

Cite this: *Chem. Sci.*, 2015, 6, 3154

## Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety†

Xiaoying Bian,<sup>ab</sup> Alberto Plaza,<sup>a</sup> Youming Zhang<sup>\*b</sup> and Rolf Müller<sup>\*a</sup>

Colibactin represents a structurally undefined class of bacterial genotoxin inducing DNA damage and genomic instability in mammalian cells, thus promoting tumour development and exacerbating lymphopenia in animal models. The colibactin biosynthetic gene cluster (*clb*) has been known for ten years and it encodes a hybrid nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) assembly line. Nevertheless, the final chemical product(s) remain unknown. Previously, we and others reported several colibactin pathway-related metabolites including *N*-myristoyl-D-asparagine (**1**) as part of a prodrug precursor that is cleaved from the putative precolibactin to form active colibactin by the peptidase ClbP. Herein, we report two new colibactin pathway-related metabolites (**2** and **3**) isolated from a *clbP* mutant of the probiotic *E. coli* Nissle 1917 strain. Their structures were established by HRMS and NMR. Compound **2** shows an additional 4-aminopenitamic acid moiety with respect to **1**, while **3** is characterized by the presence of an unusual 7-methyl-4-azaspiro[2.4]hept-6-en-5-one residue. Moreover, we propose the biosynthetic pathway towards both intermediates on the basis of extensive gene inactivation and feeding experiments. The identification of **2** and **3** provides further insight into colibactin biosynthesis including the involvement and formation of a rare 1-aminocyclopropanecarboxylic acid unit. Thus, our work establishes additional steps of the pathway forming the bacterial genotoxin colibactin.

Received 11th January 2015

Accepted 23rd March 2015

DOI: 10.1039/c5sc00101c

www.rsc.org/chemicalscience

## Introduction

Human gut microbiota play an important role in health and disease.<sup>1,2</sup> As the predominant aero-anaerobic Gram-negative bacteria in the normal intestinal microbiota, most of the *Escherichia coli* are harmless or of benefit to the mammalian host and rarely cause disease. However, some virulent *E. coli* have acquired pathogenicity islands and can induce disease. Recently, it has been shown that many commensal and extra-intestinal pathogenic *E. coli* strains of the B2 phylogenetic group and closely related *Enterobacteriaceae* harbour a *pks* genomic island that is responsible for the production of a

bacterial genotoxin termed colibactin.<sup>3,4</sup> Colibactin-producing *E. coli* (*clb*+) induce DNA damage, cell cycle arrest, and genomic instability of mammalian cells *in vitro* and *in vivo*, and thus promote development of tumours, *e.g.* colorectal cancer, and exacerbate lymphopenia in animal models.<sup>3,5–11</sup> To reveal the molecular mechanism underlying the genotoxic activity of colibactin and to design drugs that inhibit its activity, information regarding colibactin's chemical structure, biosynthetic assembly and transfer to the host cell is required.<sup>12</sup>

The colibactin biosynthetic gene cluster (*clb*) was described about a decade ago and encodes a hybrid nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) assembly line as reported for many secondary metabolite biosynthetic pathways.<sup>3,13</sup> However, the product(s) of this pathway remain a mystery which is at least in part due to the fact that the cytopathic activity is contact-dependent and therefore neither the *clb*+ bacterial culture supernatant nor the bacterial lysate are cytopathic.

Although the probiotic commensal strain *E. coli* Nissle 1917 is widely used for the treatment of intestinal disorders, it surprisingly also contains a functional *clb* locus.<sup>14</sup> Unexpectedly, the product(s) of the colibactin pathogenicity island here possess anti-inflammatory activity and are required for the

<sup>a</sup>Department of Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Department of Pharmaceutical Biotechnology, Saarland University, Campus C2 3, 66123 Saarbrücken, Germany. E-mail: rolf.mueller@helmholtz-hzi.de; Tel: +49-681-30270201

<sup>b</sup>Shandong University-Helmholtz Institute of Biotechnology, State Key Laboratory of Microbial Technology, School of Life Science, Shandong University, Zhuzhou Road 168, 266101 Qingdao, P. R. China. E-mail: zhangyouming@sdu.edu.cn; Tel: +86-531-88363082

† Electronic supplementary information (ESI) available: Experimental methods, supplementary tables and figures including information on mutants and NMR spectra of compounds. See DOI: 10.1039/c5sc00101c



probiotic activity of *E. coli* Nissle 1917.<sup>15</sup> A possible explanation for these entirely different activities might be that colibactin may not be a single compound but rather a class of metabolites differentially expressed and formed under various culture conditions. Thereby the *clb* biosynthetic gene cluster may encode for the formation of additional compounds that could mediate the probiotic activity observed in the gut.<sup>15</sup> These intriguing but ambiguous results motivated our efforts to learn more about the chemical structure and the biosynthesis of colibactin(s).

Our previous genetic and comparative metabolomic efforts led to the identification of several colibactin pathway-dependent small molecules including *N*-myristoyl-D-asparagine (**1**, Fig. 1).<sup>16</sup> Compound **1** is believed to represent a fragment of a prodrug precursor cleaved from the putative precolibactin to form active colibactin by a peptidase ClbP (see Fig. 4Ba).<sup>16–19</sup> Such a prodrug activation mechanism was described recently in several natural product biosynthetic pathways, *e.g.* xenocoumarins,<sup>20</sup> zwittermicins,<sup>21,22</sup> and didemnins.<sup>23</sup> The *in vitro* biochemical characterization of ClbP and *in vivo* identification of **1** provided evidence for a prodrug mechanism during colibactin maturation.<sup>16,17</sup> More recently, a handful of colibactin pathway intermediates including **1** and **2** (Fig. 1) were identified from the wild type and/or *clbP* mutants of *E. coli* Nissle 1917 and *E. coli* IHE3034 by comparative metabolomics and targeted structural network analyses.<sup>24</sup> However, the structure of **2** was only proposed based on MS<sup>2</sup> fragmentation analysis.

