

Exploiting genetic diversity by directed evolution: molecular breeding of type III polyketide synthases improves productivity† ‡

Wenjuan Zha,§ Sheryl B. Rubin-Pitel and Huimin Zhao*

Received 16th November 2007, Accepted 16th January 2008

First published as an Advance Article on the web 31st January 2008

DOI: 10.1039/b717705d

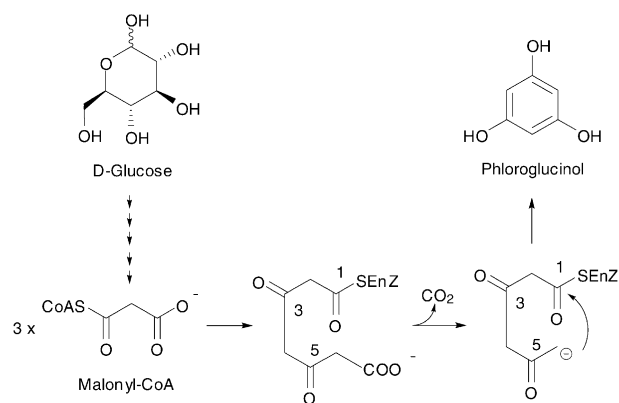
Applying directed evolution to the phloroglucinol synthase PhlD from *Pseudomonas fluorescens* Pf-5 has provided the first example of engineering enhanced productivity in a type III polyketide synthase, and a rare instance of improving the activity of a biosynthetic enzyme from secondary metabolism.

Type III polyketide synthases (PKSs) from bacteria, fungi, and plants produce an amazing array of biologically and medically important polyketides.^{1,2} These enzymes carry out complex polyketide synthesis in an assembly line fashion, which involves loading a starter Coenzyme A (CoA) thioester, iterative decarboxylation–condensation with a CoA extension unit, and cyclization of a linear chain into the final polyketide products. Similar to other enzymes involved in secondary metabolism, type III PKSs generally exhibit low productivity, *i.e.*, low activity and stability,^{2–4} thereby impeding their application as biosynthetic tools for expanding the diversity and practical synthesis of important polyketides. Protein engineering techniques are a promising means to bridge the functional gap between an enzyme's natural activity and the productivity goals for industrial application,⁵ and several examples of applying directed evolution techniques to modular, type I PKSs do exist.⁶ We chose as a target for improvement the enzyme PhlD, a type III PKS from the soil bacterium *Pseudomonas fluorescens* Pf-5. The recent characterization of PhlD as a phloroglucinol synthase⁷ creates an opportunity for biosynthesis of this compound, which is used widely in industry and medically as an antispasmodic therapy. Herein we report the first successful example of improving the productivity of type III PKS enzyme by applying directed evolution to increase the catalytic activity and thermostability of PhlD.

PhlD is a small, homodimeric protein (77 kDa), and catalyzes the synthesis of phloroglucinol from three molecules of malonyl-CoA (Scheme 1).^{7,8} We recently characterized the biochemical properties of this enzyme.⁸ It was found that non-hexahistidine tagged PhlD is unstable with a half-life of only 7.2 minutes at 37 °C, and has a low catalytic efficiency (0.56 $\mu\text{M min}^{-1}$) that is attributed largely to a low turnover

number (7.3 min^{-1}). Thus, we sought to improve the productivity of this enzyme using a recently developed directed evolution method, synthetic shuffling.⁹ In this strategy, the sequence information from a set of homologous parent genes is used to design a series of forward and reverse primers, which are subsequently assembled into the full-length gene. The genetic diversity from these parents is incorporated *via* degeneracy in the oligonucleotides, and recombines independently to create a library of chimeric progeny.⁹

A BLAST search (<http://www.ncbi.nlm.nih.gov/BLAST/>) using the PhlD gene from *P. fluorescens* strain Pf-5 as a probe identified 52 homologous PhlD proteins that share 79%–99% sequence identity. Multiple sequence alignment by Clustal W revealed a diversity of 2 amino acids difference at 30 positions, 3 amino acids difference at 4 positions, and 4 amino acids difference at 1 position, giving a total of 4×10^{19} possible combinations for synthetic shuffling (Fig. 1). In practice, two factors are crucial to the success of synthetic shuffling.⁹ The first factor concerns the quality of synthetic oligonucleotides. This led us to adopt a two-step PCR method¹⁰ to construct a PhlD synthetic shuffling library, which avoids the use of long oligonucleotides in order to achieve more accurate gene synthesis. Accordingly, 42 short forward and reverse oligonucleotides of 20–25 bases were designed to assemble the full length *phlD* gene and incorporate genetic diversity *via* degenerate oligonucleotides. The second crucial factor is the selection of parental diversity. Because deleterious mutations are over-represented in synthetic shuffling, they must be minimized. A close examination of the diversity from 52 homologous *phlD* sequences revealed several mutations present in



