Structure-guided function discovery of an NRPS-like glycine betaine reductase for choline biosynthesis in fungi

Yang Hai (海洋)², Arthur M. Huang¹, and Yi Tang¹,a,b,1

¹Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90095; and ²Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095

Edited by Wilfred A. van der Donk, Howard Hughes Medical Institute and University of Illinois, Urbana–Champaign, Urbana, IL, and accepted by Editorial Board Member Stephen J. Benkovic April 10, 2019 (received for review February 27, 2019)

Nonribosomal peptide synthetases (NRPSs) and NRPS-like enzymes have diverse functions in primary and secondary metabolisms. By using a structure-guided approach, we uncovered the function of a NRPS-like enzyme with unusual domain architecture, catalyzing two sequential two-electron reductions of glycine betaine to choline. Structural analysis based on the homology model suggests cation-π interactions as the major substrate specificity determinant, which was verified using substrate analogs and inhibitors. Bioinformatic analysis indicates this NRPS-like glycine betaine reductase is highly conserved and widespread in kingdom fungi. Genetic knockout experiments confirmed its role in choline biosynthesis and maintaining glycine betaine homeostasis in fungi. Our findings demonstrate that the oxidative choline-glycine betaine degradation pathway can operate in a fully reversible fashion and provide insight in understanding fungal choline metabolism. The use of an NRPS-like enzyme for reductive choline formation is energetically efficient compared with known pathways. Our discovery also underscores the capabilities of the structure-guided approach in assigning functions of uncharacterized multidomain proteins, which can potentially aid functional discovery of new enzymes by genome mining.

glycine betaine reductase | NRPS-like | choline metabolism | biosynthesis | fungi

The nonribosomal peptide synthetase (NRPS)-like carboxylic acid reductases (CARs) catalyze ATP- and NADPH-dependent reduction of carboxylic acids to the corresponding aldehydes (1). These multidomain enzymes consist of a N-terminal adenylation (A) domain, which selects and activates a carboxylic acid substrate by adenylation, then transfers the acyl moiety to the phosphopantetheinyl arm from the thiolation (T) domain through a thioester linkage. The T domain delivers the acyl thioester to the C-terminal thioester reductase (R) domain where the NADPH-dependent thioester reduction occurs (Fig. 1A). Unlike some off-loading R domains in polyketide synthases and NRPS assembly lines, which catalyze nonprocessive [2 + 2] electron reduction to over-reduce the aldehyde product to alcohol (2, 3), the R domains in CARs catalyze a strict two-electron reduction, and the structural basis for such specificity was reported recently (4).

The members of the CAR family are functionally diverse and play indispensable roles in both primary and secondary metabolisms, such as the α-aminoadipate reductase Lys2, required for lysine biosynthesis in fungi (5); and tyrosine reductase HqL, involved in biosynthesis of herculeine A (Fig. 4A) (6). CARs also show substantial promise as green biocatalysts for the conversion of aromatic or aliphatic carboxylic acids into the corresponding aldehydes. Indeed, synthetic applications of CARs have been reported for the biocatalytic synthesis of vanillin (7), preparation of methyl branched aliphatic aldehydes (8), production of biofuels (9, 10), and other chemical commodities (11).

Prominent metabolic roles of CARs and their biocatalytic potential encouraged us to discover uncharacterized members with different functions by genome mining. During this effort, we focused on a fungal CAR-like protein with an unknown function. It is distinguished from all other CARs due to an extra C-terminal YdTG-like short-chain dehydrogenase/reductase domain. We named this protein ATTR and the corresponding gene atrr after its unusual domain architecture (A-T-R₁-R₂). The presence of two fused R domains in ATTR implies that it could catalyze two consecutive two-electron reductions of a carboxylic acid to yield an alcohol (Fig. 1B). Moreover, a genomic survey reveals that atrr genes are widespread but exclusive to eukaryotes, mostly in fungi and found in several protist and invertebrate species. In particular, ATTR orthologs are highly conserved in kingdom fungi: Sequence identities are greater than 60% from different species, which indicates a unified and conserved role of ATTR. Transcriptomics studies also suggest atrr genes are constitutively transcribed in different fungal species, including human opportunistic pathogen Aspergillus fumigatus (12, 13); and rice blast fungus Magnaporthe oryzae (14). These features motivated us to discover and characterize its enzymatic activity and biological function.

