Supporting Information

Andersen et al. 10.1073/pnas.1205532110

SI Materials and Methods

Fungal Growth, Extraction, and Isolation of Nidulanin A. *Aspergillus nidulans* (IBT 22600) was inoculated as three-point stabs on 200 plates of MM and incubated in the dark at 30 °C for 7 d. The fungi were harvested and extracted twice overnight with EtOAc. The extract was filtered and concentrated in vacuo. The combined extract was dissolved in 100 mL of MeOH and H_2O (9:1), and 100 mL of heptane was added after the phases were separated. Eighty milliliters of H_2O was added to the MeOH/H_2O phase, and metabolites were then extracted with 5 × 100 mL of dichloromethane (DCM). The phases were then concentrated separately in vacuo. The DCM phase (0.2021 g) was absorbed onto diol column material and dried before packing into a 10-g SNAP column [coefficient of variation (CV) = 15 mL; Biotage] with diol material. The extract was then fractionated on an Isolera flash purification system (Biotage) using seven steps of heptane-DCM-EtOAc-MeOH. A flow rate of 20 mL·min^{-1} was used, and fractions were automatically collected with 2 × 2 CVs for each step. Solvents used were of HPLC grade, and H_2O was milliQ-water (purified and deionized using a Millipore system through a 0.22-μm membrane filter). Two of the Isolera fractions were subjected to further purification on separate runs on semipreparative HPLC (Waters 600 Controller with a 996-photodiode array detector). This was achieved using a Luna II C_{18} column (250 mm × 10 mm, 5 μm; Phenomenex). A linear water-MeCN gradient was used starting with 15% MeCN and increasing to 100% over 20 min using a flow rate of 4 mL·min^{-1}. MeCN was of HPLC grade, and H_2O was milliQ-water (purified and deionized using the Millipore system through a 0.22-μm membrane filter); both were added to 50 ppm of TFA. The fractions obtained from the separate runs were pooled, and a final purification using the same method yielded 1.5 mg of nidulanin A.

Marfey's Method. Stereoisometry of the amino acids was elucidated using Marfey's method (1). One hundred micrograms of the peptide was hydrolyzed with 200 μL of 6 M HCl at 110 °C for 20 h. To the hydrolysis product (or 2.5 μmol of standard D- and L-amino acids) was added 50 μL of water, 20 μL of 1 M $NaHCO_3$ solution, and 100 μL of 1% 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA) in acetone, followed by reaction at 40 °C for 1 h. The reaction mixture was removed from the heat and neutralized with 10 μL of 2 M HCl, and the solution was diluted with 820 μL of MeOH to a total volume of 1 mL. The retention times of the FDAA derivatives were compared with retention times of the standard amino acid derivatives.

Analysis. Analysis was performed using ultra-high-performance liquid chromatography (UPHLC) UV/Vis diode array detector (DAD) high-resolution MS (HRMS) on a maXis G3 orthogonal acceleration (OA) quadrupole–quadrupole time of flight (QQ-TOF) mass spectrometer (Bruker Daltonics) equipped with an electrospray injection (ESI) source and connected to an Ultimate 3000 UHPLC system (Dionex). The column used was a reverse-phase Kinetex 2.6-μm C_{18}, 100 mm × 2.1 mm (Phenomenex), and the column temperature was maintained at 40 °C. A linear water-acetonitrile gradient was used (both solvents were buffered with 20 mM formic acid) starting from 10% (vol/vol) MeCN and increased to 100% in 10 min, maintaining this rate for 3 min before returning to the starting conditions in 0.1 min and staying there for 2.4 min before the following run. A flow rate of 0.4 mL·min^{-1} was used. HRMS was performed in ESI$^+$ with a data acquisition range of 10 scans per second at *m/z* 100–1,000. The mass spectrometer was calibrated using bruker daltonics high precision calibration (HPC) by means of the use of the internal standard sodium formate, which was automatically infused before each run. UV spectra were collected at wavelengths from 200 to 700 nm. Data processing was performed using DataAnalysis software (Bruker Daltonics). HRMS analysis of nidulanin A was measured to 604.3497 Da corresponding to a molecular formula of $C_{34}H_{45}N_5O_5$ (deviation of −0.6 ppm).

