ACS Catalysis

Letter

pubs.acs.org/acscatalysis

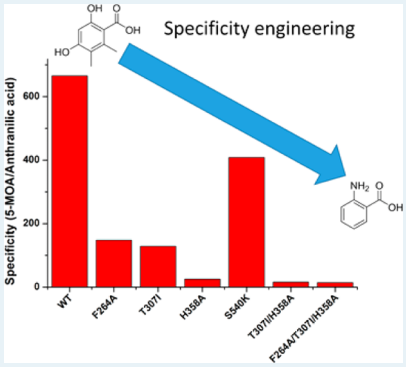
Characterization and Engineering of the Adenylation Domain of a NRPS-Like Protein: A Potential Biocatalyst for Aldehyde Generation

Meng Wang[†] and Huimin Zhao*[,†,‡]

[†]Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, [‡]Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: The adenylation (A) domain acts as the first "gate-keeper" to ensure the activation and thioesterification of the correct monomer to nonribosomal peptide synthetases (NRPSs). Our understanding of the specificity-conferring code and our ability to engineer A domains are critical for increasing the chemical diversity of nonribosomal peptides (NRPs). We recently discovered a novel NRPS-like protein (ATEG_03630) that can activate 5-methyl orsellinic acid (5-MOA) and reduce it to 2,4-dihydroxy-5,6-dimethyl benzaldehyde. A NRPS-like protein is much smaller than multidomain NRPSs, but it still represents the thioesterification half-reaction, which is otherwise missed from a stand-alone A domain. Therefore, a NRPS-like protein may serve as a better model system for A domain engineering. Here, we characterize the substrate specificity of ATEG_03630 and conclude that the hydrogen-bond donor at the 4-position is crucial for substrate recognition. Next, we show that the substrate specificity of ATEG_03630 can be engineered toward our target substrate anthranilate via bioinformatics analysis and mutagenesis. The resultant mutant H358A increased its activity toward anthranilate by 10.9-fold, which led to a 26-fold improvement in specificity. Finally, we demonstrate one-pot chemoenzymatic synthesis of 4-hydroxybenzaldoxime from 4-hydroxybenzoic acid with high yield.



KEYWORDS: *NRPS-like protein, adenylation domain, substrate specificity engineering, aldehyde, one-pot synthesis*

INTRODUCTION

Nonribosomal peptides (NRPs), such as antitumor agent bleomycin and antibiotic vancomycin, are an important group of secondary metabolites. They are synthesized by nonribosomal peptide synthetases (NRPSs), megaenzymes that consist of multiple modules.[1] Each module is usually responsible for activation, modification, and incorporation of a specific monomer into the NRP product. The adenylation (A) domain acts as the first "gate-keeper" to ensure the activation and thioesterification of the correct monomer to the assembly line. Although A domains typically have high specificity, the occasional promiscuity of certain A domains is a known source of congeners of NRPs,[2] which diversify the chemical structures of NRPs as a potential way to hedge against resistance. In light of that, researchers have been developing different approaches to artificially increase the chemical diversity of NRPs by changing the substrate specificity of A domains.

The first approach is domain-swapping,[3,4] which is achieved via replacing the original A-ACP didomains with heterologous A-ACP domains. However, initially, the success of this approach was quite limited due to the complex nature of domain−domain interactions, which may involve complicated conformational changes of the megaenzymes. Later, significant improvements were made through better understanding of the domain−domain communication[5] and the identification and engineering of important linker regions.[6] Nonetheless, the domain-swapping method is still far from optimal.

The second approach involves mutating important code-conferring residues[7,8] to alter the substrate specificity of the A domains.[1] Numerous A domains have been engineered to incorporate nonnatural substrates into the NRP products.[9,10] In addition, directed evolution[11−13] and computational redesign[14] have been used to engineer the substrate specificity of the A domains. However, most of the previous engineering works focused on proteinogenic amino acids. On the contrary, one of the hallmarks and significant advantages of NRPSs is their ability to incorporate nonproteinogenic amino acids,[15] which exponentially increases the chemical complexity of NRPs.

Aryl acids are an important class of monomers of NRPs. Aryl acid monomers are most commonly seen in siderophores,[16] but are not limited to them.[17] However, our understanding of the code-conferring residues of A domains for aryl acids is less profound. In addition, the engineering of an A domain for aryl acids is rare.[11] Recently, we discovered a novel NRPS-like protein, which consists of an A domain, an acyl carrier protein (ACP) domain and a reducing (R) domain, from *Aspergillus terreus*.[18] It is involved in a polyketide biosynthesis pathway and can convert 5-MOA, produced by a dedicated nonreducing

Received: January 12, 2014
Revised: February 17, 2014
Published: March 17, 2014

 © 2014 American Chemical Society dx.doi.org/10.1021/cs500039v | *ACS Catal.* 2014, 4, 1219−1225

polyketide synthase, to 2,4-dihydroxy-5,6-dimethyl benzaldehyde (Figure 1). Compared with their bacterial counterparts,

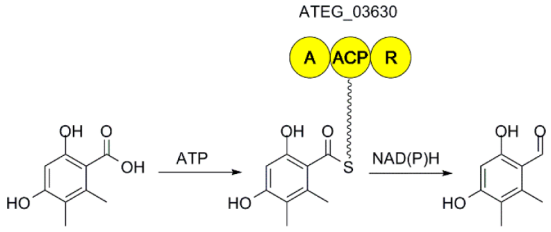


Figure 1. Mechanism of the NRPS-like protein-catalyzed reduction of 5-MOA.

the code-conferring residues of fungal NRPS A domains for aryl acids are less thoroughly studied. Moreover, 5-MOA is a novel substrate for any known A domain and cannot be predicted by current bioinformatics tools.[18] Therefore, the determination and characterization of the specificity-conferring code of ATEG_03630 is necessary for further genome mining of related biosynthetic pathways. In addition, it would lay the ground for substrate specificity engineering of fungal A domains.

