Recognition Sequences and Substrate Evolution in Cyanobactin Biosynthesis

Debosmita Sardar, Elizabeth Pierce, John A. McIntosh and Eric W. Schmidt

Department of Medicinal Chemistry, University of Utah, Salt Lake City, Utah 84112, USA

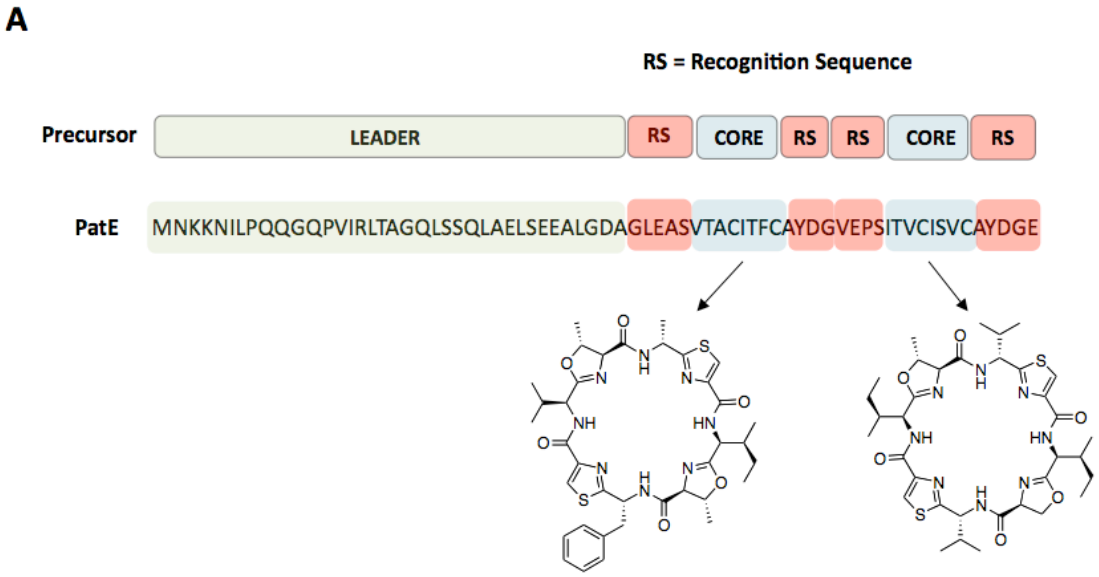
This file contains:

Supporting Figures S1 to S29

Supporting Tables S1 to S8

Supporting Information

Figure S1. Biosynthetic pathway of cyanobactins, with the specific example of patellamide biosynthesis. (A) The precursor peptide (PatE) is ribosomally translated. It contains the core peptide sequence (blue) that ends up in the final natural product. Flanking the cores are recognition sequences (RS, red) that direct posttranslational enzymes. The RS's are often found on a leader peptide (green). Ultimately the leader is proteolytically cleaved, leaving the core that matures into the final natural product. (B) Sequence of precursor peptide PatE is shown with an expansion of residues GLEASVTACITFCAYDG. At the first step, thiazolines/oxazolines (yellow circles) are installed in the cores by a heterocyclase (PatD). Then, an N-terminal protease (PatA) and a C-terminal protease (PatG) carry out proteolysis, cleaving the core peptide from the rest of the precursor. PatG also carries out macrocyclization in tandem with C-terminal proteolysis, leading to the cyclic peptide natural product patellamide C.



B

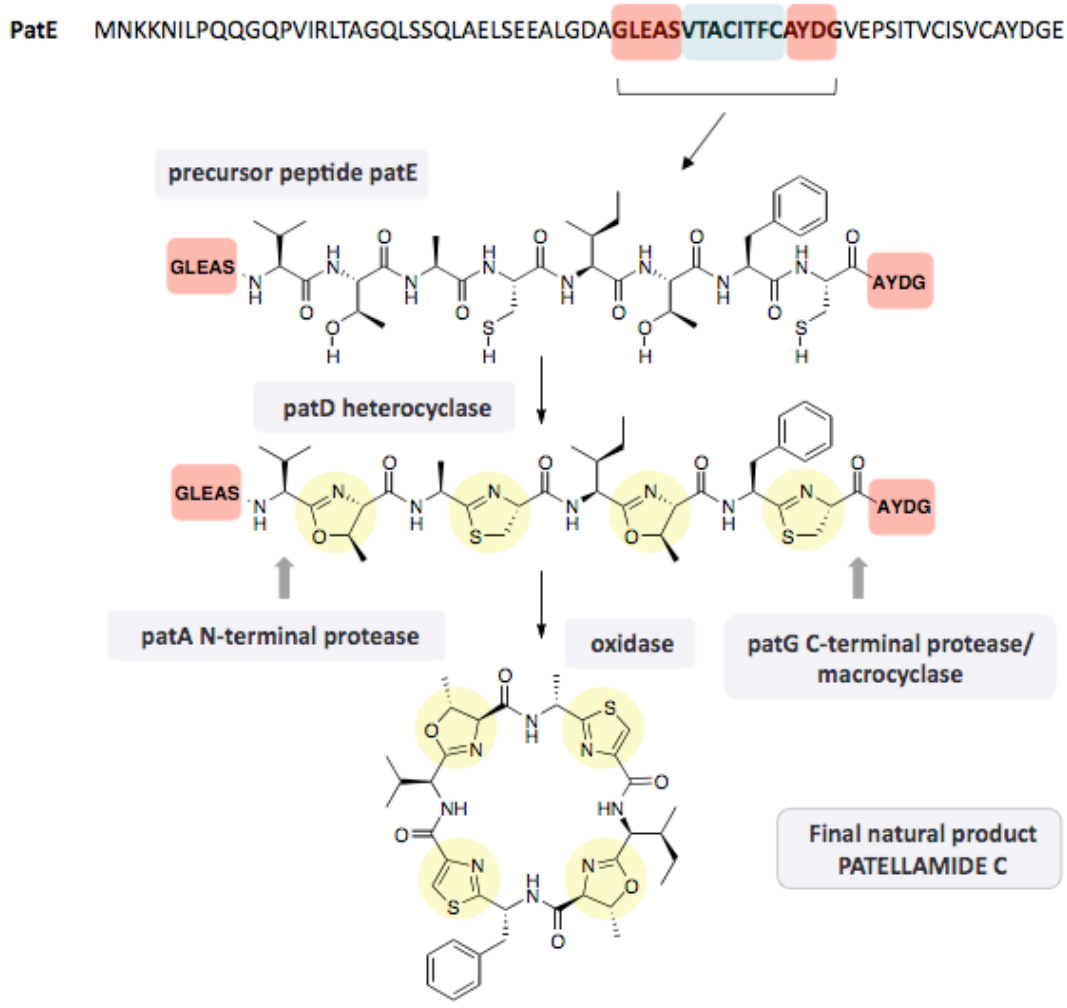
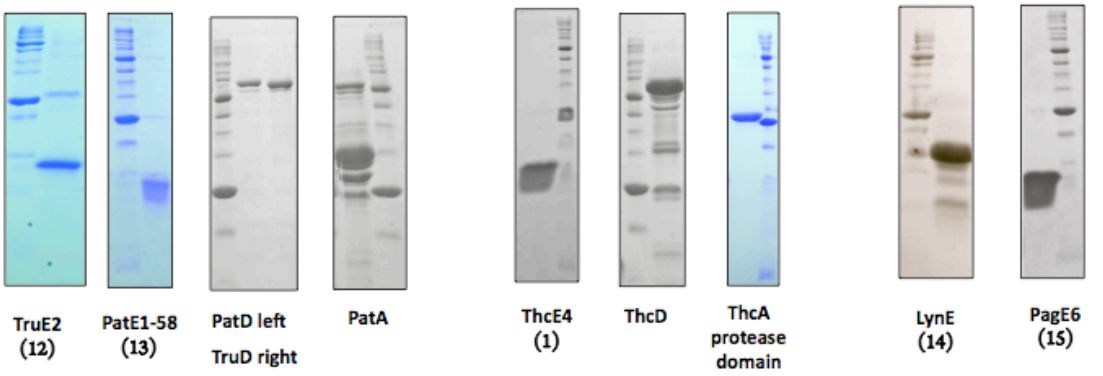

PatE MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALGDAGLEASVTACITFCAYDGVEPSITVCISVCAYDGE
precursor peptide patE
GLEAS
AYDG
patD heterocyclase
GLEAS
AYDG
patA N-terminal protease
oxidase
patG C-terminal protease/ macrocyclase
Final natural product PATELLAMIDE C

Figure S2. SDS-PAGE of purified proteins from *pat* (PatE1-58, PatA and PatD), *tru* (TruE2, TruD), *thc* (ThcE4, ThcD, ThcA protease domain), *lyn* (LynE) and *pag* (PagE6) pathways. SDS-PAGE of hybrid precursor peptides is described later. The relationships between the previously characterized *pat* and *tru* and the newly characterized *thc* enzymes is given. A schematic representation of this is also shown wherein white to black represents increasing identity from 0-100% on an increasing gray-scale.



% identity	ThcD	ThcA
TruD	58	-
PatD	57	-
PatA	-	60

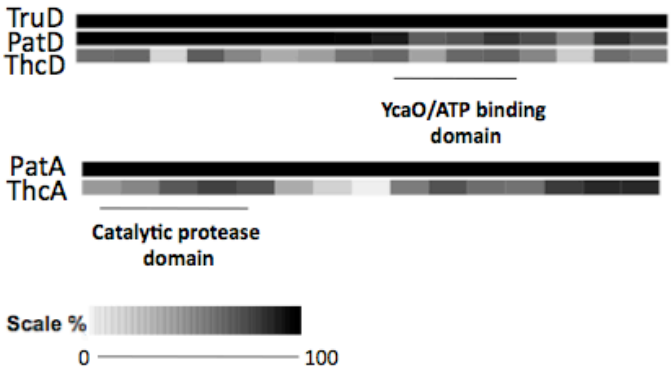
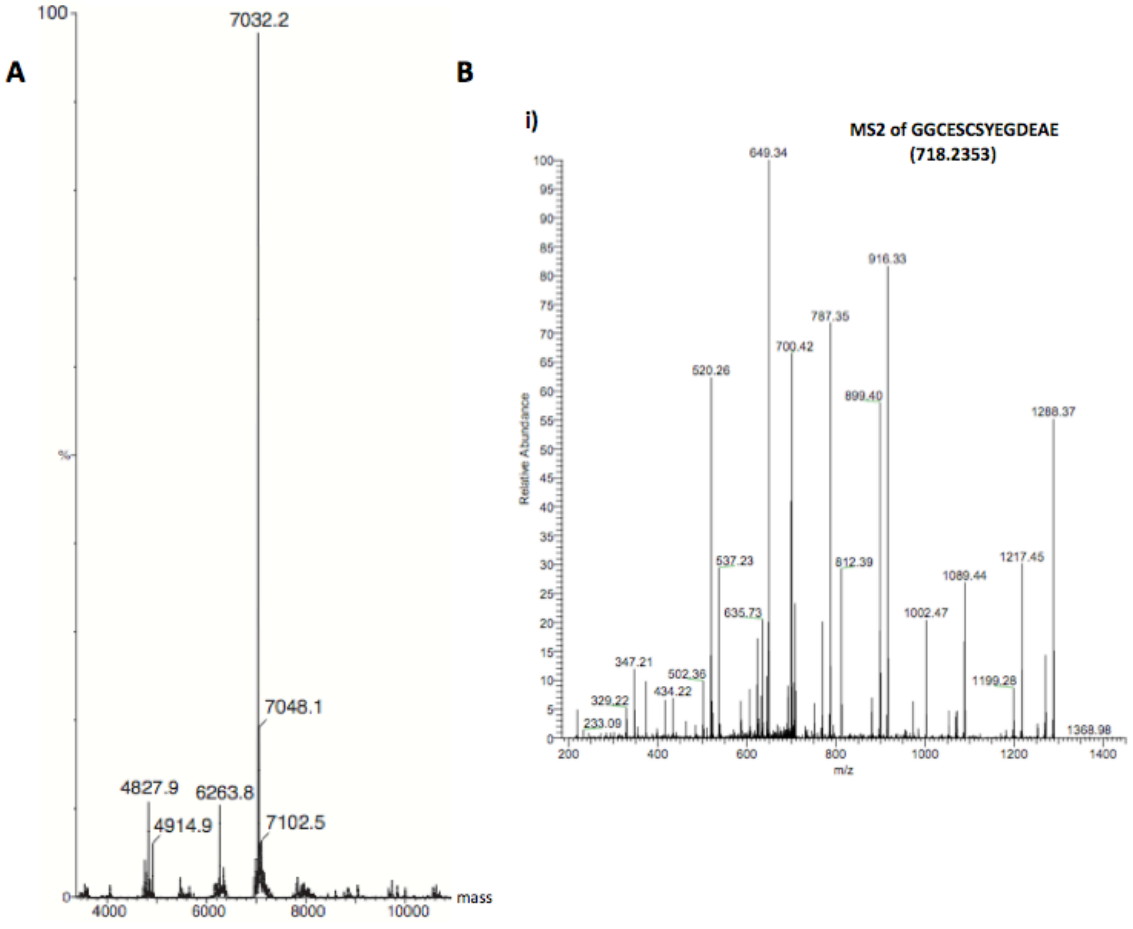
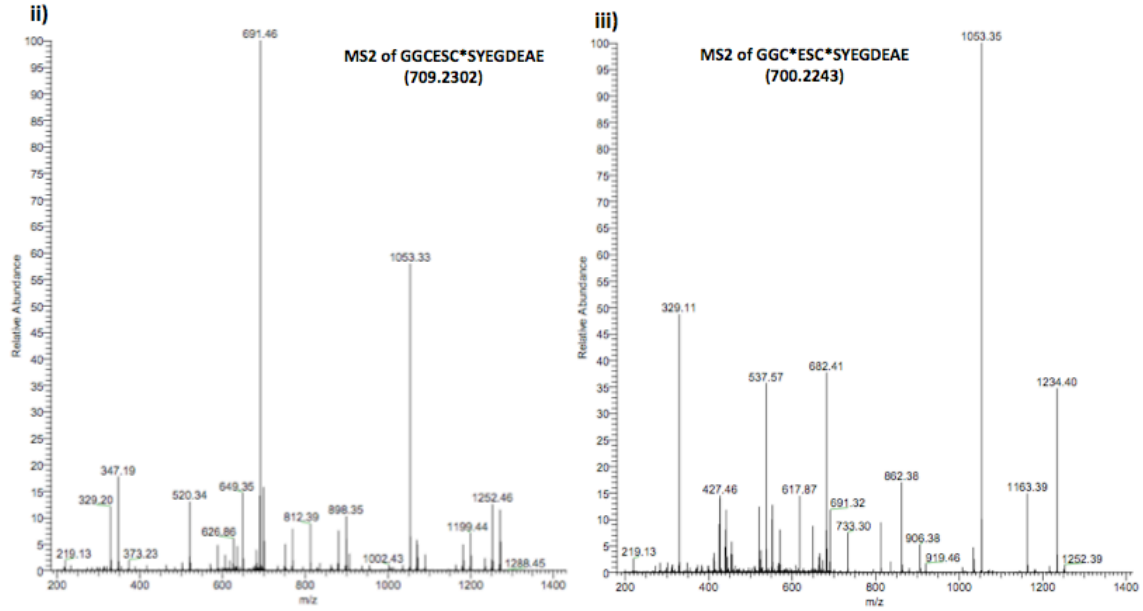


Figure S3. Functional analysis of *thc* enzymes showing MS data of the ThcE4/ThcD reaction. Each Cys modification to thiazoline is shown by C*. (A) ESI-MS of ThcE4 after modification by ThcD showing loss of $4H_2O$ $[M-H]^-$, indicated by the observed 7032.2 Da peak. (B) MS2 of the following $[M+2H]^{2+}$ parent ions resulting from fragments GGCESCSYEGDEAE: i) 718.2353 representing unmodified peptide GGCESCSYEGDEAE, ii) 709.2302 representing GGCESC*SYEGDEAE with one heterocycle modification, iii) 700.2243 representing GGC*ESC*SYEGDEAE with two heterocycle modifications. (C) MS2 of the following parent ions resulting from fragment ASSCDCSLY: i) 465.6704 representing ASSCDC*SLY with one heterocycle modification, ii) 456.6650 representing ASSC*DC*SLY with two heterocycle modifications. Localizations of all of the above modifications were obtained from analysis of the fragmentation patterns as given in Table S1.




ii)
MS2 of GGCESC*SYEGDEAE
(709.2302)
691.46
1053.33
347.19
329.20
520.34
649.35
626.86
812.39
898.35
1252.46
1199.44
219.13
373.23
1002.43
1288.45
Relative Abundance
m/z
iii)
MS2 of GGC*ESC*SYEGDEAE
(700.2243)
1053.35
329.11
537.57
682.41
1234.40
427.46
617.87
862.38
691.32
733.30
1163.39
906.38
919.46
219.13
1252.39
Relative Abundance
m/z

C

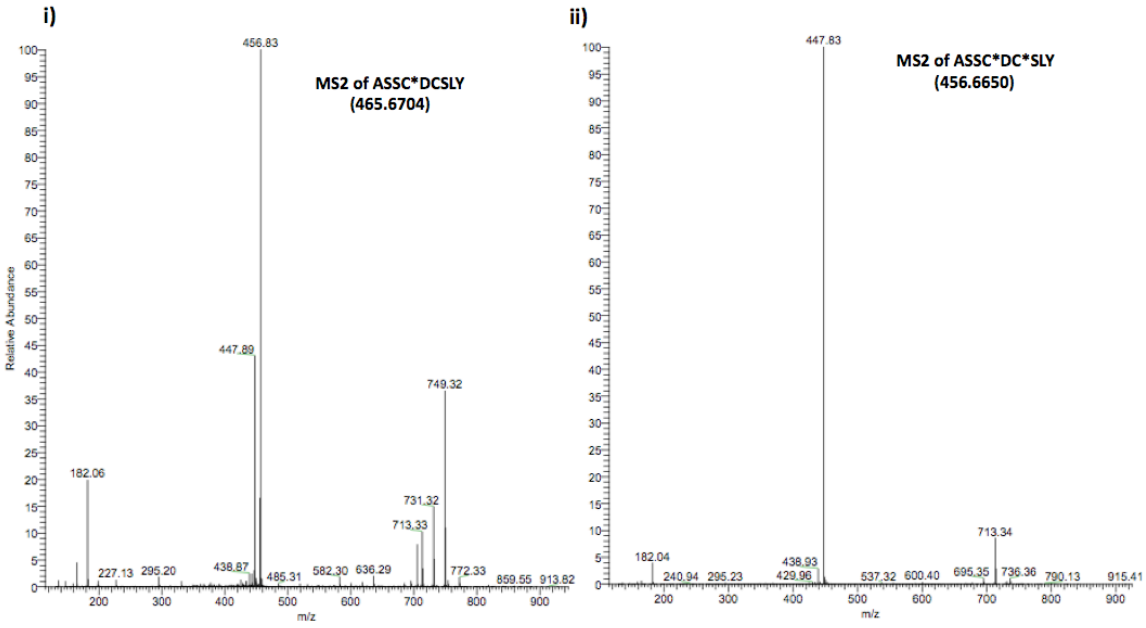

i)
MS2 of ASSC*DCSLY
(465.6704)
456.83
447.89
749.32
182.06
731.32
713.33
227.13
295.20
438.87
485.31
582.30
636.29
772.33
859.55
913.82
Relative Abundance
m/z
ii)
MS2 of ASSC*DC*SLY
(456.6650)
447.83
713.34
182.04
240.94
295.23
438.93
429.96
537.32
600.40
695.35
736.36
790.13
915.41
m/z

Table S1. Localization of heterocyclization PTMs shown in Figure S3 to specific Cys residues

A

i)

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGCESCSYEGDEAE	-	-	-	-
GG	-	-	-	-	GCESCSYEGDEAE	-	-	-	-
GGC	-	-	-	-	CESCSYEGDEAE	-	-	-	-
GGCE	347.21	-	-	329.22	ESCSYEGDEAE	-	-	-	-
GGCES	-	-	-	-	SCSYEGDEAE	1089.44	-	-	1071.42
GGCESC	-	-	-	-	CSYEGDEAE	1002.47	-	-	-
GGCESCS	624.33	-	-	606.33	SYEGDEAE	899.40	-	-	-
GGCESCSY	787.35	-	-	769.31	YEGDEAE	812.39	-	-	794.39
GGCESCSYE	916.33	-	-	-	EGDEAE	649.34	-	-	-
GGCESCSYEG	973.44	-	-	955.35	GDEAE	520.26	-	-	502.26
GGCESCSYEGD	-	-	-	-	DEAE	-	-	-	-
GGCESCSYEGDE	1217.45	-	-	1199.28	EAE	347.21	-	-	-
GGCESCSYEGDEA	1288.37	-	635.73	1270.45	AE	219.13	-	-	-
GGCESCSYEGDEAE	-	-	700.42	-	E	-	-	-	-

MS2 of GGCESCSYEGDEAE (718.2353)

ii)

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGCESC*SYEGDEAE	-	-	-	-
GG	-	-	-	-	GCESC*SYEGDEAE	-	-	-	-
GGC	-	-	-	-	CESC*SYEGDEAE	-	-	-	-
GGCE	347.19	-	-	329.20	ESC*SYEGDEAE	-	-	-	-
GGCES	-	-	-	-	SC*SYEGDEAE	-	-	-	1053.33
GGCESC*	-	-	-	-	C*SYEGDEAE	-	-	-	-
GGCESC*S	606.31	-	-	588.27	SYEGDEAE	-	-	-	-
GGCESC*SY	769.35	-	-	751.29	YEGDEAE	812.39	-	-	794.48
GGCESC*SYE	898.35	-	-	880.39	EGDEAE	649.35	-	-	-
GGCESC*SYEG	955.43	-	-	937.41	GDEAE	520.34	-	-	502.31
GGCESC*SYEGD	-	-	-	-	DEAE	-	-	-	-
GGCESC*SYEGDE	1199.44	-	-	1181.47	EAE	-	-	-	-
GGCESC*SYEGDEA	1270.44	635.84	626.86	1252.46	AE	219.13	-	-	-
GGCESC*SYEGDEAE	-	-	691.46	-	E	-	-	-	-

MS2 of GGCESC*SYEGDEAE (709.2302)

iii)

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGC*ESC*SYEGDEAE	-	700.79	691.32	-
GG	-	-	-	-	GC*ESC*SYEGDEAE	-	-	-	-
GGC*	-	-	-	-	C*ESC*SYEGDEAE	-	-	-	-
GGC*E	329.11	-	-	-	ESC*SYEGDEAE	-	-	-	-
GGC*ES	-	-	-	-	SC*SYEGDEAE	1071.57	-	-	1053.35
GGC*ESC*	-	-	-	-	C*SYEGDEAE	-	-	-	-
GGC*ESC*S	-	-	-	570.33	SYEGDEAE	899.54	-	441.46	-
GGC*ESC*SY	-	-	-	733.31	YEGDEAE	812.37	-	-	794.34
GGC*ESC*SYE	880.41	440.44	-	862.38	EGDEAE	649.41	-	-	-
GGC*ESC*SYEG	-	-	-	919.46	GDEAE	520.31	-	-	-
GGC*ESC*SYEGD	-	-	-	1034.47	DEAE	463.37	-	-	445.34
GGC*ESC*SYEGDE	1181.38	-	-	1163.39	EAE	348.22	-	-	330.22
GGC*ESC*SYEGDEA	1252.39	-	617.87	1234.41	AE	219.11	-	-	-
GGC*ESC*SYEGDEAE	-	691.32	682.41	-	E	-	-	-	-

MS2 of GGC*ESC*SYEGDEAE (700.2243)

B

i)

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ASSC*DCSLY	-	-	456.83	-
AS	-	-	-	-	SSC*DCSLY	859.55	-	-	-
ASS	-	-	-	-	SC*DCSLY	772.33	-	-	-
ASSC*	-	-	-	-	C*DCSLY	-	-	-	-
ASSC*D	-	-	-	-	DCSLY	600.28	-	-	582.30
ASSC*DC	-	-	-	-	CSLY	485.31	-	-	-
ASSC*DCS	636.29	-	-	-	SLY	-	-	-	-
ASSC*DCSL	749.32	-	-	731.32	LY	295.20	-	-	-
ASSC*DCSLY	-	456.83	447.89	-	Y	182.06	-	-	-

MS2 of ASSC*DCSLY (465.6704)

ii)

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ASSC*DC*SLY	-	-	447.83	-
AS	-	-	-	-	SSC*DC*SLY	-	-	-	-
ASS	-	-	-	-	SC*DC*SLY	-	-	-	736.36
ASSC*	-	-	-	-	C*DC*SLY	-	-	-	-
ASSC*D	-	-	-	-	DC*SLY	-	-	-	-
ASSC*DC*	-	-	-	-	C*SLY	-	-	-	-
ASSC*DC*S	-	-	-	600.40	SLY	-	-	-	-
ASSC*DC*SL	-	-	-	713.34	LY	295.23	-	-	-
ASSC*DC*SLY	-	447.83	-	-	Y	182.04	-	-	-

MS2 of ASSC*DC*SLY (456.6650)

Figure S4. (A) Peptides carrying LAELSEEAL RSI always contain azole/azoline heterocycles (yellow circles) whereas peptides lacking RSI always lack heterocycles. Interestingly, such peptides end in a Pro residue (purple residue), since an ultimate cyclic motif is required for C-terminal protease/macrocyclase activity. (B) Presence of LAELSEEAL-like RSI outside of cyanobactin families (core sequence is in red and LAELSEEAL-like sequence is highlighted in yellow).

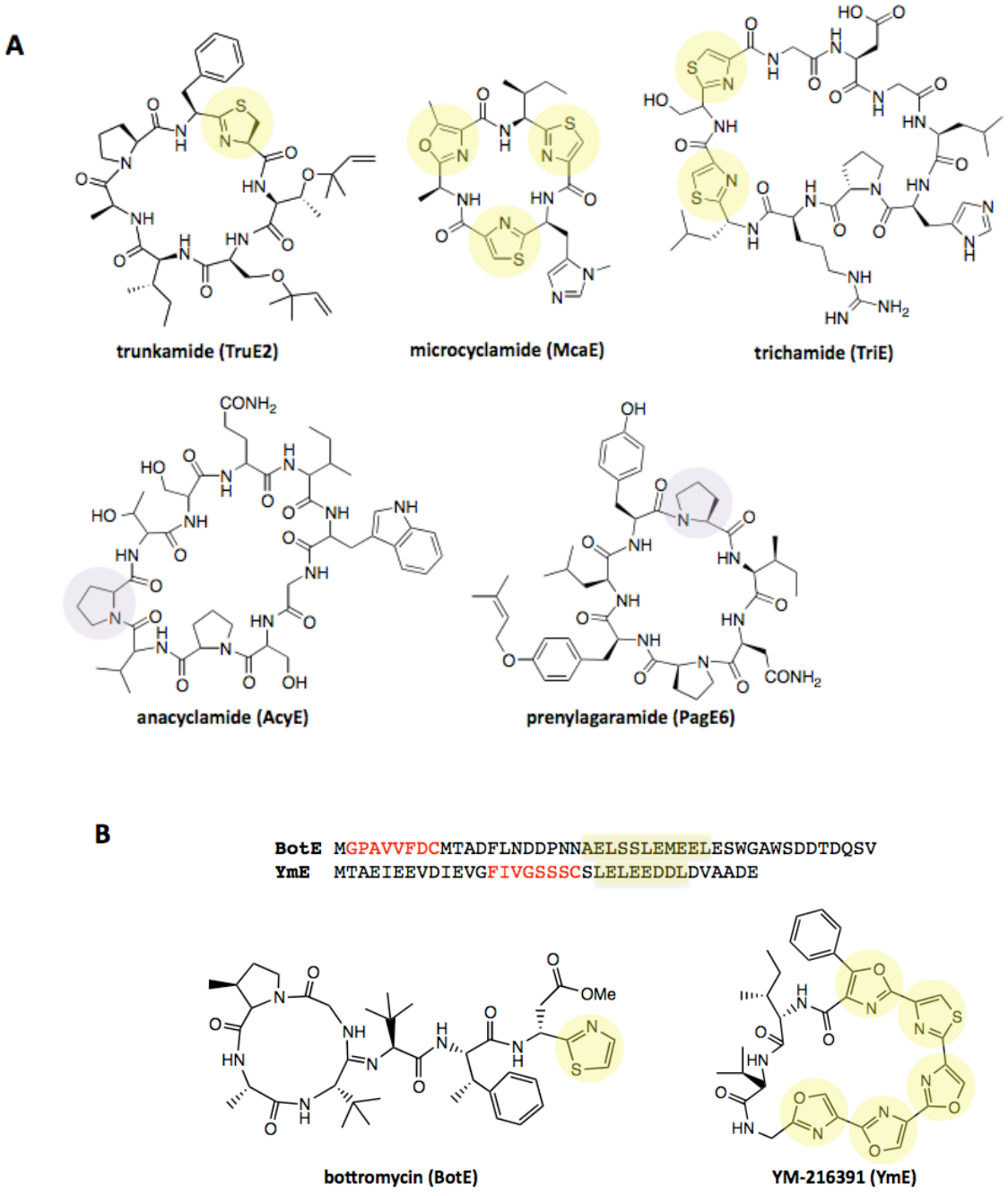


Figure S5. His-tagged expression constructs of (A) hybrid precursor peptide substrates **2**, **3** and **11**, and (B) RSI mutants **8** (of hybrid **3**) and **9-10** (of hybrid **11**). See Figure S2 for the rest of the precursors used in this study.

