

# **Identification of novel tail-anchored membrane proteins integrated by the bacterial twin-arginine translocase**

José Jesús Gallego-Parrilla<sup>1\*</sup>, Emmanuele Severi<sup>1\*</sup>, Govind Chandra<sup>2</sup> and Tracy Palmer<sup>1</sup>

<sup>1</sup> Microbes in Health and Disease Theme, Newcastle University Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

<sup>2</sup> John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK.

\* These authors contributed equally to this study.

## **SUPPLEMENTARY INFORMATION**

**Fig S1.** A. Multiple sequence alignment of seven S1 peptidase sequences from *Streptomyces*. The red box indicates the likely signal peptide region (predicted using SignalP-6.0) and the yellow the C-tail region (identified from DeepTMHMM). B. AlphaFold structure prediction for Uniprot entry D9XEL1, an S1-peptidase from *Streptomyces viridochromogenes*, coloured according to confidence (from dark blue which is very high through to red which is very low).

**Fig S2.** Multiple sequence alignment of six streptomycete D-alanyl-D-alanine carboxypeptidases from dataset S1 alongside *S. coelicolor* DacC (CAB9066.1). The twin arginines of the signal peptide are indicated in red. The C-tail regions of each protein (identified from DeepTMHMM) are shown in yellow.

**Fig S3.** A. Multiple sequence alignment of seven LPXTG/LAETG-motif containing peptidase sequences from *Streptomyces*. The red box indicates the likely signal peptide region (predicted using SignalP-6.0) and the yellow the C-tail region (identified from DeepTMHMM). The asterisks indicate the position of the sortase recognition motif.

**Fig S4.** A. Multiple sequence alignment of SCO1860 family LAETG-anchored proteins identified in dataset S2. The twin arginines of the signal peptide are indicated in red. peptidase sequences from *Streptomyces*. The yellow box indicates the C-tail region (identified from DeepTMHMM). The asterisks indicate the position of the sortase recognition motif.

**Fig S5.** No detectable export of AmiA by signal peptides of candidates 3, 5, 7, 8, 9, 14 and 16. Plasmids encoding the indicated signal peptide (SP) - AmiA fusions were introduced into *E. coli* strains MC4100  $\Delta amiA \Delta amiC$  (circles) and MC4100  $\Delta amiA \Delta amiC \Delta tatABC$  (triangles and labelled tat-) and grown for 20 h in LB in the presence of 0.5% SDS without shaking in a plate reader. Growth curves correspond to an averaged triplicate set  $\pm$  one SD (standard deviation). For clarity, growth curves are displayed either in groups of 2 or individually

alongside the positive ("SP\_TorA") and negative ("Vector", i.e., pUniAmiA) controls in each plot.

WP\_043259400.1 1 MRHARRRVRRVTRLTAAGVGLLGGAMVTOAAM SDTPARDAATLSAPQTPADKGSALVAKLSARTAGSWVDSGGRTVV  
 WP\_006380203.1 1 MRHARRRIVRRATRLTAAGCVILGGIMVTOAVA -EPPE-VNLPFSSAQAAAGTTGSELVTELGTDRTAGSWIAADGKFPVV  
 WP\_043435093.1 1 MRHARRRVRRRLTRLTAAGVGLLGGCTMVARAAV SEPPPASAR---SLSSAGNTCAAWVARLGRTARTAGTWIGSDGREVV  
 WP\_016325237.1 1 MRHARRRIVRRVARLTAAGVGLLGGCTMVTRAVA -EPPDASAAPRTFQAQTPSGAGDLVSRLLGTGRTAGHWVDSEGREVV  
 WP\_079047754.1 1 MRHARRRVRRVTRLTAAGVGLLGGAMVTRAVA -EPPADFPVYAAAEEDTGPGSALVARLGASRTAGSWLGEDGRTVV  
 WP\_086565138.1 1 MRHARRRVRRVTRLTAAGVGLLGGAMVTRAVA -ESPDPISIGVPYSARPATSMGTDLVSRLLGTSRTAGSWVDSAGREVV  
 WP\_086855524.1 1 MRHARRRVRRVTRLTAAGVGLLGGAMVTRAVA -EPPDPTIGVPYSAQPATSMGTDLVSRLLGTSRTAGSWVDSAGREVV

WP\_043259400.1 81 AVTDKATAAEVTKAGAEAKVVPYSNQLKSAATTTSSAPKIAGTAWAMDYRNKVLVEGDSSVSATSWAAMTKVAKSTGA  
 WP\_006380203.1 79 AVTDAEAATEVEKAGARAKIVDHSMNDLKSATKTTLSAPRVAGTAWAVDYTKNEVVQADSTVSAADWAKLSAVAKSIGG  
 WP\_043435093.1 78 AVTDDAAARAVRRTCAGAKIVRNSMNDLTSATATLRAAPRVAGTAWAVDYRTNTVKVQADRTVSAANWSRISQVAKKIGG  
 WP\_016325237.1 80 AVTDERAAEAARRAGATAKVRHSMSELRSAASTLSAPRVAGTAWALDYRSNEVVVRGDRTVSGSDWSRLTAVAEIGIG  
 WP\_079047754.1 80 AVTDEKAAAEVREACADPKVVRHSMRDILKSATKTLGSAAPRVPGTAWAMDYRSNEVVVRGDRTVSADWSRLTAVAEIGIG  
 WP\_086565138.1 80 AVTDEKAAAEVERAGARAKVVRHSMSDLKSATATLSAPRVPGTAWALDYKSNEVVVRGDRTVSADWSRLTAVAEIGIG  
 WP\_086855524.1 80 AVTDEKAAAEVERAGAOAKVVRHSMSDLKSATATLSAPRVPGTAWALDYKSNEVVVRGDRTVSADWSRLTAVAEIGIG

WP\_043259400.1 161 SVEMORTQCKETTRINCAQPIILSTGRCRSAGENVNTGQTDFFILTAGHCGPVGSIWFADNQDQQLGKTVSDNFPGSDFSL  
 WP\_006380203.1 159 SVRMERSEGTFTTRINAAQPIILSTGRCRSAGENVNTGSADEFFILTAGHCGPVGSIWFADTGSTETIKTISSEFPINDMSL  
 WP\_043435093.1 158 FVHMERTKCAFTTRINCAQPIILSTGRCRSAGENVNTGSTDFILTAGHCGPVGSTWFAADNQDQQLGKTVSNVFPFGHDFSL  
 WP\_016325237.1 160 FVRMERTEGAFTTRINCAEPILSTGRCRSAGENVNTDCTSDFFILTAGHCGPVGSIWFADNRPEQGVQRTVAGSFPFGDDFSL  
 WP\_079047754.1 160 AVTDEKAAAEVREACADPKVVRHSMRDILKSATKTLGSAAPRVPGTAWAMDYRSNEVVVRGDRTVSADWSRLTAVAEIGIG  
 WP\_086565138.1 160 FVRMERTEGTFTTRINCAQPIILSTGRCRSAGENVNTDCTTDYILTAGHCGPVGSIWFADNRKQVGSLLKGSFPFGDDMSL  
 WP\_086855524.1 160 FVRMERTEGTFTTRINCAEPILSTGRCRSAGENVNTDCKTDFILTAGHCGPVGSIWFADNRQVCTSLKGSFPFGDDMSL

WP\_043259400.1 241 VOYASGSAGGADVVAIGNCGSVTITGLADAAGVEKVRFSGSTSGLHDEGVTALDATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_006380203.1 239 VOYNGGSAGGADVVAIGNCGSVRITGAADPTVGQKVRFSGSTTGLRDKVITGLNATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_043435093.1 238 VOYSSGOAGGADVVAIGGCKGVRTITGAADPAVGQVRFSGSTSGLHDEGVTALNATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_016325237.1 240 VEYANGKAGGADVVAIGGCKGVRTITGAGEPAVGQVRFSGSTSGLHDEGVTALDATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_079047754.1 240 IEYDCGRAGEGAGVSVGNCGQIRITGVGDPVVGQVRFSGSTSGLHDEGVTALNATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_086565138.1 240 VKYDSGRAGEGADVVAIGGCKGVRTITGVADAEGVQKIFRSGSTSGLHDEGVTALNATVNYPEGTIVGLIETIVCAEPGDS  
 WP\_086855524.1 240 VKYDSGRAGEGADVVAIGGCKGVRTITGVADAEGVQKIFRSGSTSGLHDEGVTALNATVNYPEGTIVGLIETIVCAEPGDS

WP\_043259400.1 321 GGPIFSEGIAGLGVTSGGSGDCT-AGGTTFFQELTKALAAALNVKLLVAPKTA-----ASAPASASAAGSAAAGTGSVTPGT  
 WP\_006380203.1 319 GGPIFSEGIAGLGVTSGGSGDCK-TGGTTFFQEPVTEALTALNVRLIGQAGQ-----AAPAPSG-AATQSAVAPGEAOPGS  
 WP\_043435093.1 318 GGPIFSEGIAGLGVTSGGSGDCT-TGGTTFFQEPVTKAMTALCVRLIVSAQNAAGGTQSGAPSPGPSTPPQRAIAPSAAAPGS  
 WP\_016325237.1 320 GGPIFSEGIAGLGVTSGGSGDCA-KGGTTFFQELPEAMASLCVRLIVPGREGAAAGTSAPAGQA-P---GTRAPGASGAGA  
 WP\_079047754.1 320 GGPIFSEGIAGLGVTSGGSGDCAQAGGTTFFQEPVTKALAAALNVKLLVAPKAGGGADTAAPASAST---GAATFDAASE--  
 WP\_086565138.1 320 GGPIFSEGIAGLGVTSGGSGDCT-AGGTTFFQEPVTKALTELGVKLLIVAKQAG-G-TGASAPAPASQ---GAASPGSSAP--  
 WP\_086855524.1 320 GGPIFSEGIAGLGVTSGGSGDCA-AGGTTFFQEPVTKALTELGVKLLIVAKQAG-GATGASAPASQ---GAASPGSSAP--

WP\_043259400.1 395 SAPATNTTAGTTTTASRLTPKIVGPGLLVIAAGSLIALVATRWIRA QDRKAFQRHYSAMWG  
 WP\_006380203.1 392 VAPVTGGSAG-EALLDKLADPKIVGPGLLIAGGSLIALVATRWFRV QDRRDYRRYSATWS  
 WP\_043435093.1 397 SAPVTGTAQG-GALLARLTDPKIVGPGLLVIAAGSLIALVATRYIRA QERRAYQQHYSATWG  
 WP\_016325237.1 395 PAPVAGV-DG-STLLARLTDPKIVGPGLLVVAGSLVALVATRWIRE QDRKAYRRRYSATWG  
 WP\_079047754.1 395 GSSAAGI-NA-STLLSRLADPRIVGPGLLVIAAGSLIALVATRWIRA QDRKAYQRYYSATWS  
 WP\_086565138.1 392 ---VTGE-DV-SGLLTRLTPKIVGPGLLVIAAGSLIALVATRWIRA QDRKAYRQHYSATWG  
 WP\_086855524.1 393 ---VTGE-DV-SGLLTRLTPKIVGPGLLVIAAGSLIALVATRWIRA QDRKAYRQYYSATWG

