Supporting Information

Characterization of methanobactin from Methylosinus sp. LW4

Grace E. Kenney, Anthony W. Goering, Matthew O. Ross, Caroline J. DeHart, Paul M. Thomas, Brian M. Hoffman, Neil L. Kelleher, Amy C. Rosenzweig*

Departments of Molecular Biosciences and of Chemistry, Northwestern University, Evanston, Illinois 60208, United States

*amyr@northwestern.edu

TABLE OF CONTENTS

Supplementary Methods

Ms. sp. LW4 Mbn production and purification

Ms. sp. LW4 was grown in a 12 L fermenter in medium containing 0.1μM $CuSO_4$ but otherwise grown and harvested as previously described for *Ms. trichosporium* OB3b;[1] no isotopic supplementation was performed for Mbn used in NMR experiments. Spent medium was clarified using a Centramate PE tangential flow filtration system equipped with a 100 kDa Omega cassette (Pall) and loaded onto a 1 in x 30 in Diaion HP-20 column. *Ms.* sp. LW4 Mbn was eluted with 60% MeOH/40% 10mM NH_4OAc and either lyophilized immediately (apo Mbn) or stabilized with near equimolar levels of copper and then lyophilized (CuMbn).

Ms. sp. LW4 (Cu)Mbn was further purified on an Agilent 1100 series HPLC system equipped with a photodiode array detector, using a 30-min gradient at 1 mL/min of 10-50% Solvent B (80% HPLC-grade acetonitrile (Sigma) and 20% 10 mM NH_4OAc) against Solvent A (10 mM NH_4OAc pH 6.0) on a Grace Vydac analytical column (218TP54 300Å 5 μm, 4.6 mm i.d. x 250 mm) or preparative column (218TP1022 300Å 5 μm, 22 mm i.d. x 250 mm, runtimes adjusted for the increased column diameter and a 5mL/min flow rate).

UV-visible spectrophotometry

Spectra were collected on an Agilent 8453 spectrophotometer equipped with a Peltier temperature controller. For *Ms.* sp. LW4 Mbn degradation, 100 mM HCl (final) was added to 250 μL Mbn in a quartz cuvette, and complete spectra were collected over a 4-day period, beginning with spectrum acquisition every 15 s with intervals between acquisitions increasing by 1% each round. The resulting data were analyzed using Igor Pro 6.1 (Wavemetrics).

Inductively coupled plasma mass spectrometry

Samples were prepared in 3% trace metal grade nitric acid (Sigma-Aldrich). Samples were run on a Thermo iCap Qc ICP-MS in the Northwestern Quantiative Bio-Element Imaging Center (QBIC), and calibrated against a standard curve of .1-200ppb multi-element standards (Inorganic Ventures). Three survey scans were performed for each run, and the average intensity was used to calculate the total amount of a given metal isotope.

Electron paramagnetic resonance (EPR) spectroscopy

EPR samples were analyzed in quartz Wilmad EPR tubes via X-band continuous wave (CW) EPR on a Bruker ESP-300 spectrometer, utilizing an Oxford Instruments ESR-900 helium flow cryostat at 20 K. All spectra were collected under the following conditions: microwave frequency 9.36-9.37GHz, power attenuation 35 dB, 160 ms time constant, 16 G modulation amplitude. Background corrected spectra were double integrated in LabCalc over identical field range, and total Cu^{2+} spin concentration of Mbn samples was correlated to a standard curve of known concentration $CuSO_4$ standards in 10 mM ammonium acetate, pH 6.5 (500 μM, 250 μM, 125 μM, 62.5 μM, and 31.25 μM $CuSO_4$), $R^2 = 0.975$. Since intact CuMbn exhibited no signal, spin quantitation was not performed for this sample. *Ms.* sp. LW4 CuMbn samples were digested in 10% trace metal grade nitric acid for 1 h at 100 ºC and were then re-analyzed via EPR, with the nitric acid-derived dilution factors taken into account. After analysis, metal content was confirmed via ICP-MS.

Electrospray mass spectrometry and tandem MS

Following resuspension, *Ms.* sp. LW4 CuMbn extracts were analyzed by HR-LC-MS on an Agilent 1100 HPLC stack paired with a Thermo Q-exactive mass spectrometer in the Kelleher laboratory. The LC method was run with the flow rate of 200 µL/min for 70 min. The compound was resolved using a 35 min gradient from 2-60% solvent B (0.1% formic acid in acetonitrile) against solvent A (0.1% formic acid in water), followed by a 19 min, gradient of 60-98% B.

The MS method was also 70 min, run in positive mode, with a scan window from *m/z* 250-3750 with 35,000 resolution. After each full scan, a data dependent MS^2 cycle was initiated to acquire fragmentation spectra from the top 5 most intense ions present in the MS1 spectrum. MS2 spectra were acquired with a dynamic scan range; fragmentation was induced by higher energy collisional dissociation with normalized collisional energy value of 25. A dynamic exclusion list was used to prevent the re-fragmentation of the same ion species for a period of 20 s.

HCl-degraded *Ms.* sp. LW4 Mbn was analyzed at the Northwestern Integrated Molecular Structure Education and Research Center (IMSERC) via LC-MS on an Agilent 1100 HPLC stack with inline DAD paired with a Bruker AmazonX ion trap mass spectrometer, using electrospray injection. Samples were introduced via injection onto a Supelcosil LC-18-DB column (4.6mm x 3.3 cm, 3µ) and analyzed over a gradient of 5-55% B (80% ACN, 20% 10mM NH_4OAc) against A (10mM NH_4OAc) at .5mL/min. Masses were acquired between 100 and 2000 Da and analyzed using both positive and negative ion modes.

NMR techniques

Lyophilized *Ms.* sp. LW4 CuMbn was resuspended in 10 mM sodium phosphate pH 6.5 at either 10% or 100% D_2O, for a final Mbn concentration of 1-5 mM. All experiments were performed at 25 ºC on a 600 MHz Agilent DD2 instrument equipped with a z-gradient triple resonance (HCN) COLD probe in IMSERC at Northwestern.

In 10% D_2O samples, solvent suppression with excitation sculpting (ES) was used for 1D and homonuclear experiments, namely ^1H [^1H-^1H]-TOCSY and [^1H-^1H]-ROESY experiments; TOCSY and ROESY experiments used mixing times of 80 and 150 ms respectively, 150 ms acquisition time, a spectral width of 12ppm, 128 increments in t_1 and 1077/1258 complex points in t_2 respectively. For each t_1 value, 16 transients were averaged. Presaturation solvent suppression was used in [^1H-^{13}C]-HMBC and [^1H-^{13}C]-HSQC experiments. Proton/carbon heteronuclear two-dimensional experiments used a mixing time of 150 ms, a proton spectral width of 14 ppm, a carbon spectral width of 190 ppm, 128 increments in t_1, and 32 transients with 1258 complex data points in t_2. ^1H and ^{13}C chemical shifts were referenced to residual solvent signals. [^1H-^{15}N]-HSQC experiments used a proton spectral width of 4 ppm, 96 transients, and 599 complex points; the water flip-back approach was used to address solvent signal.

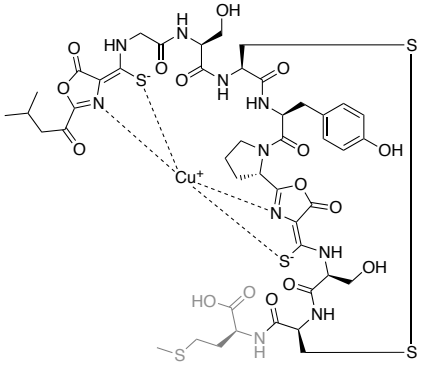
Samples resuspended in 100% D2O were analyzed under similar conditions, with the following differences. [^1H-^1H]-TOCSY used 200 t_1 increments, 300 ms acquisition time and 2155 complex data points. [^1H-^1H]-ROESY also used 200 t_1 increments and 992 complex points in t_2. Both

carbon-proton heteronuclear experiments used a proton spectral width of 11 ppm and a carbon spectral width of 200 ppm, 200 t_1 increments, and 992 complex points. [^{1}H-^{13}C]-HSQC used 8 transients, while [^{1}H-^{13}C]-HMBC involved averaging over 32 transients. No [^{1}H-^{15}N]-HSQC experiments were performed in 100% D_2O.

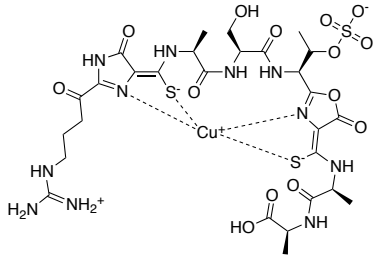
The spectra from these experiments were processed using either MestReNova 9.1.0 or ACD/Spectrus 2015. All ^{1}H, ^{13}C, and ^{15}N assignments are provided in Table S1. Figure S3 shows correlation results from individual 2D NMR experiments, and spectra from all individual experiments are available at the end of this document.

Supplementary figures

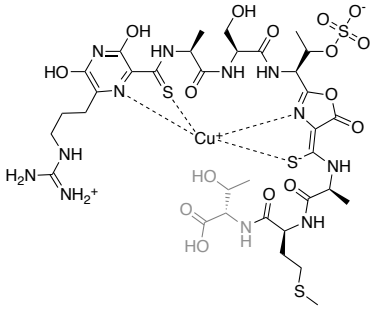
Figure S1. The published stuctures of the six structurally characterized CuMbns. The structures of CuMbns from *Mc. hirsuta* CSC1, *Mc. rosea* SV97, and *Mc.* sp. M were determined via X-ray crystallography; the structures of CuMbns from *Mc.* sp. SB2 and *Ms.* sp. LW4 were determined via NMR (the latter as described in this manuscript); the structure of CuMbn from *Ms. trichosporium* OB3b was initially determined via X-ray crystallography and was updated after analysis of NMR data. Caveats regarding the identification of *Mc.* Mbn heterocycles are discussed in the manuscript. Light gray structural components represent residues that are sometimes lost in the natural product as isolated from spent medium; based on genomic information, there are additional C-terminal residues that are frequently lost in *Mc.* Mbns. In all but *Mc.* sp. M, the residues following the sole oxazolone/thioamide moiety are AATNG, and though no genomic information is available for *Mc.* sp. M, homology to other Mbns suggests the corresponding five residues in that Mbn are AMTNG.