Herein, we identify two new colibactin pathway-related products (**2** and **3**) isolated from the *clbP* mutant of *E. coli* Nissle 1917. Their structures were established by HRMS and NMR and their biosynthetic pathway(s) was proposed by a combination of gene inactivation and feeding experiments revealing further insight into colibactin formation.

## Results and discussion

We applied Red/ET recombineering to construct multiple single or double gene mutations/deletions in the native colibactin

producer *E. coli* Nissle 1917.<sup>25,26</sup> Each inactivation of the respective *clb* biosynthetic gene by replacement with an antibiotic resistance gene was performed as previously reported (see Experimental section, Tables S1 and S2, Fig. S1†).<sup>16</sup> The gene *clbL* encodes a putative amidase that might catalyse the hydrolysis of an amide in the putative precolibactin,<sup>3,24</sup> whereas *clbP* encodes a peptidase which cleaves **1** off from precolibactin.<sup>16–19</sup> Thus, *clbL* or *clbP* single deletion mutants were constructed to investigate the presence of colibactin-related products. Mutants and wild type strains were cultured, extracted, and analysed by LC-MS. We observed that the yield of the precursor fragment *N*-myristoyl-D-asparagine (**1**) is greatly decreased in the *clbP* mutant but slightly increased in the *clbL* single mutant (Fig. S2†). These results were consistent with our previous reports.<sup>16</sup>

Intriguingly, the LC-MS chromatograms of the *clbP* mutant showed additional metabolites when compared to the wild type and the *clbL* mutant (Fig. S2†). To shed light on their molecular structures, EtOAc extracts prepared from 20 L cultures of *E. coli* Nissle 1917  $\Delta clbP$  were chromatographed by sephadex LH-20 followed by reversed-phase HPLC to yield compounds **2** (4.9 mg) and **3** (0.7 mg).

The molecular formula of compound **2** was determined to be C<sub>23</sub>H<sub>43</sub>N<sub>3</sub>O<sub>5</sub> by HR-ESI-MS ( $m/z$  442.3273 [M + H]<sup>+</sup>). Its structure was characterized by 2D NMR and HRMS experiments and it exhibits an additional 4-aminopentanoic acid (App) residue in comparison to **1** (Fig. 1, NMR data see Table S3, Fig. S3–S6†). The structure of **2** was first proposed by Vizcaino *et al.* on the basis of MS fragmentation analysis but the compound was not isolated or characterized by NMR.<sup>24</sup> Our structural elucidation of **2** supports the hypothesis that the NRPS module (C–A–T, Fig. 4Ba) found on ClbB activates alanine and incorporates this residue into an *N*-myristoyl-D-asparagine precursor (**1**) as formed by ClbN (Fig. 4Ba). This process is followed by extension of the intermediate by the PKS module (KS–AT–KR–DH–ER–T) on ClbB by one malonyl-CoA unit including  $\beta$ -carbon processing (Fig. 4Ba), and subsequent hydrolytic release of **2** by a currently unknown mechanism. This hydrolysis pattern is also described in other colibactin pathway intermediates.<sup>24</sup>

HR-ESI-MS of **3** displayed an [M + H]<sup>+</sup> ion peak at  $m/z$  547.3853 consistent with the molecular formula C<sub>30</sub>H<sub>50</sub>N<sub>4</sub>O<sub>5</sub>. A mass consistent with **3** was detected by MS before,<sup>24</sup> but it was never isolated nor structurally characterized. Its <sup>1</sup>H NMR spectrum in comparison to that of **2** showed a few additional signals including an exchangeable NH proton at  $\delta$  8.53 (1H, s) and a methyl vinyl at  $\delta$  1.97 (3H, s) (see Table S4†). Comparison of the HSQC spectrum of **3** to that of **2** indicated that the signals corresponding to the methylene C-2,4-App ( $\delta_H$  2.14, 2.17,  $\delta_C$  30.4) in **2** were shifted downfield in **3** ( $\delta_H$  2.92, 2.83;  $\delta_C$  38.5). Additionally, analysis of <sup>13</sup>C NMR, HSQC, and DET 135 spectra of **3** indicated the presence of two equivalent methylene groups ( $\delta_H$  1.49,  $\delta_C$  13.6, and  $\delta_H$  1.41,  $\delta_C$  13.7) that were absent in **2**. A detailed analysis of the 2D NMR data (HSQC, HMBC, DQF-COSY, ROESY) of **3** showed that the C-terminus of the fragment *N*-myristoyl-asparagine was connected to a 4-aminopentanoyl (4-Appn) residue. The structure of the remaining C<sub>7</sub>H<sub>8</sub>NO fragment was assembled as follows: HMBC correlations from the

