Investigating and manipulating the reaction mechanism of reductive carboxylases

Dissertation

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

Dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Gabriele Stoffel

aus Bellinzona, Schweiz

Marburg/Lahn, Deutschland, 2019

Die Untersuchungen zur vorliegenden Arbeit wurden von Mai 2015 bis März 2019 unter der Betreuung von Herrn Prof. 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: 12.09.2019

Erstgutachter: Herr Prof. Dr. Tobias Erb

Zweitgutachter: Herr Prof. Dr. Johann Heider

Weitere Mitglieder der Prüfungskommission:

Herr Prof. Lars-Oliver Essen

Herr Prof. Hans-Ulrich Mösch

Tag der mündlichen Prüfung am: 17.10.2019

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „**Investigating and manipulating the reaction mechanism of reductive carboxylases**" 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 28.6.2019

Gabriele Stoffel

"Let's get crazy!"

Bob Ross

Contents

Summary

Efficient capture and conversion of atmospheric carbon dioxide (CO_2) is a prerequisite to develop a carbon-neutral, circular future economy. Carbon fixation is the process by which inorganic carbon is fixed into biomass. In Nature, enzymes called caboxlyases are able to capture atmospheric carbon dioxide under mild conditions and catalyze its incorporation into organic molecules. It is estimated that 400 Gt of CO_2 are fixed annually solely by the enzyme ribulose-1,5-bisphophate-carboxylase/oxygenase (RuBisCO), the key enzyme of photosynthesis. In comparison, CO_2 utilization by chemical industries accounts for only 0.1 Gt of carbon annually and utilizes pressurized CO_2, which emphasizes our need to understand the molecular mechanism that allow carboxylases to selectively interact with a CO_2 at atmospheric concentrations (0.04% vol) during catalysis. Enoyl-CoA carboxylases/reductases (ECRs) represent the fastest carboxylases known to date and is, in contrast to RuBisCO, completely specific for CO_2. These enzymes catalyze the reductive carboxylation of enoyl-CoAs by oxidizing one equivalent of NADPH. ECRs represent a good case study for the understanding of the CO_2 chemistry that carboxylases use.

In this work, we try to gain a better understanding of the underlying catalytic principles that enable ECRs to achieve high catalytic rates. Initially we focus on understanding how the precise interaction between protein and CO_2 takes place at the active site of ECRs. We were able to identify and assign a function to four conserved amino acid residues found at the active site of ECRs. Three residues are responsible for the precise positioning of CO_2 for nucleophilic attack by the enolate intermediate. Additionally, one residue is able to shield the active site from water thereby preventing the irreversible protonation of the enolate. These two mechanistic principles are at the base of the efficient carboxylation in ECRs. The following chapter briefly describes how the enzyme is able to accept other electrophiles than CO_2. We show that ECRs can utilize formaldehyde as an alternative electrophile to CO_2 thereby yielding β-hydroxy thioesters. The exquisite stereospecificity together with the vast range of small electrophiles make ECR a potential biocatalyst for the production of various α-substituted thioesters.

The last two chapters of this work focus on the structural aspects of ECR catalysis. We were able to obtain four new crystal structures of an ECR from *Kitasatospora setae* and to propose

a model for the catalytic cycle of this enzyme. We show that the communication between and within the dimers that compose the functional homotetramer is crucial for the fast catalytic rates observed in this ECR. A separate study aims at developing an *in vivo* directed evolution screen to improve the catalytic properties of an ECR from *Burkholderia ambifaria*. Our approach yields an evolved variant, with mutations distant from the active site. The observed improved catalytic supports the importance of the residues for the catalytic rate. Both studies revealed the importance of the residues at the interface of the ECR monomers by their impact on catalytic rates of this enzyme.

Zusammenfassung

Effiziente Abscheidung und Umwandlung von atmosphärischen Kohlenstoffdioxid (CO_2) ist eine Voraussetzung für die Entwicklung einer zukünftigen Kohlenstoff-neutralen Resourcenvewertung. Kohlenstofffixierung ist der Prozess wodurch anorganischer Kohlenstoff in Biomasse eingebaut wird. Carboxylasen sind natürlich vorkommende Enzyme welche in der Lage sind atmosphärisches Kohlenstoffdioxid unter milden Bedingungen in organische Verbindungen einzubauen. Jedes Jahr werden circa 400 Gt CO_2 alleine von Ribulose-1,5-bisphosphat-Carboxylase/Oxygenase (RuBisCO), das Schlüsselenzym der Photosynthese, fixiert. Verglichen dazu, Verwendet die chemische Industrie nur 0.1 Gt CO_2 pro Jahr und zusätzlich benötigen diese Prozesse einen hohen Druck und Temperaturen. Deshalb ist es notwendig mehr Wissen über die molekularen Mechanismen zu erhalten, welche es Carboxylasen erlauben mit CO_2 unter atmosphärischen Konzentrationen (0.04% vol) zu interagieren. Enoyl-CoA Carboxylasen/Reductasen (ECRs) sind die schnellsten Carboxylasen die heutzutage bekannt sind und sind zusätzlich, im Vergleich zu RuBisCO, spezifisch für CO_2. Diese katalysieren die reduktive carboxylierung von Enoyl-CoAs mit der Oxidation von NADPH. Deshalb stellen ECRs ein gutes Modellsystem dar um die CO_2 Chemie von Carboxylasen zu studieren.

In dieser Arbeit, versuchen wir mehr Wissen über die grundlegenden katalytischen Prinzipien zu erlangen wodurch ECRs hohe Umschlagszahlen erreichen. Im ersten Kapitel steht die Interaktion zwischen CO_2 und den Aminosäuren im Aktiven Zentrum von ECRs im Vordergrund. Es wurden vier konservierte Aminosäuren identifiziert und für jede eine Funktion zugeschrieben. Die Seitenketten von drei Aminosäuren sind verantwortlich um CO_2 genau zu positionieren damit der nucleophile Angriff vom Enolat Intermediat stattfinden kann. Eine vierte Seitenkette ist verantwortlich das Aktive Zentrum vor Wasser zu schützen sodass das Enolat Intermediat nicht irreversibel protoniert wird. Diese sind die Zwei grundlegenden katalytischen Prinzipien wodurch ECRs eine effiziente Carboxylierung katalysieren. Im Anschluss wird in einem Kapitel beschrieben, ob es möglich ist, dass ECRs zusätzlich zu CO_2, andere Elektrophile inkorporieren können. Wir zeigen, dass ECRs Formaldehyd verwenden und somit ein β-hydroxy Thioester bilden können. Die exzellente Stereospezifizität und die große Vielfalt an kleinen elektrophilen Verbindungen machen ECR zu einem potentiellen Biokatalysator für die Produktion von α-substituierten Thioestern.

In den letzten beiden Kapiteln dieser Arbeit werden strukturelle Eigenschaften von ECRs betrachtet. Vier neue Kristallstrukturen der ECR von *Kitasatospora setae* wurden gelöst und ein neuer katalytischer Zyklus für dieses Enzym vorgeschlagen. Wir haben gezeigt, dass die Kommunikation zwischen, und innerhalb der Dimere welche das funktionelle Homotetramer bilden, die Grundlage für eine schnelle Katalyse darstellt. Im letzten Kapitel versuchen wir eine *in vivo* gerichtete Evolution Strategie zu entwickeln um die katalytischen Parameter einer ECR von *Burkholderia ambifaria* zu verbessern. Mit dieser Methode haben wir eine evolvierte ECR Variante erhalten die Aminosäure Mutationen außerhalb des Aktiven Zentrums aufweist. Wir konnten feststellen, dass die Umschlagszahl verbessert wurde und somit die mutierten Aminosäuren von großer Bedeutung für die Katalyse sind. In beiden Kapiteln konnte festgestellt werden, dass Aminosäuren an der Kontaktoberfläche von Monomeren bezüglich der Katalyse in ECRs eine wichtige Rolle spielen.

Parts of this thesis are published or are in preparation

Stoffel G. M. M.[*], Saez D. A., DeMirci H., Vögeli B., Rao Y., Zarzycki J., Yoshikuni Y., Wakatsuki S., Vöhringer-Martinez E., Erb T. J.

(2019) Four amino acids define the CO_2 binding pocket of enoyl-CoA carboxylases/reductases. *Proc. Natl. Acad. Sci. USA 116 (28) 13964-13969*

Hasan DeMirci[*], Yashas Rao[*], Gabriele M.M. Stoffel[*], Bastian Vögeli[*], Kristina Schell[*], Alexander Batyuk, Cornelius Gati, Raymond G. Sierra, Mark S. Hunter, E. Han Dao, Halil I. Ciftci, Brandon Hayes, Fredric Poitevin, Kensuke Tono, David Adrian Saez, Esteban Vöhringer-Martinez, Samuel Deutsch, Yasuo Yoshikuni, Tobias J. Erb, Soichi Wakatsuki

(2019) Coupled inter-subunit dynamics enable the fastest CO_2-fixation by reductive carboxylases. Available as preprint on bioRxiv doi 10.1101/607101

Publications that are not discussed in this thesis

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. *J. Biol. Chem.* 293(44):17200-17207

1. Introduction

1.1. Carbon dioxide in the atmosphere

Our planet has seen increasing amounts of atmospheric CO_2 since the industrial revolution. Since the 18th century, CO_2 levels have increased by 200 ppm up to 400 ppm recorded in 2018 (Fig. 1) due to anthropogenic activities. Burning of fossil fuels such as coal, natural gases and oil is the major contributor of CO_2 emissions followed by land change and usage and industrial activities[1]. Carbon dioxide is a naturally occurring greenhouse gas and among others such as CH_4 and N_2O contributes to global warming[2].
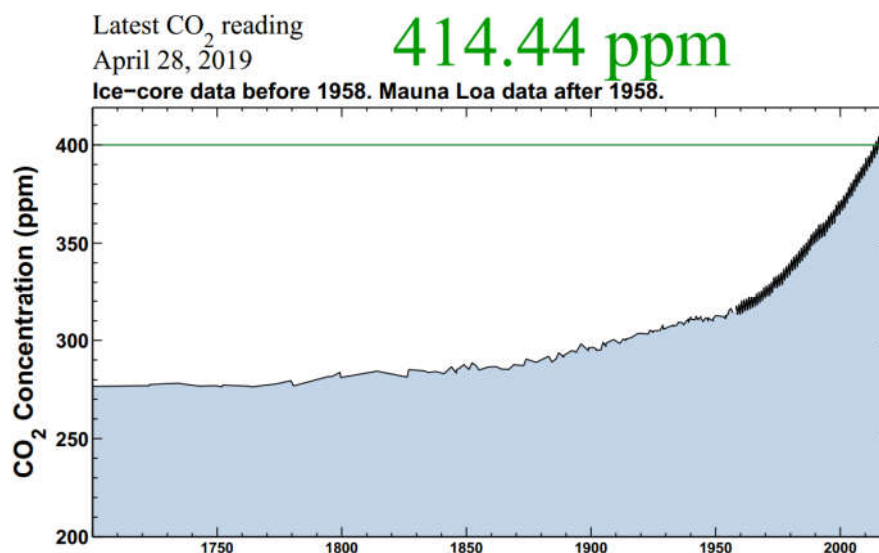

Figure 1: Atmospheric CO₂ concentration from the 18th century until present. The data was recorded at Mauna Loa, Hawaii by the National Oceanic & Atmospheric Administration (NOAA)/ EarthSystem Research Laboratory (ESRL). Picture adapted from https://scripps.ucsd.edu/programs/keelingcurve/.

The implications of these environmental changes are numerous. Thermal expansion of ocean waters as well as melting of land ice both contribute to sea-level rise and causes flooding of coastal regions[3]. Moreover, atmospheric CO_2, which dissolves in water, causes acidification of ocean waters. Over a long period this leads to a decrease in atmospheric CO_2 uptake by oceans[4] and has a negative impact on marine ecosystems[5]. In the light of this, to counteract the rise in atmospheric CO_2 concentration has become a major challenge for humanity in the past decades. This can be achieved by a combined approach of reducing emissions and the

development of efficient carbon capture strategies (CCSs). Pumping CO_2 into large underground reservoirs has been a widely adopted CCS, but in a scenario where the CO_2 emissions need to be reduced this is not a sustainable alternative[6]. Other CCSs aim at capturing atmospheric CO_2 and incorporating it into organic molecules, thus storing energy in chemical bonds. While on one hand CO_2 appears to be the cause of many environmental problems we are currently facing, on the other side it represents a readily available carbon source. Currently, chemistry is still struggling at developing efficient catalysts for the capture of carbon dioxide. Large scale industrial processes still require harsh conditions for the production of chemicals such as urea, salicylic acid and polycarbonates[7]. In contrast, Nature offers much more efficient catalytic strategies for the capture and conversion of CO_2. It is estimated that plants, specifically the enzyme ribulose-1,5-bisphosphate-carboxylase/-oxygenase (RuBisCO), are able to fix 400 Gt of CO_2 each year, thereby greatly exceeding the value for atmospheric CO_2 fixed in industrial processes (0.1 Gt/year)[8]. It is obvious that Nature possesses more efficient carbon capture mechanisms that also operate under mild physiological conditions. Understanding catalytic principles on Nature's CO_2 capturing catalysts, namely carboxylases will provide more insight into the mechanistic principles of CO_2 chemistry and provide the tools for the rational engineering of enzymes towards more efficient CO_2 fixation.

1.2. Carboxylases in natural CO_2 fixation pathways

In Nature, enzymes are able to capture carbon dioxide and convert it into biomass under mild conditions. The catalysts responsible for this reaction are enzymes known as carboxylases. They are among the most important enzymes because they catalyze the fixation of inorganic carbon. In this way, they contribute to the global carbon cycle by fixing atmospheric carbon into biomass. They are present in all domains of life and function in different metabolic contexts.

Carboxylases in autotrophic pathways are alone responsible for the complete carbon fixation into biomass of the organism. One example is the reductive tricarboxylic acid (TCA) cycle, which represents the most energy efficient CO_2 fixation cycle. The efficiency directly correlates with the carboxylases found in this pathway, namely the isocitrate dehydrogenase, α-

ketoglutarate:ferredoxin oxidoreductase and pyruvate ferredoxin oxidoreductase. All of these are reductive carboxylases and except isocitrate dehydrogenase, utilize the strong reducing power of ferredoxin to drive an energy efficient carboxylation reaction[9]. The distribution of this pathway is, despite its efficiency, limited to anoxic environments due to the inherent oxygen sensitivity of enzymes involved in these pathways. In the Calvin-Benson-Bassham cycle[10], found in plants algae and bacteria, RuBisCO is responsible for fixation of CO_2. This process is estimated to fix 400 Gt of CO_2 each year and represents an essential process for all life on earth. It is therefore not surprising that the estimated amount of RuBisCO on earth is 5 kg of soluble protein per person. A disadvantage this enzyme possesses is its inherent reactivity with oxygen instead of CO_2. About one in every five catalytic cycles RuBisCO utilizes oxygen and produces 2-phosphoglycolate, a toxic intermediate that has to be metabolized by the organism through an energy demanding process known as photorespiration. The carboxylase involved in the 3-hydroxypropionate bicycle and the hydroxypropionate/hydroxybutyrate cycle[11], acetyl-CoA/propionyl-CoA carboxylase, is a bifunctional biotin dependent carboxylase that is able to carboxylate acetyl-CoA and propionyl-CoA to malonyl-CoA and methylmalonyl-CoA respectively. Biotin dependent carboxylases utilize bicarbonate (HCO_3^-) as carboxylating species. Bicarbonate has to be activated prior to covalent attachment onto biotin, and this step requires the formation of carboxyphosphate via nucleophilic attack of bicarbonate onto ATP.

Apart from autotrophic carbon fixation, carboxylases also operate in non-autotrophic assimilatory strategies. There, the carboxylation reaction introduces a reactive functional group into the inert substrate molecule which facilitates further downstream processing. Examples for this are NADPH:2-ketopropyl-CoM oxidoreductase/carboxylase of *Xanthobacter autotrophicus Py2*[12] and acetone carboxylase which convert epoxypropane and acetone, respectively to acetoacetate. This compound is then activated to the corresponding CoA ester and is converted to central metabolites such as acetyl-CoA (Scheme 2e). The same strategy is employed by the carboxylases that generate precursor molecules for the biosynthesis of fatty acids[13, 14]. These biotin carboxylases generate malonyl-CoA from acetyl-CoA. Here the carboxylate group is important for the elongation of fatty acids through decarboxylative Claisen condensation.
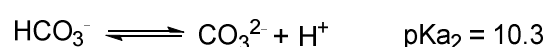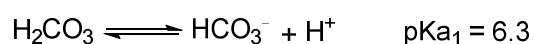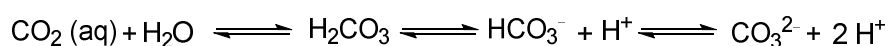
In the assimilation of C_2 units in the form of acetate the ethylmalonyl-CoA pathway (EMCP) was initially characterized as an alternative to the well know glyoxylate shunt. The EMCP is characterized by crotonyl-CoA carboxylase/reductase (Ccr)[15] that is able to reductively carboxylate enoyl thioesters. Ccr converts crotonyl-CoA to ethylmalonyl-CoA by oxidizing one equivalent of NADPH and using CO_2 as carboxylating species. Further steps in the pathway finally yield succinyl-CoA and (L)-malate which are both C_4 units and can be fed back into central carbon metabolism. It was additionally demonstrated that these carboxylases also operate in secondary metabolism by producing alkylmalonyl-CoA precursors for the biosynthesis of polyketides[16].

Despite the occurrence of carboxylases in various metabolic pathways, they are unified by common mechanistic principles, which will be discussed in the following chapter.

1.3. Mechanistic principles employed by carboxylases

Carboxylases differ greatly in their requirements of substrates, cofactors and mechanisms, but they all face the same challenge of activating CO_2. CO_2 represents the most oxidized form of carbon and is thus a thermodynamically and kinetically very stable compound[7, 17]. Carboxylases have evolved different strategies to overcome this problem. A common strategy is that of substrate activation. This generates reactive intermediates called enolates, which represent strong carbon nucleophiles that are able to attack CO_2 and form a stable C-C bond. Formation of these intermediates is very difficult in aqueous solution given that they are easily protonated, but carboxylases possess active sites that can stabilize these compounds by interactions with the protein backbone, sidechain of amino acids or metal cofactors[18]. The formation of enolates is predicted from structures of substrates of carboxylases such as acyl-thioesters[12, 13], ketones[19] and α,β-unsaturated enoyl-thioesters[15]. The strategies to generate these enolate intermediates vary among carboxylases. In RuBisCO, for example, the sidechain of an active site lysine, which is carboxylated to form a carbamate as a post translational modification, abstracts a proton from C3 of the substrate ribulose 1,5-bisphosphate and generates a dienolate which is stabilized by an active site bound magnesium ion[20] (Scheme 2a). In acetone carboxylase, formation of the enolate is dependent on ATP in the presence of

Mg^{2+} and yields phosphoenol acetone as the activated nucleophile responsible for the attack of CO_2[19]. An even different approach is used by ECRs, which reduce their enoyl-CoA substrates by the use of NADPH[15] and the resulting thioester enolate is stabilized by interaction of the 2'-OH group of the NADPH ribose ring (Scheme 2c). An alternative route has been proposed in ECRs which involves the formation of a covalent adduct between substrate and cofactor which is then resolved by the enzyme to form the enolate[21]. Substrate activation provides a great driving force for the reaction of carboxylases, but the preferred carboxylating species (HCO_3^- or CO_2) is not always present in the form required by the enzyme. This led to a diverse set of strategies that allowed carboxylases to adapt to the environmental conditions in order to capture and transform their preferred carboxylating species. As enzymes operate in aqueous solution, such as the cytoplasm of a cell, it is important to consider in what form CO_2 will be available.

$$CO_2\,(aq) + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons HCO_3^- + H^+ \rightleftharpoons CO_3^{2-} + 2\,H^+$$

$$H_2CO_3 \rightleftharpoons HCO_3^- + H^+ \qquad pKa_1 = 6.3$$

$$HCO_3^- \rightleftharpoons CO_3^{2-} + H^+ \qquad pKa_2 = 10.3$$

Scheme 1: Equilibrium of carbon dioxide dissolved in water and pK_a values for the dissociation of carbonic acid and bicarbonate (pKa_1 and pKa_2) at 25 °C.

Given that CO_2 and HCO_3^- are at equilibrium in aqueous solution (Scheme 1)[22] the question arises on what strategies carboxylases have evolved to utilize the preferred species depending on the availability. Bicarbonate dependent carboxylases have to overcome the inherent poor electrophilicity of bicarbonate. To activate this species an energy input is required and the energy is provided by the exergonic hydrolysis of ATP. Nucleophilic attack of bicarbonate onto the γ-phosphate of ATP yields carboxyphosphate where the carbon atom is more electrophilic. One class of enzymes that utilize this strategy are biotin carboxylases. In these enzymes the biotin cofactor is linked to an active site lysine and is able to swing between the biotin carboxylase (BC) and carboxyl transfer (CT) domain. In the BC domain, carboxyphopshate acts as a CO_2 donor and carboxylates the nitrogen of the ureido moiety of biotin. After translocation of to the CT domain a molecule of CO_2 is liberated and carboxylates the enolate

of the corresponding substrate (e.g. acetyl-, propionyl-CoA) (Scheme 2b).
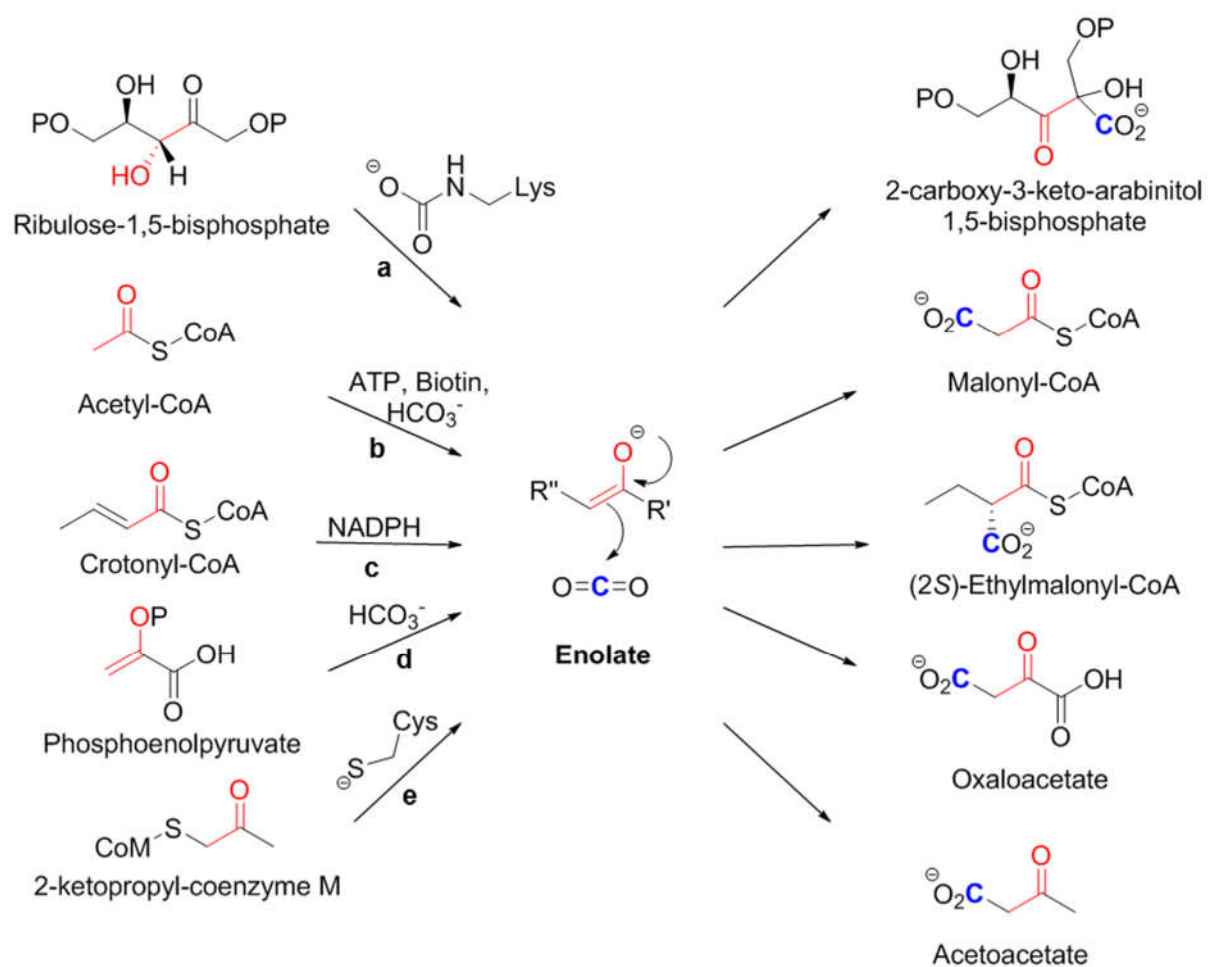
Scheme 2: Reaction mechanism of carboxylases displaying the structure of the common enolate intermediate. Highlighted in red are the three atom centers, which stabilize the enolate. Enzymes catalyzing each reaction are **a**: RuBisCO, **b**: Acetyl-CoA carboxylase, **c**: ECR, **d**: PEP carboxykinase and **e**: 2-ketopropyl coenzyme M oxidoreductase/carboxylase.

Another example is phosphoenolpyruvate (PEP) carboxykinase where carboxyphophsate is generated by transferring the phosphate group of PEP onto bicarbonate. Subsequently enolpyruvate is carboxylated to yield oxaloacetate[23] (Scheme 2d). Another strategy carboxylases employ is that of electrophilic activation of CO_2. A structural study on PEP carboxykinase from *E. coli* proposed that active site residues form a CO_2 binding pocket[24] in which positively charged active site residues interact with the oxygen atoms of CO_2. This interaction polarizes the C-O double bond and withdraws electron density from the central carbon atom thus increasing its electrophilicity. This may potentially provide a different strategy to activate the central carbon atom and similar amino acids have been observed in active sites of other (de)carboxylases as well[25-27].

Lastly, the increase of effective amount of CO_2 in or around the active site of carboxylases is another strategy to promote carboxylation turnovers. An increased presence of CO_2 in the surroundings of an enzyme represents a thermodynamic driving force by pushing the equilibrium to the product side. For example, carboxysomes are macroscopic protein structures found within cyanobacterial cells[28-30]. They encapsulate RuBisCO and carbonic anhydrase thereby creating a compartmentalization for these enzymes. Bicarbonate is able to diffuse through the protein layer and once it reaches the inside, carbonic anhydrase converts it to CO_2. This increases the local concentration of CO_2 around RuBisCO. Additionally these microcompartments are not permeable to oxygen. The overall result is an environment where RuBisCO can perform carboxylation reaction without the presence of oxygen, which leads to undesired side reactions.

Mechanistic studies on carboxylases have elucidated the fundamental mechanistic principles underlying the catalysis of carboxylases. The next chapter focuses on the family of ECRs, which represent one of the fastest carboxylases described today.

1.4. Enoyl-CoA carboxylase/reductase (ECR)

ECRs represent a class of carboxylases, which catalyze the unique reductive carboxylation of α,β-unsaturated enoyl thioesters by oxidizing one equivalent of NADPH (Scheme 3). This class of enzymes was initially described in the EMCP in which it converts crotonyl-CoA to (2S)-ethylmalonyl-CoA, the namesake compound of the pathway.
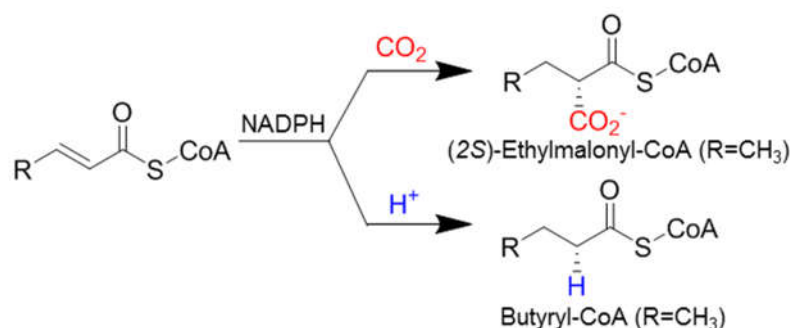
Scheme 3: Reaction catalyzed by ECRs in the presence and absence of CO_2. The residue "R" can vary considerably[16, 31].

This anaplerotic pathway is found in many bacteria and represents an acetyl-CoA assimilation strategy in the absence of isocitrate lyase (ICL), a key enzyme of the glyoxylate cycle. Operation of this pathway was first described in the Alphaproteobacterium *Rhodobacter sphaeroides*, which, fed with acetate under aerobic conditions, is able to grow despite lacking ICL. Moreover, the EMCP is important for methylotrophs, which can grow on C_1 carbon sources such as methanol via the serine cycle[32]. Outside of their metabolic contexts, ECRs also represent an interesting model enzyme to study mechanistic principles of carboxylation. These enzymes outcompete other carboxylases in turnover rate and/or specificity for their substrate[33]. Mechanistic studies on ECRs have demonstrated that the initial hydride transfer occurs from the pro-(4*R*) hydrogen of NADPH. The nucleophilic attack of the enolate intermediate onto CO_2 yields a (*S*) configuration at the C_α of ethylmalonyl-CoA. When omitting CO_2 from the reaction ECR catalyzes the reduction of its substrate, yielding, in the case of crotonyl-CoA as starting material, butyryl-CoA[15]. This reactivity is equivalent to enoyl thioester reductases (ETRs) which operate, for example, in fatty acid biosynthesis[34]. In fact, ECRs belong to the same superfamily as ETRs, namely, the medium-chain dehydrogenase/reductase (MDR) superfamily. In the light of this phylogenetic analysis, it was proposed that the reduction reaction catalyzed by ECRs was present in an early evolutionary state of the enzyme and that the carboxylation function, was acquired only later[35]. ECRs possess a characteristic active site able to accommodate CO_2 and exclude water to prevent the reduction side reaction[36], but this hypothesis still lacks experimental evidence. Recent mechanistic studies revealed that, in the absence of CO_2, a reaction intermediate accumulates during the reaction catalyzed by ECRs. Structural characterization using NMR spectroscopy and mass spectrometry showed that this compound is a covalent adduct between NADPH and crotonyl-CoA covalently linked between the C_α of crotonyl-CoA and the C2 carbon of the nicotinamide ring of NADPH. It was also shown that, in the presence of CO_2, the enzyme is able to convert this adduct to (2*S*)-ethylmalonyl-CoA and $NADP^+$. This suggests that this intermediate might be relevant for the catalytic cycle of ECRs but the accumulation of this intermediate during the carboxylation reaction has only scarce evidence[21]. Many recent studies demonstrated that the accumulation of this intermediate is not limited to ECRs but is observed in the MDR and other enzyme families as well[34, 37-39] which hints towards a more widespread occurrence of this type of NADPH catalysis in oxidoreductases.

Since its first description it was clear that ECR was also operative in secondary metabolism of streptomcyetes[32] by providing extender units for the biosynthesis of antibiotics. The fact that many ECR homologs are located within gene clusters together with polyketide synthases (PKSs) and given the complexity and diversity of polyketide structures, the substrate scope of these enzymes must extend beyond that of crotonyl-CoA[16]. A study on substrate promiscuity of ECRs revealed three active site residues that define the size of the acyl moiety of the thioester that can be accommodated in the active site. The active site of ECRs from primary metabolism is lined by three bulky residues (C146, I169 and F373 in the ECR from *Caulobacter crescentus*) which restricts activity to crotonyl- and pentenoyl-CoA as substrates. In contrast, in ECRs from secondary metabolism the active site is lined by small residues (P141, A163, G362 in CinF (PDB: 4A0S)) which allow for bulky and branched (e.g. octenoyl-, cinnamoyl-, 5-methylhexenoyl-CoA) enoyl-CoA esters as substrates. By site-directed mutagenesis of the active site residues of the ECR from *C. crescentus* it was possible to increase the substrate promiscuity of this enzyme. This study revealed consensus amino acid residues that determine the substrate promiscuity in ECRs and provided the basis for rational manipulation of substrate specificity in these enzymes. Recently this concept was used to generate unnatural malonyl-CoA extender units for polyketide synthases[40]. The successful production of six different alkylmalonyl-CoA derivatives allowed the authors to generate diverse triketides in a modified *in vitro* PKS system.

The rational design approach to engineer enzymes is greatly aided by X-ray crystallography. Initially, crystal structures of ECRs were limited to apo structures and/or structures in complex with cofactor (e.g. ECR from *Streptomyces collinus* (PDB 3HZZ), ECR from *Streptomyces coelicolor* (PDB 3KRT), ECR from *Methylobacterium extorquens* (PDB 4GI2), AntE from *Streptomyces albus* (PDB 4Y0K)[41]. This limited the structural understanding with respect to how this enzyme interacts with enoyl-CoA substrates and CO_2. Later, a structural study on CinF, an ECR form *Streptomyces cinnabarigriseus*, became the base to gain more insights into the structural aspects of catalysis. CinF (PDB 4A0S) is responsible to generate hexylmalonyl-CoA extender units from octenoyl-CoA for the biosynthesis of cinnabaramide natural products. This study showed the first ECR crystal structure, which contained both $NADP^+$ and octenoyl-CoA[36]. The presence of both substrate and cofactor in this structure shows an active site preordered for catalysis and allowed for the *in silico* docking of CO_2. This revealed a potential CO_2 binding site composed of highly conserved residues among the whole ECR

family. The founding hypothesis was that in the active site of CinF (PDB 4A0S) the sidechains of N77 and E167 would interact with the two oxygens of CO_2 through hydrogen bonding whereas the sidechain of F166 would undergo hydrophobic interactions with CO_2 and shield the active site from water. No conclusive data is available to validate this hypothesis and it still remains unclear how ECRs are able to interact with CO_2 at their active site.

1.5. Evolution within enzyme families

Studying the evolution enzyme families allows us to gain a better understanding of the fundamental mechanisms that drive functional and structural evolutionary divergence in enzymes. Laboratory evolution offers the great possibility to observe intermediary stages along the evolutionary coordinate. In this respect, directed evolution represents a powerful methodology to recreate natural selection and enables evolution on a shorter timescale compared to Nature.

It is believed that the specificity displayed by modern enzymes is a consequence of selective pressure on more promiscuous enzymes[42]. When exposed to an environment with a specific set of conditions, (e.g. temperature, pH, salinity, limited nutrient availability) enzymes are put under strong selective pressure and have to rapidly adapt. This principle of natural selection drives the evolution of enzymes toward variants with specific functions and kinetic parameters. Knowledge of these parameters allows us to gain a better understanding of the biochemistry, evolutionary history and the role of enzymes in various physiological contexts.

For example, enzymes operating in metabolic pathways represent a large library of kinetically characterized catalysts for studies on evolution within enzyme families. In central metabolism (carbohydrates, fatty acids, nucleotides and amino acids, energy conservation), in order to sustain the high flux rates of metabolites through these pathways, enzymes have to become catalytically efficient and fast. In kinetic terms, this translates to a high k_{cat}/K_M and k_{cat} value, respectively. In contrast, enzymes from secondary metabolism (the metabolism of non-essential metabolites that are produced in low amounts) have to sustain a lower metabolic flux and are thus on average slow. This is reflected in a 30-fold higher average k_{cat} and 6-fold higher k_{cat}/K_M value of enzymes in central compared to secondary metabolism[43]. In reality,

other factors than metabolic flux rates also play an important role in the evolution of enzymes. The physicochemical properties of substrate molecules, such as molecular mass and hydrophobicity, have a major impact on their K_M values. The smaller size of a substrate translates to less interaction surface with the enzyme which in turn is reflected in high K_M values (e.g. hydrogen peroxide in catalase[44] and carbon dioxide in carbonic anhydrase[45]). Moreover, it was shown that the intracellular concentration of metabolites often exceeds K_M values in metabolic enzymes. This was explained by the necessity of keeping flux rates high throughout central metabolic pathways. Low K_M values are not needed when enzymes operate in the presence of saturating substrate concentrations. Under these circumstances, enzymes can operate at maximal turnover (k_{cat}) and maintain a high flux of metabolites[46]. This scenario confirms the concept of the inherent accuracy-speed tradeoff that enzymes display[47], which describes to what degree an improvement in accuracy corresponds to a decrease in catalytic efficiency. Finally, adaptation might also occur on the genetic level, where the regulation of the expression level changes the amount of enzyme present in the cell.

1.6. Aims of this thesis

The general goal of this thesis was to gain a better understanding of the underlying principles of ECR catalysis. Of particular interest was the interaction between the protein and the CO_2 molecule at the active site of the enzyme and how this knowledge can be exploited to rationally identify new electrophiles, other than CO_2, for this enzyme. Moreover, this thesis focuses on understanding the structural determinants that are at the basis of the fast catalytic rates observed in ECRs.

1.6.1 Mechanistic Principles of CO_2 interaction at the active site of ECRs

The second and third chapter delve into the fundamental aspects of CO_2-binding in ECRs. As ECRs represent one of the fastest carboxylases known today and are highly specific for CO_2, they represent a good model system to study CO_2 chemistry employed by enzymes. Using experimental biochemistry, X-Ray crystallography and QM/MM simulations, we were able to identify and assign a function to four conserved amino acid residues. ECRs appear to utilize residues, which anchor and precisely position CO_2 for nucleophilic attack by the enolate intermediate. Additionally, one residue is able to shield the active site from water thereby preventing the irreversible protonation of the enolate. These two mechanistic principles are at the base of the efficient carboxylation observed in ECRs. Our gained knowledge in this study prompted us to test whether ECR was promiscuous towards other electrophiles than CO_2. We showed that ECR can utilize formaldehyde instead of carbon dioxide and yield 2-hydroxy thioesters. Further development of this system has the potential to yield a more diversified product range which could be employed as building blocks for organic synthesis.

1.6.2. Structural determinants of ECR catalysis

The fourth and fifth chapter of this thesis focus on the structural aspects of ECR catalysis. We were able to obtain four new crystal structures of an ECR from *Kitasatospora setae* and to propose a model for the catalytic cycle of this enzyme. The functional homotetramer, composed of two dimers, breaks its symmetry upon cofactor binding and yields a set of two distinct functional dimers. We showed that the communication between and within the dimer is crucial for the fast catalytic rates observed in this ECR. We elucidated how this precise communication functions on the molecular level by a combined approach of analytical methods and steady state analysis. A separate study aimed at developing an *in vivo* directed

evolution screen to improve the catalytic properties of a slow ECR from *Burkholderia ambifaria*. An ECR library was expressed in an organism lacking the genomic *ecr*, and under specific conditions, growth of the organism was dependent on a functional ECR. Through our approach we obtained an improved variant, which revealed mutations distant from the active site. These have a major impact on the catalytic rate of these enzymes and represent new structural hotspots of this enzyme family, which were previously unknown. Both studies revealed the importance of the residues at the interface of the ECR monomers by their high impact on catalytic rates of this enzyme.

1.4. References

1.   Le Quere C*, et al.*, Global Carbon Budget 2018. *Earth Syst Sci Data* 10(4):2141-2194 **2018**.
2.   Pierrehumbert RT, Infrared radiation and planetary temperature. *Phys Today* 64(1):33-38 **2011**.
3.   Nicholls RJ & Cazenave A, Sea-level rise and its impact on coastal zones. *Science* 328(5985):1517-1520 **2010**.
4.   Sabine CL*, et al.*, The oceanic sink for anthropogenic CO2. *Science* 305(5682):367-371 **2004**.
5.   Riebesell U, Climate change: Acid test for marine biodiversity. *Nature* 454(7200):46-47 **2008**.
6.   van Vuuren DP*, et al.*, Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies. *Nat Clim Change* **2018**.
7.   Alper E & Orhan OY, CO_2 utilization: Developments in conversion processes. *Petroleum* 3(1):109-126 **2017**.
8.   Phillips R & Milo R, A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. U.S.A.* 106(51):21465-21471 **2009**.
9.   Evans MC, Buchanan BB, & Arnon DI, A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 55(4):928-934 **1966**.
10.   Calvin M & Massini P, The path of carbon in photosynthesis. XX. The steady state. *Experientia* 8(12):445-457 **1952**.
11.   Berg IA, Kockelkorn D, Buckel W, & Fuchs G, A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318(5857):1782-1786 **2007**.
12.   Pandey AS, Mulder DW, Ensign SA, & Peters JW, Structural basis for carbon dioxide binding by 2-ketopropyl coenzyme M oxidoreductase/carboxylase. *FEBS Lett.* 585(3):459-464 **2011**.
13.   Tong L, Acetyl-coenzyme A carboxylase: crucial metabolic enzyme and attractive target for drug discovery. *Cell. Mol. Life Sci.* 62(16):1784-1803 **2005**.
14.   Tong L, Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* 70(5):863-891 **2013**.
15.   Erb TJ, Brecht V, Fuchs G, Muller M, & Alber BE, Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Natl. Acad. Sci. U.S.A.* 106(22):8871-8876 **2009**.
16.   Wilson MC & Moore BS, Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* 29(1):72-86 **2012**.
17.   Liu Q, Wu L, Jackstell R, & Beller M, Using carbon dioxide as a building block in organic synthesis. *Nature communications* 6:5933 **2015**.
18.   Hamed RB, Batchelar ET, Clifton IJ, & Schofield CJ, Mechanisms and structures of crotonase superfamily enzymes--how nature controls enolate and oxyanion reactivity. *Cell. Mol. Life Sci.* 65(16):2507-2527 **2008**.
19.   Boyd JM & Ensign SA, ATP-dependent enolization of acetone by acetone carboxylase from Rhodobacter capsulatus. *Biochemistry* 44(23):8543-8553 **2005**.
20.   Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, & Lorimer GH, Mechanism of Rubisco: The Carbamate as General Base. *Chem. Rev.* 98(2):549-562 **1998**.

21. Rosenthal RG*, et al.*, Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* 10(1):50-55 **2014**.

22. Appel AM*, et al.*, Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO2 fixation. *Chem. Rev.* 113(8):6621-6658 **2013**.

23. Janc JW, Cleland WW, & O'Leary MH, Mechanistic studies of phosphoenolpyruvate carboxylase from Zea mays utilizing formate as an alternate substrate for bicarbonate. *Biochemistry* 31(28):6441-6446 **1992**.

24. Cotelesage JJH*, et al.*, How does an enzyme recognize CO2? *Int. J. Biochem. Cell. B.* 39(6):1204-1210 **2007**.

25. Chabriere E*, et al.*, Crystal structure of the free radical intermediate of pyruvate:ferredoxin oxidoreductase. *Science* 294(5551):2559-2563 **2001**.

26. Phillips JD, Whitby FG, Kushner JP, & Hill CP, Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase. *EMBO J.* 22(23):6225-6233 **2003**.

27. Lee HJ*, et al.*, Kinetic and crystallographic studies on deacetoxycephalosporin C synthase (DAOCS). *J. Mol. Biol.* 308(5):937-948 **2001**.

28. Turmo A, Gonzalez-Esquer CR, & Kerfeld CA, Carboxysomes: metabolic modules for CO2 fixation. *FEMS Microbiol. Lett.* 364(18) **2017**.

29. Rae BD, Long BM, Badger MR, & Price GD, Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77(3):357-379 **2013**.

30. Zarzycki J, Axen SD, Kinney JN, & Kerfeld CA, Cyanobacterial-based approaches to improving photosynthesis in plants. *J. Exp. Bot.* 64(3):787-798 **2013**.

31. Peter DM*, et al.*, Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13457-13461 **2015**.

32. Erb TJ*, et al.*, Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A.* 104(25):10631-10636 **2007**.

33. Bar-Even A, Noor E, Lewis NE, & Milo R, Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A.* 107(19):8889-8894 **2010**.

34. Rosenthal RG*, et al.*, The use of ene adducts to study and engineer enoyl-thioester reductases. *Nat. Chem. Biol.* 11(6):398-400 **2015**.

35. Schada von Borzyskowski L, Rosenthal RG, & Erb TJ, Evolutionary history and biotechnological future of carboxylases. *J. Biotechnol.* 168(3):243-251 **2013**.

36. Quade N, Huo L, Rachid S, Heinz DW, & Muller R, Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* 8(1):117-124 **2011**.

37. Vogeli B*, et al.*, InhA, the enoyl-thioester reductase from Mycobacterium tuberculosis forms a covalent adduct during catalysis. *J. Biol. Chem.* 293(44):17200-17207 **2018**.

