Propionyl-CoA Synthase:

Characterization, Engineering and Physiological Role of a Trifunctional Fusion Enzyme

Dissertation

kumulativ

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

dem Fachbereich Biologie der Philipps-Universität Marburg
vorgelegt von

Iria Grundling
(geb. Bernhardsgrütter)

aus Zug, Schweiz

Marburg/Lahn, Deutschland, 2020

Die vorliegende Dissertation wurde von Mai 2016 bis Februar 2020 unter der Betreuung von Herrn Prof. Dr. Tobias Jürgen Erb am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg in der Abteilung „Biochemie und Synthetischer Metabolismus“ angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am 25.08.2020

Erstgutachter: Prof. Dr. Tobias Jürgen Erb
Zweitgutachter: Prof. Dr. Johann Heider

Weitere Mitglieder der Prüfungskommission:
Prof. Dr. Lars-Oliver Essen
Prof. Dr. Hans-Ulrich Mösch

Tag der Disputation: 07.09.2020

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel **„Propionyl-CoA Synthase: Characterization, Engineering and Physiological Role of a Trifunctional Fusion Enzyme"** selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als den 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 17. Juni 2020

Iria Grundling

"I do not know what I may appear to the world; but to myself I seem to have been only like a [girl] playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, while the great ocean of truth lay all undiscovered before me."

Isaac Newton

in Memoirs of the Life, Writings, and Discoveries of Sir Isaac Newton (1885) by Sir David Brewster

Table of Contents

Summary

Anthropogenic carbon dioxide (CO_2) emissions cause an imbalance in the global carbon cycle that consequently leads to global warming. Besides the indisputable role of CO_2 as harmful greenhouse gas, this small molecule harbors great potential as a simple and accessible carbon source. To exploit this potential, biotechnological strategies need to be established to convert CO_2 into value-added products, like fuels or antibiotics. It is therefore indispensable to identify and characterize efficient carboxylases. To date, the members of the enoyl-CoA carboxylase/reductase (Ecr) family account for the most efficient carboxylases found in nature. Their efficiency partly depends on the effective stabilization of the CO_2 molecule within the active site. The conserved CO_2-binding motif is characteristic for Ecrs.

This work deals with the thorough study of the three-domain fusion enzyme propionyl-CoA synthase (Pcs) of *Erythrobacter* sp. NAP1. This complex enzyme comprises the Ecr family CO_2-binding motif in its reductase domain, suggesting a potential carboxylase activity and therefore deserves detailed investigation.

The first part sets a focus on the biochemical features of Pcs. Combined kinetic and structural analysis proposed that Pcs uses a highly synchronized catalytic mechanism to sequester its reactive intermediate acrylyl-CoA. X-ray crystallography revealed an enclosed reaction chamber that features all three active sites of the fusion enzyme. This allows for the catalysis of the three subsequent reactions within the chamber. Kinetic data supported the idea that conformational changes in the Pcs ligase domain regulate the opening and closing of the catalytic compartment. Additional structural elements in Pcs either mimic domains of neighboring protomers in stand-alone homologues that contribute essential residues for catalysis or seal the reaction chamber. The presumed carboxylation potential of the reductase domain was demonstrated albeit at a very low efficiency in Pcs wildtype. Rational design was used to implement the two principles of efficient carboxylation known from Ecrs into the Pcs reductase domain. Improved CO_2-binding and shielding of the active site from water converted the reductase domain into a carboxylase domain. The engineered trifunctional, substrate-channeling carboxylase could prove advantageous in synthetic CO_2-fixation pathways.

In the second part of this work, light is shed on the physiological and ecological role of Pcs. While well described in the context of the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*, the presence of Pcs in the genome of several (potential photo-) heterotrophic microorganisms suggests

an alternative function. The genome of the aerobic anoxygenic phototrophic bacterium *Erythrobacter* sp. NAP1 encodes homologous enzymes of a partial 3-hydroxypropionate bi-cycle able to convert acetyl-CoA and two bicarbonate molecules into succinyl-CoA. The two key enzymes, Pcs and malonyl-CoA reductase (Mcr), were shown to be upregulated when the cells were grown in the light. Hence, it was suggested that this pathway might be involved in the adjustment of photosynthesis-induced redox imbalance.

Zusammenfassung

Anthropogene Kohlenstoffdioxid (CO_2) Emissionen stören das Gleichgewicht im globalen Kohlenstoffzyklus und führen schließlich zur Erderwärmung. Abgesehen von der Funktion als schädliches Treibhausgas, steckt in diesem kleinen Molekül großes Potential als einfache und zugängliche Kohlenstoffquelle. Um dieses Potential jedoch ausnutzen zu können, müssen biotechnologische Strategien entwickelt werden für die Umsetzung von CO_2 in höherwertige Produkte, wie Treibstoff oder Antibiotika. Die Identifizierung und Charakterisierung von effizienten Carboxylasen ist daher unabdingbar. Die Familie der Enoyl-CoA Carboxylasen/Reduktasen (Ecr) umfasst die am schnellsten CO_2-fixierenden Enzyme, die bisher in der Natur identifiziert wurden. Ihre Effizienz beruht unter anderem auf der erfolgreichen Stabilisierung des CO_2-Moleküls im aktiven Zentrum des Enzyms. Das konservierte CO_2-Bindungsmotiv ist charakteristisch für Ecrs.

Diese Arbeit befasst sich mit der gründlichen Untersuchung des Fusionsenzyms Propionyl-CoA Synthase (Pcs) von *Erythrobacter* sp. NAP1, welches aus drei Domänen besteht. Die Reduktase-Domäne dieses komplexen Enzyms enthält das CO_2-Bindungsmotiv der Ecr Familie, was eine potentielle Aktivität als Carboxylase vermuten lässt und daher genauer untersucht werden soll.

Der erste Teil dieser Arbeit konzentriert sich auf die biochemischen Eigenschaften von Pcs. Anhand kinetischer und struktureller Analysen wurde vorgeschlagen, dass Pcs einen hoch-synchronisierten Katalysemechanismus nutzt, um das Entweichen des reaktiven Zwischenprodukts Acrylyl-CoA zu verhindern. Proteinkristallographie offenbarte eine geschlossene Reaktionskammer, welche alle drei aktiven Zentren des Fusionsenzyms enthält. Somit wird die Katalyse der drei aufeinanderfolgenden Reaktionen im Innern der Kammer ermöglicht. Kinetische Daten stützten die Idee, dass Konformationsänderungen in der Pcs Ligase-Domäne das Öffnen und Schließen der Kammer regulieren. Zusätzliche strukturelle Elemente in Pcs imitieren entweder Domänen benachbarter Protomere in eigenständigen Homologen, welche wesentliche Aminosäurereste zur Katalyse beitragen, oder verschließen die Reaktionskammer. Das vorgeschlagene Carboxylierungspotential der Reduktase-Domäne konnte nachgewiesen werden, wenn auch nur mit sehr geringer Effizienz im Pcs Wildtyp. Rationales Design wurde angewandt, um die zwei Prinzipien der effizienten Carboxylierung, welche aus den Ecrs bekannt sind, in der Reduktase-Domäne der Pcs anzuwenden. Verbesserte CO_2-Bindung und Abschirmung des aktiven Zentrums gegen das Einströmen von Wasser überführten die Reduktase-

Domäne in eine Carboxylase-Domäne. Die verbesserte trifunktionelle Carboxylase, welche den direkten Transfer reaktiver Zwischenprodukte von einem aktiven Zentrum zum nächsten ermöglicht, könnte sich in einem synthetischen Stoffwechselweg zur CO_2-Fixierung als sehr erfolgreich erweisen.

Der zweite Teil dieser Arbeit gibt Aufschluss über die physiologische und ökologische Rolle von Pcs. Pcs wurde ursprünglich im Kontext des autotrophen 3-Hydroxypropionat Bizyklus in *Chloroflexus aurantiacus* entdeckt und detailliert beschrieben. Die Verbreitung des *pcs* Gens in mehreren (potentiell photo-) heterotrophen Mikroorganismen suggeriert jedoch noch eine alternative Funktion. Im Genom des aeroben anoxygenen phototrophen Bakteriums *Erythrobacter* sp. NAP1 sind neben *pcs* auch alle Gene des 3-Hydroxypropionat-Bizyklus Abschnittes vertreten, um Acetyl-CoA und zwei Bicarbonat-Moleküle in Succinyl-CoA umzuwandeln. Die zwei Schlüsselenzyme Pcs und Malonyl-CoA Reduktase (Mcr) waren in höheren Mengen vertreten, wenn die Zellen bei Licht wuchsen. Folglich wurde vorgeschlagen, dass dieser Stoffwechselweg in der Regulierung eines Redox-Ungleichgewichts beteiligt ist, welches durch photosynthetische Aktivität ausgelöst wird.

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

Iria Bernhardsgrütter, Bastian Vögeli, Tristan Wagner, Dominik M. Peter, Niña Socorro Cortina, Jörg Kahnt, Gert Bange, Sylvain Engilberge, Eric Girard, François Riobé, Olivier Maury, Seigo Shima, Jan Zarzycki and Tobias J. Erb. The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite. ***Nat. Chem. Biol.*** **2018**, 14, 1127.

In this work, I.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the collection of X-ray datasets
the writing of the manuscript

Iria Bernhardsgrütter, Kristina Schell, Dominik M. Peter, Farshad Borjian, David Adrian Saez, Esteban Vöhringer-Martinez and Tobias J. Erb. Awakening the sleeping carboxylase function of enzymes: engineering the natural CO_2-binding potential of reductases. ***J. Am. Chem. Soc.*** **2019**, 141 (25), 9778-9782

In this work, I.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the establishment and performance of CoA-ester analyses
the writing of the manuscript

Iria Bernhardsgrütter, Timo Glatter, Tobias J. Erb. Propionyl-CoA synthase pathway in Erythrobacter sp. NAP1: a light-induced metabolism to maintain the redox homeostasis. *(in preparation)*

In this work, I.B. contributed to:
the conception of the project
the design and performance of experiments
the analysis of data
the writing of the manuscript

The above stated contributions of I.B. to the publications or manuscript in preparation, which are part of this thesis, are herewith endorsed by:

Iria Grundling (geb. Bernhardsgrütter) Prof. Dr. Tobias J. Erb

Publications that are not discussed in this thesis:

Katharina Kremer, Murial C. F. van Teeseling, Lennart Schada von Borzyskowski, **Iria Bernhardsgrütter**, Rob J. M. van Spanning, Andrew J. Gates, Mitja N. P. Remus-Emsermann, Martin Thanbichler, Tobias J. Erb. Dynamic metabolic rewiring enables efficient acetyl coenzyme A assimilation in *Paracoccus denitrificans.* ***mBio*** **2019**, 10 (4), e00805-19

Iria Bernhardsgrütter, Gabriele Stoffel, Tobias J. Erb. Schöne neue Biologie? Synthetisch-biologische Ansätze zur CO_2-Umwandlung. ***Biospektrum*** **2020**, 26, 24-27

Charles A. R. Cotton, **Iria Bernhardsgrütter**, Simon Burgener, Luca Schulz, Hai He, Stepan Toman, Marian Dempfle, Nicole Paczia, Joachim Kopka, Steffen N. Lindner, Tobias J. Erb and Arren Bar-Even. Underground aerobic and anaerobic isoleucine biosynthesis pathways in *E. coli*. *(under review)*

Lennart Schada von Borzyskowski, **Iria Bernhardsgrütter**, Tobias J. Erb. Biochemical unity revisited: Microbial central carbon metabolism holds new discoveries, multi-tasking pathways, and redundancies with a reason. *(submitted)*

1 Introduction

1.1 Carbon dioxide – a curse and a chance

Carbon is a fundamental element in biology. It constitutes the backbone for all living matter on earth. Due to its chemistry, carbon can form covalent bonds with up to four other atoms. This makes carbon compounds highly versatile, ranging from simple one-carbon molecules like carbon dioxide (CO_2) to highly complex and branched carbon scaffolds as in lipids, carbohydrates, proteins and nucleic acids. Primary producers like plants and autotrophic microorganisms can build all these building blocks from CO_2 as sole carbon source, while heterotrophic organisms or so-called consumers require organic compounds for growth. The energy to drive growth can be either harvested from light or from chemical energy stored in the bonds of organic and inorganic compounds. Independent of the lifestyle, the metabolism of all living organisms constantly moves the carbon between the different reservoirs on earth (**Figure 1**). Even dead organic matter does not leave the carbon cycle. Over millions of years, buried organic matter is decomposed and transformed into fossil fuel, which would slowly recur through volcanic activities. This exemplifies how the carbon cycle keeps the balance between the different reservoirs over the long term. However, since the industrial era, an excess of CO_2 is emitted into the atmosphere by human activities, primarily by fossil fuel combustion. This has led to continuous increase of the atmospheric CO_2 concentration.

CO_2 represents one of the major greenhouse gases next to methane and nitric oxide and the increase in its atmospheric concentration initiates a vicious cycle of global warming. CO_2 molecules absorb the infrared radiation from reflected sunlight and thereby heat up the atmosphere. Increased temperatures in turn make CO_2 leak from the oceans and melt the pole ice, exposing surface that absorbs more heat[1]. With an atmospheric CO_2 level of about 410 ppm as of 2019[2], even halting the CO_2 emissions completely would not be sufficient to stop climate change[3]. Inevitably, we have to establish strategies to capture CO_2 efficiently from the atmosphere. A promising approach in the mitigation of climate change is the use of negative CO_2 emission technologies, like the so-called 'bioenergy with carbon-capture and storage' (BECCS) technology. This technology envisions that energy is gained from combustion of CO_2-derived biomass and released CO_2 is then captured and stored indefinitely. Despite the potential of BECCS, there are many concerns. Growth of required biomass would compete with crop plants for land, water and fertilizer and safe geological storage reservoirs would need to be established. However, the advantage of

BECCS lies in employing natural CO_2-fixation for biomass generation, which works at atmospheric CO_2 concentration and ambient conditions. In contrast, chemical CO_2 reduction reactions usually require increased CO_2 concentrations and high energy input[4]. Another strategy for climate change mitigation based on natural CO_2-fixation is reforestation. An increase of 25 % in forested area could represent a sink for about 25 % of the current atmospheric CO_2 pool[5]. However, a recent model predicts that the potential of our tropical forest trees to sequester additional CO_2 is declining[6]. Taken together, there is a high demand for alternative innovative technologies. Synthetic biology, while still in its infancy, has a great potential to harness and even improve nature's ability to capture CO_2 from the atmosphere. Moreover, synthetic biology could be employed to convert CO_2 not only into biomass but also into various complex carbon compounds such as fuels, bioplastics, or antibiotics. Towards that aim, carboxylases are the key players to provide substrates to biosynthetic pathways in a synthetic setup. In order to exploit the full potential of carboxylases, thorough understanding of the natural CO_2-fixation processes needs to proceed its application.

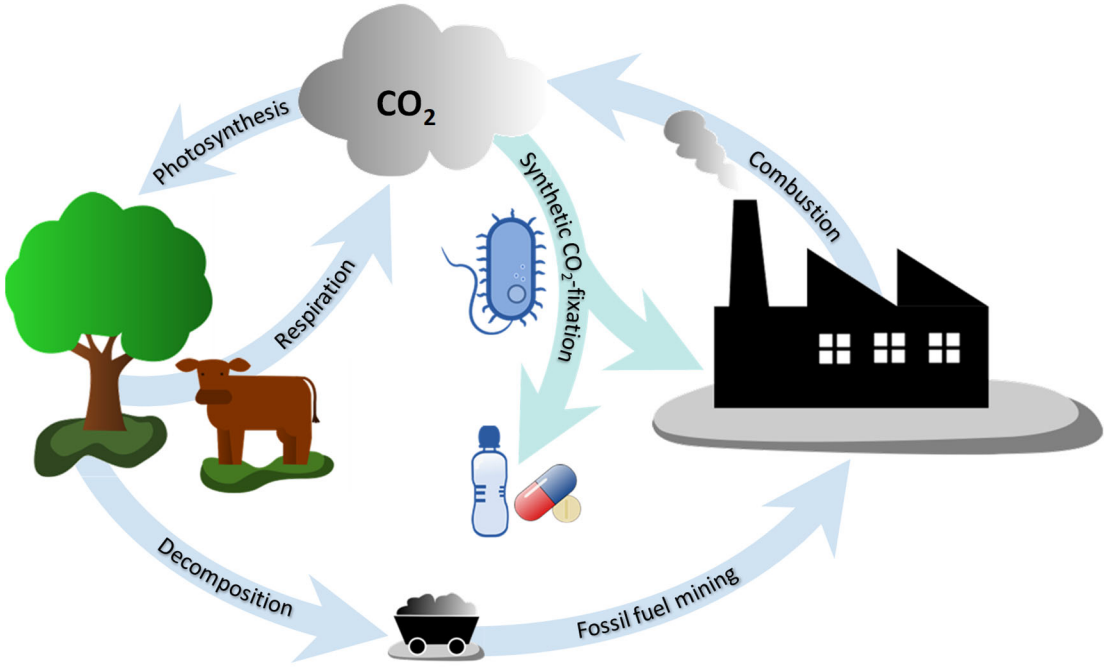


Figure 1: Simplified schematic of the global carbon cycle. Primary producers like plants fix atmospheric CO_2. This biomass serves as feed for the whole ecosystem and its decomposition under pressure yields energy-rich fossil fuels. Augmented mining of these fossil fuels and the combustion thereof lead to a disproportional release of CO_2. Synthetic biological strategies to harvest CO_2 and convert it into value-added products like biofuel, bioplastic or antibiotics might partially compensate for the increased CO_2 emissions.

1.2 Natural CO_2-fixation pathways

Biological fixation of CO_2 into organic matter is primarily catalyzed by carboxylating enzymes, so-called carboxylases. In autotrophic organisms, carboxylases are essential to assimilate biomass, which in turn feeds the ecosystem. To date, we know of six pathways for autotrophic CO_2 assimilation. The most prominent pathway is the Calvin-Benson-Bassham (CBB) cycle, which performs in all plants, algae and some bacteria[7]. Not surprisingly, the key enzyme of the CBB cycle, ribulose-1,5-bisphosphate-carboxylase/-oxygenase (RubisCO), is probably the most abundant enzyme on earth[8-10] and accounts for almost all biological carbon fixation in aquatic environments[11]. Its importance is also reflected in numerous studies undertaken to elucidate structures and the reaction mechanism in detail. In a first step, RubisCO binds the substrate ribulose-1,5-bisphosphate and forms the enediolate[12]. Proton abstraction is required for enolization and is likely performed by the carbamoyl group of an active site lysine. This carbamylation has been reported to be a crucial modification in order to activate RubisCO[13,14]. The same carbamoyl group is also required to coordinate a magnesium ion in the active site, which in turn stabilizes the enediolate. In this state, the nucleophilic enediolate attacks the CO_2 to form a new C-C bond (**Figure 2**). The C6 compound is subsequently hydrolyzed into two molecules of 3-phosphoglycerate. Despite RubisCO's predominance in biological carbon fixation, its carboxylation reaction per se is not very efficient. RubisCO is a rather slow catalyst with an average turnover rate of less than ten per second in plant homologues and with a chance of erroneously use oxygen (O_2) instead of CO_2 as substrate[15]. This oxygenation reaction is catalyzed in roughly every fifth turnover and yields the rather toxic 2-phosphoglycolate[16]. Photorespiration, the detoxification pathway for 2-phosphoglycolate, is a wasteful process that uses energy and releases CO_2 and ammonia.

The carboxylases in the autotrophic CO_2 assimilation pathways 3-hydroxypropionate (3HP) bi-cycle[17,18] and 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle[19] are acetyl-CoA and propionyl-CoA carboxylase (Acc and Pcc), which belong to the family of biotin-dependent carboxylases[20-23]. These enzymes consist of three functional domains, the biotin carboxylation (BC) domain, the biotin carboxyl carrier protein (BCCP) domain and the carboxyltransferase (CT) domain. The biotin cofactor is covalently linked through an amide bond to a lysine residue in the BCCP domain. This link serves as a swinging-arm that can swing the biotin moiety from the BC domain to the CT domain and back. The BC domain catalyzes the MgATP-dependent activation of bicarbonate to carboxyphosphate. Subsequent release of CO_2 from carboxyphosphate and enolization of biotin promote the C-N bond formation

between these substrates to yield carboxybiotin (**Figure 2**). The carboxyltransferase reaction is most likely initiated by decarboxylation of carboxybiotin in the active site of the CT domain, followed by deprotonation and enolization of the substrate, which then combines with the CO_2 through nucleophilic attack (**Figure 2**).

Similarly, the carboxylation mechanism of phosphoenolpyruvate carboxykinase (Pck), a carboxylase employed in the autotrophic dicarboxylate/hydroxybutyrate (DC/HB) cycle[24], also depends on activation of bicarbonate as CO_2 species and enolate formation within the active site[25-28]. The activated carboxyphosphate is formed from bicarbonate and the phosphate group of phosphoenolpyruvate. Subsequently, the CO_2 moiety from carboxyphosphate is released and attacked by the enolate form of pyruvate (**Figure 2**). The carboxylation reaction of the second carboxylase in the DC/HB cycle, the pyruvate:ferredoxin oxidoreductase (Por), underlies a more complex radical mechanism[29,30]. The reactive intermediate to react with the CO_2 molecule is stabilized as an enamine formed between acetyl-CoA and the cofactor thiamine pyrophosphate (**Figure 2**). The same enzyme also functions in the reductive tricarboxylic acid (rTCA) cycle[31,32] together with the 2-oxoglutarate:ferredoxin oxidoreductase, another member of the same superfamily of enzymes. Isocitrate dehydrogenase (Idh) is the third carboxylase in the rTCA cycle. The reversible isocitrate dehydrogenase catalyzes the $NAD(P)^+$-dependent oxidative decarboxylation of isocitrate in the TCA cycle and the NAD(P)H- and CO_2-dependent reductive carboxylation of 2-oxoglutarate in the rTCA cycle. The catalytic reaction occurs in three steps[33,34]. During reductive carboxylation the 2-oxoglutarate substrate is first deprotonated within the active site and stabilized in the enolate form. The enolate subsequently undergoes C-C bond formation with the CO_2 to yield oxalosuccinate, which is then reduced by NAD(P)H in the second step of catalysis (**Figure 2**).

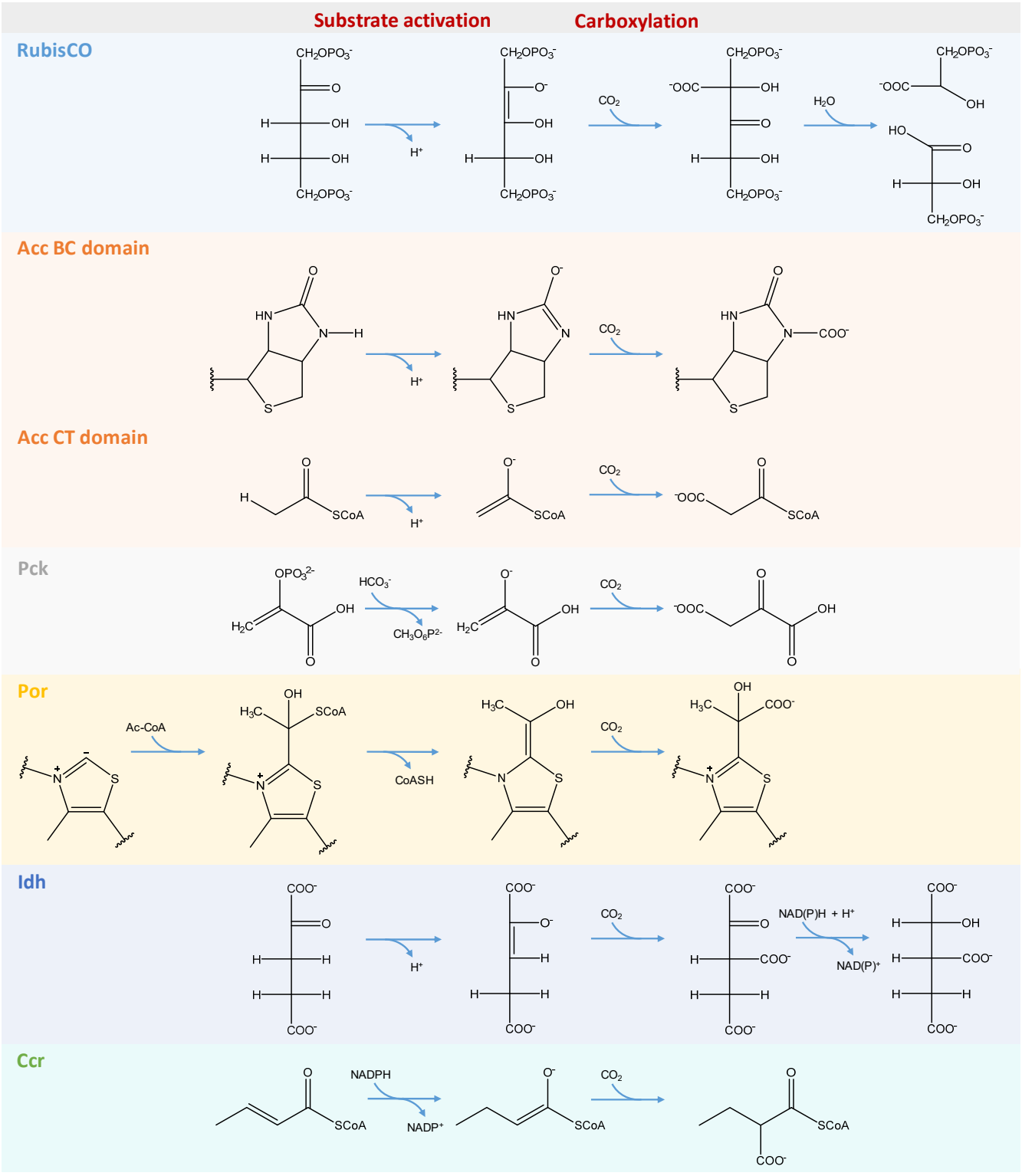


Figure 2: Simplified reaction mechanism of some selected carboxylases. The two crucial steps of carboxylation, activation of the substrate to its enolate or enamine form and C-C (or C-N) bond formation between the activated substrate and CO_2, are depicted for the selected carboxylases.

The difficulty in carboxylation reactions is the low reactivity of CO_2. Despite the very few carboxylases that have been described in this chapter, a common principle to overcome this challenge becomes apparent. The carboxylase enhances the nucleophilic reactivity of the substrate by converting it into its enolate or enamine form. These intermediates are highly reactive and could simply resolve by abstracting a proton from free water molecules or any proton donor in its vicinity. Therefore, carboxylases need to (i) stabilize the enolate within their active sites, (ii) position the resolving electrophile in close proximity to the enolate and (iii) exclude undesired electrophiles from the active site. While these principles apply to all carboxylases, there is a difference in their preference for the CO_2 species. In an aqueous solution, CO_2 and bicarbonate are at equilibrium. While bicarbonate is the predominant species, it is an even weaker electrophile than CO_2. Thus, carboxylases that use bicarbonate have to invest energy to activate it in a first step, e.g., biotin-dependent carboxylases or PEP carboxylase convert bicarbonate to carboxyphosphate.

This chapter has set a focus on carboxylases in autotrophic CO_2 fixation pathways. These "autotrophic carboxylases" [35] are essential as they supply most of the carbon for life on earth. Nevertheless, carboxylases play a crucial role in many other biological processes like organic carbon assimilation, biosynthesis or redox balancing. A few of the above-mentioned carboxylases serve in multiple pathways, e.g., biotin-dependent carboxylases that operate in fatty acid biosynthesis, polyketide biosynthesis and acetyl-CoA assimilation. Considering the broad functional spectrum of carboxylases, it will not be surprising if future studies uncover novel mechanisms and/or novel candidates of this enzyme class.

1.3 Efficient carboxylation by the enoyl-CoA carboxylase/reductase family

A significant discovery in the class of carboxylases was the enoyl-CoA carboxylase/reductase (Ecr) enzyme family. Its members catalyze the reductive carboxylation of an enoyl-CoA thioester to the corresponding malonyl-CoA derivative using CO_2 and NADPH as reducing agent. This unprecedented biochemical reaction achieves one of the fastest turnover rates among all carboxylases known to date and complete specificity for CO_2 in presence of oxygen. How is this carboxylation reaction so efficient? Comparable to the carboxylases described in chapter 1.2, Ecrs convert the enoyl-CoA substrate into its enolate form within the active site and stabilize it to increase the nucleophilic reactivity[36]. However, in the previously described carboxylases, substrate enolization is usually induced by a reversible deprotonation reaction (ΔG of about 0.5 to 1 kcal/mol)[37]. In contrast, Ecrs achieve enolization through a

hydride transfer from NADPH onto the β-position of the enoyl-CoA thioester. This step is nearly irreversible and thereby commits the enolate towards product formation rather than resolving into the substrate. The enolate then attacks the CO_2 molecule and the resulting α-carboxylated acyl-CoA (malonyl-CoA derivative) is released from the active site (**Figure 2**)[36,38]. Interestingly, this efficient carboxylation reaction remained unnoticed for a long time, because in the absence of CO_2 the enzyme catalyzes the "usual" reduction of enoyl-CoA by protonating the enolate. Only later, it has been uncovered that the reduction reaction per se, i.e., the protonation of the enolate, proceeds nonenzymatically[36]. In absence of CO_2, a covalent ene intermediate between NADPH and the enoyl-CoA ester accumulates, is released from the active site and subsequently decays into $NADP^+$ and the reduced acyl-CoA[36].

The Ecr family is involved in a wide range of physiological processes. Crotonyl-CoA carboxylase/reductase, the first Ecr that has been discovered, is the key enzyme of the ethylmalonyl-CoA pathway[38,39]. It plays a crucial role in the assimilation of two-carbon compounds like acetyl-CoA in bacteria that lack the glyoxylate cycle[34,39]. Additionally, some members of the Ecr family operate in secondary metabolism where they provide extender units for polyketide biosynthesis[40-43]. Taken together, the physiological diversity and high catalytic efficiency make Ecrs valuable candidates for applications in biotechnology and synthetic biology. This is impressively reflected in the CETCH cycle, a synthetic CO_2-fixation cycle established around the highly efficient Ecr carboxylation reaction[44]. The *in vitro* reconstituted cycle consists of 17 enzymes and compares favorably with natural autotrophic CO_2 fixation. Furthermore, engineered Ecrs have been successfully implemented in polyketide synthase systems where they provide unusual extender units for the structural diversification of polyketides[45-47].

Successful engineering of Ecrs underlies the fundamental understanding of the catalytic mechanism in general and of substrates and cofactor binding in particular. Crystal structures combined with kinetic and mutational studies offer insight into the catalytic mechanism of Ecrs. To date, there are two crystal structures of Ecrs available that include the respective CoA-ester substrate or product and the NADP cofactor in the active site (PDB 4A0S and 6OWE)[48,49]. Additionally, the CO_2 molecule was positioned within the active site by *in silico* docking and Quantum Mechanics/Molecular Mechanics (QM/MM) simulations, respectively. This structural insights allowed for speculations about binding and stabilization of substrates. Only recently, a thorough study using experimental biochemistry and computational simulations revealed molecular details about the role of the individual active site amino

acids (**Figure 3**)[49]. As suggested before[48], the side chain carboxamide NH_2 group of the active site asparagine is directly involved in CO_2 binding. The asparagine in turn is coordinated by two second shell residues, which seems to be necessary to properly orient the asparagine in respect to the CO_2 molecule. On the opposite side, the CO_2 molecule is stabilized through a complex hydrogen bonding network. An ordered water molecule coordinated by the active site histidine and glutamine is in direct contact to the CO_2 molecule. Mutation of these residues result in loss of the ordered water molecule and hence to increased tumbling of the CO_2 molecule. Nevertheless, carboxylation was fully maintained in both variants. A fourth amino acid was shown to be critical for efficient carboxylation in Ecrs. An aromatic residue (phenylalanine or tyrosine) beneath the CO_2 molecule shields the active site from water. The mutation to a non-aromatic residue allows water diffusion into the active site where it displaces the CO_2 molecule and quenches the enolate intermediate. Taken together, the four active site residues constitute the two basic principles for efficient carboxylation in Ecrs: (i) stabilization of CO_2 and proper positioning in respect to the substrate and (ii) exclusion of water from the active site.

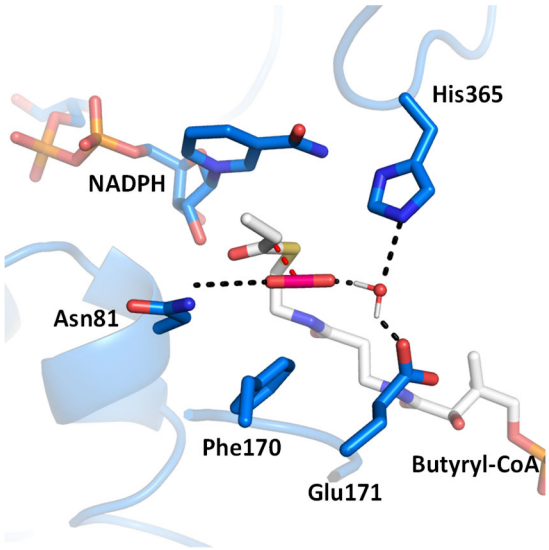


Figure 3: CO_2-binding pocket of Ecr from *Kitasatospora setae*[49]. The CO_2-binding pocket is defined by four conserved residues (Asn81, Phe170, Glu171, His365). CO_2 was modeled into the structure (PDB 6OWE).

Taken together, it was suggested that Ecr homologues containing these four conserved residues could catalyze the carboxylation of their substrates. One such homologue has been identified as the reductase domain of the multidomain propionyl-CoA synthase (Pcs). This enzyme has been described to catalyze the three step reaction sequence from 3-hydroxypropionate to propionyl-CoA in the 3-hydroxypropionate bi-cycle of *Chloroflexus aurantiacus*. There, carboxylase activity of the Pcs was unexpected, as the

reduced product propionyl-CoA is in turn substrate for subsequent enzymes in this autotrophic pathway. However, Pcs homologues are encoded beyond *C. aurantiacus* in a few potentially photosynthetic organisms, many of which belong to the aerobic anoxygenic phototrophic (AAP) bacteria[18]. The physiological role of Pcs in these bacteria and its potential to carboxylate is yet to be discovered.

1.4 Aerobic anoxygenic phototrophic bacteria

AAP bacteria are obligate aerobic organisms with a heterotrophic lifestyle that are capable of additionally harvesting light energy. Astonishingly, these bacteria have been overlooked until recently despite their wide distribution and potential ecological relevance. In marine surface water, they account for up to 15 % of the total microbial community and thus play a significant role in the marine carbon cycle[50-52]. About 2-5 % of photosynthetic electron flux in the upper ocean have been attributed to AAP bacteria[50]. The photosynthetic apparatus of AAP bacteria is similar to that of purple non-sulfur bacteria; however, their 16S ribosomal genes show a polyphyletic origin within the non-phototrophic proteobacteria. The evolutionary emergence of AAP bacteria therefore remains an enigma. It has recently been suggested that anaerobic phototrophic purple non-sulfur bacteria might have evolved independently into the proteobacterial AAP clades, in which most members lost the phototrophy during evolution[53].

The metabolic diversity across AAP bacteria exacerbates the general study of this functional group. Common to all members is the photosynthetic machinery, which consists of a type-II photosynthetic reaction center similar to that of purple non-sulfur bacteria and light-harvesting complexes that contain the pigment bacteriochlorophyll a (BChl *a*). Harvesting light energy under aerobic conditions harbors the great danger of the emergence of harmful reactive oxygen species (ROS). However, AAP bacteria have developed two major strategies to overcome this threat. First, AAP bacteria express a high amount of carotenoids, which are able to capture high-energy radiation and quench ROS[54]. And second, the BChl *a* expression underlies strong regulation and is only activated in the dark[55-57]. This regulation seems contradictory. However, regarding the bacteria's rather long generation time and the continuous day/night cycle they experience in their natural environment, BChl *a* should always be available at sufficient amounts to enable photosynthesis throughout the light period. In line with this strategy, growth promoting effects by light could only be observed when the AAP bacteria were incubated in a dark/light cycle and not if continuous illumination was applied[55,56]. While reported several times under laboratory cultivation conditions, only recently, direct evidence was also collected for the light-dependent

stimulation of natural populations of marine AAP bacteria[58]. Some AAP bacteria have been shown to accumulate as much as 25 – 110 % more carbon under light/dark conditions compared to the dark[59]. The increased growth arises mainly from the ability to invest organic compounds into biomass production rather than respiring them for energy conservation. Upon illumination, respiration was shown to decrease to about a quarter of its rate in the dark[59,60]. Additionally, increased CO_2-fixation rates in the light have been observed to provide for up to 11 % of the cellular carbon content[59]. Despite the significant contribution of carbon fixation, genetic as well as experimental evidence for autotrophic growth is lacking[61-63]. *Dinoroseobacter shibae* was shown to up-regulate the ethylmalonyl-CoA pathway in light[64]. Hence, the two carboxylases of this pathway, crotonyl-CoA carboxylase/reductase and propionyl-CoA carboxylase, could account for the increased CO_2-fixation rate. However, other AAP bacteria like *Erythrobacter* sp. NAP1 do not encode the genes for the ethylmalonyl-CoA pathway but still demonstrate enhanced CO_2-fixation in light[59]. This light-promoted efficiency in carbon utilization provides a competitive advantage over purely heterotrophic bacteria. Moreover, it has been demonstrated that the exposure to light-dark cycles extends the survival of AAP bacteria under nutrient starving conditions[64,65]. The strategies and pathways underlying these phenomena remain to be thoroughly investigated.

1.5 Aim of this thesis:

A bioinformatics inquiry revealed that one domain of the multidomain Pcs contains the CO_2-binding motif characteristic to Ecrs. So far, Ecrs had only been known as lone-standing enzymes. Therefore, this finding attracted interest and prompted for further investigations. A general aim of this study was to characterize the multidomain enzyme and to engineer it for its application in (synthetic) CO_2-fixation pathways. A second objective was the understanding of the physiological role of Pcs in its host *Erythrobacter* sp. NAP1.

Pcs is a fascinating enzyme as it comprises three catalytic functions in one scaffold. Chapter 2 aimed to comprehensively characterize Pcs using biochemical and structural approaches. These data could build the base to understand the mechanistic principles of such a complex enzyme. Thorough kinetic assays could provide insight into the synchronization of the three-step reaction sequence. Besides, the evolutionary emergence and advantage of such a fusion enzyme deserved study. Structural arrangements required to fuse three stand-alone enzymes into a fully functional complex could in turn hint towards essential regulatory or catalytic contributions. Canonical motives for the evolution of Pcs needed to be probed, e.g., substrate channeling or increased catalytic efficiency compared to stand-alone enzymes. Regarding the highly reactive acrylyl-CoA intermediate, substrate channeling was a highly likely outcome. However, substrate channeling of the bulky and electrically charged CoA-thioester intermediate would presumably require a mechanism that differentiates from canonical substrate channeling enzymes. The elaborate study could deepen our understanding of general principles in fusion and substrate channeling enzymes. Furthermore, it would allow for detailed comparison of the Pcs reductase domain with Ecrs. The question was raised whether the two principles for efficient carboxylation in Ecrs, i.e., effective stabilization of CO_2 and shielding of the active site from water, could be universally applied to Ecr homologues. Chapter 3 aimed to address this question on the example of the Pcs reductase domain. A carboxylating Pcs would be a highly attractive candidate to be employed in (synthetic) CO_2-fixation pathways, like the HOPAC cycle that includes the same three-step reaction sequence. Furthermore, this proof-of-principle would unlock a great potential to engineer carboxylation activity within Ecr-related reductases.

Pcs had been identified as part of the autotrophic 3-hydroxypropionate bi-cycle in *C. aurantiacus*. The genomic distribution of *pcs* in several (potential photo-) heterotrophs raised the question of the

physiological role of this enzyme within these organisms. The objective in chapter 4 was to investigate the physiological and ecological role of Pcs in the aerobic anoxygenic phototrophic bacterium *Erythrobacter* sp. NAP1. *In vivo* approaches across different growth conditions could reveal the regulation and function of Pcs in the metabolism of this marine bacterium.

1.6 References

1 Perovich, D. K. The changing Arctic sea ice cover. *Oceanography* **24**, 162-173 (2011).

2 NOAA Office of Ocean Exploration and Research. *Monthly Average Mauna Loa CO_2*, <https://www.esrl.noaa.gov/gmd/ccgg/trends/index.html> (2020).

3 Archer, D. *et al.* Atmospheric lifetime of fossil fuel carbon dioxide. *Annu. Rev. Earth Pl. Sc.* **37**, 117-134 (2009).

4 Sakakura, T., Choi, J.-C. & Yasuda, H. Transformation of carbon dioxide. *Chem. Rev.* **107**, 2365-2387 (2007).

5 Bastin, J.-F. *et al.* The global tree restoration potential. *Science* **365**, 76-79 (2019).

6 Hubau, W. *et al.* Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* **579**, 80-87 (2020).

7 Bassham, J. A. *et al.* The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc.* **76**, 1760-1770 (1954).

8 Ellis, R. J. The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241-244 (1979).

9 Raven, J. A. Rubisco: still the most abundant protein of Earth? *New Phytologist* **198**, 1-3 (2013).

10 Bar-On, Y. M. & Milo, R. The global mass and average rate of rubisco. *PNAS* **116**, 4738-4743 (2019).

11 Raven, J. A. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**, 177-192 (2009).

12 Taylor, T. C. & Andersson, I. The structure of the complex between rubisco and its natural substrate ribulose 1, 5-bisphosphate. *J. Mol. Biol.* **265**, 432-444 (1997).

13 Lorimer, G. H., Badger, M. R. & Andrews, T. J. The activation of ribulose-1,5-bisphosphate carboxylase by carbon dioxide and magnesium ions. Equilibria, kinetics, a suggested mechanism, and physiological implications. *Biochemistry* **15**, 529-536 (1976).

14 Lorimer, G. H. & Miziorko, H. M. Carbamate formation on the .epsilon.-amino group of a lysyl residue as the basis for the activation of ribulosebisphosphate carboxylase by carbon dioxide and magnesium(2+). *Biochemistry* **19**, 5321-5328 (1980).

15 Orr, D. J. *et al.* Surveying Rubisco diversity and temperature response to improve crop photosynthetic efficiency. *Plant Physiol.* **172**, 707-717 (2016).

16 Sharkey, T. D. Estimating the rate of photorespiration in leaves. *Physiol. Plant.* **73**, 147-152 (1988).

17 Alber, B. E. & Fuchs, G. Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Biol. Chem.* **277**, 12137-12143 (2002).

18 Zarzycki, J., Brecht, V., Müller, M. & Fuchs, G. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *PNAS* **106**, 21317-21322 (2009).

19 Berg, I. A., Kockelkorn, D., Buckel, W. & Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318**, 1782-1786 (2007).

20 Knowles, J. R. The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* **58**, 195-221 (1989).

21 Zeczycki, T. N. *et al.* Insight into the carboxyl transferase domain mechanism of pyruvate carboxylase from *Rhizobium etli*. *Biochemistry* **48**, 4305-4313 (2009).

22 Attwood, P. V. & Wallace, J. C. Chemical and catalytic mechanisms of carboxyl transfer reactions in biotin-dependent enzymes. *Acc. Chem. Res.* **35**, 113-120 (2002).

23 Chou, C.-Y., Linda, P. & Tong, L. Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism. *J. Biol. Chem.* **284**, 11690-11697 (2009).

24 Huber, H. *et al.* A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. *PNAS* **105**, 7851-7856 (2008).

25 Kai, Y., Matsumura, H. & Izui, K. Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms. *Arch. Biochem. Biophys.* **414**, 170-179 (2003).

26 O'Leary, M. H., Rife, J. E. & Slater, J. D. Kinetic and isotope effect studies of maize phosphoenolpyruvate carboxylase. *Biochemistry* **20**, 7308-7314 (1981).

27 Hansen, D. E. & Knowles, J. The stereochemical course at phosphorus of the reaction catalyzed by phosphoenolpyruvate carboxylase. *J. Biol. Chem.* **257**, 14795-14798 (1982).

28 Fujita, N., Izui, K., Nishino, T. & Katsuki, H. Reaction mechanism of phosphoenolpyruvate carboxylase. Bicarbonate-dependent dephosphorylation of phosphoenol-α-ketobutyrate. *Biochemistry* **23**, 1774-1779 (1984).

29 Ragsdale, S. W. Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* **103**, 2333-2346 (2003).

30 Menon, S. & Ragsdale, S. W. Mechanism of the *Clostridium thermoaceticum* pyruvate: ferredoxin oxidoreductase: evidence for the common catalytic intermediacy of the hydroxyethylthiamine pyropyrosphate radical. *Biochemistry* **36**, 8484-8494 (1997).

31 Evans, M., Buchanan, B. B. & Arnon, D. I. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *PNAS* **55**, 928 (1966).

32 Buchanan, B. B. & Arnon, D. I. A reverse KREBS cycle in photosynthesis: consensus at last. *Photosynth. Res.* **24**, 47-53 (1990).

33 Hurley, J. H., Dean, A. M., Koshland Jr, D. E. & Stroud, R. M. Catalytic mechanism of NADP+-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP+ complexes. *Biochemistry* **30**, 8671-8678 (1991).

34 Aktas, D. F. & Cook, P. F. A Lysine-Tyrosine Pair Carries Out Acid− Base Chemistry in the Metal Ion-Dependent Pyridine Dinucleotide-Linked β-Hydroxyacid Oxidative Decarboxylases. *Biochemistry* **48**, 3565-3577 (2009).

35 Erb, T. J. Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* **77**, 8466-8477 (2011).

36 Rosenthal, R. G. *et al.* Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* **10**, 50 (2014).

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

38 Erb, T. J., Brecht, V., Fuchs, G., Müller, M. & Alber, B. E. Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase. *PNAS* **106**, 8871-8876 (2009).

39 Erb, T. J. *et al.* Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. *PNAS* **104**, 10631-10636 (2007).

40 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.* **26**, 90-114 (2009).

41 Wilson, M. C. & Moore, B. S. Beyond ethylmalonyl-CoA: the functional role of crotonyl-CoA carboxylase/reductase homologs in expanding polyketide diversity. *Nat. Prod. Rep.* **29**, 72-86 (2012).

42 Eustáquio, A. S. *et al.* Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *PNAS* **106**, 12295-12300 (2009).

43 Liu, Y., Hazzard, C., Eustáquio, A. S., Reynolds, K. A. & Moore, B. S. Biosynthesis of salinosporamides from α, β-unsaturated fatty acids: implications for extending polyketide synthase diversity. *J. Am. Chem. Soc.* **131**, 10376-10377 (2009).

44 Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354**, 900-904 (2016).

45 Peter, D. M. *et al.* Screening and engineering the synthetic potential of carboxylating reductases from central metabolism and polyketide biosynthesis. *Angew. Chem. Int. Ed.* **54**, 13457-13461 (2015).

46 Kundert, J. & Gulder, T. A. Extending polyketide structural diversity by using engineered carboxylase/reductase enzymes. *Angew. Chem. Int. Ed.* **55**, 858-860 (2016).

47 Vögeli, B. *et al.* Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis. *Cell Chem. Biol.* **25**, 833-839. e834 (2018).

48 Quade, N., Huo, L., Rachid, S., Heinz, D. W. & Müller, R. Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* **8**, 117 (2012).

49 Stoffel, G. M. *et al.* Four amino acids define the CO_2 binding pocket of enoyl-CoA carboxylases/reductases. *PNAS* **116**, 13964-13969 (2019).

50 Kolber, Z. S. *et al.* Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**, 2492-2495 (2001).

51 Yutin, N. *et al.* Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ. Microbiol.* **9**, 1464-1475 (2007).

52 Koblížek, M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **39**, 854-870 (2015).

53 Yurkov, V. & Csotonyi, J. T. in *The purple phototrophic bacteria* Vol. 28 *Advances in Photosynthesis and Respiration* (eds C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer, & J. Thomas Beatty) Ch. 3, 31-55 (Springer, 2009).

54 Fraser, N. J., Hashimoto, H. & Cogdell, R. J. Carotenoids and bacterial photosynthesis: The story so far. *Photosynth. Res.* **70**, 249-256 (2001).

55 Harashima, K., Kawazoe, K., Yoshida, I. & Kamata, H. Light-stimultated aerobic growth of *Erythrobacter* species OCh 114. *Plant Cell Physiol.* **28**, 365-374 (1987).

56 Yurkov, V. V. & van Gemerden, H. Impact of light/dark regimen on growth rate, biomass formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch. Microbiol.* **159**, 84-89 (1993).

57 Biebl, H. & Wagner-Döbler, I. Growth and bacteriochlorophyll a formation in taxonomically diverse aerobic anoxygenic phototrophic bacteria in chemostat culture: influence of light regimen and starvation. *Process Biochem.* **41**, 2153-2159 (2006).

58 Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M. & Gasol, J. M. Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME J.* **11**, 2391-2393 (2017).

