Isolation and structure determination of bioactive peptides from actinobacteria based on genome mining

Thesis

Isolation and structure determination of bioactive peptides from actinobacteria based on genome mining

June, 2020

Graduate School of Science and Technology, Bioscience Course,
Shizuoka University

Issara Kaweewan

CONTENTS

ACKNOWLEDGEMENTS

I wish to express gratitude to Associate Professor Shinya Kodani of Graduate School of Science and Technology, Shizuoka University for his inspiring guidance, valuable advice, suggestion and encouragement throughout my Doctoral course.

I wish to thank Dr. Hisayuki Komaki of Biological Resource Center, National Institute of Technology and Evaluation (NBRC), Dr. Hikaru Hemmi of Food Research Institute, NARO, Dr. Shigeyoshi Harada of AIDS Research Center, National Institute of Infectious Diseases, Associate Professor Takeshi Hosaka of Institute for Biomedical Sciences, Shinshu University, Kanata Hoshino of Graduate School of Science and Technology, Shinshu University, Associate Professor Takanori Oyoshi of Academic Institute, Shizuoka University, and Gouchi Isokawa of Graduate School of Integrated Science and Technology, Shizuoka University, for their collaborations and assistances.

I appreciate supports from Graduate school of Science and Technology and Faculty of Agriculture, Shizuoka University. Special thanks to Environmental Leaders Program (ELSU), Shizuoka University and Mitsubishi Corporation International Scholarship for financial support throughout the Doctoral course. Thanks to Sasakawa Scientific Research Grant from The Japan Science Society (grant number 2018-3001).

I would like to thank all members in the laboratory of Applied Microbiology, Faculty of Agriculture, Shizuoka University for their sincere assistances and friendships. Thanks to all staffs of Graduate School of Science and Technology, Shizuoka University for their helps and guidance.

Finally, I am very grateful to my parents, Rungrueng Kaweewan and Jeerapun Kaweewan, and my sister, Rungrujee Kaweewan, for their understanding and supporting my daily life. I wish to thank Jirayu Boonyakida and all my friends for their encouragement and moral support.

LIST OF ABBREVIATIONS

A	adenylation
antiSMASH	antibiotics and secondary metabolite analysis shell
ATCC	American type culture collection
BGCs	biosynthetic gene clusters
BLAST	basic local alignment search tool
C	condensation
CFU	colony-forming unit
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
DAD	diode array detector
D-MEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DQF-COSY	double quantum filtered correlated spectroscopy
E	epimerization
E-MEM	Eagle's minimal essential medium
ESI-MS	electrospray ionization mass spectrometry
ESI-TOF-MS	electrospray ionization time-of-flight mass spectrometry
FBS	fetal bovine serum
GBAP	gelatinase biosynthesis-activating pheromone
HCl	hydrochloric acid
HI	hydriodic acid
HIV	human immunodeficiency virus
HMBC	heteronuclear multiple-bond correlation spectroscopy
HPLC	high performance liquid chromatography
HSQC	heteronuclear single-quantum correlation spectroscopy
IC_{50}	the half maximal inhibitory concentration
ISP2	international *Streptomyces* project-2
JCM	Japan collection of microorganisms
L-FDLA	*N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide
MeCN	acetonitrile
MeOH	methanol

MHB	Mueller-Hinton broth
MICs	minimum inhibitory concentrations
MS	mass spectrometry
$NaHCO_3$	sodium bicarbonate
NBRC	NITE biological resource center
NMR	nuclear magnetic resonance spectroscopy
NOESY	nuclear overhauser effect spectroscopy
NRPs	nonribosomal peptides
NRPSs	nonribosomal peptide synthetases
ODS	octadecyl silica
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PDB	protein data bank
PPI	protein−protein interactions
RiPPs	ribosomally synthesized and post-translationally modified peptides
RODEO	rapid orf description and evaluation online
rRNA	ribosomal ribonucleic acid
TE	thioesterase
TFA	trifluoroacetic acid
TOCSY	total correlation spectroscopy
TPPI	time proportional phase incrementation
UV	ultraviolet
WT	wild type

LIST OF FIGURES

Chapter I: Isolation and structure determination of a new lasso peptide, specialicin, from *Streptomyces specialis* based on genome mining

LIST OF TABLES

General introduction

1. Actinomycetes

Actinomycetes are Gram-positive bacteria containing high guanine and cytosine content in DNA which belong to phylum *Actinobacteria*, one of the largest taxonomic units among *Bacteria* domain.[1] They are fungus-like bacteria found in both terrestrial environments, mainly in soil, and aquatic environments including freshwater and marine. As actinomycetes have adapted to wide range of environments, some of them can survive in extreme conditions, such as thermophilic, acidophilic, halophilic and endophytic actinobacteria.[2] Like filamentous fungi, actinomycetes exhibit a complex life cycle comprising of many type of cells including spores, vegetative and aerial mycelium. Like all bacteria, actinobacterial cells are enclosed by peptidoglycan cell wall and their chromosomes are organized in nucleoid.[3]

Actinomycetes have been known as important bioresources for wide variety of secondary metabolites, such as antibacterials, antivirals, antitumors, immunosuppressives and enzyme inhibitors.[4] More than half of commercial antibiotics are produced by actinomycetes especially *Streptomyces* species. Actinomycetes existing in the soil play an important role in carbon cycle, decomposition of organic matters and humus formation.[2] Some actinomycetes reveal an antagonistic activity against plant pathogens and can be applied as biocontrol.[2-5]

The phylum *Actinobacteria* was classified based on morphological and chemical classification into 5 subclasses, 6 orders and 14 suborders. However, genome-based taxonomic classification using 16s rRNA sequence was recently applied to bacterial classification. Some taxonomic units have been reclassified based on 16S rRNA gene.[6] Most actinobacteria with medical significance are classified into subclass *Actinobacteridae*, and belong to the order *Actinomycetales*. The *Streptomycetaceae*, a family of *Actinobacteria*, includes the important genus *Streptomyces* which notable as original source of many antibiotics.[1, 3]

1.1 *Streptomyces*

Streptomyces, a type genus of the family *Streptomycetaceae*, is the largest genus of *Actinobacteria* that includes more than 500 species. Although streptomycetes widely distribute in terrestrial and aquatic environments, they have been reported to exist in other habitats such as plants, animals and humans.[7] Most streptomycetes produce vegetative hyphae (approximately 0.5-1.0 μm in diameter) with branches similar to fungi. The morphological differentiation occurs during the formation of hyphae so that a layer of hyphae can develop into chains of spores. Cell wall of streptomycetes mainly contain LL-diaminopimelic acid and glycine which is different from other actinomycetes.[7-9]

Streptomyces spp. have received much attention because of their ability to produce a wide variety of potential antibiotics. Historically, streptomycin, an antibiotic from *Streptomyces griseus*, was discovered in 1943 by American biochemists Selman Waksman, Albert Schatz, and Elizabeth Bugie. Since then, streptomycin has been used in the treatment of numerous infectious diseases including tuberculosis.[10] In 1950, the first antifungal, called nystatin, was isolated from *S. noursei* by Rachel Fuller Brown and Elizabeth Lee Hazen. It has been used for *Candida* infections involving skin, mouth, esophagus and vagina infections.[11] *S. coelicolor* A3 (2) is one of important strains which has been used as a model actinobacterium for studying complex regulating system of antibiotic production and morphological differentiation. The sequenced genome of *S. coelicolor* A3 (2) revealed the largest number of genes including more than 20 secondary metabolites biosynthetic gene clusters.[12]

2. Secondary metabolites

In recent years, microbial resources, e.g. terrestrial, marine, and endophytic microbes, have been reported to be rich sources for novel secondary metabolites with diverse biological functions. Microbial secondary metabolites, also known as natural products, are low molecular organic compounds which are not essential for growth, development and reproduction of microorganisms, however, some of them exhibit high potential for human health.

Microbial secondary metabolites are mostly produced by microorganisms during late growth phase within a limited range of culture conditions. The main producers belong to phylum *Actinobacteria*, including *Streptomyces* species and rare actinomycetes. The microbial secondary metabolites usually possess diverse chemical structures and biological activities, such as antimicrobials, antitumor, immunosuppressive, and enzyme inhibitory activities. Microbial secondary metabolites which exhibit any biological activity are defined as bioactive metabolites or “bioactive compounds”, distinguishing from other inactive secondary metabolites. The bioactive compounds are classified based on their bioactivities into two main groups, including antibiotics and other bioactive compounds.[13-15]

2.1 Antibiotics

Initially, secondary metabolites derived from microorganisms which exhibit any antimicrobial activities, e.g. antibacterial, antiviral and antifungal, are called “antibiotics”.[14] In the present, antibiotics refer to an antibacterial substance produced by microorganisms or others that is used in the treatment of bacterial infections by killing and/or inhibiting the growth with low concentration. In 1928, Alexander Fleming discovered penicillin, an antibacterial agent produced by *Penicillium* sp., which is clinical used to treat bacterial infections against

staphylococci and other Gram-positive pathogens.[16, 17] The discovery of penicillin resulted in a number of new antibiotics discovery in the period between 1940s and 1960s, known as the golden era of antibiotic discovery. During this period, important drugs such as streptomycin, tetracycline, chloramphenicol, rifamycin, vancomycin, and kanamycin were discovered. In 1970s, the rate of novel drug discovery suddenly dropped and several pathogens have acquired resistance to commercial antibiotics.[16, 18] Since then, researchers around the world have been trying to find new sources of natural products to overcome antibiotic resistance problems.

2.2 Other bioactive compounds

Besides antibiotic activity, microbial secondary metabolites have been reported to exhibit other bioactivities such as antiviral, antitumor, immunosuppressive and enzyme inhibitory activity. Many bioactive compounds have contributed to the development of antiviral drugs.[19] For examples, formycin, an antiviral agent isolated from *Streptomyces* sp. S-685, was found to be active against poliovirus and influenza virus as well as inhibited the multiplication of two myxoviruses, WSN and VSV.[20] Pepstatin was isolated from *Streptomyces* sp. as an aspartic proteinase inhibitor which inhibit HIV-l proteinase.[21] Siastatins A and B were isolated from *S. verticillus* var. *quantum* as potent sialidase inhibitors. Siastatin B analogs showed antiviral activity by inhibiting influenza virus neuraminidase (NA) *in vitro*.[22, 23]

Many bioactive compounds have been utilized as anticancer agents. For instance, actinomycin D, also known as dactinomycin, is an effective anticancer originally isolated from *S. parvulus*. It has been used in the treatment of Ewing's sarcoma, Wilms' tumor, gestational trophoblastic tumor and rhabdomyosarcoma for infants.[24] Some compounds possess various bioactivities. The compound named FK506, a clinically used immunosuppressive agent, exhibits not only immunosuppressive activity, but also antifungal, anti-inflammatory and neuroprotective activities.[25] Rapamycin exhibits immunosuppressive activity, along with anticancer, neuroprotective and neurogenerative properties.[26, 27]

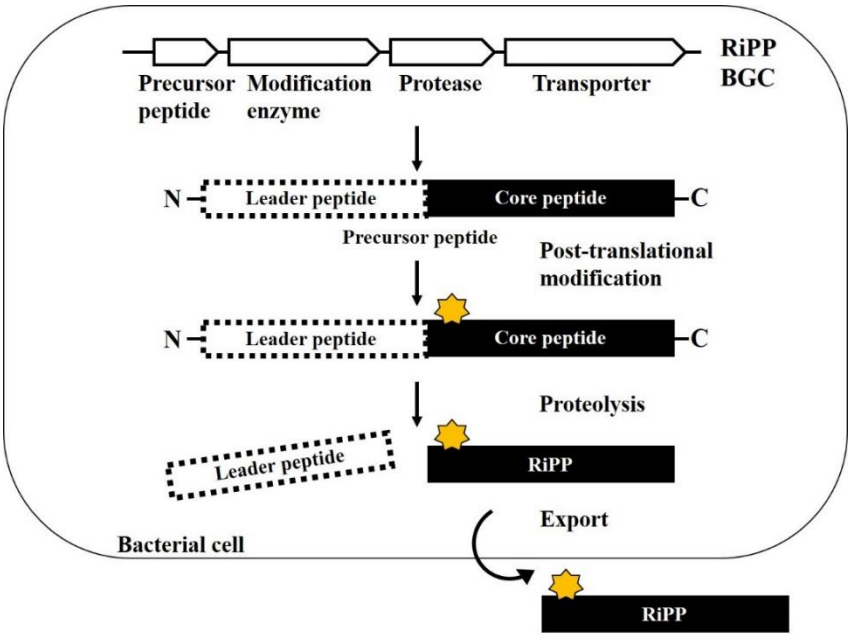
3. Peptides

Peptides are polymers of amino acids which can have almost infinite variety of sequence. Among peptides, a group of peptide which shows biological functions and provides positive influence on human health are defined as "bioactive peptides".[28] Bioactive peptides have received much attention because many of them play an importance role in the pharmaceutical industry. For example, insulin is a peptide hormone which controls blood sugar level used in the treatment of diabetes.[29] Liraglutide is a glucagon-like peptide-1 receptor

agonist used to treat diabetes mellitus type II and obesity.[30, 31] Moreover, peptide antibiotics of bacterial origins, for example vancomycin, chloramphenicol and daptomycin have been developed as commercial drugs for several years. Bioactive peptides can be classified based on their function into hormones, neuropeptides, alkaloids, etc. Bacterial peptides can be classified according to their biosynthetic pathways into two major groups, including ribosomally synthesized and post-translationally modified peptides (RiPPs) and peptides synthesized by nonribosomal peptide synthetases (NRPSs).[30]

3.1 Ribosomally synthesized and post-translationally modified peptides (RiPPs)

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a major class of bioactive peptide natural products which have been recognized as potential drug seed compounds.[30] Initially, RiPPs are ribosomally biosynthesized as long length precursor peptides, approximately 20-110 amino acids, encoded by structural peptide gene. A precursor peptide normally consists of two peptide regions, including leader peptide and core peptide.[32] The leader peptide involves in recognition by post-translational modification enzymes. The core peptide is modified by post-translational modification enzymes as a final product. In a late step of maturation process, the leader peptide is removed from the modified core peptide by proteolytic cleavage and the matured peptide is finally exported to outside of the cell. The biosynthesis of RiPPs through extensive post-translational modifications results in diverse chemical structures and bioactivities. RiPPs are divided into more than 20 subclasses according to their structural motifs and biosynthetic pathways, such as lanthipeptides, cyanobactins, thiopeptides and lasso peptides.[32-34]

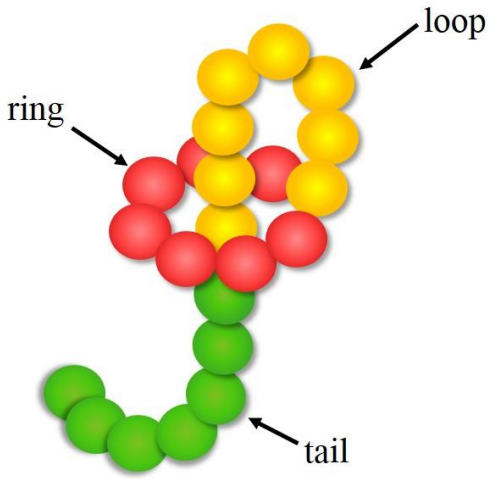


RiPP biosynthetic pathway

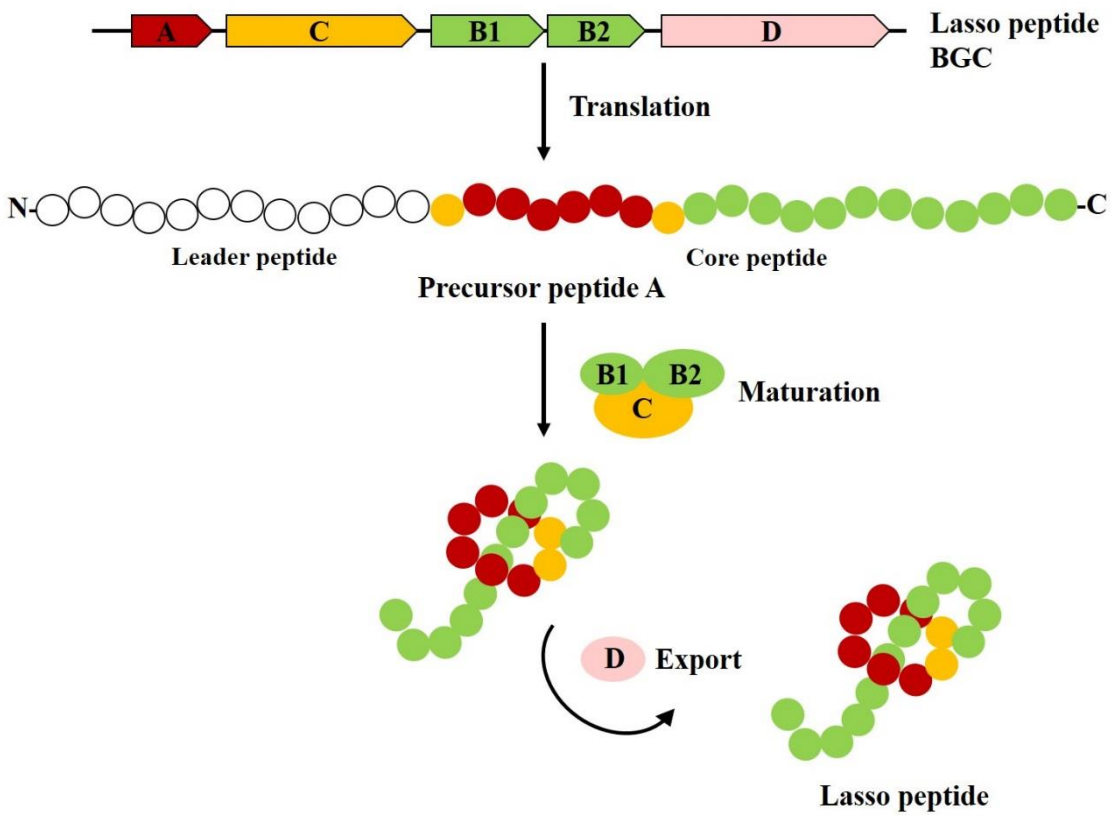
3.1.1 Lasso peptides

Lasso peptides are a class of ribosomally synthesized and post-translationally modified peptides or RiPPs that are characterized by a unique knotted topology. Lasso peptides usually consist of *N*-terminal macrolactam ring and linear *C*-terminal peptide tail threaded through the ring. The macrolactam ring comprises of 7-9 amino acids, formed by an isopeptide bond between *N*-terminus amino acid and side chain of 7^{th}-9^{th} aspartate or glutamate.[35-37] The lasso topology is maintained by the presence of disulfide bridges and/or steric interactions between ring and side chain of specific residues, called the plugs.[35] In general, lasso peptides are classified into 3 classes according to a number of disulfide bonds. Class I lasso peptides (e.g. sviceucin,[38] aborycin,[39] siamycins I and II[40]) are a group of lasso peptides containing two disulfide bonds. Class II lasso peptides (e.g. microcin J25,[41] capistruin,[42] propeptin,[43] chaxapeptin[44] and RES-701-1[45]) are devoid of disulfide bond. Class III lasso peptide, for example BI-32169,[46] contains one disulfide bond.[36, 37]

The archetype lasso peptide, microcin J25 (MccJ25), was originally isolated from *Escherichia coli* in 1992.[41] The biosynthetic gene cluster of MccJ25 was reported to contain four genes including *mcjA*, *mcjB*, *mcjC* and *mcjD*. MccJ25 production was accomplished by heterologous expression.[47] *In vitro* experiments revealed that McjA encodes a precursor peptide which is further processed by the ATP-dependent cysteine protease, McjB, cleaving off the leader peptide from the core peptide. McjC is characterized as a macrolactam synthetase which is responsible for isopeptide bond formation. After modifications, the mature MccJ25 is exported by ABC-transporter, McjD.[35, 41, 47, 48]



Lasso peptide structure



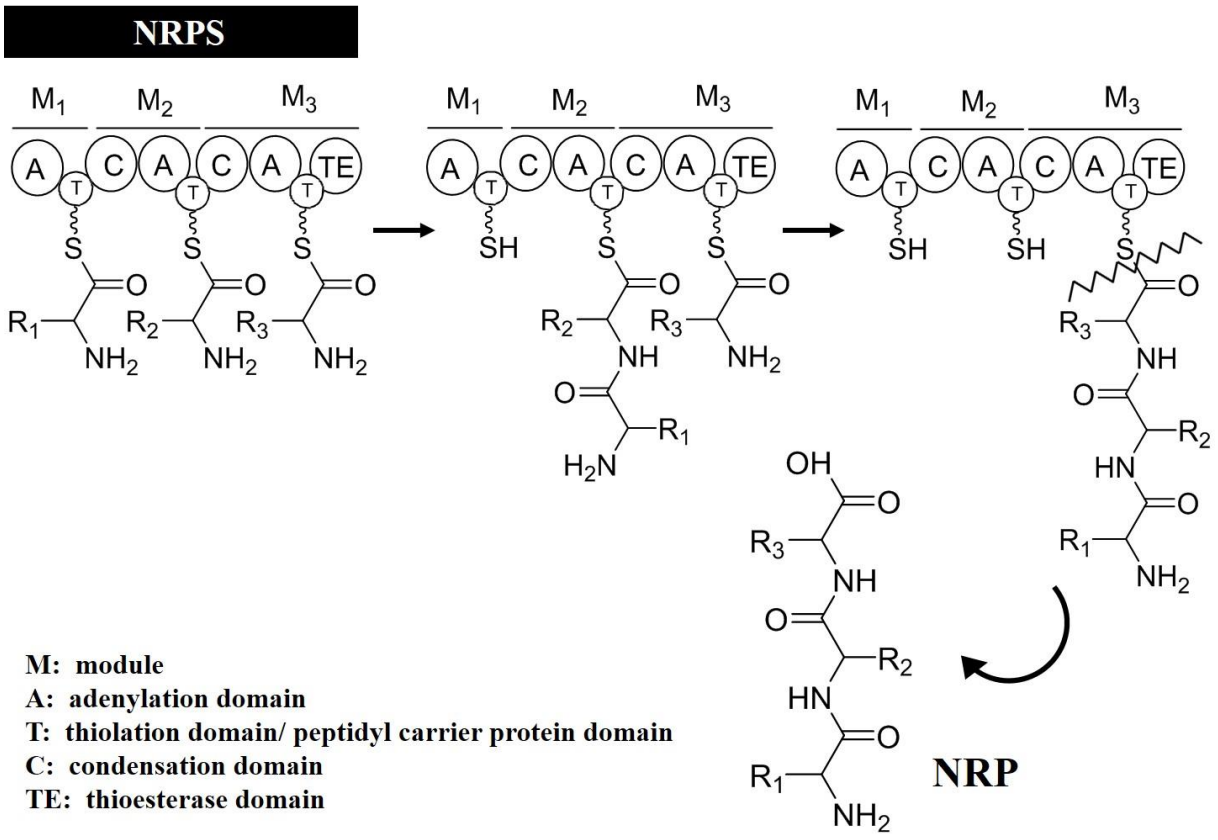
Lasso peptide biosynthetic pathway

Lasso peptides possess a wide range of biological activities such as enzyme inhibitors, receptor antagonists, anti-HIV and antibacterial activities.[32] Capistruin, a lasso peptide derived from *Burkholderia thailandensis*, displays antibacterial activities against *Burkholderia* and *Pseudomonas* strains.[49] Siamycin I, also known as MS-271, is an inhibitor of the myosin light chain kinase and anti-HIV agent.[50] Sungsanpin, a lasso peptide from deep sea streptomycete, and chaxapeptin, a lasso peptide from extremotolerant streptomycete, showed inhibitory activity against cell invasion of human lung cancer cell line A549.[44, 51]

3.2 Nonribosomal peptides (NRPs)

In the past, research on peptide antibiotics was mostly focused on peptides of nonribosomal origin called nonribosomal peptides (NRPs). Nonribosomal peptides are a class of peptide natural products which are biosynthesized by nonribosomal peptide synthetases (NRPSs). The NRPSs are defined as large multienzyme machineries that assemble peptides with diverse properties such as siderophores, pigments, antibacterials, antitumor and immunosuppressants. NRP biosynthesis generally occurs in several steps, including the processes of chain initiation, elongation and termination on the assembly line as well as post translational modification by tailoring enzymes, resulting in the mature peptide scaffolds.[52-55]

The module in an NRPS is responsible for the incorporation of one specific amino acid into the final product like building blocks. Each module normally consists of a set of three domains, including adenylation (A) domain, peptidyl carrier protein (PCP) domain and condensation (C) domain, for substrate recognition, transport to the respective catalytic centers and formation of the peptide bond, respectively. After the linear peptide chain has reached the final module, the peptide is released by the function of a thioesterase (TE) domain which exists at the end of the NRPS assembly line. Some modules contain epimerization (E) domain to carry out epimerization of the C_{α}-carbon of the PCP-tethered aminoacyl substrate to afford a D- or L-configuration equilibrium.[52] In addition, optional modifications of peptides such as halogenation, glycosylation, and hydroxylation were also found during the steps of NRP biosynthesis.[55, 56] In contrast to peptides synthesized by ribosomal origin, the NRPs contain not only 20 normal amino acids, but also hundreds of unusual modified amino acids. Therefore, the NRP chemical structures reveal high diversity and complexity which results in a wide variety of biological activities.[57-59] So far, several commercial antibiotics (e.g. chloramphenicol,[60] daptomycin,[61] teicoplanin[62] and vancomycin[63, 64]), antitumors (e.g. actinomycin D[65] and bleomycin A2, B2[66]) and immunosuppressants (cyclosporine A[67]) have been reported to biosynthesized by NRPS.[52]



NRP biosynthetic pathway

4. Genome mining approach

Genome mining approach has become a useful tool for natural product research, and could lead to the discovery of new drugs. This approach is a genetic screening strategy to predict potential biosynthetic gene clusters (BGCs) of natural products in the whole genome sequences accumulated in genome sequence database. In general, peptides in the same class are produced by biosynthetic pathways which contain highly homologous essential gene set and optional modification enzymes. Basically, differential analysis on the biosynthetic genes leads to find new BGCs which contain different modification enzyme coding genes. Therefore, genes that encode such enzymes are suitable targets for genome mining. Among them, tailoring enzyme coding genes, for example halogenase genes, are successfully used to identify several novel BGCs of natural products which belong to broad range of structural classes.[68]

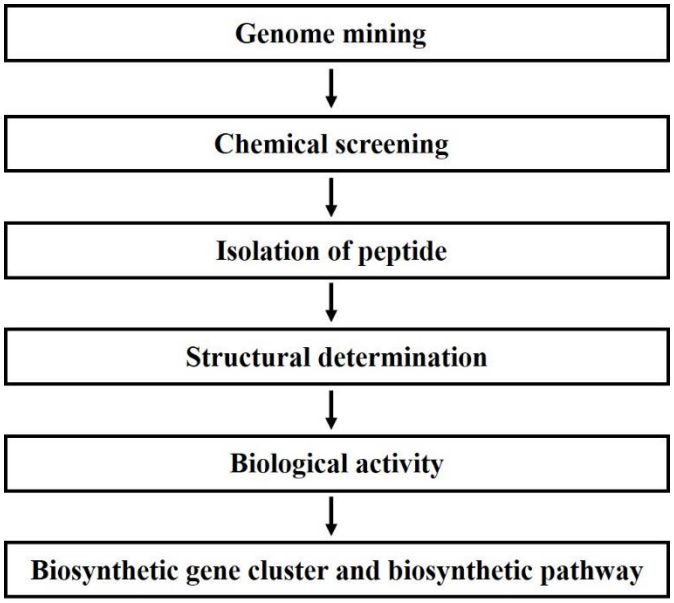
In case of peptides produced by biosynthetic system like assembly line, for example polyketide and nonribosomal peptides, the presence or absence of domains with tailoring enzyme activities in the individual modules can be used to predict the way that building block is modified during the process of incorporation of each amino acid.[69, 70] Ribosomally synthesized and post-translationally modified peptides or RiPPs is gene-encoded peptides whose sequence can be used to predict novelty and structural elucidation.[32] The precursor peptide contains leader peptide and core peptide regions involved in binding and modification of tailoring enzyme, respectively. RiPP tailoring enzymes can be selective for leader sequence but can process different core sequences. Thus, genome mining of genes encoded tailoring enzyme provides a direct route to discover new natural products.[71, 72]

In 2011, the genome mining software named antiSMASH was initially released to identify secondary metabolite BGCs in bacterial and fungal genome sequences.[73] AntiSMASH version 5 which contains many improvements was developed in 2019. It is currently used by academic and industrial scientists for natural products research and other related fields, such as metagenomics and environmental biology.[74] In addition, a new genome mining tool, Rapid ORF Description and Evaluation Online or RODEO, was developed to analyze BGCs of RiPPs and predict RiPP precursor peptides. Thousands of putative precursors and BGCs were identified based on RODEO system.[72]

5. Aim of the study

Bioactive compounds derived from actinobacteria have been utilized in pharmaceutical applications and scientific research for several years. Actinobacteria in *Streptomyces* genus are a major group that produces bioactive compounds with diverse biological functions. Recently, many bioactive peptides from *Streptomyces* spp. have been discovered by our group based on genome mining approach, such as achromosin,[75] globimycin[76] and curacomycin[77]. We reported bioactive peptides, along with its chemical structures, biological activities, and biosynthetic gene clusters.

Herein, this study introduces structures and biological activities of three new bioactive peptides from *Streptomyces* species. The study contains three chapters as following; Chapter I: Isolation and structure determination of a new lasso peptide, specialicin, from *Streptomyces specialis* based on genome mining, Chapter II: Isolation and structure determination of a new cytotoxic peptide, curacozole, from *Streptomyces curacoi* based on genome mining, and Chapter III: Isolation and structure determination of a new antibacterial peptide, pentaminomycin C, from *Streptomyces cacaoi* subsp. *cacaoi*.



Experimental design flowchart

Chapter I

Isolation and structure determination of a new lasso peptide, specialicin, from *Streptomyces specialis* based on genome mining

1.1 Introduction

Lasso peptides are ribosomally synthesized and post-translationally modified peptides (RiPPs) which are characterized by the presence of unique lasso topology in the molecule.[36, 37] The basic structure of lasso peptides consist of *N*-terminal macrolactam ring comprised of seven to nine residues and linear *C*-terminal peptide tail that threads through the ring.[35] The lasso structure is maintained by an isopeptide bond formed between amino residue of *N*-terminus of peptide and β- or γ- carboxyl residue of side chain of aspartate or glutamate. This lasso structure makes lasso peptide stable against protease and gives a potential to utilize as pharmaceutical reagents. Moreover, many of lasso peptides have been reported to exhibit antibacterial activity mainly against Gram-positive bacteria. For example, achromosin, a lasso peptide isolated from *Streptomyces achromogenes*, showed antibacterial activity against *Micrococcus luteus*.[75]

There are four essential genes that involved in lasso peptide biosynthesis, including gene A encoding lasso peptide precursor, genes B and C encoding maturation enzymes, and gene D encoding an ABC transporter.[35, 78] Thus, genome mining for lasso peptides has focused on these four genes in lasso peptide biosynthetic gene cluster (BGC). Based on the combination of genome mining and heterologous expression, a number of new lasso peptides have been discovered.[35] For example, a lasso peptide sviceucin was originally discovered from *Streptomyces sviceus* DSM 924^{T} by genome mining and heterologously expressed in *S. coelicolor*. Complete structural characterization indicated that sviceucin contains two disulfide bridges and belongs to class I lasso peptide. Sviceucin displayed antimicrobial activity against Gram-positive bacteria and inhibition of *fsr* quorum sensing in *Enterococcus faecalis*.[38] Moreover, the BGC of a new lasso peptide chaxapeptin was found in the genome of *S. leeuwenhoekii* strain C58 by genome mining. Chaxapeptin was isolated and characterized as class II lasso peptide which possesses inhibitory activity against human lung cancer cell invasion.[44] Lasso peptides siamycins I and II were isolated from soil actinobacterium, *Streptomyces* sp. AA3861 and *Streptomyces* sp. AA6532, respectively. They exhibited antibacterial activity against Gram-positive bacteria and anti-HIV activity.[50] Siamycin I, also known as MS-271, was further reported as a novel inhibitor of calmodulin-activated myosin light chain kinase.[79] Moreover, Siamycin I was also found to be an inhibitor of *fsr* quorum

sensing inhibited both gelatinase production and gelatinase biosynthesis-activating pheromone (GBAP) production in *Enterococcus faecalis*.[80] Siamycin I (MS-271) biosynthetic gene cluster (BGC) was recently identified in the genome sequence of *Streptomyces* sp. M-271 and function of genes encoding in the BGC were characterized.[81] Based on this background, BGC of a new lasso peptide was found in the genome sequence of *Streptomyces specialis* GW41-1564^T by genome mining. A lasso peptide named specialicin was isolated from the extract of *S. specialis*. Structure elucidation by the combination of electrospray ionization mass spectrometry (ESI-MS) and nuclear magnetic resonance (NMR) analyses was performed on the peptide. The BGC for specialicin was proposed in the genome sequence of *S. specialis*. Specialicin showed biological activities including antibacterial activity and moderate anti-HIV activity against HIV-1 NL4-3. The isolation and structure determination of a new lasso peptide, specialicin, were described in the chapter I.

1.2 Materials and Methods

1.2.1 Bacterial strains

Streptomyces specialis JCM 16611^T (= *S. specialis* GW41-1564^T) was obtained from Japan Collection of Microorganisms (JCM) culture collection, Japan. Bacterial strains including *Escherichia coli* NBRC 102203^T, *Pseudomonas aeruginosa* NBRC 12689^T, *Micrococcus luteus* NBRC 3333^T, *Staphylococcus aureus* NBRC 100910^T and *Bacillus subtilis* NBRC 13719^T were obtained from NITE Biological Resource Center (NBRC) culture collection, Japan.

1.2.2 Isolation of peptide

Streptomyces specialis JCM 16611^T was cultured in 4 L of International *Streptomyces* Project-2 Medium or ISP2 agar medium[82] and incubated at 30°C for 10 days. Bacterial cells on the agar medium surface were harvested using stainless steel spatula. For extraction, 200 ml of methanol (MeOH) was added to the harvested cells, followed by filtration using filter paper (Whatman No. 1, GE Healthcare Life Sciences, Little Chalfont, UK). The MeOH extract was concentrated using rotary evaporator and subjected to open column chromatography (styrene-divinylbenzene polymer resin, CHP-20P, Mitsubishi Chemical Corp., Tokyo, Japan), eluted with 10%, 60%, and 100% MeOH, respectively. The 100% MeOH fraction was concentrated using rotary evaporator and injected to high performance liquid chromatography (HPLC). The peptide was isolated by HPLC using octadecyl silica (ODS) column (4.6×250 mm, Wakopak Handy ODS, WAKO). The UV detector was set at wave length of 220 nm. Isocratic elution mode, 45% acetonitrile (MeCN) containing 0.05% trifluoroacetic acid (TFA) at flow rate 1 ml/min, was used to isolate the peptide specialicin (Retention time; 14.2 min).

1.2.3 ESI-MS analysis

ESI-MS analysis was performed using a JEOL JMS-T100LP mass spectrometer. Poly ethylene glycol was used as an internal standard for accurate MS analysis.

1.2.4 NMR analysis

The peptide was dissolved in 500 µl of DMSO-d_6. NMR spectra were obtained on Bruker Avance 600 and Avance III HD 800 spectrometers with quadrature detection in the phase-sensitive mode by States-TPPI (time proportional phase incrementation) and in the echo-antiecho mode. One-dimensional (1D) ^{1}H, ^{13}C and DEPT-135 spectra were recorded at 25 °C with 15 ppm for proton and 239 ppm or 222 ppm for carbon. The two-dimensional (2D) double quantum filtered correlated spectroscopy (DQF-COSY) was recorded with 512 and 1024 complex points in t1 and t2 dimensions. The 2D homonuclear total correlation spectroscopy (TOCSY) with DIPSI2 mixing sequence, was recorded with mixing time of 80 ms, 512 and 1024 complex points in t1 and t2 dimensions. The 2D nuclear overhauser effect spectroscopy (NOESY) was recorded with mixing times of 60, 100 and 200 ms, 512 and 1024 complex points in t1 and t2 dimensions. The 2D ^{1}H-^{13}C heteronuclear single quantum correlation (HSQC) and heteronuclear multiple bond connectivity (HMBC) spectra were acquired at 25 °C in the echo-antiecho mode. The ^{1}H-^{13}C HSQC and HMBC spectra were recorded with 1024 and 512 complex points for 12 ppm in the ^{1}H dimension and 160 ppm or 222 ppm in the ^{13}C dimension, respectively, at a natural isotope abundance. All NMR spectra were processed using TOPSPIN 3.5 (Bruker). Peak-picking and assignment were performed with Sparky program. Before Fourier transformation, the shifted sinebell window function was applied to t1 and t2 dimensions. All ^{1}H and ^{13}C dimensions were referenced to DMSO-d_6 at 25 °C.

1.2.5 Structure calculation

Distance restraints were constructed from intensities of NOE cross peaks in 2D NOESY spectra with mixing times of 100 ms, which were classified into four distance categories including 2.9, 3.5, 5.0, and 6.0 Å. Pseudo-atom corrections were made for non-stereospecifically assigned methylene and methyl resonance.[83] An additional 0.5 Å were added to the upper bounds for methyl protons.[84] Backbone Φ dihedral angle restraints were evaluated from $^{3}J_{HN\alpha}$ values obtained from the high digital resolution 2D DQF-COSY spectrum and intraresidue and sequential NOEs. Backbone ϕ-angles were restrained to –120° ± 30° for $^{3}J_{HN\text{-}H\alpha} > 9$ Hz. The solution structure was calculated by simulated annealing protocol using distant and dihedral angle restraints with the program CNS version 1.1.[85] The Cys1-Asp9 isopeptide linkage was generated using a manual patch of the “protein-allhdg.top” CNS file

modified from the manual patch of the "protein1.0.top" XPLOR-NIH file.[86] Three hundred structures were calculated, of which the 15 structures of lowest energy were selected for structural analysis. The final 15 lowest-energy ensemble structures were analyzed by MOLMOL[87] and PROCHECK-NMR,[88] and graphics were created by MOLMOL. The atomic coordinate data was deposited in the Protein Data Bank (PDB ID: 6AK0).

1.2.6 Modified Marfey's analysis

Marfey's analysis was carried out in a sealed vacuum hydrolysis tube. The peptide (1.0 mg) was hydrolyzed with 500 µl of 6 N hydrochloric acid (HCl) containing 3% phenol at 166 °C for 25 min in order to recover tryptophan (Trp) according to the previous report.[89] The peptide (1.0 mg) was hydrolyzed with 500 µl of 6 N HCl containing 0.2% sodium azide at 110 °C for 24 h to detect cysteic acid by oxidation of cysteine residues according to the previous report.[90] After cooling down at room temperature, the hydrolysate was completely evaporated using rotary evaporator and freeze-drying under vacuum. The hydrolysate was resuspended in 200 µl of water, followed by adding 10 µl of the solution of *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA, Tokyo Chemical Industry Co., LTD, Tokyo, Japan) in acetone (10 µg/µl). The 100 µl of 1 M sodium bicarbonate ($NaHCO_3$) was added to the hydrolysate and the mixture was incubated at 80 °C for 3 min. The mixture was cooled down to room temperature and then neutralized with 50 µl of 2 N HCl. The 1 ml of 50% MeCN/water was added to the mixture before it was subjected to HPLC. Standard amino acids (1.0 mg) were derivertized with L-FDLA and D-FDLA in the same manner. Approximately 30 µl of FDLA-derivertive peptide was analyzed by HPLC (C18 column, 4.6 × 250 mm, Wakopak Handy ODS, Wako Pure Chemical Industries, Tokyo, Japan), compared to FDLA-derivertive standard amino acids. The DAD detector (MD-2018, JASCO, Tokyo, Japan) was used to detect FDLA-derivertives, accumulating data of absorbance from 220 nm to 420 nm. For standard amino acids including cysteic acid (Cys), aspartic acid (Asp), tyrosine (Tyr), serine (Ser), alanine (Ala), leucine (Leu), tryptophan (Trp) and phenylalanine (Phe), HPLC analyses were performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 70 min, increasing percentage of solvent B from 25% to 60% (HPLC condition 1). HPLC analysis for valine (Val) was performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05% TFA) with a linear gradient mode from 0 min to 70 min, increasing percentage of solvent B from 30% to 50% (HPLC condition 2).

1.2.7 Antibacterial activity

Antibacterial activity of peptide was assessed by minimum inhibitory concentrations (MICs) test according to the previous reports.[91, 92] Tetracycline was used as a positive control. The 96-well microplate containing peptide or tetracycline (50 µl/well) with concentration ranging from 64 to 0.0625 µg/ml was prepared by serial twofold dilution technique using Mueller-Hinton broth (MHB). Testing strains including *E. coli*, *P. aeruginosa*, *S. aureus*, *B. subtilis* and *M. luteus* were cultured in nutrient agar medium and incubated at 30 °C for 24 h. Testing bacteria was inoculated into 0.85% saline solution to give a final inoculum of 10^5 CFU/ml. The 5 µl of bacterial solution was transferred to 10 ml of MHB and then 50 µl of the MHB containing testing bacteria was added into each well. The 96-well microplates were incubated at 30 °C for 24 h. The MICs were determined at the lowest concentration that gave no visible growth of testing bacteria.

1.2.8 Anti-HIV assay

Anti-HIV assay was established in PM1/CCR5 cells[93] based on viability of cells that had been infected or not infected with $100TCID_{50}$ of an NL4-3 strain exposed to various concentrations of the peptide. PM1/CCR5 cells were incubated at 37 °C for 7 days. The 50% inhibitory concentration (IC_{50}) values were measured using Cell Counting Kit-8 assay (Dojindo Laboratories, Kumamoto, Japan). The assay was performed in triplicate.

1.3 Results and discussion

1.3.1 Genome mining for new peptide discovery

Lasso peptides siamycin I and II were previously isolated from *Streptomyces* sp. as anti-HIV compounds. Siamycin I, also known as MS-271, was reported as a new inhibitor of calmodulin-activated myosin light chain kinase and inhibitor of *fsr* quorum sensing which inhibited gelatinase production and gelatinase biosynthesis-activating pheromone (GBAP) production in *Enterococcus faecalis*. Recently, the BGC of MS-271 was identified as *msl* in the genome sequence of *Streptomyces* sp. M-271 by Feng et al.[79-81, 94] Genes encoding in the *msl* BGC were also characterized. For example, the precursor peptide gene, *mslA*, encoded 42 amino acid residues, including a leader peptide and a core peptide comprising of 21 amino acids. The genes, *mslB1*, *mslB2* and *mslC* were identified as essential genes required for biosynthesis of a lasso peptide MS-271. The proteins MslD1-D4 were deduced to be ABC transporters, and MslH was proposed to be epimerase enzyme.[81] Based on this data, we performed genome mining using amino acid sequence of *mslA* as a query sequence to find structural gene of siamycin I (MS-271). BLAST result showed possible structural genes of siamycin I (MS-271)

in several *Streptomyces* spp. including *S. mirabilis* OK461 (FONR01000003.1), *S. diastatochromogenes* ATCC 12309 (LRTP01000767.1), *S. phaepurpureus* DSM 40125^T, which were recently reclassified into *S. griseorubiginosus* (LMWK01000007.1)[95], and *S. griseorubiginosus* DSM 40469^T (LMWV010000015.1) as shown in Fig. 1.1. Structural genes consist of leader peptide region (letters without underline) and core peptide region (underlined letters) with cleavage site indicated by an arrow. Moreover, we found a possible new lasso peptide structural gene homologous to siamycin I (MS-271) structural gene in the genome sequence of *S. specialis* GW41-1564^T (FAXE01000003.1). Amino acid sequence in the core peptide region of a new lasso peptide, named specialicin, was predicted to be CLGVGSCVDFAGCGYAVVCFW, indicated by underlined letters in Fig. 1.1. By comparison between amino acid sequence of siamycin I (MS-271) and specialicin, residues Asn8 and Ile17 in siamycin I (MS-271) are substituted by Val8 and Val17 in specialicin, respectively.

