JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS
Open Access on 06/16/2015

Comparative Metabolomics and Structural Characterizations Illuminate Colibactin Pathway-Dependent Small Molecules

Maria I. Vizcaino,[†,‡] Philipp Engel,[†,‡,⊥] Eric Trautman,[†,‡] and Jason M. Crawford*[,†,‡,∥]

[†]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States
[‡]Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States
[∥]Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, Connecticut 06510, United States

Supporting Information

ABSTRACT: The gene cluster responsible for synthesis of the unknown molecule "colibactin" has been identified in mutualistic and pathogenic *Escherichia coli*. The pathway endows its producer with a long-term persistence phenotype in the human bowel, a probiotic activity used in the treatment of ulcerative colitis, and a carcinogenic activity under host inflammatory conditions. To date, functional small molecules from this pathway have not been reported. Here we implemented a comparative metabolomics and targeted structural network analyses approach to identify a catalog of small molecules dependent on the colibactin pathway from the meningitis isolate *E. coli* IHE3034 and the probiotic *E. coli* Nissle 1917. The structures of 10 pathway-dependent small molecules are proposed based on structural characterizations and network relationships. The network will provide a roadmap for the structural and functional elucidation of a variety of other small molecules encoded by the pathway. From the characterized small molecule set, *in vitro* bacterial growth inhibitory and mammalian CNS receptor antagonist activities are presented.

Human health and disease are profoundly affected by the microbiota, a collection of resident microbial members on or in the human body.[1−3] Such intimate interactions are influenced by microbial metabolism and the metabolic exchange of small molecules between host and microbiota.[4,5] An important member of the symbiotic consortium is *Escherichia coli*, a common inhabitant of the human gut, whose presence at high abundance has been correlated to inflammation-induced dysbiosis (i.e., microbiota imbalance)[6] and colitis-associated colorectal cancer.[7,8] *E. coli* strains experimentally linked to tumor formation under inflammatory conditions[6] harbor the 54-kilobase genomic island "*pks*" (referred to here as "*clb*") responsible for the production of an unknown genotoxin(s) termed "colibactin".[9] The gene cluster encodes a hybrid polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) biosynthetic system.[9] A revised domain architecture for the pathway is illustrated in Figure S1. Presence of the *clb* gene cluster leads to DNA damage-mediated chromosome instability of mammalian cells *in vitro*[9] and *in vivo*[6,10] facilitating inflammation-induced colorectal cancer[6,7,11] and senescence-induced tumor growth.[12]

The *clb* genomic island has been identified in *E. coli* strains from the phylogenetic B2 group and closely related Enterobacteriaceae and correlates to B2 group long-term persistence in the intestine.[13−15] Acting as a virulence factor, *clb's* presence in *E. coli* also enhances lymphocyte death *in vivo* and sepsis-associated lymphopenia.[16] The locus has also been found in *E. coli* Nissle 1917 (EcN),[9] a probiotic bacterium used in the treatment of select inflammatory bowel diseases including ulcerative colitis.[17] Its role in EcN appears 2-fold, where pathway inactivation not only attenuated EcN's probiotic activity in colitis-induced mice but also its ability to inflict DNA damage on eukaryotic cells.[14] While EcN is cited as lacking defined virulence factors,[18] the island's carcinogenic role in closely related strains highlights the importance of characterizing small molecules encoded by this pathway. However, to date, attempts to identify genotoxic molecules from the colibactin pathway have failed. We analyzed whole-culture organic extracts to holistically identify colibactin pathway-dependent small molecules encoded by the meningitis isolate *E. coli* IHE3034 and the probiotic EcN using comparative metabolomics. Analyses among mutant and wildtype organisms provided a catalog of small molecules dependent on the colibactin pathway. We isolated and elucidated the structures of three major molecules from this catalog using detailed NMR analyses and chemical degradation experiments and propose the structures for seven additional molecules based on their connectivity within the structural network. Two of the isolated molecules and a third hydrolytic product identified were synthesized to assess relevant *in vitro* bioactivities and confirm absolute configurations.