59 Hauruseu, D. & Koblížek, M. Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl. Environ. Microbiol.* **78**, 7414-7419 (2012).

60 Koblížek, M., Mlčoušková, J., Kolber, Z. & Kopecký, J. On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch. Microbiol.* **192**, 41-49 (2010).

61 Koblizek, M. *et al.* Genome sequence of the marine photoheterotrophic bacterium *Erythrobacter* sp. strain NAP1. *J Bacteriol* **193**, 5881-5882 (2011).

62 Swingley, W. D. *et al.* The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. *J. Bacteriol.* **189**, 683-690 (2007).

63 Fuchs, B. M. *et al.* Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *PNAS* **104**, 2891-2896 (2007).

64 Bill, N. *et al.* Fixation of CO_2 using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ. Microbiol.* **19**, 2645-2660 (2017).

65 Soora, M. & Cypionka, H. Light enhances survival of *Dinoroseobacter shibae* during long-term starvation. *PLoS One* **8** (2013).

66 SHIBA, T. Utilization of light energy by the strictly aerobic bacterium *Erythrobacter* sp. OCh 114. *J. Gen. Appl. Microbiol.* **30**, 239-244 (1984).

Chapter II

The multi-catalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite

Authors:

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

* These authors contributed equally to this work.

Published in:

Nature Chemical Biology **14**, 1127–1132(2018)

https://doi.org/10.1038/s41589-018-0153-x

Author contributions:

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

2 The multi-catalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite

2.1 Abstract

Cells need to cope with toxic or reactive intermediates formed during metabolism. One strategy is to sequester reactions that produce such intermediates within specialized compartments or tunnels connecting different active sites. Here we show that propionyl-CoA synthase (PCS), a ~400 kDa homodimer, three-domain fusion protein and the key enzyme of the 3-hydroxypropionate bi-cycle for CO_2-fixation, sequesters its reactive intermediate acrylyl-CoA. Structural analysis showed that PCS forms a multi-catalytic reaction chamber. Kinetic analysis suggests that access to the reaction chamber and catalysis are synchronized by interdomain communication. The reaction chamber of PCS features three active sites and has a volume of only 33 nm^3. As one of the smallest multi-reaction chambers described in biology, PCS could inspire the engineering of a new class of dynamically regulated nanoreactors.

2.2 Introduction

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

Here, we report on the enzyme reaction cascade from 3-hydroxypropionate to propionyl-CoA, which is the key sequence in the 3-hydroxypropionate bi-cycle for autotrophic CO_2 assimilation [3,4]. The overall reaction sequence comprises three enzymatic steps, during which a highly reactive, toxic and unstable intermediate, acrylyl-CoA, is formed[5]. In autotrophic *Sulfolobales* the three reactions are

catalyzed by individual enzymes [6,7]. However, in several phyla (e.g., Proteobacteria and Chloroflexi, **Supplementary Figure 1**) the three reactions are catalyzed by a fusion enzyme of about 1850 amino acids, that comprises three catalytic domains, the propionyl-CoA synthase (PCS, **Figure 1** and **S2**) [3], suggesting that PCS specifically evolved as a fusion enzyme to overcome the free diffusion of reactive acrylyl-CoA. Here we show that the PCS from *Erythrobacter* sp. NAP1 (GenBank accession no. EAQ29651) is a multi-catalytic "nanoreactor" that features a central reaction chamber to sequester its reaction intermediates acrylyl-CoA during catalysis. Biochemical experiments show that access to the reaction chamber is dynamically controlled during the catalytic cycle and that catalysis is synchronized through interdomain communication.

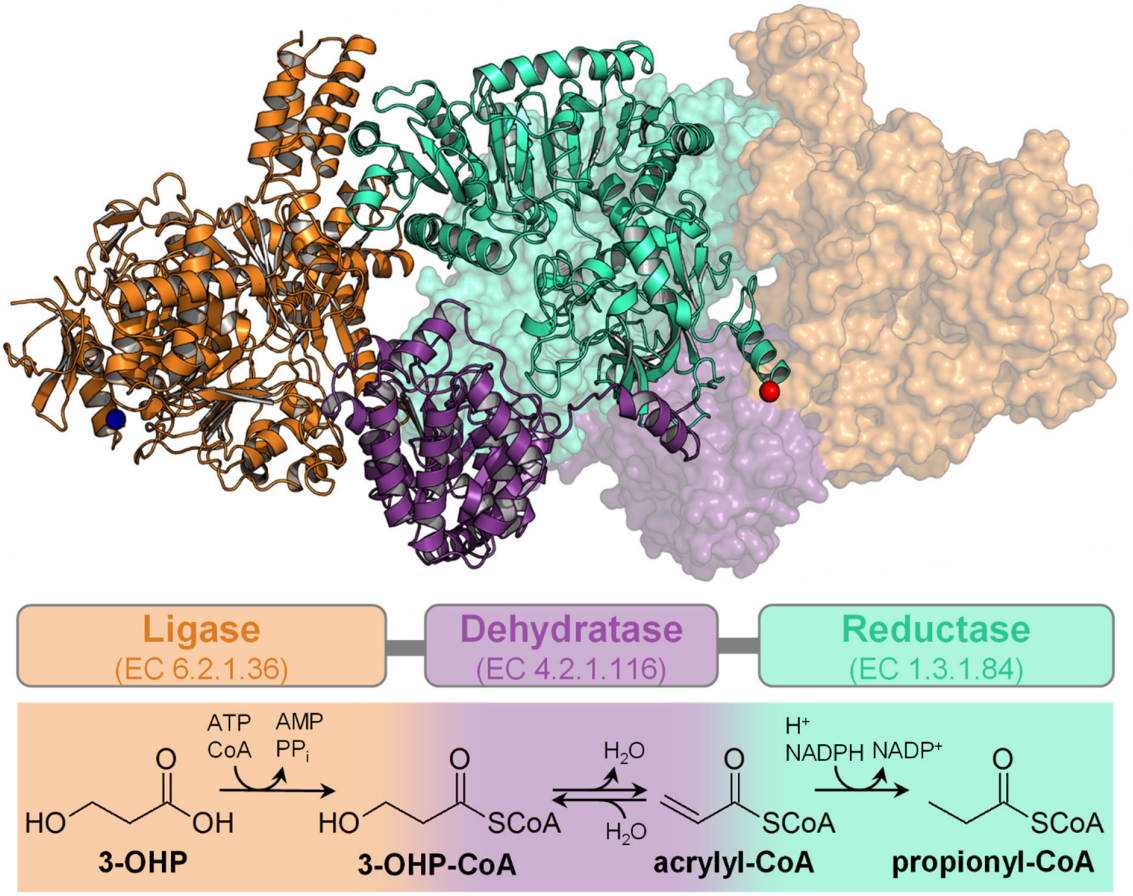


Figure 1. Trifunctional PCS: Structure and reaction sequence. Dimeric structure of PCS from *Erythrobacter* sp. NAP1 (PDB 6EQO). One protomer is depicted in cartoon and one in surface representation. The multi-domain organization is highlighted by different colors: orange, ligase domain; purple, dehydratase domain; cyan, reductase domain; blue sphere, N-terminus; red sphere, C-terminus. Schematic arrangement of the three domains and their individual reactions are shown using the same color code.

2.3 Results

PCS sequesters the reactive intermediate acrylyl-CoA

To study the catalytic cycle of PCS from 3-hydroxypropionyl-CoA to propionyl-CoA (**Figure 1**), we produced PCS from *Erythrobacter* sp. NAP1 (GenBank accession no. EAQ29651) and characterized the enzyme biochemically. When we followed the overall reaction of PCS, we could not detect acrylyl-CoA in the assay mixture, while the other intermediates 3-hydroxypropionyl-CoA and propionyl-CoA accumulated over time (**Figure 2a**). Only when we increased the concentration of the enzyme by two orders of magnitude, acrylyl-CoA became detectable, albeit at very low levels (**Figure 2b**). The concentration of acrylyl-CoA corresponded to 1.8 ± 0.1 % of PCS monomers and stayed constant during steady-state. This demonstrated that acrylyl-CoA was formed *in situ* by the dehydratase domain, but presumably stayed sequestered within PCS, where it was quickly consumed by the reductase domain. Notably, when the reduction reaction was prevented (by omitting NADPH), acrylyl-CoA did yet not accumulate, while 3-hydroxypropionyl-CoA still did (**Figure 2c**). This indicated that even when interrupting the catalytic sequence, acrylyl-CoA remained effectively sequestered within PCS. We could also show that externally added intermediates were not or only to a minor amount converted by the enzyme (**Figure 2d** and **2e**), as discussed in detail later in this manuscript.

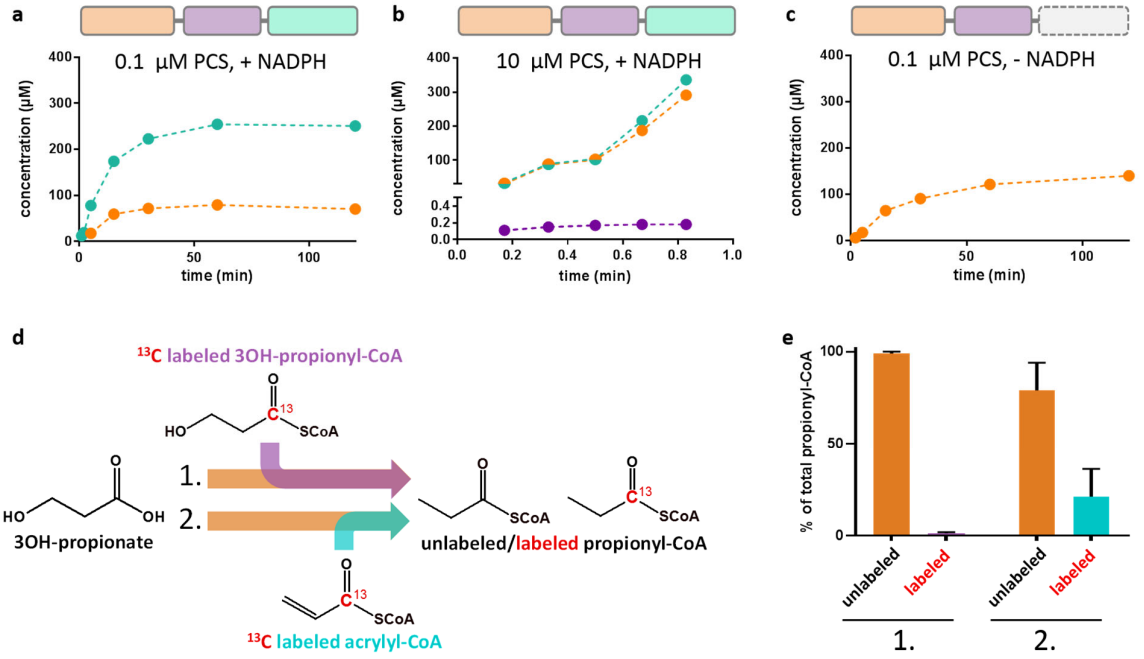


Figure 2. PCS sequesters the reactive intermediate acrylyl-CoA. a, Time course of the overall reaction with 0.1 µM PCS, 800 µM CoA, 500 µM 3-hydroxypropionate, 800 µM ATP and 300 µM NADPH. Production of the 3-hydroxypropionyl-CoA intermediate (orange) and the final product propionyl-CoA (cyan) was observed. In contrast no free acrylyl-CoA was detectable. **b**, Time course of the reaction containing 10 µM PCS, 5 mM CoA, 5 mM 3-hydroxypropionate, 5 mM ATP and 5 mM NADPH. At these high enzyme

concentrations acrylyl-CoA (purple) was detected at 0.18 µM during steady-state corresponding to 1.8% occupancy of reductase active sites. 3-Hydroxypropionyl-CoA und propionyl-CoA accumulate over time. **c**, as in a, but without NADPH. Again, formation of 3-hydroxypropionyl-CoA was observed, but not of free acrylyl-CoA. **d**, Isotopic labeling competition experiment containing unlabeled 3-hydroxypropionate and either 13C-labeled 3-hydroxypropionyl-CoA (experiment 1) or acrylyl-CoA (experiment 2). The reaction was started by the addition of PCS. Products were analyzed by LC-MS (see **Supplementary Table 4** for detailed assay conditions). **e**, Results of the isotopic labeling competition experiment. Only 0.8 ± 0.4 % of propionyl-CoA was produced from exogenous 13C-labeled 3-hydroxypropionyl-CoA during steady state (experiment 1). Approximately every fifth propionyl-CoA (21 ± 15%) was formed from exogenous 13C-labeled acrylyl-CoA during steady-state (experiment 2). a – c, data of a representative single experiment. e, data mean ± s.d. (n=3).

PCS forms a multi-catalytic reaction chamber

To understand the structural basis of acrylyl-CoA sequestration in PCS, we solved the crystal structure of the enzyme in the presence of CoA, $NADP^+$ and an ATP analog (phosphomethylphosphonic acid adenylate ester) using a new phasing compound [8] (PDB 6EQO, **Supplementary Table 1** and **Supplementary Figure 3** and **4**). PCS forms a dimer (of ~400 kDa) around a central core of reductase domains (**Figure 1**). The ligase and dehydratase domains extend to both sides, enclosing spherical compartments ('reaction chambers', **Supplementary Figure 5**) that each feature three internal active sites (**Figure 3**). The active sites are not connected through individual tunnels but all open into the central cavity of the reaction chamber. The surface of the central cavity is positively charged, which may help retaining the CoA-ester intermediates during catalysis or even guide them between active sites as shown in the example of the malate dehydrogenase / citrate synthase cascade[9-11]. Escape of intermediates from the reaction chamber is presumably disfavored by negative charges surrounding any small openings.

The inner diameter of the reaction chamber is between 3.5 – 5.5 nm with a total volume of 33 nm³. This volume is between three and six orders of magnitude smaller than that of bacterial microcompartments [12] (14×10^3 to ~10×10^7 nm^3, calculated from inner diameters ranging from 30 to 600 nm) [13], two orders of magnitude smaller than of described nanocompartments, such as encapsulins (5×10^3 nm^3, calculated from inner diameter of 22 nm) [14], and even half of that of proteasomes (59 and 84 nm^3) [15]. Thus, PCS forms one of the smallest multi-catalytic reaction chambers observed in Nature.

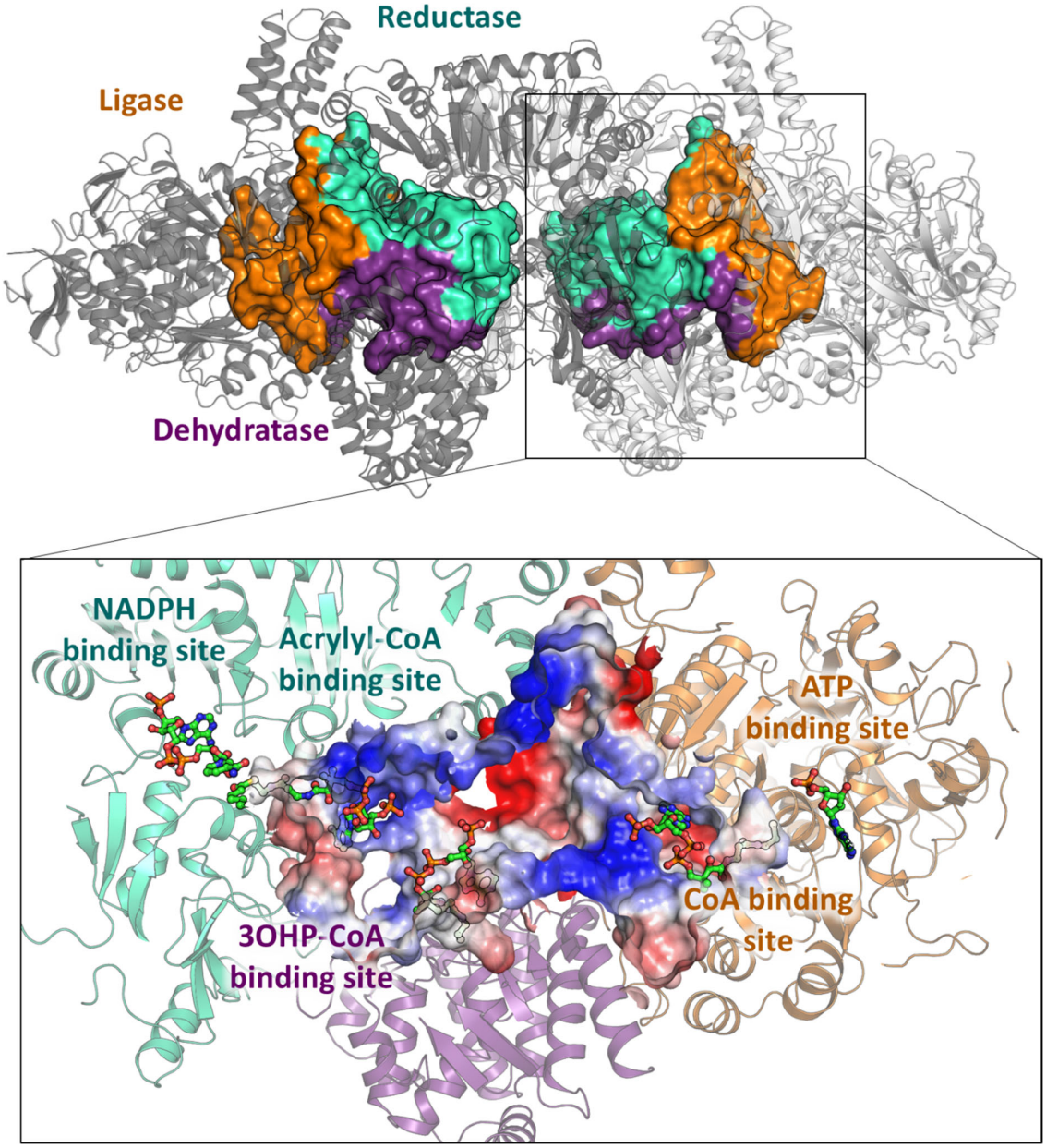


Figure 3. Multi-catalytic reaction chamber of PCS. Volume filling representation[16] of the reaction chambers enclosed by PCS. The central catalytic reaction chamber of each protomer is formed through the contribution of all three domains. Orange, contribution of the ligase domain; purple, contribution of the dehydratase domain; cyan, contribution of the reductase domain. The close up shows a cross section through the reaction chamber. Electrostatic charge distribution is shown as a gradient from red – negatively charged to blue – positively charged. The three active sites are well connected within the reaction chamber. Large positively charged patches may help retaining the CoA-ester intermediates inside during catalysis or even guide them between active sites. Negative charges around the small openings may also prevent leakage of the negatively charged CoA-derivatives. The PCS structure co-crystalized with CoA (no density), an ATP analog and $NADP^+$ is depicted in cartoon showing the ligase domain in orange, the dehydratase domain in purple and the reductase domain in cyan. CoA binding sites have been modelled based on the superposition of the structures of lone-standing CoA ligase (PDB 2P2F) [17], dehydratase (PDB 5JBX) [18] and reductase (PDB 4A0S) [19] onto PCS (compare **Supplementary Figure 6, 7** and **8**). Distances between the active sites have been determined by measuring the distance between sulfur atoms of modelled CoA moieties to be: ligase - dehydratase 42.5 Å, dehydratase – reductase 33.7 Å, ligase – reductase 63.5 Å.

Structural elements in PCS mimic protomer contributions

The active sites of the acyl-CoA ligase, dehydratase and reductase domain of PCS align well with the ones of their corresponding lone-standing homologues (i.e., individual acyl-CoA ligase, dehydratase and reductase enzymes **Supplementary Figure 6**, **7** and **8**). However, these lone-standing enzyme pendants are organized as homo-oligomers that require contributions from neighboring protomers in the oligomeric complex, whereas the individual domains of PCS seem to be organized as functional monomers within the enzyme [17-19]. Notably, in PCS the individual "monomeric" domains carry additional structural elements that apparently compensate (or "mimic") the essential contributions from the missing neighboring protomers in the lone-standing pendants (**Supplementary Figure 9**, **10** and **11**).

As an example, compared to lone-standing dehydratases, the dehydratase domain of PCS features additional helices (**Supplementary Figure 10**) containing two highly conserved residues (F1220 and K1223 [18]) that are involved in stabilization of the CoA-ester. In case of the lone-standing dehydratases these residues usually protrude from one protomeric subunit helix into the active site of a neighboring protomeric subunit of the homo-trimeric complex. Another example is the reductase domain of PCS that carries a structural extension compensating the part of the CoA binding site, which is provided by a neighboring protomeric subunit in lone-standing homo-tetrameric enoyl-CoA reductases/carboxylase homologues (**Supplementary Figure 11**). Although the additional elements in the monomeric PCS domains structurally mimic the missing protomeric contributions from the lone-standing counterparts, they seem not to be related on the primary sequence level, which raises the question about the evolutionary history and significance of these additional structural elements.

In addition to these protomer-mimicking structural elements, the ligase domain of PCS carries structural extensions that are absent in any lone-standing acyl-CoA synthetase homologues (**Supplementary Figure 9**), most prominently, an additional four helix bundle. This four-helix bundle appears to be unique to PCS and is exclusively found in PCS homologues (based on BLASTP analysis). The four helix bundle caps one side of the reaction chamber. Its absence would leave the reaction compartment wide open (**Supplementary Figure 14**). Taken together, PCS is more than just a simple fusion of individual catalytic domains, but contains structural extensions that allow to form a highly organized multi-functional enzyme compartment.

Access to the PCS reaction chamber is controlled

How is the sequence of reactions orchestrated within the compartment? The three enzyme reactions of PCS can be measured individually, when the appropriate substrates and cofactors are provided, demonstrating that all active sites are in principle accessible for their respective substrates [3]. We

determined the kinetic parameters for the overall reaction of PCS as well as for each catalytic domain. While the ligase and the dehydratase domain had apparent turnover frequencies (k_{cat}) comparable to the overall reaction of PCS, the k_{cat} of the reductase domain was almost 30-fold higher (**Supplementary Figure 12**, **Supplementary Table 2** and **3**). This suggests that acrylyl-CoA is immediately consumed upon its formation *in situ*. To study whether externally provided intermediates can access the reaction chamber of PCS during steady-state, we performed an isotopic labeling competition experiment [20,21]. When starting from 3-hydroxypropionate, PCS preferentially catalyzed the overall reaction. Externally added 3-hydroxypropionyl-CoA was unable to enter PCS under these conditions, while some externally added acrylyl-CoA was converted to propionyl-CoA. However, the amount of isotopically labeled propionyl-CoA corresponded well to the amount of acrylyl-CoA expected to be turned over in the pre-steady-state of PCS (i.e., within the first 0.6 s before 3-hydroxypropionyl-CoA is formed by the ligase domain) (**Figure 2d**, **2e** and **Supplementary Table 4**). In other words: despite the high catalytic efficiency of the reductase domain, only a minor amount of externally added acrylyl-CoA was reduced by PCS in the steady-state. These results demonstrated that catalysis in PCS is consecutive, and that internally produced 3-hydroxypropionyl-CoA and acrylyl-CoA are effectively channeled within the enzyme.

Interdomain communication regulates access to the chamber

Apparently, external acrylyl-CoA is prevented from entering the reaction chamber of PCS, indicating that the enzyme assumes a "closed state" during catalysis. Therefore, we wondered, if any of the substrates (CoA, ATP, 3-hydroxypropionate) or products (AMP) would restrict access of exogenous acrylyl-CoA to the reductase domain during steady-state. Indeed, CoA had a strong effect on both the K_M and the k_{cat} of acrylyl-CoA consumption by the reductase domain. Most notably, binding of CoA to the ligase domain directly lowered activity of the reductase domain. We showed that lysine K783 plays a crucial role in conferring this interdomain communication (**Supplementary Figure 12**, **13** and **Supplementary Table 2**). The synchronization of domains in PCS was additionally supported by the fact that NADPH (the co-substrate of the third reaction) had a strong effect on the kinetic parameters of the first reaction. Notably, this effect was independent from an active reductase domain. NADPH had an ever stronger effect on the kinetic parameters of the ligase reaction when the cofactor was added to a dehydratase mutant of PCS (E1027Q variant), where the reductase domain is still functional but not provided with substrate due to the inactive dehydratase domain. These observations suggested that interdomain communication in PCS works in both directions (**Supplementary Figure 12** and **Supplementary Table 2**) and that the individual domains in PCS do not act independently from each other, but catalyze the three-step reaction sequence in a concerted fashion.

Taken together, our experiments demonstrated a functional coupling of the last reaction step in PCS to the first one and *vice versa*. Apparently, PCS undergoes synchronized conformational changes during catalysis allowing substrates and products to enter and leave the reaction chamber. The gatekeeper to the reaction chamber of PCS is presumably the ligase domain. Stand-alone CoA ligases undergo significant conformational changes between "open" and "closed" states during catalysis [17,22]. When we superposed a "closed-state" *Salmonella enterica* ligase (PDB 2P2F) [17] with the PCS ligase domain, the structures aligned almost perfectly (rmsd of 0.932 Å over 441 Cα-atoms) with the exception of above mentioned additional four-helix bundle extension that is only present in PCS (**Supplementary Figure 9**). An "open-state" ligase of *Saccharomyces cerevisiae* (PDB 1RY2) [22] still aligned well to the PCS ligase domain with its N-terminal domain (rmsd of 0.955 Å over 374 Cα-atoms), while the C-terminal domain (~130 residues) appeared to be rotated outwards. Modelling a corresponding conformational change onto PCS resulted in the exposure of a "hole" that would provide access to the interior of an "open state" compartment (**Supplementary Figure 14**). Upon binding of CoA and formation of 3-hydroxypropionyl-CoA the ligase domain would switch back to the closed state, sealing the reaction chamber and sequestering the CoA-bound intermediates.

We performed limited proteolysis on PCS with trypsin in the absence and presence of different substrates and products to directly test for conformational changes of PCS during catalysis (**Supplementary Figure 15**). When incubated in the presence of CoA, which restricts access to the active site of the reductase and forces the enzyme in the "closed" confirmation, PCS was fragmented within 90 min. In contrast, the simultaneous addition of 3-hydroxypropionate, $NADP^+$ and ATP protected PCS from total proteolysis (**Supplementary Figure 16**). Peptide fragment analysis localized the changes in proteolysis to the flexible parts of the ligase domain that get solvent exposed in the "closed" formation and presumably buried in the "open" confirmation of the enzyme. In summary, limited proteolysis confirmed that the enzyme assumed different conformations depending on the presence of different substrates or products. Additional SAXS analyses of PCS in the presence of different substrates supported this conclusion (**Supplementary Table 5**).

2.4 Discussion

PCS, a key enzyme in the 3-hydroxypropionate bi-cycle for CO_2 fixation, catalyzes the three-step reaction sequence from 3-hydroxypropionate to propionyl-CoA. The enclosed reaction chamber, the observed interdomain communication and the proposed conformational changes suggest a highly complex and synchronized catalytic mechanism according to the following model (**Figure 4**). First, the ligase forms 3-hydroxypropionyl-AMP from ATP and 3-hydroxypropionate in the open

conformation. Binding of CoA then closes the reaction chamber leading to the formation of 3-hydroxypropionyl-CoA inside the enzyme. CoA-binding is facilitated by NADPH, which is consistent with the lowered $K_{M,CoA}$ observed upon NADPH addition and might ensure that the enzyme contains all necessary cofactors to catalyze the complete reaction sequence after closing the chamber. The formed 3-hydroxypropionyl-CoA is subsequently dehydrated to acrylyl-CoA within the closed reaction chamber, isolated from the external environment. The final reduction reaction triggers the re-opening of the reaction chamber for product release and prepares the enzyme for the next catalytic cycle. This mechanism fits well with the observation that the k_{cat} of the ligase reaction alone drops significantly in presence of NADPH, indicating that NADPH stabilizes the closed conformation and limits the catalytic rate of the ligase domain.

Our experiments show that the three domains of PCS follow a synchronized reaction mechanism to sequester and channel the toxic intermediate acrylyl-CoA between active sites. Thus, PCS is not the simple fusion product of three individual enzymes, but represents a sophisticated three-dimensional arrangement of three different domains enclosing a central reaction chamber connecting all three active sites. The volume of the reaction chamber of PCS (33 nm^3) is several orders of magnitude smaller than those of known bacterial micro- or nanocompartments. Usually proteinaceous compartments consist of self-assembling shell proteins that encapsulate their enzymes [23-26]. In PCS, these elements are integrated all into one polypeptide, which fulfills both the structural role of forming the reaction chamber, as well as the catalytic role of driving the multi-reaction sequence. This minimizes biosynthetic costs, because no additional proteins are required to build the compartment.

The three-reaction sequence in PCS is orchestrated through interdomain communication and conformational changes. Access to the reaction chamber is catalytically controlled by the ligase domain that is the only entry and exit site. This makes PCS a complex 'nanoreactor' and differentiates the enzyme from canonical channeling enzymes (eg. tryptophane synthase[27,28], amidotransferases[29,30]), which usually connect two active sites through a narrow channel where conformational changes of single residues suffice to gate separate entry and exit sites[29,31,32]. Compared to dimethylglycine oxidase (DMGO) that has also been described to possess an internal cavity and a single entry and exit funnel, PCS appears to be more complex, featuring a reaction chamber of three active sites and a more complex opening and closing mechanism [33,34].

PCS catalyzes three very fundamental chemical reactions in CoA-ester biochemistry and is able to retain free CoA-esters within its reaction chamber. Known CoA-ester enzyme cascades, such as polyketide, fatty acid or HMG-CoA synthases, require covalently attached CoA-intermediates or shared binding sites of the CoA moiety between two active sites to direct intermediates along a

defined multi-reaction sequence[35-37]. Compared to those PCS employs an intriguingly 'simple' design principle to catalyze a consecutive reaction sequence within a controlled environment. The natural example of a minimal self-assembling nanoreactor that is dynamically regulated could serve as a model for the engineering of spatially and temporally controlled reaction sequences [38-43], especially such that proceed via toxic, reactive or unstable intermediates.

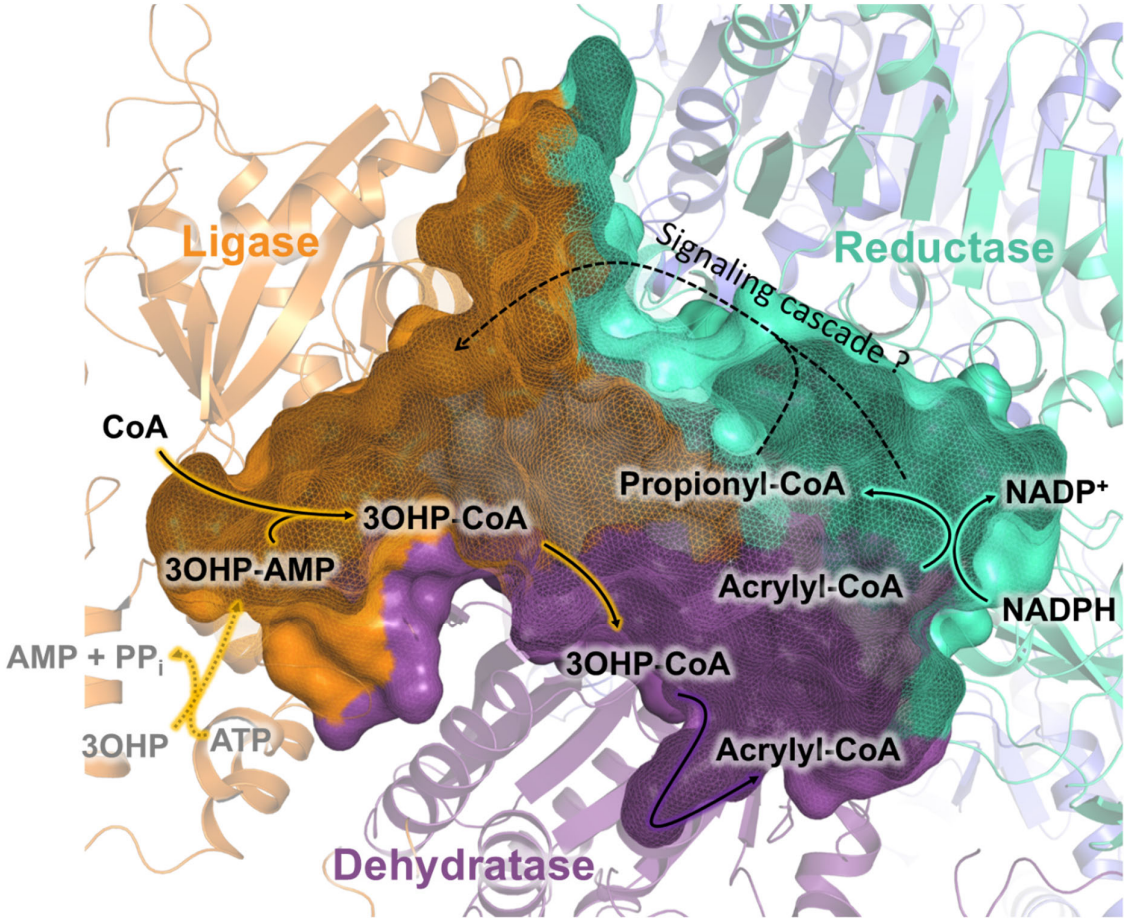


Figure 4. Proposed catalytic cycle of PCS. In the open conformation 3-hydroxypropionate (3OHP) and ATP are converted to 3-hydroxypropionyl-AMP (3OHP-AMP) through the ligase domain (orange). The binding of CoA induces closing of the enzyme and the formation of 3-hydroxypropionyl-CoA (3OHP-CoA). 3OHP-CoA is released into the reaction chamber, where it is converted by the dehydratase domain (purple) to acrylyl-CoA. Acrylyl-CoA then enters the active site of the reductase domain (cyan). Following the reduction of acrylyl-CoA to propionyl-CoA the reaction chamber reopens to release propionyl-CoA, which leaves PCS ready for the next catalytic cycle.

2.5 Methods

Chemicals

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

Synthesis of 3-hydroxypropionyl-CoA, ^{13}C-acrylyl-CoA and ^{13}C-3-hydroxypropionyl-CoA

For the synthesis of unlabeled 3-hydroxypropionyl-CoA a previously described method using carbonyldiimidazole coupling of the precursor acid with coenzyme A was used [44]. Unlabeled acrylyl-CoA was synthesized using a previously described mixed anhydride coupling via ethylchloroformate[44]. ^{13}C-acrylyl-CoA and ^{13}C-3-hydroxypropionyl-CoA were synthesized in two steps from ^{13}C-propionate. In the first step the CDI coupling method was adapted for the synthesis of ^{13}C-propionyl-CoA by a protonation step. ^{13}C-propionate (0.156 mmol, 4.8 eq.) was dissolved into 1 mL THF containing pTsOH (0.156 mmol, 4.8 eq.) for 15 min, the mixture was centrifuged and CDI (0.130 mmol, 4 eq.) added to the supernatant. The mixture was stirred at RT for 1 h. CoA (0.0325 mmol, 1 eq.) dissolved in 250 µL 0.5M $NaHCO_3$ was added and stirred for 1h. The mixture was lyophilized, HPLC purified and again lyophilized. ^{13}C-acrylyl-CoA was synthesized enzymatically using the acyl-CoA oxidase Acx4 from *Arabidopsis thaliana* [45]. A 1 mL assay contained 100 µL 1 M $KHPO_4$ 200 µL 30 mM ^{13}C-propiony-CoA and 600 µL 1mg/mL Acx4. The reaction was quenched after 1 h by adding 20 µL 50% formic acid and directly injected into the HPLC-MS for purification using a previously described purification protocol [44]. In case of ^{13}C-3-hydroxypropionyl-CoA the assay contained additionally 50 µL of the dehydratase PhaJ from *Pseudomonas aeruginosa* for direct hydration of the *in situ* generated acrlyl-CoA.

Bacterial strains and growth conditions

E. coli DH5α (Thermo Scientific™) strains were used for cloning and grown in LB medium [46]. For protein expression *E. coli* BL21-AI™ (Invitrogen) or Arctic-Express (DE3) RIL (Agilent Technologies) were grown in TB medium[47]. Incubation temperature was 37°C. Antibiotics for selection purposes were used accordingly: 100 µg/ml ampicillin, 15 µg/ml gentamycin.

Cloning. All *in silico* cloning was performed with Clone Manager 9 (Scientific & Educational Software). For purification, preparation, cloning, transformation and amplification of DNA, standard protocols were used [48]. Plasmid isolation and PCR product purification was performed with kits from Macherey Nagel (Düren, Germany) according to the manufacturer's protocols.

The gene encoding for PCS (GenBank accession no. EAQ29651) with an N-terminal 10x His tag was synthesized by the DOE Joint Genome Institute. The construct was cloned into the expression backbone pET-16b by restriction cloning, resulting in the plasmid pTE1005. Point mutants were generated by QuickChange® Site-Directed mutagenesis (Stratagene, La Jolla, USA). Following primers were used: forward primer (5'-CGT TTC GGT CAA CCA CAA ATC AAT CTT CGC-3') and reverse primer (5'-GCG AAG ATT GAT TTG TGG TTG ACC GAA ACG-3') for the E1027Q variant; forward primer (5'-CGG AAA TTT TTG GCA CAG CGC TGT GCA ATG CTT ATG AG-3') and reverse primer (5'-CTC ATA AGC ATT GCA CAG CGC TGT GCC AAA AAT TTC CG-3') for the H1769A variant; forward primer (5'-CCT CAC AGC CAG ATG GGT GTA ACT CC-3') and reverse primer (5'-GGA GTT ACA CCC ATC TGG CTG TGA GG-3') for the K783M variant.

Expression and purification of PCS

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

ArcticExpress (DE3) RIL Cpn60 chaperone, an additional wash step was performed with 15 mL removal buffer (50 mM Tris-HCl pH 7.5, 50 mM KCl, 20 mM $MgCl_2$, 5 mM ATP). PCS was eluted with 100% buffer B and applied to a pre-equilibrated HiLoad 16/60 200 pg superdex (GE Life Science) column (150 mM NaCl, 20 mM Tris HCl pH 7.9). The purity of the PCS was tested by SDS-PAGE (**Supplementary Figure 17**).

Kinetic characterization of PCS

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

All reactions were measured by following the consumption of NADPH or NADH (ligase coupling assay) at 340 nm ($\varepsilon_{NAD(P)H}$ = 6.22 mM^{-1} cm^{-1}) or at 365 nm (ε_{NADH} = 3.4 mM^{-1} cm^{-1}). The detailed conditions for all assays can be found in **Supplementary Table 2**.

Crystallization of PCS

All crystallization was performed at 18°C using the sitting drop method in 96-well 2-drop MRC Crystallization Plates in polystyrene (Molecular Dimensions, Suffolk, UK). Crystallization drops (1.4 – 2 µL) contained PCS at 10 mg/ml premixed with 2mM of CoA, $NADP^+$ and phosphomethylphosphonic acid adenylate ester each mixed with reservoir solution in a 1:1 ratio. First thin needle-shaped crystals appeared after several weeks in 100 mM BisTris pH 6.5, 200 mM NaAc,

25 % (w/v) polyethylene glycol (PEG) 3350 supplemented with 3 % (w/v) trimethylamine N-oxide dihydrate as additive (condition 1). These crystals had a *C*2 symmetry and the best resolution obtained was 2.7 Å, which was used for the final structure model. Increasing the additive trimethylamine N-oxide dihydrate to 6 % still lead to crystal formation (condition 2) but exhibited strong twinning. Crystals could also be reproduced in the same condition replacing the additive with 100 mM D-(-)-fructose (condition 3) or 4 % (v/v) tert-butanol (condition 4). Crystals of the space group $P\ 2_1\ 2_1\ 2$ were obtained in the same condition supplemented with 2 % (w/v) benzamidine hydrochloride as additive (condition 5). Phasing was achieved by soaking crystals of condition 3 in 100 mM Xo4 [8] for 4 minutes. All crystals were cryo-protected with the respective crystallization solution supplemented with 20 – 30 % ethylene glycol.

X-ray crystallography analysis

Numerous heavy atom derivatives have been tested attempting to solve the structure of PCS such as: Potassium tetrachloroplatinate(II), organic mercury derivatives, 5-amino-2,4,6-triiodoisophthalic acid and lanthanide phasing compounds (NatX-ray SAS, Saint Martin d'Hères, France). These derivatives resulted in either very low occupancy of the heavy atoms or a significant decrease in diffraction. A recently developed lanthanide complex, Xo4, containing a terbium ion [8] gave the best results with high anomalous signal not interfering with diffraction quality for short time soakings. We solved PCS in the *C*2 crystalline form using the single-wavelength anomalous scattering method. Datasets were collected at beamline Proxima-2A of the SOLEIL synchrotron (Paris, France) at the Tb LIII absorption edge (wavelength of 1.649165 Å) on two different crystals (condition 3, see above) soaked in Tb-Xo4. Merging provided a 3.45 Å resolution data set with high redundancy facilitating the location of 18 terbium sites using the SHELX [49] software and PHASER [50]. After multiple cycles of phasing, electron density modification, and secondary structure building using AUTOSOL from the PHENIX package [51], the electron density quality was sufficient to build a model with Buccaneer from the CCP4 package [50]. The initial model was then used as template for molecular replacement with a dataset of a native crystal (condition 1, see above) using PHASER-MR. The native dataset was collected (wavelength of 0.97625 Å) at beamline ID29 of the ESRF (Grenoble, France). Manual extension of the model was done using COOT [52]. Several rounds of manual and automatic refinements were performed using COOT and PHENIX-Refine. The model (PDB 6EQO) structure was refined with Ramachandran statistics of 94.93 % favored, 4.82 % allowed, and 0.25 % outliers.

Limited proteolysis

PCS at 0.5 mg/mL was forced into a supposedly closed or open state by adding 3 mM CoA or a combination of 3.4 mM ATP, 2 mM 3-hydroxypropionate and 2 mM $NADP^+$, respectively. A zero time sample of 10 µL was taken. Trypsin (Promega, diluted in 25 mM Tris-HCl pH 7.5, 10 mM $CaCl_2$) was added in a protein:protease ratio of 200:1. Samples of 10 µL were taken at different time points. The sample was quenched with 10 µL 4x SDS buffer and heated at 90°C for 10 min. Samples were applied onto an SDS-PAGE gel. For peptide quantification, the limited proteolysis was repeated as described. However, the samples were quenched by adding PMSF protease inhibitor (dissolved in 2-propanol) to a final concentration of 1 mM. The propylation was performed overnight in the dark (at RT) in 100 mM HEPES buffer pH 7.5, 400 mM sodium cyanoborohydride and 5 % (v/v) acetone [53]. The open state PCS samples were treated with unlabeled acetone while the closed state PCS samples were incubated with D6-labeled acetone. The reaction was stopped with 0.07 % TFA. Samples were concentrated and dried in a Speedvac. The open and closed state samples of each time point were combined and purified over a C18 membrane (cut from Empore™ SPE disks). Peptides were eluted with 0.1 % trifluoroacetic acid (TFA) in 50 % acetonitrile (ACN). Samples were dried in the Speedvac. Peptides were resuspended in 50 µl 0.1 % TFA. 1 µl of the peptide sample was mixed with 1 µl solution of 3 g/L alpha-Cyano-4-hydroxycinnamic acid in 80 % ACN (v/v) containing 0.3 % TFA onto a MALDI plate. The dried spots were measured automatically for MS and MSMS in a MALDI TOF/TOF analyzer (Applied Biosystems/MDS Sciex, Framingham, MA, USA) and the 4800 Series Explorer Software.

Time course assays

The time course assays with 0.1 µM PCS contained 0.8 mM CoA, 0.5 mM 3-hydroxypropionate, 0.8 mM ATP, 4 mM $MgCl_2$, 40 mM KCl, 50 mM $KHCO_3$ and 0.3 mM NADPH, if stated, in 100 mM potassium phosphate buffer pH 8. At specific time points 20 µL of the assay were quenched with 20 µL of 50 % formic acid. The samples were centrifuged at 17´000 × *g* and frozen in liquid nitrogen. Samples were immediately thawed before application to hrLC-MS. The time course assay with 10 µM PCS contained 5 mM CoA, 5 mM 3-hydroxypropionate, 5 mM ATP, 6 mM $MgCl_2$, 60 mM KCl and 2 mM NADPH in 100 mM Tris-HCl pH 7.8. At specific time points 10 µL of the assay were quenched with 10 µL acetonitrile and 10 % formic acid and directly injected into the hrLC-MS. Standard curves (0.05 µM to 500 µM) for quantification for 3-hydroxypropionyl-CoA, acrylyl-CoA and propionyl-CoA were prepared in the corresponding buffer conditions and treated in parallel to the samples. Acrylyl-CoA concentrations in PCS samples were too low for UV/Vis quantification, therefore MS based quantification using the extracted ion count of the standard curve the was used.

Acrylyl-CoA could only be detected via MS in the samples containing high amounts of PCS (acrylyl-CoA detection limit was 5nM).

Isotopic labeling competition experiment

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

High resolution LC-MS (hrLC-MS)

3-hydroxypropionyl-Coa, acrylyl-CoA and propionyl-CoA were analyzed using an Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode through a 1290 Infinity UPLC (Agilent Technologies Inc. Santa Clara, CA, USA). Compounds were separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex XB-C18, Phenomenex, Aschaffenburg, Germany) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 (A) and methanol (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 µl/min: 0 min 0% B; 1 min 0% B, 3 min 2.5% B; 9 min 23% B; 14 min 80 %B; 16 min 80%; 17 min 0 % B; 18 min 0 % B.

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

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

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

SAXS analysis

PCS was freshly purified as described above two days before SAXS analysis was performed. The protein was stored on ice until measurements. Gel filtration buffer for dilutions and blank measurements was treated equally. SAXS data were recorded at the European Synchrotron Radiation

Facility (Grenoble, France) on beamline BM29. The protein was up-concentrated at the beamline. If stated, cofactors were added to the concentrated protein at following concentrations: 3-hydroxypropionate, 2mM; CoA, 3 mM; ATP, 3.4 mM. Two-fold dilution series (4 mg/mL to 0.125 mg/ml) were prepared by dilution with gel filtration buffer containing the corresponding cofactors. The different dilutions were measured to investigate sample quality. Sample storage and measurement temperature was set to 20°C. The ESRF BM29 online software was employed for primary data reduction. PrimusQt (version 4.8.1 [54]) was used for data analysis.

2.6 References

1 Wheeldon, I. *et al.* Substrate channelling as an approach to cascade reactions. *Nat. Chem.* **8**, 299-309 (2016).

2 Linster, C. L., Van Schaftingen, E. & Hanson, A. D. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* **9**, 72-80 (2013).

3 Alber, B. E. & Fuchs, G. Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Biol. Chem.* **277**, 12137-12143 (2002).

4 Zarzycki, J., Brecht, V., Müller, M. & Fuchs, G. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *PNAS* **106**, 21317-21322 (2009).

5 Todd, J. D., Curson, A. R. J., Sullivan, M. J., Kirkwood, M. & Johnston, A. W. B. The *Ruegeria pomeroyi acuI* Gene Has a Role in DMSP Catabolism and Resembles *yhdH* of *E. coli* and Other Bacteria in Conferring Resistance to Acrylate. *Plos One* **7**, e35947 (2012).

6 Teufel, R., Kung, J. W., Kockelkorn, D., Alber, B. E. & Fuchs, G. 3-hydroxypropionyl-coenzyme A dehydratase and acryloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the Sulfolobales. *J. Bacteriol.* **191**, 4572-4581 (2009).

7 Berg, I. A., Kockelkorn, D., Buckel, W. & Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. *Science* **318**, 1782-1786 (2007).

8 Engilberge, S. *et al.* Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effect, phasing properties and luminescence. *Chem. Sci.* **8**, 5909-5917 (2017).

9 Lindbladh, C. *et al.* Preparation and kinetic characterization of a fusion protein of yeast mitochondrial citrate synthase and malate dehydrogenase. *Biochemistry* **33**, 11692-11698 (1994).

10 Shatalin, K., Lebreton, S., Rault-Leonardon, M., Vélot, C. & Srere, P. A. Electrostatic channeling of oxaloacetate in a fusion protein of porcine citrate synthase and porcine mitochondrial malate dehydrogenase. *Biochemistry* **38**, 881-889 (1999).

11 Datta, A., Merz, J. M. & Spivey, H. O. Substrate channeling of oxalacetate in solid-state complexes of malate dehydrogenase and citrate synthase. *J. Biol. Chem.* **260**, 15008-15012 (1985).

12 Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse Bacterial Microcompartment Organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438-468 (2014).

13 Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293-1297 (2017).

14 Sutter, M. *et al.* Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat. Struct. Mol. Biol.* **15**, 939-947 (2008).

15 Jung, T. & Grune, T. The proteasome and the degradation of oxidized proteins: part I—structure of proteasomes. *Redox Biol.* **1**, 178-182 (2013).

16 Ho, B. K. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct. Biol.* **8**, 49 (2008).

17 Reger, A. S., Carney, J. M. & Gulick, A. M. Biochemical and Crystallographic Analysis of Substrate Binding and Conformational Changes in Acetyl-CoA Synthetase. *Biochemistry* **46**, 6536-6546 (2007).