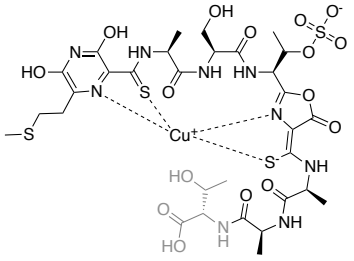
Ms. trichosporium OB3b



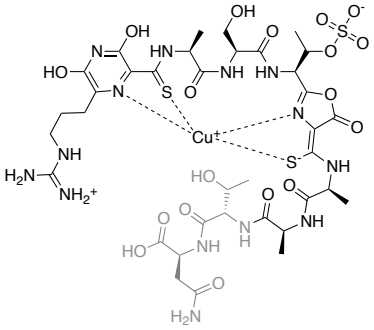
Mc. sp. SB2



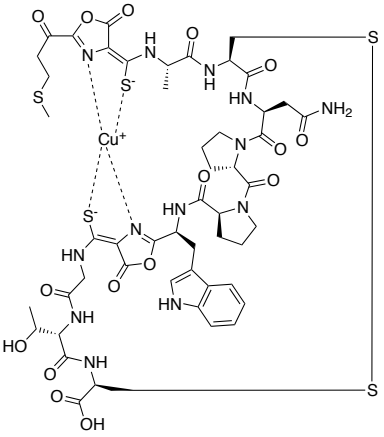
Mc. sp. M



Mc. hirsuta CSC1

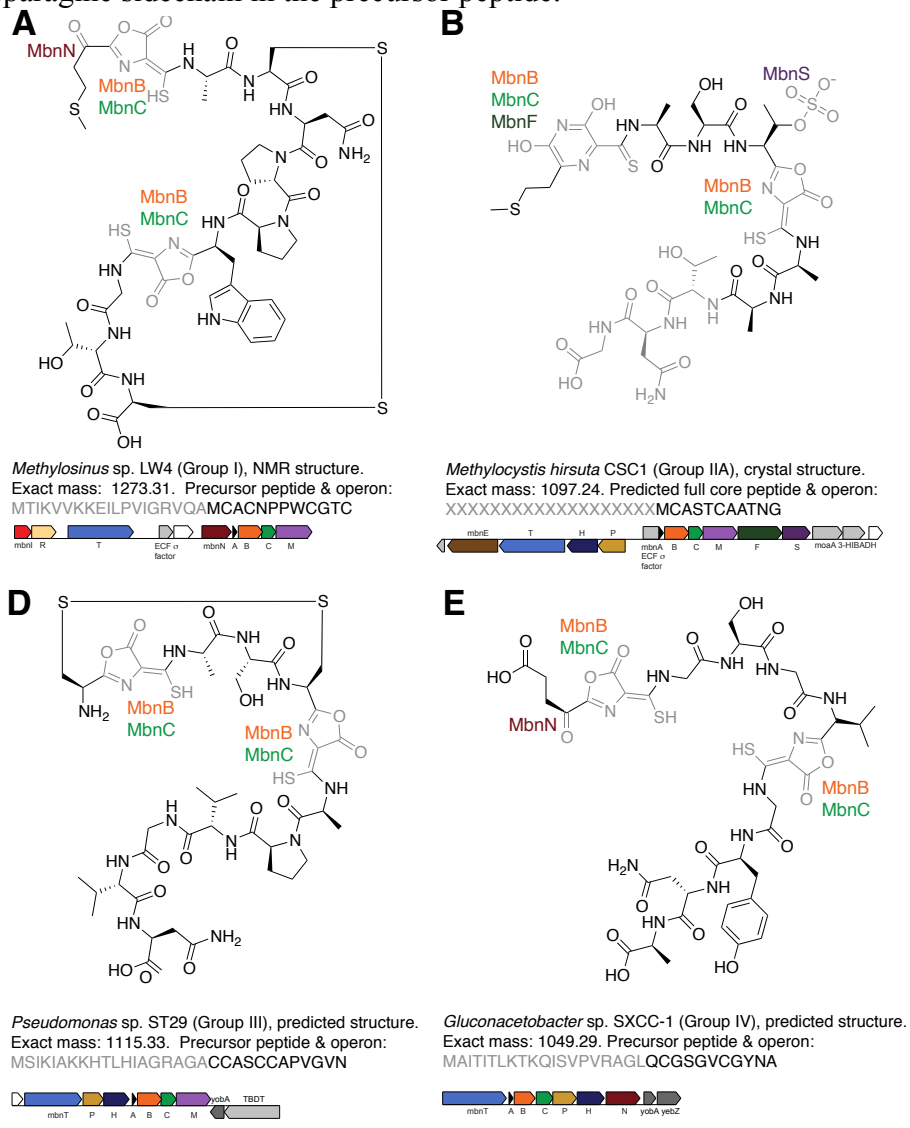


Mc. rosea SV97



Ms. sp. LW4

Figure S2. Predicted correlations between precursor peptide sequence, operon content, and Mbn structure. In the operons, genes encoding MbnB, MbnC, MbnN, MbnS, MbnF, MbnD, and MbnX are colored orange, green, brown, purple, dark green, teal, and pink, respectively. Genes shown in other colors are involved in transport or have unknown functions and are described in detail in references 10 and 11 of the manuscript. (A) *Ms.* sp. LW4, characterized in this manuscript. (B) *Mc. hirsuta* sp. CSC1. Precursor peptide sequence, operon content, and additional C-terminal residues predicted on the basis of the observed crystal structure, and recently confirmed via sequencing of the *Mc. hirsuta* CSC1 Mbn operon. (C) *Pseudomonas* sp. ST29. No Group III Mbns have yet been characterized, and some ambiguity is present due to the pairs of adjacent cysteines; the locations of the two oxazolone/thioamide moieties have been proposed on the basis of the preference for smaller residues immediately adjacent to the modified cysteine as well as spacing between cysteine modifications observed in characterized Mbns. (D) *Gluconacetobacter* sp. SXCC-1 from Group IV. The MbnN enzyme in these species is not closely related to that found in Group I; there is some possibility it may target the highly conserved asparagine sidechain in the precursor peptide.



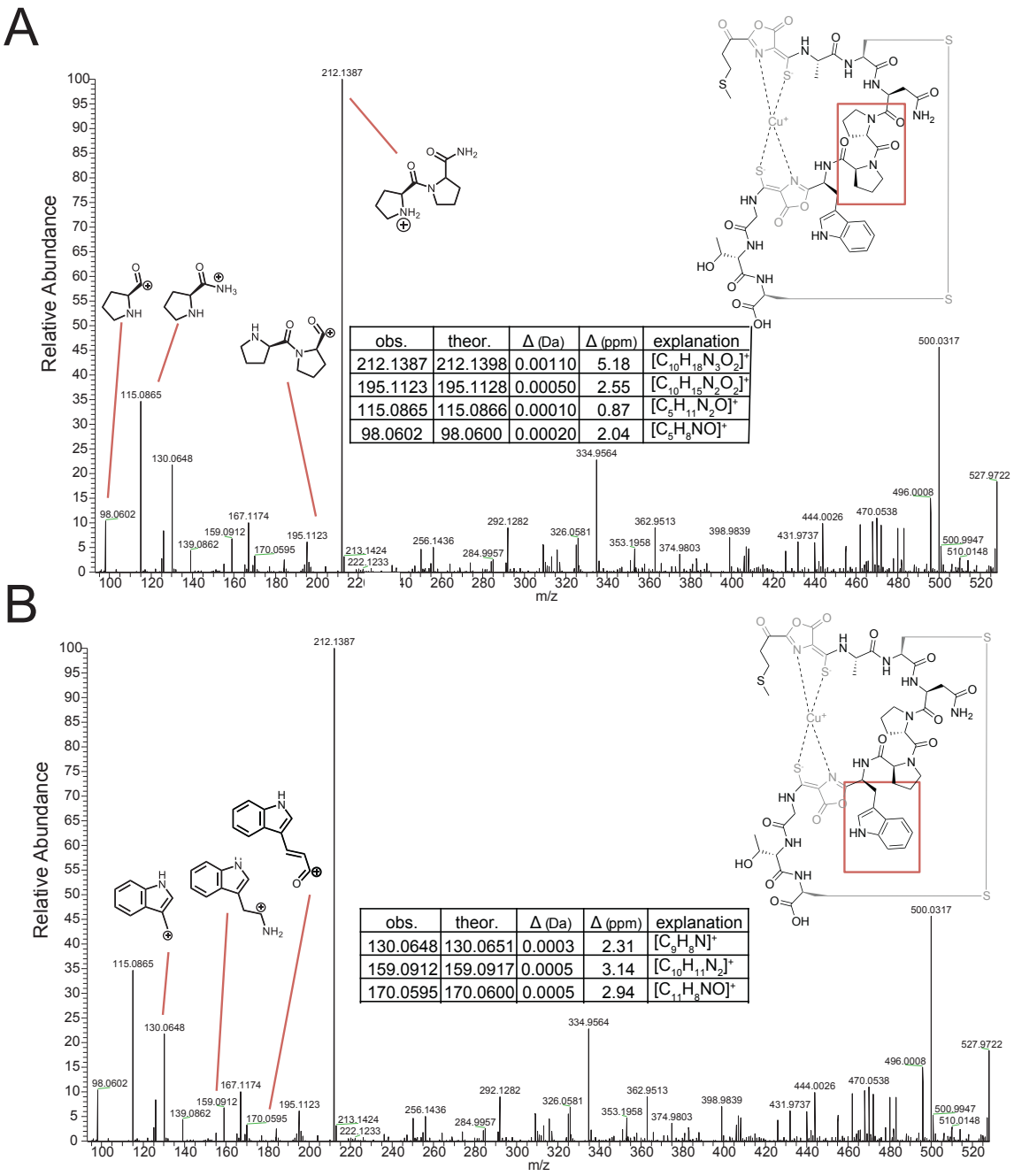
Methylosinus sp. LW4 (Group I), NMR structure.
Exact mass: 1273.31. Precursor peptide & operon:
MTIKVVKKEILPVIGRVQAMCACNPPWCGTC

Methylocystis hirsuta CSC1 (Group IIA), crystal structure.
Exact mass: 1097.24. Predicted full core peptide & operon:
XXXXXXXXXXXXXXXXXXMCASTCAATNG

Pseudomonas sp. ST29 (Group III), predicted structure.
Exact mass: 1115.33. Precursor peptide & operon:
MSIKIAKKHTLHIAGRAGACCASCCAPVGVN

Gluconacetobacter sp. SXCC-1 (Group IV), predicted structure.
Exact mass: 1049.29. Precursor peptide & operon:
MAITITLKTKQISVPVRAGLQCGSGVCGYNA

Figure S3. Fourier-transform tandem MS of *Ms.* sp. LW4 CuMbn. (A) Tryptophan low mass indicator ions are observed, consistent with the presence of a tryptophan residue in the *Ms.* sp. LW4 CuMbn backbone. (B) Proline and diproline low mass indicator ions are also observed, again consistent with the diproline moiety predicted to be present in *Ms.* sp. LW4 CuMbn.