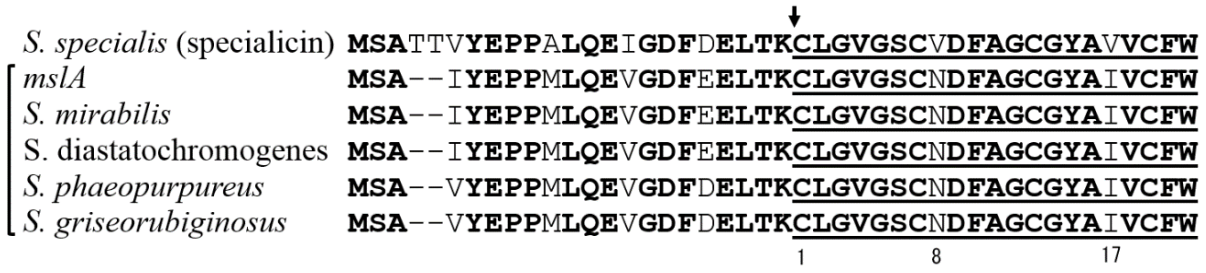


Figure 1.1 Amino acid sequences of specialicin precursor peptide and siamycin I (MS-271) precursor peptides found by BLAST search; Underline indicates core peptide region; bold letters indicate conserved amino acids; arrow indicates possible cleavage site; left parenthesis indicates siamycin I (MS-271) precursor peptides

1.3.2 Isolation of specialicin

Preliminary chemical investigation using HPLC and ESI-MS analyses revealed the presence of a new lasso peptide, named specialicin, in the crude extract of *S. specialis* JCM 16611^T. The peptide was detected by HPLC using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 30 min, increasing percentage of solvent B from 20% to 60% (Fig. 1.2). ESI-MS analysis on the expected lasso peptide was observed at *m/z* 1067.33 as shown in Fig. 1.3. Therefore, *S. specialis* JCM 16611^T was cultured in a large scale, approximately 4 L of ISP2 agar medium, in order to obtain enough amount of peptide for structure determination and biological activity test. A new lasso peptide specialicin (4.0 mg) was finally isolated by HPLC purification using isocratic elution mode, 45% MeCN/water containing 0.05% TFA at the flow rate of 1 ml/min. Specialicin was detected by HPLC at the retention time 14.2 min (Fig. 1.4)

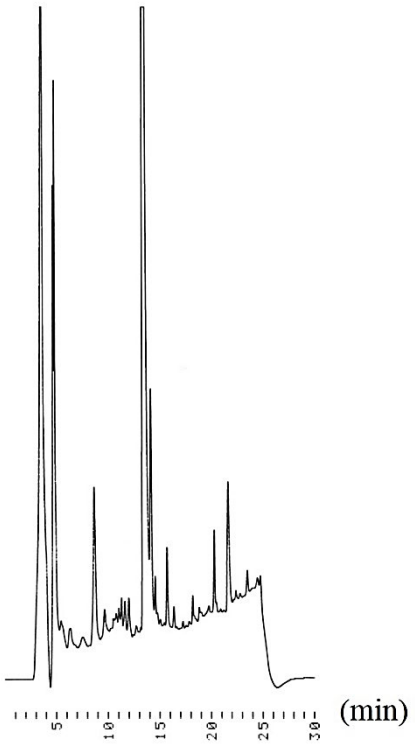


Figure 1.2 HPLC chromatogram of crude extract of *S. specialis* JCM 16611^{T}; The 100 µl of the extract was analyzed using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 30 min, increasing percentage of solvent B from 20% to 60%

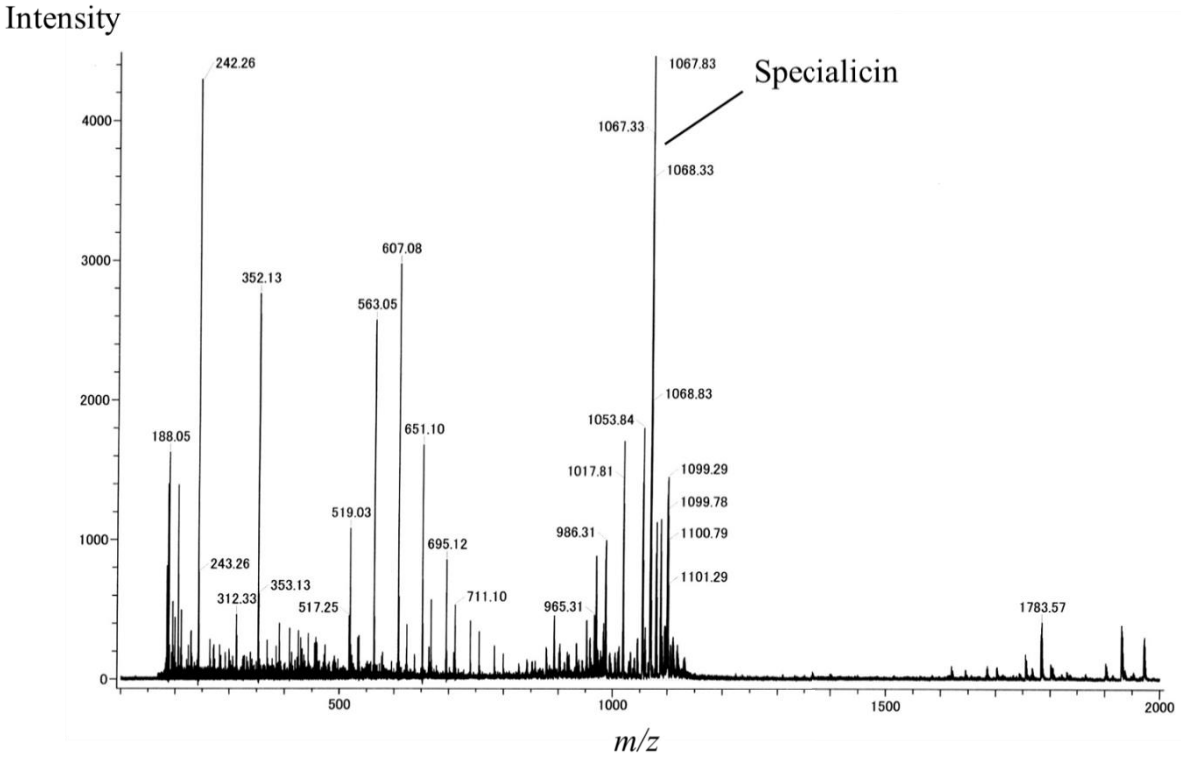


Figure 1.3 ESI-TOF-MS of specialicin; The $[M+2H]^{2+}$ was observed at *m/z* 1067.33

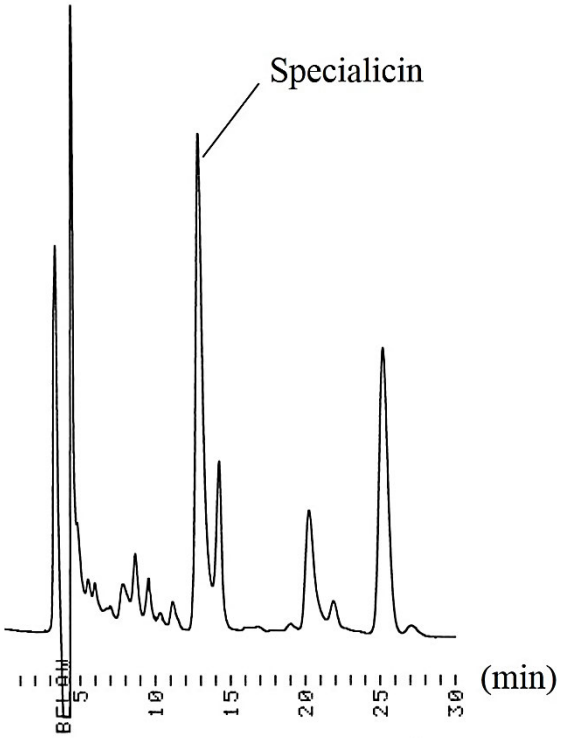


Figure 1.4 HPLC chromatogram of 100% MeOH extract of *S. specialis* JCM 16611^T; The 100 µl of the extract was analyzed using an isocratic elution mode, 45% MeCN/water containing 0.05% TFA, at flow rate of 1 ml/min; specialicin was detected by HPLC at the retention time 14.2 min

1.3.3 Structure determination of specialicin

The chemical structure of specialicin was determined by combination of ESI-MS and NMR analyses. Molecular formula of specialicin was established as $C_{97}H_{132}N_{22}O_{25}S_4$ by accurate ESI-MS analysis. The ion corresponding to $[M+2H]^{2+}$ was observed at *m/z* 1067.4372 (calculated *m/z* value was 1067.4386) (data not shown). Considering the molecular formula and core peptide sequence, specialicin was predicted to be a peptide containing two disulfide bonds. The planar chemical structure was completed by NMR analyses, including ^1H, ^{13}C, DEPT-135, DQF-COSY, TOCSY, NOESY, HMBC and HSQC (Fig. 1.5 - 1.12). The α-protons of amino acid residues at 3.02-5.23 ppm and amide protons at 6.62-9.34 ppm were observed in ^1H NMR spectrum (Fig. 1.5). The chemical shift of secondary amine proton (10.80 ppm) of indole moiety indicated the presence of Trp. The assignments of constituent 21 amino acids were determined by spin-system identification as shown in Fig. 1.13 and Table 1.1. The sequential analysis of amino acids was performed by NOESY correlations between α-proton and amino proton in adjacent amino acids indicated by double-end arrows in Fig. 1.13. A disulfide bond between

Cys7 and Cys19 was indicated since NOESY correlation was observed between β-protons of Cys7 and Cys19. The other disulfide bond was predicted to present between the remaining cysteine residues, Cys1 and Cys13, although NOESY correlation was not observed between β-protons of Cys1 and Cys13. Overall, specialicin was established to be a lasso peptide comprising of 21 amino acids with an isopeptide bond and two disulfide bonds in the molecule.

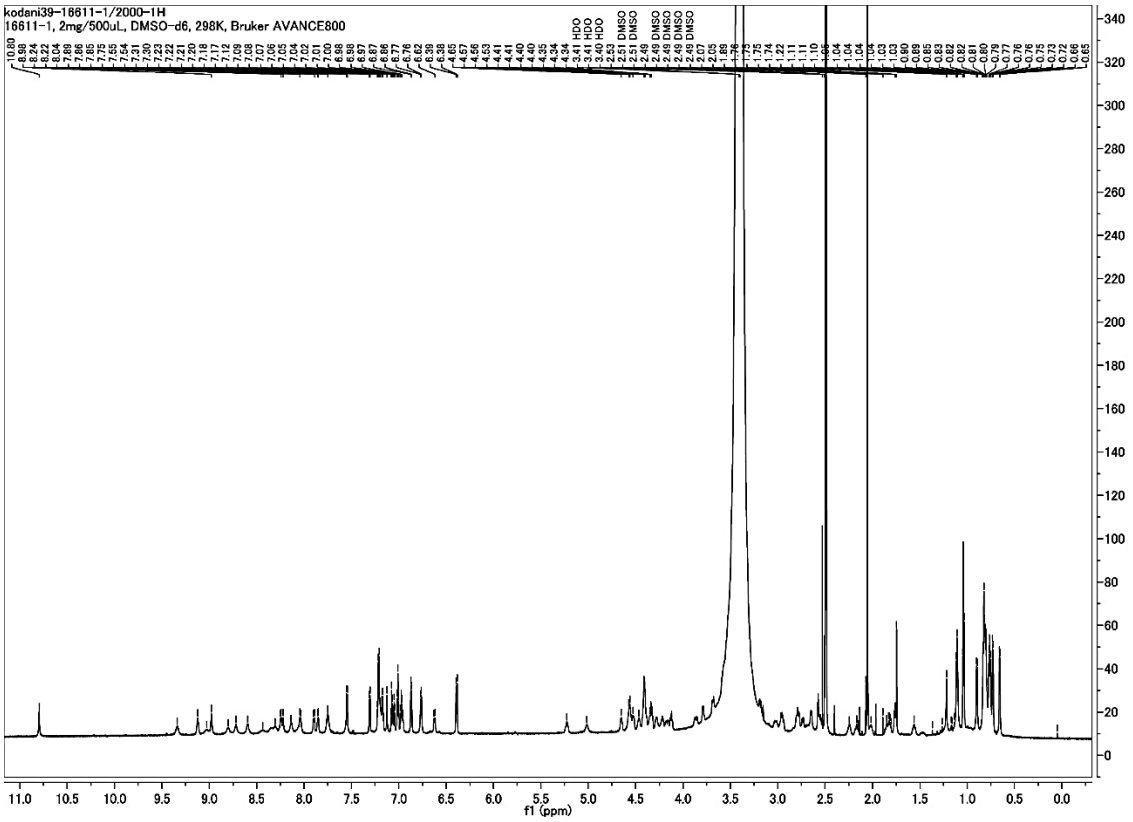


Figure 1.5 ^{1}H NMR spectrum of specialicin

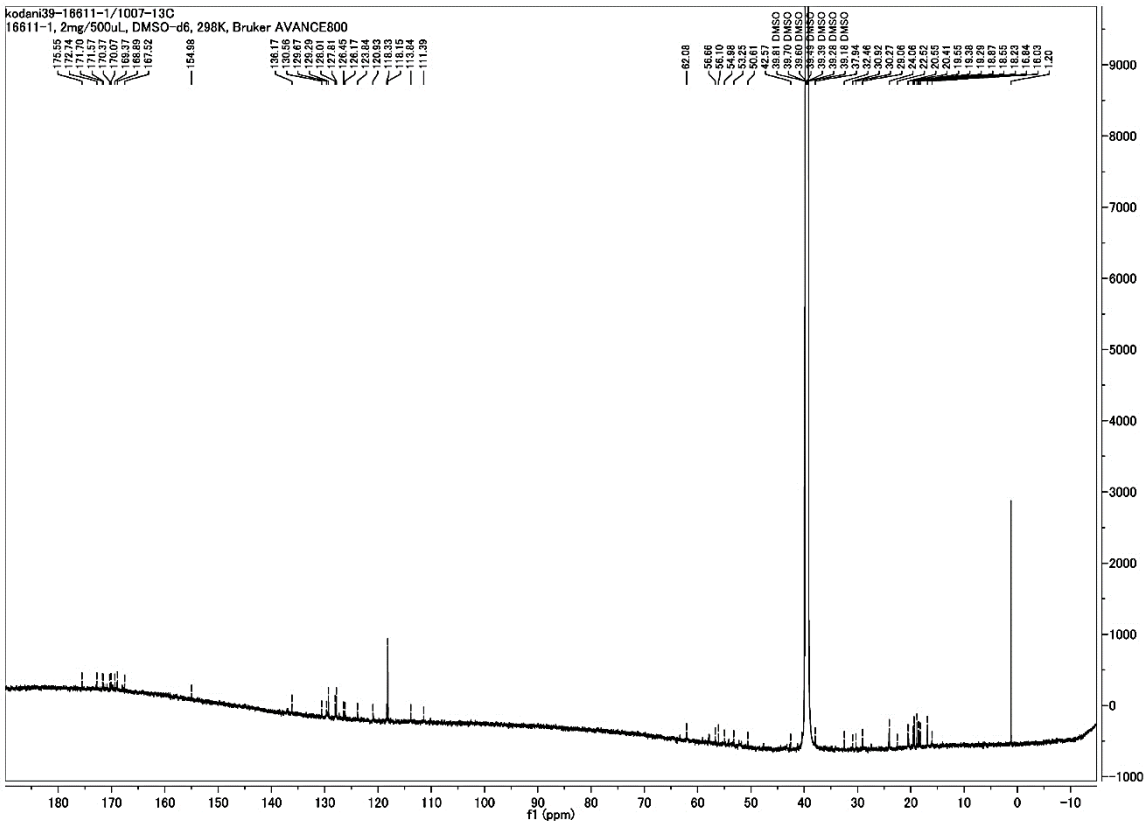


Figure 1.6 ^{13}C NMR spectrum of specialicin

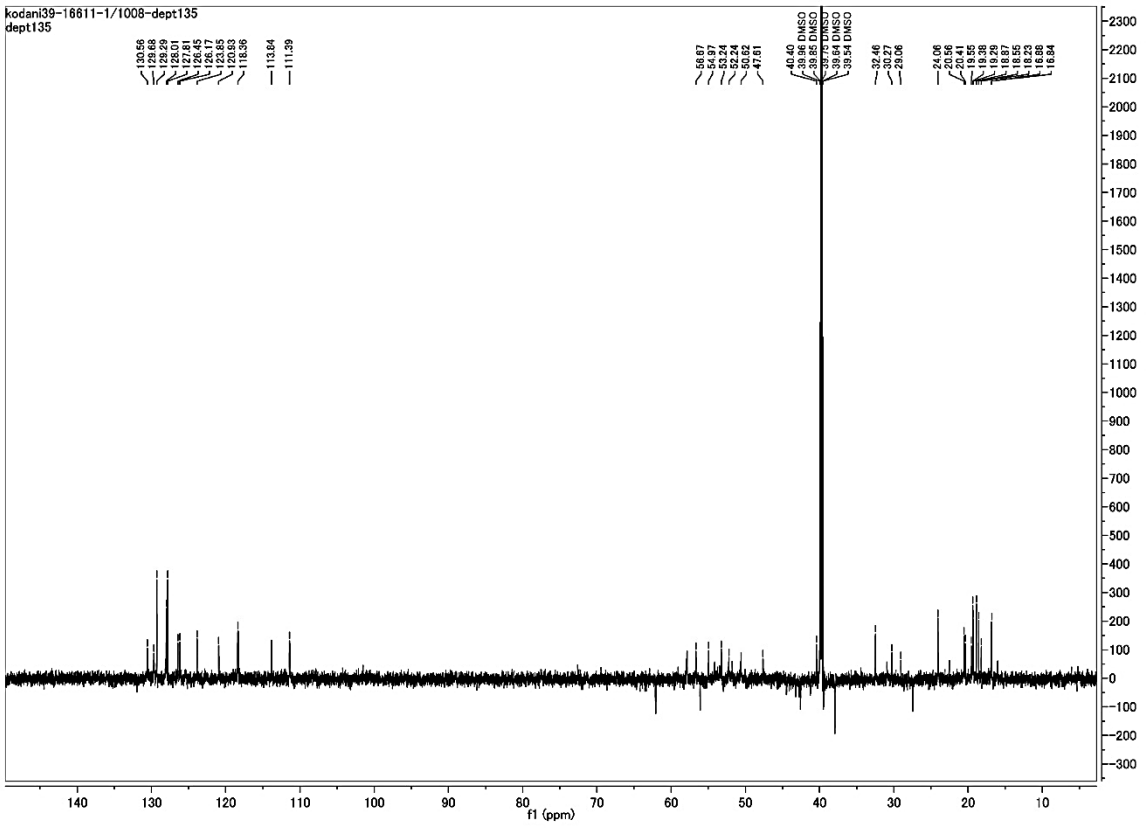


Figure 1.7 DEPT-135 spectrum of specialicin

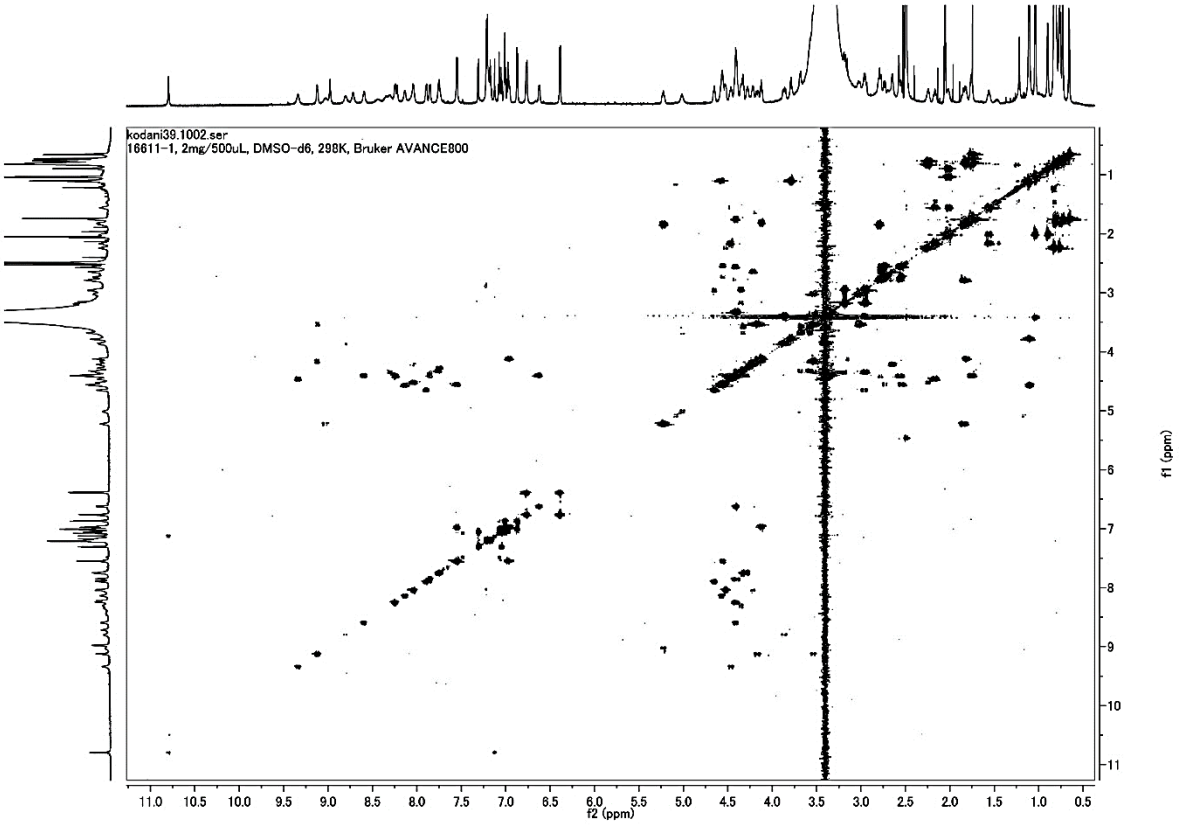


Figure 1.8 DQF-COSY spectrum of specialicin

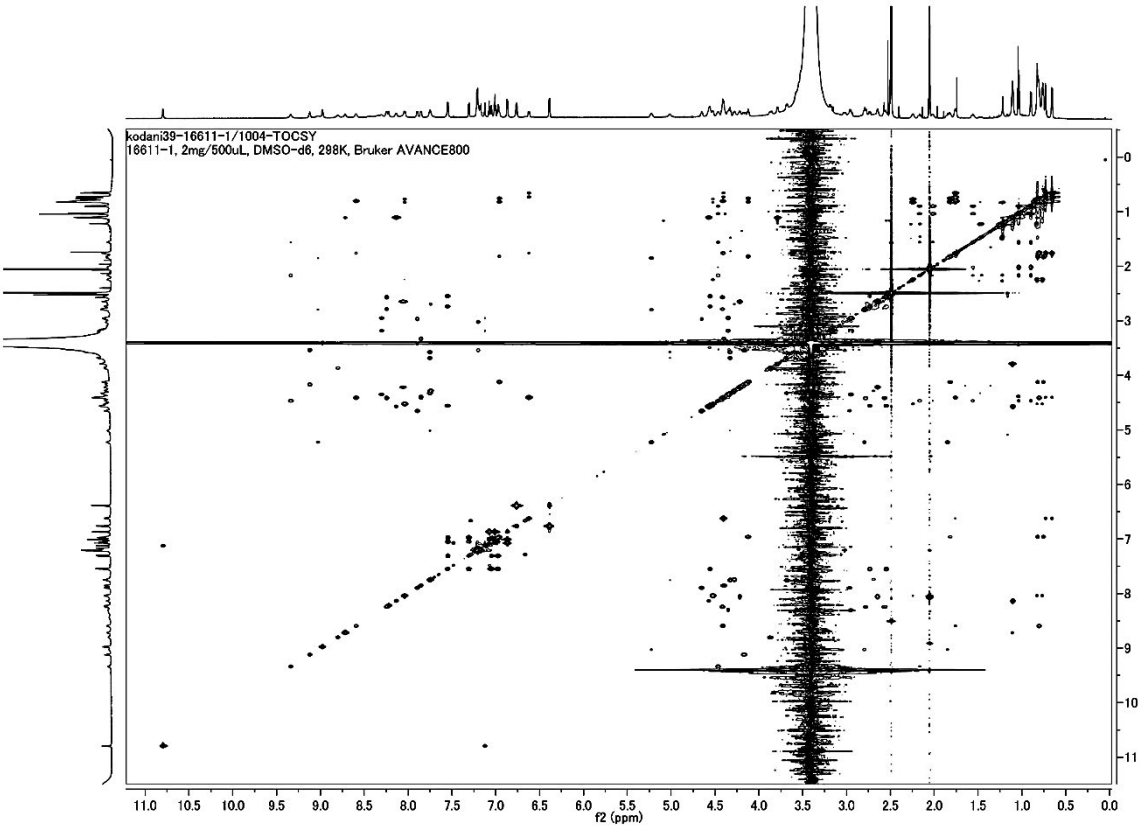


Figure 1.9 TOCSY spectrum of specialicin

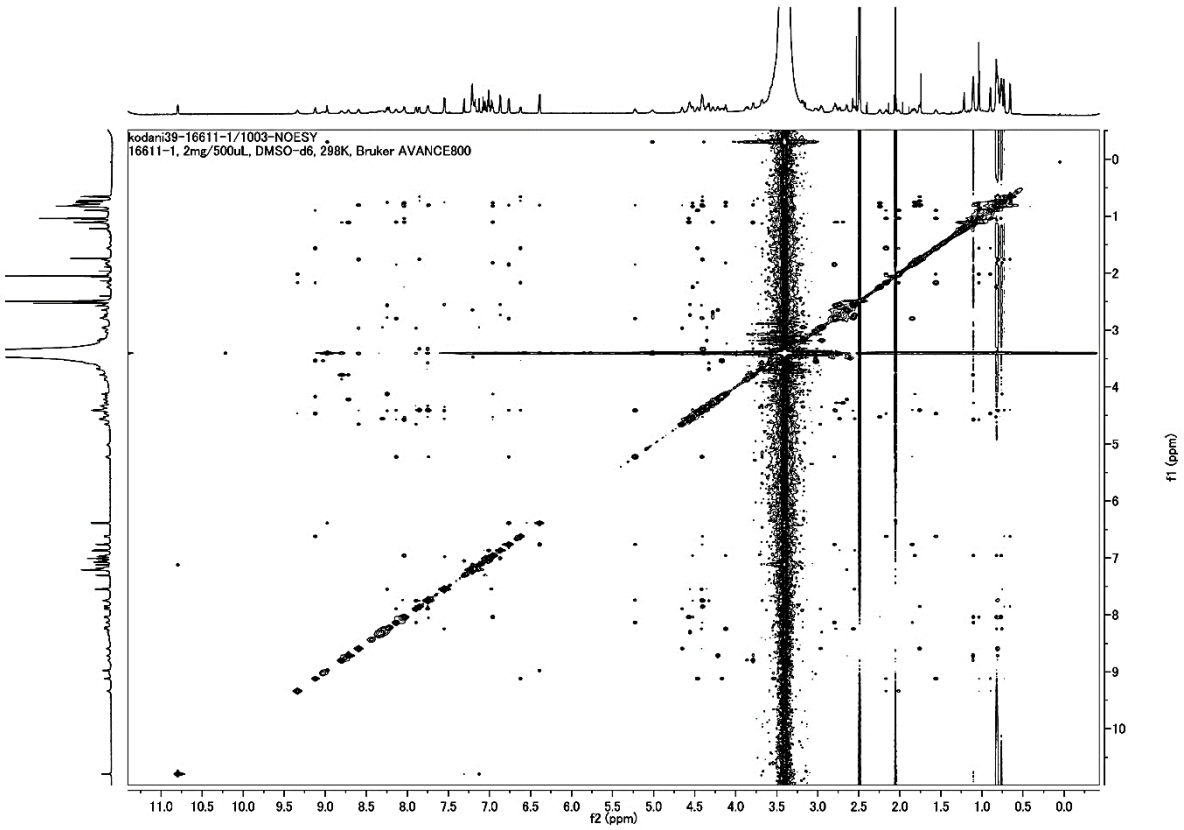


Figure 1.10 NOESY spectrum of specialicin

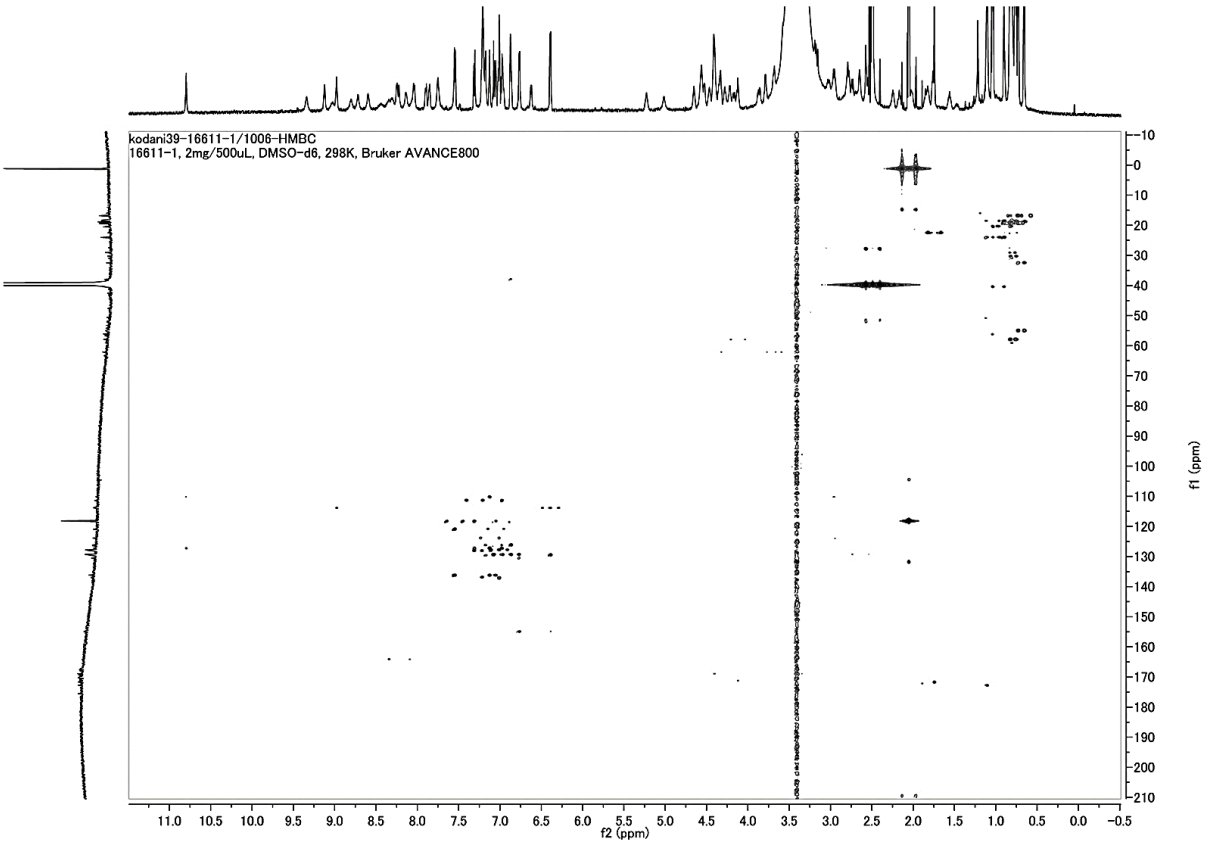


Figure 1.11 HMBC spectrum of specialicin

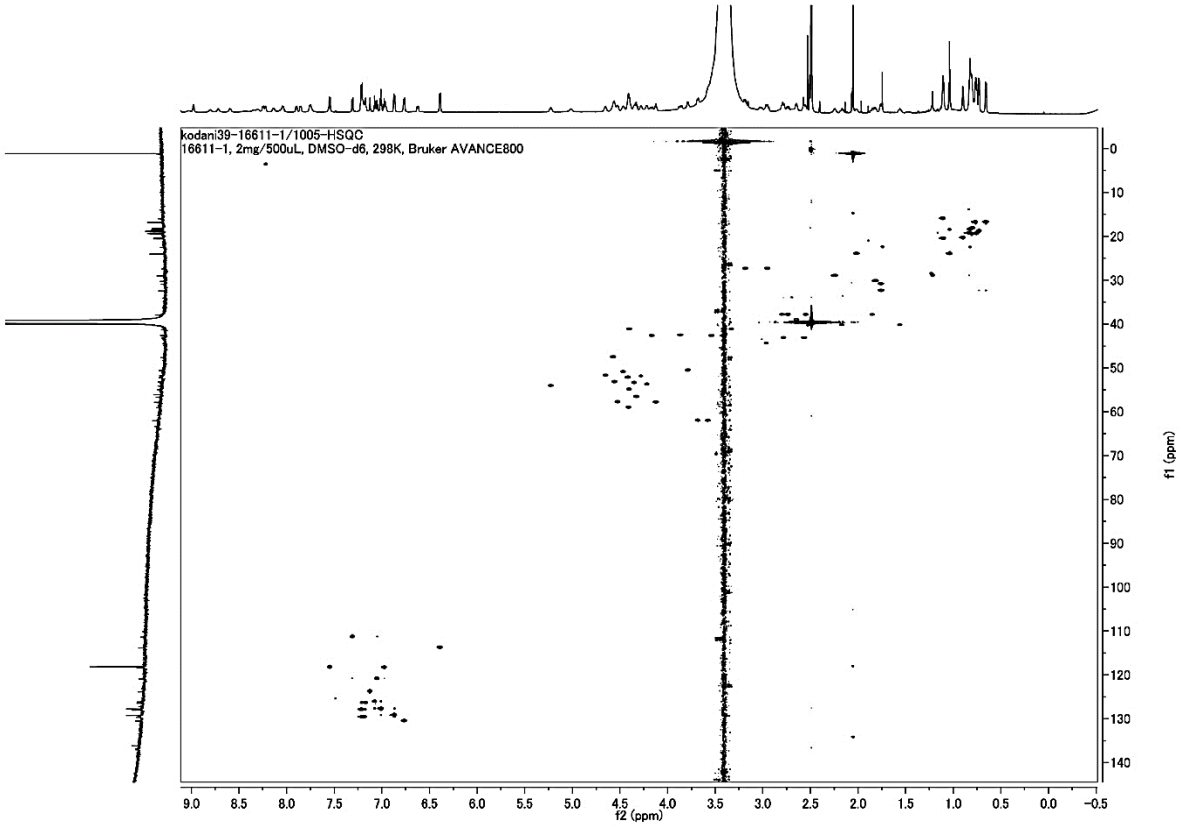


Figure 1.12 HSQC spectrum of specialicin

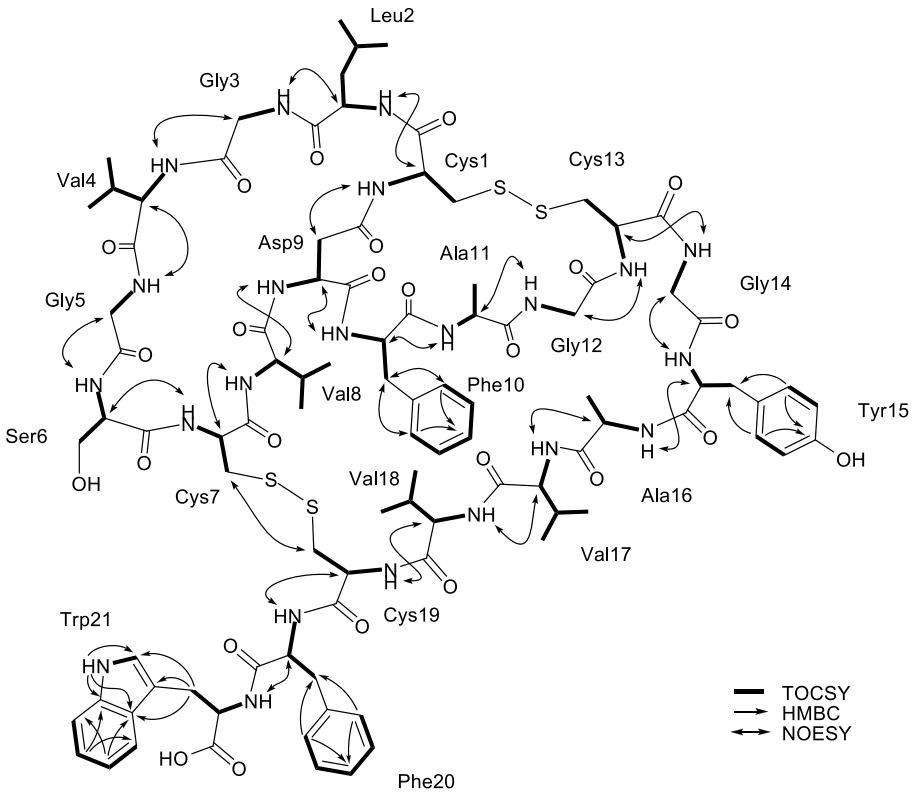


Figure 1.13 Key NMR correlations for structure determination of specialicin

Table 1.1 NMR chemical shift values of specialicin in DMSO-d_6

AA	POS	δH	δC
Cys1	CO		ND
	NH	8.44	
	α	4.57	50.2
	β	2.61, 3.49	ND
Leu2	CO		ND
	NH	9.34	
	α	4.47	51.0
	β	1.56, 2.17	40.3
	γ	2.02	24.0
	δ1	1.04	20.4
	δ2	0.90	24.1
Gly3	CO		ND
	NH	9.12	
	α	3.54, 4.17	42.8
Val4	CO		168.9
	NH	6.62	
	α	4.41	55.0
	β	1.76	32.5
	γ1	0.76	16.8
	γ2	0.66	18.9
Gly5	CO		ND
	NH	7.85	
	α	3.33, 4.40	41.3
Ser6	CO		169.7
	NH	7.75	
	α	4.33	56.7
	β	3.58, 3.68	62.1
	OH	5.02	
Cys7	CO		ND
	NH	7.90	
	α	4.65	51.8
	β	3.00, 3.42	44.5
Val8	CO		ND
	NH	8.60	
	α	4.41	59.1
	β	1.76	31.0
	γ1	0.80	18.3
	γ2	0.80	19.6
Asp9	α-CO		ND
	NH	7.74	
	α	4.28	52.0
	β	2.69, 2.78	34.1
	β-CO		ND
Phe10	CO		ND
	NH	8.06	
	α	4.22	53.8
	β	2.65	39.2
	γ		136.9
	δ	7.21	129.7
	ε	7.22	128.0
	ζ	7.17	126.5
Ala11	CO		ND
	NH	8.72	
	α	3.79	50.6
	β	1.11	16.0
Gly12	CO		ND
	NH	8.80	
	α	3.86, 3.41	42.6
Cys13	CO		ND
	NH	8.35	
	α	4.25	ND
	β	3.47, 2.83	ND
Gly14	CO		ND
	NH	7.20	
	α	3.02, 3.54	43.6
Tyr15	CO		ND
	NH	9.03	
	α	5.23	54.2
	β	1.85, 2.80	38.0
	γ		129.5
	δ	6.76	130.6
	ε	6.39	113.9
	ζ		155.0
	OH	8.98	
Ala16	CO		ND
	NH	8.14	
	α	4.57	47.6
	β	1.10	20.6
Val17	CO		ND
	NH	8.04	
	α	4.53	57.9
	β	2.24	29.1
	γ1	0.83	16.9
	γ2	0.77	19.3
Val18	CO		ND
	NH	6.96	
	α	4.12	57.9
	β	1.82	30.3
	γ1	0.82	18.6
	γ2	0.76	19.4
Cys19	CO		ND
	NH	8.25	
	α	4.42	52.3
	β	2.57, 2.78	43.2
Phe20	CO		ND
	NH	7.55	
	α	4.56	53.3
	β	2.55, 2.74	38.0
	γ		137.1
	δ	6.87	129.3
	ε	7.01	127.8
	ζ	7.08	126.2
Trp21	CO		ND
	NH	8.30	
	α	4.35	53.5
	β	2.95, 3.18	27.5
	γ		110.2
	δ1	7.13	123.9
	δ2		127.3
	ε1	10.80	
	ε2		136.2
	ε3	7.55	118.3
	ζ2	7.31	111.4
	ζ3	6.98	118.4
	η2	7.05	120.9

ND: not determined

1.3.4 Modified Marfey's analysis of specialicin

Stereochemistry of constituent amino acids in specialicin was clarified by Modified Marfey's analysis, according to the previous report.[96] The hydrolysate of specialicin was derivatized using *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA) and the derivatives were analyzed by HPLC, compared to derivatives of standard amino acids.

HPLC condition 1 was applied for standard amino acids including cysteic acid (Cys), aspartic acid (Asp), tyrosine (Tyr), serine (Ser), alanine (Ala), leucine (Leu), tryptophan (Trp) and phenylalanine (Phe). Retention times (min) of L- or D-FDLA derivatized amino acids in HPLC condition 1 were as follows; L-Cysteic acid-L-FDLA (13.73 min), L-Cysteic acid-D-FDLA (12.74 min), L-Asp-L-FDLA (23.43 min), L-Asp-D-FDLA (26.99 min), L-Tyr-L-FDLA (34.70 min), L-Tyr-D-FDLA (24.09 min), L-Ser-L-FDLA (24.39 min), L-Ser-D-FDLA (25.80 min), L-Ala-L-FDLA (32.09 min), L-Ala-D-FDLA (39.13 min), L-Leu-L-FDLA (43.71 min), L-Leu-D-FDLA (59.16 min), L-Trp-L-FDLA (44.16 min), L-Trp-D-FDLA (51.12 min), L-Phe-L-FDLA (46.22 min) and L-Phe-D-FDLA (54.16 min) (Fig. 1.14 – 1.21). To detect cysteic acid, the peptide specialicin was hydrolyzed with 6 N HCl containing 0.2 % sodium azide and was analyzed by HPLC condition 1, compared to standard amino acids, as shown in Fig. 1.22. To recover Trp, the peptide was hydrolyzed with 6 N HCl containing 3 % phenol and was analyzed by HPLC condition 1, compared to standard amino acids, as shown in Fig. 1.23. However, Trp was still difficult to detect in this HPLC condition. We performed co-injection experiment on L-FDLA derivative of specialicin by adding L-Trp-L-FDLA or L-Trp-D-FDLA, as shown in Fig. 1.24.

HPLC condition 2 was applied for valine (Val) analysis. Retention times (min) of L- or D-FDLA derivatized amino acids in HPLC condition 2 were as follows; L-Val-L-FDLA (35.90 min) and L-Val-D-FDLA (56.40 min) (Fig. 1.25). HPLC chromatogram of L-FDLA derivative of specialicin hydrolyzed with 6 N HCl containing 3% phenol in HPLC condition 2 is shown in Fig. 1.26. Overall, stereochemistries of all constituent amino acids in specialicin, except for Trp, were determined to be L. The stereochemistry of Trp was determined to be D, same as siamycin I (MS-271) which was also reported to contain D-Trp at the C-terminus. Thus, the planar chemical structure of specialicin is shown in Fig. 1.27.