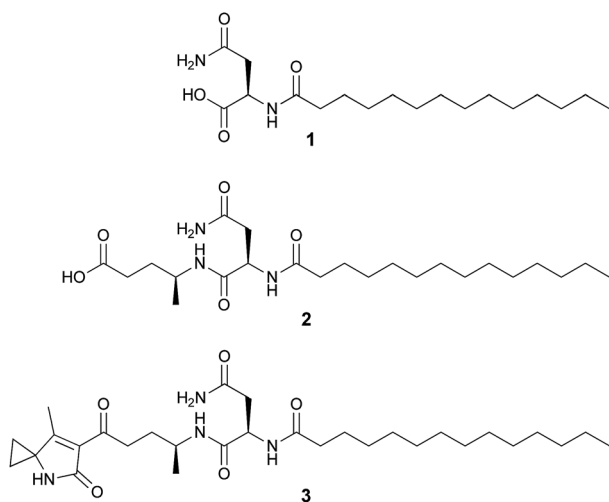


Fig. 1 Structures of colibactin pathway-related metabolites **1**–**3**.



methyl signal at  $\delta$  1.97 (Me-7<sub>Azh</sub>) to the quaternary carbon resonance at  $\delta$  45.3 (C-3<sub>Azh</sub>) and to the olefinic carbon resonances at  $\delta$  169.5 (C-7<sub>Azh</sub>) and 128.9 (C-6<sub>Azh</sub>), together with HMBC correlations from the amide proton at  $\delta$  8.53 (NH-4<sub>Azh</sub>) to C-3<sub>Azh</sub>, C-6<sub>Azh</sub>, and to the carbonyl resonance at  $\delta$  169.2 (C-5<sub>Azh</sub>), determined the presence of a 4-methyl-1,5-dihydro-2H-pyrrol-2-one ring (see Fig. 2). Furthermore, long-range correlations from the methylene protons at  $\delta$  1.49 (H-1<sub>Azh</sub>) and 1.41 (H-2<sub>Azh</sub>) to C-3<sub>Azh</sub> and C-7<sub>Azh</sub>, COSY correlations between H-1<sub>Azh</sub> and H-2<sub>Azh</sub>, together with ROESY correlations from H-1<sub>Azh</sub> to Me-7<sub>Azh</sub> and from H-2<sub>Azh</sub> to H-4<sub>Azh</sub> indicated that a cyclopropane was spiro fused to the dihydropyrrolone ring forming a 7-methyl-4-azaspiro[2.4]hept-6-en-5-one residue (Azh). Finally, connectivity between Azh and 4-Apn residues was deduced from four-bond HMBC correlations from Me-7<sub>Azh</sub> and NH-4<sub>Azh</sub> to the ketone carbon at  $\delta$  197.7 (C-1<sub>4-Apn</sub>) thereby completing the structure of **3** as depicted in Fig. 1 (NMR spectra see Fig. S7–S15†). Tandem mass spectrometry provided further evidence to support the structure of **3** (Fig. S16†). MS<sup>2</sup> fragmentation of the ion peak at  $m/z$  547 displayed ions at  $m/z$  530 [ $M + H - NH_3$ ]<sup>+</sup>, 320 [ $M + H - NH_2 - Myr$ ]<sup>+</sup>, and 205 [ $M + H - Myr - Asn - NH_2$ ]<sup>+</sup>. Thus the MS<sup>2</sup> fragmentation patterns in complete agreement with the structure of **3** determined by NMR.

The absolute configuration of D-Asn was determined by MS detected chromatographic comparison with appropriate amino acid standards of the acid hydrolysate of **2** and **3** after derivatizing with L- and D-FDLA (L-fluoro-2,4-dinitro-phenyl-5-L/D-leucinamide).<sup>27,28</sup> Similarly, the absolute configuration of C-4 of App in **2** was established as *S*. The *S* configuration of C-4<sub>App</sub> in **3** was assumed as *S* on the basis the structural and NMR data similarities to those of **2**.

To investigate the question which gene(s) from the *clb* pathway are responsible for biosynthesis of the 7-methyl-4-azaspiro[2.4]hept-6-en-5-one moiety in **3**, eleven mutations were performed on the *clbP* mutant of *E. coli* Nissle 1917 by one more step of Red/ET recombination. Each target gene was replaced by a kanamycin resistance gene (*Km<sup>R</sup>*) for *clbC* to *clbK* and *clbO*, whereas inactivation of *clbL* was performed by replacement with a spectinomycin resistance gene (*Spect<sup>R</sup>*) (Table S1†). All mutants were verified by colony PCR (data now shown). LC-MS analyses of the *clbP* mutant and the double mutants showed that inactivation of *clbC*, *clbH* or *clbI* resulted in abolishment of the production of **3** indicating that these three genes were involved in the biosynthesis of **3** (Fig. 3). On the other hand, gene inactivation of each of *clbD*, *clbE*, *clbF*, *clbG*, *clbJ*, *clbK* or *clbO* had no significant effect on the production of **3** (Fig. 3). The

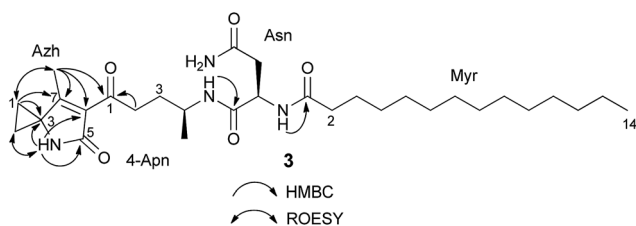


Fig. 2 NMR-based connectivity of the fragments of **3**.