38. Sandor R*, et al.*, Sanguinarine is reduced by NADH through a covalent adduct. *Phytochemistry* 145:77-84 **2018**.

39. Khare D*, et al.*, Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure* 23(12):2213-2223 **2015**.

40. Vogeli B*, et al.*, Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem. Biol.* 25(7):833-839 e834 **2018**.

41.   Zhang L*, et al.*, Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13462-13465 **2015**.

42.   Yoshikuni Y, Ferrin TE, & Keasling JD, Designed divergent evolution of enzyme function. *Nature* 440(7087):1078-1082 **2006**.

43.   Bar-Even A*, et al.*, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50(21):4402-4410 **2011**.

44.   Varnado CL, Hertwig K, Thomas R, Roberts JK, & Goodwin DC, Spectral and kinetic properties of a novel periplasmic catalase-peroxidase. *Biochemistry* 42(28):8621-8622 **2003**.

45.   Wistrand J, Lindahl S, & Wahlstrand T, Human renal carbonic anhydrase. Purification and properties. *Eur. J. Biochem.* 57(1):189-195 **1975**.

46.   Park JO*, et al.*, Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* 12(7):482-489 **2016**.

47.   Tawfik DS, Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21:73-80 **2014**.

CHAPTER II

Four amino acids define the CO$_2$ binding pocket of enoyl-CoA carboxylases/reductases

2.1. Abstract

Carboxylases are biocatalysts that capture and convert carbon dioxide (CO_2) under mild conditions and atmospheric concentrations at a scale of more than 400 Gt annually. However, how these enzymes bind and control the gaseous CO_2 molecule during catalysis is only poorly understood. One of the most efficient classes of carboxylating enzymes are enoyl-CoA carboxylases/reductases (Ecrs), which outcompete the plant enzyme RubisCO in catalytic efficiency and fidelity by more than an order of magnitude. Here we investigated the interactions of CO_2 within the active site of Ecr from *Kitasatospora setae*. Combining experimental biochemistry, protein crystallography and advanced computer simulations we show that four amino acids, N81, F170, E171 and H365 are required to create a highly efficient CO_2-fixing enzyme. Together, these four residues anchor and position the CO_2 molecule for the attack by a reactive enolate created during the catalytic cycle. Notably, a highly ordered water molecule plays an important role in an active site that is otherwise carefully shielded from water, which is detrimental to CO_2-fixation. Altogether our study reveal unprecedented molecular details of selective CO_2 binding and C-C bond formation during the catalytic cycle of nature's most efficient CO_2-fixing enzyme. This knowledge provides the basis for the future development of novel catalytic frameworks for the capture and conversion of CO_2 in biology and chemistry.

Enzyme catalysis | Reaction mechanism | Carboxylase | Enoyl Reductase | Oxidoreductase

Significance

Carboxylases capture and convert CO_2, which makes them key enzymes in photosynthesis and the global carbon cycle. However, the question how enzymes bind atmospheric CO_2 is still unsolved. We studied enoyl-CoA carboxylases/reductases (ECRs) the fastest CO_2-fixing enzymes in Nature, using structural biology, biochemistry and advanced computational methods. ECRs create a highly specific CO_2-binding pocket with four amino acids at the active site. The pocket controls the fate of the gaseous molecule during catalysis and shields the catalytic center from oxygen and water. This exquisite control makes ECRs highly efficient carboxylases outcompeting RubisCO, the key enzyme of photosynthesis, by an order of magnitude. Our findings define the atomic framework for the future development of CO_2-converting catalysts in biology and chemistry.

2.2. Introduction

The efficient capture and conversion of atmospheric carbon dioxide (CO_2) is a prerequisite to develop a carbon-neutral, circular future economy. In biology, carbon fixation is performed under mild conditions and at atmospheric concentrations of CO_2 (0.04 vol%) by enzymes called carboxylases[1]. It is estimated that 400 Gt of CO_2 are fixed annually by a single biocatalyst, ribulose-1,5-bisphophate-carboxylase/oxygenase (RuBisCO), the key enzyme of photosynthesis[2, 3]. In comparison, the chemical conversion of CO_2 in industry accounts for only 0.1 Gt of carbon annually and uses pressurized CO_2[4], which emphasizes our need to understand the molecular mechanism that allow (bio)catalysts to selectively interact with a low-concentrated CO_2 molecule during catalysis.

Carboxylases catalyze the formation of a C-C bond between an acceptor substrate and a CO_2 molecule [1] where the latter represents the electrophile[5]. To facilitate C-C bond formation, most carboxylases activate their respective nucleophilic substrate (usually a thioester, α-ketoacid or ketone) by converting it into an enol(ate)[6]. Enol(ate)s are strong nucleophiles and highly reactive. A key requirement of CO_2-fixation catalysis is the tight control of the reaction between the activated acceptor substrate and CO_2. Any loss of catalytic control over the enol(ate) or the CO_2 molecule bears the danger of side reactions and reduces the efficiency of carbon fixation[7]. The most prominent example is RuBisCO, which is known to feature several side reactions, most notably an oxygenation reaction[8]. One in every five turnovers RuBisCO will incorporate an oxygen (O_2) molecule instead of CO_2, which leads to the formation of 2-phosphoglycolate, a side-product that is toxic to the cell and has to be recycled in an energy demanding process, highlighting the need of carboxylases to control the reaction of the activated acceptor with CO_2.

Another challenge in this respect is the accessibility of water (or protic amino acids) to the active site of carboxylases. Protons are better electrophiles than the CO_2 molecule, which can directly quench the enolate. As a consequence, it is not sufficient that carboxylases enrich a low abundant gaseous CO_2 molecule, they also need to efficiently suppress any competing (re-)protonation reactions. Altogether, these examples show that controlling the fate of CO_2 at the molecular level is a crucial feature of carboxylases. However, to date only very limited

biochemical, let alone structural information on CO_2 binding in carboxylases (and other proteins) is available, besides some theoretical considerations[9-14].

Here, we focused on a class of carboxylases, Enoyl-CoA carboxylases/reductases (Ecrs)[15, 16] that show the fastest turnover frequencies among all carboxylases to date and exclusively react with CO_2 in the presence of O_2. These features make Ecrs excellent model systems to understand the details of selective CO_2 binding and C-C-bond formation in proteins.

The best-studied Ecr is Crotonyl-CoA carboxylase/reductase (Ccr) that catalyzes the NADPH-dependent reductive carboxylation of crotonyl-CoA into (2S)-ethylmalonyl-CoA. While Ccr does not show side reactivity with O_2, the enzyme catalyzes the reduction of crotonyl-CoA to butyryl-CoA as a side reaction, but only in the absence of CO_2 and at low catalytic efficiency (Scheme 1)[17, 18]. It has been suggested that this side reaction is an evolutionary remnant of Ecrs, which are evolutionary related to enoyl thioester reductases (ETRs) that catalyze the ordinary reduction of enoyl-CoA esters[6, 19]. Apparently, Ecrs evolved from simple reductases into reductive carboxylases by acquiring a CO_2-fixation function along their evolutionary trajectory.
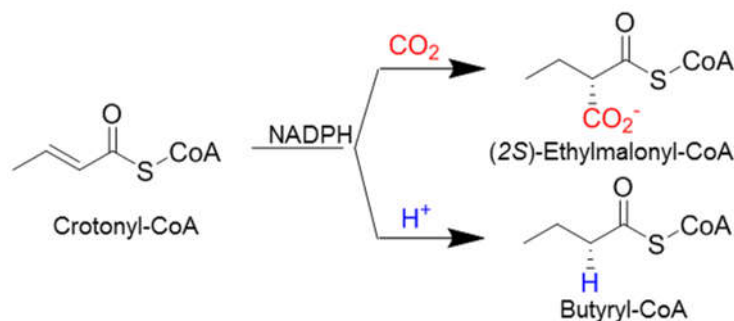
Scheme 1: Reaction catalyzed by Ccr. Carboxylation to (2S)-ethylmalonyl-CoA in the presence of CO_2 and reduction to butyryl-CoA in the absence of CO_2.

In a previous structural study, CinF, an Ecr from *Streptomyces sp. JS360*, was crystallized with NADP$^+$ and octenoyl-CoA (PDB 4A0S[20]). A putative CO_2 binding pocket was proposed to be composed of Asn77, Phe166 and Glu167, which are all highly conserved in Ecrs. It was suggested that CO_2 is held in position by hydrogen bonding to Asn77 and Glu167 while Phe166 would undergo hydrophobic interactions with CO_2. Mutation of these residues suppressed the carboxylation of octenoyl-CoA[20]. However, the exact role of the individual residues directing and controlling the carboxylation reaction remains enigmatic. In particular, how the gaseous

CO_2 molecule is aligned at the active site and how the reduction side reaction is efficiently suppressed.

Here we combine experimental biochemistry, protein crystallography and computer simulations to define the molecular interactions of CO_2 during C-C bond formation at the active site of Ecrs. Our results suggest that four amino acids are sufficient to convert an ordinary reductase into a highly efficient carboxylase. Together, these four residues anchor and lock the CO_2 molecule in a favorable position for the attack by the reactive enolate created during catalysis. Notably, a highly ordered water molecule plays an essential role in coordinating the CO_2 molecule, while the active site is otherwise effectively shielded from water to suppress the reduction side reaction. Altogether, our computational and experimental studies reveal the details of selective CO_2 binding and C-C bond formation in the catalytic cycle of nature's most efficient CO_2-fixing enzyme.

2.3. Results

Crystal structure of KsCcr with Ethylmalonyl-CoA and NADPH
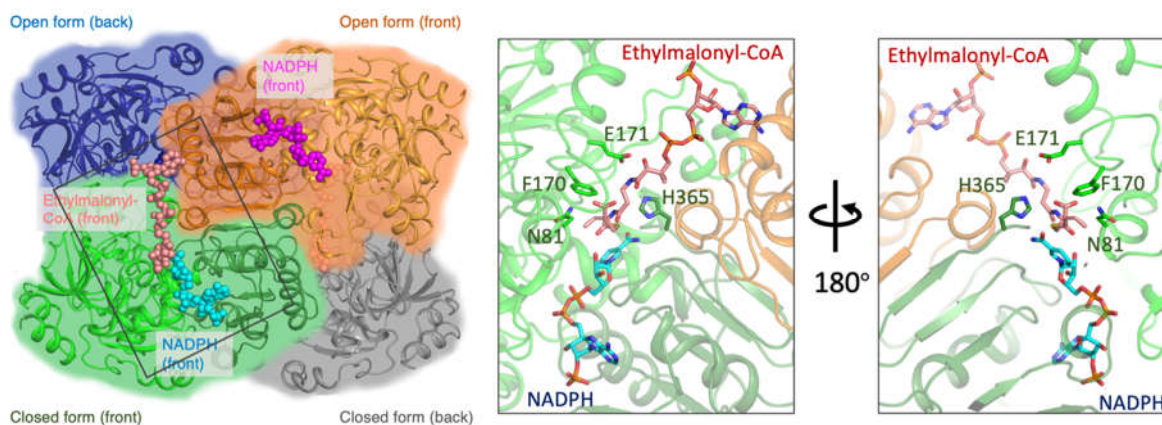
Figure 1: Structure of KsCcr complexed with NADPH and ethylmalonyl-CoA. Left: KsCcr forms a dimer of dimers of open- and closed-form subunits. The subunits are highlighted in grey and green (closed form with NADPH and ethylmalonyl-CoA, both represented in spheres) and blue and orange (open form with NADPH only, represented in spheres). The rectangle represents close-up of the active site shown in the middle and right panel. Middle: the active site with the CO_2-binding residues His365, Glu171, Phe170 and Asn81 in green, ethylmalonyl-CoA (salmon) and NADPH (cyan); oxygen and nitrogen atoms are colored in red and blue, respectively. Right: same as middle panel but rotated by 180 degrees about the viewing direction.

We solved the structure of Ccr from *Kitasatospora setae* (KsCcr) co-crystallized with NADPH and soaked with ethylmalonyl-CoA, the product of the carboxylation reaction, at 1.7 Å

resolution (Fig. 1, SI Appendix, Table S1, PDB: 6OWE). The active site of KsCcr shares similar features with the CinF homologue previously reported[20] (PDB 4A0S[20], 4Y0K[21], 4GI2). In contrast to these structures that did not capture the interaction of the protein with CO_2, neither as free gas or covalently bound to the acyl-CoA moiety, our structure shows densities that can be interpreted as carboxylated product (SI Appendix, Fig. S1). This enabled us to identify four residues that potentially interact with CO_2, namely Asn81, Phe170, Glu171 and His365. His365 also interacts with NADPH via hydrogen bonding (3.0 Å) to the carboxamide group of the nicotinamide, indicating a second function of His365 in coordinating the NADPH cofactor during catalysis. Similar to previously published structures we also observed an ordered water molecule between His365 and Glu171 at a distance of 2.9 Å and 2.7 Å, respectively. This feature is absent in structures lacking substrate and cofactor (PDB 3HZZ, 3KRT) suggesting a role of the water molecule in the active enzyme complex. To test the function of these residues during catalysis, we characterized different active site variants and addressed the reaction mechanism with molecular dynamics simulations along the minimum free energy path within the QM/MM methodology.

Kinetic characterization of KsCcr WT

KsCcr WT showed an apparent turnover frequency (k_{cat}) of 103 ± 3 s^{-1} which is well in line with previously reported value of 104 s^{-1} for the Ccr of *Rhodobacter sphaeroides*[16] . Apparent K_M values where 21 ± 2 μM (crotonyl-CoA), 37 ± 4 μM (NADPH) and 90 ± 10 μM (CO_2), respectively, and substrate inhibition for crotonyl-CoA was observed at a K_i of $3{,}650 \pm 810$ μM. Under saturating amounts of CO_2, the enzyme showed 100% carboxylation activity and exclusively formed (2*S*)-ethylmalonyl-CoA. In the absence of CO_2, the enzyme catalyzed the reduction of crotonyl-CoA to butyryl-CoA (Table 1). Stereochemical analysis of the butyryl-CoA in D_2O showed that 94 ± 2% of the deuterium label was retained (Table 1), demonstrating that the reduction side reaction took place in a stereospecific manner.

Table 1: Apparent steady state parameters for KsCcr and its mutants expressed as mean value ± standard error.

Enzyme	Substrate	K_M (µM)	K_i (µM)	k_{cat} (s⁻¹)	% EMC[b]	% Label retention[d]
Wild-type (WT)	Crotonyl-CoA	21 ± 2	3650 ± 810	103 ± 3	100	94.3 ± 1.8
	NADPH	37 ± 4	-	86 ± 2		
	CO_2[a]	90 ± 10	-	78 ± 2		
N81L	Crotonyl-CoA	ND[c]	ND	ND	19	58.9 ± 1.1
	NADPH	ND	ND	ND		
F170Y	Crotonyl-CoA	10 ± 1	558 ± 80	83 ± 4	100	ND
	NADPH	36 ± 3	-	56 ± 1		
	CO_2[a]	150 ± 20	-	56 ± 2		
F170A	Crotonyl-CoA	31 ± 6	-	8.3 ± 0.4	17	87 ± 0.6
	NADPH	11 ± 0.6	-	11 ± 0.1		
H365N	Crotonyl-CoA	29.8 ± 4.2	-	5.0 ± 0.2	93	63.3 ± 0.5
	NADPH	22 ± 2	-	8.1 ± 0.3		
	CO_2[a]	1310 ± 220	-	7.4 ± 0.7		
E171A	Crotonyl-CoA	500 ± 62	-	5.1 ± 0.2	97	91.4 ± 0.2
	NADPH	112 ± 8	-	6.0 ± 0.2		
	CO_2	155 ± 30	-	5.1 ± 0.3		

SI Appendix, Fig. S2 shows the Michaelis-Menten graphs of the original data.[a] Apparent K_M values for CO_2 were calculated from the HCO_3^- concentration in solution at pH = 8.[b] Percentage of (2S)-ethylmalonyl-CoA over total amount of products. [c] Not determined due to accumulation of the covalent C2-ene adduct.[d] Deuterium label retention at the α-position of crotonyl-CoA expressed as mean value ± standard deviation.

Asn81 anchors the CO_2 molecule

How is the CO_2 molecule bound in the active site? A key residue is Asn81, which defines one end of the putative CO_2 binding pocket. Simulations of the WT enzyme exhibited a hydrogen bond interaction between the carboxamide NH_2 group of Asn81 and the CO_2 molecule (SI Appendix, Supplementary Video 1 and Fig. 2, panel A). When we experimentally characterized the reaction of the N81L variant in more detail, we observed a strongly decreased carboxylation reaction, as well as accumulation of a covalent reaction intermediate, a so-called C2-ene adduct (SI Appendix, Fig. S3). C2-ene adducts are also observed in WT Ecrs when the catalytic cycle is stalled, for example when CO_2 is omitted from the reaction mixture. The fact that a C2-ene adduct is observed in the reaction of the N81L variant even under saturating CO_2 conditions suggests that the interaction of the enzyme with CO_2 is severely disturbed by the N81L mutation. Simulations of the N81L variant revealed that most of the catalytic residues and water molecules remain in the same position, while the CO_2 molecule appears increasingly disordered (Fig. 2, panel D, SI Appendix, Supplementary Video 2). As a consequence, the minimum energy profile from the C2-adduct to the product (2S)-

ethylmalonyl-CoA of N81L becomes endothermic and shows significantly higher barriers for the C-C bond formation step compared to the WT. This explains the reduced carboxylation activity, as well as the accumulation of the C2-ene adduct that we experimentally observed in the N81L variant (Fig. 2, panel C and F).
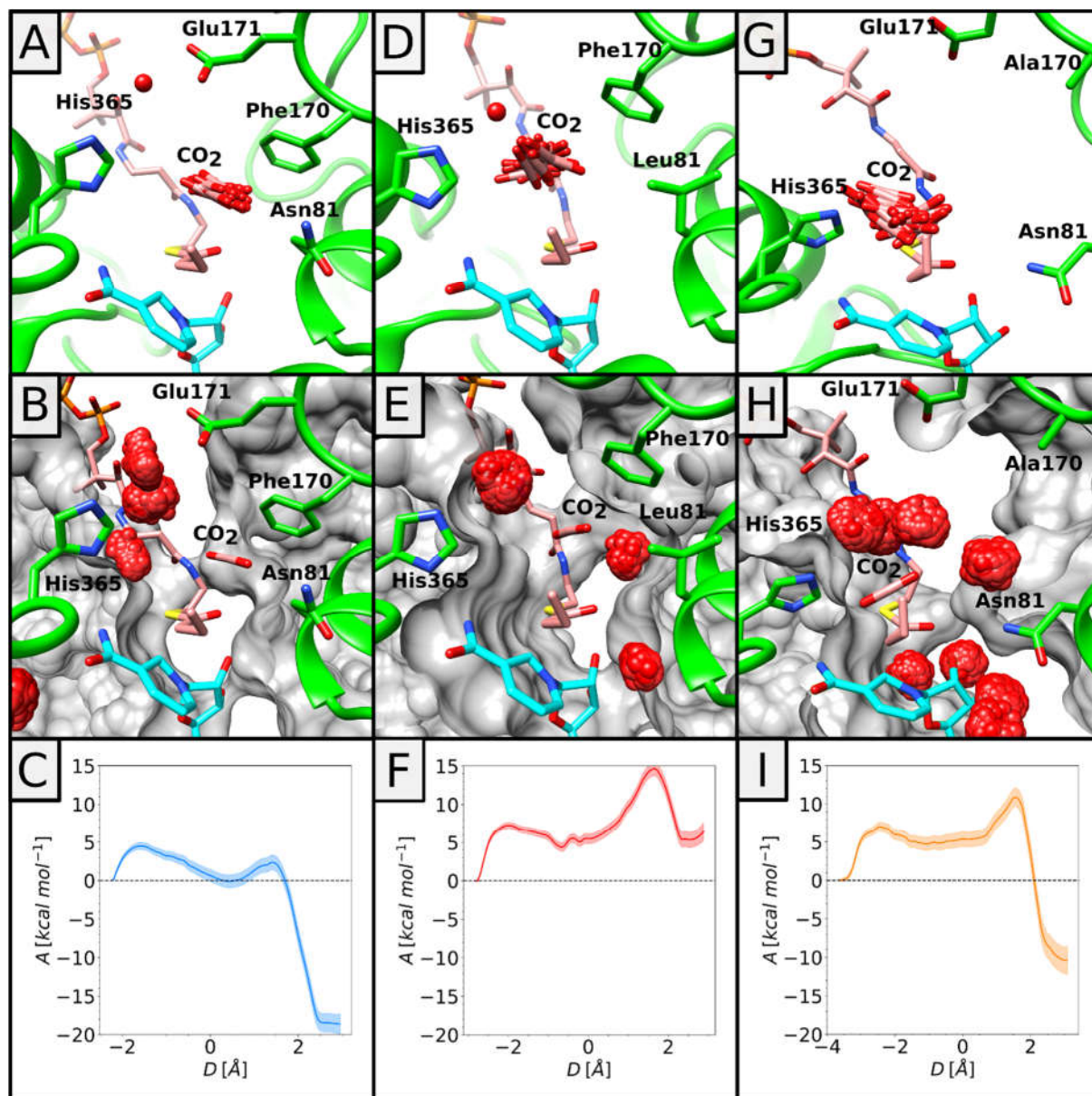

Figure 2: Left, middle and right column represent the WT, N81L and F170A variants respectively. **A**, **D**, **G**: Active site of KsCcr variants prior to the nucleophilic attack of the enolate onto CO_2. An overlay of different conformations of the CO_2 molecule visualizes the tumbling motion in the different enzyme variants. CoA-ester is shown in salmon and NADP$^+$ in cyan, and the red sphere represents the conserved water molecule coordinated by His365 and Glu171. **B**, **E**, **H**: Hydration sites located within 5 Å of the C_α obtained with SSTMap. For each enzyme, representative structures of the reactant state were subjected to 1 ns of constrained QM/MM simulations. The different positions sampled by the water molecules allowed the determination of clusters showing the preferential location of the solvent within the active site. **C**, **F**, **I** Minimum free energy

path from the C2-ene adduct through the enolate to the products (2S)-ethylmalonyl-CoA and NADP$^+$. Parameter D represents the distance between C_α and C_2 minus the distance between C_α and CO_2.

Our simulations show that the active site of N81L does not become more solvent accessible, so that the reduction side reaction is still suppressed at the enzyme's active site. But why does N81L show 89% butyryl-CoA formation? Note that C2-ene adducts are unstable and prone to spontaneous decay into butyryl-CoA and NADP$^+$ in free solution. Accordingly, the "apparent" side reaction in N81L is non-enzymatic and caused by release of the C2-ene
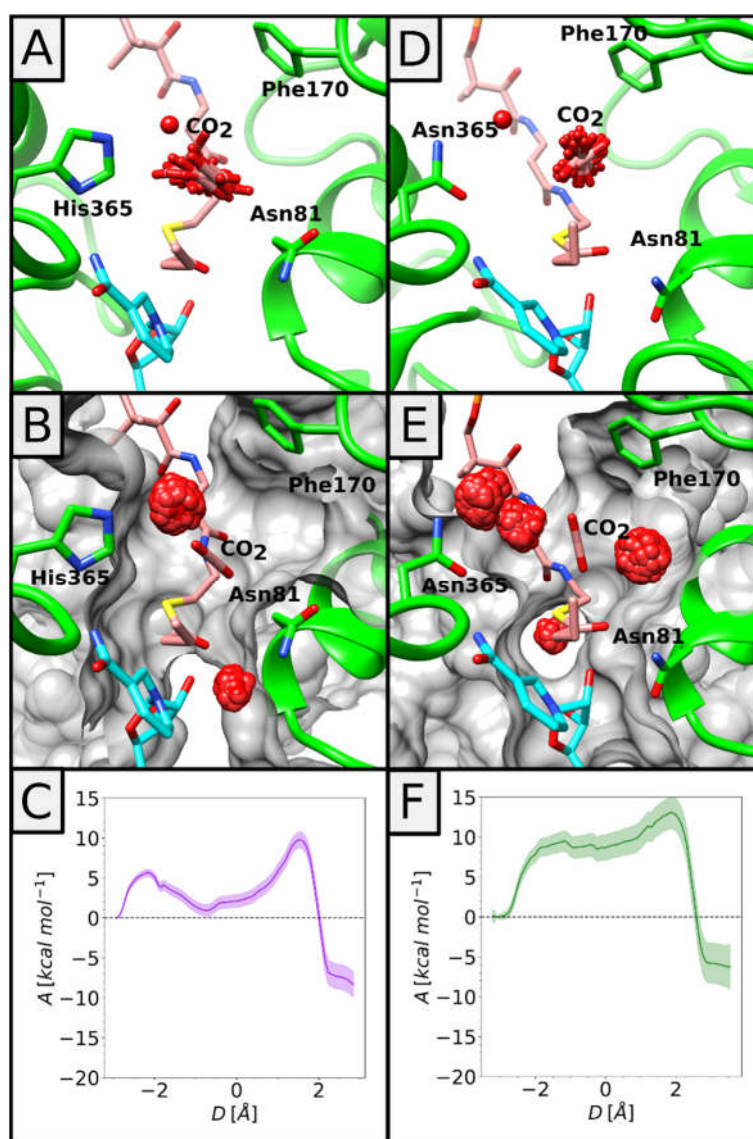
Figure 3: Left and right column represent the E171A and H365 N variants respectively **A**, **D**: Active site of KsCcr variants prior to the nucleophilic attack of the enolate onto CO_2. The CoA-ester is shown in salmon and NADP$^+$ in cyan. Different orientations of carbon dioxide in the simulations represent the tumbling motion of the molecule prior to product formation. **B**, **E**: Hydration sites located within 5 Å of the C_α obtained with SSTMap. **C**, **F**: Minimum free energy paths from the C2-ene adduct through the enolate to the products (2S)-ethylmalonyl-CoA and NADP$^+$. Parameter D represents the distance between C_α and C_2 minus the distance between C_α and CO_2.

adduct from the active site followed by its spontaneous decay in the solvent, as described before[18, 22, 23]. This hypothesis is supported by our observation that stereospecificity of butyryl-CoA formation is almost completely lost in N81L compared to the WT (Table 1). Altogether, our experimental findings are well in line with the higher calculated free energy barriers for the carboxylation step and endothermic product formation in the N81L variant (Fig. 2, panel F), highlighting how crucial Asn81 is for correct positioning of CO_2.

How is the interaction of the carboxamide group of Asn81 with CO_2 controlled? Analysis of the interaction network of the amino acid shows hydrogen bonding of Asn81 to two residues in the second shell of the active site, Thr82 and Ser119. We hypothesized that these interactions are essential in pointing the carboxamide NH_2 group Asn81 towards the active site to enforce its interaction with the CO_2 molecule. Indeed, the variants T82D and S119A showed almost full carboxylation, but at more than 50-fold reduced turnover frequency (SI Appendix, Supplementary Table 3), demonstrating the importance of these secondary shell residues in increasing catalytic activity of Asn81. In summary, both simulation and experimental data suggest that Asn81 is crucial to position CO_2 and to establish favorable interactions of the gas molecule with the enzyme during catalysis. Absence of this residue leads to lowered carboxylation efficiency and increased formation of the labile C2-ene adduct, which is not further processed by the enzyme and leaves the active site upon which it spontaneously decays in solution.

Phe170 shields the active site from water

The reactive nature of the enolate that is formed during the catalytic cycle of Ccr mandates that the enolate does not get into contact with water at the active site, which would inevitably lead to its protonation and formation of the butyryl-CoA side product. A role in shielding the active site from water had been previously suggested for Phe170[20]. Simulations of a F170A variant reveal conformational changes that result in a disorganization of the active site and an increase in the number of hydration sites compared to the WT (Fig. 2, compare panel B and H, SI Appendix, Supplementary Video 3 and Supplementary Pymol and Chimera Files 1a & b and 3a & b). These changes also perturb the interaction of the CO_2 molecule with Asn81 (Fig. 2, panel G). Accordingly, the CO_2 molecule loses its favorable position for the reaction with the enolate so that both activation barriers are increased (Fig. 2, panel I), which is in agreement with our experimental data (Table 1). In the F170A variant, the carboxylation

activity is decreased to 17% at the expense of increased reduction side reactivity. Unlike in the N81L variant, however, protonation takes place in F170A with almost WT stereospecificity (see Table 1), confirming that in the F170A enzyme water is able to reach the active site and directly protonate the enolate.

Some Ecrs feature a tyrosine at position 170 instead of the phenylalanine. When we tested a F170Y variant, the enzyme showed a slightly increased substrate inhibition, but otherwise very similar kinetic parameters as the WT. Most importantly, the F170Y variant displayed full conversion of crotonyl-CoA to (2S)-ethylmalonyl-CoA in the presence of saturating amounts of CO_2, indicating that the presence of the hydroxyl group does not affect carboxylation activity. In summary, these experiments together with the simulations showed that the phenyl rings of phenylalanine (and tyrosine) play an important role in suppressing the reduction reaction of Ecrs by water shielding.

His365 and Glu171 coordinate an ordered water molecule interacting with CO_2

Opposite of Asn81 and at the other end of the putative CO_2 binding pocket the residues His365 and Glu171 are located. Together, these two residues coordinate a water molecule. In our simulations, the ordered water molecule participates in hydrogen bonded network of three water molecules, which interact directly with the CO_2 molecule during the carboxylation step (Fig. 3). What is the exact contribution of these two residues to catalysis, in particular in respect to CO_2 binding and water accessibility?

His365 serves a dual role by also coordinating the nicotinamide ring of NADPH. To preserve interaction of residue 365 with the NADPH cofactor, but interrupt its coordination of the ordered water molecule, we generated KsCcr H365N. The H365N variant showed a 20-fold decreased activity compared to the WT enzyme but still displayed 93% of carboxylated product, even though the K_M for CO_2 was raised by more than one order of magnitude. This data suggests that a defect in water coordination negatively affects C-C bond formation activity in the H365N variant. However, this does not lead to a complete hydration of the active site. Simulations of the H365N mutant revealed broken interactions of the CO_2 molecule with Asn81 (Fig. 3, panel D, SI Appendix, Supplementary Video 4). The coordination to the ordered water molecule that bridges to Glu171 is lost, and the latter residue is rotated out of the active site disfavoring CO_2 binding, which explains the experimentally observed increased K_M for CO_2. In our simulations, the CO_2 molecule shows an increased rotational tumbling at

the active site of the H365N variant compared to the WT, which is reflected by an increased RMSD for CO_2 (2.4 Å versus 0.46 Å). Together, experiment and simulations indicated reduced carboxylation efficiency because of decreased control of CO_2 (but not water) at the active site in the H365N variant, lowering the chances of productive Michaelis complex formation.

Similar to the H365N variant, replacement of Glu171 with alanine also resulted in a k_{cat} decrease (17-fold), and carboxylation of crotonyl-CoA was also maintained (97%). Additionally, however, the K_M for crotonyl-CoA increased 25-fold, indicating an additional role of Glu171 in positioning the CoA-substrate. In our simulations of the E171A mutant, interaction of CO_2 with Asn81 is partially maintained, while Phe170 adopts the position of the mutated glutamate in the WT (Fig. 3, panel A, SI Appendix, Supplementary Video 5). As for H365N, the E171A variant showed increased CO_2 tumbling at the active site, which is reflected by a RMSD of 2.0 Å (compared to 0.46 Å of the WT). Altogether, these results suggested that in the E171A variant, similar to the H365N variant, control of the CO_2 molecule (but not water) is affected.

While H365N and E171A show a similar carboxylation behavior, the free energy paths of the two variants show distinct differences (Fig. 3, panel C and F). In both variants, the two main barriers along the minimum free energy path of the reaction – the one that leads to the enolate from the C2-ene adduct and the one that adds CO_2 to the enolate – appear increased compared to the WT (for WT see Fig.2, panel C). This is experimentally reflected by the decreased catalytic activity of the two variants and can be related to a distorted water network and increased tumbling of CO_2 at the active site of these enzymes. In H365N, however, the first energy barrier is higher compared to E171A. This indicates that enolate formation is disfavored and suggests that the C2-ene adduct accumulates in the H365N variant. This was experimentally confirmed by measuring the stereospecificity of the reduction reaction in the absence of CO_2. In the H365N variant, stereospecificity was lost (see Table 1), indicating that the H365N is additionally affected in enolate formation compared to the E171A mutant and the WT, respectively.

In summary, His365 and Glu171 can partially compensate each other so that carboxylation function is maintained. However only the combined action of the two residues allows full control over the CO_2 molecule and thus a fast carboxylation rate.

2.4. Discussion

Ccr from *Kitasatospora setae* carboxylates crotonyl-CoA at a turnover frequency of more than 100 s^{-1}. This is almost one order of magnitude faster than an average RubisCO homolog and one of the fastest CO_2-conversion rates described to date. Combining X-ray crystallography, experimental biochemistry and molecular dynamics simulations we characterized the role of individual amino acids at the active site of KsCcr in CO_2 binding and C-C-bond formation. In KsCcr the active site is optimized to accommodate CO_2 and at the same time exclude water to suppress the competing reduction reaction. All this is apparently achieved by only four amino acids: Asn81, His365, Glu171 and Phe170. The amide group of Asn81 is responsible for anchoring the CO_2 from one side, while a water network organized from an ordered water molecule coordinated between His365 and Glu171 serves as an additional anchor point for CO_2 from the opposite side. The aromatic ring of Phe170 side chain actively prevents the diffusion of water into the active site.

Why is the CO_2-fixation reaction of Ecrs so much faster but still more specific compared to the reaction catalyzed by RubisCO, although both enzymes react through an enolate? Note that there are fundamental differences in the catalytic mechanisms of RubisCO and Ecrs. In RuBisCO, enolization of the substrate ribulose 1,5-bisphosphate is achieved by abstraction of the H3 proton through an active site carbamylated lysine, which is a reversible process and very close to equilibrium[24]. (Re-)protonation of the enolate yields the substrate again, which can undergo another round of activation until it reacts with a CO_2 (or O_2) molecule, which pulls the reaction further[25]. In contrast, in Ecrs the enolate is formed by hydride transfer from NADPH to the β-position of crotonyl-CoA[17-19]. This provides a more "unidirectional" reaction path and leaves the enolate committed for a nucleophilic attack. Accordingly, Ecrs are required to carefully control the further fate of the enolate and especially to prevent its protonation, which would irreversibly quench the reaction (in stark contrast to the case of RubisCO). When CO_2 is absent from the active site of WT Ecr[18], or its positioning is disturbed (e.g. by H365N or N81L mutation, this study), the enolate is not simply transformed back into the starting substrates as in RubisCO, but tends to collapse into the C2-ene adduct, which is in line with the idea of "unidirectionality" in Ecr catalysis. It might be tempting to speculate that the C2-ene adduct serves as a way to "store" the reactive enolate until a resolving CO_2

electrophile becomes available, thereby increasing the overall reactivity of Ecrs compared to RubisCOs.

RubisCO and Ecrs probably both evolved from non-CO_2-fixing ancestors. While it has been speculated that RubisCO emerged from an ancestral sugar phosphate isomerase it is thought that Ecrs evolved from a primordial enoyl-CoA reductase[6, 26]. Clearly, this put more constraints onto the active site topology of Ecrs, because unlike the CO_2 fixation reaction in RubisCO that could be simply built on top of a reversible isomerization reaction, the reduction reaction needed to be suppressed and replaced by the carboxylation reaction in the Ecr scaffold because of the "unidirectionality" of Ecr's catalytic mechanism. The situation is reminiscent of 2-ketopropyl coenzyme M oxidoreductase/carboxylase that evolved within the superfamily of NAD(P)H disulfide oxidoreductases for which the enzyme also had to replace an active site topology prone to reduction reactions by a CO_2-fixing active site[11]. In both cases, the active site transformation was achieved with only little changes and notably without introduction of a competing oxygenation reaction, posing the question whether the evolutionary circumstances or the nature of the catalytic cycle of both enzymes were responsible to achieve this superior selectivity against oxygen compared to RubisCO.

In summary, our findings provide detailed insights into the molecular control of CO_2 at one of nature's most efficient carbon fixing enzymes. These insights will be helpful in the future design of catalytic frameworks for the capture and conversion of CO_2 in chemistry and biology[27], but also for efforts that aim at using Ecrs as key enzymes in the development of synthetic cycles for the sustainable and efficient fixation of CO_2[28, 29].

2.5. Materials and Methods

Chemicals

Crotonic Anhydride and Carbonic anhydrase from bovine erythrocytes were purchased from Sigma Aldrich AG, Coenzyme A trilithium salt and DNAse I from Roche Diagnostics, NADPH Na_4 (98%) from Carl Roth GmbH. Solvents and salts were all analytical grade or better. Crotonyl-CoA was synthesized as previously reported[30].

Cloning and Mutagenesis

The KsCcr gene was provided by the JGI. Enzyme variants were generated with the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA) using primer pairs listed (SI Appendix, Supplementary Table 2).

Protein Expression and Purification

His-tagged protein was expressed in *E. coli* BL21 AI™ (DE3). Cells in terrific broth were grown at 37 °C to an OD_{600} = 0.8-1.0 upon which expression for 12-16 h at 23 °C was induced by the addition of 500 µM IPTG (Isopropyl-D-β-thiogalactopyranoside) and 0.02% L-Arabinose. Cells were harvested for 15 min at 7'500 g at 4 °C then resuspended in 2 mL of Buffer A (50 mM Tris, 500 mM NaCl, 1M L-Proline, pH= 7.5) per gram of pellet. The suspension was treated with 10 mg/mL of DNAse I and 5 mM $MgCl_2$ on ice for 20 min upon which cells were lysed at 16000 Psi using a LM10 Microfluidizer™. The lysate was clarified at 45'000 g at 4°C for 45 min and then loaded onto a pre-equilibrated 1 mL HisTrap FF column and washed with 12 % Buffer B (50 mM Tris, 500 mM NaCl, 1 M L-Proline, 500 mM imidazole, pH = 7.5) for 20-30 column volumes until the UV 280 nm reached the baseline level. The protein was eluted by applying 100% buffer B, collected then pooled and desalted into 12.5 mM Tris, 125 mM NaCl, 1 M L-Proline. The addition of L-Proline to each buffer increased the yields for the protein purification by making the protein more soluble as previously reported[31]. The protein was frozen in N_2 (l) and stored at -80°C if not immediately used for assays.

Quantification of reaction products

Reactions for product analysis of KsCcr contained saturating amounts (at least 10 times the K_M) of Crotonyl-CoA, NADPH, 100 mM K_2HPO_4 pH = 8.0, 50 mM $NaHCO_3$, 2 ug/mL carbonic

anhydrase and each enzyme was added at a final concentration of 1 μM. The reaction procedure was monitored by decrease in absorbance of NADPH at 340 nm, quenched with 10 μL of 50% formic acid at completion and spinned at 17'000 g for 10 min to precipitate the protein. The reaction was diluted 10 times into 5% methanol/Buffer 8.1 and analyzed by UHPLC over a Sonoma C18(2), 3 μm 100 Å, 100 x 2.1 mm using a 5 to 45% methanol gradient over 14.5 min. Quantification was performed by peak integration at 260 nm: ethylmalonyl-(4 min), crotonyl-(7.4 min) and butyryl-CoA (8.6 min).

Spectrophotometric enzyme assays

Assays were performed on a Cary-60 UV/Vis spectrophotometer (Agilent) at 30°C using quartz cuvettes (1 or 10 mm path length; Hellma). Reactions contained 20 μg/mL carbonic anhydrase and were performed in 100 mM K_2HPO_4 pH = 8.0. Kinetic parameters for one substrate were determined by varying its concentration while the others were kept constant at 10 times their K_M value. Reaction procedure was monitored by following the oxidation of NADPH at 365 nm ($\varepsilon_{NADPH,365nm} = 3.33$ mM^{-1} cm^{-1}). Each concentration was measured in triplicates and the obtained curves were fit using GraphPad Prism 8. Hyperbolic curves were fit to the Michaelis-Menten equation to obtain apparent k_{cat} and K_M values. For mutants revealing substrate inhibition, the data was fit to $v_0 = (V_{Max} [S])/(K_M + [S] ((1+[S])/K_i))$.

Determining the stereochemistry of protonation

Isotopic label incorporation experiments were performed analogous to a previously described method (see SI Appendix, Fig. S4)[22].

Molecular Dynamics Simulations

From the crystal structure, chains A and C along with the corresponding cofactors and substrates were solvated inside a cubic box of TIP3P water molecules, with a distance of 12 Å between the enzyme and the edge of the box. To neutralize the system, sodium and chloride ions were added to reach a concentration of 0.125 M. All molecular dynamics simulations were performed with the software Amber16[32] and the force field CHARMM22/CMAP[33, 34]. Details are described in the SI Appendix.

Acknowledgements

2.6. References

1. Erb TJ, Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* 77(24):8466-8477 **2011**.
2. Ellis RJ, Most Abundant Protein in the World. *Trends Biochem. Sci.* 4(11):241-244 **1979**.
3. Phillips R & Milo R, A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. U.S.A.* 106(51):21465-21471 **2009**.
4. Alper E & Orhan OY, CO_2 utilization: Developments in conversion processes. *Petroleum* 3(1):109-126 **2017**.
5. Aresta M & Dibenedetto A, Utilisation of CO_2 as a chemical feedstock: opportunities and challenges. *Dalton Trans.* (28):2975-2992 **2007**.
6. Schada von Borzyskowski L, Rosenthal RG, & Erb TJ, Evolutionary history and biotechnological future of carboxylases. *J. Biotechnol.* 168(3):243-251 **2013**.
7. Edmondson DL, Kane HJ, & Andrews TJ, Substrate Isomerization Inhibits Ribulosebisphospate Carboxylase-Oxygenase during Catalysis. *FEBS Lett.* 260(1):62-66 **1990**.
8. Bowes G, Ogren WL, & Hageman RH, Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem. Biophys. Res. Commun.* 45(3):716-722 **1971**.
9. Prussia GA*, et al.*, Substitution of a conserved catalytic dyad into 2-KPCC causes loss of carboxylation activity. *FEBS Lett.* 590(17):2991-2996 **2016**.
10. Kofoed MA, Wampler DA, Pandey AS, Peters JW, & Ensign SA, Roles of the redox-active disulfide and histidine residues forming a catalytic dyad in reactions catalyzed by 2-ketopropyl coenzyme M oxidoreductase/carboxylase. *J. Bacteriol.* 193(18):4904-4913 **2011**.
11. Pandey AS, Mulder DW, Ensign SA, & Peters JW, Structural basis for carbon dioxide binding by 2-ketopropyl coenzyme M oxidoreductase/carboxylase. *FEBS Lett.* 585(3):459-464 **2011**.
12. Stec B, Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *Proc. Natl. Acad. Sci. U.S.A.* 109(46):18785-18790 **2012**.
13. van Lun M, Hub JS, van der Spoel D, & Andersson I, CO2 and O2 distribution in Rubisco suggests the small subunit functions as a CO2 reservoir. *J. Am. Chem. Soc.* 136(8):3165-3171 **2014**.

14. Ito Y, Kondo H, Shiota Y, & Yoshizawa K, Theoretical analysis of the reaction mechanism of biotin carboxylase. *J. Chem. Theory Comput.* 4(2):366-374 **2008**.

15. Peter DM*, et al.*, Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13457-13461 **2015**.

16. Erb TJ*, et al.*, Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A.* 104(25):10631-10636 **2007**.

17. Erb TJ, Brecht V, Fuchs G, Muller M, & Alber BE, Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Natl. Acad. Sci. U.S.A.* 106(22):8871-8876 **2009**.

18. Rosenthal RG*, et al.*, Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* 10(1):50-55 **2014**.

19. Rosenthal RG*, et al.*, The use of ene adducts to study and engineer enoyl-thioester reductases. *Nat. Chem. Biol.* 11(6):398-400 **2015**.

20. Quade N, Huo L, Rachid S, Heinz DW, & Muller R, Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* 8(1):117-124 **2011**.

21. Zhang L*, et al.*, Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13462-13465 **2015**.

22. Vogeli B*, et al.*, InhA, the enoyl-thioester reductase from Mycobacterium tuberculosis forms a covalent adduct during catalysis. *J. Biol. Chem.* 293(44):17200-17207 **2018**.

23. Libby RD & Mehl RA, Characterization of covalent Ene adduct intermediates in "hydride equivalent" transfers in a dihydropyridine model for NADH reduction reactions. *Bioorg. Chem.* 40(1):57-66 **2012**.