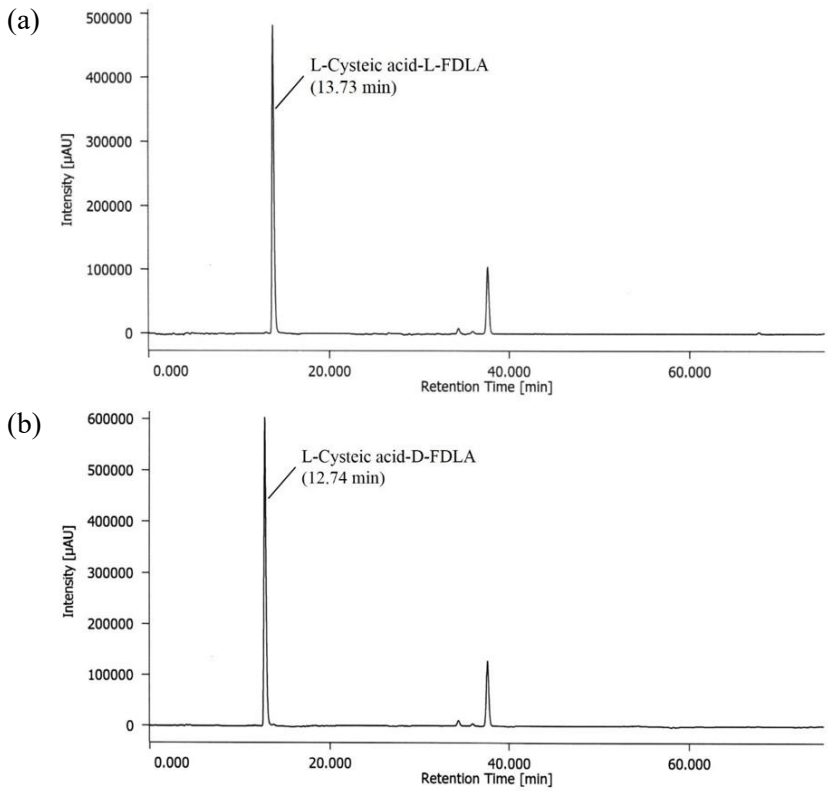


Figure 1.14 HPLC chromatograms of (a) L-Cysteic acid-L-FDLA and (b) L-Cysteic acid-D-FDLA (HPLC condition 1)

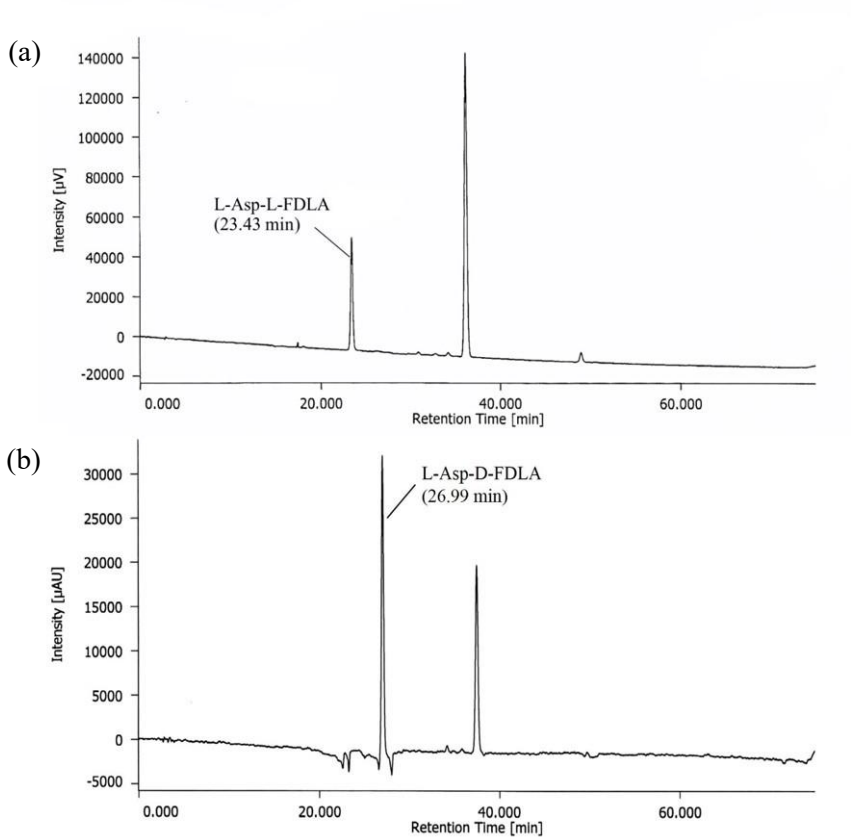


Figure 1.15 HPLC chromatograms of (a) L-Asp-L-FDLA and (b) L-Asp-D-FDLA (HPLC condition 1)

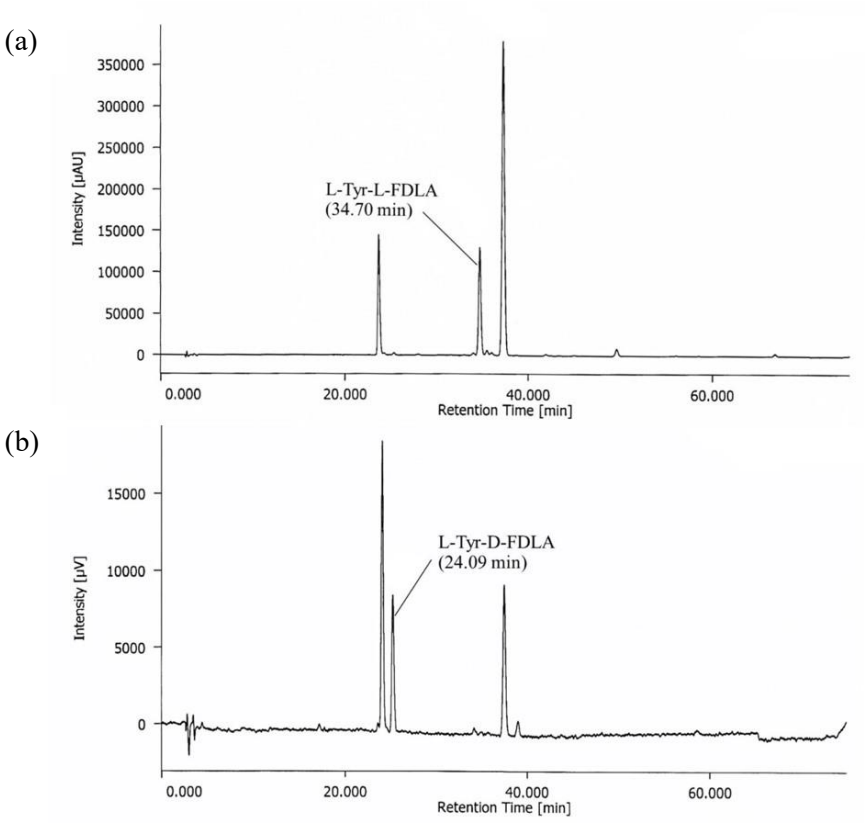


Figure 1.16 HPLC chromatograms of (a) L-Tyr-L-FDLA and (b) L-Tyr-D-FDLA (HPLC condition 1)

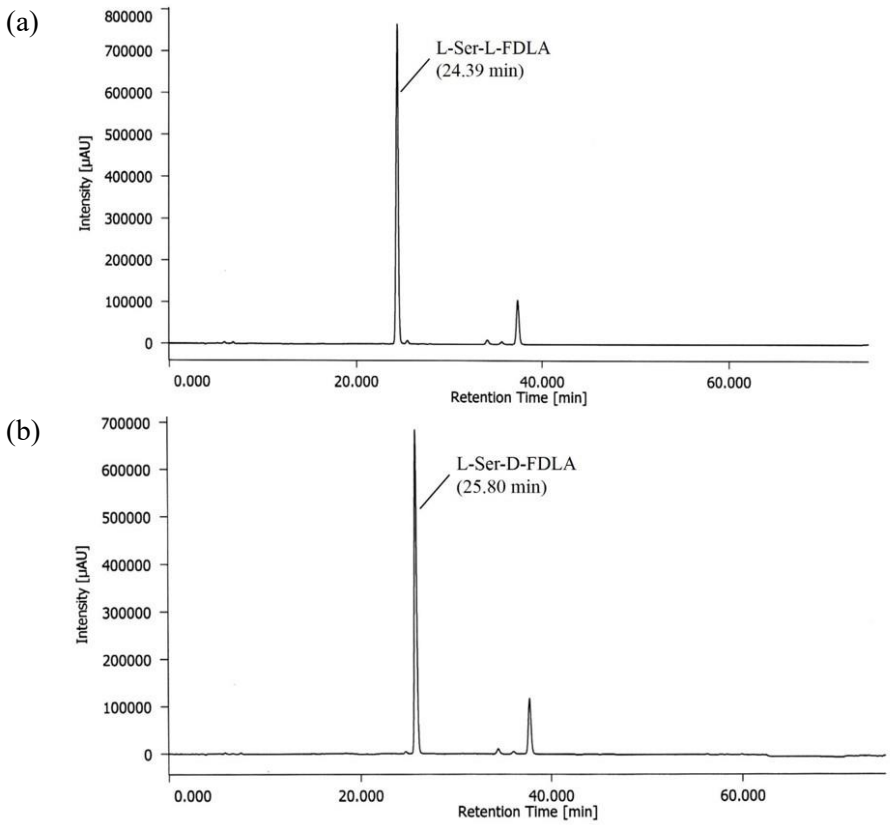


Figure 1.17 HPLC chromatograms of (a) L-Ser-L-FDLA and (b) L-Ser-D-FDLA (HPLC condition 1)

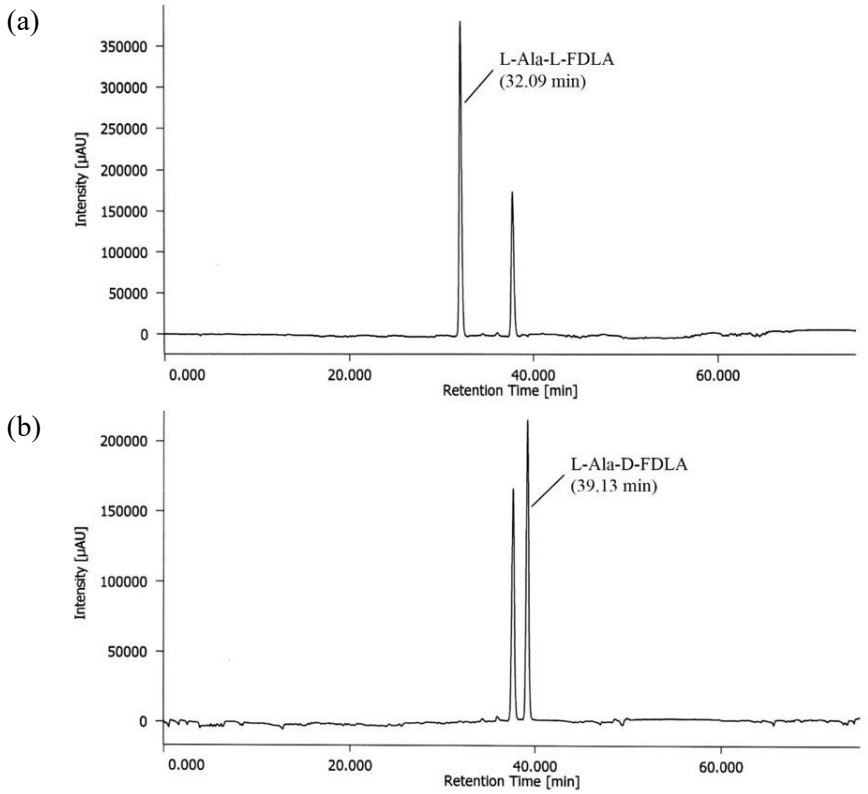


Figure 1.18 HPLC chromatograms of (a) L-Ala-L-FDLA and (b) L-Ala-D-FDLA (HPLC condition 1)

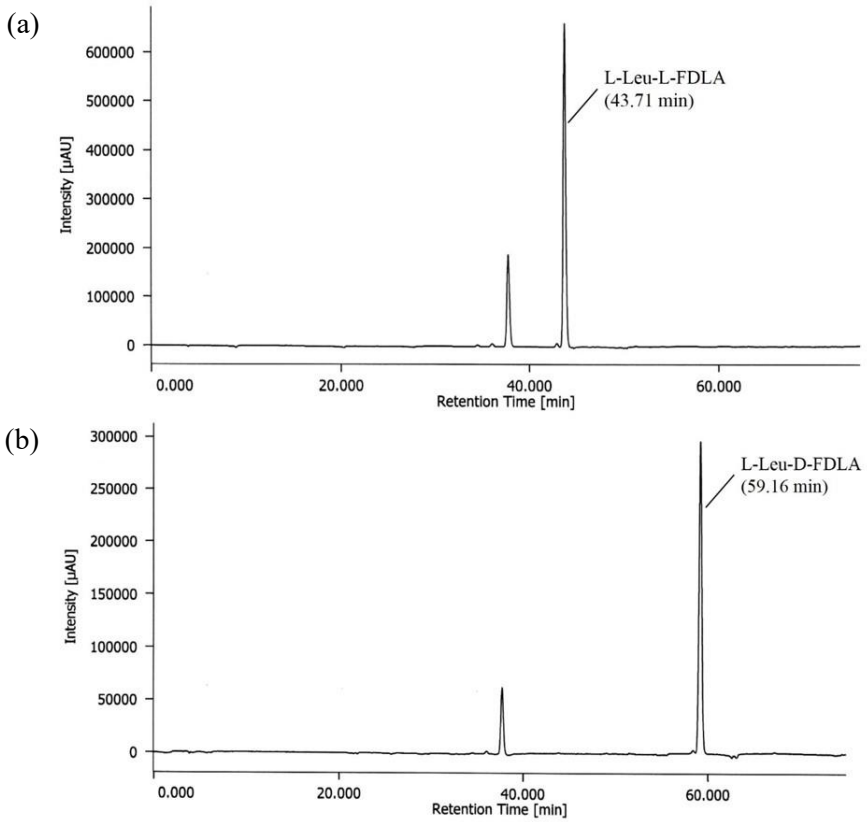


Figure 1.19 HPLC chromatograms of (a) L-Leu-L-FDLA and (b) L-Leu-D-FDLA (HPLC condition 1)

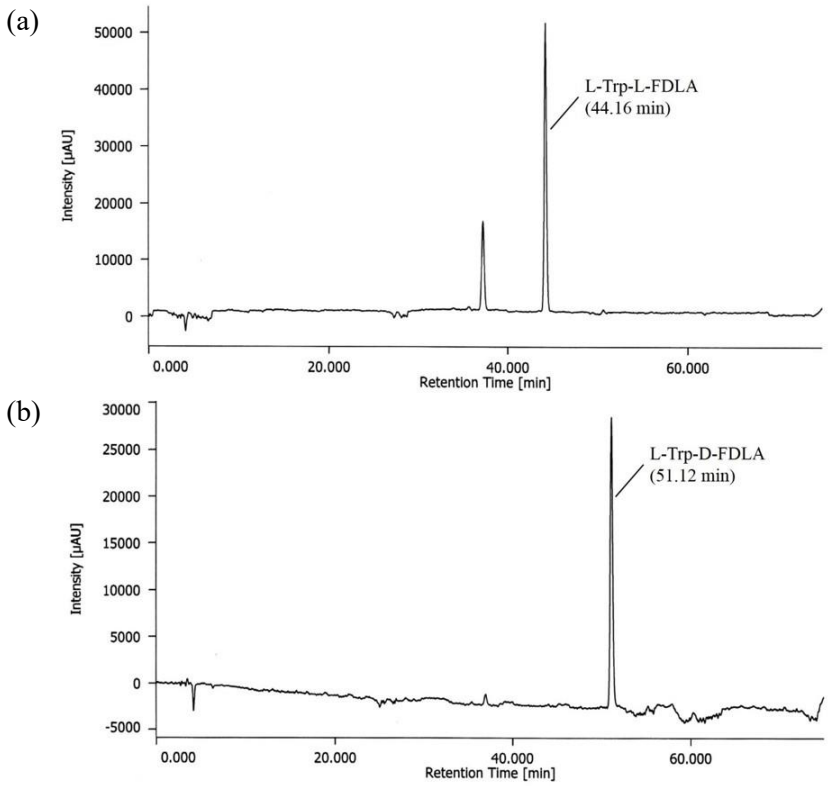


Figure 1.20 HPLC chromatograms of (a) L-Trp-L-FDLA and (b) L-Trp-D-FDLA (HPLC condition 1)

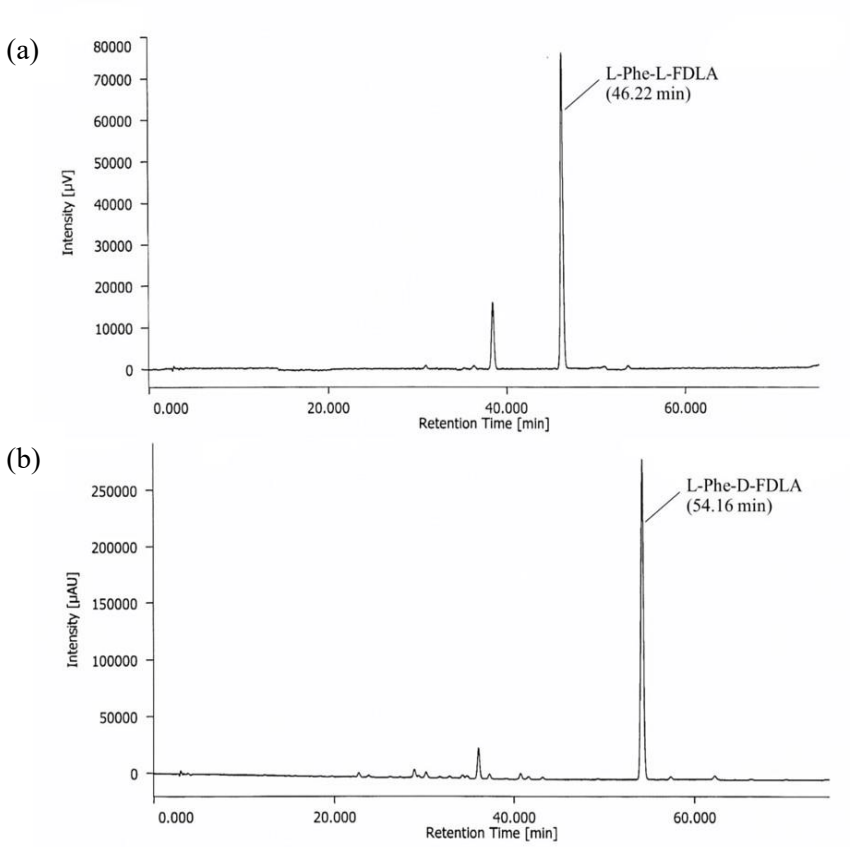


Figure 1.21 HPLC chromatograms of (a) L-Phe-L-FDLA and (b) L-Phe-D-FDLA (HPLC condition 1)

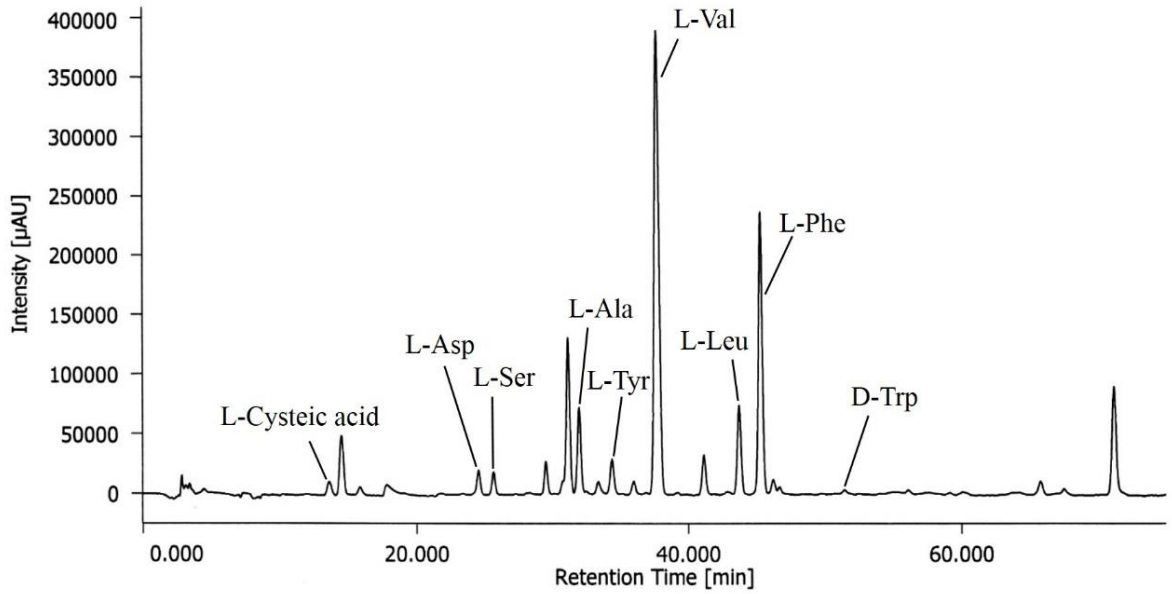


Figure 1.22 HPLC chromatogram of L-FDLA derivative of specialicin hydrolyzed with 6N HCl containing 0.2 % sodium azide (HPLC condition 1)

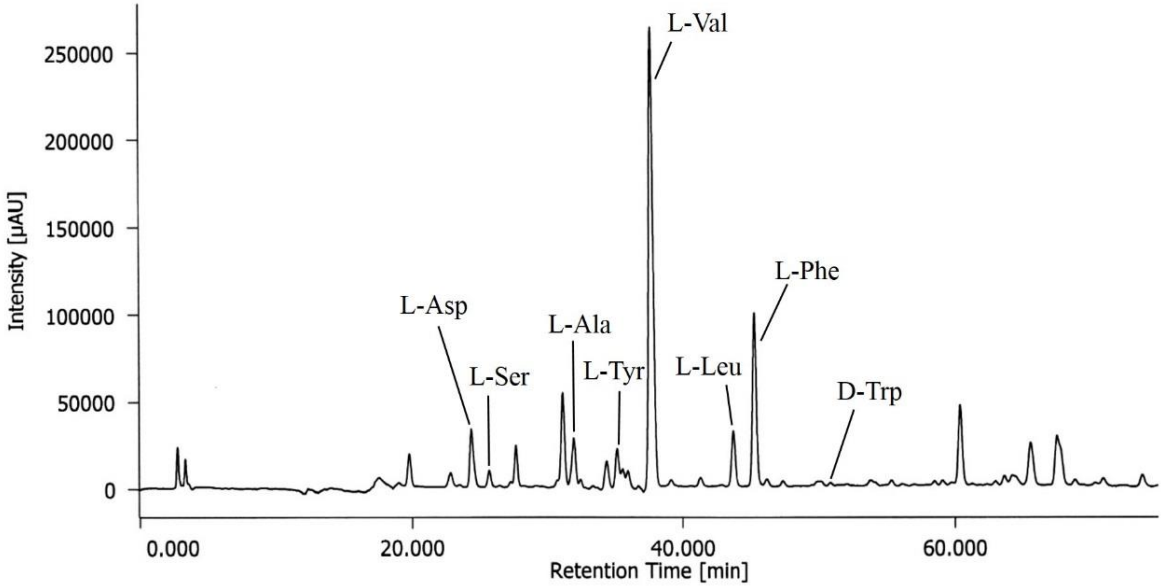


Figure 1.23 HPLC chromatogram of L-FDLA derivative of specialicin hydrolyzed with 6N HCl containing 3 % phenol (HPLC condition 1)

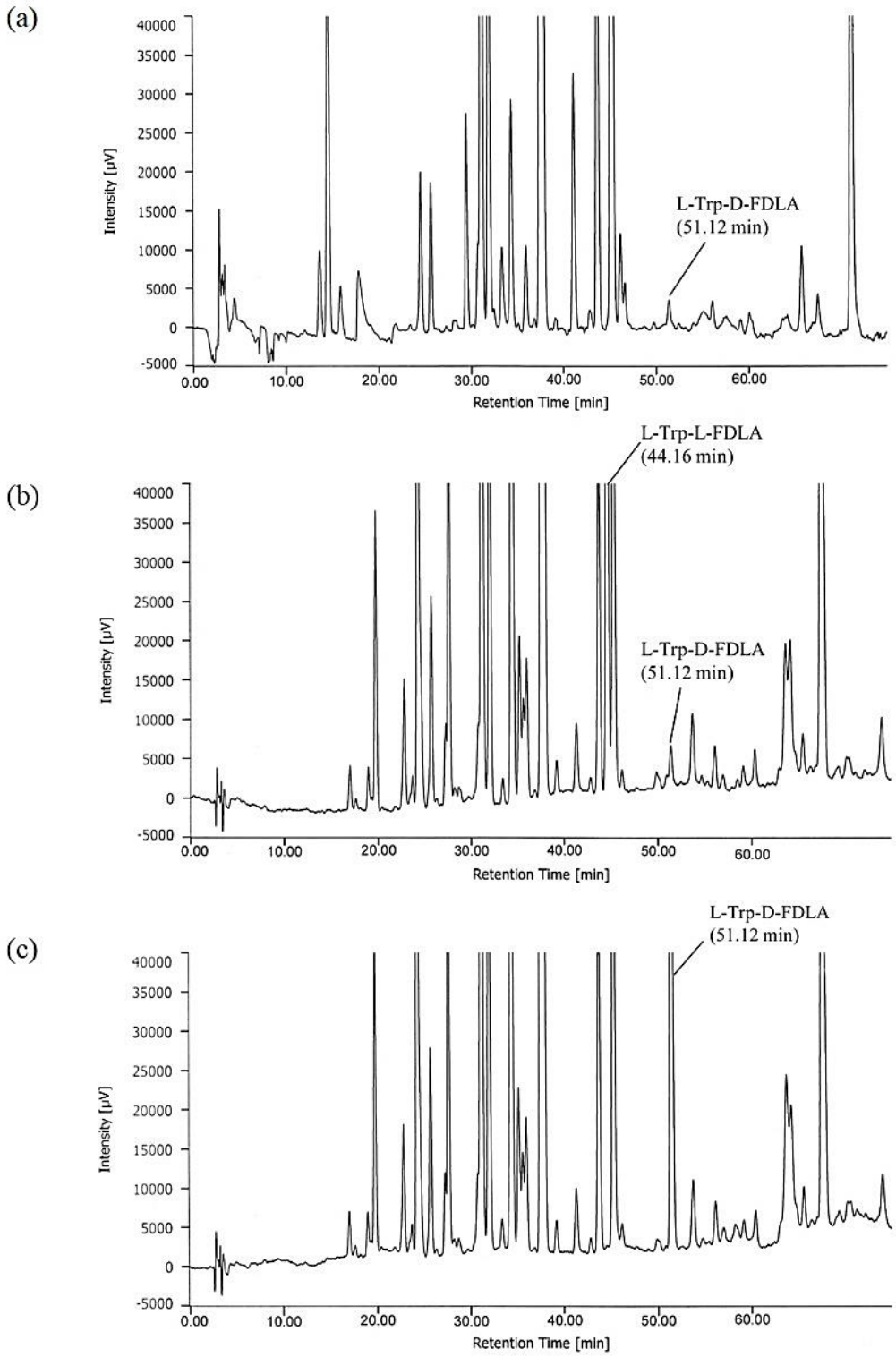


Figure 1.24 HPLC chromatograms of (a) L-FDLA derivative of specialicin hydrolyzed with 6N HCl containing 3 % phenol (b) co-injection between L-FDLA derivative of specialicin and L-Trp-L-FDLA (c) co-injection between L-FDLA derivative of specialicin and L-Trp-D-FDLA (HPLC condition 1)

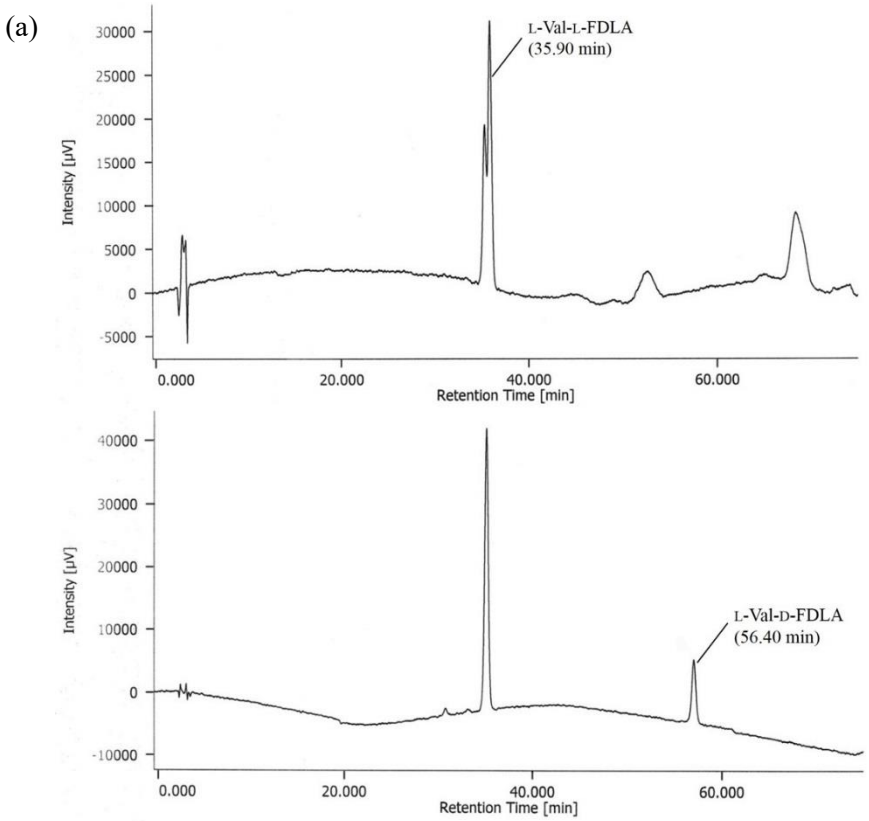


Figure 1.25 HPLC chromatograms of (a) L-Val-L-FDLA and (b) L-Val-D-FDLA (HPLC condition 2)

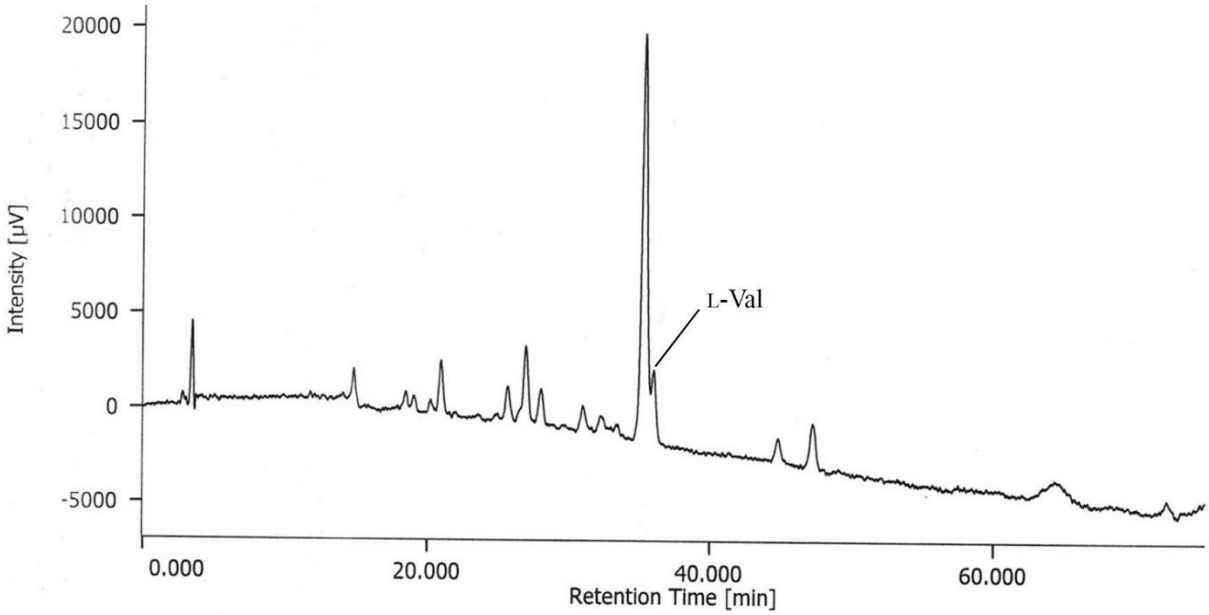


Figure 1.26 HPLC chromatogram of L-FDLA derivative of specialicin hydrolyzed with 6N HCl containing 3 % phenol (HPLC condition 2)

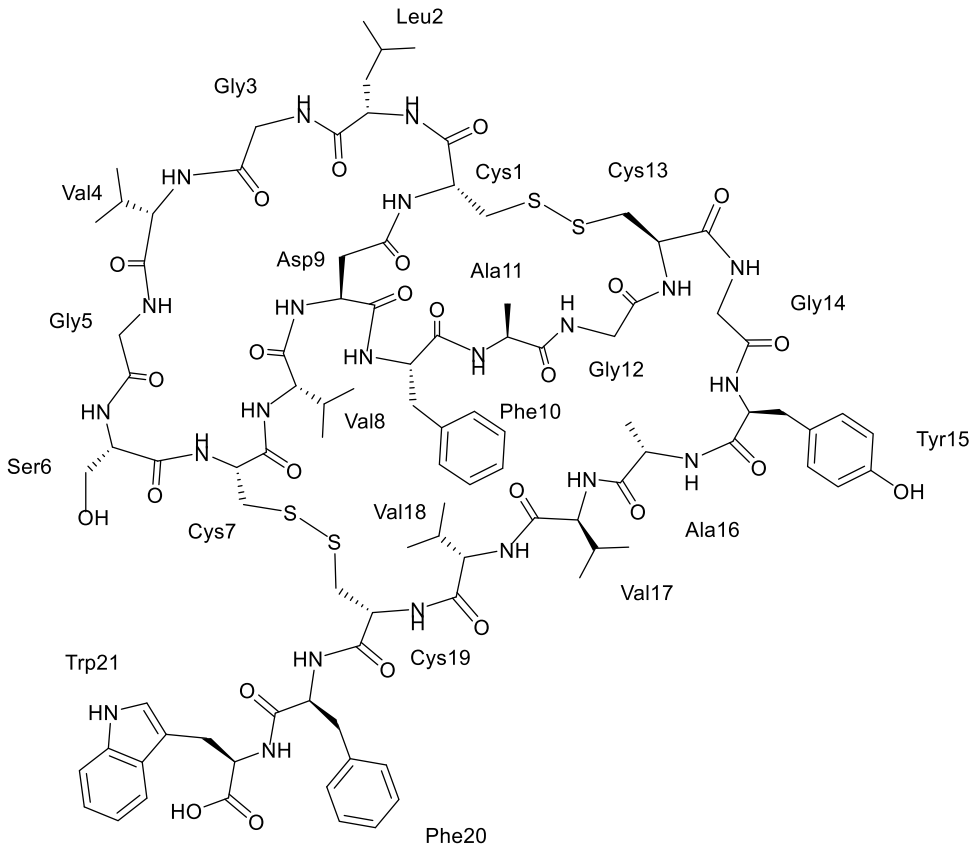


Figure 1.27 Chemical structure of specialicin

1.3.5 Three-dimensional (3D) structure of specialicin

Three-dimensional solution structure of specialicin was generated by CNS (version 1.1), based on distance and dihedral angle restraints determined from NOE intensities and $^{3}J_{HN\alpha}$ coupling constants (Fig. 1.28). Superposition of the 15 lowest-energy structures and the lowest-energy structure of specialicin are shown in Fig. 1.28a and b, respectively. As shown in Fig. 1.28, specialicin comprise of 21 amino acids. Macrolactam ring was formed by nine amino acids, from Cys1 to Asp9, indicated by green. The isopeptide bond was formed between Cys1 and Asp9, indicated by red. Amino acid residues Phe10 to Cys13 form loop structure and residues Gly14 to Trp21 form tail structure, indicated by blue. Disulfide bond between Cys1 and Cys13, and disulfide bond between Cys7 and Cys19 are indicated by yellow. The three-dimensional structure of specialicin is very similar to that of siamycin I (MS-271).[94]

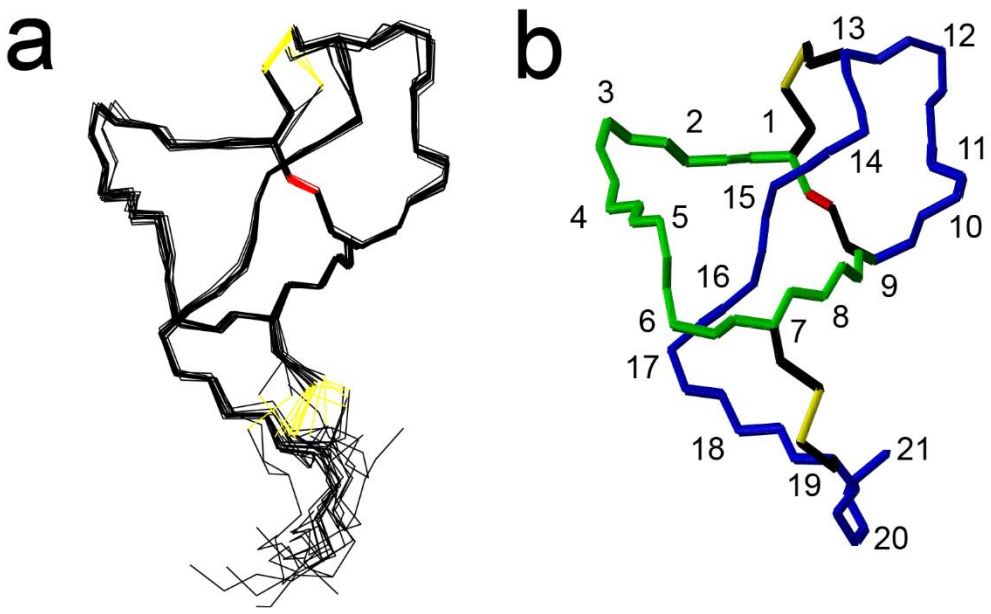


Figure 1.28 Three-dimensional (3D) structure of specialicin; (a) superposition of the 15 lowest-energy structures and (b) the lowest-energy structure of specialicin; red indicates an isopeptide bond; yellow indicates two disulfide bonds; green indicates macrolactam ring; blue indicates loop and tail structure.

1.3.6 Biological activities of specialicin

Antibacterial activity of specialicin was tested against several Gram-positive and Gram-negative bacteria including *E. coli*, *P. aeruginosa*, *S. aureus*, *B. subtilis* and *M. luteus* using minimum inhibitory concentrations (MICs) assay. As a result, specialicin exhibited an antibacterial activity against *M. luteus* with MIC of 8 μg/ml. However, the peptide did not show antibacterial activity against the other testing bacteria at the concentration of 32 μg/ml (Table 1.2). In addition, the inhibitory activity of specialicin was tested against HIV-1 NL4-3 strain using the WST-8 assay performed in PM1/CCR5 cells. Specialicin exhibited moderate activity against HIV-1 NL4-3 with 50% inhibitory concentration (IC_{50}) of 7.2 ± 2.1 μM (15 ± 4.5 μg/ml).

Table 1.2 Minimum inhibitory concentrations (MICs) of specialicin

Test microorganisms	MIC (μg/ mL)	
	Tetracycline	Specialicin
Bacillus subtilis	1	>32
Micrococcus luteus	0.25	8
Escherichia coli	2	>32
Staphylococcus aureus	0.125	>32
Pseudomonas aeruginosa	4	>32

1.3.7 Biosynthetic gene cluster of specialicin

Since a new lasso peptide structural gene was found in the genome sequence of *S. specialis* JCM 16611^T, the gene cluster, containing specialicin structural gene, was considered as specialicin BGC, named *spe* cluster. Previously, *msl* cluster which responsible for biosynthesis of MS-271 (siamycin I) was reported by Feng et al.[81] The *spe* cluster reveals gene organization similar to that of *msl* cluster, as shown in Fig. 1.29. Therefore, we annotated genes within *spe* cluster and proposed biosynthetic pathway for specialicin based on the similarity and mechanism of MS-271. The *spe* cluster includes genes encoding a specialicin precursor peptide (SpeA), maturation enzymes (SpeB1, SpeB2 and SpeC) and proteins involved in transport (SpeD1, SpeD2, SpeD3 and SpeD4), disulfide bond formation (SpeE and SpeF), regulation (SpeG, SpeR1 and SpeR2) and epimerization (SpeH). Considering specialicin biosynthetic pathway, a specialicin precursor peptide is ribosomally synthesized by SpeA. SpeB1 and SpeB2 are responsible for recognition and proteolysis, respectively, to generate a linear peptide comprising of 21 amino acids. A macrolactam synthetase (SpeC) forms macrolactam ring between amino group of N-terminal Cys1 and carboxyl group of Asp9. SpeE and SpeF form disulfide bonds between Cys1 and Cys13 and between Cys7 and Cys19,

respectively. The peptide specialicin is finally exported by ABC transporters (SpeD1, D2, D3 and D4). SpeG and SpeR1/R2 are two-component histidine kinases and response regulators, respectively, which involve in regulation of gene expression of *spe* cluster. SpeH, homologous to MslH, is responsible for epimerization of Trp21.

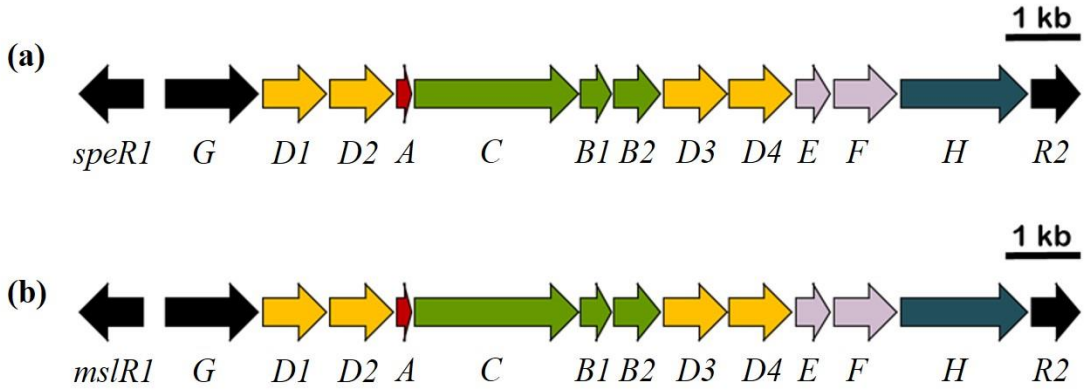


Figure 1.29 Gene organization of biosynthetic gene clusters for (a) specialicin and (b) MS-271; Function of genes are indicated by colors as following; red indicates precursor peptide; green indicates lasso topology formation; purple indicates disulfide bonds formation; dark blue indicates epimerization at C-terminal Trp residue; yellow indicates transport; black indicates regulation

1.4 Summary

In this study, a new lasso peptide structural gene was found in the genome sequence of *S. specialis* by genome mining. The peptide specialicin was isolated from the extract of *S. specialis* JCM 16611^{T}. Specialicin was a new lasso peptide with a length of 21 amino acids, containing isopeptide bond and two disulfide bonds. Interestingly, stereochemistry of Trp at C-terminus of specialicin was determined to be D-configuration. Three-dimensional structure indicated that specialicin possessed conformational structure similar to siamycin I (MS-271). The biosynthetic gene cluster for specialicin, *spe* cluster, was proposed in the genome sequence of *S. specialis* and genes involved in specialicin biosynthesis were annotated based on the similarity to *msl* cluster. Specialicin showed moderate anti-HIV activity against HIV-1 NL4-3 and antibacterial activity against *M. luteus*.

