Electronic Supplementary Material (ESI) for ChemComm.

Albusnodin: an acetylated lasso peptide from *Streptomyces albus*

Chuhan Zong[a], Wai Ling Cheung-Lee[b], Hader E. Elashal[c], Monika Raj[c] and A. James Link[b,d]

Department of Chemistry [a], Department of Chemical and Biological Engineering [b] and Department of Molecular Biology [d], Princeton University, Princeton, NJ 08544 USA

Department of Chemistry [c], Seton Hall University, 400 South Orange Ave, South Orange, NJ 07079, USA

To whom correspondence should be addressed: ajlink@princeton.edu

Contents

Methods

Bioinformatics

The genome sequences in fasta format from bacteria and archaea were downloaded from NCBI genome database. An updated version of the original genome mining script developed by Maksimov et al[1] was used in this study. In this version, a more flexible lasso peptide precursor pattern was used, allowing for variation beyond glycine in the first amino acid of the core peptide. Additionally, a new MEME training set containing 28 known lasso peptide clusters was used to generate four motifs from B enzymes and three motifs from C enzymes by applying the MEME searching algorithm[2]. The updated motif information was then incorporated into the script for maturation enzyme motif matching using MAST[2]. Putative lasso peptide clusters were ranked from 0 to 7 based on the numbers of motifs found in maturation enzymes.

Growth of *S. albus* and genomic DNA extraction

The freeze dried culture of *Streptomyces albus* DSM 41398 was purchased from DSMZ. The culture pellet was rehydrated in GYM medium (glucose 4.0 g, yeast extract 4.0 g, malt extract 10.0 g per liter of distilled water, pH = 7.2) and grown on a GYM plate (glucose 4.0g, yeast extract 4.0g, malt extract 10.0g, $CaCO_3$ 2.0g, agar 12.0 g per liter of water, pH = 7.2) at 30 $^{\circ}$C for 6 days. Spores were harvested and subcultured into 10 mL of GYM medium for genomic DNA extraction. The whole genomic DNA of *Streptomyces albus* DSM 41398 was extracted using Qiagen DNeasy Blood & Tissue Kit (Qiagen). PicoMaxx

DNA polymerase (Agilent Technologies) was used for albusnodin gene cluster amplification with primers purchased from Integrated DNA Technologies.

Plasmid construction and conjugation of albusnodin gene clusters

Molecular cloning was done according to standard protocols. The plasmids used in this study is shown in Table S1. The fragments containing the *albACB* or *albACBT* gene clusters with a native RBS upstream of *albA* were amplified using primers from Table S2. This PCR amplification also introduced the restriction sites *Eco*RI and *Xho*I at 5' end and *Hin*dIII at 3' end of the fragments. Both *albACB* and *albACBT* were cloned into the multiple cloning sites (*Eco*RI and *Hin*dIII) of pQE80 (Qiagen) to create pCZ64 and pCZ65 accordingly. The plasmids pCZ64 and pCZ65 were then digested with *Xho*I and *Hin*dIII to generate fragments *XhoI-albACB-HindIII* and *XhoI-albACBT-HindIII*. The two fragments were inserted and ligated into the multiple cloning sites (*Xho*I and *Hin*dIII) of *Streptomyces* shuttle vector pIJ10257[3] to generate pCZ66 and pCZ68.

Table S1: Plasmids used in this study

Plasmid Name	Description	Primers	Resitriction Sites	Vector
pCZ64	*albACB*	1,2	*EcoR*I, *Xho*I, *Hind*III	pQE80
pCZ65	*albACBT*	1,3	*EcoR*I, *Xho*I, *Hind*III	pQE80
pCZ66	*albACB*	-	*Xho*I, *Hind*III	pIJ10257
pCZ68	*albACBT*	-	*Xho*I, *Hind*III	pIJ10257

Table S2: Primers used in this study

Primer Name	Sequence (5’ to 3’)
Primer 1	GATCAACAATTCTCGAGCGCACTCCGAAAGGAAGATCATGGACAGCC
Primer 2	GAGCTAAGCTTTCAGGGCTGGTGCTGATGATGGCC
Primer 3	GAGCTAAGCTTCTACGGTGCTCGGCGGAGCTGA

The resulting plasmids were transformed into *S. albus* J1074 *and S. lividans* 66 by conjugation with *E. coli* S17-1 using standard protocols.[4, 5] Due to methyl-specific restriction mechanisms of *S. coelicolor* M1146, methylation deficient strain *E. coli* ET12567/pUZ8002 was used for transformation and conjugation with *S. coelicolor* M1146. Exconjugants were selected with hygromycin (50 µg/mL) and nalidixic acid (25 µg/mL)[4]. Successful exconjugants were restreaked onto fresh SFM plates (20 g soy flour, 20 g mannitol, 20 g agar, 10 mM $MgCl_2$ per liter of water) with selection for an additional 7 days before harvesting spores.

Albusnodin expression and purification

For a 1 L scale heterologous production of albusnodin, *Streptomyces* cultures were started by inoculating 2 x 10^9 spores from a frozen stock in 150 mL 2YT medium (tryptone 16.0 g, yeast extract 10.0 g, sodium chloride 5.0 g per liter of water, pH = 7.0) and incubated at 30 °C with orbital shaking (350 rpm) for 8 h. The germlings were harvested by centrifugation at 2,000 x *g*, resuspended in 25 mL of GYM medium, dispersed by a

quick sonication pulse and used to inoculate production cultures at a starting OD_{450} of 0.005. All strains were cultured in Erlenmeyer flasks that had been treated overnight with 1 mL dimethyldichlorosilane in toluene to minimize the culture from sticking to the glass wall. Flasks were filled to no more than $1/5^{th}$ of the flask volume with GYM medium. A stainless steel spring was placed at the bottom of each flask to improve aeration. Production cultures were incubated at 30 °C with orbital shaking (250 rpm) for an additional 7 days.

The cell lysate and supernatant of the samples were separated by centrifugation at 8,000 x *g* at 4 °C for 20 minutes. The cells were lysed in methanol. The insoluble fraction was removed by centrifugation at 8,000 x *g* at 4 °C for 10 minutes, and the methanol-soluble fraction was dried using rotary evaporation and then resuspended in 500 µL of 50% acetonitrile (ACN) and 50% water solution. The supernatant was extracted using Strata C8 column (Phenomenex 6 mL size). The crude extract was eluted from the column using methanol, and dried using rotary evaporation, and then reconstituted in 500 µL of 50% acetonitrile (ACN) and 50% water solution.

HPLC analysis

The peptide samples from cell lysate and supernatant were injected onto a Zorbax 300SB-C18 Analytical-Prep HPLC Column (3.0 x 150mm, Agilent Technologies) using solvent gradient 1 (see Table S3). An HPLC fraction from 1 to 2 minutes contained the peptide of interest as judged by MALDI-MS and was collected. The fraction was dried under reduced pressure, resuspended into 100 µL of water, and then subjected onto a Zorbax 300SB-C18 Semi-Prep HPLC Column (9.4 x 250mm, Agilent Technologies) using

solvent gradient 2 (See Table S4). Solvent A consisted of H_2O with 0.1 % trifluoroacetic acid (TFA) and solvent B consisted of acetonitrile with 0.1% TFA.

