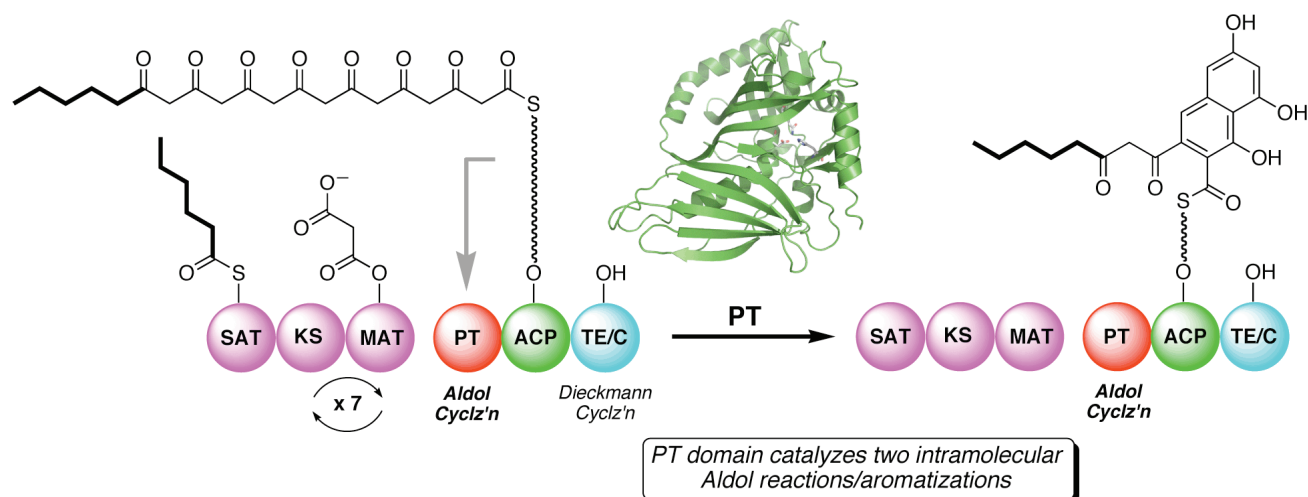


SUPPLEMENTARY INFORMATION

**This pdf file includes:**

- 1) Supplementary Discussion
 - (a) Crystallization and structure determination
 - (b) Detailed structural description of the PksA PT domain
 - (c) A unique dimer interface
 - (d) Three unusual insertion regions
 - (e) Structural analysis of the palmitate and HC8-bound PT structures
 - (f) PT has evolved a distinct active site relative to other DHD enzymes
 - (g) Docking simulations of linear, monocyclic and bicyclic natural substrates
 - (h) Structure-based mutations support the proposed mechanism
 - (i) Details of the global architecture study
 - (j) The molecular basis of cyclization specificity in F-fold promoting PT domains.
- 2) Supplementary Methods
- 3) Supplemental Tables
 - Table S1. Crystallographic statistics of PT crystals
 - Table S2. Domain association studies
 - Table S3. Oligonucleotide sequences for cloning
- 4) Supplementary Figures
 - Figure S1. Sequence alignment comparing PT domains that catalyze C4-C9 and C2-C7 cyclization
 - Figure S2. Topology representation of PT compared to other DHD, SHD, and helix grip proteins
 - Figure S3. *In silico* docking images representing the progression of linear, monocyclic, and bicyclic intermediates in the PT pocket
 - Figure S4. HPLC quantification of norpyrone **4** from PT site-directed mutagenesis experiments
 - Figure S5. Examples from three classes of fungal NR-PKS products with three different first ring cyclization patterns (C4-C9, C2-C7, and C6-C11).
 - Figure S6. Comparison of pocket shapes between ARO/CYC and PT domains.
- 5) Supplementary Notes

1. Supplementary Discussion

(a) Crystallization and structure determination.

Analysis of the rod-like crystals arising from a dissected PT-ACP construct by SDS-PAGE revealed that the didomain had been cleaved in its inter-domainal linker region to yield the two free domains, from which the PT domain selectively crystallized. The PT monodomain was then produced for crystallographic studies. The PT domain crystallizes as a dimer in two crystal forms, with the space groups $P2_12_12_1$ and $P4_12_12$ (Table S1). Initial phases were determined by multi-wavelength anomalous dispersion (MAD) using selenomethionine-derivatized PT followed by phase extension with native data to 1.8 Å. The crystal structures from both crystal forms showed a distinct electron density for the C₁₆ fatty acid, palmitate, which presumably bound to the enzyme during overproduction and purification from *E. coli*. The palmitate was difficult to remove by chromatography or chemical exchange, suggesting tight binding to the hydrophobic active site. Eventually, the palmitate was removed using the lipid-binding resin, XAD-2 (Amberlite), followed by immediate incubation with the product analog HC8 (**5**, Fig. 1) with PT (10:1 ratio). Without this chemical stabilization, however, the XAD-2 treated protein would rapidly denature and precipitate from solution. Orange crystals, indicative of substrate binding, grew in the same condition and space group as the palmitate-bound structure. The HC8-bound PT structure was solved by molecular replacement using the palmitate-bound structure. The dissected 333-residue PT monodomain (39 kDa) was isolated as a homodimer (~73 kDa) as estimated by gel filtration chromatography, and the protein also exists as a dimer in the crystal structure (Fig. 2a).

(b) Detailed structural description of the PksA PT domain

The PT crystal structure shows a “double hot dog” (DHD) fold (Fig. 2a, Fig. S2)⁸ that is also observed in the dehydrase (DH) domains of animal FASs and bacterial modular PKSs (Fig. 2b, Fig. S2)^{9,10}, the eukaryotic MFE-2 hydratase, and a subclass of CoA-hydrolysing thioesterases (such as Rv0216)^{8,11}. The canonical “hot dog” fold has a central helix (the “hot dog”) that is packed in the middle of a curved 5- to 7-stranded β -sheet (the “bun”). Bacterial “single hot dog” (SHD) enzymes dimerize to form a DHD across a 10-14 stranded β -sheet, with two identical active sites forming at the dimer interface^{12,31}. In comparison, the “double hot dog” is a pseudodimeric protein that contains two SHD repeats fused together with only a single active site per monomer residing at the pseudodimer interface.

The mammalian didomain dimeric DHs have two SHD sub-domains that employ helix-helix contacts in addition to the extensive interactions *via* the 14-strand β sheet (7 strands from each sub-domain) (Fig. S2) to stabilize the dimer⁹. Similar to the mammalian DH, PT also has two SHD sub-domains (Fig. 2a, Fig. S2): in sub-domain A, the core β sheet follows the β -strand order **1-3-4-5-2** with the long helix α A inserted between strands β 1 and β 2. Similarly, in sub-domain B, the core β sheet follows the β -strand order **6-8-9-10-7** with the long helix α B inserted between strands β 6 and β 7. Additionally, there are insertion regions between β 1 and α A (α xA, β x2 and β x3), as well as between sub-domains A and B (α xC, β x4, α xD, β x5 and β x6) (Fig. 2a, Fig. S2) that deviate from the typical hot dog fold. Strands β 1 to β 10 from both sub-domains, together with the *N*-terminal β x1, β x4 and β x6, form a 13-strand β sheet (Fig. S2) where sub-domains A and B are spatially related by a two-fold rotation axis perpendicular to the center of the β sheet (Fig. 2a, S2). These sub-domains interact *via* anti-parallel β - β interactions between the backbones of β 2 from sub-domain A and β 7 from sub-domain B (Fig. S2). Taken together, these structural characteristics define the unique conserved DHD fold of fungal PT domains that shape the highly conserved PT active site core (β 1 to β 10, α A and α B).

