

Integrative biosynthetic gene cluster mining for optimization of metabolic pathway to efficiently produce L-homophenylalanine in *Escherichia coli*

Zhenning Liu[†], Dengwei Lei[†], Bin Qiao[†], Shilin Li[†], Jianjun Qiao^{†,‡}, Guang-Rong Zhao^{*,†,‡}

[†]Frontier Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Yaguan Road 135, Jinnan District, Tianjin 300350, China

[‡]SynBio Research Platform, Collaborative Innovation Centre of Chemical Science and Engineering (Tianjin), Tianjin University, Yaguan Road 135, Jinnan District, Tianjin 300350, China

*Corresponding author: Guang-Rong Zhao (grzhao@tju.edu.cn)

Tel: +86-22-85356580; Fax: +86-22-27403389

Supporting Information

Table S1: Accession number of amino acid sequences of 39 L-Hph biosynthetic cluster used for phylogenetic analysis in this study.

Table S2: The abundance data of target nonribosomal peptides (NRPs) used for depiction of the heatmap in this study.

Table S3: The main primers used in this study.

Table S4: Nucleotide sequences of codon optimized genes used in this study.

Fig. S1: SDS-PAGE analysis of *E. coli* BW25113 (control), BWH13, BWH14 and BWH15.

Fig. S2: HPLC analysis of fermentation broth of strain BWH01 (expression of L-Hph BGC from *N. punctiforme* PCC73102).

Fig. S3: LC-MS analysis for identification of 2-BMA and L-Hph.

Table S1. Accession number of amino acid sequences of 39 L-Hph biosynthetic cluster used for phylogenetic analysis in this study.

| Strains | Accession No. | | |
|---|----------------|----------------|----------------|
| | HphA | HphCD | HphB |
| <i>Nostoc punctiforme</i> PCC 73102 | WP_012409019.1 | WP_012409013.1 | WP_012409012.1 |
| <i>Nostoc</i> sp. ATCC 53789 | WP_114081681.1 | WP_114081674.1 | WP_114081673.1 |
| <i>Nostoc</i> sp. HIID-D1B | AVK43370.1 | AVK43361.2 | AVK43360.1 |
| <i>Nostoc</i> sp. XHIID C2 | AVK43396.1 | AVK43390.1 | AVK43389.1 |
| <i>Nostoc</i> sp. 268 | AVK43342.1 | AVK43335.2 | AVK43334.1 |
| <i>Nostoc</i> sp. KVJ20 | WP_069068163.1 | WP_069068157.1 | WP_069068156.1 |
| <i>Nostoc</i> sp. 213 | WP_094332088.1 | WP_094332086.1 | WP_094332087.1 |
| <i>Nostoc</i> sp. N6 | WP_104901026.1 | WP_104901028.1 | WP_104901027.1 |
| <i>Nostoc</i> sp. 232 | WP_094346109.1 | WP_094346111.1 | WP_094346110.1 |
| <i>Nostoc</i> sp. XPORK14A | AVK43401.1 | AVK43403.2 | AVK43402.1 |
| <i>Nostoc</i> sp. N135.9.1 | AVK43376.1 | AVK43378.1 | AVK43377.1 |
| <i>Nostoc calcicola</i> FACHB-389 | OKH39823.1 | OKH39821.1 | OKH39822.1 |
| <i>Scytonema</i> sp. HK-05 | WP_073634529.1 | WP_073634541.1 | WP_073634542.1 |
| <i>Nostoc</i> sp. CENA543 | WP_103139157.1 | WP_103139164.1 | WP_103139165.1 |
| <i>Nostoc</i> sp. FSN-E | AVK43354.1 | AVK43347.2 | AVK43346.1 |
| <i>Anabaena</i> sp. BIR260 | AVK43216.1 | AVK43207.1 | AVK43206.1 |
| <i>Anabaena</i> sp. 90 | WP_015078980.1 | WP_015078972.1 | WP_015078971.1 |
| <i>Nodularia spumigena</i> UHCC 0039 | WP_006196343.1 | WP_006196336.1 | WP_006196335.1 |
| <i>Nodularia spumigena</i> CENA596 | WP_063874468.1 | WP_063874473.1 | WP_063874474.1 |
| <i>Sphaerospermopsis torques-reginae</i> ITEP-024 | API83199.1 | API83191.1 | API83190.1 |
| <i>Dolichospermum compactum</i> NIES-806 | WP_096669565.1 | WP_096669549.1 | WP_096669547.1 |
| <i>Anabaena</i> sp. LE011-02 | OBQ10118.1 | OBQ10125.1 | OBQ10126.1 |
| <i>Anabaena</i> sp. AL09 | OBQ12658.1 | OBQ12650.1 | OBQ12649.1 |
| <i>Snowella</i> sp. 249/25 | AVK43467.1 | AVK43465.1 | AVK43466.1 |
| <i>Planktothrix agardhii</i> No758 | AQY61077.1 | AQY61075.1 | AQY61076.1 |
| <i>Planktothrix rubescens</i> NIVA-CYA 407 | WP_026785824.1 | WP_026785826.1 | WP_026785825.1 |
| <i>Planktothrix rubescens</i> No108 | AQY60314.1 | AQY60312.1 | AQY60313.1 |
| <i>Planktothrix agardhii</i> NIVA-CYA 126/8 | AQY60510.1 | AQY60508.1 | AQY60509.1 |
| <i>Microcystis aeruginosa</i> PCC 9432 | WP_002752284.1 | WP_002752281.1 | WP_002752283.1 |
| <i>Planktothrix agardhii</i> NIVA-CYA 68 | AQY61269.1 | AQY61267.1 | AQY61268.1 |
| <i>Planktothrix agardhii</i> No66 | AQY60686.1 | AQY60684.1 | AQY60685.1 |
| <i>Microcystis aeruginosa</i> PCC 7005 | WP_024969395.1 | WP_002752281.1 | WP_002752283.1 |
| <i>Microcystis aeruginosa</i> PCC 9701 | WP_002804152.1 | WP_043998115.1 | WP_002752283.1 |
| <i>Planktothrix agardhii</i> NIVA-CYA 56/3 | WP_026785824.1 | WP_027255437.1 | WP_027255438.1 |
| <i>Planktothrix agardhii</i> NIVA-CYA 34 | WP_026785824.1 | WP_026795616.1 | WP_026795615.1 |
| <i>Planktothrix agardhii</i> NIVA-CYA 15 | WP_027248863.1 | WP_026795616.1 | WP_027248864.1 |
| <i>Planktothrix rubescens</i> NIES-610 | WP_026798230.1 | WP_026798228.1 | WP_026798229.1 |

| | | | |
|---------------------------------------|----------------|----------------|----------------|
| <i>Planktothrix agardhii</i> NIES-204 | BBD53655.1 | BBD53653.1 | BBD53654.1 |
| <i>Planktothrix agardhii</i> NIES-905 | WP_141293771.1 | WP_026785826.1 | WP_141293773.1 |

Table S2. The abundance data of target nonribosomal peptides (NRPs) used for depiction of the heatmap in this study.

| Strains | Target NRPs | Abundance (mg/kg) | Reference |
|------------------------------------|--------------------|-------------------|-----------|
| <i>Anabaena</i> sp. strain 90 | Anabaenopeptin A | 1383 | 1 |
| | Anabaenopeptin B | 872 | |
| | Anabaenopeptin C | 553 | |
| <i>S. torques-reginae</i> ITEP-024 | Namalide B | 400 | 2 |
| | Namalide C | 70 | |
| | Spumigin K | 170 | |
| | Spumigin L | 140 | |
| | Spumigin M | 130 | |
| | Spumigin N | 150 | |
| | | | |
| <i>N. spumigena</i> UHCC 0039 | Nodulapeptin A | 426 | 3 |
| | Nodulapeptin B | 1277 | |
| | Spumigin A | 319 | |
| | Spumigin B1 | 638 | |
| | Spumigin B2 | 447 | |
| | Spumigin C | 426 | |
| <i>P. agardhii</i> CYA126/8 | Anabaenopeptin 908 | 3063 | 4 |
| | Anabaenopeptin 915 | 2438 | |
| | Cyanopeptolin 880 | 1813 | |
| | Cyanopeptolin 960 | 813 | |
| <i>P. agardhii</i> NIES-204 | Anabaenopeptin B | 1541 | 5 |
| | Anabaenopeptin E | 178 | |
| | Anabaenopeptin F | 64 | |
| <i>P. agardhii</i> NIES-905 | Oscillamide B | 2000 | 6 |
| <i>P. agardhii</i> NIES-610 | Anabaenopeptin F | 300 | 6, 7 |
| | Oscillamide Y | 515 | |