18 Bock, T., Reichelt, J., Müller, R. & Blankenfeldt, W. The Structure of LiuC, a 3-Hydroxy-3-Methylglutaconyl CoA Dehydratase Involved in Isovaleryl-CoA Biosynthesis in *Myxococcus xanthus*, Reveals Insights into Specificity and Catalysis. *ChemBioChem* **17**, 1658-1664 (2016).

19 Quade, N., Huo, L., Rachid, S., Heinz, D. W. & Müller, R. Unusual carbon fixation gives rise to diverse polyketide extender units. *Nat. Chem. Biol.* **8**, 117-124 (2012).

20 Spivey, H. O. & Ovadi, J. Substrate channeling. *Methods* **19**, 306-321 (1999).

21 Lyle, S., Ozeran, J. D., Stanczak, J., Westley, J. & Schwartz, N. B. Intermediate Channeling between ATP Sulfurylase and Adenosine 5'-Phosphosulfate Kinase from Rat Chondrosarcoma. *Biochemistry* **33**, 6822-6827 (1994).

22 Jogl, G. & Tong, L. Crystal structure of yeast acetyl-coenzyme A synthetase in complex with AMP. *Biochemistry* **43**, 1425-1431 (2004).

23 Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and Mechanisms of a Protein-Based Organelle in *Escherichia coli*. *Science* **327**, 81-84 (2010).

24 Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat. Microbiol.* **2**, 17029 (2017).

25 Fan, C. G., Cheng, S. Q., Sinha, S. & Bobik, T. A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *PNAS* **109**, 14995-15000 (2012).

26 Kerfeld, C. A. *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936-938 (2005).

27 Pan, P., Woehl, E. & Dunn, M. F. Protein architecture, dynamics and allostery in tryptophan synthase channeling. *Trends Biochem. Sci.* **22**, 22-27 (1997).

28 Hyde, C., Ahmed, S., Padlan, E., Miles, E. W. & Davies, D. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263**, 17857-17871 (1988).

29 Mouilleron, S., Badet-Denisot, M.-A. & Golinelli-Pimpaneau, B. Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. *J. Biol. Chem.* **281**, 4404-4412 (2006).

30 Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M. & Rayment, I. Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry* **36**, 6305-6316 (1997).

31 Singh, H., Arentson, B. W., Becker, D. F. & Tanner, J. J. Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site. *PNAS* **111**, 3389-3394 (2014).

32 Smith, N. E., Vrielink, A., Attwood, P. V. & Corry, B. Biological channeling of a reactive intermediate in the bifunctional enzyme DmpFG. *Biophys. J.* **102**, 868-877 (2012).

33 Leys, D., Basran, J. & Scrutton, N. S. Channelling and formation of 'active'formaldehyde in dimethylglycine oxidase. *EMBO J.* **22**, 4038-4048 (2003).

34 Tralau, T. *et al.* An internal reaction chamber in dimethylglycine oxidase provides efficient protection from exposure to toxic formaldehyde. *J. Biol. Chem.* **284**, 17826-17834 (2009).

35 Ishikawa, M., Tsuchiya, D., Oyama, T., Tsunaka, Y. & Morikawa, K. Structural basis for channelling mechanism of a fatty acid β-oxidation multienzyme complex. *EMBO J.* **23**, 2745-2754 (2004).

36 Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041-1072 (2007).

37 Vögeli, B. *et al.* Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *PNAS*, 3380-3385 (2018).

38 Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389-394 (2016).

39 Aussignargues, C. *et al.* Structure and Function of a Bacterial Microcompartment Shell Protein Engineered to Bind a [4Fe-4S] Cluster. *J. Am. Chem. Soc.* **138**, 5262-5270 (2016).

40 Giessen, T. W. & Silver, P. A. A Catalytic Nanoreactor Based on *in Vivo* Encapsulation of Multiple Enzymes in an Engineered Protein Nanocompartment. *ChemBioChem* **17**, 1931-1935 (2016).

41 Azuma, Y., Zschoche, R., Tinzl, M. & Hilvert, D. Quantitative Packaging of Active Enzymes into a Protein Cage. *Angew. Chem. Int. Ed.* **55**, 1531-1534 (2016).

42 Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely *de novo* protein framework. *Nat. Chem.* **8**, 837-844 (2016).

43 Brasch, M. *et al.* Assembling Enzymatic Cascade Pathways inside Virus-Based Nanocages Using Dual-Tasking Nucleic Acid Tags. *J. Am. Chem. Soc.* **139**, 1512-1519 (2017).

44 Peter, D. M., Vögeli, B., Cortina, N. S. & Erb, T. J. A chemo-enzymatic road map to the synthesis of CoA esters. *Molecules* **21**, 517 (2016).

45 Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide *in vitro*. *Science* **354**, 900-904 (2016).

46 Bertani, G. STUDIES ON LYSOGENESIS I.: The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293 (1951).

47 Tartof, K. & Hobbs, C. Improved media for growing plasmid and cosmid clones. *Focus* **9**, 12 (1987).

48 Sambrook, J. & Russel, D. *Molecular cloning: a laboratory manual*. 3 edn, (Cold Spring Laboratory Press, New York, 2001).

49 Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122 (2008).

50 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235-242 (2011).

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

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

53 Eichacker, L., Granvogl, B. & Gruber, P. Method for quantitative comparison of two or more proteins. Germany patent EP1947461 (2008).

54 Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277-1282 (2003).

55 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307-1320 (2008).

56 Sudhir Kumar, G. S., and Koichiro Tamura. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Mol. Biol. Evol.* (2015).

57 Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791 (1985).

2.7 Supplementary Information

Supplementary Tables

Supplementary Table 1. X-ray diffraction data collection and model refinement statistics

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

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

Supplementary Table 2. Kinetic parameters of the reactions of propionyl-CoA synthase of *Erythrobacter* sp. NAP1.

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

Data are shown as the mean ± 95% confidence intervals, as determined from a non-linear fit of at least 18 data points. Michaelis-Menten graphs of the original data are shown in **Supplementary Figure 11**. *NA*, not applicable. For detailed reaction compositions, see **Supplementary Table 3**.

Supplementary Table 3. Composition of enzyme assays to determine kinetic parameters of PCS.

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

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

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

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

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

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

Supplementary Table 4. Isotopic labeling experiments.

WT control			
	% of 3OHP-CoA	% of acrylyl-CoA	% of propionyl-CoA
unlabeled	100	ND	100
labeled	0	ND	0
WT + 100µM ^{13}C-3-hydroxypropionyl-CoA			
	% of 3OHP-CoA	% of acrylyl-CoA	% of propionyl-CoA
unlabeled	39 ± 9	ND	99.2 ± 0.3
labeled	61 ± 9	ND	0.8 ± 0.3
WT + 100µM ^{13}C-acrylyl-CoA			
	% of 3OHP-CoA	% of acrylyl-CoA	% of propionyl-CoA
unlabeled	96.4 ± 0.9	7 ± 0.5	79 ± 15
labeled	3.6 ± 0.9	93 ± 0.5	21 ± 15

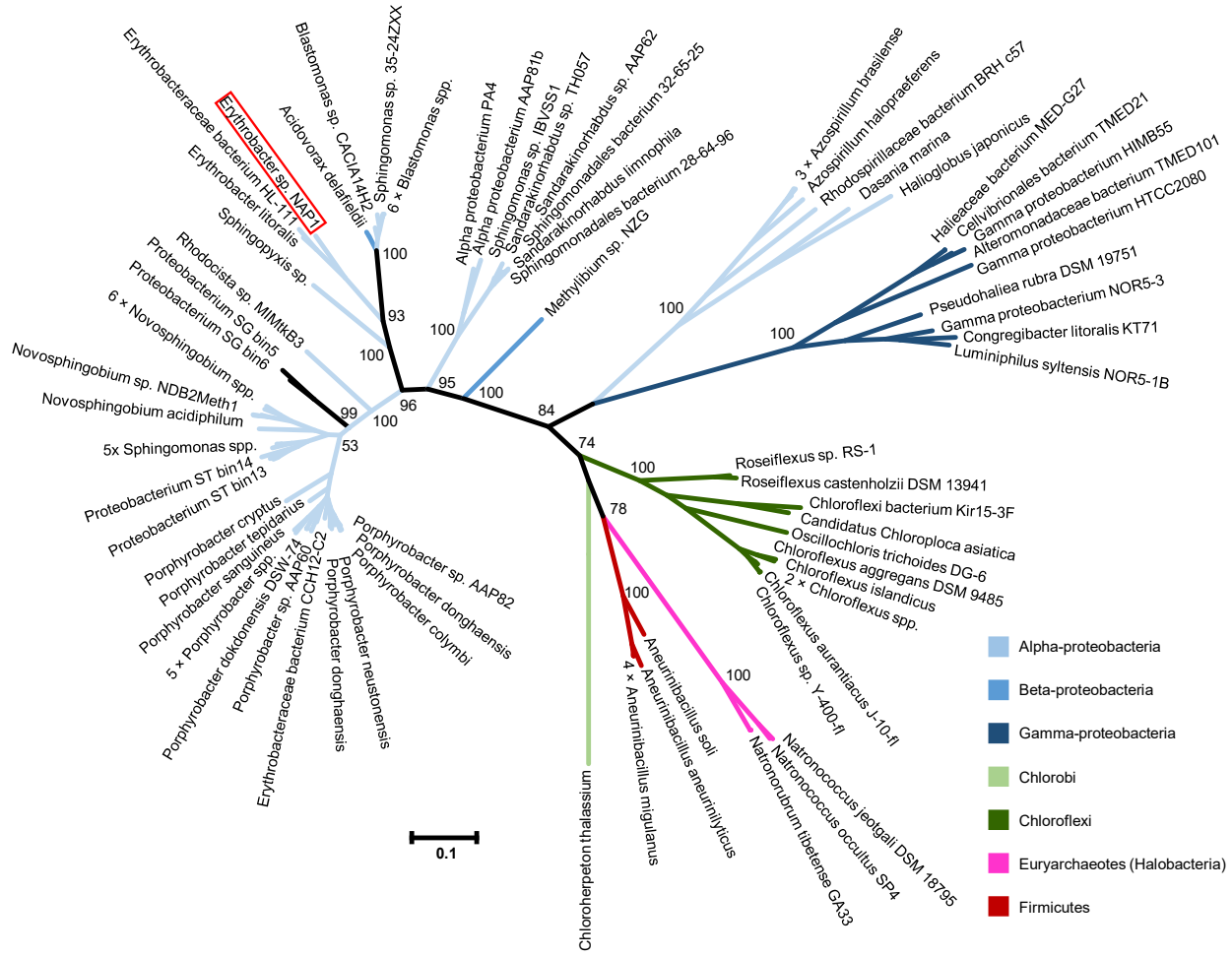
100µL assays contained 2 mM 3-hydroxypropionate, 5 mM ATP, 3 mM CoA, 200 µM NADPH, and labeled intermediate as indicated. Reactions were started by adding 0.1 µM of PCS. Consumption of NADPH (last reaction step forming propionyl-CoA) was monitored at 340 nm. Reactions were stopped by adding 20 µL of 50% formic acid after 20% of NADPH was consumed. Fractions of labeled and unlabeled 3-hydroxypropionyl-CoA (3OHP-CoA), acrylyl-CoA and propionyl-CoA were determined by hrLC-MS. All assays were performed in triplicates. Errors are given as standard deviation. ND, not detectable.

Supplementary Table 5. Small angle X-ray scattering (SAXS) data.

PCS control		
PCS concentration (mg/mL)	I_0 (cm^{-1})	R_g (nm)
4	288.95 ± 0.55	6.35 ± 0.21
2	282.72 ± 0.54	6.53 ± 0.22
1	284.30 ± 0.78	6.47 ± 0.21
PCS + CoA		
PCS concentration (mg/mL)	I_0 (cm^{-1})	R_g (nm)
4	322.45 ± 0.58	6.30 ± 0.34
2	305.86 ± 0.51	6.56 ± 0.20
1	309.86 ± 0.62	6.86 ± 0.28
PCS + ATP + 3OHP		
PCS concentration (mg/mL)	I_0 (cm^{-1})	R_g (nm)
4	406.75 ± 1.65	7.91 ± 2.30
2	393.42 ± 1.46	8.00 ± 2.30
1	371.30 ± 1.67	7.86 ± 2.03

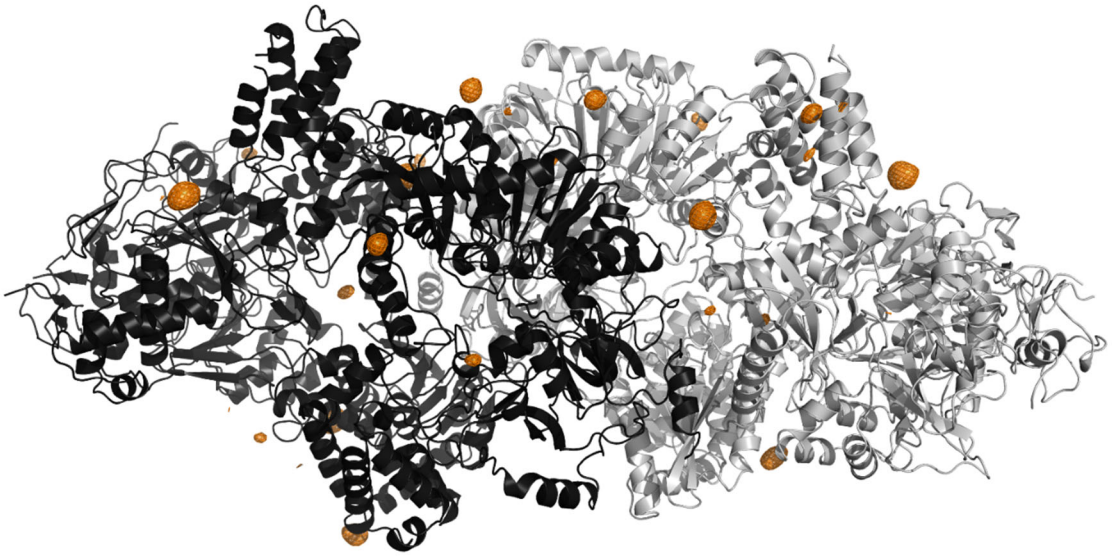
SAXS data were recorded for PCS in the presence or absence (PCS control) of any cofactors/substrates. If stated, cofactors were added to the concentrated protein at following concentrations: 3-hydroxypropionate (3OHP), 2mM; CoA, 3 mM; ATP, 3.4 mM; $NADP^+$, 2 mM. Sample and measurement temperature was 20°C. The ESRF BM29 online software was employed for primary data reduction. PrimusQt (version 4.8.1 [54]) was used for data analysis. Standard deviation as provided by PrimusQT for individual datasets. The changes of the radii of gyration (Rg) demonstrate that ATP and 3OHP induce a global relaxation of PCS, while the structure of apo PCS or PCS in presence of CoA is more compact.

Supplementary Figures

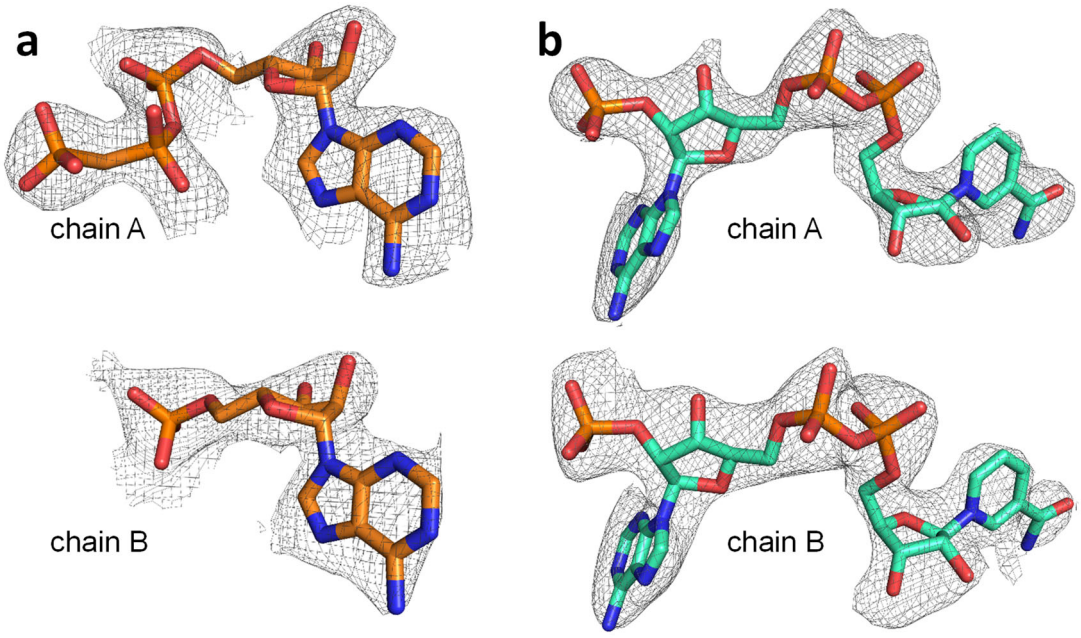


Supplementary Figure 1. Maximum likelihood phylogenetic tree of PCS homologues. The evolutionary history was inferred based on the Le-Gascuel model [55]. The scale of the branch lengths is measured in number of substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 [56]. Bootstrap values [57] for confidence limits are given at important nodes. The position of the PCS homologue of *Eryhrobacter sp*. NAP1 is highlighted with a red box.

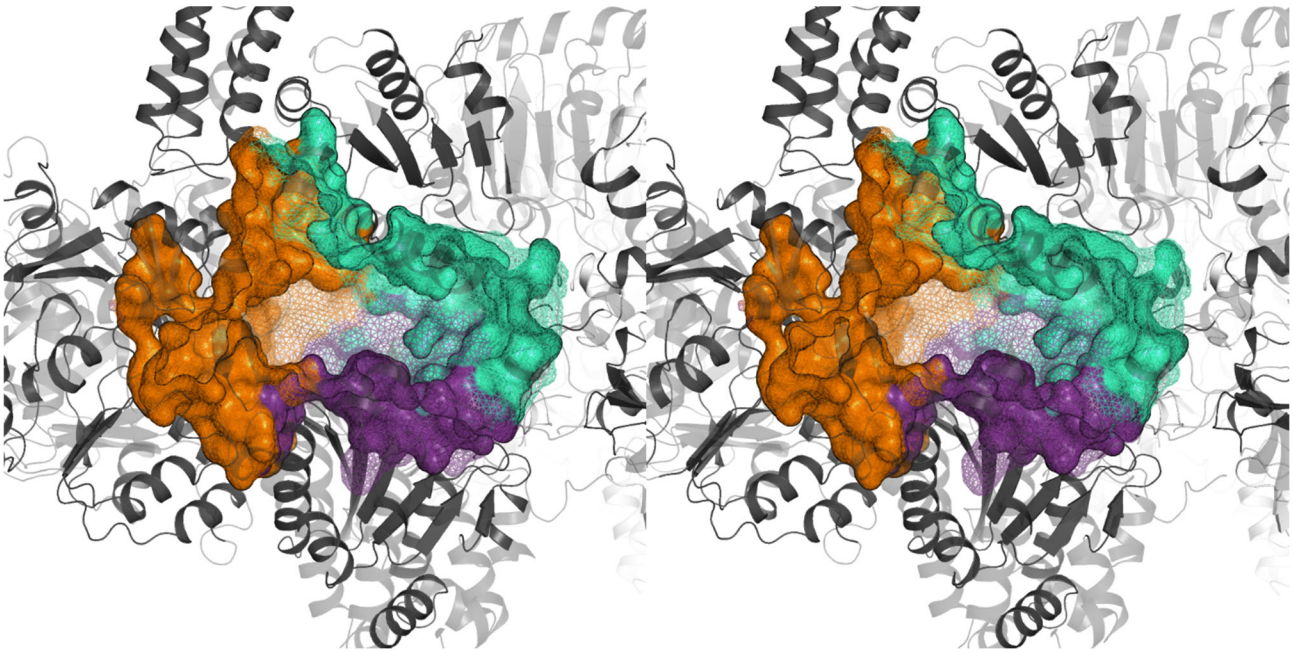
Supplementary Figure 2. PCS domain distribution. Sequence alignment of *Erythrobacter* sp. NAP1 PCS (PDB 6EQO) with its closest lone-standing homologues; the acetyl-CoA ligase of *Salmonella enterica* (PDB 2P2F) [17], the HMG-CoA dehydratase of *Myxococcus xanthus* (PDB 5JBX) [18] and the enoyl-CoA carboxylase/reductase of *Streptomyces* sp. JS360 (PDB 4A0S) [19]. The secondary structure of PCS is represented above the sequence and colored according to their domain contribution; orange - ligase domain, purple – dehydratase domain, cyan - reductase domain.



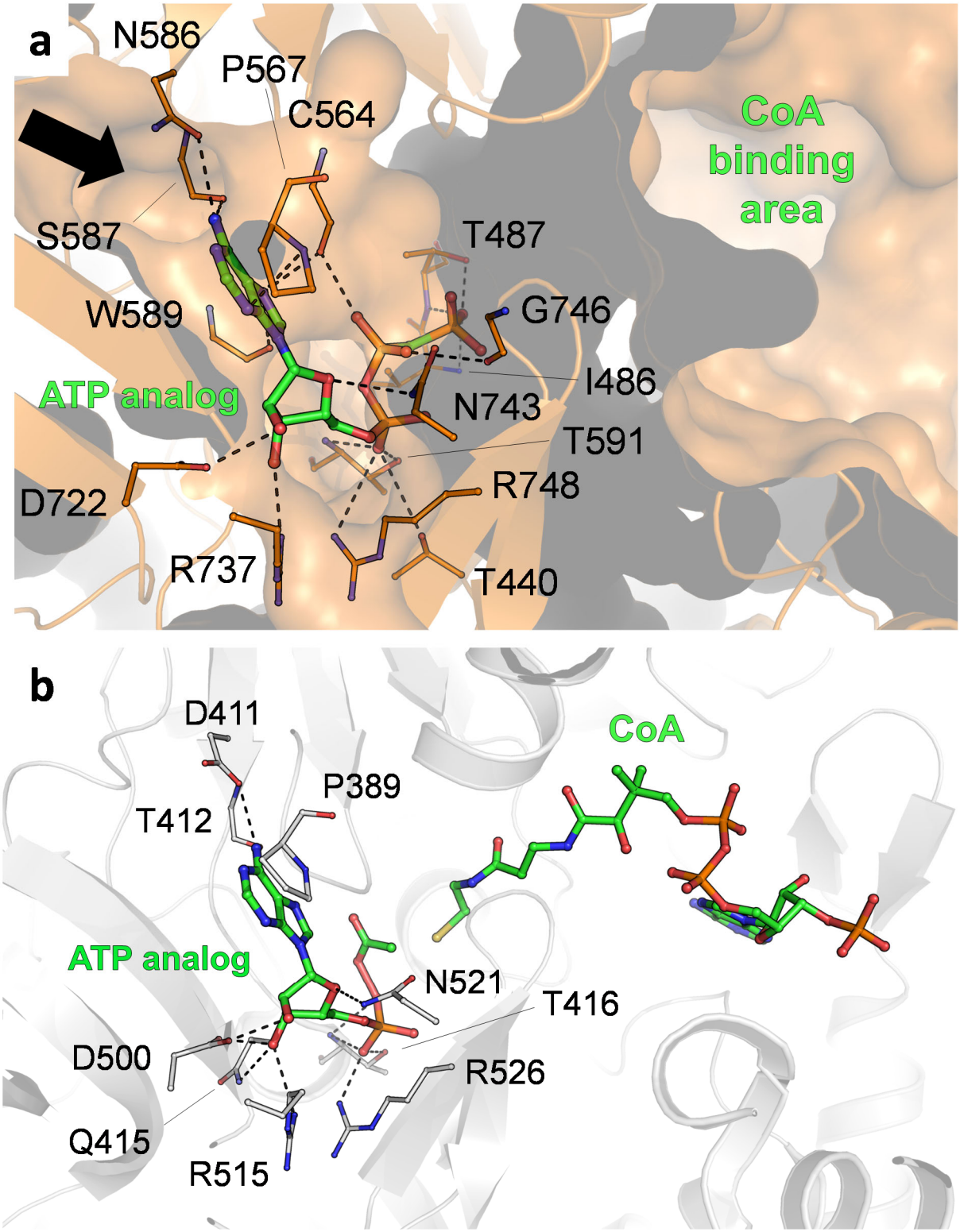
Supplementary Figure 3. Anomalous map contoured at 6 σ indicating the positions (in orange) of terbium contained in the new phasing compound Tb-Xo4 [8].



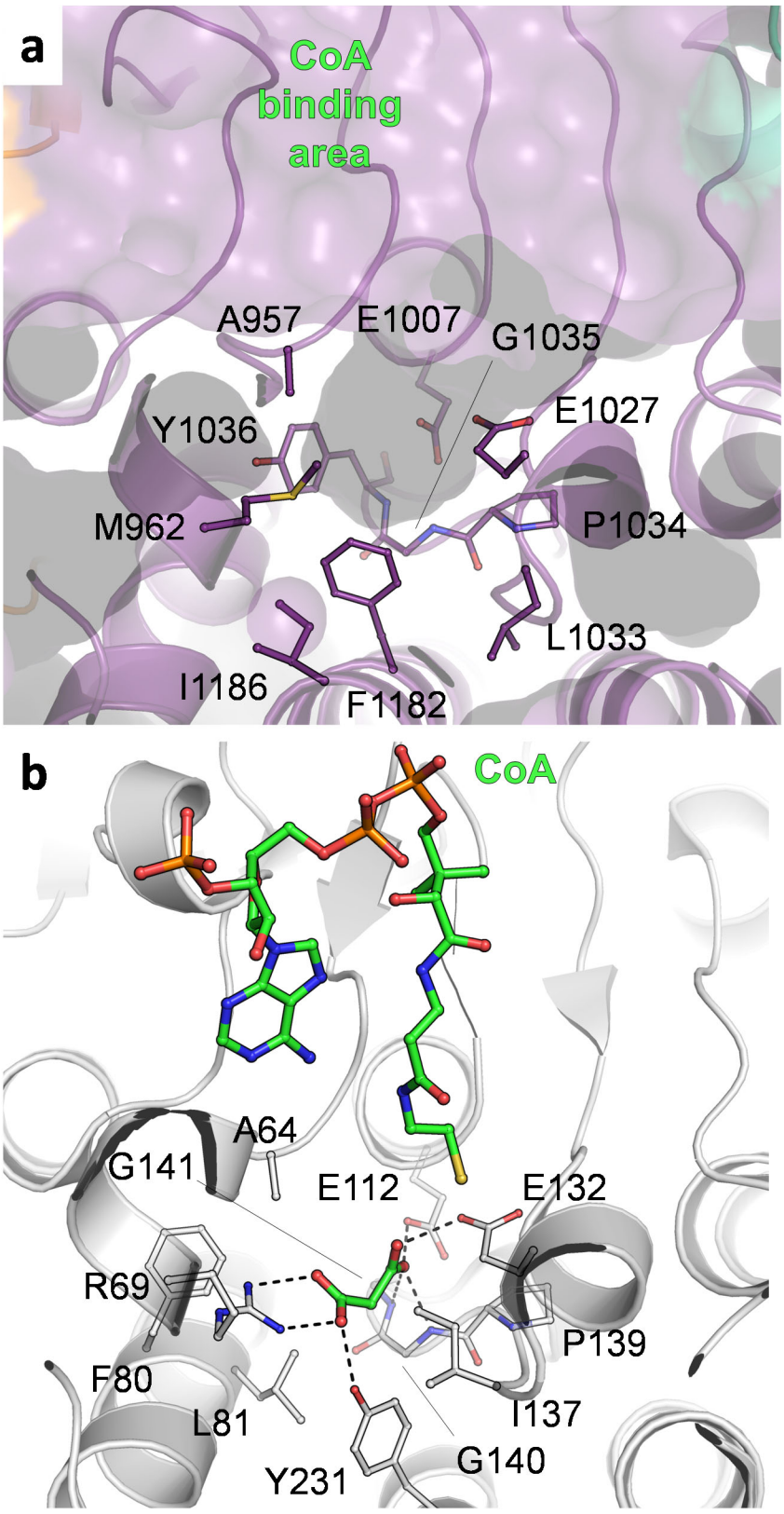
Supplementary Figure 4. Superpositions of Fo-Fc electron density simulated annealing omit maps on refined ligands. a, Omit maps at 1.5 σ for the refined phosphomethylphosphonic acid adenylate ester (ACP) in both of the polypeptide chains of the dimeric PCS model (PDB 6EQO). **b**, Omit maps at 3.0 σ for the refined $NADP^+$ (NAP) in both of the polypeptide chains of the dimeric PCS model (PDB 6EQO).



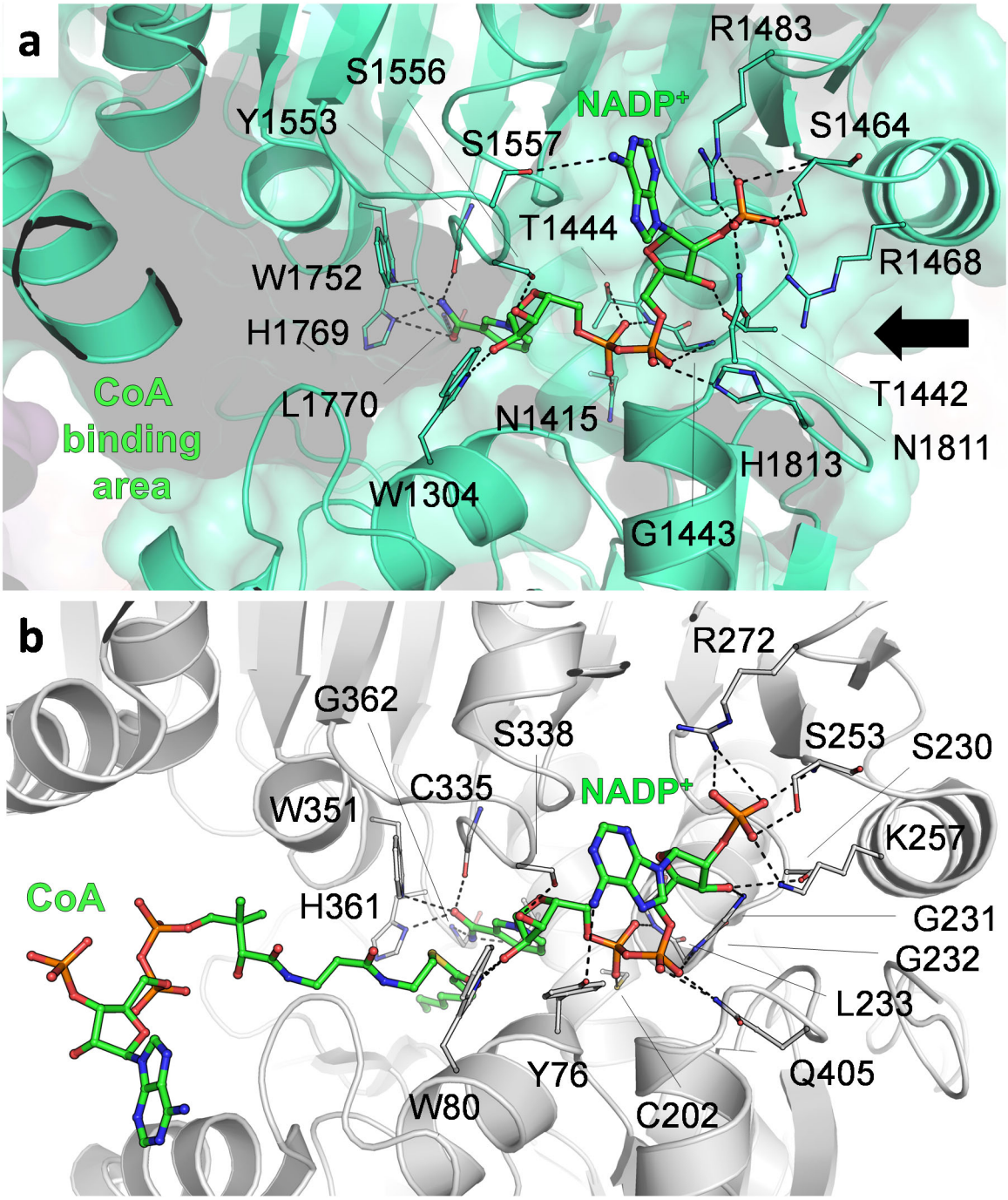
Supplementary Figure 5. Stereodiagram of the PCS reaction chamber. The central catalytic reaction chamber is formed through the contribution of all three domains. Orange, contribution of the ligase domain; purple, contribution of the dehydratase domain; cyan, contribution of the reductase domain.



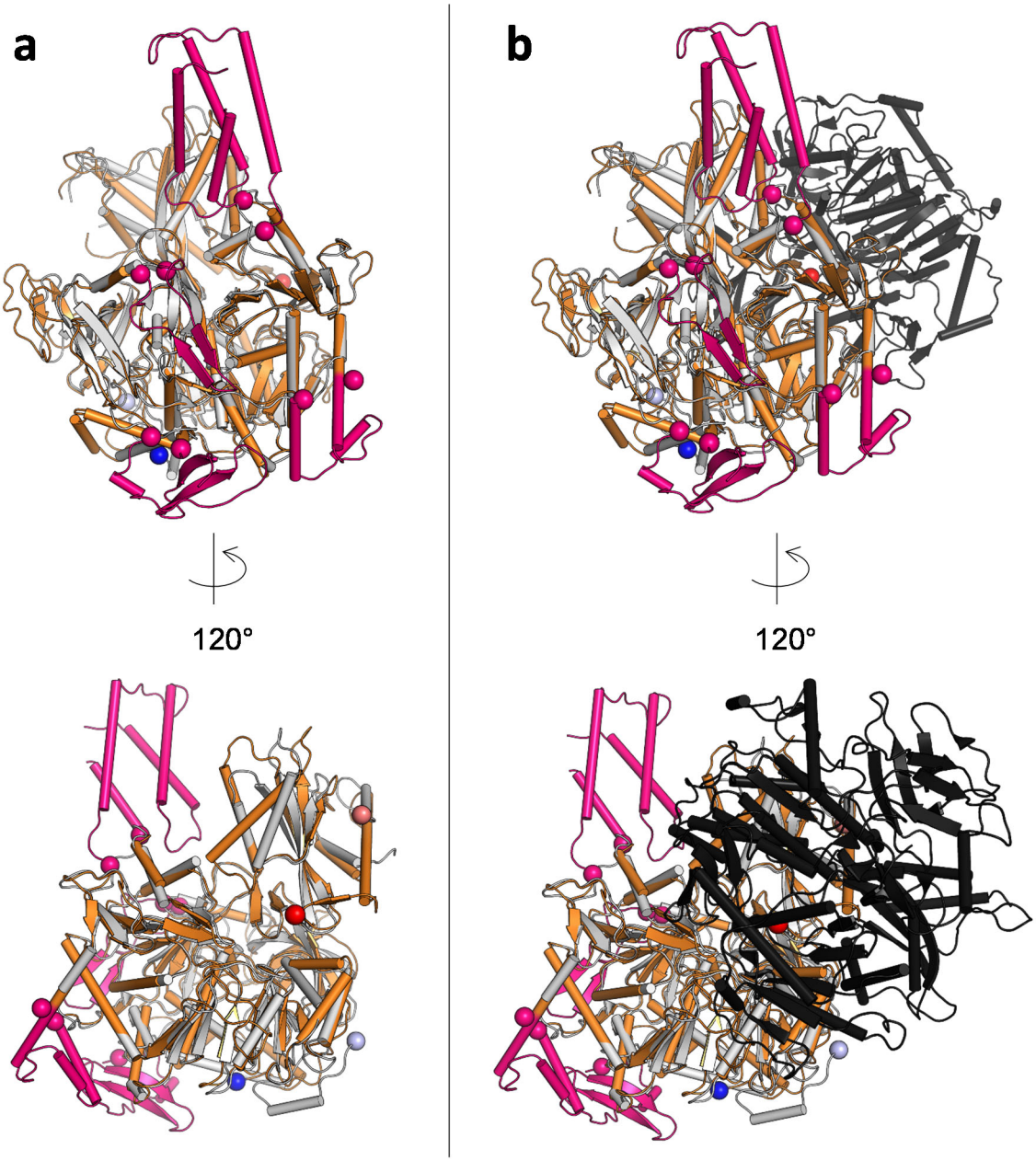
Supplementary Figure 6. Close-ups of the active sites of the ligase domain of PCS and a close homologue. a, Active site of the PCS ligase domain co-crystalized with an ATP analog (green) and with the predicted CoA binding site indicated (green label). The ATP entry site is indicated by a black arrow. **b**, Active site of a lone-standing CoA ligase of *Salmonella enterica* (PDB 2P2F) [17] co-crystalized with an ATP analog and CoA (green).



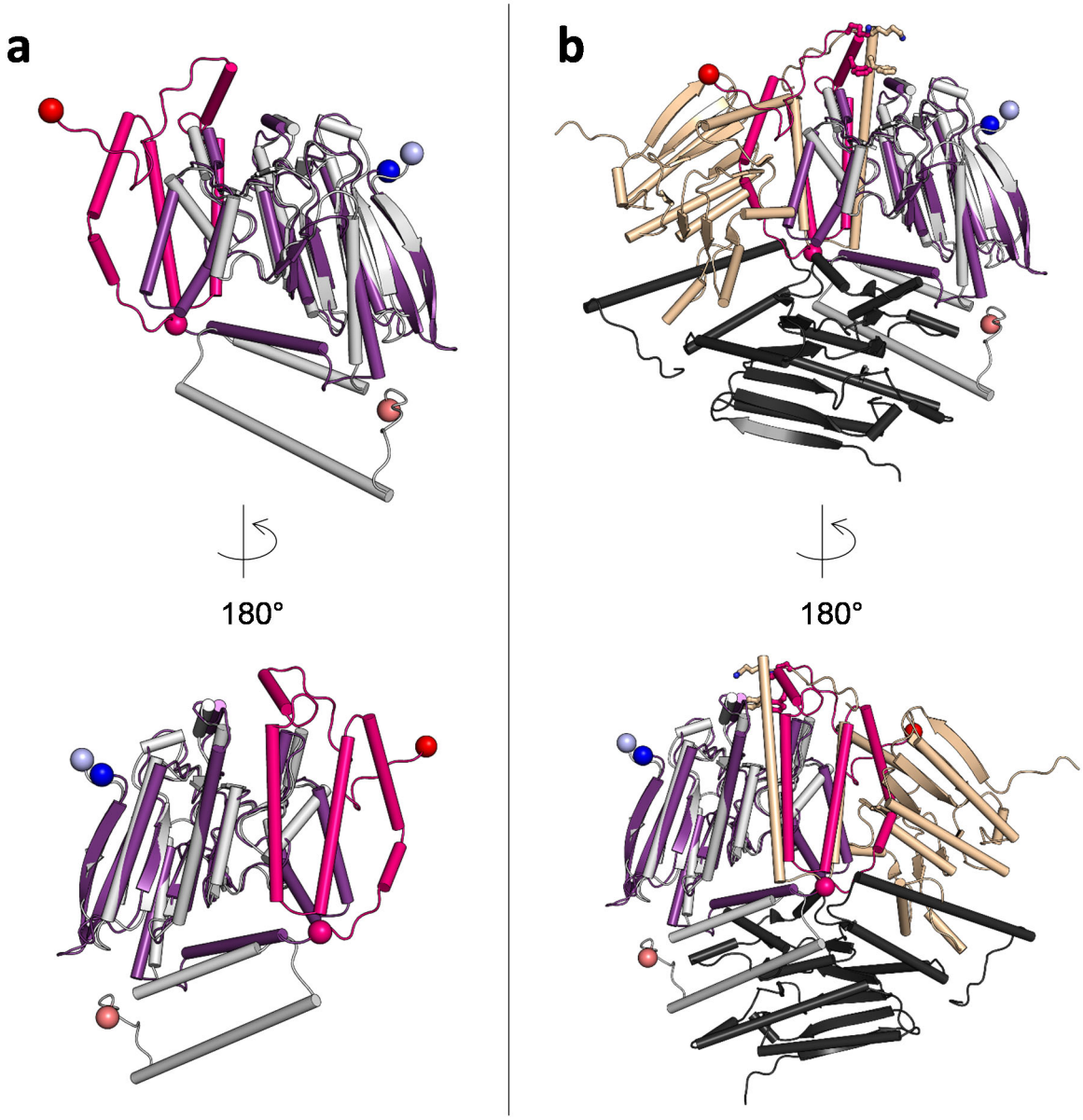
Supplementary Figure 7. Close-ups of the active sites of the dehydratase domain of PCS and a close homologue. a, Active site of the PCS dehydratase domain with the essential active site glutamate E1027. **b**, Active site of a lone-standing dehydratase of *Myxococcus xanthus* co-crystalized with CoA and malonic acid (green) (PDB 5JBX) [18] with the essential active site glutamate E132.



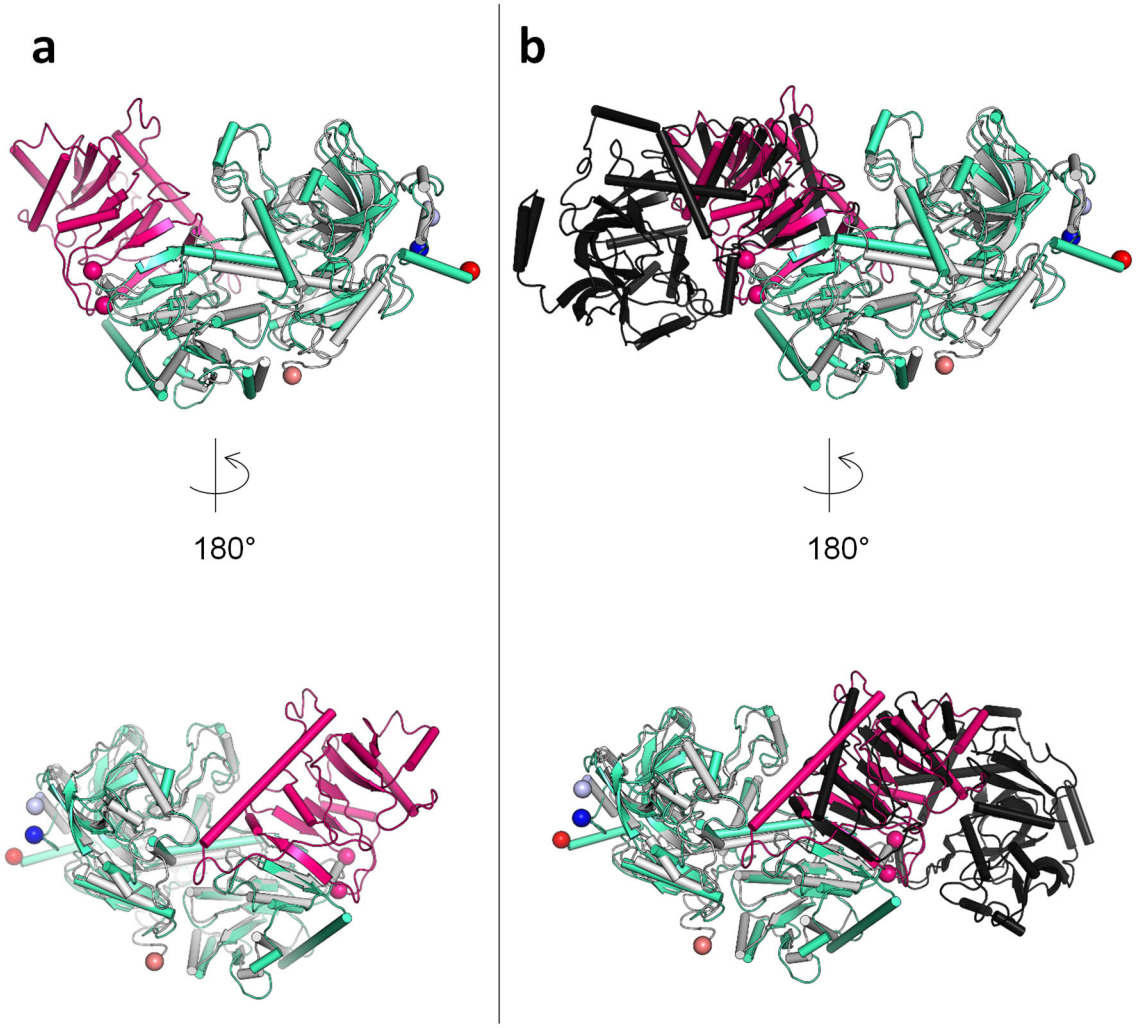
Supplementary Figure 8. Close-ups of the active sites of the reductase domain of PCS and a close homologue. a, Active site of the PCS reductase domain with essential active site histidine H1769 co-crystalized with $NADP^+$ (green). The $NADP^+$ entry site is indicated by a black arrow. **b**, Active site of a lone-standing enoyl-CoA carboxylase/reductase of *Streptomyces* sp. JS360 co-crystalized with octenoyl-CoA and $NADP^+$ (green) (PDB 4A0S) [19] with the essential active site histidine H361.



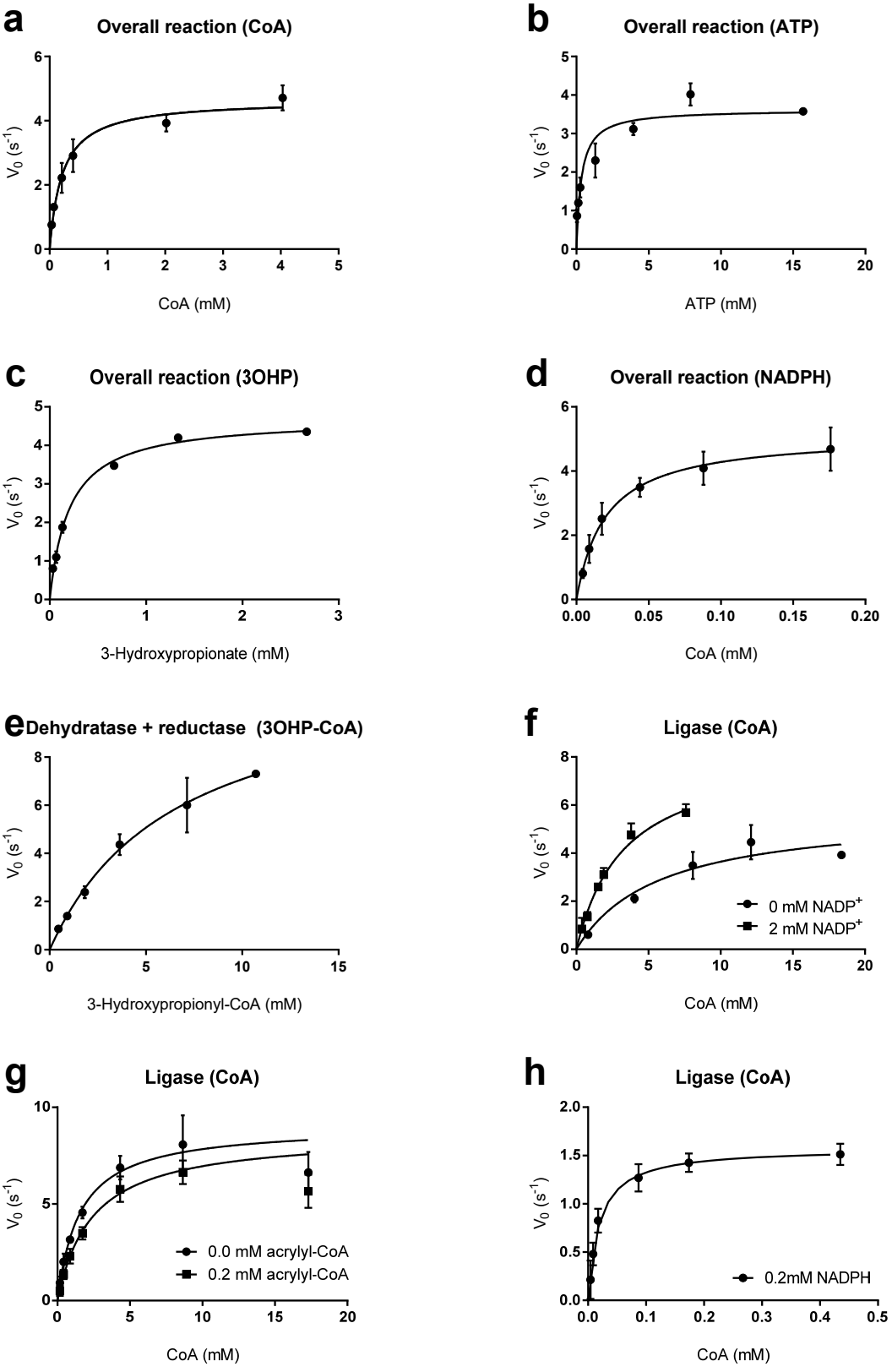
Supplementary Figure 9. Superposition of the ligase domain with a close homologue, the acetyl-CoA synthetase of Salmonella enterica (PDB 2P2F) [17]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. **a**, Superposition of a PCS ligase domain (orange) with the monomer of the homologue (gray, rmsd of 0.930 Å over 447 Cα-atoms). Additional structural elements of the PCS monomer are shown in pink. **b**, Superposition of a PCS ligase domain with the dimer of the homologue (gray and black). The peripheral extensions of the PCS ligase domain are not aligning with the second chain of the dimeric homologue. The largest additional structural element is a four helix bundle on top of the structure.