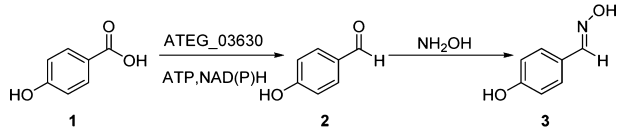
Traditionally, A domain characterization and engineering relies on two methods: (1) ATP-pyrophosphate (ATP-PPi) exchange assays using purified A domain proteins[19] and (2) genetic modification of A domains followed by characterization of the final NRP products.[10] However, the former method characterizes only the first half-reaction catalyzed by an A domain (i.e., the acyl-adenylate formation reaction), but not the second half-reaction (i.e., the thioesterification reaction involving an ACP domain). Therefore, an apparent active substrate in this assay may form only an acyl adenylate intermediate that cannot be transferred to the ACP domain for subsequent reactions.[11] The second method is more comprehensive and usually leads to novel modified natural products.[10,12] However, because of the complex substrate compatibility issue of multiple rounds of condensation in an NRPS assembly line, this method has met with limited success. Here, we consider the NRPS-like protein as a better model system for A domain engineering. Compared with multimodule NRPSs, it is a much smaller protein (~120 kDa), which is more convenient for genetic manipulation, heterologous expression, and protein purification. In addition, an aldehyde instead of an acyl adenylate is the final product, which can circumvent the above-mentioned inherent disadvantage of ATP-PPi exchange assays.

In the present study, 20 different substituted aryl acids were used to investigate the substrate specificity of the NRPS-like protein ATEG_03630. Combined with multiple sequence alignment, the specificity-conferring code for the novel substrate 5-MOA is established. Through protein engineering, the substrate specificity of the A domain is significantly altered toward our target substrate, anthranillic acid. Finally, as proof of concept, a one-pot chemoenzymatic synthesis of aldoxime from in-situ-generated aldehyde catalyzed by the NRPS-like protein is demonstrated (Scheme 1).

RESULTS AND DISCUSSION

Characterization of the Substrate Specificity of ATEG_03630. To the best of our knowledge, no previously discovered aryl acid substrates of A domains contain a methyl

Scheme 1. One-Pot Chemoenzymatic Synthesis of Aldoxime from 4-Hydroxybenzoic Acid



substituent, which makes the A domain of ATEG_03630 a unique opportunity to study the effect of the methylation pattern on substrate specificity of an A domain. We determined the specific activity of ATEG_03630 toward 5-MOA, 3-MOA, orsellinic acid, and 2,4-dihydroxybenzoic acid. As shown in Table 1, when the 5-methyl group was removed, the specific

Table 1. Specific Activities of WT ATEG_03630 with Various Aryl Acids[a]

substrate	specific activity (μmol/h/mg)	relative activity (%)
5-methyl orsellinic acid (**4**)	4.56 ± 0.21	100
3-methyl orsellinic acid (**5**)	1.15 ± 0.00	25
orsellinic acid (**6**)	1.32 ± 0.01	29
2,4-dihydroxybenzoic acid (**7**)	1.66 ± 0.09	36
salicylic acid (**8**)	0.0331 ± 0.0014	0.7
4-hydroxybenzoic acid (**9**)	1.98 ± 0.023	43
4-methoxybenzoic acid (**10**)	0.165 ± 0.006	4
protocatechuic acid (**11**)	0.664 ± 0.006	15
vanillic acid (**12**)	0.263 ± 0.004	6
isovanillic acid (**13**)	0.0340 ± 0.0005	0.7
4-amino-2-chlorobenzoic acid (**14**)	0.439 ± 0.008	10
anthranilic acid (**15**)	0.00685 ± 0.00140	0.2
2-nitrobenzoic acid (**16**)	N.D.[b]	0
3-nitrobenzoic acid (**17**)	N.D.	0
4-nitrobenzoic acid (**18**)	N.D.	0
4-amino-3-nitrobenzoic acid (**19**)	trace[c]	trace[c]
phthalic acid (**20**)	N.D.	0
4-aminophenylacetic acid (**21**)	N.D.	0
3,4-diaminobenzoic acid (**22**)	N.D.	0
4-butoxybenzoic acid (**23**)	N.D.	0

[a]See Supporting Information Figure S2 for the chemical structures of tested substrates. [b]N.D.: Not detected by HPLC/MS. [c]Only detected by MS after 24 h incubation; not quantifiable.

activity decreased by ~4.5-fold. It is likely that the 5-methyl group provides substrate selectivity via hydrophobic interactions with the A domain. However, methyl substituents at the 3- or 6-positions have no effect on the specific activity.