Chapter II

Isolation and structure determination of a new cytotoxic peptide, curacozole, from *Streptomyces curacoi* based on genome mining

2.1 Introduction

Peptide natural products have been known as promising therapeutic agents in the treatment of cancer. They can be used directly as a cytotoxic agents or can be a carrier of cytotoxic agents and radioisotopes by specifically targeting cancer cells.[97] So far, several cytotoxic peptides have been discovered and characterized. Most of them possess a macrocyclic structure. Macrocyclization is a common occurrence found in two major classes of peptide natural products, including nonribosomal peptides (NRPs) and ribosomally synthesized and post-translationally modified peptides (RiPPs).[32, 52, 59] The macrocyclization has been reported to improve biological activity of peptides. For instance, cyclization of RGD peptides can enhance their binding affinity for integrin cell receptors.[98] In the recent years, macrocyclic peptides become potential candidates for pharmacological applications since their properties, for example binding affinity, proteolytic stability and solubility, can be easily optimized and tailored to a specific application.[99] Many macrocyclic peptides have been utilized as selective biding peptides involving in clinically protein−protein interactions (PPI) for treatment of human diseases, such as cancer, neurodegeneration and autoimmunity.[98-100]

Telomestatin is a macrocyclic peptide isolated from *Streptomyces anulatus* 3533-SV4 as a potent telomerase inhibitor.[101] Since telomestatin's structure is very similar to G-tetrad, the telomerase inhibition might be caused by its ability to facilitate and/or stabilize intermolecular G-quadruplex structures.[102] The gene cluster which responsible for telomestatin biosynthesis was identified by whole genome sequence analysis of *S. anulatus* 3533-SV4 and heterologous expression using the engineered host strain *S. avermitilis* SUKA. The result indicated that telomestatin was biosynthesized by RiPP biosynthetic system.[103] In addition, the macrocyclic peptide, YM-216391, was isolated from *Streptomyces nobilis* JCM 4274 as a potent cytotoxic compound against a human cancer cell line.[104] Structure elucidation based on 1D and 2D NMR analyses revealed that YM-216391 structure contains a polyoxazole-thiazole moiety which is similar to telomestatin.[105] The biosynthetic gene cluster for YM-216391 was identified from genome sequence data of *S. nobilis* JCM 4274. Bioinformatic analysis indicated the presence of a precursor peptide coding gene, *ymA*, encoding 36 amino acids containing the FIVGSSSC amino acid sequence, which is identical to predicted YM-216319 core peptide sequence. Maturation enzyme coding genes were also found

nearby. This suggested that YM-216391 is biosynthesized by RiPP system.[106]

Previously, antibacterial peptide curacomycin was isolated from *Streptomyces curacoi* NBRC 12761^T by our group.[77] Further genome mining on a genome sequence of this strain resulted in discovery of a new gene cluster homologous to YM-216391 gene cluster. By chemical investigation on the extract of *S. curacoi* NBRC 12761^T, we found a new cytotoxic peptide, named curacozole. We successfully isolated the peptide, curacozole, and determined its structure using ESI-MS and NMR analyses. The BGC for curacozole was clarified in the genome sequence of *S. curacoi*, and compared to YM-216391 BGC. Curacozole exhibited cytotoxic activity against HCT116 and HOS cancer cells. The isolation and structure determination of a new cytotoxic peptide, curacozole, were described in the chapter II.

2.2 Materials and Methods

2.2.1 Bacterial strains

Streptomyces curacoi NBRC 12761^T was obtained from NITE Biological Resource Center (NBRC) culture collection, Japan. Rifampicin-resistant (rif^r) mutants of *S. curacoi* NBRC 12761^T were obtained by spontaneous mutation. Bacterial spores or hyphae (approximately 2×10^9 colony forming units) were spread on GYM agar medium[107] containing 10 μg/ml of rifampicin, corresponding to an amount that is 2-fold greater than the minimum inhibitory concentration. Rifampicin-resistant (rif^r) mutants of *S. curacoi* NBRC 12761^T were harvested after 10 days of cultivation.

2.2.2 Mutation analysis of *rpoB* gene

The partial *rpoB* gene fragment of rif^r mutants was obtained by polymerase chain reaction (PCR) using the primers as following; 5'-GGCGCTCGGCTGGACGACCG-3' (forward primer) and 5'-CGATCAGACCGATGTTCGGG-3' (reverse primer), which were designed based on sequence of *S. curacoi* NBRC 12761^T. The PCR amplification was accomplished using Tks Gflex DNA polymerase (TaKaRa Bio, Inc., Shiga, Japan). The purified PCR product was subjected to DNA sequencing analysis, performed by Eurofins Genomics K. K. (Tokyo, Japan). DNA sequence data was aligned using the ClustalW program at DDBJ (http://clustalw.ddbj.nig.ac.jp/).

2.2.3 Isolation of peptide

Streptomyces curacoi rifampicin-resistant (rif^r) mutant strain R25, also called *S. curacoi* R25, was cultured in 2 L of ISP2 agar medium[82] incubated at 30°C for 7 days. After cultivation, bacterial cells and aerial hyphae were harvested and approximately 200 ml of

MeOH was added to the harvested cells for extraction. The extract was filtered using filter paper (Whatman No.1, GE Healthcare Life Sciences, Little Chalfont, UK), followed by concentrating using rotary evaporator. The concentrated extract was purified by open column chromatography (styrene-divinylbenzene resin, CHP-20P, Mitsubishi Chemical Corp., Tokyo, Japan), eluted with 10% MeOH, 60% MeOH and 100% MeOH, respectively. The 100% MeOH fraction which contains expected peptide was concentrated using rotary evaporator and then subjected to HPLC separation using an ODS column (4.6 × 250 mm, Wakopak Handy ODS, WAKO). The UV detector was set at wave length of 220 nm. Isocratic elution mode, 73% acetonitrile (MeCN) containing 0.05% trifluoroacetic acid (TFA) at flow rate 1 ml/min, was used to isolate the peptide curacozole (Retention time; 12.0 min).

2.2.4 ESI-MS analysis

ESI-MS analysis was performed using a JEOL JMS-T100LP mass spectrometer. Reserpine was used as an internal standard for accurate MS analysis.

2.2.5 NMR analysis

The peptide was dissolved in 500 µl of DMSO-d_6. NMR spectra were obtained on Bruker Avance 600 and Avance III HD 800 spectrometers with quadrature detection in the phase-sensitive mode by States-TPPI (time proportional phase incrementation) and in the echo-antiecho mode. One-dimensional (1D) ^{1}H, ^{13}C and DEPT-135 spectra were recorded at 25 °C with 12 ppm for proton and 239 ppm or 222 ppm for carbon. The two-dimensional (2D) DQF-COSY was recorded with 512 and 1024 complex points in t1 and t2 dimensions. The 2D TOCSY with MLEV-17 mixing sequence, was recorded with mixing time of 80 ms, 256 and 1024 complex points in t1 and t2 dimensions. The 2D NOESY was recorded with mixing times of 200 and 400 ms, 256 and 1024 complex points in t1 and t2 dimensions. The 2D ^{1}H-^{13}C HSQC and HMBC spectra were acquired at 25 °C in the echo-antiecho mode. The ^{1}H-^{13}C HSQC and HMBC spectra were recorded with 1024 and 512 complex points for 12 ppm in the ^{1}H dimension and 160 ppm or 222 ppm in the ^{13}C dimension, respectively, at a natural isotope abundance. All NMR spectra were processed using TOPSPIN 3.5 (Bruker). Peak-picking and assignment were performed with Sparky program. Before Fourier transformation, the shifted sinebell window function was applied to t1 and t2 dimensions. All ^{1}H and ^{13}C dimensions were referenced to DMSO-d_6 at 25 °C.

2.2.6 Modified Marfey's analysis

Marfey's analysis was carried out in a sealed vacuum hydrolysis tube. The peptide (1.0 mg) was hydrolyzed with 500 µl of 6 N hydrochloric acid (HCl) at 110˚C for 16 h. After cooling down at room temperature, the hydrolysate was completely evaporated using rotary evaporator and freeze-drying under vacuum. The hydrolysate was resuspended in 200 µl of water, followed by adding 10 µl of the solution of *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA, Tokyo Chemical Industry Co., LTD, Tokyo, Japan) in acetone (10 µg/µL). The 100 µl of 1 M sodium bicarbonate ($NaHCO_3$) was added to the hydrolysate and the mixture was incubated at 80˚C for 3 min. The mixture was cooled down to room temperature and then neutralized with 50 µl of 2 N HCl. The 1 ml of 50% MeCN/water was added to the mixture before it was subjected to HPLC. Standard amino acids (1.0 mg) were derivertized with L-FDLA or D-FDLA in the same manner. Approximately 30 µl of FDLA-derivertive peptide was analyzed by HPLC (C18 column, 4.6 × 250 mm, Wakopak Handy ODS, Wako Pure Chemical Industries, Tokyo, Japan), compared to FDLA-derivertive standard amino acids. The DAD detector (MD-2018, JASCO, Tokyo, Japan) was used to detect FDLA-derivertives, accumulating data of absorbance from 220 nm to 420 nm. For standard amino acids including allo-isoleucine (allo-Ile) and isoleucine (Ile), HPLC analyses were performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 70 min, increasing percentage of solvent B from 25% to 60%.

2.2.7 Chiral HPLC analysis

The peptide (1.0 mg) was hydrolyzed with 500 µl of 6 N hydrochloric acid (HCl) at 110˚C for 16 h. The hydrolysate of peptide was analyzed using chiral HPLC on a SUMICHIRAL OA5000 (150×4.6 mm, Sumika Chemical Analyervice). The HPLC analysis was performed with UV detection at 254 nm, using 2 mM $CuSO_4$ containing 5% MeCN as a mobile phase at a flow rate of 1 mL/min.

2.2.8 Cytotoxic assay

A human colon carcinoma cell line (HCT116) and human osteosarcoma (HOS) cell (approximately 2.5×10^3) were prepared in 96-well microplates and treated with various concentrations of peptide curacozole (5-100 nM) in D-MEM and E-MEM containing FBS (10%) for HCT116 and HOS cells, respectively. The cell viability was examined after 72 h incubation using CellTiter-Glo luminescent cell viability assay (Promega, Madison, USA) with a JNR Luminescencer (ATTO, Tokyo, Japan), according to the manufacturer's protocol.

2.3 Results and discussion

2.3.1 Genome mining for new peptide discovery

YM-216391, a peptide isolated from *Streptomyces nobilis* JCM 4274, was reported to be a novel cytotoxic peptide and has been characterized as a new class of cyclic peptides containing a polyoxazole-thiazole moiety.[104, 105] The biosynthetic gene cluster (BGC) for YM-216391 was identified in genome sequence of *S. nobilis* JCM 4274 based on heterologous expression.[106] The peptide YM-216391 was biosynthesized from the precursor peptide (accession number: AFJ68074.1) and modified by several modification enzymes. Amino acids Ser and Cys in the core peptide region FIVGSSSC were converted to oxazoles and thiazoles by modification enzymes. BLASTp search using amino acid sequence of YM-216391 precursor peptide resulted in discovery of analogous precursor peptide coding genes in genome sequences of *S. curacoi* DSM 40107^T (accession number: WP_107116988.1), *S. viridochromogenes* Tü57 (accession number: WP_003997107.1) and *S. aurantiacus* JA 4570 (accession number: WP_016640788.1) as shown in Fig. 2.1.

S. nobilis (YM-216391)	M-----TAEIEEVDIEVGFIVGSSSCSLELEEDDLDVAADE
S. curacoi (curacozole)	MFENT-TAEIEEVDIEVGFIIGSTCCSLEMEEDDLDVAAEETEAV
S. viridochromogenes	MFENT-TAEIEEVDIEVGFIIGSTCCSLEMEEDDLDVAAEETETV
S. aurantiacus	MSENITTAEIEEVDIEVGFIIGSSCCSLEMEEDDLDVAADE

Figure 2.1 Amino acid sequences of YM-216391 precursor peptide and analogous precursor peptides found by BLASTp search; underlined letters indicate core peptide region; bold letters indicate conserved amino acids

According to the precursor peptide sequence found in *S. curacoi* DSM 40107^T, a new peptide was expected to be biosynthesized from the core peptide sequence FIIGSTCC (underlined letters in Fig. 2.1). By comparison between amino acid sequence of YM-216391 and the expected peptide, residues Val3, Ser6 and Ser7 of YM-216391 are substituted by Ile3, Thr6 and Cys7 in expected peptide, respectively (Fig. 2.1). Chemical investigations using HPLC and ESI-MS were performed on extract of *S. curacoi* NBRC 12761^T (=DSM 40107^T) to examine production of the expected peptide. As a result, the expected peptide named curacozole was detected in extract of *S. curacoi* NBRC 12761^T. However, curacozole productivity was found to be very low on HPLC chromatogram (data not shown).

2.3.2 Mutation analysis of *rpoB* gene

There are reports indicated that *Streptomyces* strain which possess specific RNA polymerase β gene (*rpoB*) mutation representing rifampicin resistance can produce abundant quantities of secondary metabolites.[108, 109] To increase productivity of the peptide curacozole, rifampicin-resistant (rif^r) mutants of *S. curacoi* NBRC 12761^T were constructed by spontaneous mutation, according to the previous report.[110] The extract of rif^r mutants of *S. curacoi* (30 strains) were analyzed by HPLC to evaluate curacozole production compared to wild type (WT) strain. As a result, three rif^r mutants harboring a 1298 C>T (Ser433Leu: S433L) mutation in RNA polymerase β subunit coding gene displayed curacozole production of 3-5 times higher than WT strain (Fig. 2.2). Furthermore, all curacozole-overproducing rif^r mutants were found to contain S433L mutation in the *rpoB* gene, corresponding to the report of Hu et al. (2002) which indicated that *S. lividans* with S433L mutation in *rpoB* gene overproduced blue-pigmented antibiotic actinorhodin.[111] These results suggested that the *rpoB* S433L mutation can improve productivity of secondary metabolites in a variety of *Streptomyces* spp.

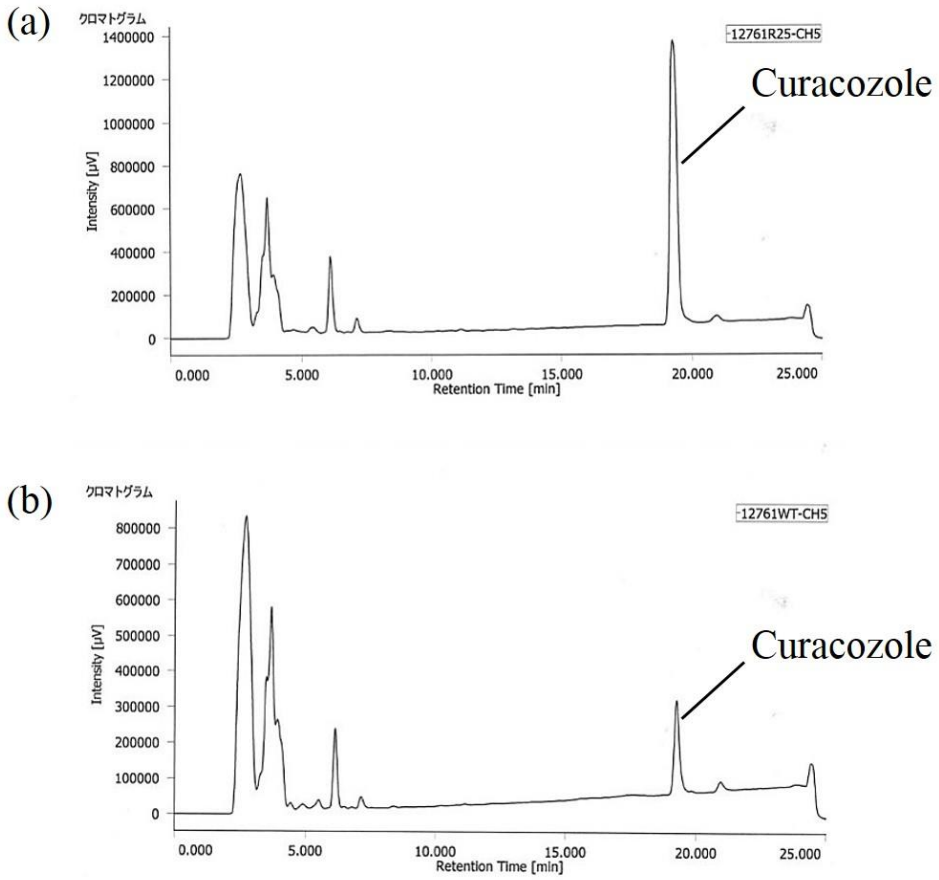


Figure 2.2 HPLC chromatogram of extract of *S. curacoi* NBRC 12761^T; (a) rifampicin-resistant (rif^r) mutant and (b) wild type strain

2.3.3 Isolation of curacozole

Streptomyces curacoi rif^r mutant strain number 25, also called *S. curacoi* R25, was cultured in 2 L of ISP2 agar medium and incubated at 30˚C for 7 days. After cultivation, *S. curacoi* R25 cells were harvested and extracted with MeOH, followed by evaporation using rotary evaporator. The extract was subjected to open column chromatography eluted with 10%, 60% and 100% MeOH, respectively. The peptide curacozole was detected in 100% MeOH fraction by HPLC at the retention time 12.0 min. To isolate curacozole, the 100% MeOH fraction was repeatedly subjected to HPLC separation using isocratic elution mode, 73% MeCN/water containing 0.05% TFA at the flow rate of 1 ml/min (Fig. 2.3).

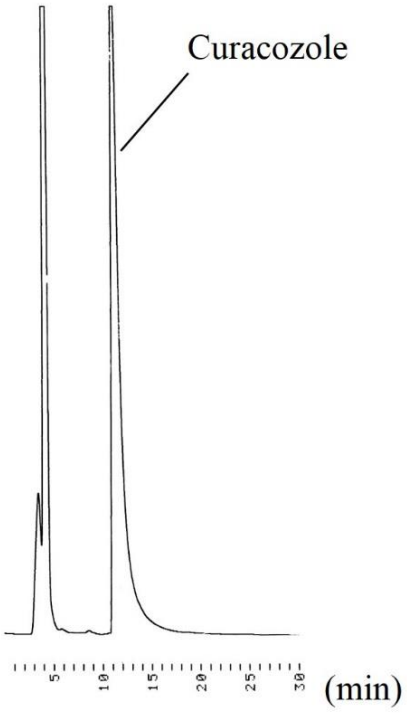


Figure 2.3 HPLC chromatogram of 100% MeOH extract of *S. curacoi* R25. The 100 µl of the extract was analyzed using an isocratic elution mode; 73% MeCN/water containing 0.05% TFA at flow rate of 1 ml/min

2.3.4 Structure determination of curacozole

Structure of curacozole was elucidated by combination of ESI-MS and NMR analyses. The ions corresponding to $[M+H]^+$ and $[2M+H]^+$ were observed at *m/z* 741.20 and 1481.39, respectively, by ESI-TOF-MS as shown in Fig. 2.4. By accurate mass analysis, the ion corresponding to $[M+H]^+$ was observed at *m/z* 741.2284 (calculated *m/z* value: 741.2277) (Fig. 2.5). Therefore, molecular formula of curacozole was established to be $C_{36}H_{36}N_8O_6S_2$. NMR analyses including ^1H, ^{13}C, DEPT-135, DQF-COSY, HSQC, TOCSY and HMBC were

performed on curacozole dissolved in DMSO-d_6 (Table 2.1 and Fig. 2.6 – 2.11). The ^{1}H NMR spectrum showed 34 proton signals including three amide protons, two aminomethines and one aminomethylene, while HSQC spectrum indicated ^{1}H-^{13}C connections. The 2D NMR correlations based on TOCSY and HMBC spectra analyses suggested the structure of curacozole as shown in Fig. 2.12. Overall, NMR spectra and chemical shift values of curacozole were very similar to those of YM-216391 (Table 2.1). However, there are three major differences between NMR spectra of curacozole and YM-216391. The chemical shift value of position 31 of curacozole (δH 8.60, δC 120.8) indicated the presence of thiazole unlike YM-216391 which contains oxazole in this position. HMBC correlations from methyl residue (H28) to C25, C26, C29 and C23 were observed, indicating that the methyl attached to carbon (C25). The H-H spin system based on TOCSY and HMBC correlations from H14 to C13 and C11 indicated the presence of Ile containing C14. Although connections between C39, C41 and C48 were not accomplished due to absence of HMBC correlation, oxazole was proposed to present in this position based on similarity of chemical shifts between curacozole and YM-216391 (Table 2.1).

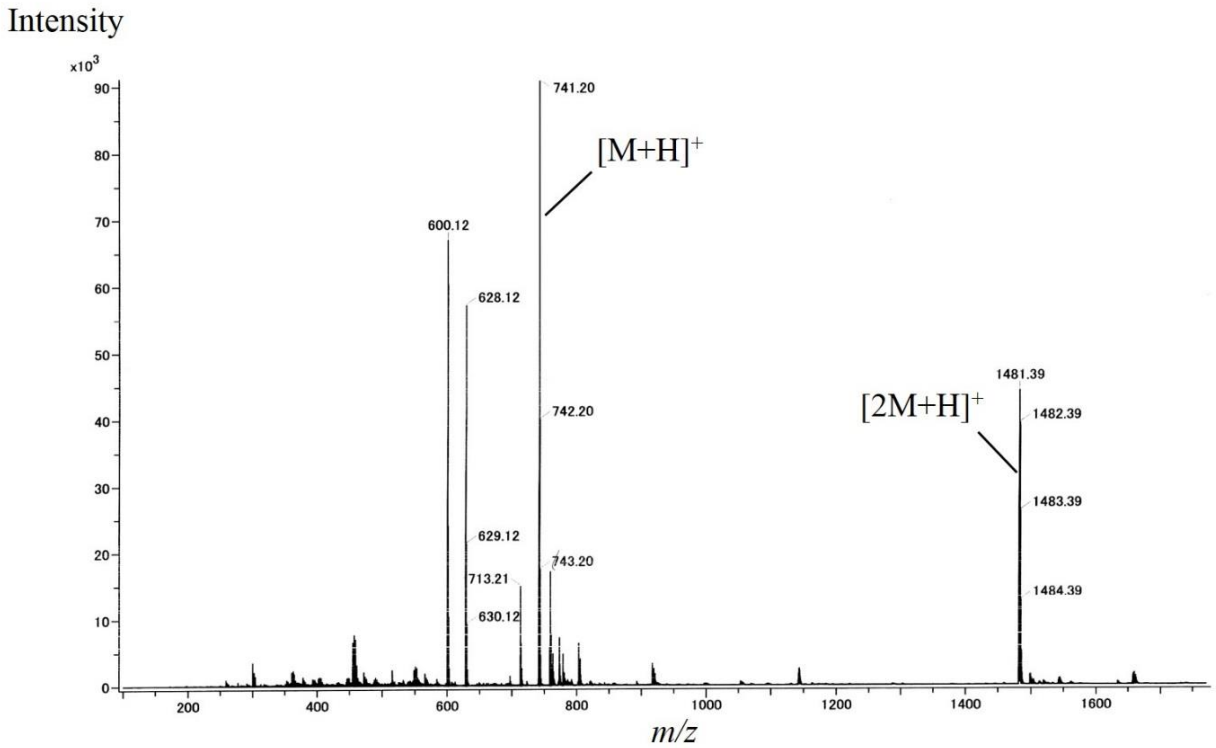


Figure 2.4 ESI-TOF-MS of curacozole. The $[M+H]^+$ and $[2M+H]^+$ were observed at *m/z* 741.20 and 1481.39, respectively

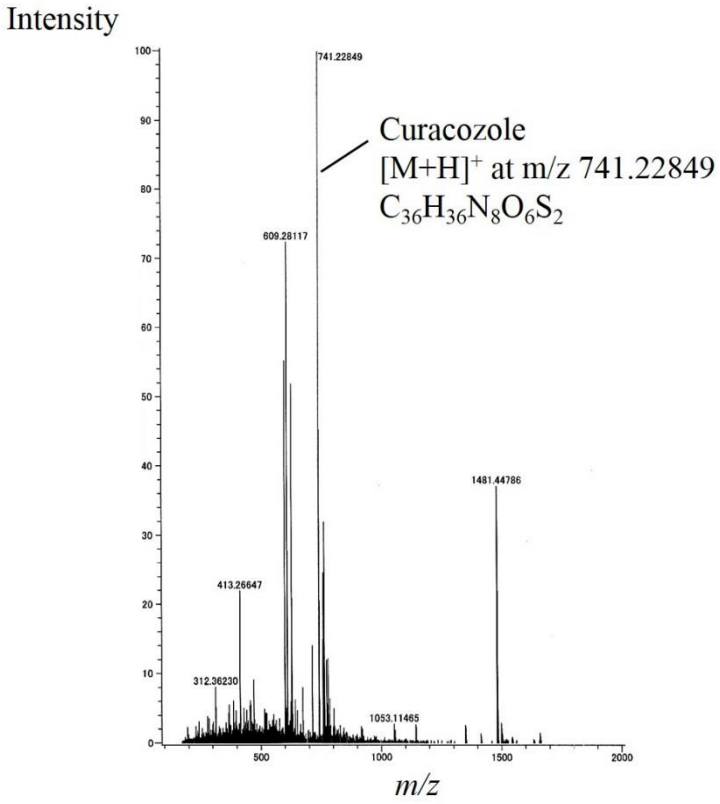


Figure 2.5 Accurate mass analysis of curacozole. The $[M+H]^+$ was observed at *m/z* 741.22849

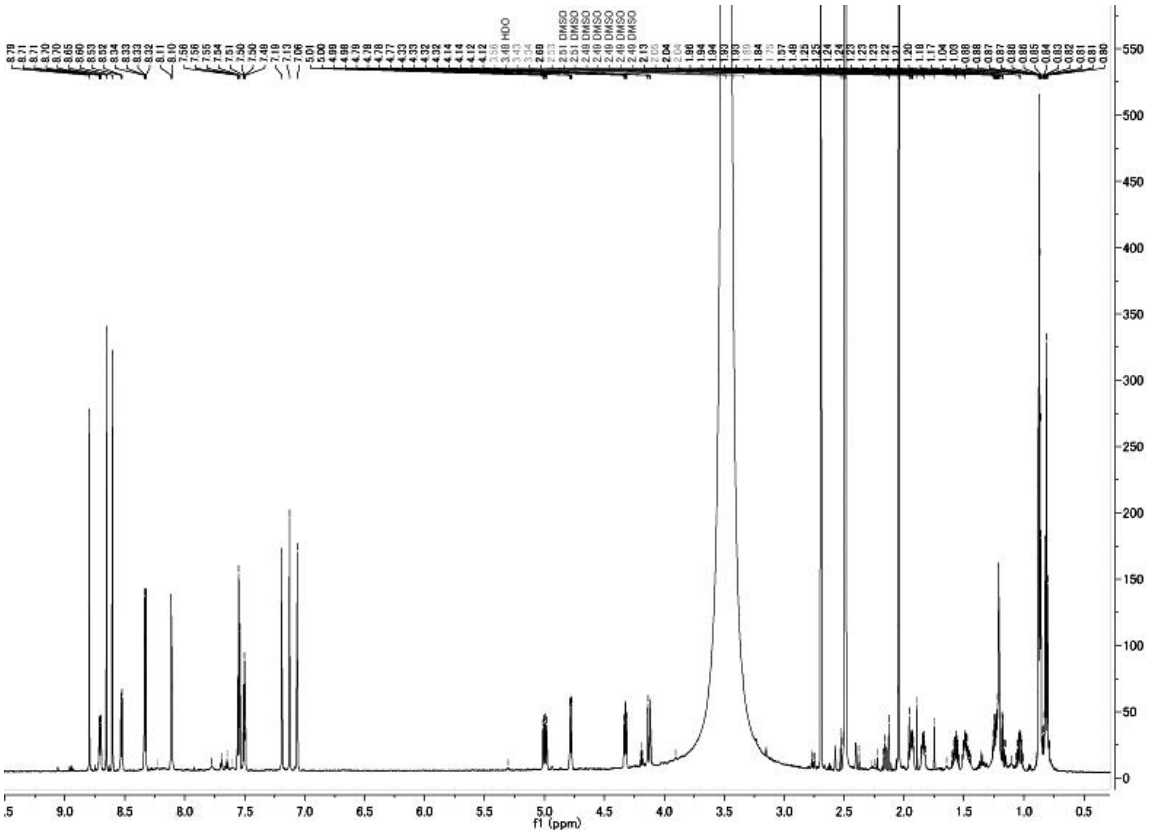


Figure 2.6 ^{1}H NMR spectrum of curacozole

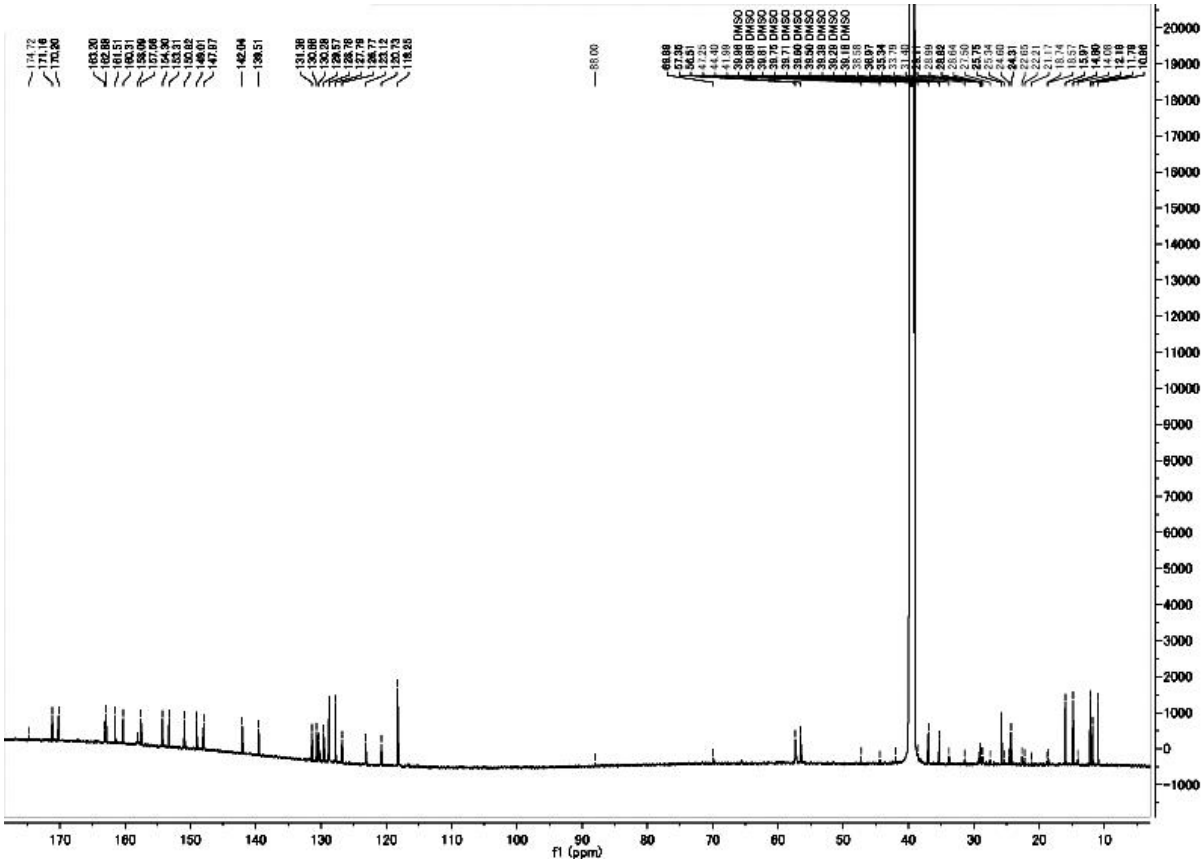


Figure 2.7 ^{13}C NMR spectrum of curacozole

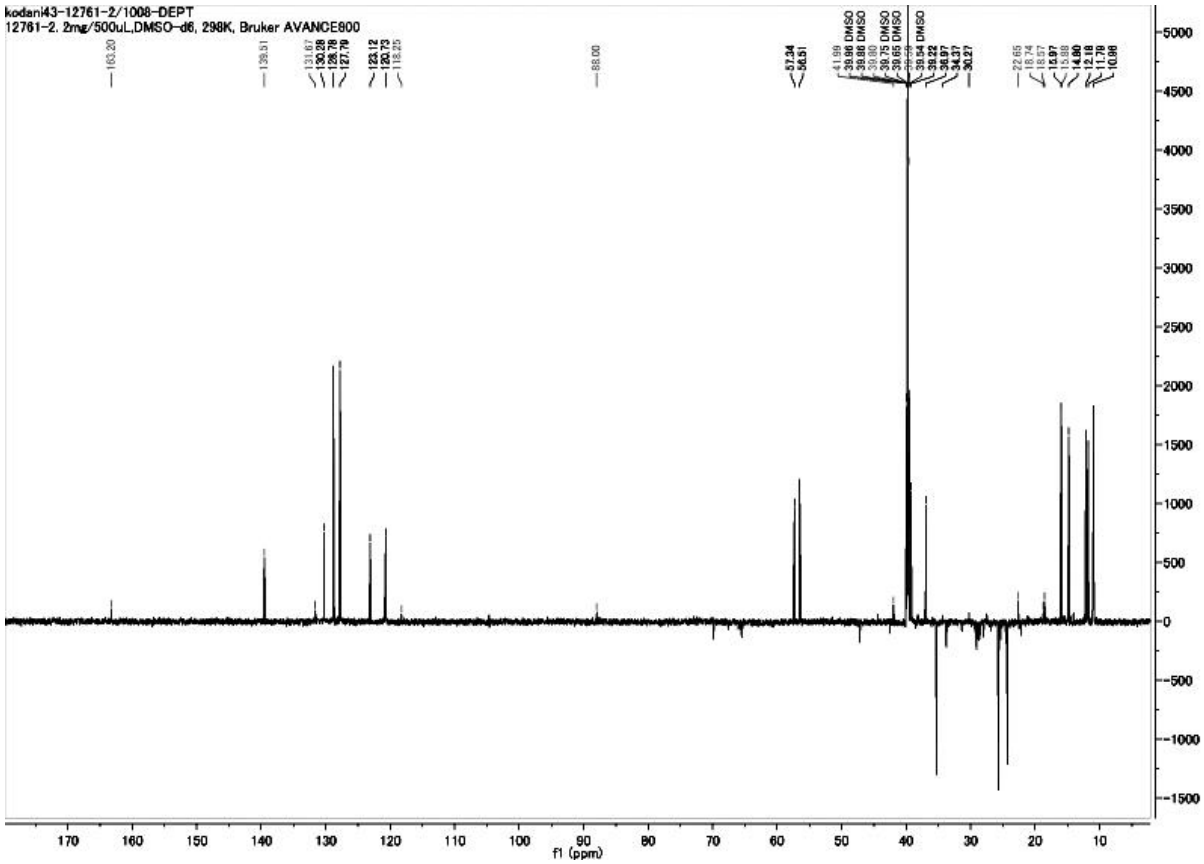


Figure 2.8 DEPT-135 spectrum of curacozole

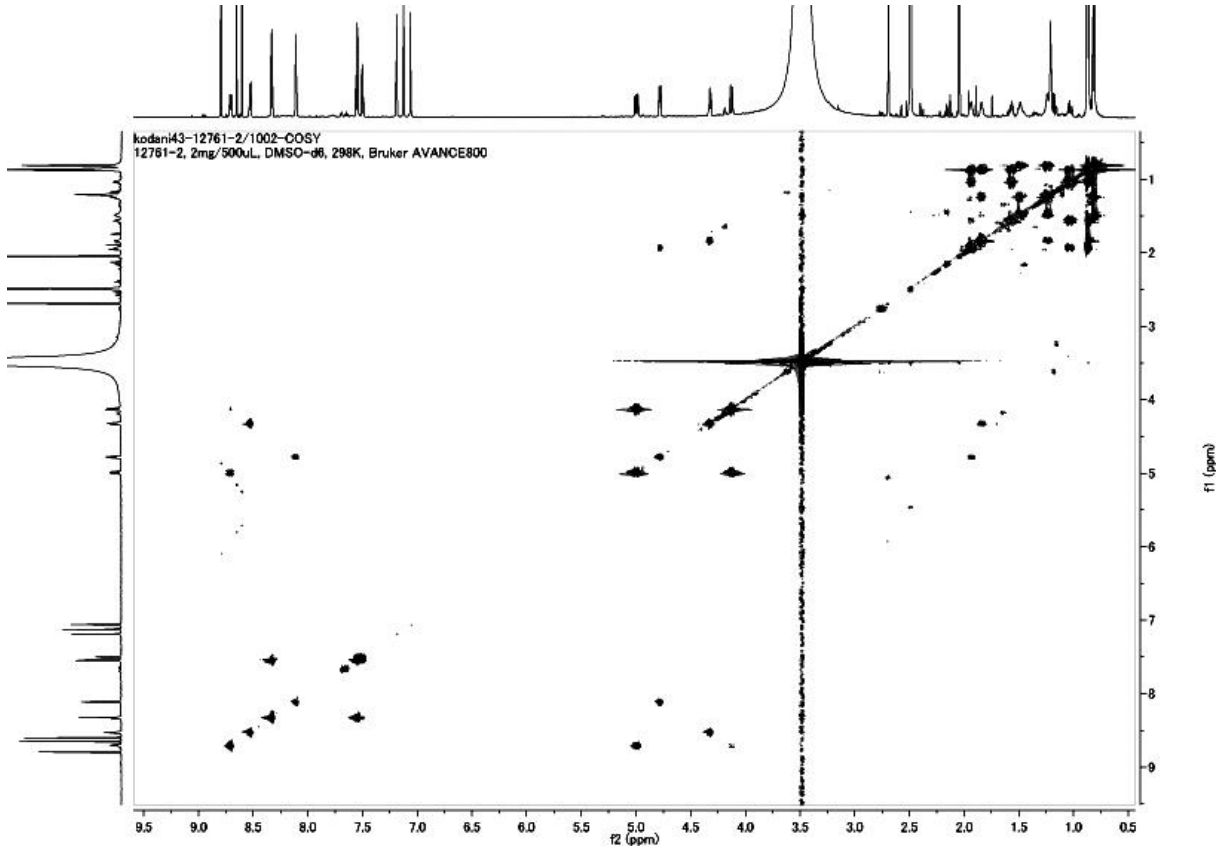


Figure 2.9 DQF-COSY spectrum of curacozole

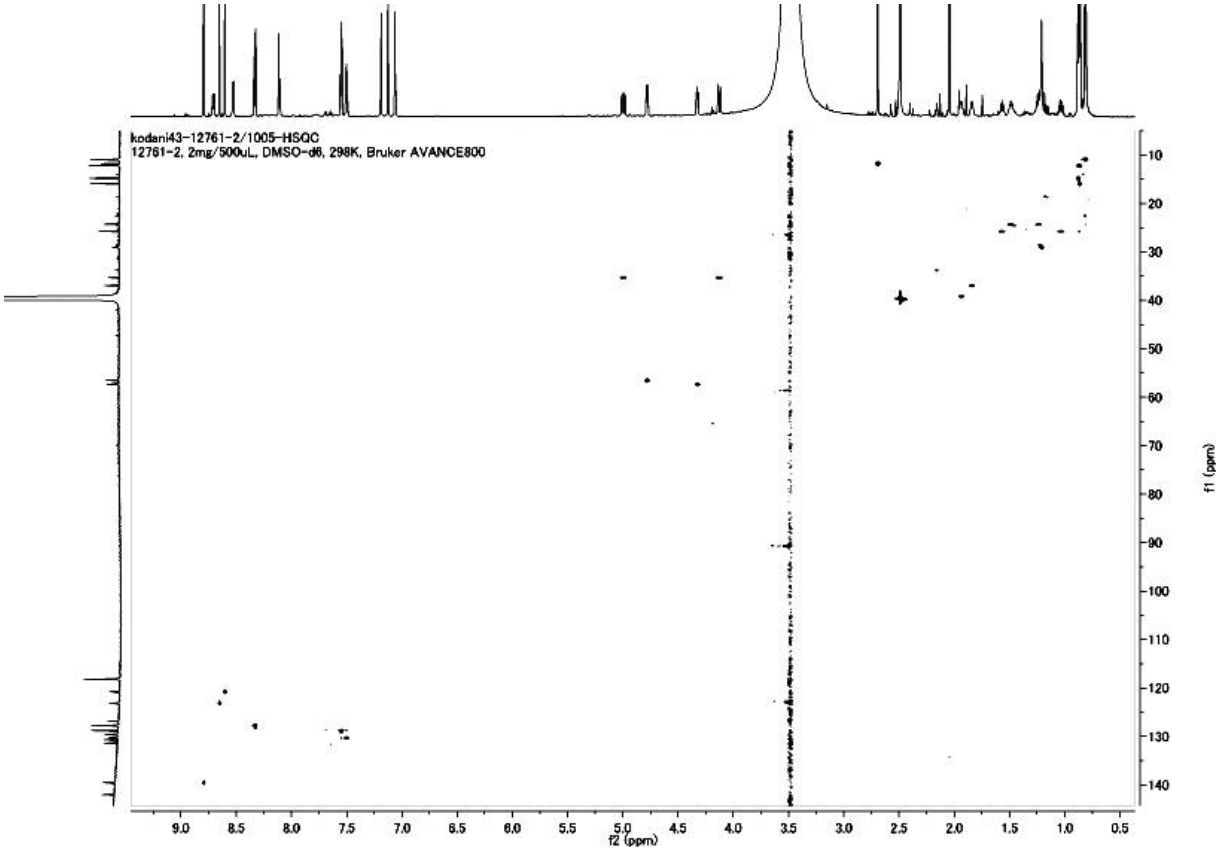


Figure 2.10 HSQC spectrum of curacozole

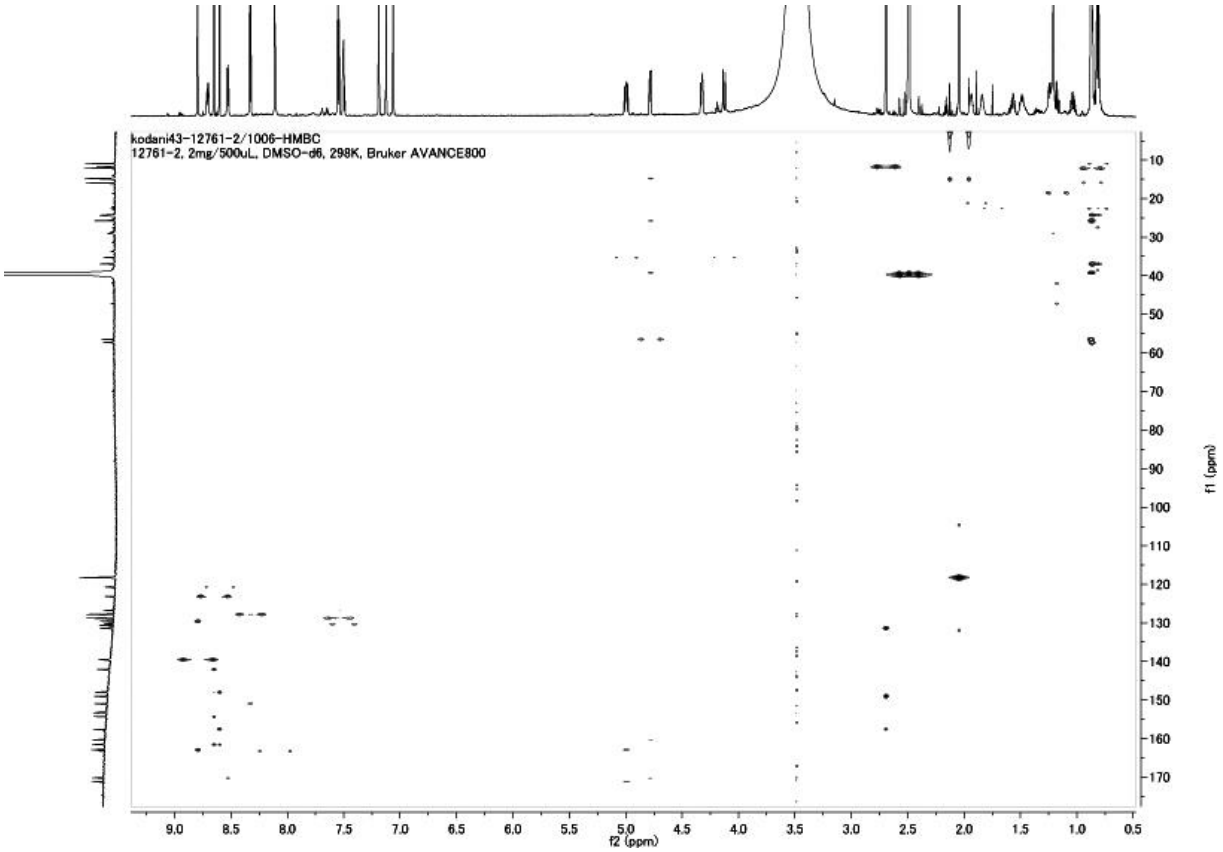


Figure 2.11 HMBC spectrum of curacozole

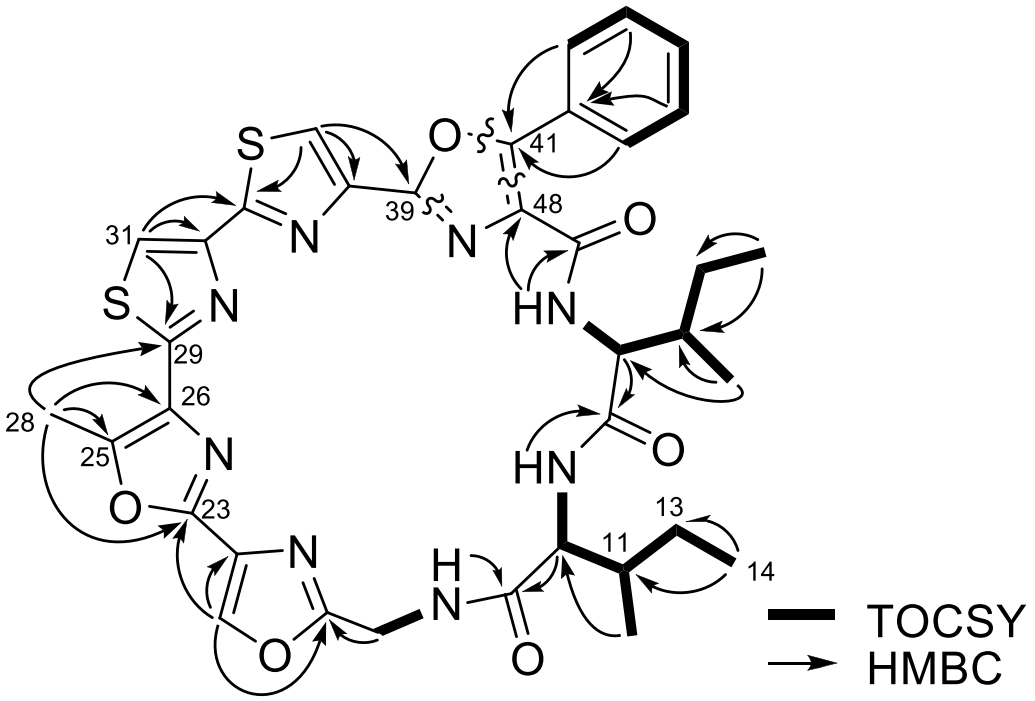


Figure 2.12 Selected 2D NMR correlations of curacozole

Table 2.1 NMR chemical shift values of curacozole (**1)** and YM-216391(**2**) in DMSO-d_6