Scheme 1 Phloroglucinol synthesis catalyzed by PhlD. In *Escherichia coli* expressing PhlD, cellular metabolism produces malonyl-CoA substrate from glucose *via* the intermediates pyruvate and acetyl-CoA.

Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 S. Mathews Ave., Urbana, IL 61801, USA. E-mail: zhao5@uiuc.edu; Fax: (001) 217-333-5052

† Electronic supplementary information (ESI) available: experimental, supplementary figures and tables. See DOI: 10.1039/b717705d

‡ This research was supported by a grant from the Office of Naval Research (N00014-02-1-0725).

§ Present Address: 200 Penobscot Drive, Redwood City, CA 94063, USA

MSTLCLPHVMFPQHKITQQQMVDHLENLHADHPRMALAKRMIANTEVNERHLVLPIDELA 60
VHTGFTHRSIVYEREARQMSSAAARQAIENAGLQISDIRMVIIVTSCTGFMMPSLTAHLIN 120
R I TTD A P
L VN V
138
DLALPTSTVQLPIAQLGCVAGAAAINRANDFARLDARNHVLIVSLEFSSLCYQPDDTKLH 180
GMRA A D ARSPD A Q
D N SQA Q N
KM
AFISAALFGDAVSACVLRADDQAGGFKIKKTESYFLPKSEHYIKYDVKDTGFHF[†]TLDKAV 240
M N KVP AN G F D SD
A HS N N
272
MNSIKDVAPVMERLNYESFEQNCAHNDFFIFHTGGRKILDELVMHLLDLASNRVSQSRSSL 300
M E FDT N HR Q V T LQ EPG A D
I K V D T
305
SEAGNIASVVVFDVLKRQFDSNLNRGDIGLLAAFGPGFTAEMAVGEWTA 350

Fig. 1 Diversity encoded in 52 PhlD homologs. Amino acid alternatives colored in blue were incorporated into the synthetic shuffling library; those colored in green were presumably errors and thus were discarded. Residues colored in red are the catalytic triad (C138, H272, N305).

only one or two of the parental sequences (Fig. 1), which we speculated were likely the result of PCR or sequencing errors associated with the discovery of these genes rather than natural diversity. Exclusion of these mutations reduced the library size to 2×10^{11} , and more importantly, increased the active fraction of the resulting library from 0.1% to 20% (data not shown).

A two-tiered colorimetric screening system was developed to identify PhlD mutants with improved phloroglucinol production. Throughout the screening, a modified M9 minimal medium (supplemented with 10 g L^{-1} of tryptone) was used to reduce background reaction with colorimetric reagents. Further, the BL21 *lacY* deletion strain Tuner(DE3) was chosen for expression of the PhlD mutant library, to ensure a linear and homogenous response to IPTG inducer during screening. As shown in Supplemental Scheme S1[†], after transformation of the library, a membrane-based pre-screening was used to rapidly isolate active clones in a high throughput manner, based on the colorimetric reaction between Gibb's reagent (*N*,2,6-trichloro-*p*-benzoquinoneimide) and phenols.¹¹ All active mutants, indicated by the appearance of a purple halo around the colony, were picked directly from

the membrane into 96-well plates and subjected to more quantitative analysis. Phloroglucinol production in cell-free supernatant was quantified after induction using the specific and sensitive colorimetric reaction between cinnamaldehyde and phloroglucinol, a method adapted from the Weisner test, a classical assay in plant science which utilizes phloroglucinol for the histochemical staining of lignin.¹² Using this method we screened approximately 40 000 clones from the synthetic shuffling library. Two mutants, 23D9 and 27B5, were identified, which gave 4.2- and 7.8-fold more phloroglucinol production than wild type PhlD, respectively. The kinetic parameters of purified 23D9 and 27B5 were determined and are listed in Table 1 (first entry). Mutant 23D9 had more than 4-fold increase in k_{cat} over that of the wild type PhlD, and an essentially unchanged K_{M} , resulting in catalytic efficiency that was 4-fold higher than that of the wild type PhlD. Meanwhile, mutant 27B5 had only a 2-fold increase in k_{cat} , and a slight decrease in K_{M} compared with those of the wild type enzyme. As a result, 27B5 gave almost 3-fold improvement in catalytic efficiency. The thermostabilities of these two mutants were also examined, and were expressed as their half-lives of thermal inactivation ($t_{1/2}$) at 37 °C.