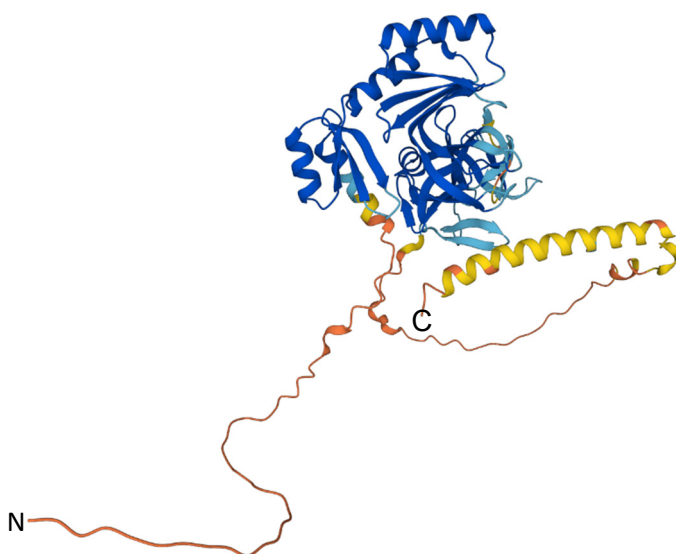


Fig S1



```

WP_086778847.1 1 MRITPSRRRTLGRSTAAVAATALVALG-AAPATADEPEPDLGVGTIAPIKGVQAGSAESTPLTELNKGTEEVVSYVTYA
WP_076967992.1 1 MGIAASCRRTLLSATAVSATAALIALG-AAPACAIAIKPDLGVRLAPVTGIAAGSGFGLPVSLLNKGTEELPKVWVSYT
WP_030568954.1 1 MGIAASCRRTLLSATAVSATAALIALG-AAPACAIAIKPDLGVRLAPVTGIAAGSGFGLPVSLLNKGTEELPKVWVSYT
WP_084749454.1 1 MGIIVSCRRTPLSGGAVLASAALIALATAPACAAGTGPDLGVGALTPTVTGIAFGGTGLPVSVLNKGTEEVKVVLSYS
WP_032758621.1 1 MGNVASCRRTLLSGTAVLAVALIALG-SAPACADETEPDLGVRLPTVTGIAFGSGFGLPVSLLNKGTEELPKAYVTYS
WP_031122887.1 1 MGNVASCRRTLLSGTAVLAVALIALG-SAPACADETEPDLGVRLPTVTGITFGSGFGLPVSLLNKGTEELPKAYVTYS
WP_055558699.1 1 MGNVASCRRTLLSGTAVLAVALIALG-SAPACADETEPDLGVRLPTVTGITFGSGFGLPVSLLNKGTEELPKAYVTYS

WP_086778847.1 80 VSPGLQADETYKNCSTYITPSHDEIPDEQVAACKIDQPLKPGVVYGTETKPLSLKALDLSALNDDLRTVTGVTEPDPSGDT
WP_076967992.1 80 YSSGIAPAETYSNCVYITLSPGGEWPEEHRACTVDQPLKPGVLYGTETQPIALATARANALYDEVGLGLLAEEPVEHP-DI
WP_030568954.1 80 YSSGIAPAETYSNCVYITLSPGGEWPEEHRACTVDQPLKPGVLYGTETQPIALATARANALYDEVGLGLLAEEPVEHP-DI
WP_084749454.1 81 FSVGLETAETYSNCVYFRINSYDERBERNEALCTLDQPLKPGVLYGTETKPLGLKAGADALYDRAGYSVSADKAPDD--G
WP_032758621.1 80 FSSGLATAETYSNCLYSTIGSYDEMPERHRAVCTVDQPLKPGVLYRTETKPIGLKAGANALYDTAGLGLWASPPCPDD--G
WP_031122887.1 80 FSSGLATAETYSNCLYSTIGSYDEMPERHRAVCTVDQPLKPGVLYGTETKPIGLKAGANALYDTAGLGLWASPPCPDD--G
WP_055558699.1 80 FSSGLATAETYSNCLYSTIGSYDEMPERHRAVCTVDQPLKPGVLYGTETKPIGLKAGANALYDTAGLGLWASPPCPDD--G

WP_086778847.1 160 GADEPVPGTGDPITLVEKGPATDADRANHPEESAASADVTAANTADFGLTGAELKGEPEKVTATVKFTNAGPARVQOARD
WP_076967992.1 159 HHGEPEPGTGCPPLKLVERGEATAADRAAHAEPSRTEVTAVENTADYALTGAELQKVGDKVTALVTFTEGPARVYRDQG
WP_030568954.1 159 HHGEPEPGTGCPPLKLVERGEATAADRAAHAEPSRTEVTAVENTADYALTGAELQKVGDKVTALVTFTEGPARVYRDQG
WP_084749454.1 159 GESEHVPGTGAPLKLVERGEATEADRAAHDPDSVVGADVTAVENTADFALTGAELKGVGDKVTAQVKFTNKGPAVVMGERD
WP_032758621.1 158 GETDPVPGTGAPLTLVERGEATEADRAAHAEPSGWTVDVTAVENTADYALTGAELKGVGDKVTAQVKFTNKGPAVVIYRELG
WP_031122887.1 158 GETDPVPGTGAPLTLVERGEATEADRAAHAEPSGARTDVTAVENTADYALTGAELKGVGDKVTAQVKFTNKGPAVVIYRELG
WP_055558699.1 158 GETDPVPGTGAPLTLVERGEATEADRAAHAEPSGARTDVTAVENTADYALTGAELKGVGDKVTAQVKFTNKGPAVVIYRELG

WP_086778847.1 240 QSVATVDVHIKGTSTVKAHGYCDAVTKTHYRCGTSSQWVDVNGGESYDFVLRIDKAVGRTTGKVSFGCPERPFDNRNAN
WP_076967992.1 239 TSAANVDVHIKGTSTVKAHGYCRPVITQTHYRCGTSSQWVEAKGGETYSFVLRIDKAVGRTTGKVSFGCAARPFDKNAAN
WP_030568954.1 239 TSAANVDVHIKGTSTVKAHGYCRPVITQTHYRCGTSSQWVEAKGGETYSFVLRIDKAVGRTTGKVSFGCAARPFDKNAAN
WP_084749454.1 239 TSVTTVDVRIKGTSTVKAHGYCDAVTKTHYRCGTSSQWVDVPTDCATYFVLRIDKAVGRTTGKVSFGCAARPFDKNAAN
WP_032758621.1 238 TGAASVDVRIKGTSTVKAHGYCDAVTKTHYRCGTSSQWVDEEGGETYSFVLRIDKAVGRTTGKVSFGCQSRPFDRNPAN
WP_031122887.1 238 TGAASVDVRIKGTSTVKAHGYCDAVTKTHYRCGTSSQWVDEEGGETYSFVLRIDKAVGRTTGKVSFGCQSRPFDRNPAN
WP_055558699.1 238 TGAASVDVRIKGTSTVKAHGYCDAVTKTHYRCGTSSQWVDEEGGETYSFVLRIDKAVGRTTGKVSFGCQSRPFDRNPAN

WP_086778847.1 320 DTAQIVIDTGGKSDTSGSSGSS--GS-----TG-GTGGSSSTGGSD-----TTGGTSGTGGSSGSSGAGTGCT
WP_076967992.1 319 DTAQIVVDTPPEPFDGTSTGGTSSSTG-G--SAGPTGSTDGCTGSGGSTSGSSGSASSGSGTGGSDTGTGGSATSGSCS
WP_030568954.1 319 DTAQIVVDTPPEPFDGTSTGGTSSSTG-G--SAGPTGSTDGCTGSGGSTSGSSGSASSGSGTGGSDTGTGGSATSGSCS
WP_084749454.1 319 DTAQIVVETGEPATTGGTASTGGSGTSGSSGSSGSGTGGSDGCTGSGGSTSGSSGSGTATG-----GTTGSDSTTSGSCS
WP_032758621.1 318 DTAQIVVETGKAGSTGGTSSGSS-----ANGGSDGCTGSGGSTSGSSGSGTGGGSTD---TTGGSGSATSGGCS
WP_031122887.1 318 DTAQIVVETGKAGSTGGTSSGSS-----ANGGSDGCTGSGGSTSGSSGSGTGGGSTD---TTGGSGSATSGGCS
WP_055558699.1 318 DTAQIVVETGKAGSTGGTSSGSS-----ANGGSDGCTGSGGSTSGSSGSGTGGGSTD---TTGGSGSATSGGCS

WP_086778847.1 381 TTGGATPQTGCNLAATGSDSTLELAGAAGAAVAVAGGGIVWAVRRSAARGN
WP_076967992.1 396 TSGGTTPTQTCGGLAATGSDSTLELAGAAGAAVAVAGGGIVWAVRRSAARGN
WP_030568954.1 396 TSGGTTPTQTCGGLAATGSDSTLELAGAAGAAVAVAGGGIVWAVRRSAARGN
WP_084749454.1 393 TTGGTTPTQTCGGLAATGSDSTLELVAGAAGAAVAVAGAGLFYAMRRTAARKG
WP_032758621.1 384 TTGGATPQTGCNLAATGSDSTLELVAGAAGAAVAVAGGGIVWAVRRTAARNG
WP_031122887.1 384 TACGATPQTGCNLAATGSDSTLELVAGAAGAAVAVAGGGIVWAVRRMAARNG
WP_055558699.1 384 TACGATPQTGCNLAATGSDSTLELVAGAAGAAVAVAGGGIVWAVRRTAARNG

