http://dx.doi.org/10.5935/0103-5053.20130240

J. Braz. Chem. Soc., Vol. 24, No. 12, 1926-1932, 2013.
Printed in Brazil - ©2013 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Article

Cyclodipeptides from Metagenomic Library of a Japanese Marine Sponge

Rui He,[a,b,c] Bochu Wang,*[,b] Toshiyuki Wakimoto,[c] Manyuan Wang,[a] Liancai Zhu[b] and Ikuro Abe*[,c]

[a]*School of Traditional Chinese Medicine, Capital University of Medical Sciences, No. 10 Xitoutiao, You An Men, 100069 Beijing, P. R. China*

[b]*Bioengineering College, Chongqing University, No. 174, Shanpingba Main Street, 400030 Chongqing, P. R. China*

[c]*Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-0033 Tokyo, Japan*

A metagenômica independente de cultura é um meio atraente e promissor para explorar pequenas moléculas bioativas únicas de esponjas marinhas que abrigam micro-organismos simbiontes não cultiváveis. Foi realizada uma triagem funcional da biblioteca metagenômica da esponja marinha japonesa *Discodermia calyx*. O fracionamento bio-guiado do extrato cultivado em placas do clone bactericida pDC113 produziu onze ciclodipeptídeos: Ciclo(L-Thr-L-Leu) (**1**), Ciclo(L-Val-D-Pro) (**2**), Ciclo(L-Ile-D-Pro) (**3**), Ciclo(L-Leu-L-Pro) (**4**), Ciclo(L-Val-L-Leu) (**5**), Ciclo(L-Leu-L-Ile) (**6**), Ciclo(L-Leu-L-Leu) (**7**), Ciclo(L-Phe-L-Tyr) (**8**), Ciclo(L-Trp-L-Pro) (**9**), Ciclo(L-Val-L-Trp) (**10**) e Ciclo(L-Ile-L-Trp) (**11**). Eles são os primeiros ciclodipeptídeos isolados a partir de uma biblioteca metagenômica. A análise sequencial indicou que os ciclodipeptídeos isolados não foram sintetizados por peptídeo sintetases não ribossomais e não havia indícios significativos de sintetases ciclodipeptídicas.

Culture-independent metagenomics is an attractive and promising approach to explore unique bioactive small molecules from marine sponges harboring uncultured symbiotic microbes. Therefore, we conducted functional screening of the metagenomic library constructed from the Japanese marine sponge *Discodermia calyx*. Bioassay-guided fractionation of plate culture extract of antibacterial clone pDC113 afforded eleven cyclodipeptides: Cyclo(L-Thr-L-Leu) (**1**), Cyclo(L-Val-D-Pro) (**2**), Cyclo(L-Ile-D-Pro) (**3**), Cyclo(L-Leu-L-Pro) (**4**), Cyclo(L-Val-L-Leu) (**5**), Cyclo(L-Leu-L-Ile) (**6**), Cyclo(L-Leu-L-Leu) (**7**), Cyclo(L-Phe-L-Tyr) (**8**), Cyclo(L-Trp-L-Pro) (**9**), Cyclo(L-Val-L-Trp) (**10**) and Cyclo(L-Ile-L-Trp) (**11**). To the best of our knowledge, these are first cyclodepeptides isolated from metagenomic library. Sequence analysis suggested that isolated cyclodipeptides were not synthesized by nonribosomal peptide synthetases and there was no significant indication of cyclodipeptide synthetases.

Keywords: cyclodipeptides, diketopiperazines, metagenomics, marine sponge

Introduction

Marine sponges are rich and important sources for a broad range of secondary metabolites. Many of these biologically active compounds could be produced by symbiotic bacteria.[1] However, the vast majority of the sponge microbial community remains uncultured on laboratory conditions.[2] Functional metagenomics, exploring uncultured environmental microorganisms by extracting genomic DNA directly from samples without any culture or isolation steps, has been proven to be a practical approach to search for unique bioactive small molecules from interesting resources, such as soil[3,4] and marine sponges.[5] Therefore, searching for bioactive small molecular compounds from metagenomic library of marine sponges is promising and attractive.

The marine sponge *Discodermia calyx* (*D. calyx*), containing calyculins[6] as the major cytotoxic compounds and calyxamides[7] as the cytotoxic cyclic peptides, would be an attractive source of metagenomic library for functional

*e-mail: wangbc2000@126.com, abei@mol.f.u-tokyo.ac.jp

screening of small molecules. Recently, four porphyrin pigments[8] and three antibacterial β-hydroxyl fatty acids[9] were identified from positive clones by functional screening from metagenomic library of this marine sponge. This implicated that the metagenomic library of this sponge would be worthy of further study. Therefore, we conducted the antibacterial screening of the metagenomic library of the marine sponge, *D. calyx*, which resulted in the detection of eleven cyclodipeptides (CDPs) from plate culture of active clone pDC113.

Results and Discussion

The metagenomic library of the marine sponge *D. calyx*, containing 2.5×10^5 clones harboring ca. 40 kb insert DNA, was constructed and screened for antibacterial activity using the two-layer overlay method. An active clone, pDC113, was detected by the clear inhibition zone against *Bacillus cereus* (*B. cereus*) on Luria-Bertani (LB) agar medium.

Bioassay-guided fractionation by Sephadex LH-20 column chromatography yielded two active fractions obtained from the EtOH extract of 50 plate (∅ 150 mm, 100 mL $plate^{-1}$) cultures of pDC113, along with a chloramphenicol containing active fraction. Both active fractions were further purified by reverse phase high performance liquid chromatography with diode array detector (RP-HPLC-DAD) to afford seven compounds (**1-7**) from F8 (Figure 1) and four compounds (**8-11**) from F14 (Figure 2). All other HPLC eluting fractions f1-f4 except for compounds **1-11** were collected and fractionated by time (0-10 min, 10-20 min, 20-30 min, 30 min) and showed no antibacterial activity against *B. cereus*. Therefore, antibacterial activities of both F8 and F14 can be ascribed to the isolated compounds. Besides, the plate culture of the negative control (strain EPI300 carrying the pCC1FOS fosmid vector) was also fractionated and the corresponding fractions showed no antibacterial activity, suggesting that active compounds might be specific to clone pDC113. In addition, comparison of the production of cyclodipeptides **1-7** from clone pDC113 and negative control showed that cyclodipeptides **1-7** were only produced by clone pDC113 (Supplementary Information Figure S1). This indicated that cyclodipeptides **1-7** were clone-specific.

The identification of CDPs **1-11** (Figure 3) was based on the analysis of nuclear magnetic resonance (^{1}H NMR, ^{13}C NMR, ^{1}H-^{1}H COSY, HMQC, HMBC of compound **4** and ^{1}H NMR, ^{13}C NMR, and ^{1}H-^{1}H COSY of others) spectra (Figures S2-S37) and electrospray ionization mass spectrometry (ESI-MS) data (Table 1 and Figure S38). The dipeptide structures were evident from the observation of characteristic ^{13}C signals of two amide carbonyl groups (CONH, δ_C 165-172) and ^{1}H signals of two α-protons (δ_H 3.5-4.2). Proline as a common counterpart of compounds **2-4** and **9** was easily deduced from the presence of broad methylene multiplets (δ_H 1.7-3.7). The NMR spectra clearly showed that valine, isoleucine, leucine and tryptophan were another counterpart in compounds **2-4** and **9**, respectively. The presence of threonine, tyrosine and phenylalanine residues in other compounds was also clear based on the NMR data. To verify the diketopiperadine ring (DKP, Figure 4) formation, the HMBC spectrum of the major compound **4** (4.24 mg) (Table 1) was measured in $CDCl_3$ (Figure 5 and Figure S8). The HMBC signals H-3 to C-1, H-8 NH to C-6 and C-7, H-6 to C-1 were strong evidences of the cyclic system of compound **4**. The HMBC correlations of other CDPs were not detected due to the scarcity of materials. However, the NMR data in accordance with the MS data (Table 1) can elucidate the structures of **1-11** as Cyclo(Thr-Leu) (**1**),[10] Cyclo(Val-Pro) (**2**),[11] Cyclo(Ile-Pro) (**3**),[12] Cyclo(Leu-Pro) (**4**),[13] Cyclo(Val-Leu) (**5**),[11] Cyclo(Leu-Ile) (**6**),[10] Cyclo(Leu-Leu) (**7**),[10] Cyclo(Phe-Try) (**8**),[14] Cyclo(Trp-Pro) (**9**),[15] Cyclo(Val-Trp) (**10**) and Cyclo(Ile-Trp) (**11**).

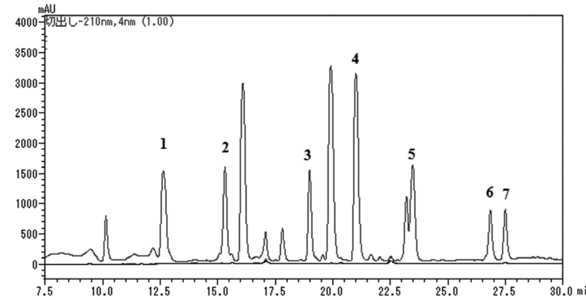


Figure 1. Compounds (**1-7**) from active F8 of LH-20 by semi-preparative RP-HPLC-DAD. HPLC Conditions: linear gradient with a mixture of H_2O and MeCN, both containing 0.05% TFA. 0-20 min, 5-35% MeCN; 20-28 min, 35-56% MeCN; 28-29 min, 56-100% MeCN; and 29-32 min, 100% MeCN. Column: Cosmosil $5C_{18}$-PAQ-Waters, 10 × 250 mm. 2.5 mL min^{-1}. DAD profiles were measured with a Shimadzu HPLC System: LC-20AD and SPD-20A Prominence Diode Array Detector.

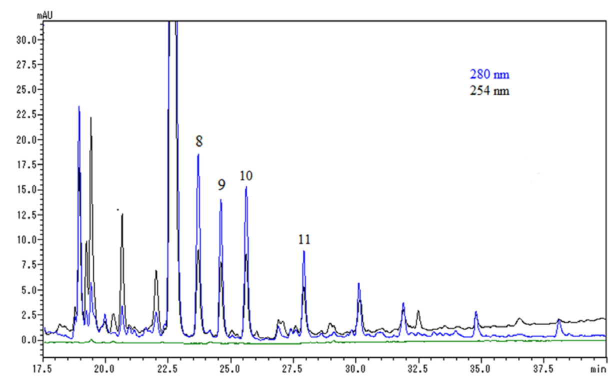


Figure 2. Compounds (**8-11**) from active F14 of LH-20 by semi-preparative RP-HPLC-DAD. HPLC conditions were the same as that of F8 in Figure 1.

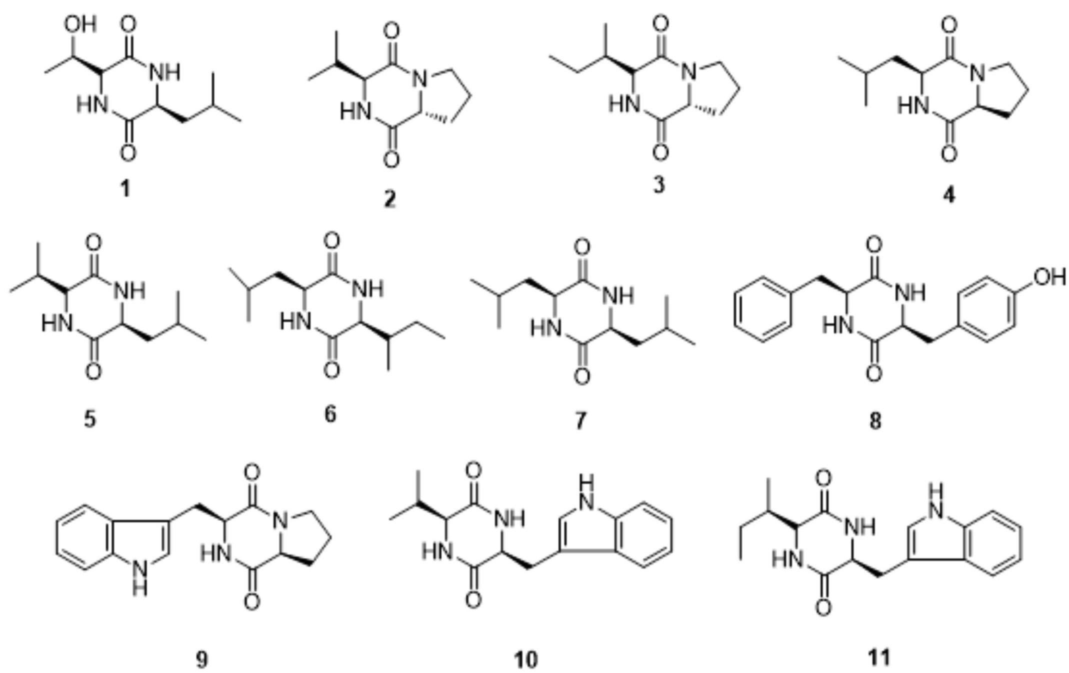


Figure 3. Structures of isolated cyclodipeptides (**1**-**11**) from metagenomic library of the marine sponge *D. calyx*: Cyclo(L-Thr-L-Leu) (**1**), Cyclo(L-Val-D-Pro) (**2**), Cyclo(L-Ile-D-Pro) (**3**), Cyclo(L-Leu-L-Pro) (**4**), Cyclo(L-Val-L-Leu) (**5**), Cyclo(L-Leu-L-Ile) (**6**), Cyclo(L-Leu-L-Leu) (**7**), Cyclo(L-Phe-L-Tyr) (**8**), Cyclo(L-Trp-L-Pro) (**9**), Cyclo(L-Val-L-Trp) (**10**) and Cyclo(L-Ile-L-Trp) (**11**).

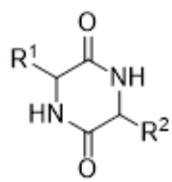
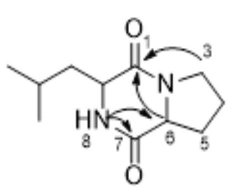


Figure 4. Structure of diketopiperazines.



4

Figure 5. Key HMBC correlations, evidence of the DKP ring formation of compound **4** in $CDCl_3$.

The configurations of CDPs **2**, **3**, **4**, and **5** were determined by chiral-phase gas chromatography (GC) analysis of amino acids. Retention times (min) of standard amino acids were as follows: L-Leu (10.0), D-Leu (11.3), L-Val (6.2), D-Val (11.8), L-Ile (8.3), D-Ile (8.9), L-Pro (8.8), D-Pro (9.2). Thus, after hydrolysis, the presence of L-Val (6.2) and D-Pro (9.2) in compound **2**, L-Ile (8.3) and D-Pro (9.3) in compound **3**, L-Leu (10.0) and L-Pro (8.9) in compound **4**, L-Val (6.1) and L-Leu (10.0) in compound **5** were confirmed. Stereochemistry of other compounds was suggested by optical rotation values (Table 1) comparing with reported data: Cyclo(L-Thr-L-Leu) (**1**),[10] Cyclo(L-Leu-L-Ile) (**6**),[10] Cyclo(L-Leu-L-Leu) (**7**),[10] Cyclo(L-Phe-L-Tyr) (**8**),[16] Cyclo(L-Trp-L-Pro) (**9**),[17] Cyclo(L-Val-L-Trp) (**10**)[18] and Cyclo(L-Ile-L-Trp) (**11**).[19]

CDPs occur in numerous natural products and are often found alone or embedded in larger, more complex architectures in a variety of natural products from fungi, bacteria, marine sponges, plants, and mammals.[20] Due to their significant and diverse biological activities, such as antimicrobial,[21,12] antitumor,[21,22] antifouling,[13] antiprion,[23] antioxidant,[10] Quorum sensing signals,[24] immunosuppressive and anti-inflammatory activities, there has been an increasing interest in natural CDPs in recent years. Most CDPs isolated from natural sources were in the LL form. Interestingly, D-Proline existed in compound **2** and **4**. There were also some reports of DD and DL enantiomers as natural products[12,13] and showed very strong activity against the pathogen *Vibrio anguillarum* (MIC, 0.03-0.14 mg mL^{-1}).[12] There was no consistency in the biological activity of the LL-enantiomers, which depended on the assay systems.[13,25,26]

CDPs are catalyzed by two kinds of reported enzymes: nonribosomal peptide synthetase (NRPS) and small cyclodipeptide syntheases (CDPSs), a newly defined family of class-I aminoacyl-tRNA synthetase-like enzymes.[27]

Maiya and Li reported a bimodular NRPS enzyme FtmPS that used L-tryptophan and L-proline as substrates to synthesize cyclodipeptide brevianamide F from the fumitremorgin gene cluster of *Aspergillus fumigatus.*[28]

Table 1. Yields, optical rotation and ESI MS data of cyclodipeptides

Compound	Yields / (mg (50 plates)$^{-1}$)	$[\alpha]_D^{25}$ / degree (c, solvent)	Reported $[\alpha]_D^{25}$/ degree (c, solvent)	m/z [M]$^+$
Cyclo(L-Thr-L-Leu) (**1**)	1.04	–53.7 (0.06, MeOH)	–56.5 (0.07, MeOH)[10]	215.1
Cyclo(L-Val-D-Pro) (**2**)	0.95	+34.8 (0.08, EtOH)		197.0
Cyclo(L-Ile-D-Pro) (**3**)	1.03	+71 (0.08, EtOH)		211.1
Cyclo(L-Leu-L-Pro) (**4**)	4.24	–88 (0.32, EtOH)	–133 (0.3, EtOH)[11]	211.0
Cyclo(L-Val-L-Leu) (**5**)	1.88	–53 (0.16, MeOH)	–71.2 (0.10, MeOH)[10]	213.0
Cyclo(L-Leu-L-Ile) (**6**)	0.64	–52 (0.05, MeOH)	–56.6 (0.10, MeOH)[10]	227.0
Cyclo(L-Leu-L-Leu) (**7**)	0.68	–45 (0.03, MeOH)	–46.4 (0.10, MeOH)[10]	227.1
Cyclo(L-Phe-L-Tyr) (**8**)	1.68	–81 (0.10, DMSO)	–117.6 (0.3, DMSO)[16]	311.1
Cyclo(L-Trp-L-Pro) (**9**)	1.21	–48 (0.06, MeOH)	–64 (0.69, MeOH)[17]	284.0
Cyclo(L-Val-L-Trp) (**10**)	0.51	–59 (0.04, MeOH)	–65 (0.11, MeOH)[18]	286.1
Cyclo(L-Ile-L-Trp) (**11**)	0.14	–98 (0.01, EtOH)	+82 (0.5, EtOH)[19,a]	300.1

[a]Optical rotation of cyclo(D-Ile-L-Trp) measured at 20 ºC

Ding *et al.* also identified a bimodular NRPS named notE (2241 aa) based on the whole genome sequence of a marine-derived *Aspergillus* sp.[29] However, sequence analysis of clone pDC113 showed that there was no Adenylation (A) domain (required in an NRPS module), through blast research or NRPS predictor of 42 open reading frames (ORFs) encoded in 43.32 kb (Table 2). This indicated that the isolated CDPs were not synthesized by NRPS.