Table S3: Solvent gradient 1

Flow Rate	0.750 mL/min
Time (min)	**% Solvent B**
0.00	10%
1.00	10%
20.00	50%
25.00	90%
30.00	90%
32.00	10%
36.00	10%

Table S4: Solvent gradient 2

Flow Rate	2.667 mL/min
Time (min)	**% Solvent B**
0.00	1%
1.00	1%
60	20%
70	90%
80	1%

100	1%

MALDI analysis

The fraction from 29.5 to 31.5 min (gradient 2) was collected, and 1 µL of the collected peak was spotted on a MTP 384 target plate ground steel BC plate. Subsequently, 1 µL of 2.5 mg/mL solution of α-Cyano-4-hydroxycinnamic acid (Sigma) in 50% acetonitrile/water was mixed with the spotted sample and allowed to dry. Molecular weight determination was performed in the *m/z* 300–3,000 range using an UltraFlextreme MALDI TOF/TOF (Bruker). MS/MS product ion analysis was performed with collision energies of 20 to 100 eV. Both MS and MS/MS analyses were performed in positive-ion mode.

LC-MS analysis

LC-MS analysis was performed on an Agilent 1260 Infinity II HPLC system coupled to an Agilent 6530 Accurate-Mass Q-TOF spectrometer with an electrospray ionization (ESI) source operated in positive ion mode. The analytes were separated with an Agilent Zorbax 300SB C18 column with the solvent gradient 3 (Table S5). Solvent A consisted of H_2O with 0.1 % formic acid (FA) and solvent B consisted of acetonitrile with 0.1% FA. The analytes from the column were sent to the MS and spectra were acquired in profile mode. When running tandem MS/MS, the +2 and +3 charge states (Z) of albusnodin were used as the target ion with a specific retention time of 11.5 minutes +/- 0.5 minutes. Collision energies of 25-90 eV were used to obtain MS/MS spectra.

Table S5: solvent gradient 3

Flow Rate	0.400 mL/min
Post Time	8 min
Time (min)	**% Solvent B**
0.00	1%
1.00	1%
60.00	20%
70.00	90%
80.00	1%
100.00	1%

Carboxypeptidase Assay

The peptide substrate was digested with 1 µL of 0.1 µg/µL carboxypeptidase B (Sigma-Aldrich) and 1 µL of 0.1 µg/µL carboxypeptidase Y (Affymetrix) in carboxypeptidase digestion buffer (50 mM sodium acetate, pH = 6.0) with a total volume of 20 µL for 16 hours at 20 °C. The digested sample was further analyzed using an LC-MS (Agilent 6530 Accurate-Mass Q-TOF LC/MS).

Trypsin Assay

Trypsin digestion was conducted in trypsin digestion buffer (50 mM NH_4HCO_3). The peptide substrate was treated with 1µl of 0.1 µg/µL of sequencing grade trypsin (Promega) in a total volume of 20 µL. The reaction was carried out for 16 hours at 37 °C. The digested

sample was subsequently analyzed using an LC-MS (Agilent 6530 Accurate-Mass Q-TOF LC/MS).

Chemical Cleavage

To activate Ser and accounting for the possible activation of Glu, excess N,N'-disuccinimidyl carbonate (50 mg), diisopropylethylamine (40 µl of a 20% vol/vol solution in dimethylformamide (DMF)), and a crystal of dimethylaminopyridine were added to partially purified, lyophilized albusnodin in 450 µl of DMF in a 5 mL round bottom flask. The reaction mixture was left on a shaker for 5 minutes to ensure solubility of the reagents followed by the addition of 450 µl of water to allow for simultaneous amino acid modification and amide bond cleavage. The reaction was left on a shaker at room temperature for 48 hours which was followed by the analysis of the reaction mixture using LC-MS as described above.

References

1 M. O. Maksimov, I. Pelczer and A. J. Link, *Proc. Natl. Acad. Sci.*, 2012, **109**, 15223–15228.

2 T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li and W. S. Noble, *Nucleic Acids Res.*, 2009, **37**, W202-208.

3 H. J. Hong, M. I. Hutchings, L. M. Hill and M. J. Buttner, *J. Biol. Chem.*, 2005, **280**, 13055–13061.

4. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, 2000.

5. J. P. Gomez-Escribano and M. J. Bibb, in *Methods Enzymol.*, ed. A. H. David, Academic Press, 2012, vol. Volume 517, pp. 279-300.

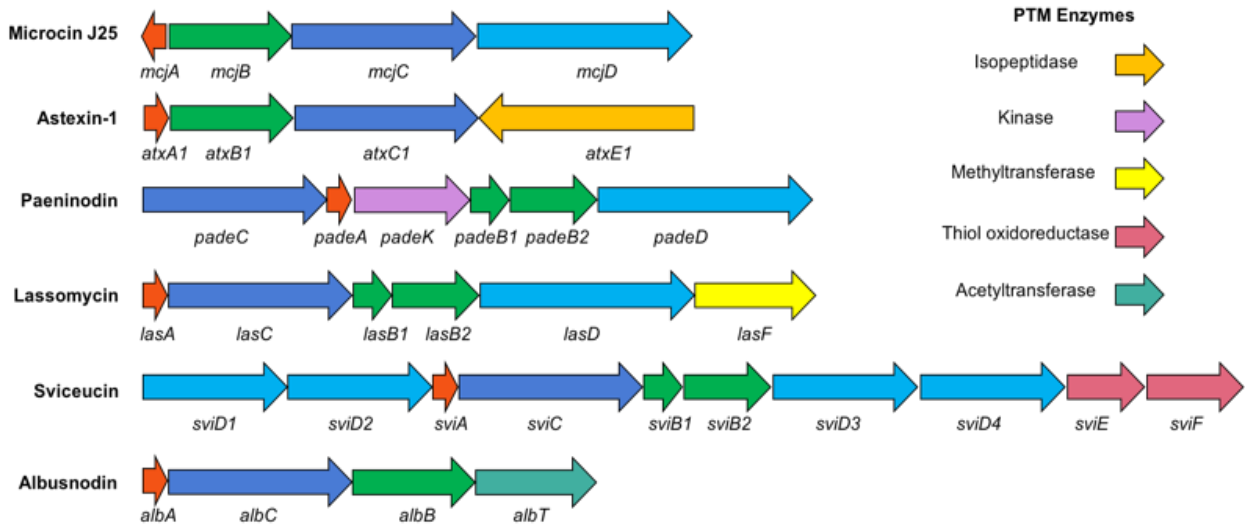


Figure S1, Examples of lasso peptide gene clusters with and without posttranslational modification enzymes.

