Electronic Supplementary Material (ESI) for Chemical Science.

SUPPORTING INFORMATION

Bioinformatics-Guided Discovery of Biaryl-Linked Lasso Peptides

Hamada Saad, Thomas Majer, Keshab Bhattarai, Sarah Lampe, Dinh T. Nguyen, Markus Kramer, Jan Straetener, Heike Brötz-Oesterhelt, Douglas A. Mitchell, and Harald Gross

Table of Contents

Relevant known scaffolds containing the PTMs under investigation.....

Supplemental References

1. Experimental Procedures

1.1 General Experimental Procedures

Solvents were all HPLC grade. Chemical reagents and standards were purchased from Sigma Aldrich unless indicated otherwise. The isotopically labeled substrates [L-tyrosine (D7, 98%), and L-tryptophan (D8, 97-98%)] were purchased from Cambridge Isotope Laboratories. Optical rotation values were measured on a Jasco P-2000 polarimeter, using a 3.5 mm × 10 mm cylindrical quartz cell. UV spectra were recorded on a PerkinElmer Lambda 25 UV/vis spectrometer. Infrared spectra were obtained by employing a Jasco FT/IR 4200 spectrometer, interfaced with a MIRacle ATR device (ZnSe crystal). ECD spectra were measured using a Jasco J-720 spectropolarimeter and a standard rectangular quartz cell (light path 2 mm, 700 µL volume).

For Liquid Chromatography/Low-resolution Electron Spray Ionization Mass Spectrometry (LC/LRESI-MS) measurements an 1100 Series HPLC system (Agilent Technologies, Waldbronn, Germany) was used. The Agilent HPLC components (G1322A degasser, G1312A binary pump, G1329A autosampler, G1315A diode array detector) were connected with an ABSCIEX 3200 QTRAP LC/MS/MS mass spectrometer (AB Sciex, Germany GmbH, Darmstadt, Germany). The resultant fractions were dissolved in methanol and profiled by LC-MS using a MeCN/H_2O (0.1% TFA) gradient, increasing the MeCN portion from 10% to 50% over 16 min and from 50% to 100% over 6 min, followed by isocratic elution at 100% MeCN for 8 min (Phenomenex Luna C18 (2)−100 Å column, 2 × 250 mm, 5 µm; 0.2 mL/min flow rate; with total ion current [positive mode Q3] and photodiode array monitoring).

For Liquid Chromatography/High-resolution Electron Spray Ionization Mass Spectrometry (LC/HRESI-MSMS) measurements, an Ultimate 3000 HPLC (Thermo Fisher Scientific) system united with MaXis-4G instrument (Bruker Daltonics, Bremen, Germany) was used. The developed HPLC-method was (0.1% FA in H_2O as solvent A and CH_3CN as solvent B), a gradient of 10% B to 100% B in 30 min ending with 100% B for an additional 10 min, with a flow rate of 0.3 mL/min, 5 µL injection volume and UV detector (UV/VIS) wavelength monitoring at 210, 254, 280 and 360 nm. Integrating Phenomenex Luna Omega polar C18 (3 µm, 150 × 3 mm) column enabled the separation with MS acquisition range of *m/z* 50-1800. A capillary voltage of 4500 V, nebulizer gas pressure (nitrogen) of 2 (1.6) bar, ion source temperature of 200 °C, the dry gas flow of 9 l/min source temperature, and spectral rates of 3 Hz for MS^1 and 10 Hz for MSMS were used. For MS/MS fragmentation, the 10 most intense ions per MS^1 were chosen for subsequent collision-induced dissociation (CID) with the stepped recommended CID energies.[1] For the mass calibration, sodium formate was directly infused before each sample measurement.

Vacuum liquid chromatography (VLC) was accomplished using the reversed-phase (RP) C18 column (dimensions: 10 × 5 cm; material: Macherey-Nagel Polygoprep 50–60 C18 RP silica gel). HPLC profiling was carried out using a system consisting of Waters 1525 Binary Pump with a 7725i Rheodyne injection port, a Kromega Solvent Degasser, Waters 2998 Photodiode Array Detector, and a Luna Omega polar C18 (5 µm, 250 × 4.6 mm, Phenomenex). ACN (solvent A) and H_2O + 0.1% TFA (solvent B) were used for the gradient elution of the analytes with a steady flow rate of 0.5 mL/min with an injection volume of 7 µL. For the main separation and purification, the same previous RP-HPLC setup was recalled using a Phenomenex Kinetex PFP column (5 µm, 4.6 × 250 mm); 1 mL/min flow rate, and UV monitoring at 211, 250 and 280 nm.

1.2 NMR Spectroscopy

For nocapeptin A (**1**) and longipeptin A (**3**), 1D and 2D NMR spectra were measured at 297 K on a Bruker Avance III HDX spectrometer (700, 176 and 71 MHz for ^{1}H, ^{13}C and ^{15}N NMR, respectively) in d_3-CH_3OH/H_2O (96:4), equipped with a 5 mm Prodigy TCI CryoProbe head. The NMR spectra were processed with TopSpin 3.5 and MestReNova 12.0.4 and calibrated to the residual solvent signals ($\delta_{H/C}$ 3.31/49.15). Mixing times were 80 ms for ^{1}H-^{1}H TOCSY and 300/500 ms for ^{1}H-^{1}H NOESY, respectively. Band-Selective constant time ^{1}H-^{13}C HMBC spectra were recorded to dissect the peptide carbonyl region.^{15}N unreferenced chemical shifts were reported in ppm (spectrometer default values). All ^{1}H-^{13}C LR-HSQMBC experiments were performed according to Williamson *et al*.[2] employing an adiabatic broadband ^{13}C decoupling sequence. For nocapeptin A (**1**), we recorded LR-HSQMBC spectra with an optimized delay Δ for a long-range coupling constant of 1, 2 and 4 Hz. The obtained spectra were processed as outlined by Oberlies and coworkers.[3]

1.3 Bacterial Strains

Nocardia terpenica IFM 0406 was obtained from the Medical Mycology Research Center (MMRC) culture collection, Chiba University, Chiba, Japan, while *N. terpenica* IFM 0706 (DSM 44935), and *Longimycelium tulufanense* CGMCC 4.5737 (DSM 46696) were purchased from the DSMZ (German collection of microorganisms and cell cultures).

1.4 Bioinformatics

The initial bioinformatic analysis was carried out using antiSMASH 5.1[4] and RODEO[5] to detect the putative biosynthetic gene clusters (BGCs) of nocapeptins from *Nocardia terpenica* and longipeptins from *Longimycelium tulufanense*. Using RODEO (https://rodeo.igb.illinois.edu) annotation, the assignment of the possible functions of each biosynthetic gene was achievable. The retrieval of further homologous nocapeptins was also facilitated by a manual BLAST-P query of the nocapeptin precursor peptide (NopA) against the non-redundant NCBI protein database in tandem with RODEO.

LassoHTP
The software LassoHTP was downloaded from github (https://github.com/ChemBioHTP/LassoHTP/) and implemented on a Unix-system and the parameters ´seq´ = GWYGSQWDGPTGQRD, ´ring_length´ = 8 and ´upper_plug´ = 5 were used to describe the input file.

AlphaFold Structural Prediction of LopH
AlphaFold was used to predict the structure of LopH with each of the five trained model parameters.[6] The multiple sequence alignment (MSA) generation and the AlphaFold predictions were enabled by ColabFold, a publicly available Jupyter notebook,[7] on a Google Colab GPU cluster.

Docking S-adenosyl-L-homocysteine (SAH) to AlphaFold-predicted LopH
The highest ranking AlphaFold structure of LopH was aligned using PyMOL Align using the top DALI[8] match as the template: human 5,10-methylenetetrahydrofolate reductase (PDB code: 6fcx, chain A), which was co-crystallized with SAH.[9] The resulting coordinates from the alignment of LopH and SAH were then utilized for molecular docking using the Molecular Operating Environment 2020.0901 (MOE) software (Chemical

Computing Group Inc.). This initial LopH-SAH structure was protonated with the Protonate 3D module of MOE.[10] Specifically, the generalized Born/volume integral (GB/VI) approach[11] was utilized to model hydration with electrostatic interactions described with Coloumb's law with a 15 Å cutoff and dielectric of 2 inside of the protein. A dielectric constant of 80 was used to model the implicit solvent. The utilized temperature was 300K, pH was 7.0, and salt concentration was 0.1 mM. The van der Waals interactions were set with the 800R3 potential and a cutoff of 10 Å. After protonation, partial charges were assigned using the CHARMM27 forcefield.[12] Energy minimization was then performed with the protonated ligand structure with convergence criteria of RMS 0.00001 kcal/mol/ Å2, with a constraint of rigid water molecules. All protein structures were visualized with either PyMOL or UCSF Chimera.[13]

Generating Ligand Interaction Network using LigPlot

The ligand interaction networks for SAH in AlphaFold-predicted LopH docked with SAH, and crystallized human 5,10-methylenetetrahydrofolate reductase (PDB code: 6fcx), were generated using LigPlot plus using standard parameters.[14]

Retrieval of cytochrome P450 proteins associated with RiPP precursor recognition elements (RRE)

Using the genome-mining tool RRE-Finder operating in precision mode[15], we first compiled an updated list of all identifiable RRE domains using the most recent release of UniProt (2022_04; released on 12-Oct-2022). Members of protein family PF00067 (cytochrome P450)[16] that co-occur within 10 open-reading frames of any detected RRE domain were collected (n = 932). Removal of identical sequences (including cases were two distinct strains produce identical proteins) yielded a set of 883 RRE-associated cytochrome P450 proteins.

Sequence Similarity Network (SSN), Similarity and Identity Matrices Generation

The cytochrome P450 SSN (Figure S10) was generated using EFI-EST (http://efi.igb.illinois.edu/efi-est) using an alignment score of 99 and repnode 100 parameters to represent any sequences that are 100% identical as a single node in the network.[17] The sequences used were from the above-described RRE-associated cytochrome P450 dataset (n = 883 proteins), UniProt BLAST-P hits to NopF (n = 1000 proteins, LopG as a presumed Trp hydroxylase was not retrieved in the list supporting its 31/32% identity, and 42/43% similarity scoring to NopF/LopF respectively), a previously reported set of atropopeptide- and biaryltide-associated cytochrome P450 proteins,[18] and cytochrome P450 proteins in lasso peptide BGCs predicted in this study (Figure S9). Accession identifiers of cytochrome P450 proteins annotated in this network are provided in Supplementary Dataset 1 and 2.

NopF, LopF, LopG, annotated cytochrome P450 proteins in the SSN (Figure S49), and the five top BLAST-P hits to NopF (UniProt) were aligned using the E-INS-i algorithm in MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/index.html).[19] The alignment was then submitted to SIAS (http://imed.med.ucm.es/Tools/sias.html) to generate similarity and identity matrices using the BLOSUM62 matrix.

1.5 Isotopic Labeling Experiments

N. terpenica IFM 0406 was revived on Brain Heart Infusion (BHI) broth agar plates (2%) at 37 °C. Colony growth was detected after 3 d of cultivation. Using fresh spores of IFM 0406, triplicates of seed cultures were prepared in Brain heart infusion (BHI) media, BHI broth 3.7%, (80 ml) in 250 ml baffled Erlenmeyer flasks at 37 °C with 150 rpm for 4 d. Starter cultures (0.4 mL) were used to inoculate 20 mL of the production medium, consisting of a modified R4 medium [glucose 0.5%, yeast extract 0.1%, $MgCl_2$•$6H_2O$ 0.5%, $CaCl_2$•$2H_2O$ 0.2%, K_2SO_4 0.1%, casamino acids 0.05%, L-proline 0.07%, L-valine 0.12%, TES (N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid) 0.28%, and 50 µl trace elements solution ($ZnCl_2$ 40 mg/L, $FeCl_3$•$6H_2O$ 20 mg/L, $CuCl_2$•$2H_2O$ 10 mg/L, $MnCl_2$•$4H_2O$ 10 mg/L, $Na_2B_4O_7$•$10H_2O$ 10 mg/L and $(NH_4)_6Mo_7O_{24}$•$4H_2O$ 10 mg/L)] in a 50 mL Erlenmeyer baffled flask at 32 °C with 150 rpm. The supplementation of the labeled amino acids, [$^2H_7$] L-tyrosine and [$^2H_8$] L-tryptophan into to the production medium was adjusted to a 2 mM as a final concentration of each. After 6-7 d, the cell-free supernatants were prepared by centrifugation

and extracted twice with 50 ml of *n*-BuOH. The organic phases were combined, dried *in vacuo*, dissolved in MeOH, and submitted to LC/HRESI-MSMS.

1.6 Large Scale Fermentation, Extraction Scheme, Fractionation, and Isolation

N. terpenica IFM 0406 and *Longimycelium tulufanense* were grown using the nutrients recipe and the growth parameters as previously described in the above section, except the 250 ml Erlenmeyer baffled flasks were filled with 120 ml of the production medium. A 25 L cultivation was done in 6 batches in the case of nocapeptin A (**1**) and around 50 L in 12 batches for longipeptin A (**3**). To remove the cells, the liquid cultures were centrifuged twice in a Thermo Scientific Heraeus Multifuge 4KR centrifuge at 4,000 × g at 4 °C for 30 min. Subsequently and using n-BuOH (1:1), the cell free supernatants (SN) were extracted twice. Under reduced pressure, the *n*-BuOH extracts were evaporated affording the crude extracts (BuOH-SN extracts), which were resuspended in methanol followed by centrifugation to get rid of debris prior to LC/MS analysis, HPLC profiling, and column chromatography. Fractionation of the BuOH-SN extracts was accomplished either through a VLC- or an open column chromatography system by stepwise elution of H_2O mixed with methanol with a decreased polarity fashion, shifting from H_2O/MeOH (90:10) to pure MeOH in ten steps (750 mL per fraction). Guided by LCMS (see section 1.1), the 60% MeOH fraction of the *N. terpenica* BuOH-SN extract and the 80% plus 90% MeOH fraction of the *L. tulufanense* BuOH-SN extract were prioritized, redissolved in MeOH/H_2O and filtered prior to HPLC separation. For the purification of **1** and **3** PFP-HPLC (Phenomenex Kinetex pentafluorophenyl (PFP) column, 4.6 × 250 mm, 5 µm) and UV monitoring (at 215, 254 and 280 nm) was employed and performed in gradient elution mode with deionised H_2O containing 0.1% TFA (A) and MeCN (B) as mobile phases, however different gradient elution schemes were applied in each case.

For isolation of nocapeptin A (**1**), the gradient started at 10% B, increasing the percentage of B in 5 min to 20%, from 20% to 25% in 3 min, from 25% to 30% in 12 min and 30% to 100% in 1 min, followed by isocratic elution of 100% B for 4 min (1.1 mL/min flow rate). Nocapeptin A (**1**) eluted at t_R = 18 min and yielded ≈ 17 mg.

For the isolation of pure longipeptin A (**3**), the gradient started at 20% B, increasing the percentage of B in 10 min to 60%, followed by isocratic elution of 60% B for 5 min and subsequently, increasing the percentage of B to 100% in 1 min, followed by isocratic elution of 100% B for an additional 2 min (1.0 mL/min flow rate). Longipeptin A (**3**) eluted at t_R = 9.5 min and yielded ≈ 2 mg.

1.7 Biological and Physical Assays

Antibacterial assay

The minimal inhibitory concentration (MIC) was determined as described previously[20] in cation-adjusted Mueller–Hinton medium according to the standards and guidelines of the Clinical and Laboratory Standards Institute.[21] A 2-fold serial dilution of the test compound was prepared in microtiter plates and seeded with a final test bacterial inoculum of 5 × 10^5 colony-forming units (CFU)/ml. After an overnight incubation at 37 °C, the MIC was read as the lowest compound concentration preventing visible bacterial growth. The strain panel included representative "**ESKAPE**" human pathogenic bacterial species. Specifically, the following strains were used: ***E****nterococcus faecium* BM 4147-1, ***S****taphylococcus aureus* ATCC 29213, ***K****lebsiella pneumoniae* ATCC 12657, ***A****cinetobacter baumannii* 09987, ***P****seudomonas aeruginosa* ATCC 27853, and ***E****nterobacter aerogenes* ATCC 13048. Additional MIC testing was performed on: *Bacillus subtilis* 168, *Enterococcus faecalis* ATCC 29212, *Escherichia coli* ATCC 25922, *Escherichia coli* HN 818, *Micrococcus luteus* ATCC 4698, *Neisseria gonorrhoeae* ATCC 19424, *Neisseria gonorrhoeae* S1441, and *Mycobacterium smegmatis* mc^2 155 ATCC 700084.

Cytotoxicity assay (one dose NCI-60 panel)
Nocapeptin A (**1**) was selected for the anticancer drug screening service as a part of the Developmental Therapeutics Program at the National Cancer Institute (NCI). *In vitro* tumor growth inhibitory effects were explored using a standard protocol with a single high dose test against a panel comprising 60 human cancer cell lines.[22] Results for cell line NCI-H23 are omitted based on NCI staff informing that authentication was not performed during the screening time frame.

Thermal stability assay
To investigate the thermal stability of nocapeptin A (**1**) was dissolved in ddH_2O, and a solution of 10 µg/100 µL purified lasso peptide was incubated at 95°C for 1, 2, 4, or 8 hr. Samples were cooled to 4°C and were subsequently analyzed via LR-HPLC-MS using the following MeCN/H_2O (0.5% formic acid) gradient, increasing the MeCN portion from 10% to 50% over 16 min and from 50% to 100% over 6 min, followed by isocratic elution at 100% MeCN for 3 min (Phenomenex Luna C18 (2)−100 Å column, 2 × 250 mm, 5 µm; 0.2 mL/min flow rate; with total ion current [positive mode Q3] and photodiode array monitoring). The assay was conducted in triplicate. As reference, untreated nocapeptin A was analyzed.

Proteolytic stability assay
Stability against proteolytic degradation of nocapeptin A (**1**) was investigated by incubating 10 µg of the purified lasso peptide with two different proteases. The carboxypeptidase Y assays were performed with 0.5 U carboxypeptidase Y in a buffer containing 50 mM MES and 1 mM $CaCl_2$ at a pH of 6.75 for 4 hr at 25°C. 10 µL of a 1 mg/mL solution of **1** was mixed with 10 µL of a 0.1 U/mL protease solution.The chymotrypsin assays were performed with 0.5 µg protease in a buffer containing 100 mM Tris-HCl and 10 mM $CaCl_2$ at a pH of 8.5 for 4 hr at 25°C. Subsequently, the samples were analyzed via LR-HPLC-MS using the same gradient as in the thermal stability assay. As reference, untreated nocapeptin A was analyzed.

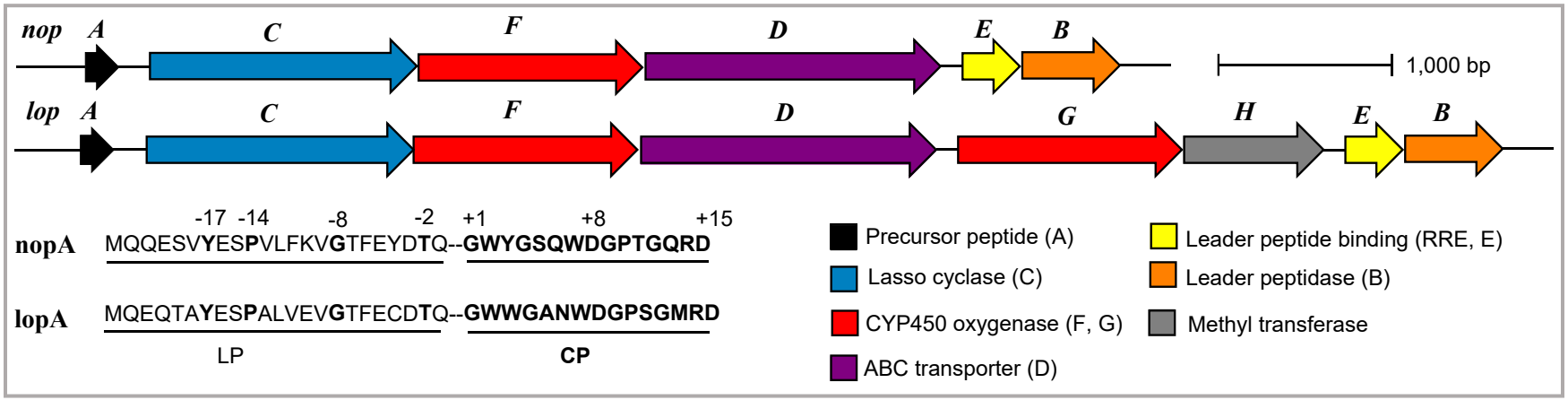


Figure S1. (Top) Putative biosynthetic gene clusters (BGCs) of nocapeptin, *nop* and longipeptin, *lop* from *Nocardia terpenica* and *Longimycelium tulufanense*, respectively. (Bottom left) Products of *nop*A and *lop*A, respectively, including prediction of the cleavage site between the leader peptide (LP) and core peptide (CP).

Table S1. Putative functions of proteins from the *nop* and *lop* BGCs using RODEO. Length is given as number of amino acids (aa).

RODEO Analysis						
	Protein	**NCBI Accession ID**	**Length [aa]**	**PFAM**	**Description**	**E-value**
nop BGC	NopA	WP_195116738.1	43	none	none	none
	NopC	WP_171983240.1	547	PF00733	Asparagine synthase	$4.20E^{-36}$
	NopF	WP_156674500.1	404	PF00067	Cytochrome P450	$4.00E^{-50}$
	NopD	WP_171983239.1	575	PF00005	ABC transporter	$9.40E^{-96}$
	NopE	WP_067588831.1	86	PF05402	Stand alone lasso RRE	$2.10E^{-27}$
	NopB	WP_082871451.1	157	PF13471	Transglutaminase	$7.60E^{-24}$
lop BGC	LopA	WP_189061731.1	38	none	none	none
	LopC	WP_194500064.1	536	PF00733	Asparagine synthase	$1.40E^{-38}$
	LopF	WP_189061729.1	404	PF00067	Cytochrome P450	$1.10E^{-50}$
	LopD	WP_189061728.1	596	PF00005	ABC transporter	$1.20E^{-94}$
	LopG	WP_189061727.1	406	PF00067	Cytochrome P450	$1.00E^{-46}$
	LopH	WP_189061726.1	191	none	Putative Methyl transferase	none
	LopE	WP_189061725.1	85	PF05402	Stand alone lasso RRE	$3.50E^{-27}$
	LopB	WP_189061724.1	137	PF13471	Transglutaminase	$1.80E^{-26}$

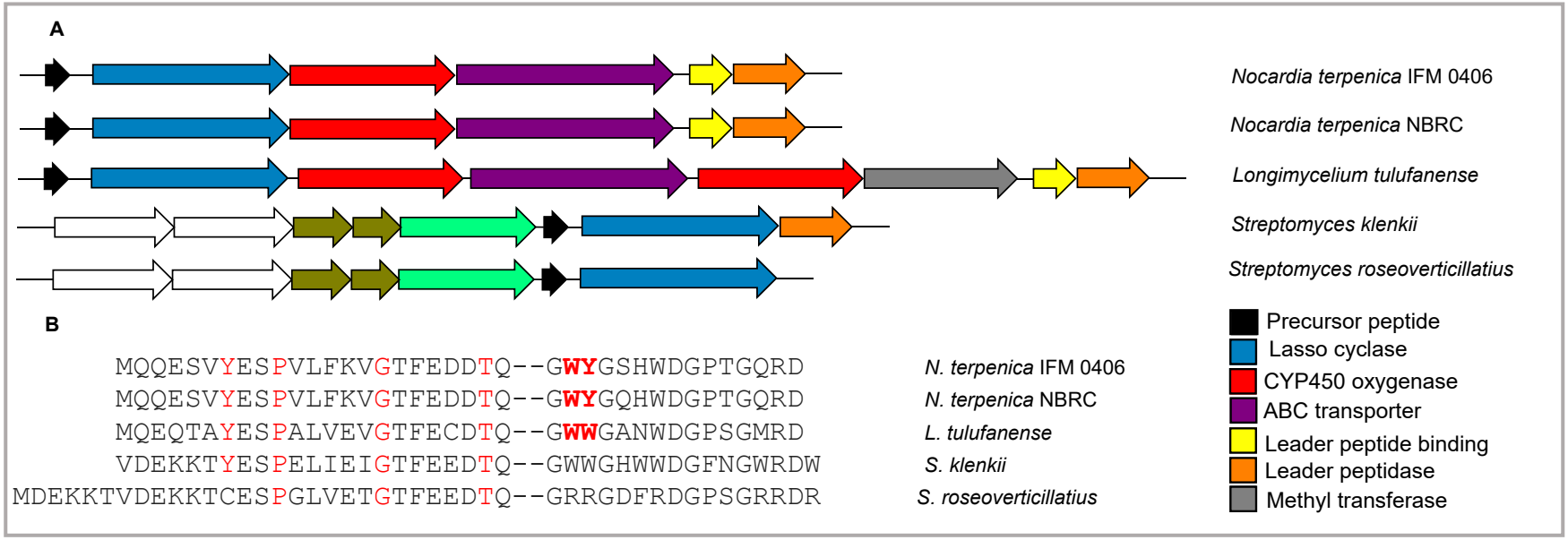


Figure S2. Putative list of homologous lasso peptides BGCs using BLAST-P query of Nop B/C. (**A**) Predicted biosynthetic gene clusters and bacterial strains (**B**) sequence alignment of their lasso peptide precursors. The red-colored residues in the leader peptide portion are predicted to be either the recognition sequence (YxxP) or the conserved residues in lasso peptide precursors (G and T). The red-colored bolded residues in the core peptide portion (WY and WW) are predicted to be the motifs of the biaryl crosslink.

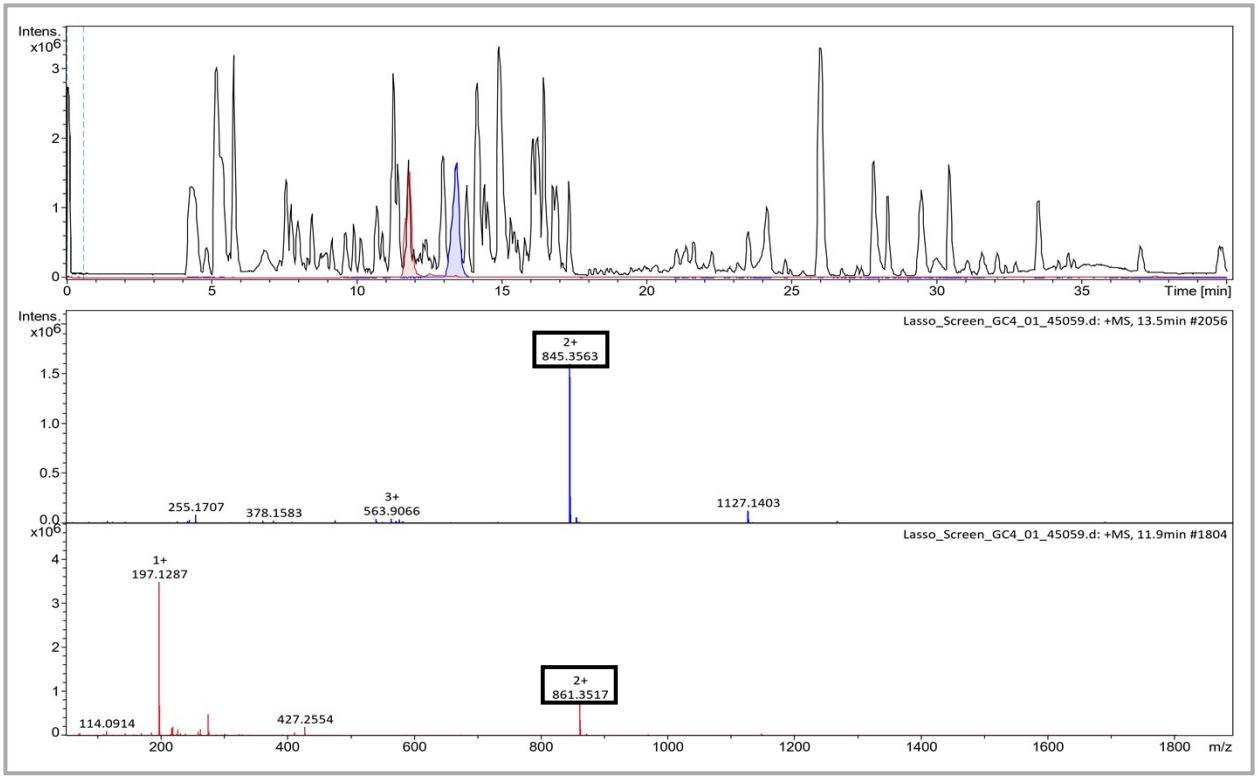


Figure S3. LCMS profile of modified R4 medium-based cultivation highlighting the production of nocapeptin A (**1**) [845 Da] and nocapeptin B (**2**) [861 Da].

Meas. *m/z*	$[M+2H]^{2+}$ Formula	Theoretical *m/z*	Absolute error [ppm]	rdb	mSigma
845.356405	$C_{75}H_{98}N_{22}O_{24}$	845.355667	0.9	39.0	5.6
	$C_{74}H_{102}N_{18}O_{28}$	845.354999	1.7	34.0	8.3
	$C_{77}H_{110}N_{8}O_{34}$	845.355673	0.9	28.0	8.5
	$C_{76}H_{114}N_{4}O_{38}$	845.355004	1.7	23.0	17.2
	$C_{73}H_{106}N_{14}O_{32}$	845.354330	2.5	29.0	18.5
	$C_{75}H_{118}O_{42}$	845.354335	2.4	18.0	27.5
861.351721	$C_{75}H_{98}N_{22}O_{26}$	861.350582	1.3	39.0	13.4
	$C_{77}H_{110}N_{8}O_{36}$	861.350587	1.3	28.0	19.7
	$C_{74}H_{102}N_{18}O_{30}$	861.349913	2.1	34.0	22.8
	$C_{76}H_{114}N_{4}O_{40}$	861.349919	2.1	23.0	30.2

Figure S3A. Molecular formula predictions of nocapeptin A (**1**) [845 Da] and nocapeptin B (**2**) [861 Da]

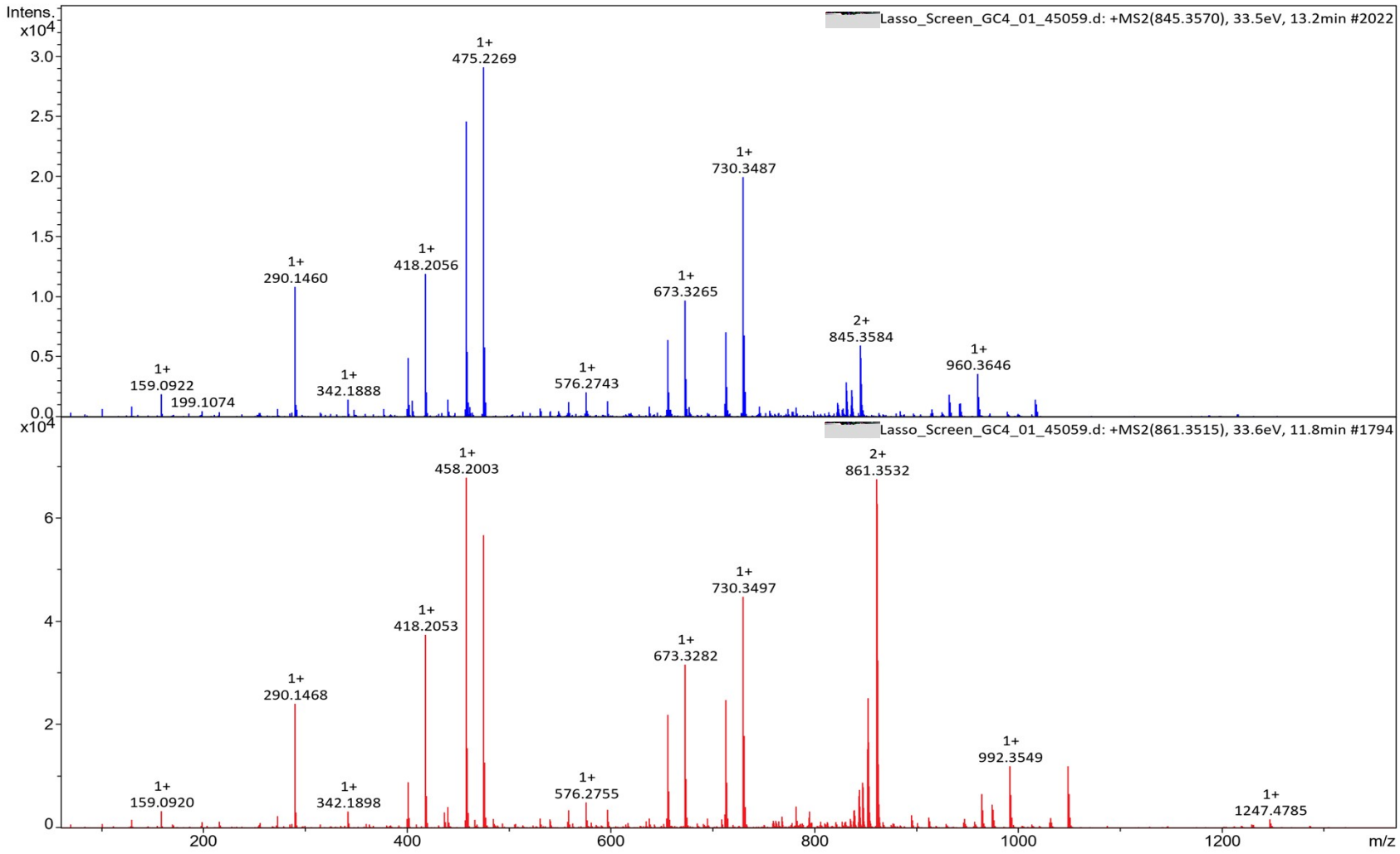


Figure S4. Comparative MS^2 spectra of nocapeptin A (**1**) [845 Da] and nocapeptin B (**2**) [861 Da]

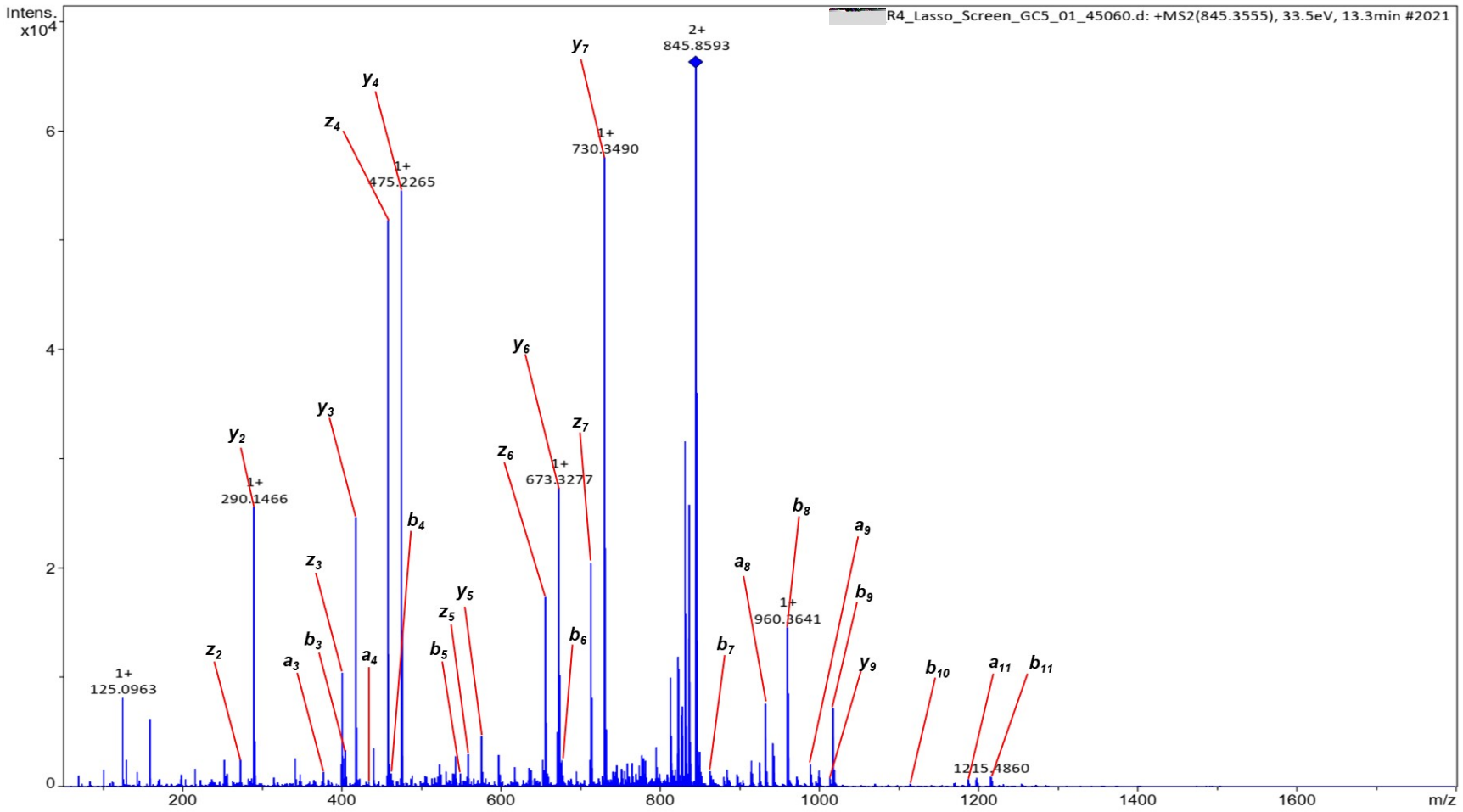


Figure S4A. Annotated MS^2 spectrum of nocapeptin A (**1**) [845 Da]

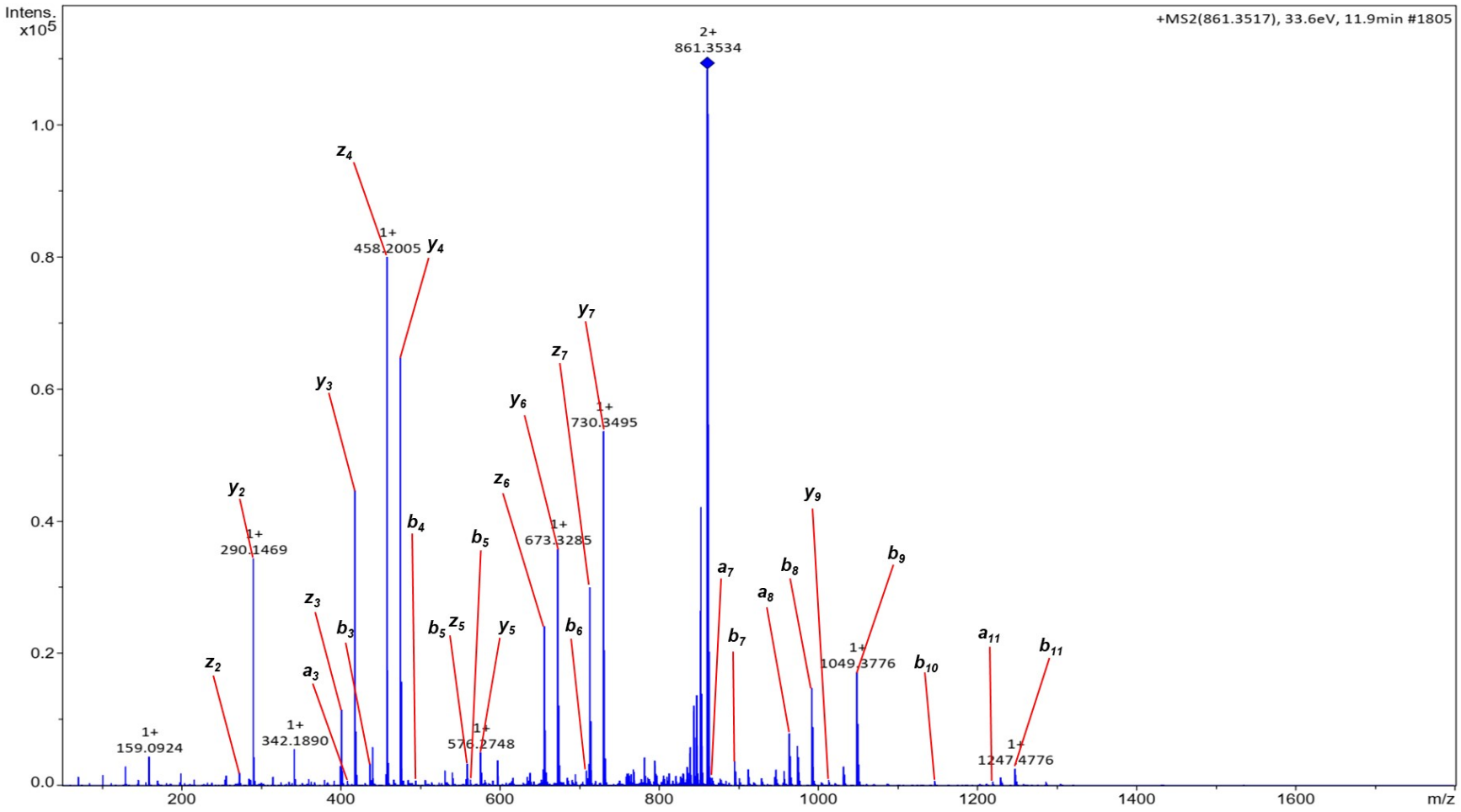


Figure S4B. Annotated MS^2 spectrum of nocapeptin B (**2**) [861 Da]

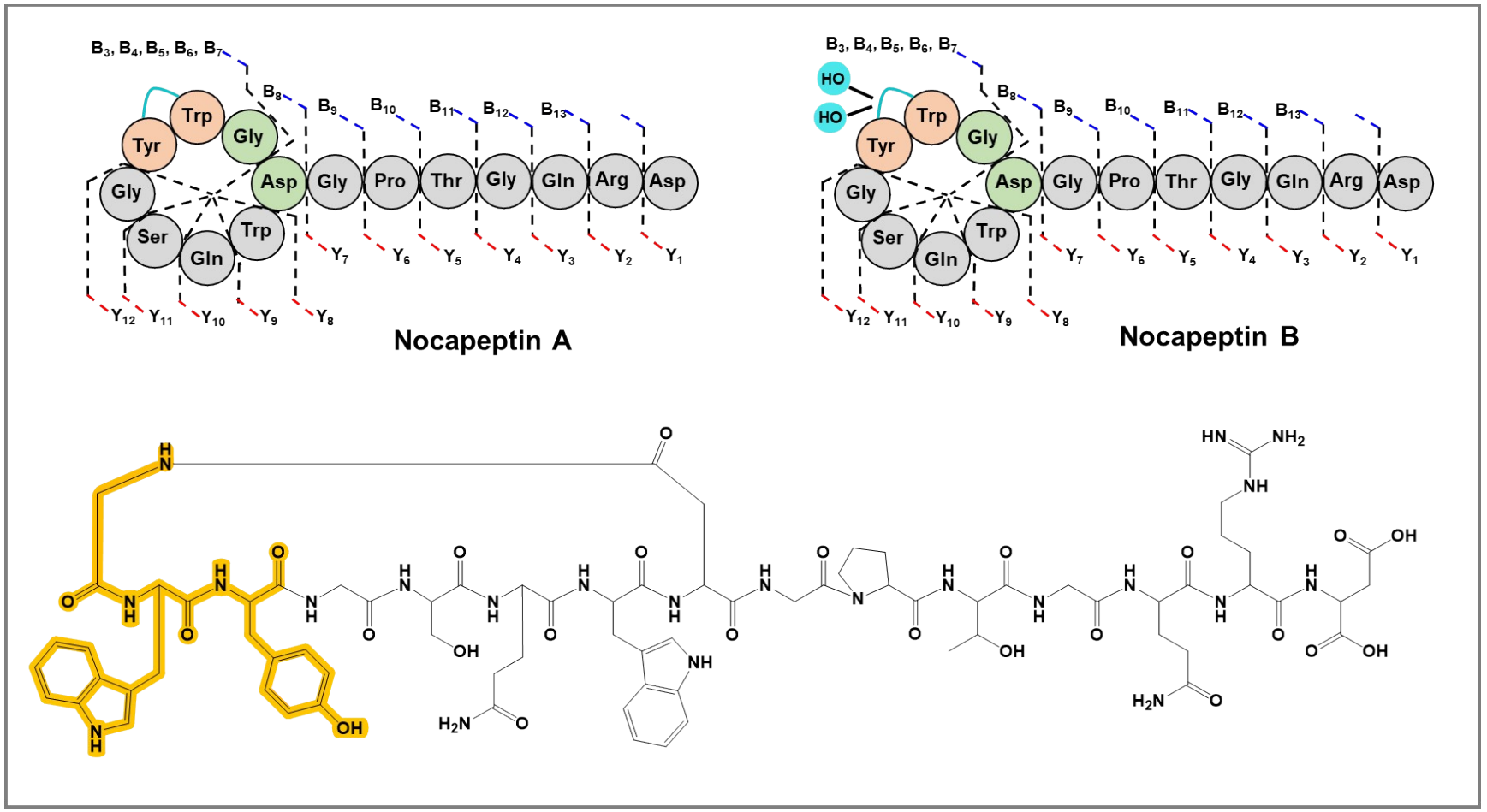


Figure S4C. Schematic structures of nocapeptins A (**1**) and B (**2**) illustrating their annotated fragments.