Hydroxyl substituents of aryl acids are known to be crucial for substrate specificity because of their ability to form hydrogen bonds with critical residues of the A domains.[20] Interestingly, 2-hydroxyl and 4-hydroxyl groups are not equally important for the substrate recognition of ATEG_03630. When the 2-hydroxyl moiety was removed, the specific activity toward 4-hydroxybenzoic acid was actually slightly increased compared with that of 2,4-dihyxroxy benzoic acid; however, the activity of ATEG_03630 dropped by 143-fold when salicylic acid was used as a substrate. These results suggest that the 4-hydroxyl group provides critical substrate recognition, likely via hydrogen-bonding, and the 2-hydroxyl group has no interaction with the A domain. The cocrystal structure of DhbE and its substrate 2,3-dihydroxybenzoic acid (DHB) reveals that hydroxyl groups

dx.doi.org/10.1021/cs500039v | ACS Catal. 2014, 4, 1219−1225

Table 2. Comparison of the 10-Residue Specificity Code for Selected NRPS Adenylation Domains

		NRPS code									
name/locus	substrate[a]	235 (263)[b]	236 (264)	239 (267)	278 (307)	299 (334)	301 (336)	322 (358)	330 (366)	331 (367)	517 (540)
GrsA−PheA	L-Phe	D	A	W	L	I	A	A	T	I	K
ATEG_03630	5-MOA	G	F	V	T	L	G	H	P	L	S
AN_6444	5-MOA(?)	G	F	V	T	T	G	H	P	L	S
PMAA_062890	5-MOA(?)	G	F	L	T	V	G	H	P	L	T
ATEG_07380	5-MOA(?)	G	F	L	T	A	G	H	A	I	S
consensus[c]	5-MOA	G	F	V/L	T	X_h	G	H	P/A	I/L	S/T
AnaPS_A1	Ant	G	A	L	I	I	A	A	G	V	K
NFIA_057960_A1	Ant	G	I	I	M	G	A	A	G	I	K
AFUA_6g12080_A1	Ant	G	V	I	L	L	A	A	G	I	K
NFIA_043670_A1	Ant(?)	G	M	I	L	V	A	A	G	I	K
ACLA_017890_A1	Ant(?)	G	V	I	V	L	A	A	G	V	K
ACLA_095980_A1	Ant(?)	G	V	I	M	L	A	G	G	L	K
ATEG_07358	Ant(?)	G	I	I	I	F	G	V	G	V	K
consensus	Ant	G	X_h	I/L	X_h	X_h	A/G	A/G	G	X_h	K

[a]Abbreviations:5-MOA, 5-methyl orsellinic acid; Ant, anthranilic acid; "(?)" indicates putative substrate. [b]Residue positions are numbered according to GrsA−PheA. Numbers in parentheses indicate residues numbered according to ATEG_03630. [c]The abbreviation "Xh" stands for variable hydrophobic residues.

of DHB are hydrogen-bonded to S240Oγ (3′-OH) and N235Nδ2 (bivalent hydrogen bonds to 2′- and 3′-OH).[20] Multiple sequence alignment (Table 2, Figure S1 and Table S2) indicates that G263 and V267 of ATEG_03630 occupied the same positions as N235 and S240 of DhbE, respectively. Therefore, the lack of appropriate hydrogen-bonding partners on the side chains of G263 and V267 is consistent with the conclusion that the 2-hydroxyl group has no effect on substrate recognition. To determine whether the 4-hydroxyl group acts as a hydrogen-bond donor or acceptor, we used 4-methoxybenzoic acid as a substrate. Approximately 11-fold decrease in specific activity was observed, which indicates that a hydrogen-bond donor is likely required at the 4-position. This is further supported by the fact that 4-amino-2-chlorobenzoic acid can be converted to an aldehyde with only 4-fold decrease in specific activity compared with 4-hydroxybenzoic acid.

Substituted aryl aldehydes are important fine chemicals for large-scale industrial applications. For example, vanillin (4-hydroxy-3-methoxybenzaldehyde), the main component of vanilla flavor,[21] is the global leader in aroma compounds. The search for an alternative route to chemically synthesize vanillin via bioconversion[22,23] or de novo biosynthesis[24] is inspired by significant financial incentives.[25,26] De novo biosynthesis of vanillin is of particular interest because the starting material can be either glucose or even cellulosic biomass. It has been shown that primary metabolites can be diverted from the shikimic acid pathway toward vanillin biosynthesis.[24] A critical step in the de novo pathway is to convert vanillic acid to vanillin by an aromatic carboxylic acid reductase (ACAR).[27,28] So far, ACAR from *Nocardia* sp. is the only viable protein.[24,29] In addition, because of the promiscuity of ACAR, isovanillin is suspected to be an undesired byproduct from the engineered pathway.[29] Because of the structural similarity between vanillic acid and 4-hydroxybenzoic acid, we set out to investigate whether vanillic acid is a substrate for ATEG_03630. As a result, vanillic acid and its precursor, protocatechuic acid, can be efficiently converted to aldehyde products (Table 1). Moreover, there is a ∼8-fold difference in specific activity between vanillic acid and isovanillic acid. Therefore, ATEG_03630 may be a suitable or even superior alternative to ACAR in vanillin de novo biosynthesis.