obs.	theor.	Δ (Da)	Δ (ppm)	explanation
212.1387	212.1398	0.00110	5.18	$[C_{10}H_{18}N_3O_2]^+$
195.1123	195.1128	0.00050	2.55	$[C_{10}H_{15}N_2O_2]^+$
115.0865	115.0866	0.00010	0.87	$[C_5H_{11}N_2O]^+$
98.0602	98.0600	0.00020	2.04	$[C_5H_8NO]^+$

obs.	theor.	Δ (Da)	Δ (ppm)	explanation
130.0648	130.0651	0.0003	2.31	$[C_9H_8N]^+$
159.0912	159.0917	0.0005	3.14	$[C_{10}H_{11}N_2]^+$
170.0595	170.0600	0.0005	2.94	$[C_{11}H_8NO]^+$

Figure S4. Correlations from individual 2D NMR experiments. (A) [^{1}H-^{1}H]-TOCSY connectivity in 100% D_2O, depicted via blue arrows. (B) [^{1}H-^{13}C]-HMBC in 100% D_2O, depicted by directional green arrows. (C) [^{1}H-^{1}H]-ROESY in 100% D_2O, inter-residue NOEs depicted via purple arrows. (D) [^{1}H-^{13}C]-HSQC in 100% D_2O, CH bonds indicated by yellow circles. (E) [^{1}H-^{1}H]-TOCSY in 10% D_2O. (F) [^{1}H-^{13}C]-HMBC in 10% D_2O, proton-carbon connectivity depicted by directional green arrows. (G) [^{1}H-^{1}H]-ROESY in 10% D_2O, inter-residue NOEs depicted via purple arrows. (H) [^{1}H-^{13}C]-HSQC in 10% D_2O, CH pairs indicated by yellow circles. (I) [^{1}H-^{15}N]-HSQC in 10% D_2O, NH pairs in red.

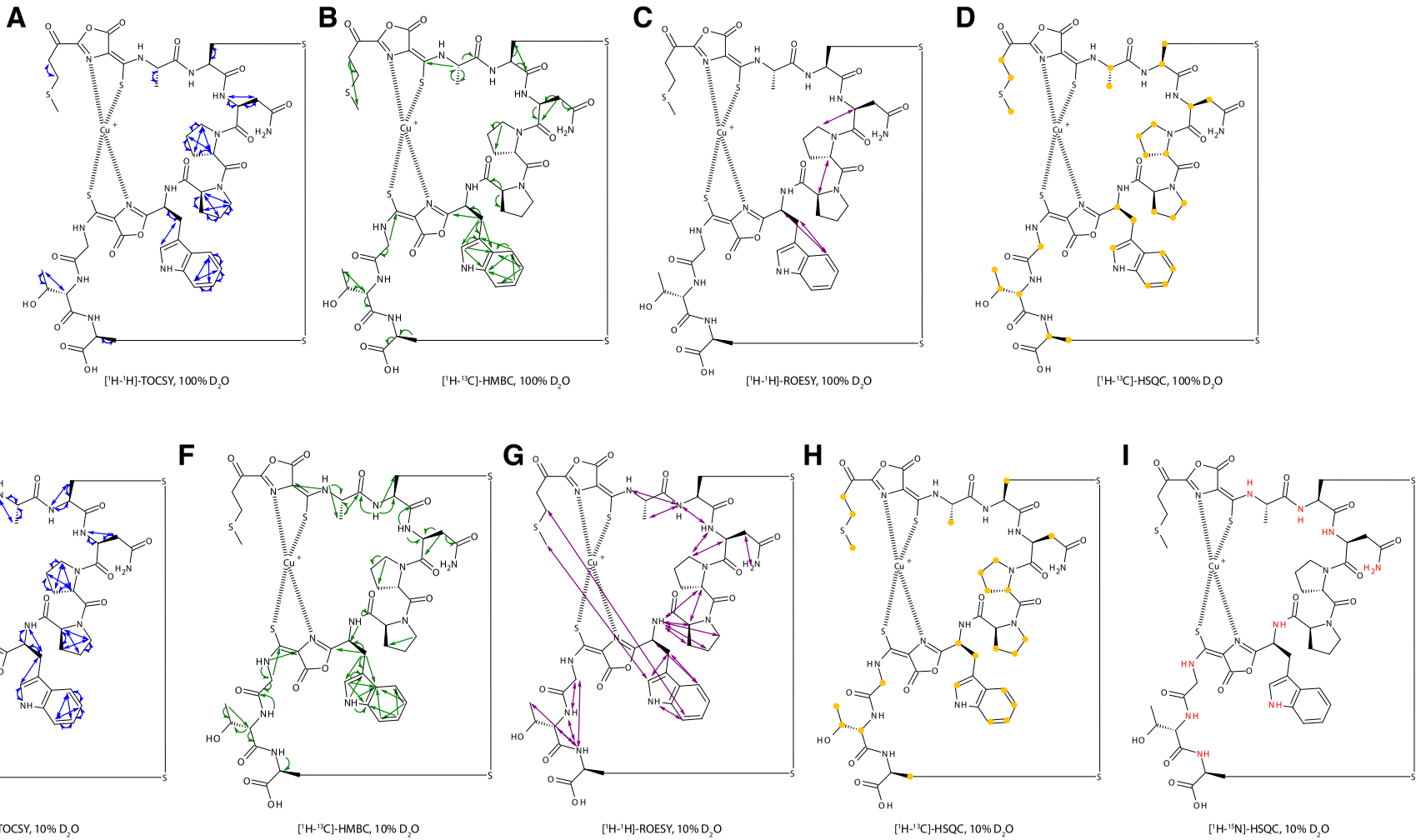

A
B
C
D
[1H-1H]-TOCSY, 100% D2O
[1H-13C]-HMBC, 100% D2O
[1H-1H]-ROESY, 100% D2O
[1H-13C]-HSQC, 100% D2O
E
F
G
H
I
[1H-1H]-TOCSY, 10% D2O
[1H-13C]-HMBC, 10% D2O
[1H-1H]-ROESY, 10% D2O
[1H-13C]-HSQC, 10% D2O
[1H-15N]-HSQC, 10% D2O

Figure S5. Acid hydrolysis of *Ms.* sp. LW4 Mbn. (A) UV-vis spectra of the compound taken at 10 ks intervals over a 96 h period after addition of 100 mM HCl. Due to a drop during the initial 10 ks, there is no true isosbestic point observed during this process, although absorbance appears to change the least at 277 nm; this is similar to what has been observed for *Ms. trichosporium* OB3b Mbn. (B) Hydrolytic degradation pathway of *Ms.* sp. LW4 Mbn.

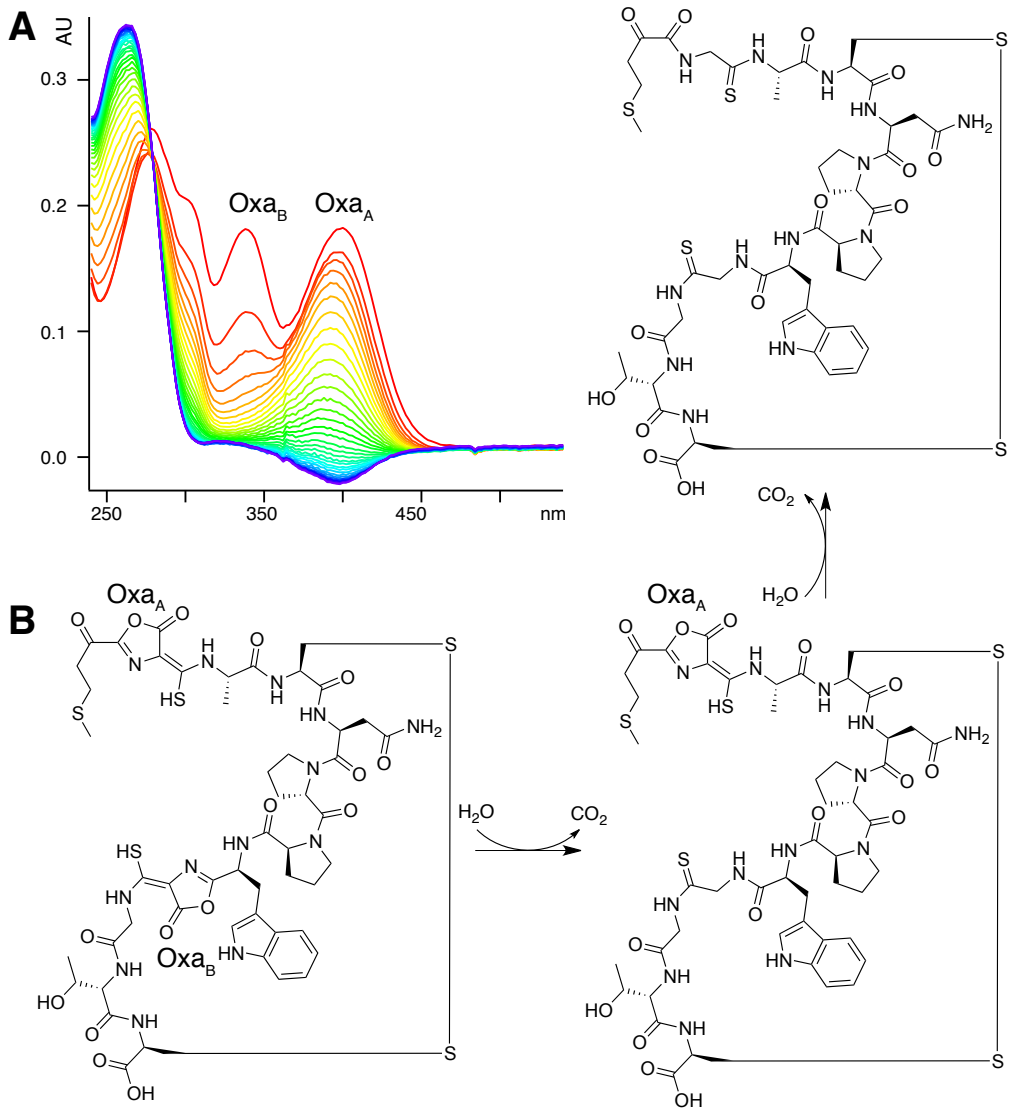


Figure S6. (A) Imidazolone moiety present in the published NMR-derived *Mc.* sp. SB2 CuMbn structure. (B) Pyrazinedione moiety present in the other three published *Mc.* CuMbn structures, as determined via X-ray crystallography. (C) A hydroxy-oxo tautomer of the pyrazinedione moiety, which would exhibit 2D NMR connectivity similar to that observed for the *Mc.* sp. SB2 compound, including the presence of an additional secondary amine group. (D) $2F_o$-F_c electron density map of *Mc.* sp. M CuMbn showing the amino terminal 3-guanidinopropane-pyrazinedione group (PDB code 2YGJ, calculated using the Electron Density Server (eds.bmc.uu.se) contoured at 0.7σ (gray) and 3 σ (blue)). Extra density consistent with protonation is indicated with black arrows, and is similar in magnitude and shape to that observed for the protons on the propane chain (blue arrows).

