Published in final edited form as:

Org Lett. 2017 July 07; 19(13): 3560–3563. doi:10.1021/acs.orglett.7b01534.

Identification and Heterologous Production of a Benzoyl-Primed Tricarboxylic Acid Polyketide Intermediate from the Zaragozic Acid A Biosynthetic Pathway

Nicholas Liu¹, Yiu-Sun Hung¹, Shu-Shan Gao¹, Leibniz Hang², Yi Zou¹, Yit-Heng Chooi^{3,*}, and Yi Tang^{1,2,*}

¹Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California, USA

²Department of Chemistry and Biochemistry, University of California, Los Angeles, California, USA

³School of Molecular Sciences, The University of Western Australia, Perth, Australia

Abstract

Zaragozic acid A (1) is a potent cholesterol lowering, polyketide natural product made by various filamentous fungi. The reconstitution of enzymes responsible for the initial steps of the biosynthetic pathway of 1 is accomplished using an engineered fungal heterologous host. These initial steps feature the priming of a benzoic acid starter unit onto a highly reducing polyketide synthase (HRPKS), followed by oxaloacetate extension and product release to generate a tricarboxylic-acid containing product 2. The reconstitution studies demonstrated only three enzymes, HRPKS, citrate synthase and hydrolase, are needed in *A. nidulans* to produce the structurally complex product.

Fungal polyketide natural products have been an important source of pharmaceutical drugs due to their wide range of bioactivities.¹ The diverse and complex structural features have also attracted intense research efforts towards understanding the biosynthetic logic.² The carbon scaffolds of many fungal polyketide natural products, including lovastatin³ and cytochalasans,⁴ are built from the iterative functions of highly reducing polyketide synthases (HRPKSs). These scaffolds are typically oxidatively modified by subsequent downstream tailoring enzymes, such as oxidases and oxygenases, to furnish the mature product.⁵ Zaragozic acid (ZA) A (1) (also known as squalestatin S1), first discovered in 1992,⁶ is a heavily oxidized fungal polyketide that offers potent cholesterol lowering activity.⁷ Members of the ZA family of molecules share a 2,8-dioxobicyclic[3.2.1]octane-3,4,5-tricarboxylic acid core that is connected to two lipophilic polyketide or fatty acid derived arms (Figure

Corresponding Author: yitang@ucla.edu, yitheng.chooi@uwa.edu.au.

Notes

The authors declare no competing financial interests.

ASSOCIATED CONTENT SUPPORTING INFORMATION

Experimental details, spectroscopic and computational data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

1).⁸ The unique structural features of ZA mimic presqualene diphosphate, the product of the head-head condensation of farnesyl-diphosphate, and make ZA potent inhibitors of squalene synthase.⁹ Though various total syntheses of **1** have been reported,¹⁰ a complete understanding of its biosynthesis, especially with regard to formation of the bicyclic core remains elusive. Labeling studies have shown the tricarboxylic acid core is partially-derived from oxaloacetate, an intermediate found in the citric acid cycle.¹¹ Cox et al. have shown the tetraketide arm (left in Figure 1) in **1** is synthesized by a HRPKS¹² and enzymatically esterified to the core in the last biosynthetic step.¹³ However, formation of the other polyketide arm of **1** is unresolved. Here, we utilize an engineered *Aspergillus nidulans* heterologous host to reconstitute the biosynthesis of the benzoyl-containing portion of **1**, and identified a tricarboxylic acid-containing product **2** that may serve as the starting point in the oxidative maturation of the dioxobicyclic core.

We sequenced the genome of the fungal pathogen *Curvularia lunata* (also known as *Cochliobolus lunatus* ATCC 74067), which was previously identified as a producer of **1**.¹⁴ Searching through the assembled genome led to identification of a gene cluster (*clz*) that is likely responsible for the biosynthesis of **1** (Figure 2A). This cluster is nearly identical in gene organization (Figure S1 and Table S4) to the recently reported squalastatin S1 cluster identified from *Phoma* sp. C2932 and unidentified fungus MF5453 by Cox and coworkers.¹³ Among the ~20 genes in the cluster (Figure S1), there is a gene that encodes squalene synthase (*Clz20*) as a potential resistance enzyme to **1** and two HRPKSs (*Clz2*) and (*Clz14*). *Clz2* is highly homologous to the previously identified squalastatin tetraketide synthase.^{12, 15} Also present is the acyltransferase *Clz6* that catalyzes transferring of tetraketide product to the hydroxyl group in the tricarboxylic acid core.¹³ Therefore, we designated the *clz* cluster to be responsible for **1** in *C. lunata*.

