Supplementary Information for:

Ancient plant-like terpene biosynthesis in corals

Immo Burkhardt[1], Tristan de Rond[1§], Percival Yang-Ting Chen[1,#], Bradley S. Moore[1,2]*

[1]Scripps Institution of Oceanography, University of California San Diego; 9500 Gilman Dr., La Jolla, CA 92093, USA.
[2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego; 9500 Gilman Dr., La Jolla, CA 92093, USA.

§Present address: School of Chemical Sciences, University of Auckland; Private Bag 92019, Auckland 1142, New Zealand.

#Present address: Morphic Therapeutics; 35 Gatehouse Drive, A2, Waltham, MA 02451, USA

*Corresponding author. Email: bsmoore@ucsd.edu

Contents

Supplementary Tables

Supplementary Table 1. Uniprot accession numbers of TC sequences used for constructing the TC profile HMM.

A0A1L7U8F2
A0A1L7VZE7
E4MYY0
A0A0H5BN61
A0A0H5BN57
D9XDR8
K0K750
B5GS26
C7PLV2
A0A0L6WFN4
A0A2N3Y098
A0A0L6WRL8
E8W6C7
B1W019
E4N7E5
S0DX56
A9FZ87
A8NU13
A7NH01
A0A0L6WB32
B5GW45
D5SL78
B5GMG2
A4FG19
A3KI17
Q9F1Y6
D2B747
D9XD61
A0A1L9UKS1
E9F8R9
A1C8C3
A8NE23
Q96WT2
Q9UR08
Q03471
B6H063
Q82RR7
M4VQY9
E4N7E5
A0A0B4G3Q7
A0A0B4IF96
E9E766
A0A0B4GN88
A0A0B4FWC3
A0A1B4XBG5
P9WEV7
C9K1X5
Q9UVY5
S0EA85
O13284
Q5KSN4
Q9K499
A0A169T193
A0A1Y1C7Q5
A0A0P0ZEM1
A0A0P0ZD79
A0A1U9WTL2
C9K2Q3
A2PZA5
Q9X839
A8NCM9
B2J4A4
B5GRC8
I1S104
M1V9Q0
M1VDX3
A0A0B4EB91
A0A1B8A5N9
A1DN30
A0A2Z6AQX7
A0A348DU52
E3VWJ0
Q82IY4
Q55012
B2DBF0
A0A2L0VXR0
Q6WP50
P0DL13
B5H7H3
A0A2Z6AQX6
B5HDJ6
P0DPK6
A0A5S9I252
Q9AJE3
O34707
P13513
I2N045
A2PZA5
A0A1Q5UFC3

Supplementary Table 2. Public genome assemblies used for genome mining.

Coral species	Number of TC genes (number including paralogs)	Acc no.	reference
Dendronephthya gigantea	2 (2)	NCBI acc no.: GCA_004324835.1	[1]
Renilla muelleri	1 (1)	Downloaded from: http://rmue.reefgenomics.org/	[2]
Renilla reniformis	1 (1)	NCBI acc no.: GCA_900177555.1	[3]
Xenia sp.	1 (2)	Downloaded from: https://cmo.carnegiescience.edu/data	[4]
Trachythela sp. YZ-2020	6 (12)	NCBI acc no.: GCA_016169945.1	[5]
Paramuricea clavata	4 (8)	Downloaded from: https://denovo.cnag.cat/pclavata	[6]

Supplementary Table 3. Public data used for transcriptome mining.

Species	Number of TC	Assembly info	SRA dataset	reference	Terpenoids reported
Alcyonium palmatum	0	Assembled in this work	SRR3407216	[7]	Yes [8]
Acanthogorgia aspera	1 (1 partial)	NCBI acc no.: GEXC01	SRR3303654	-	From genus [9]
Briareum asbestinum	2	NCBI acc no.: GHBD02	SRR7895344	[10]	Yes [11,12]
Callogorgia delta	3	Assembled in this work	SRR13925246	[13]	No
Clavularia sp.	1 (1 partial)	NCBI acc no.: GHAW01	SRR8102779, SRR8102780	[14]	From genus [15,16]
Corallium rubrum	1	Downloaded from: https://doi.org/10.5061/dryad.31f77	SRR1552945	[17]	From genus [18]
Dendronephthya gigantea	2 (1 partial)	Assembled in this work	SRR8293935	[1]	From genus [19,20]
Eleutherobia rubra	2	NCBI acc no.: GHFI01	SRR8486076	[21]	From genus [22]
Eunicea calyculata	1	Assembled in this work	SRR6039601	[23]	Yes [24]
Eunicea flexuosa	1	Assembled in this work	SRR12876635	-	From genus [24]
Eunicella cavolinii	1 (1 partial)	Assembled in this work	SRR1324943	[25]	Yes [26]
Gorgonia ventalina	3	Assembled in this work	SRR935083	-	Yes [27]
Heliopora coerulea	3	NCBI acc no.: IABP01	DRR072217, DRR072218, DRR072219	[28]	Yes [29]
Paramuricea biscaya	2	Assembled in this work	SRR7521178	[30]	From genus [31]
Paramuricea clavata	3 (1 partial)	Assembled in this work	ERR3664728	[6]	From genus [31]
Paramuricea type B3	1 (1 partial)	Assembled in this work	SRR13925244	[13]	From genus [31]
Pinnigorgia flava	0	Assembled in this work	ERX4193767	[32]	no
Renilla koellikeri	1 (1 partial)	Assembled in this work	SRR7495167	-	From genus [33]
Renilla reniformis	0	Assembled in this work	ERR2808321	[2]	Yes [33]
Sarcophyton trocheliophorum	1	Assembled in this work	SRR12573946	-	Yes [34]
Sinularia cruciata	1 (1 partial)	Assembled in this work	ERR3026434	[35]	From genus [36]
Tubipora musica	2	Assembled in this work	ERR3026435	[35]	Yes [37]
Xenia sp.	1	Assembled in this work	SRR9278441	[4]	From genus [38]

Supplementary Table 4. Details for TC coding contigs from octocoral genomes.

scaffold	Organism and reference	length (kbp)	coded TC	GC content (%)	coverage (avg. Reads per base)	Dataset used for transcript mapping	Dataset used for long read mapping
NW_021163036	*D. gigantea*[1]	2263	DgTC-1	38	90 (pacbio)	SRR8293935	SRR8293699
NW_021163042	*D. gigantea*[1]	2228	DgTC-2	37	100 (pacbio)	SRR8293935	SRR8293699
HiC_scaffold_25	*Xenia sp.*[4]	4749	XsTC-1	35	88 (ONP)	SRR9278441	SRR9278438

Supplementary Table 5. Genetic context of XsTC-1 on contig HiC_scaffold_25.

Sequence	GC%	# exons	Region (bp)	Best scoring BLAST hit:	Sequence name	organism	E-value	NCBI accession no.
XsTC-1 -1	39	8	402750-397351		uncharacterized protein LOC122962660	*Acropora millepora*	1x10-^{128}	XP_044181808.1
XsTC-1	42	1	421582-420354		terpene synthase family protein	*Streptosporangium nondiastaticum*	0.1	WP_106680131.1
XsTC-1 +1	38	1	426242-425778		hypothetical protein KOW79	*Hemibagrus wyckioides*	0.076	KAG7326961.1
XsTC-1 +2	40	1	426516-427094		no hit			
XsTC-1 +3	53	5	436341-433328		no hit			
XsTC-1 +4	50	5	442821-440118		microspherule protein 1-like	*Plakobranchus ocellatus*	8x10-^{31}	GFN97869.1
XsTC-1 +5	39	23	445833-486627		transmembrane protein 87A-like	*Actinia tenebrosa*	0.0	XP_031573117.1

Supplementary Table 6. Genetic context of DgTC-1 on contig NW_021163036.

Sequence	GC%	# exons	Region (bp)	Best scoring BLAST hit:	Sequence name	organism	E-value	NCBI accession no.
-3	47	15	1163131-1171572		rho GTPase-activating protein 39-like	*Acropora millepora*	5x10^{-154}	XP_029180732.2
-2	43	10	1178909-1174089		ubiquitin thioesterase otulin-like	*Zootermopsis nevadensis*	3x10^{-37}	XP_021939107.1
-1	45	1	1180563-1179595		retinol dehydrogenase 12-like	*Crassostrea gigas*	5x10^{-111}	XP_011420247.2
DgTC-1	46	1	1181590-1182822		HEAT repeat domain-containing protein	*Cand. Lokiarchaeota archaeon*	0.57	NMC08674.1
+1	46	14	1191885-1199119		suppressor of hairless protein	*Saccoglossus kowalevskii*	0.0	NP_001158483.1
+2	46	16	1201421-1221069		ABC transporter G family member 24	*Nematostella vectensis*	0.0	XP_001642016.2

Supplementary Table 7. Genetic context of DgTC-2 on contig NW_021163042.

Sequence	**GC%**	**# exons**	**Region (bp)**	**Best scoring BLAST hit:**	**Sequence name**	**organism**	**E-value**	**NCBI accession no.**
-2	41	10	1129642-1131952		protein-glutamine gamma-glutamyltransferase K	*Phallusia mammillata*	7x10^{-173}	CAB3266928.1
-1	42	3	1144401-1150450		uncharacterized protein LOC110068857	*Orbicella faveolata*	1x10^{-18}	XP_020631939.1
DgTC-2	39	1	1154905-1153709		nucleotidyltransferase family protein	*Oscillospiraceae bacterium*	8.9	MBR3419154.1
+1	41	6	1156399-1158527		15-hydroxyprostaglandin dehydrogenase	*Cyprinodon variegatus*	4x10^{-51}	XP_015226913.1
+2	37	6	1164817-1167808		peroxisomal biogenesis factor 19-like	*Acropora millepora*	1x10^{-56}	XP_029188873.2
+3	37	1	1169969-1171558		zinc finger protein 271-like	*Acyrthosiphon pisum*	5x10^{-108}	XP_008180596.1
+4	39	9	1172532-1177997		proton-coupled amino acid transporter 1-like	*Orbicella faveolata*	6x10-^{110}	XP_031553692.1

Supplementary Table 8. PISA server results.

Enzyme	$\Delta^i G$ (kcal/mol)	Interface (Å^2)	State in solution	reference
ErTC-2 (7s5l)	-6.5	597		This work
epi-isozizaene synthase (4lz0)	-2.2	485	monomeric	39
2-MIB synthase (4la6)	-37.2	2549	parallel homodimeric	40
trichodiene synthase (2q9z)	-40.9	2606	antiparallel homodimeric	41

Supplementary Table 9. RSMDs of ErTC2 with microbial (red) plant (green) and IDS (black) structures.

RMSD (Å)^{a}	4lz0 EIZS	4okm	4kux	5eau	3p5r TXDS	1rqj	4jzb	5ja0 FPPS
ErTC2	4.190	4.630	4.707	4.221	5.011	9.723	7.665	16.417

.

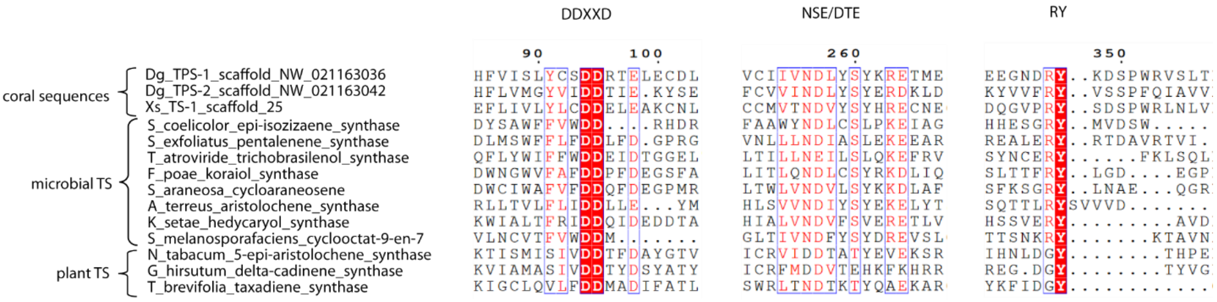
Supplementary Table 10. Accession numbers for non-coral sequences used for phylogenetic analysis.

E2E2N7
B6SCF5
I6RAQ6
J7LQ09
Q5SBP3
P59287
P0CJ43
O48935
Q4KSH9
Q84UU4
B3TPQ6
Q41594
B5A434
Q9FXY7
H8ZM70
C7E5V7
O24475
Q6PWU2
A4FVP2
C5YHH7
O49853
Q84LF0
J7LH11
B1B1U3
O81191
O64404
G1JUH1
Q6Q3H2
B9S9Z3
Q94G53
Q8GUE4
B9RXW0
I6QPS5
G5CV45
B0FGA9
E2E2P0
O64961
Q29VN2
C7ASI9
XP_041688448
XP_031085895
WP_014133196
D9XDR8
WP_015102836
WP_003954606
WP_012792334
KNZ74377
WP_010314578
WP_020663213
WP_014153723
B1W019
XP_023427124
WP_012241161
KAG5727529
KAG5733850
XP_001836356
WP_012119179
WP_003956090
WP_003963519
WP_003952918
WP_011874125
WP_053126184
WP_011031839
D2B747
WP_003994861
XP_007824857
XP_001832925
Q9UR08
XP_038933630
XP_002557473
WP_010981512
BAJ27126
KID95099
A0A1B4XBG5
4OMG_A
WP_011030119
AQY56778
KFG77771
EAU89322
WP_012410187
KAG8664989
XP_011319367
M1VDX3
WP_143632835
WP_010984429
Q55012
XP_024552383
WP_005320742
QJQ03973
WP_005317515
WP_030261827
A0A5S9I252
P13513
WP_006348711
QGA30878
XP_001832573
XP_001836556
XP_001832925
XP_001836356
EIM83755
EIM91001
EIM88705
EIM82223
EIM91236
P0DL13
BBD74517
XP_645958
XP_645125
XP_644874
XP_642676
XP_640697
XP_638489
XP_642261
XP_629084
XP_642260
XP_002679043
XP_002682659
AXN72983
AZO92733
AZO92734
ASV63464
ASV63466
ASV63465
Q43315
Q54BK1
Q9LIA0
B2DBE9
A0A2L0VXR5
P22939
Q8WMY2
P0AD57
J7FIX8
A0A1B4XBK0
A0A140JWS2
Q7XYS9
O64905
O24241
P49351
O53507
Q3IPL1
P48368
P54976
P39464
P45204
XP_001958742
XP_022808076
ALL35400
KAE9533555
AAX78434
XP_029182012
XP_028399803
P14324
P08836
APB88773
APB88774
APB88775
APB88776
APB88777
APB88778
APB88779
APB88780
AVZ23977
AYD76069
ALL35411
ALL35417
ALL35420
ALL35414

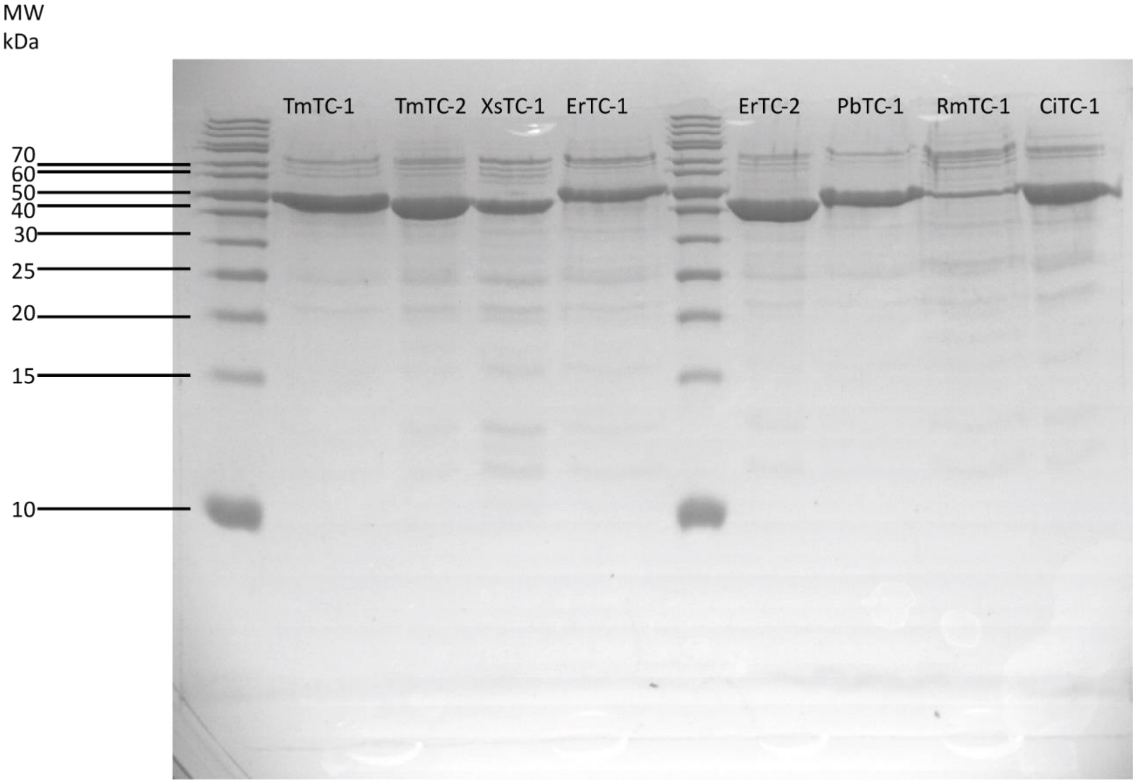
Supplementary Table 11. Data collection and model refinement statistics of coral terpene synthases.

	Cd(II)-ErTC-2	ErTC-2
PDB ID	-	7S5L
Beamline	UCSD X-ray Crystallography Facility	SSRL 12-2
Wavelength (Å)	1.54178	0.97946
Space Group	C2	C2
Cell dimensions (Å)	a = 82.75, b = 99.70, c = 46.48, β = 92.29°	a = 85.69, b = 101.65, c = 46.34, β = 90.60°
Resolution	33.98-2.57 (2.67-2.57)	40.0-1.58 (1.67-1.58)
# unique reflections	23522 (2453)	54425 (8769)
Completeness (%)	99.6 (96.8)	96.1 (94.1)
Redundancy	15.3 (8.8)	3.9 (3.9)
<I/sigI>	20.8 (3.9)	14.7 (2.6)
R_{sym}	0.073 (0.376)	0.046 (0.390)
$CC_{1/2}$	*	(0.90)
Resolution		37.97-1.58
# unique reflections		52284
R_{work} (%) / R_{free} (%)		16.1/19.2
RMS bond length (Å)		0.005
RMS bond angles (°)		0.697
# of atoms		
Protein Atoms		3152
Diphosphate		-
Ethylene glycol		12
Water molecules		188
Mg^{2+}		-
Cl^-		1
Average B-factor (Å^2)		
Protein Atoms		30
Diphosphate		-
Ethylene glycol		38
Water molecules		39
Mg^{2+}		-
Cl^-		28
Ramachandran plot		
Favored (%)		98.12
Allowed (%)		1.88
Outliers (%)		0.00
Rotamer outliers (%)		0.00
Residues modeled (1-393)		2-195, 201-381, 389-391

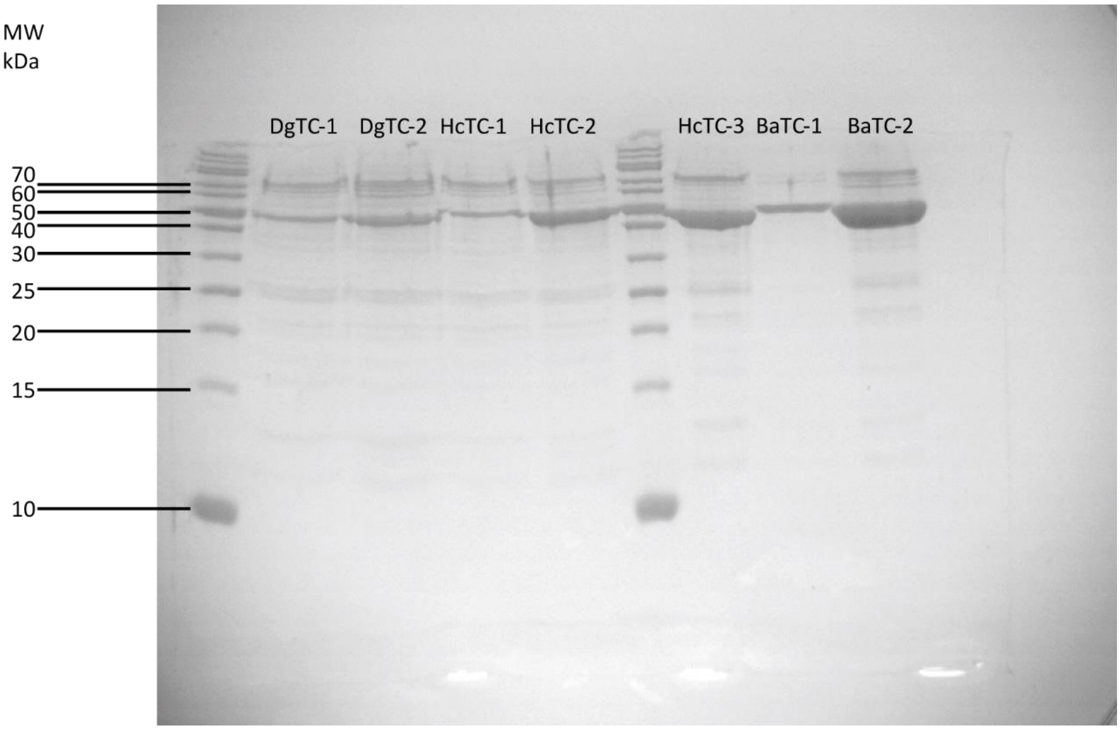
Supplementary Figures



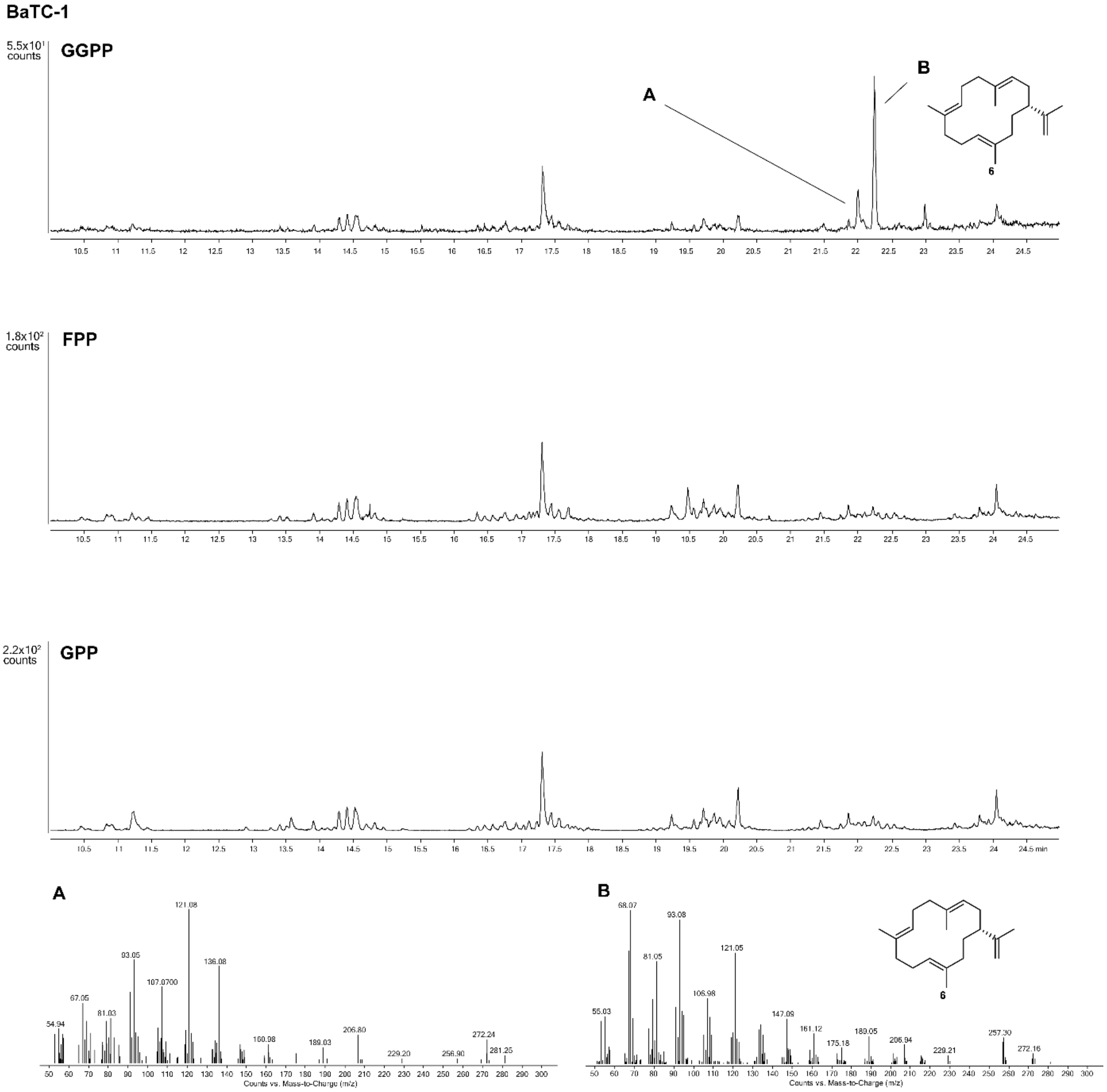
Supplementary Fig. 1. Regions from alignments with newly identified coral sequences with microbial and plant terpene synthase (TS) sequences that show the conserved motifs. Plant-like terpene synthases show more diversity in the NSD/DTE motif and do not have the RY motif.



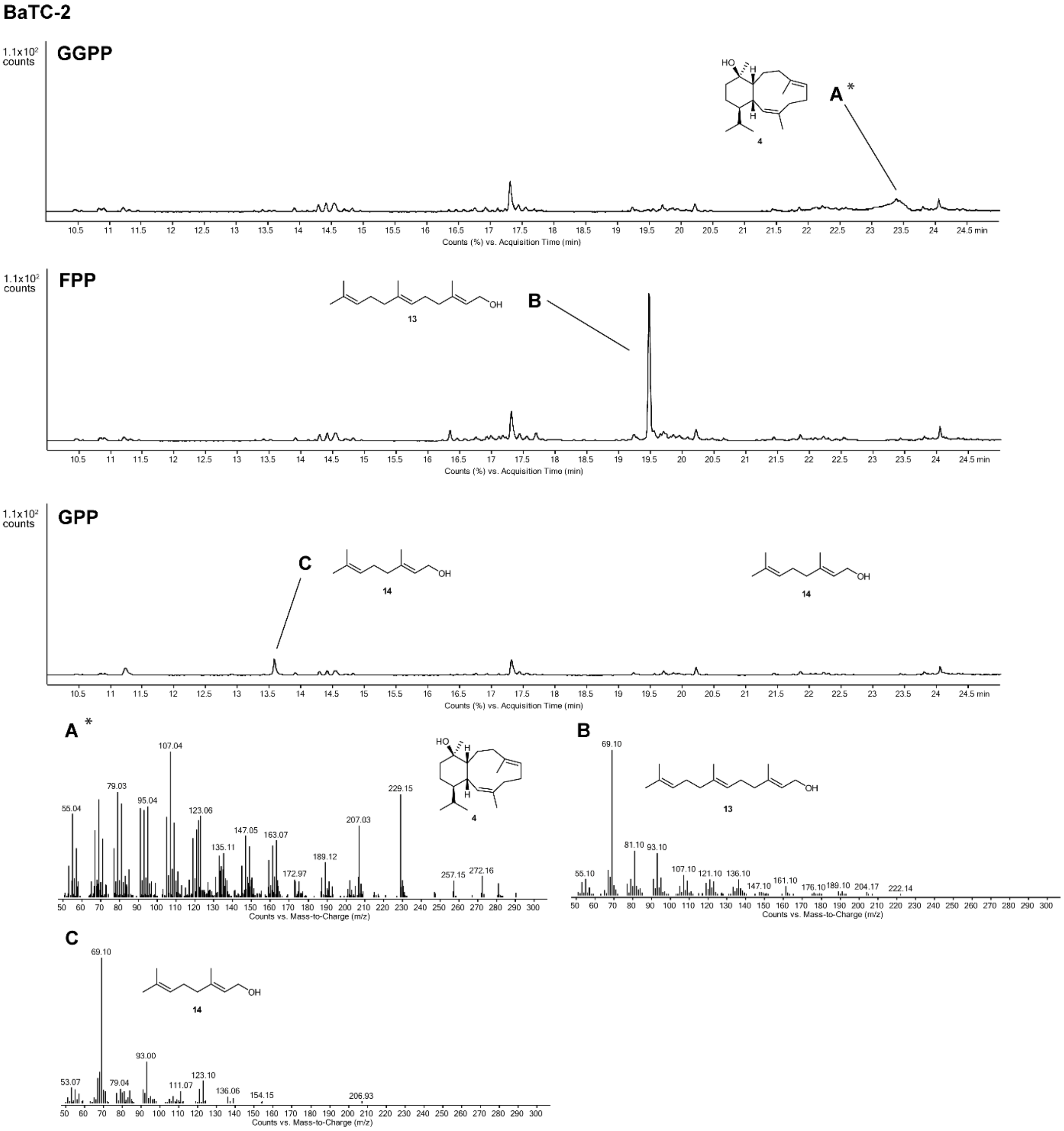
Supplementary Fig. 2. SDS PAGE analysis of recombinant TmTC-1, TmTC-2, XsTC-1, ErTC-1, ErTC-2, PbTC-1, RmTC-1, CiTC-1. The image shows representative bands after Ni-NTA purification. All expressions were carried out at least twice on different days. No attempts failed.



Supplementary Fig. 3. SDS PAGE analysis of recombinant DgTC-1, DgTC-2, HcTC-1, HcTC-2, HcTC-3, BaTC-1, BaTC-2. The image shows representative bands after Ni-NTA purification. All expressions were carried out at least twice on different days. No attempts failed.

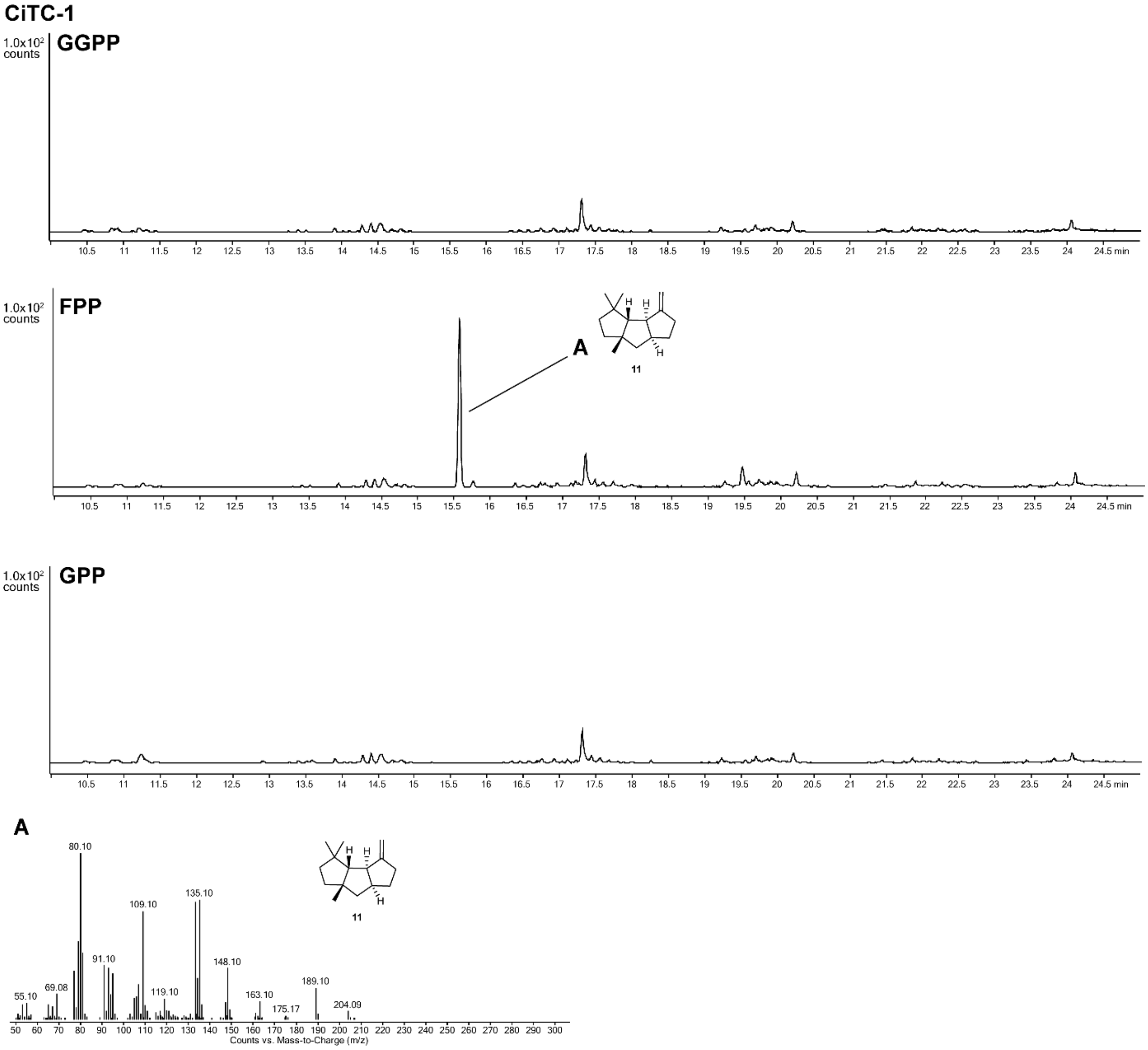


Supplementary Fig. 4. GCMS analysis of BaTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown. The absolute configuration of **6** produced by this enzyme was not determined.

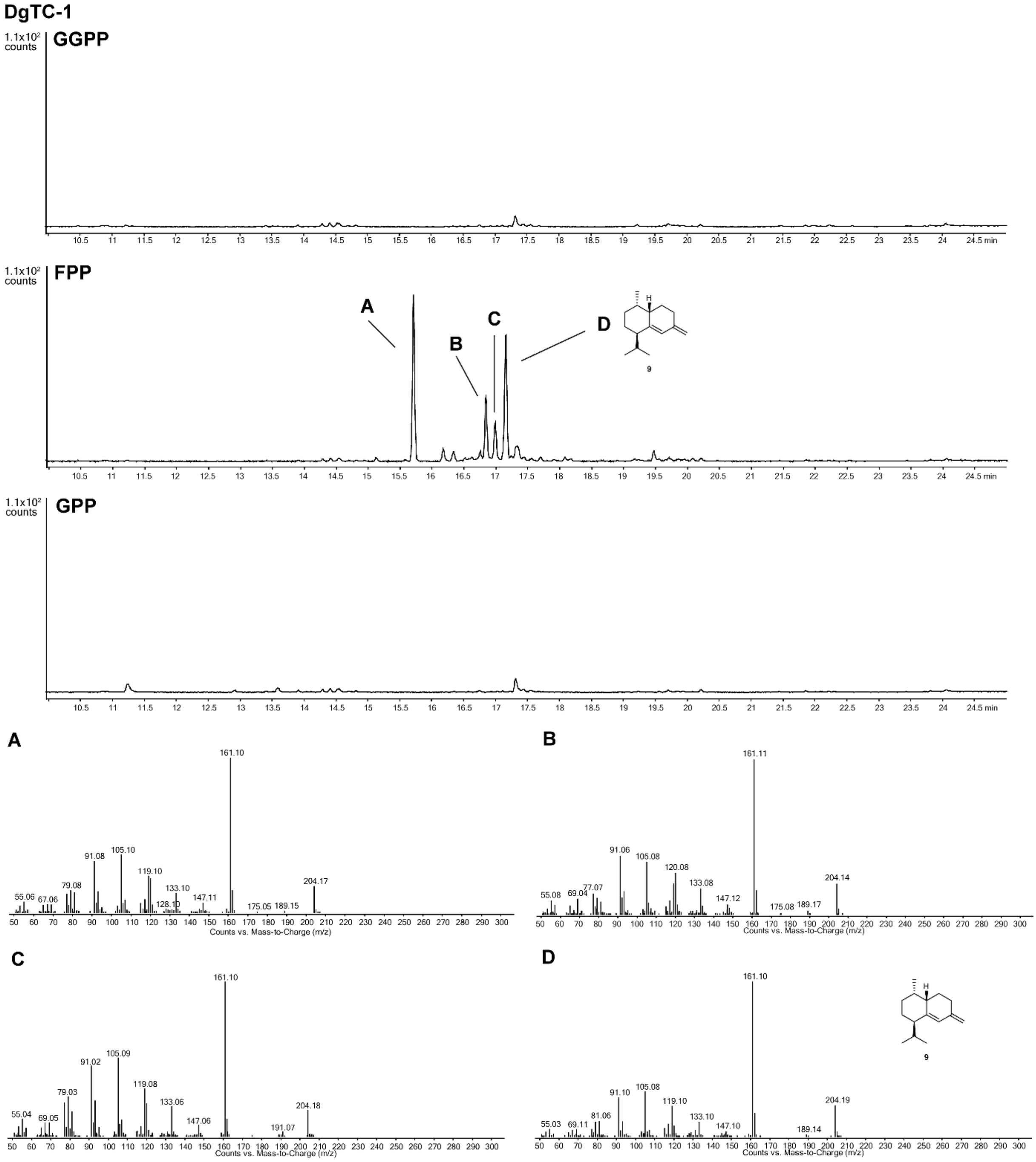


Supplementary Fig. 5. GCMS analysis of BaTC-2 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

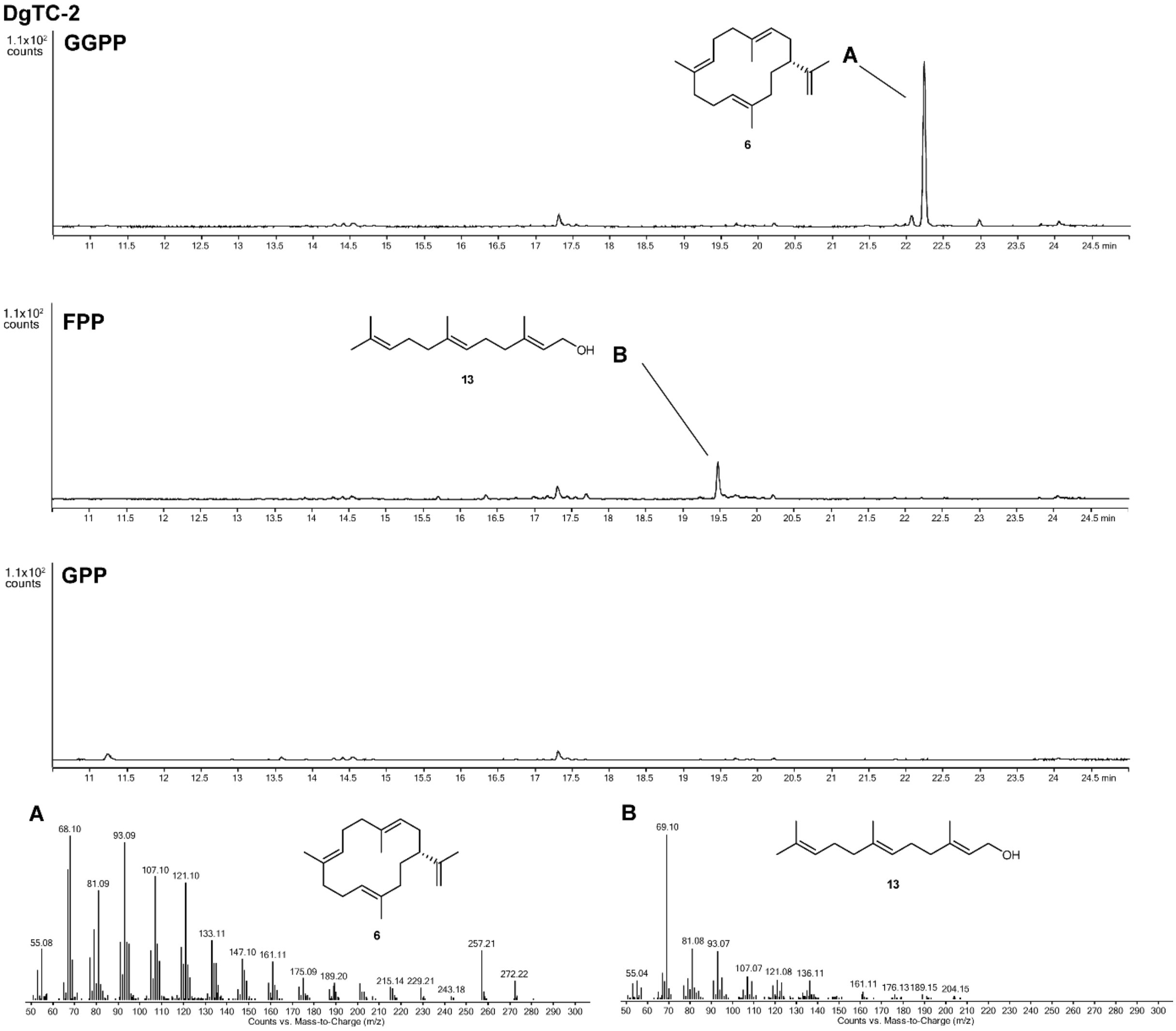
*: Compound **4** rearranges/decomposes in the GC injector and elutes as multiple products as a broad peak. Mass spectrum A is averaged over the whole signal.



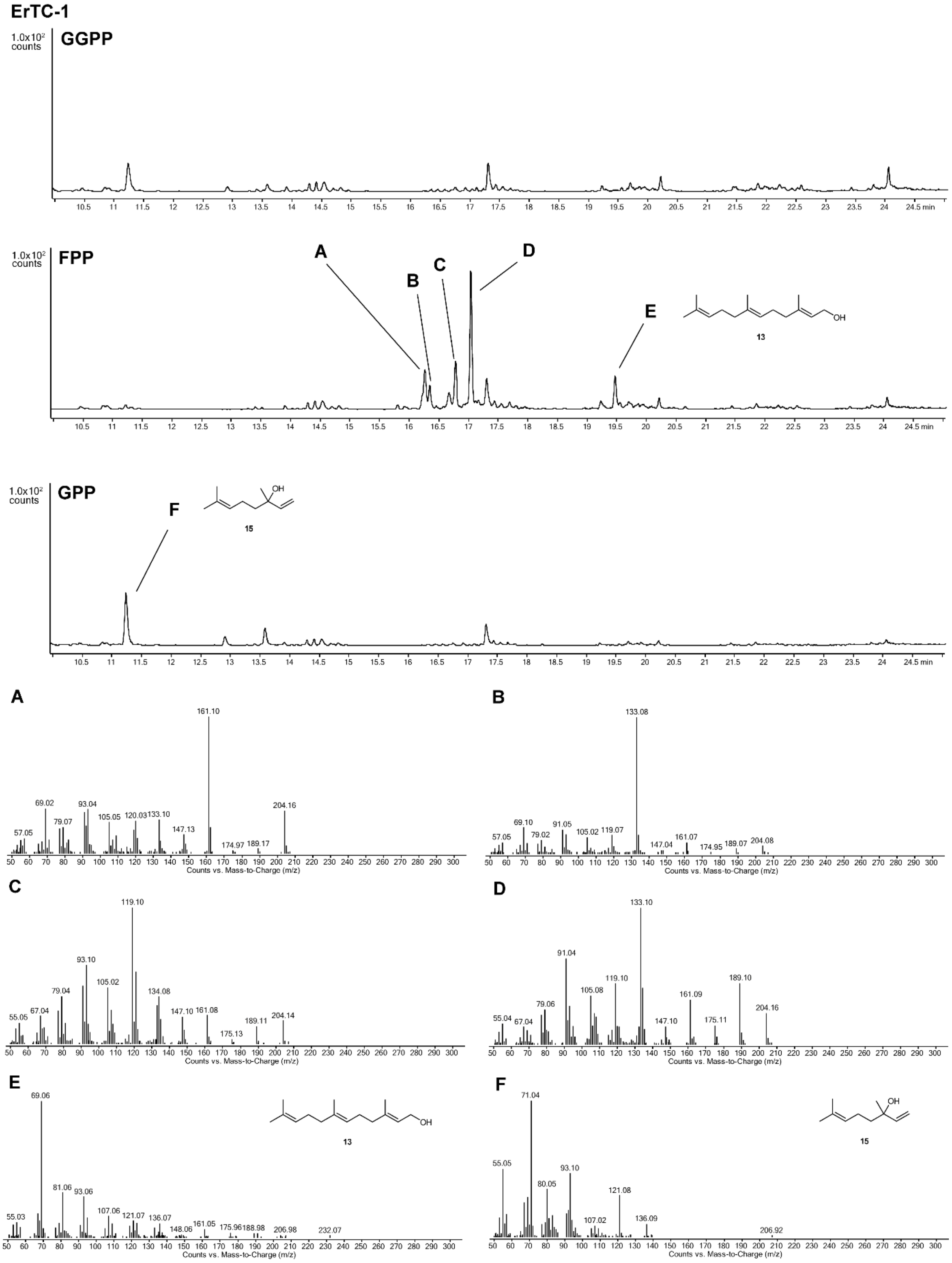
Supplementary Fig. 6. GCMS analysis of CiTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.



