Rational Reprogramming of Fungal Polyketide First Ring Cyclization

Yuquan Xu[1,6], Tong Zhou[2,6], Zhengfu Zhou[1,3], Shiyou Su[1,3], Sue A. Roberts[4], William R. Montfort[4,5], Jia Zeng[2], Ming Chen[3], Wei Zhang[3], Min Lin[3], Jixun Zhan[2,*], and István Molnár[1]

[1] Natural Products Center, School of Natural Resources and the Environment, The University of Arizona, 250 E. Valencia Rd., Tucson, AZ 85706, USA [2]Department of Biological Engineering, Utah State University, 4105 Old Main Hill, Logan, UT 84322, USA [3] Biotechnology Research Institute, The Chinese Academy of Agricultural Sciences, 12 Zhongguancun South St., Beijing 100081, P. R. China [4] Department of Chemistry and Biochemistry, The University of Arizona, 1041 E. Lowell St., Tucson, AZ 85721, USA [5] Bio5 Institute, The University of Arizona, 1657 E. Helen St., Tucson, AZ 85721, USA [6] These authors contributed equally to this work. * Corresponding authors: imolnar@email.arizona.edu (I. Molnár); jixun.zhan@usu.edu (J. Zhan)

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Resorcylic acid lactones (RAL) and dihydroxyphenylacetic acid lactones (DAL) represent important pharmacophores with heat shock response and immune system modulatory activities. The biosynthesis of these fungal polyketides involves a pair of collaborating iterative polyketide synthases (iPKSs): a highly reducing iPKS (hrPKS) whose product is further elaborated by a nonreducing iPKS (nrPKS) to yield a 1,3-benzenediol moiety bridged by a macrolactone. Biosynthesis of unreduced polyketides requires the sequestration and programmed cyclization of highly reactive poly-β-ketoacyl intermediates to channel these uncommitted, pluripotent substrates towards defined subsets of the polyketide structural space. Catalyzed by product template (PT) domains of the fungal nrPKSs and discrete aromatase/cyclase enzymes in bacteria, regiospecific first-ring aldol cyclizations result in characteristically different polyketide folding modes. However, a few fungal polyketides, including the DAL dehydrocurvularin, derive from a folding event that is analogous to the bacterial folding mode. The structural basis of such a drastic difference in the way a PT domain acts has not been investigated until now. We report here that the fungal versus the bacterial folding mode difference is portable upon creating hybrid enzymes, and structurally characterize the resulting unnatural products. Using structure-guided active site engineering, we unravel structural contributions to regiospecific aldol condensations, and show that reshaping the cyclization chamber of a PT domain by only three selected point mutations is sufficient to reprogram the dehydrocurvularin nrPKS to produce polyketides with a fungal fold. Such rational control of first ring cyclizations will facilitate efforts towards the engineered biosynthesis of novel chemical diversity from natural unreduced polyketides.

biosynthesis | cyclase | iterative catalysis | natural products | polyketides

Introduction

Fungal polyketides are one of the largest families of structurally diverse natural products with antibiotic, antiproliferative, immunosuppressive, and enzyme inhibitory activities. Importantly, they also provide lead compounds and inspiration for pharmaceutical drug discovery, as evidenced by the statin cholesterol-lowering agents (1, 2). Fungal polyketides are biosynthesized by multi-domain megasynthases (Type I iterative polyketide synthases, iPKSs) that employ ketoacyl synthase (KS), acyl transferase (AT), and acyl carrier protein (ACP) domains to catalyze recursive thio-Claisen condensations using malonyl-CoA extender units. While the architecture of these enzymes is similar to a single module of the bacterial Type I modular PKSs (3), fungal iPKSs use a single set of active sites iteratively, analogous to dissociated bacterial type II PKSs (4). Fungal iPKSs may be classified into three subgroups (5). Highly reducing iPKSs (hrPKSs) generate complex linear or non-aromatic cyclic products by reducing the nascent β-ketoacyl intermediates to the β-alcohol, the alkene, or the alkane after each condensation step, using their ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains to execute a cryptic biosynthetic program (2, 6-8). Partially reducing iPKSs omit enoyl reduction to generate simple cyclic structures (5). Finally, nonreducing iPKSs (nrPKSs) feature no reducing domains, and generate a wide variety of aromatic products. nrPKSs select different starter units by a starter unit:ACP transacylase (SAT) domain (9), and mold the polyketide chains into cyclic products by regiospecific cyclizations. First-ring cyclizations are catalyzed by the product template domains (PT) (10), while the polyketide chains are terminated by Claisen cyclase (11), macrolactone synthase (12) (thioesterase, TE) or reductive release (R) domains (2).

While the biosynthesis of most fungal polyketides requires a single iPKS enzyme, the assembly of the resorcylic acid lactones (RALs) involves a pair of collaborating hrPKS and nrPKS, acting in sequence (12-17). Fungal RALs are rich pharmacophores with estrogen agonist (zearalenone), mitogen-activated protein kinase inhibitory (hypothemycin), and heat shock response modulatory activities (radicicol and monocillin II (**1**), Fig. 1) (18, 19). For these RALs, the hrPKS produces a reduced linear polyketide chain that is directly transferred to the nrPKS (9). The nrPKS further extends the polyketide, closes the first 6-membered ring by aldol condensation, and releases the RAL product by macrolactone formation (**Fig. 1**). We have recently shown that the assembly of 10,11-dehydrocurvularin (**2**), a phytotoxic dihydroxyphenylacetic acid lactone (DAL) from *Aspergillus terreus*, employs a similar chemical modularity principle (20). Curvularins modulate the mammalian immune system by repressing the inducible nitric oxide synthase (iNOS) (21, 22). In addition, both monocillin II (**1**) and 10,11-dehydrocurvularin (**2**) act as promising broad spectrum inhibitors of various cancer cell lines *in vitro* by overwhelming the heat shock response, an evolutionarily conserved coping mechanism of eukaryotic cells that maintains protein homeostasis (23-26).