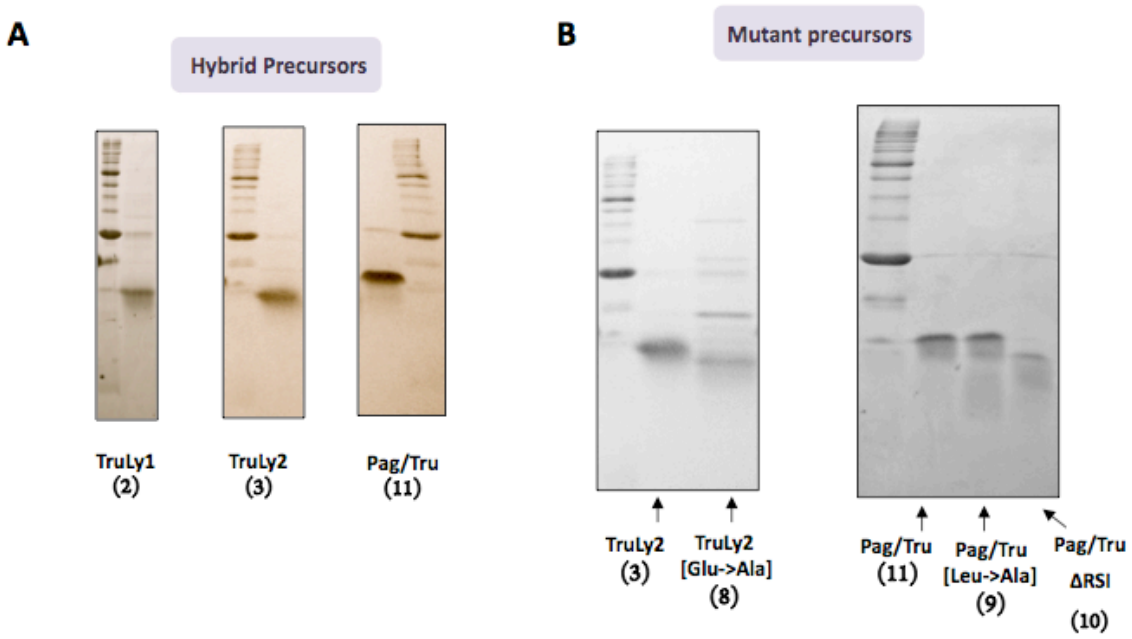
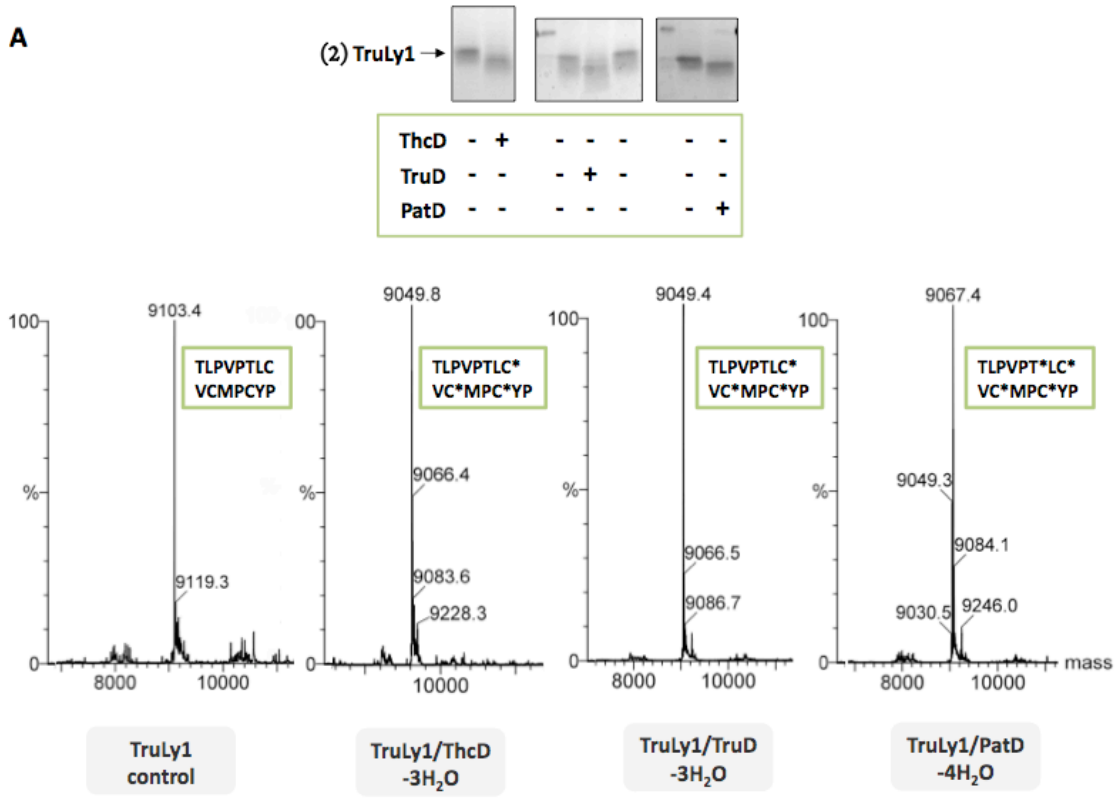


Figure S6. Deconvoluted ESI-MS of heterocyclization reactions of (A) TruLy1 (**2**) and (B) TruLy2 (**3**) with 25 μM substrate and 1 μM each of ThcD, TruD and PatD (18 h), showing $[M+H]^+$ of substrate (control) and its dehydrated products in presence of enzyme. With TruLy2, PatD was relatively slower and the doubly dehydrated product was the major product. The core sequence is boxed in green and heterocycle PTM is shown by C* or T*. SDS-PAGE band-shift visualization is also shown for the same reactions.



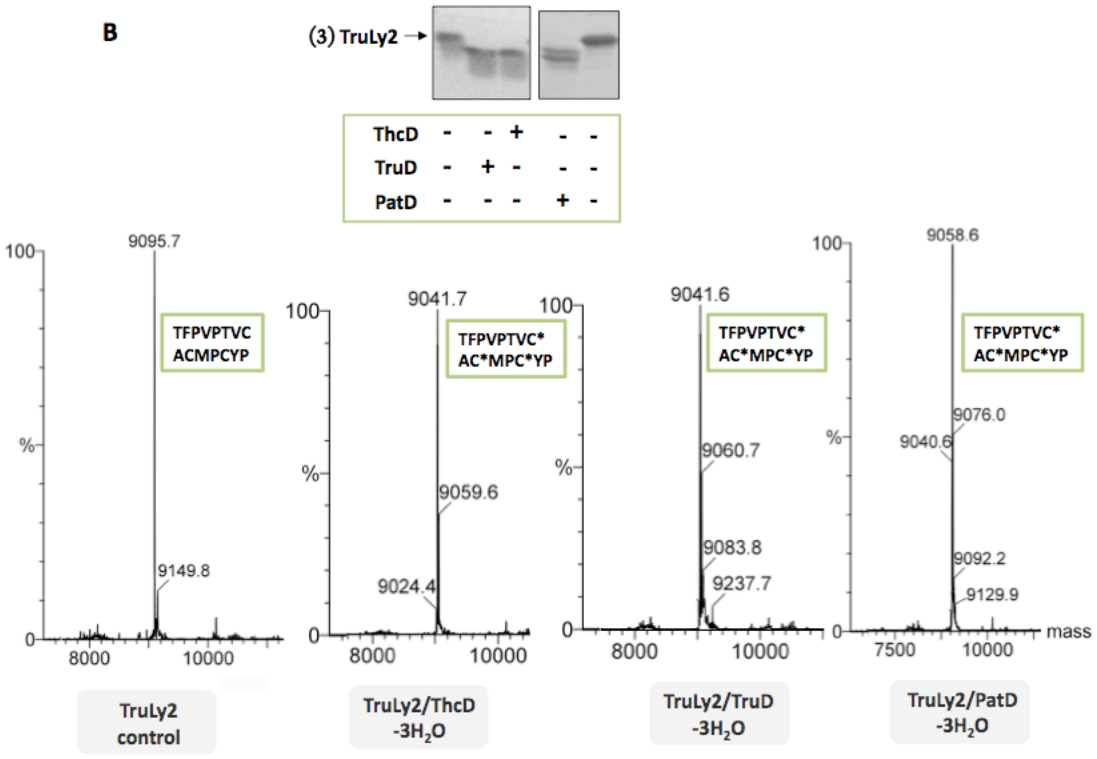
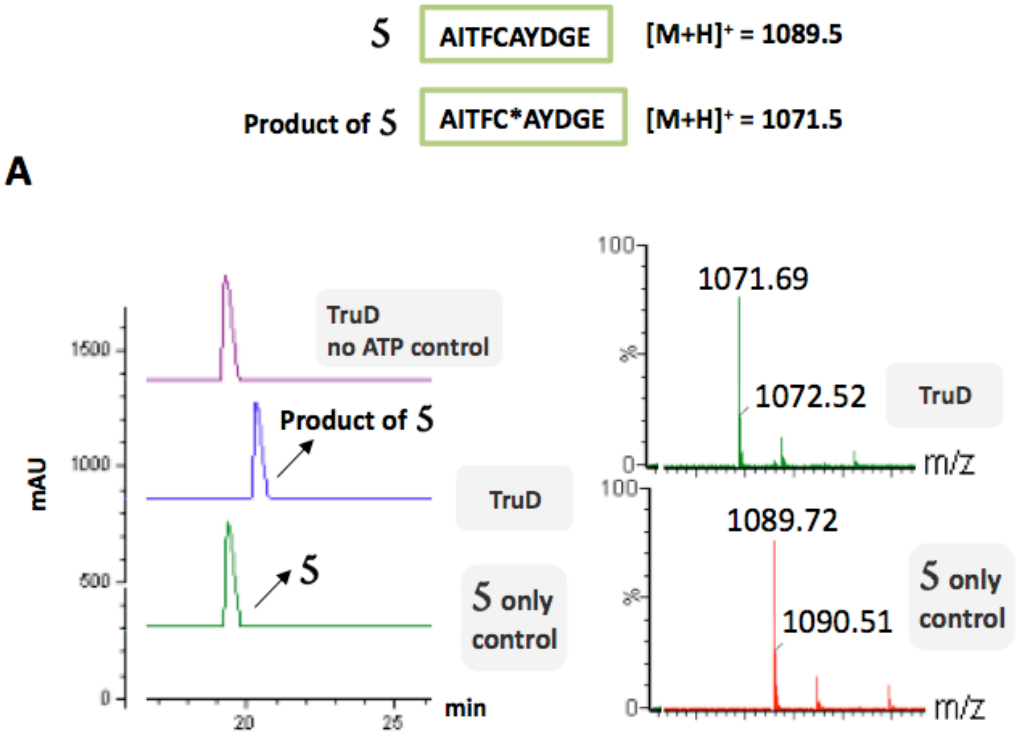

B
(3) TruLy2
ThcD - - + - -
TruD - + - - -
PatD - - - + -
100
%
0
9095.7
9149.8
TFPVPTVC
ACMPCYP
8000
10000
TruLy2
control
9041.7
9059.6
9024.4
TFPVPTVC*
AC*MPC*YP
TruLy2/ThcD
-3H2O
9041.6
9060.7
9083.8
9237.7
TruLy2/TruD
-3H2O
9058.6
9076.0
9040.6
9092.2
9129.9
7500
mass
TruLy2/PatD
-3H2O

Figure S7. HPLC and ESI-MS analysis of heterocyclization reaction of **5** with TruD (A), ThcD and PatD (B). 200 μM substrate was used with 2 μM of each enzyme (21 h reaction). The sequence is boxed is green and thiazoline PTM shown by C* along with expected masses.



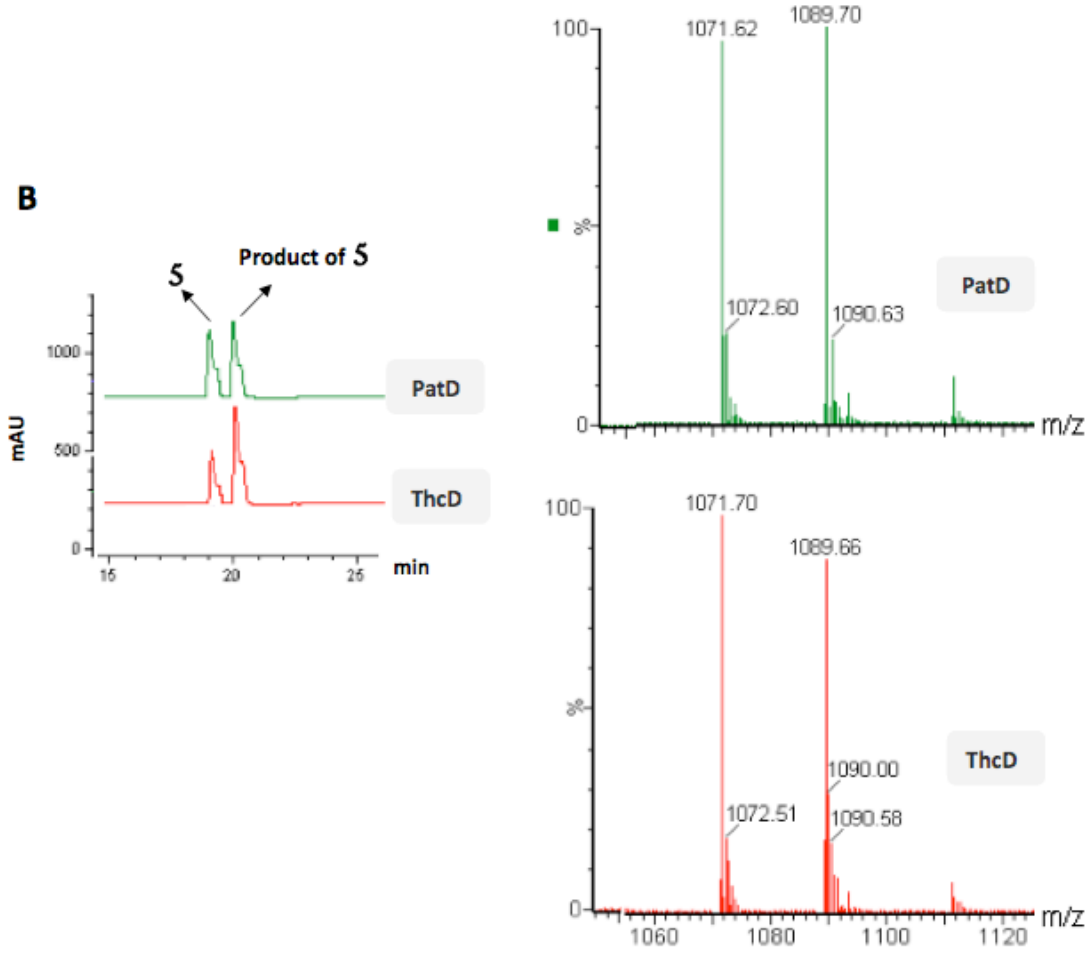

B
5
Product of 5
1000
500
0
mAU
15
20
25
min
PatD
ThcD
100
%
0
1071.62
1089.70
1072.60
1090.63
PatD
m/z
100
%
0
1071.70
1089.66
1090.00
1072.51
1090.58
ThcD
1060
1080
1100
1120
m/z

Figure S8. HPLC and ESI-MS analysis of heterocyclization reaction of **6** with TruD, ThcD and PatD. 200 μM substrate was used with 2 μM of each enzyme (30 h reaction). The sequence is boxed is green and thiazoline PTM shown by C* along with expected masses. Note that an additional +12 mass (of ~1115) is due to thioproline adduct of the N-terminal Cys residue with formaldehyde during MS. The mass peak of 1119.43 is the oxidized form of **6**.

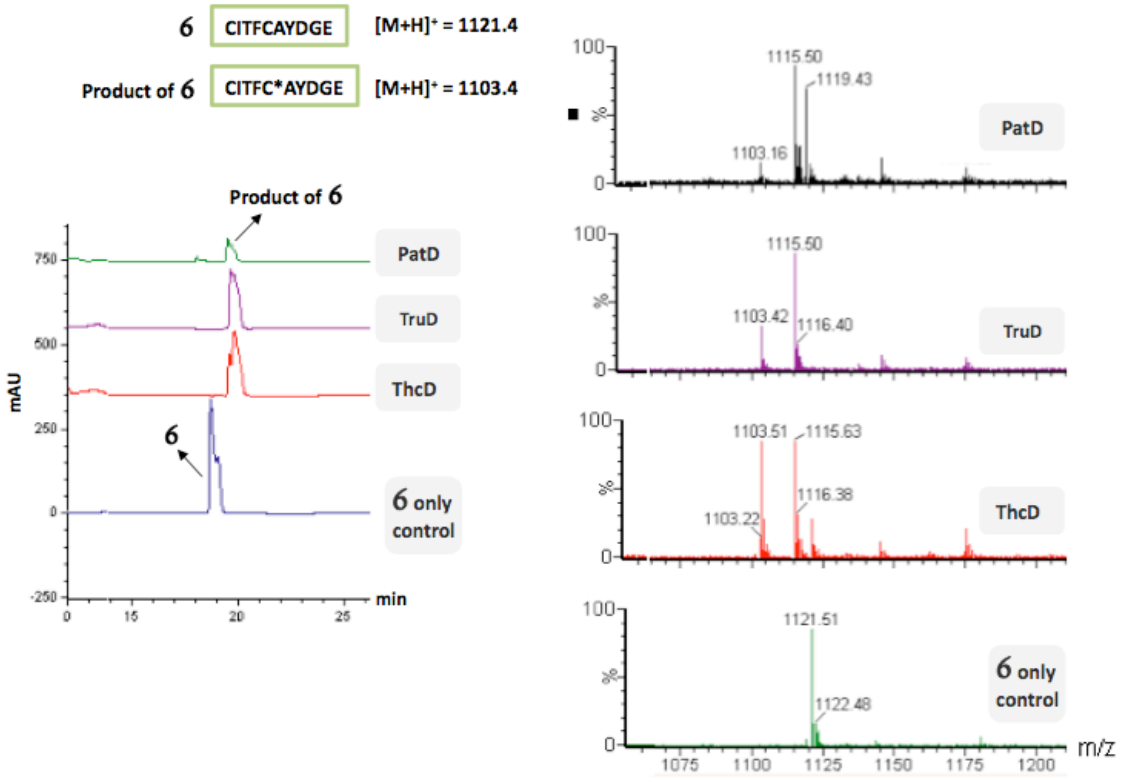
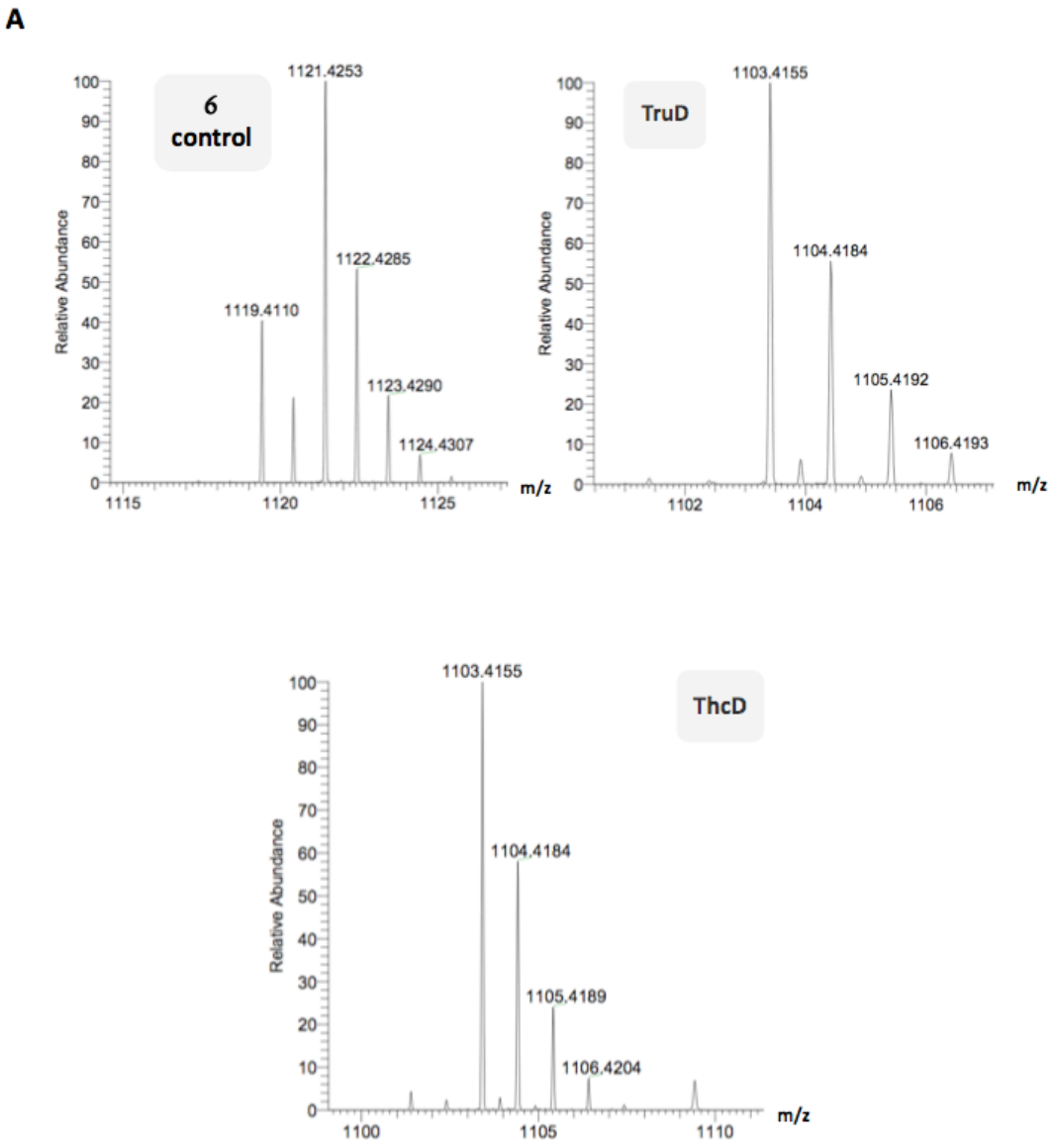


Figure S9. (A) FTMS spectral characterization of the dehydration product of **6** from the above TruD and ThcD reactions confirming thiazoline formation, showing $[M+H]^+$ of substrate (control) and its products in presence of enzyme. (B) MS2 of the same.

Table S2. MS2 fragmentation analysis of reactions shown in Figure S9.



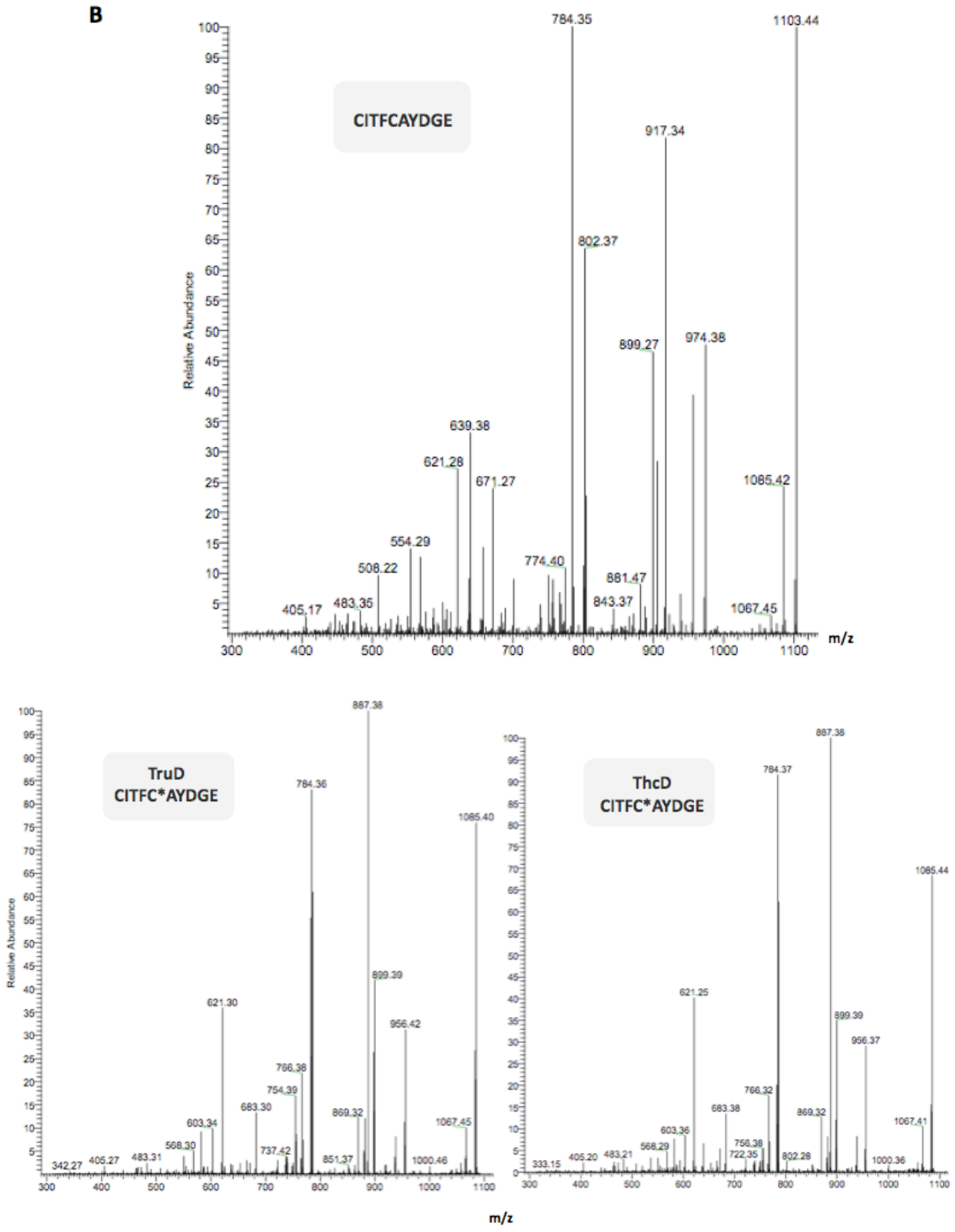

B
CITFCAYDGE
Relative Abundance
m/z
784.35
1103.44
917.34
802.37
899.27
974.38
639.38
621.28
671.27
1085.42
554.29
508.22
774.40
881.47
843.37
405.17
483.35
1067.45
TruD
CITFC*AYDGE
887.38
784.36
1085.40
899.39
621.30
956.42
766.38
754.39
683.30
869.32
603.34
568.30
737.42
1067.45
342.27
405.27
483.31
851.37
1000.46
ThcD
CITFC*AYDGE
887.38
784.37
1085.44
621.25
899.39
956.37
766.32
683.38
869.32
603.36
568.29
756.38
722.35
1067.41
333.15
405.20
483.21
802.28
1000.36

