Engineering Regioselectivity of a P450 Monooxygenase Enables the Synthesis of Ursodeoxycholic Acid via 7β-Hydroxylation of Lithocholic Acid

Sascha Grobe, Christoffel P. S. Badenhorst, Thomas Bayer, Emil Hamnevik, Shuke Wu, Christoph W. Grathwol, Andreas Link, Sven Koban, Henrike Brundiek, Beatrice Großjohann, and Uwe T. Bornscheuer*

Abstract: *We engineered the cytochrome P450 monooxygenase CYP107D1 (OleP) from Streptomyces antibioticus for the stereo- and regioselective 7β-hydroxylation of lithocholic acid (LCA) to yield ursodeoxycholic acid (UDCA). OleP was previously shown to hydroxylate testosterone at the 7β-position but LCA is exclusively hydroxylated at the 6β-position, forming murideoxycholic acid (MDCA). Structural and 3DM analysis, and molecular docking were used to identify amino acid residues F84, S240, and V291 as specificity-determining residues. Alanine scanning identified S240A as a UDCA-producing variant. A synthetic "small but smart" library based on these positions was screened using a colorimetric assay for UDCA. We identified a nearly perfectly regio- and stereoselective triple mutant (F84Q/S240A/V291G) that produces 10-fold higher levels of UDCA than the S240A variant. This biocatalyst opens up new possibilities for the environmentally friendly synthesis of UDCA from the biological waste product LCA.*

Ursodeoxycholic acid (UDCA) is a valuable bile acid frequently prescribed for the treatment of cholecystitis as it can solubilize cholesterol gallstones with fewer side effects than chenodeoxycholic acid (CDCA).[1] UDCA also has anti-

[*] M. Sc. S. Grobe, Dr. C. P. S. Badenhorst, Dr. T. Bayer,
Dr. E. Hamnevik, Dr. S. Wu, Prof. Dr. U. T. Bornscheuer
Department of Biotechnology and Enzyme Catalysis, Institute of
Biochemistry, University of Greifswald
Felix Hausdorff-Str. 4, 17487 Greifswald (Germany)
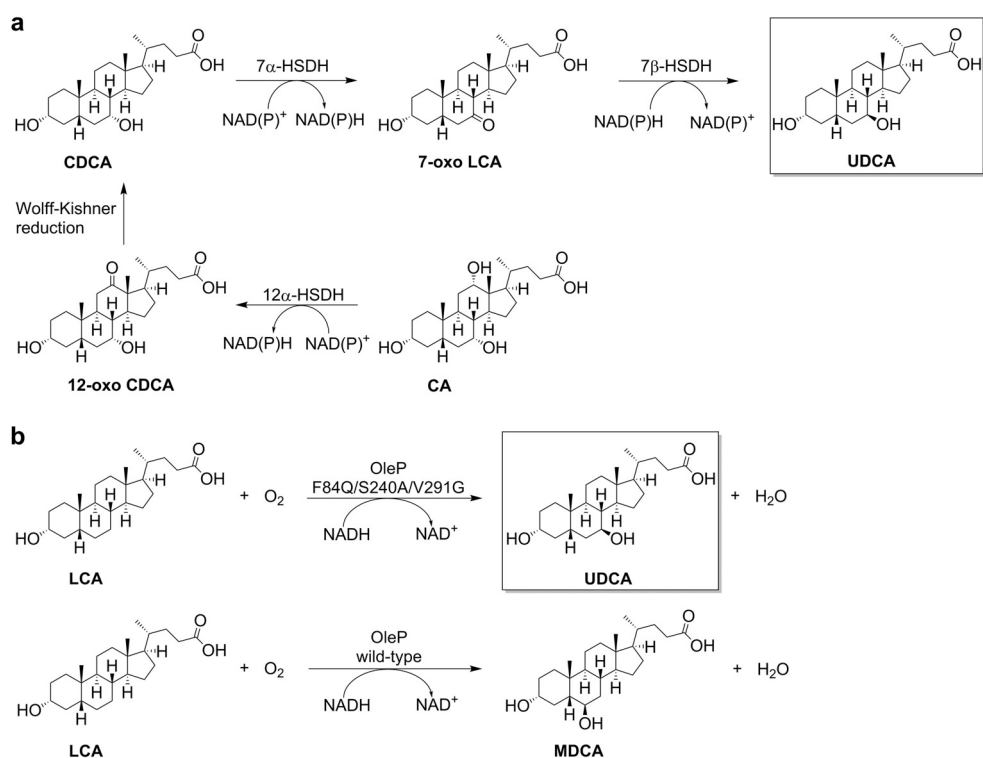E-mail: uwe.bornscheuer@uni-greifswald.de