Subsequently, we compared the 42 ORFs to reported CDPSs to check whether there were any ORFs sharing homology with CDPSs. CDPSs used aminoacyl-tRNAs as substrates to synthesize the two peptide bonds of various CDPs.[30] Until now, there were nine CDPSs using L amino acids reported.[31] However, only three of them (AlbC, Rv2275 and YvmC-Blic) have been fully elucidated including the crystallographic structures. AlbC (239 aa) was firstly reported to form cyclo(L-Phe-L-Leu) in the biosynthesis of albonoursin from *Streptomyces noursei*[32] through ping-pong catalytic mechanism.[30] Rv2275 (289 aa) synthesized Cyclo(L-Tyr-L-Tyr) in the first step of biosynthesis of mycocyclosin.[33] YvmC (249 aa) formed LL cyclodileucine in the biosynthetic pathway of pilcherrimin.[34] Interestingly, the CDPSs shared only moderate sequence similarity (19-27% sequence identity). Sequence alignment of nine reported CDPSs showed only seven conserved residues at positions Gly35, Ser37, Gly79, Tyr128, Tyr178, Glu182 and Tyr202 (AlbC numbering) and shared only three catalytic residues (Ser37, Tyr178 and Glu182).[30,31] Therefore, we aligned the 42 ORFs in clone pDC113 with reported CDPSs to check whether any ORF contained the nine conserved regions or the three catalytic residues (Ser37, Tyr178 and Glu182). Unfortunately, there was no potential ORF candidate either sharing all conserved regions or the three catalytic residues of reported CDPSs. Through sequencing alignments it was difficult to discover significant indications of potential candidate ORFs related to CDPSs involving in the biosynthesis of isolated cyclodipeptides.

The isolated CDPs were not biosynthesized by NRPS and there were no obvious potential CDPSs candidates through sequence analysis of the insert DNA of clone pDC113. It had high possibility that they were biosynthesized by new enzymes encoded by new genes. This result favors the most attractive theoretical potential of metagenomics – to be powerful for the finding of new genes with enhanced chances. The cyclodipeptides producing clone pDC113 were detected and the insert DNA of pDC113 was sequenced and analyzed. Although there were no indications of the potential CDPSs candidates, there is high possibility to discover the functional genes from 42 ORFs encoded in 43.32 kb by subcloning and mutation. The isolated CDPs **1-11** were combination of L and D amino acids residues. Identification of the functional genes involving in the biosynthesis of isolated CDPs is currently under investigation.

Conclusions

Eleven CDPs (**1-11**) were isolated by bioassay-guided fractionation from LB agar plate culture of positive clone pDC113 screened from metagenomic library of marine sponge *D. calyx*. To the best of our knowledge this is the first report of CDPs from metagenomic library. Based on the protein BLAST of the sequence, the biosynthesis of the isolated CDPs, some of which containing D-proline residue, was not through NRPS. Sequencing alignments of 42 ORFs to reported CDPSs indicated that there was no significant potential ORF candidate related to CDPSs. It

Table 2. The enzyme homology analysis of pDC113 (18 ORFs encoded in 18.507 kb of 43.32 kb)

ORF	Size / aa	Enzyme	E value	Identity (100%)
1	131	zinc finger protein [*Syntrophus aciditrophicus* SB]	9e-22	43
		conserved hypothetical protein [*Stigmatella aurantiaca* DW4/3-1]	6e-21	38
		putative metal-binding protein [*Eggerthella* sp. YY7918]	5e-17	41
2-24[a]				
25	302	hypothetical protein CHU_0606 [*Cytophaga hutchinsonii* ATCC 33406]	3e-59	39
26	392	ring-hydroxylating dioxygenase, large terminal subunit [*gamma proteobacterium* HIMB55]	1e-107	41
		phenylpropionate dioxygenase and related ring-hydroxylating dioxygenases, large terminal subunit [uncultured *gamma proteobacterium* HF0010_05D02]	2e-103	40
		Rieske (2Fe-2S) domain-containing protein [*Parvibaculum lavamentivorans* DS-1]	3e-86	38
27	348	5,10-methylenetetrahydromethanopterin reductase [*Phenylobacterium zucineum* HLK1]	8e-143	60
		Luciferase-like, subgroup [*Frankia* sp. CN3]	3e-120	54
		F420-dependent oxidoreductase [*Frankia* sp. EuI1c]	6e-119	55
28	232	sensory box histidine kinase/response regulator [*Synechococcus* sp. JA-2-3B'a(2-13)]	1e-41	42
		PAS fold family [*Microcoleus chthonoplastes* PCC 7420]	2e-41	40
		unnamed protein product [*Desulfobacterium autotrophicum* HRM2]	4e-41	36
29	573	PAS/PAC sensor hybrid histidine kinase [*Opitutus terrae* PB90-1]	1e-125	54
		multi-sensor hybrid histidine kinase [*Chthoniobacter flavus* Ellin428]	1e-122	54
		unnamed protein product [*Desulfatibacillum alkenivorans* AK-01]	2e-113	51
30	97	heme NO binding domain-containing protein [*Nostoc punctiforme* PCC 73102]	3e-36	63
		unnamed protein product [*Acaryochloris marina* MBIC11017]	7e-31	55
		Chain A, Crystal Structure Of An H-Nox Protein From Nostoc Sp. Pcc 7120, L66wL67W DOUBLE MUTANT	8e-30	53
31	278	transposase, IS4 family protein [*Roseiflexus* sp. RS-1]	5e-67	45
32	59	transposase, IS4 family protein [*Roseiflexus* sp. RS-1]	1e-14	59
33	448	transposase [*marine psychrotrophic bacterium* Mst37]	3e-79	35
		IS element transposase [*Pseudoalteromonas haloplanktis* ANT/505]	9e-66	31
		putative transposase [uncultured bacterium]	8e-58	38
34	86	heme NO binding domain-containing protein [*Nostoc punctiforme* PCC 73102]	9e-28	58
		unnamed protein product [*Cyanothece* sp. PCC 7425]	3e-25	61
		unnamed protein product [*Nostoc* sp. PCC 7120]	2e-24	59
35	359	lipopolysaccharide heptosyltransferase II [*Flexistipes sinusarabici* DSM 4947]	1e-63	32
		ADP-heptose:LPS heptosyltransferase II [*Fusobacterium* sp. 3_1_5R]	4e-63	31
		glycosyl transferase family protein [*Denitrovibrio acetiphilus* DSM 12809]	1e-61	35
36	366	putative glycosyl transferase [*Candidatus Cloacamonas acidaminovorans*]	5e-81	40
		glycosyltransferase [*Leptospira borgpetersenii serovar Hardjo-bovis* JB197]	9e-65	34
37	346	glycosyl transferase family 9 [*Caldithrix abyssi* DSM 13497]	3e-63	36
		family 9 glycosyl transferase [*Chloroherpeton thalassium* ATCC 35110]	7e-58	37
38	402	CDP-glycerol:poly(glycerophosphate)glycerophosph otransferase [*Caldithrix abyssi* DSM 13497]	1e-120	49
		hypothetical protein CLOAM0422 [*Candidatus Cloacamonas acidaminovorans*]	3e-115	47
39	75	transposase, truncation [*Synechococcus* sp. JA-3-3Ab]	1e-17	59
40	288	unnamed protein product [*Meiothermus ruber* DSM 1279]	4e-85	49
		unnamed protein product [*Truepera radiovictrix* DSM 17093]	1e-79	46
		IS605 OrfB family transposase [*Nitrosococcus watsonii* C-113]	1e-79	47
41	486	unnamed protein product [*Geobacter metallireducens* GS-15]	3e-152	51
		unnamed protein product [*Pelobacter propionicus* DSM 2379]	4e-151	52
		D-glycero-D-mannoheptose-7-phosphate inase and D-glycero-D-mannoheptose-1-phosphate adenylyltransferase [*Geobacter sulfurreducens* KN400]	9e-149	52
42	85	membrane-bound nitrate reductase large subunit [uncultured bacterium]	0.017	57

[a]ORFs 2-24 encoded in 24.813 kb were reported as ORFs 1-23 involved in the biosynthesis of fatty acids.[9]

was highly possible that they were biosynthesized by novel enzymes encoded by interesting genes. This result will surely be helpful for discovering new genes by attractive metagenomics. Subcloning and mutation are under investigation to search for the functional genes.

Experimental

General experimental procedures

^{1}H and ^{13}C NMR spectra were recorded on a JEOL ECX-500 spectrometer in DMSO-d_6, CD_3OD and $CDCl_3$. ^{1}H and ^{13}C NMR chemical shifts were reported in parts per million and referenced to solvent peaks (ppm): δ_H 2.50 and δ_C 39.50 for DMSO-d_6; δ_H 3.31 and δ_C 49.00 for CD_3OD; δ_H 7.26 and δ_C 77.16 for $CDCl_3$. Optical rotations were measured on a JASCO DIP-1000 digital polarimeter.

Construction and screening of the metagenomics library

The marine sponge *D. calyx* was collected by hand using SCUBA from a depth of approximately 10 m off Shikine-jima Islands in Japan. Samples were kept frozen at –80 °C until use. The total sponge DNA was extracted and purified as previously described.[8] The library was constructed according to the manufacturer's protocol. In brief, the purified DNA larger than 35 kb was blunt-ended with an End-It DNA End-Repair Kit (Epicentre, Madison, WI), and ligated into the pCC1FOS fosmid vector (Epicentre). Then, this vector was packaged with a MaxPlax Lambda Packaging Extract (Epicentre) and transfected into *Escherichia coli* EPI300-T1R (Epicentre). Mixtures were plated on the LB agar containing 12.5 µg mL^{-1} of chloramphenicol and grown cells were collected. Two-layer top agar diffusion method[35] with *B. cereus* as test bacterium was used for screening the antibacterial clones by observation of inhibition zones.

Production and isolation of CDPs by bioassay-guided separation

The active clone was cultured on LB agar plates (∅ 150 mm) supplemented with chloramphenicol (12.5 µg mL^{-1}) at 30 °C for 3 days. The LB agar containing cells was extracted with EtOH overnight. The resulting mixture solution of EtOH and water was filtered and evaporated *in vacuo* to remove the EtOH. The resulting water solution (about 500 mL) was extracted with same volume of ethyl acetate three times. The active ethyl acetate extract (1.0 g) was subsequently separated by Sephadex LH-20 gel filtration chromatography eluting with MeOH. Except the chloramphenicol containing fraction, two active fractions F8 and F14 were subjected to semi-preparative RP-HPLC-DAD separation (linear gradient with a mixture of H_2O and MeCN, both containing 0.05% TFA. 0-20 min, 5-35% MeCN; 20-28 min, 35-56% MeCN; 28-29 min, 56-100% MeCN; and 29-32 min, 100% MeCN. Column: Cosmosil 5C_{18}-PAQ-Waters, 10 × 250 mm, 2.5 mL min^{-1}. DAD profiles were measured with a Shimadzu HPLC System: LC-20AD and SPD-20A Prominence Diode Array Detector.). Eleven CDPs were finally isolated.

Antibacterial assay

Standardized agar disc diffusion test using *B. cereus* as a test bacterium was used for bioassay guided separation. LB agar plates (∅ 90 mm) containing overnight cultured *B. cereus* were freshly prepared and divided into four or six quadrants, with a disc paper (6 mm, Tokyo Roshi Kaisha, Ltd) carrying samples (2 mg $paper^{-1}$ for crude extract or 100 µg $paper^{-1}$ for fractions) or positive control chloramphenicol (2 µg $paper^{-1}$) on each quadrant. The plates were incubated at 37 ºC for 12-16 h. Inhibition zone around the paper was observed as indication of anti-*B. cereus* activity.

Determination of the configurations of CDPs by chiral-phase GC

Amino acid analysis of CDPs was performed on a Shimadzu GC-MS-QP 2010 plus gas chromatograph mass spectrometer (GC-MS).[7] In brief, the compound (100 µg) was hydrolyzed with 6 mol L^{-1} HCl (500 µL) at 110 °C for 24 h, treated with 5-10% HCl/MeOH (500 µL) at 100 °C for 30 min and then dried under nitrogen gas before being treated with trifluoroacetic anhydride (TFAA)/CH_2Cl_2 (1:1, 500 µL) at 100 °C for 5 min. Finally, each reaction mixture was dried under nitrogen gas, dissolved in $CHCl_2$ and 1 µL was injected for GC analysis. The chiral-phase GC analysis of the N-trifluoroacetyl (TFA)/methyl ester derivatives was performed using a CP-Chirasil-D-Val column (Alltech, 0.25 mm × 25 m; N_2 as the carrier gas; program rate 50-200 °C at 4 °C min^{-1}). Standard amino acids were also converted to the TFA/Me derivatives by the same procedure. Retention times (min) were compared.

DNA sequencing and analysis

DNA sequencing was performed with a Genome analyzer II (Illumina). Small gaps were closed by primer walking on an ABI 15 PRISM 3100 Genetic Analyzer (Applied Biosystems). Analysis of the ORFs was performed using Geneious Pro 5.5.6, in combination with FramePlot

2.3.2 (http://www0.nih.go.jp/~jun/cgi-bin/frameplot.pl) Blast analysis and NRPS predictor.

Supplementary Information

Supplementary information (Figure S1-S38) is available free of charge at http://jbcs.sbq.org.br as a PDF file.

Acknowledgements

R. H. thanks the China Scholarship Council for the visiting Ph.D. student program. This work was partly supported by The Mitsubishi Foundation, Strategic Research Foundation Grant-aided Project for Private Universities, Grants-in-Aids from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, and the National Natural Science Foundation of China (Grant No.11172337).