Engineering the Substrate Specificity of ATEG_03630 toward Anthranilate. Anthranilate, a primary metabolite in tryptophan biosynthesis, is a very rare monomer in bacterial NRPs or polyketides.[30,31] However, it is quite common in fungal secondary metabolites, including asperlicin[32] and acetylaszonalenin.[33] Its planar 1,3-substituted carboxy and amino group are critical for the formation of the benzodiazepinone and the quinazolinone core structures. The anthranilate activating A domain and its specificity-conferring code have been proposed.[34] Our study indicates that anthranilate is a very poor substrate for ATEG_03630 (Table 1) with 666-fold lower activity compared with the native substrate. It would be intriguing to see whether the substrate specificity of ATEG_03630 can be enhanced for anthranilate. The substrate specificity of DhbE, a stand-alone A domain from the siderophore bacillibactin biosynthesis pathway, has been engineered from 2,3-dihydroxybenzoic acid toward anthranilate with a 206-fold specificity switch.[11] It was achieved via 10 rounds of directed evolution based on the binding affinity of the A domain with a surrogate intermediate (acyl-adenosine monosulfate). Although directed evolution is a powerful high-throughput protein engineering method, the same drawback can be observed as that of ATP−PPi exchange assays because only the adenylation half-reaction is used in such engineering strategy. As a result, a universal His to Trp mutation in all isolated positive clones had to be mutated back to His to restore their ability to catalyze the second half thioesterification reaction. On the contrary, ATEG_03630 produces aldehydes as final products. Therefore, our engineered results would reflect both the adenylation and the thioesterification half reactions of the A domain.

Because ATEG_03630 is a newly discovered protein with novel substrate specificity,[18] we first performed multiple sequence alignment of ATEG_03630 with known anthranilate activating A domains as well as other A domains with similar substrate specificity. The 10-residue specificity-conferring codes are summarized in Table 2 and Supporting Information Table S2. Both 5-MOA and anthranilate activating A domains are

dx.doi.org/10.1021/cs500039v | ACS Catal. 2014, 4, 1219−1225

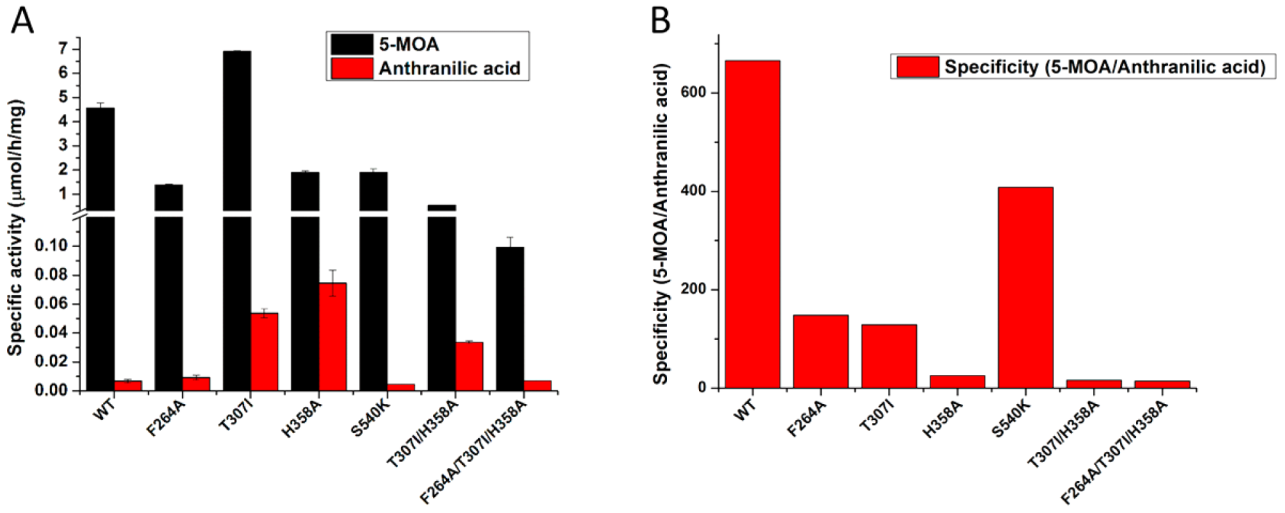


Figure 2. Substrate specificity engineering of the A domain. (A) Specific activities of the WT and mutant ATEG_03630 toward the native substrate 5-MOA and the target substrate anthranilate. (B) Substrate specificity of the engineered ATEG_03630 mutants (specificity is defined as the ratio of the specific activity of 5-MOA versus that of anthranilic acid).

highly conserved at positions 235 and 301 (positions are numbered according to GrsA-PheA). In addition, both classes of A domains contain variable hydrophobic residues at positions 239, 299, and 331. However, at position 236, the 5-MOA-activating A domains have exclusively a bulky aromatic amino acid phenylalanine, and the anthranilate activating A domains have a highly variable hydrophobic residue. At positions 278 and 322, the exclusively conserved tyrosine and histidine may provide critical interactions with the native substrate 5-MOA. Proline is highly enriched at position 330 compared with glycine in the anthranilate activating A domains. Finally, it is very interesting that the 5-MOA activating A domains have either serine or threonine, whereas all other A domains (including bacterial or fungal origin) have a conserved lysine at position 517.