NMR. The 1D and 2D spectra were recorded on a Bruker Daltonics Avance 800-MHz spectrometer equipped with a 5-mm TCI Cryoprobe at the Danish Instrument Centre for NMR Spectroscopy of Biological Macromolecules at Carlsberg Laboratory. Spectra were acquired using standard pulse sequences, and a 1H spectrum, as well as COSY, NOESY, heteronuclear single quantum coherence (HSQC), and heteronuclear multiple bond correlation (HMBC) spectra, were acquired. The deuterated solvent was acetonitrile-d_3, and signals were referenced by solvent signals for acetonitrile-d_3 at δ_H = 1.94 ppm and δ_C = 1.32/118.26 ppm. The NMR data were processed using Topspin 3.1 (Bruker Daltonics). Chemical shifts are reported in parts per million (δ), and scalar couplings are reported in hertz. The sizes of the *J* coupling constants reported in the tables are the experimentally measured values from the spectra. There are minor variations in the measurements, which may be explained by the uncertainty of *J*. NMR data for nidulanin A are presented in Table S3, and the structure is shown in Fig. S6.

Protein Domain Predictions. Nonribosomal peptide synthase (NRPS) protein domains were predicted using the analysis tool of Bachmann and Ravel (2) with the standard settings. Only domains with significant *P* values ($P < 0.05$) were included in the analysis. Adenylation domain specificities were predicted using NRPSpredictor (3).

Structural Elucidation. The 1H NMR spectrum of nidulanin A displayed four resonances at δ_H 8.16, 7.91, 7.64, and 7.51 ppm, which were identified as amide protons indicative of a nonribosomal peptide type of compound. For each resonance, a COSY correlation to a proton further up-field in the α-proton area could be observed. This coupled each of the amide protons to Hα protons at resonances of δ_H 4.82, 3.92, 4.56, and 3.85 ppm, respectively. Investigation of the NOESY connectivities allowed for assembling of the peptide backbone, which revealed a cyclical tetrapeptide as illustrated in Fig. S7.

The two protons at δ_H 7.64 and 4.56 ppm were part of a larger spin system with correlations to a couple of diastereotopic protons at δ_H 3.02 [1H, doublet of doublets (dd), 14.4, 8.0] and 2.82 (1H, dd, 14.3, 7.5) ppm, as well as five aromatic protons at δ_H 7.14 [1H, multiplet (m)], 7.21 (2H, m), and 7.22 (2H, m). HMBC correlations from the diastereotopic pair as well as the aromatic protons revealed a quaternary carbon with a carbon chemical shift of 137.5 ppm. This information, put together, led to the amino acid phenylalanine. The protons at δ_H 7.91 and 3.92 ppm, as well as the protons at δ_H 7.51 and 3.85 ppm, had very similar spin systems. In both spin systems, a single proton appeared (δ_H 1.93 and 1.96, both multiplets), as well as two methyl groups as doublets (δ_H 0.71/0.78 ppm and 0.84/0.79 ppm). In both cases, the amino acid could be established as valine. Elucidation of the final part of the structure showed that this was not one of the standard proteinogenic amino acids. For this final part, three different spin systems, as well as two

isolated methyl groups, were present, which could be linked together by HMBC correlations as well as NOESY connectivities. The first spin system consisted of the amide proton at δ_H 8.16 ppm, the H_α proton at 4.82 ppm, and a diastereotopic pair of protons at δ_H 3.63 (1H, dd, 17.7, 9.7) and 3.09 (1H, dd, 17.6, 4.9) ppm. The second spin system consisted of four aromatic protons at δ_H 7.79 (1H, dd, 8.2, 1.5), 7.28 [1H, doublet of doublets of doublets (ddd), 8.6, 7.0, 1.5], 6.81 (1H, dd, 8.7, 0.7), and 6.57 (1H, ddd, 8.6, 7.0, 1.1) ppm, whereas the third and final spin system contained three protons located in the double-bond area at δ_H 5.95 (1H, dd, 17.6, 10.7), 5.13 (1H, dd, 10.7, 1.0), and 5.15 (1H, dd, 17.6, 1.0) ppm. The latter was shown to be connected to the two methyl groups at δ_H 1.39 [3H, singlet (s)] and 1.38 (3H, s) ppm, and the presence of a quaternary carbon at δ_C 53.7 ppm linked this part as an isoprene unit. The entire residue and key HMBC correlations for the structural elucidation of this part are shown in Fig. S8. The residue contains the amino acid L-kynurenine, which is an intermediate in the tryptophan degradation pathway. In this structure, L-kynurenine has been further modified, because the aforementioned isoprene unit has been incorporated onto the amine located at the aromatic ring.

