Supporting Information for

Engineered production of hapalindole alkaloids in the cyanobacterium *Synechococcus* sp. UTEX 2973

Cory J. Knoot^a, Yogan Khatri^b, Robert M. Hohlman^b, David H. Sherman^b, and Himadri B. Pakrasi^a

^a Department of Biology, Washington University, St. Louis, Missouri, USA

^b Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

Supplementary Methods

Sequences of synthesized DNA fragments for pSL3228 assembly. The P_{J23119} and *lacl* W220F coding sequence were synthesized with terminal overlaps for Gibson assembly. Sequence:

gagcagaagaggctactaggtaggatcaaaaataatcaagagggcaaaaaactcaatttggttatcaaaaatgaagag gaagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcag accgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcgatggcgga gctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagtct ggccctgcacgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgat ggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcatt agacacccatcaacagtattattttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgcattgggtcacc aatcaaattcagccgatagcggaacgggaaggcgactttagtgccatgtccggttttcaacaaaccatgcaaatgctgaa tgagggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatgcgcgccattaccgagtccg ggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgttaacca ccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaactctctcaggggccaggcggtgaag cqttqqccqattcattaatqcaqctqqcacqacaqqtttcccqactqqaaaqcqqqcaqtqaaqatctaactaccqcatta aagctagacatggtacctgaattcgaccttgcaggag

2

The P_{trc2O} promoter was synthesized with the following sequence:

ttgatcctacctagtagcctcttctgctcctgcagaattgtgagcgctcacaattgatattttggtctgtcgttgcgatcgcccgtt gcaggccgacatgaaggattgacaattaatcatccggctcgtataatgaattgtgagcgctcacaattggtaccggtgata ccgctaaggaggtaagtgcaccatgattagtgaaaaaattctc

Extraction and detection of *L*-tryptophan. SL3228 cells were cultured in a photobioreactor at 38 °C under 500 μ E m⁻² s⁻¹ while bubbling with 5% CO₂-supplemented air. After three days of growth, the cells were harvested by centrifugation, washed with water, and cell pellets resuspended in 50 mM sodium acetate pH 4.8. The cells were then boiled for one hour and pelleted by centrifugation. The lysate was transferred to HPLC vials and analyzed using an Agilent 1200 series instrument equipped with a fluorescence detector. Extracts were loaded onto a 150 x 4.6 mm Phenomenex Luna 5 μ m C₁₈ column and separated using an isocratic mobile phase containing 85% 50 mM sodium acetate pH 4.8 and 15% acetonitrile. We tracked tryptophan fluorescence using an excitation wavelength of 287 nm and detection wavelength of 352 nm.

Supplementary Figures



Figure S1. Maps of RSF1010-type plasmids used to express *fam* genes in S2973. Genes are shown as arrows and genetic elements are labeled.



Figure S2. All three *famH* genes are necessary for the production of *cis* indole-isonitrile in S2973. The figure shows HPLC traces of organic S2973 extracts and standards monitored at 222 nm. (A) authentic *trans* and *cis* indole-isonitrile (**1**) standards. (B) S2973 expressing *famH1*, *famH2*, and *famH3* (SL3228). (C) S2973 expressing *famH2* and *famH3*. (D) S2973 expressing *famH1* and *famH3*. In B – D the *famH* operons are shown diagrammatically above the corresponding trace.



Figure S3. Extracted ion chromatograms (EICs) of S2973 extracts compared with authentic hapalindole standards. (A) Wild-type S2973 (*m/z* of 168). (B) SL3228 (*m/z* of 168). (C) SL3285 (*m/z* of 305). (D) SL3251 (*m/z* of 305). (E) Authentic hapalindole H (**3**, *m/z* of 305). (F) SL3283 (*m/z* of 305). (G) Authentic 12-*epi* hapalindole U (**4**, *m/z* of 305). (305).







