WARWICK
THE UNIVERSITY OF WARWICK

Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/169512

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Diene Incorporation by a Dehydratase Domain Variant in Modular Polyketide Synthases

Christian Hobson[1†‡], Matthew Jenner[1, 2†], Xinyun Jian[1, 2], Daniel Griffiths[1‡], Douglas M. Roberts[1], Matias Rey-Carrizo[1‡] and Gregory L. Challis[1, 2, 3, 4*]

[1]Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

[2]Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK

[3]Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia

[4]ARC Centre of Excellence for Innovations in Peptide and Protein Science, Monash University, Clayton, VIC 3800, Australia

*Corresponding Author; Email: G.L.Challis@warwick.ac.uk

†These authors contributed equally

‡Current addresses: Willow Biosciences Inc, Vancouver, British Columbia, Canada (C.H.); Monash University Accident Research Centre, Clayton, VIC 3800, Australia (D.G.); BCN Medical Writing, Sabadell, Catalunya 08203, Spain (M.R.-C.); Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC 3800, Australia (X.J.)

Abstract

Modular polyketide synthases (PKSs) are biosynthetic assembly lines that construct structurally diverse natural products with wide-ranging applications in medicine and agriculture. Various mechanisms contribute to structural diversification during PKS-mediated chain assembly, including dehydratase (DH) domain-mediated elimination of water from *R* and *S*-configured 3-hydroxy thioesters to introduce *E* and *Z*-configured carbon-carbon double bonds, respectively. Here we report the discovery of a novel DH domain variant that catalyses the sequential elimination of two molecules of water from a (3*R,* 5*S*)-3, 5-dihydroxy thioester during polyketide chain assembly, introducing a conjugated *E*, *Z*-diene into various modular PKS products. We show that the reaction proceeds via a (2*E,* 5*S*)-2-enoyl-5-hydroxy thioester intermediate and involves an additional universally conserved histidine residue that is absent from the active site of most conventional DH domains. These findings expand the diverse range of chemistries mediated by DH-like domains in modular PKSs, highlighting the catalytic versatility of the double hot dog fold.

Introduction

Polyketides are an important family of natural products, with numerous applications in human medicine, animal health and crop protection[1]. Although they have exceptionally complex and diverse structures,

the chemical transformations used to assemble their carbon chains are deceptively straightforward. A thioester starter unit is elongated with a series of (alkyl)malonyl thioester extender units via a decarboxylative condensation reaction, and the β-ketothioesters resulting from each chain elongation undergo one or more α- and/or β-carbon modifications.[2,3]

In bacteria, structurally complex polyketides are typically assembled by type I modular polyketide synthase (PKS) multienzymes. These remarkable molecular machines employ sets of catalytic domains grouped into modules, each of which is usually responsible for a single round of carbon chain extension and subsequent α/β-carbon modification (Fig. 1a).[4] An acyl transferase (AT) loads the extender unit onto the post-translationally installed phosphopantetheine (ppant) prosthetic group of an acyl carrier protein (ACP) domain in each module (Fig. 1a). In *cis*-AT PKSs these ATs are integrated into the modules and are typically able to utilise a range of (alkyl)malonyl-CoA thioesters, whereas in *trans*-AT PKSs a single standalone AT usually acylates the ACP domain in each module with malonyl-CoA. Each chain extension module also contains a ketosynthase (KS) domain that receives the acyl chain attached to the ACP domain in the upstream module (via transthioesterification onto a conserved active site Cys residue) and elongates it with the extender unit (Fig. 1a). Optional α/β-carbon modifying domains alter the structure of the resulting β-ketothioester prior to chain elongation by the downstream module. These include C-methyltransferase (MT) domains, which methylate the α-carbon, ketoreductase (KR) domains, which reduce the β-keto group, dehydratase (DH) domains that eliminate water from the resulting β-hydroxy thioester and enoyl reductase (ER) domains, which saturate the alkene formed by the dehydration reaction (Fig. 1a).[4,5] O-methyltransferase (OMT) domains catalyse O-methylation of the β-hydroxy thioester intermediate and several *trans*-acting enzymes are responsible for other types of modification.[6–8]

Many α/β-carbon modifying domains are highly stereoselective and sequence analyses can predict the stereochemistry of the products they form. For example, “type A” and “type B” KR domains generate (3*S*) and (3*R*)-3-hydroxythioester products,[4] respectively (Fig. 1b), and are distinguishable using sequence alignments. Similarly, DH domains typically convert (3*R*)-3-hydroxythioesters to *E*-configured alkenes, although it has recently been established that some DH domains can convert *S*-configured substrates to *Z*-configured products (Fig. 1b).[5,9]

Several *trans*-AT PKSs assemble polyketides containing a conjugated *E*, *Z*-diene. Examples include gladiolin **1**, leinamycin **2**, kirromycin **3** and elansolid A **4** (Fig. 1c). Analysis of the catalytic domains in the module pairs proposed to be responsible for incorporating these dienes (e.g., modules 12 and 13 of the gladiolin PKS; Fig. 1d and Supplementary Fig. 1 and 2) show that the first contains a type A KR domain, whereas the second contains a type B KR domain. This is consistent with formation of a (3*R*, 5*S*)-3, 5-dihydroxy thioester intermediate. However, only the second module contains a DH-like domain, raising intriguing questions about the mechanism for *E*, *Z*-diene formation.

Here we elucidate the catalytic function and mechanism of the DH-like domain from module 13 of the gladiolin PKS (Supplementary Fig. 2) using chemically synthesised analogues of intermediates in polyketide chain assembly, intact protein mass spectrometry and site-directed mutagenesis. Our data show that this domain can convert an ACP-bound (3*R*, 5*S*)-3, 5-dihydroxyacyl thioester to a (2*E*, 4*Z*)-2, 4-dienoyl thioester, via a (2*E*, 5*S*)-2-enoyl-5-hydroxy thioester intermediate, further expanding the catalytic repertoire of DH-like domains in type I PKSs.

