

Cite this: *RSC Chem. Biol.*, 2025,
6, 721

Synthetic biology routes to new and extinct natural products

Thomas J. Simpson 

Recent developments in genome sequencing and genetic engineering have revolutionised elucidation of biosynthetic pathways in bacteria and fungi and allowed production of new natural products and engineered strains with optimised production of new and/or preferred metabolites. The clinically important antibiotic mupirocin is a mixture of closely related pseudomonic acids produced by *Pseudomonas fluorescens* via a *trans*-AT modular PKS. Extensive gene knock-out experiments have led to the isolation of a plethora of new metabolites: both biosynthetic intermediates and shunt products. Parallel experiments, along with swapping of biosynthetic genes, with a *Pseudoalteromonas* sp. which produces the closely related thiomarinols give similar results and many new products. A genetically engineered strain of *P. fluorescens* produces high titres of a single pseudomonic acid with improved stability and antibiotic properties. Tenellin and bassianin are insecticidal fungal metabolites produced by *Beauveria* species via multi-domain PKS-NRPSs. Heterologous expression in *Aspergillus oryzae* of hybrid systems produced by domain swapping between the two biosynthetic gene clusters produce many new metabolites in high yields and reveal the key elements in control of polyketide chain length and methylation, showing the potential for combinatorial biosynthesis of these and related metabolites. *Cryptosporiopsis* sp. 8999 produces three related dimeric xanthenes. Gene knock-outs allow elucidation of the full biosynthetic pathway, isolation of the monomeric precursor and engineering of a strain producing only the major component of the wild-type mixture.

Received 26th February 2025,
Accepted 19th March 2025

DOI: 10.1039/d5cb00047e

rsc.li/rsc-chembio

Introduction

Natural products isolated from plants, marine organisms, fungi and bacteria have long been a major source of bioactive compounds with a wide range of biological activities making them a highly important source of medicines and agrochemicals.¹ The combination of their intriguing structures and interesting properties have made them important targets for synthetic and medicinal chemists.² Apart from the technical and intellectual challenges, this is often justified by the limited amounts sometimes available from the natural sources, and the development of synthetic routes that would allow access to structural variations for structure activity studies and optimisation of biological activities and other desirable properties, e.g., stability and solubility.

In the same way that they have attracted the attention of the synthetic community, studies of their biosynthesis – the metabolic pathways by which they are elaborated in their producing organisms – has long excited the biosynthetic community. These studies have taken many forms – initially isolation of structurally related compounds which might be intermediates or side (shunt) products that may provide evidence of the

biosynthetic sequence.³ These were soon augmented by the increasing availability and use of radioactive isotopes (carbon-14 and tritium) which allowed postulated early or advanced intermediates to be selectively labelled and their incorporation into the final metabolite determined.⁴ This again often necessitated development of sophisticated synthetic methods to strategically place the isotopes in critical positions in the molecule to be incorporated, and elaborate degradation routes to establish that the label had been incorporated site specifically and not randomly. Classical biosynthetic studies on e.g. alkaloid biosynthesis in higher plants relied heavily on this methodology. These, however, were typified by often extremely low levels of incorporations of precursors. The advent of the ready availability of stable isotopes, ¹³C, ²H, ¹⁵N, ¹⁸O and the contemporaneous development of NMR and MS methods for structure elucidation and specific location of labels (and multi-labelling) largely superseded radioisotopes especially in microorganisms where the necessary higher levels of incorporation were more readily achieved. This led to a huge advance in the understanding and elucidation of the details of biosynthetic pathways.⁵ These methods obviated the need for complicated degradations and paralleled the increasing dominance of NMR methods for natural product structure elucidation.

In sharp contrast to primary metabolism, and apart from a few heroic efforts in higher plants,⁶ relatively little was done on

School of Chemistry, University of Bristol, Bristol BS8 1TS, UK.
E-mail: tom.simpson@bristol.ac.uk



the detailed enzymology of secondary metabolic pathways. The advent of molecular genetics and the realisation that all, or at least most, of the genes involved in natural product biosynthesis were normally “clustered” on contiguous stretches of DNA in fungi and bacteria was the next and arguably most significant advance in biosynthetic studies.⁷ This was further augmented by the increasing ease of whole genome sequencing, computational methods of analysis, *e.g.* anti-SMASH⁸ allowing assignment of catalytic function to gene products, and the ability to carry out genetic manipulation. These include overexpression of genes allowing access to biosynthetic enzymes, selective gene deletions to isolate/trap intermediates, methods for heterologous expression of genes, and addition and combination of genes from related pathways. In this review we shall illustrate some applications of these methods drawn from our own work on bacterial and fungal polyketides to the production of novel – new to nature or not so new to nature – metabolites, the production of “improved” structures and generation of (multiple) analogues and structural variations not readily amenable to synthetic methodologies. In addition, we can develop rational improvements of overall titres, and engineer strains which give selective production of only a desired compound rather than mixtures which may be problematic to separate.

Mupirocin/thiomarinol

Mupirocin is a clinically important antibiotic, produced by the bacterium *Pseudomonas fluorescens*.⁹ It is mainly used for topical treatments such as acne and as a nasal spray for pre-op prophylactic treatment against MRSA, multiple resistant *Staphylococcus aureus*.¹⁰ It acts by inhibiting isoleucyl-tRNA synthetase, but its use has been limited by its inherent instability and susceptibility to hydrolysis. It consists of a mixture of pseudomonic acids (PAs). The main constituent is PA-A (*ca.* 90%) **1**, PA-B (7%) **2** and PA-C < (2%) **3**. A main cause of its instability is the intramolecular addition of the 7-hydroxyl onto either C-10 or C-11 of the 10,11-epoxide to give cyclic ethers which are inactive. Biogenetic analysis based on

precedent and logic would suggest that initially formed PA-C is epoxidized to give PA-A followed by further hydroxylation at C-8 to give PA-B. We thus postulated that if the gene controlling epoxidation could be identified, gene knock-out would give PA-C, which lacking the epoxide would be more stable and thus should be an improved and more widely useable antibiotic. An alternative strategy to achieve the same aim uses the closely related thiomarinol A **4**. Thiomarinols are a group of related antibiotics produced by a marine bacterium *Pseudoalteromonas* SANK73390.^{11,12} The structures of the major metabolites, thiomarinol A and pseudomonic acid A (PA-A), differ by the presence of a 10,11-alkene *versus* an epoxide, an additional 4-hydroxyl, an 8-hydroxyoctanoic acid rather than 9-hydroxynonanoic acid side chain, attached to an additional pyrroline moiety *via* an amide linkage to the hydroxy acid moiety.

The mupirocin biosynthetic gene cluster (BGC), designated *mup*, was identified in 2003 (Fig. 2A).¹³ It shows an unprecedented complexity, with a series of PKS and related modules and an unexpectedly large set of tailoring genes, itself suggestive of an unusually complex biosynthetic pathway. Mutational analysis showed that all the genes were essential for mupirocin production.¹⁴

However, mutation of each of a series of apparently unrelated genes all resulted in a surprising total switch from production of mainly PA-A to give high titres of exclusively PA-B, indicating that PA-B must be a precursor to PA-A.¹⁵ This conundrum was eventually solved by the pathways shown in Scheme 1¹⁶ in which a methionine derived methyl (*) in a linear polyketide intermediate **I** is converted by dehydrogenation, epoxidation and an intriguing anti-Baldwin cyclisation to form **III** with the THP core essential for activity.¹⁷ After MmpB mediated production and addition¹⁸ of the fatty acid to give PA-B, subsequent steps analogous to those found in deoxysugar biosynthesis then remove the 8-hydroxyl from PA-B to give PA-A.¹⁶ All the intermediates shown between PA-B and PA-A have been isolated and characterised from the appropriate mutant KO strains (Fig. 2A). Intriguingly it is only when the last step of formation of PA-A takes place that full antibiotic activity is revealed.