Table 1 Characterization of kinetic parameters and thermostability for PhlD wild type and mutants^a

Enzyme	Kinetics				$t_{1/2}$ at 37 °C (min)	$t_{1/2}$ at 37 °C Improvement
	k_{cat} (min ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (min ⁻¹ μM ⁻¹)	$k_{\text{cat}}/K_{\text{M}}$ Improvement		
WT	7.3 ± 0.2	13.1 ± 1.4	0.56	—	7.2	—
23D9	32.3 ± 1.0	13.3 ± 2.5	2.43	4.3	9	1.3
27B5	14.3 ± 0.5	12.0 ± 1.7	1.19	2.1	24	3.3
K210L	12.5 ± 0.9	11.9 ± 2.7	1.05	1.9	10	1.4
Y256R	11.9 ± 0.2	11.2 ± 0.7	1.06	1.9	10.2	1.4
A289R	7.8 ± 0.3	10.1 ± 1.5	0.77	1.4	20.8	2.9
K210L + Y256R	6.8 ± 0.3	13.5 ± 2.4	0.50	0.9	11.9	1.7
Y256R + A289R	4.7 ± 0.3	10.3 ± 2.9	0.45	0.8	13.5	1.9
K210L + Y256R	13.1 ± 0.5	6.9 ± 1.1	1.89	3.4	11.5	1.6

^a The kinetic parameters and thermal inactivation for PhlD enzyme were determined as described in ref. 7. See Supplementary Table S1[†] for a list of the amino acid and nucleotide substitutions found in mutants 23D9 and 27B5, and the nucleotide substitutions for mutants K210L, Y256R, and A289R.

Wild type PhlD has a half-life of only 7.2 min at 37 °C. The stability of 23D9 at 37 °C was comparable to that of the wild type enzyme, but 27B5 showed more than 3-fold improved stability at 37 °C.

DNA sequencing revealed that 23D9 and 27B5 have 14 and 10 amino acid changes, respectively (Supplemental Table S1†), which were further mapped onto a structural model of PhlD⁸ (Supplemental Fig. S2†). Surprisingly, all of the 17 different mutations were located on the surface of the protein, away from the active site. Due to the large number of mutations, it was difficult to decipher which mutations were responsible for the improved enzymatic activity and stability. Thus, we sought to evaluate the effect of individual mutations using saturation mutagenesis. Each of the 17 sites was randomly replaced with all 20 amino acids, and the resulting libraries were examined for phloroglucinol production. Three single mutants were identified which improved phloroglucinol production by approximately 4-fold. DNA sequencing revealed three distinct mutations, K210L (from both 23D9 and 27B5), Y256R (from 23D9), and A289R (from 27B5). Mutants K210L and Y256R showed almost 2-fold improvement in k_{cat} , while A289R had only moderate improvement compared with the k_{cat} of the wild type PhlD (Table 1, second entry). All three mutations had positive effects on the substrate binding, giving slightly lower K_{M} value than that of the wild type enzyme. As a consequence, K210L and Y256R improved catalytic efficiency by 2-fold, and A289R by 1.5-fold. As for thermostability, K210L and Y256R slightly increased the half-life of PhlD at 37 °C, but A289R made a significant improvement in the thermostability of PhlD by almost 3-fold.