(c) A unique dimer interface

The dimer interface in PT is unique among DHD proteins. DHD proteins not only have two nearly identical SHD folds, but most of them form a dimer (four SHD folds). Whereas the dimer interface of other DHD proteins consists mainly of the core β -sheet (such as the “back-to-back” interface in the *E. coli* TE II³² or the “side-to-side” interface in porcine FAS DH⁹), the dimer interface of PT is formed by a PT-specific insertion region through interactions between α xC- α xC, β x4- β x4, and α xE- α xE, as well as the region between α B and β 7 (Fig. 2, Fig. S1, orange circles, and Fig. S2). Further, because the insertions β x2, β x3

and $\beta x5$ form a three-stranded β sheet in each monomer, the dimer interface is reinforced by the creation of a six-stranded anti-parallel β sheet (three strands per monomer). In addition, electrostatic interactions between αxE , $\beta x4$ and αB further stabilize the dimer and include R1500-E'1562, K1493-N'1554(C=O), K1489-N'1558, H1478-G'1481(C=O), N1487-E'1515 (through a bridging water molecule), T1490(NH)-D'1557. This unique dimer interface includes several highly conserved residues, such as G1481, K1493, N1554 and D1557. G1481 is necessary due to its small size, while K1493-D1557 reinforces the dimer interface with two pairs (from both monomers) of electrostatic interactions. These interactions lead to a dimeric molecule with dimensions 116 x 53 x 32 Å. The dimer interface residues are also highly conserved among PT domains of fungal NR-PKSs (Fig. S1).

(d) Three unusual insertion regions

In distantly related DHD proteins the two hot dog sub-domains typically have a nearly identical structural arrangement, in which the two central helices can be overlapped completely when superimposed. The two sub-domains in PT, however, are markedly different where the central helix αA is relatively linear compared to the kinked helix αB . The kink in αB is the result of a highly conserved G1550 positioned in the middle of the helix causing a significant bend that molds part of the substrate-binding pocket (Fig. 2a).

The PT structure harbors two additional distinct insertions that are important for structural and functional integrity (Fig. 2a, S2). The first insertion, between $\beta 1$ and αA , includes αxA , $\beta x2$ and $\beta x3$, and forms the lower part of the substrate channel important for cyclization specificity. The proposed catalytic base, His1345, is located on $\beta x3$ and is highly conserved. The second insertion, between sub-domains A and B, includes αxC , $\beta x4$, αxD , $\beta x5$, and $\beta x6$, which defines a binding-pocket to accommodate the hexyl side chain of the incoming substrate (the “hexyl-binding region” discussed in the main text), in addition to the PT dimer interface as discussed above. The third unique element is a three-turn helix at the C-terminus of PT corresponding to a portion of the linker domain that connects PT to the mobile ACP in the full length PksA. It has been shown for modular Type I PKSs that the inter-domain linker regions contain structural elements that are important for protein-protein interactions and substrate specificities³³. Because this flexible structural element is close to the PT substrate binding pocket, this linker region is likely important for interactions necessary for catalytic fidelity in fungal multidomainal IPKSs such as PksA.

(e) Structural analyses of the palmitate and HC8-bound PT structures

The palmitate-bound PT structure exemplifies the tight aliphatic binding capabilities with the tail of the 16-carbon fatty acid positioned directly in the hexyl-binding region (Fig. 3b). In the HC8-bound structure, six ordered waters replace the absent side chain in the hexyl-binding region (Fig. 3c), although the propionyl group of HC8 is pointing towards the hexyl-binding region. The slightly off-positioned HC8 in the structure suggests that the hexyl group is important for proper binding and orientation of the natural substrate in PksA. Residues in $\beta x4$, αxE , and loop αxE - $\beta x5$, important for binding the substrate 6 hexyl group, are only conserved in PTs that bind hexyl-primed substrates, leading to the production of norsolorinic acid (Fig. S1). In sum, the hexyl-binding and PPT-binding regions anchor two ends of the natural substrate, and regiospecific cyclization/aromatization chemistry occurs in the internal cyclization chamber.

Both the palmitate- and HC8-bound PT structures identify a deep pocket that extends 30 Å from the surface to the bottom (Fig. 3a). The co-crystal structures contain well-defined electron density for palmitate (PLM, Fig. 3b) or HC8 (5, Fig. 3c), which identify key pocket residues capable of binding the poly- β -keto substrate and holding it through cyclization/aromatization. The two crystal structures are highly similar with an RMSD of 0.38 Å, indicating that little conformational change occurs in the rigid PT binding pocket in the presence of either linear or bicyclic compounds. The only large difference is Asn1568, which has rotated into the active site in the PLM structure to hydrogen-bond with the carboxylate of PLM, whereas in the HC8-bound structure, Asn1568 is involved in a hydrogen-bonding network to water. The palmitate- and HC8-bound PT structures now unambiguously demonstrate that PT can indeed bind a linear polyketide substrate and its cyclization products.

(f) PT has evolved a distinct active site relative to other DHD enzymes

The SHD and DHD DH and hydratase enzymes all contain a catalytic histidine residue. This residue is located on a tight turn just before the start of the “hot dog” helix. In all structurally known “hot dog” DH and hydratase enzymes, except for the PksA PT, the histidine is polarized by hydrogen bonding to the backbone carbonyl oxygen of a residue located downstream on the same loop⁸. In PksA PT, however, the

polarization of the catalytic histidine (His1345) derives from an alternative source, Asp1543, and it is N δ , instead of N ϵ , that must serve as the basic nitrogen (Fig. 4, Fig. S3).

The known “hot dog” structures that perform hydrations and dehydrations all contain a conserved acidic residue (Asp or Glu) in the active site, and Asp1543 is the corresponding residue in PksA PT. Unlike Asp1543 in PT, such an acidic residue is not found within hydrogen bonding distance of the catalytic histidine in any of these other “hot dog” structures. An aspartate was first suggested to function as a catalytic acid for the fatty acid DH reactions³¹, but this has been questioned by subsequent work on hydratase enzymes¹¹. The latter proposal, based on the structure of the *C. tropicalis* MFE-2 hydratase 2 domain, suggests that the carboxylate acts rather to position a water molecule, which may serve as one half of the oxyanion hole that stabilizes the enolate formed either in the first step of the E1cb dehydration (in the DHs), or from the conjugate attack of water (in the hydratases). The proposed mechanism invokes the above oxyanion hole and implicates the catalytic histidine as both an acid and a base for a concerted *syn* addition (for the hydratases) or elimination (for the DHs).

