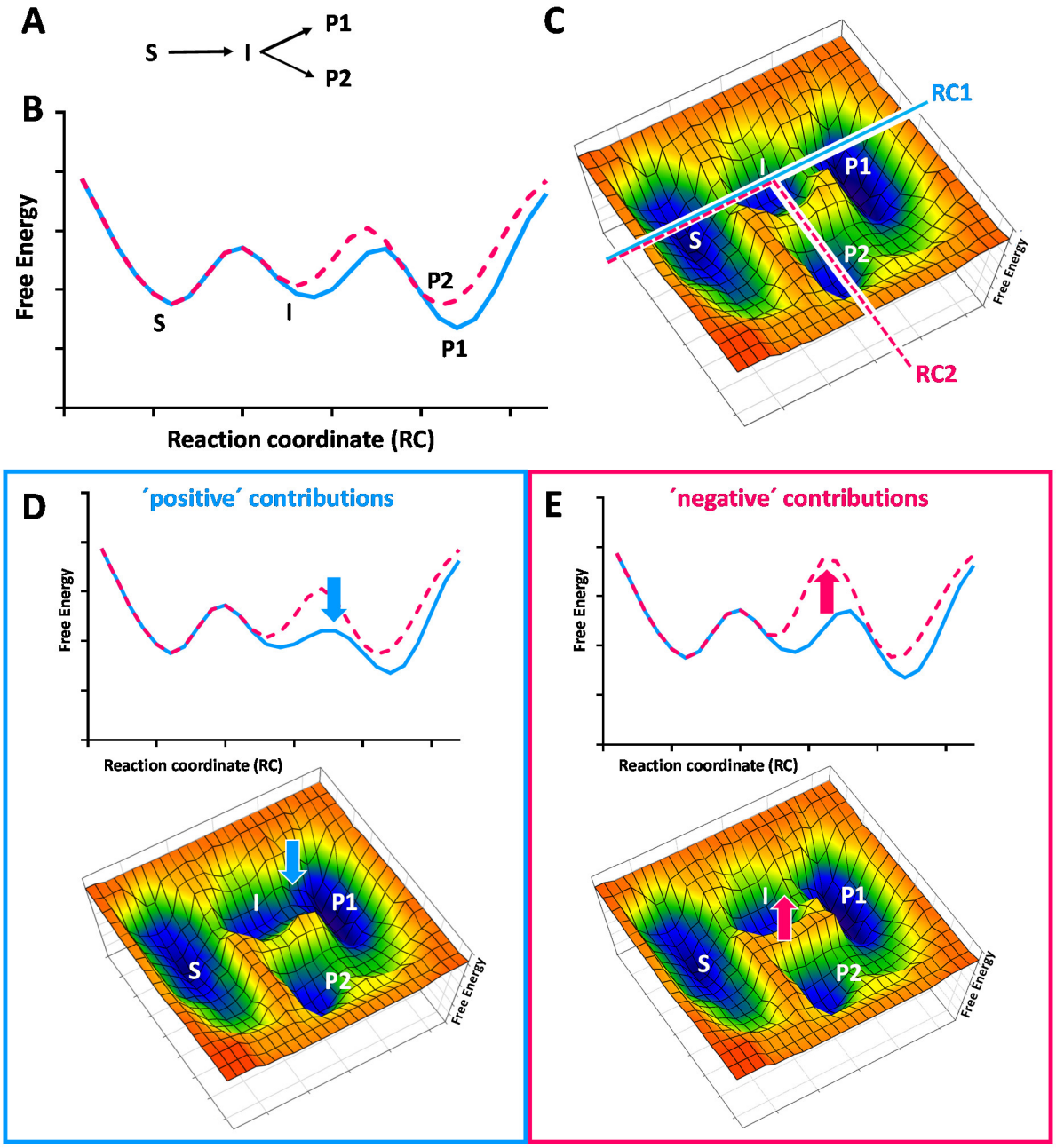


Figure 1 Visualization of a chemical reaction. A) Simple reaction scheme depicting the possible outcomes of a chemical reaction. **B)** Two dimensional free energy plot along the reaction coordinate (RC). RC1 (light blue) shows the RC for the formation of the desired product P1. RC2 (pink) shows the RC for the formation of the side product P2. Transition states are shown as local maxima. **C)** Three dimensional free energy landscape showing the transition states as saddle points along the RCs. **D)** ´Positive´ contributions that stabilization of the productive transition states lead to increased formation of the desired product (P1). **E)** ´Negative´contributions that destabilize the unwanted transition states also lead to the increased formation of the desired product (P1) compared to the side product (P2)