References

1. Faulkner, D. J.; *Nat. Prod. Rep.* **2000**, *17*, 1.
2. Singh, B. K.; Macdonald, C. A.; *Drug Discov. Today* **2010**, *15*, 792.
3. Handelsman, J.; *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 669.
4. Banik, J. J.; Brady, S. F.; *Curr. Opin. Microbiol.* **2010**, *13*, 603.
5. Kennedy, J.; Flemer, B.; Jackson, S. A.; Lejon, D. P. H.; Morrissey, J. P.; O'Gara, F.; Dobson, A. D. W.; *Mar. Drugs* **2010**, *8*, 608.
6. Kato, Y.; Fusetani, N.; Matsunaga, S.; Hashimoto, K.; Fujita, S.; Furuya, T.; *J. Am. Chem. Soc.* **1986**, *108*, 2780.
7. Kimura, M.; Wakimoto, T.; Egami, Y.; Co, K. T.; Ise, Y.; Abe, I.; *J. Nat. Prod.* **2012**, *75*, 290.
8. He, R.; Wakimoto, T.; Takeshige, Y.; Egami, Y.; Kenmoku, H.; Ito, T.; Wang, B. C.; Asakawa, Y.; Abe, I.; *Mol Biosyst.* **2012**, *8*, 2334.
9. He, R.; Wakimoto, T.; Egami, Y.; Kenmoku, H.; Ito, T.; Asakawa, Y.; Abe, I.; *Bioorg. Med. Chem. Lett.* **2012**, *22*, 7322.
10. Furukawa, T.; Akutagawa, T.; Funatani, H.; Uchida, T.; Hotta, Y.; Niwa, M.; Takaya, Y.; *Bioorgan. Med. Chem.* **2012**, *20*, 2002.
11. Pedras, M. S. C.; Yu, Y.; Liu, J.; Tandron-Moya, Y. A.; *Z. Naturforsch., C: J. Biosci.* **2005**, *60*, 717.
12. Fdhila, F.; Vázquez, V.; Sánchez, J. L.; Riguera, R.; *J. Nat. Prod.* **2003**, *66*, 1299.
13. Li, X.; Dobretsov, S.; Xu, Y.; Xiao, X.; Hung, O. S.; Qian, P. Y.; *Biofouling* **2006**, *22*, 187.
14. Xiang, L.; Guo, D. L.; Ju, R.; Ma, B.; Lei, F.; Du, L. J.; *Chin. Tradit. Herbal Drugs* **2007**, *38*, 1622.
15. Ivanova, V.; Graefe, U.; Schlegel, R.; Schlegel, B.; Gusterova, A.; Kolarova, M.; Aleksieva, K.; *Biotechnol. Biotec. Eq.* **2003**, *17*, 2128.
16. Cheng, Z. H.; Wu, T.; Yu, B. Y.; *Nat. Prod. Res. Dev.* **2005**, *17*, 1.
17. Kobayashi, M.; Aoki, S.; Gato, K.; Matsunami, K.; Kurosu, M.; Kitagawa, I.; *Chem. Pharm. Bull.* **1994**, *42*, 2449.
18. Pedras, M. S. C.; Smith, K. C.; Taylor, J. L.; *Phytochemistry* **1998**, *49*, 1575.
19. Kimura, Y.; Sawada, A.; Kuramata, M.; Kusano, M.; Fujioka, S.; Kawano, T.; Shimada, A.; *J. Nat. Prod.* **2005**, *68*, 237.
20. Borthwick, A. D.; *Chem. Rev.* **2012**, *112*, 3641.
21. de Carvalho, M. P.; Abraham, W. R.; *Curr. Med. Chem.* **2012**, *19*, 3564.
22. Martins, M. B.; Carvalho, I.; *Tetrahedron* **2007**, *63*, 9923.
23. Bolognesi, M. L.; Ai Tran, H. N.; Staderini, M.; Monaco, A.; López-Cobeñas, A.; Bongarzone, S.; Biarnés, X.; López-Alvarado, P.; Cabezas, N.; Caramelli, M.; Carloni, P.; Menéndez, J. C.; Legname, G.; *ChemMedChem.* **2010**, *5*, 1324.
24. Campbell, J.; Lin, Q.; Geske, G. D.; Blackwell, H. E.; *ACS Chem. Biol.* **2009**, *4*, 1051.
25. Furtado, N. A. J. C.; Pupo, M. T.; Carvalho, I.; Campo, V. L.; Duarte, M. C. T.; Bastos, J. K.; *J. Braz. Chem. Soc.* **2005**,*16*, 1448.
26. Cheenpracha, S.; Borris, R. P.; Tran, T. T.; Jee, J. M.; Seow, H. F.; Cheah, H. Y.; Ho, C. C.; Chang, L. C.; *J. Braz. Chem. Soc.* **2011**, *22*, 223.
27 Gondry, M.; Sauguet, L.; Belin, P.; Thai, R.; Amouroux, R.; Tellier, C.; Tuphile, K.; Jacquet, M.; Braud, S.; Courçon, M.; Masson, C.; Dubois, S.; Lautru, S.; Lecoq, A.; Hashimoto, S.; Genet, R.; Pernodet, J.; *Nat. Chem. Biol.* **2009**, *5*, 414.
28. Maiya, S.; Grundmann, A.; Li, S. M.; Turner, G.; *ChemBioChem.* **2006**, *7*, 1062.
29. Ding, Y.; de Wet, J. R.; Cavalcoli, J.; Li, S.; Greshock, T. J.; Miller, K. A.; Finefield, J. M.; Sunderhaus, J. D.; McAfoos, T. J.; Tsukamoto, S.; Williams, R. M.; Sherman, D. H.; *J. Am. Chem. Soc.* **2010**, *132*, 12733.
30. Sauguet, L.; Moutiez, M.; Li, Y.; Belin, P.; Seguin, J.; Le Du, M. H.; Thai, R.; Masson, C.; Fonvielle, M.; Pernodet, J. L.; Charbonnier, J. B.; Gondry, M.; *Nucleic Acids Res.* **2011**, *39*,4475.
31. Belin, P.; Moutiez, M.; Lautru, S.; Seguin, J.; Pernodet, J. L.; Gondry, M.; *Nat. Prod. Rep.* **2012**, *29*, 961.
32. Lautru, S.; Gondry, M.; Genet, R.; Pernodet, J. L.; *Chem. Biol.* **2002**, *9*, 1355.
33. Vetting, M. W.; Hegde, S. S.; Blanchard, J. S.; *Nat. Chem. Biol.* **2010**, *6*, 797.
34. Bonnefond, L.; Arai, T.; Sakaguchi, Y.; Suzuki, T.; Ishitani, R.; Nureki, O.; *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3912.
35. Brady, S. F.; *Nat. Protoc.* **2007**, *2*, 1297.

Submitted: February 27, 2013
Published online: October 2, 2013

J. Braz. Chem. Soc., Vol. 24, No. 12, S1-S21, 2013.
Printed in Brazil - ©2013 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Supplementary Information

Cyclodipeptides from Metagenomic Library of a Japanese Marine Sponge

Rui He,[a,b,c] Bochu Wang,*,[b] Toshiyuki Wakimoto,[c] Manyuan Wang,[a] Liancai Zhu[b] and Ikuro Abe*,[c]

[a]*School of Traditional Chinese Medicine, Capital University of Medical Sciences, No. 10 Xitoutiao, You An Men, 100069 Beijing, P. R. China*

[b]*Bioengineering College, Chongqing University, No. 174, Shanpingba Main Street, 400030 Chongqing, P. R. China*

[c]*Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-0033 Tokyo, Japan*

Comparative data of cyclodipeptide 1-7 production in clone pDC113 and negative control (NC, strain EPI300 carrying pCC1FOS fosmid vector)

The respective 2 plates of pDC113 and NC were cultured in the same conditions (30 °C, **3d**) and subjected to the same extraction and separation procedures and finally using same volume of MeOH to dissolve the LH-20 cyclodipeptides fraction before injection (both 5 μL) to RP-HPLC-DAD. HPLC analysis was performed on ODS column (Cosmosil $5C_{18}$ PAQ waters, 4.6 × 250 mm) with a mixture of H_2O and MeCN, both containing 0.05% TFA: 0-20 min, 5-35% MeCN; 20-45 min, 35-100% MeCN; and 45-55 min, 100% MeCN, 0.8 mL min^{-1}. DAD profiles were measured with a Shimadzu HPLC System: LC-20AD and SPD-20A Prominence Diode Array Detector.

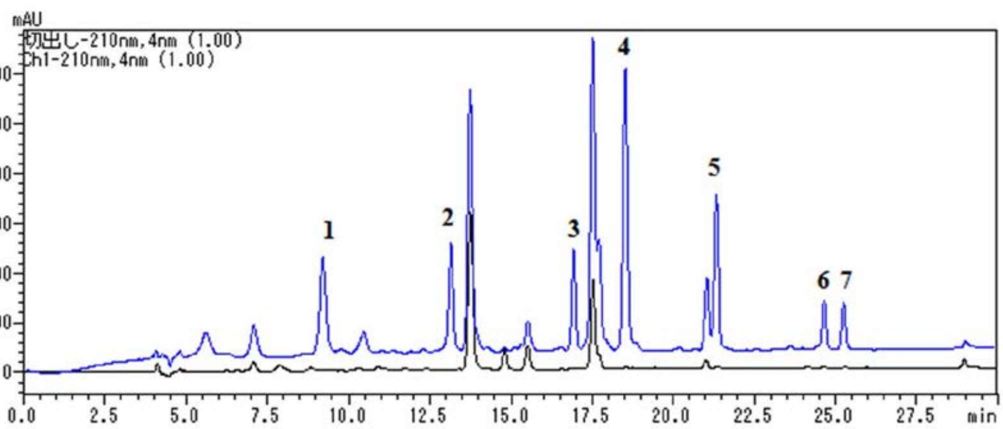


Figure S1. RP-HPLC-DAD profile of LH-20 fraction of pDC113 and NC. Blue line indicates profile of pDC113; black line indicates profile of negative control.

*e-mail: wangbc2000@126.com, abei@mol.f.u-tokyo.ac.jp

^{1}H, ^{13}C chemical shifts, and ^{1}H-^{1}H COSY data of cyclodipeptides 1-7 and 9

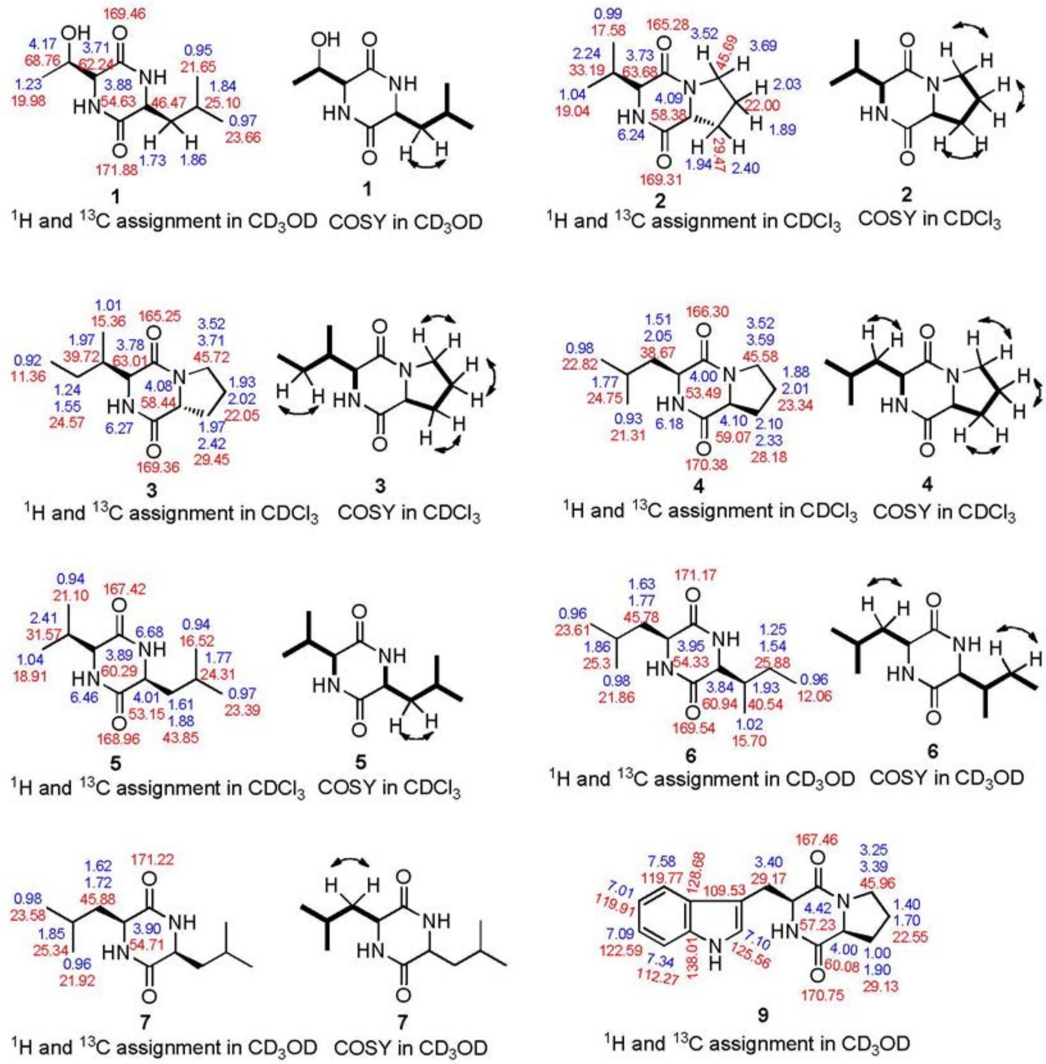


Figure S2. Chemical shifts and COSY of cyclodipeptides **1**-**7** and **9**. ^{1}H NMR (500 MHz) chemical shifts (blue), ^{13}C NMR (125 MHz) chemical shifts (red) and ^{1}H-^{1}H COSY (bold line and arrows) are shown.

^{1}H, ^{13}C chemical shifts, and ^{1}H-^{1}H COSY data of cyclodipeptides 8, 10 and 11

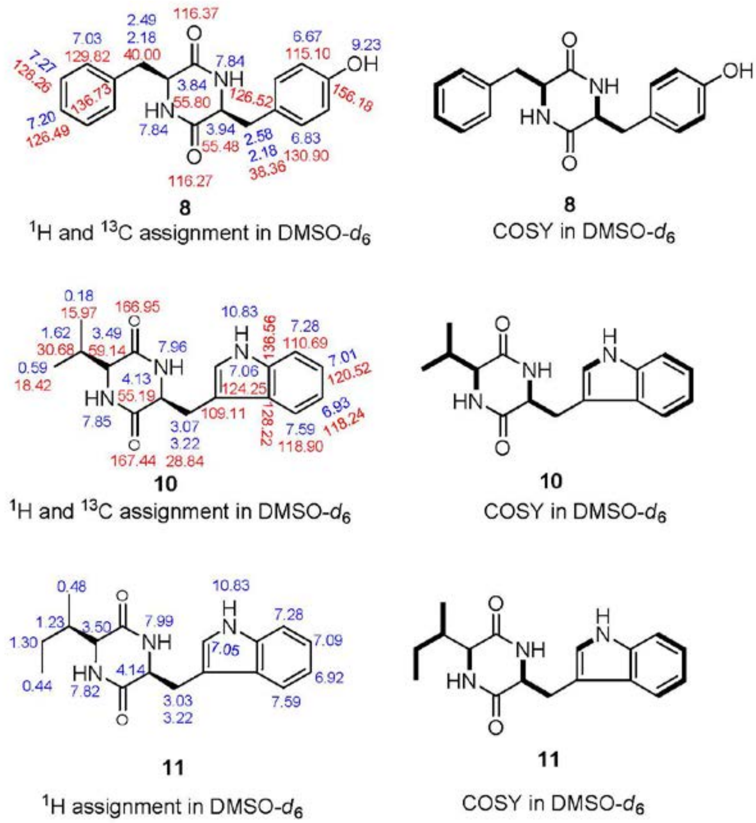


Figure S3. Chemical shifts and COSY of cyclodipeptides **8**, **10** and **11**. ^{1}H NMR (500 MHz) chemical shifts (blue), ^{13}C NMR (125 MHz) chemical shifts (red), and main ^{1}H-^{1}H COSY (bold line) correlations are shown. ^{13}C NMR of **10** were inferred from its HMQC and some of HMBC data.

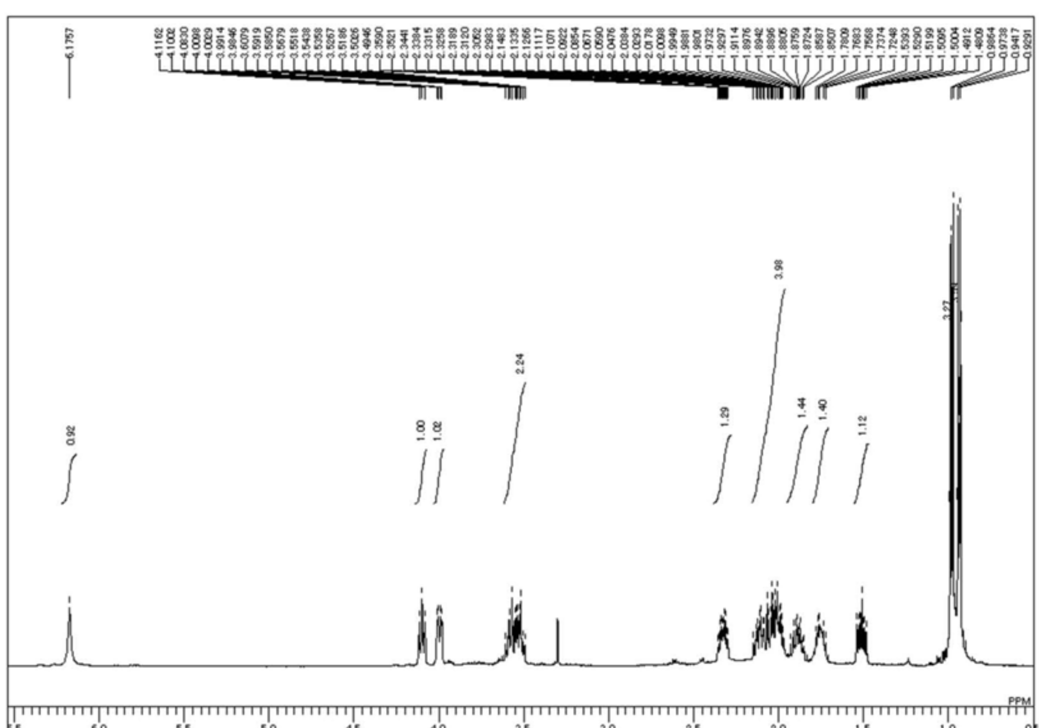


Figure S4. ^{1}H NMR spectrum of Cyclo(L-Leu-L-Pro) (**4**) (500 MHz, $CDCl_3$).