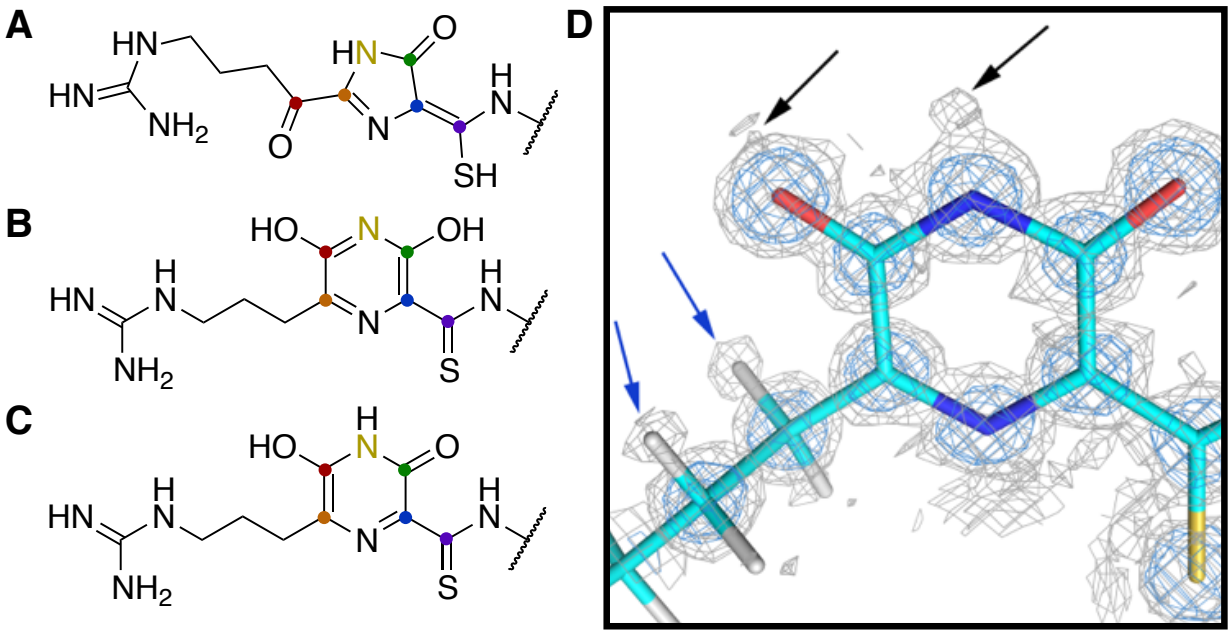


Figure S7. ESI LC-MS analysis of *Ms.* sp. LW4 Mbn after acid hydrolysis at 4 h (A) and 4 d (B). As with Mbn from *Ms. trichosporium* OB3b, hydrolysis of the *Ms.* sp. LW4 heterocycles results in a mass decrease of 26 Da, consistent with a combined hydrolysis/decarboxylation step that yields thioamide-adjacent glycines in the place of the oxazolone rings. Evidence for two consecutive hydrolysis/decarboxylation steps can be found given the presence of two degradation products at earlier timepoints, with the mass shifting towards the fully lysed compound at later timepoints.

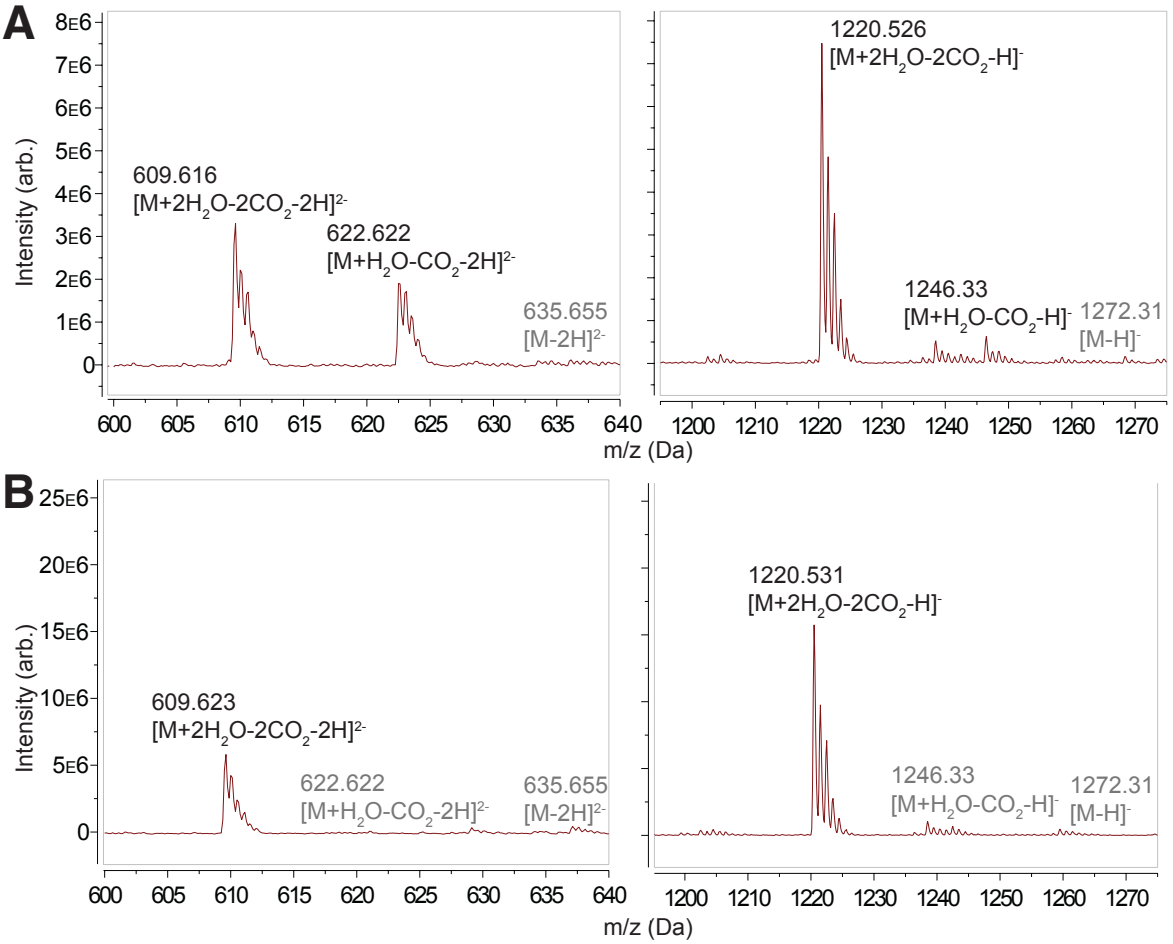
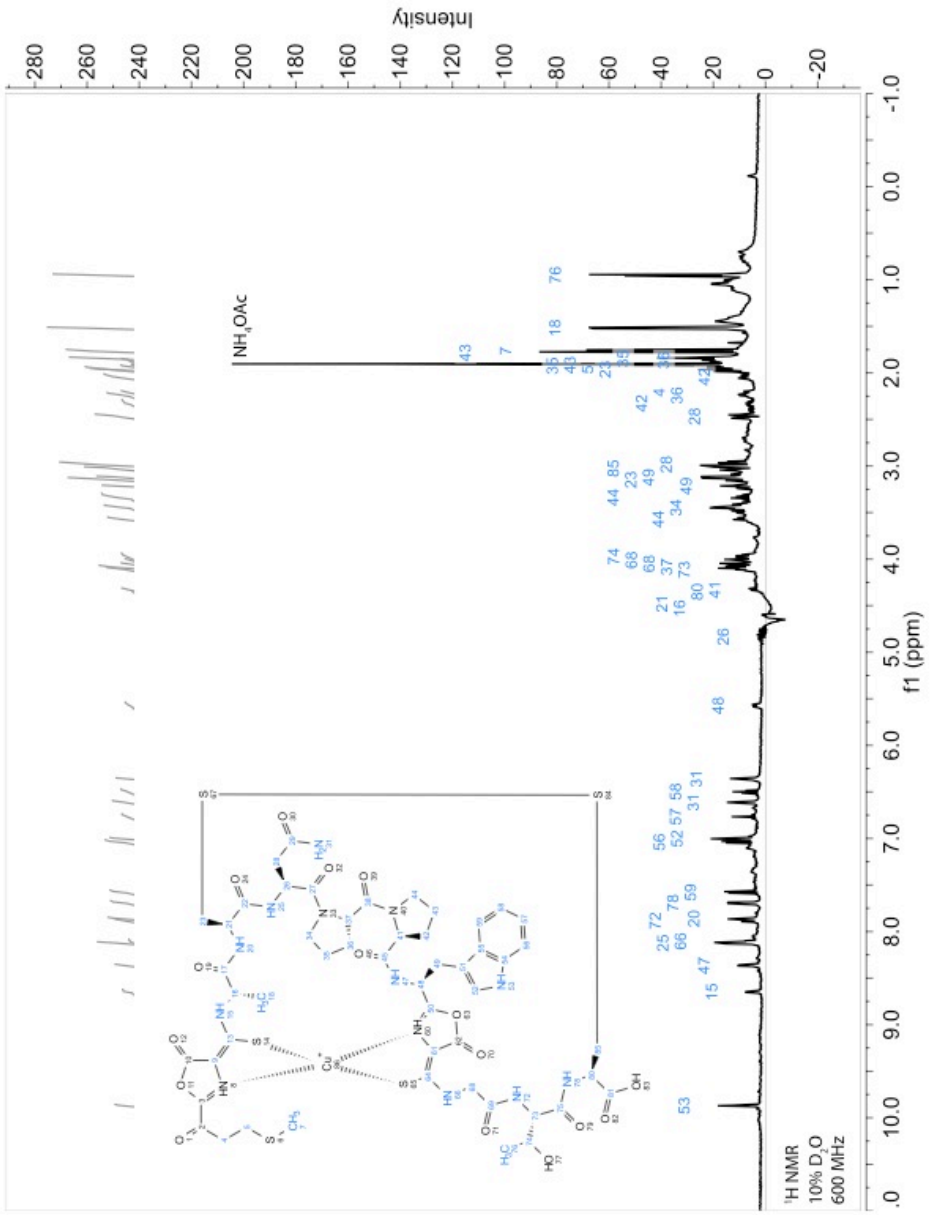


