Supporting Information

Complete Elucidation of the Late Steps of Bafilomycin Biosynthesis in *Streptomyces lohii*

Zhong Li[1,2], Lei Du[1,2] Wei Zhang[1], Xingwang Zhang[1], Yuanyuan Jiang[1,2], Kun Liu[1], Ping Men[1], Huifang Xu[1], Jeffrey L. Fortman[3], David H. Sherman[3], Bing Yu[4], Song Gao[4], and Shengying Li[1]*

[1]Shandong Provincial Key Laboratory of Synthetic Biology, and CAS Key Laboratory of Biofuels at Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong 266101,
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Life Sciences Institute, Departments of Medicinal Chemistry, Chemistry, and Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA.
[4]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510060, China.

Running title: Post-PKS tailoring steps in bafilomycin biosynthesis

*To whom correspondence may be addressed. E-mail: lishengying@qibebt.ac.cn

Table of Contents

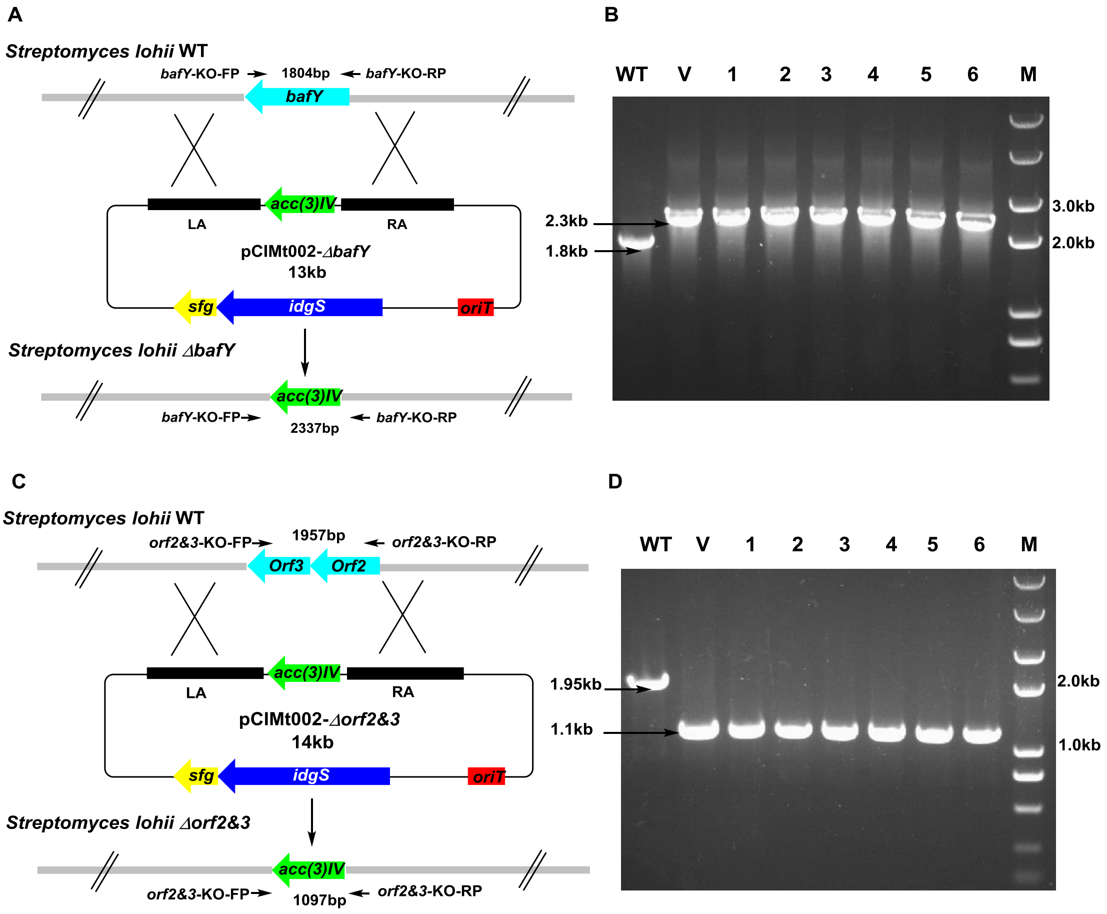
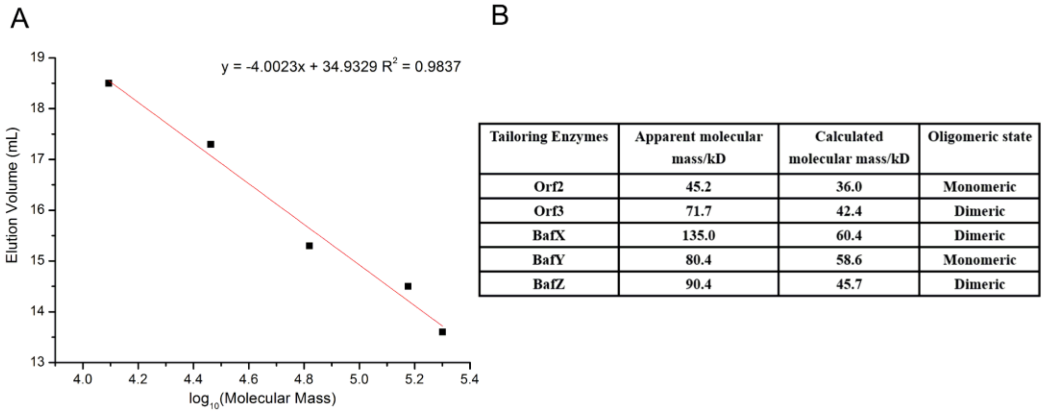


Figure S1. Inactivation of *bafY* and *orf2&orf3*. (A) Construction of the *bafY* inactivation mutant *S. lohii* Δ*bafY*. (B) PCR confirmation of the double-crossover mutants. WT: *S. lohii* wild type; M: DNA marker; V: pCIMt002-Δ*bafY*; 1-6: *bafY* inactivation mutants. (C) Construction of the *orf2* and *orf3* inactivation mutant *S. lohii* Δ*orf2&orf3*. (D) PCR confirmation of the double-crossover mutants. WT: *S. lohii* wild type; M: DNA marker; V: pCIMt002-Δ*orf2&orf3*; 1-6: *orf2 and orf3* inactivation mutants. *Note*： LA： left homologous arm; RA： right homologous arm.