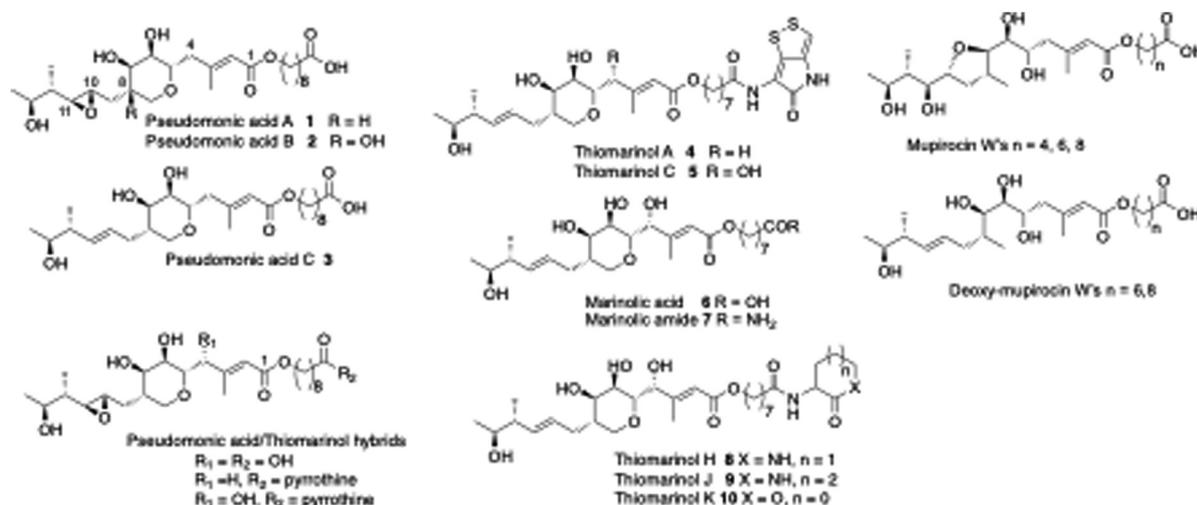
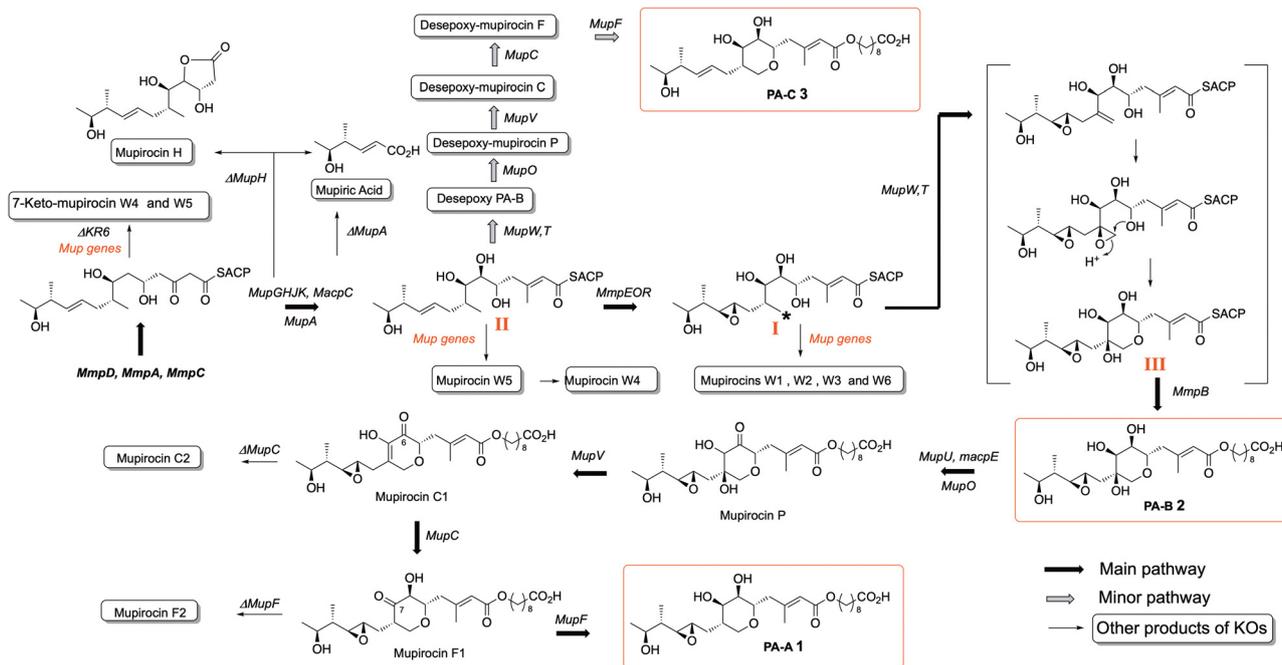


Fig. 1 Selected pseudomonic acid, thiomarinol and mupirocin metabolites.





Scheme 1 Mupirocin biosynthetic pathway and products of gene KOs.

In a minor pathway the non-epoxidised analogue **II** of linear intermediate **I** follows exactly the same sequence to produce the minor PA-C product, again all the intermediates arising from KOs of *mupF*, *C*, *V* and *O* have been isolated in good yields.¹⁹ With this information in hand it was postulated that identifying and blocking the 10,11 epoxidation should divert the biosynthesis down the minor pathway to give the desired PA-C as the main product. The gene responsible was identified as an oxidase domain in the multidomain protein encoded by *mmpE* and KO of the oxidase blocked PA-A production and did indeed give PA-C albeit in low titre. However, minor modifications of the fermentation conditions gave a very high producing strain with PA-C as the sole main product (Fig. 3).²⁰ As summarised in Scheme 1, systematic gene knock-out

experiments have resulted in isolation of a very large number of linear and ring containing mupirocin analogues, but only PA-A and PA-C show significant antibiotic activity.