Curacozole (**1**)			YM-216391(**2**)[#]		
Position	δH (*J* = Hz)	δC	Position	δH (*J* = Hz)	δC
1		160.3	1		160.3
2	8.10, 1H (d, 7.7)		2	8.22, 1H (d, 6.5)	
3	4.78, 1H (dd, 7.7, 4.2)	56.5	3	4.82, 1H (dd, 6.5, 3.5)	57.1
4	1.94, 1H (m)	39.2	4	2.09, 1H (m)	38.8
5	0.87, 3H (ov*)	14.8	5	0.95, 3H (d, 7.5)	14.7
6	1.57, 1H (m)	25.8	6	1.66, 1H (m)	25.6
	1.03, 1H (m)			1.09, 1H (m)	
7	0.87, 3H (ov*)	12.2	7	0.91, 3H (t, 7.5)	12.1
8		170.2	8		170.2
9	8.53, 1H (d, 8.5)		9	8.57, 1H (d, 9.0)	
10	4.32, 1H (dd, 8.5, 6.5)	57.4	10	4.60, 1H (dd, 9.0, 4.5)	57.5
11	1.84, 1H (m)	37.0	11	2.14, 1H (m)	31.5
12	0.86, 3H (ov*)	16.0	12	0.97, 3H (d, 6.5)	17.4
13	1.23, 1H (m)	24.3	13	0.93, 3H (d, 6.5)	19.7
	1.49, 1H (m)				
14	0.81, 3H (t, 7.3)	11.0			
15		171.2	14		170.8
16	8.71, 1H (dd, 9.1, 2.5)		15	8.67, 1H (dd, 9.0, 2.0)	
17	4.99, 1H (dd, 16.6, 9.1)	35.9	16	5.05, 1H (dd, 17.0, 9.0)	35.2
	4.12, 1H (dd, 16.6, 2.3)			4.19, 1H (dd, 16.5, 2.5)	
18		162.9	17		163.0
20	8.79, 1H (s)	139.5	19	8.89, 1H (s)	139.6
21		129.6	20		129.1
23		153.4	22		155.5
25		149.1	24	8.98, 1H (s)	139.1
26		131.4	25		129.9
28	2.69, 3H (s)	11.7			
29		157.6	27		155.0
31	8.60, 1H (s)	120.8	29	9.07, 1H (s)	139.4
32		148.0	30		135.6
34		161.5	32		157.5
36	8.65, 1H (s)	123.1	34	8.65, 1H (s)	122.2
37		142.1	35		141.5
39		154.3	37		154.1
41		150.9	39		150.6
42		126.7	40		126.7
43/47	8.33, 1H (d, 7.5)	127.8	41/45	8.35, 1H (d, 7.5〉	127.5
44/46	7.55, 1H (t, 7.5)	128.8	42/44	7.57, 1H (t, 7.0)	128.6
45	7.50, 1H (t, 7.5)	130.3	43	7.52, 1H (t, 7.0)	130.0
48		130.6	46		130.8

ov*: overlapped, #chemical shift values of **2** was excerpted from Sohda et al. (2005)[105]

2.3.5 Modified Marfey's analysis of curacozole

Modified Marfey's analysis was performed to determine absolute stereochemistries of two Ile groups in curacozole. The hydrolysate of curacozole was derivatized using *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA) and the derivatives were analyzed by HPLC, compared to derivatives of standard amino acids. Retention times (min) of L- or D-FDLA derivatized standard amino acids were as follows; L-allo-Ile-L-FDLA (41.20 min), L-allo-Ile-D-FDLA (58.16 min), L-Ile-L-FDLA (42.22 min) and L-Ile-D-FDLA (58.16 min) (Fig. 2.13 and 2.14). The modified Marfey's analysis on curacozole showed the presence of L-Ile and the mixture containing D-Ile and D-allo-Ile in 1: 1 molar ratio (Fig. 2.15).

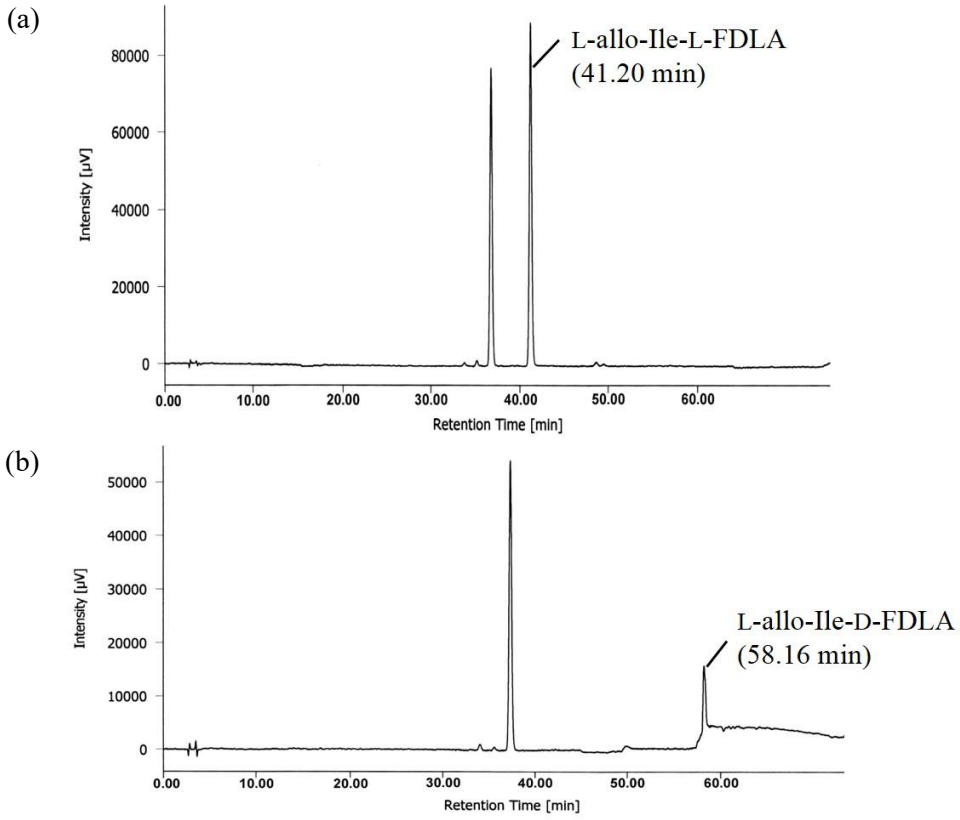


Figure 2.13 HPLC chromatograms of (a) L-allo-Ile-L-FDLA and (b) L-allo-Ile-D-FDLA

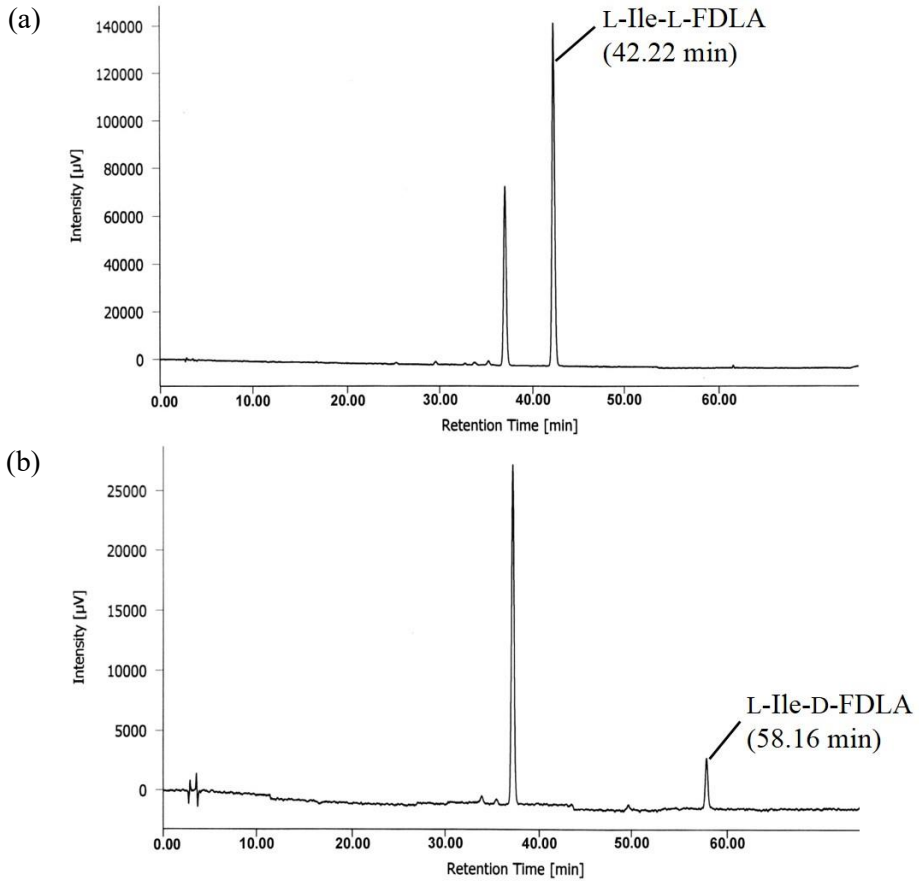


Figure 2.14 HPLC chromatograms of (a) L-Ile-L-FDLA and (b) L-Ile-D-FDLA

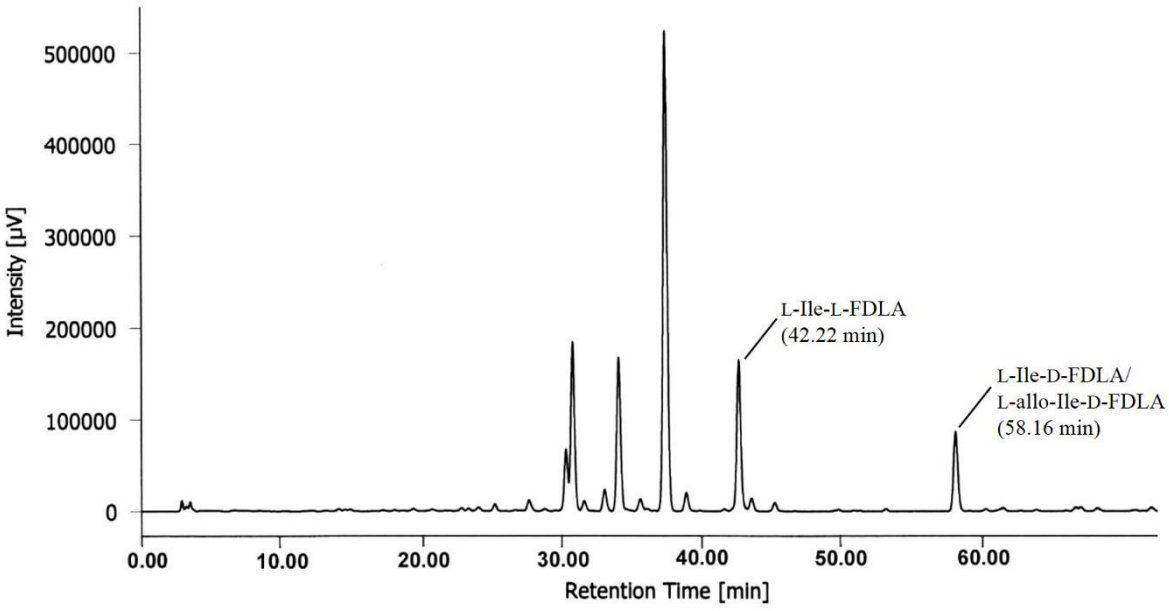


Figure 2.15 HPLC chromatogram of L-FDLA derivative of curacozole hydrolyzed with 6N HCl

2.3.6 Chiral HPLC analysis of curacozole

Since D-Ile and D-allo-Ile could not be distinguished due to identical retention times in modified Marfey's analysis, chiral HPLC analysis was performed on the hydrolysate of curacozole to clarify the proportions of D-Ile and D-allo-Ile according to the previous report.[105] Retention times (min) of standard amino acids were as follows; D-allo-Ile (13.00 min) and D-Ile (15.90 min) (Fig. 2.16). Chiral HPLC analysis of curacozole revealed the presence of D-allo-Ile and D-Ile in 6: 4 molar ratio (Fig. 2.17). Overall, chemical structure of curacozole was determined as shown in Fig. 2.18.

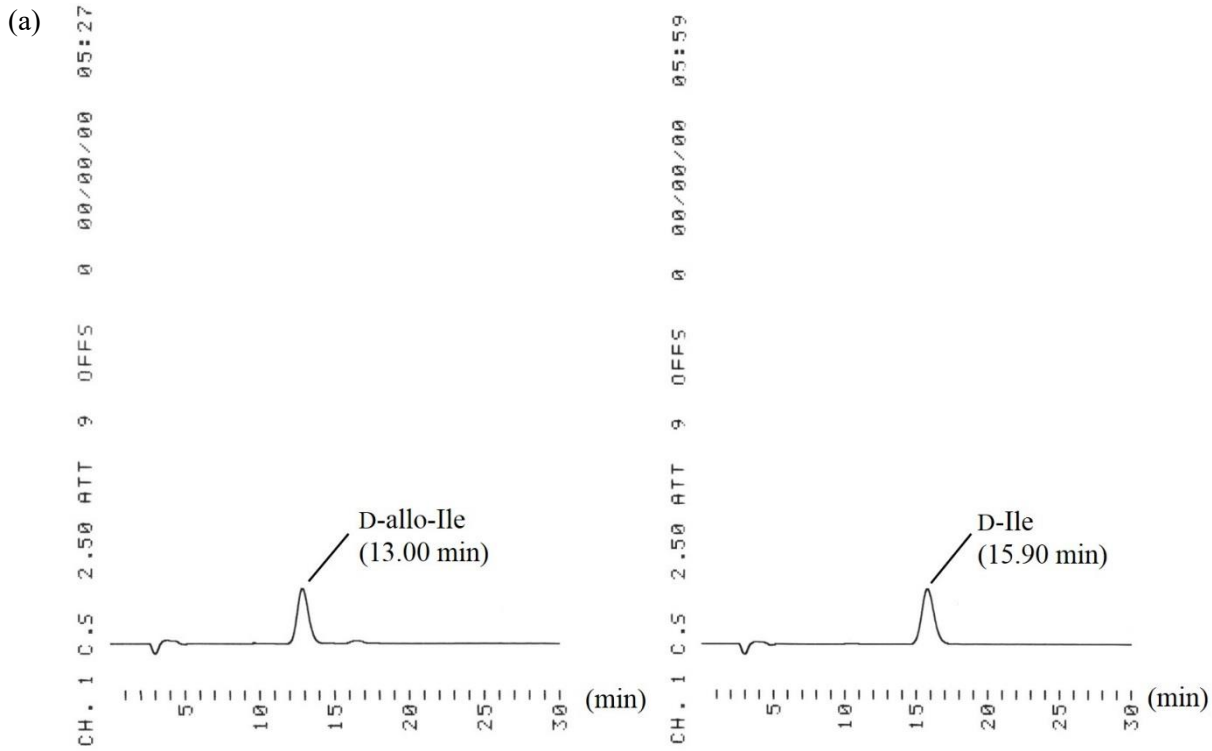


Figure 2.16 Chiral analysis of (a) D-allo-Ile and (b) D-Ile

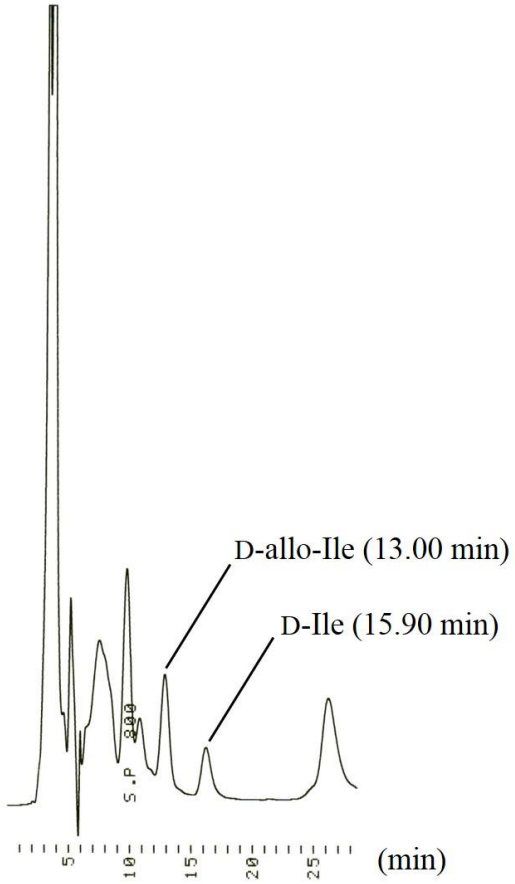


Figure 2.17 Chiral HPLC analysis of hydrolysate of curacozole

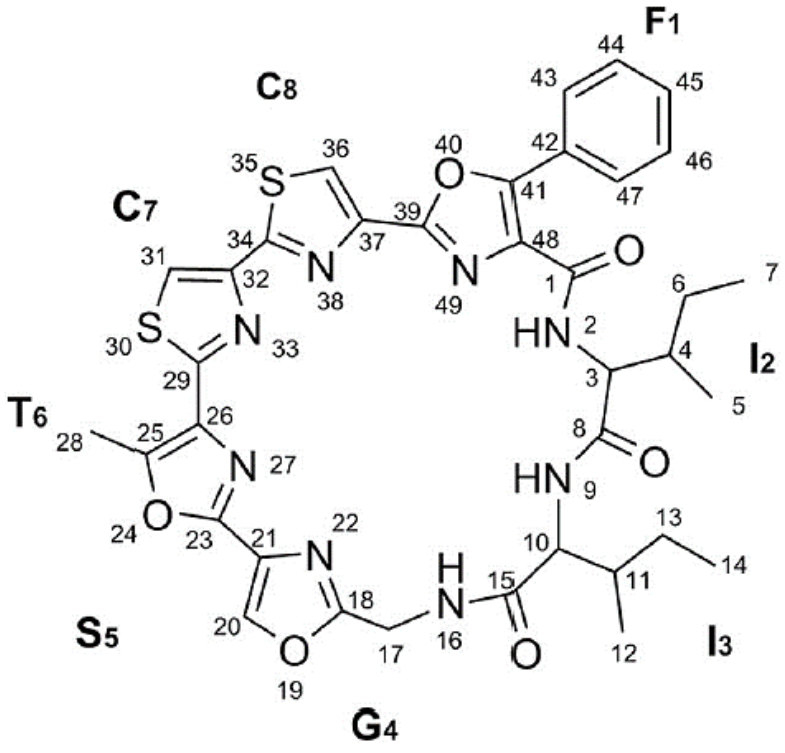


Figure 2.18 Chemical structure of curacozole; bold letters indicate precursor amino acids

2.3.7 Cytotoxic assay of curacozole

To evaluate cytotoxicity of curacozole, cancer cell lines including HCT116 and HOS cells were incubated with various concentrations of curacozole for 72 h and cell viability was measured by the CellTiter-Glo luminescent cell viability assay. As a result, a dose-dependent cytotoxicity against the HCT116 and HOS cells were observed. Curacozole exhibited cytotoxicity with IC_{50} values of 8.6 nM and 10.5 nM for HCT116 and HOS cells, respectively (Fig. 2.19). These results indicated that curacozole is highly toxic towards HCT116 and HOS cancer cells. Moreover, curacozole at the concentration of 50 nM could induce HCT116 cancer cell apoptosis within 24 h, as shown in Fig. 2.20.

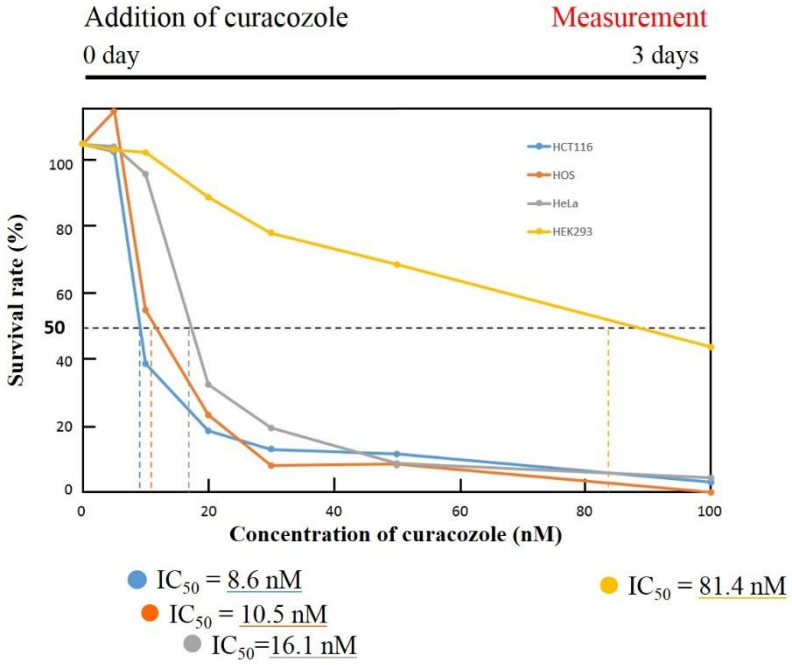


Figure 2.19 Cytotoxic assay of curacozole against cancer cells

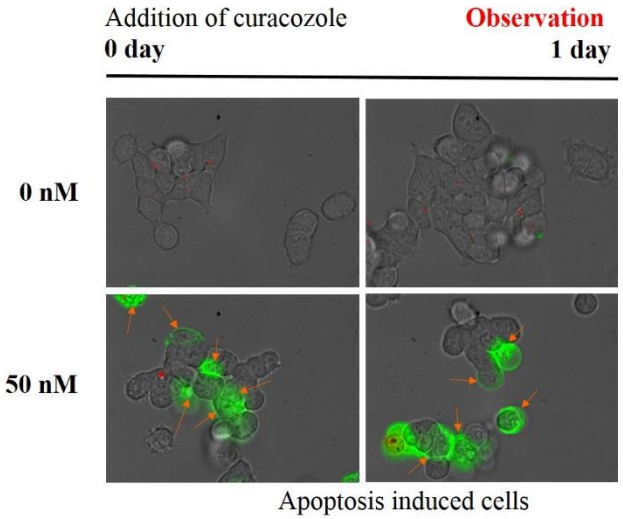


Figure 2.20 Curacozole induced HCT116 cell apoptosis

2.3.8 Biosynthetic gene cluster of curacozole

By BLASTp search, a new precursor peptide coding gene (accession number: WP_107116988.1), analogous to YM-216391 precursor peptide coding gene, was found in the genome sequence of *Streptomyces curacoi* and the peptide named curacozole was successfully isolated from *S. curacoi* NBRC 12761^{T}. Therefore, biosynthetic gene cluster (BGC) containing a new precursor peptide coding gene was proposed to be responsible for biosynthesis of curacozole. The gene cluster consists of ten genes which shares gene organization similar to that of YM-216391 (Fig. 2.21).[106] The genes, excepted for AQI70_RS14565, are orthologues of *ymI*, *ymA*, *ymD*, *ymE*, *ymB1*, *ymC1*, *ymF*, *ymBC* and *ymR1* genes responsible for biosynthesis of YM-216391. The amino acid sequences translated from those genes showed 69% to 97% similarity to Ym proteins (Table 2.2). Due to the similarity between curacozole and YM216391, the biosynthetic pathway of curacozole was proposed based on YM-216391 biosynthetic pathway, as shown in Fig. 2.22. A linear precursor peptide containing 44 amino acids (AQI70_RS38515) is biosynthesized via RiPP system. Docking protein (AQI70_RS14600) and cyclodehydratase-oxidoreductase enzyme (AQI70_RS37415) convert G-S-T-C-C in core peptide region into Gly-oxazole-methyloxazole-thiazole-thiazole moiety. P-450 hydroxylase (AQI70_RS14595) hydroxylates Phe residue in the intermediate peptide. Protease (AQI70_RS14580) cleaves a core peptide, followed by N-C terminal cyclization. A hypothetical protein (AQI70_RS14605) might be required for cyclic peptide formation. Cyclodehydratase (AQI70_RS14590) cyclizes thiazole-β-hydroxy-Phe and FAD-binding monooxygenase (AQI70_RS14585) oxidizes the ring resulting in a thiazole-phenyloxyazole moiety formation. Interestingly, the proposed gene cluster in *S. curacoi* does not contain epimerase coding gene homologous to YmG, however, one of the two Ile in curacozole was found to be D-Ile/D-allo-Ile. The epimerase responsible for conversion of L-Ile to D-Ile/D-allo-Ile still remains to be identified in the gene cluster. According to cypemycin BGC, CypI was reported to isomerize L-Ile residues.[112] The gene encoded homolog of CypI (AQI70_17525) was found in the genome sequence of *S. curacoi* at the locus distant from the gene cluster, and was proposed to catalyze isomerization of Ile residue in curacozole.

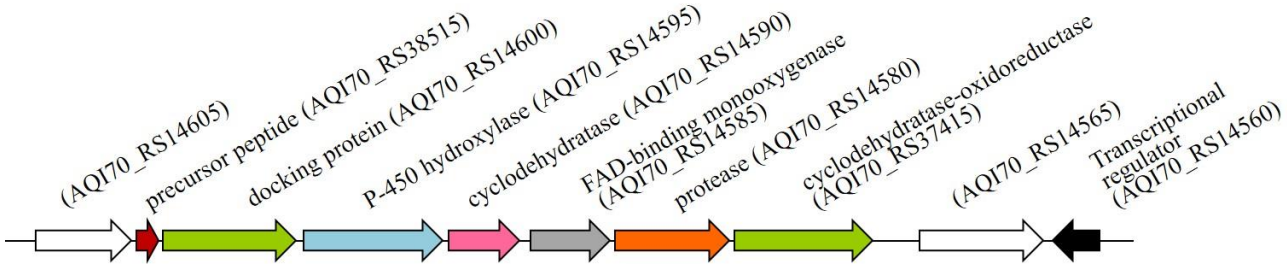


Figure 2.21 Biosynthetic gene cluster of curacozole

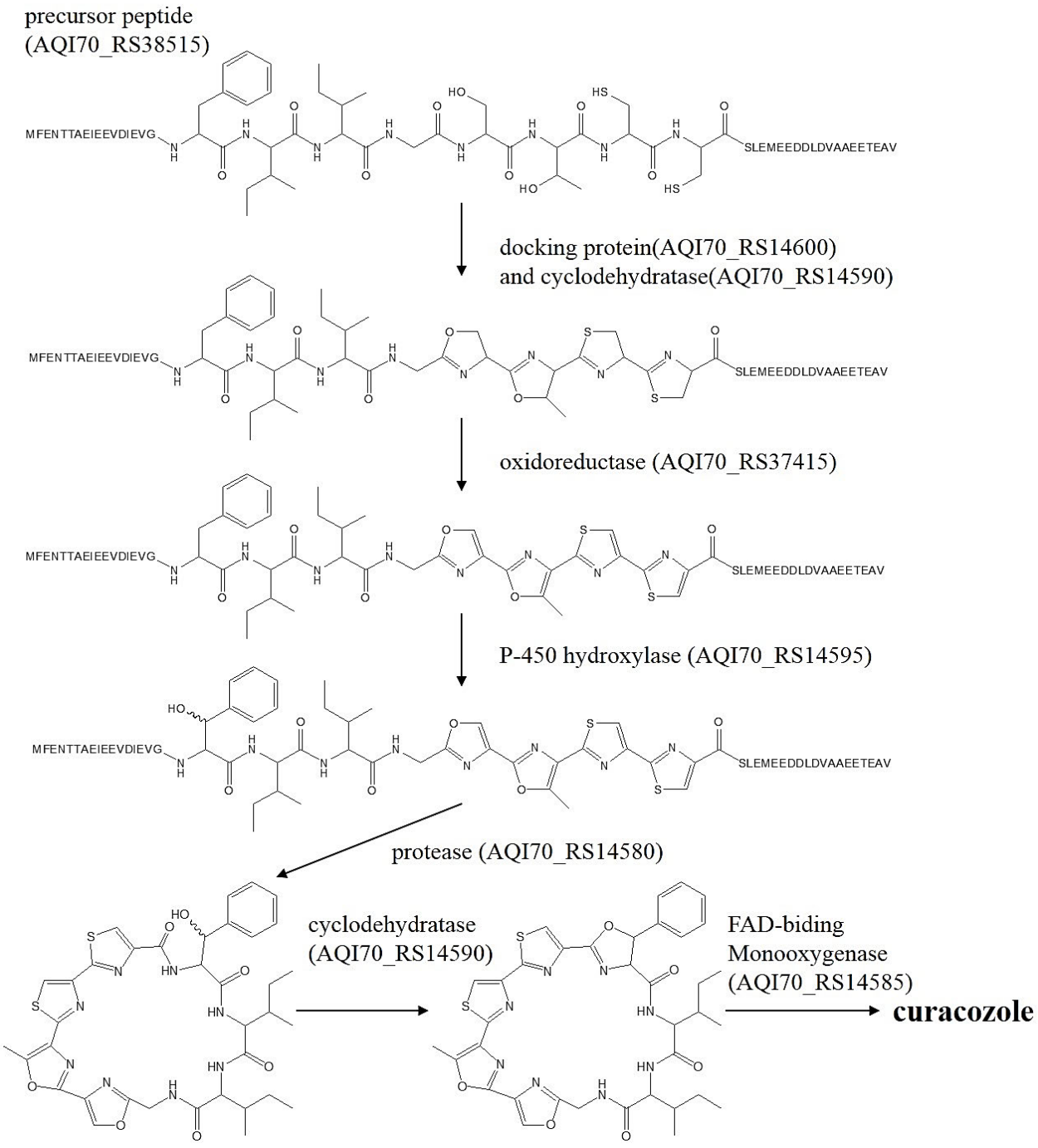


Figure 2.22 Biosynthetic pathway of curacozole

Table 2.2 Comparison of amino acid sequences of biosynthetic genes for biosynthesis of curacozole and YM-216391

Putative function	Biosynthetic gene for YM-216391	Biosynthetic gene for curacozole	Similarity (%)
precursor peptide	Gene name: *ymA* Locus tag: AFJ68074.1 Accession number: JN411915.1 (36 amino acids)	Locus tag: AQI70_RS38515 Accession number: WP_107116988.1 (36 amino acids)	86
docking protein	Gene name: *ymD* Locus tag: AFJ68075.1 Accession number: JN411915.1 (476 amino acids)	Locus tag: AQI70_RS14600 Accession number: WP_079051299.1 (470 amino acids)	74
P-450 hydroxylase	Gene name: *ymE* Locus tag: AFJ68076.1 Accession number: JN411915.1 (500 amino acids)	Locus tag: AQI70_RS14595 Accession number: WP_062148882.1 (469 amino acids)	66
cyclodehydratase	Gene name: *ymB1* Locus tag: AFJ68086.1 Accession number: JN411915.1 (261 amino acids)	Locus tag: AQI70_RS14590 Accession number: WP_062148878.1 (261 amino acids)	85
FAD-binding monooxygenase	Gene name: *ymC1* Locus tag: AFJ68077.1 Accession number: JN411915.1 (318 amino acids)	Locus tag: AQI70_RS14585 Accession number: WP_062148875.1 (287 amino acids)	59
protease	Gene name: *ymF* Locus tag: AFJ68078.1 Accession number: JN411915.1 (381 amino acids)	Locus tag: AQI70_RS14580 Accession number: WP_062148872.1 (362 amino acids)	69
cyclodehydratase-oxidoreductase	Gene name: *ymBC* Locus tag: AFJ68085.1 Accession number: JN411915.1 (492 amino acids)	Locus tag: AQI70_RS37415 Accession number: WP_062148869.1 (465 amino acids)	67
transcriptional regulator	Gene name: *ymR1* Locus tag: AFJ68079.1 Accession number: JN411915.1 (239 amino acids)	Locus tag: AQI70_RS14560 Accession number: WP_062148863.1 (138 amino acids)	65
epimerase	Gene name: *ymG* Locus tag: AFJ68082.1 Accession number: JN411915.1 (210 amino acids)	not detected	-

2.4 Summary

A new precursor peptide coding gene was found in the genome sequence of *S. curacoi* by genome mining. The peptide named curacozole was isolated from the extract of rif^r mutant of *S. curacoi* NBRC 12761^T. Chemical structure of curacozole was determined to be macrocyclic peptide containing two isoleucine, two thiazole and three oxazole, similar to YM-216391. Curacozole showed potent cytotoxicity against HCT116 and HOS cancer cells as well as induced HCT116 cell apoptosis. The biosynthetic gene cluster for curacozole was identified in the genome sequence of *S. curacoi* based on similarity to *ym* gene cluster which is responsible for biosynthesis of YM-216391. To the best of our knowledge, this is the first report of a new macrocyclic peptide containing sequential oxazole/methyloxazole/thiazole ring discovered from actinomycetes by genome mining.

Chapter III

Isolation and structure determination of a new antibacterial peptide, pentaminomycin C, from *Streptomyces cacaoi* subsp. *cacaoi*

3.1 Introduction

Natural products have been considered as abundant source of bioactive compounds for pharmaceutical applications. In the past, bacterial and fungal natural products research was mainly focused on peptide of nonribosomal origin. Nonribosomal peptides (NRPs) are a class of peptide secondary metabolites synthesized by nonribosomal peptide synthetases (NRPSs), a large multienzyme machineries that assemble a numerous peptides.[52] NRP biosynthesis generally occurs in a several steps, including the processes of chain initiation, elongation and termination on the assembly line as well as a post translational modification by tailoring enzymes, resulting in the mature peptide scaffolds.[52, 54, 55]

The module in an NRPS is responsible for the incorporation of one specific amino acid to build up the final mature peptide. Each module normally consists of a set of three domains, including adenylation (A) domain responsible for substrate recognition, peptidyl carrier protein (PCP) domain for transport to the respective catalytic centers and condensation (C) domain for the formation of peptide bond. After the linear peptide chain has reached the final module, the peptide is released by the function of a thioesterase (TE) domain which exists at the end of the NRPS assembly line. Some modules contain epimerization (E) domain to carry out epimerization of the C_{α}-carbon of the PCP-tethered aminoacyl substrate to afford D- or L-configuration equilibrium.[52] In addition, several modification enzyme coding genes were found located nearby NRPS biosynthetic gene cluster (BGC) and further modifications of peptides such as halogenation[56, 113], glycosylation, and hydroxylation were found to be occurred during the final step of NRP biosynthesis.

In contrast to peptides synthesized by ribosomal origin, the NRPs contain not only the common 20 amino acids, but hundreds of amino acid building blocks. Therefore, the NRP chemical structures reveal the diversity and complexity which relate to its biological functions.[57-59] The relationship between structures and biological activities of many NRPs have been investigated. For instance, vancomycin, a glycopeptide antibiotic produced by *Streptomyces*, possesses a unique structure and exhibited biological activity inhibiting bacterial peptidoglycan cross-linking in bacterial cell wall biosynthesis.[59, 63, 64] Moreover, several commercial antibiotics (e.g. chloramphenicol[60], daptomycin[61] and teicoplanin[62]), antitumors (actinomycin D[65] and bleomycin A2 and B2[66]), and immunosuppressants

(cyclosporine A[67]) have been reported to biosynthesized via NRPS system.[52]

Nonribosomal peptides (NRPs) possess diverse chemical structures including a linear peptides and cyclic peptides.[114] Cyclic peptides derived from microorganisms have been reported to exhibit a wide variety of biological activities. A cyclic pentapeptides, malformins, are a group of peptides isolated from *Aspergillus* sp. as inducers of malformations in bean plants and curvatures in corn roots.[115, 116] Malfornins (e.g. malformin A1, C and E) showed strong cytotoxic, fibrinolytic and antimicrobial activities.[117-119] Kakeromamide A, a cyclic pentapeptide isolated from marine cyanobacterium showed inducing activity in the differentiation of neural stem cells into astrocytes *in vitro*.[120] Cyclic pentapeptides, BE-18257A and B, were isolated from *Streptomyces misakiensis* BA18257 as endothelin-binding inhibitors.[121, 122] Recently, pentaminomycins A and B, a hydroxyarginine containing cyclic pentapeptides, were isolated from *Streptomyces* sp. RK88-1441. Pentaminomycin A showed an antimelanogenic activity against alpha-melanocyte stimulating hormone (α-MSH)-stimulated B16F10 melanoma cells by suppressing the expression of melanogenic enzymes, tyrosinase-related protein-1 (TRP-1) and tyrosinase-related protein-2 (TRP-2).[123]

By chemical screening on extracts of actinobacteria, we found a new cyclic pentapeptide, named pentaminomycin C, and a known peptide BE-18257A in the extract of *Streptomyces cacaoi* subsp. *cacaoi* NBRC 12748^{T}. We successfully isolated the peptide, pentaminomycin C, and determined its structure using ESI-MS and NMR analyses. Putative BGC containing two NRPSs for pentaminomycin C and BE-18257A was clarified in the genome sequence of *S. cacaoi*. Pentaminomycin C exhibited antibacterial activity against Gram-positive bacteria. The isolation and structure determination of a new antibacterial peptide, pentaminomycin C, were described in the chapter III.

3.2 Materials and Methods

3.2.1 Bacterial strains

Bacterial strains including *Streptomyces cacaoi* subsp. *cacaoi* NBRC 12748^{T}, *Escherichia coli* NBRC 102203^{T}, *Pseudomonas aeruginosa* NBRC 12689^{T}, *Micrococcus luteus* NBRC 3333^{T}, *Staphylococcus aureus* NBRC 100910^{T} and *Bacillus subtilis* NBRC 13719^{T} were obtained from NITE Biological Resource Center (NBRC) culture collection, Japan.

3.2.2 Isolation of peptides

Streptomyces cacaoi subsp. *cacaoi* NBRC 12748^{T} was cultured in 5 L of modified ISP2 agar medium[82] containing 2 g of leucine (Leu). After incubation at 30 °C for 9 days, bacterial cells were harvested using stainless steel spatula and were extracted by adding 200 ml

of MeOH. The extract was filtered using filter paper (Whatman No. 1, GE Healthcare Life Sciences, Little Chalfont, UK), followed by centrifugation at 4000 rpm for 10 min to remove insoluble compounds. The MeOH extract was concentrated using rotary evaporator and subjected to open column chromatography (styrene-divinylbenzene polymer resin, CHP-20P, Mitsubishi Chemical Corp., Tokyo, Japan), eluted with 10%, 60%, and 100% MeOH, respectively. The 100% MeOH fraction was concentrated using rotary evaporator and injected to HPLC. The peptide was isolated by HPLC using Navi C30-5 column (4.6×250 mm, Wakopak, Wako Pure Chemical Industries, Tokyo, Japan). The UV detector was set at wave length of 220 nm. Isocratic elution mode, 37 % MeCN/water containing 0.05% TFA at flow rate of 1 ml/min, was used to isolate the peptide pentaminomycin C (Retention time: 26.70 min), along with BE-18257A.

3.2.3 ESI-MS analysis

ESI-MS analysis was performed using a JEOL JMS-T100LP mass spectrometer. Reserpine was used as an internal standard for accurate MS analysis.