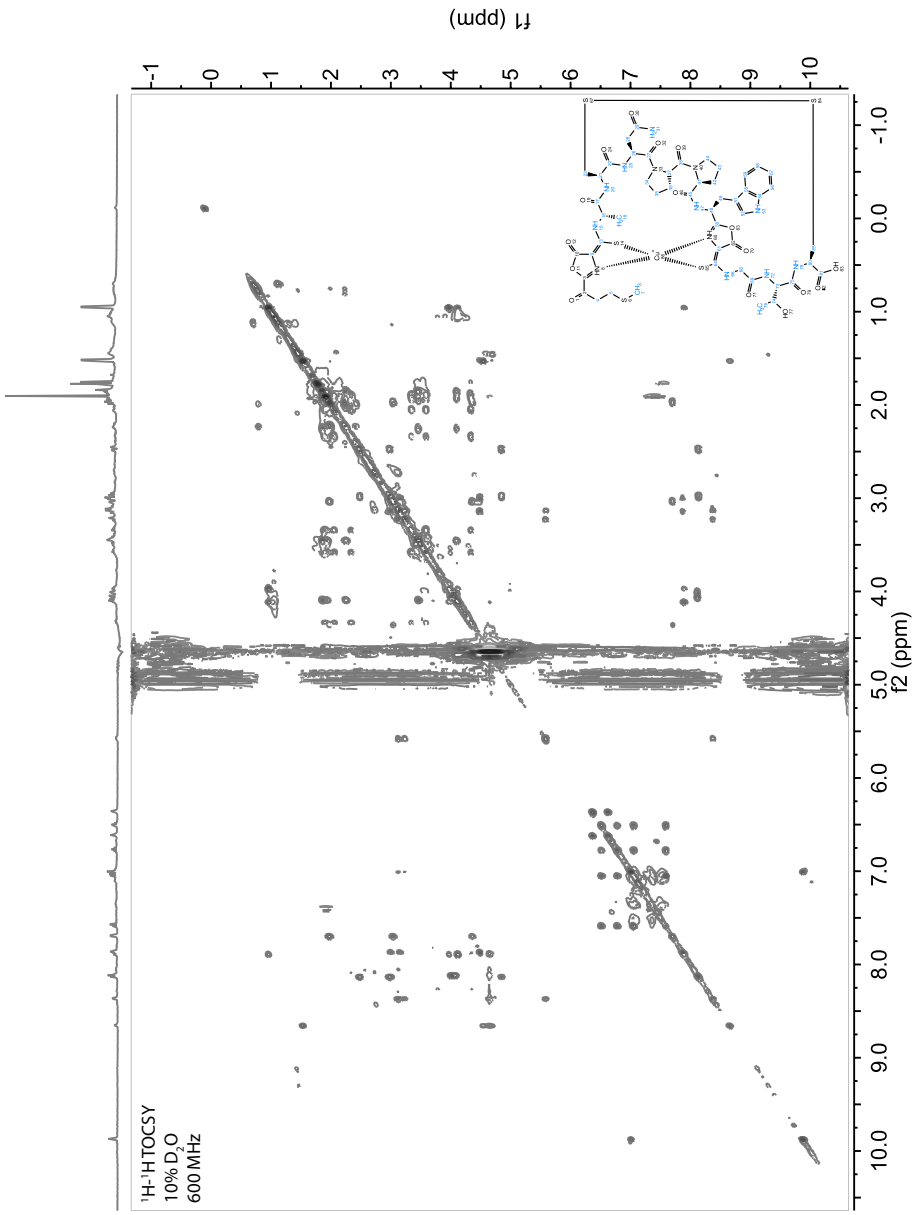
Table S1. ^{1}H, ^{15}N, and ^{13}C nuclear magnetic resonance assignments (ppm) for *Ms.* sp. LW4 CuMbn, collected at 600 MHz in either 10% D_2O, 10 mM phosphate pH 6.5 or 100% D_2O, 10mM phosphate pH 6.5 at 25ºC. Chemical shifts from 10% D_2O experiments indicated first, 100% D_2O values in parentheses. "N.D." means "not detected" and "---" means no signal expected for a given experiment.

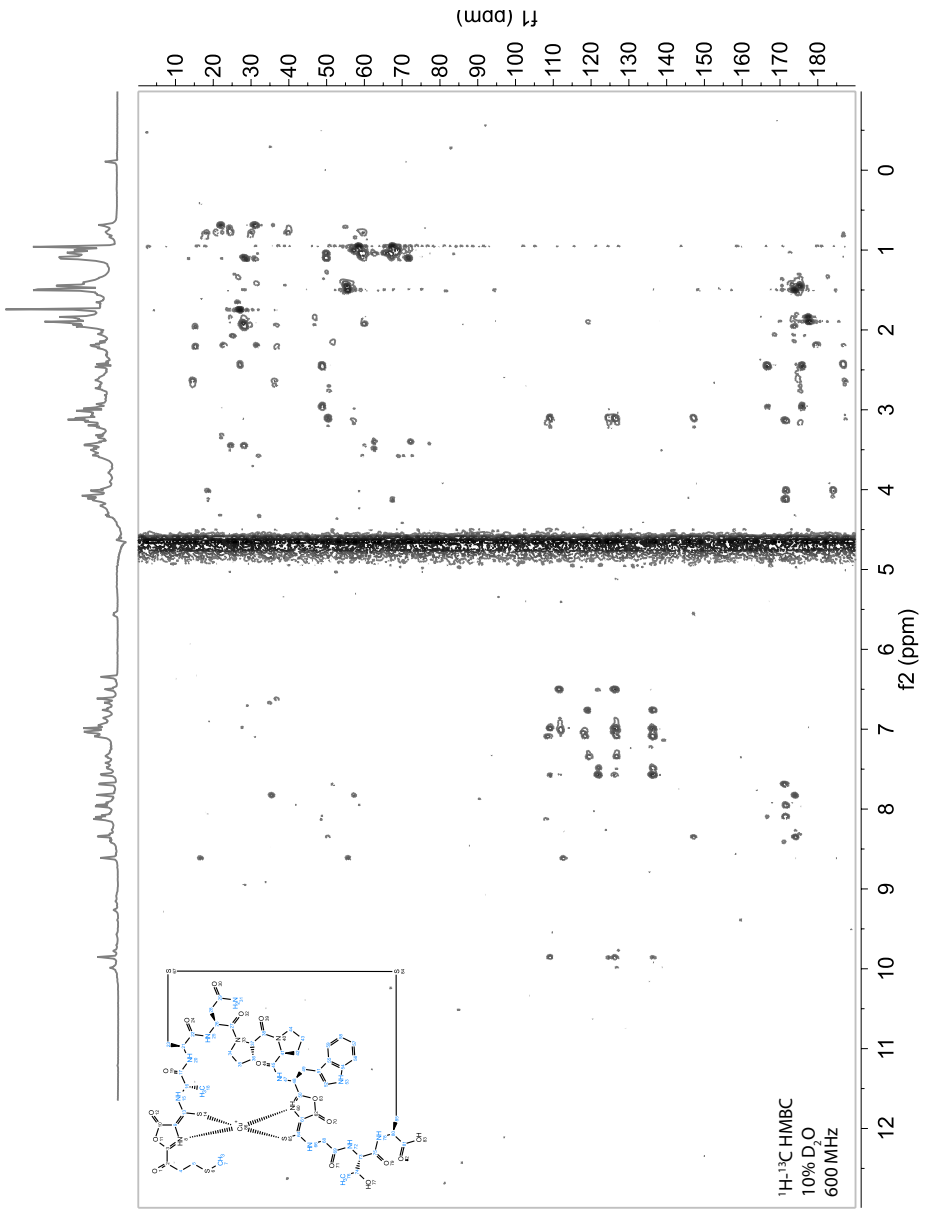
Residue	Position	^{1}H	^{15}N	^{13}C
Met_1	$C_\alpha O$	---	---	N.D.
	C_β /$2H_\beta$	2.22 (2.22)	---	27.96 (26.96)
	C_γ /$2H_\gamma$	1.96 (1.99)	---	37.43 (36.62)
	C_ε /$2H_\varepsilon$	1.78 (1.77)	---	15.12 (15.12)
Oxa_A	C_2	---	---	N.D.
	N_3	---	N.D.	---
	C_4	---	---	112.51
	C_5 (O)	---	---	N.D.
	C_6 (S)	---	---	(185.26)
Ala_3	HN	8.65	134.68	---
	C_α /H_α	4.53 (4.51)	---	55.28 (55.46)
	C_β /$3H_\beta$	1.52 (1.49)	---	16.32 (16.35)
	CO	---	---	173.92 (173.77)
Cys_4	HN	7.86	115.1	---
	C_α /H_α	4.47(4.48)	---	57.12 (57.00)
	C_β /$2H_\beta$	2.96, 3.13 (3.00, 3.15)	---	35.12 (35.26)
	CO	---	---	171.32 (171.36)
Asn_5	HN	8.12	124.91	---
	C_α /H_α	4.84 (4.84)	---	48.56 (48.76)
	C_β /$2H_\beta$	2.46, 2.96 (2.48, 2.98)	---	36.72 (36.62)
	$C_\gamma O$	---	---	175.70 (175.83)
	N_ε /$2H_\varepsilon$	6.36, 6.61	111.96	---
	CO	---	---	166.37 (166.59)
Pro_6	N	---	N.D.	---
	C_α /H_α	4.09 (4.11)	---	60.10 (59.93)
	C_β /$2H_\beta$	2.23, 1.86 (2.25, 1.86)	---	27.02 (27.94)
	C_γ /$2H_\gamma$	1.85, 1.93 (1.88, 1.93)	---	24.43 (24.55)
	C_δ /$2H_\delta$	3.45 (3.46)	---	46.79 (46.88)
	CO	---	---	171.14 (171.36)