Figure S5. *L*-tryptophan content in SL3228 is comparable in uninduced and induced cells. The figure shows HPLC traces of SL3228 lysate from cells induced with 0, 100, and 250 μ M IPTG and an *L*-tryptophan standard. Elution was monitored using excitation and detection wavelengths of 287 nm and 352 nm, respectively. The amount of extract loaded for each sample was normalized to dry cell weight.



Figure S6. Analysis of UV-spectra of C3-geranylated indole-isonitrile (**2**, RT at 7.7 min) and the C2-geranylated isomer (**5**, RT at 11.7 min) obtained by the *in vitro* aromatic prenyltransferase assay using purified FamD2 enzyme. The spectrum of product **2** obtained from SL3285 (A) is compared with the product at pH10.0 (B) and the authentic standard of 3-GC (C). The product peak (RT of 11.7 min) obtained by the chemoenzymatic method at pH 6.0 (D) is corroborated with the intact-ring absorbance of *cis* indole-isonitrile (**1**) (E).



Figure S7. 1H NMR spectrum of hapalindole H (3) isolated from strain SL3251.



Figure S8. *Escherichia coli* is unable to produce hapalindoles but accumulates *cis* indole-isonitrile. The figure shows HPLC traces of organic cell extracts from XL1 Blue *E. coli* hosting the *fam* plasmids cultured in LB media, S2973 strain extracts ("SL") cultured in BG11 media, and authentic *cis* indole-isonitrile standard monitored at 222 nm. All *E. coli* strains in the figure were induced with 200 μ M IPTG and S2973 was induced with 50 – 75 μ M IPTG. (A) authentic *cis* indole-isonitrile (**1**). (B) *E. coli* with pSL3084, a control plasmid. (C) *E. coli* with pSL3228. (D) *E. coli* with pSL3325 (P_{lac} replacing P_{cpcB}). (E) *E. coli* with pSL3251. (F) SL3251, hapalindole H peak is marked (**3**). (G) *E. coli* with pSL3283. (H) SL3283, 12-*epi* hapalindole U peak is marked (**4**). The peak at 7.8 minutes in the *E. coli* extracts is an unknown metabolite not related to the hapalindoles.



Figure S9. 1H NMR spectrum of 12-epi hapalindole U (4) isolated from strain SL3283.

Supplementary Tables

Table S1. Promoter Information.

Promoter	S2973 Gene	S7942 Gene	Gene product/ Notes	Promoter, 5' UTR and RBS Sequence*
	ID	Locus Tag		
0195	M744_RS01990	Synpcc7942_0195	hypothetical protein	ccacctgcgttcaaccgcattggaagcaatcgccaagggaacactaaaactaaggcctcagcctc
				cttcctcacccgccgctccacttcggagagtcacc
1613	M744_RS08700	Synpcc7942_1613	hypothetical protein	gaagcgtaacaattcaggaaccgtaattctggcaagctcgtcataataatcgtgtttgcccaccccc
				aatccgct
0398	M744_RS00950	Synpcc7942_0398	hypothetical protein	agggcattcaactgctgagtctgcggggcttaccctaagatttgtgaagacctttgctgccgagcagg
				cagtagcctgtcctagcccggcgaggagaacacgcc
1616	M744_RS08715	Synpcc7942_1616	ribonuclease P	tcacggtcaagccaggccttgacctcgacaccagcacgagtcagtttttgcaagaattagaagacc
				ttttaactcgggcggagatcatcc
1219	M744_RS10045	Synpcc7942_1219	50S ribosomal protein L21	gatcgctttgtccacagtttgacagttctcaaaagtttctgagatgatagagggctgttgtttaaaaggc
				ggctcggaacaggtttcttcaagtgctctctctgccctcggcaggtgtagccacctctacggctgaaa
				atgc
1999	M744_RS06155	Synpcc7942_1999	RNA-binding protein	gtgacgagagtgaaataatctggcactttcagtttttttt
				gcgtggtttgttcggcagcgcgcaggttactttggtcggcccatccgacccagcgtctgagaggttca
				ggttcagagtttgctttccgttttcccgaacagtccgattctcgctttaggatacgcaa