Detailed examination of the single mutations K210L, Y256R, and A289R in the PhlD homology model shows that they are all located far away from the active site (Supplemental Fig. S2†). For example, mutation K210L, located on the N-terminus of β -sheet 8, is approximately 17 Å from the catalytic cysteine (C138), whereas mutation A289R, located on a solvent-exposed turn between helices 12 and 13, is more than 15 Å from the active site. As a result, the structural basis for the improved activity and thermostability remains unclear. It should be noted that in directed evolution, it has become a common theme to find mutations that are far from the active site of an enzyme, yet affect enzymatic activity.¹³

To further investigate whether these single mutations are additive, a series of double mutants were created by site-directed mutagenesis. The combination of K210L/A289R or Y256R/A289R slightly impaired the activity of PhlD, resulting in lower turnover number than either of the three single mutants, or even the wild type enzyme itself (Table 1, third entry). This implies a conflicting effect of A289R with the other two mutations. In contrast, the combination of Y256R and K210L improved not only the k_{cat} but also the K_{M} of PhlD to a greater extent than by the two single mutations alone, giving rise to more than 3-fold increase of the catalytic

efficiency. In terms of thermostability, none of the three double mutants showed additive improvement on the enzyme's half-life at 37 °C compared with the single mutants. Rather, the thermostability improvement from A289R alone was impaired by the addition of a second mutation, either K210L or Y256R. Nevertheless, the parental mutant 23D9 initially identified from synthetic shuffling library was still the most active PhlD variant with more than 4-fold improvement in catalytic efficiency, and mutant 27B5 gave the most improved thermostability at 37 °C.

In conclusion, we used directed evolution to improve the productivity of PhlD *via* synthetic shuffling of 52 homologous PhlD genes in combination with high-throughput screening. Two PhlD mutants were identified that increased phloroglucinol production as a result of improved activity and/or thermostability. Mutational analysis by saturation mutagenesis coupled with homology modeling revealed that subtle changes in the protein structure, brought about by single mutations far away from the active site, can significantly improve the catalysis or stability of PhlD. This work represents the first example of improving the productivity of a type III polyketide synthase, only the second successful application of synthetic shuffling, and a rare example of improving the activity of a biosynthetic enzyme involved in secondary metabolism.

Notes and references

- 1 M. B. Austin and J. P. Noel, *Nat. Prod. Rep.*, 2003, **20**, 79.
- 2 K. Watanabe, A. P. Praseuth and C. C. C. Wang, *Curr. Opin. Chem. Biol.*, 2007, **11**, 279.
- 3 B. S. Moore and J. N. Hopke, *ChemBioChem*, 2001, **2**, 35.
- 4 N. Funa, Y. Ohnishi, Y. Ebizuka and S. Horinouchi, *J. Biol. Chem.*, 2002, **277**, 4628.
- 5 (a) T. W. Johannes and H. Zhao, *Curr. Opin. Microbiol.*, 2006, **9**, 261; (b) S. B. Rubin-Pitel and H. Zhao, *Comb. Chem. High Throughput Screening*, 2006, **9**, 247; (c) H. Zhao, K. Chockalingam and Z. Chen, *Curr. Opin. Biotechnol.*, 2002, **13**, 104.
- 6 (a) B. S. Kim, D. H. Sherman and K. A. Reynolds, *Protein Eng. Des. Sel.*, 2004, **17**, 277–284; (b) H. M. O'Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer and P. F. Leadlay, *Chem. Biol.*, 2006, **13**, 277–285.
- 7 J. Achkar, M. Xian, H. Zhao and J. W. Frost, *J. Am. Chem. Soc.*, 2005, **127**, 5332.
- 8 W. Zha, S. B. Rubin-Pitel and H. Zhao, *J. Biol. Chem.*, 2006, **281**, 32036.
- 9 J. E. Ness, S. Kim, A. Gottman, R. Pak, A. Krebber, T. V. Borchert, S. Govindarajan, E. C. Mundorff and J. Minshall, *Nat. Biotechnol.*, 2002, **20**, 1251.
- 10 L. Young and Q. Dong, *Nucleic Acids Res.*, 2004, **32**, e59.
- 11 H. D. Gibbs, *J. Biol. Chem.*, 1927, **72**, 649.
- 12 (a) E. Adler, K. J. Bjorkkqvist and S. Haggroth, *Acta Chem. Scand.*, **2**, 93; (b) F. Pomer, F. Merino and A. Ros Barcelo, *Protoplasma*, 2002, **220**, 17.
- 13 (a) K. L. Morley and R. J. Kazlauskas, *Trends Biotechnol.*, 2005, **23**, 2310; (b) T. Yano, S. Oue and H. Kagamiyama, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5511; (c) R. Woodyer, W. A. van der Donk and H. Zhao, *Comb. Chem. High Throughput Screening*, 2006, **9**, 237.