SUPPLEMENTAL INFORMATION

ArsAB, a Novel Enzyme from *Sporomusa ovata* Activates Phenolic Bases for Adenosylcobamide Biosynthesis

Chi Ho Chan, and Jorge C. Escalante-Semerena*
Department of Bacteriology, University of Wisconsin
1550 Linden Drive, Madison, WI 53706-1521 USA

Running title: the α-ribotide synthesis (*arsAB*) genes of *S. ovata*

*For correspondence: E-mail: escalante@bact.wisc.edu, Tel: (+1) 608 262 7379; Fax (+1) 608 265 7909

Figure S1. Scheme for the identification of *S. ovata arsA* and *arsB* genes.

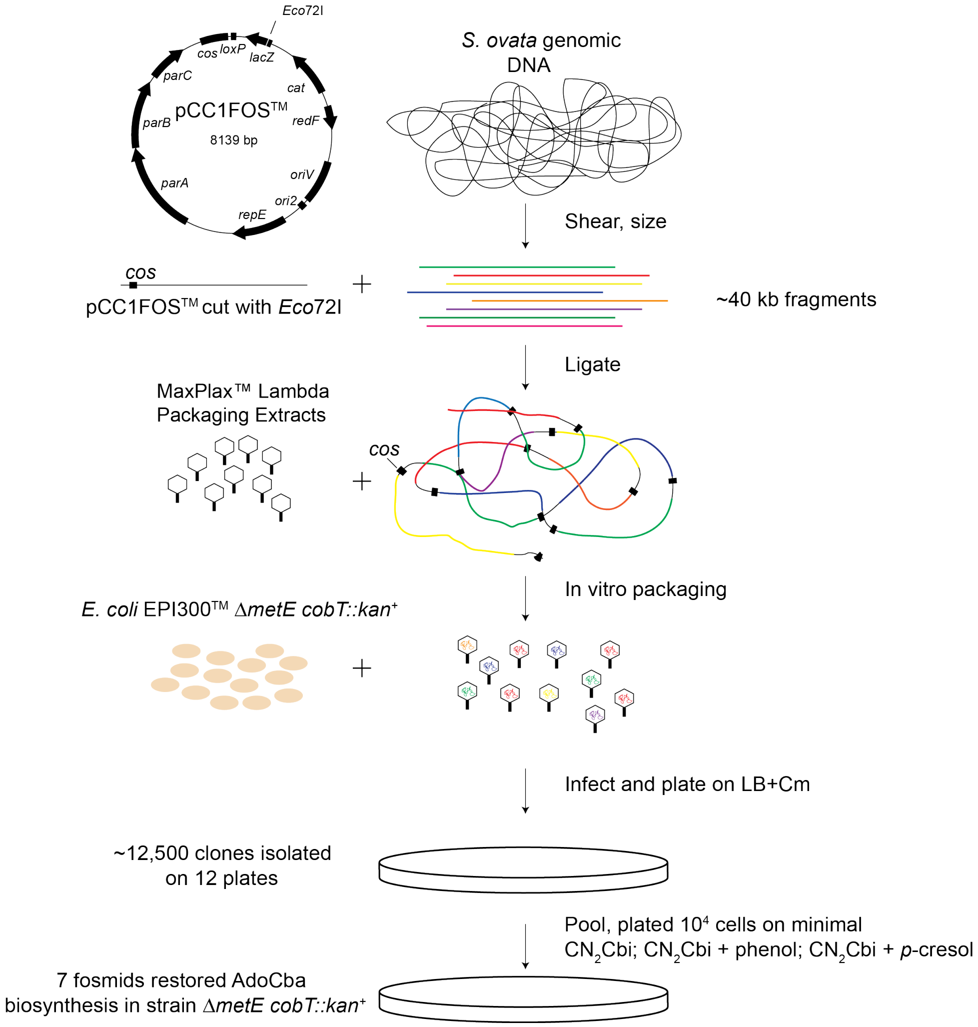

Figure S2. Nucleotide sequence of the *arsA* and *arsB* genes of *S. ovata*. Start codons (ATG) are highlighted in green, stop (TAG, TGA) codons are highlighted in red. *arsA* and *arsB* are 2,098 bp long (including stop codons); the entire sequence (as shown) = 2,267 bp

```
CAAGGGGAAGGCGTAAGAGTTAATGAACTGGGAGCCAGGAGACCTGCCTTATACCGCAACCATCCT
TCATGGACAAGAAGGCTTTTTTGGGCTTAACCAAAAAAGCCTTTTATTTTTATCCAGATGGAAAGT
GGATAAAAACAAAAGACTTCTTTTAAGTATACTAGTAATGGAGGTTATTATGAGTTTACTGCAAGC
AACAGTAGCGAAAATTATGAGGCCGGATACAGTAATTAAAGATCAGGTAAAGACTAAACTGGCAGG
TGTGTTACAGTCAGCAGGCAGCCTGGGGCGATTGGAAGATATGGTTGAGCAGTATGCGGGAATAAC
AGGCGAACTGAATCCGGCATTGCCTAAGCCTTGTATGGTGGTAGCTTCAGCTGACCATGGGGTGGC
CAGACGAGTGGTGAGTGCCTATCCGATAGAAACAACAATTCATATGACAGCTAATTATCTTATCTC
CCAAGGTGCAAGTGCCAATGCCTTTGCGAACTTCTGCGGGGCTGATATGGTTGTGGTTGATATGGG
CGTGGCTGGCGATTTGTCTTATGTTCCGGGACTGTGGCATCGCAAGATTGCCTATGGCACACAGGA
TTTTACCGAAGGCCCGGCAATGACGCGTGAGCAGGCCATCCAAGCGGTGGAAACAGGCATTGATAT
TGTTAATGACCGGGTGAAGCACGGCAATCGCTGTTTTTGTCTGGGAGAAATGGGCATTGGTAATAC