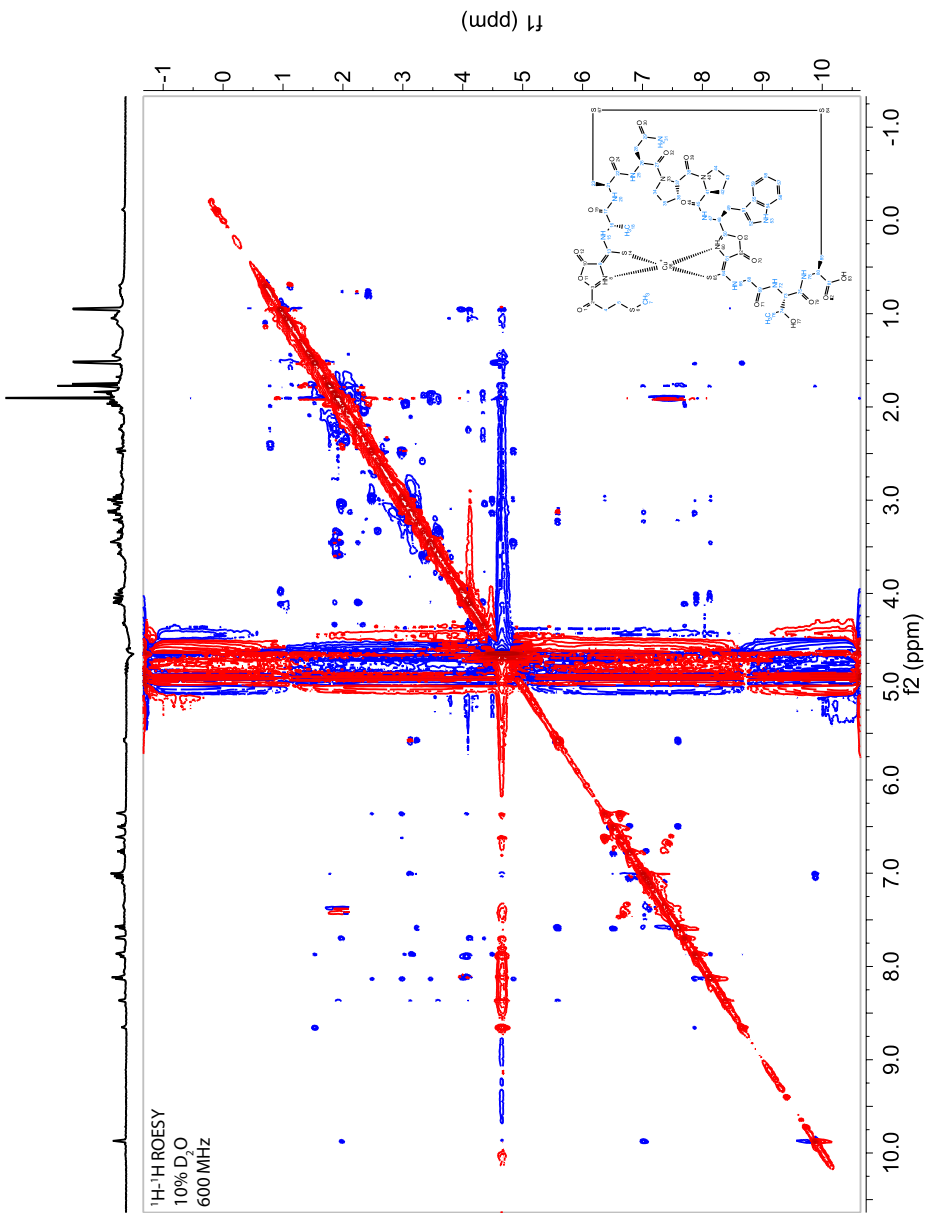
Pro_7	N	---	N.D.	---
	C_α /H_α	4.33 (4.37)	---	(56.97)
	C_β /2H_β	2.32, 2.03 (2.33, 2.05)	---	31.93 (32.06)
	C_γ /2H_γ	1.92, 1.89 (1.91)	---	(21.85))
	C_δ /2H_δ	3.58, 3.33	---	48.11 (47.72)
	CO	---	---	173.92 (170.61)
Trp_8	HN	8.36	124.91	---
	C_α /H_α	5.57 (5.58)	---	50.040 (50.22)
	C_β /2H_β	3.11, 3.22 (3.12, 3.23)	---	27.46 (27.61)
	C_γ	---	---	108.98 (109.14)
	$C_{\delta1}$ /$H_{\delta1}$	7.00 (7.00)	---	124.58 (124.43)
	$C_{\delta2}$	---	---	125.95 (126.17)
	$N_{\varepsilon1}$ /$H_{\varepsilon1}$	9.87	128.47	---
	C_ε	---	---	135.94 (136.22)
	$C_{\varepsilon3}$ /$H_{\varepsilon3}$	7.58 (7.59)	---	118.90 (119.09)
	$C_{\zeta2}$ /$H_{\zeta2}$	7.04 (7.05)	---	111.35 (111.38)
	$C_{\zeta3}$ /$H_{\zeta3}$	6.50 (6.51)	---	117.92 (117.98)
	$C_{\eta2}$ /$H_{\eta2}$	6.77 (6.77)	---	121.68 (121.70)
Oxa_B	C_2	---	---	147.00 (147.07)
	N_3	---	N.D.	---
	C_4	---	---	107.91
	C_5 (O)	---	---	N.D.
	C_6 (S)	---	---	183.78 (184.00)
Gly_{10}	HN	8.11	114.68	---
	C_α /2H_α	4.00, 4.05 (4.01, 4.09)	---	48.50 (48.26)
	CO	---	---	171.28 (171.36)
Thr_{11}	HN	7.89	117.30	---
	C_α /H_α	4.10 (4.13)	---	58.28 (58.30)
	C_β /H_β	3.97 (3.99)	---	67.38 (67.28)
	C_γ /3H_γ	.95	---	18.19 (18.54)
	CO	---	---	171.03 (171.00)
Cys_{12}	HN	7.69	127.53	---
	C_α /H_α	4.35 (4.41)	---	51.99 (51.70)
	C_β /2H_β	1.96, 3.01 (2.07, 3.03)	---	37.43 (37.62)
	CO	---	---	170.76 (175.83)