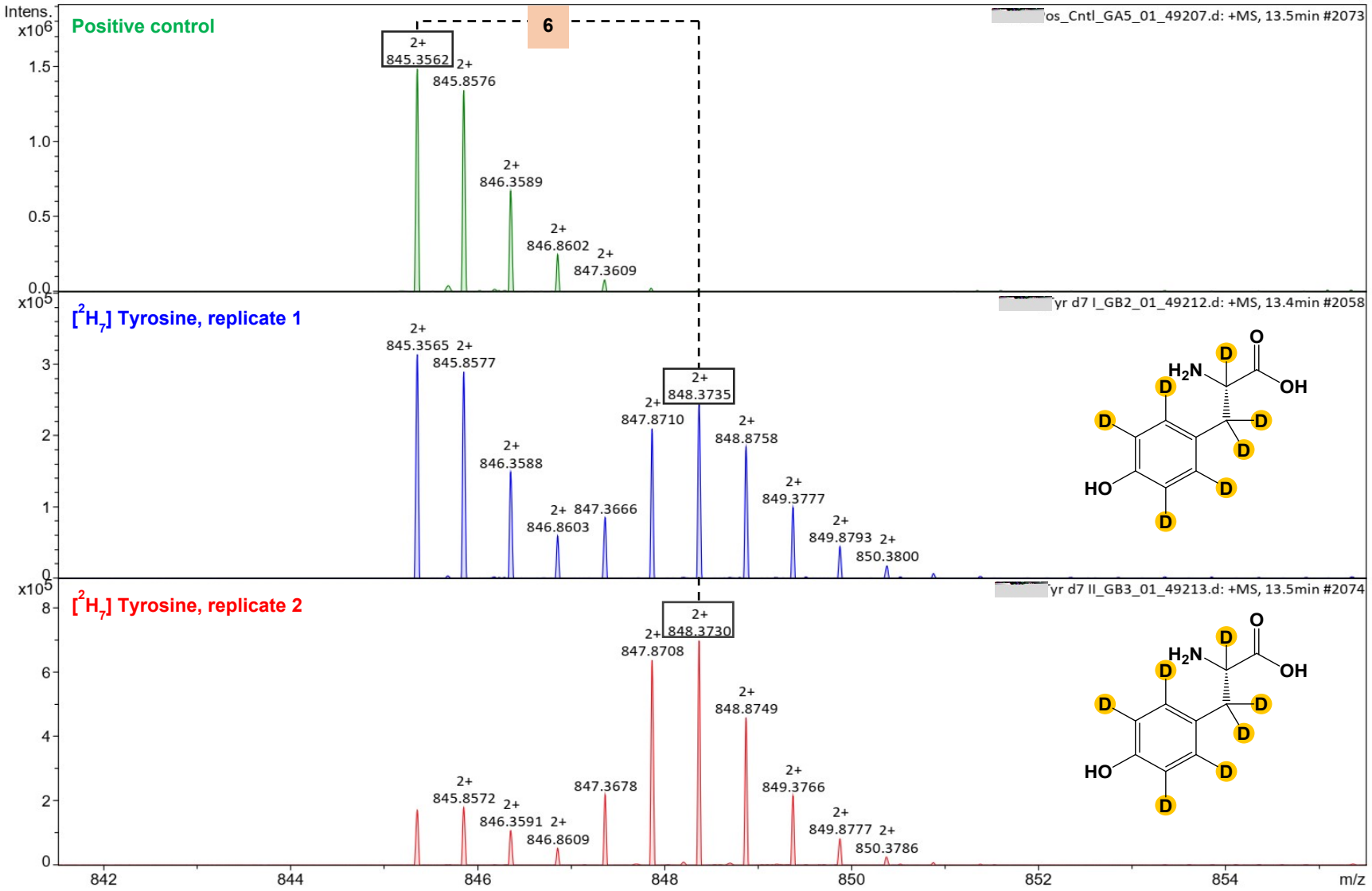


Figure S5. Comparative MS[1] spectra of nocapeptin A (**1**) and its [$^2H_7$] L-tyrosine-labeled version

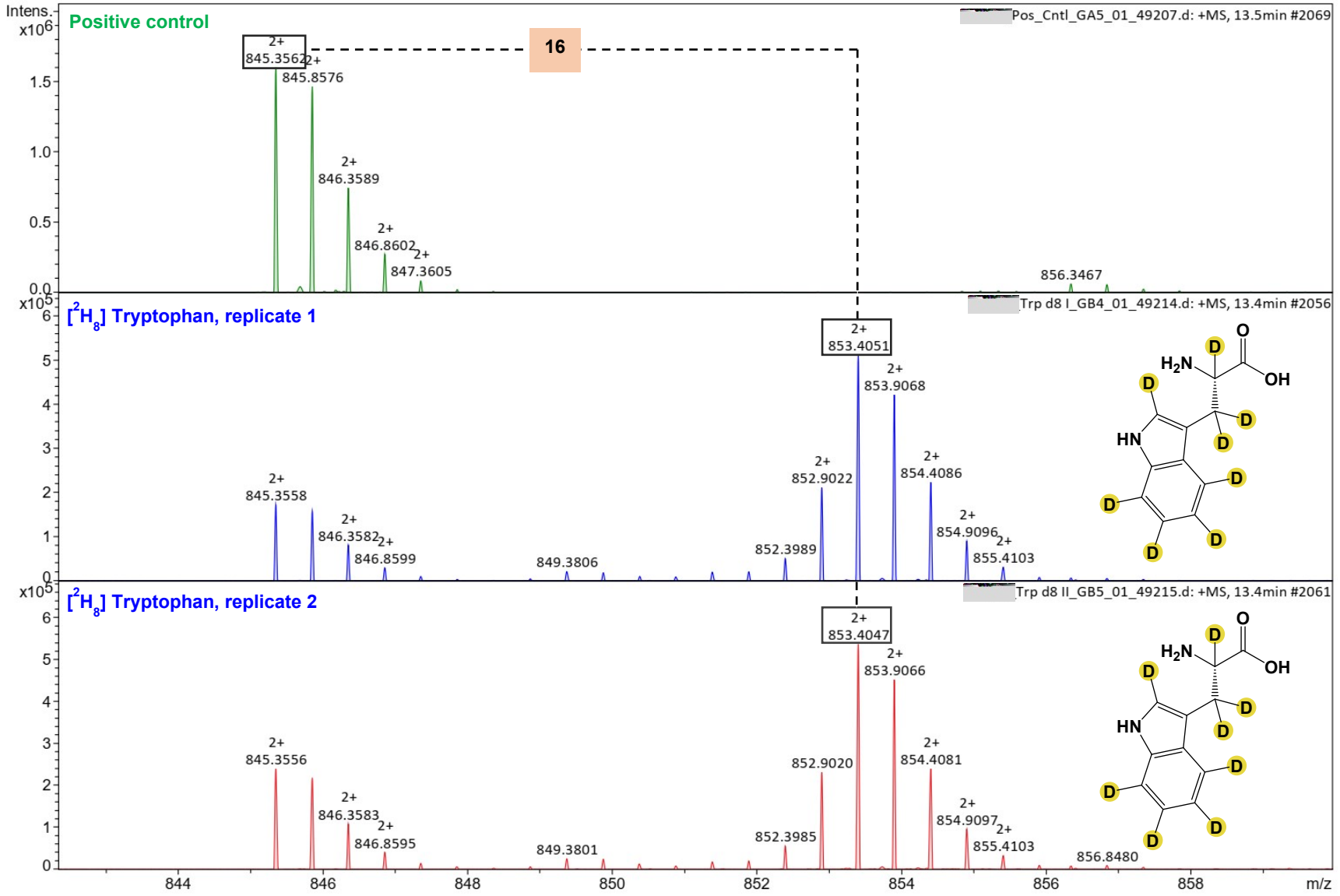


Figure S6. Comparative MS[1] spectra of nocapeptin A (**1**) and its [$^2H_8$] L-tryptophan-labeled version

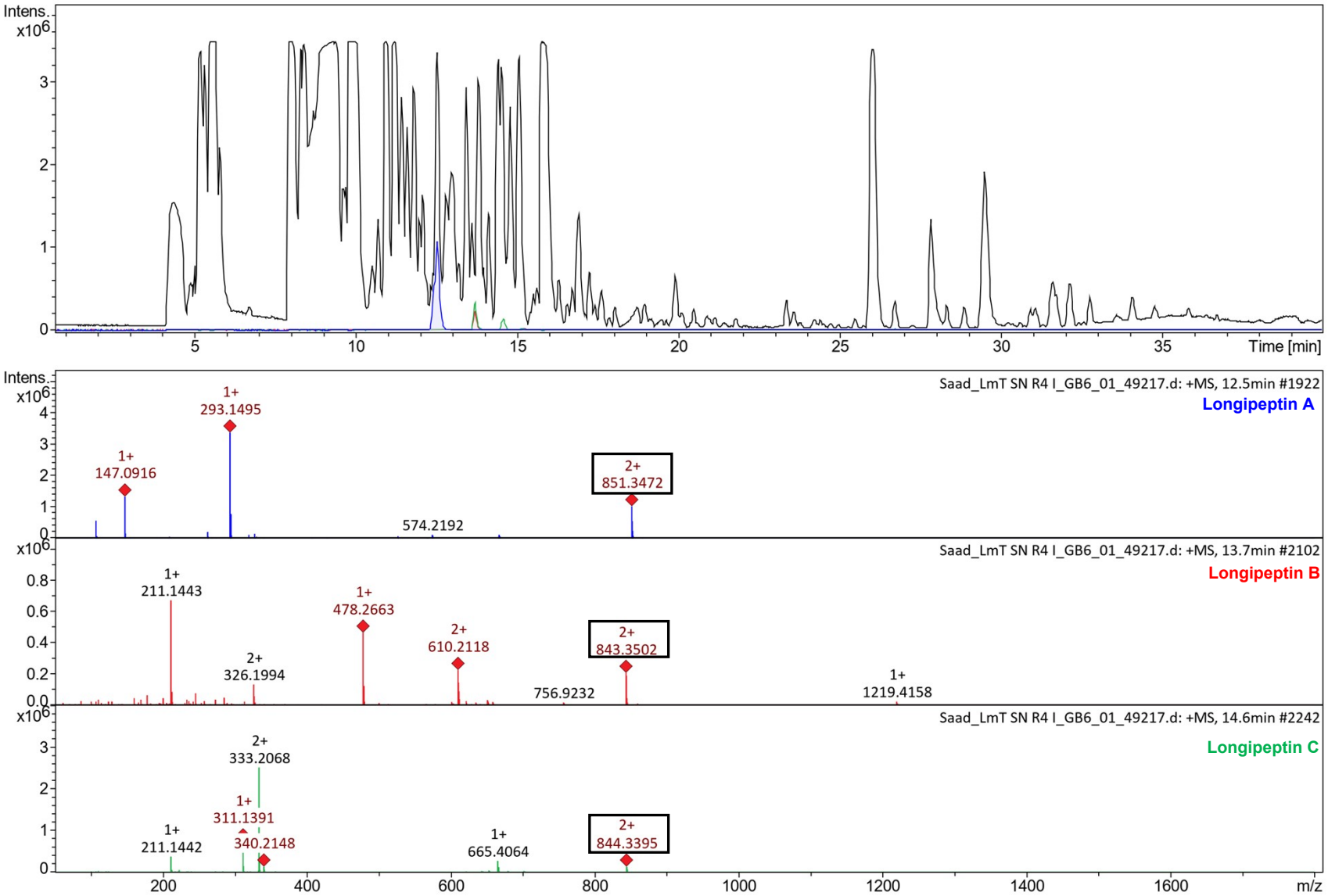


Figure S7. LCMS profile of modified R4 medium-based cultivation highlighting the production of longipeptin A (**3**), longipeptin B (**4**) and longipeptin C (**5**)

Meas. *m/z*	[M+2H]$^{+2}$ Formula	Theoretical *m/z*	Absolute error [ppm]	rdb	mSigma
851.347216	$C_{76}H_{98}N_{22}O_{22}S$	851.346788	0.5	40.0	2.1
	$C_{78}H_{110}N_{8}O_{32}S$	851.346794	0.5	29.0	10.9
	$C_{75}H_{102}N_{18}O_{26}S$	851.346120	1.3	35.0	12.4
	$C_{77}H_{114}N_{4}O_{36}S$	851.346125	1.3	24.0	21.3
	$C_{74}H_{106}N_{14}O_{30}S$	851.345451	2.1	30.0	23.3
	$C_{76}H_{118}O_{40}S$	851.345456	2.1	19.0	32.0
843.350141	$C_{76}H_{98}N_{22}O_{21}S$	843.349331	1.0	40.0	10.0
	$C_{78}H_{110}N_{8}O_{31}S$	843.349336	1.0	29.0	12.9
	$C_{75}H_{102}N_{18}O_{25}S$	843.348662	1.8	35.0	16.6
	$C_{77}H_{114}N_{4}O_{35}S$	843.348668	1.7	24.0	23.0
844.339481	$C_{75}H_{96}N_{22}O_{22}S$	844.338963	0.6	40.0	11.2
	$C_{78}H_{104}N_{12}O_{28}S$	844.339637	0.2	34.0	12.3
	$C_{77}H_{108}N_{8}O_{32}S$	844.338969	0.6	29.0	16.8
	$C_{74}H_{100}N_{18}O_{26}S$	844.338295	1.4	35.0	17.8
	$C_{76}H_{112}N_{4}O_{36}S$	844.338300	1.4	24.0	25.7
	$C_{73}H_{104}N_{14}O_{30}S$	844.337626	2.2	30.0	27.4
	$C_{75}H_{116}O_{40}S$	844.337631	2.2	19.0	35.7

Figure S7A. Molecular formula predictions of longipeptin A (**3**) [851 Da], longipeptin B (**4**) [843 Da] and longipeptin C (**5**) [844 Da]

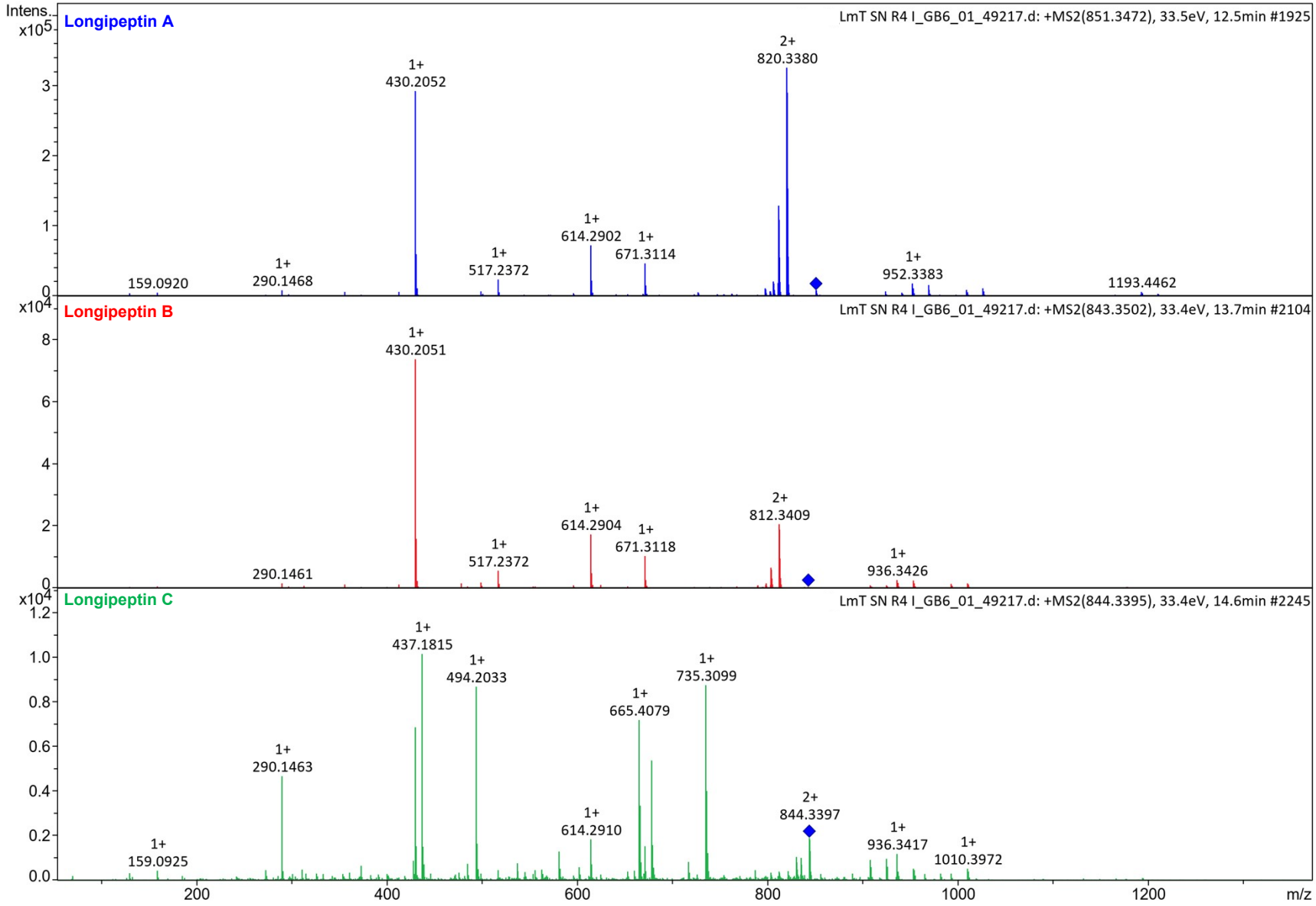


Figure S8. Comparative MS^2 spectra of longipeptins A (**3**) [851 Da], B (**4**) [843 Da] and C (**5**) [844 Da]

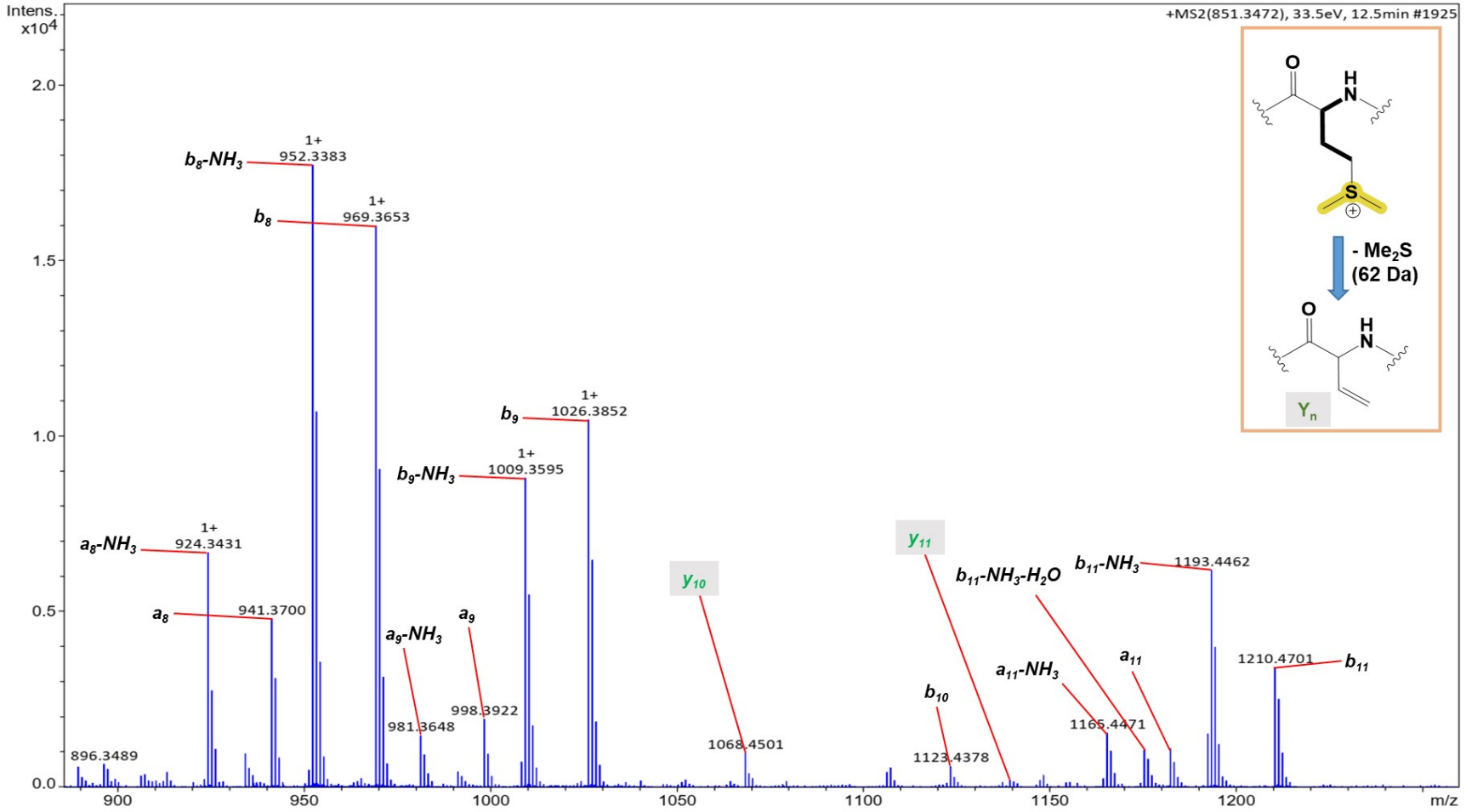


Figure S9. Annotated MS^2 spectrum of longipeptin A (**3**) [851 Da]

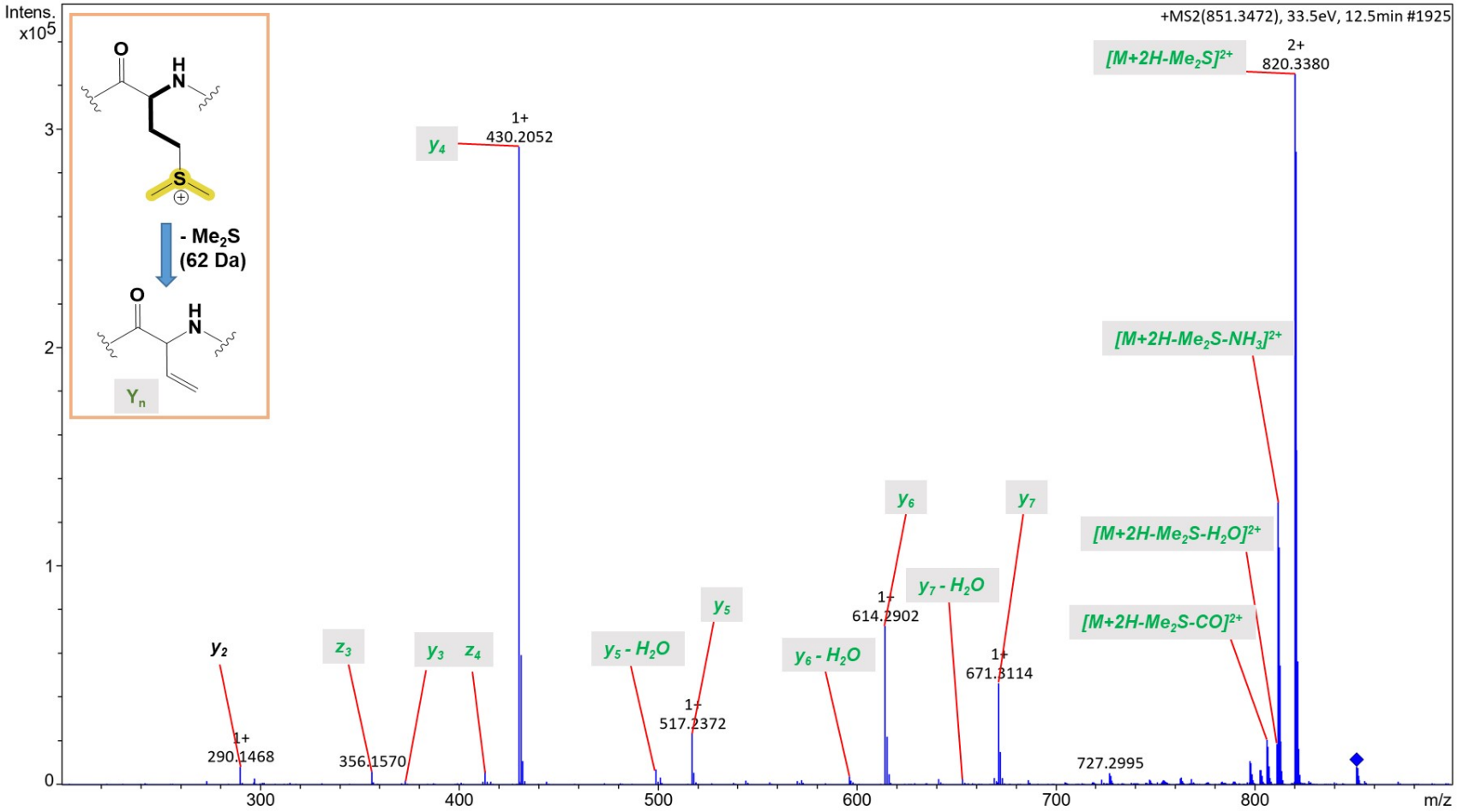


Figure S9A. Annotated MS^2 spectrum of longipeptin A (**3**) [851 Da]

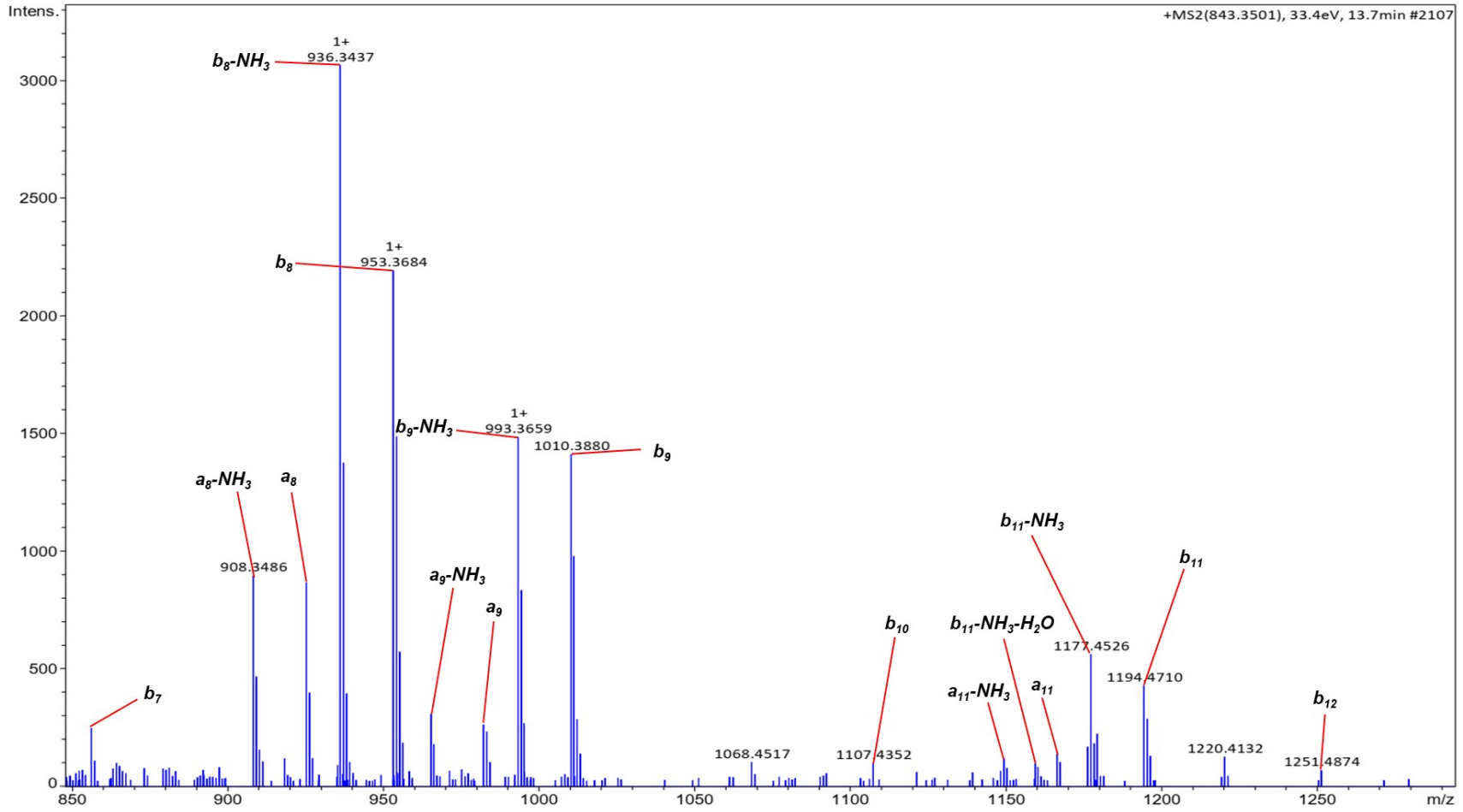


Figure S10. Annotated MS^2 spectrum of longipeptin B (**4**) [843 Da]

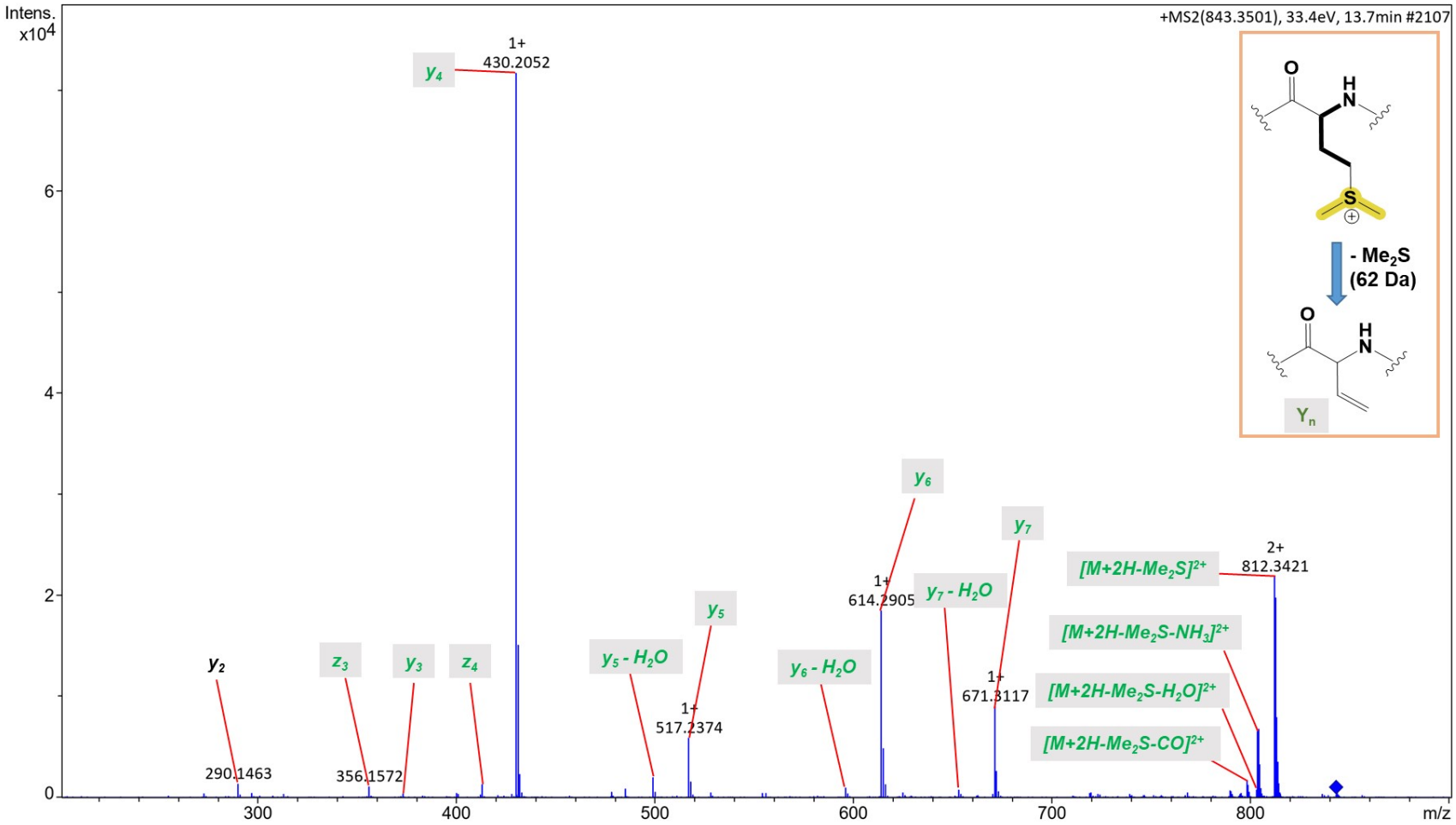


Figure S10A. Annotated MS^2 spectrum of longipeptin B (**4**) [843 Da]

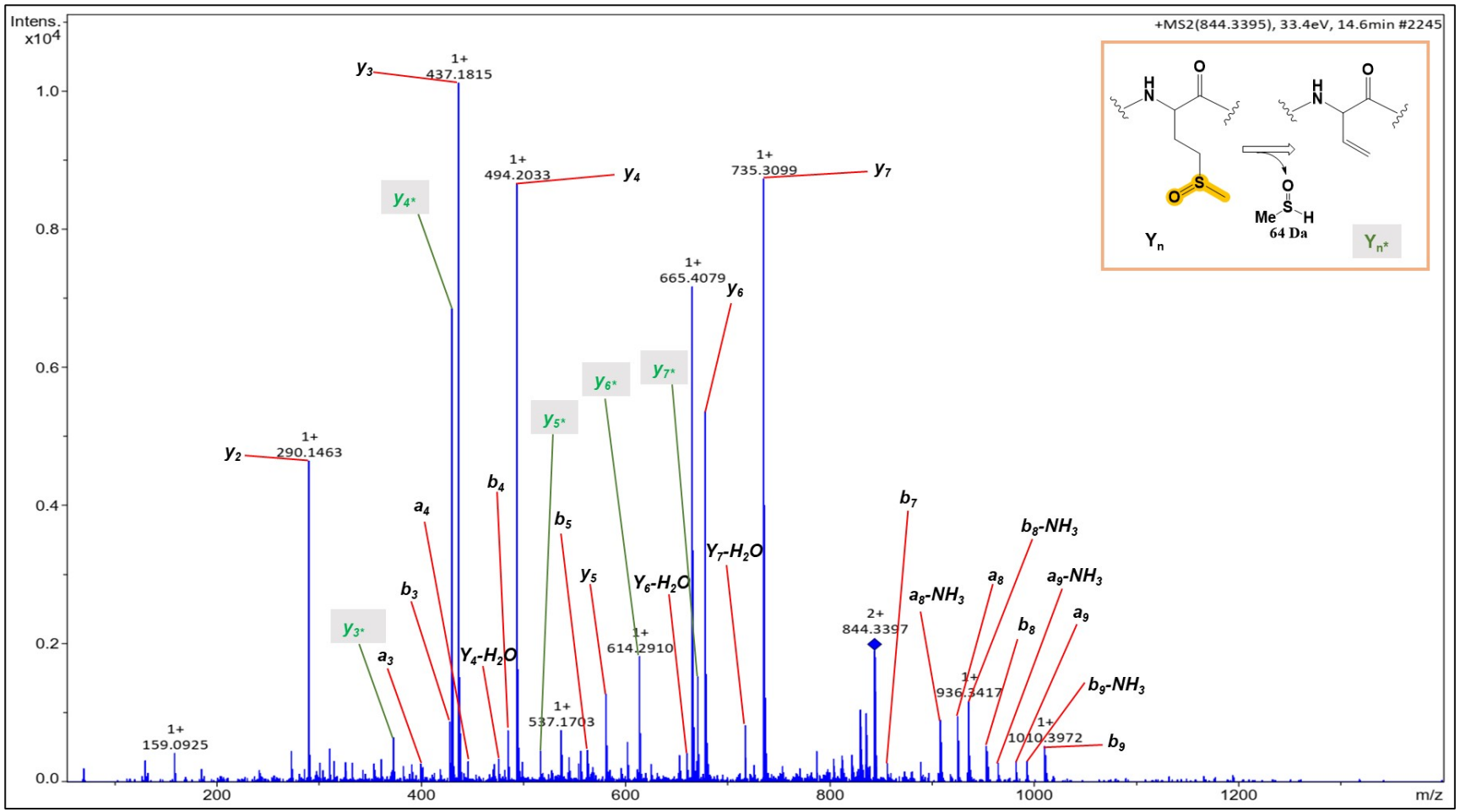


Figure S11. Annotated MS^2 spectrum of longipeptin C (**5**) [844 Da]

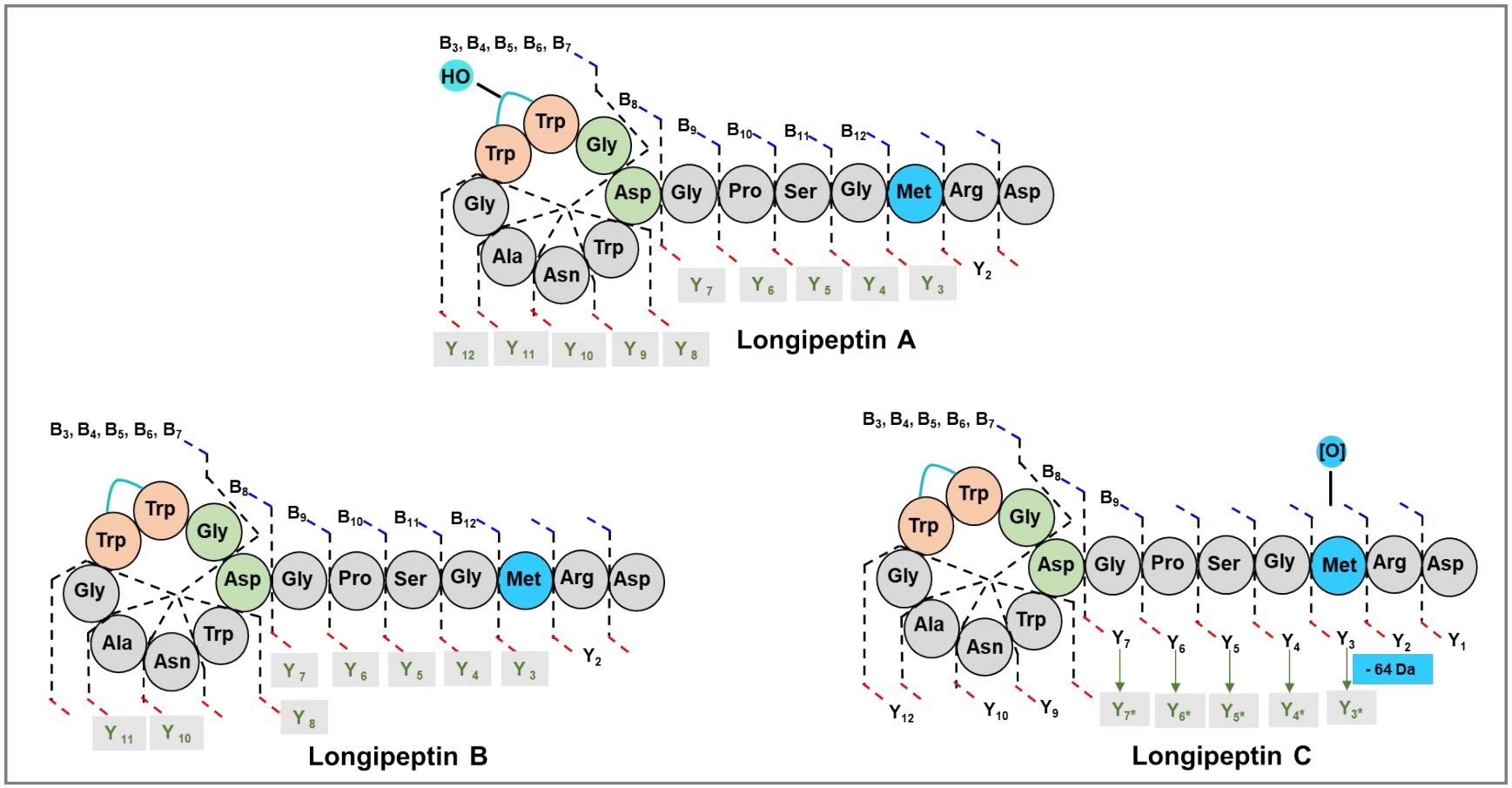


Figure S12. Schematic structures of longipeptins A, B and C (**3**-**5**) illustrating their annotated fragments

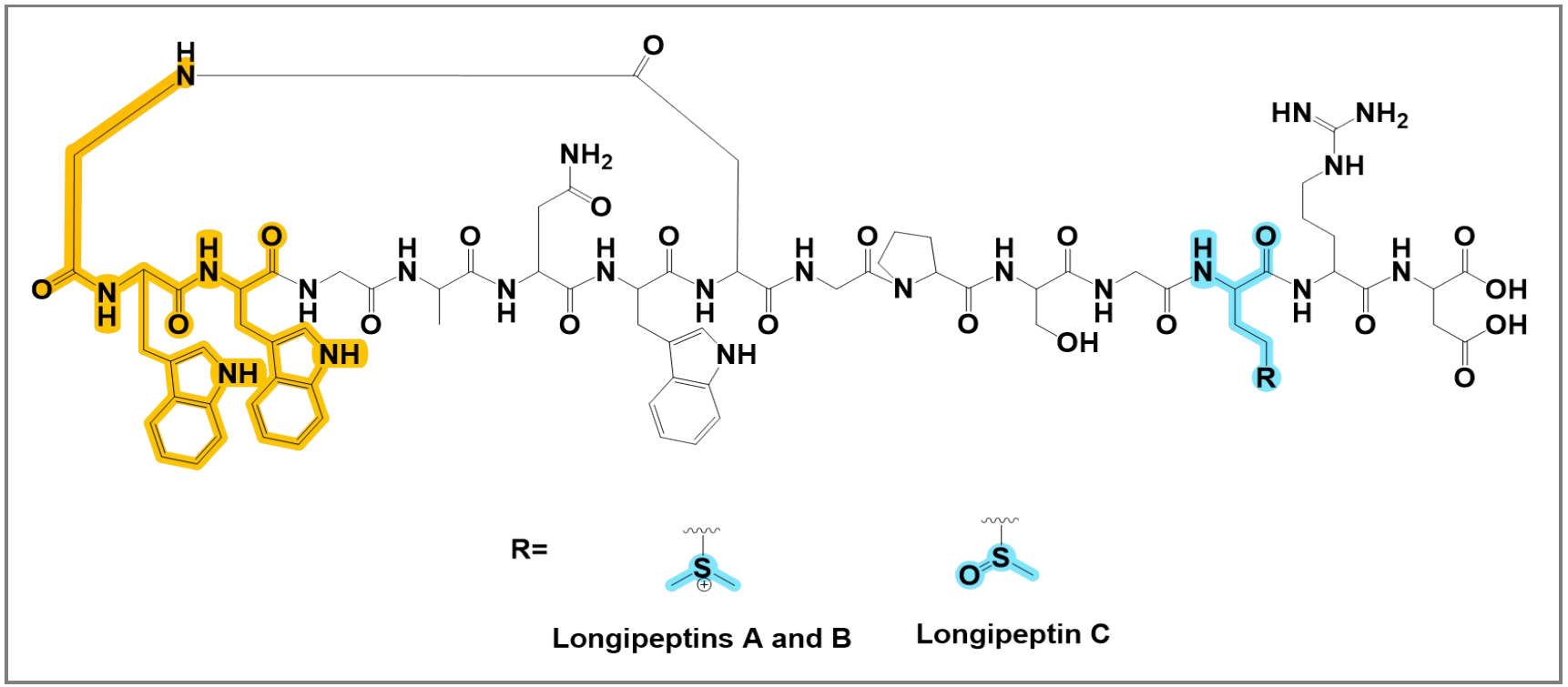


Figure S13. Detailed structures of longipeptins A, B and C (**3**-**5**) illustrating the different PTMs localities

Table S2. The assigned fragments of nocapeptin A (**1**) [845 Da], degree of unsaturation (RDB, rings and double bonds)

Fragment	Ion Formula	RDB	Calc. *m/z*	Meas. *m/z*	Error, ppm	Sequence
y_1	$C_4H_8NO_4$	2	134.0453	134.0450	2.24	D
y_2	$C_{10}H_{20}N_5O_5$	4	290.1464	290.1466	0.69	RD
z_2	$C_{10}H_{17}N_4O_5$	5	273.1199	273.1198	0.37	
y_3	$C_{15}H_{28}N_7O_7$	6	418.2050	418.2053	0.72	QRD
z_3	$C_{15}H_{25}N_6O_7$	7	401.1785	401.1784	0.25	
y_4	$C_{17}H_{31}N_8O_8$	7	475.2265	475.2265	0.00	GQRD
z_4	$C_{17}H_{28}N_7O_8$	8	458.1999	458.2002	0.65	
y_5	$C_{21}H_{38}N_9O_{10}$	8	576.2742	576.2741	0.17	TGQRD
z_5	$C_{21}H_{35}N_8O_{10}$	9	559.2476	559.2482	1.07	
y_6	$C_{26}H_{45}N_{10}O_{11}$	10	673.3269	673.3277	1.19	PTGQRD
z_6	$C_{26}H_{42}N_9O_{11}$	11	656.3004	656.3009	0.76	
y_7	$C_{28}H_{48}N_{11}O_{12}$	11	730.3484	730.3490	0.82	GPTGQRD
z_7	$C_{28}H_{45}N_{10}O_{12}$	12	713.3218	713.3223	0.70	
y_8	$C_{32}H_{51}N_{12}O_{14}$	14	827.3648	---------	---------	**D**GPTGQRD
y_9	$C_{43}H_{61}N_{14}O_{15}$	21	1013.4441	1013.4434	0.69	W**D**GPTGQRD
z_9	$C_{43}H_{58}N_{13}O_{15}$	22	996.4175	996.4169	0.60	
y_{10}	$C_{48}H_{69}N_{16}O_{17}$	23	1141.5027	1141.5036	0.79	QW**D**GPTGQRD
y_{11}	$C_{51}H_{74}N_{17}O_{19}$	24	1228.5347	---------	---------	
y_{12}	$C_{53}H_{77}N_{18}O_{20}$	25	1285.5562	1285.5639	5.98	GSQW**D**GPTGQRD
z_{12}	$C_{53}H_{74}N_{17}O_{20}$	26	1268.5296	1268.5147	11.74	
y_{13}	---------	---	---------	---------	---------	
y_{14}	---------	---	---------	---------	---------	
b_1	---------	---	---------	---------	---------	
b_2	---------	---	---------	---------	---------	
b_3	$C_{22}H_{21}N_4O_4$	15	405.1563	405.1562	0.25	**G**WY
a_3	$C_{21}H_{21}N_4O_3$	14	377.1614	377.1610	1.06	
b_4	$C_{24}H_{24}N_5O_5$	16	462.1777	462.1778	0.22	**G**WYG
a_4	$C_{23}H_{24}N_5O_4$	15	434.1828	434.1842	3.22	
b_5	$C_{27}H_{29}N_6O_7$	17	549.2098	549.2099	0.18	**G**WYGS
a_5	$C_{26}H_{29}N_6O_6$	16	521.2149	521.2142	1.34	
b_6	$C_{32}H_{37}N_8O_9$	19	677.2683	677.2675	1.18	**G**WYGSQ
a_6	$C_{31}H_{37}N_8O_8$	18	649.2734	649.2746	1.84	
b_7	$C_{43}H_{47}N_{10}O_{10}$	26	863.3477	863.3494	1.97	**G**WYGSQW
b_8	$C_{47}H_{50}N_{11}O_{12}$	29	960.3640	960.3641	0.10	**G**WYGSQW**D**
a_8	$C_{46}H_{50}N_{11}O_{11}$	28	932.3691	932.3705	1.50	
b_9	$C_{49}H_{53}N_{12}O_{13}$	30	1017.3855	1017.3865	0.98	**G**WYGSQW**D**G
a_9	$C_{48}H_{53}N_{12}O_{12}$	29	989.3906	989.3904	0.20	
b_{10}	$C_{54}H_{60}N_{13}O_{14}$	32	1114.4383	1114.4405	1.97	**G**WYGSQW**D**GP
a_{10}	$C_{53}H_{60}N_{13}O_{13}$	31	1086.4434	1086.4480	4.23	
b_{11}	$C_{58}H_{67}N_{14}O_{16}$	33	1215.4859	1215.4860	0.08	**G**WYGSQW**D**GPT
a_{11}	$C_{57}H_{67}N_{14}O_{15}$	32	1187.4910	1187.4883	2.27	

b_{12}	$C_{60}H_{70}N_{15}O_{17}$	34	1272.5074	1272.5056	1.41	**G**WY**G**SQW**D**GPTG
a_{12}	$C_{59}H_{70}N_{15}O_{16}$	33	1244.5125	1244.5106	1.53	
b_{13}	$C_{65}H_{78}N_{17}O_{19}$	36	1400.5660	1400.5628	2.28	**G**WY**G**SQW**D**GPTGQ
a_{13}	$C_{64}H_{78}N_{17}O_{18}$	35	1372.5711	1372.5808	7.07	
b_{14}	$C_{71}H_{90}N_{21}O_{20}$	38	1556.6671	---	---	

Table S3. The assigned fragments of nocapeptin B (**2**) [861 Da], degree of unsaturation (RDB)

Fragment	Ion Formula	RDB	Calc. *m/z*	Meas. *m/z*	Error, ppm	Sequence
y_1	$C_{10}H_{20}N_5O_5$	2	134.0453	134.0450	2.24	D
y_2	$C_{10}H_{20}N_5O_5$	4	290.1464	290.1469	1.72	RD
z_2	$C_{10}H_{17}N_4O_5$	5	273.1199	273.1207	2.93	
y_3	$C_{15}H_{28}N_7O_7$	6	418.2050	418.2057	1.67	QRD
z_3	$C_{15}H_{25}N_6O_7$	7	401.1785	401.1795	2.49	
y_4	$C_{17}H_{31}N_8O_8$	7	475.2265	475.2273	1.68	GQRD
z_4	$C_{17}H_{28}N_7O_8$	8	458.1999	458.2005	1.31	
y_5	$C_{21}H_{38}N_9O_{10}$	8	576.2742	576.2748	1.04	TGQRD
z_5	$C_{21}H_{35}N_8O_{10}$	9	559.2476	559.2488	2.15	
y_6	$C_{26}H_{45}N_{10}O_{11}$	10	673.3269	673.3285	2.38	PTGQRD
z_6	$C_{26}H_{42}N_9O_{11}$	11	656.3004	656.3018	2.13	
y_7	$C_{28}H_{48}N_{11}O_{12}$	11	730.3484	730.3495	1.50	GPTGQRD
z_7	$C_{28}H_{45}N_{10}O_{12}$	12	713.3218	713.3232	1.96	
y_8	$C_{32}H_{51}N_{12}O_{14}$	14	827.3648	827.3652	0.48	**D**GPTGQRD
y_9	$C_{43}H_{61}N_{14}O_{15}$	21	1013.4441	1013.4437	0.39	W**D**GPTGQRD
z_9	$C_{43}H_{58}N_{13}O_{15}$	22	996.4175	996.4167	0.80	
y_{10}	$C_{48}H_{69}N_{16}O_{17}$	23	1141.5027	1141.4966	5.34	QW**D**GPTGQRD
y_{11}	$C_{51}H_{74}N_{17}O_{19}$	24	1228.5347	---	---	
y_{12}	$C_{53}H_{77}N_{18}O_{20}$	25	1285.5562	1285.5593	2.41	GSQW**D**GPTGQRD
y_{13}	---	---	---	---	---	
y_{14}	---	---	---	---	---	
b_1	---	---	---	---	---	
b_2	---	---	---	---	---	
b_3	$C_{22}H_{21}N_4O_6$	15	437.1461	437.1465	0.92	**G**WY
a_3	$C_{21}H_{21}N_4O_5$	14	409.1512	409.1508	0.98	
b_4	$C_{24}H_{24}N_5O_7$	16	494.1676	494.1684	1.62	**G**WY**G**
b_5	$C_{27}H_{29}N_6O_9$	17	581.1996	581.2012	2.75	**G**WY**G**S
a_5	$C_{26}H_{29}N_6O_8$	16	553.2047	553.2033	2.53	
b_6	$C_{32}H_{37}N_8O_{11}$	19	709.2582	709.2588	0.85	**G**WY**G**SQ
a_6	$C_{31}H_{37}N_8O_{10}$	18	681.2633	681.2683	7.34	
b_7	$C_{43}H_{47}N_{10}O_{12}$	26	895.3375	895.3382	0.78	**G**WY**G**SQW
a_7	$C_{42}H_{47}N_{10}O_{11}$	25	867.3426	867.3448	2.54	
b_8	$C_{47}H_{50}N_{11}O_{14}$	29	992.3539	992.3546	0.71	**G**WY**G**SQW**D**
a_8	$C_{46}H_{50}N_{11}O_{13}$	28	964.3590	964.3606	1.66	

b_9	$C_{49}H_{53}N_{12}O_{15}$	30	1049.3753	1049.3776	2.19	**G**WY**D**G
a_9	$C_{48}H_{53}N_{12}O_{14}$	29	1021.3804	1021.3828	2.35	
b_{10}	$C_{54}H_{60}N_{13}O_{16}$	32	1146.4281	1146.4279	0.17	**G**WY**D**GP
a_{10}	$C_{53}H_{60}N_{13}O_{15}$	31	1118.4332	1118.4271	5.45	
b_{11}	$C_{58}H_{67}N_{14}O_{18}$	33	1247.4758	1247.4776	1.44	**G**WY**D**GPT
a_{11}	$C_{57}H_{67}N_{14}O_{17}$	32	1219.4809	1219.4809	0.00	
b_{12}	$C_{60}H_{70}N_{15}O_{19}$	34	1304.4972	1304.4935	2.84	**G**WY**D**GPTG
a_{12}	$C_{59}H_{70}N_{15}O_{18}$	33	1276.5023	1276.4985	2.98	
b_{13}	$C_{65}H_{78}N_{17}O_{21}$	36	1432.5558	1432.5660	7.12	**G**WY**D**GPTGQ
b_{14}	$C_{71}H_{90}N_{21}O_{22}$	38	1588.6569	---------	---------	

Table S4. The assigned fragments of longipeptin A (**3**) [851 Da]; *Y*, *Z*, *Y-H_2O* indicate ions upon Me_2S (62 Da) loss, degree of unsaturation (RDB)

Fragment	Ion Formula	RDB	Calc. *m/z*	Meas. *m/z*	Error, ppm	Sequence
y_1	$C_4H_8NO_4$	2	134.0453	---------	---------	
y_2	$C_{10}H_{20}N_5O_5$	4	290.1464	290.1468	1.37	RD
y_3	$C_{14}H_{25}N_6O_6$	6	373.1836	373.1837	0.27	MRD
z_3	$C_{14}H_{22}N_5O_6$	7	356.1570	356.1570	0.00	
y_4	$C_{16}H_{28}N_7O_7$	7	430.2050	430.2052	0.46	GMRD
z_4	$C_{16}H_{25}N_6O_7$	8	413.1785	413.1783	0.48	
y_5	$C_{19}H_{33}N_8O_9$	8	517.2370	517.2372	0.39	SGMRD
y_5-H_2O	$C_{19}H_{31}N_8O_8$	9	499.2265	499.2265	0.00	
y_6	$C_{24}H_{40}N_9O_{10}$	10	614.2898	614.2902	0.65	PSGMRD
y_6-H_2O	$C_{24}H_{38}N_9O_9$	11	596.2792	596.2786	1.00	
y_7	$C_{26}H_{43}N_{10}O_{11}$	11	671.3113	671.3114	0.15	GPSGMRD
y_7-H_2O	$C_{26}H_{41}N_{10}O_{10}$	12	653.3007	653.3015	1.22	
y_8	$C_{30}H_{46}N_{11}O_{13}$	14	768.3277	768.3240	4.81	**D**GPSGMRD
y_9	$C_{41}H_{56}N_{13}O_{14}$	21	954.4070	954.4048	2.30	W**D**GPSGMRD
y_{10}	$C_{45}H_{62}N_{15}O_{16}$	23	1068.4499	1068.4501	0.19	NW**D**GPSGMRD
y_{11}	$C_{48}H_{67}N_{16}O_{17}$	24	1139.4870	1139.4801	6.05	ANW**D**GPSGMRD
y_{12}	$C_{50}H_{70}N_{17}O_{18}$	25	1196.5085	1196.5090	0.42	GANW**D**GPSGMRD
y_{13}	---------	---	---------	---------	---------	
y_{14}	---------	---	---------	---------	---------	
b_1	---------	---	---------	---------	---------	
b_2	---------	---	---------	---------	---------	
b_3	$C_{24}H_{22}N_5O_4$	17	444.1672	444.1681	2.03	**G**WW
a_3	$C_{23}H_{22}N_5O_3$	16	416.1723	416.1728	1.20	
b_4	$C_{26}H_{25}N_6O_5$	18	501.1886	501.1892	1.20	**G**WWG
a_4	$C_{25}H_{25}N_6O_4$	17	473.1937	473.1945	1.70	
b_5	$C_{29}H_{30}N_7O_6$	19	572.2258	572.2252	1.05	**G**WWGA
a_5	$C_{28}H_{30}N_7O_5$	18	544.2308	544.2306	0.37	
b_6	$C_{33}H_{36}N_9O_8$	21	686.2687	686.2701	2.04	**G**WWGAN
b_6-NH_3	$C_{33}H_{33}N_8O_8$	22	669.2421	669.2424	0.45	

a_6	$C_{32}H_{36}N_9O_7$	20	658.2738	658.2715	3.50	
a_6-NH_3	$C_{32}H_{33}N_8O_7$	21	641.2472	641.2470	0.31	
b_7	$C_{44}H_{46}N_{11}O_9$	28	872.3480	872.3483	0.34	**G**WW**G**ANW
b_8	$C_{48}H_{49}N_{12}O_{11}$	31	969.3644	969.3653	0.93	**G**WW**G**ANW**D**
b_8-NH_3	$C_{48}H_{46}N_{11}O_{11}$	32	952.3378	952.3383	0.53	
a_8	$C_{47}H_{49}N_{12}O_{10}$	30	941.3695	941.3700	0.53	
a_8-NH_3	$C_{47}H_{46}N_{11}O_{10}$	31	924.3429	924.3431	0.22	
b_9	$C_{50}H_{52}N_{13}O_{12}$	32	1026.3858	1026.3852	0.58	**G**WW**G**ANW**D**G
b_9-NH_3	$C_{50}H_{49}N_{12}O_{12}$	33	1009.3593	1009.3595	0.20	
a_9	$C_{49}H_{52}N_{13}O_{11}$	31	998.3909	998.3922	1.30	
a_9-NH_3	$C_{49}H_{49}N_{12}O_{11}$	32	981.3644	981.3648	0.41	
b_{10}	$C_{55}H_{59}N_{14}O_{13}$	34	1123.4386	1123.4378	0.71	**G**WW**G**ANW**D**GP
b_{11}	$C_{58}H_{64}N_{15}O_{15}$	35	1210.4706	1210.4701	0.41	**G**WW**G**ANW**D**GPS
b_{11}-NH_3	$C_{58}H_{61}N_{14}O_{15}$	36	1193.4441	1193.4462	1.76	
b_{11}-NH_3-H_2O	$C_{58}H_{59}N_{14}O_{14}$	37	1175.4335	1175.4315	1.70	
a_{11}	$C_{57}H_{64}N_{15}O_{14}$	34	1182.4757	1182.4743	1.18	
a_{11}-NH_3	$C_{57}H_{61}N_{14}O_{14}$	35	1165.4492	1165.4471	1.80	
b_{12}	$C_{60}H_{67}N_{16}O_{16}$	36	1267.4921	1267.5161	18.94	**G**WW**G**ANW**D**GPSG
a_{12}	$C_{59}H_{67}N_{16}O_{15}$	35	1239.4972	1239.5103	10.56	
$[M+H-Me_2S]^{2+}$	$C_{74}H_{92}N_{22}O_{22}$	40	820.3379	820.3380	0.12	**G**WW**G**ANW**D**GPSGMRD
$[M+H-Me_2S-NH_3]^{2+}$	$C_{74}H_{89}N_{21}O_{22}$	41	811.8246	811.8268	2.70	**G**WW**G**ANW**D**GPSGMRD
$[M+H-Me_2S-H_2O]^{2+}$	$C_{74}H_{90}N_{22}O_{21}$	41	811.3326	811.3340	1.70	**G**WW**G**ANW**D**GPSGMRD
$[M+H-Me_2S-CO]^{2+}$	$C_{73}H_{92}N_{22}O_{21}$	39	806.3404	806.3415	1.36	**G**WW**G**ANW**D**GPSGMRD

