SUPPORTING INFORMATION

Combining Mass Spectrometric Metabolic Profiling with Genomic Analysis: A Powerful Approach for Discovering Natural Products in Cyanobacteria

Karin Kleigrewe[†], *Jehad Almaliti*[†], *Isaac Yuheng Tian*[†,‡], *Robin B. Kinnel*[§], *Anton Korobeynikov*[⊥‖∇], *Emily A. Monroe*[†O], *Brendan M. Duggan*[□], *Vincenzo Di Marzo*[−], *David H. Sherman*[◊], *Pieter C. Dorrestein*[□], *Lena Gerwick*[†] *and William H. Gerwick*[*†□]

[†]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, USA;

[‡]University of California Berkeley, USA; §Hamilton College, Clinton, NY, USA;

[⊥]Faculty of Mathematics and Mechanics, Saint Petersburg State University, Russia;

[‖]Center for Algorithmic Biotechnology, Saint Petersburg State University, Russia;

[∇]Algorithmic Biology Laboratory, Saint Petersburg Academic University, Russia;

[O]Department of Biology, William Paterson University of New Jersey, USA;

[−]Institute of Biomolecular Chemistry, National Research Council, Pozzuoli, Italy;

[◊]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan;

[□]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, USA

Table 1: Comparison of the residues in the amino acid binding pockets of the adenylation domains of ColG with the consensus sequence. The number represent the positions of the amino acids of GrsA.[1] The red letters indicate conservative substitutions in ColG and NosA relative to the Stachelhaus consensus sequence.

Position	**235**	**236**	**239**	**278**	**299**	**301**	**322**	**330**
Consensus (Ser)	D	V	W	H	L	S	L	I
ColG	D	V	W	H	I	S	L	I
NosA	D	V	W	H	I	S	L	I

*Table 2: Multiple sequence alignment of N-methyltransferases. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix. A . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix. The sequence information was obtained from http://www.nii.ac.in/~pksdb/sbspks/master.html.*[2]

```
CLUSTAL FORMAT: MUSCLE (3.7) multiple sequence alignment


barba006_N      LSNYDKQPIPEAQMRDWAEDIVTQVLANKPNSVWEVGCGTGMLLFKIAPHTRAYYGTDIS
ColG_NM         -----NQPIPVEQMRIWAGDIVTQVLAQKPESVWEIGCGTGMLLFQIAPQTQNYYGTDIS
tubul002_N      ----TGQPVAVEEMRDWLRHRVERVRGLRPRRILEVGCGTGLMLFALLPHCERYVGTDFS
nodul005_N      -----GLPIAQEQMGQWLNSTVERILALQPERVLEIGSGTGMLLFRIAPQCLRYCGTDIS
anaba002_N      NSSYTGKAIPDSEMREWVESTVSRILLGKPQRVLEIGCGSGLLLFRVAPHCQEYWGADYS
tubul001_N      -----GEPLPPEQMREWVETTVERLMELVPRRVLELGCGSGLLLRRLAPRCESYWGTELS
thaxt001_N      --SYGGRPI--EGMREWREQTVRQIRELAPRRVLEIGCGSGLLLSQLAGDCESYWGTDIS
prist002_M      ------------------------------------------------------------
thaxt002_N      NSTYDGEPIPVPQMQAWRDATVDSIRALRPRRVLEIGVGTGLLLSRLAGDCEAYWATDFS
actin003_T      ------------------------------------GVGTGLLLSRLAPHCEEYWGTDFS
actin003_r      ----------------------------------------------LAPECEEYWGTDLS
compl003_M      ------------------------------------------------------------


barba006_N      EVSLKYIQTQIAQQPD-KYAHVTLAQKAAEEMADIADNSFDVVLLSSIVQYFPSVEYLLQ
ColG_NM         NVSLEYIKQQIEQEPD-KYGDVSLAQKRADNMADIADNSFDVVLLSSIVQYFPSVEYLLQ
tubul002_N      PAALDYVRRYL--PPE-HPGRVELLHRTADEWSGVAAGSFDAVLLNSVVQYFPSQEYLRQ
nodul005_N      DTAIRYVETQMQKVGS-AWSQVQLYNQPAHNLQGFEPKTFDAVIINSVVQYFPSIDYLVS
anaba002_N      SATIRNLERLCG-EIQ-GLENVRLLHKTADNFEGIPQGAFDTVVVNSVVQYFPSVDYLLQ
tubul001_N      PVAVERLREQLQTGGSPLAQRVRLMAQPADDFSGLPEAGFDTVILNSVTQLFPSVDYLLR
thaxt001_N      GALIERLRGQVAERPG-LADRVVLHQLSAHELGSLPSGGFDTVVLNSVIQYFPSGDYLFD
prist002_M      -----------------------------------------------VAQYFPDARYLAG
thaxt002_N      AEVIETLGKKVDVDPV-LREKVHLLHGPAHDLPGLPEGYFDTVVLNSVIQYFPSADYLVS
actin003_T      PTVIADLRGHVEADPE-LAARVQLRTQPAHDFDQLPHGHFDTVVLNSVVQYFPNAGYLEQ
actin003_r      PTVIEALSRHVDADPE-LARRVTLRAGAAHEHEGLPVGHFDTVVLNSVVQYFPNADYLAQ
compl003_M      ---------------------------AADETDGLPEGHFDTVVLNSVVQYFPNADYLRG
                                                               : * **.   **

barba006_N      VISNSIRVVKPGGMIFLGDIRSLPLMRAFHTSVQLHKAP-PSLSVQQLKQGIYRLMQQET
ColG_NM         VIEESIRVVKPGGMIVLGDIRSFPLMRAFHSSVQLY------------------------
tubul002_N      VLARCVEAVEDGGFVFVGDVRSLPLLESFHASVELERAA-PSMPLEAWRERVRRAVLEDN
nodul005_N      VLEGAVEMVAPGGWIFVGDVRSLPLLPAFHADIVLHQSS-HDLPTADWWQRVQKNLQEDQ
anaba002_N      VLEGAMTAIASQGKIFVGDVRSLPLLLPYHAAVQLARAE-SDKTVEQWQQQVHQTVAAEE
tubul001_N      VVEGALRVLQPGGTLFIGDVQNLRLFELFHASVALEQAS-ADLEAPALLARTRQRMLLDE
thaxt001_N      LLREVSRLLVPGGAVFLGDVRNLRLLRTFHAGGLLAAAT-HTDTPQTVCAAIDRAMAQEK
prist002_M      ILHRAAELLAPGGTIFLGD-----------------------------------------
thaxt002_N      VLREAARLLAPGGRVFVGDIRHLRLLRPLRSAVRLRSATRREASASAVRAAVEQDLVDEK
actin003_T      VLDHALRILAPGGTVFI-------------------------------------------
actin003_r      VIEQALRLLAPGGAVFI-------------------------------------------
compl003_M      VLEWALRLVAPGGAVFV-------------------------------------------
                ::      :   * :.:

barba006_N      ELLVSPELFVALKDTYPEITHVQIRLQRGSEHNELNKYRYSVLLHIQAKPT-------
ColG_NM         ----------------------------------------------------------
tubul002_N      ELVVDPALFVALAHQHPRVSHVDIELTRGTHPNEMARFRYNAVLHIGPRTP-------
nodul005_N      ELVIDPAFFTALMQHLPQIRRVQIQLKRGRDRNELTRFRYDVILHIETEVVPPIESQD
anaba002_N      ELLIDPRFFIALQQRFPQITWVEIQPKRGHAQNELTQFRYDVTLHLVLMWGKGSSLVK
tubul001_N      RLYVDPDFFAALATHFPQLGAVRLHLKRGSGRN-------------------------
thaxt001_N      ELLVDPEFFTTAVGALPGMTLESCTLKRGG----------------------------
prist002_M      ----------------------------------------------------------
thaxt002_N      ELLLDPAFFAAVPRWIPQLRGVRTAVQRGTHHNELTRYRYDAVLIKEPVETGTAAPDA
actin003_T      ----------------------------------------------------------
actin003_r      ----------------------------------------------------------
compl003_M      ----------------------------------------------------------
```