The other half of the oxyanion hole can be the backbone N–H of a glycyl residue at the tip of the “hot dog” helix, whose positive dipole confers additional stabilization of the oxyanion. The acyl-CoA TEs, for example, are in a subfamily of the “hot dog” proteins but do not have histidine as a catalytic residue. The location of the catalytic carboxylate residue in these TEs is not conserved, yet the glycine at the tip of the major helix remains conserved. Consequently, the one common feature among all previously existing “hot dog” structures appears to be the presence of this residue serving as one part of the oxyanion hole. In stark contrast, the tip of the “hot dog” helix in PksA PT is not this conserved glycine but rather is replaced by a proline (Pro1355) — the only residue incapable of providing an amide hydrogen bond. Thus, the oxyanion hole in TEs with a DHD fold cannot exist in the same location in PksA PT. In summary, the alternate arrangement of His1345 in the active site, the alternative function of Asp1543 in polarizing His1345, and the replacement of the canonical glycine with Pro1355 suggest that PksA PT has evolved substantially from that of a simple DH or hydratase.

(g) Docking simulations of linear, monocyclic and bicyclic natural substrates

HC8 (**5**), lacking the constraints of hexyl and pantetheinyl binding, does not precisely mimic the natural bicyclic PT product (**7**, Fig. 1). Instead, it has an acetyl moiety in a sterically constrained position (assuming this mimics the PPT attachment site), while the first ring methoxy group clashes with the wall of the cyclization chamber. As a result, it is more likely that HC8 binds in an alternate conformation that is skewed slightly from the native substrate-binding orientation (Fig. 3c). Nevertheless, HC8 is useful to visualize the bicyclic binding motif, especially since the natural polyketide substrate **6**, intermediates **8–10** (aldol additions/dehydrations), and product **7** (Fig. 1c) are too unstable to isolate for co-crystallization studies. Furthermore, steric constraints in the binding pocket thwarted co-crystallization attempts with the tricyclic norpyrone (**4**) or norsolorinic acid (**2**). To assess the native substrate-binding motif, we conducted *in silico* substrate docking between PT and the linear (**6**), monocyclic **8**, aromatized monocyclic **9**, bicyclic **10** and aromatized bicyclic **7**. Intermediates were tethered to a PPT arm. Consistent results were obtained by manual docking followed by CHARMM minimization or computer-simulated docking using the program GOLD. We found that: (1) the hexyl and PPT groups of the linear intermediate **6** are peripherally clamped in the hexyl-binding and PPT-binding regions, respectively (Fig. S3a). In particular, Asn1554 is within hydrogen bonding distance to the carbonyl of C11. (2) Docking of the C4–C9 mono-cyclized intermediate **8** orients the C9 oxyanion towards a pair of crystallographic water molecules in a network with residues Asp1543, Asn1568, and Thr1546, implicating the water network as an altered location of the oxyanion hole for the C–C bond formation in PksA PT (Fig. S3b). (3) Docking of the C4–C9 first-ring cyclized and aromatized intermediate **9** shows that C2 (required for the C2–C11 cyclization event) is positioned towards the active site His1345. Additionally, the hydrogen bond from N1554 has translocated from the C11 to the C13 carbonyl, pulling the substrate chain out of the hexyl-binding region by a two-carbon unit (Fig. S3c). (4) Docking of the aromatized bicyclic intermediate **7** shows that the cyclization chamber can accommodate two aromatic rings (Fig. S3d). (5) Third- or fourth-ring cyclized polyketides cannot be docked in the PT cyclization chamber, owing to prohibitive steric clashes. Therefore, the substrate-PT docking simulation results are consistent with previous *in vitro* reconstitution results⁴, further supporting that PksA PT catalyzes C4–C9 first-ring and C2–C11 second-ring cyclizations and aromatizations.

(h) Structure-based mutations support the proposed mechanism

The new structural insights guided mutagenesis studies to probe the proposed mechanism for PT domains. Seven independent PT monodomain mutants (*) including His1345Ala, Asp1543Ala, Gln1547Ala, Asn1554Ala, Thr1546Ala, Asn1568Ala, and Gly1491Leu were analyzed in 3-part multidomain combination experiments (SAT-KS-MAT + PT* + ACP) to monitor their involvement in catalytic flux to produce the ACP-bound bicyclic product, which was then detected as its spontaneous cyclization product, norpyrone **4** (Fig. S4). We found that His1345Ala, Asp1543Ala, Gln1547Ala, Asn1554Ala and Gly1491Leu resulted in no detectable activity above spontaneous poly- β -keto cyclization observed in control experiments. Thr1546Ala only retained partial activity, and, surprisingly, Asn1568Ala showed slightly elevated activity.

The rationale for the selection of these mutations and their catalytic outcome is as follows: (1) His1345Ala destroyed activity, as this His is the key active site base; (2) Asp1543Ala abolished activity, because Asp1543 is involved in a critical hydrogen-bonding network to polarize His1345 (catalytic diad); (3) Gln1547Ala drastically diminished activity being involved in a hydrogen bonding network with Asp1543 and Tyr1492, and is thought to orient Asp1543 for interaction with His1345 (corresponding glutamine residues have been implicated for a similar function in known DH and hydratase structures and shown to be crucial to activity in those enzymes by mutagenesis experiments^{34,35}); (4) Asn1554 is proposed to associate with the ketone oxygen atoms to position and “kink” the intermediates for proper ring cyclizations; (5) Gly1491Leu represents a bulky mutation to close the hexyl binding pocket resulting in no apparent activity over the controls, as expected; and (6/7) Thr1546 and Asn1568 are implicated in the water network. (6) Thr1546Ala only retained partial activity (~40%), supporting the importance of its side chain hydroxyl directly bound to one of two ordered water molecules that stabilize the oxyanion formed in the aldol cyclization (Fig.4); (7) Asn1568Ala resulted in elevated activity (~120%). In the PT-HC8 structure the Asn1568 side chain amide is hydrogen-bonded to an ordered water molecule that simulations show is the second to stabilize the oxyanion in the cyclization reaction(s) (Fig. 4). On the other hand, in the PT-PLM structure the Asn side chain has moved to hydrogen-bond to the carboxylate of the bound fatty acid, yet the water molecule remains ordered owing to other favorable interactions in the active site. Thus the Asn amide-water hydrogen bond does not govern its placement in the active site, and, as it is the amide backbone of this residue that is suggested to assist stabilization of the substrate anion for the aldol cyclization, this latter interaction will be unaffected by mutation to Ala. Alanine, however, is a less sterically demanding residue and can be visualized to interfere less with proximal substrate movement and ring-forming events in the active site.