Table S5. The assigned fragments of longipeptin B (**4**) [843 Da]; *Y*, *Z*, *Y*-H_2O indicate ions upon Me_2S (62 Da) loss, degree of unsaturation (RDB)

Fragment	Ion Formula	RDB	Calc. *m/z*	Meas. *m/z*	Error, ppm	Sequence
y_1	$C_4H_8NO_4$	2	134.0453	---------	---------	
y_2	$C_{10}H_{20}N_5O_5$	4	290.1464	290.1463	0.34	RD
y_3	$C_{14}H_{25}N_6O_6$	6	373.1836	373.1830	1.61	MRD
z_3	$C_{14}H_{22}N_5O_6$	7	356.1570	356.1572	0.56	
y_4	$C_{16}H_{28}N_7O_7$	7	430.2050	430.2052	0.46	GMRD
z_4	$C_{16}H_{25}N_6O_7$	8	413.1785	413.1782	0.73	
y_5	$C_{19}H_{33}N_8O_9$	8	517.2370	517.2374	0.77	SGMRD
y_5-H_2O	$C_{19}H_{31}N_8O_8$	9	499.2265	499.2267	0.40	
y_6	$C_{24}H_{40}N_9O_{10}$	10	614.2898	614.2905	1.14	PSGMRD
y_6-H_2O	$C_{24}H_{38}N_9O_9$	11	596.2792	596.2788	0.67	
y_7	$C_{26}H_{43}N_{10}O_{11}$	11	671.3113	671.3117	0.60	GPSGMRD
y_7-H_2O	$C_{26}H_{41}N_{10}O_{10}$	12	653.3007	653.3009	0.31	
y_8	$C_{30}H_{46}N_{11}O_{13}$	14	768.3277	768.3253	3.12	**D**GPSGMRD
y_9	$C_{41}H_{56}N_{13}O_{14}$	21	954.4070	---------	---------	W**D**GPSGMRD
y_{10}	$C_{45}H_{62}N_{15}O_{16}$	23	1068.4499	1068.4517	1.68	NW**D**GPSGMRD
y_{11}	$C_{48}H_{67}N_{16}O_{17}$	24	1139.4870	1139.4790	7.02	ANW**D**GPSGMRD

y_{12}	$C_{50}H_{70}N_{17}O_{18}$	25	1196.5085	---------	---------	
y_{13}	---------	---	---------	---------	---------	
y_{14}	---------	---	---------	---------	---------	
b_1	---------	---	---------	---------	---------	
b_2	---------	---	---------	---------	---------	
b_3	$C_{24}H_{22}N_5O_3$	17	428.1723	428.1730	1.63	**G**WW
a_3	$C_{23}H_{22}N_5O_2$	16	400.1773	400.1775	0.50	
b_4	$C_{26}H_{25}N_6O_4$	18	485.1937	485.1936	0.21	**G**WWG
a_4	$C_{25}H_{25}N_6O_3$	17	457.1988	457.1994	1.31	
b_5	$C_{29}H_{30}N_7O_5$	19	556.2308	556.2309	0.18	**G**WWGA
a_5	$C_{28}H_{30}N_7O_4$	18	528.2359	528.2360	0.19	
b_6	$C_{33}H_{36}N_9O_7$	21	670.2738	670.2753	2.24	**G**WWGAN
b_7	$C_{44}H_{46}N_{11}O_8$	28	856.3531	856.3499	3.73	**G**WWGANW
b_8	$C_{48}H_{49}N_{12}O_{10}$	31	953.3695	953.3684	1.15	**G**WWGANW**D**
b_8-NH_3	$C_{48}H_{46}N_{11}O_{10}$	32	936.3429	936.3437	0.85	
a_8	$C_{47}H_{49}N_{12}O_9$	30	925.3745	925.3723	2.38	
a_8-NH_3	$C_{47}H_{46}N_{11}O_9$	31	908.3480	908.3486	0.67	
b_9	$C_{50}H_{52}N_{13}O_{11}$	32	1010.3909	1010.3880	2.87	**G**WWGANW**D**G
b_9-NH_3	$C_{50}H_{49}N_{12}O_{11}$	33	993.3644	993.3659	1.51	
a_9	$C_{49}H_{52}N_{13}O_{10}$	31	982.3960	982.4099	14.15	
a_9-NH_3	$C_{49}H_{49}N_{12}O_{10}$	32	965.3695	965.3748	5.49	
b_{10}	$C_{55}H_{59}N_{14}O_{12}$	34	1107.4437	1107.4352	7.68	**G**WWGANW**D**GP
b_{11}	$C_{58}H_{64}N_{15}O_{14}$	35	1194.4757	1194.4710	3.93	**G**WWGANW**D**GPS
b_{11}-NH_3	$C_{58}H_{61}N_{14}O_{14}$	36	1177.4492	1177.4526	2.89	
b_{11}-NH_3-H_2O	$C_{58}H_{59}N_{14}O_{13}$	37	1159.4386	1159.4420	2.93	
a_{11}	$C_{57}H_{64}N_{15}O_{13}$	34	1166.4808	1166.4720	7.54	
a_{11}-NH_3	$C_{57}H_{61}N_{14}O_{13}$	35	1149.4543	1149.4795	21.92	
b_{12}	$C_{60}H_{67}N_{16}O_{15}$	36	1251.4972	1251.4874	7.83	**G**WWGANW**D**GPSG
$[M+H-Me_2S]^{2+}$	$C_{74}H_{92}N_{22}O_{21}$	40	812.3404	812.3421	2.09	**G**WWGANW**D**GPSGMRD
$[M+H-Me_2S-NH_3]^{2+}$	$C_{74}H_{89}N_{21}O_{21}$	41	803.8271	803.8285	1.74	**G**WWGANW**D**GPSGMRD
$[M+H-Me_2S-H_2O]^{2+}$	$C_{74}H_{90}N_{22}O_{20}$	41	803.3351	803.3379	3.48	**G**WWGANW**D**GPSGMRD
$[M+H-Me_2S-CO]^{2+}$	$C_{73}H_{92}N_{22}O_{20}$	39	798.3429	798.3429	0.00	**G**WWGANW**D**GPSGMRD

Table S6. The assigned fragments of longipeptin C (**5**) [844 Da]; Y_n· indicates *Y* ions upon MeSOH (64 Da) loss, degree of unsaturation (RDB)

Fragment	Ion Formula	RDB	Calc. *m/z*	Meas. *m/z*	Error, ppm	Sequence
y_1	$C_4H_8NO_4$	2	134.0453	---------	---------	
y_2	$C_{10}H_{20}N_5O_5$	4	290.1464	290.1463	0.34	RD
z_2	$C_{10}H_{17}N_4O_5$	5	273.1199	273.1191	2.93	
y_3	$C_{15}H_{29}N_6O_7S$	5	437.1818	437.1821	0.69	MRD
y_3·	$C_{14}H_{25}N_6O_6$	6	373.1836	373.1831	1.34	
y_4	$C_{17}H_{32}N_7O_8S$	6	494.2033	494.2035	0.40	GMRD
y_4·	$C_{16}H_{28}N_7O_7$	7	430.2050	430.2056	1.39	
y_5	$C_{20}H_{37}N_8O_{10}S$	7	581.2353	581.2343	1.72	SGMRD
y_5·	$C_{19}H_{33}N_8O_9$	8	517.2370	517.2372	0.39	
y_5-H_2O	$C_{20}H_{35}N_8O_9S$	8	563.2248	563.2241	1.24	
y_6	$C_{25}H_{44}N_9O_{11}S$	9	678.2881	678.2879	0.30	PSGMRD
y_6·	$C_{24}H_{40}N_9O_{10}$	10	614.2898	614.2910	1.95	
y_6-H_2O	$C_{25}H_{42}N_9O_{10}S$	10	660.2775	660.2776	0.15	
y_7	$C_{27}H_{47}N_{10}O_{12}S$	10	735.3096	735.3099	0.41	GPSGMRD
y_7·	$C_{26}H_{43}N_{10}O_{11}$	11	671.3113	671.3112	0.15	
y_7-H_2O	$C_{27}H_{45}N_{10}O_{11}S$	11	717.2990	717.3002	1.67	
y_8	$C_{31}H_{50}N_{11}O_{14}S$	13	832.3259	---------	---------	**D**GPSGMRD
y_9	$C_{42}H_{60}N_{13}O_{15}S$	20	1018.4053	1018.4202	14.63	W**D**GPSGMRD
y_{10}	$C_{46}H_{66}N_{15}O_{17}S$	22	1132.4482	1132.4430	4.60	NW**D**GPSGMRD
y_{11}	$C_{49}H_{71}N_{16}O_{18}S$	23	1203.4853	---------	---------	
y_{12}	$C_{51}H_{74}N_{17}O_{19}S$	24	1260.5068	---------	---------	
y_{13}	---------	---	---------	---------	---------	
y_{14}	---------	---	---------	---------	---------	
b_1	---------	---	---------	---------	---------	
b_2	---------	---	---------	---------	---------	
b_3	$C_{24}H_{22}N_5O_3$	17	428.1723	428.1737	3.26	**G**WW
a_3	$C_{23}H_{22}N_5O_2$	16	400.1773	400.1754	4.74	
b_4	$C_{26}H_{25}N_6O_4$	18	485.1937	485.1947	2.06	**G**WWG
a_4	$C_{25}H_{25}N_6O_3$	17	457.1988	---------	---------	
b_5	$C_{29}H_{30}N_7O_5$	19	556.2308	556.2293	2.70	**G**WWGA
a_5	$C_{28}H_{30}N_7O_4$	18	528.2359	---------	---------	
b_6	$C_{33}H_{36}N_9O_7$	21	670.2738	670.2753	2.23	**G**WWGAN
a_6	$C_{32}H_{36}N_9O_6$	20	642.2789	---------	---------	
b_7	$C_{44}H_{46}N_{11}O_8$	28	856.3531	856.3567	4.20	**G**WWGANW
b_8	$C_{48}H_{49}N_{12}O_{10}$	31	953.3695	953.3727	3.35	**G**WWGANW**D**
b_8-NH_3	$C_{48}H_{46}N_{11}O_{10}$	32	936.3429	936.3417	1.28	
a_8	$C_{47}H_{49}N_{12}O_9$	30	925.3745	925.3743	0.21	
a_8-NH_3	$C_{47}H_{46}N_{11}O_9$	31	908.3480	908.3497	1.87	
b_9	$C_{50}H_{52}N_{13}O_{11}$	32	1010.3909	1010.3972	6.23	**G**WWGANW**D**G
b_9-NH_3	$C_{50}H_{49}N_{12}O_{11}$	33	993.3644	993.3627	1.71	
a_9	$C_{49}H_{52}N_{13}O_{10}$	31	982.3960	982.3906	5.49	
a_9-NH_3	$C_{49}H_{49}N_{12}O_{10}$	32	965.3695	965.3679	1.66	

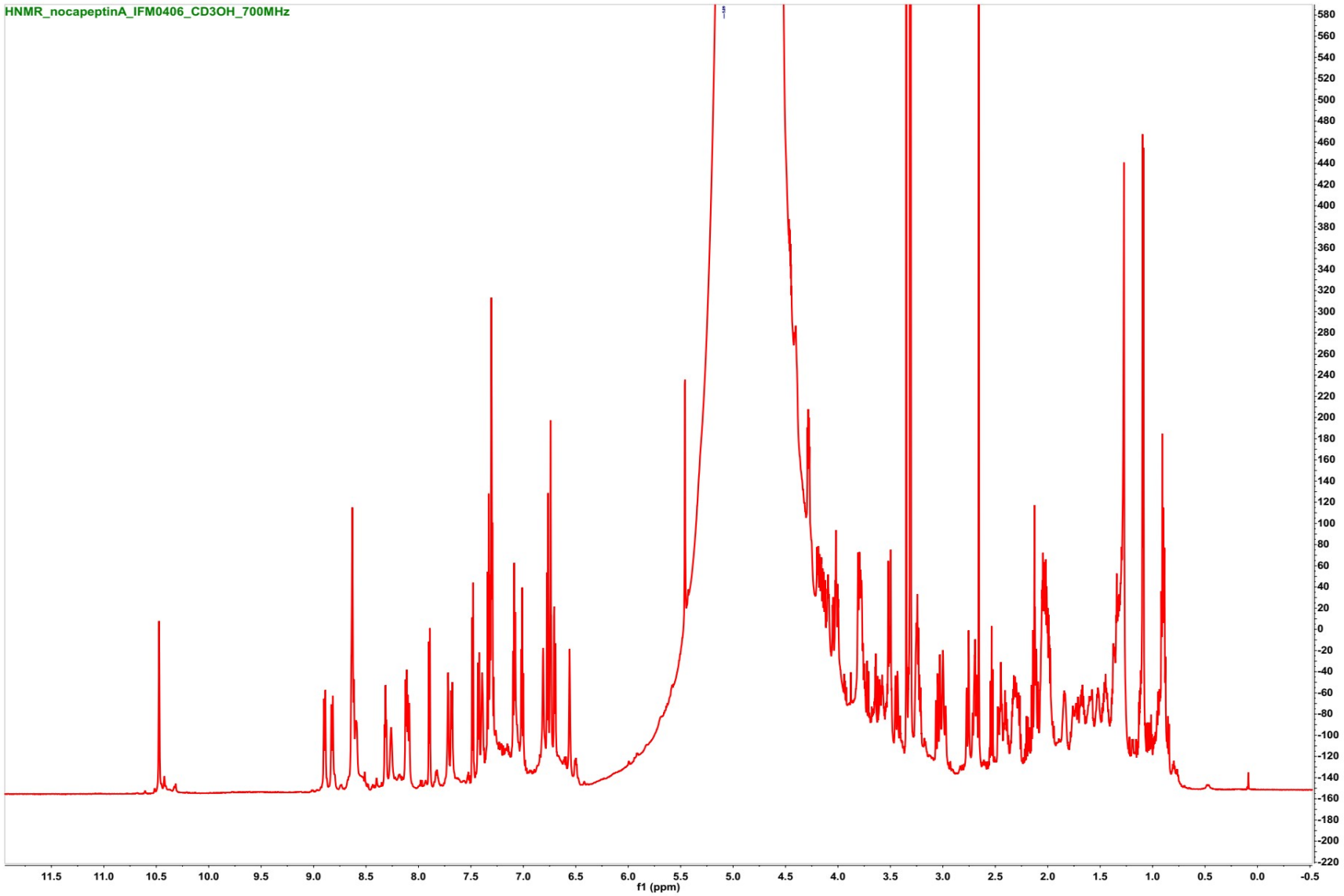


Figure S14. 700 MHz ^{1}H-NMR spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

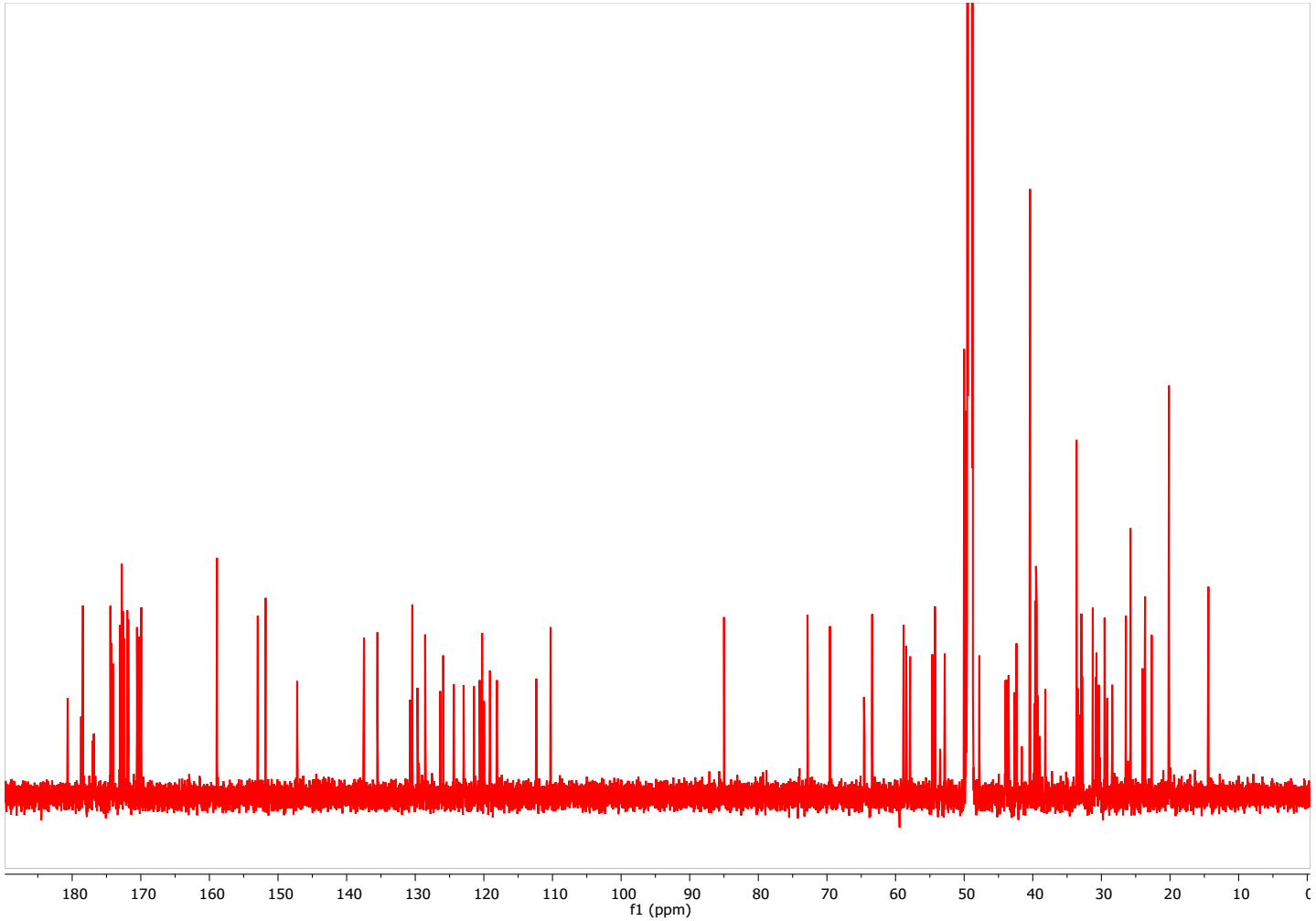


Figure S15. 176 MHz ^{13}C-NMR spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

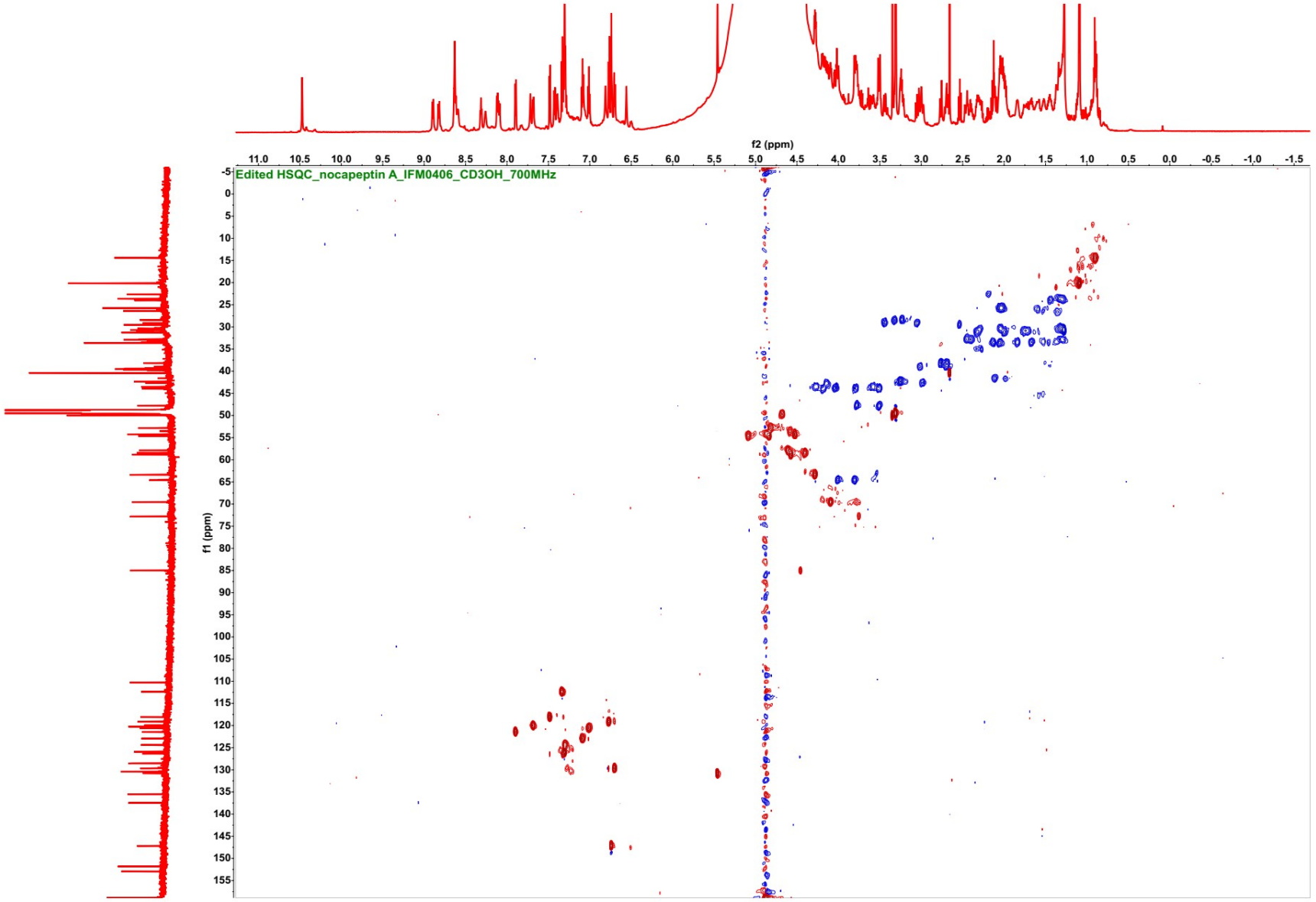


Figure S16. 700 MHz ^{1}H-^{13}C edited HSQC spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

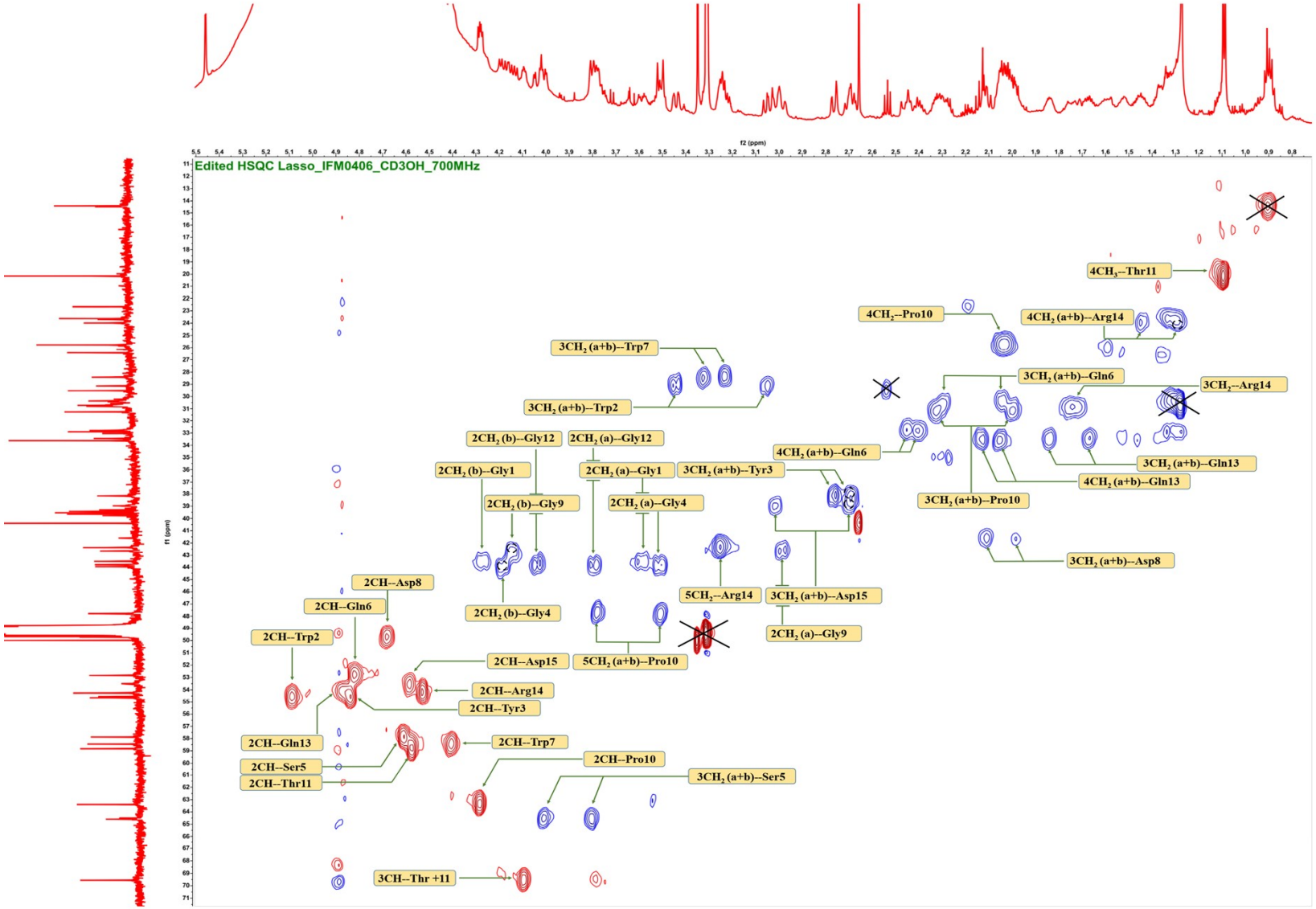


Figure S16A. Annotated ^{1}H-^{13}C edited HSQC spectrum of nocapeptin A (**1**)

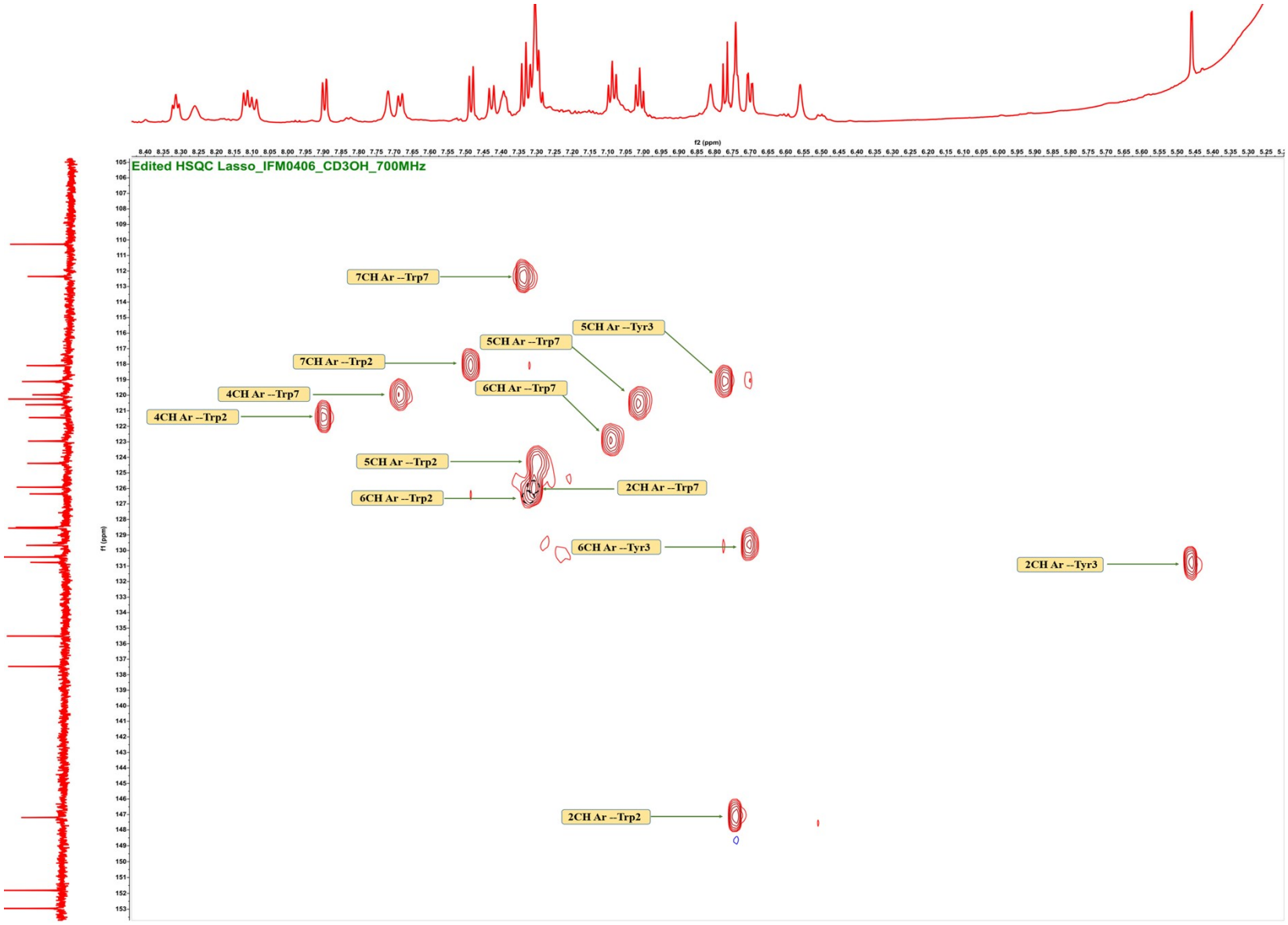


Figure S16B. Annotated ^{1}H-^{13}C edited HSQC spectrum of nocapeptin A (**1**)

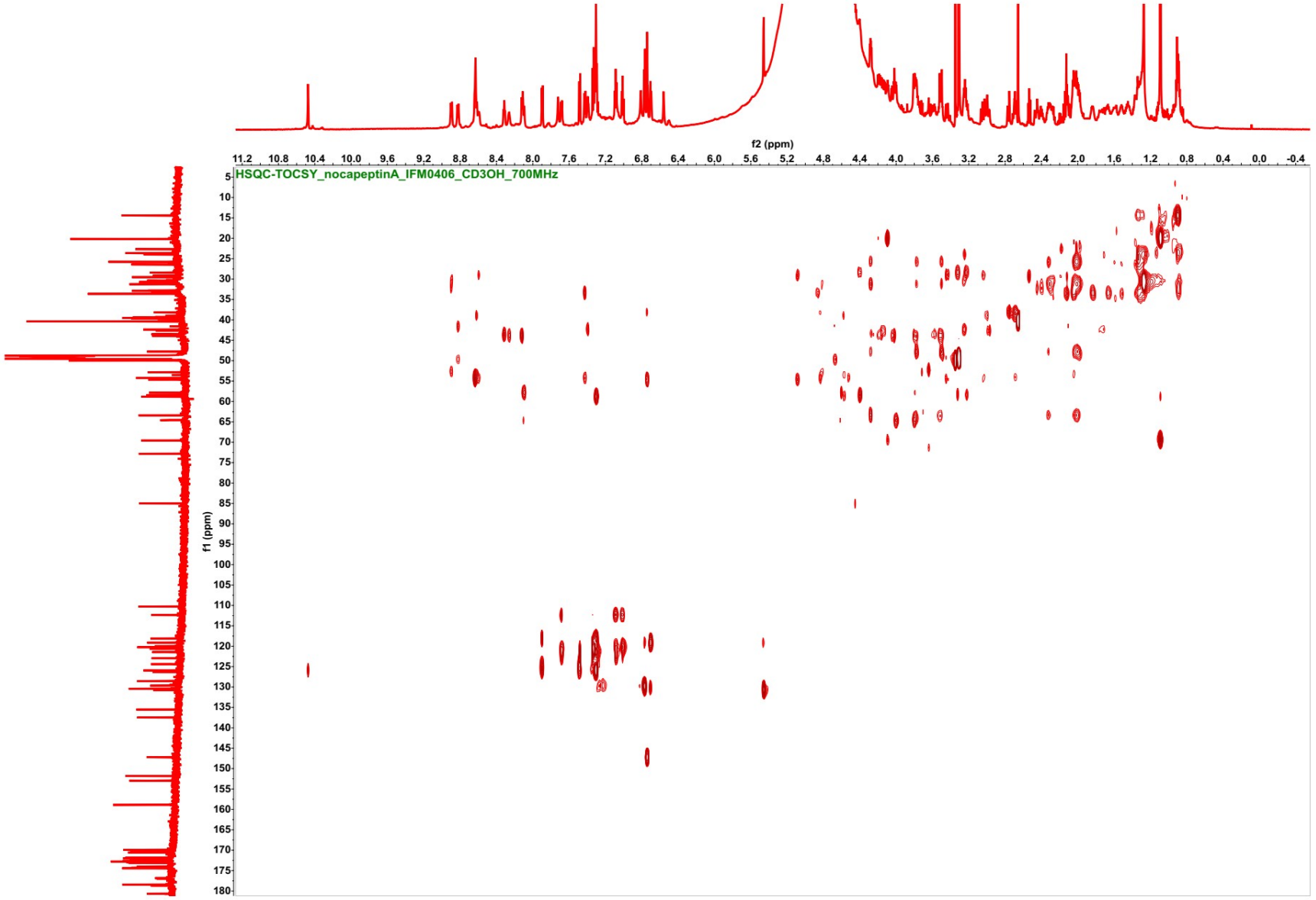


Figure S17. 700 MHz ^{1}H-^{13}C HSQC-TOCSY spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

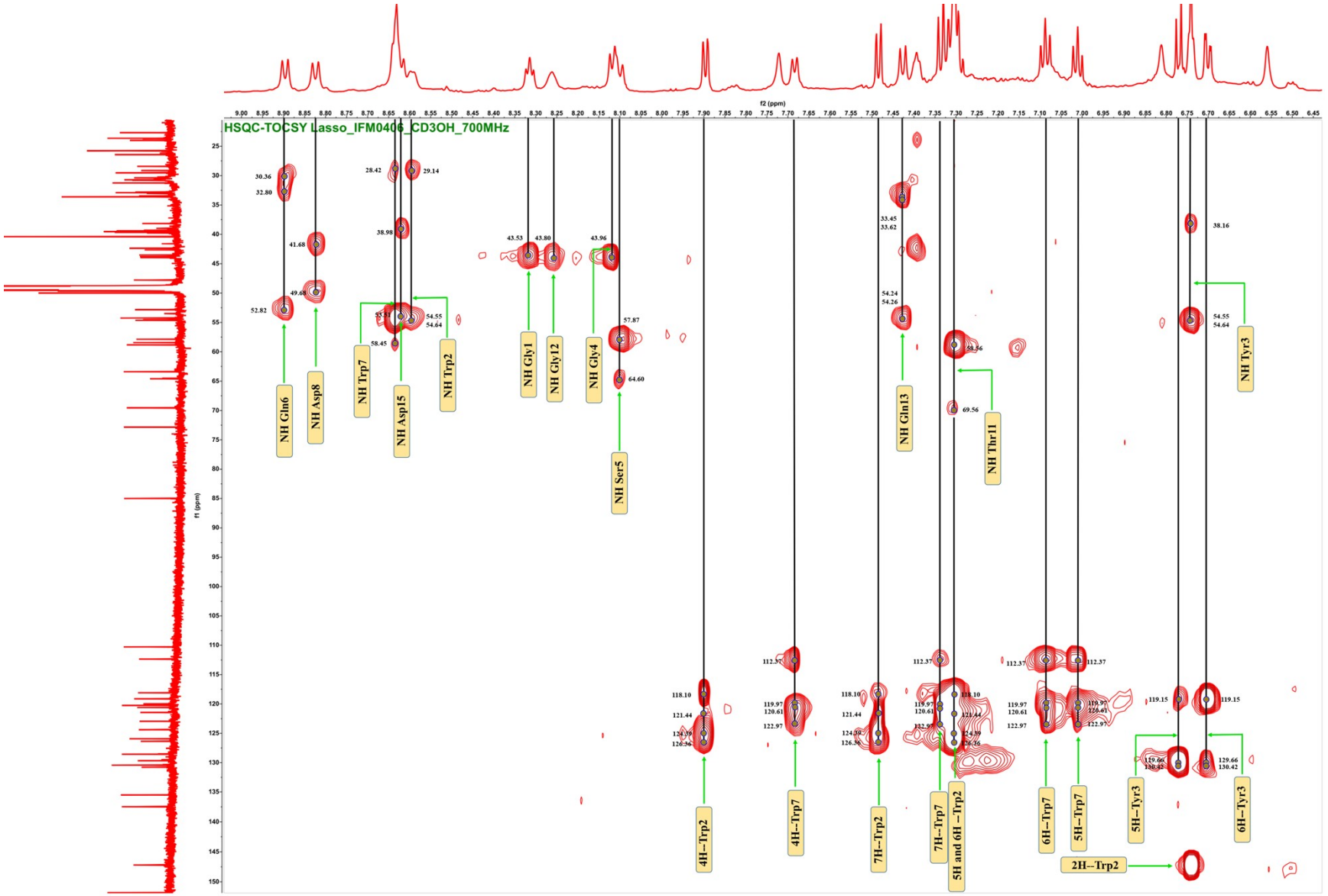


Figure S17A. Annotated ^{1}H-^{13}C HSQC-TOCSY spectrum of nocapeptin A (**1**)

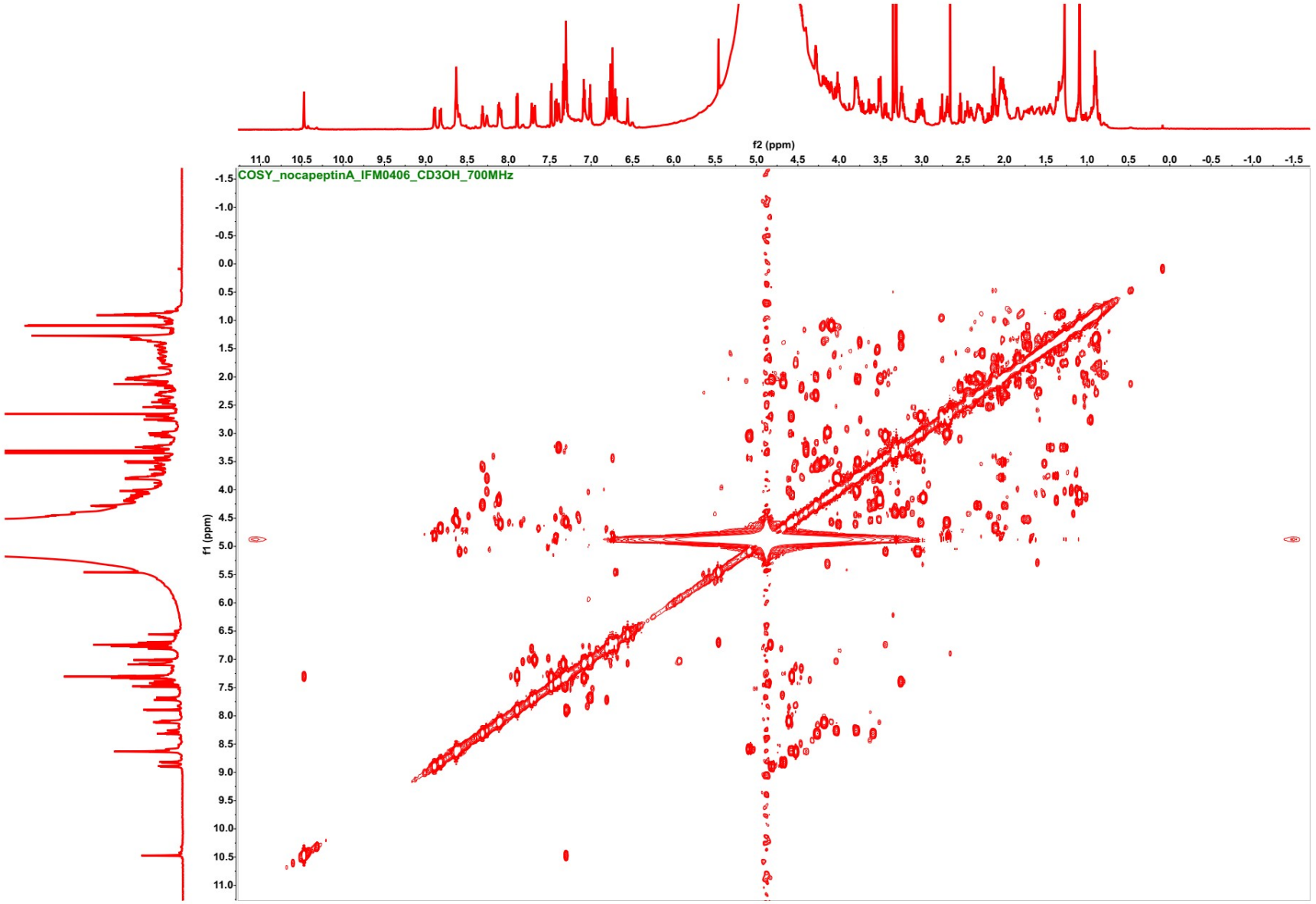


Figure S18. 700 MHz ^{1}H-^{1}H COSY spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

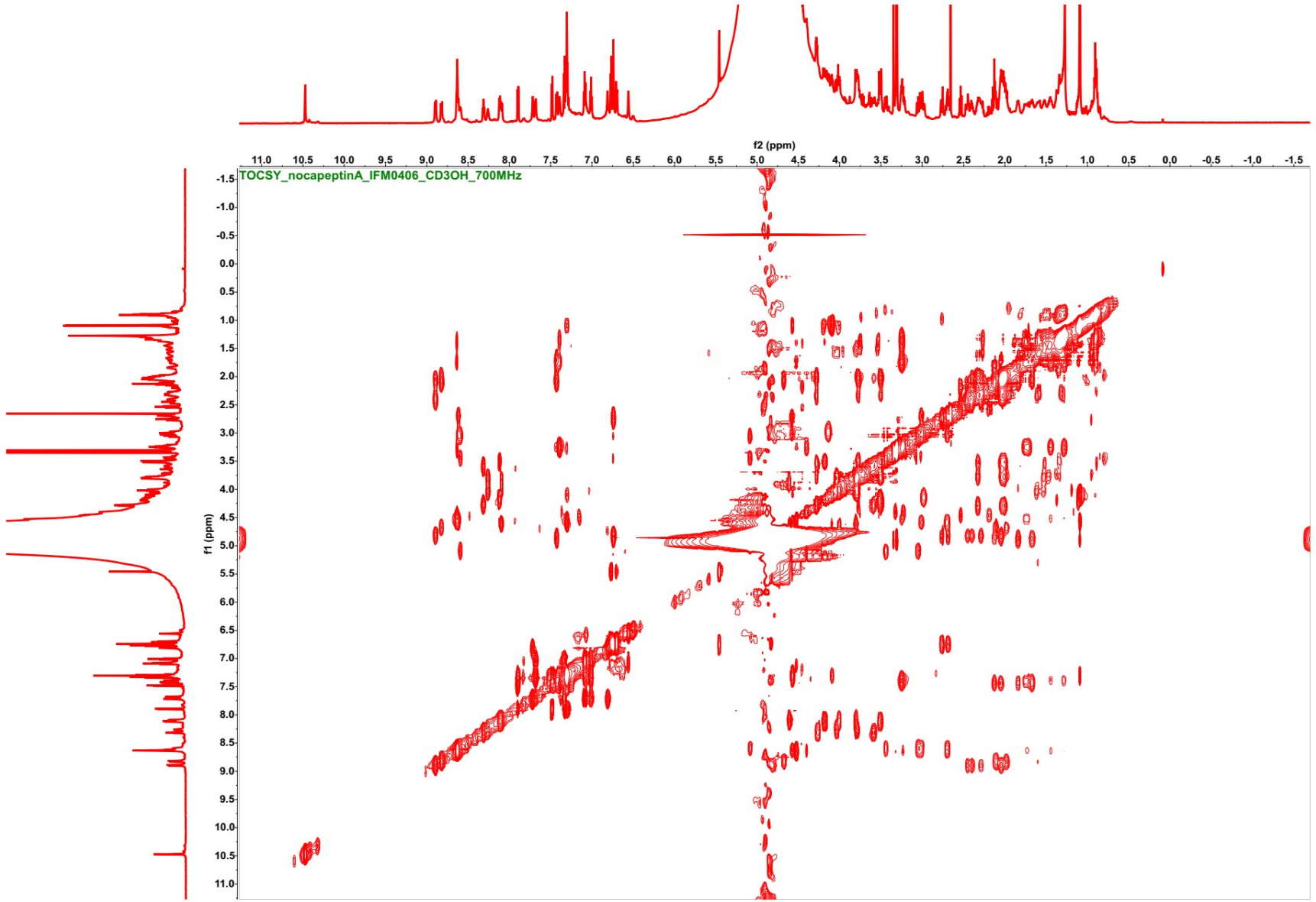


Figure S19. 700 MHz ^{1}H-^{1}H TOCSY spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

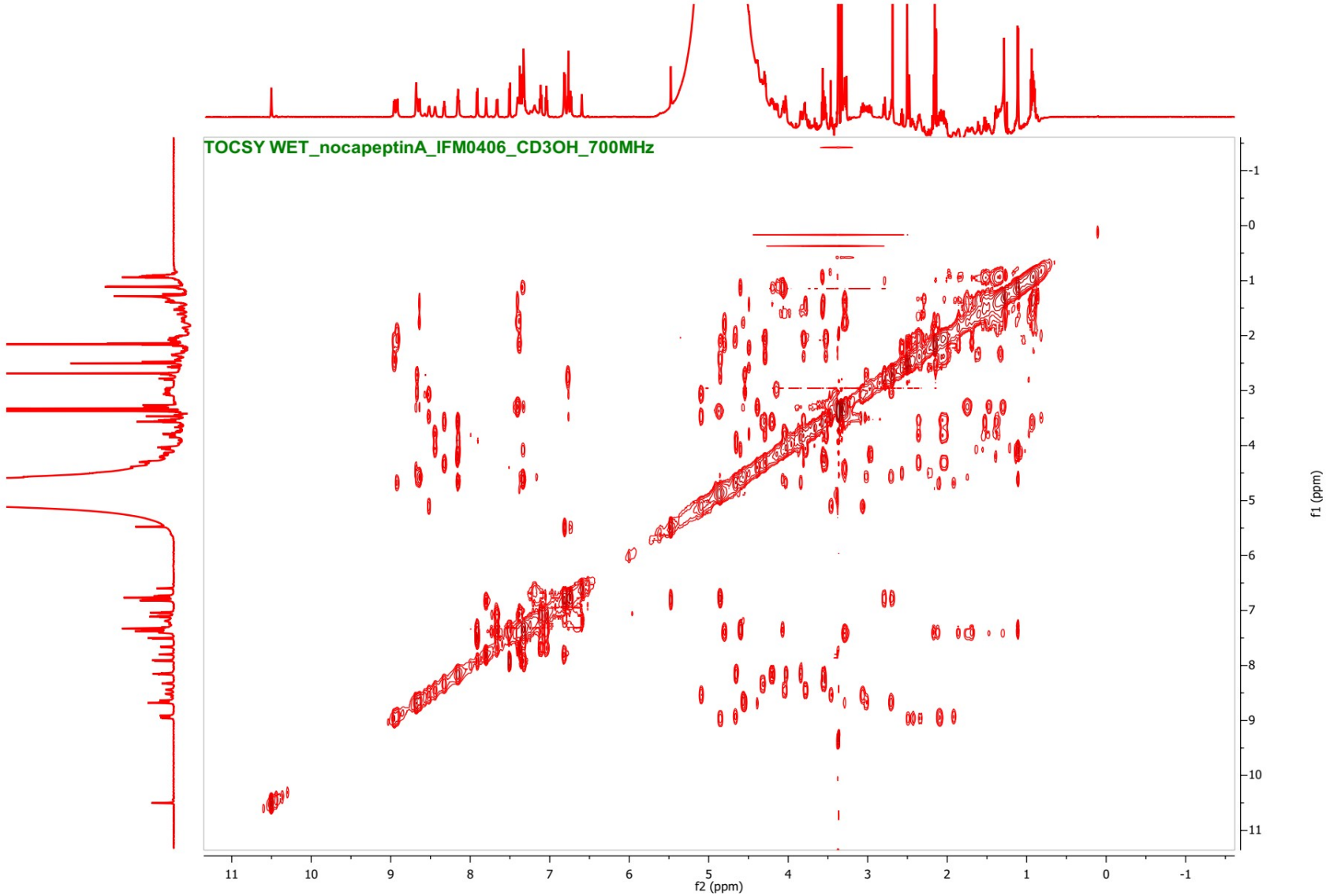


Figure S19A. 700 MHz ^{1}H-^{1}H TOCSY spectrum of nocapeptin A (**1**) with WET solvent suppression in d_3-CH_3OH/H_2O (96:4)

