CO_2 Fixation

International Edition: DOI: 10.1002/anie.201505282
German Edition: DOI: 10.1002/ange.201505282

Screening and Engineering the Synthetic Potential of Carboxylating Reductases from Central Metabolism and Polyketide Biosynthesis

*Dominik M. Peter, Lennart Schada von Borzyskowski, Patrick Kiefer, Philipp Christen, Julia A. Vorholt, and Tobias J. Erb**

Abstract: *Carboxylating enoyl-thioester reductases (ECRs) are a recently discovered class of enzymes. They catalyze the highly efficient addition of CO_2 to the double bond of α,β-unsaturated CoA-thioesters and serve two biological functions. In primary metabolism of many bacteria they produce ethylmalonyl-CoA during assimilation of the central metabolite acetyl-CoA. In secondary metabolism they provide distinct α-carboxyl-acyl-thioesters to vary the backbone of numerous polyketide natural products. Different ECRs were systematically assessed with a diverse library of potential substrates. We identified three active site residues that distinguish ECRs restricted to C4 and C5-enoyl-CoAs from highly promiscuous ECRs and successfully engineered a selected ECR as proof-of-principle. This study defines the molecular basis of ECR reactivity, allowing for predicting and manipulating a key reaction in natural product diversification.*

The biosynthetic potential of nature is impressively reflected in the diversity of secondary metabolism. More than 325 000 natural products have been described to date.[1] All of these compounds differ strongly with respect to chemical structure and biological activity,[2] yet the biosynthesis of their structural backbone is based on simple elongation reactions from basic building blocks, the so-called extender units.[3] A prominent example is the large class of polyketides that are assembled through subsequent Claisen condensation reactions from α-carboxylacylthioester units.[4] The standard building blocks in polyketide assembly lines are malonyl-coenzyme A (CoA) and methylmalonyl-CoA, which are mainly provided by enzymes of fatty acid metabolism that α-carboxylate acetyl-CoA and propionyl-CoA, respectively.[5] However, an increasing number of polyketides show variations from this common principle, because their side-chains suggest non-standard extender units to vary the structural backbone.[6] Additional extender units include ethylmalonyl-CoA[7] chloroethylmalonyl-CoA,[8] propylmalonyl-CoA[9] and longer chain derivatives,[10] as well as isobutylmalonyl-CoA[7b] and other branched-chain analogues.[11]

These non-traditional extender units are provided by the reductive carboxylation of α,β-unsaturated acyl-CoA thioesters by carboxylating enoyl-thioester reductases (ECR), a novel class of enzymes that was described only recently. The prime example in the ECR family is crotonyl-CoA carboxylase/reductase (Ccr) that catalyzes the NADPH-dependent carboxylation of crotonyl-CoA into ethylmalonyl-CoA and that is one of the most efficient CO_2-fixing enzymes described to date.[12] As of 2014, the ECR family features more than 900 homologues that can be physiologically divided into two subfamilies, a large ECR subfamily of so-called primary metabolism Ccrs (ECR-1) that function in the ethylmalonyl-CoA pathway, a recently discovered central metabolic pathway for acetyl-CoA assimilation,[13] and a subfamily of secondary metabolism ECRs (ECR-2) that are associated with polyketide biosynthesis.[6]

Given the fact that ECRs are key enzymes to alter the polyketide backbone, surprisingly little is known on the structure and catalytic mechanism of these proteins.[14] Yet such information is indispensable 1) to understand what factors direct and control the incorporation of non-traditional extender units into the growing polyketide chain, 2) to assign the biosynthetic function of ECR homologues for correctly predicting the polyketide product structure, and 3) to manipulate ECR reactivity to rationally engineer polyketide biosynthesis.