A crucial step during the programmed biosynthesis of aromatic polyketide natural products is the cyclization of the first ring, catalyzed by the PT domains of the nrPKS (10, 27). This event commits the highly reactive, pluripotent poly-β-ketoacyl

Reserved for Publication Footnotes

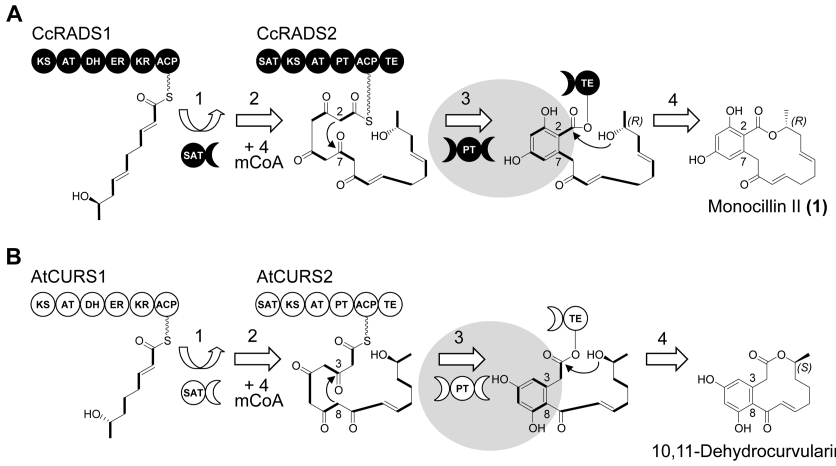

Fig. 1. Biosynthesis of monocillin II (1) and 10,11-dehydrocurvularin (2). **(A)** During the biosynthesis of the radicicol intermediate monocillin II (**1**) in *Chaetomium chiversii* (14), the hrPKS CcRADS1 produces a reduced pentaketide starter unit (13) that is transferred to the nrPKS CcRADS2 by the SAT domain (9) (step 1). After a further four successive condensation events with malonyl-CoA (mCoA, step 2), catalyzed by the KS of the nrPKS CcRADS2, the linear ACP-bound polyketide chain undergoes a C2-C7 aldol condensation catalyzed by the PT domain (10) (step 3). This condensation follows an F-type folding mode (28, 29). **1** is released by macrolactone formation catalyzed by the TE domain (2) (step 4). **(B)** Assembly of **2** in *Aspergillus terreus* AH-02-30-F7 also involves sequentially acting collaborating iPKSs. However, the hrPKS AtCURS1 produces a reduced tetraketide starter, while the AtCURS2 PT domain catalyzes aldol condensation in the C8-C3 register, following an S-type folding mode (28). C-C bonds in bold indicate intact acetate equivalents (malonate-derived C_2 units) incorporated into the polyketide chain by the iPKSs.

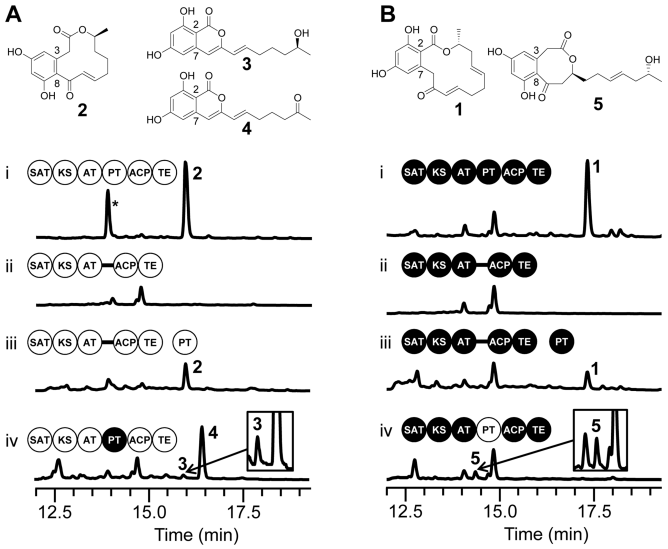

Fig. 2. Engineering first ring cyclization regiospecificity by domain replacements.(**A**) Product profiles (HPLC traces recorded at 300 nm) of *S. cerevisiae* BJ5464-NpgA(13, 34) co-transformed with YEpAtCURS1 and the indicated YEpATCURS2 derivatives (**SI Methods**): (i) YEpATCURS2; (ii) YEpATCURS2-ΔPT; (iii) YEpATCURS2-ΔPT+$PT_{AtCURS2}$; (iv) YEpATCURS2-$PT_{CcRADS2}$. The peak in trace (i) labeled with the asterisk corresponds to 11-hydroxycurvularin, a spontaneous hydration product of **2**. (**B**) Product profiles (HPLC traces recorded at 300 nm) of *S. cerevisiae* BJ5464-NpgA (13, 34) co-transformed with YEpCcRADS1 and the indicated YEpCcRADS2 derivatives (**SI Methods**): (i) YEpCcRADS2; (ii) YEpCcRADS2-ΔPT; (iii) YEpCcRADS2-ΔPT+$PT_{CcRADS2}$; (iv) YEpCcRADS2-$PT_{AtCURS2}$. Polyketide products were characterized based on their UV, ESI-MS, NMR and CD spectra, as well as Mosher's method (see **SI Methods**for details on isolation and chemical characterization).