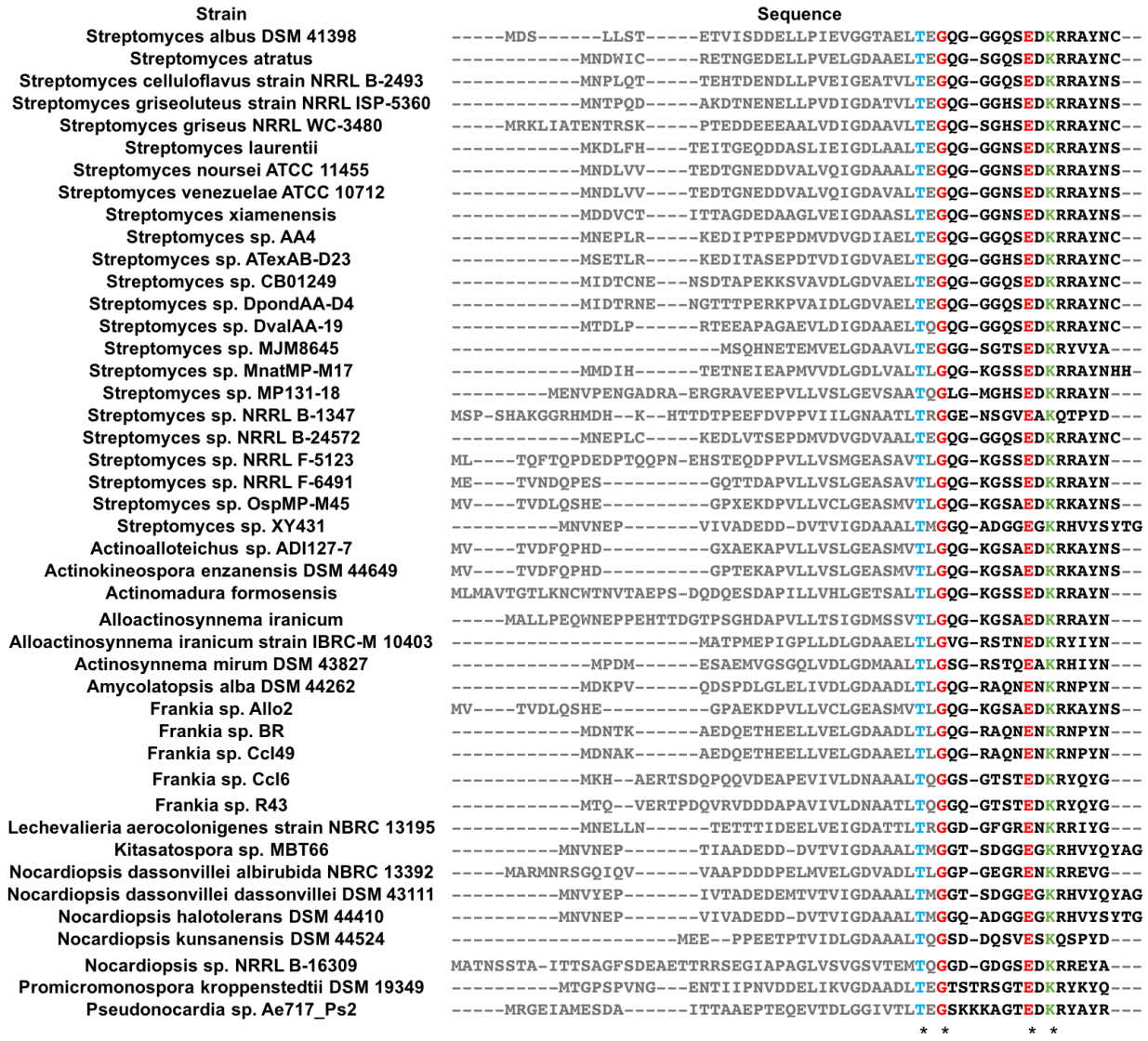


Figure S2: Complete list of distinct lasso peptide precursors found in actinobacteria with an *ACBT* gene cluster architecture. Asterisks from left to right highlight 1) the universally conserved Thr residue in the penultimate position of the leader peptide, the Gly residue at the N-terminus of the core peptide, the universally conserved Glu residue that forms the isopeptide bond, and a universally conserved Lys that is the site of acetylation. Note that some precursors include additional Lys residues and may be polyacetylated.

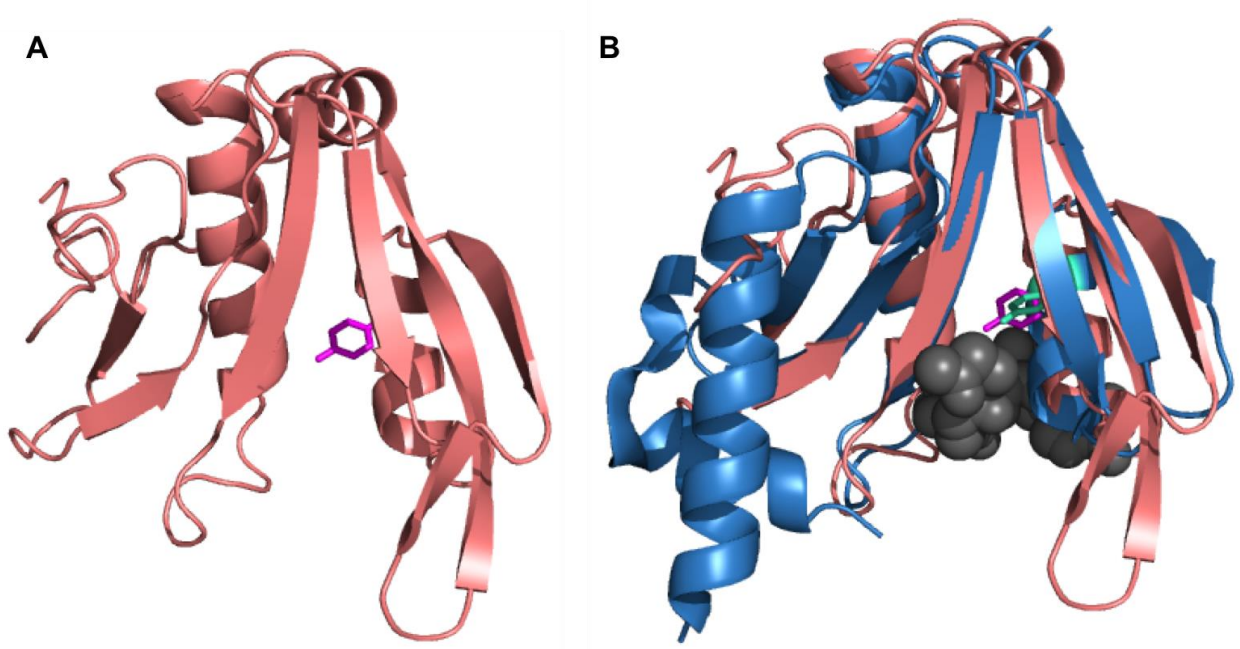


Figure S3: Structure prediction of the acetyltransferase found in the albusnodin gene cluster. A: Homology model of the AlbT acetyltransferase with catalytically important Tyr residue shown in stick representation. B: Alignment of the homology model in A with the *C. jejuni* PseH crystal structure (PDB 4XPL). The catalytically important Tyr residue is shown in sticks and acetyl-CoA is shown in space filling. The acetyl-CoA binding pocket is structurally well-conserved between the AlbT homology model and PseH though other regions of the protein are more poorly conserved.