Dr. C. W. Grathwol, Prof. Dr. A. Link
Institute of Pharmacy, University of Greifswald
Friedrich-Ludwig-Jahn-Str. 17, 17487 Greifswald (Germany)

M. Sc. S. Koban, Dr. H. Brundiek
Enzymicals AG
Walther-Rathenau-Str. 49, 17487 Greifswald (Germany)

Prof. Dr. B. Großjohann
HERBRAND PharmaChemicals GmbH, Betriebsstätte Anklam
An der Redoute 1, 17390 Murchin (Germany)

inflammatory properties[2] and is applied in the therapy of cystic fibrosis[3] and liver diseases like primary biliary cholangitis.[4] The major natural source of UDCA is bear bile,[5] a popular traditional medicine obtained by biliary catheterization of farmed bears. Alternatively, semi-synthetic UDCA can be produced from cholic acid (CA)[6] or CDCA.[7,8] The synthesis route starting from CA forms CDCA within 5 steps, including a Wolff–Kishner reduction, and an epimerization at C7 to produce UDCA (Scheme 1a, Scheme S1).[9] The yields of this pathway do not exceed 30%. To overcome these limitations, a shorter synthesis route based on the biocatalytic epimerization of CDCA to UDCA (Scheme 1a) has been developed.[7,10]

LCA is an abundant and inexpensive waste product of meat production[11] as this bile acid is found in farmed animals like sheep,[12] cattle,[12] and pigs.[13] Currently, no biotechnological process[14] utilizing LCA originating from these sources is known, making it a desirable starting material for the synthesis of UDCA. A few microbial organisms have been reported to form UDCA from LCA.[15] For example, the fungus *Fusarium equiseti* converts LCA to a number of products, including UDCA at 35% yield.[16]

However, there is currently no enzyme known to selectively hydroxylate LCA at the 7β-position to form UDCA and its synthesis pathway in microbial organisms, starting from LCA, remains enigmatic. An enzyme for direct 7β-hydroxylation would be a valuable tool for direct conversion of LCA to UDCA without involving the complex metabolism of fungi that invariably[16] produce multiple undesired side products,[15] complicating downstream processing.

A major challenge in the enzymatic conversion of LCA to UDCA is the hydrophobicity and thus, extremely low water solubility of LCA compared to the more widely used CA or CDCA.[17]

P450s are heme-containing enzymes capable of stereo- and regioselective hydroxylation reactions of a wide variety of substrates, using molecular oxygen as oxidant.[18]

This enzyme class has been intensively investigated due to its potential for late-stage hydroxylation of industrially relevant precursor compounds.[19] Protein engineering is often employed to produce P450 monooxygenases that meet the requirements of excellent regio- and stereoselectivity for the desired applications.[20]

The monooxygenase CYP107D1 (OleP), originally described for an epoxidation[21,22] in the oleandomycin biosynthesis pathway,[23] was found to accept 12-membered macro-

a

b

Scheme 1. a) Chemoenzymatic synthesis routes towards UDCA starting from CA or CDCA. b) The desired 7β-hydroxylation of LCA towards UDCA by an engineered OleP variant and wild-type OleP mediated formation of MDCA, starting from LCA.

lactone substrates[24] as well. OleP can also hydroxylate testosterone at the positions 6β, 7β, 12β, and 15β.[25] However, bile acids like LCA and deoxycholic acid (DCA) are hydroxylated exclusively at the 6β-position, forming MDCA (Scheme 1b) and 3α-, 6β-, 12α-trihydroxy-5β-cholan-24-oic acid, respectively.[26]

To open up a novel reaction pathway towards UDCA, as no enzyme is known for selective 7β-hydroxylation, and to avoid using whole cell fungi, wherein the action of multiple P450s frequently results in side-product formation, we decided to engineer OleP and create an *Escherichia coli* (*E. coli*)-based whole-cell system for the regio- and stereoselective hydroxylation of LCA at the 7β-position to form UDCA (Scheme 1b).