Results

Dehydration of 3-hydroxy thioesters

The *apo*-DH-ACP di-domain from module 13 of the gladiolin PKS was overproduced and purified as reported previously (Supplementary Fig. 3).[10] Initially we investigated *N*-acetylcysteamine and pantetheine thioester mimics of the proposed intermediates in chain assembly attached to the ACP domains in modules 12 and 13 of the PKS as substrates for the DH-like domain. Consistent with studies of excised DH domains from other modular PKSs,[11] they were found to be poor substrates for the DH-like domain in our construct and no dehydration products could be detected. To circumvent this problem, we exploited the CoaA, CoaD and CoaE enzymes to convert the pantetheine thioesters to the corresponding coenzyme A thioesters,[12–14] which were then used, in conjunction with $MgCl_2$, ATP and the substrate tolerant 4'-phosphopantethienyl transferase Sfp,[15] to convert the *apo*-ACP domain of the DH-ACP di-domain to the corresponding acylated *holo*-forms (Fig. 2a). This allowed dehydration of the ACP-bound thioesters, resulting in a -18 Da mass shift, to be detected directly by intact protein mass spectrometry (Fig. 2a). The pantetheine thioesters were also used to convert the excised *apo*-ACP

domain to the corresponding acylated *holo*-forms, which were analysed by intact protein mass spectrometry as a negative control (Supplementary Fig. 3 and 4)

As initial probes of dehydratase activity, we synthesised (3*R*) and (3*S*)-3-hydroxyhexanoyl pantetheine thioesters, **5** and **6** (Supplementary Note) and incubated them separately with purified recombinant CoaA, CoaD, CoaE, Sfp and the *apo*-DH-ACP di-domain. Given the structural diversity of the predicted substrates of the putative diene-forming DH-like domains in *trans*-AT PKSs (Supplementary Fig. 1), we reasoned they are likely to have broad substrate tolerance and that the ACP-bound thioesters derived from **5** and **6** would be accepted by the module 13 DH-like domain. After 3 hours, the mixtures were analysed using UHPLC-ESI-Q-TOF-MS, which showed that the *R*-configured thioester had undergone dehydration, whereas the *S*-configured thioester had not (Fig. 2b, 2c and Supplementary Table 1). To elucidate the stereochemical outcome of the transformation with the *R*-configured thioester, we exploited the intrinsic reversibility of the dehydration reaction, following the precedent set by previous studies of dehydratases involved in bacterial fatty acid biosynthesis and a DH domain from the pikromycin PKS.[11,16] The (2*E*) and (2*Z*)-2-hexenoyl panetheine thioesters, **7** and **8**, were synthesised (Supplementary Note) and loaded via their CoA thioesters onto the *apo*-ACP domain of the DH-ACP di-domain construct. UHPLC-ESI-Q-TOF-MS analyses of the reaction mixtures showed that only the *E*-configured thioester underwent rehydration (Fig. 2d, 2e and Supplementary Table 1). Based on the principle of microscopic reversibility,[16] these results indicate that the DH-like domain in module 13 of the gladiolin PKS, like canonical modular PKS DH domains, catalyses the dehydration of (3*R*)-3-hydroxyacyl thioesters to form (2*E*)-configured products. The DH-like domain cannot, however, catalyse the formation of a *Z*-configured product from a (3*S*)-3-hydroxyacyl thioester, as reported recently for the DH domains from module 2 of the fostriecin *cis*-AT PKS,[5] and modules 2 and 5 of the bongkrekic acid and oxazolomycin *trans*-AT PKSs, respectively.[9] Formation of the *Z*-configured C12-C13 double bond in gladiolin is therefore very unlikely to involve direct dehydration of the (3*S*)-3-hydroxyacyl thioester produced by the type A KR domain in module 12 of the PKS.

Double dehydration of a (3*R*, 5*S*)-3,5-dihydroxy thioester

The organisation of catalytic domains in modules 12 and 13 of the gladiolin PKS suggests a (3*R*, 5*S*)-3, 5-dihydroxyacyl thioester attached to the module 13 ACP domain might be an intermediate in

polyketide chain assembly. It is therefore conceivable that the module 13 DH-like domain catalyses two successive dehydrations of this diol to form the (10*E*, 12*Z*)-diene of gladiolin. To test this possibility, we synthesised (3*R,* 5*S*)-3, 5-dihydroxyhexanoyl pantetheine thioester **9** (Supplementary Note), converted it to the corresponding CoA thioester, loaded it onto the *apo*-DH-ACP di-domain and determined its fate by intact protein MS. In addition to the diol, singly and doubly dehydrated species were observed (Fig. 3a and Supplementary Table 1), indicating that the DH-like domain is indeed capable of diene formation. To gain further insight into this process, we synthesised (2*E,* 4*Z*)-2, 4-hexadienoyl pantetheine thioester **10** (Supplementary Note), which corresponds to the presumed product of the double dehydration reaction and loaded it onto the *apo*-DH-ACP di-domain. Intact protein MS analysis showed that singly and doubly hydrated products are formed (Fig. 3b and Supplementary Table 1). Significant quantities of the *holo*-DH-ACP di-domain were observed in both experiments, likely resulting from facile offloading of the diol from the ACP domain via δ-lactonisation. Taken together, these data are consistent with formation of the (10*E*, 12*Z*)-diene of gladiolin via two sequential dehydrations of a (3*R,* 5*S*)-3, 5-dihydroxyacyl thioester by the DH-like domain in module 13 of the PKS. To further substantiate this hypothesis, we used cysteamine to cleave the products of the reaction with the (3*R,* 5*S*)-3, 5-dihydroxyhexanoyl thioester from the ACP domain.[17] LC-MS comparisons with the cysteamine adducts derived from the synthetic pantetheine thioesters of (2*E,* 4*Z*) and (2*E,* 4*E*)-2, 4-hexadienoate (**10** and **11**, respectively; Supplementary Note), showed that the (2*E,* 4*Z*)-hexadienoyl thioester is the major diene-containing product of the reaction (Fig. 3c). Significant quantities of the (2*E,* 4*E*)-2, 4-hexadienoyl cysteamine adduct were also observed. This is likely due primarily to facile isomerisation of the (2*E,* 4*Z*)-2, 4-hexadienoyl thioester to the thermodynamically more stable *E, E*-stereoisomer, as observed during the synthesis of (2*E,* 4*Z*)-2, 4-hexadienoyl pantetheine thioester **10**. However, we cannot exclude that the dehydration leading to formation of the 4, 5-double bond results in small amounts of the 4*E*-configured product, in addition to the major 4*Z*-configured product of the reaction.