*clbL/clbP* double mutant showed a great decrease of **3** implying that the putative peptide amidase ClbL might be involved in hydrolysis of **3** from a larger precolibactin molecule. Unfortunately, we were not able to detect any larger colibactin-pathway dependent metabolites in these mutants.

According to the gene inactivation experiments and bioinformatics analysis of the modular organization of the pathway, a proposal for the biosynthesis of **3** is shown in Fig. 4B. Initially the presence of the 7-methyl-4-azaspiro[2.4]hept-6-en-5-one residue in **3** seemed difficult to rationalize biosynthetically. However, when taking into account the possibility of incorporation of a nonproteinogenic 1-aminocyclopropane-carboxylic

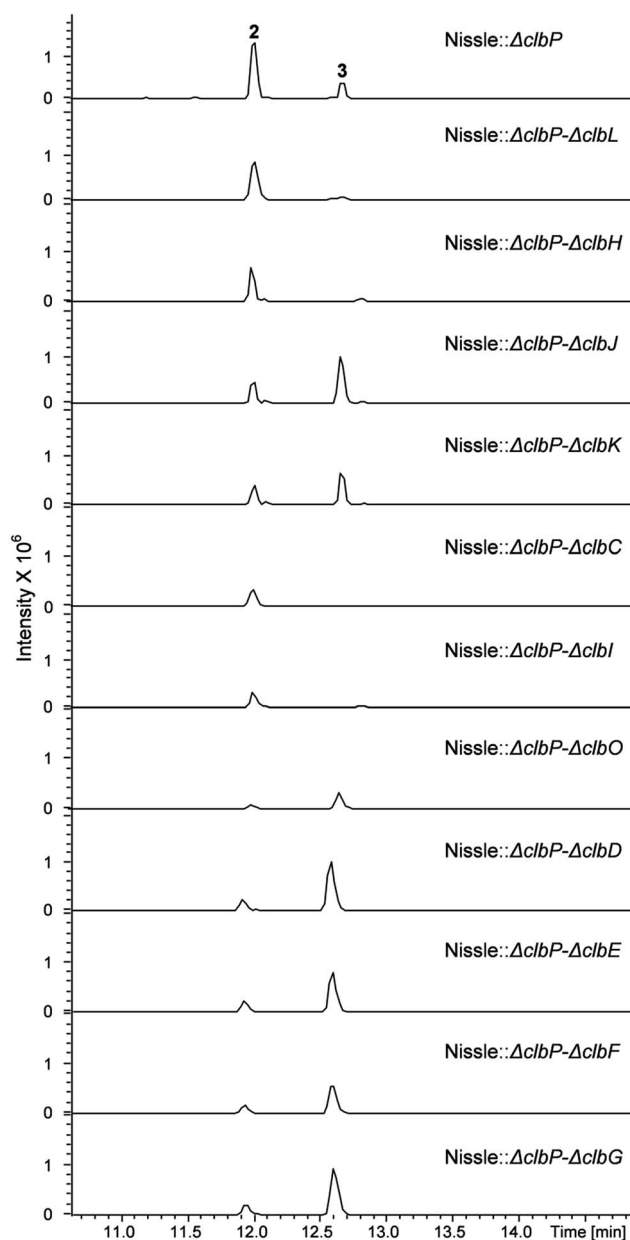
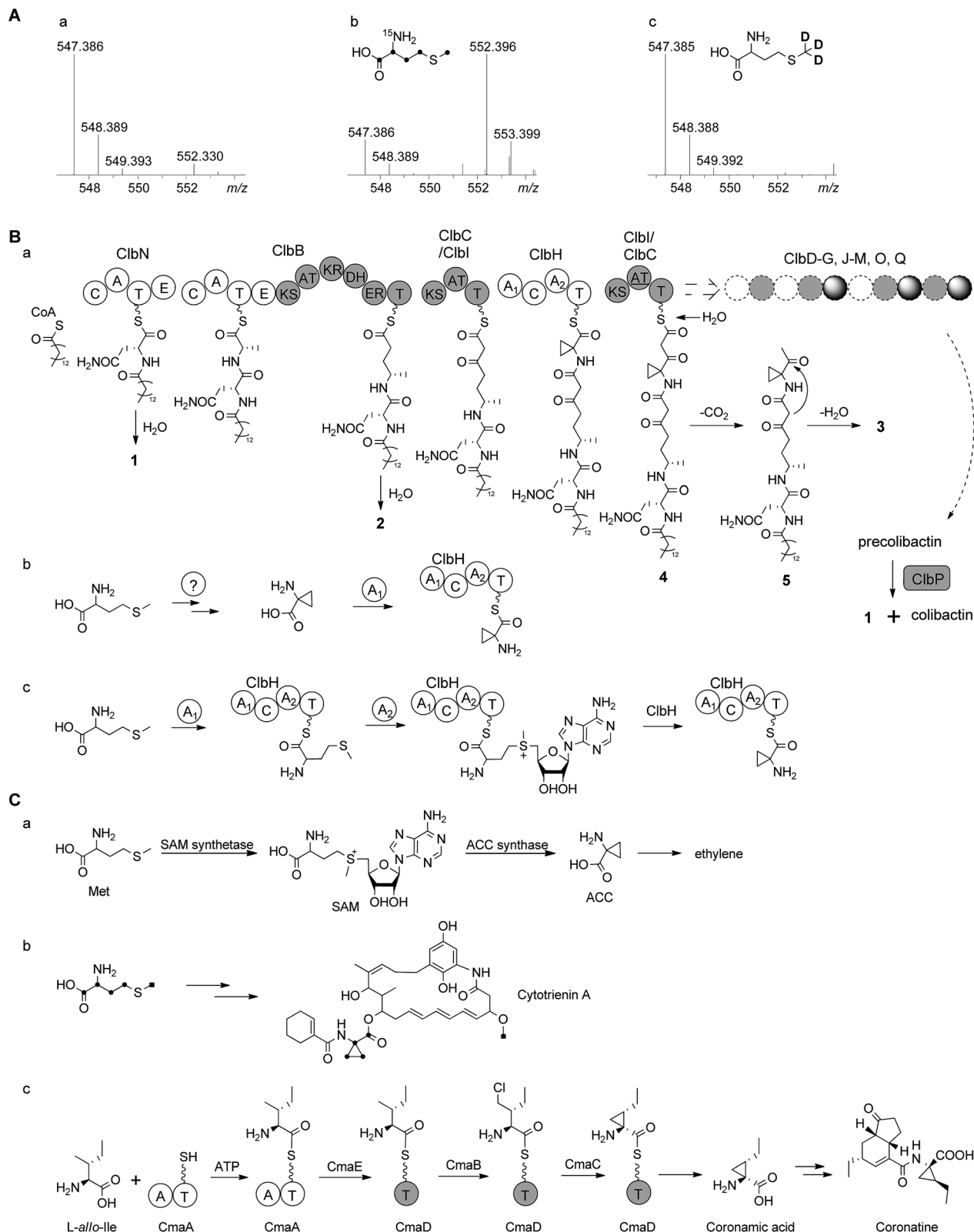


