doi.org/10.26434/chemrxiv.12752459.v1

Structural Determination of an Orphan Natural Product Using Microcrystal Electron Diffraction and Genome Mining

Lee Joon Kim, Masao Ohashi, Dan Tan, Matthew Asay, Duilio Cascio, Jose Rodriguez, Yi Tang, Hosea Nelson

Submitted date: 03/08/2020 • Posted date: 04/08/2020

Citation information: Kim, Lee Joon; Ohashi, Masao; Tan, Dan; Asay, Matthew; Cascio, Duilio; Rodriguez, Jose; et al. (2020): Structural Determination of an Orphan Natural Product Using Microcrystal Electron Diffraction and Genome Mining. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12752459.v1

More than 60% of pharmaceuticals are related to natural products (NPs), chemicals produced by living organisms. Hence, new methods that accelerate natural product discovery are poised to profoundly impact human health. Of the many challenges that remain in natural product discovery, none are as pervasive as structural elucidation, as determination of the molecular structure of a newly discovered natural product can take months, years, or in some cases be altogether unachievable. This challenge can be fueled by lack of sufficient material for spectroscopic analysis, or difficulties in sourcing the producing organism. Even in cases where the analyte is abundant, its physical properties, including molecular structure, can prevent unambiguous structural determination. Here we report the use of microcrystal electron diffraction (MicroED), an emerging cryogenic electron microscopy (CryoEM) technique, in combination with genome mining, to address these challenges. As proof-of-principle, we apply these techniques to fischerin (1), an orphan NP isolated more than 30 years ago, with potent cytotoxicity but ambiguous structural assignment. We utilize genome mining methods to reconstruct its biosynthetic pathway and highlight the sensitivity of MicroED through the precise determination of the solid-state structure of 1 from sub-micron thick crystals. This structural solution serves as a powerful demonstration of the synergy of MicroED and synthetic biology in NP discovery, technologies that when taken together will ultimately accelerate the rate at which new drugs are discovered.

File list (2)

SUPPORTING INFORMATION

Title:

Structural Determination of an Orphan Natural Product Using Microcrystal Electron Diffraction and Genome Mining

Authors:

Lee Joon Kim,[1$] Masao Ohashi,[2$] Dan Tan,[2] Matthew Asay,[1] Duilio Cascio,[1,3] José A. Rodriguez,[1,3] Yi Tang,[1,2*] and Hosea M. Nelson[1*]

Affiliations:

1. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States.
2. Department of Chemical and Biomolecular Engineering and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States.
3. UCLA-DOE Institute for Genomics & Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States.

[$]*These authors contributed equally.*
**Corresponding authors.*

TABLE OF CONTENTS

Supplementary methods

1. Material, fungal strains and culture conditions.

Aspergillus carbonarius NRRL346 was obtained from Agricultural Research Service Culture Collection (NRRL). *A. carbonarius* was maintained on PDA (potato dextrose agar, BD) for 3 days for sporulation or in liquid PDB medium (PDA medium without agar) for isolation of genomic DNA. *A. nidulans* was maintained on Czapek-Dox (CD) agar for sporulation or on CD–ST agar for gene overexpression, compound production and RNA extraction (http://www.fgsc.net/).

2. General DNA manipulation technique.

All DNA manipulation technique was performed as described previously.[1] *E. coli* TOP10 was used for cloning, following standard recombinant DNA techniques as described. DNA restriction enzymes were used as recommended by the manufacturer (New England Biolabs, NEB). PCR was performed using Phusion High-Fidelity DNA Polymerase (NEB). The gene-specific primers are listed in Supplementary Information. PCR products were confirmed by DNA sequencing. *In vivo* homologous recombination using *Saccharomyces cerevisiae* was used for the construction of the *A. nidulans* overexpression plasmids.

3. Preparation of overexpression plasmids for *A. nidulans*.

Full length *finD* was amplified by PCR with three sets of primers of pMO30001-f1/r1, pMO30001-f2/r2, pMO30001-f3/r3 using the genomic DNA of *Aspergillus carbonarius* NRRL346 as the template. The three overlapping DNA fragments and NotI/PacI-digested pYTU[1] expression vector were transformed into yeast to generate pMO30001 by yeast homologous recombination. The genes of *finC* and *finA* were individually amplified with primer pairs of pMO30004-f1/r1 and pMO30004-f3/r3 using genomic DNA of *A. carbonarius* as the template. The glaA promoter was amplified with primer pairs of pMO30004-f2/r2 using pYTU as the template. The three overlapping DNA fragments were ligated into expression vector pYTR[1] in the sites of BamHI/PacI to create plasmid pMO30004 (*finC* and *finA*). The genes of *finJ* and *finI* were amplified with the primer pairs of pMO30009-f1/r1 and pMO30009-f3/r3 using genomic DNA of *A. carbonarius*, respectively. The gpdA promoter was amplified with primer pairs of pMO30009-f2/r2 using pYTR as the template. Each amplified fragment was ligated into expression vector pYTP in the sites of BamHI/HindIII to create plasmid pMO30009 (*finJ* and *finI*). The gene of *finH* was amplified with primer pairs of pMO300010-f2/r2 using genomic DNA of *A. carbonarius* as the template. The amyB promoter was amplified with primer pairs of pMO30010-f1/r1 using pYTP as the template. The two overlapping DNA fragments were ligated into NotI-digested pMO30004 to create plasmid pMO30010 (*finC*, *finA*, and *finH*). The gene of *finE* was amplified with primer pairs of pMO300011-f2/r2 using genomic DNA of *A. carbonarius* as the template. The glaA promoter was amplified with primer pairs of pMO30011-f1/r1 using pYTU as the template. The two overlapping DNA fragments were ligated into SwaI-digested pMO30009 to create plasmid pMO30011 (*finJ*, *finI*, and *finE*).

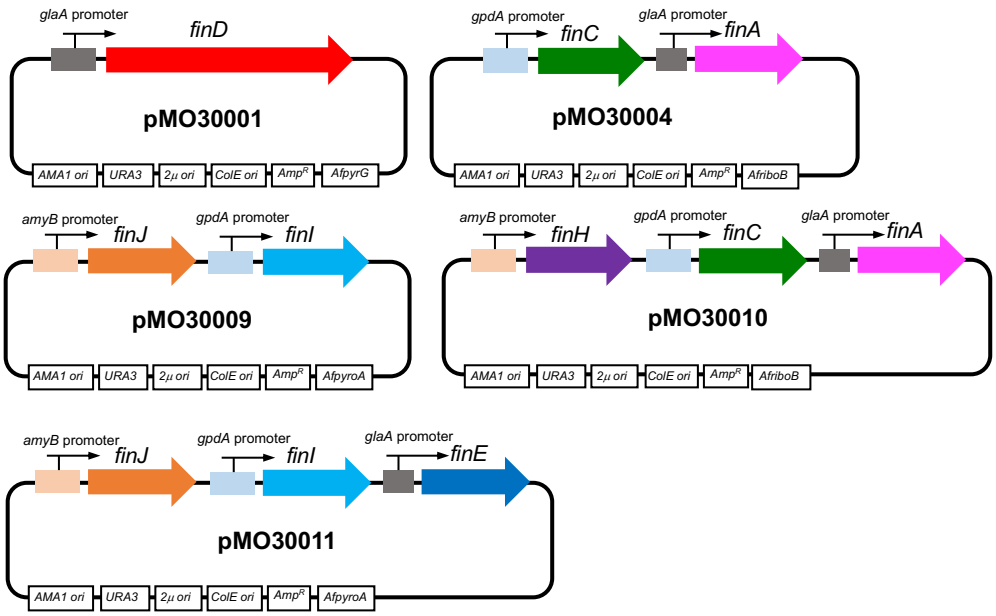


Figure S1. Plasmids used for heterologous expression of *fin* genes in *A. nidulans*.

4. Heterologous expression in *A. nidulans.*

A. nidulans A1145ΔEMΔST[2] was initially grown on oatmeal agar plates containing 10 mM uracil, 0.5 μg/mL pyridoxine HCl and 2.5 μg/mL riboflavin at 30 °C for 5 days. Fresh spores of *A. nidulans* were inoculated into 45 mL liquid CD media (1 L: 10 g glucose, 50 mL 20 x nitrate salts, 1 mL trace elements, pH 6.5) in 125 mL Erlenmeyer flask and germinated at 30 °C and 250 rpm for approximately 16 h. For the preparation of 20 x Nitrate salts, 120 g $NaNO_3$, 10.4 g KCl, 10.4 g $MgSO_4 \cdot 7H_2O$, 30.4 g KH_2PO_4 were dissolved in 1 L double distilled water. The 100 mL trace elements with pH 6.5 contains 2.20 g $ZnSO_4 \cdot 7H_2O$, 1.10 g H_3BO_3, 0.50 g, $MnCl_2 \cdot 4H_2O$, 0.16 g $FeSO_4 \cdot 7H_2O$, 0.16 g $CoCl_2 \cdot 5H_2O$, 0.16 g $CuSO_4 \cdot 5H_2O$, and 0.11 g $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$. Mycelia were harvested by centrifugation at 4000 g for 10 min, and washed with 10 mL Osmotic buffer (1.2 M $MgSO_4$, 10 mM sodium phosphate, pH 5.8). Then the mycelia were transferred into 10 mL of Osmotic buffer containing 100 mg lysing enzymes from *Trichoderma* and 60 mg Yatalase in a 125 mL flask. The flask was kept in shaker at 80 rpm for 4 h at 30 °C. Cells were collected in a 30 mL Corex tube and overlaid gently by 10 mL of Trapping buffer (0.6 M sorbitol, 0.1 M Tris-HCl, pH 7.0). After centrifugation at 3,500 rpm for 15 min at 4 °C, protoplasts were collected in the interface of the two buffers. The protoplasts were then transferred to a sterile 50 mL falcon tube and washed by 10 mL STC buffer (1.2 M sorbitol, 10 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5). The protoplasts were resuspended in 1 mL STC buffer for transformation. Then, the plasmids (see Fig. S1) were added to 60 μL protoplast suspension and the mixture incubated 60 min on ice. Then 600 μL of PEG solution at pH 7.5 (60% PEG, 50 mM $CaCl_2$ and 50 mM Tris-HCl) was added to the protoplast mixture, and the mixture was incubated at room temperature for additional 20 min. The mixture was spread on the regeneration dropout solid medium (CD solid medium with 1.2 M sorbitol and appropriate supplements) and incubated at 30 °C for 2 to 3 days. The transformants were grown on CD-ST (1 L: 20 g starch, 20 g casamino acids, 50 mL 20 x nitrate salts, 1 mL trace elements, 15 g agar) agar for 3 to 4 days before extraction.

5. Analysis of metabolites and isolation of fischerin (1).

For small scale analysis, the transformants of *A. nidulans* were grown for 3 to 4 days on 25 mL CD-ST agar. Then, a chunk of agar with fungal mycelia was scraped and vortexed with 500 µL of acetone for 30 minutes. The organic phase was dried by speed vacuum and dissolved in methanol for analysis. LC-MS analyses were performed on a Shimadzu 2020 EV LC-MS (Kinetex 1.7 µm C18 100 Å, LC Column 100 × 2.1 mm) using positive- and negative-mode electrospray ionization with a linear gradient of 5–95% acetonitrile MeCN–H_2O with 0.5% formic acid in 15 min followed by 95% MeCN for 3 min with a flow rate of 0.3 ml/min.

