Structural and evolutionary relationships of “AT-less” type I polyketide synthase ketosynthases

Jeremy R. Lohman^a, Ming Ma^a, Jerzy Osipiuk^b, Boguslaw Nocek^b, Youngchang Kim^b, Changsoo Chang^b, Marianne Cuff^b, Jamey Mack^b, Lance Bigelow^b, Hui Li^b, Michael Endres^b, Gyorgy Babnigg^b, Andrzej Joachimiak^b, George N. Phillips Jr.^{c,d}, and Ben Shen^{a,e,f,1}

^aDepartment of Chemistry, The Scripps Research Institute, Jupiter, FL 33458; ^bMidwest Center for Structural Genomics and Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, IL 60439; ^cBioSciences at Rice, Rice University, Houston, TX 77251; ^dDepartment of Chemistry, Rice University, Houston, TX 77251; ^eDepartment of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL 33458; and ^fNatural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, FL 33458

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 1, 2015 (received for review August 4, 2015)

Acyltransferase (AT)-less type I polyketide synthases (PKSs) break the type I PKS paradigm. They lack the integrated AT domains within their modules and instead use a discrete AT that acts in trans, whereas a type I PKS module minimally contains AT, acyl carrier protein (ACP), and ketosynthase (KS) domains. Structures of canonical type I PKS KS-AT didomains reveal structured linkers that connect the two domains. AT-less type I PKS KSs have remnants of these linkers, which have been hypothesized to be AT docking domains. Natural products produced by AT-less type I PKSs are very complex because of an increased representation of unique modifying domains. AT-less type I PKS KSs possess substrate specificity and fall into phylogenetic clades that correlate with their substrates, whereas canonical type I PKS KSs are monophyletic. We have solved crystal structures of seven AT-less type I PKS KS domains that represent various sequence clusters, revealing insight into the large structural and subtle amino acid residue differences that lead to unique active site topologies and substrate specificities. One set of structures represents a larger group of KS domains from both canonical and AT-less type I PKSs that accept amino acid-containing substrates. One structure has a partial AT-domain, revealing the structural consequences of a type I PKS KS evolving into an AT-less type I PKS KS. These structures highlight the structural diversity within the AT-less type I PKS KS family, and most important, provide a unique opportunity to study the molecular evolution of substrate specificity within the type I PKSs.

biosynthesis | secondary metabolism | iso-migrastatin | leinamycin | oxazolomycin

Acytransferase (AT)-less type I polyketide synthases (PKSs) violate the paradigm for canonical type I PKS architecture (1, 2). Canonical type I PKSs have an assembly line architecture with modules minimally composed of a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) (Fig. 1). AT-less type I PKSs have modules that lack AT domains; instead, AT is a discrete enzyme that acts in trans to iteratively load malonate units onto the ACPs (3). This activity was first demonstrated with domains from the leinamycin (LNM) biosynthetic pathway, in which the discrete AT is able to load all of the ACPs with malonate (4).

In the report of the LNM gene cluster, bioinformatics analysis of the regions C terminal to the KS domains revealed conserved regions that were hypothesized to be “AT-docking domains” (ATds) (5). Structural studies of canonical type I PKS KS-AT didomains revealed that KSs form a dimeric core structure that is flanked by AT domains connected through structured linkers or “KS/AT adapters” (6). Recent structural studies of AT-less type I PKSs have revealed that the KS domains indeed have a structured ATd or “flanking subdomain” that is equivalent in structure to the “KS/AT adapter” (Fig. 1) (7). Whether the ATd has function or is vestigial, and whether AT-less type I PKSs evolved from canonical type I PKSs or vice versa, remains to be determined.

The products of AT-less type I PKSs are typically more complex than those of the type I PKSs because of the presence of unique

AT-less type I PKS domains, producing diverse substrates for the KS domains. KS domains carry out two half-reactions in the process of polyketide chain elongation (Fig. 1C). The first reaction, trans-thiolation, is transfer of a substrate acyl chain carried by the pantetheine moiety of an ACP onto the KS active site cysteine. The second reaction is a decarboxylative Claisen-condensation of a malonyl-S-ACP on the acyl-S-KS thioester to generate β -ketoacyl-S-ACP. Common polyketide intermediates are ACP-tethered thioesters with β -keto, β -hydroxy, α/β -unsaturated (alkene), and alkane functionalities that are generated by KS, ketoreductase, dehydratase, and enoylreductase domains, respectively (SI Appendix, Fig. S1A). The α -carbon of β -ketoacyl-S-ACP can be alkylated by additional methyltransferase domains. Another common modification to β -ketoacyl-S-ACP is β -alkylation, which is carried out by hydroxymethylglutaryl-CoA synthase homologs, and the resulting β -hydroxyacyl-S-ACP is dehydrated and decarboxylated by enoyl-CoA-hydratase homologs (SI Appendix, Fig. S1B) (8). Two other modifications found within AT-less type I PKSs are pyran/furan formation by pyran synthase domains (9) and migration of α/β alkenes to β/γ alkenes by enoyl-isomerases (SI Appendix, Fig. S1A and C) (10).