Table S2

MS2 of CITFCAYDGE

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
C	-	-	-	-	CITFCAYDGE	-	-	-	-
CI	-	-	-	-	ITFCAYDGE	-	-	-	-
CIT	-	-	-	-	TFCAYDGE	-	-	-	-
CITF	-	-	-	-	FCAYDGE	-	-	-	-
CITFC	-	-	-	-	CAYDGE	-	-	-	-
CITFCA	639.38	-	-	621.25	AYDGE	554.29	-	-	-
CITFCAY	802.37	-	-	784.35	YDGE	483.35	-	-	-
CITFCAYD	917.34	-	-	899.27	DGE	-	-	-	-
CITFCAYDG	974.38	-	-	-	GE	-	-	-	-
CITFCAYDGE	1103.44	-	-	1085.42	E	-	-	-	-

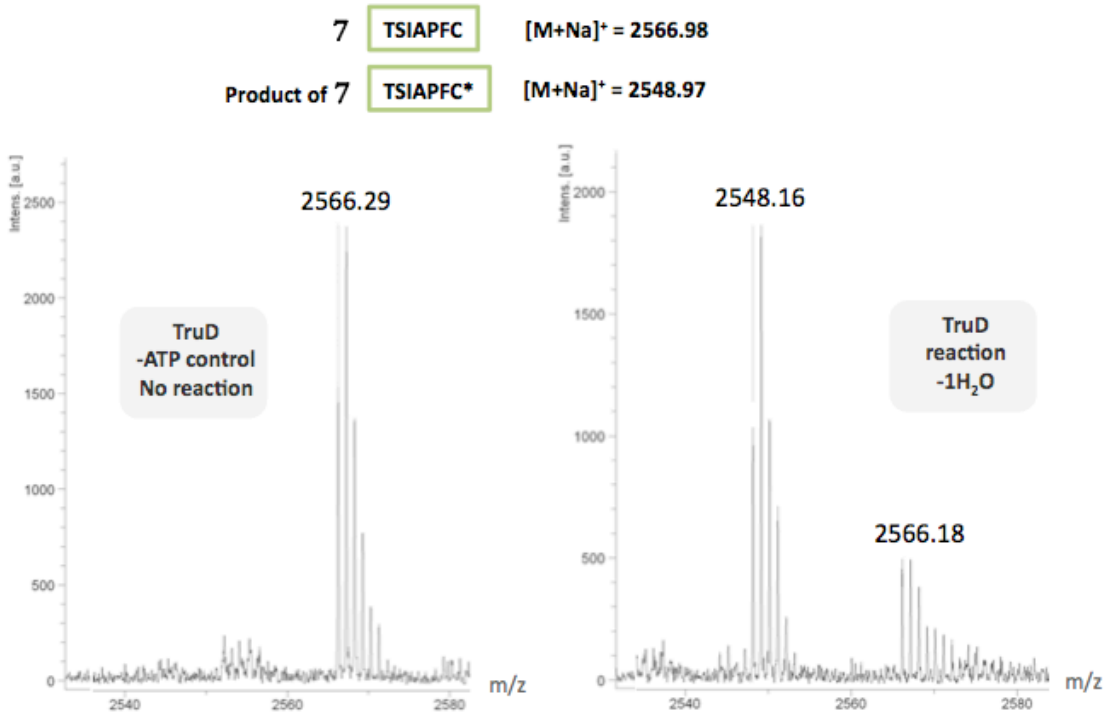
MS2 of TruD reaction CITFC*AYDGE

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
C	-	-	-	-	CITFC*AYDGE	-	-	-	-
CI	-	-	-	-	ITFC*AYDGE	1000.46	-	-	-
CIT	-	-	-	-	TFC*AYDGE	887.38	-	-	869.32
CITF	-	-	-	-	FC*AYDGE	-	-	-	-
CITFC*	-	-	-	-	C*AYDGE	-	-	-	-
CITFC*A	621.25	-	-	603.24	AYDGE	-	-	-	-
CITFC*AY	784.36	-	-	766.38	YDGE	483.31	-	-	-
CITFC*AYD	899.39	-	-	-	DGE	-	-	-	-
CITFC*AYDG	956.42	-	-	-	GE	-	-	-	-
CITFC*AYDGE	1085.40	-	-	1067.45	E	-	-	-	-

MS2 of ThcD reaction CITFC*AYDGE

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
C	-	-	-	-	CITFC*AYDGE	-	-	-	-
CI	-	-	-	-	ITFC*AYDGE	1000.36	-	-	-
CIT	-	-	-	-	TFC*AYDGE	887.38	-	-	869.32
CITF	-	-	-	-	FC*AYDGE	-	-	-	-
CITFC*	-	-	-	-	C*AYDGE	-	-	-	-
CITFC*A	621.25	-	-	603.36	AYDGE	-	-	-	-
CITFC*AY	784.37	-	-	766.32	YDGE	483.21	-	-	-
CITFC*AYD	899.39	-	-	-	DGE	-	-	-	-
CITFC*AYDG	956.37	-	-	-	GE	-	-	-	-
CITFC*AYDGE	1085.44	-	-	1067.41	E	-	-	-	-

Figure S10. MALDI-MS analysis of heterocyclization of **7** by TruD, ThcD and PatD. The core sequence of **7** is given in a green box along with expected masses of substrate and product. Substrate was used at 120 μM with 0.5 μM of each enzyme. The 254 nm UV absorbance shoulder from HPLC analysis (see main text) provided confirmation that the dehydration seen is thiazoline modification.



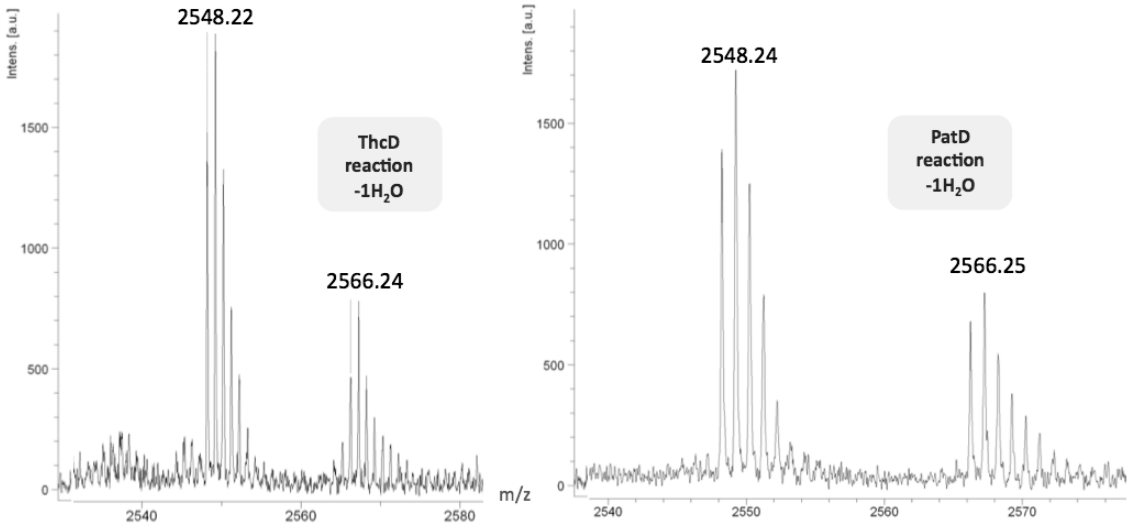

Intens. [a.u.]
2548.22
ThcD reaction -1H2O
2566.24
m/z
Intens. [a.u.]
2548.24
PatD reaction -1H2O
2566.25

Figure S11. Comparison of the heterocyclization reaction progress of substrates **5-7** (200 μM each) with 2 μM ThcD enzyme. ESI-MS (A) was used to follow heterocyclization of **7** due to difficulty in separating product and substrate peaks by HPLC. (B) and (C) show time-courses of substrates **5-6** by HPLC analysis. Confirmation of the identity of the product peak was provided by ESI-MS.

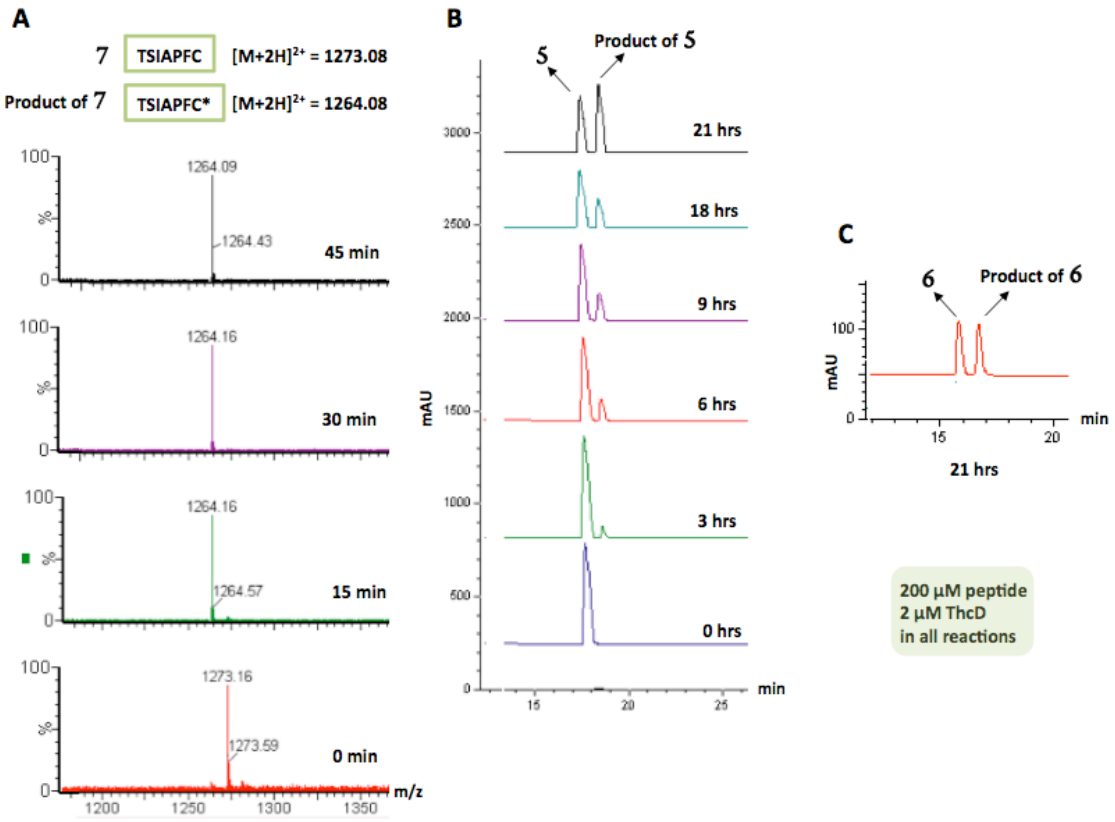


Figure S12. Comparison of the reaction progress of heterocyclization of substrates **2-3** (A) with **7** (B). Substrates were 50 μM each with 2 μM ThcD. **2-3** were analysed by SDS-PAGE band-shift due to clear visibility (the 18-h time-point was confirmed by ESI-MS to be the final product), and **7** was analysed by ESI-MS.

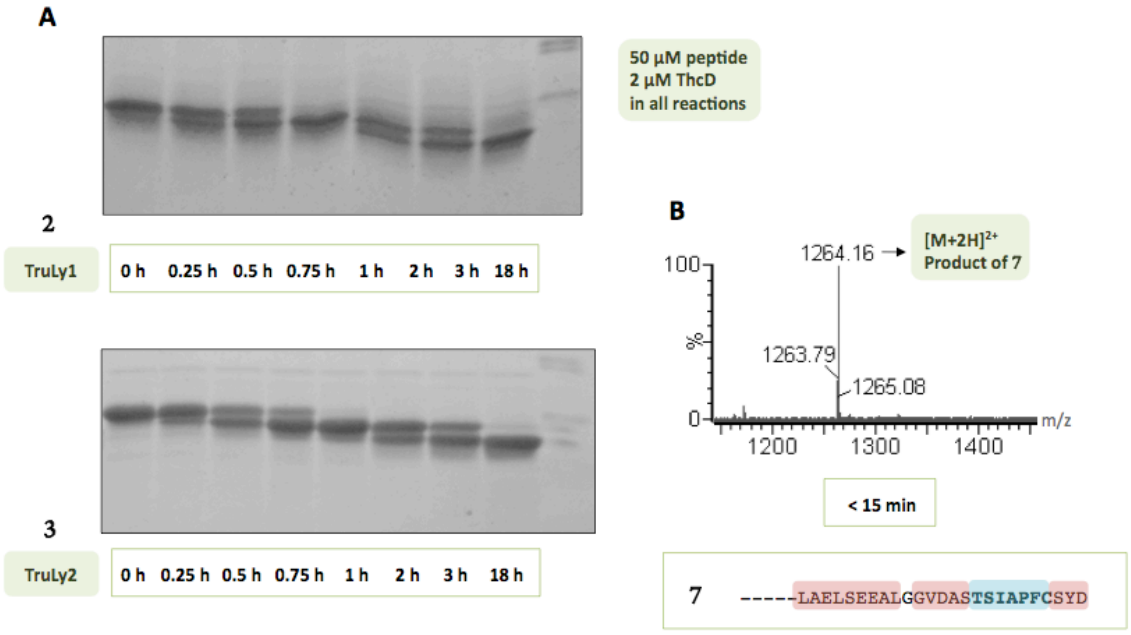
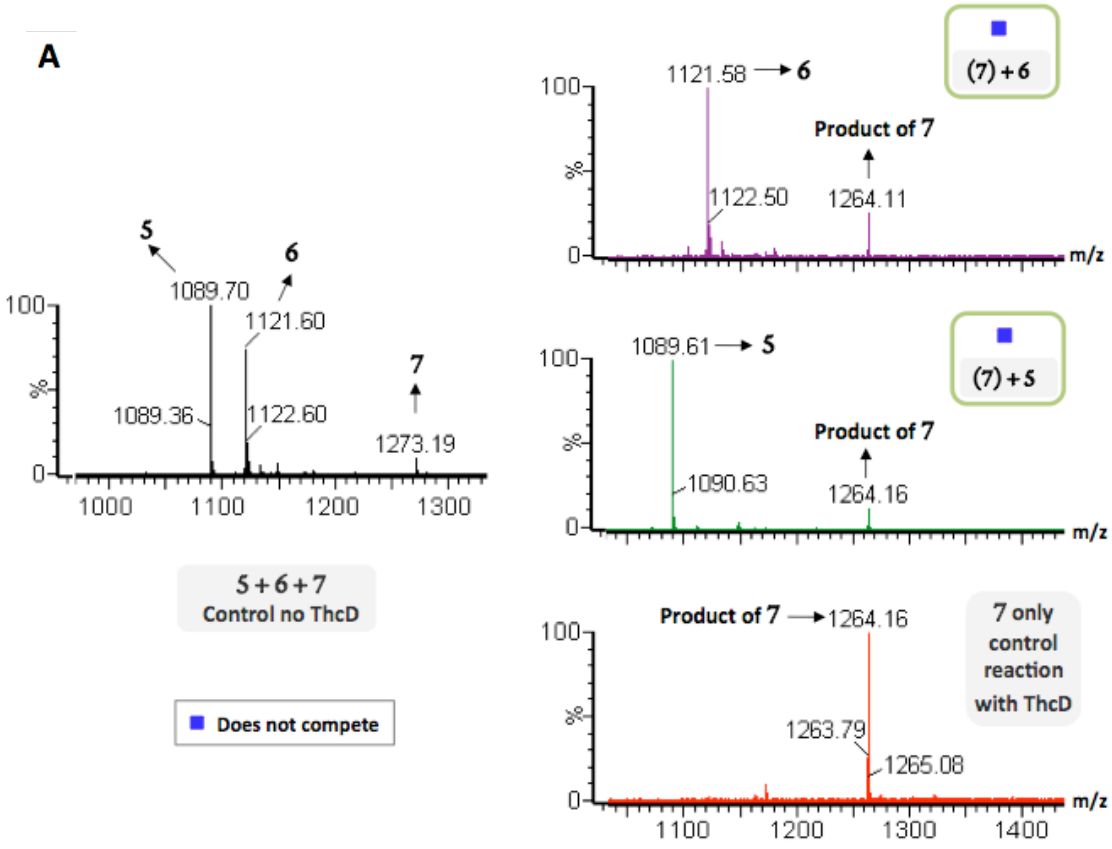


Figure S13. (A) Competition reactions of substrates **5-6** (both lacking RSI) with **7**, containing RSI show no inhibition of modification of **7** by these substrates. Each substrate was maintained at a concentration of 200 μM, and reactions were carried out with 2 μM ThcD (15 m), and analysed by ESI-MS. The masses for substrates **5-6** is $[M+H]^+$ while those for **7** and its product is $[M+2H]^{2+}$. Control reactions of peptides only without enzyme and ThcD/**7** reaction in absence of competing substrates at the same time-point are shown. (B) A 3 h time-point of reaction of **7** in presence of **5**, shows that **5** is still capable of being a substrate in presence of **7**, and modification does occur at a later time-point.



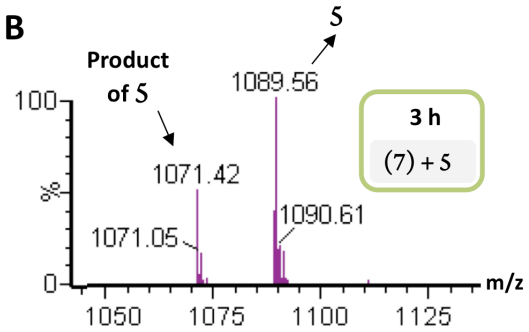


Figure S14. (A) WebLogo analysis of a stretch of sequence containing RSI in addition to four residues on either end. The sequences were taken from an alignment of all known cyanobactin precursors. (B) Predominant residues at each position were used to create the helical wheel projection using a web-based program (http://rzlab.ucr.edu/scripts/wheel/wheel.cgi). Residues from RSI are highlighted in yellow, and the conserved Leu and Glu can be seen to be on opposite faces of the helix.

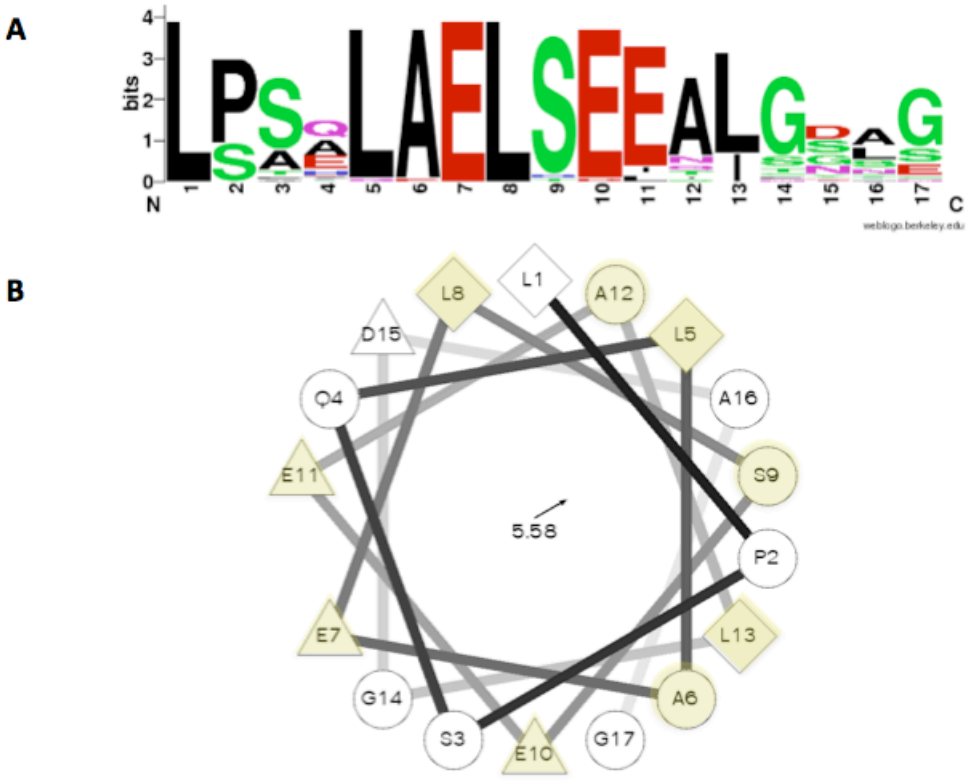


Figure S15. Deconvoluted ESI-MS of TruLy2 (3[Glu->Ala]) mutant (**8**) reaction with TruD and ThcD in comparison to a control reaction without enzyme, showing $[M+H]^+$ of substrate. In presence of enzyme, up to two 18 Da mass-shift was observed corresponding to heterocycle formation. Peptides were used at a concentration of 25 μM with 1 μM of each enzyme (18 h).

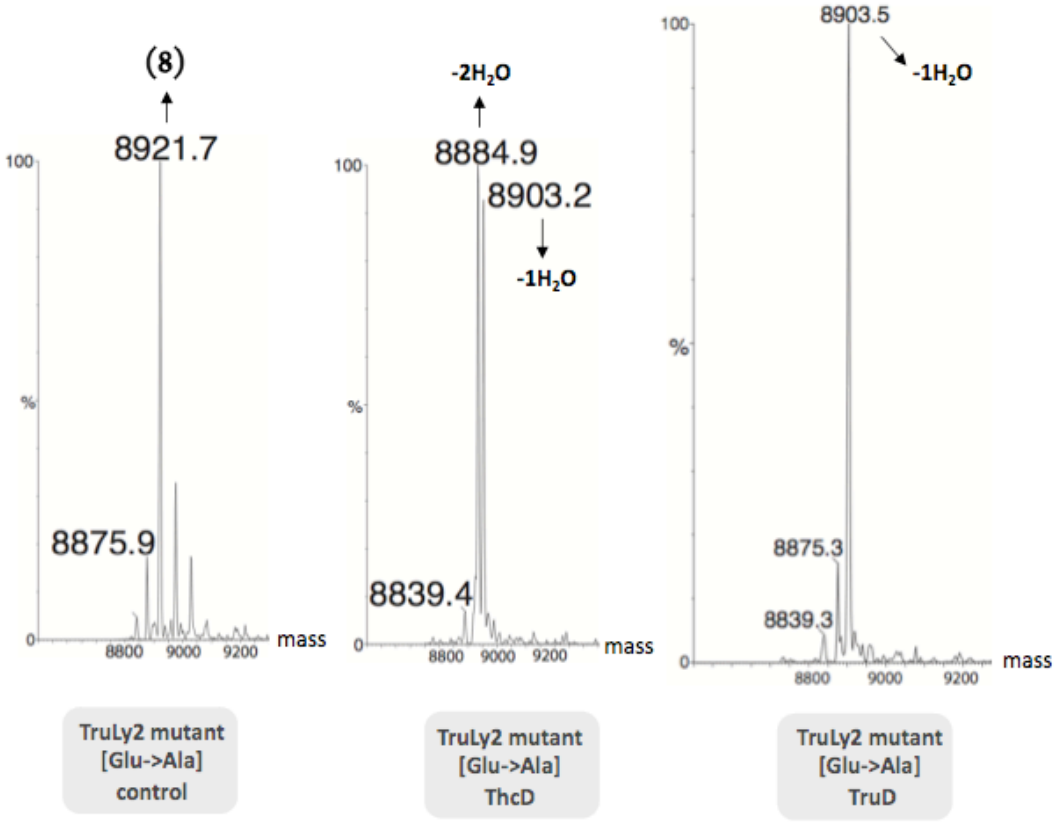


Figure S16. Deconvoluted ESI-MS of Pag/Tru (3[Leu->Ala]) (**9**) reaction with TruD and ThcD in comparison to a control reaction without enzyme, showing $[M+H]^+$ of substrate. In presence of enzyme, the 18 Da mass-shift was not observed, indicating lack of heterocycle formation. Peptides were used at a concentration of 25 μM with 1 μM of each enzyme (18 h).

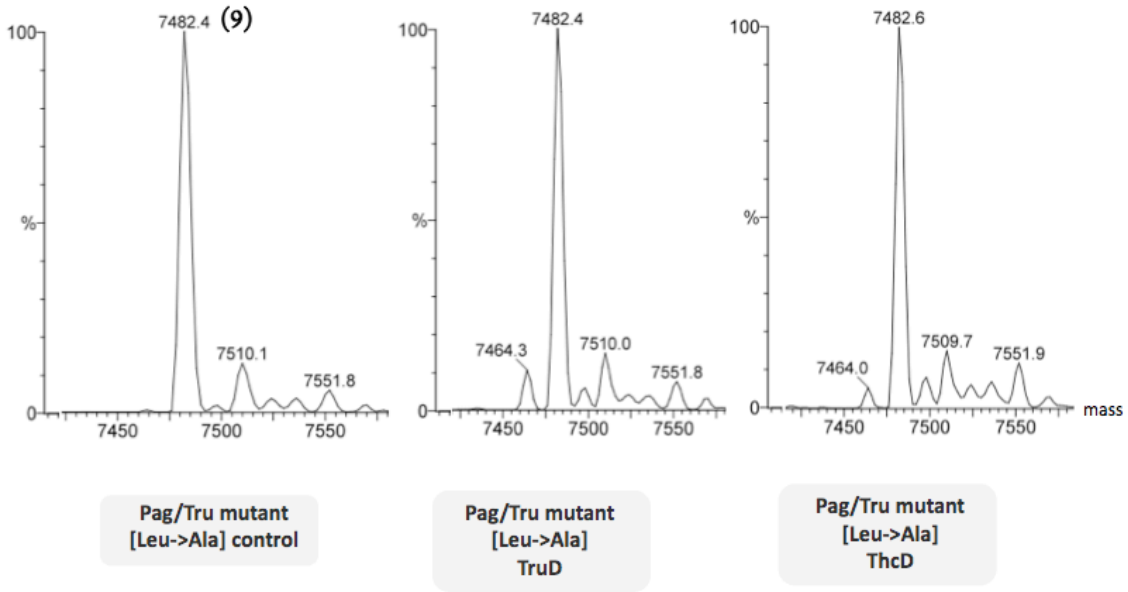


Figure S17. Deconvoluted ESI-MS of (A) Pag/Tru ΔRSI mutant (**10**) vs (B) Pag/Tru RSI insertion (**11**) reactions in comparison to control reactions without enzyme, showing $[M+H]^+$ of substrate. The RSI could be imported into natively non-heterocyclized *pag*-based core (green box) to introduce heterocyclization. Thiazoline modification is shown by C*. Peptides were used at a concentration of 25 μM with 1 μM of each enzyme (18 h).

