Exploring the biosynthetic potential of *Cystobacter* sp. SBCb004

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Hu Zeng

Saarbrücken

2021

Tag des Kolloquiums: 31.05.2021

Dekan:	Prof. Dr. Jörn Walter
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Andriy Luzhetskyy
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Michael Kohlstedt

Die vorliegende Arbeit wurde von September 2016 bis Februar 2021 unter der Anleitung von Herrn Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) angefertigt.

Acknowledgement

I would like to thank my supervisor Prof. Rolf Mueller for offering me the chance to have my PhD in HIPS. You also established a good example of scientific researcher for me.

I would like to thank my second supervisor Prof. Andriy Luzhetskyy for always giving good advices to my projects.

I would like to thank my mentor Dr. Carsten Volz for supervising me in the last four years. Thank you for your patience to answer my endless questions. Your sense of humor also changed my stereotype about German and brought a lot of fun to the lab.

I would like to thank all the Chinese colleagues including Jun. Prof. Dr. Chengzhang Fu, Dr. Feng Xie, Dr. Yunkun Liu, Dr. Qiushui Wang, Yunsheng Gao, Tingting Wang and Jingjun Mo. Your focus and devotion to scientific work influenced me a lot and inspired me to work hard. I also want to thank you guys for always being so nice to share your work experience, to discuss with me and to encourage me.

I would like to thank Dr. Alexander Popoff, Dr. Asfandyar Sikandar, Dr. Bastien Schnell, Jan Dastbaz, Dr. Joachim Hug, Joy Birkelbach, Laura Franz, Dr. Maja Remskar, Patrick Haack, Dr. Sebastian Adam and Sebastian Walesch for always inviting me to go hiking and biking. I had the most relaxing time during those hiking and biking tours with you guys. I would like to thank Selina Deckarm for always bringing delicious cakes to us.

I would like to thank Dr. Alexander Popoff, Joy Birkelbach, Dr. Lena Keller and Dr. Judith Hoffmann for elucidating the structure of the compound for me. Thanks to Chantal Bader, Christine Walt, Dr. Fabian Panter, Laura Franz, Nicolas Frank and Patrick Haack for measuring my samples on Maxis4G, as well as giving me suggestions on compound isolation and purification. Thanks to Daniel Sauer for teaching me to manipulate semipreparative HPLC. Thanks to Claudia Helbig for providing me all the experimental materials and kits in time. Thanks to Stefanie Schmitt for doing the bioassay for me. Thanks to Markus Neuber for helping me manipulating fermenter. Thanks to Dr. Domen Pogorevc, Jan Schlemmer and Dr. Sebastian Groß for sharing your experience on myxobacteria and work out. Thanks to Christina Decker and Ellen Merckel for providing all the information I need and arranging all the appointments.

I would like to thank Dr. Hui Shuai for leading me to the most beautiful sceneries and the most delicious restaurants in Saarland and being my travel partner. Thanks for being my barber in the last four years.

I would like to thank my parents for their understanding and support in the last four years.

I would like to thank the financial support from China Scholarship Council in the last four years and thank the substantial support from Chinese Embassy during corona pandemic.

Zusammenfassung

Myxobakterien sind eine ergiebige Quelle für Naturstoffe mit einzigartigen chemischen Strukturen und pharmazeutischem Anwendungspotenzial. Ihr biosynthetisches Potenzial ist jedoch noch wenig erforscht. Diese Arbeit konzentriert sich auf die Erforschung des verborgenen biosynthetischen Potentials des Myxobakteriums *Cystobacter* sp. SBCb004. Im ersten Kapitel wurden durch die Kombination von Genom- und Metabolomanalysen ein Ajudazol-ähnliches biosynthetisches Gencluster (BGC) und eine Reihe neuartiger Ajudazol-Derivate entdeckt, in deren Biosynthese vermutlich eine 3,3-Dimethylacrylyl-CoA Startereinheit verwendet wird. Außerdem unterscheiden sich die Derivate im Grad der Redoxreaktion und im Glykosylierungsmuster. Mithilfe eines Tailoring-Modifikationsnetzwerks wurden zwei P450-abhängige Enzyme und eine Glykosyltransferase postuliert, deren Beteiligung an der Biosynthese nach einem Mutagenese-Experiment bewiesen wurden. Im zweiten Kapitel wurde durch Metabolomanalyse die neuartige Peptidfamilie der Cystopipecotide identifiziert, welche die bislang unbekannte 5-Hydroxyl-6-hydroxymethylpipecolinsäure (HHMPA) als Baustein enthalten. Das entsprechende BGC konnte ebenfalls identifiziert werden. Darüber hinaus konnte durch eine *in vitro* Rekonstruktion des gesamten Stoffwechselweges gezeigt werden, dass die HHMPA-Biosynthese von aufeinander folgenden Reaktionen der Transketolasen und Lysin-biosynthetischen Enzymen abhängt. Dies stellt einen bislang unbekannten Weg zur Biosynthese von substituierten Pipecolaten in Mikroben dar.

Abstract

Myxobacteria are a fruitful source of natural products, which often feature unique chemical structures and potential pharmaceutical applications. However, their biosynthetic potential is underexplored. This thesis focused on exploring the hidden biosynthetic potential of myxobacterium *Cystobacter* sp. SBCb004. In chapter one, genomic analyses coupled with metabolomics analyses revealed an ajudazol like BGC and a set of novel ajudazol derivatives, which presumably use 3,3-dimethylacrylyl-CoA as a starter unit and are characterized by varying degrees of oxidoreduction and different glycosylation patterns. A promiscuous tailoring modification network was proposed and two P450 dependent enzymes and one glycosyltransferase were postulated to be involved according to the mutagenesis work. In chapter two, metabolic profiling uncovered a family of novel peptides later termed cystopipecotides, which exhibit the previously unknown 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) as a building block. The corresponding BGC was identified and the HHMPA biosynthesis was shown to be dependent on concerted actions of transketolases and lysine biosynthetic enzymes by reconstitution of the full pathway *in vitro*. It represents an unprecedented route to biosynthesize substituted pipecolate in microbes.

Table of contents

1. Introduction

1.1 Natural products

Natural products represent a large family of diverse chemical compounds produced by living organisms. They have been widely used throughout human history, notably in the field of pharmaceutics and medicine. The earliest documented use of natural products as medicine could be dated back to 2600 BCE as 1000 plant-derived substances were used in traditional medicine in Mesopotamia[1]. To date, approximately 50 % of the marketed drugs in human therapy are either direct natural products or such based on natural product scaffolds[2].

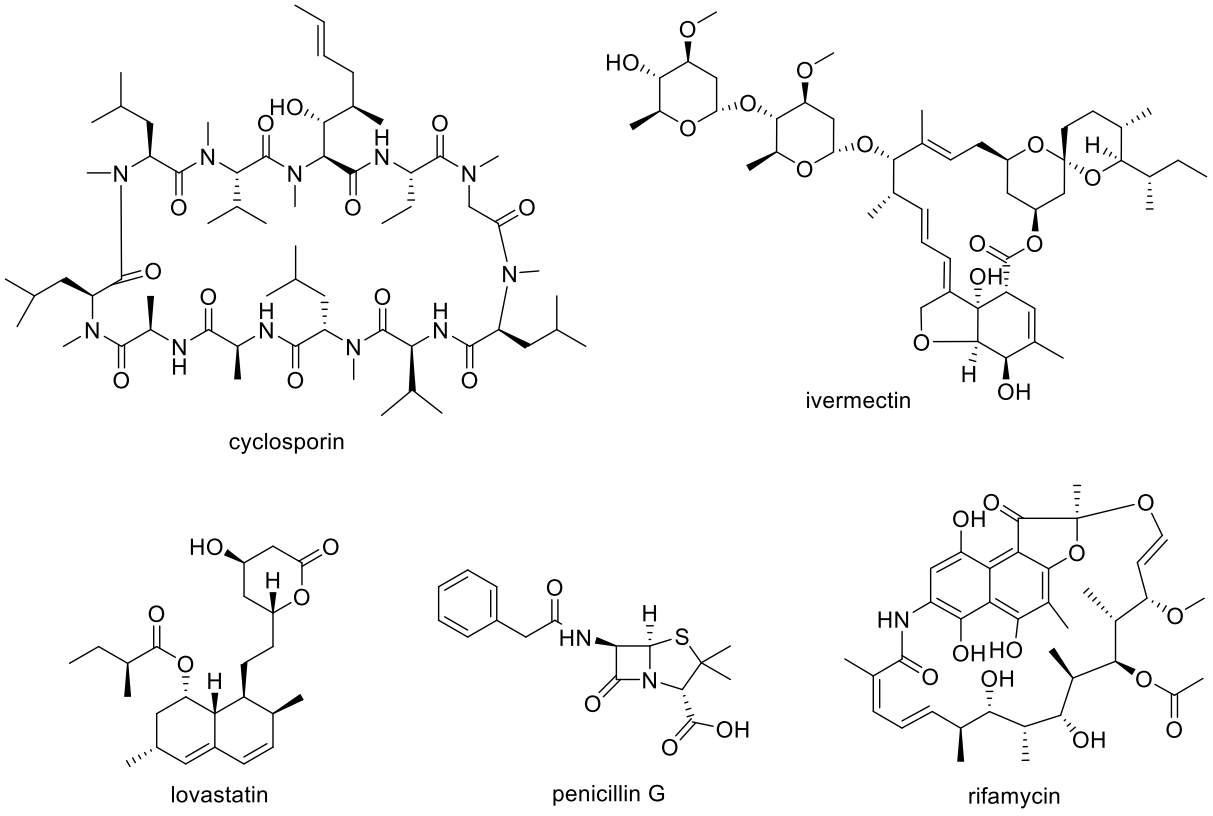


Figure 1.1.1 Chemical structures of some pharmaceutically valuable natural products from fungi and bacteria.

The discovery of penicillin from the filamentous fungus *Penicillium* in 1929, followed by the discoveries of actinomycin, streptothricin, and streptomycin by systematic screening of actinomycetes in the 1940s, prompted the screening of microorganisms' extracts and

ushered the so-called "Golden Age of Antibiotics"[3]. Natural products from bacteria and fungi are also called secondary metabolites. In many cases, these compounds are not essential to the host under laboratory conditions. Nevertheless, secondary metabolites have been shown to be involved in numerous important processes in the natural environment. In such an environment, their producers benefit from the action of such natural compounds. These compounds exhibit both, an impressive chemical diversity and complexity. Based on the mechanism of action, they may possess highly selective and specific biological activities, what makes them valuable candidates in drug development. Well-known representatives of such natural compounds are penicillin (antibacterial agent), cyclosporin (immunosuppressive agent), lovastatin (cholesterol lowering agents), rifamycin (anti-tuberculosis drug) and ivermectin (antiparasitic drug)[2] (Figure 1.1.1).

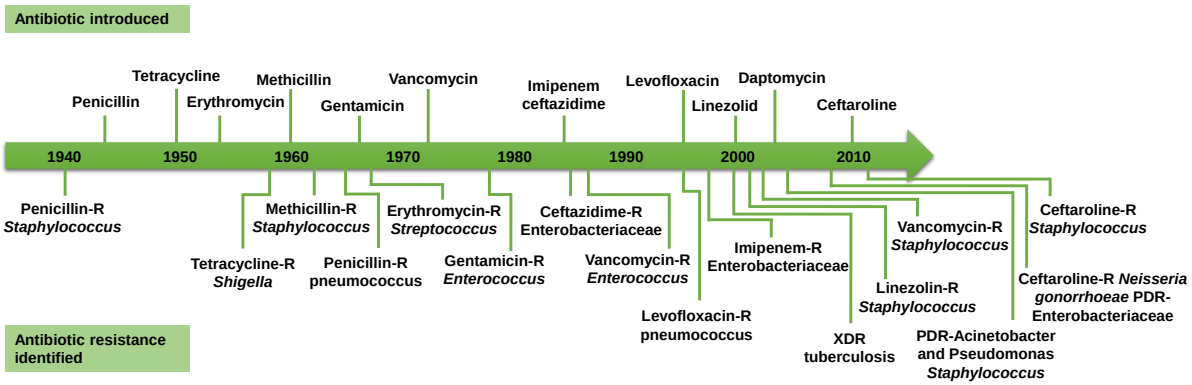


Figure 1.1.2 Developing antibiotic and antibiotic resistance: a timeline of key events[4]. R: resistant. XDR: extensively drug-resistance. PDR: pan-drug-resistant. Note: before widespread usage in 1943, penicillin was in limited use.

However, after the 1960s, there was a strong decrease in discovery of novel antibiotics (Figure 1.1.2). Only a few new chemical lead structures were described in the last decades and the majority was variations of known structures. Even actinomycetes, which produce most of the natural products characterized so far from bacteria, seemed to be exploited. Due to the increasing number of antibiotic resistances resulting from the widespread and uncontrolled use of antibiotics (Figure 1.1.2), there is an urgent need to describe new pharmaceutical lead structures. However, it has been estimated that microorganisms from all origins which have been cultivated successfully in the laboratory depict less than 1 % of all the species present in these habitats[5]. Consequently, this might imply that we only see

the tip of the iceberg regarding the number of all possibly produced secondary metabolites. Moreover, genome sequencing efforts revealed that the biosynthetic potential of microorganisms seems to be by far untapped as there are much more biosynthetic gene clusters (BGCs) than known secondary metabolites and microbes possessing large genomes might have a much higher potential to produce secondary metabolites than ever recognized. Amongst these microorganisms, myxobacteria may also depict a promising underexplored biosynthetic reservoir.

1.2 Myxobacteria, a prolific source of microbial natural products

Myxobacteria are Gram-negative δ-proteobacteria, which ubiquitously exist in soil, while they have also been isolated from other environments including oceans, freshwater lakes, decaying plants and herbivore dung[6]. They are well-known for their unique social behaviors, including the ability to move by gliding or creeping on solid surfaces and to form multicellular fruiting bodies harboring propagative spores under starvation conditions[7] (Figure 1.2.1). By secreting exo-enzymes they are able to use a wide variety of biological macromolecules as a food source and are also able to prey actively on microorganisms.

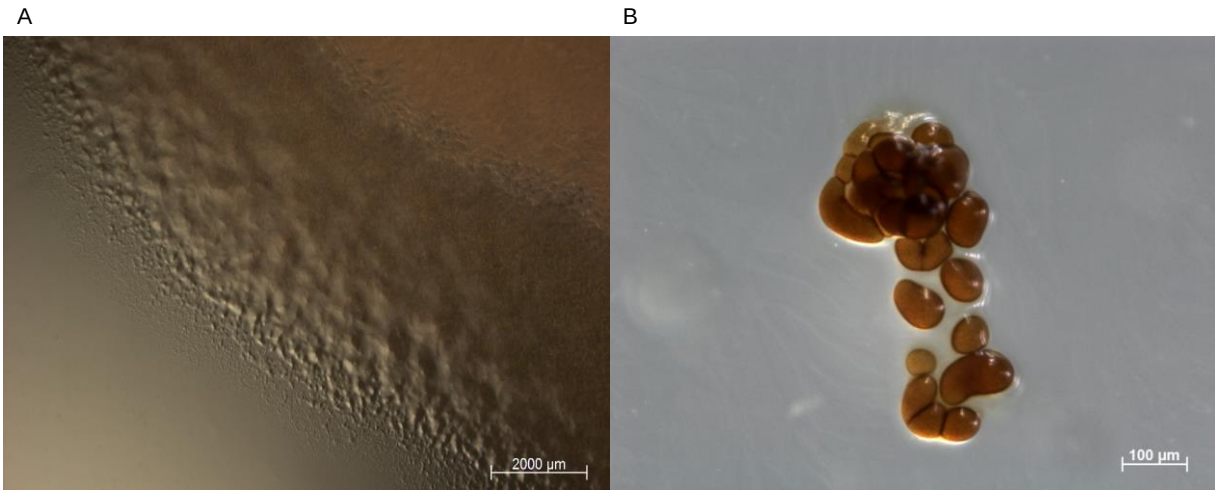


Figure 1.2.1 Swarming behavior (A) and fruiting body (B) of *Cystobacter* sp. SBCb004.

Another notable feature of myxobacteria is their extraordinary ability to produce secondary metabolites, which makes them one of the most important producers of natural products[8,9]. In the past 35 years, over 100 distinct core structures and over 600 derivatives thereof have

been isolated from over 9000 myxobacterial strains[10]. These structures exhibit antifungal (e.g. myxothiazol), antibacterial (e.g. cystobactamid), antimalarial (e.g. myxovalargin), antitumor (e.g. epothilon), antiviral (e.g. etnangien) and anti-immunomodulatory (e.g. argyrin) properties[11] and belong to numerous classes of secondary metabolites (Figure 1.2). The majority are polyketides (e.g. pyxidicycline[12]), non-ribosomal peptides (e.g. cystobactamid and argyrin), and their hybrids (e.g. epothilone, which is clinically used in chemotherapy)[11], while others include terpenoids (e.g. cystodienoeic acid[13]), and ribosomally synthesized and post-translationally modified peptides (RiPPs) (e.g. cittilin[14] and crocagin[15]) (Figure 1.2.2). In addition, many myxobacteria could produce compounds belonging to multiple structural families and many of the compounds exhibit both, unique structural features and novel modes-of-actions which are rarely observed from other microorganisms[9,16]. Furthermore, myxobacteria exhibit genome sizes ranging from 9 to 16 Mbp and thus possess some of the largest genomes known so far from prokaryotes[17]. As bigger genome size usually means the potential capacity of more gene clusters, the unexplored biosynthetic potential of myxobacteria is beyond characterized compounds described.

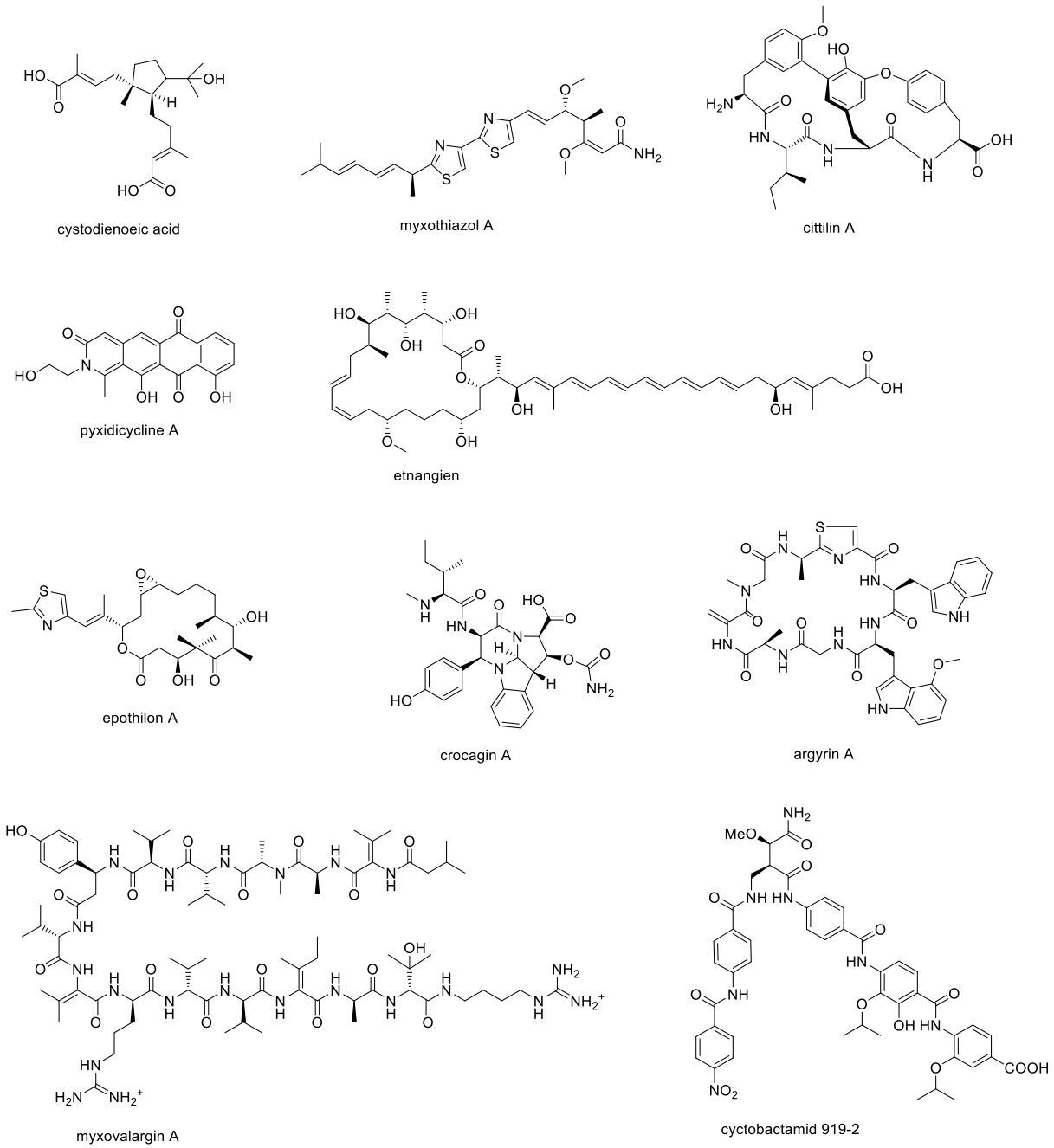


Figure 1.2.2 Chemical structures of myxobacterial secondary metabolites.

1.3 Biosynthesis of microbial secondary metabolites

1.3.1 Biosynthesis of polyketides

Polyketides represent a highly diverse group of natural products comprising polyphenols, macrolides, polyenes, enediynes, and polyethers. These compounds are synthesized by modular megaenzymes termed polyketide synthases (PKS). In general, polyketides biosynthesis is initiated with the loading of an activated acyl starter unit and the polyketide

chain is elongated via stepwise Claisen-like condensation steps. Typically, each chain elongation module minimally contains a ketosynthase (KS) domain which links the acyl unit to the growing polyketide chain via Claisen-like chemistry, an acyl transferase (AT) domain which recruits the building block to the module and an acyl carrier protein (ACP) which acts as an anchor for the growing chain. In addition, PKS modules can harbor further accessory domains varying the polyketide skeleton, including ketoreductase (KR) domain which reduces the nascent β-keto functions to give hydroxyl groups, dehydratase (DH) domain that eliminates water to generate olefinic moieties, and enoylreductase (ER) domain that further reduces the double bond to saturated moieties. At the end of the assembly line, the finished chain is generally released from the assembly line by a thioesterase (TE) domain by hydrolysis or more commonly macrolactonization using an internal hydroxyl nucleophile[17]. The products released from PKS are usually tailored by various post-PKS enzymes such as oxidoreductases, glycosyltransferases, methyltransferases, aminotransferases and halogenases, which diversify the structures and confer bioactivity on the final metabolites[18,19] (Figure 1.3.1.1).

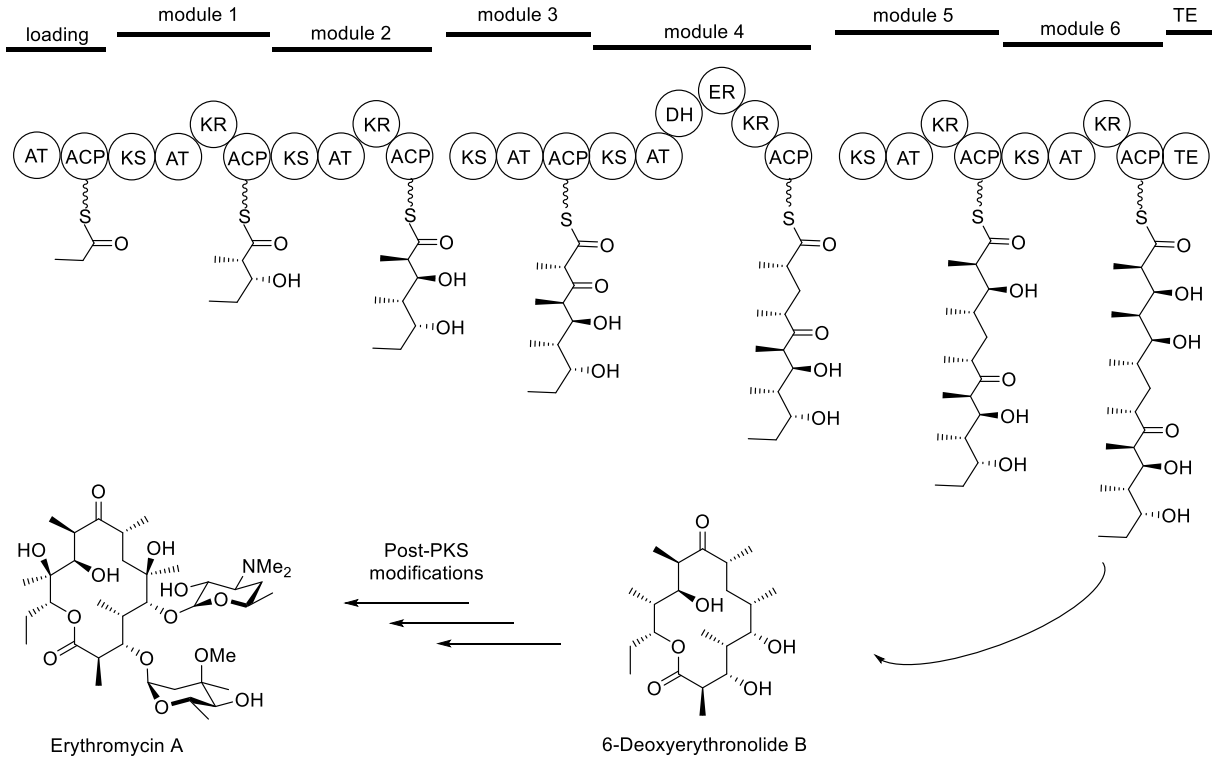


Figure 1.3.1.1 Erythromycin polyketide synthase, the canonical *cis*-AT modular PKS[20]. One starter unit of propionyl coenzyme A (CoA) and six extender units of (2S)-methylmalonyl-CoA are required for the biosynthesis of erythromycin A. The erythromycin PKS is comprised of a loading

module, six chain-extension modules and a chain-terminating thioesterase (TE). Each chain-elongation module contains an acyl transferase (AT), a ketosynthase (KS) and an acyl carrier protein (ACP) domain, and additional accessary domains, as required. Following TE-catalysed liberation of the intermediate, aglycone 6-deoxyerythronolide B, it is further processed by a set of post-PKS modifications (twice hydroxylation, twice glycosylation and once methylation on a sugar residue) to yield the biologically active erythromycin A. DH: dehydratase. ER: enoyl reductase. KR: ketoreductase.

