How to cite: *Angew. Chem. Int. Ed.* **2022,** *61,* e202113405
International Edition: doi.org/10.1002/anie.202113405
German Edition: doi.org/10.1002/ange.202113405

Enzyme Catalysis

Unifying Scheme for the Biosynthesis of Acyl-Branched Sugars: Extended Substrate Scope of Thiamine-Dependent Enzymes

*Jan-Patrick Steitz, Leonhard Krug, Lydia Walter, Karel Hernández, Caroline Röhr, Pere Clapés, and Michael Müller**

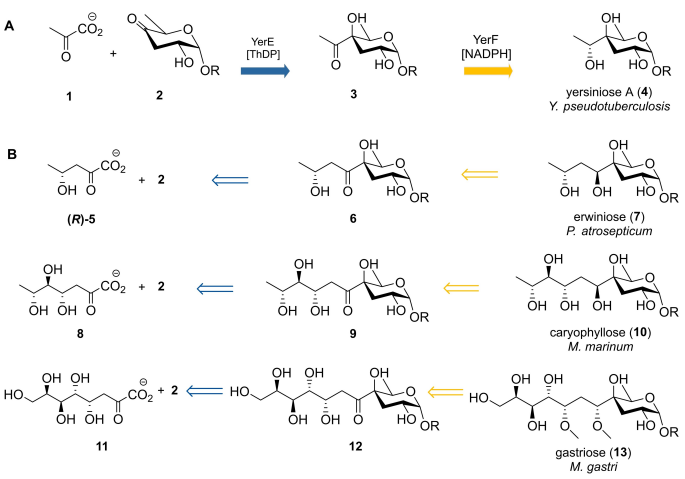
Dedicated to Maria-Regina Kula and Martina Pohl

Abstract: *Thiamine diphosphate (ThDP) dependent enzymes are useful catalysts for asymmetric C–C bond formation through benzoin-type condensation reactions that result in α-hydroxy ketones. A wide range of aldehydes and ketones can be used as acceptor substrates; however, the donor substrate range is mostly limited to achiral α-keto acids and simple aldehydes. By using a unifying retro-biosynthetic approach towards acyl-branched sugars, we identified a subclass of (myco)bacterial ThDP-dependent enzymes with a greatly extended donor substrate range, namely functionalized chiral α-keto acids with a chain length from C_4 to C_8. Highly enantioenriched acyloin products were obtained in good to high yields and several reactions were performed on a preparative scale. The newly introduced functionalized α-keto acids, accessible by known aldolase-catalyzed transformations, substantially broaden the donor substrate range of ThDP-dependent enzymes, thus enabling a more general use of these already valuable catalysts.*

Branched sugars are components of diverse antibiotics and are also found in the cell-wall polysaccharides of various organisms.[1] The most prevalent variant is C_1 branching, which is established by methylation mediated by *S*-adenosyl-L-methionine.[2,3]

The biosynthesis of acetyl-branched sugars, a presumed small group of carbohydrates which are mostly part of secondary metabolites or present in the structure of the bacterial cell wall, was investigated with regard to the biosynthesis of yersiniose A (**4**) in *Yersinia pseudotuberculosis*.[3–5] Chen et al. identified the ThDP-dependent enzyme YerE as catalyzing the carboligation of pyruvate with CDP-4-keto-3,6-dideoxy-D-glucose (**2**; Scheme 1A).[4] Acetyl branches established in this way are the starting point for further modifications implemented by reductases, transaminases, or oxygenases in the biosynthesis of several secondary metabolites. However, high specificity was observed for YerE with respect to the donor substrates tested, which limits the chain length of the branches formed.[6]

Nevertheless, the variety of modified acyl-branched (other than acetyl) carbohydrates found in the outer membrane of (myco)bacteria is remarkable: significantly longer and multifunctionalized branches have been described as being present there.[7–10] Hence, the biosynthesis of longer ($>C_2$)-branched sugars, present in such lipopoly-