We focused our comparative metabolomics analysis on the *clb* pathway derived from the pathogen *E. coli* IHE3034 and the probiotic EcN. For the *E. coli* IHE3034 pathway, the locus was previously cloned on a bacterial artificial chromosome (BAC, pBAC *clb*+) and heterologously expressed in *E. coli* DH10B cells.[9] We deleted the hydrolytic enzyme ClbP (Table S1), which has been proposed to catalyze an amide bond hydrolysis of an unknown precolibactin molecule.[19,20] For EcN, we deleted the entire locus as a control (Δ*clb*) and generated an analogous nonpolar Δ*clbP* strain. In this manner, we could identify unique small molecules present in cells harboring wildtype and mutant pathways compared to control cells

Received: April 7, 2014
Published: June 16, 2014

   dx.doi.org/10.1021/ja503450q | J. Am. Chem. Soc. 2014, 136, 9244−9247

lacking the gene cluster. The wildtype pathway from IHE3034 exhibited the characteristic megalocytosis phenotype against HeLa cells,[9] whereas the Δ*clbP* mutant, which could be complemented in *trans*, and the pBAC vector lacking the pathway (pBAC *clb*−) were inactive (Figure S2). We profiled whole-culture organic extracts from five biological replicates of each strain using untargeted HR-QTOF MS analysis. The complex chromatograms were data mined using a molecular feature extraction algorithm and molecular features (MOFs) were extracted, filtered, and statistically analyzed with Mass Profiler Professional (Agilent Technologies, see SI Methods, Figure S3). After removing MOFs detected in any of the corresponding control spectra, a conservative total of 101 unique MOFs (Table S2) were identified as *clb* pathway-dependent metabolites. The MOF distribution summary is shown in Figure 1, where a larger number of molecular features could be detected in the *clbP* mutants. This is likely due to the accumulation and detection of additional ClbP substrates. A number of MOFs were specific to a given strain, and only three were found in all four, which included ions *m*/*z* 260.2223, *m*/*z* 440.3485, and *m*/*z* 442.3280. While the *clb* genes in both IHE3034 and EcN strains share high sequence homology,[9] the unique small molecule distributions between strains could help to delineate the pathway's pathogenic and mutualistic roles in the host.

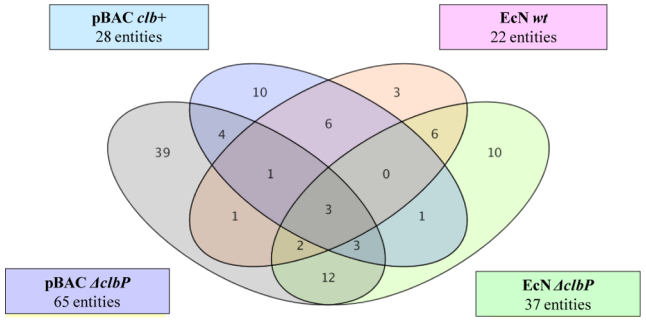

Figure 1. Venn diagram depicting colibactin pathway-dependent molecular features encoded by the *clb* locus of *E. coli* IHE3034 and EcN.

To further characterize the MOFs unique to the *clb* pathway, we ran a targeted metabolomics experiment, which provided enhanced fragmentation coverage of pathway-dependent small molecules compared to untargeted analysis (Figure S4). Targeted tandem MS (MS^2) molecular networking was performed using the generated unique ion list to cluster individual molecules based on their characteristic MS fragmentation patterns (Figure 2).[21,22] The three most abundant ions (pink nodes, metabolites **1**−**3**) in this network were selected for structural characterization, which allowed us to then propose structures for other interconnected small abundance metabolites (purple nodes, metabolites **4**−**9**) (Figure 2).