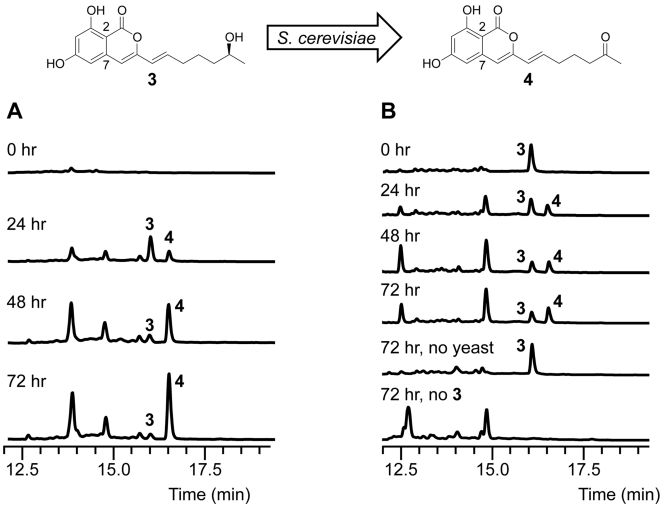

Fig. 3. Fortuitous oxidation of isocoumarin 3 by *S. cerevisiae* BJ5464-NpgA. (**A**) Product profiles (HPLC traces recorded at 300 nm) of *S. cerevisiae* BJ5464-NpgA (13, 34) co-transformed with YEpAtCURS1 and YEpATCURS2-$PT_{CcRADS2}$ and cultivated for the indicated amount of time. (**B**) HPLC analysis of the bioconversion of **3** into **4** by *S. cerevisiae* BJ5464-NpgA.

chains towards defined structural classes of the possible polyketide scaffold space. PT-catalyzed cyclizations most often follow an F-type pattern whereby the benzene ring is assembled from two intact malonate-derived C_2 units and from two bridging carbons from two additional acetate equivalents (28, 29) (**Fig. 1**). F-type first ring cyclizations typically result from aldol condensations in the C2-C7 (as in **1**), C4-C9 or C6-C11 register. In contrast, bacterial polyketide cyclase/aromatase enzymes (parts of Type II PKS multienzyme complexes) typically direct an S-type folding event whereby the carbons of the benzene ring are derived from three intact malonate-derived C_2 units (1, 28). Nevertheless, a select few fungal polyketides, including DALs like **2** feature a first ring connectivity that is analogous to the S-type folding mode, resulting from an unorthodox C8-C3 aldol cyclization event (20). The bacterial aromatase/cylcase enzymes show little sequence similarity to fungal PT domains, and feature a different protein fold and active site architecture as a prominent example of convergent evolution (10, 30-32). The sequences of fungal PT domains catalyzing F-type cyclization can be classified into seven clades according to their regiospecificity and the length of their product (29, 33). In spite of catalyzing an atypical S-type folding,

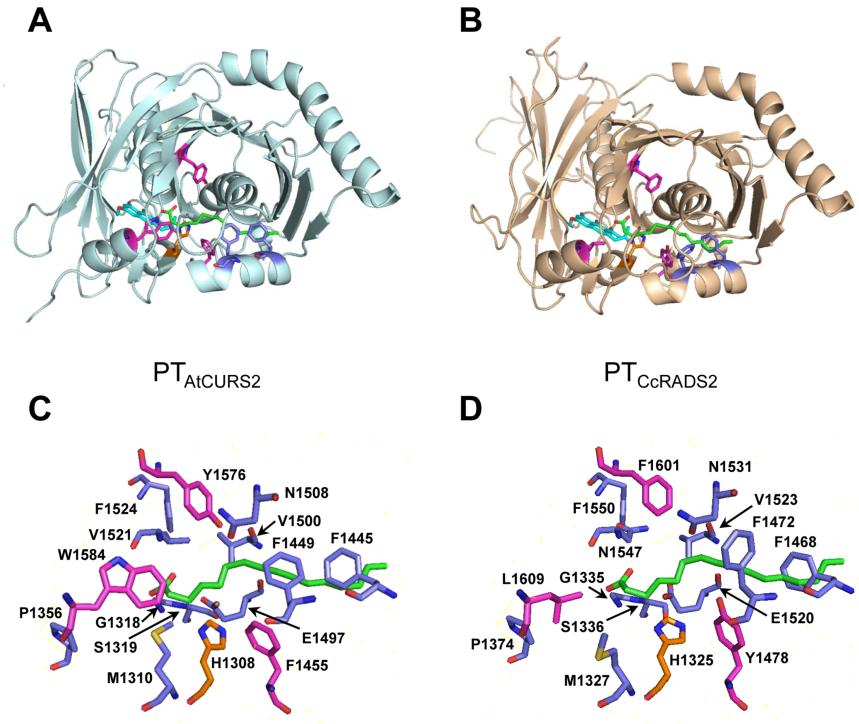

Fig. 4. Homology models of $PT_{CcRADS2}$ and $PT_{AtCURS2}$.Cartoon views of the homology models, and stick representation of the amino acids lining the cyclization chambers of (**A, C**) $PT_{AtCURS2}$ and (**B, D**) $PT_{CcRADS2}$. The side chains of residues discussed in the text are shown as golden sticks (catalytic histidine), magenta sticks (residues implicated in differentiating substrate orientation), and blue sticks (other significant residues). Green sticks: The substrate analog palmitic acid resident in the PT_{NSAS} structure 3HRQ (10).