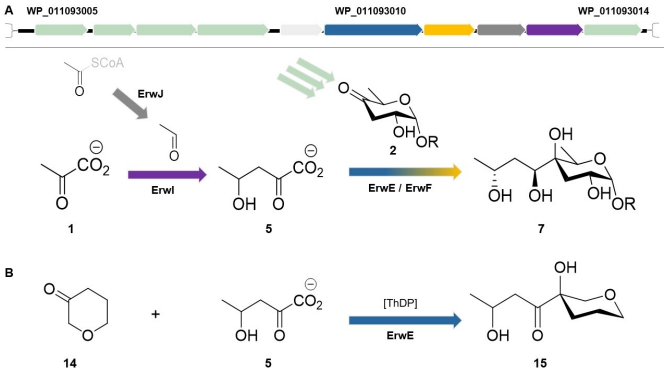
Scheme 1. Unifying scheme for the biosynthesis of acyl-branched sugars. A) Biosynthesis of yersiniose A in *Y. pseudotuberculosis*; the YerE-catalyzed condensation of **1** and **2** is the key step in the formation of the C_2 branch, followed by YerF-catalyzed reduction.[4] B) Retro-biosynthesis of ThDP-dependent reactions (blue) and oxidoreductions (yellow) in the formation of the longer-branched sugars erwiniose (**7**) from *P. atrosepticum*, caryophyllose (**10**) from *M. marinum*, and gastriose (**13**) from *M. gastri*. The (poly)hydroxylated α-keto acids (*R*)-**5**, **8**, and **11** are proposed to be the physiological substrates for condensation reactions with deoxy keto sugar **2**.

[*] Dr. J.-P. Steitz, L. Krug, Dr. L. Walter, Prof. Dr. M. Müller
Institut für Pharmazeutische Wissenschaften
Albert-Ludwigs-Universität Freiburg
Albertstrasse 25, 79104 Freiburg (Germany)
E-mail: michael.mueller@pharmazie.uni-freiburg.de
Dr. K. Hernández, Prof. Dr. P. Clapés
Chemical Biology Department
Institute for Advanced Chemistry of Catalonia (IQAC-CSIC)
Jordi Girona 18–26, 08034 Barcelona (Spain)
Prof. Dr. C. Röhr
Institute for Inorganic Chemistry
Albert-Ludwigs-Universität Freiburg
Albertstrasse 21, 79104 Freiburg (Germany)

saccharides (LPS), seemed to be a promising starting point to search for new ThDP-dependent enzymes putatively catalyzing reactions with unidentified donor substrates. For the biosynthesis of the branched sugars erwiniose (C_4, **7**) from *Pectobacterium atrosepticum*, caryophyllose (C_6, **10**) from *Mycobacterium marinum*, and a C_7-branched sugar from the LPS of *Mycobacterium gastri*, herein named "gastriose" (**13**), we propose ThDP-dependent C–C bond formation reactions of acceptor substrate **2** with functionalized donor substrates **5**, **8**, and **11** (Scheme 1B).[7–9] Accordingly, similar to the biosynthesis of yersiniose A (**4**, Scheme 1A), a ThDP-dependent (cf. YerE) branching, followed by an NAD(P)H-dependent (cf. YerF) reduction of the side chain, would take place.

Starting from the known genes *yerE* and *yerF*, we identified the two genes *erwe* and *erwf* in the genome of *P. atrosepticum* SCRI1043 putatively encoding a ThDP-dependent enzyme (ErwE, WP_011093010) and an NAD(P)H-dependent reductase (ErwF, WP_011093011), respectively (Scheme 2A). Both genes are located in direct vicinity to genes predicted to code for a pyruvate aldolase (ErwI, WP_011093013) and an acetaldehyde dehydrogenase (ErwJ, WP_011093012). These two genes show pairwise amino acid sequence similarity to BphI/BphJ (43 %, 41 %), two complex-forming enzymes from *Paraburkholderia xenovorans* which catalyze the cleavage of 4-hydroxy-2-oxopentanoate (**5**).[11,12] This reinforced our hypothesis that (*R*)-**5**, the aldol product of pyruvate (**1**) and acetaldehyde, might be the physiological donor substrate of the ThDP-dependent enzyme ErwE, which would catalyze an addition reaction of (*R*)-**5** with deoxy keto sugar **2** to form **6**, the direct precursor of erwiniose (**7**). Furthermore, the putative biosynthetic gene cluster (BGC) harbors the necessary genes coding for enzymes most probably involved in the biosynthesis of **2** (Supporting Information).