CACGTCCAGCGCAACCATTGTTGGGGCATTCACCGGTCTCGCTCCGGAGAAAGTTACCGGGCGGGG
CACCGGGATATCAGACAGCCGTCTTAAAACCAAGATGGAGATTGTCGGTCGGGCCTTAGCTGTGAA
TAAACCGAATCCTCAGGATGGATTAGATGTTCTGGCCAAAGTGGGTGGATTTGAACTAGGTGCATT
AGCCGGCGTAATTCTTGGCTCAGCCGCCAACCGCTGCGCTGTCGTTATTGATGGGCTTAACACGAC
GGCGGCGGCTTTAATCGCCAATGTCATTCATCCTCTGAGCAAAGAGTATATGTTTGCGTCGCACCT
GTCCGGTGAACCAGCTCATAGTATTGCGTTGCGTCAGCTTCAGTTAGAAGCTTGTCTGGAGCTGGG
AGTCCGGCTGGGAGAGGGTATCGGGGCTTCTATGGTTGTCGATATGCTCTATGTGGCCATTAAGCT
GCTAAATAATAGGGGAGGTAAAGCCAATGCTTGAAGAACTAATTGCAGCCATAAAACCACTTGATA
GTATAGCGATGGAACAGTGTCAGCGGCGAGTCGATAATCTGACCAAACCGTTAAACAGTCTGCATT
CCTTTGAACACATTGCCTGTAAGCTGGCAGGAATTAGCGGCAATCCCCGGCCACGGGCGCTGGAAA
AAAGCATCATTATTATGGCCGCTGATAATGGCGTAGCAATGGCAACTGACCAACAGCAGATGACAA
CAGCAGCCCGACTGACCGGCTTCTGCCAGGGGCAGGCGCCCATCCAGGTGTTTGCCGCCCATGTTC
AGGCCAGGCTGATTATGGTGGATATTGGTGTTGCTGCCGACCTGCCGCATTCTCCGGCAGTATGCC
GGAAAAAGCTGGCTTATGGCAGTCGCAATAGTACCGAAGGTCCGGCTATGACCCGCCAACAAGCGA
TACAAGCCATTGAAGTGGGAGTAAGAATAGCACAAGCAGAGATTGCGCGCGGCTGCCAGGTGATCG
GCTTGGGCGAAATGGGGCTGGGCGGTTTGGCTGCCGCTATGGCAATTGTCGCCTGCTGCCATGGTC
AGCCACTACCCGGCTTAGCCGGACGGGAGGCAGAACTTGTCAATACAGCTATCGCAGTCAACCGCC
CGAATGCTGCTGATCCACTGGATATTCTGACTAAGGTAGGTGGTTTGGCTATCGCCGGGTTAGTTG
GCGTTATTCTGGGGGCTGCGGCCGGGCGGGCGGCCGTGGTGTTAGACGGTTTAGCCACAAGTACGG
CCGCCCTGATTGCTATAAACCTGGTTCCAGACGTTAAGCCTTATTTGATAGGCTCCCACTTTGCTG
CCGAGCCAGCCCATGAGACAGCCTTGGCATTGCTTGACGTACCGGCCTATCTCCAACTTAAAATGA
ACCTGGGAGAAGGGACAGGTGCAGCACTCGGGATGTCAGTAATTAACGCCACGCTGCATATGCTAA
ATGACATGAAAACCTTTGGTGAGGCCGAAGTCGCTGTGGCGCAGGATGGACCAGGTGCTCTGCGGC
AAAGCAAGGATGTTAGAGATTAGCAAGC
```