The most abundant wildtype ion was *m*/*z* 343.2592 (Table S2), which we structurally characterized early in our analysis from IHE3034 as *N*-myristoyl-D-Asn (**1**, Figure 2) using NMR (^{1}H, gCOSY, gHSQCAD, gHMBCAD; Figures S5−S8, Table S3) and MS^2 fragmentation (detailed in Figure S9). Two recent studies identified the *N*-myristoyl-D-Asn colibactin fragment by biochemical reconstitution[20] and structural characterization from EcN.[23] The second most abundant ion, *m*/*z* 341.2440 consistent with the formula $[C_{18}H_{33}N_2O_4]^+$ (theoretical *m*/*z* 341.2440, Figure S10), was similarly isolated and characterized (Figures S11−S14, Table S4) as a new colibactin metabolite,

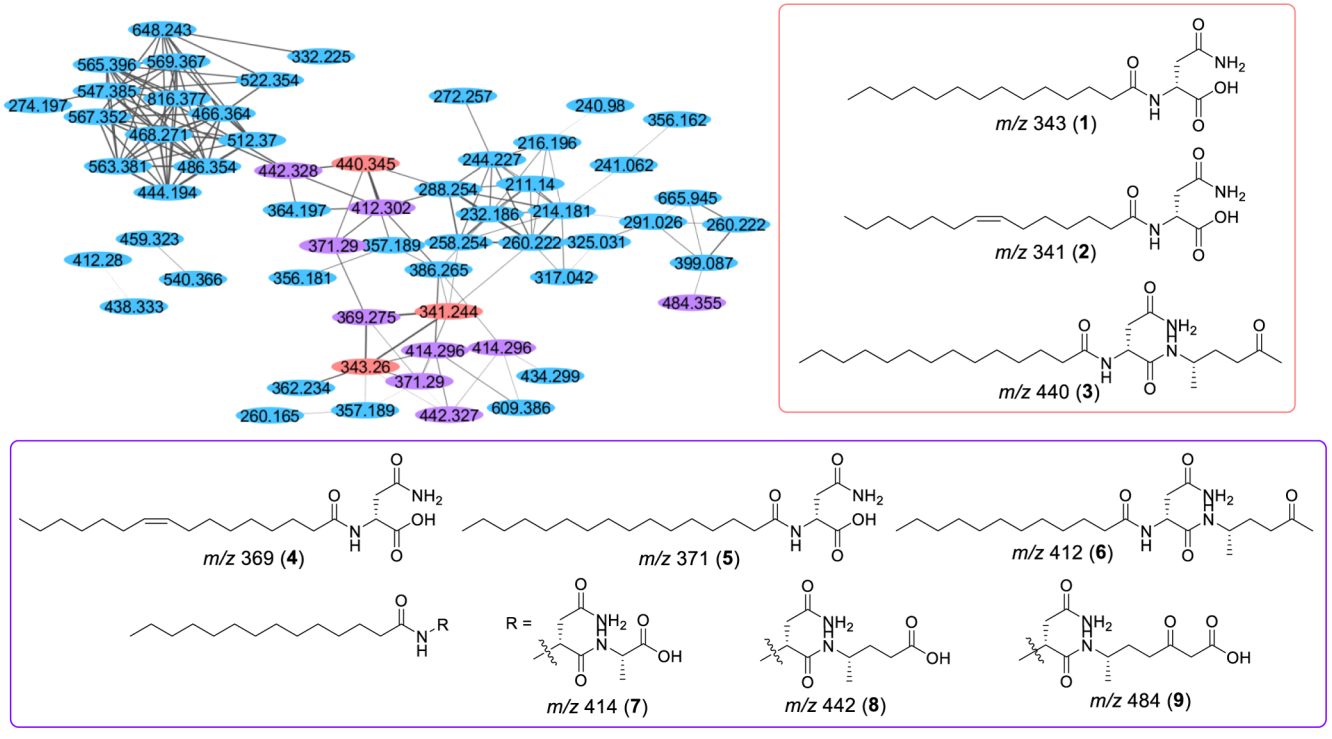

Figure 2. Colibactin pathway-dependent metabolites clustered based on their MS^2 fingerprint. Pink nodes (circles) were structurally characterized, and purple nodes indicate ions whose structures are proposed based on HRMS, fragmentation patterns, and network connections. Connectivity strength is represented by the thickness of lines connecting individual nodes.