In the past, most of the research focused on understanding the principles that are used by active site residues to promote catalysis. In contrast, much less is understood about the mechanisms that suppress the formation of competing transition states leading to unwanted side reactions, even though a concept of 'negative catalysis' was developed by Rétey almost 30 years ago [13]. According to Rétey the suppression of unwanted side reactions ('negative catalysis') can be equally important as the promotion of the desired reaction ('positive catalysis').

8.2.4. A simple way to visualize positive and negative contributions of active site residues to catalysis

How can positive and negative contributions of active site amino acids to catalysis be assessed? In principle, the role of a given amino acid can be described in two dimensions. One dimension is the contribution of an active site residue to the apparent turnover frequency of an enzyme (k_{cat}), while the other is the contribution of an active site residue to reaction specificity, expressed as the relative ratio of side product (*P2*) to reaction product (*P1*), expressed as *log[P2/P1]*.

In a first approximation residues conferring positive contributions would mainly affect the k_{cat} of an enzyme (measured as rate of substrate consumption or formation of all products). Residues conferring negative contributions on the other hand would mainly affect *log[P2/P1].* Note that positive and negative contributions are of course not completely independent (e.g., an increase in k_{cat} for the desired reaction would also decrease the amount of side product formed). Yet, plotting the effect of an active site mutation onto both dimensions (k_{cat} versus *log[P2/P1])* in an activity/specificity-graph allows for an intuitive way to visualize and understand its contribution to catalysis, as exemplified below.

In principle, the activity/specificity-graph is an intriguingly simple way to assess the catalytic functions of an active site residue of an enzyme. However, in practice, while the turnover frequencies of an enzyme and its active site variants are usually well documented, the relative ratios of side product to product for the same set of enzyme variants are very often not reported. In many cases the side reactions of enzymes remain unknown, because they are not systematically identified, hard to measure (due to their low occurrence in the WT), or simply not considered in the study design and thus not investigated, let alone quantified. It is mainly this lack of data that prevents the possibility to assess the different contributions of active site amino acids to catalysis more systematically. For this review, we have surveyed the literature to identify studies that explicitly report the function of amino acids in both catalytic dimensions (Figure 2).

8.2.5. A growing body of evidence for the existence of 'negative catalysis'

Mutation of conserved active site residues very often affect the catalytic rate of an enzyme, meaning that these residues serve in promoting the catalyzed reaction. But how much do they additionally contribute to controlling the formation of the correct reaction product, *i.e.*, how big is there contribution on the *log[P2/P1]* coordinate? And even more extreme: Are there dedicated amino acid residues at the active site of enzymes that only exist to suppress the formation of side products without affecting the overall catalytic rate of an enzyme? In other words: Are there amino acids that almost exclusively operate in 'negative catalysis'? In the k_{cat} versus *log[P2/P1]* plot of our selected examples, such residues should be shifted left on the x-axis, while appearing mostly unchanged on the y-axis. In the following, we will discuss several examples of amino acids that guide and control the outcome of catalysis in different enzymes by 'negative catalysis' and discuss their mode of action.