For isolation of compounds, transformants of *A. nidulans* were grown for 96 h on 4 L CD-ST agar and then solid culture was cut into small pieces and was extracted with acetone. The organic layer was evaporated and the remaining water phase was extracted with 2 x 1 L ethyl acetate. After evaporation of the organic phase, the crude extracts were separated by silica gel chromatography. Fractions containing fischerin (**1**) were combined and used for further purification by HPLC with a COSMOSIL column (Nacalai Tesque Inc., C18 MS-II, 5 µm, 10 × 250 mm, flowrate 4 mL/min). For elucidation of **1**, 1D and 2D NMR spectra were obtained on Bruker AV500 spectrometer at the UCLA Molecular Instrumentation Center. High resolution mass spectra were obtained from an Agilent Technologies 6545 Accurate Mass QTOF LC/MS at the UCLA Molecular Instrumentation Center.

6. Recrystallization of fischerin (1) and austinol (4).

Dried and purified **1** was suspended in tetrahydrofuran (THF), and H_2O was added until the solution became clear and homogeneous, changing color to light yellow. Crystals were grown at room temperature by slow evaporation in a dram vial.

Sample containing **4** was dissolved in 1:1 mixture of MeCN and H_2O, then precipitated at ambient temperature under a stream of air.

7. Serial dilution of 4.

0.4 mg of vacuum-dried sample of **4** was redissolved in a 2:1 MeCN/H_2O solution at 333 µM concentration. The sample was then diluted every 100-fold in H_2O, vortexed, and centrifuged before the subsequent dilution.

8. Electron diffraction data acquisition and processing.

Fischerin (**1**) microcrystals were diluted to optimize crystal density and deposited onto a Quantifoil holey-carbon EM grid as a 2 µL drop before blotting and plunge freezing in liquid ethane using Vitrobot Mark IV (FEI). The vitrified grid was then transferred to the transmission electron microscope (TEM) on Gatan 626 cryo-holder, maintaining cryogenic temperature (100 K) throughout data acquisition.

2 µL of austinol (**4**) sample was aliquotted from the top of the solution to account for crystal density before deposition onto a continuous carbon EM grid. 2 µL of serial dilution sample was obtained from the bottom of the solution and deposited onto a Quantifoil holey-carbon EM grid. In both cases, solvent was wicked away manually, and the dried grid was transferred to the TEM for ambient temperature data acquisition on a single-tilt sample holder.

All diffraction data was collected on FEI Tecnai F200C electron microscope with an operating

voltage of 200 keV, corresponding to a wavelength of 0.025 Å. During data acquisition, the crystal of interest was isolated using a selected area aperture and continuously rotated at a rate of –0.3° s^{-1} over a tilt range of 50–100°. Continuous rotation diffraction data was recorded using rolling shutter mode with a Ceta-D CMOS 4k x 4k camera, integrating at a rate of 3 s per frame and binning by 2 to produce final images of 2k x 2k.[3]

Diffraction movies saved as SER files were converted to SMV format using ser2smv software as described previously.[4] Frames were indexed and integrated in XDS.[5] Data from four and two crystals were scaled and merged together using XSCALE[6] to produce the final data sets for **1** and **4**, respectively. Finally, intensities were converted to SHELX format using XDSCONV.[6]

9. Structure determination and refinement

Structures of **1** and **4** were solved *ab initio* using direct methods in SHELXD[7] and SHELXT,[8] respectively, and refined with SHELXL[9] in ShelXle.[10] All non-hydrogen atoms were refined anisotropically, and hydrogen atoms were placed using the riding model. Residual density corresponding to solvent molecules was observed during refinement; however, solvent disorder hindered unambiguous identification, and oxygen atoms were placed in lieu of water molecules. Overall data quality and refinement statistics are reported in Table S1. Structure overlays were illustrated using Pymol.[11]

Table S1. MicroED data collection and refinement statistics for compounds **1** and **4**.

	Fischerin (1)	**Austinol (4)**
Stoichiometric Formula	$C_{23}H_{29}NO_7$	$C_{25}H_{30}O_8$
Data Collection		
Temperature (K)	100(2)	293(2)
Space group	C2	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	36.59(2), 20.800(9), 18.9(2)	7.910(6), 11.110(8), 23.89(1)
α, β, γ (°)	90.00, 96.3(2), 90.00	90.00, 90.00, 90.00
Resolution (Å)	1.05 (1.10–1.05)	1.00 (1.05–1.00)
Observed reflections (#)	52478 (2220)	6179 (856)

Unique reflections (#)	5789 (621)	1315 (173)
R_{obs} (%)	19.0 (94.6)	16.0 (71.8)
R_{meas} (%)	20.1 (111.2)	18.3 (80.4)
I/σI	5.87 (1.19)	5.20 (1.74)
$CC_{1/2}$ (%)	98.9 (64.9)	98.8 (67.4)
Completeness (%)	86.5 (72.5)	99.3 (100.0)
Refinement		
Resolution (Å)	1.05	1.00
R_1 (%)	13.78	14.69
wR_2 (%)	32.23	33.56
GooF	1.432	1.401

*Highest resolution shell is shown in parenthesis.

10. Data availability

Crystallographic information files (CIF) for compounds **1** and **4** containing atomic coordinates and structure factors have been deposited at the Cambridge Crystallographic Data Center (Deposit number: 2020516 and 2020510, respectively).

Supplementary tables

Table S2. Bioinformatics analysis of *fin and api* gene clusters.

Protein	Size (aa)	Proposed function	Homolog (identity)	Organism	api cluster homolog (identity)[1]
FinA	328	short-chain dehydrogenase/reductase	IccH (39%)	Penicillium variabile	ApiA (41%)
FinB	788	transcriptional factor	ApdR (31%)	Aspergillus nidulans	N/A[2]
FinC	370	enoylreductase	IccB (65%)	Penicillium variabile	ApiC (63%)
FinD	4099	polyketide synthase nonribosomal peptide synthetase	TenS (49%)	Beauveria bassiana	ApiD (59%)
FinE	509	cytochrome P450	TenA (61%)	Beauveria bassiana	ApiE (76%)
FinF	540	transcriptional factor	LepE (31%)	Aspergillus flavus	ApiF (41%)
FinG	441	major facilitator superfamily	ApdF (47%)	Aspergillus nidulans	ApiG (62%)
FinH	662	old yellow enzyme	IccG (54%)	Penicillium variabile	ApiH (51%)
FinI	402	*O*-methyltransferase	EqxD (33%)	Fusarium heterosporum	ApiI (55%)
FinJ	521	cytochrome P450	IccF (51%)	Penicillium variabile	ApiJ (78%)

[1]: % identity with *fin* homolog.

[2]: homolog not found in *api* BGC.

Table S3. Primers used in this study.

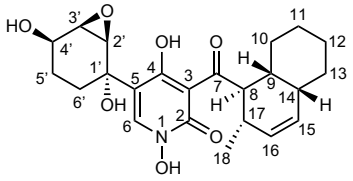
Primers	Sequences (5' → 3')
pMO30001-f1	CTGAGCTTCATCCCCAGCATCATTACACCTCAGCAATGGAGACTCTAGCAAGCTGTCG
pMO30001-r1	AAGTCTTCGGCCTGTTGGAGAAC
pMO30001-f2	AGGGGCTGACGCTCAACATGA
pMO30001-r2	AAGCACATCCGCAGACACTCG
pMO30001-f3	AATTTGGTTCCTCAGCAAGCATGC
pMO30001-r3	TGGAGGACATACCCGTAATTTTCTGGGCATTTAAATTGTCCCTCCTAGCATTTATCCACC
pMO30004-f1	TAACCATTACCCCGCCACATAGACACATCTAAACAATGACCTTGTCCGTCCCCC
pMO30004-r1	AGGTCGCCAGGTACGACCAGTTCGGAAGATCAGGCGGGATTGCGTGTTATACTTACTCAG
pMO30004-f2	GGGGCTGAGTAAGTATAACACGCAATCCCGCCTGATCTTCCGAACTGGTCGTA
pMO30004-r2	AGGATACTGCCTTGCAGGGGAGAGAGGGGTGCCATTGCTGAGGTGTAATGATGCTGG
pMO30004-f3	CAATTAAACCACTCCACACCCTGC
pMO30004-r3	AGGGTATCATCGAAAGGGAGTCATCCAATTTAAATTTGAACGGCTTTGAAACCACGCTG
pMO30009-f1	CCTTCTCTGAACAATAAACCCCACAGAAGGCATTTATGGCCATCATCGACCTCCTCA
pMO30009-r1	GGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTCATAGCAGCATCCGCCTCTTGT
pMO30009-f2	ATTGATTTGGGTGACGGGAGAG
pMO30009-r2	TGTTTAGATGTGTCTATGTGGCG
pMO30009-f3	TAACCATTACCCCGCCACATAGACACATCTAAACAATGGCGTCGGATGCCAGTGTA
pMO30009-r3	TGAGACCCAACAACCATGATACCAGGGGATTTAAATTCTTGATTGCGGTGGAGCTCC
pMO30010-f1	AGTAACCTCGCGGGTGTTCTTGACGATGGCATCCTGATTAAAGGTGCCGAACGAGCTA
pMO30010-r1	AGGGTAAGAGGCTGAGCAATGTTCAGATGGGACATAAATGCCTTCTGTGGGGTTTATTG
pMO30010-f2	ATGTCCCATCTGAACATTGCTCAGCC
pMO30010-r2	GGTCTCTCCCGTCACCCAAATCAATTCACCGGAGTTTAGCGGTGCGATGGCAGTGTTTG
pMO30011-f1	CTGCAACTGCGTCCGGAGCTCCACCGCAATCAAGACCTGATCTTCCGAACTGGTCGTAC
pMO30011-r1	GCGAGAACACCAAACACCGCGGCGGGTAGGATCATTGCTGAGGTGTAATGATGCTGGG
pMO30011-f2	ATGATCCTACCCGCCGCGGTGTTTG
pMO30011-r2	GAGACCCAACAACCATGATACCAGGGGATTTAAATCCTCCTTTTGTCCTCCCTCTGCTGG

Table S4. Plasmids used in this study.

Plasmids	Vector	Genes
pMO30001	pYTU	*finD* (PKS-NRPS)
pMO30010	pYTR	*finC* (ER)-*finA* (SDR)-*finH* (OYE)
pMO30011	pYTP	*finE* (P450)-finI (O-MT)-*finJ* (P450)

Table S5. Spectroscopic data of compound **1.**

^{1}H NMR spectrum (500 MHz), ^{13}C NMR spectrum (125 MHz), in $CDCl_3$



reported[2]			this study		
position	δ_H, mult (J in Hz)	δ_c	position	δ_H, mult (*J* in Hz)	δ_c
1-OH	10.0, br s		1-OH	9.94, br s	
2		157.4	2		157.0
3		107.4	3		107.2
4		173.2	4		173.0
5		113.8	5		113.5
6	7.93, s, s	134.2	6	8.01, s	133.4
7		215.6	7		215.6
8	4.68, t (10.2)	48.9	8	4.76, dd (10.0, 10.0)	48.9
9	2.24, dd (10.2, 3.9)	38.1	9	2.31, m	38.1
10	1.39, m	29.0	10	1.47, m	29.0
11	1.21, m	22.0	11	1.28, m	22.0
12	1.21, m	26.1	12	1.28, m	26.1
	1.60, m			1.68, m	
13	1.21, m	30.2	13	1.28, m	30.2
	1.60, m			1.68, m	
14	2.02, m	37.0	14	2.09, m	37.0
15	5.59, ddd (9.9, 4.7, 2.8)	131.3	15	5.66, ddd (10.0, 5.0, 2.5)	131.2
16	5.40, dt (9.9, 1.7)	131.2	16	5.47, ddd (10.0, 2.0, 2.0)	131.2
17	2.52, m	37.6	17	2.60, m	37.5
18	0.91, d (7.2)	19.8	18	0.99, d (7.0)	19.8
1'		70.6	1'		70.6
2'	3.52, d (3.9)	58.7	2'	3.59, d (4.0)	58.7
3'	3.50, t (3.9)	56.2	3'	3.58, dd (7.5, 3.5)	56.1
4'	4.15, dd (8.0, 3.9)	63.8	4'	4.22, br s	63.7
5'	1.92, m	25.7	5'	2.01, m	25.7
	1.45, m			1.53, m	
6'	1.82, m	28.4	6'	1.90, m	28.3
1'-OH	4.15, s		1'-OH		
4'-OH			4'-OH		

HRMS (ESI, MH^+) calculated for $C_{23}H_{30}NO_7$ 432.2017, found 432.2030.