To establish the stereochemistry of nidulanin A, Marfey's analysis (1) was performed. This technique enables one to determine the absolute configuration of amino acids in peptides (1). The analysis showed the phenylalanine residue present was L-phenylalanine, whereas the analysis for valine showed equal amounts of L- and D-valine.

We used bioinformatics prediction algorithms to identify the stereochemistry of the added amino acids further. Both NRPS protein domain predictions and adenylation domain specificity predictors identify four adenylation domains, corresponding to the four amino acids of the cyclopeptide. By comparison of predictions and the known sequence, the specificity and sequence of the adenylation domains were assigned as predicted to Phe-Kyn-Val-Val. The last two adenylation domains give similar predictions, further supporting both to be specific for valine.

The structure with the proposed absolute chemistry is given in Fig. 4. The absolute configuration of the kynurenine, as well as the order of the L- and D-valine, which is based solely on the bioinformatic studies, has not been verified chemically.

1. Marfey P (1984) Determination of D-amino acids. II. Use of a bifunctional reagent,1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res Commun* 49:591.
2. Bachmann BO, Ravel J (2009) Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol* 458:181–217.
3. Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* 33(18): 5799–5808.

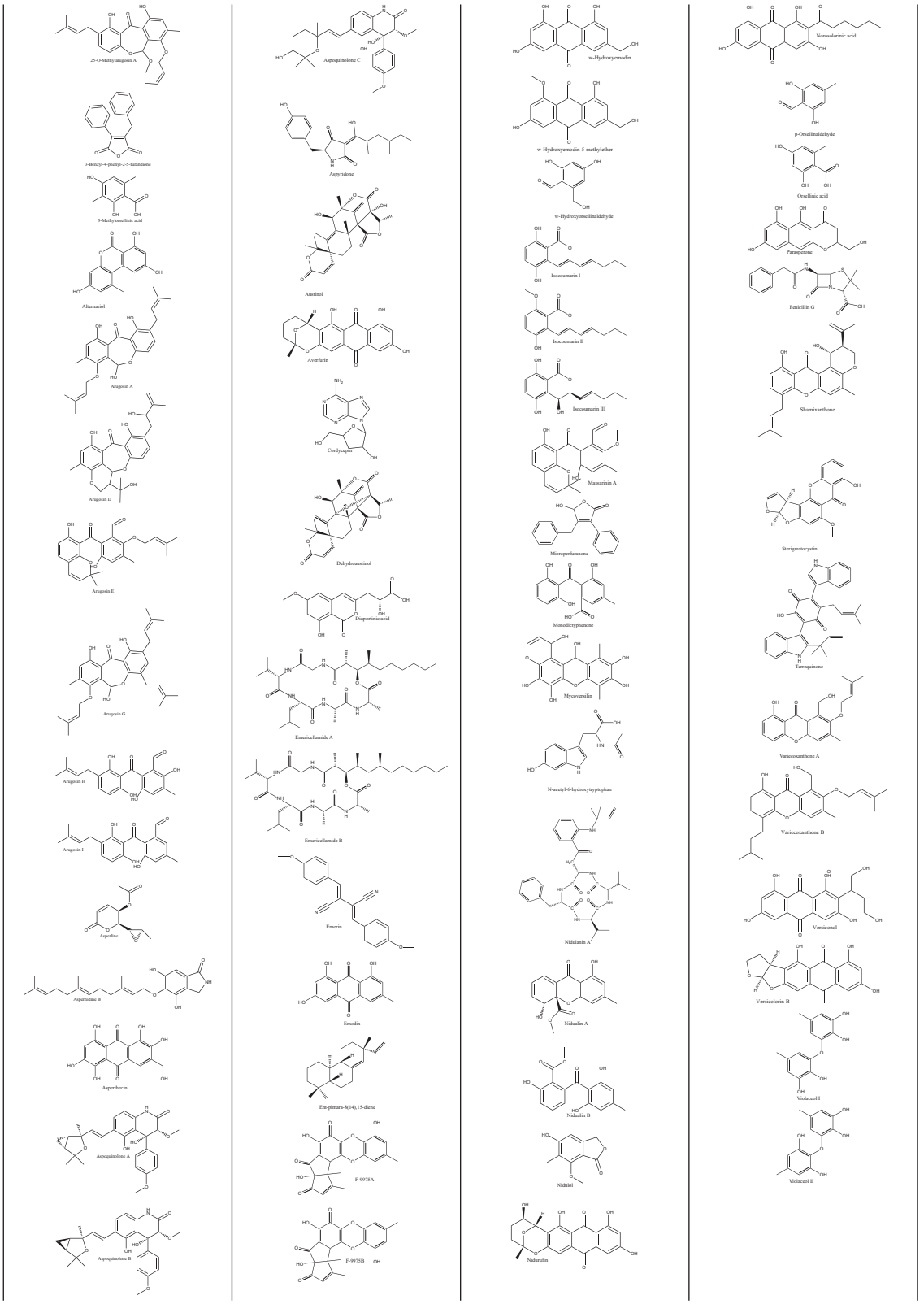

Fig. S1. Chemical structures of secondary metabolites. Structures are shown in alphabetical order in columns from left to right.

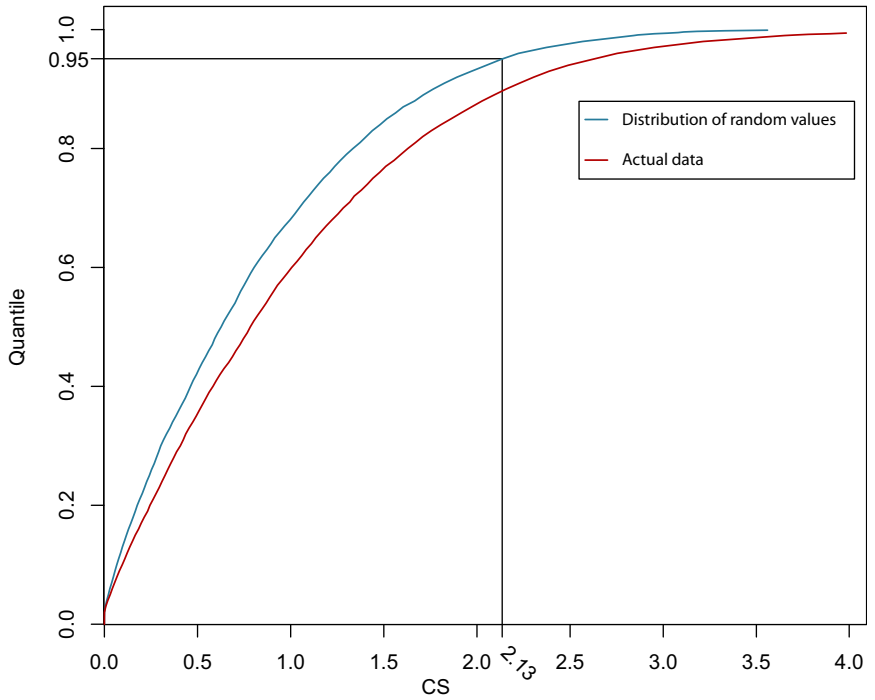

Fig. S2. Quantile plot of clustering scores (CSs). The gray line plots the quantile for a given value of the CS based on a random combination of genes (*Materials and Methods*). Ninety-five percent of the values attained are 2.13 or below (as shown). The red line is a plot of the quantiles of actual values for the genes, as can be found in Dataset S2.

