Supplementary Infromation

Characterization of the post-assembly line tailoring processes in teicoplanin biosynthesis

Oleksandr Yushchuk[a], Bohdan Ostash[a], Thu H. Pham[b], Andriy Luzhetskyy[c,d], Victor Fedorenko[a], Andrew W. Truman[b*], Liliya Horbal[a,c*]

[a] Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Ukraine

[b] Department of Molecular Microbiology, John Innes Centre, Colney Lane, Norwich, UK

[c] Department of Pharmaceutical Biotechnology, Saarland University, Campus, Saarbrucken, Germany

[d]Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS) Helmholtz Center for Infectious Research (HZI), Saarbrucken, Germany

*Corresponding author for genetics – Liliya Horbal (lihorbal@gmail.com) and for chemistry – Andrew W. Truman (andrew.truman@jic.ac.uk)

Inventory of Supporting Information

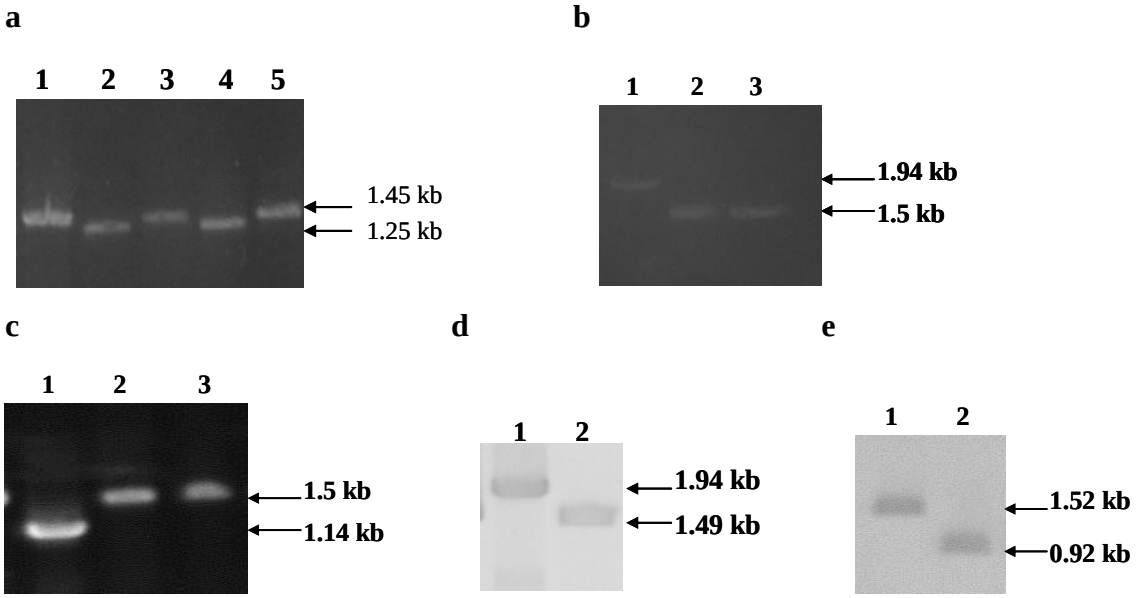


Figure S1. Inactivation of the *tei10**, *tei3**, *tei11**, *tei13** and *tei30** genes. **A** Results of the PCR amplification of the *tei10** gene from the chromosomal DNA of the Δtei10* mutant (1, 3, 5) and the wild type strain (2, 4). **B** Results of the PCR amplification of the *tei3** gene from the chromosomal DNA of the wild type strain (1) and the Δ*tei3** mutants (2, 3). **C** Results of the PCR amplification of the *tei11** gene from the chromosomal DNA of the wild type strain (1) and the Δ*tei3** mutants (2, 3). **D** Results of the PCR amplification of the *tei13** gene from the chromosomal DNA of the wild type strain (1) and the Δ*tei13** mutant (2). **E** Results of the PCR amplification of the *tei30** gene from the chromosomal DNA of the Δ*tei30** mutant (1) and the wild type strain (2).