Initially, we determined which amino acid residues could influence the positioning of LCA in the OleP active site. Using crystal structures[21,22] of OleP with bound inhibitors and a natural substrate analog, we were able to dock LCA into the OleP active site in multiple conformations (Figure 1). In this evaluation process, a docked structure was found matching three desired criteria. First, LCA had to be positioned at most 5 Å away from the heme iron. Second, the 6β-hydrogen had to be oriented towards the heme, mimicking the binding of LCA for MDCA formation. This resulted in the elimination of perpendicular orientations. Finally, since the steroid nucleus can flip in the horizontal orientation, we used the most frequently occurring conformations in our docking experiment (Figure 1a). Based on this docking pose, residues in a zone of 5–14 Å around the heme

were chosen. Important P450 specificity-determining residues known from literature[21,24,27] were also selected. Finally, residues interacting with the substrate analog[24] were also selected. In total, 24 active site residues were identified for further investigation.

Alanine scanning was used to generate space in the active site to allow LCA to shift from the 6β- towards 7β-hydroxylation orientation and to investigate the contributions of residues to hydroxylation activity.

Of the 24 residues selected (Figure 1b), four were already alanine or glycine. The remaining 20 were exchanged to alanine by site-directed mutagenesis. The variants were co-expressed with the redox partner proteins Putidaredoxin (PdX) and Putidaredoxin reductase (PdR), employing a two-plasmid system in *E. coli* C43-

(DE3). Whole-cell biotransformations were performed with 5 mM LCA. HPLC analysis revealed five UDCA-producing variants (F84A, V93A, L94A, S240A, and V291A). L179A and S295A produced no product and the remaining variants produced exclusively MDCA, like wild-type OleP.

The variants F84A, S240A, and V291A were most promising as they produced mostly MDCA and UDCA, compared to V93A and L94A, which produced several unidentified products.

The residue F84 of the BC loop is associated with substrate recognition.[22] S240 is part of the I-helix and coordinates parts of the hydrogen-bond network, which connects OleP with the substrate. V291 is within a β-hairpin (β3) and part of a hydrophobic bulge facilitating coordination of the substrate by van der Waals interactions.[22] This suggested that manipulation of the water network and taking advantage of the hydrophobic interactions, which have significant effects in the positioning of the substrate, could enhance the 7β-hydroxylation of LCA to form UDCA.

A 3DM[28] database of P450 monooxygenases was used to evaluate the most frequently occurring amino acids at each of the residues altered in the initial UDCA-forming variants (F84A, V93A, L94A, S240A, and V291A). Each residue was mutated to the four most frequently occurring amino acids. The aim was to diversify these positions without the introduction of amino acids which are evolutionarily unfavored and likely to result in inactive variants. With these variants created, UDCA formation and the occurrence of side products were analyzed by HPLC. Substitutions of F84, S240,
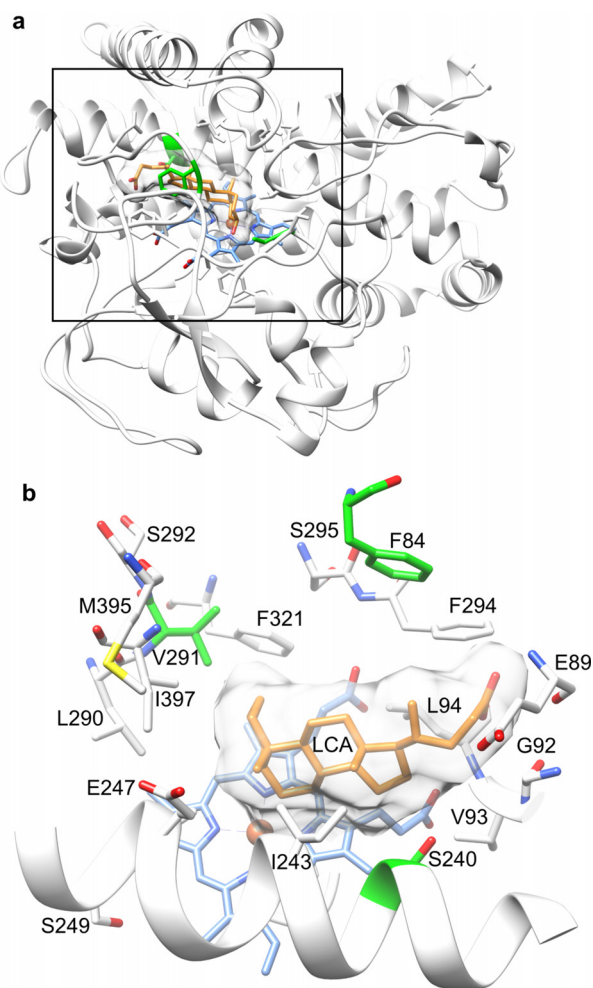
Figure 1. a) Structure of OleP (PDB ID: 4XE3) with docked LCA. b) Active site residues surrounding the docked LCA. The heme cofactor is colored blue and the iron atom in red, LCA is orange with its surface shown in white. Residues selected for alanine scanning are colored white and the selectivity influencing residues F84, S240, and V291 are green.