24. Tcherkez G, Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters. *Plant Cell Environ.* 36(9):1586-1596 **2013**.

25. Tcherkez GG, Farquhar GD, & Andrews TJ, Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U.S.A.* 103(19):7246-7251 **2006**.

26. Erb TJ & Zarzycki J, A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Curr. Opin. Biotechnol.* 49:100-107 **2018**.

27. Appel AM*, et al.*, Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO2 fixation. *Chem. Rev.* 113(8):6621-6658 **2013**.

28. Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS, & Erb TJ, A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354(6314):900-904 **2016**.

29. Bar-Even A, Noor E, Lewis NE, & Milo R, Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U.S.A.* 107(19):8889-8894 **2010**.

30. Peter DM, Vogeli B, Cortina NS, & Erb TJ, A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters. *Molecules* 21(4):517 **2016**.

31. Samuel D, Kumar TK, Jayaraman G, Yang PW, & Yu C, Proline is a protein solubilizing solute. *Biochem. Mol. Biol. Int.* 41(2):235-242 **1997**.

32. Case DA*, et al.* **2017** Amber 2017.

33. MacKerell AD*, et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616 **1998**.

34. Mackerell AD, Jr., Feig M, & Brooks CL, 3rd, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in

reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11):1400-1415 **2004**.

35.   Vogeli B*, et al.*, Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem. Biol.* 25(7):833-839 e834 **2018**.

36.   Kabsch W, Xds. *Acta Crystallogr. D* 66:125-132 **2010**.

37.   Kabsch W, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* 66:133-144 **2010**.

38.   Adams PD*, et al.*, PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* 58:1948-1954 **2002**.

39.   Adams PD*, et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66:213-221 **2010**.

40.   Mccoy AJ*, et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* 40:658-674 **2007**.

41.   McCoy AJ, Acknowledging Errors: Advanced Molecular Replacement with Phaser. *Methods Mol. Biol.* 1607:421-453 **2017**.

42.   Waterhouse A*, et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46(W1):W296-W303 **2018**.

43.   Winn MD, Isupov MN, & Murshudov GN, Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* 57(Pt 1):122-133 **2001**.

44.   Winn MD, Murshudov GN, & Papiz MZ, Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* 374:300-321 **2003**.

45.   Emsley P & Cowtan K, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* 60(Pt 12 Pt 1):2126-2132 **2004**.

46.   Emsley P, Lohkamp B, Scott WG, & Cowtan K, Features and development of Coot. *Acta Crystallogr. D* 66(Pt 4):486-501 **2010**.

47.   Pavelites JJ, Gao JL, Bash PA, & Mackerell AD, A molecular mechanics force field for NAD(+), NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem.* 18(2):221-239 **1997**.

48.   Aleksandrov A, Zvereva E, & Field M, The mechanism of citryl-coenzyme A formation catalyzed by citrate synthase. *J. Phys. Chem. B* 118(17):4505-4513 **2014**.

49.   Gaus M, Cui QA, & Elstner M, DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* 7(4):931-948 **2011**.

50.   Gaus M, Goez A, & Elstner M, Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* 9(1):338-354 **2013**.

51.   Gaus M, Lu XY, Elstner M, & Cui Q, Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* 10(4):1518-1537 **2014**.

52.   Lu XY, Gaus M, Elstner M, & Cui Q, Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* 119(3):1062-1082 **2015**.

53.   Zinovjev K & Tunon I, Adaptive Finite Temperature String Method in Collective Variables. *J. Phys. Chem. A* 121(51):9764-9772 **2017**.

54.   Kastner J & Thiel W, Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.* 124(23) **2006**.

55.   Haider K, Cruz A, Ramsey S, Gilson MK, & Kurtzman T, Solvation Structure and Thermodynamic Mapping (SSTMap): An Open-Source, Flexible Package for the Analysis

of Water in Molecular Dynamics Trajectories. *J. Chem. Theory Comput.* 14(1):418-425 **2018**.

2.7. Supplementary Appendix
Supplementary Tables

Supplementary Table 1. Data collection and refinement statistics.

	K. setae ECR NADPH Ethylmalonyl-CoA
Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	109.1, 78.7, 138.9
α, β, γ (°)	90.0, 108.1, 90.0
Resolution (Å)	30 – 1.72
	(1.83 – 1.72)
R_{sym} or R_{merge}	0.15 (0.51)
$I / \sigma I$	4.7 (2.26)
Completeness (%)	97.2 (97.8)
Redundancy	3.34
Refinement	
Resolution (Å)	29.89 – 1.72
	(1.79 – 1.72)
No. reflections	230524 (24347)
R_{work} / R_{free}	0.16/0.17
No. atoms	
Protein	14014
Ligand/ion	324
Water	1112
B-factors	
Protein	25.4
Ligand/ion	35.1
Water	33.5
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.23

*Single crystal used for the dataset.
**Values in parentheses are for highest-resolution shell.

Supplementary Table 2. Forward and reverse primer pairs used for site-directed mutagenesis of the KsCcr gene.

Mutation	Forward Primer	Reverse Primer
N81L	CGTCTTCCGTAAACTATCTGACCGTTTGGTCTTCCATC	GATGGAAGACCAAACGGTCAGATAGTTTACGGAAGACG
N81D	GTCTTCCGTAAACTATGACACCGTTTGGTCTTC	GAAGACCAAACGGTGTCATAGTTTACGGAAGAC
T82D	GTCTTCCGTAAACTATAACGACGTTTGGTCTTCCATCTTC	GAAGATGGAAGACCAAACGTCGTTATAGTTTACGGAAGAC
S119A	CACGTTCTGGGTGCTGATCTGGCTG	CAGCCAGATCAGCACCCAGAACGTG
F170A	GCGCATCTGGGGCGGCGAAACCAACTTTG	CAAAGTTGGTTTCGCCGCCCCAGATGCGC
E171L	CGCATCTGGGGCTTCCTGACCAACTTTGGTGG	CCACCAAAGTTGGTCAGGAAGCCCCAGATGCG
H365N	CGTATCGTAGGCTCTAACTTCGCTAACTATC	GATAGTTAGCGAAGTTAGAGCCTACGATACG

Supplementary Table 3. Apparent steady state parameters for KsCcr S119A and T82D mutants expressed as mean value ± standard error.

Enzyme	Substrate	K_M (µM)	K_i (µM)	k_{cat} (s^{-1})	% EMC[b]
S119A	Crotonyl-CoA	36 ± 5	-	1.79 ± 0.05	98
	NADPH	30 ± 7	-	1.47 ± 0.08	
	CO_2[a]	80 ± 20	-	1.32 ± 0.09	
T82D	Crotonyl-CoA	13.4 ± 2.4	446 ± 63	0.172 ± 0.001	94
	NADPH	52 ± 6	-	0.157 ± 0.005	
	CO_2[a]	30 ± 0.003	-	0.170 ± 0.004	

SI Appendix, Fig. S2 shows the Michaelis-Menten graphs of the original data.[a] Apparent K_M values for CO_2 were calculated from the HCO_3^- concentration in solution at pH = 8.[b] Percentage of (2*S*)-ethylmalonyl-CoA over total amount of products.

Supplementary Figures

Supplementary Figure S1: Simulated annealing Fo-Fc omit-maps at 3.0 σ for NADPH (cyan) and at 1.25 σ for ethylmalonyl-CoA (salmon) bound to KsCcr. A high occupancy was obtained for NADPH, which was co-crystallized with KsCcr. Ethylmalonyl-CoA was soaked into the crystals before freezing resulting in a much lower occupancy. The very weak electron density at the carboxy group of ethylmalonyl-CoA can be explained by enzyme mediated and/or partial spontaneous decarboxylation, as reported before[35].
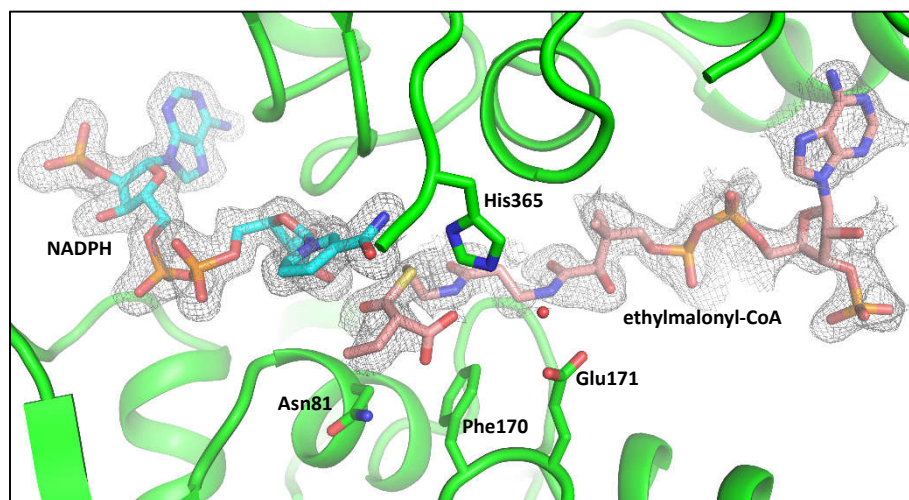
Supplementary Figure S2: Steady state parameters of KsCCR and its variants. The data are summarized in **Table 1.** All reactions contained 50 mM NaHCO₃ and 20 µg/ml of carbonic anhydrase and were performed in 100 mM K₂HPO₄ pH = 8. For The H365N mutant 100 mM NaHCO₃ were used.
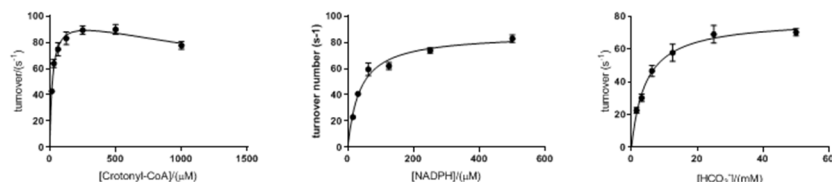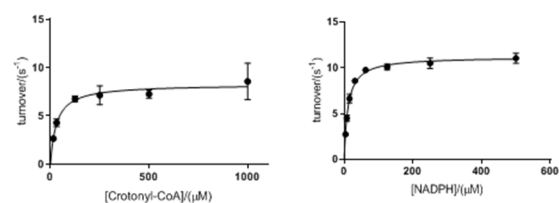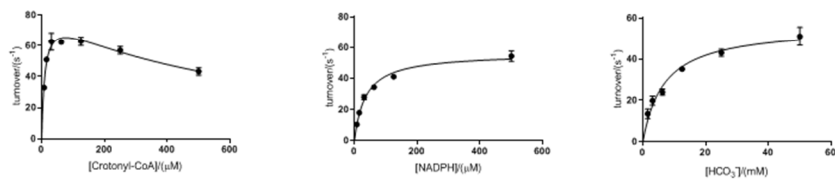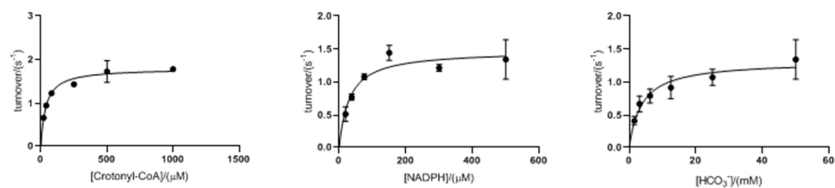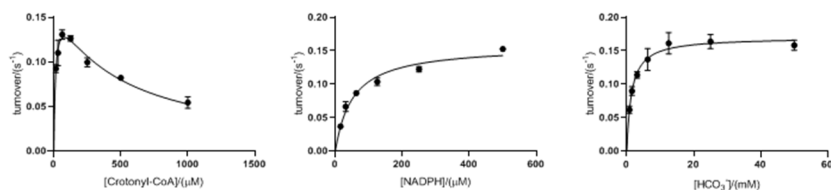
KsCcr E171A
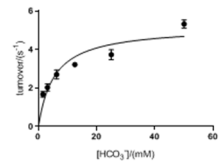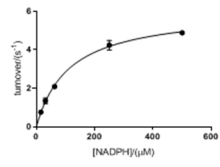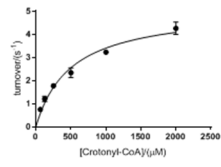
KsCcr H365N
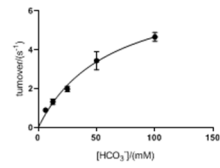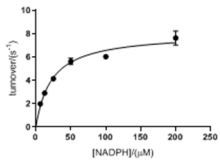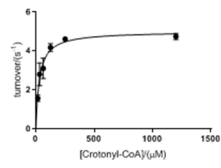
Supplementary Figure S3: Formation of the C2-ene adduct in KsCCR N81L. In the absence or presence of 50 mM HCO_3^- (blue and red trace respectively) KsCcr N81L accumulates the C2-ene adduct. The formation was followed at 385 nm and the reaction contained 300 μM NADPH, 1mM Crotonyl-CoA, 5μM KsCcr N81L and was performed in 100 mM K_2HPO_4 pH = 8 at 30°C.

C2-ene adduct
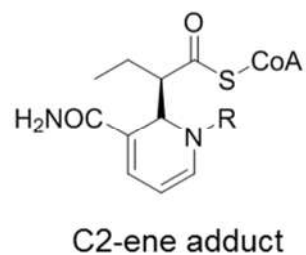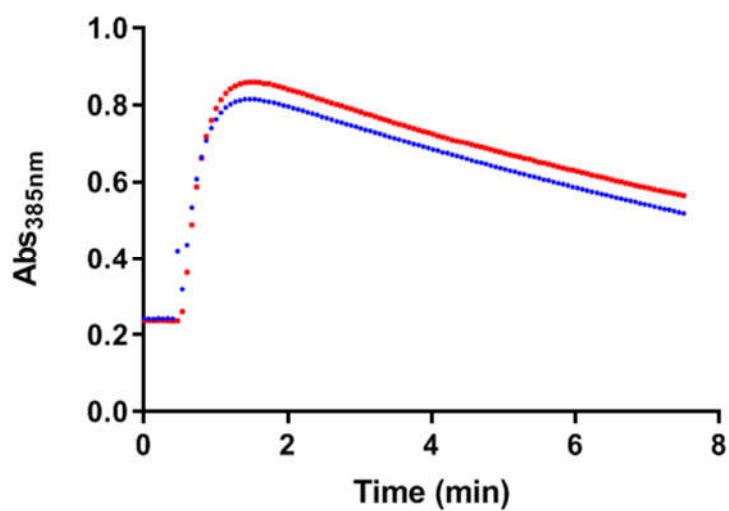
Supplementary Figure S4: Isotopic labelling experiment. The reduction reaction is performed in D_2O in the presence of Ccr and the reaction product is purified by HPLC. The product is then incubated with Acx4 to generate crotonyl-CoA.
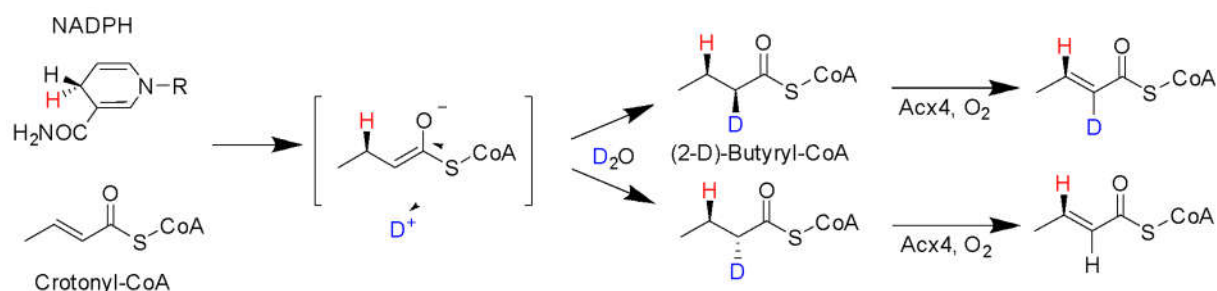
Supplementary Figure S5: Quantum mechanical region used in the QM/MM calculations shown for the C2-adduct representing the reactant state. The indexed atoms are those involved in the collective coordinates. L= link atom.
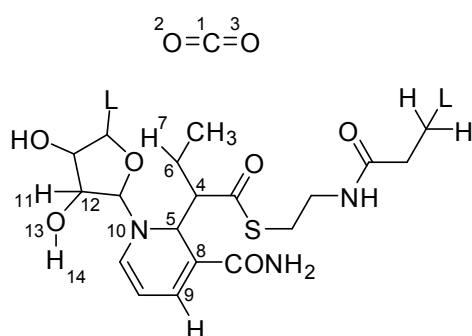
Collective coordinates used to trace the minimum free energy path in the adaptive string method. Asterisks indicate collective coordinates important in energetic terms.

CV type	Atoms involved
Distance between atoms	1-4*, 4-5*, 6-7, 7-9, 4-6, 6-9, 5-8,
Angle	2-1-3
Torsion	11-12-13-14
Point-Plane distance (central atom)	4, 5, 9, 6, 8, 10

Supplementary Figure S6: Dissociation of CO_2 from the product ethylmalonyl-CoA shown as snapshots. Panels **A** to **D** visualize the sequential increase in distance between the C_α of the substrate and the carbon atom of CO_2. Starting from ethylmalonyl-CoA, where the distance between C_α and the carboxylate carbon atom is 1.5 Å, the system reaches the state in panel **D** where the CO_2 carbon atom is 3.6 Å away from C_α and the C2-adduct is formed.
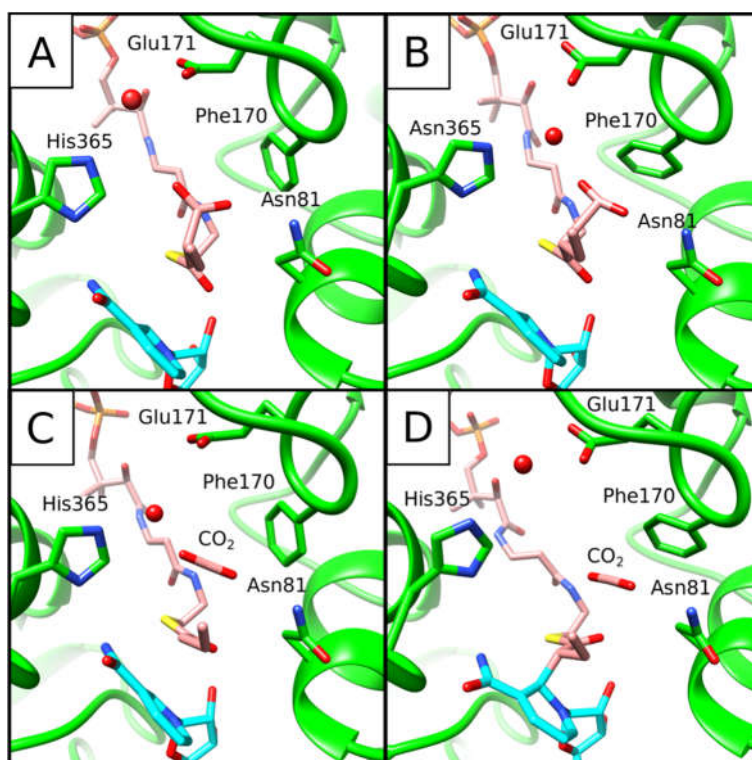
Supplementary Figure S7: Root-mean-squared-deviation (RMSD, left) and Root-mean squared-fluctuation (RMSF, right) of the protein backbone C$_\alpha$. 100 ns of MM equilibration shows an increased but converged RMSD for the F170A variant, compared to the other mutants. The RMSF measured for the last 25 ns of the same equilibration trajectory shows increased fluctuations in F170A and smaller variations in the E171A mutant.
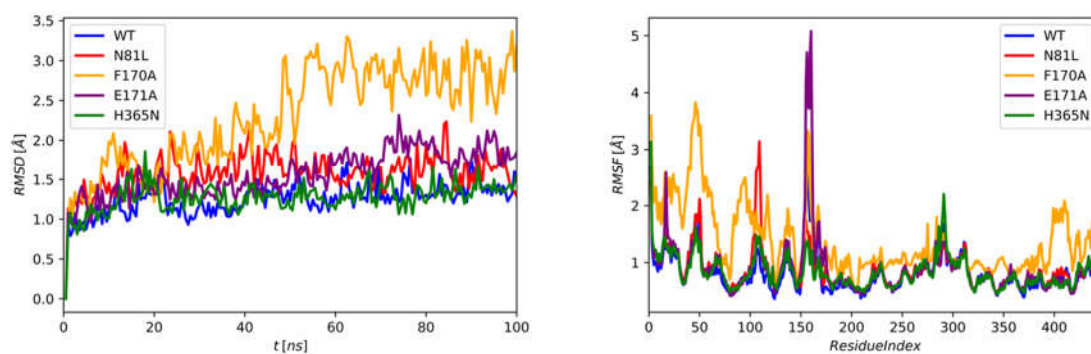
Protein Crystallization and Structural Analysis

72-well sitting-drop crystallization trays (Terasaki) were set up and screened against a library of various crystallization conditions (Molecular Dimensions, Hampton). Each crystallization well contained 0.77 µl of 10 mg/ml *K. setae* ECR protein kept in 500 mM Imidazole, 300mM NaCl, 1M proline and TRIS-HCl pH = 8.5 mixed with 0.77 µL of the various crystallization buffers. Each well was sealed with 16.6 µL of 100% paraffin oil (Hampton Research) to slow the crystallization process. Crystals of apo ECR protein were observed in various morphologies after 24 hours of incubation. The initial crystallization conditions were from various MIDAS, Crystal Screen, and PGA-LM screening conditions (Molecular Dimensions, Hampton Research). The apo ECR was crystallized from a solution containing 100 mM TRIS pH = 8.0 and 20% w/v poly (acrylic acid sodium salt) 5100 and resulted in 30-micron plate-like crystals. The ternary ECR complex were co-crystallized with final concentration of 5 mM of NADPH and soaked with 5 mM ethylmalonyl-CoA. Crystals were harvested after 30 minutes incubation with 30% (v/v) glycerol as a cryoprotectant and then flash cooled in liquid nitrogen.

Diffraction data of the ternary complex with NADPH and ethylmalonyl-CoA were collected on 23-ID-B beamline of The Advanced Photon Source (APS). The *K. setae* ternary complex belonged to the space group $P2_1$ with unit cell dimensions a= 109.1 Å, b= 78.7 Å, c= 138.9 Å and $\alpha = \gamma = 90°$ $\beta = 108.1°$. The data processing for synchrotron structures were carried out using *autoXDS* and scaling was done with XSCALE[36, 37]. A set of 5% of randomly chosen reflections were set aside for the calculation of the free Rfactor (Rfree). The apo structure was solved using by *PHENIX*[38, 39] and *PHASER*[40, 41] molecular replacement program. Initial search model for molecular replacement was generated by using *SWISS-MODEL*[42] server against a structure of a crotonyl-CoA carboxylase/reductase (PDB 4GI2). The *K. setae* apo structure served as the model for solving the ternary-complex structure structure. This resulted in four monomers in the asymmetric unit. The refinement was carried out using *PHENIX* refinement, utilizing automatically generated TLS groups based on the structure and ordered solvent to place the water molecules[43, 44]. Following the first round of refinement, the structure was manually adjusted to the electron density and waters were added using *COOT*[45, 46].

Molecular Dynamics Simulations Workflow Description

Molecular mechanics parameters for NADPH and for the Coenzyme A fraction of Crotonyl-CoA were taken from Pavelites[47] and Aleksandrov[48], respectively. The carboxylic region of the product ethylmalonyl-CoA was parameterized (bonded and non-bonded terms) using as model the glutamate residue present in CHARMM22/CMAP[33, 34]. The same force field was used at all the stages of this work.

The first step of the molecular dynamics simulations was the modification of the crystal structure to create ethylmanlonyl-CoA (EMC), the carboxylated product, inside the active cavity. For the mutants, this step was followed by the modification of the target residue. After the solvation, each system was subject to energy minimization and three equilibration stages: 500ps NVT using the Langevin thermostat (300 K, 1.0 ps^{-1}, 2 kcal mol^{-1}Å^{-2} restraint on non-solvent atoms), 5 ns NPT using Langevin thermostat and Monte Carlo barostat (300 K, 1 bar, 2 kcal mol^{-1} Å^{-2} restraint on non-solvent atoms) and a final equilibration of 100 ns NVT (2 kcal mol^{-1} Å^{-2} restraint on NADP$^+$ and EMC atoms), in order to relax the protein (SI Appendix, Fig. S7).

QM/MM Molecular Dynamics

The final equilibrated structure at 100 ns of molecular dynamics, as well as frames corresponding to 80 and 90 ns, were subject to the following QM/MM scheme of simulations. The QM region comprised 56 atoms, corresponding to the reactive portions of both NADPH and crotonyl-CoA (SI Appendix, Fig. S5), and was described using the semiempirical hamiltonian DFTB3, with the 3ob-3-1 set of parameters[49-52]. The semiempiricial DFTB3 hamiltonian was validated for representative structures of reactant, transition states and intermediate in the WT MFEP with ωB97X/def2-TZVP QM/MM potential energy calculations presenting an energy difference between 0.1 and 3.7 kcal/mol to the DFTB3 results.

After an energy minimization of the carboxylated system, it was equilibrated at 300 K for 100 ps. Steered molecular dynamics (SMD) was used to form the decarboxylated reactant state, pulling the carboxyl group from its bonded position at 1.5 Å from the C_α atom to a distance of 3.6 Å in 100 ps (SI Appendix, Fig. S6). During the SMD simulation, spontaneous formation of the C2-adduct was observed. Then the reactant system, formed by the C2-adduct and the dissociated CO_2 molecule, was equilibrated during 100 ps, applying a small restraint of 200

kcal mol^{-1} $Å^{-2}$ on the distance between the carbon atom of CO_2 and C_α, to maintain carbon dioxide in the catalytic cavity.

Minimum free energy path determination

The equilibrated reactant and product structures of the system were used as a guess for the initial and final points in the adaptive string method, implemented in Amber16[53]. This method allows the exploration of the free energy surface projected on key collective coordinates (CVs) in order to find the minimum free energy path. Fifteen relevant structural parameters (detailed in the SI) were selected as collective variables to describe the progress of the reaction.

The string was formed by forty nodes connecting reactant and product states, and each replica was given a period of 2 ps to reach the configuration corresponding to a linear interpolation between initial and final values of each CV. Then, every node drifted, following the direction of the mean force in the space spanned by the CVs. No nodes were fixed allowing the reactant and product state to reach their stationary points. For every enzymatic system simulated the convergence of the free energy path was monitored by the mean distance between successive paths and the sampling was improved attempting replica exchange every 20 steps. After reaching convergence, this final path was used as a reaction coordinate to perform umbrella integration[54], obtaining a free energy profile. The string method provides a measure of the 95% confidence interval for the free energy; the umbrella integration was stopped after obtaining a constant value (around 1.0 kcal mol^{-1}) for this parameter. The distribution of water molecules inside the active site was examined with the SSTMap software[55] to determine the hydration sites. For this purpose, representative frames of the reactant, enolate and product states were subject to 1 ns of constrained QM/MM molecular dynamics.

Supplementary Video Files Descriptions

Supplementary Videos 1-5: The videos show molecular dynamics of key points along reaction coordinate for KsCcr WT and its variants.

SV1: KsCcr WT

SV2: KsCcr N81L

SV3: KsCcr F170A

SV4: KsCcr E171A

SV5: KsCcr H365N

Supplementary Pymol and Chimera Session Files Descriptions

Supplementary data 1-6: Pymol and Chimera files corresponding to figure 2 panels B, E, H and figure 3 panels B and E

SD1a: Pymol file of figure 2 panel B

SD1b: Chimera file of figure 2 panel B

SD2a: Pymol file of figure 2 panel E

SD2b: Chimera file of figure 2 panel E

SD3a: Pymol file of figure 2 panel H

SD3b: Chimera file of figure 2 panel H

SD4a: Pymol file of figure 3 panel B

SD4b: Chimera file of figure 3 panel B

SD5a: Pymol file of figure 3 panel E

SD5b: Chimera file of figure 3 panel E

Files can be downloaded under the following link:

https://www.mpi-marburg.mpg.de/547072/Data

CHAPTER III

Alternative electrophiles for ECR

3.1. Abstract

Carboxylases are enzymes responsible for the fixation of carbon into biomass and are essential to the global carbon cycle. They are able to utilize atmospheric CO_2 and incorporating it into organic molecules. Generally, they form a C-C bond by creating an activated substrate nucleophile, called an enolate, which can then be carboxylated with CO_2. Enoyl-CoA cabroxylases/reductases (ECRs) catalyze the reductive carboxylation of enoyl-CoAs and represent the fastest carboxylases known today. We tested whether an aldehyde could replace CO_2 as the resolving electrophile in the final step of the reaction. We show that ECR from *Kitasatospora setae* is able to utilize formaldehyde as an alternative electrophile thereby forming 2-(hydroxymethyl) butyryl-CoA (2-HMB-CoA). This compound was characterized by mass spectrometry and NMR spectroscopy and confirmed the predicted structure. This opens up the possibility for more electrophiles to be tested and employ ECRs as biocatalyst for the production of α-substituted carboxylic acids.

3.2. Introduction

Enoyl-CoA carboxylases/reductases (ECRs) perform the unique reductive carboxylation of enoyl-CoA thioesters by oxidizing one equivalent of NADPH. ECRs catalyze the carboxylation of crotonyl-CoA to ethylmalonyl-CoA in the ethylmalonyl-CoA pathway, an important pathway for the assimilation of C_2 units in many ecologically relevant bacteria[1]. ECRs are also found associated with the biosynthesis of natural products where they produce malonyl-CoA derivatives as extender units[2]. A previous study on substrate promiscuity within this enzyme family also determined the active site residues responsible for the substrate specificity in ECRs[3]. Carboxylases in general activate their substrates and form a reactive enolate intermediate which readily reacts with CO_2[4]. ECR is known to utilize CO_2 as the resolving electrophile for the enolate, but is also able to catalyze the normal reduction reduction by using H^+ [5]. We therefore thought to interrogate the potential of alternative electrophiles that this enzyme can accept.
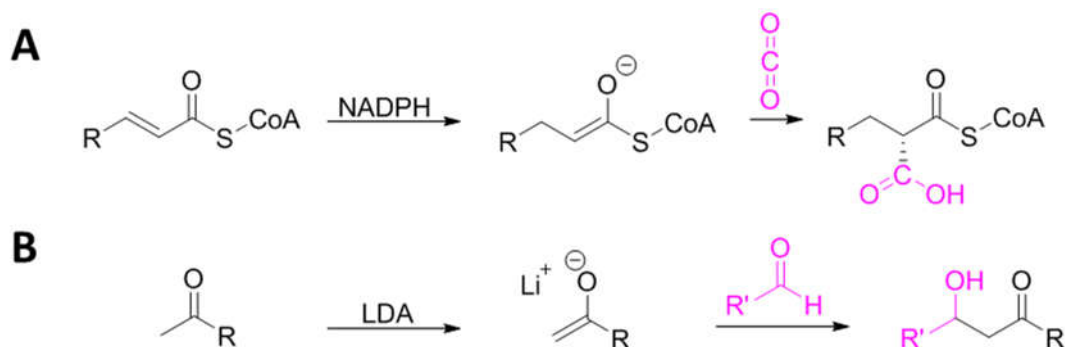
Scheme 1: **A** Reductive carboxylation catalyzed by ECRs. **B** Aldol addition of a ketone to an aldehyde yielding a 3-hydroxy ketone compound. LDA: Lithium diisopropylamide.

The reactivity of ECRs, which reminds of an aldol reaction (Scheme 1B), prompted us to consider alternative electrophiles for the ECR catalyzed reaction. Formaldehyde (FALD) is a small one carbon compound with a high electrophilic character and contains a carbonyl function for the coordination through hydrogen bonding in the active site. Moreover, its size is well suited to be accommodate din the active site given that it is smaller than CO_2. Following reaction in scheme 1B the reaction of ECR with formaldehyde would yield a β-hydroxy-thioester as the product of an aldol addition reaction. Such intermediates represent valuable building blocks for organic synthesis[6] after cleavage with a thioesterase.

Here we show that ECR can accept formaldehyde as an alternative electrophile to CO_2 and form 2-(hydroxymethyl) butyryl-CoA (2-HMB-CoA). The proposed structure was confirmed by mass spectrometry and 2D-NMR spectroscopy. The conversion to 2-HMB-CoA was stoichiometric with no formation of butyryl-CoA at a concentration of 20 mM FALD. This opens up the possibility to higher substituted aldehydes instead of CO_2 in addition

3.3. Results

Structural characterization of 2-HMB-CoA

ECR from *Kitasatospora setae* was able to utilize formaldehyde as a substrate in the presence of crotonyl-CoA and NADPH. In the absence of CO_2 ECR is known to catalyze the normal reduction reaction to butyryl-CoA. We therefore determined how much of the aldol reaction product (2-HMB-CoA)would be formed compared to the reduction product (butyryl-CoA). ECR was able to produce 94% of aldol reaction product at 20 mM FALD (Supplementary Table 1) at 1/10 of the speed of the carboxylation reaction. This data also yielded an apparent K_M value of 25 mM which is 3 orders of magnitude higher than the apparent K_M value for CO_2.

57

The MS2 spectra of 2-HMB-CoA confirmed the expected mass of the acyl-pantethenoyl fragment (Fig. 2) and other characteristic peaks, which derive from the fragmentation of CoA esters[7] (Supplementary figure S2)
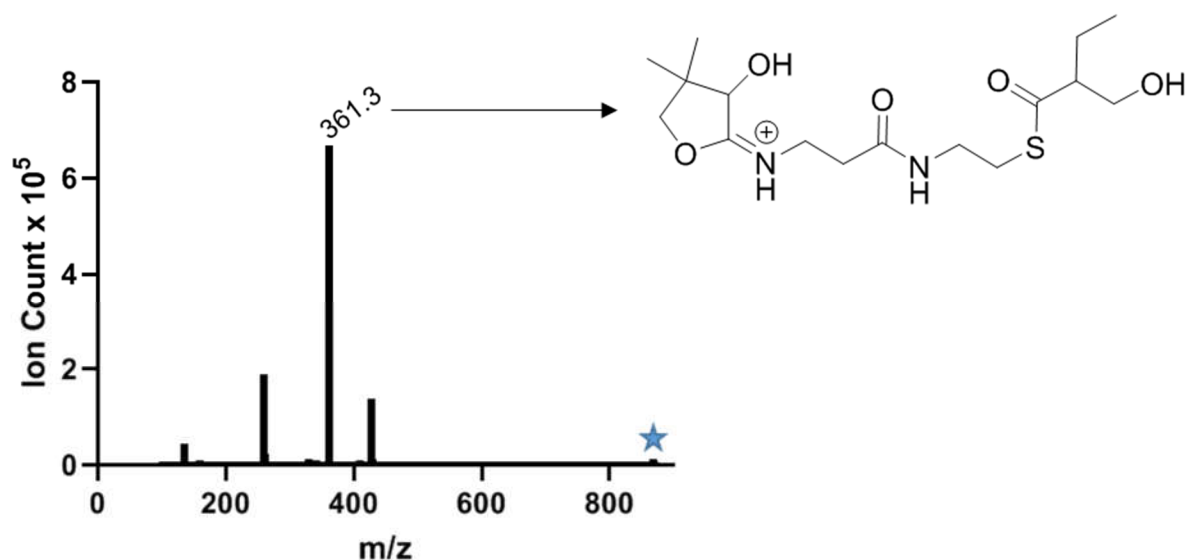

Figure 2: MS2 spectrum of 2-HMB-CoA. The peak at *m/z* 868.2 (blue star) generates the peak at *m/z* 361.3 represents the characteristic acyl-S-(cyclo)pantetheine MS2 product. The structure of this product is highlighted on the right of the spectrum.

We additionally performed structural characterization by NMR spectroscopy. The newly formed C-C bond is confirmed by the observed spin coupling between protons at the C_α and the $C_{\alpha'}$ positions in the DQF COSY spectrum. The heteronuclear multiple bond correlation (HMBC) experiment further confirms the result by showing the coupling over the carbon backbone (supplementary figure S1). These findings allowed us to identify the structure of the acyl moiety of 2-HMB-CoA.

3.4. Discussion

We showed that ECR from *Kitasatospora setae* is able to utilize FALD as an alternative electrophile to CO_2. The structure of the product was confirmed by mass spectrometry and 2D-NMR spectroscopy. We confirmed that the predicted structure of the acyl moiety is a β-hydroxy thioester. We exploited the inherent reactivity of an enolate intermediate produced by a carboxylase under mild conditions and quenched this species with a potent electrophile. This approach represents a novel biocatalytic alternative to already known routes for the preparation of enantiomerically pure hydroxy-acids [6, 8]. A previous study reported that an engineered fructose-1,6-bisphosphate aldolase[9] variant was able to connect a variety of ketones with aldehydes yielding β-hydroxy ketones and aldehydes with high stereoselectivity. Analogous to that ECRs could be used as a biolcatalyst to promote stereoselective C-C bond formation. The well studied substrate promiscuity of the ECR family[3] and the synthetic accessibility of enoyl-CoAs[7, 10] provide a good starting point for the development of this biocatalytic strategy.
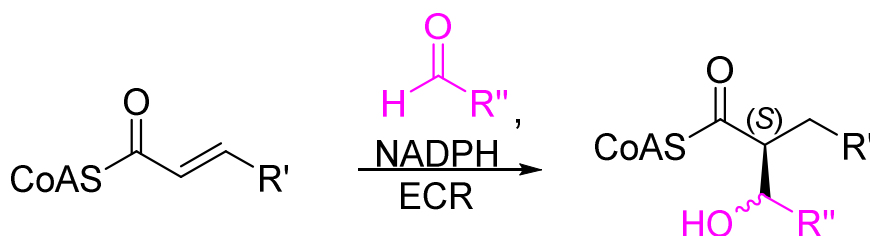
Scheme 2: Expansion of the product range of the ECR catalyzed aldol addition reaction. The rest R' can be varied by employing different ECR homologs whereas the rest R'' by employing a different aldehyde

On the other side the simple case of FALD as the alternative electrophile could be expanded to higher substituted aldehydes (Scheme 2). The reaction of enoyl-CoAs with higher substituted aldehydes would generate 2 stereocenters in one reaction.

ECRs represent a new biocatalytic strategy for the fast and stereospecific C-C bond formation between and enoyl-CoA and formaldehyde. Reactions can be performed under mild conditions and do not require protective group strategies. Further experiments are needed to address the promiscuity towards additional electrophiles and clarification of the reaction mechanism.

3.5. Materials and Methods

Large scale synthesis of 2-HMB-CoA

15 mg of crotonyl-CoA, 16 mg of NADPH, were dissolved in 100 mM KH_2PO_4 pH = 8, 100 mM formaldehyde and the reaction started by addition of 10 µM ECR form *Kitasatorspora setae* in a final volume of 3 mL and incubated at 30 °C for 20 min. The reaction was quenched by addition of 100 µL 50% formic acid and centrifuged at 17000x g to precipitate the protein. 2HMB-CoA was purified by preparative RFLC/MS over a Gemini 10 µm NX-C18 110 Å, 100 x 21.2 mm, AXIA packed column (Phenomenex) using a methanol gradient from 5 % to 30 % over 10.5 min with 25 mM ammonium formate pH = 8.1 as the aqueous phase. Fractions containing the product were pooled, lyophilized and stored at -20°C. For NMR experiments the sample was dissolved in 600 µL 50 mM KH_2PO_4 pH = 7.0 and was resuspended in 600 µL of D_2O. The sample was measured at a final concentration of 9 mM in 50 mM KD_2PO_4 pD = 7.4.

Spectrophotometric enzyme assays

Assays were performed on a Cary-60 UV/Vis spectrophotometer (Agilent) at 30°C using quartz cuvettes (1 or 10 mm path length; Hellma). Reactions were performed in 100 mM K_2HPO_4 pH = 8.0. Kinetic parameters for one substrate were determined by varying its concentration while the others were kept constant at 10 times their K_M value. Reaction procedure was monitored by following the oxidation of NADPH at 365 nm ($\varepsilon_{NADPH,365nm}$ = 3.33 $mM^{-1}\,cm^{-1}$).

Analysis of reaction products

Reactions for product analysis contained saturating amounts (at least 10 times the K_M) of Crotonyl-CoA, NADPH, 100 mM K_2HPO_4 pH = 8.0, and varying amounts of formaldehyde in a final volume of 100 µL. The reaction procedure was monitored by decrease in absorbance of NADPH at 365 nm, quenched with 10 µL of 50% formic acid at completion and spinned at 17'000 g for 10 min to precipitate the protein. The reaction was diluted 10 times into 5% methanol/Buffer 8.1 and analyzed by UHPLC over a Sonoma C18(2), 3 µm 100 Å, 100 x 2.1 mm using a 5 to 45% methanol gradient over 14.5 min. Retention times were 2HMB-(6.4 min), crotonyl-(7.2 min) and butyryl-CoA (8.1 min)

3.6. References

1. Erb TJ*, et al.*, Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A.* 104(25):10631-10636 **2007**.
2. Wilson MC & Moore BS, Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* 29(1):72-86 **2012**.
3. Peter DM*, et al.*, Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13457-13461 **2015**.
4. Schada von Borzyskowski L, Rosenthal RG, & Erb TJ, Evolutionary history and biotechnological future of carboxylases. *J. Biotechnol.* 168(3):243-251 **2013**.
5. Erb TJ, Brecht V, Fuchs G, Muller M, & Alber BE, Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Natl. Acad. Sci. U.S.A.* 106(22):8871-8876 **2009**.
6. Ren Q, Ruth K, Thony-Meyer L, & Zinn M, Enatiomerically pure hydroxycarboxylic acids: current approaches and future perspectives. *Appl. Microbiol. Biotechnol.* 87(1):41-52 **2010**.
7. Agarwal V*, et al.*, Chemoenzymatic Synthesis of Acyl Coenzyme A Substrates Enables in Situ Labeling of Small Molecules and Proteins. *Org. Lett.* 17(18):4452-4455 **2015**.
8. Roldan R*, et al.*, Biocatalytic Aldol Addition of Simple Aliphatic Nucleophiles to Hydroxyaldehydes. *ACS Catal.* 8(9):8804-8809 **2018**.
9. Samland AK, Rale M, Sprenger GA, & Fessner WD, The transaldolase family: new synthetic opportunities from an ancient enzyme scaffold. *ChemBioChem* 12(10):1454-1474 **2011**.
10. Peter DM, Vogeli B, Cortina NS, & Erb TJ, A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters. *Molecules* 21(4):517 **2016**.