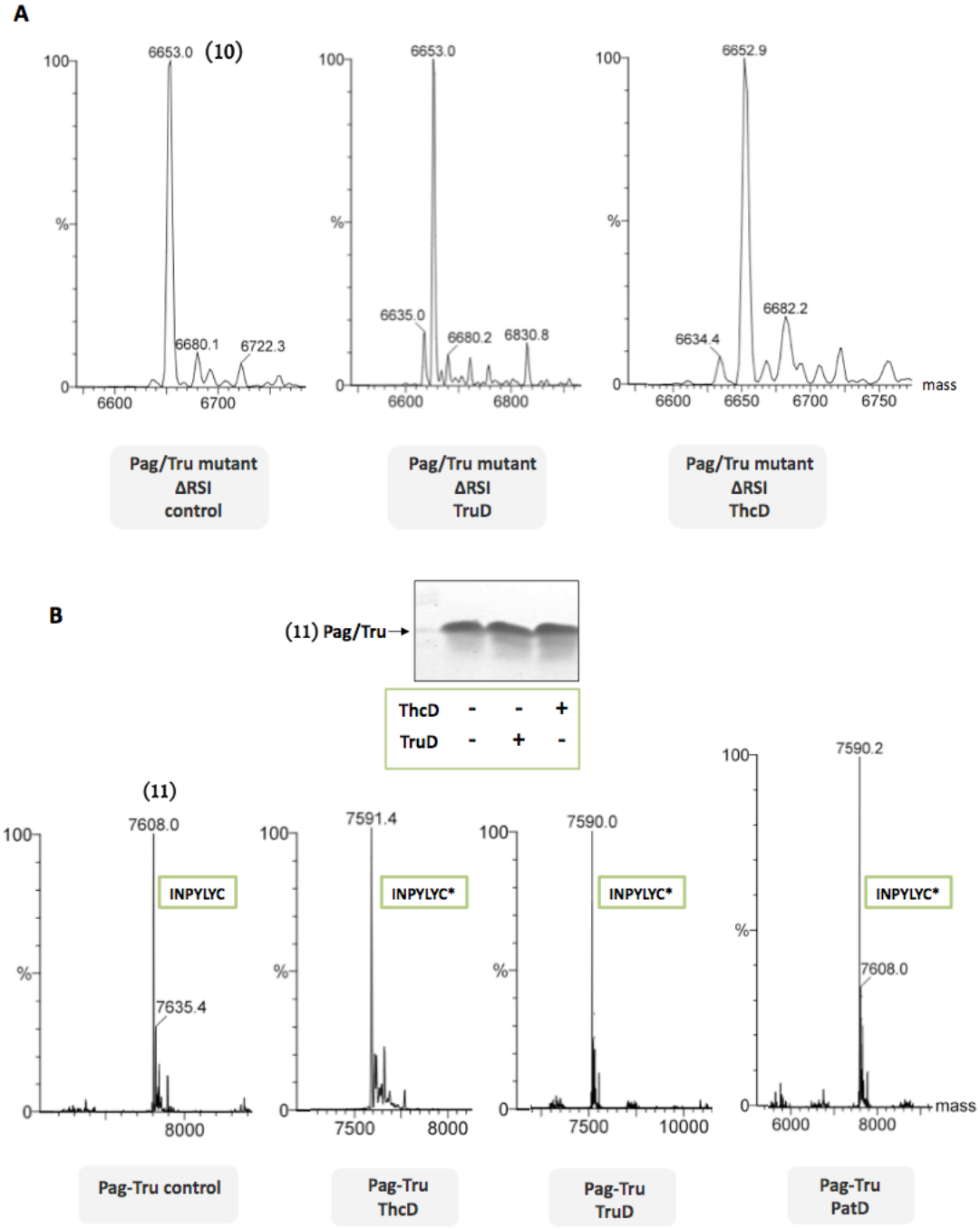


Figure S18. MALDI-MS analysis of competition reactions of substrate **7** with RSI mutants **8** and **10** (competition reactions with substrates **9** and **11** is shown in Figure 4). Reactions containing **8** or **10** did not compete with modification of **7**, and are comparable to a control reaction of **7** with ThcD, which was carried out in absence of competing substrates (where only a minor baseline level of starting substrate mass peak is visible). All substrates were maintained at a concentration of 50 μM with 2 μM ThcD (15 m). The mass peak 2566 corresponds to $[M+Na]^+$ of substrate **7** and 2548 is the $[M+Na]^+$ of its product.

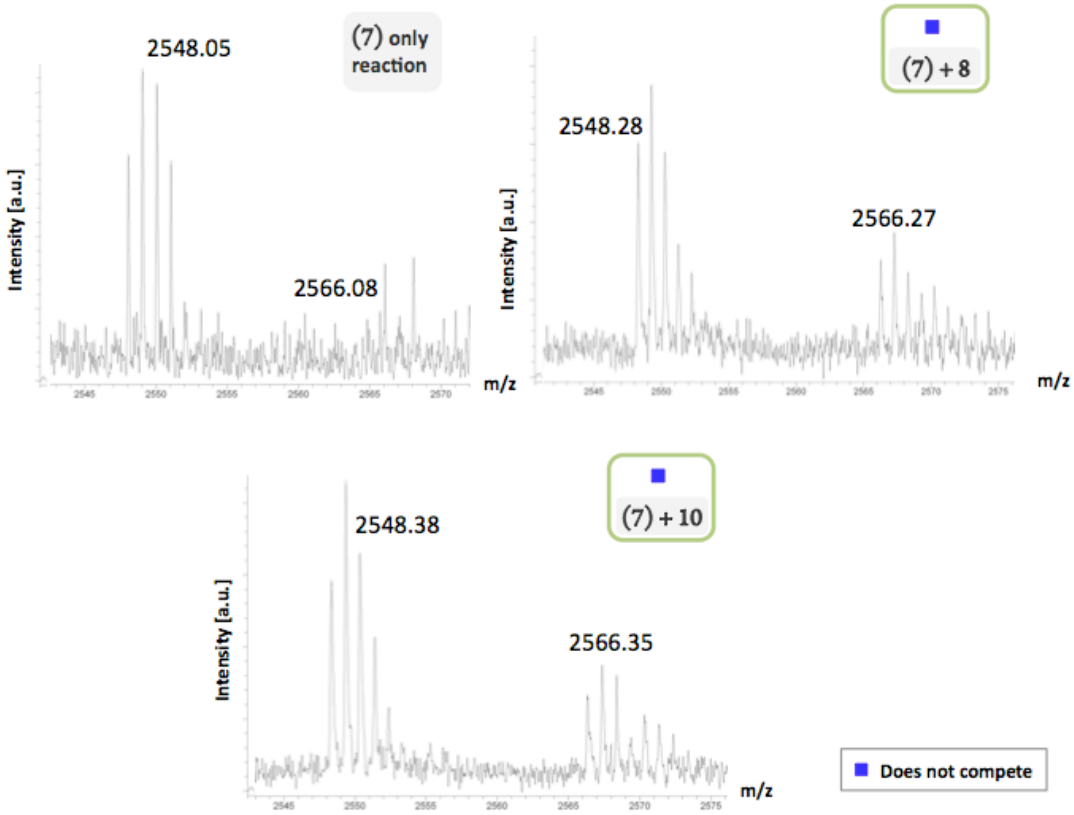
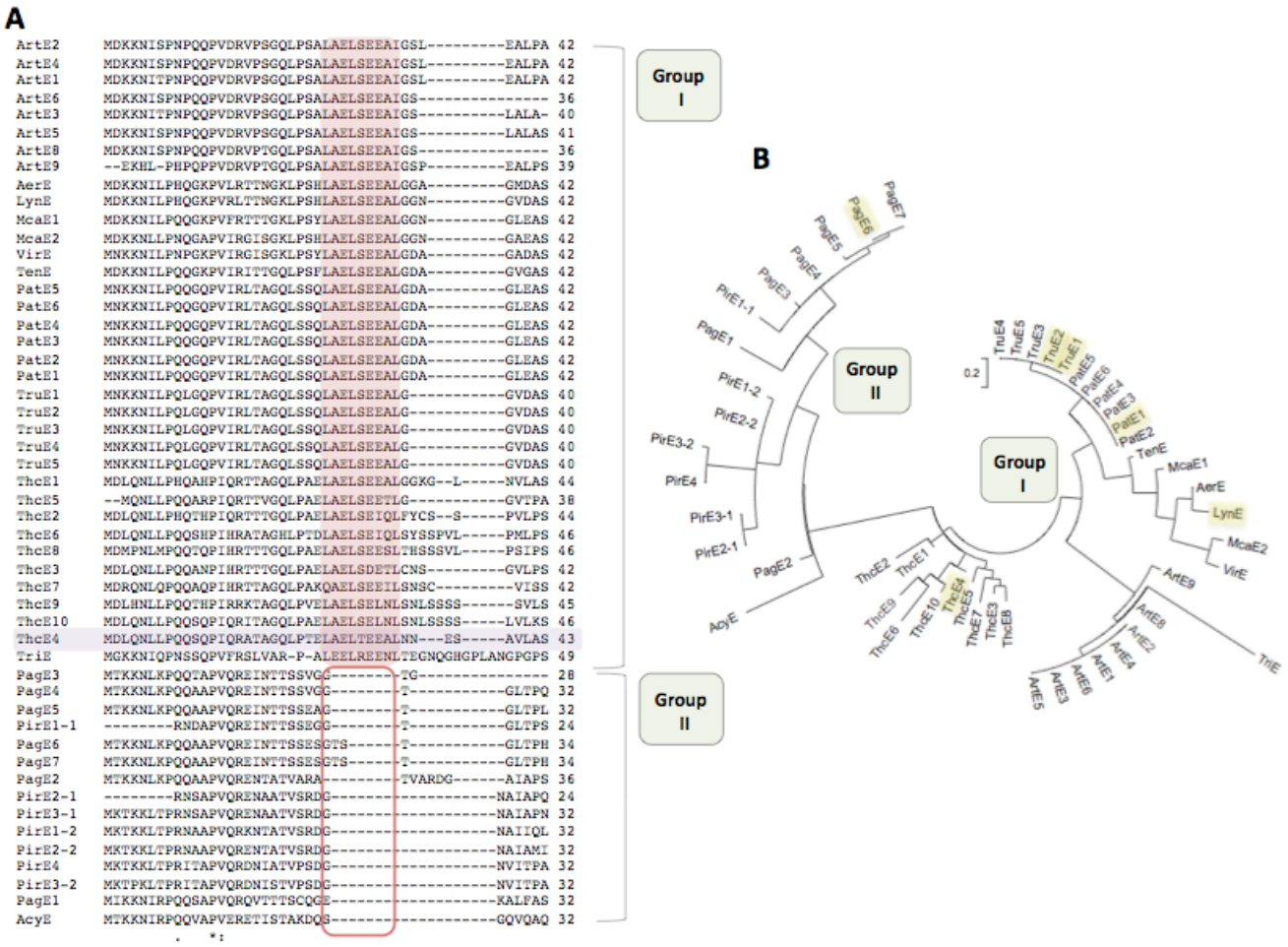


Figure S19. (A) Sequence alignment of precursor peptide sequences used in constructing the phylogenetic tree shown in (B). Only those precursors are shown that are from gene clusters with characterized cyanobactins. The recently characterized linear cyanobactins aeruginosamide (AerE) and viridisamide (VirE) are also included. Alignments were done starting from the N-terminal end to RSII. The RSII sequences are based on previously published studies (see Methods). The ThcE1-10 recognition sequences are predicted based on our *in vitro* analysis of ThcE4 proteolysis (ThcE4 is highlighted). (B) Phylogenetic analysis of precursor peptides using Maximum Likelihood. Note that the tree can be divided into two distinct clades based on the presence/absence of the LAELSEEAL-like RSI element (Group I/Group II). Precursors used in this study are highlighted in yellow. (C) Phylogenetic tree based on alignment of leader sequences only, from which both RSI and RSII were deleted.



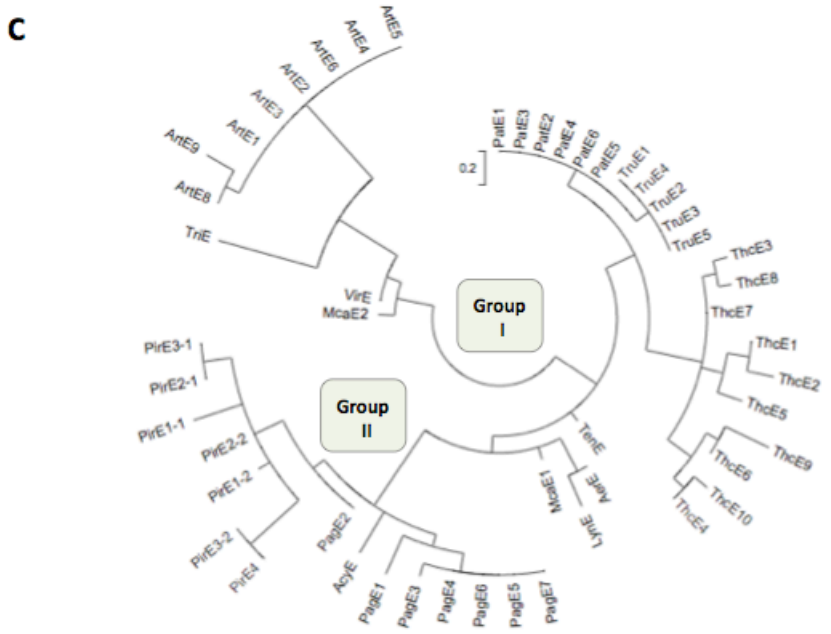


Figure S20. Phylogenetic tree of D-heterocyclase enzymes of the respective Group I precursor peptides from pathways shown in Figure S19.

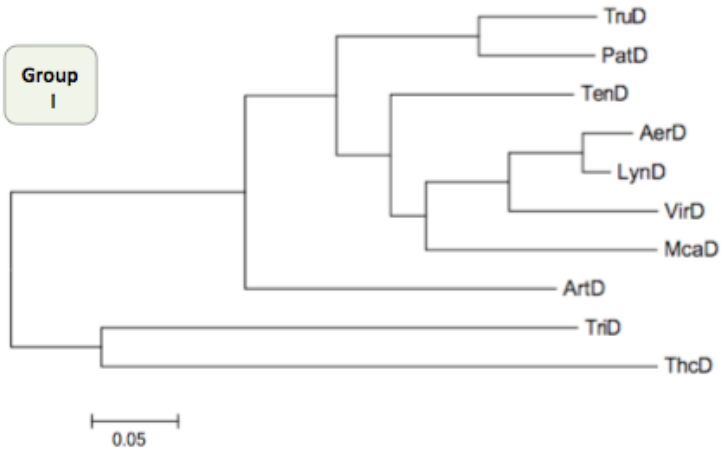
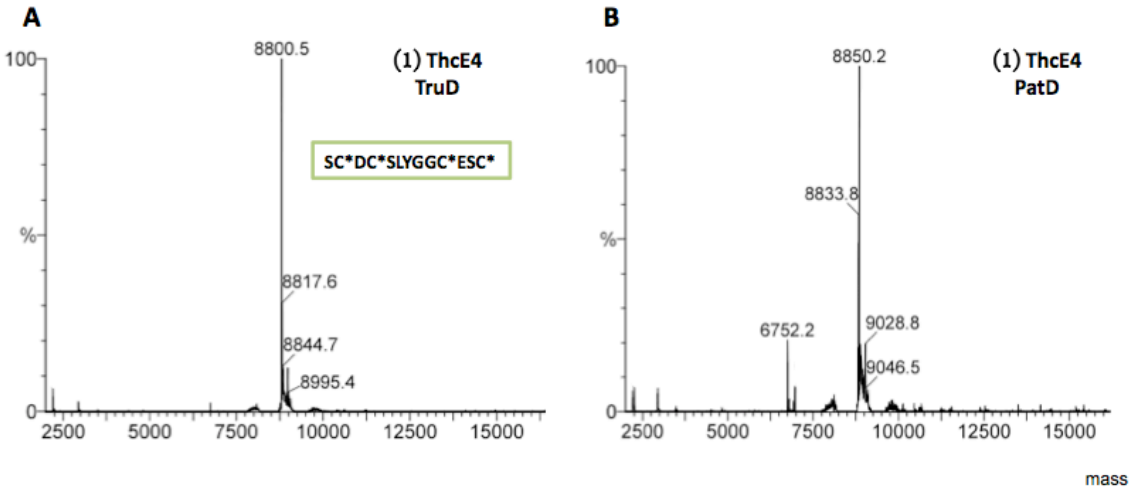
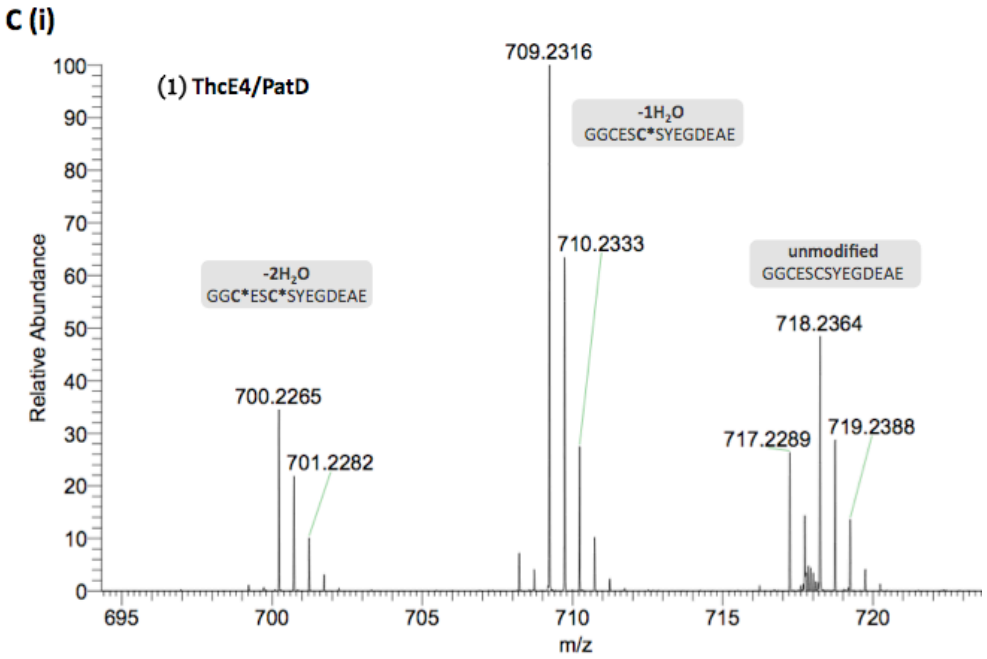
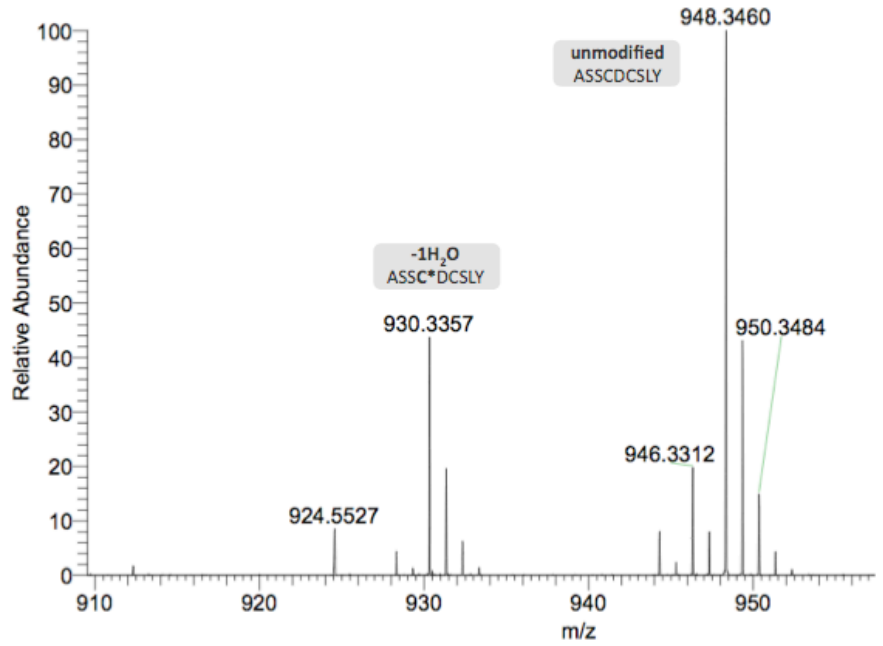


Figure S21. Deconvoluted ESI-MS spectra of ThcE4 (**1**) reaction with TruD (A) and PatD (B). Reaction with ThcD is shown in Figure 2. The core sequence of ThcE4 is given in a green box, and the mass peaks correspond to $[M\text{-}H]^-$. For PatD, confirmation of dehydration is provided by FTMS and MS/MS spectra of chymotryptic digests shown in (C), where the parent ions correspond to $[M+2H]^{2+}$ species. Reactions were carried out with 30 μM of each substrate with 0.4 μM enzyme (18 h). Note that under these conditions PatD was slow and showed only one heterocycle formation (B), which led us to carry the reaction for 45 h (0.6 μM PatD) before FTMS analysis. Thiazoline modification is shown by **C***.

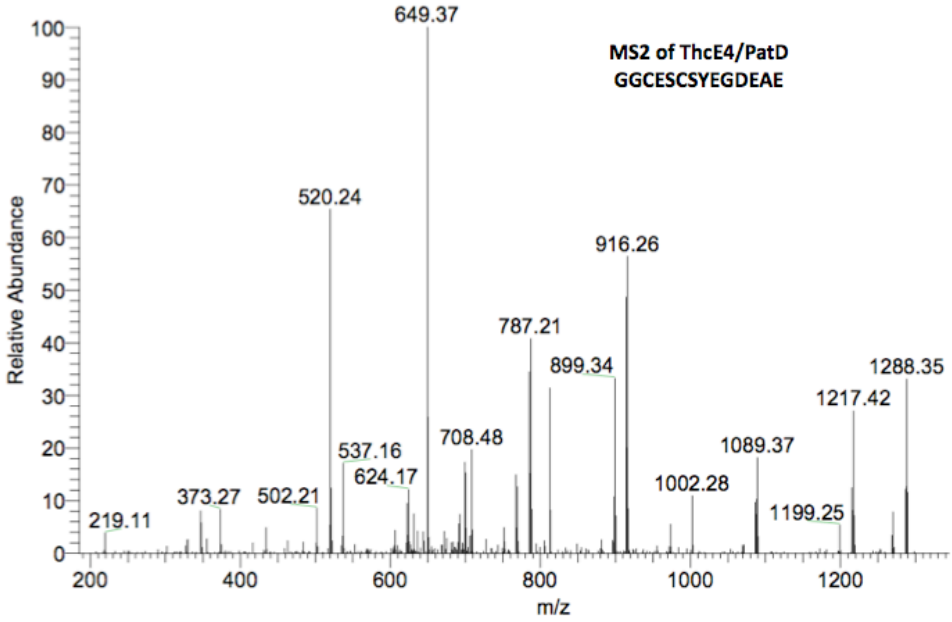




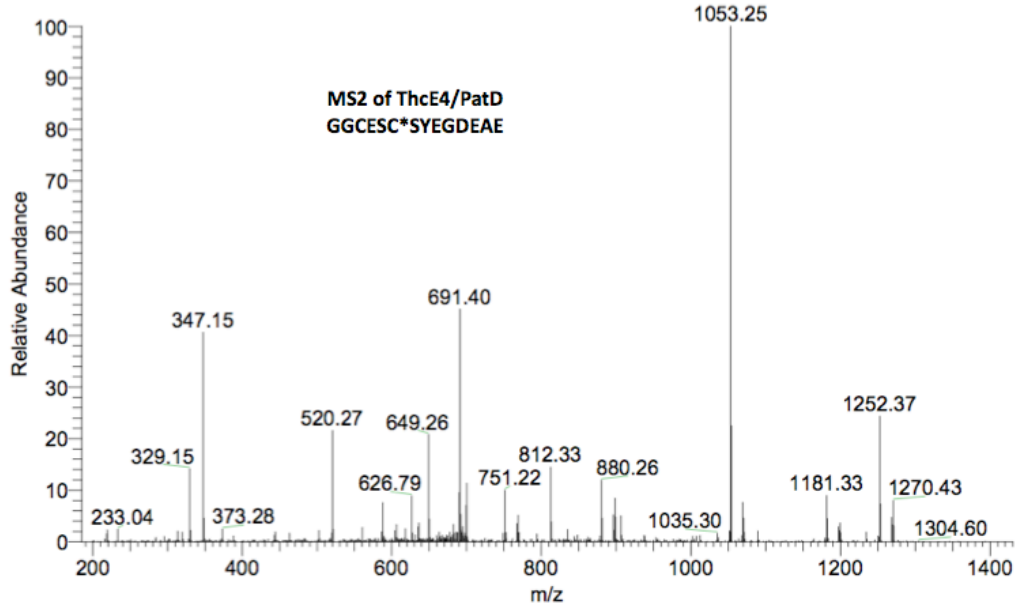
(ii)


948.3460
unmodified
ASSCDCSLY
-1H2O
ASSC*DCSLY
930.3357
950.3484
946.3312
924.5527
Relative Abundance
m/z

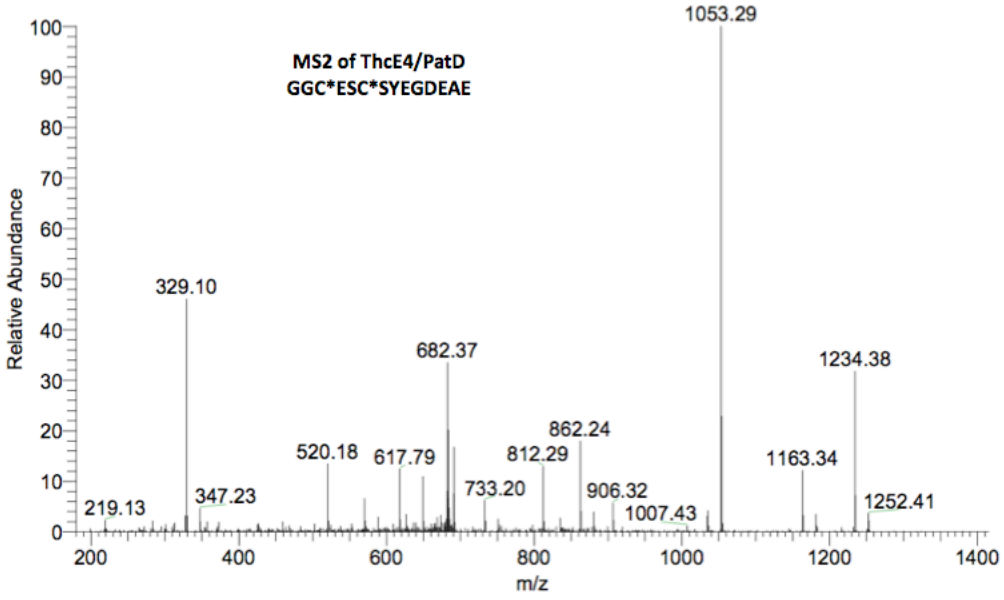
(iii)


649.37
MS2 of ThcE4/PatD
GGCESCSYEGDEAE
520.24
916.26
787.21
899.34
1288.35
1217.42
708.48
537.16
624.17
1089.37
1002.28
373.27
502.21
1199.25
219.11
Relative Abundance
m/z

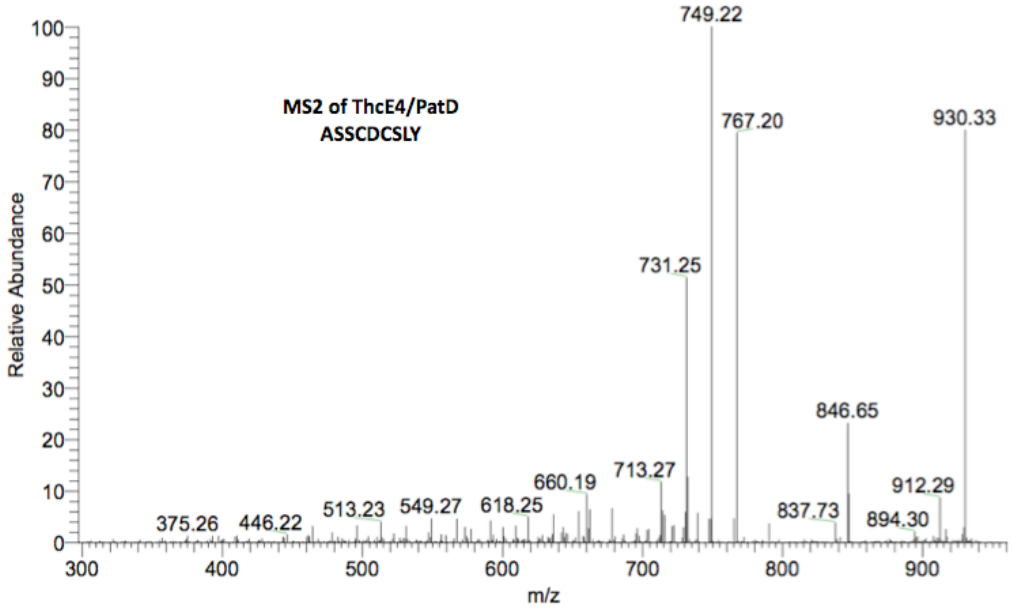
(iv)


MS2 of ThcE4/PatD
GGCESC*SYEGDEAE
Relative Abundance
m/z
1053.25
691.40
347.15
1252.37
520.27
649.26
329.15
812.33
880.26
626.79
751.22
1181.33
1270.43
233.04
373.28
1035.30
1304.60

(v)


MS2 of ThcE4/PatD
GGC*ESC*SYEGDEAE
Relative Abundance
m/z
1053.29
329.10
682.37
1234.38
862.24
520.18
617.79
812.29
1163.34
733.20
906.32
219.13
347.23
1007.43
1252.41

(vi)


MS2 of ThcE4/PatD
ASSCDCSLY
Relative Abundance
m/z
749.22
767.20
930.33
731.25
846.65
713.27
660.19
912.29
375.26
446.22
513.23
549.27
618.25
837.73
894.30

(vii)

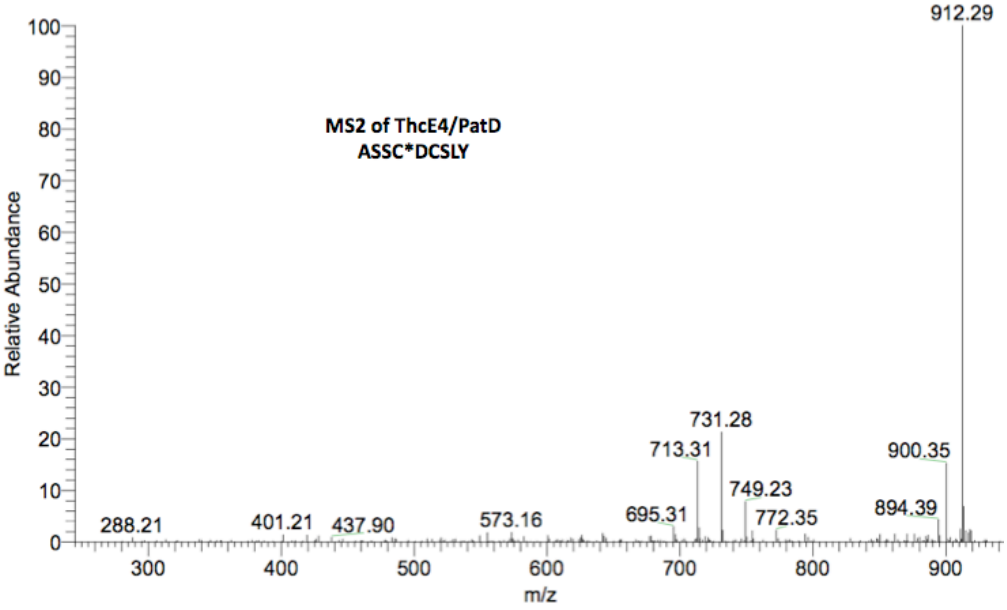

MS2 of ThcE4/PatD
ASSC*DCSLY
Relative Abundance
m/z
912.29
731.28
713.31
900.35
749.23
894.39
772.35
695.31
288.21
401.21
437.90
573.16

Table S3. Confirmation of dehydration as thiazoline modification of ThcE4 (**1**)/PatD reaction shown in Figure S21-C.