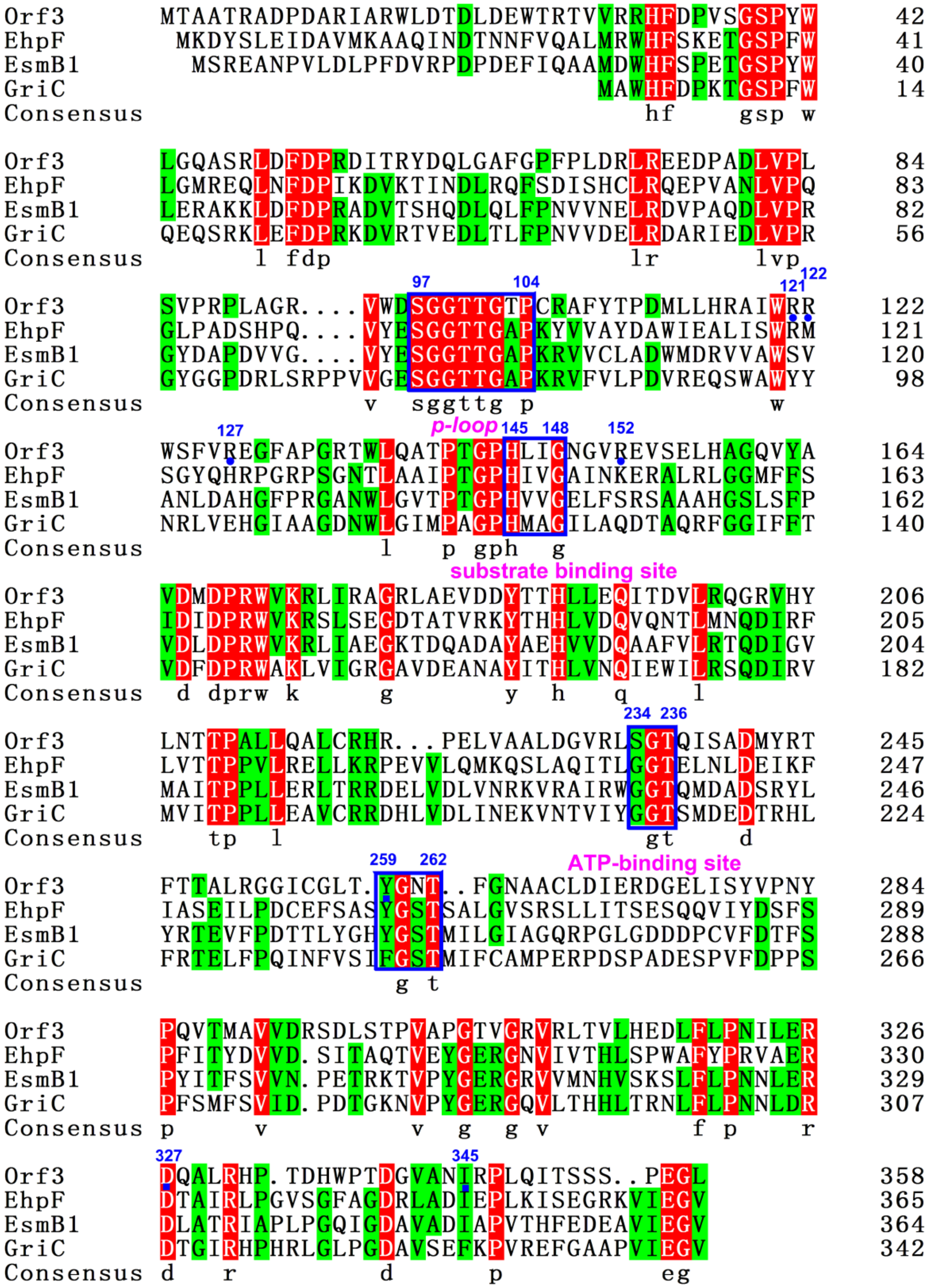
```
N-truncated Orf3                               MVRRHFDPVSGSPYWLGQAS   20
BafCI              MTAATRADPDDRIARWLDTDLDTWTRTVVRRHFDPVSGSPYWLGQAP   47
P1                 MTAATRADPDARIARWLETDLDEWTRTVVRRHFDPVSGSPYWLDQAS   47
P2                 MTAATRADPDARIARWLDTDLDVWTRTVVRRHFDPVSGSPYWLGQAP   47
Consensus                                       vrrhfdpvsgspywl qa

N-truncated Orf3   RLDFDPRDITRYDQLGAFGPFPLDRLREEDPADLVPLSVPRPLAGRV   67
BafCI              RLDFDPRDITRYDQLGAFGPFPLDRLRHEDPADLVPLSVPRPLAGRV   94
P1                 RLDFDPRDITRYDQLGAFGPFPLDRLREEDPADLVPLSVPRPLAGRV   94
P2                 RLDFDPRDITRYDQLGAFGPFPLDRLRHEDPADLVPLSVPRPLAGRV   94
Consensus          rldfdprditrydqlgafgpfpldrlr edpadlvplsvprplagrv

N-truncated Orf3   WDSGGTTGTPCRAFYTPDMLLHRAIWRRWSFVREGFAPGRTWLQATP  114
BafCI              WDSGGTTGTPCRAFYTPDMLLHRAVWRRWSFVREGFAPGRTWLQATP  141
P1                 WDSGGTTGTPCRAFYTPDMLLHRAIWRRWSFVREGFAPGRTWLQATP  141
P2                 WDSGGTTGTPCRAFYTPDMLLHRAVWRRWSFVREGFAPGRTWLQATP  141
Consensus          wdsggttgtpcrafytpdmllhra wrrwsfvregfapgrtwlqatp
```

Figure S2. The protein sequence alignment of the *N*-terminal 110-140 amino acids of the previously annotated Orf3 and some homologous proteins with sequence identity >90%. *Note:* BafCI (GenBank accession number: AGK25205.1), P1 (WP_028419150.1P2), and P2 (WP_028419150.1). The amino acids highlighted in red stand for 100% homology; the residues highlighted in green stand for 75% < homology <100%.



Tailoring Enzymes	Apparent molecular mass/kD	Calculated molecular mass/kD	Oligomeric state
Orf2	45.2	36.0	Monomeric
Orf3	71.7	42.4	Dimeric
BafX	135.0	60.4	Dimeric
BafY	80.4	58.6	Monomeric
BafZ	90.4	45.7	Dimeric

Figure S3. Gel filtration analysis for oligomeric state determination of all Baf tailoring enzymes. (A) The standard curve of gel filtration chromatography. (B) The calculation of the apparent molecular mass of each tailoring enzyme. Experimentally, purified BafX, BafY, BafZ, Orf2, and Orf3 were individually applied to size-exclusion chromatography using a Superdex200 10/300 column (GE Healthcare) in a buffer containing 20 mM HEPES, 150 mM NaCl, and 1 mM DTT, pH 7.5. The apparent molecular mass of each sample was calculated based on the calibration of the column using protein standards.



```
Orf3       MTAATRADPDARIARWLDTDLDEWTRTVVRRHFDPVSGSPYW     42
EhpF        MKDYSLEIDAVMKAAQINDTNNFVQALMRWHFSKETGSPFW     41
EsmB1        MSREANPVLDLPFDVRPDPDEFIQAAMDWHFSPETGSPYW     40
GriC                                       MAWHFDPKTGSPFW     14
Consensus                                  hf    gsp w

Orf3       LGQASRLDFDPRDITRYDQLGAFGPFPLDRLREEDPADLVPL     84
EhpF       LGMREQLNFDPIKDVKTINDLRQFSDISHCLRQEPVANLVPQ     83
EsmB1      LERAKKLDFDPRADVTSHQDLQLFPNVVNELRDVPAQDLVPR     82
GriC       QEQSRKLEFDPRKDVRTVEDLTLFPNVVDELRDARIEDLVPR     56
Consensus        l fdp                   lr      lvp

                           97     104            121 122
Orf3       SVPRPLAGR....VWDSGGTTGTPCRAFYTPDMLLHRAIWRR     122
EhpF       GLPADSHPQ....VYESGGTTGAPKYVVAYDAWIEALISWRM     121
EsmB1      GYDAPDVVG....VYESGGTTGAPKRVVCLADWMDRVVAWSV     120
GriC       GYGGPDRLSRPPVVGESGGTTGAPKRVFVLPDVREQSWAWYY     98
Consensus               v  sggttg p               w

               127        p-loop 145 148 152
Orf3       WSFVREGFAPGRTWLQATPTGPHLIGNGVREVSELHAGQVYA     164
EhpF       SGYQHRPGRPSGNTLAAIPTGPHIVGAINKERALRLGGMFFS     163
EsmB1      ANLDAHGFPRGANWLGVTPTGPHVVGELFSRSAAAHGSLSFP     162
GriC       NRLVEHGIAAGDNWLGIMPAGPHMAGILAQDTAQRFGGIFFT     140
Consensus                l   p gph  g

                             substrate binding site
Orf3       VDMDPRWVKRLIRAGRLAEVDDYTTHLLEQITDVLRQGRVHY     206
EhpF       IDIDPRWVKRSLSEGDTATVRKYTHHLVDQVQNTLMNQDIRF     205
EsmB1      VDLDPRWVKRLIAEGKTDQADAYAEHVVDQAAFVLRTQDIGV     204
GriC       VDFDPRWAKLVIGRGAVDEANAYITHLVNQIEWILRSQDIRV     182
Consensus   d dprw k     g       y  h   q    l

                                        234 236
Orf3       LNTTPALLQALCRHR...PELVAALDGVRLSGTQISADMYRT     245
EhpF       LVTTPPVLRELLKRPEVVLQMKQSLAQITLGGTELNLDEIKF     247
EsmB1      MAITPPLLERLTRRDELVDLVNRKVRAIRWGGTQMDADSRYL     246
GriC       MVITPPLLEAVCRRDHLVDLINEKVNTVIYGGTSMDEDTRHL     224
Consensus     tp  l                       gt    d

                        259 262      ATP-binding site
Orf3       FTTALRGGICGLT.YGNT..FGNAACLDIERDGELISYVPNY     284
EhpF       IASEILPDCEFSASYGSTSALGVSRSLLITSESQQVIYDSFS     289
EsmB1      YRTEVFPDTTLYGHYGSTMILGIAGQRPGLGDDDPCVFDTFS     288
GriC       FRTELFPQINFVSIFGSTMIFCAMPERPDSPADESPVFDPPS     266
Consensus                 g t

Orf3       PQVTMAVVDRSDLSTPVAPGTVGRVRLTVLHEDLFLPNILER     326
EhpF       PFITYDVVD.SITAQTVEYGERGNVIVTHLSPWAFYPRVAER     330
EsmB1      PYITFSVVN.PETRKTVPYGERGRVVMNHVSKSLFLPNNLER     329
GriC       PFSMFSVID.PDTGKNVPYGERGQVLTHHLTRNLFLPNNLDR     307
Consensus  p     v         v  g  g v         f p    r

           327                345
Orf3       DQALRHP.TDHWPTDGVANIRPLQITSSS..PEGL            358
EhpF       DTAIRLPGVSGFAGDRLADIEPLKISEGRKVIEGV            365
EsmB1      DLATRIAPLPGQIGDAVADIAPVTHFEDEAVIEGV            364
GriC       DTGIRHPHRLGLPGDAVSEFKPVREFGAAPVIEGV            342
Consensus  d   r         d      p          eg
```