The thiomarinols are structurally very similar to pseudomonic acids and are also powerful anti-MRSA antibiotics but were not developed clinically due to toxicity issues. The strain also produces a series of xenorhabdins, pyrrothines acylated with a range of long chain (C₈–C₁₆) saturated and unsaturated fatty acids (Fig. 4a).²¹ These acylpyrrothines show antibiotic activities in their own right and are thought to inhibit RNA polymerase. The BGC (*tml*) was identified *via* genome sequencing²² and contains, as expected, *trans*-AT PKS and associated tailoring genes with high homology to the mupirocin (*mup*) cluster, along with genes (*holA-H*) encoding a non-ribosomal peptide synthetase

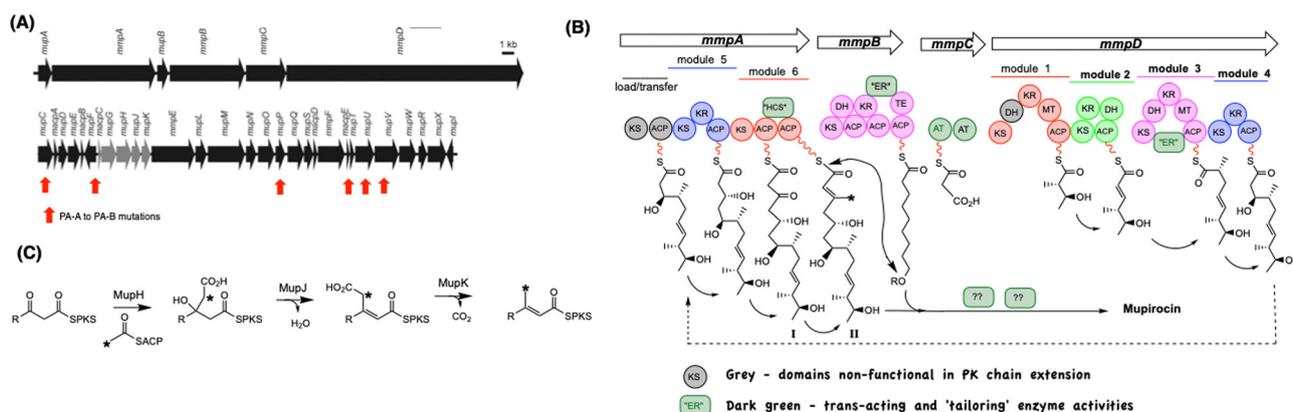


Fig. 2 (A) Summary of the mupirocin biosynthetic gene cluster. The HCS cassette responsible for β -branching is shown in grey. (B) Formalized scheme for the biosynthesis of monic acid and its elaboration to mupirocin. "Grey" domains are inactive or non-functional in chain elongation. Tailoring proteins acting *in trans* are indicated as elongated domains. (C) Scheme for incorporation of a C-15 methyl group by HCS cassette.



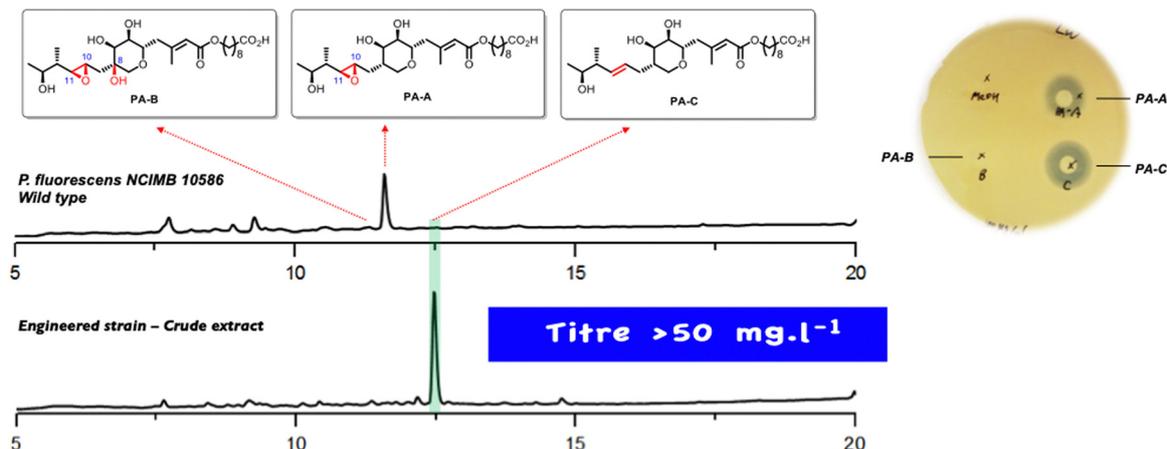


Fig. 3 HPLC trace of the *P. fluorescens* wild type and *mmpE* Δ *OR* mutant cultured in baffled flasks. Plate assay of PAs against MRSA.

(NRPS) linked to a set of tailoring enzymes, similar to those shown to control holomycin (acetyl-pyrrothine) biosynthesis in *Streptomyces clavuligerus*.²³ An internal segment of the NRPS gene (*holA*) was used to generate a Δ NRPS strain.²⁴ This abolished biosynthesis of both thiomarinols and the acylpyrrothines and produced mainly a PA-C analogue, marinolic acid **6** lacking the pyrrothine moiety (Fig. 4c), and two minor analogues with truncated (C_6 and C_4) fatty acid chains and marinolic amide **7**. A Δ PKS mutant produced no thiomarinols but retained the xenorhabdins (Fig. 4d). Finally an in frame deletion of *tmlU* a putative acyl CoA synthase²⁵ resulted in a strain (Δ *tmlU*) which did not produce thiomarinols but produced both marinolic acid and the xenorhabdins (Fig. 4e).

A series of mutasynthesis experiments was carried out with these mutants.²⁶ When pseudomonic acid A was fed to the Δ PKS mutant, pyrrothine addition and/or 4-hydroxylation occurred to give its corresponding thiomarinol A and thiomarinol C

analogues and to 4-hydroxy-PA-A, lacking addition of pyrrothine (Fig. 1). Thus TmlU and the 4-hydroxylase are able to accept close analogues of thiomarinol. PA-C was similarly metabolised. In contrast, feeding PA-B or desepoxy-PA-B simply led to pyrrothine addition, and no 4-hydroxylated metabolites were detected. Hence it is apparent that the presence of an 8-hydroxy group inhibits 4-hydroxylation. A comparison of the mupirocin and thiomarinol gene clusters to identify genes that lack paralogues in the mupirocin cluster and therefore might encode the 4-hydroxylase identified TmuB as an α KG-dependent dioxygenase.²⁷ The *tmuB* coding region was inserted into wild-type *P. fluorescens* via the broad-host-range vector pJH10. HPLC analysis of the culture supernatant after *tmuB* expression showed a main metabolite more polar than PA-A, confirmed to be 4-hydroxy-PA-A. Similar transformation of the PA-C producing mutant gave 4-hydroxy-PA-C, but mutant strains which produced PA-B, showed no conversion to the 4-hydroxylated

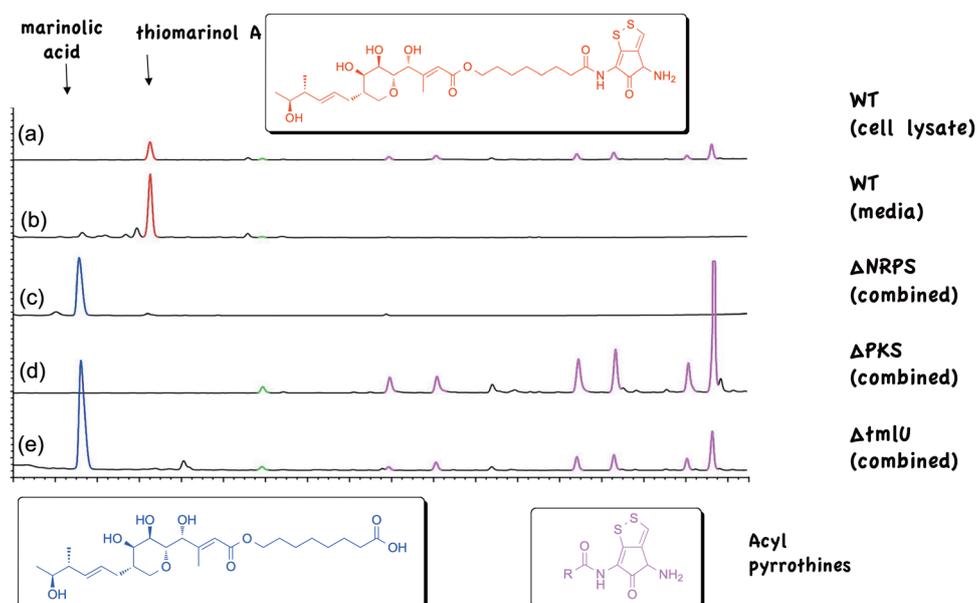


Fig. 4 Products of wild type and mutant strains of *Pseudoalteromonas* SANK73390.