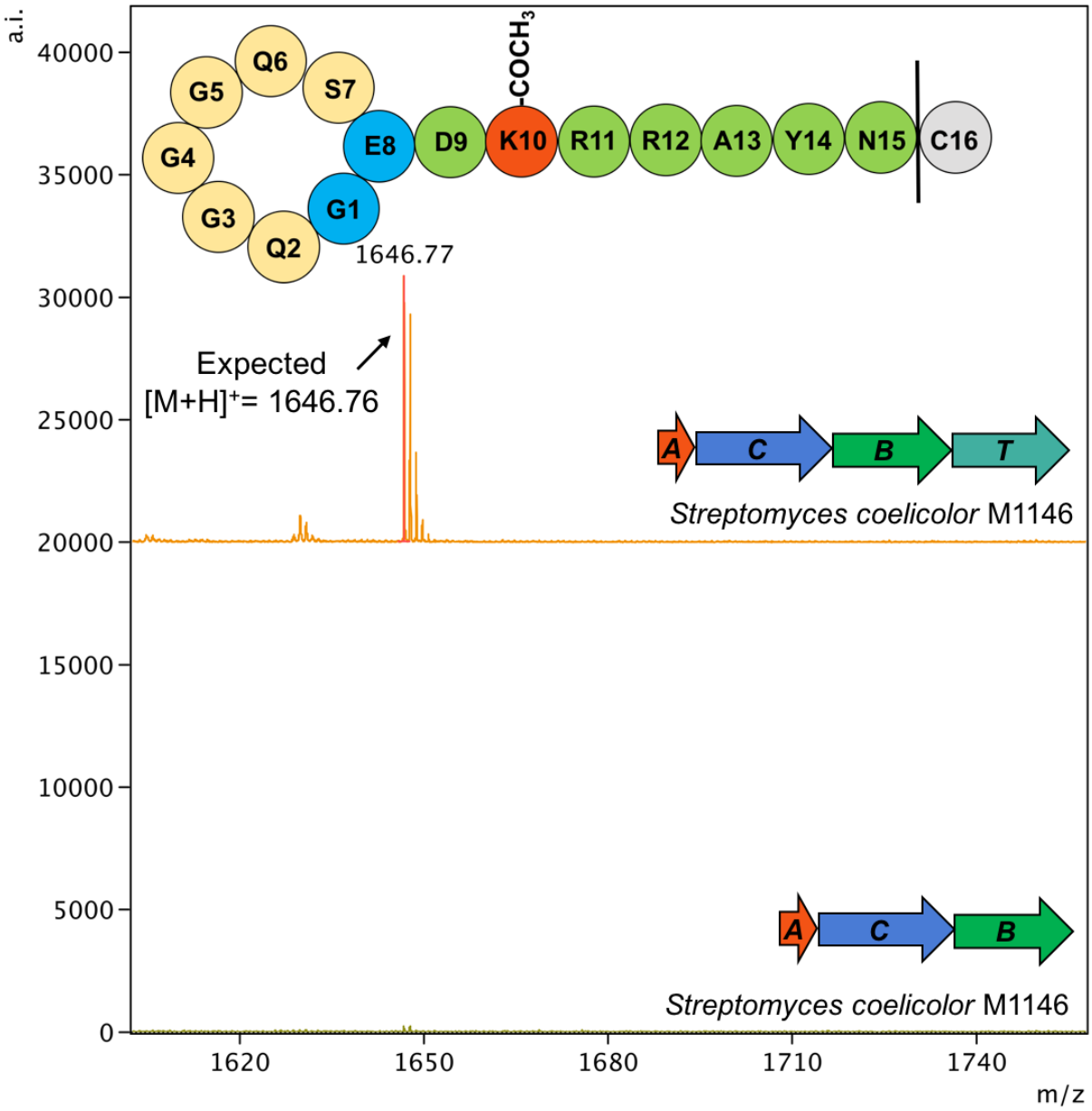


Figure S4, MALDI-MS spectra of partially purified supernatants from cultures heterologously expressing albusnodin in *S. coelicolor M1146*. When the acetyltransferase (T) gene is not included, no production of albusnodin is observed.