Figure S4. Protein sequence alignment of Orf3 with EphF, EsmB1 and GriC. The amino acids highlighted in red stand for 100% homology; the residues highlighted in green stand for 75% < homology <100%; ■ denotes the key residues of Orf3 involved in ATP-binding; ● denotes the arginine residues of Orf3 related to substrate-binding; *p*-loop, substrate binding site; the ATP binding sites are boxed.

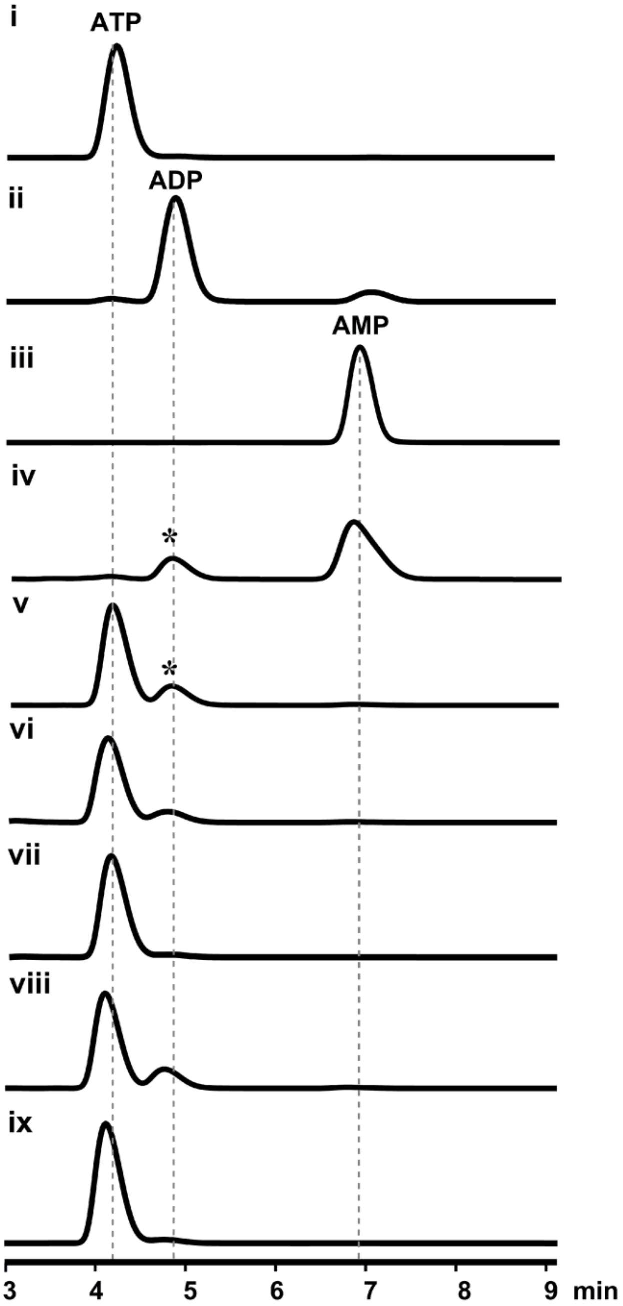


Figure S5. HPLC analysis (250 nm) of the ATP transformation by Orf3 or BafY. Trace (i): ATP standard. Trace (ii): ADP standard. Trace (iii): AMP standard. Trace (iv): 40 μM Orf3 + 10 mM $MgCl_2$ + 1 mM disodium fumarate + 1 mM ATP. Trace (v): the negative control for trace iii, in which Orf3 was boiling inactivated. Trace (vi): 40 μM BafY + 10 mM $MgCl_2$ + 200 μM **2** + 1 mM ATP. Trace (vii): the negative control for trace vi, in which BafY was boiling inactivated. Trace (viii): 40 μM BafY + 10 mM $MgCl_2$ and 1mM ATP. Trace (ix): the negative control for trace viii, in which BafY was boiling inactivated. *Note*: all reactions were carried out at 28°C for 12 h. The HPLC analysis was performed on a Waters Atlantis® T3 column (2.1 × 150 mm) with isocratic 100 % 50 mM KH_2PO_4 buffer (pH= 6.0) at a flow rate of 0.25 ml/min for 10 min. The asterisk stands for the artifact (coincidently with the same retention time of ADP) in Orf3 preparation.

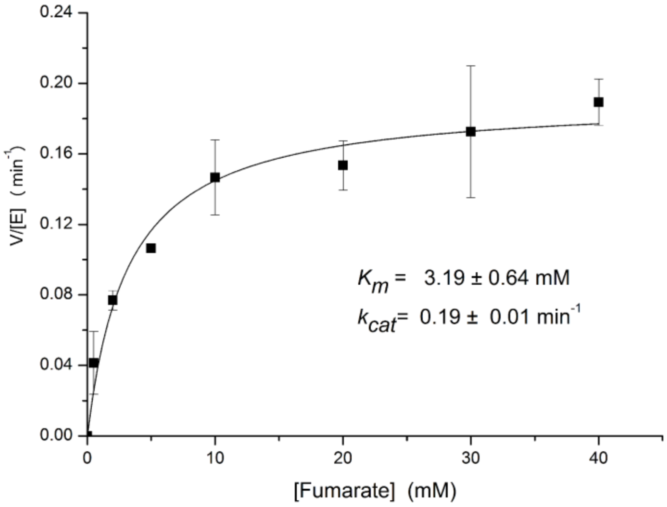


Figure S6. The kinetic curve of Orf3 using fumarate as substrate.

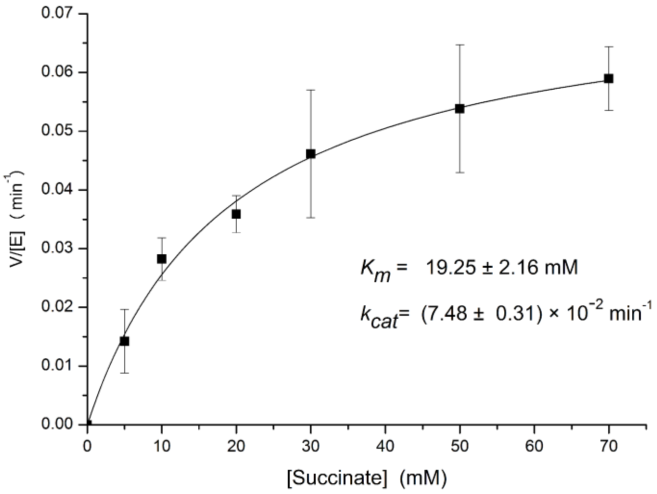


Figure S7. The kinetic curve of Orf3 using succinate as substrate.