Primary sequences were obtained using ExPASy Translate tool available online at http://ca.expasy.org/tools/dna.html. The putative catalytic Glu residue is highlighted in blue. Each protein is 350 residues long.

ArsA (calculated mass = 36.65 kDa):

M E V I M S L L Q A T V A K I M R P D T V I K D Q V K T K L A G V L Q S A G S L G R L E D M V
E Q Y A G I T G E L N P A L P K P C M V V A S A D H G V A R R V V S A Y P I E T T I H M T A N
Y L I S Q G A S A N A F A N F C G A D M V V V D M G V A G D L S Y V P G L W H R K I A Y G T
Q D F T E G P A M T R E Q A I Q A V E T G I D I V N D R V K H G N R C F C L G E M G I G N T
T S S A T I V G A F T G L A P E K V T G R G T G I S D S R L K T K M E I V G R A L A V N K P N
P Q D G L D V L A K V G G F E L G A L A G V I L G S A A N R C A V V I D G L N T T A A A L I A
N V I H P L S K E Y M F A S H L S G E P A H S I A L R Q L Q L E A C L E L G V R L G E G I G A
S M V V D M L Y V A I K L L N N R G G K A N A

ArsB (calculated mass = 36.18 kDa):

M L E E L I A A I K P L D S I A M E Q C Q R R V D N L T K P L N S L H S F E H I A C K L A G I S
G N P R P R A L E K S I I I M A A D N G V A M A T D Q Q Q M T T A A R L T G F C Q G Q A P I
Q V F A A H V Q A R L I M V D I G V A A D L P H S P A V C R K K L A Y G S R N S T E G P A M
T R Q Q A I Q A I E V G V R I A Q A E I A R G C Q V I G L G E M G L G G L A A A M A I V A C C
H G Q P L P G L A G R E A E L V N T A I A V N R P N A A D P L D I L T K V G G L A I A G L V G
V I L G A A A G R A A V V L D G L A T S T A A L I A I N L V P D V K P Y L I G S H F A A E P A H
E T A L A L L D V P A Y L Q L K M N L G E G T G A A L G M S V I N A T L H M L N D M K T F G
E A E V A V A Q D G P G A L R Q S K D V R D

Figure S3. UV-visible spectrum of CNCbl (vitamin B_{12}) provided for comparison with the spectra of phenolic Cbas shown in figure 4 in the text.

