ORGANIC LETTERS

2009
Vol. 11, No. 6
1241−1244

Protostadienol Biosynthesis and Metabolism in the Pathogenic Fungus *Aspergillus fumigatus*

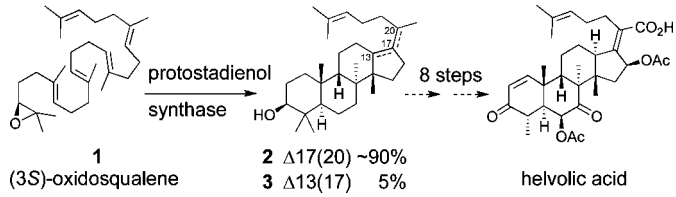
Silvia Lodeiro,[†] Quanbo Xiong,[‡] William K. Wilson,[‡] Yulia Ivanova,[†] McKenzie L. Smith,[†] Gregory S. May,[§] and Seiichi P. T. Matsuda*[,†,‡]

Departments of Chemistry and of Biochemistry and Cell Biology, Rice University, Houston, Texas 77005, and Division of Pathology and Laboratory Medicine, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

matsuda@rice.edu

Received November 21, 2008

ABSTRACT



Details of the fungal biosynthetic pathway to helvolic acid and other fusidane antibiotics remain obscure. During product characterization of oxidosqualene cyclases in *Aspergillus fumigatus*, we found the long-sought cyclase that makes (17*Z*)-protosta-17(20),24-dien-3β-ol, the precursor of helvolic acid. We then identified a gene cluster encoding the pathway to helvolic acid, which is controlled by a transcription regulator (LaeA) associated with fungal virulence. Evidence regarding the evolutionary origin and taxonomic distribution of fusidane biosynthesis is also presented.

The pathogenic fungus *Aspergillus fumigatus* is pervasive in air and soil.[1] Its airborne spores readily invade the human lung, resulting in high mortality among immunocompromised individuals.[2] This virulence can be promoted by secondary metabolites,[3,4] such as the antibiotic helvolic acid. While characterizing the triterpene synthases[5] of *As. fumigatus*, we found the enzyme that cyclizes (3*S*)-oxidosqualene (**1**) to (17*Z*)-protosta-17(20),24-dien-3β-ol (**2**, protostadienol), the biosynthetic precursor of helvolic acid. Described herein is the cloning, heterologous expression, and product characterization of protostadienol synthase.[6] We further identified a gene cluster associated with the metabolism of **2** to helvolic acid and discuss its evolutionary origins.

The genome of *As. fumigatus* strain Af293 contains six annotated (oxido)squalene cyclase genes.[7] We obtained one of these genes (Afu4g14770) by PCR amplification from genomic DNA. Subcloning into the high-copy expression vector pRS426GALR gave the plasmid pXQ4.2, which was used to transform the *Saccharomyces cerevisiae* strain SMY21, a derivative of the lanosterol synthase mutant SMY8.[8] A culture of the resultant SMY21[pXQ4.2] strain yielded a major triterpene product that was purified chromatographically and char-

[†] Department of Chemistry.
[‡] Department of Biochemistry and Cell Biology.
[§] Division of Pathology and Laboratory Medicine.