Here we focused on identifying the enzymes that assemble the benzyl containing polyketide arm. We reasoned that the uncharacterized HRPKS, *Clz14*, should be involved, but the structure of the potential polyketide product is unknown, especially with respect to the extent of reduction at the different C β -carbons. One anticipated feature of the product is the presence of the benzyl group, which is proposed to be a starter unit for the HRPKS.¹⁶ Although nonacetate starter units, such as propionate, have been observed in priming of fungal HRPKSs,¹⁷ a benzoate unit would represent the most significant departure from the canonical starter unit acetate. While the benzoate starter unit is used by some bacterial PKS pathways such as in the biosynthesis of soraphen¹⁸ or enterocin,¹⁹ it is very rarely used by fungal PKS systems. The only other proposed pathway in addition to **1** is in the biosynthesis of the strobilurin/oudemansin family of molecules.²⁰ The *N*-terminus of *Clz14* contains a ~90 residue segment before the KS domain that bears no secondary structure or signal peptide sequences, and is not found in other fungal PKSs (Figure S2). Two genes in the *clz* cluster, *clz10* and *clz12*, encode phenylalanine ammonia lyase (PAL) and 4-coumarate-CoA ligase, respectively. These two enzymes may be involved in transforming phenylalanine into benzoyl-CoA, a strategy that is used in the biosynthesis of the enterocin natural products in *Streptomyces*.²¹ Also present in the gene cluster is a homolog of citrate synthase, *Clz17*, which may be involved in connecting the C α carbons of the polyketide chain and oxaloacetate to afford the tricarboxylic acid unit. Homologs of this enzyme are found in

gene clusters of nonadride-containing polyketides, and often function in tandem with an alkylcitrate dehydratase to form maleic anhydride.²² No dehydratase homolog is found in the *clz* cluster, consistent with the presence of the tricarboxylic acid moiety in **1**. Instead two potential hydrolytic enzymes, Clz11 and Clz13, which are α/β hydrolase and β -lactamase, respectively, are in close proximity to Clz14, and may participate in product release.

To elucidate the function of Clz14, we used an episomally based heterologous system in *A. nidulans* A1145 that is capable of expressing up to 12 genes using 3 plasmids.²³ We first introduced five genes (*clz*10, 11, 12, 14 and 17) and monitored total ion count (TIC) to detect formation of new products. However, the plasmid system induced significant production of both sterigmatocystin²⁴ and emericellamide,²⁵ which contributed to a high background TIC (Figures S7 and S8). This amalgam of endogenous metabolites, compounded by the unknown product mass, made identification of any new product related to **1** difficult. To remove these metabolites, we used homologous recombination facilitated by CRISPR-Cas9 to delete *stcA* and *easA* and yield the strain *A. nidulans* A1145 ST EM (Figures S7 and S8).²⁶ Reintroduction of the *clz* genes into this engineered strain led to the identification of a new mass peak (m/z [M-H]⁻ 419) that was previously buried under the signals of sterigmatocystin and emericellamide B (Figure 2B, iii and Figure S7). Exclusion of *clz14* abolished the production of **2**, confirming it is derived from the HRPKS.

Large-scale fermentation was performed to isolate sufficient amount of **2** (titer of ~ 0.1 mg/L) for NMR characterization and structural elucidation. **2** was found to have the molecular formula C₂₃H₃₂O₇ based on positive HRESIMS data (Figure S5). A database search for ZA related compounds revealed that a previous characterized compound, L-731.120, was isolated from the ZAA producing strain MF5453 (ATCC 20986),²⁷ which could potentially match **2**. Detailed analysis of the 1D and 2D NMR data of **2**, particularly the COSY and HMBC spectra, revealed the presence of a monosubstituted phenyl ring and one trisubstituted double bond, which led to the assignment of C-8 to C-19 fragment (Table S6 and Figures S9–14). Further analysis of the ¹H NMR data of **2** revealed the methylene group C-2 is bonded to two quaternary carbons, supported by the AB geminal coupling (J = 16 Hz). This evidence combined with HMBC correlations from H-2 to C-20 and from H-4 to C-3, C-20 and C-21 established the tricarboxylic acid substructure. This moiety was connected to the benzyl-containing fragment via three methylene groups, which accounted for the rest of the unassigned atoms in the molecular formula. Thus, the planar structure of **2** was assigned. EIC analysis of the metabolite extract from *C. lunata* also revealed the presence of **2** (Figure 2B, ii) along with **1**, further corroborating the relationship of **2** to the set of *clz* genes.