Motivated by the above, we sought to investigate the molecular basis for substrate specificity of ECRs in more detail. We first established a diverse substrate library of enoyl-CoA thioesters to cover the natural and non-natural chemical space of polyketide extender units (**1**–**19**, Figure 1). This library was used to test eight phylogenetically diverse ECRs. From the ECR-1 subfamily that is supposedly specific for crotonyl-CoA (**1**) as substrate, we included four homologues: CcrCc from *Caulobacter crescentus*, CcrSg from *Streptomyces griseus*, CcrSb from *Streptomyces bottropensis*, and CcrPd from *Paracoccus denitrificans*. From the ECR-2 subfamily, we included CinF from *Streptomyces* sp. JS360, RevT from *Streptomyces* sp. SN-593, SalG from *Salinispora tropica*, and EcrSh from *Streptomyces hygroscopicus* that were reported to catalyze the reductive carboxylation of octenoyl-CoA (CinF),[14a] hexenoyl-CoA (RevT)[11a] and chlorocrotonyl-CoA, as well as pentenoyl-CoA, repectively (SalG).[8] Above candidates were heterologously expressed in *Escherichia coli* BL21 cells and purified to homogeneity for individual screens of above substrate library in a HPLC-MS based in vitro assay (Supporting Information, Figure S1).

[*] D. M. Peter, L. Schada von Borzyskowski, Dr. T. J. Erb
Biochemistry and Synthetic Biology of Microbial Metabolism Group
Max-Planck-Institute for terrestrial Microbiology
Karl-von-Frisch-Strasse 10, 35043 Marburg (Germany)
E-mail: toerb@mpi-marburg.mpg.de
D. M. Peter, L. Schada von Borzyskowski, Dr. P. Kiefer, P. Christen, Prof. J. A. Vorholt, Dr. T. J. Erb
Institute of Microbiology
Eidgenössisch Technische Hochschule (ETH) Zürich
Vladimir-Prelog-Weg 4, 8050 Zürich (Switzerland)

Supporting information and ORCID(s) from the author(s) for this article are available on the WWW under http://dx.doi.org/10.1002/anie.201505282.

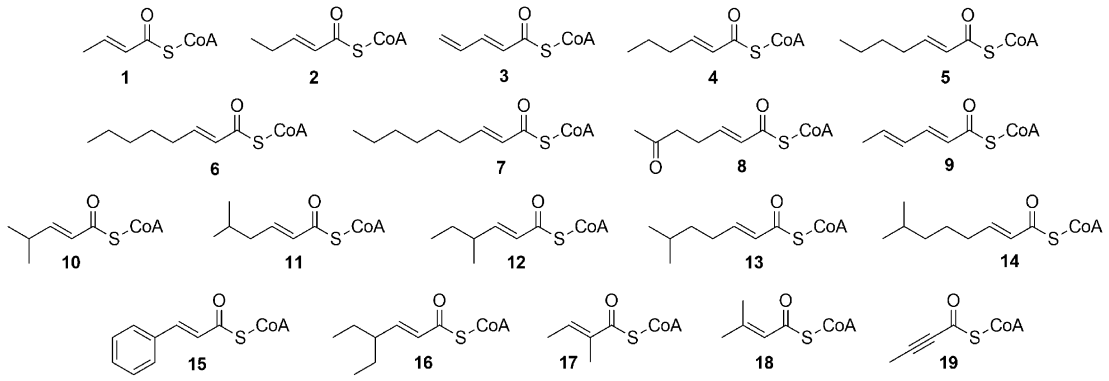

Figure 1. Enoyl-CoA thioesters used in this study as potential ECR substrates.

Notably, all of the ECRs accepted crotonyl-CoA as substrate, independent of their physiological context. Besides this common feat, clear differences were observed between the two subfamilies: ECR-1s were confined to crotonyl-CoA, accepting only pentenoyl-CoA and pent-2,4-dienoyl-CoA as additional substrates. In contrast, ECR-2s showed a surprisingly broad substrate tolerance, accepting unbranched CoA thioesters of all lengths (C4 to C9) and branched enoyl-CoA derivatives. An exception was SalG that was reported to provide (chloro)ethylmalonyl-CoA and propylmalonyl-CoA in Salinosporamide biosynthesis.[8] This ECR accepted CoA thioesters only up to C7 and as branched substrate only 4-methylpentenoyl-CoA. Our results thus indicate that ECR-1 and ECR-2 differ strongly with respect to substrate promiscuity: While ECR-1s seem to be restricted to short, unbranched CoA esters, ECR-2s are apparently much more promiscuous, accepting longer chain and even branched-chain substrates (Figure 2).