Table S3. The main primers used in this study.

| Primer | Sequence |
|-----------------------|--|
| Candidates assessment | |
| ACD01 F | GGAATAAACCATGGAAACCCTGCCGATCATCA |
| ACD01 R | TTCGGGTTACCGGAATTCTTATTATTAATCAAAGCGATCGCT AATTTTACCGTTTACAG |
| PE01 F | AAGAATTCCGGTAACCCGAAGCTTGTTTAAACGGT |
| PE01 R | GGGTTTCCATGGTTTATTCCCTCCTTATTTAATCGATACATTAA TATATACCTCTTTAA |
| B02 F | GGAATAAACCATGGAAATCTCTGTCTAAGGCATCTTATCG |
| B02 R | AGGGGTTACGCGGATCCTTATTATTATTTACGAGACTCATGC AGTACACCCA |
| PA02 F | TAAGGATCCGCGTAACCCCTCGA |
| PA02 R | GAGATTCCATGGTTTATTCCCTCCTTATTTAATCGATACATTAA TATATACCTCTTTAA |
| ACD03 F | ATGCAAACCTCTGCCGATTAATAATTTCCG |
| ACD03 R | CCGGAATTCTCACGGTATTTTTTCATCCAGCAGGTGGCTAATT TTAC |
| PE03 F | AACCGTGAGAATTCCGGTAACCCGAAGCTTG |
| PE03 R | TCGGCAGAGTTTGCATGGTTTATTCCCTCCTTATTTAATCGATA CATTAAATATATACCT |
| ACD04 F | GTAAATTATGCAGACCCTGCCGATCAAGA |
| ACD04 R | AATTCAGTCCGTTAATCCAGAATGTGGCTGATTTTACCCTTA ACCG |
| PE04 F | CTGGATTAACGGACTGAATTCCGGTAACCCGAAGCTTG |
| PE04 R | CAGGGTCTGCATAATTTAACCTCCTTAATTTAATCGATACATT AATATATACCTCTTTAA |
| ACD05 F | ATGCAAACCTCTGCCGATTAATAATTTCTG |
| ACD05 R | CCGGAATTCTCACGGTAAATCCAGGTTGGCAGTAATTTTACC TTAATAGCAG |
| PE05 F | AACCGTGAGAATTCCGGTAACCCGAAGCTTG |
| PE05 R | TCGGCAGAGTTTGCATGGTTTATTCCCTCCTTATTTAATCGATA CATTAAATATATACCT |
| A06 F | ATGCAAACCTCTGCCGATTAATAATTTCCG |
| A06 R | CGGAATTCTCACGGTAAATCAAAGCGATCGCTAATTTTACCG TTTACAGCA |
| PE06 F | GATTAACCGTGAGAATTCCGGTAACCCGA |
| PE06 R | TCGGCAGAGTTTGCATGGTTTATTCCCTCCTTATTTAATCGATA CATTAAATATATACCT |

A07 F AGGTAAATTATGCAGACCCTGCCGATCAA
A07 R GGATCCTGATGAATTACTCGAAAATACCGTC
PE07 F CGAGTAATTCATCAGGATCCAAGAGGAGAAATTA
PE07 R GGGTCTGCATAATTTAACCTCCTTAATTTAATCGATACATTA
TATATACCTCTTTAA
A08 F ATGCAAACCTCTGCCGATTAATAATTTCTG
A08 R TCGAAGATGCCATTCTTCCACAGCT
PE08 F TGAAGAATGGCATCTTCGAATAaTTCATCAGGATCCAAGAG
GAGAAATTA
PE08 R TCGGCAGAGTTTGCATGGTTTATTCTCCTTATTTAATCGATA
CATTAAATATACCT
CD09 F TCAGGATCCATGAGGAGAATCAGAACATGTCTTCTATCAAAC
AGATTCTGTACCTG
CD09 R CCGGAATTCTCACGGTTATTTTTTCATCCAGCAGGTGGCTAATT
TTAC
PE09 F AACCGTGAGAATTCCGGTAACCCGAAGCTTG
PE09 R CTCCTCATGGATCCTGATGAATTACTTAAACACACCG
CD10 F GTAATTCATCAGGATCCAAGAGGAGAAATTA
CD10 R AATTCAGTCCGTTAATCCAGAATGTGGCTGATTTTACCCTTA
ACCG
PE10 F CTGGATTAACGGACTGAATTCCGGTAACCCGAAGCTTG
PE10 R CTTGGATCCTGATGAATTACTTAAACACACCG
CD11 F TCAGGATCCATGAGGAGAAAACACGGATGACTGATGCCAAC
AAAATCCTGAAC
CD11 R CCGGAATTCTCACGGTTAATCCAGGTTGGCAGTAATTTTACC
TTAATAGCAG
PE11 F AACCGTGAGAATTCCGGTAACCCGAAGCTTG
PE11 R CTCCTCATGGATCCTGATGAATTACTTAAACACACCG
B12 F TAATTTATGGGTAGCCTGAGCAAGCCGA
B12 R ACGCGGATCCTAACGGTTAGCTGCTCAGTTCCTCCAGCA
PA12 F TAACCGTTAGGATCCGCGTAACCCCTCGA
PA12 R CTCAGGCTACCCATAAATTATTCCTCCTTATTTAATCGATACA
TTAATATATACCTCT
B13 F ATGCGTAGCAGCAGCAACCAGA
B13 R GGATCCTCACGGTTACTTATAAAATTCGTGCAGCACACCAAT
C
PA13 F ATAAGTAACCGTGAGGATCCGCGTAACCCCTCGA
PA13 R TGGTTGCTGCTGCTACGCATAAATTATTCCTCCTTATTTAATC
GATACATTAATATATAC
B14 F GAATAATTTATGGCGAGCCTGAACAAGAGCAGCT

B14 R ATCTCACGGTTATTCGGTCGGTTGCTCCAGGCTGA
 PA14 F GACCGAATAACCGTGAGGATCCGCGTAACC
 PA14 R CAGGCTCGCCATAAATTATTCCTCCTTATTTAATCGATACATT
 AATATATAC
 B15 F GGAATAAACCATGTCTGAAGAATTACCATATTGCCGTATTG
 B15 R TCCTTATTACACCCCTTCTGCTACATAGCG
 PA15 F GCAGAAGGGGTGTAATAAGGATCCGCGTAACCCCTCGA
 PA15 R CTTCGACATGGTTTATTCCTCCTTATTTAATCGATACATTAAT
 ATATAC

Plasmid copy number variation

B16 F CGGGGTACCAGCTGCAGATCTTCCTTAACCGCTC
 B16 R GGAATTCCATATGGTTTGACAGCTTATCATCGACTGCACGGT
 G
 P1016 F GGAATTCCATATGTAGGCTGCTGCCACCGCTGAGCAAT
 P1016 R CGGGGTACCCCGAAGGCCAGTCTTTCG
 CL1719 F TGA CTTCGCGTTTCCAGACTTTACGAAAC
 CL1719 R AGTTATTGCTCAGCGGTGGCAGCA
 PCDF19 R TGGGGCTGACTTCAGGTGCTACATTTGAAG
 18 F TGAGACTGCTGACCAAATCCCTTAACGTGAGTTTTC
 18 R TTCAAGCCGATGTCAGACCAAGTTTACTCATATATAC
 STR F TGGTCTGACATCGGCTTGAACGAATTGTTAGACA
 STR R GATTTTGGTCAGCAGTCTCACGCCCGGAGCGTA

Chromosomal integration and deletion

feaBU F TGGCACGTTTTTCGCTCTGTGAAATGTA
 feaBU R TAGGCAAAGATCGCCAACCGTTTTTTCAC
 Fap F GTTGGCGATCTTTGCCTAATCTAGAGTTTGACAGCTTATCATC
 GACTGC
 Fap R CAATCTGTGGCAGCAGCCTAGGTACC
 Chl F TAGGCTGCTGCCACAGATTGCAGCATTACACGTCTTGAG
 Chl R CGATCGCTATATTCCAGAGACAGGAACACTTAACGGCTGACA
 TG
 feaBD F CTCTGGAATATAGCGATCGCTTACAGGC
 feaBD R GTTGTCTGGATGACCACCGGCACCA
 gRNA F CACACCAGGTCTCACCAGTCATGTTGCACTAGTATTATACCT
 AGGACTGAGC
 gRNA R CACACCAGGTCTCACTGGGATCGAGGTTTTAGAGCTAGAAAT
 AGCAAGTTAA
 tyrAU F CCGAATAGCCCGCAATACCT
 tyrAU R TCGGAAGACTGATTGCCCTCG

tyrAD F AGGGCAATCAGTCTTCCGAGGATTACCGTGAATCGGGCG
tyrAD R TCCGGTGCCATAAACGACG
leuB-Chl F GAGTTGCAACGCAAAGCTCAACACAACGAAAACAACAAGGA
AACCGTGTGAGATTGCAGCATTACACGTCTTGA
leuB-Chl R CAATTTTTCGTATAACGTCTTAGCCATGATTACACCCCTTCTG
CTACATAACAGGAACACTTAACGGCTGACAT