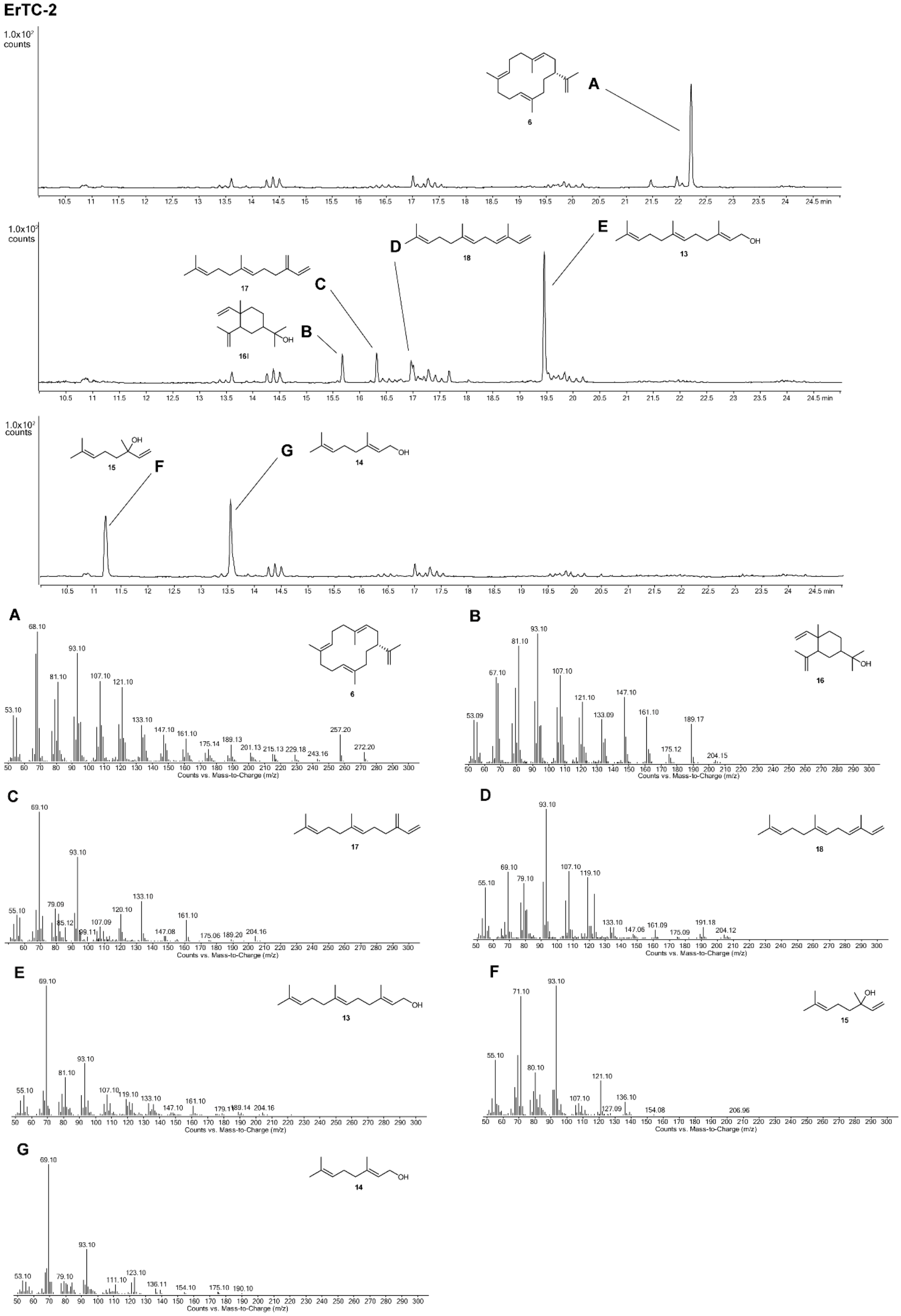
Supplementary Fig. 7. GCMS analysis of DgTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown. The absolute configuration of **9** produced by this enzyme was not determined.



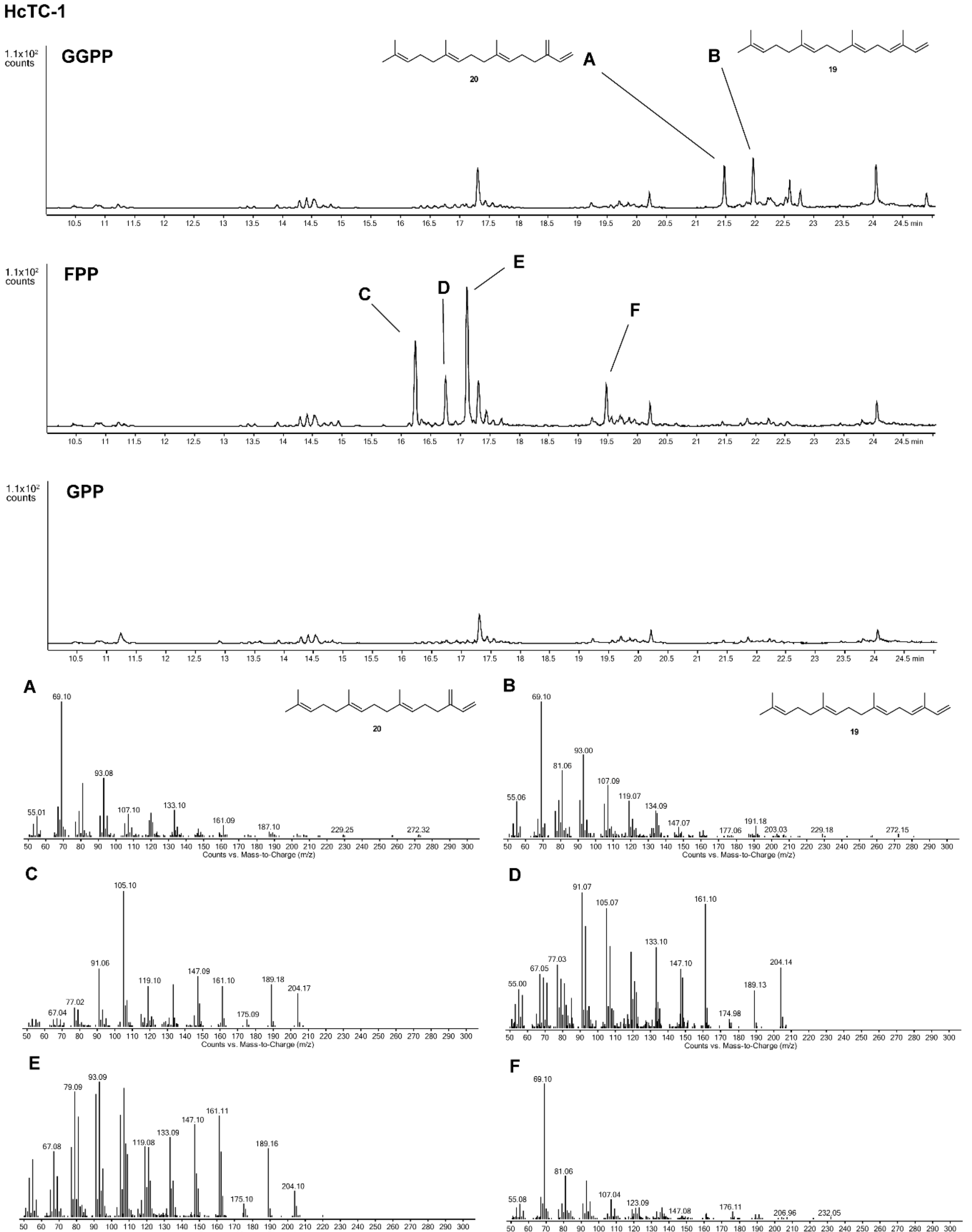
Supplementary Fig. 8. GCMS analysis of DgTC-2 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown. The absolute configuration of **6** produced by this enzyme was not determined.



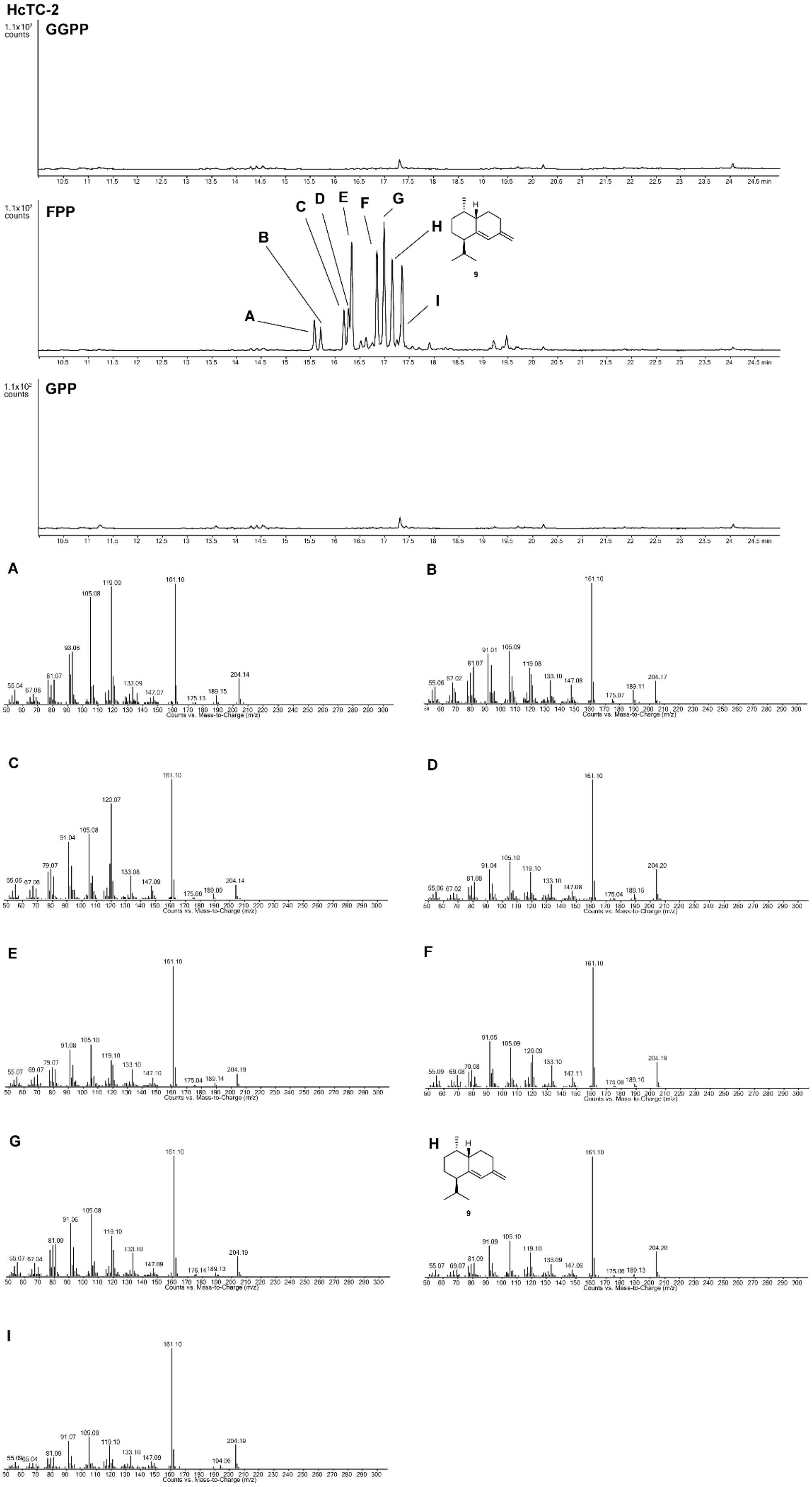
Supplementary Fig. 9. GCMS analysis of ErTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.



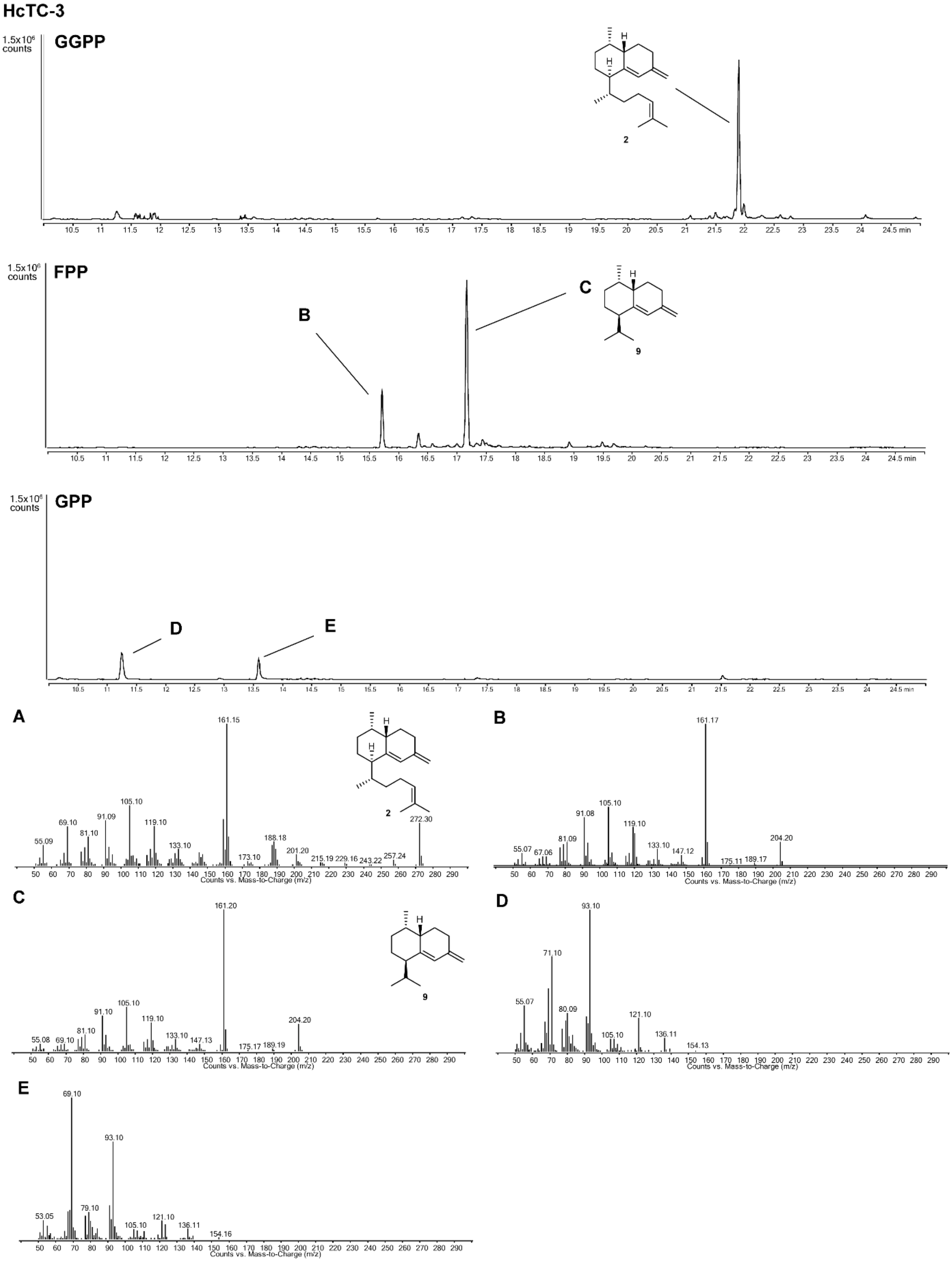
Supplementary Fig. 10. GCMS analysis of ErTC-2 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.



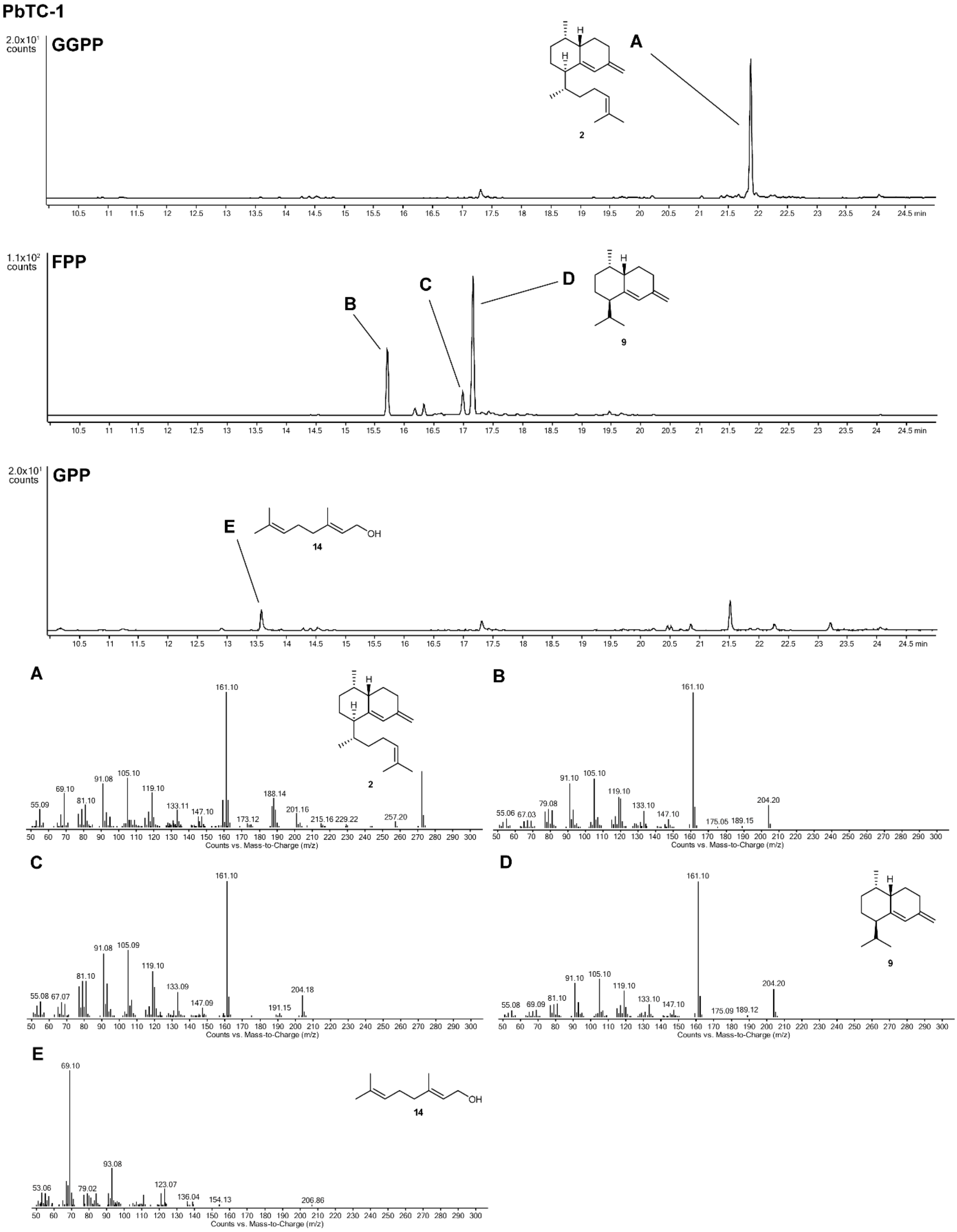
Supplementary Fig. 11. GCMS analysis of HcTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.



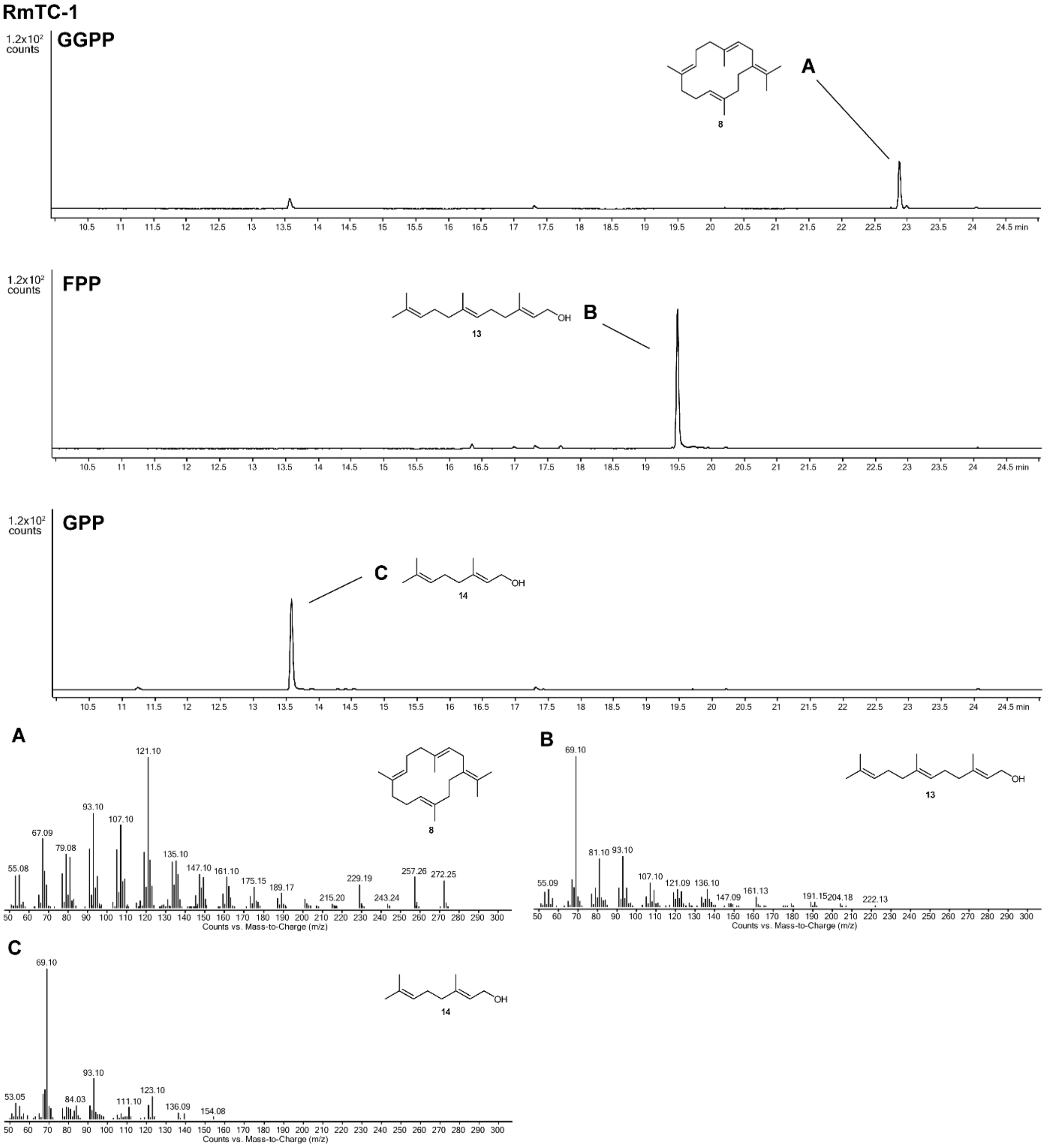
Supplementary Fig. 12. GCMS analysis of HcTC-2 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown. The absolute configuration of **9** produced by this enzyme was not determined.



Supplementary Fig. 13. GCMS analysis of HcTC-3 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

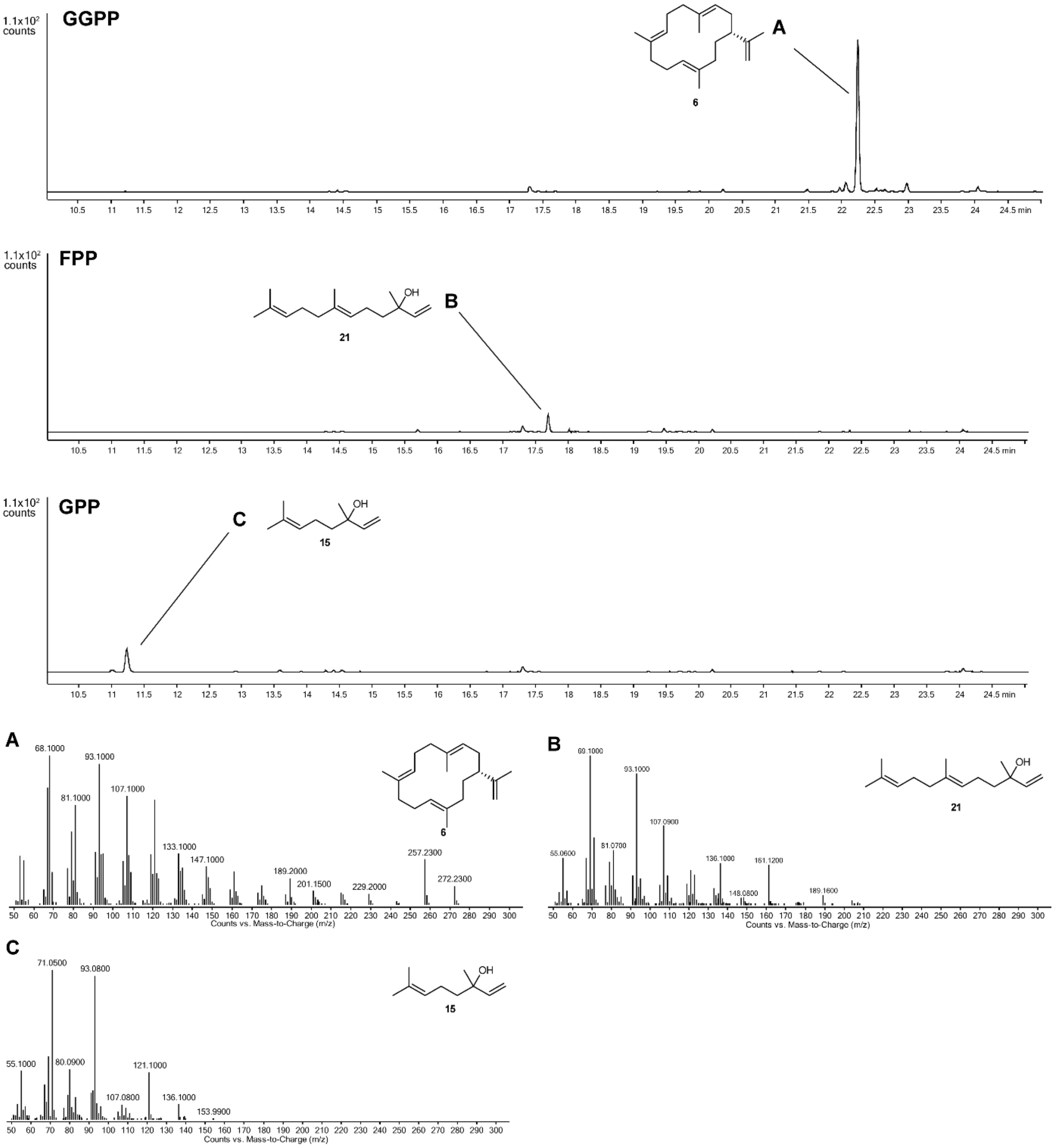


Supplementary Fig. 14. GCMS analysis of PbTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

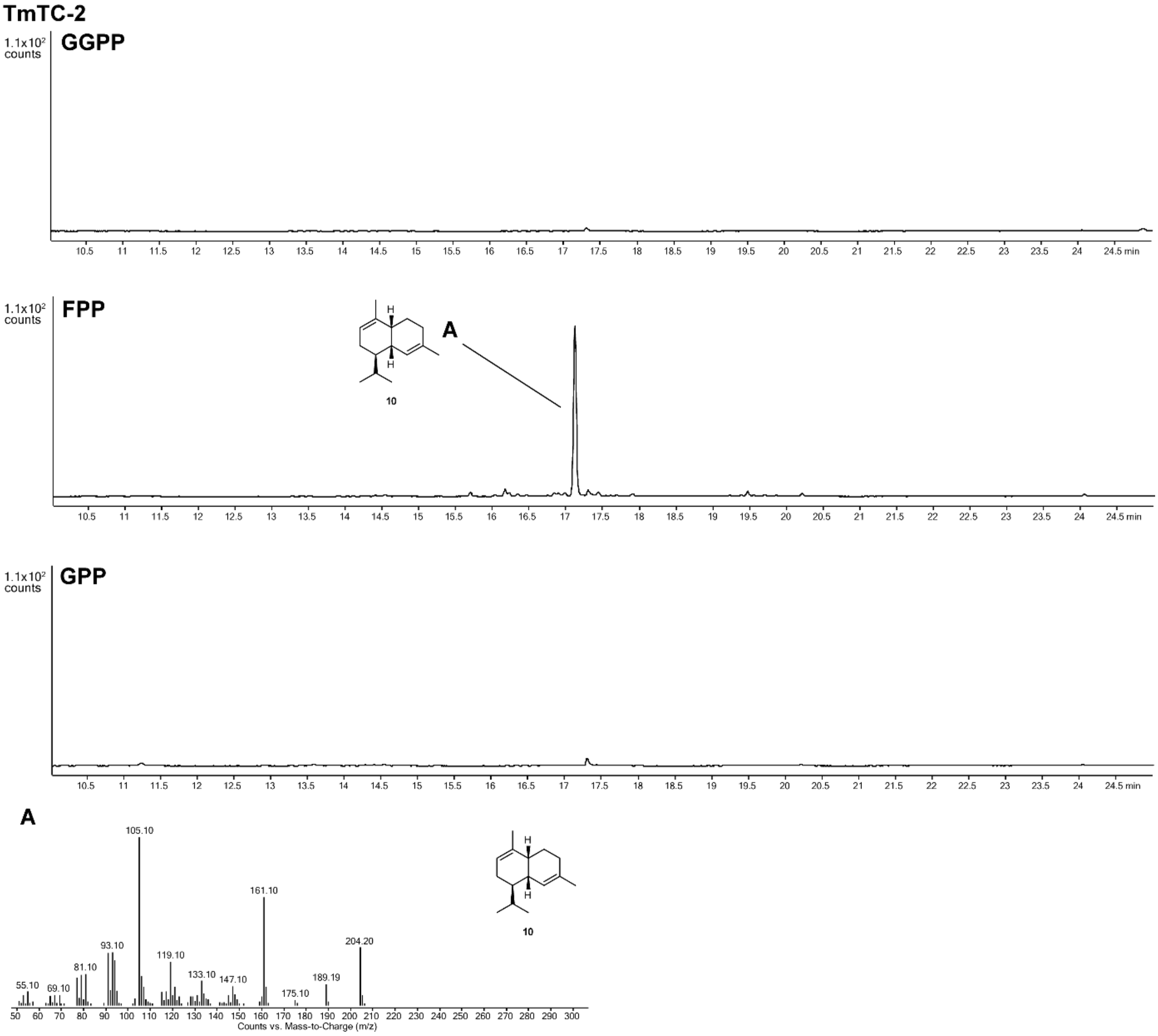


Supplementary Fig. 15. GCMS analysis of RmTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

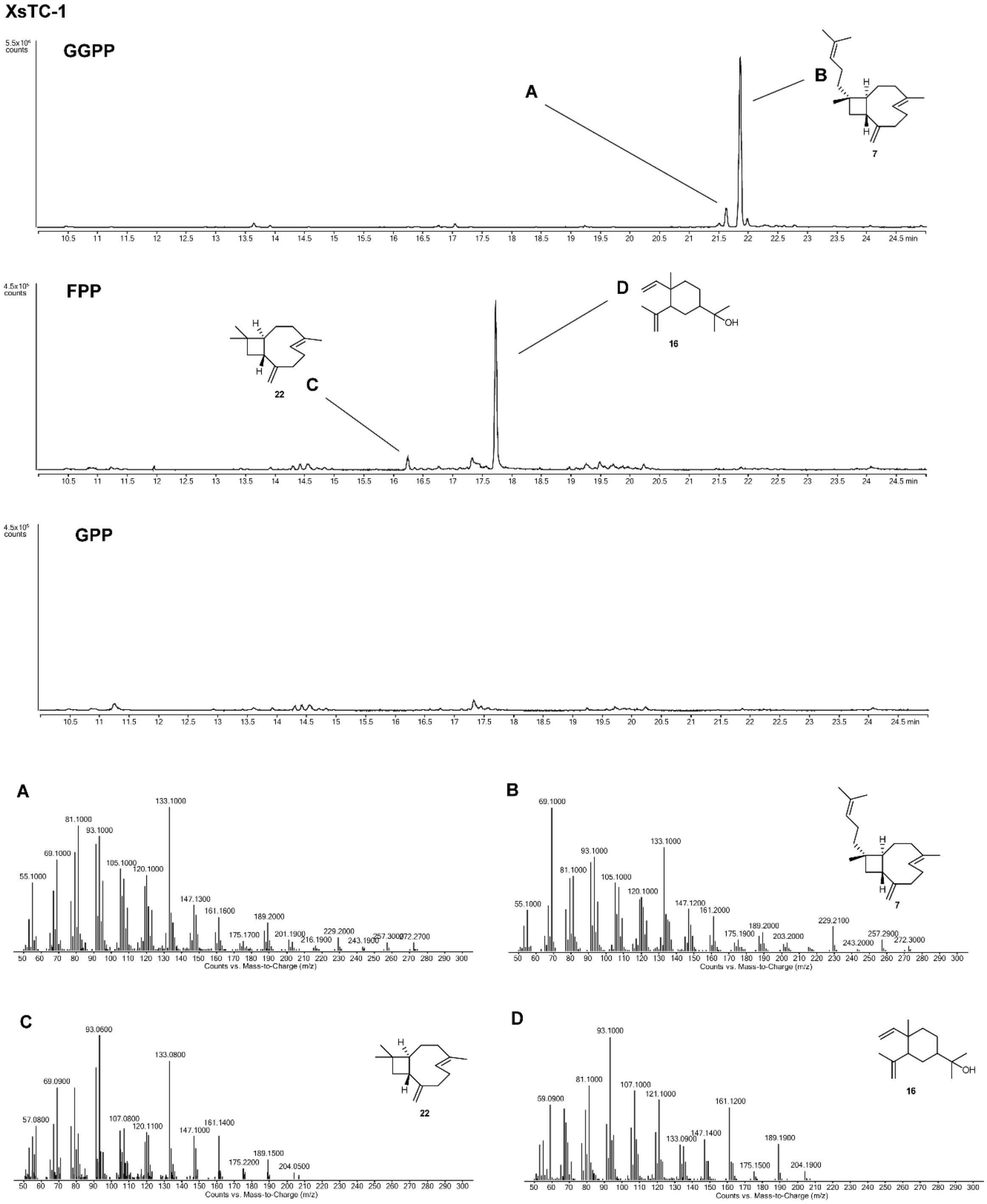
TmTC-1



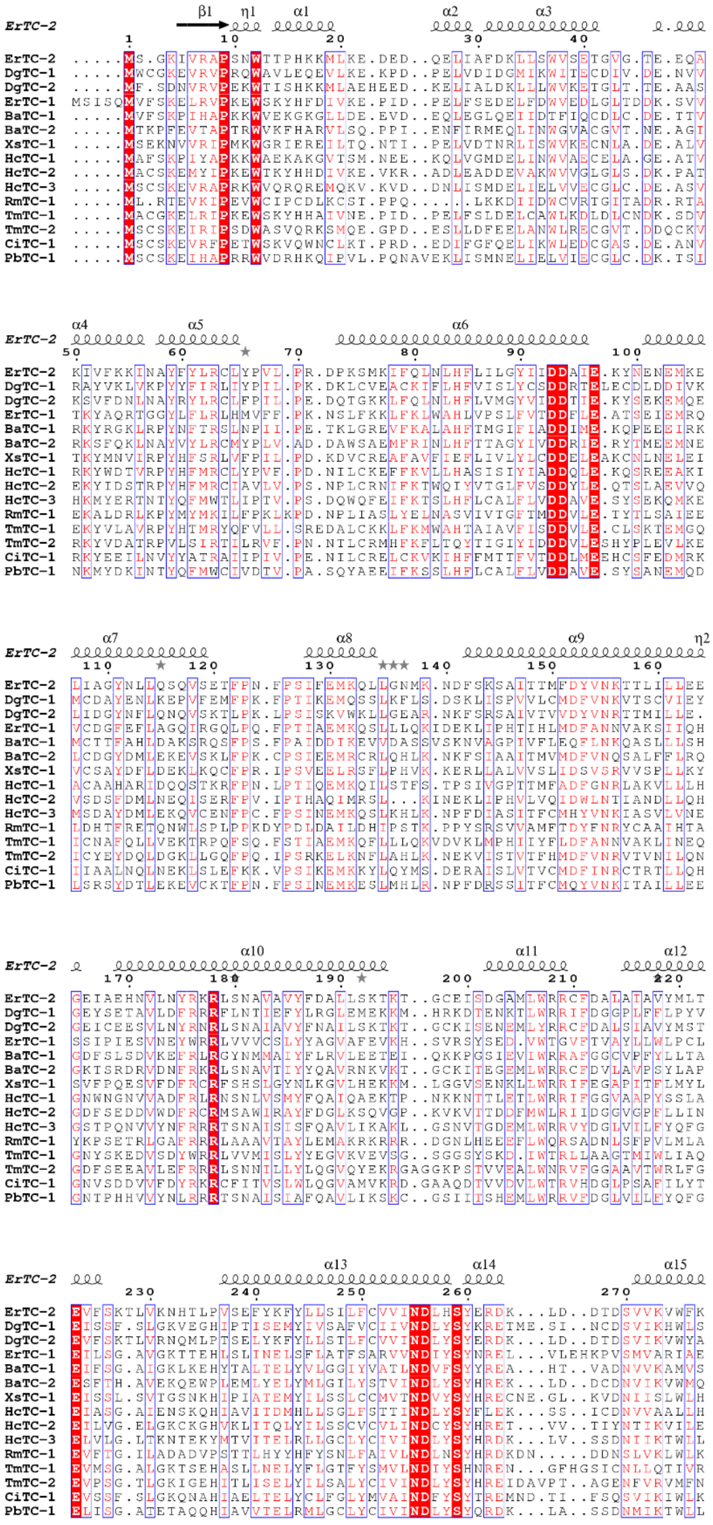
Supplementary Fig. 16. GCMS analysis of TmTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown. The absolute configuration of **6** produced by this enzyme was not determined.

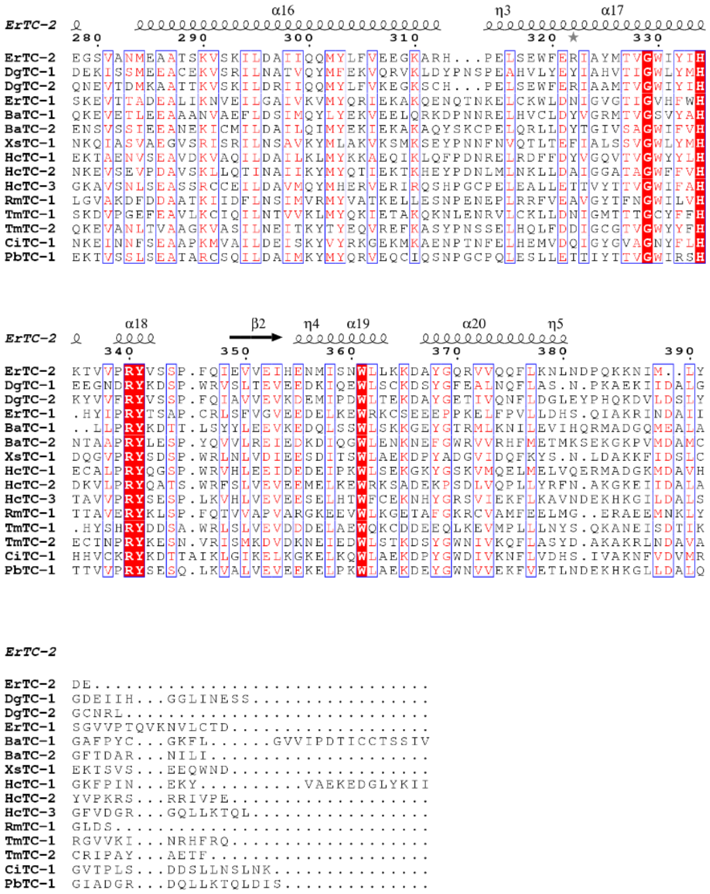


Supplementary Fig. 17. GCMS analysis of TmTC-2 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

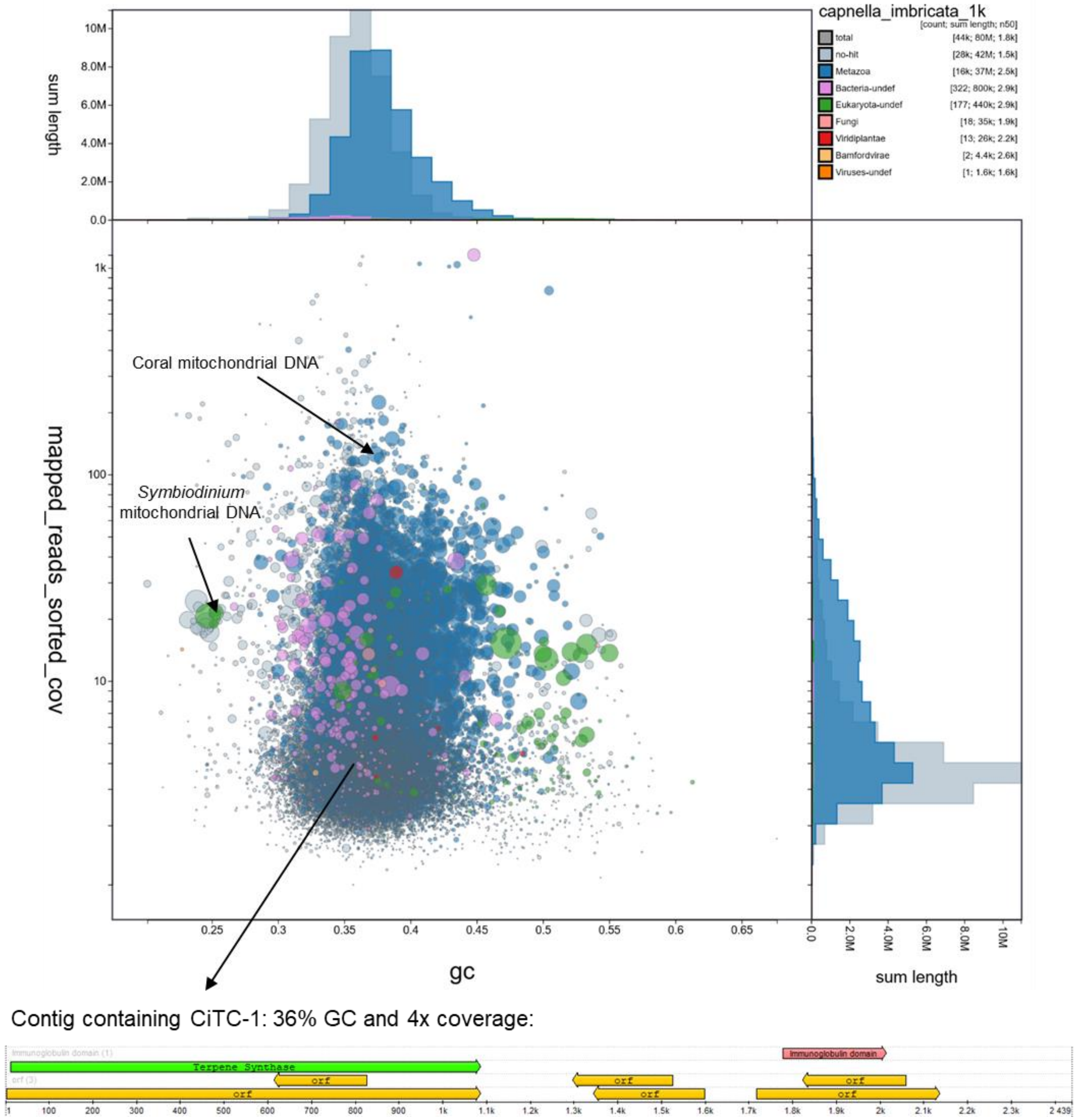


Supplementary Fig. 18. GCMS analysis of XsTC-1 with the substrates GGPP (top) FPP (middle) and GPP (bottom). Mass spectra for major peaks are shown.

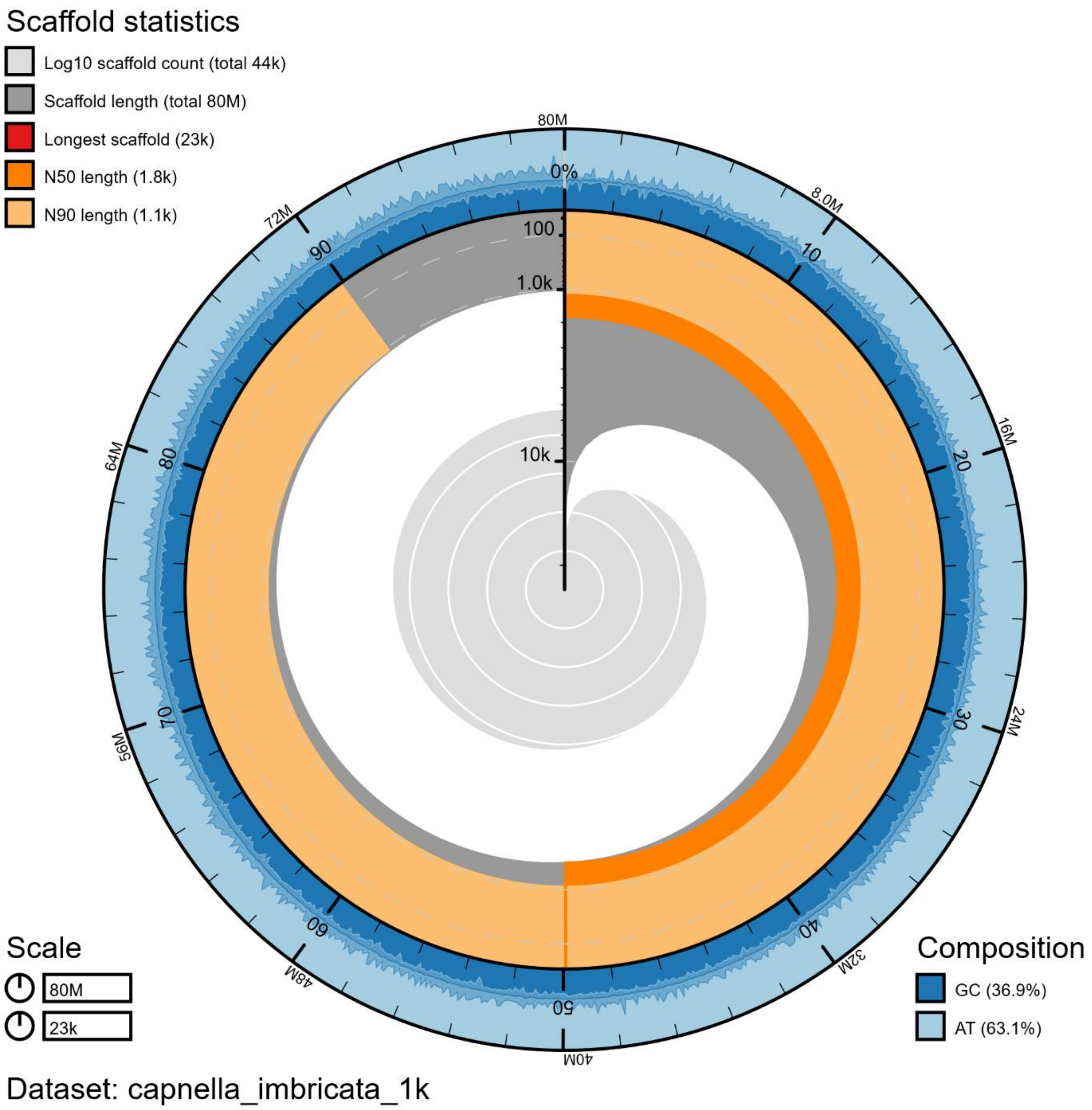

ErTC-2
β1
η1
α1
α2
α3
α4
α5
α6
α7
α8
α9
η2
α10
α11
α12
α13
α14
α15
ErTC-2
DgTC-1
DgTC-2
ErTC-1
BaTC-1
BaTC-2
XsTC-1
HcTC-1
HcTC-2
HcTC-3
RmTC-1
TmTC-1
TmTC-2
CiTC-1
PbTC-1



Supplementary Fig. 19. Sequence alignment of octocoral TC genes characterized in this work. Structural elements observed in the ErTC-2 structure are indicated at the top.