Fig. 3 UPLC-HRMS analysis of the *E. coli* Nissle double mutants. Shown is a base peak chromatogram (BPC)  $m/z$  for 442.3 [ $M + H$ ]<sup>+</sup> and 547.4 [ $M + H$ ]<sup>+</sup> of compounds **2** and **3**, respectively.



**Fig. 4** (A) HR-ESI-MS spectra of **3** obtained from feeding experiments. (a) Control; (b) after feeding of  $L$ -[ $^{13}\text{C}$ ,  $^{15}\text{N}$ ]methionine (only four carbons and one nitrogen are found in the final product, Fig. S17†); (c) after feeding of  $L$ -[methyl- $^2\text{H}_3$ ]methionine (no  $^2\text{H}$  incorporated) showing that the aminobutyryl moiety of methionine (without  $S$ -methyl) is incorporated into the final product **3**. (B) (a) Proposed biosynthetic pathway of colibactin up to intermediate **4** involving **1** and **2** tethered to the assembly line. After release of **4** and decarboxylation of the  $\beta$ -keto-acid, **3** is formed by intramolecular cyclization analogous to Dieckmann condensation. (b) A hypothesis for ACC biosynthesis and then loading on ClbH; (c) alternative hypothesis for loading of methionine onto ClbH and then conversion to ACC on the T domain catalyzed by  $A_2$  domain. (C) Biosynthesis of ACC moiety in other organisms. (a) Biosynthesis of ACC in the ethylene pathway in higher plants;<sup>30,31</sup> (b) incorporation of  $L$ -[ $^{13}\text{C}$ ]



acid (ACC) residue in conjunction with hydrolysis and possible shunt product formation the following scenario was developed: compound **2** undergoes extension by one PKS module (encoded by ClbC or ClbI; KS-AT-T) to incorporate one malonyl-CoA unit, one NRPS module (ClbH, A<sub>1</sub>-C-A<sub>2</sub>-T) is needed for incorporation of the unusual ACC, and one more PKS module (encoded by ClbI or ClbC; KS-AT-T) is needed to incorporate an additional malonyl-CoA unit eventually giving rise to **4** attached to the biosynthetic assembly line (Fig. 4Ba). We hypothesize that **4** might be hydrolysed enzymatically (e.g. TE catalysed, like **2**) to form the corresponding free acid undergoing decarboxylation to build **5**, which in turn can engage in an intramolecular cyclization analogous to a Dieckmann condensation forming the dihydropyrrolone ring found in **3** (Fig. 4Ba). A similar Dieckmann cyclization catalyzed by a Dieckmann cyclization domain (DKC) was proposed to form the pyrrolone ring in pyranonigrin **E**,<sup>29</sup> but no homologous DKC domain was found in the *clb* gene cluster leading to the speculation that the condensation in **3** may occur spontaneously.

