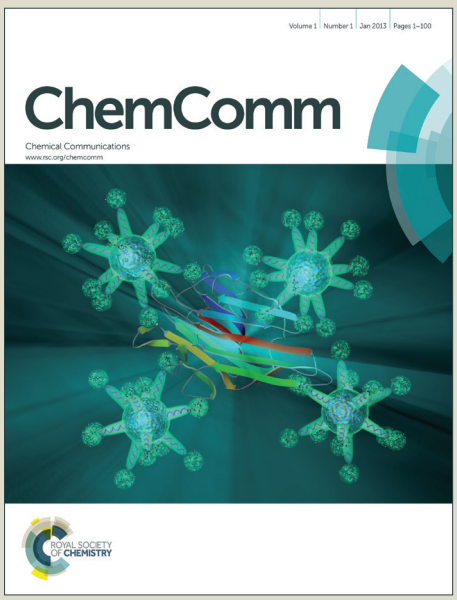
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Colibactin biosynthesis and biological activity depends on the rare aminomalonyl polyketide precursor†

Cite this: DOI: 10.1039/x0xx00000x

A. O. Brachmann,[a] C. Garcie,[b] V. Wu,[a] P. Martin,[b] R. Ueoka,[a] E. Oswald[b] and J. Piel[a]

Received 00th January 2012,
Accepted 00th January 2012

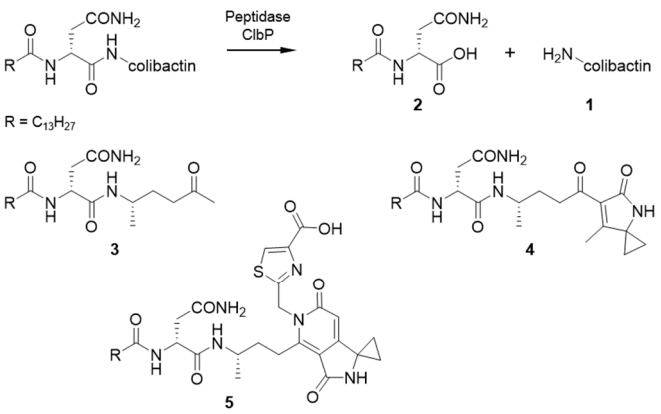
DOI: 10.1039/x0xx00000x

www.rsc.org/

The as-yet unidentified *E. coli* metabolite colibactin induces DNA damages in eukaryotic cells and promotes tumorigenesis. Its wide distribution in pathogenic and probiotic strains has raised great interest in its structure and biosynthetis. Here we show that colibactin formation involves a rare aminomalonyl unit used as building block.

The cryptic natural product colibactin (**1**) has received much attention due to its association with human disease, widespread occurrence, and elusive identity. In various pathogenic *E. coli* strains, the colibactin biosynthetic genes (termed *clb* or *pks*) are positively correlated with DNA double-strand breaks in host cells,[1] colon inflammation and damages in mammalian epithelial tissues.[2] They are also involved in cellular senescence accompanied by tumor promotion in colitis-associated colorectal cancer.[3] The *clb* gene cluster encodes a hybrid nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS) system and is widely distributed among different enterobacteria,[4] including coral[5] and honeybee symbionts.[6] Considering the toxic effect in pathogens, it is remarkable that the probiotic *E. coli* Nissle 1917 strain, which also carries the *clb* genes, is being used extensively for treatment of inflammatory bowel disease.[7]

Despite substantial efforts by several laboratories to identify **1**,[1b,5,7-8] its structure remains unknown. First insights into the biosynthesis were provided by the identification of ClbP as a membrane bound peptidase that is essential in maturation of a hypothetical pre-colibactin to induce genotoxicity.[8a] Based on precedence from other pathways,[9] biochemical experiments,[8c] and *in vivo* studies,[8b] the prodrug activation mechanism should involve removal of an *N*-fatty acyl-D-asparaginyl moiety by the peptidase, releasing the active compound (Scheme 1). Recently, MS-guided identification of intermediate-derived metabolites **2**-**5** revealed further structural features.[8d-8h] **4** and **5** contain an unusual 1-aminocyclopropane-1-carboxylic acid (ACC) unit that contributes to the instability of **1** and also the mechanism of action.[8e,8g] Comparison of **4** and thiazol-containing **5** with the NRPS-PKS architecture (Fig. S1) suggests that two further polyketide synthase modules (ClbO and first module of ClbK) are involved in incorporation of uncharacterized polyketide building blocks.