dx.doi.org/10.1021/ja503450q | *J. Am. Chem. Soc.* 2014, 136, 9244−9247

cis-7-tetradecenoyl-D-Asn (**2**, Figure 2). A mass consistent with **2** was detected by MS in EcN,[23] but it was never isolated nor structurally characterized. Presence of the unsaturated double bond was characterized by an overlapped downfield methine shift at δ5.32 with a coupling constant of 7.29 Hz, indicating a cis conformation. The double bond location at *cis*Δ^7 was determined by ozonolysis and HRMS (Figure S15). Production of the unsaturated fatty acid chain can be attributed to one complete round of β-oxidation of the CoA thioester of palmitoleic acid (16:1 Δ^9), the second most commonly used fatty acid in *E. coli* membrane biosynthesis.[24] Based on these structures, the connected node m/z 369.2748 supports the C16 unsaturated fatty acyl-derivative **4** and the m/z 371.2900 node as the saturated fatty acyl-derivative **5** (Figures 2 and S16 and S17 show MS^2 fingerprints, respectively).

Previous studies have described the release of *N*-acyl-D-Asn small molecules by amide or ester hydrolysis through a prodrug activation mechanism in the xenocoumacins,[25] the zwittermicins,[26] the didemnins,[27] and the unknown colibactins.[20] Interestingly, deletion of ClbP led to the accumulation of ion m/z 440.3485 as the major network product in both mutants, which we identified as new colibactin metabolite **3** (Figure 2, detailed in SI). Structural characterization of **3** was determined by NMR correlations (^{1}H, gCOSY, gHSQCAD, and gHMBCAD; Figures S18–S22, Table S5), HRMS, network analysis, and synthesis. The HRMS showed a major ion peak of m/z 462.3306, which was consistent with the ion formula $[C_{24}H_{45}N_3O_4+Na]^+$ (theoretical m/z 462.3308) and supported by MS^2 fragmentation (Figure S18). Marfey's analysis established the D-configuration of Asn for **1**–**3**. The *R*, *S* absolute configuration of **3** was determined by chromatographic comparisons to synthetic standards of the *R*, *R* and *R*, *S* diastereomers (detailed in Figure S23). We validated **3** as an authentic ClbP substrate *in vivo* by feeding synthetic **3** to *E. coli* DH10B cells overexpressing an arabinose-inducible ClbP (Table S6).

A biosynthetic proposal for **3** is shown in Figure 3. Structural characterizations support that ClbB activates Ala *in vivo* and

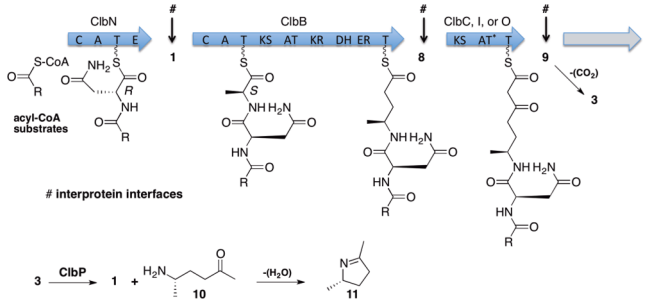

Figure 3. Biosynthetic proposal for formation of *clb* metabolites up to the structural complexity represented by **9**. Proposed intramolecular cyclization of **10** after being released from **3** by ClbP is also shown. C, condensation; A, adenylation; T, thiolation; E, epimerization; KS, ketosynthase; AT, acyl-transferase; KR, ketoreductase; DH, dehydratase; ER, enoyl-reductase.