(1) Tekaia, F.; Latge, J. P. *Curr. Opin. Microbiol.* **2005**, *8*, 385–392.
(2) Maschmeyer, G.; Haas, A.; Cornely, O. A. *Drugs* **2007**, *67*, 1567–1601.
(3) (a) Keller, N. P.; Turner, G.; Bennett, J. W. *Nat. Rev. Microbiol.* **2005**, *3*, 937–947. (b) Fedorova, N. D.; et al. *PLoS Genet.* **2008**, *4*, e1000046 (see Supporting Information for full reference). (c) McDonagh, A.; et al. *PLoS Pathog.* **2008**, *4*, e1000154 (see Supporting Information for full reference). (d) Schoberle, T.; May, G. S. *Adv. Genet.* **2007**, *57*, 263–283. (e) Rementeria, A.; López-Molina, N.; Ludwig, A.; Vivanco, A. B.; Bikandi, J.; Pontón, J.; Garaizar, J. *Rev. Iberoam. Micol.* **2005**, *22*, 1–23.
(4) Larsen, T. O.; Smedsgaard, J.; Nielsen, K. F.; Hansen, M. A. E.; Samson, R. A.; Frisvad, J. C. *Med. Mycol.* **2007**, *45*, 225–232.
(5) Reviews of triterpene biosynthesis: (a) Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305–314. (b) Abe, I. *Nat. Prod. Rep.* **2007**, *24*, 1311–1331.
(6) Meeting announcements by us and others: (a) Ivanova, Y. M.; Xiong, Q.; Matsuda, S. P. T. *American Society of Pharmacognosy Annual Meeting*, Arlington, VA, 2006; p P-013. (b) Reference 5b, p 1329.
(7) Nierman, W. C.; et al. *Nature* **2005**, *438*, 1151–1156 (see Supporting Information for full reference).

10.1021/ol802696a CCC: $40.75
Published on Web 02/13/2009

acterized by GC-MS and NMR. The molecular ion of *m/z* 498 (TMS ether) in conjunction with NMR data indicated a $C_{30}H_{50}O$ triterpene alcohol. This result was consistent with the amino acid sequence, which had motifs characteristic of an oxidosqualene cyclase rather than the squalene cyclase function proposed by machine-generated annotation.[7]

The structure of the product was elucidated from 1D NMR, HSQC, HMBC, COSYDEC, and NOESY spectra. NMR showed two C=C bonds and five upfield methyl singlets, suggesting a tetracyclic triterpene. ^{1}H and ^{13}C NMR chemical shifts (Table 1) were incompatible with NMR data for known

Table 1. ^{1}H and ^{13}C NMR Chemical Shifts for Protostadienol (**2**)[a]

atom	^{13}C	^{1}H α	^{1}H β	atom	^{13}C	^{1}H α	^{1}H β
1	32.96	*1.432*	*1.419*	16	29.27	2.083	2.157
2	29.15	1.717	1.611	17	136.75		
3	79.35	3.249		18	16.93	0.758	
4	39.19			19	22.62	0.933	
5	47.57	1.472		20	126.37		
6	18.33	1.527	*1.210*	21	20.80	1.581	
7	34.70	*1.205*	1.952	22	33.70	2.06	2.07
8	38.89			23	28.06	*1.975*	*2.052*
9	45.36	1.512		24	124.71	5.112	
10	36.73			25	131.13		
11	22.90	1.256	1.486	26	25.75	1.683	
12	27.26	2.212	1.364	27	17.67	1.603	
13	46.83	2.306		28	29.07	0.988	
14	50.61			29	16.07	0.794	
15	30.32	1.465	1.130	30	22.02	1.127	

[a] ^{1}H NMR: ~2 mM in $CDCl_3$, 25 °C, 800 MHz; chemical shifts referenced to $SiMe_4$ and corrected for strong coupling effects; ca. ± 0.001 ppm accuracy (±0.003 ppm for values in italics; ± 0.01 ppm for the C22 protons). ^{13}C NMR: ~5 mM in $CDCl_3$, 25 °C, 125 MHz; referenced to $CDCl_3$ (77.0 ppm); ca. ± 0.03 ppm accuracy. Stereochemical assignments are not given for the C22 and C23 protons.

products of the dammarenyl and baccharenyl cations[9] but closely matched values reported for protosta-20,24-dien-3*β*-ol[10] in rings A and B and part of ring C. Connectivities from COSYDEC and HMBC spectra (Figure 1B, C) confirmed this AB-ring fragment and extended the substructure to a tetracycle bearing a C_8 side chain attached by a Δ17(20) double bond. The ring junction configurations were established from NOESY results (Figure 1D), ^{1}H−^{1}H coupling constants, and chemical shift comparisons (see Supporting Information).

