

Biosynthetic Pathways

DOI: 10.1002/anie.201201445

A Sweet Origin for the Key Congocidine Precursor 4-Acetamidopyrrole-2-carboxylate**

Sylvie Lautru,* Lijiang Song, Luc Demange, Thomas Lombès, Hervé Galons, Gregory L. Challis, and Jean-Luc Pernodet

Pyrrole groups are found in nature in two primary metabolites (heme and tryptophan) and in a large variety of secondary metabolites (e.g. pyoluteorin, clorobiocin, congocidine, and prodigiosins). The diversity of these metabolites is mirrored in the multiple biosynthetic pathways leading to pyrrole groups. Six different biosynthetic pathways have been characterized to date, [1,2] involving various primary metabolite precursors, such as amino acids (glycine, proline, serine, threonine, and tryptophan), [1-10] dicarboxylic acids (malonate, oxaloacetate, and succinate), [1,2,7,8] or N-(5'-phosphoribosyl)anthranilate.[1]

Pyrrolamides are a family of natural products, synthesized by Streptomyces and related actinobacteria, that all contain one or more pyrrole-2-carboxamide moieties in their structure. Most pyrrolamides, such as the well-characterized congocidine (1, also called netropsin; Figure 1) and distamycin, bind noncovalently in the DNA minor groove with some sequence specificity,[11] but pyrronamycin B has been suggested to bind DNA covalently. [12] This capacity to bind DNA confers on them a variety of biological activities, such as antiviral, antibacterial, antitumor, and anthelmintic activities,

[*] Dr. S. Lautru, Dr. J.-L. Pernodet Institut de Génétique et Microbiologie Université Paris-Sud, CNRS, UMR8621, 91405 Orsay (France) E-mail: sylvie.lautru@igmors.u-psud.fr

Dr. L. Song, Prof. G. L. Challis Department of Chemistry, University of Warwick, Coventry CV4 7AL (UK)

Dr. L. Demange, Prof. H. Galons

Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques UMR 8601 CNRS

Université Paris Descartes, Sorbonne Paris Cité, UFR Biomédicale des Saints Pères

45 rue des Saints Pères, 75006 Paris (France)

T. Lombès

UMR 8638 CNRS-Paris Descartes, Faculté des Sciences Pharmaceutiques et Biologiques

4 av. de l'Observatoire, 75006 Paris (France)

[**] We would like to thank Dr. Christophe Corre for helpful discussions, Dr. Laurent Micouin for his help with the synthesis of the 4acetamidopyrrole-2-methanol, and Dr. Cyrille Kouklovsky, Géraldine Le Goff, and Dr. Robert Thaï for their help with mass spectrometry. This work was supported in part by the European Union through the Integrated Project ActinoGEN (CT-2004-0005224) and the reintegration grant BOPA (CT-2005-029154), by the Pôle de Recherche et d'Enseignement Supérieur UniverSud Paris and by the Bettencourt Schueller Foundation.



Supporting information for this article (experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie. 201201445.

but also renders them too toxic for clinical use. Nonetheless, because molecules binding to DNA at specific positions can manipulate the expression of genes involved in various diseases (such as cancer), congocidine and distamycin have prompted the synthesis of many structurally related molecules that bind in the minor groove of DNA at various defined sequences.[13]

The nature and biosynthetic origin of the pyrrolamide pyrrole precursor are unknown. A retrobiosynthetic analysis of pyrrolamide structures suggests 4-aminopyrrole-2-carboxylate as the potential pyrrole precursor common to all pyrrolamides. However, this remains to be established and no biosynthetic pathway has been reported for the synthesis of this molecule. We report herein that 4-acetamidopyrrole-2carboxylate (10) is the true precursor of congocidine and propose for this compound a biosynthetic pathway starting from N-acetylglucosamine-1-phosphate (2), involving carbohydrate metabolizing enzymes, and differing entirely from known pyrrole biosynthetic pathways.

We recently reported the identification, analysis, and heterologous expression of the first pyrrolamide gene cluster; the cgc gene cluster directs congocidine biosynthesis in Streptomyces ambofaciens.[14] Sequence analyses of the proteins encoded by the cgc cluster led us to propose a pathway for congocidine biosynthesis involving a noncanonical nonribosomal peptide synthetase (NRPS) and three putative precursors: guanidinoacetate, 3-aminopropionamidine, and 4aminopyrrole-2-carboxylate. However, the biosynthetic origins of these precursors could not easily be inferred from the analysis of the cgc gene cluster. In particular, we could not identify any gene encoding homologues of known pyrrole biosynthetic enzymes, suggesting that the pyrrole groups in pyrrolamides could be synthesized through an entirely new pathway.