Table S4. Nucleotide sequences of codon optimized genes used in this study.

| | |
|---------------------------|---|
| <i>Np_{hphA}</i> | <p>ATGGAAACCCTGCCGATCATCATTTCTGATGAAACTCTGCGTG ATGGTGAACAACAAGCAGGTATCTTCTTCTCTTATCCAATAA GCAAAAACCTGGCCCATCTGATTGCCCAAACCTGGTGTAGATCGT CTGTCTATGATGCCGCGTGTTTCATGAGCAAGAAGAAGAGCTGG TTAAATCTCTGGTATCTGAAGGTCTGGATTCTCTGATCGCTGCT GCTACCATGACCGGTAAGCAGTATATCGATCAAGCCAAAATGT ACGGTGTAAAGCGCATCATCTGTTCTACGGCCTGTCCGATCGC CTGCTGTTTCTGCGTGATCCGCAGATCCGTCTCCTCAACGAGTT CAAAGGGAAAACCATTTGATGATGATATCCCAAAAAAAGTAAT TGAGCGTATTCGCCAAAACGCCATTGATCTGATTGTTGAAAAC CTGCACTATGCCACTACCGTTGCTGGTCTCACTGTAGACTTCGC CGCCGAAGACGCCTCTCGTGCCGATTTTGACTTTCTGGTGCAAT GTATCCGCTCTTTTAGCCCGTACCTCGACCACTTCATGCTGTGT GATACCGTTGGTGTCTGACTCCAGAAAAGAGCTATATCTGGA TTAACGACCTGCTGCAACGTACCACTGGTGTGCTTTCGGTGTA CACTACCACAACGACATGGGCCTGGCTCTGGAAAACACTCTGC AATCTGTTATCGCGGGTGGCGACTCTGGTAAGCGGTACTTTCTGC GGGATTGGTGAACGTGCCGGTAACGTTGCCATTGAGCAGGTAC TGAACGGTCTGCGTGTTTCGCTTTGGTATTGAAGTTAAAGGTATT AACTATGATGCGATCGCCGCCGTGACTGATTACATTGAGCAAC TGGGTATTCGTCCGGCTCCGCCGTACTCTCAAACCTGCTCAACGC CACGAATCTGGCATCCATGTTAACTCTCTGTTTAGCGACCCGCA AAGCTATGCTGCTCTGCCGTACAACACCATCGAGGTTCTGTTT GGTAAATGGTCTGGTATTTCTAACTTTCAATATCTGTTTAAAA GCAACTGCAAAACCCGCAACCGCGTAACCAATATGAAAAAAT GCGCTCTGTGATTAAATCTCTGGCTGTTGAGCAAGAACGCTAT TTTACCGCTAAGGAGGTACTGGAGCTGTGAAAAACGGTGTGT TTAAGTAA</p> |
| <i>Np_{hphCD}</i> | <p>ATGTCTCACGAAAACAACGTTCTGCACCTGGGTGATGATATCA ACACTGATGATATTATCCGGCTAACCGTGCTACCAACGATGA CCTGGATTCTCTGAAACATTATGCTCTGGAGCATGTTATTGGTG CAGGTGAACTGCTGAAATATGATGTGATTGAAGCAGGTGAAA ACTTCGGCTGTGGTTCTCTCGTGAAATTGCTCCGATTGCCCTG CAAGCTGCTGGTATTACCAAGATTCGTGCGCGTTCTTTCGCCGA GATTTTTTATCGTAACAGCATCAACATTGGTCTGTCTCTGAAA TTCTGGGTGAGAAACAGGAGAACCCGGTAGTTGATGCGATCGC AGCAGCAGGTGGCCTCATGTCTTCAACCAGCTGAGACGTCAG GGTCAGATTAGCATTCCGGGTAGCCTGACTCCGCCACGTCCGA</p> |

| | |
|---------------|--|
| | <p>TGACTCTGGTAGAAAACTGCTGGGCCTGGCCTCTCGTAACTC CTACGTTCA GCCGGGTGAAGTGGTTTTTGTCTCAGGTTGATCTGG CACTGTCCCATGATGCTGTGGCCGGTCCGGTAGCTCAAACCTT CTATAAACAGTATGGTGAAGGTAAGCTGTGGGACTCCCAA AAGGTGGTTCTGGTAGCAGATCACTTCATCCAAATCAACGACA TCCGTGCTGATGACAAAGCATATCTGATGTACGAACAGATGAT TGAATTTGCTAAATTCCAAGGTTGCCACCTGTTTGTATGTGGTTT CCCCAGGTGAAGCAACCGGCATCTGTCACGTTCTGCTGCCAGA AAAAGGTTTTGTTCGTCCAGGTATGATCATTGCCGGTACTGATT CCCATACCTGCACCTACGGTGCATTAGGCGCCTTCAGCACCGG CGTGGGTA CTACCGATATGGCTAACATTTATGCCACCGGTGAT ATGTGGATTCGTATTCCGCCAACCTGGTATTTGAGCTGTCTGG TACTCTGCCGCCGAAAATCTCTGCCAAAGATATTATCCTGTTTA TTCTGGGTCAAATCGGCTGTGCTGGCGCTACTTCTAAGGTAAT GGAGTTTCGTGGTTCTATCCTGGCACAGCTCCCGTTCGACGAA CGTCTGACTCTGGCCAACATGGCCGTGGAGTGTGGTGCAATCT GTGGTCTGATTGTTCCGGATGAAGTAACCCGTGAATATCTGAA AAACCGTGTCTCTGAAGAATTTCAAGAGATTATCGGCGATGCT GATGCTGAGTACGAACGTGTTTATCAATTTGACCTGAGCTCTCT GGAACCACAAGTTGCCCGTCCGCCGAAACCAGATCGTGTAGTT GCCATTTCTCAGCTGGAAAACATTCCGATTACTAAAGCCTTTAT TGGCTCTTGTACTGGTGGTAAACTGTATGACCTGGCTCAAGCA GCCGAAGTTCTGAAGGGTCCGCAACTGGCAGACGGTGTAACC TGTTTATTGTACCGGCAACTATTGAAATTCGTGAACAAGCTCA AGAACTGGGCTATATCGACATTTTTGAACAAGCAGGTGCACAA ATTCTGAAGACCGTTGTGGTGCTTGTATCAACTCTGGTTTTGG TGTTCTGGCGAAAGAAGAAACCGGTATTTATGCAACCAACCGC AACTTTAAGGGTCGCTCTGGCGATCCAACCGGCAAAA ACTATC TGGCCTCTCCGCGTACCGTAGCTATCTCTGCTGTAAACGGTAA AATTAGCGATCGCTTTGATTAA</p> |
| <i>NphphB</i> | <p>ATGGAATCTCTGTCTAAGGCATCTTATCGTATTGTTGCCATTCC GGGTGAAGGTATTGGTCCAGAAGTTGTAGAAGCTAGTCTGAAA ATTCTGCAACACGTTGCTAAAATTCAAGGTTTTACCCTGCAAGT AGACTATGGTTGGCTGGGCGCGACTGCCTTCGAACA ACTGGGA TCTTACTTCCCGCAAGCCACTACCGAACTGTGCGATGGTGCCG ATGGTATTGTATTTGGTGCAGTTTCTCAAGGCGGTCTGCTGGAA CTGCGTAAACACTTCGACTTTTTTTGCAACCTGCGTCCAATCCG TAGCGTTAACAGCCTGCTGCATAAATCTTCTCTGCGTCCAGAG AAAATTCAAGGTCTGGATATTCTGATCGTTCGTGAACTGGTGT CTGGTATCTATTTTGGTCCGTCTGGTCGCTCCTCTGATGAACGT GGTAACTATGGTTACCATAACCATGCGCTATTACGATGAAGAAA TTCGTGCTATTGCTCGTAAAGCTCTGGAGCAAGCCCAATACCG TCGTGGTCTGCTGACCGTAGCTCACAAAGAAAACGCCCTGCCG CATCTGCATTGGACTCGTCTGGTTCAAGAAGAAGCCACCAAGT</p> |