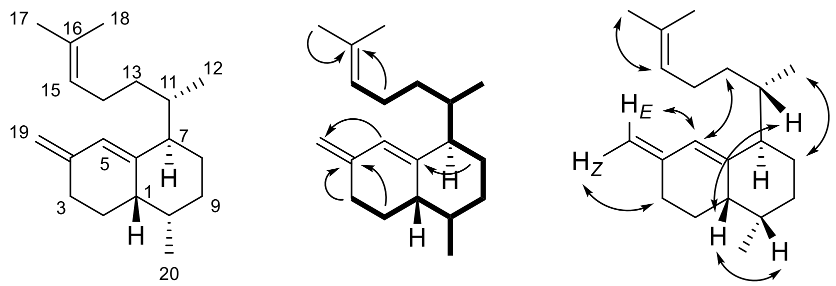
Supplementary Fig. 20: GC-vs-coverage blob plot created using Blobtools[42] for contigs 1 kilobasepairs and larger in our assembly for Capnella imbricata. Contigs are colored by kingdom as classified by Diamond[43] blastx against the NCBI nr database. Circles are sized in proportion to contig length on a square-root scale, ranging from 1,000 to 22,827. Histograms show the distribution of contig length sum along each axis. Most Symbiodinium genomic and plastid DNA did not assemble into contigs > 1 kilobasepairs. The large spread in coverage is likely due to contigs representing repetitive elements. The contig containing CiTC-1 is shown. It also contains a fragment of a gene with an immunoglobulin domain, however this is sometimes not present when assembled using other parameters or assemblers. Only one full-length terpene cyclase gene could be detected in this and any other assembly of *C. imbricata*, and the sequence of CiTC-1 synthesized was a consensus of that in various assemblies.



Supplementary Fig. 21: Snail plot created using Blobtools for contigs 1 kilobasepairs and larger in our assembly for *Capnella imbricata*. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 79,762,010 bp assembly. The distribution of record lengths is shown in dark grey with the plot radius scaled to the longest record present in the assembly (22,827 bp). Orange and pale-orange arcs show the N50 and N90 record lengths (1,772 and 1,104 bp), respectively. The pale grey spiral shows the cumulative record count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. The assembly has been filtered to only include contigs classified as Metazoa or Unclassified by Diamond blastx against the NCBI nr database.

Supplementary Note 1 – Structure elucidation

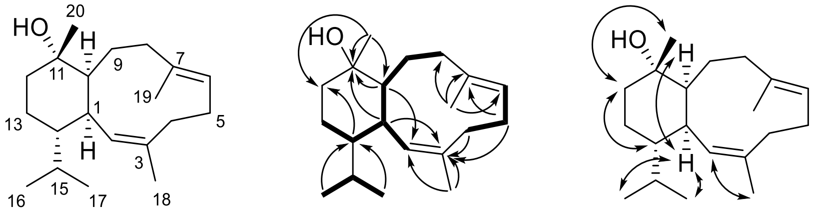
Diterpene product of HcTC-3 (**2**)



Structure of elisabethatriene (**2**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows. **Purification by column chromatography**: Pentanes (100%), **yield**: 1.9 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = +56.8 (*c* 0.055, MeOH). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 273 (7), 272 (31), 257 (4), 201 (8), 190 (4), 189 (11), 188 (17), 187 (14), 162 (20), 161 (100), 160 (11), 159 (33), 147 (7), 146 (6), 145 (9), 134 (4), 133 (11), 132 (5), 131 (8), 129 (5), 128 (4), 121 (4), 120 (8), 119 (27), 118 (5), 117 (11), 115 (7), 109 (5), 107 (7), 106 (6), 105 (42), 104 (3), 103 (4), 95 (7), 93 (13), 92 (4), 91 (31), 82 (5), 81 (20), 79 (12), 77 (9), 69 (27), 67 (9), 65 (4), 55 (14), 53 (5). The relative configuration in agreement with structural revision[44] after first the report [45]. The absolute configuration derived from enantioselectively synthesized oxidized derivatives [46]

C	^{13}C	^{1}H
1	37.1	2.24 (q, ^{3}J = 6.0)
2	27.5	1.63-1.58 (m) 1.53-1.44 (m)
3	29.8	2.36 (ddd, ^{2}J = 14.1, ^{3}J = 9.1, ^{3}J = 4.6) 2.21-2.15 (m)
4	143.9	-
5	127.5	6.14 (s)
6	144.02	-
7	50.0	1.83-1.79 (m)
8	23.2	1.67-1.69 (m, 2H)
9	29.6	1.75-1.67 (m) 1.21 (ddt, ^{2}J = 12.4, ^{3}J = 3.1, ^{3}J = 3.1)
10	35.5	1.79-1.75 (m)
11	31.9	1.74-1.70 (m)
12	17.7	0.9 (d, ^{3}J = 6.6, 3H)
13	35.2	1.56-1.52 1.04 (ddt, ^{2}J = 13.2, ^{3}J = 9.4, ^{3}J = 4.9)
14	25.9	2.15-2.09 (m) 2.00-1.90 (m)
15	125.6	5.12 (ddsept, ^{3}J = 7.1, ^{3}J = 7.1, ^{4}J = 1.5)
16	131	-
17	25.9	1.65 (s, 3H)
18	17.8	1.59 (s, 3H)
19	109. 0	4.86 (s, H_E) 4.78 (s, H_Z)
20	15.2	0.86 (s, 3H)

Diterpene product of BaTC-2 (**4**)

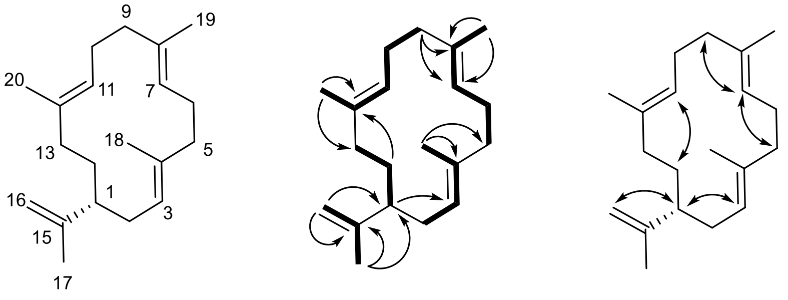


Structure of klysimplexin R (**4**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows. **Purification by column chromatography**: Hexanes/EtOAc (4:1), **yield**: 1.1 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = -23.6 (*c* 0.09, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 290 (5), 281 (9), 272 (17), 257 (11), 230 (13), 229 (83), 229 (12), 207 (29), 207 (15), 205 (14), 202 (16), 201 (9), 191 (9), 190 (9), 189 (42), 187 (16), 175 (9), 173 (21), 164 (17), 164 (10), 163 (47), 162 (14), 162 (10), 161 (38), 161 (14), 160 (9), 159 (35), 155 (9), 150 (19), 149 (40), 148 (19), 148 (15), 147 (33), 147 (21), 145 (24), 137 (10), 136 (15), 136 (12), 135 (48), 134 (19), 134 (14), 133 (50), 131 (17), 127 (15), 124 (9), 123 (46), 123 (27), 122 (42), 122 (32), 121 (100), 120 (15), 119 (46), 119 (20), 117 (13), 113 (13), 111 (26), 110 (16), 109 (56), 109 (50), 108 (26), 108 (12), 107 (69), 107 (46), 106 (14), 105 (82), 99 (20), 97 (12), 96 (13), 95 (72), 94 (13), 93 (48), 93 (46), 92 (9), 92 (11), 91 (58), 91 (25), 85 (44), 84 (15), 83 (19), 83 (11), 82 (15), 81 (77), 81 (31), 80 (12), 79 (80), 77 (35), 77 (16), 71 (82), 70 (14), 69 (77), 68 (13), 68 (10), 67 (45), 67 (11), 65 (10), 58 (15), 57 (54), 56 (14), 55 (9), 55 (64), 53 (24).

Absolute configuration set after optical rotation [47]. NMR spectra show several minor conformers.

C	^{13}C	^{1}H
1	37.1	2.94-2.88 (m)
2	131.5	5.23-5.27 (m)
3	133.7	-
4	32.9	2.18-2.11 (m) 1.77-1.71 (m)
5	26.1	2.25-2.20 (m) 1.98-1.91 (m)
6	125.2	5.34-5.29 (m)
7	138.0	-
8	40.5	2.20-2.14 (m) 1.91-1.84 (m)
9	24.6	1.50-1.45 (m) 1.27-1.22 (m)
10	47.3	1.96-1.91 (m)
11	72.7	-
12	36.3	1.41-1.35 (m, 2H)
13	20.6	1.57-1.50 (m) 1.42-1.36 (m)
14	45.6	0.99-0.94 (m)
15	26.9	1.92-1.86 (m)
16	22.4	0.94 (d, ^{3}J = 6.8, 3H)
17	18.2	0.84 (d, ^{3}J = 6.7, 3H)
18	25.2	1.68 (s, 3H)
19	17.5	1.54 (s, 3H)
20	26.7	1.13 (s, 3H)

Diterpene product of ErTC-2 (**6**)

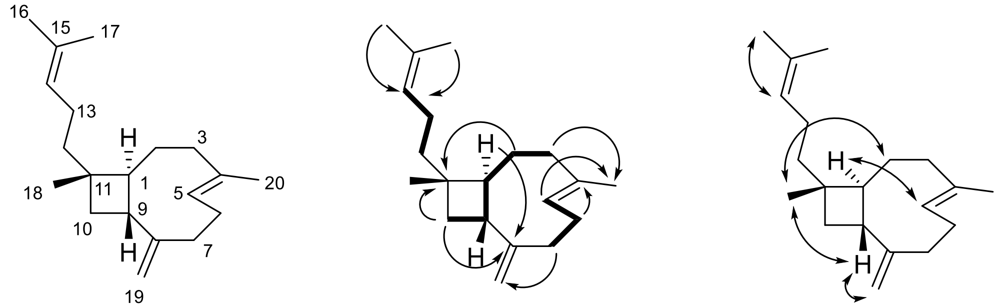


Structure of cembrene A (**6**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows.
Purification by column chromatography: Pentanes (100%), yield: 3.4 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = -28.0 (*c* 0.02, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 272 (7), 272 (3), 258 (6), 257 (21), 229 (4), 216 (6), 215 (6), 207 (5), 203 (4), 201 (7), 191 (4), 190 (4), 189 (17), 187 (6), 176 (5), 175 (7), 175 (6), 173 (6), 163 (5), 162 (8), 161 (4), 161 (17), 160 (4), 159 (5), 159 (4), 149 (9), 148 (15), 147 (25), 145 (6), 137 (3), 136 (6), 136 (4), 135 (21), 134 (20), 133 (33), 131 (4), 123 (16), 122 (17), 121 (66), 120 (22), 119 (6), 119 (25), 117 (3), 109 (25), 108 (33), 107 (70), 106 (13), 105 (28), 96 (5), 95 (33), 94 (36), 93 (93), 92 (4), 92 (15), 91 (37), 82 (9), 81 (67), 80 (17), 79 (48), 78 (3), 77 (25), 69 (22), 68 (100), 67 (76), 66 (4), 65 (11), 55 (31), 53 (33).

The absolute configuration was set after optical rotation [48]

C	^{13}C (δ in ppm)	^{1}H (δ in ppm,)
1	46.5	2.25-2.19 (m)
2	33.0	2.19-2.13 (m) 2.05-1.99 (m)
3	124.5	5.35-5.30 (m)
4	134.9	-
5	39.4	2.18-2.12 (m, 2H)
6	25.3	2.32-2.24 (m) 2.17-2.10 (m)
7	126.5	5.10-5.05 (m)
8	133.5	-
9	39.9	2.12-2.06 (m, 2H)
10	24.2	2.18-2.12 (m, 2H)
11	122.3	5.28-5.23 (m)
12	137.8	-
13	34.3	2.07-2.00 (m) 1.98-1.90 (m)
14	28.5	1.84-1.77 (m) 1.45-1.36 (m)
15	149.2	-
16	110.8	4.84-4.82 (m, 2H)
17	19.4	1.68 (s, 3H)
18	15.6	1.55 (s, 3H)
19	15.4	1.54 (s, 3H)
20	18.3	1.61 (s, 3H)

Diterpene product of XsTC-1 (**7**)

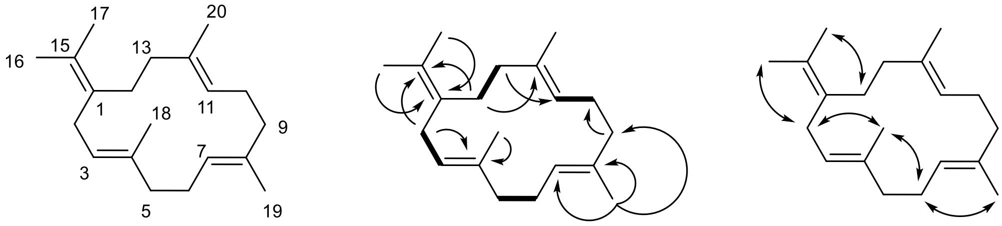


Structure of xeniaphyllene (**7**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows. **Purification by column chromatography**: Pentanes (100%), **yield**: 2.5 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = -12.1 (*c* 0.18, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 272 (3), 257 (7), 230 (4), 229 (17), 203 (6), 201 (4), 190 (5), 189 (13), 188 (4), 187 (10), 175 (7), 173 (6), 162 (6), 161 (24), 160 (4), 159 (11), 149 (10), 148 (18), 147 (29), 146 (5), 145 (9), 137 (8), 136 (20), 135 (21), 134 (25), 133 (72), 132 (3), 131 (9), 123 (21), 122 (11), 121 (30), 120 (37), 119 (36), 118 (4), 117 (8), 115 (5), 109 (23), 108 (10), 107 (45), 106 (26), 105 (48), 103 (4), 96 (4), 95 (34), 94 (13), 93 (66), 92 (18), 91 (63), 83 (6), 82 (10), 81 (52), 80 (12), 79 (51), 78 (8), 77 (29), 70 (6), 69 (100), 68 (10), 67 (32), 66 (4), 65 (10), 55 (28), 53 (17).

The absolute configuration was derived from enantioselectively synthesized oxidized derivatives [49]. NMR spectra show several minor conformers.

C	^{13}C	^{1}H
1	52.6	1.85-1.78 (m)
2	30.3	1.51-1.44 (m) 1.44-1.33 (m)
3	40.3	2.08-2.02 (m) 1.98-1.92 (m)
4	135.1	-
5	124.9	5.39 (dd, ^{3}J = 9.4, 5.4)
6	29.0	2.36-2.32 (m) 2.02-1.96 (m)
7	35.0	2.16-2.11 (m) 2.00-1.94 (m)
8	154.7	-
9	49.0	2.35-2.30 (m)
10	39.7	1.79-1.69 (m, 2H)
11	36.2	-
12	44.6	1.51-1.44 (m) 1.43-1.37 (m)
13	23.4	2.06-1.98 (m, 2H)
14	125.8	5.27-5.20 (m)
15	130.8	-
16	25.2	1.70 (s, 3H)
17	17.7	1.60 (s, 3H)
18	20.1	0.98 (s, 3H)
19	112.3	5.05 (s, H_Z) 4.87 (s, H_E)
20	16.5	1.58 (s, 3H)

Diterpene product of RmTC-1 (**8**)

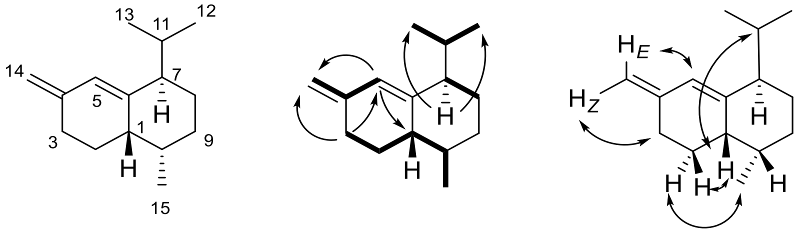


Structure of cembrene C (**8**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows.

Purification by column chromatography: Pentanes (100%), **yield**: 1.1 mg. **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 273 (4), 272 (18), 258 (3), 257 (20), 229 (15), 201 (6), 189 (9), 187 (6), 177 (4), 176 (4), 175 (14), 173 (8), 163 (6), 162 (15), 161 (21), 159 (9), 150 (5), 149 (20), 148 (13), 147 (22), 145 (8), 137 (6), 136 (22), 135 (31), 134 (16), 133 (32), 131 (5), 123 (15), 122 (32), 121 (100), 120 (13), 119 (36), 117 (5), 115 (3), 109 (19), 108 (18), 107 (55), 106 (10), 105 (39), 103 (3), 95 (21), 94 (13), 93 (64), 92 (7), 91 (39), 83 (6), 82 (4), 81 (34), 80 (8), 79 (36), 78 (3), 77 (23), 69 (15), 68 (24), 67 (46), 65 (8), 57 (3), 55 (21), 53 (22).

C	^{13}C	^{1}H
1	124.74	-
2	31.0	2.91 (d, ^{3}J = 7.0, 2H)
3	125.3	5.18 (tq, ^{3}J = 6.9, ^{4}J = 1.4)
4	133.7	-
5	39.3	2.14-2.10 (m, 2H)
6	25.3	2.17 (dt, ^{3}J = 6.3, ^{3}J = 8.7, 2H)
7	126.8	5.05 (tq, ^{3}J = 6.5, ^{4}J = 1.4)
8	133.1	-
9	40.4	2.09-2.05 (m, 2H)
10	24.3	2.11-2.08 (m, 2H)
11	124.66	5.16 (tq, ^{3}J = 6.9, ^{4}J = 1.4)
12	134.8	-
13	37.9	2.10-2.07 (m, 2H)
14	31.9	2.30 (t, ^{3}J = 7.7, 2H)
15	131.7	-
16	20.7	1.67 (s, 3H)
17	20.9	1.70 (s, 3H)
18	15.4	1.56 (s, 3H)
19	15.1	1.52 (s, 3H)
20	16.0	1.61 (s, 3H)

Sesquiterpene product of HcTC-3 (**9**)

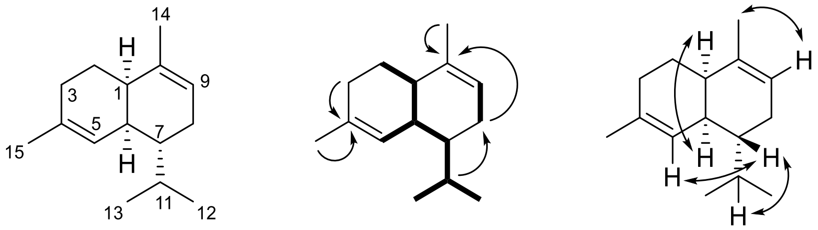


Structure of nephthene (**9**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows. **Purification by column chromatography**: Pentanes (100%), **yield**: 1.3 mg **Optical rotary power**: $[\alpha]_D^{25.0}$ = +5.3 (*c* 0.19, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 204 (20), 162 (15), 161 (100), 147 (5), 133 (9), 131 (3), 120 (7), 119 (19), 117 (8), 115 (6), 107 (4), 106 (4), 105 (27), 103 (3), 93 (10), 92 (4), 91 (24), 81 (9), 79 (9), 77 (8), 69 (4), 67 (4), 65 (3), 55 (4).

The absolute configuration was derived from **2**., sample contains a second sesquiterpene (~20%) of unknown structure, which is not separable by column chromatography.

C	^{13}C	^{1}H
1	37.1	2.13-2.07, (m)
2	27.5	1.62-1.53 (m) 1.53-1.44 (m)
3	30.1	2.35 (ddd, *J* = 14.5, 8.4, 4.6 Hz) 2.19-2.12 (m)
4	144.1	-
5	127.3	6.09 (s)
6	144.2	-
7	51.5	1.69-1.62 (m)
8	23.1	1.70-1.64 (m) 1.64-1.56 (m)
9	29.6	1.70-1.63 (m) 1.25-1.17 (m)
10	35.2	1.78-1.75 (m)
11	27.5	1.75-1.69 (m)
12	21.0	0.79 (d, *J* = 6.4, 3H)
13	21.5	0.88 (d, *J* = 6.3, 3H)
14	108.9	4.88 (s, H_E) 4.80 (s, H_Z)
15	15.0	0.84 (d, *J* = 7.0, 3H)

Sesquiterpene product of TmTC-2 (**10**)

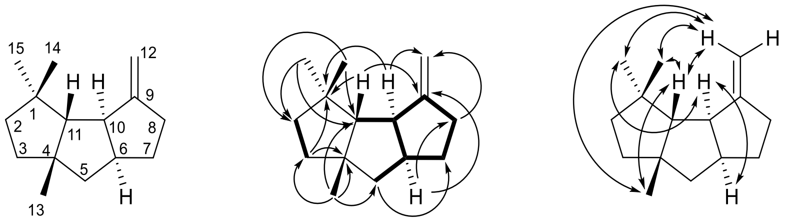


Structure of a-muuroladiene (**10**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows.
Purification by column chromatography: Pentanes (100%), **yield**: 2.5 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = +55.0 (*c* 0.25, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 205 (5), 204 (33), 189 (12), 162 (10), 161 (60), 160 (5), 149 (3), 148 (6), 147 (11), 145 (6), 135 (3), 134 (6), 133 (14), 131 (6), 129 (5), 128 (5), 123 (5), 121 (10), 120 (5), 119 (25), 117 (8), 115 (7), 107 (10), 106 (17), 105 (100), 103 (3), 95 (8), 94 (27), 93 (31), 92 (8), 91 (30), 81 (18), 80 (3), 79 (18), 78 (4), 77 (16), 69 (5), 67 (5), 65 (5), 55 (8), 53 (6).

The absolute configuration was set after optical rotation [50].

C	^{13}C	^{1}H
1	39.4	2.02-1.96 (m)
2	25.0	1.82-1.74 (m) 1.53-1.43 (m)
3	30.7	1.90-1.79 (m, 2H)
4	134.4	-
5	124.7	5.59-5.55 (m)
6	37.2	2.16-2.10 (m)
7	41.5	1.57-1.48 (m)
8	24.9	1.91-1.84 (m) 1.82-1.74 (m)
9	122.6	5.47-5.44 (m)
10	136.1	-
11	27.0	2.07 (dsept, ^3J = 6.9, ^3J = 4.4)
12	21.6	0.88 (d, ^3J = 6.9, 3H)
13	16.1	0.87 (d, ^3J = 6.8, 3H)
14	21.9	1.70 (s, 3H)
15	24.1	1.67 (s, 3H)

Sesquiterpene product of CiTC-2 (**11**)



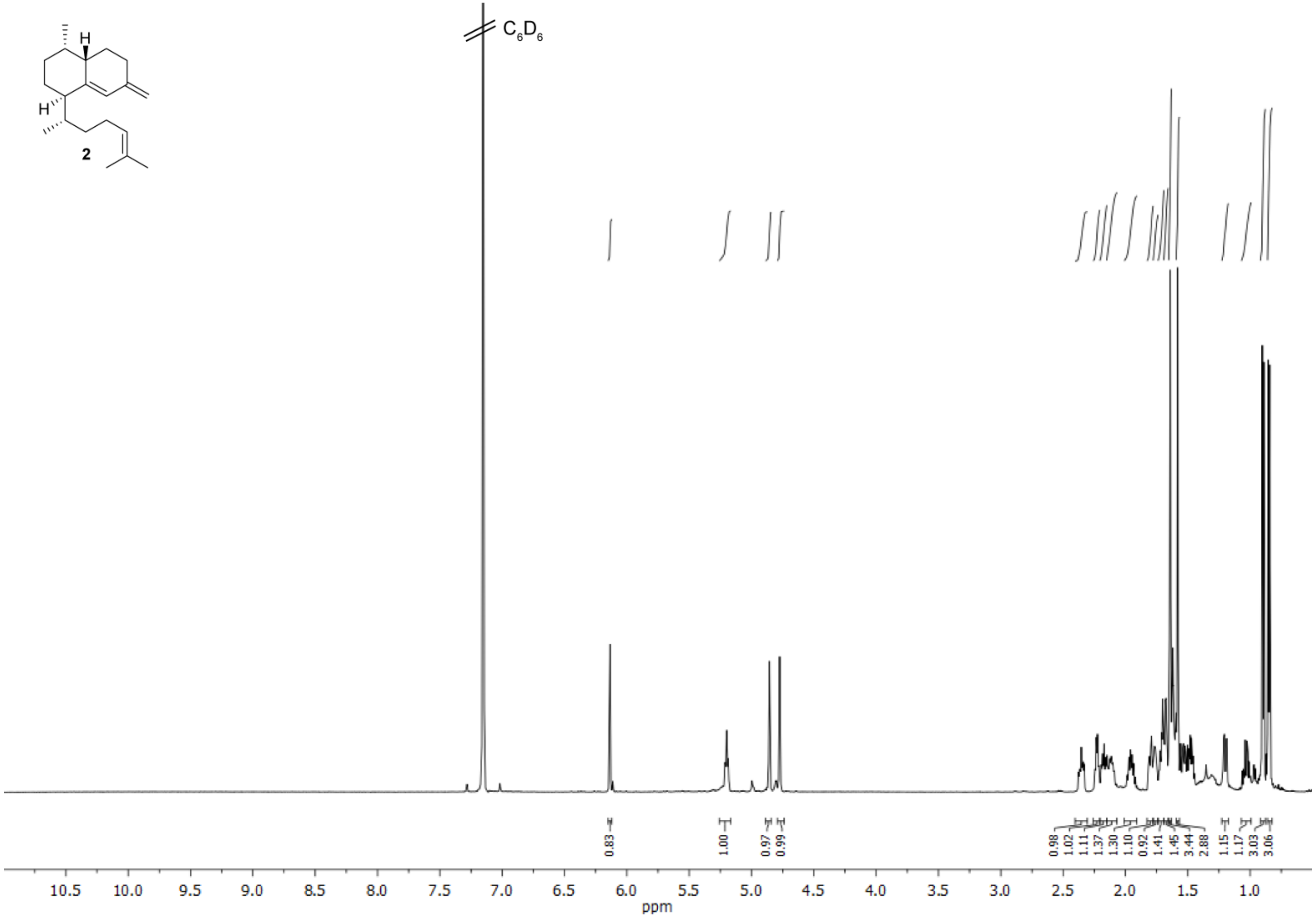
Structure of capnellene (**11**). COSY correlations are shown as bold lines, key HMBC correlations are shown as single headed arrows. Key NOE correlations are shown as double headed arrows.

Purification by column chromatography: Pentanes (100%), **yield**: 2 mg. **Optical rotary power**: $[\alpha]_D^{25.0}$ = -93.7 (*c* 0.16, $CHCl_3$). **NMR**: See table. **EI-MS** (70 eV): m/z (%) = 204 (6), 189 (14), 163 (7), 149 (3), 148 (21), 147 (6), 136 (6), 135 (46), 134 (18), 133 (45), 121 (3), 120 (3), 119 (12), 117 (4), 115 (4), 111 (4), 110 (6), 109 (42), 107 (13), 106 (9), 105 (15), 95 (18), 94 (9), 93 (19), 92 (5), 91 (28), 81 (25), 80 (100), 79 (30), 78 (5), 77 (18), 69 (10), 67 (9), 65 (6), 55 (11), 53 (7).

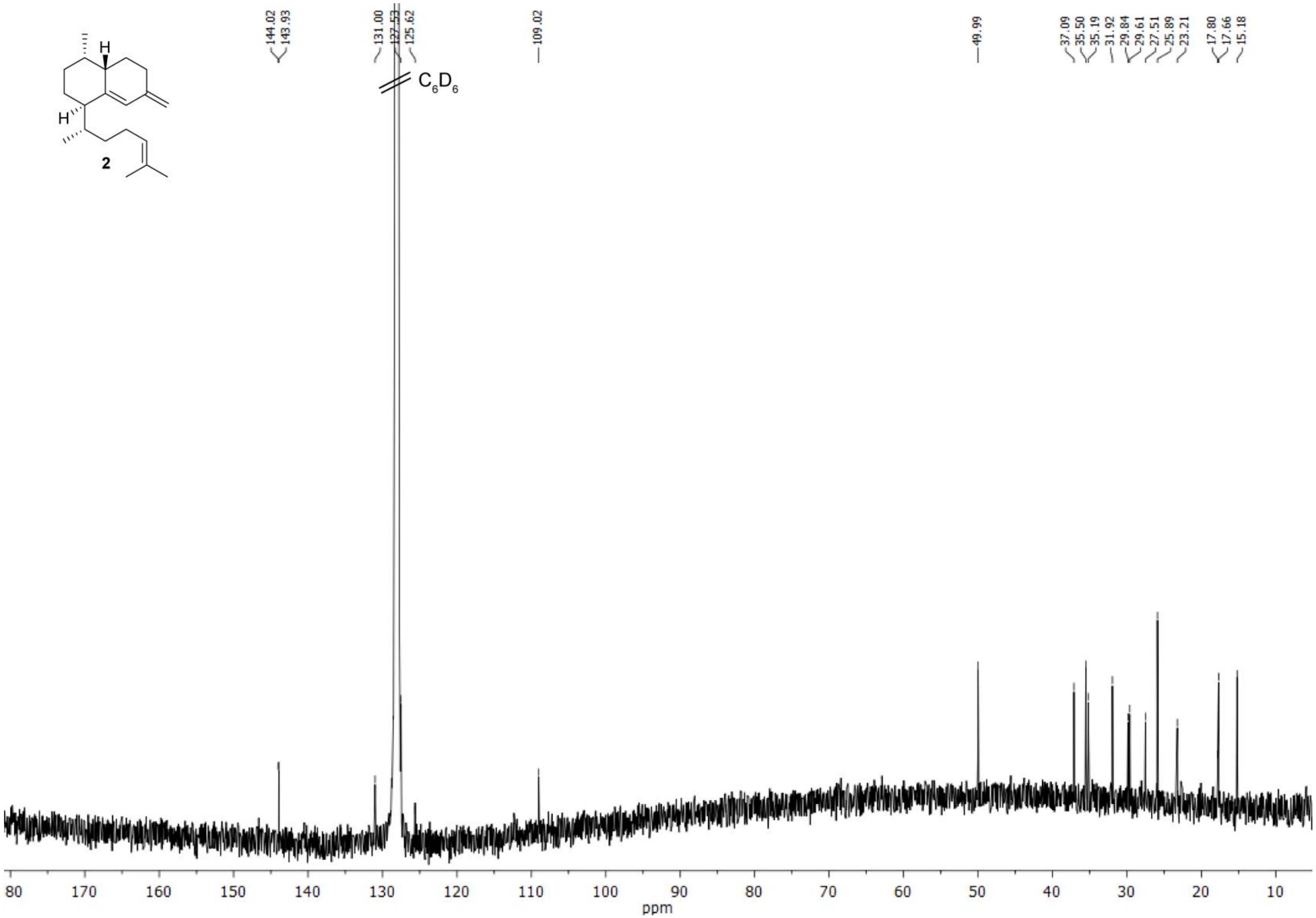
The absolute configuration set after optical rotation [51].

C	^{13}C	^{1}H
1	42.2	-
2	41.7	1.44 (m, 2H)
3	40.6	1.52 (m, 2H)
4	53.2	-
5	47.9	1.61 (m) 1.23 (dd, *J* = 13.1, 9.4 Hz)
6	46.1	2.44 (m)
7	29.0	1.63 (m) 1.43 (m)
8	31.5	2.46 (m) 2.34 (m)
9	158.3	-
10	52.3	2.67 (dt, *J* = 7.8, 2.6 Hz)
11	69.0	1.8 (d, *J* = 2.2 Hz)
12	105.4	5.04 (q, *J* = 2.4, H_E) 4.97 (q, *J* = 2.4, H_Z)
13	31.8	1.22 (s, 3H)
14	30.7	1.07 (s, 3H)
15	25.8	0.96 (s, 3H)

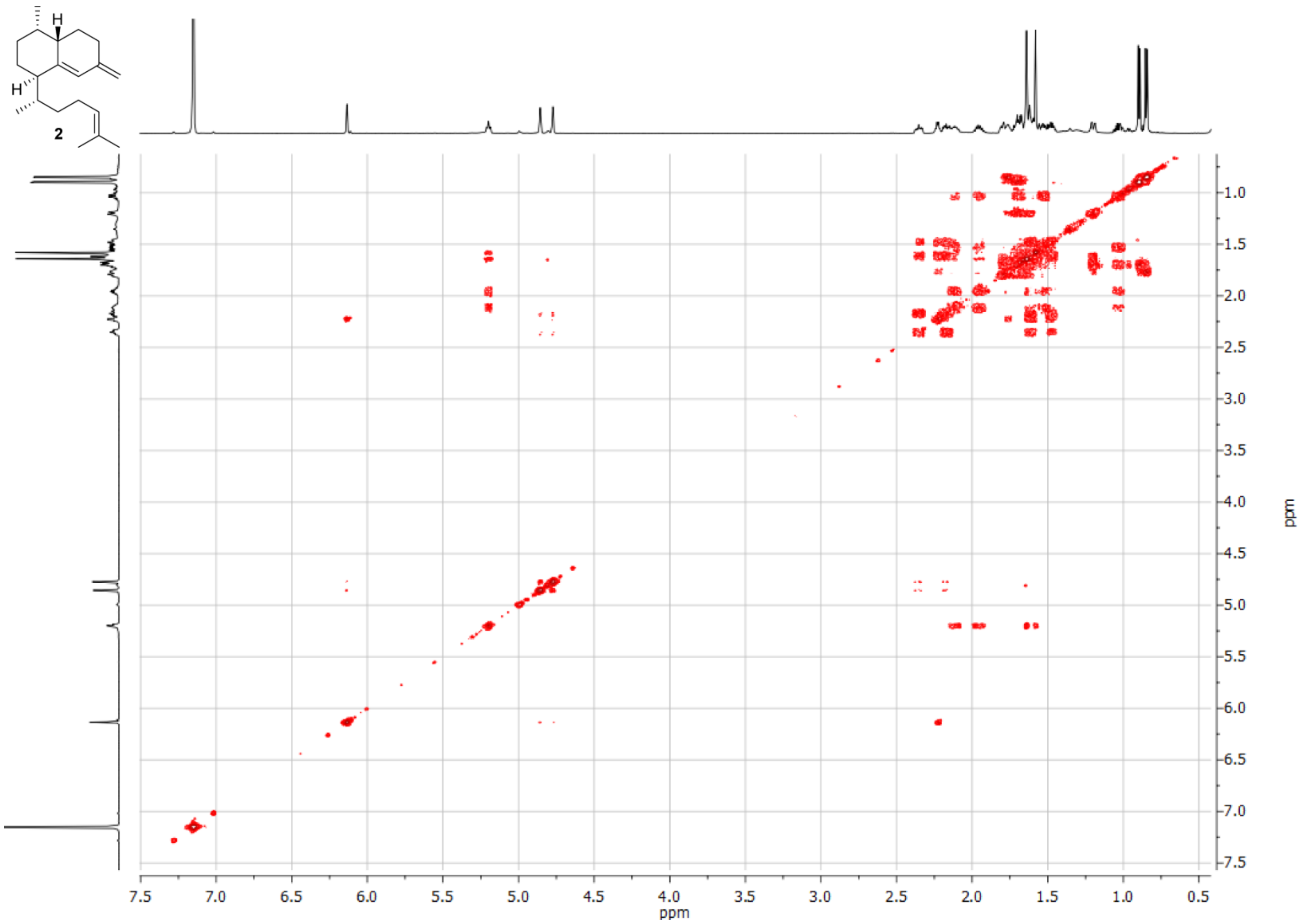
NMR spectra



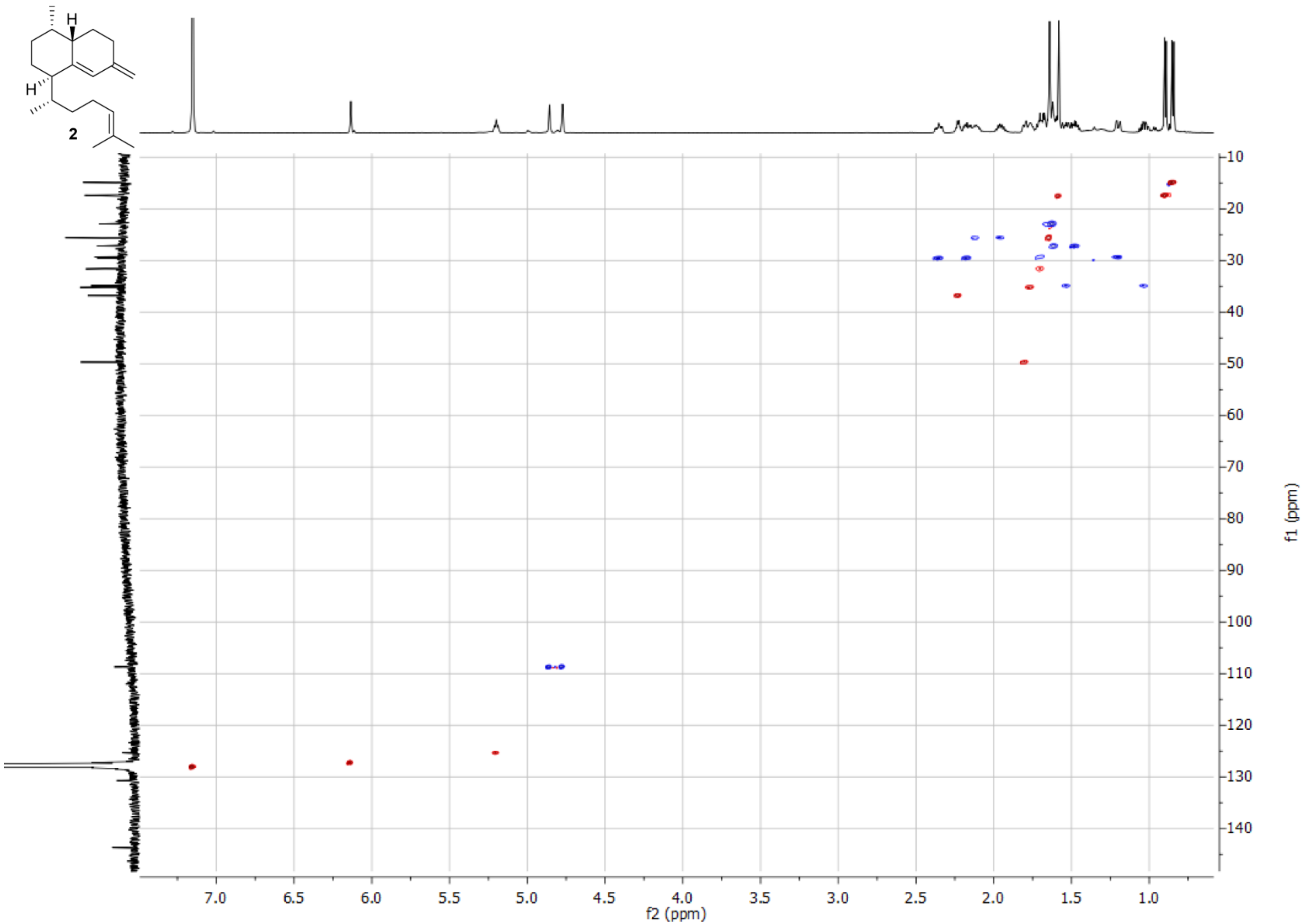
^{1}H NMR spectrum (600 MHz) of **2**.



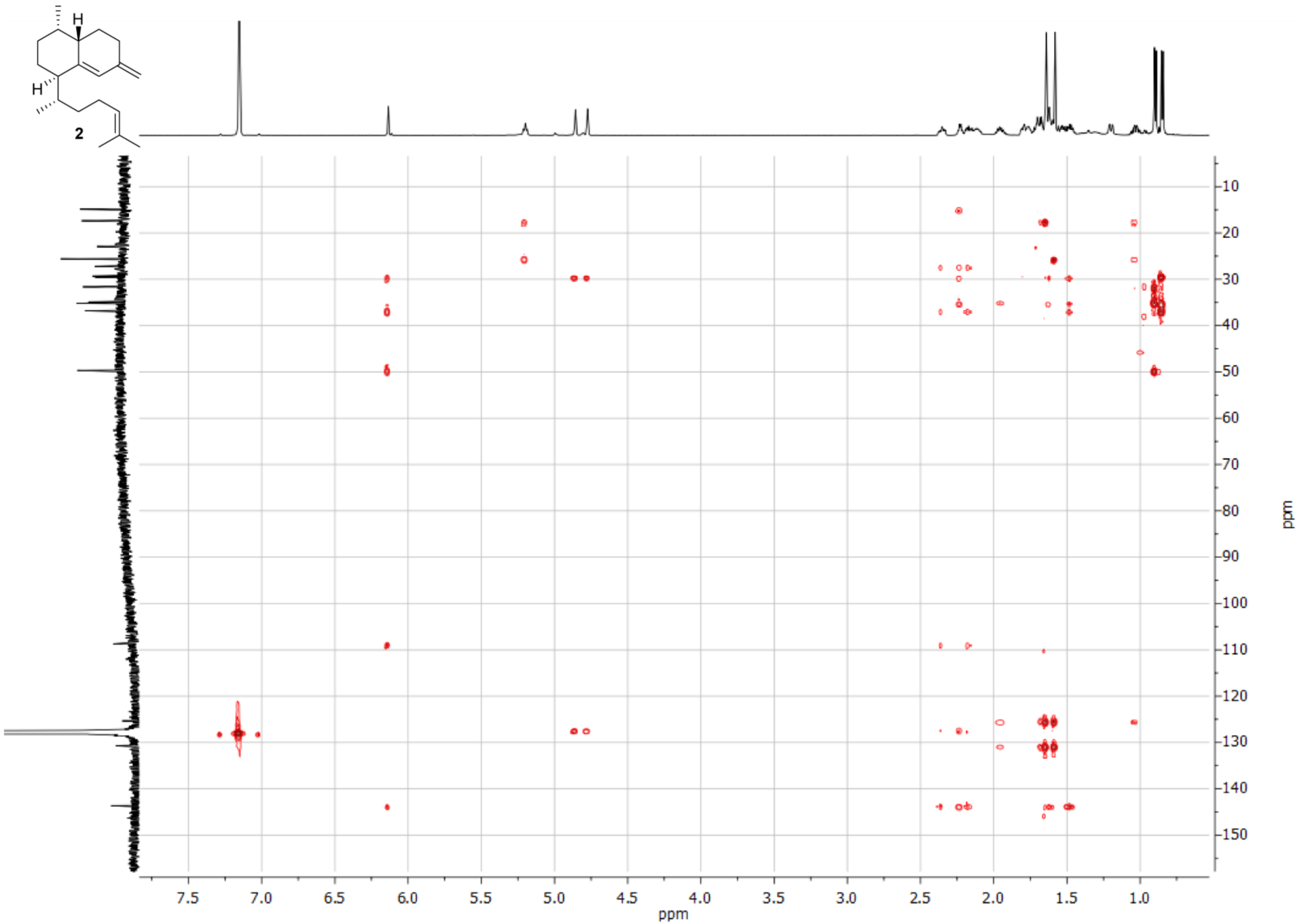
^{13}C NMR spectrum (125 MHz) of **2**.