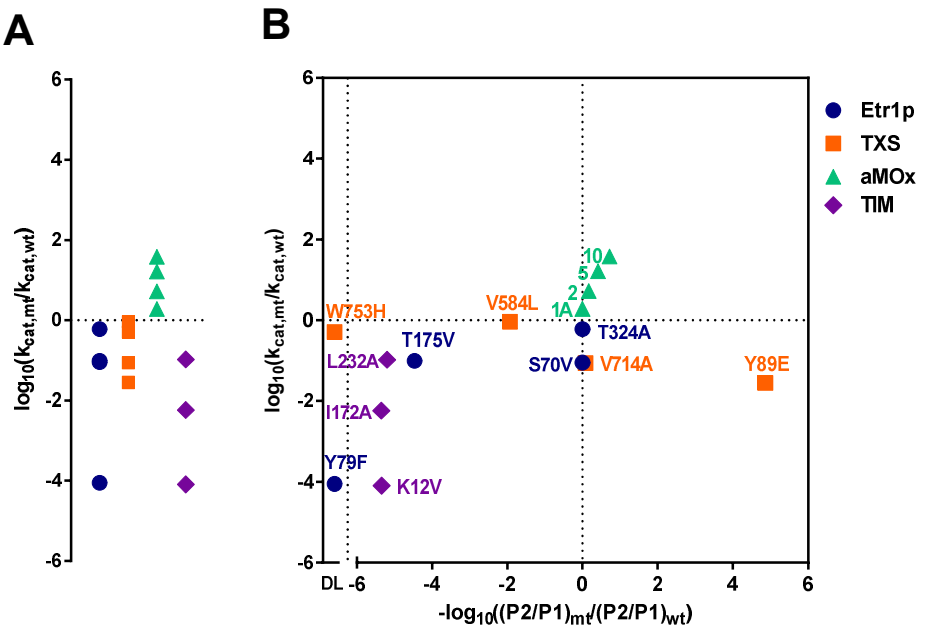


Figure 2 Relative *log(k_{cat})* vs. *log[P2/P1]* plot for active site mutants of selected enzymes. A) Conventional one dimensional plot of k_{cat} for active site mutants of Etr1p, TXS, aMOx and TIM. **B)** *log(k_{cat})* vs. *log[P2/P1]* plot of the same variants highlighting the additional dimension/information that can be gained by including a measure for side product formation/reaction selectivity of the enzyme. DL; product P1 below the detection limit.

Enoyl-thioester reductase Etr1p. Etr1p operates in mitochondrial fatty acid biosynthesis and reduces enoyl-CoA thioestesters into their saturated counterparts at the expense of NADPH. In the enzyme a conserved threonine 175 was proposed to interact with the NADPH cofactor [14]. When this threonine is converted into an isosteric valine, the catalytic rate of the enzyme is decreased 'only' by a factor of ten. The error rate of the enzyme, however, becomes dramatically increased by this mutation. The T175V variant forms a covalent adduct between the NADPH cofactor and the enoyl-CoA substrate ('C4-adduct') as a side product at high rates. While the WT enzyme does not form the side product above the detection limit (less than one per $1.7 \cdot 10^{-7}$ turnovers), the side product is formed in the T175V mutant approximately every third turnover. In other words, the T175V variant still operates at 10% WT turnover frequency, but its error rate increased by almost six orders of magnitude (Figure 2B) [15]. This strong increase in side product formation by the T175V mutation can be explained through a lowered energetic barrier to a competing transition state leading to formation of the side product. In summary, T175 probably represents one of the most striking examples of an active site residue that almost exclusively functions in 'negative catalysis'.

Note that the function of T175 is very different from the function of other amino acids at the active site of the enzyme. A conserved tyrosine 79, for instance, is the proton donor in the Etr1p reaction [16]. When this residue is mutated to phenylalanine, the Y79F variant also shows almost exclusively formation of the side product. However, at the same time the k_{cat} of the WT reaction is dramatically decreased (below the background rate of the spontaneous decay of the C4-adduct in solution). Y79 is an essential residue in catalysis and its mutation strongly increases the energetic barrier to the preferred transition state. This blocks the original trajectory of the enzyme and opens the path to the formation of the side product instead. This is represented by a decrease in both dimensions on the specificity-activity plot. Therefore, unlike T175 that increases the energetic barrier leading to unwanted side product, Y79 controls the reaction outcome by promoting formation of the preferred reaction product. A similar positive effect on catalysis can be attributed to the two residues within

hydrogen bond distance to the hydroxyl group of Y79; S70V and T324A. Both of them only affect k_{cat} by lowering the pK_a of the proton donor, thus lowering the activation barrier for protonation. In conclusion, Etr1p represents a case example of how different amino acid residues at the active site work together to control the outcome of an enzyme reaction through positive (Y79, S70V and T324A) and negative (T175) catalytic contributions.

Taxadiene synthase (TXS). TXS cyclizes geranylgeranyl diphosphate (GGPP) to taxadiene (taxa-4(5),11(12)-diene), a precursor of the antitumor compound Taxol[17]. The enzyme forms taxadiene at a yield of 93%. However, TXS also forms 5% of the taxadiene isomer taxa-4(20),11-diene, <1% of verticillia-3,7,12(13)-triene, as well as 1% of verticillia-3,7,11(12)-triene as naturally occurring side products [18]. Site-directed mutagenesis in combination with molecular mechanic simulations identified multiple residues that play an essential role in guiding the reaction towards the desired taxadiene product. Catalysis in TXS is initiated by a Mg^{2+}-mediated pyrophosphate elimination step, forming a highly reactive carbocation, which is converted over multiple other carbocation intermediates to the final product.