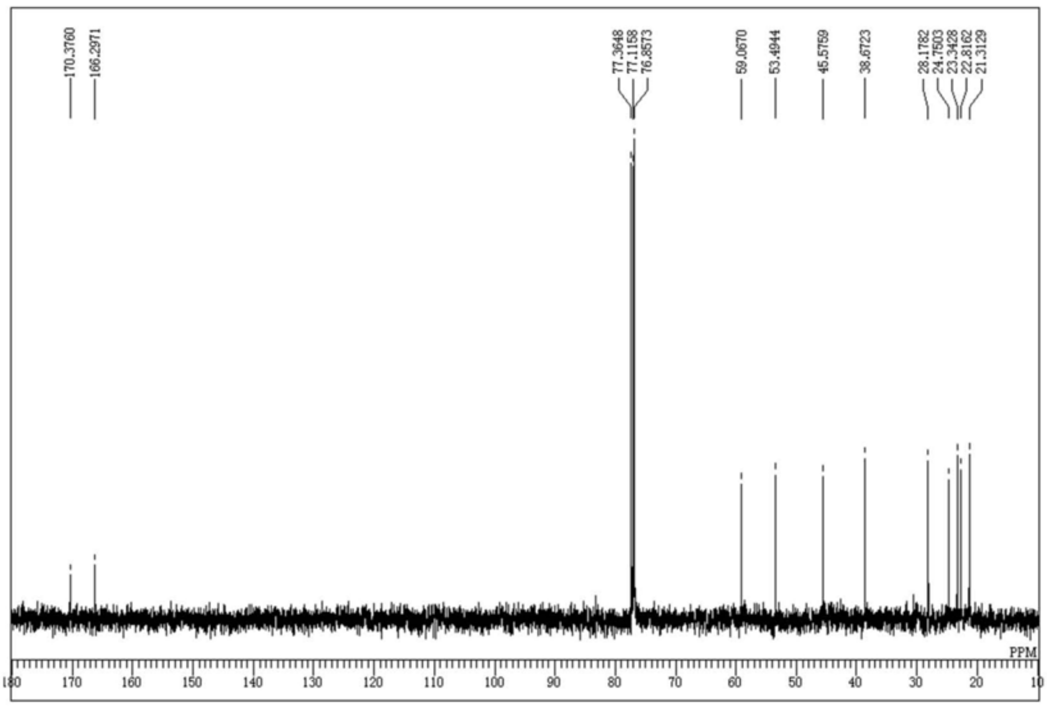


Figure S5. ^{13}C NMR spectrum of Cyclo(L-Leu-L-Pro) (**4**) (125 MHz, $CDCl_3$).

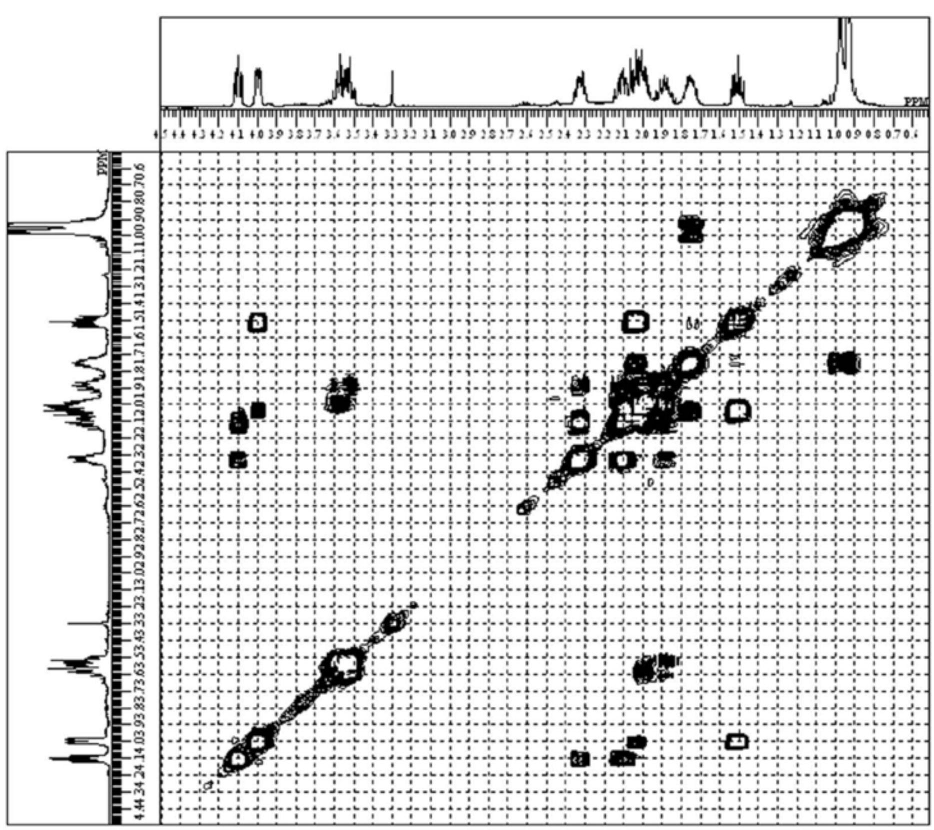


Figure S6. ^1H-^1H COSY spectrum of Cyclo(L-Leu-L-Pro) (**4**).

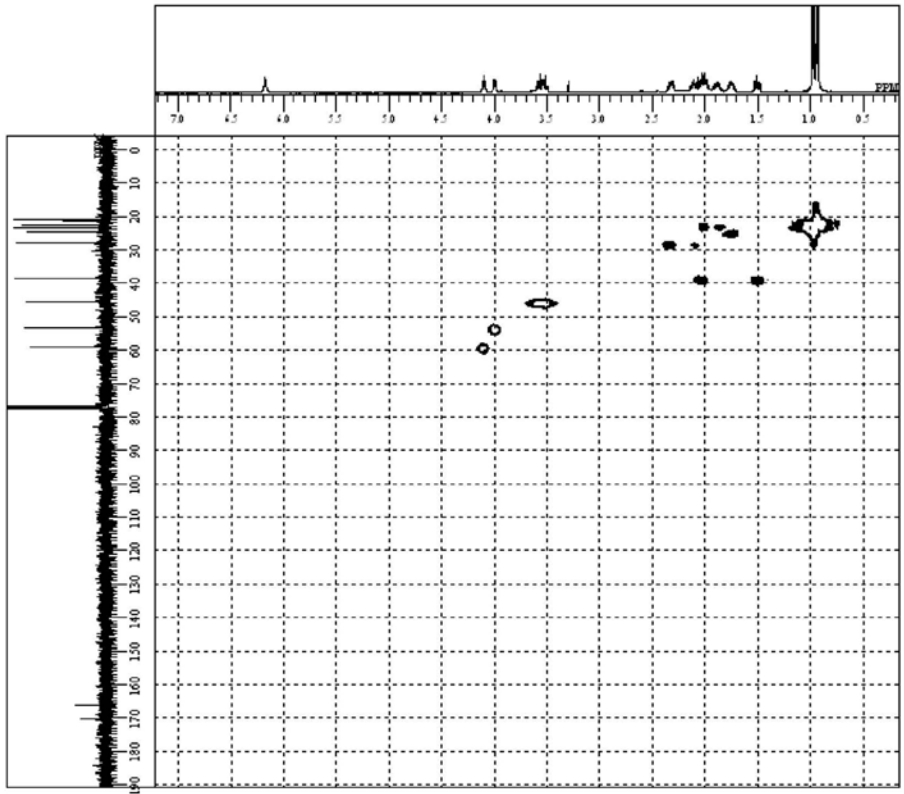


Figure S7. HMQC spectrum of Cyclo(L-Leu-L-Pro) (**4**).

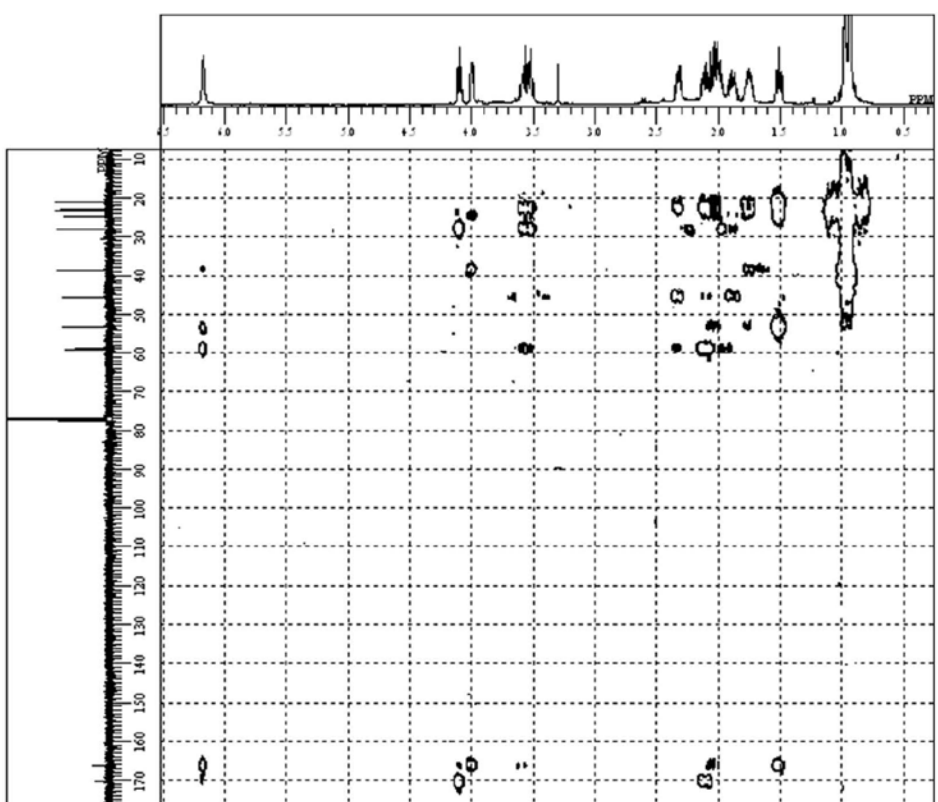


Figure S8. HMBC spectrum of Cyclo(L-Leu-L-Pro) (**4**).

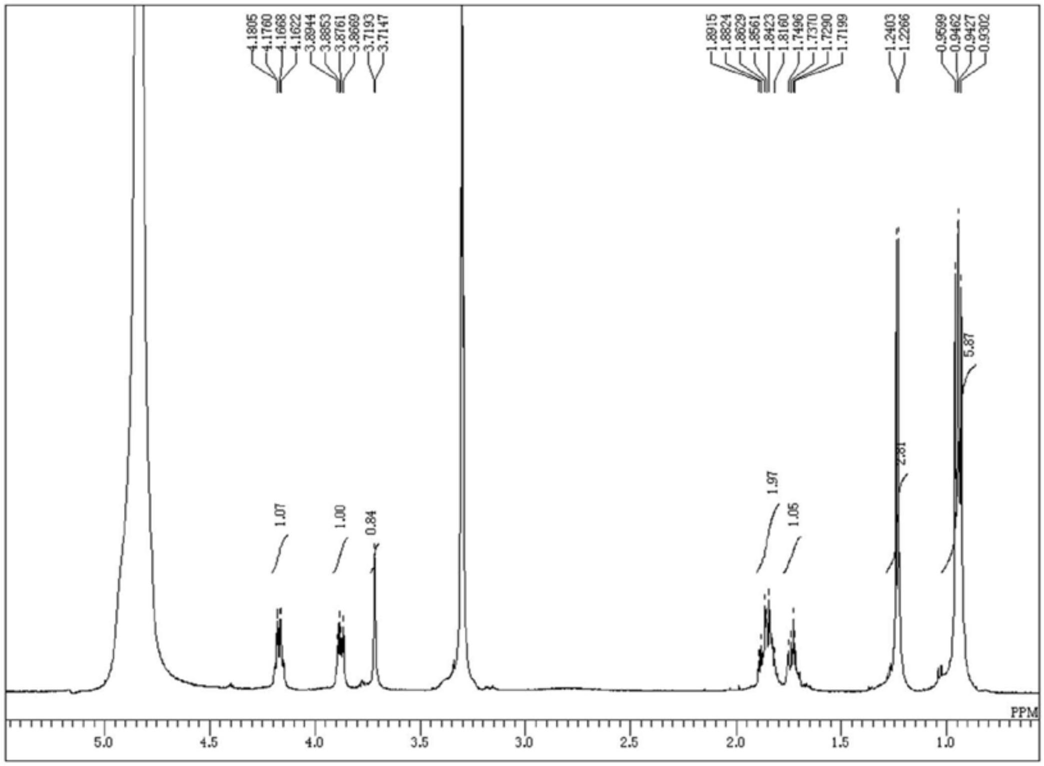


Figure S9. ^{1}H NMR of Cyclo(L-Thr-L-Leu) (**1**) (500 MHz, CD_3OD).

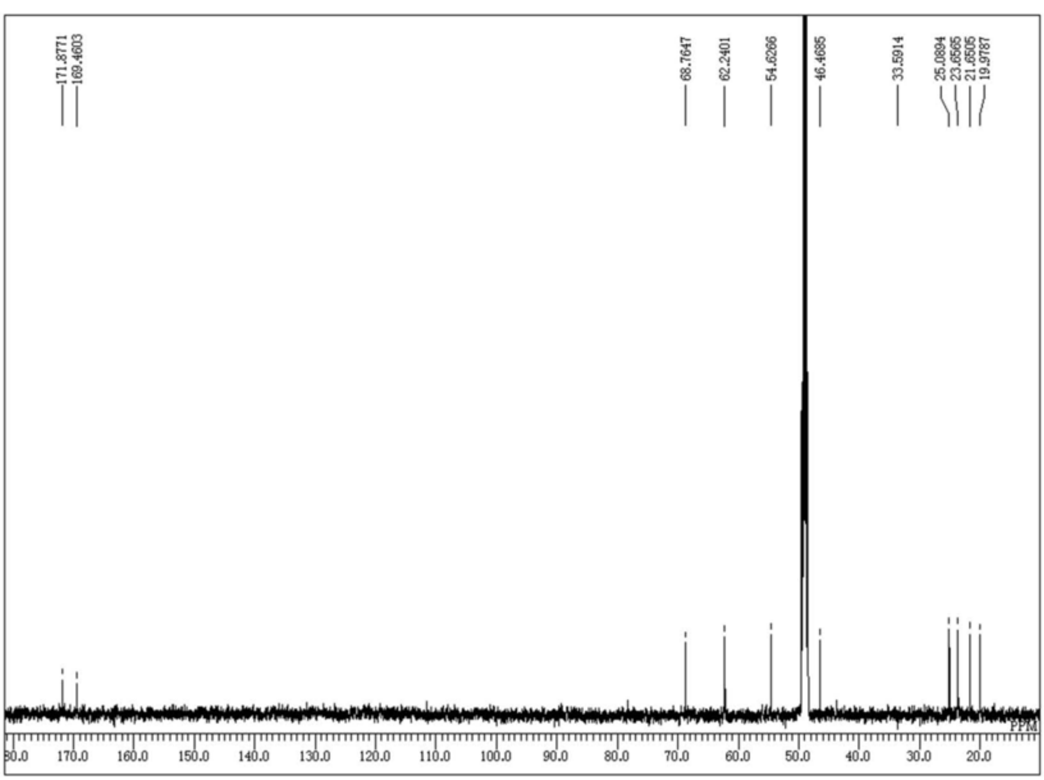


Figure S10. ^{13}C NMR of Cyclo(L-Thr-L-Leu) (**1**) (125 MHz, CD_3OD).

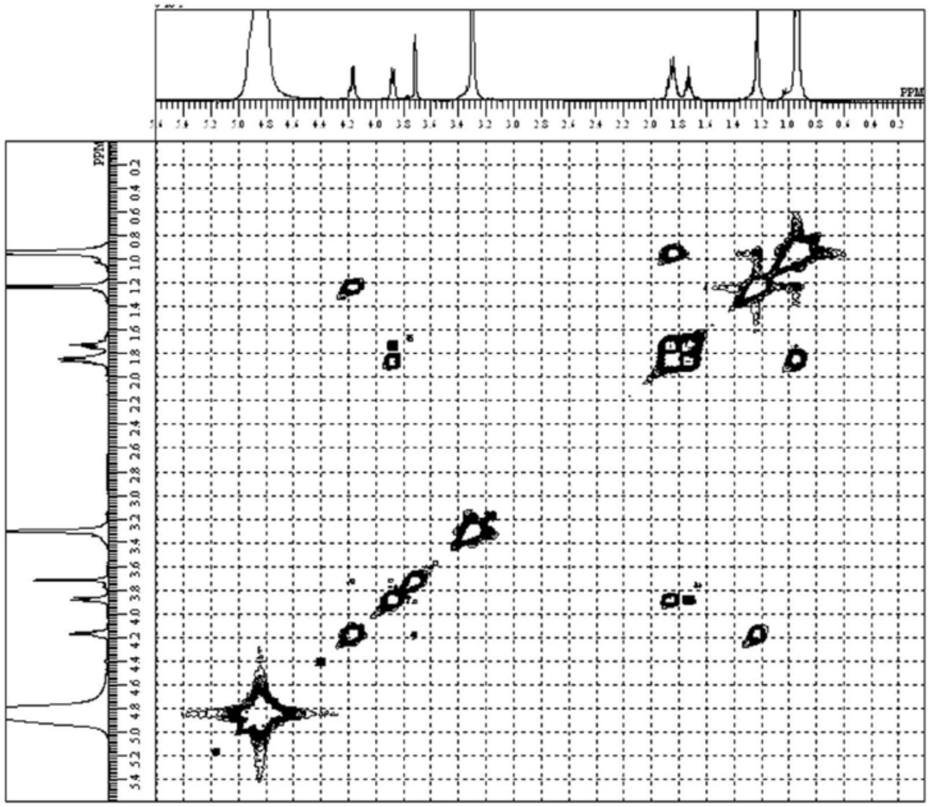


Figure S11. ^{1}H-^{1}H COSY of Cyclo(L-Thr-L-Leu) (**1**).