$[\alpha]_D^{20}$ –28 (*c* = 0.10, $CDCl_3$), reported $[\alpha]_D^{26}$ –65 (*c* = 0.52, $CDCl_3$)[12].

Supplementary Figures

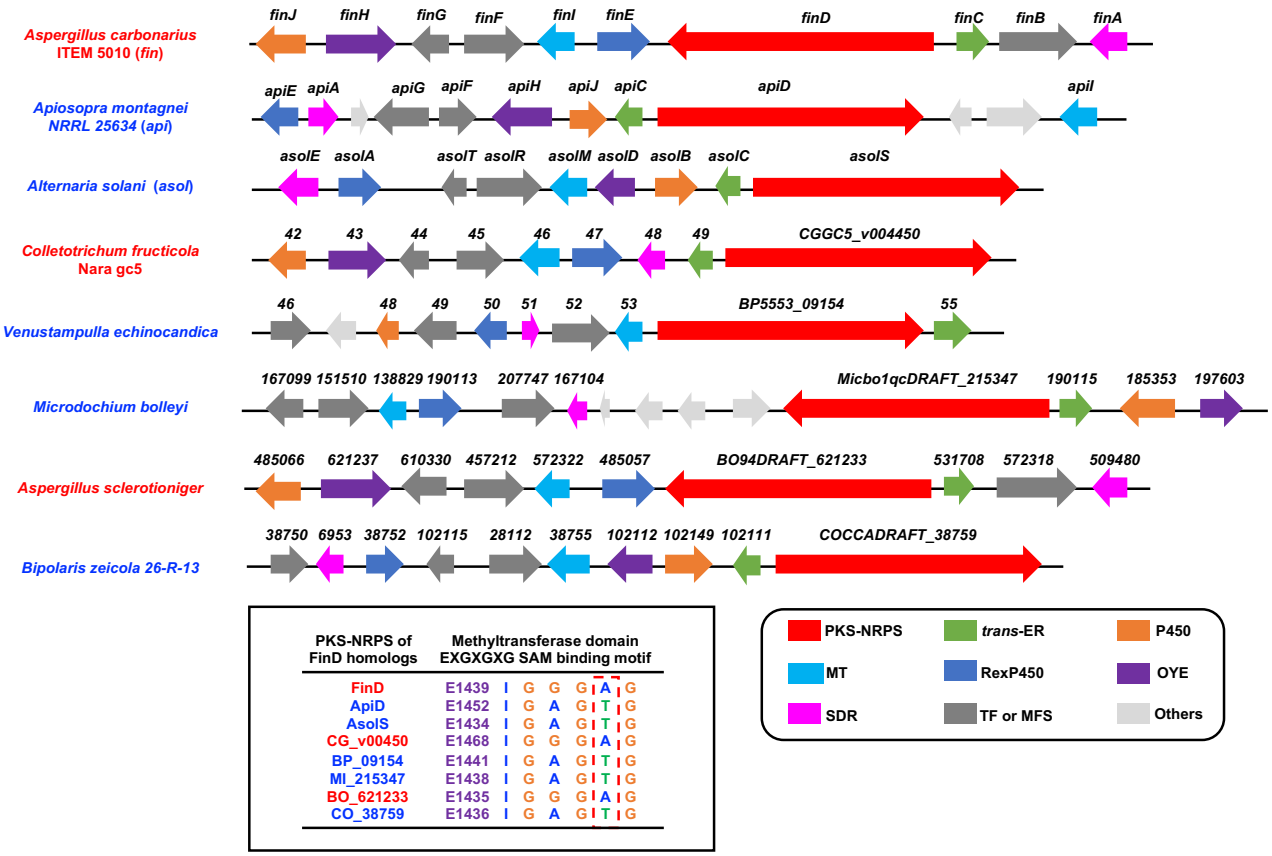


Figure S2. Biosynthetic gene clusters that are homologous to those of fischerin (**1**) and N-hydroxyapiosporamide (**2**); and multiple sequence alignment of SAM-binding motif of *cis*-MT domain in PKS-NRPSs.

Shown here are the putative biosynthetic gene clusters of **1** and **2** and their homologous biosynthetic gene clusters found in NCBI database. SAM binding motif is shown in an alignment with those from FinD and ApiD homologs. Previous study reported that active *cis*-MT domains contain conserved EXGXGTG sequence as a SAM binding motif.[2] Based on this, we hypothesized that the *cis*-MT domains in PKS-NRPSs which do not have this conserved this motif are inactive, and the biosynthetic gene clusters which contain the PKS-NRPSs could be responsible for formation of **1**. For example, the *cis*-MT domains from FinD (*Aspergillus carbonarius*), CG_v00450 (*Colletotrichum fructicola* Nara gc5), and BO_621233 (*Aspergillus sclerotioniger*) do not contain this conserved EXGXGTG motif as the threonine residues are mutated to alanines. As shown in Fig. 1d, the biosynthetic gene cluster, which contains FinD, is indeed responsible for formation of **1**.

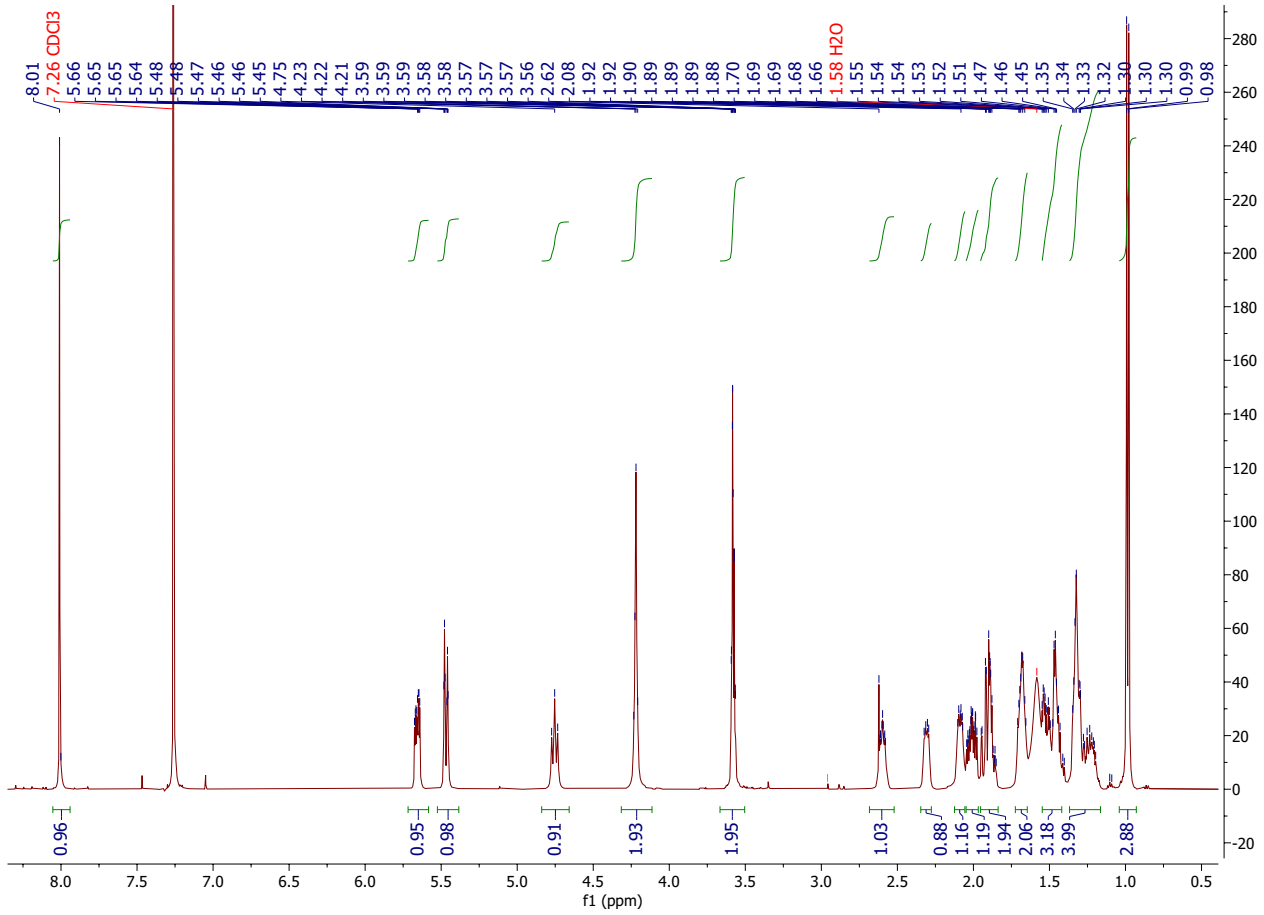


Figure S3. ^{1}H NMR spectra of **1** in $CDCl_3$, 500 MHz for ^{1}H NMR.

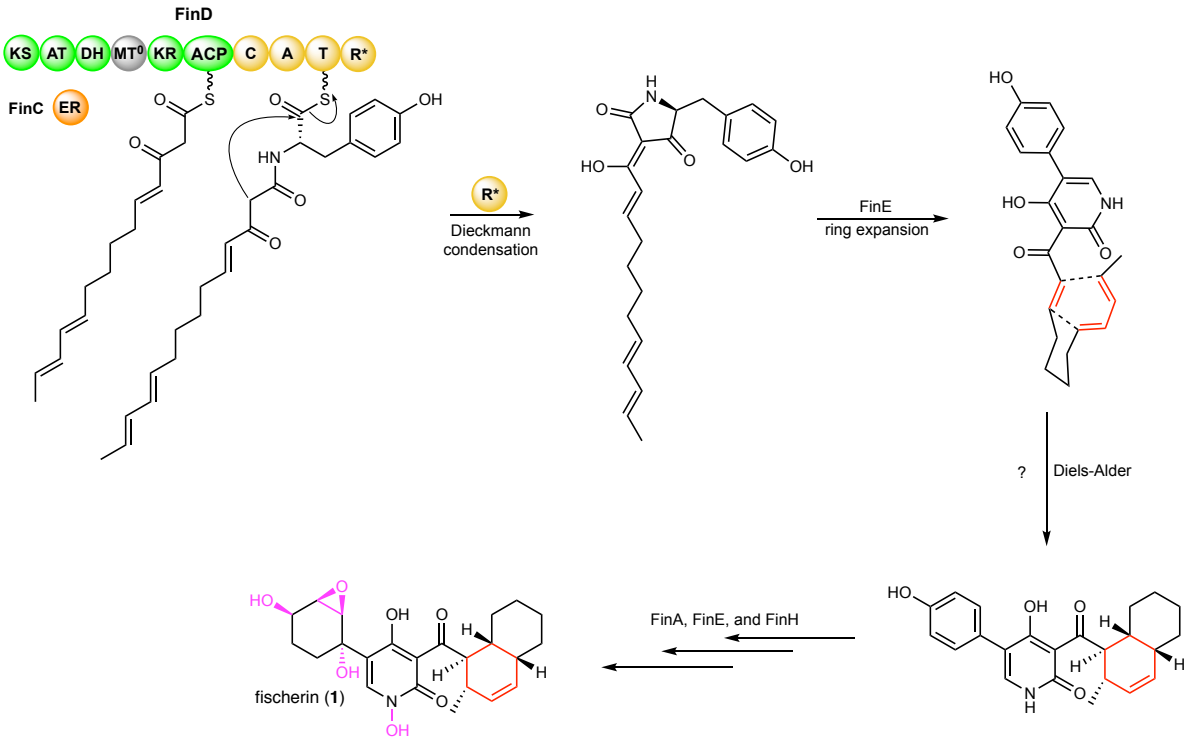


Figure S4. Proposed biosynthetic pathway of **1**.