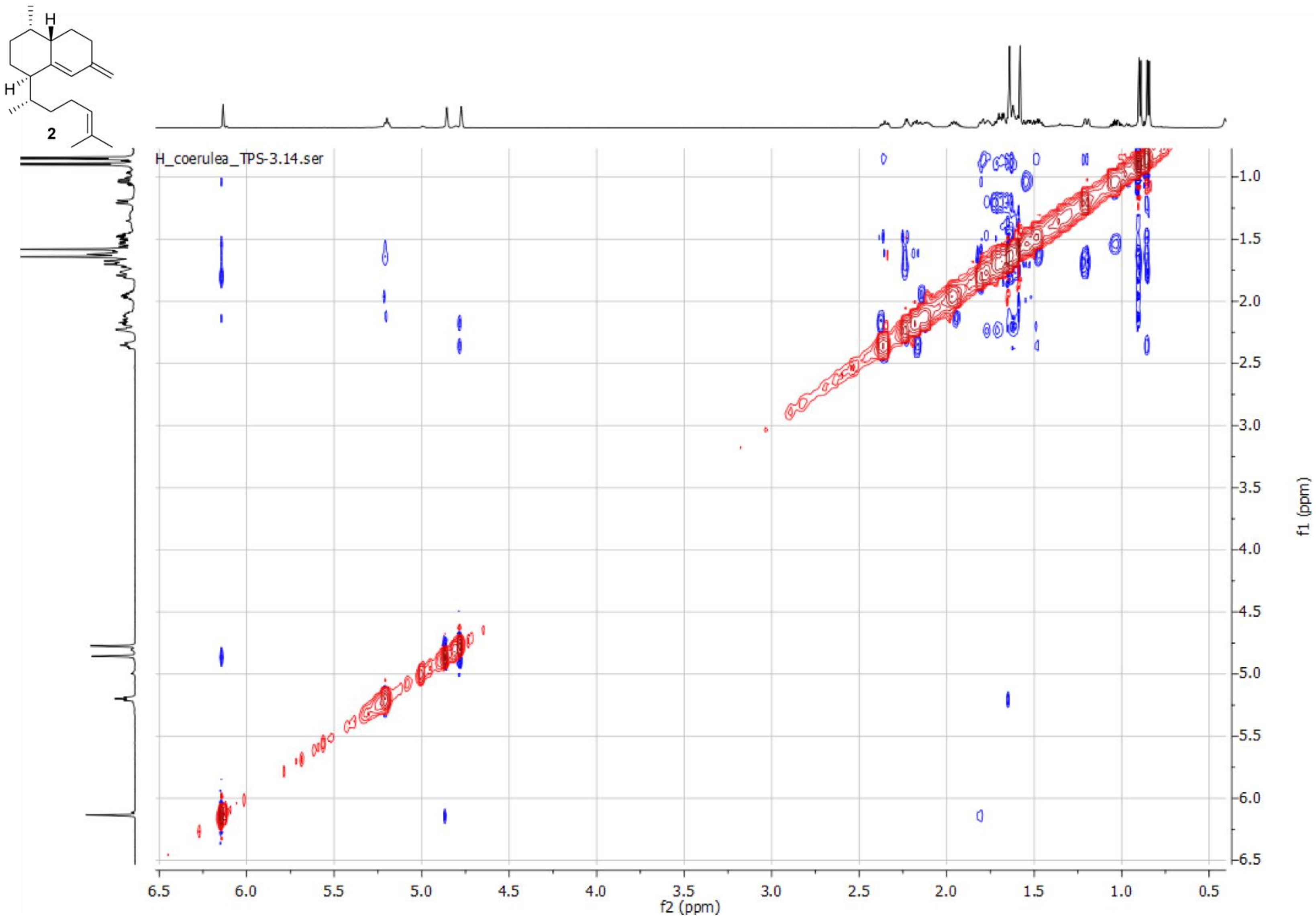
^{1}H, ^{1}H COSY spectrum (600 MHz) of **2**.



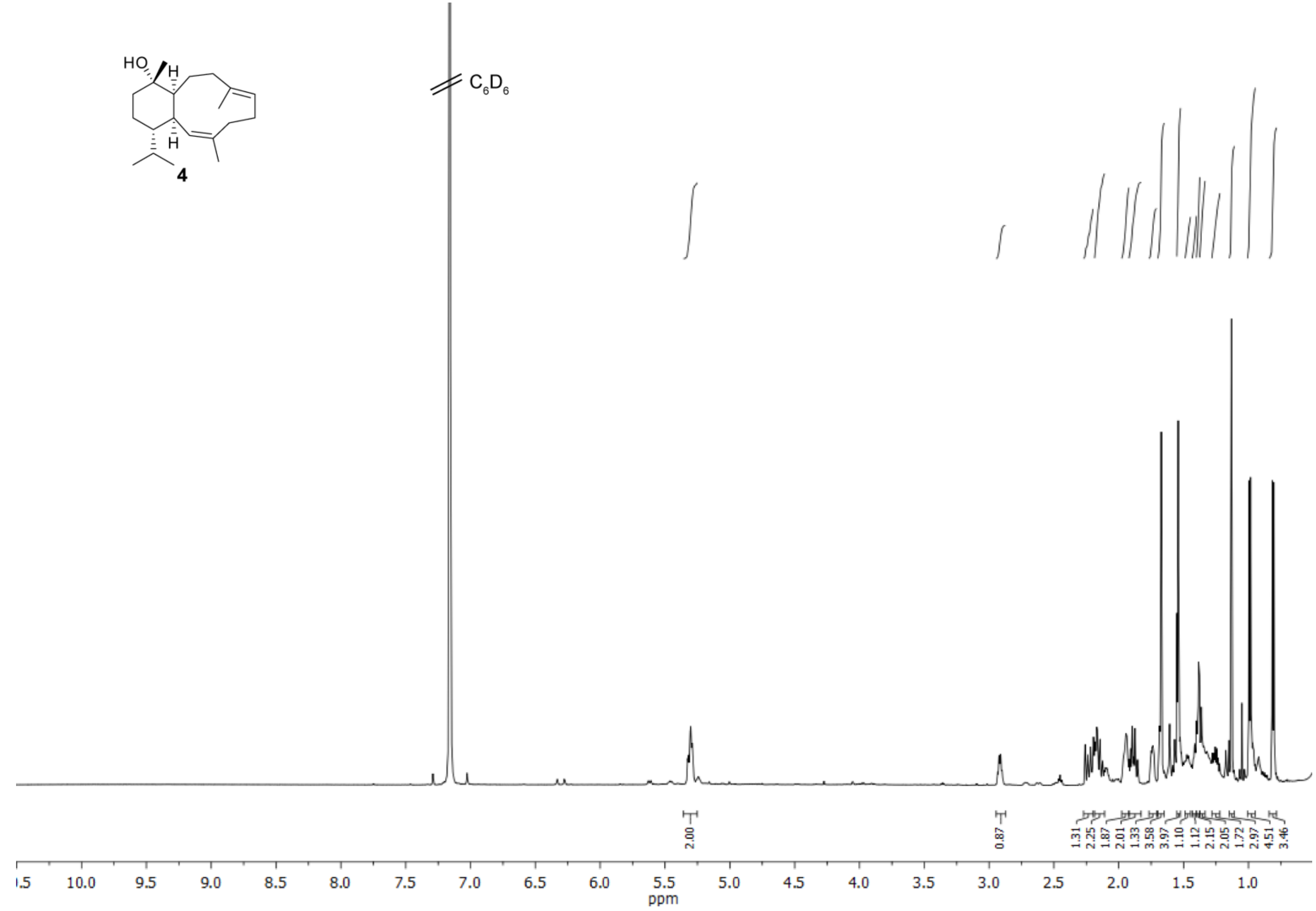
^{1}H, ^{13}C HSQC spectrum (600 MHz) of **2**



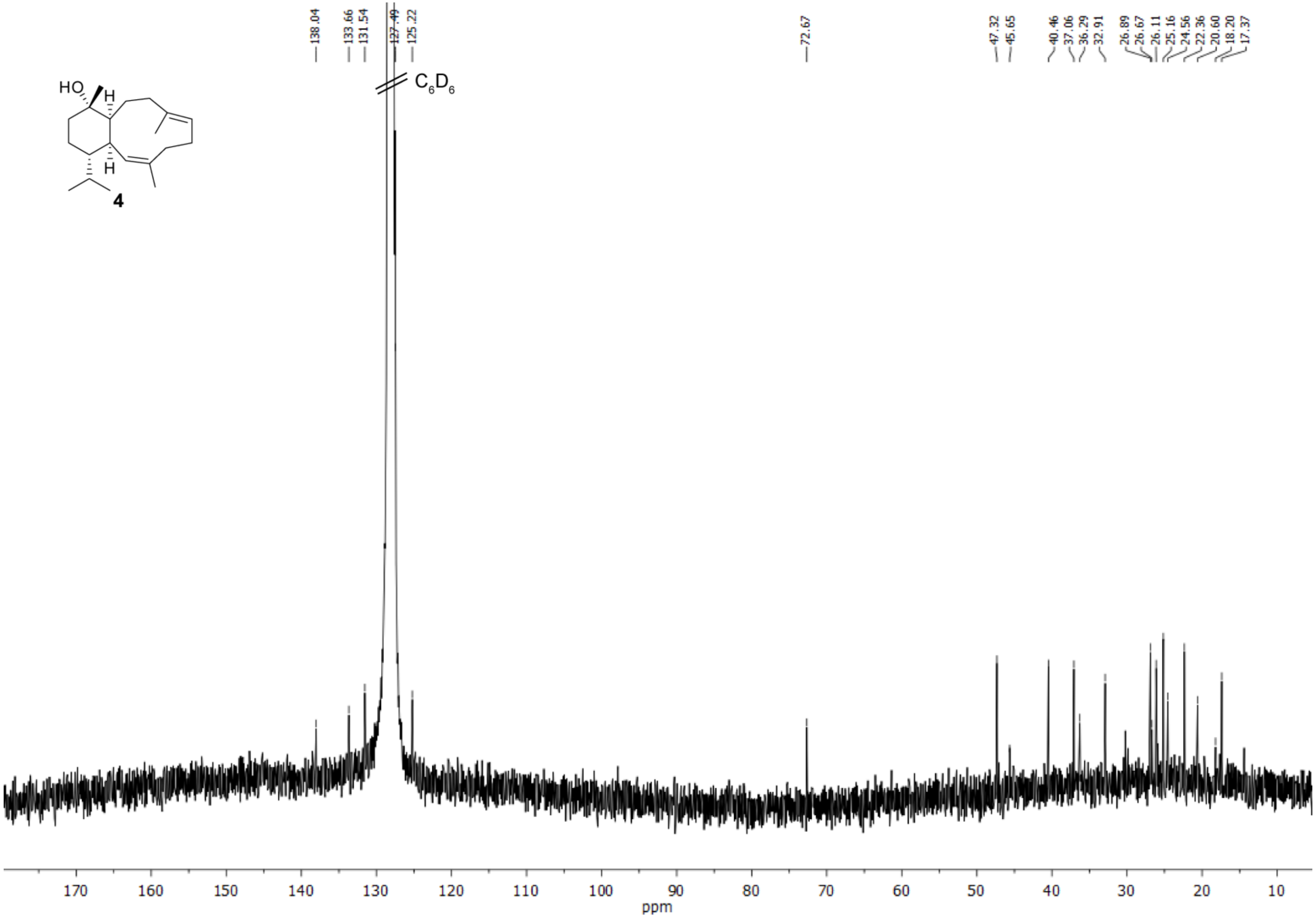
^{1}H, ^{13}C HMBC spectrum (600 MHz) of **2**



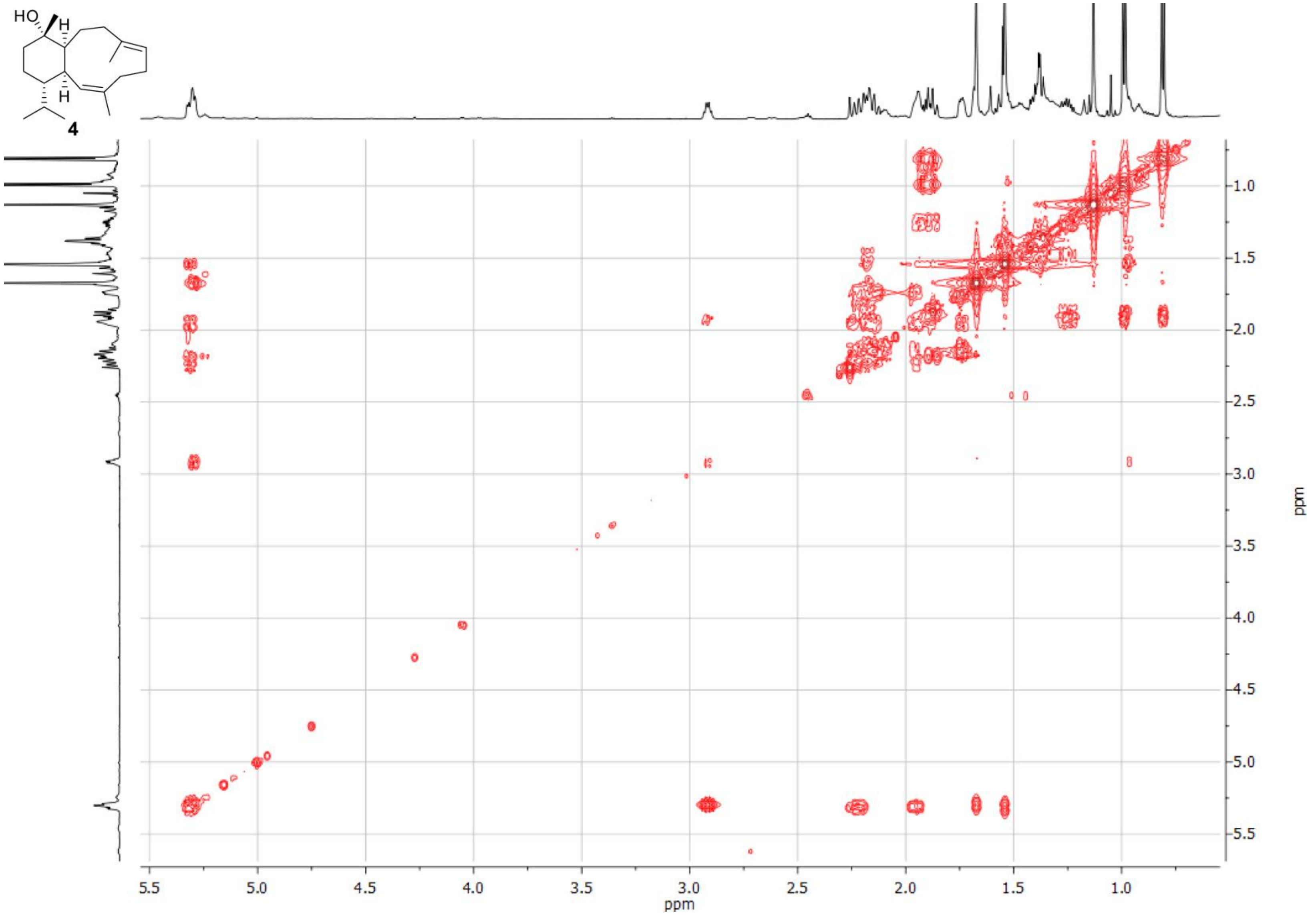
^{1}H, ^{1}H NOESY spectrum (600 MHz) of **2**.



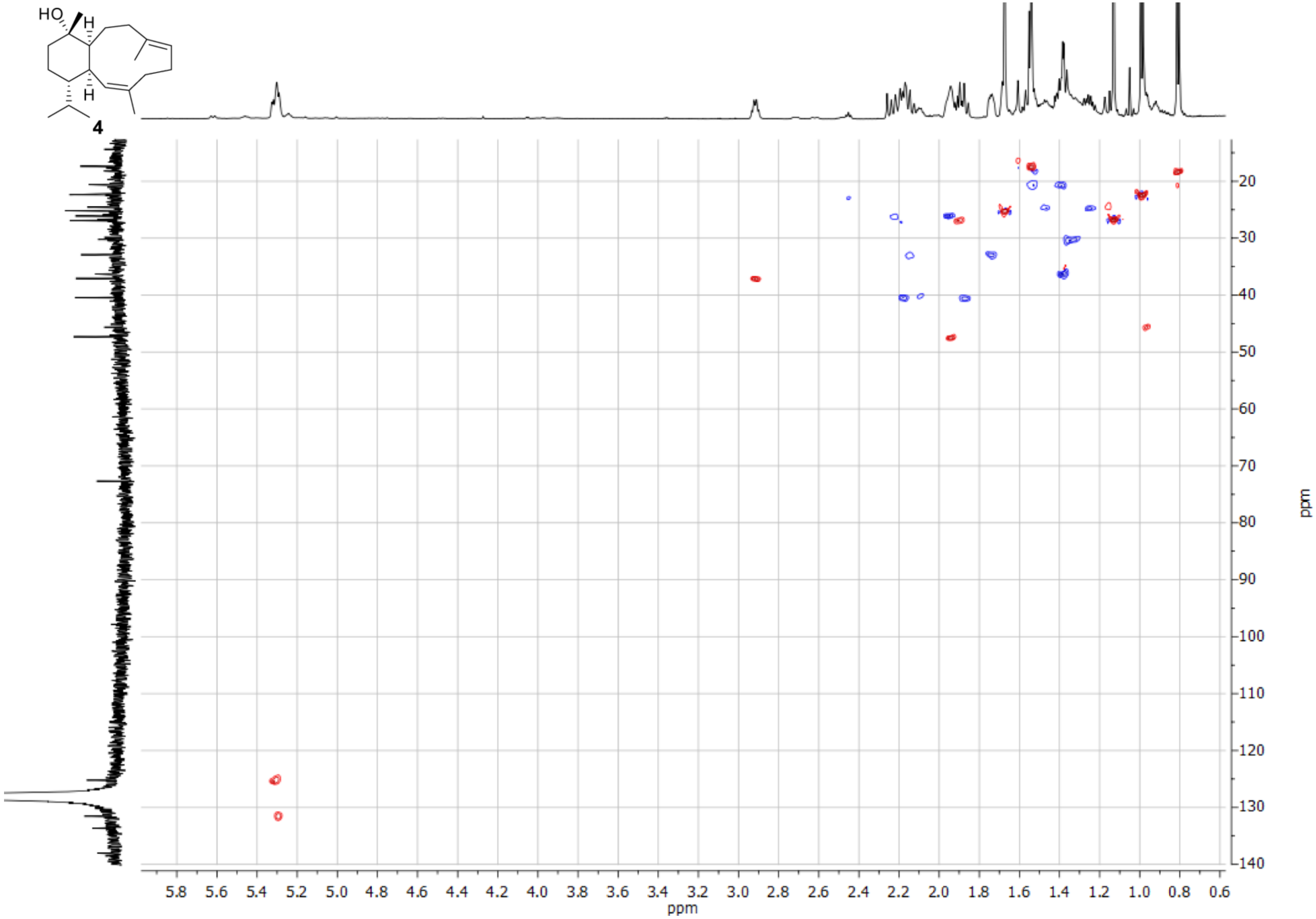
^{1}H NMR spectrum (600 MHz) of **4**.



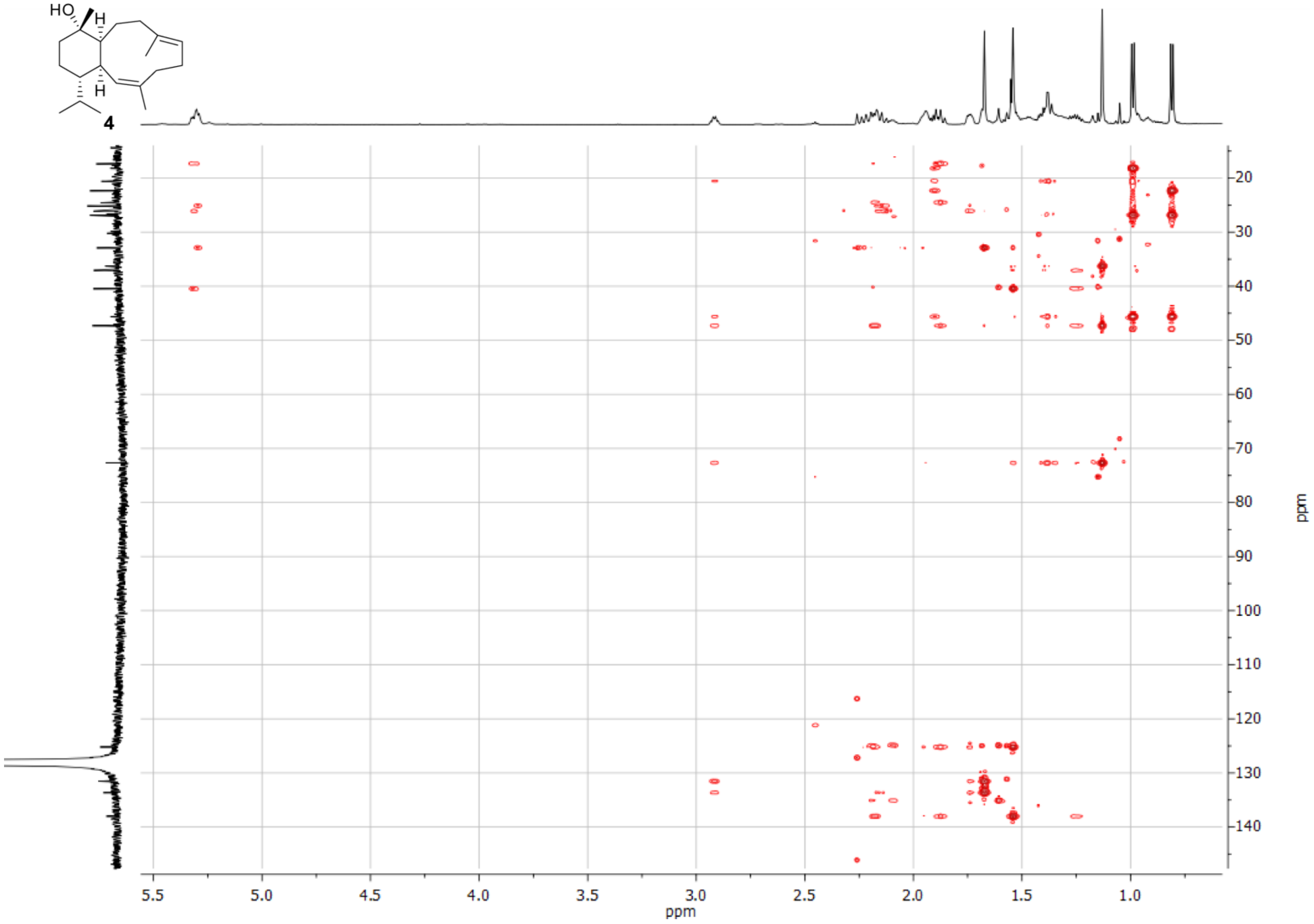
^{13}C NMR spectrum (125 MHz) of **4**.



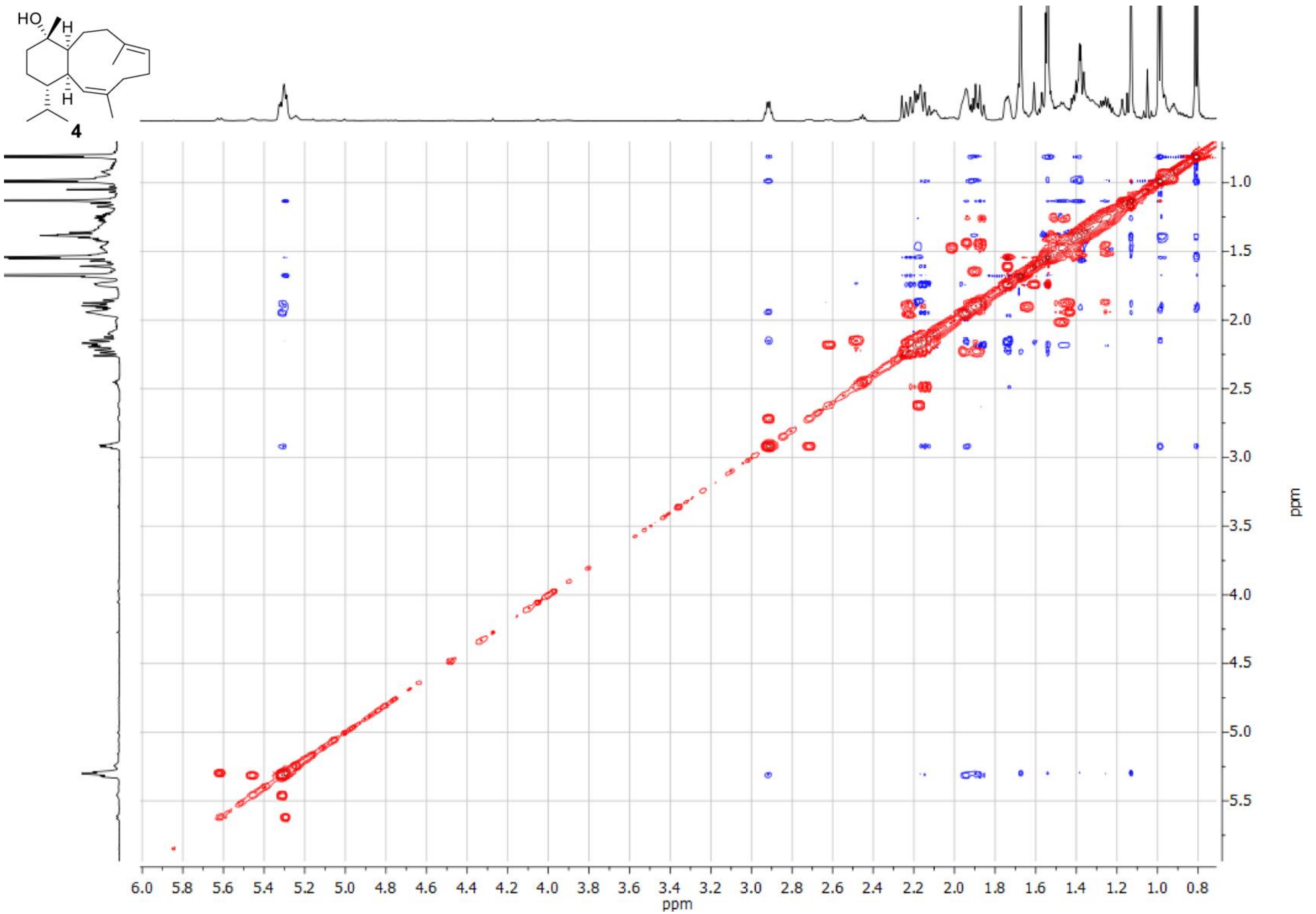
^{1}H, ^{1}H COSY spectrum (600 MHz) of **4**.



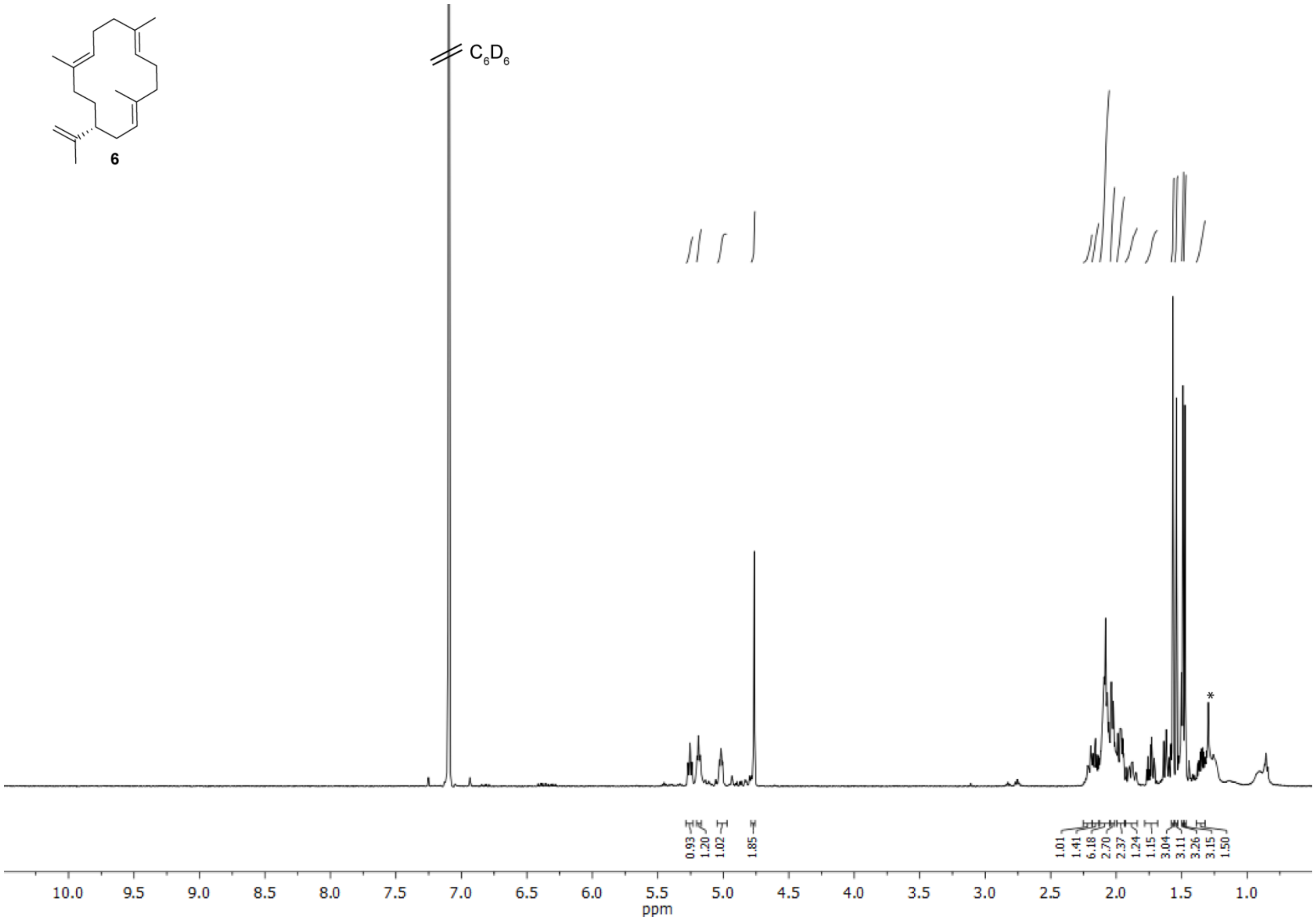
^{1}H, ^{13}C HSQC spectrum (600 MHz) of **4**.



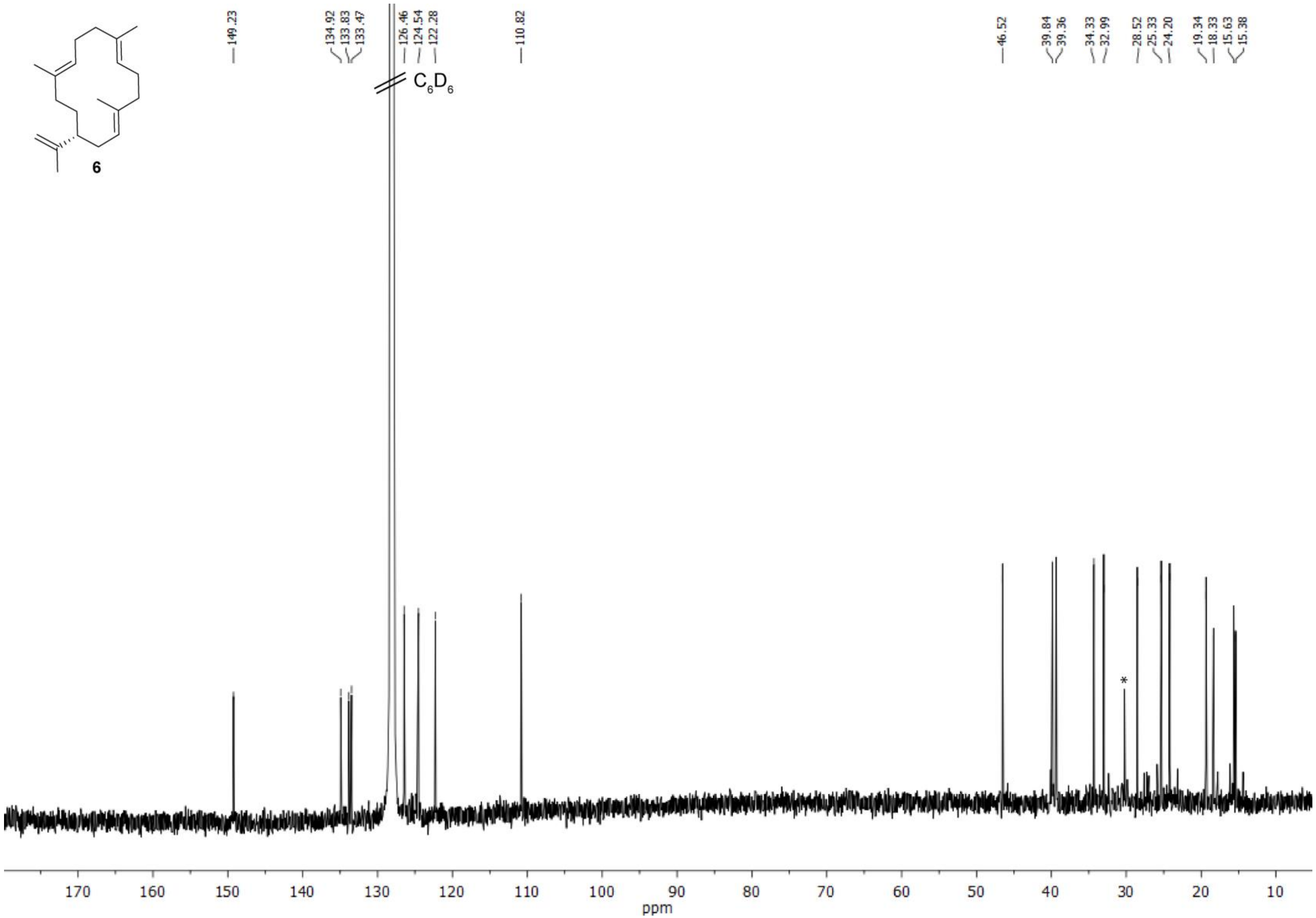
^{1}H, ^{13}C HMBC spectrum (600 MHz) of **4**.



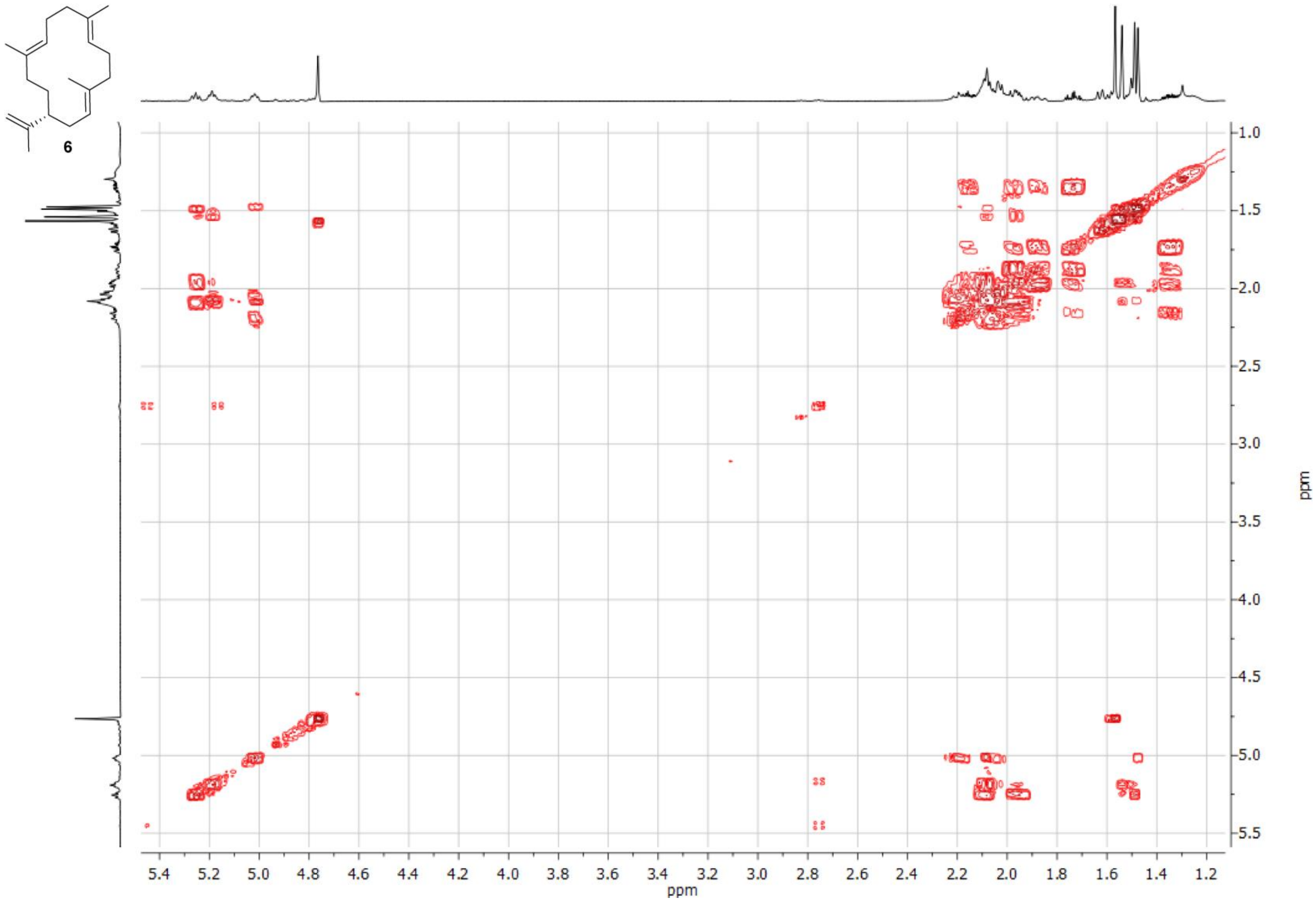
^{1}H, ^{1}H NOESY spectrum (600 MHz) of **4**.



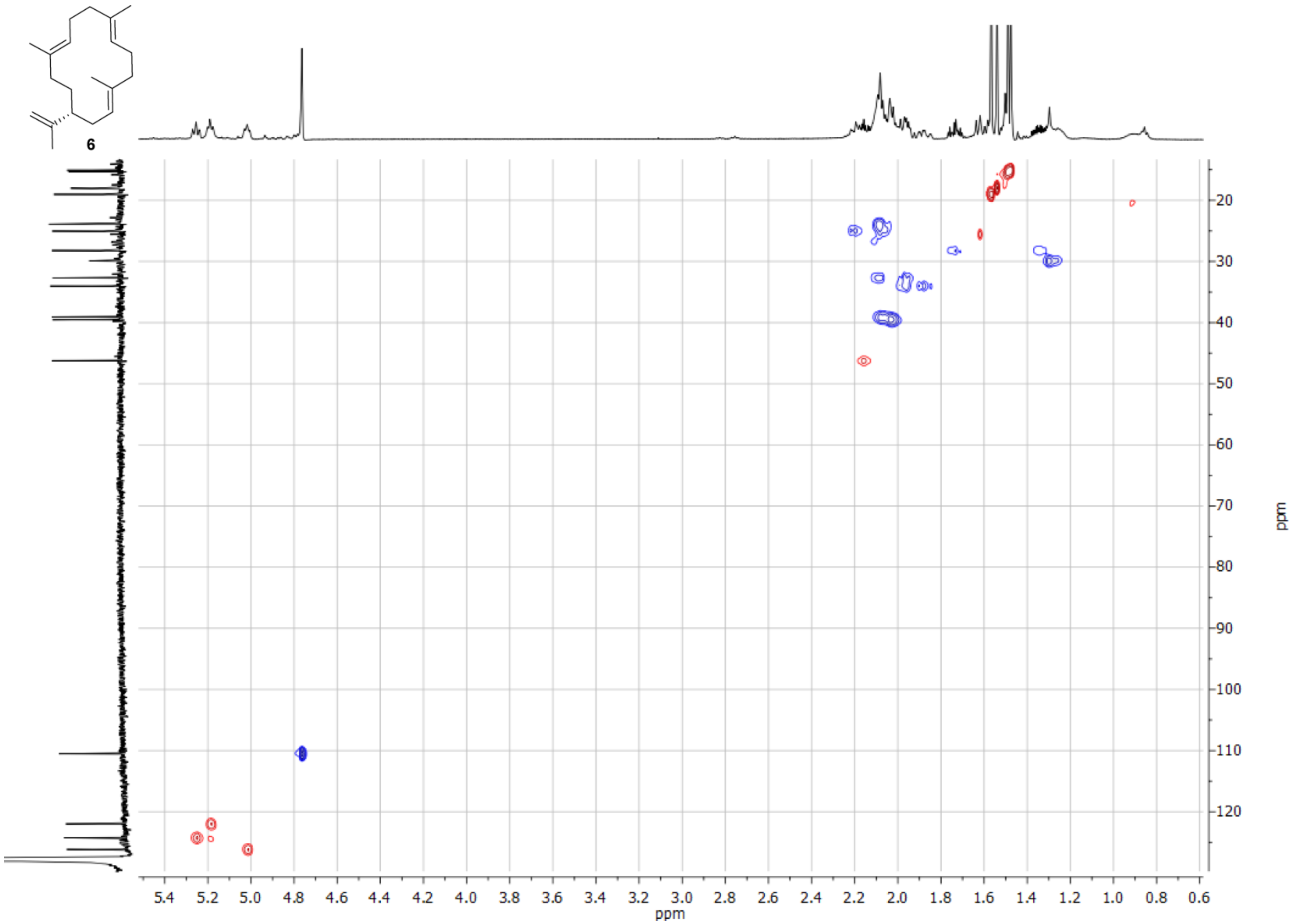
^{1}H NMR spectrum (500 MHz) of **6**.



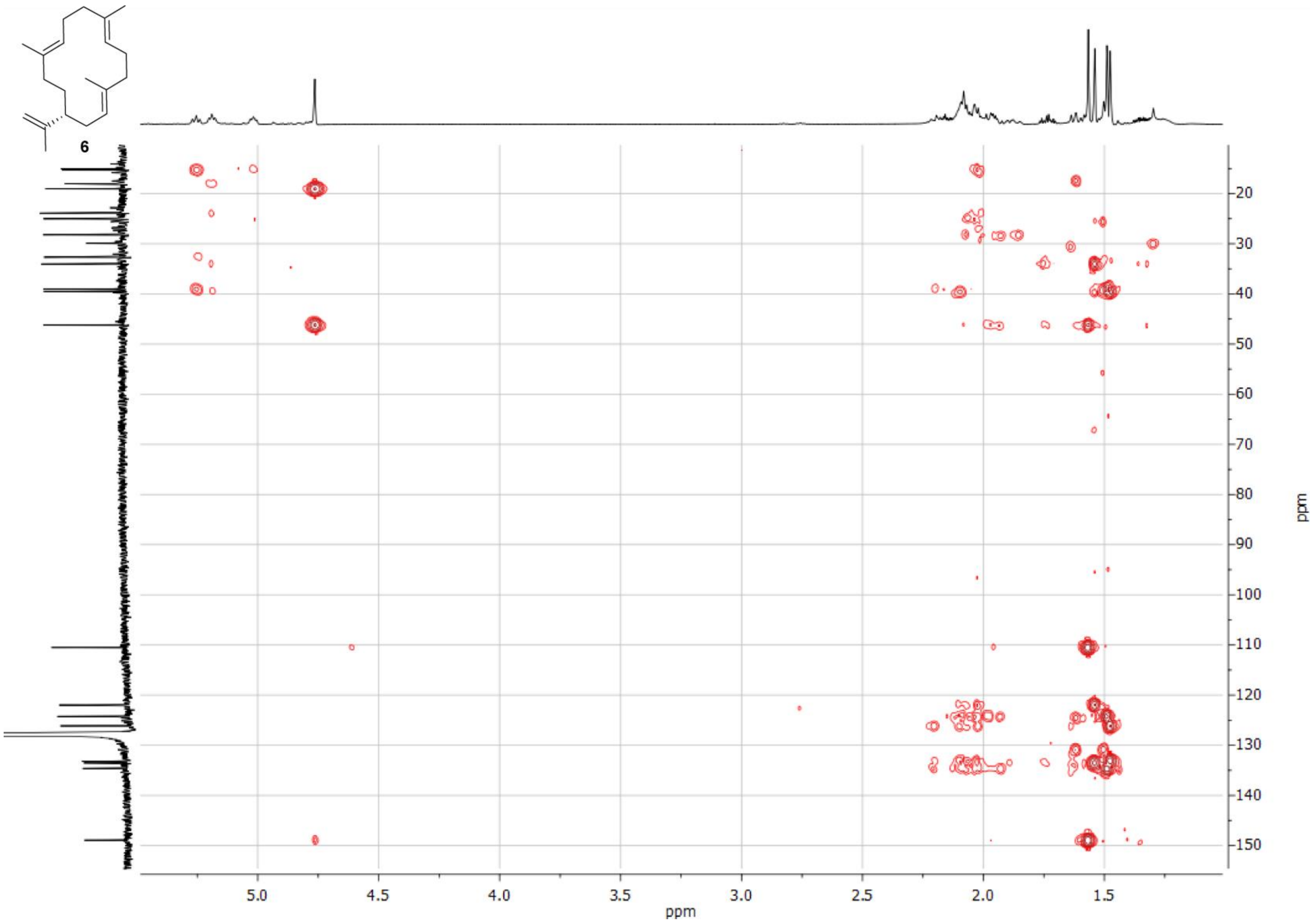
^{13}C NMR spectrum (125 MHz) of **6**.



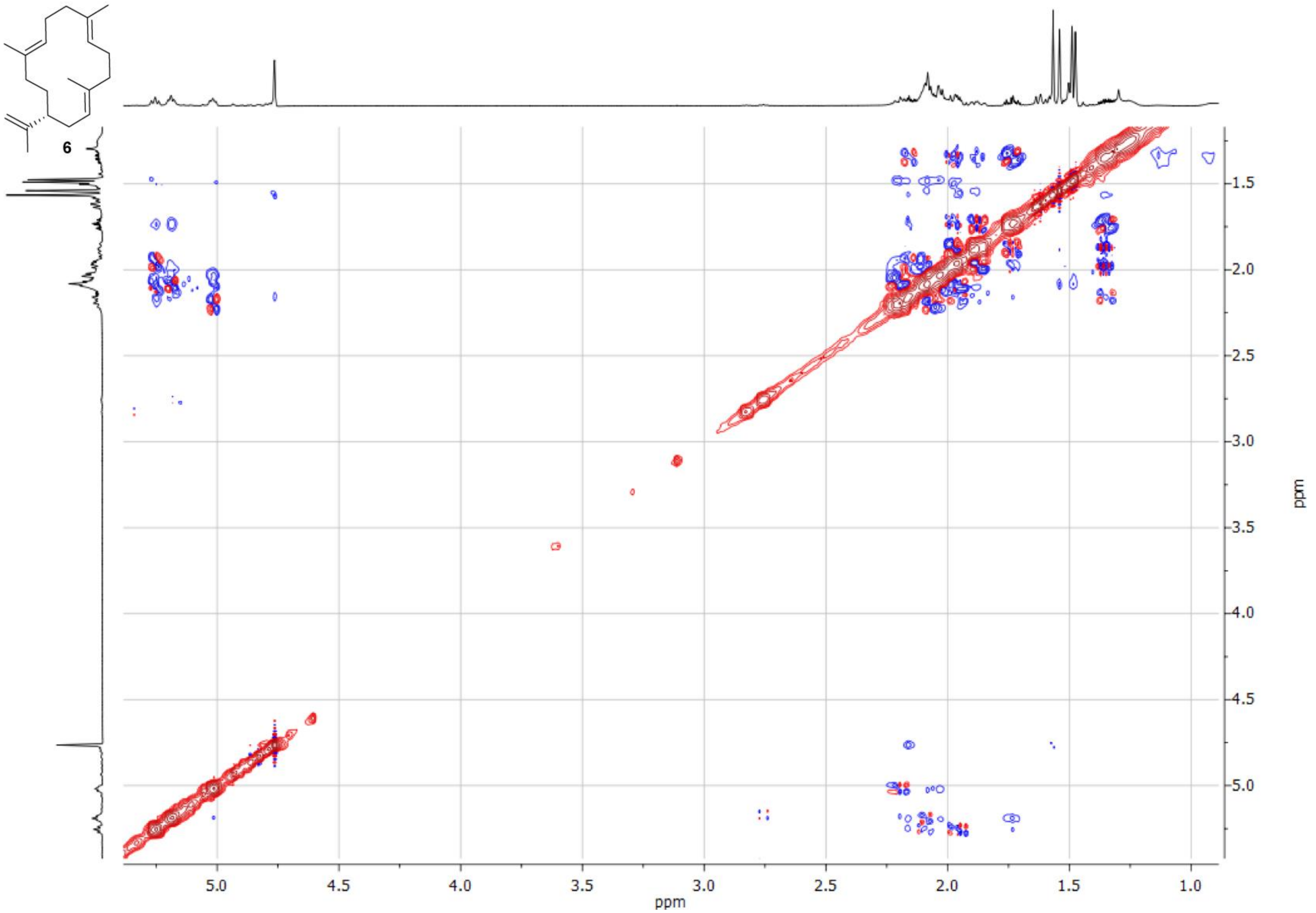
^{1}H, ^{1}H COSY spectrum (500 MHz) of **6**.