To verify the proposed biosynthesis, we prepared *rac*-**5** by use of HpaI, an already established aldolase from *E. coli* K-12. Although HpaI shows very low global protein sequence identity to BphI/ErwI (13 %, 18 %), it has been described as catalyzing the biosynthesis of **5**.[13] Subsequently, we applied a biomimetic approach with heterologously produced and purified ErwE [using *E. coli* BL21-Gold(DE3) cells] to investigate if ErwE can catalyze the proposed carboligation reaction of a functionalized α-keto acid with a ketone. Dihydro-2*H*-pyran-3(4*H*)-one (**14**), a simplified analogue of **2**, was tested as acceptor substrate with the putative donor substrate **5** (Scheme 2B). The aldol product **15** was isolated and characterized by NMR spectroscopy, thus confirming our hypothesis (Supporting Information).



Scheme 2. A) Putative BGC from *P. atrosepticum* SCRI1043; genes proposed for the biosynthesis of deoxy keto sugar **2** (pale green), ThDP-dependent enzyme (ErwE, blue), oxidoreductase (ErwF, yellow), aldolase (ErwI, violet), and dehydrogenase (ErwJ, dark gray). B) ErwE-catalyzed in vitro synthesis of α-hydroxy ketone **15**.

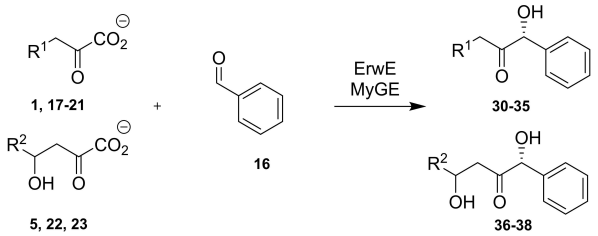
As the aldolase reaction could, in principle, be performed after a ThDP-dependent branching with pyruvate, which would refute our hypothesis, it was tested whether ErwE accepts or discriminates pyruvate. As expected, pyruvate (**1**) was not accepted by ErwE in the reaction with ketone **14**. Hence, we deduce the following reaction sequence for the biosynthesis of erwiniose (**7**), which is consistent with the identified genes in the putative BGC (Scheme 2A): the biosynthesis starts with the ErwI (violet) catalyzed aldolase reaction of **1** with acetaldehyde, which is provided by acetyl-CoA dehydrogenase ErwJ (dark gray). The formed γ-hydroxy keto acid **5** and deoxy keto sugar **2** are converted by ThDP-dependent ErwE (blue), followed by NAD(P)H-dependent reduction to **7**, catalyzed by ErwF (yellow). Finally, with a glycosyl transferase reaction, most probably encoded by WP_011093016, **7** is integrated into the LPS of *P. atrosepticum*.

We further investigated the donor and acceptor range of ErwE. To this end, the aldol synthesis of additional γ-hydroxylated keto acids was performed: **22** using MBP-YfaU and **23** using HpaI (Supporting Information).[11,14] Regarding the acceptor substrate range, we used benzaldehyde (**16**) or *p*-bromobenzaldehyde (**24**); both were converted with *rac*-**5**. Moreover, γ-functionalized 4-hydroxy-2-oxobutanoate (**22**) and -hexanoate (**23**) were converted with **16** as acceptor, whereas pyruvate (**1**) and 2-oxobutyrate (**17**), as well as 2-oxopentanoate (**18**), 4-methyl-2-oxopentanoate (**19**), and 2-oxooctanoate (**21**), were not accepted by ErwE (Table 1).

Although aldolase reactions with α-keto acids as donor are a straightforward way towards diverse functionalized α-keto acids, an alternative biocatalytic access is by transamination of the respective α-amino acids. α-Keto-γ-(methylthio)butanoate (**20**) and α-keto acid **21**, accessible by transamination of methionine and α-aminooctanoate, respectively, were tested. We identified **20** as being accepted by ErwE, but resulted in a low conversion of 10 % (Table 1).

