Microbiome

International Edition: DOI: 10.1002/anie.201505341
German Edition: DOI: 10.1002/ange.201505341

The Microbes inside Us and the Race for Colibactin**

*Helge B. Bode**

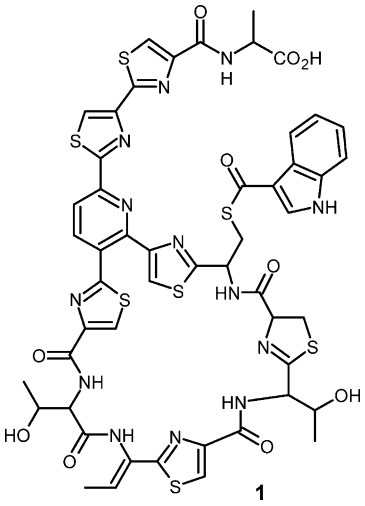
colibactin · colorectal cancer · dysbiosis · microbiome · prodrug activation

The human body is composed of about 200 different cell types (http://www.bioon.com/book/biology/mboc/mboc.cgi@code=220801800040279.htm), with overall 10^{13} cells making up the different tissues and organs, such as brain, skin, liver, and blood vessels. Although these numbers highlight at the same time the simplicity and the complexity of the human body, with its underlying biochemistry and cell–cell communication leading to multicellularity, they are nothing compared to the number and diversity of bacteria living in and on humans. Conservative estimations assume at least 2000 different species of bacteria with a total number of 10^{14} (ten times more than human cells) microbial cells on each human body (http://hmpdacc.org). They are essential for digesting food, training the immune system, protecting us against "bad" microbes by occupying their infection niche, and producing vitamins (e.g., biotin or vitamin K).

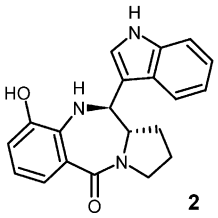
Gut microbes in particular are generally regarded as beneficial to our health, and there is increasing evidence that the gut microbiome can even influence neurological outcomes such as behavior or the onset or severity of disorders of the nervous system.[1] The mammalian immune system, which is generally thought to be designed to control microorganisms, might in fact be developed and controlled by the gut mibrobiome.[2]

With the development of modern sequencing technology, it has become obvious that not only does the organismic diversity of our microbiome exceed that of human cell types, but the number of microbial genes is also several hundred times greater than that of human genes.

These genes include small-molecule biosynthetic gene clusters (BGCs) that are involved in the biosynthesis of typical macrolide polyketides, nonribosomal peptides, or ribosomally encoded and posttranslationally modified peptides (RiPPs). An example is the thiopeptide lactocillin (**1**), derivatives of which are currently in clinical trials as antibiotics.[3] Thus it seems likely that drugs produced by our own microbiota also contribute to our health. On the other hand, an imbalance of symbionts (health-promoting bacteria), commensals (permanent residents with no benefit or detriment to the host), and pathobionts (pathogens), which is termed 'dysbiosis', can be directly correlated to obesity, diabetes, inflammation diseases like Crohn's disease, and even cancer.[2] Although the underlying mechanisms are just beginning to be revealed, there are several recent examples to suggest that typical natural products might also play a major role in these processes.[4]



1



2

Tilivalline (**2**), a pyrrolobenzodiazepine cytotoxin, is synthesized by a nonribosomal peptide synthetase (NRPS) in *Klebsiella oxytoca*.[5] *K. oxytoca* is the causative agent of antibiotic-associated hemorrhagic colitis (AAHC), a disease associated with antibiotic-driven enterobacterial overgrowth by *K. oxytoca*. This bacterium is a resident of the gut in 2–10% of healthy individuals and upon antibiotic therapy, it might be able to dominate the gut microbiome, thereby resulting in tilivalline-induced apoptosis and disrupted epithelial barrier function, which can ultimately lead to colitis.