1987	M744_RS06215	Synpcc7942_1987	hypothetical protein	aagacatctttcttcatttcaccttgaacaatcattgcgttttctcagtcctcggattagggaatctaggct
				gggagcatcctgccaaaattccgctgct
0012	M744_RS02920	Synpcc7942_0012	30S ribosomal protein S6	gatcgctgagcctcatttttcggtaaactaaaagatccttgcttcaaactcactgggtaaaaccagctg
				aagatttggagccccatgaccataaggattggttctg
0657	M744_RS12900	Synpcc7942_0657	hypothetical protein	gaccccaacctgcatcttcccctgcgatcgcgggaatactgagcccatcgttaacagccacccgca
				cagttgacagaaaaaggggatcacgctaaaccaagtcaaaatctaagaggttagccagcaatctt
				aagaaagtgtaaaattgtgactagcccaggagaacgatc
1809	M744_RS07105	Synpcc7942_1809	diflavin flavoprotein A	aaccggctcattcacccagacgcgggcgatcgcgtcacaataacaaaacagcgaatttacagcct
				cggagggattc
1479	M744_RS08030	Synpcc7942_1479	cytochrome <i>b</i> ₆ <i>f</i> complex	gtttgcaggacgtgtctcagaatgatcatacacctttacaattagaggggtagcagcagcaatcgcc
			subunit 5 PetG	ccttgccttgggcttttcggggagctagctcaatcgctacactgcttaagcaccctgtatttccaagagg
				cagccaatc
lacZ	-	-	E coli lacZ beta-D-	tagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtgaaattgtgag
			galactosidase	cggataacaatttcacacaggaaacagct
1830	M744_RS07010	Synpcc7942_1830	thioredoxin	ttaacccacttatgaattctttcgtttcctagcagacagggtcatgagacatccgctatcttaggattttaa
				tgacctgatccaagcaaaggctgaacct
2486	-	Synpcc7942_2486	hypothetical protein	cggatcaggttgactaaagtccggcagaatagaattagatttaagagaagagggccagacggtct
nsaA	M744 RS05900	Syppec7942 2049	nhotosystem L P700	
poan				
			chlorophyll a apoprotein	gctccttcccgtgtggaatcctgcgtcgctgtttttgttttcagagattgcaggtccttcct
			A1 (PsaA1)	cccaccgctcccttgctgagccaactgcgctatccgggttcatggtgctgtttagtctcggagttgccg

				agggggtggccccgagaaatctgaggtcgccacgtgtggctctcagaggaagga
				ccctgtctgaagagagagagtctca
2352	M744_RS04355	Synpcc7942_2352	sigma 54 modulation	at caata attgt caa cag a caa a a ggtt ctt a a cg cact cttt cct at cag t gat a ggg t ggg t ggt gat a ggg t gat a ggg t gat a ggg t
			protein	gggggtagaaaacttgtctgggaggcgattt
laclq	-	-	laclq is a mutated version	gacaccatcgaatggtgcaaaacctttcgcggtatggcatgatagcgcccggaagagagtcaattc
			of the <i>E. coli</i> lacl promoter	agggtggtgaat
cpcB†	M744_RS10895	Synpcc7942_1052	phycocyanin beta subunit	cccgtcagtcagagcttcacaatttttagcgaatcttgtggccgcgatcgttgtataagaatgccaagg
			(срсВ)	caactggataaggttcaggagactggttga
J23119	-	-	synthetic promoter,	aattettgacagetagetcagteetaggtataatgetageatetataetggaagagagteaatteagg
			Biobrick BBa_J23119	gtggtgaat
trc1O	-	-	derivative of synthetic tac	ttgacaattaatcatccggctcgtataatgtgtggaattgtgagcggataacaatttcacacatactag
			promoter	agaaagaggagaaatactag

* 5' UTR, 5' untranslated region. RBS, ribosome binding site.