In general, PKSs could be classified into various types based on the architecture and mode of action of the enzymatic assembly lines, while transition states between different classified systems exist[21]. Similar to the nomenclature of fatty acid synthases, type I PKSs are multifunctional enzymes that are linearly organized with fused catalytic domains, whereas type II refers to a dissociate complex of discrete and usually monofunctional enzymes. Type III indicates chalcone synthase type PKSs. In addition, PKSs could be also categorized as iterative or non-iterative according to the repeat use of KS domain during elongation[22]. Further, a special type I PKSs, designating as *trans*-AT PKS apart from the canonical *cis*-AT PKS, were found lacking AT domains within the modules (Figure 1.3.1.2), for which AT activity is provided at each elongation step by one or few free-standing proteins usually within the BGCs[23]. The aggregate genomic data suggests that *trans*-AT PKS represent 38 % of all bacterial modular PKSs, indicating that this type of polyketides constitute a major natural product class[24]. Despite the AT architecture, *trans*-AT PKS have numerous peculiarities including the presence of modules with unusual domain orders or unique domains, non-elongating modules, action of domains across modules, split modules that are located on two proteins, and functions provided *in trans* in addition to ATs. *Trans*-ATs can also function within *cis*-AT PKS pathways[23]. Thus, on one hand, make the *cis*-AT co-linearity rule not applicable. On the other hand, these features increase the structural diversification and complexity.

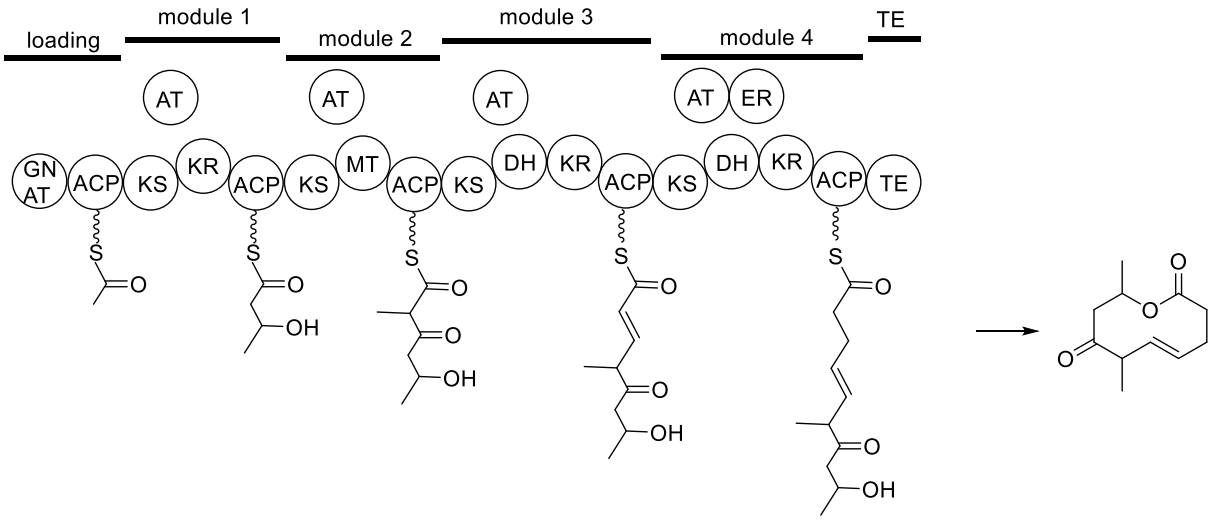


Figure 1.3.1.2 Biosynthesis of a fictional complex polyketide by a *trans*-AT PKS[23]. AT activity at each elongation step is provided by one or few free-standing proteins usually within the BGC. GNAT: acetyl-loading AT of the GCN5-related N-acetyl transferase superfamily. MT: methyltransferase.

Polyketides structural diversity could be attributed to a number of determinants throughout the chain assembly. Firstly, the building blocks incorporated into the polyketide chain have a wide variety of sources, such as the commonly used malonyl coenzyme A (CoA), their 2-substituted derivatives and even products from other PKSs, fatty acid synthases, terpene synthases or nonribosomal peptide synthases[25]. Then, the optional β-keto processing activities by KR, ER, DH domains define the stereochemistry and configuration, which largely contribute to the structurally diversity of polyketides. Thirdly, the products may vary by different chain releasing mechanisms, resulting in various structures such as linear polyketides or macrolide, *etc*. In addition, secondary cyclization and rearrangement may occur and lead to novel carba- and heterocycles. Finally, the structure may be decorated by a wide range of post-PKS tailoring enzymes[18,19,22].

1.3.2 Biosynthesis of non-ribosomal peptides

Nonribosomal peptides (NRPs) are a large group of structurally diversified peptides with variable sizes[26]. NRPs consist of head-to-tail cyclized peptides (e.g. argyrins), lipocyclopeptides (e.g. myxochromides) and linear peptides (e.g. cystobactamids) (Figure 1.3.2.1). Typically, NRPs are generated by modular multifunctional megaenzymes termed non-ribosomal peptide synthetases (NRPS). Similar to PKSs, modules in NRPS comprise

at least three distinct domains: adenylation (A) domains, peptidyl carrier protein (PCP) domains (also referred to as thiolation (T) domains), and condensation (C) domains. In the assembly line, each A domain is responsible for activating and transferring specific amino acids onto the phosphopantetheine (Ppant) prosthetic group of PCP within the same module and the C domain achieve the elongation of polypeptide chain via catalyzing the amide bond formation between the aminoacyl and the peptide-chain tethered on the PCP of the last module. Further structural modifications including epimerization (by E domains), formylation (by F domains), methylation (by M domains) and oxidation (by Ox domains) could optionally occur after each chain extension step. Finally, the mature oligopeptide is released from the assembly line by a thioesterase (TE), reductase or Dickmann cyclase domain[27], followed by post-NRPS modification and decoration (Figure 1.3.2.2).

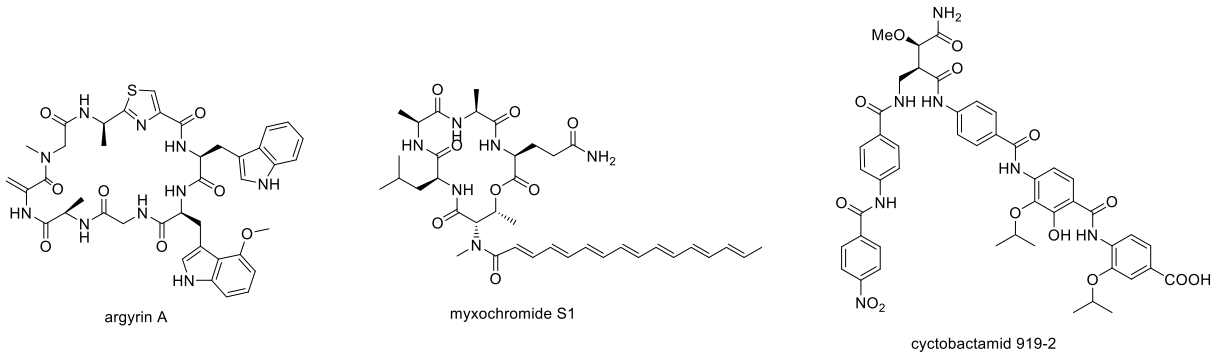


Figure 1.3.2.1 Chemical structures of myxobacterial NRPs representatives.

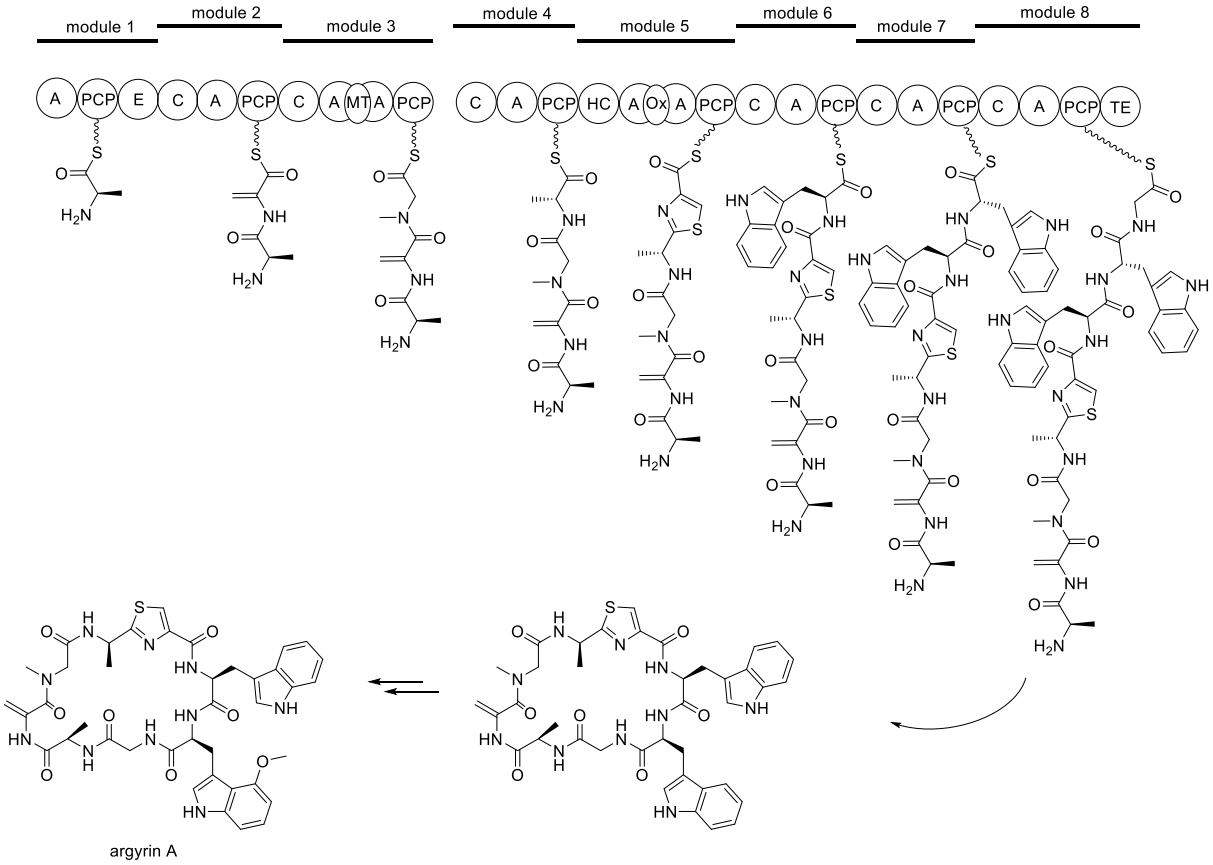


Figure 1.3.2.2 NRPS principle illustrated by the machinery of argyrin A[28]. The argyrin NRPS consists of eight modules, which activate and condense in order alanine, serine, glycine, alanine, cysteine, tryptophan, tryptophan and glycine. Dehydroalanine is proposed to be generated from thioester-bound seryl residue by C domain in module 3. The thiazole ring originates from cysteine via heterocyclization and subsequent oxidation. Once the intermediate is released from the assembly line, it is further processed by a tryptophan 2,3-dioxygenase and an O-methyltransferase to form argyrin A. A: adenylation domain. C: condensation domain. PCP: peptidyl carrier protein domain. E: epimerization domain. MT: methyltransferase domain. HC: heterocyclization domain. Ox: oxidation domain. TE: thioesterase domain.

Although NRPSs are commonly described as multimodular megaenzymes, there exist dissociated nonmodular NRPSs[29]. Therefore, NRPSs could be classified into different types based on their organizations. Type I NRPSs systems refer to the large modular NRPSs whereby each module incorporates one designated building block. Type II contains standalone or minimal domains and it usually encompasses specific tailoring enzymes that introduce modified amino acids into the products.

The structural diversity of NRPs basically derives from the building blocks, including not only 20 proteinogenic amino acids but also abundant nonproteinogenic amino acids such as

hydroxyl amino acids and methyl amino acids. Besides, the intermediates could be decorated by either optional domains incorporated in each modules such as E, M and Ox domains or auxiliary tailoring enzymes such as P450 monooxygenases. These additional modifications installed during and after NRPs assembly contribute to the structural diversity as well as biological activity.

1.3.3 PKS-NRPS hybrid

PKs and NRPs are two distinct classes of natural products and their biosynthesis systems show many similarities. For example, both systems consist of three core domains (KS, AT, ACP in PKS and C, A, PCP in NRPS) and both of them use carrier proteins to tether the growing chain via thioester bonds. A multitude of PKS-NRPS hybrid assembly lines, which use both PKS and NRPS system to build structurally complex polyketide-amino acid/peptide hybrid molecules, have been identified[30–33]. When an NRPS module is located in the upstream of a PKS module, the KS domain of the downstream PKS module receives the aminoacyl chain from the PCP domain of the upstream NRPS module via a catalytic cysteine residue and mediates C-C bond formation via decarboxylative condensation in the elongation step (Figure 1.3.3.1A). When a PKS module is located in the upstream of an NRPS module, the NRPS C domain accommodates the acyl group bound to the PKS ACP domain and catalyzes the amide bond-forming condensation reaction (Figure 1.3.3.1B).

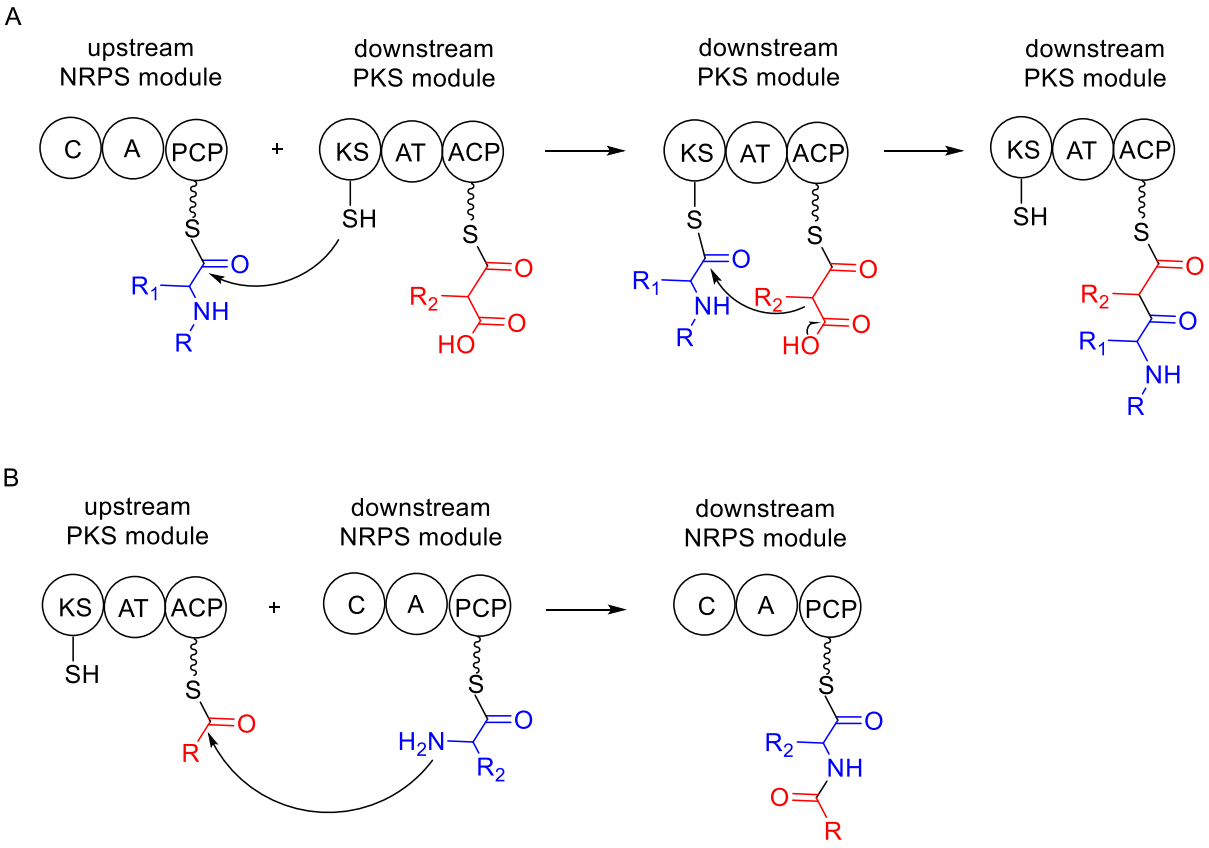


Figure 1.3.3.1 The condensation reactions catalyzed by the KS and C domains. A C-C bond formation by NRPS and PKS modules. **B** Amide bond formation by PKS and NRPS modules. A: adenylation. ACP: acyl carrier protein. AT: acyltransferase. C: condensation. KS: ketosynthase. PCP: peptidyl carrier protein.

In PKS-NRPS hybrid systems, the PKS and NRPS modules are present in the same polypeptide or as separate subunits. For example, the frontalamide biosynthetic system contains one iterative PKS module and an NRPS module, which are present in the same protein[34] (Figure 1.3.3.2A). In this case, an inter-modular linker is usually formed to connect PKS and NRPS modules. A notable example of separate hybrid PKS-NRPS modular type is the biosynthetic machinery of epothilone, which consists of nine PKS modules and one NRPS module. All the PKS modules are distributed over five proteins and the NRPS module is present as a separate protein[35,36] (Figure 1.3.3.2B). In this case, the interactions between the upstream ACP/PCP domains and the downstream C/KS domains are believed to be mediated by folded docking domains that are located at the termini of the polypeptides[37]. In some cases, standalone enzymes are employed. For instance, the A domain and the PCP domain are both standalone in the biosynthetic machinery of pyoluteorin (Figure 1.3.3.2C).

The KS domain of the PKS module directly accepts the modified amino acid building block from the dissociated PCP domain[38].

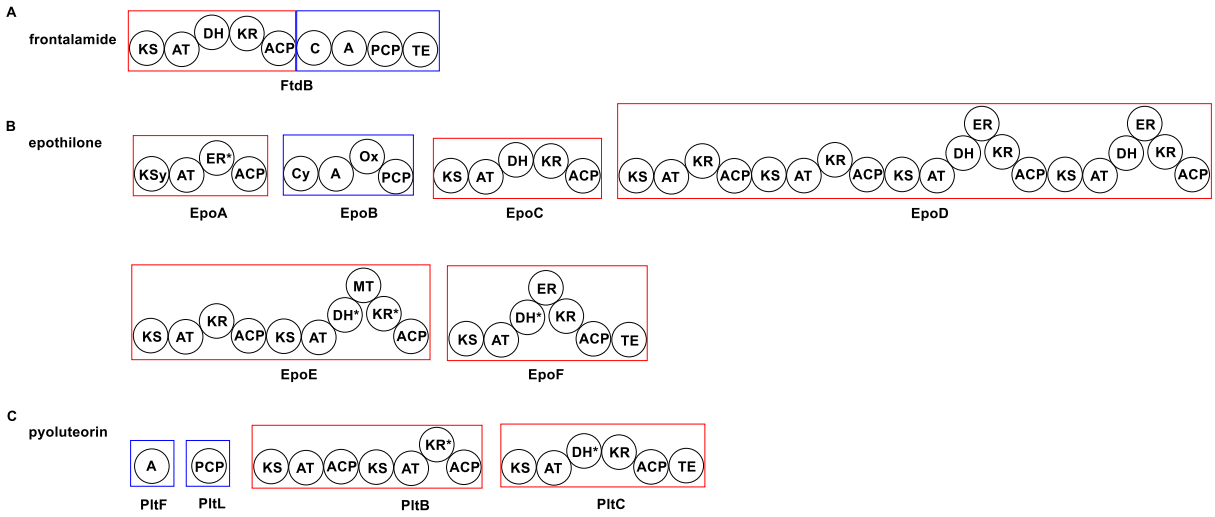


Figure 1.3.3.2 Domain organization of PKS-NRPS hybrids. A Tethered PKS-NRPS hybrid frontalamide. **B** Separate PKS-NRPS hybrid epothilone. **C** Standalone PKS-NRPS hybrid pyoluteorin. Inactive domains are denoted with an asterisk. Cy: cyclization. DH: dehydratase. ER: enoylreductase. KR: ketoreductase. KSy: ketosynthase-like decarboxylase. MT: methyltransferase. Ox: oxidase. TE: thioesterase.

Recent bioinformatic analysis showed that approximately one third of more than 3,300 identified gene clusters attribute to PKS-NRPS hybrids[29]. As exhibiting the chemical properties of both PKs and NRPs, PK-NRP hybrid products expand their biological activities. Moreover, the combination of these two assembly lines also prompts the capacity of microorganisms to produce structurally diversified compounds remarkably.

1.3.4 RiPPs

Ribosomally synthesized and post-translationally modified peptides (RiPPs) depict another major class of natural products despite all the other prevalent classes like terpenoids, alkaloids, polyketides, and non-ribosomal peptides. RiPPs are produced in all three domains of life and show a vast structural diversity and intriguing biological activities[32]. RiPPs also have a profound impact on human life, veterinary medicine and academia. For example, one of the most well-known RiPPs, nisin (Figure 1.3.4.1), has been used as food preservative

for more than 50 years[33]. Thiostrepton (Figure 1.3.4.1), not only shows antimicrobial properties, but also demonstrated antimalarial and anticancer activities[34].

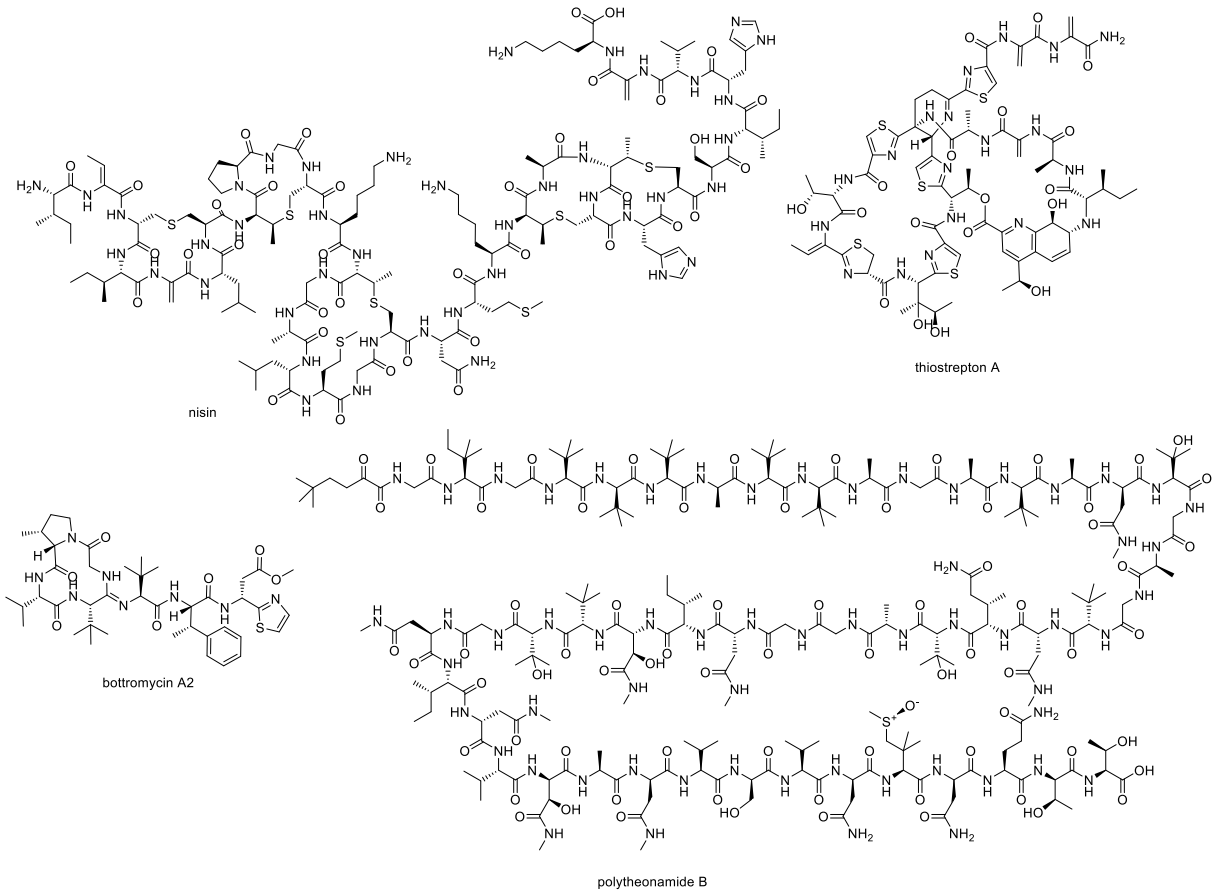


Figure 1.3.4.1 Examples of RiPPs.

Unlike NRPSs which usually utilize multimodular megaenzymes to incorporate amino acids into a peptide backbone, the biosynthesis of RiPPs is initiated as a ribosomally synthesized precursor peptide which comprises a leader peptide at the N-terminus and a core peptide at the C-terminus. The leader peptide is usually important for recognition by enzymes mediating post-translational modifications (PTM) and for export[35]. In contrast, the core peptide will be transformed into the final natural products. In some rare cases, such as bottromycins (Figure 1.3.4.1), the leader peptide is attached at the C-terminus of the core peptide and therefore termed as the follower peptide[36,37]. Peptides of eukaryotic origin such as cyclotides possess a signal sequence appended to the N-terminus of the leader peptide, which directs the peptide to the specific cell compartment where the peptide is subsequently modified. For some peptides, a recognition sequence is found at the C-terminus of the core

peptide, which is important for excision and cyclization[38,39]. The PTM could occur before or after the leader peptide is removed by proteolysis and the mature active compound transformed from core peptide is sometimes cyclized (Figure 1.3.4.2).

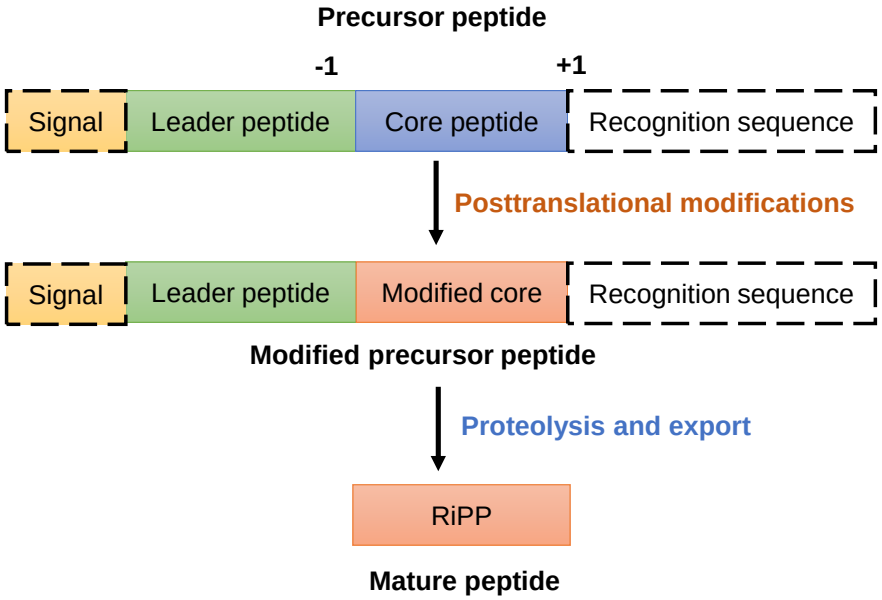


Figure 1.3.4.2 Schematic of RiPPs biosynthesis. The core region contained in the precursor peptide is transformed into the mature product. Many of the posttranslational modifications are guided by the leader peptide and the recognition sequences. Products originating from eukaryotes often contain a signal peptide at the N-terminus to direct the peptide to specialized compartments for modification and secretion.