and V291 resulted in increased UDCA production. We then designed a "small but smart" 3DM library[29] focused on F84, S240, and V291 (Table S2). The library, containing 4,480 unique variants, was ordered from Twist Bioscience, thereby eliminating bias, and cloned into the pET-28a vector.[30] *E. coli* C43(DE3) was co-transformed with the library and the redox partner system pACYCDuet-1-*pdR/pdX*. Sequencing a random sample revealed only the desired codons. Cultivation and whole-cell biotransformations took place in 24 deep-well plates, which allowed us to standardize the process. Detection of the product initially depended on HPLC measurements.

Screening of libraries towards desired regio- and stereo-selective hydroxylation of steroids and bile acids has—with some exceptions[31]—mostly been done by chromatographic methods. A colorimetric UDCA assay, employing an NADP+-dependent 7β-hydroxysteroid dehydrogenase,[32] allowed us to screen 600 variants per round (Scheme S2).

The S240A variant, which was the most active variant resulting from the initial alanine scanning, was used as a positive control during screening (Figure S5). Screening a set of 3,400 clones actually led to the identification of S240A as a hit. In total, 34 clones, with up to 18-fold higher absorbance than that of the S240A variant, were identified (Figure S5). Of these 34 hits, 32 were confirmed to produce UDCA by HPLC in comparison to a commercial standard.

We observed a clear correlation between the colorimetric UDCA assay signals and the number of side products formed. An increased signal in the 7β-HSDH assay correlated with fewer side products as confirmed by HPLC (Figure 2). The variants showing the highest colorimetric signal mostly produced MDCA and UDCA in different ratios ranging from 25.9% UDCA formation (F84Q/S240A), through 31.9% UDCA formation (F84C/S240A/V291A), up to 72.8% UDCA formation (F84Q/S240A/V291D) (Table S3). The specificity of F84Q/S240A/V291D for UDCA formation was remarkable, considering that wild-type OleP exclusively produces MDCA and the S240A variant forms only 8.2% UDCA. None of the UDCA-producing mutants formed the 7α-hydroxylation product (CDCA), demonstrating perfect stereoselectivity.