extends an *N*-acyl-D-Asn-ClbN precursor, which is followed by successive polyketide extensions, reductive processing, and hydrolytic release of intermediate **9**. ClbB specificity for Ala is consistent with previous adenylation domain *in vitro* analysis.[20] In support of the intermediacy of **9**, an ion with m/z 484.3388 (**9**, Figure 2) was detected at low abundance in pBAC *clb*+ and EcN Δ*clbP* (Table S2) and transiently in pBAC Δ*clbP* by LC-MS. Its HRMS and fragmentation pattern (Figure S24) support a carboxylated form of **3** and is consistent with the biosynthetic pathway (Figure 3). These data indicate that a hydrolytic cleavage event releases **9**, which undergoes decarboxylation to form **3** as the predominant Δ*clbP* enrichment product (Figure 3). We deleted ClbL, a second protease in the pathway, as a possible catalyst responsible for this cleavage. However, Δ*clbL* produced similar levels of **1** and **3** compared to wildtype (Figure S25), suggesting that **3** is a preferential assembly line derailment product, in which the assembly line acyl-carrier protein thioester of **9** is hydrolyzed. ClbP then cleaves **3** liberating **1** and fragment **10**, which accumulates in our extracts as cyclic imine **11** by decay through intramolecular cyclization (Figure 3). This cyclic cleavage product **11** was identified in wildtype cultures but not in the *clbP* mutants (Table S2). Co-migration of **11** to a synthetic standard was confirmed by HRMS.

Interprotein junctions in multimodule assembly lines represent variable hydrolytic points through incomplete substrate transfer events.[28] While these events are generally regarded as being nonproductive, we speculate that the evolution of specific protein–protein junctions could provide a route for pathway output diversification. Indeed, small molecules identified in our metabolomics analyses can be correlated to interprotein junction sites of the proposed biosynthesis (e.g., **1**, **8**, and **9**, Figure 3).

To characterize possible functional roles of the colibactin metabolites, we subjected synthetic **1**, **3**, and **11** (Schemes S1 and S2) to a variety of relevant *in vitro* biological assays. We determined that the molecules did not induce the HeLa cell megalocytosis phenotype (detailed in Figure S2), suggesting that other minor metabolites within the colibactin network could possess this activity. Next, we tested whether synthetic **1** had antibacterial or growth inhibitory activities, as other *N*-acyl amino acids have been reported as antibiotics.[29−31] **1** exhibited mild growth inhibitory activity against *B. subtilis* NCIB 3610, whereas its synthetic L-enantiomer (**12**) lacked detectable activity (detailed in Figure S26). Growth inhibition of competitor bacteria in the gut could provide the producing host with a niche advantage as observed in other symbioses.[32] The molecules were also submitted to the Psychoactive Drug Screening Program for activity analysis against a panel of central nervous system receptors, transporters, and channels (http://pdsp.med.unc.edu/). A summary of the significant results are shown in Table S7. Briefly, **1** exhibited moderate antagonist activity against the dopamine 5 (D5) and 5-hydroxytryptamine-7 (5-HT7) receptors in primary binding assays at 10 μM. Quantitative LC-MS indicates that **1** accumulates at a physiologically relevant concentration of ∼27 μM in the extracellular environment of suspension cultures (Figure S27). D5 transcripts are detected in the gut,[33] but it is unclear how D5 inhibition in the gut could contribute to the observed phenotypes. Interestingly, inhibition of 5-HT7 alleviates colitis symptoms in DSS-colitis mouse models.[34]

We also noticed that the linear cleavage product **10** has structural similarities to the decongestant drug tuamine. Tuamine inhibits the norepinephrine transporter (NET) indirectly leading to vasoconstriction.[35] A variety of other structurally related molecules have similar sympathomimetic effects.[35] All attempts to synthesize fragment **10**, however, led to cyclic imine **11** (Scheme S2), which did not exhibit significant NET activity. This raises an open question as to

dx.doi.org/10.1021/ja503450q | J. Am. Chem. Soc. 2014, 136, 9244–9247

whether linear fragment **10** has transient activity before decay through intramolecular cyclization.