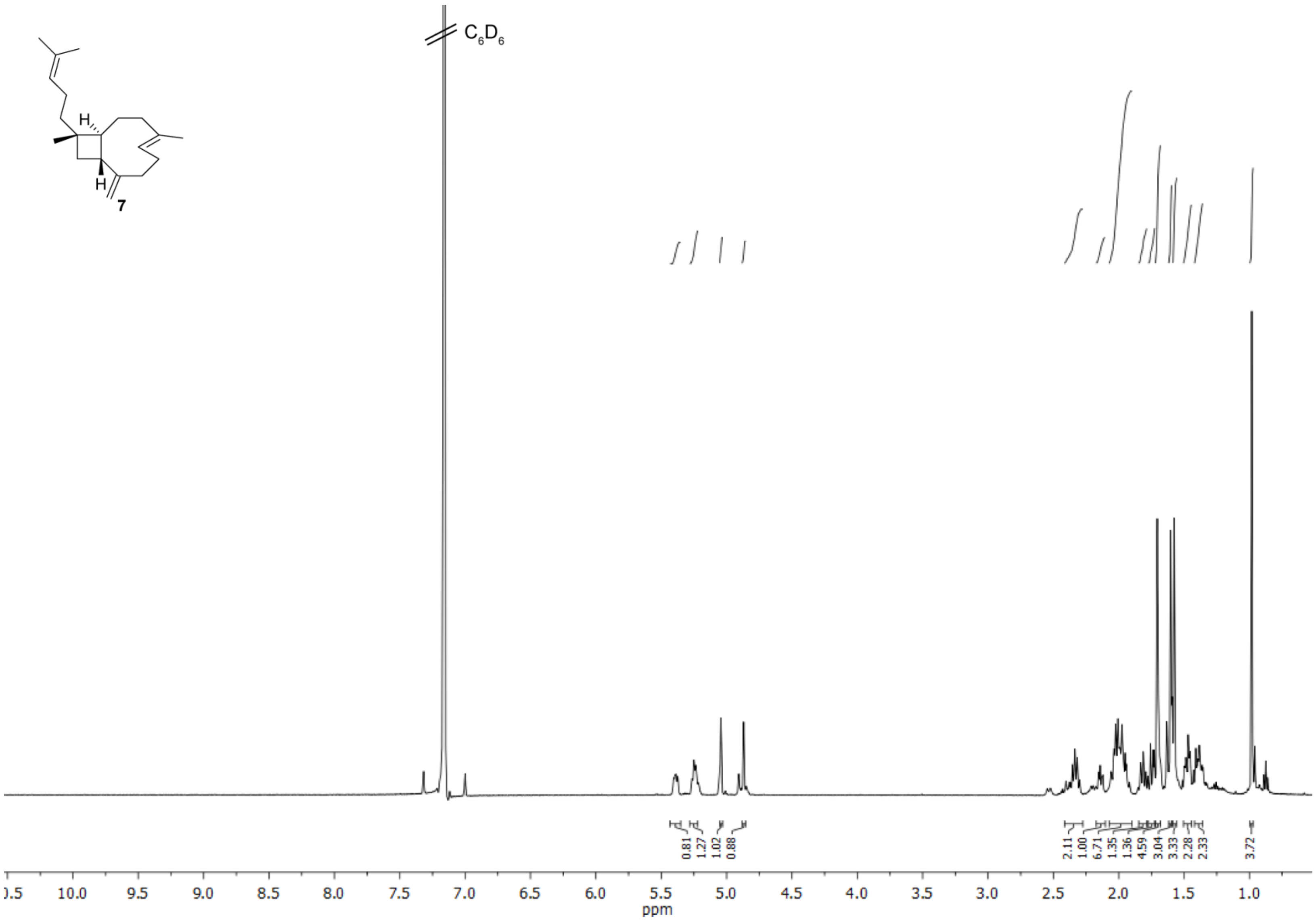
^{1}H, ^{13}C HSQC spectrum (500 MHz) of **6**.



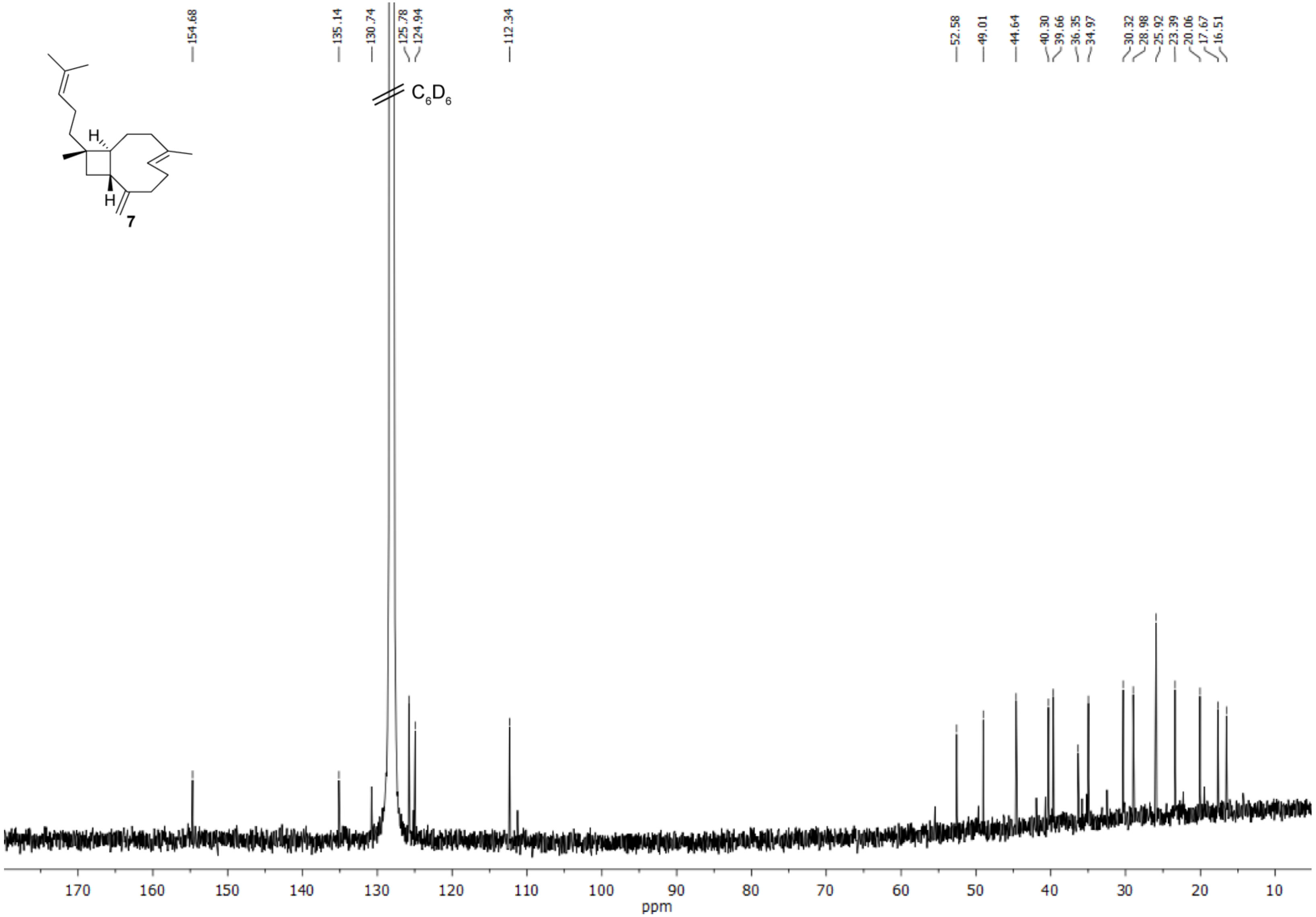
^{1}H, ^{13}C HMBC spectrum (500 MHz) of **6**.



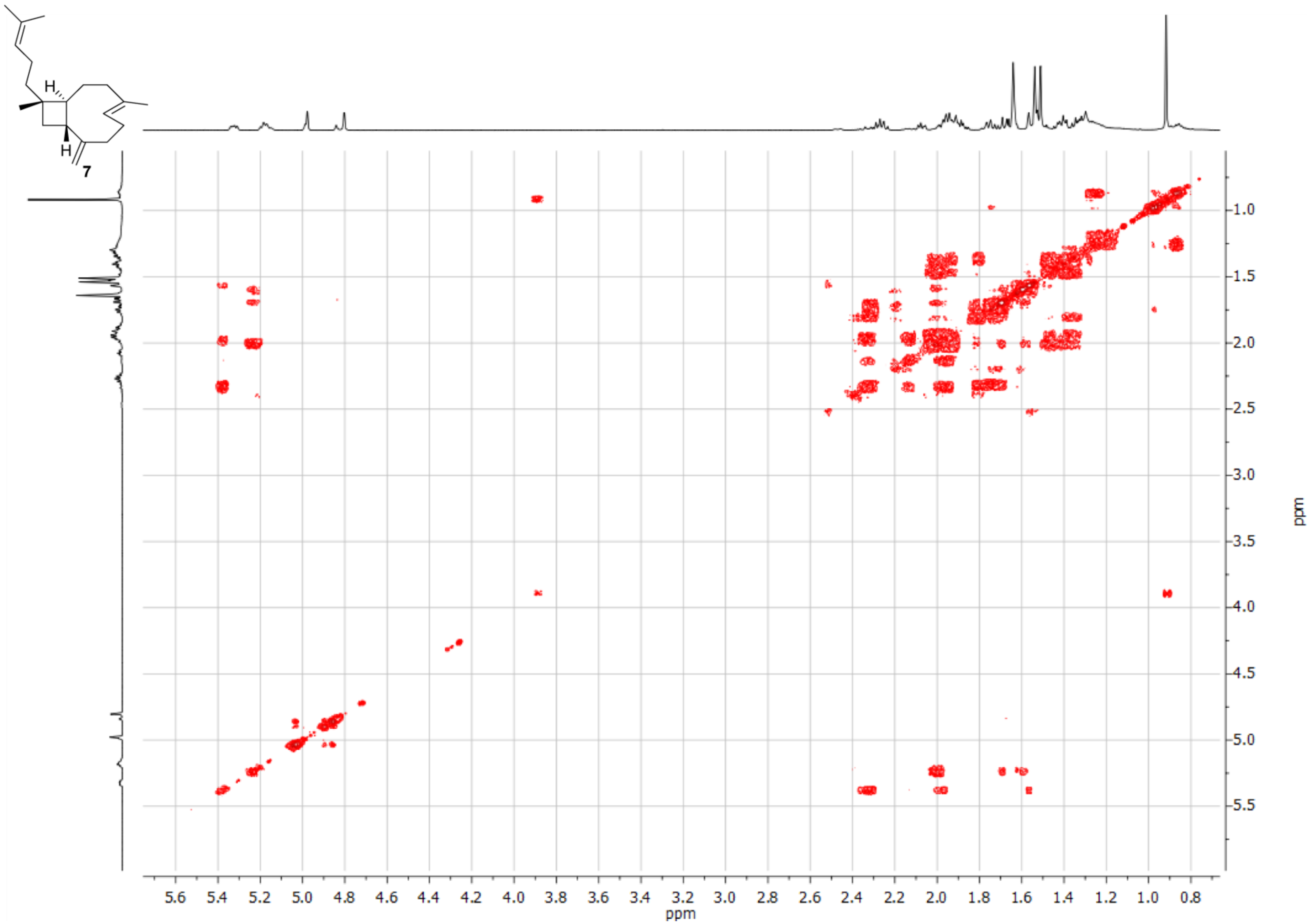
^{1}H, ^{1}H NOESY spectrum (600 MHz) of **6**.



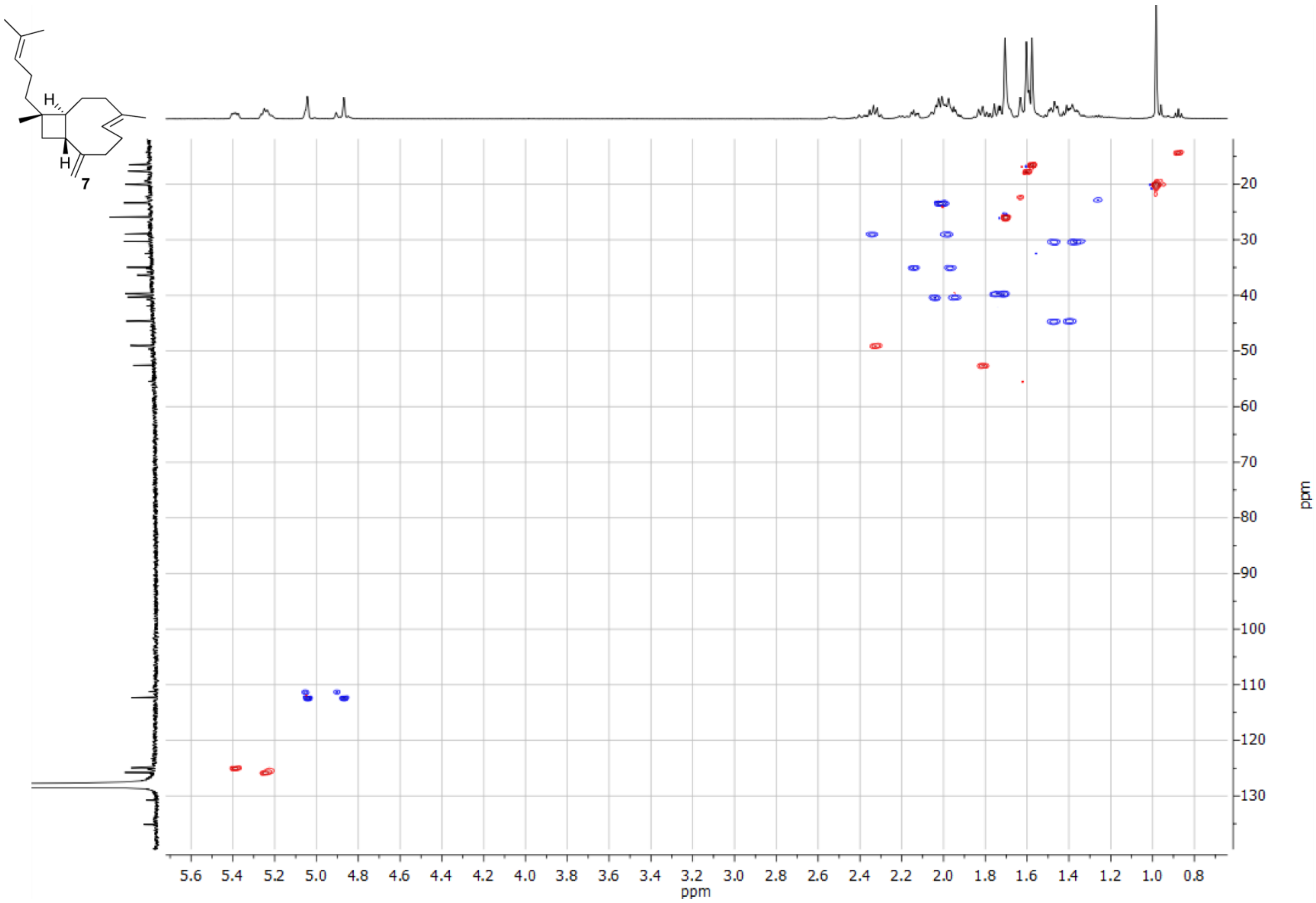
^{1}H NMR spectrum (600 MHz) of **7**.



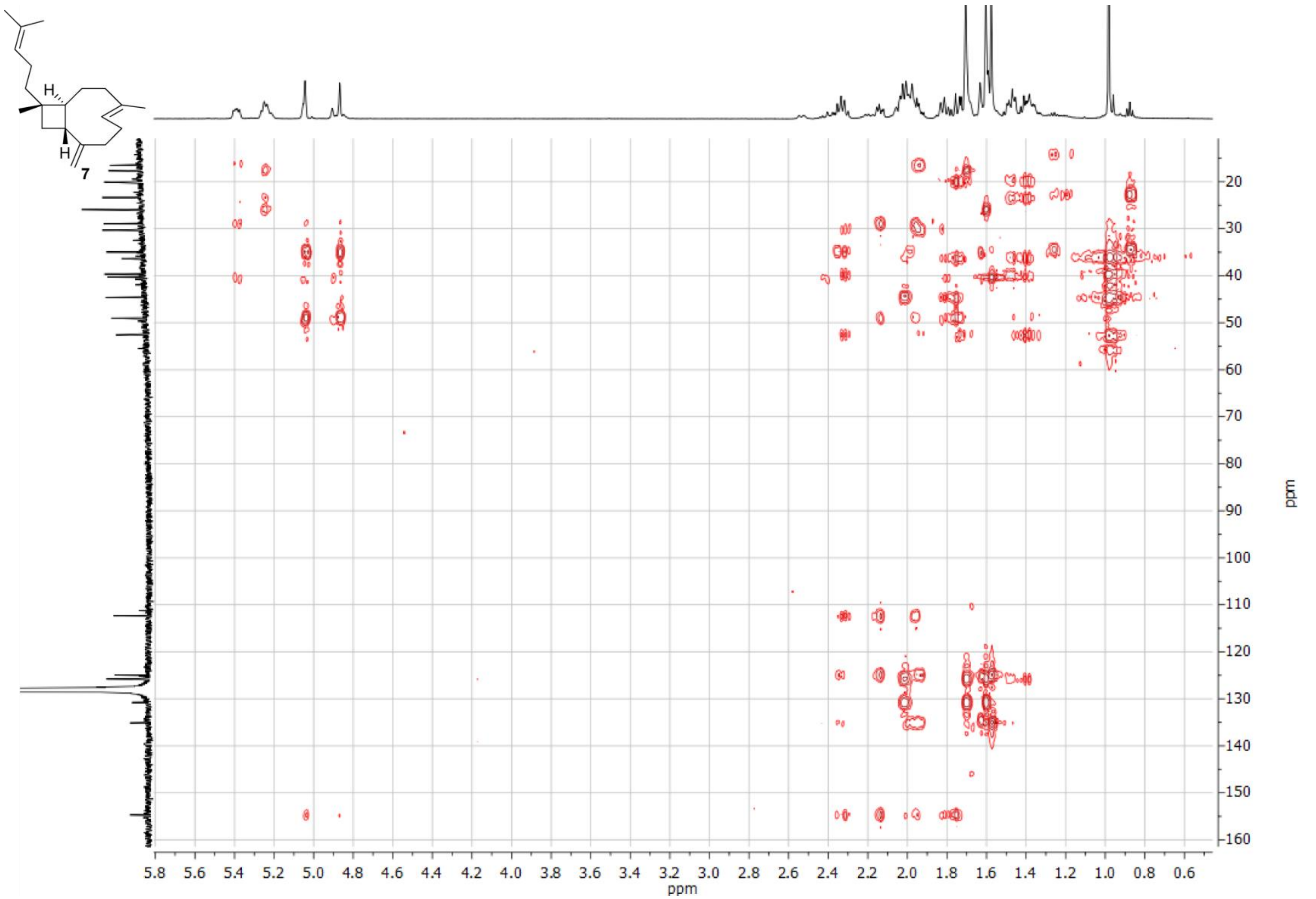
^{13}C NMR spectrum (125 MHz) of **7**.



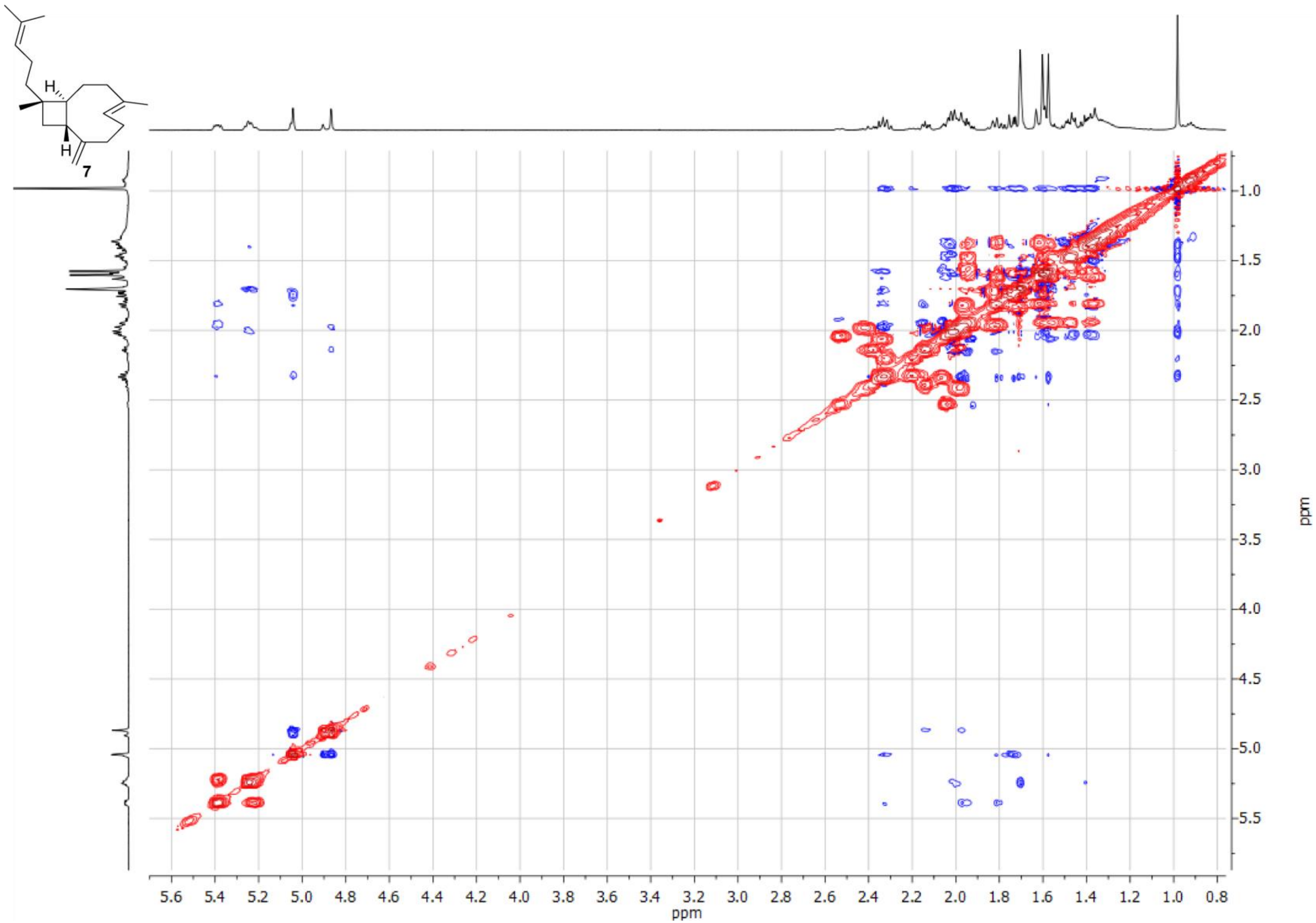
^{1}H, ^{1}H COSY spectrum (600 MHz) of **7**.



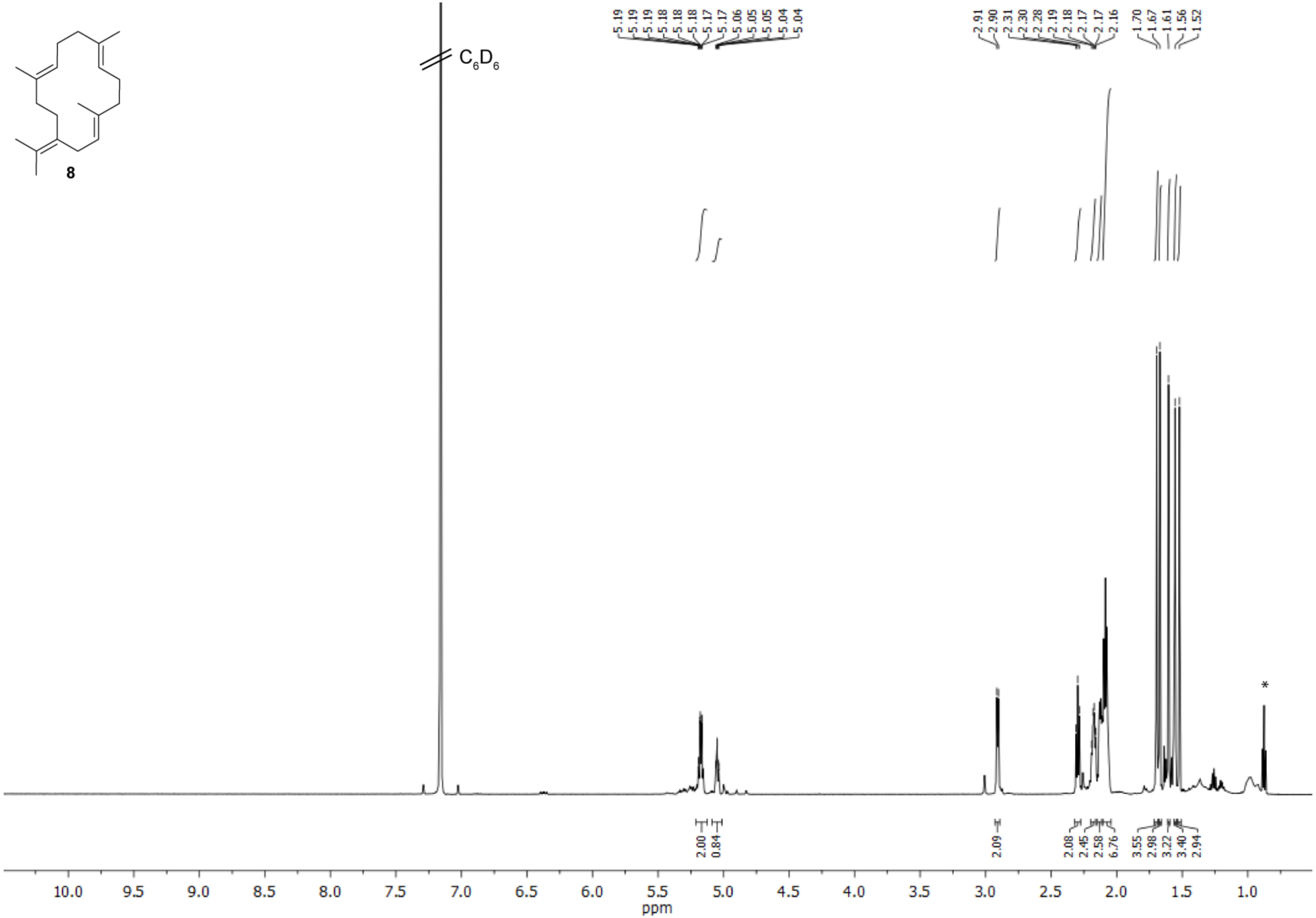
^{1}H, ^{13}C HSQC spectrum (600 MHz) of **7**.



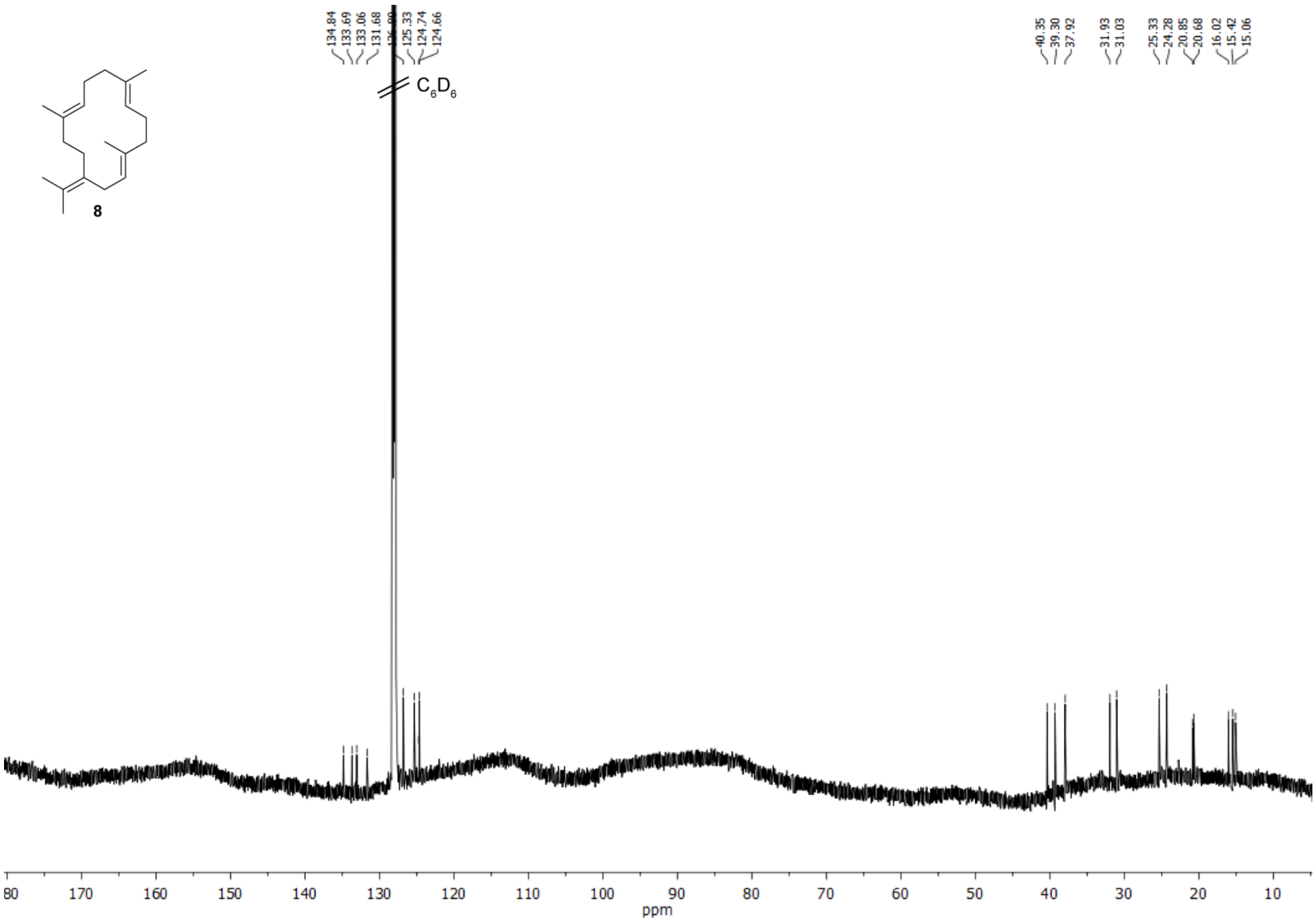
^{1}H, ^{13}C HMBC spectrum (600 MHz) of **7**.



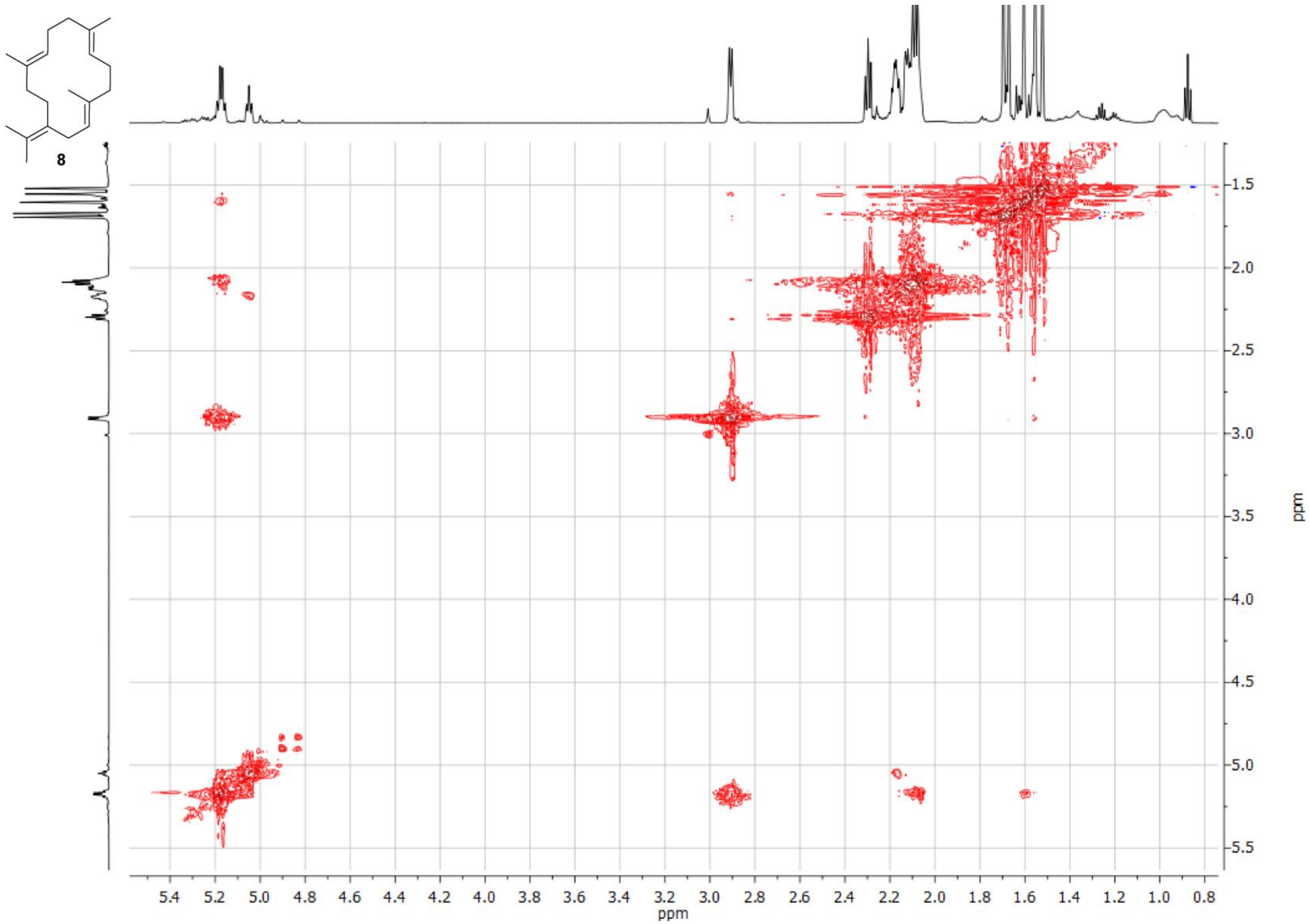
^{1}H, ^{1}H NOESY spectrum (500 MHz) of **7**.



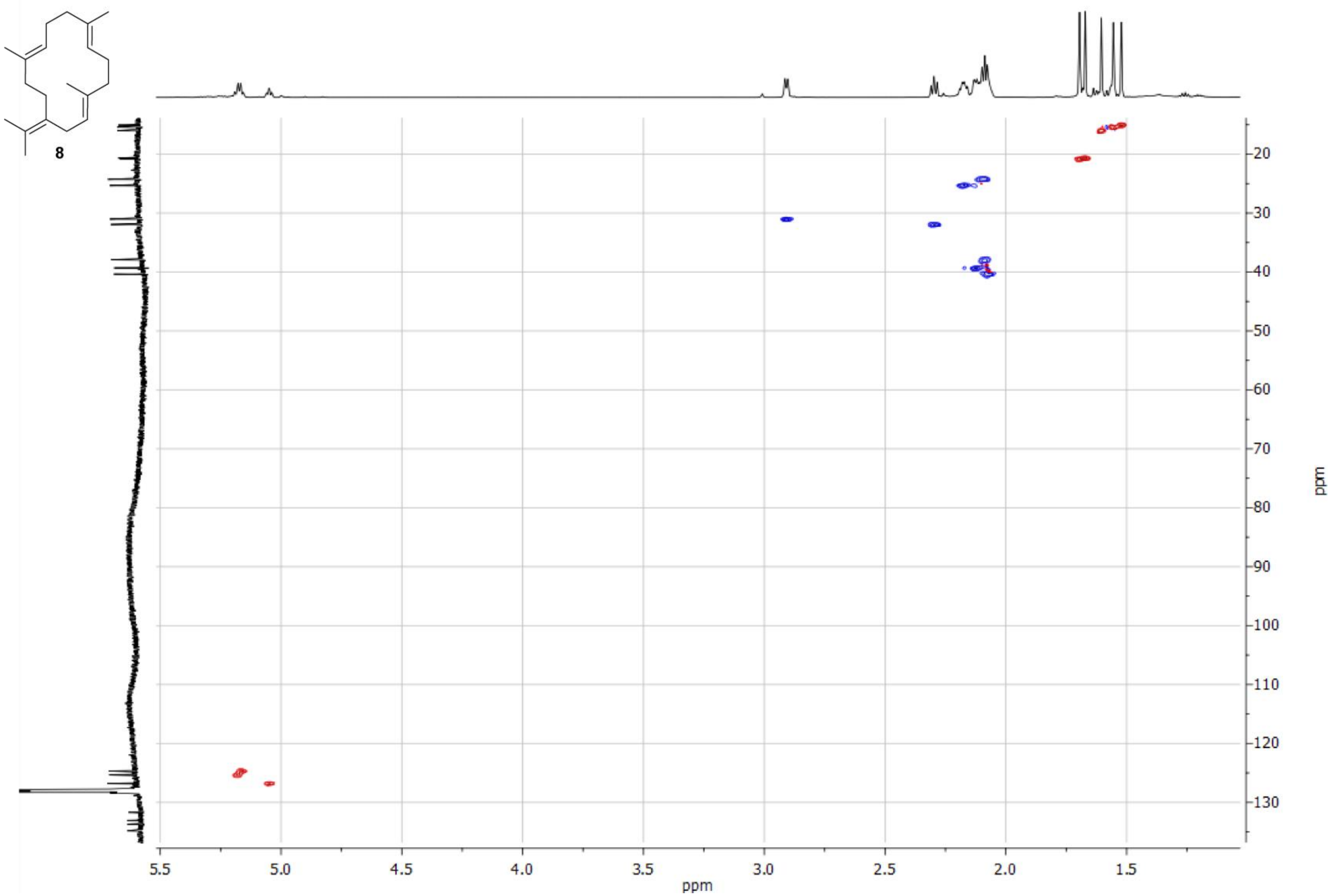
^{1}H NMR spectrum (600 MHz) of **8**. Asterisk denotes residual pentanes.



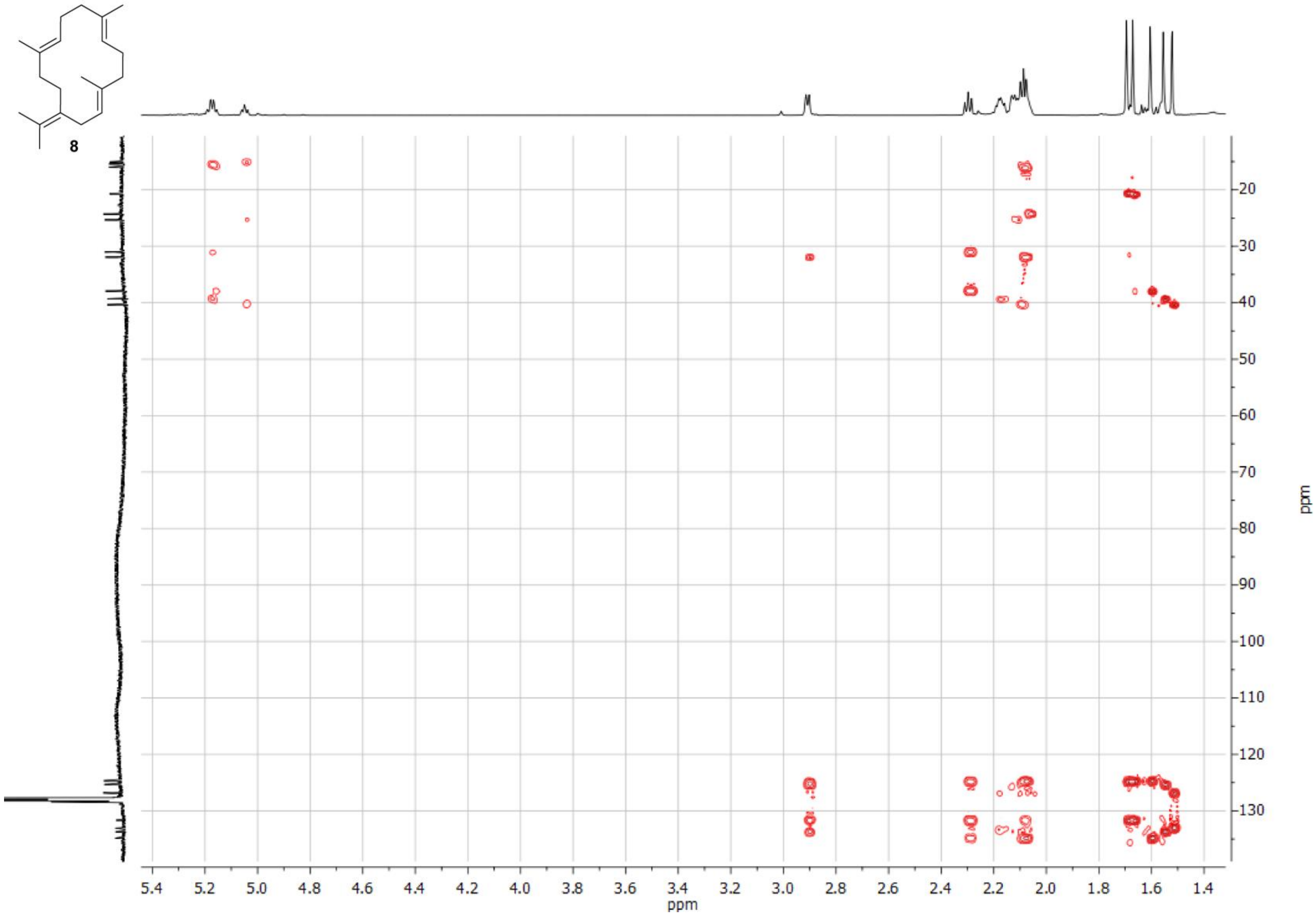
^{13}C NMR spectrum (125 MHz) of **8**.



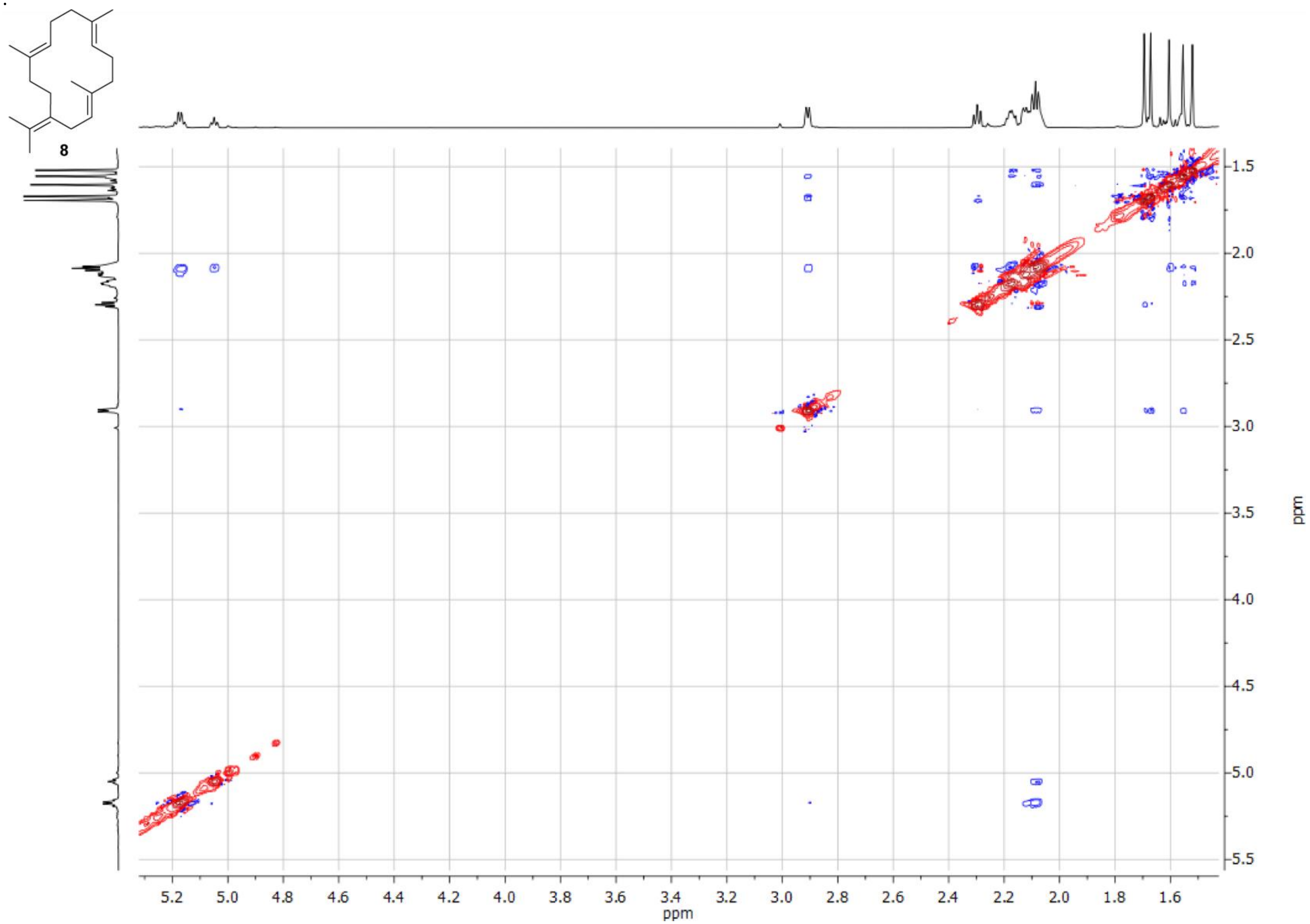
^{1}H, ^{1}H COSY spectrum (600 MHz) of **8**.