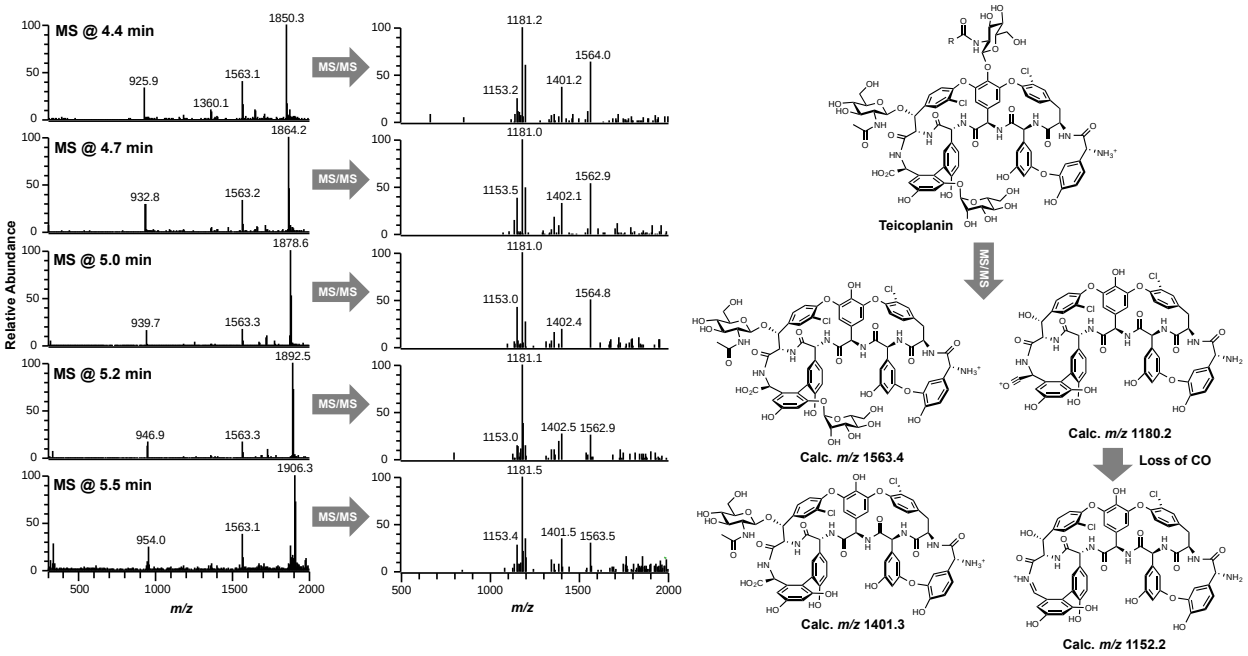


Figure S2. MS/MS analysis of teicoplanins shown in Figure 1 illustrating that they all fragment in an identical manner, and therefore differ in the nature of the acyl chain attached to the glucosaminyl group on amino acid 4. Fragment peaks differ by ± 1 due to the variety of teicoplanin isotope peaks that can be selected for fragmentation.