Results

Bioinformatic Analysis Provides Initial Insight. We first applied the “genomic enzymology” strategy to obtain clues regarding the ATTR function from the atrr genomic context since enzymes in a microbial metabolic pathway often are encoded by a gene cluster (15, 16). However, comparative genomic analysis of atrr orthologs in fungi reveals that there are no well-defined and conserved

Significance

Oxidation of choline to glycine betaine is a major pathway for choline degradation and glycine betaine biosynthesis. Although this oxidative process is found in all domains of life, enzymes involved in the reduction of glycine betaine to choline have not been found. Here, we found this missing link in choline metabolism through a structure-guided function discovery of a NRPS-like glycine betaine reductase, which catalyzes reduction of glycine betaine to choline via the intermediate glycine betaine aldehyde. Bioinformatic analysis suggests this enzyme is highly conserved and widespread in the fungal kingdom, and our discovery of this reductive choline biosynthesis pathway provides insight into fungal choline metabolism.

Author contributions: Y.H. and Y.T. designed research; Y.H. and A.M.H. performed research; Y.H. analyzed data; and Y.H. and Y.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. W.A.v.d.D. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

1To whom correspondence should be addressed. Email: yitang@ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1903282116/-/DCSupplemental.

Published online May 6, 2019.
gene neighborhoods. Even in some closely related species of the same genus (e.g., Aspergillus), the genome environment of atrr differs substantially (SI Appendix, Figs. S2 and S3). The diversified genomic context and lack of obvious co-occurrence with signature metabolic pathway genes prevented us from reliable functional assignment.

We then focused on characterization of the A domain of ATRR (ATRR-A), which selects a substrate and is, therefore, the entry point to ATRR catalysis. Because the substrate specificity of A domains is dictated by the amino acid residues lining the active site pocket (10 AA code), deciphering this “nonribosomal code” of ATRR-A would, in theory, allow us to predict its function (17, 18). Bioinformatic analysis showed the ATRR-A has a unique 10 AA code, and it is conserved among ATRR orthologs from different fungal species (SI Appendix, Fig. S4). However, substrate prediction by online servers was unsuccessful due to the lack of characterized homologs with significant similarity (19). Closer inspection reveals that the highly conserved aspartate residue (D235 in PheA), which typically interacts with the substrate α-amino group via a salt bridge (20, 21), is not conserved in ATRR-A (SI Appendix, Fig. S5). Loss of this key aspartate residue suggests that the substrate is not a typical α-amino acid or an unusual amino acid substrate accommodated in an unprecedented fashion as in the case of ClbH-A2 which activates S-adenosyl-l-methionine (SAM) (22). Indeed, most characterized A domains without D235 equivalent residues (Asp/Glu) are known to activate nonamino acid-type substrates, including aryl acid, keto acid, and hydroxy acid (SI Appendix, Fig. S5) (23–26). Based on this prediction, we screened the CAR activity of ATRR in vitro against a library of common carboxylic acids from primary metabolism as well as the 20 proteinogenic amino acids (SI Appendix, Fig. S6). The Aspergillus nidulans ATRR protein (UniProt AN53182.2) was expressed from Escherichia coli BL21(DE3) and purified to homogeneity by affinity chromatography and size-exclusion chromatography (SI Appendix, Fig. S7). Aporo-ATRR was enzymatically converted to holo-form by using phosphopantetheinylation transferase NpgA and CoA (SI Appendix, Fig. S8) (27). The reductase activity of holo-ATRR was examined by monitoring the oxidation of NADPH at 340 nm. ATRR showed no activity toward any of the substrates in this library.

**Structure-Based Approach Predicts a Domain Substrate.** We then used a structure-based approach to gain more insight about the substrate specificity of ATRR-A, a strategy that has led to many successful predictions for enzymes of unknown function (28, 29). We constructed an active site homology model of ATRR-A by using a TycA variant (PDB entry 5N82) (21) as the template (34% sequence identity to ATRR-A) and manually replacing the 10 AA code residues with the corresponding ones in A. nidulans ATRR-A (Fig. 2A). In comparison with the structure of the archetypical A domain PheA (17), three aromatic residues (F201, W204, and W270) stood out in the ATRR-A model. These residues could potentially form an aromatic cage to bind a quaternary ammonium group through ionic interactions (30, 31). This quaternary ammonium group can be further electrostatically stabilized by a nearby anionic residue E304. Such an aromatic box is reminiscent of those observed in quaternary ammonium moiety binding proteins, such as periplasmic betaine binding protein Prox (32); cis-4-hydroxy-D-proline betaine epimerase HpbD, and demethylase HpbJ (28); histidine triad thymyllysine reader proteins (33); and acetylcholine esterase (30, 31).