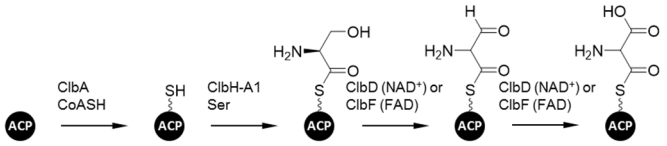


Scheme 1 Peptidase activation mechanism for colibactin (top) and detected metabolites **2**-**5** from *E. coli* cultures.

Here we present biochemical data on eight proteins of the biosynthetic pathway. The results suggest that the colibactin pathway employs, in addition to the amino acids glycine and cysteine, a rare 2-aminomalonyl (AM) precursor. Based on MS studies and bioactivity profiles of mutant strains, the AM building block likely belongs to an as-yet unknown portion of the colibactin structure.

Initial experiments were motivated by bioinformatic analysis of the *clb* cluster in the meningitis-associated strain *E. coli* IHE3034 (Fig. S1) that suggested homology of ClbD and ClbF to dehydrogenases (Table S1). Close relatives are ZmaG and ZmaI from zwittermicin (*zma*) biosynthesis, which were shown to jointly catalyze the formation of the unusual PKS extender AM-ACP from seryl-ACP, indicating related functions in the *clb* pathway (Scheme 2).[10] Interestingly, zwittermicin biosynthesis also features a similar

prodrug activation mechanism as proposed for **1**.[10a] In the *zma* pathway, seryl-ACP is generated by an isolated adenylation enzyme that loads a cognate, free-standing ACP. A close *clb* homolog of this ACP was identified as ClbE. The only matching adenylation enzyme with predicted[11] serine specificity, however, was an A domain (ClbH-A1) at the N terminus of the unusual NRPS protein ClbH. In contrast to the canonical NRPS architecture, ClbH lacks a peptidyl carrier protein (PCP) domain downstream of the A domain, but is directly connected to a condensation (C) domain (overall architecture A-C-A-PCP). Resequencing of the NRPS region confirmed the correct architecture of ClbH.



Scheme 2 Proposed formation of aminomalonyl-ACP and role of colibactin biosynthetic enzymes.

To biochemically determine the amino acid substrate of ClbH-A1, we expressed and characterized the domain using a previously described γ-[$^{18}O_4$]-ATP pyrophosphate exchange assay.[12] In addition to proteinogenic amino acids, we also tested several non-standard amino acids including ACC and *S*-adenosyl- L-methionine (SAM). These were previously considered together with L-Met as possible substrates for either ClbH-A1 or ClbH-A2, the second A domain on ClbH, to account for the presence of ACC in **4** and **5**.[8f-g] However, among all tested amino acids, significant conversion was observed only for L-Ser (Fig. 1) in agreement with the prediction. We measured low exchange rates for ACC, but previous feeding experiments suggest that this compound is not directly incorporated[8g] and should instead be derived from NRPS-activated L-Met or SAM.[8f-g] Our data show that these two amino acids are unlikely substrates for ClbH-A1. We also expressed and assayed ClbH-A2, but did not detect significant exchange rates with any test substrate, including L-Met, SAM, and ACC (Fig. S2c). In conclusion, our data support a role of ClbH-A1 in the provision of a seryl rather than an ACC unit.

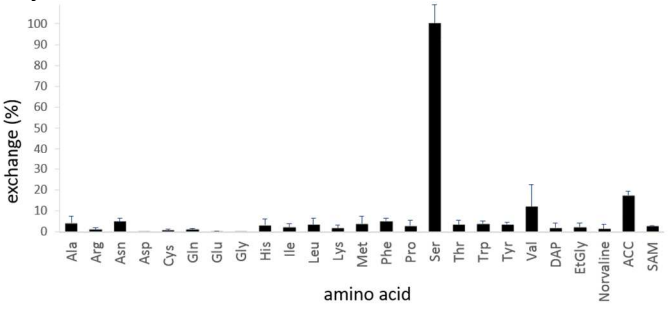


Fig. 1 Exchange rates for the A domain ClbH-A1 and various amino acids. EtGly, ethylglycine; DAP, 2,3-diaminopropionic acid.