Next we aimed at identifying the molecular basis for the observed substrate spectrum in the ECR family. Inspection of the active site of different ECR crystal structures allowed us to pinpoint candidate active site residues for substrate selectivity. Conservation of these active site residues was confirmed by analyzing a multiple sequence alignment of 925 ECR homologues, resulting in the identification of three key residues that seemed to be responsible for restricting the active site of ECR-1s to short-chain substrates: C146, I169, and F373 (numbering according to the ECR-1 CcrCc; Figure 3).

To confirm the role of these active site amino acids in substrate recognition, we mutated the three sites in CcrCc to

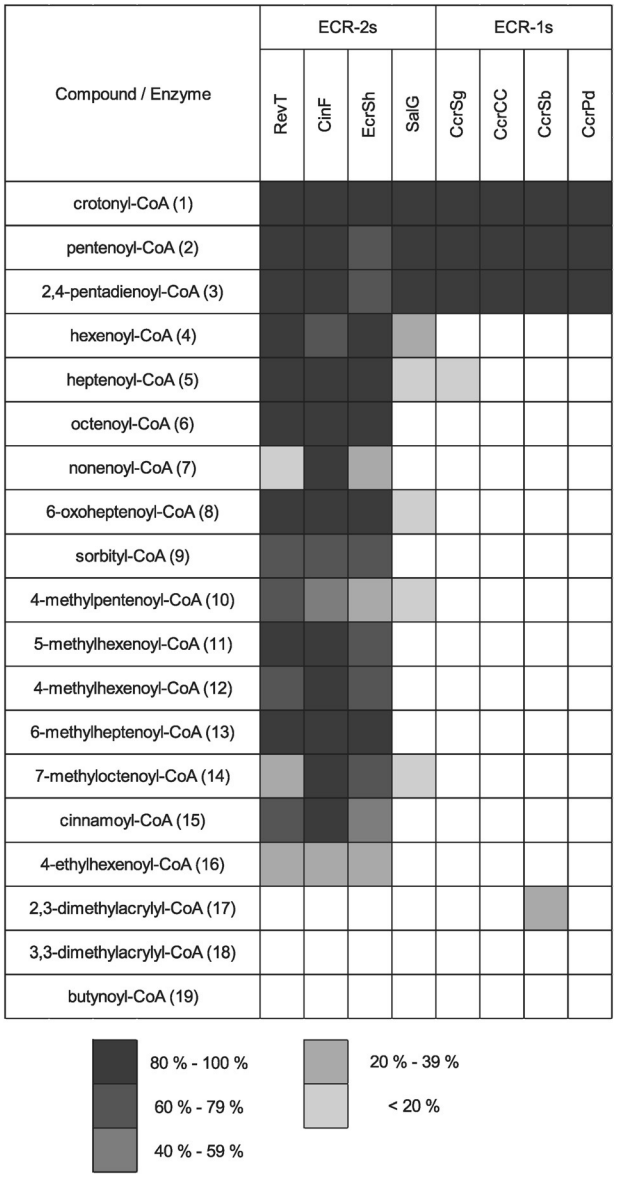

Figure 2. Substrate library screen of ECRs measured with LC-MS. All possible enzyme substrate combinations were assayed independently with 1 mM substrate, 0.4–2.5 μM enzyme for 3 h to maximize detection of poor substrates and to minimize product decarboxylation (see additional text in the Supporting Information); the incubation time for CinF assays with **5**, **6**, and **7** was shortened to 10 min. Shown in different shades of gray is the relative product formation for samples with detected carboxylated products exceeding the threshold area of 10^4 (approximately 10 μM reaction product). For complete data refer to the Supporting Information.