(i) Details of the global architecture study

To gain some insight into the role of the PT domain in global domain organization of Type I, NR-PKSs, a series of truncations within the linker region (~100 aa) between the two halves of PksA (MAT-PT linker) was evaluated for the ability to form higher order complexes. In N-terminal half constructs of PksA, only the SAT-KS-MAT tridomain migrated as a dimer, which is consistent with KS dimerization among all FAS and PKS enzymes. For C-terminal half constructs of PksA, the ACP⁷ and TE/CLC monodomains and the ACP-TE/CLC didomain migrated as apparent monomers (Table S2) while the PT monodomain and PT-ACP didomain, migrated as dimers suggesting that the PT domain is responsible for the observed dimerization, and is consistent with the crystal structure. Unexpectedly, dynamic light scattering showed ACP containing substantial flanking linker sequences exists as a dimer, presumably through linker-linker association that dissociated under the native PAGE conditions. The PT-ACP-TE/CLC tridomain migrated as both monomeric and dimeric forms with association/dissociation likely influenced by the mobile ACP and TE domains. In sum, the results support a structural organization of PksA with similarity to the animal FAS (Fig. 2b).

Phylogenetic results¹⁸ carried out on the broad family of fungal PKSs suggests that the SAT domain in fungal NR-PKSs was acquired by *N*-terminal fusion³⁶, while the reductive domains were lost from its reducing PKS ancestors. The recent porcine FAS crystal structure displays a bipedal head-to-head arrangement facilitated by the dimeric KS central core with AT acting as the “feet”⁹. By analogy, fusion of the SAT domain adds two extra “feet” in PksA and likely results in a quadriped-like structure as opposed to the bipedal animal FAS. On the other hand, the structural analysis here coupled with previous phylogenetics results¹⁸ support the view that the dimeric PT domain evolved from ancient DH domains in reducing PKSs and would be positioned at the DH locus in animal FAS. The location of the inter-SHD linker region is also similar between PT and DH from animal FAS and DEBS module 4 DH domain (Fig. 2b, Fig. S2)^{9,10},

although the PT linker has a higher percentage of helical content. The PT C-terminal region, the linker region between PT and ACP, is discussed in section (d), and the role of this PT-ACP linker region may also be similar to the inter-domain linkers in animal FAS and modular PKS, despite having an alternative dimerization motif. Furthermore, the animal FAS harbors a pseudomethyltransferase domain, a structural relic maintained from fungal reducing PKS ancestors⁹, further supporting such close structural relationships among fungal PKSs, bacterial modular PKS, and animal FASs (Fig. 2b)

(j) The molecular basis of cyclization specificity in F-fold promoting PT domains.

The PT domains in fungal PKSs can be divided into three classes based on their first-ring cyclization specificity: the C4-C9 PT (as PksA in aflatoxin biosynthesis⁴), the C2-C7 PT (such as WAS³⁷ and PKS4^{19,38}), and C6-C11 PT³⁹ (Fig. S5). We were able to obtain the protein sequence of C4-C9 and C2-C7 PTs, but not C6-C11. Based on the sequence alignment and structural analysis, the protein fold, di-domain arrangement, dimer interface, ACP docking site and catalytic mechanism for PT domains in fungal PKSs are conserved. As expected, the largest variation lies within the cyclization chamber, where residues are conserved only within C4-C9 and C2-C7 PTs. These differences provide excellent opportunities to predict the cyclization specificities observed in the two PT domain classes. Conservation of active site residues in C2-C7 versus C4-C9 cyclizing PTs (Fig. S1) will provide an additional characteristic to predicting cyclization specificities in the many “cryptic” nonreducing IPKSs emerging from genome sequencing efforts.

To probe the difference between C4-C9 and C2-C7 PTs, we generated a homology model of the PKS4 PT from *Gibberella fujikuroi*⁴⁰. PKS4 has the same domain architecture as PksA; however, its product bikaverin contains a C2-C7 cyclized ring. The homology model of PKS4 PT was generated by Swiss Model in batch mode, using the palmitate-bound PksA PT as the template. The sequence of PKS4 PT was then mapped onto the PksA PT structure based on sequence alignment generated by the program ClustalW2⁴¹, followed by energy minimization in Swiss Model. Using the program GOLD, we then added the natural substrates of the PKS4 PT (C2-C7 first-ring and C1-C10 second-ring cyclized intermediates) into the minimized PKS4 PT substrate pocket.

Significantly, the PKS4 PT homology model has a bent substrate pocket (Fig. S6c), compared to the long, straight PksA PT pocket (Fig. S6b). This dramatic change is originated by the following change of residues in the cyclization chamber: Q1547 (a catalytic residue), A1499 and M1495 in PksA become H, V and F in PKS4. Due to the presence of these bulkier residues, access to the hexyl-binding region is blocked (Fig. S6c). These three residues are highly conserved in C2-C7 PTs (Fig. S1). Further, G1491 at the bottom of the hexyl finger becomes a bulky residue (L, I or V) in C2-C7 PTs, further collapsing the hexyl finger. Lack of the hexyl finger region in C2-C7 PTs is not surprising, because their corresponding substrate(s) do not have a hexyl group. A second striking feature of the modeled PKS4 PT pocket is the enlargement of the cyclization chamber. In PksA PT, this is blocked by conserved T1546, F1574, F1617, L1620. These residues are changed to G, L, C, V, respectively in Pks4, and are also conserved among C2-C7 PTs. As a result, the predicted pocket in C2-C7 cyclizing PTs is bent in a fashion that could be more accommodating for C2-C7 first ring cyclization (Fig. S6c), instead of the C4-C9 cyclization promoted by the straight pocket in PksA PT. Additionally, at the back of the cyclization chamber, T1354 in PksA is changed to Pro in C2-C7 PTs. Proline steric and structural effects may block the ring storage region, disabling first ring translocation in the C2-C7 PTs. Consistent with this prediction, domain reconstitution results of PKS4 supports that the PKS4 PT controls the C2-C7 first ring cyclization, but not the second or third ring cyclizations^{19,38}. The above analyses suggest that it could be possible to change the pocket size and shape (and therefore cyclization specificity) by mutations of PT pocket residues.

2. Supplementary Methods

Substrate and protein docking simulations. Intermediates **8-11** (Figure 1C) with PPT arm attached were generated using PRODRG and hydrogens added to PT using Chimera⁴². Docking between PksA PT and intermediates were performed using GOLD²¹. The A chain from the PksA PT structure bound with HC8 or palmitate was used, with all waters and HC8/palmitate removed. The substrate-binding pocket of PT was determined by the automatic cavity-search program in GOLD and is consistent with the pocket from visual search. The docked substrate was constrained so that C4 and C9 (where first ring cyclization occurs) were within 3.5 Å of the proposed catalytic base H1345. Different ligand conformations were searched using the GOLD default settings. Each compound was docked 100 times to ensure consistent docking simulation.