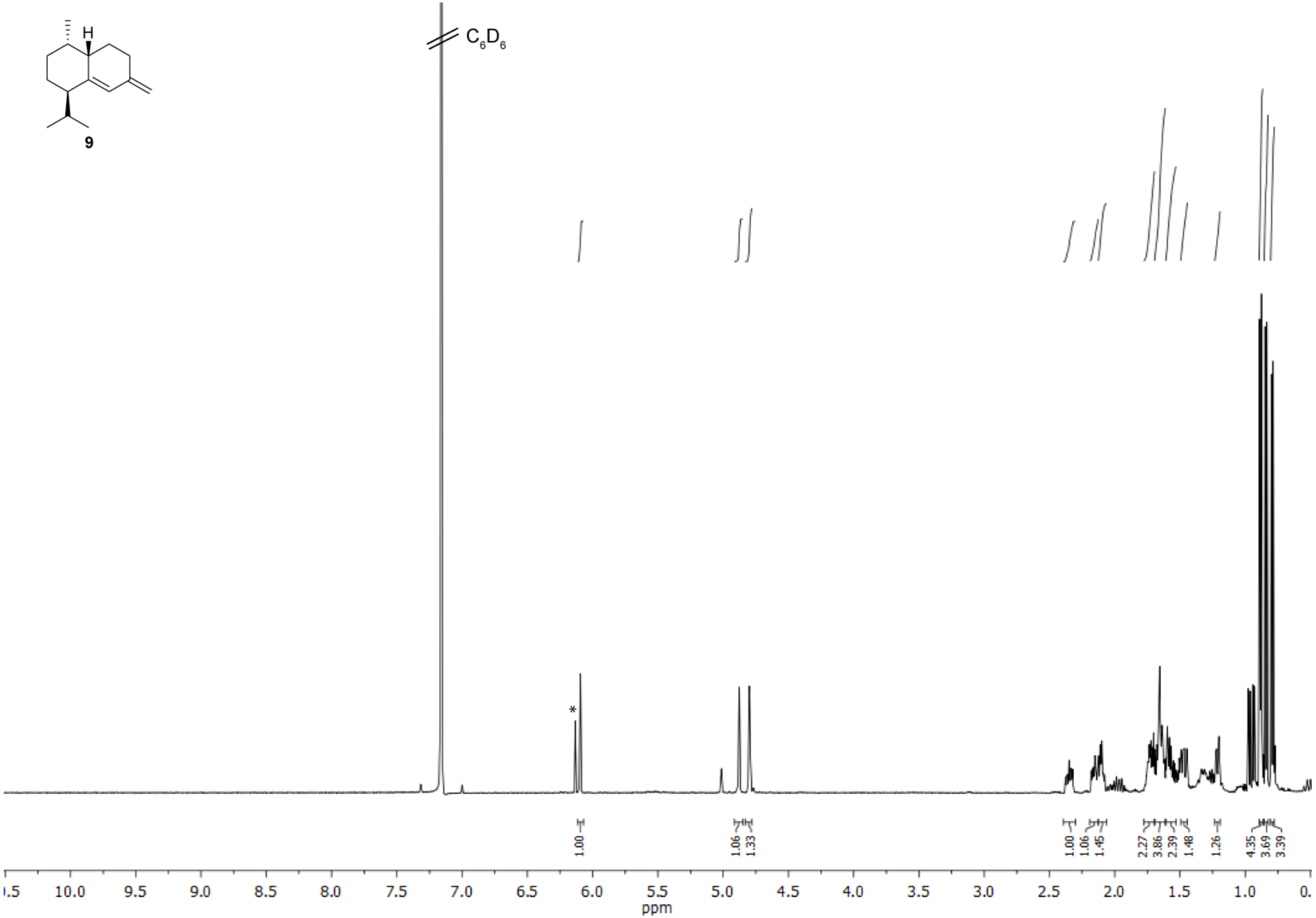
^{1}H, ^{13}C HSQC spectrum (600 MHz) of **8**.



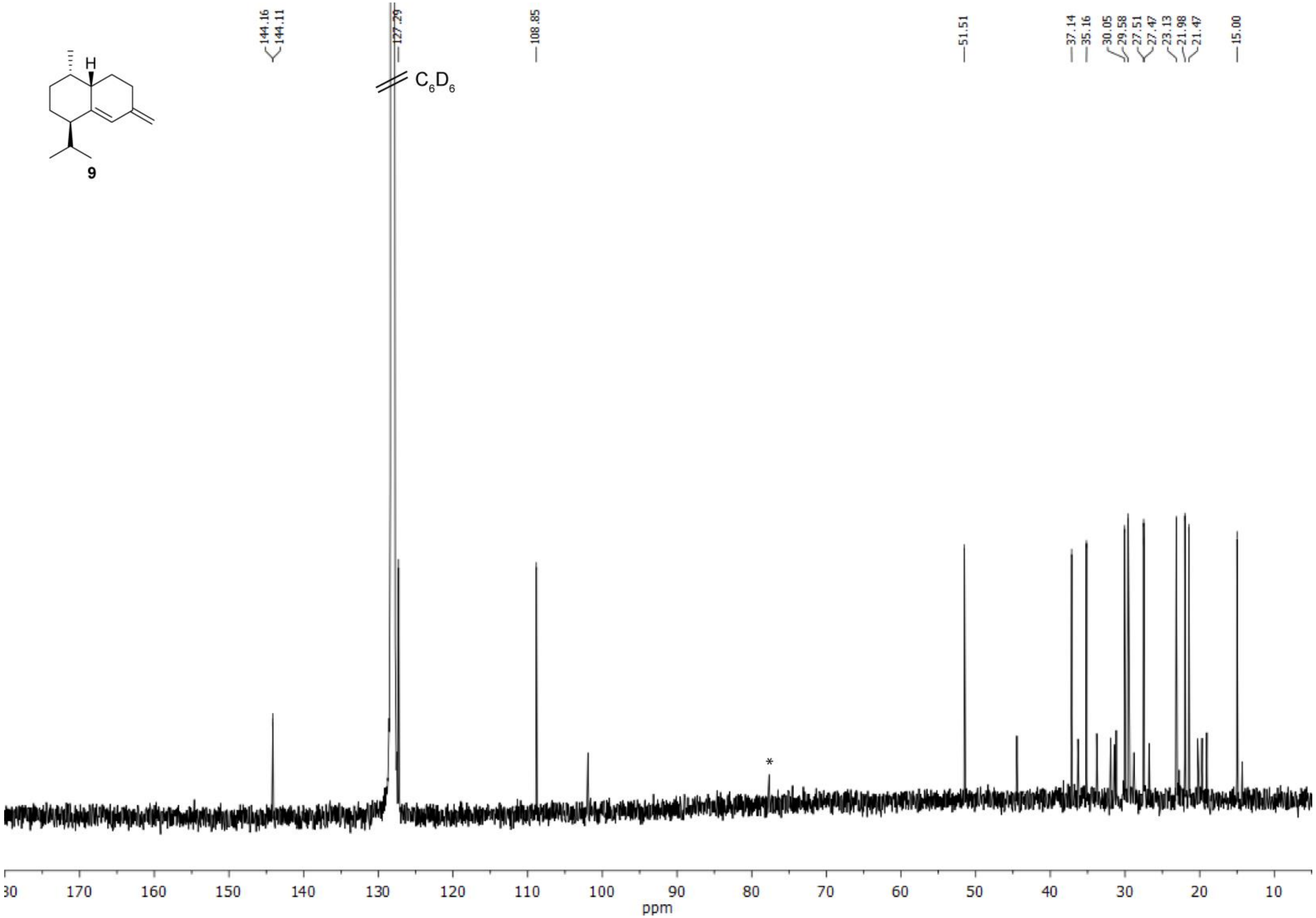
^{1}H, ^{13}C HMBC spectrum (600 MHz) of **8**



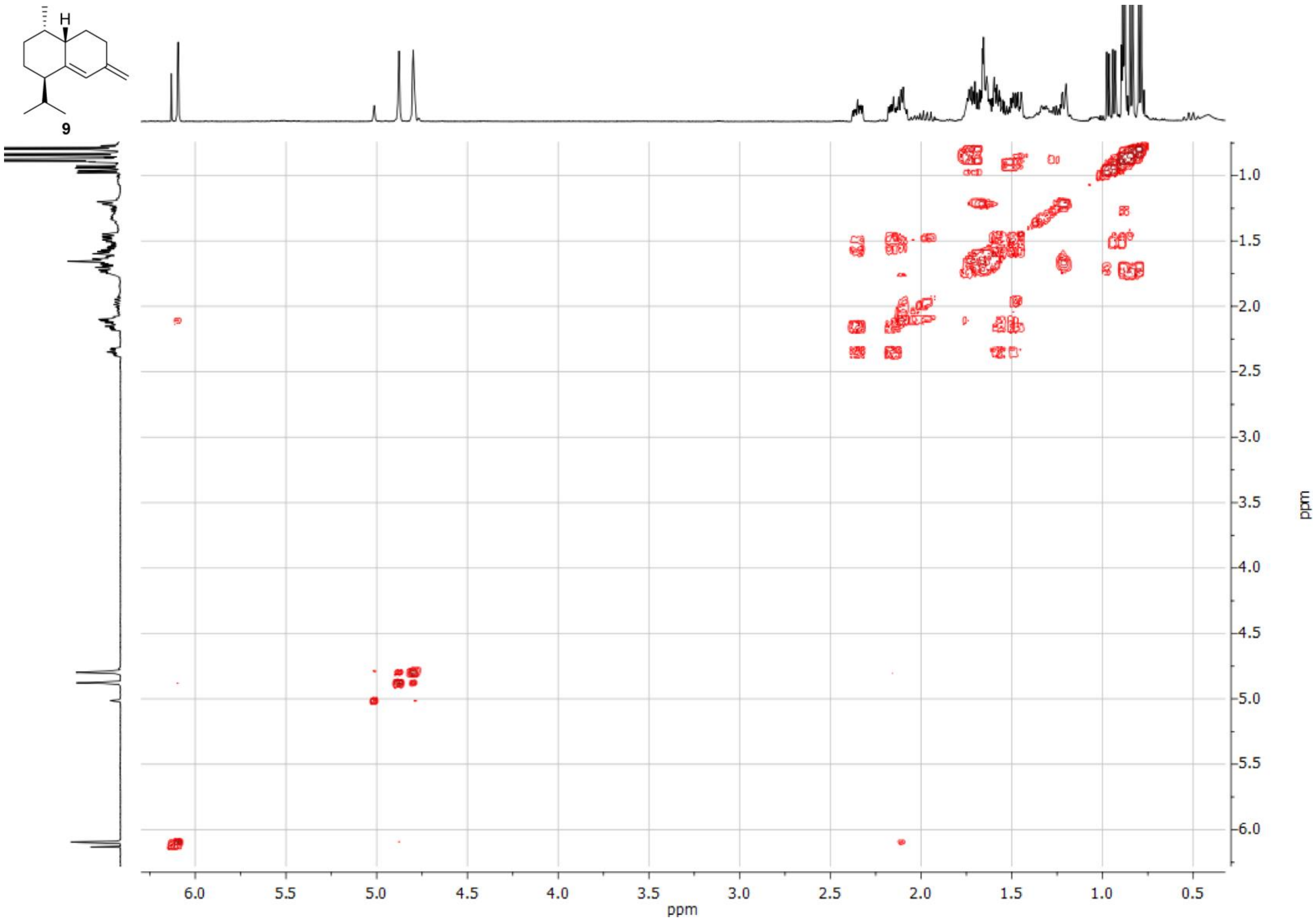
^{1}H, ^{1}H NOESY spectrum (500 MHz) of **8**.



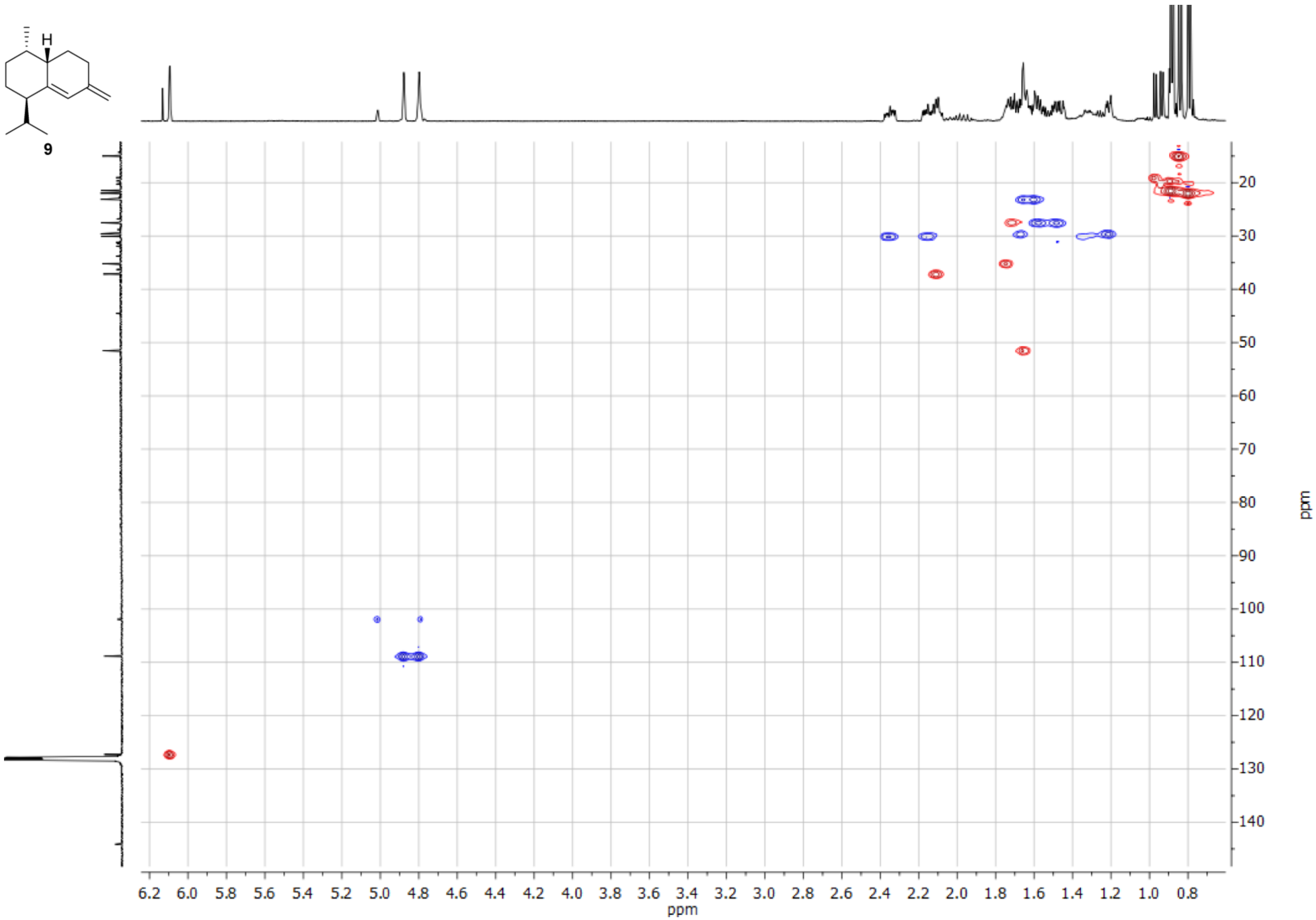
^{1}H NMR spectrum (500 MHz) of **9**. Asterisk denotes residual chloroform. Other minor signals originate from an inseparable second unidentified sesquiterpene.



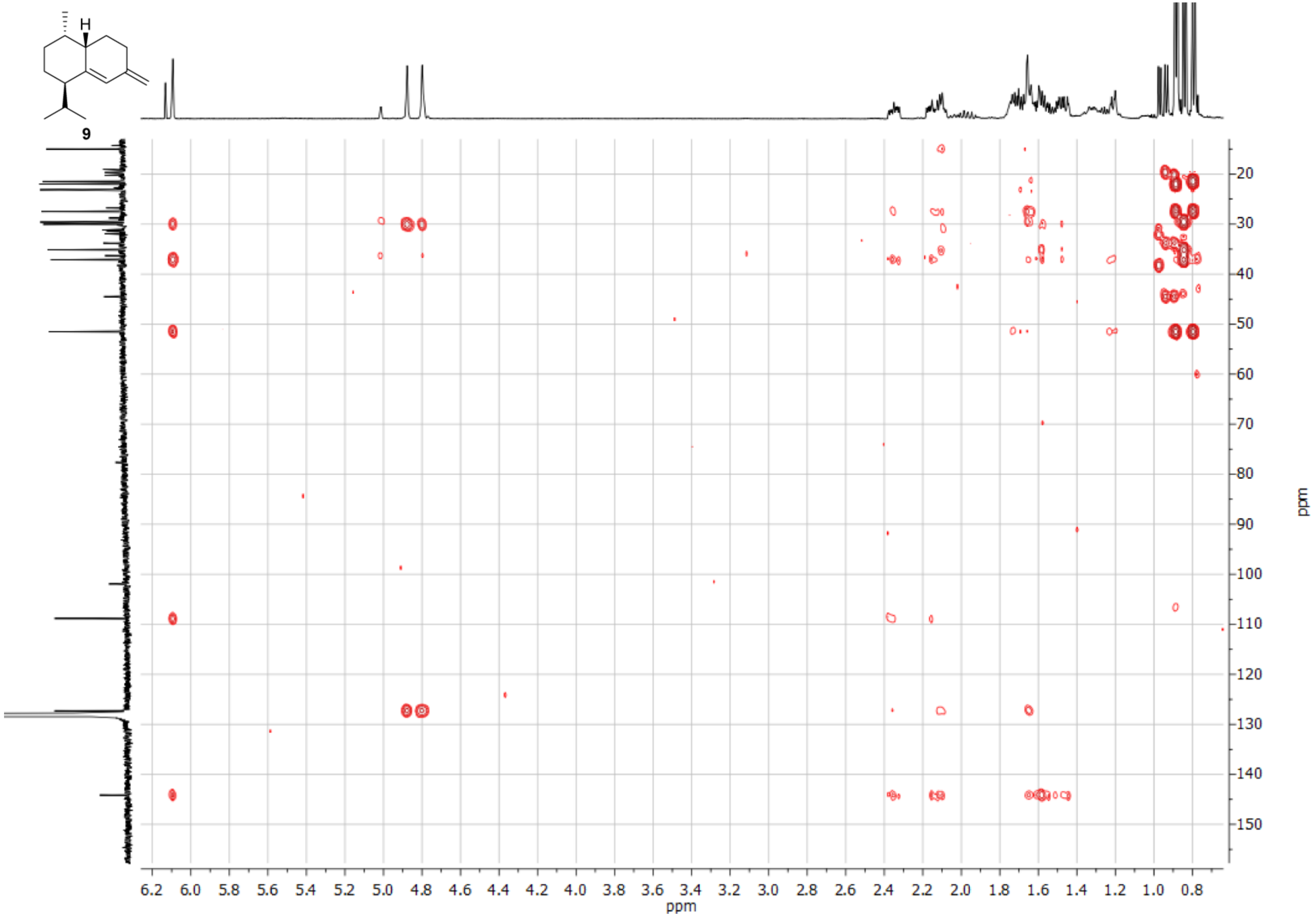
^{13}C NMR spectrum (125 MHz) of **9**. Asterisk denotes residual chloroform. Other minor signals originate from an inseparable second unidentified sesquiterpene.



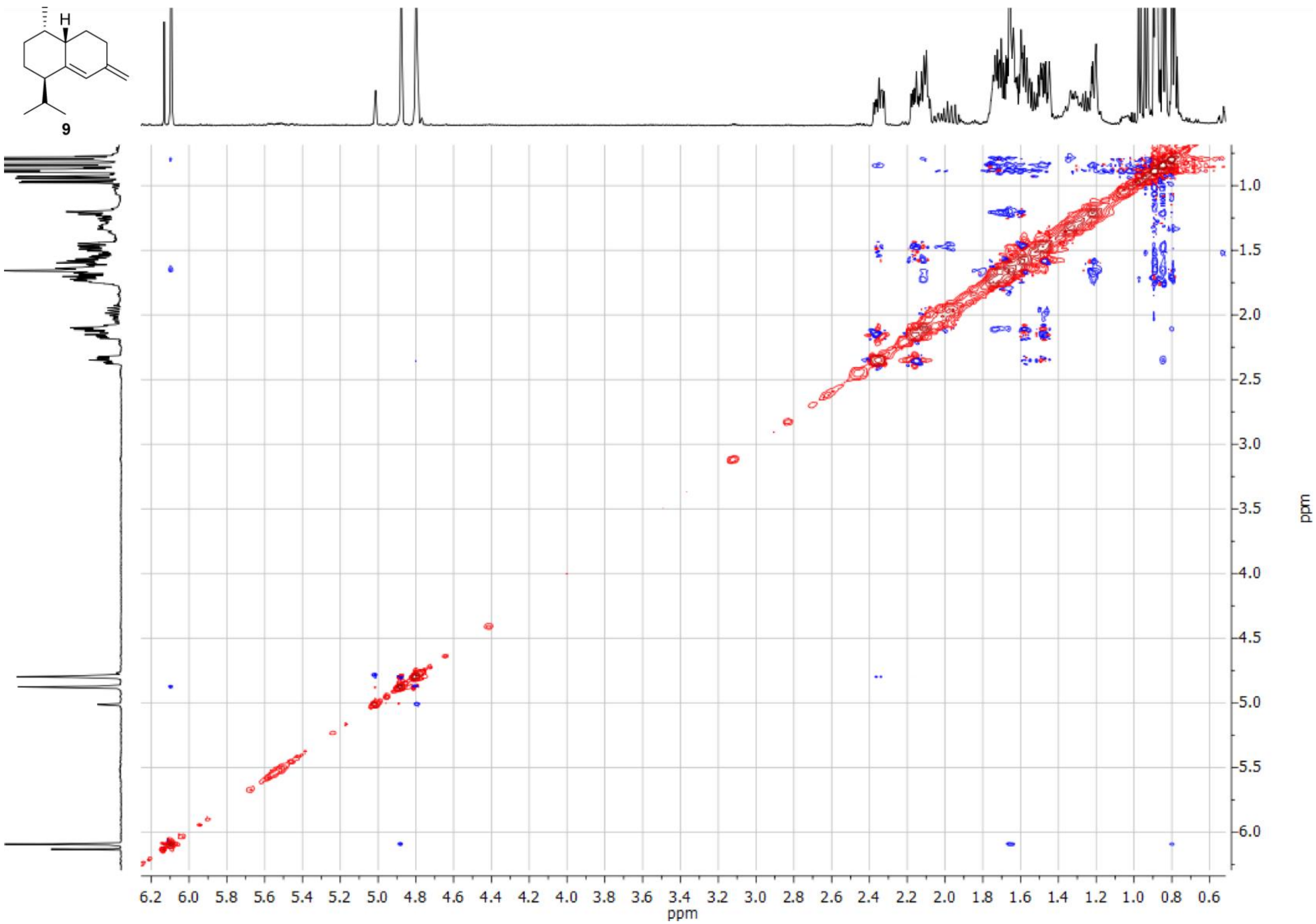
^{1}H, ^{1}H COSY spectrum (500 MHz) of **9**.



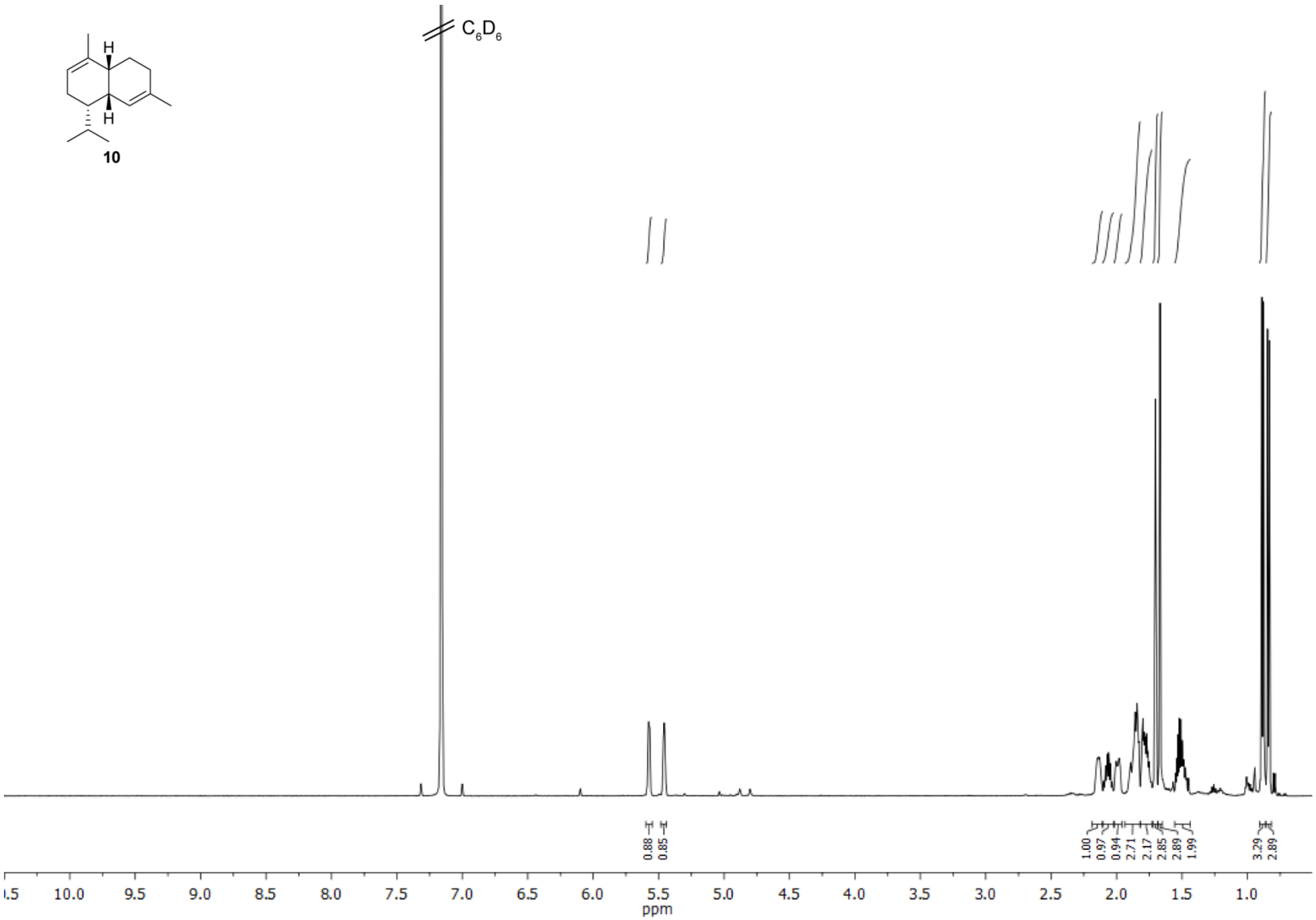
^{1}H, ^{13}C HSQC spectrum (500 MHz) of **9**.



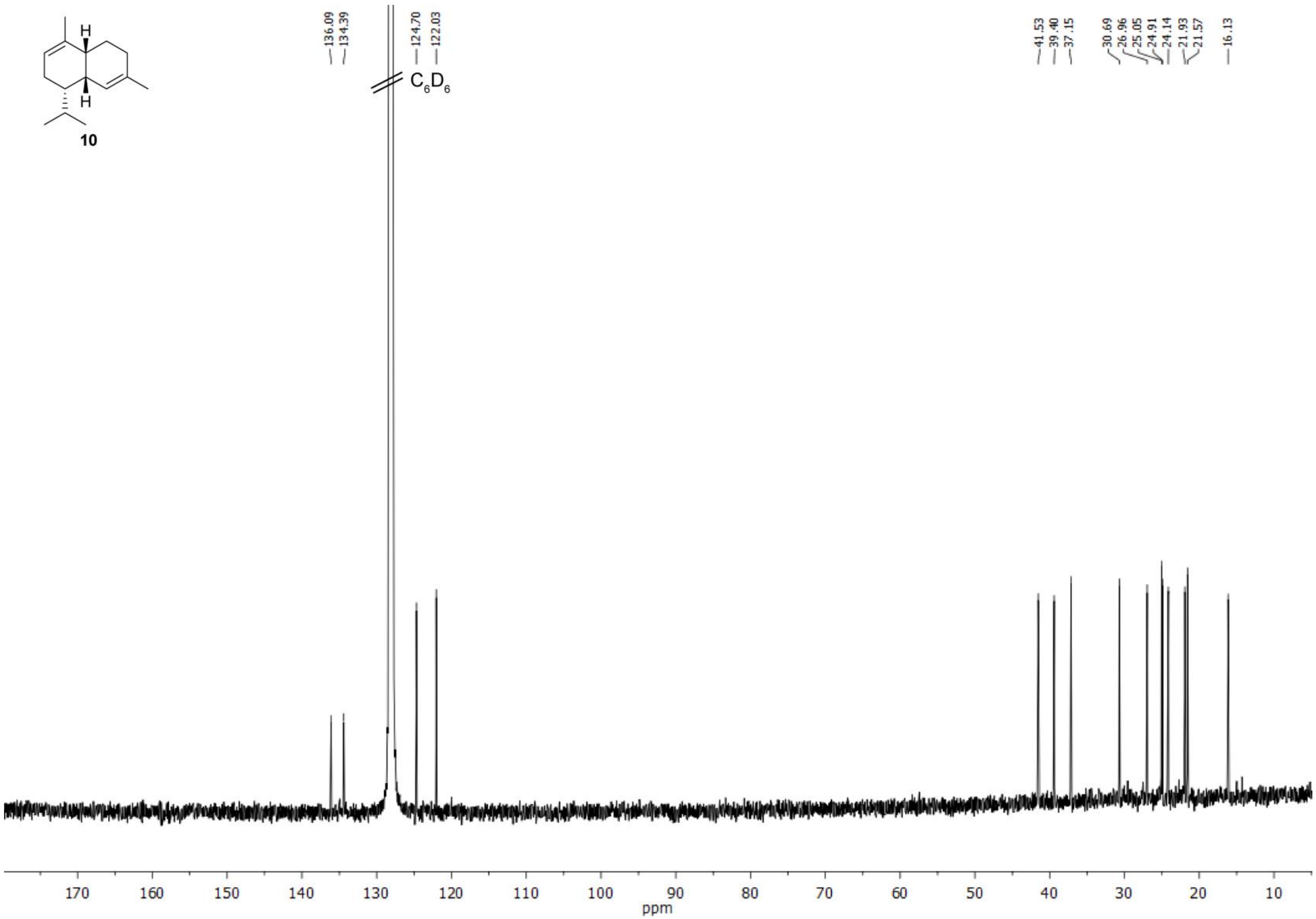
^{1}H, ^{13}C HMBC spectrum (500 MHz) of **9**.



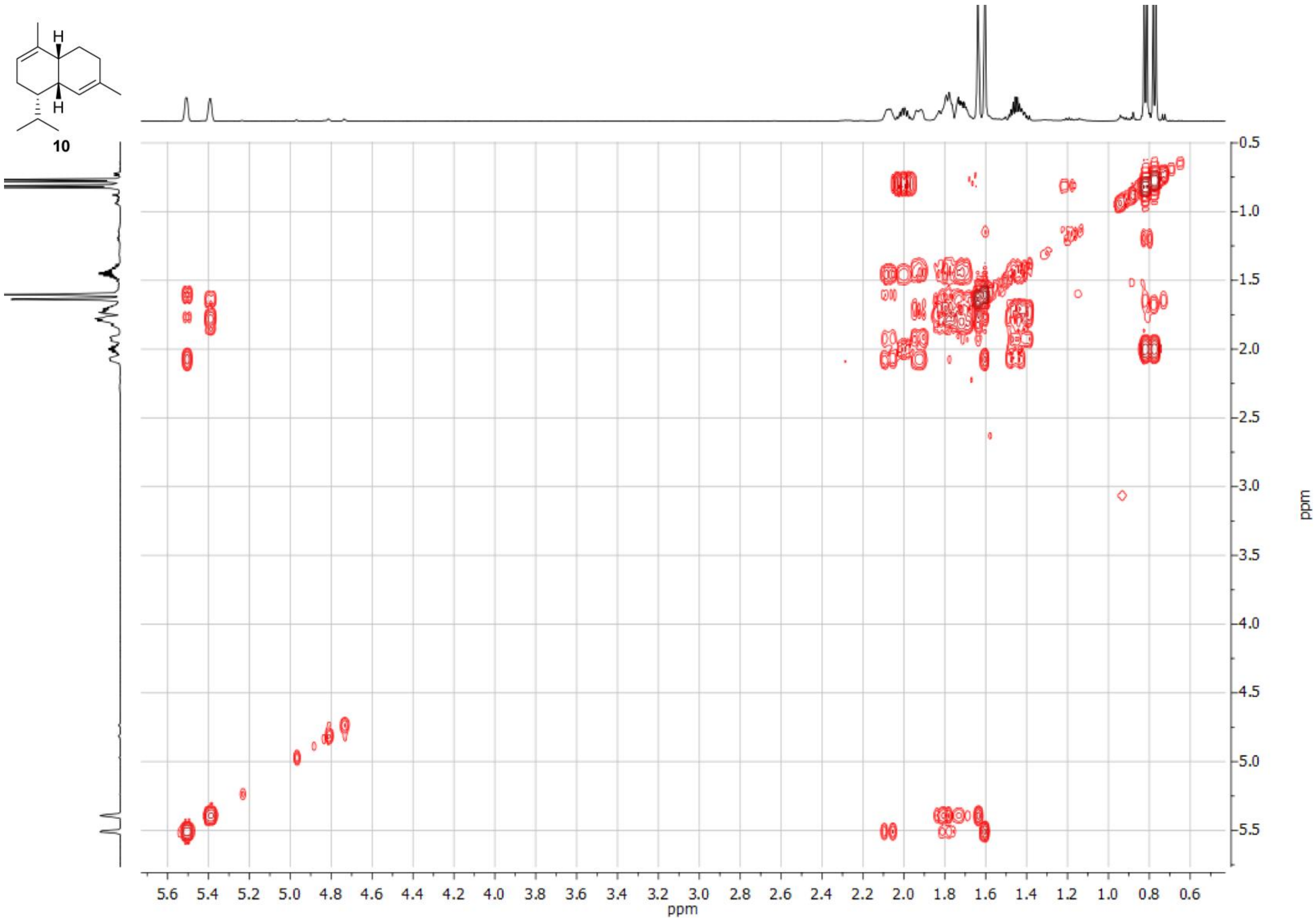
^{1}H, ^{1}H NOESY spectrum (500 MHz) of **9**.



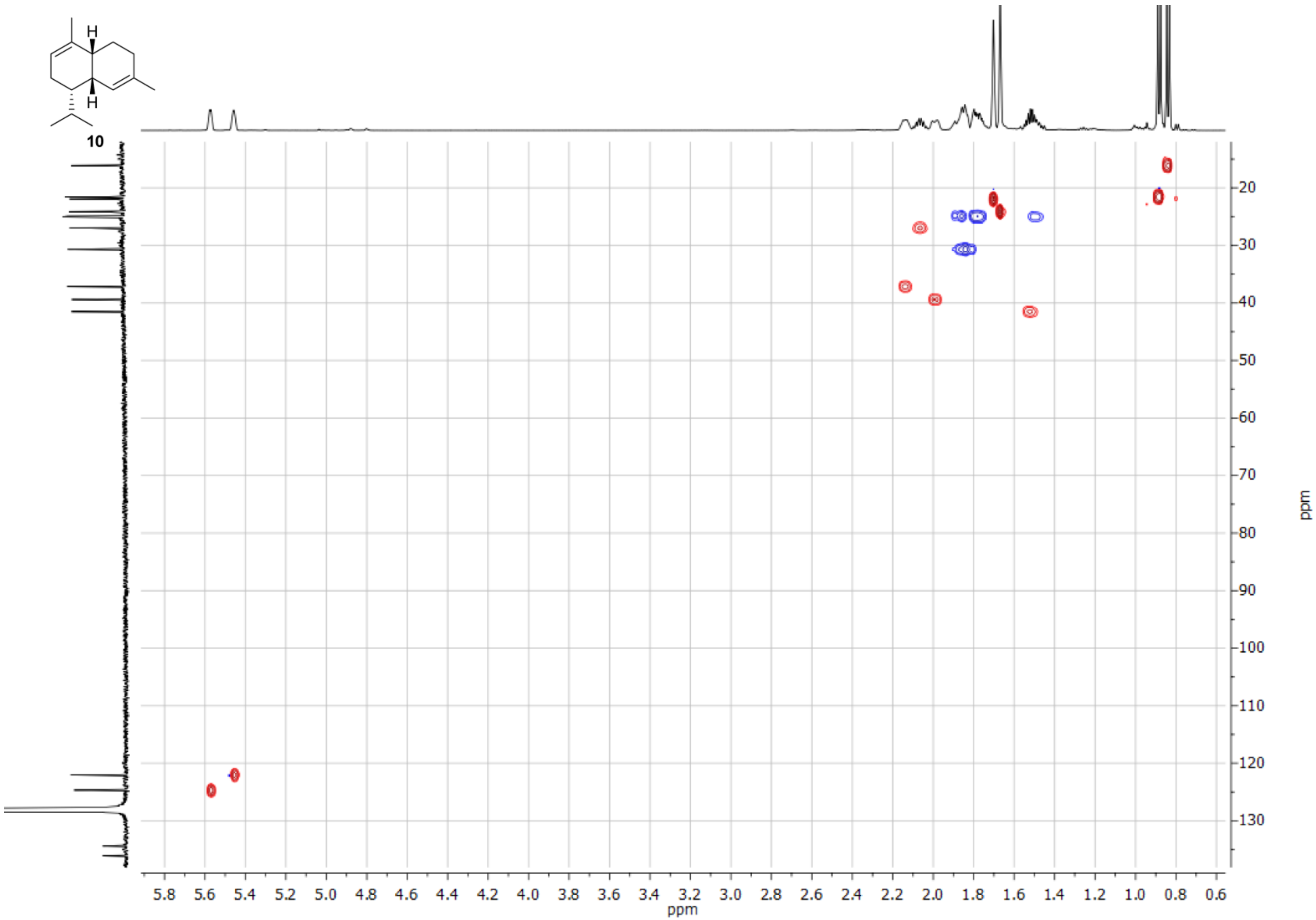
^{1}H NMR spectrum (500 MHz) of **10**.



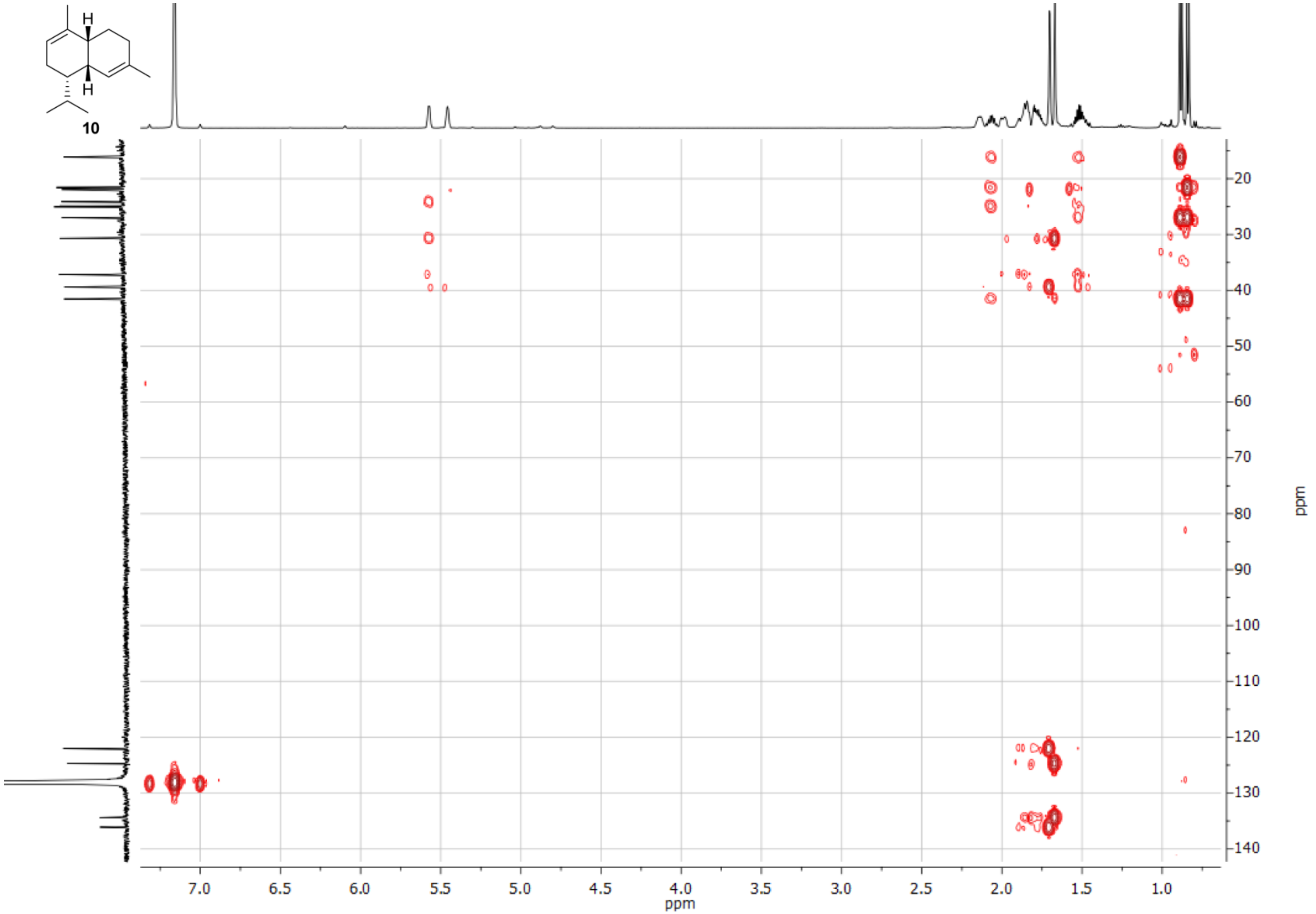
^{13}C NMR spectrum (125 MHz) of **10**.



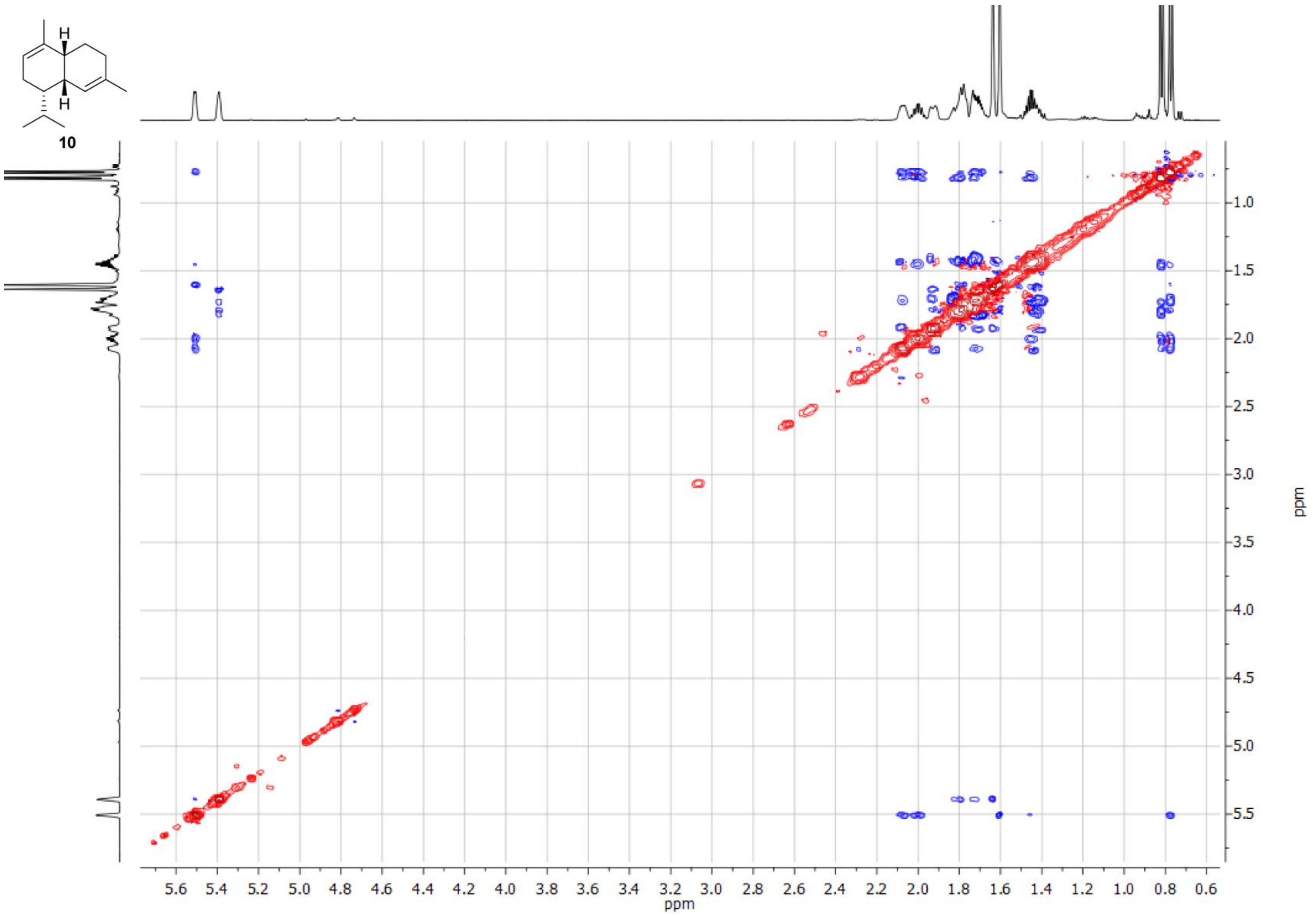
^{1}H, ^{1}H COSY spectrum (500 MHz) of **10**.