We next investigated whether ClbH-A1 in combination with other enzymes is also able to provide AM-ACP. As additional components we individually expressed the ACP ClbE, predicted dehydrogenases ClbD and ClbF, and the cluster-encoded PPTase ClbA. The PPTase was suspected to transform the *apo*-ACP into the active *holo*-ACP similar to other thiotemplate pathways.[13] The PPTase activity was assayed in mixtures containing ClbA, ClbE, and the substrate coenzyme A. Analysis by liquid chromatography-high-resolution MS (LC-HRMS) showed the appearance of a new species with a characteristic mass shift of +340 Da as compared to unmodified ClbA, consistent with a phosphopantetheinyl moiety attached to the ACP (Fig. S3). ClbA was previously shown to replace the siderophore-associated NRPS PPTase EntD in enterobactin biosynthesis.[14] The activation of ClbE suggests that ClbA belongs to the promiscuous PPTases that can activate PCP as well as ACP domains.

With the *holo*-ACP in hand, we tested whether it is loaded with L-Ser after addition of ClbH-A1. The formation of seryl-ACP was supported by a mass shift of +87 Da for one of the major ACP species (Fig. 2a and S4). This interpretation was also corroborated by detecting an additional mass shift of +3 Da after using $^{13}C_3$-labeled L-Ser (Fig. S5). Further oxidation to AM-ACP was tested by adding both dehydrogenases ClbF and ClbD and their respective cofactors NAD^+ and FAD (Fig. 2). Here, additional ions were only observed when both dehydrogenases were added (Fig. 2e, Table S2), whereas neither ClbF nor ClbD alone were able to yield new ACP intermediates (Fig. 2c and 2d). For assays containing both dehydrogenases, MS analysis reveiled a major new species with a detected mass corresponding to glycyl-ACP. This product was previously shown to be indicative for AM-ACP due to spontaneous decarboxylation.[10a]

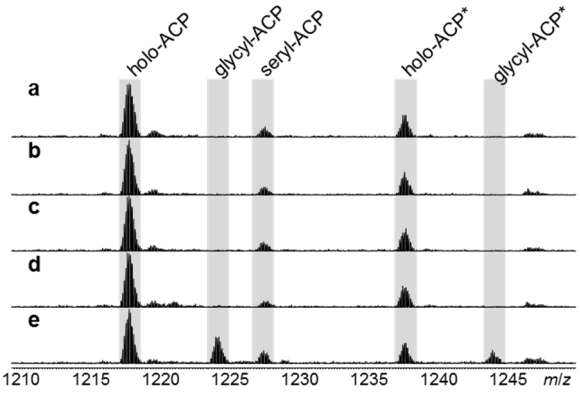


Fig. 2 Isotopically resolved spectra of ACP species $[M+9H]^{9+}$. (a) ACP species after addition of PPTase ClbA, ACP ClbE, adenylation domain ClbH A1 and their respective cofactors. The further spectra show the same experiment with the addition of: (b) boiled dehydrogenases ClbD and ClbF, (c) only ClbD, (d) only ClbF, and (e) both dehydrogenases ClbD and ClbF. ACPs labeled with an asterisk correspond to N-glycosylated species. Table S2 shows the identified ACP species in (e), their detected mass after deconvolution, and relative abdundance.

The data raised the question of whether the AM unit serves as precursor to an alternative ACC pathway or is incorporated into an as-yet unknown portion of **1**. Based on the position of *clbH* in the gene cluster, which corresponds to the timing of ACC incorporation, the latter hypothesis is counterintuitive. To nevertheless test both possibilities and to further test the relevance of AM for colibactin production, we generated two deletion mutants in the producer *E. coli* MG1655 pBAC*pks*$^+$. The mutant *pks*$^+$ Δ*clbE* lacked the ACP gene, whereas mutant *pks*$^+$ Δ*clbDEF* carried additional deletions of both dehydrogenase genes. Production of **1** was assessed in the resulting strains through quantification of megalocytosis, which correlate with DNA double strand breaks resulting from the genotoxic effect of colibactin[1]. HeLa cells were infected with the different strains for 4 hours, fixed, and stained with methylene blue in order to quantify the megalocytosis effect. These assays revealed that the inactivation of the *clbE* gene or the *clbDEF* operon abrogated the colibactin effect (Fig. 3). Transformation of these

mutants with plasmids carrying the functional wild type *clbE* or *clbDEF* genes resulted in a complete restoration of megalocytosis (Fig. 3).