AT-less type I PKSs are rich in domains catalyzing the biosynthesis of hybrid peptide–polyketide natural products. Two general strategies are known for hybrid peptide–polyketide natural product biosynthesis: hybrid nonribosomal peptide synthetase (NRPS)/PKS modules containing a KS domain that accepts *N*-acyl amino acid or peptide from a peptidyl carrier protein and performs C–C bond formation with a malonyl-S-ACP, or hybrid

Significance

There are many differences in the sequences of ketosynthase (KS) domains from the well-studied type I polyketide synthases (PKSs) and the more recently discovered acyltransferase (AT)-less type I PKSs. The AT-less type I PKSs generate polyketides with a high degree of structural diversity, which stems from their evolution by horizontal gene transfer. In comparison, canonical type I PKSs evolve by gene duplication. The seven structures of AT-less type I PKS KSs reveal the molecular details surrounding the evolution of substrate specificity and structural diversity, and their overall differences with canonical type I PKS KSs. Understanding the mechanism of substrate specificity will allow reprogramming of the KS active sites to generate polyketide analogues by PKS and polyketide biosynthetic pathway engineering.

Author contributions: J.R.L., G.N.P., and B.S. designed research; J.R.L., M.M., J.O., B.N., Y.K., C.C., M.C., J.M., L.B., H.L., and M.E. performed research; G.B., A.J., and G.N.P. contributed new reagents/analytic tools; J.R.L., M.M., J.O., B.N., Y.K., C.C., M.C., J.M., L.B., H.L., M.E., G.B., G.N.P., and B.S. analyzed data; and J.R.L., G.N.P., and B.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. Email: shenb@scripps.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515460112/-DCSupplemental.

Fig. 1. Canonical and AT-less type I PKS module architecture and activity. The domain labeled "?" represents any number of modification domains. (A) Canonical type I PKS module activation, with the AT domain acting in cis. (B) AT-less type I PKS module activation, with a discrete AT acting in trans. (C) Canonical activities for KS domains catalyzing transthiolation and decarboxylative condensation. Magenta oval represents the ATd or KS/AT adapter. See Fig. 2 for representative KS transthiolation substrates.

PKS/NRPS modules containing a condensation domain that forms an amide bond between acyl-S-ACP and the amine of an aminoacyl-S-peptidyl carrier protein (*SI Appendix, Fig. S2A and D*) (11). We previously revealed that hybrid NRPS/PKS KS domains are phylogenetically distinct from canonical type I PKS KS domains (11); however, their relationships to AT-less type I PKS KSs were not examined.

Some AT-less type I PKS KS domains have activities other than β -keto-acyl-S-ACP synthesis. The rhizoxin AT-less type I PKS contains a β -branching KS with an appended B domain, which produces a β -branch lactone (*SI Appendix, Fig. S1C*) (12, 13). In analogy, the glutarimide ring of iso-migrastatin (MGS) is installed by a β -branching KS performing similar β -branching chemistry (14). Although most KS domains possess a conserved active site cysteine and two histidines, some have mutated histidine residues, making them catalytically incompetent for C–C bond formation, yet capable of transthiolation activity, and are referred to as KS⁰. Recently, one of these KS⁰ domains from the FR901465 biosynthetic pathway was confirmed to have "gatekeeping" activity and would only transfer the product of the upstream module if correctly processed (*SI Appendix, Fig. S2B*) (15).