Based on the reported proposed biosynthetic pathway of other 2-pyridone alkaloids such as leporins and ilicicolin H, we proposed the biosynthetic pathway of **1**. FinD (PKS-NRPS) and the partnering FinC (ER) form the tetramic acid intermediate. A P450 FinE catalyzes the oxidative ring-expansion reaction of the tetramic acid to the 2-pyridone compound. Then, a Diels-Alderase likely catalyze the Diels-Alder reaction to form the energetically disfavored *cis*-decalin ring, since the previous study (ref 7) showed that nonenzymatic Diels-Alder reaction of the analog of 2-pyridone compound in water only led to formation of the *trans*-decalin compound. Further redox modification by FinA, FinE, and FinH forms **1**.

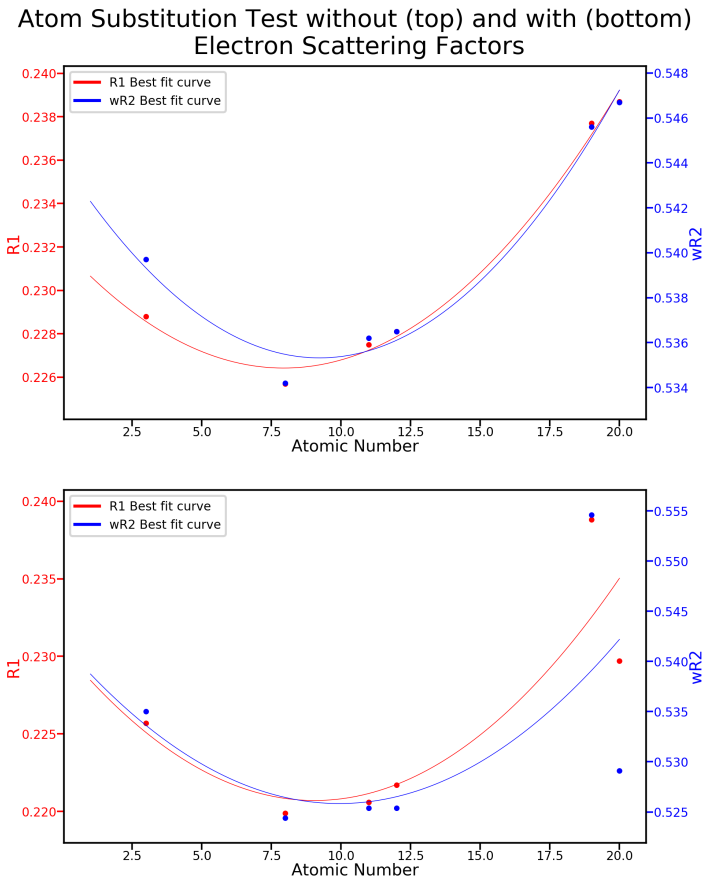


Figure S5. Atom substitution test for fischerin with (top) and without (bottom) electron scattering factors.

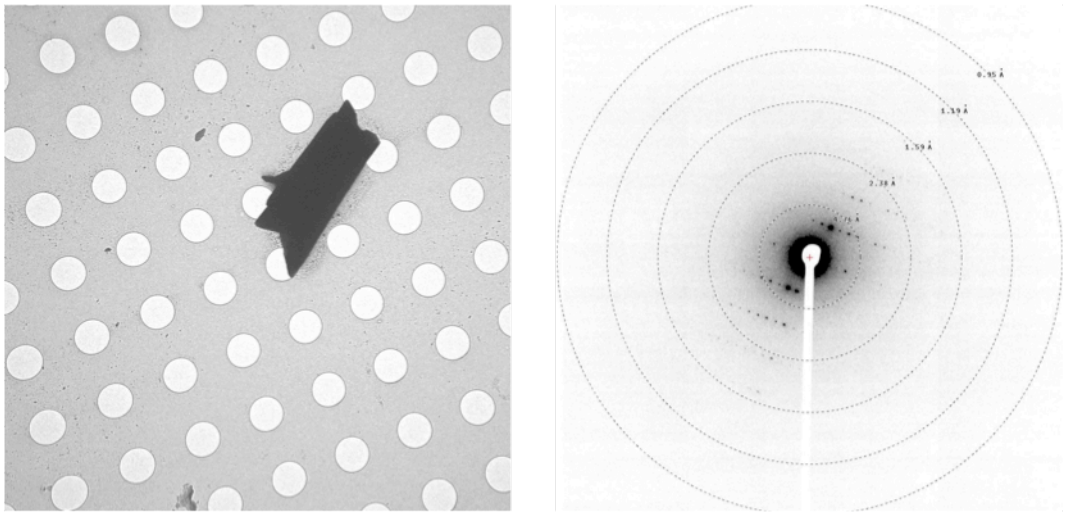


Figure S6. Electron micrograph of austinol crystal (left) and its diffraction pattern (right) from 3 ng of sample; holes are 1 μm wide in diameter.

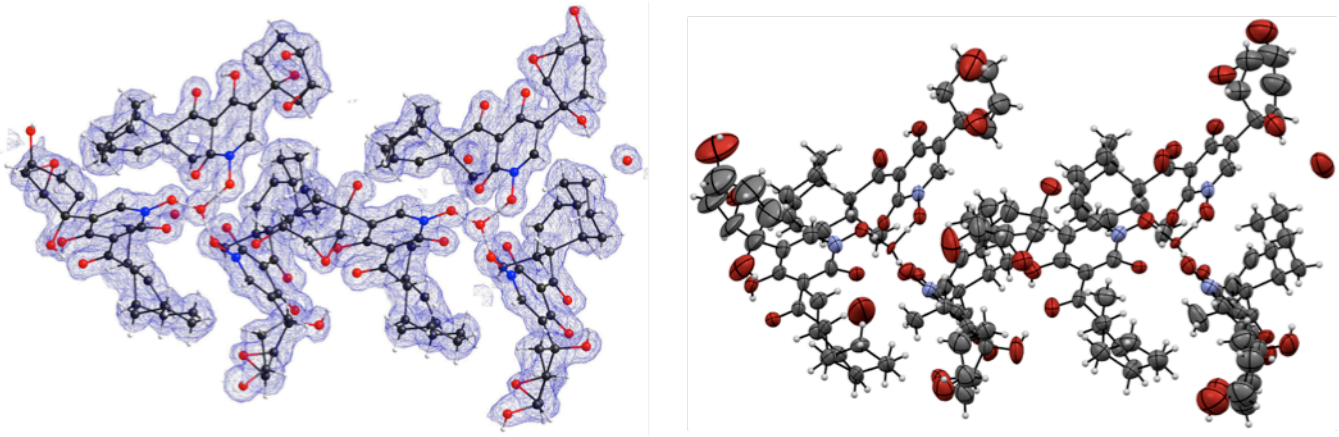


Figure S7. Asymmetric unit of fischerin (**1**) crystal structure. The structure is shown as ball-and-stick model overlaid with electron potential map (left) and ORTEP diagram (right). Thermal ellipsoids are drawn at 30% probability.

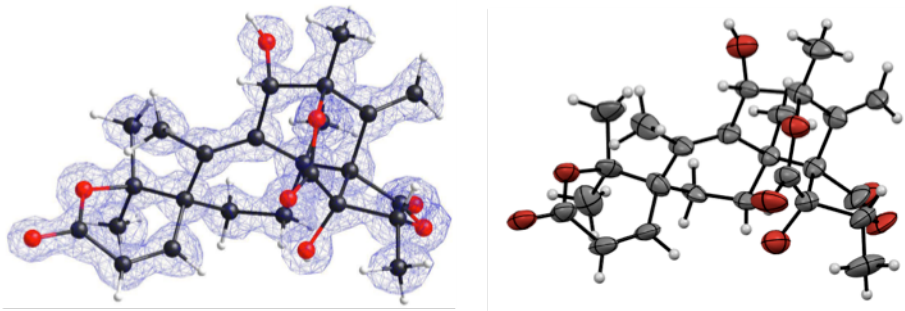


Figure S8. Asymmetric unit of austinol (**4**) crystal structure. The structure is shown as ball-and-stick model overlaid with electron potential map (left) and ORTEP diagram (right). Thermal ellipsoids are drawn at 30% probability.

NMR spectra

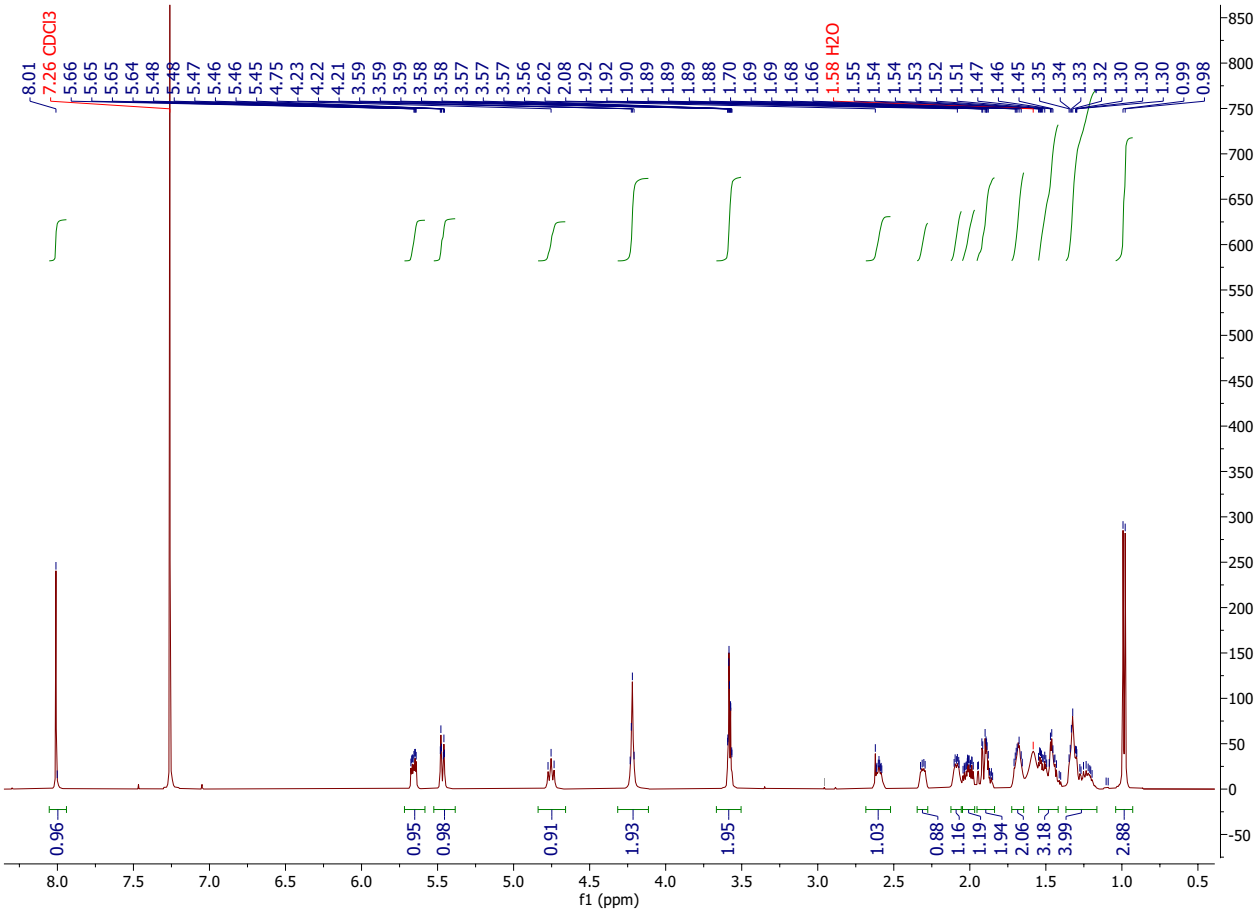


Figure S9. ^{1}H NMR of compound **1** in $CHCl_3$, 500 MHz for ^{1}H NMR.