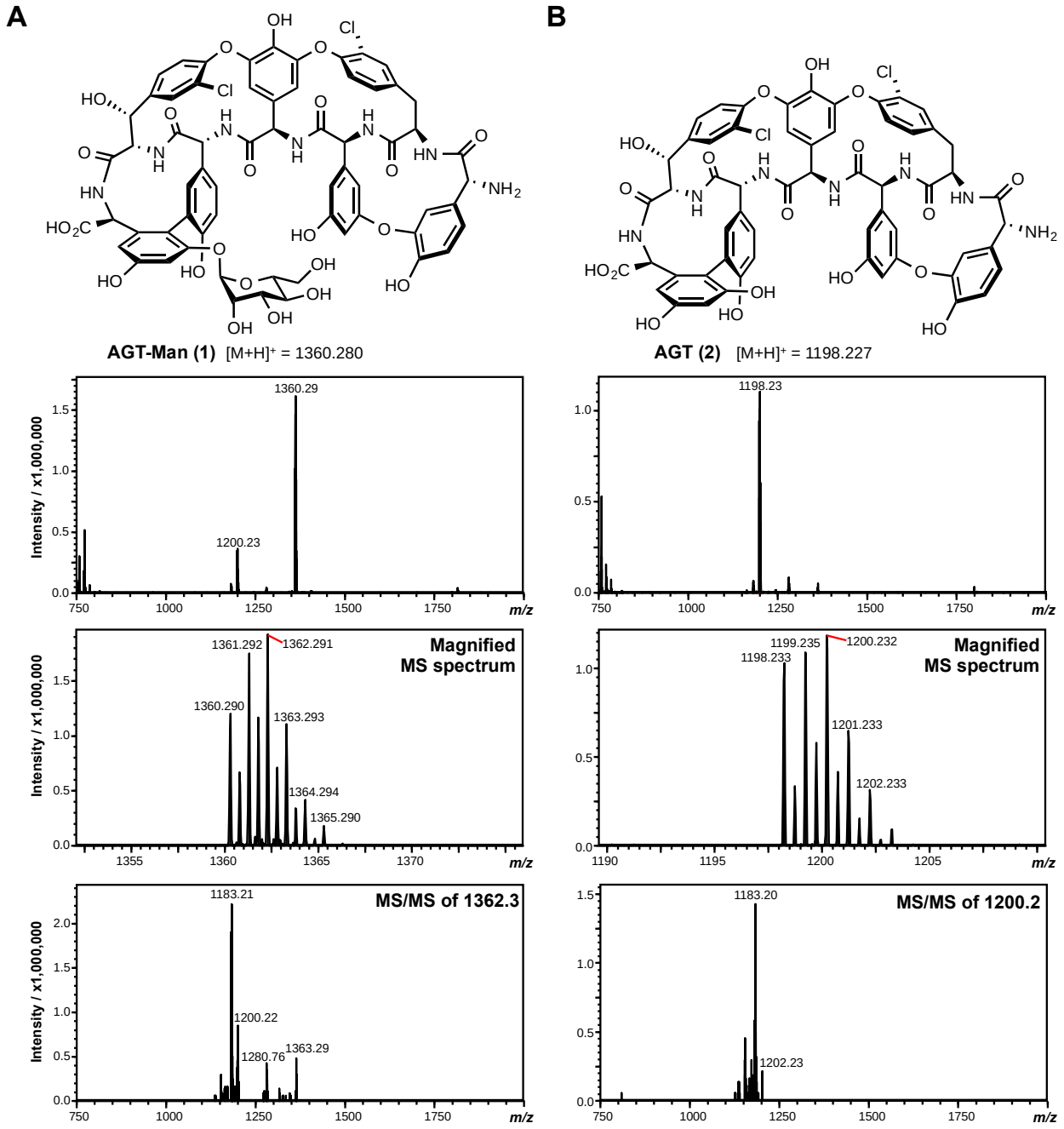
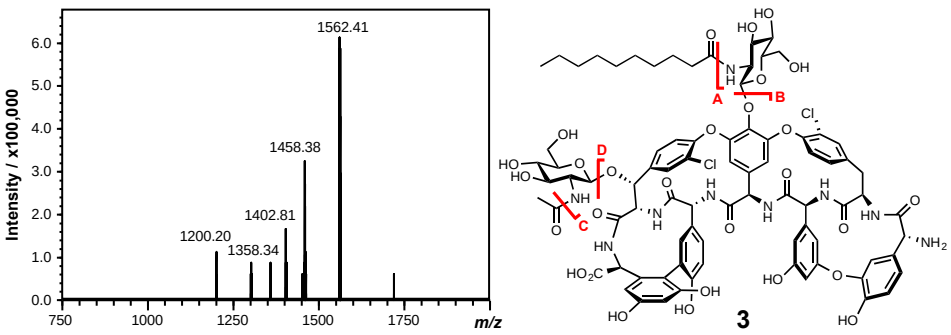
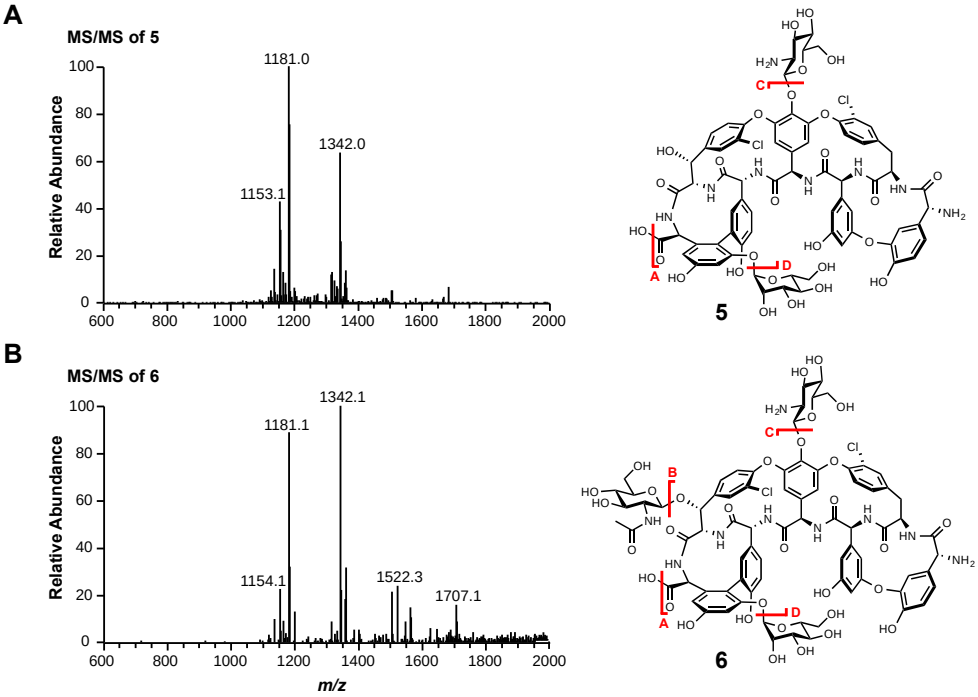