Site directed mutagenesis studies revealed multiple residues in TXS that contribute mainly to reaction specificity and do not affect the k_{cat} of the reaction [18]. Substitution of W753 by a histidine, for instance, lowers enzyme activity by 50%, but completely redirects the reaction to the formation of one side product only, (-)-(R)-cembrene A, which is formed below the detection limit in the WT enzyme. This strongly suggests that W753 functions mainly in preventing the premature termination of the reaction. Another essential residue in catalysis is V584, which is positioned very closely to the positively charged monocyclic carbocation. The V584L variant still operates with 90% activity, while formation of the side product verticillia-3,7,12(13)-triene is increased by two orders of magnitude (from 0.8% to 84%). This suggests that V584 and W753, mainly guide the reactive intermediate along the reaction coordinate by actively preventing competing side reactions to take place. Note that alternatively the introduced variant could also lower the activation energy for the side reaction (e.g. the histidine could serve as a base for premature deprotonation), this would however be expected to be accompanied by an increase in k_{cat} for this side reaction. Therefore, these residues most likely play a similar role as T175 in Etr1p by directing, but not accelerating the chemistry of the enzyme to control the catalytic outcome, as suggested recently [19].

Triosephosphate isomerase (TIM). During the interconversion of GA3P and DHAP, TIM forms a reactive enediol(ate) phosphate intermediate as part of the catalytic cycle in the WT enzyme. This enediol(ate) phosphate intermediates are unstable in water and prone for β-elimination of the phosphate with estimated rate constants between $8 \cdot 10^6$ and $8 \cdot 10^8$ s^{-1}, which is at least 100 times faster than the protonation in solution [20]. The side product of this elimination is methylglyoxal which is only formed once every million reactions in the WT enzyme. Knowles and coworkers showed that a loop in the enzyme closes the active site upon substrate binding, which stabilizes the enediol(ate) intermediate [21]. This induced fit allows the enzyme to bind the enediol(ate) intermediate stronger than both substrate and product, which locks the phosphate group in a conformation disfavoring elimination[22-24]. A critical residue in catalysis is lysine 12 that forms direct contacts with both the phosphate and one of the hydroxyl groups of the enediol(ate). Mutagenesis of the L12 by a glycine lead to a 12,000 fold decrease in k_{cat} of the enzyme, accompanied by a 180,000 fold increase of methylglyoxal formation [25]. Consequently, L12 is important in stabilizing the enediol(ate) phosphate intermediate transition state. At the same time L12 also destabilizes the transition state leading to the β-elimination side reaction. L12 seems to exert both positive, as well as negative control on the outcome of the reaction of TIM (Figure 2B).

Studies on the phosphate binding pocket of TIM have additionally revealed the importance of leucine 232 and isoleucine 172 in binding and positioning the phosphate group. Mutation of either L232A or I172A lead to the increased formation of methylglyoxal (around one every ten turnover in both cases) but a less pronounced loss in k_{cat} than for the L12G variant (10% of WT activity for L234A and 0.5% activity for I172A) [26]. These residues are therefore all influencing both the promotion of the transition state stabilizing the enediol(ate) intermediate for the isomeration reaction and the suppression of the transition state leading to the unwanted β-elimination, but to a different extent. Future theoretical and experimental investigations on TIM might lead to a clearer separation between the individual contributions of these residues towards catalysis.

8.2.6. Implications of the existence of 'negative catalysis' in enzymes

Above examples provide evidence that the active destabilization of unfavorable transition states is a strategy used by enzymes to control the outcome of catalysis. This strategy is complementary to the well-known principle of promoting favorable transition states, which is commonly used to explain enzyme catalysis. Accordingly, enzyme catalysis is the combined result of both positive and negative constraints on the substrate-to-product reaction landscape, which are determined by the amino acid residues at the active site. Notably, positive and negative contributions can be unified in one and the same active site residue (*e.g.*, L12 of TIM), but also separated onto different amino acids (*e.g.*, T175 and Y79 in Etr1p). Assessing the contribution of individual active site residue onto turnover frequency as well as reaction specificity is an important prerequisite to understand the full catalytic landscape of an enzyme.