CHARMm minimizations. CHARMm minimizations were performed with the Discovery Studio (DS) 2.1 suite. Intermediate **6**-PPT was generated with the structure drawing functionality of DS 2.1 and manually docked in the active site of PksA PT domain (HC8) crystal structure. The terminal 11 carbons of the intermediate were manually superimposed over the locations of the terminal 11 carbons of PLM in the "hexyl-binding region". All waters were removed, excepting the three active site waters common to both the HC8 and PLM structures. The entire system was typed with the CHARMm forcefield and then minimized using an adopted basis NR (ABNR) algorithm with a generalized Born implicit solvent model with molecular volume (GBMV) and constraining hydrogen-bonding partners from the crystal structure to ± 0.25 Å of their measured values. On the ligand, C4 was constrained between 2.5 and 3.5 Å of the basic nitrogen of H1345. C4 was also constrained to within 5 Å of C9. The algorithm exited after achieving an RMS gradient of less than 0.1 kcal/(mol Å). All other settings were default. The first-ring-cyclized intermediate was drawn from the results of the minimization of **6**-PPT and was minimized first by steepest descent, then by ABNR with GBMV as above but with no ligand constraints. The first-ring-cyclized aromatic intermediate was drawn from the results of the minimization of the first-ring-cyclized intermediate and was minimized first by steepest descent, then by ABNR with GBMV as above, with ligand constraints of C2 within 3.5 Å of His1345 and 5 Å of C11. The second-ring-cyclized intermediate (**7**) was drawn from the results of the minimization of the first-ring-cyclized aromatic intermediate and was minimized first by steepest descent, then by ABNR with GBMV as above but with no ligand constraints.

PksA deconstruction. A predicted linker region (~100 aa) exists between the two halves of PksA (MAT-PT linker). The majority of this linker sequence between the two halves of PksA (SAT-KS-MAT and PT-ACP-TE/CLC) was removed for successful expression and purification of downstream PT containing constructs. Three PT monodomain constructs (PT #1, 2⁴, and 3) were generated with varying lengths of the C-terminal linker sequence including a C-terminal His₆-tag fusion (Table S3). All constructs showed comparable activity in 3-part multidomain combinations (The PT monodomain was amplified with *pt5.4* and *pt3.1*, *pt3.2*, or *pt3.3*, and the products were inserted into the NdeI/NotI sites in pET28a, generating the C-terminal His₆-tagged constructs, pEPT1, pEPT2, and pEPT3, respectively). In order to test how much linker was necessary for the SAT-KS-MAT tridomain, five expression constructs were generated (NKA #1-3, 4⁴, and 5) with varying lengths of the C-terminal linker incorporated (Table S3). All five constructs gave comparable total expression levels as determined by SDS-PAGE (the SAT-KS-MAT tridomain was amplified using *nka5* and *nka-c3.1*, *nka-c3.2*, *nka-c3.3*, *nka-c3.4*, or *nka-c3.5* and inserted into the NcoI/NotI sites in pET28a, generating five C-terminal His₆-tagged constructs pENKA1, pENKA2, pENKA3, pENKA4, and pENKA5, respectively). The first truncation site, NKA #1, resulted in insoluble protein, and NKA #2 yielded minutely soluble protein under the expression conditions tested (protein production and purification was carried out as previously described⁴). However, the last three truncation sites, NKA #3-5, resulted in highly stable, soluble proteins.

Table S1 Data collection, phasing and refinement statistics for **PT** structures

	Native PT-plm	Native PT-HC8	SeMet PT		
Data collection					
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P4 ₁ 2 ₁ 2		
Cell dimensions					
<i>a</i> , <i>b</i> , <i>c</i> (Å)	88.79, 91.25, 91.23	88.81, 90.61, 90.66	73.73, 73.73, 321.21		
<i>α</i> , <i>β</i> , <i>γ</i> (°)	<i>α</i> = <i>β</i> = <i>γ</i> =90	<i>α</i> = <i>β</i> = <i>γ</i> =90	<i>α</i> = <i>β</i> = <i>γ</i> =90		
			<i>Peak</i>	<i>Inflection</i>	<i>Remote</i>
Wavelength	1.0	1.0	0.98000	0.98025	0.96482
Resolution (Å)	1.8	1.9	3.2	3.3	3.3
<i>R</i> _{sym} or <i>R</i> _{merge}	4.9 (40.8)	6.0 (28.9)	18.1(50.3)	15.3 (44.1)	15.1(42.0)
<i>I</i> / <i>σI</i>	43.7 (5.6)	30.6 (4.2)	24.8 (8.8)	25.8 (10.8)	25.2 (10.7)
Completeness (%)	100.0 (100.0)	99.8 (89.7)	99.8 (100.0)	99.7 (100.0)	99.8 (100.0)
Redundancy	7.4 (7.4)	7.0 (5.1)	26.7 (26.7)	26.9 (28.5)	26.9 (28.5)
Refinement					
Resolution (Å)	1.8	1.9			
No. reflections	67206	54675			
<i>R</i> _{work} / <i>R</i> _{free}	20.2/22.2	19.0/22.3			
No. atoms					
Protein	5000	4959			
Ligand/ion	36	46			
Water	491	391			
B-factors					
Protein	26.3	24.7			
Ligand/ion	29.8	34.3			
Water	36.4	30.1			
R.m.s deviations					
Bond lengths (Å)	0.005	0.016			
Bond angles (°)	1.34	1.52			

Table S2. Domain association studies

Construct	Association
SAT*	Monomer
SAT-KS-MAT	Dimer
PT	Dimer
ACP*	Monomer
TE	Monomer
PT-ACP	Dimer
ACP-TE	Monomer
PT-ACP-TE	Monomer/dimer

*Association determined previously⁷.

Table S3. Oligonucleotide sequences for cloning

Primer	Sequence
<i>PT monodomain</i>	
PT5.4	5'-CGCATATGGAAATCAAGACCACCACGAC-3'
PT3.1	5'-GTAAGCGGCCGCGAGGCGATATTGGCATTGTCGTTG-3'
PT3.2	5'-GTAAGCGGCCGCAATAGCTTGGCTGGAAACCTGAGTATT-3'
PT3.3	5'-GTAAGCGGCCGCGATCGGCCTTTTCACTGTAGTCTGC-3'
<i>SAT-KS-MAT tridomain</i>	
NKA5	5'-CATATCCATGGCTCAATCACGTCAACTCTTTCTC-3'
NKA-C3.1	5'-CAAGCGGCCGCGGAGGAGGCGAAAGGGGC-3'
NKA-C3.2	5'-CAAGCGGCCGCGCTCGGCAGATCCACCACC-3'
NKA-C3.3	5'-CAAGCGGCCGCGATCCTGCTGGTGGCGATGCAG-3'
NKA-C3.4	5'-CAAGCGGCCGCGGATGGACGGTGAGGCGTC-3'
NKA-C3.5	5'-CAAGCGGCCGCTTCGGGGAAGGCCTCCTTGC-3'