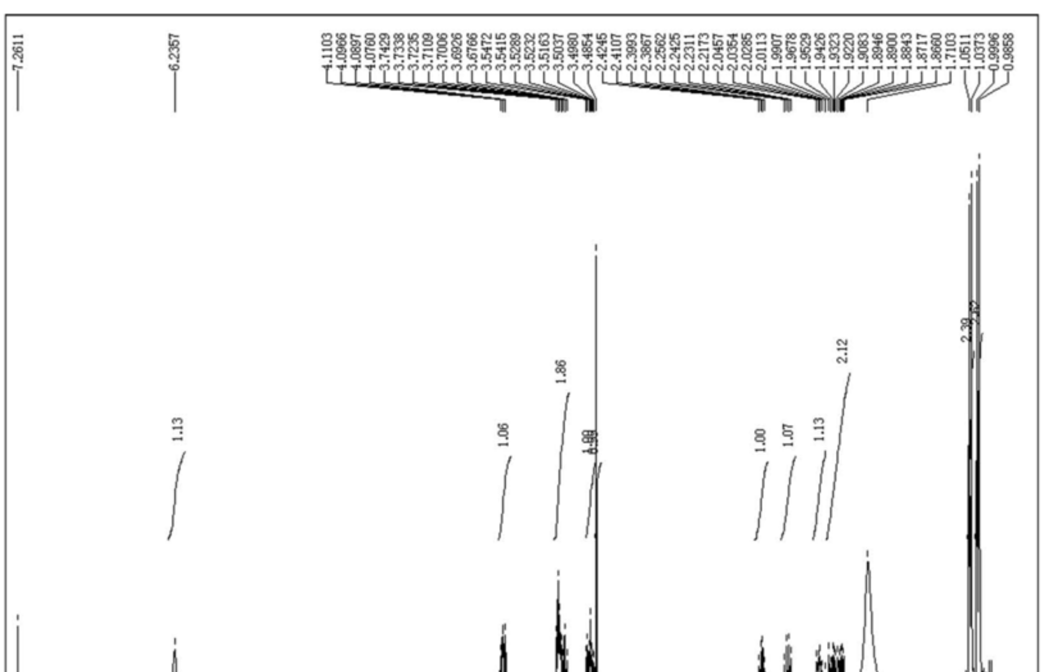


Figure S12. ^{1}H NMR of Cyclo(L-Val-D-Pro) (**2**) (500 MHz, $CDCl_3$).

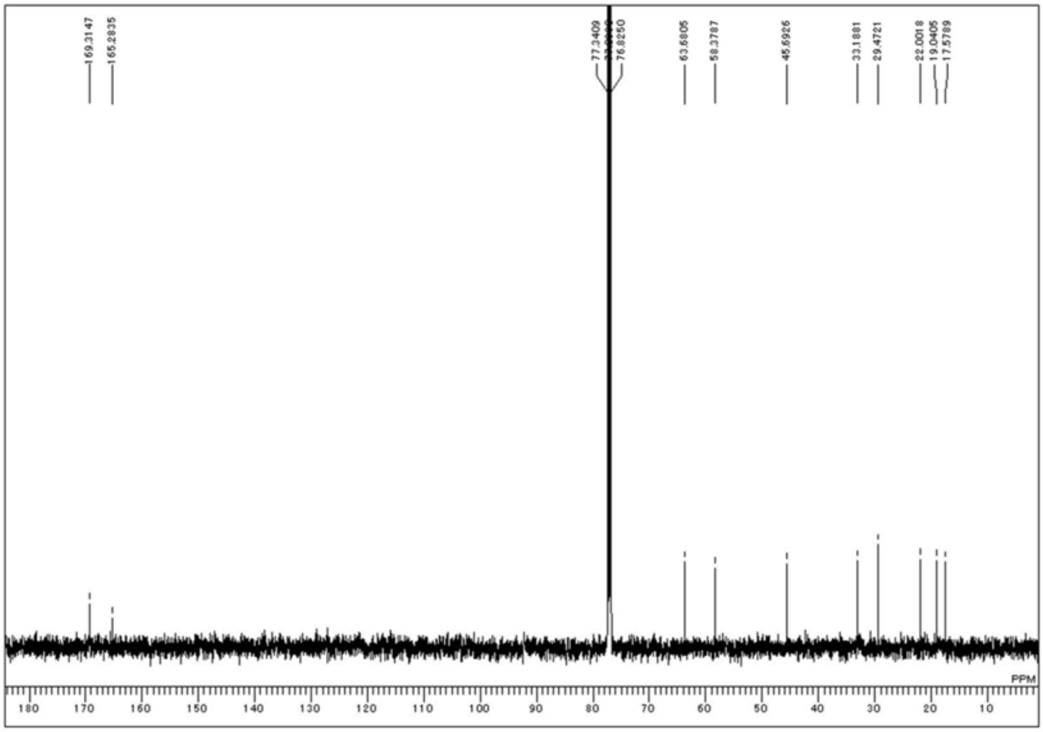


Figure S13. ^{13}C NMR of Cyclo(L-Val-D-Pro) (**2**) (125 MHz, $CDCl_3$).

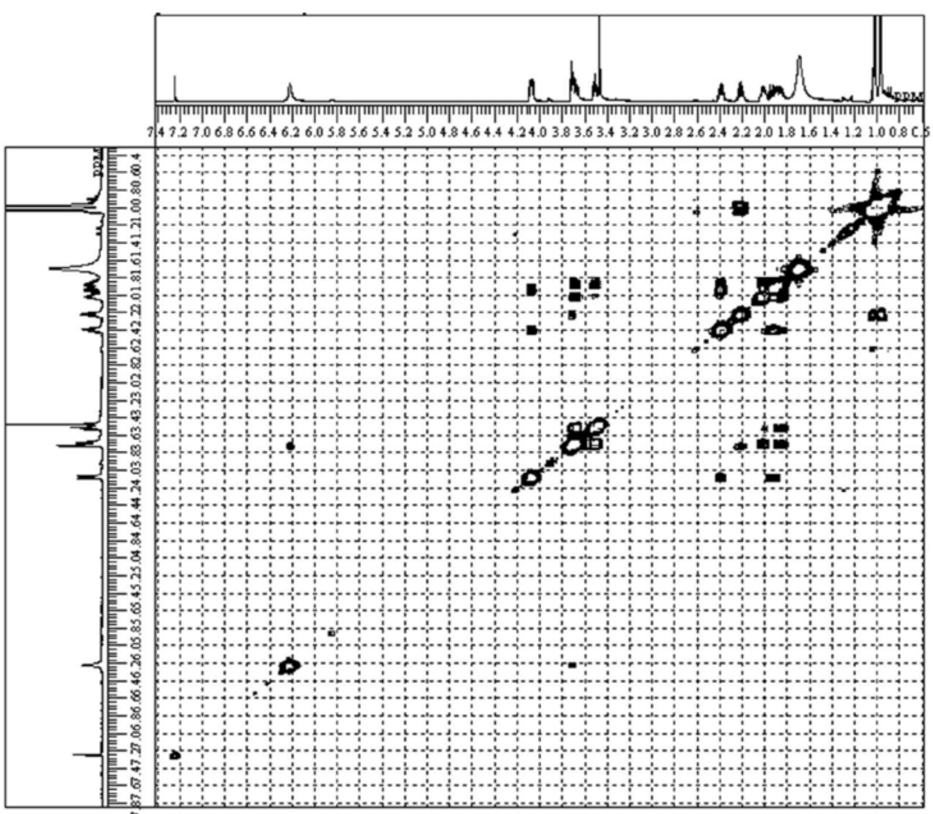


Figure S14. ^{1}H-^{1}H COSY of Cyclo(L-Val-D-Pro) (**2**).

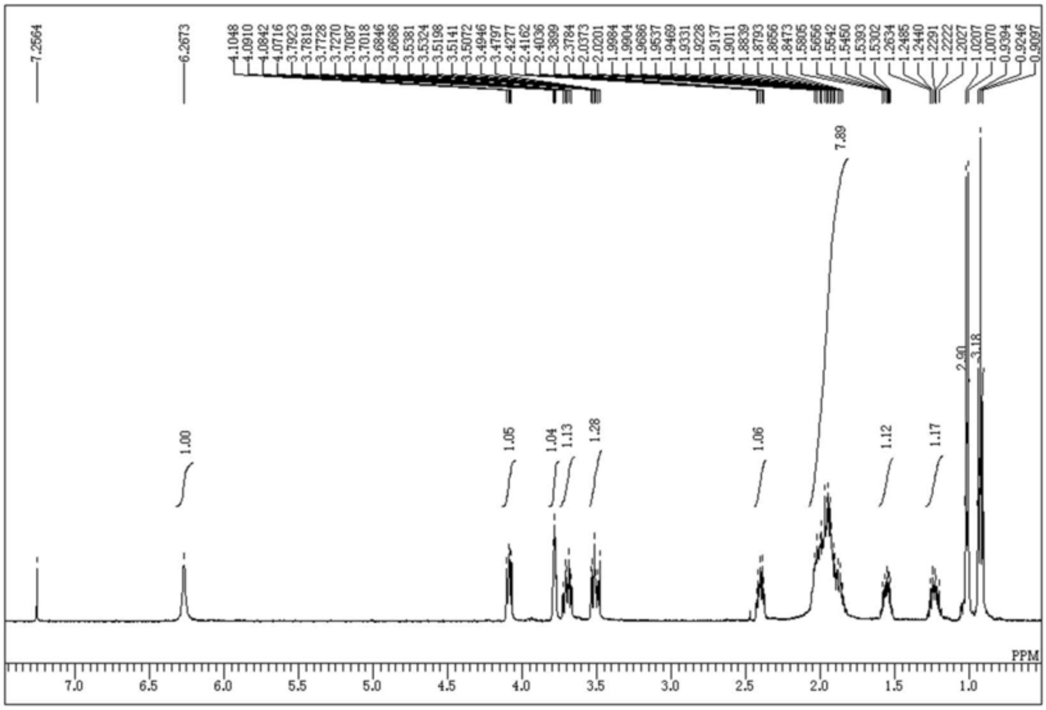


Figure S15. ^{1}H NMR of Cyclo(L-Ile-D-Pro) (**3**) (500 MHz, $CDCl_3$).

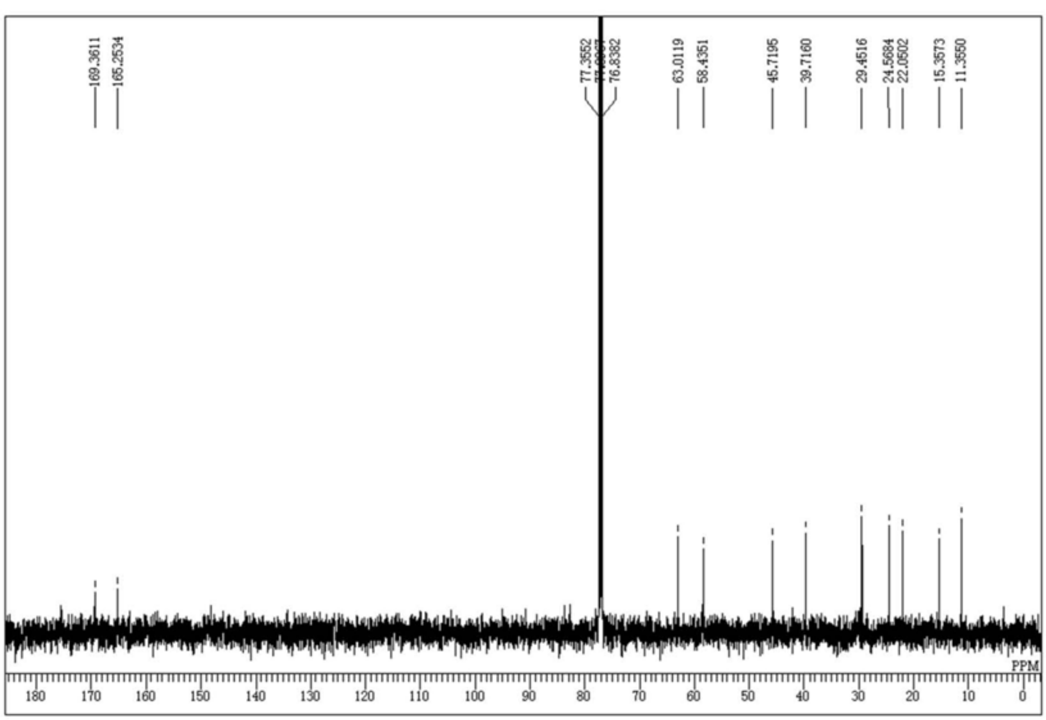


Figure S16. ^{13}C NMR of Cyclo(L-Ile-D-Pro) (**3**) (125 MHz, $CDCl_3$).

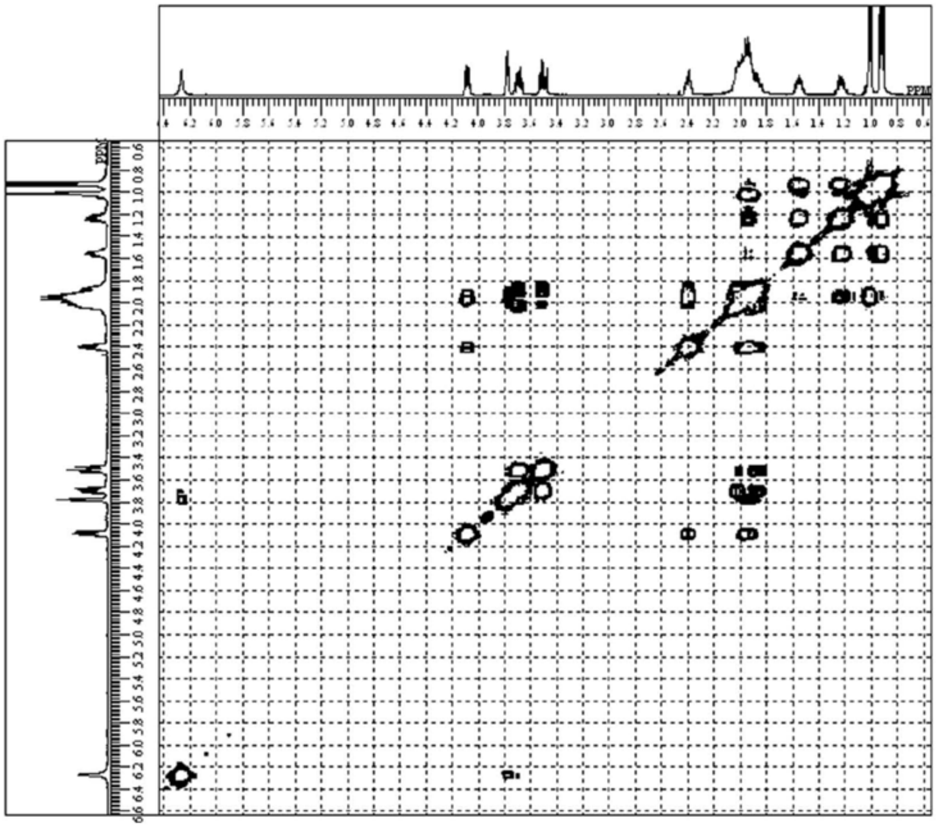


Figure S17. [1]H-[1]H COSY of Cyclo(L-Ile-D-Pro) (**3**).

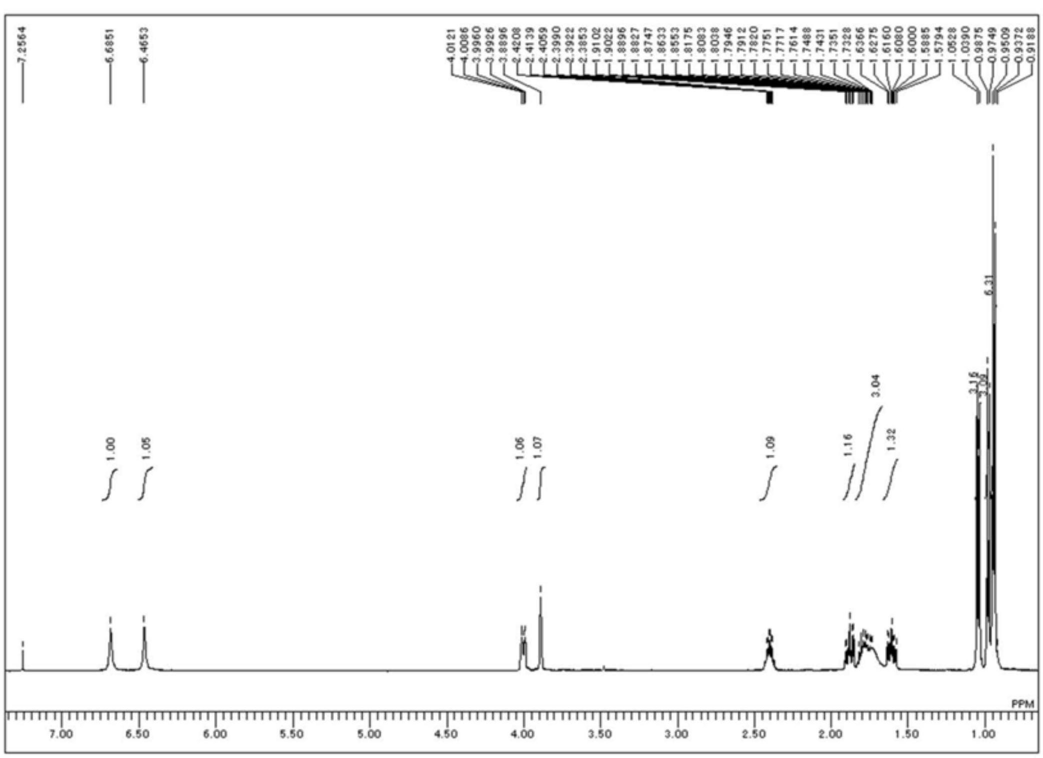


Figure S18. [1]H NMR of Cyclo(L-Val-L-Leu) (**5**) (500 MHz, $CDCl_3$).