We, therefore, speculated that the carboxylic acid substrate of ATRR could be a small betaine, such as the physiologically abundant carnitine, glycine betaine, etc. To test this hypothesis, we assayed ATRR activity with a panel of naturally occurring betaines (Fig. 2B). Gratifyingly, rapid oxidation of NADPH was observed only in the presence of the smallest betaine, glycine betaine. Consumption of NADPH indicates that glycine betaine was reduced. In accordance with our early hypothesis about ATRR function (Fig. 1B), the predicted corresponding two sequential two-electron reduction product choline was indeed accumulated in the reaction mixture (SI Appendix, Fig. S9). We then determined the apparent steady-state kinetic constants for ATRR glycine betaine reductase activity (SI Appendix, Fig. S10 and Table S1). The overall Kcat value (~1 s−1) is modest but similar to that of previously characterized CARs (4), whereas the Km value (~2 mM) matches the intracellular level of glycine betaine (~3 mM) in fungi (34). Therefore, glycine betaine is a physiologically relevant substrate for ATRR.

**Mechanistic Study of ATRR.** We next studied the catalytic mechanism of the glycine betaine reductase function of ATRR. By using a hydroxylamine-trapping assay designed for testing deamination activity (35), we confirmed that glycine betaine was indeed adenylated by ATRR-A (Fig. 3A). Furthermore, similar to all CARs, the glycine betaine reductase activity of ATRR

---

**Fig. 1.** Comparison between typical CARs and ATRR. (A) Domain architecture of typical CARs and CAR-catalyzed reactions. (B) Domain architecture of ATRR and the postulated reaction based on domain function.
In Atrr, we introduced another mutation at R2 (G1036A) to compromise binding of the cosubstrate NADPH. As expected, double mutation (G1036A/Y1178F) in R2 further decreased the aldehyde reductase activity by 4,000-fold, whereas the carboxylic acid reductase activity was not compromised. The role of R2 as a dedicated aldehyde reductase was also supported by the robust reduction of betaine aldehyde to choline using the excited stand-alone R2 domain (SI Appendix, Table S1). To summarize, the thioester reductase domain R2 strictly carries out a single two-electron reduction of glycine betainoyl thioester to glycine betaine aldehyde, which is then reduced to choline at the aldehyde reductase domain R1.

Substrate Specificity and Inhibition of ATRR. Establishing the function of ATRR as glycine betaine reductase enabled us to further characterize its substrate specificity. We first studied ATRR specificity toward glycine betaine by assaying both A and carboxylic acid reductase activity of ATRR against different glycine betaine analogs (Fig. 3 A and B). The isosteric neutral analog 3,3-dimethylbutyrate completely abolished the activity. In contrast, the tertiary sulfonium analog sulfo betaine, which also bears a positive charge on its head group but lacks one methyl group

---

**Fig. 2.** Structure-based prediction of the ATRR substrate. (A) Comparison of the homology model of A. nidulans ATRR-A domain with the structure of PheA in complex with its cognate substrate L-phenylalanine. The conserved salt bridges recognizing the α-carboxylate and α-amino groups are shown as black dashes. The proposed cation–π and electrostatic interactions favoring binding of betaine-like substrates in the ATRR A domain are shown as orange and red dashes, respectively. The simplest betaine, glycine betaine, was shown to indicate the plausible binding mode. (B) CAR activity of ATRR against a panel of betaine substrates.

---

**Fig. 3.** ATRR substrate specificity and inhibition studies. (A) Substrate specificity of ATRR-A profiled with different substrates. The A activity was monitored here by continuous trapping with phenylhydrazine (SI Appendix, Fig. S12). Direct reduction of glycine betaine aldehyde to choline by ATRR was also observed, confirming its role as an intermediate in the stepwise reduction (SI Appendix, Fig. S9). Kinetic assay showed that glycine betaine aldehyde reduction by ATRR is more efficient: The catalytic efficiency (kcat/Km = 3.0 × 10^4 M⁻¹ s⁻¹) is 66-fold greater than that of the overall carboxylic acid reduction, whereas the turnover number (kcat = 7 s⁻¹) is sixfold higher. Rapid reduction of glycine betaine aldehyde shows that the aldehyde reduction step is not rate limiting thereby avoiding accumulation of the reactive aldehyde species.