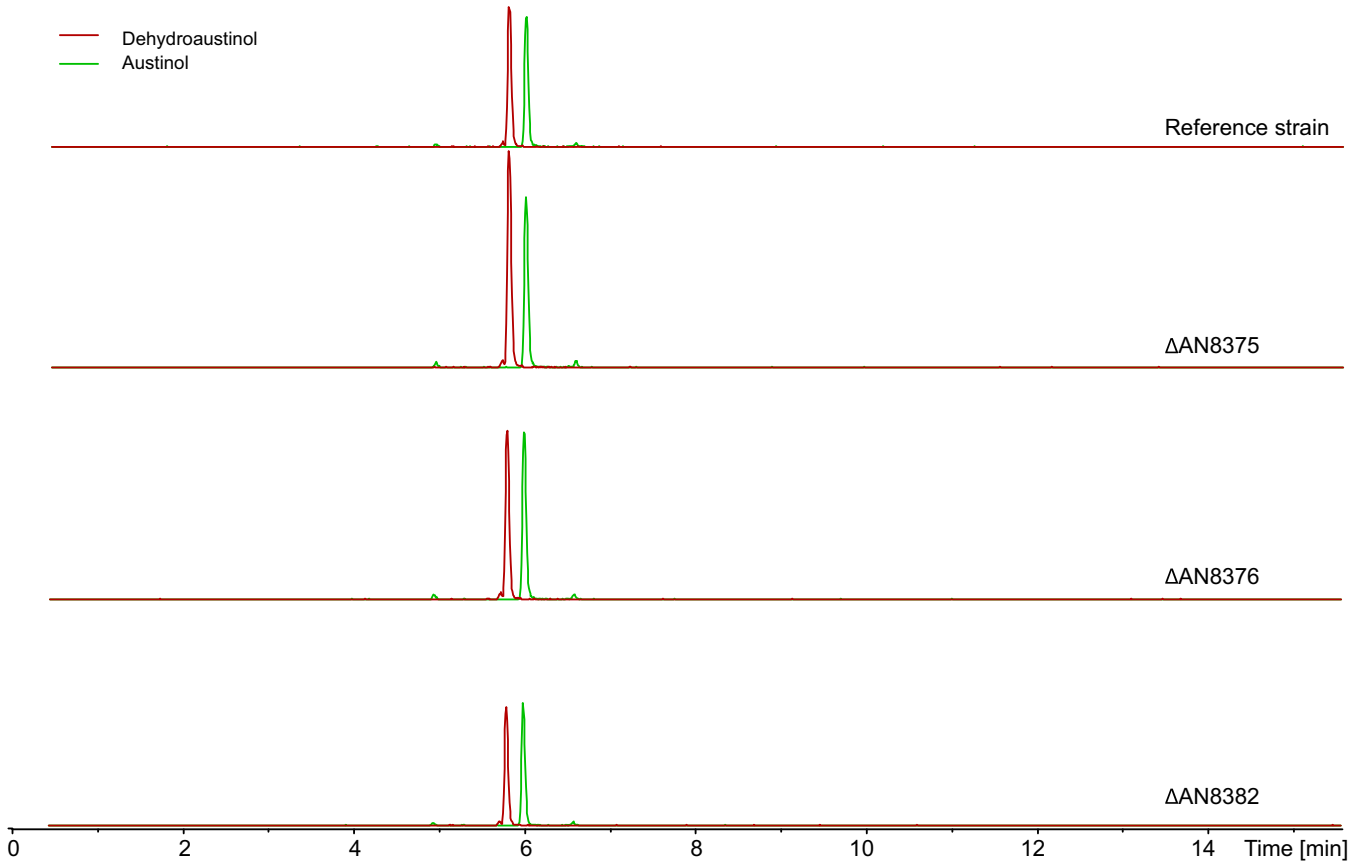

Fig. S3. Extracted ion chromatograms (EICs) for austinol and dehydroaustinol (mass tolerance ± 0.005 Da) from UHPLC-DAD-HRMS of chemical extractions from the reference strain and the ΔAN8375, ΔAN8376 and ΔAN8382 strains. DAD, diode array detector.

Fig. S4. Overview of the gene expression profiles for all predicted members of the biosynthetic gene clusters (Tables 1–3 and Dataset S2). The *y* axis indicates the gene expression index on a $\log_2$ scale, and the *x* axis represents the 44 experimental conditions included in the microarray compendium. The biosynthetic clusters are sorted into the Superclusters indicated in Fig. 3.