Even though simple in principle, the assignment of positive and negative catalytic contributions to individual amino acids is not straight forward in practice. This becomes especially apparent in cases like V584 in TXS that almost exclusively serve a negative catalytic function. When using standard enzyme assays that are based on measuring substrate or cofactor consumption, such residues would obviously not show a difference compared to the WT. Only the complete characterization of the enzyme reaction, *i.e.* the quantification of all reaction products formed, will allow to understand the contribution of these residues to catalysis. In this context, it might be worthwhile to speculate how many of the active site residues that have been investigated in respect to their effect onto k_{cat} but were found to be non-important, might actually serve a function in suppressing unwanted side reactions.

Another difficulty is that even when an active site mutant is strongly affected in k_{cat} upon mutation, this does not exclude the (additional) function of the particular residue in negative catalysis. One example is L12 in TIM that supposedly serves both a positive and a negative function. Another example is actually T175 in Etr1p, whose function in negative catalysis was initially masked by the fact that the T175V variant appeared to be catalytically dead on first glance, when using conventional spectrophotometric assays. Only close inspection using stopped-flow spectroscopy demonstrated that the T175V variant is basically still completely functional, but quickly accumulates the C4-adduct side product, which in turn appeared to be a very strong inhibitor of the enzyme at nM concentration [15]. In summary, these examples show that the quantification of the positive and negative contributions of active site residues onto catalysis is not a simple task that can be achieved through a standard procedure, but requires the careful attention and experimental planning of the researcher.

8.2.7. Using 'negative catalysis' in enzyme engineering, *de novo*-design and directed evolution

In *de novo* design and engineering of enzymes towards new chemical reactions, the focus almost exclusively lies on the introduction of residues that lower the activation energy for the desired reaction. Based on the examples discussed above, it might be of equal importance to consider the introduction of residues conferring a negative contribution towards undesired reactions. Similar to the

concept of negative design that is already used in the *de novo* design of proteins [27], including negative catalytic elements as a basic design constraint could pave the way for new efficient and more precise catalysts in biology and chemistry. A recent example emphasizing these aspects is the successful engineering of an *anti*-Markovnikov alkene oxidase (aMOx) from a P450 alkene epoxidase form *Labrenzia aggregate*, which displayed some oxidase side reactivity [28]. The desired oxidase reaction strongly competes with an energetically favorable epoxidation reaction. To achieve a change in chemical reactivity, the path of the reaction needs to be shifted from a concerted oxo transfer to a stepwise oxidation pathway involving a 1,2-hydride migration. In principle, this switch could be achieved by stabilizing the desired transition states along the *anti*-Markovnikov oxidation coordinate or on the destabilization of the epoxidation reaction. Initial directed evolution did only increase the overall enzyme activity but not the reaction selectivity towards the *anti*-Markovnikov oxidation (Figure 1B, points 1A and 2). Further evolution selecting on the ratio between *anti*-Markovnikov oxidation and epoxidation led to the desired switch in chemical reactivity (Figure 1B, points 5 and 10). The amino acids mutated during the directed evolution rounds are widely distributed over the enzyme and their detailed functions are not yet characterized. It will be interesting to assess the positive and negative contributions of these residues to catalysis.

8.2.8. Final remarks

In summary, we hope that this work revives the discussion about the negative contributions in enzyme catalysis. The concept of active site residues that function in destabilizing unwanted transition states may not only help to identify the role of conserved residues, who lack an apparent function in catalysis so far, but might also help in the design and engineering of enzymes with new, unusual or altered reaction specificity.

8.3. Substrate channeling

Millions of years of evolution produced biological systems able to catalyze a myriad of reactions in parallel and in one compartment, the cell. In that process nature had to find solutions that are able to manage cross talk between pathways, competing side reactions, toxic and unstable intermediates and can overcome unfavorable thermodynamic bottlenecks to ensure efficient metabolic flux through metabolism to sustain life. It has been frequently observed that enzymes within a metabolic pathway are able to transfer intermediates directly from one enzyme to the next without releasing it into the bulk solvent, thus avoiding the above mentioned problems of pathway cross-talk and toxicity. This process is termed substrate channeling. The design of similar cascade reactions with controllable product distribution, excellent molecular efficiencies and high selectivity are a great challenge in chemistry and synthetic biology, there is therefore great interest in understanding natural systems as well as in applying those to novel cascade reactions [29].