[†] truncated version of the natural *cpcB* gene promoter from S2973 and S7942.

 Table S2. PCR primers and oligos.

Oligo/Primer Name	Sequence (5' to 3')
P0012-F	tcgaattcaggtaccatcacctgcatcggatcgctgagcctcatttttc
P0012-R	tgaacagctcctcgcccttgctcaccatcagaaccaatccttatggtc
P0195-F	tcgaattcaggtaccatcacctgcatcgccacctgcgttcaacc
P0195-R	tgaacagctcctcgcccttgctcaccatggtgactctccgaagtgg
P0398-F	tcgaattcaggtaccatcacctgcatcgagggcattcaactgctgag
P0398-R	tgaacagctcctcgcccttgctcaccatggcgtgttctcctcg
P0657-F	tcgaattcaggtaccatcacctgcatcggaccccaacctgcatc
P0657-R	tgaacagctcctcgcccttgctcaccatgatcgttctcctgggc
P1219-F	gaattcaggtaccatcacctgcatcggatcgctttgtccacag
P1219-R	aacagctcctcgcccttgctcaccatgcattttcagccgtagag
P1479-F	tcgaattcaggtaccatcacctgcatcggtttgcaggacgtgtctcag
P1479-R	tgaacagctcctcgcccttgctcaccatgattggctgcctcttg
P1613-F	tcgaattcaggtaccatcacctgcatcggaagcgtaacaattcagga
P1613-R	tgaacagctcctcgcccttgctcaccatagcggattgggggt
P1616-F	tcgaattcaggtaccatcacctgcatcgtcacggtcaagccagg
P1616-R	tgaacagctcctcgcccttgctcaccatggatgatctccgcccg
P1809-F	tcgaattcaggtaccatcacctgcatcgaaccggctcattcaccc
P1809-R	tgaacagctcctcgcccttgctcaccatgaatccctccgag
P1830-F	accctaacgggtgttttttgtttctggtctcccttaacccacttatgaattctttc
P1830-R	ttacatacatccataggttcagcctttgcttg
P1987-F	gaattcaggtaccatcacctgcatcgaagacatctttctt

P1987-R	aacagctcctcgcccttgctcaccatagcagcggaattttgg
P1999-F	gaattcaggtaccatcacctgcatcggtgacgagagtgaaataatctg
P1999-R	aacagctcctcgcccttgctcaccatttgcgtatcctaaagcg
P2352-F	gaattcaggtaccatcacctgcatcgatcaataattgtcaacagacaca
P2352-R	aacagctcctcgcccttgctcaccataaatcgcctcccagac
P2486-F	tcctgcaaggtcgaattcaggtaccatcggatcaggttgactaaagtc
P2486-R	aaagctttatttatcgacatagaccgtctggccctc
PcpcB-F	tcgaattcaggtaccatcacctgcatcgcccgtcagtcag
PcpcB-R	acageteetegeeettgeteaceatteaaceagteteetgaacettateeagttgee
PJ23119-F	tcgaattcaggtaccatcacctgcatcgaattcttgacagctagct
PJ23119-R	tgaacagctcctcgcccttgctcaccatattcaccaccctgaattg
Plac-F	tcgaattcaggtaccatcacctgcatcgtagctcactcattaggcacc
Plac-R	tgaacagctcctcgcccttgctcaccatagctgtttcctgtgtgaaa
Placlq-F	tcgaattcaggtaccatcacctgcatcggacaccatcgaatggtg
Placlq-R	tgaacagctcctcgcccttgctcaccatattcaccaccctgaattg
PpsaA-F	tcgaattcaggtaccatcacctgcatcgaagctctttgaattattgttaatcg
PpsaA-R	tgaacagctcctcgcccttgctcaccattgagactctcctcttcagac
Ptrc-F	ggtcgaattcaggtaccatcacctgcatcgttgacaattaatcatccgg
Ptrc-R	ggtgaacagctcctcg
3084-1F	atctagaggatctgtaacaacaaagccgccgtcc
3084-1R	ccaccgtaggcatcatgg
3084-2F	ctggaagctatggctaaccg
3084-2R	ctttgttgttacagatcctctagatggtacctgaattc