analogue, consistent with the feeding experiments to the Δ PKS mutant described above. Kinetic experiments with purified TmuB indicated that PA-B is a competitive inhibitor of the enzyme. Deletion of *tmlF*, similar to *mupF* which catalyses the last step to PA-A, led to isolation of new thiomarinol metabolites analogous to PA-B and mupirocins C and F, and their deoxy analogues (Scheme 1) all containing the 4-hydroxy and pyrrothine. As might have been anticipated, however, 8-hydroxy-thiomarinol C fails to gain the 4-hydroxyl.

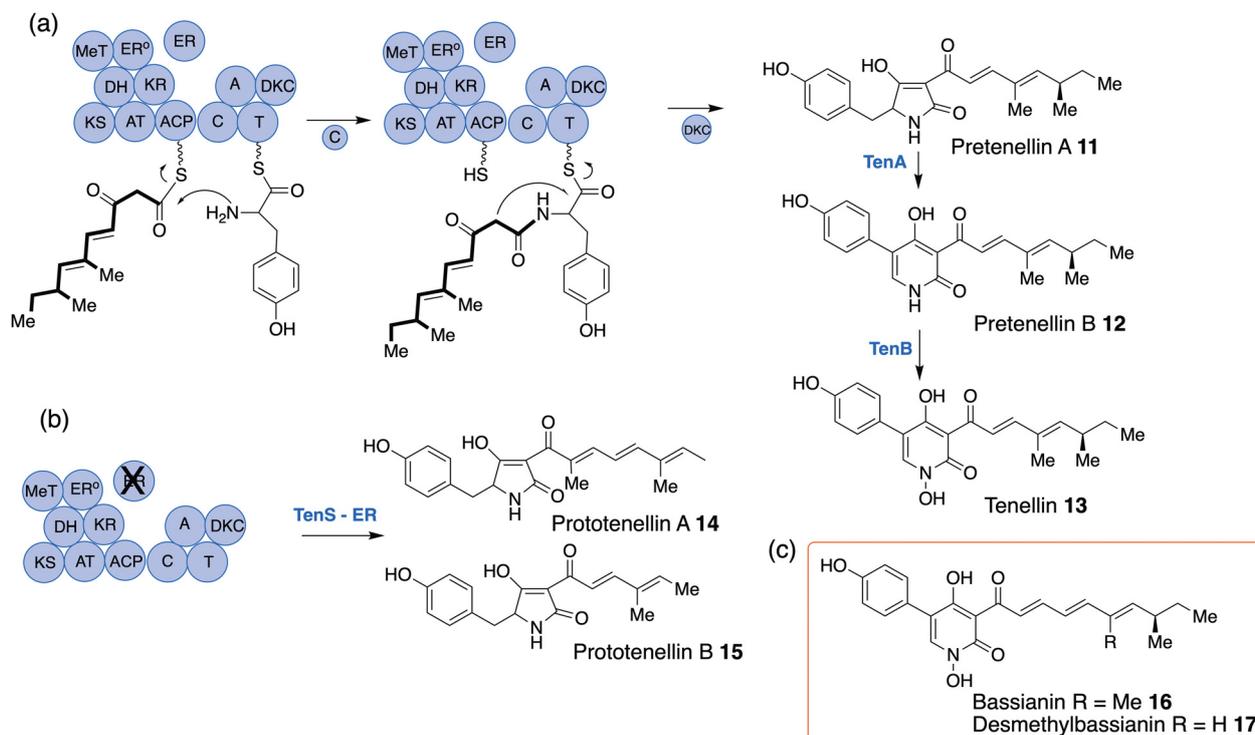
The mutasynthetic incorporation of alternative amines in the Δ NRPS mutant was also examined. The isolation of thiomarinol H **8** was reported from a related organism, now unavailable, *Altermonas rava*,²⁸ but not previously from SANK73390, prompted us to feed anhydro-ornithine and related amines. Ornithine itself has no effect on SANK73390 which may have lost its ability to cyclise ornithine to its anhydro-form. When this was fed, both thiomarinol H and thiomarinol A were produced, indicating that anhydro-ornithine can compete with pyrrothine as a coupling partner in thiomarinol production. Similarly anhydrolysine gave a small amount of the corresponding 7-membered lactam, a “new to nature” analogue, which was named thiomarinol J **9**. In contrast α -aminobutyrolactone was not incorporated but the linear homoserine did yield the novel 5-membered lactone analogue, by analogy called thiomarinol K **10**.

These studies show the value of mutagenesis in producing strains with enhanced and selective production of metabolites with optimised activities. Along with parallel mutasynthesis a very wide range of novel metabolites can be obtained. Bioassays of most of the novel compounds indicate that PA-A and PA-C,

their 4-hydroxy and pyrrothine analogues, thiomarinol A, and marinolic acid all show similar levels of activity against MRSA and a range of other bacteria, but PA-B and all compounds lacking the mature dihydroxy-THP ring display only limited activities. The availability of compounds with truncated chain length fatty acid moieties also confirmed the need for at least a C₆ fatty acid chain for activity, in agreement with X-ray studies of PA-A bound to isoleucyl tRNA synthetase.¹⁰

Tenellin/bassianin PKS-NRPS domain swaps

The fungal metabolites tenellin **13** and bassianin **16** (Scheme 2) are fungal metabolites produced somewhat confusingly from the insect pathogenic fungus *Beauveria bassiana* and *Bassiana tenacellus* respectively.²⁹ They are produced *via* fungal polyketide non-ribosomal peptide synthetase (PKS-NRPS) proteins, first described for biosynthesis of the mycotoxin fusarin C in *Fusarium moniliforme*³⁰ but now known to be widespread. Tenellin, pentaketide, and bassianin, hexaketide, differ only in the chain length of the polyketide moiety, bassianin containing one extra *trans*-olefinic unit. The tenellin BGC is relatively simple, consisting of genes encoding the PKS-NRPS (TenS), an enoyl reductase (TenC) which works in tandem with TenS to produce the first isolable intermediate, pretenellin A **11**, and two P₄₅₀ enzymes, TenA and TenB.³¹ The tenellin PKS belongs to the highly-reducing Type 1 iterative family (hr-PKS)³² and catalyses formation of a bis-*C*-methylated pentaketide with two



Scheme 2 (a) Tenellin biosynthetic pathway. (b) Products of TenS in absence of *in trans* ER, TenC. (c) Bassianin metabolites.