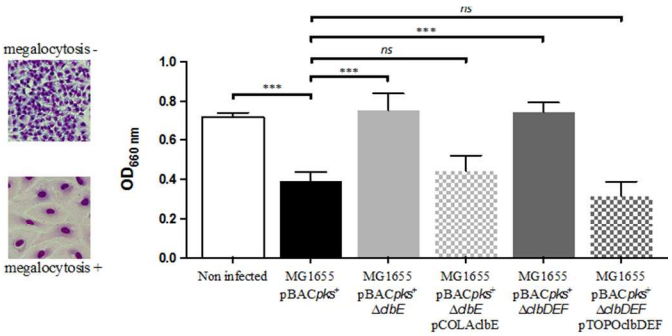


Fig. 3 Megalocytosis quantification for *clbE* and *clbDEF* mutants, and complemented derivatives (right). HeLa cells infected with the *pks*+ strain harbor a megalocytosis phenotype (left, bottom), while those infected with *clbE* or *clbDEF* mutants show a normal phenotype (left, top). Complementation of these mutants by functional *clbE* or *clbDEF* genes restores the megalocytosis effect. Multiplicity of infection: MOI = 400. Statistical analysis: one-way Anova. ***: $p < 0.001$, ns: not significant.

Further high-resolution mass spectrometric analysis of the deletion mutants revealed that the ACC-containing compounds **4** and **5** are still produced with comparable intensities to colibactin-positive strains (Fig. S6). These results suggest that *clbDEF* and thus AM are not involved in ACC formation, but on the other hand are indispensable for colibactin production as assessed by the megalocytosis assay. It is therefore likely that AM is incorporated into the unknown part of the colibactin structure. This is also supported by Ser feeding experiments of the Müller and Crawford groups that resulted in poor labeling of the known intermediates.[8f-g] Candidates for PKS modules that could process the AM unit are present in ClbK and ClbO (Fig. S1). Both contain highly aberrant domains resembling acyltransferases (ATs) but lacking the conserved active site serine residue within a GSHxG motif that is normally needed for building block transfer to the ACP (Fig. S7).[15] In addition to these integrated AT-like domains, the cluster encodes a candidate for a free-standing AT, ClbG. The protein features the GHSxG motif, but a HAFH motif, which in malonyl-specific ATs correlates with substrate specificity,[16] is changed to VPYH. A related motif (GPFH) is also present in the standalone AT ZmaF that transfers an aminomalonyl unit in zwittermicin biosynthesis.[10] Thus, ClbG is a good candidate for aminomalonyl transfer in colibactin biosynthesis. Aminomalonyl building blocks rarely occur in polyketide biosynthesis and have to our knowledge only been reported for zwittermicin[10] and guadinomine.[17]

Finally, we also wished to obtain biochemical data for other uncharacterized colibactin enzymes and therefore expressed the A domains of NRPS modules ClbJ-A1, ClbJ-A2, and ClbK-A2 (numbers refer to module number within the protein). While ClbJ-A2 yielded only insoluble protein under various conditions, the other A domains were obtained in soluble, active form. The adenylation assays showed clear preference for Gly by ClbJ-A1 and for Cys by ClbK-A2, but other amino acids were not or poorly accepted (Fig. S2). Gly is component of intermediate **5**, thus assigning ClbJ to this moiety. ClbK corresponds to an as-yet uncharacterized portion.

In summary, our functional study of eight protein components of the *clb* pathway provide insights into the missing regions of colibactin. These are predicted to contain aminomalonyl as well as Cys units. The NRPS architecture further suggests that the Cys residue is converted to a thiazol moiety. This information could aid in the targeted isolation and characterization of colibactin as a basis to study the pharmacology of this elusive and intriguing metabolite.

Notes and references

[a] Institute of Microbiology, Eigenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland.
[b] Inserm, UMR1043, INRA, USC 1360, CNRS, UMR5282, Toulouse, F-31000, France; Université de Toulouse III, Toulouse, F-31000, France; CHU Toulouse, Service de Bactériologie-Hygiène, Toulouse, F-31000, France.