*Table 3: Multiple sequence alignment of O-methyltransferases. An * (asterisk) indicates positions which have a single, fully conserved residue. A : (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix. A . (period) indicates conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix. The sequence information was obtained from http://www.nii.ac.in/~pksdb/sbspks/master.html.*[2]

```
CLUSTAL FORMAT: MUSCLE (3.7) multiple sequence alignment


ColG_OM         ----------------------------DVGANIGMFSL--FASQQVKDL----EIFAFE
KtzH_OM         --------------------------VVDVGGHVGLFSL--FVKTRRPDC----RIYAFE
stigm_004       -------------------------LNEKQPGFSWLRVA--YGLDPSEER----MRLLLH
stigm_005       ----------------------WGVFQEIVPGFSWIRTV--FRPSERPEG----RERLAV
melit004_O      ------------------VYLTFGIMRRPVPGFSWLLNV--YGMSETPEH----KDVLLD
barba005_O      ------------------DYARFAPFSEVVPDFSWLEVI--LNPDQHPEQ----TSLSLQ
melit005_O      ----------------------------------WISL---YRPEPQEQSYRAYYDIALK
onnam001_O      ------------------------SSMTLVEGIYKNNLVSDYFNQVLGDV----------
onnam002_O      ---------------------------EKIEGLYKNNHVCDYFNQVVAEV----------
peder001_O      ---------------------------EKIEGLYKNNLICDYFNDVVAGV----------
nodul002_O      ---------QVSELYTGIAKRDKDLNNPNIPQ--WLNFG--YWQEETTYN---------G
onnam005_O      --MSRSHLEEIAELYDS---AEGHVGNLIFDGQ--VHWG--YWDERNADA------SLAE
peder004_O      MQTAIADVEKVATLYDS---AEGQVGPILFGGH--MHWG--YWDEVTGEG------NFAN
anaba003_O      --------------FDSLIFNT--STRDYYGEKEFFNVG--YWHSDTQNQ--------HE
peder003_O      ---------HINQHYDHTFFSE-GLTSLLVDGSDYRNIG--YWDETTTTQ--------HE
onnam003_O      -------------HYDRFFYEQHGVERLIREETDFKNLG--YWDDTTLDL--------NA
onnam004_O      --ARIPLTDEINSFYDHQFYSQDSIFGLLLGDTKFRNIG--YWDETTPDQ--------NA


ColG_OM         PIPPVFKVLEMNTELYIS-KVKLFECGLSNQTRMETFTYYP-ENSVVSGLYADQNQEQEM
KtzH_OM         PIPELAEMFRINAELHDI-DAVVTNCGVGATAGTARFTYYP-DMSMLSGRFADEREERRM
stigm_004       SQRALRNVLLDSVDFSRAKSVWDFGCGYASDIIA-LGERHS-HLKLHGHTLSSEQAELGL
stigm_005       AQRELRRVLFRAVDLSAIKNVMDFGCGHGSDLII-LGEQNE-HLKLDGYTISGKQAEVCK
melit004_O      GQRGLRSVLFGGVRWESVRKVLDFGCGYASDLLS-LARRHP-HLKLHGYTISAEQAAVDA
barba005_O      AQAEMKAVLFRGIDFSSIKKVMDIGCGYSHDLID-LATNHV-HLQLDGYNISPEQVKAGE
melit005_O      ANQEMARILYRGVDFSTRTRVLDIGCGHAADLVD-LARAHP-HLELHGCNISPDQIEVGR
onnam001_O      ----LVAFMTTRGHEEPV-RILEIGAGTGGTTATLLEKLRPFQEQIAAYCYTDVSKAFLF
onnam002_O      -VQTYIHQRLAVNPKATI-RILEIGAGTGGTTSMVLPALRPFQDHIDTYSYTDLSKSFFI
peder001_O      -AQAYIQRRLENEPNAEI-RLLEVGAGTGGTTSTVLPQLNLWRAFIAEYAYTDLSKSFFN
nodul002_O      ACAALARKLGEVAELSPGEQVLDVGFGFAEQDIL-WMRENN-LGAITGINTTELQVKIAQ
onnam005_O      GADRLTQIMIDKTTIEKGQKFCDLGCGWGGPAVA-LAKAKG--CYIDGITCSGQQQQNAV
peder004_O      AAERLAQIMIAKAPIKAGQKFIDMGCGFGESALK-LAKAKG--CFVDGITISKEQQLSAI
anaba003_O      ACFNLMEKLLEFIPRKQG-NILDVGCGLGATTSH-LLNYYS-PADVVGINISRKQIERSI
peder003_O      ASERLQDALLDFIPEKSG-RILDAACGMGASTRH-LLEYYP-ADNIWAINISEKQIEATR
onnam003_O      AAERLFKTLMAMIPKKSG-RILDAGCGTGGATRR-LLESYP-PENVWAINISAKQIETTK
onnam004_O      AAEKLQDMLLEMIPEKTG-RILDVACGMGASTRR-LAELYS-PENVWAINISEKQIESTR
                                          * .                ::     :


ColG_OM         -----------MKTFLSNKQKETGE------KSKLSSQELEQVSSYMFQTQQQVNCQLRT
KtzH_OM         -----------LERVLRNERLADLD------DGVLDELLAERL------RGQQVDVELRT
stigm_004       RKIEARGLGGRVQVLRRDSSKDAPL------ESAYDVILGFEVATHIKEKRSLFQNLSSH
stigm_005       QRVRTRGLQNRIRIFQRDSAKDDFP-------GMYDLVLGFEVAGLIPDKDALFSNIDRH
melit004_O      RRVRERGFEDRIRVFARDSAKDAFP-------DRYDVAFGFEVATHIADKDALFSNLARS
barba005_O      QKIQGLGYSDRIYLYNRDSAKQPLP-------DTYDLIFSCQVIHHIKRKEDVFLNISQH
melit005_O      QRIRALGLDGRVLLHYQDSSRDQFP-------STYDLVIAYQVIHHIRAKSDLFANISRS
onnam001_O      HAEEHFAPEHPFIGTAIFDVEQPLANGVIK-PADYDIVIATNVLHATKNICETLRNAKAA
onnam002_O      HAKERYGTAYPFVEYKILNIEKPLAKQDVTLLGSYDIAIATNVLHATKSMRNTIRNVKAA
peder001_O      HARLRYGTDYPYITYRLLNIEEPLIQQDIE-IGTYDILIATNVLHATRNMRNTLRNAKAA
nodul002_O      ERVARAGLEERINLQVGSATKIPFA------ENSFDKVTALECAFHFNTREDFFAEAFRV
onnam005_O      KKAQELGMDDLLNFIHGDALNMPCK------DQTYDGGWFFESIFHMGHREALL-EANRI
peder004_O      TRAEAEQLQERVRFIHGSALNIPCE------DQSYDGGWFFESIFHMGHRKA-LHEAARV
anaba003_O      VNAPG------CKFICMDAVQMEFE------DDFFDNIICVEAAFYFNTREKFLKEAMRV
peder003_O      RNVPG------CHAQVMNAVDLSFE------EGFFDNILCIEAAFHFETRQKFLEEARRI
onnam003_O      QNVKG------CHAIVMNAVDMTFE------DNFFDTVLSIEAAMHFETRRKFLEESFRV
onnam004_O      ENAKG------CHVQVMSAVEMTFD------NDFFDTIMCIEAAFHFETRRKFFDDSLRV
                                                    .                 .
```

```
ColG_OM         L--------SEVIREQGVEQIDLLKIDVEKSELEVLE------GIESEDWSKIKQ----I
KtzH_OM         L--------SDLIREQGIDRIDLLKIDAEKSELDVVR------GIEPEHWAIV-------
stigm_004       LREGGFMLLADFIANS-GSGVDVQDIASYNV--------------TPSQWVELLSEHGLR
stigm_005       LTNGGLLIMADFVANT-LSPIEVQETSTFSS--------------TREQWNKLFSSNHLR
melit004_O      LNNGGFLLLADFIAAG-VSAINIEETASYNS--------------SAEEWADVLSRHNFR
barba005_O      LNDSGFFVAAEIISNLPLTPIDDAKSTAYYV--------------TRSKWAQLLARNNLK
melit005_O      MKPGGLLIMAETMSNM-VSPIEHPESTTQFV--------------PVGEWAELLARNHLR
onnam001_O      LKQHGLMLLNELSDQSLFAHLTFGLLEGWWRHEDASIRIPGSPGLFPEAWQSVLEREGFT
onnam002_O      LARNGIAIINEMTTKTVFATVLFGLIDGWSLSEDTVLRIPGSPGLYAETWHQLLEEEGFR
peder001_O      LRGNGILILNEISDKTIFASVLFGLIDGWSLAEDEHWRIPGSPGLFAENWQALLLQEGFD
nodul002_O      LRPGGKLALADCL------PRVGRDINFWLRVNSKKMCIPFVNQYDRNTYVEKLKKQGFV
onnam005_O      LKLGATLLITDAYLLS-TASEDFKEHTSRRVHSRFM---------PKDIYPGVLEETGFE
peder004_O      LKPGSTLLLTDLPLLP-ESTEAFKEF-VWEHIHSRFV--------SREDYPELLAEAEFE
anaba003_O      LKPGGNLILADLIFDT---TKYFGDLIVPENIVKDK---------DIEDYKRLYQQAGFQ
peder003_O      LRPGGRLVLSDVLFSS---SERLEQYPIFPSAINHLN--------DTEEYRRLLKDTGFS
onnam003_O      LKQDGCLVLSDILFTS---QERLEQNDYFGGVSNHIE--------TIEDYQQLMEEIGFR
onnam004_O      LKQGGRLVLSDTLFTS---KERLEQSSIFPSPENHID--------TLEEYRQVMEEAGFR
                :         :                                      :

ColG_OM         VVEVHDINGRL-------AAVEELLKAQGYQL----------------------------
KtzH_OM         ---------------------RQVVAEVH-------------------------------
stigm_004       LVECVDVSQEVANF-LFDADFDANLTQLETSVGISAIEKRNYQAMRNFGAALERKILSYV
stigm_005       LVDAVDVSNEVANC-LHNPDYAAQFEALCKELKLDEVTQRSFGSYENVYKALRGGLISYV
melit004_O      LVEGVDISREASLF-LEDPSFDQNLERVTERFKLNELV----------------------
barba005_O      VVEGVDASLGIANY-LYDPNF---------------------------------------
melit005_O      VVECVDATQEI-------ANFLHDAE----------------------------------
onnam001_O      SVCFP----------------ARAAHQLGQQIIVAESN----------------------
onnam002_O      SILFP----------------AHPARELGQQVIVSESN----------------------
peder001_O      KVSFP----------------AQVAHDLGQQIIVAQTN----------------------
nodul002_O      NIQAIPIGEYV--WPAVVHYFAQVGQGISKHDLVINLQKDNPGLEAWSRDRGWFMAFDDY
onnam005_O      AVEVLDVTQYV--MRPLAQKLKDACVAYREEILKLVPE---EAIDDWLWGFEDFCANLGY
peder004_O      LIEIDDITDNVMPW--LEPKLKEAIELHRPQVEAIIPNDTEKAIDDWLYLFEYMSENLGY
anaba003_O      PIEFVEATEVC--W---KIHYRDLKSSIIEEFNTGKIDEETYNFNVVAIDALLDSSSIDY
peder003_O      QVEIEDVSDEV--W---GAHFIYAVKRVHEAFYKGE------------------------
onnam003_O      NVVVKDVSKAV--W---GSNFLYNINKLHKEFYHGR------------------------
onnam004_O      NIVVKDVSKNV--W---EAHFLYVINKIHEGFYHGR------------------------


ColG_OM         -------------------------------
KtzH_OM         -------------------------------
stigm_004       LFIAQKDSHVRSTYLRHINQKWVEAPAPYAA
stigm_005       LFHVQKDRFSRSDELFHLNAKQFEQLTP---
melit004_O      -------------------------------
barba005_O      -------------------------------
melit005_O      -------------------------------
onnam001_O      -------------------------------
onnam002_O      -------------------------------
peder001_O      -------------------------------
nodul002_O      ILFSGEKP-----------------------
onnam005_O      LLVTARKK-----------------------
peder004_O      MIVMAKKL-----------------------
anaba003_O      LLVSVKKP-----------------------
peder003_O      -------------------------------
onnam003_O      -------------------------------
onnam004_O      -------------------------------
```

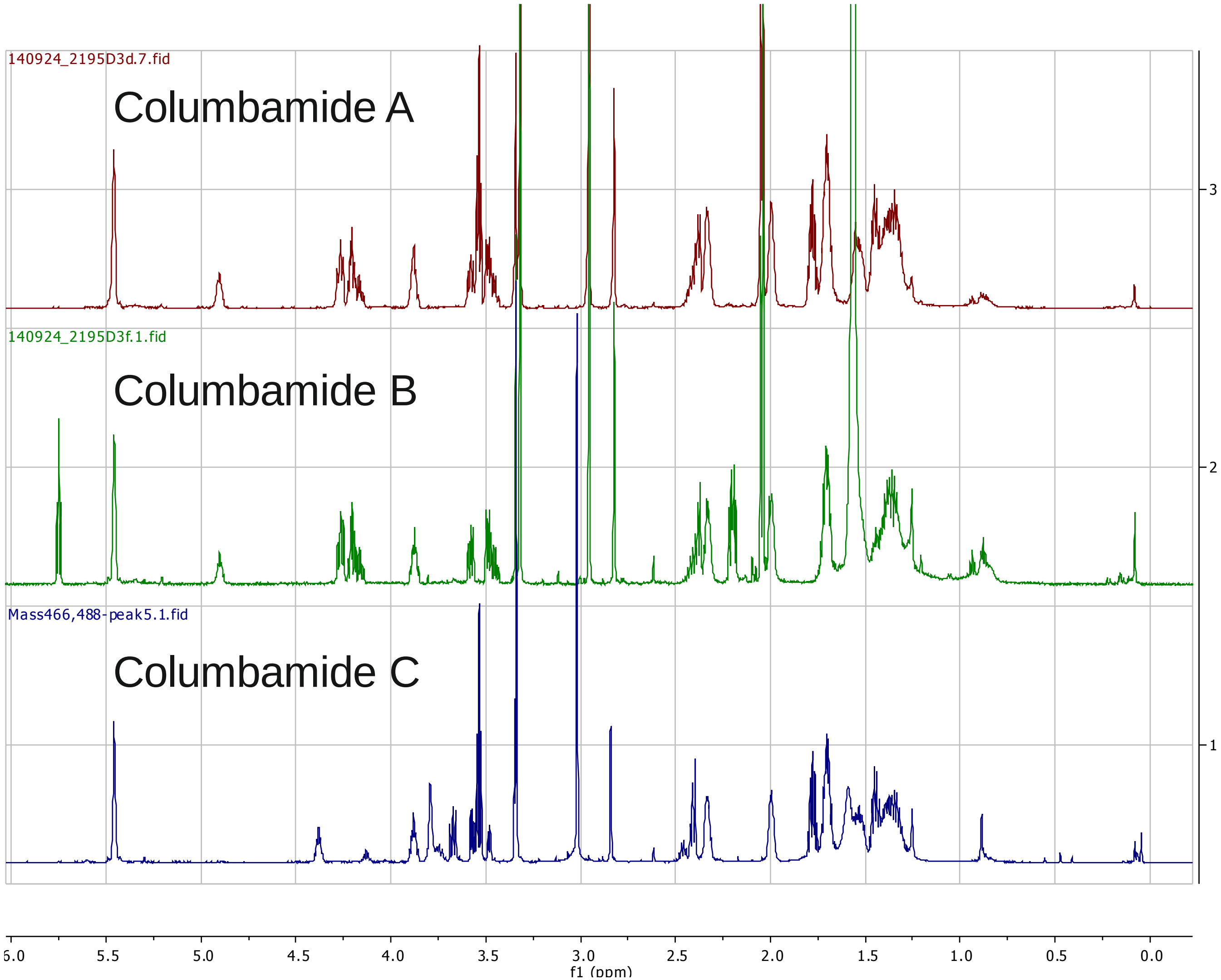

Figure S 1:Stacked spectra of ^{1}H-NMR of Columbamide A, B and C (recorded on a 600 MHz NMR with cryoplatform in $CDCl_3$).