The cgc18 gene encodes an NRPS module that has been proposed to recognize 4-aminopyrrole-2-carboxylate and catalyze its ATP-dependent activation as the corresponding peptidyl carrier protein (PCP)-bound thioester. [14] To investigate whether 4-aminopyrrole-2-carboxylate is indeed a precursor of congocidine, we used HPLC to compare the profile of metabolites in the culture supernatants of the "wild type" SPM110 Streptomyces ambofaciens strain and the CGCA004 mutant strain, in which cgc18 has been disrupted. [14] A new compound accumulated in the culture supernatant of CGCA004 (retention time = 12.3 min, Figure 2a), which we expected to be 4-aminopyrrole-2-carboxylate. This metabolite peak was not detected in the CGCA004 chromatogram of our previous report[14] and on average, it was detected in approximately half of the cultures of strains deleted in cgc

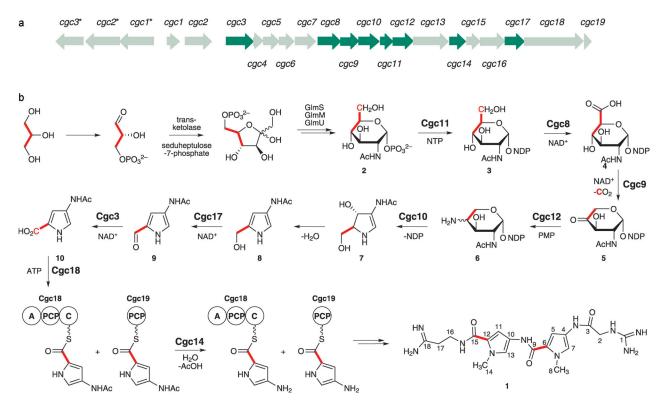


Figure 1. Congocidine gene cluster (a) and proposed pathway for biosynthesis of 4-acetamidopyrrole-2-carboxylate (10) from fructose-6-phosphate by way of an intermediate 2 (b). [U-13C]Glycerol is proposed to be converted into [U-13C]glyceraldehyde-3-phosphate, which undergoes transaldolase-catalyzed condensation with seduheptulose-7-phosphate with loss of erythrose-4-phosphate to yield fructose-6-phosphate bearing ¹³C labels at C-4, C-5, and C-6. The positions of the ¹³C labels are highlighted in red. The primary metabolic enzymes GlmSMU catalyze conversion of fructose-6-phosphate into 2. Genes involved in 10 biosynthesis are indicated in green.

genes not involved in the biosynthesis of 10. It is likely that this variability is related to the variability of secondary metabolite production often seen in Streptomyces. [15] LC-MS analysis of the CGCA004 culture supernatant showed that this metabolite produces an m/z 168.9 ion rather than the expected m/z 127.1 ion (Supporting Information, Figure S1 a). MS/MS analyses suggested that the accumulated metabolite was 10 (Supporting Information, Figure S1 a). This hypothesis was confirmed by the preparation of a standard of 10, whose retention time (Figure 2b), UV/Vis spectrum, and MS/MS fragmentation pattern (Supporting Information, Figure S1b) were identical to those recorded for the metabolite accumulated in the CGCA004 culture supernatant.

We next confirmed that 10 is an intermediate in congocidine biosynthesis and investigated the pathway for its assembly. Among the 22 genes constituting the cgc gene cluster, 13 could not be assigned a function in regulation of gene expression, congocidine resistance, or assembly of the putative precursors, guanidinoacetate, 3-aminopropionamidine, and 10.^[14] These 13 genes encode proteins with similarity to enzymes of known function (with the exception of cgc7) and are likely to be involved in the biosynthesis of congocidine precursors. Six of the genes encode proteins that are similar to enzymes involved in carbohydrate biosynthesis or transfer (Cgc8, putative NDP-hexose dehydrogenase; Cgc9, putative NDP-hexose epimerase/dehydrogenase; Cgc10, putative glycosyltransferase; Cgc11, putative hexose nucleotidyltransferase; Cgc12, putative DegT/DnrJ/EryC1/StrS aminotransferase; Cgc13, putative glycoside hydrolase). The presence of these genes in the cgc cluster is intriguing because congocidine does not contain a carbohydrate moiety. Yet, we have previously shown that at least one of these genes, cgc10, is involved in congocidine biosynthesis.^[14] To determine whether the other five genes are also involved in congocidine biosynthesis, we constructed in-frame deletions of each within a bacterial artificial chromosome (BAC; pCGC002) containing the entire cgc cluster. The resulting constructs were heterologously expressed in S. lividans TK23. None of the S. lividans strains with the genetically modified BACs produced congocidine, which would elute at 15.1 min, (Figure 2 dand Supporting Information, Figure S3, left panels) except the CGCL047 strain ($\triangle cgc13$), which produced residual congocidine (15% of the "wild type" strain CGCL006). This indicates that the genes encoding putative carbohydrate biosynthesis/transfer enzymes are involved in congocidine biosynthesis and, therefore, are likely to participate in biosynthesis of a congocidine precursor.