CK138	ctcctgcaaggtcgaattcaggtaccatgtctagctttaatgcggtagtt
СК139	gagcagaagaggctactaggtaggatcaaaaataatc
CK140	ttgatcctacctagtagcctcttctgctcctgcag
CK141	gagaattttttcactaatcatggtgcacttacctccttagc
CK142	aagtgcaccatgattagtgaaaaaattctcagac
CK065	tgactagtacttactaactcttgttgtcaaggg
CK066	ttgacaacaagagttagtaagtactagtcacaaagcatagg
CK112	gatgcctggtttataaaatatgtacccgttgcaa
CK113	aacgggtacatattttataaaccaggcatcaaataaaacg
CK143	gggacggcggctttgttgttacagatcctgtttaaactataaacgcagaaaggcc
CK170	cttaataaattaaaaggtaaatatcatac
CK171	atttacctttaatttttaatttattaagtcattgagcaagagcaaaat
СК073	cgttgatacctaagattatctaagttgaaattgaggtaag
СК074	tttcaacttagataatcttaggtatcaacggtttctg
CK075	taccgtaccacgaaggagaaatagatgaagcgaaatttgattg
СК076	cttcatctatttctccttcgtggtacggtactaaattacagccgattcaac
CK041	aggagactggttgaatgttatcaaaattggtgaagg
СК037	caattttgataacattcaaccagtctcctgaacc
СК077	cgggacggcggctttgttgttacagatcctttagcgaatcttgtggc
СК217	tgttccttcactaacattcaaccagtctcctgaacctt
CK172	gctttgttgttacagatcctgtttgtttaaacttagcgaatcttgtggc
CK188	ataagaatgccaagggattatctaagttgaaattgaggtaag
CK189	tcaacttagataatcccttggcattcttatacaacg

СК270	cacaggaaacagctatgttatcaaaattggtgaagg
CK271	aattttgataacatagctgtttcctgtgtgaaattg
CK272	ggcggctttgttgttacagatcctgttttagctcactcattaggcacc

Strain	Strain Genotype
SL3228	pSL3228[<i>lacl</i> (W220F), <i>famH1, famH2, famH3, aph3</i> (Kan ^R)]
SL3251	pSL3251[<i>lacl</i> (W220F), <i>famH1, famH2, famH3, famE2, famD2, famC2, famC3, aph3</i> (Kan ^R)]
SL3285	pSL3285[<i>lacl</i> (W220F), <i>famH1, famH2, famH3, famE2, famD2, aph3</i> (Kan ^R)]
SL3283	pSL3283[<i>lacl</i> (W220F), <i>famH1, famH2, famH3, famE2, famD2, famC1,</i> <i>aph3</i> (Kan ^R)]

Li's Reported Spectra ¹ (all shifts in	Observed Spectra (all shifts in ppm,
ppm, C ₆ D ₆)	C ₆ D ₆)
6.76 (bs) 1H	6.74 (bs) 1H
7.62 (dd, <i>J</i> =1.4, 2.2 Hz) 1H	7.62 (dd, <i>J</i> =1.7 Hz) 1H
7.05 (d, <i>J</i> =7.2 Hz) 1H	7.06 (d, <i>J</i> =7.3 Hz) 1H
7.26 (t, <i>J</i> =7.8 Hz) 1H	7.27 (t, <i>J</i> =7.7 Hz) 1H
6.96 (d, <i>J</i> =8.0 Hz) 1H	6.96 (d, <i>J</i> =8.1 Hz) 1H
3.14-3.06 (m) 1H	3.13-3.07 (m) 1H
3.12-3.10 (m) 1H	3.11-3.10 (m) 1H
1.62 (td, <i>J</i> =3.4, 13.8 Hz) 2H	1.61 (m) 1H
0.81 (dt, <i>J</i> =5.6, 12.7 Hz)	0.80 (td, <i>J</i> =5.7, 13.2 Hz) 1H
1.35-1.30 (m) 2H	1.36-1.30 (m) 2H
1.13-1.07 (m) 1H	1.11-1.07 (m) 1H
0.96 (s) 3H	0.96 (s) 3H
1.29 (s) 3H	1.29 (s) 3H
1.09 (s) 3H	1.09 (s) 3H
6.18 (dd, <i>J</i> =11.1, 17.6 Hz) 1H	6.18 (dd, <i>J</i> =11.0, 17.6 Hz) 1H
5.14 & 5.09 (dd, <i>J</i> =11.1, 1.3 Hz) 2H	5.14 (d, <i>J</i> =11.0 Hz) 1H & 5.09 (d, <i>J</i> =17.5
	Hz) 1H