A feature of canonical type I PKS architecture is colinearity; that is, the PKS amino acid sequence correlates with the linear extension of the polyketide product. Colinearity allows prediction of polyketide structures from PKS sequence and vice versa. Conversely, AT-less type I PKSs frequently deviate from colinearity as a result of the presence of unique domains, non-canonical KS activities, and cryptic modifying domain activities. Phylogenetic analysis of the canonical type I PKS KS domains reveals that they share high homology and mainly clade with members of their biosynthetic pathway (16). In stark contrast, phylogenetic analysis of the AT-less type I PKS KS domains reveals that they form clades that correlate with the functional groups of the α - and β -carbons of the acyl-S-ACP substrate (17). This phylogenetic substrate correlation has been used to predict the products of AT-less type I PKSs, as colinearity analysis often fails (18). These observations have led to the hypothesis that AT-less KS domains have specificity for their substrates. Characterization of the substrate specificity of six AT-less type I PKS KSs from the psymberin and bacillaene pathways demonstrate that the KSs have substrate preference (7, 19–21). Furthermore, a structure of a bacillaene KS, with the natural substrate or a mimic bound, reveals that there are specific interactions between the KS and natural substrate that further strengthen the argument for AT-less type I PKS KS substrate specificity (7).

To obtain a more detailed understanding of how features in amino acid sequences relate to structure–function relationships, we subjected the KS domains from the LNM, oxazolomycin (OZM) (22),

and MGS (23) biosynthetic pathways to high-throughput structural genomics analysis. This effort resulted in four structures from the OZM and three structures from the MGS biosynthetic pathways. Sequence similarity network analysis (SSA) of more than 600 KS domains was used to reveal sets of KSs with similarity to the structurally characterized members. The structures and SSA reveal the molecular details of the evolutionary relationships between the KS domains of canonical, hybrid NRPS/PKS, and AT-less type I PKSs. Structural and sequence alignment analysis gives further insight into the molecular details accounting for AT-less type I PKS KS substrate specificity. The structural details revealed here can be leveraged toward redesign of the KS active sites (by protein engineering) to enable combinatorial biosynthesis of AT-less type I PKS products.

Results and Discussion

Sequence Analysis of AT-Less Type I PKS KSs and High-Throughput Structural Genomics Results.

Alignment of 456 KS domains from 43 AT-less type I PKSs with 165 KS-AT domains from 23 canonical type I PKS pathways (12 from hybrid NRPS/PKS pathways) revealed conservation in the KS cores and weak homology between the KS/AT adapters and ATd; however, some AT-less KS domains lacked clear ATd, and others contain an ACP-like insertion within the ATd (*SI Appendix, Fig. S3 and Tables S1 and S2*). A conserved tryptophan that is inserted into a pocket on the surface of the KS as found in the structures of the canonical type I PKS KS-AT didomains was used to define the C terminus (6). KS domains were produced with 10 residues C terminal to the conserved tryptophan or truncated to the KS core. In general, KS domain constructs that include the ATd are more soluble than the corresponding KS core constructs. Of the 29 KS domains examined for large-scale expression, 20 were soluble and 16 produced crystals, yielding seven structures (*SI Appendix, Table S3*). The structurally characterized KS domains are MgsE-KS3, MgsF-KS4, MgsF-KS6, OzmQ-KS1, OzmN-KS2, OzmH-KS7, and OzmH-KS8 (*SI Appendix, Tables S4 and S5 and Fig. S4*).

SSA of AT-Less Type I PKS KSs Reveals that Sequence Clusters Correlate with Predicted Substrates.

To determine whether the substrate specificities of our KSs with solved structures correlated with other KSs having similar sequences, we performed phylogenetic analysis and SSA (*SI Appendix*) (24). The KS domains from each pathway were assigned a substrate based on colinearity and a retrobiosynthetic analysis of the major product from each pathway (*SI Appendix, Table S1 and Fig. S5*), and the substrate structures were overlaid on the SSA (Fig. 2 and *SI Appendix, Fig. S6*). In the SSA, we will refer to E-value cutoffs close to zero ($>E-200$) as high stringency and E-value cutoffs further from zero ($E-165$, $E-185$) as low stringency. The SSA reveals clear correlation between clusters of sequences and clusters of substrates. At higher stringency, highly related clusters break apart, revealing detailed relationships and subclusters that have evolved through subfunctionalization. At lower stringency, new relationships between the clusters start to emerge, revealing clusters of sequences that have likely evolved by gene duplication. Overlaying clade names from a previous phylogenetic analysis (18) onto the SSA reveals a high correlation with phylogenetic clades and the sequence clusters, but also reveals many more clades than previously described, in part because we have captured a much larger dataset (*SI Appendix, Fig. S6B*). Overall, the SSA is appropriate for analyzing this very large set of proteins, where there is a high likelihood of very different rates of evolution as a result of a mixture of horizontal gene transfer and gene duplication events.