On the basis of our bioinformatics analysis, we first decided to generate mutants F294A, T307I, H358A, P366G, and S540K (mutants are numbered according to ATEG_03630). Each purified protein was tested against both 5-MOA and anthranilate to determine its specific activity. First, mutant P366G showed no detectable activity toward 5-MOA or anthranilate. In addition, the protein yield of P366G was relatively low compared with that of other mutants. Therefore, it is likely that mutating proline to glycine may destabilize the protein and render it inactive. Surprisingly, mutating T307 to isoleucine increased its activity toward the native substrate 5-MOA by ~1.5-fold (Figure 2A), which indicates that T307 may not be involved in any direct interactions with 5-MOA. However, the specific activity of T307I toward anthranilate was also improved 7.8-fold, which resulted in a moderate 5.2-fold improvement in specificity (Figure 2B). Therefore, T307I may increase the catalytic activity of ATEG_03630 toward all substrates. Mutant F264A has a quite moderate effect on specific activity. The combination of a 3.3-fold decrease in specific activity to 5-MOA with an only 1.4-fold increase in specific activity to anthranilate gave a moderate 4.5-fold improvement in specificity. The most significant specificity improvement came from mutant H358A. Its specific activity with 5-MOA was decreased by 2.4-fold, and its activity to anthranilate was improved by 10.9-fold. Therefore, its specificity was increased by 26-fold. In all previously reported A domains, lysine is absolutely conserved at position 517. The crystal structures of PheA and DhbE revealed that lysine 517 provided key interactions with the carboxylate groups of the amino acid/aryl acid substrates as well as the adenosine moiety.[20,35] Surprisingly, when we mutated the distinct S540, which occupies position 517, to lysine, only a very moderate activity decrease for both 5-MOA and anthranilate was observed.

To shed some light on the significance of the conserved S540 in the 5-MOA activating A domains compared with the exclusive lysine in the other A domains, we attempted to build a homology model of the A domain of ATEG_03630. Because the crystal structure of aryl acid activating A domain from fungi is not available, a homology model was created using DhbE (PDB ID: 1MDB) as the template.[20] We chose DhbE as the template because it can activate an aryl acid instead of an amino acid. DhbE and the A domain of ATEG_03630 share only 22.3% sequence identity. However, a BLAST search indicated that they have better sequence homology (27% identity/36% positive) in the region from S220 to L367 (residue number according to ATEG_03630). Notably, this region contains the first nine code-conferring residues. Only the position 517 is not included. This suggests that the majority of their substrate-binding pockets may be similar to each other. Superposition of DhbE with the A domain of ATEG_03630 suggested that their overall structures are similar (Supporting Information Figure S3). After superposition, the ligand DHB adenylate from the DhbE crystal structure became out of position and clashed with the homology model. Therefore, we modified the ligand to 5-MOA adenylate, followed by energy minimization to investigate the potential interactions between the structural model and the 5-MOA adenylate intermediate.

Interestingly, albeit the relative low quality of the model due to the low sequence homology, some of the predicted protein–ligand interactions are consistent with our in vitro characterization (Figure 3). For example, we concluded that the 4-hydroxyl group of 5-MOA is critical for substrate recognition and the 2-hydroxyl group does not contribute to substrate specificity. It is consistent with the homology modeling results, which only predicted a strong hydrogen-bonding between the 4-hydroxyl group and the backbone carbonyl group of A268; no

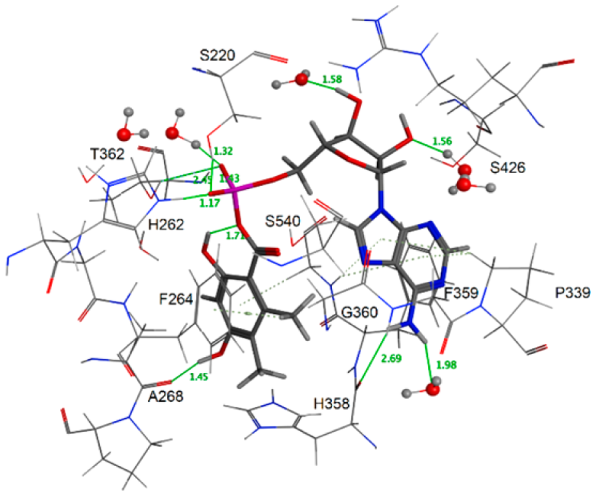


Figure 3. Ligand−protein interactions predicted in the homology model of the A domain of ATEG_03630. Predicted hydrogen bonds are shown in green solid lines with proton−oxygen distance (Å). Predicted hydrophobic interactions are shown in dashed lines.

protein−ligand interaction was predicted for the 2-hydroxyl group. F264A was predicted to have hydrophobic interaction with the 6-methyl group of the 5-MOA moiety, which may explain the moderate substrate specificity change of the mutant F264A; however, the homology model predicted that the backbone carbonyl group of H358 only had strong hydrogen-bonding with the adenylate moiety, which cannot explain the dramatic specificity switch of the mutant H358A. It is likely the low sequence homology rendered the homology model inaccurate. Finally, the model suggested that, without the extended side chain like in lysine, S540 is not involved in any interaction with the ligand. Therefore, the significance of the conserved S540 in the 5-MOA-activating A domains need to be further investigated. For instance, the cocrystal structure of ATEG_03630 with 5-MOA may provide a clearer picture.

Because three single mutants, F264A, T307I, and H358A, have some effects on substrate specificity, we set out to construct the double mutant T307I/H358A and the triple mutant F264A/T307I/H358A to determine whether there is a synergistic effect. Intriguingly, the double and triple mutants have significant synergistic effects on decreasing their activities toward the native substrate 5-MOA. Their specific activities to 5-MOA dropped by 8.5-fold and 45.9-fold, respectively. Unfortunately, double mutant T307I/H358A also decreased its activity with anthranilate by 2.2-fold compared with the best parent H358A. In addition, the specific activity of triple mutant F264A/T307I/H358A with anthranilate returned almost to the wild-type enzyme level. Although the double and triple mutants were not as efficient as H358A, they achieved highest specificity switch by 42- and 47-fold, respectively.