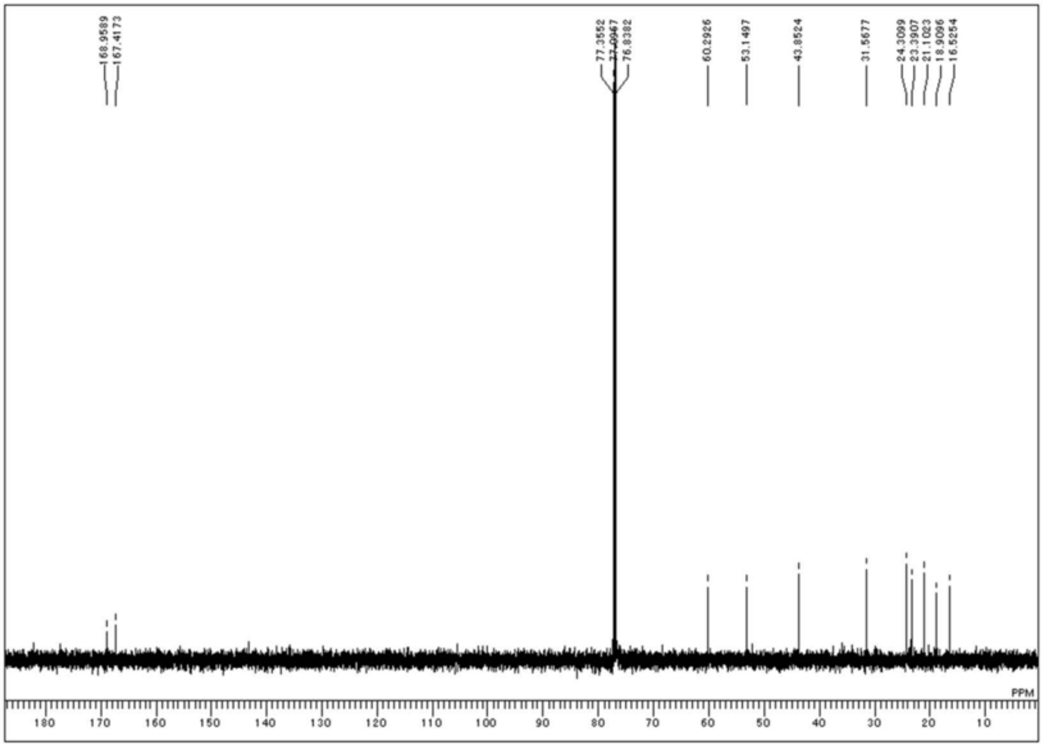


Figure S19. ^{13}C NMR of Cyclo(L-Val-L-Leu) (**5**) (125 MHz, $CDCl_3$).

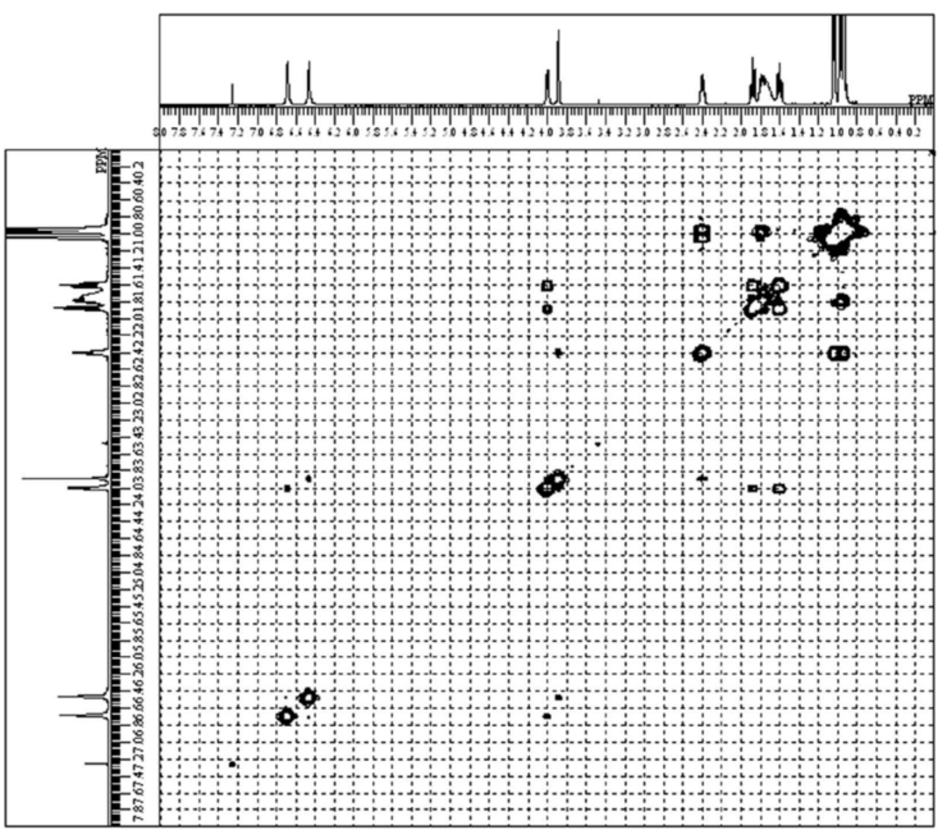


Figure S20. ^{1}H-^{1}H COSY of Cyclo(L-Val-L-Leu) (**5**).

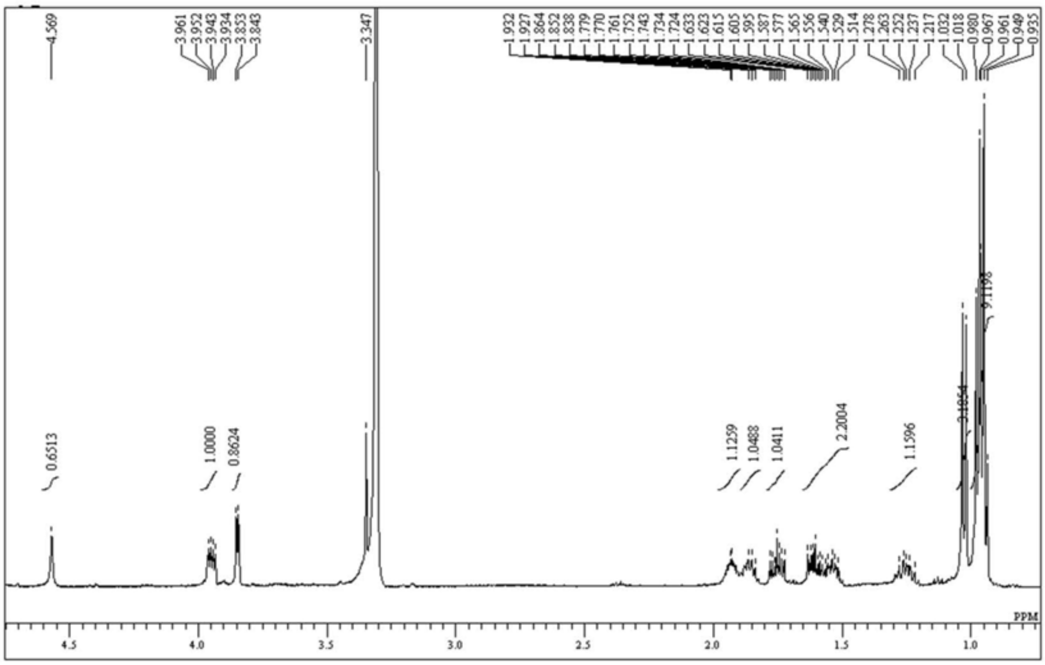


Figure S21. ^{1}H NMR of Cyclo(L-Leu-L-Ile) (**6**) (500 MHz, CD_3OD).

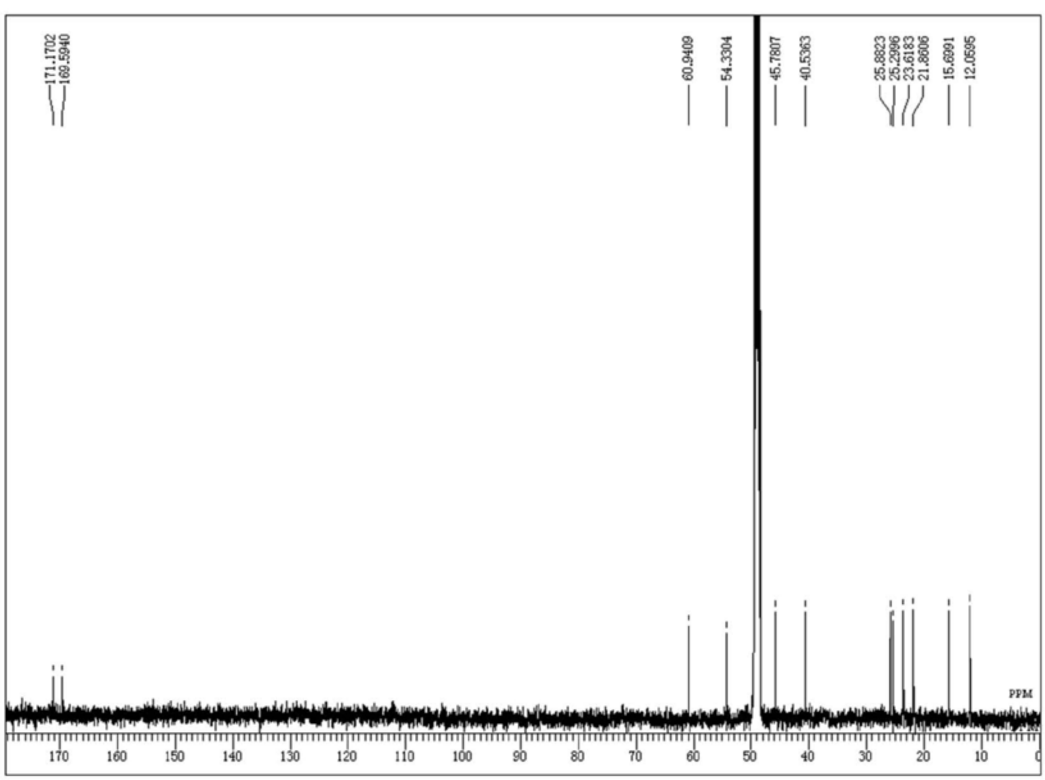


Figure S22. ^{13}C NMR of Cyclo(L-Leu-L-Ile) (**6**) (125 MHz, CD_3OD).

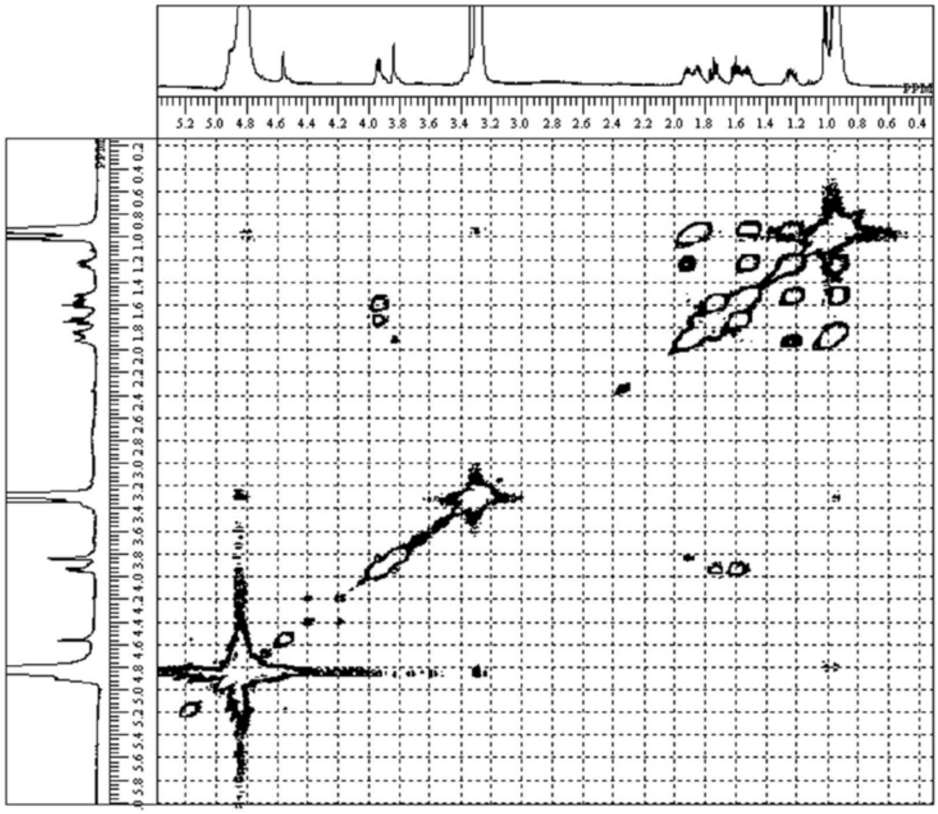


Figure S23. ^{1}H-^{1}H COSY of Cyclo(L-Leu-L-Ile) (**6**).

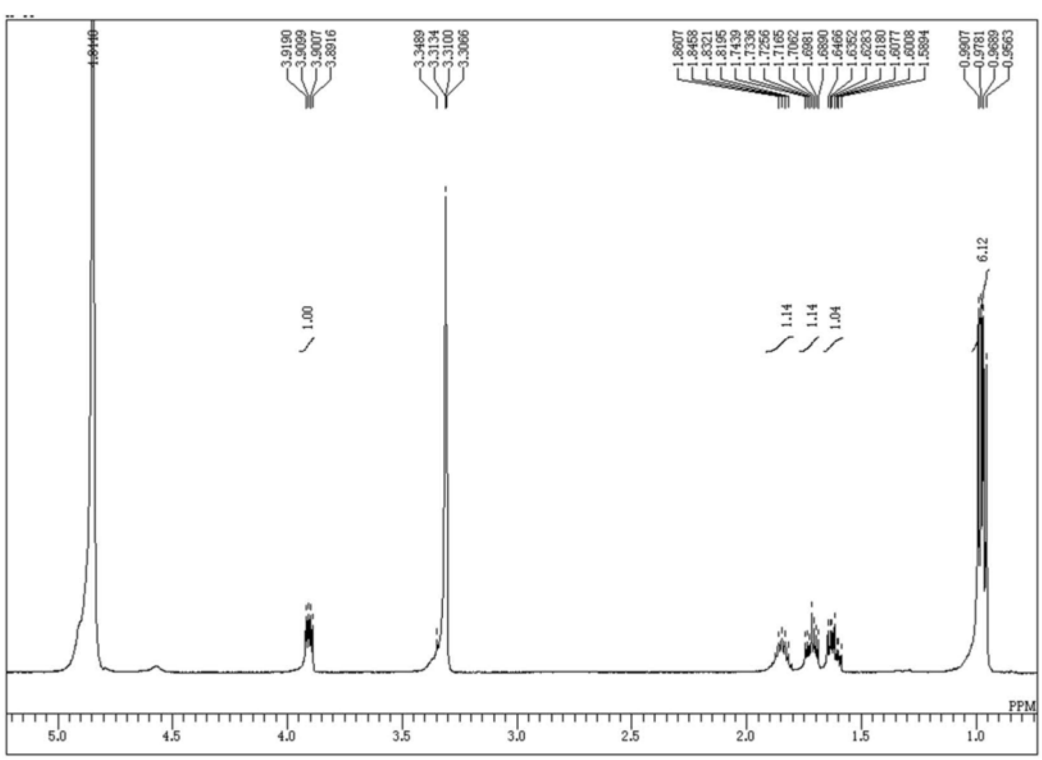


Figure S24. ^{1}H NMR of Cyclo(L-Leu-L-Leu) (**7**) (500 MHz, CD_3OD).

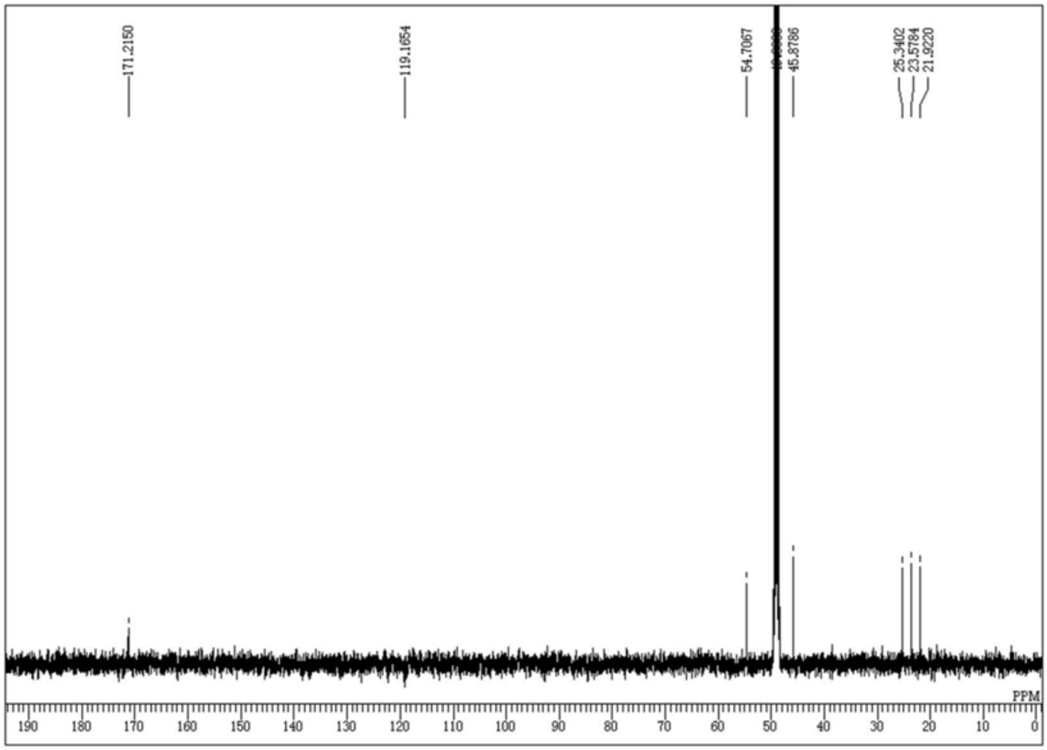


Figure S25. ^{13}C NMR of Cyclo(L-Leu-L-Leu) (**7**) (125 MHz, CD_3OD).