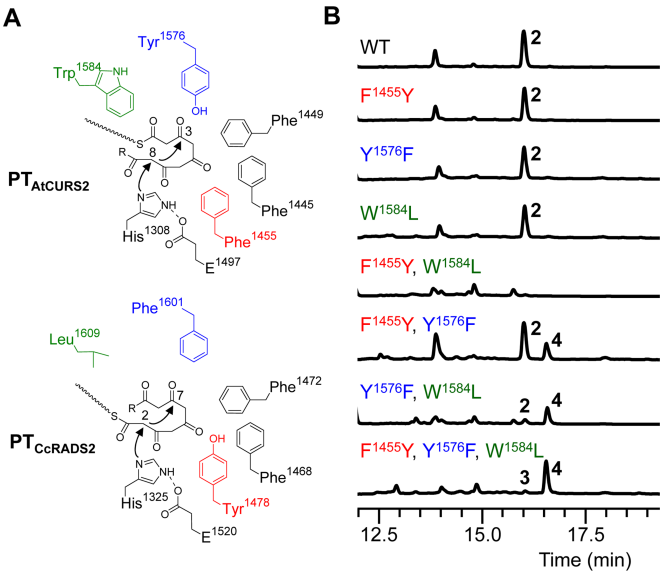

Fig. 5. Reprogramming first ring cyclization regiospecificity by site directed mutagenesis. (**A**) Proposed mechanism of regiospecific cyclizations catalyzed by the $PT_{AtCURS2}$ and the $PT_{CcRADS2}$ domains. Enolate formation by deprotonation of C8 (AtCURS2) or C2 (CcRADS2) is promoted by the histidine catalytic base, polarized by a conserved glutamic acid (AtCURS2: H^{1308} and E^{1497}, CcRADS2: H^{1325} and E^{1520}). The substrates are proposed to thread via alternate routes as directed by different gating residues at the entrance of the cyclization chamber (AtCURS2: W^{1584}, CcRADS2: L^{1609}), and further oriented by an inverted hydrophobic/H-bond donor residue pair lining the rear end of the cyclization chamber (AtCURS2: F^{1455} and Y^{1576}, CcRADS2: Y^{1478} and F^{1601}). (**B**) Product profiles (HPLC traces recorded at 300 nm) of *S. cerevisiae* BJ5464-NpgA (13, 34) co-transformed with YEpAtCURS1 and the YEpATCURS2 derivatives encoding the indicated $PT_{AtCURS2}$ domain mutations in AtCURS2.

the PT domain of the AtCURS2 curvularin synthase nrPKS is firmly rooted in the C2-C7 clade of PTs which yield RALs like **1** (20). This result concurs with previous observations that fungal iPKS evolve orthogonal product specificities primarily by point mutations, and not by domain shuffling amongst distinct enzymes (2).

The present work aimed to define how different regiospecific outcomes for first ring cyclization are programmed into nrPKS enzymes. By exploiting the orthogonal aldol condensation regiospecificities of the related PT domains of the nrPKSs for **1** and **2**, we attempted to alter this program to switch F and S type cyclization modes. Achieving precise control of regiospecificity during the engineered biosynthesis of fungal polyketides is central to producing biologically active "unnatural products", and may guide efforts to generate novel chemical diversity from natural non-reduced fungal polyketides.

Results and Discussion

PT domains are necessary for programmed polyketide formation. Throughout this study, we have used an *in vivo* reconstituted system for polyketide production, whereby recombinant hrPKS + nrPKS pairs are expressed from compatible plasmids in the host *Saccharomyces cerevisiae* BJ5464-NpgA to produce **1** (9 mg/l, isolated yield), **2** (6 mg/l), and their derivatives (13, 34). Deletion of PT domains has previously been shown in the Tang and Townsend laboratories to yield shunt metabolites whose backbones have undergone spontaneous cyclizations (12, 27, 29). Thus, at the start of this work we considered it to be possible that $PT_{AtCURS2}$ is simply an inactive enzyme that does not contribute to the folding of the nascent polyketide chain, and thus the DAL scaffold of **2** is a serendipitous derailment product retained by evolution. To exclude this possibility, we have deleted the PT domain of AtCURS2 as well as the PT domain of CcRADS2, the radicicol/monocillin II nrPKS of *Chaetomium chiversii* (14). However, the corresponding yeast expression strains produced no polyketides (**Fig. 2**). Complementation of the PT-less nrPKSs with their dissected PT domains, expressed as separate ORFs *in trans*, led to the rescue of the production of the native products **1** (1 mg/l) and **2** (2 mg/l), respectively (**Fig. 2**). The success

of this experiment contrasts previous *in trans* domain complementation attempts that did not yield observable products *in vivo* (1, 29), and emphasizes that appropriate protein-protein docking, substrate recognition and processing still can take place with dissected domains expressed in heterologous hosts. Taken together, these experiments show that the PT-less nrPKSs and the freestanding PT domains are all catalytically active. The absence of RAL/DAL product formation in the absence of PT domains thus indicates that these domains fulfill an essential role in both systems, and that un-templated poly-β-ketones are not released by these nrPKSs.

Portability of aldol cyclization programs. Recent experiments to replace PT domains in nrPKS model systems showed that the register of F-type aldol condensations may be switched (29, 35), with the incoming PTs enforcing a subtle shift of the substrate chains in the catalytic chambers to expose different carbons (C2, C4 or C6) to the catalytic histidine for deprotonation and enolate formation (10). Thus, we were interested to see whether first ring cyclizations may also be reconfigured *between* F- and S-type folding modes by exchanging PT domains. This would require radical re-routing of the polyketide substrate chain to expose a carbon to the catalytic base that is distal (S-type) or proximal (F-type) to the phosphopantetheine thioester. As shown above and observed previously by others (1, 29, 36), *in trans* reconstitution of iPKSs from dissected domains incurs a penalty for catalysis, in terms of product yield and/or fidelity. Thus, we elected to conduct these experiments with nrPKSs where the heterologous PT domains replace their native equivalents *in cis* (29). Replacement of the PT domain of AtCURS2 with $PT_{CcRADS2}$ led to the production of two isocoumarins, both resulting from C2-C7 aldol cyclizations. The priming acyl chain of **3** (0.3 mg/l) corresponds to the expected product of the hrPKS AtCURS1, while the major product **4** (3 mg/l) features a carbonyl at C15 (**Fig. 2A**). The formation of **3** and **4** indicates that $PT_{CcRADS2}$ is able to process a shorter substrate (an octaketide as opposed to its native nonaketide) while faithfully executing an F-type first ring closure. Time course analysis of the fermentation with *S. cerevisiae* BJ5464-NpgA [YEpAtCURS1 and YEpATCURS2-$PT_{CcRADS2}$] (**Fig. 3A**) shows that **3** is the primary product 24 hr after the induction of polyketide production, but by 48 hr the formerly minor product **4** becomes dominant. Extending the cultivation to 72 hr and beyond increased only the production of **4** but did not eliminate **3**, nor did it lead to the production of additional polyketide products. Similarly, incubation of purified **3** with the untransformed yeast host strain *S. cerevisiae* BJ5464-NpgA led to the gradual, albeit not complete, biotransformation of **3** to **4** (Fig. 3B). It was confirmed that the untransformed yeast host strain does not produce **3** or **4**, nor is **3** converted to **4** by spontaneous oxidation in the culture medium in the absence of yeast cells **(Fig. 3B)**. Thus, **4** derives from a chance oxidation of the 15-OH of **3** by an endogenous enzyme of the yeast host.