| | |
|----------------|--|
| | <p>TTCCGAACGTTATTGTTGAACCAATGCTGGTGGATAACCTGGC CATGCAAATGGTTCTGAACCCGCAACGTTTTGATGTAATTCTG GCGTCTAACCTGTTTGGTGATATTCTGAGCGATATTGGCGGTGC ACTGGTTGGCTCTATTGGCCTGCTGGGTTCTGCATCTCTGAACG CTAACGGTTTTGGTCTGTACGAGGCCATTACGGTACTGCGCC AGATATTGCAGGCCTGGGTATTGCTAACCCGCTGGGTACTCTG GGTGCTTGTGTTCTGATGCTGCAACAATGGGGTGAAGTGCAGG CTGCCAACAGATTATTGCAGCGCAAGACCGTATTCTGGCTAA GGGTACCCTACCGCTGATCTGTTTACCCAAGGCGAAGAAATT CTGGTTAACACTGCAACCCTGGTTGACCTGCTGCTGGAAGAAC TGCTGTTGTGCAACATTCTCAACTGGGTGACTGCATGAGTCT CGTAAATAA</p> |
| <i>AshphA</i> | <p>ATGCAAACCTCTGCCGATTA AAAATTTCCGACGAAACCCTGCGTG ATGGTGAACAGCAAGTTGGCATT TTTCTTCAGCTATCCGACCAA ACACAAACTGGCTCATCTGATTGCCCAAACCGGTGTTGATGAA CTGTGCATCATGCCGATGTTTATGAACAGGAGGCTAAACTGG TTAAAACCCTGGTATCTGAGGGTCTGGATGGTGTGTTACTGCT TCTACCATGATGGGTAAAGAGTATATTGATCATGCGAAGA ACT GCGGTGTTAAGCGCATTATTCTGTTTCATGCTGTTTCTGATCGT CTGCTGTTTCTGCGTGATCCGCAAATTCGCCTGATGAGCGAATT TCAAGGTAAAACCATTGATGATGATATCCCATCTGCAATCATC AACCGTATTCGTCAGAACGCTATTAACGTTATTGTAGAAAACC TGCACTATGCCACCAAAGTGGCCGGTCTGCGTGTTAACTTTGCT GCGGAGGATGCTTCTCGTGCGGATTTTGACTTTCTGGTGGAAT GTATTCGCTCTTGCTCTCCGTATATTGAAAAC TTTCTGCTGTGT GATACCGTAGGTGTGCTGAACCCAGAAAAGTCTTATATCTGGA TTCATGATCTGCTGCAATCCACTACTGGTGTAGGTCTGGGTGTA CACTACCACAACGATATCGGTCTGGCACTGGAAAACACTCTGC AATCTCTGATGGCGGGCGCGACTCTGGTAAGCGGTACTTTCTG CGGGATTGGTGAACGTGCCGGTAACGTTGCCATTGAGCAGGTG CTGAACGGTCTGCGTGTTCTGTTTGGCATTGAAGTTCAGGGTAT CAACTATGATGCGATCGCCG CAGTGACTAACTATATTGAGCAA CTGGGTATTCGTCCGGCATCTCCG TACTCTCAA ACTGCACAAC GCCACGAATCTGGTATTCATGTTAACTCTCTGTTTTCGTGATCCG AAGAGCTATGCAGTATTTCCACATAACAAAATCGATGTTGTTT TTGGCAAATGGTCTGGTGT TTTCTAACTTCCAATATCTGTTTGAA AAACA ACTGCAAAAACCCGCAATCTCGTAAGCAGTATGAAAAA ATGCGTTCGGTTATCAAATCTCTGGCAATGGAAAAAGAATGTT ACTTTACCGCTAACGAAGTTCTGGA ACTGTGGAAAGATGGTAT CTTTGAATAA</p> |
| <i>AshphCD</i> | <p>ATGTCTTCTATCAAACAGATTCTGTACCTGGGTGATGATATCAA CACTGATGATATTATGCCGGCCAACCGTGCGACCAACGATGAC CCGGATAACCTGAAACACTATGCTCTGGAACATCTGATTGGTG TAGGTGAGCTGCTGAAATACGATGTGATTGAAGCCGGTGAAAA</p> |

| | |
|---------------|---|
| | <p>CTTCGGCTGTGGTTCTTCTCGTGAAATTGCCCCGATTGCCATCA AAGCTGCCGGTATCGAACTGATTCGTGCGAAATCTTTCGCAGA AATTTTTTATCGCAACAGCATTAAACATTGGTCTGCCGCTGGAA ATTCTGGGTGAAAAACCAGAGAACCCGGTAGTTAACGTGATCG CCGCAGCCGGTGGTCTGATCCCGTTTAACCAAAGCGTCGTCA GGTAAAATTACCATTCGGGCCAGCGCGACTTCCCCGCGTCCG ATGACTCTGGTGGAAAACTGCTGGCCAAATCCTCTGGTAACA CCTACGTTACTCCAGGTGAAGTGGTTTTTGTCTCAAGTTGATCTG GCCCTGTCCCACGATGCTGTGGCCAGCCCGGTTGCTAAAATCT TTTATAAACACTATGGTGAAGATGCAAAGCTGTGGGATGCTCA ACGTGTAGTTCTGGTTGCAGATCACTTCATCCAAGTTAACGAC ATCCGTGTTGATCACAAAGCTGAACTGATGTATCAACAGATGG TTCAGTTTGCCAAAGACCAAGAATGTTACCTGTTTGATGTAGTT TCCCAGGTGAAGCCGCAGGTATCTGTCACGTTCTGCTGCCAG AAAAAGGTTTTGTTTCGTCCAGGTATGATCATTGCCGGCACCGA TTCCCATACCTGCACCTATGGTGCACCTCGGCGCGTTCAGCACC GGTGTGGTACTACCGATATGGCGAACATCTATGCTATGGGTG ATATGTGGATTTCGTGTTCCATCTACCCTGGTATTTGATCTGTCT GGTACTCTGCCGCACCATATTTCTGCCAAAGATATTATCCTGTT TATCCTGGGTCAAATTGGTGTGCTGGTGTACCCGGTAAGGTA ATGGAGTTTCGTGGTTCTATCCTGGCACAGCTCCCGTTCGACGA ACGTCTGACTCTGGCTAACATGGCTATTGAGTGTGGTGCAGTTT GTGGTCTGATTGTTCCGGATGAAACCACCCGTA ACTATGTTTCGT TCTCGTTCTACCCAAGGTTTTGTAGAAATGATTGCCGATCCGGA TGCTGAATACGAAAAAGTATATCAATTTGACCTGTCTAACCTG GAACCACAAGTTGCCCGTCCGCCAAAACCAGATCAGGTTATCC CGATTAGCCAGCTGGAGGAGACTCCGATTACCAAAGCGTTTAT TGGCTCTTGTACCGGTGGCAAGCTGTATGATCTGGCCCAGGCT GCTGAAGTTCTGCAAGGTCGTCAACTGGCAGATGGCGTGAGCC TGTTTGTGGTTCCGGCTTCCGTAGAGATTTCGTGAACAAGCTGA AGCCCTGGGTATCTGGAAATTTTTGCACAAGCAGGCGCACAG ATTCTGAAGTCTGGTTGTGGTGCTTGTATTA ACTCTGGTATGGG TGTGCTGGATAAAGAAGAAACCCGGTGTATTATGCAACCAACCGC AACTTTAAAGGTCGTTCTGGCGATCCGACCGGTA AAAACTATC TGGCCTCTCCACGTACCGTGGCAATCTCTGCCATCAAGGGTAA AATTAGCCACCTGCTGGATGAAAAATAA</p> |
| <i>AshphB</i> | <p>ATGGGTAGCCTGAGCAAGCCGAGCTACCGTATCGTTGCGATTC CGGGCGAGGGCATCGGTCCGGAAGTGGTTGAGGCGAGCCTGC AAATCCTGCAACACGTTGCGAAAATCGAAGGTTTCACCCTGCG TGTGGACTATGGTTGGCTGGGTGCGACCGGTTCCCTGCAAATG GGCAGCTATTTTCCGCAGGCGACCGTGCAACTGTGCGATGGCG CGAACGGTATTGTTTTTGGTGCTGTGAGCCAAGGTGGCCTGCT GGAGCTGCGTAAGCACTACGACTTCTTTTGCAACCTGCGTCCG ATCCGTATCATTGATAGCCTGGTTCACAAGAGCAGCCTGCGTC</p> |