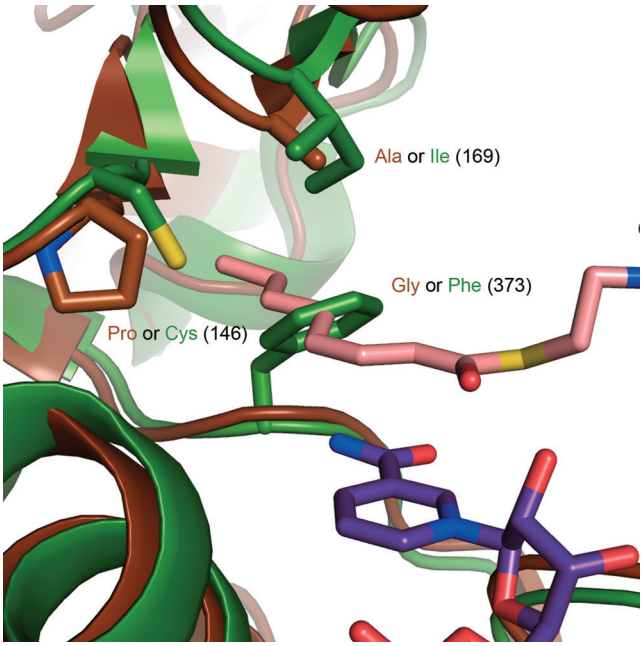

Figure 3. Superimposed structures of CinF (brown) crystallized with octenoyl-CoA (salmon) and NADPH (violet)[14a] and a Ccr from *S. collinus* accession code: 3HZZ (unpublished). The three substrate-promiscuity determining residues Pro or Cys 146, Ala or Ile 169, and Gly or Phe 373 are labeled.

Compound / Enzyme	CcrCc	CP	FG	CP / FG	IA	IA / CP	IA / FG	IA / CP / FG
crotonyl-CoA (1)								
pentenoyl-CoA (**2**)								
2,4-pentadienoyl-CoA (**3**)								
hexenoyl-CoA (**4**)								
heptenoyl-CoA (**5**)								
octenoyl-CoA (6)								
nonenoyl-CoA (**7**)								
6-oxoheptenoyl-CoA (**8**)								
sorbityl-CoA (**9**)								
4-methylpentenoyl-CoA (**10**)								
5-methylhexenoyl-CoA (11)								
4-methylhexenoyl-CoA (**12**)								
6-methylheptenoyl-CoA (**13**)								
7-methyloctenoyl-CoA (**14**)								
cinnamoyl-CoA (**15**)								
4-ethylhexenoyl-CoA (**16**)								
2,3-dimethylacrylyl-CoA (**17**)								
3,3-dimethylacrylyl-CoA (**18**)								
butynoyl-CoA (**19**)								

80 % - 100 %
60 % - 79 %
40 % - 59 %
20 % - 39 %
< 20 %

Figure 4. Substrate library screen of CcrCc wt and CcrCc single, double, and triple mutants measured with LC-MS. All possible enzyme substrate combinations were assayed independently with 1 mM substrate and 0.4–2.5 μM enzyme for 3 h to maximize detection of poor substrates and to minimize product decarboxylation (see additional text in the Supporting Information); the incubation time for CcrCc CP assays for **1** was shortened to 10 min. Shown in different shades of gray is the relative product formation for samples with detected carboxylated products exceeding the threshold area of 10^4 (approximately 10 μM reaction product). For complete data refer to the Supporting Information. Substrates marked in bold were kinetically characterized in more detail using UV/Vis assays (see Figure 5).

the major consensus residues of the ECR-2 subfamily (namely C146P, I169A, and F373G) and screened the substrate promiscuity of the resulting single and double mutants of CcrCc, as well as the corresponding triple mutant. The substrate spectrum of the enzyme increased incrementally from individual single mutants to the triple mutant (Figure 4), the latter being essentially indistinguishable from a true member of the ECR-2 subfamily. The CcrCc triple mutant converted short-chain (**1**), long-chain (**11**), branched-chain (**8**), and also bulky substrates (**16**), demonstrating the importance of all three candidate residues in opening up the substrate binding site.