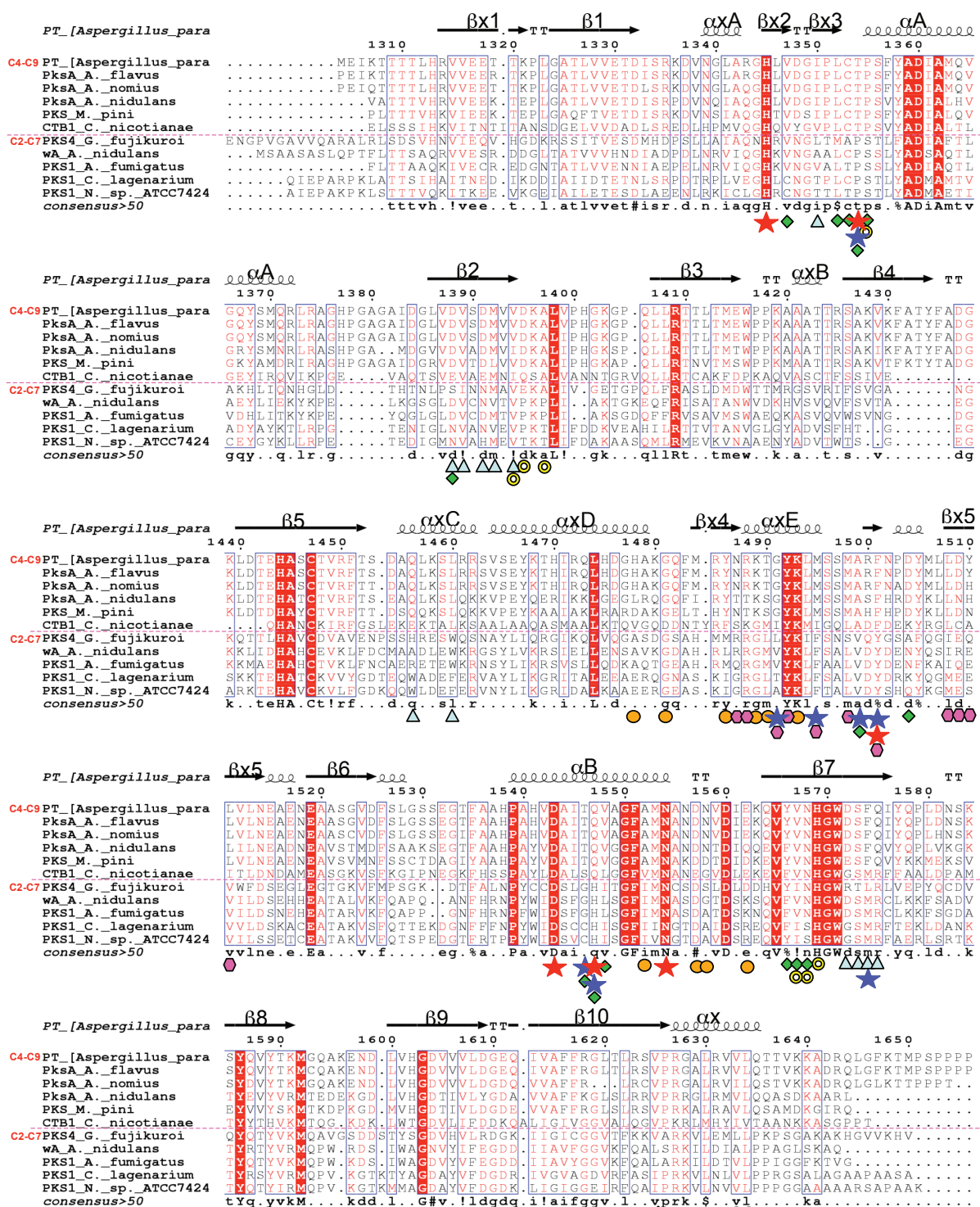


Figure S1: Sequence alignment comparing PT domains that catalyze C4-C9 and C2-C7 cyclization. Red star, C4-C9 active site residues; Blue star, C2-C7 active site residues; Blue triangle, DHD subdomain stabilization; Orange circle, dimer interface; Purple hexagon, conserved residues in C4-C9 hexyl binding PTs; Green diamond, cyclization chamber residues; Yellow open circle, PPT binding region.

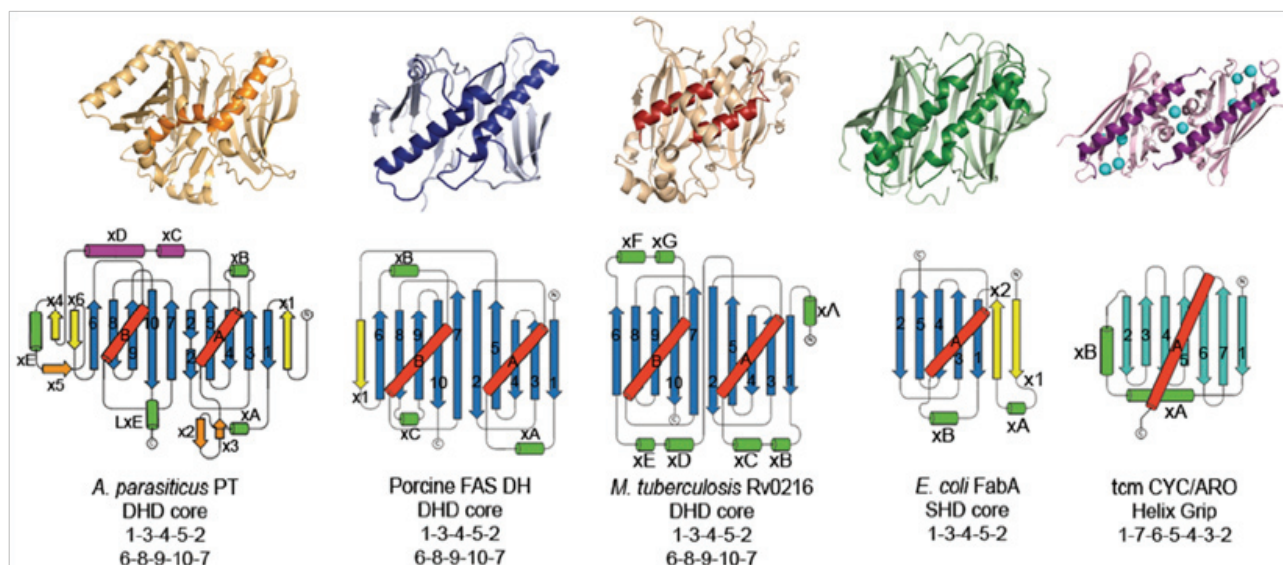


Figure S2: Topology representation of PT compared to other DHD, SHD, and helix grip proteins. The top panel highlights the location of the central helices with respect to the surrounding β sheets that characterize this fold. The bottom panel shows the core domain of the DHD (or SHD) fold in blue and red. Auxiliary sheets (yellow or orange) and helices (green or purple) highlight the differences of PT compared to the canonical DHD fold. The tcm ARO/CYC dimer is not a SHD or DHD enzyme but adopts a structurally similar helix-grip fold and is shown as teal sheets. This structure is shown here as it has been proposed to carryout a similar catalytic reaction to the PT domain, exemplifying the catalytic relationships between fungal and bacterial aromatic polyketide biosynthesis.