There are more than 20 classes of RiPPs including lanthipeptides, thiopeptides, linear azol(in)e-containing peptides (LAPs) and microcins[40]. Each class is characterized by its PTMs which usually protect the products from proteolysis and confer proper biological activities on mature peptides. One common PTM is cyclization includes thioether cross-links and heterocyclization to form thiazolines and oxazolines[41]. However, cyclization is not always indispensable and linear RiPPs also exist, such as polytheonamides[42] (Figure 1.3.4.1), the most extensively modified RiPPs known to date. Despite cyclization which usually defines the type of RiPPs, there are some other tailoring modifications often occurring on amino acid side chains including hydroxylation, methylation, halogenation, prenylation and acylation. Finally, some tailoring modifications may occur within the backbone including epimerization as well as N- or C- terminal modifications. These

modifications endow individual amino acids with structure alteration and extend structural diversity of RiPPs, although the precursor peptides are limited to proteinogenic amino acids. Furthermore, many PTM enzymes show plasticity and are highly permissive to the mutations in core peptides, which facilitate the engineering of customized RiPPs[43].

1.4 Microbial natural products discovery

1.4.1 Bioactivity-based screening

The traditional bioactivity-guided screening of extracts from microorganisms involves acquisition of microorganisms, fermentation, products isolation and bioassay of fermentation broths and purified compounds (Figure 1.4.1). This strategy was prevailing during 1940s-1970s, the so-called "Golden Age of Antibiotics", and had been used to identify more than 1000 natural products with antimicrobial or antifungal activity[3]. The most notably examples include streptomycin, erythromycin and avermectin. However, this approach was more and more exhaustive in the recent decades because of the tendency to only rediscover known metabolites[44]. One reason is that many strains that have been investigated are, in most cases, related species, which may obtain the same BGCs during evolution. One strategy to overcome this problem is to extend the scope of biological resources to those that are rarely explored, such as samples from marine environment and human gut microbiota, or even new strain families and clades. Another pertinent reason, uncovered by the whole genome sequencing, is that many BGCs are cryptic or silent under conventional laboratory growth conditions.

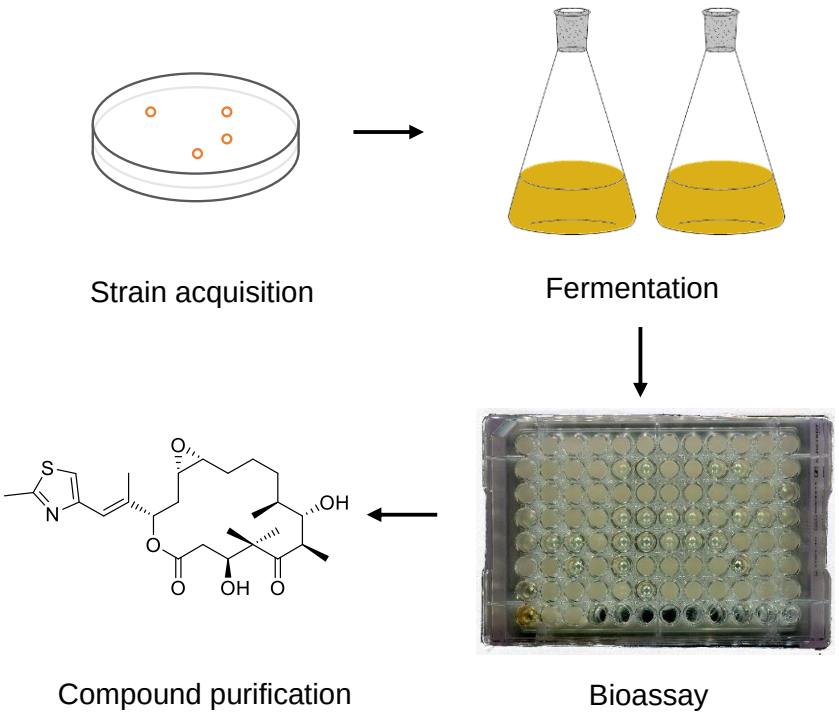


Figure 1.4.1 Schematic of bioactivity-guided screening which involves strain collection, fermentation, activity-guided fractionation of crude extracts and compound purification.

1.4.2 One strain-many compounds (OSMAC) and metabolomics-based discovery

An approach that focusses on a systematic alteration of the cultivation parameters in order to increase the number of secondary metabolites available from one microbial source is termed OSMAC (One Strain-Many Compounds)[61]. Genome mining described below usually includes alterations on genetic elements and therefore the need for the establishment of genetic manipulation. In contrast, the control and manipulation of cultivation parameters is much easier to achieve (for example, media composition, aeration, culture vessels, addition of chemical inducers or other additives). Sometimes, such manipulation may lead to a desired pleiotropic effect on production of secondary metabolites. One good example for such an effect is closthioamide, a polythioamide with various antimicrobial activities, whose production was induced by the addition of an aqueous soil extract to the fermentation medium of *Clostridium cellulolyticum*[62]. This approach provides an alternative way to explore nature's chemical diversity. However, as many organisms could only grow under limited conditions in the laboratory, the number of parameters that could be adjusted are limited to a large extent in such cases. In addition, it is almost impossible to predict what kind of influence such harsh changes in culture conditions will have on the gene expression

of BGCs. Thus, sensitive and convenient analytics to detect metabolic changes responding to the cultivation alteration are in need. To solve this problem, the OSMAC approach is usually combined with metabolomics analyses (Figure 1.4.2).

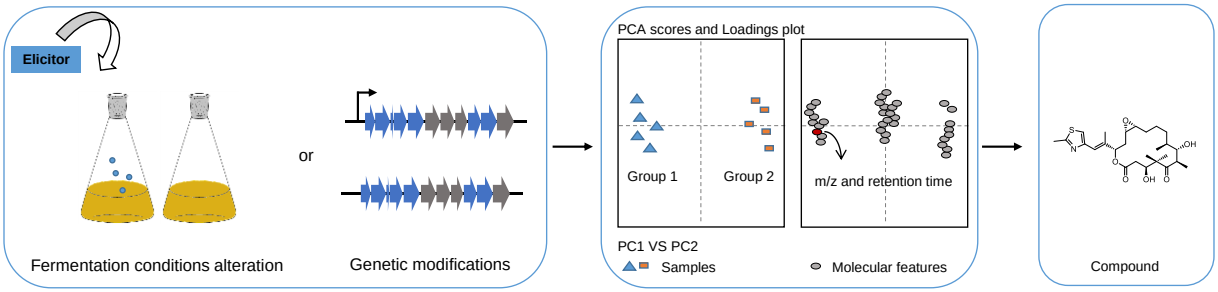


Figure 1.4.2 Schematic of discovering novel metabolites by metabolomics strategy coupled with OSMAC and genomic strategies. The crude extracts of strain growing under different conditions or with genetic modifications are subjected to principle component analysis (PCA). The molecular features triggered by the alteration of growth conditions or genetic elements are indicated by PCA, followed by compound isolation and purification.

Modern mass spectrometry techniques can provide an global overview of low molecular weight metabolites while multivariate statistical analyses including PCA (principal component analysis[63]) and GNPS (MS/MS database Global Natural Product Social molecular networking platform[64]) facilitate the metabolic comparison of different biological samples. Therefore, an enhanced overview of secondary metabolites profiling could be provided by metabolomics analyses which facilitates both, a rapid dereplication of known metabolites and prioritization of strains. In addition, metabolomics sensitively reflects the response of microorganisms to growth condition's alterations and genetic modifications, which makes the metabolomics approach a powerful tool to discover secondary metabolites while coupled to the strategies described above[65]. For instance, the comprehensive metabolite profiling analysis coupled to statistical data evaluation by PCA enabled the discovery of myxoprincomide, which was simply missed in the complex chromatogram[66].

1.4.3 Genomic driven discovery

At the beginning of the new millennium, complete genome sequences of two well-studied natural products producers, *Streptomyces coelicolor* and *Streptomyces avermitilis*, unveiled

a great unexplored biosynthetic potential in these strains[45,46]. This observation could also be done in other natural products producing microorganisms with large genomes[47–49] and stimulated the birth of genome mining: predicting and isolating a natural product based on genetic information prior to structure elucidation. This approach usually uses genes either encoding core biosynthetic enzymes, or acting as resistance or regulators to mine and identify BGCs. This is possible as the biosynthetic principles and biosynthetic machineries of many secondary metabolite classes are often highly conserved and the related biosynthetic genes are usually clustered[50]. Correspondingly, plenty of bioinformatics tools were developed to facilitate *in silico* analysis of the genetic data and connecting them to secondary metabolites. One notable example is the antiSMASH platform[51], which has been updated to the fifth version recently. However, as the mining is based on the known biosynthetic types and families, some novel secondary metabolites originating from alternative biosynthetic machineries may be missed. Besides, connecting new metabolites to their respective BGCs is not yet easy to achieve, even with the help of available bioinformatics tools.

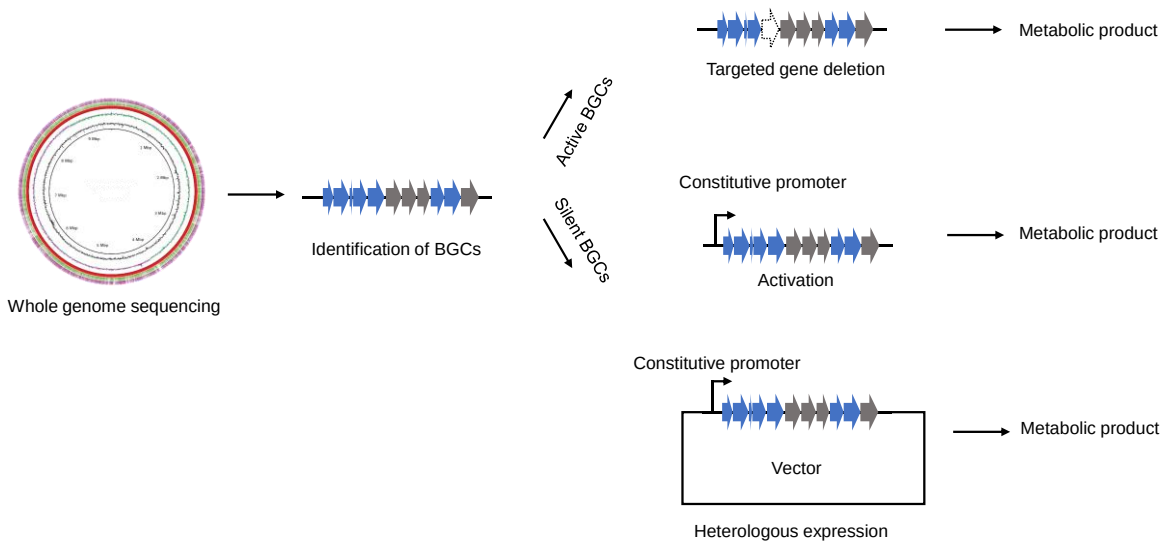


Figure 1.4.3 Schematic of genomic-based discovery of natural products. The biosynthetic gene clusters (BGCs) are predicted by bioinformatics tools. When the BGCs are active and the strain is amenable to genetic manipulation, the corresponding metabolites could be indicated by metabolic comparison between the targeted gene inactivation mutant and the wild type strains. When the BGCs are silent, strategies including promoter replacement and heterologous expression could be used to activate the BGCs.

For strains that are amenable to genetic manipulation, the core biosynthetic genes could be directly inactivated and the respective secondary metabolite could be correlated using metabolic comparisons of wild type and mutant strains when the BGC is actively expressed (Figure 1.4.3). Nevertheless, this strategy is not able to connect all the gene clusters to their corresponding compounds. For example, until recently, less than half of the gene clusters identified by antiSMASH could be connected to the respective metabolites in the most intensively studied myxobacterium *Myxococcus xanthus* DK1622[8]. One explanation is that the majority of the gene clusters are silent or low expressed under the conventional laboratory conditions, which make their products undetectable due to a too low concentration. Hence, multiple approaches are developed to activate these silent gene clusters, including replacing the native promoter within the BGC with constitutive or readily inducible promoters, inhibiting the negative regulators, enhancing positive regulators and overexpressing resistance genes[52] (Figure 1.4.3). In some cases, when one biosynthetic pathway is blocked as a consequence of genetic manipulation, the products of other pathways which may compete for the same precursors may pop up[53].

In other cases like slow growing bacteria, producers only exhibiting a low level of resistance towards a bioactive secondary metabolite or bacteria which are not genetically amenable, heterologous expression of putative BGCs becomes a pivotal tool to discover, to activate, to engineer and to characterize secondary metabolites[54]. The identification and isolation of secondary metabolites may be simplified when the BGC is expressed in a "chassis host" owning a limited natural metabolic profile (Figure 1.4.3). Furthermore, recent development of cloning technologies such as RecET direct cloning, Redαβ recombineering[55], transformation-associated recombination (TAR)[56,57] and CRISPR/Cas9[58–60] facilitate cloning of large gene clusters, as well as refactoring and reprogramming such gene clusters. However, many factors may lead to a failure of heterologous expression including the lack of specific precursors in the chassis host, missing standalone genes in the cloned BGCs or more general effects like non-compatible codon preference.

1.5 Outline of the thesis

Cystobacter sp. SBCb004 is the producer of multiple known compounds including stigmatellin, tubulysin, argyrin and geosmin. Although its genome harbors dozens of BGCs without any assigned secondary metabolites, strain SBCb004 was not included in a priority

list before. In this thesis, a genomic strategy coupled with a metabolomic approach was used to explore the hidden biosynthetic potential of *Cystobacter* sp. SBCb004.

In Chapter one, the discovery of novel ajudazol derivatives is described. Interrogation of the SBCb004 genome uncovered one BGC which is highly similar to the known cluster encoding essential functions necessary for the biosynthesis of ajudazol. However, neither ajudazol A nor B was detected in the metabolic profiling, either indicating a silent BGC or the presence of an unknown BGC. The targeted gene inactivation coupled with PCA helped to correlate a set of novel secondary metabolites with the BGC, followed by compound isolation and structure elucidation. These secondary metabolites turned out to be novel ajudazol derivatives, referred to as ajudazols C-J. These novel ajudazols were proposed to incorporate 3,3-dimethylacrylyl-CoA as a starter unit. Ajudazols C to J are characterized by both, varying degrees of hydroxylation and different glycosylation patterns. A biosynthetic pathway was proposed and targeted gene inactivation led to the identification of candidate enzymes as being involved in these post-assembly line modifications. Two of these candidate enzymes are not encoded within the BGC, showing the diversity and complexity of BGCs in myxobacteria. All the new ajudazol derivatives exhibit different levels of cytotoxicity.

In Chapter two, the isolation and characterization of two novel secondary metabolites, cystopipecotides A and B is described. The analyses of metabolic profiling of SBCb004 extracts helped the identification of two novel secondary metabolites which were shown to harbour non-proteinogenic amino acids. These two novel metabolites, later named cystopipecotides A and B were isolated and identified. Cystopipecotides are hexapeptides containing the non-proteinogenic amino acid 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) at the N-terminus and a γ-lactam at the C-terminus. A retro-biosynthetic analysis enabled the identification of the corresponding BGC, a NRPS-PKS hybrid, which was confirmed by gene inactivation. *In silico* analysis of the BGC revealed an operon which comprises homologs of transketolases and some lysine biosynthetic enzymes involved in the α-aminoadipate pathway and is attributed to the synthesis of HHMPA. The biosynthetic pathway of HHMPA was proposed and reconstructed *in vitro*.

Taken together, a well-balanced combination of both genomic and metabolomic methods unraveled two families of novel secondary metabolites hidden in the biosynthetic potential of the SBCb004 genome. Work summarized in this thesis demonstrates in practice that the

biosynthetic potential of myxobacteria is indeed largely untapped, even in heavily explored strains like SBCb004.

1.6 Reference

(1) Cragg, G. M.; Newman, D. J. Natural products: a continuing source of novel drug leads. Biochim. Biophys. Acta **2013**, 1830 (6), 3670–3695. DOI: 10.1016/j.bbagen.2013.02.008.

(2) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83* (3), 770–803. DOI: 10.1021/acs.jnatprod.9b01285.

(3) Katz, L.; Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43* (2-3), 155–176. DOI: 10.1007/s10295-015-1723-5.

(4) Ventola, C. L. The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics* **2015**, *40* (4), 277–283.

(5) Stewart, E. J. Growing Unculturable Bacteria. *J. Bacteriol.* **2012**, *194* (16), 4151–4160. DOI: 10.1128/JB.00345-12.

(6) Dawid, W. Biology and global distribution of myxobacteria in soils. *FEMS Microbiol. Rev.* **2000**, *24* (4), 403–427.

(7) Whitworth, D., Ed. *Myxobacteria: Multicellularity and differentiation*; ASM Press, 2007.

(8) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480. DOI: 10.1016/j.biotechadv.2019.107480.

(9) Kunze, B.; Reichenbach, H.; Müller, R.; Höfle, G. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (myxobacteria). *J. Antibiot.* **2005**, *58* (4), 244–251.

(10) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34* (2), 135–160. DOI: 10.1039/C6NP00106H.

(11) Schäberle, T. F.; Lohr, F.; Schmitz, A.; König, G. M. Antibiotics from myxobacteria. *Nat. Prod. Rep.* **2014**, *31* (7), 953–972. DOI: 10.1039/c4np00011k.

(12) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9* (21), 4898–4908. DOI: 10.1039/C8SC01325J.

(13) Raju, R.; Mohr, K. I.; Bernecker, S.; Herrmann, J.; Müller, R. Cystodienoic acid: A new diterpene isolated from the myxobacterium *Cystobacter sp. J. Antibiot.* **2015**, *68*, 473–475. DOI: 10.1038/ja.2015.8.

(14) Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of cittilins, unusual ribosomally synthesized and post-translationally modified peptides from Myxococcus xanthus. *ACS Chem. Biol.* **2020**, *accepted.* DOI: 10.1021/acschembio.0c00430.

(15) Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Kohnke, J.; Trauner, D.; Müller, R. Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.* **2017** (56), 1–5. DOI: 10.1002/anie.201612640.

(16) Weissman, K. J.; Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, *27* (9), 1276–1295. DOI: 10.1039/c001260m.

(17) Wenzel, S. C.; Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **2009**, *26* (11), 1385–1407. DOI: 10.1039/b817073h.

(18) Pang, B.; Wang, M.; Liu, W. Cyclization of polyketides and non-ribosomal peptides on and off their assembly lines. *Nat. Prod. Rep.* **2016**, *33* (2), 162–173. DOI: 10.1039/C5NP00095E.

(19) Olano, C.; Mendez, C.; Salas, J. A. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Nat. Prod. Rep.* **2010**, *DOI: 10.1039/b911956f.*

(20) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **2002**, *19* (5), 542–580. DOI: 10.1039/b103920m.

(21) Müller, R. Don't classify polyketide synthases. *Chem. Biol.* **2004**, *11* (1), 4–6. DOI: 10.1016/j.chembiol.2004.01.005.

(22) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (26), 4688–4716. DOI: 10.1002/anie.200806121.

(23) Helfrich, E. J. N.; Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **2016**, *33* (2), 231–316. DOI: 10.1039/c5np00125k.

(24) O'Brien, R. V.; Davis, R. W.; Khosla, C.; Hillenmeyer, M. E. Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiot.* **2014**, *67* (1), 89–97. DOI: 10.1038/ja.2013.125.

(25) Khosla, C.; Gokhale, R. S.; Jacobsen, J. R.; Cane, D. E. Tolerance and specificity of polyketide synthases. *Annu. Rev. Biochem.* **1999**, *68*, 219–253. DOI: 10.1146/annurev.biochem.68.1.219.

(26) Süssmuth, R. D.; Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **2017**, *56* (14), 3770–3821. DOI: 10.1002/anie.201609079.

(27) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. USA* **2014**, *111* (25), 9259–9264. DOI: 10.1073/pnas.1401734111.

(28) Perlova, O.; Fu, J.; Kuhlmann, S.; Krug, D.; Stewart, F.; Zhang, Y.; Müller, R. Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **2006**, *72* (12), 7485–7494. DOI: 10.1128/AEM.01503-06.

(29) Sandmann, A.; Sasse, F.; Müller, R. Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity. *Chem. Biol.* **2004**, *11* (8), 1071–1079. DOI: 10.1016/j.chembiol.2004.05.014.

(30) Buntin, K.; Rachid, S.; Scharfe, M.; Blöcker, H.; Weissman, K. J.; Müller, R. Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (24), 4595–4599. DOI: 10.1002/anie.200705569.

(31) Liu, F.; Garneau, S.; Walsh, C. T. Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chem. Biol.* **2004**, *11* (11), 1533–1542. DOI: 10.1016/j.chembiol.2004.08.017.

(32) Velásquez, J. E.; van der Donk, Wilfred A. Genome mining for ribosomally synthesized natural products. *Current opinion in chemical biology* **2011**, *15* (1), 11–21. DOI: 10.1016/j.cbpa.2010.10.027.

(33) Lubelski, J.; Rink, R.; Khusainov, R.; Moll, G. N.; Kuipers, O. P. Biosynthesis, immunity, regulation, mode of action and engineering of the model lantibiotic nisin. *Cell. Mol. Life Sci.* **2008**, *65* (3), 455–476. DOI: 10.1007/s00018-007-7171-2.

(34) Bagley, M. C.; Dale, J. W.; Merritt, E. A.; Xiong, X. Thiopeptide antibiotics. *Chem Rev* **2005**, *105* (2), 685–714. DOI: 10.1021/cr0300441.

(35) Oman, T. J.; van der Donk,W. A. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat. Chem. Biol.* **2010**, *6* (1), 9–18. DOI: 10.1038/nCHeMBIo.286.

(36) Huo, L.; Rachid, S.; Stadler, M.; Wenzel, S. C.; Müller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **2012**, *19* (10), 1278–1287. DOI: 10.1016/j.chembiol.2012.08.013.

(37) William J. K. Crone; Finian J. Leeper; Andrew W. Truman. Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides. *Chem. Sci.* **2012**, *3* (12), 3516–3521. DOI: 10.1039/C2SC21190D.

(38) Hallen, H. E.; Luo, H.; Scott-Craig, J. S.; Walton, J. D. Gene family encoding the major toxins of lethal Amanita mushrooms. *Proc. Natl. Acad. Sci. USA* **2007**, *104* (48), 19097–19101.

(39) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (20), 7315–7320. DOI: 10.1073/pnas.0501424102.

(40) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D. D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link,

A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. E.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108–160. DOI: 10.1039/C2NP20085F.

(41) Truman, A. W. Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. *Beilstein J. Org. Chem.* **2016**, *12*, 1250–1268. DOI: 10.3762/bjoc.12.120.

(42) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H.-G.; Matsunaga, S.; Piel, J. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* **2012**, *338* (6105), 387–390. DOI: 10.1126/science.1226121.

(43) Funk, M. A.; van der Donk, Wilfred A. Ribosomal Natural Products, Tailored To Fit. *Accounts of chemical research* **2017**, *50* (7), 1577–1586. DOI: 10.1021/acs.accounts.7b00175.

(44) Tulp, M.; Bohlin, L. Rediscovery of known natural compounds: nuisance or goldmine? *Bioorg Med Chem* **2005**, *13* (17), 5274–5282. DOI: 10.1016/j.bmc.2005.05.067.

(45) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat. Biotechnol.* **2003**, *21* (5), 526–531. DOI: 10.1038/nbt820.

(46) Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417* (6885), 141–147.

(47) Bachmann, B. O.; van Lanen, S. G.; Baltz, R. H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **2014**, *41* (2), 175–184. DOI: 10.1007/s10295-013-1389-9.

(48) Wenzel, S. C.; Muller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Molecular bioSystems* **2009**, *5* (6), 567–574. DOI: 10.1039/b901287g.

(49) Keller, N. P.; Turner, G.; Bennett, J. W. Fungal secondary metabolism - from biochemistry to genomics. *Nature reviews. Microbiology* **2005**, *3* (12), 937–947. DOI: 10.1038/nrmicro1286.

(50) Ziemert, N.; Alanjary, M.; Weber, T. The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016**, *33* (8), 988–1005. DOI: 10.1039/c6np00025h.

(51) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019** (47), W81-W87. DOI: 10.1093/nar/gkz310.

(52) Rutledge, P. J.; Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015**, *13* (8), 509–523. DOI: 10.1038/nrmicro3496.

(53) Culp, E. J.; Yim, G.; Waglechner, N.; Wang, W.; Pawlowski, A. C.; Wright, G. D. Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nature biotechnology* **2019**, *80*, 1. DOI: 10.1038/s41587-019-0241-9.

(54) Huo, L.; Hug, J. J.; Fu, C.; Bian, X.; Zhang, Y.; Müller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, *36*, 1412-1436. DOI: 10.1039/C8NP00091C.

(55) Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Muller, R.; Stewart, A. F.; Fu, J.; Zhang, Y. RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **2016**, *11* (7), 1175–1190. DOI: 10.1038/nprot.2016.054.

(56) Zhang, J. J.; Yamanaka, K.; Tang, X.; Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods in enzymology* **2019**, *621*, 87–110. DOI: 10.1016/bs.mie.2019.02.026.

(57) Kouprina, N.; Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **2016**, *125* (4), 621–632. DOI: 10.1007/s00412-016-0588-3.

(58) Liu, Y.; Tao, W.; Wen, S.; Li, Z.; Yang, A.; Deng, Z.; Sun, Y. In Vitro CRISPR/Cas9 System for Efficient Targeted DNA Editing. *mBio* **2015**, *6* (6), e01714-15. DOI: 10.1128/mBio.01714-15.

(59) Jiang, W.; Zhao, X.; Gabrieli, T.; Lou, C.; Ebenstein, Y.; Zhu, T. F. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nature communications* **2015**, *6*, 8101. DOI: 10.1038/ncomms9101.

(60) Lee, N. C. O.; Larionov, V.; Kouprina, N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic acids research* **2015**, *43* (8), e55. DOI: 10.1093/nar/gkv112.

(61) Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **2002**, *3* (7), 619–627.