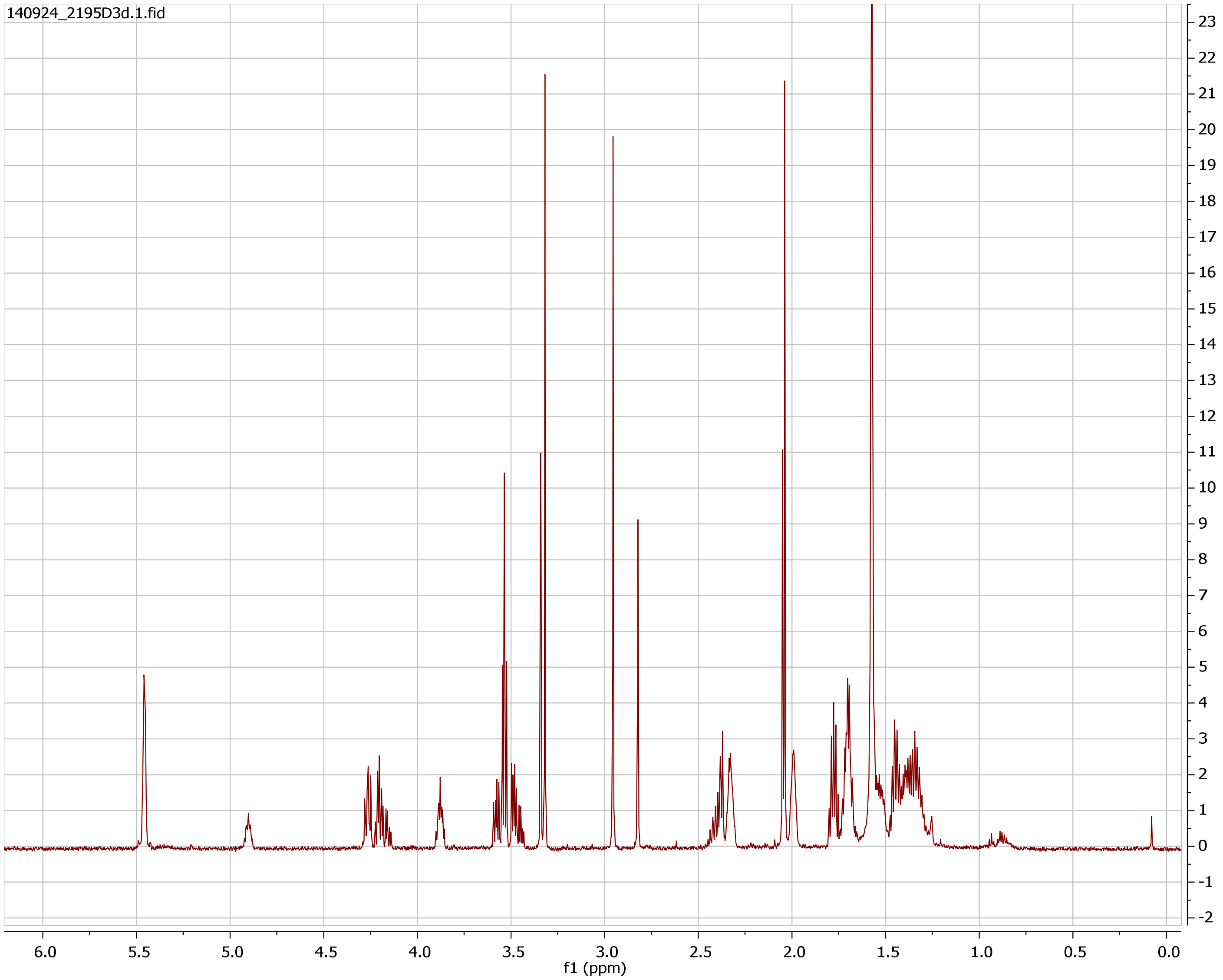

Figure S 2: ^{1}H NMR (600 MHz, CDCl$_3$) spectrum of columbamide A.

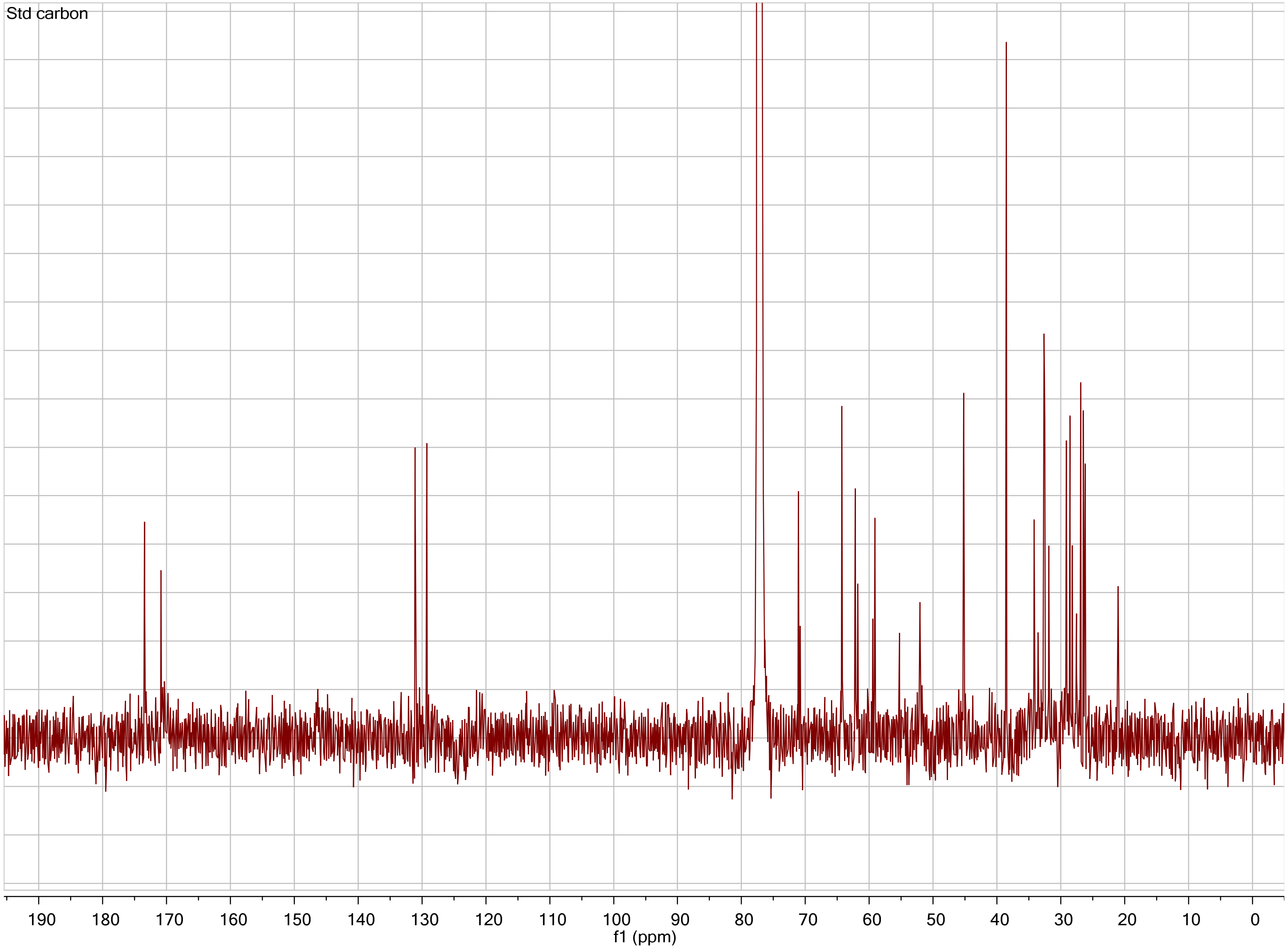

Figure S 3: ^{13}C NMR (125 MHz, CDCl$_3$) spectrum of columbamide A.

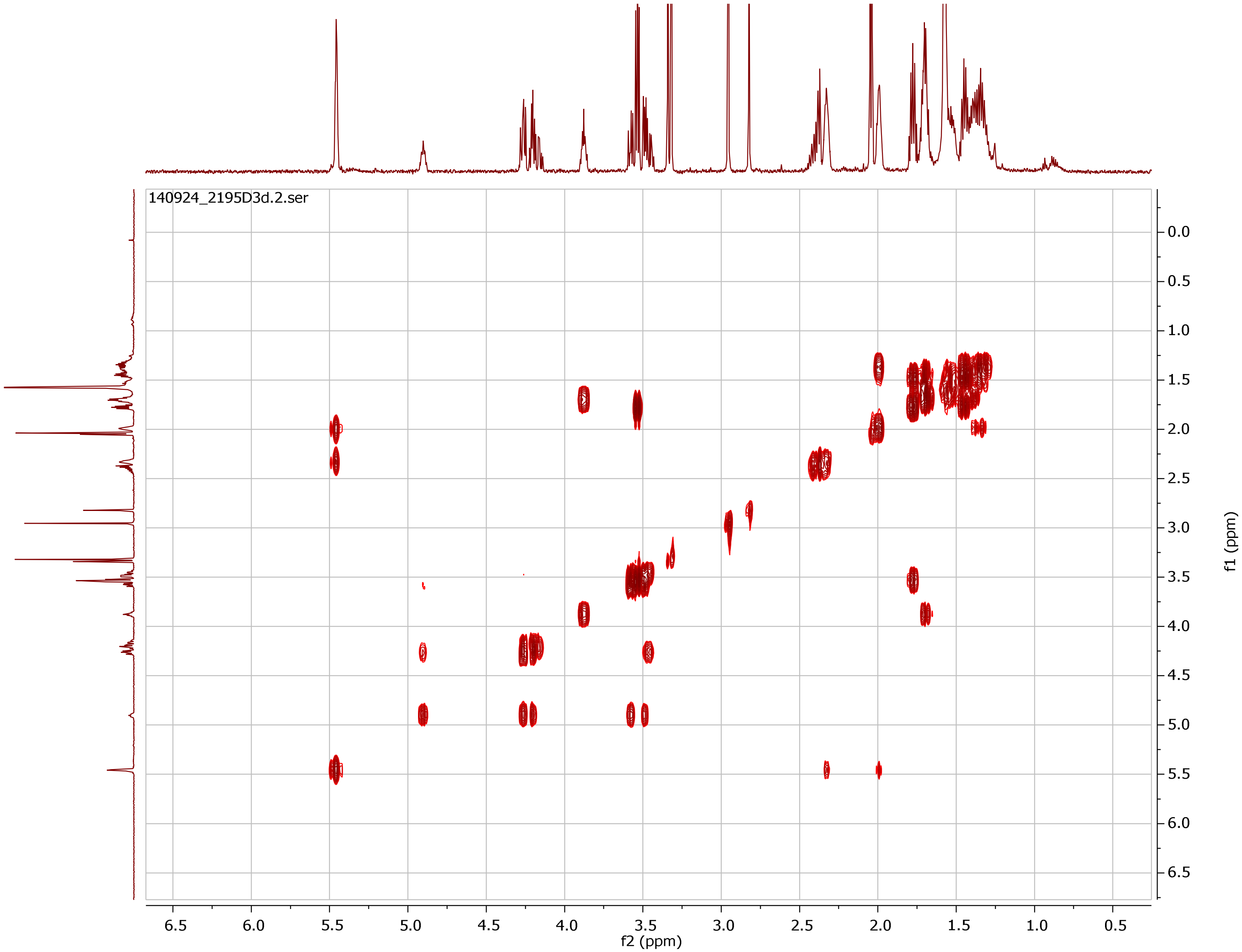

Figure S 4: COSY (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide A.

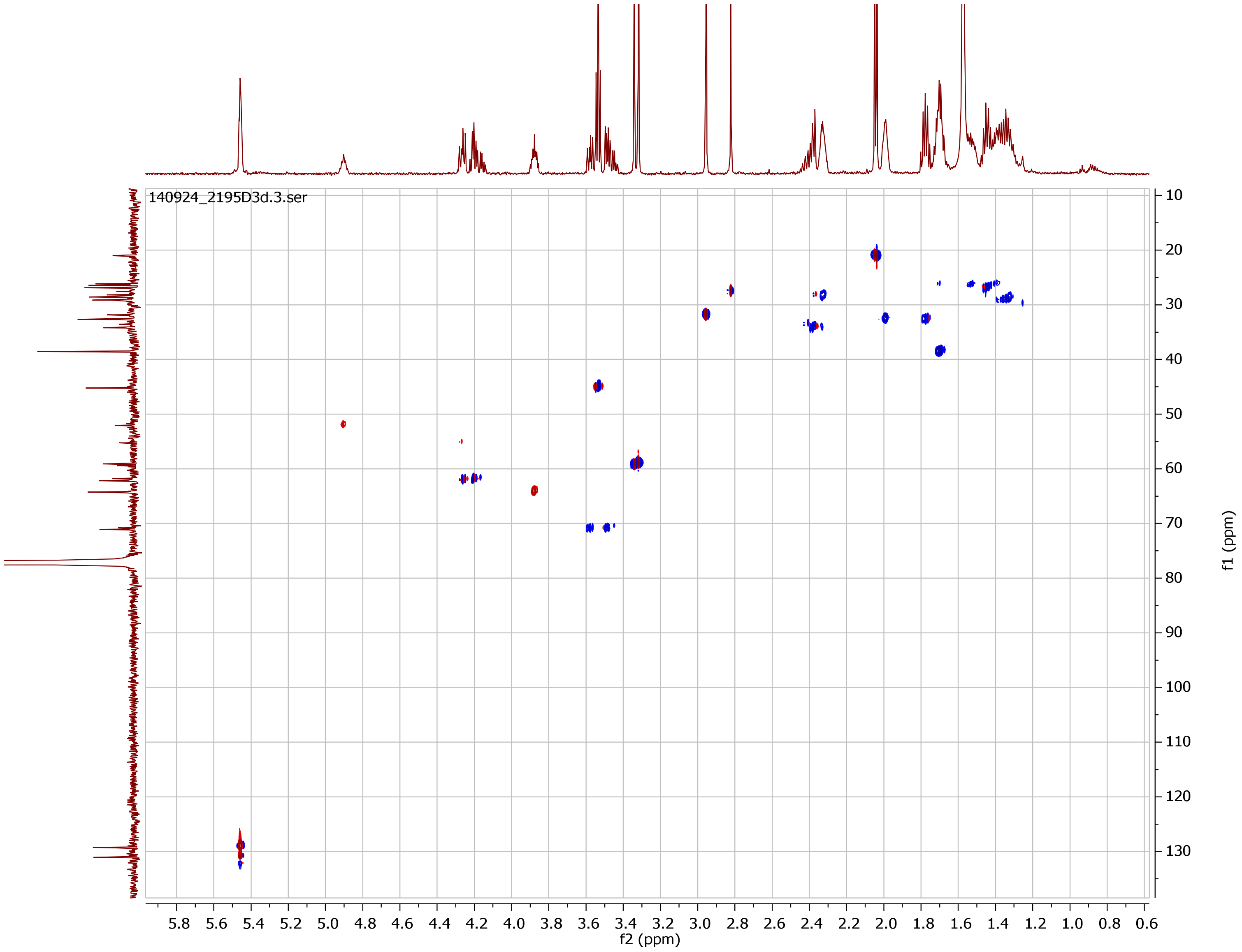

Figure S 5: HSQC (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide A.