```

\*\*\*\*\*

```

WP_070013327.1 1 -----MITHV-----LSSRARRSV--LLSAATVMV--GGAMVSV--AHASADGHSSTENATAVA
WP_070014799.1 1 -----MITHV-----SGSRARRSA--LLGAAAVLA--GGTFLATP--AHAPAHGHSGSSKAGATA
WP_004943253.1 1 -----MSSTT--RSSFGMPVSVRRAV--VPSAAALAAALTALSAS--PASAPAT--GDHRTCAADAVV
WP_055573436.1 1 -----MNSTNSRV-----LARRRT--VVAAAAVLAAGSAALAGAGTAQATGE-----PCSASAAV
WP_055694013.1 1 -----MNSTNSRV-----PARRRT--VVAAAAVLAAGSAALAGAGTAQATGE-----PCSASAAV
WP_055606274.1 1 -----MNSTDSRA-----SARRRA--VVATATVLAAGPVALAGAGAAQATGE-----HCSASAAV
WP_055688174.1 1 -----MNSTDSRA-----SARRRA--VVATVTVLAAGPVALAGAGAAQATGE-----HCSASAAV
WP_065002969.1 1 -----MNSTDSRA-----SARRRA--VVATVTVLAAGPVALAGAGAAQATGE-----HCSASAAV
WP_078852820.1 1 MSPIGRRRWLHVSCGGSTILNSNTFRLAALA-----VAAAPVALLVAAPAHATGAT--TTTGGCKASAVV
WP_030791129.1 1 -----MNSFTFRMPACRRHRR--LVVATATALAAGPAALAGAGSARATGG--DHDRCGRADAVV
WP_064073155.1 1 -----MNSTT-----SRPSARRRTI--AAAVAAFAAGPALLTAPAA--HAT--G-----GECRASAVV
WP_073752756.1 1 -----MNSNTVRLPARLPAARRRALSTAAAVAALAGPALLAAPMA--HATGG-----GECRASAVV

WP_070013327.1 52 ARADLVQSVAD-AAKVEVKASLNSVSAEDAEETLLSAEVKCANKGQPLKLVQADVARSGAKADARRSTGHVKLAGLRAF
WP_070014799.1 52 ARADLVQSVAD-AAKVEVRASLNEVSABRSAORTLLTAEVNCANKGRELKLVQAEVARSAAKTGADRSSGHVKLAGLRAF
WP_004943253.1 56 IRAGLDVSLLDKTVRVEVNASLNDVHAFGNAARTALSVRVGCAEQGRPEVLLRADAATARATADEGTATGRAELVRAKVH
WP_055573436.1 49 IRTGLDVALLDETVNVELTVSLNEVHAFRSAHRTALTADLEGVDGGRFFSVLRAEVADARATVEGNRAEASTTLAEARIH
WP_055694013.1 49 IRTGLDVALLDGTVNVELTVSLNEVHAFRSAHRTALTADLEGVDGGRFFSVLRAEVADARATVEGNRAEASTTLAEARIH
WP_055606274.1 49 IRTGLDVALLDRTVNVELTVSLNEVHAFRSAHRTALTADLEGVDRGRFFPVLRAEADARATVEGNRAEASTTLAKARIH
WP_055688174.1 49 IRTGLDVALLDRTVNVELTVSLNEVHAFRSARRTALTADLEGVDRGRFFPVLRAEADARATVEGNRAEASTTLAKARIH
WP_065002969.1 49 IRTGLDVALLDRTVNVELTVSLNEVHAFRSARRTALTADLEGVDRGRFFPVLRAEADARATVEGNRAEASTTLAKARIH
WP_078852820.1 64 IRTALDVSLLSKTVDVELKTTLNEVHAFASAEQTALVRLDVNHCGRFFSVLRADVATANATVDKHRAEGSYNLAKARVH
WP_030791129.1 58 IRTGLDVALLNKTVNVELKVSLNEVQABASAEKTALTADLEGVDRGRFFPVLRADVATANATVKGGRAEGSYNLARAEMH
WP_064073155.1 48 IRTDLDVALLNKAVHVELTATLNEVQABASARRTALTVTLDGVEQGRFSVLMADVATAQATDGRRGEGRTELARAKVH
WP_073752756.1 56 IRTDLDVSLLDKTVHVELKATLNEVQABASARRTALTVTLDGVEQGRFSVLADVATAKATADARRAEGHTNLAHATLH

WP_070013327.1 131 APGLLGKPLVSADLLTATATCKAGEKMAKADI-ENSVLGRPVTVNGVGSQQVKVPVGVTVDVALEQETTKAASGAATA
WP_070014799.1 131 APGLLGKPLVSADLLTATATCEAGAKETAQADI-ENSVLGRPVNVNAPCGQQIKLPVGVTVDVSLEQETTKTASGAATA
WP_004943253.1 136 VPGLPLLSLIEVRQVTSEAVCRAGQRTATSNLLGPTVVLGKITLTSGGTTDVDVPGVGRVRLDLSKTATTSHTGAATA
WP_055573436.1 129 LPGLPLLSLVELEAAGSRAVCETGRTPLASATLPAAVTVLGKRVLTASGPTEVEVPGVGEVRLDLSKTETTSRTAAATA
WP_055694013.1 129 LPGLPLLSLVELEAAGSRAVCETGRTPLASATLPAAVTVLGKRVLTASGPTEVEVPGVGEVRLDLSKTETTSRTAAATA
WP_055606274.1 129 LPGLPLLSLVELEAVSSKAVCEMGRTPLASATLPAAVTVLGKRVLTASGPTEVEVPGVGEVRLDLSKTETTSRTAAATA
WP_055688174.1 129 LPGLPLLSLVELEAVSSKAVCEMGRTPLASATLPAAVTVLGKRVLTASGPTEVEVPGVGEVRLDLSKTETTSRTAAATA
WP_065002969.1 129 LPGLPLLSLVELEAVSSKAVCEMGRTPLASATLPAAVTVLGKRVLTASGPTEVEVPGVGEVRLDLSKTETTSRTAAATA
WP_078852820.1 144 LPGLPLLSLIEVEKVTSAKCEVGRRRPAESHNVLGHSVLGKRVTLTSGGTTRVDVPGVGEVRLDLSKTETTSRTAAATA
WP_030791129.1 138 VPGLPLLSLIEVEKVTSAKCEAGKRPVAEANLLGGTVLGKKVTLTAGGTTEVKVPGVGEVTLALSKRRTTSRTAVAAA
WP_064073155.1 128 VPGLPLLSLIELEKVTARALCEAGGQPVAEANVQGRVTVLGKRVTLTAGGPTKVEVPGVGEVTLTLSDTTRTTTSRTAAATA
WP_073752756.1 136 VPGLPLLSLIEVEKVTSAKCEAGQPVAEANLLGRVTVLGKRVTLTAGGPTKVDVPGVGEVTLLDSTTRTTTSRTAAATA

WP_070013327.1 210 LKLSYEVNPAKLNIVKAAGEIVLSEATCETPAGGGEETGGTGGNEPGPQTGE-----
WP_070014799.1 210 LKLSYEVNPAKLNIVKATGDIVLSEATCETPKGGSD-----GNEPGPQTGGGD-----
WP_004943253.1 216 LELAVDVNPLKLGVAEVHGRVTLVRACETPAGAGKPAEQPPAPARQAPA---K-----EATRTSGDGIKP
WP_055573436.1 209 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_055694013.1 209 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_055606274.1 209 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_055688174.1 209 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_065002969.1 209 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_078852820.1 224 LRLEVSVDPLKLNVAEVDGTVTLAEATCEAPGGDAAPA---PGGQREPERREEPQEQGRERERADAVPPA--GSGTGVEP
WP_030791129.1 218 LELKVSVDNPLKLNVAEVEGTVTLAGACCEAPRAERERTGGG---TKPQG-----KPAKPV---
WP_064073155.1 208 LALTVSVDNPLKANVADVTGTTTLAEATCESPKAPAAAAAPA---APTAP-----AAKPADVRP
WP_073752756.1 216 LALTVSVDNPLKANVADVTGTTTLAEATCESPTAP-----PKP-----AEARSAGVQF

WP_070013327.1 263 -----DGGDGLAETGGSSATPFIAGAGALLVAGCGAMYLIRRGARSTS--
WP_070014799.1 258 -GGDAGSNNGDGGGLAETGGDATPFIATAGALLVAGGTTMYLIVRRRAARNS---
WP_004943253.1 279 QTV-----ADKRDLAETGGSSATPFIAGAGALLVAGCGGAVIARRRAAARGRG
WP_055573436.1 284 RGARAGRAA---PGLAETGGDEAPYVAGGALALLAAGGAAVLLRRGRRS---
WP_055694013.1 284 RGARAGRAA---PGLAETGGDEAPYVAGGALALLAAGGAAVLLRRGRRS---
WP_055606274.1 289 QGARAERAAPARTGLAETGGDSTPPYVAGGALALLAAGGAAVLLRRGRRT---
WP_055688174.1 288 QGARAERAAPARTGLAETGGDSTPPYVAGGALALLAAGGAAVLLRRGRRT---
WP_065002969.1 289 QGTRAERAAPTRTGLAETGGDSTPPYVAGGALALLAAGGAAVLLRRGRRT---
WP_078852820.1 285 QTGTGHAPAAVDTSLAETGGSSTPPYIAGGALLLALCAGAMVVRRGAQD---
WP_030791129.1 270 --G-----HKATEANLAETGGNSTPPYIAGGAVVLLIAGGGAVAMRRSRSRG---
WP_064073155.1 264 QTA-----GEPPEANLAETGGSSLPPYLAGGSLALLACGAGALALRRARAARARG
WP_073752756.1 263 QSA-----GEPQANLAETGGSSLPPYLAGGSLALLTACAATFALRRARTTRERG