To better understand the contribution of each of the three residues to substrate promiscuity, we spectrophotometrically determined individual kinetic parameters of different model sPeterubstrates for CcrCc, all seven CcrCc mutants, as well as EcrSh, a promiscuous ECR-2 member, for comparison. As model substrates we chose a short-chain (**1**), a long-chain (**6**), and a branched-chain (**11**) substrate (Table 1; Supporting Information, Figure S2).

The catalytic efficiency for the short-chain model substrate **1** dropped gradually and nearly 2000-fold from CcrCc wt ($27 \pm 10 \times 10^5\ L\,mol^{-1}\,s^{-1}$) to the CcrCc triple mutant ($14 \pm 6 \times 10^2\ L\,mol^{-1}\,s^{-1}$). In contrast, the catalytic efficiency for the long-chain model substrate **6** increased to a maximum of $11 \pm 2 \times 10^4\ L\,mol^{-1}\,s^{-1}$ for the CcrCc triple mutant, which is almost identical to the ECR-2 subfamily member EcrSh ($13 \pm 4 \times 10^4\ L\,mol^{-1}\,s^{-1}$). Notably, the bulky model substrate **11** was converted even one order of magnitude more efficiently by the CcrCc triple mutant ($6 \pm 2 \times 10^3\ L\,mol^{-1}\,s^{-1}$) compared to ECR-2 member EcrSh ($44 \pm 7 \times 10^1\ L\,mol^{-1}\,s^{-1}$). In summary, our results point to a tradeoff in CcrCc (and likely the ECR enzyme family) between catalytic efficiency for the short natural substrate and promiscuity towards longer and bulkier substrates (Figure 5).

Finally, we sought to investigate the relevance of this tradeoff affecting catalytic efficiency in CcrCc, when the enzyme and its mutant variants are put back in its natural physiological context of primary carbon metabolism. To that

Table 1: Spectrophotometrically determined catalytic efficiency and K_M of CcrCc wt and mutants with crotonyl-CoA in relation to doubling time of the complemented *Δccr* strains of *Methylobacterium extorquens* AM1.[a]

Mutation	k_{cat}/K_M [L mol^{-1} s^{-1}]	Doubling time [h]
wt	$27\pm10\times10^5$	5.3 ± 0.3
CP	$32\pm11\times10^4$	5.4 ± 0.2
FG	$15\pm9\times10^4$	5.3 ± 0.1
IA	$18\pm8\times10^3$	5.6 ± 0.3
CP/FG	$39\pm16\times10^3$	6.2 ± 1.3
FG/IA	$92\pm2\times10^2$	13 ± 2
CP/IA	$18\pm6\times10^2$	50 ± 27
CP/IA/FG	$14\pm6\times10^2$	27 ± 5

[a] Data is shown as mean with 95% confidence interval.