The 17*Z* configuration of the Δ17(20) double bond was anticipated from the shape of the active-site cavity in crystal structures of human lanosterol synthase[11] and from the double bond geometry in fusidane natural products.[12] The

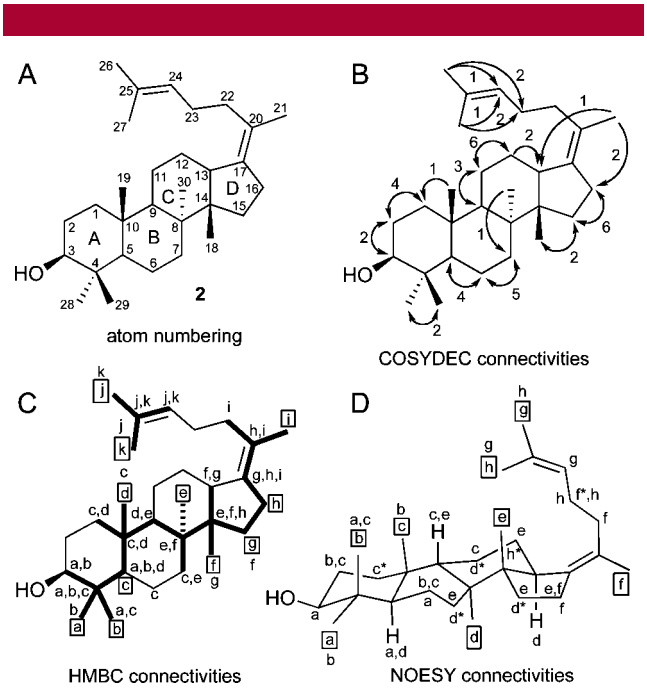


Figure 1. NMR evidence for the structure of **2**. (A) Atom numbering. (B) COSYDEC connectivities indicated by arrows labeled with the number of 2D cross peaks observed. (C) HMBC connectivities, shown from ^{1}H (boxed letters) to ^{13}C (plain letters); thick bonds denote deduced connectivities. (D) NOESY connectivities, shown from ^{1}H methyls (boxed letters) to ^{1}H (plain letters); asterisks denote ambiguous connectivities.

17*Z* geometry was confirmed by the H16/H21 NOESY results (Figure 1D). Thus, the major enzymatic product is (17*Z*)-protosta-17(20),24-dien-3*β*-ol (**2**),[13] and Afu4g14770 encodes a protostadienol synthase (PdsA).

Like other cyclases,[9a] PdsA also makes minor products. The most abundant byproduct, the Δ13(17) isomer **3** (5% of total products), was identified from distinctive 1D and 2D NMR signals that closely matched literature values.[14] Lanosterol, its 14-demethylated Δ8,24 metabolite, and parkeol were detected at 0.1−0.2% levels.[15]

Like the well-studied lanosterol synthases, PdsA cyclizes oxidosqualene to the chair-boat-chair tetracycle. However, PdsA then promotes direct deprotonation to a structure with more ring strain than lanosterol. PdsA consequently is an important example of enzymes that kinetically form an energetically unfavorable product.

(8) Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327–331.

(9) (a) Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova, M. D.; Onak, C. S.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2007**, *129*, 11213–11222. (b) Goad, L. J.; Akihisa, T. *Analysis of sterols*; Blackie (Chapman & Hall): London, 1997; appendix 3, sections 9g and 9h.

(10) Wu, T.-K.; Liu, Y.-T.; Yu, M.-T.; Wang, H.-J. *J. Am. Chem. Soc.* **2006**, *128*, 6414–6419.

(11) Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. *Nature* **2004**, *432*, 118–122.

(12) Cole, R. J.; Schweikert, M. A. *Handbook of Secondary Fungal Metabolites*, Vol 2; Academic Press: Amsterdam, 2003; chapter 16, "Fusidanes and Protostanes", pp 461−492.