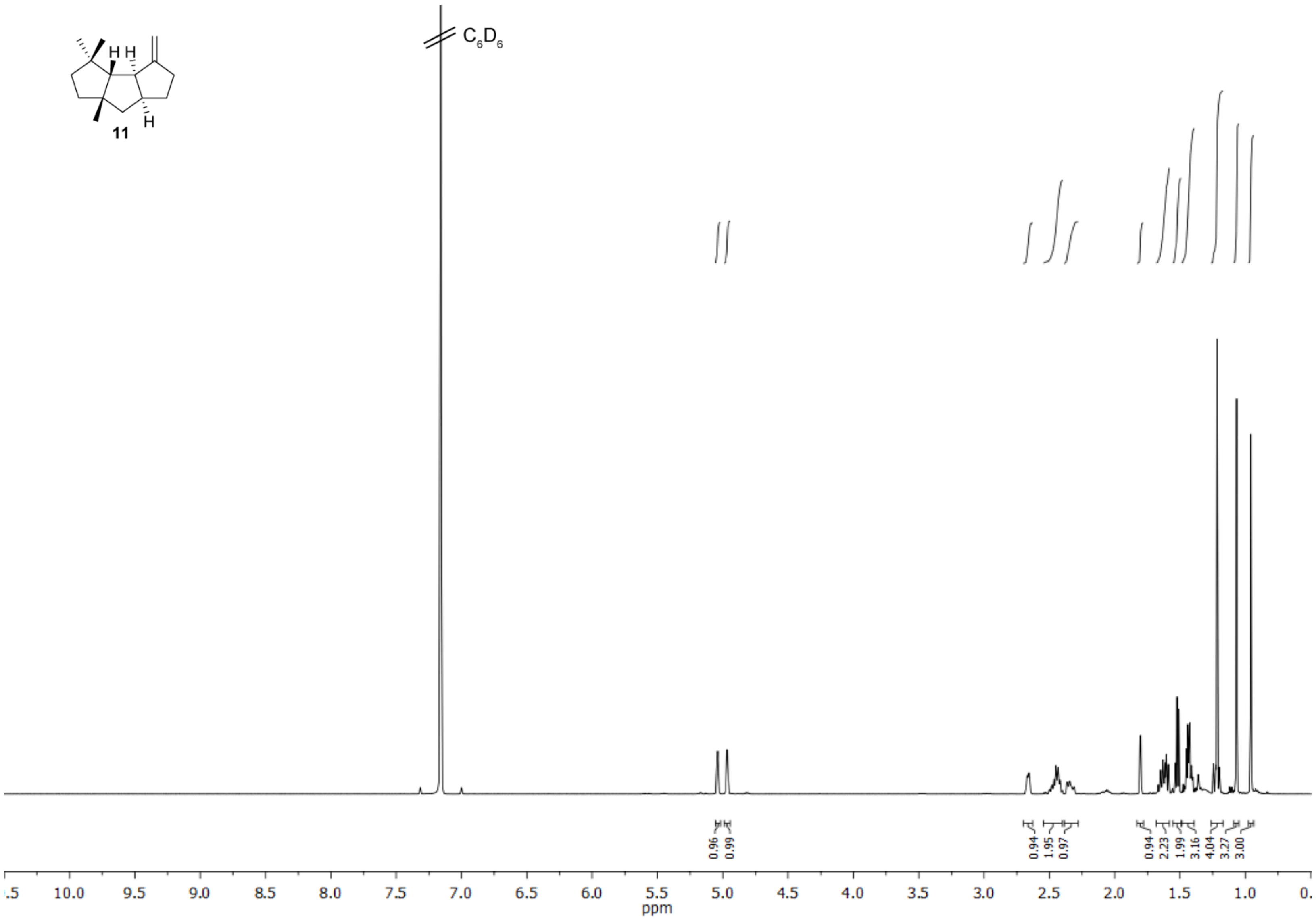
^{1}H, ^{13}C HSQC spectrum (500 MHz) of **10**.



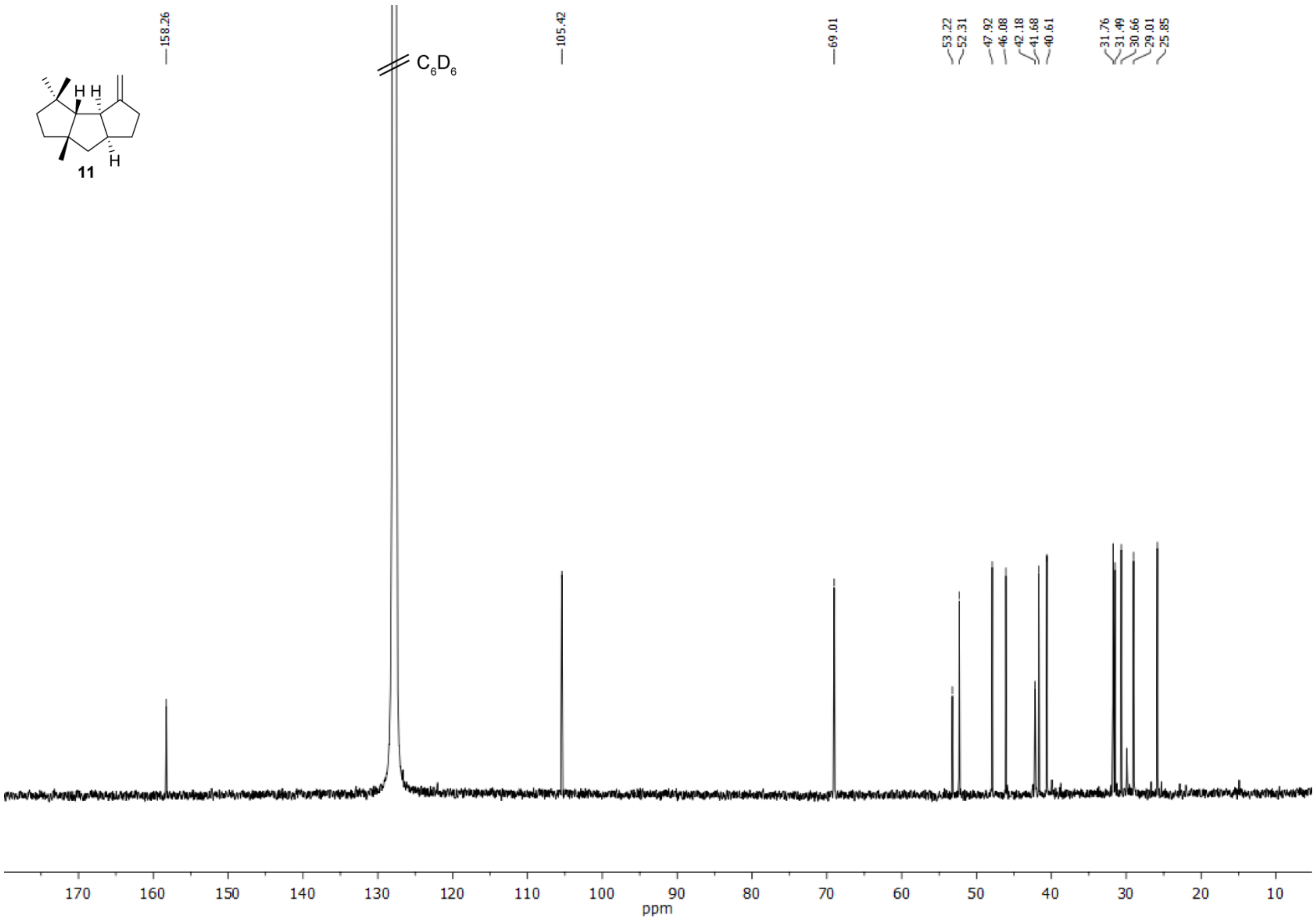
^{1}H, ^{13}C HMBC spectrum (500 MHz) of **10**.



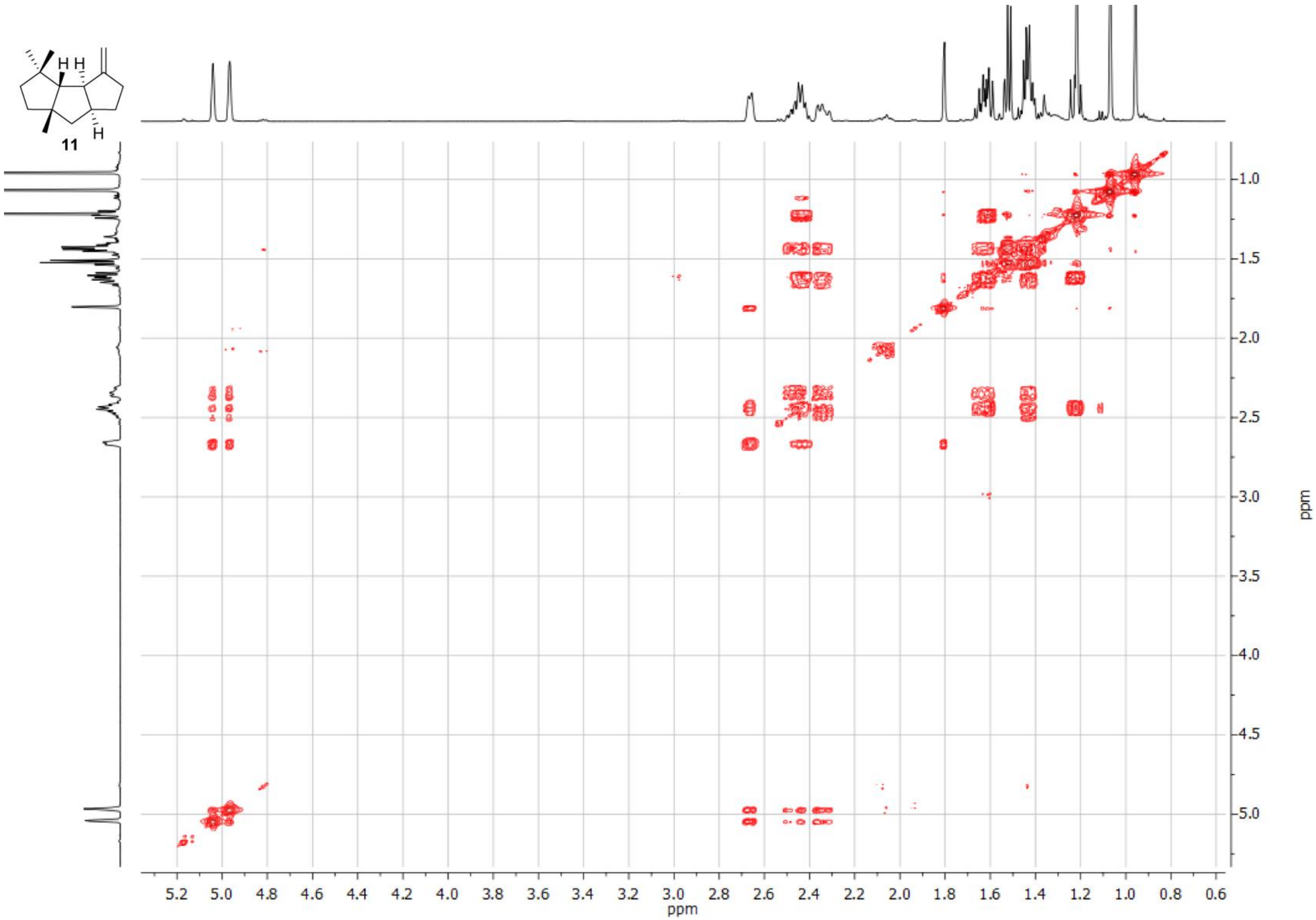
^{1}H, ^{1}H NOESY spectrum (500 MHz) of **10**.



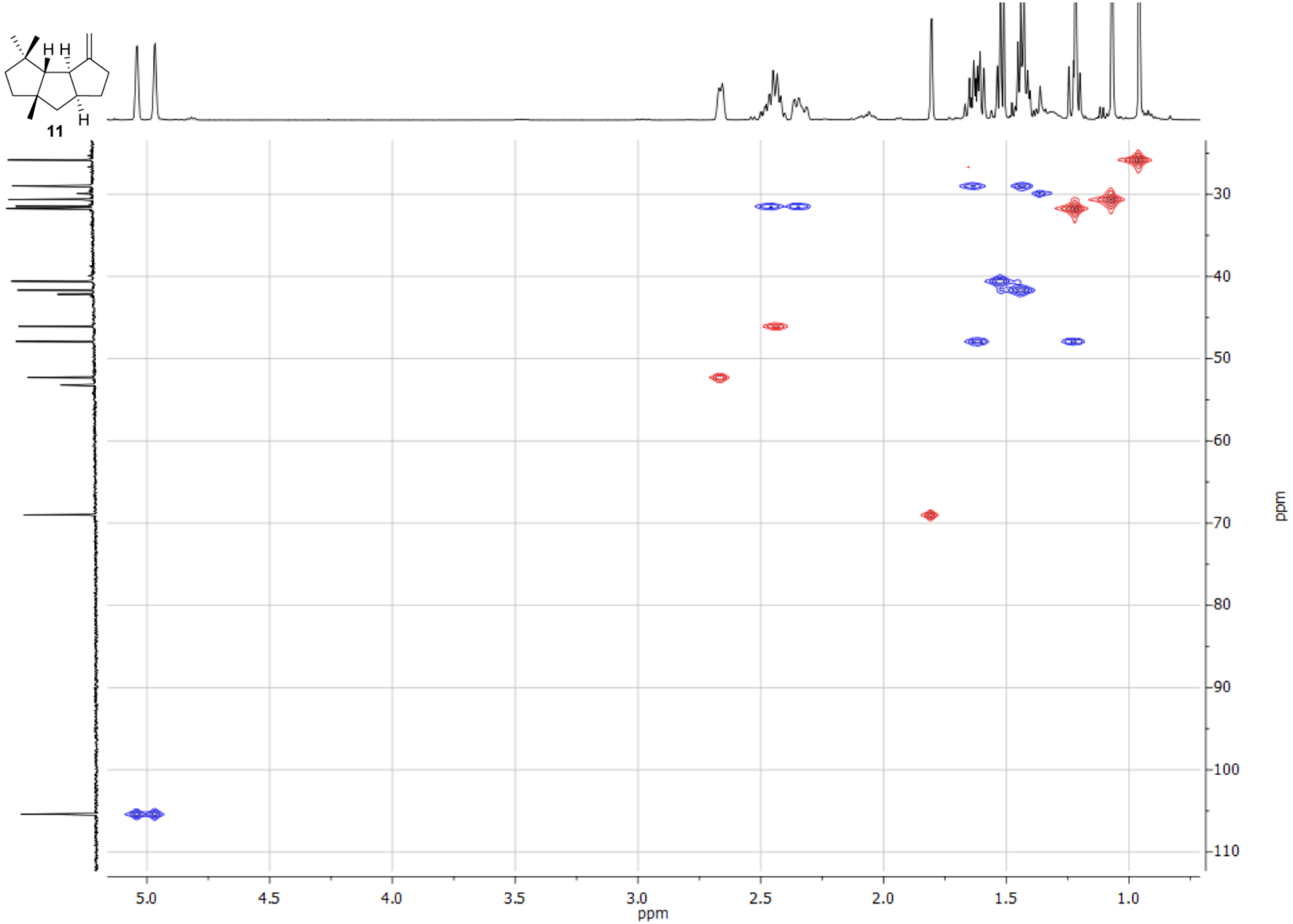
^1H NMR spectrum (500 MHz) of **11**.



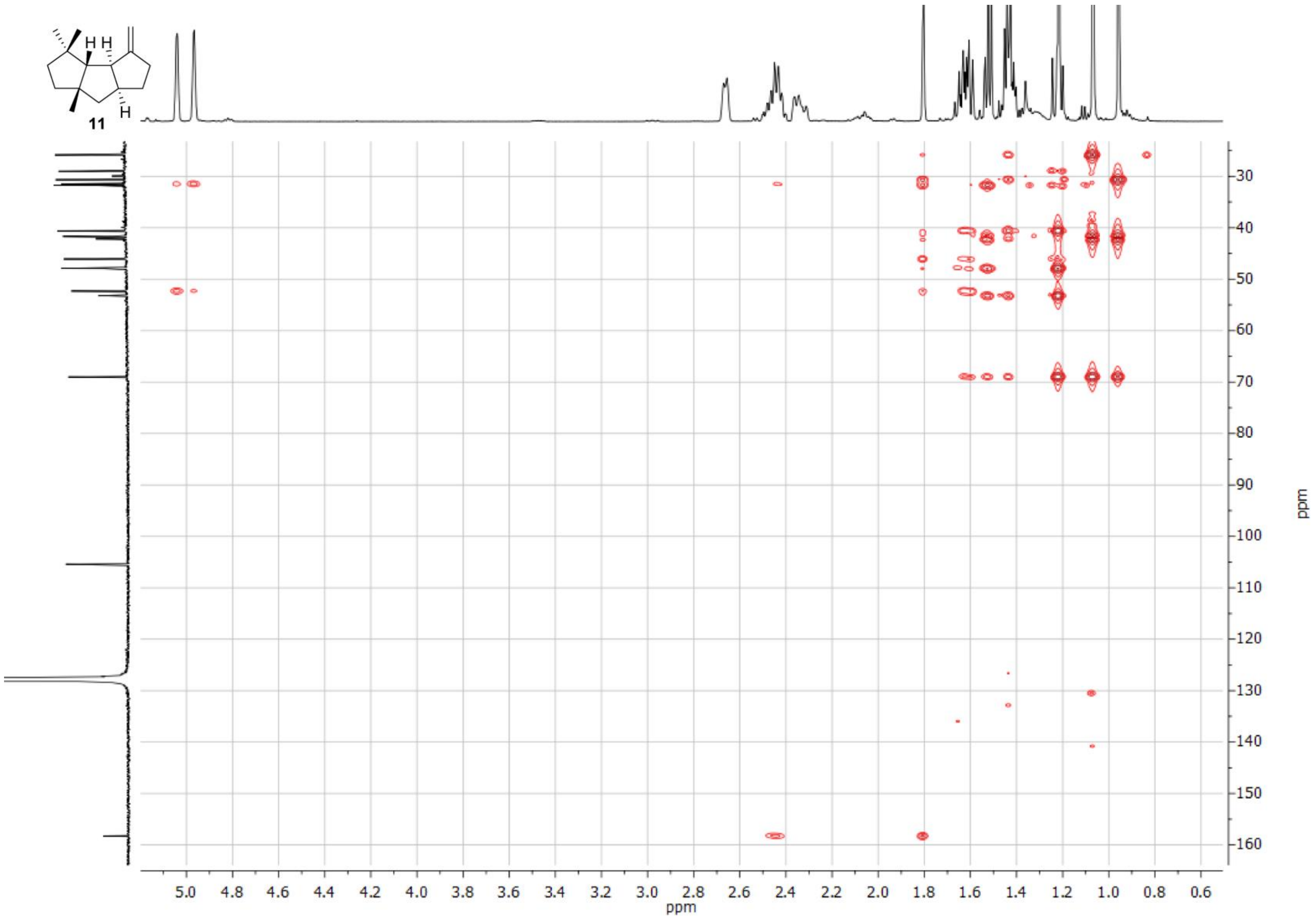
^{13}C NMR spectrum (125 MHz) of **11**.



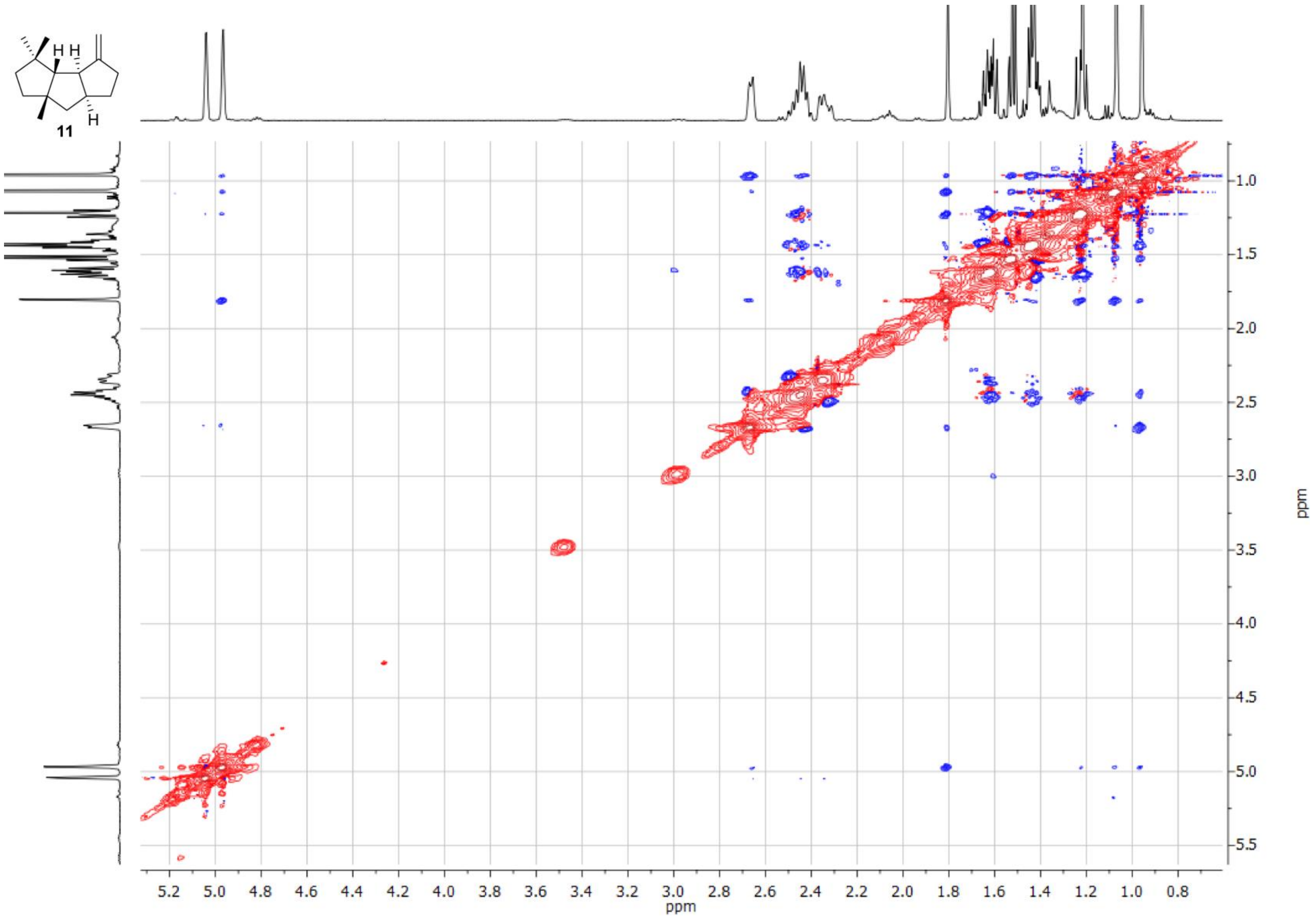
^{1}H, ^{1}H COSY spectrum (500 MHz) of **11**.



^{1}H, ^{13}C HSQC spectrum (500 MHz) of **11**.



^{1}H, ^{13}C HMBC spectrum (500 MHz) of **11**.



^{1}H, ^{1}H NOESY spectrum (500 MHz) of **11**

Supplementary Note 2 – Sequences

Capnella imbricata mutS sequence.

```
>C. imbricata mutS
ATACCTATGCAATATTTTAACTTAGCGGAGGAGAATTATTCTAAGTATGGATTATCAGTAATCCAGCTTATCCAA
ATTGGTAAGTTTTATGAACTTTGGCATGAGCCCGATACTTTTAGTAGGCACCAAGTTTACTTTCAAGCCGATTTA
TTAGTTGATCCATCCAGGCGAAGTGGGCCTTTGGGGGTAACGCCCCCGATTGAACAAGTTGCCTCGTTACTTGAT
ATGAGAATAACGTCTCCCGGTAAAAGATCCTTGCTTCAAATGGGGTTTCCAACTTATTCCCTTACTACCCACTTA
AGTACTCTGCTGGATAAAGGTTGGACTGTTATCGTTATCGATGAATTAGTCACTGGTAAATCCGGGCCAAAACAA
CGCGCAGTATCTCAGGTTTATTCTCCTAGTTGTAATTTAGAGGAATGTTTGGAATTGCCCTATGTGTTATCAATT
TATTTTTCTCGAGATGACTTATTAGGTATCACTTTATTTTCGGCCATGAATGGGCATAGTATAATGTTCCCCGTC
TCCTGGACGGACCGAGACAAAGTAGCCCGACTATTAATCGGCTATCATATTAGAGAAATAGTGATTTGGGCCCAC
TTGGGGGCTGGCTCGGATATTTTAATATATAGAATATATGATCTATTAATTGATTGGAATCTATTCCCCACTGAG
CCCAATGCTAAAATTGAAGTTATGG
```

TC Sequences

All characterized sequences are given as nucleotide sequences that are codon optimized for *E. coli* as used in this work and as amino acid sequences of the resulting proteins.

```
>BaTC-1_nucleotide_codon_opt
ATGGTGTTTAGCAAACCGATTCATGCGCCGAAAAAATGGGTGGAAAAAGGCAAAGGTCTTTTAGATGAAGAAGTT
GATGAGCAGTTAGAAGGCTTGCAAGAAATTATTGACACATTCATTCAGTGCGATCTGTGTGACGAGACCACCGTG
CGTAAATACCGTGGTAAATTGCGTCCGTATAACTTTACCCGTAGCTTAAATCCGATTATTCCTGAGACCAAACTT
GGCCGTGAAGTATTCAAAGCCCTGGCGCATTTTACAATGGGGATTTTTATCGCTGATGACATTATGGAAAAACAA
CCAGAGGAGGAAATCCGCAAGATGTGCACCACGTTTGCCCACTTGGACGCAAAGTCTCGCCAGAGTTTTCCGAGC
TTTCCGGCGATTGATGATATTAAGGAAGTGGTGGATGCGAGCAGCGTTTCTAAAAATGTGGCGGGTCCCATCGTA
TTTCTTGAACAGTTTCTTAACAAACAGGCAAGCTTGCTCCTGAGCCATGGTGATTTTTCGCTGTCTGATGTCAAG
GAATTCCGCTTACGGGGCTATAACATGATGGCCATTTATTTTCTGCGCGTTTTAGAAGAGACTGAGATTCAAAAA
AAACCGGGAAGCATCGAAGTAATTTGGCGTCGTGCGTTCGGCGGTTGCGTCCCATTTTATTTACTCACCGCTGAA
ATCTTTAGTGGTGCGATCGGCAAACTGAAAGAACATTATACCGCTCTGACCGAACTCTACGTGCTGGGAGGCATT
TATGTGGCAACCTTAAATGATGTATTCTCCTATTATCGCGAGGCGCATACCGTGGCGGATAATGTTGTTAAAGCA
ATGGTATCACAGAAAGAAGTGGAAACATTGGAAGAGGCCGCCGCGAACGTTGCTGAGTTTTTAGATAGCATTATG
CAGTACCTGTATGAAAAAGTCGAAGAACTGCAACGCAAAGACCCTAACAATCGTGAACTCCACGTTTGCCTTGAC
TATGTGGGCCGCATGACTGTAGGTTCCGTATATGCCCATTTATTGCCGCGTTACAAAGATACAACACTTTCATAT
TACCTGGAAGAAGTCAAGGAAGACCAGTTATCGTCGTGGCTGAGCAAAAAGGGTGAGTACGGCACACGGATGCTG
AAAAATATCCTTGAGGTCATTCACCAACGCATGGCAGACGGCCAGATGGAAGCATTGGCCGGTGCATTTCCGTAT
TGCGGGAAATTTCTTGGAGTTGTAATTCCGGATACTATTTGCTGTACCTCTAGCATCGTGTAA
```

```
>BaTC-1_amino_acid
MVFSKPIHAPKKWVEKGKGLLDEEVDEQLEGLQEIIDTFIQCDLCDETTVRKYRGKLRPYNFTRSLNPIIPETKL
GREVFKALAHFTMGIFIADDIMEKQPEEEIRKMCTTFAHLDAKSRQSFPSFPAIDDIKEVVDASSVSKNVAGPIV
FLEQFLNKQASLLLSHGDFSLSDVKEFRLRGYNMMAIYFLRVLEETEIQKKPGSIEVIWRRAFGGCVPFYLLTAE
IFSGAIGKLKEHYTALTELYVLGGIYVATLNDVFSYYREAHTVADNVVKAMVSQKEVETLEEAAANVAEFLDSIM
QYLYEKVEELQRKDPNNRELHVCLDYVGRMTVGSVYAHLLPRYKDTTLSYYLEEVKEDQLSSWLSKKGEYGTRML
KNILEVIHQRMADGQMEALAGAFPYCGKFLGVVIPDTICCTSSIV*
```

```
>BaTC-2_nucleotide_codon_opt
ATGACGAAACCATTTGAGGTTACCGCTCCAACGCGTTGGGTTAAATTCCATGCGCGCGTTCTGTCTCAGCCGCCG
ATCGAGAACTTCATCCGTATGGAACAACTGATCAACTGGGGCGTTGCTTGTGGAGTAACGAATGAAGCCGGCATC
CGGAAATCCTTTCAAAAGTTGAACGCATATGTGTATCTGCGCTGTATGTACCCGTTGGTGGCGGACGATGCGTGG
AGCGCGGAAATGTTCCGCATTAACCTTCACTTTACGACAGCCGGCTATATCGTGGATGACCGGATTGAGCGCTAC
ACGATGGAAGAAATGAATGAACTCTGCGACGGATACGACATGTTAGAAAAAGAAGTTTCTAAATTATTTCCGAAA
TGCCCATCGATTGAAGAAATGCGGTGTCGCCTCCAGCATCTGAAAAATAAATTTTCTATCGCCGCGATCACGATG
GTGATGGACTTTGTAAACCAATCAGCGTTATTTTTCTTACGTCAAGGCAAAACAAGCCGCGACCGTGTTGATAAT
TTCCGTAAGCGTCTGTCCAATGCAGTGACCATTTATTATCAGGCGGTCCGTAATAAGGTTAAAACGGGTTGCAAA
ATCACTGAAGGTGAAATGCTGTGGCGCCGTTGCTTCGATGTTTTGGCCGTACCTTCATATCTGGCGCCGGAGAGC
TTCACGCACGCCGTTGAAAAACAGGAATGGCCACTGGAAATGCTGTACGAATTGTACATGCTCGGCATTCTGTAT
TCCACAGTTATTAATGACCTGTATAGCTATCACCGTGAGAAGCTTGATGATTGTGATAATGTCATTAAAGTGTGG
ATGCAAGAAAACAGTGTCTCTAGCATTGAAGAAGCTAACGAAAAAATTTGTATGATCTTGGATGCCATCCTGCAA
ATTATGTATGAAAAAATTGAAAAGGCAAAAGCGCAGTATTCCAAGTGCCCGGAACTTCAGCGCCTGTTAGACTAT
ACCGGCATCGTCTCTGCTGGATGGATTTTTGTCCATAACACGGCGGCGCCGCGGTATCTTGAGAGCCCGTACCAG
GTTGTCCTTCGTGAAATCGAAGACAAGGATATCCAGGGGTGGCTTGAAAACAAGAACGAGTTTGGTTGGCGCGTG
GTACGCCATTTTATGGAAACGATGAAGTCCGAGAAAGGCAAACCGGTTATGGATGCTATGTGCGGCTTCACAGAT
GCCCGGAATATTCTTATCTGA
```

```
>BaTC-2_amino_acid
MTKPFEVTAPTRWVKFHARVLSQPPIENFIRMEQLINWGVACGVTNEAGIRKSFQKLNAYVYLRCMYPLVADDAW
SAEMFRINLHFTTAGYIVDDRIERYTMEEMNELCDGYDMLEKEVSKLFPKCPSIEEMRCRLQHLKNKFSIAAITM
VMDFVNQSALFFLRQGKTSRDRVDNFRKRLSNAVTIYYQAVRNKVKTGCKITEGEMLWRRCFDVLAVPSYLAPES
FTHAVEKQEWPLEMLYELYMLGILYSTVINDLYSYHREKLDDCDNVIKVWMQENSVSSIEEANEKICMILDAILQ
IMYEKIEKAKAQYSKCPELQRLLDYTGIVSAGWIFVHNTAAPRYLESPYQVVLREIEDKDIQGWLENKNEFGWRV
VRHFMETMKSEKGKPVMDAMCGFTDARNILI*
```

```
>CiTC-1_nucleotide_codon_opt
ATGTCCTGTTCCAAAGAGGTTCGTTTCCCGGAAACATGGTCCAAGGTTCAGTGGAATTGCTTAAAAACACCACGT
GATGAAGATATCTTTGGCTTCCAGGAACTGATTAAATGGTTGGAAGATTGTGGGGCGTCAGATGAAGCCAACGTC
CGGAAGTACGAAGAAATTCTGAACGTGTACTATGCAACCCGTGCCATCATTCCGATCGTGCCTGAAAACATTCTG
TGTCGCGAACTGTGCAAAGTCAAAATTCACTTCTTCATGACGACGTTCGTTACCGATGATTTGATGGAGGAGCAC
TGTAGTTTTGAAGATATGCGCAAAATCATTGCGGCGCTGAACCAGCTGAACGAGAAACTTTCACTGGAGTTTAAA
AAGGTACCGAGCATTAAAGAAATGAAAAAGTATCTTCAGTACATGAGTGATGAACGTGCGATTTCTCTGGTTACC
GTGTGCATGGATTTCATCAACCGTTGTACGCGCACGCTGCTCCAGCACGGTAATGTCTCCGATGACGTGGTTTTT
GATTACCGCAAACGTTGTTTTATCACCGTGAGTTTGTGGCTGCAAGGCGTGGCCATGGTCAAACGCGATGGTGCA
GCCCAGGATACCGTAGTCGATGTCTTGTGGACACGCGTTCACGATGGCCTCCCAAGCGCCTTCATCTTATATACC
GAAGTCAGCTCTTTTAGTTTGGGCAAACAGAATGCCCACATCGCGGAACTTACTGAACTGTATTGTTTATTTGGC
CTGTATATGGTGGCAATTAACGATTTCTATTCTTACACTCGCGAGATGAATGATACGATCTTCTCGCAATCCGTT
ATTAAAATTTGGCTGAAAAACAAAGAGATTAATAACTTTAGCGAAGCGGCGCCGAAGATGGTCGCCATTTTAGAC
GAGATCTCTAAATATGTTTACCGTAAAGGTGAGAAAATGAAAGCCGAAAACCCGACCAATTTCGAACTCCATGAG
ATGGTTGACCAGATCGGTTACGGTGTTGCAGGCAACTATTTCCTGCACCATCACGTTTGTAAACGGTACAAAGAT
ACGACAGCCATCAAATTGGGTATTAAAGAACTTAAAGGTAAGGAACTGAAACAGTGGCTGGCGGAAAAAGACCCG
TACGGCTGGGATATTGTCAAGAATTTTCTGGTTGATCACAGCATCGTAGCCAAGAACTTTGTCGACGTTATGCGC
GGCGTAACCCCACTGTCGGACGATTCACTGTTAAACAGCTTGAACAAATAA
```

```
>CiTC-1_amino_acid
MSCSKEVRFPETWSKVQWNCLKTPRDEDIFGFQELIKWLEDCGASDEANVRKYEEILNVYYATRAIIPIVPENIL
CRELCKVKIHFFMTTFVTDDLMEEHCSFEDMRKIIAALNQLNEKLSLEFKKVPSIKEMKKYLQYMSDERAISLVT
VCMDFINRCTRTLLQHGNVSDDVVFDYRKRCFITVSLWLQGVAMVKRDGAAQDTVVDVLWTRVHDGLPSAFILYT
EVSSFSLGKQNAHIAELTELYCLFGLYMVAINDFYSYTREMNDTIFSQSVIKIWLKNKEINNFSEAAPKMVAILD
EISKYVYRKGEKMKAENPTNFELHEMVDQIGYGVAGNYFLHHHVCKRYKDTTAIKLGIKELKGKELKQWLAEKDP
YGWDIVKNFLVDHSIVAKNFVDVMRGVTPLSDDSLLNSLNK*
```

```
>DgTC-1_nucleotide_codon_opt
ATGTGGTGCGGAAAAGAAGTTCGTGTGCCCCGTCAATGGGCTGTGCTGGAGCAAGAGGTGTTAAAGGAGAAACCC
GATCCGGAATTGGTCGATATCGACGGCATGATCAAATGGATTACGGAATGCGATATCGTAGACGAAAACGTCGTT
CGTGCCTATGTGAAGTTAGTGAAACCGTACTACTTTATTCGCCTGATTTATCCAATTCTTCCGAAAGATAAGTTG
TGTGTCGAAGCCTGTAAAATTTTCCTGCATTTCGTTATTAGCCTTTACTGCAGTGATGACCGCACAGAACTGGAA
TGCGACCTGGATGATATCGTGAAAATGTGCGACGCATATGAAAATTTAAAGGAACCGGTTTTCGAGATGTTTCCG
AAATTCCCGACAATTAAAGAAATGCAGAGTAGCTTAAAATTTCTGTCGGACTCGAAATTGATTTCGCCAGTTGTA
CTCTGTATGGATTTCGTTAACAAAGTTACCAGCTGCGTCATCGAATACGGTGAATATAGCGAAACAGCCGTTCTG
GATTTTCGTCGCCGTTTCCTGAATACGATTGAATTCTACCTGCGGGGTCTGGAGATGGAAAAGAAAATGCACCGC
AAAGACACGGAAAACAAGACTCTTTGGCGCCGTATTTTCGACGGTGGCCCGCTGTTTTTCTTCCTTATGTAGAA
ATTAGTAGCTTTTCATTGGGCAAAGTCGAGGGGCACATCCCGACCATCTCTGAAATGTACATCGTTTCTGCTTTT
GTCTGTATCATTGTGAACGACCTGTATTCGTACAAACGTGAAACCATGGAATCAATCAACTGCGATAGCGTAATC
AAACATTGGCTGTCGGATGAAAAGATCTCTTCGATGGAAGAAGCGTGCGAAAAAGTAAGTCGTATCCTGAACGCG
ACAGTACAGTATATGTTCGAAAAAGTGCAGCGTGTCAAGCTTGACTATCCTAATAGCCCAGAAGCGCACGTCCTG
TATGAATACATCGCGCACGTGACTATCGGTTGGCTCTACATGCATGAAGAAGGCAACGATCGGTACAAAGACAGC
CCATGGCGTGTCTCGCTGACCGAGGTGGAGGAGGATAAAATTCAGGAATGGCTGAGTTGCAAAGATAGCTACGGA
TTCGAAGCCTTGAACCAGTTCCTGGCCTCAAACCCAAAGGCCGAAAAAATTATTGACGCGCTGGGCGGCGACGAA
ATTATTCATGGCGGTCTGATCAACGAATCCTCCTAA

>DgTC-1_amino_acid
MWCGKEVRVPRQWAVLEQEVLKEKPDPELVDIDGMIKWITECDIVDENVVRAYVKLVKPYYFIRLIYPILPKDKL
CVEACKIFLHFVISLYCSDDRTELECDLDDIVKMCDAYENLKEPVFEMFPKFPTIKEMQSSLKFLSDSKLISPVV
LCMDFVNKVTSCVIEYGEYSETAVLDFRRRFLNTIEFYLRGLEMEKKMHRKDTENKTLWRRIFDGGPLFFLPYVE
ISSFSLGKVEGHIPTISEMYIVSAFVCIIVNDLYSYKRETMESINCDSVIKHWLSDEKISSMEEACEKVSRILNA
TVQYMFEKVQRVKLDYPNSPEAHVLYEYIAHVTIGWLYMHEEGNDRYKDSPWRVSLTEVEEDKIQEWLSCKDSYG
FEALNQFLASNPKAEKIIDALGGDEIIHGGLINESS*


>DgTC-2_nucleotide_codon_opt

ATGTTTTCTGACAATGTGCGGGTGCCGGAAAAATGGACGATTTCGCACAAAAAAATGCTTGCGGAACATGAGGAA
GATAAGGAATTAATCGCATTAGACAAATTATTGTTGTGGGTGAAAGAAACCGGTTTGACCACCGAAGCGGCGAGT
AAATCTGTCTTCGACAATCTGAACGCGTATCGTTATCTGCGCTGTCTGTTTCCAATCCTGCCAGAAGATCAGACT
GGCAAAAAACTTTTTCAGCTGAATCTGCACTTTCTGGTGATGGGCTACGTGATTGATGATACAATTGAAAAATAT
AGCGAAAAGGAAATGCGGGAACTGATTGACGGTTACAATTTTCTCCAGAATCAGGTATCTAAGACGTTCCCGAAA
TTGCCGAGCATCTCAAAAGTTTGGAAACTTCTGGGTGAAGCCCGTAACAAGTTTAGTCGTTCGGCAATTGTCACC
GTTGTAGATTACGTGAATCGTACCACCATGATTCTCTTGGAAGGCGAAATCTGCGAAGAAAGCGTTCTGAATTAT
CGTAACCGTCTGTCTAACGCTATTGCGGTGTATTTTAACGCAATTCTTTCAAAAACTAAAACCGGCTGTAAAATT
TCCGAAAACGAAATGCTGTACCGCCGTTGTTTCGATGCGTTAAGTATTGCAGTATATATGAGTACCGAAGTCTTC
AGTAAAACTCTGGTGCGCAACCAGATGTTGCCGACGAGCGAACTGTATAAGTTCTACCTGCTGAGCACGCTGTTC
TGTGTTGTGATTAATGATCTGTATTCATACGAACGTGACAAACTGGACGATACCGACTCTGTAATCAAAGTATGG
TATGCACAGAACGAGGTCACGGATATGAAAGCCGCGACCACTAAAGTGAGCAAAATTCTGGATCGTATCATCCAA
CAGATGTACCTGTTTGTGAAAGAAGGTAAATCGTGCCATCCAGAGCTGAGCGAATGGTTTGAACGCATTGCCGCT
ATGACGGTTGGTTGGATCTACATTCACAAATACGTCGTTTTTCGTTATGTGTCGTCTCCGTTTCAAATCGCCGTG
GTTGAAGTGAAAGATGAAATGATCCCGGATTGGTTGACAGAAAAAGACGCATATGGAGAGACCATTGTGCAGAAT
TTCCTGGATGGTTTAGAATATCCTCATCAGAAAGATGTCCTGGATAGCCTGTACGGTTGCAATCGTCTGTAA


>DgTC-2_amino_acid
MFSDNVRVPEKWTISHKKMLAEHEEDKELIALDKLLLWVKETGLTTEAASKSVFDNLNAYRYLRCLFPILPEDQT
GKKLFQLNLHFLVMGYVIDDTIEKYSEKEMRELIDGYNFLQNQVSKTFPKLPSISKVWKLLGEARNKFSRSAIVT
VVDYVNRTTMILLEGEICEESVLNYRNRLSNAIAVYFNAILSKTKTGCKISENEMLYRRCFDALSIAVYMSTEVF
SKTLVRNQMLPTSELYKFYLLSTLFCVVINDLYSYERDKLDDTDSVIKVWYAQNEVTDMKAATTKVSKILDRIIQ
QMYLFVKEGKSCHPELSEWFERIAAMTVGWIYIHKYVVFRYVSSPFQIAVVEVKDEMIPDWLTEKDAYGETIVQN
FLDGLEYPHQKDVLDSLYGCNRL*

>ErTC-1_nucleotide_codon_opt
ATGAGCATCTCACAAATGGTTTTCTCAAAAGAACTTCGCGTTCCGAAAGAGTGGTCGAAGTATCATTTCGATATC
GTCAAAGAGCCGATCGACCCGGAATTGTTCTCCGAGGACGAGCTGTTCGATTGGGTGGAAGATCTGGGGCTGACA
GATGATAAATCCGTCGTTACCAAATATGCACAACGCACCGGCGGCTACCTCTTCCTGCGTCTCCATATGGTCTTC
TTTCCGAAAAATTCCCTGTTTAAAAAACTTTTTAAACTGTGGGCCCATCTGGTCCCGTCGTTGTTCGTTACCGAT
GACTTCTTGGAGGCGACCAGCGAGATCGAAATGCGTCAGGTCTGTGATGGATTCGAGTTCTTGGCCGGCCAGATT
```

```
CGGGGACAACTGCCACAGTTCCCTACGATTGCGGAGATGAAACAGAGCCTGCTGTTACAGAAAATTGACGAAAAA
CTGATTCCGCACACGATTCACCTGATGGATTTCGCTAACAACGTAGCCAAAAGTATTATTCAGCATTCGAGTATT
CCGATCGAATCGGTGAACGAATATTGGCGCCGTCTTGTCGTTGTCTGCAGTCTGTATTATGCAGGGGTGGCATTT
GAGGTGAAACATTCAGTGCGTTCCTATAGCGAAGACGTCTGGACCGGCGTCTTCACCGTGGCCTATTTACTGTGG
CTTCCATGCCTGGAGATTCTGAGCGGCGCCGTTGGAAAAACCACCGAACACCTGTCACTGATTAACGAATTATCT
TTTTTAGCCACCTTTAGCGCACGCGTAGTGAATGATATCTATAGCTATAACCGTGAATTGGTACTCGAACACAAA
CCGGTTTCGATGGTGGCGCGTATCGCCGAATCGAAGGAGGTTACGACAGCGGACGAAGCGCTCATTAAAAACGTT
GAAATCCTGGGTGCGATCGTGAAGGTAATGTACCAGCGTATCGAAAAAGCAAAGCAGGAAAATCAAACCAACAAG
GAACTGTGCAAATGGCTGGATAATATTGGCGTCGGGACGATCGGCGTCCACTTTTGGCACCATTATATTCCCCGC
TATACCTCGGCGCCGTGCCGTCTGTCCTTCGTGGGAGTCGAGGAGGATGAATTGAAGGAGTGGCGCAAATGCTCT
GAAGAAGAACCTCCCAAAGAACTGTTTCCGGTGCTGTTGGACCACTCCCAGATCGCGAAACGCATCAACGATGCA
ATCATCTCCGGCGTCGTGCCGACGCAAGTTAAAAACGTCCTGTGTACCGATTAA

>ErTC-1_amino_acid
MSISQMVFSKELRVPKEWSKYHFDIVKEPIDPELFSEDELFDWVEDLGLTDDKSVVTKYAQRTGGYLFLRLHMVF
FPKNSLFKKLFKLWAHLVPSLFVTDDFLEATSEIEMRQVCDGFEFLAGQIRGQLPQFPTIAEMKQSLLLQKIDEK
LIPHTIHLMDFANNVAKSIIQHSSIPIESVNEYWRRLVVVCSLYYAGVAFEVKHSVRSYSEDVWTGVFTVAYLLW
LPCLEILSGAVGKTTEHLSLINELSFLATFSARVVNDIYSYNRELVLEHKPVSMVARIAESKEVTTADEALIKNV
EILGAIVKVMYQRIEKAKQENQTNKELCKWLDNIGVGTIGVHFWHHYIPRYTSAPCRLSFVGVEEDELKEWRKCS
EEEPPKELFPVLLDHSQIAKRINDAIISGVVPTQVKNVLCTD*


>ErTC-2_nucleotide_codon_opt
ATGTCGGGCAAAATTGTCCGTGCTCCATCGAACTGGACTACCCCACATAAAAAAATGCTTAAAGAAGACGAAGAC
CAGGAACTGATTGCATTTGATAAACTGCTCAGCTGGGTGAGCGAGACAGGGGTGGGTACGGAAGAACAGGCAAAA
ATCGTGTTCAAAAAAATCAACGCGTATTTCTACCTCCGGTGTCTTTATCCAGTGCTCCCTCGTGATCCAAAATCC
ATGAAAATTTTCCAGTTGAATCTCCACTTCTTAATTCTGGGCTACATTATCGATGATGCTATCGAAAAGTACAAC
GAGAATGAAATGAAAGAGTTAATCGCGGGTTATAATCTTCTGCAGTCTCAGGTGTCCGAGACGTTTCCCAATTTT
CCAAGTATCTTTGAAATGAAACAGCTGCTGGGAAATATGAAGAATGACTTTTCCAAAAGCGCAATTACCACGATG
TTCGATTACGTGAATAAAACTACACTGATTCTGCTGGAAGAAGGTGAAATTGCAGAACACAACGTGCTGAACTAT
CGTAAACGCTTGAGCAACGCAGTTGCTGTATATTTCGATGCTCTGTTGAGCAAGACCAAAACCGGGTGCGAAATC
TCAGACGGTGCGATGTTATGGCGCCGTTGTTTTGATGCACTGGCTATCGCTGTGTACATGTTAACGGAGGTATTT
TCTAAAACCCTGGTTAAGAATCACACGTTGCCGGTCAGCGAATTTTATAAATTTTACCTCCTGTCTATTCTGTTT
TGCGTGGTTATTAACGATTTACATTCTTACGAGCGTGACAAATTGGATGATACCGACTCCGTGGTTAAAGTCTGG
TTCAAGGAAGGATCCGTGGCAAATATGGAAGCTGCGACCTCCAAAGTTTCCAAGATCCTGGATGCCATTATCCAG
CAGATGTACCTCTTCGTGGAAGAAGGCAAAGCGCGCCATCCGGAACTCAGTGAATGGTTTGAACGTATTGCGTAT
ATGACCGTGGGTTGGATTTACATCCACAAAACCGTGGTACCGCGCTACGTGAGTAGCCCGTTTCAAATCGAGGTC
GTGGAAATTCACGAGAACATGATTTCTAACTGGCTTCTCAAGAAGGACGCTTACGGCCAACGCGTGGTGCAGCAG
TTTTTAAAAAATTTGAACGACCCTCAAAAGAAAAACATTATGCTGTATGACGAGTGA

>ErTC-2_amino_acid
MSGKIVRAPSNWTTPHKKMLKEDEDQELIAFDKLLSWVSETGVGTEEQAKIVFKKINAYFYLRCLYPVLPRDPKS
MKIFQLNLHFLILGYIIDDAIEKYNENEMKELIAGYNLLQSQVSETFPNFPSIFEMKQLLGNMKNDFSKSAITTM
FDYVNKTTLILLEEGEIAEHNVLNYRKRLSNAVAVYFDALLSKTKTGCEISDGAMLWRRCFDALAIAVYMLTEVF
SKTLVKNHTLPVSEFYKFYLLSILFCVVINDLHSYERDKLDDTDSVVKVWFKEGSVANMEAATSKVSKILDAIIQ
QMYLFVEEGKARHPELSEWFERIAYMTVGWIYIHKTVVPRYVSSPFQIEVVEIHENMISNWLLKKDAYGQRVVQQ
FLKNLNDPQKKNIMLYDE*
```

```
>HcTC-1_nucleotide_codon_opt
ATGGCATTCAGTAAGCCGATTTACGCGCCTAAAAAATGGGCTGAAAAAGCCAAAGGAGTTACATCAATGAACGAA
GAAAAACAATTAGTGGGCATGGATGAACTGATTAACTGGGTCGCAGAGTGTGAACTGGCAGGGGAAGCCACGGTG
CGGAAATATTGGGATACGGTGCGCCCTTATCATTTTATGCGTTGCTTGTATCCAGTGTTTCCGGACAACATCCTG
TGCAAGGAATTTTTTAAAGTGTTGTTACATGCCAGTATTTCTATCTATATTGCTGACGACCAGTTGGAAAAACAG
AGCCGTGAAGAAGCTAAAATTGCATGTGCCGCGCATGCACGTATTGATCAGCAGAGCACGAAGCGTTTTCCGAAT
CTGCCCACCATCCAGGAGATGAAACAGATCCTTAGCACGTTTTCTACGCCGTCGATTGTTGGCCCGACAACGATG
TTTGCTGACTTCGGCAACCGCCTCGCAAAGGTGCTGCTGCTGCATGGTAATTGGAACGGGAACGTGGTGGCGGAT
TTTCGTCTCCGTAATTCTAACTTAGTTAGTATGTATTTTCAAGCGATCCAGGCCGAAAAGACTCCGAATAAAAAG
AACACCACGTTAGAGACCCTGTGGCGTCGGATCTTTGGTGGTGTCGCAGCCCCTTACTCTTCCCTGGCCGAAATC
GCAAGCGGGGCAATTGAAAATTCAAAACAACACATTGCGGTCATCACTGATATGCATCTGCTGAGCGGCTTGTTT
TCGACCACAATTAATGATCTGTATTCATATTTCTTGGAGAAGTCAAGTATCTGCGACAATGTAGTCGCAGCCCTG
CTGCATGAAAAAACAGCGGAAAACGTGAGCGAAGCCGTTGATAAAGTTGCACAGATTCTGGACGCCATTTTGAAA
CTGATGTATAAAAAAGCGGAACAAATTAAACTGCAGTTCCCGGATAATCGCGAGCTGCGCGATTTTTTCGATTAC
GTTGGTCAGGTGACAGTCGGCTGGTACTATTTACACGAGTGCGCATTACCCCGTTATCAGGGTTCGCCTTGGCGT
GTACATCTGGAAGAGATTGACGAAGACGAAATCCCGAAATGGTTAAGTGAGAAGGGCAAATACGGCTCTAAAGTC
ATGCAGGAGTTGATGGAATTAGTGCAGGAACGCATGGCGGACGGTAAAATGGACGCGGTTCACGGGAAATTTCCG
ATTAACGAAAAATACGTGGCAGAAAAAGAAGATGGCCTGTATAAAATTATTTAA
```

```
>HcTC-1_amino_acid
MAFSKPIYAPKKWAEKAKGVTSMNEEKQLVGMDELINWVAECELAGEATVRKYWDTVRPYHFMRCLYPVFPDNIL
CKEFFKVLLHASISIYIADDQLEKQSREEAKIACAAHARIDQQSTKRFPNLPTIQEMKQILSTFSTPSIVGPTTM
FADFGNRLAKVLLLHGNWNGNVVADFRLRNSNLVSMYFQAIQAEKTPNKKNTTLETLWRRIFGGVAAPYSSLAEI
ASGAIENSKQHIAVITDMHLLSGLFSTTINDLYSYFLEKSSICDNVVAALLHEKTAENVSEAVDKVAQILDAILK
LMYKKAEQIKLQFPDNRELRDFFDYVGQVTVGWYYLHECALPRYQGSPWRVHLEEIDEDEIPKWLSEKGKYGSKV
MQELMELVQERMADGKMDAVHGKFPINEKYVAEKEDGLYKII*
```

```
>HcTC-2_nucleotide_codon_opt
ATGGCATGCTCCAAAGAGATGTACATTCCTAAAGAGTGGACGAAATACCACCACGACATCGTGAAAGAGGTGAAA
CGTGCAGACCTGGAAGCCGATGATGAGGTCGCGAAATGGGTGGTGGGTTTAGGCCTGAGCGATAAACCGGCAACC
GAGAAATATATCGATAGTACTCGTCCTTATCATTTTATGCGCTGCATCGCTGTCCTGGTGCCATCTAATCCACTG
TGTCGTAACATTTTCAAAACCTGGCAGATCTACGTCACGGGTCTGTTCGTATCCGACGACTATCTGGAACAGACC
TCCCTGGCGGAGGTAGTCCAAGTGAGCGATTCTTTCGATATGCTTAATGAACAGATCAGTGAGCGTTTTCCCGTG
ATTCCAACCCACGCGCAAATTATGCGGTCGCTTAAAATCAACGAAAAATTGATCCCGCATGTGCTGGTCCAGATC
GATTGGTTAAATACCATTGCTAACGACCTTCTGCAACACGGCGATTTTTCAGAAGATGATGTATGGGATTTTCGG
TGTCGTATGTCGGCGTGGATCCGTGCGTATTTTGATGGGCTGAAAAGCCAGGTCGGTCCGAAAGTAAAAGTTACG
ACCGATGACTTTATGTGGCTGCGCATCATTGATGGAGGCGTCGGCCCGTTCCTGCTGATTAATGAAATCCTGGTG
GGCGAGCTGGGTAAGTGCAAAGGTCATGTTAAACTTATTACGCAGCTGTACATCCTCTCTAGCTGCGTTTGCGTA
CTGATCAACGATTGCTATTCGTATCATCGTGAAACCGTAGTGACCATCTACAACACAATTAAAGTCATTTTAGAA
AATAAAGAAGTGTCGGAAGTGCCAGATGCGGTTAGTAAACTGTTGCAGACGATTAACGCGATCATTAAATACATG
TATCAGACCATTGAAAAAACGAAACATGAATATCCGGATAACCTGATGCTGAACAAACTGCTGGATGCTATCGGT
GGTGCGACGGCGGGTTGGTTCTTTGTCCACGACAAAGTCTTGCCGCGTTATCAGGCGACCTCTTGGCGCTTCTCA
CTGGTGGAAGTCGAAGAAATGGAGCTGAAAGAGTGGCGCAAAAGCGCTGATGAAAAGCCCTCAGATCTTGTACAA
CCTCTGCTCTACCGCTTTAATGCTAAAGGCAAAGAAATCATTGATGCGCTTGCTTACGTCCCCAAACGTAGTCGG
CGTATTGTGCCAGAGTGA
```

```
>HcTC-2_amino_acid
MACSKEMYIPKEWTKYHHDIVKEVKRADLEADDEVAKWVVGLGLSDKPATEKYIDSTRPYHFMRCIAVLVPSNPL
CRNIFKTWQIYVTGLFVSDDYLEQTSLAEVVQVSDSFDMLNEQISERFPVIPTHAQIMRSLKINEKLIPHVLVQI
DWLNTIANDLLQHGDFSEDDVWDFRCRMSAWIRAYFDGLKSQVGPKVKVTTDDFMWLRIIDGGVGPFLLINEILV
GELGKCKGHVKLITQLYILSSCVCVLINDCYSYHRETVVTIYNTIKVILENKEVSEVPDAVSKLLQTINAIIKYM
YQTIEKTKHEYPDNLMLNKLLDAIGGATAGWFFVHDKVLPRYQATSWRFSLVEVEEMELKEWRKSADEKPSDLVQ
PLLYRFNAKGKEIIDALAYVPKRSRRIVPE*
```