3.2.4 NMR analysis

NMR sample was prepared by dissolving pentaminomycin C in 500 µl of DMSO-d_6. All NMR spectra were obtained on Bruker Avance III HD 800 spectrometers with quadrature detection in the phase-sensitive mode by States-TPPI (time-proportional phase incrementation) and in the echo-antiecho mode. One-dimensional (1D) ^{1}H, ^{13}C, and DEPT-135 spectra were recorded at 25 °C with a 12 ppm window for proton and 239 ppm or 222 ppm windows for carbon. The following 2D ^{1}H-NMR spectra were recorded at 25 °C with 12 ppm spectral widths in the t1 and t2 dimensions: 2D DQF-COSY, recorded with 512 and 1024 complex points in the t1 and t2 dimensions; 2D TOCSY with DIPSI-2 mixing sequence, recorded with mixing time of 80 ms, 512 and 1024 complex points in the t1 and t2 dimensions; 2D NOESY, recorded with mixing time of 200 ms, 256 and 1024 complex points in the t1 and t2 dimensions. 2D ^{1}H-^{13}C HSQC and HMBC spectra were acquired at 25 °C in the echo-antiecho mode or in the absolute mode, respectively. The ^{1}H-^{13}C HSQC and HMBC spectra were recorded with 1024 × 512 complex points for 12 ppm in ^{1}H dimension and 160 ppm or 222 ppm in ^{13}C dimension, respectively, at a natural isotope abundance. 2D ^{1}H-^{15}N HSQC spectrum was recorded with 1024 × 128 complex points for 12 ppm in ^{1}H dimension and 99 ppm in ^{15}N dimension at a natural isotope abundance. All NMR spectra were processed using TOPSPIN 3.5 (Bruker). Before Fourier transformation, the shifted sinebell window function was applied to the t1 and t2 dimensions. All ^{1}H and ^{13}C dimensions were referenced to DMSO-d_6 at 25 °C.

3.2.5 Modified Marfey's analysis

Marfey's analysis was carried out in a sealed vacuum hydrolysis tube. The peptide (1.0 mg) was hydrolyzed with 500 µl of hydriodic (HI) acid at 155˚C for 24 h in order to remove the hydroxy group of hydroxyarginine (OHArg) according to the previous report.[124] To recover tryptophan (Trp), the peptide (1.0 mg) was hydrolyzed with 500 µl of 6 N hydrochloric acid (HCl) containing 3% phenol at 166˚C for 25 min according to the previous report.[89] After cooling down at room temperature, the hydrolysate was completely evaporated using rotary evaporator and freeze-drying under vacuum. The hydrolysate was resuspended in 200 µl of water, followed by adding 10 µl of the solution of *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA, Tokyo Chemical Industry Co., LTD, Tokyo, Japan) in acetone (10 µg/µL). The 100 µl of 1 M sodium bicarbonate ($NaHCO_3$) was added to the hydrolysate and the mixture was incubated at 80˚C for 3 min. The mixture was cooled down to room temperature and then neutralized with 50 µl of 2 N HCl. The 1 ml of 50% MeCN/water was added to the mixture before it was subjected to HPLC. Standard amino acids (1.0 mg) were derivertized with L-FDLA or D-FDLA in the same manner. Approximately 30 µl of FDLA-derivertive peptide was analyzed by HPLC (C18 column, 4.6 × 250 mm, Wakopak Handy ODS, Wako Pure Chemical Industries, Tokyo, Japan), compared to FDLA-derivertive standard amino acids. The DAD detector (MD-2018, JASCO, Tokyo, Japan) was used to detect FDLA-derivertives, accumulating data of absorbance from 220 nm to 420 nm. For standard amino acids including leucine (Leu), valine (Val), tryptophan (Trp), arginine (Arg) and phenylalanine (Phe), HPLC analyses were performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 70 min, increasing percentage of solvent B from 25% to 60% (HPLC condition 1). HPLC analysis for Val was performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05% TFA) with an isocratic mode 40 % of solvent B for 55 min (HPLC condition 2). HPLC analysis for Trp was performed at a flow rate of 1 ml/min using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05% TFA) with an isocratic mode 45 % of solvent B for 55 min (HPLC condition 3).

3.2.6 Antibacterial activity

Antibacterial activity of peptide was assessed by minimum inhibitory concentrations (MICs) test according to the previous reports.[91] Tetracycline was used as a positive control. The 96-well microplate containing peptide and tetracycline (50 µl/well) with concentration ranging from 64 to 0.0625 µg/ml was prepared by serial twofold dilution technique using

Mueller-Hinton broth (MHB). Testing strains including *E. coli*, *P. aeruginosa*, *S. aureus*, *B. subtilis* and *M. luteus* were cultured in nutrient agar medium and incubated at 30˚C for 24 h. Testing bacteria was inoculated into 0.85% saline solution to give a final inoculum of 10^5 CFU/ml. The 5 µl of bacterial solution was transferred to 10 ml of MHB and then 50 µl of the MHB containing testing bacteria was added into each well. The 96-well microplates were incubated at 30˚C for 24 h. The MICs was determined at the lowest concentration that gave no visible growth of testing bacteria.

3.3 Results and discussion

3.3.1 Chemical investigation of actinobacteria

Chemical investigation on extract of actinobacterial cells were conducted using HPLC and ESI-MS analyses. As a result, *Streptomyces cacaoi* subsp. *cacaoi* NBRC 12748^T was found to be a producer of new bioactive peptides. To examine peptide production, *S. cacaoi* subsp. *cacaoi* NBRC 12748^T was cultured in ISP2 agar medium at 30˚C for 9 days and bacterial cells were harvested and extracted with MeOH, followed by subjecting the crude MeOH extract to HPLC analysis using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 30 min, increasing percentage of solvent B from 20% to 70%, as shown in Fig. 3.1. Each peptide was preliminary isolated by HPLC using Navi C30-5 column with an isocratic elution mode, 37 % MeCN/water containing 0.05% TFA at flow rate of 1 ml/min and was analyzed by ESI-MS. The peptides including BE-18257A and a new peptide named pentaminomycin C were detected by HPLC as peak no. 2 and 4, respectively (Fig. 3.2).

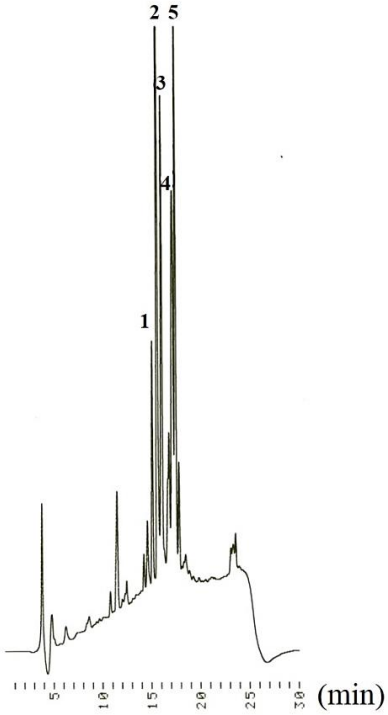


Figure 3.1 HPLC chromatogram of crude extract of *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T}; The 100 µl of the extract was analyzed using solvent A (distilled water containing 0.05% TFA) and solvent B (MeCN containing 0.05%TFA) with a linear gradient mode from 0 min to 30 min, increasing percentage of solvent B from 20% to 70%

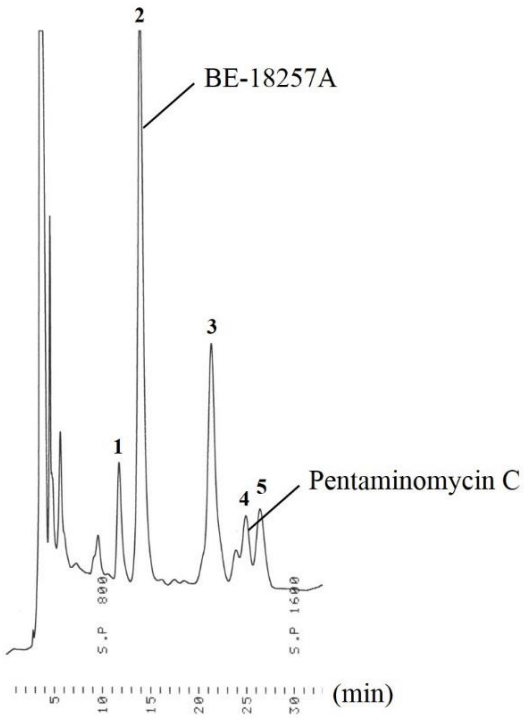


Figure 3.2 HPLC chromatogram of crude extract of *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T} using isocratic elution mode, 37% MeCN/water containing 0.05% TFA, at flow rate of 1 ml/min; BE-18257A and pentaminomycin C were detected by HPLC as peak no. **2** and **4**, respectively.

3.3.2 Isolation of pentaminomycin C

To increase the yield of pentaminomycin C, *Streptomyces cacaoi* subsp. *cacaoi* NBRC 12748^{T} was cultured in modified ISP2 agar medium containing leucine (2 g Leu, 4 g yeast extract, 10 g malt extract, 4 g glucose, and 15 g agar in 1 L distilled water, adjusted pH 7.3). After 9 days of cultivation, the harvested cells were extracted with MeOH and subjected to HPLC analysis using Navi C30-5 column with isocratic elution mode 37% MeCN/water containing 0.05% TFA at the flow rate of 1 ml/min. Pentaminomycin C was detected by HPLC at the retention time 26.70 min and HPLC chromatogram revealed the increasing of pentaminomycin C production in the extract of *S. cacaoi* subsp. *cacaoi*, as shown in Fig. 3.3. Therefore, *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T} was cultured in a large scale, 5 L of modified ISP2 agar medium containing leucine, in order to obtain pentaminomycin C (3.0 mg) and BE-18257A (5.0 mg) for structure determination.

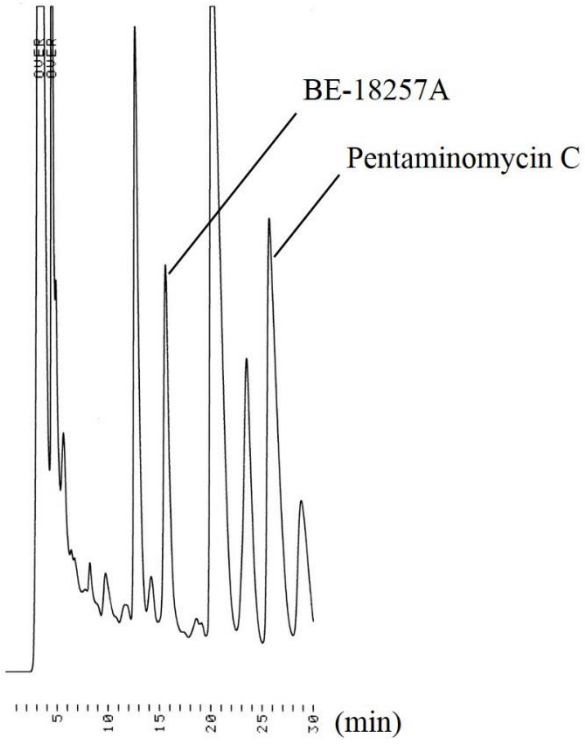


Figure 3.3 HPLC chromatogram of crude extract of *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T} cultured in modified ISP2 agar medium containing leucine; The 100 µl of the extract was analyzed using an isocratic elution mode, 37% MeCN/water containing 0.05% TFA, at flow rate of 1 ml/min; pentaminomycin C was detected by HPLC at the retention time 26.70 min

3.3.3 Structure determination of pentaminomycin C

Chemical structure of pentaminomycin C was elucidated by combination of ESI-MS and NMR analyses. The molecular formula of pentaminomycin C was established as $C_{37}H_{51}N_9O_6$ due to the ion corresponding to $[M+H]^+$ was observed at *m/z* 718.38 by ESI-TOF-MS (Fig. 3.4) and the ion corresponding to $[M+H]^+$ was observed at *m/z* 718.4010 (calculated *m/z* value: 718.4040) by accurate mass analysis (Fig. 3.5). NMR analyses including ^1H, ^{13}C, DEPT-135, DQF-COSY, TOCSY, NOESY, HMBC, HSQC and ^{15}N-^1H HSQC were performed on pentaminomycin C dissolved in DMSO-d_6 (Fig. 3.6 – 3.14). The assignment of constituent amino acids including Leu, Val, Trp, 5-OHArg and Phe was accomplished using spin-system identification (Table 3.1). The amino acid sequence of pentaminomycin C was clarified based on HMBC correlations. HMBC correlations from α-proton and amide proton to same carbonyl carbon were observed, indicated by half ended arrows in Fig. 3.15. The results confirmed the sequence of 5-OHArg-Phe-Leu-Val-Trp in pentaminomycin C. Although HMBC correlation from α-proton (δH 4.27) to carbonyl carbon (δC 171.9) in Trp was not observed, NOESY correlation which was observed between α-proton of Trp (δH 4.27) and amide proton of 5-OHArg (δH 7.31), indicated by double ended arrows in Fig. 3.15, revealed the connection between Trp and 5-OHArg. An unusual amino acid 5-OHArg was established by HMBC correlation detected from proton (δH 10.52) to carbon of C=N bond (157.4) as well as the chemical shifts of protons (δH 3.40) and carbon at δ position (δC 50.6) in 5-OHArg were shifted to downfield due to the presence of hydroxy residue of δ-amide. The chemical shift values of 5-OHArg in pentaminomycin C were also similar to the chemical shift values of 5-OHArg in pentaminomycin A and B which were previously reported by Jang et al. (2018).[123] Structure of pentaminomycin C was determined to be a cyclic pentapeptide containing Leu, Val, Trp, 5-OHArg, and Phe. In addition, ESI-MS and NMR analyses were performed on BE-18257A isolated from *S. cacaoi* subsp. *cacaoi* NBRC 12748^T in the same manner. The chemical shift values were very similar to the literature values of BE-18257A reported by Nakajima et al. (1991) (data not shown).[122]

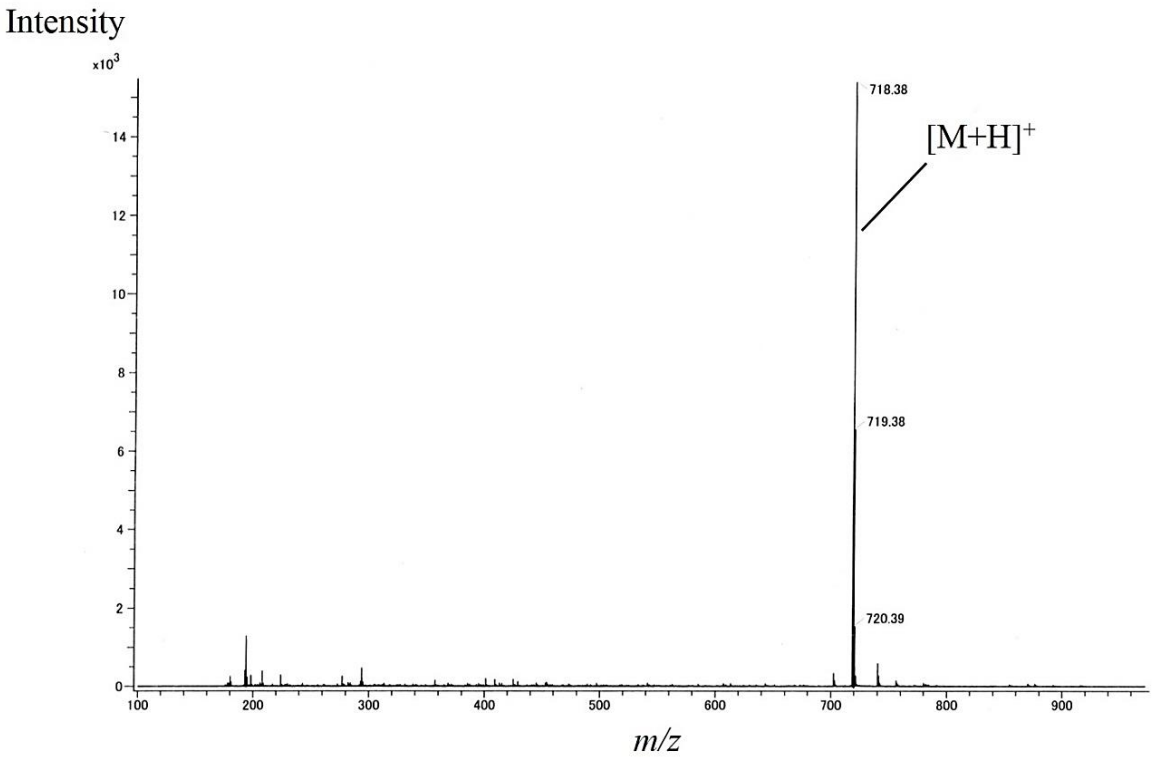


Figure 3.4 ESI-TOF-MS of pentaminomycin C. The $[M+H]^+$ was observed at *m/z* 718.38

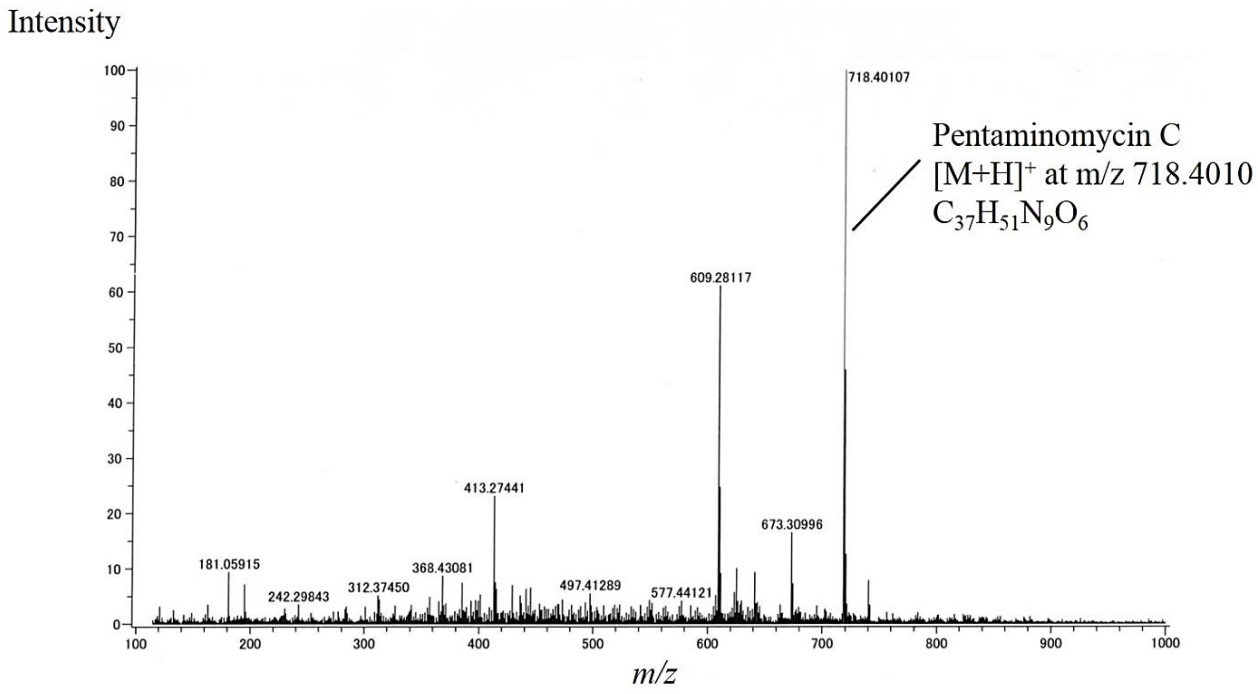


Figure 3.5 Accurate mass analysis of pentaminomycin C. The $[M+H]^+$ was observed at *m/z* 718.4010

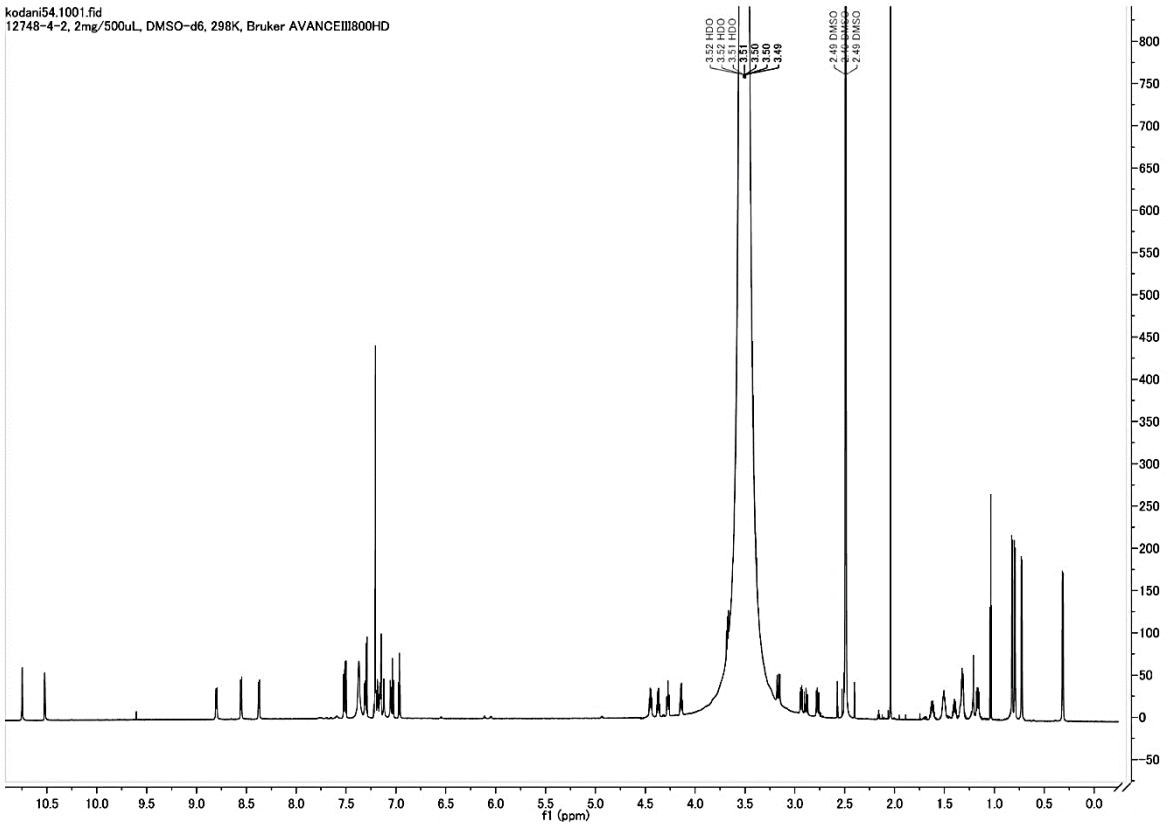


Figure 3.6 ^{1}H NMR spectrum of pentaminomycin C

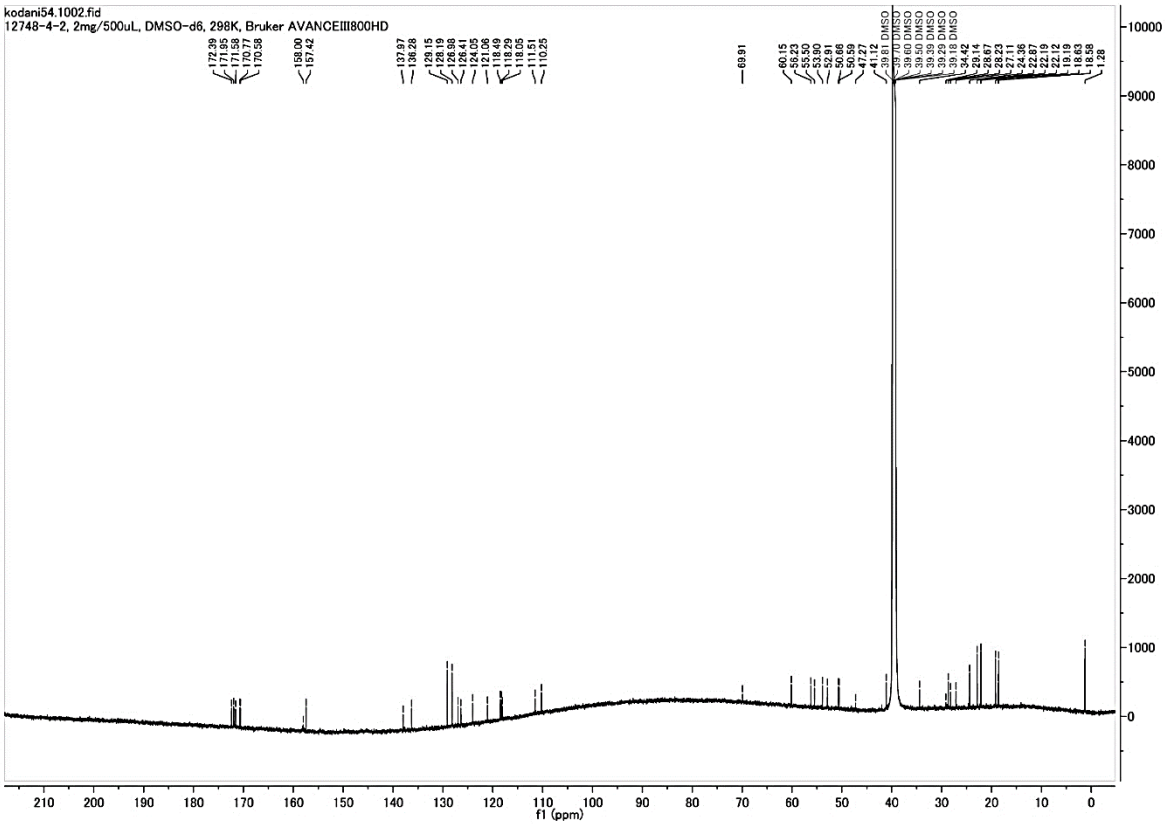


Figure 3.7 ^{13}C NMR spectrum of pentaminomycin C

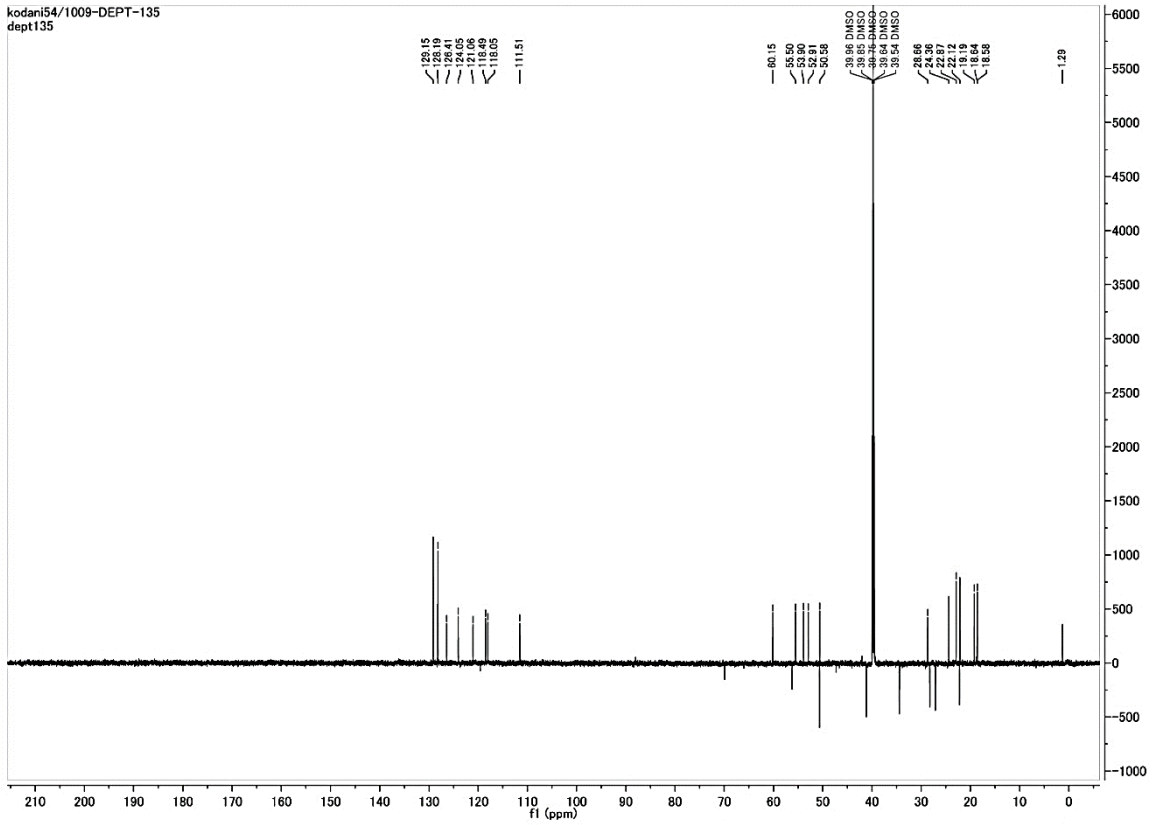


Figure 3.8 DEPT-135 spectrum of pentaminomycin C

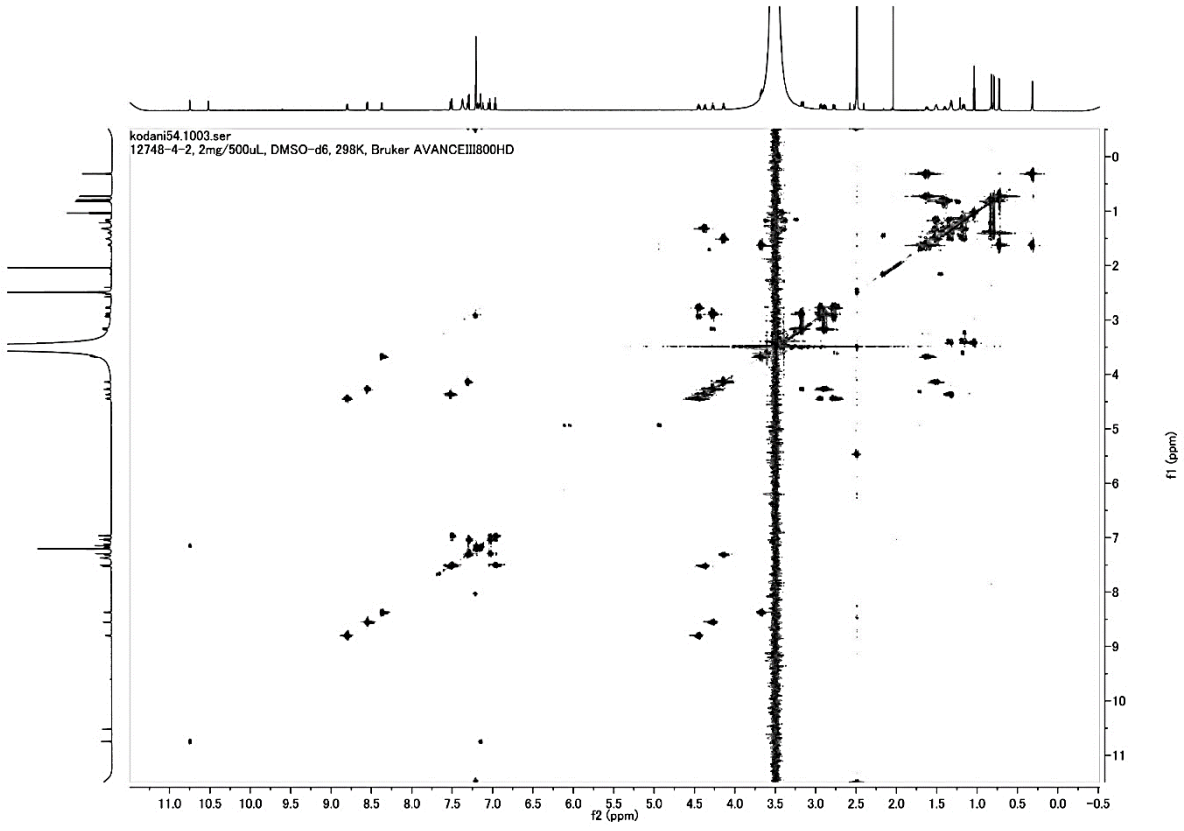


Figure 3.9 DQF-COSY spectrum of pentaminomycin C

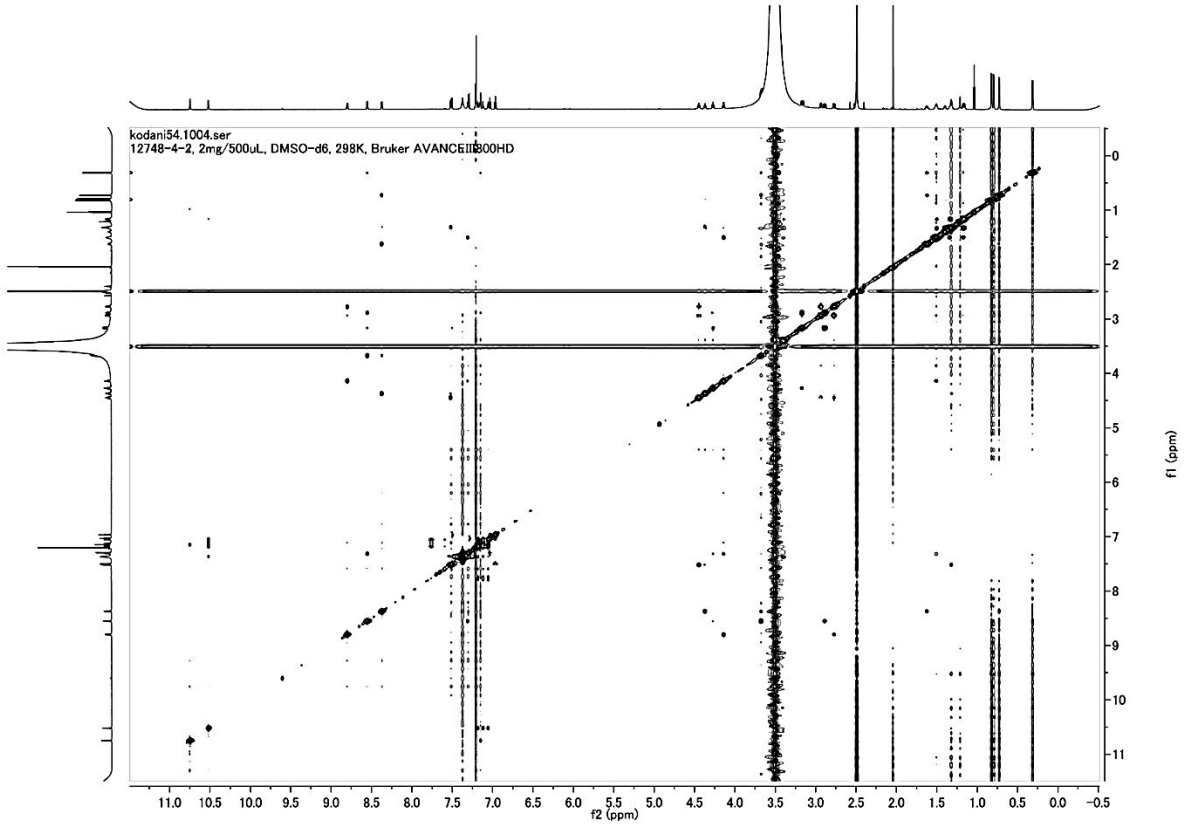


Figure 3.10 TOCSY spectrum of pentaminomycin C

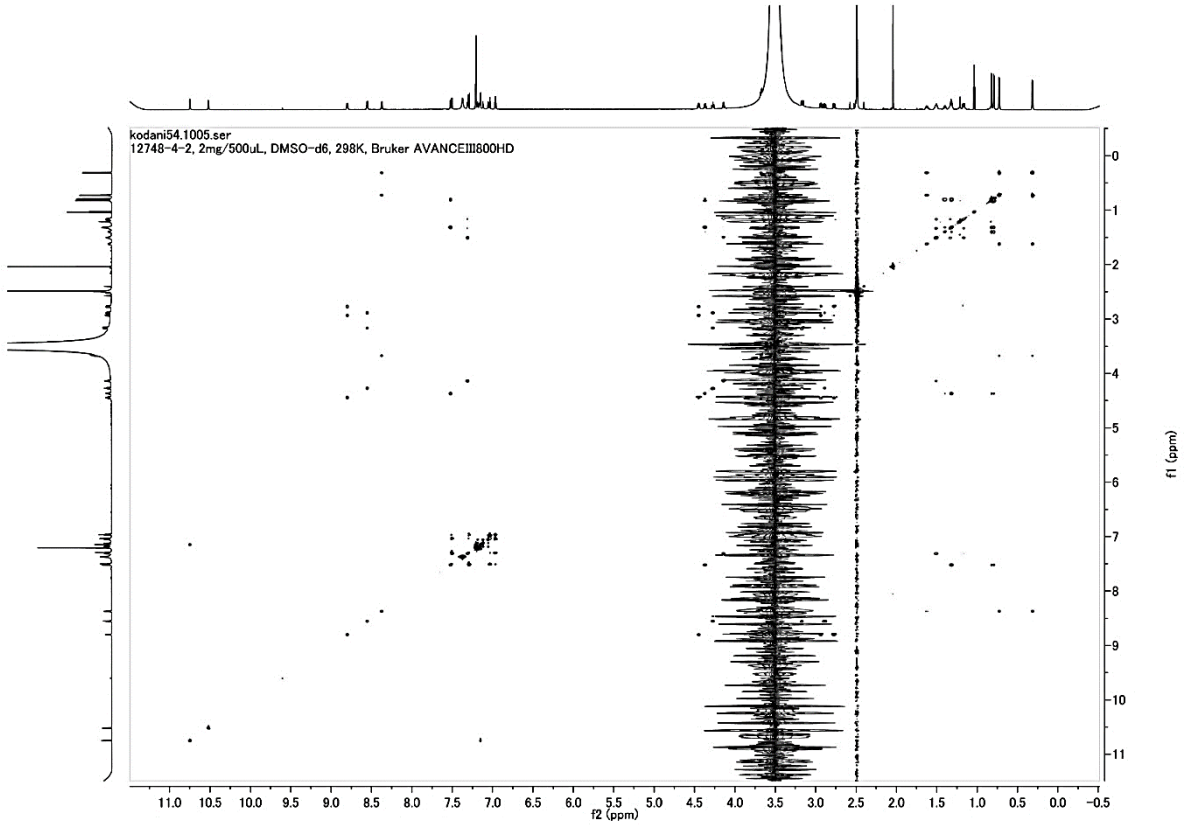


Figure 3.11 NOESY spectrum of pentaminomycin C

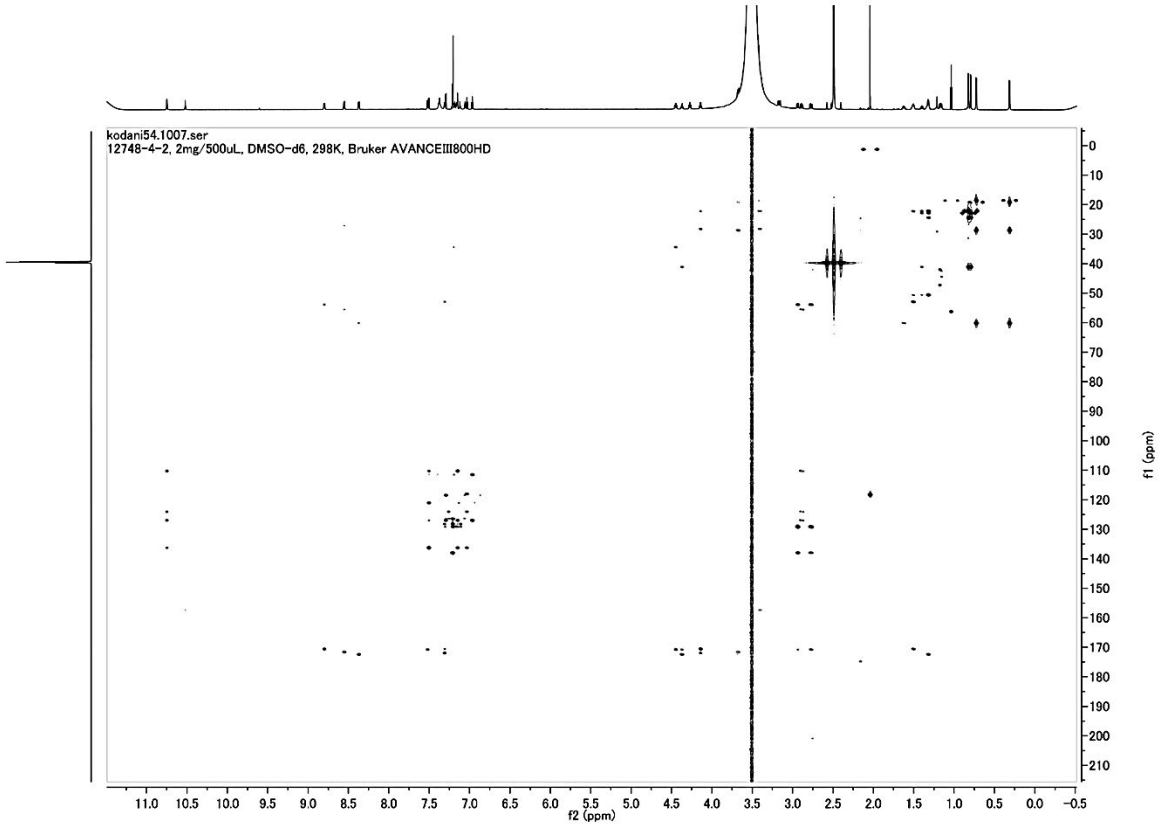


Figure 3.12 HMBC spectrum of pentaminomycin C

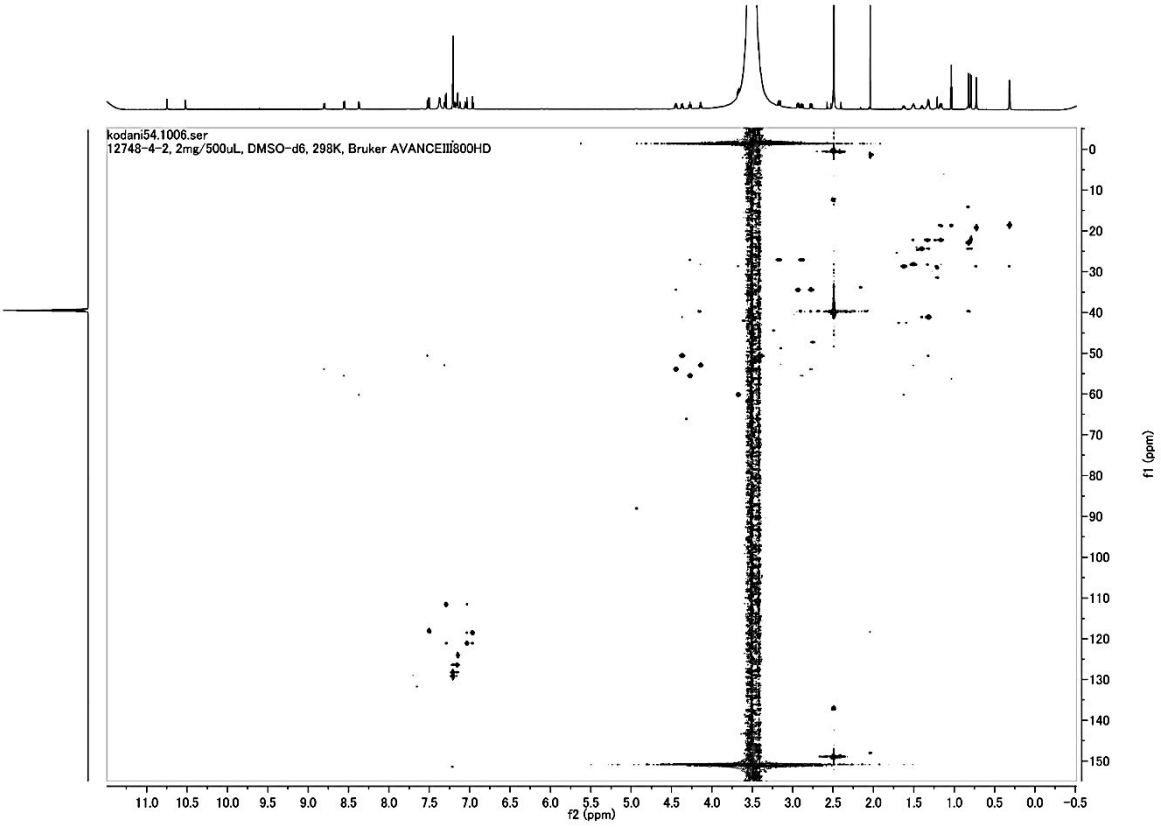


Figure 3.13 HSQC spectrum of pentaminomycin C

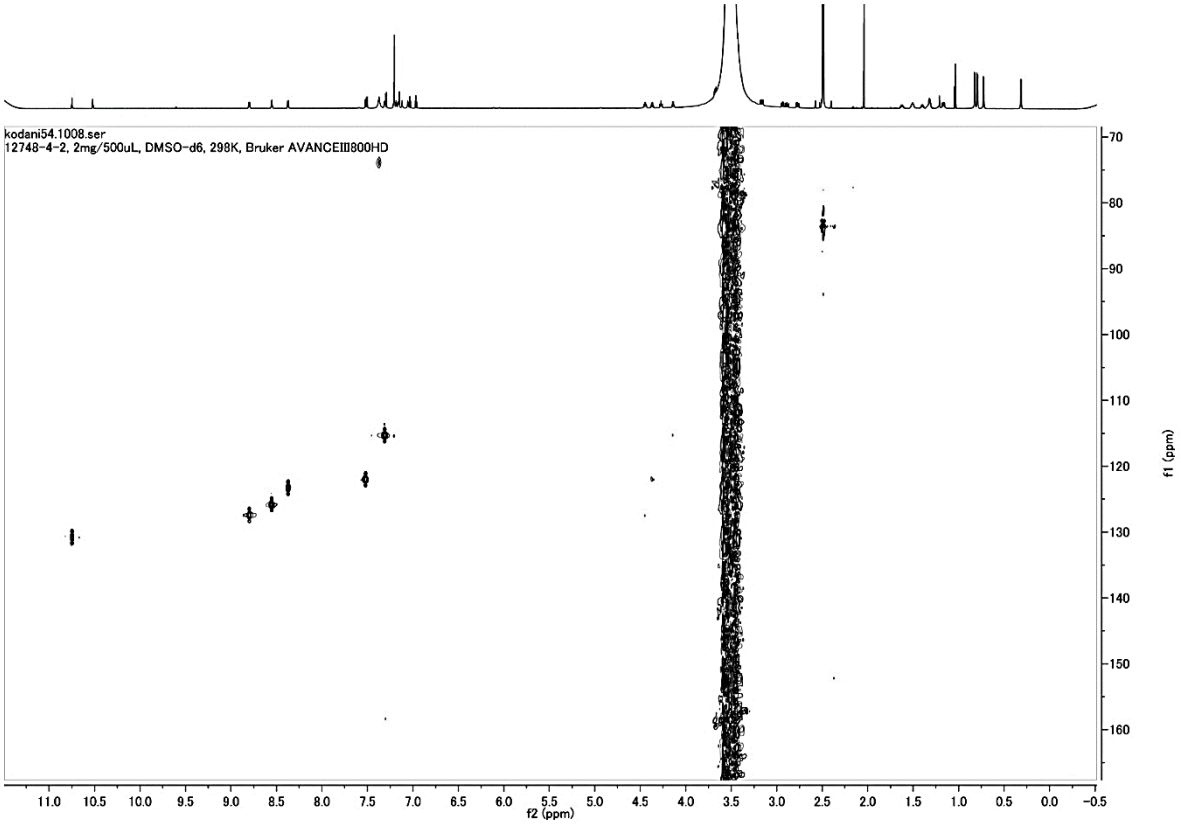


Figure 3.14 ^{15}N-^{1}H HSQC spectrum of pentaminomycin C

Table 3.1 NMR chemical shift values of pentaminomycin C in DMSO-d_6

AA	Position	δH (*J* = Hz)	δC	δN
Leu	CO		172.4	
	NH	7.52 (d, 9.0)		122.0
	α	4.37 (m)	50.6	
	β	1.32 (m)	41.1	
	γ	1.40 (m)	24.4	
	δ1	0.79 (d, 6.5)	22.1	
	δ2	0.82 (d, 6.6)	22.9	
Phe	CO		170.8	
	NH	8.80 (d, 8.2)		127.4
	α	4.45 (m)	53.9	
	β	2.78 (d, 14.0, 9.4)	34.5	
		2.93 (d, 14.0, 5.9)		
	γ		138.0	
	δ	7.21 (m)	129.2	
	ε	7.21 (m)	128.2	
	ζ	7.16 (m)	126.4	
OHArg	CO		170.6	
	NH	7.31 (d, 7.2)		115.3
	α	4.14 (m)	52.9	
	β	1.51 (m)	28.2	
	γ	1.16 (m), 1.33 (m)	22.2	
	δ	3.40 (m)	50.6	
	N-OH	10.52 (s)		
	C=NH	ND	157.4	
	NH_2	ND		
Trp	CO		171.9	
	NH	8.55 (d, 7.8)		125.8
	α	4.27 (m)	55.5	
	β	2.89 (dd, 14.6, 11.3)	27.1	
		3.17 (dd, 14.6, 3.6)		
	ε1	10.75 (s)		130.8
	δ1	7.15 (m)	124.0	
	γ		110.2	
	ε3	7.50 (d, 7.9)	118.1	
	ζ3	6.96 (m)	118.5	
	η2	7.03 (m)	121.1	
	ζ2	7.29 (d, 8.1)	111.5	
	ε2		136.3	
	δ2		127.0	
Val	CO		171.6	
	NH	8.37 (d, 7.5)		123.3
	α	3.67 (dd, 9.9, 7.5)	60.1	
	β	1.62 (m)	28.7	
	γ1	0.31 (d, 6.7)	18.6	
	γ2	0.73 (d, 6.6)	19.2	