Theoretical calculations show that proximity of enzymes alone is usually insufficient to produce effective channeling, as the diffusion of small molecules in the cytoplasm is often orders of magnitudes faster than the reaction rates. To be able to observe a channeling effect purely based on proximity it has been calculated that the distance between active site should not exceed 1 nm [30]. Longer distances require bonded diffusion (e.g. chemical gradients or intermolecular interactions) or sequestration (e.g. compartmentalization or a molecular tunnel) [29]. Methods to evaluate substrate channeling are largely indirect and often difficult to interpret. This makes it challenging to generate clear evidence of channeling for a designated reaction cascade and it is often advised to use multiple independent methods. These include transient time analysis, isotope dilution and enrichment studies, cascade resistance to a competing side reaction or to a reaction inhibitor [29]. Accompanying structural information is also very helpful to analyze natural and artificial systems for their capacity to channel substrates.

A recent study on the proximity effect of enzyme cascades scaffolded on DNA nicely highlights these difficulties. Multiple studies had previously shown that scaffolding enzyme cascades on DNA enhances the overall activity and a channeling effect was evoked to explain the observed improvement. Theoretical calculations and models however predicted that in these cases proximity cannot explain the observed rate enhancement. Careful reanalysis of the scaffolded cascade revealed that the reason for the observed enhancement is that the pH near the surface of negatively charged DNA nanostructures is lower than in the bulk solution, which created a more optimal pH environment for the scaffolded enzymes [31].

There are many different ways natural systems control the diffusion of cascade intermediates, including intramolecular tunnels (eg. tryptophane synthase , carbamoyl-phosphate synthase), electrostatic guidance (eg. malate dehydrogenase – citrate synthase complex), covalent bonding of intermediates (eg. polyketide synthases, pyruvate dehydrogenase complex) and micro-compartmentalization (eg. carboxysomes, bacterial micro-compartments) [29]. In this work we characterize two natural enzyme cascades and show biochemically as well as structurally that they channel their intermediates.

The trifunctional propionyl-CoA synthase (PCS) forms a closed reaction chamber to sequester the reactive acrylyl-CoA intermediate (Chapter 6). PCS therefore seems to use a slightly different mechanism than the ones described above. The reaction chamber encloses all three active sites, but does not show the directionality of a conventional tunnel, and the CoA ester intermediates are not covalently attached to the enzyme but freely diffuse within the compartment. The reaction chamber is completely formed by the enzyme itself and does not require any structural proteins to form the compartment, which is in strong contrast to other described micro-compartments. The volume of the

reaction chamber is about an order of magnitude smaller than the smallest described bacterial micro compartment. PCS therefore represents an intriguing, alternative mechanism of substrate channeling which harbors great potential for the engineering of a new class of dynamically regulated nanoreactors.

The substrate channeling mechanism of the thiolase/HMG-CoA synthase complex of archaea (Chapter 7) most closely resembles the covalent swinging arm fatty acid and polyketide synthases use to channel their intermediates. In the thiolase/HMG-CoA synthase complex the intermediate is however not covalently attached, but instead tightly bound in a shared CoA binding site, enabling the pantothenyl-arm of CoA to swing from the thiolase active site to the HMG-CoA synthase active site. The two channeling systems described in this work therefore represent two alternative ways of channeling CoA ester intermediates in a non-covalent fashion.

8.4. Outlook

The goal of this thesis was to study the mechanisms that enzymes use to control reactive intermediates, both inside one active site and between enzymes of a metabolic pathway. The recently discovered catalytically competent C2 intermediate formed at the active site of enoyl-thioester reductases and reductive carboxylases served as a starting point to study reactive intermediates within the active site of an enzyme. We were able to assign the function of most active site residues in Etr1p and used the C2 adduct successfully to engineer the stereo specificity of protonation. It is however still unclear whether the C2 adduct is part of the reaction coordinate or just an energetically close side product. Measurements of the secondary kinetic isotope effect of NAD(P)H containing a deuterium label at the C2 position could help to clarify this question. If the C2 adduct lays on the reaction coordinate and a covalent bond between the C2 of NAD(P)H and the substrate is formed during the reaction, which changes the hybridization of the C2 carbon, a secondary kinetic isotope effect should be detectable. QM/MM modeling could additionally help to estimate the activation energies involved in either mechanism.