[*] Prof. Dr. H. B. Bode
Merck Stiftungsprofessur für Molekulare Biotechnologie
Fachbereich Biowissenschaften, Goethe Universität Frankfurt
Max-von-Laue-Str. 9, 60438 Frankfurt am Main (Germany)
and
Buchmann Institute for Molecular Life Sciences (BMLS)
Goethe Universität Frankfurt
Max-von-Laue-Str. 15, 60438 Frankfurt a. M. (Germany)
E-mail: h.bode@bio.uni-frankfurt.de
Homepage: http://www.uni-frankfurt.de/fb/fb15/institute/inst-3-mol-biowiss/AK-Bode

[**] This work was supported by the DFG.

Another example for a famous compound produced by a gut bacterium is colibactin. Almost 10 years ago, it was shown that *E. coli* strains carrying the colibactin biosynthesis gene cluster can induce DNA double-strand breaks in eukaryotic cells, thereby leading to a block in mitosis and thus to megalocytosis, which ultimately contributes to colorectal cancer under host inflammation conditions.[6] The complete structure or biosynthesis of colibactin is not known yet but has triggered research efforts by several different groups during recent years. Colibactin is derived from a hybrid polyketides synthase (PKS)/NRPS and all of the genes involved in its biosynthesis were identified in the initial publication. Whereas 53 % of all extraintestinal pathogenic *E. coli* (ExPEC) strains carry the required genomic island, surprisingly, it was also found in 34 % of fecal samples from healthy individuals.[6] This gene cluster is also present in *Klebsiella pneumoniae*, *Enterobacter aerogenes, Citrobacter koseri*, a marine *Pseudovibrio* strain, and the bee gut symbiont *Fischerella perrara*.[7] Interestingly, one of the *E. coli* strains is *E. coli* Nissle 1917, a commensal strain that has been widely used as a probiotic treatment for intestinal disorders such as Crohn's disease. However, the biosynthesis gene cluster is complex and does not allow simple prediction of the final natural product. The starting point for the elucidation of colibactin biosynthesis was the identification of ClbP as a new class of peptidase involved in its activation during the biosynthesis.[8] Therefore, inhibition of ClbP by small molecules could confer protection against colibactin-induced colorectal cancers.[9]

The identification of similar prodrug activation mechanisms in other natural products derived from NRPS and PKS/NRPS hybrids revealed a widespread occurrence of such a mechanism,[10] which was originally identified for the antibiotic xenocoumacin from *Xenorhabdus nematophila*.[11] There, it was shown that acylated prexenocoumacins are cleaved after a conserved D-asparagine upon secretion from the cell, which results in the active antibiotic xenocoumacin 1. Subsequent analysis of different colibactin producers led to the identification of the analogous cleavage product acyl-D-Asn (**3**) and an elongated product **4** (Figure 1), for which a biosynthetic scheme was proposed.[12–14]

In 2015, the search for the colibactin culminated in the publication of a unique 7-methyl-4-azaspiro[2.4]hept-6-en-5-one (**5**) structure by the groups of Balskus, Crawford, and Müller.[15–17] Whereas Brotherton et al. were able to show ClbP cleavage of **5**,[15] Vizcaino and Crawford could additionally demonstrated DNA cross-linking activity for **5** in vitro, thus suggesting a molecular mechanism for its toxicity after ClbP activation.[17] Moreover, they also proposed a structure for a precolibactin (**6**; Figure 1) based on extensive MS analysis in combination with labelling experiments that allowed them to propose a biosynthetic model for colibactin biosynthesis.

Very recently, Li et al. added another piece to the colibactin puzzle.[18] They cloned the whole biosynthesis gene cluster from *E. coli* CFT073 and expressed it in classical *E. coli* expression strains, thereby leading to a colibactin overproducer that produces colibactin at 12-fold higher levels than the original strain Through large-scale cultivation (200 L) of this strain, they were able to identify a 1*H*-pyrrolo[3,4-*c*]pyridine-3,6(2*H*,5*H*)-dione moiety (**7**), which is connected to a thiazole unit in a late-step intermediate (**8**), and the structures of these compounds were solved by NMR. Compounds **7** and **8** were produced only in minute quantities of 6 and 0.5 mg, respectively, and unfortunately no experimental details on the cultivation and isolation procedure were given. MS analysis also allowed them to predict a structure for the precolibactin (**9**), which was isolated in even smaller amounts (0.1 mg), and this structure differs from that proposed by Crawford (Figure 1). Individual deletions of several *clb* genes additionally allowed them to predict a biosynthesis pathway