We examined which of the carbon atoms of congocidine are carbohydrate-derived using incorporation experiments with labeled precursors. Cultures of S. ambofaciens SPM110 were pulse-fed with uniformly 13C-labeled glycerol ([U-¹³C]glycerol; 4×100 μmol) and congocidine was purified by cation exchange chromatography. The DEPTQ-135 NMR spectrum (Supporting Information, Figure S4) of the isolated congocidine showed coupled doublets flanking the natural abundance signals for C-2/C-3, C-6/C-9, and C-12/C-15,

7455



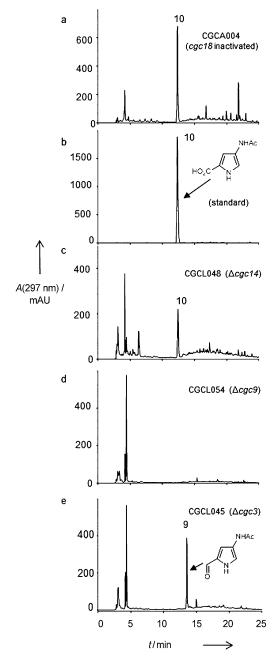


Figure 2. LC analysis of culture supernatents of mutants with deletions in congocidine biosynthetic genes. Chromatograms from a) CGCA004 (disruption of cgc18); b) synthetic **10**; c) CGCL048 ($\Delta cgc14$); d) CGCL054 ($\Delta cgc9$), and e) CGCL045 ($\Delta cgc3$).

showing that a two-carbon fragment of glycerol was incorporated intact into C-2/C-3, C-6/C-9, and C-12/C-15 (see Figure 1 for carbon atom numbering). The guanidinoacetate precursor of congocidine has been suggested to originate from glycine by way of transguanylation. This is consistent with the incorporation pattern seen for C-2/C-3 (Supporting Information, Figure S5). More significant, however, is the incorporation of two carbons from glycerol into the same position (C-2 and the carboxamide group) in both 4-amino-pyrrole-2-carboxamide units of congocidine. This incorporation pattern is consistent with derivation of the 4-amino-

pyrrole-2-carboxamide units from fructose-6-phosphate by way of a **2** intermediate (Figure 1).^[17] To verify this hypothesis, we next attempted to feed cultures of *S. ambofaciens* with [¹⁵N,¹³C-1] *N*-acetylglucosamine. Unfortunately, the addition of *N*-acetylglucosamine strongly suppresses congocidine production (presumably because of DasR-mediated regulation).^[18] Thus, we were not able to obtain enough congocidine to determine whether *N*-acetylglucosamine was incorporated.^[19]

Taken together these data suggest that the putative carbohydrate-processing enzymes encoded within the cgc cluster are responsible for the assembly of 10 from 2. To further investigate this hypothesis and examine whether 10 is an intermediate in congocidine biosynthesis, we fed chemically synthesized **10** to cultures of the cgc11, cgc8, cgc9, cgc12, and cgc10 deletion mutants. LC-MS analyses showed that production of congocidine was restored in all of the mutants (Supporting Information, Figure S3, central and right panels), indicating that **10** is a congocidine precursor and that *cgc11*, cgc8, cgc9, cgc12, and cgc10 are all involved in its biosynthesis. When the cgc13 deletion mutant was fed with 10, congocidine production was increased by a factor of 2.5 (Supporting Information, Figure S3h). This result suggests that the reduced congocidine production in the cgc13 deletion mutant is due to impaired synthesis of **10**.