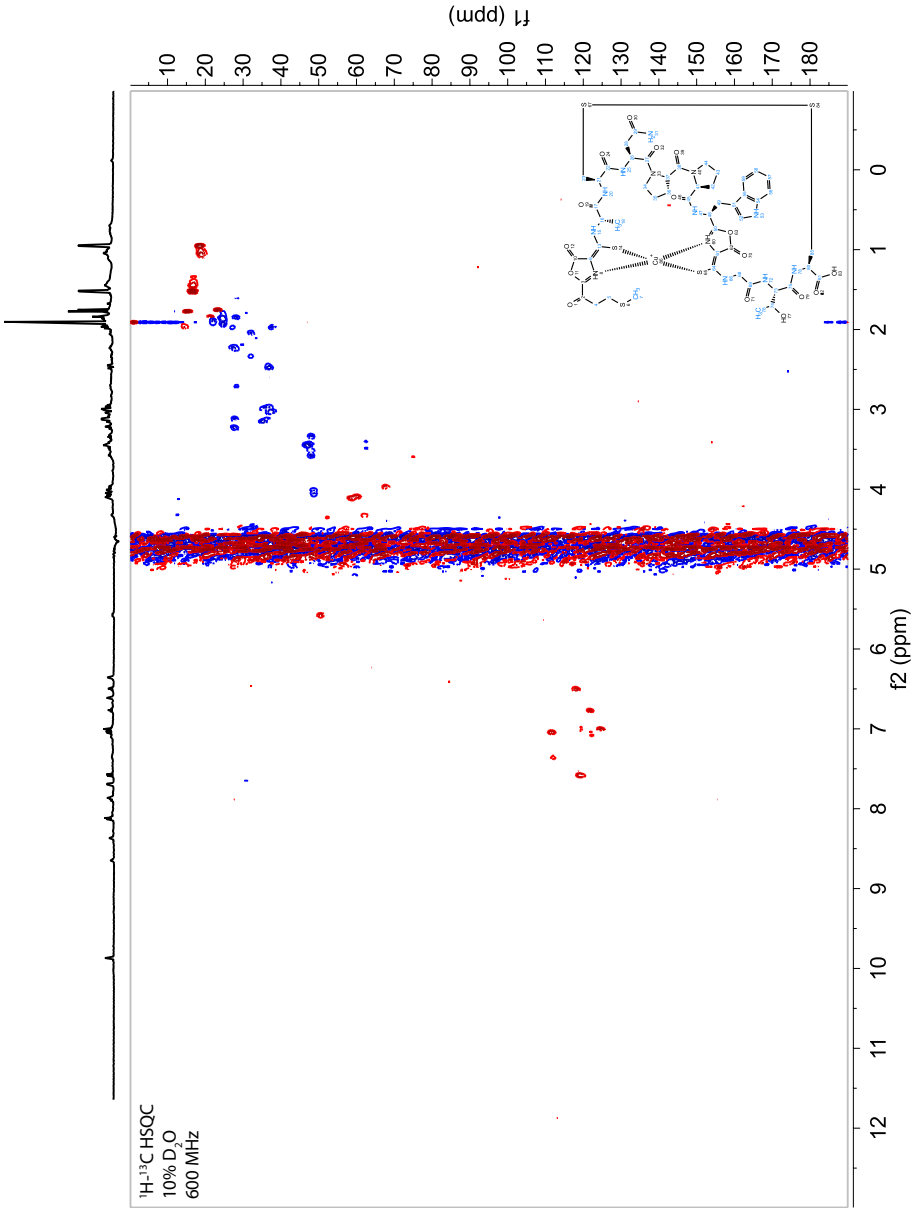
NMR Data

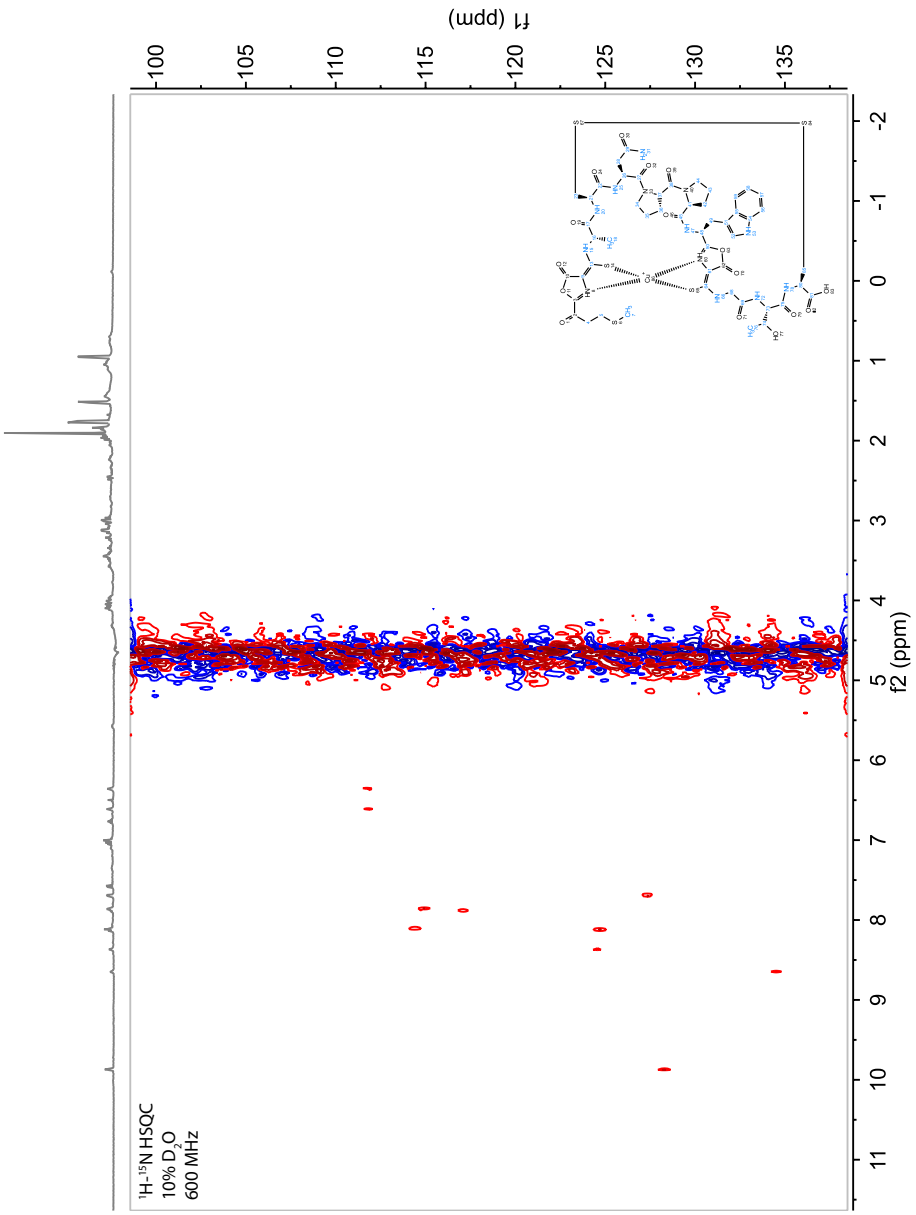


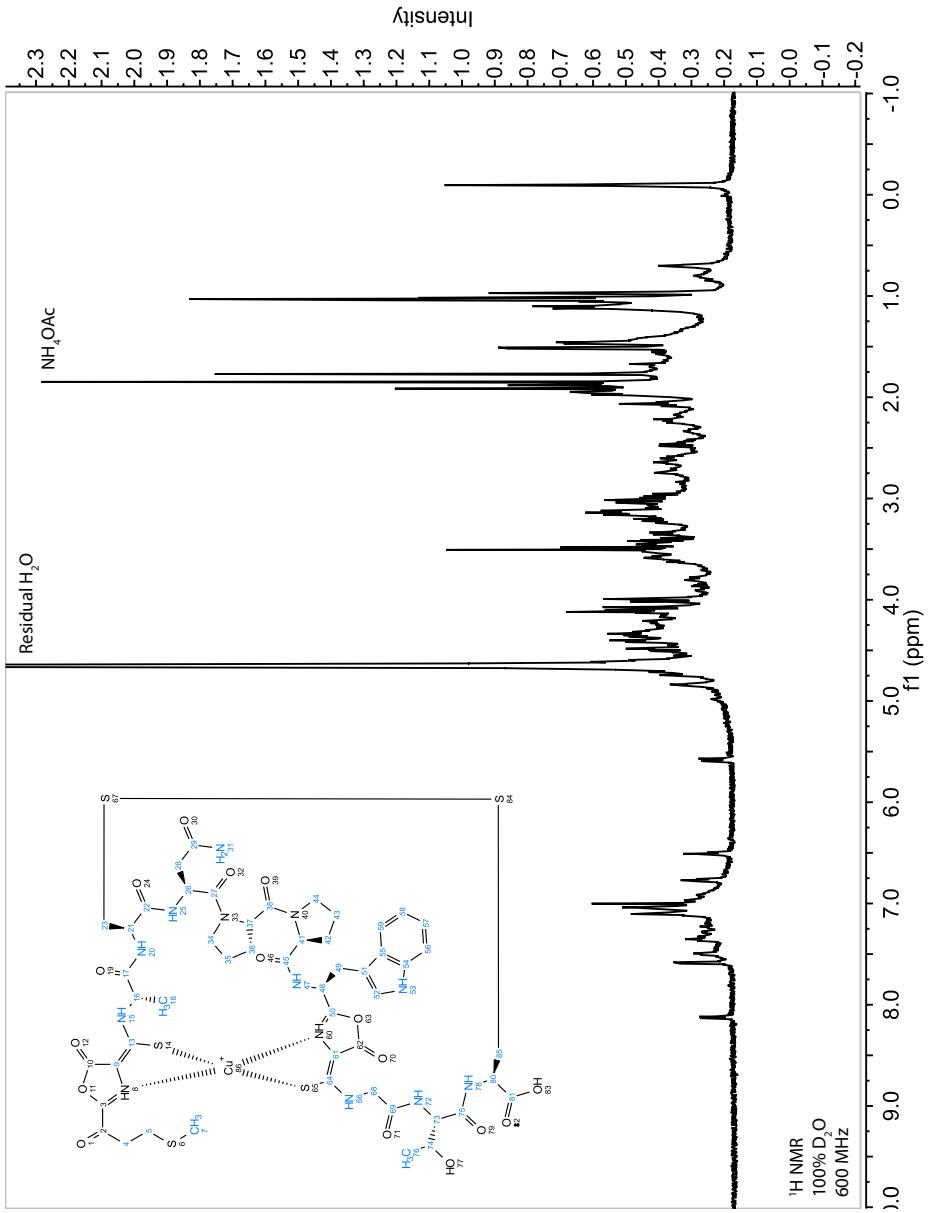

f1 (ppm)
-1
0
1
2
3
4
5
6
7
8
9
10
f2 (ppm)
-1.0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
^{1}H-^{1}H TOCSY
10% D_2O
600 MHz