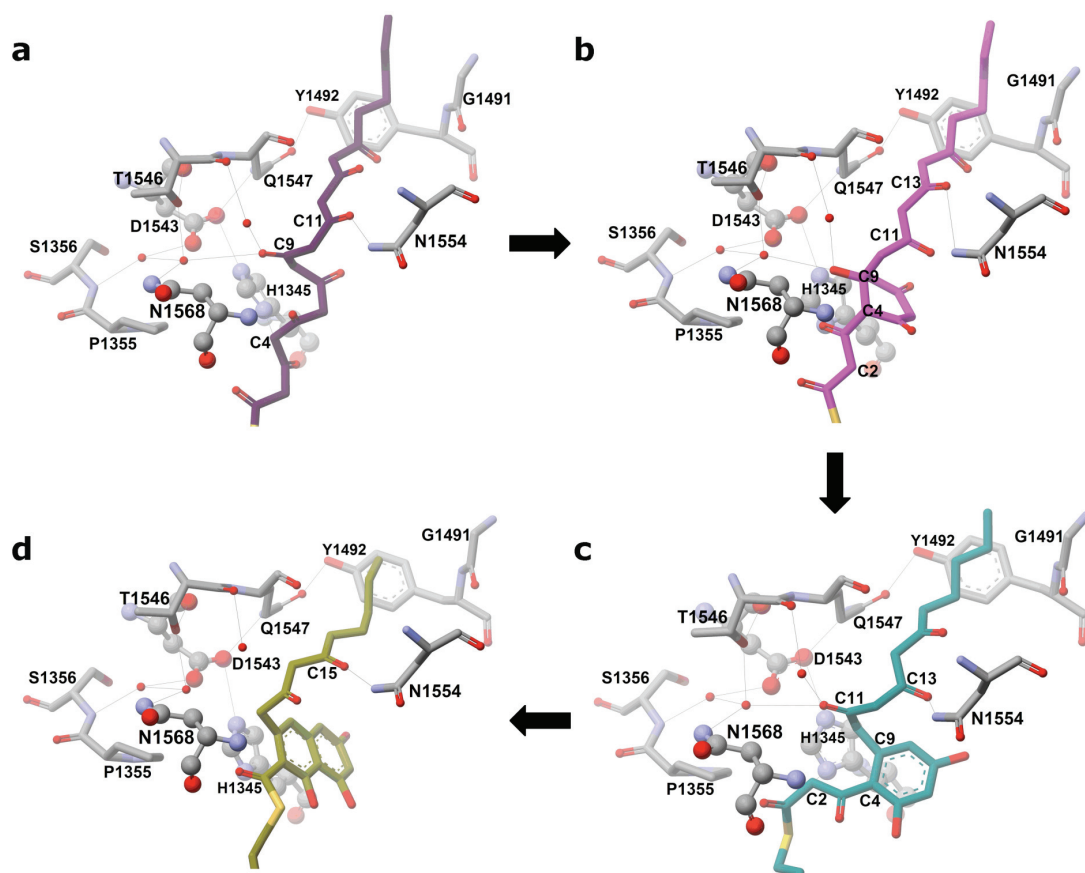


Figure S3: *In silico* docking images representing the progression of linear, monocyclic, and bicyclic intermediates in the PT pocket. **(a)** Docking of the linear intermediate **6** (purple sticks) with C4 in proximity of the active site H1345. Atoms within hydrogen bonding distance and interacting atoms are highlighted by grey lines. **(b)** Docking of hypothetical intermediate **8** (pink sticks). **(c)** Same as (b) but with presumed intermediate **9** **(d)** Docking of the intermediate **10** that has undergone second ring formation between C2 and C11.

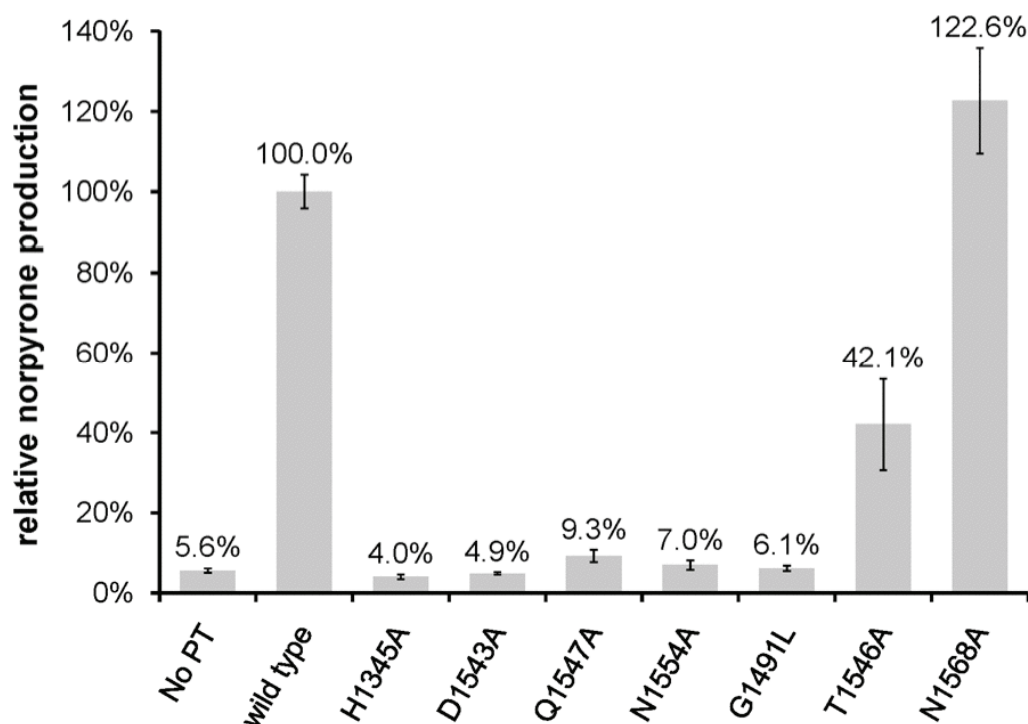


Figure S4. HPLC quantification of norpyrone **4** from PT site-directed mutagenesis experiments. Norpyrone production is reported normalized to “wild-type” production, which is set to 100% (error = s.d.). Individual mutations are listed on the x-axis.

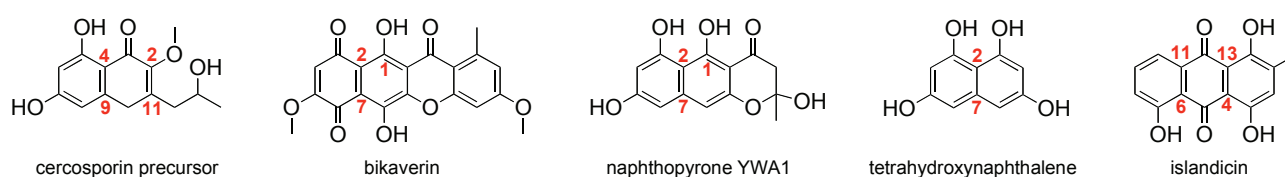


Figure S5. Examples from three classes of fungal NR-PKS products with three different first ring cyclization patterns (C4-C9, C2-C7, and C6-C11).