We continued by sequencing the HPLC-confirmed UDCA-producing variants and found very clear amino acid preferences at each of the randomized positions (Figure S6). In UDCA-producing variants, F84 was most frequently mutated to Q (44%) or T (19%). For S240, a clear preference for A (87%) was observed. For V291, the most common variants were G (34%), A (14%) and E (14%). By the introduction of a polar and more flexible amino acid like Q or T at position F84, the flexibility of the BC-loop is enhanced. This loop is involved in substrate recognition and stabilization
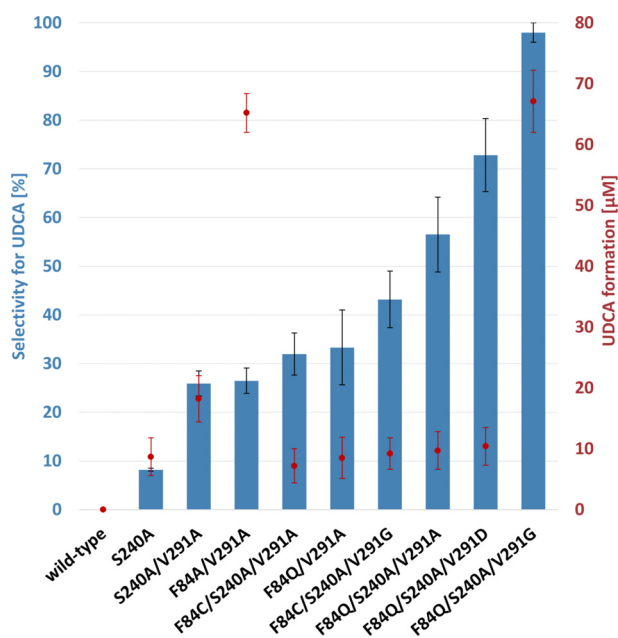

Figure 2. UDCA-producing variants with increased selectivity. OleP variants show increasing selectivity towards UDCA (blue) and increased UDCA formation (red). UDCA was quantified by HPLC of three replicate reactions.

of the bile acid. Since S240 is both a key residue establishing the water network of the active site and associated with the water-mediated coordination of substrates, we hypothesize that with the introduction of a nonpolar and small residue, the water network in the active site is remodeled, favoring the 7β-hydroxylation. This concept is in line with recent findings reported for a testosterone hydroxylation using the P450 monooxygenase BM3.[34] As V291 is part of a hydrophobic cleft, its mutation to glycine could be beneficial for the 7β-hydroxylation by the generation of space to position the more dented bile acid in the active site better compared to flat steroids such as testosterone.

To get further insights into the contribution of each position and to find the most selective variant, we created a set of 37 single, double, and triple mutants at residues F84, S240A, and V291. This selection was based on UDCA:MDCA ratios of the individual variants and sequencing of UDCA producing variants (Figure S6 and Table S3). All newly created variants were analyzed by the colorimetric assay as well as HPLC. Several single, double, and triple mutants (Figure 2) were capable of producing UDCA (Table S3) although the selectivity towards UDCA formation varied significantly. The most striking variant, F84Q/S240A/V291G, produced 67 μM of UDCA and exhibited outstanding selectivity. Only trace amounts of MDCA were formed as the sole side-product. Interestingly, this mutant combines the most frequently occurring residues in UDCA-producing variants (Figure S6). By scaling the reaction to 500 mL using OleP F84Q/S240A/V291G, we were able to isolate the product and confirmed it as UDCA by NMR-spectroscopy (Figures S11 and S12). The side-product formed only in traces was confirmed as MDCA (Figure S13 and S14).

We were thus able to engineer OleP, which exclusively formed MDCA from LCA, towards UDCA production. The variant F84Q/S240A/V291G showed perfect stereoselectivity and excellent regioselectivity for the 7β-hydroxylation of LCA. This discovery opens up new synthesis routes towards 7β-hydroxylated therapeutic agents like UDCA. These results also expand our knowledge about P450-mediated hydroxylations.

We successfully engineered the P450 monooxygenase OleP for the regio- and stereoselective 7β-hydroxylation of LCA to form UDCA. We started with alanine scanning of important active site residues identified from literature, inspection of crystal structures, and molecular docking. The bioinformatic tool 3DM was then used to design a "small but smart" synthetic library, which was efficiently screened employing a high-throughput colorimetric UDCA assay. Starting from an enzyme which did not even have trace activity in the formation of the target compound, we were able to create a variant that produces mainly UDCA, with only traces of MDCA formed. The near complete inversion of regioselectivity from 6β- (MDCA formation) to 7β-hydroxylation (UDCA production) demonstrates how protein engineering can be used to create custom biocatalysts, even for reactions where no naturally occurring counterpart has yet been identified.