Activity with stereoisomeric 3, 5-dihydroxy thioesters

To further probe the stereochemical preferences of the DH-like domain, we synthesised all three stereoisomers (**12**, **13** and **14**) of (3*R,* 5*S*)-3, 5-dihydroxyhexanoyl pantetheine thioester **9**

(Supplementary Note) and studied their fate upon conversion to the corresponding CoA thioesters and subsequent loading onto the *apo*-DH-ACP di-domain (Extended Data Fig. 1 and 2). Negligible dehydration was observed for stereoisomers **12** and **13** with 3*S*-configured hydroxyl groups, relative to the control reactions employing the excised ACP domain (Extended Data Fig. 1a, 1b, Supplementary Table 1 and Supplementary Fig. 4). Interestingly, the major product of the reaction employing **13** is the *holo*-DH-ACP di-domain, which likely results from offloading of the substrate via δ-lactonisation (Extended Data Fig. 1b) These data are consistent with that obtained for the (3*S*)-3-hydroxyhexanoyl pantetheine thioester **6** (Fig. 2) and indicate that diene formation proceeds via elimination of the C-3 hydroxyl group, followed by the C-5 hydroxyl group.

In contrast, the (3*R,* 5*R*)-3, 5-dihydroxyacyl thioester **14** underwent almost complete conversion to a mixture of singly and doubly dehydrated products (Extended Data Fig. 2a and Supplementary Table 1), indicating that the DH-like domain is not stereoselective towards the C-5 hydroxyl group. Comparison of the cysteamine cleavage adducts from this reaction with that for the panthetheine thioesters of (2*E,* 4*Z*) and (2*E,* 4*E*)-2, 4-hexadienoate (**10** and **11**, respectively) showed that the (2*E*, 4*E*)-2, 4-dienoyl thioester is the sole diene-containing product (Extended Data Fig. 2b). Accordingly, conversion of the (2*E,* 4*E*)-2, 4-dienoyl pantetheine thioester **11** to the corresponding CoA thioester and subsequent loading onto the *apo*-DH-ACP di-domain yielded predominantly a mono-rehydrated product (Extended Data Fig. 2c and Supplementary Table 1). The differences in product profile observed for the reactions employing the (3*R,* 5*S*)-3, 5-dihydroxyhexanoyl / (2*E,* 4*Z*)-2, 4-dienoyl panthetheine thioesters (**9** and **10**), where a significant proportion of the diol remains / is produced, and the (3*S,* 5*S*)-3, 5-dihydroxyhexanoyl / (2*E,* 4*E*)-2, 4-dienoyl pantetheine thioesters (**14** and **11**), where very little of the diol remains / is produced, can be explained by the greater thermodynamic stability of the *E*, *E*-diene than the *E*, *Z*-diene, and are consistent with the hypothesis that the (2*E,* 4*Z*)-2, 4-dienoyl thioester is the major product of the enzyme-catalysed dehydration of (3*R,* 5*S*)-3, 5-dihydroxyhexanoyl thioester.

Mechanism of diene formation

Consistent with the observation that 3*S*-configured 3,5-dihydroxy thioesters are poor substrates for the DH-like domain, mechanistic logic suggests that diene formation likely proceeds via initial elimination of the C-3 hydroxyl group. This is because the protons attached to C-2 are significantly more acidic

than those attached to C-4 in the 3, 5-dihydroxy thioester and the acidity of the protons attached to C-4 increases dramatically in the 5-hydroxy-2-enoyl thioester resulting from elimination of the C-3 hydroxyl group. To test this hypothesis, we synthesised (2*E,* 5*S*)-2-hexenoyl-5-hydroxy pantetheine thioester **15** (Supplementary Note), which corresponds to the putative mono-dehydrated intermediate in diene formation from the (3*R,* 5*S*)-3,5-dihydroxyhexanoyl thioester. Loading of thioester **15**, via the corresponding CoA thioester, onto the *apo*-DH-ACP di-domain produced a mixture of hydrated and dehydrated products (Fig. 4 and Supplementary Table 1), providing clear evidence that it is an intermediate in the interconversion of the diol and the diene. In an analogous experiment with the synthetic (2*E*, 5*R*)-2-hexenoyl-5-hydroxy pantetheine thioester **16** (Supplementary Note) the diene was produced (Extended Data Fig. 2d and Supplementary Table 1). However, as for the (2*E*, 4*E*)-2,4-hexadienoyl thioester **11**, negligible diol formation was observed (Extended Data Fig. 2d and Supplementary Table 1).