Canonical Type I PKS KS Domains Accepting β -Branched and *N*-Acyl Amino Acid Substrates Are More Similar to AT-Less Type I PKS KS Domains.

The SSA reveals that canonical type I PKS KS domains form a distinct cluster with members that are all very related, with an average identity of $\sim 67\%$, whereas the AT-less type I PKS KSs have an average identity of $\sim 42\%$ (*SI Appendix, Fig. S7C*). Some AT-less type I PKSs occasionally contain canonical KS-AT didomains, and not surprisingly, they cluster with the majority of canonical type I PKS KSs (Fig. 2 and *SI Appendix, Fig. S6*). At higher stringency,

Fig. 2. Sequence similarity network of canonical and AT-less type I PKS KS core domains, at various E-value cutoffs. Large symbols represent structurally characterized KSs, and their labels are placed nearby. Triangles represent canonical type I PKS and hybrid NRPS/PKS KS-AT didomains and are not colored. Colored triangles represent KS-AT didomains from AT-less type I PKS, circles represent AT-less type I PKS KSs, diamonds represent KS⁰ and nonelongating AT-less type I PKS KSs, and vees represent β -branching AT-less type I PKS KSs. A graphical legend depicting the color and symbol scheme for the KS substrates is shown below the sequence similarity networks.

cyanobacterial and myxobacterial type I PKS KS domains form distinct clusters near actinobacterial type I PKS KSs. However, the two β -branch accepting type I PKS KS domains, CurF-KS2 and JamJ-KS2, which are from cyanobacteria, cluster with the AT-less type I PKS KSs. Thus, the cyanobacterial type I PKSs appear to possess KS domains from AT-less type I PKS origins.

Hybrid NRPS/PKS KS domains form two clusters at medium stringency, representing canonical and AT-less type I PKSs, respectively (Fig. 1 and *SI Appendix*, Fig. S6). At lower stringency, they cluster closer to the AT-less type I PKS KSs. Previous phylogenetic analyses also revealed that hybrid PKS/NRPS domains form clades distinct from canonical type I PKS KS domains (25–27). However, to our knowledge, this is the first study revealing that they are more closely related to the AT-less type I PKS KS family, suggesting that the evolutionary relationships between canonical and AT-less type I PKSs are more complex than previously reported.

The Structures of OzmQ-KS1 and OzmN-KS2 Represent KS Domains from Hybrid NRPS/PKS. The OzmQ-KS1 and OzmN-KS2 domains, which accept β -formamide and oxazole substrates, respectively, cluster most closely with *N*-acyl amino acid-accepting domains from type I PKSs rather than AT-less type I PKSs (*SI Appendix*, Fig. S6). This suggests that these domains have recently been incorporated into the OZM biosynthetic gene cluster from a type I PKS cluster by horizontal gene transfer. The structures of OzmQ-KS1 and OzmN-KS2 therefore represent not only *N*-acyl amino acid-accepting KSs from AT-less type I PKSs but also a much larger class of KS domains found in hybrid NRPS/PKS gene clusters from bacteria across diverse phyla.

The Structure of OzmH-KS8 Represents a Subset of KS⁰ Domains Typically at the Interface Between AT-Less Type I PKS Proteins. OzmH-KS8 is predicted to be nonfunctional for C–C bond formation, as there is an asparagine in place of the first catalytic histidine. Many of the KS⁰ domains are highly divergent and only cluster at low stringency, and in phylogenetic analyses, these domains tend to form clades, but this is likely because of long-branch attraction (28, 29). However, some KS⁰ domains form clusters in the SSA, and OzmH-KS8 clusters with KS⁰ domains that are a part of split bimodules, such as PksJ-KS4, PksJ-KS8, DifG-KS5, and TaiK-KS8, where these KS⁰ domains are at the end of AT-less type I PKS proteins and lack ATd domains (*SI Appendix*, Fig. S1C).

The Structure of MgsE-KS3 Represents KS Domains Accepting Glutarimides and β -Branch Lactones. MgsE-KS3 accepts a glutarimide and phylogenetically clades with other glutarimide-accepting KS domains such as ChxE-KS3, SmdI-KS3, and LtmE-KS3. MgsE-KS3 is in a cluster with RhiF-KS15, which follows the β -branch lactone introducing RhiE-KS14, and other domains that accept six-membered rings at the β -position. Most of the other domains that cluster with MgsE-KS3 accept alkane bearing substrates, including PksJ-KS2, which was previously structurally characterized (Fig. 2).