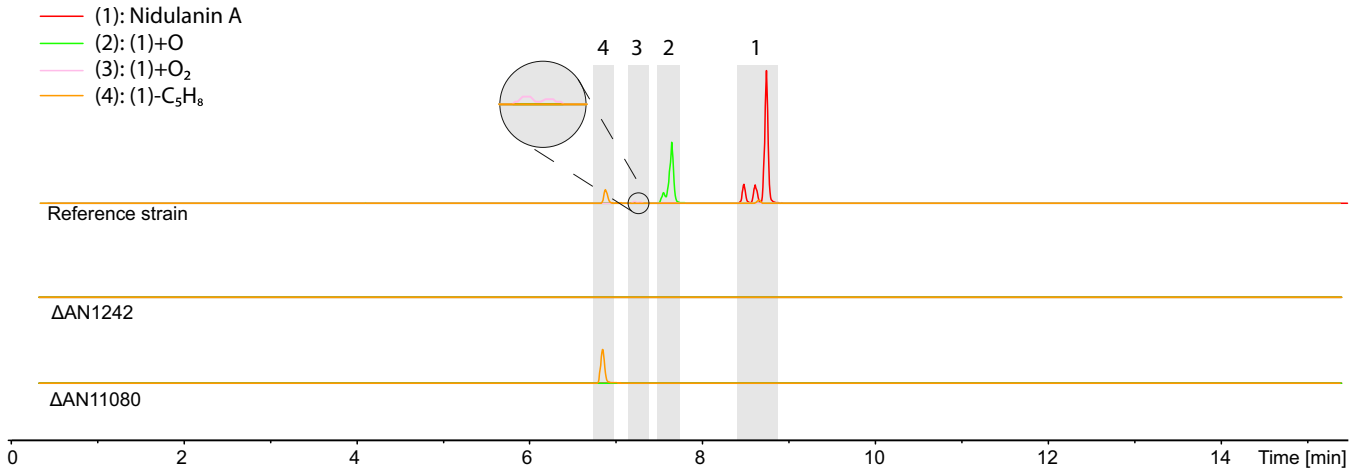

Fig. S5. Extracted ion chromatograms (EICs) for compounds 1–4. Mass tolerance ± 0.005 Da from UHPLC-DAD-HRMS of chemical extractions from the reference strain and ΔAN1242 and ΔAN11080 strains. DAD, diode array detector.

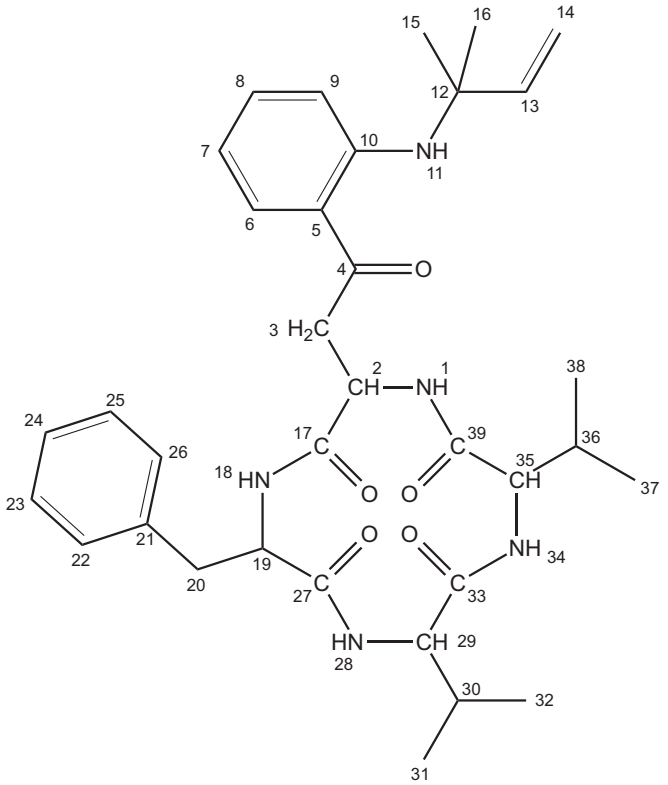

Fig. S6. Structure of nidulanin A, including numbering of individual atoms.

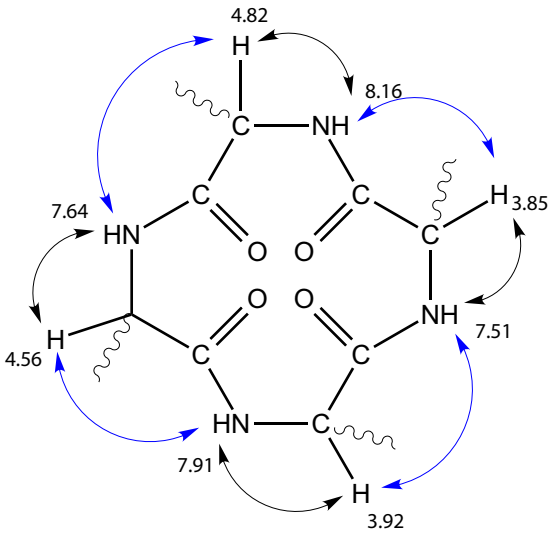

Fig. S7. COSY (black) and NOESY (blue) connectivities lead to the cyclical tetrapeptide.