Previously, all A domain engineering studies were performed with bacterial A domains.[9−14] In addition, most studies relied on the ATP-PPi exchange assay for substrate specificity characterization or as a screening assay for the directed evolution strategy.[13] However, in this study, we used a novel fungal NRPS-like protein as a template for A domain engineering. Its relatively small size renders it more accessible for manipulation compared with multidomain NRPS mega-enzymes. More importantly, unlike the ATP-PPi exchange assay, we rely on the detection of aldehyde products to quantify engineering results, which represents both the adenylation and thioesterification reactions. In addition, aldehydes are highly active compounds for various chemical reactions that have been developed as screening assays for directed evolution.[36,37] Therefore, we consider NRPS-like proteins such as ATEG_03630 as a better model system for A domain engineering.

One-Pot Chemoenzymatic Synthesis of Aldoxime. In addition to being fine chemicals themselves, substituted aryl aldehydes are also important intermediates for organic synthesis and industrial applications.[38−40] For instance, aldoximes synthesized from aryl aldehydes have various applications in analytical chemistry, industry, and medicine.[34,41] They can serve as versatile precursors for the synthesis of amides,[42] nitrile oxides,[43] and amines.[44] Nonetheless, because of the low stability of aldehydes, it is difficult to prepare aldehydes in high yields.[45] Therefore, one-pot synthesis of aldoximes from in situ enzymatically generated aldehydes has numerous advantages.[46] It is more efficient because aldehyde isolations are not necessary. It is also more environmentally friendly because the reactions are performed in mild buffer systems without heavy metals.

As proof of concept, we selected 4-hydroxybenzoic acid as the substrate for the one-pot chemoenzymatic synthesis of 4-hydroxybenzaldoxime. Initial test indicated that the enzymatic reduction of 4-hydroxybenzoic acid to 4-hydroxybenzaldehyde is much faster than the subsequent chemical conversion to 4-hydroxybenzaldoxime. Therefore, we optimized two parameters, including the hydroxylamine concentration and the reaction time, to strike a balance between yield and productivity. As shown in Table 3, when 1.5 equiv of hydroxylamine was used, the yield of 4-hydroxybenzaldoxime reached 94.4% in 24 h. However, when 10 equiv of hydroxylamine was used, a yield of as high as 91.6% was obtained within 8 h. It is interesting that 10 mM hydroxylamine seems to have no detrimental effect on ATEG_03630-catalyzed enzymatic reduction, but it was reported that hydroxylamine can be toxic to enzymatic reactions.[46,47]

Table 3. Optimization of the Preparation of 4-Hydroxybenzaldoxime

entry	substrate (mM)	equiv NH_2OH	reaction time (h)	yield **3** (%)
1	1	1.5	8	35.1
2	1	1.5	24	94.4
3	1	10	6	85.3
4	1	10	8	91.6
5	1	10	24	98.3

CONCLUSION

We characterized the substrate specificity of WT ATEG_03630 protein and concluded that the hydrogen bond donor at the 4-position is crucial for substrate recognition. In addition, we demonstrated that it could be a superior candidate for de novo biosynthesis of vanillin. We have shown that the substrate specificity of ATEG_03630 can be engineered toward our target substrate anthranilate via bioinformatics analysis and mutagenesis. As a result, mutant H358A increased its activity toward anthranilate by 10.9-fold, which led to a 26-fold improvement in specificity. NRPS-like proteins are much smaller than multidomain NRPSs, but they still represent the thioesterification half-reaction, which is otherwise missed from the stand-alone A domains. Therefore, the NRPS-like protein

dx.doi.org/10.1021/cs500039v | ACS Catal. 2014, 4, 1219−1225

may serve as a better model system for A domain engineering. Finally, we demonstrated the one-pot chemoenzymatic synthesis of 4-hydroxybenzaldoxime from 4-hydroxybenzoic acid with high yield. Combined with the above-mentioned amenability to substrate specificity engineering, the NRPS-like protein could serve as a promising biocatalyst for in situ aldehyde generation for multistep one-pot synthesis.

EXPERIMENTAL PROCEDURES

Materials and Reagents. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Failsafe 2× PreMix buffer G was purchased from Epicentre Biotechnologies (Madison, WI). Synthetic complete drop-out medium lacking uracil (SC-Ura) from MP Biomedicals (Solon, OH) was used to select *Saccharomyces cerevisiae* transformants containing the assembled plasmids. *S. cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1 Δ1.6R can1 GAL*) was used as the host for DNA assembly and protein overexpression.

Plasmid Construction. Cloning of ATEG_03630 is described previously.[18] Mutants were constructed by the same method using primers listed in Supporting Information Table S1.

Protein Expression and Purification from *S. cerevisiae*. Each expression vector was transformed to *S. cerevisiae* strain BJ5464-NpgA using the lithium acetate/single stranded carrier DNA/polyethylene glycol (PEG) method.[48] Transformants were grown in 20 mL of SC-Ura dropout media for 2 days and inoculated to 1 L of YPAD medium. Cells were grown at 30 °C and 250 rpm for 72 h. The cells were harvested by centrifugation (4000*g*, 15 min, 4 °C), resuspended in 30 mL lysis buffer (50 mM NaH_2PO_4, pH 8.0, 0.15 M NaCl, 10 mM imidazole), and lysed by sonication on ice. His-tagged proteins were purified by using Ni-NTA agarose (QIAGEN, Valencia, CA) following the supplied protocols. The cleared cell lysate following sonication and centrifugation was directly applied onto a column that was packed with Ni-NTA agarose. After washing with washing buffer containing 20 mM imidazole, the protein was eluted with the elution buffer containing 250 mM imidazole. Purified proteins were concentrated and buffer-exchanged into a storage buffer (50 mM Tris−HCl, 100 mM NaCl, pH = 7.9) containing 10% glycerol. The concentrated enzyme solutions were aliquoted and flash-frozen. Protein concentrations were determined with the Bradford assay (BioRad) using BSA as a standard.