Figure S3. High-resolution MS/MS analysis of AGT-mannose (**1**, panel A) and AGT (**2**, panel B) produced by *A. teichomyceticus* Δ*tei10**. The magnified MS spectrum indicates some unusual peaks spaced by 0.5 mass units, which could reflect the presence of $[2M+2H]^{2+}$ in the mass spectrometer. This would be consistent with the ability of many glycopeptides to dimerize, although this phenomenon was not observed for all compounds identified in this study.



Loss	Predicted fragment *m/z*
A	1562.37
B	1401.31
B+C	1359.30
B+D	1198.22

Figure S4. MS/MS analysis of **3** produced by *A. teichomyceticus* Δ*tei3**. Differences between predicted and observed masses are likely to arise from the number of isotopes that can be selected for fragmentation (*m/z* 3 isolation width).



Loss	Predicted fragment *m/z*
A+C	1342.3
A+C+D	1180.2
A+C+D - CO	1152.2

Loss	Predicted fragment *m/z*
A	1706.4
B	1521.3
A+B+C	1342.3
A+B+C+D	1180.2
A+B+C+D - CO	1152.2

Figure S5. MS/MS analysis of **5** (panel A) and **6** (panel B) produced by *A. teichomyceticus* Δ*tei11**.

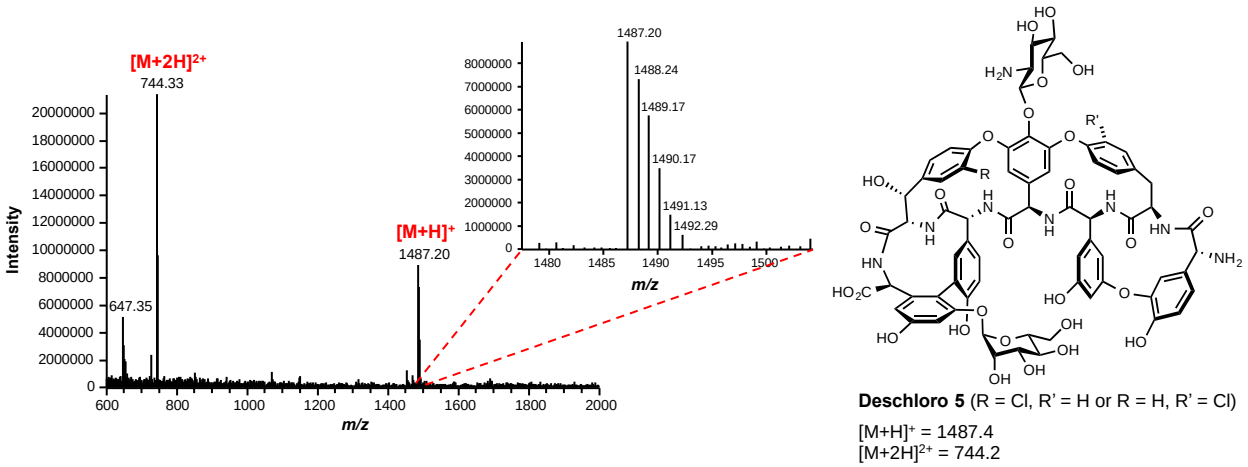


Figure S6. MS spectrum of deschloro **5** produced by *A. teichomyceticus* Δ*tei11**. The isotope pattern is consistent with the illustrated monochlorinated compound ($[M+H]^+$ = $C_{70}H_{68}ClN_8O_{27}^+$).