† Electronic Supplementary Information (ESI) available: Experimental procedures and additional figures and tables. See DOI: 10.1039/c000000x/

1. (a) 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**, 848-851; (b) G. Cuevas-Ramos, C. R. Petit, I. Marcq, M. Boury, E. Oswald and J. P. Nougayrede, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 11537-11542.
2. J. Raisch, E. Buc, M. Bonnet, P. Sauvanet, E. Vazeille, A. de Vallee, P. Dechelotte, C. Darcha, D. Pezet, R. Bonnet, M. A. Bringer and A. Darfeuille-Michaud, *World J. Gastroenterol.*, 2014, **20**, 6560-6572.
3. (a) T. Secher, A. Samba-Louaka, E. Oswald and J. P. Nougayrede, *PLoS One*, 2013, **8**, e77157; (b) 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**, 1932-1942.
4. (a) J. Putze, C. Hennequin, J. P. Nougayrede, 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**, 4696-4703; (b) J. R. Johnson, B. Johnston, M. A. Kuskowski, J. P. Nougayrede and E. Oswald, *J. Clin. Microbiol.*, 2008, **46**, 3906-3911.
5. V. Bondarev, M. Richter, S. Romano, J. Piel, A. Schwedt and H. N. Schulz-Vogt, *Environ. Microbiol.*, 2013, **15**, 2095-2113.
6. P. Engel, M. I. Vizcaino and J. M. Crawford, *Appl. Environ. Microbiol.*, 2014.
7. S. Homburg, E. Oswald, J. Hacker and U. Dobrindt, *FEMS Microbiol. Lett.*, 2007, **275**, 255-262.
8. (a) 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**, 203-214; (b) X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang and R. Müller, *ChemBioChem*, 2013, **14**, 1194-1197; (c) C. A. Brotherton and E. P. Balskus, *J. Am. Chem. Soc.*, 2013, **135**, 3359-3362; (d) M. I. Vizcaino, P. Engel, E. Trautman and J. M. Crawford, *J. Am. Chem. Soc.*, 2014, **136**, 9244-9247; (e) C. A. Brotherton, M. Wilson, G. Byrd and E. P. Balskus, *Org. Lett.*, 2015, **17**, 2294; (f) X. Bian, A. Plaza, Y. Zhang and R. Müller, *Chem. Sci.*, online early version, doi: 10.1039/C5SC00101C; (g) M. I. Vizcaino and J. M. Crawford, *Nat. Chem.*, 2015, **7**, 411-417; (h) Z. R. Li, Y. Li, J. Y. Lai, J. Tang, B. Wang, L. Lu, G. Zhu, X. Wu, Y. Xu and P. Y. Qian, *ChemBioChem*, 2015, online early version, doi: 10.1002/cbic.201500239.
9. D. Reimer, K. M. Pos, M. Thines, P. Grun and H. B. Bode, *Nat. Chem. Biol.*, 2011, **7**, 888-890.
10. (a) Y. A. Chan and M. G. Thomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 14349-14354; (b) B. M. Kevany, D. A. Rasko and M. G. Thomas, *Appl. Environ. Microbiol.*, 2009, **75**, 1144-1155; (c) Y. A. Chan and M. G. Thomas, *Biochemistry*, 2010, **49**, 3667-3677.
11. (a) T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493-505; (b) G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211-224.
12. V. V. Phelan, Y. Du, J. A. McLean and B. O. Bachmann, *Chem. Biol.*, 2009, **16**, 473-478.
13. R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923-936.

14. P. Martin, I. Marcq, G. Magistro, M. Penary, C. Garcie, D. Payros, M. Boury, M. Olier, J. P. Nougayrede, M. Audebert, C. Chalut, S. Schubert and E. Oswald, *PLoS Path.*, 2013, **9**, e1003437.
15. S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041-1072.
16. F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortes, B. A. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489-494.
17. T. C. Holmes, A. E. May, K. Zaleta-Rivera, J. G. Ruby, P. Skewes-Cox, M. A. Fischbach, J. L. DeRisi, M. Iwatsuki, S. Omura and C. Khosla, *J. Am. Chem. Soc.*, 2012, **134**, 17797-17806.