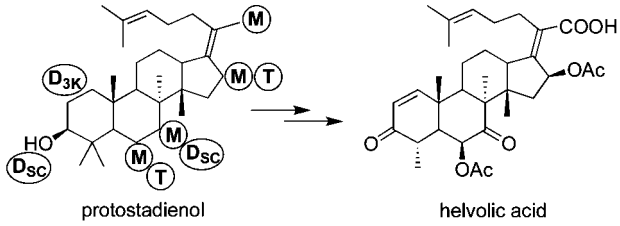
(13) (a) The NMR chemical shifts in Table 1 are consistent with partial ^{1}H NMR data reported for **2** (prepared synthetically from fusidic acid): von Daehne, W.; Godtfredsen, W. O. U.S. Patent 6177418, 2001. (b) **2** and **3** were isolated from *Cephalosporium caerulens* (*Sarocladium oryzae*), and their structures were elucidated from chemical correlations and limited spectral characterization: Hattori, T.; Igarashi, H.; Iwasaki, S.; Okuda, S. *Tetrahedron Lett.* **1969**, 1023–1026.

(14) Wu, T.-K.; Wen, H.-Y.; Chang, C.-H.; Liu, Y.-T. *Org. Lett.* **2008**, *10*, 2529–2532.

(15) Because of metabolism by yeast, the true levels of lanosterol and parkeol formation may be an order of magnitude higher: Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439–442.

PdsA is the cyclase that was postulated[16] to create the distinctive ring skeleton of fusidane antibiotics. The first fusidane, helvolic acid, was isolated from *As. fumigatus* as part of the 1940s war effort to develop antibiotics.[17] The higher potency of penicillin apparently relegated the initial X-ray studies of helvolic acid by Hodgkin (née Crowfoot)[18] to a low priority,[19] and the first crystal structure of a fusidane was not published until 1968.[20] After a flurry of activity in the 1960s and 70s, which included the isolation and structure determination of **2** and **3**,[13b] work on fusidanes languished until the recent interest in their potential role in fungal virulence.[3e,4]

Virulence is associated with secondary metabolism, whose biosynthetic pathways are commonly encoded by gene clusters concentrated in subtelomeric regions of the genome.[3] Cluster boundaries may be refined by examining the expression pattern induced by transcription regulators.[21] One such regulator, LaeA, affects much of secondary metabolism in aspergilli by selectively regulating specific clusters without affecting neighboring genes.[22] A recent microarray study of LaeA regulation[23] in *As. fumigatus* established cluster boundaries for synthesis of gliotoxin, melanin, and ergot alkaloids. Of many annotated[7] clusters that make unidentified products, one cluster spanning genes Afu4g14380 to 14850 at the end of chromosome 4 included *pdsA* (14770). LaeA had modest effects on the expression of ~40 genes preceding 14770 but strongly regulated genes 14770 to 14850. These genes encode putative functions that correspond precisely to the oxidation, dehydrogenation, and acetylation (acyl transfer) steps needed to convert protostadienol to helvolic acid[24] (Figure 2). We designate these nine genes as the helvolic acid cluster.[25]



gene		annotated enzyme function		ΔLaeA/wt
Afu4g14380-14760		[various]		0.6 (MAD)
Afu4g14770	helvolic acid cluster	squalene-hopene-cyclase [sic]		3.4
Afu4g14780		cytochrome P450 monooxygenase	M	2.7
Afu4g14790		cytochrome P450 monooxygenase	M	3.1
Afu4g14800		short chain dehydrogenase	DSC	3.0
Afu4g14810		cytochrome P450 monooxygenase	M	1.7
Afu4g14820		transferase family protein	T	1.5
Afu4g14830		cytochrome P450 monooxygenase	M	1.7
Afu4g14840		transferase family protein	T	1.3
Afu4g14850		3-ketosteroid 1-dehydrogenase	D3K	1.6