(62) Lincke, T.; Behnken, S.; Ishida, K.; Roth, M.; Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem. Int. Ed.* **2010**, *49* (11), 2011–2013. DOI: 10.1002/anie.200906114.

(63) Ringner, M. What is principal component analysis? *Nat. Biotechnol.* **2008**, *26* (3), 303–304.

(64) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.;

Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828–837. DOI: 10.1038/nbt.3597.

(65) Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, *31* (6), 768–783. DOI: 10.1039/c3np70127a.

(66) Cortina, N. S.; Krug, D.; Plaza, A.; Revermann, O.; Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (3), 811–816. DOI: 10.1002/anie.201106305.

Chapter 1

Discovery and characterization of novel ajudazol derivatives

Hu Zeng†, Joy Birkelbach†, Judith Hoffmann, Alexander Popoff, Carsten Volz and Rolf Müller

†These authors contributed equally to this work

Contributions to the presented work

Author's contribution

The author significantly contributed to the conception of the study, designed and performed experiments, evaluated and interpreted the resulting data. The laboratory and *in silico* work regarding gene cluster analysis, isolation of ajudazols C-J, as well as the mutagenesis and feeding experiments were performed by the author. The biosynthetic pathway and post-assembly line modification network of ajudazols were proposed by the author. Furthermore, the author contributed significantly to conceiving and writing this chapter.

Contribution by co-workers

Joy Birkelbach contributed by elucidating the structures of ajudazols C, D, E, H and I, as well as by conceiving, writing and interpreting all the NMR data. Judith Hoffmann performed the structure elucidation of ajudazols F and J. Alexander Popoff performed the structure elucidation of ajudazol G. Carsten Volz contributed to conception and supervision of this study, contributed to conceiving, writing and editing of this chapter. Rolf Müller was responsible for the conception of the project and contributed to writing and proofreading of this chapte

2.1 Abstract

In the past decades, myxobacteria have proven to be a rich source of natural products. However, their biosynthetic potential still seems to be underexplored given the high number of biosynthetic gene clusters present in myxobacterial genomes. In this study we identified a truncated ajudazol biosynthetic gene cluster in *Cystobacter* sp. SBCb004. In a combined approach using mutagenesis and metabolomics analyses, a set of novel ajudazols (named ajudazol C-J) were identified and subsequently isolated. Their structures were elucidated using comprehensive HR-MS and NMR spectroscopy. These novel ajudazols were proposed to incorporate 3,3-dimethylacrylyl-CoA as a starter unit. Ajudazols C to J are characterized by both, varying degrees of hydroxylation, desaturation and different glycosylation patterns. Two P450 dependent enzymes and one glycosyl transferase were speculated to be responsible for the hydroxylation at C_8, desaturation at C_{15}-C_{33} and the transfer of a hexose, respectively, based on mutagenesis results. One of the cytochrome P450 dependent enzymes and the glycosyl transferase are encoded by genes located elsewhere in the genome and therefore were shown not to be part of the biosynthetic gene cluster. Ajudazols C-H exhibit good to intermediate levels of cytotoxicity against various cancer cell lines.

2.2 Introduction

Myxobacteria are Gram-negative δ-proteobacteria ubiquitously occurring on earth[1]. They depict an extraordinary order of bacteria due to their ability to swarm, their social behavior upon starvation and their ability to prey on other microorganisms[2]. In the last decades, myxobacteria have emerged as a rich reservoir of secondary metabolites which exhibit antimicrobial (e.g. argyrin), antifungal (e.g. myxalamid) and antitumor (e.g. epothilone) activities[3]. Amongst them, many own unique structural features as well as novel mode of actions[4]. In addition, myxobacteria harbour the largest genomes among prokaryotes, known so far[5]. These genomes possess numerous regions encoding biosynthetic gene clusters (BGCs), and the corresponding secondary metabolites of most of them are still unknown. This observation indicates a large unexplored biosynthetic potential in the order of the myxobacteria.

Many of the known secondary metabolites from myxobacteria are synthesized by type I polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs). Both types are multi-modular megaenzymes which are responsible for the biosynthesis of a vast number of

highly diverse natural products. This diversity may arise from the incorporation of unusual building blocks, the comprehensive combination of these two types of assembly lines and numerous possible modifications of the core structure introduced by specialized enzymes[6,7].

Cystobacter sp. SBCb004 possesses a large genome which comprises 49 biosynthetic gene clusters predicted by antiSMAH 5.0[8] and is able to produce multiple bioactive compounds including tubulysins, stigmatellins and argyrins[9–11]. However, no natural product could be assigned yet to 43 of the predicted clusters. In this study, we identified an ajudazol-like BGC in SBCb004 which has two modules missing compared to the BGC of the known ajudazols A and B, and subsequently identified a set of novel ajudazol derivatives, which were proposed to incorporate a different starter unit, 3,3-dimethylacrylyl-CoA (DMA-CoA). By contrast, the biosynthetic machinery of the known ajudazols A and B condenses one acetyl-CoA and one malonyl-CoA followed with *O*-methylation to form the 3-methoxy-2-butenyl-S-ACP. The diversity of these novel ajudazol derivatives was shown to result from varying degrees of hydroxylation and different glycosylation patterns. Finally, a cytotoxic activity of the new derivatives C, D, E, F, G, and H could be shown.

2.3 Materials and methods

2.3.1 Bacterial strains, culture conditions and reagents

Escherichia coli DH10B was used for cloning and was cultivated at 37 °C in LB medium (1 % trypton, 0.5 % yeast extract, 0.5 % NaCl). *Cystobacter* sp. SBCb004 was cultivated at 30 °C and 180 rpm in CYH medium (0.3 % casitone, 0.3 % yeast extract, 0.2 % $CaCl_2 \cdot 2H_2O$, 0.2 % soy meal, 0.2 % glucose, 0.8 % soluble starch, 0.1 % $MgSO_4 \cdot 7H_2O$, 1.19 % HEPES, 8 mg/l Fe-EDTA, adjusted to pH 7.3 with 10 N KOH) for statistically analysis. M medium (1 % phytone, 1 % maltose, 0.1 % $CaCl_2 \cdot 2H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/l Fe-EDTA, 1.19 % HEPES, adjusted to pH 7.2 with 10 N KOH) was used for transformation and feeding. VY2M medium (0.5 % Baker's Yeast, 0.05 % $CaCl_2 \cdot 2H_2O$, 0.2 % maltose, 0.119 % HEPES, 0.5 mg/l Vitamin B12, adjusted to pH 7.0 with 10 N KOH) was used for large scale fermentation and compound isolation. 50 µg/ml Kanamycin was added to the culture when necessary.

Restriction endonucleases, Taq DNA polymerase, Phusion DNA polymerase, FastAP, and T4 DNA ligase was purchased from Thermo scientific. NEBuilder HiFi DNA assembly Master Mix was purchased from New England Biolabs. Oligonucleotide primers were

synthesized by Sigma-Aldrich and DNA sequencing of PCR products was performed by LGC Biosearch Technologies.

2.3.2 Gene inactivation in SBCb004

Gene inactivation in SBCb004 was accomplished by homologous recombination of a suitable plasmid backbone in the encoding region. At this, internal DNA fragments of the target genes served as homologous regions to mediate the insertion of the constructed plasmids into the target gene via homologous recombination. Internal DNA regions of the target gene with a size of 800 bp to 1200 bp were amplified from SBCb004 genomic DNA by using corresponding primers 9101-sc-gene-F and 9101-sc-gene-R (Table S1). Subsequently, these DNA fragments were cloned into plasmid pbluekan kindly provided by Dr. Carsten Volz (unpublished). pbluekan encodes a aminoglycoside phosphotransferase (*aph*(3')-II) conferring kanamycin resistance[12] as well as a β-lactam resistance cassette (*bla*) and a *lacZ* gene under the control of the T7A1 promoter[13] (Figure S1). Stop codons were introduced at both ends of the cloned homologous regions upon PCR amplification by the use of mutated oligonucleotide primer. The constructed plasmids (Table S2) were confirmed by Sanger sequencing using primer seq-pblueCN-R. Correct plasmids were introduced into SBCb004 by electroporation following a previously published protocol[9]. The correct transformants were confirmed by PCR reactions with primer CN-check-R which binds the plasmid backbone and primer gene_CK_F (where "gene" stands for the designation of the target gene) which binds the upstream of the target gene.

2.3.3 UPLC-MS measurement

All the measurements were performed on a Dionex UltiMate 3000 RSLC system equipped with Waters BEH C18 column (50 × 2.1 mm, 1.7 µm). The flow rate was 600 µl/min and the column temperature was 45 °C. The gradient was as follows: 0-0.5 min, 95 % water with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 0.5-18.5 min, 5-95 % B; 18.5-20.5 min, 95% B; 20.5-21 min, 95-5% B; 21-22.5 min, 5% B. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the Bruker Daltonics maXis 4G hr-qToF mass spectrometer using the Apollo II ESI source. Mass spectra were acquired in centroid mode in the range from 150-2500 m/z at a 2 Hz full scan rate. Full scan and MS spectra were acquired at 2 Hz.

2.3.4 Principal Component Analysis

SBCb004 wild type strain and inactivation mutants were cultivated in 50 ml CYH medium supplemented with 2% Amberlite XAD-16 resin. The cultivation were performed in triplicates. A CYH medium blank was treated and extracted in parallel as a negative control. After 10 days, the cells and the Amberlite XAD-16 resin were harvested by centrifugation and extracted with 40 ml acetone. The solvent was dried by rotary evaporation and the extract was resolved in MeOH followed by measurement by UPLC-MS. The MS data was processed by the molecular feature finder implemented in Bruker Compass Data Analysis 4.2 with the compound detection parameters SN threshold 1; Correlation coefficient 0.9; minimum compound length 10 spectra and smoothing width of 3 spectra. Bucketing was generated using Bruker Compass Profile Analysis 2.1 with advanced bucketing and window parameters of 30s and 15 ppm. The PCA t-Test function was then used to distinguish the features derived from each group. The features not derived from the medium and disappearing in a mutant were proposed to be related to the inactivated biosynthetic gene cluster.

2.3.5 Compound isolation

SBCb004 was grown in 20 L VY2M medium supplemented with 2 % Amberlite XAD-16 resin at 30 °C for 10 days. Cells and XAD-16 resin were harvested by centrifugation and extracted with 3 × 1.5 L acetone. The extract was partitioned between hexane and methanol (3:1, v/v) and the methanol extract was subsequently partitioned between ethyl acetate and deionized water (3:1, v/v). The ethyl acetate extract was fractionated on a Sephadex LH-20 column (GE healthcare) eluted with methanol to afford two fractions.

Fraction containing ajudazol C, D, E, G and H was separated by semi-preparative HPLC (Waters Xbridge Peptide BEH C18, 5 µm, 250 × 10 mm; 5ml/min; UV 220 nm) using a gradient as follows: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 58 % B; 8-31 min, 58 % to 75 % B; 31-33 min, 75 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B. 2.04 mg ajudazol C (t_R=23.2 min) and 11.91 mg ajudazol D (t_R=25.4 min) were obtained. Impure fraction collected at t_R=13.7 min was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5ml/min; UV 220 nm) using the following gradient: 0-5 min, 5 % H_2O with

0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 43 % B; 8-31 min, 43 % to 44 % B; 31-33 min, 44 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B. 1.21 mg ajudazol H (t_R=22.0 min) was obtained. Impure fraction collected from C18 column at t_R=16.8 min was further purified on a Waters XSelect CSH Fluro-Phenyl HPLC column (10 × 250 mm, 5 µm; 5ml/min; UV 220 nm) using the following gradient: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 45 % B; 8-31 min, 45 % to 45.8 % B; 31-33 min, 45.8 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B. 0.4 mg ajudazol G (t_R=21.0 min) was obtained. Impure fraction collected from C18 column at t_R=17.2 min was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5ml/min; UV 220 nm) using the following gradient: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 50 % B; 8-31 min, 50 % to 50.4 % B; 31-33 min, 50.4 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B. 0.7 mg ajudazol E (t_R=23.6 min) was obtained.

Fraction containing ajudazol F, I and J was separated by semi-preparative HPLC (Waters Xbridge Peptide BEH C18, 5 µm, 250 × 10 mm; 5ml/min; UV 220 nm). The following gradient was used: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 42 % B; 8-31 min, 42 % to 52 % B; 31-33 min, 52 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B. Ajudazol I containing fraction was further purified on a Phenomenex Synergi™ Fusion-RP (10 × 250 mm, 4 µm; 5ml/min; UV 220 nm) and eluted with the following gradient: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 40 % B; 8-31 min, 40 % to 40.4 % B; 31-33 min, 40.4 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B, affording ajudazol I (0.8 mg, t_R=23.0 min). Fraction contains ajudazol J was further purified using a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5ml/min; UV 220 nm) and eluted with the following gradient: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 40 % B; 8-31 min, 40 % to 40.7 % B; 31-33 min, 46.4 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B, affording ajudazol J (0.9 mg, t_R=20.9 min). Ajudazol F containing fraction was further purified on a Phenomenex Synergi™ Fusion-RP (10 × 250 mm, 4 µm; 5ml/min; UV 220 nm) and eluted with the following gradient: 0-5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5-8 min, 5 % B to 45 % B; 8-31 min, 45 % to 46.4 % B; 31-33 min, 46.4 % to 95 % B; 33-34 min, 95 % to 5 % B; 34-36 min, 5 % B, affording ajudazol F (2.4 mg, t_R=20.4 min).

2.3.6 Structure elucidation

UV/Vis data were obtained using the DAD of the Dionex Ultimate 3000 SL system linked to a Bruker maXis 4G UHRqTOF for mass spectrometric analysis.

NMR data were recorded on an UltraShield 500 MHz (^{1}H at 500 MHz, ^{13}C at 125 MHz) or an AVANCE III 700 MHz NMR (^{1}H at 700 MHz, ^{13}C at 175 MHz) equipped with a 5 mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) are calculated in ppm and coupling constants (J) in Hz. Spectra were recorded in methanol-d_4, acetone-d_6 or chloroform-d and adjusted to the solvent signals (methanol-d_4 δ_H 3.31, δ_C 49.2, acetone-d_6 δ_H 2.05, δ_C 29.9 and chloroform-d δ_H 7.27, δ_C 77.0). Measurements were conducted in 5 mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA). For the two dimensional experiments HMBC, HSQC, ROESY and gCOSY standard pulse programs were used. HMBC experiments were optimized for $^{2,3}J$C-H = 6 Hz and HSQC experiments for ^{1}JC-H = 145 Hz.

Ajudazol C (2): colorless solid; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S4-6; HR-ESIMS m/z 577.3270 $[M+H]^+$ (calc. for $C_{34}H_{45}N_2O_6$ 577.3272).

Ajudazol D (3): colorless solid; M = 574.30; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S7; HR-ESIMS m/z 575.3128 $[M+H]^+$ (calc. for $C_{34}H_{45}N_2O_6$ 577.3116).

Ajudazol E (4): colorless solid; M = 590.30; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S8; HR-ESIMS m/z 591.3077 $[M+H]^+$ (calc. for $C_{34}H_{42}N_2O_7$ 591.3065).

Ajudazol F (5): colorless solid; M = 736.36; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S9; HR-ESIMS m/z 737.3648 $[M+H]^+$ (calc. for $C_{40}H_{52}N_2O_{11}$ 737.3644).

Ajudazol G (6): colorless solid; M = 592.31; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S10; HR-ESIMS m/z 593.3228 $[M+H]^+$ (calc. for $C_{34}H_{44}N_2O_7$ 593.3221).

Ajudazol H (7): colorless solid; M = 608.31; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S11; HR-ESIMS m/z 609.3171 $[M+H]^+$ (calc. for $C_{34}H_{44}N_2O_8$ 609.3170).

Ajudazol I (8): colorless solid; M = 770.36; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S12; HR-ESIMS m/z 771.3699 $[M+H]^+$ (calc. for $C_{40}H_{54}N_2O_{13}$ 771.3699).

Ajudazol J (9): colorless solid; M = 740.35; UV spectra see Figure S5; ^{1}H and ^{13}C NMR data see Table S13; HR-ESIMS m/z 741.3593 $[M+H]^+$ (calc. for $C_{39}H_{52}N_2O_{12}$ 741.3593).

2.3.7 Feeding experiment

The feeding studies of SBCb004 were performed in 50 ml M medium. 5.3 mg 3-Methyl-2-butenoic acid-*d6* were resolved in 1 ml water and sterilized by filtration. Equal aliquots (250 µl) were added to the culture after 24 h, 36 h, 48 h and 60 h to a final concentration of 1 mM. A control experiment without isotopically labelled precursor was done in parallel. 2% Amberlite XAD-16 resin was added after 72 h. The culture was harvested after 90 h by centrifugation. Cells including XAD were extracted with 50 ml acetone. The solvents were evaporated and the extract was resolved in MeOH and analyzed by UPLC-MS.

2.3.8 Antimicrobial activity test

All the tested pathogens were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). For micro dilution assays, overnight cultures of each strain were diluted in the growth medium to approximately 10^6 cfu/ml and added to sterile 96-well plates. The cell suspension was treated with the respective compounds in serial dilution. After incubation on a microplate shaker (750 rpm, 37 °C) for 16 h, the growth inhibition was assessed by visual inspection.

3.3.9 Cytotoxicity assay

Cell lines were obtained from the DSMZ and were cultured under conditions as recommended by the depositor. 6×10^3 cells were seeded in 180 µl complete medium in each well of the 96-well plates and treated with compounds in serial dilution after 2 h of equilibration. Both compounds and the internal solvent control was tested in duplicates. After 5 d of incubation, 20 µl of 5 mg/mL MTT (Thiazolyl blue tetrazolium bromide) in PBS was added per well and it was continued to be at 37 °C for 2 h. The medium was then discarded and cells were washed with 100 µl PBS before adding 100 µl 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader and the percentage relative to the respective methanol control was used to express cell viability. IC_{50} values were determined by sigmoidal curve fitting.

2.3.10 Sequence of ajudazol biosynthetic gene cluster in *Cystobacter* sp. SBCb004

The major sequence of the ajudazol BGC in SBCb004 was provided by 454 technology[14,15]. The gaps in genes *ajuA*, *ajuC*, *ajuE*, *ajuG1* and *ajuL* and the gap between *orf31* and *orf32* are refilled by PCR and Sanger sequencing with the primers 9104-gene-gap-F and 9104-gene-gap- R in Table S1 ("gene" refers to the gene with a gap). The complete sequence will be deposited in GenBank when the manuscript is submitted.

2.4 Results and discussion

2.4.1 Genomic-driven discovery of novel ajudazol derivatives

In the screening for known biosynthetic gene clusters of the SBCb004 genome *in silico*, we identified a PKS-NRPS encoding region with high similarity to the known BGC of ajudazol in *Chondromyces crocatus* Cm c5[16] (Figures 2.4.1.1 A). However, gene *ajuK* and *ajuL* which together encode for the first four modules in the biosynthesis of ajudazols A and B are missing in this region. Only the homologue of *ajuL* was found 20 kb upstream of the region (Figure 2.4.1.1 A). Furthermore, none of the two known ajudazols A or B could be detected in any of the extracts obtained from SBCb004. Due to this discrepancy we decided to elucidate this further and therefore inactivated the genes which are homologous to *ajuD* and *ajuL* (Table S3) via insertional mutagenesis. A Principal Component Analysis (PCA) of the extracts obtained from the mutants SBCb004 ajuD$^-$, SBCb004 ajuL$^-$ and the wild type strain revealed a set of compounds with an m/z ranging from 575 to 771 disappearing in both strains SBCb004 ajuD$^-$ and SBCb004 ajuL$^-$ (Figure 2.4.1.1 B and Figure S2). None of the observed compounds owning these m/z values could be identified in either, the Dictionary of Natural Products[17] or our in-house database Mxbase[18]. As these compounds were not produced in both of the mutant strains, they were considered to be novel ajudazol derivatives and were named ajudazols C-J (Figure 2.4.1.2). These compounds were subsequently isolated and purified. The interpretation of the structure elucidation will be shown in Joy Birkelbach's thesis and the NMR tables and spectra are attached in the supplementary information.

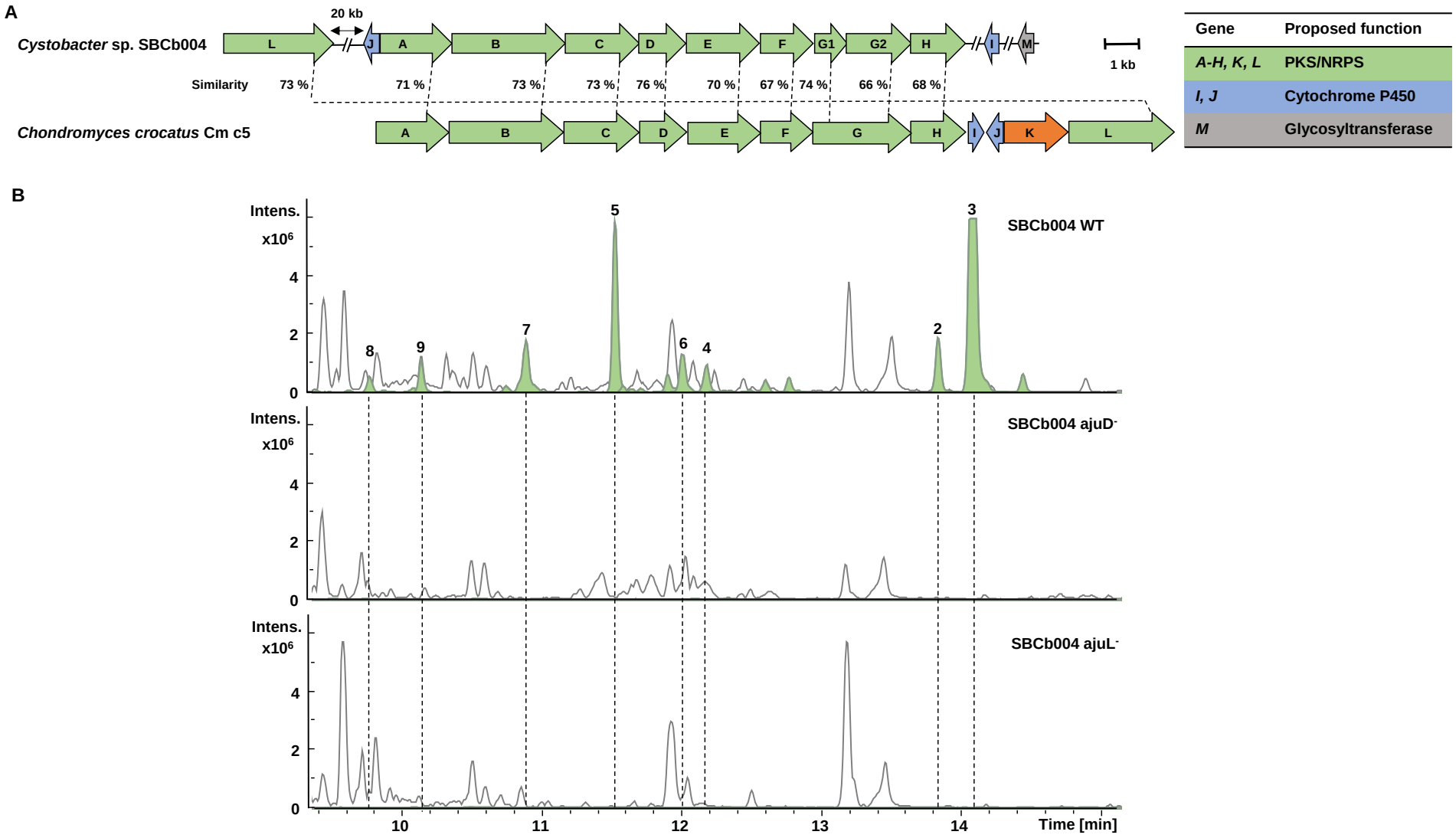
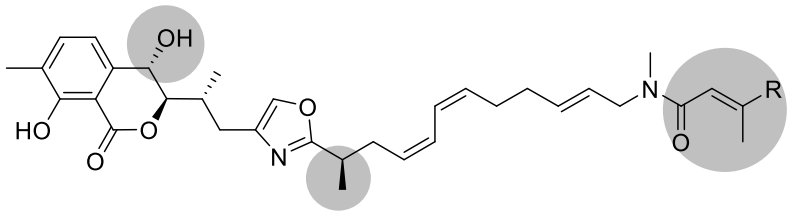
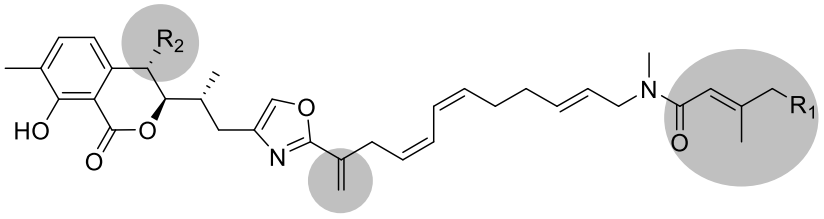

A
20 kb
Cystobacter sp. SBCb004
L
J
A
B
C
D
E
F
G1
G2
H
I
M
1 kb
Similarity
73 %
71 %
73 %
73 %
76 %
70 %
67 %
74 %
66 %
68 %
Chondromyces crocatus Cm c5
A
B
C
D
E
F
G
H
I
J
K
L
Gene
Proposed function
A-H, K, L
PKS/NRPS
I, J
Cytochrome P450
M
Glycosyltransferase
B
Intens.
x10⁶
SBCb004 WT
SBCb004 ajuD⁻
SBCb004 ajuL⁻
Time [min]

Figure 2.4.1.1 A Comparison of ajudazol biosynthetic gene cluster (BGC) in *Cystobacter* sp. SBCb004 with ajudazol BGC in *Chondromyces crocatus* Cm c5. Gene *ajuK* which has no homologue in the BGC of SBCb004 is shown in orange. Gene *ajuL* in SBCb004 is located 20 kb upstream of gene *ajuJ*. Gene *ajuI* and *ajuM* in SBCb004 are located somewhere else in the genome. **B** Comparative HPLC-MS analysis of SBCb004 ajuD$^-$, SBCb004 ajuL$^-$, and SBCb004 wild type (WT) extracts. **2**-**9**: ajudazols C-J.



ajudazol B (**1**)
$R=OCH_3$

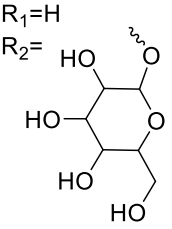
ajudazol C (**2**)
$R=CH_3$

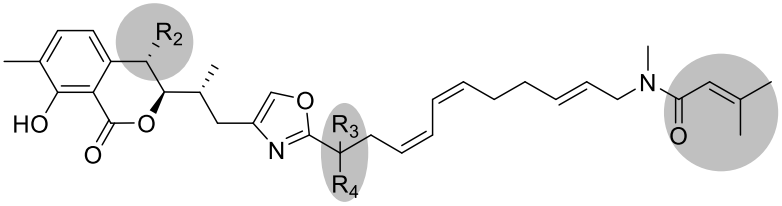


ajudazol D (**3**)
R_1=H
R_2=OH

ajudazol E (**4**)
R_1=OH
R_2=OH

ajudazol F (**5**)
R_1=H
R_2=

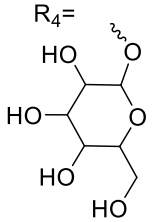




ajudazol G (**6**)
R_2=OH
R_3=OH
R_4=CH_3

ajudazol H (**7**)
R_2=OH
R_3=OH
R_4=CH_2OH

ajudazol I (**8**)
R_2=OH
R_3=OH
R_4=



ajudazol J (**9**)
R_2=OH
R_3=OH
R_4=

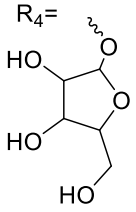


Figure 2.4.1.2 Chemical structures of ajudazols B-J (1-9). Structural differences are highlighted in grey. *relative configuration of *β*-glucopyranose.