The yeast strain co-expressing CcRADS1 and CcRADS2-$PT_{AtCURS2}$ yielded a novel compound (**5**, **Fig. 2B**), although with a low productivity (0.3 mg/l). Structural characterization of this product revealed that **5** harbors a novel carbon skeleton featuring a C8-C3 dihydroxyphenylacetic acid moiety bridged by an 8-membered lactone (**SI Methods**). Thus, $PT_{AtCURS2}$ is competent to process a longer substrate (a nonaketide as opposed to its native octaketide) while retaining its ability to direct an S-type folding and cyclization event. The 4-oxo-2-oxacyclooctanone ring of **5** may be produced by the facile attack of the C1 carboxyl on C11 of the enone; the involvement of the TE domain in this reaction cannot be excluded at this point.

Collectively, these experiments show that the F- or S-type regiospecificities of first ring cyclizations are solely programmed into the PT domains of collaborating nrPKSs (29, 35). This programming is portable amongst nrPKS platforms, without influencing starter unit choice or the number of extensions carried out by the rest of the chassis. The formation of the isocoumarins **3** and **4** and the 8-membered lactone **5** suggests that the nrPKS TE domains may hydrolyze products with switched aldol condensation patterns, but they are unable to form a macrolactone using a carbon chain with an isomeric fold. The low product yield of **5** suggests that CcRADS2 is a more stringent chassis, less amenable to combinatorial replacement of its domains.

Homology models of $PT_{AtCURS2}$ and $PT_{CcRADS2}$. To identify the structural basis of the programming of regiospecificity in PT domains, we have created homology models of the $PT_{AtCURS2}$ and the $PT_{CcRADS2}$ domains, based on the experimentally determined structure of the PT domain of NSAS from *Aspergillus parasiticus* (PT_{NSAS}, PDB ID: 3HRR and 3HRQ) (10). The nrPKS NSAS catalyzes the formation of norsorolinic acid, a C_{20} polyketide primed with hexanoic acid, with the PT_{NSAS} directing an F-type folding mode first ring cyclization event (28) in the C4-C9 register, followed by a second ring closure at C2-C11.

In spite of relatively low sequence similarities with PT_{NSAS} ($PT_{AtCURS2}$: 22% identity and 41% similarity; $PT_{CcRADS2}$: 20% identity and 40% similarity), combined structural evaluation and fold recognition scores of 3.6 ($PT_{CcRADS2}$) and 3.47 ($PT_{AtCURS2}$) obtained from the reliability assessment engine PCONS5 (37) indicated that the fold recognition is reliable for both $PT_{AtCURS2}$ and $PT_{CcRADS2}$ (a PCONS score > 2.17 is considered reliable, a score > 1.5 is considered significant). PT_{NSAS} features a long, straight cyclization chamber and a hydrophobic hexyl-binging region that accommodates the starter unit, with the substrate bound in an extended conformation (10). In contrast, homology modeling had suggested that PKS4, the zearalenone nrPKS from *Gibberella fujikuroi* (10, 16, 38) contains a PT domain with a wider, curved catalytic chamber where the substrate adopts a bent conformation (10). Similar to the model proposed for PT_{PKS4}, the hydrophobic hexyl-binding region present in PT_{NSAS} was found to be closed off in both $PT_{AtCURS2}$ and $PT_{CcRADS2}$ by the bulky side chains of two phenylalanine residues (**Fig. 4**). First, a replacement of M^{1495} of NSAS by phenylalanine (AtCURS2: F^{1449}, CcRADS2: F^{1472}) narrows the hexyl-binding region. Next, the side chain of another phenylalanine residue in place of G^{1491} (AtCURS2: F^{1445}, CcRADS2: F^{1468}) directly clashes with the tail of the palmitic acid that occupies this pocket in structure 3HRR. This latter phenylalanine is also conserved not only in all other RAL PT domains (14-16, 38), but also in clades II, III and V of characterized PT domains (29). Thus, the hrPKS-derived reduced acyl chains of the nascent intermediates for **1** and **2** may not be sequestered in a deep, buried pocket (10) in these enzymes.

The RAL/DAL PT models retain the large substrate binding chamber where cyclization occurs (10). This cyclization chamber appears constricted at the residues corresponding to V^{1347} and A^{1397} of NSAS (AtCURS2: M^{1310} and P^{1356}, CcRADS2: M^{1327} and P^{1374}) but this is compensated by substitutions with less bulky side chains corresponding to P^{1355} and W^{1571} (AtCURS2: G^{1318} and F^{1524}, CcRADS2: G^{1335} and F^{1550}, **Fig. 4**). The cyclization chamber is proposed to feature an active site dyad (AtCURS2: H^{1308}, E^{1497}; CcRADS2: H^{1325}, E^{1520}) where the aspartic acid (D^{1543}) that polarizes the catalytic base (H^{1345}) in NSAS is replaced by glutamic acid. This functionally conserved replacement is present in all known RAL and DAL PT domains (14-16, 38), but not in clades II-V of functionally characterized PT domains (29). The same D to E replacement is nonetheless common in dehydratases with a fold similar to those of PT domains, and was also found in some Type II PKS aromatase/cyclase enzymes (32). After deprotonation of C8 (AtCURS2) or C2 (CcRADS2) by the catalytic base, the enolate intermediate is thought to be stabilized by the backbone amine of V^{1521} (AtCURS2) or N^{1547} (CcRADS2) (10). After the collapse of the enolate and aldol addition to the