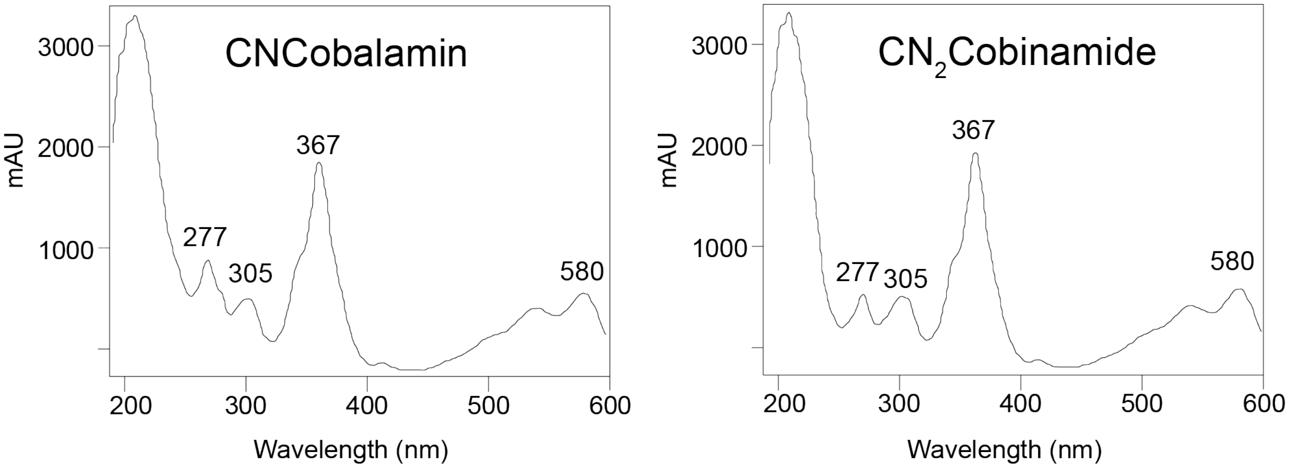

Figure S4. Isolation and mass spectrometry of α-ribosides. RP-HPLC separation and ESI mass spectra of alkaline phosphatase-treated α-ribotides synthesized by *So*ArsAB using phenol (panels A-C), *p*-cresol (panels D-F), and DMB (panels G-I). All samples were diluted 100x in 50% acetonitrile:50% water. Ammonium chloride (20 mM) was added to the samples containing α-phenolyl- and α-*p*-cresolyl-riboside before analysis. α-Phenolyl- and α-*p*-cresolyl-ribosides did not ionize without added chloride.