To illuminate the molecular basis for the unusual catalytic activity of the DH-like domain, we created a structural model using AlphaFold (Fig. 5).[18] An overlay of this model with the X-ray crystal structure of a DH domain from module 11 of the difficidin polyketide synthase, which has been proposed to convert a 3, 5-dihydroxy thioester to a diene (although no experimental evidence is available to validate this),[19] shows a very high degree of structural similarity (Extended Data Fig. 3).[19] DH domains typically contain a conserved His-Asp catalytic dyad in their active site. A one-base mechanism, in which the His residue of the catalytic dyad first deprotonates C-2 of the thioester to form an enolate, then protonates the C-3 hydroxyl group as it is eliminated from this intermediate, has been proposed for PKS DH domains (Extended Data Fig. 4),[16] and recently verified using pH-rate profiles.[20] The Asp residue of the catalytic dyad appears to position the substrate by accepting a hydrogen bond from the C-3 hydroxyl group. The AlphaFold model of the DH-like domain contains an additional active site His residue on an adjacent β-strand ~3.5 Å away from the His residue of the catalytic dyad (Fig. 5). In canonical DH domains, such as that from the EryAII subunit of the 6-deoxyerythronolide B (6-DEB) PKS, this residue is Tyr (Fig. 5 and Extended Data Fig. 5). A multiple sequence alignment of all the DH-like domains in *trans*-AT PKSs predicted to catalyse diene formation revealed that this additional His residue is universally conserved (Extended Data Fig. 3). These observations led us to propose that the 2-enoyl-5-hydroxy thioester intermediate in diene-forming DH-like domains likely results from deprotonation of C-2 by the His residue of the conserved His-Asp dyad, whereas the additional His residue (His158 in the DH-like

domain from module 13 of the gladiolin PKS) deprotonates C-4 of this intermediate to form a vinylogous enolate (Fig. 4). The C-5 hydroxyl group is then eliminated from the vinylogous enolate with concomitant protonation by the additional His residue (Fig. 4). A similar mechanism has been proposed for formation of a 4*E*-configured double bond by DH domains in the *cis*-AT PKS that assembles curacin. However, these domains lack the additional active site His residue in diene-forming DH-like domains.[21] The Asp residue in the conserved His-Asp dyad of the *trans*-AT PKS DH-like domains likely accepts hydrogen bonds from both the C-3 and C-5 hydroxyl groups of the 3, 5-dihydroxy thioester substrates. This would explain why substrates with 3*R*, 5*S* and 3*R*, 5*R* configurations give rise to predominantly 2*E*, 4*Z* and dienes 2*E*, 4*E*-configured dienes, respectively (Fig. 4b and Extended Data Fig. 2e).

To investigate the role played by the additional His residue in diene formation, a H158Y mutant of the DH-ACP di-domain from module 13 of the gladiolin PKS was constructed (Supplementary Fig. 5). Loading of (2*E*, 5*S*)-2-hexenoyl-5-hydroxy pantetheine thioester **15**, via the corresponding CoA thioester, onto the mutant di-domain resulted in formation of only the rehydrated (3*R, 5S*)-3, 5-dihydroxyacyl thioester (Fig. 5 and Supplementary Table 1). None of the diene product resulting from elimination of the C-5 hydroxyl group could be detected (Fig. 5 and Supplementary Table 1). These data are consistent with the proposed roles of the His-Asp dyad and the additional His residue in diene formation.

Discussion

In this study, we solve the conundrum of how conjugated *E*, *Z*-dienes are incorporated into the products of *trans*-AT PKSs. Our results show that a novel DH domain variant in these assembly lines containing an additional active site His residue converts (3*R,* 5*S*)-3, 5-dihydroxyacyl ACP thioester intermediates in polyketide chain assembly to the corresponding (2*E,* 4*Z*)-2, 4-dienoyl thioesters. This reaction is proposed to proceed via a (2*E,* 5*S*)-2-enoyl-5-hydroxy thioester intermediate resulting from deprotonation of C-2 in the 3, 5-dihydroxy thioester substrate by the His residue in the universally conserved His-Asp catalytic dyad of canonical DH domains. Deprotonation of C-4 in the intermediate by the additional active site His residue is hypothesised to trigger a second dehydration reaction, resulting in diene formation. We propose that the Asp residue in the His-Asp catalytic dyad forms hydrogen bonds with the C-3 and C-5 hydroxyl groups, positioning the C-2 and C-4 protons near the

conserved and additional His residues, respectively. This orients the substrate in a conformation that favours *E, Z*-diene formation via two sequential E1$_{CB}$ reactions, each of which is a net *syn*-elimination.

Interestingly, in a minority of cases (the (2*Z*, 4*E*)-dienes in basiliskamide and macrolactin, the (12*Z*, 14*E*)-diene in chivosazol, and the (8*Z*, 10*Z*)-diene in difficidin) the configuration of the diene in the final product differs from the predicted 2*E*, 4*Z*-configuration of the diene in the corresponding biosynthetic intermediate (Supplementary Figure 1). Further experiments will be required to establish the reason for this discrepancy. One possibility, is that the corresponding DH-like domains bind the substrate in a different orientation, leading to the formation of a (2*Z,* 4*E*)-2, 4-dienoyl thioester. In the case of difficidin, the 8,10-diene is known to be configurationally labile.[22] It is thus possible that the (8*Z*, 10*Z*)-configuration could result from isomerisation of an initially formed (8*E*, 10*Z*) or (8*Z*, 10*E*)-configured isomer.