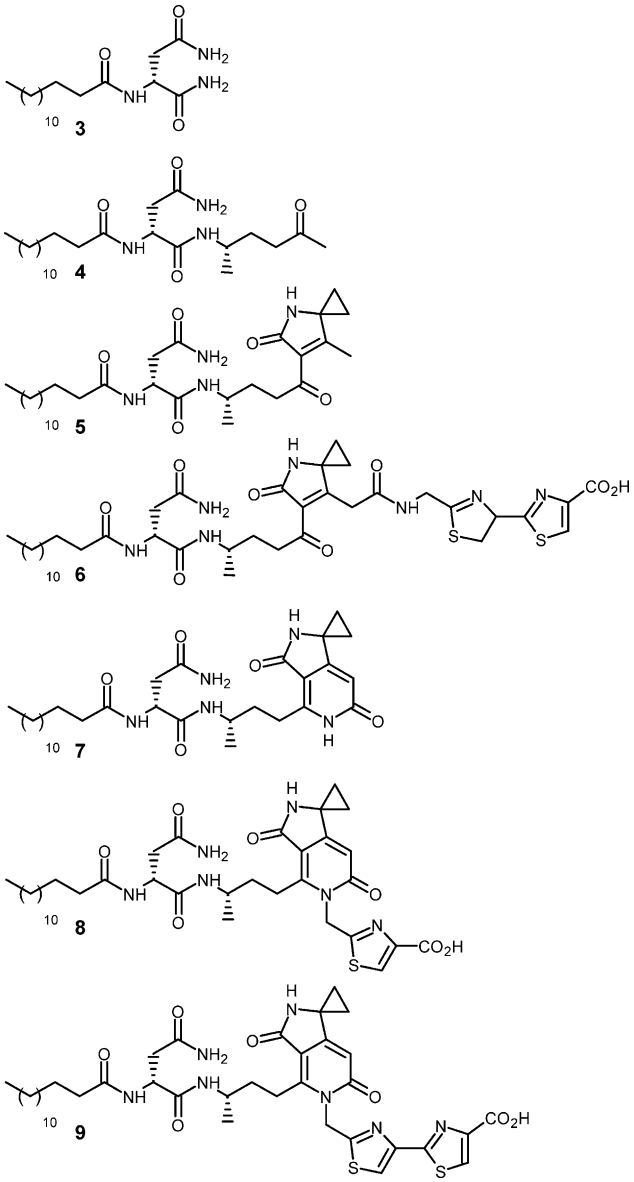


Figure 1. Major colibactin intermediates. Whereas the structures of **3**–**5**, **7**, and **8** have been elucidated based on NMR and MS data, the structures of the proposed full-length precolibactin derivatives **6** and **9** are only based on MS and labeling data and the predicted biosynthesis. The major derivatives with a myristoyl side chain are shown but other chain lengths have also been observed.