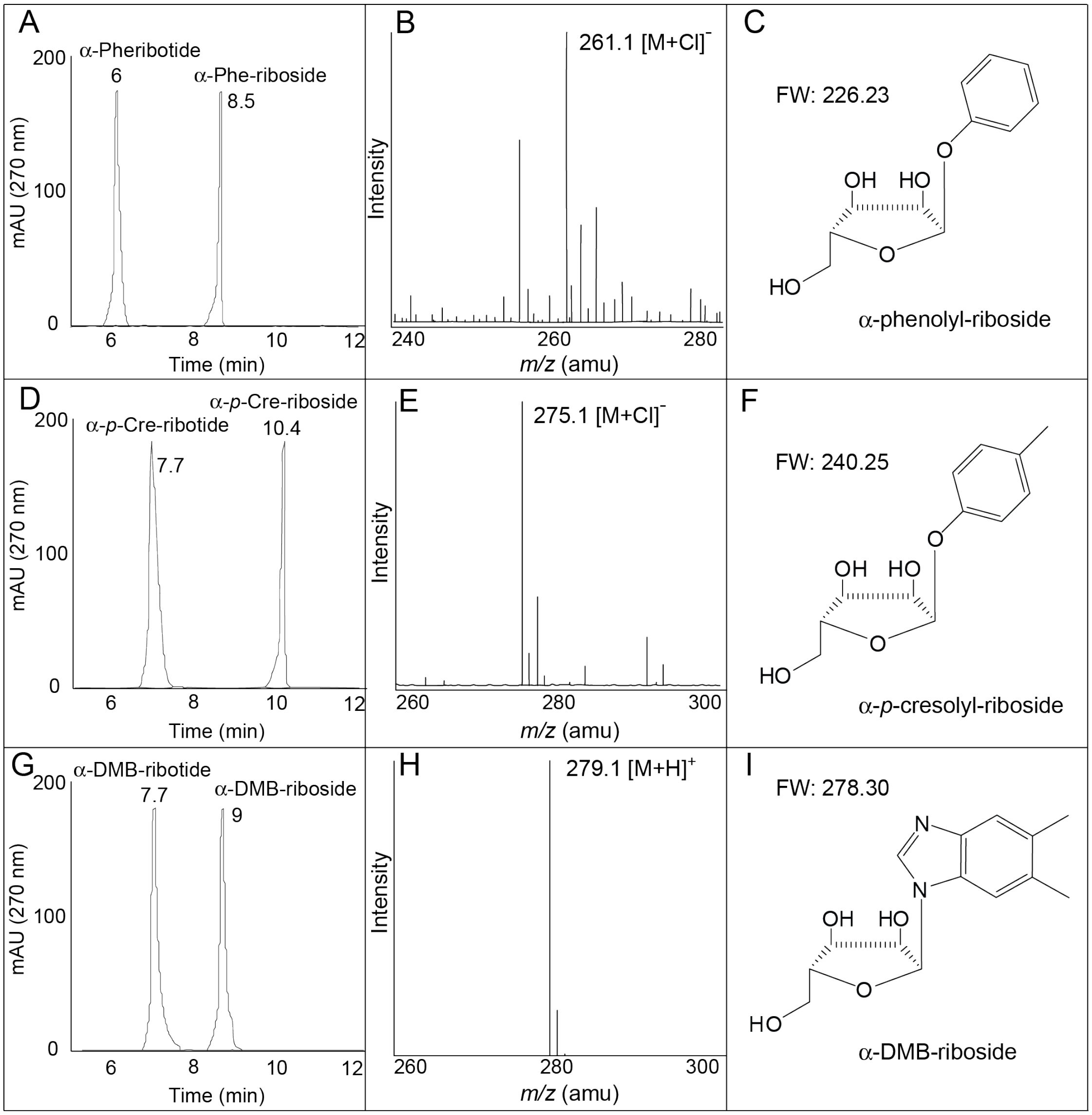

Figure S5. Alignment of ArsA and ArsB proteins of *S. ovata* with the putative proteins of *Dialister invisus* and *Veillonella parvula*. Conserved regions are highlighted. *So*ArsA is 52% identical, 70 % similar to *Di*ArsA; *Di*ArsA is 49% identical, 64% to *Vp*ArsA, and *Vp*ArsA is 49% identical, 64% similar to *So*ArsA. *So*ArsB is 33% identical, 53 % similar to *Di*ArsB; *Di*ArsB is 38% identical, 57% to *Vp*ArsB, and *Vp*ArsB is 35% identical, 52% similar to *So*ArsB.