The Structures of MgsF-KS4, MgsF-KS6, and OzmH-KS7 Represent the Largest Cluster of AT-Less Type I PKS KS Domains. MgsF-KS6 and OzmH-KS7 accept an α -methyl-alkene and alkene substrate, respectively, and are located within the largest AT-less type I PKS KS cluster (Fig. 2 and *SI Appendix*, Fig. S6). MgsF-KS4 accepts a β -hydroxy substrate and clusters near MgsF-KS6 and OzmH-KS7. At higher stringency, these KS domains clade more closely with other domains sharing their respective substrates. The previously

structurally characterized RhiE-KS14 that performs β -branching is also connected to the largest cluster.

Structural Comparison Reveals Similar Overall Structures that Roughly Correlate with the SSA Results. The overall structures of the AT-less type I PKS KS domains are very similar (Fig. 3 and *SI Appendix, Figs. S8 and S9*). Overall comparison of the KS core regions using DALI pairwise structural alignment Z-scores as a similarity matrix reveals that the structures fall into four clades (Fig. 4 and *SI Appendix*). The structure of OzmH-KS8 is unique as a result of most of the loops covering the active site lacking interpretable electron density, which is interpreted as conformational heterogeneity, as SDS/PAGE analysis

reveals the protein is intact (*SI Appendix, Fig. S4*). The structures of MgsF-KS4, OzmH-KS7, and MgsF-KS6 clade closely together with the structure of RhiE-KS14 in the same branch. MgsE-KS3 and PksJ-KS2 have similar overall structures and, surprisingly, form a clade adjacent to the type I PKS KSs. OzmQ-KS1 and OzmN-KS2 have very conserved overall structures and are the least similar to the other groups. The structural clustering is similar to patterns found in the SSA, revealing the SSA clustering captures structural features.

Differences in Structural Homology Are Mostly Attributed to the Variable Structures of Three Loops. Although the core structures of the KS domains are highly conserved, a few loops vary greatly (Fig. 3

Fig. 3. Active sites of AT-less type I PKS KSs. (Left) Overviews of the active sites. (Right) Stereograms of the active sites. The numbers correspond with regions outlined in the text and in *SI Appendix, Fig. S10*. The “clasp loop” is blue, the “dimer interface loop” is cyan for the highlighted monomer and wheat for loop emanating from the mate, and the “active site cap” is green. The conserved catalytic residues are orange (Cys199, His336, and His385), “covering residues” are red, and the residue flanking the active site cysteine is yellow. The substrate found in PksJ-KS2 active site is black. (A) OzmH-KS8, (B) MgsF-KS4, with MgsF-KS6 overlaid in stereogram, (C) MgsE-KS3, with PksJ-KS2 overlaid in stereogram, and (D) OzmN-KS2, with OzmQ-KS1 overlaid in stereogram.

Fig. 4. Structural similarity relationships of KS domains. This dendrogram is based on DALI structural alignment Z-scores.

and *SI Appendix, Figs. S8 and S9*). These loops are highlighted in a structure-based alignment that enumerates the conserved structural features (*SI Appendix, Fig. S10*). All the structures have an active site tunnel that is formed by interactions at the dimer interface and with the variable loops covering the conserved active site residues (*SI Appendix, Fig. S12*).

One of the differing loops is defined by the region between β -sheet 2 and β -sheet 3, which forms an outer “clasp loop” that is highly variable in structure and sequence. The clasp loop may interact with upstream modifying domains, as found in the structure of RhiE-K14 (*SI Appendix, Fig. S9B*). The clasp loop is short in most of the KS⁰ domains and is missing from the electron density in the structure of OzmH-KS8. Therefore, a shortened clasp loop likely influences the structural dynamics of the loops covering the active site.

A second variable “dimer interface loop” that makes up much of the dimer interface is defined by the region between β -sheet 5 and α -helix 4 in the structure-based sequence alignment. This loop is highly variable in sequence and structure and makes many contacts with itself across the twofold dimer interface. Once again this loop is missing in the structure of OzmH-KS8. For MgsE-KS3 and PksJ-KS2, this loop contains two helices; MgsF-KS4, MgsF-KS6, OzmH-KS7, and RhiE-KS14 have a single extended loop and helix; and OzmQ-KS1 and OzmN-KS2 have three helices. A comparison of this loop with the PKS KS structures reveals it is most similar to MgsE-KS3 and PksJ-KS2, which explains why the type I PKS KS domains cluster closer to MgsE-KS3 and PksJ-KS2 than MgsF-KS4, MgsF-KS6, and OzmH-KS7 in the structural homology dendrogram.