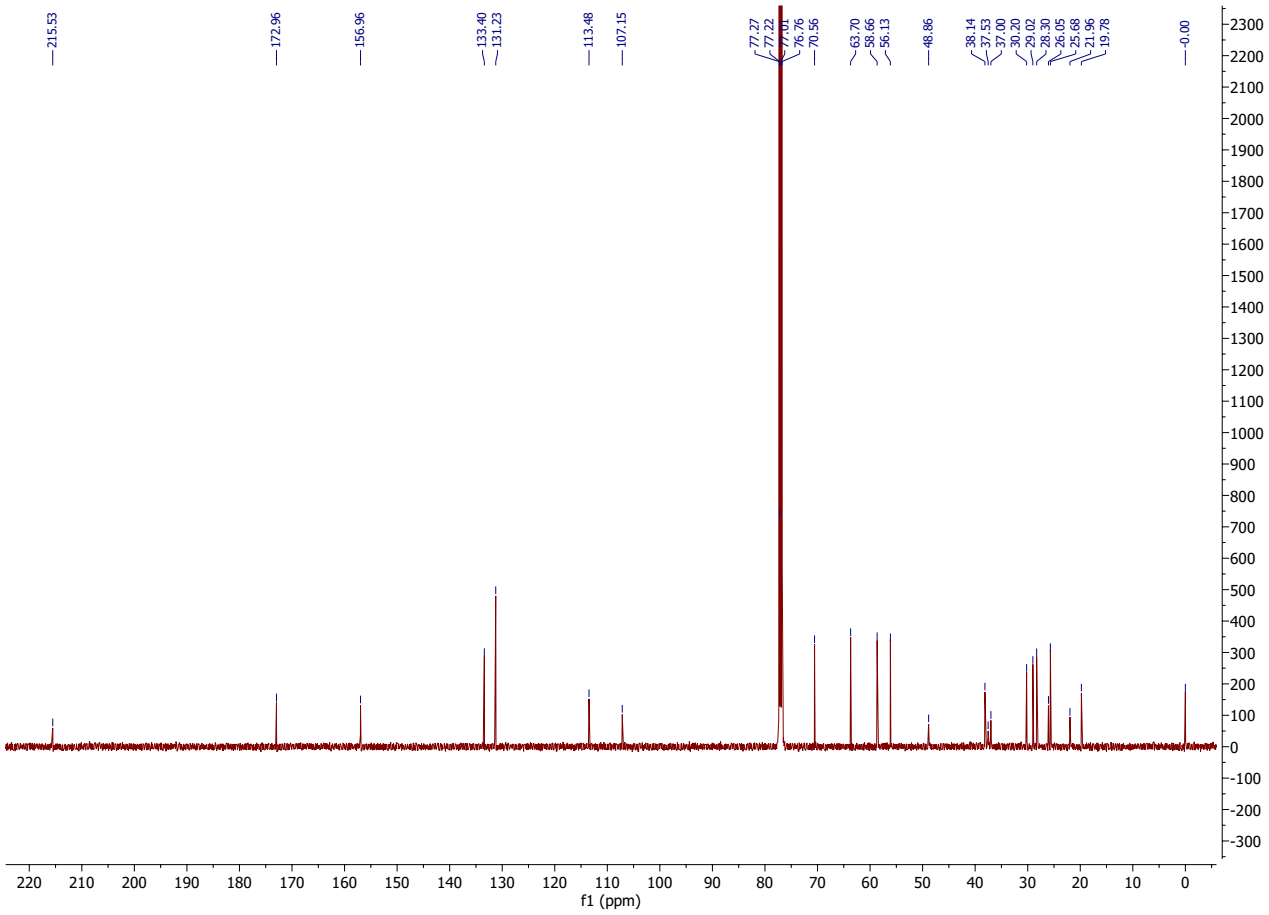


Figure S10. ^{13}C NMR of compound **1** in $CHCl_3$, 125 MHz for ^{13}C NMR.

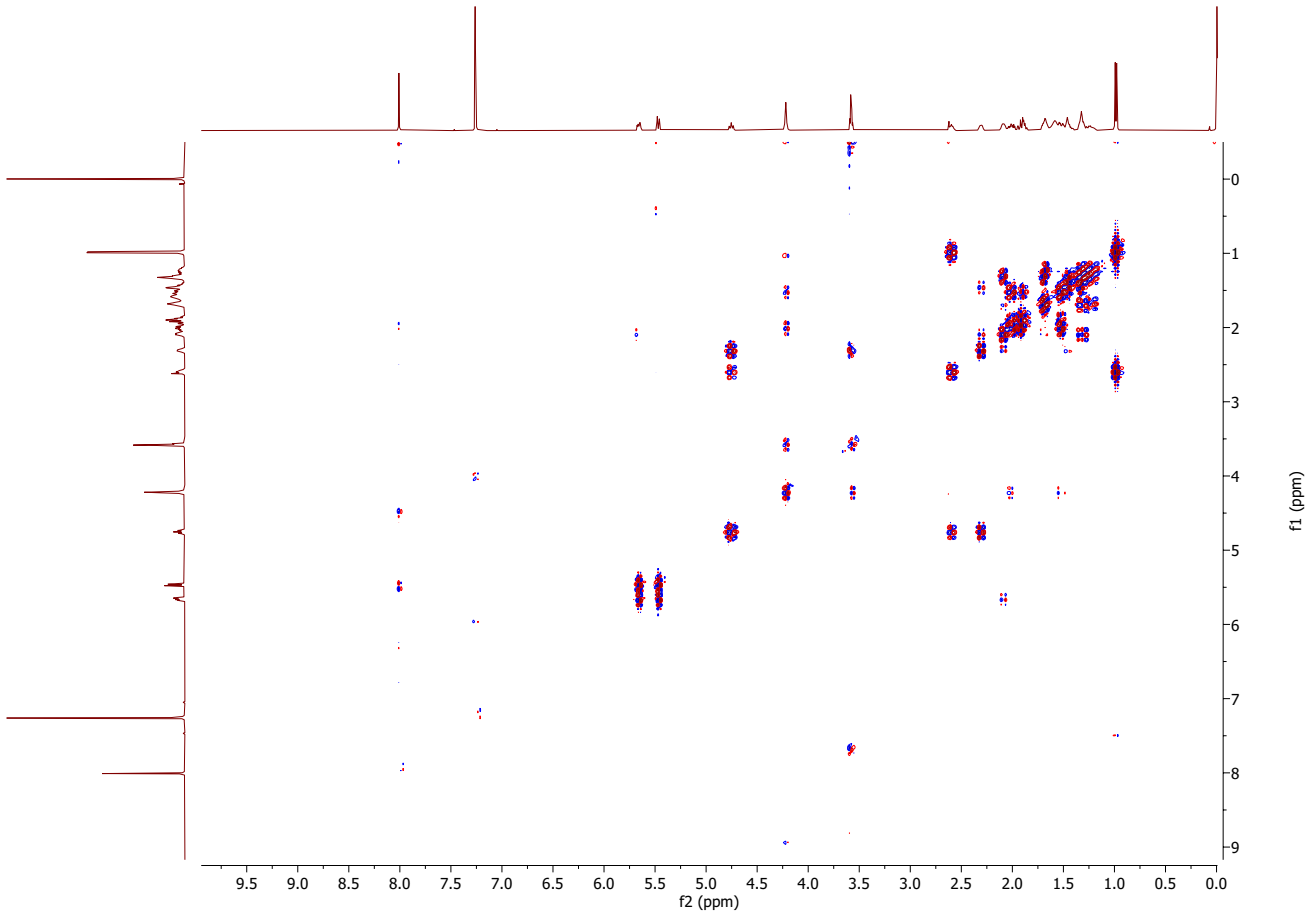


Figure S11. ^{1}H-^{1}H COSY of compound **1** in $CDCl_3$, 500 MHz for ^{1}H NMR.

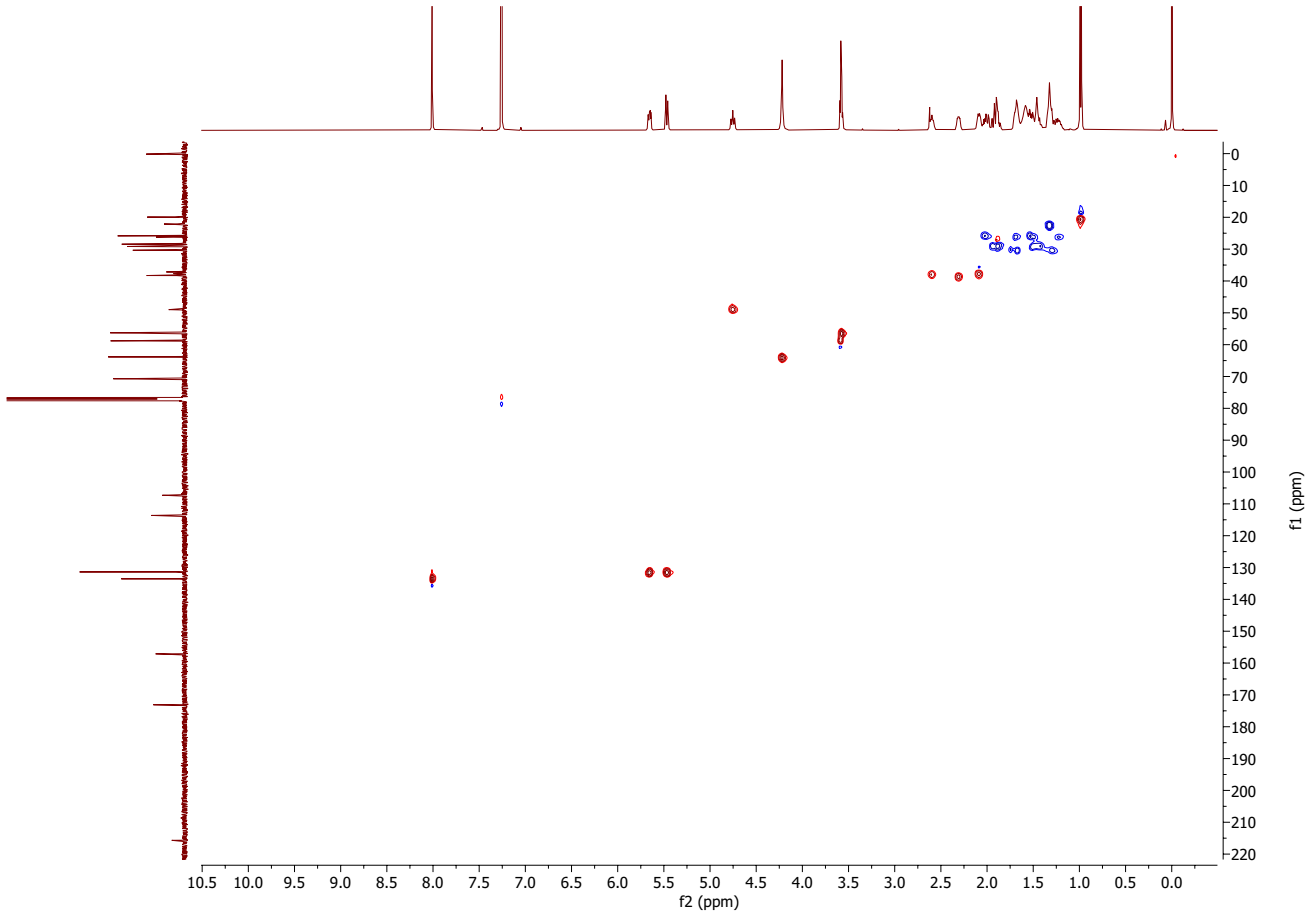


Figure S12. ^{1}H-^{13}C HSQC of compound **1** in $CDCl_3$, 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR.