Taken together, we provide a catalog of small molecules dependent on the *clb* genomic island, which has been linked to long-term persistence of *E. coli* in the bowel, probiotic activity, and carcinogenesis. From this *clb* catalog, we characterized one known and two previously unknown metabolites and propose the structures of seven additional minor molecules based on network analysis. We explore potential functional roles *in vitro* for three of the molecules that could be contributing in part to mutualistic or pathogenic host-bacteria interactions. Our data support a model, in which "colibactin" represents an ensemble of molecules that could collectively contribute to the phenotypes observed for the pathway. The *clb* structural network provides a roadmap for elucidating the roles of small molecules encoded by this important pathway. Structural and functional characterizations of other molecules in the diversified catalog are underway.

■ ASSOCIATED CONTENT

Supporting Information

Experimental and synthetic procedures, NMR and HRMS characterization of new compounds, supplementary figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

jason.crawford@yale.edu

Present Address

[⊥](P.E.) Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (1DP2CA186575) and the Damon Runyon Cancer Research Foundation (DFS:05-12) to J.M.C. We thank Prof. Eric Oswald (Inserm, France) for providing plasmid pBAC-empty and pBAC *clb*+. Binding, K_i determinations, and antagonist functional data for CNS receptors, transporters, and channels were generously provided by the National Institute of Mental Health's Psychoactive Drug Screening Program (NIMH PDSP contract no. HHSN-271-2008-00025-C).

■ REFERENCES

(1) Dethlefsen, L.; McFall-Ngai, M.; Relman, D. A. *Nature* **2007**, *449*, 811.

(2) Greenblum, S.; Turnbaugh, P. J.; Borenstein, E. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *109*, 594.

(3) Cho, I.; Blaser, M. J. *Nat. Rev. Genet.* **2012**, *13*, 260.

(4) Phelan, V. V.; Liu, W. T.; Pogliano, K.; Dorrestein, P. C. *Nat. Chem. Biol.* **2012**, *8*, 26.

(5) Rath, C. M.; Dorrestein, P. C. *Curr. Opin. Microbiol.* **2012**, *15*, 147.

(6) Arthur, J. C.; Perez-Chanona, E.; Muhlbauer, M.; Tomkovich, S.; Uronis, J. M.; Fan, T. J.; Campbell, B. J.; Abujamel, T.; Dogan, B.; Rogers, A. B.; Rhodes, J. M.; Stintzi, A.; Simpson, K. W.; Hansen, J. J.; Keku, T. O.; Fodor, A. A.; Jobin, C. *Science* **2012**, *338*, 120.

(7) Arthur, J. C.; Jobin, C. *Gut Microbes* **2013**, *4*, 253.

(8) Buc, E.; Dubois, D.; Sauvanet, P.; Raisch, J.; Delmas, J.; Darfeuille-Michaud, A.; Pezet, D.; Bonnet, R. *PloS One* **2013**, *8*, e56964.

(9) Nougayrede, J. P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Science* **2006**, *313*, 848.

(10) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrede, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11537.

(11) Elinav, E.; Nowarski, R.; Thaiss, C. A.; Hu, B.; Jin, C.; Flavell, R. A. *Nat. Rev. Cancer* **2013**, *13*, 759.

(12) Cougnoux, A.; Dalmasso, G.; Martinez, R.; Buc, E.; Delmas, J.; Gibold, L.; Sauvanet, P.; Darcha, C.; Déchelotte, P.; Bonnet, M.; Pezet, D.; Wodrich, H.; Darfeuille-Michaud, A.; Bonnet, R. *Gut* **2014**, *424*, 203.

(13) Putze, J.; Hennequin, C.; Nougayrede, J. P.; Zhang, W.; Homburg, S.; Karch, H.; Bringer, M. A.; Fayolle, C.; Carniel, E.; Rabsch, W.; Oelschlaeger, T. A.; Oswald, E.; Forestier, C.; Hacker, J.; Dobrindt, U. *Infect. Immun.* **2009**, *77*, 4696.