To determine whether other cgc genes are involved in 10 biosynthesis, the six cgc genes with no assigned function (cgc3, cgc4, cgc5, cgc6, cgc7, and cgc17) were individually deleted in frame within the pCGC002 BAC and the resulting BACs heterologously expressed in S. lividans TK23. The cgc4 deletion mutant produced residual congocidine, whereas congocidine production was abolished in the other deletion mutants (Supporting Information, Figure S3 f, g, left panels; Figure S6, left panels). The cgc4, cgc5, cgc6, and cgc7 deletion mutants accumulated 10 (confirmed by MS analyses, Supporting Information, Figure S6, right panels), indicating that these genes are not involved in 10 biosynthesis. The cgc3 and cgc17 deletion mutants, however, did not produce 10. The involvement of Cgc3 (a putative aldehyde dehydrogenase)^[14] and Cgc17 (a putative alcohol dehydrogenase)^[14] in **10** biosynthesis was confirmed by feeding 10 to the mutant strains, which restored congocidine production (Supporting Information, Figure S3 f, g, central and right panels).

The cgc3 deletion mutant accumulated a new metabolite in the culture medium that was not found in the other mutants (Supporting Information, Figure S3g). MS and NMR spectroscopic analyses of the purified metabolite (Supporting Information, Figure S7) showed that it was 4-acetamidopyrrole-2-carboxaldehyde (9), which is likely to be the substrate of Cgc3 (Figure 3).

The cgc17 gene encodes a putative alcohol dehydrogenase possibly involved in the oxidation of 4-acetamidopyrrole-2-methanol (8) to yield 9 (Figure 1). To verify this hypothesis, cultures of the cgc11, cgc8, cgc9, cgc12 cgc10, cgc17, and cgc3 deletion mutants were fed with chemically synthesized 8 (Supporting Information, methods and Figure S8). Production of congocidine was restored for the first five mutant strains (Supporting Information, Figure S9). In contrast, feeding of 8 to the cgc17 and cgc3 deletion mutants did not



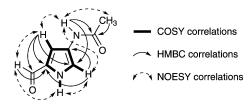


Figure 3. Summary of COSY, HMBC, and NOESY correlations recorded for 9 purified from the culture supernatant of the cgc3 mutant.

restore congocidine production. These results are consistent with 8 being the substrate of Cgc17. Accumulation of 8 in the cgc17 deletion mutant could not be detected using standard LC-MS analysis because this compound is unstable and poorly retained on reverse-phase HPLC columns.

Based on our data, we suggest the following pathway for 10 biosynthesis from 2 (Figure 1). Cgc11 catalyzes nucleotide triphosphate (NTP)-dependent conversion of 2 into the nucleotide diphosphate (NDP) derivative 3, which undergoes Cgc8-catalyzed oxidation of the C-6 hydroxy group to a carboxy group to yield NDP-2-acetamido-2-deoxyglucopyranuronic acid (4). Cgc9 catalyzes the decarboxylation of 4 to afford NDP-threo-2-acetamido-2-deoxy-pentopyran-4ulose (5). Reductive amination of the C-4 keto group of 5 is catalyzed by the putative pyridoxal-5-phosphate-dependent aminotransferase Cgc12. The next step involves intramolecular nucleophilic attack of the C-4 amino group on C-1 to displace the NDP and afford 4-acetamido-2,3-dihydropyrrole-2-methanol (7). This conversion of a carbohydrate into a pyrrole molecule is proposed to be catalyzed by Cgc10, an enzyme that resembles a glycosyltransferase. Loss of water from 7 to form 8 could then occur spontaneously (the formation of the resulting pyrrole aromatic ring is expected to be thermodynamically favorable). Finally, oxidation of the hydroxy group of 8 to the aldehyde 9 and subsequently the carboxylic acid 10 is catalyzed by Cgc17 and Cgc3, respectively.

The initial steps of our proposed pathway are catalyzed by enzymes (Cgc11, Cgc8, Cgc9, and Cgc12) that are homologous to enzymes catalyzing similar reactions in the biosynthesis of aminodeoxypentose sugars found in enediyne natural products such as maduropeptin, calicheamicin, and AT2433 (Supporting Information, Figure S10). [20,21] Carbohydrates incorporated in these natural products can be activated using UTP or TTP. By analogy, it appears likely that UTP or TTP is the NTP used to activate 2. UDP-N-acetylglucosamine, derived from 2, is an intermediate in the biosynthesis of bacterial peptidoglycan. [22,23] As the cgc11 deletion mutant does not produce any congocidine, it is unlikely that UDP-Nacetylglucosamine is an intermediate in congocidine biosynthesis. Indeed, activation of 2 by a nucleotide triphosphate other than UTP (most likely TTP) could constitute the first committed step of 10 biosynthesis.