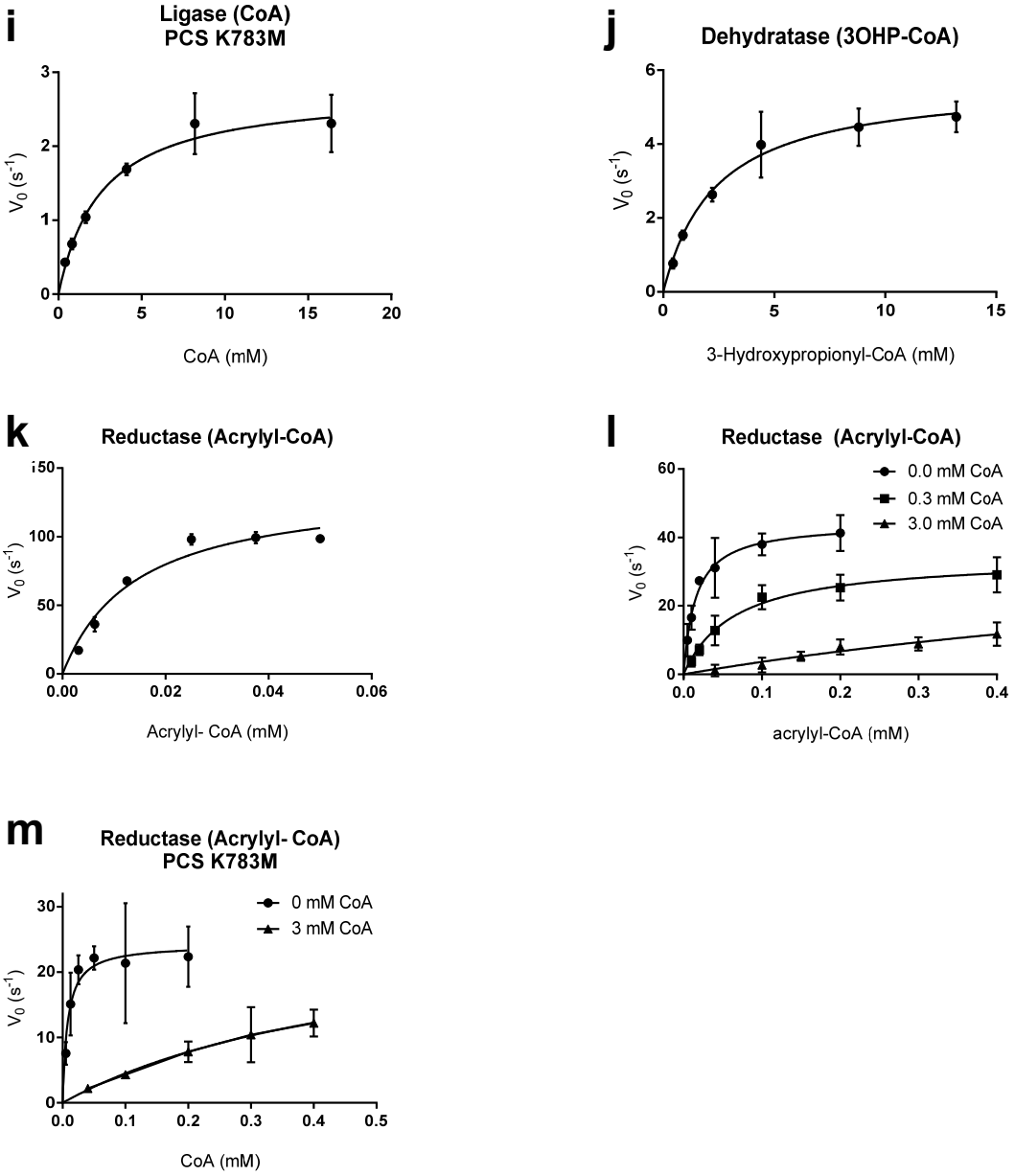


Supplementary Figure 10. Superposition of the PCS dehydratase domain with a close homologue, the HMG-CoA dehydratase of Myxococcus xanthus (PDB 5JBX) [18]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. **a**, Superposition of a PCS dehydratase domain (purple) with the monomer of the homologue (gray, rmsd of 1.085 Á over 147 Cα-atoms). Additional structural elements of PCS dehydratase domain are shown in pink. **b**, Superposition of a PCS dehydratase domain with the trimeric homologue (gray, sand and black). The additional secondary structure elements in the PCS dehydratase domain partly mimic the neighboring chain of the trimeric homologue.

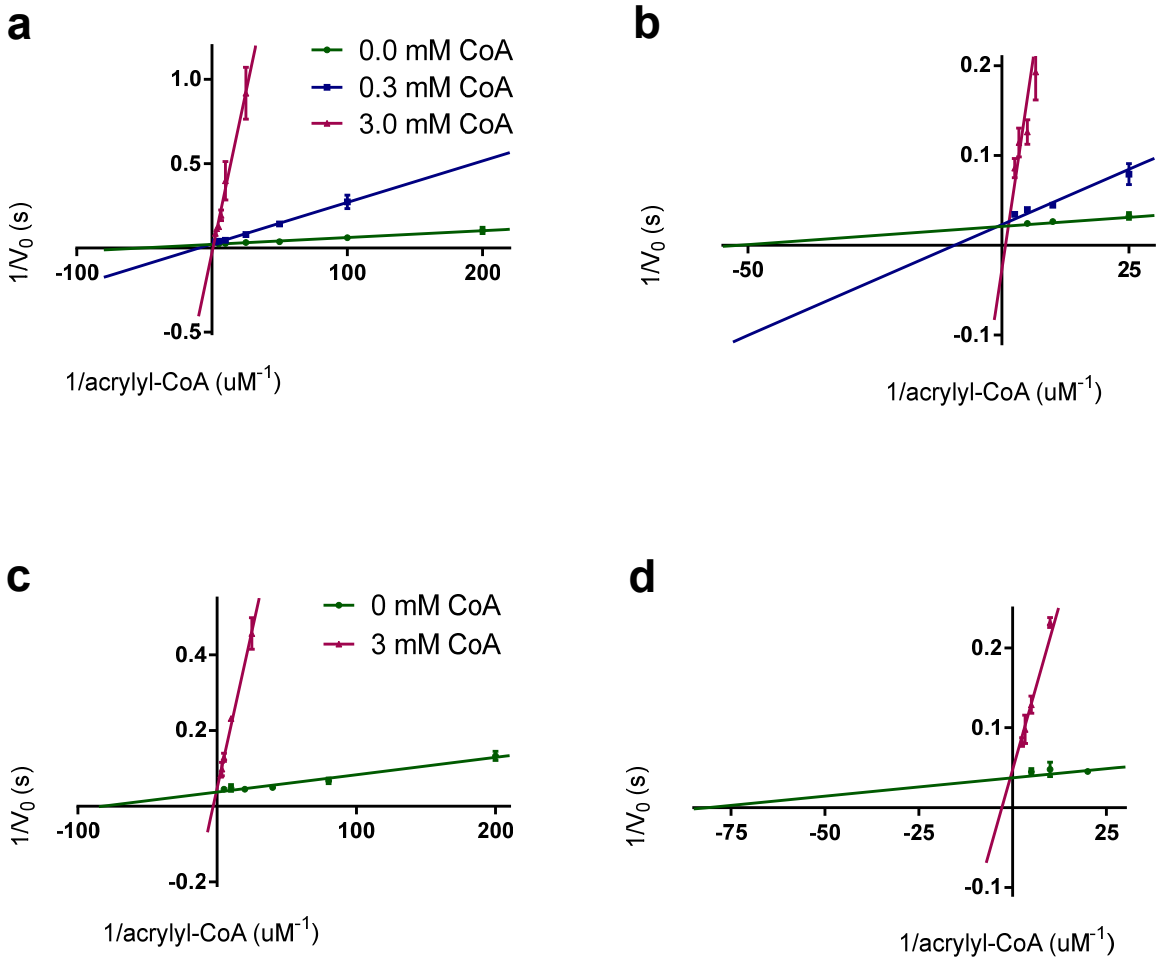


Supplementary Figure 11. Superposition of the PCS reductase domain with a close homologue, the enoyl-CoA carboxylase/reductase of Streptomyces sp. JS360 (PDB 4A0S) [19]. The blue and red spheres corresponds to the N-termini and C-termini of PCS or the homologue, respectively. The pink spheres mark the insertion points of the additional secondary structure elements in PCS. **a**, Superposition of a PCS reductase domain (cyan) with the monomer of the homologue (gray, rmsd of 1.530 Å over 291 Cα-atoms). Additional structural elements of PCS reductase domain are shown in pink. **b**, Superposition of a PCS reductase domain with the dimer of the homologue (gray and black). The additional secondary structure elements in the PCS reductase domain partly mimic the neighboring chain of the homologue dimer.

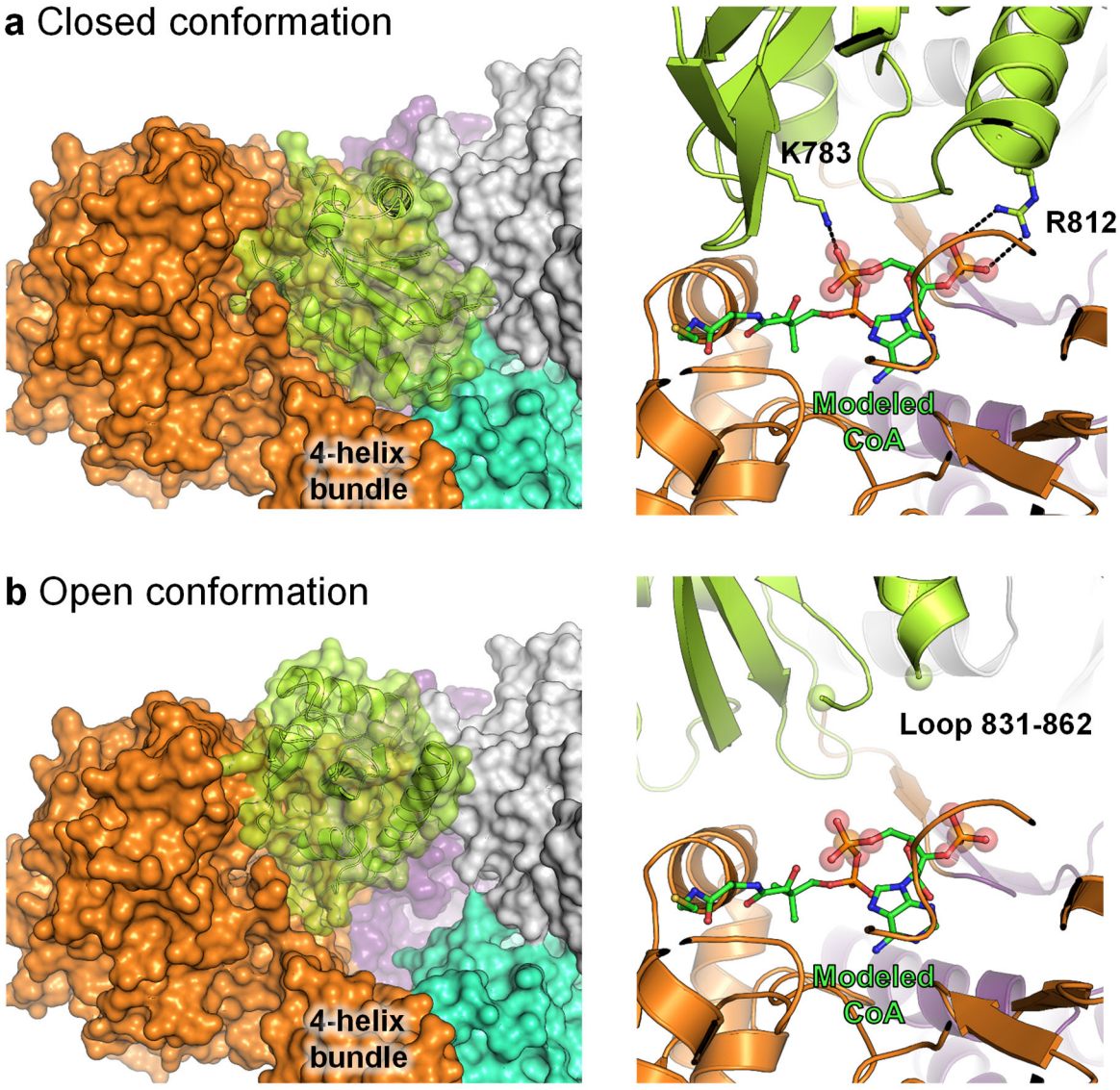

a
Overall reaction (CoA)
V0 (s-1)
CoA (mM)
b
Overall reaction (ATP)
V0 (s-1)
ATP (mM)
c
Overall reaction (3OHP)
V0 (s-1)
3-Hydroxypropionate (mM)
d
Overall reaction (NADPH)
V0 (s-1)
CoA (mM)
e
Dehydratase + reductase (3OHP-CoA)
V0 (s-1)
3-Hydroxypropionyl-CoA (mM)
f
Ligase (CoA)
V0 (s-1)
0 mM NADP+
2 mM NADP+
CoA (mM)
g
Ligase (CoA)
V0 (s-1)
0.0 mM acrylyl-CoA
0.2 mM acrylyl-CoA
CoA (mM)
h
Ligase (CoA)
V0 (s-1)
0.2mM NADPH
CoA (mM)



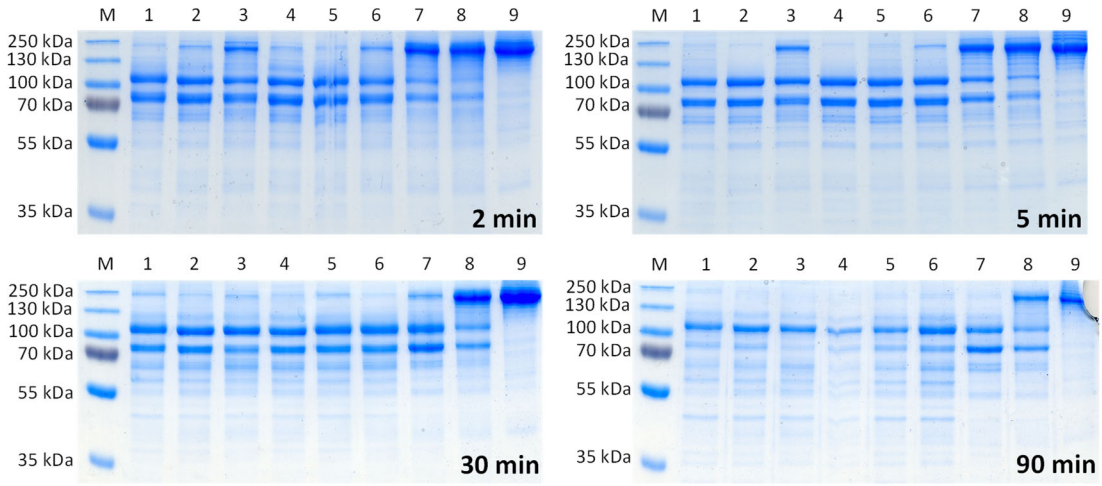
Supplementary Figure 12. Michaelis-Menten plots of the kinetic characterizations of the reactions of PCS from *Erythrobacter* sp. NAP1. All points were measured in triplicates and the 95% confidence intervals are indicated. For kinetic parameters see **Supplementary Table 2**, for assay compositions see **Supplementary Table 3**. **a**, Overall reaction of PCS dependent on CoA concentration. The curve was fitted excluding values above 5 mM CoA, which resulted in substrate inhibition. **b**, Overall reaction of PCS in dependence of ATP concentration. Substrate inhibition was observed at 10 mM ATP. **c**, Overall reaction of PCS in dependence of 3-hydroxypropionate concentration. **d**, Overall reaction of PCS in dependence of NADPH concentration. **e**, The coupled dehydratase and reductase reaction depending on the 3-hydroxypropionyl-CoA concentration. **f**, The ligase reaction alone depending on the CoA concentration in presence or absence of $NADP^+$. **g**, The ligase reaction alone using the PCS E1027Q variant deficient in the dehydratase reaction to test the effect of acrylyl-CoA. Substrate inhibition can be observed above 10 mM ATP. The curve was fitted including those values. **h**, The ligase reaction alone using the PCS E1027Q variant deficient in the dehydratase reaction to test the effect of NADPH. **i**, The ligase reaction alone using the PCS E1027Q K783M variant depending on the CoA concentration. **j**, The dehydratase reaction alone using the H1769A variant deficient in the reductase activity to allow coupling to the stand-alone reductase Etr1p (from *Saccharomyces cerevisiae*) to monitor acrylyl-CoA formation. **k**, The reductase reaction alone depending on the acrylyl-CoA concentration. **l**, The reductase reaction alone using the PCS E1027Q variant depending on the acrylyl-CoA concentration in presence of different CoA concentrations. **m**, The reductase reaction alone using the PCS E1027Q K783M variant depending on the acrylyl-CoA concentration in presence of different CoA concentrations.



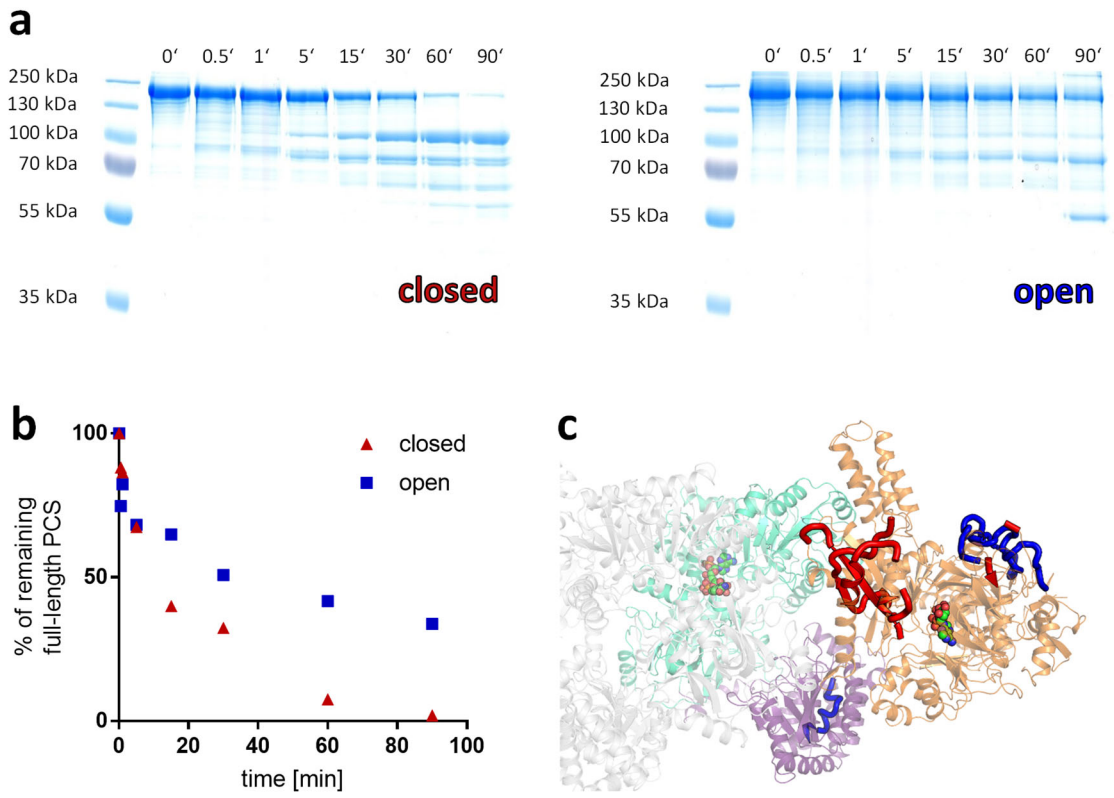
Supplementary Figure 13. Lineweaver-Burk representation of the effects of CoA on the reductase reaction of PCS. All points were measured in triplicates and the 95% confidence intervals are indicated. **a**, The reductase reaction alone catalyzed by the PCS E1027Q variant depending on the acrylyl-CoA concentration in presence of three different CoA concentrations. The inhibition effect is competitive at a CoA concentration of 0.3 mM (change in apparent $K_{M,\,CoA}$, same k_{cat}) but gains non-competitive character at the higher CoA concentration of 3 mM (changes in apparent $K_{M,\,CoA}$ and k_{cat}). **b**, Detailed view of the data of a,. **c**, The reductase reaction alone catalyzed by the PCS E1027Q K783M variant depending on the acrylyl-CoA concentration in presence of two different CoA concentrations. The interaction between K783 and CoA is suggested to stabilize the closed conformation of PCS (**Supplementary Figure 14**). The K783M mutation rendered the ligase as well as the reductase domains still functional. Notably the non-competitive inhibition of CoA on the reductase domain was completely removed in this variant. Thus, we concluded that the ligase domain and the reductase domain act not independently from each other and that the K783 plays an important role in communication between the two domains. **d**, Detailed view of the data of c,.



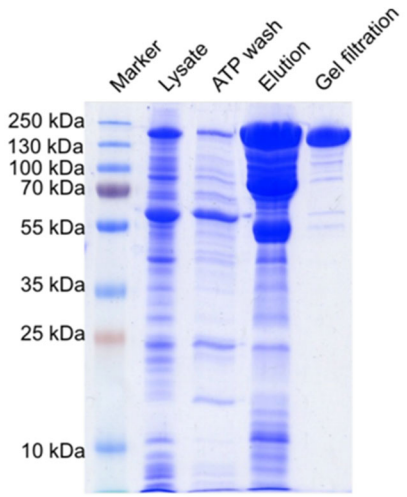
Supplementary Figure 14. Hypothetic conformational change of PCS upon CoA binding. The left panel depicts the effect of the plausible conformational change on the overall structure by surface representation, while the right panel shows a close up of the ligase domain active site in the different conformations with modeled CoA. The modeled open conformation suggests how PCS allows substrates to enter the reaction chamber. Our observation of CoA preventing access of intermediates to the reaction chamber can be explained by this hypothesis. Note, that the PCS-specific four-helix bundle is required to enclose the enzyme compartment. **a**, closed conformation of PCS as it was obtained from the crystal structure in this study. The closed conformation might be stabilized by interactions of K783 and R812 of the flexible cap domain with bound CoA. **b**, open conformation of PCS was modeled by replacing the flexible cap subdomain of PCS with the one from the open-state ligase of *Saccharomyces cerevisiae* (PDB 1RY2) [22]. The size of the opening is roughly 13 Å and 24 Å in the two dimensions. A loop corresponding to PCS residues 831-862 had to be omitted from the model because of clashes. The green spheres mark its insertion points. Orange – non-flexible part of ligase domain, light green – flexible cap of ligase domain (defined by comparing PDB 2P2F and 1RY2 [17,22]), purple – dehydratase domain, cyan and grey – reductase domains from both protomers.



Supplementary Figure 15. Limited proteolysis on PCS with different substrates/products. Limited proteolysis of PCS with trypsin (200:1 protein:protease ratio) analysed by SDS-PAGE. 1, PCS (0.5mg/ml) apo; 2, with CoA; 3, with CoA, ATP, $NADP^+$; 4, with an ATP analog, CoA, 3-hydroxypropionate; 5, with ATP; 6, with ATP, $NADP^+$; 7, with ATP, 3-hydroxypropionate; 8, with ATP, 3-hydroxypropionate, $NADP^+$; 9, PCS apo without trypsin. CoA at 3 mM, ATP and analog at 3.4 mM, 3-hydroxypropionate at 2mM, $NADP^+$ at 2mM.



Supplementary Figure 16. Limited Proteolysis of PCS in open and closed conformation. a, limited proteolysis of PCS with trypsin (200:1 protein:protease ratio) analysed by SDS-PAGE. The closed conformation (left) was induced by addition of 3 mM CoA, while the open conformation was stabilized by 2 mM 3HP, 3.4 mM ATP and 2 mM $NADP^+$ (right). **b**, percent of remaining full-length PCS over time was quantified from the SDS-PAGE gel. Samples on both gels derive from the same experiment and gels were processed in parallel. **c**, peptides from the closed and open conformation appearing after the first 30 seconds of the limited proteolysis were quantitatively compared. Peptides in blue are slightly less prominent in the sample of closed PCS (ratio of peptide in closed sample to open sample: 0.73 for peptide 866-890, 0.82 for peptide 714-724, 0.84 for peptide 1053-1065, 0.90 for peptide 121-131) while peptides in red are overrepresented in the sample of closed PCS (ratio of peptide in closed sample to open sample: 1.38 for peptide 664-673, 1.71 for peptide 866-890, 2.05 for peptide 805-818, 2.3 for peptide 837-855). Data of a representative single experiment.



Supplementary Figure 17. SDS-PAGE gel following the purification of PCS. The lysate was loaded onto a 1 mL His-Trap column. A wash step with buffer containing ATP was applied to flush away the *E.coli* ArcticExpress (DE3) RIL Cpn60 chaperone. PCS was eluted and separated from contamination or degradation products on a HiLoad 16/60 200 pg superdex column.

Chapter III

Awakening the sleeping carboxylase function of enzymes: engineering the natural CO_2-binding potential of reductases

Authors:

Iria Bernhardsgrütter, Kristina Schell, Dominik M. Peter, Farshad Borjian, David Adrian Saez, Esteban Vöhringer-Martinez, Tobias J. Erb

Published in:

Journal of the American Chemical Society **141**, 9778-9782 (2019)

doi.org/10.1021/jacs.9b03431

Author contributions:

I.B., K.S., D.M.P. and T.J.E. conceived the project. I.B., K.S. and T.J.E. designed and performed experiments and analyzed the data. I.B. and F.B. established and performed CoA-esters analyses. I.B., K.S. and T.J.E. wrote the manuscript with contributions from all other authors.

3 Awakening the sleeping carboxylase function of enzymes: engineering the natural CO_2-binding potential of reductases

3.1 Abstract

Developing new carbon dioxide (CO_2) fixing enzymes is a prerequisite to create new biocatalysts for diverse applications in chemistry, biotechnology and synthetic biology. Here we used bioinformatics to identify a "sleeping carboxylase function" in the superfamily of Medium-chain Dehydrogenases/Reductases (MDR), i.e., enzymes that possess a low carboxylation side activity next to their original enzyme reaction. We show that propionyl-CoA synthase from *Erythrobacter* sp. NAP1, as well as an acrylyl-CoA reductase from *Nitrosopumilus maritimus* possess carboxylation yields of 3 ± 1 and 4.5 ± 0.9 %. We use rational design to further engineer these enzymes into carboxylases by increasing interactions of the proteins with CO_2 and suppressing diffusion of water to the active site. The engineered carboxylases show improved CO_2-binding and kinetic parameters comparable to naturally existing CO_2-fixing enzymes. Our results provide a strategy to develop novel CO_2-fixing enzymes and shed light on the emergence of natural carboxylases during evolution.

3.2 Introduction

To harvest atmospheric CO_2 as a sustainable carbon source for (bio)catalytic and (bio)technological applications[1-5], it is necessary to extend the repertoire of CO_2-fixing reactions. One possibility is to engineer a carboxylation function into the scaffold of non-CO_2-fixing enzymes. Generally, the interaction of CO_2 with proteins is poorly understood[6]. However, for enoyl-CoA carboxylase/reductase from *Kitasatospora setae* (ECR_{Ks}), four conserved amino acids that form a CO_2-binding pocket at the active site were described recently[7] (**Figure 1a**). These four amino acids anchor and position the CO_2 molecule during catalysis, in which a reactive enolate is formed that attacks the CO_2[8].

To identify enzyme scaffolds capable of binding CO_2 beyond the ECR enzyme family, we searched homologs of the MDR superfamily for the CO_2-binding motif. Our search revealed two enzyme families that show the potential to bind CO_2, the propionyl-CoA synthase (PCS) and an archaeal enoyl-CoA reductase (AER) family (**Figure 1b**). The PCS family clusters closely to ECRs and shows a fully conserved CO_2-binding motif across individual family members (**Figure S1**). The AER family is more distantly related to the ECR family, and selected homologs only contain one or two of the

four conserved residues of the CO_2-binding pocket (**Figure S2**). We decided to test selected members of these enzyme families in their CO_2-fixing capabilities.

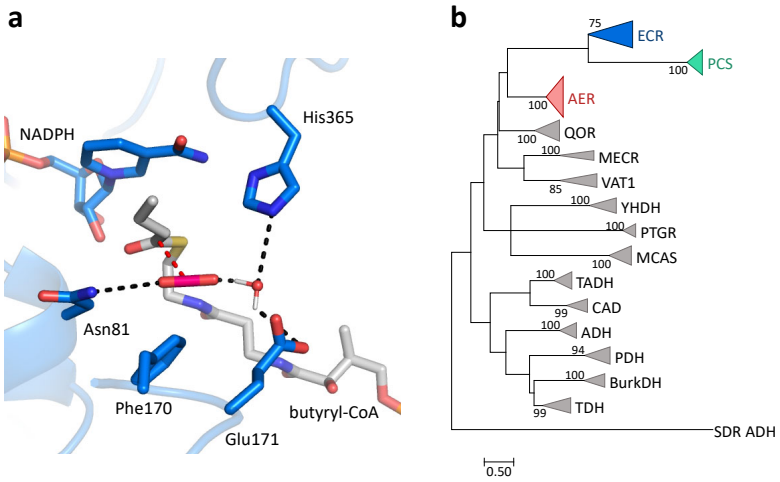


Figure 1: CO_2-binding pocket of ECR and its partial conservation in the MDR superfamily. a, active site of ECR_{Ks}[7]. The CO_2-binding pocket is defined by four conserved residues (Asn81, Phe170, Glu171, His365). CO_2 was modeled into the structure. **b,** Maximum-likelihood tree of the MDR superfamily[9] with (potential) CO_2-binding enzyme families highlighted in color.

3.3 Results

PCS is a three-domain fusion enzyme that catalyzes the overall conversion of 3-hydroxypropionate to propionyl-CoA[10] (**Figure 2a**). The enzyme forms a central reaction chamber, in which three subsequent reactions take place in a synchronized fashion[11]. When we assayed PCS from Erythrobacter sp. NAP1, PCS_{EN}, at 4.4 mM dissolved CO_2, we detected minor amounts of methylmalonyl-CoA besides the main product propionyl-CoA. Incorporation of $^{13}CO_2$-label confirmed the latent carboxylation activity of PCS_{EN} (**Figure 2b**). Notably, the carboxylation function was not limited to the Erythrobacter enzyme, but was also detected with PCS from Chloroflexus aurantiacus (PCS_{Ca}, **Table S1**).

The last reaction in the three-reaction sequence of PCS is the reduction of acrylyl-CoA to propionyl-CoA, catalyzed by a reductase domain harboring the CO_2-binding motif (**Figure 2a** and **2c**). We directly tested the reductase domain for carboxylation activity with an E1027Q variant of PCS_{EN} ($PCS_{EN\_\Delta DH}$) that is unable to generate acrylyl-CoA. When providing $PCS_{EN\_\Delta DH}$ with external acrylyl-CoA and 4.4 mM dissolved CO_2, the enzyme showed a carboxylation yield (defined as percentage yield of carboxylated product compared with total product formed, including reduced side product) of 3 ± 1 % (**Table 1**). This showed that the reductase domain is able to carboxylate acrylyl-CoA directly.

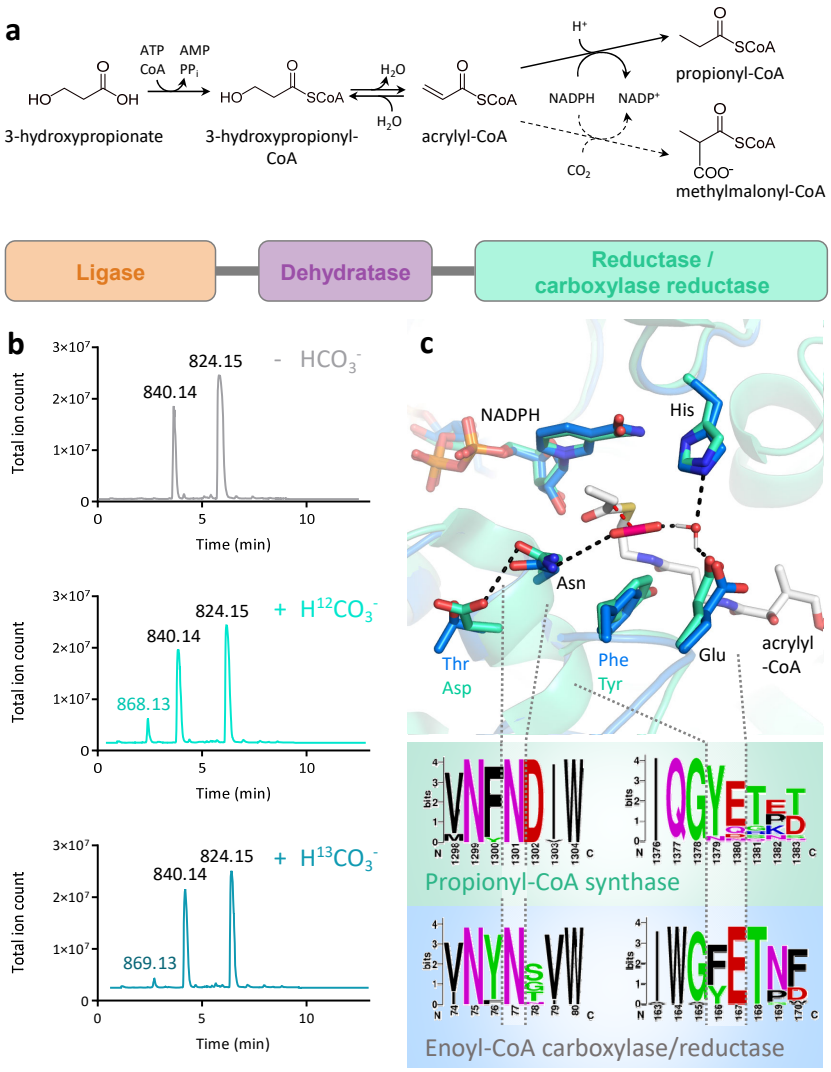


Figure 2: PCS$_{EN}$ possesses a "sleeping carboxylase function". a, Reaction sequence of PCS. PCS natively catalyzes the conversion of 3-hydroxypropionate into propionyl-CoA (solid lines) and possesses a low carboxylation activity yielding methylmalonyl-CoA (dashed line). **b,** HPLC-MS traces of the PCSEN overall reaction showing 3-hydroxypropionyl-CoA, propionyl-CoA (3-$^{12/13}$C)-methylmalonyl-CoA at m/z 840.14, 824.15 and 868.13/869.13, respectively. Methylmalonyl-CoA is only detected in presence of $^{12/13}CO_2$ (provided as bicarbonate). Data represent an individual experiment with two replicates. **c,** Active site of PCS$_{EN}$ reductase domain (cyan, PDB: 4EQO11) and ECR$_{Ks}$ (blue7), both co-crystallized with NADP$^+$. Acrylyl-CoA and CO_2 are modeled into the active site. WebLogo-Illustration[12,13] of conserved active site residues using 129 PCS and 29 ECR sequences. Numbering according to PCS$_{EN}$ or ECR$_{Ks}$, respectively.

To further improve the carboxylation efficiency of PCS$_{EN}$, we compared the active site of PCS$_{EN}$ (PDB: 6EQO) with ECR$_{Ks}$. While the NADPH binding site, as well as the four CO_2-binding pocket residues are structurally conserved (**Figure 2c**), we noticed differences in the second shell of the active site. ECR$_{Ks}$ features a small hydrophilic residue (Thr82), which interacts with Asn81 that stabilizes CO2 through its carboxyamide NH_2 group. The corresponding residue in PCS$_{EN}$ is occupied by an aspartate (Asp1302). Molecular dynamics (MD) simulations demonstrated that Asp1302 in PCS$_{EN}$ forms a strong anionic hydrogen bond to the carboxamide NH_2 group of Asn1301 (**Figure 3a**

and **S5**), locking Asn1301 in a position which prevents interactions with CO_2. This finding is in line with the fact that we could not determine an apparent K_M for CO_2 with $PCS_{EN\_\Delta DH}$ and that replacing Asn1301 by an aspartate abolished carboxylation activity.

Table 1: Reaction parameters and carboxylation yield for the reductase domain of different PCS_{EN} variants.

PCS variant	app. k_{cat} (s^{-1}) at 4.4 mM CO_2	app. $K_{M\_acrylyl\text{-}CoA}$ (mM)	% carboxylation at 4.4 mM CO_2	app. K_{M_CO2} (mM)
$PCS_{EN\_\Delta DH}$ WT	7.4 ± 1.0	0.014 ± 0.002	3 ± 1	n.m.
$PCS_{EN\_\Delta DH}$ D1302S	1.77 ± 0.09	0.027 ± 0.003	20.9 ± 0.7	27 ± 5
$PCS_{EN\_\Delta DH}$ T1753M	6.5 ± 0.6	0.0197 ± 0.0012	10 ± 2	n.m.
$PCS_{EN\_\Delta DH}$ D1302S T1753M	0.46 ± 0.03	0.026 ± 0.003	69 ± 3 %	26 ± 5

k_{cat} shows combined reduction and carboxylation activity. K_M were determined from a Michaelis-Menten fit of at least 18 data points, with fixed acrylyl-CoA concentrations for K_{M_CO2} (**Figure S3** and **S4**, **Table S2** for k_{cat} values). Carboxylation yields are calculated from mean carboxylation yields over five time points in three replicates. Data are mean ± s. d. CO_2 concentrations were calculated. n.m., not measurable.

We aimed at unlocking Asn1301 from its fixed position by replacing Asp1302 with different small hydrophilic residues. $PCS_{EN\_\Delta DH}$ variant D1302S (**Figure 3b**) showed an increased carboxylation yield of 20.9 ± 0.7 % at 4.4 mM dissolved CO_2, and notably also Michaelis-Menten-like behavior with CO_2 at an apparent K_{M_CO2} of 27 ± 5 mM (**Table 1**). Together with MD simulations that showed a more flexible asparagine residue (**Figure S5**) this demonstrated that unlocking Asn1301 improves CO_2-binding and carboxylation efficiency in PCS_{EN}.

Another, equally important catalytic principle in carboxylases is the exclusion of water from the active site to minimize protonation reactions which would prematurely quench C-C bond formation[7,14-16]. In ECR_{Ks}, a conserved methionine (Met356) restricts access of water to the CO_2-binding pocket. In PCS_{EN}, this residue is a threonine, which presumably allows water to enter the active site and displace the CO_2 molecule (**Figure 3c**). When we introduced the methionine in PCS_{EN} ($PCS_{EN\_\Delta DH}$ T1753M, **Figure 3d**), carboxylation yield increased to 10 ± 2 % at 4.4 mM dissolved CO_2. When combining the D1302S with the T1573M mutation, the carboxylation yield of $PCS_{EN\_\Delta DH}$ further increased up to 69 ± 3 % at 4.4 mM CO_2 (**Table 1**). Under saturating CO_2 concentrations (i.e., 44 mM CO_2), $PCS_{EN\_\Delta DH}$ D1302S T1753M showed a carboxylation yield of 94.5 ± 0.7 %, demonstrating that we successfully converted the reductase domain into a carboxylase. During engineering, the k_{cat} of reduction was strongly decreased, while the apparent k_{cat} for carboxylation

was maintained (**Table S2**) and falls in the range of naturally existing ECRs[4,17]. The engineered carboxylase domain also improved carboxylation yield in the context of the overall reaction of PCS$_{EN}$ (**Supplementary Information I**).

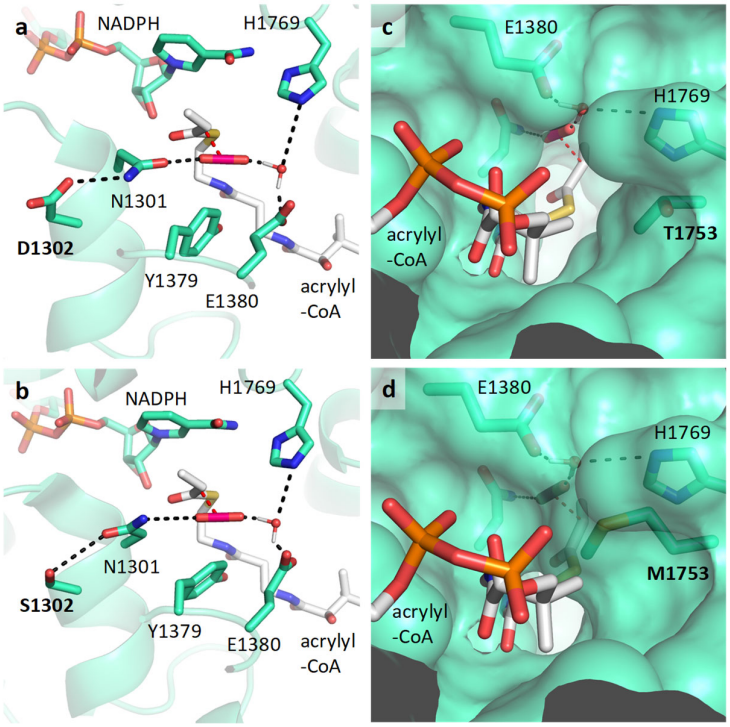


Figure 3: Directed mutagenesis to exploit the carboxylation activity of PCS$_{EN}$. a, representative snapshot from the MD simulation of the active site in wild type PCS$_{EN}$. **b**, active site model of PCS$_{EN}$ D1302S to unlock Asn1301. **c,** active site of wild type PCS$_{EN}$. **d,** active site model of PCS$_{EN}$ T1753M to restrict water access to the active site. Acrylyl-CoA and CO_2 were modeled into the active site.

We next investigated the carboxylation potential in the AER enzyme family of unknown function. We chose Nmar_1565 (AER$_{Nm}$), a homolog from *N. maritimus*, in which two of the four amino acids of the CO_2-binding motif, namely Phe122 and Glu123, are conserved (**Figure 4a** and **4b**). Although no function was assigned to AER$_{Nm}$ so far, we speculated that the enzyme might catalyze the reduction of acrylyl-CoA in the 3-hydroxypropionate/4-hyroxybutyrate cycle of *N. maritimus*[18,19]. Indeed, the enzyme reduced acrylyl-CoA to propionyl-CoA at an apparent k_{cat} of 0.99 ± 0.11 s^{-1}, confirming its reductase function.

AER$_{Nm}$ activity was very sensitive to salt and buffer composition (**Supplementary Information II**). When we incubated the enzyme with $NaHCO_3$, at concentrations corresponding to 1.31 mM free CO_2, activity dropped 10-fold. However, under these conditions AER$_{Nm}$ showed a latent carboxylation activity and converted acrylyl-CoA into methylmalonyl-CoA at a carboxylation yield of 4.5 ± 0.9 % (**Table 2**), despite the lack of two of the four amino acid residues of the CO_2-binding motif. To increase the carboxylation efficiency of AER$_{Nm}$ we decided to re-build the CO_2-binding pocket through introduction of asparagine and histidine. Re-introduction of histidine failed due to

inactive protein, which might be a result of interrupted second-shell interactions to Thr307 or steric clashes. However, replacing Asp50 by asparagine increased carboxylation yield dramatically (to 82 ± 5 %, **Figure 4c**, **Table 2**). The increase in catalytic activity in AER_{Nm} D50N was accompanied by an improved K_{M_CO2} (0.18 ± 0.03 mM), indicating increased CO_2-binding. AER_{Nm} D50N performed best in 100 mM phosphate buffer, where it showed k_{cat} and carboxylation yields comparable to those of naturally existing carboxylases, such as RubisCO[20] **(Table 2)**.

Table 2: Reaction parameters and carboxylation yield for AER_{Nm} variants in 100 mM $KHPO_4$ (pH 7.5).

AER variant	app. k_{cat} (s-1) at 1.31 mM CO_2	% carboxylation at 1.31 mM CO2	$K_{M\_}CO_2$ (mM)
WT	0.084 ± 0.011	4.5 ± 0.9	n.m.
D50N	1.6 ± 0.2	82 ± 5	0.18 ± 0.03

k_{cat} shows combined carboxylation and reduction activity. K_{M_CO2} were determined from a Michaelis-Menten fit of at least 18 data points with a fixed concentration of acrylyl-CoA (**Figure S6**, **Table S3** for k_{cat} values). Carboxylation yields were calculated from the mean carboxylation ratio over five time points in three replicates. Data are mean ± s. d. CO_2 concentrations were calculated. n.m., not measureable.

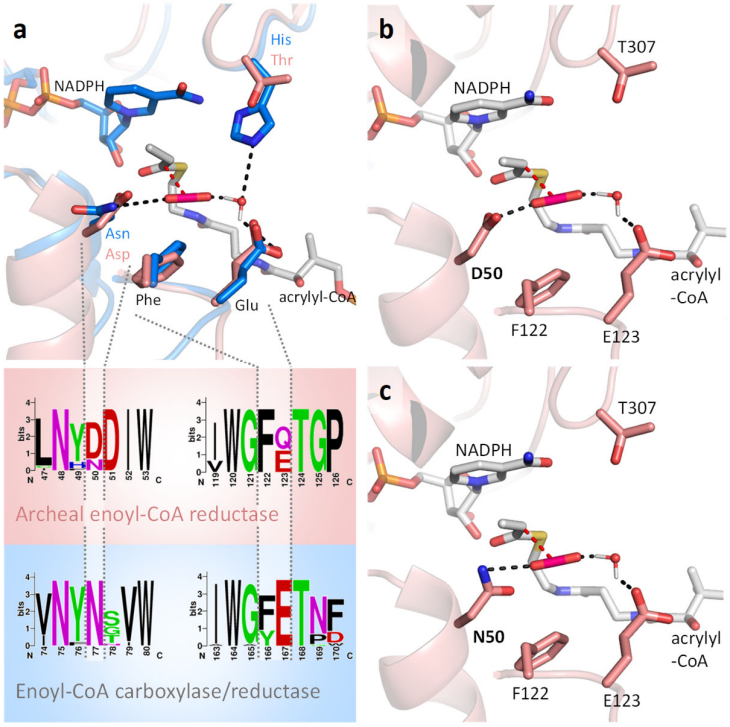


Figure 4: Awakening the "sleeping carboxylase function" in AER_{Nm}. a, Active sites of AER_{Nm} (salmon) and ECR_{Ks} (blue). Illustration of conserved active site residues, generated by WebLogo[12,13] using 21 AER_{Nm} and 29 ECR sequences. Residue numbering refers to AER_{Nm} or ECR_{Ks} sequence, respectively. **b,** Model of the AER_{Nm} wild type active site carrying an Asp50 instead of a conserved Asn. **c,** Model of the AER_{Nm} D50N active site. Homology models were created with an ECR from *Streptomyces* sp. NRRL 2288 (PDB: 4y0k[21]) using SWISS-MODEL[22]. Acrylyl-CoA and CO_2 were modeled into the active site.

3.4 Discussion

In conclusion, we successfully re-shaped the energy landscape of acrylyl-CoA reductases from the thermodynamically favored product propionyl-CoA ($\Delta_rG'^0$ ≈ -63 kJ/mol) to the disfavored methylmalonyl-CoA ($\Delta_rG'^0$ ≈ -43 kJ/mol)[23]. Our engineering efforts show that improving CO_2-binding (reduced energy barrier for carboxylation) and minimizing side reaction with water (increased energy barrier for reduction) are both required to establish a carboxylation activity in the scaffold of different reductases. This is in line with the idea that in catalysis stabilization of favorable transition states ('positive catalysis') and destabilization of unwanted transition states ('negative catalysis') are both important[24-26], as further supported by the finding that suppression of competing protonation side reactions is essential for efficient CO_2-fixation in ECR_{Ks} and 2-Ketopropyl Coenzyme M Oxidoreductase/Carboxylase[14-16].

On a broader picture, our findings also raise questions about the emergence of natural carboxylases. How did carboxylation functions naturally evolve in the scaffold of proteins, such as RubisCO or ECR? It has been suggested, that these enzymes originated from non-CO_2-fixing ancestors[27,28]. Our data provides experimental evidence for this evolutionary scenario by demonstrating that the MDR superfamily, to which ECR belongs, naturally possesses the capacity to interact with the CO_2-molecule. It apparently takes only few mutations to transform latent carboxylases that convert CO_2 at low efficiency and non-physiological CO_2 concentrations into decent CO_2-fixing enzymes.