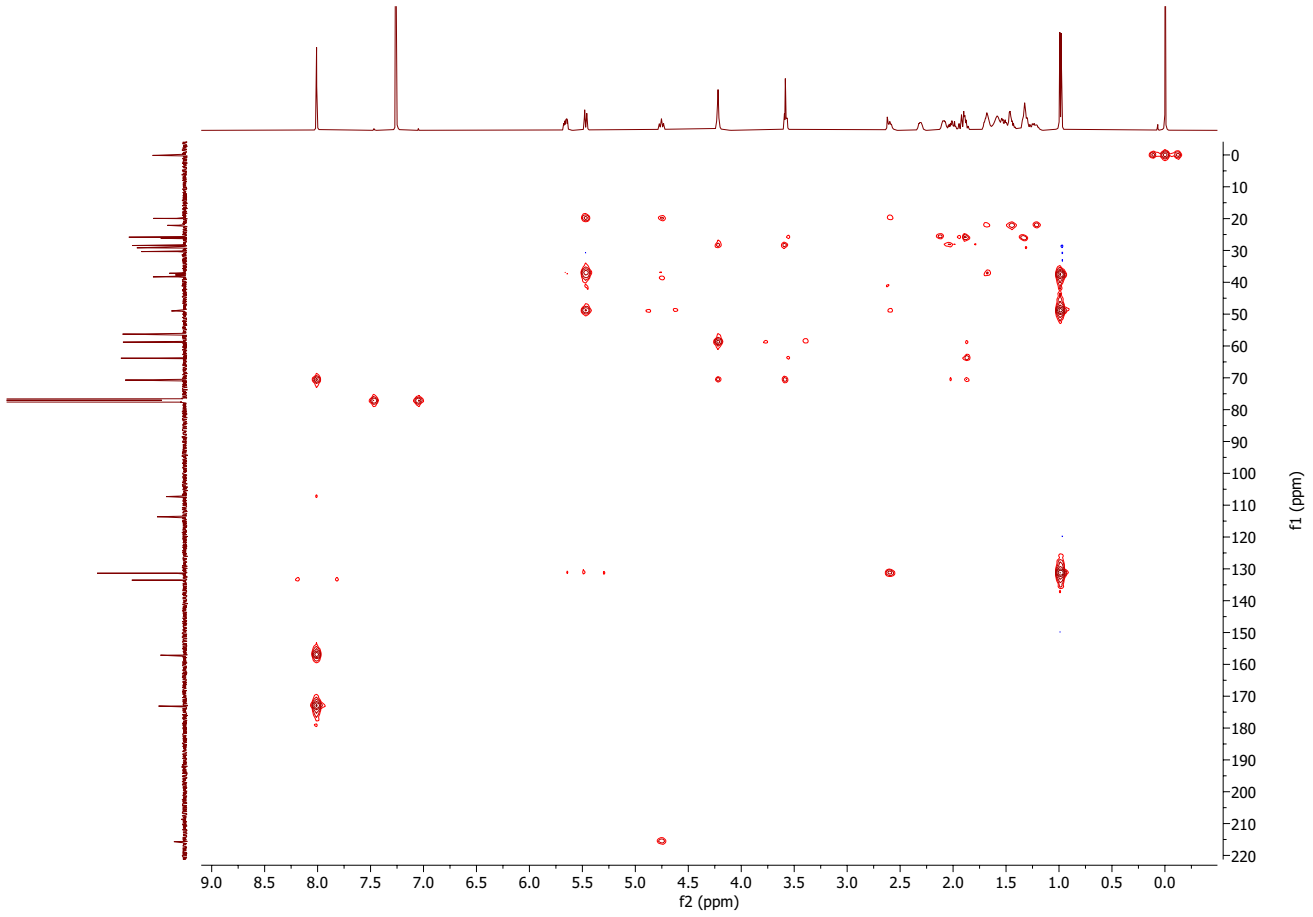


Figure S13. ^{1}H-^{13}C HMBC of compound **1** in $CDCl_3$, 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C NMR.

Supplementary references

[1] Ohashi, M. *et al.* SAM-dependent enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **549**, 502 (2017).

[2] Liu, N. *et al.* Identification and Heterologous production of a benzoyl-primed tricarboxylic acid polyketide intermediate from the zaragozic acid A biosynthetic pathway. *Org. Lett.* **19**, 3560–3563 (2017).

[3] Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods.* **11**, 927–930 (2014).

[4] Hattne, J., *et al.* MicroED data collection and processing. *Acta Cryst.* **A71**, 353–360 (2015).

[5] Kabsch, W. Xds. *Acta Cryst.* **D66**, 125–132 (2010).

[6] Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* **D66**, 133–144 (2010).

[7] Sheldrick, G. M. A short history of SHELX. *Acta Cryst.* **A64**, 112–122 (2008).

[8] Sheldrick, G. M. SHELXT – Integrated space-group and crystal-structure determination. *Acta Cryst.* **A71**, 3–8 (2015).

[9] Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Cryst.* **C71**, 3–8 (2015).

[10] Hübschle, C. B., Sheldrick, G. M. & Dittrich, B. ShelXle: A Qt graphical user interface for SHELXL. *J. Appl. Cryst.* **44**, 1281–1284 (2011).

[11] Delano, W. *The PyMOL Molecular Graphics System* (Schrödinger LLC). http://www.pymol.org.

[12] Fujimoto, H., Ikeda, M., Yamamoto, K. & Yamazaki, M. Structure of fischerin, a new toxic metabolite from an ascomycete, Neosartorya fischeri var. fischeri. *J. Nat. Prod.* **56**, 1268–1275 (1993).

Title:

Structural Determination of an Orphan Natural Product Using Microcrystal Electron Diffraction and Genome Mining

Authors:

Lee Joon Kim,[1$] Masao Ohashi,[2$] Dan Tan,[2] Matthew Asay,[1] Duilio Cascio,[1,3] José A. Rodriguez,[1,3] Yi Tang,[1,2*] and Hosea M. Nelson[1*]

Affiliations:

1. Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States.
2. Department of Chemical and Biomolecular Engineering and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, California 90095, United States.
3. UCLA-DOE Institute for Genomics & Proteomics, University of California, Los Angeles, Los Angeles, California 90095, United States.

[$]*These authors contributed equally.*

**Corresponding authors.*

Abstract:

More than 60% of pharmaceuticals are related to natural products (NPs), chemicals produced by living organisms.[1] Hence, new methods that accelerate natural product discovery are poised to profoundly impact human health. Of the many challenges that remain in natural product discovery, none are as pervasive as structural elucidation, as determination of the molecular structure of a newly discovered natural product can take months, years, or in some cases be altogether unachievable. This challenge can be fueled by lack of sufficient material for spectroscopic analysis, or difficulties in sourcing the producing organism.[2] Even in cases where the analyte is abundant, its physical properties, including molecular structure, can prevent unambiguous structural determination.[3,4] Here we report the use of microcrystal electron diffraction (MicroED),[5] an emerging cryogenic electron microscopy (CryoEM) technique, in combination with genome mining[6] to address these challenges. As proof-of-principle, we apply these techniques to fischerin (**1**), an orphan NP isolated more than 30 years ago, with potent

cytotoxicity but ambiguous structural assignment.[7] We utilize genome mining methods to reconstruct its biosynthetic pathway and highlight the sensitivity of MicroED through the precise determination of the solid-state structure of **1** from sub-micron thick crystals. This structural solution serves as a powerful demonstration of the synergy of MicroED and synthetic biology in NP discovery, technologies that when taken together will ultimately accelerate the rate at which new drugs are discovered.

Main Text:

Natural products (NPs) remain a treasure trove for the development of bioactive molecules. NPs in isolated form and derivatives arising from chemical modification form the basis set for new therapeutics and agrochemicals.[1] This role is becoming ever more important as resistance to existing therapeutic compounds is rapidly increasing as pathogenic organisms evolve in response to treatment.[8] Ideally, the fast evolution of pathogenic organisms would be met by an equally rapid rate of NP discovery; however, for structurally novel natural products, the rate of discovery is decreasing, and is currently lagging behind the emergence of resistance.[9,10] The recent explosion of sequenced microbial genomes has fueled a renaissance in NP discovery, where synthetic biology can be leveraged to produce novel metabolites or to rediscover compounds previously isolated, but no longer available for study.[11] While these cutting-edge methods in synthetic biology have advanced discovery efforts dramatically, the structural elucidation of the NPs remains a limiting step in NP discovery campaigns.

Such difficulties in structural elucidation can arise from i) the lack of sufficient quantities of material for traditional analytical methods (e.g. nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography); ii) intrinsic physical properties of the NP, such as poor solubility and stability in NMR solvents, etc.; and iii) limitations of NMR capabilities in determining relative stereochemistry, which is accentuated by analytes with distal stereocenters, especially when interrupted by rigid substructures bearing multiple rotatable bonds.[12] X-ray crystallography remains the gold-standard for unambiguous structural determination, including the assignment of stereochemistry. However, X-ray crystallographic analysis of newly-isolated natural products is often thwarted by insufficient quantities to provide crystals large enough for single-crystal diffraction (~0.1 mm^3), or poor solid-state properties that preclude the formation of large, pristine crystals when sufficient material is available. Given these challenges, we

envisioned that application of the recently reported CryoEM modality MicroED[5] could lead to vertical advances in the field of natural product discovery, as MicroED has recently been demonstrated to provide unambiguous structures from sub-micron-sized crystals of chemical compounds that had failed to yield large crystals suitable for X-ray analysis.[13,14]

As an entry into the area, we became interested in α-pyridone containing fungal metabolites. These compounds, including fischerin (**1**, Fig. 1a), bassianin,[15] tenellin,[16] ilicicolin H,[17] and apiosporamide (**3**, R = H),[18] are particularly appealing, not only because of their diverse biological activities, but because of the reported difficulties in assigning their relative stereochemistry. In fact, the relative stereochemistry of apiosporamide (**3**), tentatively assigned by the isolation chemists nearly 30 years ago, was only unambiguously assigned after a lengthy total synthesis campaign by Williams and coworkers.[19] The problem in correctly assigning the relative stereochemistry of members of this NP family is molecular topology. Here, distal, stereochemically complex ring systems are linked through freely rotating bonds to a rigid, flat α-pyridone moiety (e.g. **1**, Fig. 1a). Thus, while the relative stereochemistry of the *trans*-decalin system of apiosporamide (**3**) was apparent from 2-dimensional NMR studies, that information could not be correlated to the distal epoxydiol system.