The ACC moiety is quite an unusual moiety in bacterial secondary metabolites, and to the best of our knowledge this is the first occurrence in an *E. coli* natural product. However, it is well known from plant metabolites such as in the biosynthesis of the higher plant hormone ethylene where ACC serves as key intermediate,<sup>30</sup> ACC is synthesized from *S*-adenosylmethionine (SAM) via the pyridoxal phosphate (PLP)-dependent ACC synthase catalysing a  $\gamma$ -elimination reaction (Fig. 4Ca).<sup>30,31</sup> The ansa-bridged cytrotienin from the soil bacterium *Streptomyces* sp. RK95-74 also contains an ACC moiety<sup>32</sup> and stable isotope feeding experiments showed that the ACC moiety in this producer is also derived from *L*-methionine (Fig. 4Cb),<sup>33</sup> but the cytrotienin biosynthetic gene cluster has not been published yet. An alternative strategy was described for the enzymatic formation of the ethyl-substituted ACC moiety coronamic acid (CMA) in the phytotoxin coronatine produced by *Pseudomonas syringae* strains. Here, biosynthesis follows a cryptic chlorination pathway relying on an aminoacyl thioester intermediate involving five dedicated proteins CmaABCDE (Fig. 4Cc).<sup>34</sup> CmaB and CmaC execute  $\gamma$ -halogenation of *L*-allo-isoleucine followed by intramolecular  $\gamma$ -elimination to form the cyclopropyl ring.<sup>34,35</sup> As no homologues of ACC synthase genes or coronamic acid biosynthesis genes were found in the colibactin gene cluster, we performed feeding experiments to investigate the hypothetical biosynthesis of ACC in **3**. Methionine labelled by stable isotopes was added to the growth medium resulting in the following incorporation patterns in **3**: feeding of *L*-[U-<sup>13</sup>C, <sup>15</sup>N]methionine gave *m/z* 552.3959 [M + H]<sup>+</sup> (5 amu more than 547.3853) corresponding to the chemical formula of **3**-<sup>13</sup>C<sub>4</sub>, <sup>15</sup>N (C<sub>26</sub><sup>13</sup>C<sub>4</sub>H<sub>50</sub>N<sub>3</sub><sup>15</sup>NO<sub>5</sub>, *m/z* 552.3959 [M + H]<sup>+</sup>) while feeding of *L*-[methyl-<sup>2</sup>H<sub>3</sub>]methionine did not result in deuteration of **3** (Fig. 4A, Fig. S17A†). Taken together, these data indicated that

the aminobutyl moiety (accounting for four carbons and one nitrogen without the *S*-methyl carbon) of *L*-methionine was introduced into the ACC unit of **3** which suggests that ACC originates from *L*-methionine in *E. coli* (Fig. 4AB, Fig. S17B†). Nonetheless, the *L*-methionine-ACC pathway including ACC synthase has to the best of our knowledge not been detected in prokaryotic microorganisms.<sup>33,36</sup> As stated above we did not find a homolog of the ACC synthase gene in the *E. coli* Nissle 1917 *clb* gene cluster. An additional search for homologs in the genome of *E. coli* Nissle 1917 (GenBank: CP007799.1) using ACC synthases from apple (*Malus domestica*) (GenBank: AAB68617.1, PDB: 1B8G)<sup>37</sup> and from *Penicillium citrinum* (GenBank: AB038512)<sup>38</sup> as queries was performed using BlastP.<sup>39</sup> This resulted in the identification of some homologs belonging to the family of PLP-dependent aminotransferases/transaminases/lyases exhibiting low identities on the amino acid level (<30%, Table S5†). As either one of these proteins or any other ORF of unknown function encoded in the *E. coli* genome may be responsible for the conversion of SAM to ACC, clearly further *in vivo* gene inactivation and *in vitro* biochemical experiments are required to study the formation of ACC in *E. coli*.

Intriguingly, the NRPS module encoded on ClbH hypothesized to be in charge of ACC incorporation contains two A domains; the first A domain (A<sub>1</sub>) is predicted to activate a serine residue according to the specificity-conferring code of adenylation domains in NRPSs (Table S6†).<sup>40,41</sup> A subsequent feeding experiment employing *L*-[2,3,3-<sup>2</sup>H<sub>3</sub>]serine led to an isotopic pattern compatible with incomplete incorporation of serine into **2** and **3** (Fig. S18A†). This finding may be explained by serine degradation via pyruvate to acetyl-CoA and malonyl-CoA for incorporation into **2** and **3** by the PKS or fatty acid synthetase (FAS) which may be responsible for the fatty acid side chain (Fig. S18B†). Thus, we assume that the A<sub>1</sub> domain is not required for serine activation but may play a role in ACC formation as discussed below. The second A domain (A<sub>2</sub>) of ClbH is predicted to be specific for a hydrophobic/aromatic residue (phenylalanine/tryptophan/tyrosine/valine/leucine) or may represent an A domain with novel specificity (Table S6†). Feeding of *L*-[U-<sup>2</sup>H]valine, *L*-[U-<sup>2</sup>H]leucine, *L*-[U-<sup>13</sup>C, <sup>15</sup>N]phenylalanine and *L*-[ring-<sup>2</sup>H<sub>4</sub>]tyrosine did not show any incorporation into **2** or **3** (data not shown). The nonribosomal code of the A<sub>2</sub> domain of ClbH showed that the strictly conserved aspartic acid (D235) interacting with the  $\alpha$ -amino group of amino acids in peptide synthetases is mutated to alanine (Table S6†), indicating that the substrate of the A<sub>2</sub> domain may not have an  $\alpha$ -amino group.<sup>42</sup> Thus, according to the chemical structure of **3**, we speculate that the A<sub>1</sub> domain may load an ACC unit (Fig. 4Bb) or methionine (Fig. 4Bc) onto ClbH. The former hypothesis regarding loading of ACC by the A<sub>1</sub> domain would lead to the requirement of ACC biosynthesis in *E. coli* and make the A<sub>2</sub> domain 'superfluous' (Fig. 4Bb). The latter hypothesis

methionine into cytrotienin A showed that the ACC unit originates from methionine in *Streptomyces* sp. RK95-74;<sup>33</sup> (c) biosynthetic pathway of the coronamic acid (CMA) moiety of coronatine in *P. syringae* strains.<sup>34,35</sup> C, condensation domain; A, adenylation domain; T, thiolation domain; E, epimerization domain; KS, ketosynthase domain; AT, acyltransferase domain; DH, dehydratase domain; KR, ketoreductase domain; ER, enoyl reductase domain; SAM, *S*-adenosylmethionine; ACC, 1-aminocyclopropanecarboxylic acid.