Conflict of interest

S.K. and H.B. are employees of Enzymicals AG and B.G. is an employee of HERBRAND PharmaChemicals GmbH.

[1] U. Leuschner, M. Leuschner, J. Sieratzki, W. Kurtz, K. Hübner, *Dig. Dis. Sci.* **1985**, *30*, 642–649.

[2] a) P. Santiago, A. R. Scheinberg, C. Levy, *Ther. Adv. Gastroenter.* **2018**, *11*, 1–15; b) J.-F. Goossens, C. Bailly, *Pharmacol. Ther.* **2019**, *203*, 107396.

[3] K. Cheng, D. Ashby, R. L. Smyth, *Cochrane Db. Syst. Rev.* **2017**, *9*, CD000222.

[4] a) A. F. Gulamhusein, G. M. Hirschfield, *Nat. Rev. Gastro. Hepat.* **2020**, *17*, 93–110; b) A. Goel, W. R. Kim, *Clin. Liv. Dis.* **2018**, *22*, 563–578; c) D. M. H. Chascsa, K. D. Lindor, *J. Gastroenterol.* **2020**, *55*, 261–272.

[5] L. R. Hagey, D. L. Crombie, E. Espinosa, M. C. Carey, H. Igimi, A. F. Hofmann, *J. Lipid Res.* **1993**, *34*, 1911–1917.

[6] X.-L. He, L.-T. Wang, X.-Z. Gu, J.-X. Xiao, W.-W. Qiu, *Steroids* **2018**, *140*, 173–178.

[7] M.-M. Zheng, R.-F. Wang, C.-X. Li, J.-H. Xu, *Process Biochem.* **2015**, *50*, 598–604.

[8] F. Tonin, I. W. C. E. Arends, *Beilstein J. Org. Chem.* **2018**, *14*, 470–483.

[9] a) A. F. Hofmann, G. Lundgren, O. Theander, J. S. Brimacombe, M. C. Cook, *Acta Chem. Scand.* **1963**, *17*, 173–186; b) T. Eggert, D. Bakonyi, W. Hummel, *J. Biotechnol.* **2014**, *191*, 11–21.

[10] a) P. P. Giovannini, A. Grandini, D. Perrone, P. Pedrini, G. Fantin, M. Fogagnolo, *Steroids* **2008**, *73*, 1385–1390; b) F. Tonin, L. G. Otten, I. W. C. E. Arends, *ChemSusChem* **2019**, *12*, 3192–3203.

[11] R. J. Park, *Steroids* **1981**, *38*, 383–395.

[12] G. M. Sheriha, G. R. Waller, T. Chan, A. D. Tillman, *Lipids* **1968**, *3*, 72–78.

[13] S.-J. J. Tsai, Y.-S. Zhong, J.-F. Weng, H.-H. Huang, P.-Y. Hsieh, *J. Chromatogr. A* **2011**, *1218*, 524–533.

[14] S. Wu, R. Snajdrova, J. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2020**, *59*, https://doi.org/10.1002/anie.202006648; *Angew. Chem.* **2020**, *132*, https://doi.org/10.1002/ange.202006648.

[15] V. V. Kollerov, D. Monti, N. O. Deshcherevskaya, T. G. Lobastova, E. E. Ferrandi, A. Larovere, S. A. Gulevskaya, S. Riva, M. V. Donova, *Steroids* **2013**, *78*, 370–378.

[16] S. Kulprecha, T. Ueda, T. Nihira, T. Yoshida, H. Taguchi, *Appl. Environ. Microbiol.* **1985**, *49*, 338–344.

[17] A. F. Hofmann, L. R. Hagey, *Cell. Mol. Life Sci.* **2008**, *65*, 2461–2483.

[18] V. B. Urlacher, M. Girhard, *Trends Biotechnol.* **2019**, *37*, 882–897.

[19] a) H. Park, G. Park, W. Jeon, J.-O. Ahn, Y.-H. Yang, K.-Y. Choi, *Biotechnol. Adv.* **2020**, *40*, 107504; b) Y. Wei, E. L. Ang, H. Zhao, *Curr. Opin. Chem. Biol.* **2018**, *43*, 1–7.