We targeted fischerin (**1**) for our initial efforts in structural elucidation. First isolated more than 25 years ago from *Neosartorya fischeri,* early studies of fischerin showed that it causes acute peritonitis in mice and has potent cytotoxicity.[7] While the connectivity of fischerin (**1**) has been confirmed to be a *2*-pyridone bridging a decalin and a multiply-oxygenated cyclohexane, the relative stereochemistry between the two functional groups as well as the configuration of the epoxydiol unit have thus far eluded unambiguous assignment. Moreover, unlike apiosporamide (**3**) and other members of the family, fischerin (**1**) was hypothesized to possess a rare *cis*-decalin moiety by the isolation chemists. While the relative stereochemistry of this unusual decalin was initially proposed based on NMR analysis of the isolate, the proposed relative stereochemistry of the epoxydiol system is based on computational methods.[20] As an additional obstacle, no other isolation of fischerin (**1**) had been reported, precluding further structural or biological studies of this orphan compound.

To reconstruct the biosynthesis of fischerin (**1**) without access to the reported strain, we used a genome mining approach to find a possible biosynthetic gene cluster (BGC) from a sequenced fungi genome database. We reasoned the *fin* BGC may be present in fungi other than the original producer, albeit silent and not producing fischerin (**1**) under culturing conditions. We hypothesized the BGC should resemble that of the related compound *N*-hydroxyapiosporamide (**2**, R = OH) (Fig. 1a), which is a *trans*-decalin NP with C12 methyl substitution. The BGC of *N*-hydroxyapiosporamide (**2**, R = OH) consists of a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) pathway.[21] The lack of the C12 methyl substituent in fischerin (**1**) suggests that the methyltransferase (MT) domain in the corresponding PKS-NRPS should be inactive and may contain a mutated active site. To identify such a PKS-NRPS, we first categorized fungal BGCs that are homologous to *api* and encode the same set of accessory enzymes, such as enoylreductase (ER), P450s, ene-reductase (OYE), etc (Fig. 1 and Fig. S2). Sequence

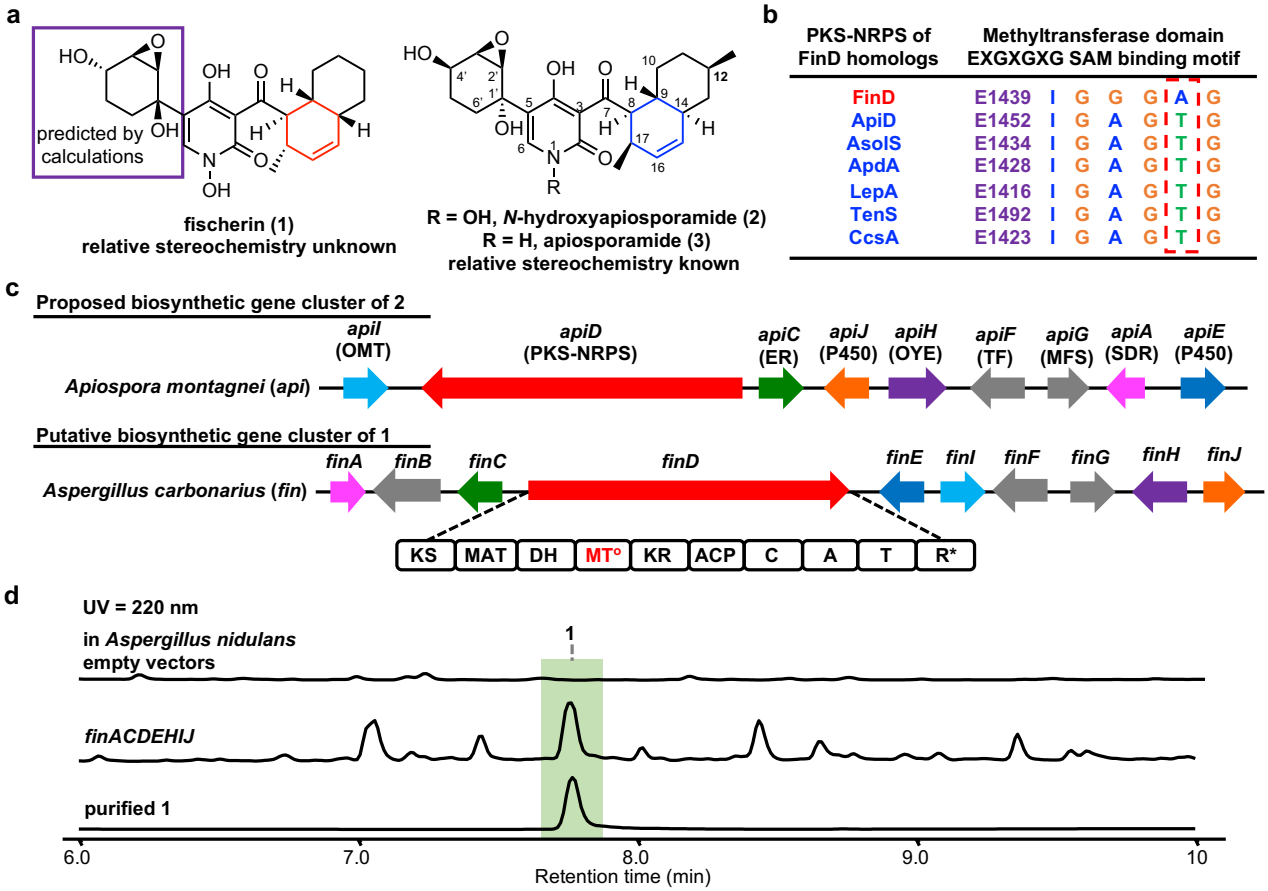


Figure 1. Rediscovery of fischerin. **a**, Proposed structure of fischerin (**1**) and related *N*-hydroxyapiosporamide (**2**). **b**, Alignment of MT domains in PKS-NRPS to identify possible fischerin PKS-NRPS. **c**, Proposed biosynthetic gene cluster of fischerin (**1**). **d**, Heterologous expression of the *fin* cluster leads to formation and isolation of **1**.

scanning of the PKS-NRPS MT domain was then performed on the candidate BGCs, especially at the conserved MT active site GXGTG motif that binds *S*-adenosyl-methionine (SAM).[22] Gratifyingly, we were able to identify one such PKS-NRPS, from a cluster in *Aspergillus carbonarius* (renamed *fin*), to contain a MT domain with a mutated and presumably inactivating GXGAG motif (Fig. 1b). This PKS-NRPS is therefore expected to be devoid of MT activity (the MT domain is designated as MT^o) and could be involved in biosynthesis of **1**.

The *fin* BGC was then completely refactored and expressed in the heterologous host *A. nidulans* A1145 ΔEMΔST[23] for metabolite production. As shown in Fig. 1d, compared to the control strain transformed with vectors only, the coexpression of *fin*ACDEHIJ produced a new metabolite with the same expected *m/z* = 431 as fischerin (**1**) with a titer of 5 mg/L. The culture was scaled up to allow purification of the sample with multiple chromatographic steps. The purified compound was judged to be >98% pure by both LC-MS (Fig. 1d) and NMR analyses (Fig. S3). Comparison of NMR peaks to those of the published data in the same solvent showed complete match of both proton and carbon signals. The purified sample also showed the same negative optical rotation with fischerin ($[\alpha]_D^{20}$ –28° (*c* = 0.10, $CHCl_3$)). Therefore, we are confident that we have correctly identified the *fin* cluster and rediscovered fischerin (**1**) (Fig. S4).

Analogous to the initial isolation study, our attempts to unambiguously establish the relative stereochemistry of fischerin (**1**) using 2D NMR or X-ray crystallography were unsuccessful.[7] We then turned to electron diffraction. While crystallization from various solvents did not yield large single crystals, electron micrographs of pale-yellow particles precipitated from a mixture of acetonitrile and water revealed microcrystals with a distinct morphology. These thin triangular plates were (~0.5–3 μm longest dimension) prone to stacking, thereby hindering the growth of larger crystalline domains and leading to complex intractable diffraction patterns in initial screens (Fig. 2a). Even in cases where we could isolate uniform crystalline domains in selected area diffraction mode, the resolution and completeness of the acquired data were too poor for *ab initio* structural determination (60% complete and 1.3–1.5 Å). Attempts to merge multiple data sets did not lead to an increase in completeness or resolution, as crystal morphology biased orientation of crystals on the grid. In MicroED studies, where stage rotation is limited,

this situation can prevent structural solution.[24] To address this challenge, we looked toward optimizing sample preparation. Re-evaluation of hundreds of crystallization trials revealed that slow evaporation from a mixture of tetrahydrofuran and water allowed for the formation of higher quality microcrystals that still possessed a similarly flat morphology, but alleviating the stacking and aggregation seen in other conditions. After optimization of vitrification conditions,[25,26] we ultimately found that embedding the microcrystals in a layer of vitreous ice promoted trapping of microcrystals in different orientations,[27] increasing the completeness of the acquired data sets.

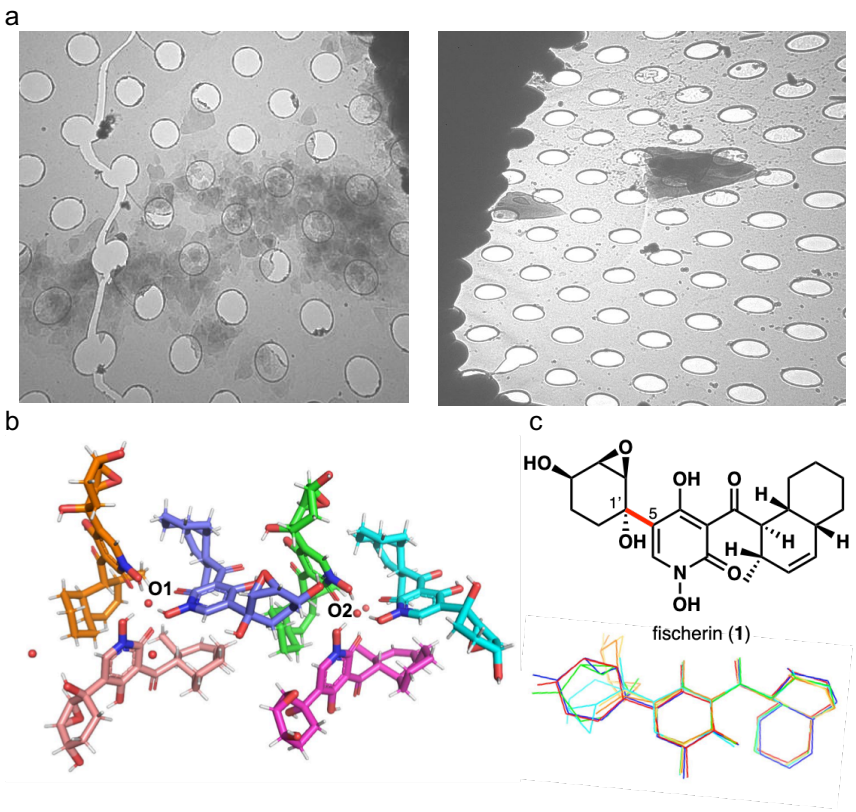


Figure 2. Structure of fischerin (**1**). **a**, Electron micrograph of fischerin (**1**) microcrystalline aggregates (left) and at maximum stage tilt (*ca.* 60°, right); holes are 1 µm wide in diameter. **b**, Asymmetric unit showing two trimers. **c**, Proposed relative stereochemistry (top) and overlay of six fischerin (**1**) molecules showing various degrees of epoxydiol ring rotation (hydrogens omitted for clarity) (bottom).