ND:not detected

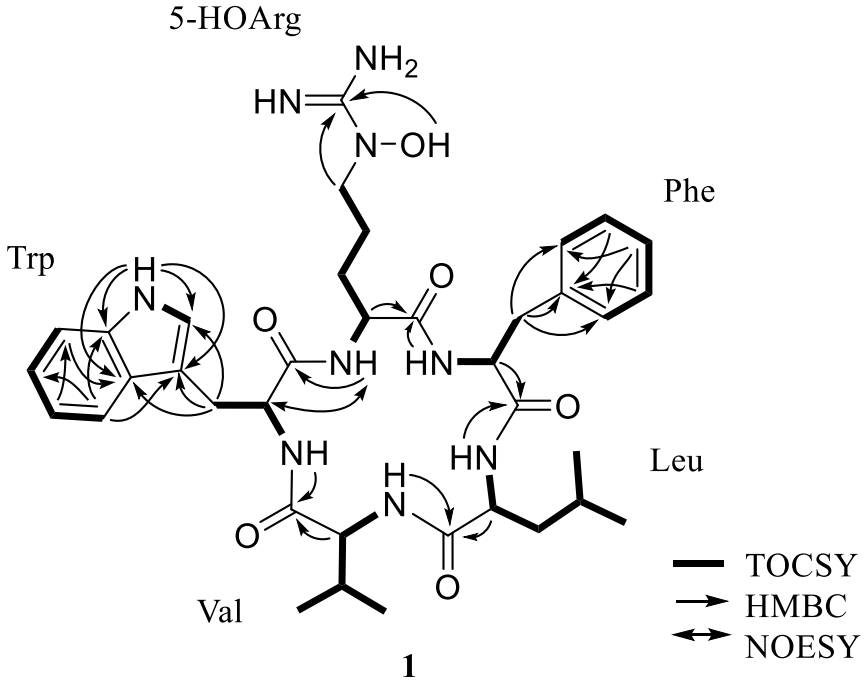


Figure 3.15 Key 2D NMR correlations of pentaminomycin C

3.3.4 Modified Marfey's analysis of pentaminomycin C

Stereochemistry of constituent amino acids in pentaminomycin C was clarified by Modified Marfey's analysis.[96] Pentaminomycin C (1.0 mg) was hydrolyzed with hydriodic acid (HI) to convert 5-OHArg to Arg according to the previous report.[124] Pentaminomycin C (1.0 mg) was hydrolyzed with 6 N HCl containing 3% phenol to recover Trp according to the previous report.[89] The hydrolysate of pentaminomycin C was derivatized using *N*α-(5-fluoro-2,4-dinitrophenyl)-L-leucinamide (L-FDLA) and the derivatives were analyzed by HPLC, compared to derivatives of standard amino acids.

HPLC condition 1 was applied for standard amino acids including arginine (Arg), valine (Val), leucine (Leu), tryptophan (Trp) and phenylalanine (Phe). Retention times (min) of L- or D-FDLA derivatized amino acids in HPLC condition 1 were as follows; L-Arg-L-FDLA (20.25 min), L-Arg-D-FDLA (18.56 min), L-Val-L-FDLA (40.68 min), L-Val-D-FDLA (53.76 min), L-Leu-L-FDLA (46.51 min), L-Leu-D-FDLA (62.57 min), L-Trp-L-FDLA (47.39 min), L-Trp-D-FDLA (54.37 min), L-Phe-L-FDLA (48.66 min) and L-Phe-D-FDLA (59.30 min) (Fig. 3.16 – 3.20). HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with HI (HPLC condition 1) and HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with 6 N HCl containing 3% phenol in HPLC condition 1 are shown in Fig. 3.21 and 3.22, respectively.

HPLC condition 2 was applied for Val analysis. Retention times (min) of L- or D-FDLA derivatized amino acids in HPLC condition 2 were as follows; L-Val-L-FDLA (21.05 min) and L-Val-D-FDLA (40.23 min) (Fig. 3.23). HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with HI in HPLC condition 2 is shown in Fig. 3.24.

HPLC condition 3 was applied for Trp analysis. Retention times (min) of L- or D-FDLA derivatized amino acids in HPLC condition 3 were as follows; L-Trp-L-FDLA (18.57 min) and L-Trp-D-FDLA (28.04 min) (Fig. 3.25). HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with 6 N HCl containing 3% phenol in HPLC condition 3 is shown in Fig. 3.26. Overall, the results indicated that stereochemistry of constituent amino acids in pentaminomycin C were L-Leu, D-Val, L-Trp, L-5-OHArg, and D-Phe. Therefore, the chemical structure of pentaminomycin C was determined to be cyclo(-L-Leu-D-Val-L-Trp-L-5-OHArg-D-Phe-), as shown in Fig. 3.27 (1). In addition, stereochemistry of constituent amino acids in BE-18257A was elucidated in the same manner. Absolute configurations of amino acids in BE-18257A were L-Leu, D-Trp, D-Glu, L-Ala and D-Val, which were identical to the previous report (data not shown).[122] The chemical structure of BE-18257A was confirmed to be cyclo(-L-Leu-D-Trp-D-Glu-L-Ala-D-Val-), as shown in Fig. 3.27 (2).

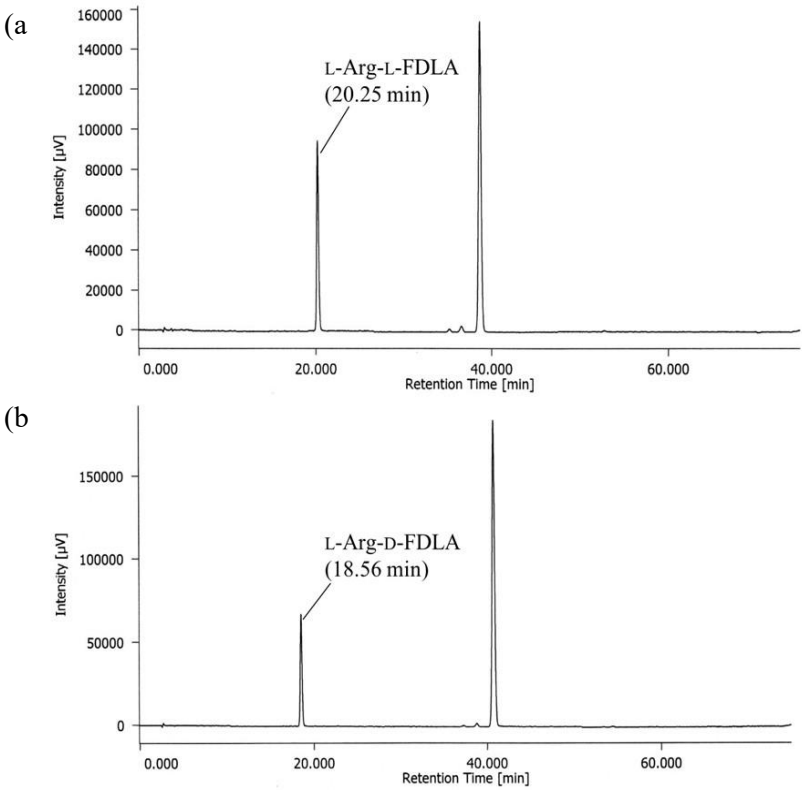


Figure 3.16 HPLC chromatograms of (a) L-Arg-L-FDLA and (b) L-Arg-D-FDLA (HPLC condition 1)

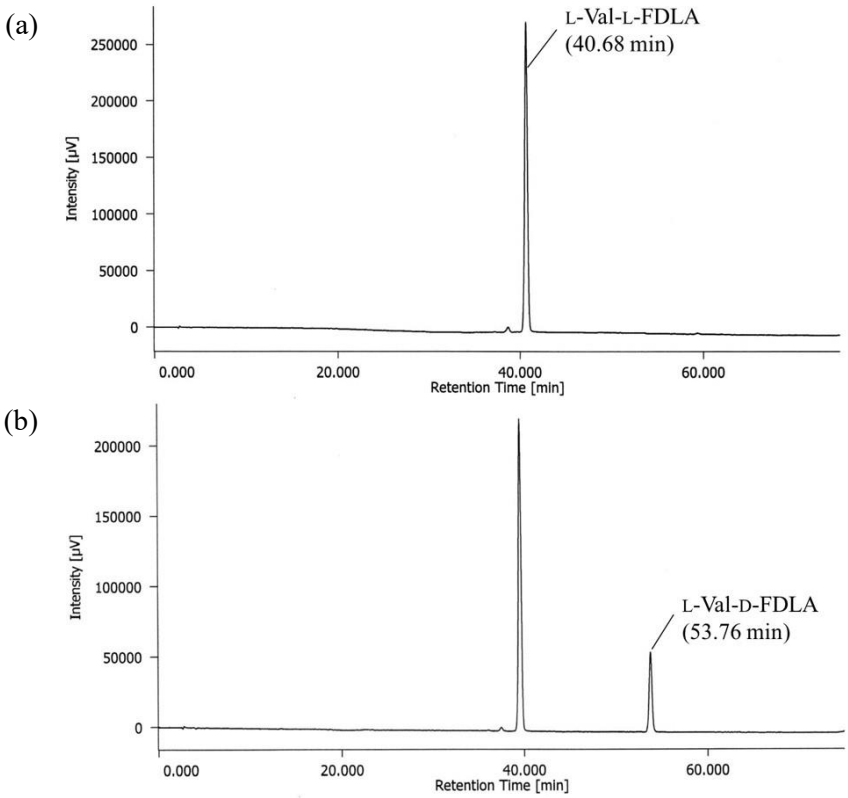


Figure 3.17 HPLC chromatograms of (a) L-Val-L-FDLA and (b) L-Val-D-FDLA (HPLC condition 1)

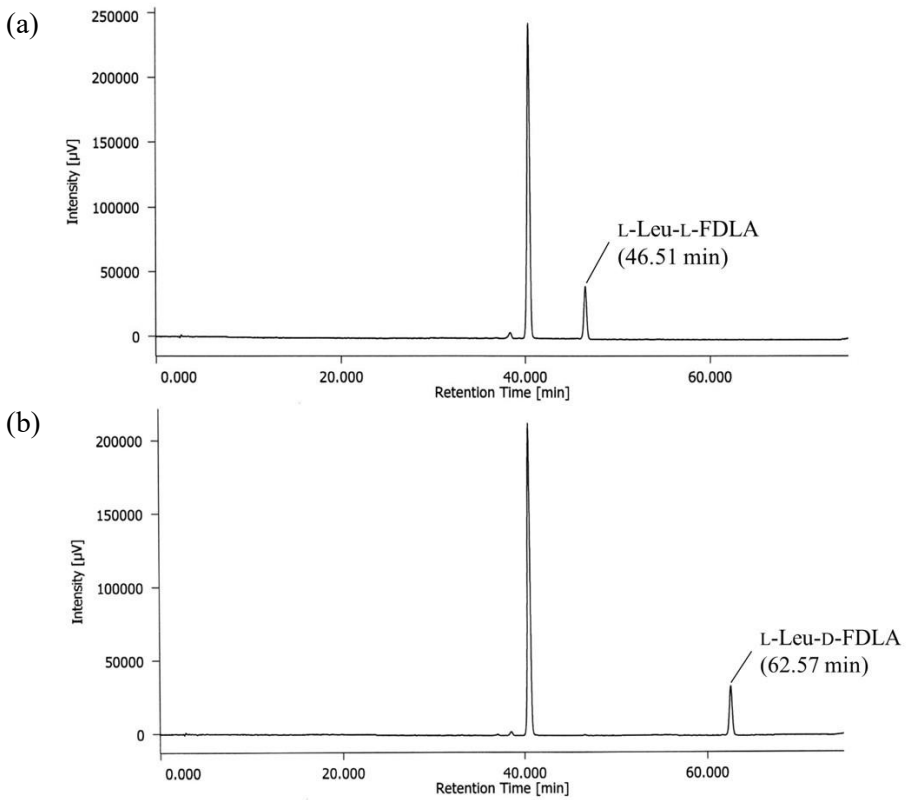


Figure 3.18 HPLC chromatograms of (a) L-Leu-L-FDLA and (b) L-Leu-D-FDLA (HPLC condition 1)

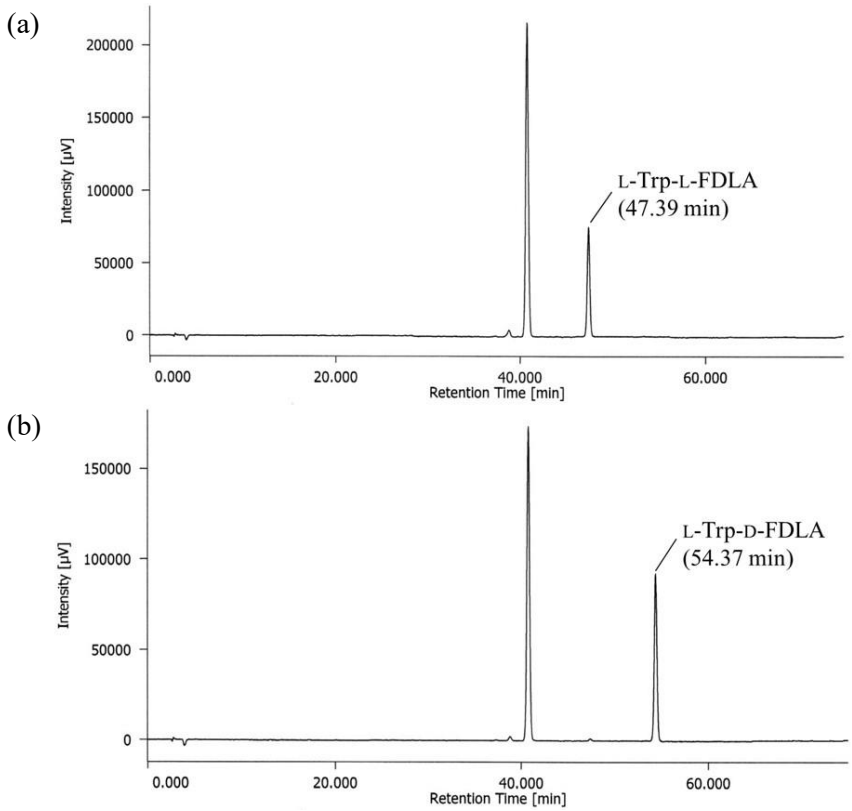


Figure 3.19 HPLC chromatograms of (a) L-Trp-L-FDLA and (b) L-Trp-D-FDLA (HPLC condition 1)

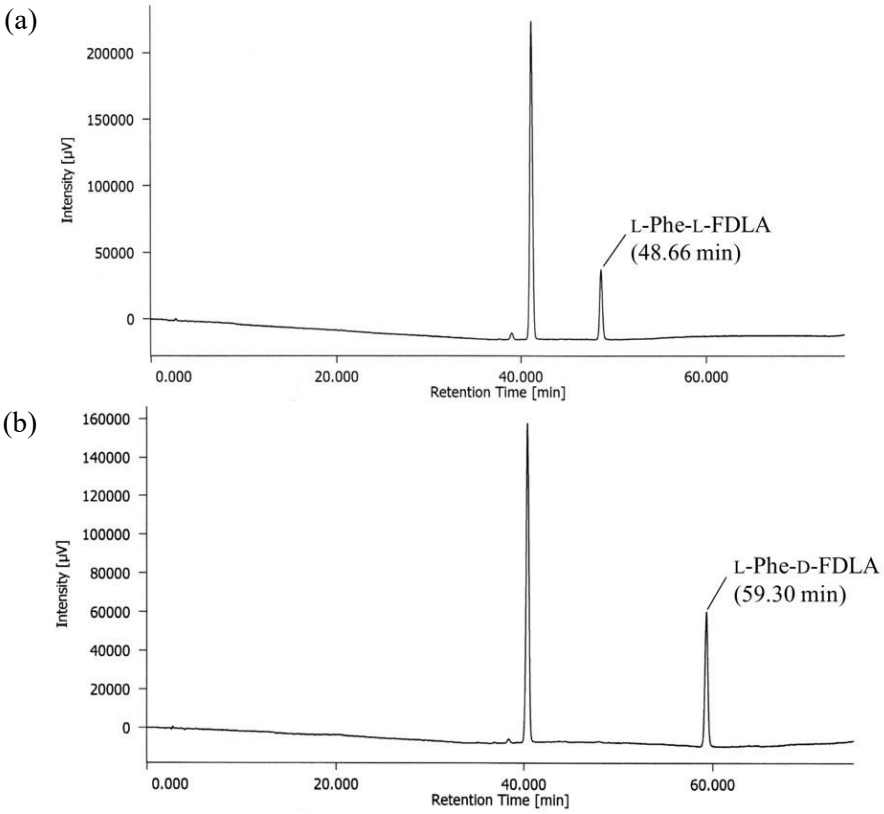


Figure 3.20 HPLC chromatograms of (a) L-Phe-L-FDLA and (b) L-Phe-D-FDLA (HPLC condition 1)

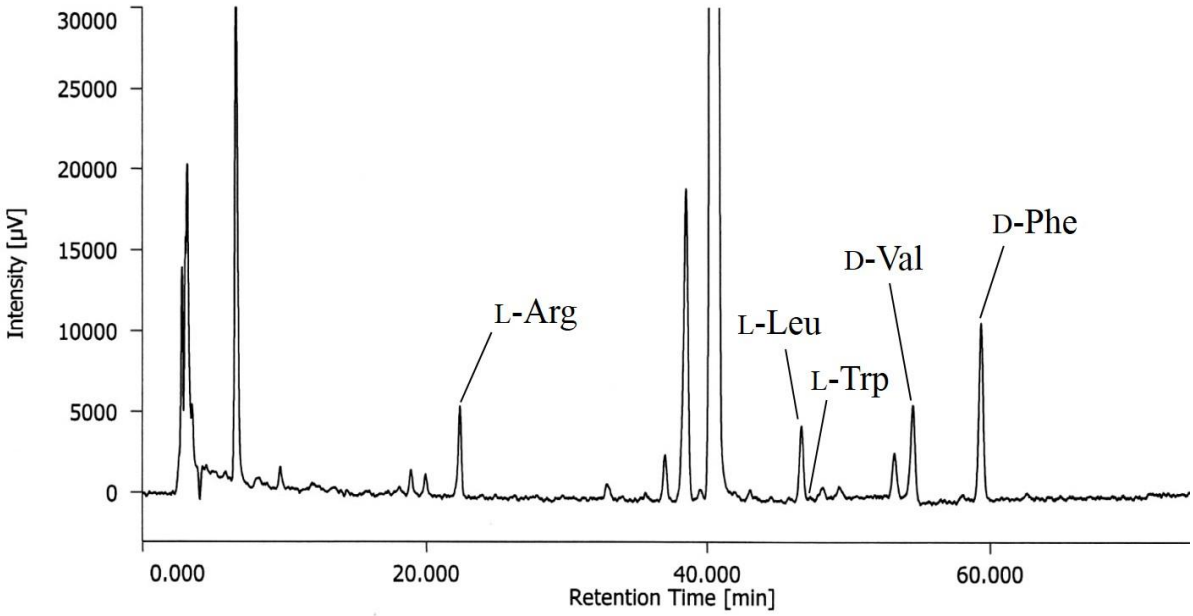


Figure 3.21 HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with HI (HPLC condition 1)

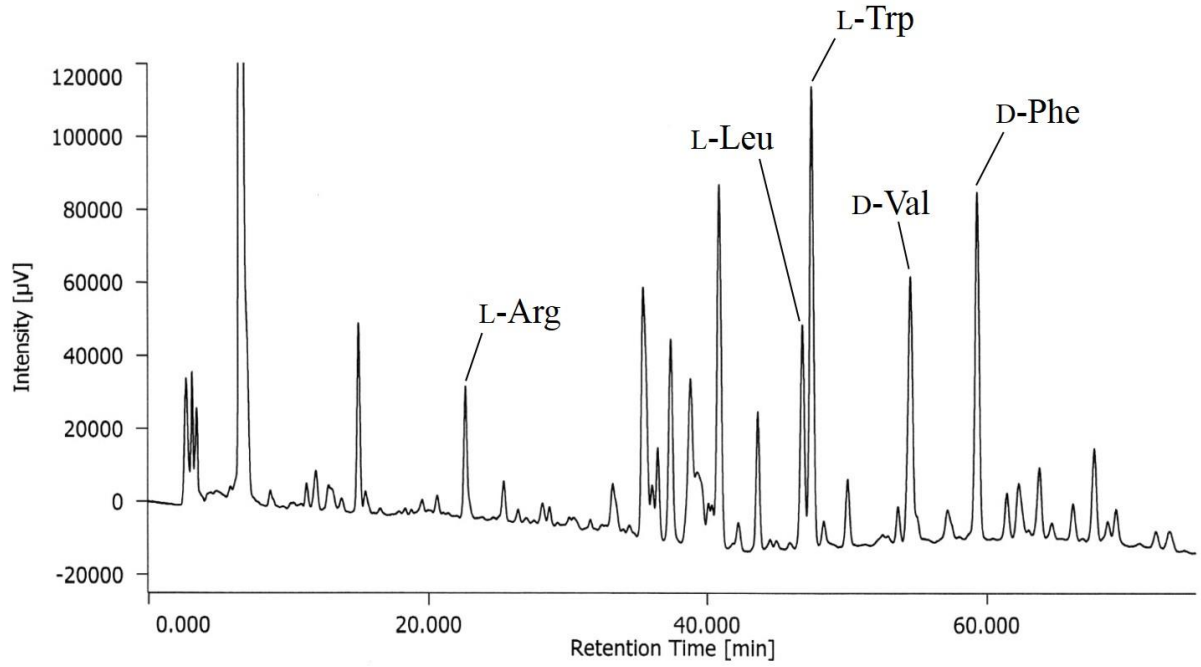


Figure 3.22 HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with 6N HCl containing 3% phenol (HPLC condition 1)

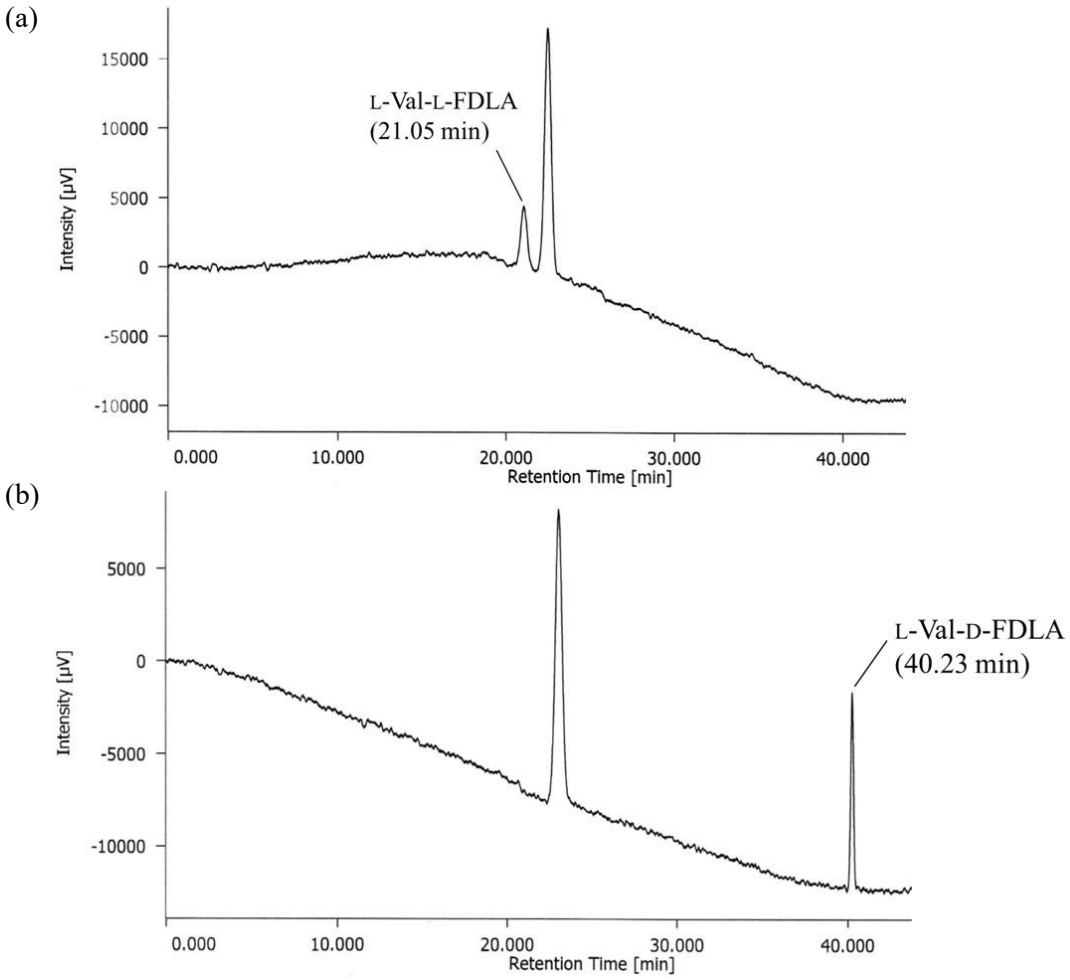


Figure 3.23 HPLC chromatograms of (a) L-Val-L-FDLA and (b) L-Val-D-FDLA (HPLC condition 2)

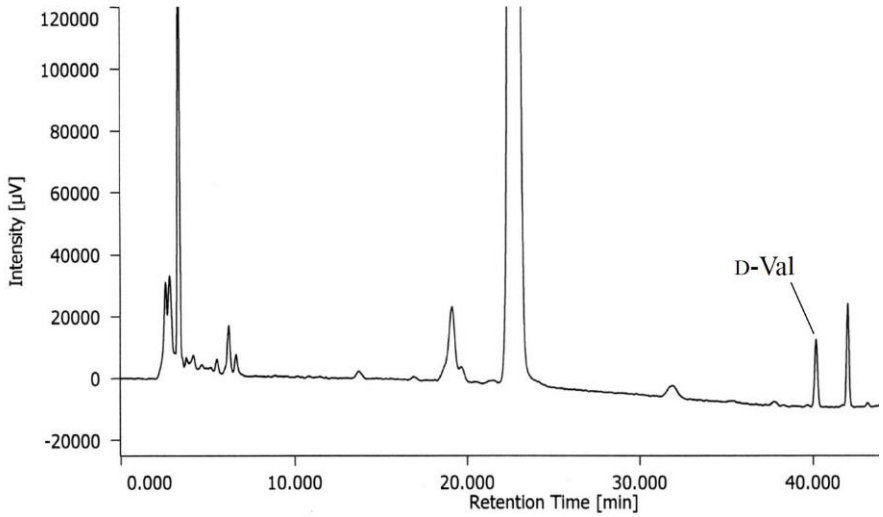


Figure 3.24 HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with HI (HPLC condition 2)

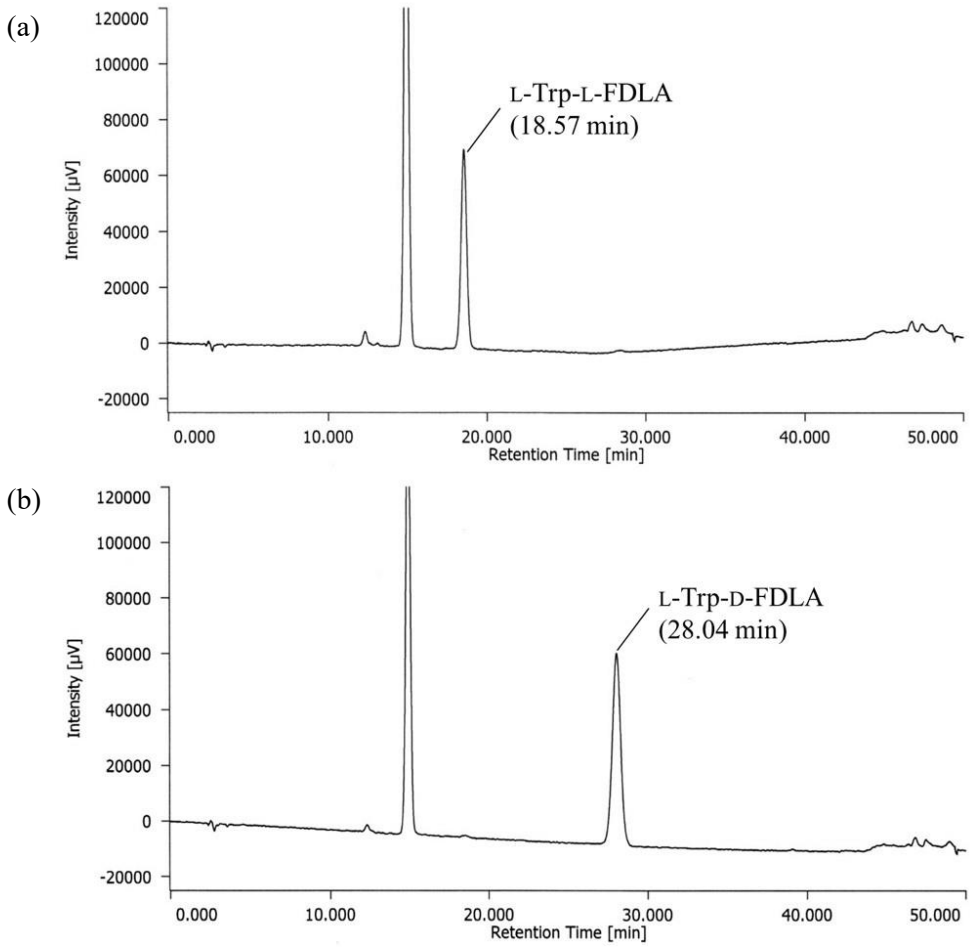


Figure 3.25 HPLC chromatograms of (a) L-Trp-L-FDLA and (b) L-Trp-D-FDLA (HPLC condition 3)

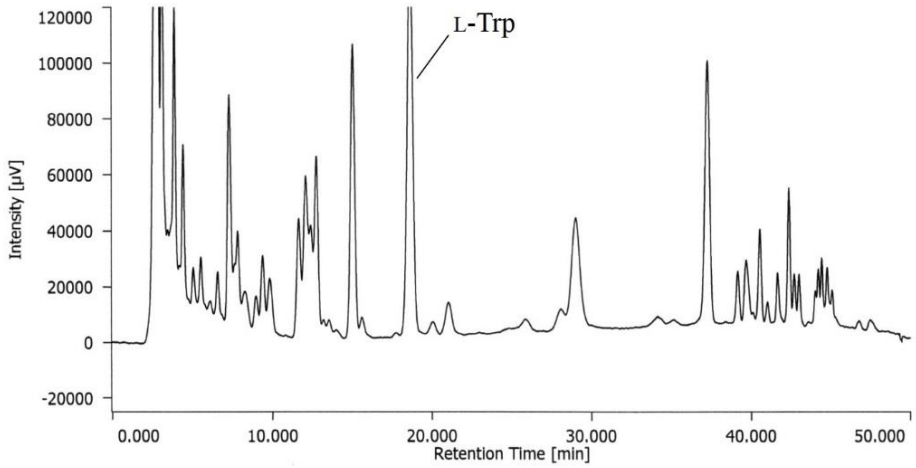


Figure 3.26 HPLC chromatogram of L-FDLA derivative of pentaminomycin C hydrolyzed with 6N HCl containing 3% phenol (HPLC condition 3)

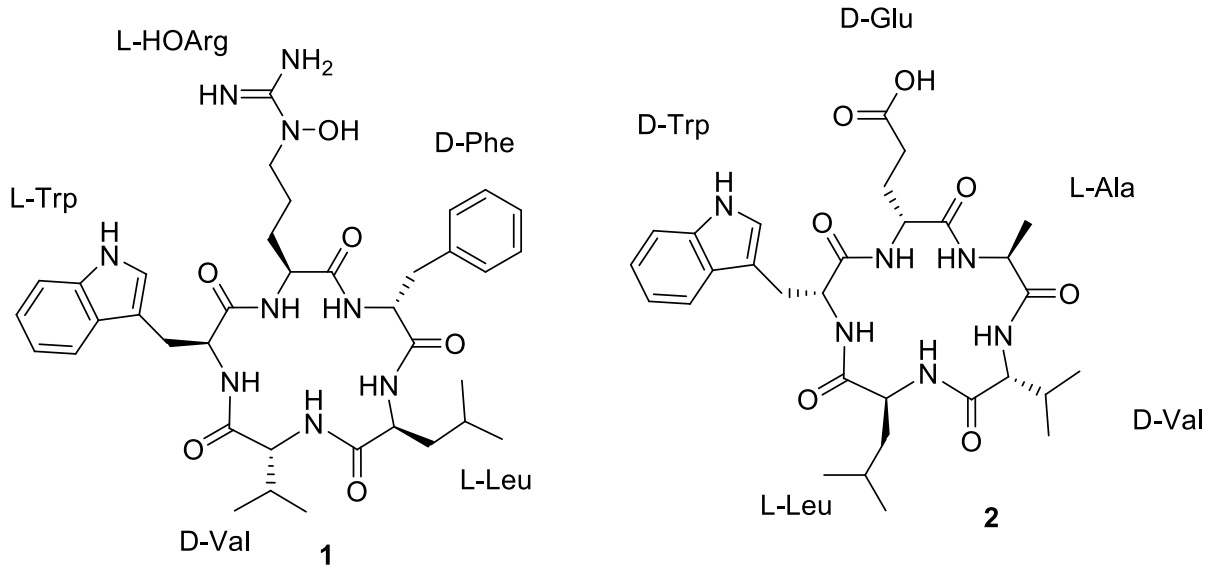


Figure 3.27 Chemical structures of pentaminomycin C **(1)** and BE-18257A **(2)**

3.3.5 Antibacterial activity of pentaminomycin C

Antibacterial activities of pentaminomycin C and BE-18257A were tested against several Gram-positive and Gram-negative bacteria including *E. coli*, *P. aeruginosa*, *S. aureus*, *B. subtilis* and *M. luteus* using minimum inhibitory concentrations (MICs) assay. As a result, pentaminomycin C showed antibacterial activity against Gram-positive bacteria including *S. aureus*, *B. subtilis* and *M. luteus* with MICs of 16 μg/ml, however, the peptide did not show antibacterial activity against Gram-negative bacteria including *E. coli* and *P. aeruginosa* at the concentration of 64 μg/ml (Table 3.2). BE-18257A did not show any antibacterial activity against any test bacteria at the concentration of 64 μg/ml (data not shown).

Table 3.2 Minimum inhibitory concentrations (MICs) of pentaminomycin C

Test microorganisms	MIC (μg/ mL)	
	Tetracycline	Pentaminomycin C
Bacillus subtilis	1	16
Micrococcus luteus	0.125	16
Escherichia coli	1	>64
Staphylococcus aureus	0.25	16
Pseudomonas aeruginosa	16	>64

3.3.6 Biosynthetic gene cluster of pentaminomycin C

Since pentaminomycin C and BE-18257A consist of five amino acids containing two and three D-amino acids, respectively, we examined biosynthetic gene cluster (BGC) for pentaminomycin C and BE-18257A in draft genome sequence of *S. cacaoi* subsp. *cacaoi* NRRL B-1220^T (=NBRC 12748^T) by search for NRPS genes harboring five modules and comprising two and three epimerase (E) domains. As a result, two NRPS genes were found in the genome sequence of *S. cacaoi* subsp. *cacaoi*, as shown in Fig 3.28. The BGC includes two NRPS genes, regulatory genes, transporter genes, and biosynthetic genes including *mtb*, putative hydrolase, and P450 enzyme coding genes. The NRPS gene (BZY53_RS10935) was considered as a gene responsible for BE-18257A biosynthesis because second, third and fifth modules comprise of E domain, and substrates of adenylation (A) domains of first, fourth and fifth modules were predicted to be Leu, Ala and Val, respectively, by bioinformatic analysis using online analysis tools[74, 125], corresponding to amino acid residues in BE-18257A (L-Leu-D-Trp-D-Glu-L-Ala-D-Val). The NRPS gene (BZY53_RS10880 to BZY53_RS01305) was considered as a gene responsible for pentaminomycin C biosynthesis because second and fifth modules comprise of E domain, and substrates of A domains of second, fourth and fifth modules were predicted to be Val, Arg and Phe, respectively, corresponding to amino acid residues in pentaminomycin C (L-Leu-D-Val-L-Trp-L-5-OHArg-D-Phe). For the biosynthetic pathway, pentaminomycin C and BE-18257A were proposed to be biosynthesized by NRPS system. The gene(s) BZY53_RS10990 and/or BZY53_RS10985 encoded for P450 enzyme hydroxylate Arg residue to 5-OHArg in pentaminomycin C. According to NRPS pathway, peptide chains are typically released from peptidyl carrier protein (PCP) domain in the last module of NRPSs by function of thioesterase (TE) domain. However, TE domain was not found in both NRPSs for pentaminomycin C and BE-18257A. Putative hydrases, stand-alone enzymes in penicillin-binding protein family, such as SurE, MppK and DsaJ, were recently reported to be involved in releasing of elongated peptide chains from PCP domain and cyclization to afford cyclic peptides.[126, 127] A putative hydrolase adjacent, BZY53_RS10935, showed about 40% similarities and 50% identities in amino acid sequence to these enzymes. Since SurE was reported as a trans-acting thioesterase cyclizing two distinct nonribosomal peptides, the putative hydrolase, BZY53_RS10940, might be involved in cyclization of the two cyclic peptides pentaminomycin C and BE-18257A. We proposed biosynthetic pathways of pentaminomycin C and BE-18257A as shown in Fig. 3.29-30. Initially, A domain within each module converts each amino acid into aminoacyl adenylate and transfers them to PCP domain to form aminoacyl thioseter. E domains in the second, third and fifth modules epimerize α-carbon of Trp, Glu and Val to form D-configuration in BE-18257A. E domains in the second and fifth modules

epimerize α-carbon of Val and Phe to form D-configuration in pentaminomycin C. C domains catalyze for successive *N*-acylation to afford L, D, D, L, D- and L, D, L, L, D-pentapeptidyl thioesters attached to PCPs of the last modules of BE-18257A and pentaminomycin C, respectively. Arg residue in pentaminomycin C is hydroxylated into 5-OHArg by P450 enzymes. Finally, putative hydrolase releases two linear peptide chains from PCPs and cyclizes them to yield two cyclic pentapeptides, pentaminomycin C and BE-18257A. Normally, a single BGC is believed to synthesize peptides containing same backbone. However, this BGC produces two peptides whose backbones are different. This study suggested that structure determination of actual products together with identification of their BGCs provide useful information to natural product research and bioinformatics analysis.