3.7. Supplementary Information

Supplementary figure S1: **Two-dimensional NMR analysis of 2-HMB-CoA recorded at 600 MHz in 50 mM NaD$_2$PO$_4$ pD = 7.4 at 25°C.** Recorded spectra are **A)** Structure and numbering of 2-HMBCoA **B)** Assignment table for 2-HMBCoA **C)** ^{1}H-NMR **D)** DQF-COSY **E)** HSQC **F)** HMBC
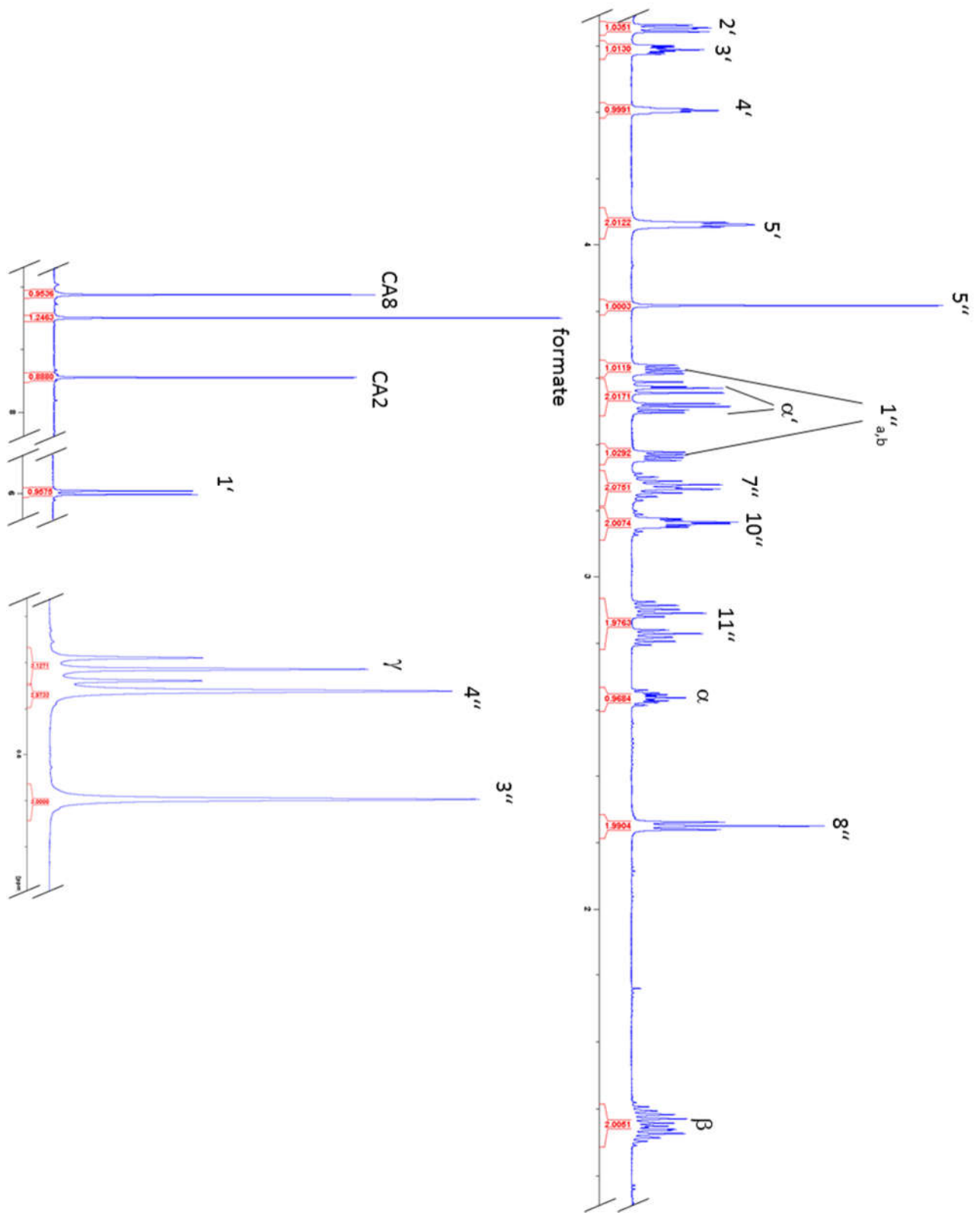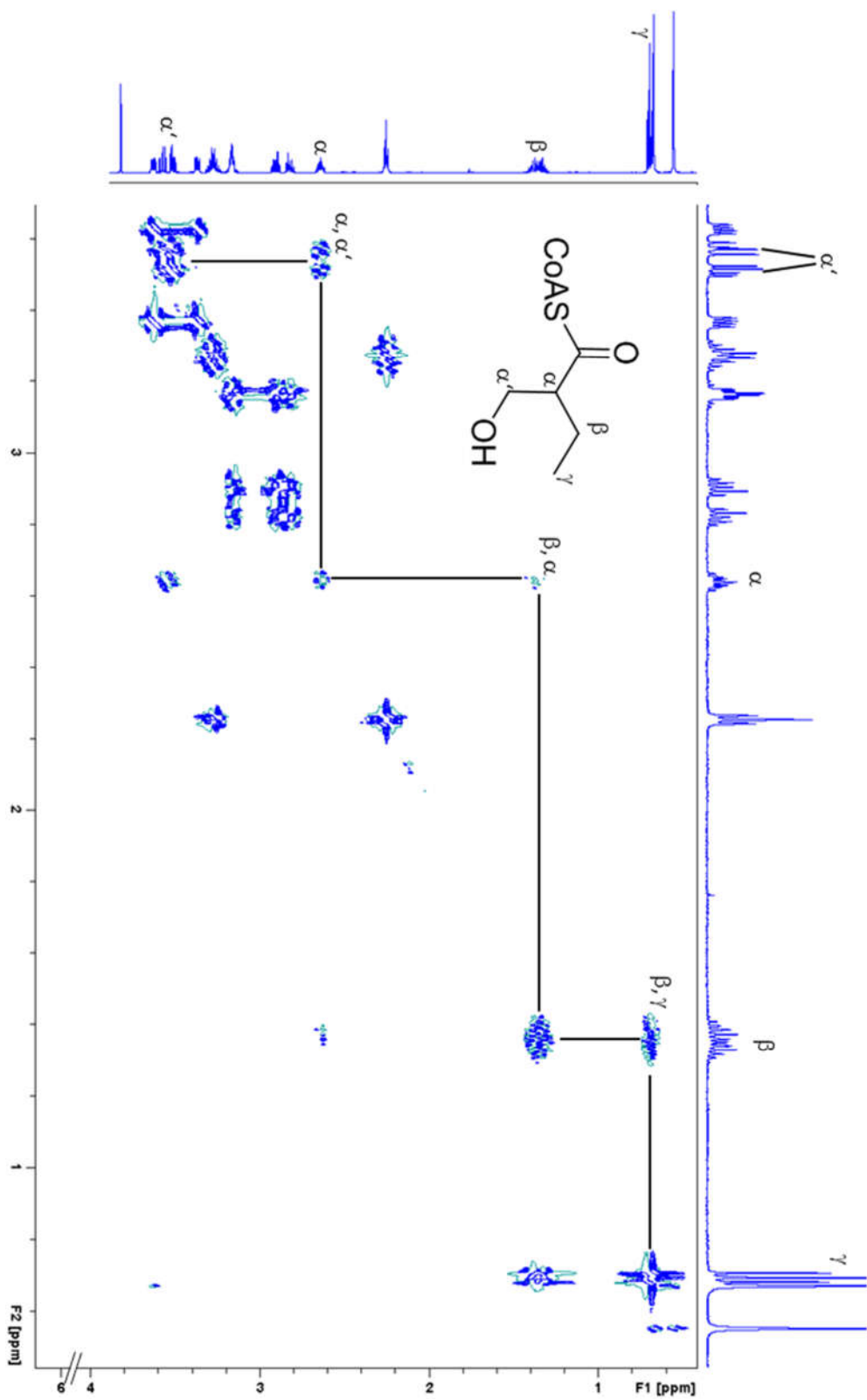
A
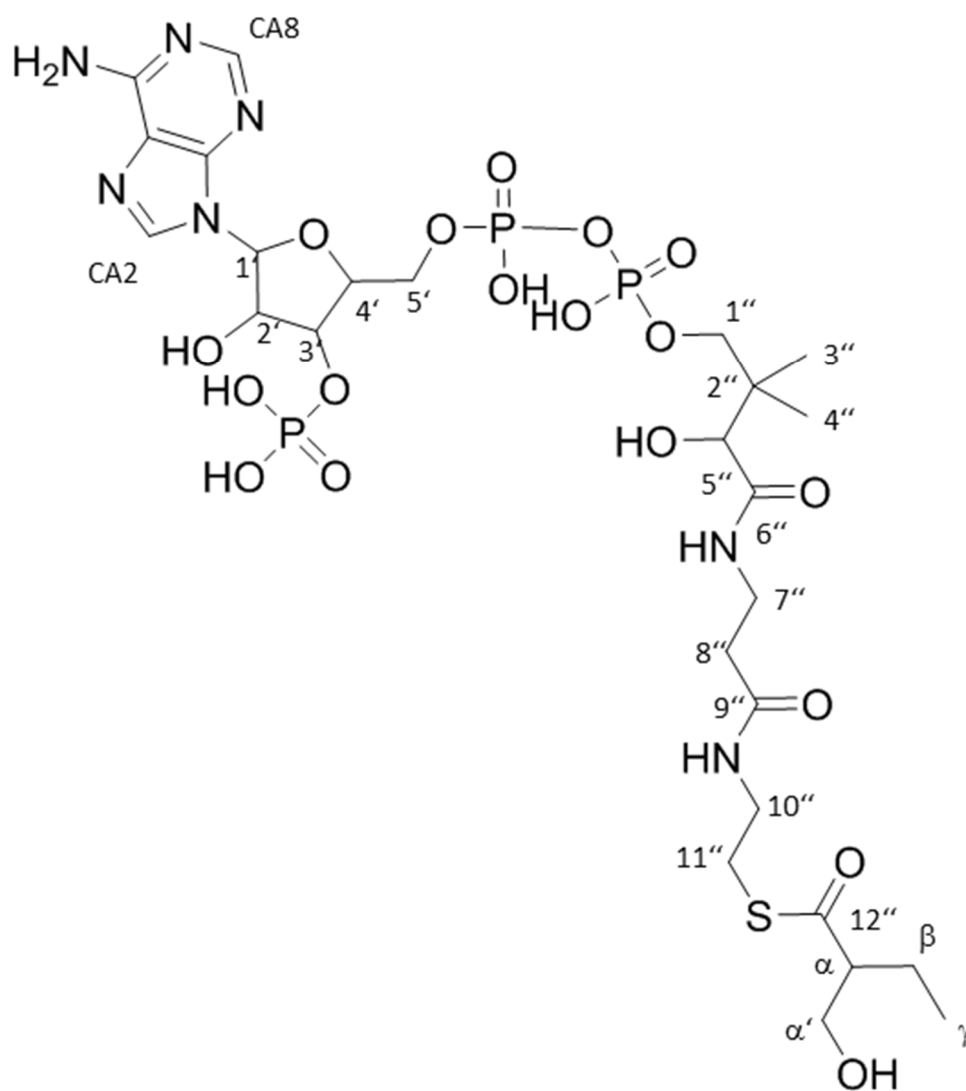
B

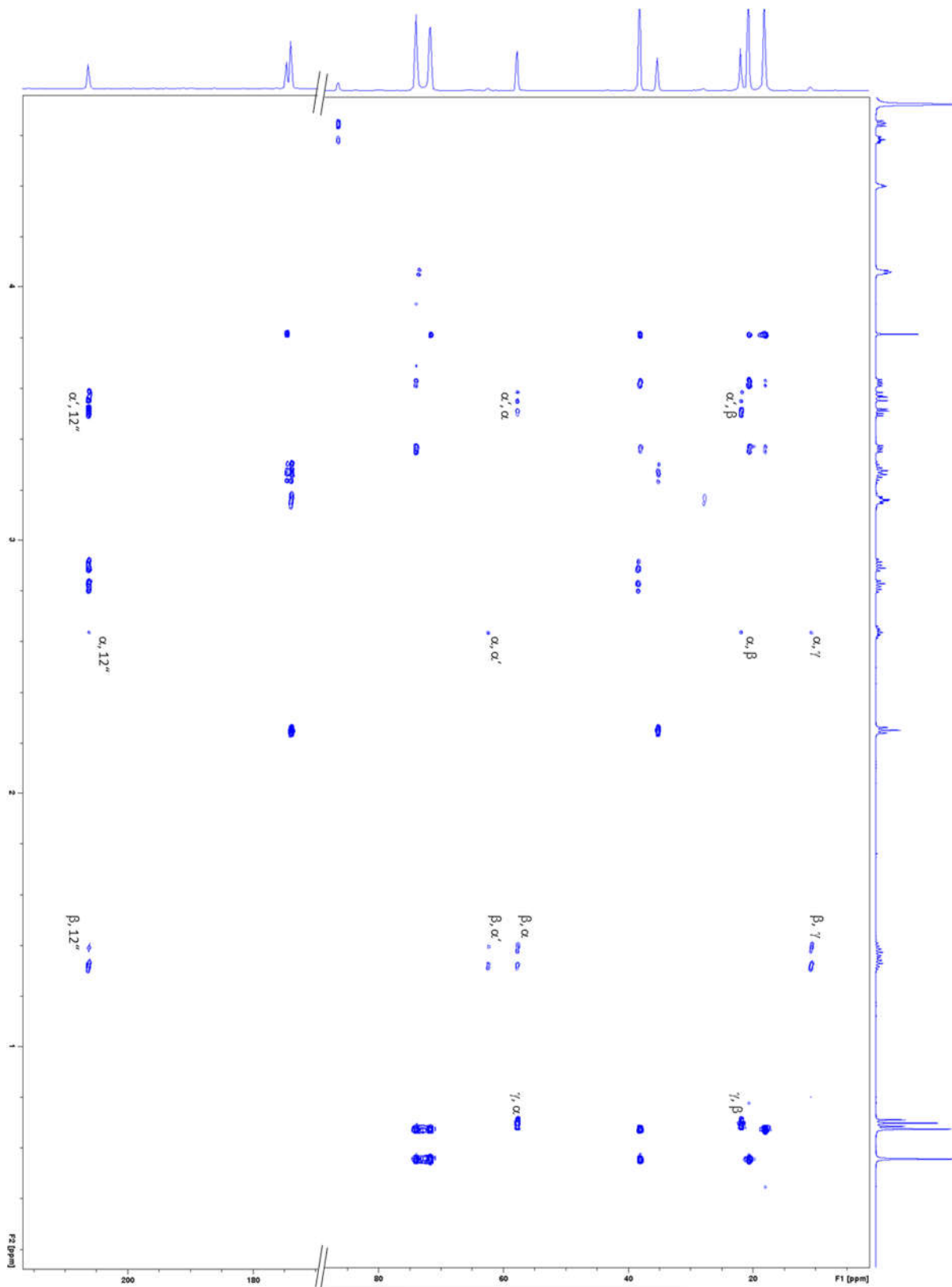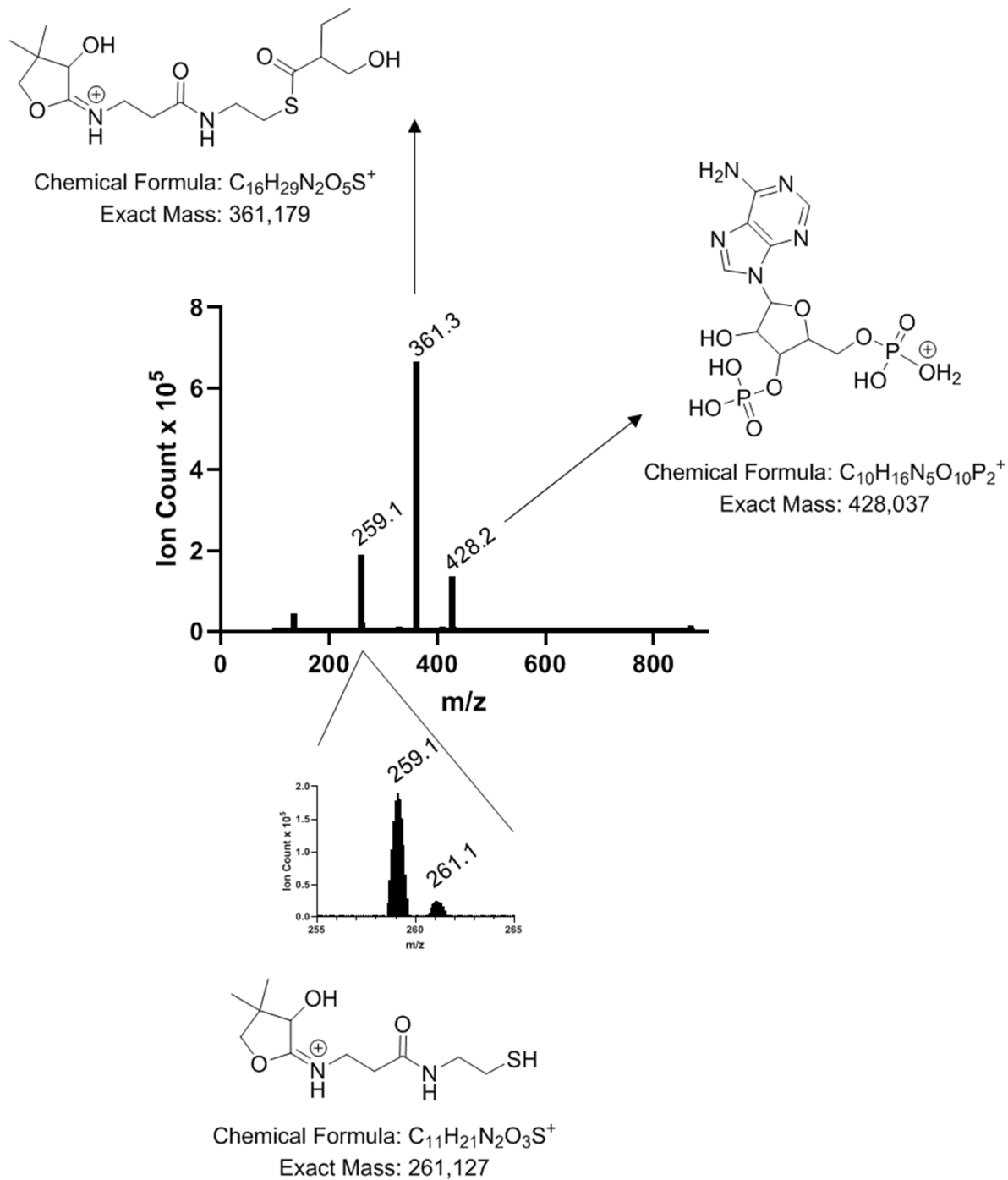| Position	¹H-shift (ppm)	¹³C-shift (ppm)	HMBC	DQF-COSY
α	2.64	57.68	$C_{\alpha'}$, C_{β}, C_{γ}	$C_{\alpha'}$, C_{β}
α'_a	3.51	62.26	C_{α}, C_{β}, $C_{12''}$	C_{α}
α'_b	3.57	62.26	C_{α}, C_{β}, $C_{12''}$	C_{α}
β	1.34	21.82	C_{α}, $C_{\alpha'}$, C_{β}, $C_{12''}$	C_{γ}, C_{α}
γ	0.69	10.75	C_{α}, C_{β}	C_{β}
$1''_a$	3.36	71.70		-
$1''_b$	3.63	71.70		-
2''	-	38		-
3''	0.55	18.15		-
4''	0.67	20.72		-
5''	3.81	74.01		-
6''	-	175		-
7''	3.22	35.21		$C_{8''}$
8''	2.25	35.21		$C_{7''}$
9''	-	173		-
10''	3.16	38.45		$C_{11''a,b}$
$11''_a$	2.81	27.62		$C_{10''}$
$11''_b$	2.91	27.62		$C_{10''}$
1'	6.00	86.59		$C_{2'}$
2'	4.65	74.05		$C_{1'}$, $C_{3'}$
3'	4.59	73.32		$C_{2'}$, $C_{4'}$
4'	4.41	83.69		$C_{3'}$, $C_{5'}$
5'	4.06	65.78		$C_{4'}$
CA2	8.11	152		
CA4	-	149		
CA5	-	119		
CA6	-	156		
CA8	8.37	140		

C

D

E
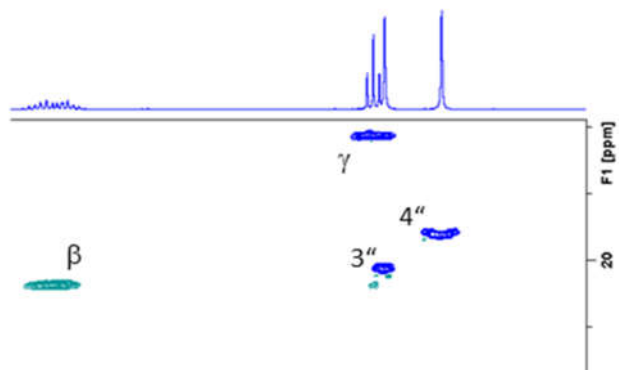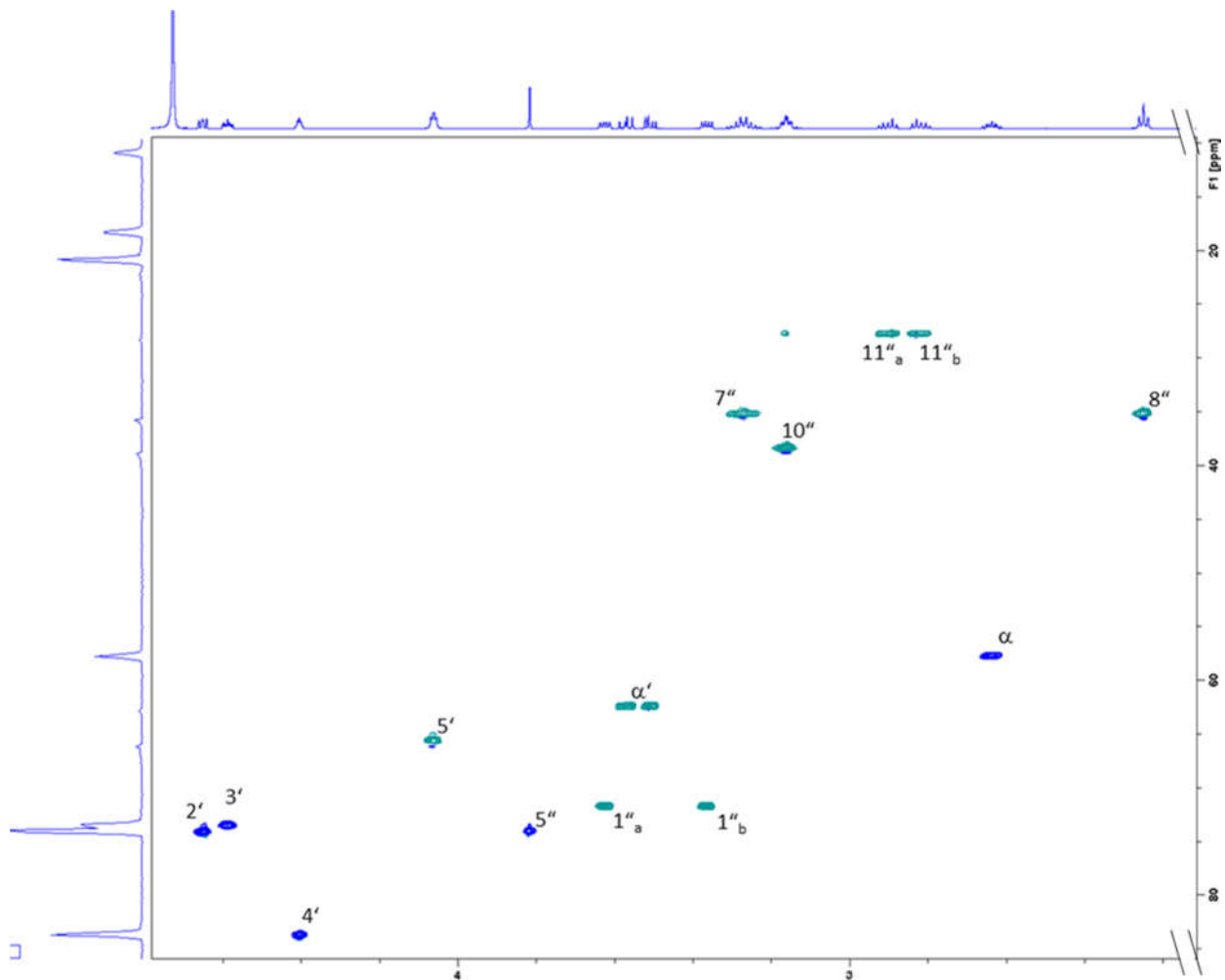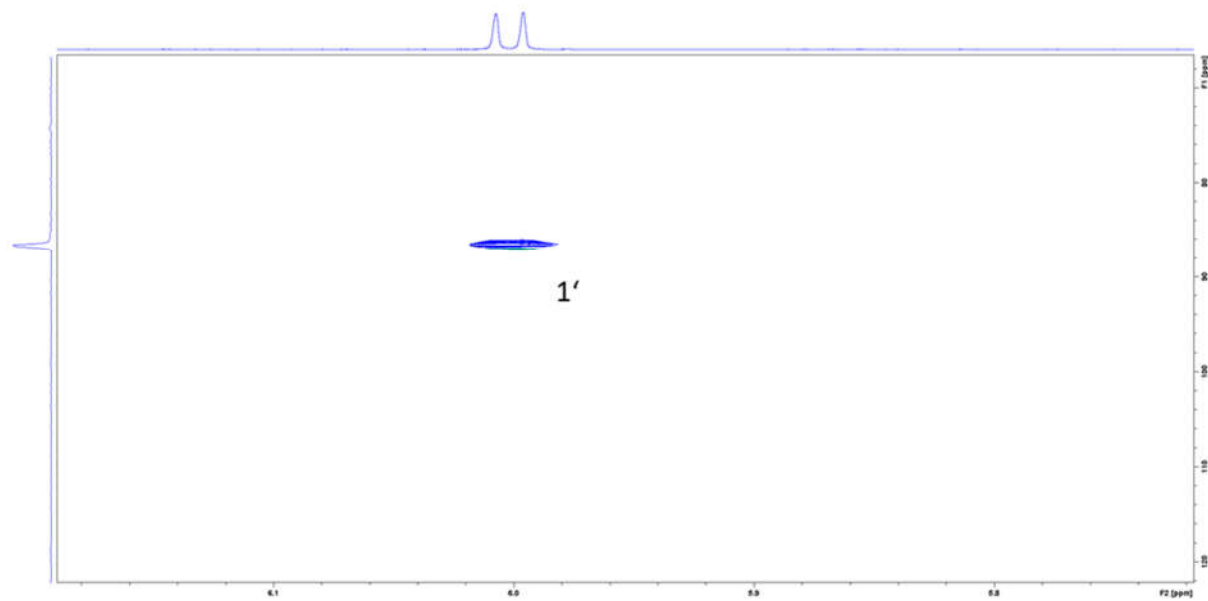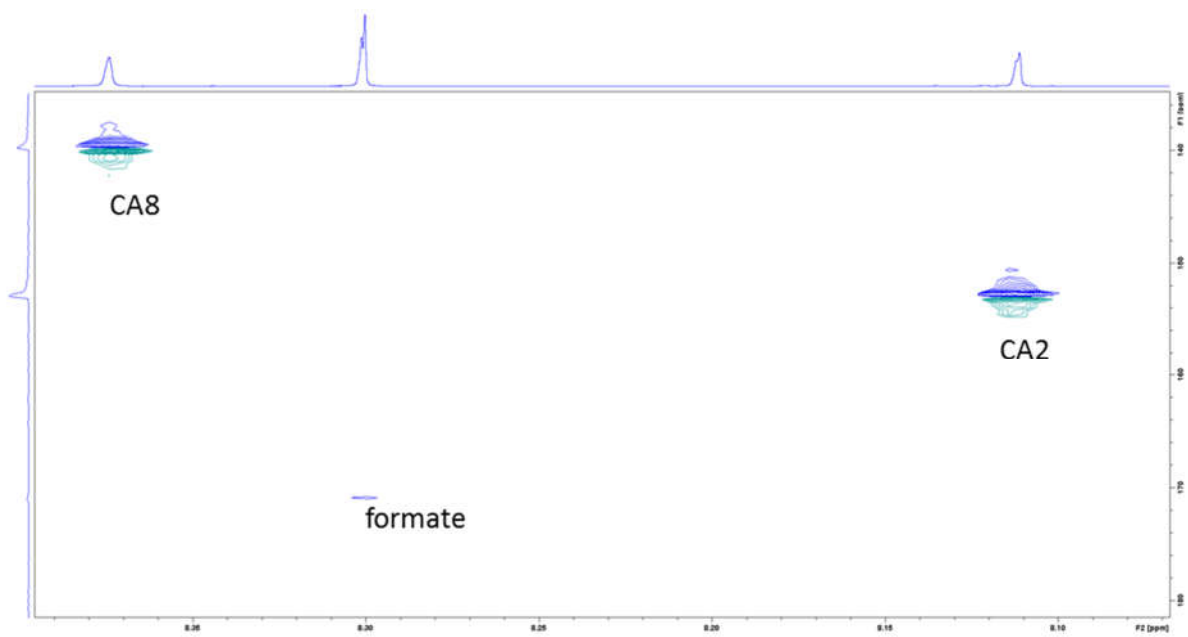
F

Supplementary Figure S2: **MS/MS spectrum of 2HMB-CoA.**

The peak observed at m/z 259.1 is likely an oxidation product of the compound found at *m/z* 261.1

Chemical Formula: $C_{16}H_{29}N_2O_5S^+$
Exact Mass: 361,179

Chemical Formula: $C_{10}H_{16}N_5O_{10}P_2^+$
Exact Mass: 428,037

Chemical Formula: $C_{11}H_{21}N_2O_3S^+$
Exact Mass: 261,127

Supplementary Table 1: Product analysis of reaction catalyzed by KsECR with formaldehyde. The initial slopes were fitted to the Michaelis Menten equation to estimate an apparent K_M value for FALD. Prolonged incubation of the reaction with [FALD] > 50 mM rvealed formylation of the adenosine moiety of The CoA ester and cofactor as previsouly described

[FALD]/(mM)	2-HMB-CoA/%	$V_0/[E]/(s^{-1})$
10	88	8
20	94	12
50	97	20
100	98	22

CHAPTER IV

Coupled inter-subunit dynamics enable the fastest CO₂-fixation by reductive carboxylases

4.1. Abstract

Enoyl-CoA carboxylases/reductases (ECRs) are the most efficient CO_2-fixing enzymes described to date, outcompeting RubisCO, the key enzyme in photosynthesis in catalytic activity by more than an order of magnitude. However, the molecular mechanisms underlying ECR's extraordinary catalytic activity remain elusive. Here we used different crystallographic approaches, including ambient temperature X-ray Free Electron Laser (XFEL) experiments, to study the dynamic structural organization of the ECR from *Kitasatospora setae*. *K. setae* ECR is a homotetramer that differentiates into a dimer of dimers of open- and closed-form subunits in the catalytically active state, suggesting that the enzyme operates with "half-site reactivity" to achieve high catalytic rates. Using structure-based mutagenesis, we show that catalysis is synchronized in *K. setae* ECR across the pair of dimers by conformational coupling of catalytic domains and within individual dimers by shared substrate binding sites. Our results provide unprecedented insights into the dynamic organization and synchronized inter- and intra-subunit communications of nature's most efficient CO_2-fixing enzyme during catalysis.

4.2. Introduction

The capture and conversion of atmospheric CO_2 remains a challenging task for chemistry, resulting in an ever-increasing interest to understand and exploit CO_2 fixation mechanisms offered by biology[1]. The recently described family of enoyl-CoA carboxylases/reductases (ECRs) represent the most efficient CO_2-fixing enzymes found in nature to date[2, 3]. ECRs catalyze the reductive carboxylation of a variety of enoyl-CoA thioester substrates at catalytic rates that are up to 20-fold higher than Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), an enzyme involved in the first carbon fixation step in the Calvin-Benson cycle of photosynthesis[1, 4].

ECRs catalyze the reduction of α,β-unsaturated enoyl-CoAs using the reduced form of the cofactor nicotinamide adenine dinucleotide phosphate (NADPH). This generates a reactive enolate species, which acts as a nucleophile to attack a CO_2 molecule[2, 3, 5]. The structural details of the carboxylation reaction have remained elusive, due in part to the lack of high-resolution structures of ECRs containing catalytic intermediates and carboxylated products. Currently, there are five available ECR structures. However, they all have different substrate

specificities, ranging from short- (PDB: 3HZZ, 3KRT) to long-chain (4A0S[6]) and aromatic enoyl-CoA substrates (4Y0K[7]), and are from different biological backgrounds including primary (i.e. central carbon) metabolism (PDB: 4GI2) and secondary metabolism. Moreover, most of them were co-crystalized with NADPH or NADP+only and do not contain CO_2, enoyl-CoA substrates or acyl-CoA products. This significantly limits our structural understanding of the enzyme's catalytic mechanism.

The aim of this study was to provide a detailed structural understanding of the carboxylation reaction of ECRs at the level of the oligomeric protein complex. To this end, we chose the ECR from *K. setae*, which shows high substrate specificity for crotonyl-CoA and superior catalytic efficiency (see Table 1). Using cryogenic X-ray crystallography at synchrotrons and room temperature serial femtosecond X-ray crystallography (SFX) at an XFEL, four high-resolution ECR structures were determined in different conformational states: the apo form and three holo forms, in binary complex with the reduced cofactor NADPH, in ternary complex with NADPH and butyryl-CoA, and in binary complex with the oxidized cofactor NADP+ (**Figure 1a**).

Here we show that the tetrameric complex assumes a dimer-of-dimers ("a pair of dimers") configuration during catalysis. The central oligomerization domains of ECR remain largely unchanged, while the peripheral catalytic domains move drastically to provide two sets of active site conformations, open- and closed-form, upon binding of the NADPH cofactor alone or in the presence of substrates. This coordinated motion is enabled by a tight coupling of catalytic domains across the pairs of dimers. Structure based mutagenesis of the interface of the catalytic domains supports this notion and provides compelling evidence that synchronization across the pair of dimers is a crucial factor in *K. setae* ECR to achieve the high catalytic efficiency. Further kinetic experiments demonstrate that subunit communication within the pair of dimers is important to synchronize open- and closed-states. Altogether, our data unveil a detailed picture of the dynamic structural organization and subunit synchronization of the ECR complexes, providing unprecedented insights into the functional organization of nature's most efficient CO_2-fixing enzyme during catalysis.

4.3. Results

Apo ECR is a symmetric homotetramer, readily accessible for NADPH binding

We first determined the apo form of the ECR crystal structure from *K. setae* at 1.8 Å resolution by using synchrotron X-ray crystallography at cryogenic temperature (**Supplementary Table 1**). The asymmetric unit contains one homotetramer composed of four subunits arranged in a dimer of dimers geometry ("pair of dimers") similar to those of the previously reported binary (PDB: 4Y0K) and ternary (PDB: 4A0S) ECR structures. Overall, the tetramer shows a non-crystallographic, close to D2 (dihedral) symmetry (**Supplementary Figure S1, top right panel**) with four conformationally identical subunits (**Supplementary Figures S1&S2,** RMSD = 0.1 Å). The tetrameric structure of *K. setae* ECR is further supported by size-exclusion chromatography which showed that the apo enzyme eluted as a single peak at 205 kDa compared to the expected monomer molecular weight of 51.2 KDa corresponding to a functional complex of four subunits (**Supplementary Figure S3**).

Each ECR subunit consists of two domains – a larger catalytic domain formed by residues 1-212 and 364-445, and a smaller oligomerization domain formed by residues 212 to 363 (**Supplementary Figure S4**). The oligomerization domain comprises a Rossmann fold[8] with repeating αβ-motifs that forms a 6-stranded β-sheet (β12 to β17). The 6-stranded β-sheets of two neighboring subunits are combined into one 12-stranded β-sheet, forming the core of one dimer, A/C or B/D. Two of these 12-stranded β-sheets then form the core of the tetrameric complex (**Supplementary Figure S4**).

The catalytic domains of *K. setae* ECR are located at the periphery of the tetrameric complex. The active site of ECR is formed by helix 8 and surrounding loops at the interface with an adjacent subunit in the tetramer (**Supplementary Figure S4**). The active site cavities in the apo form are open and accessible for both the cofactors and substrates.

NADPH binding induces ECR into a dimer of dimers with distinct open and closed form subunits

To understand how cofactor binding affects the enzyme, we determined the crystal structure of the *K. setae* ECR-NADPH binary complex at 2.4 Å resolution by using serial femtosecond X-ray crystallography (SFX) at ambient temperature (**Figure 2, Supplementary Table 1**)[9-12]. The simple F_o-F_c difference electron density map allowed us to unambiguously place NADPH molecules in all four subunits. NADPH binds with its adenine moiety in the oligomerization domain and spans the catalytic domain, where its nicotinamide moiety is located (**Supplementary Figure S5**).

Notably, binding of NADPH breaks the dihedral D2 symmetry observed in the apo-form tetramer structure, while symmetry about the y-axis is retained, resulting in a non-crystallographic, almost cyclic C2 symmetry (**Supplementary Figure S1 bottom right**). In the NADPH-ECR binary complex, the four subunits of ECR differentiate into two forms (A & B and C & D), which are structurally distinct from each other (**Supplementary Figure S1**, RMSD = 0.5 Å between A & B and C & D, 1.8 Å between A & C, A & D, B & C and B & D respectively) (**Figure 2a&b**). The A & B subunits show cofactor-binding pockets that are open, referred hereafter as "open-form" state (**Figure 2b**). On the other hand, in the C & D subunits, the cofactor binding pocket is compressed inwards, which seals the NADPH cofactor within the catalytic domain, resulting in a "closed-form" state (**Figure 2b**).

The bulk of the NADPH cofactor is bound almost identically in the two closed-form subunits, C & D (**Figure 2d,e&f**). However, the nicotinamide moiety adopts two alternate conformations in the two open-form subunits A & B (**Figure 2e&f**), indicating a more flexible cofactor binding than in the closed-form subunits (**Figure 2d**). Possible conformations of the NADPH cofactor in the open and closed binding cavities and its flexibility were studied with molecular dynamics (MD) simulations in a dimer of subunits A and C. In the closed-form subunit C the cofactor kept its position in the binding site as observed in the X-ray crystal structure. When we placed the NADPH cofactor in the same position in the open subunit A and performed similar MD simulations, NADPH left this initial confirmation in all three 200 ns trajectories and adopted various alternate conformations in the open cavity, including the two that were observed in our high-resolution crystal structure (**Supplementary Figure S6**). These variable conformations in the open A subunit are allowed by a substrate binding pocket that is more than 5 Å wider than the closed C subunit (**Figure 2c**).

In summary, binding of the NADPH cofactors to the apo enzyme induces the four subunits of the enzyme to differentiate into open- and closed-form states in both dimers (**Figure 2b**) thus breaking the dihedral D2 symmetry to cyclic C2 symmetry. This coupled subunit rearrangement of *K. setae* ECR and the large active site differences within each pair of dimers suggest that catalysis is synchronized between the individual subunits of the complex, which will become clearer in the subsequent analysis sections below.

Ternary complex supports half-site reactivity in ECR catalysis

We next attempted to determine the structure of the *K. setae* ECR ternary complex crystallized in the presence of spent cofactor $NADP^+$ and the reaction product ethylmalonyl-CoA. The structure of the ternary complex, however, indicated that the carboxylate group was lost during the crystallization process, which resulted in butyryl-CoA, which is in line with the finding that ethylmalonyl-CoA is unstable and tends to decarboxylate at the active site of ECR into butyryl-CoA and CO_2 over time[2, 13] (**Supplementary Figure S7**). Numerous attempts of preserving ethylmalonyl-CoA in the crystal structure proved to be extremely challenging and therefore we co-crystallized ECR with butyryl-CoA and NADPH and determined its structure at 1.7 Å resolution (**Figure 3**). This structure revealed that two butyryl-CoA molecules are bound at the active sites of the closed-form subunits B & D.

This ternary complex structure is overall very similar to the structure of ECR-NADPH binary complex. It also displays the non-crystallographic, pseudo C2 cyclic symmetry (**Figure 3a** and **Supplementary Figure S1 bottom right panel**) and comprises of open- and closed-form subunits that overlay very well with the open- and closed-form subunits of the ECR-NADPH binary complex (**Supplementary Figures S1&S2**, RMSD = 0.1, 2.1 Å respectively). The NADPH cofactor appears bound to all active sites, however, only the closed-form subunits B & D also contain the completely intact butyryl-CoA thioester (**Figure 3a&b**). This strongly suggests that the closed-form subunits represent the Michaelis complex in which substrate and cofactor are positioned for catalysis, while the open-form subunits represent catalytically incompetent complexes that are in place to perform the next round of catalysis.

ECR uses an elegant mechanism to align CoA-ester for catalysis in the closed-form subunit pairs. The active site of the closed-form subunits is sealed by the collective motion of loops 37-44, 88-94, 338-350, and helices 6, 7 and 21 of the catalytic domain (**Figure 3c**), which creates multiple interactions of the protein with the CoA-ester (**Figure 3d,e&f**). Notably, the CoA-ester extends from the catalytic closed-form domain into the neighboring open-form subunit within the same dimer pair, where Arg352, and Tyr353 interact with the phosphate backbone of CoA. The CoA moiety stretches further into the neighboring open-form subunit, where its adenosine tail interacts with a small binding pocket formed by three residues on the surface of the dimerization domain, Tyr328, Lys296, and Arg303 (**Figure 3g**).

When we inspected the adenine binding pocket of the closed-form subunits, we also observed electron density of adenine, indicating that the CoA-ester was bound (**Figure 3g right inset**).

The electron density beyond the adenine ring, however, becomes disordered, suggesting that the part of the CoA molecule that extends into the active site of the neighboring open-form subunit remains flexible, which is corroborated by the higher anisotropy of the CoA binding site (**Figure 1c**). Quantum mechanical/molecular mechanics (QM/MM) simulations on a dimer of subunits A and C were performed to further evaluate the flexibility of the substrate in the open- and closed-form subunits. These simulations showed that the substrate residing in the closed subunit had significantly lower B-factors than the simulated substrate in the open subunit (**Supplementary Videos 1a-e, 2a-e**). In the open subunit, the acyl moiety was found to have a high degree of flexibility within the active site (**Supplementary Figure S8, Supplementary Videos 1&2**). Taken together, both the crystallographic analyses and QM/MM simulations are agreeing with the idea that the closed-, but not the open-form subunits represent the catalytically competent subunits. Furthermore, the organization of the ternary *K. setae* ECR complex into catalytically competent and incompetent subunits, suggests that the enzyme operates with half-site reactivity, in which active sites alternate during catalysis[14–16].

Product release returns ECR back into a symmetric homotetramer

Following catalysis, *K. setae* ECR has to release the product and oxidized cofactor. In order to understand the structural basis for this part of the catalytic cycle, the protein was co-crystallized with $NADP^+$ and its structure was solved at 1.8 Å resolution. The enzyme transitioned back to the D2 symmetry of the apo enzyme, with four conformationally identical subunits (**Supplementary Figures S1&S2**) all containing bound $NADP^+$ molecule (**Figure 4a**). Compared to the homotetrameric apo enzyme in which helix α6 and loop 88-93 of the catalytic domain stabilize the phosphate backbone of the $NADP^+$ molecule(**Figure 4b**), the corresponding helix and loop moved closer to each other (**Figure 4a**), similar to the ternary complex (**Figure 3d**), which leaves the homotetramer in an "all-closed" state. Notably, while the $NADP^+$ binding mode is comparable between the ternary complex and the complex with $NADP^+$ alone, the B-factors of the latter are larger than those in the ternary complexes (**Figure 1b**), indicating that in the $NADP^+$ bound enzyme the atoms become in general more mobile, which may be advantageous for discharging the oxidized, spent cofactor. The configuration of the nicotinamide group of $NADP^+$ in the $NADP^+$ bound structure is similar to those of the NADPH in the cofactor-bound, closed form subunits (**Figure 4c**).

Swing motion of the peripheral catalytic domain during catalysis

Comparison of the high-resolution structures of apo, NADPH-bound, NADPH/butyryl-CoA-bound and NADP$^+$-bound *K. setae* tetrameric ECRs suggests that there are coordinated motions of the catalytic domains which are peripheral to the more rigid oligomerization domains (**Supplementary videos 3 & 4**). The apo-form and NADP$^+$-bound form show 4 equivalent subunits, while NADPH-bound and NADPH/butyryl-CoA-bound forms divide into two groups of open- (*i.e.*, catalytically incompetent) and closed-form (*i.e.*, catalytically competent) subunit pairs (**Figure 1a**). In order to understand how these global structural changes affect catalysis, a principal component analysis (PCA) was used to extract major structural differences among the four structures.

The PCA revealed 8 major contributions based on their singular values (**Supplementary Figure S9a**) and the movement of the catalytic domains described above was the strongest, followed by other less significant structural changes. The 8 PCA components were used to analyze the contributions of PC1 to PC8 to the structural changes between each of the 4 structures and the average tetramer structure. This analysis showed that the first three PCA components, PC1-3 (**Supplementary Videos 5a,b,&c**), can explain more than 50% of the structural changes (**Supplementary Videos 5d**). PC1 shows that the peripheral catalytic domains are coupled and swing up and down on either side of the central oligomerization domains; PC2 shows that each of the catalytic domains moves away from its partner catalytic domain, and finally the catalytic domains undergo subtle tilt motions in PC3. The deviations of the NADPH-bound and NADPH/butyryl-CoA-bound structures from the average structure is explained mainly by PC1, the NADP$^+$-bound form by PC2, and the apo form by PC3 (**Supplementary Figure S9b).**

Communication between pairs of dimers promotes catalysis

Given the coordinated motions of the peripheral catalytic domains during the catalysis, how is catalysis synchronized across the enzyme complex? One intriguing aspect of the ECR tetramer structures is that the catalytic domains share a common interface of 1636 Å^2 between the pairs of dimers (between the catalytic domains of A and D, and those of B and C, **Figure 5a&b**) suggesting that they move together as rigid bodies. A comparison of the overall domain movements between the apo-form and NADPH/product-bound ternary complex shows that the enzyme tetramer changes from the homotetrameric apo state to the open and closed-form subunit dimer pairs (**Figure 5a**). Upon binding of CoA and NADPH, neighboring

catalytic domains rotate, which couples the widening of one active site with the compression of the other active site across the pair of dimers (**Figure 5a**). Thus, it seems to be a direct consequence of the rigid structure of the inter-catalytic domain interface that the enzyme will adopt two distinct conformational states in each dimer when it becomes catalytically active.