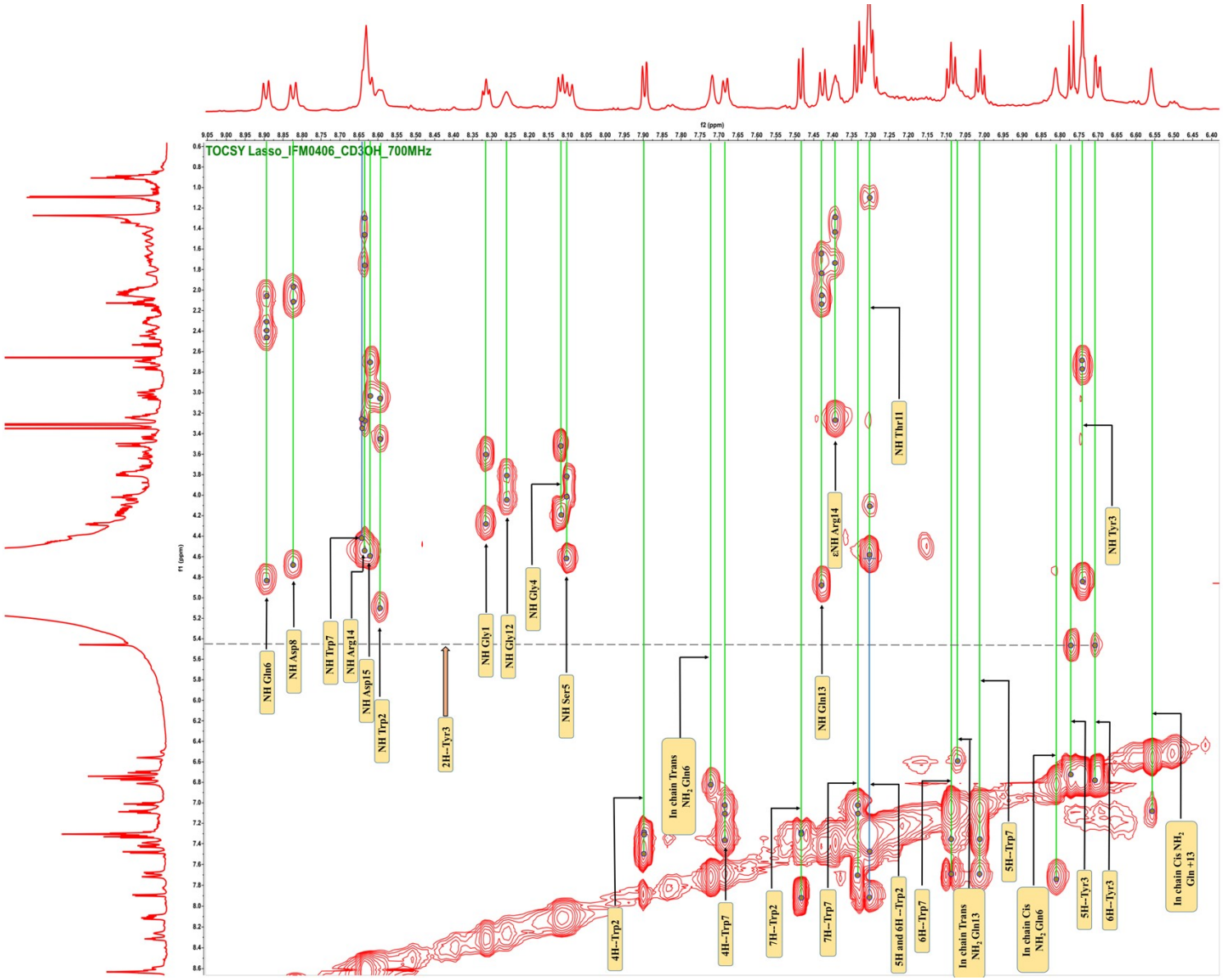


Figure S19B. Annotated ^{1}H-^{1}H TOCSY spectrum of nocapeptin A (**1**)

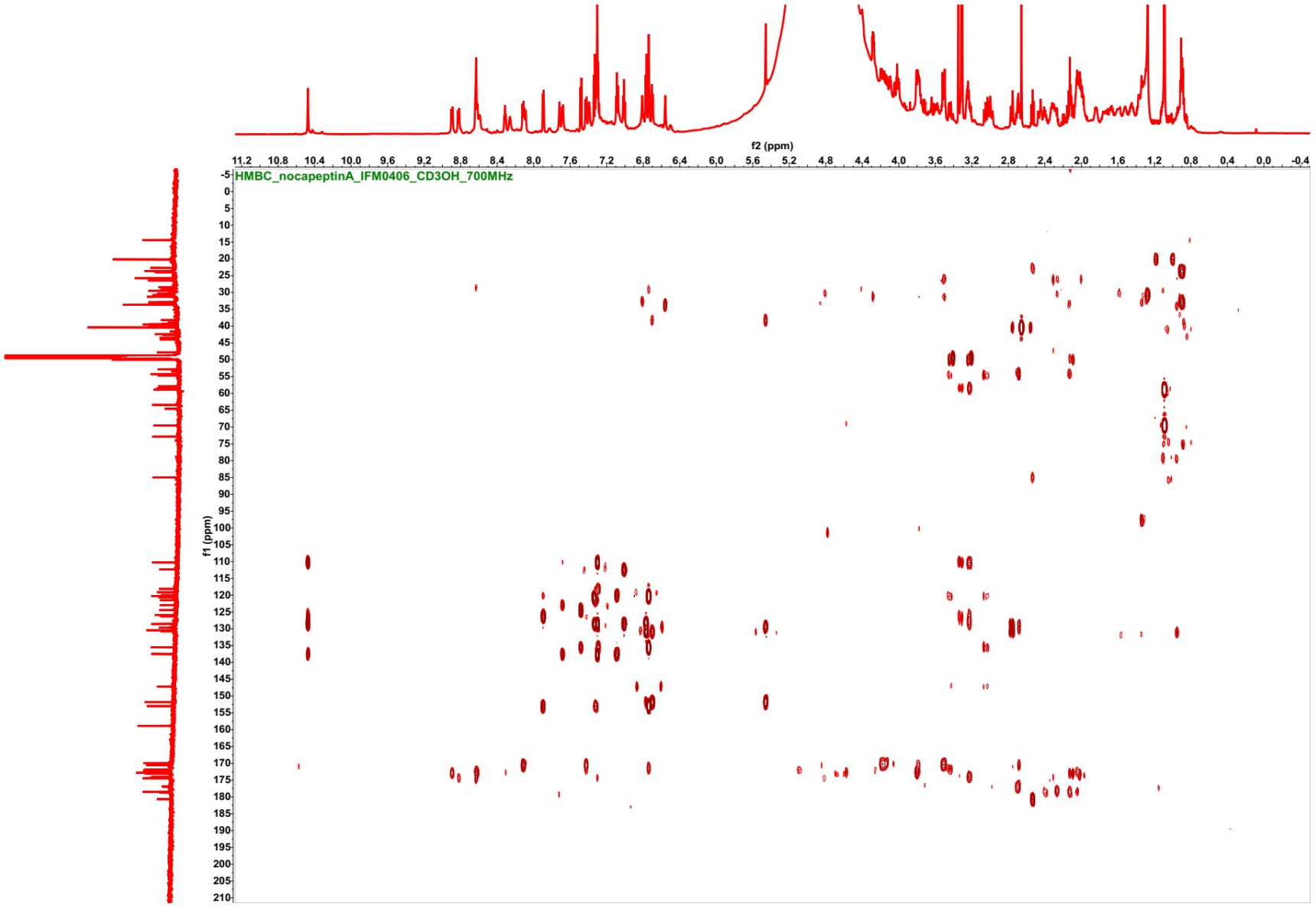


Figure S20A. 700 MHz ^{1}H-^{13}C HMBC spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

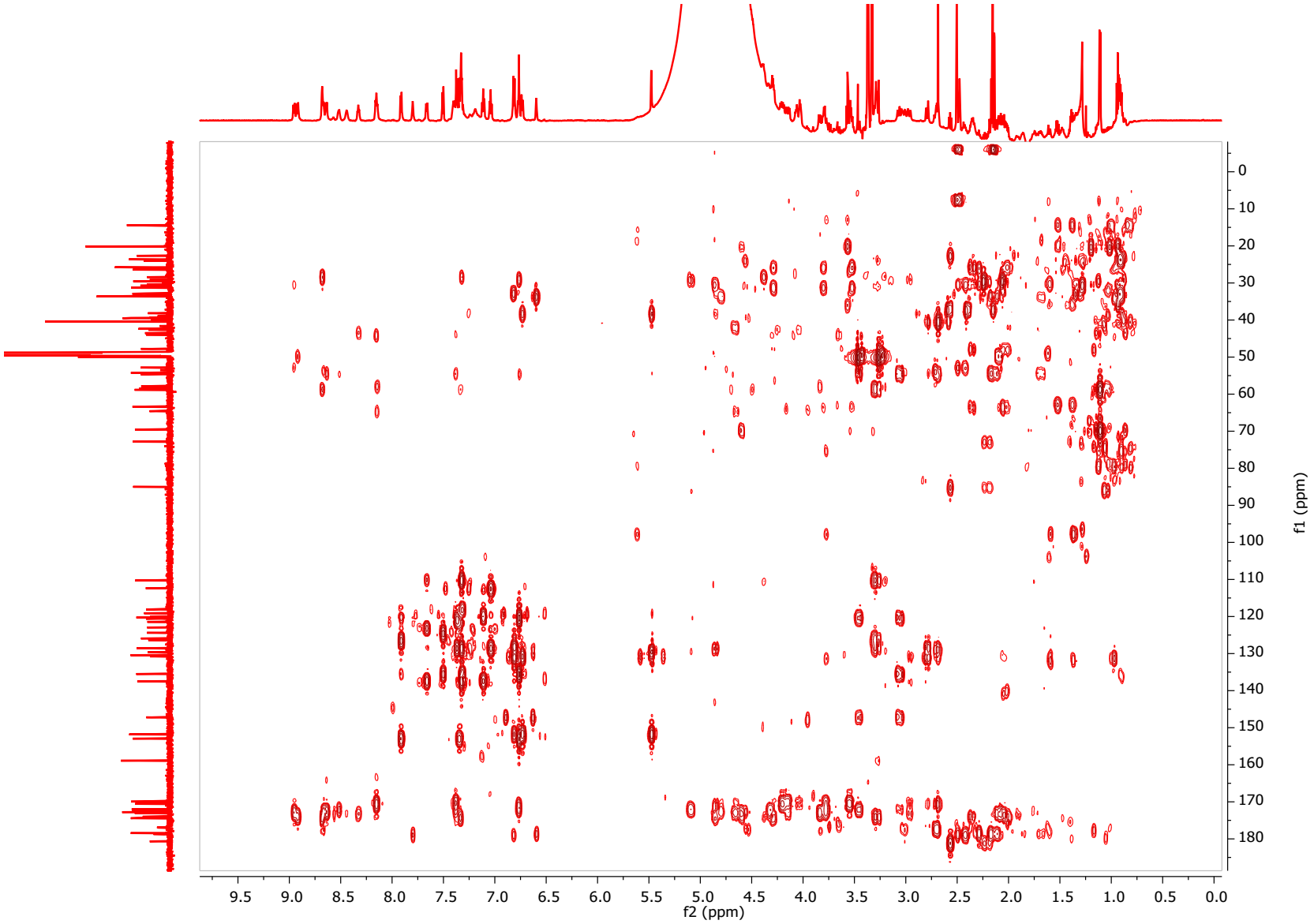


Figure S20B. 700 MHz ^{1}H-^{13}C HMBC NMR spectrum of a higher concentrated sample of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4) which allowed the observation of further long-range correlations in the arginine side chain and the overall assignment of the carbonyl carbons.

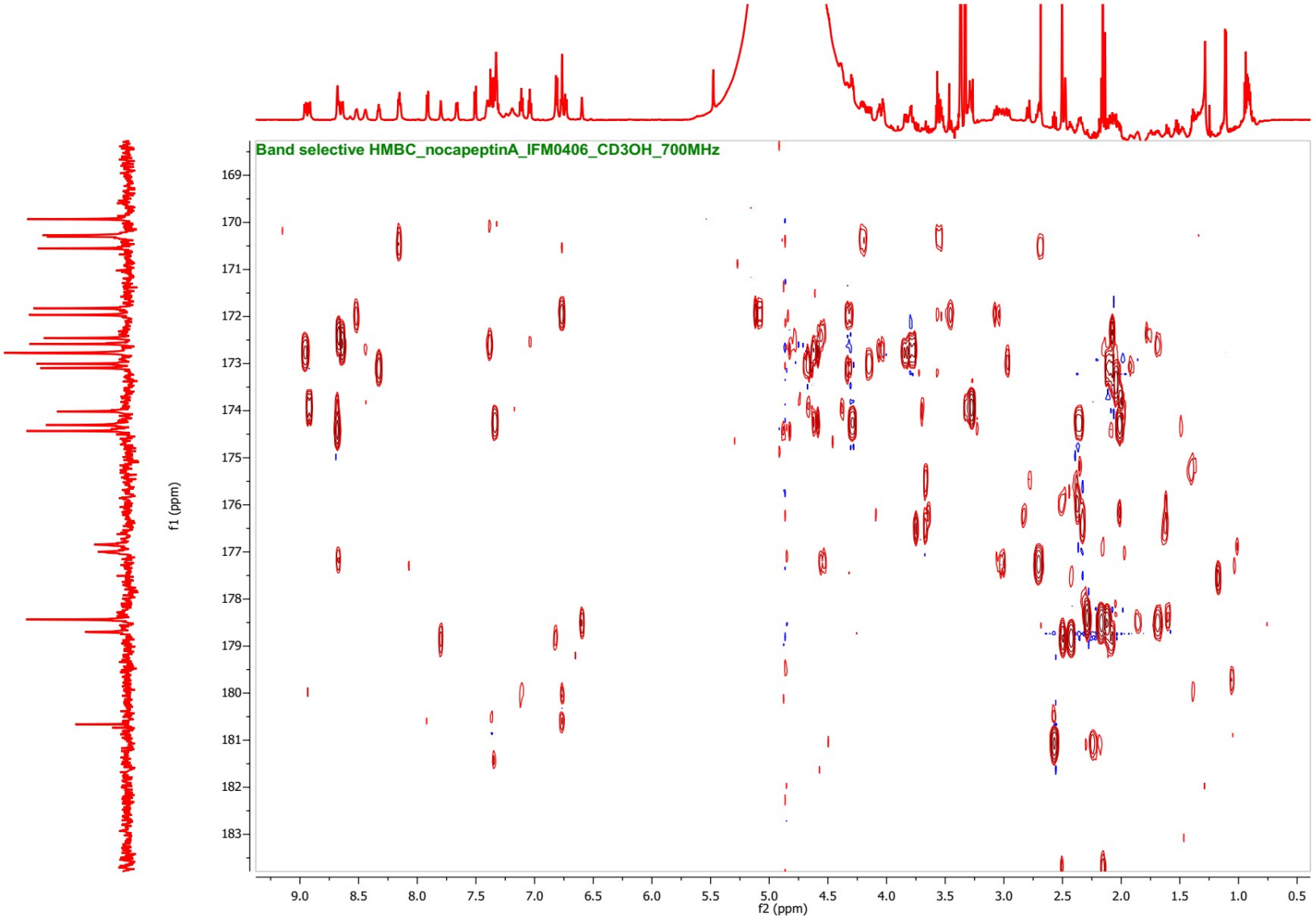


Figure S20C. 700 MHz ^{1}H-^{13}C Band-selective HMBC spectrum of nocapeptin A (**1**) focused on δ_C 161-191 in d_3-CH_3OH/H_2O (96:4)

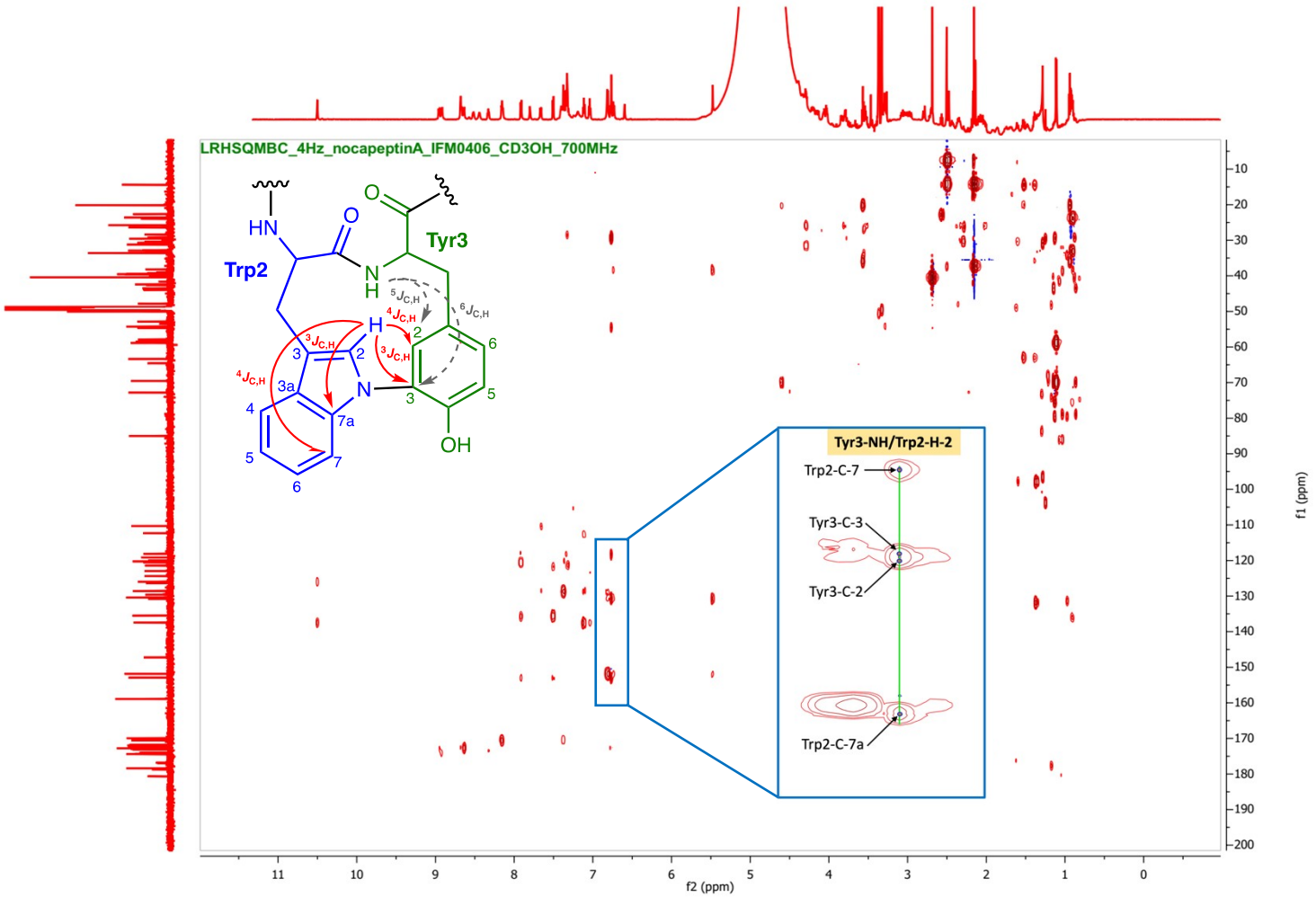


Figure S21. 700 MHz ^{1}H-^{13}C LR-HSQMBC spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4). Notably, the δ_H-resonances of Trp2-H-2 and Tyr3-NH are overlapping at 6.74 ppm and show cross correlations with Tyr3-C-2 and Tyr3-C-3 (see blue box). The experiment was adjusted for 4 Hz, which provides access to predominantly $^{3}J_{CH}$-$^{5}J_{CH}$ long-range correlations. Thus, theoretically, $^{5}J_{CH}$-based Tyr3-NH/Tyr3-C-2 and $^{6}J_{CH}$-based Tyr3-NH/Tyr3-C-3 cross correlations are possible (grey dashed arrows) but results commonly in very weak signals.[2] However, since strong cross-peaks were observed, the presence of very long-range correlations ($\geq {}^{5}J_{CH}$) was excluded and the observed cross-peaks of interest were interpretated as $^{3}J_{CH}$-based (Trp2-H-2 / Tyr3-C-3) and $^{4}J_{CH}$-based (Trp2-H-2 / Tyr3-C-2) couplings (red solid arrows). This conclusion was furthermore supported by the fact, that in the 4-Hz-LR-HSQMBC NMR spectrum overall no very long-range correlations ($\geq {}^{5}J_{CH}$) were detectable.

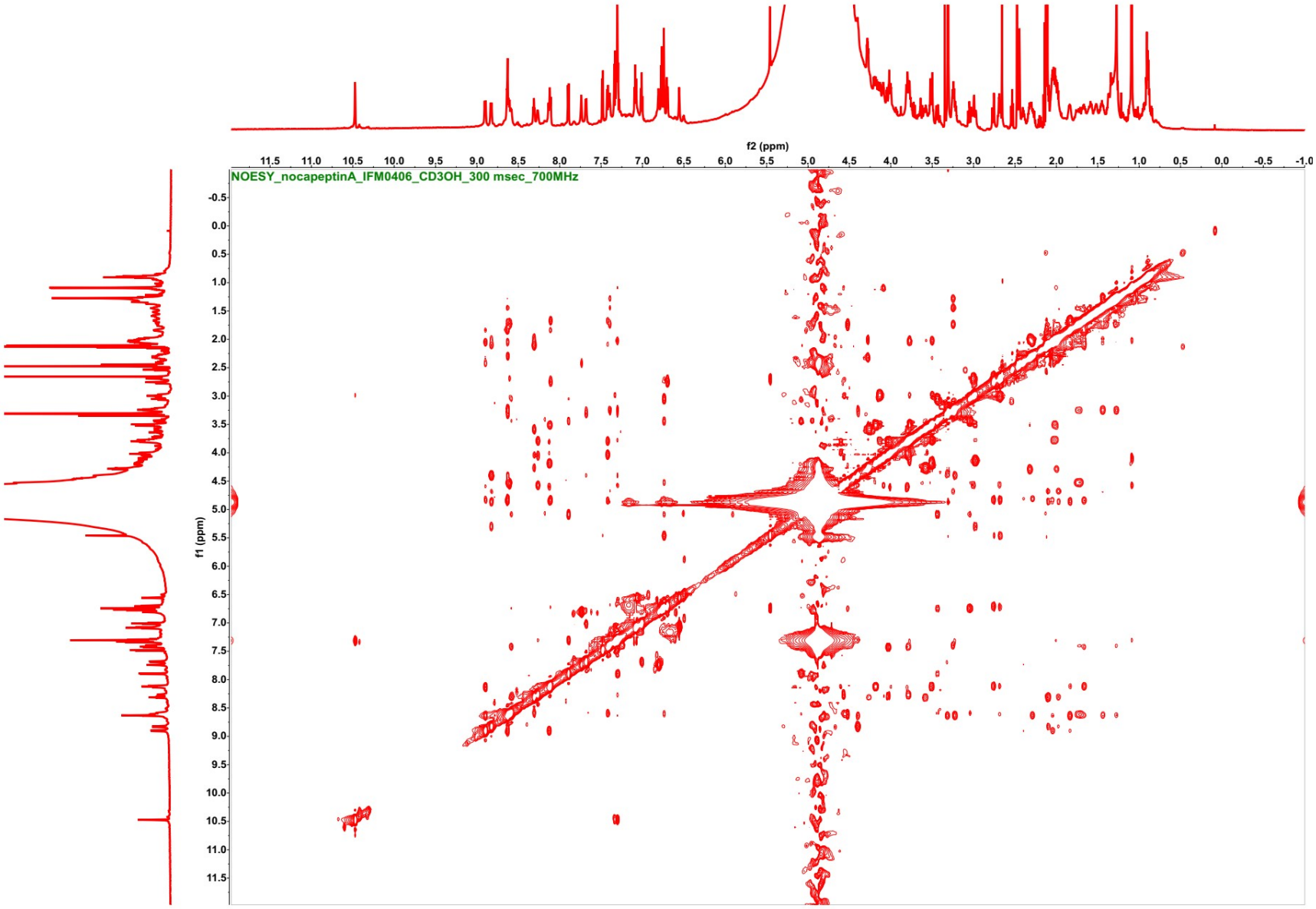


Figure S22. 700 MHz ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4) and a mixing time of 300 msec

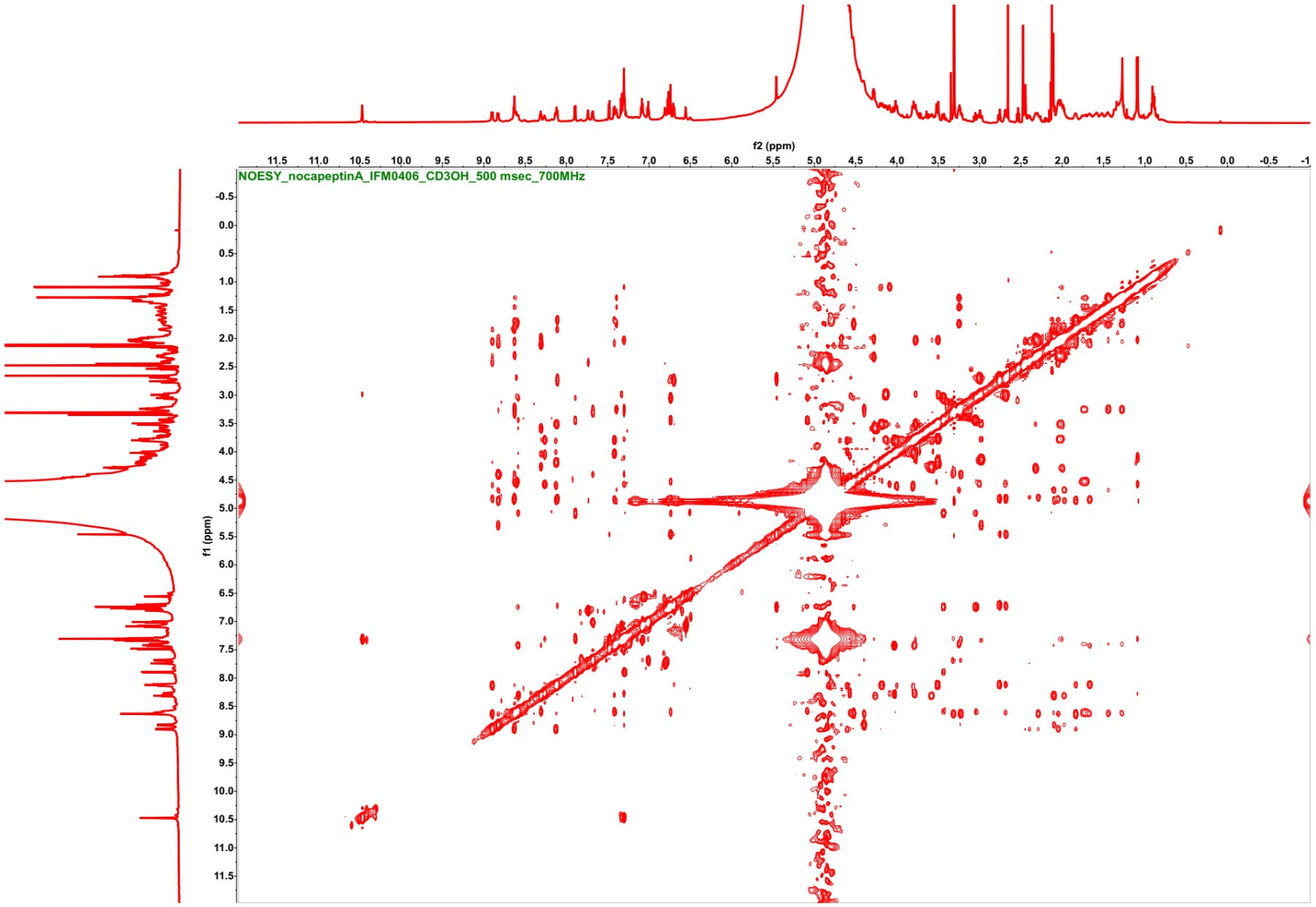


Figure S23. 700 MHz ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4) and a mixing time of 500 msec

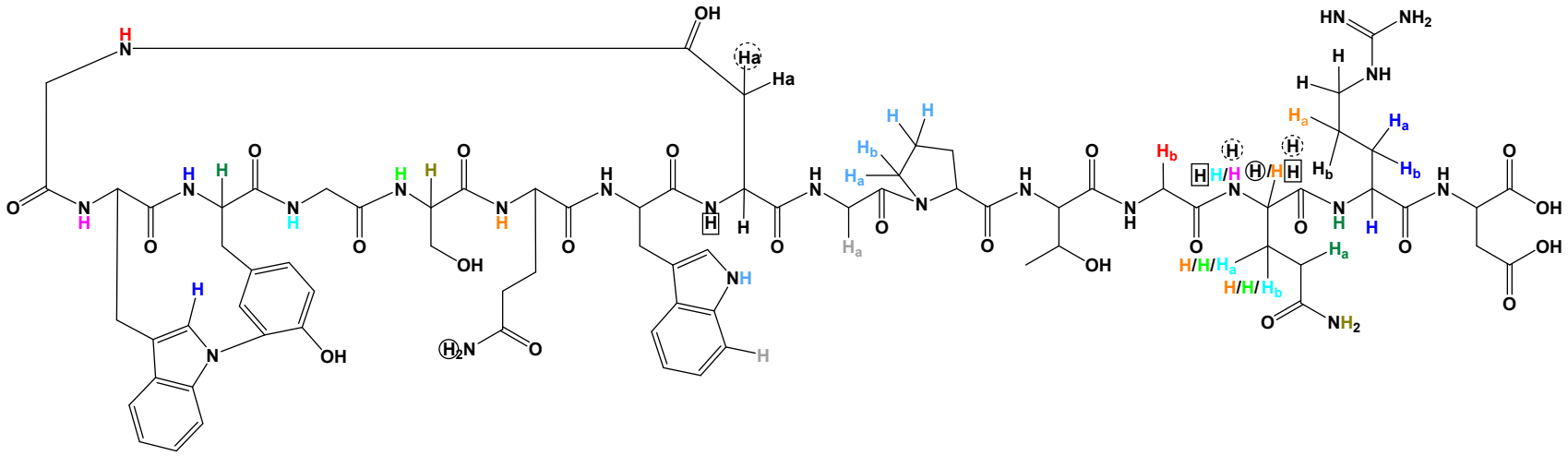


Figure S23A. NOE correlations, highlighting the threading property of nocapeptin A (**1**). Protons having similar patterns are in close proximity to each other

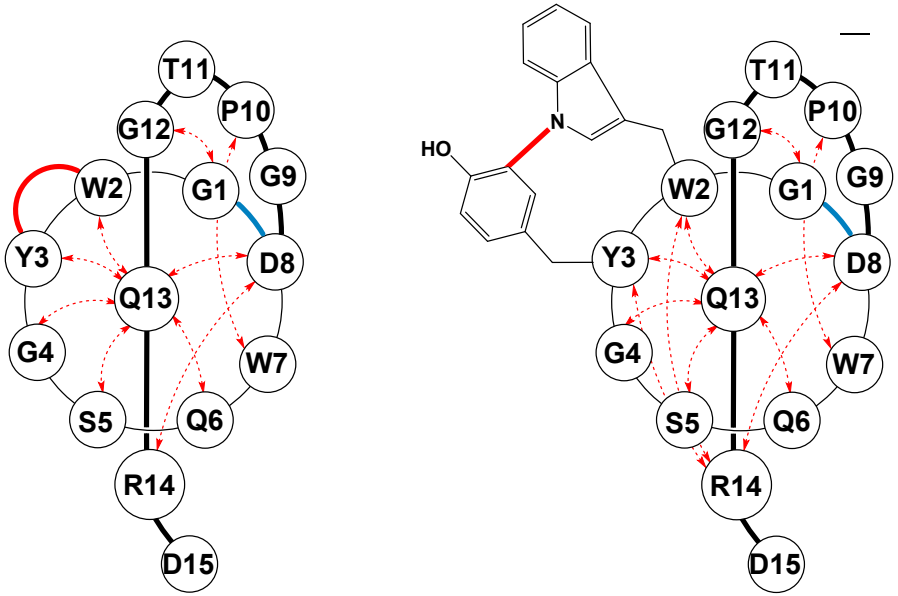


Figure S23B. Schematic representation of the lasso conformation of nocapeptin A (**1**) based on NOE correlations

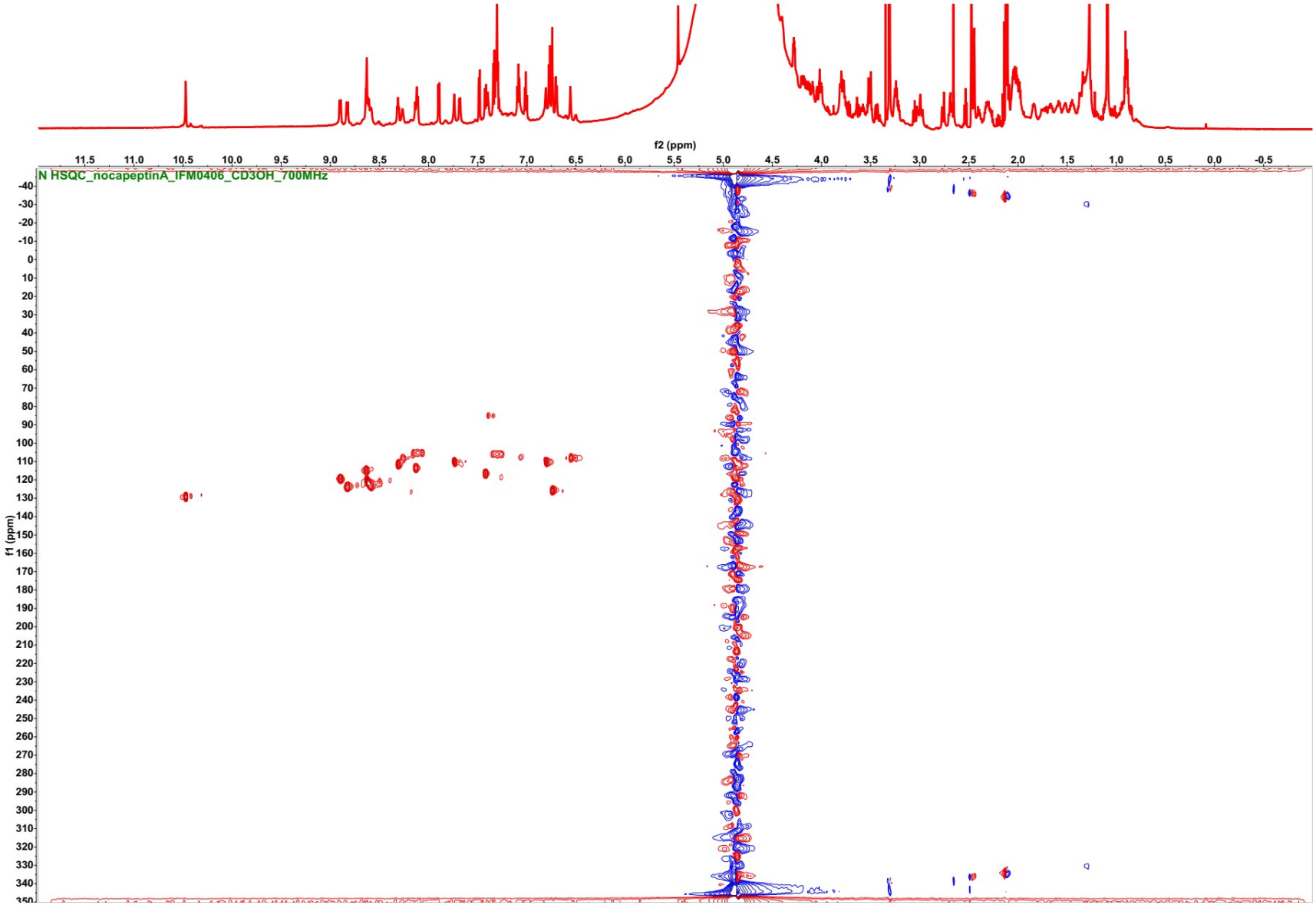


Figure S24. 700 MHz ^{1}H-^{15}N HSQC spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

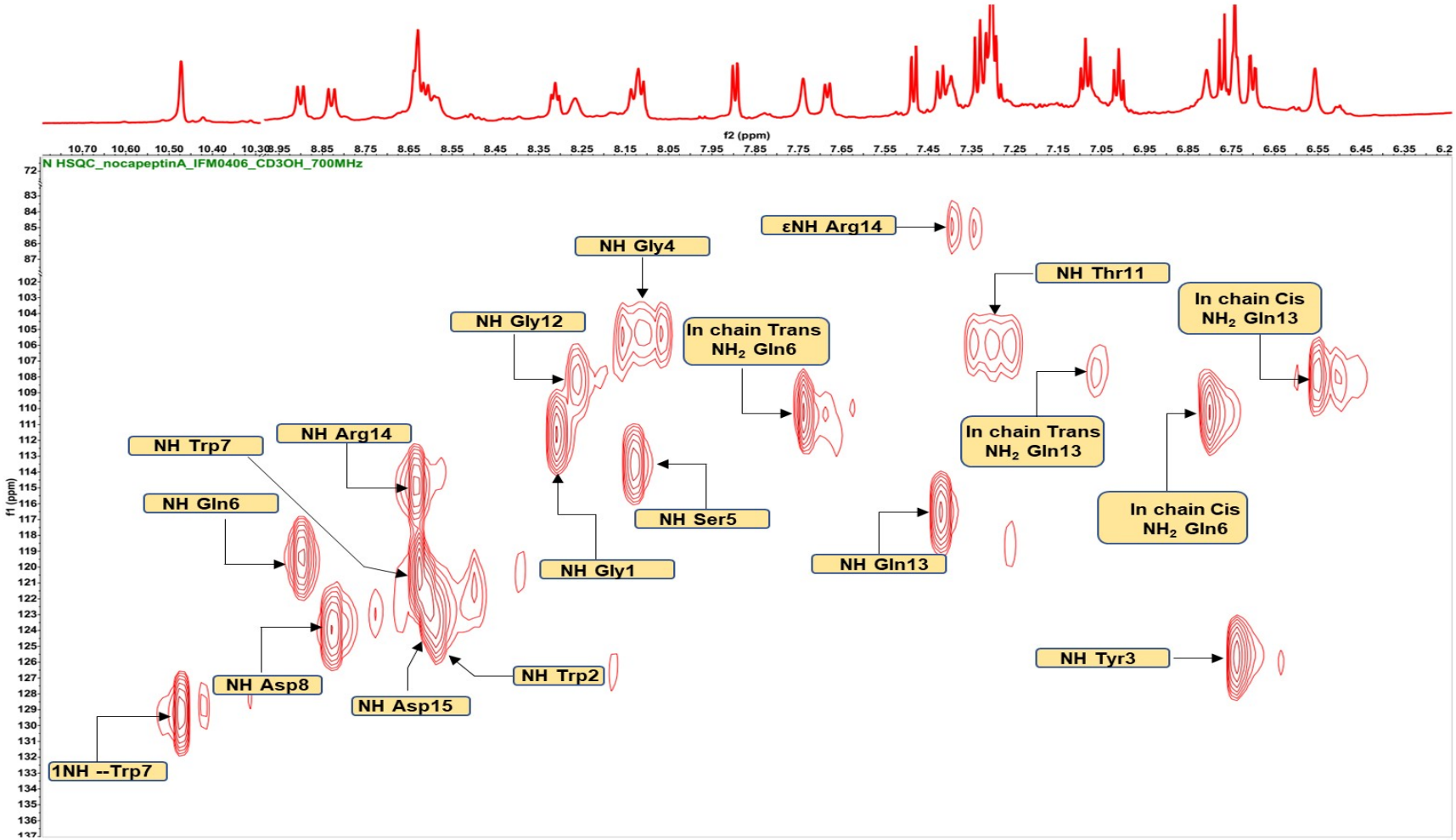


Figure S24A. Annotated ^{1}H-^{15}N HSQC spectrum of nocapeptin A (**1**)

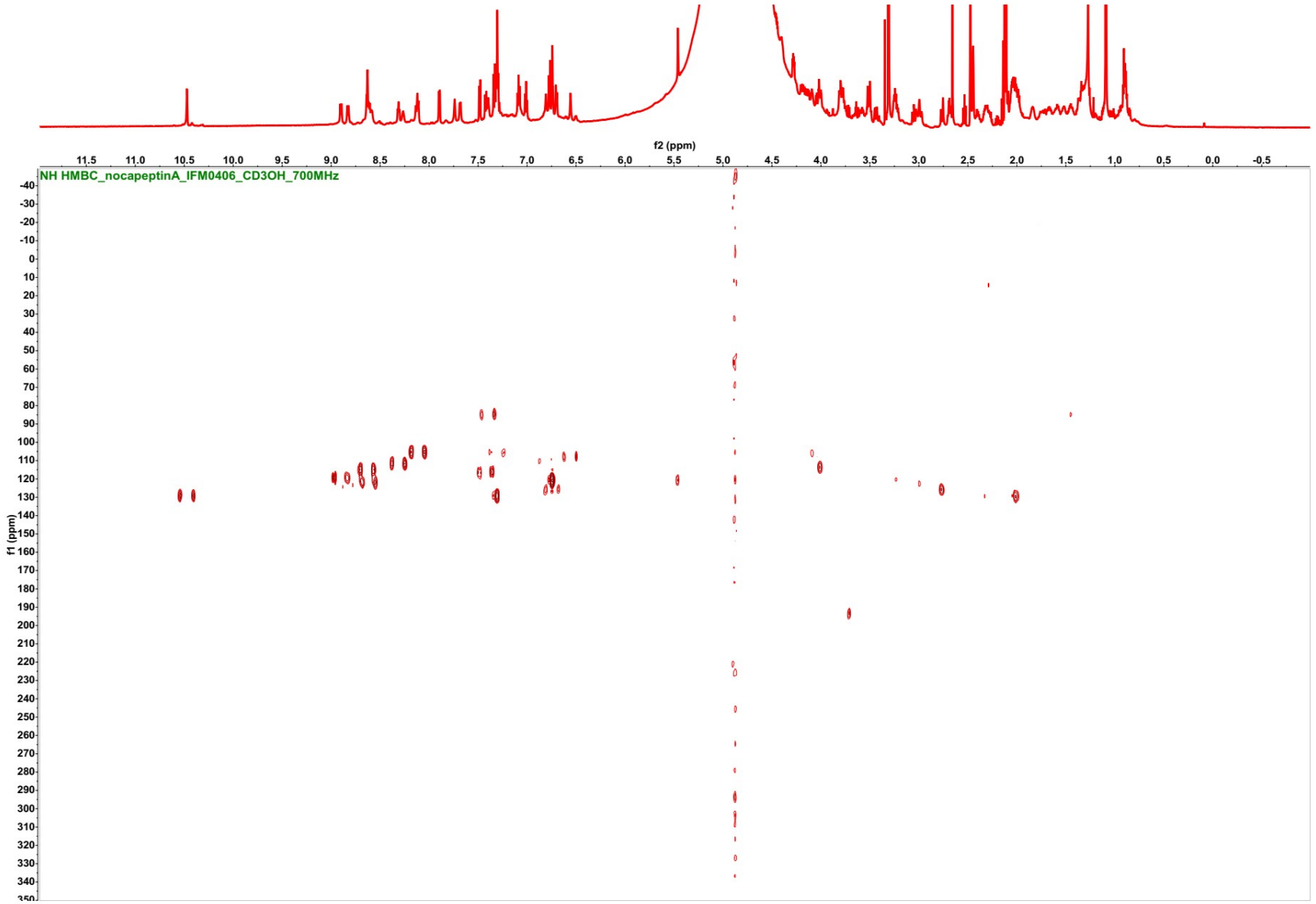


Figure S25. 700 MHz ^{1}H-^{15}N HMBC spectrum of nocapeptin A (**1**) in d_3-CH_3OH/H_2O (96:4)

Table S7. ^{1}H, ^{13}C, and ^{15}N-NMR data of nocapeptin A (**1**) [d_3-CH_3OH/H_2O (96:4); 700/176/71 MHz]. All chemical shifts are given in ppm.

Residue	Position	δ_C/δ_N	δ_H, mult (*J* in Hz)	Residue	Position	δ_C/δ_N[i]	δ_H, mult (*J* in Hz)
Gly1	CO	171.96, C	---	**Trp7**	CO	174.01, C	---
	α	43.53, CH_2	3.59, dd (11.6, 5.0) + 4.27, m		α	58.45, CH	4.41, m
	-NH-	111.70	8.31, t (6.3)		β	28.42, CH_2	3.24, m + 3.32, m
Trp2[a]	CO	171.82, C	---		1(NH)	129.20	10.47, s
	α	54.55,[c] CH	5.09, m		2	125.93, CH	7.30, overlapped
	β	29.14, CH_2	3.05, m + 3.44, dm (13.4)		3	110.28, C	---
	1(N)	120.80	---		3a	128.56,[e] C	---
	2	147.20, CH	6.74, overlapped		4	119.97, CH	7.68, d (7.7)
	3	120.25, C	---		5	120.61, CH	7.01, t (7.5)
	3a	135.51, C	---		6	122.95, CH	7.09, dd (7.7, 7.6)
	4	121.44, CH	7.90, d (7.7)		7	112.36, CH	7.34, d (8.3)
	5	124.39, CH	7.30, overlapped		7a	137.49, C	---
	6	126.36, CH	7.32, overlapped		-NH-	120.40	8.64, overlapped
	7	118.10, CH	7.48, d (7.6)	**Asp8**	CO	173.00, C	---
	7a	152.95, C	---		α	49.68; CH	4.67, overlapped by H_2O-peak
	-NH-	123.70	8.59, m		β	41.60, CH_2	1.99, m + 2.11, m
Tyr3[a]	CO	170.55, C	---		γ	173.09, C	---
	α	54.64,[c] CH	4.84, overlapped by H_2O-peak		-NH-	123.99	8.82, d (9,8)
	β	38.16, CH_2	2.69, m + 2.76, m	**Gly9**	CO	n.d.[g]	---
	1	128.51,[e] C	---		α	42.66, CH_2	2.99, m + 4.15, dd (16.3, 8.3)
	2	130.77, CH	5.46, d (1.9)		-NH-	n.d.[g]	5.30, overlapped by H_2O-peak
	3	130.42, C	---	**Pro10**[h]	CO	174.30, C	---
	4	151.81, C	---		α	63.40, CH	4.28, m
	5	119.12, CH	6.77, d (8.3)		β	31.25, CH_2	2.00, m + 2.32, m
	6	129.66, CH	6.70, dd (8.3, 1.9)		γ	25.76, CH_2	2.03, m
	OH	---	n.d.[g]		δ	47.78, CH_2	3.51, m + 3.78, m
	-NH-	125.78	6.74, m		=N-	129.72	---
Gly4	CO	170.27,[d] C	---	**Thr11**	CO	172.77, C	---
	α	43.96, CH_2	3.51, m + 4.18, dd (17.4, 7.8)		α	58.82, CH	4.58, overlapped by H_2O-peak
	-NH-	105.43	8.12, d (7.8)		β	69.56, CH	4.09, m
Ser5	CO	172.77, C	---		γ	20.16, CH_3	1.09, d (6.3)
	α	57.87, CH	4.61, overlapped by H_2O-peak		OH	---	n.d.
	β	64.60, CH_2	3.80, m + 4.01, dd (11.3, 4.1)		-NH-	106.11	7.31, m
	OH	---	n.d.[g]	**Gly12**	CO	170.30,[d] C	---
	-NH-	113.60	8.09, d (9.3)		α	43.80, CH_2	3.80, m + 4.04, dd (17.6, 5.0)
Gln6	CO	174.43, C	---		-NH-	108.21	8.26, br s
	α	52.83, CH	4.82, overlapped by H_2O-peak	**Gln13**	CO	172.58, C	---
	β	30.36, CH_2	2.04, m + 2.29, m		α	54.24,[f] CH	4.87, m
	γ	32.80, CH_2	2.40, m + 2.45, m		β	33.45, CH_2	1.66, m + 1.84, m
	δ	178.70, C	---		γ	33.62, CH_2	2.05, m + 2.12, m
	$-NH_2$	110.25	6.81, br s (*cis*)[b]		δ	178.44, C	---
			7.72, br s (*trans*)[b]		$-NH_2$	107.92	6.56, br s (*cis*)[b]
	-NH-	119.40	8.89, d (9.5)				7.07, overlapped (*trans*)[b]
					-NH-	116.58	7.43, d (9.1)

Arg14	CO	172.46, C	---------
	α	54.26,[f] CH	4.53, overlapped by H_2O-peak
	β	30.80, CH_2	1.75, m
	γ	23.99, CH_2	1.28, m + 1.44, m
	δ	42.37, CH_2	3.25, m
	εNH	85.05	7.39, m
	ζ	158.90, C	---------
	NHη1	n.d.[g]	n.d.[g]
	NH_2η1	n.d.[g]	n.d.[g]
	-NH-	114.90	8.64, overlapped
Asp15	CO_2H	176.85, C	---------
	α	53.51, CH	4.58, m
	β	38.98, CH_2	2.70, m + 3.01, m
	γ	177.00, C	---------
	-NH-	122.77	8.62, overlapped

[a] Residues involved in the C-N biaryl crosslink. [b] The partial double-bond character of the carboxamide bond leads to the observation of two distinct resonances for the corresponding NH_2-protons of Gln.[23] [c,d,e,f] interchangeable resonances. [g] n.d.: not detected. [h] Amides attached through a proline nitrogen are known to be capable of existing as equilibrating rotamers, with literature empirical rules establishing that ^{13}C NMR chemical shift differences between proline β and γ carbon resonances are characteristic of *cis* ($\Delta\beta\gamma \sim 8$–11 ppm) vs. *trans* $\Delta\beta\gamma \sim 3$–6 ppm) rotamers, respectively. Thus, on the basis of the calculated shift difference $\Delta\beta\gamma$ = 5.5 ppm, Pro10 in **1** was determined to adopt a *trans* configuration.[24] However, since in proteins it was observed that the above mentioned empirical rule can fail, particularly in the $\Delta\beta\gamma$-range of 4.8 to 9.15 ppm,[25] a conventional NOE-based analysis is recommended for final confirmation. Inspection of the ^1H-^1H-NOESY spectrum revealed crosspeaks between Pro10 H_2-δ and Gly9 H_2-α, which unambiguously corroborated the *trans* conformation for Pro10.[i] ^{15}N-shift values were extracted from the corresponding ^1H-^{15}N HSQC NMR spectrum.