| | |
|--------------|---|
| | <p>CGGAGAAAGTGCAGGGTCTGGACATTCTGATCATTCTGTGAAC GGTTAGCGGCATCTATTTTCGGCCCGGCGGGTCGTAGCAGCGAC GAAAAGGCGCGTACGGTTATCACACCATGCTGTACTATGATG AGGAAATCCGTCGTATTGCGCGTATCGCGCTGCAAAGGCGCA AGAGCGTCGTGGTCTGCTGACCGTTGCGCACAAAGAGAACGCG CTGCCGCACCTGCCGTGGACCCGTATTGTGCAGGAAGAGGCGA TCAAGTTCCCGGGCGTGGTTGTGGAGCCGATGCTGGTTGACAA CCTGGCGATGCAGATGGTGTGTAACCCGCAACGTTTCGATACC GTTCTGGCGGGTAACATGTTTGGCGACATTCTGAGCGATATCG GTGGTGCCTGGTGGTAGCATTGGTCTGCTGGGCAGCGCGAG CCTGAACGCGGACGGCTTTGGTCTGTATGAGGCGATCCACGGC ACCGCGCCGGATATTGCGGGCAAGGGTATTGCGAACCCGCTGG GCACCCTGGGTGCGTGCCTTCTGATGCTGCAACAATGGGGTGA AGTGGGTGCGGCGCAGCGTATCATTGCGGCGCAGGAGCAAGTT CTGAGCGAAGGTTATCGTACCGCGGACCTGTTCCCGCAAGGCA GCGAGATTTTGTGAACACCGAAAACTGATCGACCTGCTGCT GGAGGAACTGAGCAGCTAA</p> |
| <i>NshpA</i> | <p>ATGCAGACCCTGCCGATCAAGATTAGCGATGAAACCCTGCGTG ACGGCGAGCAGCAAGTGGGCGTTTTCTTTAGCCTGCCGACCAA GCACAAACTGGCGCACCTGATCGCGCAGACCCGGTGTGGATGA ACTGTGCATTATGCCGAGCGTGTGCGAGCAAGAAGCGGAGCTG GTTAAGACCCTGATCAGCGAGGGCTGGGACGATCTGATTACCG CGGACACCATGATGGGTAAAGATTACATCGACTATGCGAAGA ACTGCGGCGTGAAACGTATCATTCTGTTCCACGCGGTTAGCGA TCGTCTGCTGTTTCTGCGTAACCCGCACGTGCGTCTGATGAGCG AATTCAGGACAAGACCATCGACGATGACATTCCGAGCCACAT CATTAAACCAGATCCGTCAAGATGCGCTGGACGTGATTGTTGAA AACCTGCACTATGCGACCAAAGTGGCGGGTCTGCGTGTTGATT TTGCGGCGGAGGATGCGAGCCGTGCGGATTTGACTTTCTGGT TCAATGCATCCGTAGCTGTAGCCCGTACATTGAACACTTCCTGC TGTGCGATAACCGTGGGTGTTCTGAGCCCGGAGAAGAGCTATAT CTGGATTAACGACCTGCTGCAAAGCACCACCGGTGTGGCGTTC GGCGTTCACTACCACAACGATATGGGCCTGGCGCTGGAAAACA CCCTGCAAAGCCTGATGGCGGGTGCAGCCCTGGTTAGCGGCAC CTTTTGCGGTCTGGGCGAACGTGCGGGTAACGTGGCGCTGGAG CAGGTTCTGAACGGCCTGCGTGTGCGTTTTCGGTATCGAAGTTG AGGGCATTAACTACGATGCGGTGGACGCGGTTACCAACTATAT CGAGGAAATGGGTATTCGTCCGGCTCCGCCGTATAGCCAGACC AGCCAACGTCACGAGAGCGGCATCCACGTGAACAGCCTGTTCC GTGATCCGCAGAGCTACGCGATCTTTCCGCACAACAAGATTGA CGTGGTTTTTCGGTAAATGGAGCGGCGTTAGCAACTTCCAGTAC CTGTTTGA AAAACAGCTGCAAACCCGCAAACCCGTCAGCAAT ATGAGAAGATGCGTAGCGTGATCAAAGCCTGGCGACCGAAC AAGAGTGCTACTTTACCGCGAACGAAGTTCTGAAGCTGTGGAA</p> |

| | |
|----------------|--|
| | AGACGGTATTTTCGAGTAA |
| <i>NshphCD</i> | <p> ATGAGCAGCATCAACAAGGAAGTGACCAAGAAAGTTCTGTAC CTGGGTGACGATATTAACACCGACGATATCATTCCGGCGAACCC GTGCGACCACCTACGACCCGGATATCCTGAAGCAGTATGCGCT GGAACACCTGATTGGTGTGGGCGAGCTGCTGAAATACAACGTG ATCGTTGCGGGCGAAAACCTTCGGTTGCGGCAGCAGCCGTGAGA TCGCGCCGGTTGCGCTGAAGGCGGGGATCGAAAAAATTCA GGCGCGTAGCTTCGCGGAAATCTTCTACCGTAACAGCATCAAC ATTGGCCTGAACCTGGAGATTCTGGAGGAAAAACAGGAAAAAC CCGGTGGTTGAGGCGATTGCGCAAGCGGGTGGCCTGATTCCGT TTAACCAGATGCGTCGTCAAGGCAAGATCACCACCCCGAAAAG CGTGACCCCGCAGCGTCCGATGACCCTGGTTGAAAAGCTGATC GCGAAAGCGAGCGGTAACACCTACGTGCGTCCGGGCGAGGTG GTTTTGCGCAAGTTGACCTGGCGCTGAGCCATGATGCGGTGG CGGGTCCGGTTGCGACCACCTTTTACAAGCACTATGGCAAGGA AGCGAAACTGTGGGATGCGCAGCGTGTGGTTCTGGTGGCGGAT CACTTCATCCAAGTTAACGACATTCGTAACGATCAGAAAGCGA ACGTGATGTATCAAGAGATGATCGACTTCGCGAAGAAACAGG GTTGCCACCTGTTTGATATCGTTAGCCCGGGTGAAGCGGCGGG TATTTGCCACGTGCTGCTGCCGGAGAAGGGCTTCGTTCTGCCG GGTATGATCATTGCGGGCACCGACAGCCATACCTGCACCTACG GTGCGCTGGGTGCGTTTAGCGCGGGTGTGGGCACCACCGACAT GGCGAACATCTATGCGATGGGTGATATGTGGATTCTGTGCGG AGCACCTGGTTTTCGAACTGAGCGGCACCCTGCCGCCGACA TCAGCGCGAAGGACATCATTCTGTTTATCCTGGGTCAGATTGG CTGCGCGGGTGCGACCAGCAAAGTTATGGAGTTCCGTGGTAGC ATCCTGGAACAAATGCCGTTTGATGAGCGTCTGACCCTGGCGA ACATGGCGATTGAATGCGGGCGGATCTGCGGTCTGATTGTGCC GGACGAGACCACCCGTAACCTACGTTTCGTAGCCGTAGCGATCAG GAGTTCGAGGAACTGATCGCGGACCCGGATGCGGAGTACGAA AAGGTGTATAAATTTGACATCAGCAACCTGGAGCCGCAAATTG CGCGTCCGCCGAAACCGGATCAGGTGGTTCCGATCAGCCAAT TGAGGAAACCCCGATACCAAGGCGTTCATTGGCAGCTGCACC GGTGGCAAACCTGTACGACCTGGCGCAAGCGGCGGAAGTTCTGC AAGACCGTCAAGTGGCGGATGGTGTAACTGTTTCATCGTGCC GGCGAGCGTTGAGATTCGTGAAAAGGCGGAGGCGCTGGGTTA TATGGATATCTTTGCGCAAGCGGGCGCGAAGATTCTGAAAAGC GGTTGCGGCGCGTGCATCAACAGCGGTATTGGCGTGCTGAACA AAGAGGAAACCGGTGTTTACGCGACCAACCGTAACTTAAGGG CCGTAGCGGTGACCCGACCGGCAAAAACCTATCTGGCGAGCCCG CGTACCGTGGCGATCAGCGCGGTTAAGGGTAAAATCAGCCACA TTCTGGATTAA </p> |
| <i>NshphB</i> | <p> ATGCGTAGCAGCAGCAACCAGAGCTACCGTATCGTTGCGATTCC CGGGCGAGGGCATCGGTCCGGAAGTGGTTGCGGCGAGCCTGC </p> |