Another apparent question is why PCS and AER would possess a "sleeping carboxylase function"? One explanation might be that the latent carboxylation activity was selected for. PCS operates in the 3-hydroxypropionate bicycle in *C. aurantiacus* and a modified version thereof in *E.* sp. NAP1 (**Figure S7a**)[29,30], while AER_{Nm} presumably works in the 3-hydroxypropionate/4-hyroxybutyrate cycle in *N. maritimus* (**Figure S7b**)[18]. Bioenergetic considerations suggest that even a low carboxylation activity would increase biomass yield of these organisms, which thrive at a constantly low energy supply[18] (**Supplementary Information III**).

In summary, our proof-of-principle study demonstrates that it is possible to exploit the active site of reductases to create novel carboxylases. This opens the possibility for the future engineering of novel CO_2-fixing enzymes that could find application in biocatalysis and synthetic biology (e.g. in artificial pathways for the conversion of CO_2[31,32]).

3.5 Methods

Chemicals

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

Synthesis of CoA-esters

For the synthesis of 3-hydroxypropionyl-CoA a previously described method using carbonyldiimidazole coupling of the precursor acid with coenzyme A was used[33]. Acrylyl-CoA was synthesized using a previously described ethylchloroformate method[33]. Propionyl-CoA was produced from propionic anhydride using the symmetric anhydride method[33]. Methylmalonyl-CoA was enzymatically synthesized using the ligase MatB[33]. All CoA-esters were purified by HPLC-MS as previously described[33].

Bacterial strains and growth conditions

E. coli DH5α (Thermo Scientific™) strains were used for cloning and grown in LB medium[34]. For protein expression *E. coli* BL21-AI™ (Invitrogen) were grown in TB medium[35]. Incubation temperature was 37°C. Antibiotics for selection purposes were used accordingly: 100 µg/ml ampicillin, 20 µg/ml streptomycin.

Cloning. All *in silico* cloning was performed with Clone Manager 9 (Scientific & Educational Software). For purification, preparation, cloning, transformation and amplification of DNA, standard protocols were used[36]. Plasmid isolation and PCR product purification was performed with kits from Macherey Nagel (Düren, Germany) according to the manufacturer's protocols.

The *Eryhtrobacter* sp. NAP1 PCS gene (PCS_{EN}, GenBank accession no. EAQ29651) with an N-terminal 10x His tag codon-optimized for *E. coli*, the *Chloroflexus aurantiacus* PCS gene (PCS_{Ca}, GenBank accession no. AAL47820.2) with an N-terminal 10x His tag and the *Nitrosopumilus maritimus* SCM1 Nmar_1565 gene (AER_{Nm}, GenBank accession no. ABX13461) with N-terminal

6x His tag were synthesized by the DOE Joint Genome Institute. The PCS_{EN} gene was integrated into the pSEVA471 expression backbone. The plasmid is referred to as pTE1012. The PCS_{Ca} gene was cloned into the pET-16b expression vector, resulting in pTE1006. The pRSET B backbone was used for the integration of the AER_{Nm} gene. The plasmid is referred to as pTE421. Point mutants were generated by QuickChange® Site-Directed mutagenesis (Stratagene, La Jolla, USA). Following primers were used for Site-Directed mutagenesis on the PCS_{EN} gene: forward primer (5'-CGA AGT TAA CTT CAA CAG TAT CTG GGC TCT GAC TG-3') and reverse primer (5'-CAG TCA GAG CCC AGA TAC TGT TGA AGT TAA CTT CG-3') for the D1302S variant; forward primer (5'-CCG AAG TTA ACT TCG ACG ATA TCT GGG CTC-3') and reverse primer (5'-GAG CCC AGA TAT CGT CGA AGT TAA CTT CGG-3') for the N1301D variant; forward primer (5'-CGC AGG TAT GGA TGC GTC AGC GCC GC-3') and reverse primer (5'-GCG GCG CTG ACG CAT CCA TAC CTG CG-3') for the T1753M variant.

For mutagenesis of the AER_{Nm} gene following primers were used: forward primer (5'- CAG CCC TGA ACT ACA ACG ATA TCT GGG G-3') and reverse primer (5'-GCC CCA GAT ATC GTT GTA GTT CAG GGC TGC-3') for the D50N variant; forward primer (5'-CAT TCT GGG CTC TCA TCA AGG GAC GCG TG-3') and reverse primer (5'-CAC GCG TCC CTT GAT GAG AGC CCA GAA TG-3') for the T307H variant.

The *Pseudomonas aeruginosa* PhaJ gene, encoding for an enoyl-CoA hydratase, was synthesized by Eurofins MWG GmbH to be codon-optimized for *E. coli.* It was cloned into a pET-16b expression vector with an N-terminal strep tag. The plasmid is referred to as pTE656.

Expression and purification

The two PCS homologs (PCS_{EN} and PCS_{Ca}) and variants thereof as well as PhaJ were expressed from the plasmid pTE1012, pTE1006 or pTE656 using *E.coli* BL21 (DE3) AI as expression host. AER_{Nm} variants were expressed from pTE421 using *E.coli* BL21 (DE3). The cells were transformed with the expression plasmid and plated on LB agar containing selective antibiotic and grown overnight. The colonies were used to inoculate 1 L TB medium. The expression culture was incubated at 37°C while shaking at 110 rpm until an OD_{600} of 0.7 – 0.9 was reached. The *E.coli* culture was cooled down to 20°C before induction. Expression was induced by adding 0.25 mM IPTG. When using *E.coli* BL21 (DE3) AI as expression host 0.02 % L-arabinose was added additionally. The culture was incubated for 16-20 h. The cells were harvested by centrifugation at 5000 *x g* for 10 min. The pellet was stored at -20°C, optionally. Cells expressing PCS homologs and variants thereof as well as PhaJ were resuspended in a 1:3 ratio (w/w) in buffer A (50 mM Tris-HCl pH 7.9, 500 mM NaCl) containing SIGMAFAST™ protease inhibitor (Sigma-Aldrich, Munich, Germany), 5 mM $MgCl_2$ and 10 µg/mL

DNaseI. Per gram cells expressing AER_{Nm} variants, 2 mL buffer C (500 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1 M L-proline) containing 5 mM $MgCl_2$ and 10 µg/mL DNaseI was used to resuspend the pellet. Cells were lysed by ultrasonication. A heat precipitation step for 15 min at 60°C followed for the lysate containing the *C. aurantiacus* PCS homolog. The lysate was cleared by ultracentrifugation at 50'000 *x g* for 45 min at 4°C followed by filtration through a 0.45 µm syringe filter. The lysates from the different PCS and AER overexpression strains were loaded onto a 1 mL His-Trap column (GE Healthcare). Unspecifically bound proteins were washed off with 15 mL of 5 % buffer B (50 mM Tris-HCl pH 7.9, 500 mM NaCl, 500 mM imidazole) during purification of PCS and PhaJ and buffer D (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 500 mM imidazole, 1M L-proline) was used for AER_{Nm}. PCS homologs and AER_{Nm} were eluted with 100% buffer B or D, respectively. Desalting columns with Sephadex G-25 resin (HiTrapTM, GE Life Science, USA) and desalting buffer (125 mM NaCl, 12.5 mM Tris, pH 7.5, 1 M L-proline) were used for size exclusion chromatography with AER_{Nm}, while PCS homologs were applied to a pre-equilibrated HiLoad 16/60 200 pg superdex (GE Life Science) column (150 mM NaCl, 20 mM Tris HCl pH 7.9). The lysate of PhaJ overexpression strain was loaded onto a 1 mL Strep-Trap column (GE Healthcare). PhaJ was eluted with 2.5 mM desthibiotin in buffer A. The purity of the proteins was tested by SDS-PAGE.

Kinetic characterization

Spectrophotometric assays were set up to measure the activity of PCS_{EN} and AER_{Nm}. The assays were performed in 10 mm quartz cuvettes (Hellma Analytics) on a Cary-60 UV/Vis spectrometer (Agilent Technologies Inc. Santa Clara, CA, USA). The assay temperature was set to 30 °C. The PCS_{EN} overall reaction and the reduction reaction alone as well as the reaction of AER_{Nm} were measured by following the consumption of NADPH at 340 nm ($\varepsilon_{NADPH} = 6.22\ mM^{-1}\ cm^{-1}$). The parameters for the CoA ligase domain alone of PCS_{EN} were measured using a coupling assay via myokinase (purified from ASKA JW1375), pyruvate kinase and lactate dehydrogenase (SigmaAldrich P02694). Data is illustrated in **Figure S3** and **S10** and listed in **Table S1**. Detailed assay compositions to determine biochemical characteristics are listed in **Table S4** and **S7**. The influence of ionic strength on the AER_{Nm} variants was measured under compositions listed in **Table S8**. The *kcat* values for the AER_{Nm} WT in dependence on NaCl concentrations ranging from 0-200 mM were measured in 100 mM $K_xH_yPO_4$ (pH 7.5). The ion screen was performed with the AER_{Nm} D50N variant in 20 mM Tris / 200mM proline (pH 7.5) with addition of 0, 20 or 200 mM of the displayed ions **Figure S9**.

To determine the K_M for CO_2 of the different PCS_{EN} and AER_{Nm} variants, a discontinuous assay had to be performed. Detailed assay compositions are listed in **Table S5** and **S9**. The assays with PCS_{EN} variants were performed in 100 mM Tris buffer (pH 8.0) and $KHCO_3$ concentrations ranging from 0

– 2 M (potassium concentration was filled up to 2 M in every assay sample using KCl) and the corresponding dehydratase deficient $PCS_{EN\_\Delta DH}$ variant. Assays with AER_{Nm} contained either buffer with 100 mM $K_xH_yPO_4$ (pH 7.5) or 20 mM tris - 200 mM prolin (pH 7.5) and $NaHCO_3$ concentrations ranging from 0 – 80 mM (sodium concentration was filled up to 80 mM in every assay sample using NaCl). Each reaction contained 69 nM carbonic anhydrase and 2 mM NADPH. Impurities in chemically synthesized acrylyl-CoA interfered with assay analysis. Therefore, the substrate acrylyl-CoA was generated *in situ* using 2 mM 3-hydroxypropionyl-CoA and the dehydratase PhaJ. The samples were quenched with pre-cooled formic acid to a final concentration of 10 % (v/v), centrifuged for 10 minutes at 17'000 *x g* and 4°C and immediately analyzed by UHPLC. Triplicates were measured at every CO_2 concentration, each replicate comprising six time points to determine the slope for methylmalonyl-CoA and propionyl-CoA generation by UHPLC analysis. The methylmalonyl-CoA generation rates and the carboxylation efficiencies were plotted against the CO_2 concentration (**Figure S4** and **S6**). Separate k_{cat} values for carboxylation and reduction are listed in **Table S2** and **S3**.

The CO_2 concentration was calculated from the following formula[37]: $[CO_2] = ([HCO_3^-] \times [H^+]^2) / ([H^+]^2 + K_{a1} \times [H^+] + K_{a1} \times K_{a2})$, with $K_{a1} = 4.45 \times 10^{-7}$, $K_{a2} = 4.69 \times 10^{-11}$.

Carboxylation in PCS overall reaction

PCS_{EN} WT was assayed for its carboxylation capability by running the overall reaction and analyzing the generated products by hrLC-MS. The overall reaction was performed in 100 mM phosphate Buffer (pH = 8), 10 mM $MgCl_2$ and 40 mM KCl in the presence or absence of 200 mM $NaHCO_3$ or $NaH^{13}CO_3$ and 69 nM carbonic anhydrase. 1.45 µM of PCS_{EN} WT was supplied with 8 mM of NADPH, 8 mM of ATP, 8 mM of 3-hydroxypropionate and 5 mM of CoA.

Carboxylation ratios of the different PCS homologs and variants thereof in the overall reaction were determined in time course assays. Detailed assay compositions are listed in **Table S6**. A master mix was prepared for all different reactions that contained everything except the starting substrate 3-hydroxypropionate. The master mix was then split into three replicates which were started individually by adding 3-hydroxypropionate. Samples were taken at specific time points and quenched with pre-cooled formic acid to a final concentration of 10 % (v/v). Samples were centrifuged for 10 minutes at 17'000 *x g* and 4°C and the supernatant was immediately frozen in liquid nitrogen and stored at -80°C until UHPLC analysis. Triplicates were measured, each replicate comprising six time points. The 3-hydroxypropionyl-CoA, methylmalonyl-CoA and propionyl-CoA concentration in each sample was determined and plotted against the reaction time (**Figure S8**).

UHPLC product analysis

To measure the product or intermediate formation, 1 µL of the assay sample was injected into a 1290 Infinity II UPLC-UV system (Agilent Technologies Inc. Santa Clara, USA). The samples were separated on Eurospher II 100-2 C18 column (Knauer Wissenschaftliche Geräte, Berlin, Germany). To detect the CoA-thioesters via UV absorbance an InfinityLab Max-Light cartridge cell was used (10 mm detector length for the carboxylation assay samples of PCS_{EN} WT, D1302S and N1301D variants, 60 mm detector length for all other samples, Agilent Technologies Inc. Santa Clara, USA). To separate the CoA-thioesters in the overall reaction samples, a gradient of 2 – 10 % (v/v) acetonitrile in 10 mM potassium phosphate buffer (pH 6.8) over 6.5 min at a flow rate of 0.2 mL/min was applied (retention times: methylmalonyl-CoA 3.9 min, propionyl-CoA 7.1 min, 3-hydroxypropionyl-CoA 5.0 min, CoA 4.2 min). In the longer method with a gradient of 1.5 – 10 % (v/v) acetonitrile in 10 mM potassium phosphate buffer (pH 6.8) over 8 min at a flow rate of 0.2 mL/min the compounds eluted as followed; methylmalonyl-CoA 4.1 min, propionyl-CoA 8.0 min, 3-hydroxypropionyl-CoA 5.1 min, CoA 4.4 min. Standard series for free CoA, 3-hydroxypropionyl-CoA, propionyl-CoA and methylmalonyl-CoA in a range from 5 µM to 500 µM were run on both separation methods. The standards were dissolved in the assay matrix (80 mM potassium phosphate (pH 8.0), 40 mM potassium bicarbonate, 10 % (v/v) formic acid). For quantification of the CoA-thioesters in the samples, standard curves for all compounds were generated and a linear regression model was applied for the correlation of compound concentration and UV_{260} peak area.

High resolution LC-MS (hrLC-MS)

3-hydroxypropionyl-Coa, propionyl-CoA and methylmalonyl-CoA were analyzed using an Agilent 6550 iFunnel Q-TOF LC-MS system equipped with an electrospray ionization source set to positive ionization mode through a 1290 Infinity UPLC (Agilent Technologies Inc. Santa Clara, CA, USA). Compounds were separated on a RP-18 column (50 mm x 2.1 mm, particle size 1.7 µm, Kinetex XB-C18, Phenomenex, Aschaffenburg, Germany) using a mobile phase system comprised of 50 mM ammonium formate pH 8.1 (A) and methanol (B). Chromatographic separation was carried out using the following gradient condition at a flow rate of 250 µl/min: 0 min 0% B; 1 min 0% B, 3 min 2.5% B; 9 min 23% B; 14 min 80 %B; 16 min 80%; 17 min 0 % B; 18 min 0 % B.

Capillary voltage was set at 3.5 kV and nitrogen gas was used as nebulizing (20 psig), drying (13 l/min, 225 °C) and sheath gas (12 l/min, 400°C). The TOF was calibrated using an ESI-L Low Concentration Tuning Mix (Agilent Technologies Inc. Santa Clara, CA, USA) before measurement

(residuals less than 2 ppm for five reference ions). MS data were acquired with a scan range of 750-1200 m/z.

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

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

Molecular dynamic simulations

The starting structure for all simulations was the PCS_{EN} crystal structure (PDB: 6EQO[11]) with the two $NADP^+$ cofactors. Missing residues in the crystal structure were added with the Modeller software[38] and the protein dimer with both cofactors and the water molecules in the crystal structure were solvated inside a rectangular box of TIP3P water molecules with a distance of 10 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 0.150 mol/L. All molecular dynamics simulations were performed with the software Amber17[39] and CHARMM36m force field[40] for the protein. Molecular mechanics parameters for $NADP^+$ were taken from Pavelites[41].

For the D1302S mutant the aspartate residue in the crystal structure was modified to serine. 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} $Å^{-2}$ center of mass restraint on $NADP^+$), 5 ns NPT using Langevin thermostat and Monte Carlo barostat (300 K, 1 bar) and a final equilibration of 30 ns NVT. Electrostatic interactions were calculated with the PME method with a cut-off of 8 Å and a vdw interaction were calculated up to 8 Å using a time step of 2 fs. All bonds were constraint with the SHAKE algorithm.

Three independent simulations of the WT and the D1302S mutant were carried out and analyzed with respect to the orientation of Asp1301 in the active site.

Phylogenetic tree construction

Homologous sequences of PCS_{EN} and AER_{Nm} were searched using the ProtBlast/PSI-Blast tool available from the MPI Bioinformatics Toolkit (https://toolkit.tuebingen.mpg.de)[42]. The nr90 database was screened for 250 sequences each, which were then forwarded to the HHfilter tool. Sequences with a minimal coverage of 90 % and an identity in the lower range from 45-70 % (14 PCS sequences, 20 AER sequences) or in the higher range from 75 – 90 % for AER (21 sequences) or from 70 – 90 % for PCS (13 sequences) were selected. Similar phylogenetic tree were obtained with both sets of sequences (high vs. low sequence identity to AER_{Nm} or PCS_{EN}, respectively). The PCS sequences were shortened to the reductase domain (analysis with full-length PCS gave similar

results). The PCS reductase and AER sequences were combined with all sequences used to construct a previous MDR superfamily tree[28]. Phylogenetic analysis was performed using MEGA7 by the Maximum Likelihood method based on the Le_Gascues_2008 model. The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analyzed[43]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 145 amino acid sequences. There were a total of 746 positions in the final dataset.

3.6 References

1 Glueck, S. M., Gümüs, S., Fabian, W. M. & Faber, K. Biocatalytic carboxylation. *Chem. Soc. Rev.* **39**, 313-328 (2010).

2 Martin, J., Eisoldt, L. & Skerra, A. Fixation of gaseous CO_2 by reversing a decarboxylase for the biocatalytic synthesis of the essential amino acid L-methionine. *Nat. Catal.* **1**, 555 (2018).

3 Plasch, K. *et al.* Pressurized CO_2 as a carboxylating agent for the biocatalytic *ortho*-carboxylation of resorcinol. *Green Chem.* **20**, 1754-1759 (2018).

4 Peter, D. M. *et al.* Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis. *Angew. Chem. Int. Ed.* **54**, 13457-13461 (2015).

5 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. Ed.* **54**, 13462-13465 (2015).

6 Cundari, T. R. *et al.* CO_2-formatics: how do proteins bind carbon dioxide? *J. Chem. Inf. Model.* **49**, 2111-2115 (2009).

7 Stoffel, G. M. M. *et al.* Four amino acids define the CO_2 binding pocket of enoyl-CoA carboxylases/reductases. *Proc Natl Acad Sci U S A* **116**, 13964-13969 (2019).

8 Rosenthal, R. G. *et al.* Direct evidence for a covalent ene adduct intermediate in NAD(P)H-dependent enzymes. *Nat. Chem. Biol.* **10**, 50 (2014).

9 Hedlund, J., Jörnvall, H. & Persson, B. Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement. *BMC bioinformatics* **11**, 534 (2010).

10 Alber, B. E. & Fuchs, G. Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Biol. Chem.* **277**, 12137-12143 (2002).

11 Bernhardsgrütter, I. *et al.* The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite. *Nat. Chem. Biol.* **14**, 1127 (2018).

12 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

13 Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097-6100 (1990).

14 Pandey, A., Mulder, D., Ensign, S. & Peters, J. Structural basis for carbon dioxide binding by 2-ketopropyl coenzyme M oxidoreductase/carboxylase. *FEBS Lett.* **585**, 459 (2011).

15 Kofoed, M. A., Wampler, D. A., Pandey, A. S., Peters, J. W. & Ensign, S. A. 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**, 4904-4913 (2011).

16 Prussia, G. A. *et al.* Substitution of a conserved catalytic dyad into 2-KPCC causes loss of carboxylation activity. *FEBS Lett.* **590**, 2991-2996 (2016).

17 Vögeli, B. *et al.* Combining promiscuous acyl-CoA oxidase and enoyl-CoA carboxylase/reductases for atypical polyketide extender unit biosynthesis. *Cell Chem. Biol.* **25**, 833-839. e834 (2018).

18 Könneke, M. *et al.* Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO_2 fixation. *PNAS* **111**, 8239-8244 (2014).

19 Berg, I. A., Kockelkorn, D., Buckel, W. & Fuchs, G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. *Science* **318**, 1782-1786 (2007).

20 Tcherkez, G. G., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *PNAS* **103**, 7246-7251 (2006).

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. Ed.* **127**, 13664-13667 (2015).

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

23 Flamholz, A., Noor, E., Bar-Even, A. & Milo, R. eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res.* **40**, D770-D775 (2011).

24 Vögeli, B. & Erb, T. J. 'Negative' and 'positive catalysis': complementary principles that shape the catalytic landscape of enzymes. *Curr Opin Chem Biol* **47**, 94-100 (2018).

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

26 Rétey, J. Enzymic reaction selectivity by negative catalysis or how do enzymes deal with highly reactive intermediates? *Angew. Chem. Int. Ed.* **29**, 355-361 (1990).

27 Erb, T. J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr Opin Biotechnol* **49**, 100-107 (2018).

28 Schada von Borzyskowski, L., Rosenthal, R. G. & Erb, T. J. Evolutionary history and biotechnological future of carboxylases. *J Biotechnol* **168**, 243-251 (2013).

29 Zarzycki, J., Brecht, V., Muller, M. & Fuchs, G. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *PNAS* **106**, 21317-21322 (2009).

30 Zarzycki, J. & Fuchs, G. Coassimilation of organic substrates *via* the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*. *Applied and Environmental Microbiology* **77**, 6181-6188 (2011).

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

32 Trudeau, D. L. *et al.* Design and *in vitro* realization of carbon-conserving photorespiration. *PNAS* **115**, E11455-E11464 (2018).

33 Peter, D. M., Vögeli, B., Cortina, N. S. & Erb, T. J. A chemo-enzymatic road map to the synthesis of CoA esters. *Molecules* **21**, 517 (2016).

34 Bertani, G. STUDIES ON LYSOGENESIS I.: The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *J- Bacteriol.* **62**, 293 (1951).

35 Tartof, K. & Hobbs, C. Improved media for growing plasmid and cosmid clones. *Focus* **9**, 12 (1987).

36 Sambrook, J. & Russel, D. *Molecular cloning: a laboratory manual*. 3 edn, (Cold Spring Laboratory Press, New York, 2001).

37 Lower, S. K. Carbonate equilibria in natural waters. *Simon Fraser University* **544** (1999).

38 Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).

39 Case, D. *et al.* AMBER 2017, University of California. *San Francisco* (2017).

40 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14**, 71 (2017).

41 Pavelites, J. J., Gao, J., Bash, P. A. & Mackerell Jr, A. D. A molecular mechanics force field for NAD+ NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem.* **18**, 221-239 (1997).

42 Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237-2243 (2018).

43 Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791 (1985).

3.7 Supplementary Information

Supplementary Information

I Engineered PCS_{EN} carboxylase domain in the context of the overall reaction

Our engineering approaches successfully increased the carboxylation efficiency from 3 ± 1 % in $PCS_{EN\_\Delta DH}$ wild type to 69 ± 3 % in the $PCS_{EN\_\Delta DH}$ D1302S T1753M variant when at 4.4 mM CO_2 (Table 1). We also tested, whether the engineered carboxylase domain would function in the context of the overall reaction of PCS_{EN}. Indeed, carboxylation efficiency was dramatically improved from 4.3 ± 0.6 % in PCS_{EN} wild type to 54 ± 2 % at physiological, non-saturating CO_2-concentrations in PCS_{EN} D1302S T1753M (i.e., 1.1 mM CO_2). However, the overall k_{cat} was strongly decreased and even below the k_{cat} of the engineered carboxylase domain alone (**Table S1**). Catalysis in PCS is highly synchronized across the enzyme's three domains[11]. Detailed analysis showed that the surprising loss of catalytic efficiency in the overall reaction was based in the ligase domain (from 5.7 ± 1.2 s^{-1} in PCS_{EN} WT[11] to 0.0060 ± 0.0002 s^{-1} in PCS_{EN} D1302S T1753M), pointing toward an impaired domain communication in PCS_{EN} D1302S T1753M.

II AER_{NM} sensitivity to buffer composition and ionic strength

Buffer composition and ionic strength had an important influence on the carboxylation efficiency of AER_{Nm}, which showed a narrow activity optimum in phosphate buffer and was sensitive to ion concentrations above 60 mM (**Figure S9** and **S10**). The enzyme performed best in 100 mM phosphate buffer pH 7.5.

III Physiological context of latent carboxylases

We suggested that the latent carboxylation activity in PCS and AER was positively selected for because of the physiological context of these enzymes. PCS operates in the 3-hydroxypropionate bicycle (3-HPBC) of *C. aurantiacus* and a modified version of the same pathway in *E.* sp. NAP1[29,30]. In the 3-HPBC 3-hydroxypropionate is converted into propionyl-CoA by PCS (**Figure S7a**). Half of the propionyl-CoA pool is further carboxylated into methylmalonyl-CoA to feed the autotrophic branch of the cycle by an ATP-dependent propionyl-CoA carboxylase. The other half of the propionyl-CoA is used to feed the assimilation branch. A PCS variant that is able to reductively carboxylate acrylyl-CoA directly into methylmalonyl-CoA would make the ATP-dependent propionyl-CoA carboxylation step dispensable. With five ATP per three CO2 fixed into pyruvate, a 50 % carboxylating PCS variant would save 20 % ATP in the cycle. Even a PCS with a carboxylation efficiency of only 5 % would make a distinct difference in biomass yield. The argument becomes even stronger, when the 3-HPBC is used during mixotrophic growth, e.g., when acetate is assimilated,

where a fully carboxylating PCS variant would save 25 % ATP. Similar arguments hold true for AER_{Nm} that presumably works in the 3-hydroxypropionate / 4-hydroxybutyrate cycle in *N. maritimus*[18] (**Figure S7b**). A fully carboxylating AER_{Nm} would save 25% ATP per two CO_2 converted into acetyl-CoA, while a partially carboxylating enzyme would save biomass yield according to the grade of carboxylation efficiency. In the physiological context of ammonia oxidizing archaea, such as *N. maritimus*, which thrive at a constantly low energy supply[18], it seems unlikely that the sleeping carboxylation function of AER_{Nm} in *N. maritimus* would not play a role in contributing to the lifestyle of these organisms.

Supplementary Figures

Figure S1: PCS sequence alignment of CO_2-binding residues. The *Erythrobacter* sp. NAP1 homolog PCS_{EN} is marked with a blue star. The four essential residues to bind the CO_2 molecule are framed. The residue numbering refers to the PCS_{EN} sequence. The four spheres behind the sequence represent the conservation of the four essential residues (green, conserved; orange, not conserved).

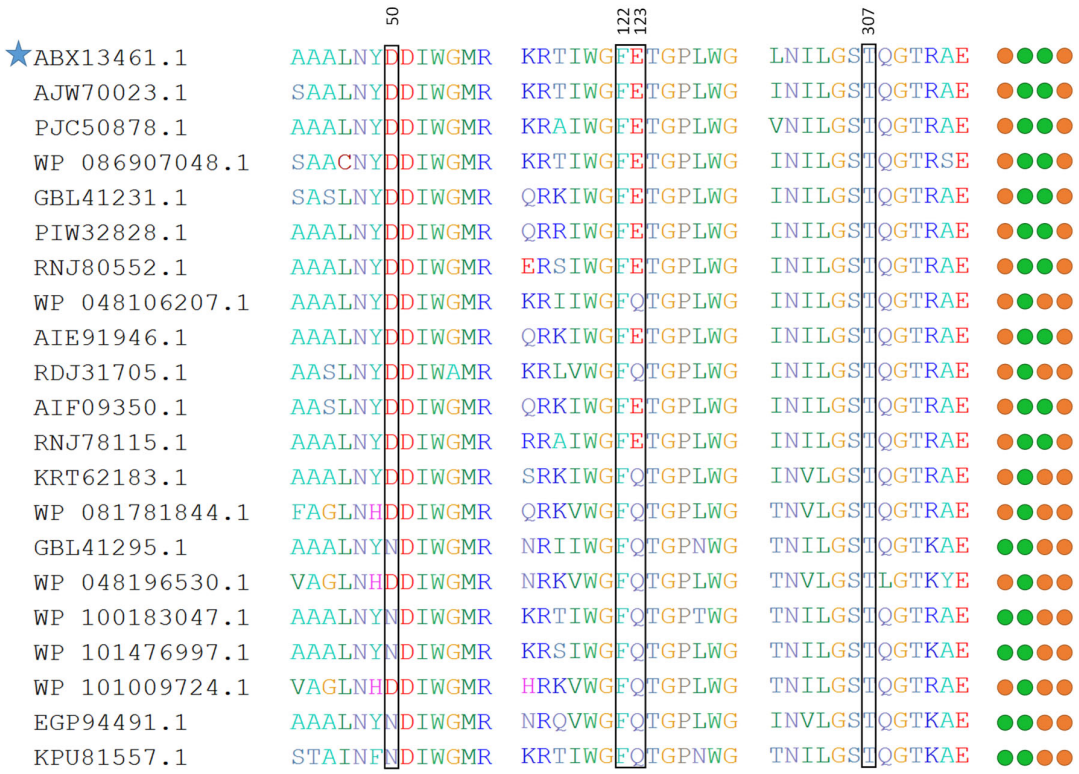


Figure S2: AER sequence alignment of CO_2-binding residues. The *Nitrosopumilus maritimus* homolog AER_{Nm} is marked with a blue star. The four essential residues to bind the CO_2 molecule are framed. The residue numbering refers to the AER_{Nm} sequence. The four spheres behind the sequence represent the conservation of the four essential residues (green, conserved; orange, not conserved).

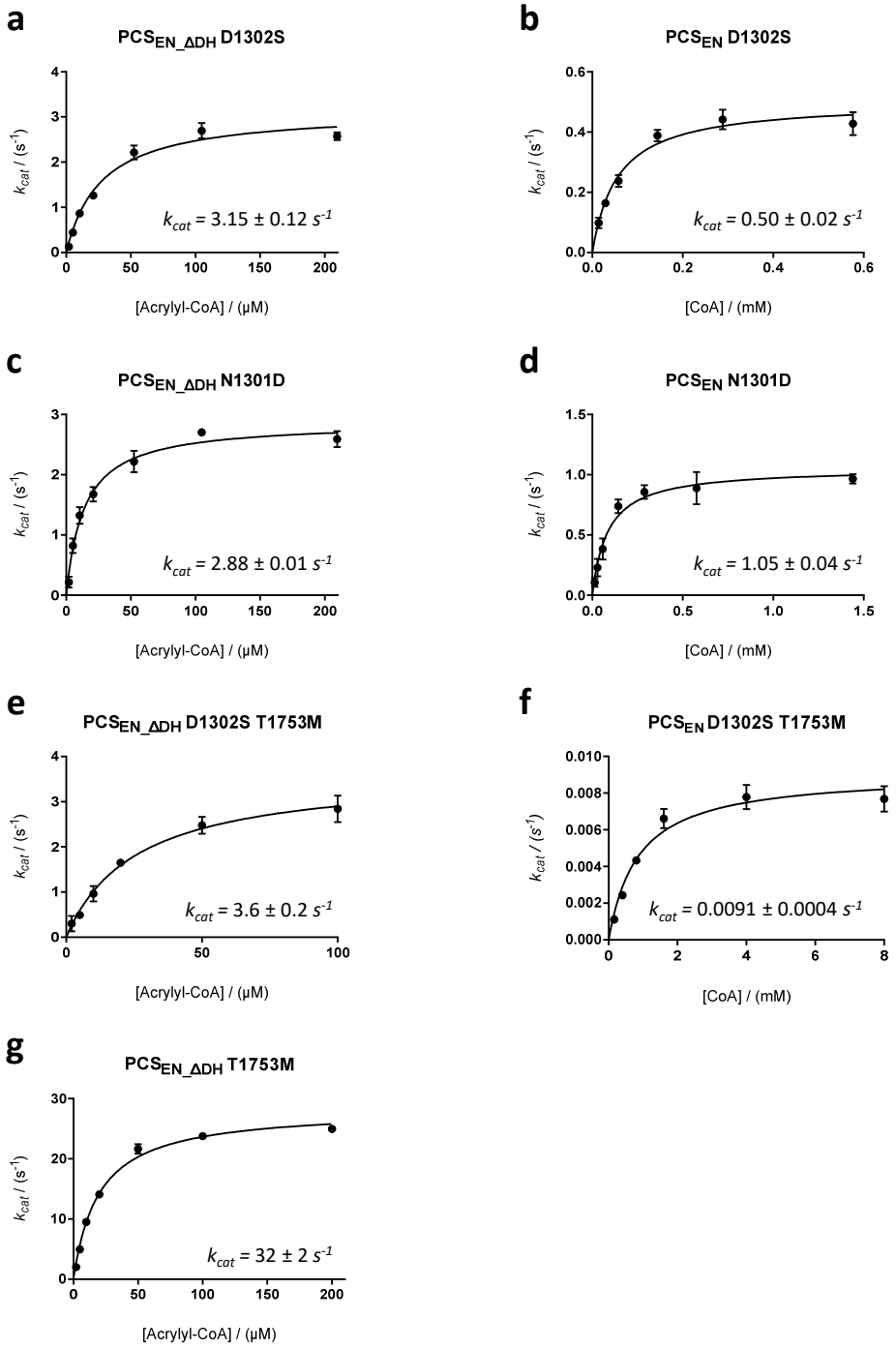


Figure S3: Michaelis-Menten plots of the kinetic characterizations of PCS$_{EN}$ variants. All points were measured in triplicates, mean ± s.d. is schown. For detailed assay conditions see **Table S1**. The reductase domain alone is assayed depending on the acrylyl-CoA concentration in **a, c, e** and **g** using the corresponding variant in a dehydratase deficient mutant. **b. b, d** and **f** depict the turnover number of the overall reaction in dependence of the CoA concentration. Assays were performed with a, PCS$_{EN_\Delta DH}$ D1302S, b, PCS$_{EN}$ D1302S, c, PCS$_{EN_\Delta DH}$ N1301D, d, PCS$_{EN}$ N1301D, e, PCS$_{EN_\Delta DH}$ D1302S T1753M, f, PCS$_{EN}$ D1302S T1753M, g, PCS$_{EN_\Delta DH}$ T1753M.

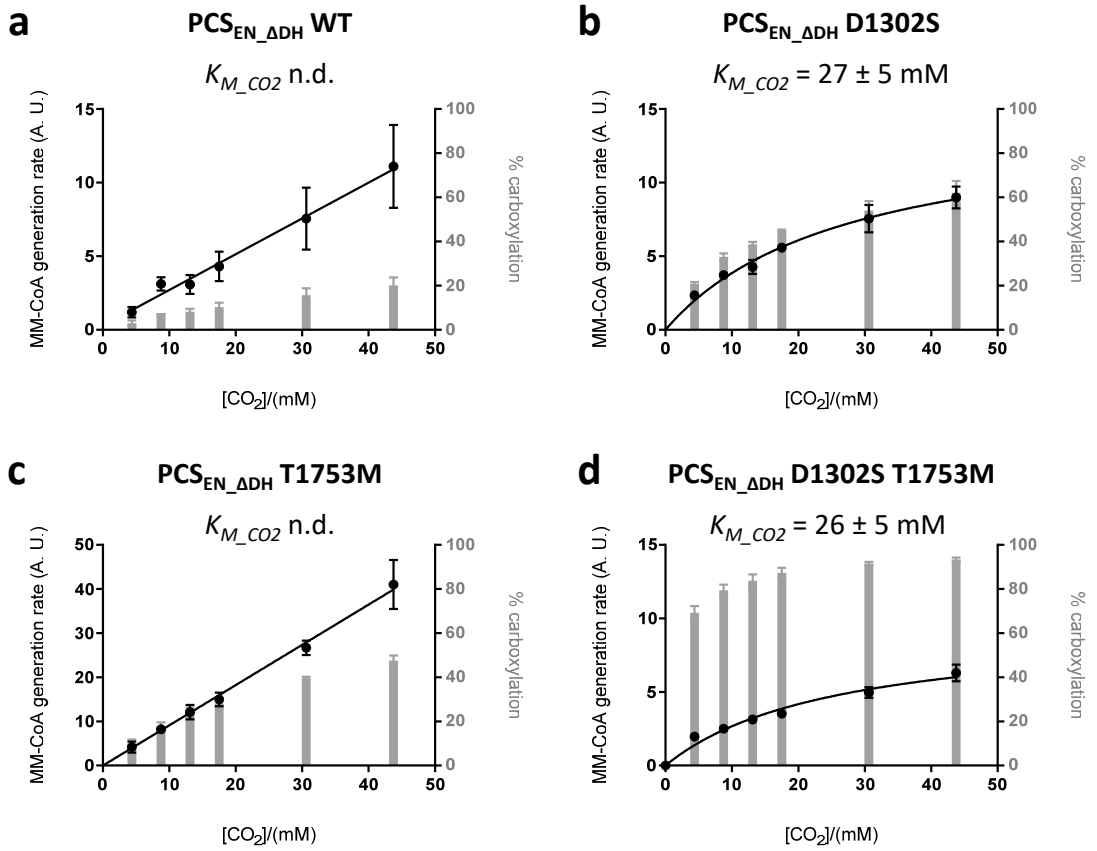


Figure S4: Determination of K_{M_CO2} and carboxylation efficiency of PCS$_{EN}$ variants at different CO_2 concentrations. The methylmalonyl-CoA (MM-CoA) generation rate is represented as black dots, the carboxylation efficiency in grey bars. MM-CoA generation rate over the reaction time course (5 time points) was determined by UHPLC (UV-peak area) measurements in triplicates and was corrected for 1 nM of enzyme. Carboxylation efficiencies are given as mean ± s. d. as calculated from the mean carboxylation ratio over the reaction time course in three replicates. CO_2 concentrations were calculated from bicarbonate concentration. Michaelis-Menten fit was plotted where applicable and the associated K_{M_CO2} is shown above the graph. For detailed assay conditions see **Table S2**. Assays were performed with **a,** PCS$_{EN_\Delta DH}$ WT, **b,** PCS$_{EN_\Delta DH}$ D1302S, **c,** PCS$_{EN_\Delta DH}$ T1753M, **d,** PCS$_{EN_\Delta DH}$ D1302S T1753M in 100 mM Tris buffer (pH 8.0) at fixed concentration of acrylyl-CoA.

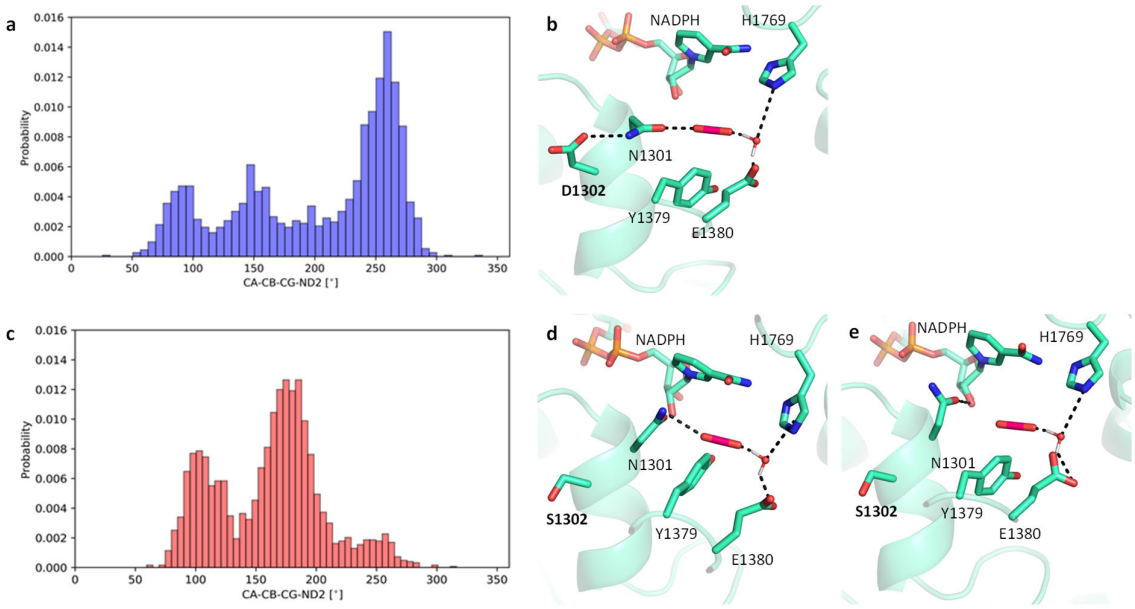


Figure S5: Molecular dynamics simulation of the PCS$_{EN}$ wt and D1302S active site. a shows the probability distribution of the Asn1301 dihedral angle in the PCS$_{EN}$ wt. The model at the dominant Asn1301 dihedral angle (250°) is depicted in **b**. **c** shows the probability distribution of the Asn1301 dihedral angle in the PCS$_{EN}$ D1302S variant. The models of the two predominant Asn1301 dihedral angles, at 100° and 180°, are depicted in d and e, respectively.

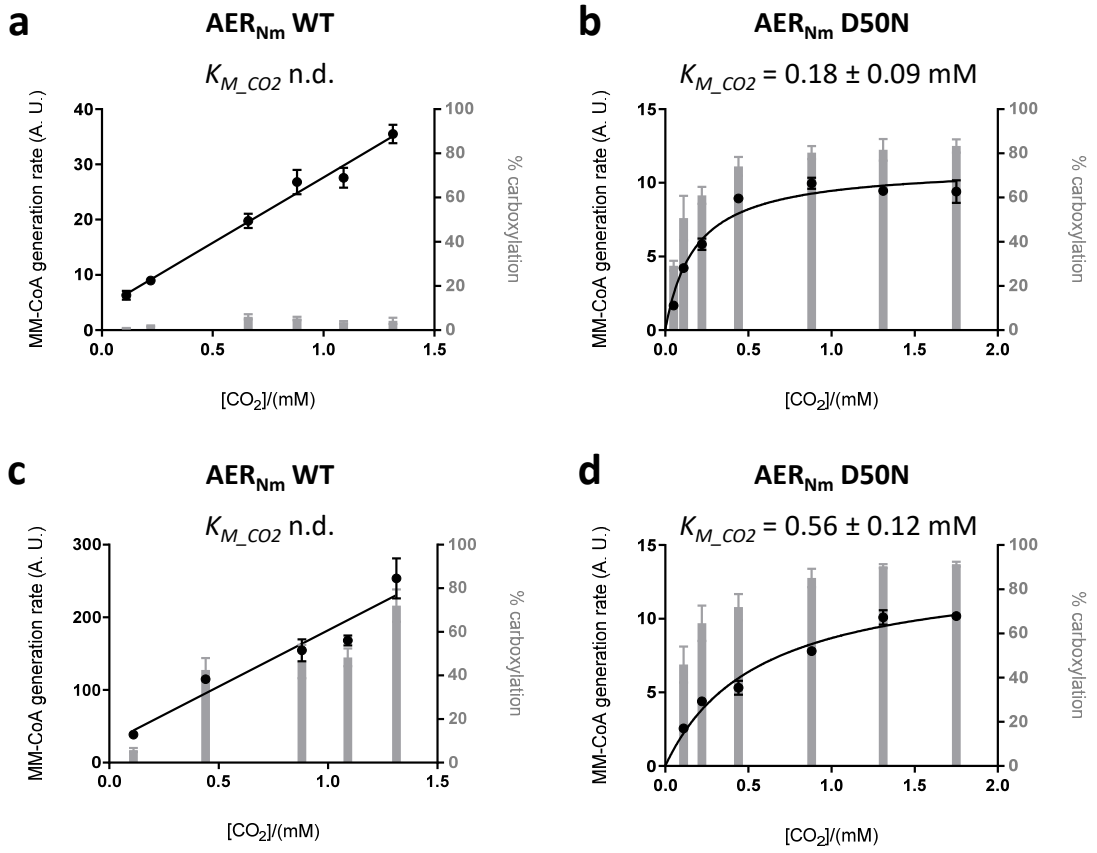


Figure S6: Determination of K_{M_CO2} and carboxylation efficiency of AER$_{Nm}$ variants at different CO2 concentrations. The methylmalonyl-CoA (MM-CoA) generation rate is represented as black dots, the carboxylation efficiency in grey bars. MM-CoA generation rate over the reaction time course (5 time points) was determined by UHPLC (UV-peak area) measurements in triplicates and was corrected for 1 µM of AER$_{Nm}$ WT and for 1 nM of AER$_{Nm}$ D50N. Carboxylation efficiencies are given as mean ± s. d. as calculated from the mean carboxylation ratio over the reaction time course in three replicates. CO_2 concentrations were calculated from bicarbonate concentration. Michaelis-Menten fit was plotted where applicable and the associated K_{M_CO2} is shown above the graph. For detailed assay conditions see **Table S7**. Assays were performed with **a, c,** AER$_{Nm}$ WT, **b, d,** AER$_{Nm}$ D50N in either 100 mM KxHyPO4 (**a,b**) or 20mM Tris/200 mM proline buffer (**c,d**) at fixed concentration of acrylyl-CoA.

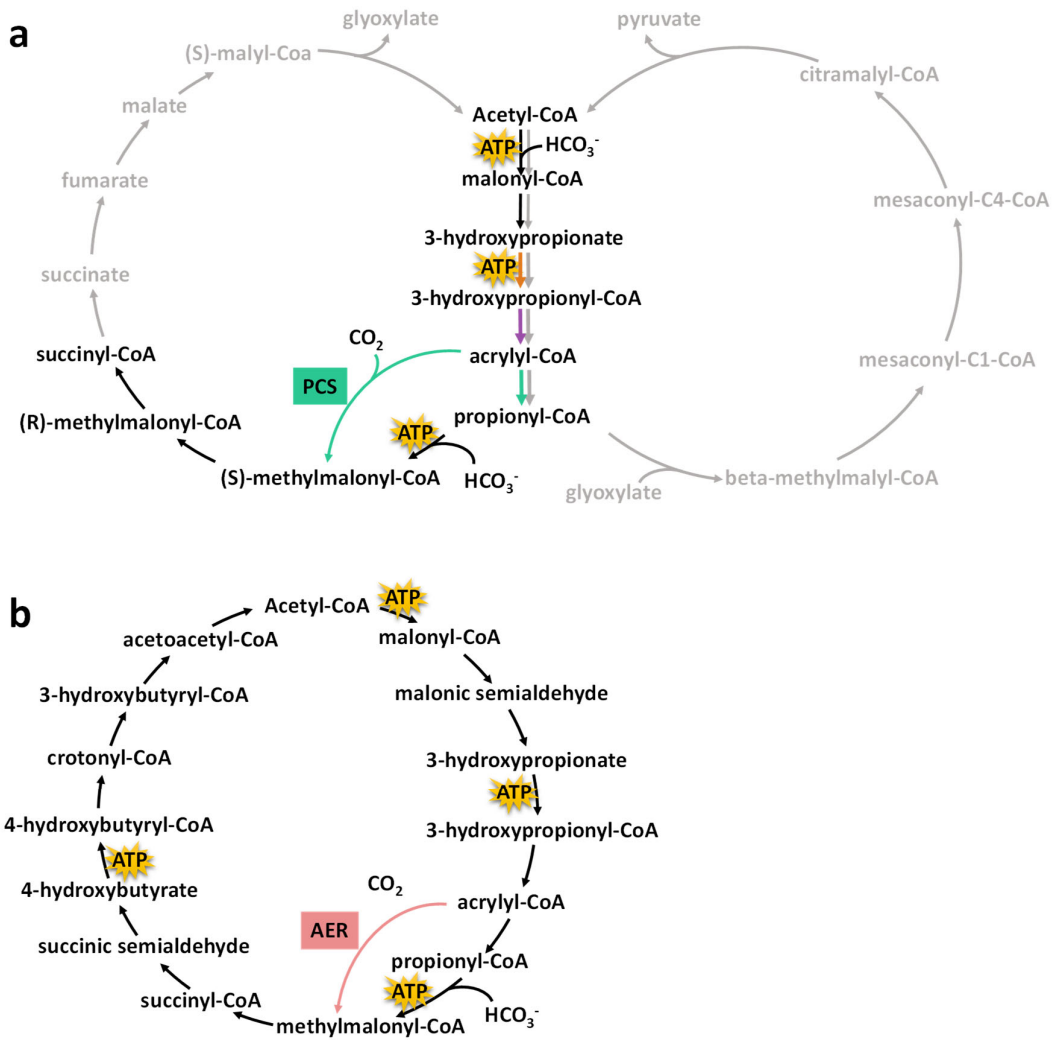


Figure S7: Physiological context of PCS & AER as sleeping carboxylases. a, PCS operates in the 3 hydroxypropionate bicycle (3-HPBC, black and grey reactions) of *C. aurantiacus* and a modified version of the same pathway in *Erythrobacter* sp. NAP1 (black reactions). b, AER_{Nm} presumably operates in the 3-hydroxypropionate / 4-hydroxybutyrate cycle in *N. maritimus*.