```

\*\*\*\*\*

Fig S4



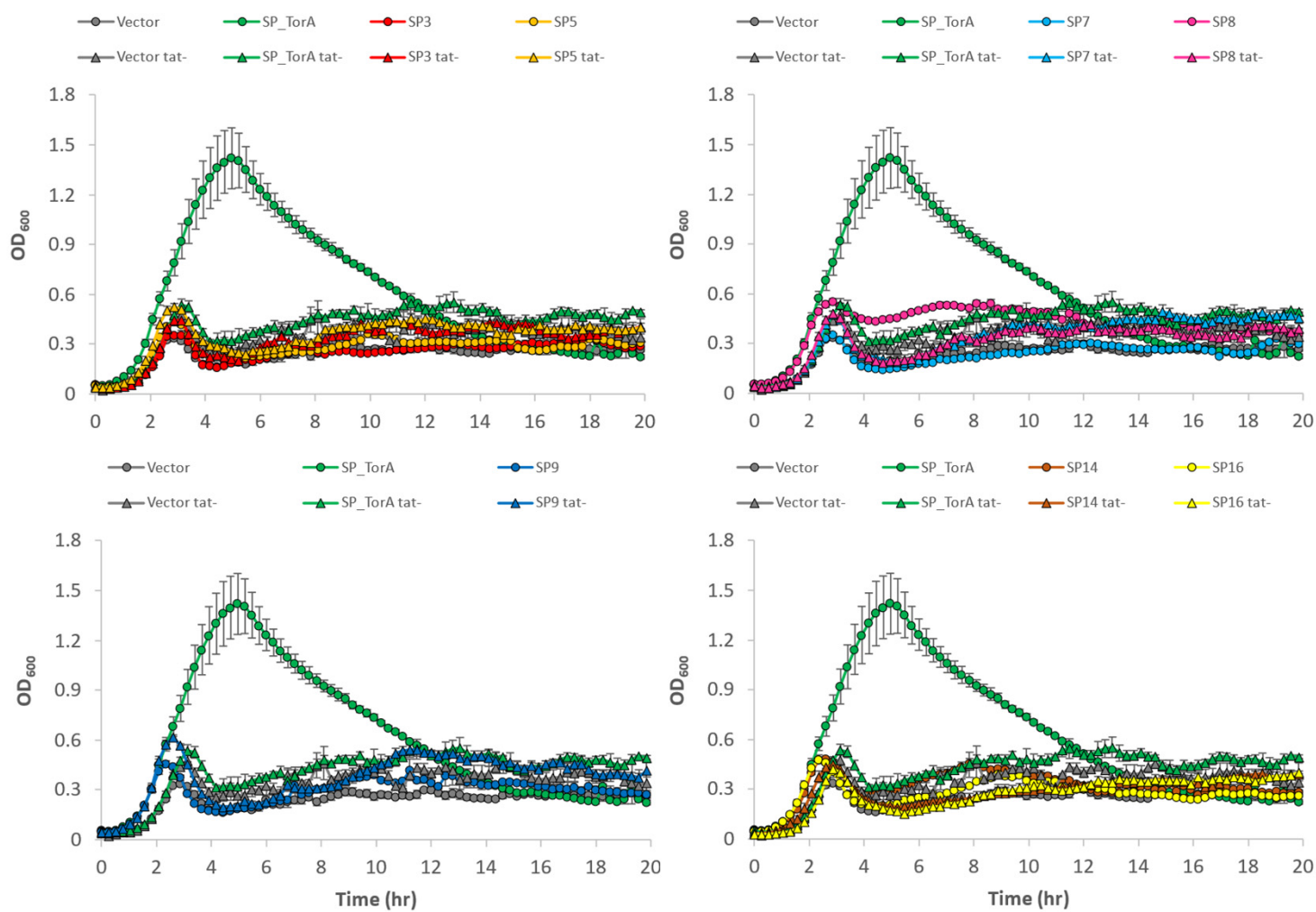


Fig S5



NAME	SEQUENCE 5'→3' <sup>a</sup>	USE <sup>b</sup>
JES3	ACAGAGGAGGATCCATGTCACTCAGTCGGCGTCAG	RL
JES4	TTGCGCTCGAGCTCCGGTACCGGATTGACCAACAG	RL
C1F	TATAGAGCTCCTGACCGACCCGCGTA	RL
C1R	TATAGCATGCTTAACCCCAGGTAGCA	RL
C2F	TATAGAGCTCACCGCTGCTCCGGACA	RL
C2R	TATAGCATGCTTATTAAGCCTGACCACCACC	RL
C6F	TATAGAGCTCTCTACCCTGCCGCTGG	RL
C6R	TATAGCATGCTTAAGAAGCACGAGCA	RL
C8F	TATAGAGCTCGCTCTGGGTCCGGCTG	RL
C8R	TATAGCATGCTTACGGACGACGACGTT	RL
SP1F	TATAGGATCCATGCGTCACGCTCG	RL
SP1R	TATATCTAGAAGCAACAGCGTTGGTA	RL
SP2F	TATAGGATCCATGACCACCTCTTCTC	RL
SP2R	TATATCTAGATAGCAGCAGCCTGGTCC	RL
SP4F	TATAGGATCCATGGGTAACGCTGTTT	RL
SP4R	TATATCTAGAAGCCTGAGCCGGAGCAGAAC	RL
SP8F	TATAGGATCCATGTCTCTGCGTCGTC	RL
SP8R	TATATCTAGAAGCCAGAGCAGCCGGAGCAGCCA	RL
SP9F	TATAGGATCCATGTCTGGTATGGC	RL
SP9R	TATATCTAGAAGCAACTTCCGGAGCC	RL
ESN102	TGTAAAACGACGGCCAGT	RL
ESN103	CAGGAAACAGCTATGAC	RL
ESN237	AGCTGGCACACCACCTTCCAC	RL
ESN434	GTGGAAGGTGGTGTGCCAGCT	RL
ESN526	CGTATGTTGTGTGGAATTGTGAGC	SEQ
ESN527	GGGTAACGCCAGGGTTTTCCC	SEQ
ESN538	GTGACCTGGATCGAGAACACG	SEQ
ESN539	CGGACAAGTGGAGCTGCTTG	SEQ
ESN540	CAGTGAGTGGATCCATGACCTCTCGTCGTGGTACCTTC	RL
ESN541	CCTAGTTCCCTCTAGAAGCCAGAGCCGGCGGAGCCAG	RL
ESN542	CAGTGAGTGGATCCATGCCGGCGACCAACCGGTGATC	RL
ESN543	CCTAGTTCCCTCTAGACGCCGCCGCGGTGGTCGGCG	RL
ESN544	CAGTGAGTGGATCCATGGGTATCGCTGCTTCTGGTCGTC	RL
ESN545	CCTAGTTCCCTCTAGAAGCCAGAGCACGAACACCCAG	RL
ESN546	CAGTGAGTGGATCCATGAGCCGTCGTCTGACCGCGCTG	RL
ESN547	CCTAGTTCCCTCTAGACGCGCTCGCGCTGGTCGCC	RL
ESN550	CAGTGAGTGAGCTCGCTGGTACCCCGCTGGCTGTTCTG	RL
ESN551	CCTAGTTCCGCATGCTTACAGACCCAGGGTACGCTGAACC	RL
ESN552	CAGTGAGTGAGCTCCAGACCGGTGGTGACCTGGCTGC	RL
ESN553	CCTAGTTCCGCATGCTTAAACCGTTACGAGCAGCCATACGACGAC	RL
ESN554	CAGTGAGTGAGCTCGGTAGCGACGCGCCGGTTGCG	RL
ESN555	CCTAGTTCCGCATGCTTACGCCTGCGCATCACGCGCACC	RL
ESN556	CAGTGAGTGAGCTCGACGATGCGACCCCGGCGGCG	RL
ESN557	CCTAGTTCCGCATGCTTACGCACCCGCCGCACGACGAC	RL
ESN565	CAGTGAGTGAGCTCGTTGCTTCTCTGCCGTGGCTG	RL
ESN566	CCTAGTTCCGCATGCTTACGGCTGAGCTTCGGTCAGC	RL
ESN584	CAGTGAGTGAGCTCCCGGAGCTGTATCACGGTC	RL
ESN585	CCTAGTTCCGCATGCTTACTCATGATGATCCTCCTCGTC	RL
ESN591	GGTCAATCCGGTACCGTAAGCATGCAAGCTTGGCGTAATCATGG	QC
ESN592	CTTACGGTACCGGATTGACCAACAGTTGCCCAATCGAGCCAC	QC
ESN669	AACGCAGCTACTCTGTGGTATC	SEQ
ESN675	CAGAGCAGCCAGACCCAGAGC	KLD

ESN676	CTGGGTGGT <b>GCCT</b> CTGACGCTC	KLD
--------	---------------------------------	-----

**Table S1. Oligonucleotides used in this study.**

a = restriction sites used for restriction-ligation cloning of SP/CT-encoding PCR products are underlined (please note that, unlike the gBlocks for SP/CT 1-9, those for SP/CT 10-16 were synthesised to include these sites already and the target sequences were amplified from conserved extensions for the re-usable oligos ESN102, ESN103, ESN237, and ESN434). The start codons of SP-coding sequences and the stop codons of CT-coding sequences are in italics. The C24A mutant codon of SP9 is in bold.

b = molecular application of each primer. RL: restriction-ligation cloning; QC: QuickChange whole-plasmid PCR; KLD: whole-plasmid PCR circularised with NEB's KLD kit; SEQ: sequencing (ESN526 and ESN527: annealing to opposite ends of the multiple cloning site of the pSU40 backbone; ESN538: internal *amiA* primer; ESN539 and ESN669: internal *sufI* primers).

NAME	SEQUENCE
SP1	ATGCGTCACGCTCGTCGTGTTGTTTCGTTCGTGTTACCCGTCTGGCTGCTGTTGGTGGTCTGCTGCTGGGTGGTGCTATGG TTACCAACGCTGTTGCT
SP2	ATGACCACCTCTTCTCCGTCTGCTCCGCGTCGTTCGTATCCTGCGTTCTGCTACCGCTCTGGTTGGTGGTGTGCTCTGGCTGT TGCTGTTCCGCTGGCTGCTTCTGCTCACGTTTCGTGTTTCTCCGGACCAGGCTGCTGCT
SP3	ATGACCTCTCGTCGTGGTACCTTCCTGGCTGCTCTGGTTACCGCTTCTCTGATCCCGCTGGCTCCGCCGGCTCTGGCT
SP4	ATGGGTAACGCTGTTTCTGGTCGTTCGTACCTGCTGTCTGGTACCGCTGTTCTGGCTGCTGTTGCTCTGATCGCTCTGGGTTC TGCTCCGGCTCAGGCT
SP5	ATGCCGGCGACCAACCGGTGATCGTAGCCGTTCGTTCGTCCGCTGGCGTTTGCGGCGGCGGTTGCGACCGCGGCGGCGATTGG CGCGGCGAGCTTGGCGGCGGCGCCGACCAACCGCGGCGGCG
SP6	ATGGGTATCGCTGCTTCTGGTCGTTCGTACCTGCTGTCTGCTACCGCTGTTTCTGCTACCGCTGCTCTGATCGCTCTGGGTGC TGCTCCGGCTCAGGCTGACGCTATCAAACCGGACCTGGGTGTTTCGTGCTCTGGCT
SP7	ATGAGCCGTTCGTTCGTACCGCGCTGCTGACCGCGCCGGCTGCTGCGGCGGCGGCGGCGCTGACCCTGGTGGGCGCGGCGA CCAGCGCGAGCGCG
SP8	ATGTCTCTGCGTCGTTCGTGCTGCTCTGCTGGTTACCGCTGCTTCTACCCTGACCGCTCTGGCTGCTCCGGCTGCTCTGGCT
SP9	ATGTCTGGTATGGCTGGTCGTTCGTTCGTGGTTGCTGCTGCTCTGGGTCTGGCTGCTCTGCTGGGTGGTTGCTCTGACGCTC GTCCGCCGGCTCCGGAAGTTGCT
SP10	ATGACGCCTAACCGCCGTCTGTGGTACTTGCCGCAGCGGTAGCAACCGCTTTGGGGTTAGGGGCTACTGCTCTTACGTTAC CGGCTGTGGCG
SP11	ATGTTAACACGCCGCAACGCTTTAAGCGCGCTTGCCTTAGGCGTCGTTCGGAGTGTTCGATCGCTCCAACAGCCTTGGCG
SP12	ATGAAGAACCTTAAGCGCGTACCTTTGGCAGGGCGTCGCACAGCTGTCGCCGCGGCCGCCGACCCACTGCGGTGGCATT GCAGGCCCTGCTTACGCG
SP13	ATGAACAGTACAACCTCGCGTCCAAGTGCACGTGCCGCACTATTGCCGCTGCTGTAGCCGCATTGCTGCTGGTCCAGCTC TTCTTACTGCACCCGCCGCACATGCG
SP14	ATGAGTCCTATTGGTCGCCGCCGCTGGTTACACGTGTCTTGTGGGGGATCAACTATCTTAACTCTAATACCTTTGCGCTTGCT GCCCTTGCTGTAGCTGCTGCTCCTGTAGCCCTTCTGGTGGCCGCACCTGCCCATGCG
SP15	ATGCCTGCGCCGAAGAAAGCGCTTCGCCGTTCTTTGACTGTTACCTCGGCCGCGTTAACAGCTTTGGCATTGCTGTCTCCTGC AGCTCTTGCG
SP16	ATGCGCGACTCCTCTGGCCTTTGCGGCCGTGCAGCCTTGGGACTTGCAGCGGCCTTGCCCTTAGCTGCAGCAACGGCAACG ACAGCTGATGCG
CT1	CTGACCGACCCGCGTAACGTTGGTCCGGGTCTGCTGGTTATCGCTGGTTCTCTGGTTGCTCTGGTTGCTACCCGTTGGATCC GTGCTGAACAGGACCGTAAAGCTTACCGTCAGCACTACTCTGCTACCTGGGGTTAA