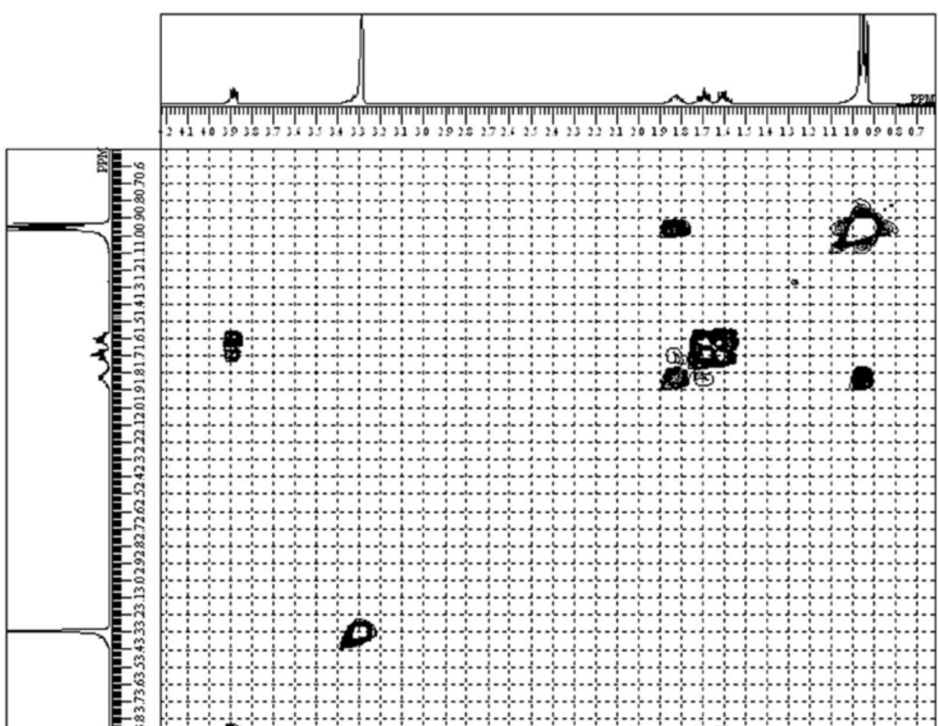


Figure S26. ^{1}H-^{1}HCOSY of Cyclo(L-Leu-L-Leu) (**7**).

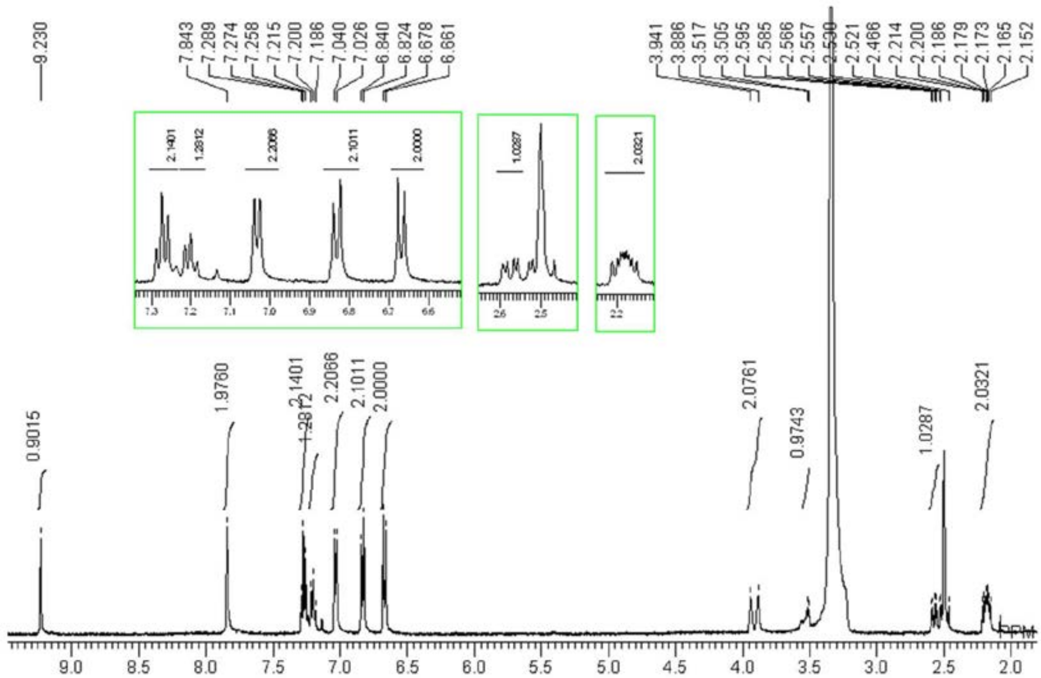


Figure S27. ^{1}H NMR of Cyclo(L-Phe-L-Tyr) (**8**) (500 MHz, DMSO).

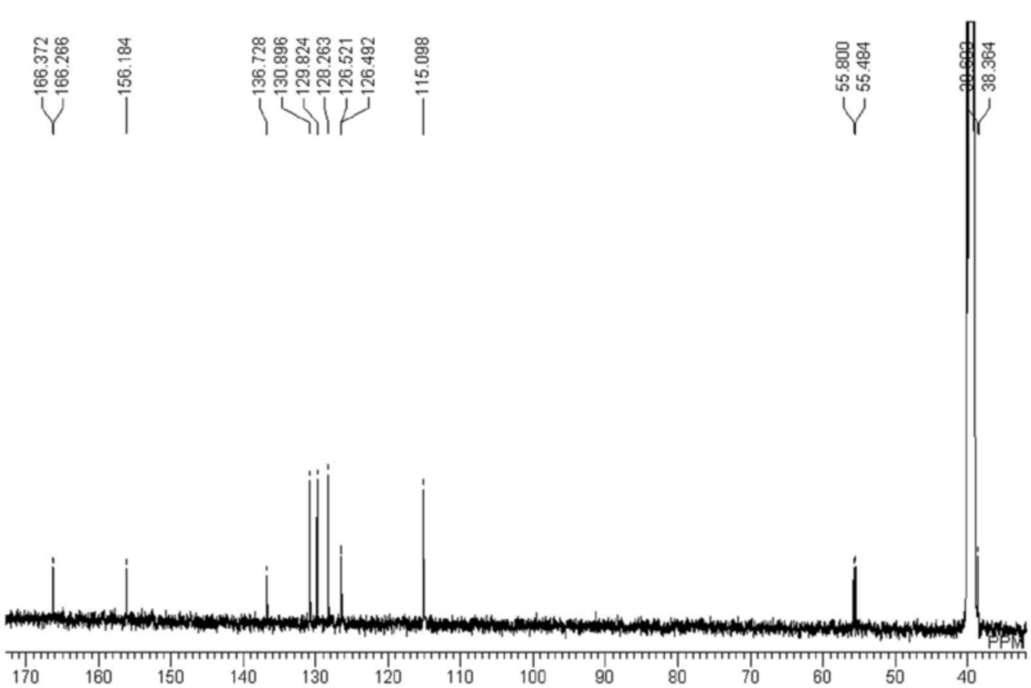


Figure S28. ^{13}C NMR of Cyclo(L-Phe-L-Tyr) (**8**) (125 MHz, DMSO).

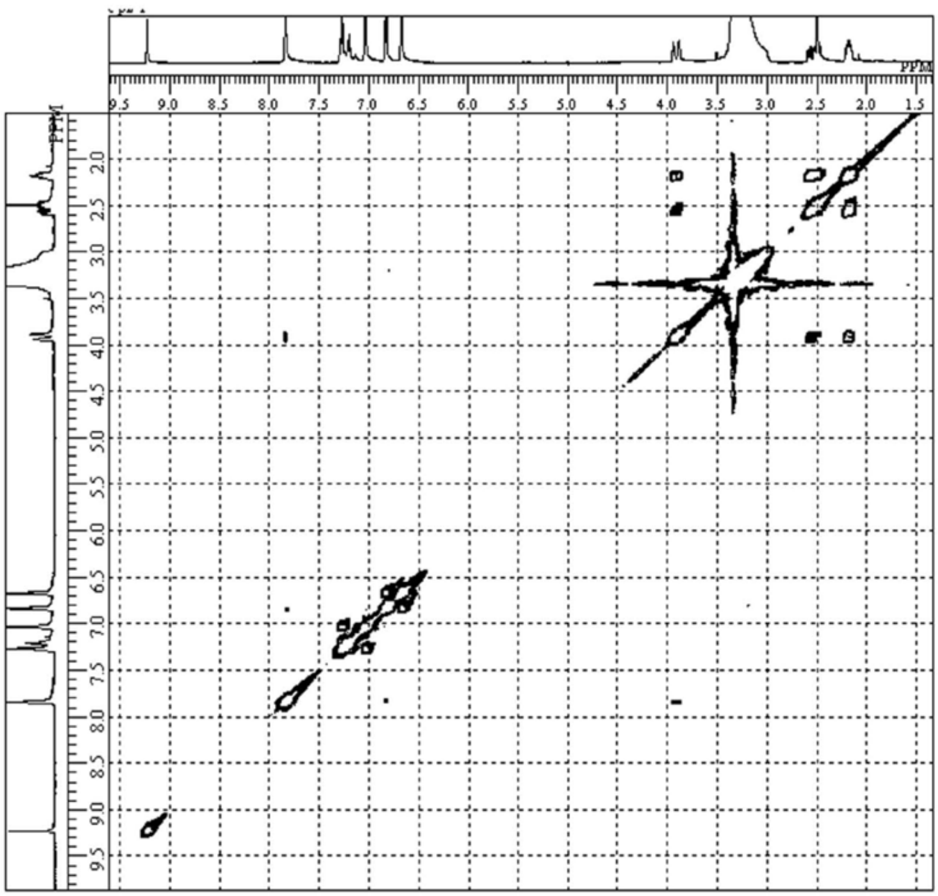


Figure S29. ^{1}H-^{1}H COSY of Cyclo(L-Phe-L-Tyr) (**8**).

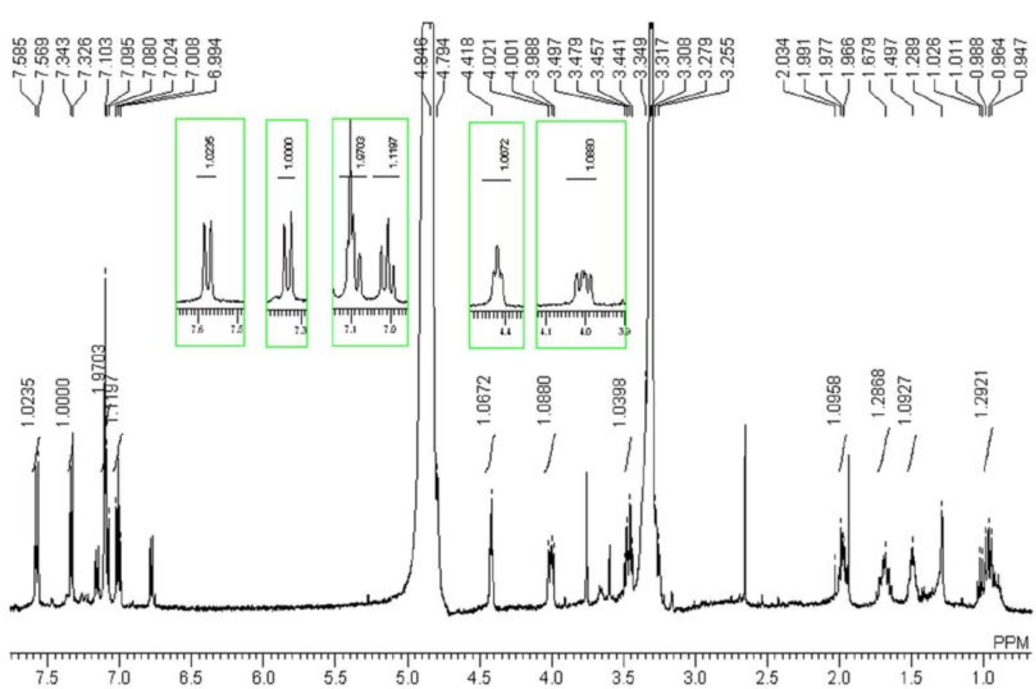


Figure S30. ^{1}H NMR of Cyclo(L-Trp-L-Pro) (**9**) (500 MHz, CD_3OD).

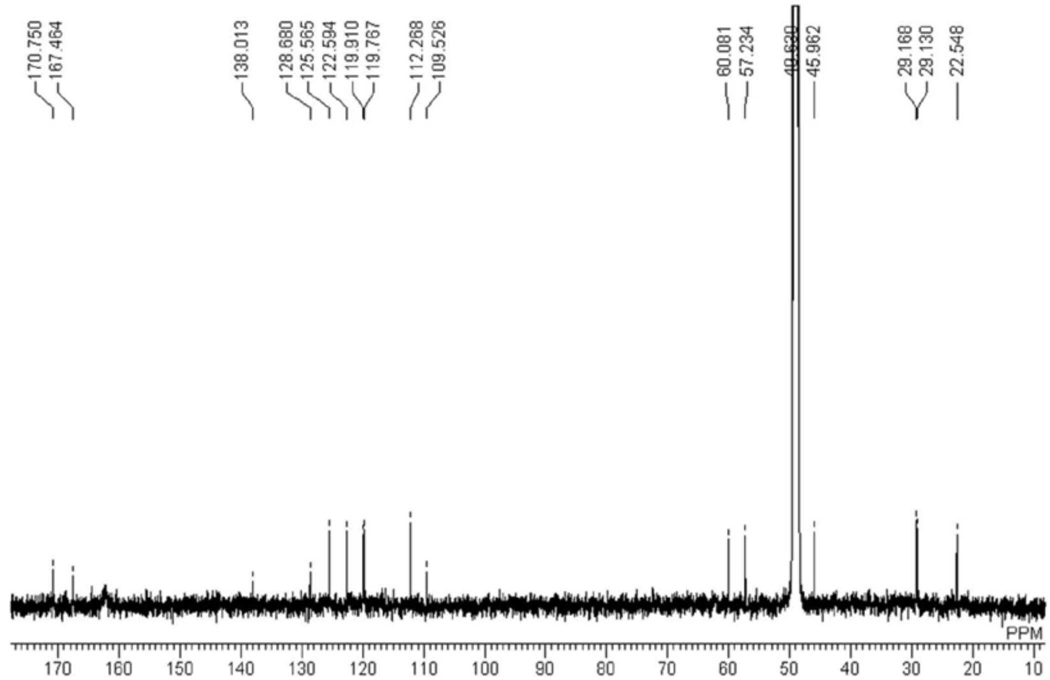


Figure S31. ^{13}C NMR of Cyclo(L-Trp-L-Pro) (**9**) (125 MHz, CD_3OD).

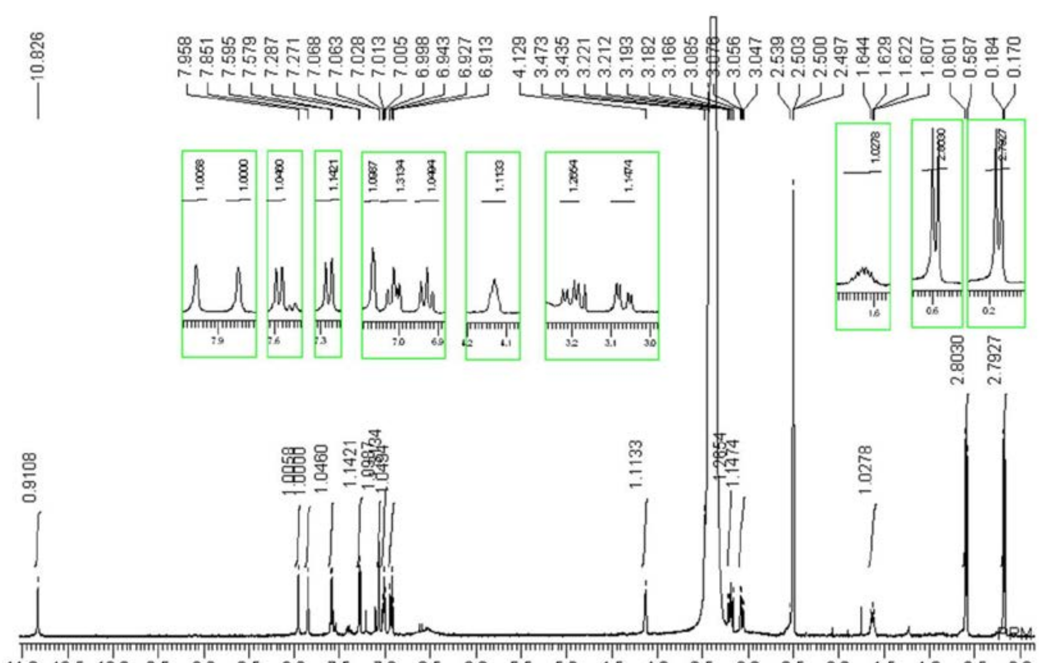


Figure S32. ^{1}H NMR of Cyclo(L-Val-L-Trp) (**10**) (500 MHz, DMSO).