| | |
|--------------|---|
| | <p>AACTGCTGCAACAAGTTGCGAAGCTGGAGGGTTTCACCCTGCA AGTGGACTATGGTTGGCTGGGCACCACCGCGCTGGAGAAATC GGCACCTATTTCCGCAAGCGACCGCGGAACTGTGCAACGGCA GCGATGGTATTGTGTTTGGTGGCGTTACCCAGGGTGGCCTGCT GGAAGTGCCTAAGCACTACGACTTCTTTTGCACCTGCGTCCG ATCCGTATTGTGGATAGCCTGGTTAACAAAAGCAGCCTGCGTC CGGAGAAGATCAAAGGTCTGGACATCCTGGTGATTCTGTAACT GGTTAGCGGCATTTATTTCCGGCAGCGCGGGTCTGTGCGAGCGAT GAGAAGGGTGGTACGGTTATCACACCATGCTGTACTATGATC ACGAAATCCGTCTGCTGGCGCGTCAGGCGCTGCAAAAAGCGCA GCAACGTCGTGGCAAGCTGACCGTTGCGCACAAAGAAAACGC GCTGCCGAACCTGCCGTGGACCCGTCTGGTGCAGGAAGAGGCG GCGCAATTCCCGGACGTGATCGTTGAACCGATGCTGGTTGACA ACCTGGCGATGCAGATGGTTATGAACCCGCAACGTTTCGATGT GATTCTGGCGAGCAACCTGTTTGGCGACATCCTGAGCGATATT GGTGGCGCGCTGGTTGGCAGCCTGGGTCTGCTGGGCAGCGCGA GCCTGAACGCGGACGGCTTTGGTCTGTATGAGGCGATCCACGG CACCGCGCCGGATATTGCGGGCAAGGGTATTGCGAACCCGCTG GGCACCATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG AAAAACGTGCGGCGCAGATCATTATGGTGGCGCAAGACCGTGT TCTGGAGCAGGGTTATCGTACCGCGGATCTGAGCCCGCAAGGC AGCGAGATCCTGGTTAACACCGAACAGCTGATCCAAGTCTGCTG TGACCGAAATTTACAACCTGCAAAAACACCGAGATTGGTGTGCT GCACGAATTTTATAAGTAA</p> |
| <i>PahpA</i> | <p>ATGCAAACCTCTGCCGATTAATAATTTCTGATACCACCCTGCGTG ATGGTGAGCAACAGGCTGGTCTGTTCTTCCCGTATGAAATGAA ACAAGAATTTGCTCATCTGATTGCCCAAACCGGTGTGTCTGAG ATCGAAATCATGCCGTGTGTTTGTGATGACGAAACCAAACCTGG TTAAAACCCCTGGTAGCGCAAGGTCTGAAGCATCAAATTGTGGC TGCCACCCCGATCAAAAAGGACTTTATCAACATCGCTAAAGAC TGTGATCTGGGTCAGATCGTACTGTTTAAAGGTCTGTCTGATCG CCTGCTGTTTCTGCGCGATCGTGAAATCTCTCAGATGAAAGAG TTTCAAGGTAAGACCATTGATGATCATATTCCGGACCACATTA TTAACCGTATCCGTCAAAAACGCCATTGATGAAATTGTTAACCA TCTGCGCTATGCAACTCAGGAAGCAGGTCTGCGTGTAATAATTT GCGATCGAGGATGCCTCTCGTGCCGATTTTGACTTTCTGGTTGA GTGCCTGCGTACCTTCTCTCCGTATGTTGAAGACTTTTACCTGT CTGATACCGTTGGTGTCTGACTCCGGAAAAATCTTATATTTGG GTTAACGATCTGCTGCAAAGCACTACCGAGGTAAACCTGGGTG TACACTACCATAACGATCTGGGCCTGGCTCTGGAAAACACCCT GCAATCTGTTATGGCAGGCGCGACCATCGTTTCTGGTACTTTTG GCGGCATTGGTGAGCGTTCTGGTAACGTTGCCATTGAGCAGGT ACTGAACGGCCTGAAAATTCGCTTTGGCATTGAAGTTGAAGGC ATTAACATGAAGCGATCGCCCCAGTTCTGGATTATATGGAAC</p> |

| | |
|----------------|---|
| | <p>AAAAAGGCATTCGTCCGGCCGCTCCGTA CTCCCAACAAGCACA ACGCCATGAATCTGGCATTACGTTAACTCTCTGCTGGTAGATC CGAAATCTTATGCCGTTCTGCCGCACAACACCATTGAAATTCT GTTTGGTAAATGGTCTGGTGCCAGCAACTTCCGTTATCTGTTTG AAAAAGAACTGAACCATCCGCAACCGAAGGAACAATATGAGA AAATGCGCTCTGTGATCAAAACCTGGCTGTAGAACAAGAACG TTATTTTACCGCCGAGGAGGTGGTGGAGCTGTGGAAGAATGGC ATCTTCGAATAa</p> |
| <i>PahphCD</i> | <p>ATGACTGATGCCAACAAAATCCTGAACCTGGGTAACGATATTA ACACTGACGATATTATTCCGGCTAAACGTGGTACTAACGATGA TCCAGACCACCTGAAACAATATGCTCTGGAACACATCATTGGC GTGGGTGAACTGCTGCAATATGATGAGATCGAAGCAGGTAAC AACTTTGGCTGTGGTTCTTCTCGTGAAATTGCCCCGATTGCCCT GAAAGCGGCTGGCATCAAAAAAGTTAAAGCGCGTTCTTTCGCC GAAATTTTCTATCGCAACTCTATTAACATTGGTCTGCCGCTGGA AATTATTGGTGAGACCGAACAAAACCCGGTGGTAGAAGCGAT CGCCACTGCTGGTGGGCTCATGGCCTTCAACCAAAGAGACGT CAAGGCAAGGTTAAAATTCCGCCGTCTGTTACCCCGGCCCGTC CGATGACCCGTGTGGAAAAGATGCTGGCAAAGCTAGTGGCA ACGATTATGTGAAACCAGGTGAAGGTGTGTTTGCCAAGATTGA TCTGGCTCTGTCCCATGACGCGGTGGCTTCTAGTGTGCGAAG GTATTCTACGACA ACTATGGTAAAACCGCCCAACTGTGGGACC CACAACGTGTGGTACTGGTTGCTGATCACTTCATCCAAATCAA CGACATTCGCCTGGATAACAAAGCCCCGGTTATGTATGAGCAG ATGGTTAAATTTGCCCAAGCTCAAGGCTGTCATCTGTTTGATCT GGTTTCCCAGGTGAAGCGGCTGGTATCTGTCATGTTCTGCTGC CAGAAAAAGGTTTTATTCCGCCGGGTATGATCATTGCTGGTAC TGACTCCCATACTGTACCTACGGCGCTCTCGGCGCCTTCTCGA CCGGTGTGGTACTACCGATATGGCCAACATTTTGGCAATGGG TGATATGTGGATCAGGGTTCGCCCACTCTGGTGTTCGAGCTG AGCGGTACTCTCCCACCGCAGATAAGCGCCAAGGACATAATCC TGTTTCATCCTGGGCAAACCTGGGCTGTGGTGGTGCCACCTCTAA AGTTATGGAATTCGTGGCTCTATCATTGAACA ACTGCCGCTG GATGAACGCCTGACCCTGGCTAACATGGCCATTGAATGTGGTG CTATGTGTGGTCTGATTGCGGCGGATGAGGTGACCAACCATTA TGTTACCTCTCGCACCCCAATTGGTTTTGAAGACGCGATCGCTG ACCCGGATGCCGAGTATGAGGCTGTTTATAAGTTTGACCTGAG CCATCTGGAACCGCAGGTTGCCCGTCCGCCGAAACCAGATCAG GTAGTTAACATTTCTCAACTGGAAGATGTTCCGATCACC AAAG CCTTCATCGGCTCTTGTACTGGCGGTAAACTGTACGATCTGGCC CAAGCGGCGGCTGTTCTGAAAGATCGTCAGGTTGCCAGGGTG TAGATTTATTTGTAGTGCCAGCTTCTATGGAGGTACGTCAAAA AGCCGAAGA ACTGGGCTATCTGGCTATTTTGAAGAATCTGGT GCGCAACTGCTGAAATCTGGTTGTGGTGCTTGTATTA ACTCTGG</p> |