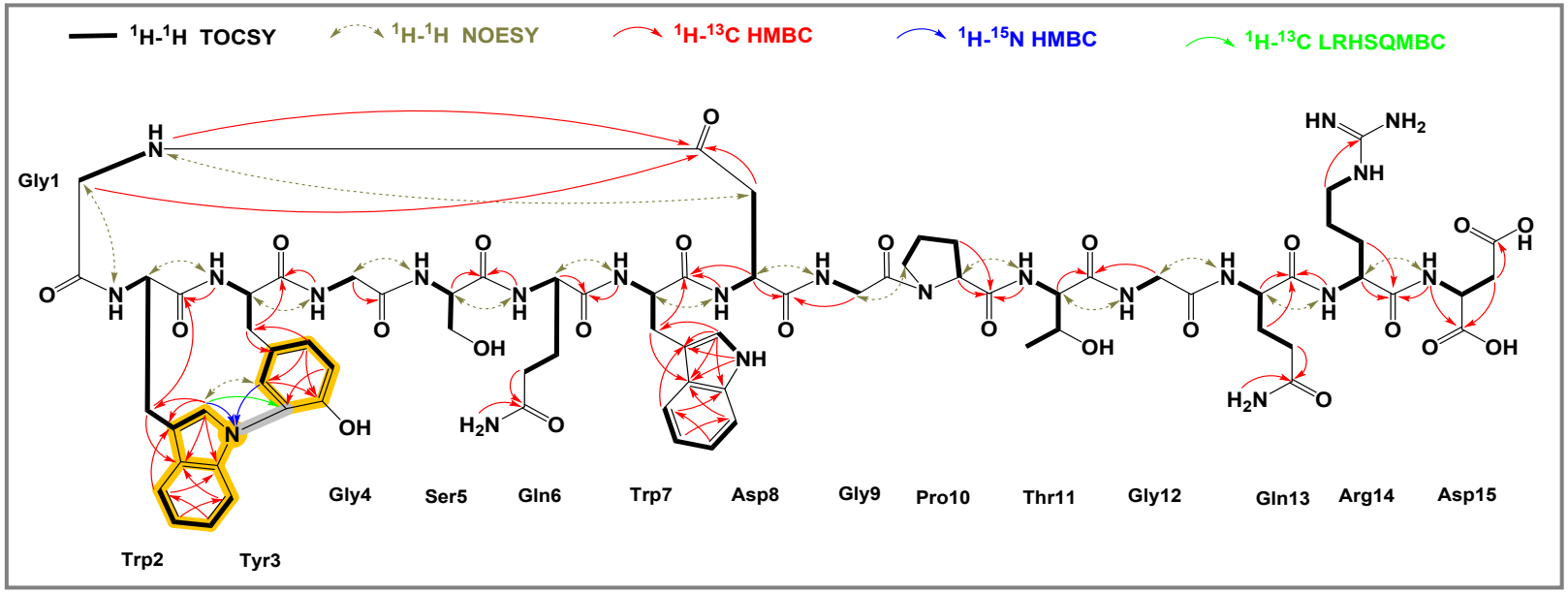


Figure S26. NMR key correlations of nocapeptin A (**1**)

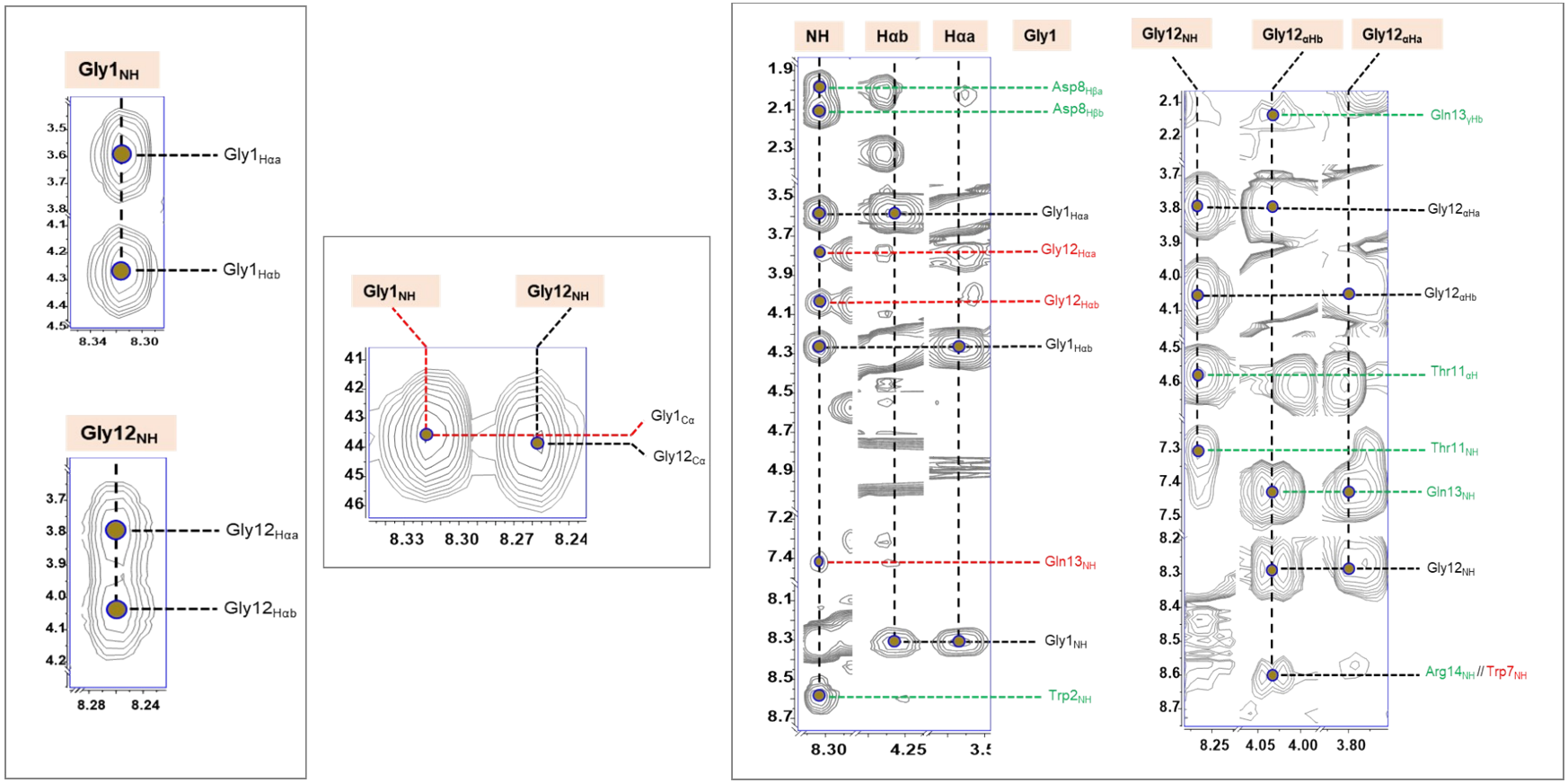


Figure S27. Annotated ^{1}H-^{1}H TOCSY (left), ^{1}H-^{13}C HSQC-TOCSY (middle) and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residues **Gly1** and **Gly12**

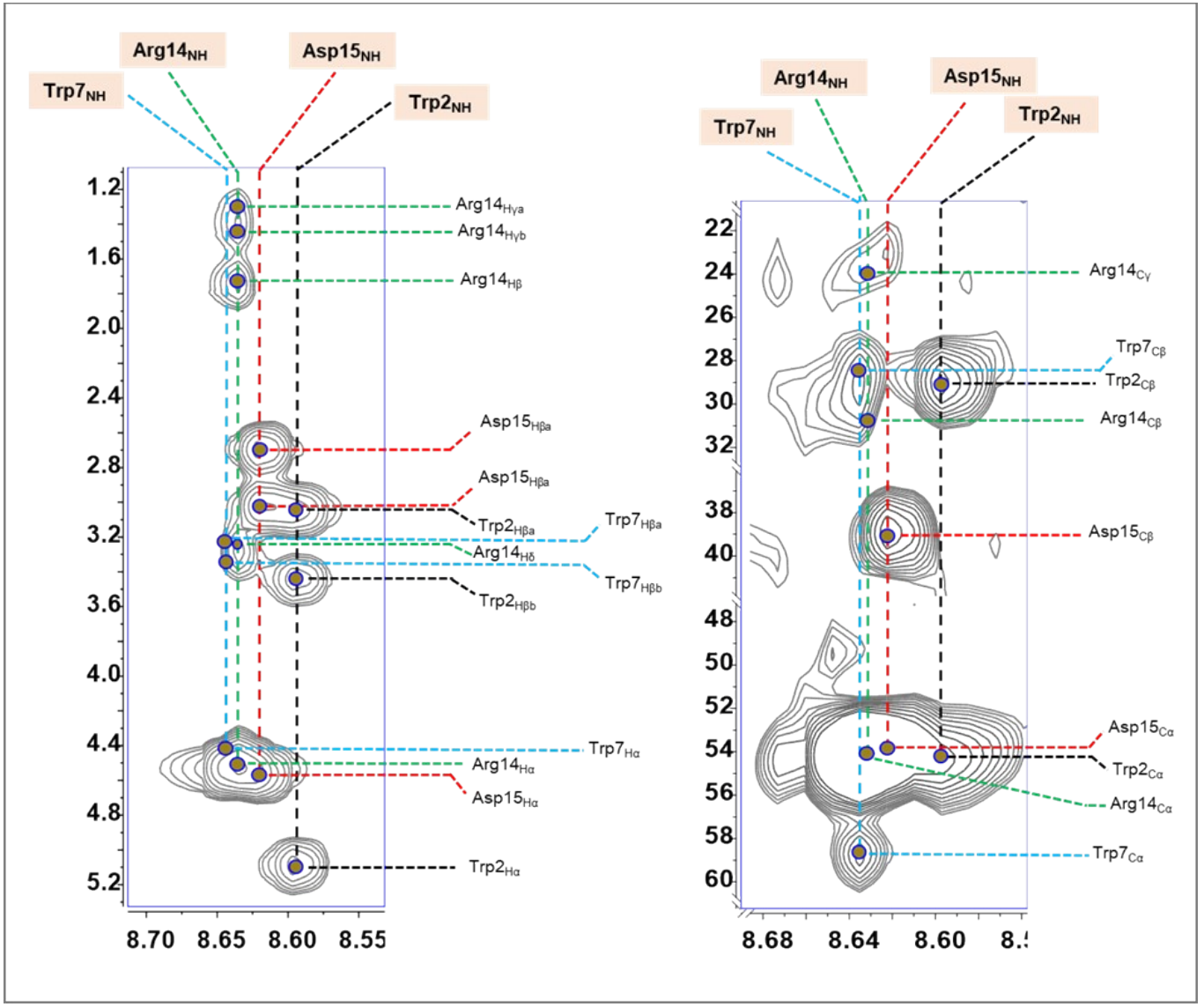


Figure S28. Annotated ^{1}H-^{1}H TOCSY (left), and ^{1}H-^{13}C HSQC-TOCSY (right) spectra of nocapeptin A (**1**) depicting residue **Trp2**

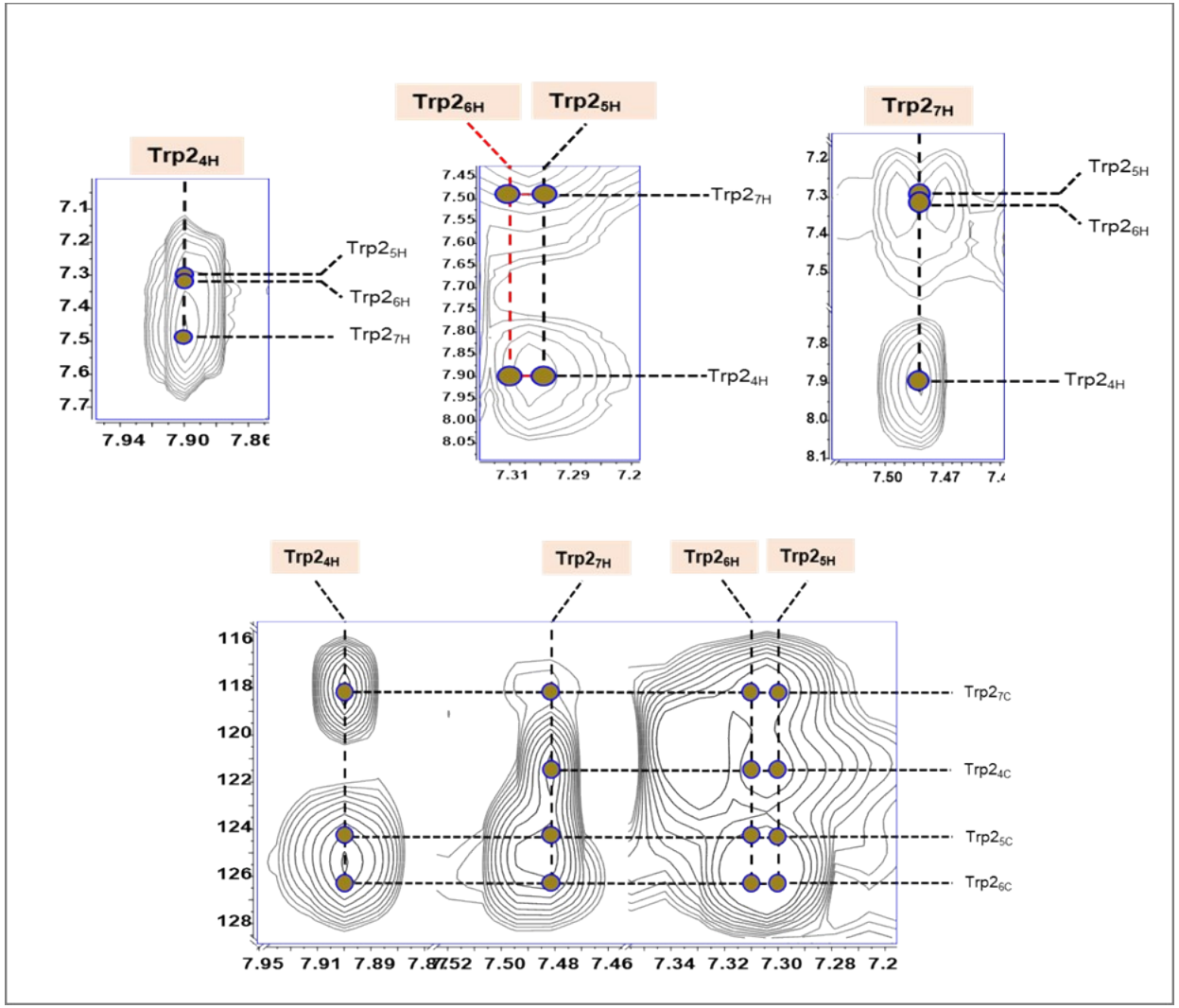


Figure S28A. Annotated ^{1}H-^{1}H TOCSY (upper), and ^{1}H-^{13}C HSQC-TOCSY (lower) spectra of nocapeptin A (**1**) depicting residue **Trp2**

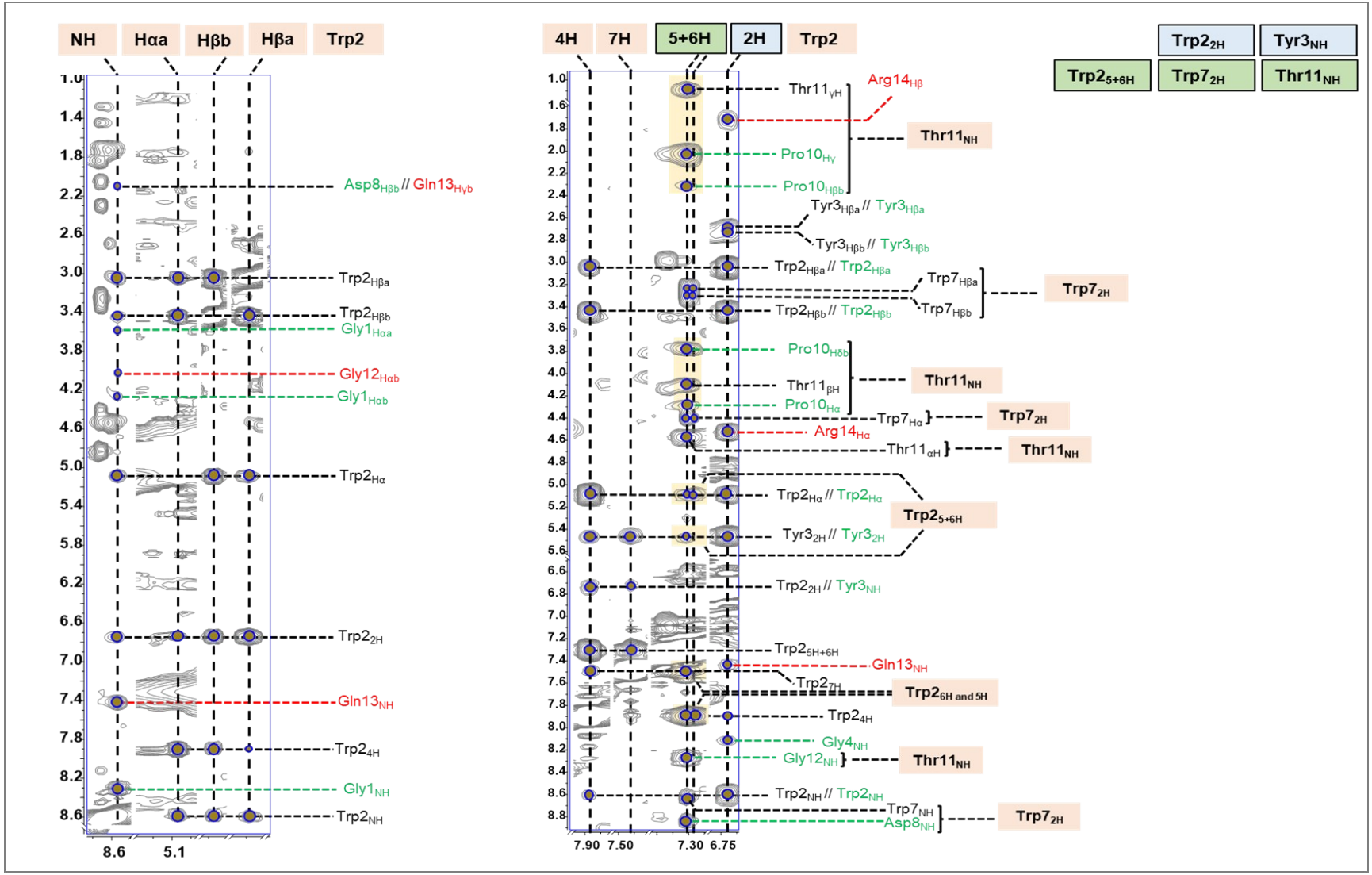


Figure S28B. Annotated ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) depicting residue **Trp2**

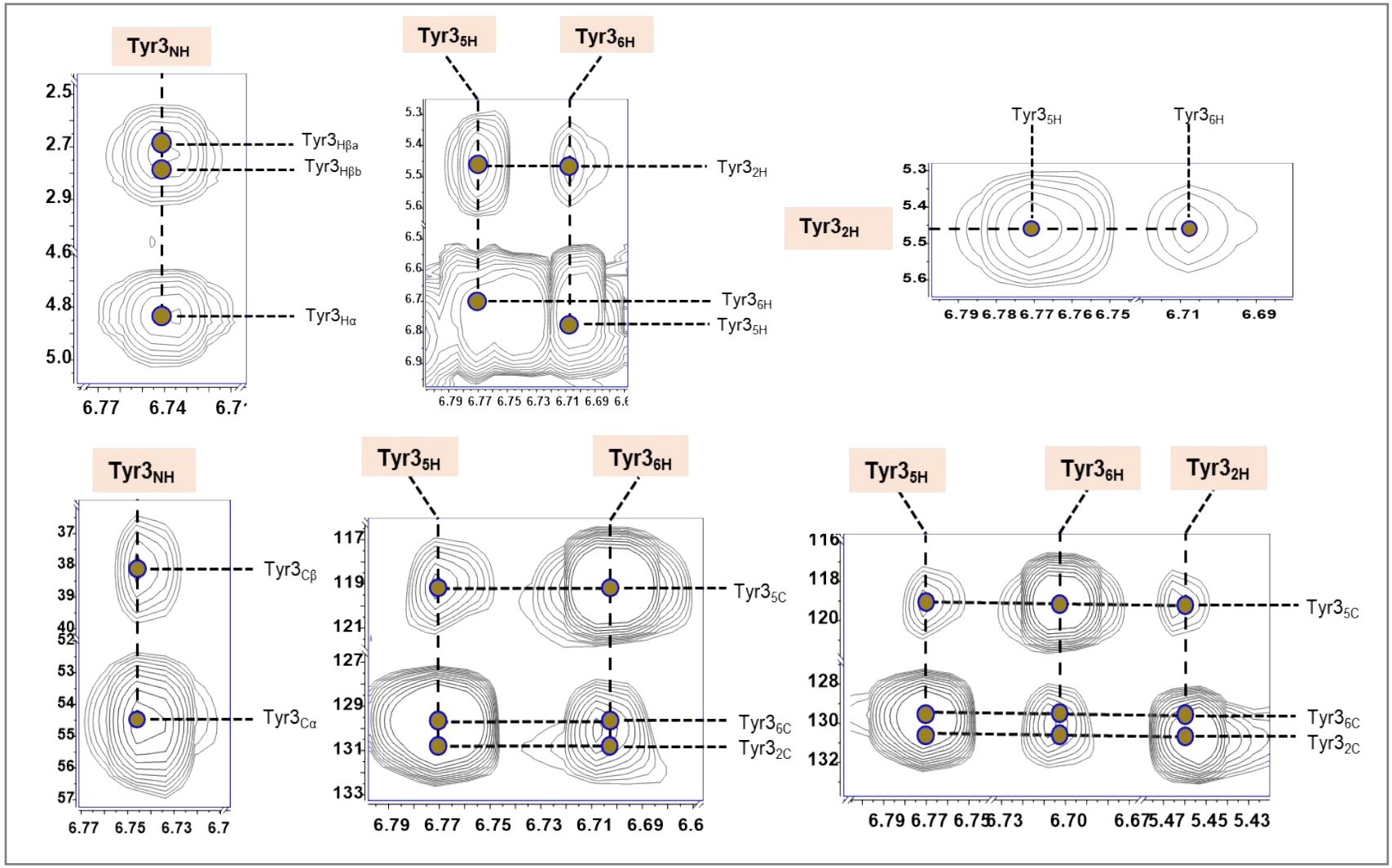


Figure S29. Annotated ^{1}H-^{1}H TOCSY (upper), and ^{1}H-^{13}C HSQC-TOCSY (lower) spectra of nocapeptin A (**1**) depicting residue **Tyr3**

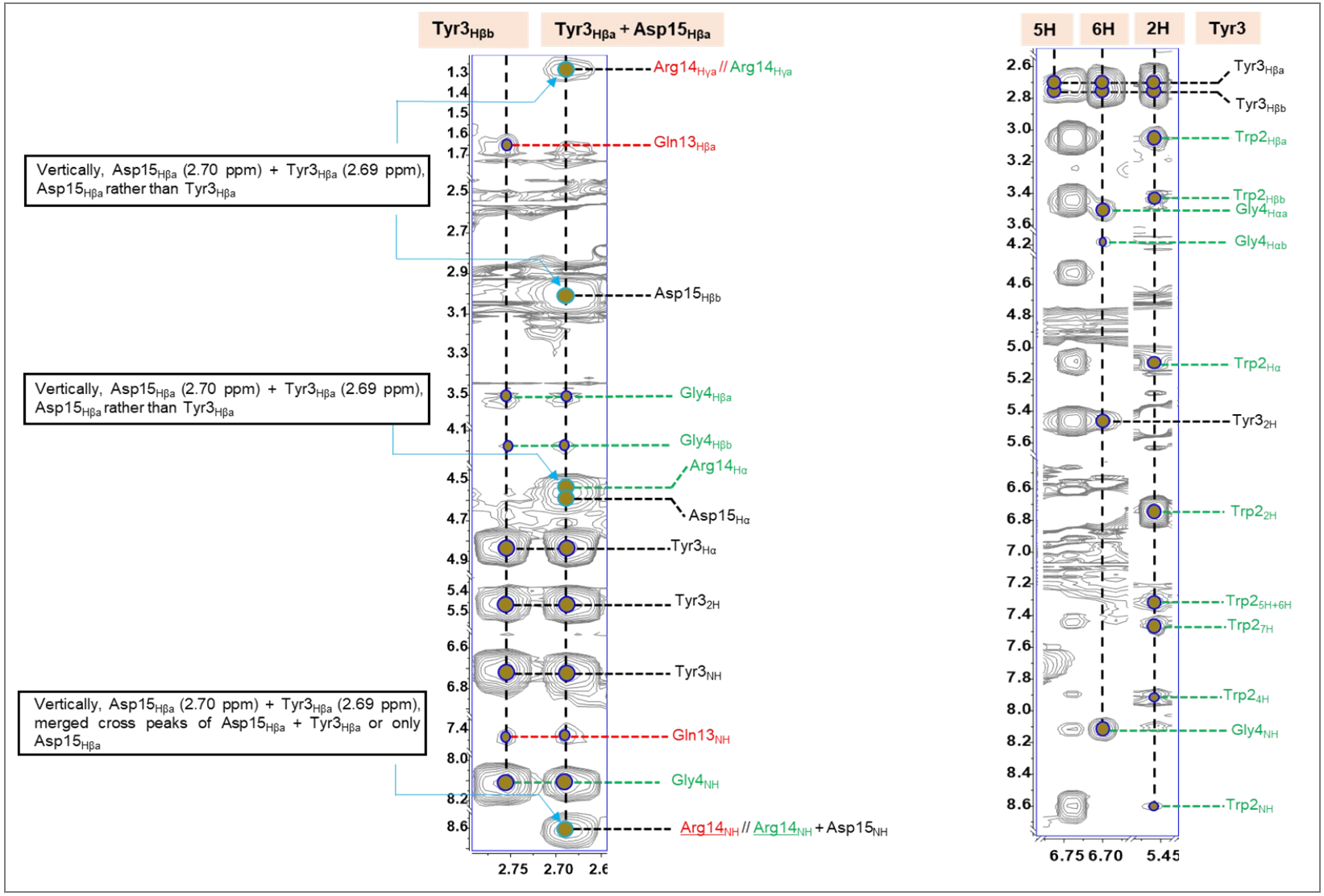


Figure S29A. Annotated ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) depicting residue **Tyr3**

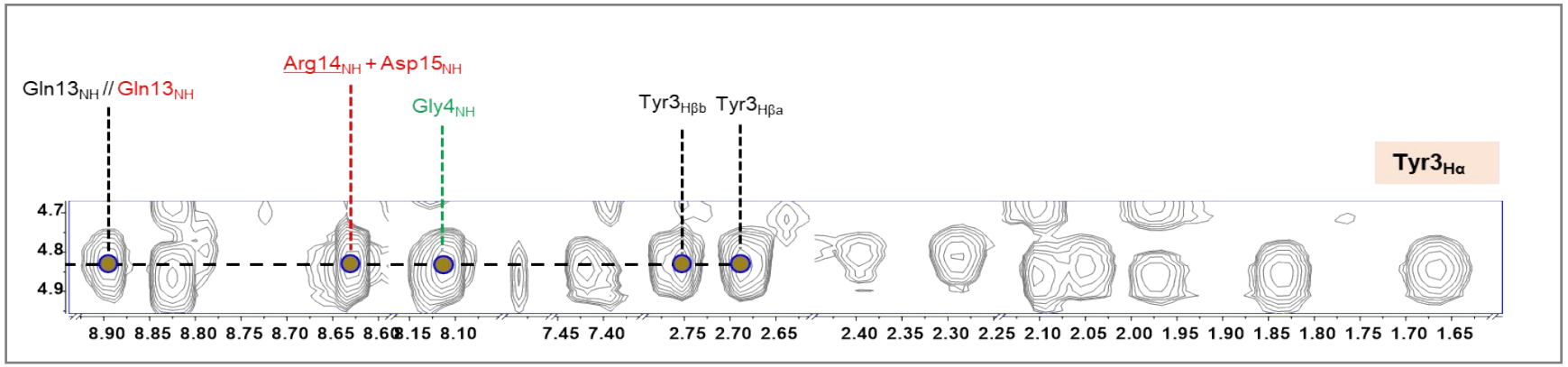


Figure S29B. Annotated ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) depicting residue **Tyr3**

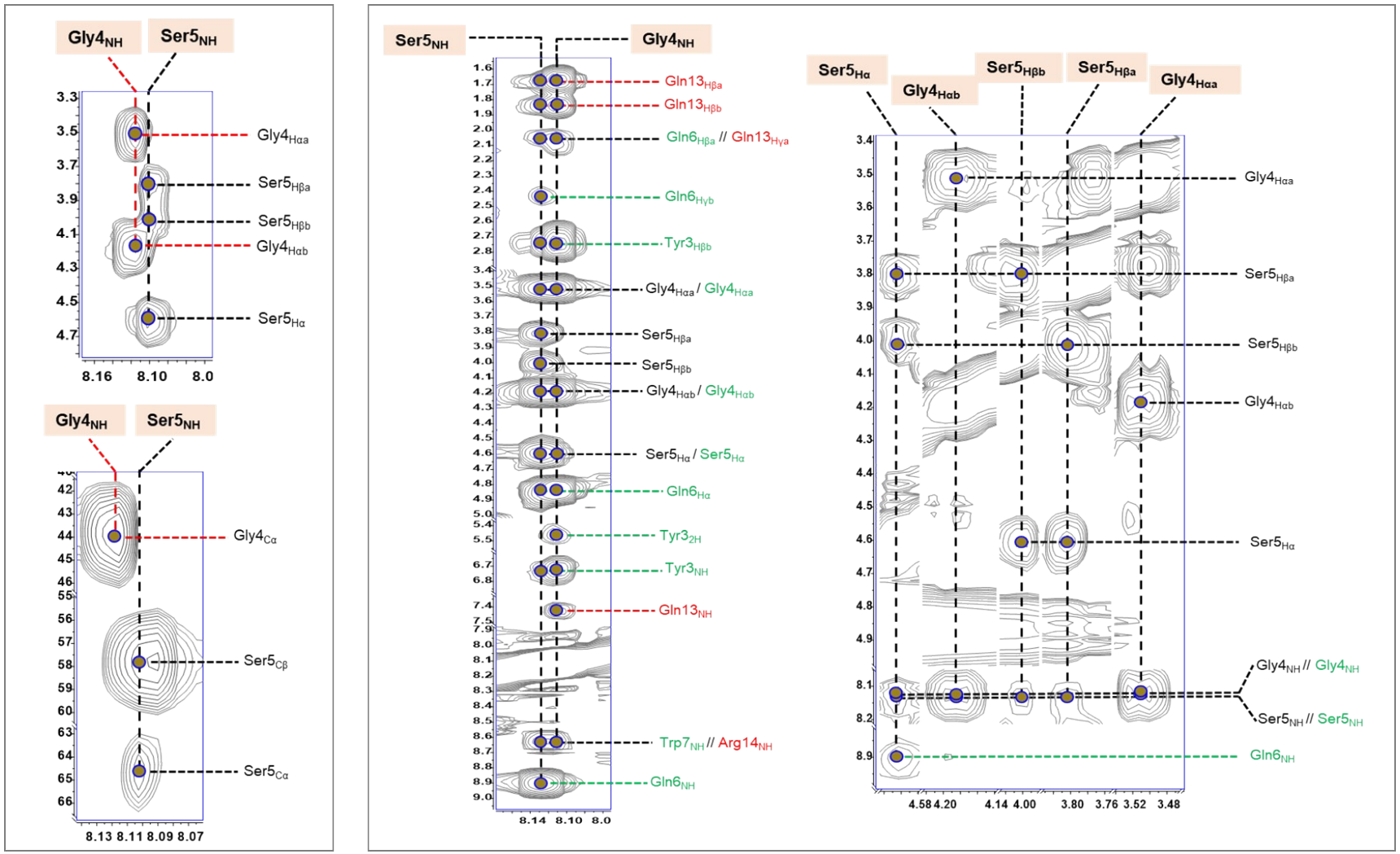


Figure S30. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residues **Gly4** and **Ser5**

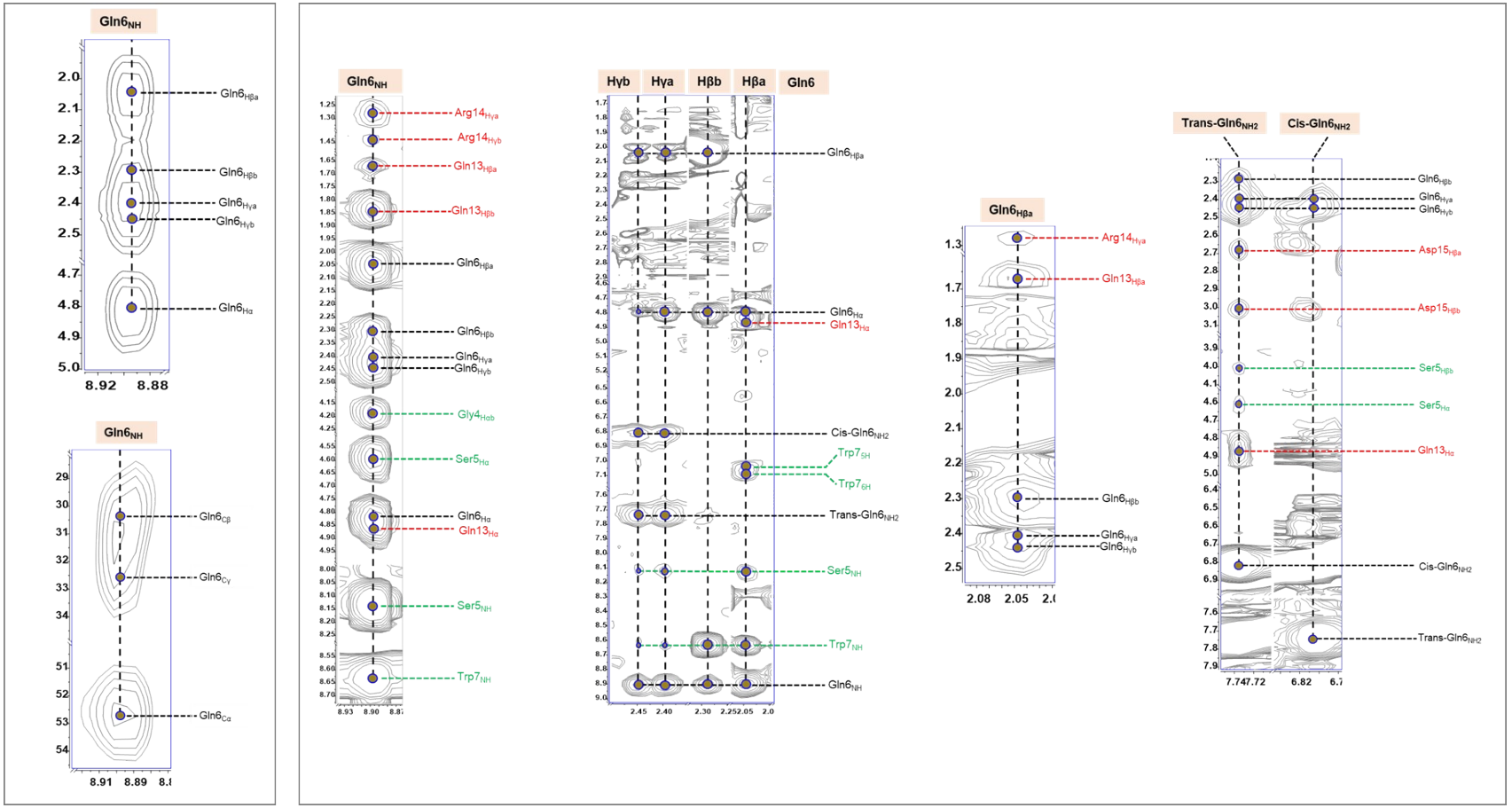
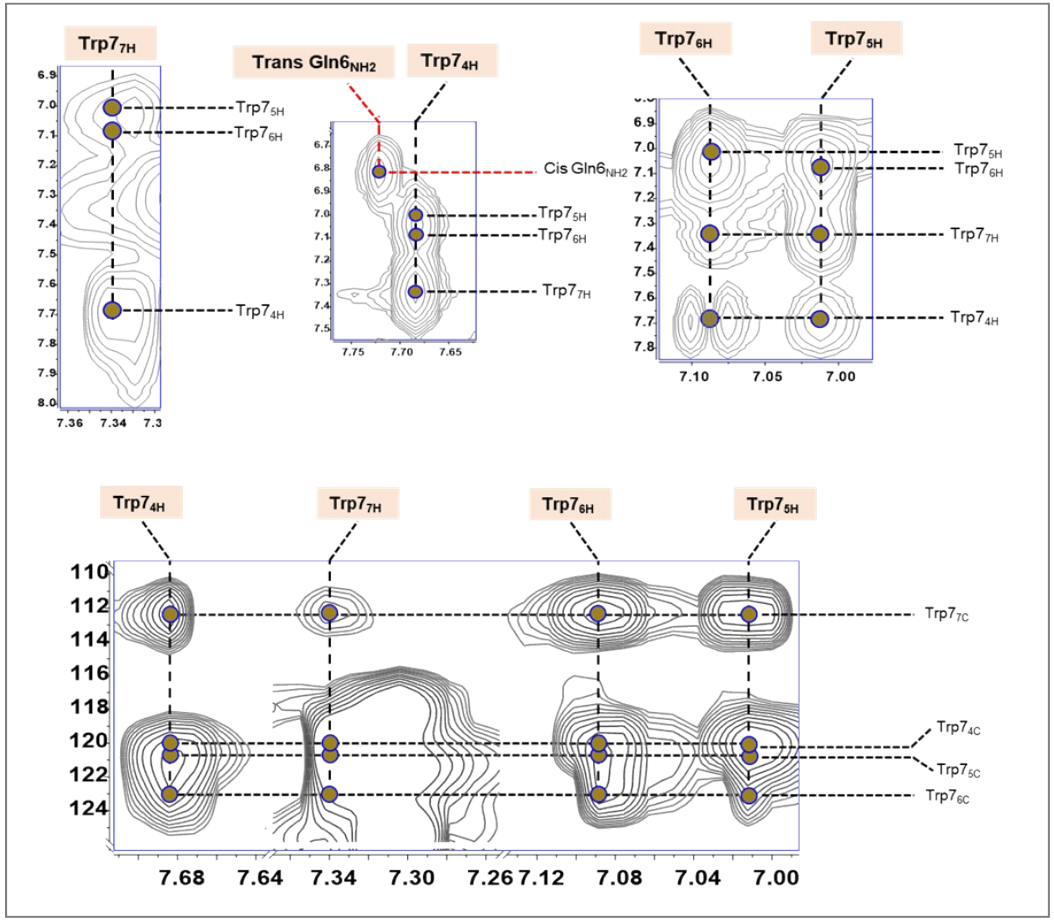


Figure S31. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Gln6**



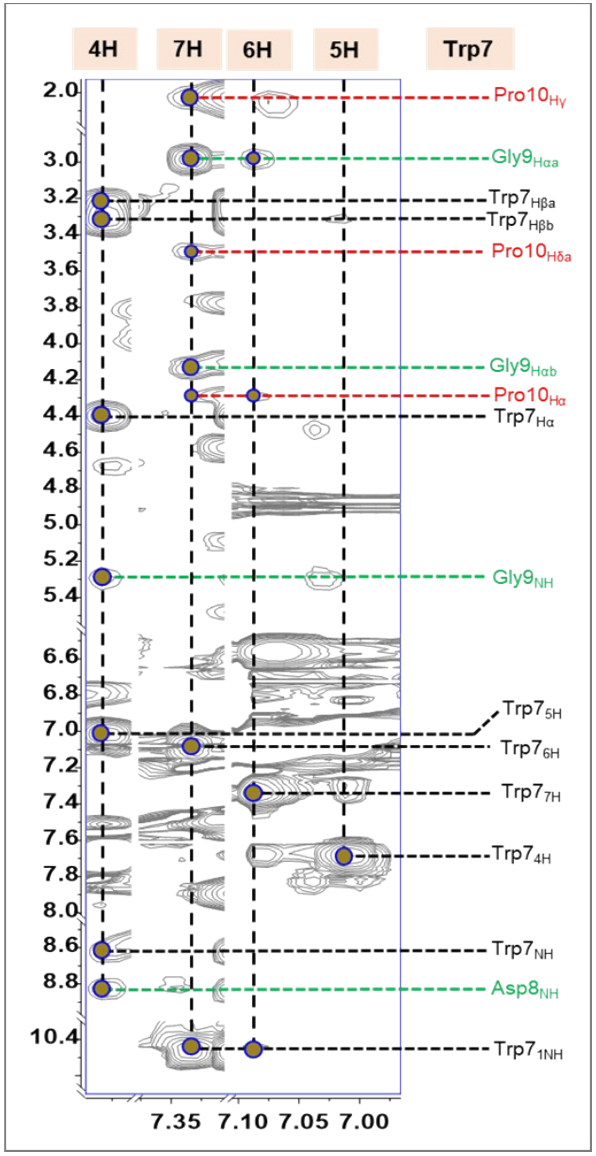


Figure S32. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Trp7**

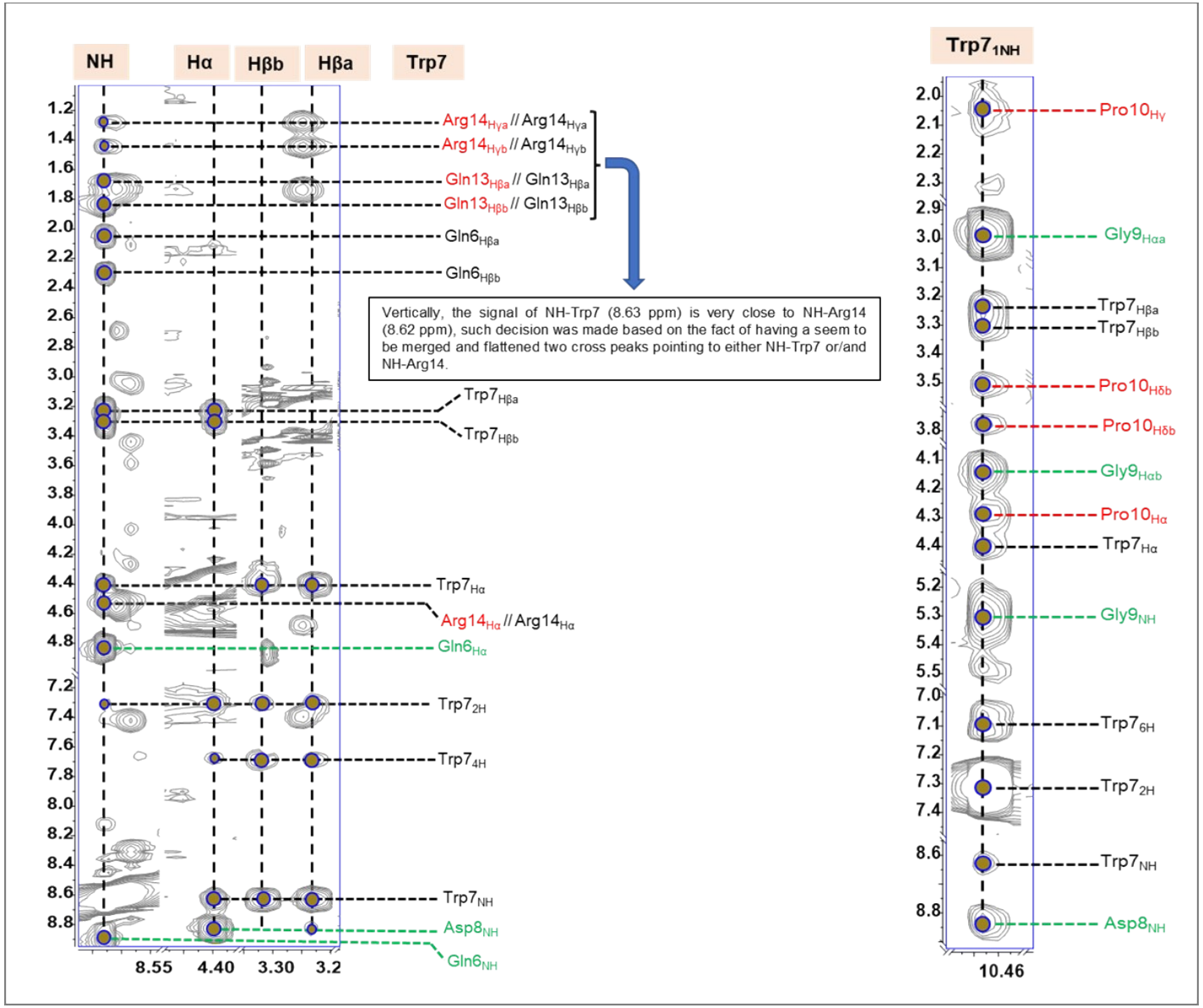


Figure S32A. Annotated ^{1}H-^{1}H NOESY spectra of nocapeptin A (**1**) depicting residue **Trp7**

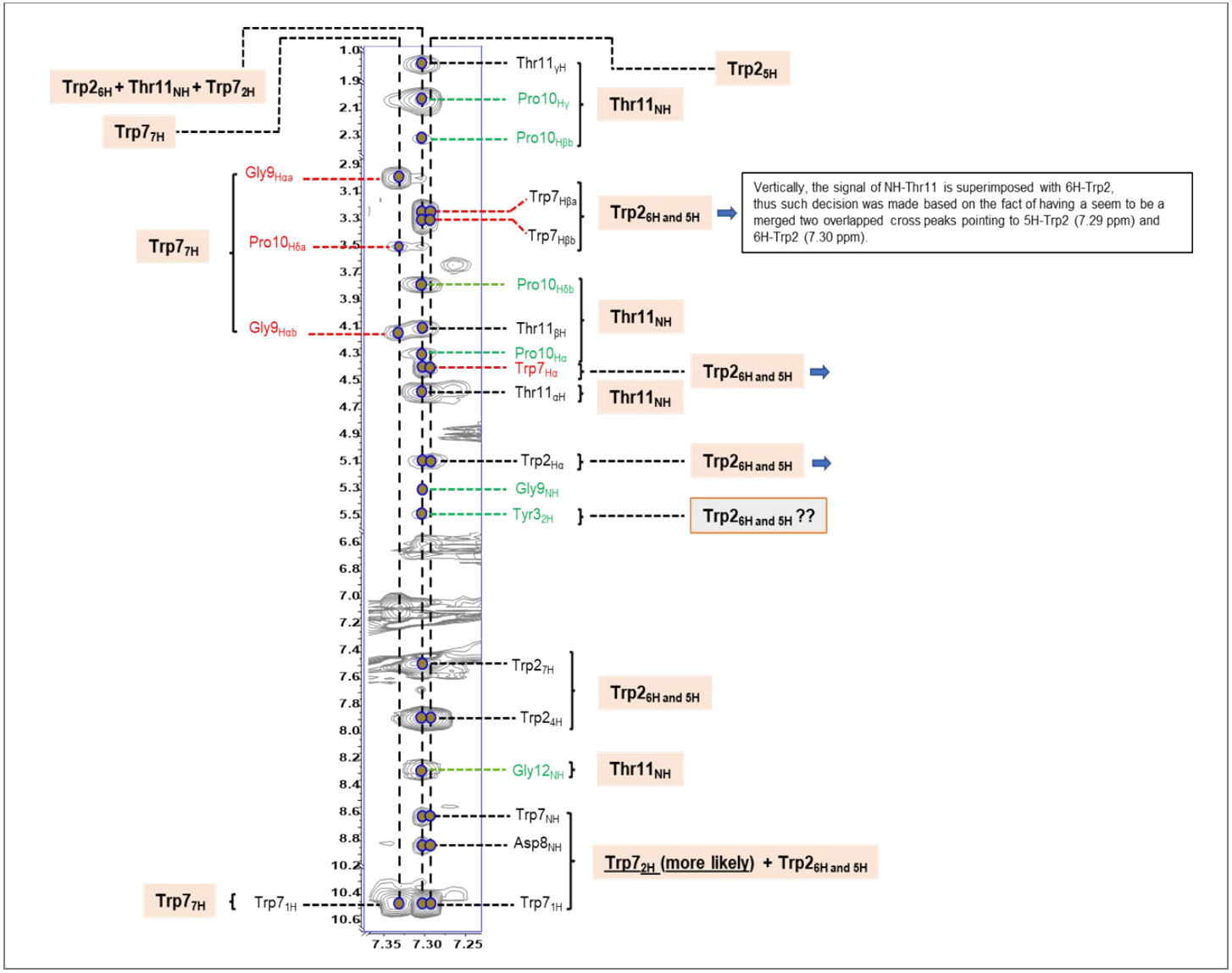


Figure S32B. Annotated ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) depicting residue **Trp7**