What are the molecular determinants that synchronize catalysis across the pair of open/closed form dimers? The inter-catalytic domain interface is mostly hydrophobic, but also features some electrostatic interactions (**Figure 5b**). Most notable are Asn218 of one subunit of one dimer that forms a hydrogen bond to Asn157 of the adjacent subunit of the other dimer, as well as Glu151 of one subunit of one dimer that forms hydrogen bonds to the main chain nitrogen of Asn133 (and/or Ala134) of the neighboring subunit of the other dimer (**Figure 5b**). Multiple sequence alignment showed that Glu151, Ans218 and Asn157 are highly conserved in ECRs from primary (i.e., central carbon) metabolism, which show faster CO_2-fixation kinetics (average k_{cat} 28 s^{-1}), but not in ECRs from secondary metabolism (**Figure 5c**), raising an interesting question about their roles in catalysis (average k_{cat} 1.2 s^{-1}).

Mutation of these residues, that are more than 20 Å away from the active site, dramatically affected the kinetic parameters of *K. setae* ECR (**Table 1**). In the E151D variant, the k_{cat} value was fivefold decreased, demonstrating that weakening the interaction of catalytic domains has profound effects on the catalytic rate of the enzyme. Mutations that targeted the asparagine interaction network showed also strong effects on the catalytic rate, but did additionally affect K_M of substrate binding. Most notable were variants N218E single and E151D/N157E/N218E triple variants that decreased the k_{cat} by more than 25- and 100-fold, respectively, highlighting that communication at the interface of the catalytic domains of the pair of dimers is an important determinant of the catalytic rate in *K. setae* ECR.

To exclude that the overall structure of the complex was not altered through these mutations, we used gel filtration, as well as native gel analysis to analyze the oligomerization state of the different enzyme variants (**Supplementary Figure S3a**). Gel filtration assays were performed under the same conditions as our kinetic measurements and showed that all mutant enzyme variants kept their tetrameric form. Only native gel analysis, which was performed under more disruptive conditions, showed slightly increase in the dimer and monomer fractions, indicating that interface interactions are weakened in these variants (**Supplementary S3b**). Overall, our mutational and kinetic data supports the hypothesis that synchronization of catalytic domains

strongly contributes to catalytic rate and is conferred through hydrogen-bond network at the interface of the pair of dimers.

Shared substrate binding within dimers is important for catalysis

While our study on the interface between catalytic domains explained how communication is conferred between different dimers through the strong coupling of the two catalytic domains, each from two dimers (inter-dimer interaction, **Figure 5a**), it did not explain how catalysis is synchronized between the open- and closed-form subunits within the same dimer. We turned our attention back to the fact that CoA substrate binding is shared between the open- and closed-form subunits in each dimer through the adenine binding pocket (intra-dimer interaction, **Figure 5a**).

To understand the role of substrate adenine binding in catalysis, we characterized the kinetics of different single, double and triple mutant variants of the adenine binding site (**Figure 3g** and **Table 2**). Mutations in the adenine binding pocket, and in particular of Arg303, strongly increased the K_M of the CoA substrate as expected, but also decreased the apparent k_{cat} of the enzyme by a factor of two to three. Notably, a comparable decrease in k_{cat} was also observed in the wild-type enzyme when we used crotonyl- panthetheine, a truncated substrate that lacks the adenosine moiety and cannot bind to the adenine binding pocket. This indicated that shared cofactor binding between neighboring subunit is important for efficient catalysis, but did not provide a conclusive answer, how catalysis is synchronized across the subunits.

We noticed that the substrate adenine binding pocket is directly followed by a loop that carries a lysine residue (Lys332), which interacts with the active site of the neighboring subunit. Lys332 residue from the open-form subunit engages in a hydrogen bonding network with the nicotine amide group of the NADPH cofactor bound to the closed-form subunit through Gln165 and His365 of the neighboring subunit (**Figure 5d**). These interactions are not observed in the active site of the open-form subunit (**Figure 5d**), raising the question whether the hydrogen bonding network connected to the adenine binding pocket might be important for catalysis. In K332A and Q165A variants, k_{cat} was decreased two- to three-fold (**Table 2**). When we tested these variants with crotonyl-panthetheine, we saw much to our surprise that catalytic activity in the K332A variant was reduced by more than two orders of magnitude, leaving us with the suggestion that adenine binding together with the loop carrying Lys332 are important to synchronize catalysis between the two subunits within the dimer. Together

with the inter-domain coupling, this intra-dimer synchronization drive fast CO_2-fixation by *K. setae* ECR.

4.4. Discussion

Our structural studies of *K. setae* ECR revealed unprecedented details on the functional organization of nature's most efficient and fastest CO_2-fixing enzyme. During catalysis, the enzyme complex differentiates into distinct functional subunits. Binding of NADPH cofactor and substrates forces the homotetrameric apo enzyme into a dimer of dimers in which each dimer is constituted of an open- and a closed-form subunits. In the closed-form subunits the NADPH cofactor and CoA substrate are aligned with each other, suggesting that this is the catalytically competent state. The open-form subunits bind cofactor and the adenine rings of the substrates but the rest of the acyl-CoA substrate remains flexible and invisible in the active site. Thus, the open-subunit active sites seem to represent a catalytically incompetent state that is pre-organized for a next round of catalysis. Altogether, this structural reorganization of ECR strongly supports the idea that the enzyme operates with "half-site reactivity", according to which catalysis is synchronized across the enzyme tetramer and alters between the open- and closed-form subunits to increase the overall catalytic efficiency of the complex[14-17].

Interaction of the catalytic domains of neighboring subunits is crucial for efficient catalysis in *K. setae* ECR. Especially important is the interaction of catalytic domains between the pairs of dimers. As soon as this interaction is disturbed, the catalytic rate of the enzyme is severely diminished. This observation is consistent with theoretical and experimental data on half-site reactivity. Synchronization of the distant catalytic subunits can enhance the catalytic rate of enzymes several-fold[18, 19]. Mutation of a single amino acid coupling the two catalytic sites of heptose isomerase GmhA reduced catalytic rate of GmhA to 6% of wild-type activity[20]. *Escherichia coli* thymidylate synthase is another example for an enzyme showing half-site reactivity[21]. Disturbing the interaction network in *E. coli* thymidylate synthase leads to a 400-fold decrease in k_{cat}[22, 23], demonstrating that domain interactions are important factors in promoting enzyme catalysis[24].

Besides the inter-dimer domain interaction, our study on ECR also suggests joint substrate-binding between neighboring subunits as another potentially important mechanism of fast synchronized catalysis. The binding of the adenine end of the CoA ester into a pocket in the neighboring subunit seems to be connected back via a hydrogen-bonding network to the

active site of the subunit where the CoA ester originated. This provides the missing link of how catalysis might be synchronized between the open- and closed-form subunits within one dimer. Taken together an attractive model of continuous turnover scheme emerges explaining the overall fast catalytic cycle of ECR; two consecutive reaction cycles alternate aided by the coupled inter-dimer catalytic domain motions (**Figure 5e**). In the first cycle (right half of **Figure 5e**), the open-form subunits A and B receive two sets of substrate and cofactor molecules, while the closed subunits C and D finish the previous reaction cycle, and release the products and $NADP^+$. As a result, the subunits A and B become closed and the subunits C and D switch to the open-subunit state. In the second cycle (left half of **Figure 5e**), subunits A and B perform the reaction and release the products and $NADP^+$ becoming open subunits, and the C and D subunits switch to closed state by acquiring a new set of substrate and NADPH.

While structural and biochemical data indicate that *K. setae* ECR achieves high catalytic rates by synchronizing active sites, this might not necessarily be true for other ECRs. A differentiation into dimers of dimers was not observed in NADPH-bound or ternary structures of other ECRs so far (e.g., PDB: 4Y0K and 4A0S respectively, which share substantial amino acid identity) (**Supplementary Figure S10**). Another reason might be that not all ECRs might perform synchronized catalysis. Note that ECRs fall into two different classes. Primary ECRs that operate in central carbon metabolism and secondary ECRs that serve in secondary metabolism, where they provide extender units for the synthesis of polyketides. Whereas primary ECRs are under strong evolutionary pressure and show on average k_{cat} values of 28 s^{-1}[25], secondary ECRs are not selected for high catalytic rates, which is also reflected by the fact that they show an average k_{cat} value of 1.2 s^{-1}[25]. Thus, it might be tempting to speculate that secondary ECRs are not selected for high turnover rates during catalysis and thus might not display synchronized "half-site reactivity".

In summary, this work provides the first overall picture of the organization of the ECR homotetrameric complex. The observation of the differentiation of the apo tetramer into open and closed form subunits upon binding of NADPH seemed to have been made possible by room temperature data collection using the XFEL beam at SACLA, highlighting the power of ambient temperature crystallography to study larger scale motions in macromolecular crystals[26, 27]. Further experiments using time-resolved X-ray crystallography at room temperature and mixing jets will be helpful to obtain a fully dynamic picture of the ECR

complex during catalysis, which will be important to fill the gaps in the mechanistic understanding of nature's most efficient CO_2-fixing principle.

Data Availability

Coordinates of the four ECR structures have been deposited in the Protein Data Bank under accession codes, 6NA3 (apo), 6NA4 (Butryl-CoA/NADPH bound), 6NA5 ($NADP^+$ bound), and 6NA6 (NADPH-bound).

Table 1. Steady state analysis of *K. setae* ECR and variants targeting the catalytic domain interface between the pair of dimers. (Michaelis-Menten curves of *K. setae* ECR and its variants are provided in **Supplementary Figure S11**)

	Crotonyl-CoA			NADPH		CO_2	
Enzyme	K_M (µM)	K_i (µM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})
Wild-type (WT)	21 ± 2	3650 ± 810	103 ± 3	37 ± 4	86 ± 2	90 ± 10	78 ± 2
E151D	28 ± 2	1958 ± 251	20 ± 1	72 ± 11	17 ± 1	80 ± 10	21 ± 1
N157E	515 ± 75	-	22 ± 1	105 ± 27	17.1 ± 0.4	40 ± 6	14 ± 0.2
N218E	272 ± 37	-	3.7 ± 0.2	66 ± 9	1.49 ± 0.06	630 ± 70	5.6 ± 0.2
E151DN157EN218E	245 ± 25	-	1.11 ± 0.04	26 ± 3	0.70 ± 0.02	440 ± 40	0.95 ± 0.03

Table 2. Apparent Michaelis-Menten parameters of *K. setae* ECR and variants targeting the adenine binding pocket and hydrogen binding network within the dimer as mean values ± standard error. (Michaelis-Menten curves of *K. setae* ECR and its variants are provided in **Supplementary Figure S11**)

	Crotonyl-CoA			Crotonyl-Pantetheine	
Enzyme	K_M (µM)	K_i (µM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})
Wild-type (WT)	21 ± 2	3650 ± 810	103 ± 3	8658 ± 531	37 ± 1
K296A	107 ± 11	-	68 ± 2	-	-
Y328F	11 ± 2	4671 ± 1693	80 ± 3	-	-
R303K	702 ± 64	-	87 ± 3	-	-
R303A	516 ± 55	-	69 ± 3	2558 ± 769	16 ± 1
R303V	334 ± 33	-	31 ± 1	6930 ± 116	16 ± 1
K296A/Y328F	192 ± 34	-	39 ± 2	-	-
K296A/R303A/Y328F	2176 ± 280	-	29 ± 2	7772 ± 106	42 ± 3
K296A/R303K/Y328F	832 ± 138	-	53 ± 2	-	-
K332A	451 ± 130	3507 ± 1809	39 ± 7	2983 ± 509	0.33 ± 0.02
Q165A	27 ± 4	1873 ± 437	56 ± 3	5632 ± 521	8.5 ± 0.3
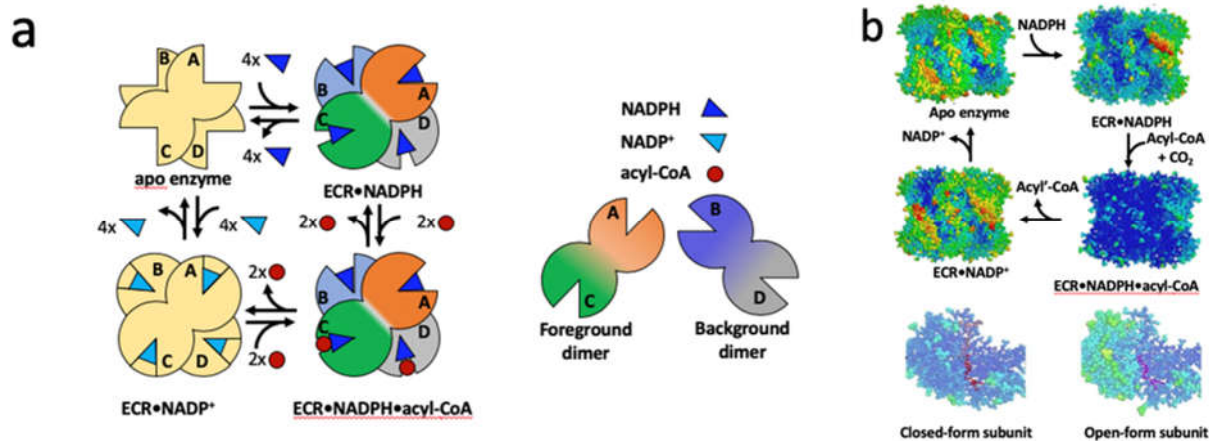
Figure 1: Structural organization of the *Kitasatospora setae* ECR complex. a. Oligomeric organization of *K. setae* ECR, as inferred from the crystal structures solved in this study. The homotetrameric apo enzyme differentiates into a dimer of dimers of open (circles with 45° wedges) and closed (full circles) subunits upon binding of NADPH cofactor. The enzyme remains a dimer of dimers in the butyryl-CoA/NADPH-ternary complex and returns back into a homotetrameric state after product release. **b**. Anisotropic B-factors of the tetramer of the different ECR complexes solved in this study. **c.** Anisotropic B-factors of the active site of the open and closed form subunits of the ternary complex. Cofactors and acyl-CoA ester are shown as stick models in red and salmon, respectively.
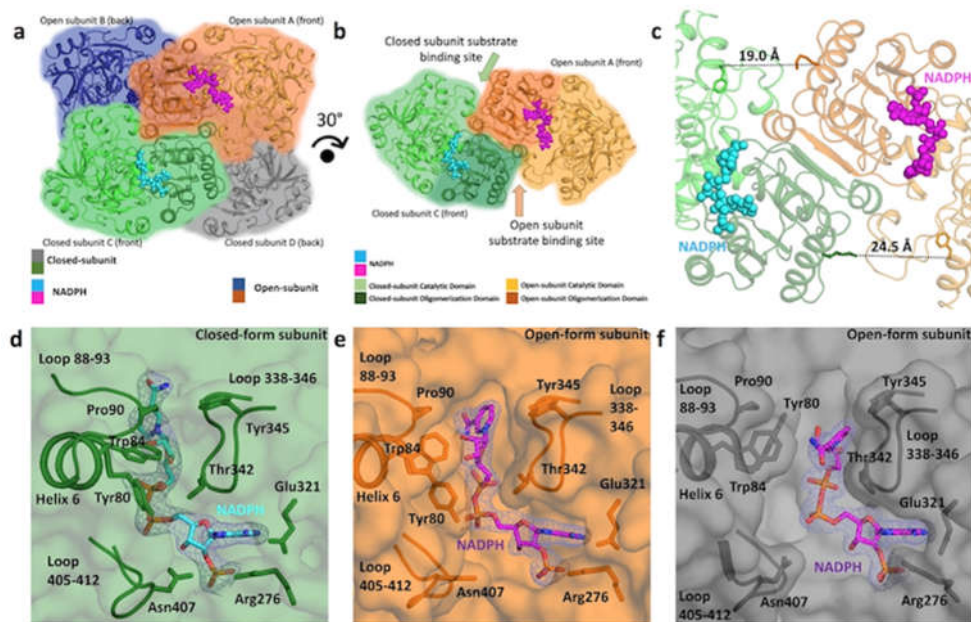
Figure 2: Binding of NADPH results in global and local conformational changes in *K. setae* ECR. a. NADPH bound tetramer complex that is organized as dimer of dimers, a pair of closed (green) and open (orange) subunits and another pair containing closed (gray) and open (blue) subunits. **b**. The foreground dimer with open- (orange) and closed-form (green) subunits rotated by 30 degrees from the view in Figure 2a. Each monomer is composed of a catalytic and an oligomerization domain. **c**. Comparison of the putative substrate binding sites between the open and closed-form subunits. **d**. Surrounding residues and loops sealing NADPH in the closed-form subunit. In panels **d** to **f**, simple $2F_O$-F_C density contoured at 1.5 sigma level is shown for NADPH within 3 Å from the molecule. **e.** Surrounding residues and loops composing the looser binding of NADPH in the open-form subunit. **f**. Alternate binding of NADPH in the open-form subunit (gray subunit in **a**.), viewed in the same direction as in **e**.).
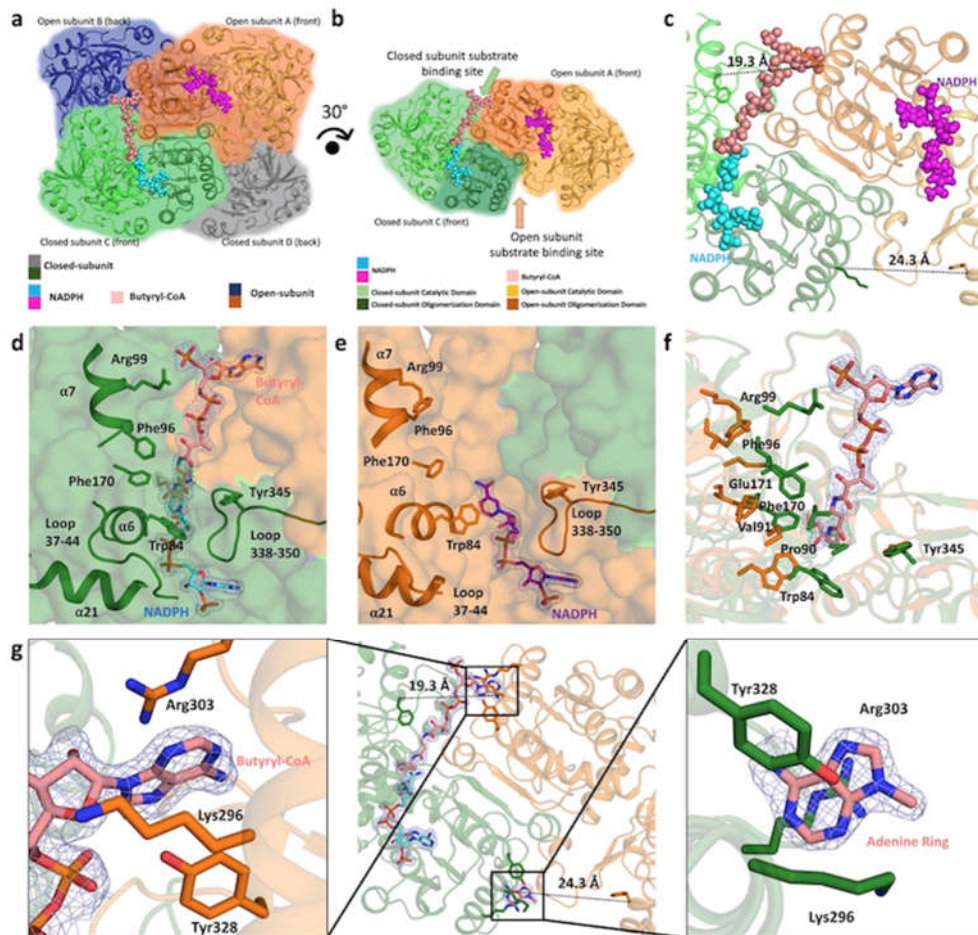
Figure 3: Structure of the ternary ECR complex. **a**. ECR tetramer in complex with NADPH and butyryl-CoA organized as dimer of dimers, the foreground dimer with one closed subunit (green) with NADPH and butyryl-CoA and open (orange) subunit containing NADPH, and another pair in the background with one closed (gray) and open (blue) subunits. **b.** The foreground dimer with closed (green) and open (orange) subunits, rotated by 30 degrees from the view in Figure 3a. Butyryl-CoA and NADPH atoms are represented as spheres. **c**. Comparison of the product binding site between the open and closed-form subunits. **d**. Cartoon and stick representation of the closed-form subunit active site. In panels **d** to **g**, simple $2F_O$-F_C density contoured at 1.5 sigma level is shown for butyryl-CoA, or potion thereof, and NADPH within 3 Å from the molecules. **e**. Cartoon and stick representation of the open-form subunit active site. **f**. Superposition of the open-form subunit onto the closed-form subunit with stick representation of the residues surrounding butyryl-CoA. **g**. Comparison of the butyryl-CoA binding sites between open and closed-form subunits with electron density of the bound butyryl-CoA and NADPH at the active site of the closed subunit (green) and the adenine ring of butyryl-CoA at the active site of the open subunit (only the adenine ring electron density is visible). Left inset: the adenine binding pocket of the open-form subunit stabilizing the adenine ring of butyryl-CoA that stretches into the adjacent closed-form subunit. Right inset: the adenine binding pocket of the closed-form subunit holding the adenine ring of butyryl-CoA. Note that only the adenine ring of butyryl-CoA is visible, while the rest of the molecule is disordered. In both cases, three residues of the adjacent subunits, Lys296, Arg303, and Tyr328 together hold the adenine ring.
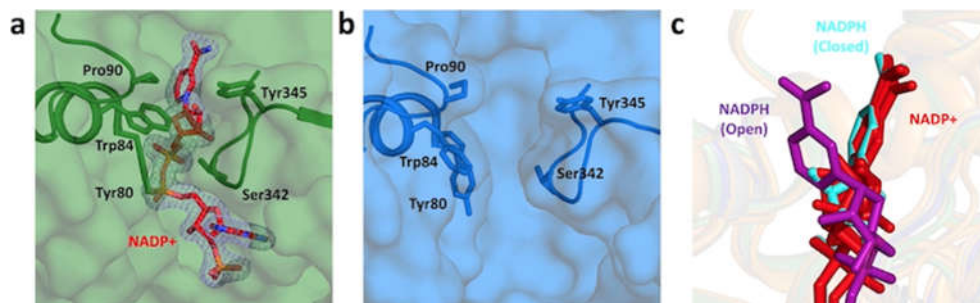
Figure 4: NADP+ binding at the active site of the NADP+ bound tetramer. a. The simple $2F_O\text{-}F_C$ electron density map of NADP+ bound in the K. setae ECR tetramer complex. Helix α6 and loop 88-93 (top left) close the binding site of NADP+. **b**. Cofactor binding site in the apo form subunit of the K. setae ECR tetramer. Helix α6 and loop 88-93 have moved away from the loop containing Ser342 and Tyr345, resulting in an open binding pocket for NADPH or NADP+. The two loops forming the cofactor binding pocket are wider apart by about 2 Å: the distances between the Cα atoms of Pro90 and Tyr345 are 13.1 Å in the open-form (blue) compared to 11.2 Å in the closed form (green). **c**. Superposition of the NADP+ molecules from all four subunits of the tetramer (red) with the open (purple) and closed-form (cyan) subunits of the NADPH/butyryl-CoA bound ternary complex.
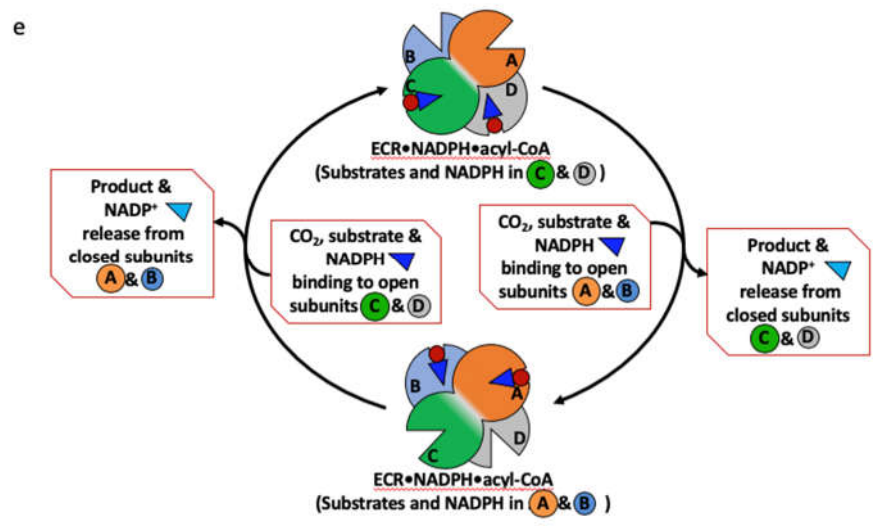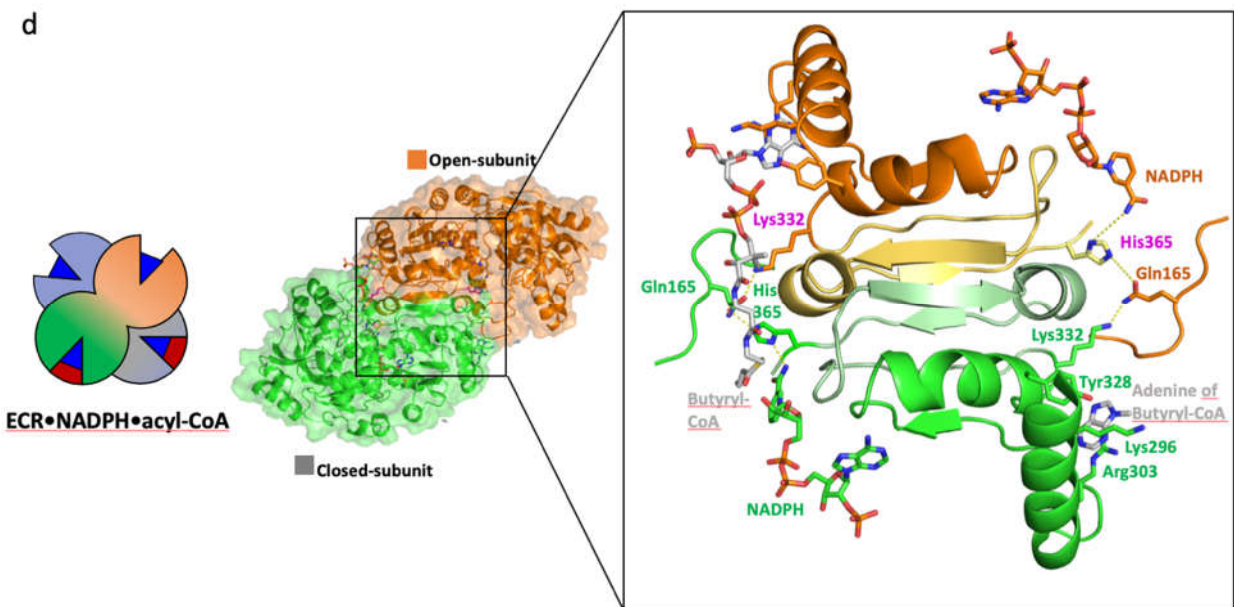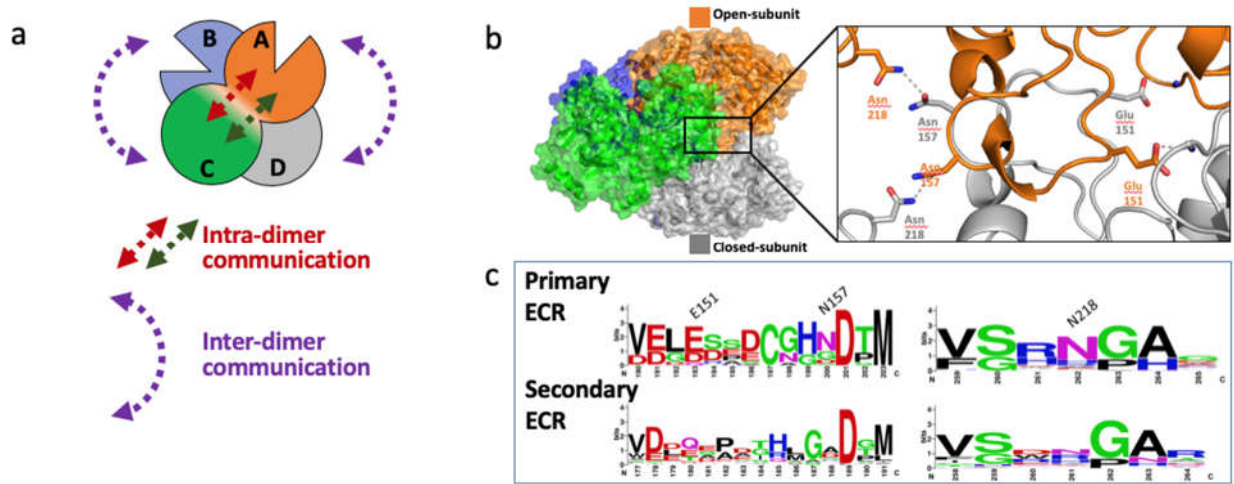
Figure 5. Inter- and intra-dimer communications drive fast CO₂ fixation by *K. setae* ECR. a. Two distinct sets of communications: inter-dimer interactions between the catalytic domains from two dimers (purple arrows) and intra-dimer communication between the open and closed subunits within each dimer (brown arrows). **b**. Inter-dimer catalytic domain interface and positions of selected amino acids that were mutated in this study in order to affect the interface between the two catalytic domains (open-form subunit in orange and closed form subunit in gray). The right panel shows the mutual H-bonding interaction between Asn218 and Asn157 from open and closed form subunits and H-bonding between Glu151 and N-atom from protein backbone. **c**. Alignment of ECR protein sequences from the primary (upper row) and the secondary (lower row) metabolism represented as sequence logos. Numbering of residues, above first row, is according to their position in K. setae ECR. **d**. Communication between the closed (green) and open (orange) subunits across the two dimers of K. setae ECR. In the closed conformation, the contacts between NADPH-His365-Glu165 and Lys332 of the adjacent open monomer allow for the correct intra-dimer communication. In the open conformation, the communication network is compromised as indicated by the increased distances between the amino acid sidechains that cause the incorrect positioning of the nicotinamide ring of NADPH. **e**. Continuous turnover scheme in which the open-form subunits (A and B) become closed by releasing the product and NADP⁺ while, concurrently, the closed subunits (C and D) bind the substrate and NADPH thus becoming closed subunits. These alternating cycles are aided by the swing motions of tightly coupled catalytic subunits like two blades swinging synchronously.

4.5. Materials and Methods

Amplification & cloning of K. setae ECR

The *K. setae* enoyl-CoA carboxylase reductase (ECR) coding sequence was codon optimized using the *E. coli* codon frequency table, and synthesis constraints were removed using the Build Optimization Software Tools (BOOST) developed by DOE-Joint Genome Institute (JGI), USA[28]. Overlapping synthetic DNA fragments were obtained from Thermo Fisher Scientific and cloned into the NdeI site of the pET16b vector (Novagen) by using the Gibson Assembly HiFi kit (SGI-DNA). The resulted colonies were sequence verified by the PacBio sequencing platform.

Site-directed Mutagenesis of K. setae ECR

Mutations are introduced by the similar methods as described in the previous section. Two fragments flanking the mutagenesis site were amplified and the Gibson assembly was performed as described above. Below is the FASTA sequence of the ECR protein and list of primers that we used to introduce catalytic site single mutations Y328F, R303K, and K296A, and the triple mutant K296A/R303K/Y328F.

tr|E4N096|E4N096_KITSK Putative crotonyl-CoA reductase OS= Kitasatospora setae

MQEILDAILSGDAASADYAALALPESYRAVTLHKGEERMFDGLASRDKDPRKSLHLDDVP

LPELGPGEALVAVMASSVNYNTVWSSIFEPVSTFGFLERYGRLSPLTARHDLPYHVLGSD

LAGVVLRTGAGVNAWKPGDEVVAHCLSVELESPDGHNDTMMDPEQRIWGFETNFGGLAQL

ALVKTNQLLPKPKHLTWEEAASPGLVNSTAYRQLVSRNGAGLKQGDNVLIWGASGGLGSY

ATQYALAGGATPICVVSSPRKADICRAMGAEAIIDRSAEGYRFWKDEHHQDPREWKRLGG

KIREFTGGEDVDIVFEHPGRETFGASVYVTRKGGTIVTCASTSGYMHQYDNRYLWMSLKR

IVGSHFANYREAFEANRLVAKGKIHPTLSKVYALEETGQAALDVHHNKHQGKVGVLCLAP

REGLGVTDPELRSKHLTKINAFRNV

Single mutations were introduced with the following *_F&*_R primer pairs:

E4N096_DMP_064_Y328F_F: catccgtgttcgtgacccgcaaaggtggcactatcg

E4N096_DMP_064_Y328F_R: gcgggtcacgaacacggatgcaccgaaggtttcgcg

E4N096_DMP_064_R303K_F: ggtggcaaaatcaaggaattcaccggtggggaagacgtgg

E4N096_DMP_064_R303K_R: aattccttgattttgccacccagacgtttccactcacg

E4N096_DMP_064_K296A_F: agtgggcccgtctgggtggcaaaatccgtgaattcaccg

E4N096_DMP_064_K296A_R: ccagacgggcccactcacgcgggtcttggtggtgttcg

The double and triple mutants were introduced in the following order: By using Y328F plasmid we introduced R303K mutation to generate double mutant Y328F/R303K

For the triple mutant we used the following special primers pair

E4N096_DMP_064_triple_F: agtgggcccgtctgggtggcaaaatcaaggaattcaccg

E4N096_DMP_064_triple_R: ccagacgggcccactcacgcgggtcttggtggtgttcg

Mutagenesis of K. setae ECR subunit interface residues

Two fragments franking the mutagenesis site were amplified and the Gibson assembly was performed as with the list of primers that we used to introduce various combinations of subunit interface single mutations N157E, N218E, E151D, E151R, E151K, E151L, E151I

Seven single mutants were introduced with the following *_F&*_R primer pairs:

ECR_N157E_F: ccggacggtcacgaagacactatgatggacccagagcagc

ECR_N157E_R: catcatagtgtcttcgtgaccgtccggagattccagttcaacagacagg

ECR_N218E_F: gctggtgtctcgtgaaggcgccggcctgaaacagggtgacaacg

ECR_N218E_R: caggccggcgccttcacgagacaccagctgacgataagcggtagagttaacg

ECR_E151D_F: gtctgttgaactggattctccggacggtcacaacgacactatgatgg

ECR_E151D_R: gaccgtccggagaatccagttcaacagacaggcagtgagcaaccacctcgtcacc

ECR_E151R_F: gtctgttgaactgaggtctccggacggtcacaacgacactatgatgg

ECR_E151R_R: gaccgtccggagacctcagttcaacagacaggcagtgagcaaccacctcgtcacc

ECR_E151K_F: gtctgttgaactgaagtctccggacggtcacaacgacactatgatgg

ECR_E151K_R: gaccgtccggagacttcagttcaacagacaggcagtgagcaaccacctcgtcacc

ECR_E151L_F: gtctgttgaactgctgtctccggacggtcacaacgacactatgatgg

ECR_E151L_R: gaccgtccggagacagcagttcaacagacaggcagtgagcaaccacctcgtcacc

ECR_E151I_F: gtctgttgaactgatctctccggacggtcacaacgacactatgatgg

ECR_E151I_R: gaccgtccggagagatcagttcaacagacaggcagtgagcaaccacctcgtcacc

Five triple mutants were obtained with the following *_F&*_R primer pairs respectively (third mutation varies):

1) Asn157Glu, Asn218Glu, Glu151Asp

   ECR_N157E_N218E_E151D_F:

   ctgtctgttgaactggattctccggacggtcacgaagacactatgatggacccagagcagcgcatctgg

   ECR_N157E_N218E_E151D_R:

   gtccatcatagtgtcttcgtgaccgtccggagaatccagttcaacagacaggcagtgagcaaccacctcg

2) Asn157Glu, Asn218Glu, Glu151Arg

   ECR_N157E_N218E_E151R_F:

   ctgtctgttgaactgaggtctccggacggtcacgaagacactatgatggacccagagcagcgcatctgg

   ECR_N157E_N218E_E151R_R:

   gtccatcatagtgtcttcgtgaccgtccggagacctcagttcaacagacaggcagtgagcaaccacctcg

3) Asn157Glu, Asn218Glu, Glu151Lys

   ECR_N157E_N218E_E151K_F:

ctgtctgttgaactgaagtctccggacggtcacgaagacactatgatggacccagagcagcgcatctgg

ECR_N157E_N218E_E151K_R:

gtccatcatagtgtcttcgtgaccgtccggagacttcagttcaacagacaggcagtgagcaaccacctcg

4) Asn157Glu, Asn218Glu, Glu151Leu

ECR_N157E_N218E_E151L_F:

ctgtctgttgaactgctgtctccggacggtcacgaagacactatgatggacccagagcagcgcatctgg

ECR_N157E_N218E_E151L_R:

gtccatcatagtgtcttcgtgaccgtccggagacagcagttcaacagacaggcagtgagcaaccacctcg

5) Asn157Glu, Asn218Glu, Glu151Ile

ECR_N157E_N218E_E151I_F:

ctgtctgttgaactgatctctccggacggtcacgaagacactatgatggacccagagcagcgcatctgg

ECR_N157E_N218E_E151I_R:

gtccatcatagtgtcttcgtgaccgtccggagagatcagttcaacagacaggcagtgagcaaccacctcg

Mutagenesis of Adenine binding residues and intra-dimer communication residues

Variants of the *K. Setae* ECR were generated with the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA) using 60 ng of template plasmid and the following forward and reverse primer pairs:

Mutation	Forward Primer	Reverse Primer
R303A	CTGGGTGGCAAAATCGCTGAATTCACCGGGTG	CTGGGTGGCAAAATCGCTGAATTCACCGGTG
R303V	CTGGGTGGCAAAATCGTGGAATTCACCGGGTGGG	CTGGGTGGCAAAATCGTGGAATTCACCGGTGGG
K332A	GTACGTGACCCGCGCAGGTGGCACTATC	GTACGTGACCCGCGCAGGTGGCACTATC
Q165A	CACTATGATGGACCCAGAGGCACGCATCTGGGGCTTCGAAAC	CACTATGATGGACCCAGAGGCACGCATCTGGGGCTTCGAAAC

Cell lysis, protein purification, and characterization

The cells were harvested by centrifugation (3000 rpm, 30 min) and the cell pellet was pooled. The pellet was resuspended in a lysis buffer containing 50 mm Tris-HCl pH 8.5, 1 M NaCl, 5% glycerol supplemented with 100µl Triton x100 per 100ml of final buffer volume (Sigma-Aldrich). The suspension was sonicated at 50% amplitude for 30 seconds three times. Immediately after the lysis, the suspension was ultra-centrifuged at 33,000 rpm for 40 minutes at 4°C.

The soluble fraction was pooled and was applied to a 10 ml Ni-NTA column and purified using an AKTA prime FPLC setup. The column was washed with 2 column volumes of HisA loading buffer (50 mM Tris-HCl pH 8.5, 300 mM NaCl, 10 mM imidazole) for equilibration. Preliminary attempts of His-tag purification were unsuccessful since the protein would precipitate out of solution during application to the column. This was remedied by adding 1 M L-proline (Sigma-Aldrich) to the lysis and HisB elution buffers to ensure the protein remains soluble. The soluble portion was then applied to Ni-NTA column, and then eluted using HisB elution buffer containing 50 mM Tris-HCl pH 8.5, 300 mM NaCl, 500 mM imidazole. The eluted fractions were collected on a fraction collector, and their purities were analyzed by SDS-PAGE, and pure fractions were pooled and concentrated to 10 mg/ml using MilliPore Amicon Ultra 30KDa molecular-weight cutoff concentrators.

Determination of the oligomeric state of KsCcr

Oligomeric state of *K. setae* ECR was determined by analytical size-exclusion chromatography. 260 µl containing 500 µg of purified protein were injected into a pre-equilibrated S200 INCREASE 10-300GL (GE Healthcare) column. Runs were performed using a 100 mM KH_2PO_4 pH=8.0 buffer at a flow of 0.75 ml/min. Protein size was determined by comparing the obtained retention volumes (RV) with a Gel filtration standard protein mixture (BioRad).

Protein	RV (mL)	MW (kDa)
K. setae ECR WT	12.98	205.1
K. setae ECR E151D	12.97	202.8
K. setae ECR N157E	13.00	199.1
K. setae ECR N218E	13.01	197.7
K. setae ECR E151DN157E N218E	12.92	207.5
Gel filtration standard		
Thyroglobulin	9.45	670

|		
Gamma-globulin	12.65	158
Ovalbumin	15.52	44
Myoglobin	17.30	17

Spectrophotometric Enzyme assays

Assays were performed on a Cary-60 UV/Vis spectrophotometer (Agilent) at 30°C using quartz cuvettes (1 or 10 mm path length; Hellma). Reactions contained 20 µg/ml carbonic anhydrase and were performed in 100 mM K_2HPO_4 pH = 8.0. Kinetic parameters for one substrate were determined by varying its concentration while the others were kept constant at 10 times their K_M value. Reaction procedure was monitored by following the oxidation of NADPH at 365 nm ($\varepsilon_{NADPH,365nm} = 3.33$ M^{-1} cm^{-1}). Each concentration was measured in triplicates and the obtained curves were fit using GraphPad Prism 8. Hyperbolic curves were fit to the Michaelis-Menten equation to obtain apparent k_{cat} and K_M values. For mutants revealing substrate inhibition, the data was fit to $v_0 = (V_{Max} [S])/(K_M + [S] ((1+[S])/K_i))$.

Chemical Synthesis of CoA-esters

Crotonic Anhydride, Carbonic anhydrase from bovine erythrocytes, 1,1-Carbonyldiimidazole (CDI) and 4-dimethylaminopyridine (DMAP) were purchased from Sigma Aldrich AG, Coenzyme A trilithium from Roche Diagnostics, NADPH Na$_4$ (98%) and pyridine from Carl Roth GmbH. Solvents and salts were all analytical grade or better. Crotonyl-CoA was synthesized as previously reported[2]. Briefly, 200 mg of CoA trilithium salt were dissolved in 4 ml of 0.4 M KHCO$_3$ and stirred on ice for 45 min. After addition of 64 µl of crotonic anhydride the reaction procedure was tested by mixing 5 µl of reaction mixture with 20 µl of an aqueous solution of DTNB (5,5'-dithio-bis-[2-nitrobenzoic acid]). Crotonyl-CoA was purified by Preparative RFLC/MS over a Gemini 10 µm NX-C18 110 Å, 100 x 21.2 mm, AXIA packed column (Phenomenex) using a methanol gradient from 5% to 35% over 15 min with 25 mM ammonium formate pH = 8.1 (Buffer 8.1) as the aqueous phase. Fractions containing the product were pooled, lyophilized and stored at -20°C.