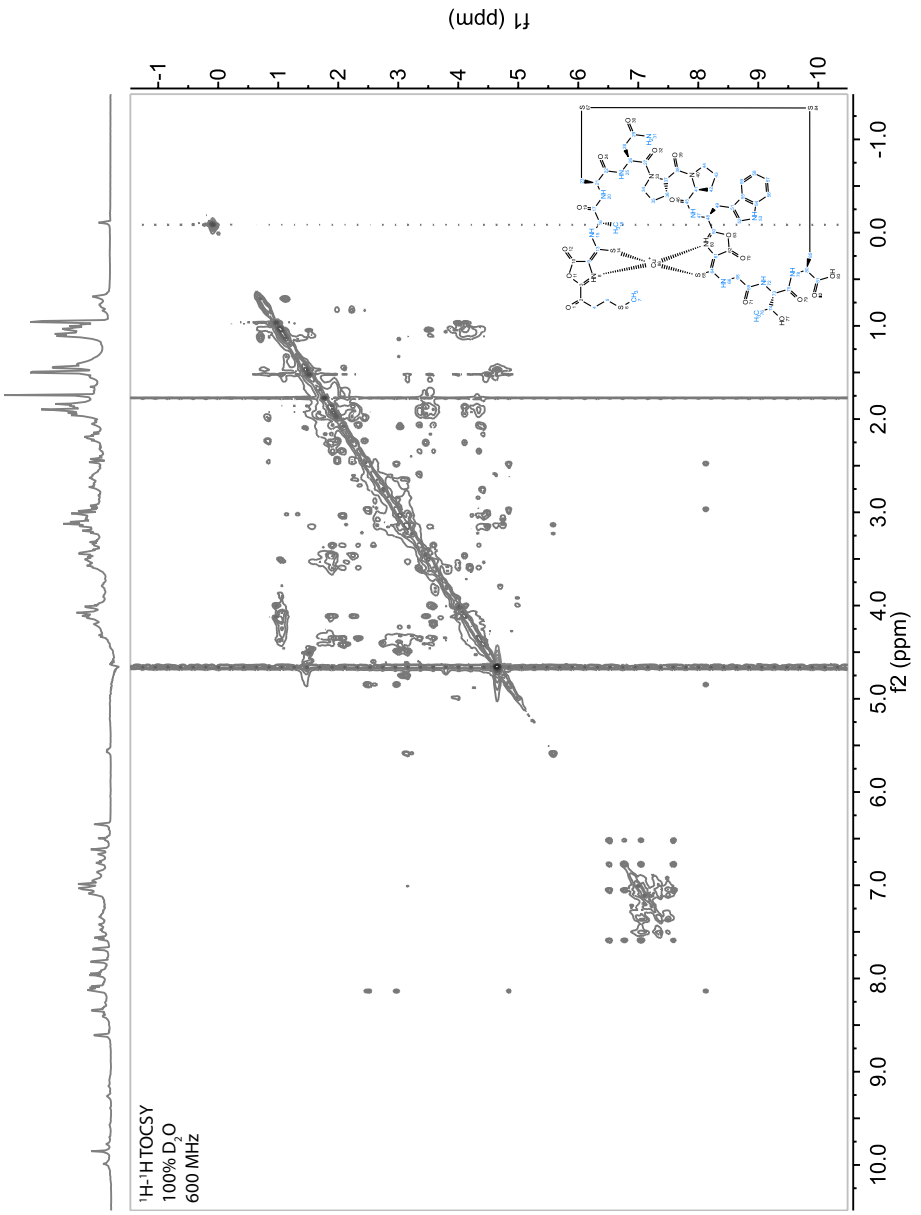

f1 (ppm)
f2 (ppm)
^{1}H-^{13}C HMBC
10% D_2O
600 MHz

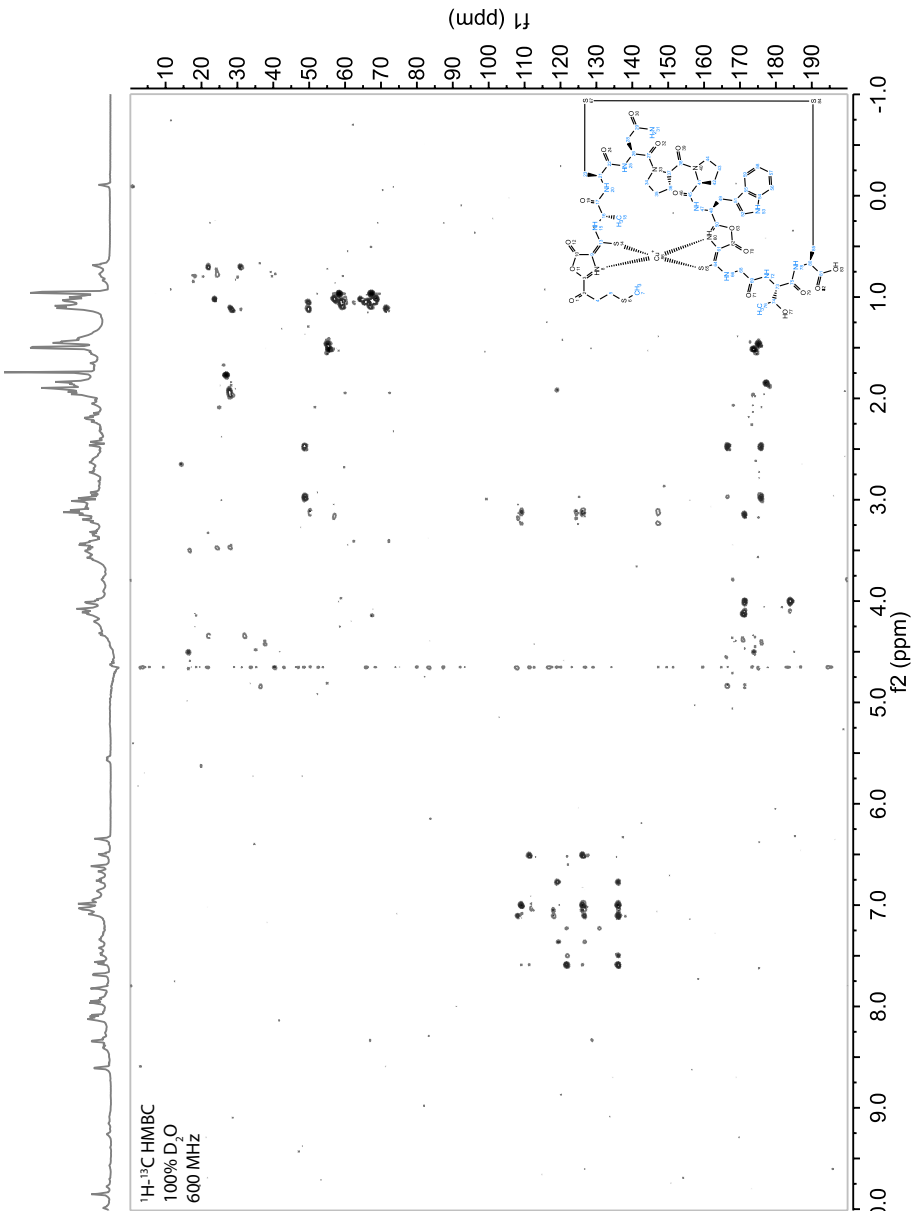

f1 (ppm)
f2 (ppm)
^{1}H-^{1}H ROESY
10% D_2O
600 MHz

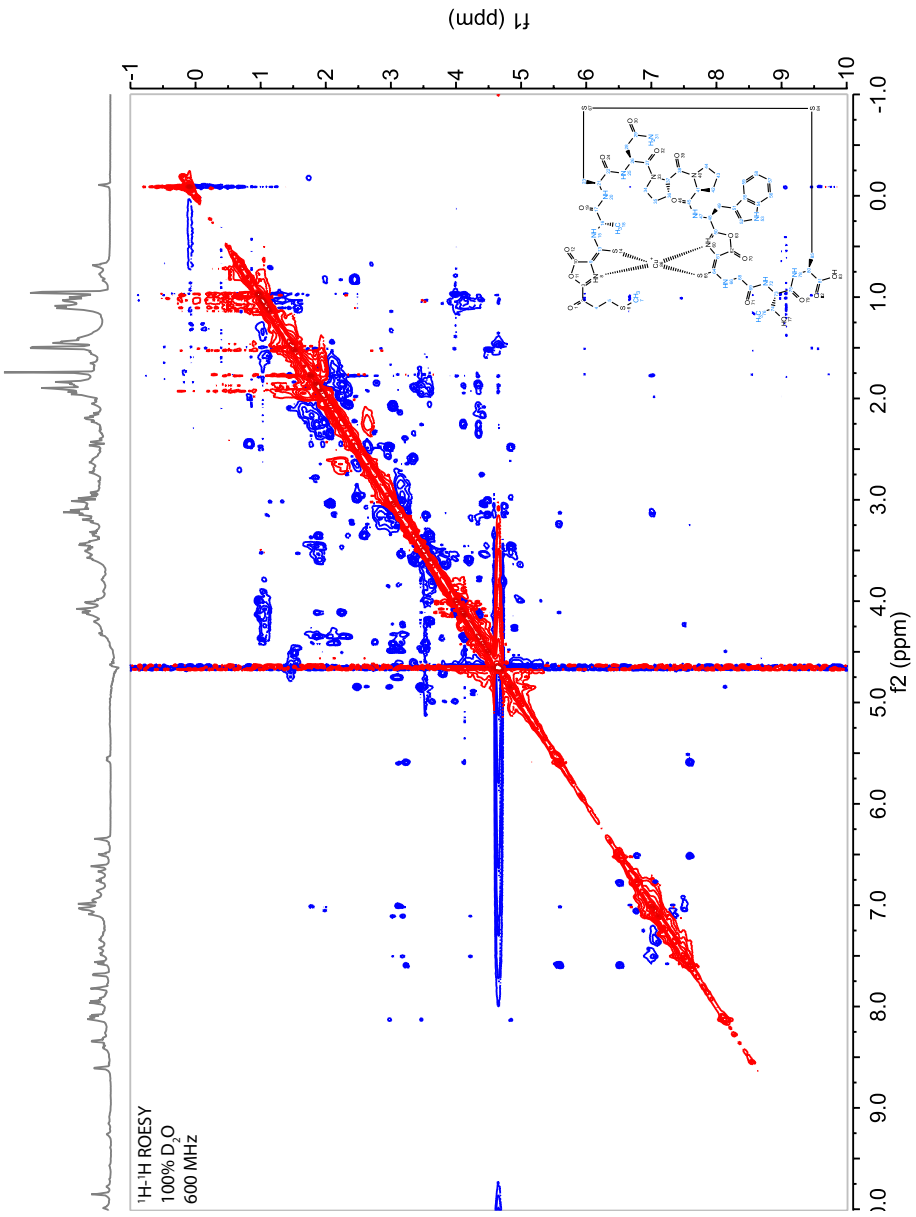

f1 (ppm)
f2 (ppm)
^{1}H-^{13}C HSQC
10% D_2O
600 MHz

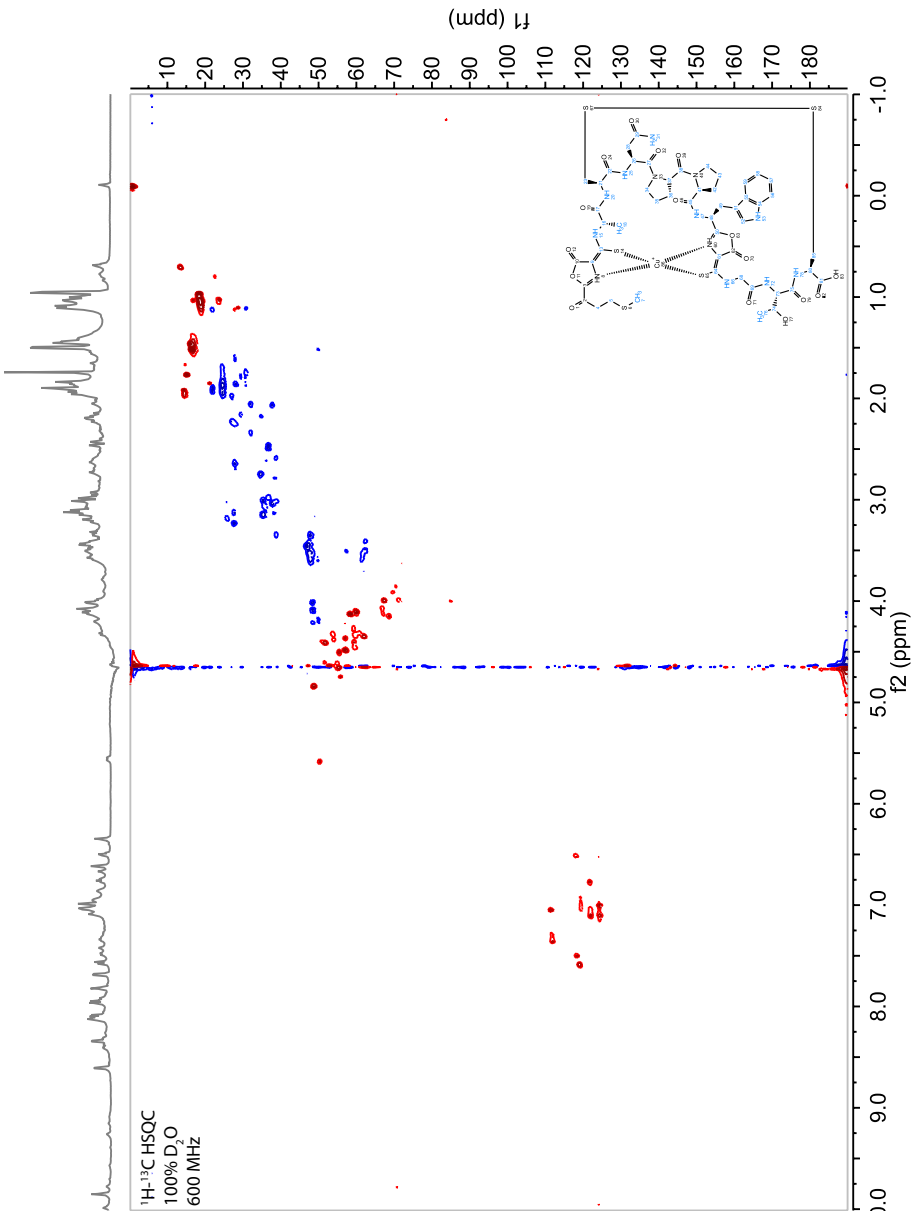

f1 (ppm)
100
105
110
115
120
125
130
135
f2 (ppm)
-2
-1
0
1
2
3
4
5
6
7
8
9
10
11
^{1}H-^{15}N HSQC
10% D_2O
600 MHz


Residual H2O
NH4OAc
Intensity
f1 (ppm)
1H NMR
100% D2O
600 MHz


f1 (ppm)
f2 (ppm)
^{1}H-^{1}H TOCSY
100% D_2O
600 MHz


f1 (ppm)
f2 (ppm)
^{1}H-^{13}C HMBC
100% D_2O
600 MHz


f1 (ppm)
f2 (ppm)
¹H-¹H ROESY
100% D2O
600 MHz


f1 (ppm)
f2 (ppm)
^{1}H-^{13}C HSQC
100% D_2O
600 MHz

Supplemental References

(1) Hakemian, A. S.; Tinberg, C. E.; Kondapalli, K. C.; Telser, J.; Hoffman, B. M.; Stemmler, T. L.; Rosenzweig, A. C. *J. Am. Chem. Soc.* **2005**, *127*, 17142.