The C2 adduct is additionally a great tool to study the detailed molecular mechanism of CO_2 activation in reductive carboxylases, a principle, which is not well understood for any CO_2 fixing enzymes. The combination of structural data gained during this work as well as the ability to probe the carboxylation step independent of the reduction/hydride transfer makes enoyl-CoA carboxylase/reductase a very good model system for such investigations. Additionally it would be interesting to look at the evolutionary history of CO2 fixation in the MDR superfamily. Initial phylogenetic analysis indicates that the CO2 fixing function evolved from simple reductases, which acquired a CO2 binding pocket and changed the resolving electrophile from a proton to CO2.

With the established biosynthesis method in this work using reductive carboxylation, the synthesis of diverse libraries of malonyl-CoA derivatives is now feasible. These libraries can now be used to systematically screen and engineer the specificity of extender unit incorporation in polyketide synthases. The change in stereochemistry of the protonation in Etr1p could be transferred to the enoyl reductase domains of these polyketide synthases to further increase the polyketide diversity.

With the description of two natural substrate channeling systems able to non-covalently channel CoA ester intermediates, it would be interesting to adapt/engineer those for the use in other challenging pathways, which involve such intermediates.

8.5. References

1. Sandor, R.; Slanina, J.; Midlik, A.; Sebrlova, K.; Novotna, L.; Carnecka, M.; Slaninova, I.; Taborsky, P.; Taborska, E.; Pes, O., *Phytochemistry* **2018,** *145*, 77-84.
2. Libby, R. D.; Mehl, R. A., *Bioorg Chem* **2012,** *40*, 57-66.
3. Macgibbon, A. K. H.; Koerber, S. C.; Pease, K.; Dunn, M. F., *Biochemistry* **1987,** *26* (11), 3058-3067.
4. Hamilton, G. A., *Progress in Bioorganic Chemistry* **1971,** *1*, 83-157.
5. Rosenthal, R. G.; Ebert, M. O.; Kiefer, P.; Peter, D. M.; Vorholt, J. A.; Erb, T. J., *Nat Chem Biol* **2014,** *10* (1), 50-U85.
6. Chen, L.; Petrelli, R.; Felczak, K.; Gao, G.; Bonnac, L.; Yu, J. S.; Bennett, E. M.; Pankiewicz, K. W., *Curr Med Chem* **2008,** *15* (7), 650-670.
7. Wolfenden, R.; Snider, M. J., *Accounts of chemical research* **2001,** *34* (12), 938-45.
8. Pauling, L., *Chemical and Engineering News* **1946,** *24* (10), 1375-1377.
9. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G., *Science* **2004,** *303* (5655), 186-95.
10. Benkovic, S. J.; Hammes-Schiffer, S., *Science* **2003,** *301* (5637), 1196-1202.
11. Zinovjev, K.; Tunon, I., *P Natl Acad Sci USA* **2017,** *114* (47), 12390-12395.
12. Boehr, D. D.; Nussinov, R.; Wright, P. E., *Nat Chem Biol* **2009,** *5* (11), 789-796.
13. Retey, J., *Angew Chem Int Edit* **1990,** *29* (4), 355-361.
14. Khare, D.; Hale, W. A.; Tripathi, A.; Gu, L. C.; Sherman, D. H.; Gerwick, W. H.; Hakansson, K.; Smith, J. L., *Structure* **2015,** *23* (12), 2213-2223.
15. Rosenthal, R. G.; Vogeli, B.; Wagner, T.; Shima, S.; Erb, T. J., *Nat Chem Biol* **2017,** *13* (7), 745-749.
16. Rosenthal, R. G.; Vogeli, B.; Quade, N.; Capitani, G.; Kiefer, P.; Vorholt, J. A.; Ebert, M. O.; Erb, T. J., *Nat Chem Biol* **2015,** *11* (6), 398-+.
17. Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; Mcphail, A. T., *Journal of the American Chemical Society* **1971,** *93* (9), 2325-&.
18. Schrepfer, P.; Buettner, A.; Goerner, C.; Hertel, M.; van Rijn, J.; Wallrapp, F.; Eisenreich, W.; Sieber, V.; Kourist, R.; Bruck, T., *P Natl Acad Sci USA* **2016,** *113* (8), E958-E967.
19. Freud, Y.; Ansbacher, T.; Major, D. T., *ACS catalysis* **2017,** *7* (11), 7653-7657.
20. Richard, J. P., *Biochemistry* **1991,** *30* (18), 4581-5.
21. Rozovsky, S.; Jogl, G.; Tong, L.; McDermott, A. E., *Journal of molecular biology* **2001,** *310* (1), 271-280.
22. Pompliano, D. L.; Peyman, A.; Knowles, J. R., *Biochemistry* **1990,** *29* (13), 3186-3194.
23. Alahuhta, M.; Wierenga, R. K., *Proteins-Structure Function and Bioinformatics* **2010,** *78* (8), 1878-1888.
24. Morell, M. K.; Paul, K.; Oshea, N. J.; Kane, H. J.; Andrews, T. J., *Journal of Biological Chemistry* **1994,** *269* (11), 8091-8098.
25. Go, M. K.; Koudelka, A.; Amyes, T. L.; Richard, J. P., *Biochemistry* **2010,** *49* (25), 5377-5389.
26. Malabanan, M. M.; Amyes, T. L.; Richard, J. P., *Journal of the American Chemical Society* **2011,** *133* (41), 16428-16431.
27. Noivirt-Brik, O.; Horovitz, A.; Unger, R., *Plos Comput Biol* **2009,** *5* (12).
28. Hammer, S. C.; Kubik, G.; Watkins, E.; Huang, S.; Minges, H.; Arnold, F. H., *Science* **2017,** *358* (6360), 215-+.
29. Wheeldon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov, P.; Sigman, M., *Nat Chem* **2016,** *8* (4), 299-309.
30. Eun, C.; Kekenes-Huskey, P. M.; Metzger, V. T.; McCammon, J. A., *J Chem Phys* **2014,** *140* (10).
31. Zhang, Y.; Tsitkov, S.; Hess, H., *Nature communications* **2016,** *7*, 13982.