We identified module pairs containing a type A KR domain in the first module and a type B KR domain alongside a DH-like domain in the second module in the *cis*-AT PKSs that assemble spirangien and thuggacin (Supplementary Fig. 1). The corresponding regions in the structures of spirangien and thuggacin contain dienes with the same configuration as the (2*E,* 4*Z*)-2, 4-dienoyl thioester intermediate predicted to be assembled by these modules, suggesting similar mechanisms may be employed for conjugated *E, Z*-diene incorporation in both *trans* and *cis*-AT PKSs. A related module pair containing an additional ER domain in the second module is present in the PKS that assembles epothilone C, which contains a *Z*-alkene in the corresponding region of its structure (Supplementary Fig. 1). Katz and co-workers proposed that the DH-like domain catalyses formation of a (2*E,* 4*Z*)-2, 4-dienoyl thioester, which is reduced by the ER domain to form the corresponding (4*Z*)-4-enoyl thioester (Supplementary Fig. 1).[23] However, while inactivation of this DH-like domain resulted in production of an epothilone derivative with the expected hydroxyl group at C-13, this compound unexpectedly contained a 10, 11-double bond.[23] It was suggested that a DH domain in a subsequent module may catalyse formation of the 10, 11-double bond, but there is no evidence to support this hypothesis. These putative diene-forming DH-like domains are phylogenetically distinct from both their counterparts in *trans*-AT PKSs and canonical DH domains in *cis*-AT PKSs (Extended Data Fig. 6). Sequence alignments with the DH domain from the EryAII subunit of the 6-DEB PKS, show that they contain the conserved His-Asp catalytic dyad (Extended Data Fig. 5). However, the additional active site His residue that plays a key

catalytic role in diene-forming DH-like domains in *trans*-AT PKSs (*vide supra*), is replaced with a conserved Trp residue (Extended Data Fig. 5). It is therefore likely that a different catalytic mechanism is employed, akin to that used by FabA for the conversion of a (3*R*)-3-hydroxydecanoyl thioester to the corresponding (3*Z*)-3-enoyl thioester,[24] involving sequential deprotonation of C-2 and C-4 by the single active site His residue. Further experiments will be required to confirm that these *cis*-AT PKS DH domain variants are responsible for *E*, *Z*-diene incorporation and to elucidate their catalytic mechanism.

In addition to dehydration of 3-hydroxyacyl thioesters, DH-like domains in *trans*-AT PKSs have been reported to catalyse the conversion of 3-enoyl-7-hydroxy thioesters and 3-enoyl-6-hydroxy thioesters to tetrahydropyrans and tetrahydrofurans, respectively, and the isomerisation of 2-enoyl thioesters to 3-enoyl thioesters.[25–29] Our finding that such domains also catalyse the conversion of 3, 5-dihydroxy thioesters to 2, 4-dienoyl thioesters further expands the catalytic repertoire of DH-like domains. Phylogenetic analysis of all such domains from *trans*-AT PKSs, which possess a common double hotdog fold, shows that they clade according to their catalytic function (Extended Data Fig. 6). This will aid the predictive analysis of novel *trans*-AT PKS systems and prompts us to propose a self-consistent nomenclature for the various classes of domain: DH^{e}, 2-enoyl thioester synthases; DH^{i}, 2-enoyl thioester isomerases; DH^{t}, tetrahydropyran synthases; DH^{d}, 2, 4-dienoyl thioester synthases. It is interesting to note that DH^{d} domains and one group of the DH^{e} domains appear to have evolved from a common ancestor with an additional His residue at position 168 (EryAII DH domain numbering) in the active site (Extended Data Fig. 5 and 6). This group of DH^{e} domains is embedded in the same type of “split module” architecture as the DH^{d} domains (Extended Data Fig. 6), but the upstream KS domain is non-elongating, and the downstream KR domain is absent. Thus, the DH^{e} domains introduce an alkene rather than a diene into the corresponding metabolic products. DH-like domains have been proposed to play yet further roles in modular PKSs, such as diene isomerisation and aromatic ring formation,[30,31] although experimental evidence for these is currently lacking. Going forward, use of this nomenclature to annotate DH-like domains should facilitate understanding of the relationship between active site architecture, catalytic function, and product structure.

In conclusion, our discovery of a novel DH domain variant that converts 3, 5-dihydroxythioesters to 2, 4-dienes expands the β-keto thioester processing toolkit of modular polyketide synthases. This may open up new opportunities for biosynthetic engineering via introduction of an additional His residue into

the active site of canonical DH domains, or domain / module swaps, and further underscores the remarkable catalytic versatility of the double hotdog fold.

Acknowledgements

This research was supported by the BBSRC (BB/L021692/1, BB/K002341/1, and BB/M017982/1 to G.L.C.). M.J. and D.G. were supported by a BBSRC Discovery Fellowship (BB/R012121/1) and the EPSRC MOAC Doctoral Training Centre (EP/F500378/1), respectively. M.R.-C. was supported by a mobility grant from the Government of Andorra (AM025-AND-2015). C.H. was supported by a PhD studentship from the University of Warwick. G.L.C. was the recipient of a Wolfson Research Merit Award from the Royal Society (WM130033).

Author Contributions

M.J., C.H. and G.L.C. designed the study. C.H. D.M.R. and M.R.-C. synthesised the pantetheine thioesters. M.J. overproduced and purified all recombinant proteins. M.J. and C.H. conducted substrate loading and LC-MS analyses of intact proteins. X.J. conducted the cysteamine-mediated product cleavage experiments. M.J. and D.G. conducted bioinformatics analyses. M.J., C.H. and G.L.C. wrote the paper with input from the other authors.

Competing Interests

D.M.R. is an employee and shareholder of Erebagen Ltd. G.L.C. is a Non-Executive Director, shareholder, and paid consultant of Erebagen Ltd. The other authors declare no competing interests.