Figure 2. Gene cluster proposed for cyclization of oxidosqualene to protostadienol, followed by monooxygenation (M), dehydrogenation (D_{SC} and D_{3K}), and acyl transfer (T) to helvolic acid. The machine-generated gene annotations are from Nierman et al.,[7] who proposed the 48-gene cluster Afu4g14380-Afu4g14850 at the end of chromosome 4. Microarray expression log values are from Perrin et al.[23] and reflect the ratio of expression in a LaeA deletion strain (ΔLaeA) to that in a wild type (wt) strain; MAD, mean absolute deviation.

tBLASTn searches of PdsA sequence revealed orthologs in the close teleomorph *Neosartorya fischeri* and three species outside aspergilli: *Arthroderma gypseum*, its close anamorph *Microsporum canis*, and *Metarhizium anisopliae*.[26] tBLASTn searches of other enzymes in the helvolic acid cluster also located orthologs mainly in these species.[27] Orthology was supported by sequence alignments and phylograms (Figures S12−S17 of the Supporting Information). Moreover, the multiple P450 monooxygenases of each species shared substantial sequence identity, and the transferases displayed similar patterns of paralogous genes. These observations suggested that the clusters evolved by gene duplication and differentiation from an ancestral P450 monooxygenase, a transferase, and two dehydrogenases.

The *As. fumigatus* and *N. fischeri* clusters showed a high degree of synteny (Figure 3A) and were >90% identical. These *Aspergillus* clusters showed limited synteny to the distant *Ar. gypseum* cluster, and sequence identities were <65%. Apparently, duplication and differentiation in the *Ar. gypseum* lineage proceeded independently of that in aspergilli, i.e. after the divergence of the Onygenales (containing *Ar. gypseum*) and Eurotiales (containing aspergilli). *Ar. gypseum* makes fusidic acid,[28] consistent with differences between its gene cluster and the helvolic acid cluster of *As. fumigatus*.

The catalytic and evolutionary origins of *pdsA* were further illuminated by a sequence alignment of 100 fungal cyclases and 16 cyclases representative of other kingdoms. Phylograms (Figure S11, Supporting Information) indicated that the protostadienol synthases (PDSs) are distant from cyclases in animals, plants, protists, and bacteria and only basally related to other fungal cyclases. This isolated status of PDSs is underscored in an alignment containing only active-site residues (Figure 3B). In this alignment, nonconsensus active-site residues of PDSs (shown in bold) are shared only sporadically with other known cyclases.

(16) Godtfredsen, W. O.; Lorck, H. O. B.; Van Tamelen, E. E.; Willett, J. D.; Clayton, R. B. *J. Am. Chem. Soc.* **1968**, *90*, 208–209.
(17) (a) Waksman, S. A.; Horning, E. S.; Spencer, E. L. *Science* **1942**, *96*, 202–203. (b) Chain, E.; Florey, H. W.; Jennings, M. A.; Williams, T. I. *Brit. J. Exptl. Path.* **1943**, *24*, 108–119.
(18) Crowfoot, D. M.; Low, B. W. *Br. J. Exp. Pathol* **1943**, *24*, 120.
(19) Eschenmoser, A.; Arigoni, D. *Helv. Chim. Acta* **2005**, *88*, 3011–3050. This reference points out how earlier knowledge of the fusidane skeleton might have spared much confusion and controversy over the mechanism of triterpene biosynthesis.
(20) Cooper, A.; Hodgkin, D. C. *Tetrahedron* **1968**, *24*, 909–922.
(21) Bouhired, S.; Weber, M.; Kempf-Sontag, A.; Keller, N. P.; Hoffmeister, D. *Fungal Genet. Biol.* **2007**, *44*, 1134–1145.
(22) Bok, J. W.; Balajee, S. A.; Marr, K. A.; Andes, D.; Nielsen, K. F.; Frisvad, J. C.; Keller, N. P. *Eukaryot. Cell* **2005**, *4*, 1574–1582.
(23) Perrin, R. M.; Fedorova, N. D.; Bok, J. W.; Cramer, R. A.; Wortman, J. R.; Kim, H. S.; Nierman, W. C.; Keller, N. P. *PLoS Pathog.* **2007**, *3*, e50.
(24) An earlier 6-keto-7α-acetoxy structure of helvolic acid is given in ref 12 (p 491) and older papers, for example, ref 13b.
(25) Reverse transcriptase RNaseHs found at the end of the cluster include gene models Afu4g14870 and 14880. The latter is affected by LaeA[23] and is upregulated during early infection of mouse lung.[3c]
(26) Helvolic acid was recently isolated from *Me. anisopliae*: Lee, S.-Y.; Kinoshita, H.; Ihara, F.; Igarashi, Y.; Nihira, T. *J. Biosci. Bioeng.* **2008**, *105*, 476–480.
(27) (a) Likewise, Nierman et al.[7] found no orthologs of any helvolic acid pathway genes in *As. oryzae* or *As. nidulans*. (b) *Me. anisopliae* sequence was available only as ESTs, which lack synteny or intron data.
(28) Perry, M. J.; Hendricks-Gittins, A.; Stacey, L. M.; Adlard, M. W.; Noble, W. C. *J. Antibiot.* **1983**, *36*, 1659–1663.