ArsA

```
S.ovata    MEVIMSLLQATVAKIMRPDTVIKDQVKTKLAGVLQSAGSLGRLEDMVEQYAGITG     55
D.invisus  ----MGLLDETIGRIHPLDREAMYKAQKRWDDLYVGVGDLGKLEEMVIQYAGVTG     51
V.parvula  ----MSLLQETCGAITGRSLKIEQHIFNSWN-AESPVELYGRLVDVVAQYGAATN     50

S.ovata    ELNPALPKPCMVVASADHGVARRVVSAYPIETTIHMTANYLISQGASANAFANFC    110
D.invisus  EVLPEIPKCCMVVACADHGVYRQKVSAYPQSTTVGMVKSYVDVKGASANALAHYC    106
V.parvula  QEQVTVPKPCMIIASADHGVADMGVSAYPKETTVGMTQNYLIPKGAGANSLANYC    105

S.ovata    GADMVVVDMGVAGDLSYVPGLWHRKIAYGTQDFTEGPAMTREQAIQAVETGIDIV    165
D.invisus  GAHMVVVDMGINADMSDVPGLLHRKITFGTKDITEGAAMSRAEAIHAIEAGIEIA    161
V.parvula  GAQMEVIDMGIDADMSWVPGLRSHKLGMGTKNFVEEPAMTREQAVEGIETGIRLV    160

S.ovata    NDRVKHGNRCFCLGEMGIGNTTSSATIVGAFTGLAPEKVTGRGTGISDSRLKTKM    220
D.invisus  ENKIKEGYRVFTVGEMGIANTTASACILGAFNRWNAVEVTGRGTNISDARLLHKI    216
V.parvula  KEKIDEGFNVFLVGEMGISNTTASALMTAKFAGLTAEEATGRGTNISDERLKLKQ    215

S.ovata    EIVGRALAVNKPNPQD-GLDVLAKVGGFELGALAGVILGSAANRCAVVIDGLNTT    274
D.invisus  EMVQKALDVNQPDPAD-GLDVLSKVGGFEFGCMTGVMLGAAANHCMTIIDGFNST    270
V.parvula  RIVHDVLEKYKDISKDDALGILSSVGGFEFACIVGVILGAAANHGLVIIDGFNTS    270

S.ovata    AAALIANVIHPLSKEYMFASHLSGEPAHSIALRQLQLEACLELGVRLGEGIGASM    329
D.invisus  ASAFIAKALSEESVQYLMASHLSLEQAHRKSLKAIGLTEYIDLDIRLGEAVGASI    325
V.parvula  ACALVAKTLAPKAMDYVMASHLSAEKAAKSSLENLGLEAYVDLGLCLGEASGGSI    325

S.ovata    VVDMLYVAIKLLNNRGGKANA                                      350
D.invisus  QKKILDMALAVYKDGADKRTGVAE                                   349
V.parvula  QMGMLDLAVHMYMAITGGNK                                       345
```

ArsB

```
S.ovata    MLEELIAAIKPLDSIAMEQCQRRVDNLTKPLNSLHSFEHIACKLAGISGNPRPRA     55
D.invisus  -----MSGIQAPDKETMEACRLYVDNLIKPIHSLGKLEDIAVRLAGITGKIKPGK     50
V.parvula  ---MRTFKVEPLDTQAMETCRFRIDNLTKPIYSLATLETIAERFAGILGDPKPNH     52

S.ovata    L K I IMAADNGVAM-ATDQQQMTTAARLTGFCQGQAPIQVFAAHVQARLIMVD    109
D.invisus  LNKAIVIMAGDTAVDG-ENKTGGKTSLTEVQMVSRGLGTVSAVARTLGAPVYLID    104
V.parvula  LRQGVLVVAADHLVDGPQNDQHGSESYAAIKRFNEGRTATQGAAAKLNAVAHVVN    107

S.ovata    IGVAADLPHSPAVCRKKLAYGSRNSTEGPAMTRQQAIQAIEVGVRIAQAEIARGC    164
D.invisus  VGLEQNTNDIEGVLTNKVVYGTHRG--NPALDQDAVSAAISIGMSVARTLAVQGI    157
V.parvula  VGLEQDTSDLTNIEQQVIRKGSHFFGVEPAISRDELERALELGFTYADKLHADGL    162

S.ovata    QVIGLGEMGLGGLAAAMAIVACCHG----------------QPLPGLAGREAELV    203
D.invisus  QAVGLGNIGERSLLSALGVTAAIM------------------KKELQENSLKDGF    194
V.parvula  QVVAIGNVGERAFLDSLVTTATITGCAYEDILVHNEYGPTIEQRAAHIHSFVDRF    217

S.ovata    NTAIAVNRPNAADP-----LDILTKVGGLAIAGLVGVILGAAAGRAAVVLDGLAT    253
D.invisus  SLHMDDVGNMANDP-----VGVLSRVGSAEIAGLFGLVVQAAREKIAIVFDNAVT    244
V.parvula  NIAVDDWSTISESERRDAVLHLLHIAGGLDIAFLTAFILGAASHRMAVVFDNIVT    272

S.ovata    STAALIAINLVPDVKPYLIGSHFAA-----EPAHETALALLDVPAYLQLKMNLGE    303
D.invisus  GAAVLAAIEVYPEVRDYVFPSAAYN-----EPVHQIQMKKLGMKPFFYYDFTVAE    294
V.parvula  GAAVLAAVTIEPLVKDYVFPSAAYDEPIHDEPIHKEQCRFLGVKPYLDYKLLIDE    327

S.ovata    GTGAALGMSVINATLHMLNDMKTFGEAEVAVAQDGPGALRQSKDVRD            350
D.invisus  GFGSAMGLSLFDAALDMVNEMKTFGQGGVDVAEDGPGKGRQREDVQ             340
V.parvula  ALGSTMGLSIIDASMHMLNDMKTFVEAEVSVAEDGAGKGRQKNKE              372
```