CT2	ACCGCTGCTCCGGACACCACCGTTACCACCGCTGCTTCTGACACCTCTGCTACCTCTTCTGCTGTTGCTGTTGGTCTGGGTGT TGGTGGTCTGGCTCTGGGTGCTGTTGCTCTGGTTGTTGCTGTTTTCGCTCTGACCCGTGTTCTGTCGTGAAGGTGGTGGTCAG GCTTAA
CT3	GCTGGTACCCCGCTGGCTGTTCTGCTGGGTCTGTTGCTGCTATCGCTGTTGCTGTTGGTGGTATCAAACCGCTGCACTCTTT CCTGCTGCAGGTTTACGCGTACCCTGGGTCTGTAA
CT4	CAGACCGGTGGTGACCTGGCTGCTACCGGTTCTGACTCTACCCTGCCGGTTGCTGGTGCTGCTGGTGCTGCTCTGCTGGCTG GTGGTGGTCTGTTCTACGCTATGCGTCGTCGTATGGCTGCTCGTAACGGTTAA
CT5	GGTAGCGACGCGCCGGTTGCGGCGCTGGGCACCGCGGCGGCGCTGGCGGTGGCGGCGGGTGCGGGCGTGTTTTTCGCGG TTCGTCGTCGTCTGGTGCGCGTGATGCGCAGGCGTAA
CT6	TCTACCCTGCCGCTGGCTGGTGCTGCTGGTGCTGCTCTGCTGGCTGGTGTTCTATCGTTTGGGCTGTTCTGTCGTCTGTTCTG CTGCTCGTGCTTCTTAA
CT7	GACGATGCGACCCCGGCGGCGCGTGCGCAAGTGCGTGCGAGCGCGCAAGAGCGTAGCGAAGTTCTGGCGGCGACCGGTG CGCGTACCGGCGTGCTGCTGGCGGCGGGTGCGCTGGCGCTGGGTCTGGGTGCGGGTCTGTTACCTGGCGTCGTCTGTCGT GCGGCGGGTGCGTAA
CT8	GCTCTGGGTCCGGCTGCTACCAAACGTGCTACCGGTGCTGAAGTGCCTTCTGACGACAAAAAGACGACGGTCTGTCTTCTT CTGCTACCACCTGGATCATCGTTGGCGTTGTATTCGTCTGCTTCTGCTGGTTTTGGTCTGCTGCTGTCTGGTCTGTAACGTCGT CGTCCGTAA
CT9	GTTGCTTCTCTGCCGTGGCTGCCGGTTGCTGCTATCGGTCTGCTGCTGGTTCGTGTTCTGTGGCGTCTGCGTCGTCTGTCGT GTGCTCGTATCGGTCTGCTCCGCTGACCGAAGCTCAGCCGTAA
CT10	GGCGCACTTGCGTCTACTGGTGCCGAAGTCAACGCTGGGGCCCTGGCAGGGGGCTGCCGCATTAGTGGTCGCCGCGGGCGC TGGCGTGGTGGTGGCGGCGTAA
CT10*	GGCGCACTTGCGTCTACTGGTGCCGAAGTCAACGCTGGGGCCCTGGCAGGGGGCTGCCGCATTAGTGGTCGCCGCGGGCGC TGGCGTGGTGGTGGCGGCGCGTCGCCGTAATGCTGCCGGGGATGCCGCTAGCTAA
CT11	GGCGCGGACGCACTGTCCTTATTAGGAATCGCAGGGGGCTGCTTCGGTAGCAGGTGCCGGGTTTATGTTATCACGTCGCAAAA ACGACTCGGAATAA
CT12	GGTAGTAGTGCGACTACCGTTGCAATGGGTGCAGCAGCGACAGGACTTGTCTTAGCAGGAGCCGGTGCTCTGTATGCTGTCC GCCGTCGTGCGAATAGCTAA
CT13	GGTAGCTCGTTGACTCCATACTTGGCTGGAGGATCGTTGGCGCTTTTAGCTGGGGGAGCCGGCGCACTTGCGCTTACGCGC CGTGCCCGTGCGGCTCGCGCTCGTGGCTAA
CT14	GGTAGCAGCACTACGCCCTACATTGCGGGCGGTGCCGCTTTGTTGTTAGCTTTAGGGGCTGGTGCAATGGTAGTAGCCCGCC GCCGTGGAGCCCAGGATTAA
CT15	GGCGCCGGCGTCGCTGCTGGCATCGCGGCTGGGGTTTTAGCATTGTTAGCAGGAGGAGCTTTTCTGATTAACCGTAAATGGC CGTTGCCAGATTTGGTCCGCCGTCGTACCCGCCCGTAA

CT16	GGCGCACCGTGGGGTTTAGTTGGAGGTGCAGGTGCTGCCGTCGCACTGCTTGCAGGCGGAGCACTGGCCTTGCGCCGTCGT CGTCGCGCAGCAGGTGAAGCAAGTGAACCAGTAGATGCTGTAGAGGAAGGTGCTGCTGAACCTAGCGAAGCTAAGGGTCGG CGTCGTCGTCGTAA
------	--

**Table S2.** DNA sequences of codon-optimised signal peptide and C-tail coding sequences used in this study.

Accession	Organism	Genomic module
WP_013506165.1	<i>Desulfurispirillum indicum</i>	FeS-CytB
WP_014856242.1	<i>Melioribacter roseus</i>	FeS-quinol oxidase-CytC3
WP_049676148.1	<i>Desulfocarbo indianensis</i>	FeS-quinol oxidase- octahaem CytC
WP_028325305.1	<i>Desulfatirhabdium butyrativorans</i>	FeS-quinol oxidase- octahaem CytC
WP_054030243.1	<i>Desulfatitalea tepidiphila</i>	FeS-quinol oxidase- CytC3- octahaem CytC
WP_020589710.1	<i>Desulfobacter curvatus</i>	FeS-CytB
WP_004071579.1	<i>Desulfobacter postgatei</i>	FeS-CytB
WP_084070237.1	<i>Desulfobacterium vacuolatum</i>	FeS-quinol oxidase- octahaem CytC
WP_014957271.1	<i>Desulfobacula toluolica</i>	FeS-CytB
WP_012173729.1	<i>Desulfococcus oleovorans</i>	FeS-CytB
WP_054692440.1	<i>Desulfosarcina cetonica</i>	FeS-quinol oxidase-CytC3-CytC3
WP_022663865.1	<i>Desulfospira joergensenii</i>	FeS-CytB
WP_015723347.1	<i>Desulfobulbus propionicus</i>	CytC3-FeS-quinol oxidase
WP_027370080.1	<i>Desulfovermiculus halophilus</i>	FeS-quinol oxidase- CytC3- octahaem CytC
WP_071544310.1	<i>Desulfovibrio dechloracetivorans</i>	FeS-CytB
WP_071546240.1	<i>Desulfovibrio dechloracetivorans</i>	FeS-quinol oxidase- octahaem CytC
WP_062251202.1	<i>Desulfovibrio fairfieldensis</i>	FeS-CytB
WP_066806246.1	<i>Desulfovibrio indicus</i>	FeS-quinol oxidase-octahaem CytC
WP_008683453.1	<i>Desulfovibrio sp. 3_1_syn3</i>	FeS-CytB
WP_009302998.1	<i>Desulfovibrio sp. 6_1_46AFAA</i>	FeS-CytB
WP_014259604.1	<i>Desulfovibrio africanus</i>	FeS-CytB
WP_027368072.1	<i>Desulfovibrio africanus</i>	FeS-CytB
WP_029459533.1	<i>Desulfovibrio alcoholivorans</i>	FeS-CytB
WP_020886939.1	<i>Desulfovibrio alkalitolerans</i>	FeS-CytB
WP_014323720.1	<i>Desulfovibrio desulfuricans</i>	FeS-quinol oxidase-octahaem CytC

WP_022659754.1	<i>Desulfovibrio desulfuricans</i>	FeS-CytB
WP_024823537.1	<i>Desulfovibrio magneticus</i>	FeS-CytB
WP_015862494.1	<i>Desulfovibrio magneticus</i>	FeS-CytB
WP_015416321.1	<i>Desulfovibrio piezophilus</i>	FeS-CytB
WP_027985824.1	<i>delta proteobacterium PSCGC 5451</i>	FeS-CytB
WP_003398576.1	<i>Anoxybacillus flavithermus</i>	FeS-CytB
WP_008409599.1	<i>Desulfotomaculum hydrothermale</i>	FeS-CytB
WP_012447020.1	<i>Natranaerobius thermophilus</i>	FeS-CytB
WP_011418038.1	<i>Syntrophus aciditrophicus</i>	FeS-CytB
YP_002248525.1	<i>Thermodesulfovibrio yellowstonii</i>	FeS-CytB
WP_051534117.1**	<i>Desulfitibacter alkalitolerans</i> DSM 16504	FeS-CytB
WP_028318203.1**	<i>Desulfobulbus elongatus</i> DSM 2908	CytC3-FeS-quinol oxidase
WP_028580723.1**	<i>Desulfogranum japonicum</i> DSM 18378	CytC3-FeS-quinol oxidase
WP_035247752.1**	<i>Desulfogranum mediterraneum</i> DSM 13871	CytC3-FeS-quinol oxidase
WP_020676593.1**	<i>Geopsychrobacter electrodiphilus</i> DSM 16401	CytC3-FeS-quinol oxidase
WP_004088007.1**	<i>Helicobacter bilis</i>	FeS-CytB
WP_072908093.1**	<i>Malonomonas rubra</i> DSM 5091	CytC3-FeS-quinol oxidase
WP_053334878.1**	<i>Thermoanaerobaculum aquaticum</i>	CytC3-FeS-quinol oxidase

Table S3. 4Fe-4S tail-anchored ferredoxin proteins, and partner proteins with which they are co-encoded at the same genomic locus. \*\* Indicates proteins that were identified in dataset S2.



FAD-containing proteins	
WP_056452081.1	FAD-dependent oxidoreductase <i>Brevundimonas</i> sp. Root1279
WP_082731613.1	FAD-dependent oxidoreductase, <i>Sphingomonas</i>
YP_001636932.1	FAD-dependent pyridine nucleotide-disulfide oxidoreductase, <i>Chloroflexus aurantiacus</i> J-10-fl
WP_082690100.1	FAD-binding monooxygenase, <i>Leptolyngbya</i> sp. NIES-2104
WP_052700213.1**	FAD/NAD(P)-binding oxidoreductase, <i>Methylococcus oryzae</i>
WP_020186848.1**	NAD(P)/FAD-dependent oxidoreductase, <i>Methylophilus</i> sp. 73B
NAD-containing proteins	
WP_046090413.1	NAD-dependent dehydratase, <i>Streptomyces antioxidantans</i>
WP_030280942.1	NAD-dependent dehydratase, <i>Streptomyces catenulae</i>
WP_052860708.1	NAD-dependent dehydratase, <i>Streptomyces celluloflavus</i>
WP_062046107.1	NAD-dependent dehydratase, <i>Streptomyces ciscaucasicus</i>
WP_078643469.1	NAD-dependent dehydratase, <i>Streptomyces hygroscopicus</i>
WP_044570187.1	NAD-dependent dehydratase, <i>Streptomyces iranensis</i>
WP_079429679.1	NAD-dependent dehydratase, <i>Streptomyces katrae</i>
WP_067236815.1	NAD-dependent dehydratase, <i>Streptomyces longwoodensis</i>
WP_055470700.1	NAD-dependent dehydratase, <i>Streptomyces pathocidini</i>
WP_086886330.1	NAD-dependent dehydratase, <i>Streptomyces rhizosphaericus</i>
WP_051779201.1	MULTISPECIES: NAD-dependent dehydratase, <i>Streptomyces</i>
WP_065964212.1	NAD-dependent dehydratase, <i>Streptomyces sparsogenes</i>
WP_053757466.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. AS58
WP_037799844.1	MULTISPECIES: NAD-dependent dehydratase, <i>Streptomyces</i>
WP_069765881.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. LUP47B
WP_073773307.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. MJM1172
WP_055547987.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. NBRC 110028
WP_031076788.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. NRRL S-118
WP_030711692.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. NRRL S-237
WP_037952333.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. PRh5
WP_057612235.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. Root369
WP_009715199.1	MULTISPECIES: NAD-dependent dehydratase, <i>Streptomyces</i>
WP_069815624.1	NAD-dependent dehydratase, <i>Streptomyces</i> sp. TP-A0874
WP_079154649.1	NAD-dependent dehydratase, <i>Streptomyces subbrutillus</i>
WP_030878525.1	NAD-dependent dehydratase, <i>Streptomyces varsoviensis</i>
WP_059145201.1	NAD-dependent dehydratase, <i>Streptomyces violaceusniger</i>
WP_051763101.1	NAD-dependent dehydratase, <i>Streptomyces virginiae</i>