The clasp loop and dimer interface loop flank an α -helix, as defined by the residues between β -sheet 7 and β -sheet 8 in the structure-based sequence alignment, which lies above the active site cysteine forming an “active site cap.” Residues emanating from this region form the top of the active site. The same residues in OzmQ-KS1 and OzmN-KS2 form an extended loop conformation with a β -sheet (β -sheet B, which interacts with β -sheet A), differentiating the family of hybrid NRPS/PKS KSs. The structure of this loop is unknown in OzmH-KS8 because of the disorder, and is lacking sequence conservation with the other structures. Nevertheless, the active site capping helix or loop generates very different active site tunnel topologies between the sets of structures (*SI Appendix, Fig. S11*).

Subtle Differences in Active Site Residues Between Members of a Cluster Likely Confer Substrate Specificity. The active site cysteine, position 199 (Fig. 3), is buttressed with the two active site histidines and the N terminus of α -helix 11, making the bottom of the active site very conserved on one side. The residue immediately N-terminal to the active site cysteine is typically either a small residue or a large bulky residue. This residue is covariant with three adjacent residues; thus, when there is a small residue at position 198, the residues at 177 and 461 are larger, as in MgsE-KS3, PksJ-KS2, OzmQ-KS1, and OzmN-KS2, whereas when this residue is a large hydrophobic residue, the others are small residues, as in OzmH-KS7, MgsF-KS4, and MgsF-KS6. These covariant residues act

together to alter the other half of the active site bottom near the tunnel (Fig. 3 and *SI Appendix, Fig. S9*).

Subtle differences in the active sites further differentiate members of MgsF-KS4 from MgsF-KS6 and OzmH-KS7. Most of the residues lining the active site pocket for MgsF-KS6 and OzmH-KS7 are identical (Met-144, Tyr-145, Ser-174, Ser-177, Met-198, Tyr-261, and Ala-461), which makes a nonpolar surface. The residues in MgsF-KS4 that differ are Thr-144, Lys-145, and Phe-261, and it is likely that the introduction of Thr-144 and Lys-145 into the active site increases the polarity of the tunnel and helps accommodate the β -hydroxy substituent over alkenes, as in MgsF-KS6 and OzmH-KS7.

MgsE-KS3 and PksJ-KS2 have modest differences in active site residues, which may help bind their respective substrates. The active site residues of MgsE-KS3 are Ser-144, Gly-145, Ala-174, Met-177, Tyr-261, and Phe-461, whereas those of PksJ-KS2 are Gly-144, Asn-145, Ile-174, Val-177, and Tyr-261. However, there are two striking differences between the active sites of MgsE-KS3 and PksJ-KS2. The first is a difference in the conformation of the residues at position 144 and 145, resulting from the bulky hydrophobic residue Leu-388 in PksJ, which is a serine in MgsE-KS3 and typically a small polar residue in other sequences. Second, PksJ-KS2 also has a methionine inserted at position 173, which is unique among members of its sequence cluster.

Comparison of the ATd Domains Reveals Evolution Caught in Action.

The defining characteristic of the AT-less type I PKS is the lack of an AT domain within each module. However, a few AT-less type I PKSs contain canonical type I PKS KS-AT didomains, suggesting occasional recombination events between these two PKS architectures. Examination of the AT-less type I PKS KS domain structures reveals that most possess an ATd that has a minimal insertion between positions 613 and 942 in the structure-based sequence alignments (*SI Appendix, Fig. S10*). However, certain variants have much larger insertions, and sequence alignments reveal that the insertions are large fragments of an AT domain and may represent evolutionary intermediates between the type I PKS and AT-less type I PKS architectures. Another set had insertions within the ATd that appears to be a degraded ACP domain near position 535. Many KS⁰ domains lacked the ATd and C-terminal tryptophan altogether, such as OzmH-KS8, and were typically found at the C terminus of AT-less type I PKS proteins (*SI Appendix, Table S1 and Fig. S2C*). The lack of ATd domains from some AT-less type I PKSs reveals that AT binding to the KS is not required for activity.