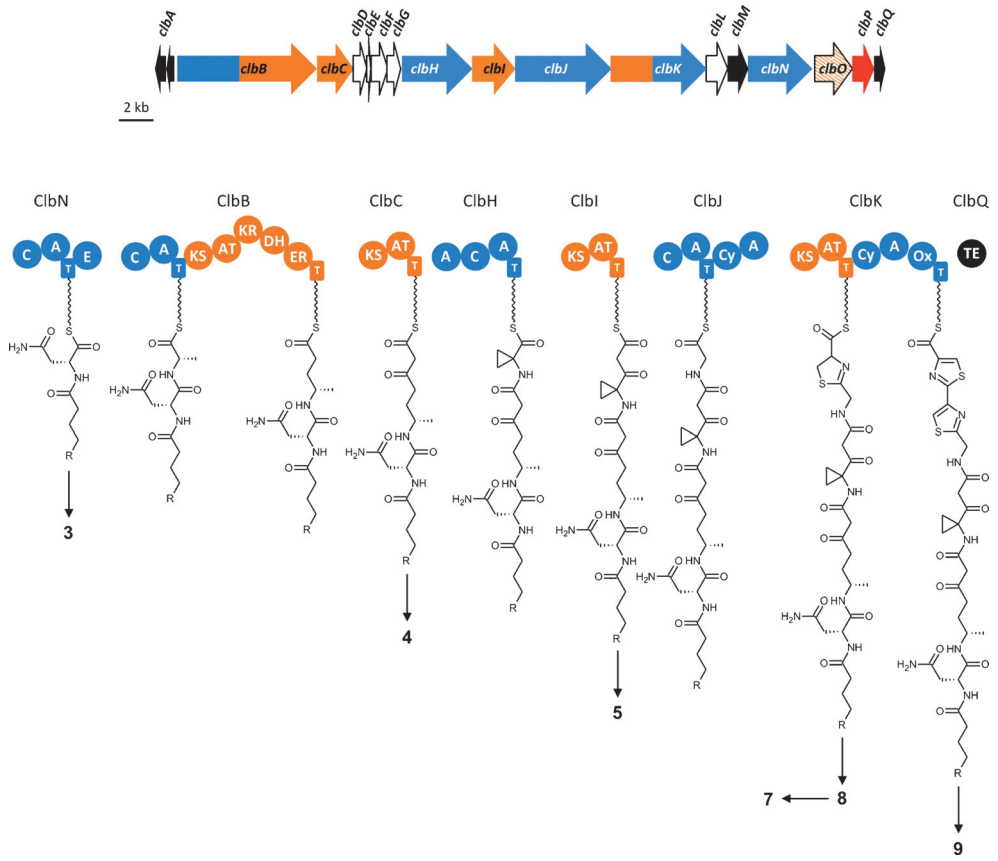


Figure 2. The colibactin biosynthesis gene cluster (*clb*) from *E. coli* (top) and the biosynthesis proposed by Li et al.[18] (bottom). The NRPS and PKS (and the respective encoding genes) are shown in blue and orange, respectively. The gene encoding the peptidase ClbP, which is responsible for prodrug activation, is highlighted in red. Genes encoding enzymes with known roles in the colibactin biosynthesis are shown in black, genes not assigned to the biosynthesis model are shown in white (*clbDEF and clbL*) or white/orange (*clbO*; encodes a PKS module).

that represents the current state of knowledge for colibactin biosynthesis (Figure 2).

Nevertheless, the race to elucidate the structure and biosynthesis of colibactin is most likely not finished yet:

1. The structure of the proposed full-length precolibactin variant(s) still needs to be proven by NMR or another MS-independent method.
2. The observation that contact between bacterial and eukaryotic cells is required for full toxicity[6] and the always very low amounts of identified intermediates might indicate that colibactin is somehow bound to the bacterial cell and thus might point to a larger colibactin structure.
3. The functions of ClbDEF are still unclear since they are not included in the current biosynthesis models but are highly conserved in all *clb* BGCs.[7] Close derivatives of these proteins are involved in the biosynthesis of the unusual polyketide extender unit aminomalonyl-ACP in zwittermicin biosynthesis,[19] but the current biosynthesis model only shows compounds derived from malonyl-ACP.
4. Similarly, the role of the amidase ClbL and the additional PKS ClbO are unclear and have not been included in any biosynthesis model, which again raises the possibility of a (pre)colibactin structure larger than or at least different to the currently proposed **6** or **9**.
5. Are all of the identified structures only non-specific "shunt" products of the biosynthesis pathway, do they contribute to the overall bioactivity, or do they even have different bioactivities? The latter might be suggested from the weak antibacterial activity of the ClbP cleavage product **3**, which might contribute to growth inhibition of other symbionts in the gut and thus might confer a niche advantage. The fact that the short product **5** shows in vitro DNA cross-linking activity[17] might also support this role, thus raising the additional question of whether the "true" colibactin might show DNA sequence selectivity arising from the combination of the 1*H*-pyrrolo[3,4-*c*]pyridine-3,6(2*H*,5*H*)-dione moiety and the thiazole tail, since the latter has been shown to be involved in DNA intercalation in other natural products.
6. The presence of closely related *clb* biosynthesis gene clusters in several different bacteria from different ecological niches might indicate similar structures and biological functions. However, this needs to be confirmed experimentally.