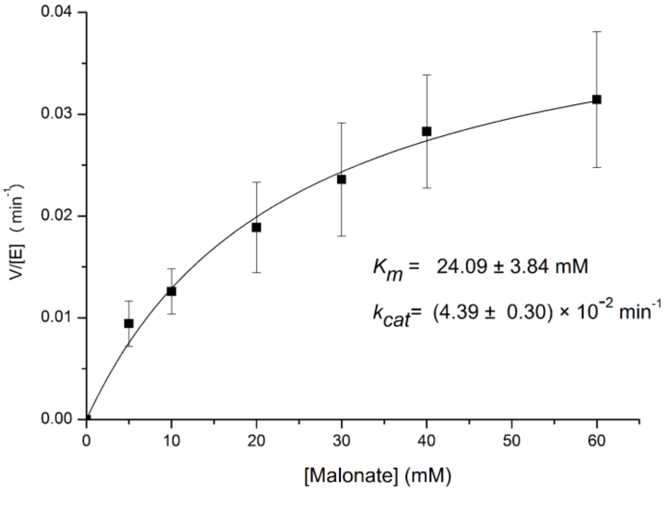


Figure S8. The kinetic curve of Orf3 using malonate as substrate.

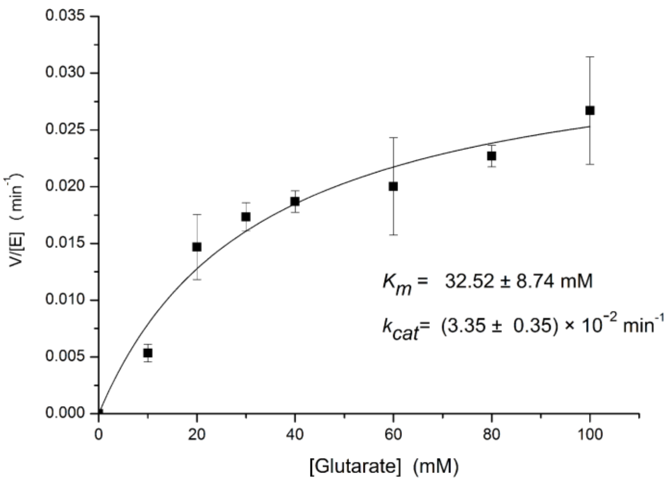


Figure S9. The kinetic curve of Orf3 using glutarate as substrate.

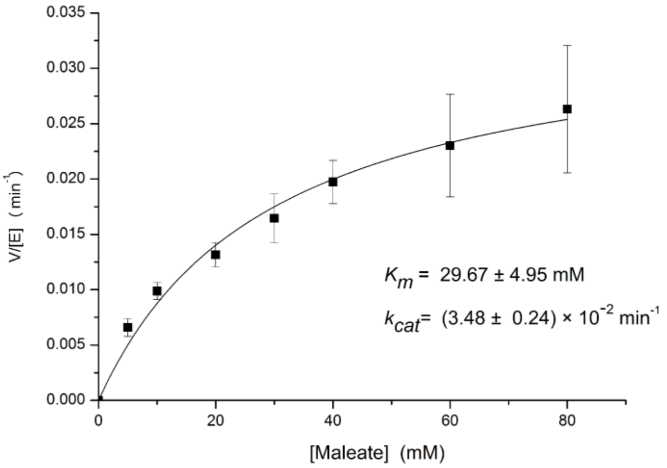


Figure S10. The kinetic curve of Orf3 using maleate as substrate.

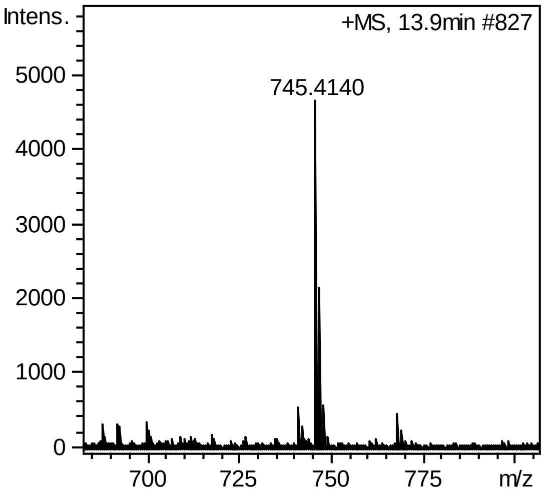
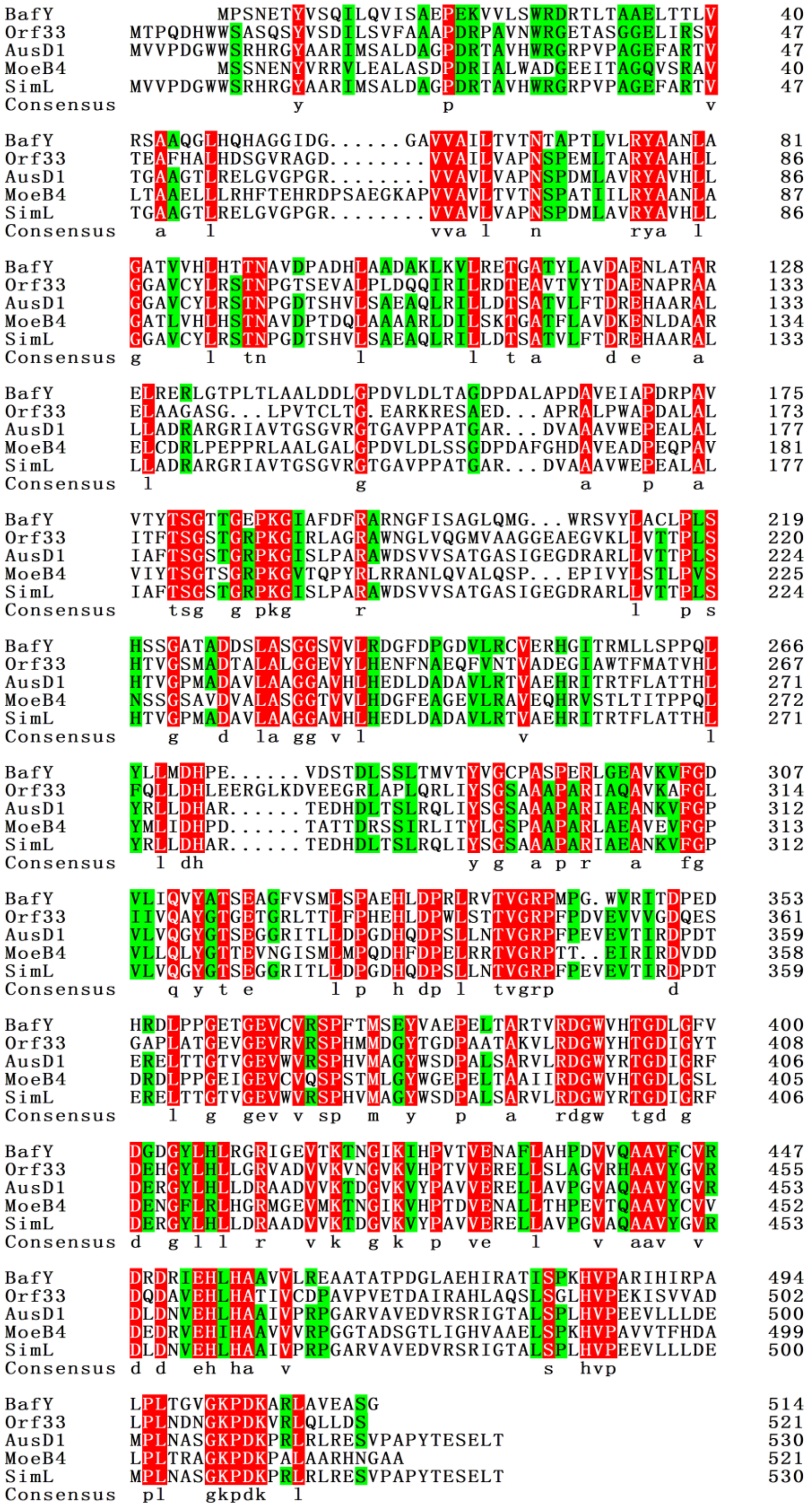