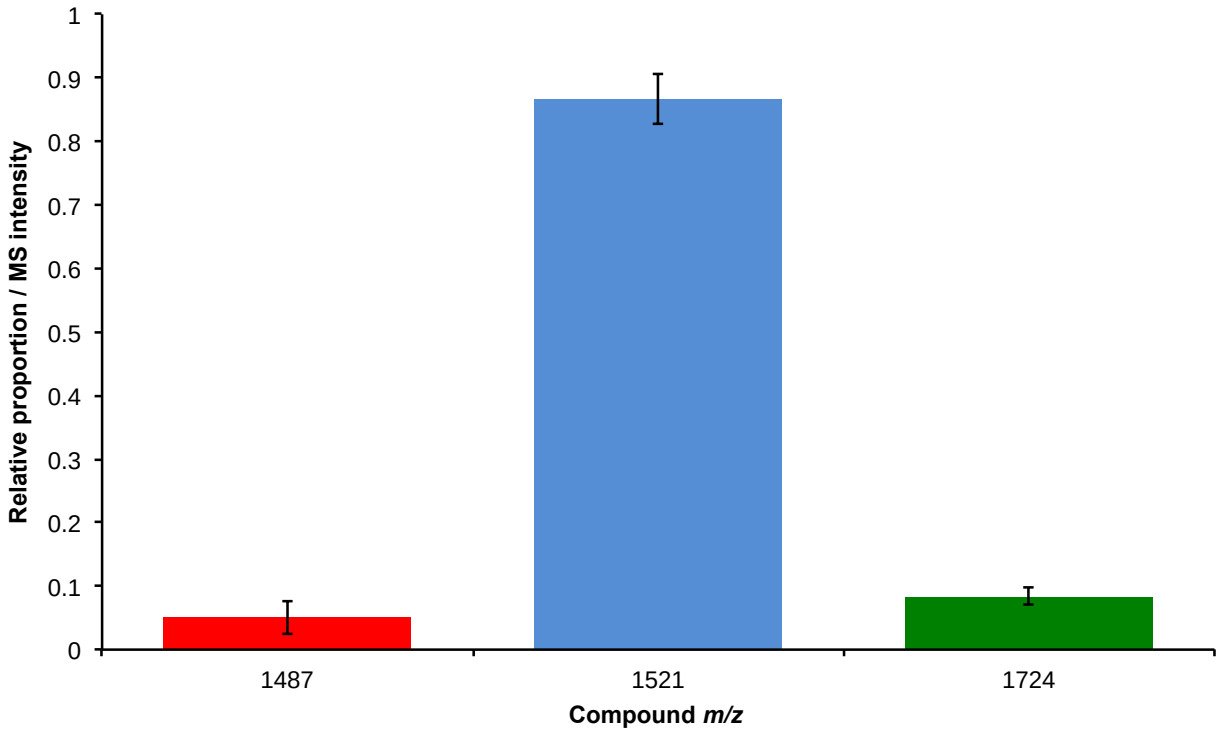


Figure S7. Relative proportion of compounds produced by *A. teichomyceticus* Δ*tei11**. *m/z* 1521 is compound **5**, *m/z* 1487 is a deschloro version of **5**, and *m/z* 1724 is compound **6**.

```
Tei13*      --------MG---YD-MATAPDLVTLIRQHVAGRPDADAVGFLTDPDDIRGGVVSWTYGQ  48
DptE        --------MS---ES-RCAGQGLVGALRTWARTRARETAVVLV-RDTGTTDDTASVDYGQ  47
RevS        MELALPAELAPT----------LPEALRLRSEQQPDTVAYVFLRDGETP---EETLTYGR  47
EcFAAL      --MSNKIFTHSLPMR-YADFPTLVDALDYAALSSA---GMNFYDR---RCQLEDQLEYQT  51
LpFAAL      ----------MKKE-YLQCQSLVDVVRLRALHSPNKKSCTFLNKE-----LEETMTYEQ  43
MtFAAL      --MALGSLLPELLPG-GSGRTPLIERAARWAREKPDAPAYTFVDYSADPAGAHVTVSWAE  57
AsFACL      --------MQ-----------TVNEMLRRAATRAPDHCALA--V-----PARGLRLTHAE  34
TtFACL      ---MEGERMNAFPSTMMDEELNLWDFLERAAALFGRKEVVSRLH-----TGEVHRTTYAE  52
                                 :

Tei13*      LDREARGYAAWLQQRL-PAGSRVLLLYPNGLDFVAAFFGCLYAGMIAVPAPLPGRYRHH-  106
DptE        LDEWARSIAVTLRQQL-APGGRALLLLPSGPEFTAAYLGCLYAGLAAVPAPLPGGRHFE-  105
RevS        LDRAARARAAALEAAG-LAGGTAVLLYPSGLEFVAALLGCMYAGTAGAPVQVPTRRRGM-  105
EcFAAL      LKARAEAGAKRLLSLNLKKGDRVALIAETSSEFVEAFFACQYAGLVAVPLAIPMGVGQRD  111
LpFAAL      LDQHAKAIAATLQAEGAKPGDRVLLLFAPGLPLIQAFLGCLYAGCIAVPIYPPAQEKLL-  102
MtFAAL      TDRRARAVAATLRQVS-GPGERAALLLPQTLEYMMTMLGAMYAHVIAVPLFSPDLPGHA-  115
AsFACL      LRARVEAVAARLHADGLRPQQRVAVVAPNSADVVIAILALHR--LGAVPALLNPRLK---  89
TtFACL      VYQRARRLMGGLRALGVGVGDRVATLGFNHFRHLEAYFAVPG--MGAVLHTANPRLS---  107
                ..    *         .  :         : :.       ..

Tei13*      --QHRVATISADARVSAVLTTAAHLGEVRDWARACGL-DHL----LIAVGDEPEFGDPAG  159
DptE        --RRRVAAIAADSGAGVVLTVAGETASVHDWLTETTAPATR----VVAVDDRAALGDPAQ  159
RevS        ---ERARRIADDAGAKTILTTTAVKREVEEHFADLL--TGL----TVIDTESLPDVPDDA  156
EcFAAL      SWSAKLQGLLASCQPAAIITGDEWLPLVNAATHDN--PEL-----HVLSHAWFKALPEAD  164