In summary, the still ongoing race to elucidate the structure and biosynthesis of colibactin shows the power as well as the limits of modern natural product research. Colibactin is a great example of the natural products encoded

by "our other genome", the microbiome, how these compounds can influence human health or disease,[20] and how a better understanding of their biosynthesis might help in the development of treatments for severe diseases.[9] The wealth of such compounds and their importance are just starting to be unveiled. With the recent development of bioinformatics, bioanalytics (especially MS methods), 3rd generation sequencing technologies, and molecular tools for manipulation of the identified BGCs,[21] it seems to be the right time to focus on the microbes in and on us for understanding human biochemistry and as a resource for finding novel and bioactive natural products.

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 10408–10411
Angew. Chem. **2015**, *127*, 10550–10553

[1] T. R. Sampson, S. K. Mazmanian, *Cell Host Microbe* **2015**, *17*, 565–576.
[2] J. L. Round, S. K. Mazmanian, *Nat. Rev. Immunol.* **2009**, *9*, 313–323.
[3] M. S. Donia, P. Cimermancic, C. J. Schulze, L. C. Wieland Brown, J. Martin, M. Mitreva, J. Clardy, R. G. Linington, M. A. Fischbach, *Cell* **2014**, *158*, 1402–1414.
[4] P. C. Dorrestein, S. K. Mazmanian, R. Knight, *Immunity* **2014**, *40*, 824–832.
[5] G. Schneditz, J. Rentner, S. Roier, J. Pletz, K. A. T. Herzog, R. Bucker, H. Troeger, S. Schild, H. Weber, R. Breinbauer, et al., *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13181–13186.
[6] J.-P. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt, E. Oswald, *Science* **2006**, *313*, 848–851.
[7] P. Engel, M. I. Vizcaino, J. M. Crawford, *Appl. Environ. Microbiol.* **2015**, *81*, 1502–1512.
[8] D. Dubois, O. Baron, A. Cougnoux, J. Delmas, N. Pradel, M. Boury, B. Bouchon, M.-A. Bringer, J.-P. Nougayrède, E. Oswald, et al., *J. Biol. Chem.* **2011**, *286*, 35562–35570.
[9] A. Cougnoux, J. Delmas, L. Gibold, T. Faïs, C. Romagnoli, F. Robin, G. Cuevas-Ramos, E. Oswald, A. Darfeuille-Michaud, F. Prati, et al., *Gut* **2015**, DOI: 10.1136/gutjnl-2014–307241.
[10] D. Reimer, H. B. Bode, *Nat. Prod. Rep.* **2014**, *31*, 154–159.
[11] D. Reimer, K. M. Pos, M. Thines, P. Grün, H. B. Bode, *Nat. Chem. Biol.* **2011**, *7*, 888–890.
[12] C. A. Brotherton, E. P. Balskus, *J. Am. Chem. Soc.* **2013**, *135*, 3359–3362.
[13] X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang, R. Müller, *ChemBioChem* **2013**, *14*, 1194–1197.
[14] M. I. Vizcaino, P. Engel, E. Trautman, J. M. Crawford, *J. Am. Chem. Soc.* **2014**, *136*, 9244–9247.
[15] C. A. Brotherton, M. Wilson, G. Byrd, E. P. Balskus, *Org. Lett.* **2015**, *17*, 2294–2294.
[16] X. Bian, A. Plaza, Y. Zhang, R. Müller, *Chem. Sci.* **2015**, *6*, 3154–3160.
[17] M. I. Vizcaino, J. M. Crawford, *Nat. Chem.* **2015**, *7*, 411–417.
[18] Z.-R. Li, Y. Li, J. Y. H. Lai, J. Tang, B. Wang, L. Lu, G. Zhu, X. Wu, Y. Xu, P.-Y. Qian, *ChemBioChem* **2015**, DOI: 10.1002/cbic.201500239.
[19] Y. A. Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher, M. G. Thomas, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14349–14354.
[20] G. Sharon, N. Garg, J. Debelius, R. Knight, P. C. Dorrestein, S. K. Mazmanian, *Cell Metab.* **2014**, *20*, 719–730.
[21] A. O. Brachmann, H. B. Bode, *Adv. Biochem. Eng./Biotechnol.* **2013**, *135*, 123–155.

Received: June 11, 2015
Published online: July 15, 2015