methyl branches being introduced by a *C*-methyl transferase (MeT) domain, the *trans* olefins by the keto-reductase (KR) and dehydratase (DH) domains and the saturated bond by both of these working in concert with the *in trans* enoyl reductase (ER) (Scheme 2a). The polyketide is bound to an ACP before being condensed with tyrosine loaded onto the thiolation (T) domain of the NRPS by the adenylation (A) domain as an amide, bond formation catalysed by the condensing (C) domain. The PKS-NRPS terminates in a “Dieckmann condensation” domain (DKC) which effects cyclisation and release from the NRPS as the tetramic acid, pretenellin A **11**. The first P₄₅₀ (TenA) then catalyses an oxidative ring expansion of the tetramic acid to a pyridone, pretenellin B **12** and the second (TenB) an *N*-hydroxylation to tenellin **13**.³³ Heterologous expression of all

4 genes in *Aspergillus oryzae* results in very high levels of tenellin production, 6 times higher than fermentation of *B. bassiana* itself.³⁴ Omitting *tenB* from the expression system results in production of pyridone, pretenellin B and omission of both *tenA* and *tenB* and gives high yields of the tetramic acid, pretenellin A.³⁵ Interestingly, expression of TenS only without the *trans*-acting ER (TenC, Scheme 2b) gives a rather complex mixture with two isolable main tetramic acids, prototenellins A and B **14** and **15** varying in chain length, number and positioning of *C*-methylation suggesting that physical association of the ER with the PKS plays a key role in fidelity of polyketide assembly.³⁶

Bassianin **16** contains a dimethylated-hexaketide in contrast to the dimethylated-pentaketide in tenellin. To investigate what

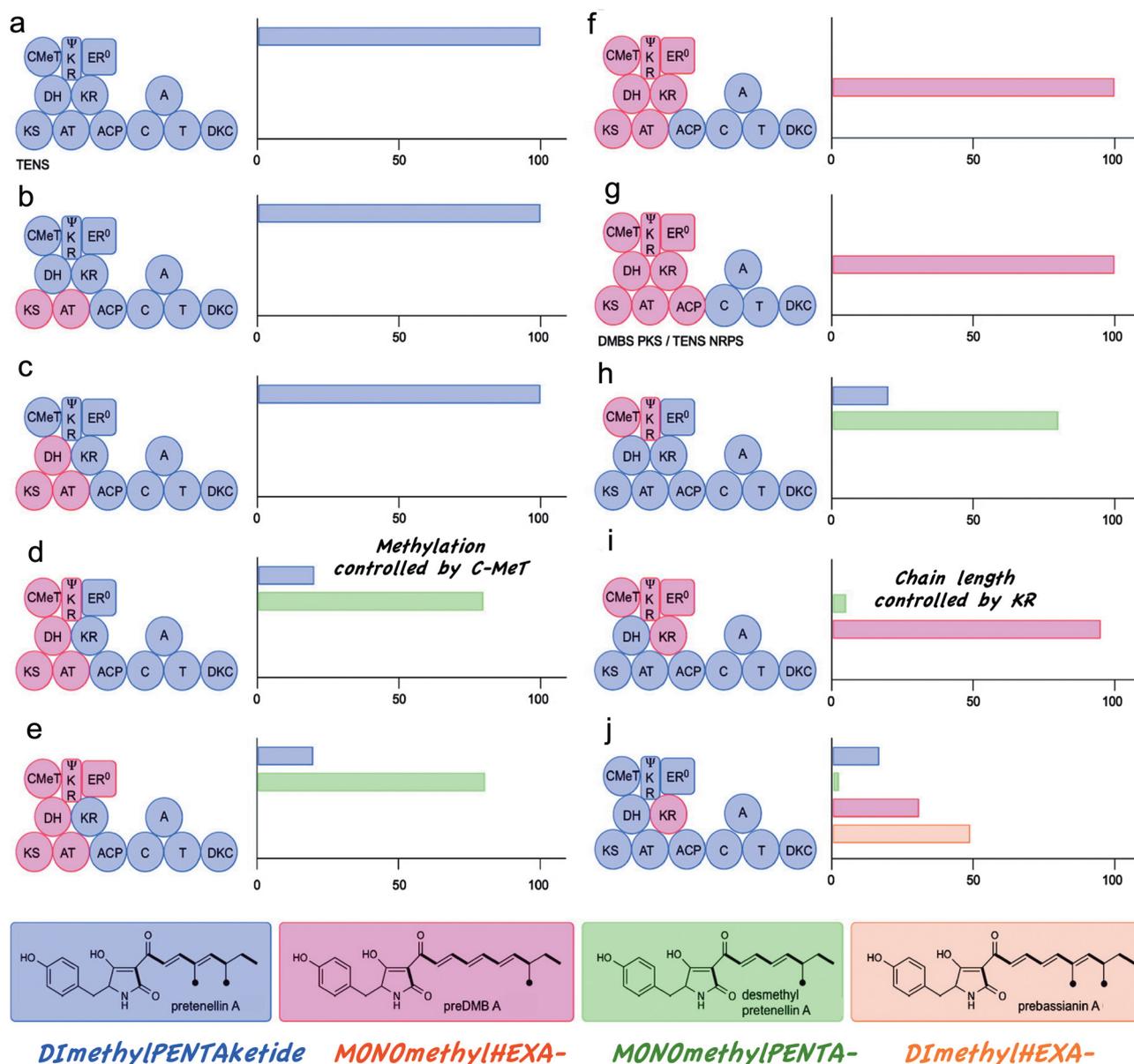


Fig. 5 Domain architectures of chimeric PKS-NRPS constructs and relative titres of resulting metabolites. Co-expression with either the Tens or Dmb *trans* ER domains were equally effective.



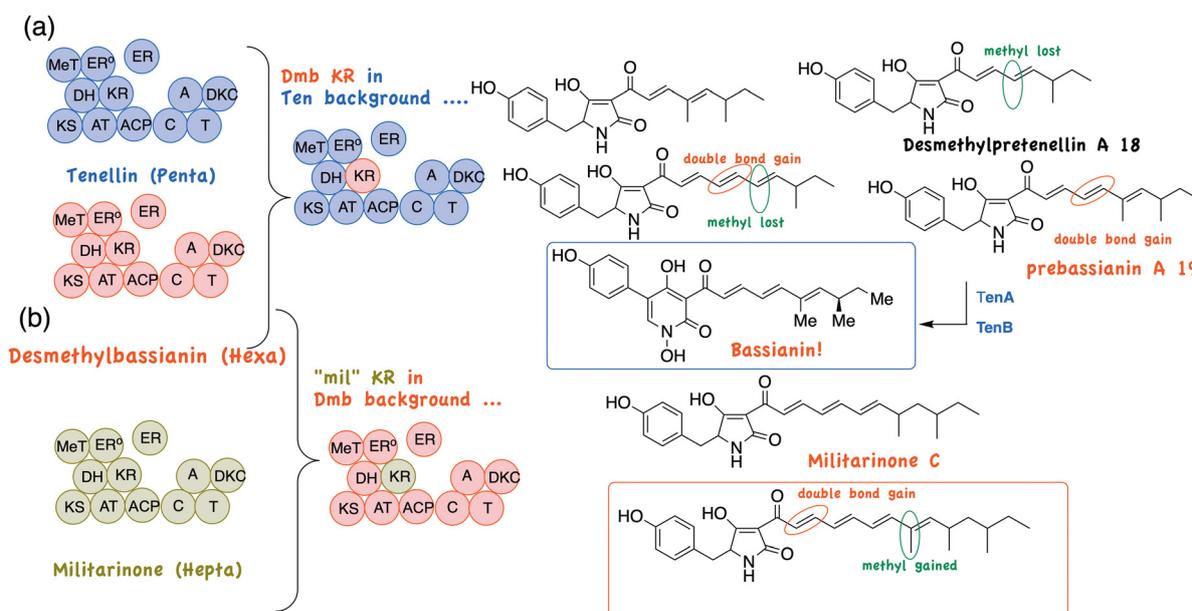
element controls chain length in these systems, we planned parallel experiments with the bassianin gene cluster. However, these were frustrated by the inability to locate a bassianin producing *Beauvaria* strain, the original producing strain no longer being available. Thus, bassianin is effectively an “extinct” metabolite. However, after extensive screening of some 30 *Beauvaria* strains,³⁶ one was found which produced a novel monomethylated derivative, desmethylbassianin **17**, simply lacking introduction of the second *C*-methylation step in bassianin (Scheme 2c), thus, allowing us to study a further control element – *C*-methylation. The genome was sequenced and the *dmb* BGC identified: this had very high homology to the ten cluster. Parallel gene expression experiments then gave exactly analogous results to those in tenellin.³⁶