Table S3-a

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGCESCSYEGDEAE	-	-	-	-
GG	-	-	-	-	GCESCSYEGDEAE	-	-	-	-
GGC	-	-	-	-	CESCSYEGDEAE	-	-	-	-
GGCE	347.20	-	-	-	ESCSYEGDEAE	-	-	-	-
GGCES	537.16	-	-	-	SCSYEGDEAE	1089.37	-	-	1071.33
GGCESC	-	-	-	-	CSYEGDEAE	1002.28	-	-	-
GGCESCS	-	-	-	606.32	SYEGDEAE	899.34	-	-	881.5
GGCESCSY	787.21	-	-	-	YEGDEAE	812.37	-	-	-
GGCESCSYE	916.26	-	-	-	EGDEAE	649.37	-	-	-
GGCESCSYEG	973.48	-	-	-	GDEAE	520.24	-	-	502.21
GGCESCSYEGD	-	-	-	-	DEAE	-	-	-	-
GGCESCSYEGDE	1217.42	-	-	1199.25	EAE	347.20	-	-	330.20
GGCESCSYEGDEA	1288.35	-	-	1270.39	AE	219.11	-	-	-
GGCESCSYEGDEAE	-	-	-	-	E	-	-	-	-

MS2 of ThcE4/PatD
GGCESCSYEGDEAE

Table S3-b

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGCESC*SYEGDEAE	-	-	-	-
GG	-	-	-	-	GCESC*SYEGDEAE	-	-	-	-
GGC	-	-	-	-	CESC*SYEGDEAE	-	-	-	-
GGCE	347.15	-	-	329.15	ESC*SYEGDEAE	-	-	-	-
GGCES	-	-	-	-	SC*SYEGDEAE	-	-	-	1053.25
GGCESC*	-	-	-	-	C*SYEGDEAE	-	-	-	-
GGCESC*S	-	-	-	588.31	SYEGDEAE	-	-	-	-
GGCESC*SY	-	-	-	751.22	YEGDEAE	812.33	-	-	-
GGCESC*SYE	-	-	-	880.26	EGDEAE	649.26	-	-	-
GGCESC*SYEG	-	-	-	-	GDEAE	520.27	-	-	502.29
GGCESC*SYEGD	-	-	-	-	DEAE	-	-	-	-
GGCESC*SYEGDE	-	-	-	1181.33	EAE	-	-	-	-
GGCESC*SYEGDEA	1270.43	-	626.79	1252.37	AE	-	-	-	-
GGCESC*SYEGDEAE	-	700.32	691.40	-	E	-	-	-	-

MS2 of ThcE4/PatD
GGCESC*SYEGDEAE

Table S3-c

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
G	-	-	-	-	GGC*ESC*SYEGDEAE	-	-	691.32	-
GG	-	-	-	-	GC*ESC*SYEGDEAE	-	-	-	-
GGC*	-	-	-	-	C*ESC*SYEGDEAE	-	-	-	-
GGC*E	329.1	-	-	-	ESC*SYEGDEAE	-	-	-	-
GGC*ES	-	-	-	-	SC*SYEGDEAE	-	-	-	1053.29
GGC*ESC*	-	-	-	-	C*SYEGDEAE	-	-	-	-
GGC*ESC*S	588.4	-	-	-	SYEGDEAE	-	-	-	-
GGC*ESC*SY	-	-	-	733.2	YEGDEAE	812.29	-	-	-
GGC*ESC*SYE	880.38	-	-	862.24	EGDEAE	649.24	-	-	-
GGC*ESC*SYEG	-	-	-	-	GDEAE	520.18	-	-	-
GGC*ESC*SYEGD	-	-	-	-	DEAE	-	-	-	-
GGC*ESC*SYEGDE	1181.42	-	-	1163.34	EAE	-	-	-	-
GGC*ESC*SYEGDEA	1252.41	-	617.79	1234.38	AE	219.13	-	-	-
GGC*ESC*SYEGDEAE	-	691.32	-	-	E	-	-	-	-

MS2 of ThcE4/PatD
GGC*ESC*SYEGDEAE

Table S3-d

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ASSCDCSLY	-	-	-	930.33
AS	-	-	-	-	SSCDCSLY	-	-	-	-
ASS	-	-	-	-	SCDCSLY	-	-	-	-
ASSC	-	-	-	-	CDCSLY	-	-	-	-
ASSCD	-	-	-	-	DCSLY	-	-	-	-
ASSCDC	-	-	-	549.27	CSLY	-	-	-	-
ASSCDCS	-	-	-	636.34	SLY	-	-	-	-
ASSCDCSL	767.27	-	-	749.22	LY	-	-	-	-
ASSCDCSLY	930.33	-	-	912.29	Y	-	-	-	-

MS2 of ThcE4/PatD
ASSCDCSLY

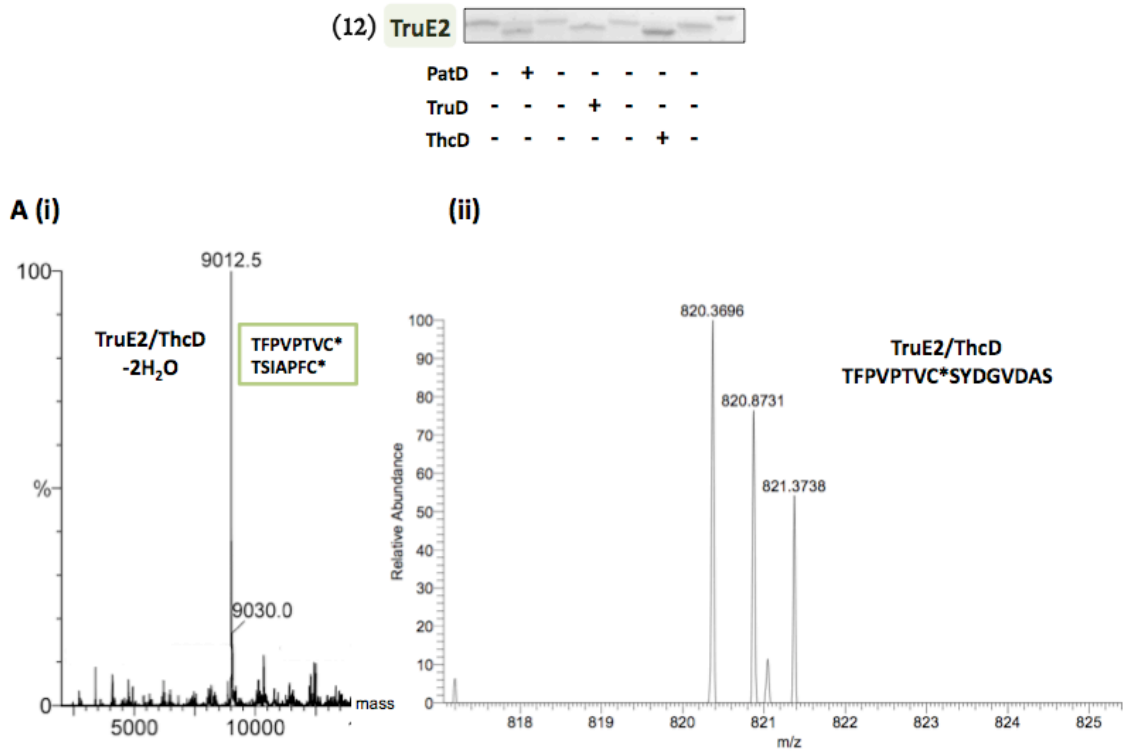
Table S3-e

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ASSC*DCSLY	-	-	-	930.33
AS	-	-	-	-	SSC*DCSLY	-	-	-	-
ASS	-	-	-	-	SC*DCSLY	-	-	-	-
ASSC*	-	-	-	-	C*DCSLY	-	-	-	-
ASSC*D	-	-	-	-	DCSLY	-	-	-	-
ASSC*DC	-	-	-	549.27	CSLY	-	-	-	-
ASSC*DCS	-	-	-	636.34	SLY	-	-	-	-
ASSC*DCSL	767.27	-	-	749.22	LY	-	-	-	-
ASSC*DCSLY	930.33	-	-	912.29	Y	-	-	-	-

MS2 of ThcE4/PatD
ASSC*DCSLY

Figure S22. (A) (i) Deconvoluted ESI-MS spectrum of 30 μM TruE2 (**12**) reaction with 0.5 μM ThcD (18 h), showing $[M+H]^+$ of substrate and its products, (ii) FTMS spectrum of the same (after PatA digest), with the $[M+2H]^{2+}$ mass corresponding to the sequence given, and (iii) MS/MS of the same. (B) ESI-MS spectrum of reaction of PatE (**13**) with ThcD using the same conditions as for TruE2/ThcD. SDS-PAGE visualization of reactions is also given. Thiazoline/oxazoline PTM is shown by C* and/or T*.

Table S4. Confirmation of dehydration as thiazoline modification which is shown by C*



(iii)

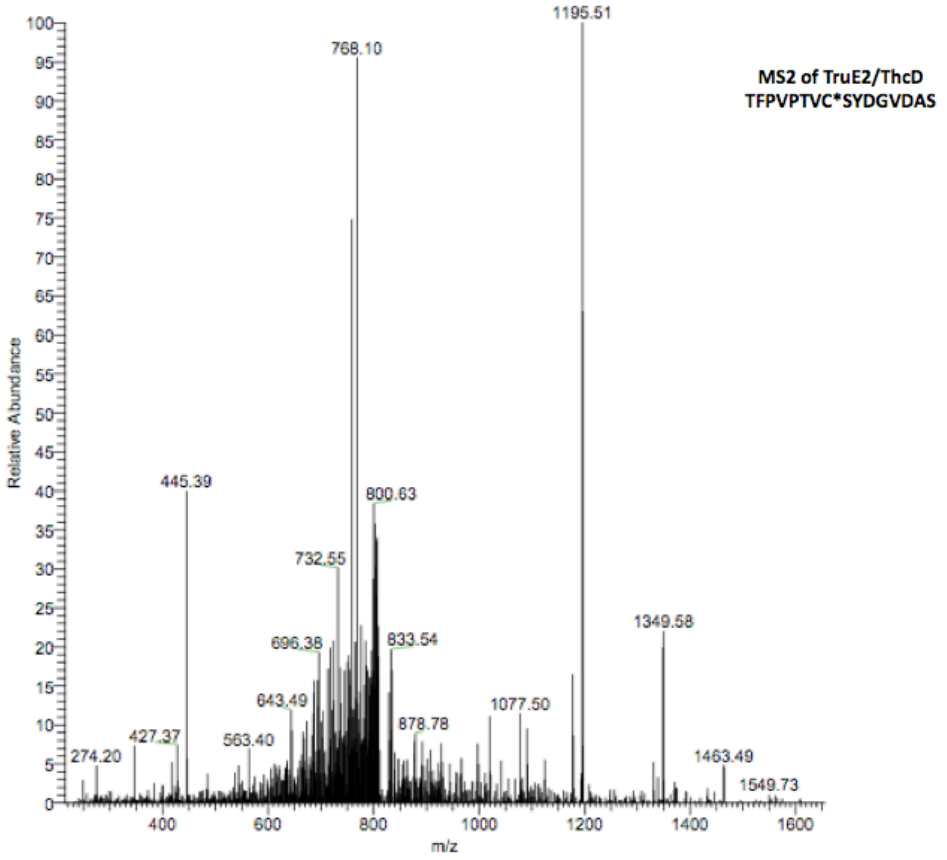


Table S4.

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
T	-	-	-	-	TFPVPTVC*SYDGVDAS	-	-	-	-
TF	-	-	-	-	FPVPTVC*SYDGVDAS	-	-	-	-
TFP	346.18	-	-	-	PVPTVC*SYDGVDAS	-	696.38	-	-
TFPV	445.39	-	-	427.37	VPTVC*SYDGVDAS	-	-	-	-
TFPVP	-	-	-	-	PTVC*SYDGVDAS	1195.51	-	-	1177.49
TFPVPT	643.49	-	-	-	TVC*SYDGVDAS	-	-	-	-
TFPVPTV	-	-	-	-	VC*SYDGVDAS	-	-	-	-
TFPVPTVC*	-	-	-	-	C*SYDGVDAS	-	-	-	-
TFPVPTVC*S	-	-	-	-	SYDGVDAS	-	-	-	-
TFPVPTVC*SY	1077.50	-	-	-	YDGVDAS	-	-	-	-
TFPVPTVC*SYD	-	-	-	-	DGVDAS	563.40	-	-	-
TFPVPTVC*SYDG	-	-	-	-	GVDAS	-	-	-	-
TFPVPTVC*SYDGV	1349.58	-	-	-	VDAS	-	-	-	-
TFPVPTVC*SYDGVD	1463.49	732.55	-	-	DAS	-	-	-	274.20
TFPVPTVC*SYDGVDA	-	768.10	759.08	-	AS	-	-	-	-
TFPVPTVC*SYDGVDAS	-	-	-	-	S	-	-	-	-

MS2 of TruE2/ThcD
TFPVPTVC*SYDGVDAS

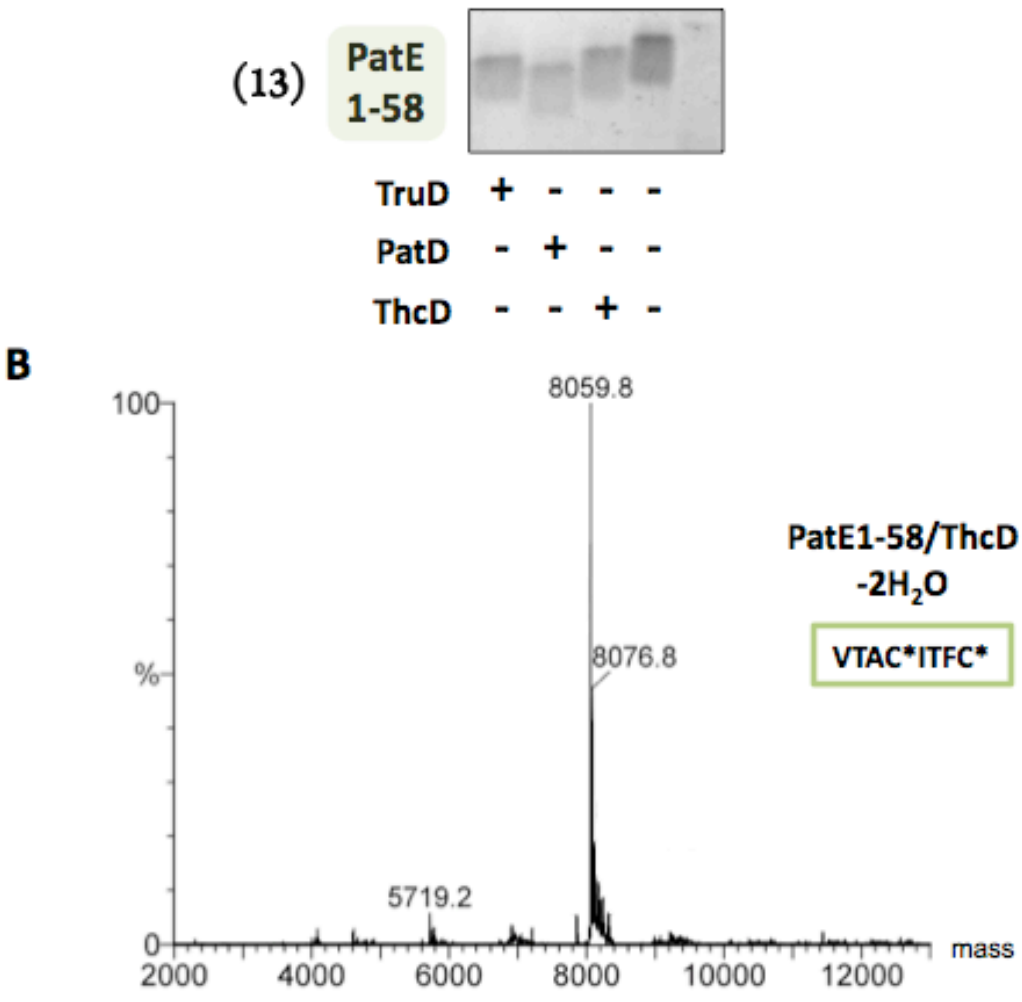
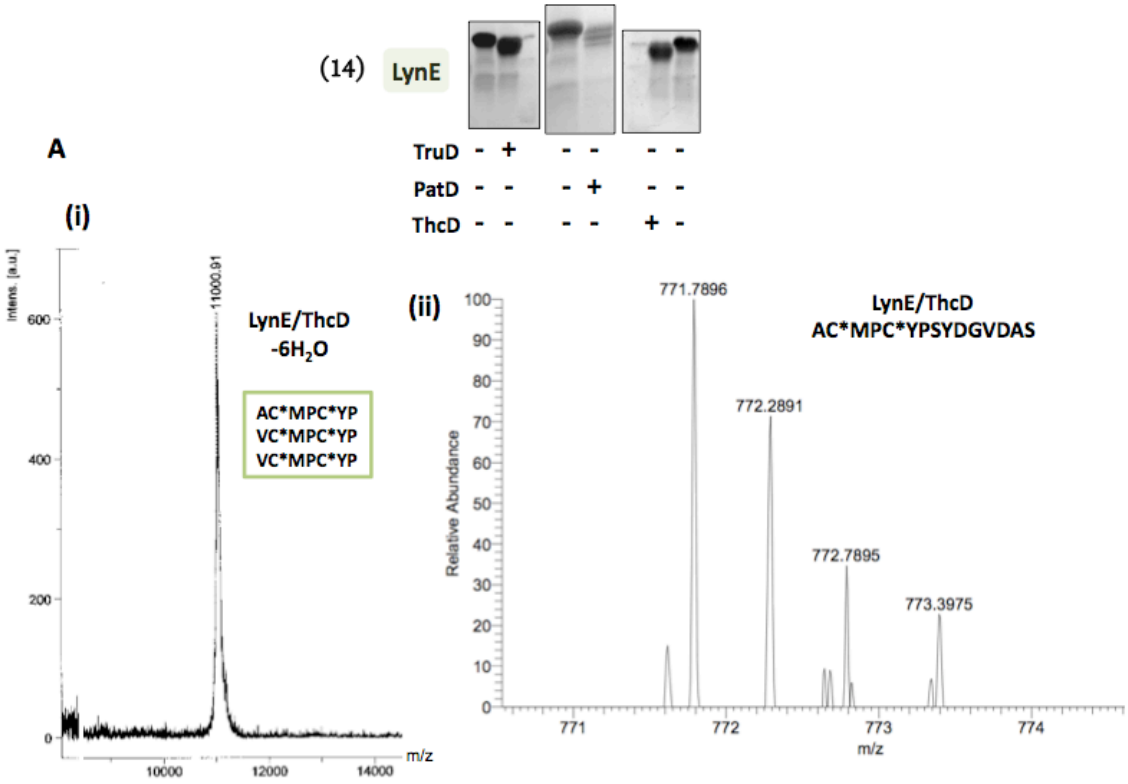

(13)
PatE
1-58
TruD + - - -
PatD - + - -
ThcD - - + -
B
100
%
0
8059.8
8076.8
5719.2
2000
4000
6000
8000
10000
12000
mass
PatE1-58/ThcD
-2H2O
VTAC*ITFC*

Figure S23. MALDI-MS of LynE (**14**) reaction with ThcD (A) and TruD (B), showing $[M+H]^+$ product peak. The core sequence of LynE is given in green boxes. The confirmation of dehydration as heterocycle modification is provided by FTMS and MS/MS spectra of PatA-digested fragments, showing $[M+2H]^{2+}$ of parent ions. Reactions were carried out with 30 μM substrate with 0.5 μM enzyme (18 h). Thiazoline modification is shown by **C***. SDS-PAGE visualization of same reactions is also shown.

Table S5-6. Confirmation of dehydrated products as thiazoline modification of reactions given in Figure S23.



(iii)

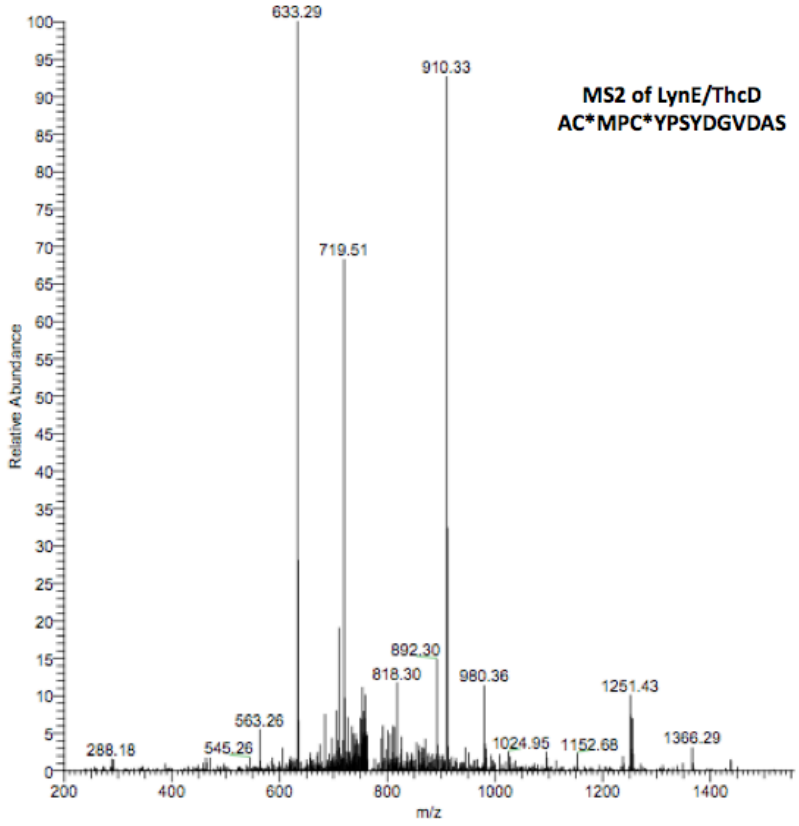
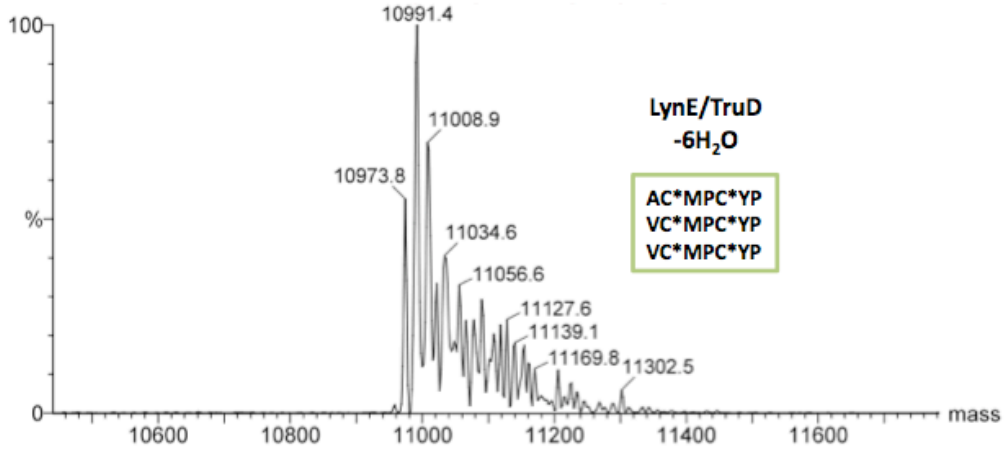

MS2 of LynE/ThcD
AC*MPC*YPSYDGVDAS
Relative Abundance
m/z
633.29
910.33
719.51
892.30
818.30
980.36
1251.43
563.26
545.26
288.18
1024.95
1152.68
1366.29

Table S5

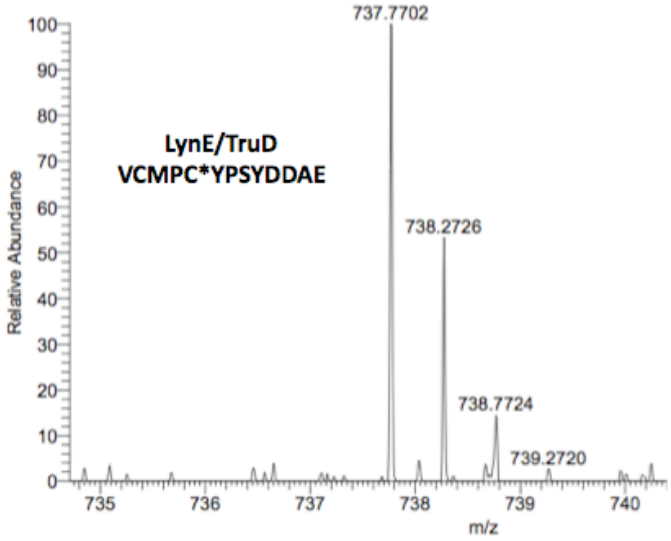
Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	AC*MPC*YPSYDGVDAS	-	-	-	-
AC*	-	-	-	-	C*MPC*YPSYDGVDAS	-	-	-	-
AC*M	288.18	-	-	-	MPC*YPSYDGVDAS	-	-	684.24	-
AC*MP	-	-	-	-	PC*YPSYDGVDAS	-	-	-	-
AC*MPC*	-	-	-	-	C*YPSYDGVDAS	-	-	-	-
AC*MPC*Y	633.29	-	-	-	YPSYDGVDAS	-	-	-	-
AC*MPC*YP	-	-	-	-	PSYDGVDAS	910.33	-	-	892.30
AC*MPC*YPS	818.30	-	-	-	SYDGVDAS	-	-	-	-
AC*MPC*YPSY	980.36	-	-	-	YDGVDAS	-	-	-	-
AC*MPC*YPSYD	-	-	-	-	DGVDAS	-	-	-	545.26
AC*MPC*YPSYDG	1152.68	-	-	-	GVDAS	563.26	-	-	-
AC*MPC*YPSYDGV	1251.43	-	-	-	VDAS	-	-	-	-
AC*MPC*YPSYDGVD	1366.29	-	-	-	DAS	-	-	-	-
AC*MPC*YPSYDGVDA	-	-	-	-	AS	-	-	-	-
AC*MPC*YPSYDGVDAS	-	719.51	710.39	-	S	-	-	-	-