To investigate the priming pathways and incorporation of the benzoate starter unit by Clz14, we performed labeling studies using either *d*₅-benzoic acid or *d*₈-phenylalanine (Figure 2B, iv and S4). In both cases, we observed the increase in molecular weight of **2** by 5 mu. In the case of *d*₈-phenylalanine feeding, retention of five deuterium labels is consistent with the proposed conversion to benzoic acid, during which PAL (Clz10) yields cinnamate, followed by esterification to yield cinnamoyl-CoA, which can undergo β -oxidation to yield benzoyl-CoA.²⁸ Incorporation of *d*₅-benzoate into **2** suggests the presence of an endogenous aryl-CoA ligase that can afford benzoyl-CoA (Figure 3). When *d*₅-benzoic acid was fed at a high

concentration (1 mg/mL), a significant amount (~30%) of **2** remained unlabeled, which represents the unlabeled benzoyl-CoA pool derived from phenylalanine. This parallel pathway to benzoyl-CoA was also observed in the priming steps of enterocin in *Streptomyces*,¹⁹ and suggests Clz10 or Clz12 may not be absolutely required in *A. nidulans*. Indeed, removing either Clz10 or Clz12 from the *A. nidulans* constructs retained production of **2**, while removing both enzymes led to ~5 fold decrease in the titer of **2** (Figure 2B, v, vi, and vii). To probe if **2** biosynthesized from the minimal construct (Clz11, Clz14 and Clz17) is derived from benzoate in *A. nidulans*, we repeated the feeding of *d*₅-benzoate into this host. As shown in Figure S4, unlabeled **2** is nearly abolished, confirming that in the absence of Clz10 and Clz12, the level of benzoyl-CoA derived from phenylalanine is very low. These studies therefore indicate the importance of Clz10 and Clz12 in the priming pathways in the biosynthesis of **2**. These two enzymes are particularly essential if the level of benzoate in the native host is low compared to that in *A. nidulans*.

We also investigated the product releasing steps that lead to biosynthesis of **2**. The citrate synthase Clz17 is essential in the biosynthesis as removing the gene from *A. nidulans* abolished production (Figure S3). We propose that following completion of the polyketide synthesis, which yields **4**, Clz17 catalyzes the addition to oxaloacetate to yield **5**. This is consistent with the role of homologous enzymes in the nonadride biosynthetic pathways, including those of phomoidride,^{22a} byssochlamic acid^{22b} and rubratoxin.^{22c} We propose that Clz11 is then responsible for directly hydrolyzing **5** to yield **2**. Clz11 shares sequence homology to other uncharacterized fungal α/β hydrolase enzymes such as in *Lepidopterella palustris* (62% id.) or *Glarea lozoyensis* (48% id.), but only very weak identity to LovG from the lovastatin pathway.²⁹ Removal of Clz11 from the *A. nidulans* construct that produced **2** led to ~99% reduction of the product level (Figure S3), indicating while spontaneous or nonspecific enzyme hydrolysis is present, Clz11 significantly accelerates product turnover. Substituting Clz11 with the β -lactamase homolog Clz13 also resulted in only background level of product formation (Figure S3). The timing of Clz11 activity is precise and only acts to release **2** from **5**. In the construct without the citrate synthase Clz17, we could not detect any trace of the free polyketide **3**. The combination of citrate synthase and hydrolase therefore represents a highly controlled mode of product release from the HRPKS.

Using the heterologous expression, we have shown that only three enzymes, Clz14, Clz17 and Clz11 are required to synthesize **2**, utilizing the benzoyl-CoA pool that is naturally present in *A. nidulans*. This represents an exceptionally concise pathway to a structurally complex, amphiphilic polyketide product. Benzoyl-priming of Clz14 is unusual among fungal PKS systems, and further biochemical characterization of the KS domain and the unique N-terminal region may reveal the molecular basis of using the non-acetate starter unit. A C-5/C-6 unsaturated, C-7 ketone version of **2** was previously proposed to be the product of the HRPKS,¹³ however this compound was not observed here. From **2**, a series of hydroxylation reactions must take place to furnish the advanced intermediate **6**, including the 2,8-dioxobicyclic[3.2.1]octane-3,4,5-tricarboxylic acid core and the C-10/C-22 *exo*-methylene. In particular, the oxidative installation of hydroxyl groups derived from atmospheric oxygen,⁶ at five *sp*³ carbons C-2, C-4, C-5, C-6, C-7 are proposed to take place, representing a remarkable cascade of C-H activation steps on vicinal carbon atoms. These