(14) Olier, M.; Marcq, I.; Salvador-Cartier, C.; Secher, T.; Dobrindt, U.; Boury, M.; Bacquie, V.; Penary, M.; Gaultier, E.; Nougayrede, J. P.; Fioramonti, J.; Oswald, E. *Gut Microbes* **2012**, *3*, 501.

(15) Nowrouzian, F. L.; Oswald, E. *Microb. Pathogenesis* **2012**, *53*, 180.

(16) Marcq, I.; Martin, P.; Payros, D.; Cuevas-Ramos, G.; Boury, M.; Watrin, C.; Nougayrede, J. P.; Olier, M.; Oswald, E. *J. Infect. Dis.* **2014**, DOI: 10.1093/infdis/jiu071.

(17) Schultz, M. *Inflammatory Bowel Dis.* **2008**, *14*, 1012.

(18) Grozdanov, L.; Raasch, C.; Schulze, J.; Sonnenborn, U.; Gottschalk, G.; Hacker, J.; Dobrindt, U. *J. Bacteriol.* **2004**, *186*, 5432.

(19) Cougnoux, A.; Gibold, L.; Robin, F.; Dubois, D.; Pradel, N.; Darfeuille-Michaud, A.; Dalmasso, G.; Delmas, J.; Bonnet, R. *J. Mol. Biol.* **2012**, *424*, 203.

(20) Brotherton, C. A.; Balskus, E. P. *J. Am. Chem. Soc.* **2013**, *135*, 3359.

(21) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1743.

(22) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J. Am. Chem. Soc.* **2012**, *134*, 1817.

(23) Bian, X.; Fu, J.; Plaza, A.; Herrmann, J.; Pistorius, D.; Stewart, A. F.; Zhang, Y.; Muller, R. *ChemBioChem* **2013**, *14*, 1194.

(24) Magnuson, K.; Jackowski, S.; Rock, C. O.; Cronan, J. E., Jr. *Microbiol. Rev.* **1993**, *57*, 522.

(25) Reimer, D.; Pos, K. M.; Thines, M.; Grün, P.; Bode, H. B. *Nat. Chem. Biol.* **2011**, *7*, 888.

(26) Kevany, B. M.; Rasko, D. A.; Thomas, M. G. *Appl. Environ. Microbiol.* **2009**, *75*, 1144.

(27) Xu, Y.; Kersten, R. D.; Nam, S. J.; Lu, L.; Al-Suwailem, A. M.; Zheng, H.; Fenical, W.; Dorrestein, P. C.; Moore, B. S.; Qian, P. Y. *J. Am. Chem. Soc.* **2012**, *134*, 8625.

(28) Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T. *Chem. Biol.* **2002**, *9*, 333.

(29) McKellar, R. C.; Paquet, A.; Ma, C. Y. *Food Microbiol.* **1992**, *9*, 67.

(30) Brady, S. F.; Chao, C. J.; Clardy, J. *Appl. Environ. Microbiol.* **2004**, *70*, 6865.

(31) Brady, S. F.; Clardy, J. *J. Am. Chem. Soc.* **2000**, *122*, 12903.

(32) Vizcaino, M. I.; Guo, X.; Crawford, J. M. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 285.

(33) Li, Z. S.; Schmauss, C.; Cuenca, A.; Ratcliffe, E.; Gershon, M. D. *J. Neurosci.* **2006**, *26*, 2798.

(34) Kim, J. J.; Bridle, B. W.; Ghia, J.-E.; Wang, H.; Syed, S. N.; Manocha, M. M.; Rengasamy, P.; Shajib, M. S.; Wan, Y.; Hedlund, P. B.; Khan, W. I. *J. Immunol.* **2013**, *190*, 4795.

(35) Schlessinger, A.; Geier, E.; Fan, H.; Irwin, J. J.; Shoichet, B. K.; Giacomini, K. M.; Sali, A. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15810.