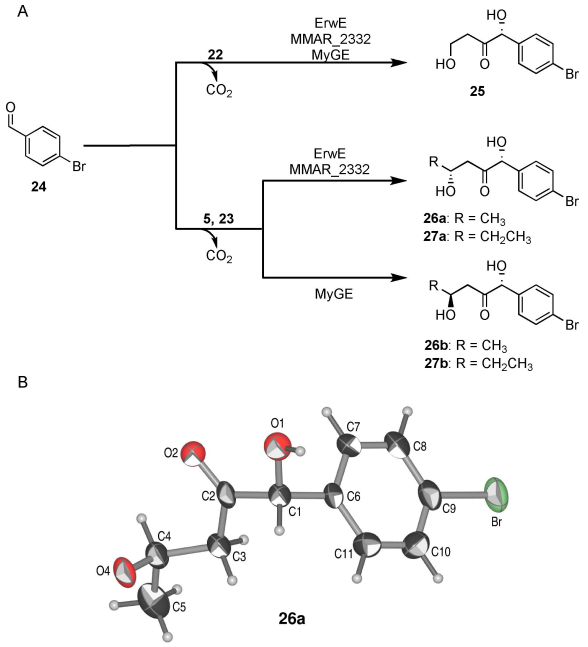
Moreover, we determined the enantioselectivity of ErwE with achiral but hydroxylated donor substrate **22** and acceptor *p*-bromobenzaldehyde (**24**). Product **25** was isolated in 11 % yield with an enantiomeric excess (*ee*) >95 %; the absolute configuration was determined as (*R*) by X-ray crystallography as well as CD spectroscopy (Scheme 3). ^{1}H NMR analysis of the synthesized products of biocatalytic reactions with the two racemic substrates **5** and **23** and benzaldehyde (**16**) as acceptor revealed the strongly favored formation of one diastereomer (Table 1). Hence, we proposed a stereoselective formation of the new C–C bond as

Table 1: Donor substrate range (**1**, **5**, **17**–**23**) with benzaldehyde (**16**) as the acceptor substrate in reactions catalyzed by ErwE and MyGE.[a] The conversion [%], absolute configuration, and *ee* value [%] of the products are given.



Donor substrate		ErwE	MyGE	Product
1	$R^1 = H$	n.a.[b]	40 (*R*) (*ee* > 98%)	**30**
17	$R^1 = CH_3$	n.a.	35 (*R*) (*ee* > 90%)	**31**
18	$R^1 = CH_2CH_3$	n.a.	85 (*R*)	**32**
19	$R^1 = CH(CH_3)_2$	n.a.	22 (*R*)	**33**
20	$R^1 = CH_2SCH_3$	10 (*R*)	77 (*R*)	**34**
21	$R^1 = (CH_2)_4CH_3$	n.a.	24 (*R*)	**35**
22	$R^2 = H$	33 (*R*)	18 (*R*)	**36**
5	$R^2 = CH_3$	49 (1*R*,4*R*)	1	**37**
23	$R^2 = CH_2CH_3$	76 (1*R*,4*R*)	2	**38**

[a] Conditions: donor substrate [20 mM (2-oxo acids **20**, **5**) or 50 mM (sodium 2-oxoalkanoates)], 7% DMSO (v/v), ErwE (Kp_i buffer 50 mM, 200 mM NaCl, 1 mM $MgCl_2$, 50 μM ThDP, pH 7.4), or MyGE (HEPES buffer 50 mM, 100 mM NaCl, 2.5 mM $MgCl_2$, 50 μM ThDP, pH 7.5), protein (1.5 mg mL^{-1}), 30 °C, 300 rpm, 20 h. Conversion was determined by ^{1}H NMR spectroscopy; the *ee* value was determined by chiral-phase HPLC. [b] Not accepted.