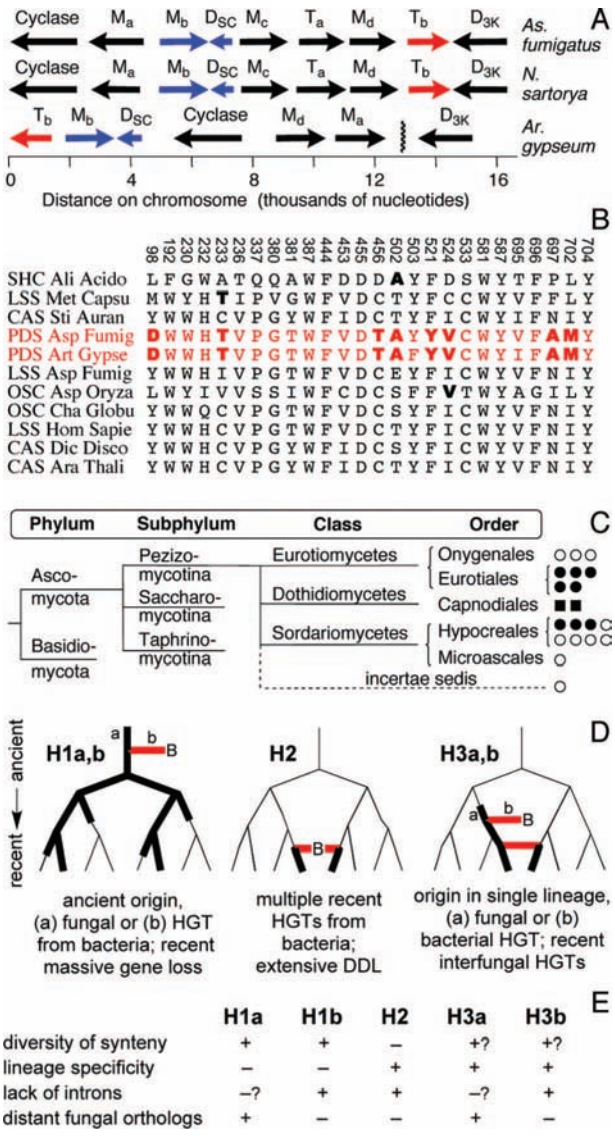


Figure 3. (A) Synteny comparison of three gene clusters encoding fusidane biosynthesis. Enzyme functions M, T, D_{SC}, and D_{3K} are defined in Figure 2. Examples of conserved (blue) and nonconserved (red) synteny are highlighted. (B) Comparison of active-site residues among cyclases (*Homo sapiens* sequence numbering[11]). See Table S8 for gene accession numbers (Supporting Information). (C) Taxonomic relationship of fungi that make helvolic acid (●), fusidic acid (○), or cephalosporin P_1 (■). Each symbol represents one species. (D) Hypotheses to explain the origin of fusidane biosynthesis: red lines, HGT; thick black lines, fungal lineages making fusidanes; B, bacteria. (E) Strengths (+) and weaknesses (−) of each hypothesis.