Our data show that Cgc13 is not essential for congocidine biosynthesis. Nonetheless the reduction of congocidine production in the cgc13 deletion mutant, and the partial restoration of the production when the mutant is fed with 10, indicates that Cgc13 may participate in the supply of 10. The similarity of Cgc13 with glycoside hydrolases suggests that Cgc13 could be involved in providing 2.

Our proposed pathway results in the biosynthesis of 10. Yet, before condensation with the other congocidine precursors can occur, this molecule must be de-acetylated to yield 4aminopyrrole-2-carboxylate. Among the proteins encoded by the cgc gene cluster, we identified Cgc14, a predicted amidohydrolase, as the enzyme likely to catalyze the deacetylation. Therefore, we constructed a cgc14 in-frame deletion mutant (CGCL048). This mutant did not produce congocidine but accumulated 10 (Figure 2c), consistent with the proposed role for Cgc14. However a mutant strain impaired in congocidine assembly (CGCA004) also accumulates 10 and not 4-aminopyrrole-2-carboxylate. These observations suggest that 10, rather than 4-aminopyrrole-2-carboxylate, may be the substrate of the Cgc18 adenylation domain and that de-acetylation occurs after 10 has been loaded onto the Cgc18 and Cgc19 PCPs (Figure 1). Consistent with this hypothesis, feeding of the cgc14 deletion mutant with commercially available 4-aminopyrrole-2-carboxylate did not restore congocidine production (Supporting Information, Figure S11). N-acetylation is a well characterized mechanism of arylamine detoxification and maintaining the 4-aminopyrrole-2-carboxylate in its N-acetylated form when free in solution would avoid the cytotoxic effects associated with arylamines.[24]

In conclusion, we have shown that 10 is the true intermediate in the biosynthesis of congocidine. Interestingly, the biosynthesis of 10 described herein is unprecedented in that it involves enzymes and precursors from carbohydrate metabolism. In particular, it bears no resemblance to the biosynthesis of other pyrrole-2-carboxylates incorporated into secondary metabolites such as clorobiocin, [6] coumermycin,[4,6] pyoluteorin, [3,25] and undecylprodiginine. [3,7,8] Although this pathway is likely to be involved in the biosynthesis of other members of the pyrrolamide family of metabolites, it is not currently known whether it is also utilized in the biosynthesis of other pyrrole-containing natural products. However, database searches identified a putative 10 biosynthetic gene cluster in the genome of Micromonospora carbonacea var. Africana ATCC 39149 (MCAG_05565 and MCAG_05568 to MCAG_05575). Moreover, it is interesting to note that this pathway, with the reductive amination reaction replaced by a ketoreduction, would be perfectly adapted to the biosynthesis of the furan precursors of proximicins, analogues of congocidine recently isolated from marine actinomycetes.[26]

Received: February 21, 2012 Revised: April 11, 2012 Published online: June 15, 2012

Keywords: biosynthesis · congocidine · natural products · pyrrolamide · pyrrole

^[1] C. T. Walsh, S. Garneau-Tsodikova, A. R. Howard-Jones, Nat. Prod. Rep. 2006, 23, 517-531.

^[2] S. Siebenberg, N. Burkard, A. Knuplesch, B. Gust, S. Grond, L. Heide, ChemBioChem 2011, 12, 2677-2685.