MS2 of LynE/ThcD
AC*MPC*YPSYDGVDAS

B (i)


10991.4
11008.9
10973.8
11034.6
11056.6
11127.6
11139.1
11169.8
11302.5
100
%
0
10600
10800
11000
11200
11400
11600
mass
LynE/TruD
-6H2O
AC*MPC*YP
VC*MPC*YP
VC*MPC*YP

(ii)


737.7702
738.2726
738.7724
739.2720
LynE/TruD
VCMPC*YPSYDDAE
Relative Abundance
100
90
80
70
60
50
40
30
20
10
0
735
736
737
738
739
740
m/z

(iii)

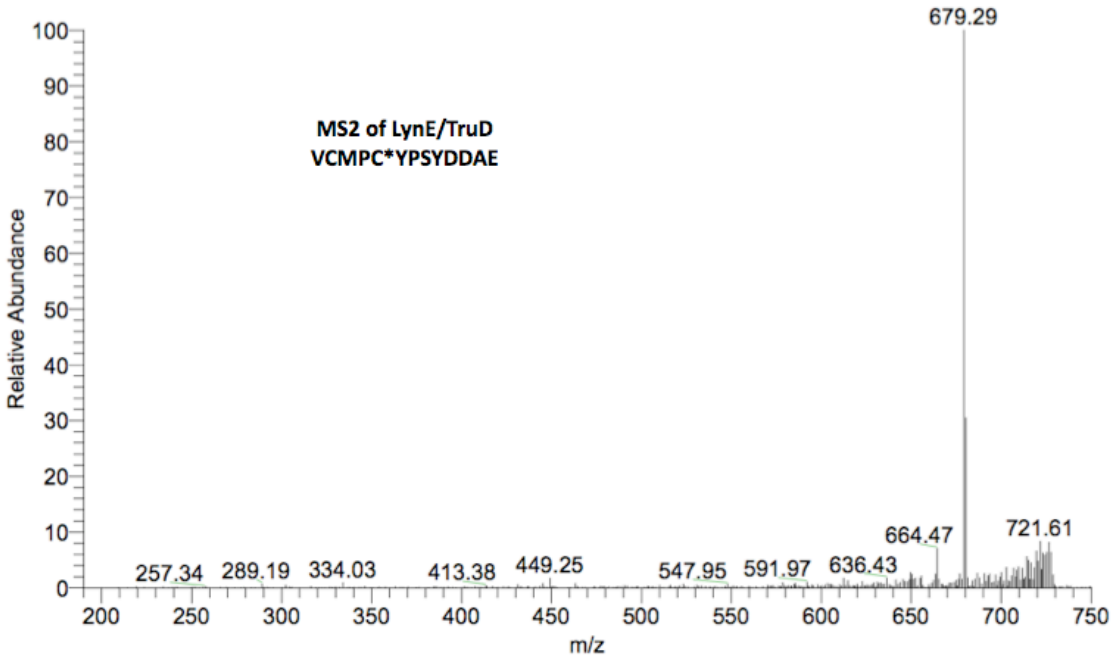


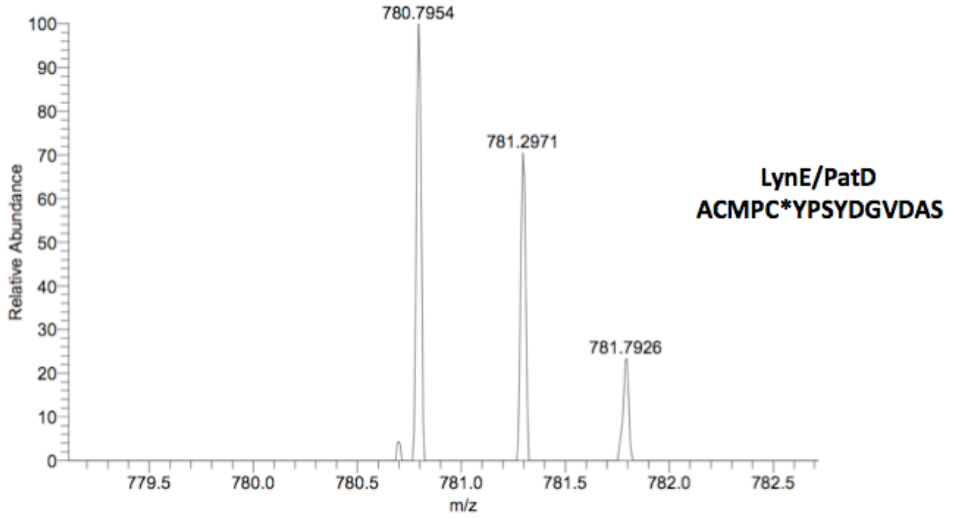
Table S6

MS2 of LynE/TruD
VCMPC*YPSYDDAE

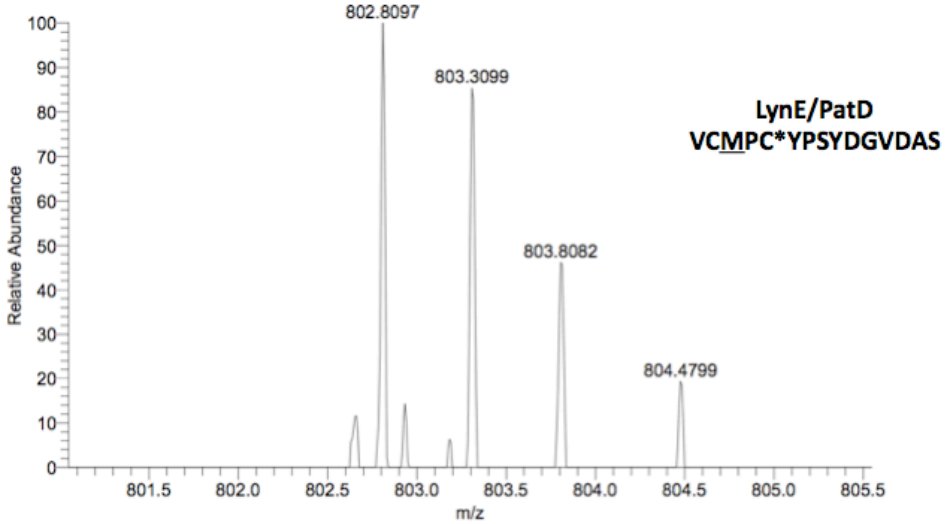
Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
V	-	-	-	-	VCMPC*YPSYDDAE	-	-	-	-
VC	-	-	-	-	CMPC*YPSYDDAE	-	-	679.29	-
VCM	334.21	-	-	-	MPC*YPSYDDAE	-	636.43	-	-
VCMP	-	-	-	-	PC*YPSYDDAE	-	-	-	-
VCMPC*	-	-	-	-	C*YPSYDDAE	-	-	-	-
VCMPC*Y	-	-	-	-	YPSYDDAE	-	-	-	-
VCMPC*YP	-	-	-	-	PSYDDAE		-	-	-
VCMPC*YPS	863.32	-	-	-	SYDDAE	-	-	-	-
VCMPC*YPSY	1026.41	-	-	-	YDDAE	-	-	-	-
VCMPC*YPSYD	1141.30	-	-	-	DDAE	449.25	-	-	-
VCMPC*YPSYDD	1256.40	-	-	-	DAE	-	-	-	-
VCMPC*YPSYDDA	1327.35	664.48	-	-	AE	-	-	-	-
VCMPC*YPSYDDAE	-	-	-	-	E	-	-	-	-

Figure S24. FTMS of LynE (**14**) reaction with PatD after PatA digest, showing $[M+2H]^{2+}$ parent ion peaks. Reactions were carried out with 30 μM substrate with 0.5 μM enzyme (18 h). Thiazoline modification is shown by **C*** and methionine oxidation is shown by underlined M.

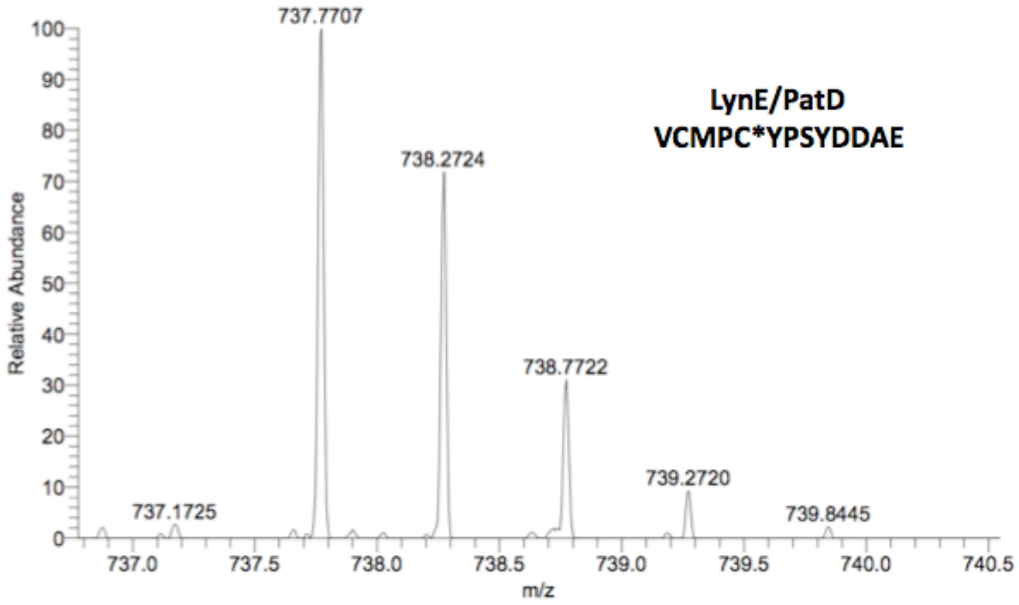
Table S7. Confirmation of dehydrated products as thiazoline modification of reactions given in Figure S24.



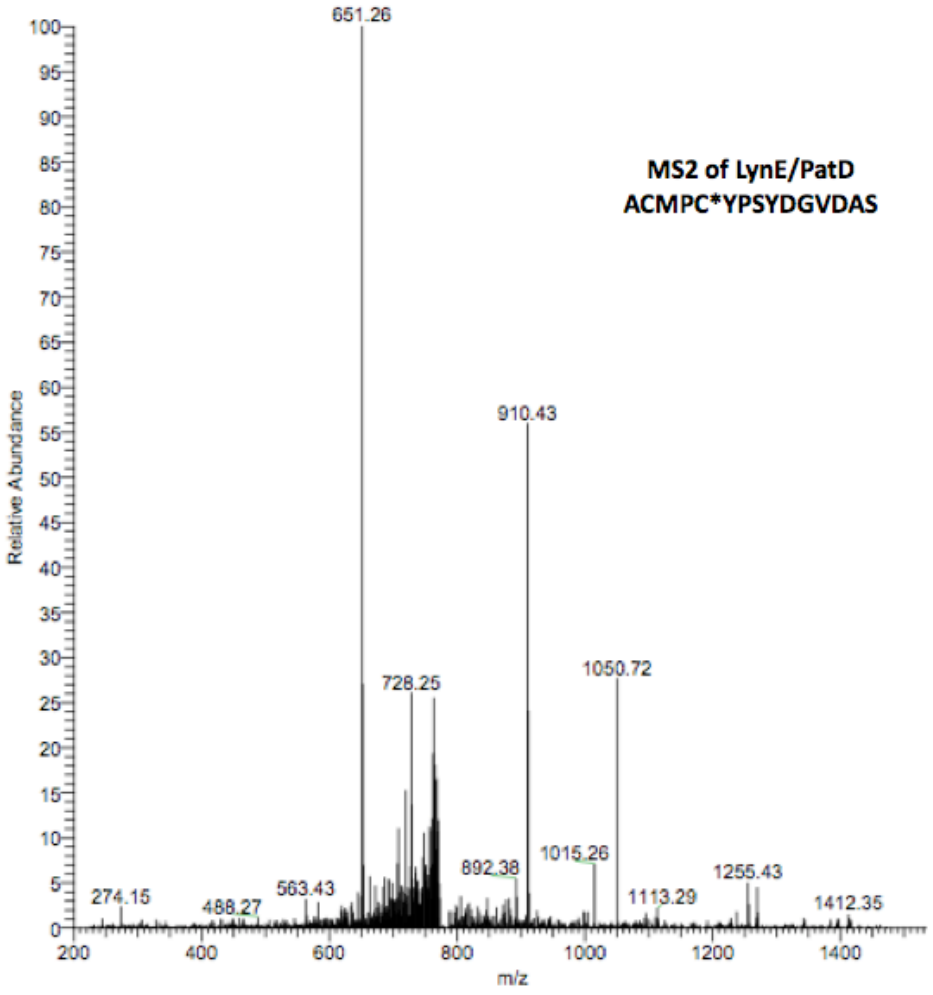
(ii)



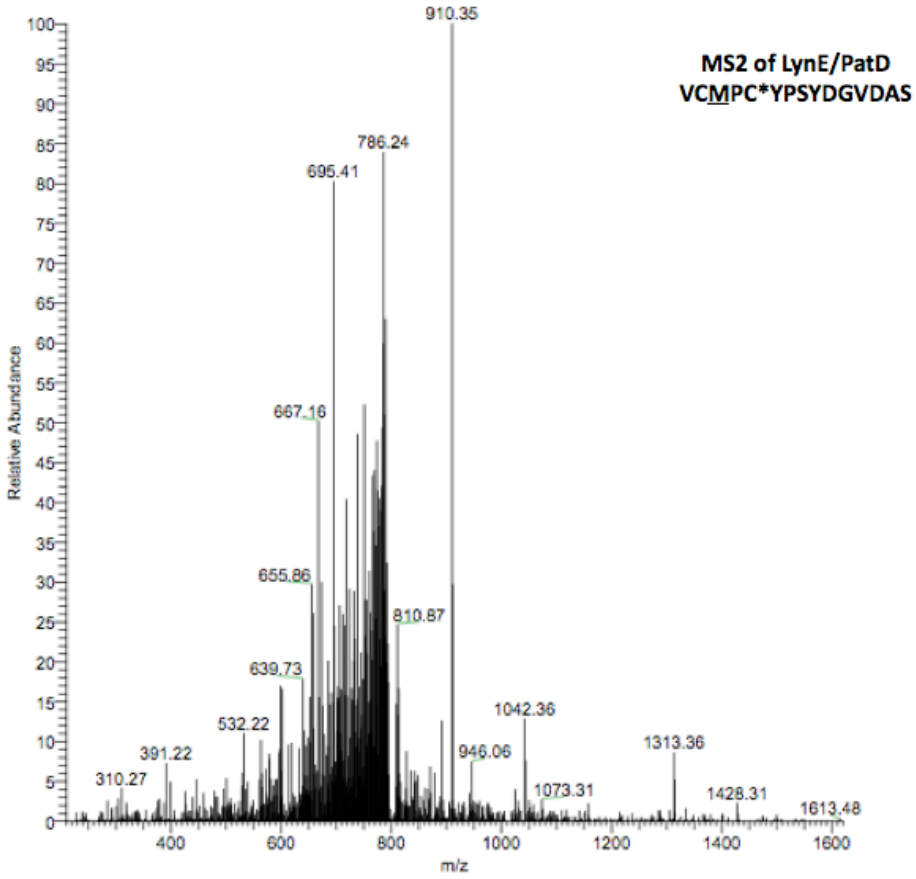
(iii)


737.7707
738.2724
738.7722
739.2720
737.1725
739.8445
LynE/PatD
VCMPC*YPSYDDAE
Relative Abundance
m/z
100
90
80
70
60
50
40
30
20
10
0
737.0
737.5
738.0
738.5
739.0
739.5
740.0
740.5

(iv)


MS2 of LynE/PatD
ACMPC*YPSYDGVDAS
Relative Abundance
m/z
651.26
910.43
728.25
1050.72
1015.26
892.38
1255.43
563.43
274.15
488.27
1113.29
1412.35

(v)


MS2 of LynE/PatD
VCMPC*YPSYDGVDAS
910.35
786.24
695.41
667.16
655.86
810.87
639.73
532.22
391.22
310.27
1042.36
946.06
1313.36
1073.31
1428.31
1613.48
Relative Abundance
m/z

(vi)

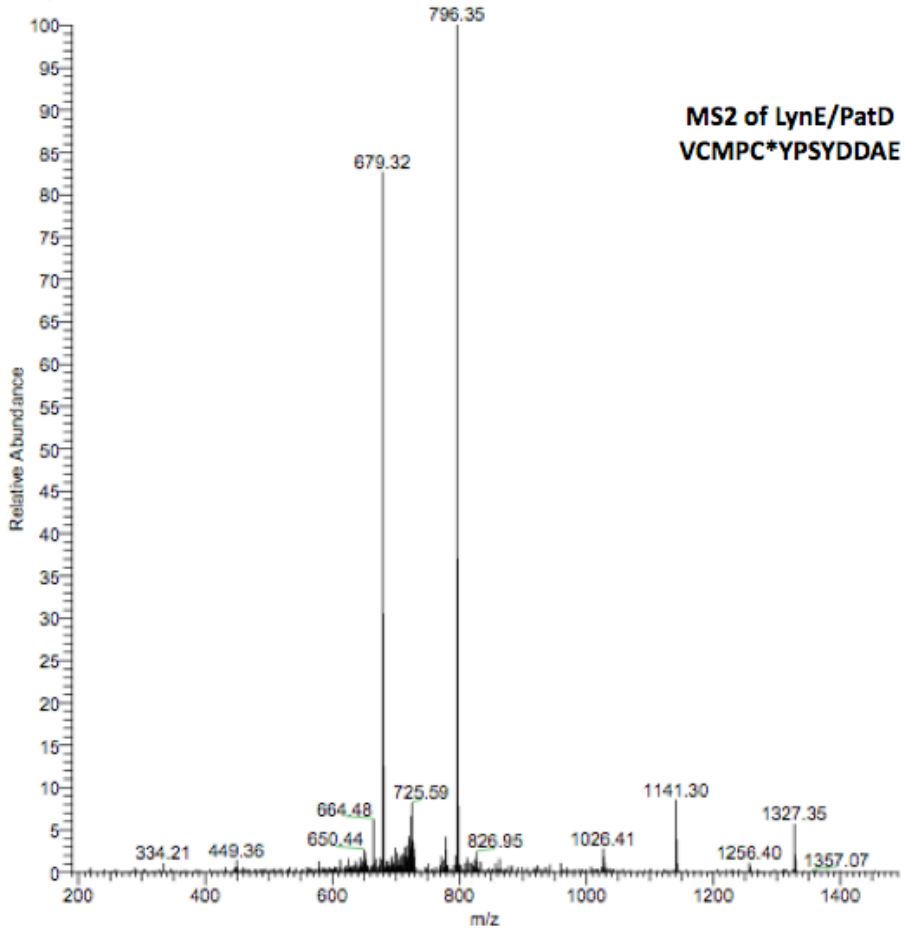

MS2 of LynE/PatD
VCMPC*YPSYDDAE
796.35
679.32
725.59
664.48
650.44
1141.30
1327.35
826.95
1026.41
1256.40
1357.07
334.21
449.36
Relative Abundance
m/z

Table S7-a

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ACMPC*YPSYDGVDAS	-	-	-	-
AC	-	-	-	-	CMPC*YPSYDGVDAS	-	-	-	-
ACM	-	-	-	-	MPC*YPSYDGVDAS	-	-	-	-
ACMP	-	-	-	-	PC*YPSYDGVDAS	1255.43	-	-	-
ACMPC*	-	-	-	-	C*YPSYDGVDAS	-	-	-	-
ACMPC*Y	651.26	-	-	-	YPSYDGVDAS	-	-	-	-
ACMPC*YP	-	-	-	-	PSYDGVDAS	910.43	-	-	-
ACMPC*YPS	-	-	-	-	SYDGVDAS	-	-	-	-
ACMPC*YPSY	-	-	-	-	YDGVDAS	-	-	-	-
ACMPC*YPSYD	1113.29	-	-	-	DGVDAS	-	-	-	-
ACMPC*YPSYDG	-	-	-	-	GVDAS	563.43	-	-	-
ACMPC*YPSYDGV	-	-	-	-	VDAS	488.27	-	-	-
ACMPC*YPSYDGVD	1269.47	-	-	-	DAS	-	-	-	-
ACMPC*YPSYDGVDA	-	-	-	-	AS	-	-	-	274.15
ACMPC*YPSYDGVDAS	-	728.25	719.55	-	S	-	-	-	-

MS2 of LynE/PatD
ACMPC*YPSYDGVDAS

Table S7-b

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
V	-	-	-	-	VCMPC*YPSYDGVDAS	-	-	-	-
VC	-	-	-	-	CMPC*YPSYDGVDAS	-	-	-	-
VCM	-	-	-	-	MPC*YPSYDGVDAS	-	-	-	-
VCMP	-	-	-	-	PC*YPSYDGVDAS	-	-	-	-
VCMPC*	532.22	-	-	-	C*YPSYDGVDAS	-	-	-	-
VCMPC*Y	695.41	-	-	-	YPSYDGVDAS	1073.31	-	-	-
VCMPC*YP	-	-	-	-	PSYDGVDAS	910.35	-	-	892.39
VCMPC*YPS	-	-	-	-	SYDGVDAS	813.33	-	-	-
VCMPC*YPSY	1042.36	-	-	1024.49	YDGVDAS	-	-	-	-
VCMPC*YPSYD	1157.55	-	-	-	DGVDAS	-	-	-	-
VCMPC*YPSYDG	1214.54	-	-	-	GVDAS	-	-	-	-
VCMPC*YPSYDGV	1313.36	-	-	-	VDAS	391.22	-	-	-
VCMPC*YPSYDGVD	1428.31	-	-	-	DAS	-	-	-	-
VCMPC*YPSYDGVDA	1499.63	750.68	-	-	AS	-	-	-	-
VCMPC*YPSYDGVDAS	-	-	-	-	S	-	-	-	-

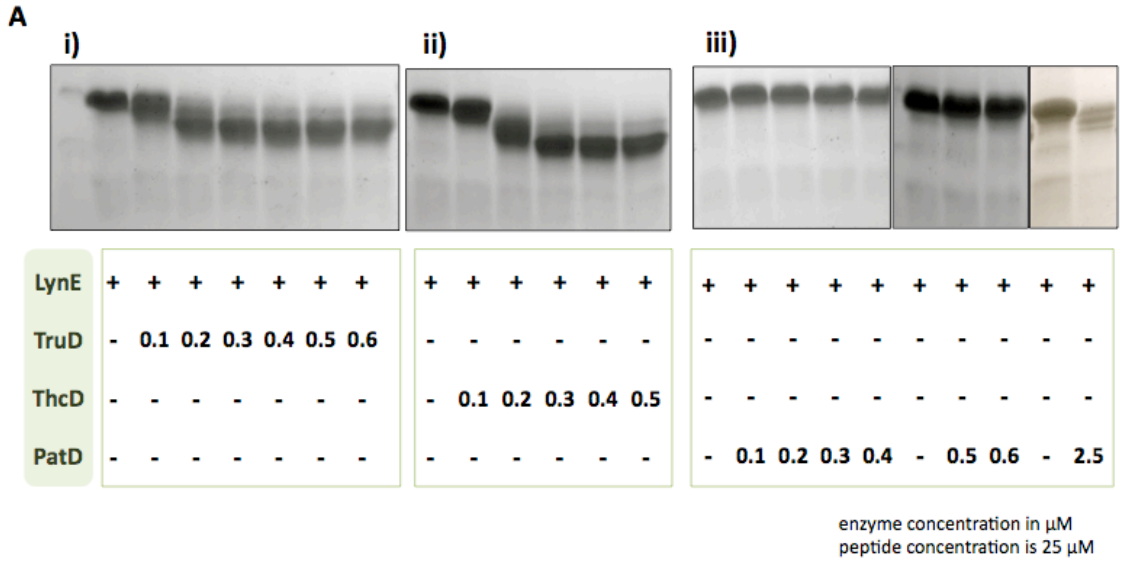
MS2 of LynE/PatD
VCMPC*YPSYDGVDAS

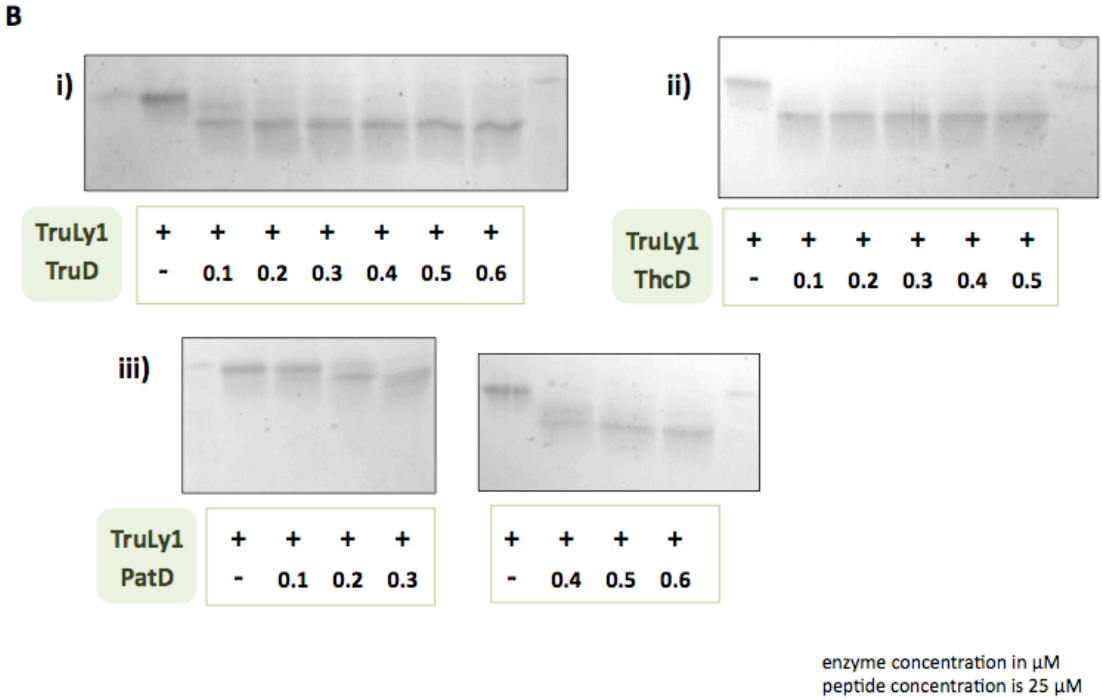
Table S7-c

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
V	-	-	-	-	VCMPC*YPSYDDAE	-	-	-	-
VC	-	-	-	-	CMPC*YPSYDDAE	-	-	679.32	-
VCM	334.21	-	-	-	MPC*YPSYDDAE	-	-	-	-
VCMP	-	-	-	-	PC*YPSYDDAE	-	-	-	-
VCMPC*	-	-	-	-	C*YPSYDDAE	-	-	-	-
VCMPC*Y	-	-	-	-	YPSYDDAE	-	-	-	-
VCMPC*YP	-	-	-	-	PSYDDAE	796.35	-	-	-
VCMPC*YPS	863.32	-	-	-	SYDDAE	-	-	-	-
VCMPC*YPSY	1026.41	-	-	-	YDDAE	-	-	-	-
VCMPC*YPSYD	1141.30	-	-	-	DDAE	449.36	-	-	-
VCMPC*YPSYDD	1256.40	-	-	-	DAE	334.21	-	-	-
VCMPC*YPSYDDA	1327.35	664.48	-	-	AE	-	-	-	-
VCMPC*YPSYDDAE	-	-	-	-	E	-	-	-	-

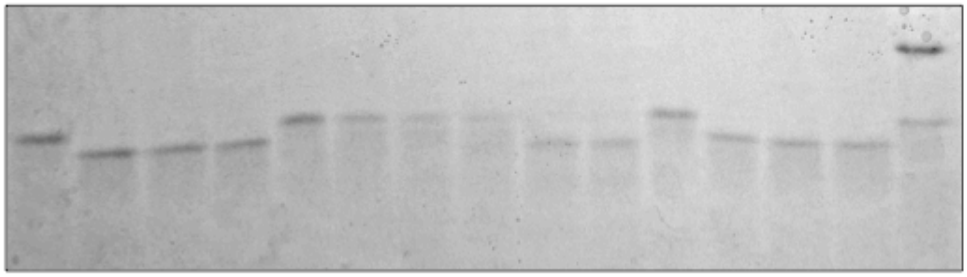
MS2 of LynE/PatD
VCMPC*YPSYDDAE

Figure S25. Comparison of reaction completion at an 18-h time-point with varying enzyme concentrations for peptides (A) LynE (**14**) (B) TruLy1 (**2**) and (C) TruE2 (**12**).