postulates that ClbH might function as a SAM synthetase after methionine loading by the A<sub>1</sub> domain to the T domain of ClbH. Subsequently, ClbH may also function as an ACC synthase (Fig. 4Bc). According to this hypothesis the SAM synthetase and ACC synthase activity may be encoded within the NRPS ClbH. In light of this hypothesis the biosynthetic function of each domain found on ClbH (A<sub>1</sub>–C–A<sub>2</sub>–T) during loading, modification and ACC incorporation needs to be characterized in upcoming experiments in detail. To achieve this goal the whole protein and/or the respective domains need to be expressed in their active forms, but this is a goal for future experiments.

We measured the *in vitro* bacterial growth inhibitory activity of 2 and 3, as 1 has been reported to exhibit mild growth inhibitory activity against *Bacillus subtilis* NCIB 3610.<sup>24</sup> Both, 2 and 3, also showed weak growth inhibitory activities (EC<sub>50</sub> 30–70 µg mL<sup>−1</sup>) against *E. coli* (TolC-deficient) and *B. subtilis* DSM-10 (Fig. S19†).

According to our gene inactivation experiments, further assembly of precolibactin and colibactin most likely requires additional proteins encoded on the *clb* pathogenicity island. Whereas this study provides evidence on the function of ClbC, ClbH and ClbI *in vivo* on route to intermediate 4, it does not determine any activity for ClbDEFG, ClbJ, ClbK and ClbO which are clearly not required for the formation of 3 according to the mutagenesis experiments reported here. Consequently, further experiments *in vitro* and/or *in vivo* are required not only to determine the structure of colibactin and precolibactin but also to decipher each biosynthetic step in greater detail.

## Conclusions

Colibactin represents a chemically unknown class of bacterial genotoxin present in most species of the B2 group of *E. coli*. Its presence in the probiotic *E. coli* Nissle 1917 not only contributes to probiotic activity but also induces chromosomal instability of mammalian cells. With the aim to eventually determine the structure of colibactin, two colibactin pathway-related metabolites were identified by HRMS and NMR analyses from *E. coli* Nissle 1917 after extensive mutagenesis studies. Compound 3 is a linear lipopeptide containing a novel 7-methyl-4-azaspiro[2.4] hept-6-en-5-one (Azh) residue and its biosynthetic pathway was proposed based on gene inactivation, *in silico* analysis of biosynthetic pathway proteins and feeding experiments. Surprisingly, the latter indicated that the uncommon non-proteinogenic amino acid ACC, which is involved in the formation of Azh, might be derived from L-methionine in *E. coli* and possible scenarios for its biosynthesis during colibactin assembly are discussed. The structural elucidation and the biosynthetic proposal regarding these two colibactin pathway-dependent biosynthetic intermediates (2 and 3) provides additional knowledge required on route to eventually establish the structure of colibactin and thus deepens our understanding of this intriguing probiotic and genotoxic compound class.

## Acknowledgements

The authors thank Dr Thomas Hoffmann, Suvd Nadmid and Fu Yan from R.M.'s laboratory for HRMS measurements and

experimental assistance. Research in the laboratory of R.M. was funded by the Deutsche Forschungsgemeinschaft (DFG) and the Bundesministerium für Bildung und Forschung (BMBF). Y.Z. is supported by Program for the Recruitment Program of Global Experts in Shandong University. The authors acknowledge Dr Kevin Sours's help in proofreading this manuscript.