Fig. 5. Stereogram overlay of the ATd of OzmQ-KS1 and EryAII-KS3 AT and linkers. The OzmQ-KS1 ATd is magenta, and the AT fragments are yellow, with the AT-KS linker in red, and EryAII-KS3 is gray. This image reveals remnants of an AT domain for OzmQ, which can also be seen in *SI Appendix, Fig. S10*.

OzmQ-KS1 contains extra sequence beyond the ATd core that forms four α -helices and three β -sheets. Structural alignments of the OzmQ-KS1 ATd with the AT domains from EryAII-KS3 and CurL-KS8 reveal structured portions of the AT that remain, and these regions are poorly aligned in sequence alignments and not recognizable as AT fragments (Fig. 5 and *SI Appendix*, Fig. S10). As OzmN-KS2 does not have this extension, we postulate that the AT domain is in the process of evolutionary degradation. In particular, this suggests that the AT-less type I PKS KS domains following NRPS modules evolved from type I PKS through the loss of the AT domain with retention of the KS/AT adapters. In general, this suggests that the other AT-less type I PKS KS domains have evolved from type I PKS through loss of the AT domain.

Conclusions

Taking a high-throughput structural genomics analysis approach to studying the family of AT-less type I PKS KS domains, which has members with varying degrees of sequence similarity and is more akin to a superfamily, revealed the architectural framework on which the various activities and substrate specificities are built. These structures reveal that three loops covering the active site have differing conformations that correlate with large clusters of KS sequences, and amino acid residues in the active sites correlate with subclusters of KS sequences and substrate specificity. As the substrates of the AT-less type I PKS KS domains are extremely diverse, the family of AT-less type I PKS KS domains gives us a unique opportunity to study the molecular details for the evolution of substrate specificity. These molecular details are beginning to be revealed by this preliminary analysis. Although this study captures four of the structural clades and a couple of the sequence subclusters, there remain many other important structures that will provide further insight into the evolution of substrate specificity within the family of AT-less type I PKS KSs (*SI Appendix*).

The SSA reveals that the evolutionary relationships among KS domains of canonical, hybrid NRPS/PKS, and AT-less type I PKSs are more complicated than previously reported. Type I PKS KSs accepting β -branched and *N*-acyl amino acids are significantly different from the canonical type I PKS KSs and more similar to those found in AT-less type I PKSs. The structures of OzmQ-KS1 and OzmN-KS2 reveal structural insights into the larger family of *N*-acyl amino acid accepting KSs; most important,

the conformation of an active site loop differentiates them from acyl-accepting KS domains. These structures will serve as a platform for understanding the unique chemistry in the biosynthesis of hybrid peptide-polyketide natural products.

Within the family of AT-less type I PKSs, there are modules that contain complete KS-AT didomains from type I PKS; have remnants of the KS/AT adapters and large portions of a nonfunctional AT domain as represented by OzmQ-KS1; have a minimal ATd, as represented by most of our structures; or completely lack the ATd, as represented by OzmH-KS8. The large diversity of structural motifs in the AT-less type I PKS family reveals the consequences of extensive horizontal exchange within the AT-less type I PKSs and with the type I PKSs. Furthermore, the structures reveal the molecular consequences of a type I PKS KS-AT didomain evolving toward an AT-less type I PKS KS, supporting the conclusion that other AT-less type I PKS KS domains may have evolved from type I PKSs through the gradual loss of the AT domain. Together, these studies reveal molecular details of type I PKS evolution, which can be used to guide engineered biosynthesis of novel polyketide natural products.

Methods

Materials, methods, and detailed experimental procedures are provided in *SI Appendix*. *SI Appendix* tables include description of KS domain regions used for bioinformatics analysis (*SI Appendix*, Tables S1 and S2), results of large-scale protein production and purification (*SI Appendix*, Table S3), primers and plasmids used in these studies (*SI Appendix*, Table S4), crystallographic data collection and structure refinement statistics (*SI Appendix*, Table S5), and results from the DALI structural alignments (*SI Appendix*, Table S6). *SI Appendix* figures include depiction of domains leading to various substrates known for KSs of AT-less type I PKSs (*SI Appendix*, Fig. S1), canonical and AT-less type I PKS architecture (*SI Appendix*, Fig. S2), subdomain schematic for the KS domains (*SI Appendix*, Fig. S3), SDS/PAGE of selected crystallized proteins (*SI Appendix*, Fig. S4), proposed OZM and MGS biosynthetic pathways (*SI Appendix*, Fig. S5), the annotated SSA of all KSs (*SI Appendix*, Fig. S6), BLAST analyses of all KSs (*SI Appendix*, Fig. S7), overview of KS-ATd structures (*SI Appendix*, Fig. S8), comparison of KS active sites (*SI Appendix*, Fig. S9), structure-based sequence alignment of all KSs (*SI Appendix*, Fig. S10), and comparison of KS active site tunnels (*SI Appendix*, Fig. S11).