In Vitro Characterization of ATEG_03630 and Its Mutants. For in vitro enzymatic assays, the final concentrations of enzymes were 1 μM, 1 mM various aryl acid substrates, with cofactor concentrations of 2 mM NADPH, 10 mM ATP and 10 mM $MgCl_2$. The assays were carried out in 50 mM Tris−HCl buffer (pH 8.5). The reaction mixtures were incubated at room temperature.

A typical volume of the reaction is 200 μL. Twenty microliters of reaction mixture was taken out at various time points and quenched with HCl. Each reaction mixture was used for HPLC−ESI-MS analysis. HPLC−ESI-MS was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA) with a reversed-phase Kinetex C18 column (Phenomenex, Torrance, CA). General HPLC parameters were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 25 min, maintain at 100% B for 10 min, return and maintain at 0% B for 7 min; flow rate 0.3 mL/min; detection by UV spectroscopy at 330 nm. The amounts of aldehyde products were quantified by area integration of the UV peak at 330 nm. A standard curve was generated using pure aldehyde standards with the same HPLC conditions. The HPLC parameters for detecting 2-amino benzaldehyde were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 60 min, maintain at 100% B for 10 min, return and maintain at 0% B for 7 min; flow rate 0.3 mL/min.

Homology Modeling. The cocrystal structure of DhbE with DHB adenylate (PDB ID: 1MDB)[20] was used as the template for homology modeling. The homology model of the A domain of ATEG_03630 was constructed using the modeling program Molecular Operating Environment (Chemical Computing Group, Montreal, Quebec, Canada). The homology model was superimposed with the cocrystal structure of DhbE and DHB adenylate. The ligand DHB adenylate was then modified to 5-MOA adenylate. After removing the crystal structure of DhbE, an energy minimization of the model and the ligand was performed before the ligand−protein interactions were investigated.

One-Pot Chemoenzymatic Synthesis of Aldoxime. For each in vitro reaction, the final concentrations of enzymes were 1 μM and 1 mM 4-hydroxybenzoic acid, with cofactor concentrations as 2 mM NADPH, 10 mM ATP, 10 mM $MgCl_2$, 1.5 mM or 10 mM NH_2OH. The assays were carried out in 50 mM Tris−HCl buffer (pH 8.5). The reaction mixtures were incubated at room temperature.

A typical volume of the reaction is 200 μL. The reaction mixture was quenched with MeOH. Each reaction mixture was used for HPLC−ESI-MS analysis. The HPLC parameters for detecting aldoxime were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 30% B in 25 min, increase and maintain at 100% B for 10 min, return and maintain at 0% B for 7 min; flow rate 0.3 mL/min; detection by UV spectroscopy at 300 nm. The amount of aldoxime was quantified by area integration of the UV peak at 300 nm. A standard curve was generated using pure 4-hydroxybenzaldoxime standards with the same HPLC condition. See the Supporting Information for the preparation of 4-hydroxybenzaldoxime standards.

ASSOCIATED CONTENT

Supporting Information

Additional tables, figures, experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: zhao5@illinois.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Academies Keck Futures Initiative on Synthetic Biology and National Institutes of Health (GM077596). We thank Professor Y. Tang for providing *S. cerevisiae* strain BJ5464-NpgA.

REFERENCES

(1) Hur, G. H.; Vickery, C. R.; Burkart, M. D. *Nat. Prod. Rep.* **2012**, *29*, 1074.
(2) Lawen, A.; Traber, R. *J. Biol. Chem.* **1993**, *268*, 20452.