| | |
|---------------|---|
| | TAAGGGTGTACTGGACAAGGAAGAAACCGGTGTTTATGCCACC AACCGTAACTTTAAAGGTCGTTCTGGCGACCCGACCGCTAAAA ACTATCTGGCTTCCCCGCGTACTGTTGCAATTTCTGCTATTTAA GGTAAAATTACTGCCAACCTGGATTAA |
| <i>PahphB</i> | ATGGCGAGCCTGAACAAGAGCAGCTACCGTATCGTGGCGATTC CGGGCGAGGGCATCGGTCCGGAAGTGGTTCAGGGCGAGCCTGA CCATTCTGCGTCAGGTTGCGCAACTGGAGGGTTTTGCGCTGAC CATCGACTATGGCCTGCTGGGTGGCATGGCGCTGGAAAACTG GGTAGCACCTTTCCGCCGGAACCAGCCAGCAATGCGAAGGC AGCGACGGTATTGTTTTTGGTGCGGTGAGCCAGGGTGGCCTGC TGGAGCTGCGTCGTCACCTTTGATTTCTTTATCAACCTGCGTCCG ATTCGTAGCTGCCCAGCCTGCTGCACACCAGCAGCCTGCGTC CGGAGAAGGTGCAAGCGCTGGATATCCTGATTGTTGCTGAACT GGTGAGCGGTATCTATTTTGGTCCGGCGGGTCGTAGCACCGAC GAGAAAGGCGCGTACGGTTATCACACCATGTGCTACTATGATC GTGAAATCCGTCGTATTGCGAAGGTGGCGCTGCAGAAAGCGCA AGAGCGTCGTGGTCTGCTGACCGTTGCGCACAAGGAAAACGCG CTGCCGCAGATCCCGTGGACCCGTCTGGTTCAAGAGGAAGCGA CCAACCTTCCGGACGTGGTTGTGAGCCCGATGCTGGTTGATAA CCTGGCGATGCAGCTGGTGCTGAACCCGCAGCAATTCGACGTT ATTCTGGCGGGTAACATGTTTGGCGACATCCTGAGCGATATCG GTGGTGCCTGGTTGGTAGCATCGGTCTGCTGGGCAGCGCGAG CCTGAACGACCAAGGCCTGGGCATGTATGAGGCGATCCACGGC ACCGCGCCGGATATTGCGGGCAAGGGTATTGCGAACCCGCTGG GCACCCTGGCGGGTTGCGTTCTGATGCTGCAGCAATGGGGCGA GATCAAGGCGGCGCAGCGTATTCTGAAAGCGCAAGAACGTCT GCTGGGCAAAGGTTATCGTACCGCGGACCTGTTCCCGCAGAAC CACGAAACCCTGGTTAACACCGCGACCCTGGTGGACCTGTTCC TGGAGGAAATCAGCCTGGAGCAACCGACCGAATAA |

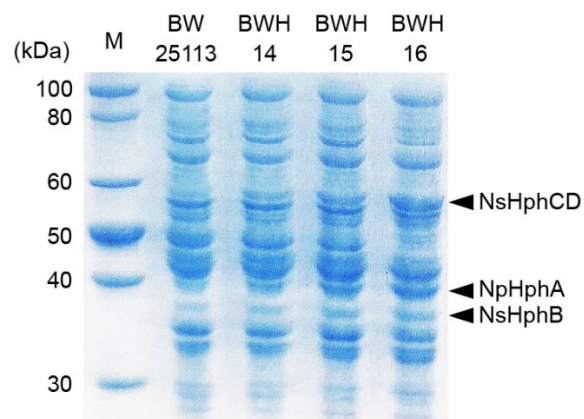


Fig. S1: SDS-PAGE of the soluble cellular proteins extracted from strains *E. coli* BW25113 (control), BWH14, BWH15 and BWH16 at 12 h cultivation. The predicted bands of the target proteins NpHphA, NsHphCD and NsHphB are arrow-pointed.

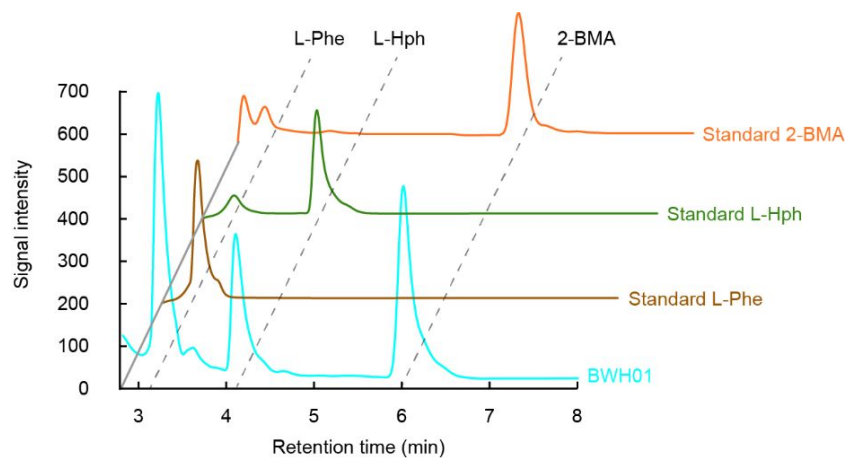
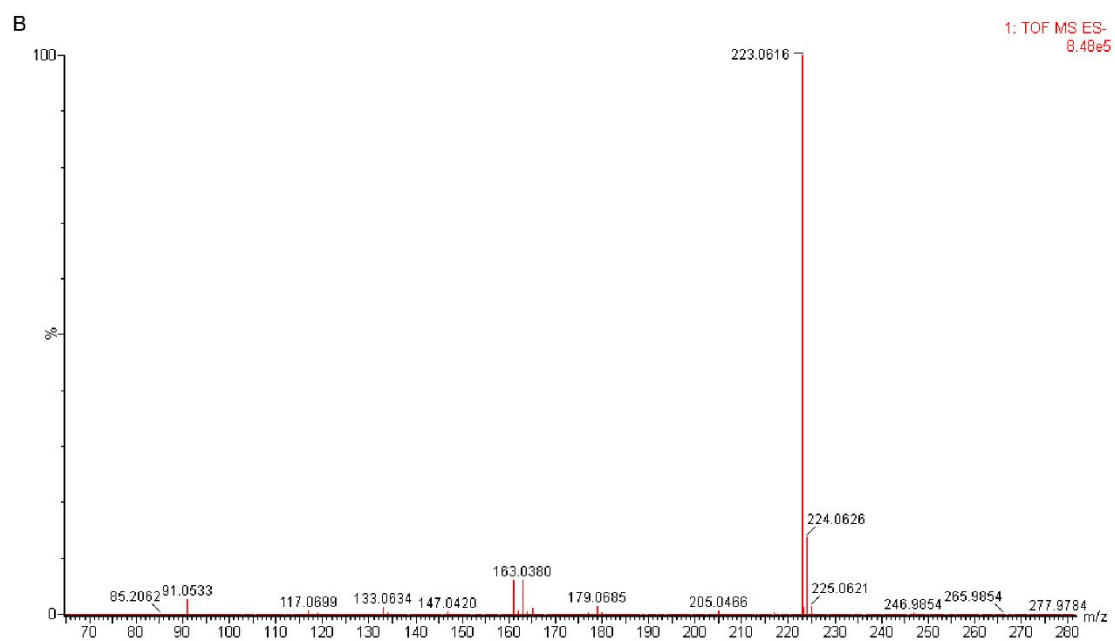
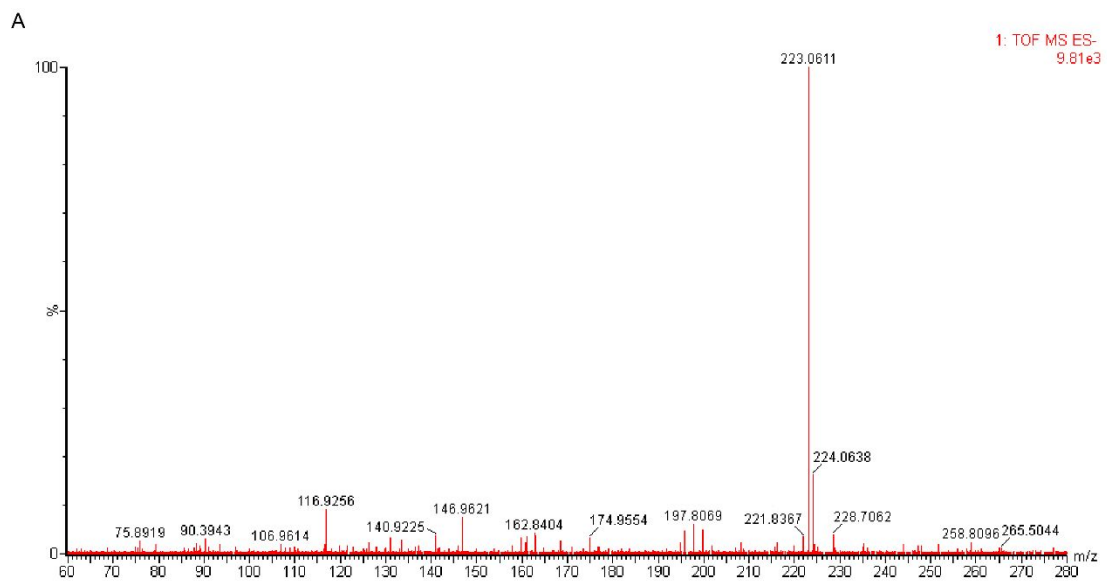


Fig. S2: HPLC analysis of fermentation broth of strain BWH01 (expression of L-Hph BGC from *N. punctiforme* PCC73102).