We next characterized the individual function of R1 and R2 domains. To isolate the catalytic activity of the individual domains, we inactivated each reductase domain by substituting the conserved catalytic tyrosine residue with phenylalanine (Y811F in R1 and Y1178F in R2). In agreement with our two consecutive two-electron reductions hypothesis (Fig. 1B), mutation of R2 essentially abolished overall carboxylic acid reductase activity, whereas aldehyde reductase activity remained intact (SI Appendix, Table S1). On the contrary, mutation of R2 did not substantially affect carboxylic acid reductase activity but resulted in the accumulation of glycine betaine aldehyde intermediate (SI Appendix, Fig. S12). The Y1178F mutation in R2 also led to a greater than 150-fold loss of aldehyde reductase activity (catalytic efficiency). (SI Appendix, Table S1). To further confirm that the remaining aldehyde reductase activity was not contributed
compared with glycine betaine, was partially active (an order of magnitude slower). Such comparison clearly demonstrates that the formal positive charge carried by the substrate is critical for activity and, thus, supports our predicted cation-π interactions conferred by ATRR-A during substrate recognition. The decreased activity of ATRR toward sulfobetaine suggests the importance of the trimethyl moiety for function. The important contribution of the trimethyl moiety of glycine betaine to ATRR catalysis was also revealed by carboxylic acid reduction assays with dimethylglycine and monomethylglycine (sarcosine), which are physiologically relevant analogs as they are intermediates in the glycine betaine degradation pathway (36). Progressively decreasing the number of N-methyl groups of glycine betaine resulted in a significant loss of carboxylic acid reductase activity: the reduction of dimethylglycine was two orders of magnitude slower than that of glycine betaine, whereas no reduction could be observed with sarcosine (Fig. 3 A and B and SI Appendix, Table S1). Since the ammonium groups of both dimethylglycine (pKₐ = 9.89) and sarcosine (pKₐ = 10.1) were thought to be protonated under our experimental condition (pH 7.5–8.0) (37), the decrease in activity per methyl group reflects the additive contribution of each methyl substitution to catalysis. ATRR-A was also shown to be strictly size selective in that γ-alanine betaine and dimethylsulfoniopropionate, each containing one additional methylene group to glycine betaine and sulfobetaine, respectively, were not substrates of ATRR (Fig. 3 A and B). Collectively, these results show that ATRR is highly specific to glycine betaine and such specificity is likely achieved by favoring binding of its head group and distinguishing different extents of methylation.

To corroborate our findings on substrate specificity, we studied the inhibition of ATRR by using different ammonium-containing inhibitors (SI Appendix, Fig. S13 and Table S2). The tetramethylammonium ion, which mimics the head group of glycine betaine, was found to be a potent competitive inhibitor of ATRR-A (Kᵢ = 69 μM), whereas the potency decreased by three orders of magnitude with dimethylammonium (Kᵢ = 21 mM). As illustrated in Fig. 3C, a monotonic increase in inhibitor affinity is correlated with the addition of each methyl group yielding an incremental free energy relationship: Addition of each methyl group contributes ~1.7 kcal/mol toward inhibitor binding. Therefore, our results reinforced the idea that the head group plays a major role in ATRR-A substrate recognition, and ATRR-A favors binding of tetra-alkylated quaternary ammonium moiety, presumably through a combination of cation-π interactions, hydrophobic effect, and maximized van der Waals interactions. Unlike ATRR-A, R₂ is not potently inhibited by the head-group mimic tetramethylammonium ion (mixed inhibitor, Kᵢ = 5, Kᵢ = 18 nM), suggesting the head group is not the dominant determinant for R₂ substrate specificity. Nonetheless, the positive charge on the substrate is still essential for R₂ catalysis as the neutral isosteric analog 3,3-dimethylbutyraldehyde (DMBA) showed a 13,500-fold loss of catalytic efficiency (SI Appendix, Table S1). It is noteworthy to mention that DMBA exists mainly as the aldehyde form (>65%) in aqueous solution favoring reductase-catalyzed NADPH-dependent reduction (38), whereas glycine betaine aldehyde predominantly exists as the hydrated gem-diol form (99%), which must then undergo dehydration to unveil the aldehyde functional group for reduction.