(3) Stachelhaus, T.; Schneider, A.; Marahiel, M. A. *Science* **1995**, *269*, 69.
(4) Doekel, S.; Marahiel, M. A. *Chem. Biol.* **2000**, *7*, 373.
(5) Hahn, M.; Stachelhaus, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15585.
(6) Doekel, S.; Coeffet-Le Gal, M. F.; Gu, J. Q.; Chu, M.; Baltz, R. H.; Brian, P. *Microbiology* **2008**, *154*, 2872.
(7) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211.
(8) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493.
(9) Eppelmann, K.; Stachelhaus, T.; Marahiel, M. A. *Biochemistry* **2002**, *41*, 9718.
(10) Thirlway, J.; Lewis, R.; Nunns, L.; Al Nakeeb, M.; Styles, M.; Struck, A. W.; Smith, C. P.; Micklefield, J. *Angew. Chem., Int. Ed.* **2012**, *51*, 7181.
(11) Zhang, K.; Nelson, K. M.; Bhuripanyo, K.; Grimes, K. D.; Zhao, B.; Aldrich, C. C.; Yin, J. *Chem. Biol.* **2013**, *20*, 92.
(12) Evans, B. S.; Chen, Y.; Metcalf, W. W.; Zhao, H.; Kelleher, N. L. *Chem. Biol.* **2011**, *18*, 601.
(13) Villiers, B.; Hollfelder, F. *Chem. Biol.* **2011**, *18*, 1290.
(14) Chen, C. Y.; Georgiev, I.; Anderson, A. C.; Donald, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3764.
(15) Sattely, E. S.; Fischbach, M. A.; Walsh, C. T. *Nat. Prod. Rep.* **2008**, *25*, 757.
(16) Crosa, J. H.; Walsh, C. T. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 223.
(17) Huang, T.; Wang, Y.; Yin, J.; Du, Y.; Tao, M.; Xu, J.; Chen, W.; Lin, S.; Deng, Z. *J. Biol. Chem.* **2011**, *286*, 20648.
(18) Wang, M.; Beissner, M.; Zhao, H. *Chem. Biol.* **2014**, *21*, 1.
(19) Lee, S. G.; Lipmann, F. *Methods Enzymol.* **1975**, *43*, 585.
(20) May, J. J.; Kessler, N.; Marahiel, M. A.; Stubbs, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12120.
(21) Gobley, M. *Recherches sur le principe odorant de la vanille*; Thunot: Paris, 1858.
(22) Barghini, P.; Di Gioia, D.; Fava, F.; Ruzzi, M. *Microb. Cell Fact.* **2007**, *6*, 13.
(23) Overhage, J.; Steinbuchel, A.; Priefert, H. *Appl. Environ. Microbiol.* **2003**, *69*, 6569.
(24) Hansen, E. H.; Moller, B. L.; Kock, G. R.; Bunner, C. M.; Kristensen, C.; Jensen, O. R.; Okkels, F. T.; Olsen, C. E.; Motawia, M. S.; Hansen, J. *Appl. Environ. Microbiol.* **2009**, *75*, 2765.
(25) Ramachandra Rao, S.; Ravishankar, G. A. *J. Sci. Food Agric.* **2000**, *80*, 289.
(26) Walton, N. J.; Mayer, M. J.; Narbad, A. *Phytochemistry* **2003**, *63*, 505.
(27) Li, T.; Rosazza, J. P. *J. Bacteriol.* **1997**, *179*, 3482.
(28) Venkitasubramanian, P.; Daniels, L.; Rosazza, J. P. *J. Biol. Chem.* **2007**, *282*, 478.
(29) Brochado, A. R.; Patil, K. R. *Biotechnol. Bioeng.* **2013**, *110*, 656.
(30) Sandmann, A.; Dickschat, J.; Jenke-Kodama, H.; Kunze, B.; Dittmann, E.; Muller, R. *Angew. Chem., Int. Ed.* **2007**, *46*, 2712.
(31) Pistorius, D.; Li, Y.; Mann, S.; Muller, R. *J. Am. Chem. Soc.* **2011**, *133*, 12362.
(32) Chang, R. S.; Lotti, V. J.; Monaghan, R. L.; Birnbaum, J.; Stapley, E. O.; Goetz, M. A.; Albers-Schonberg, G.; Patchett, A. A.; Liesch, J. M.; Hensens, O. D.; et al. *Science* **1985**, *230*, 177.
(33) Yin, W. B.; Grundmann, A.; Cheng, J.; Li, S. M. *J. Biol. Chem.* **2009**, *284*, 100.
(34) Ames, B. D.; Walsh, C. T. *Biochemistry* **2010**, *49*, 3351.
(35) Conti, E.; Stachelhaus, T.; Marahiel, M. A.; Brick, P. *EMBO J.* **1997**, *16*, 4174.
(36) Choi, K. Y.; Jung, E. O.; Yun, H.; Yang, Y. H.; Kazlauskas, R. J.; Kim, B. G. *ChemBioChem* **2013**, *14*, 1231.
(37) Rannes, J. B.; Ioannou, A.; Willies, S. C.; Grogan, G.; Behrens, C.; Flitsch, S. L.; Turner, N. J. *J. Am. Chem. Soc.* **2011**, *133*, 8436.
(38) Mallat, T.; Baiker, A. *Chem. Rev.* **2004**, *104*, 3037.
(39) Shanmuganathan, S.; Natalia, D.; Greiner, L.; Dominguez de Maria, P. *Green Chem.* **2012**, *14*, 94.
(40) Yang, J. W.; Chandler, C.; Stadler, M.; Kampen, D.; List, B. *Nature* **2008**, *452*, 453.
(41) Raptopoulou, C. P.; Psycharis, V. *Inorg. Chem. Commun.* **2008**, *11*, 1194.
(42) Allen, C. L.; Davulcu, S.; Williams, J. M. *Org. Lett.* **2010**, *12*, 5096.
(43) Yoshimura, A.; Middleton, K. R.; Todora, A. D.; Kastern, B. J.; Koski, S. R.; Maskaev, A. V.; Zhdankin, V. V. *Org. Lett.* **2013**, *15*, 4010.
(44) Bandgar, B. P.; Nikat, S. M.; Wadgaonkar, P. P. *Synth. Commun.* **1995**, *25*, 863.
(45) Kanno, H.; Taylor, R. J. K. *Synlett* **2002**, 1287.
(46) Zambelli, P.; Pinto, A.; Romano, D.; Crotti, E.; Conti, P.; Tamborini, L.; Villa, R.; Molinari, F. *Green Chem.* **2012**, *14*, 2158.
(47) Gross, P. *Crit. Rev. Toxicol.* **1985**, *14*, 87.
(48) Gietz, R. D.; Woods, R. A. *Methods Enzymol.* **2002**, *350*, 87.