Synthesis of crotonyl-pantetheine (**3**) was performed according to scheme 1 as previously reported[29].
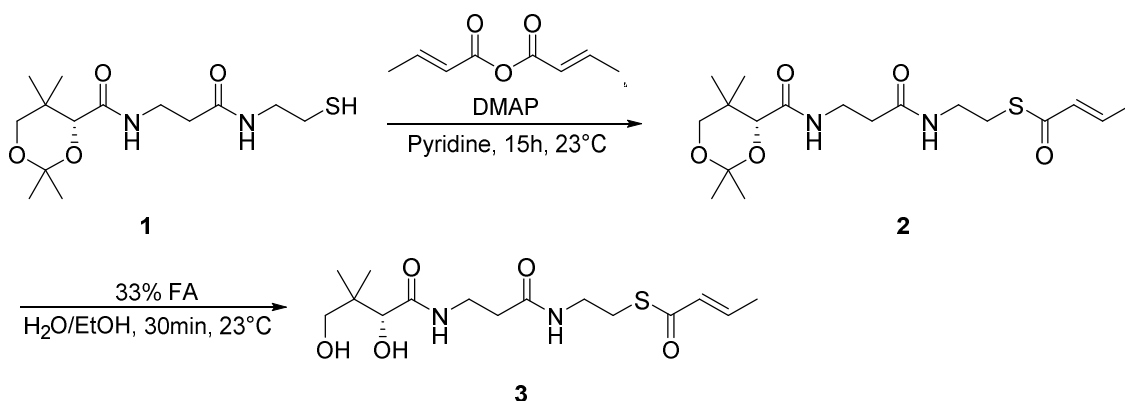
Scheme 1: Reaction conditions for the synthesis of crotonyl-pantetheine.

Pantetheine **1** (0.50 g, 1.57 mmol), DMAP (0.02 g, 0.19 mmol) and crotonic anhydride (0.50 ml, 3.37 mmol) in pyridine (12.5 ml) were stirred for 15 h at 23° C then 1 h at 50 °C. Pyridine was removed under reduced pressure and the product dissolved in saturated aqueous NaHCO₃ (1 ml) and water (1 ml). The aqueous phase was extracted with CH₂Cl₂ (3x 5 ml), dried over MgSO₄, filtered and the solvent removed under reduced pressure. The obtained product was purified over FC (SiO₂; EtOAc/hexane, 2:3 → EtOAc) to afford **2**. **2** then stirred in Water/EtOH/FA, 1:1:1 for 30 min at 23 °C. After completion the solution was lyophilized, the solid dissolved in 0.5 % aqueous TFA and then purified with HPLC (C18; 25mM Ammonium formate pH = 8.1/MeOH, 5% → 95%) and lyophilized to yield **3** as a transparent thick oil. For use in assays, the compound was resuspended in water and stored at -20°C if not used.

Analysis of Ethylmalonyl-CoA stability

Reactions to measure the stability of ethylmalonyl-CoA under crystallization conditions were performed in 200 mM TrisHCl pH = 7.5, 20% Polyacrylic acid sodium salt 5100 at 19°C. Reactions contained 600 µM Ethylmalonyl-CoA, NADP⁺ µM and 1 µM KsECR WT. Samples were quenched at different time points using 50% formic acid and spinned at 17'000 g for 10 min to precipitate the protein. The reaction was diluted 10 times into 5% methanol/Buffer 8.1 and analyzed by UHPLC over a Sonoma C18(2), 3 µm 100 Å, 100 x 2.1 mm using a 5 to 45% methanol gradient over 14.5 min.

Crystallization of K.setae ECR complexes

72-well sitting-drop crystallization trays (Terasaki) were set up and screened against a library of various crystallization conditions (Molecular Dimensions, Hampton). Each crystallization well contained 0.77 µl of 10 mg/ml *K. setae* ECR protein kept in 500 mM Imidazole, 300mM NaCl, 1M proline and TRIS-HCl pH 8.5 mixed with 0.77 µL of the various crystallization buffers. Each well was sealed with 16.6 µL of

100% paraffin oil (Hampton Research) to slow the crystallization process. Crystals of apo ECR protein were observed in various morphologies after 24 hours of incubation. The initial crystallization conditions were from various MIDAS, Crystal Screen, and PGA-LM screening conditions (Molecular Dimensions, Hampton Research). The apo ECR was crystallied from a solution containing 100 mM TRIS pH 8.0 and 20% w/v poly (acrylic acid sodium salt) 5100 and resulted in 30-micron plate-like crystals. It is important to note that all structures were solved using this condition as basis, with addition reagents as needed. The binary and ternary ECR complexes were co-crystallized with final concentration of 5 mM of each respective ligand and cofactor with a protein concentration of 10 mg/ml. Alternative crystallization conditions were used either to obtain larger crystals for higher resolution synchrotron structures or higher microcrystal density for SFX experiments. For the crystallization of *K. setae* ECR-butyrylCoA-NADPH ternary complex, the crystallization condition contained 17% w/v PEG 10000, 100 mM BisTris-HCl pH 7.5, 100 mM ammonium acetate and resulted in 50-micron plate-like crystals. For *K. setae* ECR-NADP+ binary complex, the crystallization solution contained 0.2 M ammonium formate, 10% (w/v) polyvinylpyrrolidone, 20% (w/v) PEG 4000 and resulted in 50-micron plate-like crystals. No further seeding was required for any of the synchrotron structures, and crystals were harvested after 30 minutes incubation with 30% (v/v) glycerol as a cryoprotectant. For the SFX experiments, neither of the synchrotron crystallization conditions of crystals would be sufficient due to size limitations and an optimal crystal density of 10^9 to 10^{11} crystals/ml could not be obtained. To test various crystal conditions, a batch method was employed with equal parts of protein and crystal condition to see if increased crystal densities could be achieved in 15 mL Corning conical falcon tubes. Initial tests were total volume of 1 ml (0.5 ml 10 mg/ml protein and 0.5 ml crystal condition) and incubated for 48 hours. From the tubes with crystals present, the 1 ml crystal slurry was used to seed a 10 ml total crystallization solution. The best crystals were obtained in final 10 ml sample solution consisted of the 1 ml seed crystal slurry solution, 4.5 ml of 10 mg/ml protein solution and 4.5 ml of crystallization buffer containing 0.03M Magnesium chloride hexahydrate, 0.03M Calcium chloride dihydrate, 0.05M imidazole, 0.05M MES-KOH pH 6.5, 15% v/v glycerol and 15% v/v PEG. Prior to sample injection, the crystals were filtered using a 20-micron nylon mesh filter to separate the contaminant of large crystals from the smaller ones (Millipore).

Data collection, processing and structure determination

For the apo, ternary, and NADP+ binary complex structures, the crystals were flash cooled in liquid nitrogen. The apo (1.8 Å) and NADP+ complex (1.75 Å) diffraction datasets were collected at 100 K on

Beamline 23ID-B, the Advanced Photon Source, Argonne National Laboratory (Argonne, Illinois, USA), equipped with an Eiger 16M detector. The butyryl-CoA ternary complex (1.7 Å) diffraction dataset was collected at 100 K on Beamline 12-2 at the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory (Menlo Park, California, USA), equipped with a Dectris Pilatus 6M detector. The *K. setae* apo crystals belonged to the space group $P2_12_12_1$ with unit cell dimensions a= 78.1 Å, b= 153.0 Å, c= 202.7 Å and α=β=γ=90°. The *K. setae* ternary complex crystals belonged to the space group $P2_1$ with unit cell dimensions of a = 109.3 Å, b = 78.8 Å, c = 138.8 Å and α=90°, β=108.1°, γ=90°. The *K. setae* NADP+ complex crystals belonged to the space group $P2_12_12_1$ with unit cell dimensions of a = 77.0 Å, b = 146.7 Å, c = 200.2 Å and α=β= γ=90°. The *K. setae* ECR-NADPH binary complex was determined using serial femtosecond X-ray crystallography (SFX) at an X-ray Free Electron Laser (XFEL) and was carried out on May 2017 at SACLA beamline 3 (Hyogo, Japan) (Proposal number 2017A8055)[30]. The SALCA beam had a pulse duration of 10 fs. The photon energy was 10kEV. The in air concentric Electrokinetic Microfluidic Sample Holder (coMESH) injector[27] installed at DAPHNIS[9] chamber was used to introduce samples suspended in mother liquor to the 10 fs-long X-ray pulses. X-ray diffraction data was recorded by using the multiport CCD (MPCCD)[31] detector. Data analysis was performed on the SACLA High Performance Computing Cluster consisting of several steps of parameter optimization. Diffraction images were collected with consistent experimental parameters (attenuation, transmission, detector distance etc.) during one 12-hour shift. Crystal hits were identified with the program Cheetah[32]. The raw data were processed with CrystFEL's indexamajig against given cell parameters of the *K. setae* NADPH (XFEL) complex microcrystals belonging to the space group $P2_1$ with unit cell dimensions of a = 109.8 Å, b = 78.1 Å, c = 138.9 Å and α=90°, β=107.8°, γ=90°.

The data processing for synchrotron structures were carried out using *autoXDS* and scaling was done with XSCALE[33, 34]. A set of 5% of randomly chosen reflections were set aside for the calculation of the free R factor (R_{free}). The apo structure was solved using by *PHENIX*[35, 36] and *PHASER*[37, 38] molecular replacement program. Initial search model for molecular replacement is generated by using *SWISS-MODEL*[39] server against an unpublished CCR structure of a putative crotonyl-CoA carboxylase/reductase (PDB code 4GI2, deposited by S. Weidenweber, T.J. Erb, U. Ermler). The *K. setae* apo structure served as the model for solving the binary and ternary-complex synchrotron structures and also SFX structure. This resulted in four monomers in the asymmetric unit. The refinement was carried out using *PHENIX* refinement, utilizing automatically generated TLS groups based on the structure and ordered solvent to place the water molecules[40, 41]. Following the first round of refinement, the structure was manually adjusted to the electron density and waters were added using *COOT* at one sigma cutoff[42, 43]. The *K. setae*/NADP+ complex also shares the same space group as the

apo form, and was solved directly using Phenix molecular replacement[35, 36]. The NADP+ structure and restraint files were taken from previously solved CCR/NADP+ complexes (PDB: 4Y0K).

MD and QM/MM studies of the open and closed subunits of ECR

The flexibility of NADPH in the open and closed cavity were studied with a binary complex using the CHARMM22 force field[44, 45] and parameters from Pavelites *et al.* for NADPH[46] in 200 ns explicit solvent simulations (TIP3P) at 298 K and 1 atm with the AMBER16 software package[47] (dt = 2 fs, tau = 1 ps, PME cutoff = 8.0 Å , SHAKE). Additionally, in three independent 200 ns simulations NADPH was positioned in the open cavity in the conformation observed in the closed cavity to test if these conformations are also visited in the open form.

To study crotonyl-CoA binding in the ternary complex we first extracted a dimer with one open- and another closed-subunits (subunit A & C) from the butyryl-CoA/NADPH ECR ternary X-Ray structure. We added the unresolved butyryl-CoA molecule to the open subunit aligning the protein chains from the closed subunit on the open one and shifting the butyryl-CoA coordinates from the closed subunit. The butyryl-CoA molecules were then modified to obtain the substrate crotonyl-CoA deleting the two hydrogen atoms. The resulting dimer consisted of one closed and one open subunit each with NADPH and one crotonyl-CoA molecule. The system was solvated and equilibrated (500ps NVT, 5ns NPT, 100ns NVT) as described above and substrate and NADPH were restrained to their initial configuration to relax the protein and the solvent. From these equilibrated configurations five structures were randomly extracted to study the behavior of the cofactor and substrate molecules. Five trajectories of 2 ns each for the closed and open cavity were performed with the QM/MM method using the DFTB3 Hamiltonian[48] and the 3ob parameter[49, 50] set to describe NADPH and the crotonyl fragment.

An electrostatic embedding using the link atom method at 298K and 1 atm was used together with a time step of 1 fs in AMBER16 software package. Parameters for the CoA fragment of the substrate were taken from Aleksandrov *et al*[51].

4.6. References

1.	Schwander T, Schada von Borzyskowski L, Burgener S, Cortina NS, & Erb TJ, A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* 354(6314):900-904 **2016**.
2.	Erb TJ*, et al.*, Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A.* 104(25):10631-10636 **2007**.
3.	Erb TJ, Brecht V, Fuchs G, Muller M, & Alber BE, Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Natl. Acad. Sci. U.S.A.* 106(22):8871-8876 **2009**.
4.	Erb TJ & Zarzycki J, Biochemical and synthetic biology approaches to improve photosynthetic CO2-fixation. *Curr. Opin. Chem. Biol.* 34:72-79 **2016**.
5.	Rosenthal RG*, et al.*, Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* 10(1):50-55 **2014**.
6.	Quade N, Huo L, Rachid S, Heinz DW, & Muller R, Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* 8(1):117-124 **2011**.
7.	Zhang L*, et al.*, Rational Control of Polyketide Extender Units by Structure-Based Engineering of a Crotonyl-CoA Carboxylase/Reductase in Antimycin Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13462-13465 **2015**.
8.	Rao ST & Rossmann MG, Comparison of super-secondary structures in proteins. *J. Mol. Biol.* 76(2):241-256 **1973**.
9.	Tono K*, et al.*, Diverse application platform for hard X-ray diffraction in SACLA (DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser. *J. Synchrotron Radiat.* 22(3):532-537 **2015**.
10.	Chapman HN*, et al.*, Femtosecond X-ray protein nanocrystallography. *Nature* 470(7332):73-77 **2011**.
11.	Schlichting I & Miao J, Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* 22(5):613-626 **2012**.
12.	Helliwell JR, Biochemistry. How to solve protein structures with an X-ray laser. *Science* 339(6116):146-147 **2013**.
13.	Vogeli B*, et al.*, Combining Promiscuous Acyl-CoA Oxidase and Enoyl-CoA Carboxylase/Reductases for Atypical Polyketide Extender Unit Biosynthesis. *Cell Chem. Biol.* 25(7):833-839 e834 **2018**.
14.	Bernhard SA & MacQuarrie RA, Half-site reactivity and the "induced-fit" hypothesis. *J. Mol. Biol.* 74(1):73-78 **1973**.
15.	Seydoux F, Malhotra OP, & Bernhard SA, Half-site reactivity. *CRC Crit. Rev. Biochem.* 2(2):227-257 **1974**.
16.	Levitzki A, Stallcup WB, & Koshland DE, Jr., Half-of-the-sites reactivity and the conformational states of cytidine triphosphate synthetase. *Biochemistry* 10(18):3371-3378 **1971**.
17.	Fersht AR, Mulvey RS, & Koch GL, Ligand binding and enzymic catalysis coupled through subunits in tyrosyl-tRNA synthetase. *Biochemistry* 14(1):13-18 **1975**.
18.	Hill TL & Levitzki A, Subunit neighbor interactions in enzyme kinetics: half-of-the-sites reactivity in a dimer. *Proc. Natl. Acad. Sci. U.S.A.* 77(10):5741-5745 **1980**.
19.	Levitzki A & Koshland DE, Jr., The role of negative cooperativity and half-of-the-sites reactivity in enzyme regulation. *Curr. Top. Cell. Regul.* 10:1-40 **1976**.

20. Vivoli M, Pang J, & Harmer NJ, A half-site multimeric enzyme achieves its cooperativity without conformational changes. *Sci. Rep.* 7(1):16529 **2017**.

21. Danenberg KD & Danenberg PV, Evidence for a sequential interaction of the subunits of thymidylate synthetase. *J. Biol. Chem.* 254(11):4345-4348 **1979**.

22. Anderson AC, O'Neil RH, DeLano WL, & Stroud RM, The structural mechanism for half-the-sites reactivity in an enzyme, thymidylate synthase, involves a relay of changes between subunits. *Biochemistry* 38(42):13829-13836 **1999**.

23. Finer-Moore JS, Lee TT, & Stroud RM, A Single Mutation Traps a Half-Sites Reactive Enzyme in Midstream, Explaining Asymmetry in Hydride Transfer. *Biochemistry* 57(19):2786-2795 **2018**.

24. Hammes GG, Benkovic SJ, & Hammes-Schiffer S, Flexibility, diversity, and cooperativity: pillars of enzyme catalysis. *Biochemistry* 50(48):10422-10430 **2011**.

25. Peter DM Substrate Promiscuity, Kinetics and Engineering of Enoyl-CoA Carboxylases/Reductases. Thesis (ETH Zürich) **2016**

26. Stagno JR*, et al.*, Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* 541(7636):242-246 **2017**.

27. Sierra RG*, et al.*, Concentric-flow electrokinetic injector enables serial crystallography of ribosome and photosystem II. *Nat. Methods* 13(1):59-+ **2016**.

28. Oberortner E, Cheng JF, Hillson NJ, & Deutsch S, Streamlining the Design-to-Build Transition with Build-Optimization Software Tools. *ACS Synth. Biol.* 6(3):485-496 **2017**.

29. Agarwal V*, et al.*, Chemoenzymatic Synthesis of Acyl Coenzyme A Substrates Enables in Situ Labeling of Small Molecules and Proteins. *Org. Lett.* 17(18):4452-4455 **2015**.

30. Ishikawa T*, et al.*, A compact X-ray free-electron laser emitting in the sub-angstrom region. *Nat. Photonics* 6(8):540-544 **2012**.

31. Kameshima T*, et al.*, Development of an X-ray pixel detector with multi-port charge-coupled device for X-ray free-electron laser experiments. *Rev. Sci. Instrum.* 85(3) **2014**.

32. Barty A*, et al.*, Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* 47:1118-1131 **2014**.

33. Kabsch W, Xds. *Acta Crystallogr. D* 66:125-132 **2010**.

34. Kabsch W, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* 66:133-144 **2010**.

35. Adams PD*, et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66:213-221 **2010**.

36. Adams PD*, et al.*, PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* 58:1948-1954 **2002**.

37. Mccoy AJ*, et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* 40:658-674 **2007**.

38. McCoy AJ, Acknowledging Errors: Advanced Molecular Replacement with Phaser. *Methods Mol. Biol.* 1607:421-453 **2017**.

39. Waterhouse A*, et al.*, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46(W1):W296-W303 **2018**.

40. Winn MD, Isupov MN, & Murshudov GN, Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* 57(Pt 1):122-133 **2001**.

41. Winn MD, Murshudov GN, & Papiz MZ, Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* 374:300-321 **2003**.

42. Emsley P, Lohkamp B, Scott WG, & Cowtan K, Features and development of Coot. *Acta Crystallogr. D* 66(Pt 4):486-501 **2010**.

43. Emsley P & Cowtan K, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* 60(Pt 12 Pt 1):2126-2132 **2004**.

44. Mackerell AD, Jr., Feig M, & Brooks CL, 3rd, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11):1400-1415 **2004**.

45. MacKerell AD*, et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616 **1998**.

46. Pavelites JJ, Gao JL, Bash PA, & Mackerell AD, A molecular mechanics force field for NAD(+), NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem.* 18(2):221-239 **1997**.

47. Case DA*, et al.* **2017** Amber 2017.

48. Gaus M, Cui QA, & Elstner M, DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* 7(4):931-948 **2011**.

49. Gaus M, Goez A, & Elstner M, Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* 9(1):338-354 **2013**.

50. Gaus M, Lu XY, Elstner M, & Cui Q, Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* 10(4):1518-1537 **2014**.

51. Aleksandrov A & Field M, Efficient solvent boundary potential for hybrid potential simulations. *Phys. Chem. Chem. Phys.* 13(22):10503-10509 **2011**.

4.7. Supplementary Information

Supplementary Table 1. Data collection and refinement statistics

	K. setae ECR Apo form	*K. setae* ECR NADP⁺ Butyryl-CoA	*K. setae* ECR NADP⁺	*K. setae* ECR NADPH (XFEL)
Data collection				
Space group	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_1$
Cell dimensions				
a, b, c (Å)	78.1, 153.0, 202.7	109.3, 78.8, 138.8	77.0, 146.7, 200.2	109.8, 78.1, 138.9
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 108.1, 90.0	90.0, 90.0, 90.0	90.0, 107.8, 90.0
Resolution (Å)	16.9 – 1.80	38.4 – 1.78	47.7 – 1.75	31.8 – 2.34
	(1.86 – 1.80)	(1.84 – 1.78)	(1.81 – 1.75)	(2.42 – 2.34)
R_{sym} or R_{merge}	0.12 (2.38)	0.08 (0.78)	0.17 (3.26)	-
R_{split}	-	-	-	0.31 (1.36)
I / σI	12.3 (1.56)	16.6 (2.07)	11.9 (0.89)	6.7 (0.74)
Completeness (%)	100.0 (100.0)	98.8 (98.5)	96.5 (88.5)	100.0(100.0)
Redundancy	6.14	6.67	6.21	129.2
Refinement				
Resolution (Å)	16.93 – 1.80	38.34 – 1.78	47.69 – 1.75	31.75 – 2.34
	(2.42 – 2.34)	(1.81 – 1.78)	(1.82 – 1.75)	(2.42 – 2.34)
No. reflections	223907 (18723)	227701 (16317)	220094 (24907)	93978 (9367)
R_{work} / R_{free}	0.20/0.24	0.17/0.20	0.19/0.21	0.28/0.31
No. atoms				
Protein	13791	14067	13796	13796
Ligand/ion	45	349	192	192
Water	1369	1885	1749	587
B-factors				
Protein	48.6	25.6	33.0	35.9
Ligand/ion	54.2	33.2	39.4	41.6
Water	48.2	35.3	37.3	37.1
R.m.s. deviations				
Bond lengths (Å)	0.010	0.006	0.007	0.015
Bond angles (°)	1.17	0.99	1.15	1.53

*Single crystal used for all datasets except the XFEL NADPH dataset.

**Values in parentheses are for highest-resolution shell.

Apo

Chain	RMSD (Å)
A/B	0.11
A/C	0.08
A/D	0.11
B/C	0.11
B/D	0.08
C/D	0.08

NADPH complex

Chain	RMSD (Å)
A/B	0.50
A/C	1.84
A/D	1.78
B/C	1.74
B/D	1.70
C/D	0.52

ECR + NADPH + Butyryl-CoA

Chain	RMSD (Å)
A/B	0.10
A/C	2.04
A/D	2.06
B/C	2.07
B/D	2.04
C/D	0.08

NADP+ complex

Chain	RMSD (Å)
A/B	0.18
A/C	0.70
A/D	0.25
B/C	0.85
B/D	0.35
C/D	0.49
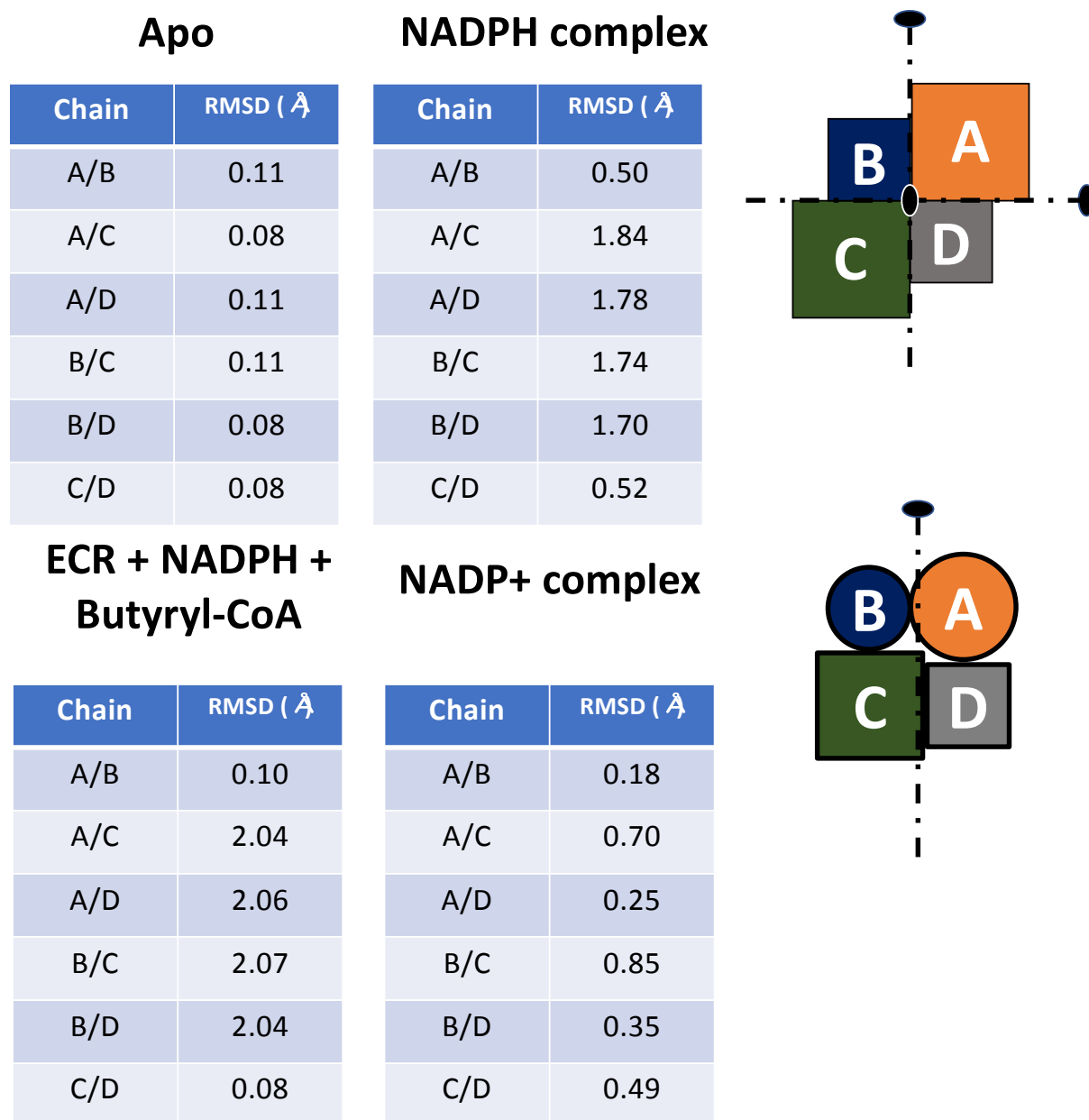
Supplementary Figure S1: Root-mean-squared deviations of the 4 subunits from the different ECR crystal structures. The Cα main chain atoms from each subunit were used for the alignments. The reported values in the left 4 tables are in Å. In the ternary complex, subunits A and B are open-form subunits having only cofactors while C and D are closed-form subunits which have the cofactors and butyryl-CoA bound. Top right: the arrangement of the 4 subunits in the apo and NADP$^+$ bound ECR with dihedral D2 symmetry, 3 two-fold symmetry axes are shown with ovals and broken lines. Bottom right: the arrangement of the 4 subunits in the NADPH and butyryl-CoA/NADPH bound ECR, with vertical cyclic C2 axis shown by a set of oval and broken line. The viewing angles of these two are the same as in Figures 2a and 3a.
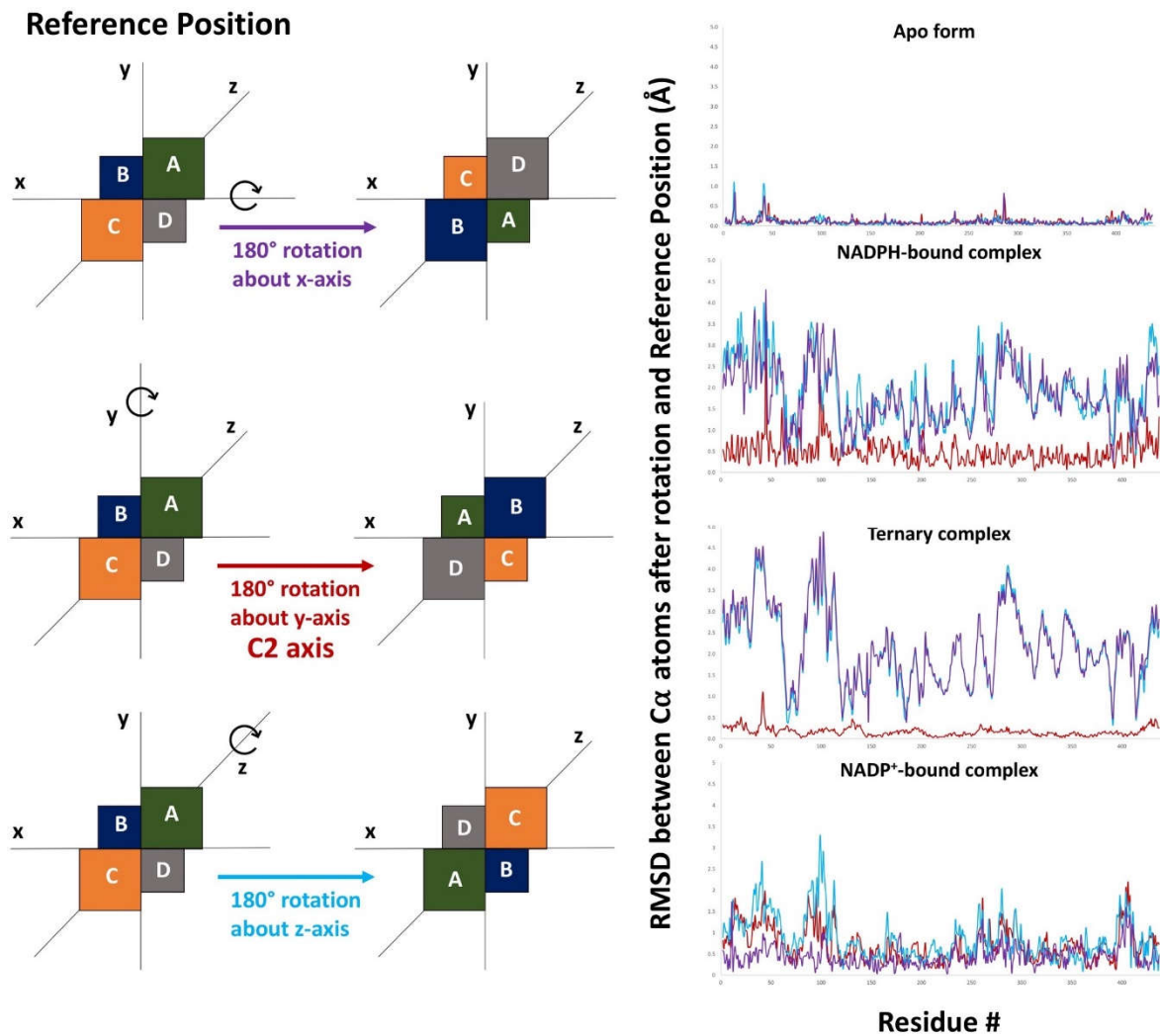
Supplementary Figure S2: This figure shows the disparity among subunits to reveal the broken symmetry. The left-hand side shows the various rotations that were used to test the symmetry. The right-hand side graphs the RMSD values of the post-rotation aligned with the reference position against each residue. The rotation axes are color coded as follows: x-axis = purple, y-axis = Red, z-axis = cyan.
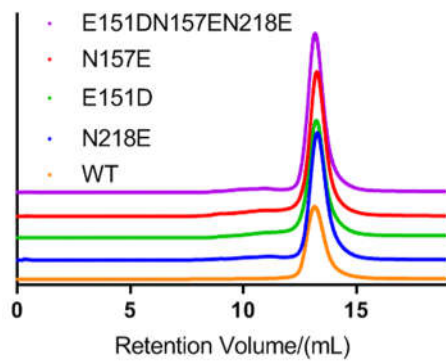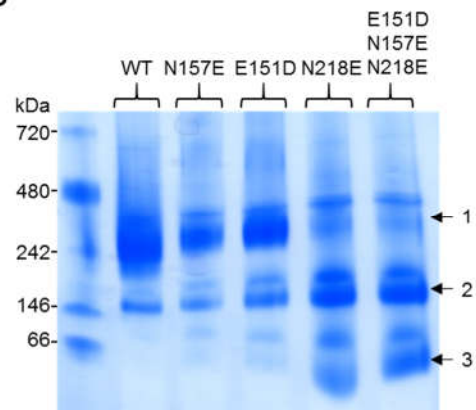
Supplementary Figure S3: A) Size-exclusion chromatography of *K. setae* ECR WT and variants. The WT protein and its variants elute at the same retention volume confirming the expected size of the protein. B) Blue native PAGE of *K. setae* ECR wildtype and variants. Oligomeric states according to numbering 1: tetramer 2: dimer 3: monomer.
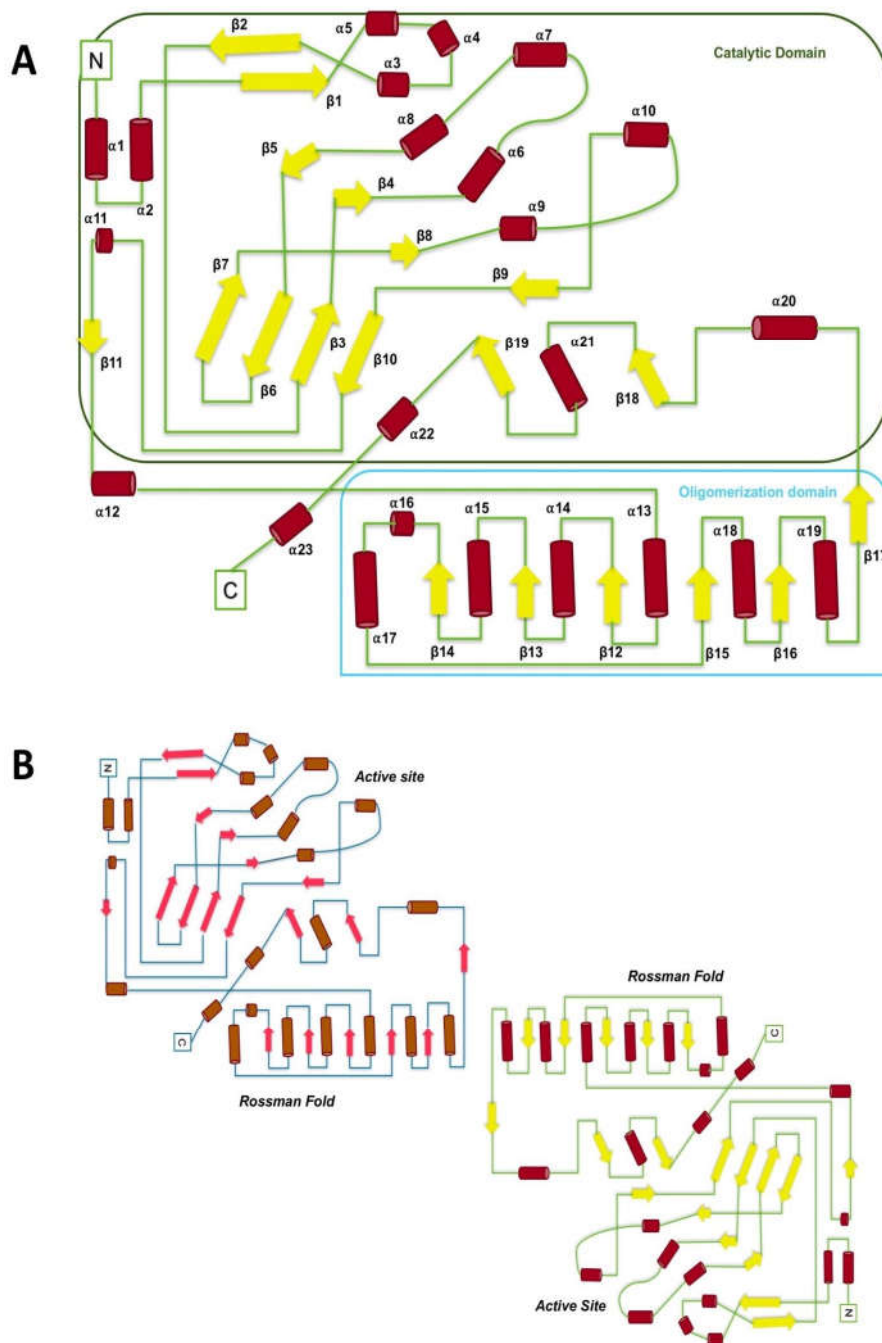
Supplementary Figure S4: Topology diagrams of both a monomer and a dimer which shows the central β-sheet network. A) The monomeric representation shows the disparity between the flexible loops of the catalytic domain and the rigid Rossmann fold in the oligomerization domain. B) This dimer representation shows how the 12 β-strand Rossmann fold network is formed which links the two subunits. The active site is also highlighted to show that is on the edge of the catalytic domain.
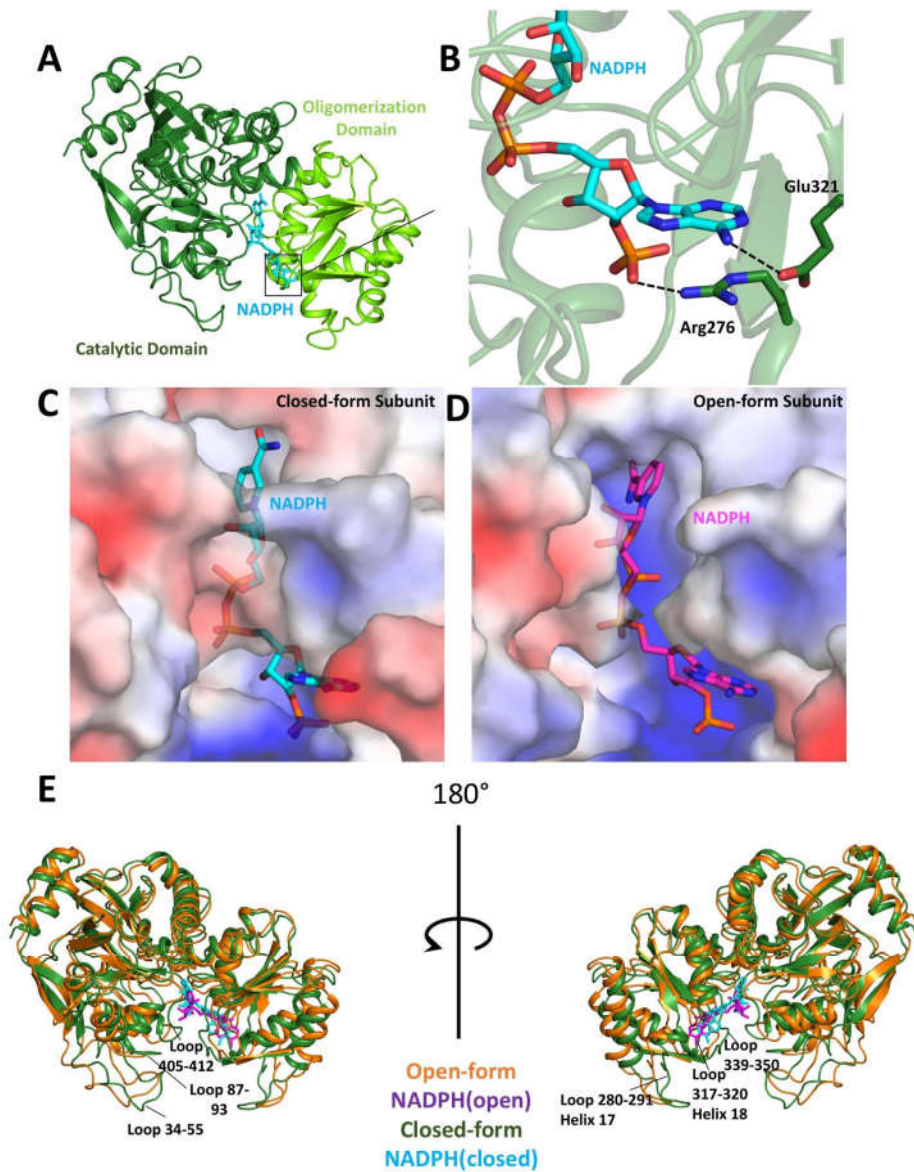
Supplementary Figure S5: A.) The NADPH spans from the adenine ring which is stabilized by the oligomerization domain to the catalytic domain, where the nicotinamide group is located. B.) The conserved stabilization of NADPH between the open and closed-form subunits is from Arg276 and Glu321. C.) The electrostatic surface of the co-factor binding pocket in the closed-form subunit. D.) The electrostatic surface of the co-factor binding pocket in the open-form subunit. E.) The conformational disparity between the open and closed-form subunits from two view angles, rotated 180°.
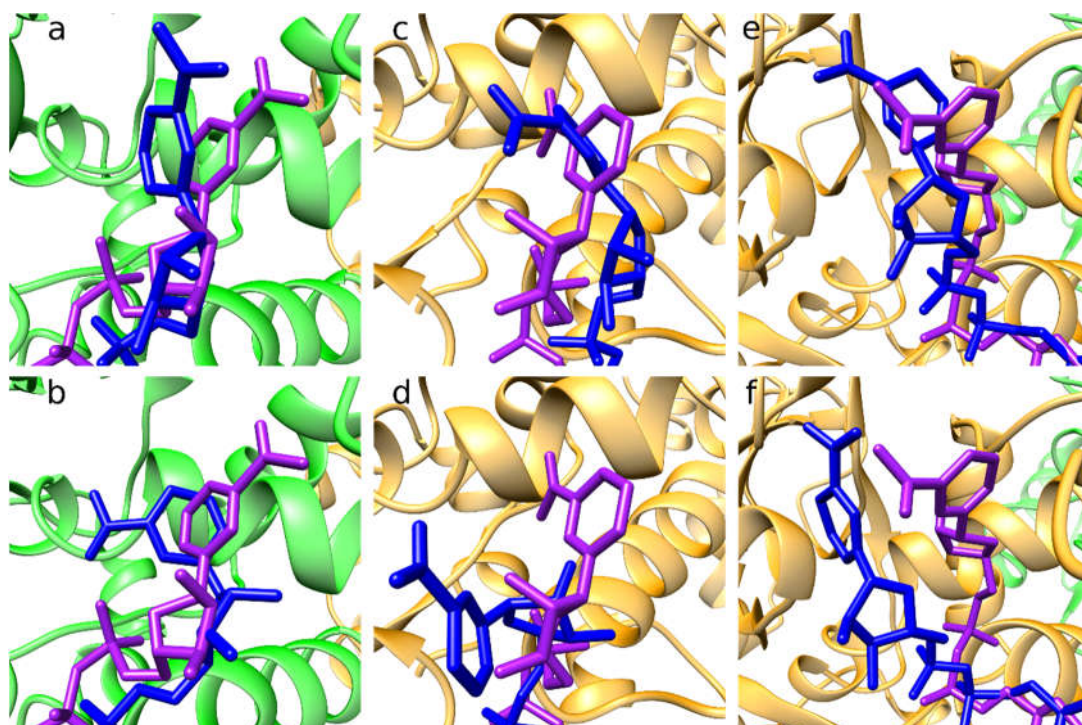
Supplementary Figure S6: Comparison between snapshots of molecular dynamics simulations (blue) and crystal structures (purple) of the binary complex in the closed and open subunits. For the closed subunits (a, b) the two predominant conformations sampled by NADPH fit very close to the crystal one. For the open subunits, the larger binding pocket allows the sampling of a wider range of conformations (d,f) including states similar to the one in the X-Ray structure (c,e).
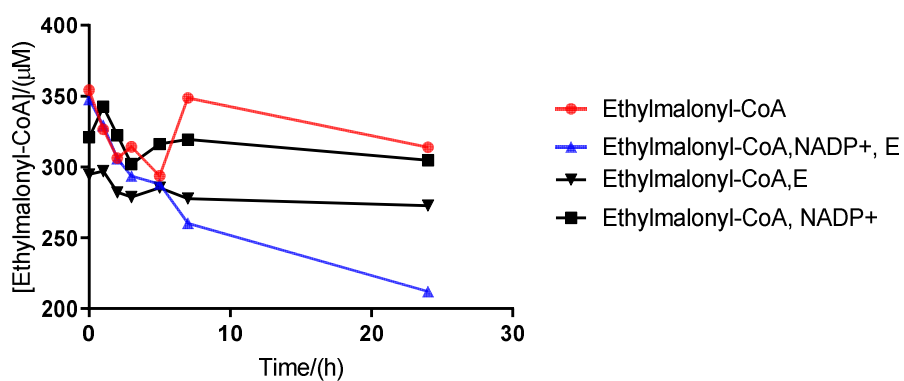
Supplementary Figure S7: Stability of ethylmalonyl-CoA stability measured over time at 19°C. The trace represented by blue triangles is the only condition in which ethylmalonyl-CoA is depleted over time. The enzyme catalyzed decarboxylation of ethylmalonyl-CoA to butyryl-CoA and CO_2 is the cause for the observed disappearance of ethylmalonyl-CoA. "E" corresponds to 1 µM *K. setae* ECR wt in the reaction mixture.