Since the product choline consists of the same head group, we next tested whether choline could act as an inhibitor competing with glycine betaine and glycine betaine aldehyde. As a result, choline was shown to be a mixed inhibitor for both glycine betaine reductase and aldehyde reductase activity but more potent (Kᵢ = 0.27, Kᵢ = 6.7 mM) in competition against glycine betaine in the first reduction step. Therefore, choline could act as a feedback inhibitor to regulate ATRR enzymatic activity (Fig. 3D). The lowered binding affinity of choline to R₂ (Kᵢ = 4 mM) favors the release of choline after glycine betaine aldehyde reduction to avoid direct product inhibition.

**Physiological Role of ATRR in Fungi.** Our discovery of ATRR as a glycine betaine reductase strongly indicates that it is the missing link in choline-glycine betaine metabolism (Fig. 4). Oxidation of choline to glycine betaine initiates the choline catabolic process. This common choline degradation pathway has been found in all domains of life, but the reverse process has only been implicated in fungi. A collection of experimental evidence, including genetics, metabolite quantification by NMR, and radioactive labeling studies, suggest glycine betaine can be reduced to choline in vivo and thereby can be an alternative source of choline for fungi (39, 40). However, the underlying biochemical mechanism
and genetic basis have remained elusive. A plausible mechanism for this conversion was postulated based on the reduction of l-aspartate to l-homoserine in which the carboxy group is first activated via phosphorylation by a kinase. This is followed by conversion to a thioester that is reduced into aldehyde by dehydrogenase and then reduced to alcohol by a separate reductase. The biosynthetic logic of ATRR discovered in this paper also uses a carboxylic activation strategy for reduction but is more concise as all steps are performed by one enzyme, and the key thioester intermediate is channeled through the phosphopantetheine arm of a T domain.

To test whether ATRR plays a role in choline biosynthesis in fungi, we deleted the atr gene in both the wild-type (wt*) A. nidulans and the A. nidulans ΔchoA strains by using the split-marker method (SI Appendix, Figs. S14 and S15) (41). Disruption of the choA gene encoding the phosphatidylethanolamine N-methyltransferase blocks the Bremer-Greenberg methylation pathway from PEA and renders the Kennedy pathway from free choline as the only option for biosynthesis of the essential phospholipid phosphatidylcholine (PC) (Fig. 4) (40). Deletion of the atr gene would abolish conversion of glycine betaine to choline, but both the Bremer-Greenberg pathway and the Kennedy pathway remain functional. Accordingly, we expect no growth defects of the Δatr strain, whereas the ΔchoA strain would show choline/betaine-dependent growth. Without the ability to convert glycine betaine to choline, the ΔatrΔchoA double knockout strain is expected to be truly choline- auxotrophic even in the presence of glycine betaine. Indeed, when cultured on solid minimal medium (MM), no major growth and morphological difference was observed between the Δatr and the wt* strains. On the contrary, neither the ΔchoA nor the ΔatrΔchoA strain was able to grow because of obstructed PC biosynthesis (Fig. 5A). The vegetative growth of the ΔchoA strain can be stimulated by supplementation of either choline or glycine betaine (2–20 μM), although full restoration of the conidiation requires higher concentrations (>100 μM). In contrast, the ΔatrΔchoA strain exhibits choline-auxotrophic character: Glycine betaine can no longer replace choline to restore growth and asexual development under the choline-limiting condition. These results confirm that ATRR rescues the phenotype of ΔchoA by converting glycine betaine to choline. Therefore, ATRR mediated reduction of glycine betaine is a “shortcut” pathway to choline.