Scheme 3. A) Enzymatic (ErwE, MMAR_2332, MyGE) synthesis of chiral **25** and diastereomeric **26a**/**b** and **27a**/**b**. The configurations of **25** and **26a**, synthesized with ErwE, were determined by X-ray crystallography. The configurations of all products were assigned by ^{1}H NMR spectroscopy and CD experiments. B) Thermal ellipsoid plot (90% level) of **26a**,showing the (*R*,*R*)-configuration. Bond distances are given in the Supporting Information.

well as enantiospecificity for one enantiomer of each of the two donor substrates. This was proven by the ErwE-catalyzed reaction of **5** and **24**, which resulted in product **26a** (36% yield after 17 h reaction time). The (1*R*,4*R*)-configuration of the major diastereomer **26a** was determined by means of X-ray crystallography; the diastereomeric ratio was 95:5 as determined by ^{1}H NMR analysis (Scheme 3). Both the decrease in the *ee* value of **25** and the formation of the minor diastereomer of **26** might result from a non-enzymatic racemization of the benzylic stereocenter, as is well-known for 1-hydroxy-1-arylpropan-2-ones.[15]

According to our proposed unifying scheme, the biosynthetic strategy should also be applicable to the other aforementioned longer-branched sugars such as caryophyllose (**10**) or gastriose (**13**). Alibaud et al. knocked out several genes of the putative BGC of caryophyllose, while studying the biosynthesis of LOS from *M. marinum*. The Δ*mmar_2331*, Δ*mmar_2332*, and Δ*mmar_2333* mutants failed to integrate a caryophyllose moiety into LOSII*, usually forming LOSIII, which is a sequence of eight sugar units containing two units of caryophyllose.[16] LOSIII is part of the cell wall of *M. marinum* and was identified to inhibit tumor necrosis factor-α secretion in macrophages.[17] *mmar_2332* and *mmar_2333* were proposed to code for a ThDP-dependent carboligase and a glycosyltransferase, respectively.[18] For *mmar_2331*, homology with enzymes from the aldolase/citrate lyase family was found, suggesting an implication in the biosynthesis of the branch.[16] With the known sequences of these genes, we identified the entire BGC as containing additional genes coding for a putative reductase and several enzymes from the biosynthesis of **2** (Supporting Information).

We cloned *mmar_2332* and heterologously [*E. coli* BL21-Gold(DE3)] produced the protein for testing. The

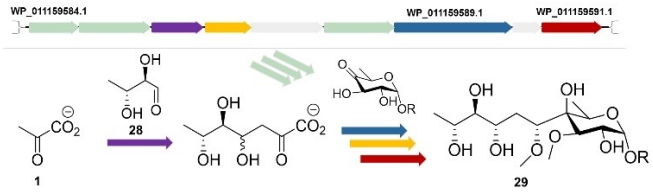
enzymatic activity of purified MMAR_2332 was shown by formation of (*R*)-benzoin from benzoyl formate and benzaldehyde (**16**), a standard activity test for ThDP carboligases; the product was identified by GC-MS and the absolute configuration was determined by CD spectroscopy. As the purification of MMAR_2332 only gave small amounts of enzyme and the stability of the purified enzyme in the tested buffer solutions was poor, subsequent conversions were performed with *E. coli* whole cells, which express *mmar_2332*. (*R*)-**25** was formed in a reaction of **22** with **24** (^{1}H NMR and CD spectroscopy). In the condensation reaction of **5** and **24** catalyzed by ErwE or MMAR_2332, the same major diastereomer **26a** was formed (^{1}H NMR analysis), suggesting a (1*R*,4*R*)-configuration (Scheme 3). The identified configuration is homochiral with the corresponding stereocenters in the natural products erwiniose (**7**) and caryophyllose (**10**) from *P. atrosepticum* and *M. marinum*, respectively (Scheme 1).

To challenge our hypothesis of a unifying biosynthetic scheme, we applied a third enzyme, MyGE, from *M. gastri.* The gene coding for ThDP-dependent MyGE was identified in a BGC in the vicinity of a gene encoding aldolase MygI and NAD(P)H-dependent oxidoreductase MygF. The side chain of the C_7-branched sugar gastriose (**13**) from *M. gastri* is highly hydroxylated and the absolute configuration has not been assigned beyond doubt; nevertheless, the relative configuration of the side chain was identified by Longépé et al.[19] According to our hypothesis, MyGE might accept hydroxylated α-keto acids with longer chain length.