>HcTC-3_nucleotide_codon_opt
ATGAGCTGTAGCAAGGAGGTGCGCGCCCCTCGCAAATGGGTACAACGCCAGCGGGAAATGCAAAAGGTGAAAGTA
GATGACAACTTGATTTCCATGGATGAGCTGATTGAGCTGGTCGTTGAGTGCGGTCTGTGTGACGAGGCCTCCGTG
CACAAAATGTATGAACGTACCAACACCTATCAATTCATGTGGACGCTGATCCCTACAGTACCTAGCGATCAGTGG
CAATTCGAGATTTTTAAAACCAGCTTGCACTTTCTTTGCGCACTGTTCCTGGTTGACGATGCGGTCGAGTCATAT
TCTGAAAAACAGATGAAAGAGATGAGCGACGCGTATGATATGTTGGAGAAACAGGTTTGTGAAAACTTCCCGTGC
TTCCCTAGTATTAACGAGATGAAACAGAGCCTGAAACACCTGAAAAATCCCTTTGATATCGCTTCAATTACGTTC
TGCATGCATTATGTGAACAAAATTGCGAGCGTGTTAGTCAATGAAGGGAGCACACCACAGAACGTTGTCTATAAC
TTCCGCCGCCGCACCAGCAATGCCATTTCAATTTCCTTTCAGGCGGTTTTGATTAAAGCTAAATTAGGGAGTAAC
GTGACCGGTGACGAAATGCTTTGGCGCCGGGTGTATGATGGCTTAGTGATTTTGTTCTACCAGTTCGGAGAACTG
GTCCTGGGCCTGACTAAAAATACTGAAAAATACATGACGGTCATCACGGAATTACGCCTGTTGGGGTGCTTGTAT
TGTATTGTAATTAACGATCTCTATAGCTACCACCGCGATAAGCTGGTCTCATCCGATAACATCATTAAGACCTGG
CTGTTGGGCAAAGCAGTTAGCAATTTGAGCGAAGCCTCTTCGCGTTGCTGCGAAATCCTCGACGCTGTCATGCAG
TATATGCACGAACGGGTGGAGCGCATTCGTCAAAGCCACCCTGGTTGTCCGGAATTGGAGGCCCTCCTGGAGACC
ACTGTTTACACTACAGTAGGGTGGATTTTTGCTCACACTGCGGTGGTGCCACGTTACAGTGAAAGCCCATTAAAA
GTACACCTTGTTGAAGTTGAGGAATCTGAGCTGCATACCTGGTTCTGTGAAAAAAACCATTATGGTCGCTCCGTT
ATCGAAAAATTCCTGAAAGCTGTGAATGATGAAAAACATAAAGGGATCCTGGACGCCCTGAGCGGGTTTGTGGAT
GGCCGTGGGCAACTGTTAAAAACTCAGCTCTAA

>HcTC-3_amino_acid
MSCSKEVRAPRKWVQRQREMQKVKVDDNLISMDELIELVVECGLCDEASVHKMYERTNTYQFMWTLIPTVPSDQW
QFEIFKTSLHFLCALFLVDDAVESYSEKQMKEMSDAYDMLEKQVCENFPCFPSINEMKQSLKHLKNPFDIASITF
CMHYVNKIASVLVNEGSTPQNVVYNFRRRTSNAISISFQAVLIKAKLGSNVTGDEMLWRRVYDGLVILFYQFGEL
VLGLTKNTEKYMTVITELRLLGCLYCIVINDLYSYHRDKLVSSDNIIKTWLLGKAVSNLSEASSRCCEILDAVMQ
YMHERVERIRQSHPGCPELEALLETTVYTTVGWIFAHTAVVPRYSESPLKVHLVEVEESELHTWFCEKNHYGRSV
IEKFLKAVNDEKHKGILDALSGFVDGRGQLLKTQL*

>PbTC-1_nucleotide_codon_opt
ATGAGCTGCTCAAAAGAGATCCATGCGCCGCGCCGCTGGGTGGATCGTCATAAACAGATTCCGGTATTGCCGCAG
AACGCGGTGGAAAAACTGATCTCTATGAATGAACTGATCGAGCTGGTCATTGAGTGTGGCCTGTGCGATAAAACG
TCGATTAATAAAATGTATGACAAGATTAACACTTACCAGTTTATGTGGTGCATCGTGGACACCGTCCCTGCGAGC
CAGTACGCGGAAGAAATTTTCAAAAGCTCATTACATTTCTTGTGCGCCTTATTCTTAGTAGACGATGCCGTAGAA
AGCTACAGTGCTAATGAGATGCAGGACCTCAGCCGTTCTTATGATACCCTGGAAAAAGAGGTATGTAAAACCTTC
CCGAACTTTCCGAGCATTAACGAAATGAAGGAGTCACTGATGCATCTGCGCAATCCGTTTGATCGTTCCTCCATT
ACTTTTTGCATGCAGTATGTTAATAAGATTACAGCAATCCTGCTGGAAGAGGGAAACACGCCGCATCATGTAGTT
TATAATTTGCGTCGCCGCACTTCGAACGCGATCTCTATTGCCTTTCAGGCAGTCCTGATTAAAAGCAAATGCGGC
AGCATCATTACCAGCCACGAAATGCTGTGGCGTCGTGTTTTCGATGGACTGGTTATTTTATTCTATCAGTTTGGA
GAACTGATCTCTGGTGCAACGGAGACGGCGCAGCAGCATATCGCAGTCGTTACGGAACTGCGCATGTTAGGTTGC
CTGTATTGTATTGTTATCAATGATCTGTATTCGTATCAGCGTGACAAGCTGGCCTCCAGCGACAATATGATTAAA
ACCTGGCTGCTGGAAAAAACAGTAAGCAGCCTGTCCGAGGCCACTGCGCGTTGTAGCCAGATCCTGGACGCCATC
ATGAAATATATGTATCAGCGCGTCGAGCAATGTATCCAATCTAATCCGGGCTGCCCGCAATTGGAAAGCCTGTTG
GAAACCACAATCTATACCACCGTTGGTTGGATTCGCTCTCATACGACGGTCGTTCCGCGTTACTCGGAAAGTCAA
TTGAAAGTTGCACTGGTCGAGGTTGAAGAGAAAGAACTGCCGAAGTGGCTCGCTGAAAAAGACGAATACGGCTGG
AACGTTGTGGAAAAAATTTGTCGAAACGCTGAATGACGAAAAACATAAGGGCCTGCTCGATGCACTCCAGGGCATT
GCTGATGGGCGTGACCAACTGCTGAAGACCCAGCTCGATATTAGTTGA

>PbTC-1_amino_acid
MSCSKEIHAPRRWVDRHKQIPVLPQNAVEKLISMNELIELVIECGLCDKTSINKMYDKINTYQFMWCIVDTVPAS
QYAEEIFKSSLHFLCALFLVDDAVESYSANEMQDLSRSYDTLEKEVCKTFPNFPSINEMKESLMHLRNPFDRSSI
TFCMQYVNKITAILLEEGNTPHHVVYNLRRRTSNAISIAFQAVLIKSKCGSIITSHEMLWRRVFDGLVILFYQFG
ELISGATETAQQHIAVVTELRMLGCLYCIVINDLYSYQRDKLASSDNMIKTWLLEKTVSSLSEATARCSQILDAI
MKYMYQRVEQCIQSNPGCPQLESLLETTIYTTVGWIRSHTTVVPRYSESQLKVALVEVEEKELPKWLAEKDEYGW
NVVEKFVETLNDEKHKGLLDALQGIADGRDQLLKTQLDIS*

```
>RmTC-1_nucleotide_codon_opt
ATGCTTCGTACGGAAGTGAAGATTCCGGAAGTATGGTGTATCCCCTGTGATCTGAAATGCAGCACGCCACCGCAG
CTGAAAAAGGATATTATCGATTGGTGCGTCCGCACCGGCATCACCGCCGACCGTCGTACAGCGGAGAAAGCCCTT
GATCGTCTGAAACCGTATATGTATATGAAAATCTTGTTCCCAAAACTGAAACCGGATAACCCGTTGATTGCTTCA
TTGTACGAACTGAACGCTAGCGTAATTGTTACCGGCTTCACCATGGACGATGTTCTGGAAACGTACACGCTGAGC
GCCATTGAGGAGCTGGATCACACTTTTCGCGAAACTCAAAACTGGTTGTCACCGCTGCCCCCGAAGGACTACCCG
GACTTGGATGCCATCCTCGACCACATTCCAAGCACAAAACCGCCTTATAGCCGGAGTGTTGTTGCCATGTTTACC
GACTACTTTAACCGTTATTGTGCCGCGATTCATACCGCGTACAAACCCTCCGAAACGCGTTTGGGCGCGTTTCGT
CGCCGCCTGGCGGCGGCGGTAACCGCGTATCTGGAAATGGCCAAACGCAAACGTCGCCGCGACGGTAATCTGCAT
GAGGAAGAATTCTTATGGCAACGCAGTGCTGATAACCTGAGCTTTCCGGTGCTCATGCTGGCAGAAGTCTTCACC
GGAATCCTGGCGGATGCAGATGTCCCGTCGACTACCTTGCACTATTATCACTTTTACTCAAATTTATTCGCGATC
GTACTGAACGATCTCAATAGCTATCATCGTGATAAAGATAACGACGACAACTCTCTCGTTAAACTGTGGCTTAAA
CTCGGCGTTGCTAAGGATTTTGATGATGCCGCCACGAAGATTATTGACTTTCTGAACTCTATTATGGTGCGGATG
TATGTTGCGACGAAAGAACTGCTGGAGTCAAATCCGGAGAACGAACCCCTCCGCCGCTTCGTGGAAGCGGTGGGC
TATACGTTCAACGGTTGGATCCTGGTACATACAACCGCGGTTGAGCGTTACAAACTGAGCCCGTTCCAGACGGTG
GTTGCGCCGGTGGCGCGCGGTAAAGAAGAAGTCTGGTTAAAGGGCGAAACGGCGTTTGGCAAACGGTGCGTTGCA
ATGTTCGAAGAGCTGATGGGTGAACGTGCCGAAGAAATGAACAAGCTGTACGGGCTGGACTCCTAA
```

```
>RmTC-1_amino_acid
MLRTEVKIPEVWCIPCDLKCSTPPQLKKDIIDWCVRTGITADRRTAEKALDRLKPYMYMKILFPKLKPDNPLIAS
LYELNASVIVTGFTMDDVLETYTLSAIEELDHTFRETQNWLSPLPPKDYPDLDAILDHIPSTKPPYSRSVVAMFT
DYFNRYCAAIHTAYKPSETRLGAFRRRLAAAVTAYLEMAKRKRRRDGNLHEEEFLWQRSADNLSFPVLMLAEVFT
GILADADVPSTTLHYYHFYSNLFAIVLNDLNSYHRDKDNDDNSLVKLWLKLGVAKDFDDAATKIIDFLNSIMVRM
YVATKELLESNPENEPLRRFVEAVGYTFNGWILVHTTAVERYKLSPFQTVVAPVARGKEEVWLKGETAFGKRCVA
MFEELMGERAEEMNKLYGLDS*
```

```
>TmTC-1_nucleotide_codon_opt
ATGGCGTGTGGGAAAGAACTGCGCATTCCTAAGGAATGGTCGAAGTACCACCATGCAATTGTGAATGAACCGATC
GATCCGGAGCTGTTCTCATTAGACGAATTGTGTGCGTGGTTAAAAGACCTGGATTTATGCAACGACAAATCTGAC
GTGGAAAAATACGTGCTGGCTGTTCGCCCATATCATACGATGCGCTATCAGTTTGTCTTGCTCAGCCGCGAAGAT
GCCTTATGTAAAAAATTATTTAAAATGTGGGCCCACACCGCGATTGCGGTTTTTATTTCAGATGACGTGCTTGAG
TGCCTTTCGAAAACCGAAATGGGCCAGATTTGCAATGCATTCCAGTTGCTGGTGGAGAAAACGCGTCCCCAATTC
TCACAATTCTCTACCATCGCGGAGATGAAACAGTTTCTCCTCTTACAGAAGGTGGATGTGAAGCTGATGCCTCAT
ATTATCTATTTTCTTGATTTCGCAAATAATGTGGCAAAGCTTATCAACGAACAGGGCAATTACAGTAAGGAAGAC
GTCTCGGATTACTGGCGCCGCCTGGTTGTTATGATTTCGCTGTATTACGAAGGTGTAAAGGTGGAGGTGAGCGGA
TCGGGGGGTAGTTACAGCAAGGACATTTGGACGCGCCTGCTCGCTGCTGGTACCATGATTTGGCTGATTGCACAG
GAAGTCATGAGCGGCGCCCTGGGCAAGACTAGCGAGCATGCTAGTTTACTGAATGAGCTGTACTTCTTAGGCACT
TTCTACAGCATGGTCCTGAACGATATCTATAGCCATAATCGCGAAAACGGTTTCCATGGTTCTATTTGTAACCTT
CTGCAGACAATTGTGCGTTCAAAGGATGTTCCCGGTGAATTTGAAGCGGTCCTGAAATGCATTCAGATCCTCAAC
ACCGTTGTTAAACTGATGTACCAGAAGATTGAAACCGCCAAACAGAAAAATCTGGAGAATCGCGTCCTGTGTAAA
TTACTGGACAACATTGGTATGACCACTACCGGATGCTATTTTTTTCATCATTACAGTCATCGTTATGATGACTCG
GCATGGCGCCTTTCGTTGGTGGAAGTTGACGACGATGAACTGGCGGAATGGCAGAAGTGCGACGATGAAGAACAG
CTGAAAGAAGTTATGCCGCTTCTGCTCAACTATTCTCAAAAAGCGAACGAAATCAGCGATACCATCAAACGGGGT
GTTGTAAAAATTAACCGCCACTTTCGGCAATGA
```

```
>TmTC-1_amino_acid
MACGKELRIPKEWSKYHHAIVNEPIDPELFSLDELCAWLKDLDLCNDKSDVEKYVLAVRPYHTMRYQFVLLSRED
ALCKKLFKMWAHTAIAVFISDDVLECLSKTEMGQICNAFQLLVEKTRPQFSQFSTIAEMKQFLLLQKVDVKLMPH
IIYFLDFANNVAKLINEQGNYSKEDVSDYWRRLVVMISLYYEGVKVEVSGSGGSYSKDIWTRLLAAGTMIWLIAQ
EVMSGALGKTSEHASLLNELYFLGTFYSMVLNDIYSHNRENGFHGSICNLLQTIVRSKDVPGEFEAVLKCIQILN
TVVKLMYQKIETAKQKNLENRVLCKLLDNIGMTTTGCYFFHHYSHRYDDSAWRLSLVEVDDDELAEWQKCDDEEQ
LKEVMPLLLNYSQKANEISDTIKRGVVKINRHFRQ*
```

```
>TmTC-2_nucleotide_codon_opt
ATGTCGTGCAGCAAGGAGATCCGCATCCCATCTGACTGGGCAAGTGTCCAGCGTAAAAGTATGCAGGAGGGCCCG
GATGAATCTCTGCTGGATTTTGAAGAACTTGCGAATTGGCTGCGTGAATGCGGCGTAACCGATGATCAATGCAAA
GTACGTAAATATGTCGACGCGACGCGTCCGGTTTTAAGCATTCGCACCATTCTGCGTGTTTTTCCGAACAATACC
CTCTGTCGCATGCATTTCAAATTCCTGACGCAGTACACCATTGGGATTTACATCGACGATGACGTGCTGGAATCC
CATTATCCGCTGGAAGTCCTGAAAGAGATTTGCTACGAGTACGATCAGCTGGACGGTAAGCTGTTGGGCCAATTT
CCGCAGATCCCGAGCCGTAAGGAACTCAAAAATTTTCTCGCACACTTAAAAAATGAAAAAGTGATCTCAACGGTT
ACCTTTCATATGGACTTCGTCAATCGCGTGACTGTTAACATCCTCCAAAATGGTGATTTTTCGGAAGAAGCGGTG
CTGGAATTTCGGCGTCGCCTGAGTAATAATATTCTTCTGTATCTGCAAGGTGTCCAGTATGAAAAACGTGGTGCG
GGGGGGAAACCATCCACCGTCGTGGAGGCCTTATGGAATCGTGTTTTCGGCGGGGCGGCTGTGACTTGGCGGCTG
TTTGGGGAAGTCCCATCAGGTACTCTGGGCAAGATTGGCGAGCATATCACCCTGATCTCTGAACTGTACATCCTT
TCAGCCCTGTACTGCATTGTGATTAATGATCTGTATAGTTACCATCGCGAAATCGATGCGGTGCCGACCGCCGGC
GAGAACTTTGTTCGCGTGATGTTTAACCAAAAAGAGGTAGCCAACCTGACAGTCGCCGCCGGTAAAGTTGCGTCT
ATCTTAAACGAGATTACCAAATACACATATGAGCAGGTTCGTGAGTTTAAAGCCTCATACCCGAACTCGTCTGAA
CTGCATCAACTGTTTGACGATATTGGGTGTGGTACCGTGGGCTGGTACTATTTCCACGAGTGTACCAACCCTCGT
TATAAAGAGAGCAATGTTCGGATTAGTATGAAAGACGTAGACAAAAACGAAATCGAAGACTGGCTGTCGACTAAA
GATAGTTATGGCTGGAATATCGTTAAACAGTTTCTGGCCTCCTACGATGCCAAAGCTAAGCGCCTGAACGATGCG
GTCGCGTGTCGCATCCCGGCCTACGCTGAGACATTCTAA
```

```
>TmTC-2_amino_acid
MSCSKEIRIPSDWASVQRKSMQEGPDESLLDFEELANWLRECGVTDDQCKVRKYVDATRPVLSIRTILRVFPNNT
LCRMHFKFLTQYTIGIYIDDDVLESHYPLEVLKEICYEYDQLDGKLLGQFPQIPSRKELKNFLAHLKNEKVISTV
TFHMDFVNRVTVNILQNGDFSEEAVLEFRRRLSNNILLYLQGVQYEKRGAGGKPSTVVEALWNRVFGGAAVTWRL
FGEVPSGTLGKIGEHITLISELYILSALYCIVINDLYSYHREIDAVPTAGENFVRVMFNQKEVANLTVAAGKVAS
ILNEITKYTYEQVREFKASYPNSSELHQLFDDIGCGTVGWYYFHECTNPRYKESNVRISMKDVDKNEIEDWLSTK
DSYGWNIVKQFLASYDAKAKRLNDAVACRIPAYAETF*
```

```
>XsTC-1_nucleotide_codon_opt
ATGTCCGAAAAAAACGTGGTGCGTATCCCGATGAAGTGGGGTCGTATTGAGCGCGAAATTTTAACACAGAACACC
ATCCCCGAGCTTGTTGATACCAACCGTCTGATTAGTTGGGTTAAGGAATGCAATCTGGCTGATGAAGCACTGGTA
ACTAAATACATGAACGTAATTCGGCCGTATCATTTTTCGCGGCTGGTTTTTCCGATTCTGCCCGACAAAGACGTC
TGTCGTGAAGCGTTTGCAGTGTTTATTGAATTTCTGATCGTATTATATCTGTGTGATGACGAACTCGAGGCAAAA
TGCAACCTGAATGAACTGGAAATCGTCTGTAGCGCGTACGATTTCTTAGATGAGAAGCTGAAACAATGTTTCCCG
CGGATTCCTTCTGTTGAGGAACTGCGTAGCTTTCTGCCTCACGTTAAAAAAGAACGCCTGCTGGCACTGGTTGTG
TCCTTGATCGACAGCGTGTCGCGGGTAGTCTCTCCGTTACTGAAGTATAGCGTGTTTCCGCAGGAATCTGTATTT
GACTTCCGTTGCCGTTTTAGTCACAGCCTGGGCTATAATCTCAAGGGCGTACTCCACGAAAAAAAAATGTTGGGC
GGTGTAAGTGAAAACAAACTGCTGTGGCGTCGTATTTTTGAAGGGGCCCCGATCACGTTTCTGATGTATTTAGAA
ATTAGCTCCCTTTCAGTGACCGGTTCCAACAAACACATCCCGATTGCAACGGAGATGTACATTTTAAGCTCGCTG
TGCTGCATGGTGACTAATGACGTGTACTCTTATCATCGTGAATGCAACGAAGGGCTGAAGGTTGATAATATTATT
AGTCTCTGGCTGCATAATAAGCAGATCGCGTCTGTAGCAGAAGGCGTGAGCCGTATTAGTCGCATTTTGAACTCC
GCGGTCAAATATATGTTAGCCAAAGTAAAATCCATGAAAAGTGAATACCCCAATAATTTCAATGTGCAGACTTTG
ACCGAGTTTATCGCGTTGAGCAGTGTGGGTTGGTTATATATGCACGACCAAGGCGTGCCTCGGTATTCCGATAGT
CCATGGCGTCTGAACCTGGTCGATATTGAAGAATCGGACATTACCTCGTGGTTGGCTGAAAAAGATCCATACGCA
GACGGCGTAATTGACCAATTTAAATACTCCAACCTGGACGCCAAAAAATTTATCGATTCTCTGTGCGAGAAGACC
AGCGTTAGCGAGGAACAGTGGAATGACTGA
```

```
>XsTC-1_amino_acid
MSEKNVVRIPMKWGRIEREILTQNTIPELVDTNRLISWVKECNLADEALVTKYMNVIRPYHFSRLVFPILPDKDV
CREAFAVFIEFLIVLYLCDDELEAKCNLNELEIVCSAYDFLDEKLKQCFPRIPSVEELRSFLPHVKKERLLALVV
SLIDSVSRVVSPLLKYSVFPQESVFDFRCRFSHSLGYNLKGVLHEKKMLGGVSENKLLWRRIFEGAPITFLMYLE
ISSLSVTGSNKHIPIATEMYILSSLCCMVTNDVYSYHRECNEGLKVDNIISLWLHNKQIASVAEGVSRISRILNS
AVKYMLAKVKSMKSEYPNNFNVQTLTEFIALSSVGWLYMHDQGVPRYSDSPWRLNLVDIEESDITSWLAEKDPYA
DGVIDQFKYSNLDAKKFIDSLCEKTSVSEEQWND*
```

Supplementary References

1. Jeon, Y. *et al.* The Draft Genome of an Octocoral, *Dendronephthya gigantea*. *Genome Biol. and Evol.* **11**, 949–953 (2019).
2. Jiang, J. B. *et al.* A hybrid de novo assembly of the sea pansy (*Renilla muelleri*) genome. *Gigascience* **8**, (2019).
3. Kayal, E. *et al.* Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evol. Biol.* **18**, 68 (2018).
4. Hu, M., Zheng, X., Fan, C.-M. & Zheng, Y. Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature* **582**, 534–538 (2020).
5. Zhou, Y. *et al.* The first draft genome of a cold-water coral *Trachythela* sp. (Alcyonacea: Stolonifera: Clavulariidae). *Genome Biol. Evol.* (2020) **13**, evaa265 (2021)..
6. Ledoux, J.-B. *et al.* The Genome Sequence of the Octocoral *Paramuricea clavata* – A Key Resource To Study the Impact of Climate Change in the Mediterranean. *G3: Genes, Genom., Genet.* **10**, 2941–2952 (2020).
7. Simion, P. *et al.* A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. *Curr. Biol.* **27**, 958–967 (2017).
8. Zubìa, E., Spinella, A., Giusto, G. B., Crispino, A. & Cimino, G. A new diterpenoid skeleton from the Mediterranean octocoral *Alcyonium palmatum*: Structure of palmatol. *Tetrahedron Lett.* **35**, 7069–7072 (1994).
9. Manzo, E. *et al.* A new xenicane norditerpene from the Indian marine gorgonian *Acanthogorgia turgida*. *Nat. Prod. Res.* **23**, 1664–1670 (2009).
10. Rivera-García, L., Rivera-Vicéns, R. E., Veglia, A. J. & Schizas, N. V. De novo transcriptome assembly of the digitate morphotype of *Briareum asbestinum* (Octocorallia: Alcyonacea) from the southwest shelf of Puerto Rico. *Mar. Genom.* **47**, 100676 (2019).

11. Coval, S. J., Cross, S., Bernardinelli, G. & Jefford, C. W. Brianthein V, a New Cytotoxic and Antiviral Diterpene Isolated from *Briareum asbestinum*. *J. Nat. Prod.* **51**, 981–984 (1988).
12. Dookran, R. *et al.* Briarane and asbestinane diterpenes from *Briareum asbestinum*. *Tetrahedron* **50**, 1983–1992 (1994).
13. DeLeo, D. M., Glazier, A., Herrera, S., Barkman, A. & Cordes, E. E. Transcriptomic Responses of Deep-Sea Corals Experimentally Exposed to Crude Oil and Dispersant. *Front. Mar. Sci.* **8**, (2021).
14. Khalturin, K. *et al.* Medusozoan genomes inform the evolution of the jellyfish body plan. *Nat. Ecol. Evol.* **3**, 811–822 (2019).
15. Iguchi, K., Fukaya, T., Yasumoto, A. & Watanabe, K. New Marine Sesquiterpenoids and Diterpenoids from the Okinawan Soft Coral *Clavularia koellikeri*. *J. Nat. Prod.* **67**, 577–583 (2004).
16. Iwashima, M., Matsumoto, Y., Takahashi, H. & Iguchi, K. New Marine Cembrane-type Diterpenoids from the Okinawan Soft Coral *Clavularia koellikeri*. *J. Nat. Prod.* **63**, 1647–1652 (2000).
17. Pratlong, M. *et al.* The red coral (Corallium rubrum) transcriptome: a new resource for population genetics and local adaptation studies. *Mol. Ecol. Resources* **15**, 1205–1215 (2015).
18. Schwartz, R. E., Scheuer, P. J., Zabel, V. & Watson, W. H. The coraxeniolides, constituents of pink coral, *Corallium* sp. *Tetrahedron* **37**, 2725–2733 (1981).
19. Ma, A. *et al.* Dendronpholides A−R, Cembranoid Diterpenes from the Chinese Soft Coral *Dendronephthya* sp. *J. Nat. Prod.* **71**, 1152–1160 (2008).
20. Elkhayat, E. S., Ibrahim, S. R. M., Fouad, M. A. & Mohamed, G. A. Dendronephthols A–C, new sesquiterpenoids from the Red Sea soft coral *Dendronephthya* sp. *Tetrahedron* **70**, 3822–3825 (2014).

21. Ryu, T. *et al.* First transcriptome assembly of the temperate azooxanthellate octocoral *Eleutherobia rubra*. *Mar. Genom.* **48**, 100682 (2019).
22. Lindel, T. *et al.* Eleutherobin, a New Cytotoxin that Mimics Paclitaxel (Taxol) by Stabilizing Microtubules. *J. Am. Chem. Soc.* **119**, 8744–8745 (1997).
23. Fuess, L. E., Mann, W. T., Jinks, L. R., Brinkhuis, V. & Mydlarz, L. D. Transcriptional analyses provide new insight into the late-stage immune response of a diseased Caribbean coral. *R. Soc. Open Sci.* **5**, 172062.
24. Look, S. A., Fenical, W., Zheng, Q. T. & Clardy, J. Calyculones, new cubitane diterpenoids from the Caribbean gorgonian octocoral *Eunicea calyculata*. *J. Org. Chem.* **49**, 1417–1423 (1984).
25. Romiguier, J. *et al.* Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).
26. Mancini, I., Guella, G., Zibrowius, H. & Pietra, F. Configuration, Conformation, and Reactivity of Highly Functionalized Eunicellane Diterpenes Isolated from the Gorgonians *Eunicella cavolinii* and *Eunicella singularis* from Marseille. *Helv. Chim. Acta* **83**, 1561–1575 (2000).
27. Tsitsimpikou, C., Vagias, C. & Roussis, V. A New Epoxymuurolane Metabolite from the Soft Coral *Gorgonia Ventalina*. *Nat. Prod. Lett.* **14**, 17–23 (1999).
28. Hongo, Y., Yasuda, N. & NagaI, S. Identification of Genes for Synthesis of the Blue Pigment, Biliverdin IXα, in the Blue Coral Heliopora coerulea. *The Biological Bulletin* (2017) doi:10.1086/692661.
29. Tanaka, J., Ogawa, N., Liang, J., Higa, T. & Gravalos, D. G. Helioporins: bioactive diterpenes from the blue coral *Heliopora coerulea*. *Tetrahedron* **49**, 811–822 (1993).
30. DeLeo, D. M. *et al.* Gene expression profiling reveals deep-sea coral response to the Deepwater Horizon oil spill. *Mol. Ecol.* **27**, 4066–4077 (2018).

31. Reddy, N. S., Reed, J. K., Longley, R. E. & Wright, A. E. Two New Cytotoxic Linderazulenes from a Deep-Sea Gorgonian of the Genus *Paramuricea*. *J. Nat. Prod.* **68**, 248–250 (2005).
32. Vargas, S., Zimmer, T., Conci, N., Lehmann, M. & Wörheide, G. Resilience to climate change in an octocoral involves the transcriptional decoupling of the calcification and stress response toolkits. *bioRxiv* 2020.07.15.202069 (2020) doi:10.1101/2020.07.15.202069.
33. Barsby, T. & Kubanek, J. Isolation and Structure Elucidation of Feeding Deterrent Diterpenoids from the Sea Pansy, *Renilla reniformis*. *J. Nat. Prod.* **68**, 511–516 (2005).
34. Liu, Z. *et al.* Capnosane-type cembranoids from the soft coral *Sarcophyton trocheliophorum* with antibacterial effects. *Tetrahedron* **70**, 8703–8713 (2014).
35. Conci, N., Wörheide, G. & Vargas, S. New Non-Bilaterian Transcriptomes Provide Novel Insights into the Evolution of Coral Skeletomes. *Genome Biol. Evol.* **11**, 3068–3081 (2019).
36. Ahmed, A. F. *et al.* Oxygenated Cembranoids from a Formosan Soft Coral *Sinularia gibberosa*. *J. Nat. Prod.* **71**, 179–185 (2008).
37. Iguchi, K., Mori, K., Suzuki, M., Takahashi, H. & Yamada, Y. New Marine Furanosesquiterpenoids, Tubipofuran and 15-Acetoxytubipofuran from the Stolonifer *Tubipora musica* Linnaeus. *Chem. Lett.* **15**, 1789–1792 (1986).
38. Cheng, Y.-B., Jang, J.-Y., Khalil, A. T., Kuo, Y.-H. & Shen, Y.-C. Xenicane-Type Diterpenes with Cytotoxicity from *Xenia florida*. *J. Nat. Prod.* **69**, 675–678 (2006).
39. Aaron, J. A., Lin, X., Cane, D. E. & Christianson, D. W. Structure of Epi-Isozizaene Synthase from *Streptomyces coelicolor* A3(2), a Platform for New Terpenoid Cyclization Templates. *Biochemistry* **49**, 1787–1797 (2010).
40. Köksal, M., Chou, W. K. W., Cane, D. E. & Christianson, D. W. Structure of 2-Methylisoborneol Synthase from *Streptomyces coelicolor* and Implications for the

Cyclization of a Noncanonical C-Methylated Monoterpenoid Substrate. *Biochemistry* **51**, 3011–3020 (2012).

41. Vedula, L. S. *et al.* Exploring biosynthetic diversity with trichodiene synthase. *Arch. Biochem. Biophys.* **466**, 260–266 (2007).
42. Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes Genom. Genet.* **10**, 1361–1374 (2020).
43. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
44. Nasuda, M., Ohmori, M., Ohyama, K. & Fujimoto, Y. Revision of the Stereochemistry of Elisabethatriene, a Putative Biosynthetic Intermediate of Pseudopterosins. *Chem. Pharm. Bulletin* **60**, 681–685 (2012).
45. Coleman, A. C. & Kerr, R. G. Radioactivity-Guided Isolation and Characterization of the Bicyclic Pseudopterosin Diterpene Cyclase Product from *Pseudopterogorgia elisabethae*. *Tetrahedron* **56**, 9569–9574 (2000).
46. Lazerwith, S. E., Johnson, T. W. & Corey, E. J. Syntheses and Stereochemical Revision of Pseudopterosin G−J Aglycon and Helioporin E. *Org. Lett.* **2**, 2389–2392 (2000).
47. Chen, B.-W. *et al.* Klysimplexins I–T, eunicellin-based diterpenoids from the cultured soft coral Klyxum simplex. *Org. Biomol. Chem.* **9**, 834–844 (2011).
48. Rinkel, J., Lauterbach, L., Rabe, P. & Dickschat, J. S. Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria albata*. *Angew. Chem. Int. Ed.* **57**, 3238–3241 (2018).
49. Huber, T., Weisheit, L. & Magauer, T. Synthesis of Xenia diterpenoids and related metabolites isolated from marine organisms. *Beilstein J. Org. Chem.* **11**, 2521–2539 (2015).

50. Angulo-Preckler, C. *et al.* Gersemiols A–C and Eunicellol A, Diterpenoids from the Arctic Soft Coral *Gersemia fruticosa*. *J. Nat. Prod.* **79**, 1132–1136 (2016).
51. Meyers, A. I. & Bienz, S. Asymmetric total synthesis of (+)-Δ–9(12)-capnellene. *J. Org. Chem.* **55**, 791–798 (1990).