**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

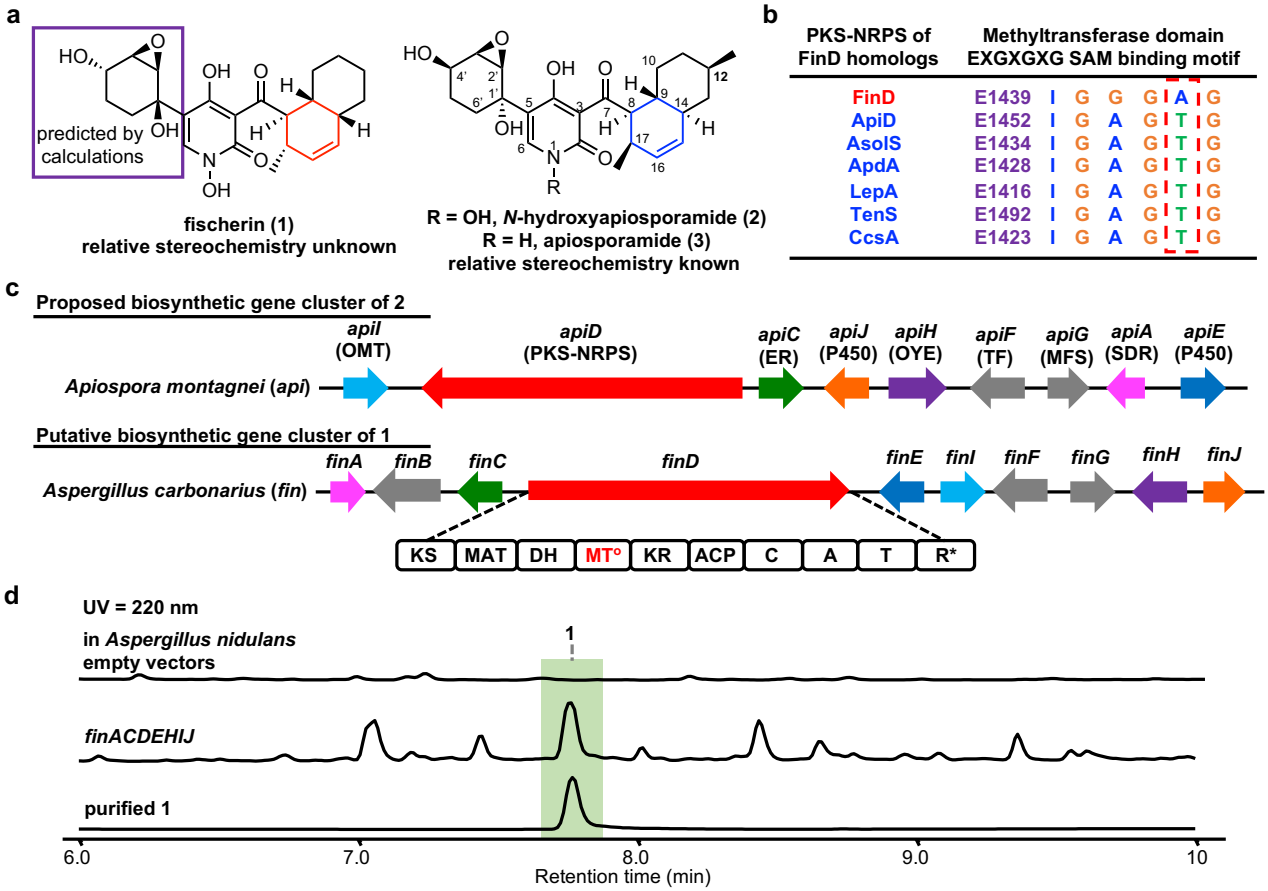


**Figure 1.** Rediscovery of fischerin. **a**, Proposed structure of fischerin (**1**) and related *N*-hydroxyapiosporamide (**2**). **b**, Alignment of MT domains in PKS-NRPS to identify possible fischerin PKS-NRPS. **c**, Proposed biosynthetic gene cluster of fischerin (**1**). **d**, Heterologous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

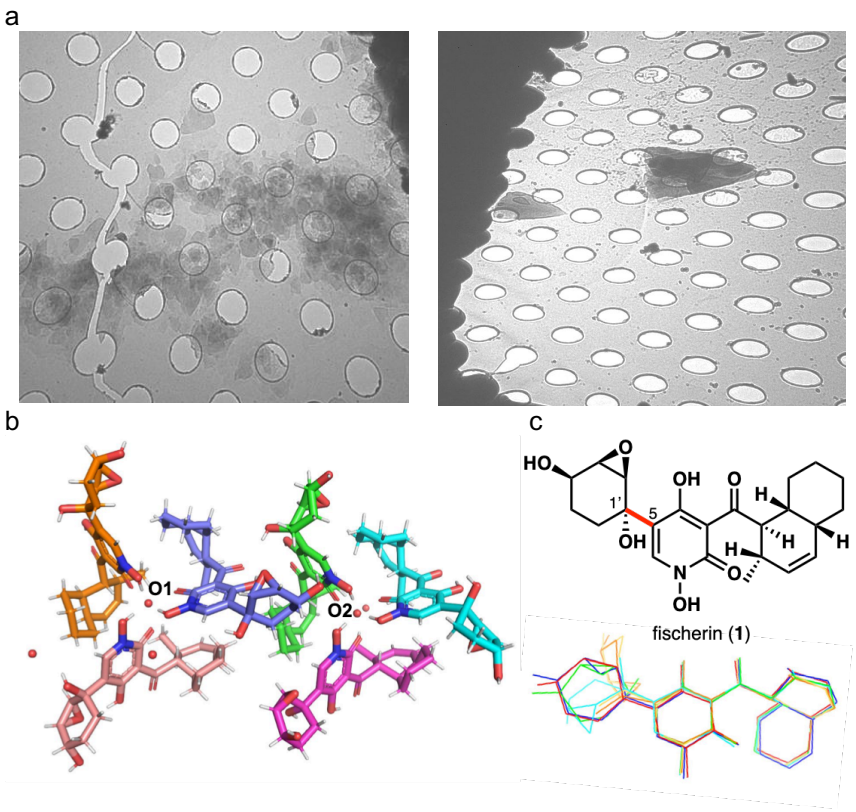


**Figure 2.** Structure of fischerin (**1**). **a**, Electron micrograph of fischerin (**1**) microcrystalline aggregates (left) and at maximum stage tilt (*ca.* 60°, right); holes are 1 µm wide in diameter. **b**, Asymmetric unit showing two trimers. **c**, Proposed relative stereochemistry (top) and 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

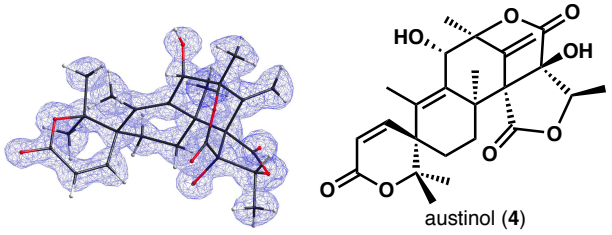


**Figure 3.** Structural determination of minor impurity austinol.

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).