To sum up, comparing the new ajudazol derivatives to ajudazols A and B, all derivatives show identical structural features like the isochromanone core, the oxazole and the E, Z, Z-dodeca-2,6,8-trien-1-yl moiety. However, new derivatives feature a 3-methylbutenamide instead of the 3-methoxy-butenamide moiety in ajudazols A and B. Furthermore, ajudazols A and B only show hydroxylation at C_8 whereas new derivatives feature hydroxylation at additional positions. Ajudazol E is hydroxylated at C_{29} and ajudazols G, H, I and J are at C_{15}. In addition, ajudazol H is hydroxylated at C_{33} as well. Moreover, ajudazols F, I and J are *O*-glycosylated with a hexose or a pentose at C_8 or C_{33}, respectively (Figure 2.4.1.2).

2.4.2 Insights into the biosynthesis of the novel ajudazol derivatives

2.4.2.1 The starter unit of ajudazols C-J

As shown in Figure 2.4.1.1 A, all genes encoding the respective PKS and NRPS homologues necessary for the biosynthesis of ajudazol are present in SBCb004, except gene *ajuK*. In the biosynthetic machinery of ajudazols A and B in *Chondromyces crocatus* Cm c5, *ajuK* encodes for both, the loading module and the first extension module which are proposed to condense one acetyl-CoA and one malonyl-CoA to form a diketide intermediate, followed with enolization and *O*-methylation occurs in the distal carbonyl group, generating 3-methoxy-butenyl-ACP[16] (Figure 2.4.2.1.1 A). Retro-biosynthesis indicated that ajudazols C-J may derive from 3,3-dimethylacrylyl CoA (DMA-CoA), which could originate from the catabolism of leucine, as well as the 3-methylglutaconyl-CoA decarboxylase catalyzed biosynthesis pathway of isovaleryl CoA in myxobacteria[22]. In order to check if DMA-CoA could be utilized by the ajudazol assembly line, *d6*-3,3-dimethylacrylic acid was fed to SBCb004 and the isotopic peak of ajudazol D which has the highest intensity in chromatogram was presented. As shown in Figure S2, the isotopic peak m/z 581 of ajudazol D (m/z 575) was enriched, indicating the involvement of DMA-CoA. Due to the fact that *ajuK* is not present in SBCb004, we proposed that DMA-CoA is used as a starter unit of the NRPS-PKS assembly line instead of the commonly incorporated starter units malonyl-CoA and acetyl-CoA. Such an alternative starter unit has also been found previously in the

biosynthesis of myxothiazol and aurafuron, where isovaleryl-CoA was shown to be used as the precursor[23,24].

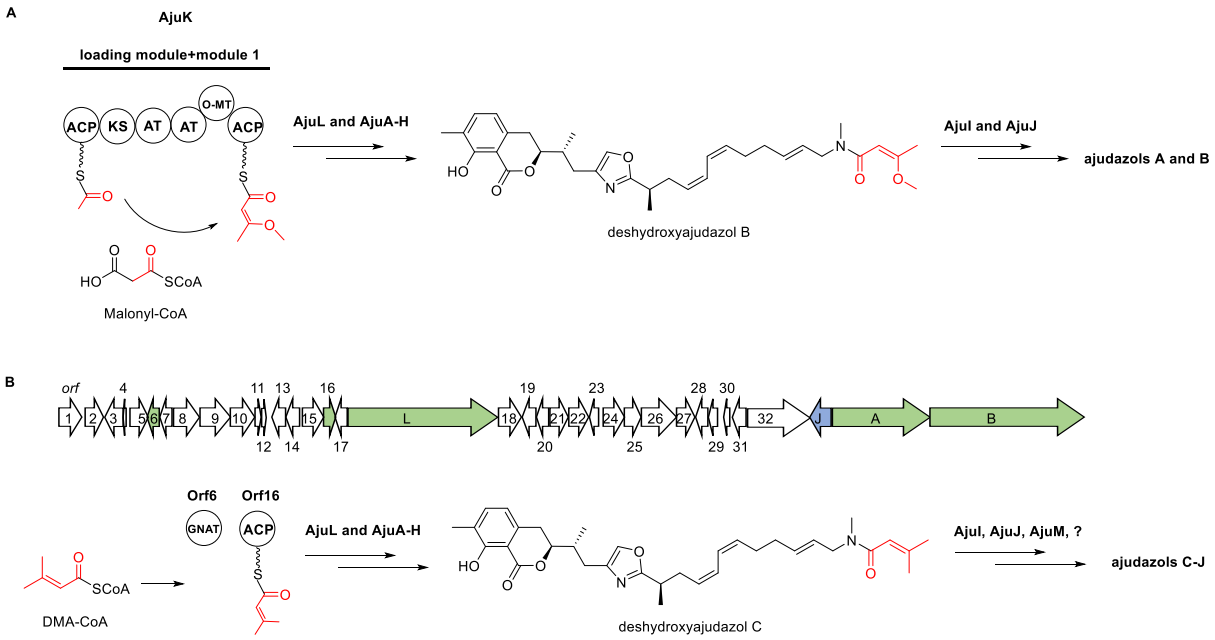


Figure 2.4.2.1.1 The initiation of ajudazol biosynthesis in *Chondromyces crocatus* Cm c5 (ajudazols A and B) and *Cystobacter* sp. SBCb004 (ajudazols C-J). A Proposed generation of 3-methoxy-butenoyl moiety in ajudazols A and B accomplished by AjuK in *Chondromyces crocatus* Cm c5. **B** The putative open reading frames (ORF) adjacent to *ajuL* in *Cystobacter* sp. SBCb004 and the proposed generation of 3-methyl-butenoyl moiety in ajudazols C-J. GNCT: GCN5-related N-acetyltransferase. ACP: acyl carrier protein. KS: ketosynthase. AT: acyltransferase. *O*-MT: *O*-methyltransferase.

To activate and load DMA-CoA to the assembly line, an AT domain and an ACP domain are needed. Our *in silico* analysis of the BGC in SBCb004 revealed that upstream of *ajuL*, *orf6* might encode a GCN5-related N-acetyltransferase. In addition, *aju16*, putatively encoding a 4'-phosphopantetheinyl transferase superfamily protein was found. However, the coding regions of these two genes are too short to perform a successful inactivation via insertion of a vector backbone as at least an internal homology fragment of approximately 1 kb is needed to be cloned into the available plasmid pbluekan to allow for homologous recombination. As no genetic tools for the construction of in-frame deletion mutants are available for this strain, we had to rely on single cross over based mutagenesis. We hypothesize that the activation and the loading of DMA-CoA to the assembly line might

either be accomplished by ORF6 and ORF16 or by alternative standalone AT-domain and ACP-domain containing proteins encoded elsewhere in the genome (Figure 2.4.2.1.1 B). The precursor would then be subjected to 12 rounds of PKS/NRPS chain extension performed by AjuL and AjuA-H generating deshydroxyajudazol C (Figure 2.4.2.1.2), which is subsequently converted into ajudazols C-J in multiple enzymatic reactions (Figure 2.4.2.2).

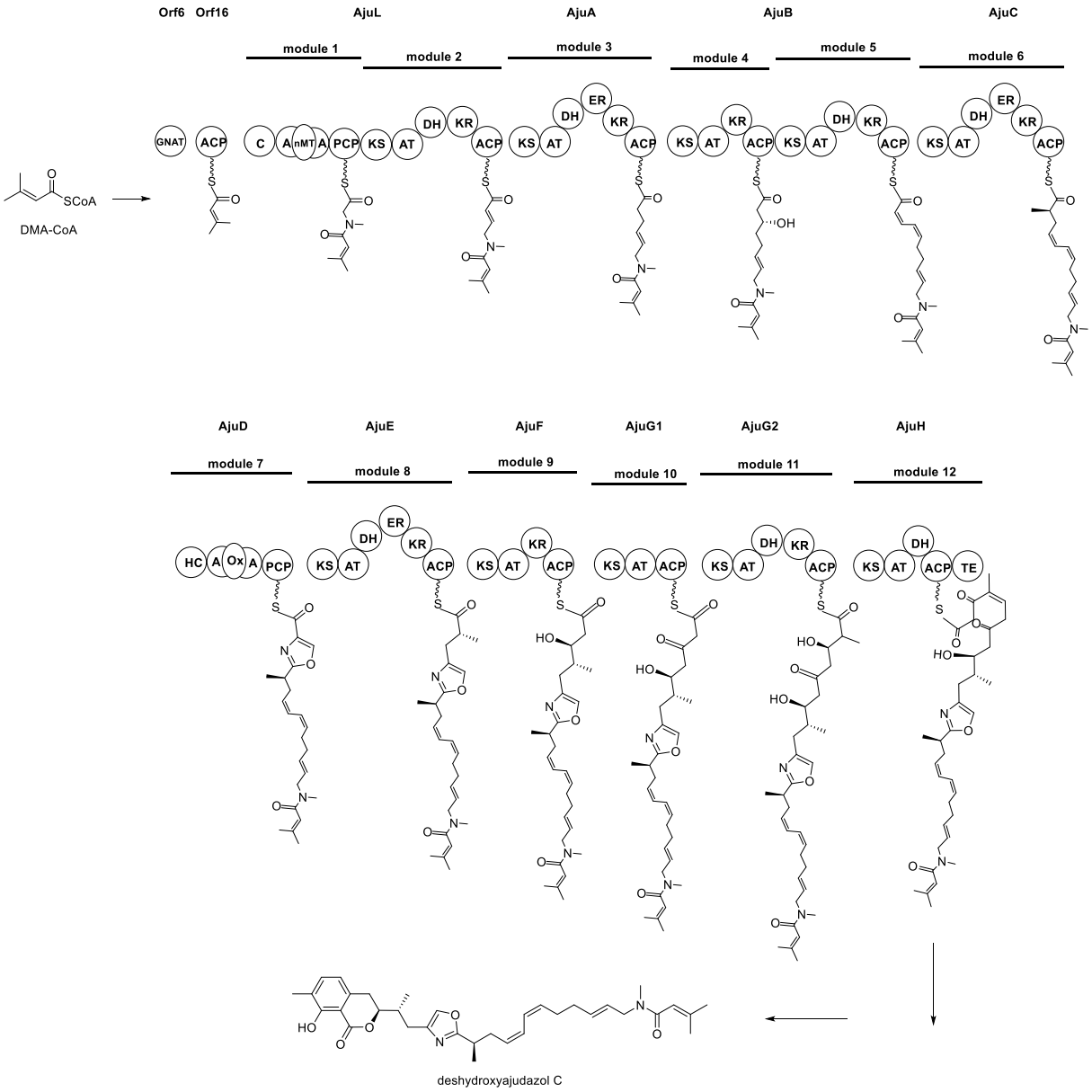


Figure 2.4.2.1.2 Model for the biosynthesis of deshydroxyajudazol C under the action of the mixed PKS–NRPS enzymes AjuL and AjuA-H. GNCT: GCN5-related N-acetyltransferase. ACP: acyl carrier protein. KS: ketosynthase. AT: acyltransferase. DH: dehydratase. ER: enoyl reductase. KR: ketoreductase. TE: thioesterase. C: condensation. A: adenylation. PCP: peptidyl carrier peptide. nMT: *N*-methyltransferase. HC: heterocyclization. Ox: oxidation.

2.4.2.2 The hydroxylation and desaturation enzymes in the biosynthesis of ajudazols C-J

In *Chondromyces crocatus* Cm c5, two cytochrome P450 dependent enzymes, AjuI and AjuJ, are described catalyzing the desaturation at C_{15}-C_{33} and hydroxylation at C_8, respectively[16]. However, only one gene homologous to *ajuJ* could be found in the ajudazol gene cluster in strain SBCb004 (Figure 2.4.1.1 A). The inactivation of gene *ajuJ* in SBCb004 results in the accumulation of the proposed deshydroxyajudazol C (m/z 561) and deshydroxyajudazol D (m/z 559). In addition, the formation of all the other derivatives seems to be abolished (Figure S4), suggesting that AjuJ is responsible for the hydroxylation at C_8 and the desaturation at C_{15}-C_{33} is performed independently of this enzyme. In order to identify candidate enzymes responsible for the desaturation between C_{15} and C_{33}, the amino acid sequence of AjuI from *Chondromyces crocatus* Cm c5 was used as a query to screen the available genome of SBCb004. Consequently, two coding regions, ctg27_313 (80 %, amino acids similarity) and ctg19_169 (60 %, amino acids similarity) were found (Table S3). In the inactivation mutant of *ctg27_313*, from now on referred to as *ajuI* (strain SBCb004 ajuI$^-$), only ajudazol C was synthesized, supporting the involvement of AjuI in the desaturation at C_{15}-C_{33} (Figure S4 and Figure 2.4.2.2). The production of ajudazols G, H, I and J was also blocked in the *ajuI* mutant (Figure S4). Consequently, the presence of a double bond at C_{15}-C_{33} seems to be a prerequisite for the hydroxylations at C_{15} and C_{33} to occur. However, the involved enzymes remain elusive to date. Deshydroxyajudazol C and deshydroxyajudazol D are only present in trace amounts in the extracts of SBCb004 wild type strain (Figure S4), indicating that the reactions catalyzed by AjuI and AjuJ are quite efficient. To locate the enzymes catalyzing the hydroxylations at C_{15} and C_{29}, all the predicted oxidoreductases within the BGC, *orf15*, *orf28* and *orf32* were inactivated (Figure 2.4.2.1.1 B and Table S3). Against all the expectations, the production of ajudazols C-J were not influenced in any of the mutants, suggesting the involvement of additional enzymes encoded elsewhere in the genome or some redundancy in functionality of single enzymes as they may complement each other.

2.4.2.3 The glycosylation in the biosynthesis of ajudazols C-J

Glycosylation is one of the most common modifications of natural products after the core structure has been synthesized. Such glycosylation may have profound effects on activities,

selectivity and pharmacokinetic properties of the mature products. This kind of modification has been observed in the biosynthesis of numerous myxobacterial compounds including sorangicins[25], disciformycins[26] and cystomanamides[27]. However, to our knowledge, only the glycosyltransferase in sorangicins, SorF, was characterized from myxobacteria while those involved in other compounds are not encoded in the clusters and therefore not identified yet. Flanking the core biosynthetic genes *ajuA-H* and *ajuL*, only one putative glycosyltransferase encoded by *orf2* could be identified (Figure 2.4.2.1.1 B and Table S3). However, a respective inactivation mutant constructed during our investigations does not impair the biosynthesis of ajudazol derivatives, implying that the involved proteins are not encoded within this region or may be complemented by other enzymes encoded in the genome. Considering the fact that SorF is able to utilize hexose as a natural donor substrate *in vivo* and has a promiscuous donor substrate specificity *in vitro*[25], it is reasonable to conclude that the glycosyl groups in ajudazols F, I and J may be transferred by similar enzymes. Thus, SorF was used as a query to identify such candidate enzymes in the available genome of SBCb004. Three hits were obtained in our search, namely *ctg1_130* (41 %, amino acids similarity), *ctg10_386* (44 %, amino acids similarity) and *ctg19_179* (50 %, amino acids similarity) (Table S3). In an inactivation mutant of *ctg19_179* (now referred to as *ajuM*; SBCb004 ajuM$^-$), the production of ajudazols F and I were abolished while the production of ajudazol J was not influenced (Figure S4). We reasoned that AjuM is able to perform the transfer of hexose to *O*-8 in ajudazol D and *O*-33 in ajudazol H, yielding ajudazol F and ajudazol I, respectively. However, the currently unknown pentose transferase is not encoded within the BGC and remains elusive.

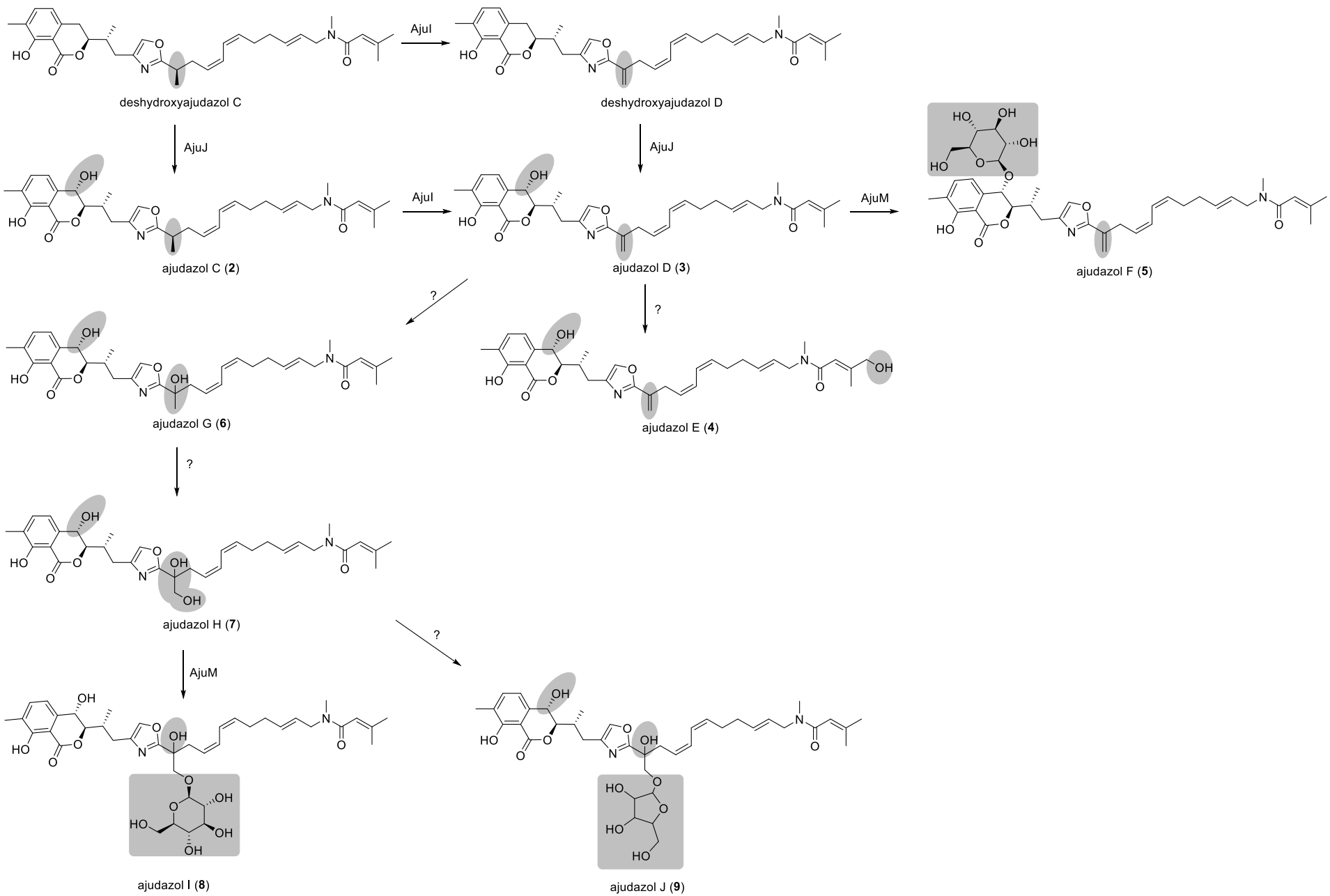

deshydroxyajudazol C
AjuI
deshydroxyajudazol D
AjuJ
AjuJ
AjuI
AjuM
ajudazol C (2)
ajudazol D (3)
ajudazol F (5)
?
?
ajudazol G (6)
ajudazol E (4)
?
ajudazol H (7)
?
AjuM
ajudazol I (8)
ajudazol J (9)

Figure 2.4.2.2 Proposed modification steps to generate ajudazols C-J in SBCb004. Modifications in each step of the biosynthesis are highlighted in grey.

Based on the results, the following proposal for the biosynthesis of ajudazols C-J is plausible. In SBCb004, DMA-CoA is presumably used as the starter unit. Thus, the enzymatic functions encoded by *orf6* and *orf16* are proposed to be involved (Figure 2.4.2.1.1 B). The subsequent steps of the biosynthesis are accomplished by NRPS-PKS extensions equal to the described ajudazol biosynthesis pathway in *Chondromyces crocatus* Cm c5 (AjuL and AjuA-H) (Figure 2.4.2.1.2). The resulting product deshydroxyajudazol C is either hydroxylated at C_8 to yield ajudazol C or desaturated at C_{15}-C_{33} to yield deshydroxyajudazol D. The latter product is subjected to further hydroxylation at C_8 to form ajudazol D. Ajudazol C could also be transformed into ajudazol D after an initial desaturation while the double bond between C_{15} and C_{33} in ajudazol D could be hydroxylated, yielding ajudazol G and ajudazol H, successively. Ajudazol D could be modified to yield ajudazol E with a hydroxylation at C_{29}. The hydroxylated products ajudazols D and H would subsequently be glycosylated with a hexose or pentose, yielding ajudazols F, I and J (Figure 2.4.2.2). In theory, ajudazols C, E and G could also be glycosylated and ajudazol D could accept a pentose. Their putative products could be observed in the chromatograms but only in very low amounts, which could be explained by the specificity and selectivity of the enzymes responsible for the subsequent modifications. Considering the structural similarity of AjuL substrates utilized in strains Cm c5 and SBCb004, the ajudazol BGCs in SBCb004 and Cm c5 are hypothesized to be evolutionary closely related. It is either that gene *ajuK* got lost in SBCb004 during evolution and the production of ajudazols C-J are enabled by a certain substrate plasticity of the residual enzymes of the assembly line encoded in the BGC, or *ajuK* was involved later, leading to the production of ajudazols A and B in Cm c5.

2.4.3 Bioactivity

The biological activities of natural products may be severely influenced by both, the nature and degree of modifications of the primary core structure. Ajudazol A was reported to only show a minor activity against a few fungi and Gram-positive bacteria[28]. As the novel derivatives ajudazols C-J reported herein exhibit a bundle of different modifications when

compared to ajudazols A and B, ajudazols C-J were tested for their activity against several pathogens including bacteria and fungi. In addition, cytotoxicity was determined using various cancer cell lines (Table S14). However, none of the novel ajudazols exhibited an antimicrobial activity against tested strains. In contrast, ajudazols G, E, D, C, F, and H exhibited significant good to moderate cytotoxicity against HepG2, HCT-116, and KB3.1 cell lines (IC_{50} varies from 0.012 µg/ml to 7.75 µg/ml, Table S14). The glycosylated derivatives ajudazols I and J exhibited only marginal cytotoxicity (Table S14). It seems that a glycosylation at position *O*-33 decreases cytotoxicity. Eventually, such glycosylation may act as a mechanism of self-resistance to convert an active compound into inactive form which can be released by a hydrolase. Similar observations could be found in the case of sorangicin, as glycosylated sorangiosides do not show biological activity[25].

2.4.4 Conclusion

Taken together, a set of novel ajudazol derivatives produced by strain SBCb004 could be discovered and characterized by combining genomic and metabolomic approaches. DMA-CoA could be shown to be used as a starter unit instead of the more commonly used malonyl-CoA and acetyl-CoA. Ajudazols C-J exhibit both different hydroxylation and glycosylation patterns. Ajudazols C, D, E, F, G, and H exhibit good to intermediate levels of cytotoxicity.

2.5 Reference

(1) Dawid, W. Biology and global distribution of myxobacteria in soils. FEMS Microbiol. Rev. **2000**, 24 (4), 403–427.

(2) Kaplan, H.B.; Whitworth, D., Eds. *Multicellularity and differentiation among the myxobacteria and their neighbors*, **2007**.

(3) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34* (2), 135–160. DOI: 10.1039/C6NP00106H.

(4) Weissman, K. J.; Müller, R. Myxobacterial secondary metabolites: bioactivities and modes-of-action. *Nat. Prod. Rep.* **2010**, *27* (9), 1276–1295. DOI: 10.1039/c001260m.

(5) Wenzel, S. C.; Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* **2009**, *26* (11), 1385–1407. DOI: 10.1039/b817073h.

(6) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (26), 4688–4716. DOI: 10.1002/anie.200806121.

(7) Süssmuth, R. D.; Mainz, A. Nonribosomal peptide synthesis - Principles and prospects. *Angew. Chem. Int. Ed.* **2017**, *56* (14), 3770–3821. DOI: 10.1002/anie.201609079.

(8) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019** (47), W81-W87. DOI: 10.1093/nar/gkz310.

(9) Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K. J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **2010**, *17* (3), 296–309. DOI: 10.1016/j.chembiol.2010.01.016.

(10) Pogorevc, D. Estabilishing and engineering heterologous production systems für argyrin and α-pyrone antibiotics. Doctoral Thesis, Saarland University, Saarbrücken, 2019.

(11) Beyer, S.; Kunze, B.; Silakowski, B.; Müller, R. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim. Biophys. Acta* **1999**, *1445* (2), 185–195.

(12) Wright, G. D.; Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in bioscience : a journal and virtual library* **1999**, *4*, D9-21.

(13) Deuschle, U.; Kammerer, W.; Gentz, R.; Bujard, H. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* **1986**, *5* (11), 2987–2994.

(14) Rothberg, J. M.; Leamon, J. H. The development and impact of 454 sequencing. *Nat. Biotechnol.* **2008**, *26* (10), 1117–1124. DOI: 10.1038/nbt1485.

(15) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.;

Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **2005.** DOI: 10.1038/nature03959.

(16) Buntin, K.; Rachid, S.; Scharfe, M.; Blöcker, H.; Weissman, K. J.; Müller, R. Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.* **2008**, *47* (24), 4595–4599. DOI: 10.1002/anie.200705569.

(17) *Dictionary of Natural Products*. http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml.

(18) Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **2014**, *31* (6), 768–783. DOI: 10.1039/c3np70127a.