Figure S11. HRMS of Compound 4 ($[M+Na]^+$: *obs.* 745.4140 *calc.* 745.4139)



```
BafY                 MPSNETYVSQILQVISAEPEKVVLSWRDRTLTAAELTTLV      40
Orf33         MTPQDHWWSASQSYVSDILSVFAAAPDRPAVNWRGETASGGELIRSV      47
AusD1         MVVPDGWWSRHRGYAARIMSALDAGPDRTAVHWRGRPVPAGEFARTV      47
MoeB4                MSSNENYVRRVLEALASDPDRIALWADGEEITAGQVSRAV      40
SimL          MVVPDGWWSRHRGYAARIMSALDAGPDRTAVHWRGRPVPAGEFARTV      47
Consensus                  y          p                     v

BafY          RSAAQGLHQHAGGIDG......GAVVAILTVTNTAPTLVLRYAANLA      81
Orf33         TEAFHALHDSGVRAGD........VVAILVAPNSPEMLTARYAAHLL      86
AusD1         TGAAGTLRELGVGPGR........VVAVLVAPNSPDMLAVRYAVHLL      86
MoeB4         LTAAELLLRHFTEHRDPSAEGKAPVVAVLTVTNSPATIILRYAANLA      87
SimL          TGAAGTLRELGVGPGR........VVAVLVAPNSPDMLAVRYAVHLL      86
Consensus       a   l                 vva l   n       rya  l

BafY          GATVVHLHTTNAVDPADHLAADAKLKVLRETGATYLAVDAENLATAR     128
Orf33         GGAVCYLRSTNPGTSEVALPLDQQIRILRDTEAVTVYTDAENAPRAA     133
AusD1         GGAVCYLRSTNPGDTSHVLSAEAQLRILLDTSATVLFTDREHAARAL     133
MoeB4         GATLVHLHSTNAVDPTDQLAAAARLDILSKTGATFLAVDKENLDAAR     134
SimL          GGAVCYLRSTNPGDTSHVLSAEAQLRILLDTSATVLFTDREHAARAL     133
Consensus     g     l  tn       l        l  t a     d e    a

BafY          ELRERLGTPLTLAALDDLGPDVLDLTAGDPDALAPDAVEIAPDRPAV     175
Orf33         ELAAGASG...LPVTCLTG.EARKRESAED...APRALPWAPDALAL     173
AusD1         LLADRARGRIAVTGSGVRGTGAVPPATGAR...DVAAAVWEPEALAL     177
MoeB4         ELCDRLPEPPRLAALGALGPDVLDLSSGDPDAFGHDAVEADPEQPAV     181
SimL          LLADRARGRIAVTGSGVRGTGAVPPATGAR...DVAAAVWEPEALAL     177
Consensus      l                g                 a    p   a

BafY          VTYTSGTTGEPKGIAFDFRARNGFISAGLQMG...WRSVYLACLPLS     219
Orf33         ITFTSGSTGRPKGIRLAGRAWNGLVQGMVAAGGEAEGVKLLVTTPLS     220
AusD1         IAFTSGSTGRPKGISLPARAWDSVVSATGASIGEGDRARLLVTTPLS     224
MoeB4         VIYTSGTSGRPKGVTQPYRLRRANLQVALQSP...EPIVYLSTLPVS     225
SimL          IAFTSGSTGRPKGISLPARAWDSVVSATGASIGEGDRARLLVTTPLS     224
Consensus        tsg  g pkg     r                     l   p s

BafY          HSSGATADDSLASGGSVVLRDGFDPGDVLRCVERHGITRMLLSPPQL     266
Orf33         HTVGSMADTALALGGEVYLHENFNAEQFVNTVADEGIAWTFMATVHL     267
AusD1         HTVGPMADAVLAAGGAVHLHEDLDADAVLRTVAEHRITRTFLATTHL     271
MoeB4         NSSGSAVDVALASGGTVVLHDGFEAGEVLRAVEQHRVSTLTITPPQL     272
SimL          HTVGPMADAVLAAGGAVHLHEDLDADAVLRTVAEHRITRTFLATTHL     271
Consensus        g  d la gg v l              v              l

BafY          YLLMDHPE......VDSTDLSSLTMVTYVGCPASPERLGEAVKVFGD     307
Orf33         FQLLDHLEERGLKDVEEGRLAPLQRLIYSGSAAAPARIAQAVKAFGL     314
AusD1         YRLLDHAR......TEDHDLTSLRQLIYSGSAAAPARIAEANKVFGP     312
MoeB4         YMLIDHPD......TATTDRSSIRLITYLGSPAAPARLAEAVEVFGP     313
SimL          YRLLDHAR......TEDHDLTSLRQLIYSGSAAAPARIAEANKVFGP     312
Consensus       l dh                     y g  a p r   a   fg

BafY          VLIQVYATSEAGFVSMLSPAEHLDPRLRVTVGRPMPG.WVRITDPED     353
Orf33         IIVQAYGTGETGRLTTLFPHEHLDPWLSTTVGRPFPDVEVVVGDQES     361
AusD1         VLVQGYGTSEGGRITLLDPGDHQDPSLLNTVGRPFPEVEVTIRDPDT     359
MoeB4         VLLQLYGTTEVNGISMLMPQDHFDPELRRTVGRPTT..EIRIRDVDD     358
SimL          VLVQGYGTSEGGRITLLDPGDHQDPSLLNTVGRPFPEVEVTIRDPDT     359
Consensus        q y t e      l p  h dp l  tvgrp         d

BafY          HRDLPPGETGEVCVRSPFTMSEYVAEPELTARTVRDGWVHTGDLGFV     400
Orf33         GAPLATGEVGEVRVRSPHMMDGYTGDPAATAKVLRDGWYHTGDIGYT     408
AusD1         ERELTTGTVGEVWVRSPHVMAGYWSDPALSARVLRDGWYRTGDIGRF     406
MoeB4         DRDLPPGEIGEVCVQSPSTMLGYWGEPELTAAIIRDGWVHTGDLGSL     405
SimL          ERELTTGTVGEVWVRSPHVMAGYWSDPALSARVLRDGWYRTGDIGRF     406
Consensus        l  g  gev v sp  m  y   p   a   rdgw  tgd g

BafY          DGDGYLHLRGRIGEVTKTNGIKIHPVTVENAFLAHPDVVQAAVFCVR     447
Orf33         DEHGYLHLLGRVADVVKVNGVKVHPTVVERELLSLAGVRHAAVYGVR     455
AusD1         DERGYLHLLDRAADVVKTDGVKVYPAVVERELLAVPGVAQAAVYGVR     453
MoeB4         DENGFLRLHGRMGEVMKTNGIKVHPTDVENALLTHPEVTQAAVYCVV     452
SimL          DERGYLHLLDRAADVVKTDGVKVYPAVVERELLAVPGVAQAAVYGVR     453
Consensus     d  g l l  r   v k  g k  p  ve   l    v  aav  v

BafY          DRDRIEHLHAAVVLREAATATPDGLAEHIRATISPKHVPARIHIRPA     494
Orf33         DQDAVEHLHATIVCDPAVPVETDAIRAHLAQSLSGLHVPEKISVVAD     502
AusD1         DLDNVEHLHAAIVPRPGARVAVEDVRSRIGTALSPLHVPEEVLLLDE     500
MoeB4         DEDRVEHIHAAVVVRPGGTADSGTLIGHVAAELSPKHVPAVVTFHDA     499
SimL          DLDNVEHLHAAIVPRPGARVAVEDVRSRIGTALSPLHVPEEVLLLDE     500
Consensus     d d  eh ha  v                     s hvp

BafY          LPLTGVGKPDKARLAVEASG                                514
Orf33         LPLNDNGKPDKVRLQLLDS                                 521
AusD1         MPLNASGKPDKPRLRLRESVPAPYTESELT                      530
MoeB4         LPLTRAGKPDKPALAARHNGAA                              521
SimL          MPLNASGKPDKPRLRLRESVPAPYTESELT                      530
Consensus      pl   gkpdk  l
```

Figure S12. Protein sequence alignment of BafY with Orf33, AusD1, MoeB4 and SimL. The amino acids highlighted in red stand for 100% homology; the residues highlighted in green stand for 75% < homology <100%.