WP_053167991.1	NAD-dependent dehydratase, <i>Streptomyces virginiae</i>
WP_014176621.1	NAD-dependent dehydratase, <i>Streptomyces bingchenggensis</i>
WP_030225883.1	NAD-dependent dehydratase, <i>Streptomyces lavendulae</i>
WP_065964212.1	NAD-dependent dehydratase, <i>Streptomyces sparsogenes</i>
WP_014054769.1	NAD-dependent dehydratase, <i>Streptomyces violaceusniger</i>
WP_087353490.1	NAD-dependent protein deacylase, <i>Olsenella</i> sp. An188
Copper-containing proteins	
WP_026166134.1	copper resistance protein, <i>Corynebacterium mastitidis</i>
WP_085235053.1	copper resistance protein CopC, <i>Mycobacterium conspicuum</i>
WP_066814184.1	copper resistance protein CopC, <i>Mycobacterium</i> sp. 1100029.7
WP_077078633.1	copper resistance protein CopC, <i>Mycobacterium</i> sp. AB215
WP_005248910.1	MULTISPECIES: copper resistance protein C, <i>Rhodococcus</i>
WP_039951718.1	copper resistance protein CopC, <i>Rhodococcus imtechensis</i>
WP_041811727.1	copper resistance protein CopC, <i>Rhodococcus jostii</i>
WP_012691057.1	copper resistance protein C, <i>Rhodococcus opacus</i>
WP_005254749.1	copper resistance protein, <i>Rhodococcus opacus</i>
WP_037228321.1	copper resistance protein CopC, <i>Rhodococcus wratislaviensis</i>
WP_006540442.1	copper resistance protein CopC, <i>Frankia</i> sp. EUN1f
WP_006594284.1	copper resistance protein C, <i>Kineosphaera limosa</i>
WP_052129147.1, WP_053773647.1, WP_081235111.1	copper resistance protein CopC, <i>Clavibacter michiganensis</i>
WP_048809232.1	MULTISPECIES: Copper resistance protein C precursor, <i>Microbacterium</i>
WP_067311181.1	copper resistance protein CopC, <i>Micromonospora rifamycinica</i>
WP_083706190.1**	copper resistance protein CopC, <i>Arthrobacter</i> sp. QXT-31
WP_051866792.1**	copper resistance protein CopC, <i>Corynebacterium atypicum</i>
WP_018023019.1**	copper resistance protein CopC, <i>Corynebacterium doosanense</i> CAU
WP_083352005.1**	copper resistance protein CopC, <i>Leucobacter chromiiresistens</i> DSM 22788
WP_058594659.1**	copper resistance protein CopC, <i>Leucobacter chromiiresistens</i> NS354
WP_082334822.1** WP_076688294.1**	copper resistance protein CopC, <i>Microbacterium aurantiacum</i>
WP_058232366.1**	copper resistance protein CopC, <i>Microbacterium enclense</i>
WP_082784215.1**	copper resistance protein CopC, <i>Microbacterium laevaniformans</i>
WP_084344888.1**	copper resistance protein CopC, <i>Microbacterium resistens</i> NBRC 103078
WP_029266231.1**	copper resistance protein CopC, <i>Microbacterium</i> sp. Cr-K32

WP_082516518.1**	copper resistance protein CopC, <i>Microbacterium</i> sp. Leaf436
WP_054678997.1**	copper resistance protein CopC, <i>Microbacterium</i> sp. No. 7
WP_076674855.1**	copper resistance protein CopC, <i>Microbacterium</i> sp. RU1D
WP_016463500.1**	copper resistance protein CopC, <i>Microbacterium</i> sp. oral taxon 186 str. F0373
WP_067359874.1**	copper resistance protein CopC, <i>Micromonospora rosaria</i>
WP_079705109.1**	copper resistance protein CopC, <i>Plantibacter cousiniae</i>
WP_079001510.1**	copper resistance protein CopC, <i>Plantibacter flavus</i>
WP_085511799.1**	copper resistance protein CopC, <i>Plantibacter flavus</i> DSM 14012
WP_064294246.1**	copper resistance protein CopC, <i>Plantibacter</i> sp. H53
WP_056779050.1**	copper resistance protein CopC, <i>Plantibacter</i> sp. Leaf171
WP_056008599.1**	copper resistance protein CopC, <i>Plantibacter</i> sp. Leaf314
WP_071259775.1**	copper resistance protein CopC, <i>Plantibacter</i> sp. MMLR14_011
WP_086474591.1**	copper resistance protein CopC, <i>Plantibacter</i> sp. VKM Ac-1784
WP_040269875.1**	copper resistance protein CopC, <i>Streptomonospora alba</i>
WP_051814695.1**	copper resistance protein CopC, <i>Streptomyces iakyrus</i>

Table S4. Candidate tail-anchored Tat substrates that are predicted to bind FAD, NAD or copper cofactors. \*\* Indicates proteins that were identified in dataset S2.

Accession	Organism	Further analysis
'Tat pathway protein signal' annotation		
WP_012680923.1	<i>Acidobacterium capsulatum</i>	Has N-terminal pyridoxal phosphate-dependent aminotransferase (PRK09105) domain. The twin arginines form part of this domain and are therefore unlikely to form part of a signal sequence.
WP_050369614.1	<i>Streptomyces acidiscabies</i>	alginate lyase family protein. The C-tail is conserved but the twin arginine motif is not conserved across close homologues
WP_067001745.1	<i>Streptomyces cellostaticus</i>	alginate lyase family protein. The C-tail is conserved but the twin arginine motif is not conserved across close homologues
WP_046705711.1	<i>Streptomyces europaeiscabiei</i>	alginate lyase family protein. The C-tail is conserved but the twin arginine motif is not conserved across close homologues
WP_086800221.1 WP_079024514.1 WP_086747795.1 WP_086754741.1	<i>Streptomyces scabiei</i>	alginate lyase family protein. The C-tail is conserved but the twin arginine motif is not conserved across close homologues
WP_070699546.1	<i>Atopobium</i> sp. HMSC064B08	Protein with an N-terminal peptidoglycan hydrolase CwIO domain and a C-terminal NLPC_P60 domain. The NLPC_P60 domain overlaps with the hydrophobic stretch and it is therefore unlikely to be a C-tail.
'Twin-arginine translocation annotation'		
WP_088345824.1	<i>Rhodomicrobium vannielii</i>	ABC transporter substrate-binding protein, FhuD family. The hydrophobic nature of the C-terminal region is not conserved across close homologues
WP_035708011.1	<i>Haematobacter massiliensis</i>	Has a C-terminal lipid-binding SYLF domain that covers the hydrophobic 'tail' region and is therefore unlikely to be a TMH.
WP_008328674.1	<i>Maritimibacter alkaliphilus</i>	Has a C-terminal lipid-binding SYLF domain that covers the hydrophobic 'tail' region and is therefore unlikely to be a TMH.
WP_072295392.1	<i>Paracoccus</i> sp. SM22M-07	Has a C-terminal lipid-binding SYLF domain that covers the hydrophobic 'tail' region and is therefore unlikely to be a TMH.
WP_068297290.1	<i>Rhodobacter</i> sp. CCB-MM2	Has a C-terminal lipid-binding SYLF domain that covers the hydrophobic 'tail' region and is therefore unlikely to be a TMH
WP_007290909.1 WP_007293159.1	delta proteobacterium MLMS-1	4Fe-4S dicluster domain-containing protein. These protein records were suppressed because they are no longer annotated on any genome and Blast analysis fails to identify any close homologues.
WP_057494184.1 WP_074726554.1	<i>Stenotrophomonas pavanii</i>	Gluconate 2-dehydrogenase subunit 3. The hydrophobic tail region is part of the gluconate_2-dh3 fold and unlikely to be a TMH.
WP_061479189.1	<i>Stenotrophomonas</i> sp. DDT-1	Gluconate 2-dehydrogenase subunit 3. The hydrophobic tail region is part of the gluconate_2-dh3 fold and unlikely to be a TMH.
WP_087921816.1	<i>Stenotrophomonas</i> sp. WZN-1	Gluconate 2-dehydrogenase subunit 3. The hydrophobic tail region is part of the gluconate_2-dh3 fold and unlikely to be a TMH.

WP_049467359.1 WP_049404954.1 WP_046429410.1 WP_065196568.1 WP_065186014.1 WP_058980795.1 WP_072168650.1 WP_057496275.1 WP_065198061.1 WP_064237859.1 WP_049434537.1 WP_087802949.1 WP_088496559.1 WP_053462412.1 WP_053505877.1 WP_053442649.1 WP_054171204.1 WP_062606679.1 WP_049403241.1	<i>Stenotrophomonas maltophilia</i>	Gluconate 2-dehydrogenase subunit 3. The hydrophobic tail region is part of the gluconate_2-dh3 fold and unlikely to be a TMH.
WP_035353858.1**	<i>Edaphobacter aggregans</i> DSM 19364	No homologues in the database.
WP_083145723.1**	<i>Mycolicibacterium parafortuitum</i>	RR conserved. However AlphaFold model suggests that the hydrophobic region is part of the protein fold.
WP_083090608.1**	<i>Rothia nasimurium</i>	Homologue of WP_083145723.1

Table S5. Protein accessions with Tat pathway protein signal or twin-arginine translocation pathway in their annotation. \*\* Indicates proteins that were identified in dataset S2.

Accession	Organism
WP_043259400.1	<i>Streptomyces hokutonensis</i>
WP_043435093.1	<i>Streptomyces pluripotens</i>
WP_006380203.1	<i>Streptomyces turgidiscabies</i>
WP_037801963.1	<i>Streptomyces</i> sp. LaPpAH-108
WP_053849945.1	<i>Streptomyces</i> sp. NRRL B-24085
WP_029393839.1	<i>Streptomyces</i> sp. TOR3209
WP_040902940.1	<i>Streptomyces chartreusis</i>
WP_015655752.1	<i>Streptomyces davawensis</i>
WP_039830567.1	<i>Streptomyces</i>
WP_037666917.1	
WP_014676716.1	<i>Streptomyces hygrosopicus</i>
WP_026151595.1	<i>Streptomyces prunicolor</i>
WP_007386666.1	<i>Streptomyces sviceus</i>
WP_003994962.1	<i>Streptomyces viridochromogenes</i>
WP_037885301.1	
WP_043733548.1	<i>Streptomyces zinciresistens</i>
WP_086565138.1 <sup>#</sup>	<i>Streptomyces africanus</i>
WP_079047754.1 <sup>#</sup>	<i>Streptomyces cyanoalbus</i>
WP_086855524.1 <sup>#</sup>	<i>Streptomyces daghestanicus</i>
WP_075028213.1 <sup>#</sup>	<i>Streptomyces mirabilis</i>
WP_070389912.1 <sup>#</sup>	<i>Streptomyces pactum</i>
WP_079040672.1 <sup>#</sup>	<i>Streptomyces prasinus</i>
WP_078878315.1 <sup>#</sup>	<i>Streptomyces</i>
WP_073890878.1 <sup>#</sup>	<i>Streptomyces</i> sp. CB01883
WP_073934166.1 <sup>#</sup>	<i>Streptomyces</i> sp. CB02400
WP_073728708.1 <sup>#</sup>	<i>Streptomyces</i> sp. CB02414
WP_087804880.1 <sup>#</sup>	<i>Streptomyces</i> sp. CS113
WP_087788440.1 <sup>#</sup>	<i>Streptomyces</i> sp. CS159
WP_079308928.1 <sup>#</sup>	<i>Streptomyces</i> sp. GKU 895
WP_079130289.1 <sup>#</sup>	<i>Streptomyces</i> sp. H-KF8
WP_076082650.1 <sup>#</sup>	<i>Streptomyces</i> sp. IMTB 2501
WP_077797205.1 <sup>#</sup>	<i>Streptomyces</i> sp. JHA26
WP_076972742.1 <sup>#</sup>	<i>Streptomyces</i> sp. M1013
WP_081220707.1 <sup>#</sup>	<i>Streptomyces</i> sp. M41(2017)
WP_071381920.1 <sup>#</sup>	<i>Streptomyces</i> sp. MUSC 1
WP_071371441.1 <sup>#</sup>	<i>Streptomyces</i> sp. MUSC 14
WP_086602166.1 <sup>#</sup>	<i>Streptomyces swartbergensis</i>
WP_086698134.1 <sup>#</sup>	<i>Streptomyces tricolor</i>
WP_086867759.1 <sup>#</sup>	<i>Streptomyces viridochromogenes</i>
WP_081238835.1 <sup>#</sup>	<i>Streptomyces viridosporus</i>
WP_016699903.1 <sup>**</sup>	<i>Actinoalloteichus spitiensis</i>
WP_058266926.1 <sup>**</sup>	<i>Arthrobacter enclensis</i>
WP_016325237.1 <sup>**</sup>	<i>Streptomyces lividans</i>

Table S6. Protein accessions for proteins of the S1 peptidase family. Accessions marked # were annotated as serine protease in dataset S1 but were assigned as S1 peptidase based on current annotation; \*\* indicates proteins that were identified in dataset S2.