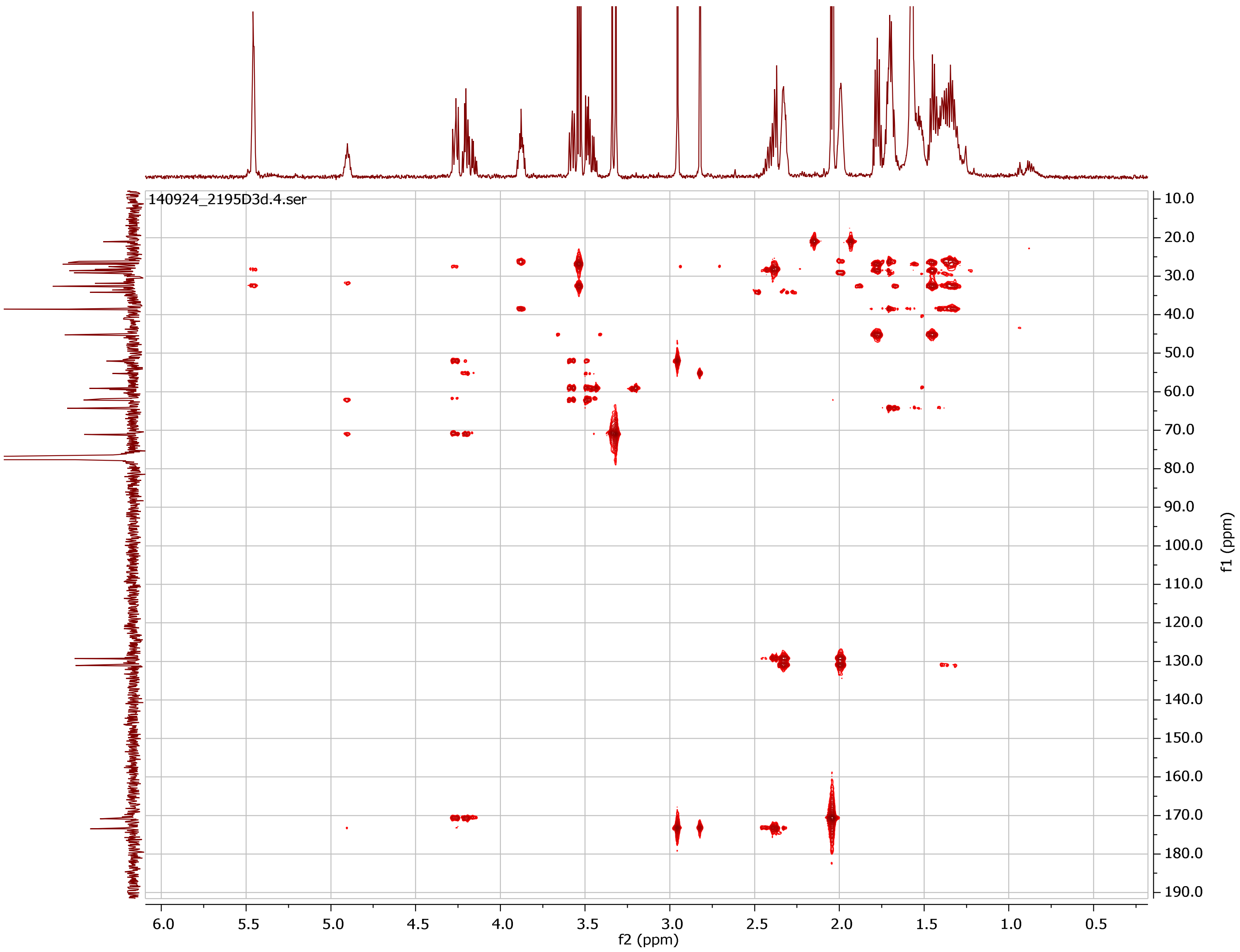

Figure S 6: HMBC (^1H 600 MHz, $CDCl_3$) spectrum of columbamide A.

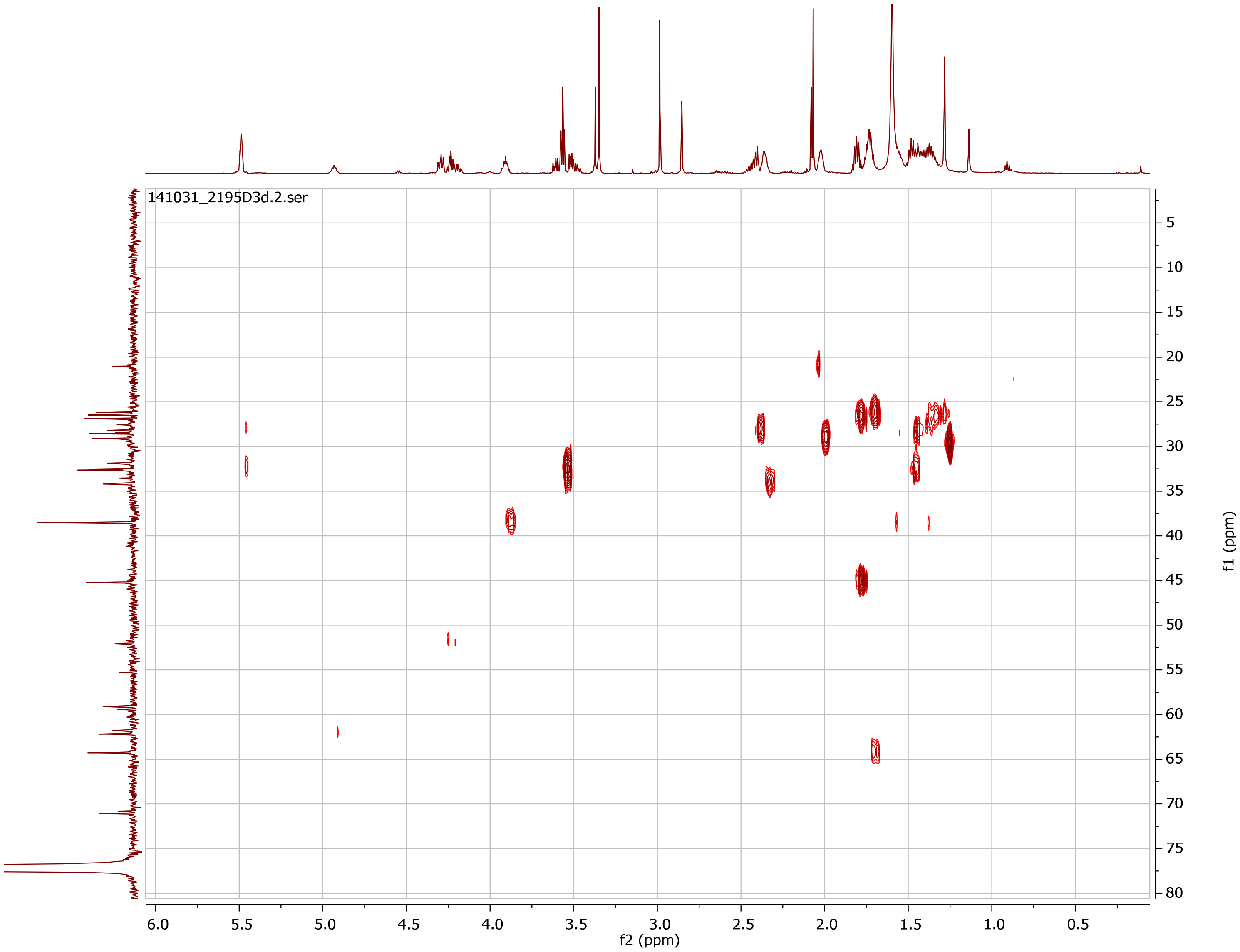

Figure S 7: H2BC (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide A

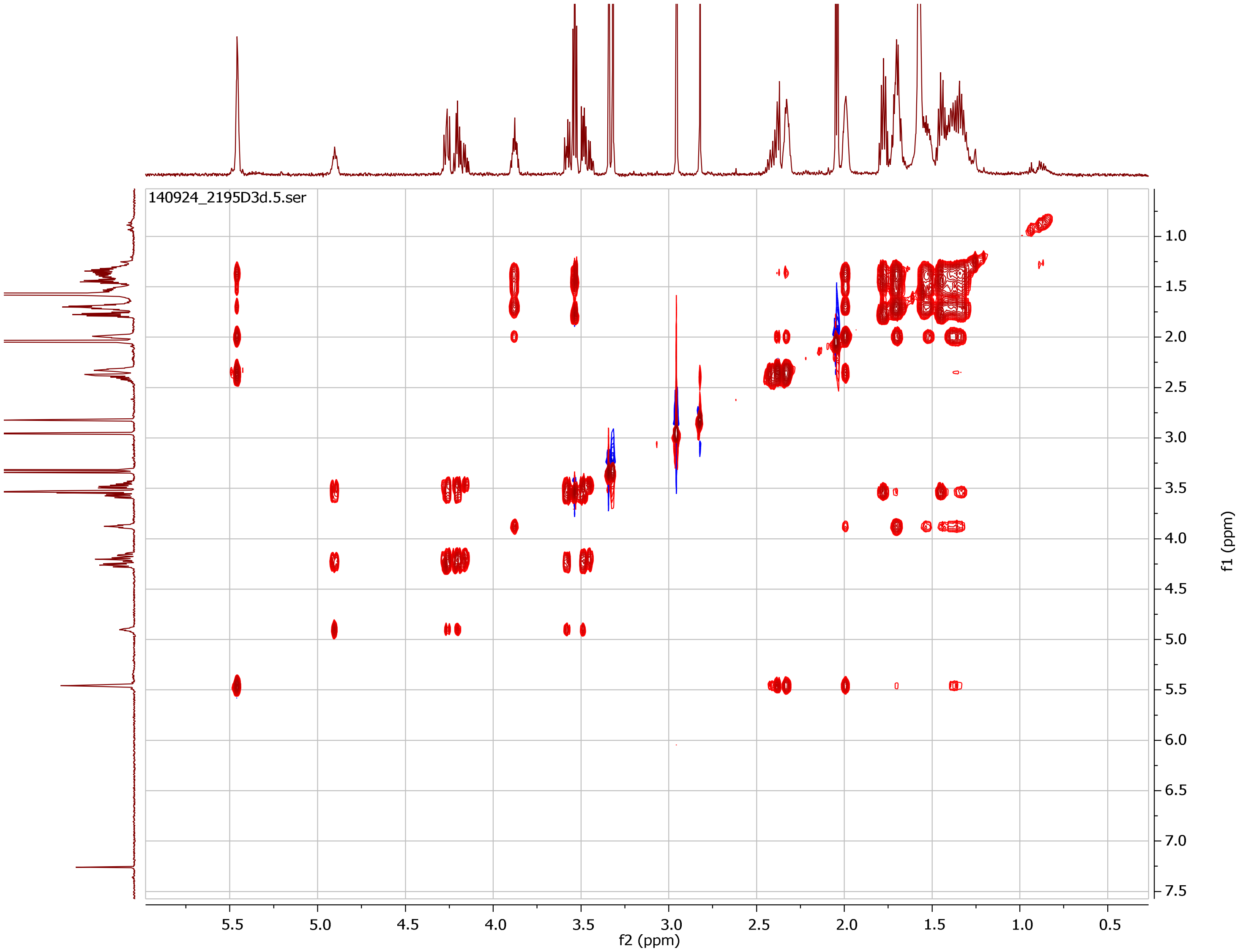

Figure S 8: TOCSY (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide A.