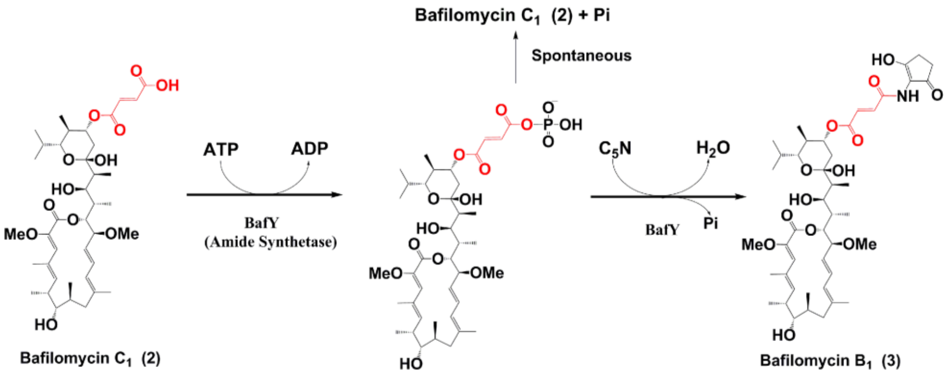


Figure S13. The proposed catalytic mechanism of BafY during the transformation from bafilomycin C_1 (**2**) to B_1 (**3**).

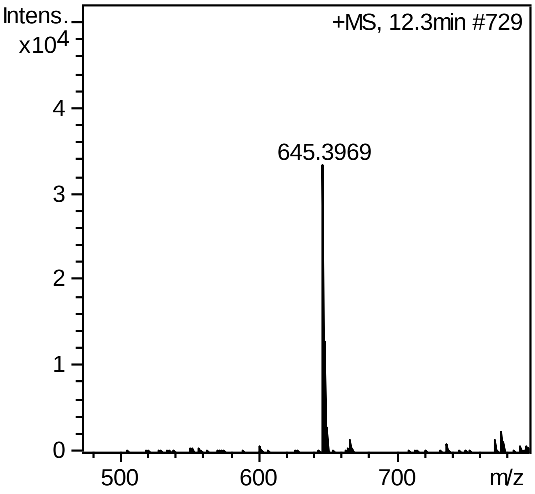


Figure S14. HRMS of bafilomycin A_1 (1) ($[M+Na]^+$: *obs.* 645.3969 *calc.* 645.3973)

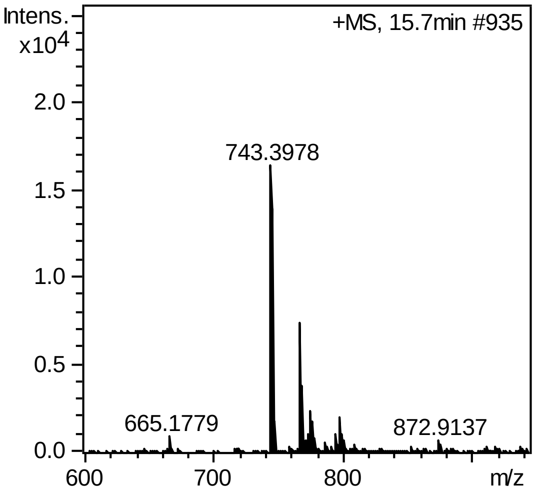


Figure S15. HRMS of bafilomycin C_1 (2) ($[M+Na]^+$: *obs.* 743.3977 *calc.* 743.3978)

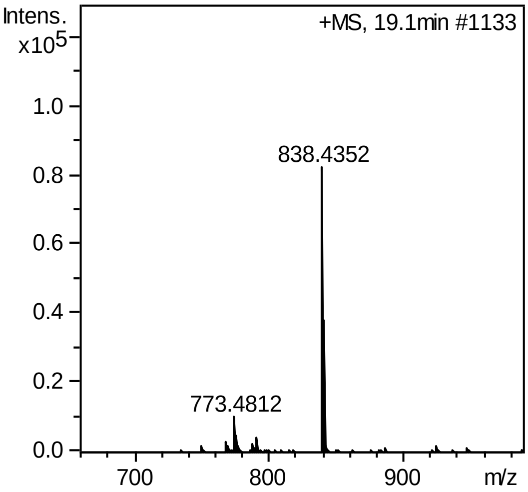


Figure S16. HRMS of bafilomycin B$_1$ (3) ([M+Na]$^+$: *obs.* 838.4352 *calc.* 838.4352)

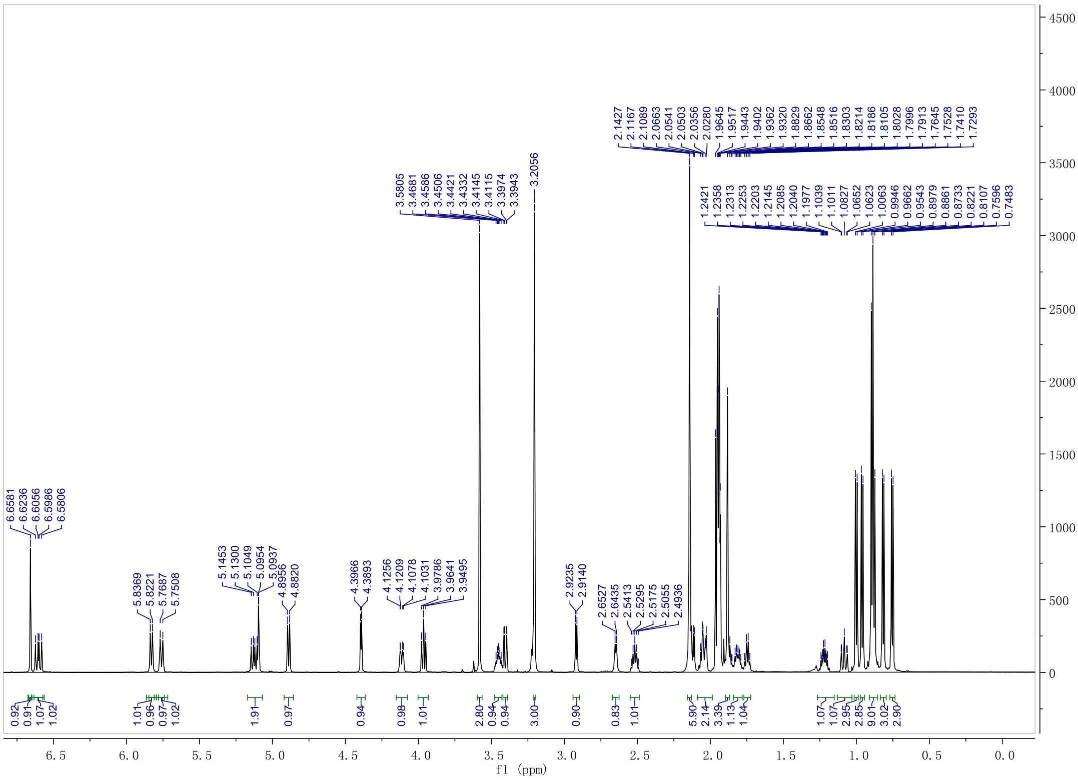


Figure S17. ^{1}H NMR spectrum of bafilomycin A$_1$ (1) in CD$_3$CN.