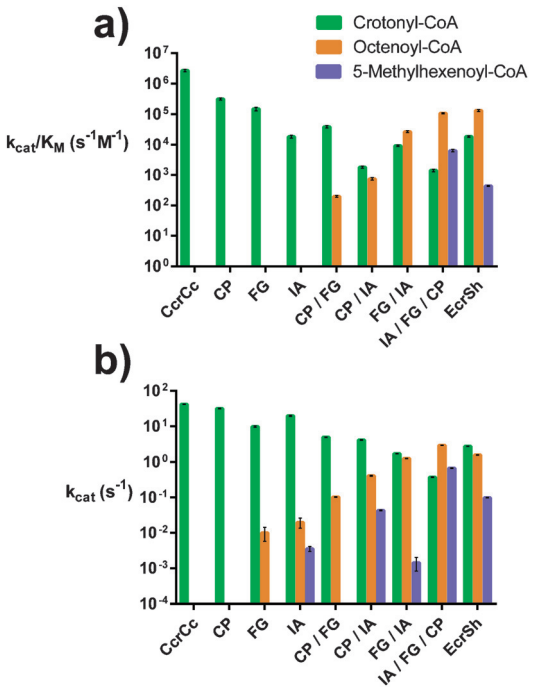

Figure 5. a) Catalytic efficiencies k_{cat}/K_M and b) turnover values k_{cat} of CcrCc wt and mutants as well as CcrSh with the three model substrates crotonyl-CoA (**1**), octenoyl-CoA (**6**), and 5-methylhexenoyl-CoA (**11**), as determined spectrophotometrically by steady-state kinetics. Catalytic efficiency for CcrCc IA and Ccr FG with **6**, as well as CcrCc FG/IA, CcrCc CP/IA, and CcrCc IA with **11** was determined in triplicates at 1 mM substrate concentration owing to low turnover rates. All kinetic values are listed in the Supporting Information, Table S1, and the respective Michaelis–Menten graphs are shown in the Supporting Information, Figure S2.

end, we used a Ccr in vivo complementation system that we established previously in a *Δccr* strain of *Methylobacterium extorquens* AM1.[16] This strain can grow on minimal medium with succinate as sole carbon source, but not with methanol as long as it lacks an operational Ccr to restore the ethylmalonyl-CoA pathway, which is indispensable for growth on C1 substrates.[15]

Growth of the complemented strains was comparable for CcrCc wt, all of the single mutants and the C146P/F373G double mutant, but was dramatically impaired for the other two double mutants and the CcrCc triple mutant (Table 1). A catalytic efficiency as low as $20–30\times10^3$ L mol^{-1} s^{-1} was sufficient to sustain wt-like growth rates. However, as soon as the catalytic efficiency fell below this threshold, the doubling time increased considerably. Because the expression levels of all CcrCc variants were of comparable range (Supporting Information, Figure S3), we propose that this threshold defines the evolutionary pressure that keeps the active site pocket closed for ECRs in primary metabolism.

Our experiments have the following implications: First, they identified three key residues that control substrate specificity in ECRs. This provides the possibility to rationally manipulate substrate selectivity of these enzymes, as demonstrated in this study with a model ECR. At the same time, our findings also allow assigning the function of ECRs in a given polyketide biosynthetic gene cluster. Using these three residues, it can be generally predicted whether the corresponding polyketide might feature a structurally diverse residue, including a longer chain, branched or bulky moiety (P146, A169, G373), or is strongly constricted to short-chain moieties (C146, I169, F373).

Second, our results suggest that the production and incorporation of non-traditional extender units in polyketides is not controlled on the level of the ECR reaction. The observed promiscuity of some ECRs in vitro, which is not reflected in the polyketide product produced in vivo, indicates that it is likely biosynthetic constraints and intracellular availability of the enoyl-CoA thioester (that is, the ECR substrate) itself that controls production of the α-acylcarboxylated extender unit.[17] Gatekeeping and proofreading mechanisms might additionally contribute to controlling extender unit incorporation.

Finally, our findings allow to identify polyketide gene clusters that are in principle amenable to varying a polyketide target structure by precursor feeding.[18] All gene clusters that feature an ECR-2 family member with the promiscuous signature motif should be suited for such feeding strategies that aim at incorporating variable precursors into the polyketide chain to obtain modified natural product variants

Acknowledgements

This work was supported by the Gebert-Rüf-Stiftung (GRS 062-12) and by the SNF-Ambizione program (PZ00P3_136828/1). We thank R. Müller (Saarbrücken) and B. Moore (San Diego) for providing *cinF* and *salG* genes.

Keywords: biotechnology · CO_2 fixation · crotonyl-CoA carboxylase/reductase · natural product engineering · polyketides

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 13457–13461
Angew. Chem. **2015**, *127*, 13659–13663

[1] P. Banerjee, J. Erehman, B. O. Gohlke, T. Wilhelm, R. Preissner, M. Dunkel, *Nucleic Acids Res.* **2015**, *43*, D935–D939.
[2] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–335.