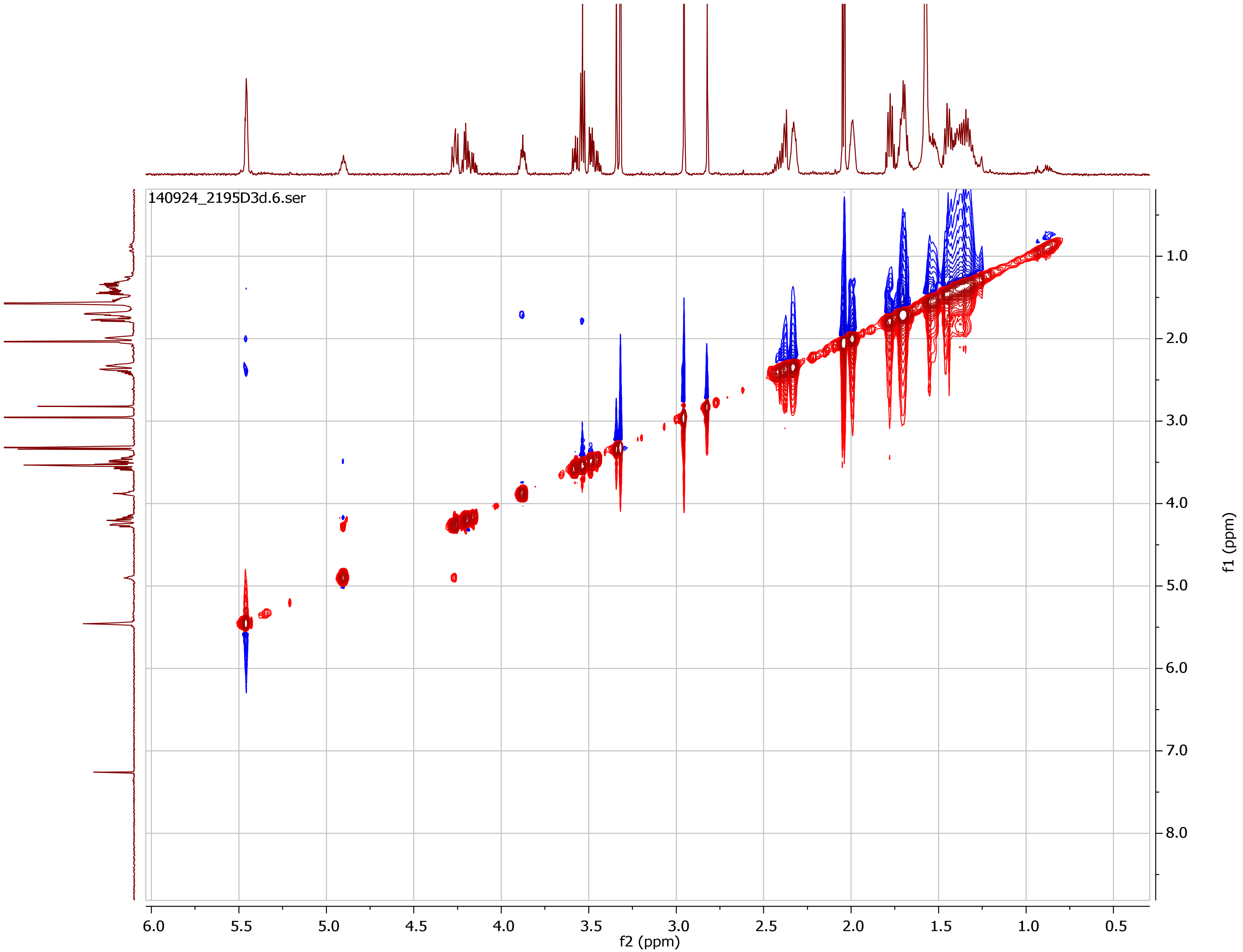

Figure S 9: NOESY (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide A.

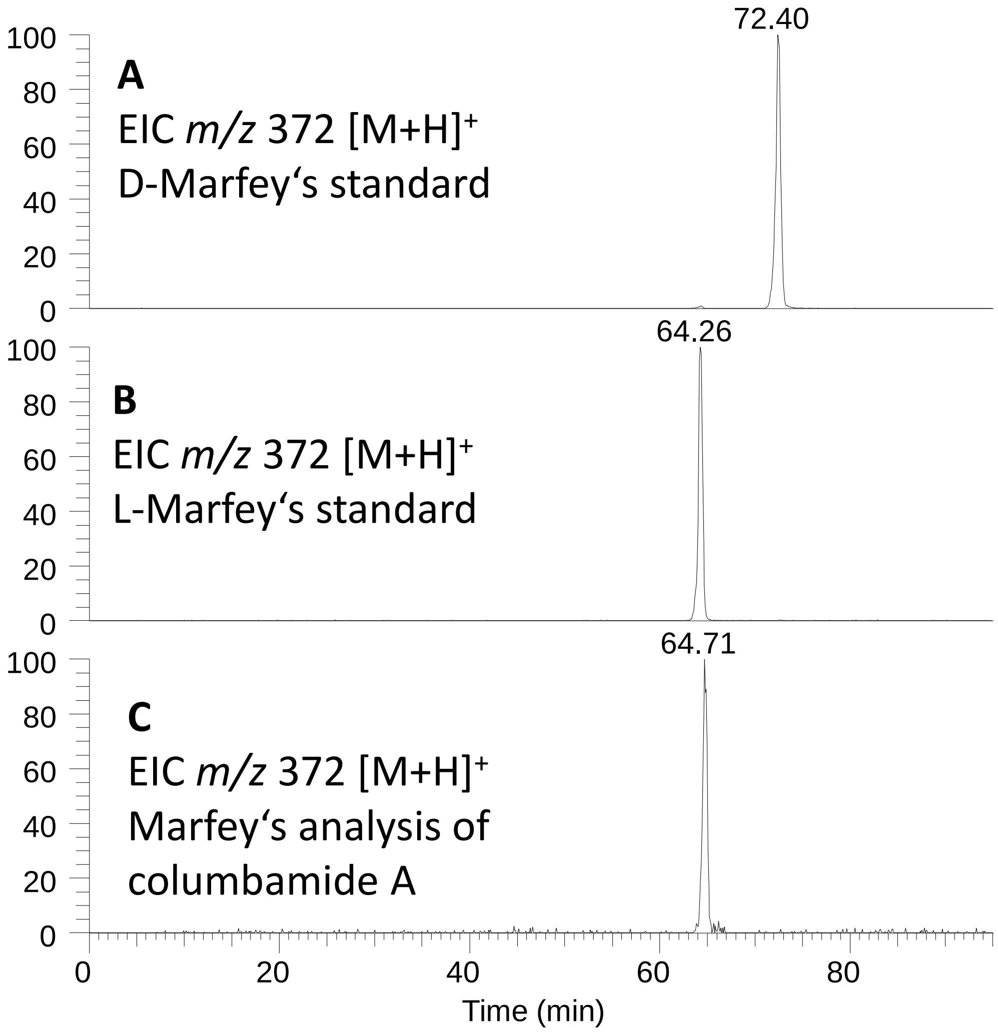

Figure S 10: Marfey's analysis of the dimethylated-serinol in columbamide A. ***A*** *shows the extracted ion chromatogram of the D-Marfey's standard and* ***B*** *shows the L-Marfey's standard.* ***C*** *depicts the extracted ion chromatogram of hydrolyzed and Marfey's derivatized columbamide A.*

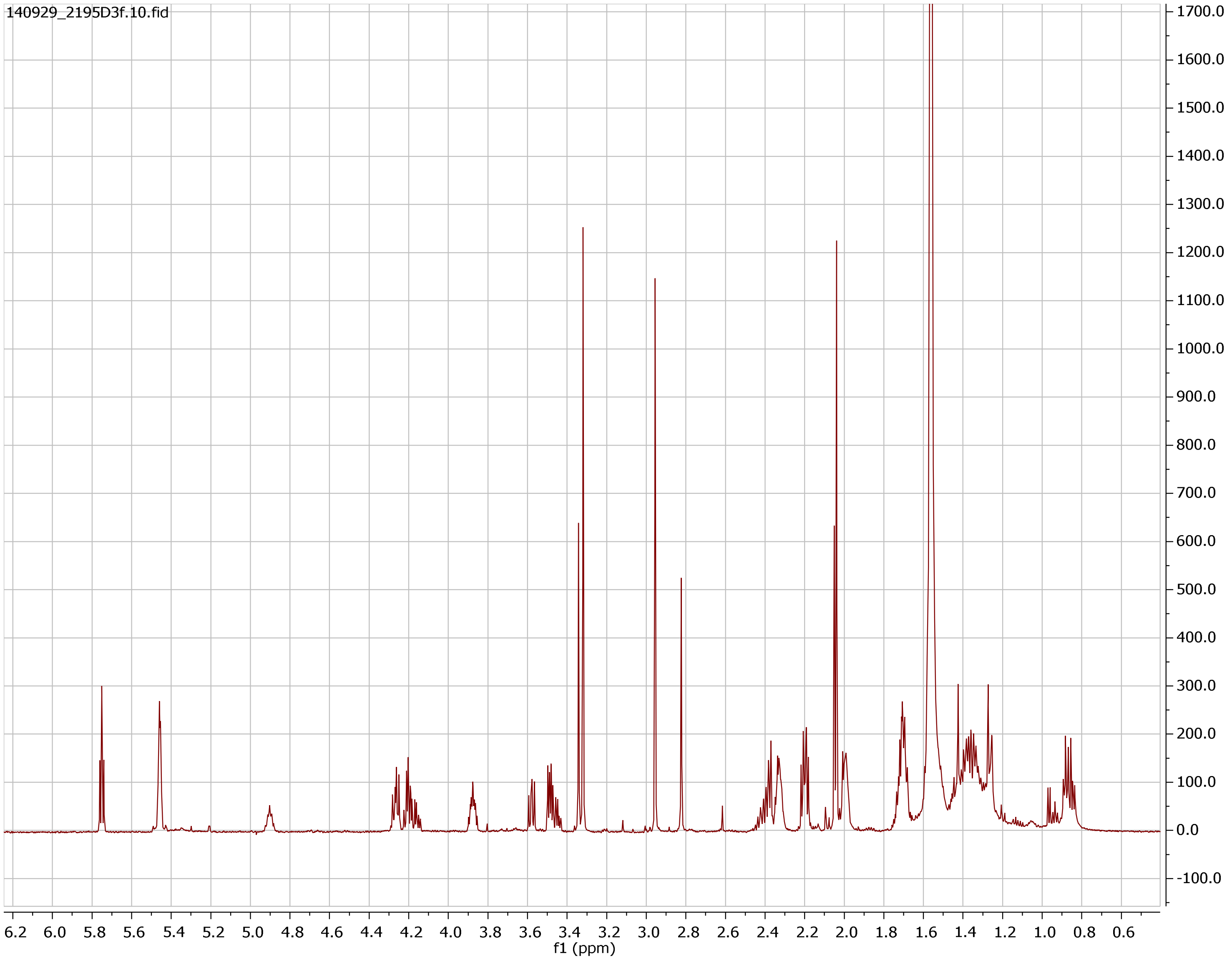

Figure S 11: ^{1}H NMR (600 MHz, $CDCl_3$) spectrum of columbamide B.

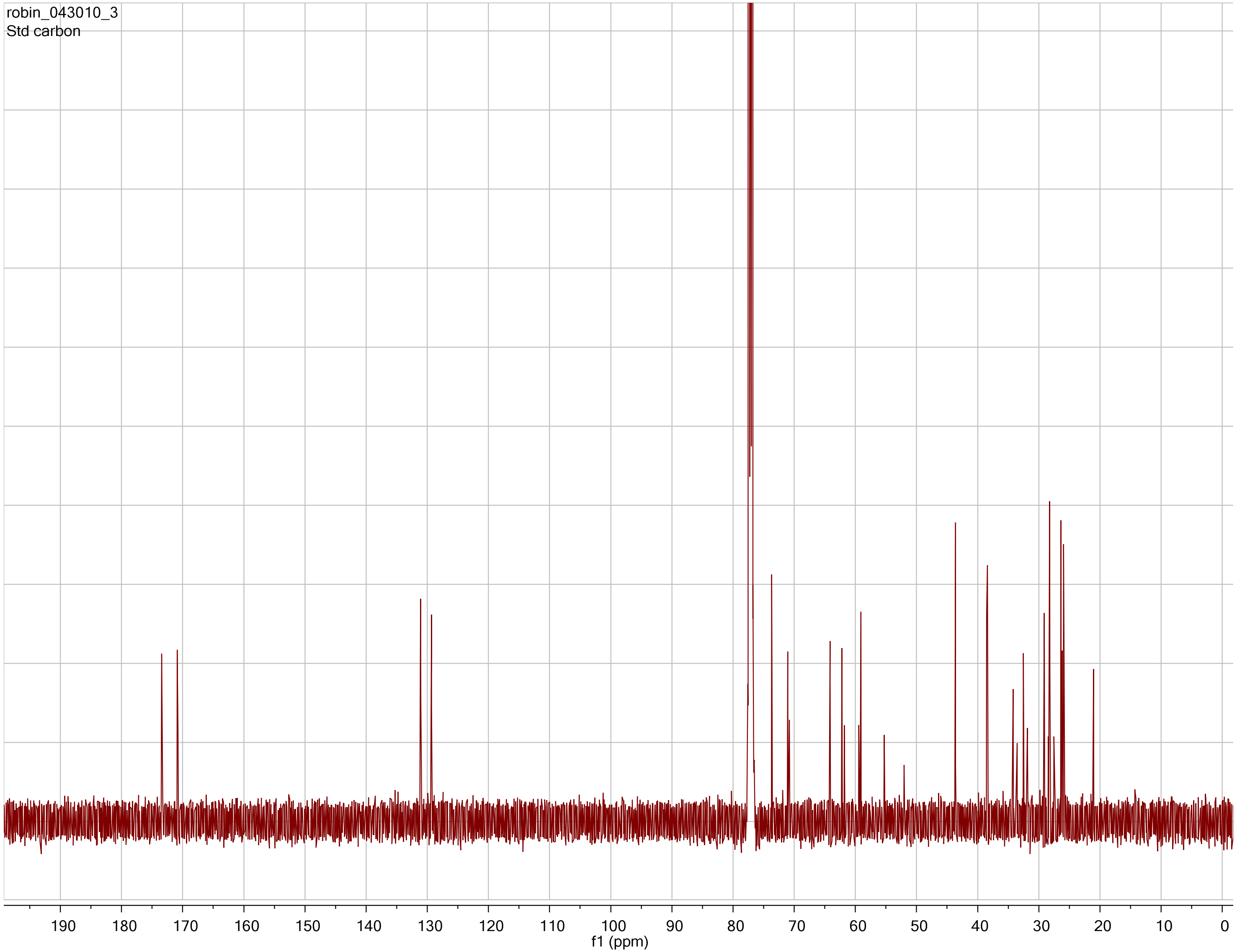

Figure S 12: ^{13}C NMR (125 MHz, $CDCl_3$) spectrum of columbamide B.