LpFAAL      ---DKAQRIVTNSKPVIVLMIADHI---KKFTADELNTNPKFLKIPAIALESIELNRSSS  156
MtFAAL      ---DRLIGAYADSEPAVIVTTRNALPHVEKFLADHDVPQPK----EILFAEEIDPALADR  168
AsFACL      --SAELAELIKRGEMTAAVIAVGRQ-VADAIFQSG-SGAR-----IIFLGDLVRDGEPYS  140
TtFACL      --PKEIAYILNHAEDKVLLFDPNLLPLVEAIRGELKTVQH-----FVVMDEKAPEG-YLA  159
                .             :

Tei13*      --------WTPASPERATIALLQYTSGSTGDPKGVVVTHDNILYNLD-ACVRGL-RWPDD  209
DptE        --------WDDPGVAPDDVALIQYTSGSTGNPKGVVVTHANLLANAR-NLAEAC-ELTAA  209
RevS        --------PAVRLPGPDDVALLQYTSGSTGDPKGVEVTHANFRANVA-ETVELW-PVRSD  206
EcFAAL      --------VALQRPVPNDIAYLQYTSGSTRFPRGVIITHREVMANLR-AISHDGIKLRPG  215
LpFAAL      --------WQPTSIKSNDIAFLQYTSGSTMHPKGVMVSHHNLLDNLN-KIFTSF-HMNDE  206
MtFAAL      --------WEDEPIGFDDVAYLQYTSGSTRRPAGVEITHGNVTANAA-QLWAGWAPERPN  219
AsFACL      YGPP---IEDPQREPA-QPAFIFYTSGTTGLPKAAIIPQRAAESRVLFMSTQVGLRHGRH  196
TtFACL      YEEALGEEADPVRVPERAACGMAYTTGTTGLPKGVVYSHRALVLHSLAASLVDGTALSEK  219
                          . : **:*:*  * ..   :     .

Tei13*      WRVGGWLPLYHDLAMQGLLNMAVVRGGYALLMEPVSFVRDPVRWLRTIAEHDI--QVTFA  267
DptE        TPMGGWLPMYHDMGLLGTLTPALYLGTTCVLMSSTAFIKRPHLWLRTIDRFGL--VWSSA  267
RevS        GTVVNWLPLFHDMGLMFGVVMPLFTGVPAYLMAPQSFIRRPARWLEAISRFRGTH--AAA  264
EcFAAL      DRCVSWLPFYHDMGLVGFLLTPVATQLSVDYLRTQDFAMRPLQWLKLISKNRG--TVSVA  273
LpFAAL      TIIFSWLPPHHDMGLIGCILTPIYGGIQAIMMSPFSFLQNPLSWLKHITKYKA--TISGS  264
MtFAAL      PELVSWLPLFHDMGLISTMALPLVNGDHAIYTDPVSFIMNPMRWLQLIASRPGRNVYTAG  279
AsFACL      NVVLGLMPLYHVVGFFAVLVAALALDGTYVVIEE----FRPVDALQLVQQEQV--TSLFA  250
TtFACL      DVVLPVVPMFHVNAWCLPYAATLVGAKQVLPGPR----LDPASLVELFDGEGV--TFTAG  273
                  :* .*  .        :                 *   :. .           .

Tei13*      PTFAYQLCLDRVT-DEQLAALDLSGWKIAGNAAEPVNPAILAAFAEKFAPAGFRPESFAP  326
DptE        PDFAYDMCLKRVT-DEQIAGLDLSRWRWAGNGAEPIRAATVRAFGERFARYGLRPEALTA  326