.



WP_027241321.1	<i>Pseudophaeobacter arcticus</i>
WP_049167940.1	<i>Corynebacterium propinquum</i>
WP_049150213.1	
WP_075811183.1	<i>Corynebacterium</i> sp. CNJ-954
WP_064834573.1	<i>Corynebacterium</i> sp. EPI-003-04-2554_SCH2473622
WP_006821438.1	<i>Corynebacterium casei</i>
WP_038546560.1	<i>Corynebacterium glyciniphilum</i>
WP_027017577.1	<i>Corynebacterium pseudodiphtheriticum</i>
WP_043687850.1	<i>Nocardia abscessus</i>
WP_039795807.1	<i>Nocardia araoensis</i>
WP_014981882.1	<i>Nocardia brasiliensis</i>
WP_036533687.1	<i>Nocardia cyriacigeorgica</i>
WP_040865004.1	<i>Nocardia exalbida</i>
WP_022597221.1	<i>Rhodococcus hoagii</i>
WP_005518562.1	
WP_013416807.1	
WP_064076622.1	
WP_013416807.1	<i>Rhodococcus equi</i>
WP_069856650.1	<i>Rhodococcus</i>
WP_051901016.1	<i>Streptomyces aureus</i>
WP_010987315.1	<i>Streptomyces avermitilis</i>
WP_059415102.1	<i>Streptomyces azureus</i>
WP_086765287.1	<i>Streptomyces bobili</i>
WP_066996028.1	<i>Streptomyces cellostaticus</i>
WP_086838583.1	<i>Streptomyces daghestanicus</i>
WP_030597466.1	<i>Streptomyces fulvoviolaceus</i>
WP_071657364.1	<i>Streptomyces gilvigriseus</i>
WP_067299969.1	<i>Streptomyces griseochromogenes</i>
WP_078967395.1	<i>Streptomyces griseofuscus</i>
WP_037656023.1	
WP_085566213.1	
WP_085568484.1	
WP_037702946.1	<i>Streptomyces lavenduligriseus</i>
WP_046417906.1	<i>Streptomyces malaysiense</i>
WP_075026534.1	<i>Streptomyces mirabilis</i>
WP_079172327.1	<i>Streptomyces</i>
WP_031119869.1	
WP_039657545.1	
WP_051707667.1	
WP_030994093.1	
WP_086806635.1	<i>Streptomyces murinus</i>
WP_052454075.1	<i>Streptomyces nodosus</i>
WP_079064839.1	<i>Streptomyces olivochromogenes</i>
WP_079101010.1	<i>Streptomyces pathocidini</i>
WP_069779106.1	<i>Streptomyces puniscabiei</i>
WP_079161659.1	
WP_079161534.1	
WP_062702374.1	<i>Streptomyces regalis</i>
WP_059249036.1	<i>Streptomyces reticuli</i>
WP_059252215.1	
WP_078629026.1	<i>Streptomyces</i> sp. 142MFCol3.1
WP_079662723.1	<i>Streptomyces</i> sp. 3124.6

WP_019981628.1	<i>Streptomyces</i> sp. Amel2xE9
WP_081546686.1	<i>Streptomyces</i> sp. B9173
WP_079192310.1	<i>Streptomyces</i> sp. CB01883
WP_043261201.1	<i>Streptomyces</i> sp. e14
WP_067039987.1	<i>Streptomyces</i> sp. FXJ1.172
WP_037801489.1	<i>Streptomyces</i> sp. LaPpAH-108
WP_081216934.1	<i>Streptomyces</i> sp. M41(2017)
WP_053657310.1	<i>Streptomyces</i> sp. MMG1121
WP_079168661.1	<i>Streptomyces</i> sp. MUSC 93
WP_078890368.1	<i>Streptomyces</i> sp. NBRC 110468
WP_053707807.1	<i>Streptomyces</i> sp. NRRL B-3648
WP_079050130.1	<i>Streptomyces</i> sp. NRRL F-5122
WP_033281943.1	<i>Streptomyces</i> sp. NRRL F-525
WP_030345652.1	<i>Streptomyces</i> sp. NRRL S-1022
WP_037861009.1	<i>Streptomyces</i> sp. NRRL S-340
WP_030842128.1	<i>Streptomyces</i> sp. NRRL S-475
WP_030228728.1	<i>Streptomyces</i> sp. NRRL WC-3626
WP_057600964.1	<i>Streptomyces</i> sp. Root1310
WP_079041571.1	<i>Streptomyces</i> sp. Root264
WP_046261706.1	<i>Streptomyces</i> sp. WM6386
WP_086697472.1	<i>Streptomyces tricolor</i>
WP_053196823.1	<i>Streptomyces viridochromogenes</i>
WP_048587024.1	<i>Streptomyces viridochromogenes</i>
WP_067119775.1	<i>Streptomyces yokosukanensis</i>
WP_051754093.1	<i>Streptomyces achromogenes</i>
WP_020273774.1	<i>Streptomyces afghaniensis</i>
WP_054230871.1	<i>Actinobacteria bacterium</i> OK006
WP_054220105.1	<i>Actinobacteria bacterium</i> OK074
WP_054238031.1	<i>Actinobacteria bacterium</i> OV320
WP_018121428.1**	<i>Corynebacterium propinquum</i>
WP_023019056.1**	<i>Corynebacterium</i>

Table S7. Protein accessions for proteins annotated as D-alanyl-D-alanine carboxypeptidase.

\*\* Indicates proteins that were identified in dataset S2.

>WP\_051901016.1 serine hydrolase [Streptomyces aureus]  
MRDSSGLSRRRAALGLAAALPLAAATATTADATSTIGGERLARAGIQVSGASGLPKHLTARSWLVADLTSG  
EVLASYNHLRLPASTLKMLFADAVLKKFERTETKYKVVADLADVPAGSSLVGVKPGITYTVHQLWQGV  
FLRSGNDAVHVL SHMNGGLPATVAEMQALAKDLQALDTHVVS PDGFDHKGQLSSAYDLTLFARHGLADAD  
FRGYCSTKTADFPAGGKKTFFQIQNTDRLLTGAWGVKPYEGLIGVKNGYTS HAGNTFTGAATRDGRTLLVT  
VMHPKSGSGVYEETAQLLDWGF GKGAQAQPVGTLVGPLSEQGASKAPTSAAKAAAAGAPASAGAAKSDGA  
PWGLVGGAGAAVALLAGGALALRRRRRAAGEASEPVD AVEEGAAEPSEAKGRRRRR

>WP\_078967395.1 serine hydrolase [Streptomyces griseofuscus]  
MRESTRLSRRAVLGLTAAALPTS AVLPLTAAAPARAASPATAIGGERLARGGLQVDGAAAVPTKLLTARA  
WLVDGTTGEVLASFNAHQRLAPASTLKMLFADTVLKKFDRTEKYTVKDS DISVPAGSSLVGVKAGITY  
TVEQLWQGVFLRSGNDAVHVL AHMNGGI AKTVAEMQARAQDLQALDTHVVS PDGFDHPGQLSSAYDLTLF  
ARHGLADPDFRGYCGTRIANFPAGGKKTFFQIQNTDRLLTGAWGLKTYDGLIGVKNGYTS HAGNTFTGAAT  
RDGRTLLVTVMHPATGSNAVYEQTAAALLDWGFTHGRTARPVGALVKPLSEGGAAA SPAPAGGQAPAGAPA  
KGPSAARLVEGAAGTAALLAAAGGFLLRRRRRAVEAREAPGDHRPPAPRAAEPDPQEPPAPTDPSGGGRH  
RR

>WP\_086806635.1 D-alanyl-D-alanine carboxypeptidase [Streptomyces murinus]  
MPAPKNTARRSL LISAATLLSLAVTAPAALAAPAAGKAPAATPPAHMSTLGGARLGLPGTQVNLAPGVPV  
LPKDLTARSWIVSDAQSGEVLASHNAHWALPPASTMKMLFADTLLPRFPRTTEHKVAAKDLAGMGESSV  
VGIKPGKTYSVHDLWLGVFLRSGNDAVHVL SAMNGGVAQTVQDMQHHADELQALDTHVVS PDGYDEPHQV  
SSAYDLTLVARSGLQKKDFRDY CSTVSADFGGQEIRNTNRLLSGD TDVPVYQGIAGVKNGNTTHAGATFT  
GVAQRGKQVLLVTVMNPQKQEHNEVYKETAKLFDWGFQAAGKVEPVGELVPPRGAAQTGTAPSGSPQASA  
PAGRAGSKPVAAAATGSRGIGTALAVAGGLLVLLAGGVFLVNRWPLPDLMRRRPRA

>WP\_079064839.1 D-alanyl-D-alanine carboxypeptidase [Streptomyces  
olivochromogenes]  
MRDSSPVPPRAAQLSRRRAALGLAAAVPLAAASPASAA TPVIGGERLARAGVQVGAAAGLPKRLTARSWLV  
ADHQSGEVLASYHAHHLAPASTLKMLFADTVLKNFERTERHRVTDADLADIPAGSSMVGVKPGITYTVE  
QLWLGVFLRSGNDAVHVL SHMNGGLARTVAQMQAQDLQALDTHVVS PDGFDHKGQVSSAYDLTLFARH  
GLADADFRAYCHTKTANFPAGGKKTFFQIQNTDRLLTGAWGVPTYDGLIGVKNGYTSNAGNTFTGAATRGG  
RTLLVTVMHPKSGSGVYEETAALLDWGFKEGAKARSVGTLVDPLSEGGASAAPTRKAARAAAGAPAAAA  
DGGGSSSWGLVGGAGGAVALLAGGAYAWRRRRRVAGDEAPDGRQERHDGQDRRDGRDRRDGRDRRDGQER  
QERHQG