Figure Legends

Figure 1: Overview of polyketide assembly by *cis*-AT and *trans*-AT modular PKSs, stereochemical details of *E*- and *Z*-alkene formation, examples of *trans*-AT PKS products containing *E*, *Z*-dienes and organisation of modules 12 and 13 of the glaldiolin PKS. a, Comparison of canonical domain architectures in *cis*-AT and *trans*-AT PKSs for assembly of a

hypothetical product (boxed). Domain abbreviations are as follows: AT = acyltransferase, ACP = acyl carrier protein, KS = ketosynthase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase, GNAT = GCN5-related acyltransferase, C-MT = C-methyltransferase. **b**, *E*- and Z-alkene formation in modular PKSs proceeds via reduction of the β-keto thioester by type B and type A KR domains to (3*R*)-3-hydroxy and (3*S*)-3-hydroxy thioesters, respectively, followed by DH domain-catalysed removal of the *pro*-S hydrogen atom at C-2 and elimination of the C-3 hydroxyl group. **c**, Structures of *trans*-AT PKS products containing *E*, *Z*-dienes (highlighted by dashed boxes). **d**, Organisation of modules 12 and 13 of the *trans*-AT PKS responsible for the assembly of gladiolin. Modules 12 and 13 contain type A and type B KR domains, respectively, which are predicted to result in formation of a (3*R*, 5*S*)-3, 5-dihydroxy thioester attached to the module 13 ACP domain. Only module 13 contains a DH-like domain, which is hypothesised to catalyse the elimination of two molecules of water from the 3, 5-dihydroxy thioester to form the corresponding *E*, *Z*-diene.

Figure 2: Stereospecificity of the DH-like domain from module 13 of the gladiolin PKS towards 3-hydroxyacyl and 2-enoyl thioesters. a, Principle of intact protein mass spectrometry assay used. A mass shift of -18 Da indicates dehydration of the loaded substrate. **B**, **c**, The (*3R*)-3-hydroxyhexanoyl thioester (**5**) undergoes dehydration, but the (*3S*)-3-hydroxyhexanoyl thioester (**6**) does not. **d**, **e**, The (2*E*)-2-hexenoyl thioester (**7**) is rehydrated, whereas the (2*Z*)-2-hexenoyl thioester (**8**) is not. Overall, the data are consistent with conversion of a (3*R*)-3-hydroxyacyl ACP thioester intermediate in gladiolin biosynthesis to the corresponding (2*E*)-2-enoyl thioester by the module 13 DH-like domain. Experiments were performed in triplicate and representative data are shown.

Figure 3: The DH-like domain from module 13 of the gladiolin PKS catalyses double dehydration of a (3*R*,5*S*)-3,5-dihydroxyacyl thioester and double rehydration of a (2*E*,4*Z*)-2,4-dienoyl thioester. a, Deconvoluted mass spectrum of the DH-ACP di-domain following conversion of pantetheine thioester **9** to the corresponding CoA thioester and loading onto the *apo*-ACP domain, showing the formation of singly and doubly dehydrated species. **b**, Deconvoluted mass spectrum of the DH-ACP di-domain following conversion of pantetheine thioester **10** to the corresponding CoA thioester and loading onto the *apo*-ACP domain, showing the formation of singly and doubly rehydrated species. Overall, the data are consistent with conversion of a (3*R*,5*S*)-3,5-dihydroxyacyl thioester intermediate

in gladiolin biosynthesis to the corresponding (2*E,*4*Z*)-2,4-dienoyl thioester by the DH-like domain in module 13 of the PKS. Experiments were performed in triplicate and representative data are shown. **c**, Extracted ion chromatograms (EICs) at *m/z* = 247.09 ± 0.02 from LC-MS comparisons of the cysteamine adducts of the 2, 4-dienoyl thioester resulting from double dehydration of the (3*R*,5*S*)-3,5-dihydroxyacyl-ACP thioester (**9**) by the DH-like domain (top) with synthetic standards of (2*E,*4*Z*) and (2*E*, 4*E*)-2,4-dienoyl pantetheine thioesters (**10** and **11**, respectively). Note that (2*E,*4*Z*)-2,4-dienoyl pantetheine thioester (**10**) undergoes facile isomerisation to the thermodynamically more stable 2*E*, 4*E*-isomer, resulting in significant quantities of the latter in the authentic standard and products of the enzymatic reaction.

Figure 4: Dehydration of a (2*E, 5S*)-2-enoyl-5-hydroxy thioester by the DH-like domain and proposed mechanism for diene formation. a, Deconvoluted mass spectrum of the DH-ACP di-domain following conversion of pantetheine thioester **15** to the corresponding CoA thioester and loading onto the *apo*-ACP domain, showing conversion to both diene and diol products. The experiment was performed in triplicate and representative data are shown. The stereochemistry of the major diene / diol formed was assigned by analogy with the data shown in Figure 3. (**b**) Proposed catalytic mechanism of the DH-like domain. Hydrogen bonding of the C-3 and C-5 hydroxyl groups to Asp_{200} orients the substrate for *E*, Z-diene formation. The catalytic His_{23} and His_{158} residues deprotonate C-2 and C-4 to form the enolate and vinylogous enolate intermediates, respectively.

Figure 5: Structural basis for diene formation by DH-like domains. a, Comparison of the active sites of the AlphaFold model of the DH-like domain from module 13 of the gladiolin PKS (top) and the EryAII DH domain (bottom; from PDB entry 3EL6) rendered using PyMol. Although both domains possess the His-Asp catalytic dyad typically found in DH domains, the DH-like domain from module 13 of the gladiolin PKS has an additional His residue (His158) adjacent to that in the catalytic dyad (His23). His158 in the DH-like domain is hypothesised to deprotonate C-4 of the (5*S*)-5-hydroxy-(2*E*)-2-enoyl thioester intermediate, forming a vinylogous enolate which eliminates the (*5S*)-configured hydroxyl group. In canonical DH domains, such as that from EryAII, the additional His residue is replaced by Tyr. **b**, Deconvoluted mass spectra of the H158Y mutant of the DH-ACP di-domain following conversion of pantetheine thioester **15** to the corresponding CoA thioester and loading onto the *apo*-ACP domain.

The mutated di-domain catalyses rehydration of the intermediate analogue to the diol but is unable to dehydrate it to the diene, confirming that His158 plays a key role in diene formation. The experiment was performed in triplicate and representative data are shown.