RevS        PSFAYELCVRSVADTGLPAGLDLSSWRVAVNGAEPVRWTAVADFTEAYAPAGFRPQAMCP  324
EcFAAL      PPFGYELCQRRVN-EKDLAELDLSCWRVAGIGAEPISAEQLHQFAECFRQVNFDNKTFMP  332
LpFAAL      PNFAYDYCVKRIR-EEKKEGLDLSSWVTAFNGAEPVREETMEHFYQAFKEFGFRKEAFYP  323
MtFAAL      PNFAFEYV-ASVATPEKIAGLDLSGLTTCLNGAEPIRPSTLATFAEVLAPAGLRPGAQAP  338
AsFACL      TPTHLDALAAAA--AHAGSSLKLDSLRHVTFAGATMPDAVLETVHQH------LPGEKVN  302
TtFACL      VPTVWLALADYL--ES--TGHRLKTLRRLVVGGSAAPRSLIARFER-------MGVEVRQ  322
                              *.       ..       :  . .

Tei13*      IYGMAEATGYISGEVGRAP-LVRTVGLDALAHGRLAD-P-APDEPVREIVSCGTPNEACD  383
DptE        GYGLAEATLFVSRSQGLHT-A--RVATAALERHEFRL-A-VPGEAAREIVSCGPVG-HFR  380
RevS        GYGLAENTLKLSGSPEDRPPTLLRADAAALQDGRVVPLT-GPGTDGVRLVGSGVTVPSSR  383
EcFAAL      CYGLAENALAVSFSDEASGVVVNEVDRDILEYQGKAVAPGAETRAVSTFVNCGKALPEHG  392
LpFAAL      CYGLAEATLLVTGGTPGSSYKTLTLAKEQFQDHRVHFAD-DNSPGSYKLVSSGNP--IQE  380
MtFAAL      GYGLAEATVFVTAAAGDGPPKVISVDREALTRGELVLRD-GG----SELVSCGG-PCGQL  392
AsFACL      IYGTTEAMNSLYMRQPKTGTE----------------------------MAPGFFSEVRI  334
TtFACL      GYGLTETSPVVVQNFVKSHLE-----------------SLSEEEKLTLKAKTGLPIPLVR  365
              ** :*    :                                        *

Tei13*      IRVVDPETSRVRPDGW-LGEIWI-RGRSVSPGYWSDAGP---AFAAVTA----EGEDGFL  434
DptE        ARIVEPGGHRVLPPGQ-VGELVL-QGAAVCAGYWQAKEETEQTFGLTLD----GEDGHWL  434
RevS        VAVVDPGTGTEQPAGR-VGEIWI-NGPCVARGYHGRPAESAESFGARIA--GQEARGTWL  439
EcFAAL      IEIR-NEAGMPVAERV-VGHICI-SGPSLMSGYFGDQVSQ-----------DEIAATGWL  438
LpFAAL      VKIIDPDTLIPCDFDQ-VGEIWV-QSNSVAKGYWNQPEETRHAFAGKIK--DDERSAIYL  436
MtFAAL      VAIVDPETRTEQPDGR-VGEIWV-HGPNTAPGYWRNSERSRDTFGGELNDPGDLPAGPWM  450
AsFACL      VRIGGGVD-EIVANGE-EGELIVAASDSAFVGYLNQPEA-T----------AEKLQDGWY  381
TtFACL      LRVADEEGRPVPKDGKALGEVQL-KGPWITGGYYGNEEATR----------SALTPDGFF  414
              :             *.: :  .     **                        :
```