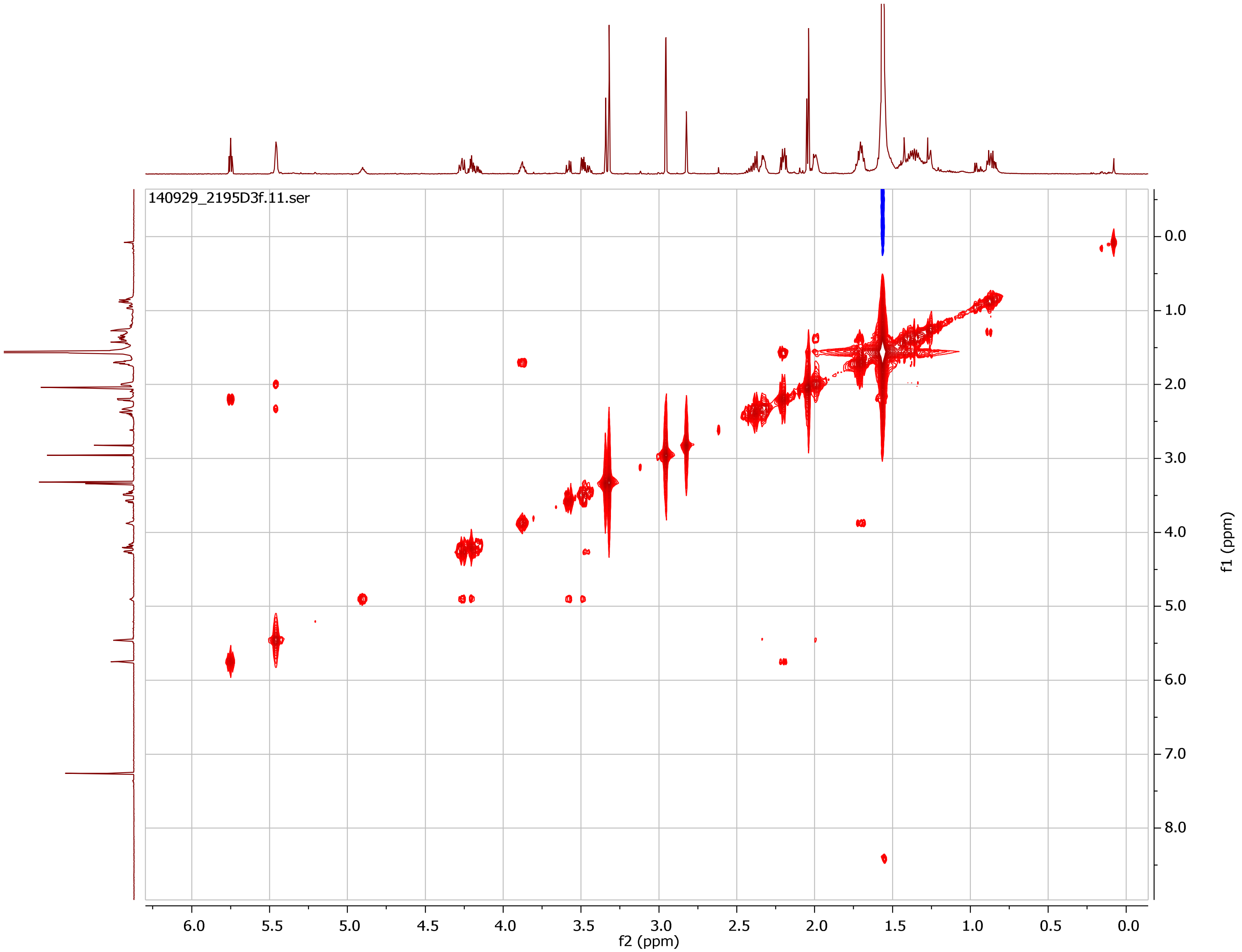

Figure S 13: COSY (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide B.

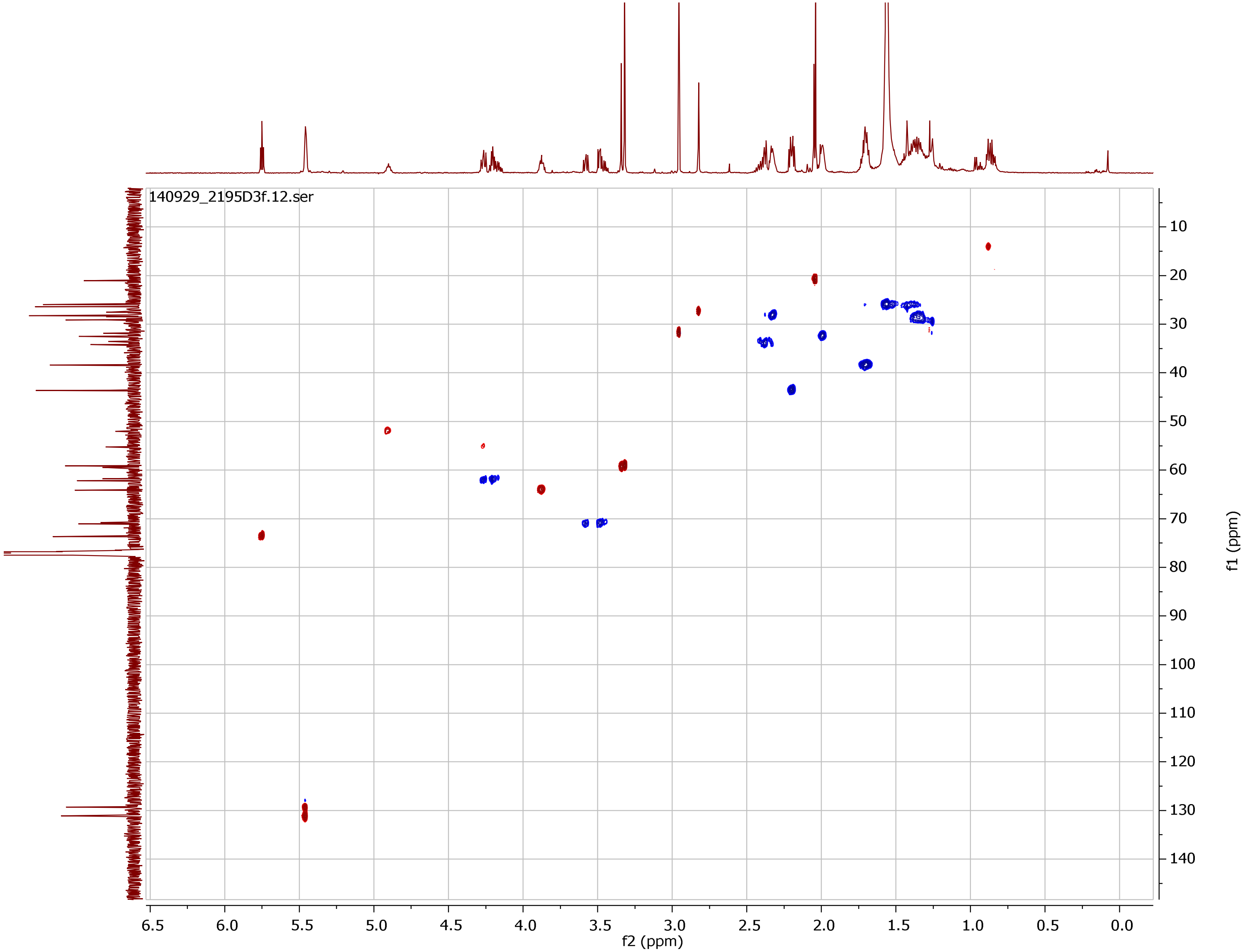

Figure S 14: HSQC (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide B.

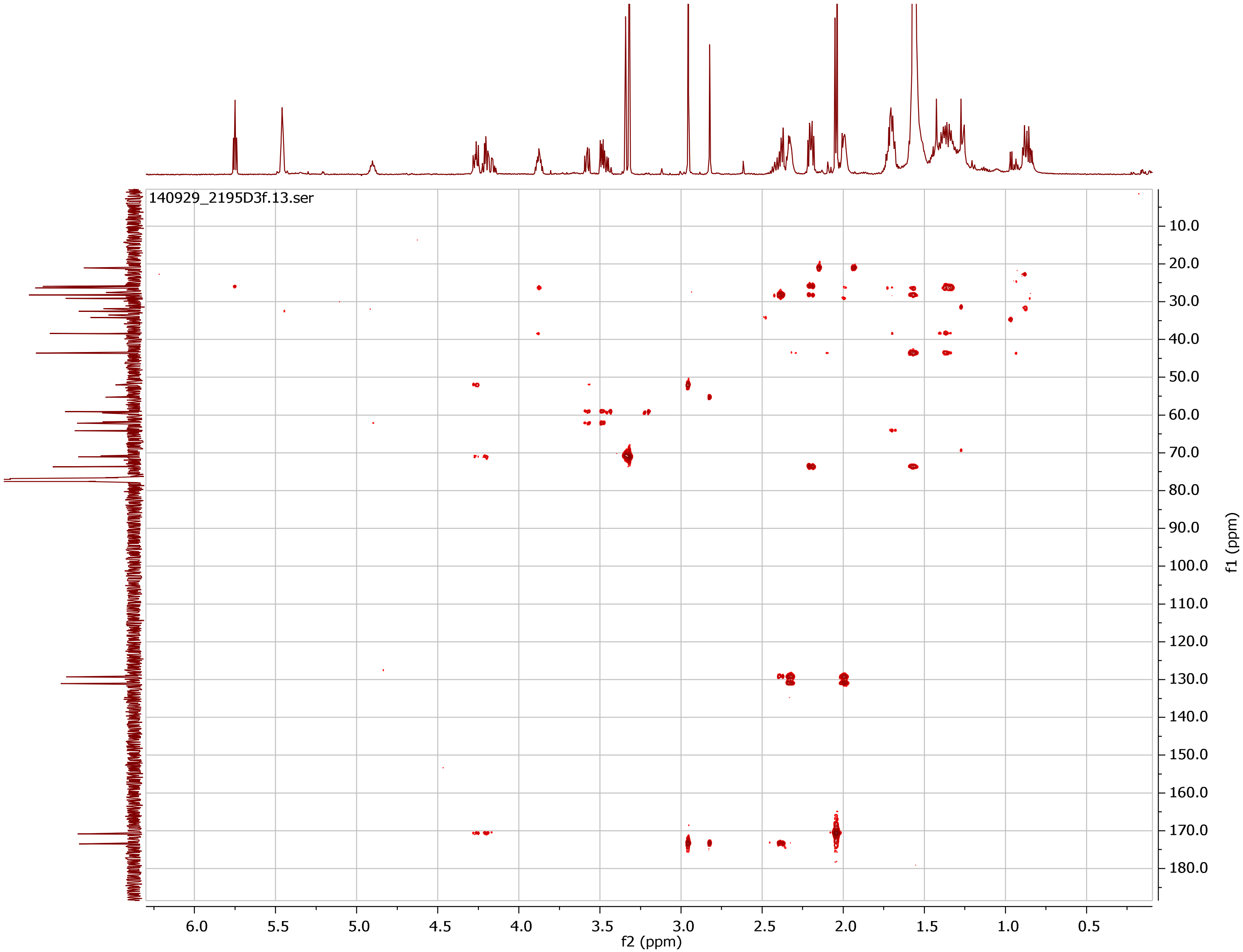

Figure S 15: HMBC (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide B.