reactions may be iteratively catalyzed by two enzymes with homology to non-heme iron and α -ketoglutarate dependent oxygenases Clz15 and Clz16, which are also found in the homologous pathway from *Phoma* sp. C2932.¹³ Homologous enzymes in the rubratoxin pathway was recently shown to hydroxylate similarly unactivated carbons, although not iteratively.^{22c} The identification and reconstitution of **2** described here set the stage for delineating the order and enzymology of these enigmatic reactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the by NIH 1R35GM118056 and 1DP1GM106413 to Y.T, and by Australian Research Council FT160100233 to Y.H.C. N.L. is supported by NIH Biotechnology Training in Biomedical Sciences and Engineering (T32GM067555). L.H. is supported by NIH Cellular and Molecular Biology (CMB) Training Grant (T32GM007185). Dr. Megan McDonald (Australian National University) is thanked for the assistance with genome assembly. We thank Professor Uffe Mortensen (Technical University of Denmark) for providing pFC330 for the CRISPR/Cas9 work.

References

1. Lim FY, Keller NP. Nat Prod Rep. 2014; 31:1277. [PubMed: 25142354]
2. (a) Chooi YH, Tang Y. J Org Chem. 2012; 77:9933. [PubMed: 22938194] (b) Cox RJ, Simpson TJ. Method Enzymol. 2009; 459:49.(c) Cox RJ. Org Biomol Chem. 2007; 5:2010. [PubMed: 17581644]
3. Ma SM, Li JW, Choi JW, Zhou H, Lee KK, Moorthie VA, Xie X, Kealey JT, Da Silva NA, Vederas JC, Tang Y. Science. 2009; 326:589. [PubMed: 19900898]
4. Scherlach K, Boettger D, Remme N, Hertweck C. Nat Prod Rep. 2010; 27:869. [PubMed: 20411198]
5. (a) Cochrane RVK, Vederas JC. Accounts Chem Res. 2014; 47:3148.(b) Cox R. Nat Prod Rep. 2014; 31:1405. [PubMed: 25060008] (c) Tang MC, Zou Y, Watanabe K, Walsh CT, Tang Y. Chem Rev. 2017; 117:5226. [PubMed: 27936626] (d) Hang L, Liu N, Tang Y. ACS Catal. 2016; 6:5935. [PubMed: 28529817]
6. Jones CA, Sidebottom PJ, Cannell RJP, Noble D, Rudd BAM. J Antibiot. 1992; 45:1492. [PubMed: 1429235]
7. Bergstrom JD, Kurtz MM, Rew DJ, Amend AM, Karkas JD, Bostedor RG, Bansal VS, Dufresne C, VanMiddlesworth FL, Hensens OD, et al. Proc Natl Acad Sci U S A. 1993; 90:80. [PubMed: 8419946]
8. Byrne KM, Arison BH, Nallinomstead M, Kaplan L. J Org Chem. 1993; 58:1019.
9. Liu CI, Jeng WY, Chang WJ, Ko TP, Wang AH. J Biol Chem. 2012; 287:18750. [PubMed: 22474324]
10. (a) Caron S, Stoermer D, Mapp AK, Heathcock CH. J Org Chem. 1996; 61:9126.(b) Freeman-Cook KD, Halcomb RL. J Org Chem. 2000; 65:6153. [PubMed: 10987953] (c) Nicolaou KC, Yue EW, Lagreca S, Nadin A, Yang Z, Leresche JE, Tsuru T, Naniwa Y, Dericcardis F. Chem-Eur J. 1995; 1:467.(d) Stoermer D, Caron S, Heathcock CH. J Org Chem. 1996; 61:9115.(e) Tomooka K, Kikuchi M, Igawa K, Suzuki M, Keong PH, Nakai T. Angew Chem Intl Ed. 2000; 39:4502.
11. Bergstrom JD, Dufresne C, Bills GF, Nallinomstead M, Byrne K. Annu Rev Microbiol. 1995; 49:607. [PubMed: 8561474]
12. Liddle E, Scott A, Han C, Ivison D, Simpson TJ, Willis CL, Cox RJ. Chem Commun. 2017; 53:1727.
13. Bonsch B, Belt V, Bartel C, Duensing N, Koziol M, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. Chem Commun. 2016; 52:6777.