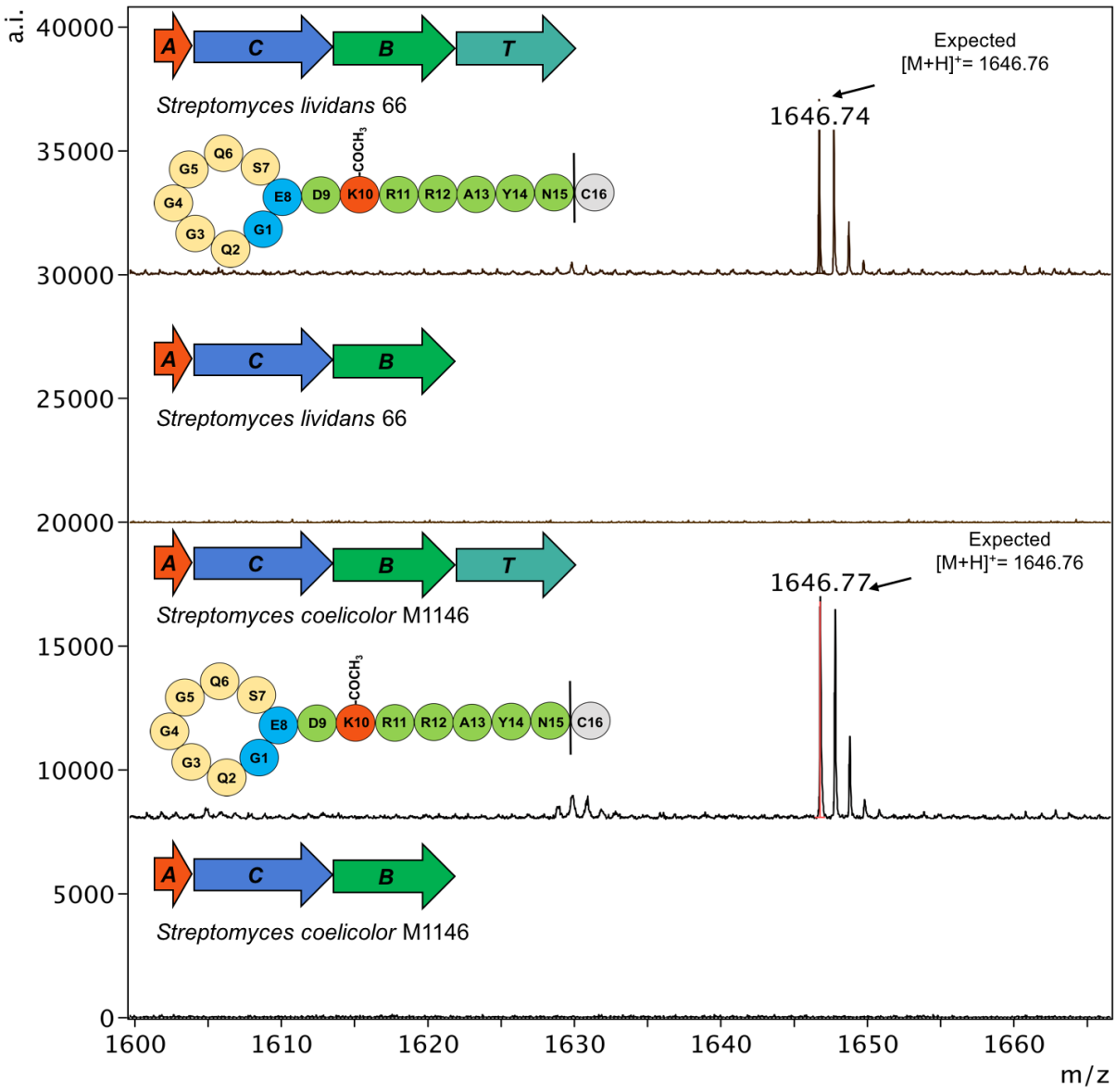
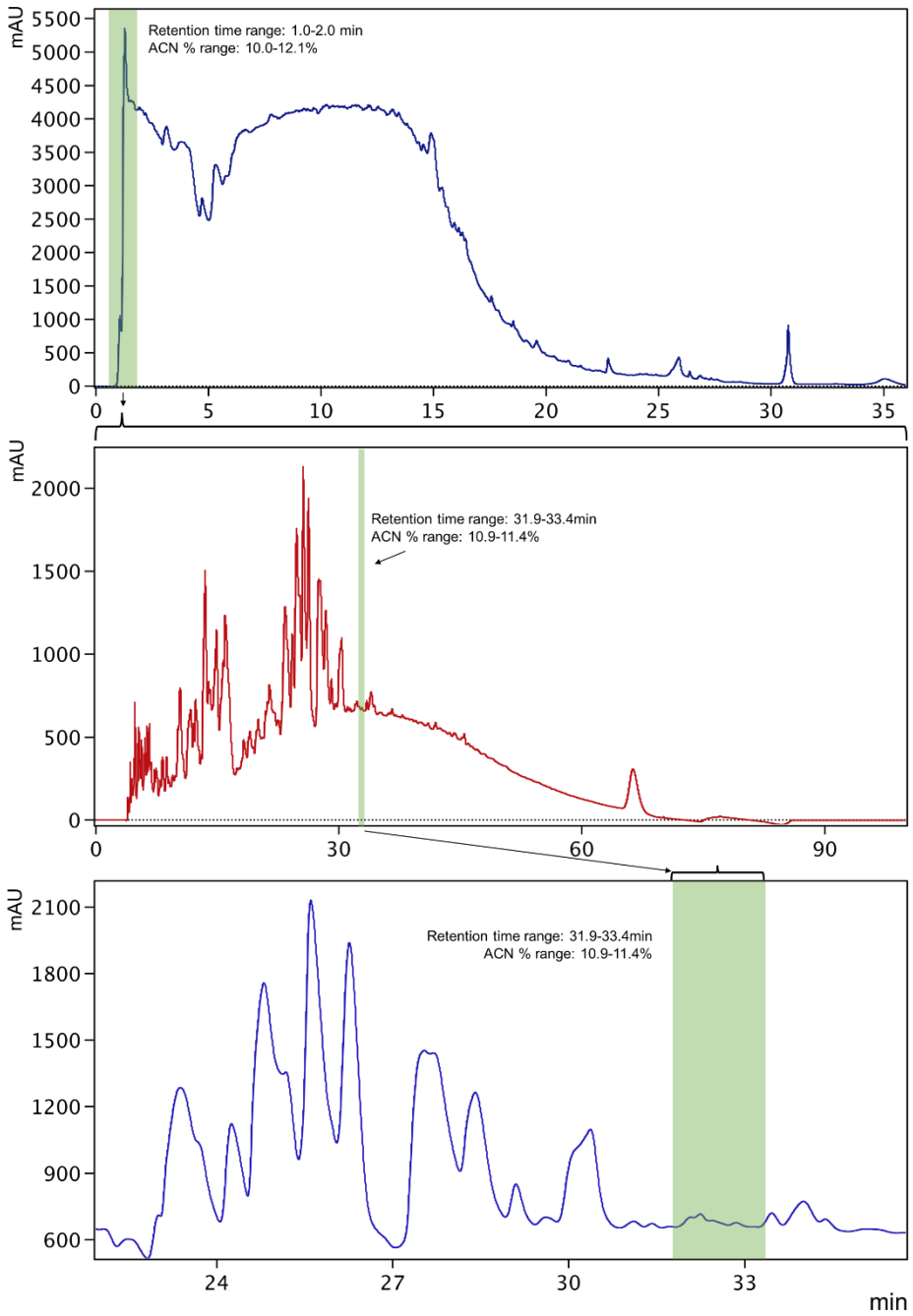


Figure S5: MALDI-MS spectra of cell lysates of *S. coelicolor* M1146 or *S. lividans* 66 harboring either the entire albusnodin gene cluster or a cluster lacking the acetyltransferase (*T*) gene. Removal of the *T* gene abrogates production of albusnodin. The same result is found in peptide isolated from the culture supernatant (Figure 3 of the main text).



4 L culture supernatant -> 4 mL of extract, 40 injections of 100 μL were done, collected peak reconstituted in 1 mL

1 mL of fractionated sample was injected 5 x 200 μL, and the indicated range was collected and injected to LC-MS (see Figure S7 below)

Figure S6: HPLC traces of albusnodin fractionation. Top: Using gradient 1, a mass corresponding to albusnodin elutes between 1 and 2 minutes. Middle: using a shallower gradient, gradient 2, the peptide elutes between 31.9 and 33.4 min. Bottom: Zoom-in of the middle HPLC trace showing no obvious peak for albusnodin