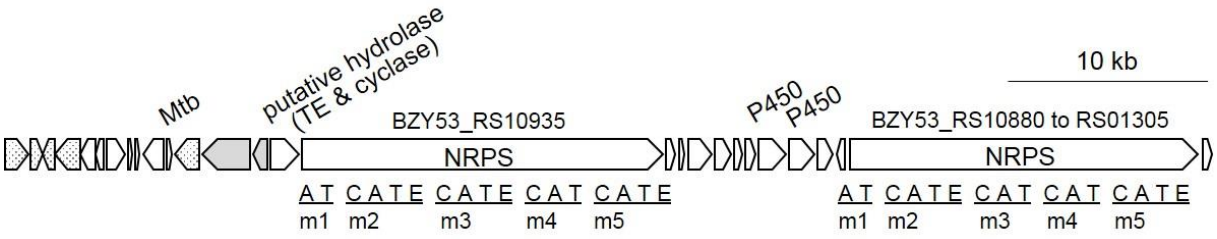


Figure 3.28 Biosynthetic gene cluster of pentaminomycin C and BE-18257A; Domain organizations are shown under NRPS genes and each module is underlined; A, adenylation domain; C, condensation domain; E, epimerase domain; T, peptidyl carrier protein, which is also called thiolation domain; m, module; TE, thioesterase; dotted color indicates regulatory genes; gray indicates transport-related genes

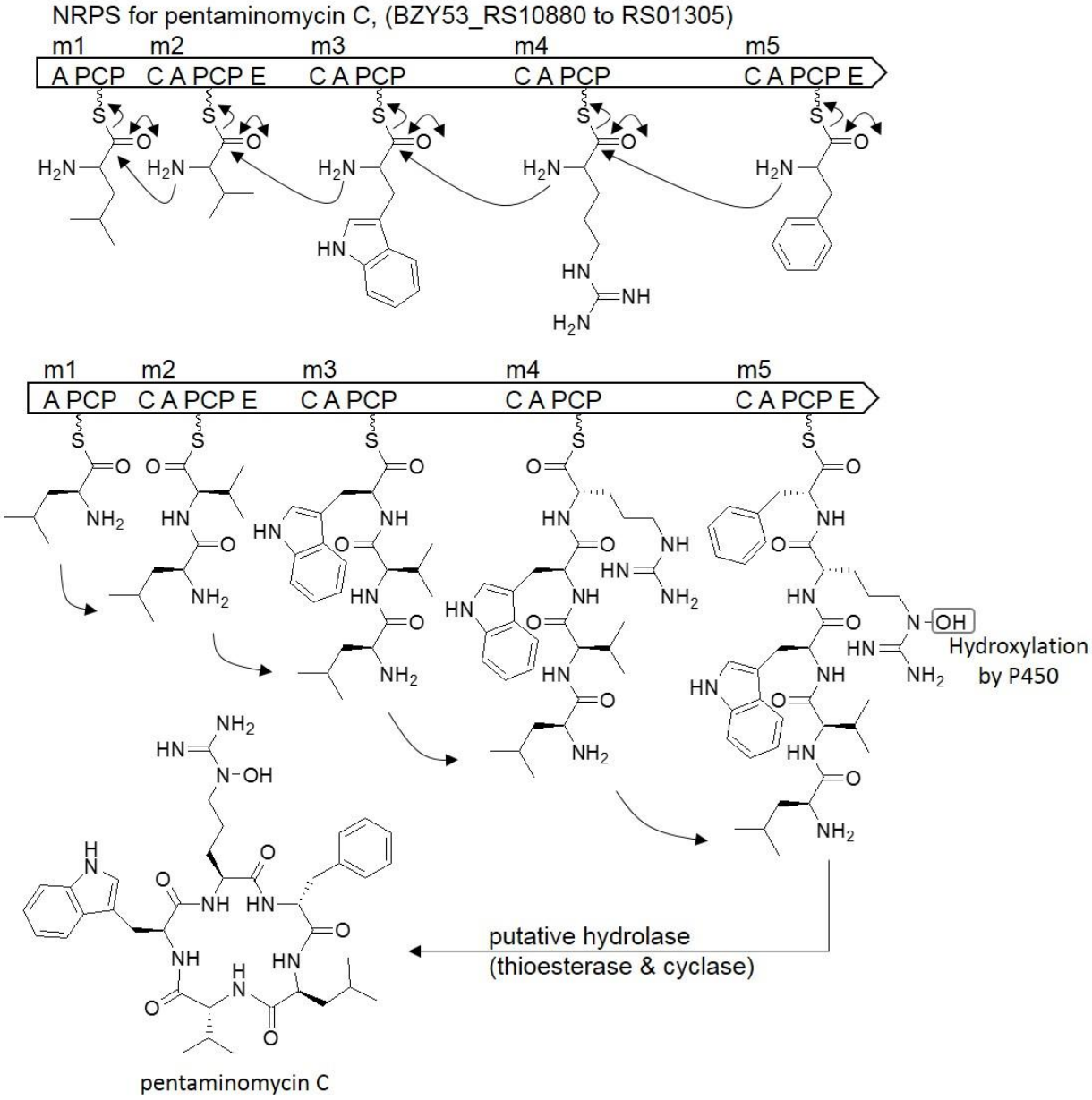


Figure 3.29 Biosynthetic pathway of pentaminomycin C. Letters in the figure represent as following: A, adenylation domain; C, condensation domain; E, epimerase domain; m, module; PCP, peptidyl carrier protein, which is also called thiolation domain

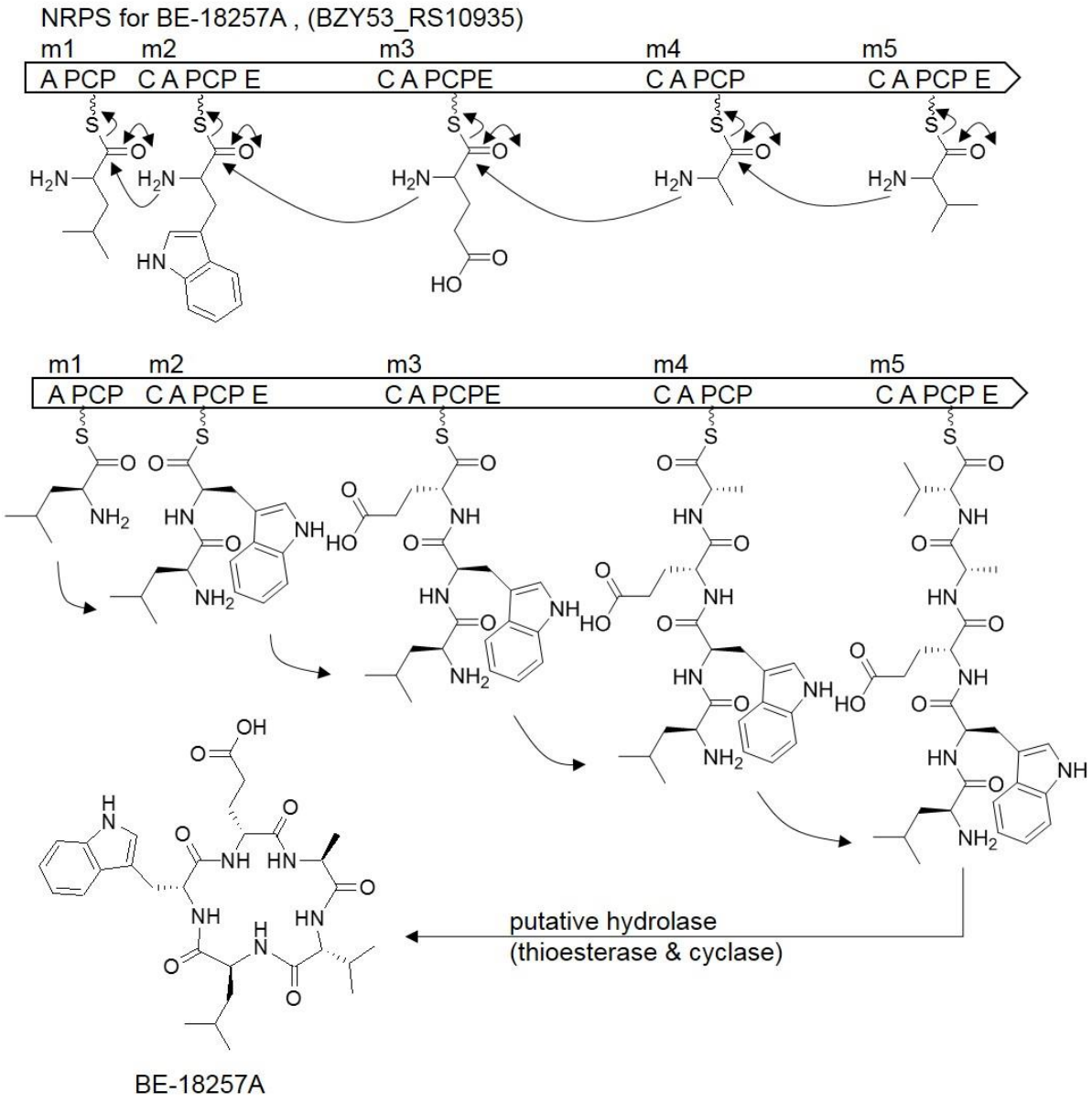


Figure 3.30 Biosynthetic pathway of BE-18257A. Letters in the figure represent as following: A, adenylation domain; C, condensation domain; E, epimerase domain; m, module; PCP, peptidyl carrier protein, which is also called thiolation domain

3.4 Summary

Chemical investigation on extract of actinobacteria leads to discovery of a new antibacterial peptide, named pentaminomycin C, from *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T}. The peptide pentaminomycin C, along with known peptide BE-18257A, were isolated from the extract of *S. cacaoi*. Pentaminomycin C was found to be a cyclic pentapeptide containing unusual amino acid, 5-OHArg, in the molecule which is similar to peptides, pentaminomycin A and B. The biosynthetic gene cluster including two NRPS genes for pentaminomycin C and BE-18257A was identified in the genome sequence of *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T} and biosynthetic pathways of pentaminomycin C and BE-18257A were also proposed in this study. Pentaminomycin C showed antibacterial activity against *S. aureus*, *B. subtilis* and *M. luteus*, but BE-18257A did not showed any antibacterial activity against test bacteria.

DISCUSSION

In recent years, antibiotic resistance has become a serious problem worldwide. Researchers around the world have been trying to find novel sources of natural products to overcome this problem. Actinobacteria are Gram-positive bacteria distributed in terrestrial and aquatic environments which have been known as promising source of bioactive peptide natural products with structural and functional diversity. So far, a number of antibiotics have been discovered from actinobacteria especially genus *Streptomyces* such as vancomycin, streptomycin, terramycin, aureomycin and so on. Moreover, wide variety of bioactive peptides from actinobacteria have been reported, including antibacterial, antiviral, anticancer and immunosuppressants.[2, 3] Bioactive peptides are a group of peptides which exhibit biological functions and provide positive influence on human health. Bioactive peptides derived from natural sources can be classified based on biosynthetic pathways into several groups. Among them, ribosomally synthesized and post-translationally modified peptides (RiPPs) and nonribosomal peptides (NRPs) attract extensive interest since they exhibit wide variety of biological activities.[28, 30] RiPPs are peptides of ribosomal origin, biosynthesized from precursor peptide coding genes. Precursor peptides typically consist of an N-terminal leader peptide and a C-terminal core peptide, including region which can be recognized by post-modification enzymes. Various modification enzymes, such as cytochrome P450, radical SAM enzyme, cyclodehydratase and oxidoreductase, modify a linear precursor peptide and install distinctive moiety onto precursor peptide to generate mature structure, resulting in different classes of RiPPs. Each class of RiPPs represents a unique structure, for instance, lasso peptide contains motif of knot structure, including macrolactam ring, loop and tail in the molecule.[32, 34, 128] NRPs are peptides biosynthesized by nonribosomal peptide synthetases (NRPSs), a large multienzyme machineries that assemble peptides with diverse properties, such as siderophores, pigments, antibacterial, and antitumor. In contrast to RiPPs, NRPS represent mRNA-free template and building machinery for NRP biosynthesis. NRPs contain not only 20 normal amino acids, but also hundreds of unusual modified amino acids. Several modification enzymes are also found in NRP biosynthesis, such as halogenase, hydroxylase and P450 enzymes.[55] Therefore, NRPs reveal high diversity and complexity in their structures compared to peptides biosynthesized by RiPP system.[57, 58]

Nowadays, new sequencing techniques have been developed and bacterial genomes have been sequenced in a large scale. Accumulation of bacterial sequenced genomes in database provides chance to discover novel natural products by genome mining. Genome mining tools, such as antiSMASH,[74] PRISM,[129] RODEO,[72] have been used to predict biosynthetic

gene clusters (BGCs) of secondary metabolites. Comparison to classical bioactivity screening, genome mining approach is more rapid and convenient. Genome mining of RiPPs mainly targets conserved tailoring enzymes which lead to structural diversity of different classes of RiPPs and precursor peptides located nearby post-modification enzymes.[71] For instance, lasso peptides comprise of conserved sequences in precursors, B and C proteins, involved in lasso topology formation. The conserved patterns are successfully used for lasso peptide genome mining. Using McjB and McjC proteins involved in maturation of well-known lasso peptide MccJ25 as a query sequence, capistruin was discovered from *Burkholderia thailandensis* E264 as the first genome mining-based discovery lasso peptide.[42] For lanthipeptides, the biosynthetic machinery and tailoring enzymes such as Ser/Thr dehydratase are highly conserved. Searching for lanthipeptide BGCs using antiSMASH results in discovery of new lanthipeptide streptocollin from *Streptomyces collinus* Tu 365.[130] However, a number of novel RiPPs classes discovered via genome mining is limited compared to NRPS and PKS. Genome mining by targeting tailoring enzymes might be unable to discover novel family of RiPPs and using precursor peptide sequences as queries may results in discovery of analogous peptides. Recently, machine learning-based tools, such as RiPPMiner,[131] RiPPER[132] and DeepRiPP,[133] have been developed for RiPPs identification and classification. These tools able to identify novel precursor peptides, classify different classes of RiPPs and predict their cleavage sites by precursor peptide sequences in a gene context-independent manner. Therefore, combination of biology, bioinformatics and computer science will provide higher chance to discover novel class of RiPPs. [128] For NRPs, amino acid sequences of core biosynthetic enzymes are conserved even though NRPs possess huge diversity in structures and bioactivities. The conserved regions in core biosynthetic enzymes are used as query sequence to find analogous BGCs. In addition, the presence or absence of domains with tailoring enzyme activities in each module can be used to predict peptide with modifications. Also, the prediction of substrate of A domain which responsible for amino acid selection in each module have been reported. [69, 70] Several genome mining tools have been developed to identify NRPS gene clusters and determine adenylation domain specificities including NRPSpredictor2,[134] antiSMASH[74] and ClustScan.[135] However, the accurate determination of adenylation domain specificities remain difficult and genome mining tool sometime fail to predict NRP gene clusters since NRPs are nonribosomal code and most of them are modified by post-modification enzymes. NRPquest, a tool integrating genomic and mass spectrometric evidence for identification of NRPs, has been developed to perform mutation-tolerant and modification-tolerant searches of spectral data sets of NRPs.[136] So far, a number of bioactive peptide natural products including RiPPs and NRPs have been discovered based on genome mining.

Although genome mining has become a powerful tool and a number of bacterial secondary metabolite BGCs have been predicted from genome sequence database, the success in isolation of bioactive peptides natural products remains challenge. Since many BGCs are not expressed in laboratory conditions, activation of the cryptic BGCs or expression in heterologous host might be needed. Several bacterial strains have been used for heterologous expression of secondary metabolite such as *E. coli*, *S. lividans* and *S. coelicolor*. Moreover, several molecular engineering strategies such as mutagenesis and recombination have been used for strain improvement to enhance expression of secondary metabolite BGCs and increase product yields.[137] Antibiotic selections for mutations have been reported to enhance transcription and translation of secondary metabolite genes. This strategy is rapid, simple and inexpensive. For example, mutations in the *rpoB* gene encoding RNA polymerase β subunit of *S. lividans* resulting in upregulation of antibiotics, actinorhodin (Act) and undecylprodigiosin (Red). Rifampicin resistance mutants of *S. coelicolor* A3(2) have been also found to produce higher secondary metabolites compared to wide type strain. [111, 138] CRISPR/Cas9 genome editing technology has recently become the most popular tool for genetic manipulation in various organisms, including bacteria. This tool was used to perform site-directed mutagenesis and gene replacement for activation of silent BGCs in native Streptomyces host.[139, 140] Above all, the key of success in isolation of peptide natural products is to find the suitable ways to obtain high productivity for structure determination and biological test.

In this study, bioactive peptides derived from actinobacteria were target to investigate based on genome mining. We succeed in isolation and structure determination of three bioactive peptides from actinobacteria. Two peptides were biosynthesized via RiPPs system, and the other was biosynthesized via NRPS. These peptides exhibited wide variety of biological activities including antibacterial, anticancer, and anti-HIV. This is an example indicated that genome mining is a powerful tool for discovery of new bioactive peptide natural products with structural and functional diversities and may lead to discovery of new drugs. This study contributed not only to discovery of new drug seed compounds, but also to accumulation of information on biosynthesis of peptides which can be applied for natural product research.

CONCLUSION

Bacterial genome sequences are available in the database containing a large number and diversity of secondary metabolite biosynthetic gene clusters that relate to their encoded natural products. Consequently, genome mining approach is very useful for novel natural product discovery. Several genome mining tools, such as RODEO and antiSMASH, were developed and bacterial bioactive peptides including ribosomally synthesized and post-translationally modified peptides (RiPPs) and nonribosomal peptides (NRPs) have been discovered based on genome mining. RiPPs, peptides of ribosomal origin, are promising bacterial natural products for drug development due to their structural and physicochemical characteristics. RiPPs normally reveal wide variety of biological activities, such as antibacterial, antitumor and enzyme inhibitory activities, and many of them possess unique structures. For example, lasso peptides contain a motif of knot structure in the molecule which provides stability and resistance to protease. Thiazole/oxazole-modified microcins (TOMMs) are a class or RiPPs classified by the presence of heterocycles derived from serine, cysteine and threonine residues. The macrocyclic peptides were reported as potent enzyme inhibitors and cytotoxic compounds. On the other hand, NRPs are peptides biosynthesized by nonribosomal peptide synthetase (NRPS) which also known as a major source of bacterial natural products. The NRPSs are defined as large multienzyme machineries that assemble peptides with complex structures and diverse properties, such as siderophores, pigments, antibacterials, antitumor and immunosuppressants. Among bacteria, actinobacteria have been recognized as main producers of bioactive peptides including RiPPs and NRPs. This study focused on finding of new bioactive peptides derived from actinobacteria based on genome mining approach. As a result, three new bioactive peptides were discovered from *Streptomyces* sp. In chapter I, a new lasso peptide named specialicin was isolated from *S. specialis* JCM 16611^{T}. Specialicin showed moderate anti-HIV activity against HIV-1 NL4-3 and antibacterial activity against Gram-positive bacteria, *Micrococcus luteus*. In chapter II, a new cytotoxic peptide containing two isoleucine, two thiazole and three oxazole, was isolated from *S. curacoi* NBRC 12761^{T}. Curacozole exhibited potent cytotoxicity against HCT116 and HOS cancer cells and induced HCT116 cell apoptosis. In chapter III, a new antibacterial peptide named pentaminomycin C was isolated from *S. cacaoi* subsp. *cacaoi* NBRC 12748^{T}. Pentaminomycin C showed antibacterial activity against *Staphylococcus aureus*, *Bacillus subtilis* and *Micrococcus luteus*. These results indicated that actinobacteria is promising source of bioactive peptides natural products and genome mining is a powerful tool for discovery of novel bioactive peptides.

REFERENCES

1. Ludwig, W., et al., Road map of the phylum Actinobacteria, in Bergey's Manual of Systematic Bacteriology, Vol 5, 2nd Edn 2012. p. 1-28.
2. Ranjani, A., D. Dharumadurai, and G. P M, An introduction to actinobacteria, in Actinobacteria: Basics and Biotechnological Applications. 2016. p. 3-37.
3. Barka, E., et al., Taxonomy, physiology, and natural products of actinobacteria. Microbiology and Molecular Biology Reviews, 2015. 80: p. 1-43.
4. Grasso, L., D. Chillura Martino, and R. Alduina, Production of antibacterial compounds from actinomycetes, in Actinobacteria: Basics and Biotechnological Applications. 2016.
5. Ventura, M., et al., Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiology and Molecular Biology Reviews, 2007. 71(3): p. 495-548.
6. Zhi, X.-Y., W.-J. Li, and E. Stackebrandt, An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. International Journal of Systematic and Evolutionary Microbiology, 2009. 59(3): p. 589-608.
7. Hasani, A., A. Kariminik, and K. Issazadeh, Streptomycetes: characteristics and their antimicrobial activities. International Journal of Advanced Biological and Biomedical Research, 2014. 2(1): p. 63-75.
8. Anderson, A.S. and E.M. Wellington, The taxonomy of *Streptomyces* and related genera. International Journal of Systematic and Evolutionary Microbiology, 2001. 51(3): p. 797-814.
9. Bergey, D.H., et al., Bergey's manual of systematic bacteriology. Vol. Vol. 5. 2012, New York: Springer.
10. Procópio, R.E.d.L., et al., Antibiotics produced by *Streptomyces*. Brazilian Journal of Infectious Diseases, 2012. 16: p. 466-471.
11. Groll, A.H., S.C. Piscitelli, and T.J. Walsh, Clinical pharmacology of systemic antifungal agents: a comprehensive review of agents in clinical use, current investigational compounds, and putative targets for antifungal drug development, in Advances in Pharmacology, J.T. August, et al., Editors. 1998, Academic Press. p. 343-500.
12. Bentley, S.D., et al., Complete genome sequence of the model actinomycete

Streptomyces coelicolor A3(2). Nature, 2002. 417(6885): p. 141-147.

13. Ruiz, B., et al., Production of microbial secondary metabolites: regulation by the carbon source. Critical Reviews in Microbiology, 2010. 36(2): p. 146-167.
14. Bérdy, J., Bioactive microbial metabolites. The Journal of Antibiotics, 2005. 58(1): p. 1-26.
15. Singh, B.P., et al., Editorial: microbial secondary metabolites: recent developments and technological challenges. Frontiers in Microbiology, 2019. 10(914).
16. Lyddiard, D., G.L. Jones, and B.W. Greatrex, Keeping it simple: lessons from the golden era of antibiotic discovery. FEMS Microbiology Letters, 2016. 363(8).
17. Chain, E., et al., Penicillin as a chemotherapeutic agent. The Lancet, 1940. 236(6104): p. 226-228.
18. Aminov, R.I., A brief history of the antibiotic era: lessons learned and challenges for the future. Frontiers in Microbiology, 2010. 1: p. 134-134.
19. Takizawa, N. and M. Yamasaki, Current landscape and future prospects of antiviral drugs derived from microbial products. The Journal of Antibiotics, 2017. 71(1): p. 45-52.
20. Ishida, N., et al., Studies on the antiviral activity of formycin. The Journal of Antibiotics, 1967. 20(1): p. 49-52.
21. Richards, A.D., et al., Effective blocking of HIV-1 proteinase activity by characteristic inhibitors of aspartic proteinases. FEBS Letters, 1989. 247(1): p. 113-117.
22. Umezawa, H., et al., Purification and characterization of a sialidase inhibitor, siastatin, produced by *Streptomyces*. The Journal of Antibiotics, 1974. 27(12): p. 963-9.
23. Nishimura, Y., et al., Siastatin B, a potent neuraminidase inhibitor: the total synthesis and absolute configuration. Journal of the American Chemical Society, 1988. 110(21): p. 7249-7250.
24. Hollstein, U., Actinomycin. Chemistry and mechanism of action. Chemical Reviews, 1974. 74(6): p. 625-652.
25. Goto, T., et al., Discovery of FK-506, a novel immunosuppressant isolated from *Streptomyces tsukubaensis*. Transplantation Proceedings, 1987. 19(5): p. 4-8.
26. Pham, J.V., et al., A review of the microbial production of bioactive natural products and biologics. Frontiers in Microbiology, 2019. 10(1404).
27. Li, J., S.G. Kim, and J. Blenis, Rapamycin: one drug, many effects. Cell metabolism, 2014. 19(3): p. 373-379.
28. Bhandari, D., et al., A review on bioactive peptides: physiological functions, bioavailability and safety. International Journal of Peptide Research and Therapeutics,

2020. 26: p. 139-150.

29. Shabanpoor, F., F. Separovic, and J.D. Wade, The human insulin superfamily of polypeptide hormones. Vitamins and Hormones, 2009. 80: p. 1-31.
30. Dang, T. and R.D. Süssmuth, Bioactive peptide natural products as lead structures for medicinal use. Accounts of Chemical Research, 2017. 50(7): p. 1566-1576.
31. Shyangdan, D., et al., Liraglutide for the treatment of type 2 diabetes. Health Technology Assessment, 2011. 15 Suppl 1: p. 77-86.
32. Arnison, P.G., et al., Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Natural Product Reports, 2013. 30(1): p. 108-60.
33. McIntosh, J.A., M.S. Donia, and E.W. Schmidt, Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. Natural Product Reports, 2009. 26(4): p. 537-559.
34. Yang, X. and W.A. van der Donk, Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. Chemistry (Weinheim an der Bergstrasse, Germany), 2013. 19(24): p. 7662-7677.
35. Hegemann, J.D., et al., Lasso peptides: an intriguing class of bacterial natural products. Accounts of Chemical Research, 2015. 48(7): p. 1909-19.
36. Maksimov, M.O., S.J. Pan, and A. James Link, Lasso peptides: structure, function, biosynthesis, and engineering. Natural Product Reports, 2012. 29(9): p. 996-1006.
37. Li, Y., S. Zirah, and S. Rebuffat, Lasso Peptides: Bacterial Strategies to Make and Maintain Bioactive Entangled Scaffolds. 2015: Springer.
38. Li, Y., et al., Characterization of sviceucin from *streptomyces* provides insight into enzyme exchangeability and disulfide bond formation in lasso peptides. ACS Chemical Biology, 2015. 10(11): p. 2641-9.
39. Potterat, O., et al., Aborycin A -tricyclic 21-peptide antibiotic isolated from *Streptomyces griseoflavus*. Liebigs Annalen der Chemie, 1994. 1994(7): p. 741-743.
40. Detlefsen, D.J., et al., Siamycins I and II, new anti-HIV-1 peptides: II. Sequence analysis and structure determination of siamycin I. The Journal of Antibiotics, 1995. 48(12): p. 1515-7.
41. Salomon, R.A. and R.N. Farias, Microcin 25, a novel antimicrobial peptide produced by *Escherichia coli*. Journal of Bacteriology, 1992. 174(22): p. 7428-35.
42. Knappe, T.A., et al., Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264.

Journal of the American Chemical Society, 2008. 130(34): p. 11446-11454.

43. Kimura, K., et al., Propeptin, a new inhibitor of prolyl endopeptidase produced by Microbispora. I. Fermentation, isolation and biological properties. The Journal of Antibiotics, 1997. 50(5): p. 373-8.

44. Elsayed, S.S., et al., Chaxapeptin, a lasso peptide from extremotolerant *Streptomyces leeuwenhoekii* strain C58 from the hyperarid atacama desert. The Journal of Organic Chemistry, 2015. 80(20): p. 10252-60.

45. Morishita, Y., et al., RES-701-1, a novel and selective endothelin type B receptor antagonist produced by *Streptomyces* sp. RE-701. I. Characterization of producing strain, fermentation, isolation, physico-chemical and biological properties. The Journal of Antibiotics, 1994. 47(3): p. 269-75.

46. Potterat, O., et al., BI-32169, a bicyclic 19-peptide with strong glucagon receptor antagonist activity from *Streptomyces* sp. Journal of Natural Products, 2004. 67(9): p. 1528-31.

47. Solbiati, J.O., et al., Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. Journal of Bacteriology, 1999. 181(8): p. 2659-62.

48. Yan, K.P., et al., Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. Chembiochem, 2012. 13(7): p. 1046-52.

49. Kuznedelov, K., et al., The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase. Journal of Molecular Biology, 2011. 412(5): p. 842-848.

50. Tsunakawa, M., et al., Siamycins I and II, new anti-HIV peptides: I. Fermentation, isolation, biological activity and initial characterization. The Journal of Antibiotics, 1995. 48(5): p. 433-4.

51. Um, S., et al., Sungsanpin, a lasso peptide from a deep-sea streptomycete. Journal of Natural Products, 2013. 76(5): p. 873-9.

52. Süssmuth, R.D. and A. Mainz, Nonribosomal peptide synthesis—principles and prospects. Angewandte Chemie International Edition in English, 2017. 56(14): p. 3770-3821.

53. Martínez-Núñez, M.A. and V.E.L.y. López, Nonribosomal peptides synthetases and their applications in industry. Sustainable Chemical Processes, 2016. 4(1): p. 13.

54. Walsh, C.T., Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. Natural Product Reports, 2016. 33(2): p. 127-35.

55. Walsh, C.T., et al., Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Current Opinion in Chemical Biology,

2001. 5(5): p. 525-534.
56. Kittilä, T., et al., Halogenation of glycopeptide antibiotics occurs at the amino acid level during non-ribosomal peptide synthesis. Chemical science, 2017. 8(9): p. 5992-6004.
57. Schwarzer, D., R. Finking, and M.A. Marahiel, Nonribosomal peptides: from genes to products. Natural Product Reports, 2003. 20(3): p. 275-87.
58. Wilkinson, B. and J. Micklefield, Chapter 14 Biosynthesis of nonribosomal peptide precursors, in Methods in Enzymology. 2009, Academic Press. p. 353-378.
59. Sieber, S.A. and M.A. Marahiel, Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chemical Reviews, 2005. 105(2): p. 715-38.
60. Dinos, G.P., et al., chloramphenicol derivatives as antibacterial and anticancer agents: historic problems and current solutions. Antibiotics (Basel), 2016. 5(2).
61. Steenbergen, J.N., et al., Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections. Journal of Antimicrobial Chemotherapy, 2005. 55(3): p. 283-288.
62. Henson, K.E., et al., Glycopeptide antibiotics: evolving resistance, pharmacology and adverse event profile. Expert Review of Anti-infective Therapy, 2015. 13(10): p. 1265-78.
63. Hubbard, B.K. and C.T. Walsh, Vancomycin assembly: nature's way. Angewandte Chemie International Edition, 2003. 42(7): p. 730-765.
64. Levine, D.P., Vancomycin: a history. Clinical Infectious Diseases, 2006. 42 Suppl 1: p. S5-12.
65. Marks, T.A. and J.M. Venditti, Potentiation of actinomycin D or adriamycin antitumor activity with DNA. Cancer Research, 1976. 36(2 Part 1): p. 496-504.
66. Dorr, R.T., et al., Analytical and biological inequivalence of two commercial formulations of the antitumor agent bleomycin. Cancer Chemotherapy and Pharmacology, 1998. 42(2): p. 149-154.
67. Tedesco, D. and L. Haragsim, Cyclosporine: a review. Journal of Transplantation, 2012. 2012: p. 230386-230386.
68. Hornung, A., et al., A genomic screening approach to the structure-guided identification of drug candidates from natural sources. Chembiochem, 2007. 8(7): p. 757-66.
69. Fischbach, M.A. and C.T. Walsh, Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. Chemical Reviews, 2006. 106(8): p. 3468-96.
70. Challis, G.L., Mining microbial genomes for new natural products and biosynthetic

pathways. Microbiology, 2008. 154(Pt 6): p. 1555-69.

71. Burkhart, B.J., et al., A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nature Chemical Biology, 2015. 11: p. 564.
72. Tietz, J.I., et al., A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nature Chemical Biology, 2017. 13(5): p. 470-478.
73. Medema, M.H., et al., antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research, 2011. 39(Web Server issue): p. W339-46.
74. Blin, K., et al., antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research, 2019. 47(W1): p. W81-W87.
75. Kaweewan, I., M. Ohnishi-Kameyama, and S. Kodani, Isolation of a new antibacterial peptide achromosin from *Streptomyces achromogenes* subsp. *achromogenes* based on genome mining. The Journal of Antibiotics, 2016. 70: p. 208.
76. Kaweewan, I., et al., Isolation and structure determination of a new thiopeptide globimycin from *Streptomyces globisporus* subsp. *globisporus* based on genome mining. Tetrahedron Letters, 2018. 59(4): p. 409-414.
77. Kaweewan, I., et al., Isolation and structure determination of new antibacterial peptide curacomycin based on genome mining. Asian Journal of Organic Chemistry, 2017. 6(12): p. 1838-1844.
78. Maksimov, M.O. and A.J. Link, Discovery and characterization of an isopeptidase that linearizes lasso peptides. Journal of the American Chemical Society, 2013. 135(32): p. 12038-47.
79. Yano, K., et al., MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from *Streptomyces* sp. I. Isolation, structural determination and biological properties of MS-271. Bioorganic & Medicinal Chemistry, 1996. 4(1): p. 115-120.
80. Nakayama, J., et al., Siamycin attenuates fsr quorum sensing mediated by a gelatinase biosynthesis-activating pheromone in Enterococcus faecalis. Journal of Bacteriology, 2007. 189(4): p. 1358-65.
81. Feng, Z., et al., Biosynthetic gene cluster of a D-tryptophan-containing lasso peptide, MS-271. ChemBioChem, 2018.
82. Shirling, E.B. and D. Gottlieb, Methods for characterization of *Streptomyces* species. International Journal of Systematic Bacteriology, 1966. 16: p. 313-340.
83. Wuthrich, K., M. Billeter, and W. Braun, Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. Journal of

Molecular Biology, 1983. 169(4): p. 949-61.

84. Clore, G.M., et al., Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. Biochemistry, 1987. 26(24): p. 8012-23.

85. Brunger, A.T., et al., Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta crystallographica. Section D, Biological crystallography, 1998. 54(Pt 5): p. 905-21.

86. Metelev, M., et al., Structure, bioactivity, and resistance mechanism of streptomonomicin, an unusual lasso peptide from an understudied halophilic actinomycete. Chemistry & Biology, 2015. 22(2): p. 241-50.

87. Koradi, R., M. Billeter, and K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures. Journal of Molecular Graphics, 1996. 14(1): p. 51-5, 29-32.

88. Laskowski, R.A., et al., AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR, 1996. 8(4): p. 477-86.

89. Muramoto, K. and H. Kamiya, Recovery of tryptophan in peptides and proteins by high-temperature and short-term acid hydrolysis in the presence of phenol. Analytical Biochemistry, 1990. 189(2): p. 223-30.

90. Manneberg, M., H.W. Lahm, and M. Fountoulakis, Quantification of cysteine residues following oxidation to cysteic acid in the presence of sodium azide. Analytical Biochemistry, 1995. 231(2): p. 349-53.

91. Di Modugno, E., et al., In vitro activity of the tribactam GV104326 against gram-positive, gram-negative, and anaerobic bacteria. Antimicrobial Agents and Chemotherapy, 1994. 38(10): p. 2362-8.

92. Yang, C.L., et al., Strepchazolins A and B: Two new alkaloids from a marine *Streptomyces chartreusis* NA02069. Marine Drugs, 2017. 15(8).

93. Yoshimura, K., et al., Enhanced exposure of human immunodeficiency virus type 1 primary isolate neutralization epitopes through binding of CD4 mimetic compounds. Journal of Virology, 2010. 84(15): p. 7558-68.

94. Katahira, R., et al., MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from *Streptomyces* sp. 2. Solution structure of MS-271: Characteristic features of the 'lasso' structure. Bioorganic & Medicinal Chemistry, 1996. 4(1): p. 121-129.

95. Kampfer, P., et al., *Streptomyces phaeopurpureus* Shinobu 1957 (Approved Lists 1980)

and *Streptomyces griseorubiginosus* (Ryabova and Preobrazhenskaya 1957) Pridham et al. 1958 (Approved Lists 1980) are heterotypic subjective synonyms. International Journal of Systematic and Evolutionary Microbiology, 2017. 67(8): p. 3111-3116.

96. Harada, K.I., et al., Application of D,L-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method. Tetrahedron Letters, 1996. 37(17): p. 3001-3004.
97. Thundimadathil, J., Cancer treatment using peptides: current therapies and future prospects. Journal of Amino Acids, 2012. 2012: p. 967347-967347.
98. Kapp, T.G., et al., A comprehensive evaluation of the activity and selectivity profile of ligands for RGD-binding integrins. Scientific Reports, 2017. 7: p. 39805-39805.
99. Vinogradov, A.A., Y. Yin, and H. Suga, Macrocyclic peptides as drug candidates: recent progress and remaining challenges. Journal of the American Chemical Society, 2019. 141(10): p. 4167-4181.
100. Rubin, S. and N. Qvit, Cyclic peptides for protein-protein interaction targets: applications to human disease. Critical Reviews in Eukaryotic Gene Expression, 2016. 26(3): p. 199-221.
101. Shin-ya, K., et al., Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. Journal of the American Chemical Society, 2001. 123(6): p. 1262-3.
102. Kim, M.Y., et al., Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. Journal of the American Chemical Society, 2002. 124(10): p. 2098-2099.
103. Amagai, K., et al., Identification of a gene cluster for telomestatin biosynthesis and heterologous expression using a specific promoter in a clean host. Scientific Reports, 2017. 7(1): p. 3382.
104. Sohda, K.Y., et al., YM-216391, a novel cytotoxic cyclic peptide from *Streptomyces nobilis*. I. fermentation, isolation and biological activities. The Journal of Antibiotics, 2005. 58(1): p. 27-31.
105. Sohda, K.Y., et al., YM-216391, a novel cytotoxic cyclic peptide from *Streptomyces nobilis*. II. Physico-chemical properties and structure elucidation. The Journal of Antibiotics, 2005. 58(1): p. 32-6.
106. Jian, X.H., et al., Analysis of YM-216391 biosynthetic gene cluster and improvement of the cyclopeptide production in a heterologous host. ACS Chemical Biology, 2012. 7(4): p. 646-51.
107. Shima, J., et al., Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and

Streptomyces coelicolor A3(2). Journal of Bacteriology, 1996. 178(24): p. 7276-7284.

108. Ochi, K. and T. Hosaka, New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. Applied Microbiology and Biotechnology, 2013. 97(1): p. 87-98.
109. Thong, W.L., et al., Methylbenzene-containing polyketides from a *Streptomyces* that spontaneously acquired rifampicin resistance: structural elucidation and biosynthesis. Journal of Natural Products, 2016. 79(4): p. 857-64.
110. Hosaka, T., et al., Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. Nature Biotechnology, 2009. 27(5): p. 462-4.
111. Hu, H., Q. Zhang, and K. Ochi, Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. Journal of Bacteriology, 2002. 184(14): p. 3984-91.
112. Claesen, J. and M. Bibb, Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proceedings of the National Academy of Sciences of the United States of America, 2010. 107(37): p. 16297-302.
113. Zeng, J. and J. Zhan, Characterization of a tryptophan 6-halogenase from *Streptomyces toxytricini*. Biotechnology Letters, 2011. 33(8): p. 1607-13.
114. Sussmuth, R.D. and A. Mainz, Nonribosomal peptide synthesis-principles and prospects. Angewandte Chemie International Edition in English, 2017. 56(14): p. 3770-3821.
115. Curtis, R.W., Curvatures and malformations in bean plants caused by culture filtrate of *Aspergillus niger.* Plant physiology, 1958. 33(1): p. 17-22.
116. Curtis, R.W., Root curvatures induced by culture filtrates of *Aspergillus niger*. Science, 1958. 128(3325): p. 661.
117. Wang, J., et al., Study of malformin C, a fungal source cyclic pentapeptide, as an anti-cancer drug. PLoS ONE, 2015. 10(11): p. e0140069.
118. Liu, Y., et al., Malformin A1 promotes cell death through induction of apoptosis, necrosis and autophagy in prostate cancer cells. Cancer Chemotherapy and Pharmacology, 2016. 77(1): p. 63-75.
119. Koizumi, Y., et al., Structure–activity relationship of cyclic pentapeptide malformins as fibrinolysis enhancers. Bioorganic & Medicinal Chemistry Letters, 2016. 26(21): p. 5267-5271.
120. Nakamura, F., et al., Kakeromamide A, a new cyclic pentapeptide inducing astrocyte differentiation isolated from the marine cyanobacterium *Moorea bouillonii*. Bioorganic

& Medicinal Chemistry Letters, 2018. 28(12): p. 2206-2209.

121. Kojiri, K., et al., Endothelin-binding inhibitors, BE-18257A and BE-18257B. I. Taxonomy, fermentation, isolation and characterization. The Journal of Antibiotics, 1991. 44(12): p. 1342-7.

122. Nakajima, S., et al., Endothelin-binding inhibitors, BE-18257A and BE-18257B II. Structure determination. The Journal of Antibiotics, 1991. 44(12): p. 1348-56.

123. Jang, J.-P., et al., Pentaminomycins A and B, hydroxyarginine-containing cyclic pentapeptides from *Streptomyces* sp. RK88-1441. Journal of Natural Products, 2018. 81(4): p. 806-810.

124. Inglis, A.S., P.W. Nicholls, and C.M. Roxburgh, Hydrolysis of the peptide bond and amino acid modification with hydriodic acid. Australian Journal of Biological Sciences, 1971. 24(6): p. 1235-40.

125. Prieto, C., et al., NRPSsp: non-ribosomal peptide synthase substrate predictor. Bioinformatics, 2012. 28(3): p. 426-7.

126. Kuranaga, T., et al., Total synthesis of the nonribosomal peptide surugamide B and identification of a new offloading cyclase family. Angewandte Chemie International Edition, 2018. 57(30): p. 9447-9451.

127. Matsuda, K., et al., SurE is a trans-acting thioesterase cyclizing two distinct non-ribosomal peptides. Organic & Biomolecular Chemistry, 2019. 17(5): p. 1058-1061.

128. Zhong, Z., et al., Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). Synthetic and Systems Biotechnology, 2020. 5(3): p. 155-172.

129. Skinnider, M.A., et al., PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Research, 2017. 45(W1): p. W49-w54.

130. Iftime, D., et al., Streptocollin, a Type IV Lanthipeptide Produced by *Streptomyces collinus* Tü 365. Chembiochem, 2015. 16(18): p. 2615-23.

131. Agrawal, P., et al., RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. Nucleic Acids Research, 2017. 45(W1): p. W80-w88.

132. Santos-Aberturas, J., et al., Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. Nucleic Acids Research, 2019. 47(9): p. 4624-4637.

133. Merwin, N.J., et al., DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. Proceedings of the National Academy

of Sciences, 2020. 117(1): p. 371-380.

134. Röttig, M., et al., NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Research, 2011. 39(Web Server issue): p. W362-7.

135. Starcevic, A., et al., ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Research, 2008. 36(21): p. 6882-6892.

136. Mohimani, H., et al., NRPquest: coupling mass spectrometry and genome mining for nonribosomal peptide discovery. Journal of natural products, 2014. 77(8): p. 1902-9.

137. Baltz, R.H., Strain improvement in actinomycetes in the postgenomic era. Journal of Industrial Microbiology & Biotechnology, 2011. 38(6): p. 657-666.

138. Xu, J., et al., A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). Molecular Genetics and Genomics, 2002. 268(2): p. 179-89.

139. Bramhachari, P.V., A.B. Pinjari, and E. Kariali, Chapter 22 - Genomics of actinobacteria with a focus on natural product biosynthetic genes, in new and future developments in microbial biotechnology and bioengineering, B.P. Singh, V.K. Gupta, and A.K. Passari, Editors. 2018, Elsevier. p. 325-335.

140. Barrangou, R. and Luciano A. Marraffini, CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Molecular Cell, 2014. 54(2): p. 234-244.