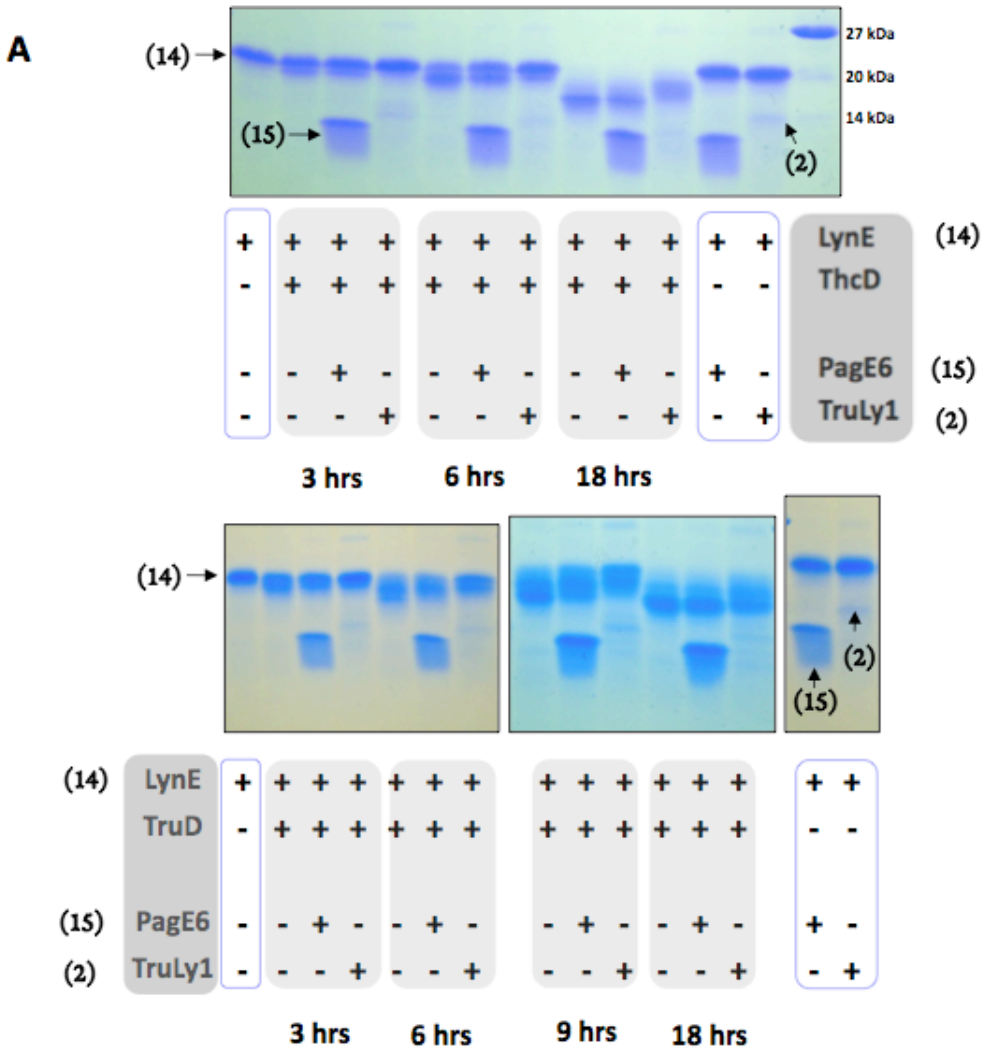
C



TruE2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TruD	-	0.1	0.2	0.3	-	-	-	-	-	-	-	-	-	-
PatD	-	-	-	-	-	0.1	0.2	0.3	0.4	0.5	-	-	-	-
ThcD	-	-	-	-	-	-	-	-	-	-	-	0.1	0.2	0.3

enzyme concentration in µM
peptide concentration is 25 µM

Figure S26. Competition reactions of LynE (**14**) by ThcD (top) and TruD (bottom) with substrates PagE6 (**15**), which lacks RSI and TruLy1 (**2**), which contains RSI. As seen by the reduction in the LynE band-shift, only **2** could inhibit LynE modification while **15** did not have much effect. Since **15** also lacked heterocyclizable residues along with RSI, it was hard to distinguish if lack of competition by **15** was because of the absence of RSI or because of the absence of binding to the core peptide. Hence, we carried out similar reactions with TruLy2 (**3**) and its RSI mutant **8** that differ only within the RSI sequence as shown in (B). All substrates were maintained at a constant concentration of 25 μM with 1 μM enzyme in each reaction. Blue boxes correspond to control bands of substrates only with no enzyme.



B

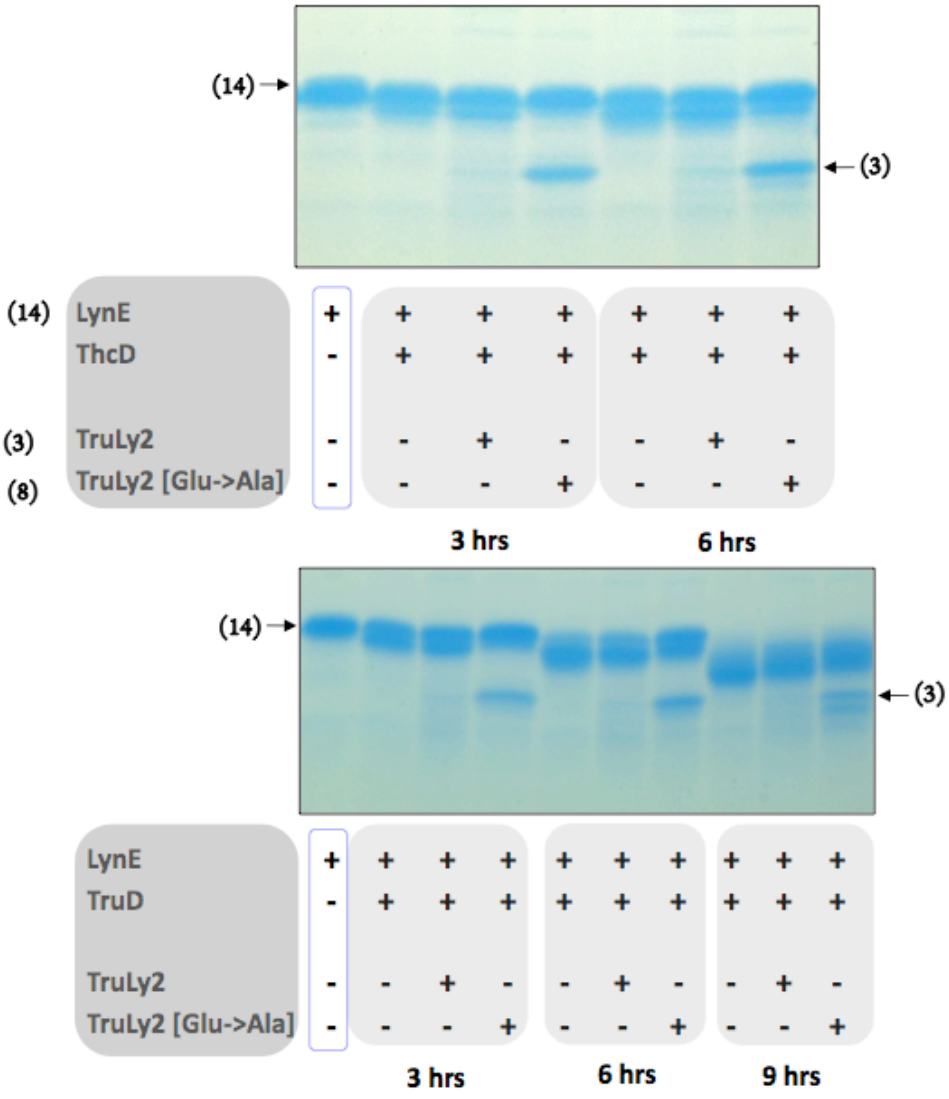

(14)
(3)
(14) LynE + + + + + + + + +
ThcD - + + + + + + + +
(3) TruLy2 - - + - - + -
(8) TruLy2 [Glu->Ala] - - - + - - +
3 hrs
6 hrs
(14)
(3)
LynE + + + + + + + + + +
TruD - + + + + + + + + +
TruLy2 - - + - - + - - + -
TruLy2 [Glu->Ala] - - - + - - + - - +
3 hrs
6 hrs
9 hrs

Figure S27. ESI-MS confirmation of the SDS-PAGE band-shifts for competition reactions of LynE (**14**) with TruLy2 mutant (**8**) (3[Glu->Ala]) and TruLy2 (**3**) is shown, along with the control reaction of **14**/ThcD in the absence of competing substrates. All mass peaks correspond to $[M+H]^+$. Substrate **8** with the mutated RSI failed to compete with LynE modification; such the ESI-MS looks about the same as control reaction. Substrate **3** inhibited the same, such that the predominant peak was only one 18 Da mass shift compared to control. Simultaneously, modification of **3** was also observed.

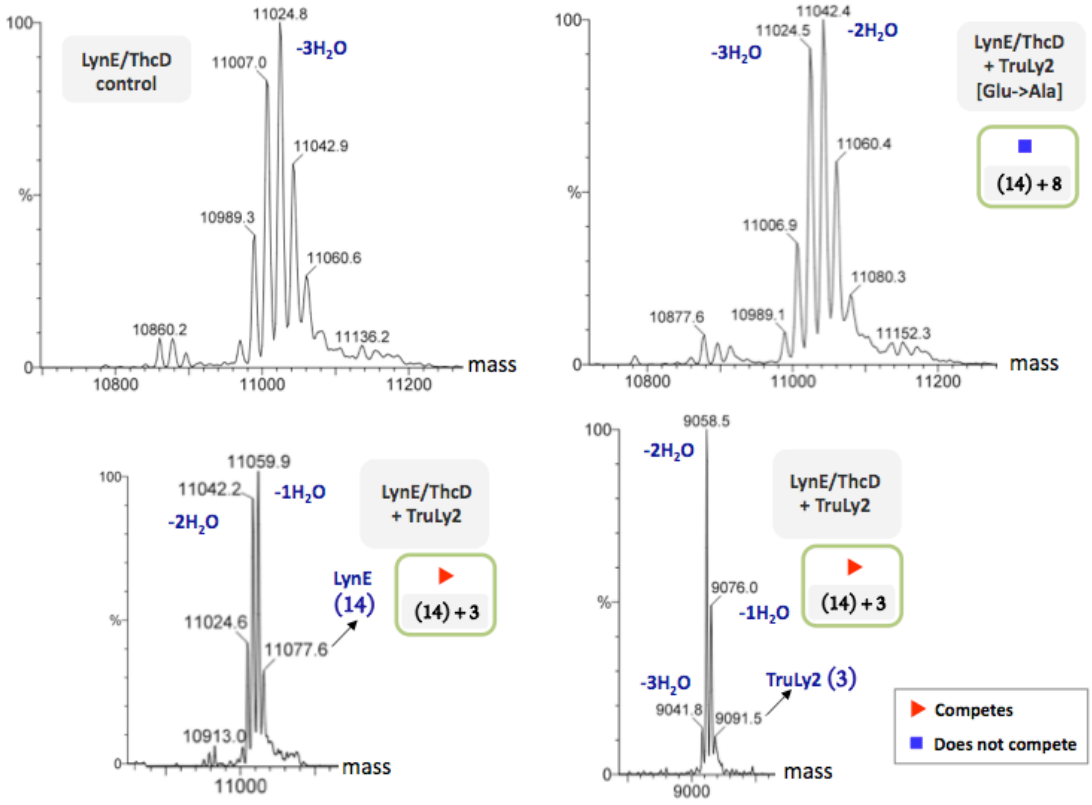
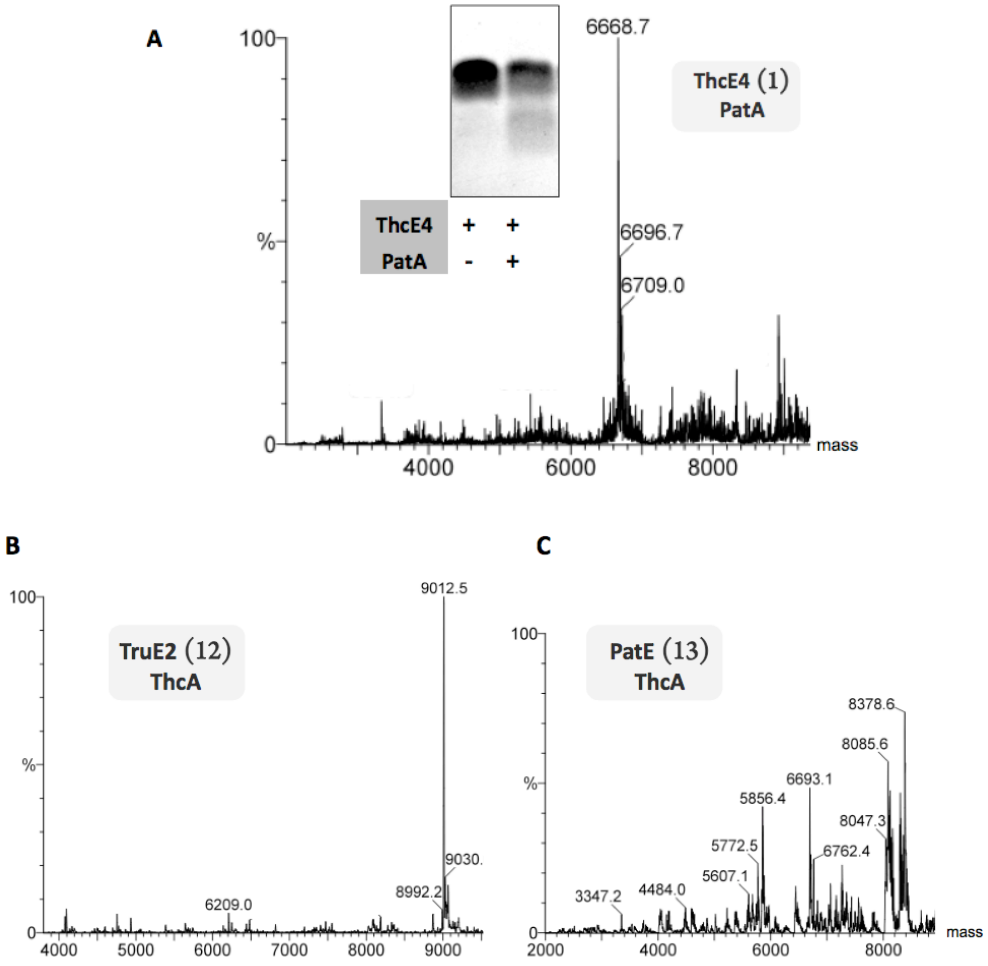


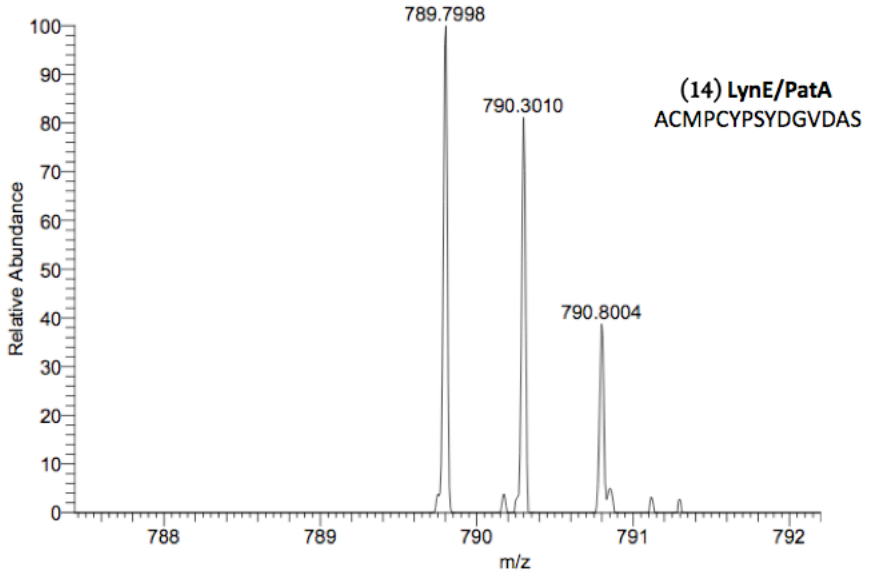
Figure S28. ESI-MS spectra of proteolysis cross-reactions of (A) ThcE4 (**1**)/PatA, (B) TruE2 (**12**)/ThcA, (C) PatE (**13**)/ThcA, (D) LynE (**14**)/PatA, and (E) TruLy1 (**2**)/PatA. All mass peaks correspond to $[M+H]^+$. Detailed interpretation of mass spectra is given in the Supporting Information.

Table S8. Analysis of MS/MS of LynE/PatA reactions shown in Figure S27-D. Parent ion peaks correspond to $[M+2H]^{2+}$.



D

(i)


789.7998
790.3010
790.8004
(14) LynE/PatA
ACMPCYPSYDGVDAS
Relative Abundance
100
90
80
70
60
50
40
30
20
10
0
788
789
790
791
792
m/z

(ii)

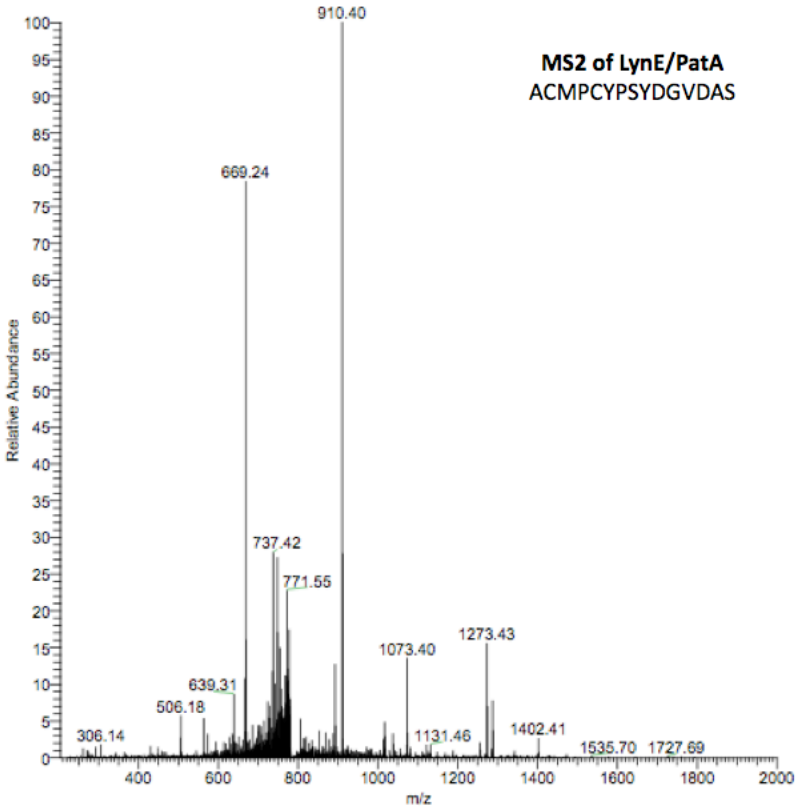
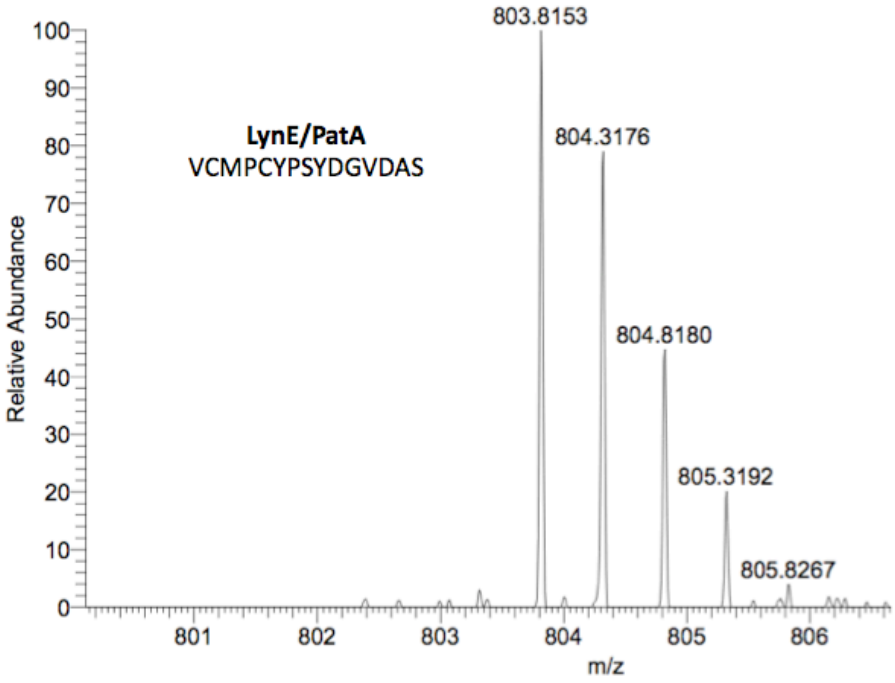


Table S8-a

Sequence	b	b^{++}	b^{++} -H_2O	b -H_2O	Sequence	y	y^{++}	y^{++} -H_2O	y -H_2O
A	-	-	-	-	ACMPCYPSYDGVDAS	-	-	-	-
AC	-	-	-	-	CMPCYPSYDGVDAS	-	-	-	-
ACM	306.14	-	-	-	MPCYPSYDGVDAS	-	-	-	-
ACMP	-	-	-	-	PCYPSYDGVDAS	1273.43	-	-	-
ACMPC	506.18	-	-	-	CYPSYDGVDAS	-	-	-	-
ACMPCY	669.24	-	-	-	YPSYDGVDAS	1073.40	-	-	-
ACMPCYP	-	-	-	748.15	PSYDGVDAS	910.40	-	-	892.40
ACMPCYPS	-	-	-	-	SYDGVDAS	-	-	-	-
ACMPCYPSY	1016.40	-	-	-	YDGVDAS	-	-	-	-
ACMPCYPSYD	1131.46	-	-	-	DGVDAS	-	-	-	-
ACMPCYPSYDG	1188.31	-	-	-	GVDAS	-	-	-	-
ACMPCYPSYDGV	1287.37	-	-	-	VDAS	-	-	-	-
ACMPCYPSYDGVD	1402.41	-	-	-	DAS	-	-	-	-
ACMPCYPSYDGVDA	1473.46	737.42	-	-	AS	-	-	-	-
ACMPCYPSYDGVDAS	-	-	771.55	-	S	-	-	-	-

MS2 of LynE/PatA
ACMPCYPSYDGVDAS

iii)


803.8153
804.3176
804.8180
805.3192
805.8267
LynE/PatA
VCMPCYPSYDGVDAS
Relative Abundance
m/z
100
90
80
70
60
50
40
30
20
10
0
801
802
803
804
805
806

iv)

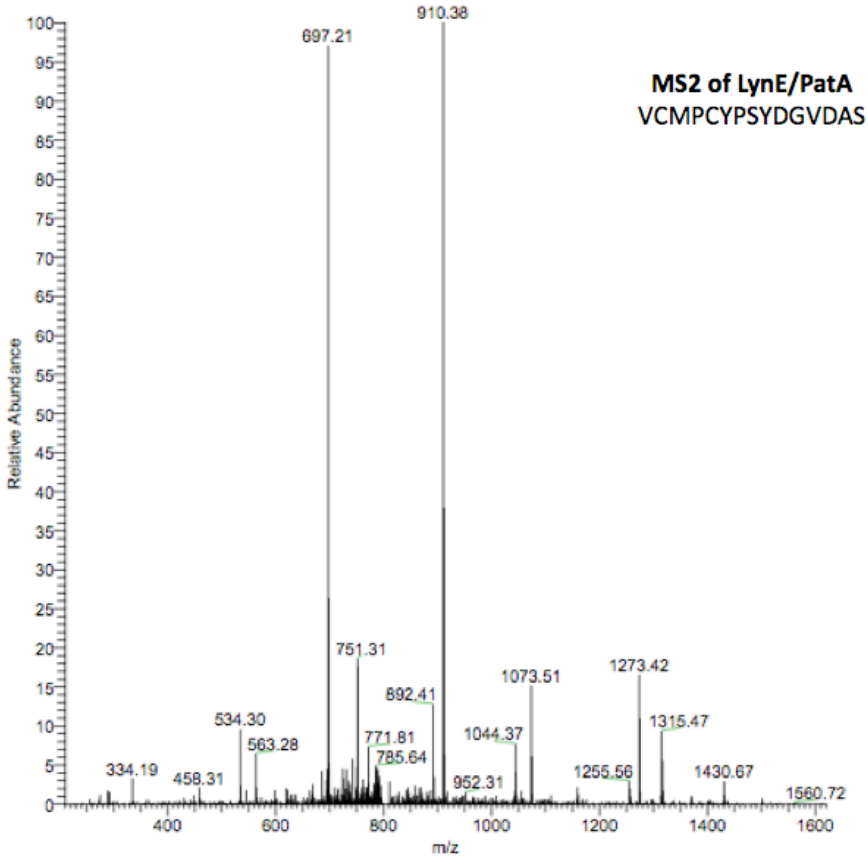
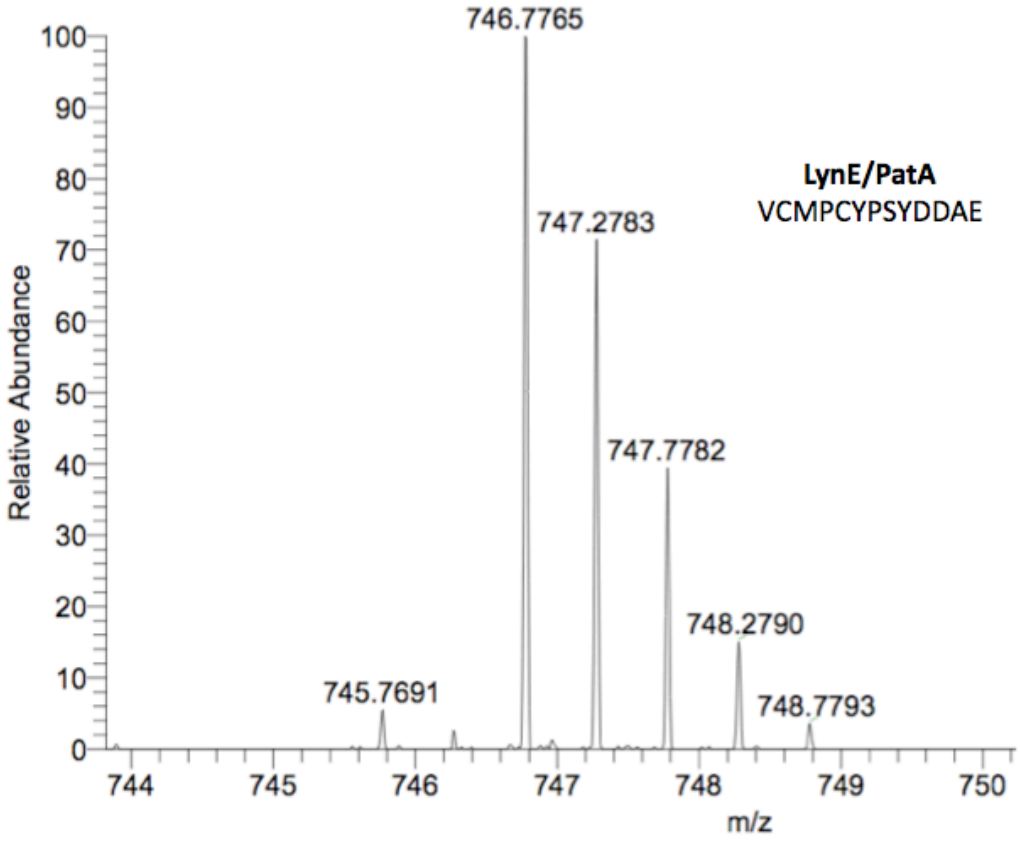