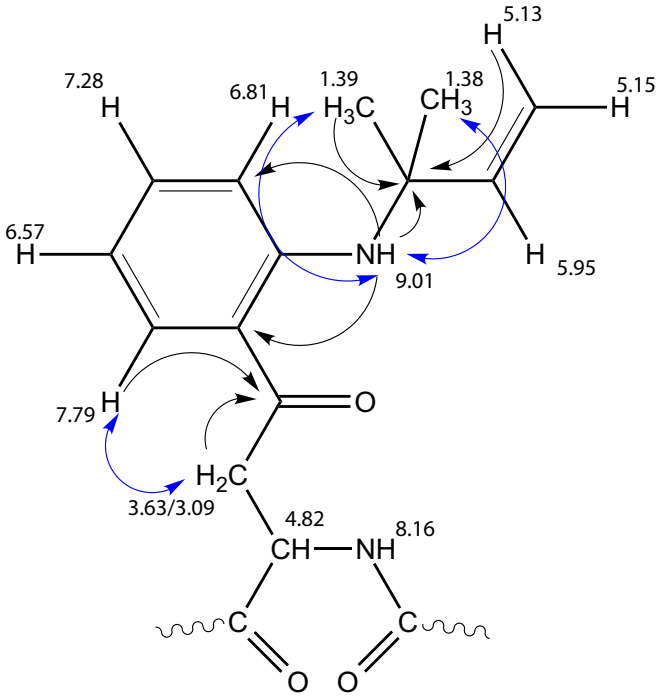

Fig. S8. Key HMBC (black) correlations and NOESY (blue) connectivities link the different spin systems.

Table S1. List of predicted biosynthetic genes where low expression indices with low variation across the 44 conditions were found

GeneID	Type	Gene name	Compound (if known)	Gene no.	Known	Refs.
AN0523	PKS				4	(1)
AN1034	PKS	*afoE*	Asperfuranone	5	7	(2, 3)
AN1036	PKS	*afoG*	Asperfuranone	5	7	(2, 3)
AN10430	PKS					
AN3273	PKS					
AN3386	PKS					
AN3612	PKS					
AN5475	PKS					
AN6961	NRPS					
AN9243	NRPS					
AN9244	NRPS					
AN6810	DTS					

These genes are assumed to be silent in all 44 conditions. DTS, diterpene synthase; NRPS, nonribosomal peptide synthase; PKS, polyketide synthase.

1. Bromann K, et al. (2012) Identification and characterization of a novel diterpene gene cluster in Aspergillus nidulans. *PLoS ONE* 7(4):e35450.
2. Chiang YM, et al. (2008) Molecular genetic mining of the Aspergillus secondary metabolome: Discovery of the emericellamide biosynthetic pathway. *Chem Biol* 15(6):527–532.
3. Bergmann S, et al. (2010) Activation of a silent fungal polyketide biosynthesis pathway through regulatory cross talk with a cryptic nonribosomal peptide synthetase gene cluster. *Appl Environ Microbiol* 76(24):8143–8149.

Table S2. Primer sequences

Primer name	Sequence
AN1242-DL-Up-F	GAGATCGTCGATGGAGTGGCG
AN1242-DL-Up-Rad	gatccccgggaattgccatgCTGCGAGGCACATCATGTTGCC
AN1242-DL-Dw-Fad	aattccagctgaccaccatgGGGTCTGGGTACGCGGGTTTG
AN1242-DL-Dw-R	GATGTGTAGGCGCGACATGGG
AN1242-CHK-Up-F	CCGTCATCATCGTTATAGCC
AN1242-CHK-Dw-R	GCACCCGCTATCACATAC
AN1242-GAPCHK-F	GGCATTATGTGAGCTGTCGTG
AN1242-GAPCHK-R	GATGGAGGGCTTGGTCTTGG
AN1242-INTCHK-R	GATCGAGACGGGTCGTTTAGG
AN11080-DL-Up-FU	GGGTTTAAUGGCAGGTACCAATAATGA
AN11080-DL-Up-RU	GGACTTAAUAGATATACGAGTATGCGG
AN11080-DL-Dw-FU	GGCATTAAUAGTGCCTGATAACTCTGC
AN11080-DL-Dw-RU	GGTCTTAAUGTTGAATCCCTCTGCCTT
AN11080-CHK-Up-F	GGACGGCCCATATTCAGA
AN11080-CHK-Dw-R	AATAAGCTGTAGCGGCGA