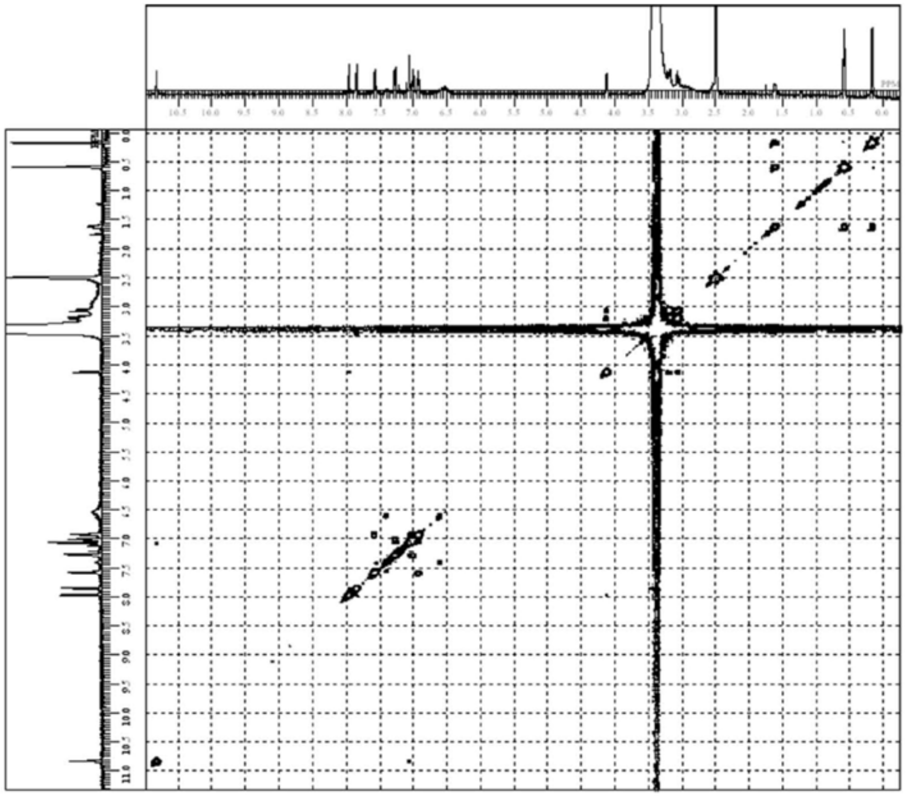


Figure S33. ^{1}H-^{1}H COSY of Cyclo(L-Val-L-Trp) (**10**).

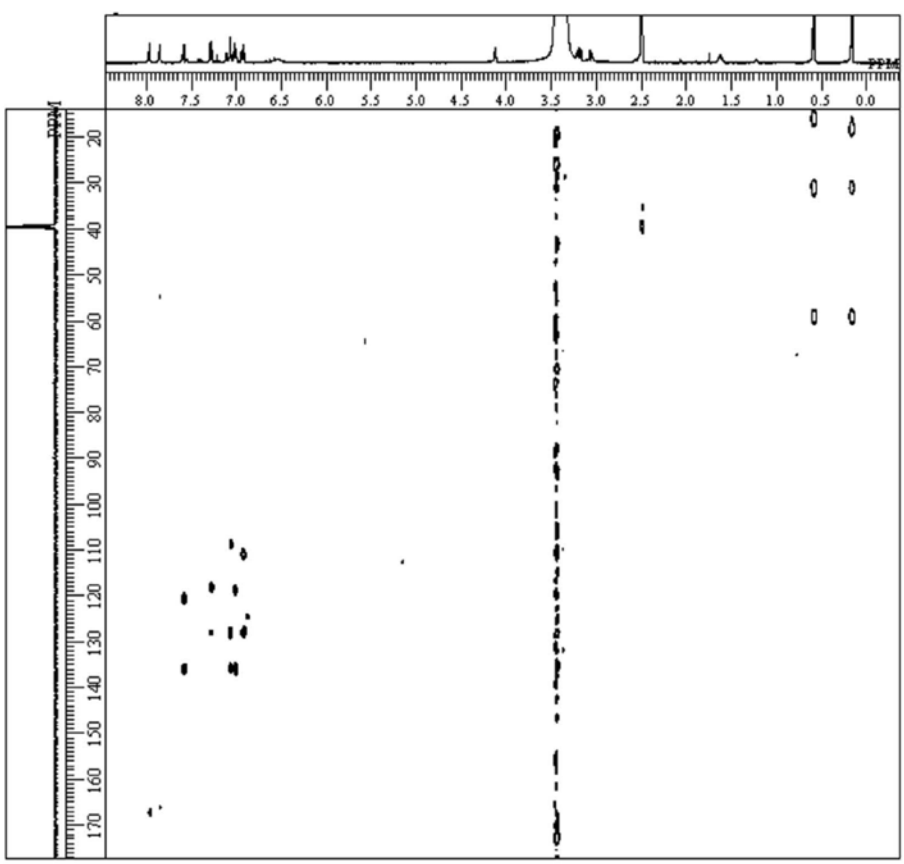


Figure S34. Some easily detected HMBC of Cyclo(L-Val-L-Trp) (**10**).

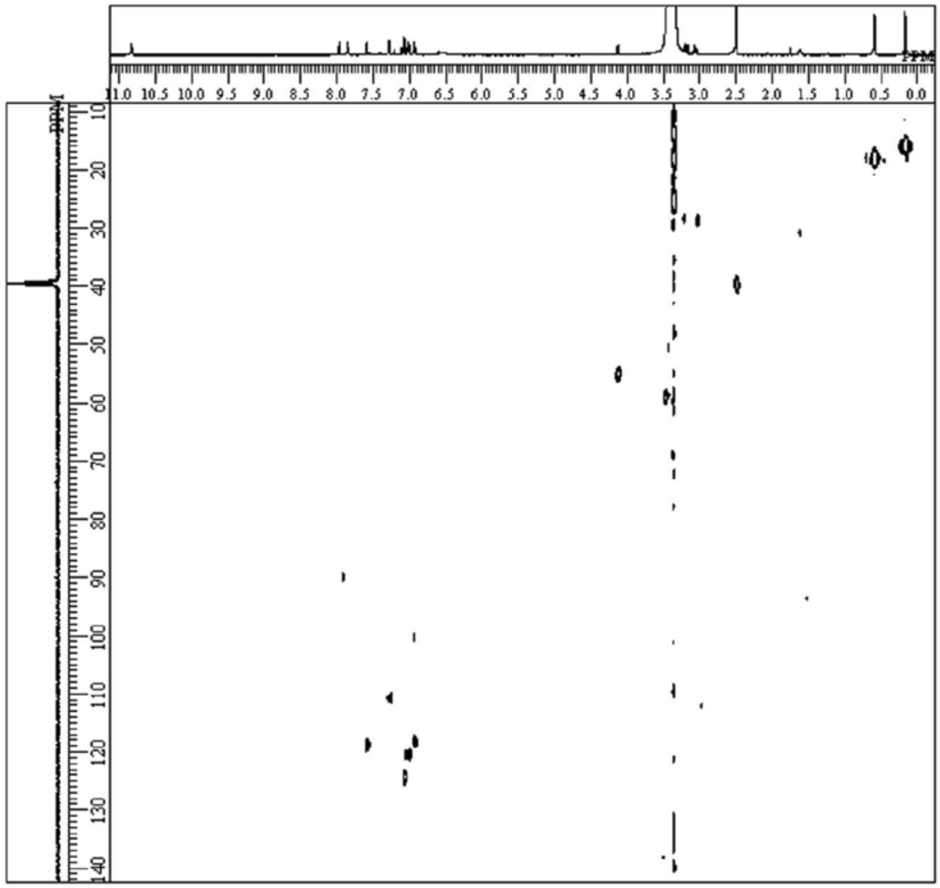


Figure S35. HMQC of Cyclo(L-Val-L-Trp) (**10**).

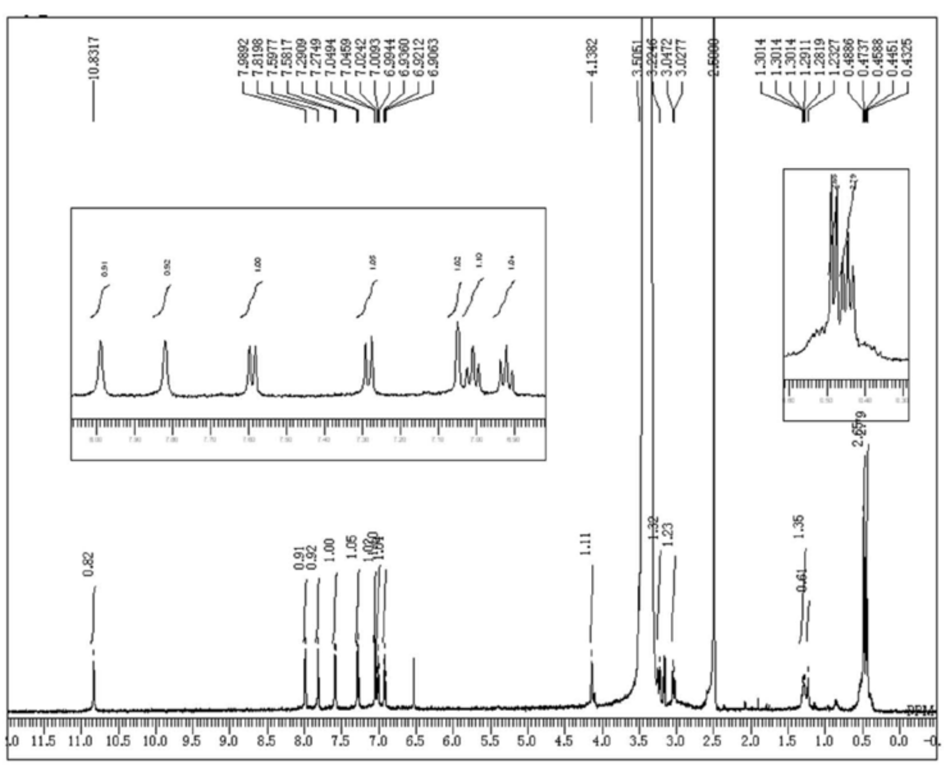


Figure S36. ^{1}H NMR of Cyclo(L-Ile-L-Trp) (**11**) (500 MHz, DMSO).

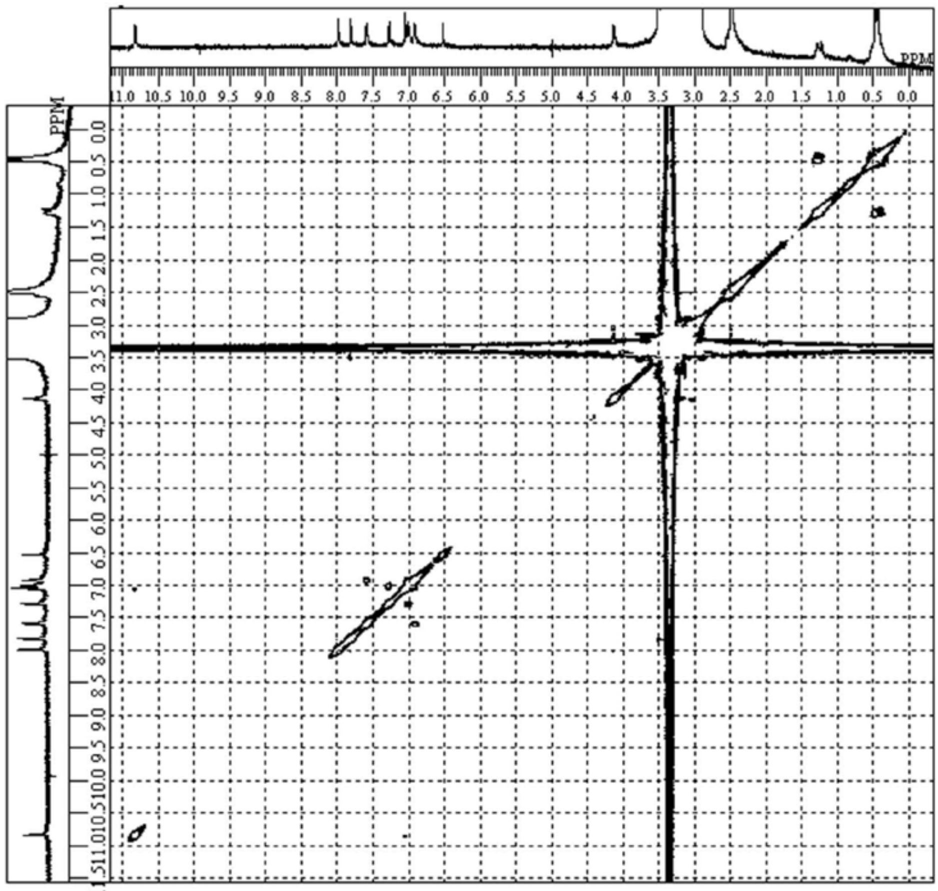


Figure S37. ^{1}H-^{1}H COSY of Cyclo(L-Ile-L-Trp) (**11**).

LC-MS data of cyclodipeptides 8-11

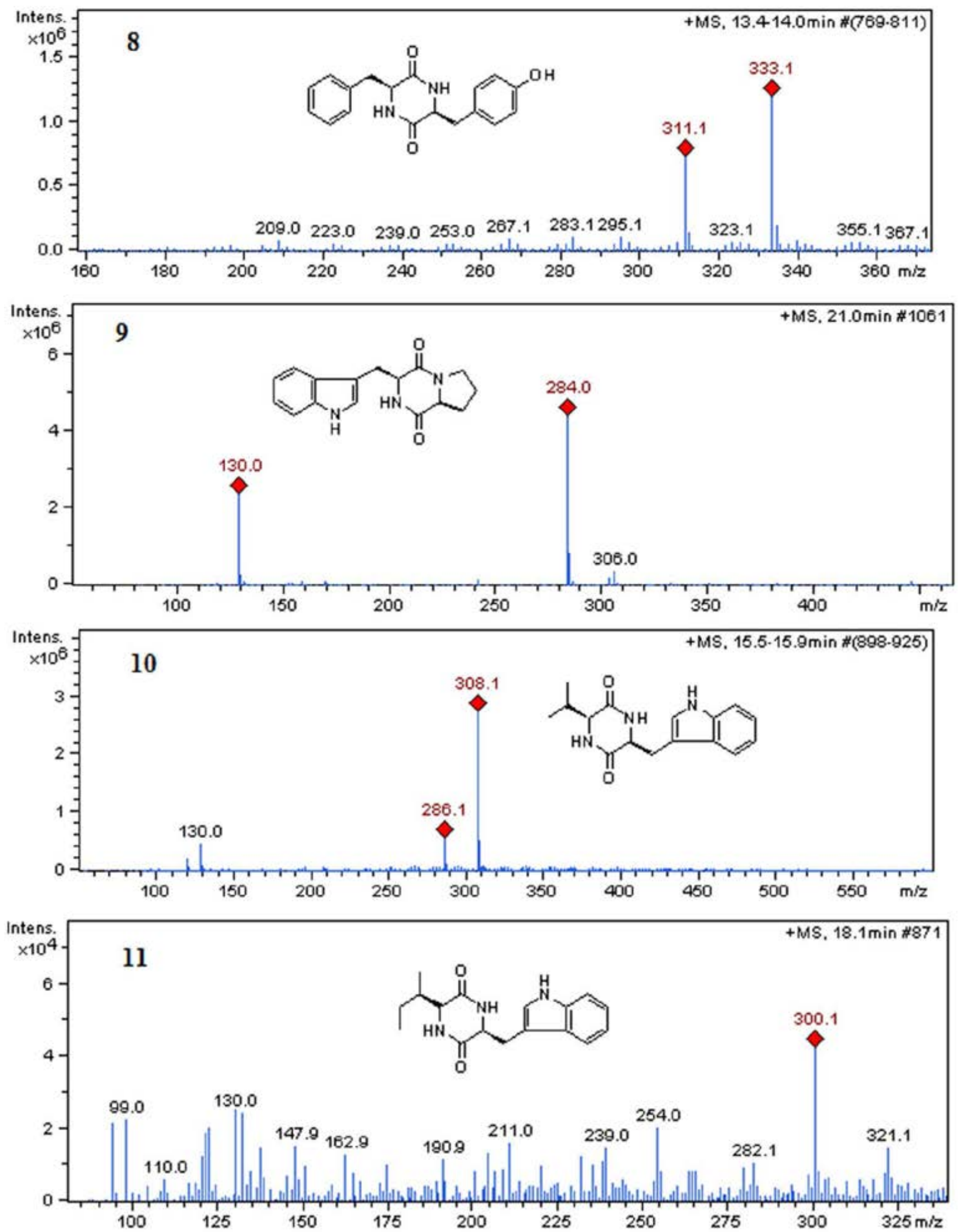


Figure S38. LC-MS data of cyclodipeptides **8**-**11**.

Liquid chromatography-mass spectrometry (LC-MS, Agilent 1100 series-Bruker Esquire 4000) analysis was performed on ODS column (TSK-Gel ODS-80Ts, 4.6 × 150 mm) with a mixture of H_2O and MeCN, both containing 0.1% acetic acid: 30-100% MeCN 30 min; 100% MeCN 10 min, 0.2 mL min^{-1}. Detected wavelength: 280 nm. Positive ESI.