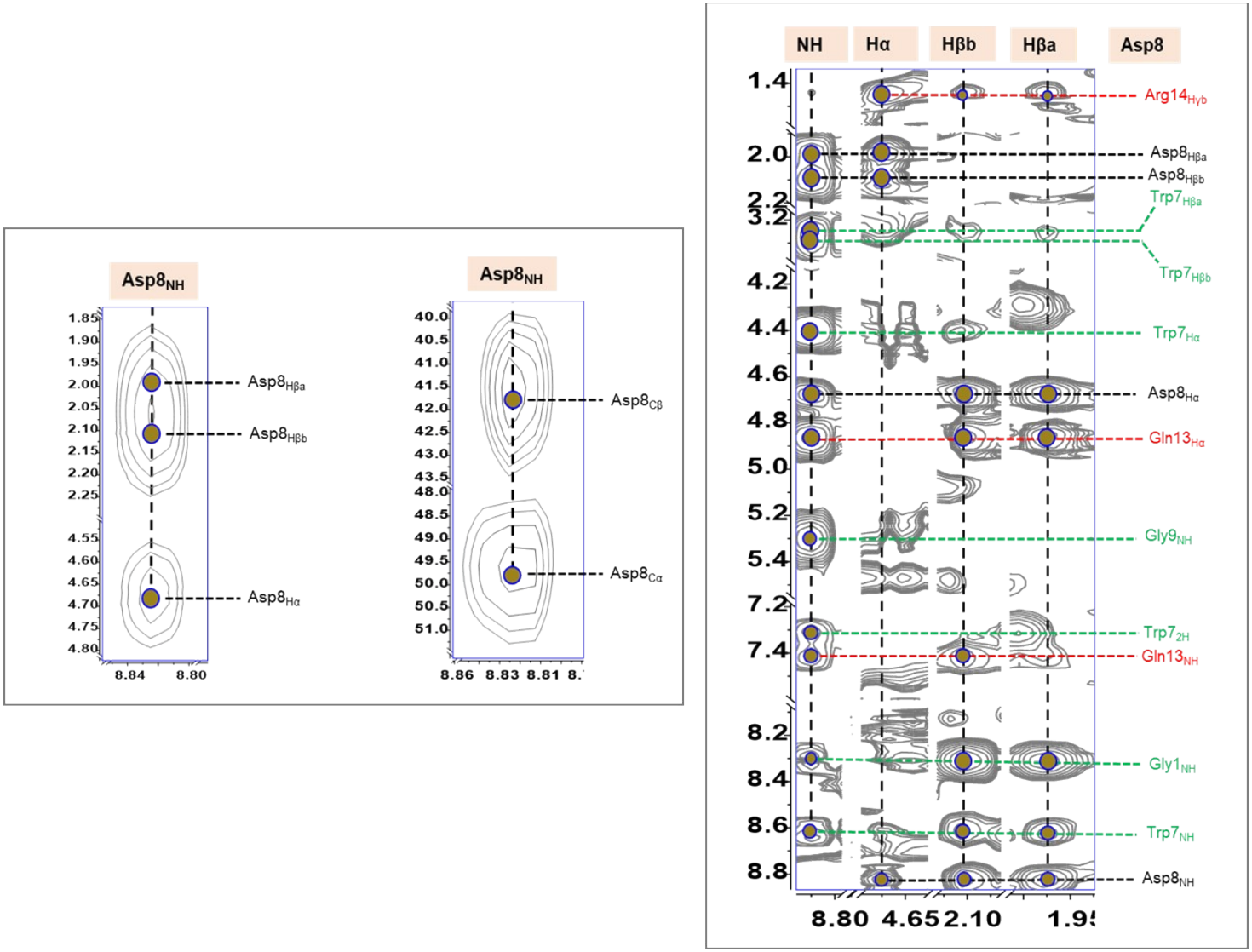


Figure S33. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Asp8**

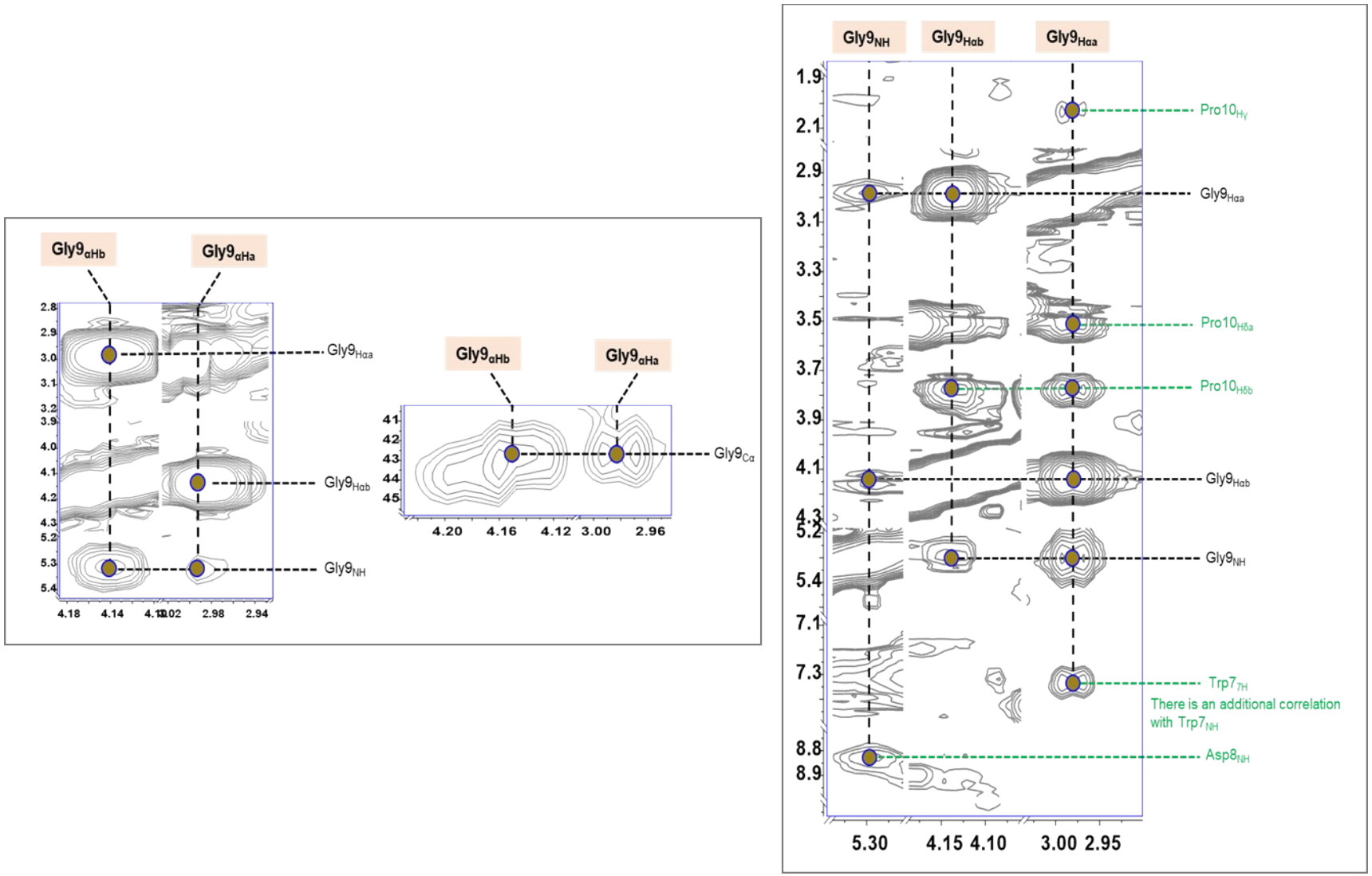


Figure S34. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Gly9**

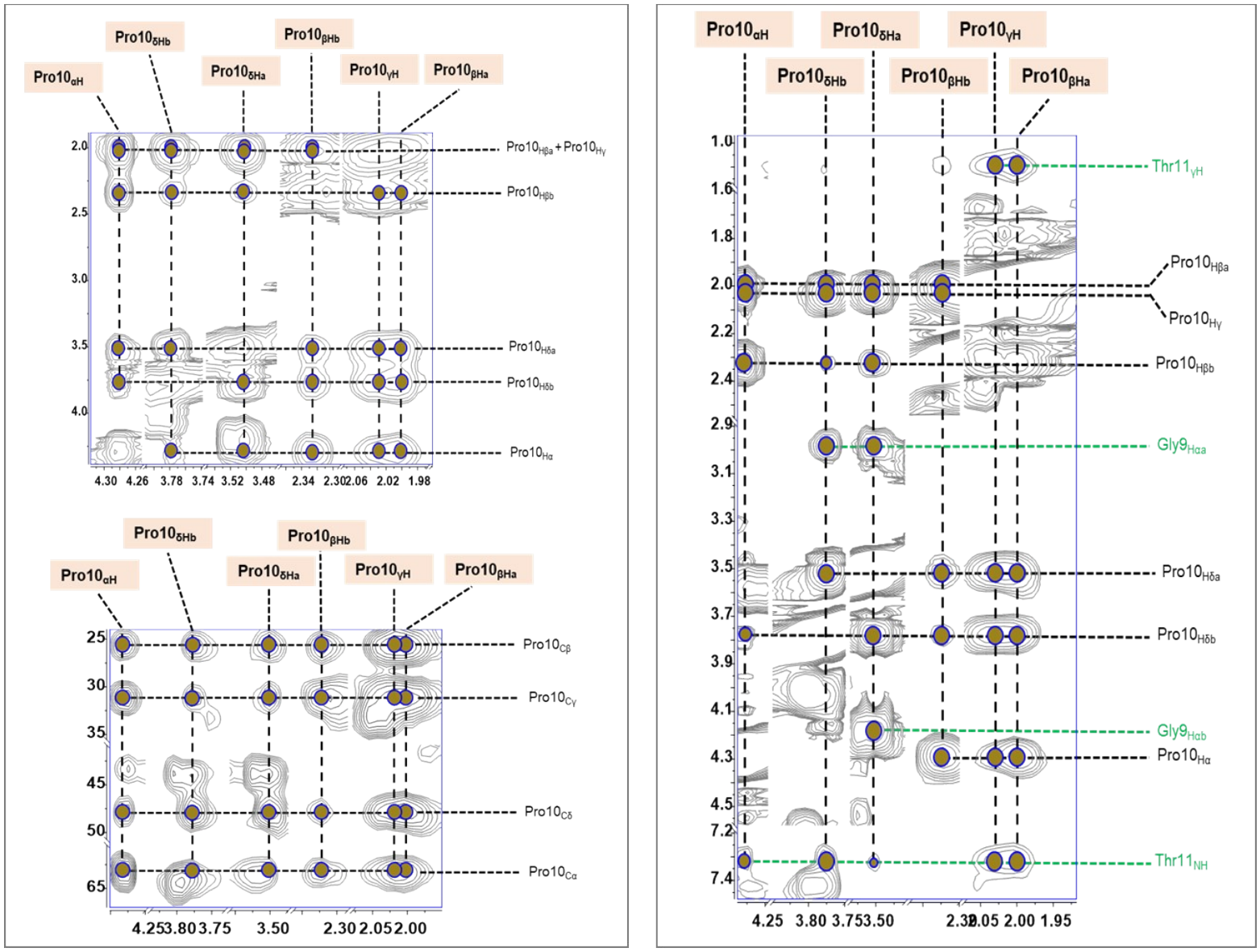


Figure S35. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Pro10**

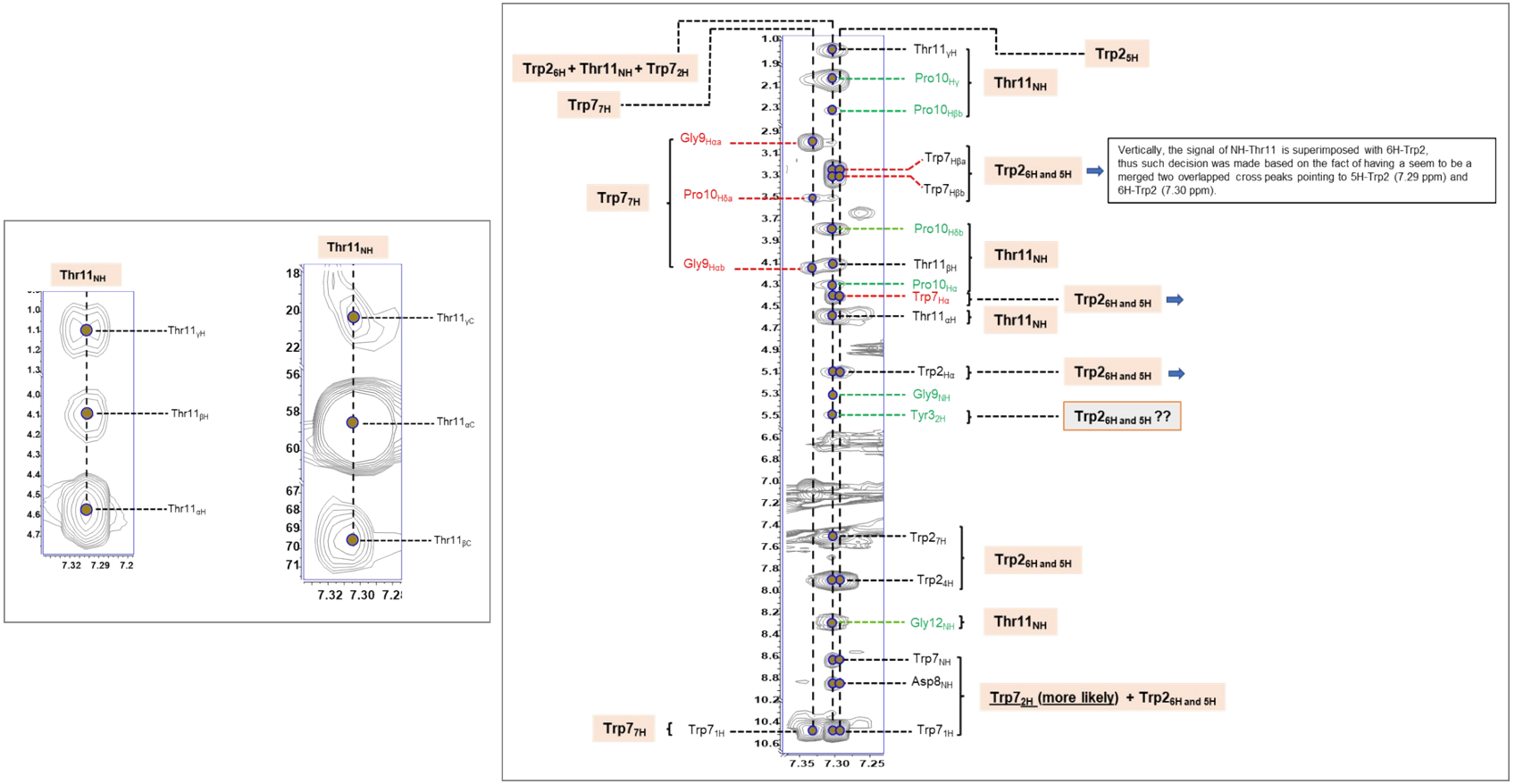


Figure S36. Annotated ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (left), and ^{1}H-^{1}H NOESY (right) spectra of nocapeptin A (**1**) depicting residue **Thr11**

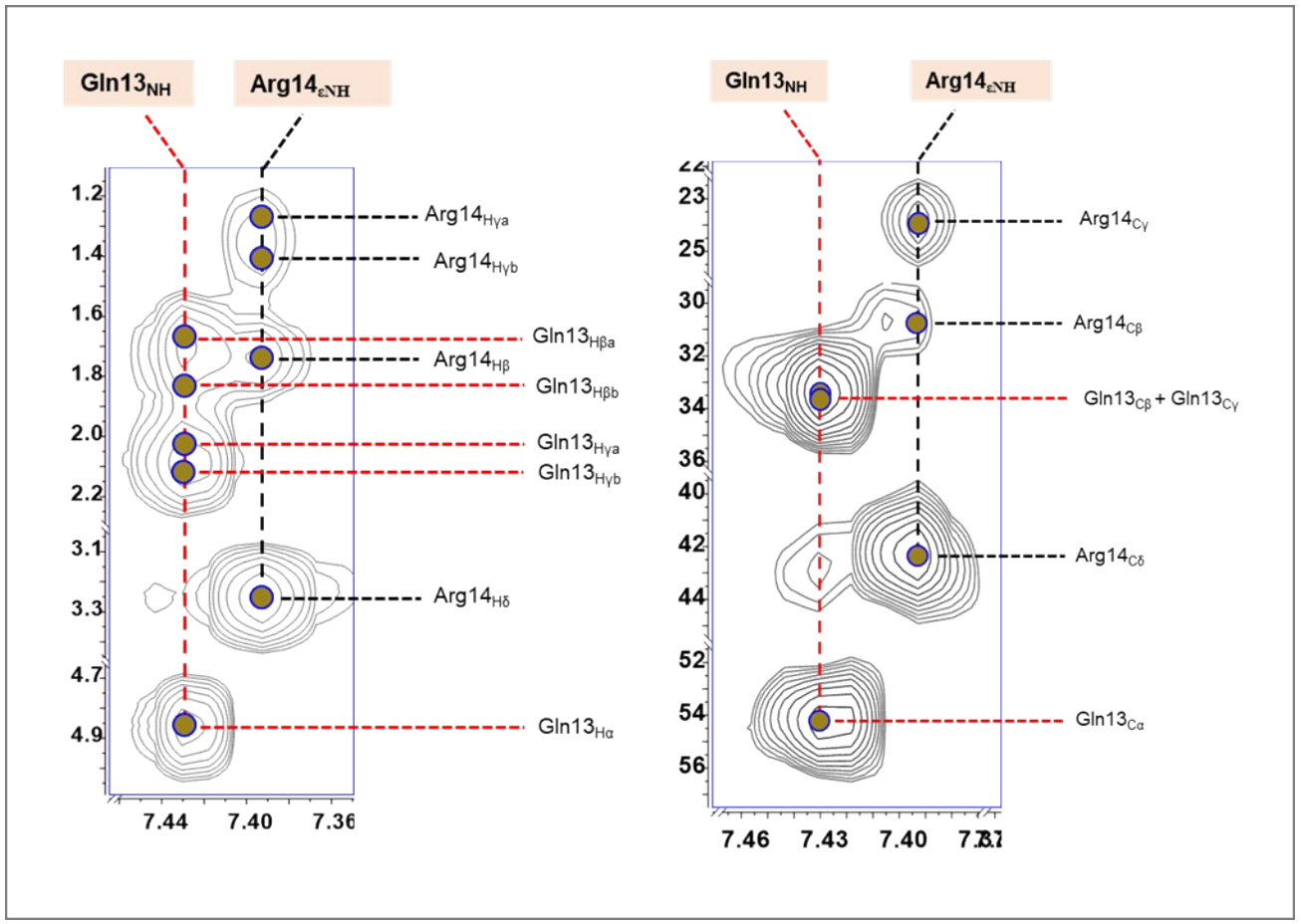


Figure S37. Annotated ^{1}H-^{1}H TOCSY (left), and ^{1}H-^{13}C HSQC-TOCSY (right) spectra of nocapeptin A (1) depicting residue **Gln13**

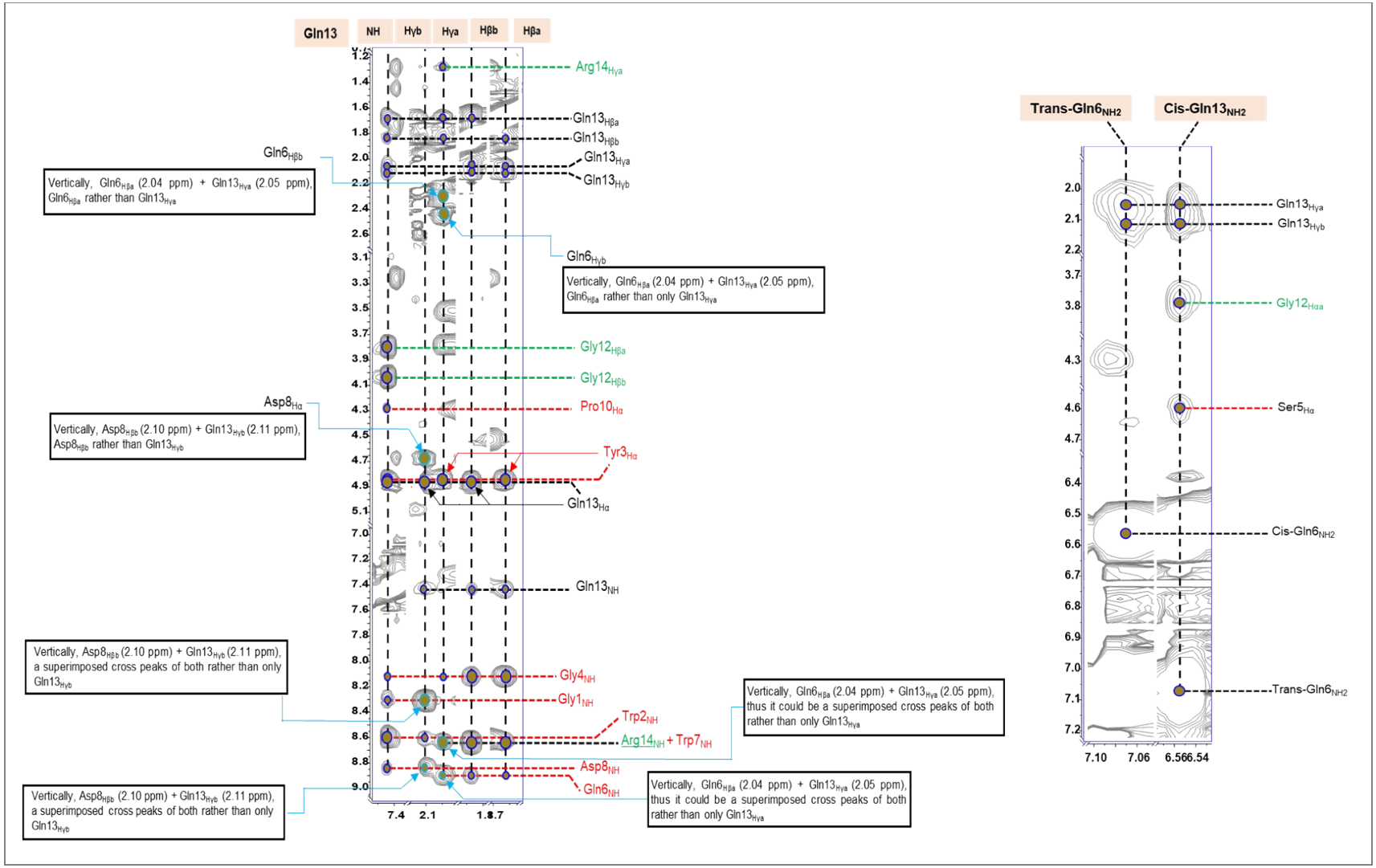


Figure S38. Annotated ^{1}H-^{1}H NOESY spectrum of nocapeptin A (**1**) depicting residue **Gln13**

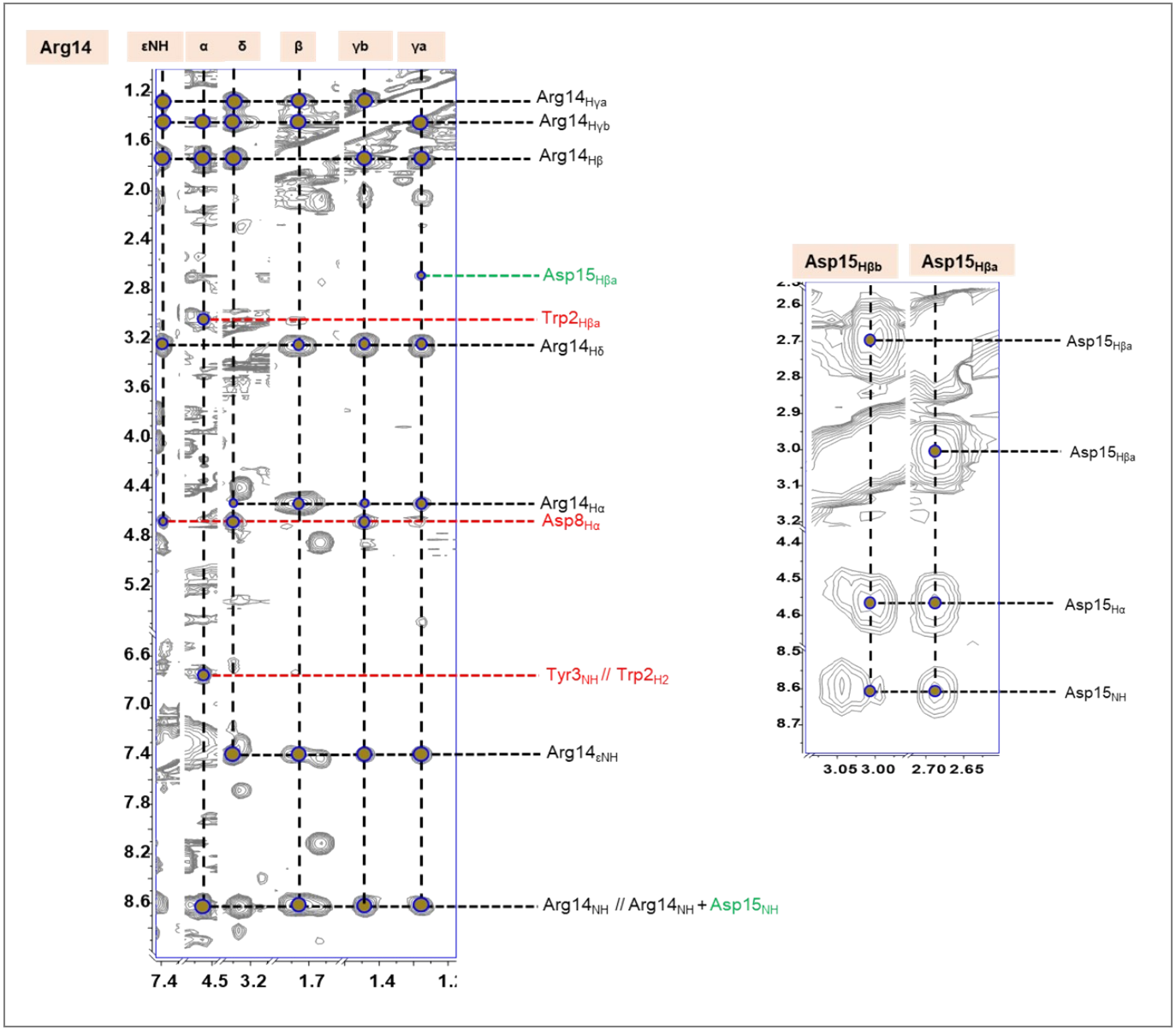


Figure S39. Annotated ^{1}H-^{1}H NOESY spectra of nocapeptin A (**1**) depicting residues **Arg14** and **Asp15**

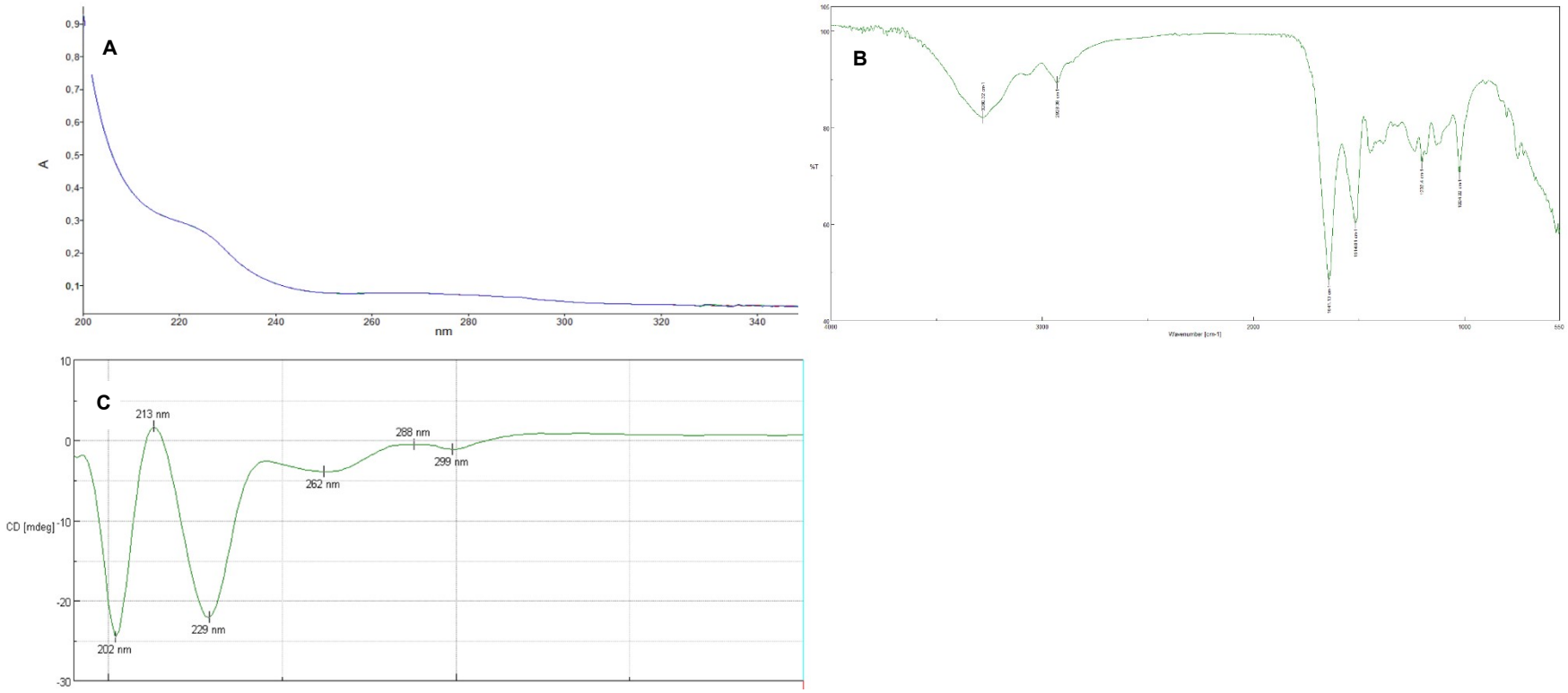


Figure S40. UV (**A**), FT-IR (**B**) and CD (**C**) spectra of nocapeptin A (**1**)

White amorphous powder

$[\alpha]_D^{24}$ -15.6 (*c* 0.64, MeOH)

UV (MeOH) λ_{max} (log ε) 220 sh (4.27), 272 (3.57) nm: see Figure S40A

FT-IR (ATR) ν_{max} (cm^{-1}): 3280, 2928, 1641, 1515, 1202, 1024: see Figure S40B

CD (conc. 0.128 mg/mL in MeOH) cotton effects (mdeg): 213 nm (+1.67), 202 nm (-24.52), 229 nm (-22.07), 262 nm (-3.96) and 299 nm (-1.16)

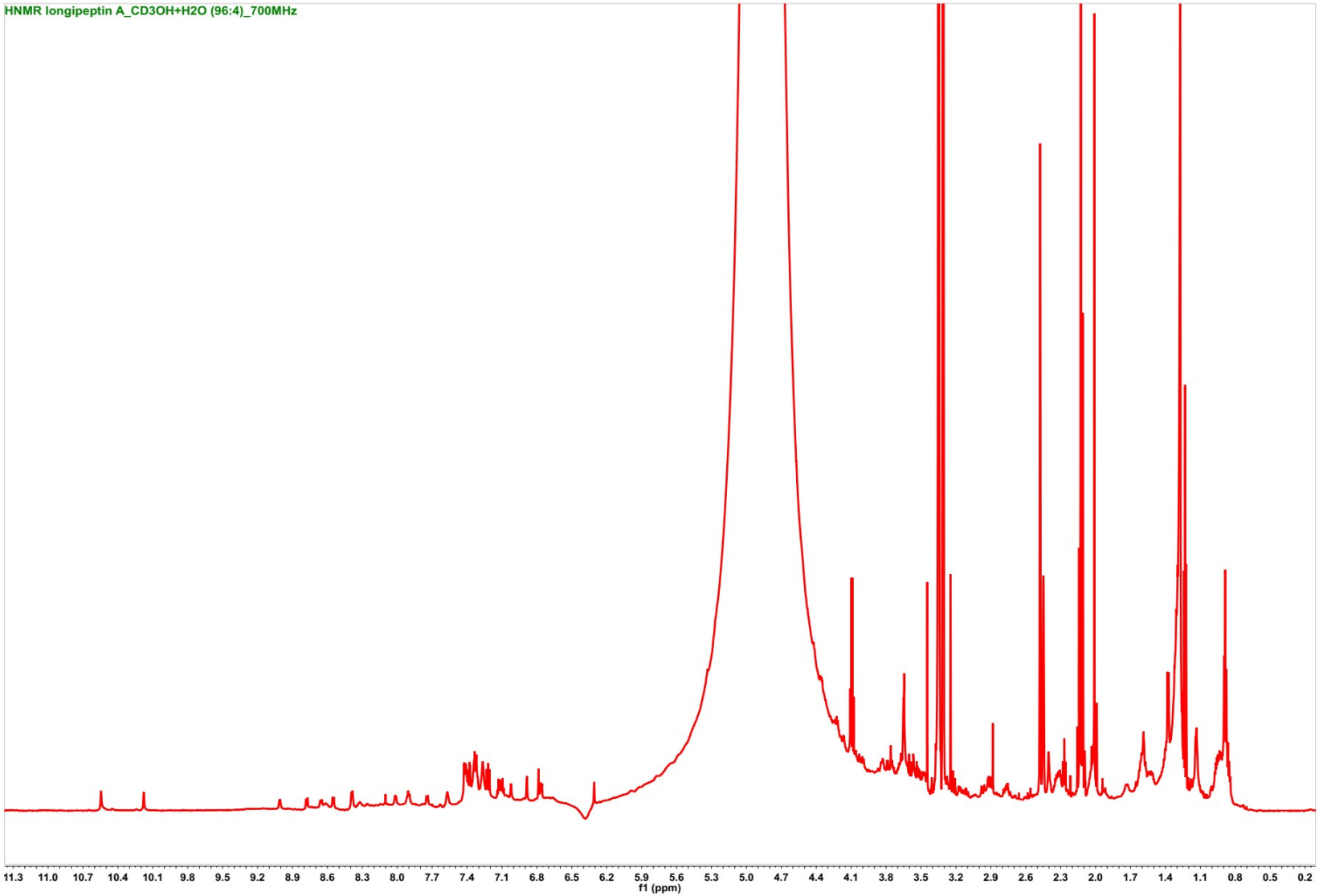


Figure S41. 700 MHz ^{1}H- NMR spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

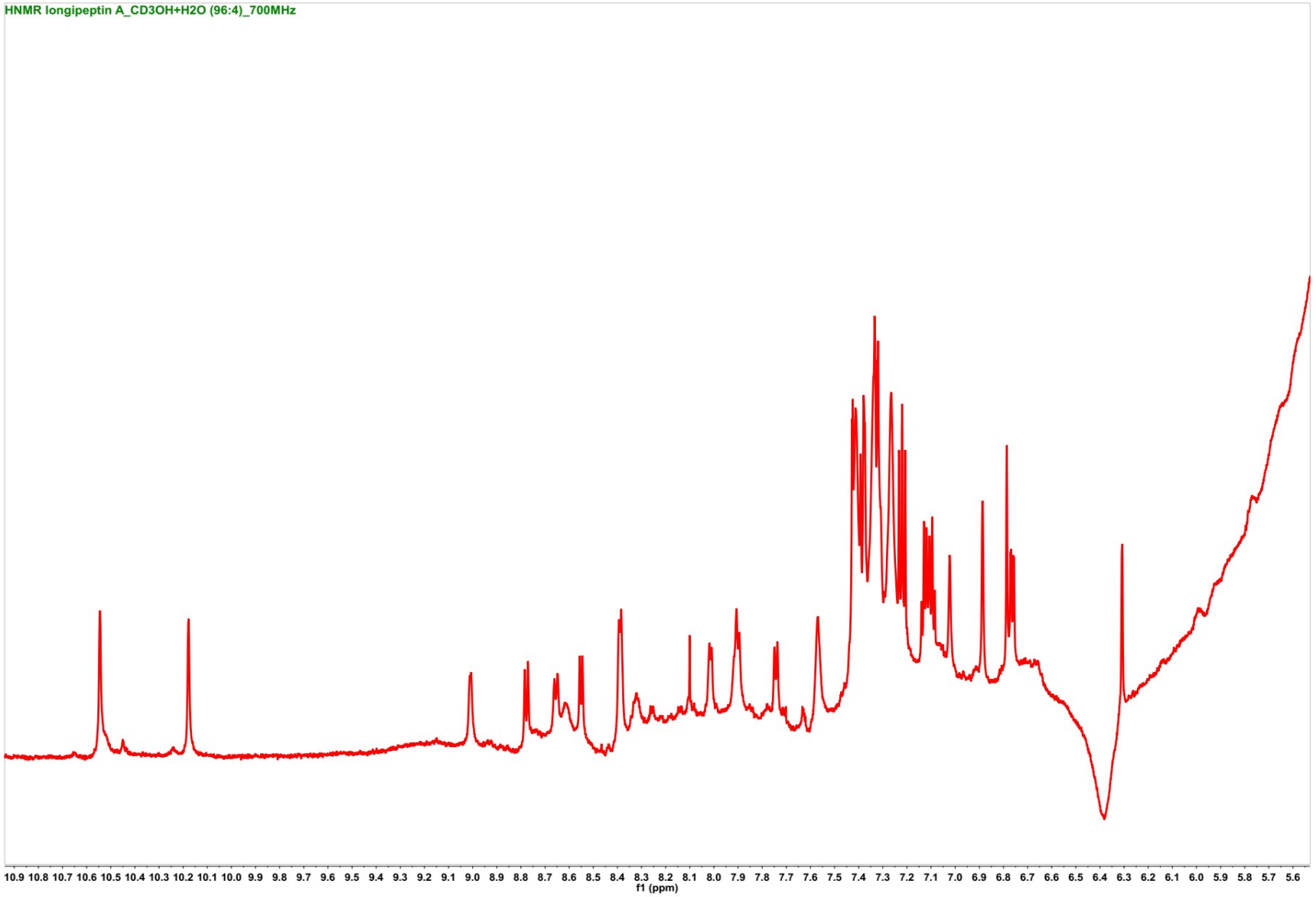


Figure S41A. Expanded 700 MHz ^{1}H- NMR spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4), detailing the aromatic region

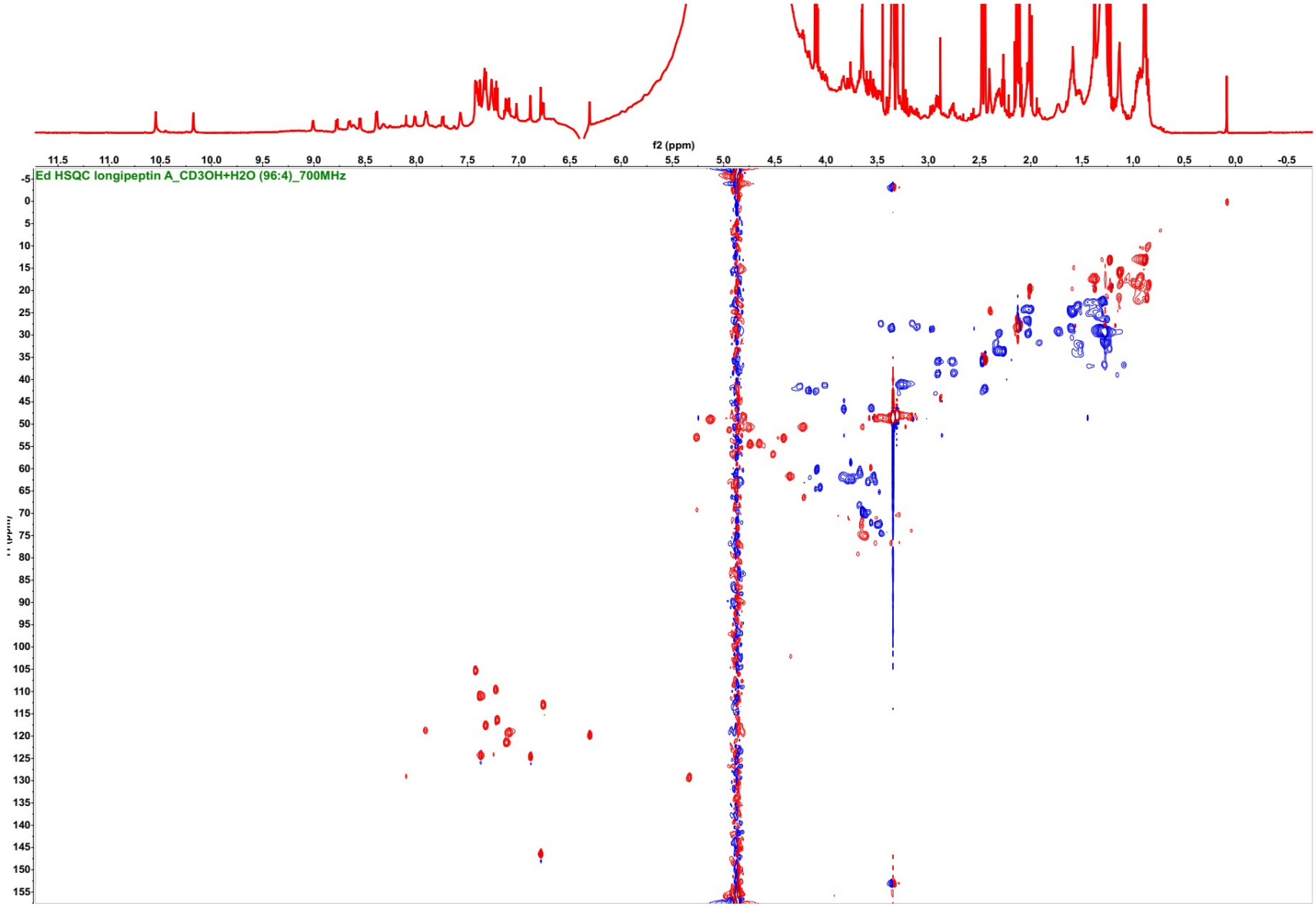


Figure S42. 700 MHz ^{1}H-^{13}C edited HSQC spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

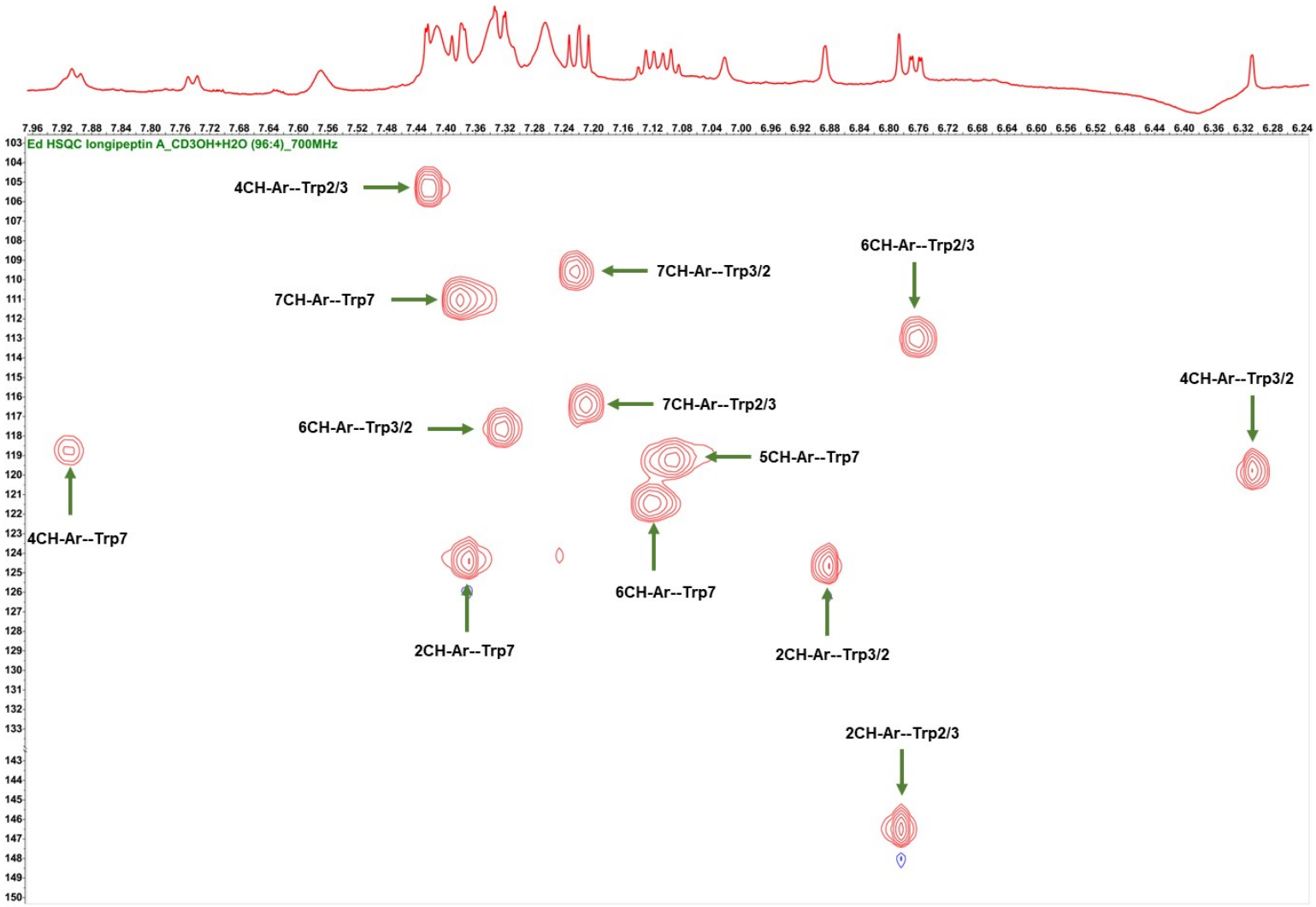


Figure S42A. Annotated ^{1}H-^{13}C edited HSQC spectrum of the assembled indolic systems (**Trp2/3**, **Trp3/2** and **Trp7**) of longipeptin A (**3**)

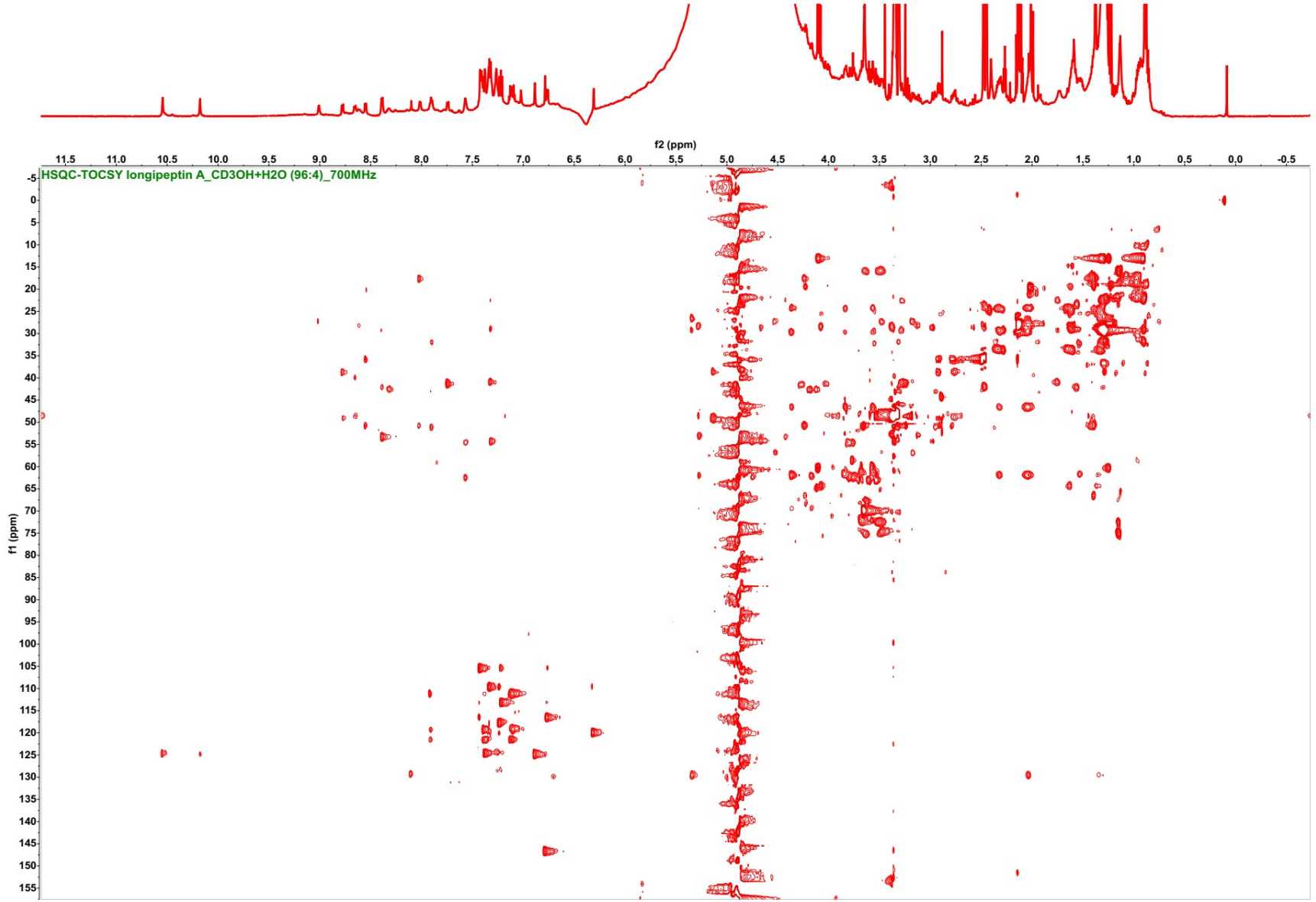


Figure S43. 700 MHz ^{1}H-^{13}C HSQC-TOCSY spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

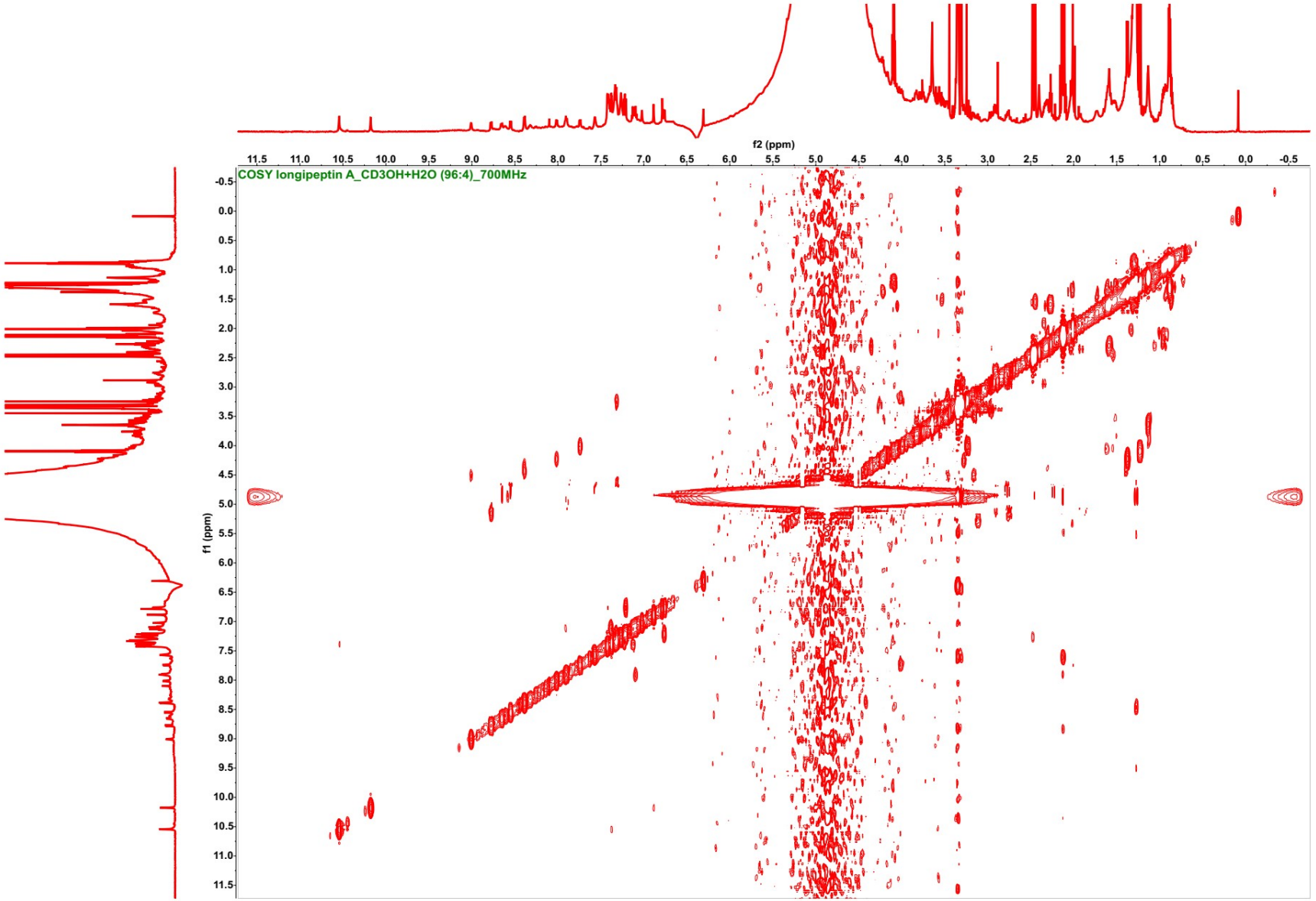


Figure S44. 700 MHz ^{1}H-^{1}H COSY spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

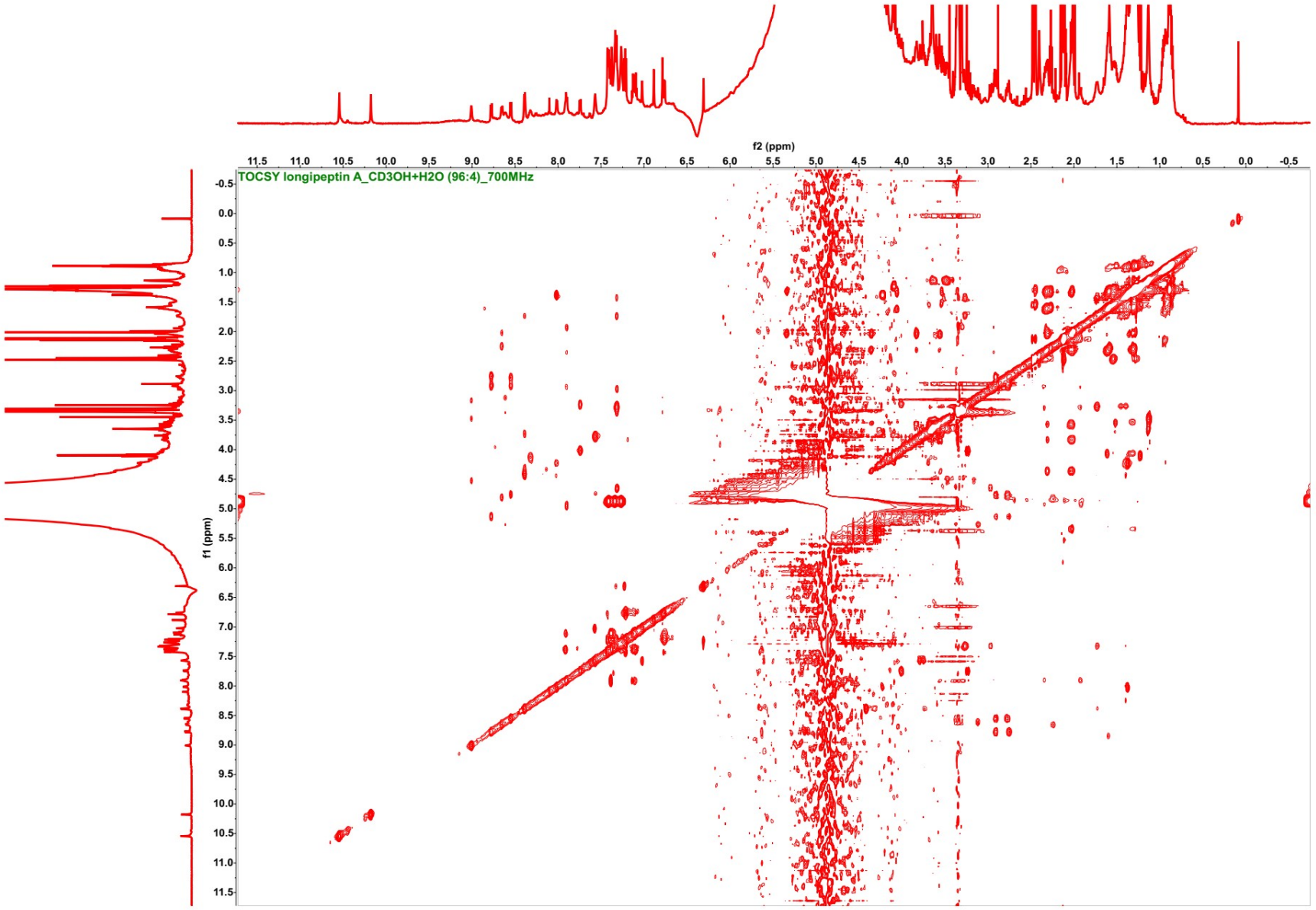


Figure S45. 700 MHz ^{1}H-^{1}H TOCSY spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