A series of gene constructs were then prepared where the DNA encoding domains in the tenellin PKS moiety (blue) were successively replaced by the corresponding desmethylbassianin (pink) domains.³⁷ The overall aim being to elucidate what controlled the differing aspects of polyketide assembly: chain length, degree of methylation, degree of reduction, *etc.* This resulted in a number of “new to nature” compounds being produced again in high titres (Fig. 5). Substitution of the first few domains had no observable effect until the *Dmb C*-MeT domain was included. The major product of this construct (Fig. 5d) was found to be a novel monomethylated-pentaketide containing metabolite, desmethylpretenellin **18** indicating not surprisingly, that the *C*-Met domain influences the number of methylation steps. Further inclusion of the *Dmb KR* domain (Fig. 5f and g), gave a complete switch to produce only the monomethylated-hexaketide, predesmethylbassianin A **19**, indicating, somewhat surprisingly, that it is the *KR* domain that primarily determines chain length.

Interestingly in heterologous PKS constructs, the *KR* domain has a small effect on the fidelity of MeT catalysed methylation and *vice versa*. This was verified by preparing single domain replacement constructs in which only the TenS MeT was included in an otherwise *Dmb* background (Fig. 5h) and one in which the *Dmb KR* domain was included in an otherwise TenS background (Fig. 5j and Scheme 3a). This gave a mixture of 4 compounds, including minor amounts of pretenellin and desmethylpretenellin indicating some loss of fidelity in these heterologous constructs. As expected total hexaketides dominated (80%) over pentaketides (20%) – *KR* control, and dimethylated products (66%) over monomethylated (34%) – MeT control. Gratifyingly of the two hexaketide derived products the major component was “prebassianin A” **19** the precursor to the “extinct” bassianin. Co-expression of *tenA* and *tenB* with this construct gave another four products, the major being the long lost bassianin. We have called such “raised from the dead” metabolites “Lazarus compounds”! Thiomarinol H above can be similarly classified. The principle can be extended to other more heterologous systems. Militarionone, a tetramic acid metabolite of *Paecilomyces militaris* contains a bis-*C*-methylated heptaketide.³⁷ Insertion of the militarionone *KR* into a *Dmb* background (Scheme 3b) now gives a trimethylated octaketide as a main product.

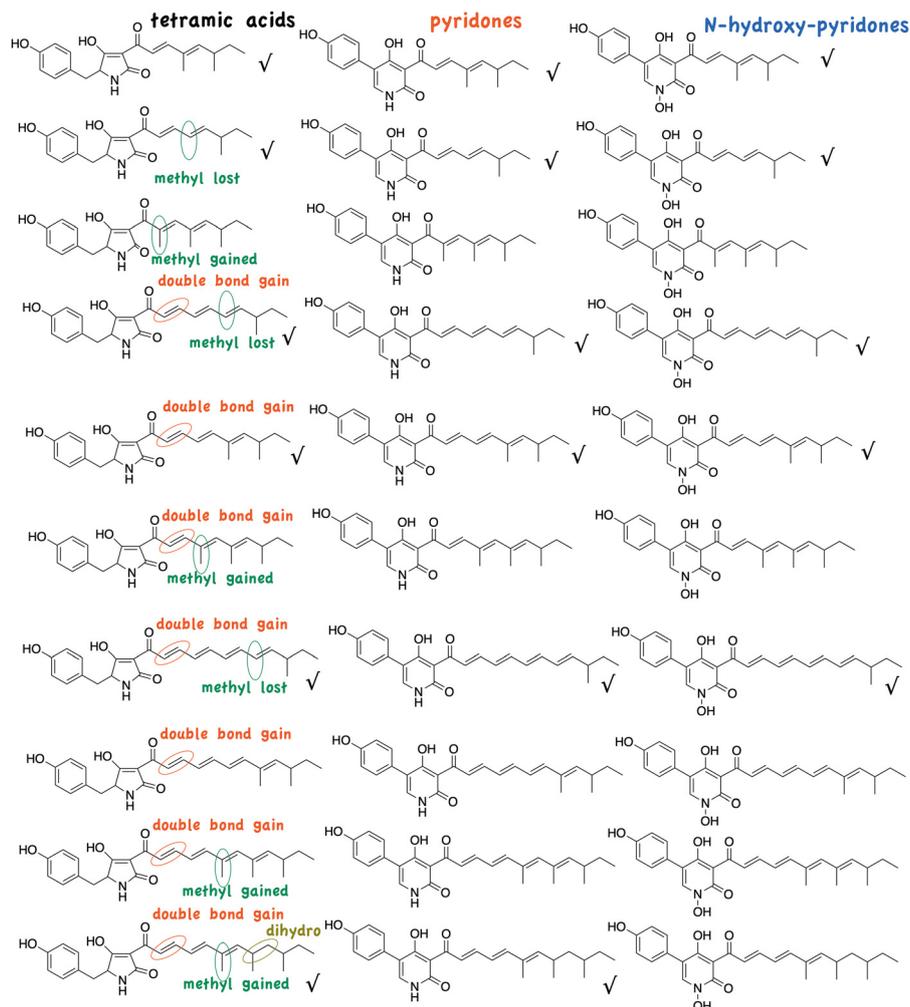
A large variety of these fungal PKS-NRPS compounds differing in polyketide assembly and amino acid component are known and so in principle, these simple domain swaps can potentially give rise to a very large number of novel compounds. As indicated in Fig. 6, many of these have already been achieved.

This ability to not only produce new natural products but also recreate extinct compounds by pathway engineering is a pertinent demonstration of the power of combinatorial biosynthesis as applied to fungal polyketides.



Scheme 3 Single KR domain replacements, (a) DMB KR in a TenS background and (b) Mil KR in a Dmb background.





30 structures shown here.
Ca. 20 are known NPs or have
been produced by domain swaps
to date.

Other variations -

Make all "dihydro" - 54
Vary methylation positions
Vary Chain Length further
Vary side chain reduction
Vary amino acid

PKS-NRPS systems for all of
these are available.

Fig. 6 Further products (actual and potential) from tenellin, desmethylbassianin and militarinone domain swaps.