Figure S8. Comparative sequence alignment of the Tei13* with well-characterized fatty acyl-AMP ligases (FAALs) and fatty acyl-coenzyme A ligases (FACLs) based on an analysis from *J. Mol. Biol.* **2011**, *406*, 313. The acyl-activating enzyme (AAE) consensus motif [LIVMFY]-X-X-[STG]-[STAG]-G-[ST]-[STEI]-[SG]-X-[PASLIVM]-[KR] (*Plant Physiol.* **2003**, *132*, 1065) is highlighted in bold, and additional motifs are color-coded as described in *J. Mol. Biol.* **2011**, *406*, 313: yellow = gate motif of adenylation active site; green = insertion motif characteristic of FAALs; cyan = hinge region between domains; red = active site loop motif. Proteins used in alignment: DptE = daptomycin pathway acyl-AMP ligase from *Streptomyces roseosporus*; RevS = reveromycin pathway FACL from *Streptomyces* sp. SN-593; EcFAAL = *Escherichia coli* FAAL; LpFAAL = *Legionella pneumophila* FAAL; MtFAAL = *Mycobacterium tuberculosis* FAAL; AsFACL = *Alcaligenes* sp. FACL; TtFACL = *Thermus thermophiles* FACL.

	Tei13*	DptE	RevS	EcFAAL	LpFAAL	MtFAAL	AsFACL	TtFACL
Tei13*	100	42.02	37.12	29.58	31.96	34.1	19.75	21.81
DptE	42.02	100	37.25	29.43	32.74	36.22	20.37	21.76
RevS	37.12	37.25	100	31.39	35.07	40.63	19.29	19.33
EcFAAL	29.58	29.43	31.39	100	30.36	30.48	20.37	19.49
LpFAAL	31.96	32.74	35.07	30.36	100	34.51	17.22	17.72
MtFAAL	34.1	36.22	40.63	30.48	34.51	100	18.72	21.62
AsFACL	19.75	20.37	19.29	20.37	17.22	18.72	100	26.31
TtFACL	21.81	21.76	19.33	19.49	17.72	21.62	26.31	100

Figure S9. Percentage identity matrix of proteins shown in Figure S8.

```
RifR            MHRPEAEKWLRRFERAPDARARLVCLPHAGGSASFFFPLAKALAPAVEVLAVQYPGRQDR 60
tei30*          MPVP----WFSAPRPLAAPRLRLVCFPYAGGNAATYRSWAGLLPPGVELVAALLPGRAER 56
SrfA            -----MSQLFKSFD--ASEKTQLICFPFAGGYSASFRPLHAFLQGECEMLAAEPPGHGTN 53
                         :      .  : :*:*:*.*** :: : .    *    *::*.  **:  .

RifR            RHEPPVDSIGGLTNRLLEVLRP-FGDRPLALFGHSMGAIIGYELALRMPEAGLPAPVHLF 119
tei30*          LDEPPLVDLDVLLSELVAAAGPLLGPVPLILFGHSLGATVAYEFGRALATEHGCVPAALL 116
SrfA            -QTSAIEDLEELT-DLYKQELNLRPDRPFVLFGHSMGGMITFRLAQKLEREG-IFPQAVI 110
                 . ..: .:  *   *           *: *****:*. : :.:.  :       *  ::

RifR            ASGRRAPSRYRDDDVRGASDERLVAELRKLGGSDAAMLADPELLAMVLPAIRSDYRAVET 179
tei30*          VSGAPAPPVRRRRTGRGLSDDDLERILGKRTDLPAGWLT-AELKPFVFPALRADLQLLDG 175
SrfA            ISAIQPPHIQRKK-VSHLPDDQFLDHIIQLGGMPAELVENKEVMSFFLPSFRSDYRALEQ 169
                 *.  .*   *       .*: :   : :  .  *  :   *: .:.:*::*:* : ::

RifR            YRHEPGRRVDCPVTVFTGDHDPRVSVGEARAWEEHTTGPADLRVLPGGHFFLVDQAA-PM 238
tei30*          YRWRPGPLPGVPVVAFGGDADPDVSAADLRAWQRCTTGPARTHVLAGDHFFIAHQPFREL 235
SrfA            FELYDLAQIQSPVHVFNG-LDDKKCIRDAEGWKKWAK-DITFHQFDGGHMFLLSQTE-EV 226
                :.         ** .* *  *   .  : ..*:. :.     : : *.*:*:  *.   :

RifR            IATMTEKLAGPALTGSTGGNS 259
tei30*          LAATVGEFAASPIAR------ 250
SrfA            AERIFAILNQHPIIQP----- 242
```

Figure S10. Comparative sequence alignment of the Tei30* with well-described type II thioesterase RifR from *Amycolatopsis mediterranei* and SrfA from *Bacillus subtilis*. Residues of the catalytic triad are highlighted in yellow, amino acids important for the activity of the catalytic center are colored in green, and residues that differ in proteins are depicted in red. Fully conserved residues are denoted with asterisk, amino acids with strongly similar properties – with colon and weakly similar – with period.