Unexpectedly, although no significant difference was observed between the Δatr and the wt* strains growing at low concentration (2–20 μM) of exogenous glycine betaine, the hyphae growth and conidiation of the Δatr strain were severely impaired at high level glycine betaine (>100 μM), Fig. 5A), which indicates ATRR function is essential to A. nidulans under such a condition. To interrogate the role of glycine betaine in the absence of ATRR, we cosupplemented choline to the medium. With an increasing amount of choline (>20 μM) added, the growth of the Δatr strain can be fully reverted to the normal state even in the presence of a millimolar level of glycine betaine (Fig. 5B). Similar but not identical results were also observed when PC and l-methionine (10 mM) were cosupplemented. These chemical complementation results suggest the impaired growth caused by the high level of exogenous glycine betaine is presumably due to the impeded Bremer-Greenberg methylation pathway, which again leads to deficient PC biosynthesis. High levels of methionine could increase SAM biosynthesis via methionine catabolism, which may further up-regulate and drive the methylation pathway against glycine betaine. In comparison, cosupplementation of l-homo cysteine (as racemic l- and d-mixtures) to favor glycine betaine degradation did not alleviate the impaired growth of the Δatr strain, which suggests the glycine betaine degradation pathway is likely not playing a major role in this phenotype. Nevertheless, inhibition of the Bremer-Greenberg methylation pathway by glycine betaine necessitates the function of ATRR in fungi, which not only can supply choline for PC biosynthesis via the Kennedy pathway, but also can help to “detoxify” high levels of glycine betaine. Therefore, our results confirmed an important physiological role of ATRR in choline and PC metabolism in A. nidulans and reveal a previously unknown effect of glycine betaine to A. nidulans growth.

Discussion

Our study here was driven by a highly conserved NRPS-like enzyme in fungi with an unusual domain architecture. Successful elucidation of its unknown function as glycine betaine reductase serendipitously led to the resolution of a long-standing question in choline-glycine betaine metabolism, how glycine betaine was converted back to choline. Now, our study provides the molecular basis: ATRR catalyzes ATP- and NADPH-dependent reductions of glycine betaine by a catalytic mechanism similar to that of well-known NRPS-dependent CARs, generating a glycine betaine aldehyde intermediate, which is further reduced at the R3 domain unique to ATRR, yielding, as a final product, choline. Orthologs of ATRR are found to be prevalent in fungi and exhibit high sequence identities with characterized A. nidulans ATRR (SI Appendix, Figs. S2 and S3), which suggests that this efficient single-enzyme catalyzed reductive choline biosynthesis may be widespread in fungi. Harboring this enzyme permits direct reutilization of endogenously stored glycine betaine (~3 mM) for on-demand biosynthesis of choline and choline derivatives, including phospholipid PC (essential role in maintaining membrane integrity and functionality) and choline-O-sulfate (a mean for intracellular sulfate storage) (42). Compared with the high metabolic cost of the de novo choline biosynthetic pathway (the Bremer-Greenberg pathway) at the expense of three consecutive methylation steps (costing 12 ATP molecules per methylation event) (43), regeneration of choline from glycine betaine is much more cost-effective (costing 1 ATP and 2 NADPH). More importantly, this shortcut pathway also acts as an “emergency safeguard” pathway to supply choline from its biosynthesis.
“reserve form” glycine betaine for PC biosynthesis when the Bremer-Greenberg methylation pathway is obstructed. Uncovering the ATRR function was rooted in understanding the biochemical logic of NRPS-like CARs. We reasoned that the A domain must play a gatekeeping role in selecting and activating the carboxylic acid substrate; and identification of the substrate would thereby reveal the function of ATRR. Although our initial search for the substrate, guided by the 10 AA code, was unsuccessful, a key insight of the A domain substrate specificity was provided by our interpretation of the structure-function relationship based on a simple active site homology model, which prioritized a pool of potential substrates, and the physiologically relevant substrate glycine betaine was identified eventually by experimental enzymatic activity screening. Our results expanded the biochemistry scope of the NRPS family: ATRR-A is shown to activate and thioesterify a betaine-type substrate; and convincing evidence supports our hypothesis that cation-π interactions play an important role in substrate recognition at the A domain. Our findings support the notion that cation-π interaction is a fundamental and universal noncovalent interaction for substrate/ligand recognition in biology (30, 31).

To summarize, our study demonstrated that the structure-guided approach can be a general strategy for functional discovery of uncharacterized NRPS-like enzymes, and orphan NRPS assembly lines from the ever-increasing size of genomic data, which may aid in the discovery of new metabolic pathways and natural products (16). In principle, such a process can be greatly facilitated by docking a library of carboxylic acid metabolites to high-quality homology models of A domains.

Materials and Methods

Materials and methods are summarized in SI Appendix, including SI Appendix, Figs. 51–56 and Tables 51–53.

ACKNOWLEDGMENTS. We thank Dr. Mengbin Chen, Dr. Matthew Jenner for helpful discussions, and Abigail Yohannes for helping with bioinformatic analysis. This work was supported by the NIH (Grant 1R35GM118056) (to Y.T.). Y.H. is a Life Sciences Research Foundation fellow sponsored by the Mark Foundation for Cancer Research.