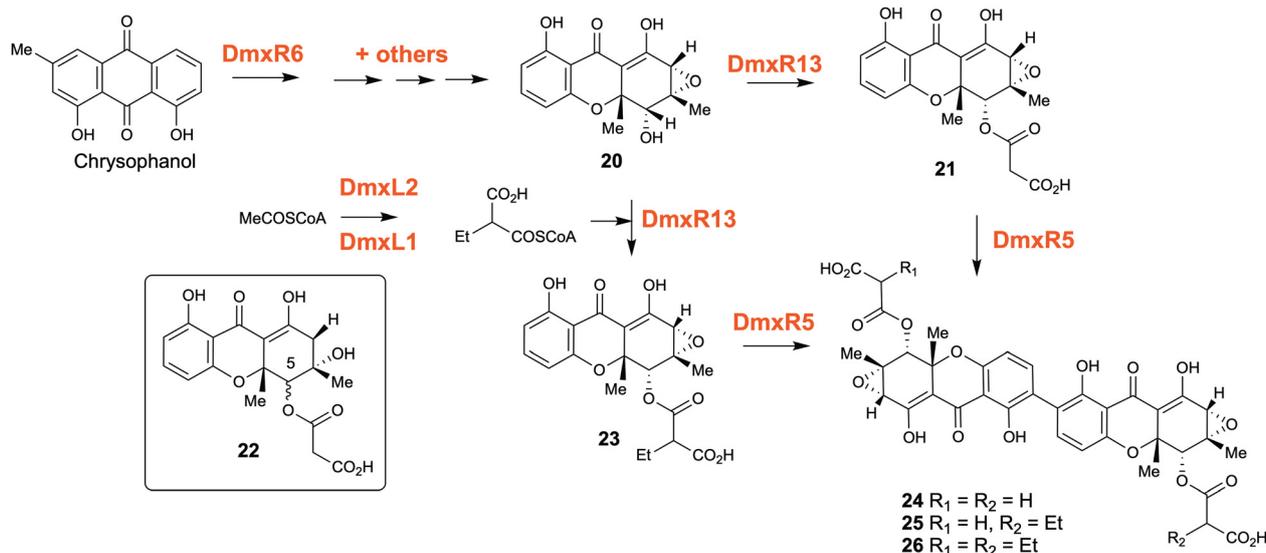
Cryptosporioptins

A large number of bioactive dimeric xanthenes are found in fungi, notably the mycotoxins secalonic acids and related ergochromes, the antimalarial ascherxanthone A and the antimicrobial and anticancer dicerandrol C.³⁸ The biosynthesis of these dimeric xanthenes have been long studied but uncertainties remain, in particular the timing of the ring cleavage of the precursor anthraquinones,³⁹ usually chrysophanol, and the timing of dimerisation of the xanthone monomers.⁴⁰ In the course of studies on the nonadrine class of natural products in of *Cryptosporiopsis* sp. 8999,⁴¹ we isolated three closely related dimeric metabolites and established the structures to be 24–26 which differ only in the presence of ethyl groups on the unusual malonyl substituent.⁴² Genome sequencing identified the cryptosporioptide (*dmx*) BGC, which was confirmed by targeted knockout of the putative cryptosporioptide PKS and subsequent abolition of all cryptosporioptide production. Disruption of *dmxR6*, a putative Baeyer–Villiger monooxygenase gave high titres of chrysophanol as expected, showing that dimerisation does occur after anthraquinone ring cleavage and xanthone formation (Scheme 4). Deletion of *dmxR5* encoding a P₄₅₀

(putative dimerase) abolished dimer production, and two novel metabolites with similar UV spectra to the cryptosporioptides but with molecular weights consistent with monomers were isolated. NMR analysis confirmed that these were hemi-cryptosporioptide A 21 and a shunt product 22 containing a tertiary alcohol at C-6 which was isolated as a mixture of epimers at C-5. When monomer 21 was re-fed to the Δ *dmxPKS* mutant, cryptosporioptide A 24 production was restored confirming this as a pathway intermediate. The *dmx* BGC has a series of unique genes not present in the other dimeric xanthone clusters such as *dmxR2*, *dmxR8*, *dmxR11–13* and *dmxL1–3*.

In particular *dmxL2* encodes a highly reducing PKS (hr-PKS) homologous to the lovastatin diketide synthase. The gene *dmxL1* encodes an acyl-CoA carboxylase. *DmxR13* encodes an *O*-acyltransferase, and again this is absent from the other clusters. Disruptions of *dmxL2* and *dmxL1* both abolished production of cryptosporioptide B 25 and C 26, to give exclusively cryptosporioptide A, proving their role in biosynthesis of butyrate and carboxylation to give ethylmalonate and its subsequent incorporation into 25 and 26, presumably by competing with endogenous malonate. It is significant that three different classes of molecule – AQ, monomeric and dimeric xanthenes





Scheme 4 Cryptosporioptin biosynthesis.

can be accessed by these means. In addition, biosynthetic flux is again redirected to a produce single metabolite from a mixture.

Conclusions

Natural products isolated from bacteria and fungi have long been an invaluable source of bioactive compounds that have been of immense value in human, animal and plant health, or have served as valuable lead compounds for further development. Limiting issues have been availability in required titres, and in particular their structural complexity making them difficult to access in quantity and to modify their structures by standard synthetic chemistry to enhance activities, stability, solubilities, *etc.* A fundamental rationale for biosynthetic studies has always been the idea that if sufficiently well delineated, biosynthetic pathways should be amenable to manipulation to produce analogues and derivatives with enhanced properties. However, achieving this has traditionally been limited to mutagenesis by classical methods, and precursor directed biosynthesis where analogues of intermediates were fed into cultures. With the advent of genome sequencing and computer-based analysis and the normal occurrence in micro-organisms of biosynthetic genes being “clustered” on contiguous stretches of the genome, myriad possibilities for genetic engineering approaches became viable. In this article some of these have been highlighted. These include targeted gene knock-outs to isolate intermediates, selected deletion of genes and addition of genes from other pathways to produce derivatives and analogues, blocking of side or other minor pathways to direct metabolite flux along a single route to avoid separation of mixtures. Heterologous expression of whole pathways in more amenable hosts can allow higher titres and more efficient isolation and purification. The results arising from applications of some of these techniques are illustrated above. Others include mining of genome sequences and switching on “silent” pathways not normally expressed and transfer of whole pathways from

non-cultivable organisms and even expression of environmental DNA. It is an exciting time for this area and after a period when exploitation of natural products became less actively pursued in academia and, in particular, in industry, the study of natural products and their biosynthetic pathways is realising a new and vigorous future.

Data availability

No primary research results, software or code have been included, and no new data were generated or analyzed as part of this review.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Work described herein has been the product of the combined efforts of several senior colleagues in the “Bristol Polyketide Group” including Professors Matt Crump, Chris Willis, Russell Cox (now Hannover), Christopher Thomas (Birmingham,) Drs Zongshu Song, Colin Lazarus and Andrew Bailey. Their substantial contributions are gratefully acknowledged. Last, but not least, the many graduate students and post-doctoral workers involved, not just for the detailed and skilled work they performed and the intellectual ideas they contributed, but also for the pleasure of associating with so many talented young researchers from around the world. There are too many to name individually, but they are listed in the original references cited below.

References

- 1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.