carbonyl, the oxyanion may be stabilized by a network of water molecules that are coordinated in NSAS by S^{1356}, D^{1543}, T^{1546} and N^{1568}. Only some of these residues are conserved in AtCURS2 (S^{1319}, E^{1497}, V^{1500}, and V^{1521}) and CcRADS2 (S^{1336}, E^{1520}, V^{1523}, and N^{1547}), and similar replacements are also present in all RAL PT domains (14-16, 38). The second half of the oxyanion hole is provided by the backbone amine of a glycine in DH domains and hydratases with which PT domains share a double hot dog fold and a proposed evolutionary origin (10). For NSAS, this glycine was seen to be replaced by P^{1355}, but this residue is restored to glycine in all known RAL PT domains (14-16, 38) (AtCURS2: G^{1318}; CcRADS2: G^{1335}). The corresponding position is occupied by serine in clades II, III and V of PT domains (29). The electrophilic carbonyl that takes part in the aldol cyclization is polarized via hydrogen bonding through the same water network, while hydrogen bonding with an asparagine that is conserved in all functionally characterized PT domains (29) may help to orient the substrate in the chamber (NSAS: N^{1554}; AtCURS2: N^{1508}, CcRADS2: N^{1531}).

Conversion of S- and F-type folding modes by structure-based site-directed mutagenesis. A superimposition of the models for $PT_{AtCURS2}$ and $PT_{CcRADS2}$ (47% sequence identity and 65% similarity) showed that the active sites and the cyclization chambers of these two enzymes are highly conserved. Nevertheless, $PT_{AtCURS2}$ forms a first ring in the C8-C3 register with a folding mode analogous to the S-type, while $PT_{CcRADS2}$ catalyzes an F-type folding mode (28) first ring closure in the C2-C7 register. We have identified three key differences that we hypothesized would result in a change in substrate orientation in the binding pocket, and lead to the orthogonal cyclization regiospecificities observed in **1** *vs*. **2** (**Fig. 5A**). First, L^{1609} of $PT_{CcRADS2}$ is replaced by W^{1584} near the substrate entrance in $PT_{AtCURS2}$. The bulky side chain of this tryptophan narrows the entrance of the cyclization chamber of $PT_{AtCURS2}$, and may serve to direct C2 of the penetrating acyl chain away from the catalytic histidine. Leucine is strictly conserved at this position in all characterized RAL PTs (14-16, 38) and predominates (with methionine as an alternative) in clade II-V PTs catalyzing various F-type cyclizations (29). Next, both $PT_{AtCURS2}$ and $PT_{CcRADS2}$ (as well as all other characterized RAL PTs (29)) display a tyrosine-phenylalanine residue pair on opposing faces at the rear of the binding pocket. Remarkably, these residues are inverted in $PT_{AtCURS2}$ *vs*. all known RAL PT domains ($PT_{AtCURS2}$: F^{1455} and Y^{1576}, $PT_{CcRADS2}$: Y^{1478} and F^{1601}). By participating in hydrogen bond networks (Y) or by contributing to a hydrophobic surface of the pocket (F), these residues may help to position the chain such that either C8 ($PT_{AtCURS2}$) or C2 ($PT_{CcRADS2}$) would be presented to the catalytic base ($PT_{AtCURS2}$: H^{1308}, $PT_{CcRADS2}$: H^{1325}), leading to S-type ($PT_{AtCURS2}$) or F-type ($PT_{CcRADS2}$) cyclization outcomes. Notably, all three distinguishing residues (W^{1584}, F^{1455} and Y^{1576}) are conserved between $PT_{AtCURS2}$ and the PT domain of an orphan hrPKS-nrPKS system in the genome of *Pyrenophora tritici-repentis* PT-1C-BFP (GenBank: EDU47225 and EDU47223): this putative DAL synthase is the closest ortholog of the dehydrocurvularin synthase AtCURS1-AtCURS2 (20). Although the highlighted residues are positioned such that substrate binding is expected to be affected, verification of the specific contacts must await the determination of the crystal structures of these domains with bound substrates/products. In the meantime, these three residues may serve as distinguishing sequence signatures to predict RAL *vs.* DAL formation by orphan biosynthetic systems found in sequenced fungal genomes.

To test our structural analysis, we systematically replaced one, two or all three of the identified residues (F^{1455}, Y^{1576}, and W^{1584}) in AtCURS2 with their counterparts of CcRADS2, and co-expressed these enzymes with AtCURS1 in yeast. Single mutations did not alter the regiospecificity of first ring closure in the product (**Fig. 5B**). Double mutations either eliminated product formation ($F^{1455}Y$ + $W^{1584}L$), or yielded a mixture of **2** and **4**, indicating a relaxed aldol condensation regiospecificity for some of these mutant enzymes. Finally, the enzyme with all three mutations produced only **3** and **4** (0.2 and 2 mg/l, respectively), with no detectable **2**. Thus, these three selected point mutations were sufficient to completely transform the native C8-C3 (S-type) regiospecificity of $PT_{AtCURS2}$ to C2-C7 (F-type, **Fig. 5B**). The converse experiment (replacing Y^{1478}, F^{1601} and L^{1609} in CcRADS2 with the corresponding residues of AtCURS2 in all combinations) reduced the yield or completely eliminated the production of **1**, but did not provide **5** or any other detectable C8-C3 DAL (**Fig. S1**). The absence of the expected product may not be surprising if we consider the low yield of **5** even with CcRADS2-$PT_{AtCURS2}$, where the incoming $PT_{AtCURS2}$ domain has presumably been optimized by evolution for the effective synthesis of C8-C3 products. A similar recalcitrance to alteration of stereocontrol has been noted for KR and ER domains during site-directed mutagenesis (but not during complete domain exchanges) in the context of modular PKS systems. Presumably, any proofreading activity from downstream domains (e.g. the TE in our system) may override the effects of subtle alterations of the active site architecture (3, 39, 40).