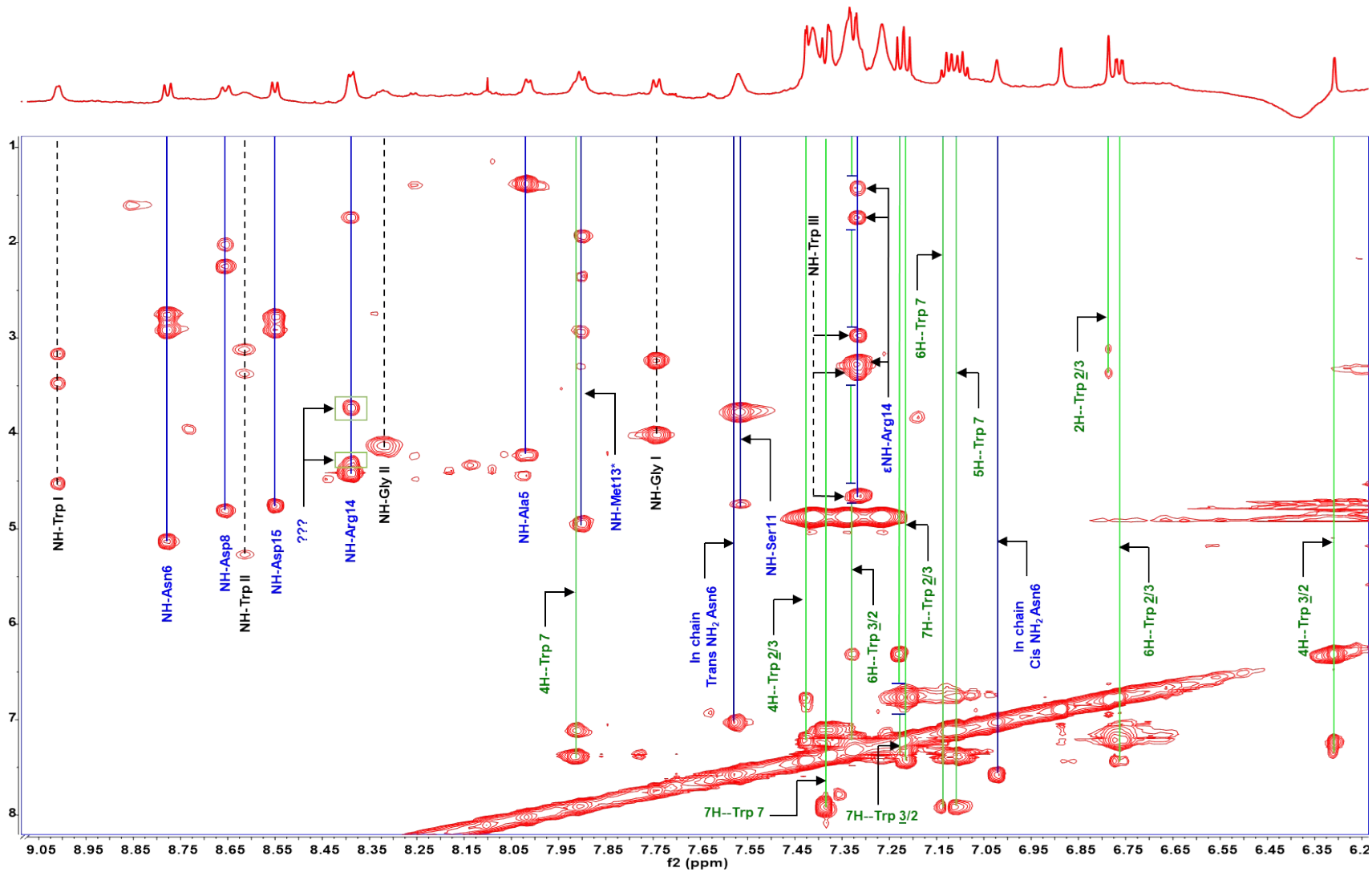


Figure S45A. Annotated ^{1}H-^{1}H TOCSY spectrum, highlighting the assembled spin systems of longipeptin A (**3**)

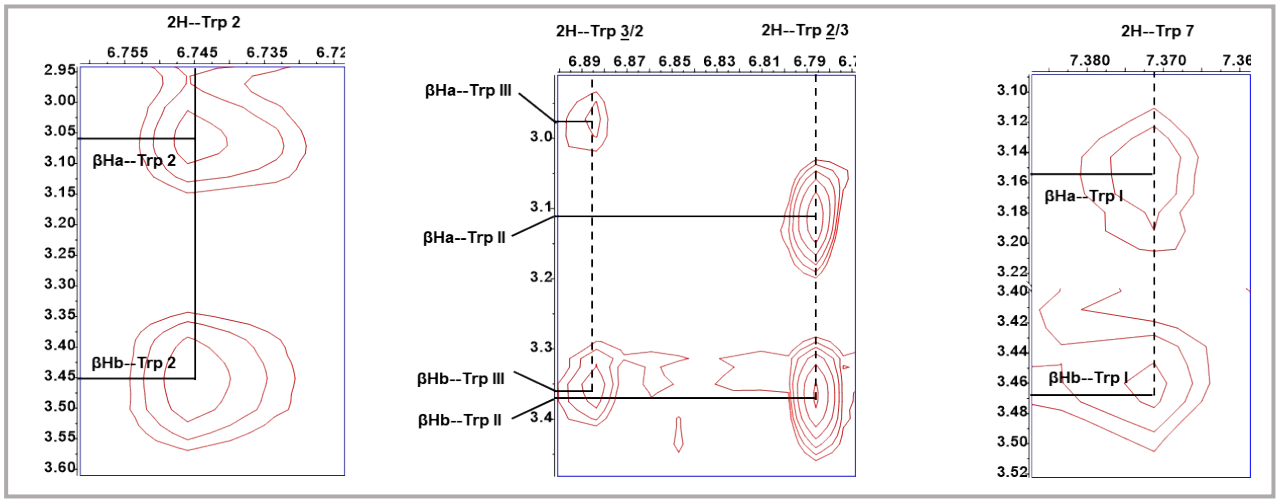


Figure S45B. Annotated ^{1}H-^{1}H TOCSY spectra, highlighting the observation of an allylic ^{4}J-coupling between the 2H and βHa+b spin systems of the Trp residues in nocapeptin A (**1**) (left) vs longipeptin A (**3**) (middle and right)

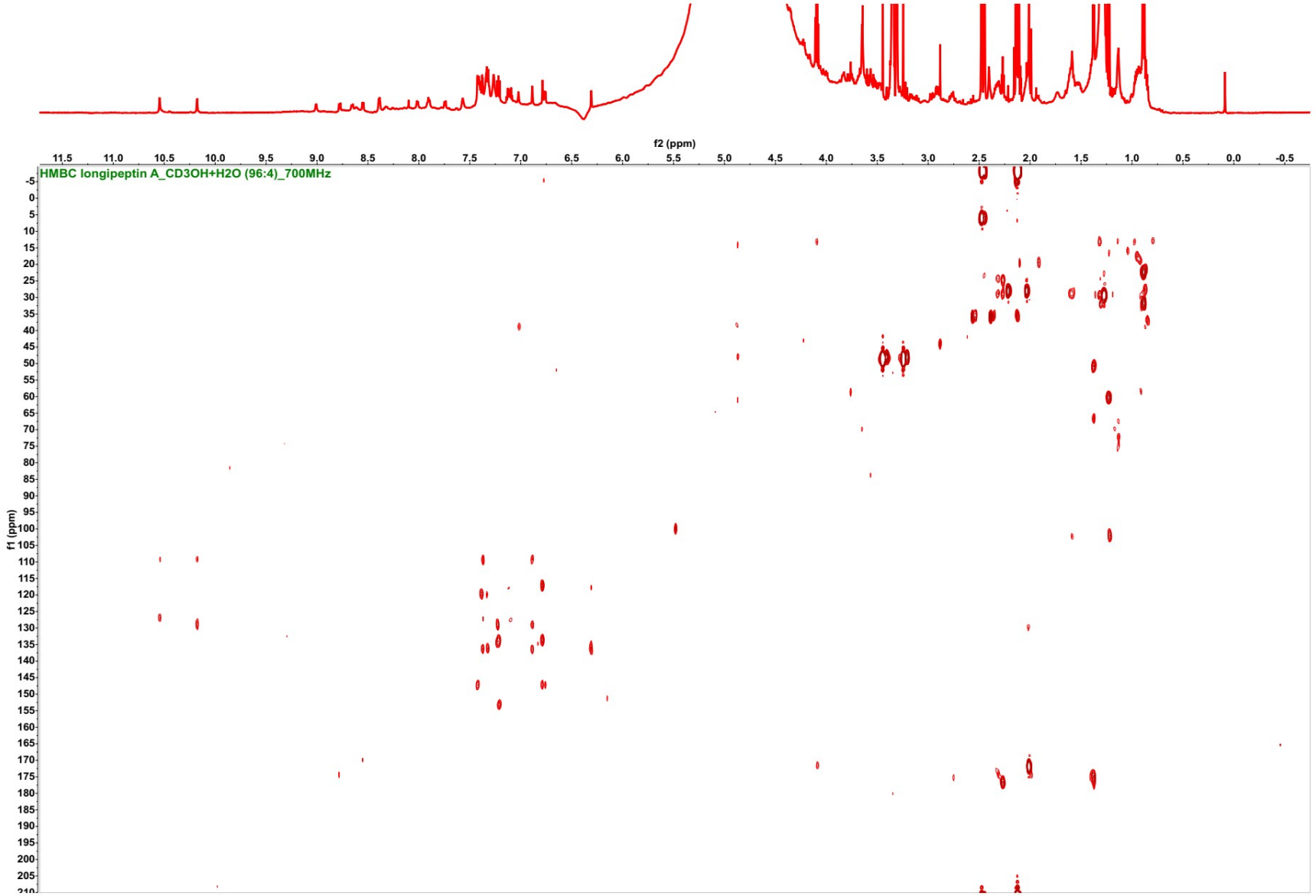


Figure S46. 700 MHz ^{1}H-^{13}C HMBC spectrum of longipeptin A (**3**) in d_3-CH_3OH/H_2O (96:4)

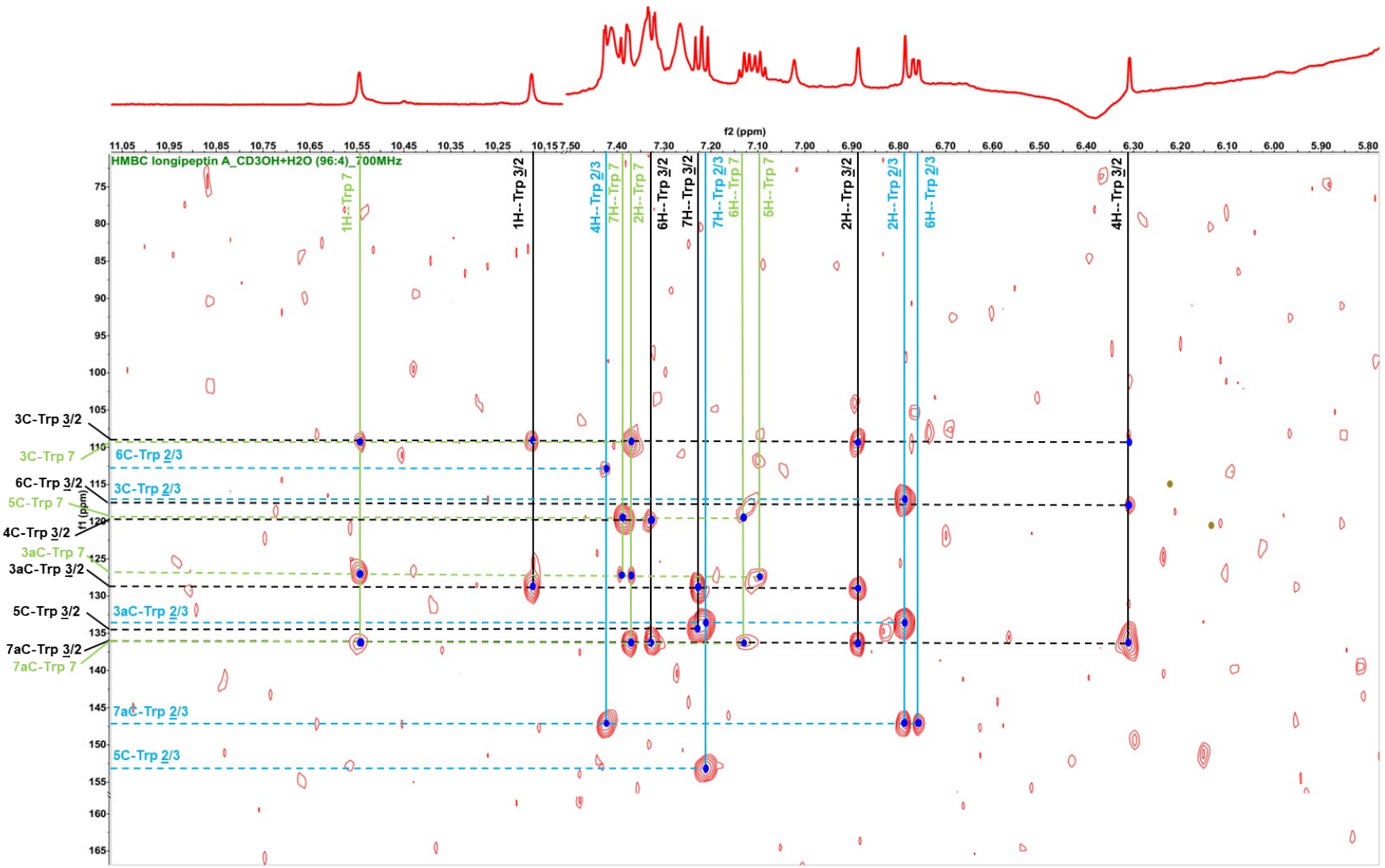


Figure S46A. Annotated ^{1}H-^{13}C HMBC spectrum highlighting the assembled indolic systems (**Trp2/3**, **Trp3/2** and **Trp7)** of longipeptin A (**3**)

Table S8. ^{1}H, and ^{13}C-NMR data of longipeptin A (**3**) [d_3-CH_3OH/H_2O (96:4); 700/176 MHz]

Residue	Position	δ_C	δ_H, mult (*J* in Hz)	Residue	Position	δ_C	δ_H, mult (*J* in Hz)
Gly1	CO	n.d.	--------	**Trp7**	CO	n.d.	--------
	α	*See Figure S47*	*See Figure S47*		α	56.78, CH	4.52
	-NH-	--------	*See Figure S47*		β	27.44, CH_2	3.16 + 3.47
Trp2/3*	CO	n.d.	--------		1 (NH)	--------	10.55 s
	α	52.97, CH	5.27		2	124.40, CH	7.37
	β	28.41, CH_2	3.12 + 3.37		3	109.44, C	--------
	1(N)	--------	--------		3a	127.26, C	--------
	2	146.46, CH	6.79 br s		4	118.74, CH	7.91
	3	117.16, C	--------		5	119.22, CH	7.09 dd (7.6, 7.2)
	3a	133.60, C	--------		6	121.46, CH	7.12 dd (7.6, 7.2)
	4	105.33, CH	7.42 d (2.2)		7	111.07, CH	7.38
	5	153.20, C	--------		7a	136.25/136.34, C	--------
	OH	--------	n.d.		-NH-	--------	9.01
	6	113.03, CH	6.76 dd (8.7, 2.2)	**Asp8**	CO	n.d.	--------
	7	116.41, CH	7.21 d (8.7)		α	48.29, CH	4.81
	7a	147.31, C	--------		β	39.91, CH_2	2.02 + 2.25
	-NH-	--------	8.61		γ	n.d.	--------
Trp3/2*	CO	n.d.	--------		-NH-	--------	8.65 d (8.6)
	α	54.33, CH	4.65	**Gly9**	CO	n.d.	--------
	β	28.69, CH_2	2.97 + 3.36		α	*See Figure S47*	*See Figure S47*
	1 (NH)	--------	10.18 s		-NH-	--------	*See Figure S47*
	2	124.70, CH	6.89 br s	**Pro10**	CO	n.d.	--------
	3	109.25, C	--------		α	61.76, CH	4.36
	3a	128.95, C	--------		β	29.63, CH_2	2.01 + 2.31
	4	119.80, CH	6.31 br s		γ	24.24, CH_2	2.02
	5	134.45, C	--------		δ	46.44, CH_2	3.56 + 3.83
	6	117.66, CH	7.32 dd (8.4, 1.8)	**Ser11**	CO	n.d.	--------
	7	109.59, CH	7.23 dd (8.4)		α	54.41, CH	4.74
	7a	136.25/136.34, C	--------		β	62.38, CH_2	3.75 + 3.79
	-NH-	--------	7.31		OH	--------	n.d.
Gly4	CO	n.d.	--------		-NH-	--------	7.57
	α	*See Figure S47*	*See Figure S47*	**Gly12**	CO	n.d.	--------
	-NH-	--------	*See Figure S47*		α	*See Figure S47*	*See Figure S47*
Ala5	CO	n.d.	--------		-NH-	--------	*See Figure S47*
	α	50.70, CH	4.22	**Met13****	CO	n.d.	--------
	β	17.57, CH_2	1.38 d (7.2)		α	51.31, CH	4.95
	-NH-	--------	8.01 d (7.0)		β	31.78, CH_2	1.93 + 2.35
Asn6	CO	n.d.	--------		γ	43.13, CH_2	2.93 + 3.29
	α	48.96, CH	5.13		S-Me_1+ Me_2	n.d.	n.d.
	β	38.70, CH_2	2.75 + 2.91		-NH-	--------	7.90 d (8.3)
	γ	n.d.	--------				
	-NH_2	--------	7.02 br s + 7.57 br s				
	-NH-	--------	8.78 d (9.4)				

Arg14	CO	n.d.	--------
	α	53.17, CH	4.41
	β	29.25, CH_2	1.73
	γ	22.70, CH_2	1.43
	δ	41.09, CH_2	3.27
	εNH	--------	7.32
	ζ	n.d.	--------
	NHη1	--------	n.d.
	NH_2η1	--------	n.d.
	-NH-	--------	8.39 d (6.7)
Asp15	CO_2H	n.d.	--------
	α	50.79, CH	4.76
	β	35.97, CH_2	2.77 + 2.91
	γ	n.d.	--------
	-NH-	--------	8.55 d (7.7)

* Residues involved in the C-N biaryl crosslink in an alternative manner.
** Residues involved in the *S*-methylation.
- CH_3, CH_2, and CH chemical shift values were extracted indirectly from ^{1}H-^{13}C HSQC.
- C chemical shift values were extracted indirectly from ^{1}H-^{13}C HMBC.

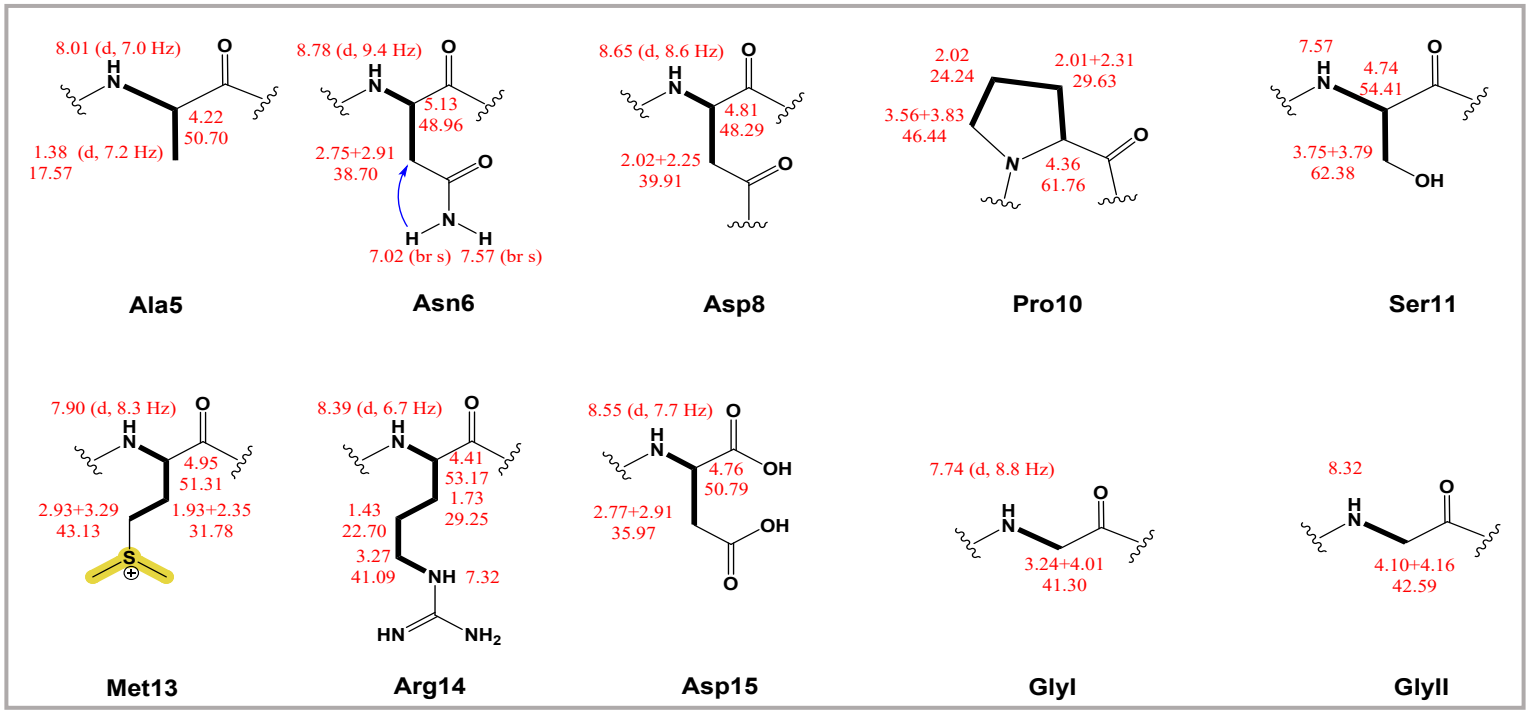


Figure S47. Schematic representation of the assembled spin systems of longipeptin A (**3**) using ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY (bold lines) and ^{1}H-^{13}C HMBC (blue arrows) correlations

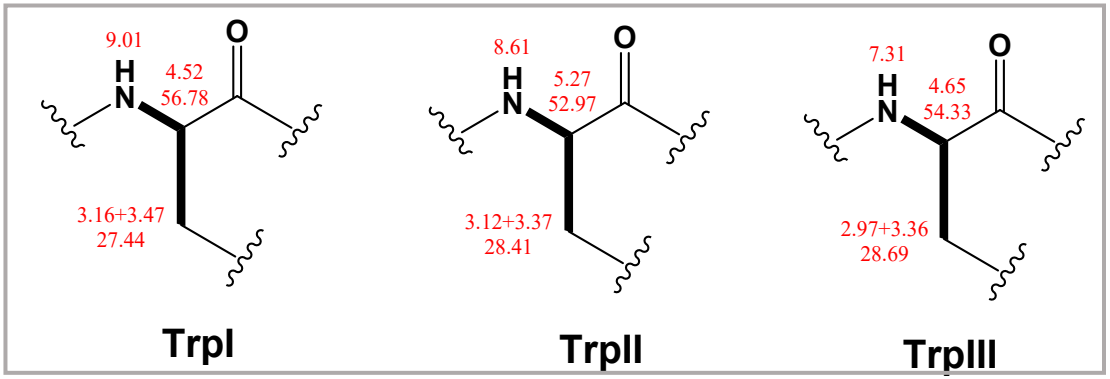


Figure S47A. Schematic representation of the assembled spin systems constituting the Trp residues of longipeptin A (**3**) using ^{1}H-^{1}H COSY and ^{1}H-^{1}H TOCSY correlations (bold lines)

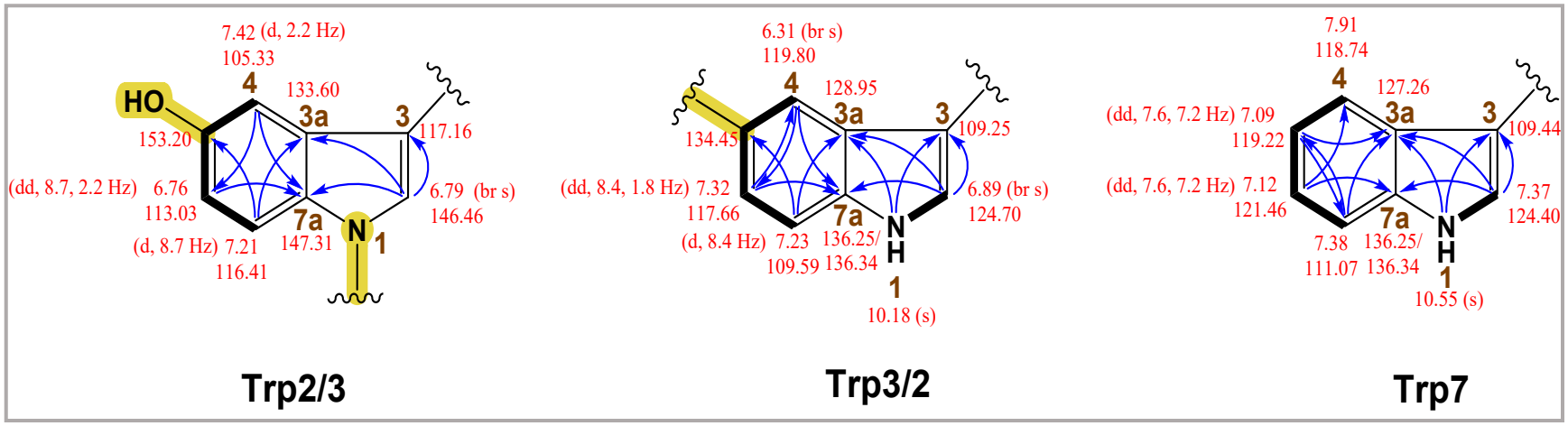


Figure S47B. Schematic representation of the assembled spin systems constituting the Trp residues of longipeptin A (**3**) using ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY (bold lines) and ^{1}H-^{13}C HMBC (blue arrows) correlations.

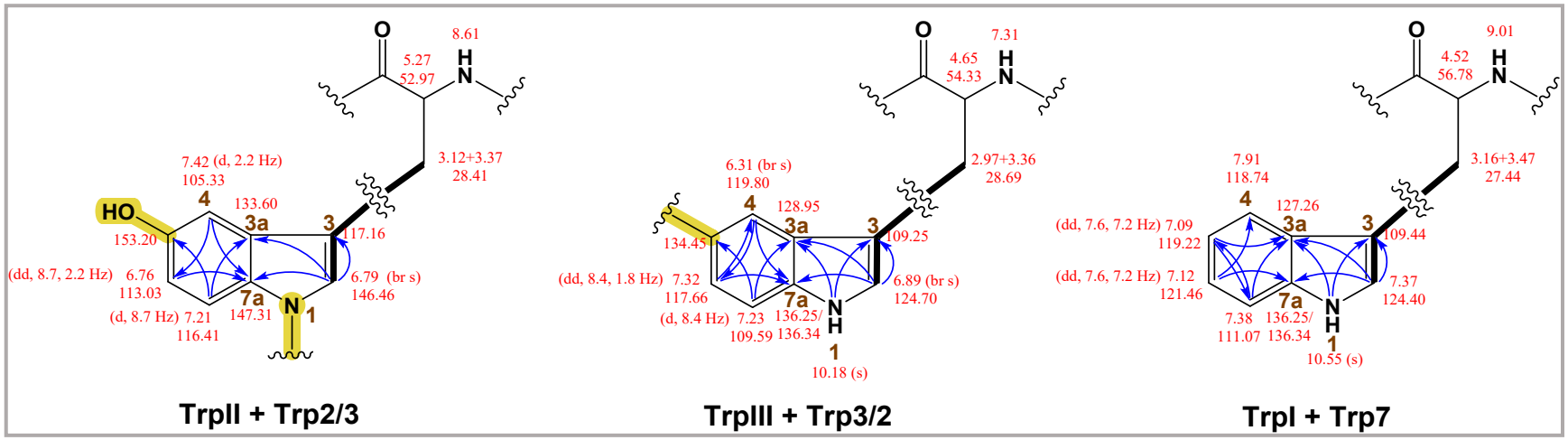


Figure S47C. Schematic representation of the connected structural units, **Trp I--Trp7**, **Trp II--Trp2/3** and **Trp III--Trp3/2** constituting the W residues of longipeptin A (**3**) based on the characteristic ^{1}H-^{1}H TOCSY allylic ^{4}J-coupling (bold lines)[26] and ^{1}H-^{13}C HMBC (blue arrows) correlations.

Partial Structure Elucidation of longipeptin A (3)

Considering the exhaustive analysis of the MS^1 and MS^2 longipeptin series (A-C, **3**-**5**), a reasonable assumption was formulated in terms of their morphed skeletons with the differently appended PTMs. The highly modified major variant, longipeptin A (**3**) was hypothesized to be the final product of the *lop* BGC loaded with three tailoring events (oxidation '+O', methylation '+CH_2' and crosslink '-2H') aligning with the unusual genetic elements featured in the BGC that consists of three ancillary processing enzymes (Figure S1).

The usage of tandem MS enabled not only sequencing **3** but also defining the nature and the positions of the multiple structural modifications. The first tailoring, the methylation event, was initially envisioned to occur at the C-terminus residue, D15 as previously reported;[27] however, such an assumption was readily ruled out after studying the MS^2 fragments. The observation of the B ion series (**b_3-b_{12}**) and the surprising inability to spot the typical intense Y fragments, normally arising from the tail, indicated an unusual fragmentation behavior upon the methylation reaction. Such an observation was further supported by the spectral similarity with variant B (**4**) which shares the same PTM of interest (methylation) (Figure S8). An orthogonal piece of evidence was also extracted from the extensive annotation of the highly informative MS^2 spectrum of the biosynthetic intermediate longipeptin C (**5**) that lacked such a structural modification (Figure S11).

As a result, and having in hand a dereplicated set of B ions from the MS^2 spectrum of **3**, several possible methylation centers were postulated, principally at the side chain heteroatoms of the tail (O => **Ser11**, N => **Arg14**, S => **Met13** or N amidic backbone). Interestingly, the S-methylation at the **Met13** residue was found to be the most convincing hypothesis which was validated by the perfect alignment with the newly observed series of Y ions, arising upon the neutral loss of a characteristic 62 Da in the form of dimethylsulfide (Me_2S) (Figures S9-9A). Similarly, the MS^2 fragments of **4**, a methylated shunt product, were also found to fit with such a structural suggestion confirming this rare C-S bond formation (Figures S10-10A).

The remaining PTMs of **3**, oxidation and crosslink, were also found to be introduced at the **G1-W2-W3** sequence, based on the MS/MS annotation in a similar pattern to nocapeptins (Figures S4A-4C). To corroborate the MS-deduced structural findings of **3**, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC spectra were utilized to construct the majority of the constituting residues of longipeptin A (**3**) scaffold (Figures S47-47B). Although the predicted CP and MS analysis highlighted the presence of 4x Gly residues, we were only limited to assembling a pair of them from the homonuclear 2D-NMR (Figure S45A). Interestingly, the tandem usage of ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC-TOCSY and ^{1}H-^{13}C HSQC supported indirectly the *S*-methylation event of M13 residue via the downfield chemical shift of its γCH_2 to be resonating at around 43.20 ppm in contrast to the typical values, 29-31 ppm (Figures S42 and S47).

In addition, three candidate spin systems, **TrpI**, **TrpII** and **TrpIII** were elucidated in concert with three indolic substructures (**Trp2/3**, **Trp3/2** and **Trp7**), mainly with the aid of ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HSQC-TOCSY and ^{1}H-^{13}C HMBC spectra, to define the three W residues featured in **3**. Despite the complete absence of a suitable ^{1}H-^{1}H NOESY and the inability to uncover any ^{1}H-^{13}C HMBC correlations to connect these structural fragments together, weaker couplings (allylic $^4J_{2H,\beta Ha+b}$ coupling) from ^{1}H-^{1}H TOCSY suggested the following possible connectivities: **TrpI--Trp7, TrpII--Trp2/3** and **TrpIII--Trp3/2**. Analogously, the observation of having allylic $^4J_{2H,\beta H}$ couplings in W residues was similarly witnessed in the ^{1}H-^{1}H TOCSY spectrum of nocapeptin A (**1**) (Figures S45B and S47C).

As expected, the ^{1}H-NMR of **3** displayed two downfield singlets, δ_H = 10.18, 10.55, indicative for only two indolic NHs of three Trp moieties, proposing a possible substitution at one indolic NH of these three W residues (Figure S41A). Tracking such downfield singlets through ^{1}H-^{1}H TOCSY, ^{1}H-^{13}C HSQC and ^{1}H-^{13}C HMBC enabled the structural elucidation of three (un)substituted aromatic systems, **Trp2/3**, **Trp3/2** and **Trp7** (Figure S46A). **Trp7** as an unsubstituted indolic fragment was readily assigned to be the aromatic part of W7 considering the CP sequence and the formerly annotated MS fragments. However, the monosubstituted (**Trp3/2**) and the disubstituted (**Trp2/3**) systems were envisioned to represent W2 and W3 residues, alternatively. In addition, the hydroxylation event was figured out to be introduced into **Trp2/3** substructure at position 5 as suggested by the ^{1}H-^{13}C HMBC correlations.

Although the structural finalization of the biaryl fragment of **3** was crippled due to the absence of a proper ^{1}H-^{1}H NOESY spectrum, the C-N crosslink was indirectly NMR deduced using the comparable shifts of nocapeptin A (**1**). The first evidence was counting on the delineated **TrpII--Trp2/3** as a disubstituted system lacking the indolic proton proposing 1N as an interlinkage atom. This was additionally supported by the characteristic δ_{2C} 146.46 which shares a comparable downfield shift to **1**. The second evidence was derived from **TrpIII--Trp3/2** as a monosubstituted system having a diagnostic δ_{4H} 6.31 with a comparable upfield drift to **1** (Figures S42A and S47C). As a result, two positional structural possibilities were laid out featuring the biaryl (C-N) installation between the structural units **1N-Trp2/3** and **5C-Trp3/2** in an alternating manner. Considering the current genomic and biosynthetic setting, the final product likely favors the substructure in which N1-W2 couples with C5-W3. (Figure S48)

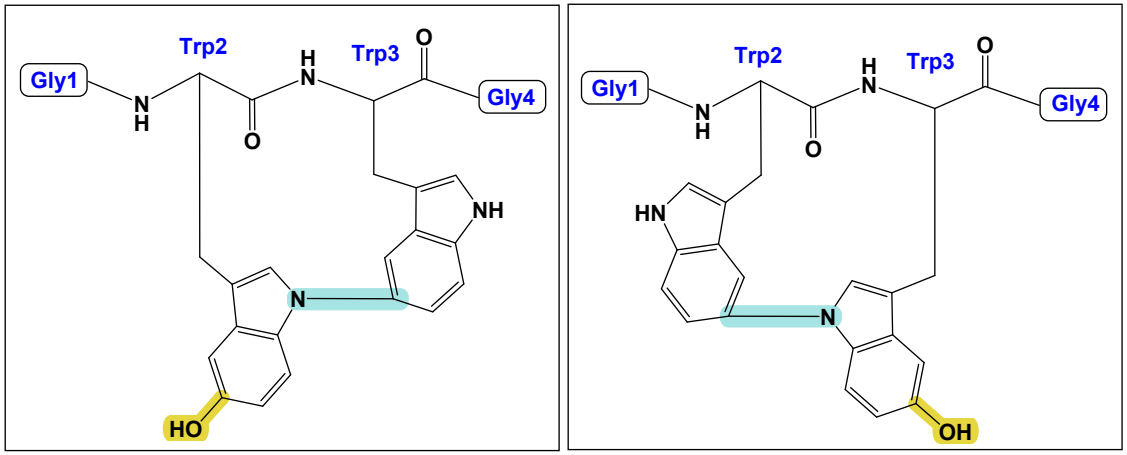


Figure S48. Schematic representation of the two possible biaryl substructures of longipeptin A (**3**) regarding the C-N linkage between **W2** and **W3** residues

Table S9. Results of the antimicrobial assays for **1**. MIC, minimum inhibitory concentration.

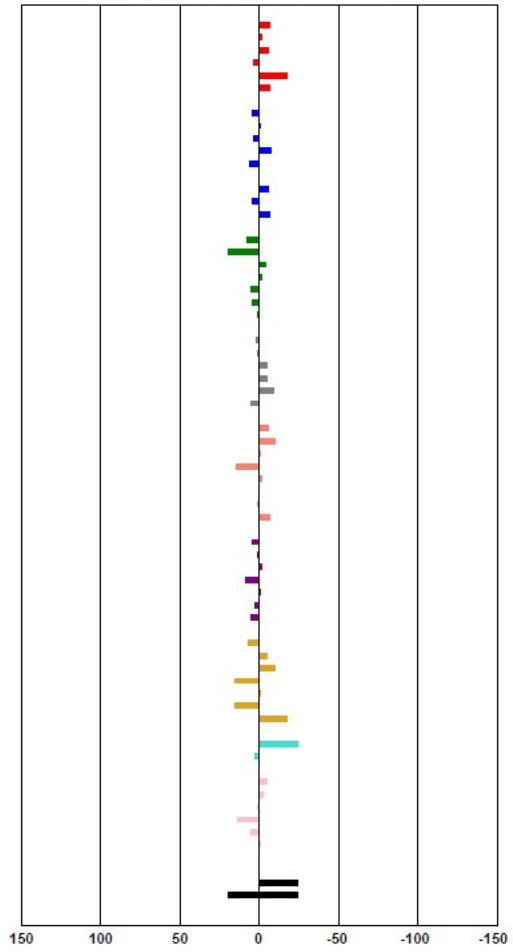
Bacterium	MIC (µg/mL)	Assay medium
Bacillus subtilis 168	>64	MH II broth
Staphylococcus aureus ATCC 29213	>64	
Enterococcus faecium BM 4147-1	>64	
Enterococcus faecalis ATCC 29212	>64	
Escherichia coli ATCC 25922	>64	
Escherichia coli HN 818	>64	
Escherichia coli HN 818 (+ 15 µg/ml PMBN)	>64	
Klebsiella pneumoniae ATCC 12657	>64	
Enterobacter aerogenes ATCC 13048	>64	
Pseudomonas aeruginosa ATCC 27853	>64	
Acinetobacter baumannii 09987	>64	
Micrococcus luteus ATCC 4698	16	
Neisseria gonorrhoeae ATCC 19424	>64	MH II broth
Neisseria gonorrhoeae S 1441	>64	+ 2,5 % FBS
Mycobacterium smegmatis mc^2 155	>64	7H9 broth

Table S10. Results of the cytotoxicity assays for **1**. Developmental Therapeutics Program (DTP)-One dose Mean Graph NCI-60 data.

Developmental Therapeutics Program
One Dose Mean Graph

NSC: D-832357 / 1	Conc: 1.00E-5 Molar	Test Date: Aug 30, 2021
Experiment ID: 2108OS84		Report Date: Oct 08, 2021

Panel/Cell Line	Growth Percent
Leukemia	
CCRF-CEM	94.72
HL-60(TB)	99.75
K-562	94.95
MOLT-4	104.37
RPMI-8226	83.54
SR	94.54
Non-Small Cell Lung Cancer	
A549/ATCC	105.31
EKVX	100.24
HOP-62	104.77
HOP-92	93.96
NCI-H226	106.83
NCI-H322M	95.24
NCI-H460	105.02
NCI-H522	94.88
Colon Cancer	
COLO 205	109.06
HCC-2998	120.16
HCT-116	97.36
HCT-15	99.45
HT29	106.25
KM12	105.18
SW-620	101.85
CNS Cancer	
SF-268	102.60
SF-295	102.24
SF-539	96.01
SNB-19	96.55
SNB-75	92.36
U251	106.43
Melanoma	
MALME-3M	95.49
M14	91.06
MDA-MB-435	100.45
SK-MEL-2	115.59
SK-MEL-28	99.62
SK-MEL-5	101.05
UACC-257	102.36
UACC-62	94.92
Ovarian Cancer	
IGROV1	105.38
OVCAR-3	101.95
OVCAR-4	99.77
OVCAR-5	109.82
OVCAR-8	100.76
NCI/ADR-RES	103.78
SK-OV-3	106.24
Renal Cancer	
786-0	107.70
ACHN	96.19
CAKI-1	91.48
RXF 393	116.04
SN12C	100.51
TK-10	116.36
UO-31	83.27
Prostate Cancer	
PC-3	76.68
DU-145	104.11
Breast Cancer	
MCF7	96.36
MDA-MB-231/ATCC	98.63
HS 578T	102.48
BT-549	114.95
T-47D	106.07
MDA-MB-468	100.30
Mean	100.90
Delta	24.22
Range	43.48



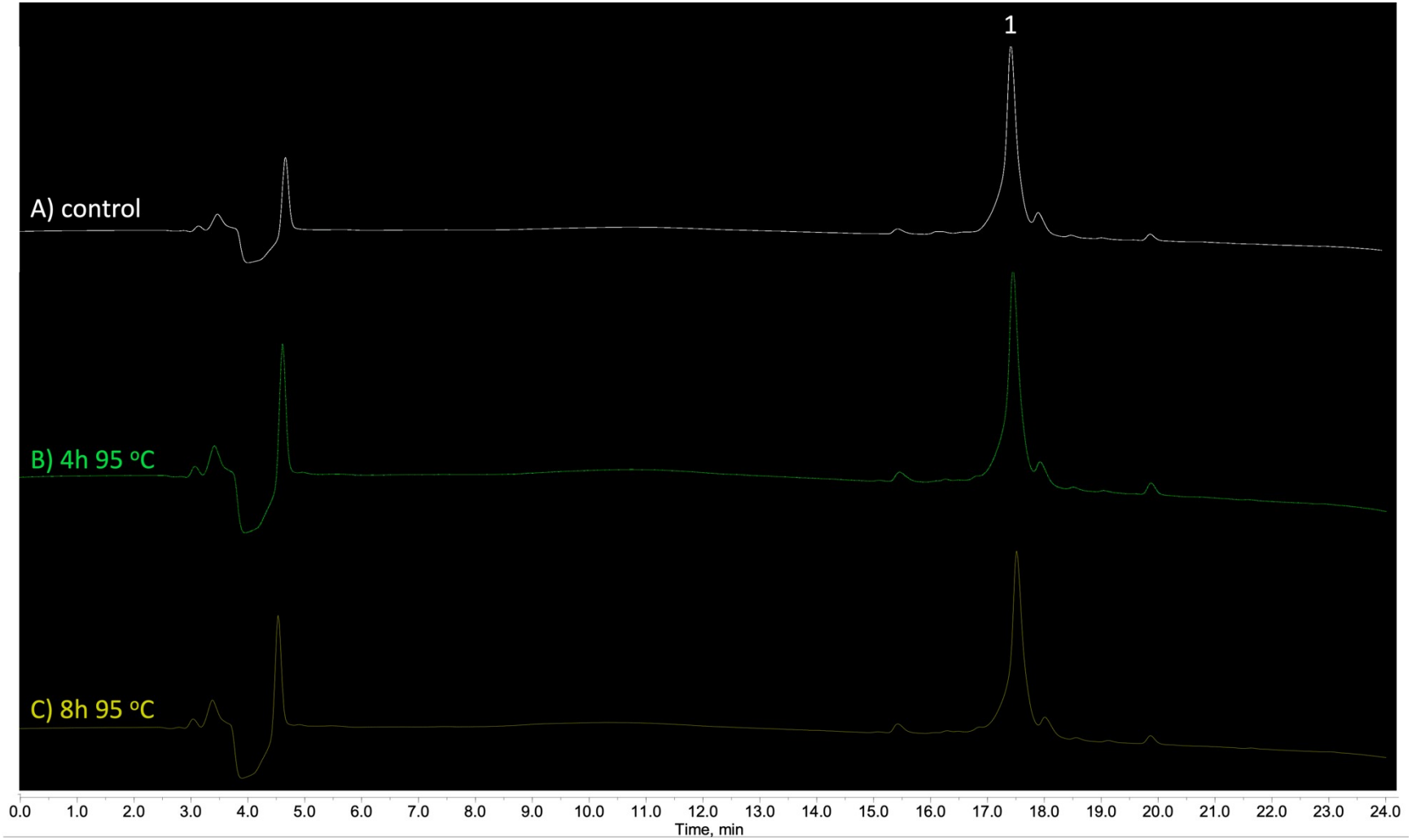


Figure S49. Thermal stability assay with nocapeptin A. Each graph represents the extracted UV chromatogram at 220 nm of the total wavelength chromatogram of an LC-MS run. The peak eluting at t_R = 17.5 min represents nocapeptin A (**1**). A) Analysis of nocapeptin of the untreated nocapeptin A; B) Analysis of nocapeptin A after 4 h of heat treatment; C) Analysis of nocapeptin A after 8 h of heat treatment.

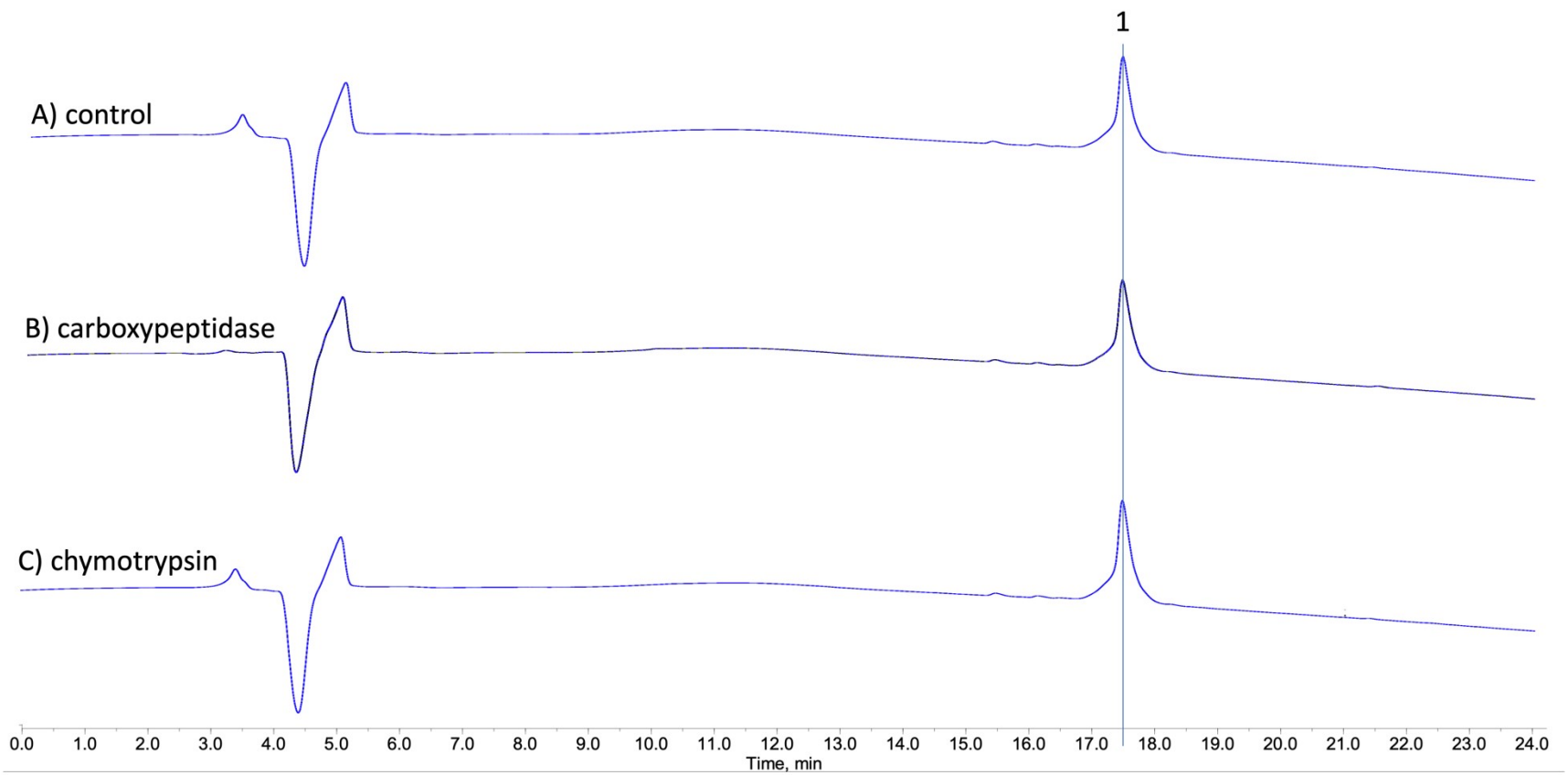


Figure S50. Proteolytic stability assay with nocapeptin A. Each graph represents the extracted UV chromatogram at 220 nm of the total wavelength chromatogram of an LC-MS run. The peak eluting at t_R = 17.5 min represents nocapeptin A (**1**). A) Analysis of the untreated nocapeptin A; B) Analysis of nocapeptin after carboxypeptidase Y treatment; C) Analysis of nocapeptin after chymotrypsin treatment.