(19) Essig, S.; Schmalzbauer, B.; Bretzke, S.; Scherer, O.; Koeberle, A.; Werz, O.; Müller, R.; Menche, D. Predictive bioinformatic assignment of methyl-bearing stereocenters, total synthesis and an additional molecular target of ajudazol B. *J. Org. Chem.* **2016**, *81* (4), 1333–1357. DOI: 10.1021/acs.joc.5b02844.

(20) Popoff, A. Exploiting the biosynthetic potential of myxobacteria for natural product discovery. Doctoral thesis, Universität des Saarlandes, Saarbrücken, 2020.

(21) Bock, K.; Pedersen, C. Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. In *Advances in Carbohydrate Chemistry and Biochemistry*; pp 27–66. DOI: 10.1016/S0065-2318(08)60055-4.

(22) Li, Y.; Luxenburger, E.; Müller, R. An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. *Angew. Chem. Int. Ed. Engl.* **2012**, *52* (4), 1304–1308. DOI: 10.1002/anie.201207984.

(23) Silakowski, B.; Schairer, H. U.; Ehret, H.; Kunze, B.; Weinig, S.; Nordsiek, G.; Brandt, P.; Blöcker, H.; Höfle, G.; Beyer, S.; Müller, R. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **1999**, *274* (52), 37391–37399. DOI: 10.1074/jbc.274.52.37391.

(24) Kunze, B.; Reichenbach, H.; Müller, R.; Höfle, G. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (myxobacteria). *J. Antibiot.* **2005**, *58* (4), 244–251.

(25) Kopp, M.; Rupprath, C.; Irschik, H.; Bechthold, A.; Elling, L.; Müller, R. SorF, a glycosyltransferase with promiscuous donor substrate specificity in vitro. *ChemBioChem* **2007**, *8* (7), 813–819. DOI: 10.1002/cbic.200700024.

(26) Surup, F.; Viehrig, K.; Mohr, K. I.; Herrmann, J.; Jansen, R.; Müller, R. Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium *Pyxidicoccus fallax* active against multiresistant staphylococci. *Angew. Chem. Int. Ed. Engl.* **2014**, *49* (53), 13588–13591. DOI: 10.1002/anie.201406973.

(27) Etzbach, L.; Plaza, A.; Garcia, R.; Baumann, S.; Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Org. Lett.* **2014**, *16* (9), 2414–2417. DOI: 10.1021/ol500779s.

(28) Kunze, B.; Jansen, R.; Hofle, G.; Reichenbach, H. Ajudazols, new inhibitors of the mitochondrial electron transport from Chondromyces crocatus. Production, antimicrobial activity and mechanism of action. *J. Antibiot.* **2004**, *57* (2), 151–155.

Chapter 1 Supplementary information

Table S1 Oligonucleotide primers used in this study

Primer	Oligonucleotide in 5'-3' direction (introduced restriction site)
9101-sc-ajuD-F	GCG**CCTAGG**GTAGGGCTGGGAGCAGGTGG (XmaJI)
9101-sc-ajuD-R	CCT**CAATTG**GGCCCAGGTCACCGCTCTTG (MfeI)
9101-sc-ajuL-F	GCG**CCTAGG**CCCTTCTAGGAGCTGGACGC (XmaJI)
9101-sc- ajuL-R	CCT**CAATTG**GCTGGGCTAGTTCCTCCTCG (MfeI)
9101-sc-orf43-F	**CCAGTCTAGCTATTAATAGGCCTAGG**GTCTCTAGCAGCTCCACTC
9101-sc-orf43-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTTAGGCATCCTTCCCCAAT
9101-sc-orf42-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TGACATCACCGGCGCCTCC
9101-sc-orf42-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTACACCACGTCGTTGAGGG
9101-sc-ajuJ-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TAGTTCGTCTTCGTGAGCGC
9101-sc-ajuJ-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TCCGGCTAGATGTCCTCGC
9101-sc-orf32-F	**CCAGTCTAGCTATTAATAGGCCTAGG**AGGCCACTAATCCTCTGTGA
9101-sc-orf32-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTACGCGGGACGGAAGATGT
9101-sc-orf28-F1	**CCAGTCTAGCTATTAATAGGCCTAGG**GCCTGGTATGATGCAGGTGA
9101-sc-orf28-R1	**GGGCTGGCTTAAAGTCGACTCAATTG**AACCTAAGCTGGGAATGTCC
9101-sc- orf28-F2	TATTCCTCCGTAGCAGAAAG
9101-sc- orf28-R2	CTTTCTGCTACGGAGGAATA
9101-sc-orf15-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAAGCGGGACTTCACCGTC
9101-sc-orf15-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GGAAGTCCTACCAGTTCAGC
9101-sc-orf2-F1	**CCAGTCTAGCTATTAATAGGCCTAGG**GCATCAGTCGCCTCAAGCT
9101-sc-orf2-R1	**GGGCTGGCTTAAAGTCGACTCAATTG**GGCCTACGTCGGGTTGTAGA
9101-sc- orf2-F2	GTGATGGCGTAACTGGTGTT

9101-sc- orf2-R2	AACACCAGTTACGCCATCAC
9101-sc-ctg19_169-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CGGTAGCGGATGGATTCCTT
9101-sc-ctg19_169-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GGAGACGTCTTCACCATCCG
9101-sc-ajuI-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CAACGTGTAGTCCGTCAGCT
9101-sc-ajuI-R	**GGGCTGGCTTAAAGTCGACTCAATTG**ATGGTCTTCATCAGCTCGCC
9101-sc-ctg1_130-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CAGATCGAGCAGATGACCGT
9101-sc-ctg1_130-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TCGGCGGATTATGAGGAGTG
9101-sc-ctg10_386-F	**CCAGTCTAGCTATTAATAGGCCTAGG**GTGACGTCCAACCCTTCGT
9101-sc-ctg10_386-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GTTGTCATCCACGTACACGC
9101-sc-ajuM-F	**CCAGTCTAGCTATTAATAGGCCTAGG**AATTGATCGAACGCGTGTGG
9101-sc-ajuM-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CGTTCGGTATCCCCTTCGG
ajuD-CK-F	CATCCCGGACAAGTCACTGC
ajuL-CK-F	CAGGAACCCGTGTCCCTTC
orf43-CK-F	CACCCTCTTCTCCTATGCGC
orf42-CK-F	GATTGGGGAAGGATGCCGAA
ajuJ-CK-F	AGATGTCGCTGTACATGGCC
orf32-CK-F	GAAACTGTAACGAGCTGCGC
orf28-CK-F	TACATGCAGCAGCCCAATGA
orf15-CK-F	GTTCGATTCAGCGTGGATGC
orf2-CK-F	GTGGCTCACGAAATCCAACG
ctg19_169-CK-F	GGCCAACACCACCTTCATCT
ajuI-CK-F	CGAGCATGATGTGGCGATTG
ctg1_130-CK-F	TCAGCTCACCCAACCTGTTC
ctg10_386-CK-F	GTCGGTGTTCTGGATGGAGC
ajuM-CK-F	CCTTCTCTCGCATCGGTCC
CN-check-R	ACCTCTGACACATGCAGCTC

seq-pblueCN-R	ACCTCTGACACATGCAGCTC
9101-ajuG1-gap-F	TCCATGATCCGGGAGATGGA
9101-ajuG1-gap-R	GGCGAGACAGTCATCTCCG
9101-ajuE-gap-F	CGTTCTTGAGGATGGGGTGG
9101-ajuE-gap-R	TTTCGACAACAGCCCCTTCT
9101-ajuC-gap-F	GTGACAGTCACCGTCGAGTA
9101-ajuC-gap-R	ATATCGATTGGGACCGGATGC
9101-ajuA-gap-F	ATCGAAACGTGACTCGCTCC
9101-ajuA-gap-R	CCAGGAGGCGAAGGTGTC
9101-ajuL-gap-F	GTGCGATACCTGCAACGAAC
9101-ajuL-gap-R	GCGGAGTACCTCGCGAAG
9101-orf31-32-gap-F	GCAGCCCATCACAGAGGATT
9101-orf31-32-gap-R	GCACTGTCCTTGCTCGTCTA

The bold letters indicate the overlap sequence to the vector pbluekan.

Table S2. Plasmids used in this study

Plasmids	**Description/derivation**	**Reference**
pbluekan	*ori (pMB1), bla, lacZ, neo*	Unpublished
pbluekan-ajuD-HR	pbluekan derivative bearing internal homologous region of gene *ajuD*	This study
pbluekan-ajuL-HR	pbluekan derivative with gene *ajuL* homologous region	This study

pbluekan-orf42-HR	pbluekan derivative with gene *orf42* homologous region	This study
pbluekan-orf43-HR	pbluekan derivative with gene *orf41* homologous region	This study
pbluekan-ajuJ-HR	pbluekan derivative with gene *ajuJ* homologous region	This study
pbluekan-orf32-HR	pbluekan derivative with gene *orf31* homologous region	This study
pbluekan-orf28-HR	pbluekan derivative with gene *orf27* homologous region	This study
pbluekan-orf15-HR	pbluekan derivative with gene *orf15* homologous region	This study
pbluekan-orf2-HR	pbluekan derivative with gene *orf2* homologous region	This study
pbluekan-ctg19_169-HR	pbluekan derivative with gene ctg19_169 homologous region	This study
pbluekan-ajuI-HR	pbluekan derivative with gene *ajuI* homologous region	This study
pbluekan-ctg1_130-HR	pbluekan derivative with gene ctg1_130 homologous region	This study
pbluekan-ctg10_386-HR	pbluekan derivative with gene ctg10_386 homologous region	This study
pbluekan-ajuM-HR	pbluekan derivative with gene *ajuM* homologous region	This study

Table S3. Predicted proteins encoded in the ajudazol BGC of strain SBCb004

Protein	**Size aa***	**Proposed function of the similar protein**	**Sequence similarity to source**	**Similarity / identity aa* level**	**Accession number of the similar protein**
Orf1	515	zinc-ribbon domain-containing protein	*Vitiosangium sp.* GDMCC 1.1324	77 / 67 %	WP_108075970.1
Orf2	394	glycosyltransferase family 2 protein	*Archangium sp.* Cb G35	96 / 94 %	WP_073567095.1

Orf3	410	GTP-binding protein	*Pyxidicoccus fallax*	85 / 79 %	WP_169349557.1
Orf4	55	MULTISPECIES: 50S ribosomal protein L33	*Archangium*	98 / 87 %	WP_043409257.1
Orf5	386	hypothetical protein	*Cystobacter ferrugineus*	71 / 60 %	WP_143177574.1
Orf6	273	GNAT family N-acetyltransferase	*Cystobacter ferrugineus*	88 / 82 %	WP_071905383.1
Orf7	251	C40 family peptidase	*Vitiosangium sp.* GDMCC 1.1324	94 / 86 %	WP_108075973.1
Orf8	563	HEAT repeat domain-containing protein	*Vitiosangium sp.* GDMCC 1.1324	95 / 89 %	WP_108075974.1
Orf9	676	ATP-dependent helicase	*Vitiosangium sp.* GDMCC 1.1324	96 / 93 %	WP_108075975.1
Orf10	544	Hsp70 family protein	*Vitiosangium sp.* GDMCC 1.1324	99 / 96 %	WP_108075976.1
Orf11	142	hypothetical protein	*Vitiosangium sp.* GDMCC 1.1324	92 / 89 %	WP_108075977.1
Orf12	93	hypothetical protein	*Archangium gephyra*	96 / 93 %	WP_047857416.1
Orf13	328	TerC family protein	*Archangium sp.* Cb G35	92 / 89 %	WP_073567110.1
Orf14	299	zinc metalloprotease HtpX	*Archangium gephyra*	97 / 91 %	WP_047857418.1
Orf15	381	MBL fold metallo-hydrolase	*Vitiosangium sp.*	92 / 85 %	WP_108075982.1

			GDMCC 1.1324		
Orf16	263	4'-phosphopantetheinyl transferase superfamily protein	*Archangium gephyra*	92 / 87 %	WP_047857428.1
Orf17	255	thioesterase	*Vitiosangium sp.* GDMCC 1.1324	91 / 87 %	WP_108075988.1
AjuL	3307	hybrid NRPS/type I PKS ($C_{7\text{-}454}$, $A_{461\text{-}980}$, $M_{927\text{-}1338}$, $PCP_{1403\text{-}1467}$, $KS_{1493\text{-}1918}$, $AT_{2023\text{-}2317}$, $DH_{2383\text{-}2542}$, $KR_{2858\text{-}3036}$, $ACP_{3137\text{-}3204}$)	*Chondromyces crocatus*	73 / 60 %	WP_050433124.1
Orf18	491	glycoside hydrolase family 6 protein	*Vitiosangium sp.* GDMCC 1.1324	88 / 80 %	WP_108069950.1
Orf19	303	DUF3396 domain-containing protein	*Archangium violaceum*	85 / 72 %	WP_152622499.1
Orf20	283	hypothetical protein	*Vitiosangium sp.* GDMCC 1.1324	79 / 69 %	WP_108072391.1
Orf21	401	hypothetical protein DAT35_32570	*Vitiosangium sp.* GDMCC 1.1324	73 / 62 %	PTL79551.1
Orf22	452	SWIM zinc finger family protein	*Vitiosangium sp.* GDMCC 1.1324	92 / 87 %	WP_108075989.1
Orf23	181	MarR family transcriptional regulator	*Vitiosangium sp.* GDMCC 1.1324	82 / 70 %	WP_108069654.1

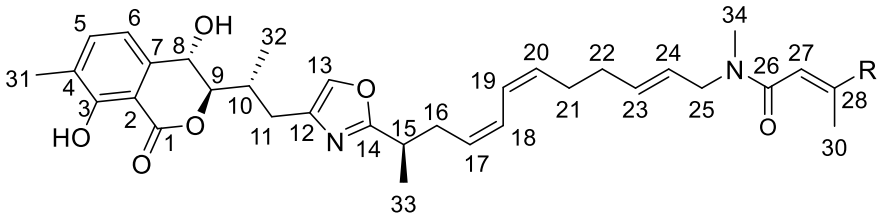
Orf24	480	hypothetical protein	*Vitiosangium sp.* GDMCC 1.1324	88 / 81 %	WP_108075990.1
Orf25	362	AAA family ATPase	*Vitiosangium sp.* GDMCC 1.1324	98 / 97 %	WP_108075991.1
Orf26	754	hypothetical protein	*Vitiosangium sp.* GDMCC 1.1324	96 / 93 %	WP_108075992.1
Orf27	384	VWA domain-containing protein	*Vitiosangium sp.* GDMCC 1.1324	97 / 93 %	WP_108075993.1
Orf28	254	alpha/beta fold hydrolase	*Chondromyces fuscus*	92 / 87 %	WP_002628336.1
Orf29	198	hypothetical protein BON30_35460	*Cystobacter ferrugineus*	84 / 76 %	OJH35915.1
Orf30	117	serine/threonine protein kinase	*Archangium sp.* Cb G35	95 / 95 %	WP_073562591.1
Orf31	311	hypothetical protein BO221_22375	*Archangium sp.* Cb G35	95 / 86 %	OJT22519.1
Orf32	1482	SIR2 family protein	*Archangium sp.* Cb G35	79 / 70 %	WP_073567252.1
AjuJ	456	cytochrome P450	*Chondromyces crocatus*	77 / 59 %	WP_169796669.1
AjuA	2147	type I PKS (KS_{30-458}, $AT_{564-861}$, $DH_{923-1088}$, $ER_{1435-1741}$, $KR_{1753-1931}$, $ACP_{2031-2096}$)	*Chondromyces crocatus*	71 / 60 %	WP_050433113.1
AjuB	3380	type I PKS (KS_{37-463}, $AT_{568-862}$, $KR_{1146-1324}$, $ACP_{1430-1499}$, $KS_{1523-1951}$, $AT_{2059-2352}$, $DH_{2422-2597}$,	*Chondromyces crocatus*	73 / 61 %	WP_050433114.1

		$KR_{2970-3148}$, $ACP_{3264-3321}$)			
AjuC	2190	type I PKS (KS_{36-465}, $AT_{570-860}$, $DH_{924-1094}$, $ER_{1475-1787}$, $KR_{1808-1987}$, $ACP_{2090-2156}$)	*Chondromyces crocatus*	73 / 61 %	WP_050433115.1
AjuD	1404	NRPS (Cy_{70-503}, $A_{514-1044}$, $PCP_{1310-1374}$)	*Chondromyces crocatus*	76 / 63 %	WP_050433116.1
AjuE	2167	type I PKS (KS_{7-433}, $AT_{538-828}$, $DH_{891-1057}$, $ER_{1437-1749}$, $KR_{1771-1950}$, $ACP_{2051-2117}$)	*Chondromyces crocatus*	70 / 57 %	WP_050433117.1
AjuF	1552	type I PKS (KS_{36-462}, $AT_{568-862}$, $KR_{1154-1332}$, $ACP_{1434-1501}$)	*Chondromyces crocatus*	67 / 55 %	WP_050433118.1
AjuG1	968	type I PKS (KS_{37-464}, $AT_{569-866}$)	*Chondromyces crocatus*	74 / 62 %	WP_050433119.1
AjuG2	1936	type I PKS (ACP_{2-69}, KS_{95-524}, $AT_{625-920}$, $DH_{988-1153}$, $KR_{1516-1705}$, $ACP_{1810-1877}$)	*Chondromyces crocatus*	66 / 53 %	WP_050433119.1
AjuH	1635	type I PKS (KS_{37-464}, $AT_{569-868}$, $DH_{934-1110}$, $ACP_{1276-1342}$, $TE_{1387-1622}$)	*Chondromyces crocatus*	68 / 56 %	WP_050433120.1
Orf33	617	asparagine synthase (glutamine-hydrolyzing)	*Vitiosangium sp.* GDMCC 1.1324	90 / 81 %	WP_108075169.1
Orf34	54	SgcJ/EcaC family oxidoreductase	*Streptosporangium sp. 'caverna'*	91 / 80 %	WP_110698014.1
Orf35	161	SRPBCC family protein	*Sandaracinus amylolyticus*	81 / 68 %	WP_053231537.1

Orf36	332	LysR family transcriptional regulator	*Polyangiaceae bacterium*	64 / 48 %	NUP07428.1
Orf37	329	hypothetical protein	*Cystobacter fuscus*	87 / 79 %	WP_155893451.1
Orf38	1054	hypothetical protein BE20_01620	*Sorangium cellulosum*	74 / 65 %	KYF92238.1
Orf39	161	anhydro-N-acetylmuramic acid kinase	*Pantoea septica*	58 / 40 %	WP_033789858.1
Orf40	353	SMP-30/gluconolactonase/LRE family protein	*Cystobacter ferrugineus*	91 / 85 %	WP_084736471.1
Orf41	268	lectin	*Chondromyces crocatus*	48 / 25 %	WP_050430387.1
Orf42	546	FAD-dependent oxidoreductase	*Archangium sp.* Cb G35	96 / 95 %	WP_083680752.1
Orf43	529	FAD-binding oxidoreductase	*Archangium sp.* Cb G35	97 / 95 %	WP_073559220.1
AjuI	459	cytochrome P450	*Chondromyces crocatus*	80 / 63 %	WP_050433121.1
Ctg19_169	476	cytochrome P450	*Chondromyces crocatus*	60 / 41 %	WP_050433121.1
AjuM	419	glycosyltransferase	*Sorangium cellulosum*	50 / 36 %	ADN68481.1
Ctg1_130	433	glycosyltransferase	*Sorangium cellulosum*	41 / 29 %	ADN68481.1
Ctg10_386	438	glycosyltransferase	*Sorangium cellulosum*	44 / 30 %	ADN68481.1

*aa: amino acids

Table S4. 1H-NMR spectroscopic Data for Ajudazol B[1] (1) vs C (2)

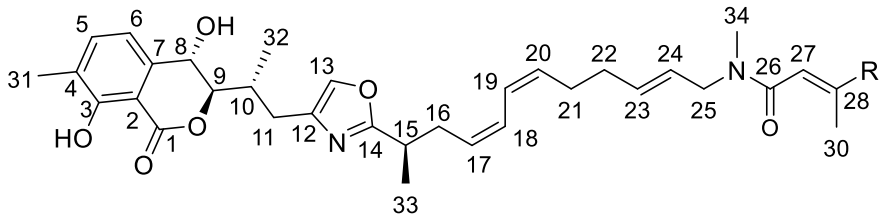


Position	Chemical shift (δ_H, in ppm), Ajudazol B[1] (acetone-d_6, 600 MHz)	Chemical shift (δ_H, in ppm), Ajudazol C (acetone-d_6, 700 MHz)	Δ ($\delta_{Ajudazol\ B}$ - $\delta_{Ajudazol\ C}$, in ppm)
1	-	-	-
2	-	-	-
3	-	-	-
3-OH	11.28	11.29	-0.01
4	-	-	-
5	7.47	7.47	0.00
6	7.08	7.07	0.01
7	-	-	-
8	4.96	4.97	0.00
9	4.43	4.43	0.00
10	2.47	2.45	0.02
11a	2.89	2.89	0.00
11b	2.49	2.50	-0.01
12	-	-	-
13	7.61	7.61	0.00
14	-	-	-
15	3.03	3.03	0.00
16a	2.62	2.61	0.01
16b	2.49	2.49	0.00
17	5.40	5.40	0.00
18	6.29	6.29	0.00
19	6.24	6.24	0.00
20	5.45	5.44	0.01
21	2.24	2.24	0.00
22	2.11	2.12	-0.01

23	5.60	5.59	0.01
24	5.47	5.43	0.04
25	3.92	3.90/3.88[b]	0.02/-
26	-	-	-
27	5.33	5.89	-0.56
28	-	-	-
29	2.13	1.81	0.32
30	-	1.89	-
31	2.22	2.22	0.00
32	1.06	1.06	0.00
33	1.29	1.292	0.00
34	2.84/2.94[b]	2.81/2.91[b]	0.03/0.03

[a]Adjusted to the solvent signal of acetone-d_6 (δ_H 2.05 ppm). [b]Due to the amide substructure signal doubling is observed.[1]

Table S5. ^{13}C-NMR spectroscopic Data for Ajudazol B[1] (1) vs C (2)

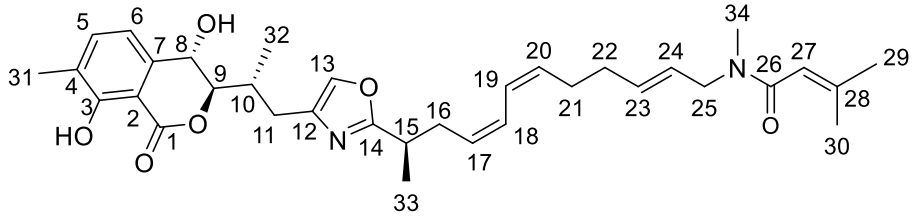


Position	Chemical shift (δ_C, in ppm), Ajudazol B[1] (acetone-d_6, 150 MHz)[a]	Chemical shift (δ_C, in ppm), Ajudazol C (acetone-d_6, 175 MHz)[a]	Δ ($\delta_{Ajudazol\ B}$ - $\delta_{Ajudazol\ C}$, in ppm)
1	170.3	170.3	0
2	107.2	107.3	-0.1
3	160.6	160.7	-0.1
4	126.1	126.2	-0.1
5	137.9	138.1	-0.3
6	116.5	116.6	-0.1
7	142.1	142.2	-0.1
8	65.3	65.4	-0.1
9	88.1	88.2	-0.1
10	33.8	33.9	-0.1

11	27.8	27.9	-0.1
12	139.4	139.5	-0.1
13	135.8	135.9	-0.1
14	168.2	168.3	-0.1
15	34.6	34.7	-0.1
16	33.5	33.6	-0.1
17	129.0	129.1	-0.1
18	126.3	126.4	-0.1
19	124.6	124.8	-0.2
20	132.4	132.4	0
21	27.8	27.9	-0.1
22	32.8	32.8	0
23	132.7	133.1	-0.4
24	127.0	126.8	0.2
25a[b]	52.5	52.4	0.1
25b[b]	49.2	48.9	0.3
26	167.9	167.9	0
27	92.3	119.1	-26.8 [c]
28	168.7	146.7	22.0 [c]
29	18.8	26.4	-7.6 [c]
30	-	20.2	-
31	15.4	15.5	-0.1
32	16.9	17.0	-0.1
33	18.3	18.4	-0.1
34a[b]	34.3	35.0	-0.7
34b[b]	33.2	32.5	0.7

[a]Adjusted to the solvent signal of acetone-d_6 (δ_C 29.9 ppm). [b]Due to the amide substructure signal doubling is observed.[1] [c]greater Δ ($\delta_{\text{Ajudazol B}}$ - $\delta_{\text{Ajudazol C}}$) due to difference in structure.

Table S6. NMR Spectroscopic Data for Ajudazol C (2)[a]

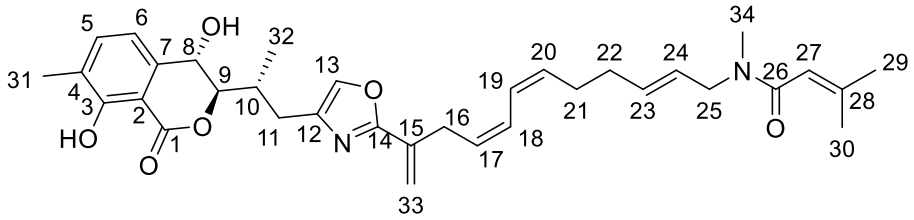


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1, 26[h]	170.8, C			30, 34, 25, 9, 27, 6

2	107.5, C			31, 8, 6, 5
3	161.2, C			31, 6, 5
4	127.3, C			31, 6
5	138.6, CH	7.46, dd (7.59, 0.72)	31, 6	31, 6, 8, 2, 7, 3
6	117.4, CH	7.01, d (7.93)	31, 8, 5	31, 8, 5, 2, 4, 3, 1
7	141.4, C			9, 8, 5
8	65.8, CH	4.89, d[g]	9, 6	9, 6, 5, 10, 2, 7
9	89.0, CH	4.35, dd (7.48, 4.96)	10, 8	32, 11, 8, 10, 7, 1
10	34.5, CH	2.29, m	32, 11, 9	32, 11, 9, 8, 13
11a	28.4, CH2	2.87, m	32, 10, 11, 13	32, 9, 10, 13, 12
11b	28.4, CH2	2.45, s	10, 11, 13	32, 9, 10, 13, 12
12	139.3, C			11
13	136.9, CH	7.58, s	11a, 11b	11, 10, 14
14	169.8, C			33, 16, 15, 13,
15	35.5, CH	3.05, m	33, 16, 16	33, 16, 17, 14
16a	34.1, CH2	2.60, m	16, 15, 17, 18	33, 15, 18, 17, 14
16b	34.1, CH2	2.49, m	16, 15, 17, 18	33, 15, 18, 17, 14
17	129.0, CH	5.35, m	16, 18	16, 15, 19
18	126.9, CH	6.28, m	16, 17, 19	16, 20
19	124.9, CH	6.20, m	20, 18	17, 22
20	132.9, CH	5.41, m	19	21, 23, 18, 25
21	33.3, CH2	2.13, m	22, 23	20, 22, 24, 23
22	28.2, CH2	2.10, m	21, 23	21, 19, 24, 23
23	134.4, CH	5.56, m	22, 21, 24	22, 21, 25, 20, 24
24	126.1, CH	5.42, m	25, 23	22, 21, 25, 23
25a[f]	50.0, CH2	3.93, d (6.26)	34, 24	34, 20, 24, 23, 26
25b[f]	53.4, CH2	3.90, dd (5.65, 0.92)	22, 24, 23	24, 23, 21
27	119.4, CH	5.88, m	30	30, 29, 28, 26
28	148.2, C			30, 27
29	26.3, CH3	1.84, m		30, 27
30	20.5, CH3	1.86, m	27	27, 29, 28, 26
31	15.7, CH3	2.25, m	6, 5	5, 2, 6, 4, 3
32	16.6, CH3	1.00, m	10, 11	11, 9, 10
33	18.5, CH3	1.32, d (7.02)	15	16, 15, 14
34a[f]	35.8, CH3	2.93, m	25	25, 26
34b[f]	33.2, CH	2.88, m	25	25, 26

[a]Recorded in methanol-d_4. [b]Acquired at 150 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1] [g]No coupling constant analysis due to signal overlap. [h]Carbons showing the same signal.