We performed analytical scale assays to compare the substrate range of ErwE and MyGE. Donor substrates **1**, **5**, and **17**–**23** were converted with benzaldehyde (**16**) as the acceptor substrate (Table 1). In addition, the achiral γ-hydroxylated donor substrate **22** and the chiral substrates **5** and **23** were converted with the acceptor *p*-bromobenzaldehyde (**24**) to compare the stereoselectivity of all three enzymes (Scheme 3).

Indeed, the experiments conducted with heterologously produced and purified MyGE indicated a broad substrate range. The three γ-hydroxylated α-keto acids (**5**, **22**, **23**) accepted by ErwE were also converted by MyGE. The *ee* value for (*R*)-**25** (>97 %) was slightly higher for MyGE, while both enzymes showed the same enantioselectivity. Interestingly, in the case of the products **26** and **27**, we identified the formation of different major diastereomers by ErwE and MyGE. Determination of the absolute configuration of those two products by ^{1}H NMR and CD spectroscopy indicated that the enantioselectivity is inverted; hence, the (1*R*,4*S*)-diastereomer is formed as the major product by MyGE. These results also give a hint of the absolute configuration of the side chain of gastriose (**13**). Moreover, linear α-keto acids with a chain length from C_3 to C_8 (**1**, **17**, **18**, **21**), as well as the branched donor substrate **19** and thioether-functionalized substrate **20**, were accepted by MyGE.

In summary, the three investigated ThDP-dependent enzymes, ErwE, MMAR_2332, and MyGE, show an extended donor substrate range encompassing longer-chain and functionalized 2-oxo carboxylic acids; remarkably, their



Scheme 4. Putative BGC from *R. palustris* strain CGA009; genes proposed for the biosynthesis of deoxy keto sugar (pale green), ThDP-dependent enzyme (blue), NAD(P)H-dependent oxidoreductase (yellow), aldolase (violet), and SAM-dependent methyltransferase (red).

respective promiscuity shows substantial differences. The broad substrate range, including ketones as acceptor substrates, as well as the distinct stereoselectivity and specificity, make these enzymes promising new catalysts for asymmetric C–C bond formations.

Recently, the new C_6-branched sugar **29** was described in the O-antigen of *Rhodopseudomonas palustris* CGA009 (Scheme 4).[10] The side chain of this sugar is equivalent to caryophyllose, but possesses an inverted configuration at the C1 carbinol of the side chain.

In accordance with our hypothesis, we tentatively identified a putative BGC in the whole genome sequence of the described strain, containing all the expected genes, including an aldolase, a ThDP-dependent enzyme, an NAD(P)H-dependent oxidoreductase, and a gene putatively coding for a SAM-dependent methyltransferase (Scheme 4). The aldolases from the BGCs in *M. marinum* and *R. palustris*, which putatively catalyze a reaction with the same combination of substrates (**1** and **28**), show a pairwise protein identity of 58 %.

The characterization of the enzymatic activity of three different ThDP-dependent enzymes, namely ErwE, MMAR_2332, and MyGE, and the identification of a putative BGC for **29** in this work confirm the hypothesis of a unifying biosynthetic scheme for longer-branched sugars (Scheme 1). As the identification and structural elucidation of uncommon carbohydrates in the LPS of (myco)bacteria is extremely difficult, our approach enables the prediction of acyl-branched sugars by sequence analysis.[5] Moreover, the structure, substitution, and stereochemistry of the side chains can be deduced by retro-biosynthetic analysis and comparison to the here-established structures.