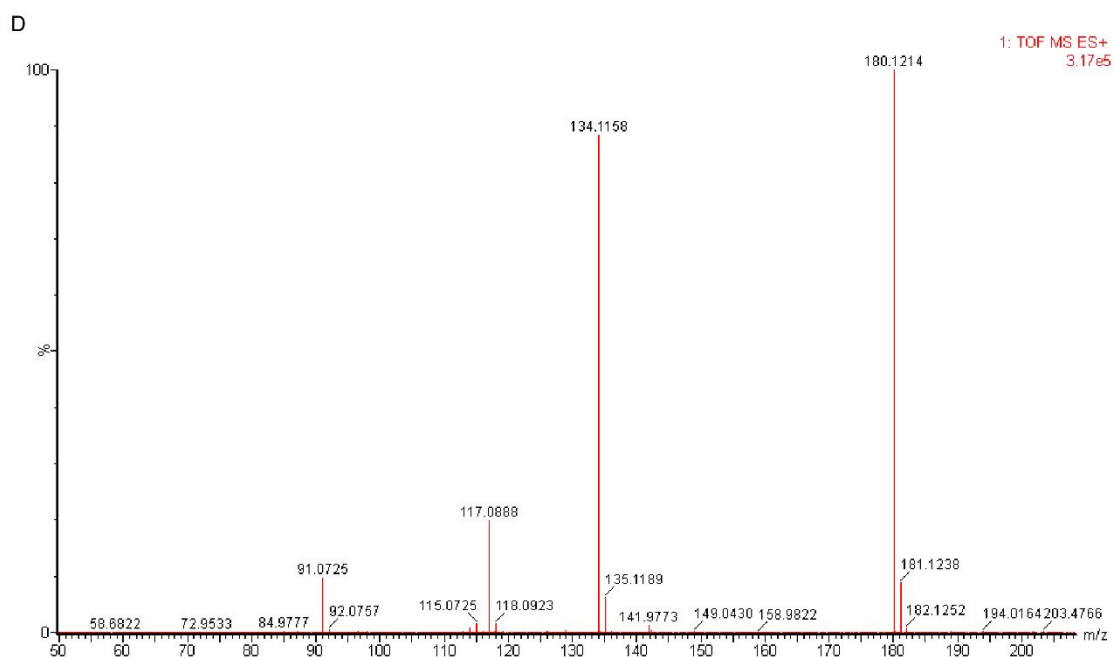
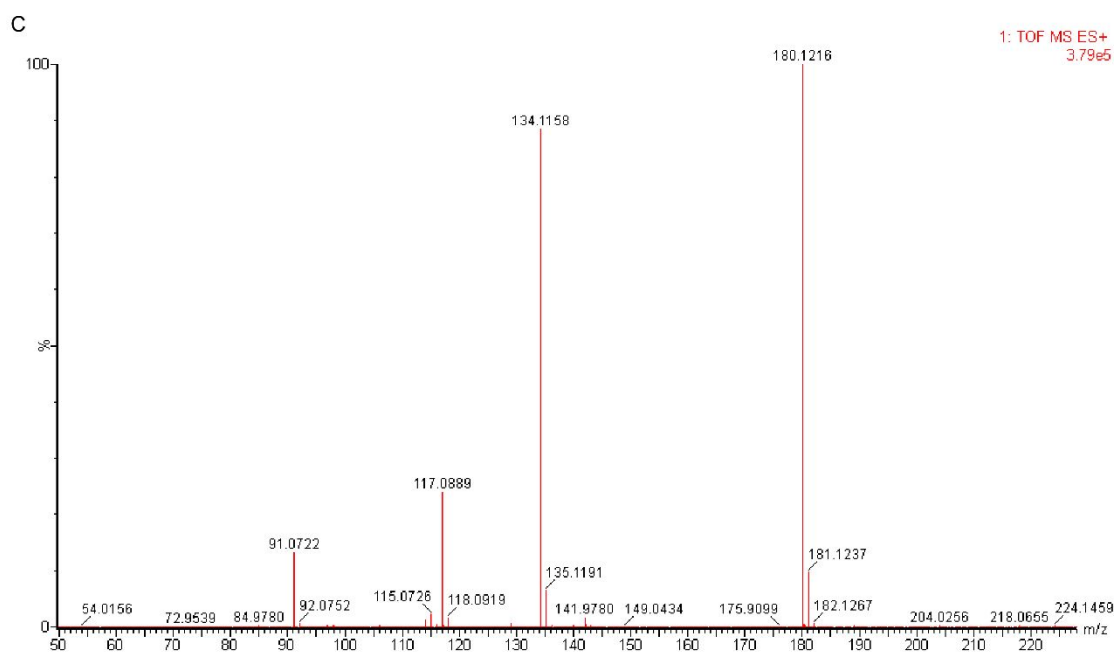


Fig. S3: LC–MS analysis for identification of 2-BMA and L-Hph. (A) LC–MS analysis of the standard 2-BMA. (B) LC–MS analysis of 2-BMA from the fermentation broth of strain BWH01. The mass of major product ($[M - H]^- = 223.1$) is identical to the theoretical mass of 2-BMA ($[M - H]^- = 223.1$). (C) LC–MS analysis of the standard L-Hph. (D) LC–MS analysis of L-Hph from the fermentation broth of

strain BWH12. The mass of major product ($[M + H]^+ = 180.1$) is identical to the theoretical mass of L-Hph ($[M + H]^+ = 180.1$).

References

- (1) Fujii, K., Sivonen, K., Nakano, T., Harada, K. I. (2002) Structural elucidation of cyanobacterial peptides encoded by peptide synthetase gene in *Anabaena* species. *Tetrahedron*. 58, 6863-6871.
- (2) Sanz, M., Salinas, R. K., Pinto, E. (2017) Namalides B and C and Spumigins K-N from the Cultured Freshwater Cyanobacterium *Sphaerospermopsis torques-reginae*. *J Nat Prod*. 80, 2492-2501.
- (3) Fujii, K., Sivonen, K., Adachi, K., Noguchi, K., Sano, H., Hirayama, K., Suzuki, M., Harada, K. I. (1997) Comparative study of toxic and non-toxic cyanobacterial products: novel peptides from toxic *Nodularia spumigena* AV1. *Tetrahedron Lett*. 38, 5525-5528.
- (4) Okumura, H. S., Philmus, B., Portmann, C., Hemscheidt, T. K. (2008) Homotyrosine-containing cyanopeptolins 880 and 960 and anabaenopeptins 908 and 915 from *Planktothrix agardhii* CYA 126/8. *J Nat Prod*. 72, 172-176.
- (5) Shin, H. J., Matsuda, H., Murakami, M., Yamaguchi, K. (1997) Anabaenopeptins E and F, two new cyclic peptides from the cyanobacterium *Oscillatoria agardhii* (NIES-204). *J Nat Prod*. 60, 139-141.
- (6) Sano, T., Usui, T., Ueda, K., Osada, H., Kaya, K. (2001) Isolation of new protein phosphatase inhibitors from two cyanobacteria species, *Planktothrix* spp. *J Nat Prod*. 64, 1052-1055.
- (7) Sano, T., Kaya, K. (1995) Oscillamide Y, a chymotrypsin inhibitor from toxic *Oscillatoria agardhii*. *Tetrahedron Lett*. 36, 5933-5936.