Table S7. NMR Spectroscopic Data for Ajudazol D (3)[a]

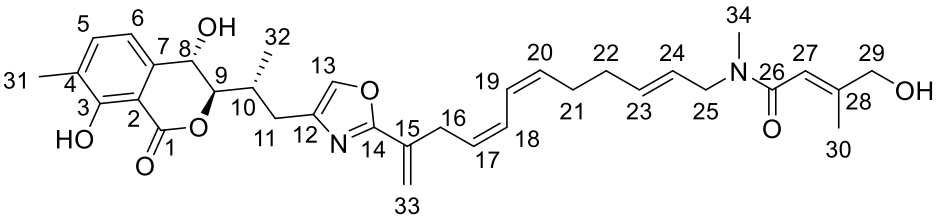


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.8, C			9, 6,
2	107.5, C			31, 8, 6, 5,
3	161.2, C			31, 6, 5,
4	127.4, C			31, 6,
5	138.6, CH	7.45, dd (7.59, 0.75)	31, 6	31, 6, 31, 2, 6, 7, 3
6	117.5, CH	7.01, d (7.48)	31, 8, 5	8, 5, 2, 4, 3, 1
7	141.4, C			9, 8, 5,
8	65.8, CH	4.91, s	31, 9, 6	9, 6, 10, 2, 7
9	89, CH	4.38, dd (7.37, 5.13)	32, 10, 8	32, 10, 11, 8, 32, 7, 1
10	34.6, CH	2.34, m	32, 11, 9	32, 11, 9, 8
11b	28.6, CH2	2.53, dd (14.69, 9.35)	32, 10, 11, 13	32, 9, 10, 13, 12
11a	28.6, CH2	2.91, m	32, 10, 11, 13	32, 9, 10, 13, 12
12	140.9, C			11, 13,
13	137.4, CH	7.66, s	11	11, 12, 14
14	163.6, C			16, 33, 13
15	136, C			16, 33
16	31.3, CH2	3.36, m	33, 17, 18	33, 18, 17, 15, 14
17	128.4, CH	5.56, m	16, 18	16, 19
18	126.9, CH	6.41, m	16, 17, 19	16
19	125, CH	6.35, m	22, 23, 18	22, 17
20	131, CH	5.34, m	22	22
21	33.3, CH2	2.16, m	22, 24	22, 23

22b	28.3, CH2	2.03, br d (6.20)	20	21, 20, 19
22a	28.3, CH2	2.30, m	21, 23, 19	21, 20, 19, 23
23	133.2, CH	5.49, m	22, 19	21, 22
24	126.1, CH	5.43, m	21, 25	25, 34
25a[f]	49.8, CH2	3.93, dd (6.20, 0.54)	21, 24	24, 26
25b[f]	53.4, CH2	3.90, dd (5.66, 1.07)	24	34, 24, 26
26	171.1, C			30, 34, 25, 27,
27	119.2, CH	5.87, m	29, 30, 34	29, 30, 28, 26
28	148.2, C			29, 30, 27
29	26.3, CH3	1.84, d (1.28)	27	30, 27, 28
30	20.6, CH3	1.86, d (1.17)	27	29, 27, 28, 26
31	15.7, CH3	2.24, s	8, 6, 5	5, 2, 4, 3
32	16.7, CH3	1.04, d (6.95)	10, 11, 9	11, 9, 10
33a	118.8, CH2	5.97, s	16, 33	16, 15, 14
33b	118.8, CH2	5.42, m	16, 33	16, 15, 14
34a[f]	35.8, CH3	2.93, m	27	25, 24, 26
34b[f]	33.2, CH3	2.87, s	27	25, 24, 26

[a]Recorded in methanol-d_4. [b]Acquired at 150 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1]

Table S8. NMR Spectroscopic Data for Ajudazol E (4)[a]

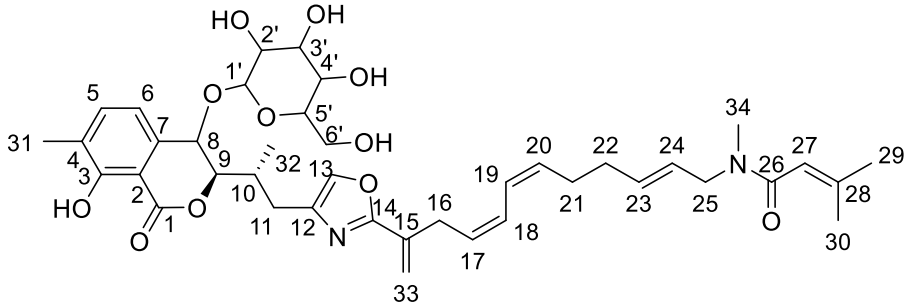


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1, 26[g]	170.8, C			30, 34, 25, 9, 27, 6
2	107.5, C			31, 8, 6, 5
3	161.2, C			31, 6, 5
4	127.4, C			31, 6,
5	138.6, CH	7.45, dd (7.54, 0.70)	31, 6	31, 6, 30, 2, 7, 3
6	117.5, CH	7.01, d (7.49)	31, 8, 5	8, 2, 4, 5, 3, 1

7	141.2, C			8, 5
8	65.9, CH	4.93, m	9, 6	9, 6, 10, 2, 7
9	89.1, CH	4.39, dd (5.42, 7.28)	10, 8	32, 11, 8, 10, 12, 1
10	34.6, CH	2.35, m	32, 11, 9	32, 11, 9, 8,
11a	28.6, CH2	2.91, m	10, 11, 13	32, 9, 10, 13, 12
11b	28.6, CH2	2.53, m	10, 11, 13	32, 9, 10, 13, 12
12	140.9, C			11, 9, 13,
13	137.3, CH	7.65, s	11	11, 12, 14
14	163.4, C			16, 33, 13,
15	136.0, C			16, 33, 17, 18,
16	31.3, CH2	3.36, m	33, 17, 33, 18	33, 17, 18, 15, 14
17	128.4, CH	5.56, m	16, 18	16, 19, 15
18	126.9, CH	6.41, m	16, 17, 19	16, 20, 15
19	125.0, CH	6.35, m	22, 20, 18	22, 17
20	133.3, CH	5.50, m	22, 19	21, 22, 18
21	33.3, CH2	2.17, m	22, 23	22, 24, 20, 23
22	28.3, CH2	2.30, q (7.06, 7.06, 7.06)	21, 20, 19	21, 20, 23, 19
23	134.6, CH	5.64, m	21, 25, 24	22, 25, 21
24	126.2, CH	5.44, m	25, 23	21, 25
25a[f]	49.9, CH2	3.96, d (6.21)	34, 24, 23	34, 24, 23, 26
25b[f]	53.4, CH2	3.92, dd (7.48, 13.26)	24, 23	24, 23, 21, 26
27	116.8, CH	6.16, m	30, 29	30, 29, 28, 26
28	150.0, C			30, 29, 27,
29	66.9, CH2	4.03, br d (9.95)	30, 27	30, 27, 28
30	15.7, CH3	1.81, br s	29, 27	29, 5, 27, 28, 26
31	15.7, CH3	2.24, s	6, 5	29, 5, 2, 4, 3
32	16.7, CH3	1.04, m	10	11, 9, 10
33a	118.8, CH2	5.97, m	16, 33	16, 15, 14
33b	118.8, CH2	5.42, m	16, 33	16, 15, 14
34a[f]	36.0, CH3	2.95, s	25	25, 26
34b[f]	33.2, CH	2.90, m	25	25, 26

[a]Recorded in methanol-d_4. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1] [g]Carbons showing the same signal.

Table S9. NMR Spectroscopic Data for Ajudazol F (5)[a]

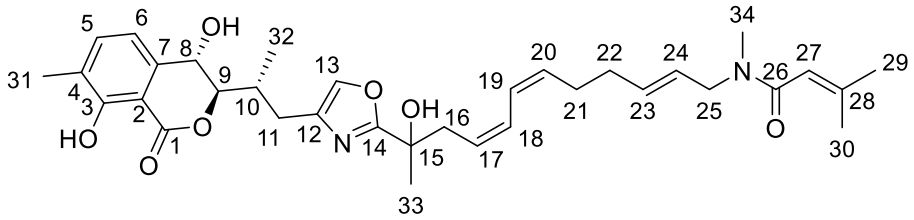


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.3, C			31, 9, 8, 6
2	108.3, C			31, 8, 6, 5
3	161.0, C			31, 6, 5
4	128.3, C			31, 6
5	138.5, CH	7.44, dd (7.56, 0.80)	31, 6	31, 8, 6, 2, 7, 3
6	120.6, CH	7.10, d (7.56)	31, 8, 5	31, 8, 5, 2, 4, 3, 1
7	137.1, C			31, 9, 8, 5
8	73.4, CH	5.10, d (3.89)	31, 9, 6	1', 9, 6, 5, 10, 2, 7, 1
9	87.6, CH	4.71, dd (7.68, 3.89)	31, 8	32, 10, 11, 8, 7, 1
10	35.0, CH	2.23, m	32, 11	32, 11, 9, 8, 13, 12
11a	29.7, CH2	2.91, m	32, 10, 11, 13	32, 10, 9, 13, 12
11b	29.7, CH2	2.49, dd (14.78, 8.71)	32, 10, 11, 13	32, 10, 9, 13, 12
12	140.8, C			10, 11, 13
13	137.3, CH	7.64, s	11, 33	11, 10, 12, 14
14	163.6, C			16, 33, 13
15	136.0, C			16, 33
16	31.3, CH2	3.33, m	33, 17, 18	33, 17, 18, 15, 14
17	128.5, CH	5.55, m	16, 18	16, 19
18	126.9, CH	6.39, d (10.08)	16, 17	16, 20, 19
19	125.1, CH	6.35, br d (11.57)	22, 20	22, 17, 18
20	133.2, CH	5.47, m	22, 19	21, 22, 23, 18
21	33.3, CH2	2.16, m	22, 24, 23	24, 22, 20, 23
22	28.3, CH2	2.29, m	21, 20, 19	21, 20, 23, 19
23	134.4, CH	5.59, m	21, 25, 24	21, 22, 24, 20
24	126.2, CH	5.43, m	21, 25, 23	21, 23, 25
25a[f]	49.8, CH2	3.93, m	24, 23	24, 26
25b[f]	53.4, CH2	3.91, m	21, 24, 23	34, 24, 23
26	170.6, C			30, 34, 25, 27

27	119.4, CH	5.87, dt (2.78, 1.30, 1.30)	30, 34	29, 30, 28, 26
28	147.8, C			30, 27,
29	26.3, CH3	1.84, m		30, 27, 30
30	20.6, CH3	1.86, m	27	29, 27, 28, 26
31	15.8, CH3	2.25, m	9, 8, 6, 5	5, 2, 6, 4, 7, 3, 1
32	16.7, CH3	0.98, d (6.87)	10, 11	10, 11, 9
33a	118.7, CH2	5.96, s	16, 33	16, 15, 14
33b	118.7, CH2	5.39, m	16, 33, 13	16, 15, 14
34a[f]	35.8, CH	2.93, m	25	25, 26
34b[f]	33.2, CH3	2.87, m	27	26
1'	103.6, CH	4.57, d (7.79)	2'	2', 8
2'	75.2, CH	3.20, dd (9.16, 7.79)	3', 1'	4', 3', 1', 5'
3'	78.2, CH	3.37, m	2', 4'	5', 4', 2'
4'	71.6, CH	3.28, m	5', 3'	2', 3', 6'
5'	78.2, CH	3.26, m	4', 6'	2', 3', 6'
6'a	62.8, CH2	3.78, dd (11.86, 1.89)	5', 6'	4', 6', 5'
6'b	62.8, CH2	3.62, dd (11.80, 5.38)	5', 6'	4', 6', 5'

[a]Recorded in methanol-d_4. [b]Acquired at 150 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1]

Table S10. NMR Spectroscopic Data for Ajudazol G (6)[a]



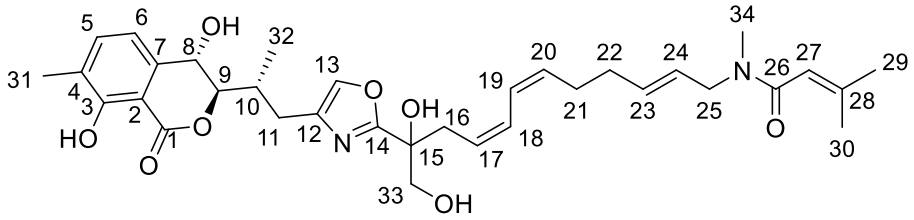
Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.7, C			9
2	107.3, C			31, 8, 6
3	161.1, C			31, 5
4	127.2, C			31, 6
5	138.6, CH	7.46, dd (7.51, 0.77)	31, 6	31, 6, 7, 3

6	117.3, CH	7.02, dd (7.51, 0.58)	5	8, 2, 4, 5
7	141.6, C			9, 8, 5
8	65.7, CH	4.9, br d (7.80)	9	9, 6, 10, 2, 7
9	89.0, CH	4.35, dd (7.71, 4.72)	10, 8	32, 11, 8, 10, 7, 1
10	34.5, CH	2.31, m	32, 11, 9	32, 11, 9, 8
11a	28.3, CH2	2.88, m	11, 13	32, 9, 10, 13, 12
11b	28.3, CH2	2.48, dd (14.55, 9.73)	10, 11	32, 9, 10, 13, 12
12	139.4, C			11, 13
13	137.3, CH	7.63, s	11	11, 12, 14
14	169.2, C			33, 16, 13
15	72.7, C			33, 16
16a	40.7, CH2	2.77, m	16, 17	33, 17, 18, 15, 18, 14
16b	40.7, CH2	2.71, m	16, 17	33, 17, 18, 15, 18, 14
17	126.2, CH	5.37, m	16, 16, 18	16
18	127.7, CH	6.31, m	17, 19	16, 19, 20
19	125.1, CH	6.19, m	20, 18	18, 24, 18
20	133.0, CH	5.41, m	22, 23, 19	21, 22, 18
21	33.3, CH2	2.12, m	22, 23	22, 20, 24, 23
22	28.2, CH2	2.22, m	21, 20	21, 20, 23
23	134.4, CH	5.57, m	21, 25, 20	21, 22, 25
24	126.2, CH	5.39, m	25	21, 25, 19
25a[f]	49.8, CH2	3.93, br d (6.36)	25, 24	24, 23
25b[f]	53.4, CH2	3.90, br dd (5.83, 1.11)	24, 23	34, 23, 24, 26
26	170.9, C			30, 34, 25, 27
27	119.4, CH	5.88, m	29, 30, 34	29, 30, 26
28	148.2, C			29, 30
29	26.3, CH3	1.84, m	27	30, 27, 28
30	20.5, CH3	1.86, m	27	29, 27, 28, 26
31	15.7, CH3	2.25, s	5	5, 2, 4, 3
32	16.6, CH3	1.00, d (6.94)	10	11, 10, 9
33	26.4, CH3	1.56, s		16, 15, 14
34a[f]	35.8, CH	2.93, s		26
34b[f]	33.3, CH3	2.87, m	27	25, 26

[a]Recorded in methanol-d_4. [b]Acquired at 175 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to

indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1]

Table S11. NMR Spectroscopic Data for Ajudazol H (7)[a]

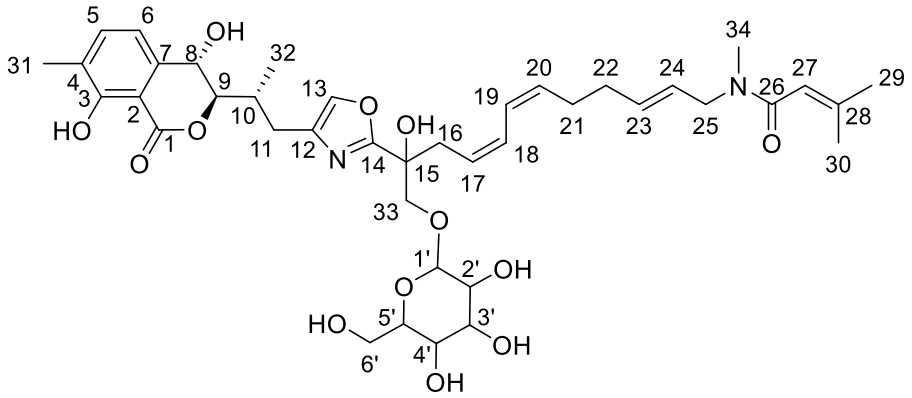


Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	170.9, C			9, 8, 6
2	107.5, C			31, 8, 6, 5
3	161.2, C			31, 6, 5
4	127.3, C			31, 6
5	138.6, CH	7.46, d (7.71)	31, 6	31, 6, 8, 2, 7, 3
6	117.4, CH	7.02, d (7.63)	31, 8, 5	31, 8, 5, 2, 4, 3, 1
7	141.6, C			9, 8, 5
8	65.7, CH	4.89, d[g]	9, 6	9, 6, 5, 10, 2, 7, 1
9	88.9, CH	4.35, dd (7.71, 4.58)	32, 10, 8	32, 10, 11, 8, 7, 1
10	34.6, CH	2.32, m	32, 11, 9	32, 11, 9, 8, 13, 12
11a	28.4, CH2	2.88, m	32, 10, 11, 13	32, 9, 10, 13, 12
11b	28.4, CH2	2.48, dd (9.61, 14.57)	32, 10, 11, 13	32, 9, 10, 13, 12
12	139.7, C			10, 11, 13,
13	137.4, CH	7.64, s	11	11, 10, 12, 14
14	167.6, C			16, 33, 13
15	76.3, C			16, 33, 17
16	35.6, CH2	2.76, m	17, 18	33, 17, 18, 15, 14
17	132.8, CH	5.42, m	16, 19, 18	21, 22, 20, 18, 16, 15, 19
18	127.7, CH	6.30, m	16, 17, 19	16, 17, 17
19	125.2, CH	6.21, m	22, 17, 18	22, 17, 21, 24
20	134.4, CH	5.56, m	21	21, 22, 17, 24
21	33.3, CH2	2.11, m	22, 25, 20, 23	22, 24, 20, 23, 19, 17
22	28.2, CH2	2.21, m	21, 24, 19	21, 24, 20, 23, 19, 17
23	135.0, CH	5.62, m	21, 25, 24	21, 25, 22
24	125.9, CH	5.41, m	22, 25, 23	21, 25, 20, 19, 22
25a[f]	49.8, CH2	3.93, d (6.18)	21, 34, 24, 23	34, 24, 23

25b[f]	53.4, CH2	3.90, br d (5.57)		24, 20, 26
26	171.1, C			30, 34, 27
27	119.1, CH	5.87, br d (8.01)	29, 30	29, 30, 28, 26
28	148.2, C			29, 30, 27
29	26.3, CH3	1.84, d (0.76)	30, 27	30, 27, 28
30	20.6, CH3	1.86, s	29, 27	29, 27, 28, 26
31	15.7, CH3	2.25, s	6, 5	5, 2, 6, 4, 3
32	16.7, CH3	1.01, m	10, 11, 9	11, 10, 9
33	68.0, CH2	3.80, dd (11.29, 33.04)		16, 15, 14
34a[f]	35.8, CH3	2.93, s	25	25, 26
34b[f]	33.2, CH3	2.87, m	25	

[a]Recorded in methanol-d_4. [b]Acquired at 150 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1] [g]No coupling constant analysis due to signal overlap.

Table S12. NMR Spectroscopic Data for Ajudazol I (8)[a]

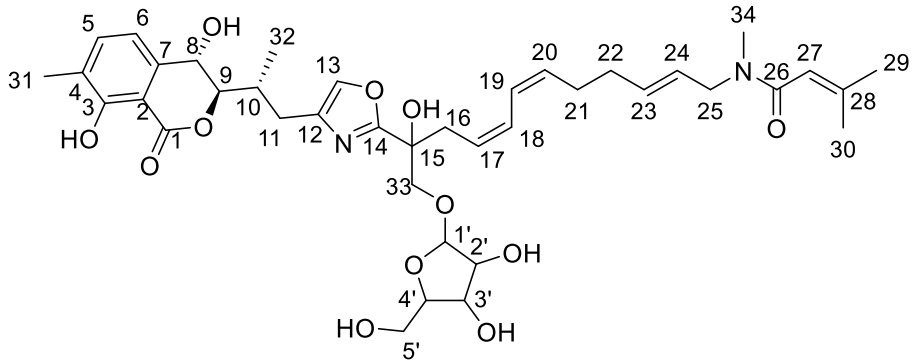


Position	**δ_C[b], type**	**δ_H[c], (*J* in Hz)**	**COSY[d]**	**HMBC[e]**
1, 26[g]	170.9, C			30, 9, 8, 27, 6
2	107.5, C			31, 8, 6, 5
3	161.3, C			31, 6, 5
4	127.4, C			31, 8, 6
5	138.7, CH	7.46, br d (7.55)	31, 6	31, 8, 6, 2, 7, 3
6	117.4, CH	7.02, d (7.48)	31, 8, 5	31, 8, 5, 2, 4, 3, 1
7	141.5, C			31, 9, 8, 5,
8	65.8, CH	4.89, d[h]	31, 9, 6	9, 6, 5, 10, 2, 4, 7, 1

9	89.0, CH	4.35, m	31, 8	32, 10, 11, 8, 7, 1
10	34.5, CH	2.32, m	32, 11	32, 11, 9, 8, 13, 12
11a	28.3, CH2	2.87, br d (4.43)	32, 10, 11, 13	32, 9, 10, 13, 12
11b	28.3, CH2	2.48, m	32, 10, 11, 13	32, 9, 10, 13, 12
12	139.8, C			10, 11, 13
13	137.4, CH	7.64, s	11, 33	11, 10, 12, 14
14	167.2, C			16, 33, 13
15	87.1, C			16
16	36.4, CH2	2.76, m	33, 17, 18	33, 18, 15, 24, 14
17	126.2, CH	5.42, m	16, 18	21
18	127.9, CH	6.29, m	16, 17	16, 19, 20
19	125.1, CH	6.17, m	22, 20	22, 18
20	133.1, CH	5.41, m	22, 19	21, 22, 18, 25
21	33.3, CH2	2.12, m	22, 24, 23	22, 23, 17, 20
22	28.2, CH2	2.20, m	21, 20, 19	21, 20, 19, 23
23	135.0, CH	5.59, m	21, 25, 24	21, 22
24	125.4, CH	5.40, m	21, 25, 23	16
25a[f]	49.9, CH2	3.93, dd[h]	24, 23	20
25b[f]	53.4, CH2	3.90, d[h]	21, 24, 23	22, 17, 26
27	119.1, CH	5.87, br d (8.70)		30, 29, 28, 26
28	148.2, C		30, 34	30, 27
29	26.3, CH3	1.84, d (0.99)		30, 27, 30
30	20.6, CH3	1.85, d (1.15)		29, 27, 28, 26
31	15.7, CH3	2.24, br s	27	5, 2, 6, 4, 5, 7, 3
32	16.7, CH3	1.00, m	9, 8, 6, 5	11, 10, 9
33a	75.7, CH2	4.33, m	10, 11	16, 33, 1', 14
33b	75.7, CH2	3.75, d (10.30)	16, 33	16, 1', 33, 14
34a[f]	35.8, CH3	2.94, u	16, 33, 13	
34b[f]	33.3, CH3	2.87, s	25	27, 26
1'	105.0, CH	4.28, m	27	2', 5', 3', 33
2'	75.2, CH	3.13, dd (9.23, 7.78)	2'	3', 1'
3'	77.9, CH	3.34, m	3', 1'	2', 5', 4', 1'
4'	71.7, CH	3.21, m	2', 4'	5', 3', 6'
5'	78.2, CH	3.24, m	5', 3'	4', 3', 6', 1'
6'	62.9, CH2	3.83, br dd (11.94, 1.95)	4', 6'	4'
6'	62.9, CH2	3.59, dd (11.98, 5.80)	5', 6'	4', 5'

[a]Recorded in chloroform-*d*. [b]Acquired at 150 MHz, adjusted to the solvent signal of chloroform-*d* (δ_H 77.00 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of chloroform-*d* (δ_H 7.27 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1] [g]Carbons showing the same signal. [h]No coupling constant analysis due to signal overlap.