14. Bills GF, Pelaez F, Polishook JD, Diezmatas MT, Harris GH, Clapp WH, Dufresne C, Byrne KM, Nallinomstead M, Jenkins RG, Mojena M, Huang LY, Bergstrom JD. *Mycol Res.* 1994; 98:733.
15. Cox RJ, Glod F, Hurley D, Lazarus CM, Nicholson TP, Rudd BAM, Simpson TJ, Wilkinson B, Zhang Y. *Chem Commun.* 2004:2260.
16. Nadin A, Nicolaou KC. *Angew Chem Intl Ed.* 1996; 35:1623.
17. Moore BS, Hertweck C. *Nat Prod Rep.* 2002; 19:70. [PubMed: 11902441]
18. Ligon J, Hill S, Beck J, Zirkle R, Molnar I, Zawodny J, Money S, Schupp T. *Gene.* 2002; 285:257. [PubMed: 12039053]
19. Piel J, Hertweck C, Shipley PR, Hunt DM, Newman MS, Moore BS. *Chem Biol.* 2000; 7:943. [PubMed: 11137817]
20. Nerud F, Sedmera P, Zouchova Z, Musilek V, Vondracek M. *Collect Czech Chem C.* 1982; 47:1020.
21. Xiang LK, Moore BS. *J Bacteriol.* 2003; 185:399. [PubMed: 12511484]
22. (a) Fujii R, Matsu Y, Minami A, Nagamine S, Takeuchi I, Gomi K, Oikawa H. *Org Lett.* 2015; 17:5658. [PubMed: 26558485] (b) Williams K, Szwalbe AJ, Mulholland NP, Vincent JL, Bailey AM, Willis CL, Simpson TJ, Cox RJ. *Angew Chem Intl Ed.* 2016; 55:6783.(c) Bai J, Yan D, Zhang T, Guo Y, Liu Y, Zou Y, Tang M, Liu B, Wu Q, Yu S, Tang Y, Hu Y. *Angew Chem Intl Ed.* 2017; 56:4782.
23. Sato M, Yagishita F, Mino T, Uchiyama N, Patel A, Chooi YH, Goda Y, Xu W, Noguchi H, Yamamoto T, Hotta K, Houk KN, Tang Y, Watanabe K. *Chembiochem.* 2015; 16:2294. [PubMed: 26360642]
24. Yu JH, Leonard TJ. *J Bacteriol.* 1995; 177:4792. [PubMed: 7642507]
25. Chiang YM, Szewczyk E, Nayak T, Davidson AD, Sanchez JF, Lo HC, Ho WY, Simityan H, Kuo E, Praseuth A, Watanabe K, Oakley BR, Wang CCC. *Chem Biol.* 2008; 15:527. [PubMed: 18559263]
26. (a) Nodvig CS, Nielsen JB, Kogle ME, Mortensen UH. *PLoS One.* 2015; 10:e0133085. [PubMed: 26177455] (b) Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. *Nucleic Acids Res.* 2014; 42:W401. [PubMed: 24861617]
27. Harris GH, Dufresne C, Joshua H, Koch LA, Zink DL, Salmon PM, Goklen KE, Kurtz MM, Rew DJ, Bergstrom JD, Wilson KE. *Bioorg Med Chem Lett.* 1995; 5:2403.
28. Qualley AV, Widhalm JR, Adebesin F, Kish CM, Dudareva N. *Proc Natl Acad Sci U S A.* 2012; 109:16383. [PubMed: 22988098]
29. Xu W, Chooi YH, Choi JW, Li S, Vederas JC, Da Silva NA, Tang Y. *Angew Chem Intl Ed.* 2013; 52:6472.

Figure 1.
Representative compounds from the zaragozic acids family of polyketides.

Figure 2.

A) Organization of the zaragozic acid A (*clz*) gene cluster found in *C. lunata*. The magnified region contains genes that were hypothesized to be responsible for the biosynthesis of the benzoyl-primed, tricarboxylic acid intermediate **2**. B) Extracted ion chromatography of LC-MS traces from both *C. lunata* (13 days) and *A. nidulans* (3 days) showing production of metabolites of interest. All masses shown correspond to m/z [M-H]⁻. i: Standard of **1**; ii: production of **1** and **2** from *C. lunata*; iii–vii: production of **2** from different combinations of genes from *C. lunata* reconstituted in the *A. nidulans* A1145 ST EM. iv: production of d_5 -labeled **2** upon feeding of d_5 -benzoic acid.

Figure 3.

Proposed biosynthesis of zaragozic acid **1**. Clz14 is the HRPKS involved in the first steps of the biosynthesis. The domains and steps of Clz14 are shown as white arrows. HRPKS domain abbreviations: KS: ketosynthase; AT: acyltransferase; DH: dehydratase; MT: methyltransferase; ER: enoylreductase; KR: ketoreductase; ACP: acyl carrier protein.