## Notes and references

- 1 I. Cho and M. J. Blaser, *Nat. Rev. Genet.*, 2012, **13**, 260.
- 2 N. Kamada, S. U. Seo, G. Y. Chen and G. Nunez, *Nat. Rev. Immunol.*, 2013, **13**, 321.
- 3 J. P. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt and E. Oswald, *Science*, 2006, **313**, 848.
- 4 J. Putze, C. Hennequin, J. P. Nougayrède, W. Zhang, S. Homburg, H. Karch, M. A. Bringer, C. Fayolle, E. Carniel, W. Rabsch, T. A. Oelschlaeger, E. Oswald, C. Forestier, J. Hacker and U. Dobrindt, *Infect. Immun.*, 2009, **77**, 4696.
- 5 G. Cuevas-Ramos, C. R. Petit, I. Marcq, M. Boury, E. Oswald and J. P. Nougayrède, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 11537.
- 6 J. C. Arthur, E. Perez-Chanona, M. Muhlbauer, S. Tomkovich, J. M. Uronis, T. J. Fan, B. J. Campbell, T. Abujamel, B. Dogan, A. B. Rogers, J. M. Rhodes, A. Stintzi, K. W. Simpson, J. J. Hansen, T. O. Keku, A. A. Fodor and C. Jobin, *Science*, 2012, **338**, 120.
- 7 A. Cougnoux, G. Dalmasso, R. Martinez, E. Buc, J. Delmas, L. Gibold, P. Sauvanet, C. Darcha, P. Dechelotte, M. Bonnet, D. Pezet, H. Wodrich, A. Darfeuille-Michaud and R. Bonnet, *Gut*, 2014, **63**, 1932.
- 8 I. Marcq, P. Martin, D. Payros, G. Cuevas-Ramos, M. Boury, C. Watrin, J. P. Nougayrède, M. Olier and E. Oswald, *J. Infect. Dis.*, 2014, **210**, 285.
- 9 T. Secher, A. Samba-Louaka, E. Oswald and J. P. Nougayrède, *PLoS One*, 2013, **8**, e77157.
- 10 J. C. Arthur and C. Jobin, *Gut Microb.*, 2013, **4**, 253.
- 11 G. Dalmasso, A. Cougnoux, J. Delmas, A. Darfeuille-Michaud and R. Bonnet, *Gut Microb.*, 2014, **5**, 675.
- 12 Y. Yang and C. Jobin, *Dis. Models & Mech.*, 2014, **7**, 1131.
- 13 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468.
- 14 S. Homburg, E. Oswald, J. Hacker and U. Dobrindt, *FEMS Microbiol. Lett.*, 2007, **275**, 255.
- 15 M. Olier, I. Marcq, C. Salvador-Cartier, T. Secher, U. Dobrindt, M. Boury, V. Bacquie, M. Penary, E. Gaultier, J. P. Nougayrède, J. Fioramonti and E. Oswald, *Gut Microb.*, 2012, **3**, 501.
- 16 X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang and R. Müller, *ChemBioChem*, 2013, **14**, 1194.
- 17 C. A. Brotherton and E. P. Balskus, *J. Am. Chem. Soc.*, 2013, **135**, 3359.
- 18 D. Dubois, O. Baron, A. Cougnoux, J. Delmas, N. Pradel, M. Boury, B. Bouchon, M. A. Bringer, J. P. Nougayrède, E. Oswald and R. Bonnet, *J. Biol. Chem.*, 2011, **286**, 35562.



- 19 A. Cougnoux, L. Gibold, F. Robin, D. Dubois, N. Pradel, A. Darfeuille-Michaud, G. Dalmasso, J. Delmas and R. Bonnet, *J. Mol. Biol.*, 2012, **424**, 203.
- 20 D. Reimer, K. M. Pos, M. Thines, P. Grün and H. B. Bode, *Nat. Chem. Biol.*, 2011, **7**, 888.
- 21 B. M. Kevany, D. A. Rasko and M. G. Thomas, *Appl. Environ. Microbiol.*, 2009, **75**, 1144.
- 22 Y. Luo, L. F. Ruan, C. M. Zhao, C. X. Wang, D. H. Peng and M. Sun, *Antimicrob. Agents Chemother.*, 2011, **55**, 4161.
- 23 Y. Xu, R. D. Kersten, S. J. Nam, L. Lu, A. M. Al-Suwailem, H. J. Zheng, W. Fenical, P. C. Dorrestein, B. S. Moore and P. Y. Qian, *J. Am. Chem. Soc.*, 2012, **134**, 8625.
- 24 M. I. Vizcaino, P. Engel, E. Trautman and J. M. Crawford, *J. Am. Chem. Soc.*, 2014, **136**, 9244.
- 25 Y. Zhang, F. Buchholz, J. P. Muyrers and F. A. Stewart, *Nat. Genet.*, 1998, **20**, 123.
- 26 J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Müller, A. F. Stewart and Y. Zhang, *Nat. Biotechnol.*, 2012, **30**, 440.
- 27 K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki and K. Harada, *Anal. Chem.*, 1997, **69**, 3346.
- 28 K. Fujii, Y. Ikai, H. Oka, M. Suzuki and K. Harada, *Anal. Chem.*, 1997, **69**, 5146.
- 29 T. Awakawa, X. L. Yang, T. Wakimoto and I. Abe, *ChemBioChem*, 2013, **14**, 2095.
- 30 K. L. C. Wang, H. Li and J. R. Ecker, *Plant Cell*, 2002, **14**, S131.
- 31 W. K. Yip, J. G. Dong, J. W. Kenny, G. A. Thompson and S. F. Yang, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 7930.
- 32 H. P. Zhang, H. Kakeya and H. Osada, *Tetrahedron Lett.*, 1997, **38**, 1789.
- 33 H. P. Zhang, H. Kakeya and H. Osada, *Tetrahedron Lett.*, 1998, **39**, 6947.
- 34 F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. E. O'Connor and C. T. Walsh, *Nature*, 2005, **436**, 1191.
- 35 W. L. Kelly, M. T. Boyne, E. Yeh, D. A. Vosburg, D. P. Galonic, N. L. Kelleher and C. T. Walsh, *Biochemistry*, 2007, **46**, 359.
- 36 M. Ueki, D. P. Galonic, F. H. Vaillancourt, S. Garneau-Tsodikova, E. Yeh, D. A. Vosburg, F. C. Schroeder, H. Osada and C. T. Walsh, *Chem. Biol.*, 2006, **13**, 1183.
- 37 G. Capitani, E. Hohenester, L. Feng, P. Storici, J. F. Kirsch and J. N. Jansonius, *J. Mol. Biol.*, 1999, **294**, 745.
- 38 Y. Kakuta, T. Igarashi, T. Murakami, H. Ito, H. Matsui and M. Honma, *Biosci., Biotechnol., Biochem.*, 2001, **65**, 1511.
- 39 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389.
- 40 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211.
- 41 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493.
- 42 M. A. Marahiel, T. Stachelhaus and H. D. Mootz, *Chem. Rev.*, 1997, **97**, 2651.