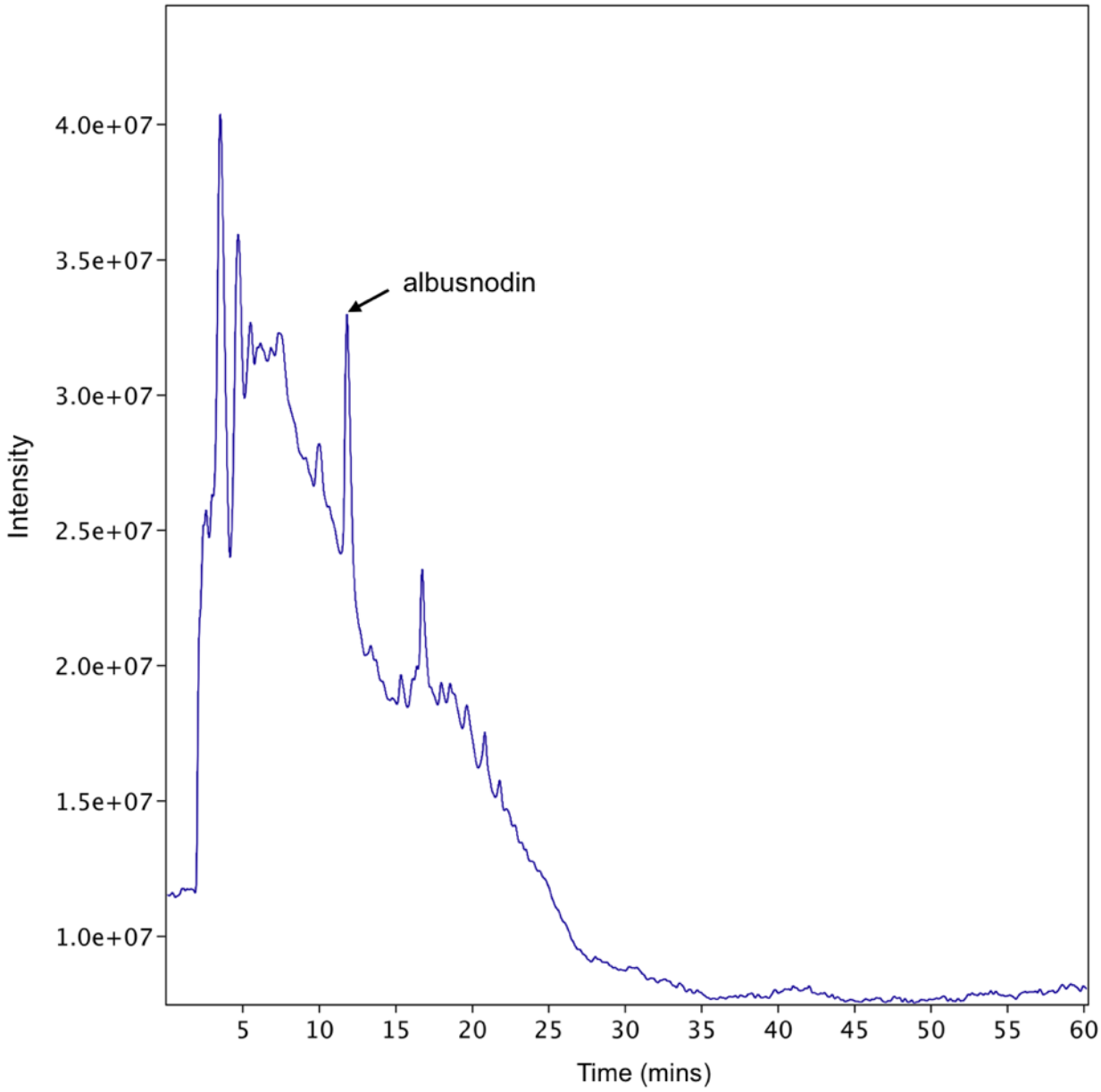


Figure S7: LC trace (total ion chromatogram) of partially purified albusnodin using qTOF mass detector. The doubly fractionated sample shown in Figure S6 was injected (13 μL). Gradient 3 (Table S5) was used to generate this trace. Intensity on the y-axis refers to the total ion counts. The position of the albusnodin peak is indicated.

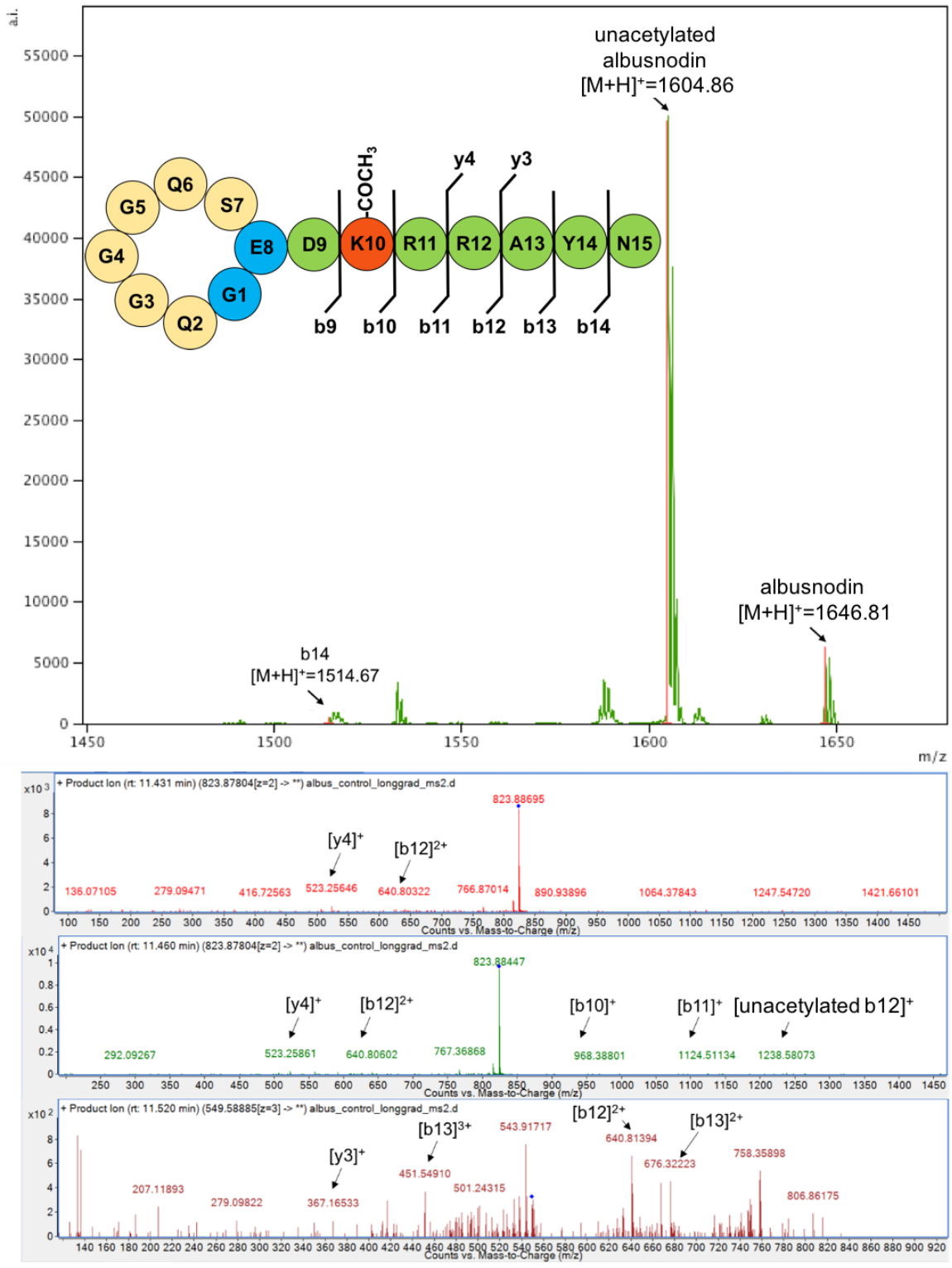


Figure S8: Additional MS/MS spectra showing fragmentation of albusnodin. Top: MALDI-TOF/TOF, bottom three panels: qTOF LC-MS/MS.