Table S4. ¹H NMR shifts for hapalindole H (**3**) produced by strain SL3251 compared to a previous report.

Table S5. ¹H NMR shifts for 12-*epi* hapalindole U (**4**) produced by strain SL3283 compared to a previous report.

Li's Reported Spectra ² (all shifts in	Observed Spectra (all shifts in ppm, C ₆ D ₆)
ppm, C ₆ D ₆)	
6.67, bs, 1H	6.67, bs, 1H
6.38, t, (<i>J</i> =1.9Hz), 1H	6.38, t, (<i>J</i> =1.9Hz), 1H
7.08, d, (<i>J</i> =7.3Hz), 1H	7.08, d, (<i>J</i> =7.3Hz), 1H
7.29, t, (<i>J</i> =7.7Hz), 1H	7.29, t, (<i>J</i> =6.7Hz), 1H
6.97, d, (<i>J</i> =8.2Hz), 1H	6.98, d, (<i>J</i> =8.1Hz), 1H
3.03, d, (<i>J</i> =11.3Hz), 1H	3.03, d, (<i>J</i> =11.7Hz), 1H
3.94, s, 1H	3.94, s, 1H
1.69, td, (<i>J</i> =13.0, 3.9Hz), 1H	1.69, td, (<i>J</i> =13.3, 4.5Hz), 1H
1.48, m, 2H	1.49, m, 1 or 2H
1.37, td, (<i>J</i> =12.9, 3.8Hz), 1H	1.38, td, (<i>J</i> =12.8, 4.1Hz), 1H
2.01, tc, (<i>J</i> =12.0, 3.5Hz), 1H	2.01, td, (<i>J</i> =11.9, 3.5Hz), 1H
1.01, s, 3H	1.01, s, 3H
1.32, s, 3H	1.32, s , 3H
5.43, dd, (<i>J</i> =17.6, 11.0Hz), 1H	5.43, dd, (<i>J</i> =17.7, 11.0Hz), 1H
4.93, dd, (<i>J</i> =11.0, 0.6Hz), 1H	4.92, dd, (<i>J</i> =11.0Hz), 1H
4.88, dd, (<i>J</i> =17.5, 0.6Hz), 1H	4.88, dd, (<i>J</i> =17.7Hz), 1H
1.19, s, 3H	1.19, s, 3H

Supplementary References

- Li, S., Lowell, A. N., Newmister, S. A., Yu, F., Williams, R. M., and Sherman, D. H. (2017) Decoding cyclase-dependent assembly of hapalindole and fischerindole alkaloids. *Nat. Chem. Biol.* 13, 467.
- (2) Li, S., Lowell, A. N., Yu, F., Raveh, A., Newmister, S. A., Bair, N., Schaub, J. M., Williams, R. M., and Sherman, D. H. (2015) Hapalindole/Ambiguine Biogenesis Is Mediated by a Cope Rearrangement, C-C Bond-Forming Cascade. *J. Am. Chem. Soc.* 137, 15366-15369.