[20] G. Qu, A. Li, Z. Sun, C. G. Acevedo-Rocha, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59*, 13204–13231; *Angew. Chem.* **2020**, *132*, 13304–13333.

[21] L. C. Montemiglio, G. Parisi, A. Scaglione, G. Sciara, C. Savino, B. Vallone, *Biochim. Biophys. Acta Gen. Subj.* **2016**, *1860*, 465–475.

[22] G. Parisi, L. C. Montemiglio, A. Giuffrè, A. Macone, A. Scaglione, G. Cerutti, C. Exertier, C. Savino, B. Vallone, *FASEB J.* **2019**, *33*, 1787–1800.

[23] a) S. Shah, Q. Xue, L. Tang, J. R. Carney, M. Betlach, R. McDaniel, *J. Antibiot.* **2000**, *53*, 502–508; b) C. Olano, A. M. Rodriguez, J. M. Michel, C. Méndez, M. C. Raynal, J. A. Salas, *Mol. Gen. Genet.* **1998**, *259*, 299–308; c) S. Gaisser, R. Lill, J. Staunton, C. Méndez, J. Salas, P. F. Leadlay, *Mol. Microbiol.* **2002**, *44*, 771–781.

[24] S. K. Lee, D. B. Basnet, J. S. J. Hong, W. S. Jung, C. Y. Choi, H. C. Lee, J. K. Sohng, K. G. Ryu, D. J. Kim, J. S. Ahn, et al., *Adv. Synth. Catal.* **2005**, *347*, 1369–1378.

[25] H. Agematu, N. Matsumoto, Y. Fujii, H. Kabumoto, S. Doi, K. Machida, J. Ishikawa, A. Arisawa, *Biosci. Biotechnol. Biochem.* **2006**, *70*, 307–311.

[26] S. Grobe, A. Wszołek, H. Brundiek, M. Fekete, U. T. Bornscheuer, *Biotechnol. Lett.* **2020**, *42*, 819–824.

[27] O. Gotoh, *J. Biol. Chem.* **1992**, *267*, 81–90.

[28] R. K. Kuipers, H.-J. Joosten, W. J. H. van Berkel, N. G. H. Leferink, E. Rooijen, E. Ittmann, F. van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. P. Martins dos Santos, P. J. Schaap, *Proteins Struct. Funct. Bioinf.* **2010**, *78*, 2101–2113.

[29] a) H. Jochens, U. T. Bornscheuer, *ChemBioChem* **2010**, *11*, 1861–1866; b) H. Jochens, M. Hesseler, K. Stiba, S. K. Padhi, R. J. Kazlauskas, U. T. Bornscheuer, *ChemBioChem* **2011**, *12*, 1508–1517; c) H. Jochens, D. Aerts, U. T. Bornscheuer, *Protein Eng. Des. Sel.* **2010**, *23*, 903–909.

[30] A. Li, C. G. Acevedo-Rocha, Z. Sun, T. Cox, J. L. Xu, M. T. Reetz, *ChemBioChem* **2018**, *19*, 221–228.

[31] a) D. Appel, R. D. Schmid, C.-A. Dragan, M. Bureik, V. B. Urlacher, *Anal. Bioanal. Chem.* **2005**, *383*, 182–186; b) M. Lisurek, M.-J. Kang, R. W. Hartmann, R. Bernhardt, *Biochem. Biophys. Res. Commun.* **2004**, *319*, 677–682.

[32] L. Liu, A. Aigner, R. D. Schmid, *Appl. Microbiol. Biotechnol.* **2011**, *90*, 127–135.

[33] F. Gielen, R. Hours, S. Emond, M. Fischlechner, U. Schell, F. Hollfelder, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E7383–E7389.

[34] A. Li, C. G. Acevedo-Rocha, L. D'Amore, J. Chen, Y. Peng, M. Garcia-Borràs, C. Gao, J. Zhu, H. Rickerby, S. Osuna, J. Zhou, M. T. Reetz, *Angew. Chem. Int. Ed.* **2020**, *59*, 12499–12505; *Angew. Chem.* **2020**, *132*, 12599–12605.