ACKNOWLEDGMENTS. This work was supported in part by NIH Grants GM094585 (to A.J.), GM098248 (to G.N.P.), and CA106150 (to B.S.) and by the US Department of Energy under Contract DE-AC02-06CH11357.

- Cheng YQ, Coughlin JM, Lim SK, Shen B (2009) Type I polyketide synthases that require discrete acyltransferases. *Methods Enzymol* 459:165–186.
- Piel J (2010) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* 27(7):996–1047.
- Musioli EM, Weber T (2012) Discrete acyltransferases involved in polyketide biosynthesis. *MedChemComm* 3(8):871–886.
- Cheng YQ, Tang GL, Shen B (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc Natl Acad Sci USA* 100(6):3149–3154.
- Tang GL, Cheng YQ, Shen B (2004) Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem Biol* 11(1):33–45.
- Tang Y, Kim CY, Mathews II, Cane DE, Khosla C (2006) The 2.7-Ångström crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci USA* 103(30):11124–11129.
- Gay DC, et al. (2014) A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure* 22(3):444–451.
- Calderone CT (2008) Isoprenoid-like alkylations in polyketide biosynthesis. *Nat Prod Rep* 25(5):845–853.
- Pöplau P, Frank S, Morinaka BI, Piel J (2013) An enzymatic domain for the formation of cyclic ethers in complex polyketides. *Angew Chem Int Ed Engl* 52(50):13215–13218.
- Gay DC, Spear PJ, Keatinge-Clay AT (2014) A double-hotdog with a new trick: Structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem Biol* 9(10):2374–2381.
- Du L, Sánchez C, Shen B (2001) Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering novel molecules. *Metab Eng* 3(1):78–95.
- Bretschneider T, et al. (2013) Vinyllogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* 502(7469):124–128.
- Kusebauch B, Busch B, Scherlach K, Roth M, Hertweck C (2009) Polyketide-chain branching by an enzymatic Michael addition. *Angew Chem Int Ed Engl* 48(27):5001–5004.
- Seo JW, et al. (2014) Comparative characterization of the lactimidomycin and isomigrastatin biosynthetic machineries revealing unusual features for acyltransferase-less type I polyketide synthases and providing an opportunity to engineer new analogues. *Biochemistry* 53(49):7854–7865.
- He HY, Tang MC, Zhang F, Tang GL (2014) Cis-Double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis. *J Am Chem Soc* 136(12):4488–4491.
- Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* 22(10):2027–2039.
- Piel J, Hui D, Fusetani N, Matsunaga S (2004) Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol* 6(9):921–927.
- Nguyen T, et al. (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* 26(2):225–233.
- Jenner M, et al. (2013) Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angew Chem Int Ed Engl* 52(4):1143–1147.
- Kohlhaas C, et al. (2013) Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. *Chem Sci* 4(8):3212–3217.
- Jenner M, et al. (2015) Acyl-chain elongation drives ketosynthase substrate selectivity in trans-acyltransferase polyketide synthases. *Angew Chem Int Ed Engl* 54(6):1817–1821.
- Zhao C, et al. (2010) Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an “acyltransferase-less” type I polyketide synthase that incorporates two distinct extender units. *J Biol Chem* 285(26):20097–20108.
- Lim SK, et al. (2009) iso-Migrastatin, migrastatin, and dorrigin production in *Streptomyces platensis* NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J Biol Chem* 284(43):29746–29756.
- Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC (2009) Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* 4(2):e4345.
- Ziemert N, et al. (2012) The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* 7(3):e34064.
- Ginolhac A, et al. (2005) Type I polyketide synthases may have evolved through horizontal gene transfer. *J Mol Evol* 60(6):716–725.
- Moffitt MC, Neilan BA (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J Mol Evol* 56(4):446–457.
- Philippe H, Laurent J (1998) How good are deep phylogenetic trees? *Curr Opin Genet Dev* 8(6):616–623.
- Kolaczowski B, Thornton JW (2009) Long-branch attraction bias and inconsistency in Bayesian phylogenetics. *PLoS One* 4(12):e7891.