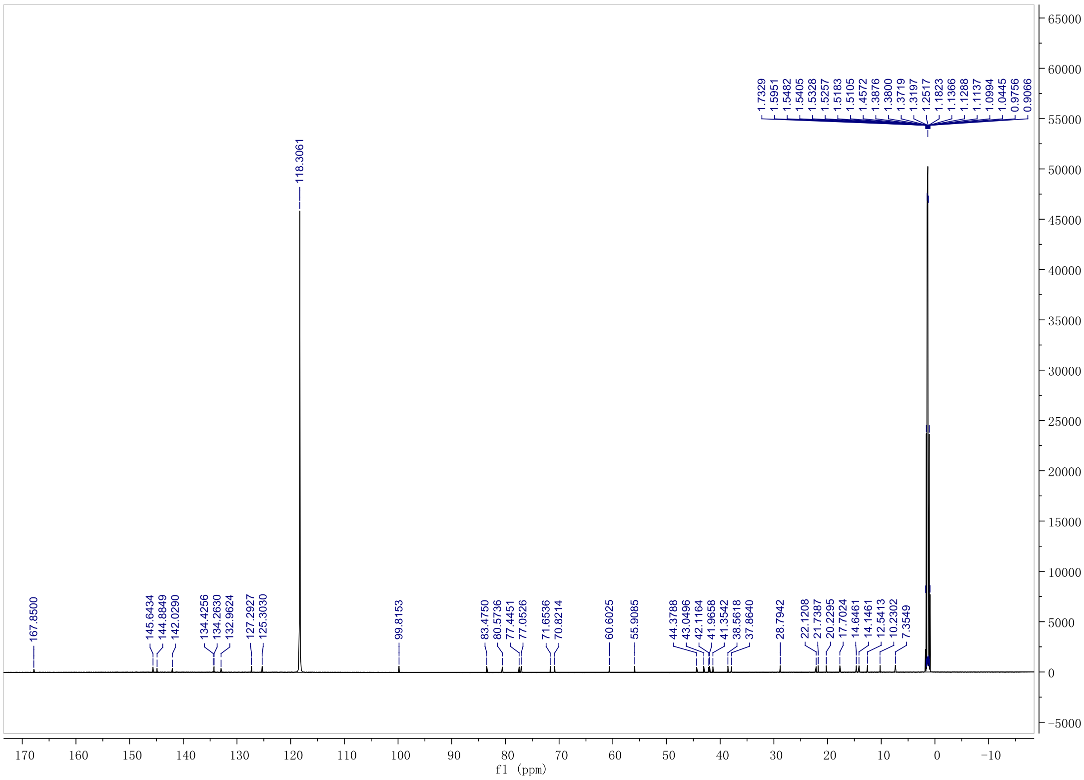


Figure S18. ^{13}C NMR spectrum of bafilomycin A$_1$ (1) in CD$_3$CN.

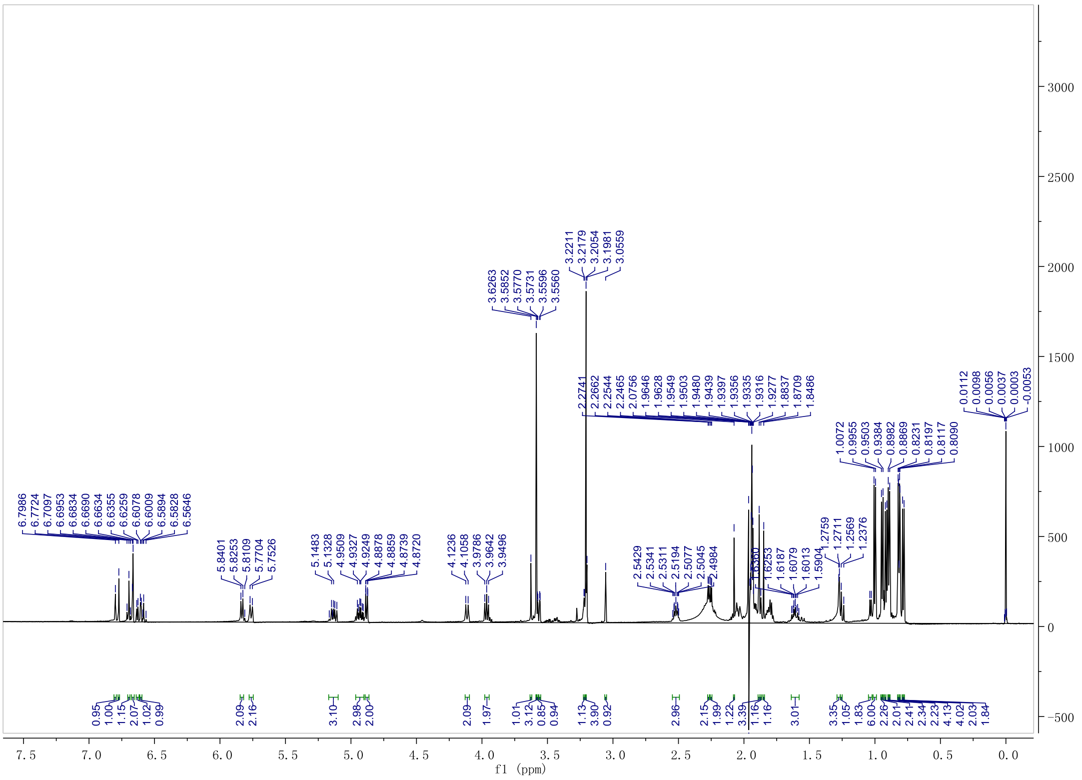


Figure S19. ^{1}H NMR spectrum of bafilomycin C$_1$ (2) in CD$_3$CN.

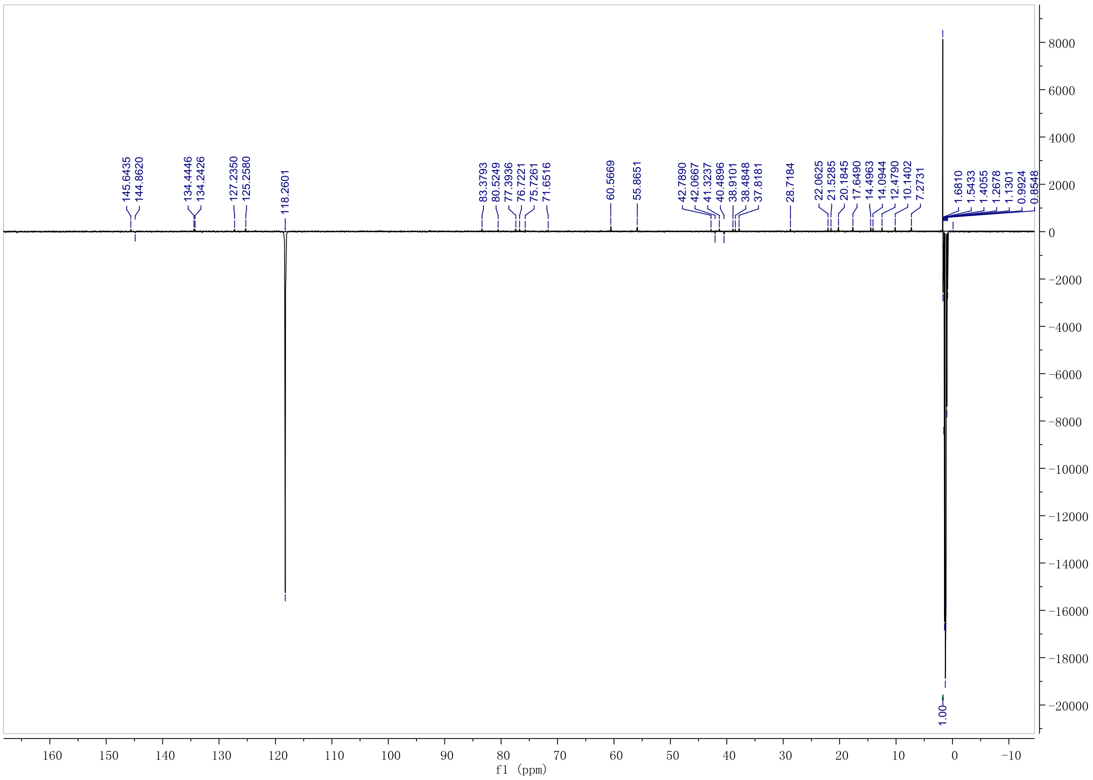


Figure S20. ^{13}C-DEPTQ spectrum of bafilomycin C$_1$ (**2**) in CD$_3$CN.