- [3] M. G. Thomas, M. D. Burkart, C. T. Walsh, Chem. Biol. 2002, 9, 171–184.
- [4] H. Xu, Z.-X. Wang, J. Schmidt, L. Heide, S.-M. Li, *Mol. Genet. Genomics* 2002, 268, 387 396.
- [5] S. Maharjan, N. Aryal, S. Bhattarai, D. Koju, J. Lamichhane, J. K. Sohng, *Appl. Microbiol. Biotechnol.* 2012, 93, 687–696.
- [6] S. Garneau, P. C. Dorrestein, N. L. Kelleher, C. T. Walsh, Biochemistry 2005, 44, 2770–2780.
- [7] S. Mo, P. K. Sydor, C. Corre, M. M. Alhamadsheh, A. E. Stanley, S. W. Haynes, L. Song, K. A. Reynolds, G. L. Challis, *Chem. Biol.* 2008, 15, 137–148.
- [8] A. E. Stanley, L. J. Walton, M. Kourdi Zerikly, C. Corre, G. L. Challis, Chem. Commun. 2006, 3981–3983.
- [9] T. Nishizawa, S. Grüschow, D.-H. E. Jayamaha, C. Nishizawa-Harada, D. H. Sherman, J. Am. Chem. Soc. 2006, 128, 724 – 725.
- [10] S. Kirner, P. E. Hammer, D. S. Hill, A. Altmann, I. Fischer, L. J. Weislo, M. Lanahan, K. H. van Pée, J. M. Ligon, J. Bacteriol. 1998, 180, 1939–1943.
- [11] S. Neidle, Nat. Prod. Rep. 2001, 18, 291-309.
- [12] A. Asai, Y. Sakai, H. Ogawa, Y. Yamashita, S. Kakita, K. Ochiai, T. Ashizawa, A. Mihara, T. Mizukami, H. Nakano, J. Antibiot. 2000, 53, 66-69.
- [13] P. B. Dervan, Bioorg. Med. Chem. 2001, 9, 2215-2235.
- [14] M. Juguet, S. Lautru, F.-X. Francou, S. Nezbedová, P. Leblond, M. Gondry, J.-L. Pernodet, *Chem. Biol.* 2009, 16, 421 – 431.
- [15] S. Siebenberg, P. M. Bapat, A. E. Lantz, B. Gust, L. Heide, J. Biosci. Bioeng. 2010, 109, 230–234.
- [16] M. E. Wildfeuer, The biosynthesis of netropsin, PhD thesis, University of Delaware (USA), 1963.
- [17] Fructose-6-phosphate can also be formed from fructose-bisphosphate aldolase-mediated condensation of glyceraldehyde-3-phosphate and dihydroxyacetone phosphate with subsequent dephosphorylation. In this case, we would expect all six carbon atoms of fructose-6-phosphate to be labeled through incorpo-

- ration of two molecules of [U-¹³C]glycerol. However, fructose-bisphosphate aldolase does not appear to be expressed/active in *Streptomyces* species cultivated in rich growth media (see C. Corre, G. L. Challis, *ChemBioChem* **2005**, *6*, 2166–2170 and G. L. Challis, K. F. Chater, *Chem. Commun.* **2001**, 935–936). Thus, only incorporation of a single molecule of [U-¹³C]glycerol into fructose-6-phosphate by way of the pentose phosphate pathway is observed.
- [18] M. A. Świątek, E. Tenconi, S. Rigali, G. P. van Wezel, J. Bacteriol. 2012, 194, 1136–1144.
- [19] We tried to feed [U-¹³C]glucose to a mutant (CGCL022) that accumulates **10**. Some of the **10** was purified and analyzed by HRMS (see Figure S12). As expected, ¹³C-5 **10** was produced. However, in addition to the ¹³C-5 labeled compound and to the unlabeled compound, ¹³C-1, ¹³C-2, ¹³C-3, ¹³C-4 and ¹³C-7 labeled molecules were also observed. This nonspecific incorporation pattern can be expected with labeled glucose feeding experiments, as many primary metabolites that could be incorporated into the **2** precursor of **10** are derived from glucose.
- [20] Q. Gao, C. Zhang, S. Blanchard, J. S. Thorson, *Chem. Biol.* 2006, 13, 733–743.
- [21] S. G. van Lanen, T.-J. Oh, W. Liu, E. Wendt-Pienkowski, B. Shen, J. Am. Chem. Soc. 2007, 129, 13082 – 13094.
- [22] D. Mengin-Lecreulx, J. van Heijenoort, J. Bacteriol. 1994, 176, 5788-5795.
- [23] A. M. Gehring, W. J. Lees, D. J. Mindiola, C. T. Walsh, E. D. Brown, *Biochemistry* 1996, 35, 579–585.
- [24] M. Payton, A. Mushtaq, T. W. Yu, L. J. Wu, J. Sinclair, E. Sim, *Microbiology* 2001, 147, 1137-1147.
- [25] B. Nowak-Thompson, N. Chaney, J. S. Wing, S. J. Gould, J. E. Loper, J. Bacteriol. 1999, 181, 2166 – 2174.
- [26] K. Schneider, et al., Angew. Chem. 2008, 120, 3302-3305; Angew. Chem. Int. Ed. 2008, 47, 3258-3261.