Fusidanes have been detected in only ~20 fungal species, all in the Pezizomycotina (Figure 3C, Table S14, Supporting Information). Of >100 sequenced fungal genomes, just 5 have homologs in the helvolic acid cluster. These clusters are distinguished by their lack of introns.[29] Among non-Saccharomycotina fungal cyclases in sequenced genomes, the protostadienol synthases are unique in lacking introns.

The helvolic acid cluster in *As. fumigatus* exemplifies the characteristics of lineage-specific subtelomeric genes. The origin of these genes is variously ascribed to horizontal gene transfer (HGT) and/or vertical transmission through gene duplication, rapid diversification, and gene loss (DDL).[3b,30] Guided by previous studies, we consider several hypotheses to account for the origin of fusidane synthesis (Figure 3D).

Hypotheses invoking HGT from bacteria (H1b, H2, H3b) explain the lack of introns in the gene clusters but appear to be at variance with an abundance of distant fungal orthologs of P450 monooxygenases, transferases, and dehydrogenases (see Figure S14, Supporting Information). An ancient origin of fusidane synthesis (H1a,b) accounts for the diverse synteny of the gene clusters, but the postulated selectivity of a recent massive gene loss throughout Pezizomycotina is doubtful. Recent HGTs from bacteria (H2) or other fungi (H3a,b) explain the lineage specificity of fusidane synthesis but marginally address (by DDL) the marked differences in synteny.[31] Hypothesis H3a is quite plausible if the intron loss can be explained as a normal process for subtelomeric clusters[3b] and if the syntenic diversity arose through incomplete transfer of the cluster during interfungal HGT, combined with extensive DDL.

In summary, we have found the gene responsible for protostadienol synthesis in *As. fumigatus* and identified a likely gene cluster for helvolic acid biosynthesis. Our findings may be relevant to the virulence of *As. fumigatus* and to the commercial production of the antibiotics fusidic and helvolic acid.[32]

Acknowledgment. The Robert A. Welch Foundation (C-1323), NIH 5R01AI58045-3, and Herman Frasch Foundation supported this work. The 800 MHz NMR spectra were acquired with instrumentation purchased by the John S. Dunn, Sr. Gulf Coast Consortium for Magnetic Resonance.

Supporting Information Available: Experimental procedures, NMR and GC-MS spectra, NMR chemical shift comparisons, and phylogenetic analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

OL802696A

(29) Fungi were originally rich in introns, which have been partially lost in some lineages and almost fully lost in the Saccharomycotina: Stajich, J. E.; Dietrich, F. S.; Roy, S. W. *Genome Biol.* **2007**, *8*, R223.

(30) (a) Dujon, B.; et al. *Nature* **2004**, *430*, 35–44 (see Supporting Information for full reference). (b) Carbone, I.; Ramirez-Prado, J. H.; Jakobek, J. L.; Horn, B. W. *BMC Evol. Biol.* **2007**, *7*, 111. (c) Khaldi, N.; Wolfe, K. H. *PLoS ONE* **2008**, *3*, e3036.

(31) Also, barring HGT of an unknown bacterial PDS, H2 requires several recent instances of convergent evolution of fungal cyclases. Cycloartenol synthase of *Stigmatella aurantiaca* is a plausible HGT candidate (Figure 3B), although its GC content (68%) is elevated relative to PDSs (55−62%); see Table S10 (Supporting Information). Penicillin synthesis and likely recent HGTs of squalene cyclases in aspergilli provide precedent for H2.

(32) Notably, nonenzymatic blocking of DNA methyl transferase and histone deacetylase can generate an abundance of otherwise silent secondary metabolites: Williams, R. B.; Henrikson, J. C.; Hoover, A. R.; Lee, A. E.; Cichewicz, R. H. *Org. Biomol. Chem.* **2008**, *6*, 1895–1897.