>WP\_053707807.1 D-alanyl-D-alanine carboxypeptidase [Streptomyces sp. NRRL  
B-3648]  
MPAPKTTGRRSLLVTSATLLSLTVTAPAALAAQAPSPSPTATPPAGMSTLGGARLGQPGTQVNLAPGVPV  
LPKELTARSWIVSDAESGEVLAAHNAHWRLPASTLKMLFADTVLPRFPKGT LHKVAPSDLKMGAGSSV  
VGIKEGGTYSVHDLWLGVFLRSGNDAVHVL SAMNGGVAQTVQDMQDHAEELQALDTHVVS PDGYDAPEQV  
SSAYDLTLFARSGLQKRDFRDY CSTVSATFGSTEIRNTNRLLSGDS DVPVYPGIAGVKNGNTTHAGATFT  
GVAERGGKVLLVTVMNPEKHEHNEVYKETAKLLDWGFQAAGKVEPVGELVPPRSAATAQPDPSASAGRDA  
AQAAGRAAGRSGTGGPGVKAAASATTPSGTSGAGIALAVAGGLLALLGGGAFLVNRWPLPDLMRRRAR  
P

>WP\_067119775.1 D-alanyl-D-alanine carboxypeptidase [Streptomyces  
yokosukanensis]  
MPAPKPTGRRSLLVTSATLLSLSATAPAALAAPPPGASPTATPPARMSTVGGERLGQPGTQVNLAPGVPV  
LPKDITARSWIVTDAESGDVLAAHNAHWQLPASTLKMLFADTLLPRFPRTMEHKVVPSDLAGMGVSSV  
VGIKEGRTYNVRDLWLGVFLRSGNDAVHVL SAMNDGVAQTVKDMQAHAEELQARDTHVISPDGYDAPGQV  
SSAYDLSLIARSGLQKKDFRDY CSTVHAKFGDAEIQNTNRLLSGDADVPVYQGIAGVKNGNTTNAGATFT  
GVAERNRVLVTAMNPEKHDHNEVYKETAKLFDWGFRAAGKVNVPV GELVPPKSALPSPGASASAGQSAG  
PHGSGSKPVADATAESGSGGVGIAAAIAGGVLLVLLAGGVFLINRRWPLPDLMRRRP RP

>CAB89066.1 putative D-alanyl-D-alanine carboxypeptidase [Streptomyces  
coelicolor A3(2)]  
MPAPKKALRRSLTVTSAALTALALLSPAALAAPSPSTSPSATPPASMSGVGGARLGKPGTQVDLAGGAPV  
LPKDLTARSWIVADAESGDVLAAHNAHWELAPASTLKMLFADTLLPKWPKTTKRKVQVSDLAGIGSGSSM  
VGIIKEEETYTVHDLWLGVFLRSGNDAVHVL SAMNDGVEKTVADMNAHAKELQAGDTTVVSPDGYDMPGQV  
SSAYDLTLFARSLQKKDFREYASTVRAKFPGETKKDKKGKTSREPFEIQNTNRLLAGDADVEVYPGIAG  
VKNGYTTNAGNTFTGVAEQGGKVLVTVMHPAEDKHNEVYKETARLFDWGFRAAGKVRPVGELVRPRSAT  
AASP SAEASGEAGGTGKGGAGTKPVAGAAAGDGAGGAGVAAGIAAGVLALLAGGAFLINRKWPLPDLVRR  
RTRP

Accession	Organism	Comments/further analysis
WP_034515529.1	<i>Actinomyces slackii</i>	Choice-of-anchor L domain-containing protein. Predicted to be a Sec signal peptide by SignalP 6.0
WP_023019151.1	<i>Corynebacterium</i>	LPXTG cell wall anchor domain-containing protein. Predicted to be a Sec signal peptide by SignalP 6.0. RR not conserved across closely related protein sequences
WP_016464758.1	<i>Microbacterium</i> sp. oral taxon 186	Tyrosine protein phosphatase. Misannotated previously
WP_051814589.1	<i>Streptomyces iakyrus</i>	LPXTG cell wall anchor domain-containing protein. Potentially wrong start codon assigned in this accession. RR not conserved.
WP_046502582.1	<i>Streptomyces</i> sp. NRRL B-24891	LAETG motif-containing sortase-dependent surface protein. This particular accession predicted to be Tat dependent by SignalP 6.0, although the RR not conserved across closely related protein sequences
WP_010637519.1	<i>Streptomyces</i> sp. S4	HtaA domain-containing protein, partial. This particular accession predicted to be Tat dependent by SignalP 6.0. Lots of closely related/identical proteins by Blast.
WP_015611441.1	<i>Streptomyces fulvissimus</i>	LPXTG cell wall anchor domain-containing protein. Strongly predicted Tat dependent by SignalP 6.0. RR motif not always conserved
WP_051840821.1	<i>Streptomyces lavendulae</i>	Annotated as LPXTG in dataset S1 but current annotation indicates it is an LAETG motif-containing sortase-dependent surface protein. The RR motif not conserved across close homologues
WP_013151425.1	<i>Nocardiopsis dassonvillei</i>	LPXTG cell wall anchor domain-containing protein. The RR motif is not conserved across homologous proteins
WP_016115629.1	<i>Bacillus cereus</i>	M4 family metallopeptidase. The RR motif not conserved across close homologues
WP_016133109.1	<i>Bacillus cereus</i>	M4 family metallopeptidase. Identical protein to WP_016115629.1
WP_007790986.1	<i>Peptostreptococcus stomatis</i>	Ig-like domain-containing protein. This bacterium lacks <i>tatA</i> or <i>tatC</i> genes so cannot be a Tat substrate
WP_067783620.1**	<i>Actinomyces vulturis</i>	LPXTG cell wall anchor domain-containing protein. Strongly predicted to be Tat dependent by SignalP 6.0. Very few related proteins, RR not conserved.
WP_034625822.1**	<i>Cellulomonas cellasea</i>	LPXTG cell wall anchor domain-containing protein. Weak prediction for Tat dependence by SignalP 6.0. Homologues generally also have paired Arg but also with weak prediction for Tat dependence
WP_087509670.1**	<i>Cellulomonas iranensis</i>	LPXTG cell wall anchor domain-containing protein. Strong prediction for Tat dependence. twin-arginine reasonably well conserved across homologous proteins
WP_070320500.1**	<i>Cellulomonas iranensis</i> NBRC 101100	LPXTG cell wall anchor domain-containing protein. Very close homologue of WP_087509670.1
WP_046529179.1**	<i>Cellulomonas</i> sp. FA1	LPXTG cell wall anchor domain-containing protein. Very close homologue of WP_087509670.1
WP_005209628.1**	<i>Clostridium celatum</i> DSM 1785	LPXTG cell wall anchor domain-containing protein. Not recognised as a Tat signal peptide by SignalP 6.0. Very few homologues
WP_087014056.1**	<i>Leucobacter</i> sp. 7(1)	SpaH/EbpB family LPXTG-anchored major pilin. Close homologues predicted to be Sec dependent
WP_057873001.1**	<i>Loigolactobacillus rennini</i> DSM 20253	LPXTG cell wall anchor domain-containing protein. This bacterium lacks <i>tatA</i> or <i>tatC</i> genes so cannot be a Tat substrate
WP_049350406.1**	<i>Rothia mucilaginosa</i>	LPXTG cell wall anchor domain-containing protein. The RR motif not conserved across close homologues
WP_037602190.1**	<i>Streptacidiphilus rugosus</i> AM-16	LPXTG cell wall anchor domain-containing protein. Partial conservation of RR motif in homologous proteins
WP_037580257.1**	<i>Streptococcus equi</i> subsp. <i>ruminatorum</i> CECT 5772	LPXTG cell wall anchor domain-containing protein. This bacterium lacks <i>tatA</i> or <i>tatC</i> genes so cannot be a Tat substrate
WP_060178439.1**	<i>Streptomyces</i> sp. IMTB 1903	LPXTG cell wall anchor domain-containing protein. RR motif well conserved across homologous proteins

WP_030231344.1**	<i>Streptomyces</i> sp. NRRL S-350	LPXTG cell wall anchor domain-containing protein. RR motif not conserved across homologous proteins
WP_018509850.1**	<i>Streptomyces</i> sp. SID5594	LPXTG cell wall anchor domain-containing protein. RR motif well conserved across homologous proteins
WP_046259605.1**	<i>Streptomyces</i> sp. WM6386	LPXTG cell wall anchor domain-containing protein. RR motif well conserved across homologous proteins
WP_086778847.1	<i>Streptomyces fimicarius</i>	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_076967992.1	<i>Streptomyces</i> sp. IB2014 011-1	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_030568954.1	<i>Streptomyces cyaneofuscatus</i>	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_084749454.1	<i>Streptomyces albolongus</i>	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_032758621.1	<i>Streptomyces albobiviridis</i>	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_031122887.1	<i>Streptomyces</i> sp. NRRL S-623	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_055558699.1	<i>Streptomyces luridiscabiei</i>	Annotated as a peptidase in dataset 1 but current annotation is LAETG motif-containing sortase-dependent surface protein
WP_070013327.1**	<i>Streptomyces abyssalis</i>	SCO1860 family LAETG-anchored protein
WP_004943253.1**	<i>Streptomyces mobaraensis</i> NBRC 13819	SCO1860 family LAETG-anchored protein
WP_070014799.1**	<i>Streptomyces nanshensis</i>	SCO1860 family LAETG-anchored protein
WP_064073155.1**	<i>Streptomyces noursei</i>	SCO1860 family LAETG-anchored protein
WP_055573436.1**	<i>Streptomyces prasinopilosus</i> CGMCC 4.3504	SCO1860 family LAETG-anchored protein
WP_055694013.1**	<i>Streptomyces prasinopilosus</i> NRRL B-2711	SCO1860 family LAETG-anchored protein
WP_055606274.1**	<i>Streptomyces prasinus</i> ATCC 13879	SCO1860 family LAETG-anchored protein
WP_055688174.1**	<i>Streptomyces prasinus</i> NRRL B-2712	SCO1860 family LAETG-anchored protein
WP_073752756.1**	<i>Streptomyces</i> sp. CB03234	SCO1860 family LAETG-anchored protein
WP_065002969.1**	<i>Streptomyces</i> sp. H-KF8	SCO1860 family LAETG-anchored protein
WP_030791129.1**	<i>Streptomyces</i> sp. NRRL S-920	SCO1860 family LAETG-anchored protein
WP_078852820.1**	<i>Streptomyces</i> sp. NRRL S-1824	SCO1860 family LAETG-anchored protein
WP_069464636.1**	<i>Actinacidiphila rubida</i>	LAETG motif-containing sortase-dependent surface protein. RR motif not conserved across homologous proteins
WP_030079189.1**	<i>Streptomyces baarnensis</i>	LAETG motif-containing sortase-dependent surface protein. RR motif not conserved across homologous proteins
WP_044384405.1**	<i>Streptomyces cyaneogriseus</i> subsp. <i>noncyanogenus</i>	LAETG motif-containing sortase-dependent surface protein. RR in wrong position relative to h-region
WP_029385521.1**	<i>Streptomyces leeuwenhoekii</i> C34	LAETG motif-containing sortase-dependent surface protein. Almost identical to WP_044384405.1
WP_048572334.1**	<i>Streptomyces leeuwenhoekii</i> C58	LAETG motif-containing sortase-dependent surface protein. Almost identical to WP_044384405.1
WP_059130410.1**	<i>Streptomyces</i> sp. NRRL F-5122	LAETG motif-containing sortase-dependent surface protein. RR motif not conserved across homologous proteins
WP_053642605.1**	<i>Streptomyces</i> sp. NRRL F-6491	LAETG motif-containing sortase-dependent surface protein. RR motif not conserved across homologous proteins

Table S8. Protein accessions with LPXTG or LAETG in their annotation. \*\* indicates the proteins were identified in dataset S2.