The current work affirms that first ring cyclization regiospecificity in fungal collaborating iPKSs is programmed in the composition and geometry of the cyclization chambers of the PT domains. Exploiting structural information on PT domains, we have gained insight into the origins of the programming of this folding specificity. Replacement of just three select residues of the product cyclization chambers in a keyhole surgery-like approach converted a PT domain from an atypical, C8-C3-specific, S-type folding mode cyclase into a typical, C2-C7-selective, F type folding mode enzyme, as predicted. The identified signature residues may be used to predict polyketide folding modes in orphan RAL/DAL biosynthetic systems. More importantly, rational reprogramming of polyketide folding modes in fungal iPKSs opens new possibilities for the engineered biosynthesis of novel unnatural polyketides with isomeric folds, including cancer cell proliferation inhibitors and immune system modulators.

Materials and Methods

Strains and Culture Conditions. *E. coli* DH10B and plasmid pJET1.2 (Fermentas) were used for routine cloning and sequencing. *Saccharomyces cerevisiae* BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*) (34, 41) was maintained on yeast extract peptone dextrose agar (YPD, Difco), and transformed using the small scale lithium chloride protocol (42). The yeast – *E. coli* shuttle vectors YEpADH2p-FLAG-URA and YEpADH2p-FLAG-TRP (20) are based on the YEpADH2p vectors with the URA3 or with the TRP1 selectable markers (13). Primers used in this study are listed in **Table S2**. Details on the construction of gene variants and expression constructs are described in the ***SI Methods***. For each recombinant yeast strain, three to five independent transformants were analyzed for the production of polyketides by small scale fermentation, and fermentations with representative isolates were repeated at least three times to confirm results.

Small scale fermentation and analysis of products. Yeast strains were cultured in 50 mL of SC medium (0.67% yeast nitrogen base, 2% glucose, and 0.72 g/L Trp/Ura DropOut supplement) at 30 °C with shaking at 250 rpm. When the OD_{600} reached 0.6, an equal volume of YP medium (1% yeast extract, 2% peptone) was added to the cultures, and the fermentation was continued at 30 °C with shaking at 250 rpm for an additional 2 days. The cultures were adjusted to pH 5.0, and extracted with equal volumes of ethyl acetate three times. The collected organic extracts were evaporated to dryness and analyzed by reversed phase HPLC (Kromasil C18 column, 5 µm, 4.6 mm × 250 mm; eluted with 5% aqueous acetonitrile for 5 min, followed by a linear gradient of 5-95% CH_3CN over 10 min, and 95% CH_3CN for 10 min at a flow rate of 0.8 mL/min; detection at 300 nm). Analysis of the time course of the production of **3** and **4**, biotransformation of **3** to **4** by *Saccharomyces cerevisiae* BJ5464-NpgA, and scale-up of fermentations and isolation of polyketide products for structure elucidation are described in the **SI Methods**.

Chemical characterization of polyketide products. Optical rotations were recorded on a Rudolph Autopol IV polarimeter with a 10-cm cell. CD spectra were acquired with a JASCO J-810 instrument using a path length of 1 cm. ^{1}H, ^{13}C, and 2D NMR (COSY, HSQC, HMBC, ROESY) spectra were recorded in DMSO-d_6, CD_3OD or C_5D_5N on a JEOL ECX-300 spectrometer. ESI-MS data were collected on an Agilent 6130 Single Quad LC-MS. See **SI Methods** for details.

Homology Modeling. Sequences of the $PT_{AtCURS2}$ and $PT_{CcRADS2}$ domains were submitted to the BioInfoBank Meta Server (3DJury) (43). Models based upon the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank PKsA structures 3HRQ and 3HRR (10) were generated by several homology modeling servers. Carbon-alpha models generated by the SAM-HMM server (44) were ranked by 3DJury as most representative and chosen for use. All atom models were created using MODELLER (45). There were no significant differences between the models based upon 3HRQ and those based upon 3HRR. PCONS5 (37) was used to evaluate fold recognition.

Acknowledgements.

This work was supported by grants from the National Science Foundation (MCB-0948751 to I. M.), the National Institutes of Health (AI065357 to J. Z., HL062969 to W. R. M.), Utah State University (Seed Program to Advance Research Collaborations, to J. Z.), the National Research and Development Project of Transgenic Crops of China (2011ZX08009-003-002 to L. M.) and the National Basic Research 973 Program of China (2013CB733903 to L. M.). Professors N. A. Da Silva (University of California, Irvine) and Y. Tang (University of California, Los Angeles) are thanked for providing the yeast host strain and expression vectors.