Table S3. NMR data for nidulanin A in acetonitrile-d_3

Atom assignment	1H-chemical shift, ppm/J coupling constants, Hz	^{13}C-chemical shift, ppm	HMBC correlations	NOE connectivities
1	8.16 (1H, d, 9.2)	—	—	2, 3, 3′, 35
2	4.82 (1H, ddd, 8.6, 7.0, 1.5)	48.2	—	1, 3, 3′, 18
3	3.63 (1H, dd, 17.7, 9.7)	50.0	2, 4, 17	1, 2, 3′, 6
3′	3.09 (1H, dd, 17.6, 4.9)	50.0	4	1, 2, 3, 6
4	—	198.9	—	—
5	—	116.6	—	—
6	7.79 (1H, dd, 8.2, 1.5)	131.8	4, 8, 10	3, 3′, 7
7	6.57 (1H, ddd, 8.6, 7.0, 1.1)	114.3	5, 9	6
8	7.28 (1H, ddd, 8.6, 7.0, 1.5)	134.0	6, 10	—
9	6.81 (1H, dd, 8.7, 0.7)	114.9	5, 7	15, 16
10	—	148.7	—	—
11	9.01 (1H, s)	—	5, 9, 12	15, 16
12	—	53.7	—	—
13	5.95 (1H, dd, 17.6, 10.7)	145.0	—	14, 14′
14	5.13 (1H, dd, 10.7, 1.0)	113.3	12	13
14′	5.15 (1H, dd, 17.6, 1.0)	113.3	12, 13	13
15	1.39 (3H, s)	27.8	12, 13, 16	9, 11
16	1.38 (3H, s)	27.5	12, 13, 15	9, 11
17	—	172.1	—	—
18	7.64 (1H, d, 8.4)	—	—	2, 19, 20, 20′
19	4.56 (1H, q, 8.0)	53.7	—	18, 28, 20, 20′
20	3.02 (1H, dd, 14.4, 8.0)	*	19, 21, 22/26	18, 19, 20′
20′	2.82 (1H, dd, 14.3, 7.5)	*	19, 21, 22/26, 27	18, 19, 20
21	—	137.5	—	—
22	7.22 (1H, m)	128.0	23/25	—
23	7.21 (1H, m)	127.7	21	—
24	7.14 (1H, m)	126.4	—	—
25	7.21 (1H, m)	127.7	21	—
26	7.22 (1H, m)	128.0	23/25	—
27	—	172.7	—	—
28	7.91 (1H, d, 9.2)	—	—	19, 29, 30, 31
29	3.92 (1H, d, 9.5)	59.1	27, 33	28, 30, 31, 32, 34
30	1.93 (1H, m)	26.3	—	28, 29, 31, 32
31	0.71 (3H, d, 6.6)	18.1	29, 30, 32	28, 29, 30
32	0.78 (3H, d, 6.6)	18.9	29, 30, 31	29, 30
33	—	172.2	172.6	—
34	7.51 (1H, d, 9.6)	—	—	29, 36
35	3.85 (1H, dd, 9.8, 10.7)	59.4	33, 39	1, 36, 37, 38
36	1.96 (1H, m)	26.9	—	34, 35, 37, 38
37	0.84 (3H, d, 6.7)	18.1	35, 36, 38	35, 36
38	0.79 (3H, d, 6.6)	18.9	35, 36, 37	35, 36
39	—	172.1	—	—

1H NMR spectrum and 2D spectra were recorded at with a Bruker Daltonics Avance 800 MHz spectrometer at Carlsberg Laboratory. Signals were referenced to the solvent signals for acetonitrile-d_3 at δ_H = 1.94 ppm and δ_C = 1.32/118.26 ppm. There are minor variations in the measurements which may be explained by the uncertainty of *J*. d, doublet; dd, doublet of doublet; ddd, doublet of doublets of doublets; m, multiplet; q, quartet; s, singlet.
*Cannot be unambiguously assigned.

Dataset S1. Overview of UHPLC-DAD-HRMS analysis of chemical extractions from the reference strain on three solid media after 4, 8, or 10 d (4d, 8d, and 10d, respectively)

Dataset S1

Values given are extracted ion chromatogram peak areas. DAD, diode array detector.

Dataset S2. Gene expression indices from 44 experimental conditions sorted according to chromosomal coordinates

Dataset S2

Locus names and annotation from the *Aspergillus* Genome Database (www.ASPGD.org) are given where available. Clustering scores and cluster members are given.