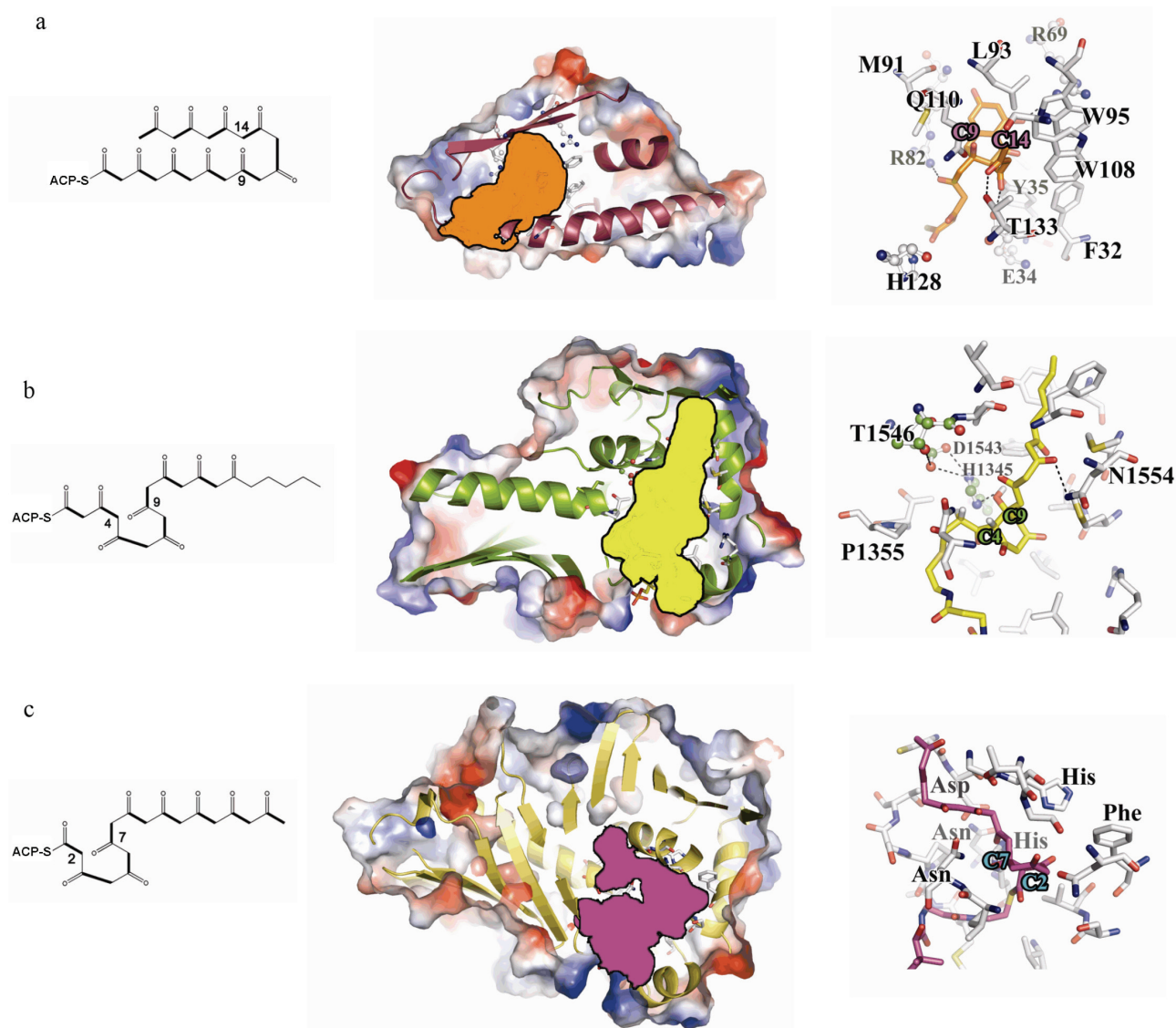


Figure S6: Comparison of pocket shapes between ARO/CYC and PT domains. (a) The tcmARO/CYC acts on a C20 polyketide to form the C9-C14 first ring. The interior pocket (orange) is comparatively wider and shorter, causing the polyketide to fold back on itself. A zoom in of the active site shows the orientation of the C9-C14 ring from a docked intermediate bound in the hydrophobic pocket. (b) The PksA PT acts on a C20 polyketide that is primed with a hexyl group. The shape of the PT pocket (yellow-green) is straight with the hexyl binding region at the end. The bound substrate is extended, and the C4-C9 first ring cyclization is facilitated by a kink in the polyketide chain. A zoom in of the active site shows the docked first ring cyclized intermediate and the orientation of the catalytic triad. (c) The PKS4 PT catalyzes C2-C7 first ring cyclization of a C18 polyketide. A homology model generated using the PksA PT as the template shows that the hexyl-binding region is closed off and the pocket now adopts a curved conformation (purple). This is proposed to be a conserved feature in PT domains that catalyze C2-C7 cyclization (Fig. S1). Similar to (b), the polyketide remains extended while the kink in the polyketide chain promotes C2-C7 cyclization. A zoom in of the active site shows that a His and Phe located above the catalytic diad closes off the hexyl-binding region and causes the interior pocket to bend towards the left.

Supplementary Notes

31. Leesong, M., Henderson, B.S., Gillig, J.R., Schwab, J.M. & Smith, J.L. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253-264 (1996).
32. Swenson, L., Green, R., Smith, S. & Derewenda, Z.S. Crystallization of thioesterase II from *Escherichia coli*. *J Mol Biol* **236**, 660-662 (1994).
33. Tang, Y., Kim, C.Y., Mathews, II, Cane, D.E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A* **103**, 11124-11129 (2006).
34. Kiema, T.R. et al. Mutagenic and enzymological studies of the hydratase and isomerase activities of 2-enoyl-CoA hydratase-1. *Biochemistry* **38**, 2991-2999 (1999).
35. Pasta, S., Witkowski, A., Joshi, A.K. & Smith, S. Catalytic residues are shared between two pseudosubunits of the dehydratase domain of the animal fatty acid synthase. *Chem Biol* **14**, 1377-1385 (2007).
36. Crawford, J.M., Vagstad, A.L., Ehrlich, K.C. & Townsend, C.A. Starter unit specificity directs genome mining of polyketide synthase pathways in fungi. *Bioorg Chem* **36**, 16-22 (2008).
37. Fujii, I., Watanabe, A., Sankawa, U. & Ebizuka, Y. Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chem Biol* **8**, 189-197 (2001).
38. Zhang, W., Li, Y. & Tang, Y. Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc Natl Acad Sci U S A* **105**, 20683-20688 (2008).
39. Paulick, R.C., Casey, M.L., Hillenbrand, D.F. & Whitlock, H.W., Jr. Letter: A ^{13}C nuclear magnetic resonance study of the biosynthesis of islandicin from $^{13}\text{CH}_3^{13}\text{CO}_2\text{Na}$. *J Am Chem Soc* **97**, 5303-5305 (1975).
40. Linnemannstons, P. et al. The polyketide synthase gene *pk4* from *Gibberella fujikuroi* encodes a key enzyme in the biosynthesis of the red pigment bikaverin. *Fungal Genet Biol* **37**, 134-148 (2002).
41. Larkin, M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
42. Pettersen, E.F., et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comp Chem* **25**, 1605-1612 (2004).