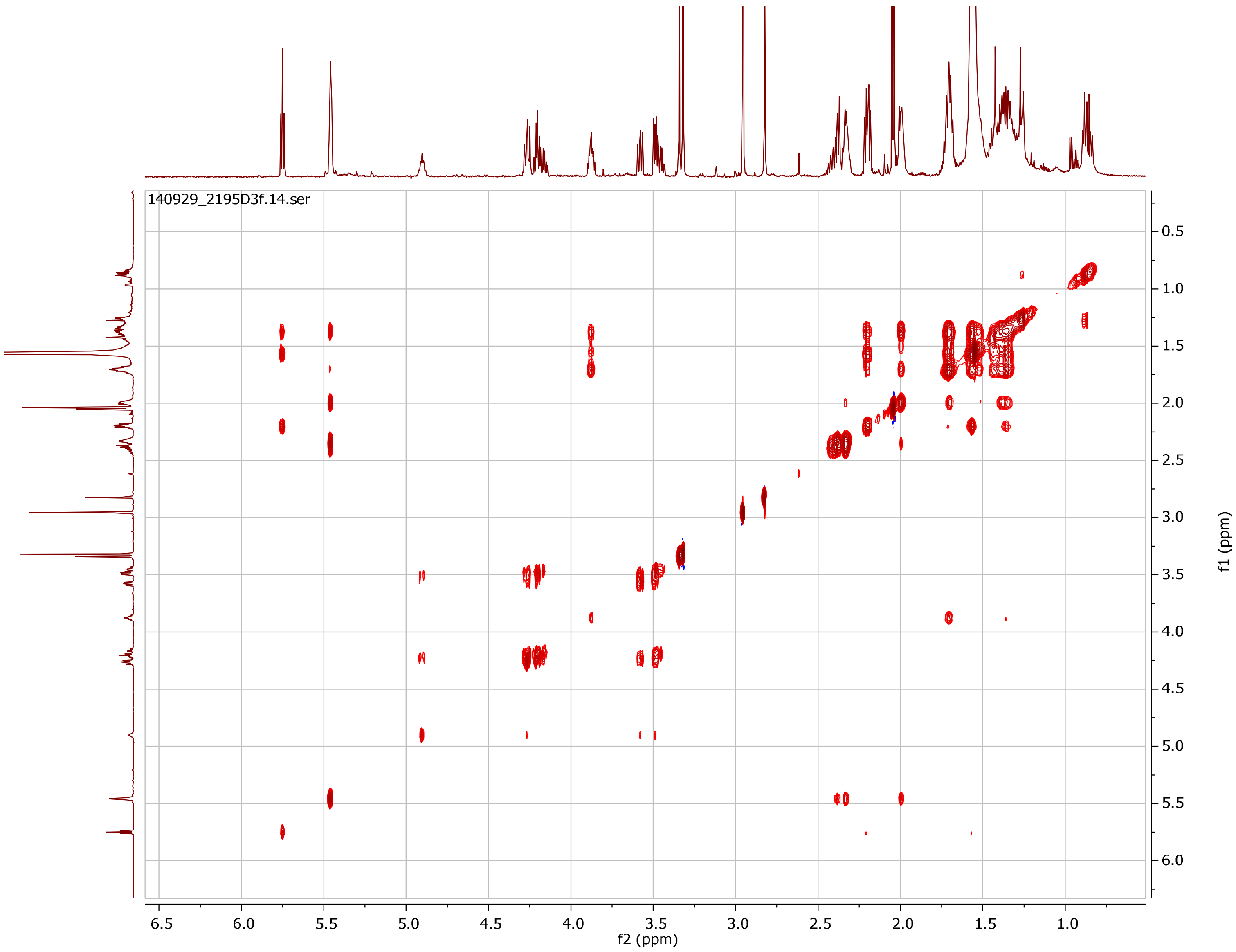

Figure S 16: TOCSY (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide B.

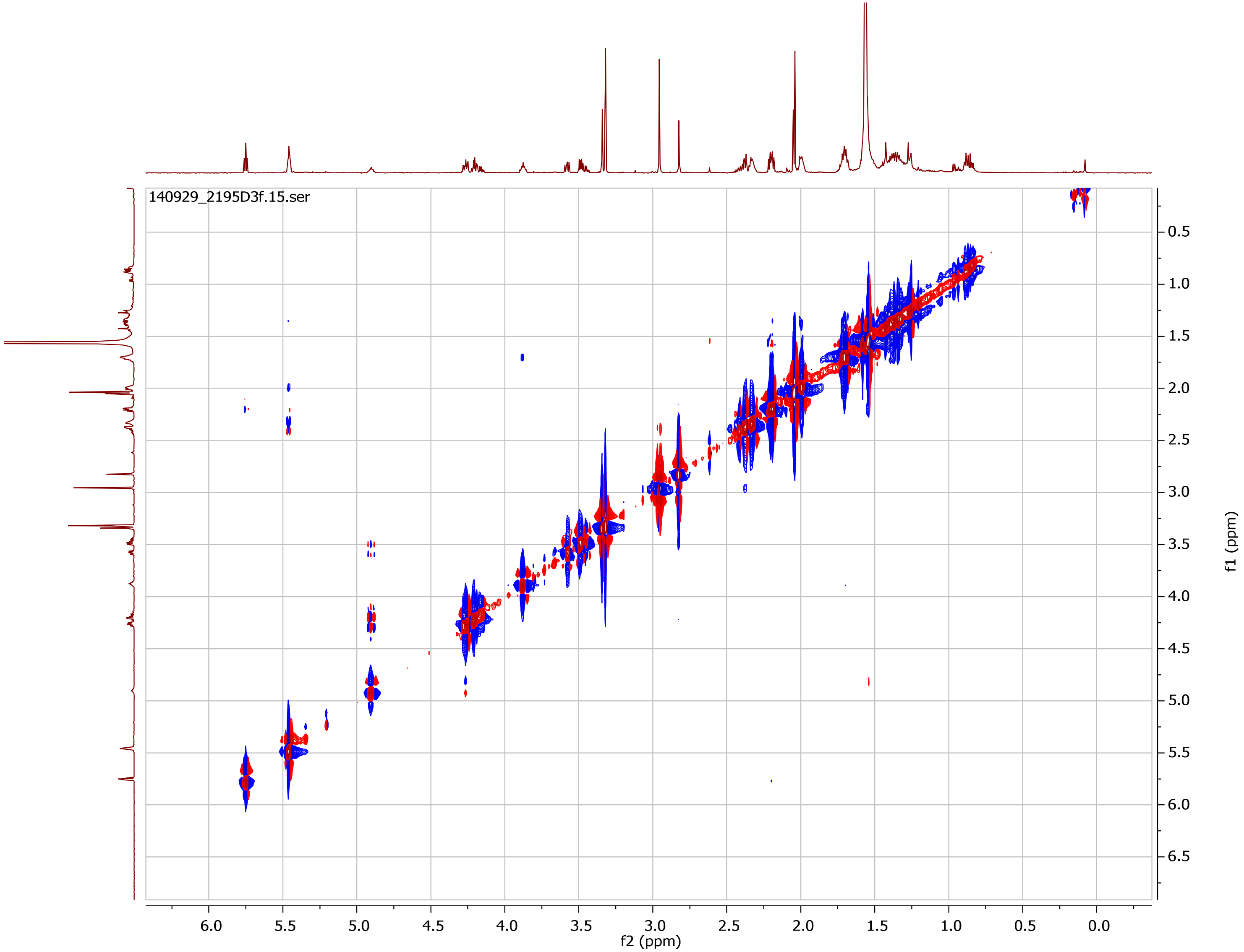

Figure S 17: NOESY (^{1}H 600 MHz, $CDCl_3$) spectrum of columbamide B.

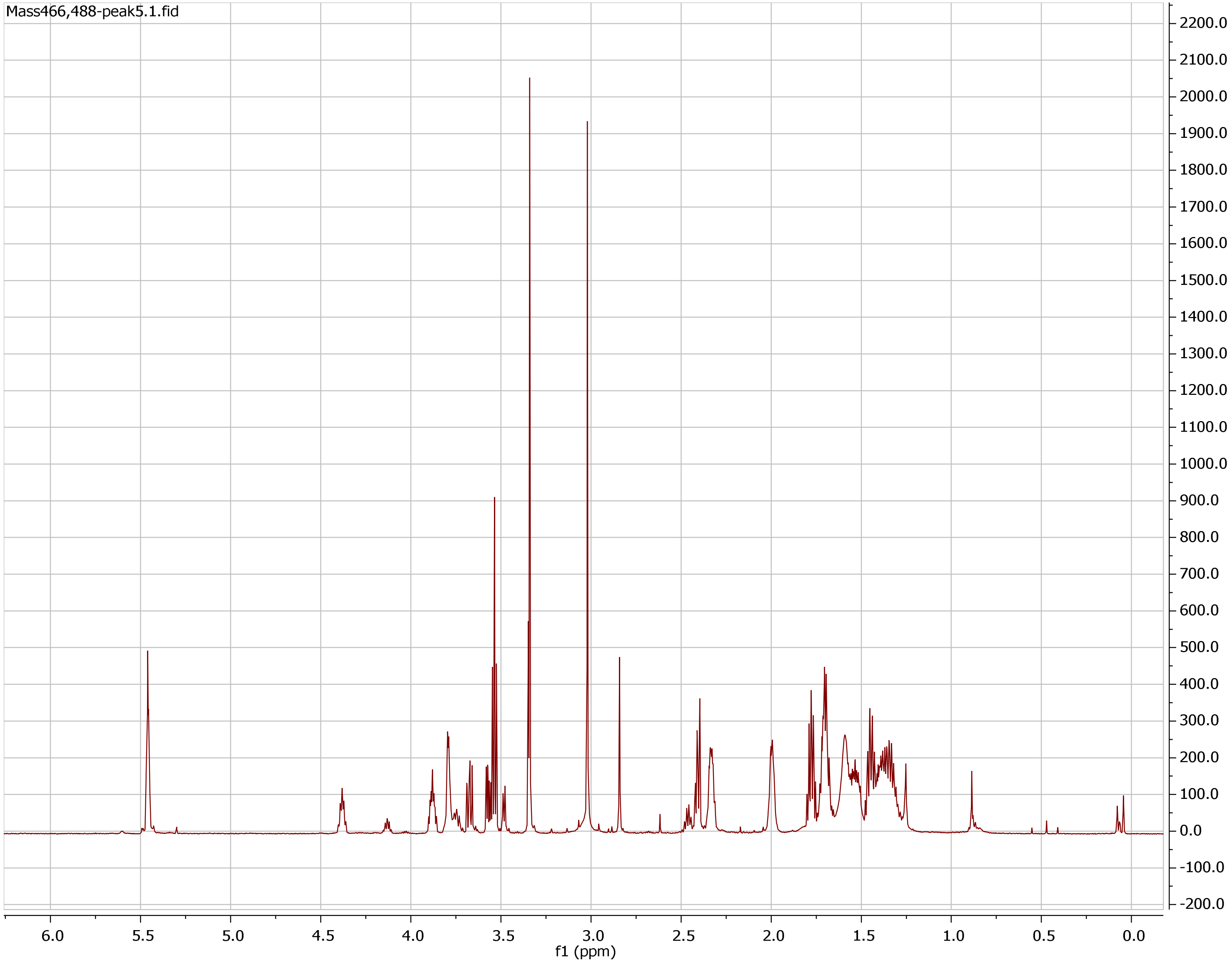

Figure S 18: ^{1}H NMR (600 MHz, $CDCl_3$) spectrum of columbamide C.

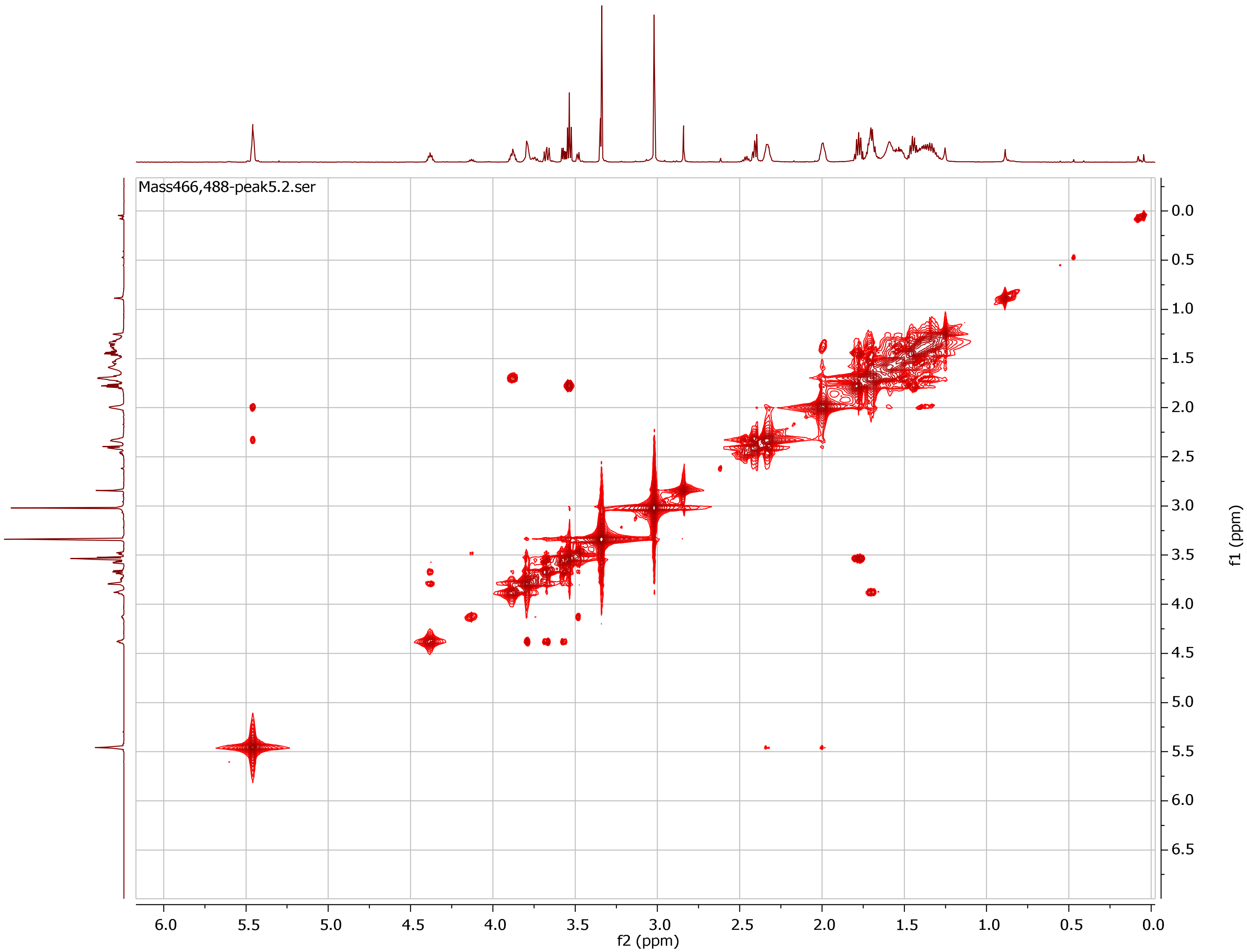

Figure S 19: COSY (¹H 600 MHz, $CDCl_3$) spectrum of columbamide C.