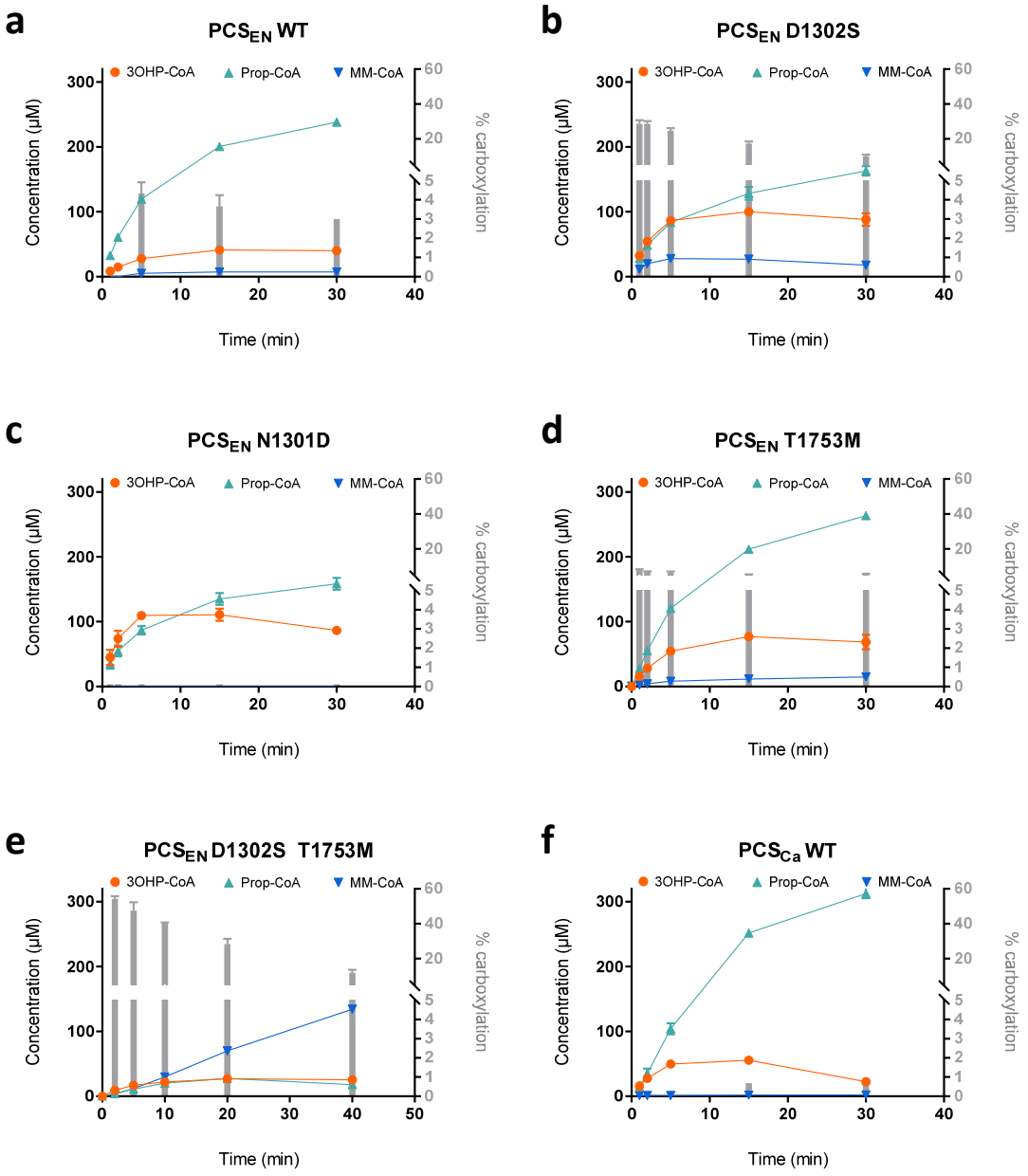


Figure S8: Carboxylation efficiency in the overall reaction of PCS homologs and variants thereof. Time course (3 replicates each) of the overall reaction with 0.8 mM 3-hydroxypropionate, 0.8 mM ATP, 0.8 mM NADPH and 0.5 mM CoA, see **Table S3** for detailed assay conditions. 3-hydroxypropionyl-CoA (3OHP-CoA), propionyl-CoA (Prop-CoA) and methylmalonyl-CoA (MM-CoA) concentrations were determined over a time course by UHPLC using an individual standard series for all three compounds (ranging from 1 – 500 µM). Carboxylation efficiency of **a,** PCS$_{EN}$ WT, **b,** PCS$_{EN}$ D1302S, **c,** PCS$_{EN}$ N1301D, **d,** PCS$_{EN}$ T1753M, **e,** PCS$_{EN}$ D1302S T1753M, **f,** PCS$_{ECaN}$ WT is illustrated as bar plot. data mean ± s.d. (n=3).

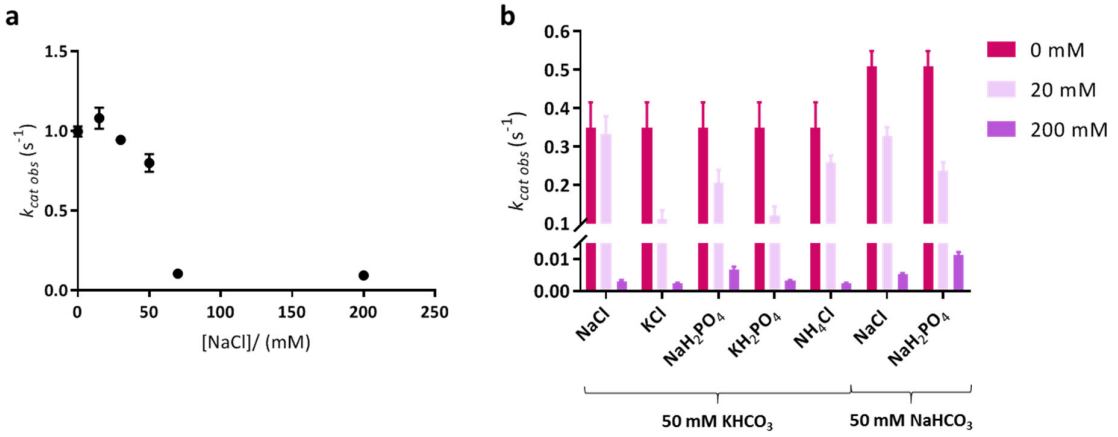


Figure S9: Influence of ionic strength on AER$_{Nm}$. a, k_{cat} in dependence on the NaCl concentration of AER$_{Nm}$ WT in 100 mM $K_xH_yPO_4$ buffer. **b,** Scan for ions with low inhibitory effect on the k_{cat} of AER$_{Nm}$. Experiments were carried out with the AER$_{Nm}$ D50N variant in 50 mM Tris/200 mM proline buffer containing 50 mM $KHCO_3$ or $NaHCO_3$ and 0, 20 or 200 mM of the displayed ions. k_{cat} is given as mean with error bars from 3 replicates measured at fixed concentrations of acrylyl-CoA and NADPH.

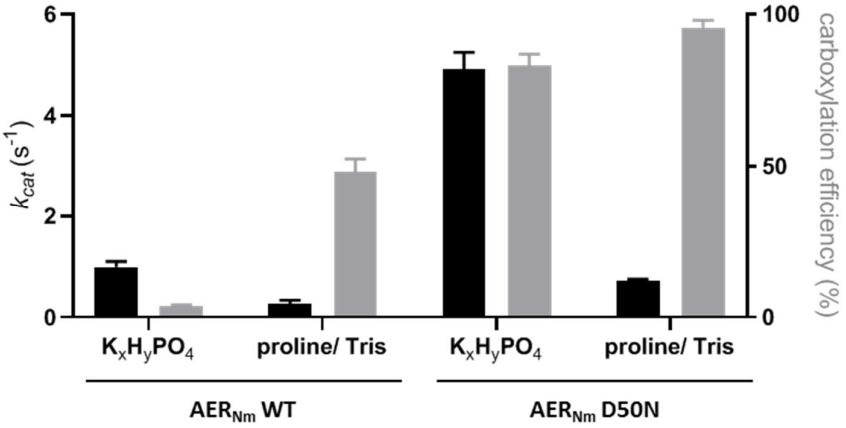


Figure S10: Influence of buffer composition on AER$_{Nm}$ WT and AER$_{Nm}$ D50N carboxylation efficiency and k_{cat}. Experiments were carried out either in 100 mM $K_xH_yPO_4$ (pH 7.5) or 200 mM proline/20 mM Tris buffer (pH 7.5) at fixed concentrations of acrylyl-CoA and NADPH. k_{cat} data is shown as mean with error bars determined from 3 replicates. Carboxylation efficiencies are given as mean ± standard deviation as calculated from the mean carboxylation ratio over the reaction time course (3 time points) in three replicates. See **Table S6** for detailes assay conditions.

Supplementary Tables

Table S1: Reaction parameters and carboxylation ratio for the overall reaction of different PCS homologs and variants thereof.

	Reductase domain reaction only	Three-domain reaction sequence	
PCS variant	app. k_{cat} (s^{-1}) at 1 mM CO_2	app. k_{cat} (s^{-1}) at 1 mM CO_2	% carboxylation at 1 mM CO_2
PCS_{EN} WT	140 ± 20 [1)]	4.7 ± 0.3 [1)]	4.3 ± 0.6
PCS_{EN} D1302S	3.15 ± 0.12	0.46 ± 0.02	29 ± 2
PCS_{EN} N1301D	2.88 ± 0.01	1.06 ± 0.04	n.d.
PCS_{EN} T1753M	32 ± 2	4.6 ± 0.2 [2)]	8.1 ± 0.2
PCS_{EN} D1302S T1753M	3.6 ± 0.2	0.0091 ± 0.0004	54 ± 2
PCS_{Ca} WT	n.d.	8.4 [1)]	0.6846 ± 0.0007

k_{cat} (combined reduction & carboxylation activity) and K_M data are shown as mean ± s. d. as determined from a Michaelis-Menten fit of at least 18 data points (Figure S3). Carboxylation ratios are given as mean ± s. d. as calculated from the mean carboxylation ratio over the reaction time course (5 time points) in three replicates (**Figure S8**). CO_2 concentrations were calculated from bicarbonate concentration. n.d., not determined.

[1)] previously reported value[10,11]

[2)] triplicate measurement at saturating conditions

Table S2: Specific activities of different PCS_{EN} variants

	PCS wt		PCS D1302S		PCS T1753M		PCS D1302S T1753M	
$[CO_2]$	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction
44 mM	1.6 ± 0.3	7.7 ± 0.6	1.3 ± 0.1	1.1 ± 0.3	5.9 ± 0.7	7.6 ± 0.4	0.91 ± 0.07	0.09 ± 0.02
30.8 mM	1.1 ± 0.3	7.0 ± 0.4	1.09 ± 0.11	1.20 ± 0.08	3.9 ±0.2	6.7 ± 0.3	0.72 ± 0.04	0.09 ± 0.02
17.6 mM	0.66 ± 0.12	6.5 ± 0.2	0.81 ± 0.02	1.27 ± 0.08	2.2 ± 0.2	6.4 ± 0.1	0.52 ± 0.03	0.11 ± 0.02
13.2 mM	0.48 ± 0.07	6.0 ± 0.1	0.62 ± 0.06	1.27 ± 0.05	1.8 ± 0.2	6.4 ± 0.3	0.46 ± 0.03	0.13 ± 0.03
8.8 mM	0.48 ± 0.05	7.5 ± 1.0	0.54 ± 0.02	1.362 ± 0.004	1.22 ± 0.04	6.1 ± 0.1	0.37 ± 0.03	0.13 ± 0.01
4.4 mM	0.21 ± 0.04	7.2 ± 1.0	0.34 ± 0.02	1.43 ± 0.07	0.6 ± 0.2	5.9 ±0.4	0.29 ± 0.01	0.17 ± 0.02

Specific activities of the reduction and carboxylation reaction, calculated from the change in UV peak over time using a previous standard curve. Note, the low k_{cat} values are due to the inhibition by high concentrations of potassium.

Table S3: Specific activities of different AER_{Nm} variants in phosphate buffer

	AER_{Nm} WT		AER_{Nm} D50N	
$[CO_2]$	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction	k_{cat} (s^{-1}) carboxylation	k_{cat} (s^{-1}) reduction
1.31 mM	0.0056 ± 0.0002	0.078 ± 0.010	1.36 ± 0.11	0.29 ± 0.07
1.09 mM	0.0045 ± 0.0002	0.081 ± 0.008	1.37 ± 0.04	0.33 ± 0.06
0.88 mM	0.0044 ± 0.0003	0.090 ± 0.003	1.45 ± 0.05	0.36 ± 0.03
0.66 mM	0.0034 ± 0.0002	0.10 ± 0.01	1.30 ± 0.03	0.43 ± 0.06
0.22 mM	0.00180 ± 0.00004	0.134 ± 0.002	0.85 ± 0.06	0.54 ± 0.06
0.11 mM	0.00141 ± 0.00009	0.136 ± 0.001	0.62 ± 0.09	0.60 ± 0.12

Specific activities of the reduction and carboxylation reaction, calculated from the change in UV peak over time using a previous standard curve.
Note, the low k_{cat} values are due to the inhibition by high ionic strength.

Table S4: Detailed assay conditions to determine the kinetic parameters for the different PCS_{EN} variants.

	overall reaction	reductase reaction	ligase reaction
Tris buffer pH 8 (mM)	100	100	100
KHCO3 (mM)	50	50	50
KCl (mM)	40	40	40
MgCl2 (mM)	10	10	10
3-hydroxypropionate (mM)	2.1		2.1
CoA (mM)	X		8
ATP (mM)	3		3
NADPH (mM)	0.3	0.3	
NADH (mM)			0.3
acrylyl-CoA (mM)		X	
PCSEN variant (nM)	1)	2)	3)
Carbonic anhydrase (nM)	69	69	69
Phosphoenolpyruvate (mM)			1
PK/LDH (U)			~ 2.4 / 3.5
Myokinase (µM)			2.7

Triplicates were measured at six different concentrations of substrate X. The ligase reaction was measured in triplicate at a fixed concentration of all substrates. PK/LDH, mixture of pyruvate kinase and lactate dehydrogenase (Sigma-Aldrich P0294).
[1] 676 nM of PCS_{EN} D1302S, 227 nM of PCS_{EN} N1301D, 9282 nM of PCS_{EN} D1302S T1753M
[2] 97 nM of $PCS_{EN\_\Delta DH}$ D1302S, 82 nM of $PCS_{EN\_\Delta DH}$ N1301D, 11 nM of $PCS_{EN\_\Delta DH}$ T1753M, 101 nM of $PCS_{EN\_\Delta DH}$ D1302S T1753M
[3] 9282 nM of PCS_{EN} D1302S T1753M

Table S5: Detailed assay conditions to determine the K_{M_CO2} for the different PCS_{EN} variants.

Tris buffer pH 8 (mM)	100
$KHCO_3$ (mM)	x
Carbonic anhydrase (nM)	69
NADPH (mM)	2
KCl (mM)	2000 - x
3OHP-CoA (mM)	2
$PCS_{EN\_\Delta DH}$ variant (µM)	[1]
PhaJ (µM)	1.5

Triplicates were measured at six different CO_2 concentrations, each replicate comprising six time points to determine the slope for methylmalonyl-CoA and propionyl-CoA generation by UHPLC analysis.
[1] 0.12 µM of $PCS_{EN\_\Delta DH}$, 0.44 µM of $PCS_{EN\_\Delta DH}$ D1302S, 0.10 µM of $PCS_{EN\_\Delta DH}$ T1753M, 0.43 µM of $PCS_{EN\_\Delta DH}$ D1302S T1753M

Table S6: Assay conditions for the PCS_{EN} overall reactions to determine the carboxylation efficiency.

Tris / Pi buffer pH 8 (mM) [1)]	100
$KHCO_3$ (mM)	50
KCl (mM)	40
$MgCl_2$ (mM)	10
3-hydroxypropionate (mM)	0.8
CoA (mM)	0.5
ATP (mM)	0.8
NADPH (mM)	0.8
PCS_{EN} variant (µM)	[2)]
Carbonic anhydrase (nM)	69

[1)] PCS_{EN} WT, PCS_{EN} D1302S, PCS_{EN} N1301D were measured in phosphate buffer. PCS_{EN} T1753M, PCS_{EN} D1302S T1753M and PCS_{Ca} were measured in Tris buffer.
[2)] 0.15 µM of PCS_{EN} WT, 1.5 µM of PCS_{EN} D1302S, 0.75 µM of PCS_{EN} N1301D, 11.6 µM of PCS_{EN} D1302S T1753M, 0.075 µM of PCS_{Ca}

Table S7: Assay conditions for the k_{cat} determination of AER_{Nm} variants.

NADPH (mM)	0.6
Acrylyl-CoA (mM)	0.5
NaHCO3 (mM)	50
Carbonic anhydrase (nM)	69
AER_{Nm}	[1)]

in phosphate or proline buffer.
[1)] 0.29 µM of AER_{Nm} WT in proline buffer (20 mM Tris/200mM proline, pH 7.5), 0.095 µM of AER_{Nm} WT in phosphate buffer (100 mM $K_xH_yPO_4$, pH 7.5), 0.053 µM of AER_{Nm} D50N in phosphate buffer, 1.4 µM of AER_{Nm} D50N in proline buffer

Table S8: Composition of assays to determine the influence of ionic strength on AER_{Nm}.

NADPH (mM)	0.3
OH-propCoA (mM)	0.5
$KaHCO_3/NaHCO_3$ (mM)	50
Carbonic anhydrase (nM)	69
PhaJ (µM)	0.4
AER_{Nm}	[1]

[1] 0.55 µM of AER_{Nm} WT in phosphate buffer (100 mM $K_xH_yPO_4$, pH 7.5), 0.39 µM of AER_{Nm} D50N in proline buffer (20 mM Tris/200 mM proline, pH 7.5)

Table S9: Detailed assay conditions to determine K_{M_CO2} for AER_{Nm} variants.

	AER_{Nm} WT	AER_{Nm} D50N
NADPH (mM)	2	2
3-OH propionyl-CoA (mM)	2	2
NaHCO3 (mM)	x	x
NaCl (mM)	60 - x	80 - x
Carbonic anhydrase (nM)	69	69
PhaJ		
AER_{Nm}	[2]	[2]

$NaHCO_3$ concentrations ranging from 0 – 60 mM in assays with the AER_{Nm} wt and ranging from 0 – 80 mM with the AER_{Nm} D50N variant.

[1] PhaJ concentration was two times the concentration of AER_{Nm}.

[2]concentration of AER_{Nm} WT in proline buffer (20 mM Tris/200mM proline) and phosphate buffer (100 mM KHPO4) was 7.8 µM. AER_{Nm} D50N concentration in Phosphate buffer was 0.33 µM and was 0.22 µM in proline buffer.

Chapter IV

Propionyl-CoA synthase module in *Erythrobacter* sp. NAP1: a light-induced pathway to maintain the redox homeostasis

Authors:
Iria Bernhardsgrütter, Timo Glatter, Tobias J. Erb

In preparation

Author contributions:
I.B. and T.J.E. conceived the project. I.B. and T.J.E. designed the experiments and analyzed the data. I.B. performed all experiments and acquired the data. T.G. performed mass spectrometry for proteomics. I.B. and T.J.E. wrote the manuscript with contributions from T.G..

4 Propionyl-CoA synthase module in *Erythrobacter* sp. NAP1: a light-induced pathway to maintain the redox homeostasis

4.1 Abstract

As photosynthetic organisms, aerobic anoxygenic phototrophic (AAP) bacteria are facing the challenge of over-reduction by light-driven electron transfer. Despite aerobic growth with oxygen serving as electron acceptor, several AAP bacteria respond to the surplus of reducing equivalents with the upregulation of a reductive metabolic pathway. Here, we describe a putative propionyl-CoA synthase (Pcs) module, a part of the 3-hydroxypropionate bi-cycle, in the AAP bacterium *Erythrobacter* sp. NAP1. Bioinformatics and proteomic approaches confirmed that *Erythrobacter* sp. NAP1 encodes and expresses all the genes required for the postulated Pcs module. A significant light-dependent upregulation of the key enzymes in *Erythrobacter* sp. NAP1 suggests the Pcs module to be involved in redox homeostasis. This finding would be in line with the phylogenetic distribution of the two key enzymes of the Pcs module among potential phototrophic microorganisms.

4.2 Introduction

Propionyl-CoA synthase (Pcs) and malonyl-CoA reductase (Mcr) have been described as key enzymes of the autotrophic 3-hydroxypropionate (3HP) bi-cycle in the green non-sulfur bacterium *Chloroflexus aurantiacus*[1,2] (**Figure 1**). Mcr reduces malonyl-CoA to 3-hydroxypropionate in a two-step reaction using two equivalents of NADPH. Pcs catalyzes the subsequent three-step reaction sequence from 3-hydroxypropionate to propionyl-CoA, which is the branching point metabolite between the first and the second part cycle[1]. Together, the two cycles generate the central metabolite pyruvate from CO_2, while regenerating the initial acceptor molecule acetyl-CoA. Although Pcs and Mcr are key to the mixotrophic and (facultative) photoautotrophic lifestyle of *C. aurantiacus*[3], the enzymes are notably spread beyond the phylum of Chloroflexi and are encoded in a variety of heterotrophic microorganisms.

Strikingly, all of the Pcs-encoding microorganisms are potential phototrophs and most of them belong to the group of aerobic anoxygenic phototrophic (AAP) bacteria. AAP bacteria are obligate aerobes that metabolize organic carbon for growth, while additionally utilizing light for energy conservation. AAP bacteria account for up to 15% of the total microbial community in the upper ocean[4-6] and contribute significantly to the marine carbon cycle. However, the physiology and metabolic repertoire of these ecologically relevant microorganisms is still poorly understood.

Our finding that Pcs is encoded by a variety of AAP bacteria encouraged us to study the role of this multi-domain enzyme in these organisms. We chose *Erythrobacter* sp. NAP1 as our model, which had been isolated from Atlantic surface water, characterized and sequenced[7,8]. We previously described the detailed structure and function of the *Erythrobacter* sp. NAP1 Pcs homologue as a multi-domain synchronized reaction-chamber[9]. In this study, we investigated the potential physiological role and regulation of Pcs in *Erythrobacter* sp. NAP1 using growth studies and whole cell proteomics.

4.3 Results

To understand the physiological role of Pcs and Mcr we first investigated the evolution and phylogenetic distribution of the two genes (**Figure S1** and **S2**). Interestingly, while the two genes seem to have spread as a unit in most cases, some organisms encode one of the two key enzymes independent of the other (**Supplementary Information I**); e.g., *Chloracidobacterium thermophilum* strain B (referred to as *C. thermophilum*) carries the *mcr* but not the *pcs* gene. However, a closer inspection of the *C. thermophilum* genome reveals two genes on different loci, one homologous to the ligase domain (56 % amino acid sequence identity, 97 % coverage) and the other to the reductase domain (60 % amino acid sequence identity, 99 % coverage) of Pcs. Moreover, the C-terminal and the N-terminal end of the *C. thermophilum* ligase and reductase homologues, respectively, align further into the dehydratase domain of Pcs. Together the two sequences cover the complete Pcs sequence (**Figure S3** and **S4**). The two *C. thermophilum* sequences clustered within the Pcs full-length sequences, when we phylogenetically analyzed the individual Pcs domains compared to their stand-alone homologues (**Figure S5**, **Supplementary Information II**). It is unclear, whether these sequences represent an intermediate stage on the evolutionary track of Pcs or if they represent the truncated remnants of a full-length Pcs sequence.

Bioinformatics analysis of the *C. thermophilum* genome revealed homologous genes for a partial 3HP bi-cycle able to convert acetyl-CoA and two bicarbonate molecules into succinyl-CoA. The same set of homologous genes was identified in the genome of *Erythrobacter* sp. NAP1, where they are arranged in two gene clusters, as noted before[2,3] (**Figure 1a**). One putative cluster encoded Pcs and Mcr, while the other one encoded propionyl-CoA carboxylase, as well as methylmalonyl-CoA epimerase and mutase. Homologues of all other genes of the 3HP-bi-cycle were absent. Altogether, these data suggested the distribution of an incomplete 3HP bi-cycle, a linear Pcs module that allows the transformation of acetyl-CoA into succinyl-CoA (**Figure 1b**). This finding is in line with the heterotrophic lifestyle of *C. thermophilum* and *Erythrobacter* sp. NAP1.

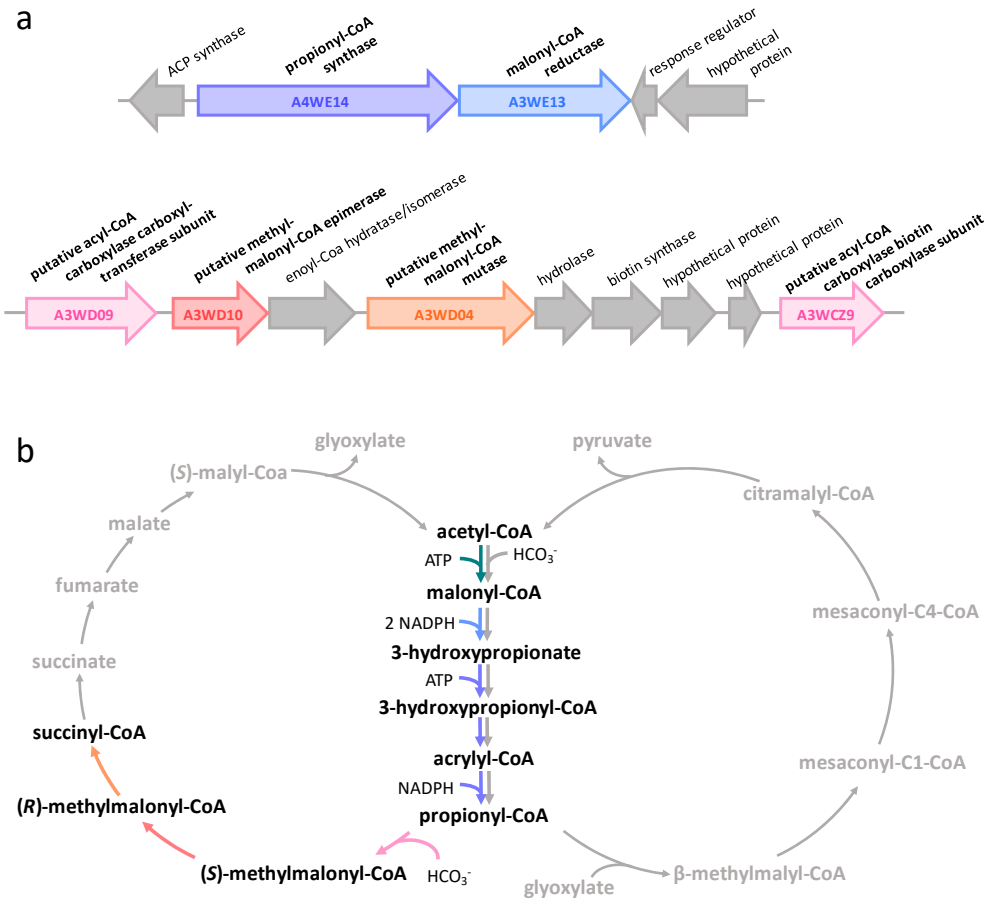


Figure 1: Putative Pcs module in *Erythrobacter* sp. NAP1. a, two *Erythrobacter* sp. NAP1 gene clusters encode for five out of the six Pcs module enzymes, illustrated by arrows labelled by the respective UniProt ID. **b,** 3-hydroxypropionate bi-cycle with the putative Pcs module enzymes highlighted in color and metabolites in black. The enzymes are color coded: acetyl-CoA carboxylase, teal; Mcr, blue; Pcs, purple; propionyl-CoA carboxylase, pink; methylmalonyl-CoA epimerase, red; methylmalonyl-CoA mutase, orange.

Using *Erythrobacter* sp. NAP1, we assessed whether the Pcs module might be involved in the assimilation of one of its intermediates (e.g., acetate, 3-hydroxypropionate or propionate). We first tested the growth of *Erythrobacter* sp. NAP1 on various carbon sources. While growth on glucose and acetate could be observed with growth rates of 0.05 ± 0.02 h^{-1} and 0.08 ± 0.02 h^{-1}, respectively, no growth was detected in minimal medium containing either 3-hydroxypropionate or propionate alone (**Figure 2**). Cells grown on acetate showed formation of intracellular storage compounds, likely polyhydroxyalkanoates granules (**Figure S6**).

We next asked whether acetate is assimilated via the Pcs module. Notably, *Erythrobacter* sp. NAP1 also encodes the enzymes of the glyoxylate cycle (isocitrate lyase (Icl) and malate synthase (Ms)), a well-studied acetate assimilation pathway. To distinguish the contribution of the two pathways, we sought to suppress the glyoxylate cycle with 3-nitropropionate (3-NP), a specific inhibitor of isocitrate lyase[10]. When testing growth of *Erythrobacter* sp. NAP1 in minimal medium containing

either glucose or acetate as carbon source and different concentrations of 3-NP (**Figure 2**), we observed that growth on acetate was completely abolished at very low 3-NP concentrations (30 µM). Notably, in glucose medium, the growth rate was unaffected at 30 µM 3-NP and only slightly affected at higher 3-NP concentrations. The lag phase prolonged with increasing 3-NP concentrations. This finding indicated that acetate assimilation proceeds via the glyoxylate cycle and that the Pcs module is not able to functionally replace the glyoxylate cycle for growth on acetate.

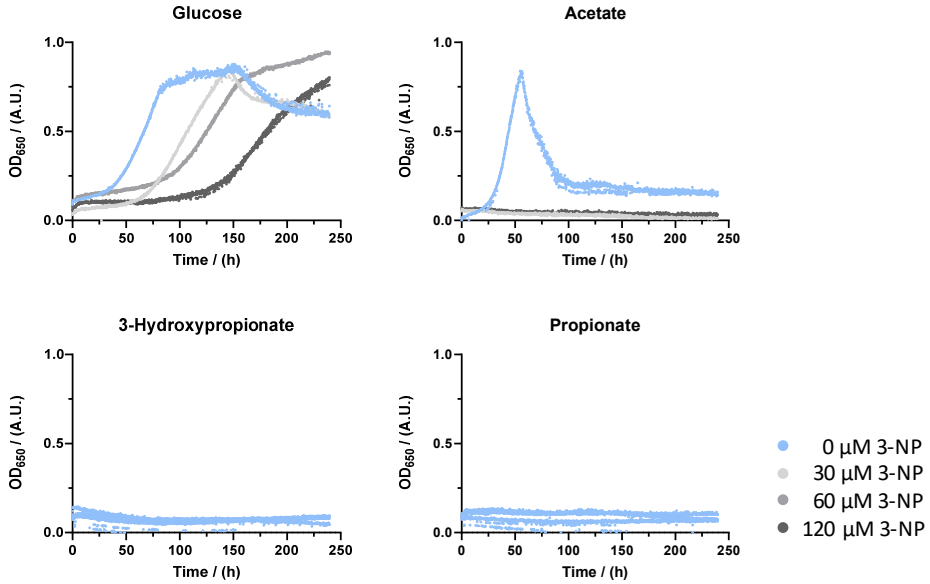


Figure 2: Influence of 3-nitropropionate on growth of *Erythrobacter* sp. NAP1 on different C-sources. *Erythrobacter* sp. NAP1 was able to grow on minimal medium with glucose or acetate as the only carbon source. 3-hydroxypropionate or propionate could not promote growth of *Erythrobacter* sp. NAP1. 3-NP extended the lag phase and slightly impaired the growth rate of *Erythrobacter* sp. NAP1 on glucose and completely abolished growth on acetate.

We then studied the protein abundance pattern of the Pcs module on glucose or acetate using shotgun proteomics and label-free quantification (**Figure S7a, Table S1**). The relative abundance of the Pcs module enzymes was inconsistent across experiments and seemed to be regulated independently of the respective carbon source. In contrast, the glyoxylate cycle was clearly upregulated (Icl: 24.3 ± 1.7 fold and Ms: 3 ± 0.4 fold) when cells were grown on acetate. Note, that in *Erythrobacter* sp. NAP1 the genes encoding for Icl and Ms are not organized in an operon, which explains the different magnitude of upregulation.

Notably, it had been shown for other AAP bacteria, that some metabolic pathways are tightly linked to the presence of light (or: are regulated by illumination)[7,11-13]. Generally, these pathways, i.e., the ethylmalonyl-CoA, polyhydroxyalkanoate (PHA) biosynthesis or anaplerotic CO_2-fixation pathways, share the common feature of being reductive pathways. To test whether light also regulates the abundance of the Pcs module enzymes, we cultivated *Erythrobacter* sp. NAP1 in batch cultures in

two incubators with an anti-parallel light/dark cycle and analyzed the proteome at specific time points (**Figure S8**). A comparison of the proteomes of cells that were incubated for 2 hours in light versus 2 hours in the dark is illustrated in **Figure 3a**. It is apparent that the key enzymes Pcs and Mcr are significantly more abundant when the cells were harvested during the light phase. Notably, the highest levels of Pcs and Mcr were detected in the beginning of the light phase, which decreased over time (**Figure 3b**). All other predicted enzymes of the Pcs module were detected but did not show a significant change in protein levels depending on the light regime (**Figure S7b, Table S2**). Significant changes were also observed in many proteins that are part of the light-harvesting complexes, which were upregulated in the light. An exception are the enzymes involved in bacteriochlorophyll *a* biosynthesis that are less abundant in light, as reported before[14-16].

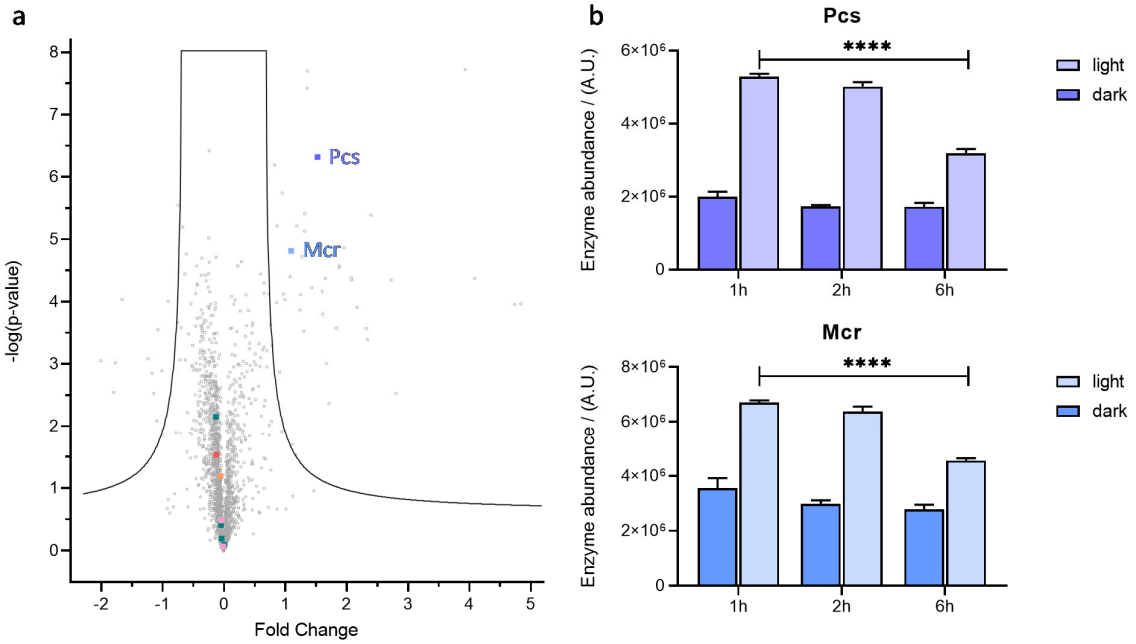


Figure 3: Differential abundance of Pcs module key enzymes Pcs and Mcr depending on light. a, volcano plot comparing the proteome of *Erythrobacter* sp. NAP1 grown for 2 hours in light versus 2 hours in the dark from cultures grown at a 12h light/dark cycle. Samples were prepared in triplicates per time point and light condition. 1998 proteins were quantified by shotgun LC-MS proteomics. The significance cutoff (black curve) was set with s0 = 0.5 and FDR = 0.001 and resulted in 56 proteins with a significantly different abundance (**Table S2**). Pcs module enzymes are highlighted using the same color code as in Figure 2. Pcs and Mcr are significantly more abundant in cells grown with light. **b,** Pcs and Mcr levels in *Erythrobacter* sp. NAP1 grown for the defined time at light/dark in a 12h light/dark cycle. Multiple t-tests were applied to probe significance. The difference between the protein levels in the light versus dark is significant for Pcs and Mcr at every time point with $p<0.001$. The decrease in Pcs and Mcr in light over time (1 h light to 6 h light) is significant with $p<0.0001$ (marked with four asterisks).

We performed growth experiments to test whether an upregulation of the Pcs module in light entailed higher flux through this pathway and thus promoted growth. We therefore let *Erythrobacter* sp. NAP1 grow in minimal medium with acetate in either a 12 h/12 h light/dark cycle or in the continuous dark. While growth was not directly promoted in cultures grown in a light/dark cycle compared to cultures

kept in the dark, there seemed to be an advantage in maintaining cell viability in nutrient-scarce conditions (**Figure S9**).

When comparing *Erythrobacter* sp. NAP1 grown on glucose or on acetate, we noticed differences in cell morphology. Transmission electron micrographs revealed substantial PHA granules within the cells grown on acetate (**Figure S6**). In line with this finding, significantly increased levels of a Phasin protein and increased levels of the β-ketothiolase (PhaA) and the acetoacetyl-CoA reductase (PhaB) were detected in cells grown on acetate compared to glucose (**Figure S10**). The PHA synthase (PhaC) on the other hand, was present at similar levels on both carbon sources. It has been previously described that PHA production in an AAP bacterium was increased under dark-light cycle conditions compared to continuous dark conditions[17]. However, in *Erythrobacter* sp. NAP1 no effect of the light regime on the abundance of PHA-related proteins could be observed in minimal medium containing acetate (**Figure S10**).

4.4 Discussion

Here we present a putative Pcs module, a linear part of the 3-hydroxypropionate bi-cycle that allows the transformation of acetyl-CoA into succinyl-CoA. The distribution of the two key enzymes in potential phototrophs might hint towards an early evolution of this pathway in photosynthetic organisms. This hypothesis might be supported by the identification of all Pcs module enzyme homologues in *C. thermophilum.* In contrast to Chloroflexi and all AAP bacteria, the microaerophilic *C. thermophilum* contains a type-I reaction center and was shown to be obligate photoheterotrophic as it required light and organic carbon sources to grow[18,19]. Furthermore, the *pcs* gene is split into two separate coding sequences, of which the gene products should be biochemically probed for activity and potential complex formation.

Using *Erythrobacter* sp. NAP1 we investigated the role of the putative Pcs module. As we showed by growth experiments with the inhibitor 3-NP, this module is not sufficient for acetate assimilation. Instead, acetate is assimilated through the canonical glyoxylate cycle. This is in contrast to *Paracoccus denitrificans* that uses two different pathways, namely the glyoxylate cycle and the ethylmalonyl-CoA pathway, to assimilate acetate[20]. This genetic redundancy allows optimal growth in a changing environment. The protein abundance of the Pcs module key enzymes was largely unchanged with different carbon sources. In contrast, we could show a light-dependent increase in Pcs and Mcr protein levels. Similarly, the abundance of these two key enzymes has also been shown to be upregulated in *Chloroflexus aurantiacus* when comparing photo(hetero)trophic to mixotrophic growth but did not change much between the different organic carbon sources provided in

photoheterotrophic growth[3]. The previously reported significant enhancement of CO_2-fixation in light in *Erythrobacter* sp. NAP1[7,13] might thus be linked to the higher activity of the Pcs module with its two carboxylases (acetyl-CoA carboxylase and propionyl-CoA carboxylase). Carbon fixation in light might also be the underlying mechanisms of prolonged cell viability, which we observed under nutrient-limiting conditions. This phenomenon was previously studied in different AAP bacteria and was even reported to occur in the absence of any organic carbon source[21,22].

The reason for the light-dependent regulation of the Pcs module remains subject of future research; some speculations are presented here. In photoheterotrophic growth, organic compounds deliver carbon and reducing power while light can additionally be used to conserve energy. Consequently, when grown under light conditions, the cell might experience a surplus of energy and reducing equivalents. One would assume that *Erythrobacter* sp. NAP1, an aerobic photoheterotrophic bacterium, could simply use oxygen as an electron sink. However, it has been described before that AAP bacteria decrease their respiration rate under light conditions[7,11,12]. Therefore, we hypothesize that the Pcs module might play a key role in re-stabilizing the redox homeostasis in *Erythrobacter* sp. NAP1. The two carboxylases of the Pcs module, acetyl-CoA and propionyl-CoA carboxylase, might thus be classified as redox-balancing carboxylases[23].

Erythrobacter sp. NAP1 is equipped with another potential redox-balancing pathway, the PHA biosynthesis pathway. Interestingly, the abundance of this pathway's enzymes changed depending on the carbon source but not on the light condition. The enhancing effect of acetate as carbon source on PHA biosynthesis has been reported previously for *Rhodobacter sphaeroides* and several other bacteria[24-27]. The reason behind this effect is yet to be investigated. A possible explanation for the enhanced PHA biosynthesis on acetate by *Erythrobacter* sp. NAP1 might be that assimilating acetate by the glyoxylate cycle generates NADH and might thus be another source of reductive stress. *Erythrobacter* sp. NAP1 might differentiate between a surplus of reducing equivalents originating from carbon metabolism or photosynthesis and responds with an upregulation of either the PHA biosynthesis pathway or the Pcs module, respectively. In contrast, the activation of both the reductive ethylmalonyl-CoA pathway as well as the PHA biosynthesis have been previously reported as a response to the redox imbalance caused by photosynthesis in *Dinoroseobacter shibae*[12,17]. Comparing these different pathways, differences in terms of bioenergetics, CO_2-fixation and consumption of reducing equivalents become obvious (**Table 1**). The Pcs module re-oxidizes the most equivalents of NADPH per converted acetyl-CoA but also requires the most energy input in form of ATP, which might be negligible under energy-rich photosynthetic conditions. The lower NADPH re-oxidation potential of the EMCP and PHA biosynthesis pathway might explain, why *D. shibae* has to upregulate

both reductive pathways in response to light. The advantage of the EMCP and the Pcs module is the co-assimilation of CO_2, which could significantly enhance growth in nutrient-scarce environments.

Table 1: Comparison of the PHA biosynthesis pathway, the EMCP and the Pcs module. Consumption of NADPH, ATP and co-assimilation of CO_2 are given in equivalents per converted acetyl-CoA into the pathways final products.

Equivalents per converted actely-CoA	NADPH	CO_2	ATP
PHA biosynthesis	1/2		
EMCP	2/3	2/3	
Pcs module	3	2/3	3

In anaerobic phototrophs, the requirement for reductive metabolic pathways has been studied in detail. For example, the upregulation of the 3-hydroxypropinate bi-cycle key enzymes Pcs and Mcr has been shown in *C. aurantiacus* when generally comparing phototrophic to chemotrophic growth[3,28,29]. In purple non-sulfur bacteria, close relatives and implied ancestors of AAP bacteria[30,31], the Calvin-Benson-Bassham cycle or the reductive tricarboxylic acid cycle are absolutely required to maintain the redox homeostasis during photosynthesis[32-34]. Even though the requirement for such reductive metabolic pathways is not expected to be found in aerobic phototrophs, it seems to be a general theme in all phototrophic bacteria. Potentially, the funneling of electrons into CO_2-co-assimilating pathways rather than respiration could serve as a strategy to maximize growth in an environment of scarce organic carbon sources like the upper ocean. The evolution of the different pathways in different phototrophic lineages and their ecological role are still to be elucidated. Extensive research efforts are necessary to further understand the function and regulation of these pathways.

4.5 Methods

Culture conditions

The *Erythrobacter* sp. NAP1 strain was kindly provided by the Koblížek lab. The minimal medium for growth of *Erythrobacter* sp. NAP1 was composed of 30 g/L hw Marinemix professional (Wiegandt GmbH, Krefeld, Germany) salt mix, 1 mM $(NH_4)_2SO_4$, 1 mM NH_4HCO_3, 0.1 mM sodium phosphate buffer pH 8.0, trace element solution 1 (15 mg/L $EDTA{\cdot}2H_2O$, 3 mg/L $FeSO_4{\cdot}7H_2O$), trace element solution 2 (3 mg/mL $CoCl_2{\cdot}6H_2O$, 4.5 mg/L $ZnSO_4{\cdot}7H_2O$, 0.64 mg/L $MnCl_2$, 1 mg/L H_3BO_3, 0.4 mg/L $Na_2MoO_4{\cdot}2H_2O$, 0.3 mg/L $CuSO_4{\cdot}5H_2O$, 3 mg/L $CaCl_2{\cdot}2H_2O$) and vitamin solution (25 µg/L thiamine, 1 µg/L riboflavin, 1 µg/L nicotinamide, 0.1 µg/L pyridoxine, 0.1 µg/L

folic acid, 0.01 µg/L cobalamin, 1 µg/L D-biotin, 1 µg/L pantothenic acid), 20 mg/L streptomycin and either 10 mM acetate, 3.3 mM glucose, 6.7 mM propionate or 6.7 mM 3-hydroxypropionate as carbon source. Erythrobacter was grown in batch cultures at 30°C while shaking. Growth was followed by measuring the optical density at 650 nm using a photometer (Spectroquant® Prove 300, Merck KGaA, Darmstadt, Germany). For the light condition, the batch cultures were illuminated with 100 µE m^{-2} s^{-1}. Growth curves were measured in 24-well plates (TC Plate, Sarstedt AG & Co. KG, Nümbrecht, Germany) using a Tecan plate reader (Tecan Group Ltd., Männedorf, Switzerland) at a wavelength of 650 nm. To avoid cell clumping within the wells, 0.02 % tyloxapol was added to the medium and the wells were filled with 1.5 mL culture each.

Sample preparation for proteomics and protein digestion

For the analysis of the proteome in dependence of different carbon sources, *Erythrobacter* sp. NAP1 was grown in minimal medium with acetate or glucose as sole carbon source. In a *Test* setup one culture per carbon source was analyzed, while in the *Experiment* setup the cultures were prepared in quadruplicates (**Figure S7a**). To investigate the effect of light on the *Erythrobacter* sp. NAP1 proteome, the cells were grown in minimal medium with acetate and incubated in parallel in two light incubators with inverse light/dark cycles (**Figure S7b** and **S8**). Samples were prepared in triplicates per time point and light condition. Cell cultures were harvested into pre-cooled tubes by centrifugation at 3800 x g and 4°C for 10 minutes. The cells were washed with one volume of 1x PBS (pre-cooled and filtered). The cells were resuspended in 1 mL of 1x PBS, transferred into Eppendorf tubes and pelleted. The pellet was snap-frozen in liquid nitrogen and kept at -80°C until further processing.

The pellet was resuspended in 4 µL lysis buffer (2 % sodium-lauroyl sarcosinate (SLS) and 5 mM tris(2-carboxyethyl)phosphine in 100 mM ammonium bicarbonate) per mg pellet and heated for 15 minutes at 95°C and 1200 rpm. The cells were lysed by ultra-sonication for 4x 30 seconds using a VialTweeter (Hielscher Ultrasonics GmbH, Teltow, Germany) and incubated for another 15 minutes at 95°C while shaking at 1200 rpm. 10 mM of freshly prepared iodacetamide were added to the cooled-down supernatant, which was incubated for another 30 minutes at 25°C and 500 rpm in the dark. The lysate was cleared by centrifugation for 5 minutes at 15´000 rpm on an Eppendorf table centrifuge and the protein concentration in the supernatant was determined by BCA assay (Pierce™ Quantitative Colorimetric Peptide Assay, Thermo Fischer Scientific). 50 µg of total protein was diluted to a final concentration of 0.5 % SLS in 100 mM ammonium bicarbonate. Digestion was performed overnight at 37°C with 1 µg of porcine trypsin (Promega, Walldorf, Deutschland). The SLS was precipitated by adding 1 % of trifluoroacetic acid (TFA). After 10 minutes incubation at

room temperature, the sample was centrifuged for 10 minutes at 15´000 rpm. The peptides were desalted using a C18 micro spincolumn (Harvard Apparatus, Holliston, United States) according to the manufacturer´s instructions. The eluted peptides were dried under vacuum and resuspended in 100 µL 0.1 % TFA.

Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis

Peptides were analyzed by HPLC-MS using a Q-Exactive Plus instrument connected to an Ultimate 3000 RSLC nano with a Prowflow upgrade and a nanospray flex ion source (all: Thermo Scientific). For chromatographic separation of the plasmids, a reverse-phase HPLC column (75 µm x 42 cm) packed in-house with C18 resin (2.4 µm, Dr. Maisch) was used. For the proteomics experiment comparing the proteome depending on different carbon sources the following separating gradient was applied: 98 % solvent A (0.15% formic acid) and 2 % solvent B (99.85 acetonitrile, 0.15 % formic acid) to 25 % solvent B over 105 minutes and to 35 % solvent B for additional 35 minutes at a flow rate of 300 nl/min. Peptides of the proteomics experiment comparing the proteome depending on light conditions were separated from 4 % to 30 % solvent B over 2h. The data acquisition mode was adjusted to obtain one high-resolution MS scan at a resolution of 70,000 full width at half maximum (at m/z 200) followed by MS/MS scans of the most intense ions. By enabling the charged state screening modus to exclude unassigned and singly charged ion, the efficiency of the MS/MS attempts was increased. The dynamic exclusion duration was set to 30 seconds. The ion accumulation time was set to 50 ms for MS and 50 ms at 17,500 resolution for MS/MS. The automatic gain control was set to $3x10^6$ for MS survey scans and $1x10^5$ for MS/MS scans.

Label-free quantification (LFQ) of the data was performed using Progenesis QIP (Waters Corporation, Milford, MA, USA) and SafeQuant. The strategy has been described in Glatter *et al.* 2015[35].

Electron Microscopy

Erythrobacter sp. NAP1 was grown in minimal medium with acetate or glucose as carbon source. The culture was concentrated by thorough centrifugation and different concentrations were high-pressure frozen (HPF Compact 02, Wohlwend, Switzerland) and freeze-substituted (AFS2, Leica, Wetzlar, Germany) according to the following protocol: -90°C for 20h, from -90°C to -60°C in 1h, -60°C for 8h, -60°C to -30°C in 1h, -30°C for 8h, -30°C to 0°C in 1h. The substitution medium contained 0.25% osmium tetroxide, 0.2% uranyl acetate and 5% ddH_2O in acetone.

The samples were washed three times with acetone at 0°C. At room temperature a 1:1 mixture of Epon 812 substitute resin (Fluka, Buchs, Switzerland) and acetone was applied and incubated for 2

hours before it was replaced with pure resin. The samples were impregnated overnight. After another substitution with fresh Epon, samples were polymerized at 60°C for 2 days.

The polymerized blocks were trimmed with razor blades. 50 nm ultrathin sections were sliced off using an ultramicrotome (UC7, Leica, Wetzlar, Germany) and a diamond knife (Diatome, Biel, Switzerland). Sections were applied onto 100 mesh copper grids coated with pioloform.

For additional contrast mounted sections were post-stained with 2% uranyl acetate for 20-30 min and subsequently with lead citrate for another 1-2 min.

The sections were analyzed and imaged using a JEM-2100 transmission electron microscope (JEOL, Tokyo, Japan) equipped with a 2k x 2k F214 fast-scan CCD camera (TVIPS, Gauting, Germany).

4.6 References

1 Alber, B. E. & Fuchs, G. Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus,* a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO_2 fixation. *J. Biol. Chem.* **277**, 12137-12143 (2002).

2 Zarzycki, J., Brecht, V., Müller, M. & Fuchs, G. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. *PNAS* **106**, 21317-21322 (2009).

3 Zarzycki, J. & Fuchs, G. Coassimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*. *Appl. Environ. Microbiol.* **77**, 6181-6188 (2011).

4 Kolber, Z. S. *et al.* Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* **292**, 2492-2495 (2001).

5 Yutin, N. *et al.* Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ. Microbiol.* **9**, 1464-1475 (2007).

6 Koblížek, M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol. Rev.* **39**, 854-870 (2015).

7 Koblížek, M. *et al.* Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol.* **180**, 327-338 (2003).

8 Koblizek, M. *et al.* Genome sequence of the marine photoheterotrophic bacterium *Erythrobacter* sp. strain NAP1. *J. Bacteriol.* **193**, 5881-5882 (2011).

9 Bernhardsgrütter, I. *et al.* The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite. *Nat. Chem. Biol.* **14**, 1127 (2018).

10 Schloss, J. V. & Cleland, W. Inhibition of isocitrate lyase by 3-nitropropionate, a reaction-intermediate analog. *Biochemistry* **21**, 4420-4427 (1982).

11 Tomasch, J., Gohl, R., Bunk, B., Diez, M. S. & Wagner-Döbler, I. Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *ISME J.* **5**, 1957 (2011).

12 Bill, N. *et al.* Fixation of CO_2 using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ. Microbiol.* **19**, 2645-2660 (2017).

13 Hauruseu, D. & Koblížek, M. Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl. Environ. Microbiol.* **78**, 7414-7419 (2012).

14 Harashima, K., Kawazoe, K., Yoshida, I. & Kamata, H. Light-stimultated aerobic growth of *Erythrobacter* species OCh 114. *Plant Cell Physiol.* **28**, 365-374 (1987).

15 Yurkov, V. V. & van Gemerden, H. Impact of light/dark regimen on growth rate, biomass formation and bacteriochlorophyll synthesis in *Erythromicrobium hydrolyticum*. *Arch. Microbiol.* **159**, 84-89 (1993).

16 Biebl, H. & Wagner-Döbler, I. Growth and bacteriochlorophyll a formation in taxonomically diverse aerobic anoxygenic phototrophic bacteria in chemostat culture: influence of light regimen and starvation. *Process Biochem.* **41**, 2153-2159 (2006).

17 Xiao, N. & Jiao, N. Formation of polyhydroxyalkanoate in aerobic anoxygenic phototrophic bacteria and its relationship to carbon source and light availability. *Appl. Environ. Microbiol.* **77**, 7445-7450 (2011).

18 Tank, M. & Bryant, D. A. *Chloracidobacterium thermophilum* gen. nov., sp. nov.: an anoxygenic microaerophilic chlorophotoheterotrophic acidobacterium. *Int. J. Syst. Evol. Micr.* **65**, 1426-1430 (2015).

19 Tank, M. & Bryant, D. A. Nutrient requirements and growth physiology of the photoheterotrophic Acidobacterium, Chloracidobacterium thermophilum. *Front. Microbiol.* **6**, 226 (2015).

20 Kremer, K. *et al.* Dynamic Metabolic Rewiring Enables Efficient Acetyl Coenzyme A Assimilation in *Paracoccus denitrificans*. *mBio.* **10**, e00805-00819 (2019).

21 Shiba, T. Utilization of light energy by the strictly aerobic bacterium *Erythrobacter* sp. OCh 114. *J. Gen. Appl. Microbiol.* **30**, 239-244 (1984).

22 Soora, M. & Cypionka, H. Light enhances survival of *Dinoroseobacter shibae* during long-term starvation. *PLoS One* **8** (2013).

23 Erb, T. J. Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* **77**, 8466-8477 (2011).

24 Khatipov, E., Miyake, M., Miyake, J. & Asada, Y. Accumulation of poly-β-hydroxybutyrate by *Rhodobacter sphaeroides* on various carbon and nitrogen substrates. *FEMS Microbiol. Lett.* **162**, 39-45 (1998).

25 Brandl, H., Gross, R. A., Lenz, R. W., Lloyd, R. & Fuller, R. C. The accumulation of poly(3-hydroxyalkanoates) in *Rhodobacter sphaeroides*. *Arch. Microbiol.* **155**, 337-340 (1991).

26 Liebergesell, M. *et al.* Formation of poly(3-hydroxyalkanoates) by phototrophic and chemolithotrophic bacteria. *Arch. Microbiol.* **155**, 415-421 (1991).

27 Idehara, K. *et al.* in *Seventeenth Symposium on Biotechnology for Fuels and Chemicals.* 361-366 (Springer).

28 Herter, S. *et al.* Autotrophic CO_2 fixation by *Chloroflexus aurantiacus*: study of glyoxylate formation and assimilation via the 3-hydroxypropionate cycle. *J. Bacteriol.* **183**, 4305-4316 (2001).

29 Cao, L. *et al.* Comparison of *Chloroflexus aurantiacus* strain J-10-fl proteomes of cells grown chemoheterotrophically and photoheterotrophically. *Photosynth. Res.* **110**, 153-168 (2012).

30 Yurkov, V. & Csotonyi, J. T. in *The purple phototrophic bacteria* Vol. 28 *Advances in Photosynthesis and Respiration* (eds C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer, & J. Thomas Beatty) Ch. 3, 31-55 (Springer, 2009).

31 Béjà, O. *et al.* Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**, 630 (2002).

32 McCully, A. L., Onyeziri, M. C., LaSarre, B., Gliessman, J. R. & McKinlay, J. B. Reductive tricarboxylic acid cycle enzymes and reductive amino acid synthesis pathways contribute to electron balance in a *Rhodospirillum rubrum* Calvin-cycle mutant. *Microbiology*, micro000877 (2019).

33 Wang, X., Falcone, D. L. & Tabita, F. R. Reductive pentose phosphate-independent CO_2 fixation in *Rhodobacter sphaeroides* and evidence that ribulose bisphosphate carboxylase/oxygenase activity serves to maintain the redox balance of the cell. *J. Bacteriol.* **175**, 3372-3379 (1993).

34 Laguna, R., Tabita, F. R. & Alber, B. E. Acetate-dependent photoheterotrophic growth and the differential requirement for the Calvin–Benson–Bassham reductive pentose phosphate cycle in Rhodobacter sphaeroides and Rhodopseudomonas palustris. *Arch. Microbiol.* **193**, 151-154 (2011).

35 Glatter, T. *et al.* Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11**, 5145-5156 (2012).

36 Alber, B. *et al.* Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal Metallosphaera and Sulfolobus spp. *J. Bacteriol.* **188**, 8551-8559 (2006).

37 Becraft, E. D. *et al.* Rokubacteria: genomic giants among the uncultured bacterial phyla. *Front. Microbiol.* **8**, 2264 (2017).

38 Anantharaman, K. *et al.* Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* **12**, 1715-1728 (2018).

39 Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237-2243 (2018).

40 Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870-1874 (2016).

41 Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307-1320 (2008).

42 Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791 (1985).

43 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547-1549 (2018).

4.7 Supplementary Information

Supplementary Information I: Phylogenetic distribution of Pcs and Mcr

Since the last comparison of the Pcs and Mcr phylogenetic distribution[3], numerous additional sequences of these two genes have been retrieved through metagenomic sequencing. However, the general statement has not changed. The phylogenetic trees for Pcs and Mcr still look very similar, suggesting that these two genes distributed as a unit (**Figure S1** and **S2**). However, two novel sub-branches appear in the phylogenetic tree of Pcs, which cluster the homologues encoded in Euryarchaeotes and Firmicutes. Interestingly, the respective candidates are all potentially able to harvest energy from light, either by rhodopsin in Euryarchaeotes or by bacteriochlorophyll in Firmicutes. The Pcs homologues in these two phyla are phylogenetically close to the homologues in the Chloroflexi phylum. As described before[3], in the genome of some members of Chloroflexi the *pcs* and *mcr* genes are located far from each other, which might explain the transfer of the *pcs* gene independent of *mcr.* However, the absence of a bifunctional Mcr does not necessarily entail the absence of malonyl-CoA reduction activity, e.g., it has been shown that this two-step reaction can be catalyzed by two individual enzymes[36]. The opposite scenario of Mcr presumably encoded in the absence of a full-length Pcs can be found in the phylum of Acidobacteria, in the genome of *Chloracidobacterium thermophilum* strain B (see chapter 4.3). A conclusive remark about the evolutionary origin of Pcs or Mcr remains elusive. The fact that the gene sequences from the same phylum/class cluster in separate sub-branches in both phylogenetic trees suggests some intricate events of horizontal gene transfer.

Supplementary Information II: Phylogenetic analysis of the individual Pcs domains

The multidomain fusion enzyme Pcs exists in bacteria and archaea, but also the stand-alone homologues of the three individual domains are known to coexist in strains of both kingdoms (e.g., in *Rhodobacter sphaeroides*, in *Metallosphaera sedula* and in *Nitrosopumilus maritimus*, **Figure S4**). This might indicate that the fusion enzyme evolved independently in both kingdoms.

Noteworthy, when we compared the individual Pcs domains to their stand-alone homologues, a few sequences apart from the *C. thermophilum* homologues became apparent. Two stand-alone homologues cluster within the full-length Pcs reductase domains (**Figure S5c**). While the *Melghirimyces algeriensis* sequence aligns with the additional Rossmann fold in the Pcs reductase sequence (**Supplementary Figure 11**, in chapter 2), the *Halococcus agarilyticus* sequence covers the complete *Eryhtrobacter* sp. NAP1 Pcs reductase sequence (97 % coverage) and shares 57 % sequence identity. Besides, all three stand-alone homologues of the Pcs domains are encoded by

neighboring genes in *H. agarilyticus* (**Figure S4**). Furthermore, the same sequences from *Candidatus* Rokubacteria are retrieved in the BLAST search of the Pcs dehydratase and reductase domain (**Figure S5b** and **c**). Interestingly, these sequences encode fusion enzymes of the two domains, however, in reverse order compared to Pcs (N-terminal reductase domain, C-terminal dehydratase domain). The additional elements that mimic contributions of neighboring protomers in Pcs (**Supplementary Figure 10** and **11** in chapter 2) also exist in the *Candidatus* Rokubacteria sequences. While homologues of the Pcs ligase domain can be found in the recorded metagenomics data of *Candidatus* Rokubacteria, none of them contains the additional 4-helix domain. However, conclusive remarks cannot be made at this point as most *Candidatus* Rokubacteria metagenomes are incomplete. Generally, the large and heterogenic genomes of individual Rokubacteria cells[37], might hint towards a high rate of genome modifications, which could put Rokubacteria into the position of a chassis for the evolution of (foreign) DNA. The genomic plasticity of *Candidatus* Rokubacteria is refelected in their sulfate reduction system, which has been shown to be of mosaic evolutionary origin[38].

Apart from the above mentioned cases, the three domains in Pcs of *Erythrobacter* sp. NAP1 cluster with stand-alone homologues from different kingdoms. While the stand-alone homologues of the ligase domain are mainly encoded in archaeal strains, the dehydratase and reductase domains seem to descend from bacterial genes (**Figure S5**). This suggests that horizontal gene transfer of the individual genes preceded the fusion event. Furthermore the full-length Pcs sequences cluster differently depending on the domain they are compared by (**Figure S5**), which might suggest that the fusion enzyme evolved several times independently.

Supplementary Figures

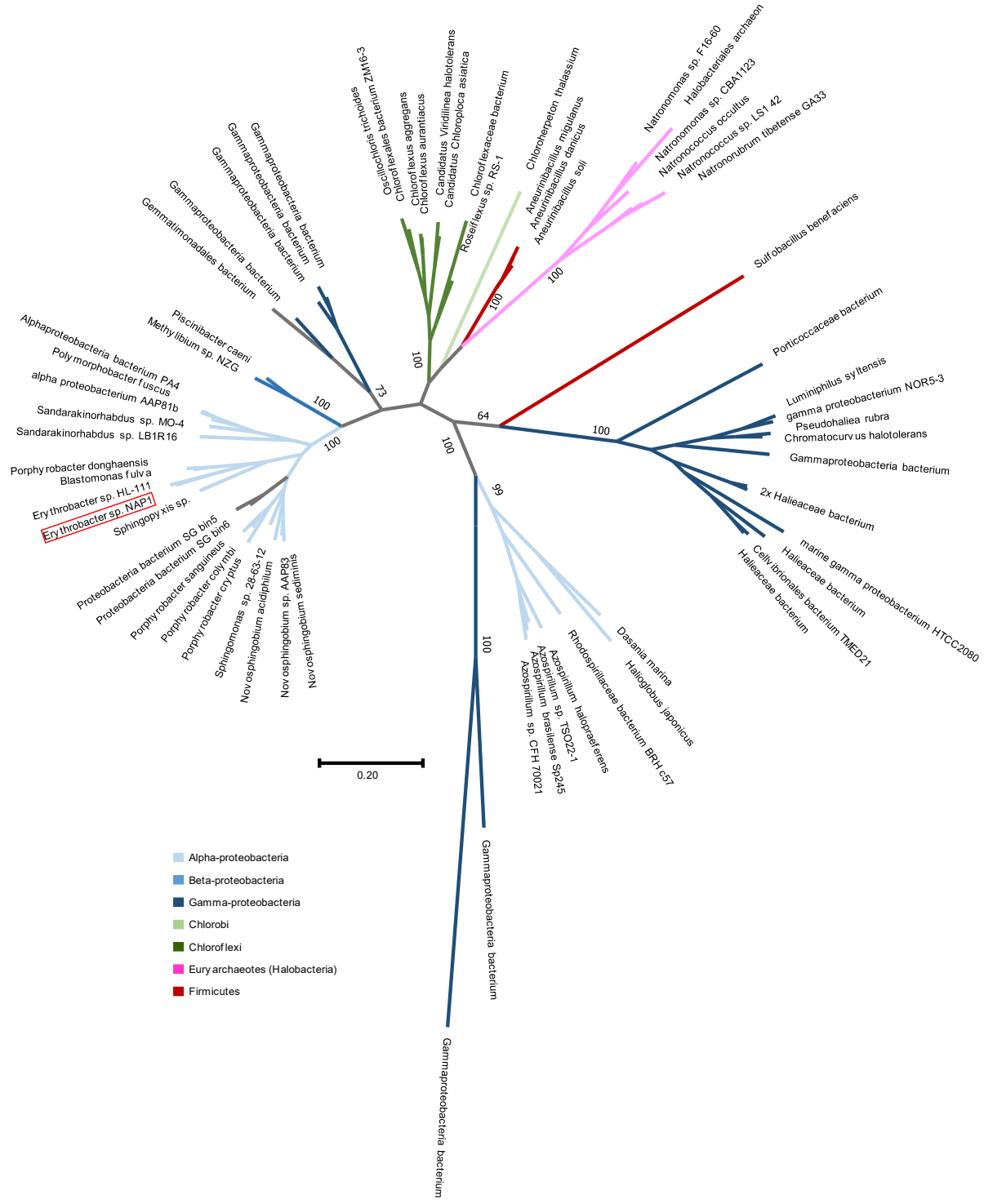


Figure S1: Maximum-likelihood tree of Pcs homologues. 66 Pcs sequences with a minimal query coverage of 90 % were retrieved using the ProtBlast/PSI-Blast tool (nr90) available from the MPI Bioinformatics Toolkit[39]. The Maximum Likelihood method based on the Le-Gascuel_2008 model in MEGA7 was used for phylogenetic analysis[40,41]. The number of substitutions per site is represented in the branch lengths (see scale bar). Bootstrap values[42] are given at important nodes. The Pcs homologue of *Erythrobacter* sp. NAP1 is highlighted with a red box. Pcs is distributed far beyond the phylum of Chloroflexi. Affiliation of the Pcs homologues to the respective phylum/class is illustrated by color code.

Figure S2: Maximum-likelihood tree of Mcr homologues. 55 Mcr sequences with a minimal query coverage of 90 % were retrieved using the ProtBlast/PSI-Blast tool (nr90) available from the MPI Bioinformatics Toolkit[39]. The Maximum Likelihood method based on the Le-Gascuel_2008 model in MEGA X was used for phylogenetic analysis[41,43]. The number of substitutions per site is represented in the branch lengths (see scale bar). Bootstrap values[42] are given at important nodes. The Mcr homologue of *Erythrobacter* sp. NAP1 is highlighted with a red box. Mcr is distributed far beyond the phylum of Chloroflexi. Affiliation of the Mcr homologues to the respective phylum/class is illustrated by color code.

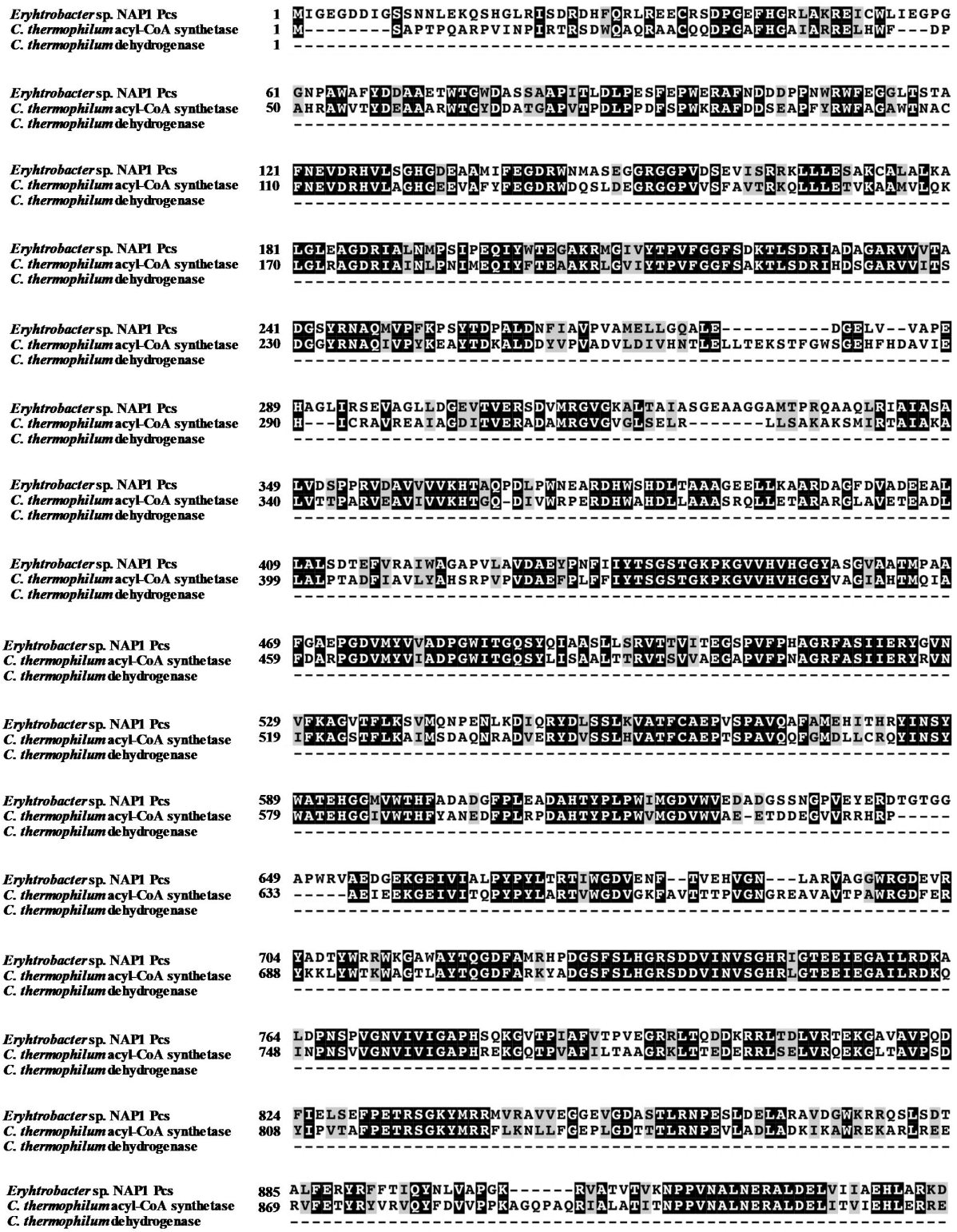


Figure S3: Alignment of the *Erythrobacter* sp. NAP1 Pcs sequence with homologues in *C. thermophilum*. The *C. thermophilum* acyl-CoA synthetase is homologous to the Pcs ligase and part of the dehydratase domain. The *C. thermophilum* dehydrogenase is homologous to part of the Pcs dehydratase domain and the reductase domain. Figure was created using BoxShape, a tool available on the ExPASy Bioinformatics Resource Portal.

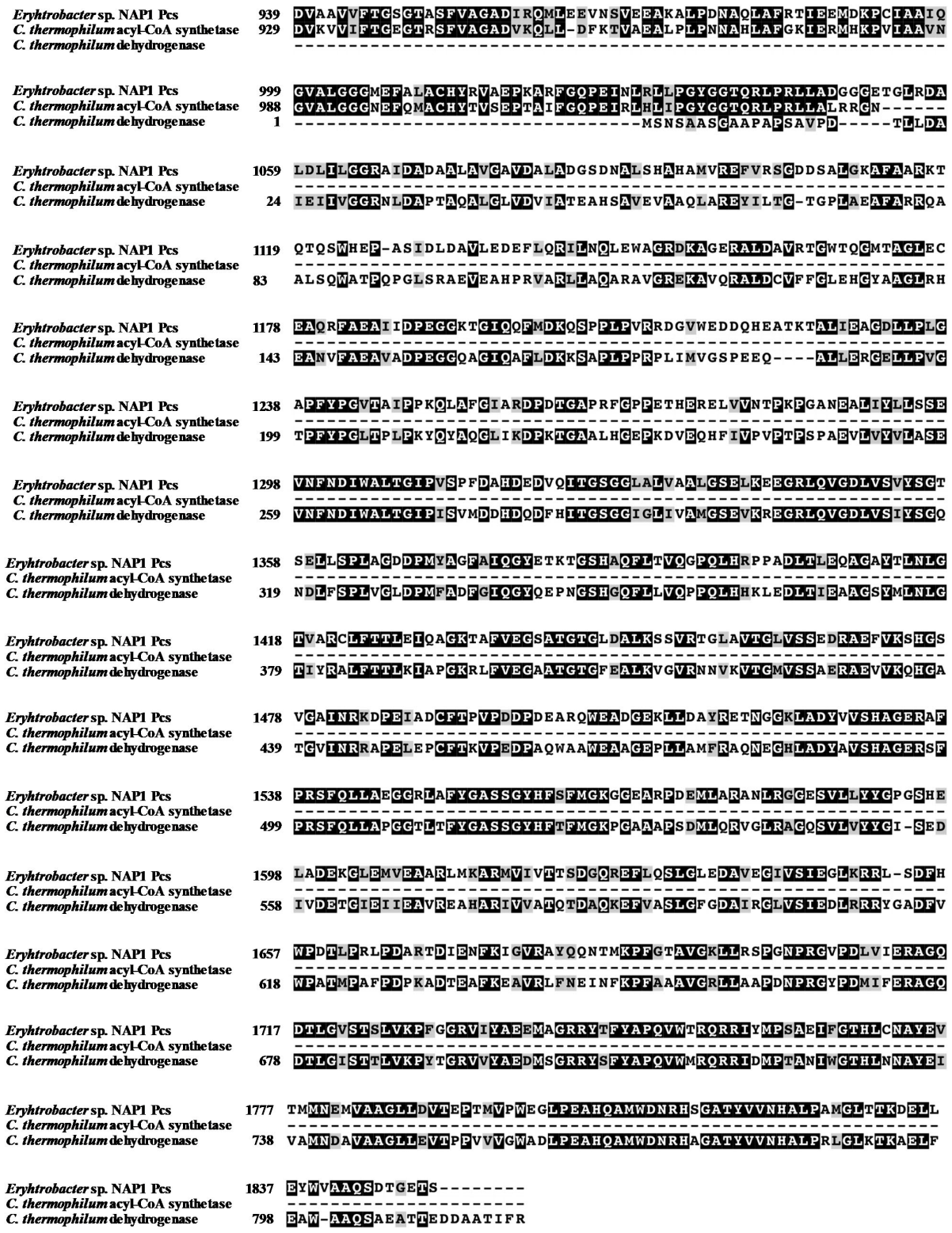


Figure S3 (continued)

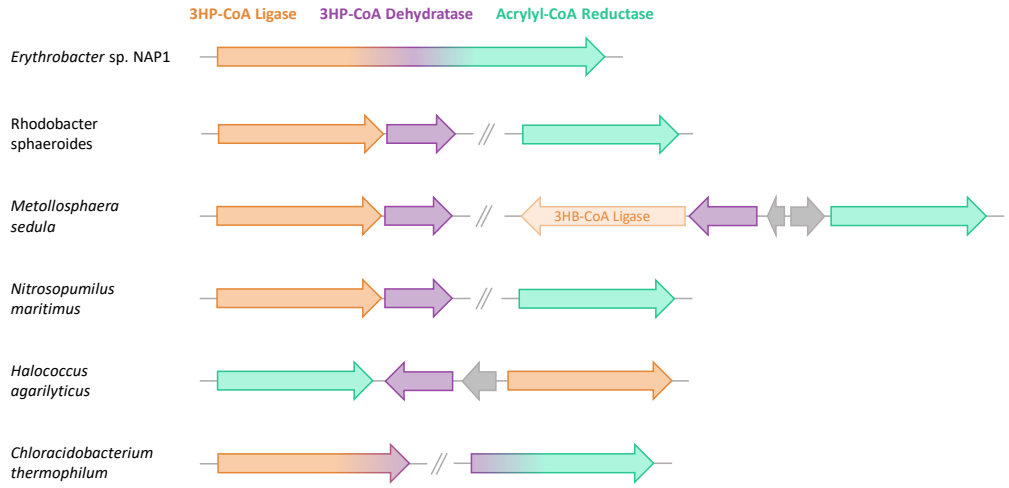


Figure S4: Organization of the genes encoding the Pcs stand-alone homologues in different bacteria. The genome neighborhood of Pcs stand-alone homologues in selected organisms. The genes for the ligase, dehydratase and reductase are depicted as arrows in orange, purple and cyan, respectively. Grey arrows show genes of unrelated function. 3HB-CoA, 3-hydroxybutyryl-CoA.

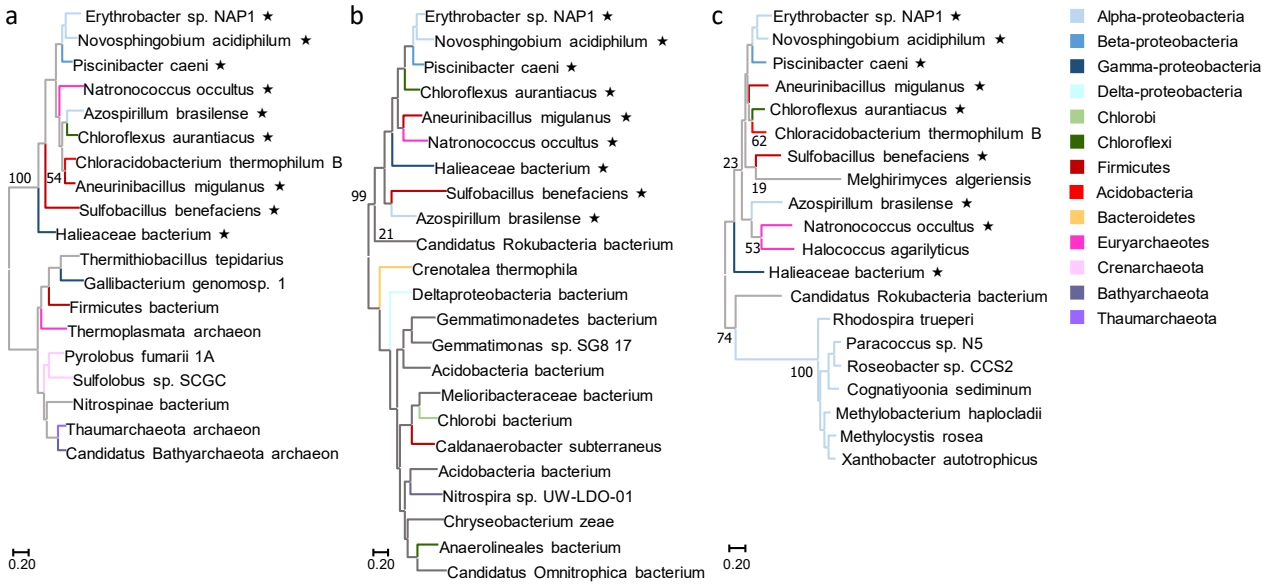


Figure S5: Maximum-likelihood tree of individual Pcs domains. Some selected full-length Pcs sequences were split into the individual domains, i.e., the ligase domain in a, the dehydratase domain in b and the reductase domain in c, and compared to stand-alone homologues. Stand-alone homologues from different phyla were manually retrieved from the ProtBlast/PSI-Blast tool (nr90) available from the MPI Bioinformatics Toolkit[39]. The Maximum Likelihood method based on the Le-Gascuel_2008 model in MEGA7 was used for phylogenetic analysis[40,41]. The number of substitutions per site is represented in the branch lengths (see scale bar). Bootstrap values [42] are given at important nodes. The star ★ marks the shortened sequences of full-length Pcs homologues. Affiliation of the homologues to the respective phylum/class is illustrated by color code.

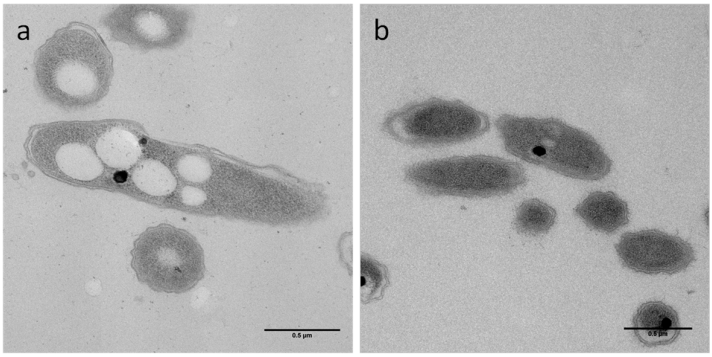


Figure S6: Transmission electron micrographs of *Erythrobacter* sp. NAP1 grown on acetate reveals PHA granules. *Erythrobacter* sp. NAP1 was grown on **a,** actetate or **b,** glucose. PHA granules are visible in cells grown on acetate. The scale bar represents 0.5 µm.

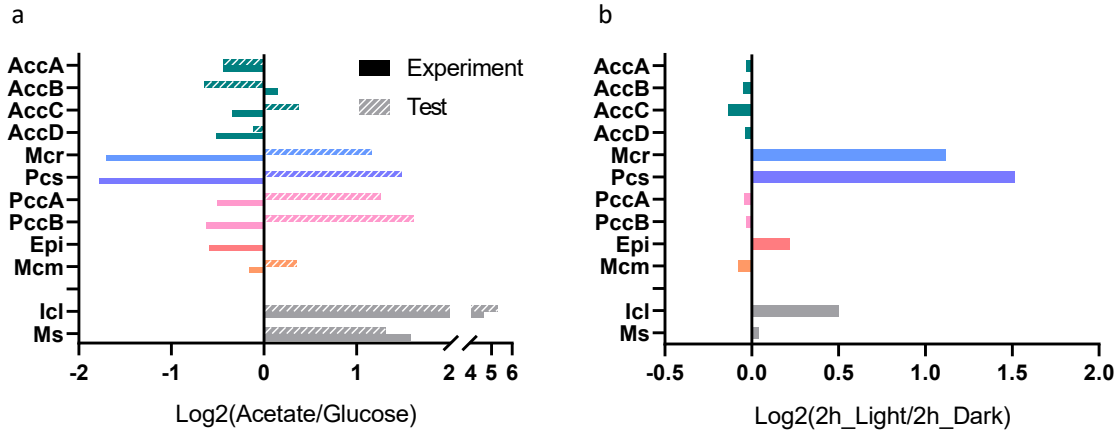


Figure S7: Abundance change of Pcs module enzymes based on proteomics data. Comparison of the protein abundance of the glyoxylate cycle and Pcs module enzymes depending on **a,** the carbon source (acetate versus glucose) and **b,** light regime (2 hours light versus 2 hours dark). In **a**, the *Test* setup comprised one sample each per carbon source, while quadruplicates were analyzed in the *Experiment* setup. The enzyme names are abbreviated as following: AccA-D, acetyl-CoA carboxylase subunit A-D; Mcr, malonyl-CoA reductase; Pcs, propionyl-CoA synthase; PccA-B, propionyl-CoA carboxylase subunit A-B; Epi, methylmalonyl-CoA epimerase; Mcm, methylmalonyl-CoA mutase; Icl, isocitrate lyase; Ms, malate synthase.

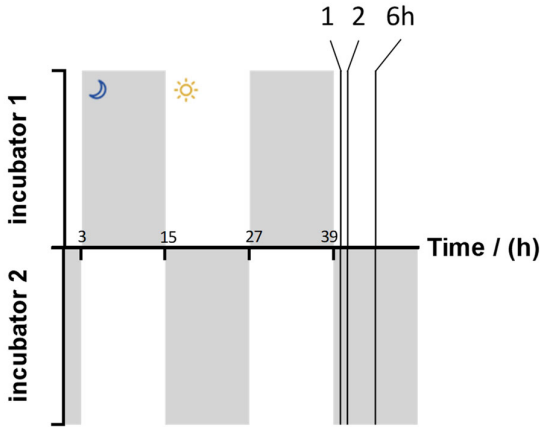


Figure S8: Sampling scheme for the light dependent proteomics experiment. *Erythrobacter* sp. NAP1 was grown in batch cultures in minimal medium with acetate and incubated in parallel in two light incubators with inverse light/dark cycles. After the light-switch at 39 hours, samples were collected in triplicates at three specific time points (1 h, 2 h, 6 h with *new* light setting).

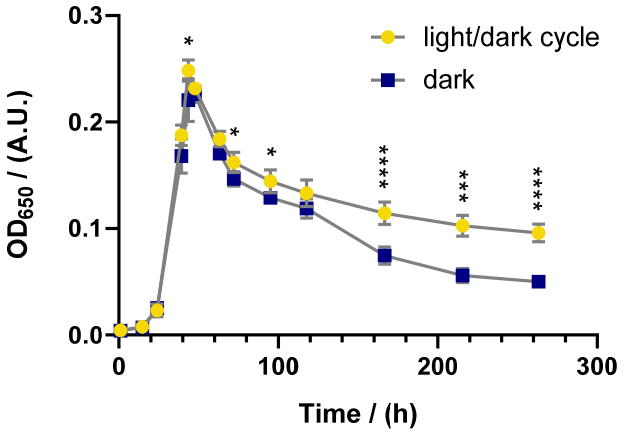


Figure S9: Light-induced Pcs module allows increased cell viability in nutrient-scarce conditions. Four cultures of *Erythrobacter* sp. NAP1 in minimal medium with acetate each were grown either with a 12h/12h light/dark cycle or in the dark. Multiple t-tests were applied to probe significance, which is represented by asterisks as following: *, p<0.05; ***, p<0.001; ****, p<0.0001. Data of a single representative experiment.

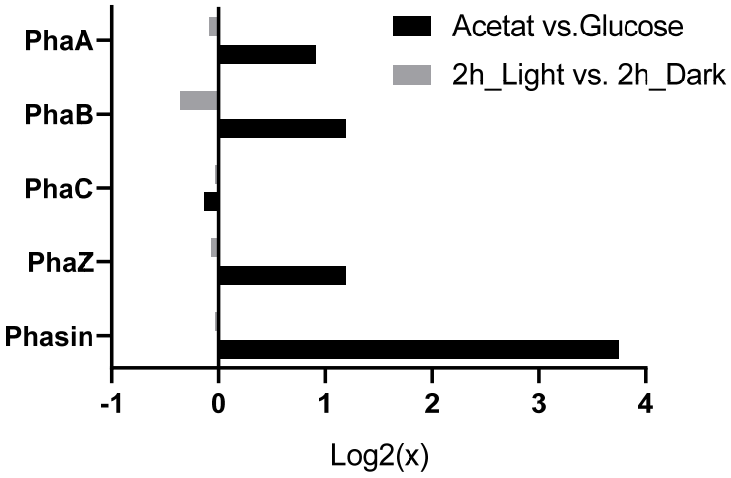


Figure S10: Abundance change of PHA-related enzymes based on proteomics data. Comparison of the abundance pattern of the glyoxylate cycle and Pcs module enzymes depending on the carbon source (acetate versus glucose) or on the light regime (2 hours light versus 2 hours dark). The enzyme names are abbreviated as following: PhaA, β-ketothiolase; PhaB, acetoacetyl-CoA reductase; PhaC, PHA synthase; PhaZ, putative Pha depolymerase.

Supplementary Tables

Table S1: List of all significantly changed protein levels depending on the carbon source (acetate versus glucose). 2100 proteins were identified by shotgun proteomics. The significance cutoff was set with s0 = 0.5 and FDR = 0.001 and resulted in 145 proteins with a significantly different abundance. Proteins that are more/less abundant on acetate are highlighted in green/red, respectively. Similar protein hits are given for uncharacterized proteins, when available.

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
7.69	3.05	A3WG46	Putative sugar nucleotide dehydratase	
7.25	4.45	A3WBU0	Putative RNA methyltransferase	
6.24	2.83	A3WAL2	Pilus assembly protein CpaB	
5.89	2.85	A3WD46	Phosphate regulon transcriptional regulatory protein	
5.28	2.93	A3WFY2	Uncharacterized protein	
5.02	2.13	A3WDV0	Acetyltransferase, GNAT family protein	
4.57	**9.56**	**A3WCR0**	**Isocitrate lyase**	
4.37	5.07	A3WBU6	Putative polysaccharide export protein	
4.27	2.26	A3WG52	ATP-grasp_3 domain-containing protein	Carbamoyl-phosphate synthase large subunit
4.23	5.47	A3WDC5	Thiol:disulfide interchange protein	
4.19	2.11	A3WG58	Exopolysaccharide biosynthesis protein	
4.16	3.36	A3WCH0	HTH marR-type domain-containing protein	MarR family transcriptional regulator
4.09	3.99	A3WDY5	Uncharacterized protein	
4.05	4.18	A3WGW0	YCII domain-containing protein	
3.79	3.70	A3WA81	Penicillin-binding protein	
3.79	**7.73**	**A3WE91**	**Phasin_2 domain-containing protein**	
3.74	4.01	A3WAB5	Uncharacterized protein	Pilus assembly protein CpaD
3.14	5.11	A3WCH8	Uncharacterized protein	Inositol monophosphatase
3.10	6.97	A3WG97	Sodium:solute symporter family protein	
3.03	3.97	A3WBT0	Peptidase_S8 domain-containing protein	
2.96	6.59	A3W9S9	Dehydrogenase	
2.94	8.37	A3WG95	Uncharacterized protein	Porin
2.86	4.27	A3WFW1	Divalent cation tolerance protein	
2.86	8.02	A3WG93	Putative regulator in two-component regulatory system LuxR family protein	
2.84	3.18	A3WHT2	5-aminolevulinate synthase	
2.84	6.43	A3WA88	Bacteriophage N4 adsorption protein B	
2.72	7.81	A3WG94	Acetyl-coenzyme A synthetase	
2.70	2.93	A3WF52	Uncharacterized protein	
2.70	3.38	A3WDY6	Putative ISXo8 transposase	
2.67	5.34	A3WB60	Uncharacterized protein	
2.64	5.58	A3WEA0	Two-component response regulator	
2.61	4.51	A3WA89	Uncharacterized protein	
2.57	4.70	A3WGA0	DNA polymerase III subunit epsilon	
2.50	3.26	A3WDD4	TraC	
2.49	5.53	A3WEX9	Uncharacterized protein	Phage major capsid protein
2.46	4.28	A3WB18	ApaG domain-containing protein	Co2+/Mg2+ efflux protein ApaG
2.44	2.59	A3WC09	Indole-3-glycerol phosphate synthase	

Table S1 (continued)

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
2.39	4.84	A3WFJ9	Peroxiredoxin	
2.37	2.98	A3WBX7	Uncharacterized protein	
2.24	3.18	A3WDL8	Uncharacterized protein	Outer membrane family protein
2.23	2.57	A3WDJ6	Uncharacterized protein	
2.16	5.09	A3WFW0	Uncharacterized protein	
2.13	4.07	A3WC32	Uncharacterized protein	Septum formation initiator
2.07	4.75	A3WGS7	Uncharacterized protein	Methyltransferase domain-containing protein
2.06	3.67	A3WGS9	Uncharacterized protein	Hydrolase
2.01	4.11	A3WAW5	Curlin-associated protein	
1.80	3.28	A3WFC5	ABC transporter, ATP-binding protein	
1.79	5.77	A3WFU3	Uncharacterized protein	Fe-S cluster assembly protein SufD
1.72	4.42	A3WAB6	Type II secretion system protein	
1.64	5.50	A3WBF3	Orotidine 5'-phosphate decarboxylase	
1.63	**5.69**	**A3WIF1**	**Malate synthase G**	
1.55	5.73	A3WCQ9	Uncharacterized protein	
1.52	4.67	A3WA29	Cell division protein ftsA	
1.47	5.78	A3WB20	30S ribosomal protein S12	
1.47	5.24	A3W9H2	Probably methylase/helicase	
1.46	4.96	A3WE22	Myeloperoxidase, thyroid peroxidase, cyclooxygenase catalytic domain	
1.42	6.01	A3WCJ1	ParA-like protein	
1.38	7.25	A3WBY5	NADH-quinone oxidoreductase subunit D	
-1.27	8.15	A3WID4	TonB-dependent receptor	
-1.30	7.17	A3WAT6	Integration host factor subunit alpha	
-1.35	6.54	A3WGH4	Enoyl CoA dehydratase/isomerase	
-1.35	6.99	A3WC35	Haloalkane dehalogenase	
-1.42	7.04	A3WI38	Glycerol-3-phosphate dehydrogenase	
-1.42	6.95	A3WEH9	Xaa-Pro dipeptidase, putative	
-1.43	5.01	A3WHS1	Acetyl-coenzyme A synthetase	
-1.45	6.32	A3WH94	Metalloendopeptidase PepO	
-1.46	4.81	A3WGZ1	DUF4440 domain-containing protein	
-1.47	4.65	A3WCE7	Acetyltransferase, GNAT family protein	
-1.48	6.10	A3WBH0	Signal recognition particle protein	
-1.49	4.49	A3WB65	Amidase	
-1.50	4.52	A3WCF3	Probable dehydrogenase/ reductase 7	
-1.51	4.53	A3WB07	Putative dolichol monophosphate mannose synthase	
-1.53	5.57	A3WAX0	Predicted transcriptional regulator	
-1.57	4.54	A3WHW0	Bacteriochlorophyllide reductase subunit BchX	
-1.58	6.48	A3WFJ7	Uncharacterized protein	Phosphoribosylformylglycinamidine synthase
-1.60	6.06	A3WHJ7	DUF547 domain-containing protein	
-1.62	5.10	A3WGM2	Uncharacterized protein	Putative methyltransferase

Table S1 (continued)

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
-1.64	4.94	A3WGV8	Tyrosine protein kinase:Aminoglycoside phosphotransferase	
-1.65	5.76	A3WH29	Uncharacterized protein	
-1.67	4.28	A3W9G7	Curlin-associated protein	
-1.67	7.51	A3WBS4	Acyl-CoA synthase	
-1.68	4.20	A3W9F4	SnoaL-like domain-containing protein	Ketosteroid isomerase
-1.69	7.02	A3WIC3	Glyoxalase	
-1.69	**5.58**	**A3WE13**	**Malonyl-CoA reductase**	
-1.70	9.00	A3WGV7	Aldehyde dehydrogenase family protein	
-1.70	4.89	A3WD65	RND efflux system, outer membrane lipoprotein, NodT	
-1.74	4.11	A3W9Z2	Guanylate cyclase-related protein	
-1.77	**6.35**	**A3WE14**	**Propionyl-CoA synthase**	
-1.77	3.78	A3WDI6	Glutathione-dependent formaldehyde-activating, GFA	
-1.79	4.34	A3WI11	Uncharacterized protein	
-1.80	4.84	A3W9U1	Uncharacterized protein	
-1.84	4.66	A3WAU1	Protein-disulfide isomerase	
-1.84	5.49	A3WAW0	Dienelactone hydrolase family protein	
-1.87	4.90	A3WFD7	Uncharacterized protein	
-1.87	3.95	A3WEH8	Lactamase_B domain-containing protein	Pyrroloquinoline quinone biosynthesis protein PqqB
-1.89	3.21	A3WDN3	Uncharacterized protein	
-1.90	3.05	A3WBX0	Bacterioferritin comigratory protein, putative	
-1.93	3.65	A3W9U8	TonB_C domain-containing protein	
-1.93	5.85	A3WEI1	Endoribonuclease L-PSP	
-1.95	3.63	A3WFA9	Glyco_trans_2-like domain-containing protein	
-1.96	3.96	A3WDA4	Uncharacterized protein	
-1.98	3.85	A3WH88	Alpha/beta hydrolase	
-1.98	3.46	A3WD01	Chloramphenicol phosphotransferase-like protein	
-1.99	3.33	A3WHW2	2-desacetyl-2-hydroxyethyl bacteriochlorophyllide a dehydrogenase	
-1.99	6.48	A3WF79	Alcohol dehydrogenase large subunit	
-1.99	4.29	A3W9L7	Putative sphingosine-1-phosphate lyase	
-2.02	4.00	A3WHW9	Putative magnesium chelatase subunit BchD	
-2.03	4.82	A3WBL9	Alpha/beta hydrolase fold protein	
-2.03	3.87	A3W9R4	Uncharacterized protein	Cytochrome c domain-containing protein
-2.08	5.37	A3WCZ5	Shikimate kinase	
-2.12	4.04	A3WAJ4	FAS1 domain-containing protein	
-2.16	2.74	A3WHV0	Putative bacteriochlorophyll synthase, 34 kDa subunit	
-2.17	3.08	A3WF21	Uncharacterized protein	
-2.18	6.43	A3WHT0	Cytochrome c family protein	
-2.23	5.15	A3WAQ8	Sensory box histidine kinase	

Table S1 (continued)

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
-2.24	4.42	A3WHV4	Photosynthetic reaction center L subunit	
-2.24	6.08	A3WHU5	Coenzyme B12-binding protein	
-2.30	2.67	A3WCW3	Uncharacterized protein	
-2.32	4.95	A3WAG9	Ferroxidase	
-2.35	5.84	A3WD33	Uncharacterized protein	
-2.46	4.84	A3WHU8	Geranylgeranyl reductase	
-2.46	3.36	A3WI60	ABC sugar (Glycerol) transporter, inner membrane subunit	
-2.58	3.57	A3WEK8	Uncharacterized protein	
-2.59	4.50	A3WIB2	Uncharacterized protein	Methyltransferase
-2.64	6.06	A3WHL8	Uncharacterized protein	
-2.64	4.65	A3WE55	Uncharacterized protein	
-2.66	5.20	A3WHU2	Light-independent protochlorophyllide reductase subunit N	
-2.69	3.47	A3WF68	UPF0145 protein NAP1_11678	
-2.71	4.72	A3WFW8	Uncharacterized protein	UrcA family protein
-2.79	7.80	A3WGH5	TonB-dependent receptor	
-2.88	3.38	A3WEN6	Uncharacterized protein	
-2.92	2.59	A3W9R3	Beta_helix domain-containing protein	
-3.03	3.53	A3WH17	Uncharacterized protein	
-3.04	3.43	A3WD27	Glycerol-3-phosphate acyltransferase	
-3.08	6.73	A3WFJ5	Catalase-peroxidase	
-3.10	4.58	A3W9G8	Uncharacterized protein	Holin of 3TMs, for gene-transfer release
-3.28	3.67	A3WHT6	Possible photosynthetic complex assembly protein	
-3.61	3.11	A3WDR5	N-carbamoylsarcosine amidase-like protein	
-3.67	4.46	A3WDA8	Putative oxidase	
-4.02	2.44	A3WH03	Uncharacterized protein	TonB family protein
-4.08	6.17	A3WA24	TonB-dependent receptor, putative	
-4.15	3.86	A3WI77	Uncharacterized protein	
-4.75	7.16	A3WEG4	Possible Dps protein family starvation-inducible DNA-binding protein	
-4.95	1.67	A3WGG2	Putative oxidoreductase	
-5.72	1.68	A3WE18	Phosphatidylserine synthase	

Table S2: List of all significantly changed protein levels depending on the light regime (2 hours light versus 2 hours dark). Proteins that are more/less abundant in the light condition are highlighted in green/red, respectively. Similar protein hits are given for uncharacterized proteins, when available.