a

50-100

10-30

200-400

Estimated B factors of Crotonyl-CoA modeled in open subunit

b

Crotonyl-CoA modeled in open subunit

c

Crotonyl-CoA in closed subunit
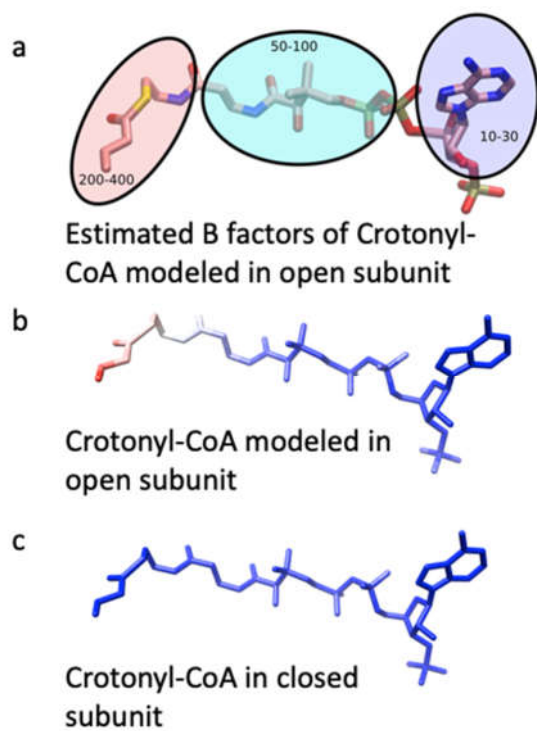
Supplementary Figure S8: QM/MM simulations of the substrate crotonyl-CoA and NADPH at the active sites. **a**. The differential flexibility of the Crotonyl-CoA substrate is reflected in the atomic fluctuations obtained from the dynamics. A. Range of B-factors (in $Å^2$) estimated from the squared atomic fluctuations (RMSF) and weighted by $(8/3)\pi^2$. **b**. Crotonyl-CoA molecule flexibility as modelled and simulated in the open subunit. **c**. Crotonyl-CoA molecule in the closed. The color reflects the estimated B-factor of each atom from low (blue) to high (red). (See also **Supplementary Videos 1&2**)
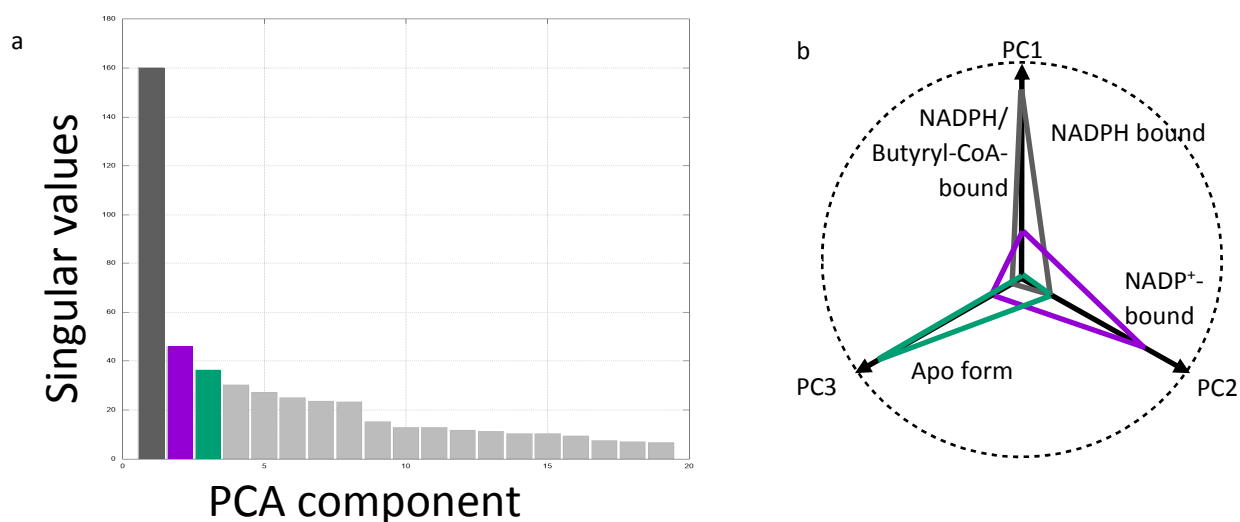
Supplementary Figure S9: Principal Component Analysis shows coupled motions of the catalytic domains to create alternating open and closed subunits. **a**. Singular values of the PCA components are plotted, showing the first component is the major contributor to the overall structural change. **b**. Contributions of PC1 to PC3 to the deviations of the NADPH-bound and NADPH/butyryl-CoA-bound structures from the average structure. See **Supplementary Videos 5** showing a morphing movie showing the structural changes between the apo form and ternary complex, and three most significant PCA motions.
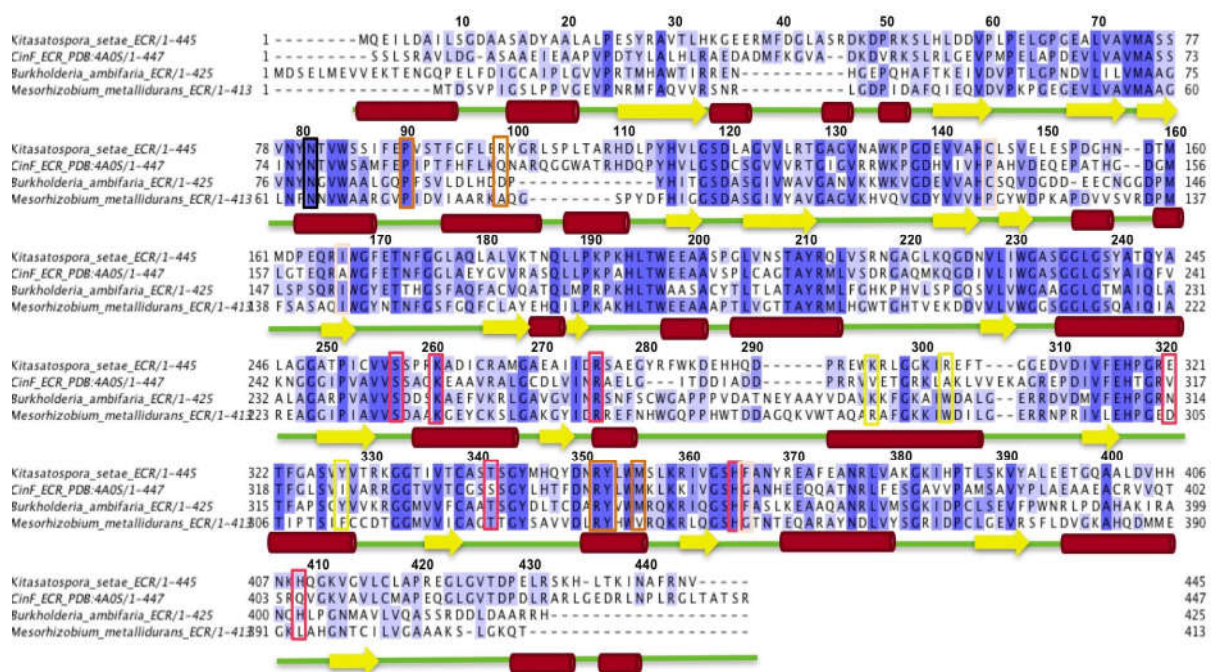
Supplementary Figure S10: The sequence alignment between two primary ECRs: *K. setae* & *M. metallidurans*, and two secondary ECRs: *B. ambifaria* & CinF from *S. collinus*. Two ECRs use crotonyl-CoA as their substrate: *K. setae* and *B. ambifaria*, while the others use octenoyl-CoA: *CinF* and *M. metallidurans*. The boxes in pink represent the residues which interact with NADP$^+$/NADPH. The residues in orange represent the residues which interact with the bound CoA. The residues in salmon represent the residues involved in substrate specificity. The residues in yellow represent the novel adenine ring CoA binding pocket discovered in the *K.setae* ECR. The residue in black (Asn81) represents the residue in question for stabilizing the CO_2 molecule for reductive carboxylation.

Supplementary Figure S11: Steady state parameters of *K. seate* ECR and its variants. The data are summarized in table 1 and table 2.
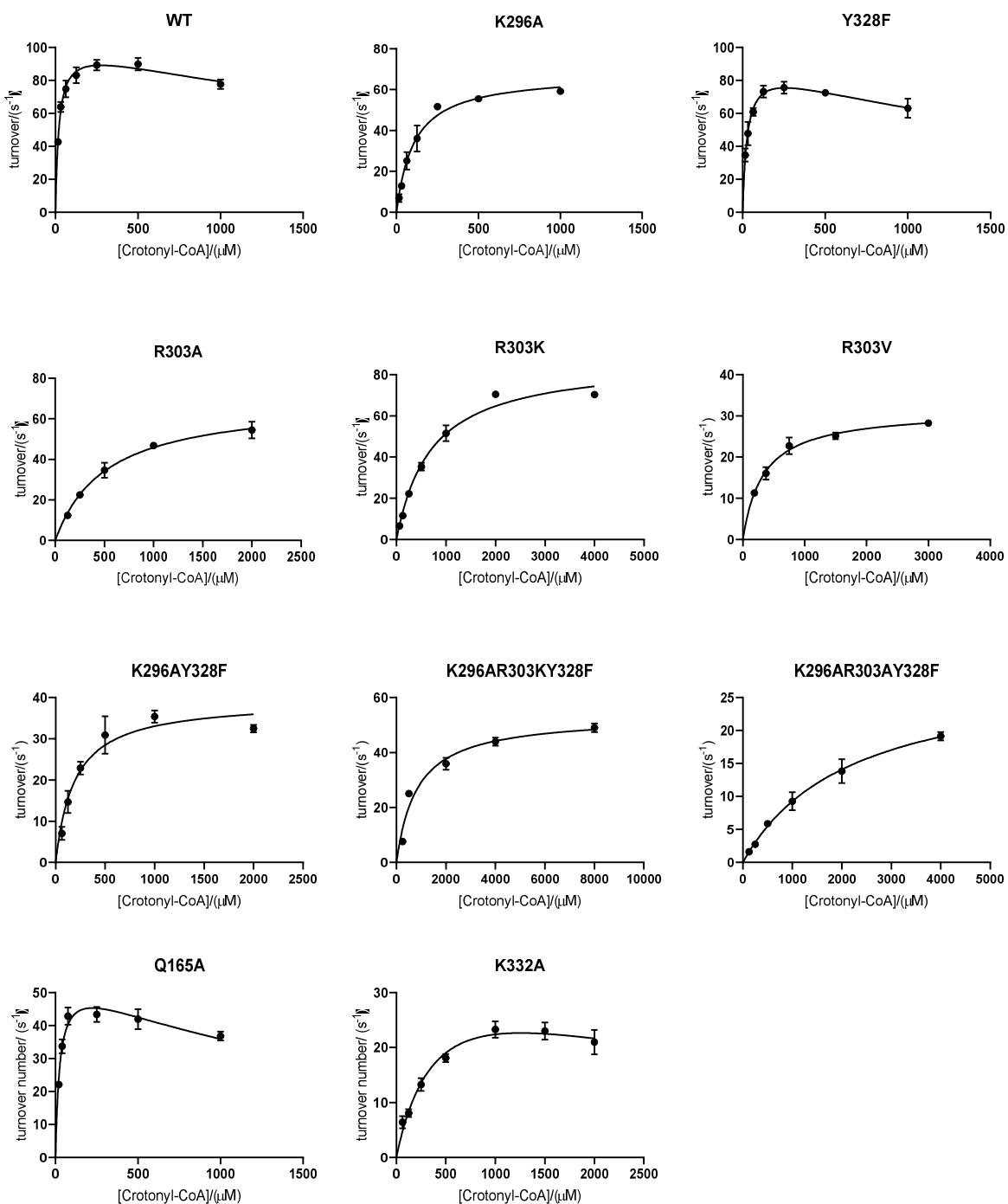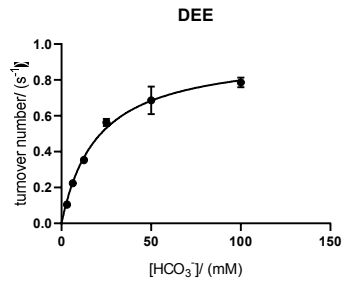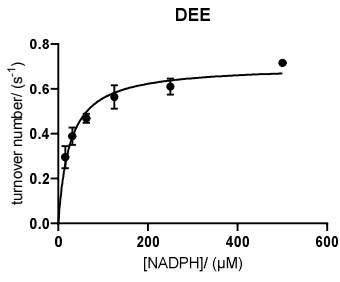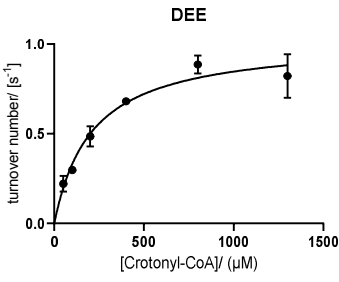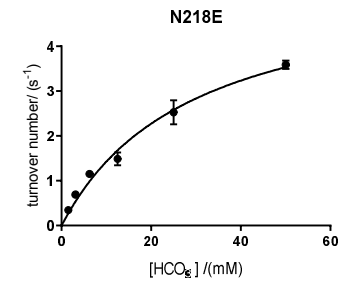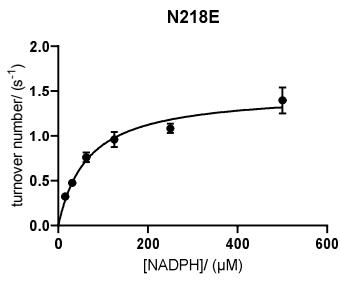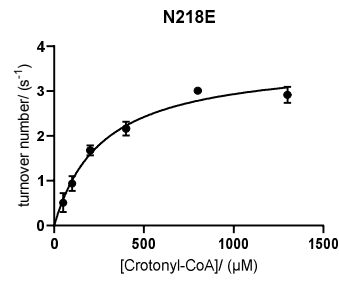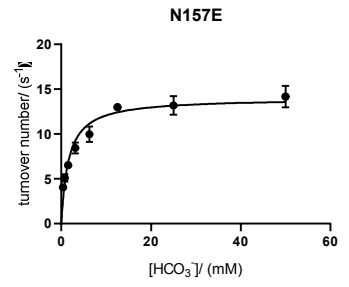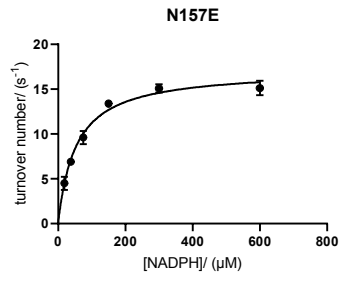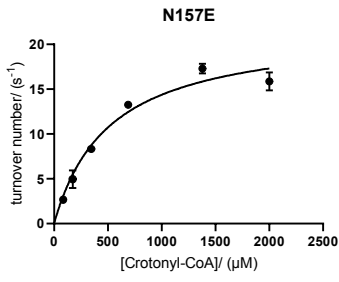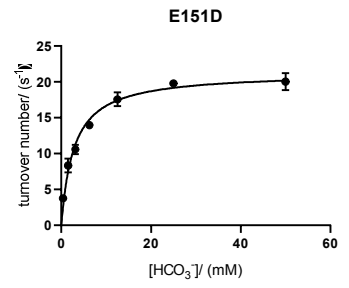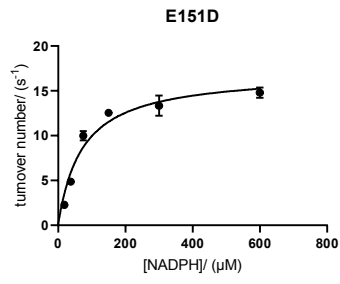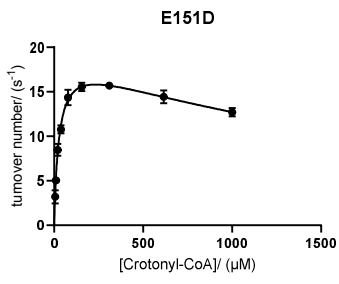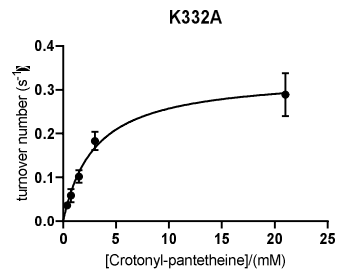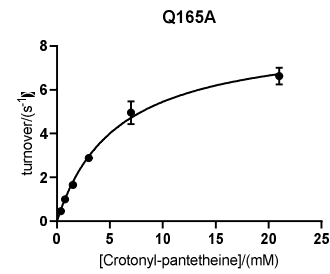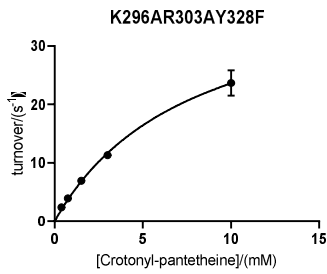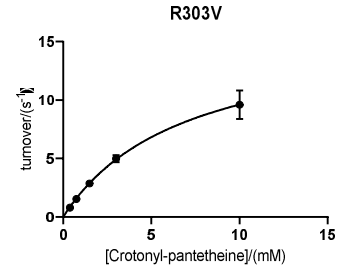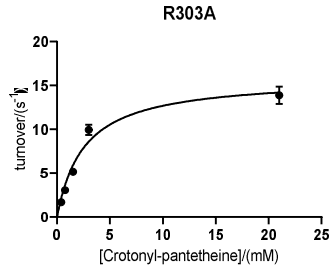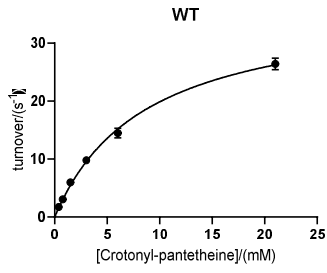
CHAPTER V

In vivo directed evolution of ECRs

5.1. Abstract

Enoyl-CoA carboxylases/reductases (ECRs) are the fastest carboxylases known to date and represent a good model enzyme for the understanding of carboxylation chemistry. ECRs are found in primary metabolism (primary ECRs) operating in the ethylmalonyl-CoA pathway (EMCP) and in secondary metabolism (secondary ECRs) where they generate precursors for natural products. Previous studies established the catalytic parameters and the structural basis for substrate specificity, which are distinctive for the two subfamilies. Here we developed an *in vivo* selection method to convert a secondary ECR from *Burkholderia ambifaria*, to an enzyme with primary ECR kinetic parameters. The turnover number of the evolved ECR (2.evo D3) increased 4-fold compared to the wildtype (WT) was able to complement growth *Methylobacterium extorquens* strain lacking the native ECR by 1.24-fold increase in growth rate. The accumulated mutations during three directed evolution cycles, are located distant form the active site and represent previously unknown structural hotspots in ECRs. Our work validates the hypothesis that strong selective pressure in primary metabolic pathways acts as a driving force to shape the kinetic parameters of enzymes.

5.2. Introduction

Carbon fixation is the process by which enzymes known as carboxylases utilize inorganic carbon, in the form of gaseous CO_2, and incorporate it into biomass. The enzyme Ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) found in plants and bacteria is responsible for fixing 400 Gt of carbon per year[1]. Nature's CO_2-fixing catalysts are of great interest because they possess the ability of utilizing atmospheric CO_2 under mild conditions. Knowledge of the underlying catalytic principles of carbon fixation is essential for the development of efficient carbon capture strategies to reduce the amount of atmospheric CO_2.

Compared to RuBisCO, reductive carboxylases from the enoyl-CoA carboxylase/reductase (ECR) family are faster and specific for CO_2[2]. The ECR family can be divided into two subfamilies based on the physiological role of the enzyme. ECRs operate in primary metabolism (primary ECRs), namely in the ethylmalonyl-CoA pathway (EMCP) which is used for the assimilation of C_2 units in many proteo- and actinobacteria[3]. In this pathway, they catalyze the NADPH-

dependent reductive carboxylation of crotonyl-CoA to (2*S*)-ethylmalonyl-CoA. These ECRs are limited to their physiological substrate crotonyl-CoA and are thus called crotonyl-CoA carboxylases/reductases (Ccrs). The *ccr* gene is usually very close to the ethylmalonyl-CoA mutase (*ecm*) gene, which is responsible for the successive reaction in the EMCP[4]. In secondary metabolism, ECRs (secondary ECRs) catalyze the reductive carboxylation of a large variety of enoyl-CoAs. The alkylmalonyl-thioesters products represent extender units for the biosynthesis of natural products[5]. The *ecr* gene is often directly associated with the polyketide synthase (PKSs) or nonribosomal peptide synthetases (NRPSs) gene clusters[6-8].

The high turnover rates, the specificity for CO_2 and their occurrence in different metabolic contexts make ECRs interesting case study for the evolution within an enzyme family. A kinetic study on the ECR enzyme family revealed that members of each subfamily display a distinct set of kinetic properties[9]. Primary ECRs are fast and specific for their substrates whereas secondary ECRs are slower and promiscuous. The findings could be rationalized based on the physiological role of the ECRs. Primary ECRs need to sustain a high flux of metabolites through the EMCP and therefore need to be fast and specific. Secondary ECRs are slower because the production of secondary metabolites, such as antibiotics, is not a frequent necessity and therefore there is no selective pressure to achieve high turnover rates. It has therefore been hypothesized that the evolutionary pressure exerted on enzymes shapes their kinetic parameters in a way to optimize their function within metabolic pathways[10]. In this regard, the previously described accuracy-rate tradeoff property that enzymes display[11] was confirmed for the ECR enzyme family *in vitro* [12]. In the light of this, we sought of developing an *in vivo* directed evolution system to improve kinetic parameters of a slow ECR variant.

Using an *in vivo* selection method in the proteobacterium *Methylobacterium extorquens* we show that a secondary ECR from *Burkholderia ambifaria* (here termed D3) can be evolved into an enzyme with improved kinetic parameters similar to primary ECRs. A gene library of D3 was expressed in the *M. extorquens* Δ*ccr* strain and selected for growth on methanol. The evolved triple mutant (2.evo D3) has a 4-fold increased k_{cat} compared to the wild type (WT) enzyme and is able to complement growth on methanol in the *M. extorquens* Δ*ccr* strain. The mutations are located far from the active site and occurred on solvent exposed patches of the protein and at the interface between monomers. Characterization of 2.evo D3 revealed that the thermal stability of the enzyme was unaffected. Steady state analysis of the carboxylation

half-reaction revealed an increased k_{cat} value compared to the WT. This study shows that the increased selection pressure exerted on ECRs in the EMC pathway represents a strong driving force towards increased catalytic rates.

5.3. Results

Gene library generation and selection of evolved variants

The directed evolution approach consisted of iterative steps of secondary *ecr* library generation followed by screening of evolved variants after growth on methanol. *M. extorquens* is a facultative methylotroph, which is able to grow on methanol [13]. The library was then transformed into the *M.extorquens* he sole source of energy and carbon. We therefore used the *M.extorquens* Δ*ccr* strain, which is dependent on a functional variant of Ccr in the EMCP for growth on methanol. The *ecr* gene library generated by epPCR and the product was cloned into the pTE103 vector Δ*ccr* strain and transformants were selected for growth on plates supplemented with methanol. Colonies typically appeared after 7 days of incubation at 30 °C. Colonies of 1-2 mm diameter were picked, plasmids isolated and sequenced (SI Fig. S1).

Steady state analysis reveals improved turnover number for evolved D3 variants

For each of the three rounds of evolution the obtained Ccrs were expressed, purified and kinetically characterized. The variant with the highest k_{cat} value each round was chosen for the following round of evolution. Kinetic parameters are summarized in table 1. D3 WT displays a k_{cat} of 18 ± 1 s^{-1} and apparent K_M values of 7 ± 1 μM (crotonyl-CoA), 64 ± 5 μM (NADPH) and 20 ± 4 μM (CO_2), respectively. Substrate inhibition for crotonyl-CoA was observed at a K_i of 741 ± 108 μM. The first round of evolution yielded the K289EL381M double mutant, which displayed a k_{cat} of 24 ± 2 s^{-1}. The K_M and K_i for crotonyl-CoA did not increase significantly, whereas the K_M for NADPH and CO_2 more than doubled. This double mutant was employed as the template for the successive round of epPCR. The fastest D3 variant obtained in this round accumulated the R325P mutation, and had a turnover number of 78 ± 8 s^{-1}. The K_M for CO_2 in the second evolution was comparable to the WT and the K_M for NADPH was similar to 1.evo D3. In an effort to increase the already 5-fold improved k_{cat} of 2.evo D3 we performed another round of directed evolution with this variant. The obtained variants either had the the same amino acid sequence as 2.evo D3 or did not show improvement in k_{cat}. The fastest variant isolated, 3.evo D3 had a k_{cat} of 60 ± 3 s^{-1}. The K_M for crotonyl-CoA, NADPH and CO_2 were not affected and the increased K_i of 2803 ± 749 μM is likely the reason for the decrease in k_{cat} compared to 2.evo D3.

Table 1: Apparent steady state parameters for crotonyl-CoA, NADPH and CO_2 for D3 WT and evolved variants.

D3 variant	Crotonyl-CoA			NADPH	CO_2
	k_{cat} (s^{-1})	K_M (µM)	K_i (µM)	K_M (µM)	K_M (µM)
WT	18 ± 1	7 ± 1	741 ± 108	64 ± 5	20 ± 4
1.evo D3	24± 2	10 ± 1	828 ± 193	147 ± 28	42 ± 7
2.evo D3	78 ± 8	87 ± 17	619 ± 143	150 ± 27	19 ± 3
3.evo D3	60 ± 3	55 ± 6	2803 ± 749	95 ± 10	40 ± 7

SI Appendix, Fig. S2 shows the Michaelis-Menten graphs of the original data [a] Calculated from bicarbonate concentration at pH = 8.

D3 variants complement growth on methanol

To confirm that the improved D3 variants could sustain the metabolic flux through the EMCP under our selection condtions we tested the evolved D3 variants for their ability to complement growth on methanol in *M. extorquens* AM1 *Δccr*. To determine the growth rate and doubling time the *M. extorquens* AM1 *Δccr* strain was transformed with the pTE103 vector containing either D3 WT, the evolved D3 variants or the native *M. extorquens* AM1 Ccr. The strain expressing D3 WT had a doubling time of 6.67 h and a growth rate of 0.1 h^{-1} whereas *M. extorquens* AM1 containing the native Ccr had a growth rate of 0.13 h^{-1} and doubling time of 5.43 h (Table 2). Complementation with evolved D3 variants improved growth on methanol in comparison with D3 WT. Growth rate for the strain containing the 1.evo D3 and the 3.evo D3 variant was 0.12 h^{-1}, thus increased 20 % in comparison with the WT, while the doubling time decreased by 13 %. With 30 % increase in growth rate and a doubling time of 5.23 h the strain with 2.evo D3 complemented growth on methanol with similar values for doubling time and growth rate as measured for *M. extorquens* AM1 Ccr (Table 2).

Table 2: *M. extorquens* AM1 *Δccr* growth rate μ_{max} and doubling time complemented with D3 variants during exponential growth.

	D3 WT	1.evo D3	2.evo D3	3.evo D3	*M. ext.* AM1 Ccr
μ_{max} (h^{-1})	0.10	0.12	0.13	0.12	0.13
doubling time (h)	6.67	5.77	5.23	5.87	5.43

Growth curves for *M. extorquens* AM1 *Δccr* with D3 WT or 2.evo D3 reached a maximal OD_{600} of approximately 6 in the stationary phase, but the onset of the exponential phase was delayed in the *M. extorquens* AM1 *Δccr* complemented with D3 WT in comparison to 2.evo D3 (Figure 1). The delayed onset of the exponential phase was also observed for

124

complementation with 1.evo D3, 3.evo D3 and the native Ccr of *M. extorquens* AM1. The maximal OD_{600} reached by the latter cultures varied between the biological replicates (Figure 1) due to the natural variance typically occurring in biological replicates. In conclusion, 2.evo D3 expressed *M. extorquens* AM1 *Δccr* was able to complement growth on methanol better than 1.evo D3 and 3.evo D3 and displayed similar growth parameters to *M. extorquens* complemented with its native Ccr.
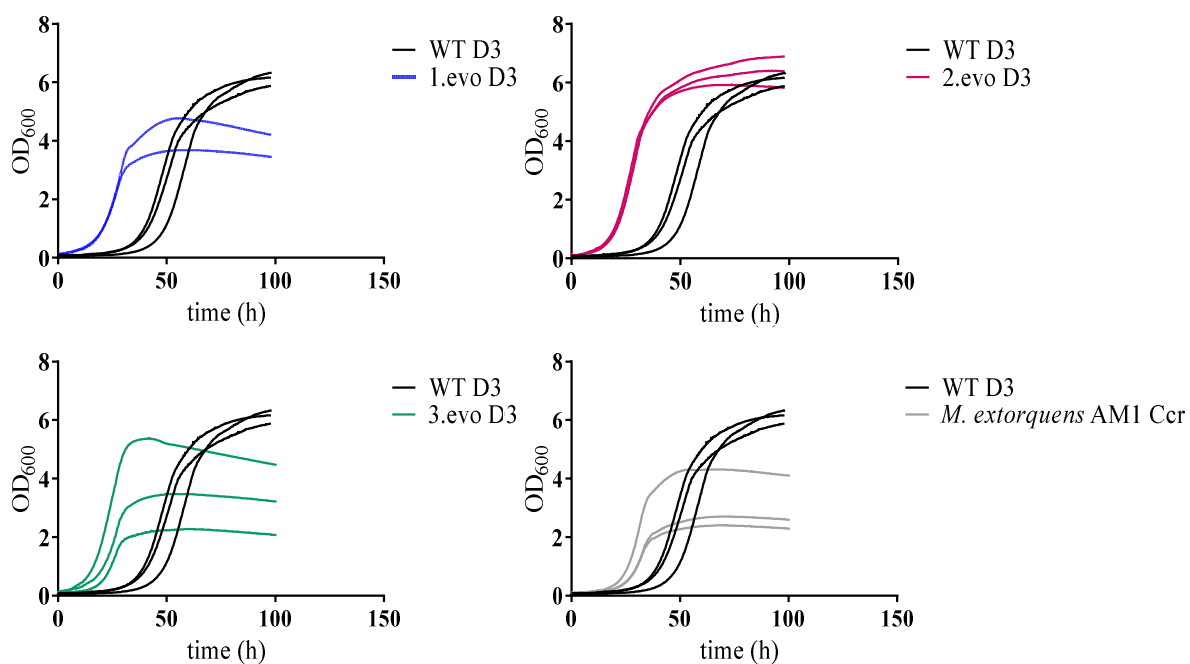

Figure 1: Growth curves for *M. extorquens* AM1 *Δccr* complemented with D3 WT, its evolved variants or native Ccr.

Characterization of 2.evo D3

The second round of evolution yielded the D3 K289EL381MR325P triple mutant which displayed an increased turnover number compared to D3 WT. Position of these mutations in the crystal structure of D3 (unpublished) are shown in figure 2. The K289E mutation is found in an α-helix, while mutations L381M and R325P are found in a loop (Figure 2). All mutations occurred on solvent exposed parts of the protein and distant from the active site. In the functional tetrameric unit of D3, R325P of chain A is located in close proximity to chain D. While Arg325 is 3.7 Å away from Leu147 of chain D, Pro 325, in the triple mutant, is 3.6 Å away from L147 (Figure 2 B and C). Thus, the R325P mutation changes the distances to the adjacent chain D and seems to influence the chains interaction of D3 in the tetrameric organization.
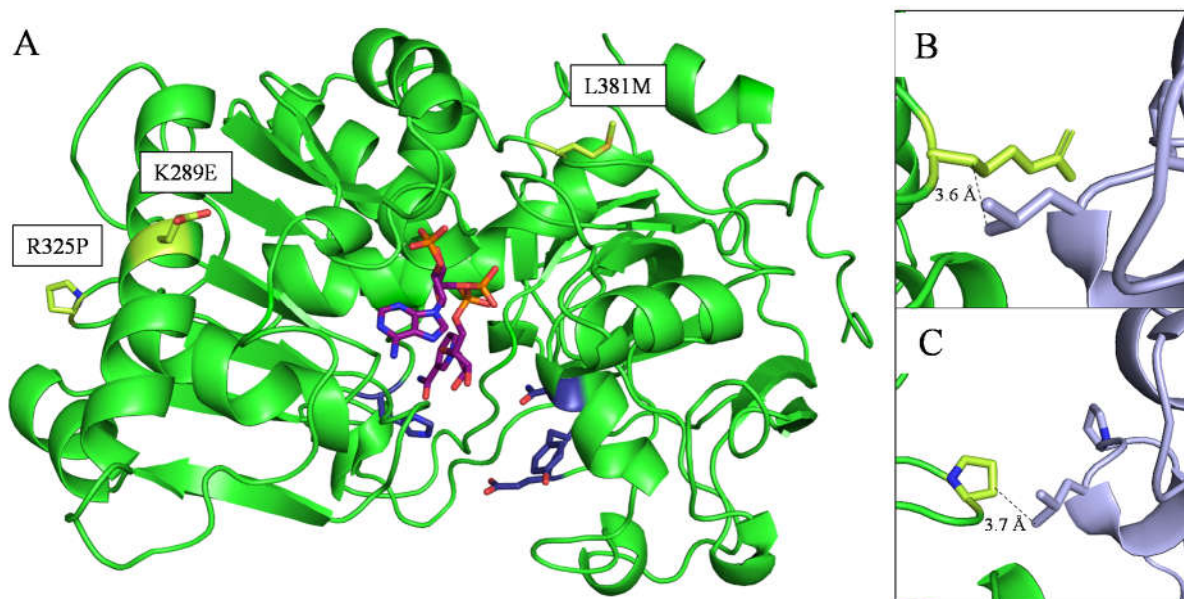
Figure 2: Structure of D3 WT with labeled mutations introduced by directed evolution. **A** ECR monomer with active site residues (deep blue) and NADP$^+$ (magenta). **B,C** Close up of residues R325 of D3 WT and P325 of 2.evo D3 respectively with indicated distance to L147 of chain D (grey).

To test the contribution of each mutation to the observed increased turnover of 2.evo D3, the possible combinations of single and double mutants of D3 were generated by site directed mutagenesis. We determined the kinetic parameters and performed product analysis to determine the ratio of carboxylated (ethylmalonyl-CoA) to reduced (butyryl-CoA) product (C/H). Single mutants D3 K289E, D3 L381M and D3 R325P all displayed increased k_{cat} values compared to D3 WT (Table 3). The R325P mutation had the biggest impact on k_{cat} among single mutants, which increased to 35 ± 3 s^{-1} compared to D3 WT. Double mutants showed both beneficial and deleterious effects on catalysis depending on the combination of mutations. Double mutant D3 K289ER381P had a turnover number similar to the WT, thus decreased turnover number in comparison to the single mutants. The highest impact on turnover among double mutants was observed in D3 R325PL381M. With a 3-fold improved turnover number of 52 ± 2 s^{-1}, compared to D3 WT, this mutant is the closest to the 2.evo D3 triple mutant in terms of k_{cat} (Table 1).

The K_M values for the substrates differed between mutants. The apparent K_M for crotonyl-CoA increased along with the turnover number leaving the catalytic efficiency similar. The K_M of NADPH was not changed significantly in all mutants compared to the WT except for the L381M mutant that displayed a K_M value of 186 ± 19 µM. Double and triple mutants containing mutation L381M increased the NADPH K_M about 2-fold. K_M values for CO_2 where mostly unaffected, only double mutant R325L381M showed a 5-fold increase compared to the WT.

The data indicates that all three mutations had additive effects on the turnover number of 2.evo D3 variant.

Table 3: Apparent steady state parameters for crotonyl-CoA, NADPH and CO_2, k_{cat}/K_M and carboxylation to reduction product ratio for D3 WT and its evolved variants.

D3 variant	Crotonyl-CoA			NADPH	CO_2		
	k_{cat} (s^{-1})	K_M (µM)	K_i (µM)	K_M (µM)	[a]K_M (µM)	k_{cat}/K_M (s^{-1} µM^{-1})[b]	C/H (%)
WT	18 ± 1	7 ± 1	741 ± 108	64 ± 5	20 ± 4	1.36	100
K289E	23 ± 1	13.4 ± 1.8	927 ± 186	65 ± 7	21 ± 2	0.625	100
L381M	29 ± 2	17.4 ± 2.5	1805 ± 597	186 ± 19	21 ± 2	0.618	100
R325P	35 ± 3	32 ± 6	1460 ± 613	55 ± 6	15 ± 2	0.618	94
K289EL381M	24 ± 2	10.4 ± 1	828 ± 193	147 ± 28	42 ± 7	1.174	100
K289ER325P	18 ± 1	42 ± 5	-	57 ± 6	22 ± 2	0.328	100
R325PL381M	52 ± 2	81.7 ± 9	-	88 ± 10	100 ± 16	0.334	100
K289EL381MR325P	64 ± 5	43 ± 8	694 ± 146	113 ± 13	49 ± 3	1.025	100

SI Appendix, Fig. S2 shows the Michaelis-Menten graphs of the original data [a] Calculated from bicarbonate concentration at pH = 8. [b] Calculated with respect to crotonyl-CoA.

All mutants where able to carboxylate more than 94 % of the starting crotonyl-CoA to ethylmalonyl-CoA indicating that the observed changes in kinetic parameters did not affect the carboxylation efficiency.

To determine if the increased turnover number was due to improved protein stability, melting curves for D3 WT and the evolved variants were measured by following changes in tryptophan fluorescence upon thermal denaturation and protein unfolding (Figure 4). D3 WT, 1.evo D3 and 2.evo D3 showed similar progress curves and inflection points, indicating that the unfolding occurred at the same temperature. The differences in peak height were due to slightly differing protein concentrations in each measurement.
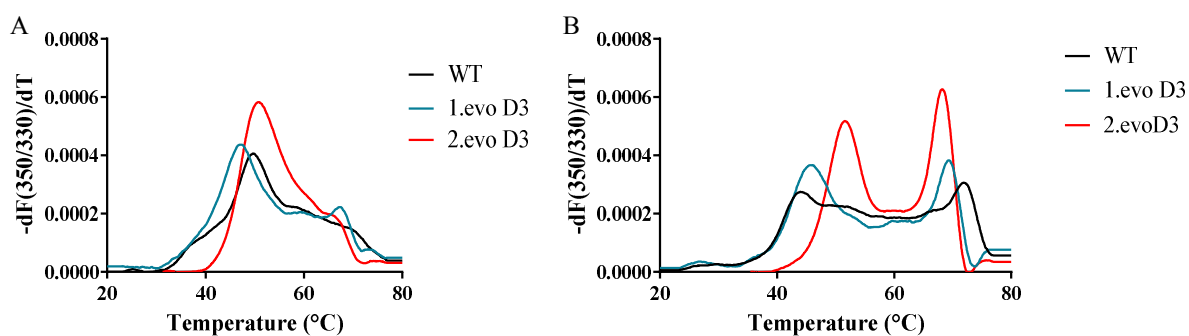
Figure 4: Melting curves with derivation of the tryptophan fluorescence against the temperature for D3 WT and variants. Samples with 0.1 µg/µL purified protein were measured in 100 mM K_2HPO_4 buffer pH = 8.

We then set out to determine the kinetic parameters of the carboxylation half reaction by starting the assay using the C2-adduct, a reaction intermediate previously described to accumulate in ECRs [14]. Steady state analysis revealed that the carboxylation half reaction is faster in 2.evo D3 than the WT (Table 4) and that this is the cause for the improved overall turnover rate.

Table 4: Steady-state parameters of the carboxylation half reaction.

Enzyme	k_{cat} (s^{-1})	$K_{M, C2}$ (µM)
D3 WT	2.8 ± 0.1	26 ± 3
2.evo D3	8.6 ± 0.4	190 ± 18

5.4. Discussion

With an *in vivo* directed evolution method, we were able to evolve a secondary ECR into a primary one. The obtained triple mutant 2.evo D3 displayed a 4-fold increased k_{cat} compared to D3 WT and was able to complement growth on methanol when expressed in *M. extorquens* AM1 Δ*ccr*. The accumulated mutations during directed evolution rounds occurred on within loops (R325P, L318M), on a solvent exposed α-helix (K298E) and were located distant form the active site. During directed evolution cycles mutations often accumulate far away from the enzymatic active site[15, 16]. These mutations may influence the protein solubility and/or the kinetic parameters of the enzyme. In the case 2.evo D3 the kinetic parameters were improved and we showed that this was due to an increased rate of the carboxylation half-reaction rather than the increased protein stability. The introduced mutation might have a beneficial impact on the overall flexibility of the protein, which might lead to a faster release rate of the product. The R325P mutation specifically, is located at the interface between two monomers and might influence the interactions between them and regulate the flexibility of the protein during catalysis.

Previous studies revealed that the difference in activity between primary an secondary ECRs is due to inherent accuracy rate tradeoff observed in enzymes. This was observed also in the ECR subfamily and was ascribed to the residues, which determine the substrate promiscuity in ECR. Not surprisingly, our evolved variants did not mutate residues in the active site because the specificity for crotonyl-CoA is essential to operate in the EMCP. Therefore, in contrast to what was previously reported[12], our case shows how residues not involved with substrate promiscuity are also important structural determinants for ECR catalysis. As these residues are well conserved throughout the ECR family they might represent structural hotspots that were not modified during evolutionary divergence of the family.

Characterization of 2.evo D3 did not improve its stability. We showed that an increased rate in the carboxylation half-reaction was the reason for the observed overall improvement in turnover. Over the rounds of directed evolution, the increase in turnover number was accompanied by an increase in K_M, which implies that the catalytic efficiency of the enzyme was not affected. Our system, in the current state, was able to improve only k_{cat} as we showed

in the steady state analysis. The K_M value could not be improved likely because the substrate and cofactor amounts were always saturating in the intracellular environment.

We showed that the selective pressure exerted on an ECR operating in the EMCP is the driving force for the evolution towards an improved turnover rate.

5.5. Materials and methods

Cloning and Mutagenesis

The error prone PCR (epPCR) performed with the GeneMorph II Random Mutagenesis Kit (Agilent Technologies, USA). The *ECR* genes were amplified with Mutazyme II polymerase by using Gibson primers G_pTE103/100_D3 fwd and rev (Table 2). The epPCR was performed with 20 ng and 100 ng template DNA and reaction conditions according to table 1.

Table 1: Reaction conditions for epPCR.

Step	Temperature [°C]	Time [min]	
Initial denaturation	95	2.0	
Denaturation	95	0.5	
Annealing	55	0.5	30 cycles
Elongation	72	2.0	
Final elongation	72	10.0	

Enzyme variants were generated with the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla, USA) using primer pairs listed (table 2).

Table 1: Primers used for the Gibson assembly and Quikchange mutagenesis. .

Primer	Sequence (5'-3')
G_pTE103_D3_fwd	AGATCTTGACTAGTCCTGCAGGTACGTTTAACTTTAAGAAGGAGATATAC
G_pTE103_D3_rev	AAACGACGGCCAGTGAATTAGGTACAACTCAGCTTCCTTTCGG
G_pTE100_D3_fwd	TCACATGGAATTCTGTACATGTTTAACTTTAAGAAGGAGATATAC
G_pTE100_D3_rev	ACGGCCAGTGAATTAGGTACCAACTCAGCTTCCTTTCG
pTE951_D3_R325P_fwd	CTACGTTGTGAAACCCGGTGGTATGGTTG
pTE951_D3_R325P_rev	CAACCATACCACCGGGTTTCACAACGTAG
pTE951_D3_K289E_fwd	CTTATGTTGACGCAGTAAAAGAATTCGGTAAAGCTATCTGGG
pTE951_D3_K289E_rev	CCCAGATAGCTTTACCGAATTCTTTTACTGCGTCAACATAAG
pTE951_D3_L381M_fwd	CAAAATCGACCCGTGCATGTCCGAAGTATTCC
pTE951_D3_L381M_rev	GGAATACTTCGGACATGCACGGGTCGATTTTG

Restriction and assembly

DNA restriction was performed by incubation of 2 µg vector DNA with appropriate restriction enzymes (NEB, USA; Thermo Scientific, USA) and buffer (FastDigest; Thermo Scientific, USA) at 37°C. The vector and insert DNA was applied in a 3:1 molar ratio to the 2x Gibson Assembly® Master Mix (NEB, USA) and was assembled in two steps. The first proceeds at 50 °C for 30 min and the second for 10 min at 70 °C.

Selection of improved D3 variants

Selection was performed by transforming the ECR library into the *M. extorquens* Δ*ccr* strain[17] and plating it on Minimal media plates supplied with methanol. Plates were incubated at 30°C and checked daily for colonies.