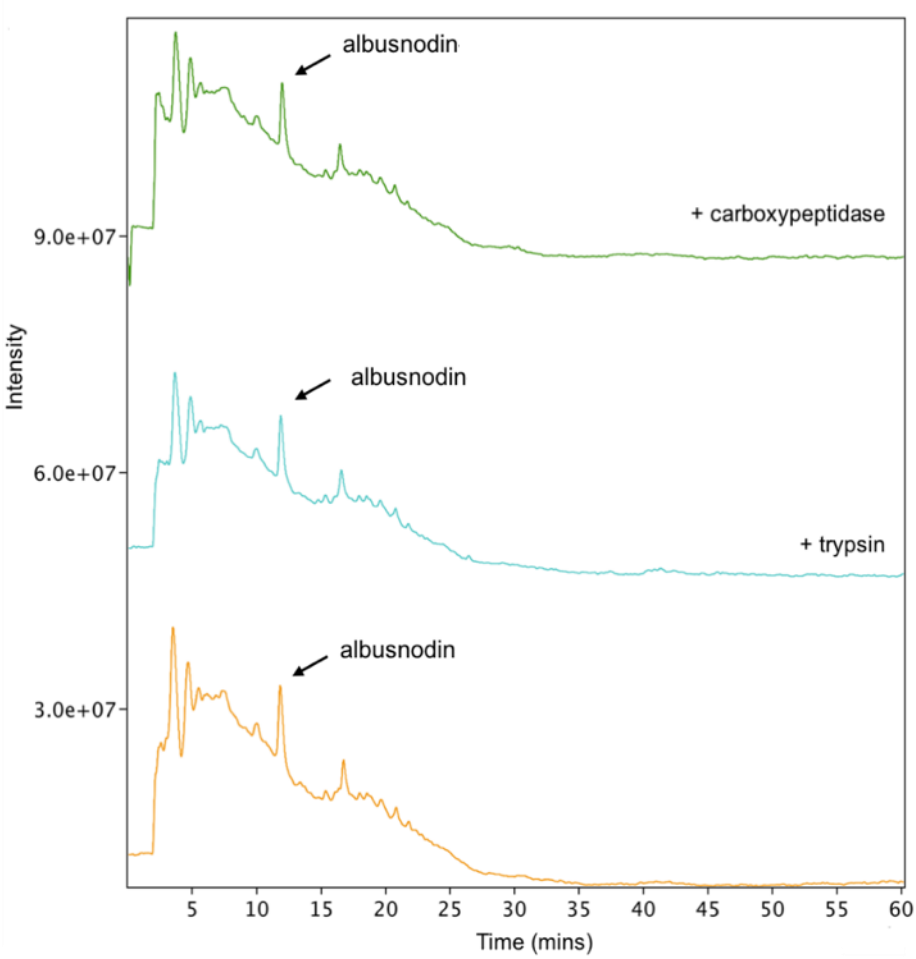


Figure S9A: LC traces (mass detector) of partially purified albusnodin (bottom, same figure as Figure S7), albusnodin after 16 h of incubation with trypsin (middle), and albusnodin after 16 h of treatment with carboxypeptidase (top). Neither trypsin nor carboxypeptidase cleaves albusnodin. Gradient 3 (Table S5) was used here.

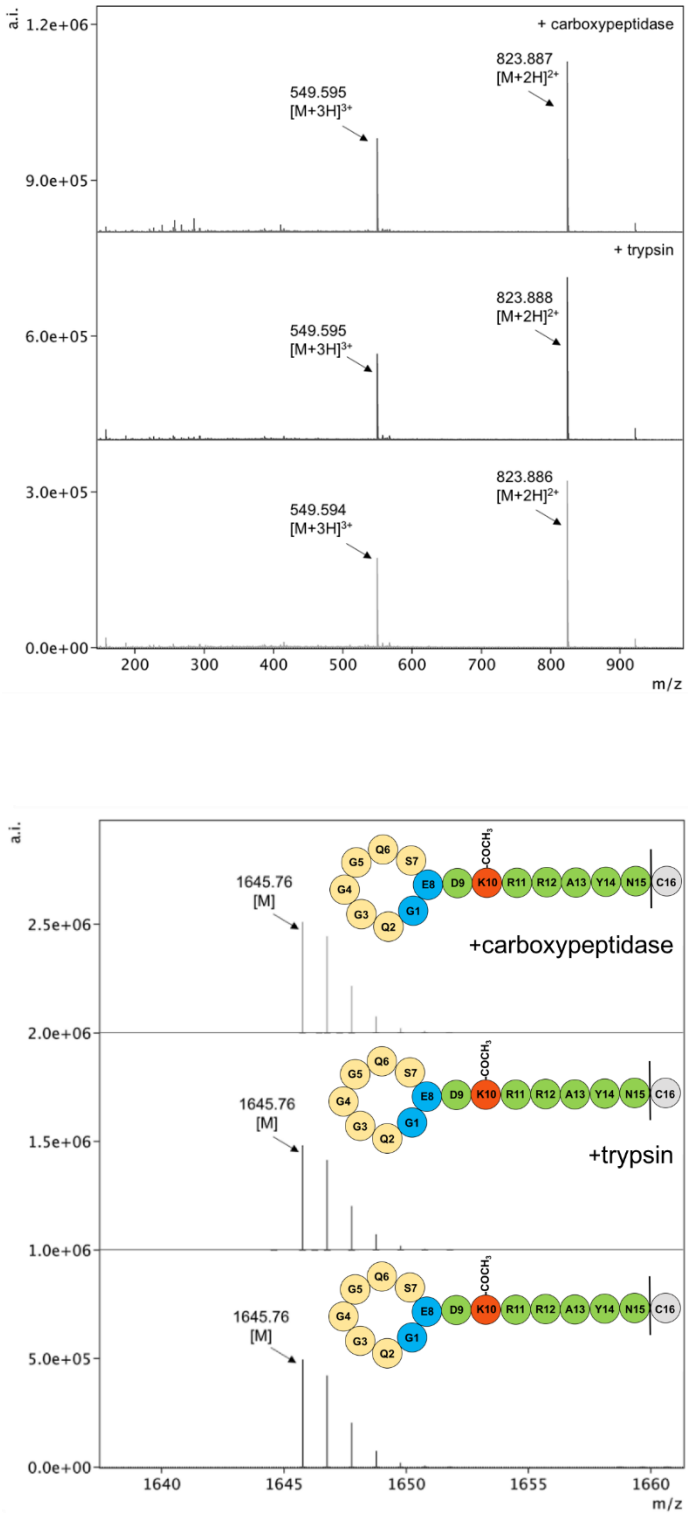


Figure S9B: Mass spectra from the samples in Figure S9A. Top: Spectra showing doubly and triply charged species. Bottom: Deconvoluted spectra.