Acknowledgements

This work could not have been accomplished without the help and support of many wonderful people which I would like to acknowledge and thank at this point.

First of all I would like to thank my supervisor **Tobi Erb** for all his guidance, trust and enthusiasm. He gave me the freedom to explore all my crazy ideas and found a way to motivate me to keep trying when they failed. I immensely enjoyed all the brainstorming sessions talking about random scientific ideas and projects and it taught me to keep an open mind for what is possible.

I would also like to thank **Prof. Mohamed Marahiel**, **Prof. Johann Heider**, **Prof. Martin Grininger** and **Dr. Hannes Link** for being part of my thesis advisory committee and for giving valuable input and advice.

Many thanks also go to **Raoul**, who was not only the supervisor of my master thesis, contributed greatly to many projects in this thesis and to improve my scientific skills, but also showed me how cool enzymes really are. **Tristan** (´the Frenchy´) taught me the ways of X-ray crystallography and solved many of the structures in this work. It was a lot of fun figuring out together, how my biochemical measurements correlate with those structures. Many of the projects in this thesis were the result of great collaborations and I would like to thank **Iria** for letting me work with her pet enzyme, **Pete** for showing me how to synthesize all the CoA esters my heart desires, **Kookie** and **Heidi** for providing MS data and analysis, **Kyra** for her support in characterizing Ccr´s and **Gabo** for his work on InhA and his appropriate jokes at appropriate times. This work would not be where it is now without the valuable help of of the two master students, whom I was privileged to supervise. Thank you **Sarah** and **Patrick**!

A scientist is nothing without a laboratory and great lab mates. I am grateful to all the former and present members of the AG Erb, who all contributed in their way to this work. It was a pleasure to see the lab grow so much over the last years and gain so many great new members. I deeply enjoyed the mixture between good science, hard work and the anarchic lighthearted atmosphere we all shared. It´s great to work with friends instead of only colleagues. I would also like to thank **Jan** and all the staff that help keeping the lab from falling into complete chaos and apologize for my mess. At this point I would especially like to thank **Thomi**, who started this great Marburg experience together with me and was a good friend and support along the way. Without you I would have never been able to overcome the challenges of German bureaucracy.

I would not be where I am today without the love and support of my **friends** and **family** and I can never thank them enough for always standing by my side and believing in me. Special thanks also goes to my **parents,** who enabled me to get to this point and were always there for me!

"*So long, and thanks for all the (back)fish!"*