Acknowledgements

We thank the Ministerio de Economía y Competitividad (MINECO), the Fondo Europeo de Desarrollo Regional (FEDER) (grant RTI2018-094637-B-I00), and Programación Conjunta Internacional (PCI2018-092937), through the initiative ERA CoBioTech (Tralaminol), and the DFG (FOR 1296) and the European Union's Horizon 2020 Research and Innovation Programm (Grant Agreement no. 635595) for financial support. Many thanks to Dr. Stephen Seah, University of Gulph, for providing the plasmid of *hpaI*. We would also like to thank Dr. Philippe Bisel and

Sascha Ferlaino, University of Freiburg, for NMR measurements and constructive suggestions for the manuscript, as well as Dr. Kay Greenfield for her careful proofreading. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Aasymmetric synthesis · Biocatalysis · Biogenesis · Carboligation · Lipopolysaccharides

[1] H. Grisebach, R. Schmid, *Angew. Chem. Int. Ed. Engl.* **1972**, *11*, 159–173; *Angew. Chem.* **1972**, *84*, 192–206.
[2] H. Achenbach, H. Grisebach, *Z. Naturforsch. B* **1964**, *19*, 561–568.
[3] H. Grisebach, *Helv. Chim. Acta* **1968**, *51*, 928–939.
[4] H. Chen, Z. Guo, H. Liu, *J. Am. Chem. Soc.* **1998**, *120*, 11796–11797.
[5] F. Di Lorenzo, K. A. Duda, R. Lanzetta, A. Silipo, C. de Castro, A. Molinaro, *Chem. Rev.* **2021**, *121*, https://doi.org/10.1021/acs.chemrev.0c01321.
[6] P. Lehwald, M. Richter, C. Röhr, H. Liu, M. Müller, *Angew. Chem. Int. Ed.* **2010**, *49*, 2389–2392; *Angew. Chem.* **2010**, *122*, 2439–2442.
[7] S. N. Senchenkova, A. S. Shashkov, Y. A. Knirel, M. Ahmed, A. Mavridis, K. Rudolph, *Russ. Chem. Bull.* **2005**, *54*, 1276–1281.
[8] M. Adinolfi, M. M. Corsaro, C. de Castro, A. Evidente, R. Lanzetta, L. Mangoni, M. Parrilli, *Carbohydr. Res.* **1995**, *274*, 223–232.
[9] M. Gilleron, J. Vercauteren, G. Puzo, *J. Biol. Chem.* **1993**, *268*, 3168–3179.
[10] R. Marchetti, E. Bedini, D. Gully, R. Lanzetta, E. Giraud, A. Molinaro, A. Silipo, *Org. Lett.* **2018**, *20*, 3656–3660.
[11] P. Baker, D. Pan, J. Carere, A. Rossi, W. Wang, S. Y. K. Seah, *Biochemistry* **2009**, *48*, 6551–6558.
[12] J. Carere, P. Baker, S. Y. K. Seah, *Biochemistry* **2011**, *50*, 8407–8416.
[13] a) W. Wang, P. Baker, S. Y. K. Seah, *Biochemistry* **2010**, *49*, 3774–3782; b) W. Wang, S. Y. K. Seah, *Biochemistry* **2005**, *44*, 9447–9455.
[14] K. Hernandez, J. Bujons, J. Joglar, S. J. Charnock, P. Domínguez de María, W. D. Fessner, P. Clapés, *ACS Catal.* **2017**, *7*, 1707–1711.
[15] S. Hampel, J.-P. Steitz, A. Baierl, P. Lehwald, L. Wiesli, M. Richter, A. Fries, M. Pohl, G. Schneider, D. Dobritzsch, et al., *ChemBioChem* **2018**, *19*, 2283–2292.
[16] L. Alibaud, J. Pawelczyk, L. Gannoun-Zaki, V. K. Singh, Y. Rombouts, M. Drancourt, J. Dziadek, Y. Guérardel, L. Kremer, *J. Biol. Chem.* **2014**, *289*, 215–228.
[17] Y. Rombouts, A. Burguière, E. Maes, B. Coddeville, E. Elass, Y. Guérardel, L. Kremer, *J. Biol. Chem.* **2009**, *284*, 20975–20988.
[18] D. Sarkar, M. Sidhu, A. Singh, J. Chen, D. A. Lammas, A. M. van der Sar, G. S. Besra, A. Bhatt, *J. Bacteriol.* **2011**, *193*, 2336–2340.
[19] J. Longépé, J. Prandi, J.-M. Beau, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 72–75; *Angew. Chem.* **1997**, *109*, 155–157.

Manuscript received: October 3, 2021
Version of record online: January 28, 2022