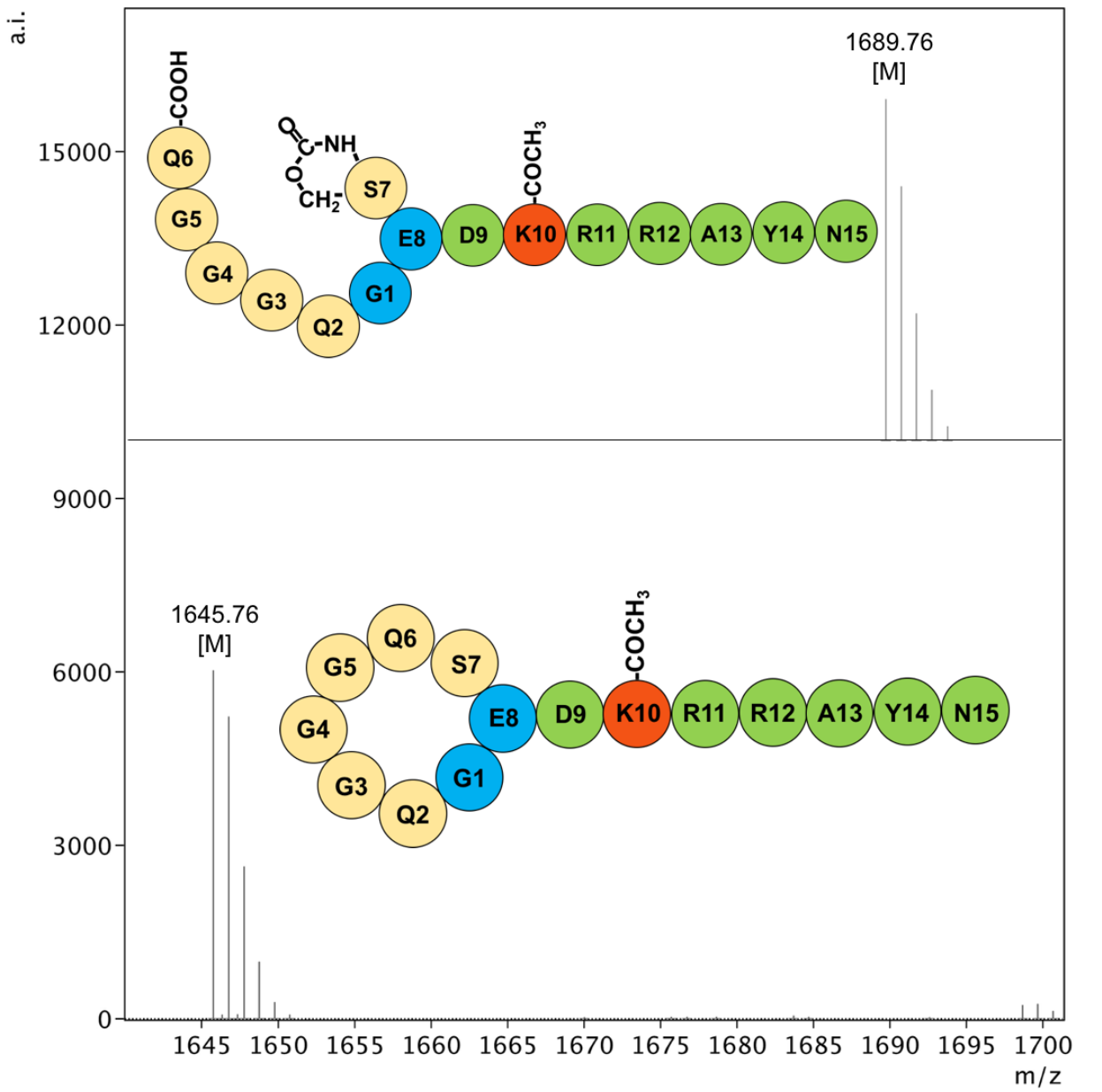


Figure S10: Chemical cleavage of albusnodin. Top: Cleavage is only observed at Ser-7, suggesting that the peptide is cyclized at Glu-8. Bottom: Mass spectrum of intact albusnodin for comparison.

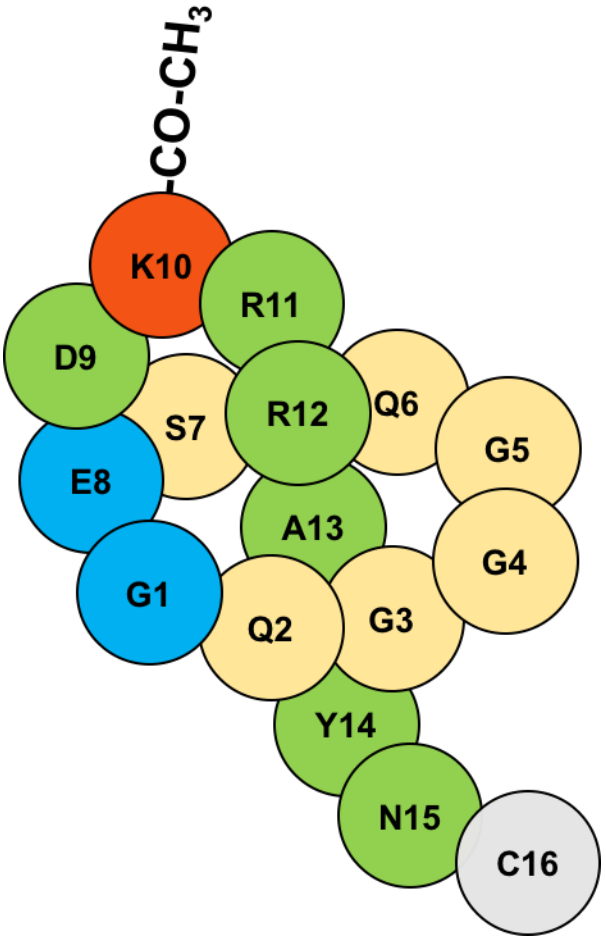


Figure S11: Cartoon model of the likely topology of full length albusnodin. The blue amino acids, G1 and E8 are isopeptide bonded. The site of acetylation, K10, is in red. C16 is encoded in the gene cluster but is cleaved off in the heterologously expressed peptide. The exact position of threading is unknown, one possibility is shown here.

Peptide Species	Expected Mass (Da)
full length albusnodin	1748.76
albusnodin	1645.76
unacetylated full length albusnodin	1706.75
unacetylated albusnodin	1603.75

Table S6, The expected masses of albusnodin variants