[3] C. T. Walsh, M. A. Fischbach, *J. Am. Chem. Soc.* **2010**, *132*, 2469 – 2493.
[4] J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380 – 416.
[5] Y. A. Chan, A. M. Podevels, B. M. Kevany, M. G. Thomas, *Nat. Prod. Rep.* **2009**, *26*, 90 – 114.
[6] M. C. Wilson, B. S. Moore, *Nat. Prod. Rep.* **2012**, *29*, 72 – 86.
[7] a) H. Liu, K. A. Reynolds, *J. Bacteriol.* **1999**, *181*, 6806 – 6813; b) Z. Xu, L. Ding, C. Hertweck, *Angew. Chem. Int. Ed.* **2011**, *50*, 4667 – 4670; *Angew. Chem.* **2011**, *123*, 4763 – 4766; c) S. F. Haydock, A. N. Appleyard, T. Mironenko, J. Lester, N. Scott, P. F. Leadlay, *Microbiology* **2005**, *151*, 3161 – 3169.
[8] A. S. Eustaquio, R. P. McGlinchey, Y. Liu, C. Hazzard, L. L. Beer, G. Florova, M. M. Alhamadsheh, A. Lechner, A. J. Kale, Y. Kobayashi, K. A. Reynolds, B. S. Moore, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12295 – 12300.
[9] H. Motamedi, A. Shafiee, *Eur. J. Biochem.* **1998**, *256*, 528 – 534.
[10] S. Rachid, L. Huo, J. Herrmann, M. Stadler, B. Kopcke, J. Bitzer, R. Müller, *ChemBioChem* **2011**, *12*, 922 – 931.
[11] a) S. Takahashi, A. Toyoda, Y. Sekiyama, H. Takagi, T. Nogawa, M. Uramoto, R. Suzuki, H. Koshino, T. Kumano, S. Panthee, T. Dairi, J. Ishikawa, H. Ikeda, Y. Sakaki, H. Osada, *Nat. Chem. Biol.* **2011**, *7*, 461 – 468; b) L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis, B. Aigle, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6258 – 6263.
[12] a) T. J. Erb, I. A. Berg, V. Brecht, M. Müller, G. Fuchs, B. E. Alber, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10631 – 10636; b) T. J. Erb, V. Brecht, G. Fuchs, M. Müller, B. E. Alber, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8871 – 8876.
[13] T. J. Erb, G. Fuchs, B. E. Alber, *Mol. Microbiol.* **2009**, *73*, 992 – 1008.
[14] a) N. Quade, L. Huo, S. Rachid, D. W. Heinz, R. Müller, *Nat. Chem. Biol.* **2012**, *8*, 117 – 124; b) R. G. Rosenthal, M. O. Ebert, P. Kiefer, D. M. Peter, J. A. Vorholt, T. J. Erb, *Nat. Chem. Biol.* **2014**, *10*, 50 – 55.
[15] R. Peyraud, P. Kiefer, P. Christen, S. Massou, J. C. Portais, J. A. Vorholt, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4846 – 4851.
[16] L. Schada von Borzyskowski, M. Remus-Emsermann, R. Weishaupt, J. A. Vorholt, T. J. Erb, *ACS Synth. Biol.* **2015**, *4*, 430 – 443.
[17] I. Koryakina, J. B. McArthur, M. D. Matthew, G. J. William, *Org. Biomol. Chem.* **2013**, *11*, 4449 – 4458.
[18] Y. Yan, J. Chen, L. Zhang, Q. Zheng, Y. Han, H. Zhang, D. Zhang, T. Awakawa, I. Abe, W. Liu, *Angew. Chem. Int. Ed.* **2013**, *52*, 12308 – 12312; *Angew. Chem.* **2013**, *125*, 12534 – 12538.

Received: June 9, 2015
Published online: September 18, 2015