- 2 K. C. Nicolau and S. Rigol, *Nat. Prod. Rep.*, 2020, **37**, 1404–1435.
- 3 W. O. James, *The Alkaloids*, ed., R. H. F. Manske and H. L. Holmes, Academic Press, New York, 1950, vol. I, p. 16.
- 4 A. R. Battersby, *Q. Rev., Chem. Soc.*, 1961, **15**, 259–286.
- 5 T. J. Simpson, *Chem. Soc. Rev.*, 1975, **4**, 497–522.
- 6 T. M. Kutchan, H. Dittrich, D. Bracher and M. H. Zenk, *Tetrahedron*, 1991, **47**, 5945–5954.
- 7 M. Oliynyk, M. Samborsky, J. B. Lester, T. Mironenko, N. Scott, S. Dickens, S. F. Haydock and P. F. Leadlay, *Nat. Biotechnol.*, 2007, **25**, 447–453.
- 8 M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano and R. Breitling, *Nucleic Acids Res.*, 2011, **39**, W339–W346.
- 9 E. B. Chain and G. Mellows, *J. Chem. Soc., Perkin Trans. 1*, 1977, 294–309.
- 10 C. M. Thomas, J. Hothersall, C. L. Willis and T. J. Simpson, *Nat. Rev. Microbiol.*, 2010, **8**, 281–289.
- 11 H. Shiozawa, A. Shimada and S. Takahashi, *J. Antibiot.*, 1997, **50**, 449–450.
- 12 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
- 13 A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–430.
- 14 J. Hothersall, J.-e Wu, A. S. Rahman, J. A. Shields, S. M. Cooper, E. Stephens, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2007, **282**, 15451–15461.
- 15 S. M. Cooper, W. Laosripaiboon, J. Hothersall, A. K. El-Sayed, C. Winfield, J. Crosby, R. J. Cox, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2005, **12**, 825–833.
- 16 L. Wang, S.-S. Gao, Z. Song, J. Hothersall, E. R. Stevens, J. Connolly, P. J. Winn, R. J. Cox, M. P. Crump, P. Race, C. M. Thomas, T. J. Simpson and C. L. Willis, *Angew. Chem., Int. Ed.*, 2017, **56**, 3930–3934.
- 17 L. Wang, A. Parnell, C. Williams, N. A. Bakar, M. Challand, M. W. van der Kamp, T. J. Simpson, P. R. Race, M. P. Crump and C. L. Willis, *Nat. Catal.*, 2018, **1**, 968–976.
- 18 A. J. Winter, M. T. Rowe, A. N. M. Weir, N. Akter, S. Z. Mbatha, P. D. Walker, C. Williams, Z. Song, P. R. Race, C. L. Willis and M. P. Crump, *Angew. Chem., Int. Ed.*, 2022, **61**, e202212393.
- 19 S.-S. Gao, J. Hothersall, Ji'en Wu, A. C. Murphy, Z. Song, E. R. Stephens, C. M. Thomas, M. P. Crump, R. J. Cox, T. J. Simpson and C. L. Willis, *J. Am. Chem. Soc.*, 2014, **136**, 5501–5507.
- 20 A. J. Winter, F. de Courcy-Ireland, A. P. Phillips, J. M. Barker, N. A. Bakar, N. Akter, L. Wang, Z. Song, J. Crosby, C. Williams, C. L. Willis and M. P. Crump, *Angew. Chem., Int. Ed.*, 2024, **63**, e202410502.
- 21 B. V. McInerney, R. P. Gregson, M. J. Lacey, R. J. Akhurst, G. R. Lyons, S. H. Rhodes, D. R. Smith, L. M. Engelhardt and A. H. White, *J. Nat. Prod.*, 1991, **54**, 774–784.
- 22 D. Fukuda, A. S. Haines, Z. Song, A. Murphy, J. Hothersall, E. R. Stephens, R. Gurney, C. Riemer, R. Marshall, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson and C. M. Thomas, *PLoS One*, 2011, **6**, e10831.
- 23 B. Li and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 19732–19735.
- 24 A. C. Murphy, D. Fukuda, Z. Song, J. Hothersall, R. J. Cox, C. L. Willis, C. M. Thomas and T. J. Simpson, *Angew. Chem., Int. Ed.*, 2011, **50**, 3271–3274.
- 25 Z. D. Dunn, W. J. Wever, N. J. Economou, A. A. Bowers and B. Li, *Angew. Chem., Int. Ed.*, 2015, **54**, 5137–5141.
- 26 L. Wang, Z. Song, P. R. Race, J. Spencer, T. J. Simpson, M. P. Crump and C. L. Willis, *Chem. Sci.*, 2020, **11**, 5221–5226.
- 27 H. H. Mohammad, J. A. Connolly, Z. Song, J. Hothersall, P. R. Race, C. L. Willis, T. J. Simpson, P. J. Winn and C. M. Thomas, *ChemBioChem*, 2018, **19**, 836–841.
- 28 D. B. Stierle and A. A. Stierle, *Experientia*, 1992, **48**, 1165–1169.
- 29 C. K. Wat, A. G. McInnes, D. G. Smith, J. L. C. Wright and L. C. Vining, *Can. J. Chem.*, 1977, **55**, 4090–4098.
- 30 Z. Song, R. J. Cox, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2004, **5**, 1196–1203.
- 31 K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2007, **8**, 289–297.
- 32 R. J. Cox, *Org. Biomol. Chem.*, 2007, **5**, 2010–2026.
- 33 L. M. Halo, M. N. Heneghan, A. A. Yakasai, Z. Song, K. Williams, A. M. Bailey, R. J. Cox, C. M. Lazarus and T. J. Simpson, *J. Am. Chem. Soc.*, 2008, **130**, 17988–17996.
- 34 M. N. Heneghan, A. A. Yakasai, L. M. Halo, Z. Song, A. M. Bailey, R. J. Cox, T. J. Simpson and C. M. Lazarus, *ChemBioChem*, 2010, **11**, 1508–1512.
- 35 L. M. Halo, J. W. Marshall, A. A. Yasaki, Z. Song, C. P. Butts, M. P. Crump, M. Heneghan, A. M. Bailey, T. J. Simpson, C. M. Lazarus and R. J. Cox, *ChemBioChem*, 2008, **9**, 585–594.
- 36 M. N. Heneghan, A. A. Yakasai, K. Williams, K. A. Kadir, W. Bakeer, K. M. Fisch, Z. Wasil, A. M. Bailey, T. J. Simpson, R. J. Cox and C. M. Lazarus, *Chem. Sci.*, 2011, **2**, 972–979.
- 37 K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, A. M. Bailey, C. M. Lazarus, T. J. Simpson and R. J. Cox, *J. Am. Chem. Soc.*, 2011, **133**, 16635–16641.
- 38 K. Schmidt, U. Riese, Z. Li and M. J. Hamburger, *J. Nat. Prod.*, 2003, **66**, 378–383.
- 39 T. Wezeman, S. Brase and K. S. Masters, *Nat. Prod. Rep.*, 2015, **32**, 6–28.
- 40 K. M. J. de Mattos-Shiple and T. J. Simpson, *Nat. Prod. Rep.*, 2023, **40**, 174–201.
- 41 K. Williams, A. J. Szwalbe, N. P. Mulholland, J. L. Vincent, A. M. Bailey, C. L. Willis, T. J. Simpson and R. J. Cox, *Angew. Chem., Int. Ed.*, 2016, **55**, 6784–6788.
- 42 C. Greco, K. de Mattos-Shiple, A. M. Bailey, N. P. Mulholland, J. L. Vincent, C. L. Willis, R. J. Cox and T. J. Simpson, *Chem. Sci.*, 2019, **10**, 2930–2939.