Determination of growth curves

Bacterial growth over the course of time was monitored using TECAN Infinite® 200 PRO plate reader systems (Tecan Trading AG, Switzerland) with TC-Plate 96 Well, Standard, F (Sarstedt, Germany). Every well contained 180 µL culture with starting OD_{600} of 0.1. The determination of OD_{600} in 10 min intervals was preceded by 530 s shaking. The measurement was continued until the cultures reached the stationary phase at OD_{600} 4-6. The analysis and fitting of growth data was done using the Prism 7 software (GraphPad Software, USA) and the growth equation $Y=Y0*exp(k*X)$.

Protein expression and purification

His-tagged proteins were expressed in *E. coli* BL21 DE3. The TB media complemented with salt buffer and antibiotics was inoculated with overnight grown cultures. After incubation at 37 °C to OD_{600} 0.5-0.8 cells were cooled for 20 min at 20 °C. Protein expression was induced by adding 500 µM IPTG (β-Isopropyl-thiogalactopyranoside) and occurred overnight at 110 rpm at 22 °C. During the following procedures cells were kept at 4°C. Cells were harvested for 20 min at 4000 g at 4°C. 2 mL lysis buffer (buffer A: 500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1 M L-proline) per gram of pellet was used to resuspend the cells. 10 mg/mL DNAse I (Roche Diagnostics, Switzerland) and 50 mM $MgCl_2$ were added to the suspension. Cells were lysed using a LM10 Microfluidizer™ at 16000 Psi or sonication for 2 times with 1 min resting time at 4 °C in between. The lysate was clarified at 45000 g for 45 min at 4 °C and the supernatant subsequently filtered with a 45 µm filter. Purification of the His-tagged protein was performed on a buffer A- equilibrated 1 mL Ni-Sepharose Fast Flow Column (HiTrap™, FF, GE Life Science, USA) on an ÄKTA Start FPLC system (GE Life Science, USA). The filtered supernatant was applied onto the column with 1 mL/min. The purification progress was followed at 280 nm. After washing steps with buffer A containing first 0 % and afterwards 12 % elution buffer B (500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 500 mM imidazole, 1 M L-proline) protein was eluted in 1 mL fractions by applying 100 % buffer B. Fractions containing the protein were desalted into 125 mM NaCl, 12.5 mM Tris pH =7.5, 1 M L-proline. L-proline was added to the purification buffers to increase the solubility and yield of the proteins[18]. The protein was concentrated

132

using Amicon® Ultra 4 mL Filters (Merck, Germany), snap-frozen in N_2 (l) and stored at -80 °C or stored at -20 °C, after addition of 50 % glycerol.

Chemicals

Crotonic Anhydride and Carbonic anhydrase from bovine erythrocytes were purchased from Sigma Aldrich AG, Coenzyme A trilithium salt and DNAse I from Roche Diagnostics, NADPH Na_4 (98%) from Carl Roth GmbH. Solvents and salts were all analytical grade or better. Crotonyl-CoA was synthesized as previously reported[19]. The C2-ene adduct was synthesized as previously reported[14]

Enzyme assays

Enzyme assays were performed on an Agilent Cary 60 UV-Vis (Agilent Technology, USA) in quartz cuvettes (1 or 10 mm path length, Hellma Analytics, Germany) by following NADPH consumption at 340 nm ($\varepsilon_{NADPH, 340nm}$= 6.22 mM^{-1} cm^{-1}) or 365 nm ($\varepsilon_{NADPH, 365nm}$= 3.3 mM^{-1} cm^{-1}). Steady state parameters were determined by varying the concentration of one substrate while keeping the other constant at 10 times its K_M value. The enzyme assays contained 100 mM K_2HPO_4 buffer pH = 8 and 2 µg/mL carbonic anhydrase in 100 µL reaction volume and were started by addition of either crotonyl-CoA or enzyme after 1 min incubation at 30 °C. Apparent k_{cat} and K_M values were obtained by fitting the data to the Michaelis-Menten equitation

$$V_0 = V_{max} \left([S]/([S] + K_M)\right).$$

or

$$V_0 = (V_{max}[S])/(K_M+[S]\,((1+[S])/K_i))$$

if substrate inhibition was observed.

Analytical methods

Purification of substrates was performed with preparative reverse phase HPLC-MS (High-performance liquid chromatography-mass spectrometry) with Gemini® 10 µm C18 110 °A, LC Column 50 x 21.2 mm (Phenomenex, USA) in an Agilent 1260 Infinity LC HPLC system. The column was equilibrated with 95 % 25 mM NH_4HCO_2, pH = 8.1 and 5 % MeOH for 2 min at 25 mL/min. The MeOH concentration was raised from 5 % to 35 % over 15 min to elute the

substrate. In a final step, the column was washed with 95% MeOH for 2 min. The eluted compounds were collected and their mass confirmed by a coupled 6130 Quadrupole mass spectrometer. The fractions were pooled, frozen in N_2 (l), lyophilized and stored at -80 °C. UHPLC (Ultra-high-performance liquid chromatography) was used to analyze the products of the enzyme assays with ECR D3 variants. Samples were prepared by addition of 50 % formic acid to quench the completed reaction and precipitation of protein at 17000 g for 10 min. The supernatant was either undiluted or diluted 5 times into 5% methanol, 95 % 25 mM NH_4HCO_2 pH = 8.1 and analyzed using a Sonoma C18 column (100 x 2.1 mm, particle size 3 μm, 100 Å pore size, ES Industries). The compounds were separated with a MeOH gradient from 5 to 90 % over 9 min. Quantification was performed by peak integration at 260 nm. Elution times of compounds were: ethylmalonyl-CoA (2.5 min), crotonyl-CoA (4.4 min) and butyryl-CoA after (4.8 min).

Protein folding and stability

The thermal stability of enzymes was determined on a Prometheus NT.48 nanoDSF technology (NanoTemper Technologies, Germany). Protein concentrations of 1 μg/μL or 0.1 μg/μL in 1 M proline pH = 7.5 or 100 mM K_2HPO_4 buffer pH = 8 were applied in capillaries to the instrument. The samples were exposed to a temperature increase from 20 °C to 80 °C in 60 min, while measuring fluorescence at 330 and 350 nm. The melting temperature is calculated from the ratio of tryptophan emission shifts at the measured wavelength.

5.6 References

1.   Phillips R & Milo R, A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. U.S.A.* 106(51):21465-21471 **2009**.
2.   Erb TJ, Brecht V, Fuchs G, Muller M, & Alber BE, Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *Proc. Natl. Acad. Sci. U.S.A.* 106(22):8871-8876 **2009**.
3.   Erb TJ*, et al.*, Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *Proc. Natl. Acad. Sci. U.S.A.* 104(25):10631-10636 **2007**.
4.   Erb TJ, Retey J, Fuchs G, & Alber BE, Ethylmalonyl-CoA mutase from Rhodobacter sphaeroides defines a new subclade of coenzyme B12-dependent acyl-CoA mutases. *J. Biol. Chem.* 283(47):32283-32293 **2008**.
5.   Wilson MC & Moore BS, Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* 29(1):72-86 **2012**.
6.   Eustaquio AS*, et al.*, Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc. Natl. Acad. Sci. U.S.A.* 106(30):12295-12300 **2009**.
7.   Blazic M, Kosec G, Baebler S, Gruden K, & Petkovic H, Roles of the crotonyl-CoA carboxylase/reductase homologues in acetate assimilation and biosynthesis of immunosuppressant FK506 in Streptomyces tsukubaensis. *Microb Cell Fact* 14:164 **2015**.
8.   Bilyk O, Samborskyy M, & Leadlay PF, The biosynthetic pathway to ossamycin, a macrocyclic polyketide bearing a spiroacetal moiety. *PloS one* 14(4):e0215958 **2019**.
9.   Peter DM Substrate Promiscuity, Kinetics and Engineering of Enoyl-CoA Carboxylases/Reductases. Thesis (ETH Zürich) **2016**
10.   Bar-Even A*, et al.*, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50(21):4402-4410 **2011**.
11.   Tawfik DS, Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21:73-80 **2014**.
12.   Peter DM*, et al.*, Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. *Angew. Chem. Int. Edit.* 54(45):13457-13461 **2015**.
13.   von Borzyskowski LS, Remus-Emsermann M, Weishaupt R, Vorholt JA, & Erb TJ, A Set of Versatile Brick Vectors and Promoters for the Assembly, Expression, and Integration of Synthetic Operons in Methylobacterium extorquens AM1 and Other Alphaproteobacteria. *ACS Synth. Biol.* 4(4):430-443 **2015**.
14.   Rosenthal RG*, et al.*, Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* 10(1):50-55 **2014**.
15.   Prier CK, Zhang RK, Buller AR, Brinkmann-Chen S, & Arnold FH, Enantioselective, intermolecular benzylic C-H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* 9(7):629-634 **2017**.
16.   Hammer SC*, et al.*, Anti-Markovnikov alkene oxidation by metal-oxo-mediated enzyme catalysis. *Science* 358(6360):215-218 **2017**.

17. Okubo Y, Yang S, Chistoserdova L, & Lidstrom ME, Alternative route for glyoxylate consumption during growth on two-carbon compounds by Methylobacterium extorquens AM1. *J. Bacteriol.* 192(7):1813-1823 **2010**.

18. Samuel D, Kumar TK, Jayaraman G, Yang PW, & Yu C, Proline is a protein solubilizing solute. *Biochem. Mol. Biol. Int.* 41(2):235-242 **1997**.

19. Peter DM, Vogeli B, Cortina NS, & Erb TJ, A Chemo-Enzymatic Road Map to the Synthesis of CoA Esters. *Molecules* 21(4):517 **2016**.

5.7. Supplementary Materials

Supplementary Figure S1: Protein sequences of obtained D3 ECR variants after directed evolution cycles. Accumulated mutations are marked in red.

Protein sequences after first round of directed evolution

```
WT       1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.1     1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.2     1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.4     1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.7     1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.9     1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.10    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.11    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.12    1 MDSELMEVVEKNENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.13    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTMEIVDVP


WT      61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.1    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGTNVKKW
20.2    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.4    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.7    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.9    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.10   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.11   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.12   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.13   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW


WT     121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.1   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.2   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.4   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.7   121 KVGDEVVAHCSQVDGDDDECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.9   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.10  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.11  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.12  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.13  121 KVGDEVVVHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMLRPKH


WT     181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.1   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.2   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.4   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.7   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.9   181 LTWAASACYNLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.10  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.11  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.12  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.13  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA


WT     241 VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.1   241 VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.2   241 VVSDDSKAEFVKRLGAAGVINRSNFSCWGAPPPVDATNEYAAYVDALKKFGKAIWDALGE
```

137

```
20.4    241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.7    241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.9    241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.10   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.11   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.12   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.13   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE


WT      301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.1    301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.2    301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.4    301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.7    301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.9    301  RRDVDMVFEHPGRNTFAPSCYVVKRGGIVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.10   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.11   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMCQKRIQGSHFA
20.12   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.13   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA


WT      361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.1    361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.2    361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.4    361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.7    361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRTNQHLPGNMAVLVQASSRDDLD
20.9    361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHQPGNMAVLVQASSRDDLD
20.10   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.11   361  SLKEAAQANRLVMSGKIDPCLSEVFPWNCLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.12   361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.13   361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD


WT      421  AARRH
20.1    421  AARRH
20.2    421  AARRH
20.4    421  AARRH
20.7    421  AARRH
20.9    421  AARRH
20.10   421  AARRH
20.11   421  AARRH
20.12   421  AARRH
20.13   421  AARRH
```

Protein sequences after second round of directed evolution. Mutations marked in green are those present in template used for ePCR

```
WT       1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.1     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.2     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.3     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.4     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.5     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.6     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.8     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.9     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.10    1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
```

```
20.11    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRDNHGEPQHAFTKEIVDVP
20.12    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.13    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.14    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.15    1 MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP


WT      61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.1    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.2    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.3    61 TLGPNDVLILVMAAGVNYNGIWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.4    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.5    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.6    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.8    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.9    61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.10   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.11   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGAKVKKW
20.12   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.13   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.14   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.15   61 TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW


WT     121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.1   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.2   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.3   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.4   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTRGSFAQFACVQATQLMPRPKH
20.5   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.6   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.8   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACAQATQLMPRPKH
20.9   121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.10  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.11  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.12  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.13  121 KVGDEVVAHCSQVDGDDEECNGGDPILSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.14  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.15  121 KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH


WT     181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.1   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.2   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.3   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.4   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.5   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.6   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.8   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.9   181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.10  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.11  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.12  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.13  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.14  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.15  181 LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
```

```
WT     241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.1   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.2   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.3   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVEATNEYAAYVDAVKEFGKAIWDALGE
20.4   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.5   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.6   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.8   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.9   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGEAIWDALGE
20.10  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.11  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.12  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.13  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.14  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.15  241  VVSDDSKAEFLKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE


WT     301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.1   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.2   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.3   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.4   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.5   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.6   301  RRDVDMVFEHPGRNTFAPSCYTVKSGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.8   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.9   301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.10  301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.11  301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.12  301  RRDVDMVFEHPGRNTFAPSCYVVKPGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.13  301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.14  301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.15  301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA


WT     361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.1   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.2   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKFRANQHLPGNMAVLVQASSRDDLD
20.3   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.4   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.5   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.6   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.8   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNTAVLVQASSRDDLD
20.9   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNSLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.10  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.11  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.12  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.13  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.14  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.15  361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD


WT     421  AARRH
20.1   421  AARHH
20.2   421  AARRH
20.3   421  AARRH
20.4   421  AARRH
20.5   421  AARRH
```

```
20.6    421  AARRH
20.8    421  AARRH
20.9    421  AARRH
20.10   421  AARRH
20.11   421  AARRH
20.12   421  AARRH
20.13   421  AARRH
20.14   421  AARRH
20.15   421  AARRH
```

Protein sequences after third round of directed evolution. Mutations marked in green are those present in template used for ePCR.

```
WT       1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.3     1  MDSELMEVVEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.5     1  MDSELMEVTEKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.9     1  MDSELMEVVEKTENGQPELFNIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP
20.10    1  MDSELMEVVKKTENGQPELFDIGCAIPLGVVPRTMHAWTIRRENHGEPQHAFTKEIVDVP


WT      61  TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.3    61  TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.5    61  TLGPNDVLILVMAAGVNYNGIWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVEKW
20.9    61  TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW
20.10   61  TLGPNDVLILVMAAGVNYNGVWAALGQPFSVLDLHDDPYHITGSDASGIVWAVGANVKKW


WT     121  KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.3   121  KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.5   121  KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.9   121  KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH
20.10  121  KVGDEVVAHCSQVDGDDEECNGGDPMLSPSQRIWGYETTHGSFAQFACVQATQLMPRPKH


WT     181  LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.3   181  LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.5   181  LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA
20.9   181  LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMALQLAALAGARPVA
20.10  181  LTWAASACYTLTLATAYRMLFGHKPHVLSPGQSVLVWGAAGGLGTMAIQLAALAGARPVA


WT     241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKKFGKAIWDALGE
20.3   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.5   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.9   241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE
20.10  241  VVSDDSKAEFVKRLGAVGVINRSNFSCWGAPPPVDATNEYAAYVDAVKEFGKAIWDALGE


WT     301  RRDVDMVFEHPGRNTFAPSCYVVKRGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.3   301  RRDVDMVFEHPGRNTFAPSCYVVKPGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.5   301  RRDVDMVFEHPGRNTFAPSCYVVKPGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.9   301  RRDVDMVFEHPGRNTFAPSCYVVKPGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA
20.10  301  RRDVDMVFEHPGRNTFAPSCYVVKPGGMVVFCAATSGYDLTCDARYVWMRQKRIQGSHFA


WT     361  SLKEAAQANRLVMSGKIDPCLSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.3   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.5   361  SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSCDDLD
```

141

```
20.9    361 SLKEVAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD
20.10   361 SLKEAAQANRLVMSGKIDPCMSEVFPWNRLPDAHAKIRANQHLPGNMAVLVQASSRDDLD


WT      421 AARRH
20.3    421 AARRH
20.5    421 AARRH
20.9    421 AARRH
20.10   421 AARRH
```

Supplementary Figure S1: Steady state parameters of ECR D3 and its variants. All reactions contained 50 mM NaHCO$_3$ and 20 µg/ml of carbonic anhydrase and were performed in 100 mM K$_2$HPO$_4$ pH = 8 at 30 °C.
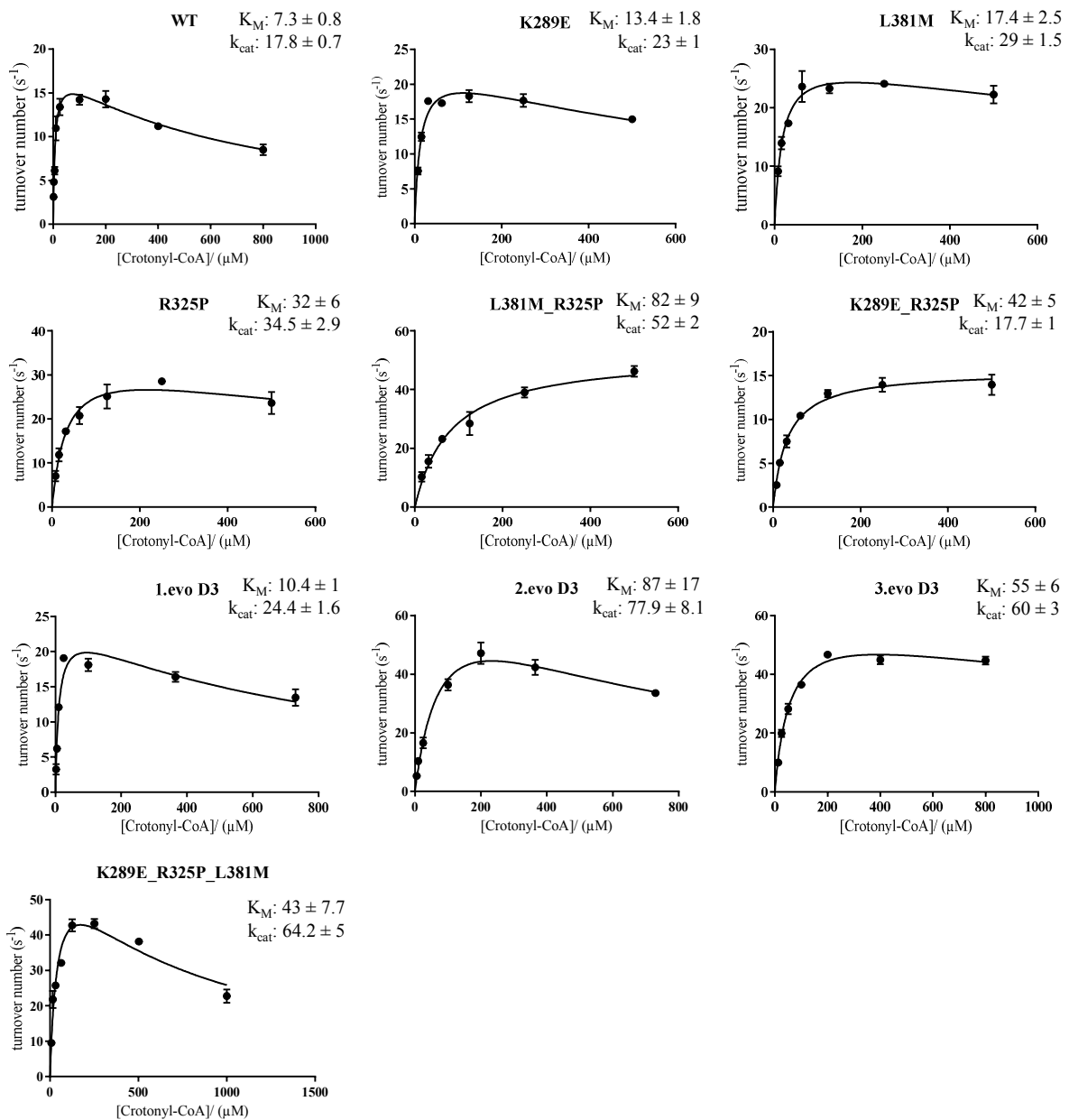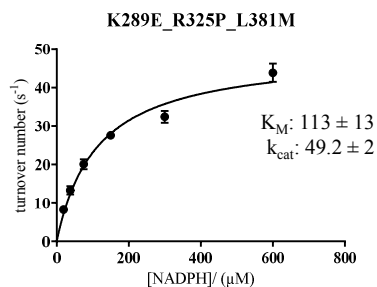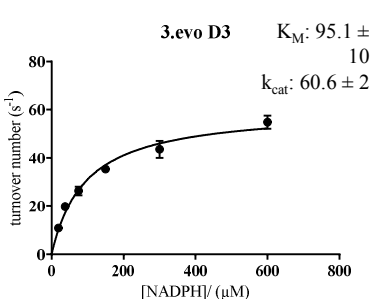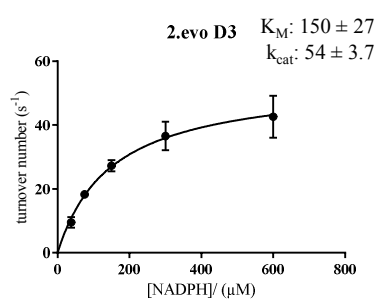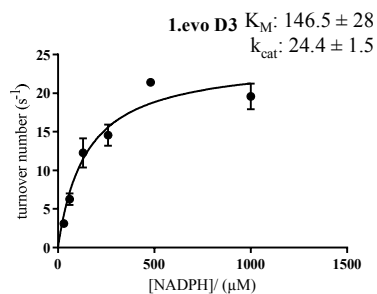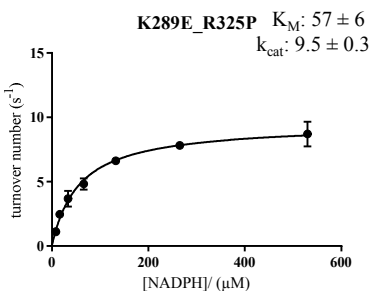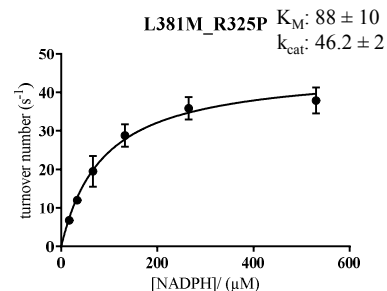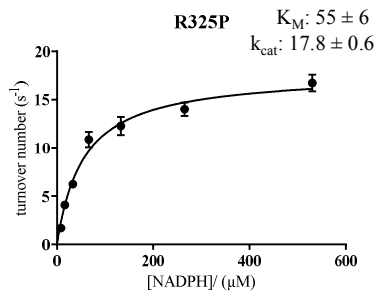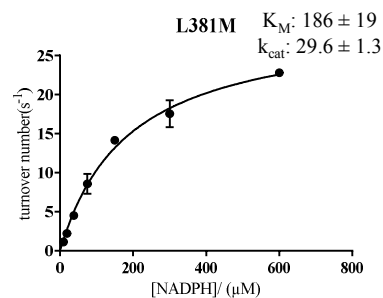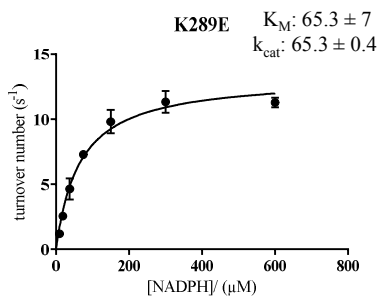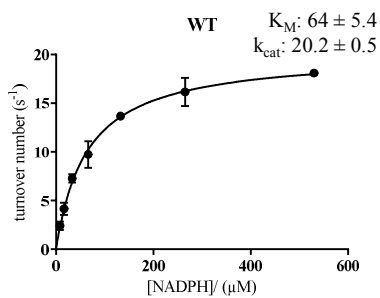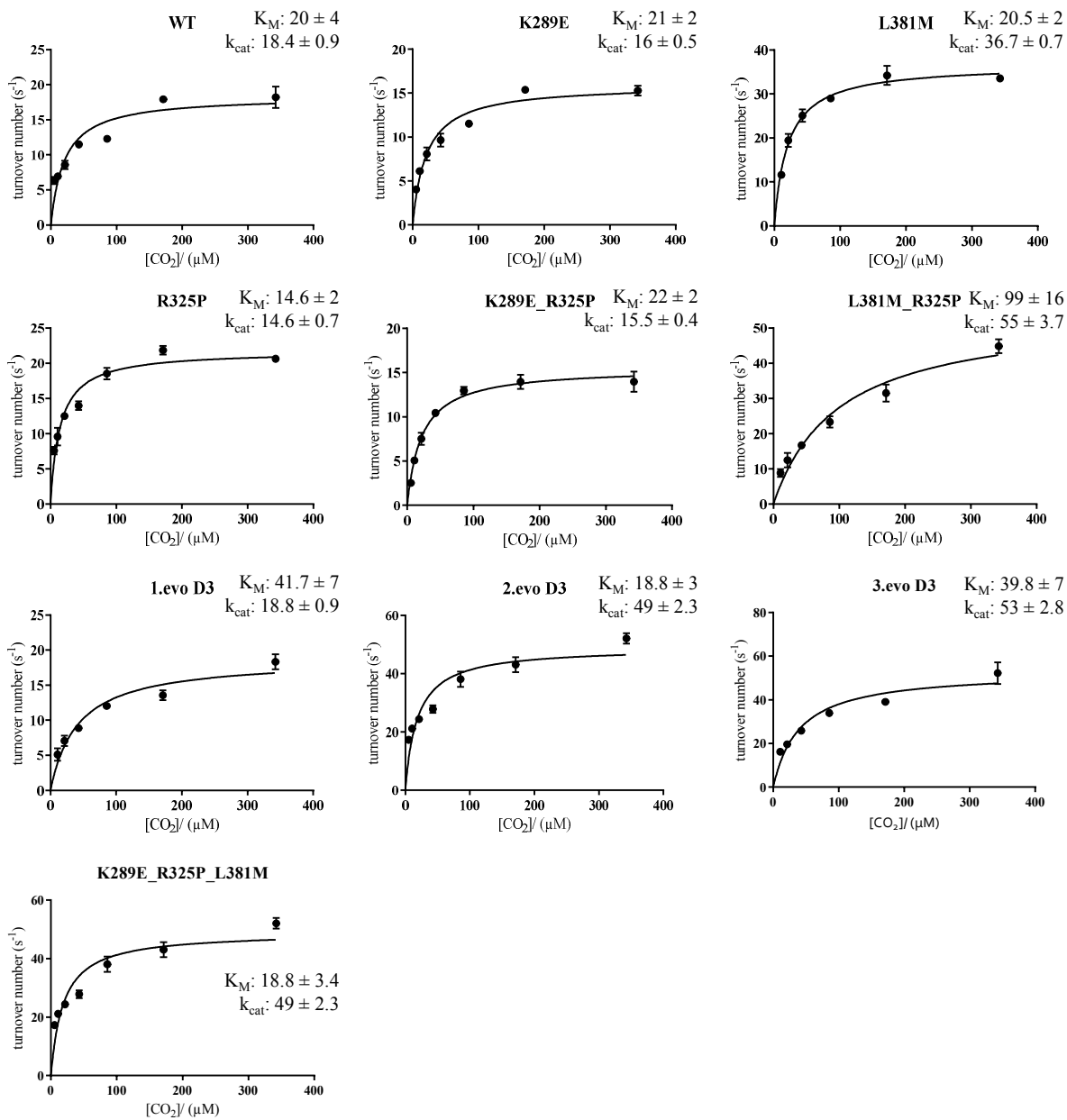
Supplementary Figure S3: Melting curves with derivation of the tryptophan fluorescence against the temperature for selected D3 variants. Samples of 0.1 µg/µL purified protein were measured in 100 mM KH$_2$PO$_4$ pH = 8. Inflection points are listed in the inset table.
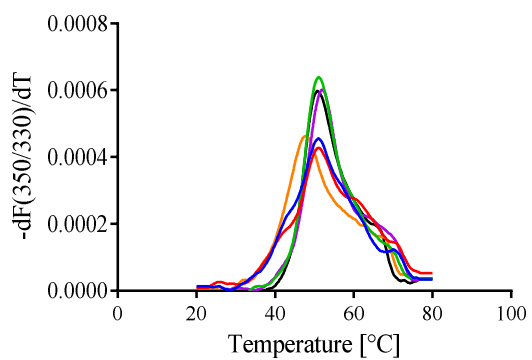
D3 variant	Inflection point
WT	50.9°C
K289E	51.0°C
R325P	50.9°C
K289ER325P	51.8°C
1.evo D3	47.8°C
2.evo D3	50.7°C

6. General Discussion and Outlook

6.1. CO_2 binding in carboxylases

One of the goals of this thesis was to gain a better understanding of the catalytic principle that ECRs employ for the interaction with CO_2. ECRs use a unique mechanism of reductive carboxylation and operate at very fast turnover rates. Moreover, ECRs do not require ATP in contrast to biotin carboxylases[1] and are fully specific for CO_2 in contrast to RuBisCO. Therefore, ECRs represent an interesting case study for the understanding of this catalytic principle. This work shows that ECRs utilize four amino acid residues in their active site to promote carboxylation. The challenging tasks carboxylases face is that of CO_2 binding and suppression of side reactivity. The enolate intermediate generated during catalysis represents an extremely potent carbon nucleophile and must therefore be efficiently protected from solvent water to prevent its protonation. Positioning of CO_2 close to the enolate intermediate is crucial to promote carboxylation. We demonstrated that the anchoring of the CO_2 molecule is performed through interaction with the sidechain of an asparagine residue and a water molecule held in position by a glutamate and a histidine sidechain. QM/MM simulations also showed that in the absence of these residues the CO_2 molecule starts to tumble and is not correctly positioned, unlike in the wild-type enzyme. Bringing substrates in close proximity in the active site of enzymes is a factor that highly contributes to catalysis [2]. Carboxylases use this principle to promote carboxylation of the enolate compared to its protonation. This is partially supported by the observation that decarboxylases have to separate the carbanion from CO_2 to efficiently complete the reaction as the reactivity of both components (enolate and CO_2) makes this step readily reversible[3]. The fourth residue important for ECR catalysis is a conserved active site phenylalanine (or tyrosine). This residue was shown to be responsible for shielding the active site from bulk water, thereby preventing the irreversible protonation of the enolate. The reason why this residue is crucial lies in the mechanism through which ECRs generate the enolate intermediate. As highlighted in the introduction, carboxylases employ various mechanisms for the generation of enolates[4]. In this respect, ECRs represent a unique example. The hydride transfer from NADPH onto the β-position of crotonyl-CoA represents the driving force for the first half-reaction in ECRs. This causes the enolate to be more committed for the subsequent steps in catalysis compared to other carboxylases. Protonation of the enolate leads to the irreversible formation of butyryl-CoA, therefore

consuming one equivalent of NADPH without productive carboxylation. In other carboxylases, the enolisation occurs by abstraction of a proton by an active site base. This generates an active site acid, which could potentially protonate the enolate and go back to the substrate, as is the case for example in RuBisCO[5]. It was shown that in RuBisCO this leads to a non negligible formation of the substrate ribulose-1,5-bisphosphate from the enolate intermediate[6]. This is not the case for ECRs as chapter 2 showed that under saturating amounts of CO_2 the conversion to (2S)-ethylmalonyl-CoA is complete. The increased commitment of the enolate towards productive carboxylation might represent an evolutionary driving force that brought ECRs to limit the water diffusion into their active site.

The precise alignment of CO_2 together with the efficient shielding of the active site from water are the most important aspects of ECR's catalysis.

PEP carboxykinase was proposed to have a CO_2 binding pocket composed of residues which wait for the incoming CO_2 molecule and anchor it for carboxylation of the enolate[7]. A protonated arginine also might have the function of polarizing the C-O bond of CO_2 thereby making the central carbon atom more electrophilic. ECRs represent a similar case in that the CO_2 molecule is anchored in the active site ready for nucleophilic attack by the enolate. Up to date there is no evidence for CO_2 binding in carboxylases and likely this is because of the apolar nature of CO_2, which makes the interactions very difficult. A bioinformatic study tried to generalize what residues are commonly found at the active site of carboxylases and speculated about the nature of the interaction these residues might undergo with CO_2[8]. It was concluded that H-bond interactions with polar residues are the preferential mode of interaction with CO_2. Other examples of interactions include hydrophobic and van der Waals interactions. Moreover they observed that CO_2 is coordinated by the protein via the oxygen atoms and not via the central carbon atom. The latter represents the point of attack of the enolate and must therefore not be sterically hindered. ECRs also obey this principle because they coordinate the CO_2 via the oxygen atoms. Interestingly, one oxygen is coordinated with a hydrogen bond to a water molecule, which has not been described yet. This work contributed in identifying new interactions and might aid our prediction of carboxylase function based on active site residues.

6.2. Structural Determinants of ECR catalysis

The fourth and fifth chapter of this thesis aimed at gaining better understanding of the structural elements that promote the fast catalytic rates observed in ECRs. A detailed structural study revealed the complex communication between subunits of the oligomeric ECR complex of *Kitasatospora setae*. Aided by four new crystal structures of ECR with different ligands (cofactor and substrate) we were able to propose the major structural rearrangements during the catalytic cycle (chapter 4) for this enzyme. The functional homotetrameric ECR complex differentiates into distinct functional units. Binding of the NADPH cofactor forces the homotetramer into a dimer of dimer conformation. Each dimer is composed of two monomers with the bound cofactor in either an open- or closed-form. The structure containing the substrate analog butyryl-CoA revealed that the disparity between subunits is still present and only one molecule of substrate is bound per dimer. Interestingly, the electron density of the CoA adenine ring can be observed in the adenosine binding pocket, suggesting that a new substrate molecule is approaching for catalysis or that a product molecule is dissociating from the enzyme as the last step in catalysis. We described how the communication between dimers and within monomers of a dimer functions on a molecular level. Inter- and intradimer communication is what enables the fast catalytic rates observed in KsECR. This synchronization was likely also optimized in the directed evolution study described in chapter 3. In fact, one mutation occurred at a residue found at the interface of two monomers namely R325P. In KsECR the analogous residue K332 was shown to have major impact on the catalytic rate when mutated to an alanine. It is tempting to speculate that only primary ECRs perform synchronized catalysis whereas secondary ECRs are simply not evolved for fast catalytic rates because of the reduced evolutionary pressure in this metabolic context. It was previously reported that enzymes do not need to have a low K_M value if the intracellular substrate concentration is always saturating and the enzyme can always operate at its k_{cat}[9]. This is a consequence of the intrinsic accuracy rate tradeoff[10] observed in enzymes. Our system might therefore provide an environment where crotonyl-CoA and NADPH are present at saturating amounts and this might limit the improvement of the K_M values with the directed evolution approach described in chapter 5.

6.3. Outlook

This work aimed at describing the underlying mechanistic principles of catalysis in ECRs. We elucidated the mechanistic principles that ECRs use to promote efficient carboxylation over unwanted side reactivity. One principle is that of preventing water from diffusing into the active site. The other principle is correct positioning of CO_2 close to the reactive enolate for optimal nucleophilic attack. In general, it is not clear, whether a true binding of CO_2 occurs during the catalytic mechanism of carboxylases and if there exists a Michaelis complex with CO_2. Efforts in identifying a Michaelis complex in RuBisCO have not been successful[11]. This was also not shown in ECRs despite the fact that a saturation kinetic was determined for this substrate. It was previously reported that the observation of saturation kinetics for CO_2 does not imply that there is a Michaelis complex with CO_2[12]. Presumably, other steps, such as product release, become limiting under saturating amounts of CO_2. An alternative possibility to a Michaelis complex is that of a bimolecular reaction between the enzyme substrate complex (ECR, crotonyl-CoA and NADPH) and CO_2. This could be determined by single turnover kinetics under varying amounts of CO_2 where a linear relationship between and CO_2 concentration is expected.

We showed that residues located at the interface between functional subunits of the ECR homotetramer are important for the fast activity observed in these carboxylases. We discussed that the nature of synchronized catalysis in ECRs might be performed solely by fast primary ECRs and that secondary ECRs are not evolved for high turnover rates. Having this structural information as well as a kinetic characterization study of the ECR family[13] at hand it would be interesting to compare the residues at interfaces of subunits between ECRs of primary and secondary metabolism and test if there is a correlation with the catalytic rate. Preliminary data show that ECR homologues that operate in primary metabolism do not possess the same residues as ECR from *K. seate* (E151, N157, N217) but still have a comparable turnover number. This indicates that there are other aspects that need to be considered in order to correctly interpret the trends in the ECR family.

We also showed that during directed evolution of a slow ECR from *Burkholderia ambifaria* in the context of primary metabolism, the enzyme accumulated three mutations, which led to a 4-fold increase in turnover number and a concomitant increase of the apparent K_M for

crotonyl-CoA. One mutation (R325P), is located at the interface between two monomers and corresponds to the same residue in the enzyme homologue from *K.setae* (K332) described in chapter 4. Mutation of R332 to proline in *K. setae* did not show an improved turnover number. This is equivalent to the R332A mutant, which showed a decreased turnover number because of the compromised communication between monomers. This might be indicative of the fact that ECR from *K. setae* is already fully optimized in terms of evolution and might therefore show only deleterious effects upon mutation.

The functional characterization of the CO_2 binding pocket in ECRs provide new knowledge about interactions between proteins and CO_2. This could be employed for the identification of CO_2 binding motifs in enzymes or for a rational design of carboxylases. One possibility would be the reversal of a decarboxylase reaction. This was demonstrated for L-methionine decarboxylase, a thiamine diphosphate-dependent enzyme[14]. Is it possible to design a *de novo* carboxylase? The general knowledge about carboxylase mechanisms and what was acquired in this study suggest that an enolase represents a good starting point, as the generation of an enolate is essential to carboxylation reactions. For this purpose aldolases represent a family of enzymes that would be well suited as a starting point for such approaches. Aldolases are enzymes that catalyze aldol reactions, which also involve enolate intermediates. The enzyme family is divided into 2 classes, based on their reaction mechanism. Class II aldolases have an active site Zn^{2+} which polarizes the carbonyl oxygen of the substrate whereas class I aldolases possess an active site lysine, which generates a Schiff base with the substrate. Both strategies facilitate deprotonation at the alpha carbon thereby forming an enolate or an enamine intermediate in the case of class I or II aldolases, respectively. Note that an enamine intermediate reacts in an analogous way to an enolate, mechanistically speaking. The enzyme D-fructose 6-phosphate aldolase belongs to class I subfamily and catalyzes the cleavage of fructose 6-phosphate to dihydroxyacetone and glyceraldehyde-3-phosphate. This enzyme was engineered to accept a variety of ketones and aldehydes and produced a variety of unusual sugar compounds[15]. Having this in mind, one could engineer this enzyme to utilize CO_2 instead of an aldehyde as the electrophile to carboxylate the substrate. In this work, it was shown that in ECR CO_2 can be exchanged with formaldehyde as an electrophilic species, suggesting that the two electrophiles are very similar. For further improvement, the carboxylation the knowledge about CO_2 binding pockets available and acquired here could be employed to

design the active site of this novel carboxylase. This enzyme variant could be used as a starting point for directed evolution of a highly proficient carboxylase.

6.4. References

1. Tong L, Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* 70(5):863-891 **2013**.
2. Menger FM, On the Source of Intramolecular and Enzymatic Reactivity. *Acc. Chem. Res.* 18(5):128-134 **1985**.
3. Kluger R, Decarboxylation, CO2 and the reversion problem. *Acc. Chem. Res.* 48(11):2843-2849 **2015**.
4. Schada von Borzyskowski L, Rosenthal RG, & Erb TJ, Evolutionary history and biotechnological future of carboxylases. *J. Biotechnol.* 168(3):243-251 **2013**.
5. Cleland WW, Andrews TJ, Gutteridge S, Hartman FC, & Lorimer GH, Mechanism of Rubisco: The Carbamate as General Base. *Chem. Rev.* 98(2):549-562 **1998**.
6. Tcherkez G, Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters. *Plant Cell Environ.* 36(9):1586-1596 **2013**.
7. Cotelesage JJH*, et al.*, How does an enzyme recognize CO2? *Int. J. Biochem. Cell. B.* 39(6):1204-1210 **2007**.
8. Cundari TR*, et al.*, CO2-formatics: how do proteins bind carbon dioxide? *J Chem Inf Model* 49(9):2111-2115 **2009**.
9. Park JO*, et al.*, Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* 12(7):482-489 **2016**.
10. Tawfik DS, Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr. Opin. Chem. Biol.* 21:73-80 **2014**.
11. Pierce J, Lorimer GH, & Reddy GS, Kinetic Mechanism of Ribulosebisphosphate Carboxylase - Evidence for an Ordered, Sequential Reaction. *Biochemistry* 25(7):1636-1644 **1986**.
12. O'Leary M, 6 Catalytic strategies in enzymic carboxylation and decarboxylation *The Enzymes*, Vol 20, pp 235-269 **1992**
13. Peter DM Substrate Promiscuity, Kinetics and Engineering of Enoyl-CoA Carboxylases/Reductases. Thesis (ETH Zürich) **2016**
14. Martin J, Eisoldt L, & Skerra A, Fixation of gaseous CO2 by reversing a decarboxylase for the biocatalytic synthesis of the essential amino acid L-methionine. *Nat. Catal.* 1(7):555-561 **2018**.
15. Roldan R*, et al.*, Biocatalytic Aldol Addition of Simple Aliphatic Nucleophiles to Hydroxyaldehydes. *ACS Catal.* 8(9):8804-8809 **2018**.

Acknowledgements

"Science is logic but sometimes stuff just happens!"

Curriculum Vitae

Gabriele Martino Marzio Stoffel
Birth Date: 22.04.1990
Nationality: Switzerland

Education

2005-2009   **Matura**, Biology and Chemistry, Liceo Cantonale Bellinzona

2010-2013   **B. Sc. Biology,** Biologie Fachrichtung Chemie, ETH Zurich

2013-2015   **M. Sc. Biology,** Biological Chemistry, ETH Zurich
                                 Thesis: Investigating the Mechanism of ACP-Enoyl Reductase
InhA

                                 Supervisor: Prof. Dr. Tobias Erb

Conferences and Workshops

March 2016   **From Chemistry to Biology,** Schwarzenberg (Austria).

March 2017   **Annual Conference of the Association for General and Applied
                                 Microbiology,** Würzburg (Germany)**.** Poster presentation:
                                 Crotonyl-CoA carboxylase/reductase as model enzyme to study
                                 protein-CO_2 interactions.

January 2018   **12th New Enzymology Kinetics Workshop,** Austin (Texas).

July 2018   **GRC conference Enzymes, Coenzymes and Metabolic
                                 Pathways,** Waterville Valley (New Hampshire). Poster
                                 presentation: Four amino acids make a carboxylase – Defining
                                 the CO_2 binding pocket of enoyl-CoA carboxylases/reductases.

Languages

Italian   Native

English   Fluent

German   Fluent

French   Fluent

Publications

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. *J. Biol. Chem.* 293(44):17200-17207

Hasan DeMirci[*], Yashas Rao[*], Gabriele M.M. Stoffel[*], Bastian Vögeli[*], Kristina Schell[*], Alexander Batyuk, Cornelius Gati, Raymond G. Sierra, Mark S. Hunter, E. Han Dao, Halil I. Ciftci, Brandon Hayes, Fredric Poitevin, Kensuke Tono, David Adrian Saez, Esteban Vöhringer-Martinez, Samuel Deutsch, Yasuo Yoshikuni, Tobias J. Erb, Soichi Wakatsuki

(2019) Coupled inter-subunit dynamics enable the fastest CO_2-fixation by reductive carboxylases. Available as preprint on bioRxiv doi 10.1101/607101

Stoffel G. M. M.[*], Saez D. A., DeMirci H., Vögeli B., Rao Y., Zarzycki J., Yoshikuni Y., Wakatsuki S., Vöhringer-Martinez E., Erb T. J.

(2019) Four amino acids define the CO_2 binding pocket of enoyl-CoA carboxylases/reductases. *Proc. Natl. Acad. Sci. USA 116 (28) 13964-13969*

Other Employment

2009-2010   Service in the Swiss Army and promoted to rank of sergeant.

Accademic Referees

Prof. Dr. Tobias Erb
Supervision Master and PhD thesis
MPI for terrestrial microbiology Marburg, Germany
+49 6421 178 700
toerb@mpi-marburg.mpg.de