1. Crawford JM & Townsend CA (2010) New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* 8:879-889.
2. Chooi YH & Tang Y (2012) Navigating the fungal polyketide chemical space: From genes to molecules. *J. Org. Chem.* 77 9933-9953.
3. Keatinge-Clay AT (2012) The structures of type I polyketide synthases. *Nat. Prod. Rep.* 29:1050-1073.
4. Zhan J (2009) Biosynthesis of bacterial aromatic polyketides. *Curr. Top. Med. Chem.* 9:1958-1610.
5. Cox RJ (2007) Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* 5:2010-2026.
6. Zhou H, *et al.* (2012) A fungal ketoreductase domain that displays substrate-dependent stereospecificity. *Nat. Chem. Biol.* 8:331-333.
7. Fisch KM, *et al.* (2011) Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* 133:16635-16641.
8. Yakasai AA, *et al.* (2011) Nongenetic reprogramming of a fungal highly reducing polyketide synthase. *J. Am. Chem. Soc.* 133:10990-10998.
9. Foulke-Abel J & Townsend CA (2012) Demonstration of starter unit Interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *ChemBioChem* 13:1880-1884.
10. Crawford JM, *et al.* (2009) Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* 461:1139-1143.
11. Korman TP, *et al.* (2010) Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc. Natl. Acad. Sci. USA* 107:6246-6251.
12. Zhou H, Zhan J, Watanabe K, Xie X, & Tang Y (2008) A polyketide macrolactone synthase from the filamentous fungus *Gibberella zeae*. *Proc. Natl. Acad. Sci. USA* 105:6249-6254.
13. Zhou H, Qiao K, Gao Z, Vederas JC, & Tang Y (2010) Insights into radicicol biosynthesis via heterologous synthesis of intermediates and analogs. *J. Biol. Chem.* 285:41412-41421.
14. Wang S, *et al.* (2008) Functional characterization of the biosynthesis of radicicol, an Hsp90 inhibitor resorcylic acid lactone from *Chaetomium chiversii*. *Chem. Biol.* 15:1328-1338.
15. Reeves CD, Hu Z, Reid R, & Kealey JT (2008) Genes for biosynthesis of the fungal polyketides hypothemycin from *Hypomyces subiculosus* and radicicol from *Pochonia chlamydosporia*. *Appl. Environ. Microbiol.* 74:5121-5129.
16. Kim YT, *et al.* (2005) Two different polyketide synthase genes are required for synthesis of zearalenone in *Gibberella zeae*. *Mol. Microbiol.* 58:1102-1113.
17. Gaffoor I, *et al.* (2005) Functional analysis of the polyketide synthase genes in the filamentous fungus *Gibberella zeae* (anamorph *Fusarium graminearum*). *Eukaryot. Cell* 4:1926-1933.
18. Winssinger N, Fontaine JG, & Barluenga S (2009) Hsp90 inhibition with resorcyclic acid lactones (RALs). *Curr. Top. Med. Chem.* 9:1419-1435.
19. Winssinger N & Barluenga S (2007) Chemistry and biology of resorcylic acid lactones. *Chem. Commun. (Camb).* 2007:22-36.
20. Xu Y, *et al.* (2013) Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat-shock response modulator anticancer fungal polyketide from *Aspergillus terreus*. *Appl. Environ. Microbiol.* DOI: 10.1128/aem.03334-12.
21. Schmidt N, *et al.* (2010) Transcriptional and post-transcriptional regulation of iNOS expression in human chondrocytes. *Biochem. Pharmacol.* 79:722-732.
22. Elzner S, *et al.* (2008) Inhibitors of inducible NO synthase expression: total synthesis of (*S*)-curvularin and its ring homologues. *ChemMedChem* 3:924-939.
23. Santagata S, *et al.* (2012) Using the heat-shock response to discover anticancer compounds that target protein homeostasis. *ACS Chem. Biol.* 7:340-349.
24. McLellan CA, *et al.* (2007) A rhizosphere fungus enhances *Arabidopsis* thermotolerance through production of an HSP90 inhibitor. *Plant Physiol.* 145:174-182.
25. Workman P, Burrows F, Neckers L, & Rosen N (2007) Drugging the cancer chaperone HSP90: Combinatorial therapeutic exploitation of oncogene addiction and tumor stress. *Ann. N.Y. Acad. Sci.* 1113:202-216.
26. McDonald E, Workman P, & Jones K (2006) Inhibitors of the HSP90 molecular chaperone: attacking the master regulator in cancer. *Curr. Top. Med. Chem.* 6:1091-1107.
27. Crawford JM, *et al.* (2008) Deconstruction of iterative multidomain polyketide synthase function. *Science* 320:243-246.
28. Thomas R (2001) A biosynthetic classification of fungal and streptomycete fused-ring aromatic polyketides. *ChemBioChem* 2:612-627.
29. Li Y, Xu W, & Tang Y (2010) Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J. Biol. Chem.* 285:22764-22773.
30. Lee MY, Ames BD, & Tsai SC (2012) Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI. *Biochemistry* 51:3079-3091.
31. Ames BD, *et al.* (2008) Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides. *Proc. Natl. Acad. Sci. USA* 105:5349-5354.
32. Ames BD, *et al.* (2011) Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway. *Biochemistry* 50:8392-8406.
33. Ahuja M, *et al.* (2012) Illuminating the diversity of aromatic polyketide synthases in *Aspergillus nidulans*. *J. Am. Chem. Soc.* 134:8212-8221.
34. Ma SM, *et al.* (2009) Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* 326:589-592.
35. Vagstad AL, *et al.* (2013) Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew. Chem. Int. Ed.* 52:1718-1721.
36. Vagstad AL, Bumpus SB, Belecki K, Kelleher NL, & Townsend CA (2012) Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J Am Chem Soc* 134:6865-6877.
37. Wallner B & Elofsson A (2005) Pcons5: combining consensus, structural evaluation and fold recognition scores. *Bioinformatics* 21:4248-4254.
38. Gaffoor I & Trail F (2006) Characterization of two polyketide synthase genes involved in zearalenone biosynthesis in *Gibberella zeae*. *Appl. Environ. Microbiol.* 72:1793-1799.
39. Kwan DH & Leadlay PF (2010) Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol.* 5:829-838.
40. Kwan DH, Tosin M, Schlager N, Schulz F, & Leadlay PF (2011) Insights into the stereospecificity of ketoreduction in a modular polyketide synthase. *Org. Biomol. Chem.* 9:2053-2056.
41. Lee KK, Da Silva NA, & Kealey JT (2009) Determination of the extent of phosphopantetheinylation of polyketide synthases expressed in *Escherichia coli* and *Saccharomyces cerevisiae*. *Anal. Biochem.* 394:75-80.
42. Gietz RD & Schiestl RH (2007) Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2:35-37.
43. Ginalski K, Elofsson A, Fischer D, & Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19:1015-1018.
44. Karplus K (2009) SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Res* 37:W492-497.
45. Eswar N, *et al.* (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5:Unit 5.6.