References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the last 25 years. *J. Nat. Prod.* **70,** 461–477 (2007).

2. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18,** 380–416 (2001).

3. Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed.* **48,** 4688–4716 (2009).

4. Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* **33,** 141–149 (2016).

5. Shah, D. D., You, Y.-O. & Cane, D. E. Stereospecific Formation of *E*-and *Z*-Disubstituted Double Bonds by Dehydratase Domains from Modules 1 and 2 of the Fostriecin Polyketide Synthase. *J. Am. Chem. Soc.* **139**, 14322-14330 (2017).

6. Skiba, M. A. *et al.* Structural Basis of Polyketide Synthase O-Methylation. *ACS Chem. Biol.* **13,** 3221–3228 (2018).

7. Walker, P. D., Weir, A. N. M., Willis, C. L. & Crump, M. P. Polyketide β-branching: diversity, mechanism and selectivity. *Nat. Prod. Rep.* **38,** 723–756 (2021).

8. Meoded, R. A. *et al.* A Polyketide Synthase Component for Oxygen Insertion into Polyketide Backbones. *Angew. Chemie Int. Ed.* **57,** 11644–11648 (2018).

9. Xie, X. & Cane, D. E. Stereospecific Formation of Z-Trisubstituted Double Bonds by the Successive Action of Ketoreductase and Dehydratase Domains from trans-AT Polyketide Synthases. *Biochemistry* **57**, 3126–3129 (2018).

10. Jenner, M. *et al.* Mechanism of intersubunit ketosynthase–dehydratase interaction in polyketide synthases. *Nat. Chem. Biol.* **14,** 270–275 (2018).

11. Li, Y. *et al.* Functional Characterization of a Dehydratase Domain from the Pikromycin Polyketide Synthase. *J. Am. Chem. Soc.* **137,** 7003–7006 (2015).

12. Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal. Biochem.* **324,** 100–5 (2004).

13. Agarwal, V. *et al.* Chemoenzymatic Synthesis of Acyl Coenzyme A Substrates Enables *in Situ* Labeling of Small Molecules and Proteins. *Org. Lett.* **17,** 4452–4455 (2015).

14. Worthington, A. S. *et al.* One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **4,** 44–46 (2006).

15. Quadri, L. E. *et al.* Characterization of Sfp, a *Bacillus subtilis* Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry* **37,** 1585–1595 (1998).

16. Labonte, J. W. & Townsend, C. A. Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chem. Rev.* **113,** 2182–2204 (2012).

17. Belecki, K. & Townsend, C. A. Biochemical Determination of Enzyme-Bound Metabolites: Preferential Accumulation of a Programmed Octaketide on the Enediyne Polyketide Synthase CalE8. *J. Am. Chem. Soc.* **135,** 14339–14348 (2013).

18. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).

19. Wagner, D. T. *et al.* Structural and Functional Trends in Dehydrating Bimodules from trans-Acyltransferase Polyketide Synthases. *Structure* **25,** 1045–1055 (2017).

20. Xie, X. & Cane, D. E. pH-Rate profiles establish that polyketide synthase dehydratase domains utilize a single-base mechanism. *Org. Biomol. Chem.* **16,** 9165–9170 (2018).

21. Fiers, W. D., Dodge, G. J., Sherman, D. H., Smith, J. L. & Aldrich, C. C. Vinylogous Dehydration by a Polyketide Dehydratase Domain in Curacin Biosynthesis. *J. Am. Chem. Soc.* **138,** 16024–16036 (2016).

22. KE, W. *et al.* Difficidin and oxydifficidin: novel broad spectrum antibacterial antibiotics produced by Bacillus subtilis. II. Isolation and physico-chemical characterization. *J. Antibiot.* **40,** 1682–1691 (1987).

23. Tang, L. *et al.* Elucidating the mechanism of cis double bond formation in epothilone biosynthesis. *J. Am. Chem. Soc.* **126,** 46–47 (2004).

24. Dodge, G. J. *et al.* Structural and dynamical rationale for fatty acid unsaturation in Escherichia coli. *Proc. Natl. Acad. Sci.* **116,** 6775–6783 (2019).

25. Pöplau, P., Frank, S., Morinaka, B. I. & Piel, J. An Enzymatic Domain for the Formation of Cyclic Ethers in Complex Polyketides. *Angew. Chem. Int. Ed.* **52,** 13215–13218 (2013).

26. Berkhan, G. & Hahn, F. A dehydratase domain in ambruticin biosynthesis displays additional activity as a pyran-forming cyclase. *Angew. Chem. Int. Ed.* **53,** 14240–14244 (2014).

27. Kusebauch, B., Busch, B., Scherlach, K., Roth, M. & Hertweck, C. Functionally Distinct Modules Operate Two Consecutive α,β→β,γ Double-Bond Shifts in the Rhizoxin Polyketide Assembly Line. *Angew. Chem. Int. Ed.* **49,** 1460–1464 (2010).

28. Gay, D. C., Spear, P. J. & Keatinge-Clay, A. T. A Double-Hotdog with a New Trick: Structure and Mechanism of the *trans* -Acyltransferase Polyketide Synthase Enoyl-isomerase. *ACS Chem. Biol.* **9,** 2374–2381 (2014).

29. Berkhan, G., Merten, C., Holec, C. & Hahn, F. The interplay between a multifunctional dehydratase domain and a C-methyltransferase effects olefin shift in ambruticin biosynthesis. *Angew. Chem. Int. Ed.* **55,** 13589–13592 (2016).

30. Wagner, D. T. *et al.* Structural and Functional Studies of a Pyran Synthase Domain from a *trans* -Acyltransferase Assembly Line. *ACS Chem. Biol.* **13**, 975-983 (2018).