Table S1. Peptide fingerprint of purified ArsAB

Purified ArsAB was digested with trypsin, and the peptide mixture was analyzed by MS/MS. Masses detected were matched to the predicted peptides generated by trypsin digest.

ArsA	
Observed mass	**Peptide**
1757.7549	IAYGTQDFTEGPAMTR
1870.8734	EQAIQAVETGIDIVNDR
1871.9658	VGGFELGALAGVILGSAANR
1885.8506	KIAYGTQDFTEGPAMTR
2115.9612	EYMFASHLSGEPAHSIALR
2989.3035	CFCLGEMGIGNTTSSATIVGAFTGLAPEK + 2 Carbamidomethyl (C)
3402.4731	CFCLGEMGIGNTTSSATIVGAFTGLAPEKVTGR + 2 Carbamidomethyl (C)
3738.6433	LEDMVEQYAGITGELNPALPKPCMVVASADHGVAR + Carbamidomethyl (C)

ArsB	
Observed mass	**Peptide**
732.5325	QSKDVR
1311.6975	QQAIQAIEVGVR
1734.9993	VGGLAIAGLVGVILGAAAGR
1787.8163	TFGEAEVAVAQDGPGALR
2134.0281	LIMVDIGVAADLPHSPAVCR + Carbamidomethyl (C)
2299.0833	LTGFCQGQAPIQVFAAHVQAR + Carbamidomethyl (C)
2479.1797	RVDNLTKPLNSLHSFEHIACK + Carbamidomethyl (C)
2575.1985	MLEELIAAIKPLDSIAMEQCQR + Carbamidomethyl (C); Oxidation (M)
2608.1443	SIIIMAADNGVAMATDQQQMTTAAR
2648.3157	EAELVNTAIAVNRPNAADPLDILTK
2789.2412	MNLGEGTGAALGMSVINATLHMLNDMK
3562.5693	GCQVIGLGEMGLGGLAAAMAIVACCHGQPLPGLAGR+ 3 Carbamidomethyl (C)

Table S2. Phosphoribosyltransferase activity of *So*ArsAB in varying buffer and pH.

Substrate	**Buffer**	**Specific activity**[a]
Phenol	pH 7 MOPS[b]	16 ± 1
Phenol	pH 8 HEPES[c]	23 ± 3
Phenol	pH 9 Glycine	26 ± 3
Phenol	pH 10 Glycine	18 ± 1
5,6-Dimethylbenzimidazole	pH 9 Glycine	19 ± 3

[a] nmol α-ribotide / min per mg of protein; average of triplicate activity measurements; the experiment was performed twice.

[b] MOPS, 3-(*N*-morpholino)propanesulfonic acid

[c] HEPES, 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid

References

Möller, B., R. Ossmer, B. H. Howard, G. Gottschalk & H. Hippe, (1984) *Sporomusa*, a new genus of Gram-negative anaerobic-bacteria including *Sporomusa-Sphaeroides* spec-nov and *Sporomusa-Ovata* spec-nov. *Arch. Microbiol.* **139**: 388-396.

Pospiech, A. & B. Neumann, (1995) A versatile quick-prep of genomic DNA from gram-positive bacteria. *Trends Genet.* **11**: 217-218.