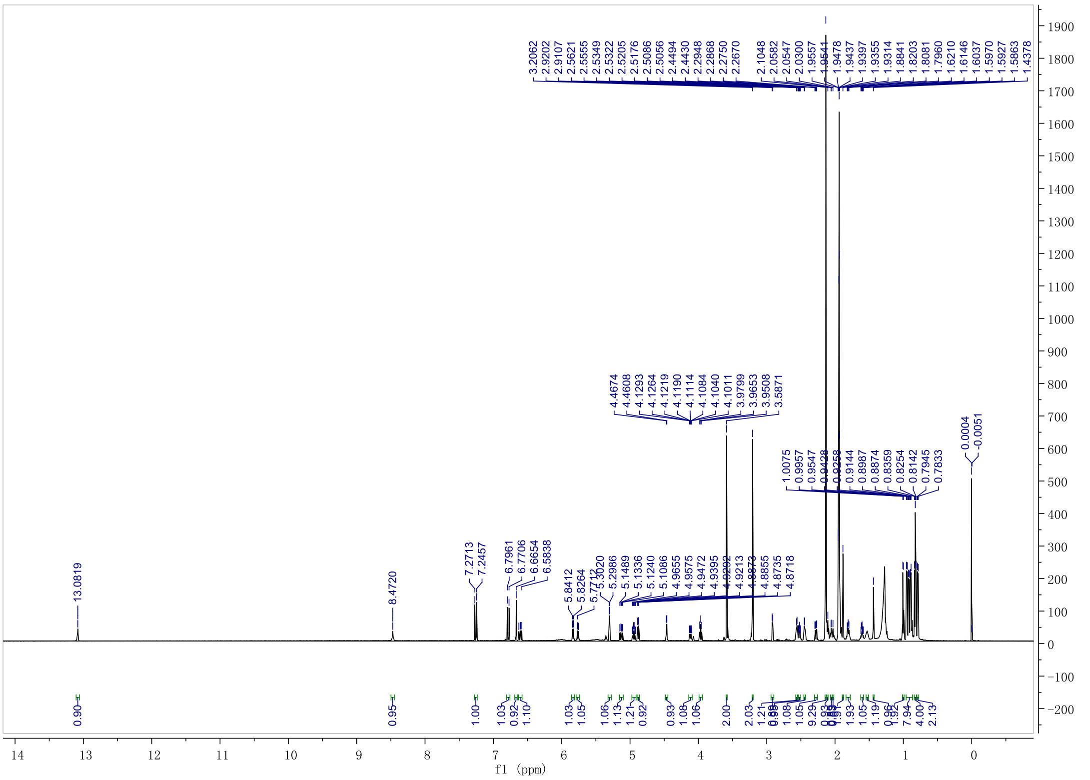


Figure S21. ^{1}H NMR spectrum of bafilomycin B$_1$ (**3**) in CD$_3$CN.

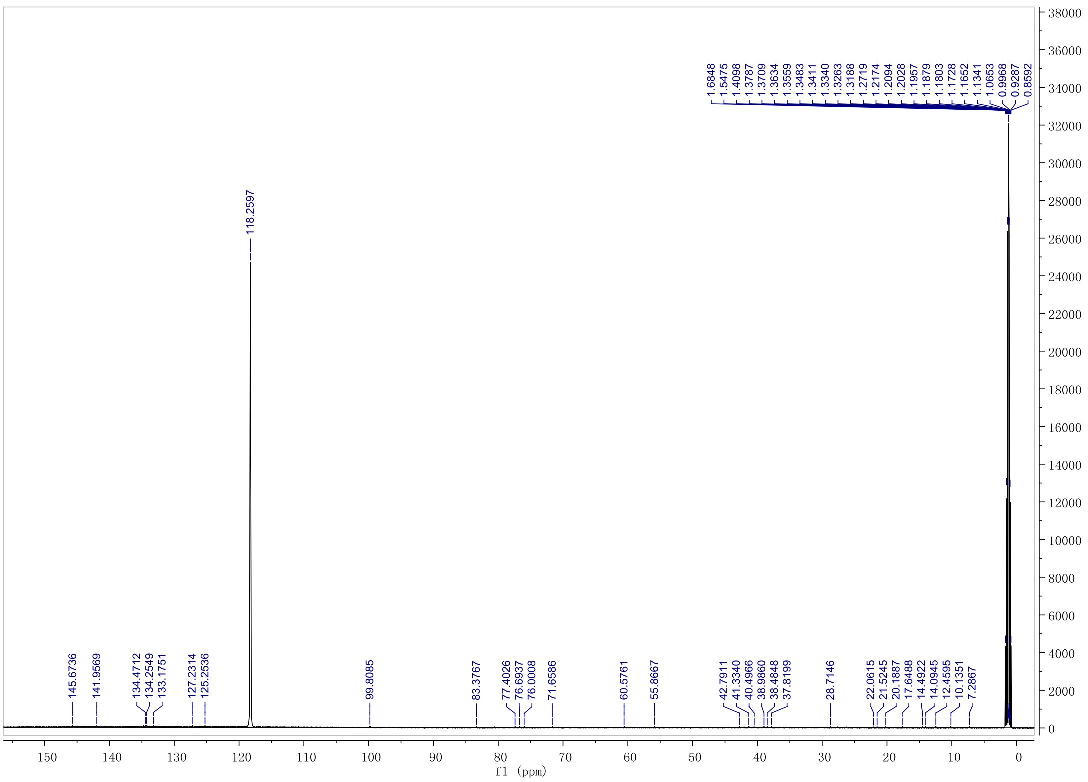


Figure S22. ^{13}C NMR spectrum of bafilomycin B$_1$ (3) in CD$_3$CN.

Table S1. The conversion percentages of 1 to 2 (fumarate as substrate) or 4 (succinate as substrate) by Orf2 (10 μM) and Orf3 (10 μM) at 28 °C for 4 h.

Compound	**Conversion Percentage**
2	83.1 ± 3.2%
4	2.2 ± 0.1%

Note: The conversion percentages were calculated based on substrate consumption.

Table S2. The primers for construction of knock-out vectors and PCR confirmation of *S. lohii* mutants.

Primers	**Sequence (5'-3')**
Δ*bafY*-left arm-FP	AAAAGCTTCCATGGGCACGCCCTAGGTCCCGCCAGATTCCGTCCA
Δ*bafY*-left arm-RP	CCAAAATCCCTTAACGTGAGCCTAGGCGGCGGAGATGAAGCCGT
Δ*bafY*-right arm-FP	CTCGCCAGTCGATTGGCTGACAATTGGGATCACCGACCCGGAGGA
Δ*bafY*-right arm-RP	CCAAGCTTGCTAGCAGATGTCAATTGTCATGCCGTCTCCTGTGCT
Δ*orf2&3*-left arm-FP	AAGAGCTTTTATAAAAGCTTCCATGGCGTAAGGGATGCTCCCGCA
Δ*orf2&3*-left arm-RP	AACGTGAGCCTAGGGCGTGCCCATGGGCCGAACAGGTACCCCAGAC
Δ*orf2&3*-right arm-FP	TTGGCTGACAATTGACATCTGCTAGCGTCGACAGGAGCGACCTCTC
Δ*orf2&3*-right arm-RP	GTGGATCCGCACCCAAGCTTGCTAGCGCGTACGCCTTCCAGCTC
Δ*bafY*-KO-FP	CGTGCCTCGTCCCACAGTT
Δ*bafY*-KO-RP	ATCACTTCGATGGCGCGGC
Δ*orf2&3*-KO-FP	GGAGGAAGAGGTCCTCGTG
Δ*orf2&3*-KO-RP	ACTGCCCCTCCATTCACA

Table S3. The primers for construction of protein expression vectors.

Primers	**Sequence (5'-3')**
BafX-NdeI-FP	GCCGGAATTCCGTGACCCTCTCCGTGGCGT
BafX-HindIII-RP	ATGCAAGCTTGGCAGACCTACGGGAGGAAATCG
BafY-EcoRI-FP	GCCGGAATTCTATGCCGTCGAACGAAACGT
BafY-HindIII-RP	ATGCAAGCTTCCACAGTTCGCGGTAGGTG
Orf2-NdeI-FP	GGAATTCCATATGAGTCTGGGGTACCTG
Orf2-EcoRI-RP	GGAATTCTCACGGCAGGGCCGGGGC
Orf3-NdeI-FP	CCTGGT GCCGCGCGGCAGCCATATGACCGCCGCCACCCGCGC
Orf3-NdeI-RP	GTCCACCAGTCATGCTAGCCATATGTCAGTAGAGGCCCTCGGGCG

Note: The underline litters indicate the restriction sites.