31. SP, N. *et al.* Mining Symbionts of a Spider-Transmitted Fungus Illuminates Uncharted Biosynthetic Pathways to Cytotoxic Benzolactones. *Angew. Chem. Int. Ed.* **59,** 7766–7771 (2020).

Methods

Cloning and Mutagenesis

The cloning of the DH-ACP di-domain and ACP(DH) domain from module 13 of the gladiolin PKS in *B. gladioli* BCC0238 was reported previously.[10] The H158Y mutant of the DH-ACP di-domain was constructed using the Q5 site-directed mutagenesis kit (NEB), with the following primers: DH-ACP(H158Y)_For-(5'-CGGCGTGGTATATGGCCCGTTCATGC-3'), DH-ACP(H158Y)_Rev-(5'-AGCTCGCGCACGCCGCGA-3'). Plasmids were isolated from cultures using a miniprep kit (Thermo), and the inserts were sequenced to verify their integrity.

Protein Overproduction and Purification

Overproduction and purification of module 13 DH-ACP di-domain and the excised module 13 ACP(DH) domain was conducted as described previously.[10] The DH-ACP(H158Y) di-domain mutant was overproduced and purified using the same protocol as that employed for the wild type di-domain.

Substrate Loading and Dehydration Assay

Pantetheine thioesters were converted to the corresponding coenzyme A thioesters and attached to wild type and mutant DH-ACP di-domain, or the excised ACP(DH) domain using previously described methodology.[12-14] Reactions were typically carried out using the DH-ACP di-domain or excised ACP(DH) domain (200 μM) in 20 mM Tris, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP (total volume: 50 μL). Assays were initiated by addition of 2 μM Sfp, 1 μM CoaA, 1 μM CoaD, 1 μM CoaE and 1 mM pantetheine thioester, and were incubated for 3 h at 25 °C. The mixture was diluted to 20 μM with dH_2O prior to UHPLC-ESI-Q-TOF-MS analysis.

UHPLC-ESI-Q-TOF-MS Analysis of Intact Proteins

All assays were analysed on a Bruker MaXis II ESI-Q-TOF-MS connected to a Dionex 3000 RS UHPLC fitted with an ACE C4-300 RP column (100 x 2.1 mm, 5 μm, 30 °C). The column was eluted with a linear gradient of 5–100% MeCN containing 0.1% formic acid over 30 min. The mass spectrometer was operated in positive ion mode with a scan range of 200–3000 m/z. Source conditions were: end plate

offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.8 bar; dry gas (N_2) at 9.0 L min^{-1} ; dry temperature at 200 °C. Ion transfer conditions were: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 200 *m/z*; collision energy at 8.0 eV; collision RF at 2000 Vpp; transfer time at 110.0 µs; pre-pulse storage time at 10.0 µs.

Cysteamine-Mediated Cleavage of Products from the DH-ACP Di-domain

Reactions were carried out using the DH-ACP di-domain (200 µM) in 20 mM Tris, 100 mM NaCl, 10 mM $MgCl_2$, 10 mM ATP (total volume: 400 µL). Assays were initiated by addition of 2 µM Sfp, 1 µM CoaA, 1 µM CoaD, 1 µM CoaE and 1 mM pantetheine thioester, and were incubated for 1.5 h at 25 °C before addition of cysteamine to a final concentration of 0.5 mM. Cystamine thioesters of (2*E*, 4*Z*)- and (2*E*, 4*E*)-hexadienoate were prepared using 1 mM of the (2*E*, 4*Z*)- or (2*E*, 4*E*)-hexadienoyl panthetheine thioesters (compounds **10** and **11**, respectively) in the same reaction conditions, except with the GbnD5 DH-ACP di-domain omitted. After 20 h at 4 °C, reaction solutions were extracted twice with EtOAc (1 mL). The combined extracts were evaporated to dryness, redissolved in 100 µL of MeOH and analysed using UHPLC-ESI-Q-TOF-MS.

UHPLC-ESI-Q-TOF-MS Analysis of Cysteamine Cleavage Products

Cystamine adducts resulting from the cleavage reactions were analysed by UHPLC-ESI-Q-TOF-MS, using a Zorbax Eclipse Plus C18 column (100 × 2.1 mm, 1.8 µm) connected to a Dionex UltiMate 3000 UHPLC coupled to a Bruker MaXis IMPACT mass spectrometer. Mobile phases consisted of water (A) and acetonitrile (B), each supplemented with 0.1% formic acid. A gradient of 5 % B to 100 % B over 42 minutes was employed at a flow rate of 0.2 mL/min. The mass spectrometer was operated in positive ion mode with a scan range of 50 - 3000 *m/z*. Source conditions were as follows: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 °C. Ion transfer conditions were: ion funnel RF at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 *m/z*; collision energy at 5.0 eV; collision RF at 600 Vpp; ion cooler RF at 50 – 350 Vpp; transfer time at 121 µs; pre-pulse storage time at 1 µs. Calibration was performed with 1 mM sodium formate through a loop injection of 20 µL at the start of each run.

Structural Model of DH-like Domain from Module 13 of the Gladiolin PKS

A structural model of the DH-like domain from module 13 of the gladiolin PKS was generated using AlphaFold.[18] The full amino acid sequence of the excised domain (minus the His-Tag) was submitted to the AlphaFold Colab notebook, which uses a slightly simplified version of AlphaFold v2.1.0, with the run_relax parameter enabled.

Data Availability

The minimum dataset necessary to interpret, verify and extend the research is provided in the main manuscript, extended data, and supplementary information. The corresponding raw data (Figures 2, 3, 4 and 5, Extended Data Figures 1 and 2, Supplementary Table 1, and Supplementary Figures 4 and 5), which were processed via standard deconvolution (intact protein mass spectrometry) and extracted ion chromatogram (LC-MS analyses) procedures, are available upon written request to the corresponding author.