Of over 200 movies collected, four were merged to provide a 1.05 Å *ab initio* solution (R_1 = 13.8%, refined anisotropically) (Fig. 2b and Table S1). The asymmetric unit revealed six *unique* fischerin (**1**) residues, each with varying degrees of rotation about the carbon-carbon bond connecting the epoxydiol ring system to the pyridone moiety (C1'–C5, Fig. 2c). The lack of symmetry and the large size of the unit cell further explain the difficulties in obtaining high quality data, and the evident facile rotation about the C1'–C5 bond points to potential reasons for challenging assignment of relative stereochemistry with NOE studies (Fig. 2b).[7] Fischerin (**1**) molecules in the asymmetric unit are arranged into two distinct trimers around a common central atom (O1 and O2, Fig. 2b). The identity of the central atom of each trimer was not apparent from initial refinement efforts. Given the challenge of accurately

discriminating elements in MicroED experiments, we devised an atom substitution test similar to the approach described by Kato and co-workers to attempt to assign this atom.[28] Minimization of R_1 and wR_2 values when refining the structure using the electron scattering factors of various atoms against measured structure factors and phases from the initial SHELXD output allowed for convergence to an oxygen atom based on residuals. This test suggested that the central atoms at each trimer are disordered water molecules (Fig. S5).

From the refined crystal structure, we unambiguously assigned the relative stereochemistry of the substituents on the highly-oxygenated cyclohexyl ring, bearing *trans* hydroxyl groups with the epoxy moiety *trans* to the tertiary C1' –OH group (e.g. **1**, Figure 2c). We also validated the predicted stereochemistry of the rare *cis*-decalin system (Fig. 2c). Importantly, the relative stereochemistry of these two stereochemically complex functionalities was apparent in the *ab initio* structural solution and differs from Amini's calculated structure derived from reported NMR chemical shifts.[20] This discordance demonstrates the challenges of using NMR shift calculations to predict stereochemistry on fluxional groups and also highlights the importance of experimental validation of computed structures. Our data collectively inform a first solid-state structure of the orphan NP fischerin (**1**) and, to the best of our knowledge, *the first published example of a structurally-complex, non-peptidyl NP elucidated via MicroED*, demonstrating this technique's potential in NP science and its synergy with other methods such as NMR.

During our crystallographic studies of fischerin, we noticed small amounts of a highly crystalline impurity, present in sufficient quantities to be detectable by electron microscopy but insufficiently abundant in solution to be detectable by ^{1}H NMR (Fig. S3). From two such microcrystals, we collected a high-completeness data set with reflections beyond 1.00 Å (R_1 value of 14.7%, see Table S1), leading to an *ab initio* structure of the complex polycyclic natural product austinol (**4**) (Fig. 3), an endogenous *A. nidulans* meroterpenoid NP that is co-purified with **1** in trace amounts.[29] Intrigued by the ability to obtain structural information from a

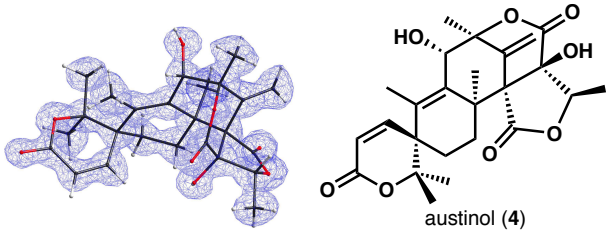


Figure 3. Structural determination of minor impurity austinol.

trace impurity, we proceeded to determine the limit of detection for austinol (**4**) in MicroED experiments. To our delight, we observed that deposition of a 1.5 ng/μL solution of austinol (**4**) in acetonitrile and water on a TEM grid followed by slow evaporation led to formation of microcrystals (Fig. S6). These crystals were of sufficient quality to obtain a high resolution structure, demonstrating the impressive sensitivity of electron crystallography to trace quantities of analytes crystallized *directly* on a TEM grid. Moreover, the ability to identify NPs from a mixture further highlights the applicability of MicroED in NP discovery and characterization, particularly for compounds produced in scarce amounts that may be overlooked by other techniques. These results showcase the exciting possibility of serendipitous discoveries in the field of NP isolation, where biological extracts could often contain minor NP impurities that may not be detectable using other established analytical methods.

In summary, we report the rediscovery of fischerin (**1**) biosynthesis using genome mining approach and its isolation for the first time since its discovery in 1993. We report the first crystal structure of this orphan NP that has eluded full structural characterization for decades, confirming the isolation chemists' proposal of a *cis*-decalin ring and *N*-hydroxy pyridone core, establishing for the first time the relative stereochemistry of the epoxydiol moiety and the stereochemical relationship of these functionalities. Importantly, this structure corrects a previously proposed structure that was based on computational analyses. We also report the structure of a co-metabolite impurity, austinol (**4**), that was present in amounts below the detection limit in our initial ^{1}H NMR experiments, and we demonstrate the exquisite sensitivity of MicroED by obtaining structural information from merely 3 ng of material. These results demonstrate the synergy of synthetic biology advances and MicroED in NP discovery and highlight the importance of developing novel characterization techniques that can complement and overcome limitations of the current state-of-the-art. Taken together, our findings provide a powerful approach toward discovery and structural determination of novel and elusive NPs.

Acknowledgements:

The authors thank Michael R. Sawaya (UCLA-DOE Institute) for assistance in crystallography with data processing and refinement. This research used resources at the UCLA-DOE Institute's X-ray

Crystallography Core Facility, which is supported by the U.S. Department of Energy (DE-FC02-02ER63421). J.A.R. acknowledges support from STROBE, an NSF Science and Technology Center through Grant DMR-1548924, DOE Grant DE-FC02-02ER63421 and NIH-NIGMS Grant R35 GM128867. J.A.R. is supported as a Pew Scholar, a Beckman Young Investigator and a Packard Fellow. Y.T. acknowledges support from the NIH (1R01AI141481). We also thank the David and Lucile Packard Foundation (Fellowship to H.M.N.) and Bristol Myers Squibb (Unrestricted Grant in Synthetic Organic Chemistry to H.M.N) for generous support.

Author Contributions:

H.M.N. and Y.T. supervised the project and designed experiments. M.O. and D.T. performed *in vivo* experiments, as well as compound isolation and characterization. L.J.K. performed crystallization experiments, collected and processed the MicroED data, and solved the structures. L.J.K. and M.A. refined the structures. D.C. assisted in structure refinement. L.J.K. and D.C. performed the atom substitution test. J.A.R. assisted in designing MicroED experiments and helped with MicroED data analysis. L.J.K. and M.O. prepared the figures. H.M.N., Y.T., L.J.K. and M.O. wrote the manuscript.

Competing Interests:

The authors declare no competing financial interests.

[1] Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).

[2] Fisch, K. M. *et al.* Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* **133**, 16635–16641 (2011).

[3] Nicolaou, K. C. & Snyder, S. A. Chasing molecules that were never there: Misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew. Chem. Int. Ed.* **44**, 1012–1044 (2005).

[4] Maier, M. E. Structural revisions of natural products by total synthesis. *Nat. Prod. Rep.* **26**, 1105–1124 (2009).

[5] Jones, C. G., *et al*. The CryoEM method MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **4**, 1587–1592 (2018).

[6] Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).

[7] Fujimoto, H., Ikeda, M., Yamamoto, K. & Yamazaki, M. Structure of fischerin, a new toxic metabolite from an ascomycete, Neosartorya fischeri var. fischeri. *J. Nat. Prod.* **56**, 1268–1275 (1993).

[8] Perfect, J. R. The antifungal pipeline: a reality check. *Nat. Rev. Drug Discov.* **16**, 603–616 (2017).

[9] Fair, R. J. & Tor, Y. Antibiotics and Bacterial Resistance in the 21st Century: *Perspectives in Medicinal Chemistry* (2014) doi:10.4137/PMC.S14459.

[10] Pye, C. R., Bertin, M. J., Lokey, R. S., Gerwick, W. H. & Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **114**, 5601–5606 (2017).

[11] Fisch, K. M. *et al.* Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* **133**, 16635–16641 (2011).

[12] Bifulco, G., Dambruoso, P., Gomez-Paloma, L. & Riccio, R. Determination of relative configuration in organic compounds by NMR spectroscopy and computational methods. *Chem. Rev.* **107**, 3744–3779 (2007).

[13] Rodriguez, J. A., *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525**, 486–490 (2015).

[14] Jones, C. J., *et al.* Characterization of reactive organometallic species via MicroED. *ACS Cent. Sci.* **5**, 1507–1513 (2019).

[15] Wat, C.-K., Mcinnes, A. G., Smith, D. G., Wright, J. L. C. & Vining, L. C. The yellow pigments of Beauveria species. Structures of tenellin and bassianin. *Can. J. Chem.* **55**, 4090–4098 (1977).

[16] Ando, K. *et al.* Funiculosin, a new antibiotic. *J. Antibiot.* **31**, 533–538 (1978).

[17] Hayakawa, S., Minato, H. & Katagiri, K. The Ilicicolins, antibiotics from Cylindrocladium ilicicola. *J. Antibiot.* **24**, 653–654 (1971).

[18] Alfatafta, A. A., Gloer, J. B., Scott, J. A. & Malloch, D. Apiosporamide, a new antifungal agent from the Coprophilous fungus Apiospora montagnei. *J. Nat. Prod.* **57**, 1696–1702 (1994).

[19] Williams, D. R., Kammler, D. C., Donnell, A. F. & Goundry, W. R. F. Total synthesis of (+)-apiosporamide: Assignment of relative and absolute configuration. *Angew. Chem. Int. Ed.* **44**, 6715–6718 (2005).

[20] Amini, S. K. Assignment of the absolute configuration of fischerin by computed nmr chemical shifts. *J. Struct. Chem.* **56**, 1334–1341 (2015).

[21] Ugai, T., Minami, A., Gomi, K. & Oikawa, H. Genome mining approach for harnessing the cryptic gene cluster in Alternaria solani: production of PKS–NRPS hybrid metabolite, didymellamide B. *Tetrahedron Letters* **57**, 2793–2796 (2016).

[22] Skiba, M. A. *et al.* Domain organization and active site architecture of a polyketide synthase C-methyltransferase. *ACS Chem. Biol.* **11**, 3319–3327 (2016).

[23] Liu, N. *et al.* Identification and Heterologous production of a benzoyl-primed tricarboxylic acid polyketide intermediate from the zaragozic acid A biosynthetic pathway. *Org. Lett.* **19**, 3560–3563 (2017).

[24] Nannenga, B. L. MicroED methodology and development. *Struct. Dyn.* **7**, 014304 https://doi.org/10.1063/1.5128226 (2020).

[25] de la Cruz, M. J., *et al.* Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods* **14**, 399–402 (2017).

[26] Dubochet, J., *et al*. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* **21**, 129–228 (1988).

[27] Natesh, R. in *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process* (ed Mohan, C. G.) 375–400 (Springer Nature Switzerland AG, 2019).

[28] Kato, K., *et al.* A vault ribonucleoprotein particle exhibiting 39-fold dihedral symmetry. *Acta Cryst.* **D64**, 525–531 (2008).

[29] Matsuda, Y. & Abe, Ikuro. Biosynthesis of fungal meroterpenoids. *Nat. Prod. Rep.* **33**, 26–53 (2016).