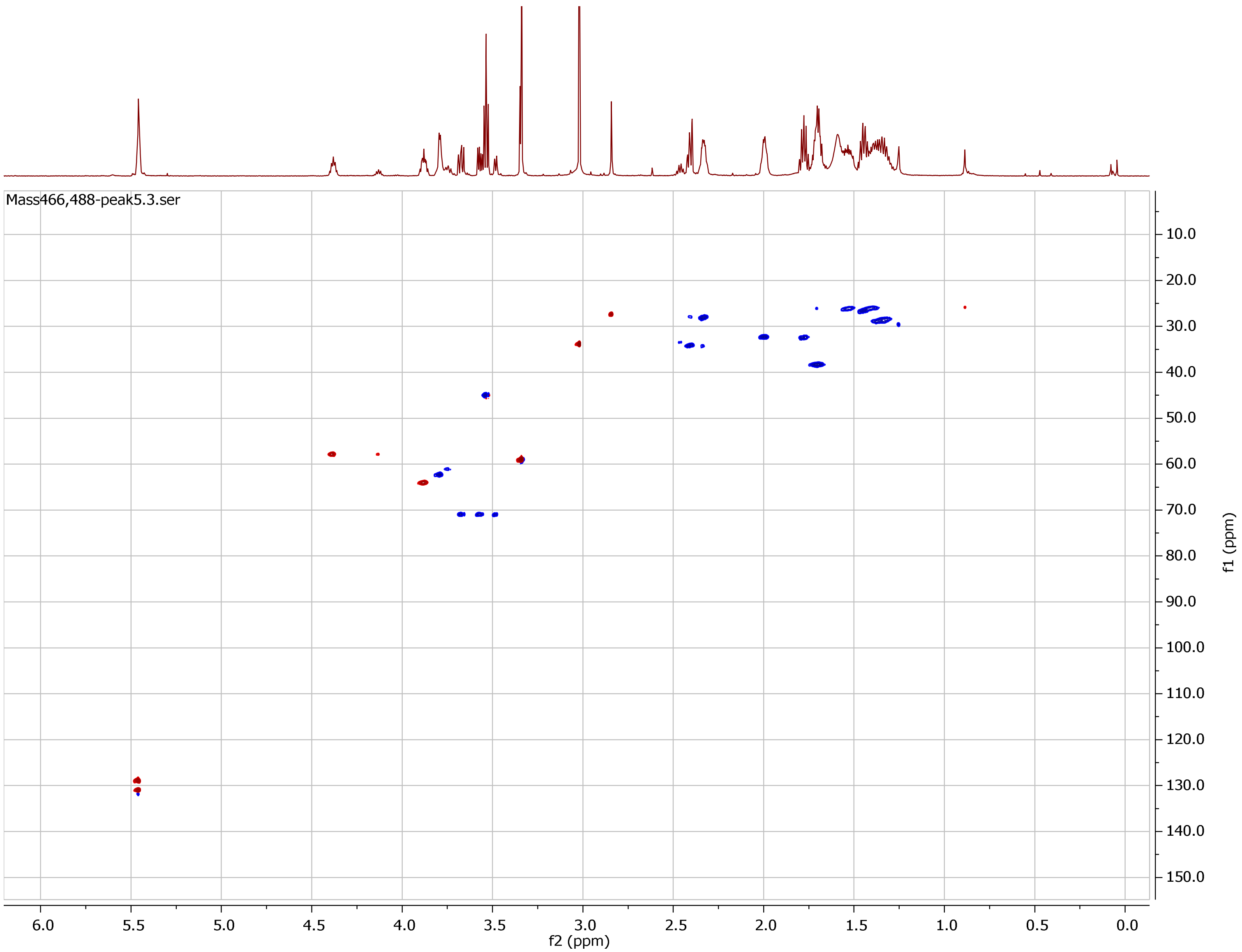

Figure S 20: HSQC (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide C.

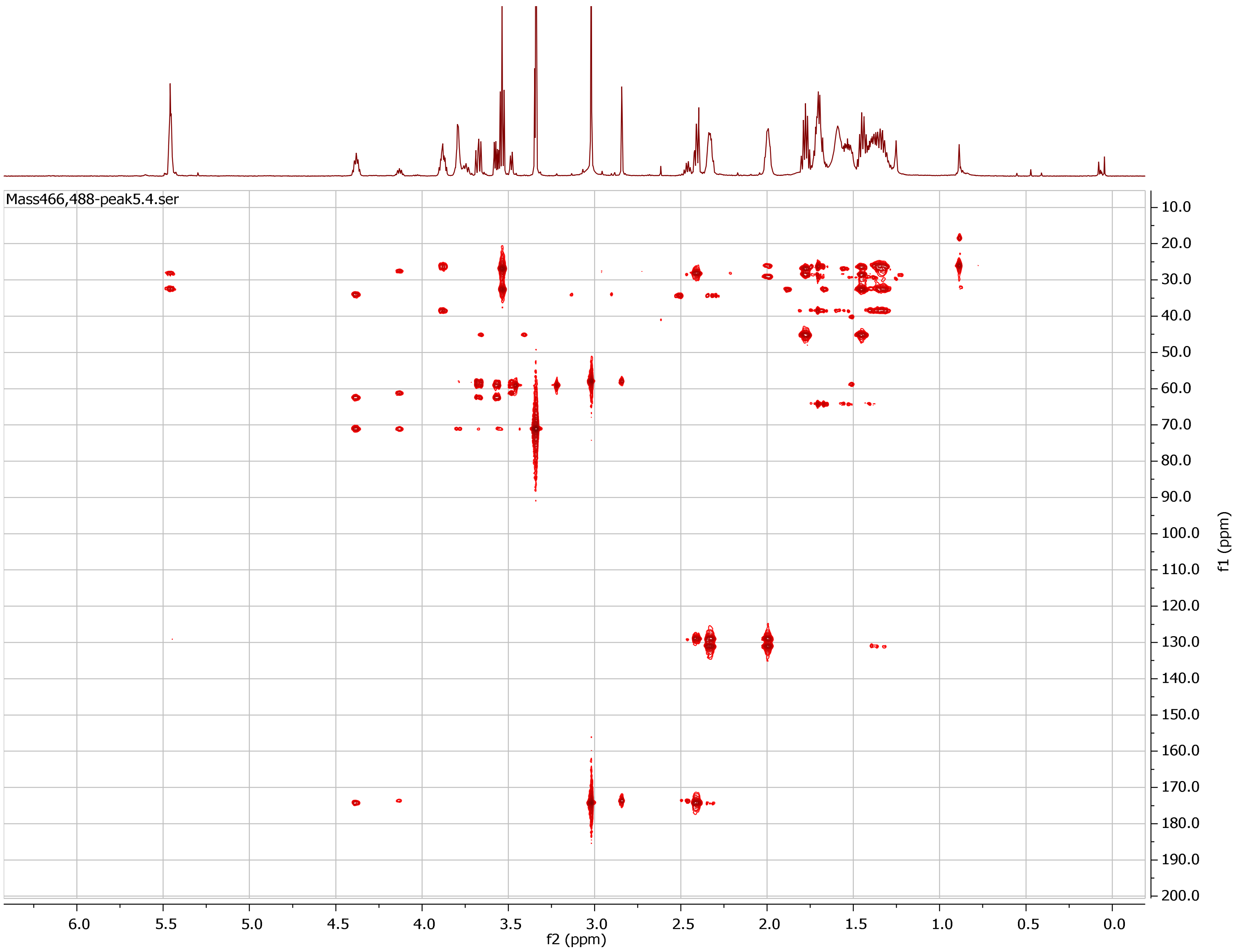

Figure S 21: HMBC (^{1}H 600 MHz, CDCl$_3$) spectrum of columbamide C.

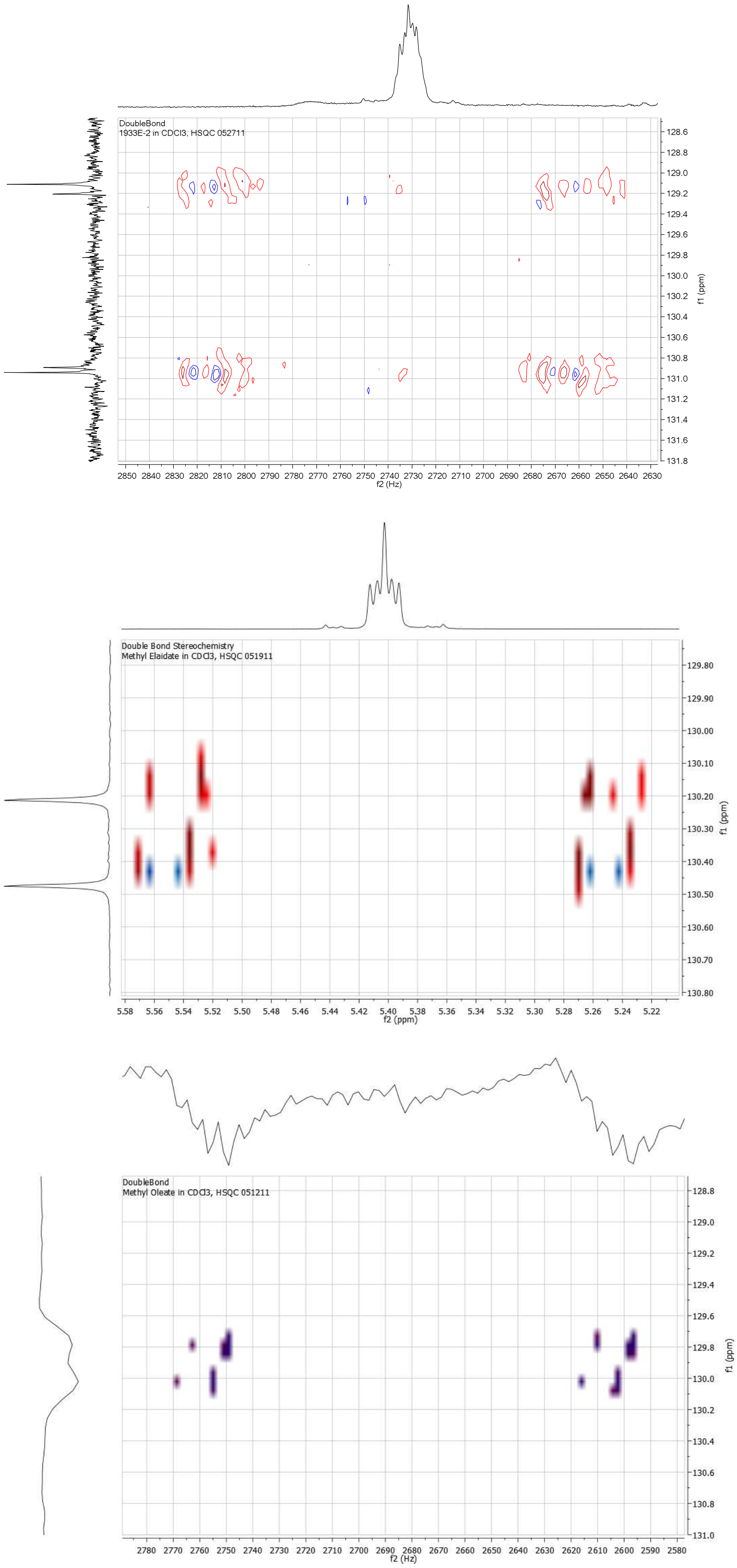

Figure S 22: Measurement of the coupling constant using HSQC and measuring the $^{3}J_{HH}$ coupling constant from the ^{13}C satellites of columbamide A, methyl elaidate and methyl oleate.

[1] G. L. Challis, J. Ravel, C. A. Townsend, *Chemistry & Biology* **2000**, *7*, 211.
[2] M. Z. Ansari, J. Sharma, R. S. Gokhale, D. Mohanty, *BMC Bioinformatics* **2008**, *9*, 454.