Table S8-b

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
V	-	-	-	-	VCMPCYPSYDGVDAS	-	-	-	-
VC	-	-	-	-	CMPCYPSYDGVDAS	-	-	-	-
VCM	**334.19**	-	-	-	MPCYPSYDGVDAS	-	-	-	-
VCMP	-	-	-	-	PCYPSYDGVDAS	**1273.42**	-	-	**1255.56**
VCMPC	**534.30**	-	-	-	CYPSYDGVDAS	-	-	-	-
VCMPCY	**697.21**	-	-	-	YPSYDGVDAS	**1073.51**	-	-	-
VCMPCYP	-	-	-	-	PSYDGVDAS	**910.38**	-	-	**892.41**
VCMPCYPS	-	-	-	-	SYDGVDAS	-	-	-	-
VCMPCYPSY	**1044.37**	-	-	-	YDGVDAS	-	-	-	-
VCMPCYPSYD	-	-	-	-	DGVDAS	**563.28**	-	-	-
VCMPCYPSYDG	-	-	-	-	GVDAS	-	-	-	-
VCMPCYPSYDGV	**1315.47**	-	-	-	VDAS	-	-	-	-
VCMPCYPSYDGVD	**1430.67**	-	-	-	DAS	-	-	-	-
VCMPCYPSYDGVDA	**1501.46**	**751.31**	-	-	AS	-	-	-	-
VCMPCYPSYDGVDAS	-	-	-	-	S	-	-	-	-

MS2 of LynE/PatA
VCMPCYPSYDGVDAS

v)


746.7765
747.2783
747.7782
748.2790
748.7793
745.7691
LynE/PatA
VCMPCYPSYDDAE
Relative Abundance
m/z
100
90
80
70
60
50
40
30
20
10
0
744
745
746
747
748
749
750

vi)

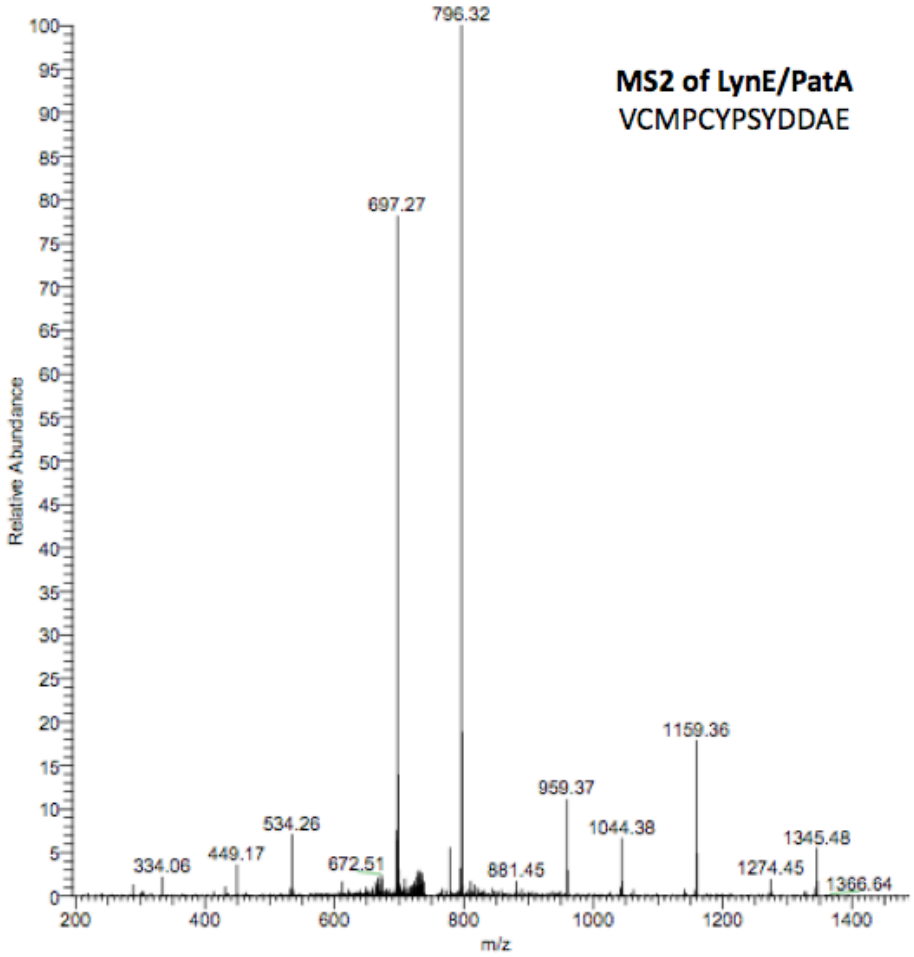


Table S8-c

Seq.	b	b^{++}	b^{++} -H_2O	b -H_2O	Seq.	y	y^{++}	y^{++} -H_2O	y -H_2O
V	-	-	-	-	VCMPCYPSYDDAE	-	-	-	-
VC	-	-	-	-	CMPCYPSYDDAE	-	697.27	-	-
VCM	334.06	-	-	-	MPCYPSYDDAE	-	-	-	-
VCMP	-	-	-	-	PCYPSYDDAE	1159.36	-	-	-
VCMP	534.26	-	-	-	CYPSYDDAE	-	-	-	1044.38
VCMPCY	697.27	-	-	-	YPSYDDAE	959.37	-	-	-
VCMPCYP	-	-	-	-	PSYDDAE	796.32	-	-	-
VCMPCYPS	881.45	-	-	-	SYDDAE	-	-	-	-
VCMPCYPSY	1044.38	-	-	-	YDDAE	-	-	-	-
VCMPCYPSYD	1159.36	-	-	-	DDAE	449.15	-	-	-
VCMPCYPSYDD	1274.45	-	-	-	DAE	334.06	-	-	-
VCMPCYPSYDDA	1345.48	-	-	-	AE	-	-	-	-
VCMPCYPSYDDAE	-	-	-	-	E	-	-	-	-

MS2 of LynE/PatA
VCMPCYPSYDDAE

E

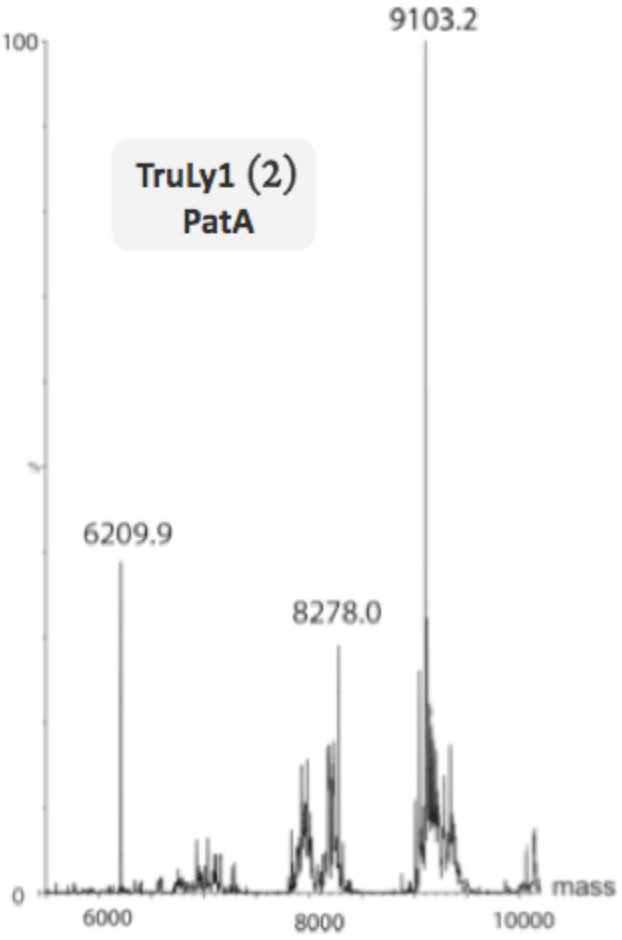
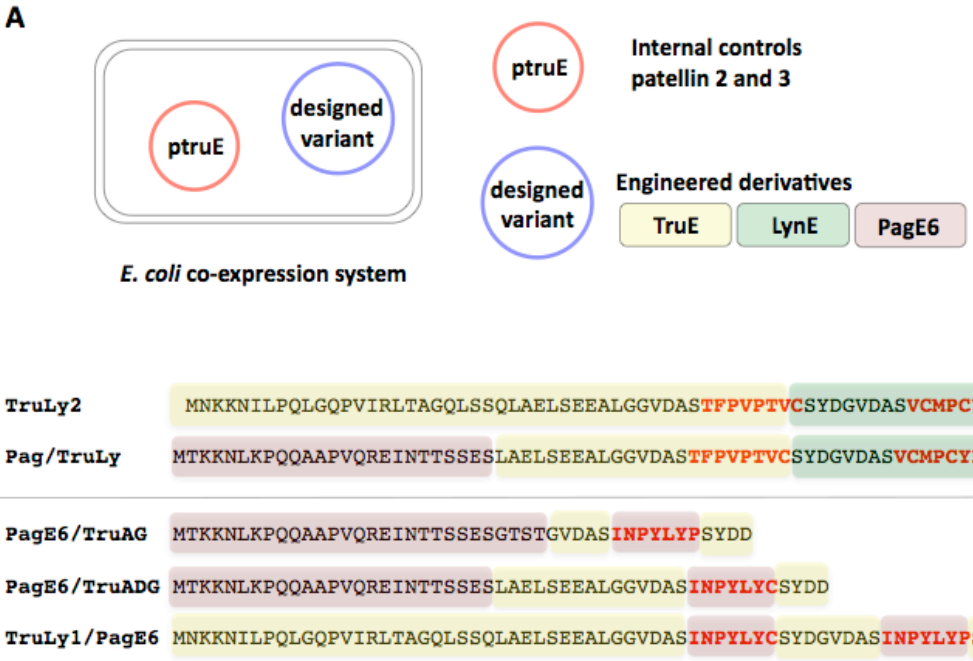
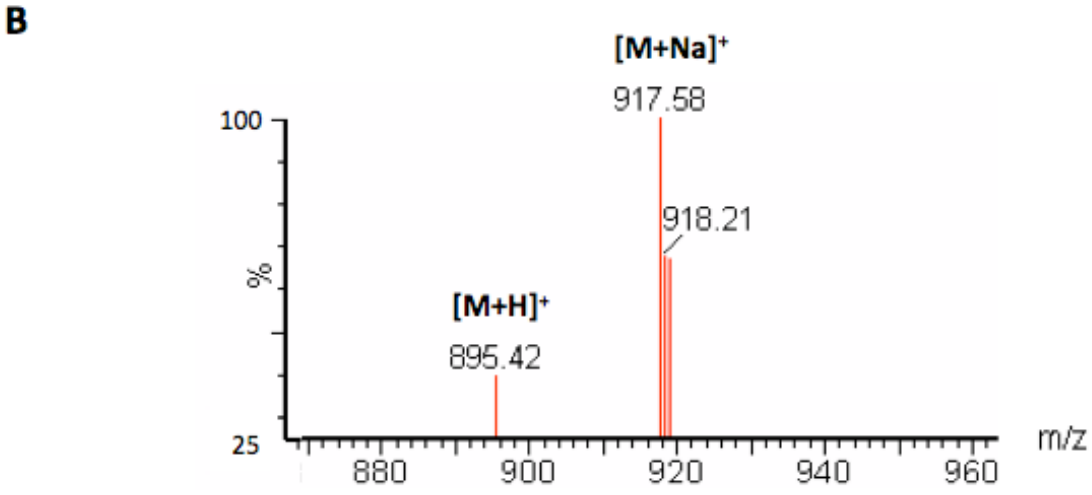


Figure S29. Design of precursor peptide hybrids used for expression in engineered *E. coli*. Only **3** and **16** led to detectable products. ESI-MS of the monoprenylated form of TFPVPTVC corresponding to the LC-MS trace shown in Figure 7 is given.





Genes and Cloning. The RSI mutants of TruLy2 (**3**) and Pag/Tru (**11**) were made by site-directed PCR mutagenesis. Codon-optimised vectors of TruLy2 (**3**) and Pag/Tru (**11**) were obtained from GenScript and were used as templates for the PCR mutagenesis. Primers were made with 5' phosphorylated ends, and after PCR amplification, the PCR reaction was digested with DpnI. The corresponding amplified bands were gel-purified from agarose gels using the QIAquick Gel Extraction Kit (Qiagen). Ligations were performed, followed by transformation and colony screening. The sequence of each plasmid was confirmed by sequencing from Genewiz, Inc. All genes for protein expression were cloned into pET28, between NdeI and XhoI sites, with an N-terminal His-tag. All precursor genes for cyanobactin production in *E. coli* were cloned into the pRSF-lac vector designed in our lab previously[1] between NdeI/KpnI restriction sites.

Protein Expression and Purification. Precursor peptides ThcE4, LynE, PagE6, TruLy2 and Pag/Tru along with their mutants were made by expression in R2D-BL21 cells in 2xYT medium supplemented with 50 μg mL^{-1} kanamycin and 25 μg mL^{-1} chloramphenicol. 10 mL of an overnight seed culture was used to inoculate each liter of media, and cultures were grown at 30°C with shaking till the OD_{600} reached 0.5-0.6, after which the cultures were induced with 1 mM IPTG and the temperature was increased to 37°C for the next 3 h. The cells were harvested and the pellets stored at -80°C till used for purification. All precursor peptides were purified using denaturing conditions. The cells were resuspended in lysis buffer (10 mM Tris, 0.1M NaH_2PO_4, 8M urea, pH 8.0) and shaken for 30 m at 4°C, centrifuged (20000 x g, 45 m) to collect supernatant, which was loaded onto a Ni-NTA affinity column. The resin was washed twice first with pH 6.3 buffer followed by another wash with pH 5.9 buffer (each 10 mM Tris, 0.1M NaH_2PO_4, 8M urea). Finally, proteins were eluted by washing with elution buffer (10 mM Tris, 0.1M NaH_2PO_4, 8M urea, pH 4.5). Elutions were subsequently dialyzed (0.5 M NaCl, 5 mM HEPES, 2-10 mM DTT, pH 6.0) over the next few days at 4°C with up to three buffer changes, aliquoted and stored flash-frozen at -80°C.

ThcD and ThcA were expressed under similar conditions as given above except that 0.1 mM IPTG was used in an overnight induction at 18°C. Note that full-length ThcA could not be expressed despite extensive attempts. The protease domain comprising residues 1-290 could be easily made instead, and exhibited functional activity as would be expected from the full-length protease. The enzymes were purified under native conditions, wherein the cell pellets were suspended in lysis buffer (50 mM Tris, 200 mM NaCl, 10 mM imidazole, 5% glycerol, pH 7.5) and stirred on ice for 1 h with lysozyme (600 μg mL^{-1}). The cells were then sonicated by a VibraCell 750 sonicator (30% amplitude for 3 m with 30s on/off cycles) and incubated for another 30 m with 20 μg mL^{-1} DNase and 10 mM $MgCl_2$ on ice. The cells were centrifuged (20000 x g, 45 m, 4°C) and the supernatant containing soluble protein was passed twice through 0.5 μm and 0.45 μm filters before loading onto a clean Ni-NTA resin affinity column. The resin was washed twice (1 M NaCl, 30 mM imidazole, pH 8.0) and finally eluted with high imidazole buffer (1 M NaCl, 200 mM imidazole, pH 8.0). The elutions were dialyzed (25 mM HEPES, 500 mM NaCl, pH 8.0) for 48-72 h with regular buffer changes, and stored as 10% glycerol (w/v) aliquots at -80°C. Only freshly thawed aliquots were used for each enzyme assay.

The precursor peptides TruE2, PatE, TruLy1 and the enzymes TruD, PatD and PatA were made as described previously.[2,3] The synthetic substrates **4-7** were made at the University of Utah Peptide Synthesis Core Facility.

Enzyme Assays and Product Characterization. The optimized reaction conditions for all PTM assays are given in Methods. Additionally, specific reaction parameters such as substrate and enzyme concentration along with reaction time-point are given in each figure legend of this manuscript. The characterization of reaction products is given in Methods. Briefly, precursor peptide PTM were analysed by SDS-PAGE for easy visualization followed by ESI-MS and/or FTMS to confirm identity of dehydration as azoline modification. Synthetic substrates were analysed by a combination of HPLC, MALDI-MS and ESI-MS based methods. HPLC analysis provided preliminary results since introduction of thiazoline corresponds to appearance of a 254 nm UV absorbance shoulder. Confirmation of this as thiazoline was done by ESI-MS, MALDI-

MS and/or FTMS. Competition reactions were analysed in the same way as single substrate reactions. Detailed interpretation of all MS spectra is given below.

Detailed interpretation of MS analysis of heterocyclization reactions

Substrates are listed below in the order that is given in Table 1 along with their detailed product characterization. In all cases an underlined Cys/Ser/Thr residue indicates heterocyclization.

A) ThcE4 (1)

ThcE4 has a core peptide sequence of SCDCSLYGGCESC (from our *in vitro* analysis of ThcE4 proteolytic cleavage). ESI-MS analysis of ThcE4/TruD reaction showed that TruD modifies four residues (Figure S21-A). Since ThcE4 has four Cys, for which TruD is known to be chemoselective, this indicates modification of these four Cys residues to thiazolines by TruD (S<u>C</u>D<u>C</u>SLYGG<u>C</u>ES<u>C</u>). PatD also modifies ThcE4, but we observed only one heterocyclization event by ESI-MS (Figure S21-B), since the reaction was slow. This led us to carry out the reaction for 45 h. The reaction product was digested with chymotrypsin and the digests were analysed by FTMS, which showed presence of three out of four possible Cys heterocycles (Figure S21-C) that were localized to S<u>C</u>DCSLYGG<u>C</u>ES<u>C</u> (Table S3). The ThcE4/ThcD reaction was carried out similarly to the TruD reaction and led to four heterocyclization events (Figures 2 and S3). This reaction was analysed by FTMS after chymotrypsin digest, localizing modifications to S<u>C</u>D<u>C</u>SLYGG<u>C</u>ES<u>C</u> (Table S1). Note that as shown in Figure S3-B, in case of the analysed fragment GGCE<u>**S**</u>CSYEGDEAE, if the Ser (shown by <u>S</u>) would have been cyclized instead of Cys, the b-ion of 347.19 (arising from fragment GGCE Figure S3B-i) or 329.11 (arising from fragment GG<u>C</u>E Figure SB-ii) and the y-H_2O ion of 1053.3 (arising from fragment S<u>C</u>SYEGDEAE Figure S3B-i,ii) would not have been observed. This is because in case of

oxazoline formation at this Ser, the resistance in fragmentation of the peptide bond before this Ser residue would prevent formation of these fragments. On the contrary presence of these fragments indicate, that the Ser have not been heterocyclized. The same principle applies for ThcE4/PatD reactions shown in Figure S21 and Table S3.

B) TruLy1 (2)

TruLy1 has core sequences of TLPVPTLC and VCMPCYP. As seen in the SDS-PAGE in Figure S6, Truly1 is modified by TruD, PatD and ThcD. TruD modification of TruLy1 leads to loss of three H_2O. These are localized to the Cys residues, since our studies show TruD to be chemoselective for Cys, and there are three Cys residues in the TruLy1 core cassettes (TLPVPTLC/VCMPCYP). PatD modification of TruLy1 leads to loss of $4H_2O$. These are localized to the three Cys residues and an additional Thr residue in the cassette TLPVPTLC/VCMPCYP. This is based on PatD activity, which is on all Cys/Thr residues present. There are only four such residues in this cassette. Lastly, ThcD modification of TruLy1 leads to loss of three H_2O, which is localized to the Cys residues, in a similar way as in the case of TruD.

C) TruLy2 (3)

TruLy2 has core sequences of TFPVPTVC and ACMPCYP. As seen in the SDS-PAGE in Figure S7, Truly2 is modified by TruD, PatD and ThcD. Similar to the above analysis for **2**, these heterocyclizations can be localized to TFPVPTVC and ACMPCYP. Given the low reactivity of PatD, loss of only three H_2O was seen, instead of the expected four.

D) CITFCA (4)

This was found to be not reactive with any enzyme. Both HPLC and ESI-MS analysis showed only starting substrate and no product formation (data not shown).

E) AITFCAYDGE (5)

As shown in Figure S7, this substrate was reactive with all enzymes. Reaction was followed both by HPLC and ESI-MS. Since there is only one Cys residue, the 18 Da mass-shift in presence of enzyme was attributed to thiazoline on this Cys (AITFCAYDGE). The UV absorbance shoulder at 254 nm provided further evidence for thiazoline ring formation.

F) CITFCAYDGE (6)

As shown in Figure S8, and similar to analysis of **5**, the heterocycle was attributed to the one Cys residue (CITFCAYDGE). Further confirmation was also provided by high-resolution FTMS, and the MS/MS fragmentation pattern confirmed thiazoline formation (Figure S9, Table S2). Due to slow reactivity of these substrates (Figure 4), these reactions were carried out for a much longer time than usual (30 h).

G) LAELSEEALGGVDASTSIAPFCSYD (7)

As shown in Figure S10, reaction of substrate **7** was analysed by MALDI-MS (also ESI-MS where doubly charged species could be seen). In all cases, with all enzymes, it was reactive to give a single 18 Da mass-shift. This was localized to TSIAPFC since there is only one Cys in the core.

H) TruLy2 (3[Glu->Ala]) mutant (8)

Substrate **8** has the same core sequences as TruLy2 (**3**), that is, TFPVPTVC and ACMPCYP. Both TruD and ThcD were reactive, with TruD leading to loss of one H_2O and ThcD to loss of

two H_2O. The exact regiospecificity of these modifications was not probed, since we only wanted to determine if this RSI mutant could be a substrate for the heterocyclase. But from precedents to other reactions, the heterocycles can be assigned to TFPVPTVC/ACMPCYP. Note that given the low efficiency of PatD in our hands, PatD was not used for these reactions.

I) Pag/Tru (3[Leu->Ala]) mutant (9)

J) Pag/Tru ΔRSI (10)

Substrates **9-10** have the same core sequence as PagE6 (**15**), that is, INPYLYC, the only difference being that the terminal Pro was mutated to a heterocyclizable Cys residue. But, they were not reactive with either TruD or ThcD, and only starting material was the major mass peak observed even after 18 h of reaction.

K) Pag/Tru (11)

Substrate **11** has the same core sequence as **9-10**, that is, INPYLYC, the only difference being it carries the RSI insertion in its leader. Unlike 9 and 10, all enzymes efficiently processed this substrate to give the 18 Da mass-shifted product of INPYLYC.

L) TruE2 (12)

TruE2 has core peptide sequences of TFPVPTVC and TSIAPFC. Earlier work has shown that TruD processes the terminal Cys of both cassettes leaving the remaining heterocyclizable Thr/Ser residues unmodified. On the other hand, PatD is also capable of modifying TruE2, but it heterocyclizes all potential residues except the N-terminal end Ser in TSIAPFC.[2] Here we show that ThcD also modifies TruE2 in a manner similar to TruD such that it converts only the terminal Cys to thiazoline. As shown in Figure S22-A, the SDS-PAGE gel band shift assay shows increased mobility of TruE2 in presence of ThcD, which is similar to reported band-shifts of

TruE2 in presence of TruD/PatD. Subsequent ESI-MS of the same reaction indicates band-shift is due to loss of two H_2O. Localization of the modifications was done by FT-ICR after PatA digestion of ThcD-modified TruE2, which gave the fragment corresponding to TFPVPTVCSYDGVDAS, which was further confirmed by the fragmentation MS2 pattern of the same (Table S4).

M) PatE1-58 (13)

PatE1-58 has a core peptide sequence of VTACITFC. It is known that TruD heterocyclizes Cys residues only, whereas PatD can heterocyclize all Cys/Thr residues.[2] Here we show that ThcD can also modify PatE1-58, as seen by the SDS-PAGE gel band-shift assay (Figure S22-B). Since PatD modifies four residues (two Cys and two Thr) in contrast to two residues (two Cys) modified by TruD, mobility of PatD-treated PatE1-58 is more than the TruD-treated peptide, as observable by a greater band-shift. In comparison, ThcD-treated PatE1-58 shows the same amount of band shift as with TruD treatment. Hence, as is the case for TruE2, ThcD also resembles TruD for PatE substrate, and performs heterocyclizations on the Cys residues only in the sequence VTACITFC, as shown by ESI-MS data of the same indicating two heterocyclization events.

N) LynE (14)

LynE has core peptide sequences of two cassettes of ACMPCYP and one cassette of VCMPCYP.[4] Here we show that LynE can be modified by all three of the enzymes TruD, PatD and ThcD as is observed by the band shift assay (Figure S23-A). For LynE/ThcD reaction, MALDI-MS analysis was done, which showed that ThcD modified LynE, and the corresponding mass indicated six heterocyclization events along with two Met oxidations. It is also possible that

this observed peak corresponds to four heterocyclization events and one disulphide linkage. The former is more likely since the SDS-PAGE gel band shift assay shows a complete shift comparable to LynE/TruD reaction, and LynE-digested fragments frequently carried Met oxidation in our MS experiments. FTMS analysis (after PatA digestion) of the reaction of LynE with ThcD showed parent ions corresponding to mass ACMPCYPSYDGVDAS, carrying two thiazoline modifications, which were localized based on MS2 pattern (Table S5). For the LynE/TruD reaction, ESI-MS analysis shows that TruD yields six modifications, corresponding to heterocyclization of all Cys residues on LynE (Figure S23-B). FT-ICR analysis after PatA digestion detected the fragment VCMPCYPSYDDAE, carrying a thiazoline ring on the Cys residue, which was localized from MS2 (Table S6). For the LynE/PatD reaction, FT-ICR analysis after PatA cleavage detected fragments corresponding to the masses of ACMPCYPSYDGVDAS, VCMPCYPSYDGVDAS and VCMPCYPSYDDAE (Figure S24). Localization of the modifications was performed based on the MS2 fragmentation pattern of each (Table S7). We also observed Met oxidation (shown by M).

O) PagE6 (15)

This substrate is a native precursor peptide from the *pag* family that lacks heterocycles. Since PagE6 lacks heterocyclizable residues in its core of INPYLYP, it could not be assayed with the heterocyclase enzymes.

P) Pag/TruLy2 (16)

Precursor peptide **16** was only used for cyanobactin production in engineered *E. coli* and no *in-vitro* analysis was performed on it.

MS analysis of proteolysis cross-reactions

PatA was seen to cleave ThcE4 (**1**) at the same site as ThcA, after AVLAS RSII (Figure S28-A). For TruE2 (**12**) cleavage by ThcA (Figure S28-B), a minor mass peak of 6209.0 Da was observed by ESI-MS. This corresponds to cleavage after the GVDAS site before the first core peptide cassette. PatE1-58 (13) cleavage by ThcA (Figure S28-C) yielded a mass peak of 6693.1 Da, which corresponds to cleavage after the GLEAS site, along with gluconoylation of the His-tag.[5] FTMS analysis showed that PatA cleaves LynE (**14**) after GVDAS before each of its three cassettes (Figure S28-D, Table S8). ESI-MS of the LynE/ThcA reaction gave smaller products, which could not be assigned to specific cleavage sites (data not shown), possibly because the samples contained a mixture of peptides arising from Met oxidation of LynE. ThcA cleavage of the hybrid TruLy1 (**2**) yielded mostly unmodified peptide with a small amount of product corresponding to cleavage after GVDAS of the second cassette. Since the reaction had a high level of noise (data not shown), this cleavage site could not be assigned definitively. The product of PatA cleavage of the same was detected after GVDAS of the first cassette corresponding to mass peak of 6209.9 Da (Figure S28-E). Activity of TruA was explored *in vivo* only. This enzyme was active on substrates **3** and **16,** both after the GVDAS RSII.

References:

(1) Tianero, M. D. B.; Donia, M. S.; Young, T. S.; Schultz, P. G.; Schmidt, E. W. *J. Am. Chem. Soc.* **2011**, 418.
(2) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. *J. Am. Chem. Soc.* **2010**, *132*, 4089.
(3) McIntosh, J. A.; Donia, M. S.; Nair, S. K.; Schmidt, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 13698.
(4) McIntosh, J. A.; Lin, Z.; Tianero, M. D. B.; Schmidt, E. W. *ACS Chem. Biol.* **2013**, *8*, 887.
(5) Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G. *Anal. Biochem.* **1999**, *267*, 169.