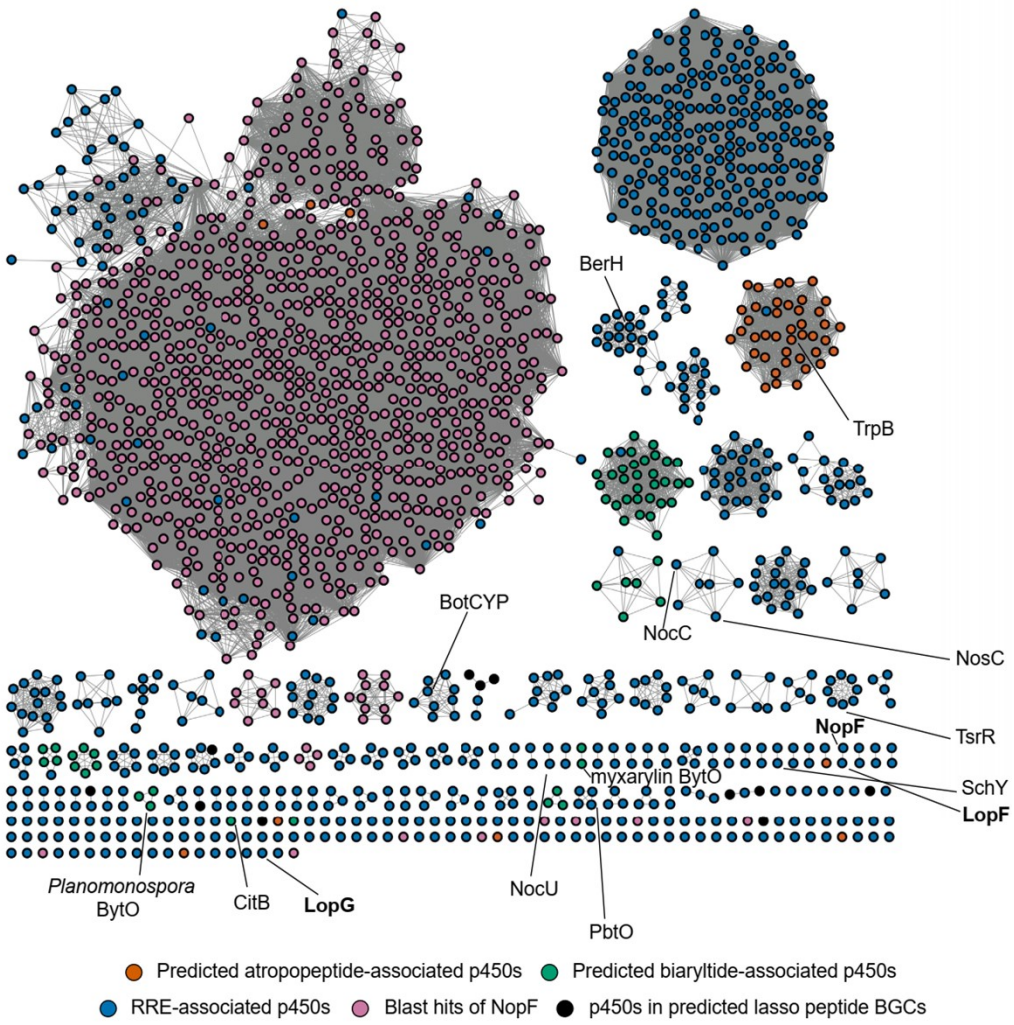


Figure S51. A sequence similarity network (SSN) displayed at RepNode 100 and alignment score = 99. Sequences included: i) top 1000 BLAST-P hits to NopF, ii) all cytochrome P450 proteins (883 sequences) within 10 open reading frames (ORFs) of a detected RRE (RRE-Finder, precision mode), iii) predicted atropopeptide- and biaryltide-associated cytochrome P450 proteins,[14] and iv) cytochrome P450 proteins in predicted lasso peptide BGCs (Figure S67). Experimentally characterized cytochrome P450 proteins (NosC, TsrR, SchY, myxarylin and *Planomospora* BytO, CitB, NocU, PbtO, BotCYP, TrpB) and P450 proteins encoded in the BGCs of RiPPs with known structures (BerH; berninamycin, and NocC; nocathiacin) are annotated. A total of 1995 proteins are represented in this SSN (Supplementary Dataset 1 and 2).

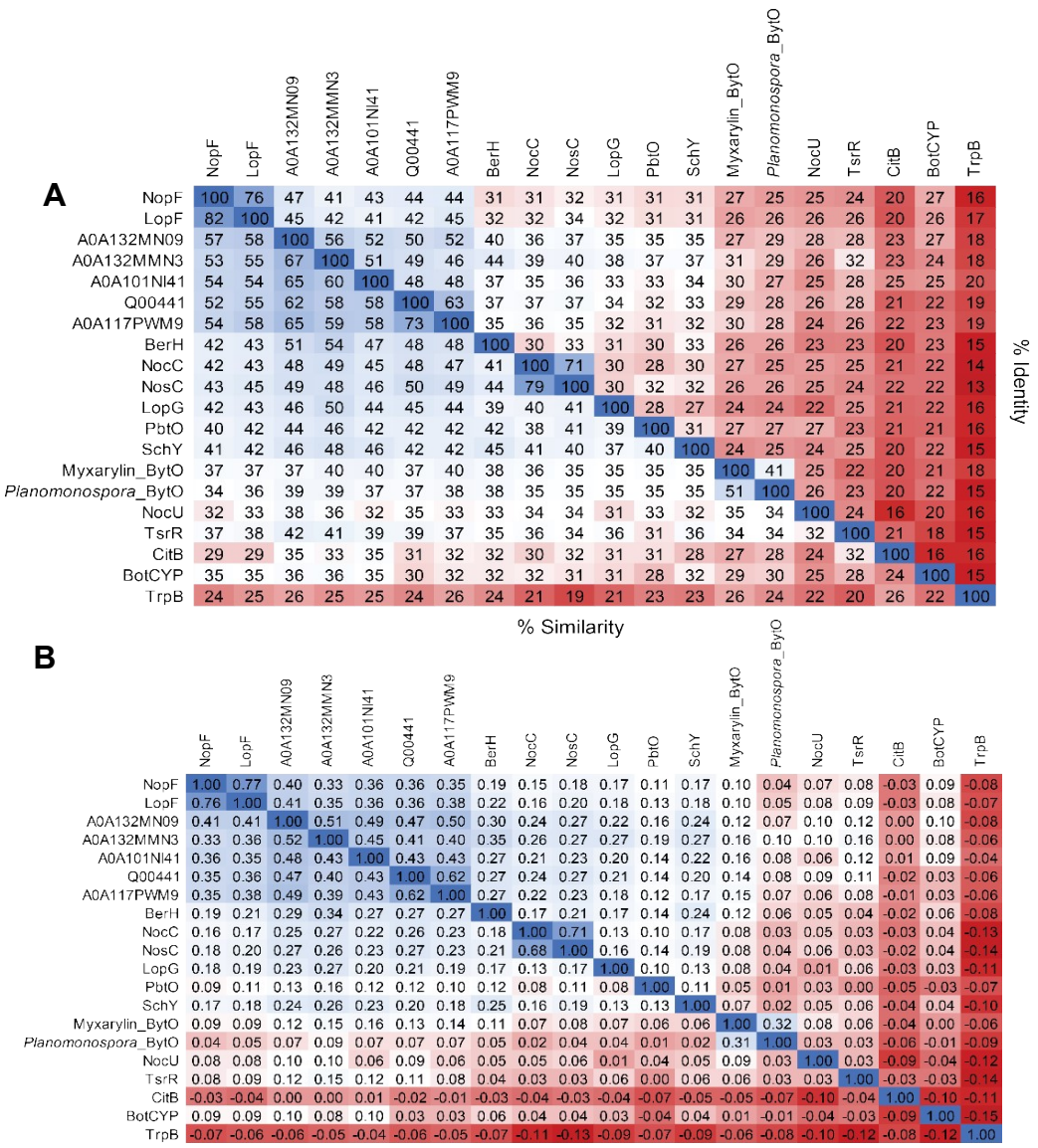


Figure S52. (**A**) Percent similarity and identity matrix and (**B**) global similarity matrix (BLOSUM62) of three cytochrome P450 proteins uncovered in this study (NopF, LopF, and LopG), experimentally characterized cytochrome P450s annotated in the SSN (Figure S49), and five top BLAST-P hits to NopF (indicated with UniProt accession codes).

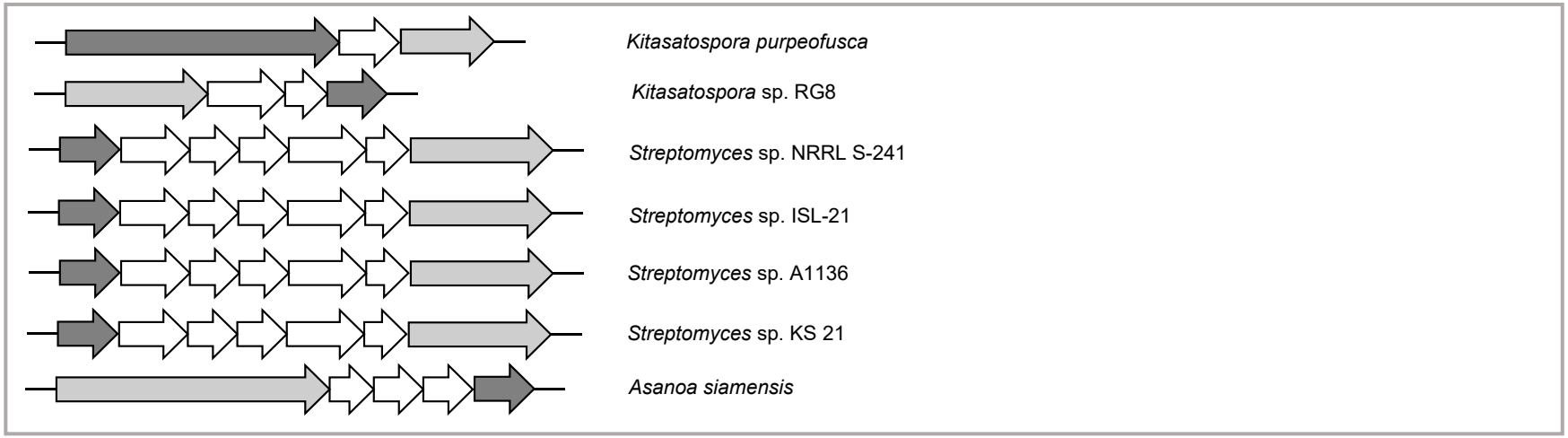


Figure S53. Representative genomic neighborhoods of selected BLAST-P hits to LopH (dark grey; WP_051757134.1, WP_209415770.1, WP_051779479.1, WP_214943134.1, WP_136214693.1, WP_133900279.1, WP_203716497.1, respectively) show a strong co-occurrence with DNA polymerase III, beta subunit (light grey, PF00712/TIGR00663).

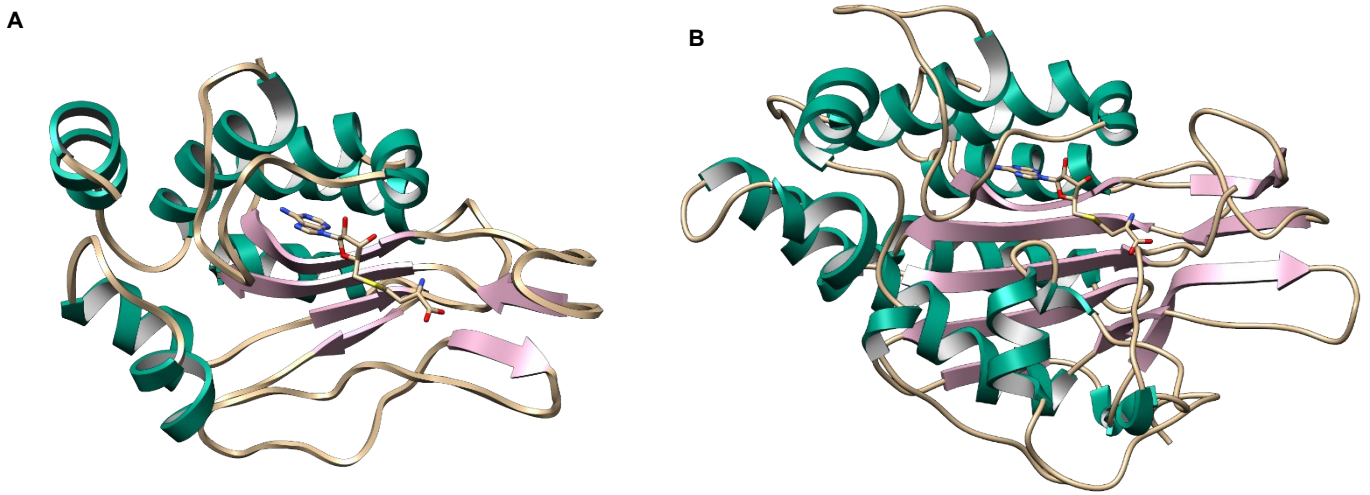


Figure S54. Tertiary structure comparison between (**A**) AlphaFold-predicted LopH with *S*-adenosyl-L-homocysteine (SAH) docked through energy minimization, and (**B**) the closest LopH match upon performing a DALI search: the SAH-binding region of human 5,10-methylenetetrahydrofolate reductase (PDB code: 6fcx, residue 344-646, chain A). Parameters generated from DALI: Z-score = 11, RMSD = 3.1, number of aligned C-alpha atoms = 150, number of residues in target structure = 252. The coordinates for panel A are given as Supplementary File 2. UCSF Chimera was used to generate these images.

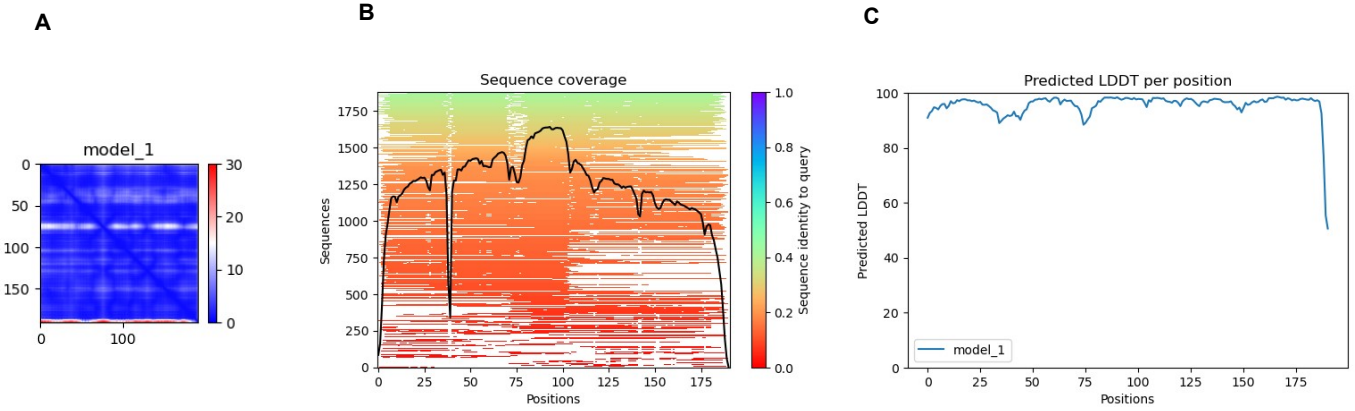


Figure S55. Confidence analysis of the AlphaFold-predicted LopH structure. (**A**) The predicted aligned error (pAE) plot, (**B**) the multiple sequence alignment summarized as a heatmap, and (**C**) the predicted LDDT per position (pLDDT) plot. The pAE and pLDDT plots correspond to the highest-ranked LopH structure generated by AlphaFold.

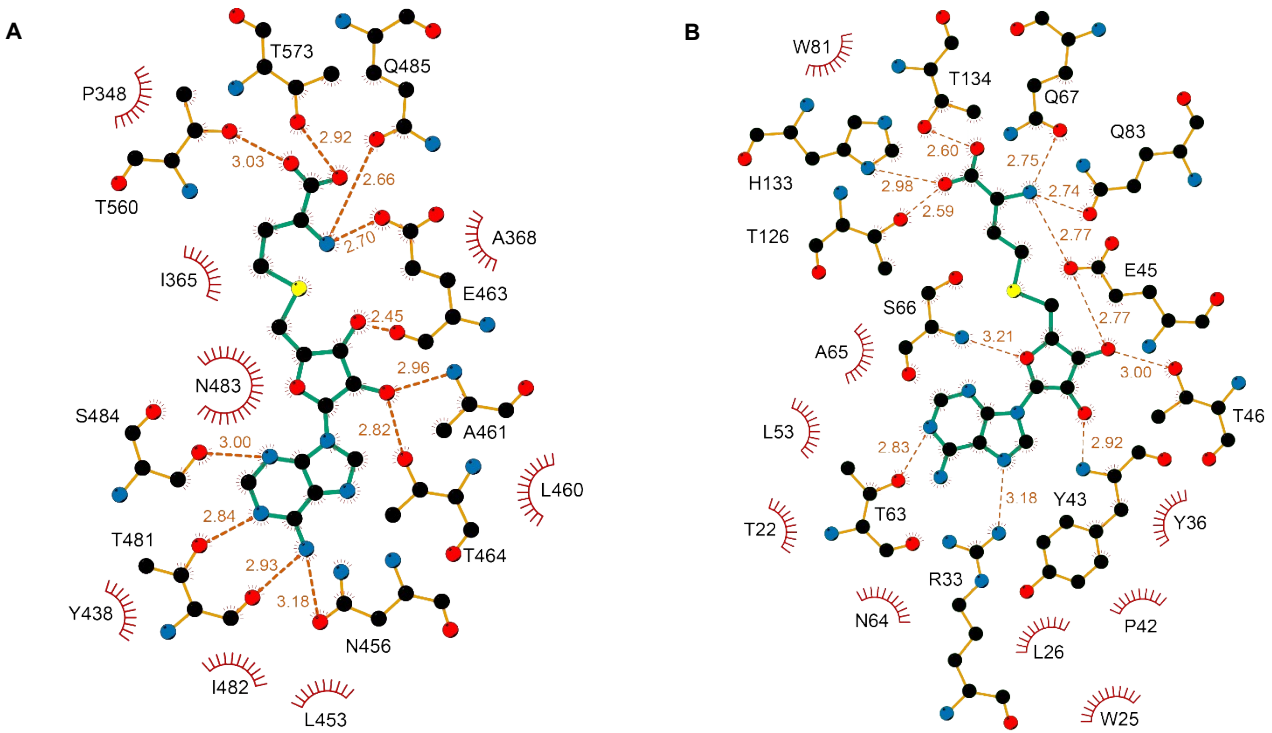


Figure S56. (**A**) A ligand interaction diagram for SAH docked with LopH through energy minimization, (**B**) A ligand interaction diagram for the crystal structure of SAH bound to human 5,10-methylenetetrahydrofolate reductase (PDB code: 6fcx, chain A). Predicted hydrogen bonds are described in orange with distances indicated, while red arcs denote hydrophobic interactions. LigPlot+ was utilized to generate these ligand interaction diagrams.

Figure S57. A secondary-structure alignment generated by DALI between LopH and human 5,10-methylenetetrahydrofolate reductase (Z-score = 11, RMSD = 3.1, number of aligned C-alpha atoms = 150). The three-state secondary structure definitions by DSSP (reduced to h= helix, e=sheet, l=coil) are utilized. Residues directly contacting the ligand are blue (for LopH) and orange (for 6fcx).

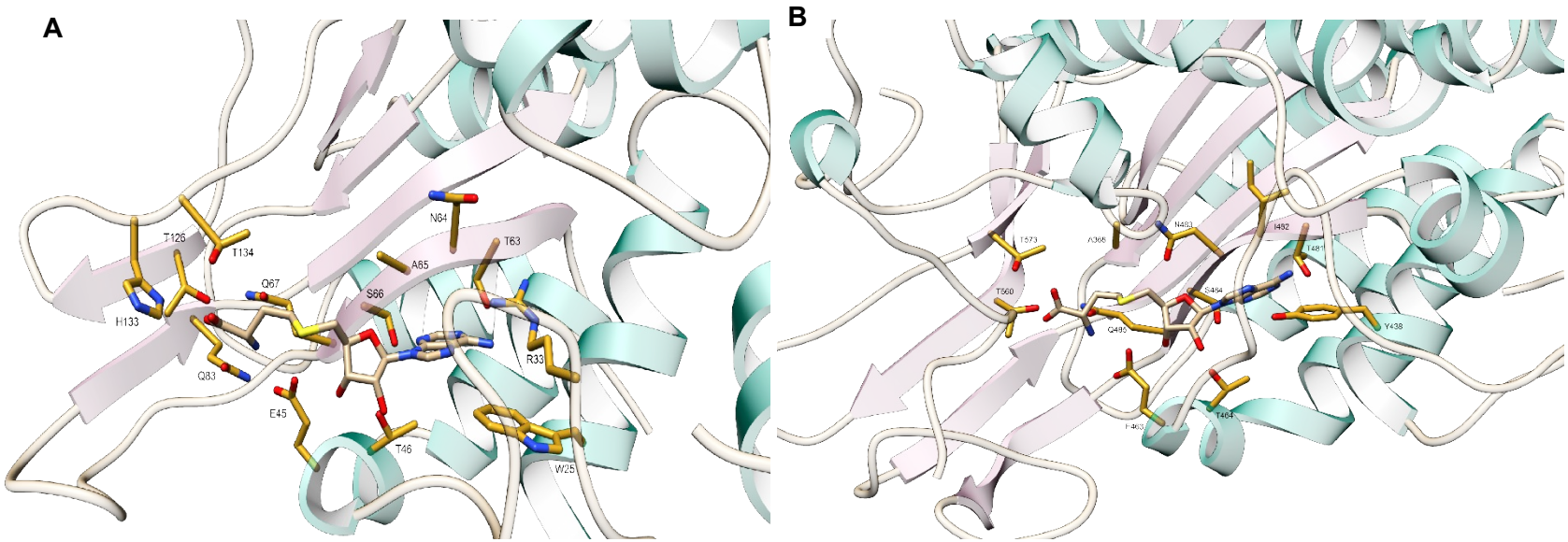


Figure S58. Structures depicting SAH interaction with amino acid residues in the (**A**) LopH structure predicted by AlphaFold, and (**B**) the crystallized human 5,10 - methylenetetrahydrofolate reductase structure (PDB code: 6fcx). The coordinates for panel A are given as Supplementary File 2. UCSF Chimera was used to generate these images.

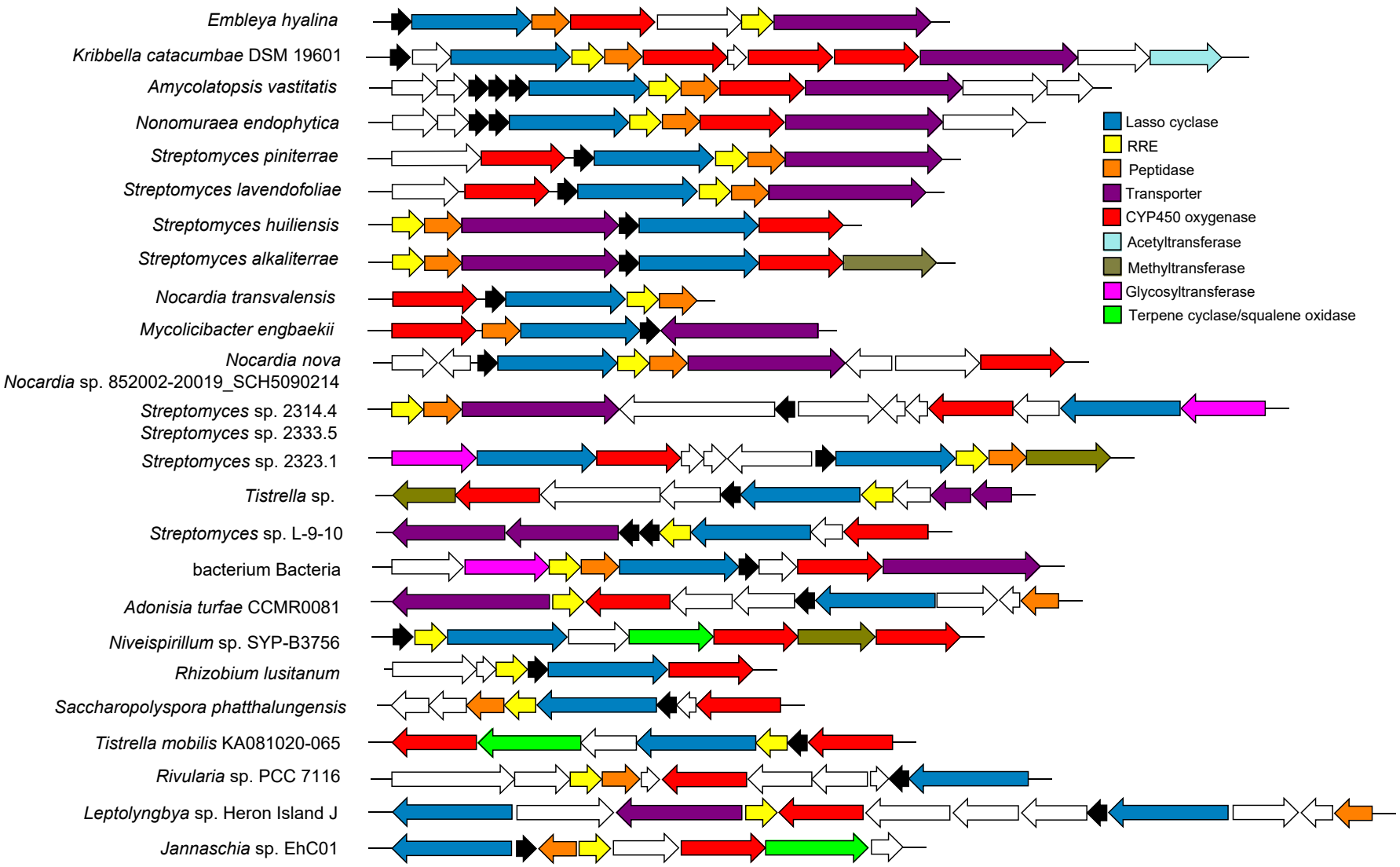


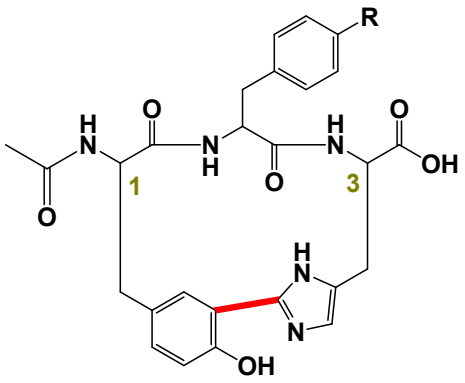
Figure S59. Expanded list of predicted lasso peptides BGCs associated with cytochrome P450 proteins. Accession codes are provided in Table S11.

Table S11. Accession codes and predicted core peptide sequences of P450-associated lasso peptides.

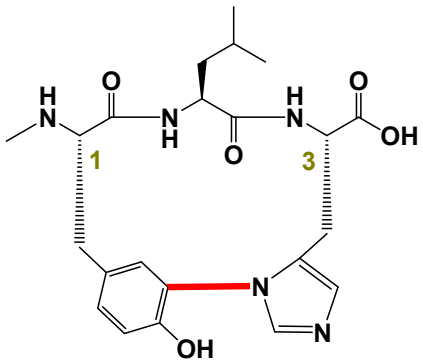
Precursor identifier	CYP450 identifier	Species	Predicted Core Peptide Sequence
GCD92712.1	GCD92710.1	*Embleya hyalina*	RKPGPWWEWIGPNYLD
WP_020387001.1	WP_020387006.1	*Kribbella catacumbae* DSM 19601	GSGGHNWEWIDAWGW
	WP_020387008.1		
	WP_020387009.1		
unannotated	OXM64042.1	*Amycolatopsis vastitatis*	SNCGRTWEWVFCGESRC
unannotated			GPCGNQWEWIACGTEC
unannotated			GGQGCNWEWIGCGWC
MBB5084694.1	MBB5084699.1	*Nonomuraea endophytica*	AGCGCVWEWLTDRLDW
MBB5084695.1			GGCGGPHWEWIYPNYCR
TJZ59061.1	TJZ59062.1	*Streptomyces piniterrae*	LLAKHGNDRLIFSKN
GGU49781.1	GGU49789.1	*Streptomyces lavendofoliae*	LAGQGSPDLLGGHSLL
WP_223766706.1	WP_223766708.1	*Streptomyces huiliensis*	ANKQGMGFDWYLTRK
MQS00863.1	MQS00861.1	*Streptomyces alkaliterrae*	ANRRGMGFDWYLTRK
MBB5915993.1	MBB5915992.1	*Nocardia transvalensis*	GSSYVIIEGYPSAWGSNY
WP_109560825.1	ORV51843.1	*Mycolicibacter engbaekii*	GSGNYLSDSSTGYGYMGWYNRHCDTPETAAPLPPRA
WP_146098681.1	PPJ28635.1	*Nocardia nova*	GVFTTESDLLVGRRGMI
WP_146098681.1	OBA54897.1	*Nocardia* sp. 852002-20019_SCH5090214	GVFTTESDLLVGRRGMI
SEE71607.1	SEE71688.1	*Streptomyces* sp. 2314.4	GTNSFDTADDFSVKSCVLELHVAAR...STCDWALASVQAPEMHGAERCGGLE
PJJ05250.1	PJJ05254.1	*Streptomyces* sp. 2333.5	GTNSFDTADDFSVKSCVLELHVAAR...STCDWALASVQAPEMHGAERCGGLE
WP_159394969.1	SOE10351.1	*Streptomyces* sp. 2323.1	GTNNFDTADDTQYKNA
unannotated	MAM73040.1	*Tistrella* sp.	GTFSGSGSDSSYS
RYJ28010.1	RYJ28015.1	*Streptomyces* sp. L-9-10	FFRNGANEAYFFFQNQND
RYJ28011.1			VFGIRNGDEITWFFDTWQ
RYX82715.1	RYX82717.1	bacterium Bacteria	ASRVGSKLDRAINAGPGTPVGPLLNEISNSLS
unannotated	NEZ55279.1	*Adonisia turfae* CCMR0081	TSTTFVGSDGGSGIFQYAS
MQP68237.1	MQP68242.1	*Niveispirillum* sp. SYP-B3756	SNGNDDGSDSMYS
unannotated	MBB6488871.1	*Rhizobium lusitanum*	GSAGPLAFDFHLSDRNT
MBB5158765.1	MBB5158767.1	*Saccharopolyspora phatthalungensis*	GNQYLYVEGFFSYLGTI
AFK55229.1	AFK55230.1	*Tistrella mobilis* KA081020-065	SGGSGPGSDNNYS
WP_107073100.1	KJS59601.1	*Streptomyces rubellomurinus*	ALGLHGAEPFFPTLHTSWW
WP_015121441.1	AFY57880.1	*Rivularia* sp. PCC 7116	ANAPNFPFNGFDGGSSPNNYAS
unannotated	ESA35183.1	*Leptolyngbya* sp. Heron Island J	TSTTFVGSDGGSGIFQYAS
WP_161489756.1	OAN82783.1	*Jannaschia* sp. EhC01	NNSGSGSDAGIYSS

Relevant Known Scaffolds Containing the PTMs under Investigation:

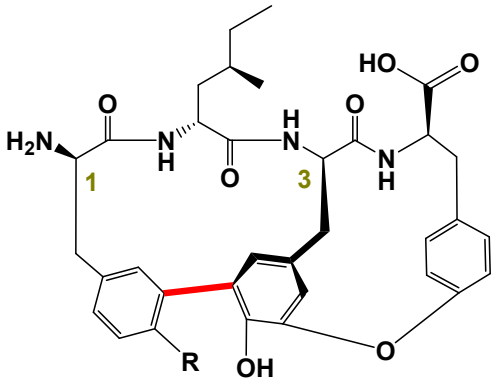
Peptides-Based Crosslinks



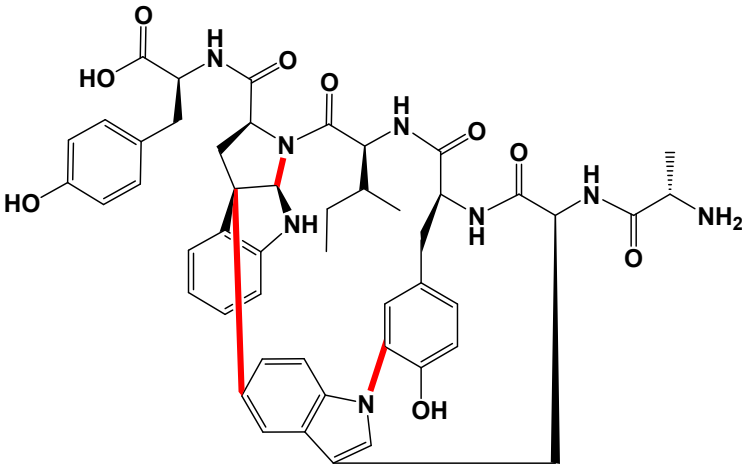
YYH (R= OH), and **YFH** (R= H), **RiPPs, P450**
Cell Chem. Bio. **2021**, *28*, 733-739



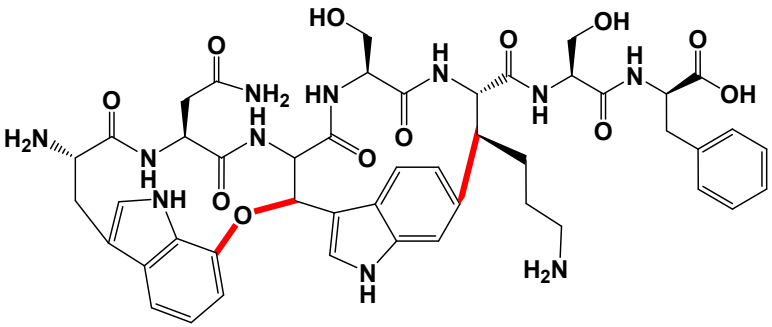
MeYLH, RiPPs, P450
Molecules **2021**, *26*, 7483



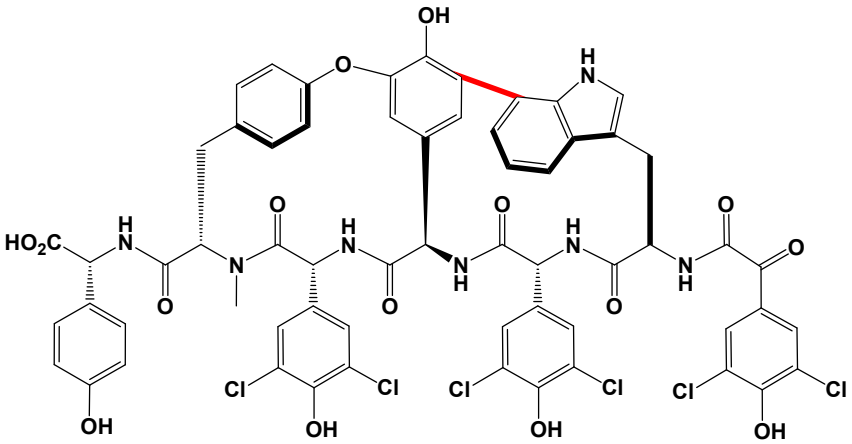
Cittilin A (R= OMe) and **B** (R= OH)**, RiPPs, P450**



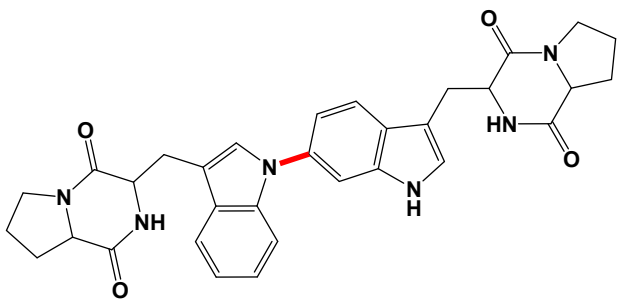
Tryptorubin A, *JACS* **2017**, *139*, 12899-12902, **RiPPs, P450**



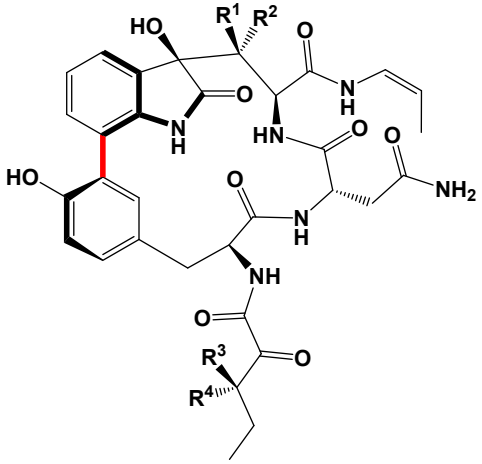
Darobactin A, *Nature* **2019**, *576*, 459-464, **RiPPs, rSAM**



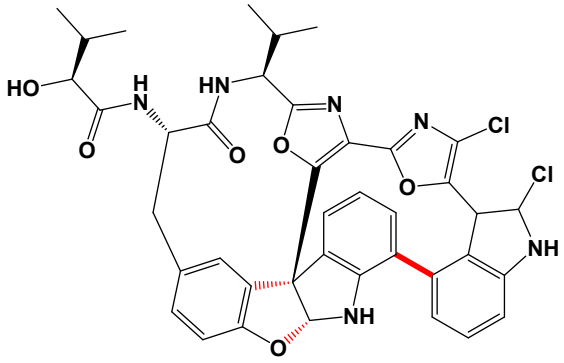
Chloropeptin I, *J. Nat. Prod.* **2001**, *64*, 874-882, **NRPS**



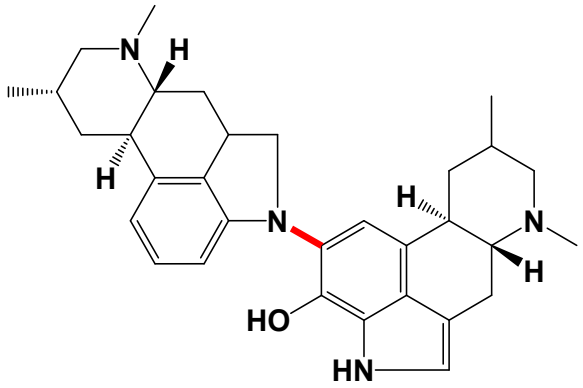
Aspergilazine A, *Tetrahedron Lett.* **2012**, *53*, 2615-2617, **NRPS**



TMC-95A, *J. Org. Chem.* **2000**, *65*, 990–995

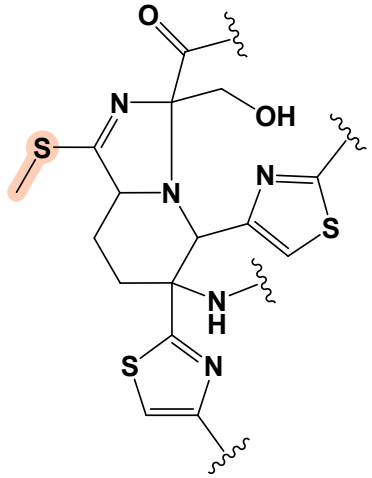


Diazonamide A, *JACS* **1991**, *113*, 2303–2304

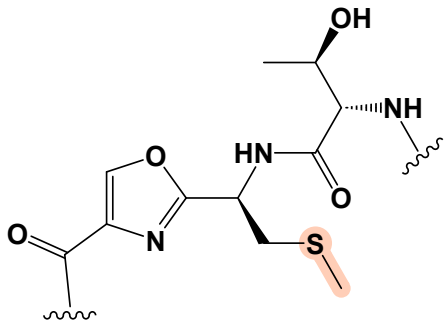


Cividiclavine, *FEMS Symp.* **1982**, *13*, 243–251., *Nat. Prod. Rep.* **1984**, *1*, 21-51., *Genes* **2017**, *8*, 342.

Known S-methylated RiPPs

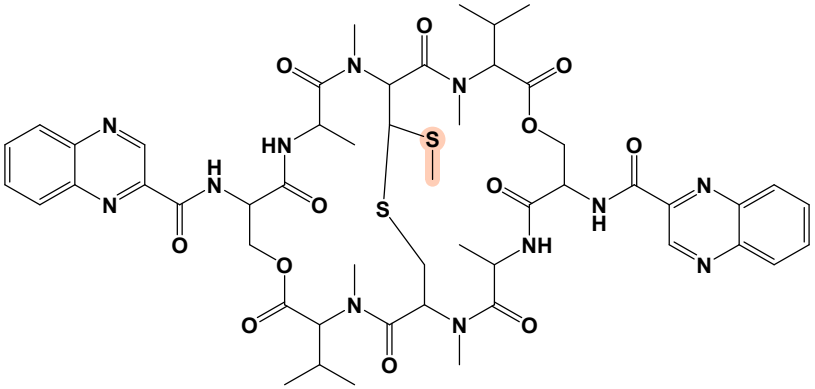


Sch 40832, *J. Antibiot.* **1998**, *51*, 221−224

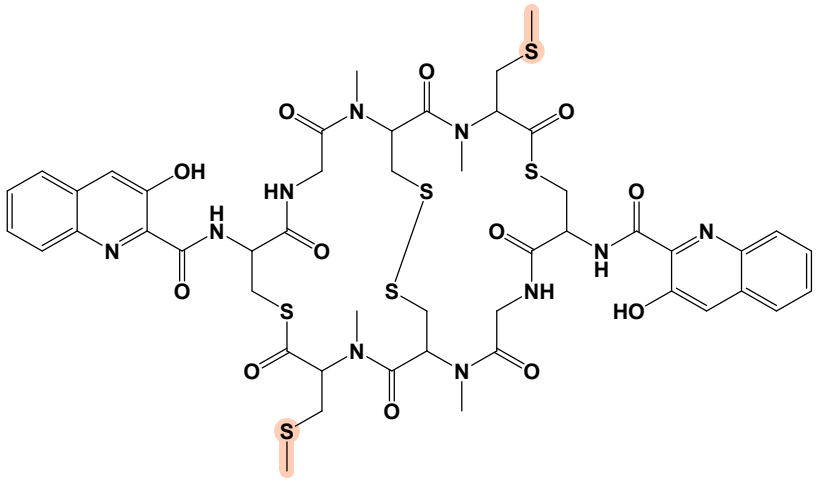


Thioxamycin and **Thioactin,** *J. Antibiot.* **1989**, *42*, 1465−1469
J. Antibiot. **1994**, *47*, 1541-1545

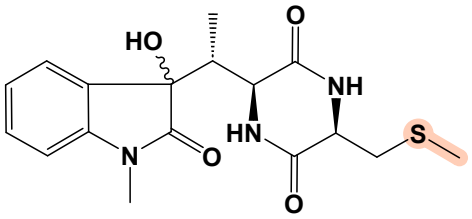
Known *S*-methylated non-RiPPs



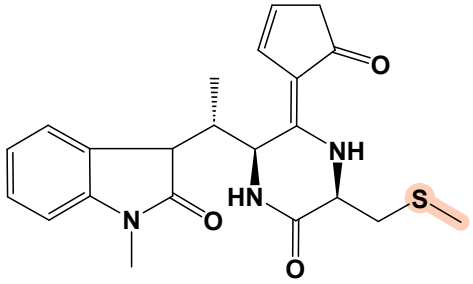
Echinomycin, *Nat. Chem. Biol.* **2006**, *2*, 423−428, **NRPS**



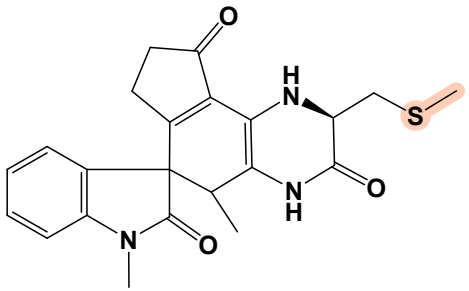
Thiocoraline, *JACS* **2014**, *136*, 17350−17354, **NRPS**



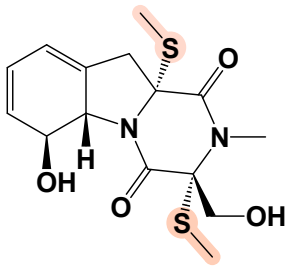
Maremycin A/B, NRPS



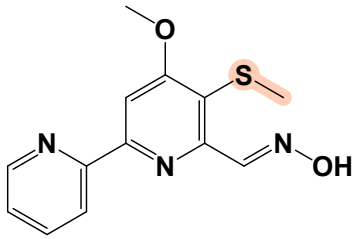
FR900452, NRPS



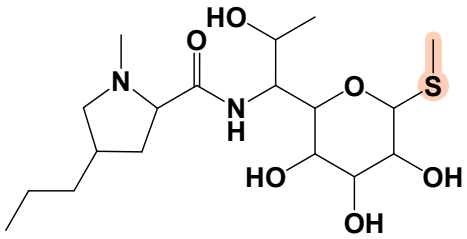
Maremycin G, *ACS Chem. Biol.* **2018**, *13*, 2387–2391, **NRPS**



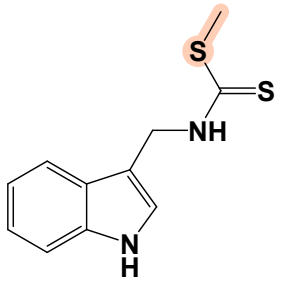
Bismethylgliotoxin
Chem. Biol. **2014**, *21*, 999−1012



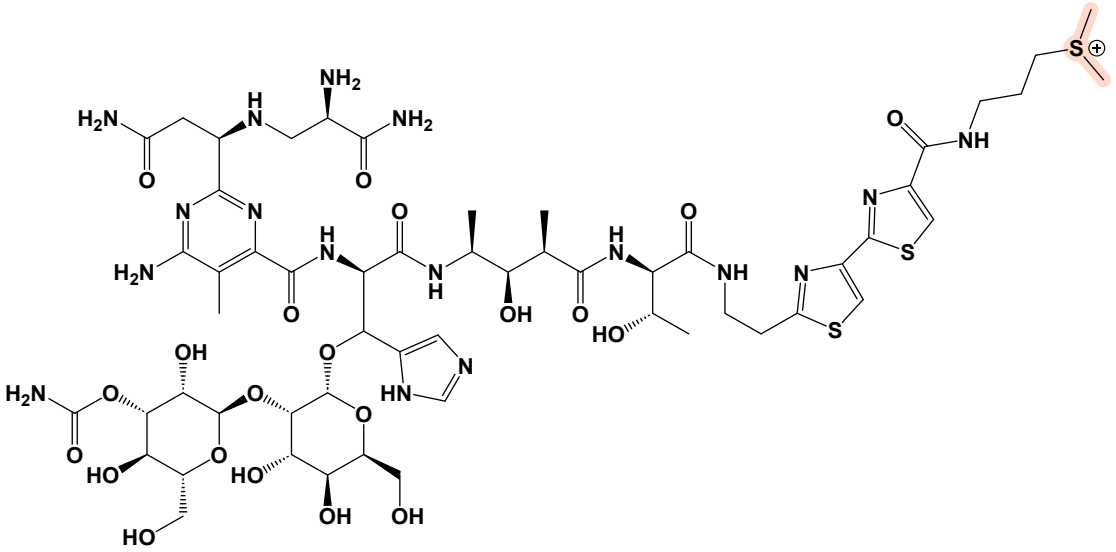
Collismycin
J. Antibiot. **1994**, *47*, 1072−1074



Lincomycin
Adv. Appl. Microbiol. **2004**, *56*, 121−154

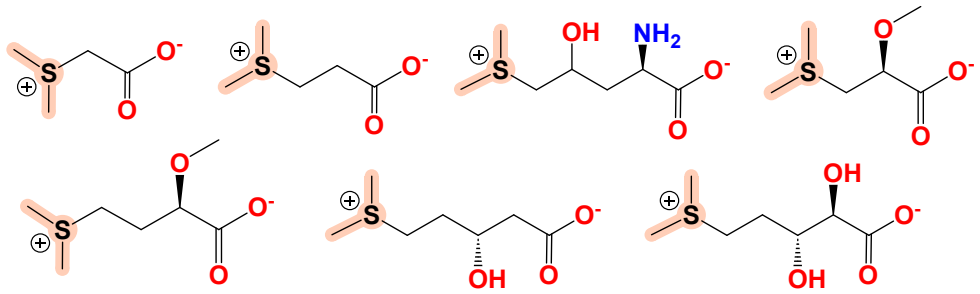


Brassinin
Nat. Prod. Rep. **2011**, *28*, 1381−1405



Bleomycin A2, *J. Antibiot.* **1966**, *19*, 200-209.

DMSP and Related Compounds



Supplemental References

[1] N. Garg, C. A. Kapono, Y. W. Lim, N. Koyama, M. J. A Vermeij, D. Conrad, F. Rohwer, P. C. Dorrestein, *Int. J. Mass Spectrom.* **2015**, *377*, 719-727.

[2] R. T. Williamson, A. V. Buevich, G. E. Martin, T. Parella, *J. Org. Chem.* **2014**, *79*, 3887-3894.

[3] L. Flores-Bocanegra, H. A. Raja. J. W. Bacon, A. C. Maldonado, J. E. Burdette, C. J. Pearce, N. H. Oberlies, *J. Nat. Prod.* **2021**, *84*, 771-778.

[4] K. Blin, S. Shaw, A. M. Kloosterman, Z. C. Powers, G. P van Weezel, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2021**, *49*, W29–W35.

[5] J. I. Tietz, C. J. Schwalen, P. S. Patel, T. Maxson, P. M. Blair, H.-C. Tai, U. I. Zakai, D. A. Mitchell, *Nat. Chem. Biol.* **2017**, *13*, 470–478.

[6] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. R. Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.

[7] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat. Methods* **2022**, *19*, 679–682.

[8] L. Holm, *Nucleic Acids Res.* **2022**, *50*, W210-W215

[9] D. S. Froese, J. Kopec, E. Rembeza, G. A. Bezerra, A. E. Oberholzer, T. Suormala, S. Lutz, R. Chalk, O. Borkowska, M. R. Baumgartner, W. W. Yue, *Nat. Comm.* **2018**, *9*, 1-13.

[10] P. Labute, *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 187-205.

[11] P. Labute, *J. Comput. Chem.* **2008**, *29*,1693-1698.

[12] A. D. Mackerell Jr., M. Feig,C. L. Brooks III, *J. Comput. Chem.* **2004**, *25*, 1400-1415.

[13] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.

[14] R. A. Laskowski, M. B. Swindells, *J. Chem. Inf. Model.* **2011**, *51*, 2778-2786.

[15] a) A. M. Kloosterman, K. E. Shelton, G. P. van Wezel, M. H. Medema, D. A. Mitchell, *mSystems* **2020**, 5, e00267-20. b) K. E. Shelton, D. A. Mitchell, *Methods in Enzymol.* **2022**, doi: 10.1016/bs.mie.2022.08.050.

[16] T. P. Lafosse, M. Blum, S. Chuguransky, T. Grego, B. L. Pinto, G. A Salazar, M. L Bileschi, P. Bork, A. Bridge, L. Colwell, J. Gough, D. H. Haft, I. Letunić, A. M. -Bauer, H. Mi, D. A. Natale, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, A. Bateman, *Nucleic Acids Res.* **2022**, doi: 10.1093/nar/gkac993

[17] R. Zallot, N. Oberg, J. A. Gerlt, Biochemistry **2019**, *58*, *41*, 4169-4182.

[18] P. Nanudorn, S. Thiengmag, F. Biermann, P. Erkoc, S. D. Dirnberger, T. N. Phan, R. Fürst, R. Ueoka, E. Helfrich, *Angew. Chem. Int. Ed.* **2022**, *61*, e202208361.

[19] K. Katoh, D. M. Standley, *Mol. Biol. Evol.* **2013**, *30*, 772-780.

[20] N. Aryal, J. Chen, K. Bhattarai, O. Hennrich, I. Handayani, M. Kramer, J. Straetener, T. Wommer, A. Berscheid, S. Peter, et al. *J. Nat. Prod.* **2022**, *85*, 530-539.

[21] J. B. Patel, F. R. Cockerill, P. A. Bradford, G. M. Eliopoulos, J. A. Hindler, S. G. Jenkins, J. S. Lewis II, B. Limbago, L. A. Miller, D. P. Nicolau, et al. *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically*, 10th ed.; Clinical and Laboratory Standards Institute: Wayne, PA, USA, **2015**; *35*.

[22] a) M. R. Boyd, K. D. Paull, *Drug Dev. Res.* **1995**, *34*, 91-109. b) R. H. Shoemaker, *Nat. Rev. Cancer* **2006**, *6*, 813-823.

[23] R. M. Zanacchi, W. J. Moore, *Aust. J. Chem.* **1980**, *33*, 1505-1510.

[24] a) D. E. Dorman, F. A. Bovey, *J. Org. Chem.* **1973**, *38*, 1719-1722. b) . E. Dorman, F. A. Bovey, *J. Org. Chem.* **1973**, *38*, 2379-2383. c) I. Z. Siemion, T. Wieland, K.-H. Pook, *Angew. Chem. Int. Ed.* **1975**, *14*, 702-703. d) H. R. Kricheldorf, E- T- K. Haupt, D. Müller, *Magn. Reson. Chem.* **1986**, *24*, 41-52.
[25] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, *J. Biomol. NMR* **2002**, *24*, 149-154.
[26] M. Sheinblatt, M. Andorn, A. Rudi, *Int. J. Peptide Protein Res.* **1988**, *31*, 373-387.
[27] E. Gavrish, C. S. Sit, S. Cao, O. Kandror, A. Spoering, A. Peoples, L. Ling, A. Fetterman, D. Hughes, A. Bissell, H. Torrey, T. Akopian, A. Müller, S. Epstein, A. Goldberg, J. Clardy, K. Lewis, *Chem. Biol.* **2014**, *21*, 509-518.