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
4.83	3.98	A3WEK0	Two-component system regulatory protein	
4.73	3.97	A3WFY0	Uncharacterized protein	Heme biosynthesis protein HemY
4.08	4.38	A3WA67	Uncharacterized protein	
3.93	7.73	A3W9H6	Uncharacterized protein	Transmembrane Protein (mercury transport)
2.79	2.53	A3WE59	UPF0276 protein	
2.72	4.36	A3WIA0	Response regulator, hypothetical	
2.40	5.40	A3W9H5	Regulatory protein	
2.33	3.40	A3WEI7	Uncharacterized protein	
2.31	3.57	A3W9T3	DJ-1/PfpI family protein	
2.16	3.92	A3WGS8	Flagellin	
2.08	4.35	A3WGT9	Flagellar basal body protein	
2.08	4.39	A3WF10	Uncharacterized protein	
1.96	4.88	A3WFK0	Alkyl hydroperoxide reductase, subunit f	
1.88	4.58	A3WAV1	Uncharacterized protein	
1.88	4.56	A3WD60	RNA polymerase sigma-24 factor	
1.83	3.59	A3WHV5	Photosynthetic reaction center M protein	
1.71	2.38	A3WHW4	Alpha/beta hydrolase fold protein	
1.67	3.02	A3WCY5	Hemerythrin	
1.65	4.38	A3WBR3	Uncharacterized protein	
1.61	3.81	A3WHV4	Photosynthetic reaction center L subunit	
1.60	4.17	A3WHW9	Putative magnesium chelatase subunit BchD	
1.59	4.91	A3WHV6	LHC domain-containing protein	
1.53	**6.33**	**A3WE14**	**Propionyl-CoA synthase**	
1.43	4.24	A3W9Q1	Uncharacterized protein	
1.36	7.43	A3WHU7	Transcriptional regulator PpsR1 Fis family protein	
1.36	7.71	A3WHT9	H subunit of photosynthetic reaction center complex	
1.31	5.14	A3WEL1	Uncharacterized protein	
1.31	5.42	A3WGK7	Bacteriophytochrome (Light-regulated signal transduction histidine kinase)	
1.30	3.53	A3WF16	YcfI, putative structural proteins	
1.27	4.74	A3WHH2	Guanylate kinase	
1.25	4.12	A3WDI2	NADH dehydrogenase	
1.24	3.87	A3WBB8	Alanyl-tRNA synthetase	
1.19	5.22	A3WEL5	Hva1_TUDOR domain-containing protein	
1.18	1.54	A3WBJ3	Methyl-accepting transducer domain-containing protein	
1.11	1.70	A3WAY2	DNA topoisomerase, type I, putative	
1.10	**4.82**	**A3WE13**	**Malonyl-CoA reductase**	
1.08	4.41	A3WEK1	HWE_HK domain-containing protein	protein histidine kinase activity
0.99	4.04	A3WG54	Uncharacterized protein	
0.97	4.13	A3WEA7	Two-component response regulator	

Table S2 (continued)

Log(Difference)	-Log(P-value)	Accession Code	Protein Description	Similar Protein
0.95	5.75	A3WHT0	Cytochrome c family protein	
0.91	2.88	A3WHW1	Hydroxyneurosporene methyltransferase CrtF	
0.86	5.22	A3WHS5	Outer membrane efflux protein	
0.86	3.99	A3WGS9	Uncharacterized protein	Hydrolase
0.83	6.20	A3WHS4	RND efflux membrane fusion protein	
-0.74	5.55	A3WEX9	Uncharacterized protein	Phage major capsid protein
-0.90	4.01	A3WDC5	Thiol:disulfide interchange protein	
-1.00	2.90	A3WDY6	Putative ISXo8 transposase	
-1.07	3.56	A3WBW5	Methyltransf_11 domain-containing protein	SAM-dependent methyltransferase
-1.09	2.09	A3WGR6	Uncharacterized protein	Flagellar biogenesis protein
-1.14	2.54	A3WHU8	Geranylgeranyl reductase	
-1.16	2.86	A3WHU5	Coenzyme B12-binding protein	
-1.23	3.09	A3WHW2	2-desacetyl-2-hydroxyethyl bacteriochlorophyllide a dehydrogenase	
-1.65	4.04	A3WHW0	Bacteriochlorophyllide reductase subunit BchX	
-1.76	3.02	A3WHU1	Mg-protoporphyrin IX methyl transferase	
-1.79	2.56	A3WHU4	Magnesium-protoporphyrin O-methyltransferase BchH subunit	
-1.99	3.06	A3WC72	DNA polymerase III delta prime subunit	

5 General Discussion and Outlook

5.1 Pcs: a substrate-channeling reaction chamber

Multidomain fusion enzymes that catalyze consecutive reactions evolve for several reasons; e.g., the coupling of biochemical reactions increases catalytic efficiency[1], intermediates are channeled between active sites[2], or the fixed stoichiometric ratio between the catalytic domains can be beneficial[3]. Substrate channeling in turn provides a strategy to avoid metabolic cross-talk between competing pathways and to cope with unstable or toxic intermediates. Nature has evolved several strategies to ensure substrate channelling from one enzyme or active site to the next, including the formation of intramolecular tunnels between active sites, covalent linking of intermediates, electrostatic guidance of intermediates over surfaces and compartmentalization[4]. As described in chapter 2, in Pcs the CoA-thioester intermediates are sequestered within an intramolecular chamber and their relay from one active site to the next is presumably facilitated by electrostatic guidance over the chamber surface. Even though, this mechanism combines known channeling strategies, several aspects set Pcs apart from previously described examples. So far, three different mechanisms have been described to channel bulky and electrically charged CoA-thioesters. In fatty acid or polyketide synthases the acyl-moiety of the acyl-CoA substrate is transferred to an acyl-carrier protein by covalent linking and guided from one enzyme to the next[5]. The second strategy has been observed in two multi-enzyme complexes where the 3'-phospho-ADP moiety of the acyl-CoA is bound at a shared CoA-binding site and the long-spanning pantetheinyl arm swings the acyl moiety between the active sites[6,7]. Similarly, in the bifunctional *Escherichia coli* enzyme PaaZ, the phosphoadenosine moiety of the CoA substrate was suggested to be guided over a short distance of positively charged surface between the two active sites followed by the hydrophobic head group[8]. The sequestration of the intermediates within a reaction chamber in Pcs presents a novel way to channel CoA-thioesters. Conformational changes induced by CoA binding in the ligase domain closes the reaction chamber and sequesters the CoA-intermediates from bulk solution. This feature of controlled access has also been described for enzymes with an intramolecular narrow substrate channel, where conformational changes upon ligand binding functions as switch to open/close the channel[2,9,10]. However, while the conformational rotation of a single residue can be sufficient to block such a substrate tunnel, the conformational changes needs to be of more drastic nature to open/close the reaction chamber in Pcs. This controlled access ensures synchronization of the catalytic cycle. The reaction chamber prevented any leakage of the reactive intermediate acrylyl-

CoA, even when the subsequent reduction reaction was inactive. This is in contrast with some other channel- or cavity-forming enzymes, where the absence of substrates leads to significant intermediate release[2,11]. This feature demonstrates the highly synchronized orchestration of the catalytic cycle and raises questions about the evolution and underlying mechanisms of such a complex metabolic machinery.

5.2 Pcs: a complex multidomain machinery

The emergence of multifunctional enzymes by gene fusion has been studied on numerous examples in nature[12-16] and has even been experimentally demonstrated in a laboratory evolution experiment[17]. In an attempt to improve the *Saccharomyces cerevisiae* glycerol pathway in an *E. coli* selection strain, the two heterologous genes, that were initially arranged in a bicistronic operon, had simply fused into a single open reading frame[17]. The increased efficiency of the bifunctional fusion enzyme was explained by partial substrate channeling. The argument of substrate channeling, however, cannot explain the fusion of the two tryptophan synthetase subunits during the evolution of fungi[18], as the bacterial heterodimeric enzyme already connects both active sites with a functional substrate tunnel[19]. This fusion event could possibly be explained by necessary genetic adaptions in the eukaryotic acquisition of bacterial operons in order to maintain the linked expression profile[20].

The multidomain fusion enzyme Pcs exists in bacteria and archaea, but also the stand-alone homologues of the three individual domains are known to coexist in strains of both kingdoms. According to the Rosetta Stone method, the three stand-alone homologues are likely to interact[21], but this has never been described so far. However, a substrate-channeling interaction as in Pcs seems unlikely due to the large structural changes, e.g., the reduction of multimeric architecture in stand-alone homologues to a single chain in Pcs. Additional structural elements that mimic protomer contributions (**Supplementary Figures 9-11** in chapter 2) indicate that Pcs is not the result of simply fusing three stand-alone enzymes. Rather, several duplication, deletion and rearrangement events were necessary to forge the parts into a functional fusion enzyme. This is also depicted in the multiple sequence alignment of all Pcs sequences, where the least conserved regions can be assigned to the linker regions between domains and to the additional structural elements. Nevertheless, these elements are essential for the complex structure and reaction mechanism of Pcs. The four-helix bundle of the ligase domain is suggested to play a crucial role in synchronizing the catalytic cycle. Its localization at the interface of the catalytic domain of the ligase and reductase domain is perfectly suited for signal transduction between the domains. This so-called interdomain communication is a phenomenon found in several multidomain enzymes and particularly in substrate-channeling

complexes[22-24]. Generally, allosteric regulation ensures directionality of the catalytic sequence by orchestrating accessibility of active sites and substrate tunnels in a step-wise manner. Substrate binding and release or even change of cofactor redox state in a distant domain have been shown to induce the underlying conformational changes[25-27].

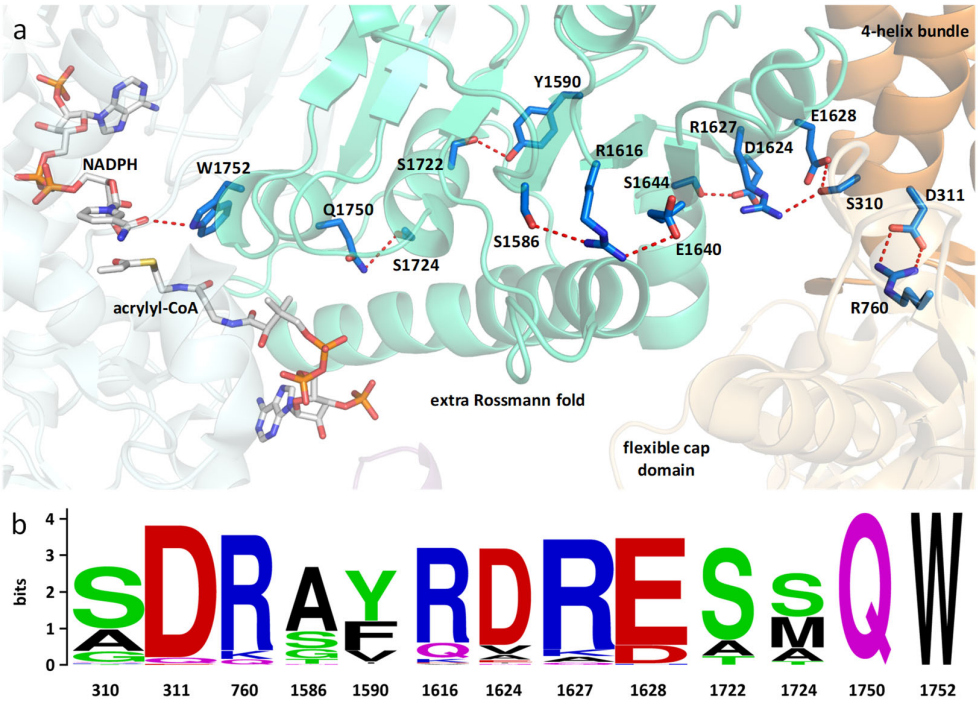


Figure 1: Suggested communication network across the ligase and reductase domain of Pcs. a, The ligase domain is shown in orange and the reductase domain in cyan. The cartoon shown in darker color represents the additional structural elements (4-helix bundle in the ligase domain and extra Rossmann fold in the reductase domain). Involved residues are shown in blue with the hydrogen bonds depicted as red interrupted lines. **b**, WebLogo-Illustration[29,30] of the residues involved in the suggested communication network.

In Pcs, binding of CoA induces closure of the reaction chamber. This state is presumably stabilized by a salt bridge between R760 in the flexible cap domain of the ligase and D311 in the four-helix bundle. The signal to re-open the reaction chamber and reset Pcs to its initial conformation after the final reduction step remains to be investigated. A hydrogen-bond network could be the key to this long-range communication (**Figure 1a**). The network stretches from the R760-D311 salt bridge all the way to the CoA-ester binding Rossmann fold. Note that this Rossmann fold is one of the additional structural elements in Pcs and mimics the neighboring subunit in stand-alone Ecrs (**Supplementary Figure 11** in chapter 2). Subunit communication has been recently reported in the Ecr of *Kitasatospora setae*[28]. There, the allosteric signal is transmitted between the subunits over a hydrogen-bond network across the shared substrate binding site. Pcs might imitate this "half-site reactivity", with the crucial difference that one of the subunits (the additional Rossmann fold) does not perform catalytic activity but might rather regulate the opening and closing of the ligase domain.

While the underlying communication network differs between the *K. setae* Ecr and Pcs, the change might be triggered by product and/or cofactor release in both scenarios. In Pcs, product release from the reductase domain active site (back into the reaction chamber) could induce conformational changes in the Rossmann fold, which would be relayed over the communication network to break the R760-D311 salt bridge. Alternatively, the oxidation of the cofactor or its release could be sensed by the interaction with W1752 (**Figure 1a**), a residue at the end of the communication network. Most residues of the suggested network are conserved throughout most of the Pcs sequences (**Figure 1b**). However, mutagenesis experiments are required to validate this elaborate mechanism. Disruption of the network would be expected to result in a drop of catalytic efficiency of the overall reaction. This effect was observed in the Pcs variant comprising an engineered carboxylase domain (PCS D1302S T1753M), where distant mutations presumably disturbed the interdomain communication (see chapter 3).

5.3 Pcs: a blueprint for the engineering of CO_2-fixation

In the light of the ever-increasing CO_2 concentration in earth's atmosphere, assimilation and recycling of CO_2 into biofuels or valuable chemicals by autotrophic microorganisms is gaining more and more attention. While many microorganisms have the ability to assimilate CO_2, the efficiency of the natural pathways is often limited by the carboxylating enzyme per se; e.g., RubisCO that has a low turnover number and performs a futile side reaction in presence of oxygen. On the other hand, the most efficient carboxylating enzymes, the Ecrs[31], are not employed in any known natural autotrophic pathways. The CETCH cycle successfully tackles this dilemma and represents a synthetic alternative for CO_2 assimilation based on the highly efficient Ecr reaction[32]. Another, theoretically designed, synthetic CO_2-fixation module is the HOPAC cycle[32]. In the HOPAC cycle, the reaction sequence from 3HP to methylmalonyl-CoA is catalyzed by three individual enzymes, whereof one is the efficient carboxylase Ecr (**Figure 2**). The same overall reaction sequence is part of the 3HP bi-cycle in *C. aurantiacus* and the Pcs module in *Erythrobacter* sp. NAP1 as described in chapter 4. In the latter two, Pcs and propionyl-CoA carboxylase (Pcc) are employed as biocatalysts. While the ATP-consuming biotin-dependent Pcc lags behind Ecr in terms of carboxylation efficiency, this set of enzymes might nevertheless prove beneficial. As pointed out in chapter 2, acrylyl-CoA is a highly reactive intermediate and its sequestration by Pcs avoids futile or harmful side reactions. Moreover, the fusion of the three catalytic domains in Pcs guarantees optimal stoichiometry in contrast to the three stand-alone enzymes. On another note, the reduced number of enzymes required to run the HOPAC cycle with Pcs could be significant regarding protein burden when implementing it *in vivo*.

In chapter 3 we use targeted mutagenesis to evoke carboxylation activity in the Pcs reductase domain. A carboxylating Pcs could spare yet another enzyme in the HOPAC cycle and cut down the energy requirement by one equivalent of ATP. For the same reason, a carboxylating Pcs could also prove beneficial within *Erythrobacter* sp. NAP1.

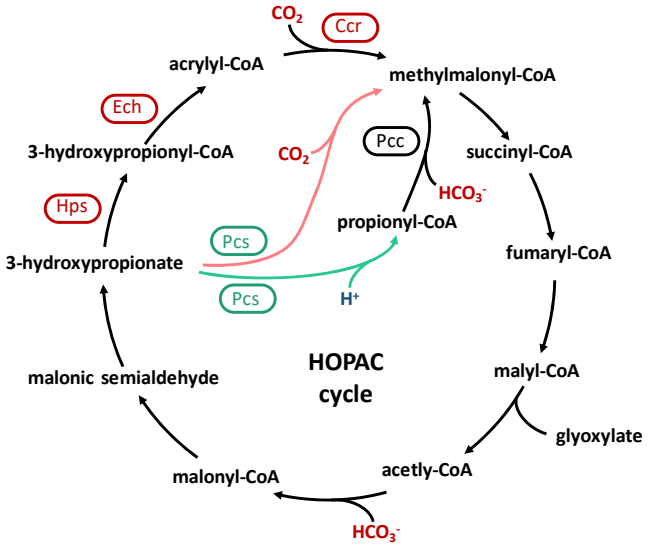


Figure 2: Potential role of Pcs in the HOPAC cycle. Pcs wildtype (shown in cyan) together with Pcc or a carboxylating Pcs variant (shown in salmon) convert 3HP directly into methylmalonyl-CoA and thereby skips the three steps catalyzed by 3HP synthetase (Hps), enoyl-CoA hydratase (Ech) and crotonyl-CoA carboxylase/reductase (Ccr). The reactive intermediate acrylyl-CoA is thereby sequestered.

The close phylogenetic relation and structural similarity of Ecrs and Pcs reductase domains paved the way for targeted mutagenesis of the latter. The two main principles for efficient carboxylation in Ecrs, namely anchoring of CO_2 and exclusion of water from the active site, were integrated into the Pcs reductase scaffold. The significantly enhanced carboxylation efficiency, however, was accompanied by a loss of reaction velocity. This trade-off between CO_2 affinity and velocity is a phenomenon often met in the attempt to optimize carboxylases, particularly in the case of RubisCO[33-36]. Insufficient understanding of structure-function relationships limits the success in rational engineering. Additionally, a systems-wide modeling study has revealed that the evolution of turnover frequency in multifunctional enzymes is constrained by strong epistasis[37]. Increasing the catalytic efficiency in one active site does not improve the overall reaction and often a neutral mutation needs to be acquired for further fitness gains. In order to free the enzyme engineering from these and other constraints, directed evolution using an activity-based selection system could be applied. This approach has yielded the greatest improvement of RubisCO in terms of carboxylation activity[38]. A selection system to directly couple improved carboxylation efficiency of Pcs to growth is presented in the Outlook (chapter 5.5).

5.4 Pcs: a physiological role in redox homeostasis within AAP bacteria

The redox state of a cell tightly regulates many biological processes and its imbalance can cause the accumulation of toxic products. It is therefore crucial to maintain the cellular redox homeostasis. For that matter, cells express several sensors that detect the redox environment and induce the corresponding regulatory responses.

Oxidative stress is usually associated with a cellular accumulation of reactive oxygen species (ROS), which cause cellular damage. ROS are formed during aerobic respiration when a single electron is transferred onto oxygen, mainly by flavoproteins[39], before reaching the end of the electron transport chain. The presence of ROS is sensed within the organism and induces a defense reaction, which includes ROS scavengers such as a superoxide dismutase, catalase and peroxidase or also non-enzymatic antioxidants[39]. Aerobic phototrophic bacteria, like purple non-sulfur and AAP bacteria, additionally produce carotenoids. The conjugated electron system in these pigments can effectively quench ROS or also an over-excited BChl *a* triplet state, which could otherwise react with molecular oxygen to form harmful singlet oxygen[40]. This seems especially crucial in AAP bacteria, where photosynthesis and thus the presence of BChl *a* occurs under aerobic conditions.

On the other hand, photosynthesis creates a surplus of reducing equivalents which induces reductive stress within the cell. Again, cells have evolved several strategies to re-establish redox homeostasis. Usually, reducing equivalents are re-oxidized in the process of cellular respiration, where electrons are relayed onto a terminal electron acceptor (oxygen in aerobes and e.g., nitrate in anaerobes) while energy is conserved in form of ATP. However, photosynthetic activity inhibits cellular respiration as reported almost a century ago for purple non-sulfur bacteria[41,42]. This inhibitory effect might be due to the membrane potential, which is higher in illuminated cells compared to respiring cells[43-45]. The same observation of light-induced inhibition of respiration has recently been made in some AAP bacteria as well[46-48]. Taken together, photosynthesis creates an excess of reducing equivalents while suppressing the main process for their regeneration. Hence, it is essential for phototrophic bacteria to activate other reductive metabolic pathways upon illumination.

We believe to have identified one of these reductive metabolic pathways in *Erythrobacter* sp. NAP1 to be the Pcs module as described in chapter 4. This pathway converts acetyl-CoA into succinyl-CoA via the two key enzymes Pcs and Mcr. The light-dependent regulation of these two enzymes supports our hypothesis. The Pcs module requires three equivalents of ATP and NADPH each to assimilate acetyl-CoA into one C4 molecule and thus has a strong impact on the cellular redox state. Furthermore, acetyl-CoA as a central metabolic intermediate independent of the carbon source

represents a convenient starting substrate for a redox state-maintaining pathway like the Pcs module. During photosynthesis, when the required energy could be (partly) provided by light, the co-assimilation of CO_2 in the Pcs module might offer a significant growth advantage in the nutrient-scarce waters where *Eryhtrobacter* sp. NAP1 resides. In the absence of light, however, this pathway seems rather costly for acetate assimilation. We showed indeed, that the glyoxylate cycle[49] is mainly used for that matter. In this canonical acetate assimilation pathway, two molecules of acetyl-CoA are assimilated into one C4 molecule while generating one equivalent of NADH.

Within their natural environment, AAP bacteria are exposed to both extremes of redox stress during a diurnal cycle: reductive stress during the photosynthetic light phase and oxidative stress in the respiratory dark phase. Adaptation to the different regimes is therefore regularly required and should be energy-efficient. The results presented in chapter 4 would be well in line with this requirement. Only the two large key enzymes Pcs and Mcr are upregulated under light conditions, while the other enzymes are maintained at a constant level.

The data in chapter 4 provide initial evidence about the role of the Pcs module within *Erythrobacter* sp. NAP1. However, further experiments will be necessary to prove its activity and regulation. Stable isotope labeling studies of cells grown in the dark or with a light/dark cycle will shed more light on the activity and regulation of the Pcs module. Furthermore, the necessity of the Pcs module for redox homeostasis during photosynthesis should be investigated. However, the perturbation of the Pcs module might prove difficult. There is no specific inhibitor to any of the enzymes and *Erythrobacter* sp. NAP1 remains genetically inaccessible despite various transformation attempts. Alternatively, the Pcs module might be perturbed in a different, genetically accessible AAP bacterium or it could be insightful to replace the reductive EMCP in *D. shibae* with the Pcs module.

5.5 Outlook: Directed evolution of Pcs

Structural and mechanistic complexity of Pcs limited the targeted evolution of the Pcs reductase domain towards improved carboxylation efficiency (chapter 3). Converting the reductase domain into a carboxylase drastically impaired the turnover rate of the Pcs overall reaction. To overcome these constraints, a carboxylase activity-based selection system for the directed evolution of Pcs needs to be set up.

Pcs wildtype catalyzes the three-step reaction sequence from 3HP to propionyl-CoA. Successful evolution of the reductase domain into a carboxylase yields (*S*-)methylmalonyl-CoA as the product of Pcs. (*S*-)Methylmalonyl-CoA can be further converted by two enzymes, methylmalonyl-CoA epimerase (Epi) and mutase (Mcm), into the central carbon metabolite succinyl-CoA. Cells require succinyl-CoA to synthesize the essential cell wall component peptidoglycan via di-aminopimelic acid (DAP)[50]. Furthermore, succinyl-CoA serves as a precursor in the biosynthesis of the amino acids lysine (via DAP) and methionine. In order to force growth depending on the Pcs carboxylation efficiency, it would be desirable to engineer an *E. coli* selection strain, where the essential succinyl-CoA exclusively derives from 3HP via the activity of Pcs, Epi and Mcm (**Figure 3**). This requires the integration of the three genes for the constitutive expression of Pcs, Epi and Mcm within a succinyl-CoA auxotrophic strain. To that end, the *Rhodobacter sphaeroides* genes encoding for Epi and Mcm organized in an operon will be genomically integrated using the no-SCAR protocol[51]. The *pcs* library, which can be generated by random mutagenesis using error-prone PCR (polymerase chain reaction), will be expressed from a plasmid. Partially carboxylating Pcs variants will still generate the reduction product propionyl-CoA, which will be metabolized by the endogenous methylcitrate cycle (**Figure 3**). Additional integration of the *Methylorubrum extorquens* propionyl-CoA carboxylase (Pcc) can serve as proofreading strategy to convert the Pcs reductase product propionyl-CoA into methylmalonyl-CoA.

E. coli can be turned auxotrophic for succinyl-CoA by deleting the genes for the two TCA cycle enzymes 2-oxoglutarate dehydrogenase (SucAB) and succinyl-CoA synthetase (SucCD)[52]. However, this strain (referred to as selection strain A) still generates succinate, which might be activated to succinyl-CoA by non-specific synthetases or CoA transferases. To avoid this, an alternative selection strain that additionally lacks succinate production could be designed (referred to as selection strain B). To this end, the genes for the TCA cycle enzymes SucAB, the succinate dehydrogenase (Sdh) and fumarate reductase

(Frd) as well as the glyoxylate cycle gene for isocitrate lyase (AceE) need to be deleted. Furthermore, the operon encoding for the methylcitrate cycle, *prpBCDE*, needs to be excised to prevent succinate generation from propionyl-CoA. Accumulation of propionyl-CoA, however, can be toxic to the cell due to competitive inhibition of the citrate synthase[53]. Both selection strains require two additional genes to be deleted; *scpB* and *scpC* encoding for a methylmalonyl-CoA decarboxylase and a propionyl-CoA:succinate CoA transferase, respectively.

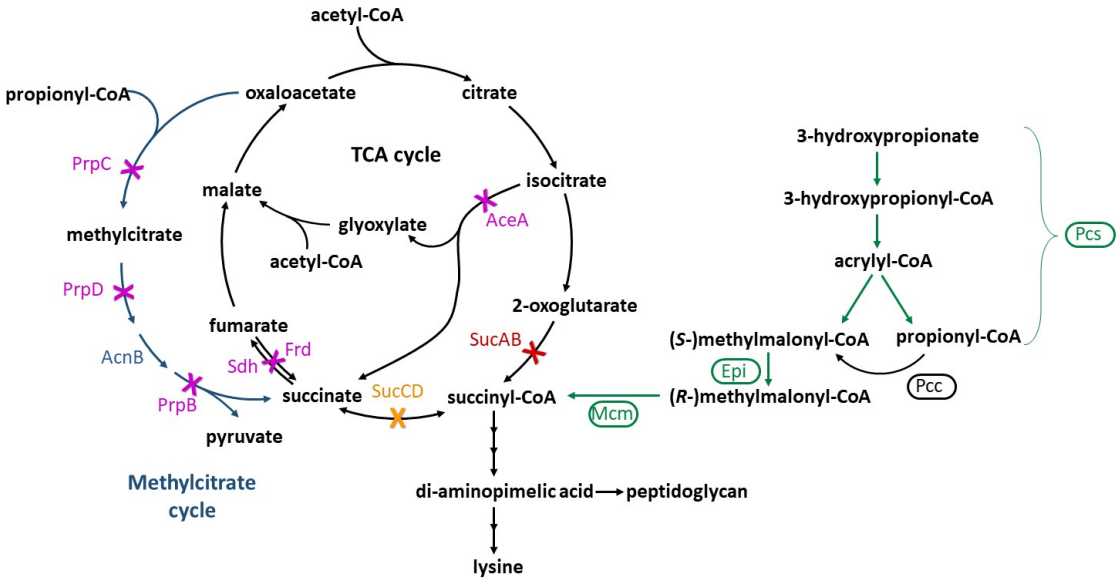


Figure 3: Schematic of two potential *E. coli* selection strains for directed evolution of Pcs. In both selection strains, succinyl-CoA can exclusively be generated from 3HP via a carboxylating Pcs variant, Epi and Mcm. The genes encoding for the enzymes depicted in red, need to be deleted in selection strain A and B; in orange; need to be deleted in selection strain A only; in pink, need to be deleted in selection strain B only; in green, need to be implemented in both selection strains. Pcc could be additionally implemented as proofreading strategy.

When designing selection strains, it is important to adapt the stringency to the potential flux through the enzyme or pathway to be evolved. Here, the Pcs carboxylation activity needs to sustain the succinyl-CoA pool only, which accounts for roughly 0.2 – 1.5 mM in *E. coli* cells[54]. The rest of the biomass can be assimilated from a defined carbon source (e.g., glycerol). Additionally, the suggested selection strains allow for clever tuning of the selection pressure by titration of lysine, methionine and DAP (or alternatively succinate in case of selection strain B) to the growth medium. Selection pressure can be further adapted by changing the expression level of the Pcs variants. The Golden Gate assembly of the expression plasmid from individual parts allows for quick exchange of promoter or ribosomal binding site to yield various expression levels. Furthermore, the expression plasmid backbone could be exchanged to vary the plasmid copy number within the cells.

Directed evolution is a powerful tool also or especially for the engineering of multidomain/enzyme complexes, where complex interactions are not obvious or easy to anticipate. In previous studies, allosteric domain communication has been improved in the three-domain chaperone DnaK[55] or even introduced *de novo* into a Bayer-Villiger monooxygenase[56] through the directed evolution approach. Moreover, the beta subunit of the canonical channeling enzyme tryptophan synthetase (TrpB), which loses most of its activity when detached from the heterodimeric complex, could be regained through directed evolution[57]. It has been proposed that the allosteric effects, which are usually implied by the alpha subunit, were reestablished in the improved variants. In the light of these previous achievements, it appears promising that the interdomain communication in Pcs could be recreated by random mutagenesis and selection. Eliminating these constraints in the carboxylating Pcs variant (D1302S T1753M) would leave the reductase as the rate-limiting domain with an apparent k_{cat} of 3.6 ± 0.2 s^{-1} (**Table S1** in chapter 3), which is in the range of the overall turnover rate of Pcs wildtype (4.7 ± 0.3 s^{-1}). Together with a decrease of the K_M for CO_2, the improved Pcs variant could be efficiently applied in the synthetic CO_2-fixing HOPAC cycle.

5.6 References

1 Crawford, J. M. *et al.* Deconstruction of iterative multidomain polyketide synthase function. *Science* **320**, 243-246 (2008).

2 Singh, H., Arentson, B. W., Becker, D. F. & Tanner, J. J. Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site. *PNAS* **111**, 3389-3394 (2014).

3 Hawkins, A. R. & Lamb, H. K. The molecular biology of multidomain proteins selected examples. *Eur. J. Biochem.* **232**, 7-18 (1995).

4 Wheeldon, I. *et al.* Substrate channelling as an approach to cascade reactions. *Nat. Chem.* **8**, 299 (2016).

5 Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041-1072 (2007).

6 Vögeli, B. *et al.* Archaeal acetoacetyl-CoA thiolase/HMG-CoA synthase complex channels the intermediate via a fused CoA-binding site. *PNAS* **115**, 3380-3385 (2018).

7 Ishikawa, M., Tsuchiya, D., Oyama, T., Tsunaka, Y. & Morikawa, K. Structural basis for channelling mechanism of a fatty acid β-oxidation multienzyme complex. *EMBO J.* **23**, 2745-2754 (2004).

8 Sathyanarayanan, N. *et al.* Molecular basis for metabolite channeling in a ring opening enzyme of the phenylacetate degradation pathway. *Nat. Commun.* **10**, 1-12 (2019).

9 Mouilleron, S., Badet-Denisot, M.-A. & Golinelli-Pimpaneau, B. Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. *J. Biol. Chem.* **281**, 4404-4412 (2006).

10 Smith, N. E., Vrielink, A., Attwood, P. V. & Corry, B. Biological channeling of a reactive intermediate in the bifunctional enzyme DmpFG. *Biophys. J.* **102**, 868-877 (2012).

11 Leys, D., Basran, J. & Scrutton, N. S. Channelling and formation of 'active' formaldehyde in dimethylglycine oxidase. *EMBO J.* **22**, 4038-4048 (2003).

12 Yourno, J., Kohno, T. & Roth, J. R. Enzyme evolution: generation of a bifunctional enzyme by fusion of adjacent genes. *Nature* **228**, 820-824 (1970).

13 Bazan, J. F., Fletterick, R. J. & Pilkis, S. J. Evolution of a bifunctional enzyme: 6-phosphofructo-2-kinase/fructose-2, 6-bisphosphatase. *PNAS* **86**, 9642-9646 (1989).

14 McCarthy, A. D. & Hardie, D. G. Fatty acid synthase—an example of protein evolution by gene fusion. *Trends Biochem. Sci.* **9**, 60-63 (1984).

15 Andersson, J. O. & Roger, A. J. Evolutionary analyses of the small subunit of glutamate synthase: gene order conservation, gene fusions, and prokaryote-to-eukaryote lateral gene transfers. *Eukaryot. Cell* **1**, 304-310 (2002).

16 Delaye, L., Becerra, A., Orgel, L. & Lazcano, A. Molecular evolution of peptide methionine sulfoxide reductases (MsrA and MsrB): on the early development of a mechanism that protects against oxidative damage. *J. Mol. Evol.* **64**, 15-32 (2007).

17 Salles, I. M., Forchhammer, N., Croux, C., Girbal, L. & Soucaille, P. Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*. *Metab. Eng.* **9**, 152-159 (2007).

18 Burns, D. M., Horn, V., Paluh, J. & Yanofsky, C. Evolution of the tryptophan synthetase of fungi. Analysis of experimentally fused *Escherichia coli* tryptophan synthetase alpha and beta chains. *J. Biol. Chem.* **265**, 2060-2069 (1990).

19 Hyde, C., Ahmed, S., Padlan, E., Miles, E. W. & Davies, D. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263**, 17857-17871 (1988).

20 Kominek, J. *et al.* Eukaryotic acquisition of a bacterial operon. *Cell* **176**, 1356-1366. e1310 (2019).

21 Marcotte, E. M. *et al.* Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**, 751-753 (1999).

22 Sharma, H., Landau, M. J., Vargo, M. A., Spasov, K. A. & Anderson, K. S. First three-dimensional structure of Toxoplasma gondii thymidylate synthase–dihydrofolate reductase: insights for catalysis, interdomain interactions, and substrate channeling. *Biochemistry* **52**, 7305-7317 (2013).

23 Myers, R. S., Amaro, R. E., Luthey-Schulten, Z. A. & Davisson, V. J. Reaction coupling through interdomain contacts in imidazole glycerol phosphate synthase. *Biochemistry* **44**, 11974-11985 (2005).

24 Tsuji, S. Y., Wu, N. & Khosla, C. Intermodular Communication in Polyketide Synthases: Comparing the Role of Protein− Protein Interactions to Those in Other Multidomain Proteins. *Biochemistry* **40**, 2317-2325 (2001).

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

26 Bahar, I. & Jernigan, R. L. Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry* **38**, 3478-3490 (1999).

27 Loutchko, D., Gonze, D. & Mikhailov, A. S. Single-molecule stochastic analysis of channeling enzyme tryptophan synthase. *J. Phys. Chem. B* **120**, 2179-2186 (2016).

28 Demirci, H. *et al.* Coupled inter-subunit dynamics enable the fastest CO_2-fixation by reductive carboxylases. *bioRxiv*, 607101 (2019).

29 Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome research* **14**, 1188-1190 (2004).

30 Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic acids research* **18**, 6097-6100 (1990).

31 Erb, T. J. Carboxylases in natural and synthetic microbial pathways. *Appl. Environ. Microbiol.* **77**, 8466-8477 (2011).

32 Schwander, T., von Borzyskowski, L. S., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354**, 900-904 (2016).

33 Long, S. P., Marshall-Colon, A. & Zhu, X.-G. Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* **161**, 56-66 (2015).

34 Parry, M., Madgwick, P., Carvalho, J. & Andralojc, P. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci.* **145**, 31 (2007).

35 Tcherkez, G. G., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *PNAS* **103**, 7246-7251 (2006).

36 Savir, Y., Noor, E., Milo, R. & Tlusty, T. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *PNAS* **107**, 3475-3480 (2010).

37 Heckmann, D., Zielinski, D. C. & Palsson, B. O. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. *Nat. Commun.* **9**, 1-9 (2018).

38 Cai, Z., Liu, G., Zhang, J. & Li, Y. Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein Cell* **5**, 552-562 (2014).

39 Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. *Nat. Rev. Microbiol.* **11**, 443-454 (2013).

40 Fraser, N. J., Hashimoto, H. & Cogdell, R. J. Carotenoids and bacterial photosynthesis: The story so far... *Photosynth. Res.* **70**, 249-256 (2001).

41 Nakamura, H. Über die Photosynthese bei der schwefelfreien Purpurbakterie, *Rhodobacillus palustris*. *Beiträge zur Stoffwechselphysiologie der Purpurbakterien, I. Acta Phytochimica* **9**, 189-229 (1937).

42 Fork, D. & Goedheer, J. Studies on light-induced inhibition of respiration in purple bacteria: action spectra for *Rhodospirillum rubrum* and *Rhodopseudomonas spheroides*. *Biochim. Biophys. Acta* **79**, 249-256 (1964).

43 Ramírez, J. & Smith, L. Synthesis of adenosine triphosphate in intact cells of *Rhodospirillum rubrum* and *Rhodopseudomonas spheroides* on oxygenation or illumination. *Biochim. Biophys. Acta* **153**, 466-475 (1968).

44 Clark, A. J., Cotton, N. P. & Jackson, J. B. Interaction between respiration and photosynthesis mediated by the protonmotive force in intact cells of Rhodopseudomonas capsulata. *Biochem. Soc. Trans.* **10**, 259-259 (1982).

45 McCarthy, J. E. & Ferguson, S. J. Respiratory control and the basis of light-induced inhibition of respiration in chromatophores from Rhodopseudomonas capsulata. *Biochem. Biophys. Res. Commun* **107**, 1406-1411 (1982).

46 Koblížek, M. *et al.* Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol.* **180**, 327-338 (2003).

47 Tomasch, J., Gohl, R., Bunk, B., Diez, M. S. & Wagner-Döbler, I. Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *ISME J.* **5**, 1957-1968 (2011).

48 Bill, N. *et al.* Fixation of CO_2 using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ. Microbiol.* **19**, 2645-2660 (2017).

49 Kornberg, H. & Krebs, e. H. Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature* **179**, 988-991 (1957).

50 Yu, B. J. *et al. sucAB* and *sucCD* are mutually essential genes in *Escherichia coli*. *FEMS Microbiol. Lett.* **254**, 245-250 (2006).

51 Reisch, C. R. & Prather, K. L. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 1-12 (2015).

52 Mattozzi, M. d., Ziesack, M., Voges, M. J., Silver, P. A. & Way, J. C. Expression of the sub-pathways of the *Chloroflexus aurantiacus* 3-hydroxypropionate carbon fixation bicycle in *E. coli*: Toward horizontal transfer of autotrophic growth. *Metab. Eng.* **16**, 130-139 (2013).

53 Man, W.-J., Li, Y., O'Connor, C. D. & Wilton, D. C. The binding of propionyl-CoA and carboxymethyl-CoA to *Escherichia coli* citrate synthase. *Biochim. Biophys. Acta* **1250**, 69-75 (1995).

54 Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593 (2009).

55 Aponte, R. A., Zimmermann, S. & Reinstein, J. Directed evolution of the DnaK chaperone: mutations in the lid domain result in enhanced chaperone activity. *J. Mol. Biol.* **399**, 154-167 (2010).

56 Wu, S., Acevedo, J. P. & Reetz, M. T. Induced allostery in the directed evolution of an enantioselective Baeyer–Villiger monooxygenase. *PNAS* **107**, 2775-2780 (2010).

57 Buller, A. R. *et al.* Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *PNAS* **112**, 14599-14604 (2015).

Acknowledgements

Together with my PhD a great and memorable chapter of my life will come to an end. I have got to know so many wonderful people during this time, who did not only touch my science but also my heart. In the following I would like to mention some of them.

First, I would like to thank my supervisor **Tobias Erb**. When you suggested this project to me, I had no idea where it will take me. You granted me the freedom to use my creativity and develop my own ideas but were always available to provide much needed advice. It was a challenging and highly rewarding journey that prepared me well for my future career as a scientist.

I would also like to thank my thesis advisory committee members, **Prof. Erhard Bremer** and **Dr. Hannes Link**, for their valuable inputs and encouragement.

I am very happy to still be in touch with previous colleagues and collaborators. Thank you **Pete** for being a great and patient supervisor during my Master thesis and introducing me to the wonderful world of carboxylases. I am grateful for **Basti** who never missed an opportunity to brainstorm about assay results with me and who awoke my fascination for enzymes. Thanks to **Thomy**, who, even in his absence, helped a lot to overcome the bureaucracy of handing-in a thesis. A special thanks goes to **Tristan**, who opened the door to protein crystallography for me and patiently helped me at every stage towards the Pcs crystal structure. I will never forget the trip to the synchrotron with you.

The whole **Erb group** deserves a great deal of credit for lifting up my mood even on bad days. I really enjoyed the mixture between great work, vivid scientific exchange and cheerful shenanigans. **Gabo**, I appreciated you sharing your expertise on Ecrs. Thank you for all the great non-hype parties – less ghost-shots would have done, too! **Simon**, thank you for your great musical support in the lab and live at your Big Band concerts. Your sleeping skills are amazing! **Tarryn**, you always had an open ear for any of my concerns; I will never forget that. **Nina**, I have found a Zug-buddy and friend in you, thank you for all the lifts back home. A special thanks goes to **Jan**, who held this lab together and was always approachable for any kind of help. You spent your well-deserved free time to solve the Pcs crystal structure and you were there when SAXS revealed my true spirit animal: the chicken!

Last but most definitely not least, I would like to thank **my friends** and **family**. You celebrated my success and cheered me up in times of disappointment, even though you could not always relate. That is more than I could have ever expected. Special thanks goes to my dear husband **Wikus**, who always supported me.

Iria Grundling
maiden name **Bernhardsgrütter**
Born: 12.06.1991 in Zug, Switzerland

Research/Work Experience

05/2016 – present **PhD**
Max Planck Institute for Terrestrial Microbiology, Marburg, Germenay, Principal investigator: Prof. Dr. Tobias J. Erb
Project: Propionyl-CoA Synthase: Characterization, Engineering and Physiological Role of a Trifunctional Fusion Enzyme

09/2014 – 01/2015 **Industrial research internship**
Vifor Pharma, Schlieren, Switzerland

09/2008 – 10/2008 **Research internship**
Institute for Research in Ophthalmology (IRO), Sion, Switzerland

Education

09/2013 – 02/2016 **Master of Science in Biology,** ETH Zürich, Zürich, Switzerland
Major in Microbiology and Immunology

09/2010 – 09/2013 **Bachelor of Science in Biology,** ETH Zürich, Zürich, Switzerland
Major in Chemistry

08/2004 – 08/2010 **Matura,** Kantonsschule Zug, Zug, Switzerland
Major in Biology and Chemistry

Publications

Iria Bernhardsgrütter, Bastian Vögeli, Tristan Wagner, Dominik M. Peter, Niña Socorro Cortina, Jörg Kahnt, Gert Bange, Sylvain Engilberge, Eric Girard, François Riobé, Olivier Maury, Seigo Shima, Jan Zarzycki and Tobias J. Erb. The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite. ***Nat. Chem. Biol.*** **2018**, 14, 1127.

Iria Bernhardsgrütter, Kristina Schell, Dominik M. Peter, Farshad Borjian, David Adrian Saez, Esteban Vöhringer-Martinez and Tobias J. Erb. Awakening the sleeping carboxylase function of enzymes: engineering the natural CO2-binding potential of reductases. ***JACS*** **2019**, 141 (25), 9778-9782

Katharina Kremer, Murial C. F. van Teeseling, Lennart Schada von Borzyskowski, **Iria Bernhardsgrütter**, Rob J. M. van Spanning, Andrew J. Gates, Mitja N. P. Remus-Emsermann, Martin Thanbichler, Tobias J. Erb. Dynamic metabolic rewiring enables efficient acetyl coenzyme A assimilation in *Paracoccus denitrificans.* ***mBio*** **2019**, 10 (4), e00805-19

Iria Bernhardsgrütter, Gabriele Stoffel, Tobias J. Erb. Schöne neue Biologie? Synthetisch-biologische Ansätze zur CO_2-Umwandlung. ***Biospektrum*** **2020**, 26, 24-27

Charles A. R. Cotton, **Iria Bernhardsgrütter**, Simon Burgener, Luca Schulz, Hai He, Stepan Toman, Marian Dempfle, Nicole Paczia, Joachim Kopka, Steffen N. Lindner, Tobias J. Erb and Arren Bar-Even. Underground aerobic and anaerobic isoleucine biosynthesis pathways in *E. coli*. *(under review)*

Iria Bernhardsgrütter, Timo Glatter, Tobias J. Erb. Propionyl-CoA synthase pathway in Erythrobacter sp. NAP1: a light-induced metabolism to maintain the redox homeostasis. *(in preparation)*

Lennart Schada von Borzyskowski, **Iria Bernhardsgrütter**, Tobias J. Erb. Biochemical unity revisited: Microbial central carbon metabolism holds new discoveries, multi-tasking pathways, and redundancies with a reason. *(submitted)*

Honors & Awards

Poster Award Awarded by ACS Division of Biological Chemistry ($500)
at GRC for Enzymes, Coenzymes and Metabolic Pathways 2018
(Waterville Valley, MA, USA)

IMPRS funding PhD student funding by the IMPRS-mic graduate school

Additional Skills

Certificate Safety in Gene Technology (according to §15 GenTSV)
Umweltinsitut Offenbach GmbH (09/2019)

Language German: Native · English: Fluent · French: Basic

Extra-Curricular Activities

04/2012 – 08/2015 **Part-time job as a waitress**
Confectionery von Rotz, Baar, Switzerland

01/2013 – 02/2013 **Volunteering** in South Africa
Animal Care and Teaching in a Township

2004 – 2012 **Several part-time jobs**
Medela AG, Steinhausen, Switzerland & Emu AG, Baar, Switzerland