Table S13. NMR Spectroscopic Data for Ajudazol J (9)[a]



Position	δ_C[b], type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
1	169.9, C			6, 47
2	105.8, C			31, 6, 5, 47
3	160.1, C			31, 6, 5, 47
4	125.5, C			31, 6, 47
5	137.5, CH	7.42, br d (7.45)	31, 6	31, 6, 47, 8, 2, 7, 3
6	114.4, CH	7.13, br d (7.71)	8, 5	31, 5, 8, 2, 4, 3, 1
7	140.8, C			9, 8, 5
8	63.8, CH	4.97, br d (10.79)	9, 48, 6	32, 9, 6, 5, 10, 7
9	86.7, CH	4.26, br d (10.92)	8	32, 11, 8, 10, 7
10	31.2, CH	2.55, br s	32	32, 11, 9, 8, 13
11a	25.4, CH2	2.80, br d (14.13)	11, 13	32, 10, 9, 13, 12
11b	25.4, CH2	2.48, br d (13.74)	11	32, 10, 9, 13, 12
12	138.1, C			11, 13
13	135.5, CH	7.45, s	11	11, 10, 12, 14
14	164.8, C			16, 33, 13
15	72.5, C			16, 33, 18
16	34.2, CH2	2.89, s	18	33, 18, 15, 18, 14
17	125.1, CH	5.38, m	20, 18	
18	127.6, CH	6.41, br d (11.69)	16, 17, 19	16, 19, 15, 20

19	123.4, CH	6.25, br d (5.01)	22, 20, 18	22, 18
20	133.1, CH	5.53, m	21, 22, 17, 19	18
21	31.9, CH2	2.18, m	22, 20	22, 23
22	26.9, CH2	2.28, m	21, 20, 19	21, 19, 23
23	133.4, CH	5.57, br d (11.56)	24	21, 22, 25
24	123.2, CH	5.34, br dd (15.48, 5.84)	34, 23	34
25a	48.7, CH2	4.02, br dd (14.32, 6.49)	25	34, 23, 26
25b	48.7, CH2	3.91, m	25	34, 23, 26
25c	52.1, CH2	3.85, m	21	34, 20, 22, 17
26	168.3, C			30, 34, 25, 27
27	117.8, CH	5.78, m	29	29, 30, 34, 28, 26
28	147.0, C			29, 30, 34, 27
29	26.3, CH3	1.84, m	27	30, 27, 28
30	20.2, CH3	1.92, s		29, 27, 28, 26
31	15.5, CH3	2.27, s	5	5, 47, 2, 6, 4, 3
32	16.3, CH3	1.07, br d (6.81)	10	11, 9, 11, 10, 8
33a	69.2, CH2	4.13, br d (10.02)	33	16, 1', 14
33b	69.2, CH2	3.48, br dd (18.82, 9.83)	33	16, 1', 15, 14
34a	34.9, CH	2.92, s		25
34b	32.5, CH3	2.90, m	24	25, 27, 24, 28, 26
1'	107.2, CH	5.10, s	2', 3'	33, 2', 3', 4'
2', 3'	78.5, CH	4.10, br s	39, 1', 38	5', 4', 1'
4'	87.9, CH	4.18, br s	5'	1', 5', 2', 3'
5'a	61.8, CH2	3.88, m	40, 4'	2', 3', 4'
5'b	61.8, CH2	3.82, m		2', 3', 4'
38	OH	6.06, m	2', 3'	
3‘-OH	OH	4.46, br d (8.35)	2', 3'	
5‘-OH	OH	3.38, m	5'	
3-OH	OH	11.08, s		31, 2, 4, 5, 3, 1
8-OH	OH	5.84, br s	8	

[a]Recorded in methanol-d_4. [b]Acquired at 150 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 49.15 ppm). [c]Acquired at 500 MHz, adjusted to the solvent signal of methanol-d_4 (δ_H 3.31 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons and carbons showing HMBC correlations to indicated protons. [f]Due to the amide substructure signal doubling is observed.[1] [g]Carbons showing the same signal.

Table S14. Minimum inhibitory concentration (MIC, µg/ml) of ajudazol C-J against common microbial pathogens and cancer cell lines

Test organisms	C	D	E	F	G	H	I	J
Staphylococcus aureus	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Mycobacterium smegmatis	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Bacillus subtilis	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Citrobacter freundii	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Acinetobacter baumannii	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Candida albicans	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
***Escherichia coli* acrB**	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
***Escherichia coli* DSM1116**	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Cryptococcus neoformans	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Pseudomonas aeruginose	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
Pichia anomala	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64	≥ 64
HepG2	0.51	2.38	0.1	NOT DONE	0.012	3.48	36.69	NOT DONE
HCT-116	0.29	0.99	0.13	1.24	0.018	1.93	46.28	18.22
KB3.1	1.25	2.76	1.03	7.75	1.41	4.11	37.05	51.24

Supplementary figures

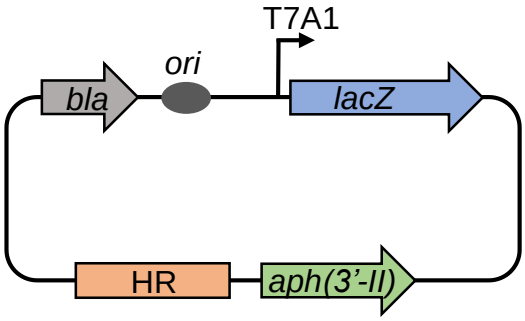
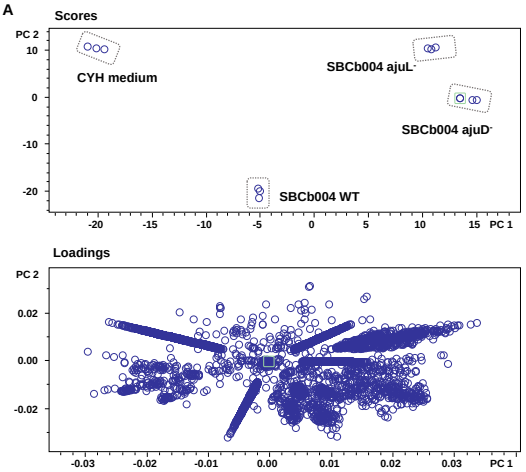


Figure S1. Schematic representation of plasmid pbluekan-HR used to perform insertional gene inactivation. *aph(3'-II)*: kanamycin resistance gene. *bla*: β-lactam resistance cassette. HR: homologous region amplified from target gene. *ori*: origin of replication of *E. coli* plasmid pMB1. The angled arrow indicates the T7A1 transcriptional promoter. *lacZ*: β-galactosidase gene.



B

Molecular features	P-value	Max. intensity	WT count (3)	Media count (3)	delta ajuD count (3)	delta ajuL count (3)
10.32 min :755.374m/z	n.a.	5871898	3	0	0	0
14.11 min : 575.315m/z	n.a.	5201352	3	0	0	0
10.14 min : 741.359m/z	n.a.	4471474	3	0	0	0
12.25min : 707.354m/z	n.a.	1894118	3	0	0	0
12.02 min : 593.322m/z	n.a.	1757720	3	0	0	0
9.77 min : 771.370m/z	n.a.	1637718	3	0	0	0
13.86 min : 577.326m/z	n.a.	1046710	3	0	0	0
11.45min : 737.364m/z	n.a.	913102	3	0	0	0
12.57min : 591.306m/z	n.a.	882834	3	0	0	0
11.19min : 723.348m/z	n.a.	786580	2	0	0	0
10.84min : 609.315m/z	n.a.	267294	3	0	0	0
10.20min : 753.357m/z	n.a.	196414	2	0	0	0
5.05min : 269.211m/z	n.a.	524536	3	0	3	3
7.42min : 249.184m/z	n.a.	160746	3	0	3	3
3.19min : 513.414m/z	n.a.	141596	3	0	3	3
9.04min : 603.333m/z	n.a.	136404	3	0	3	3
8.41min : 235.205m/z	n.a.	130962	3	0	3	3
6.04min : 515.372m/z	n.a.	124768	3	0	3	3

Figure S2. PCA and T-test models. A PCA for SBCb004 ajuL$^-$ vs. SBCb004 ajuD$^-$ vs. SBCb004 WT vs. CYH medium. **B** T-test for datasets from A. n.a.: not available.

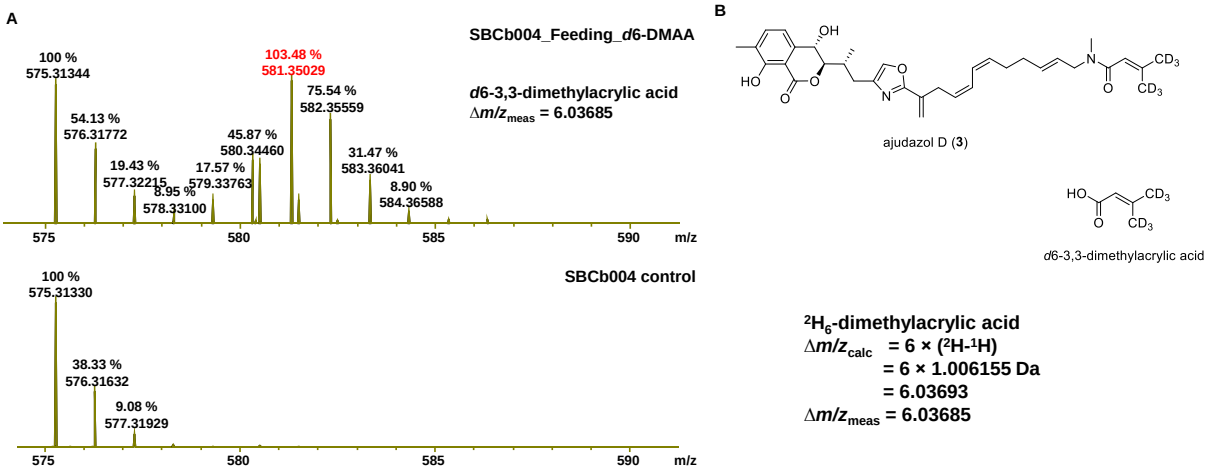


Figure S3. **Feeding of labelled dimethylacrylic acid. A** Isotopic peak pattern of the $[M+H]^+$ signals for ajudazol D (575 m/z) prove incorporation of labelled dimethylacrylic acid in SBCb004. The observed mass shifts fit the heavy isotopes that were incorporated. **B** Location of the deuterium label in the fed dimethylacrylic acid and likely incorporation pattern of dimethylacrylic acid skeleton into ajudazol D.

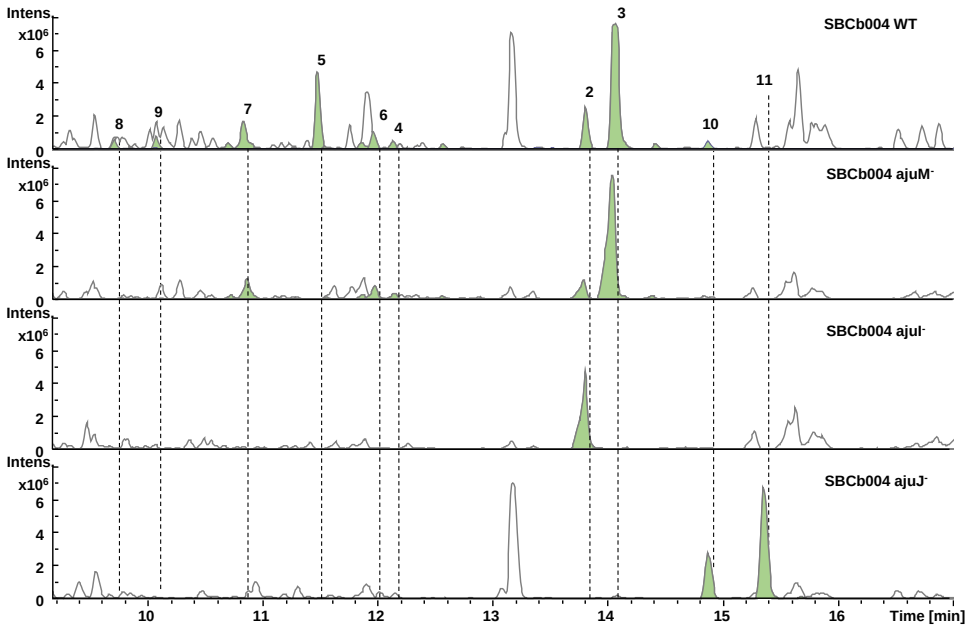
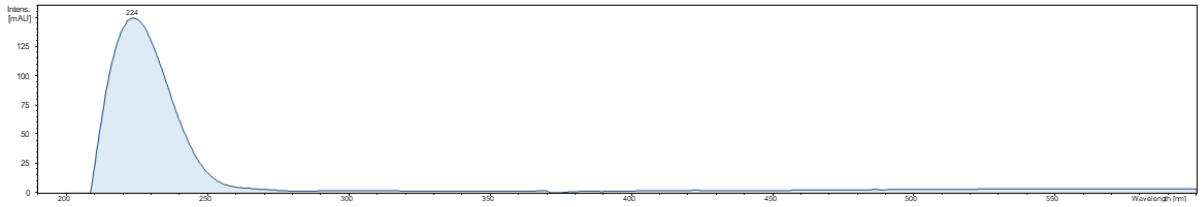


Figure S4. **HPLC-MS analysis of ajudazol production in SBCb004 wild type (WT), SBCb004 ajuM$^-$, SBCb004 ajuI$^-$ and SBCb004 ajuJ$^-$.** 2-9: ajudazol C-J. 10: deshydroxyajudazol C. 11: deshydroxyajudazol D.

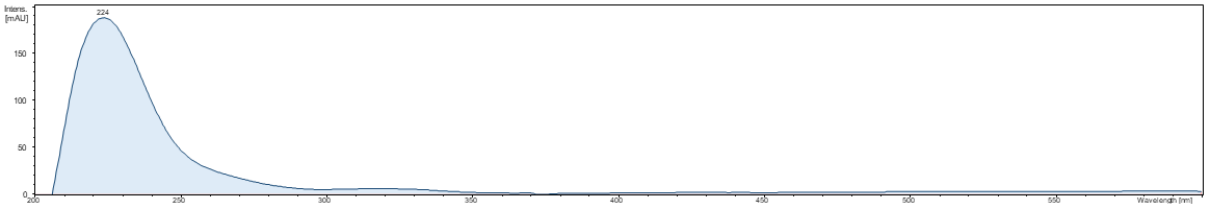
Figure S5. UV spectra

Ajudazol C – UV spectrum



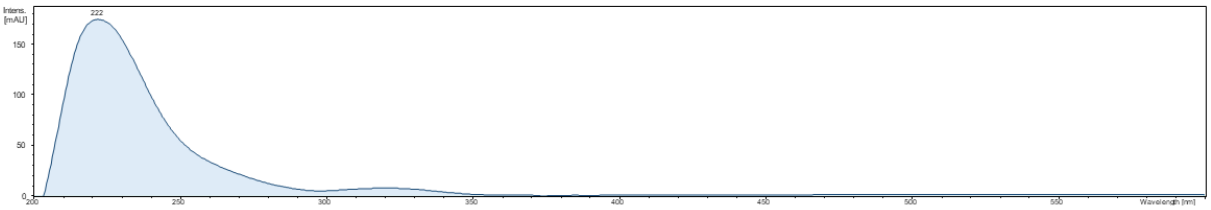
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 17.8:82.2), λ_{max} 224

Ajudazol D – UV spectrum



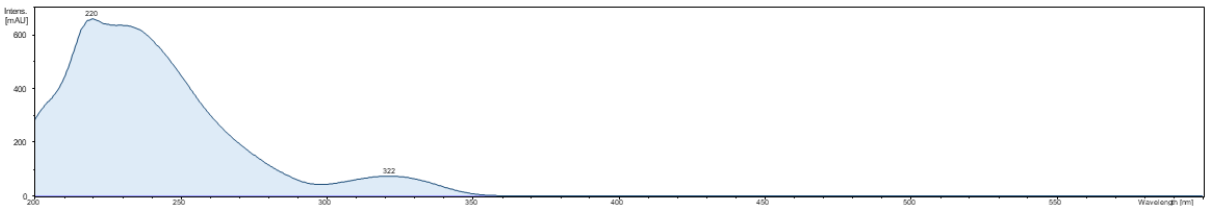
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 16.7:83.3), λ_{max} 224

Ajudazol E – UV spectrum



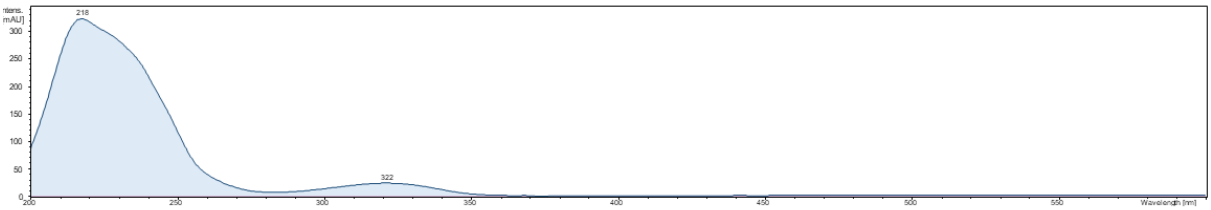
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 36.5:63.5), λ_{max} 222

Ajudazol F – UV spectrum



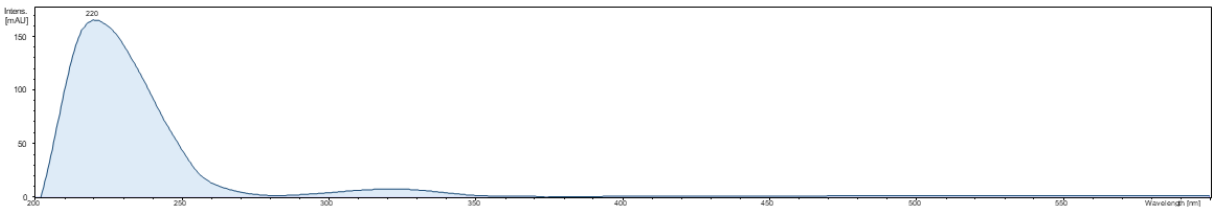
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 39.8:60.2), λ_{max} 220, 322

Ajudazol G – UV spectrum



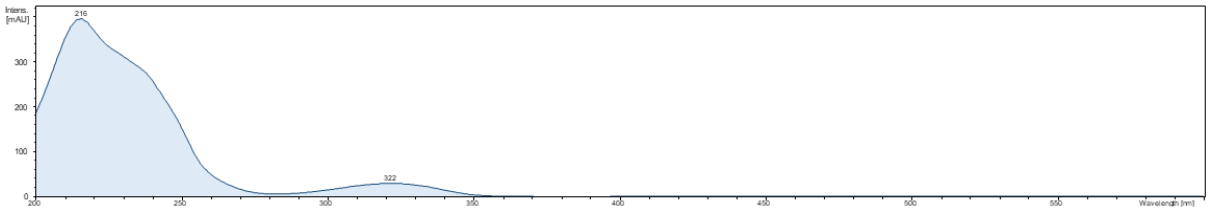
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 37.6:62.4), λ_{max} 218, 322

Ajudazol H – UV spectrum



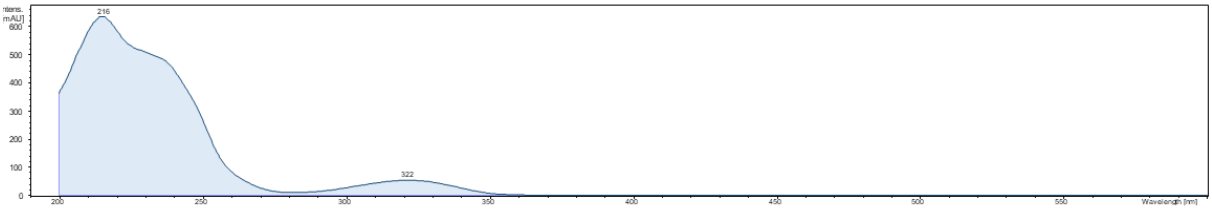
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 43.2:56.8), λ_{max} 220

Ajudazol I – UV spectrum



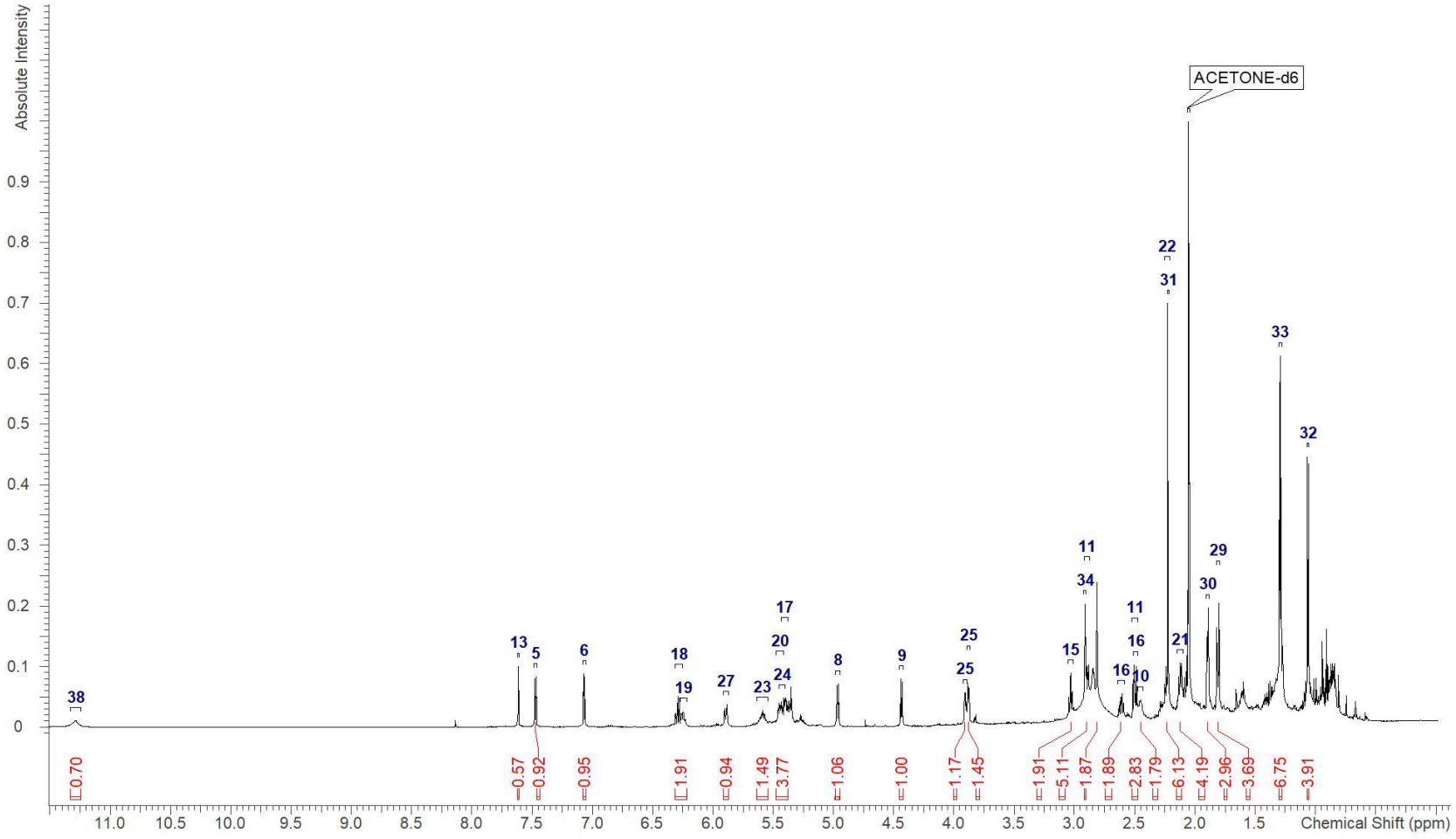
UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 48.6:51.4), λ_{max} 216, 322

Ajudazol J – UV spectrum

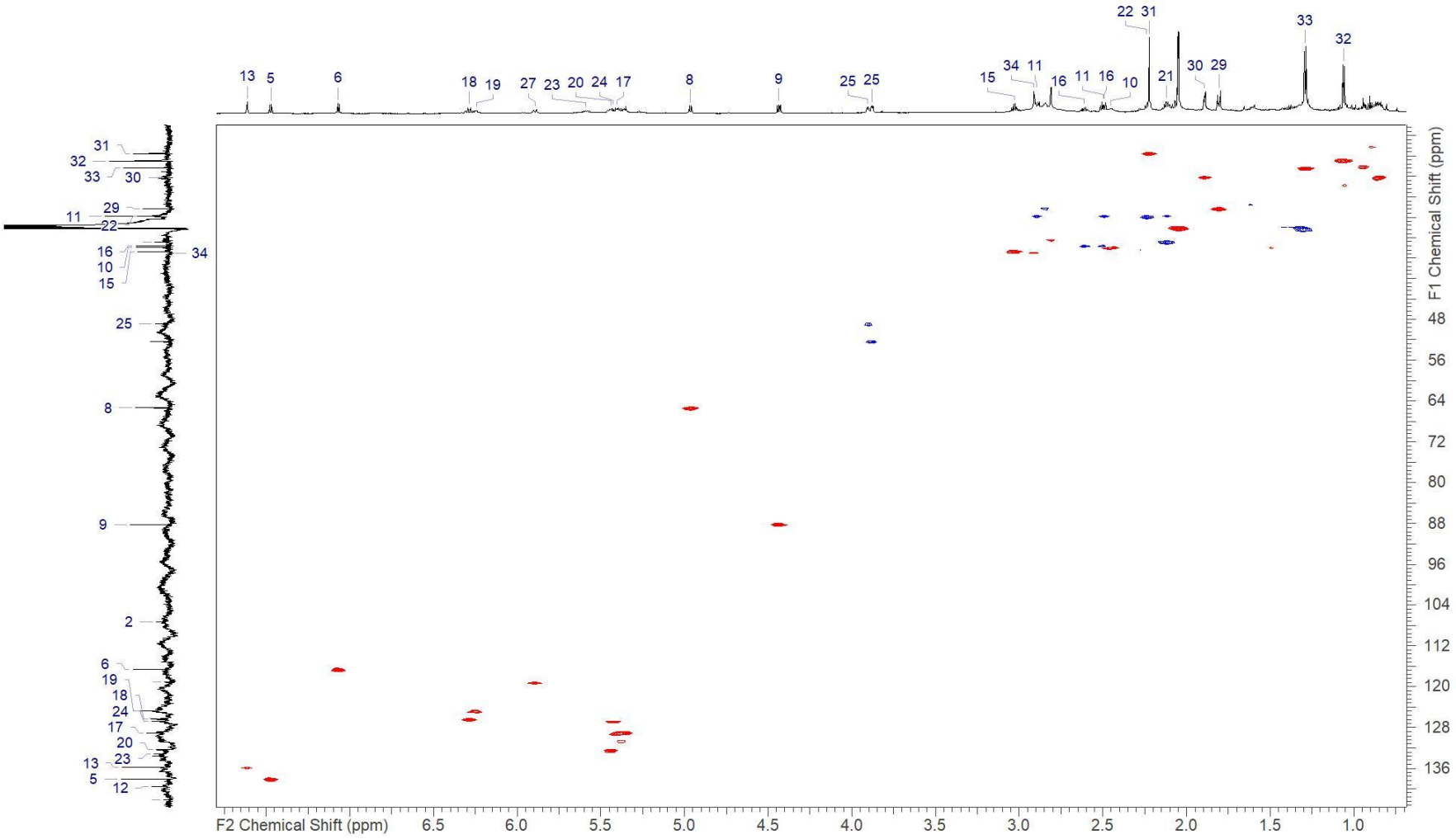


UV(H2O + 0.1% formic acid/ Acetonitrile + 0.1% formic acid, 46.7:53.3), λ_{max} 216, 322

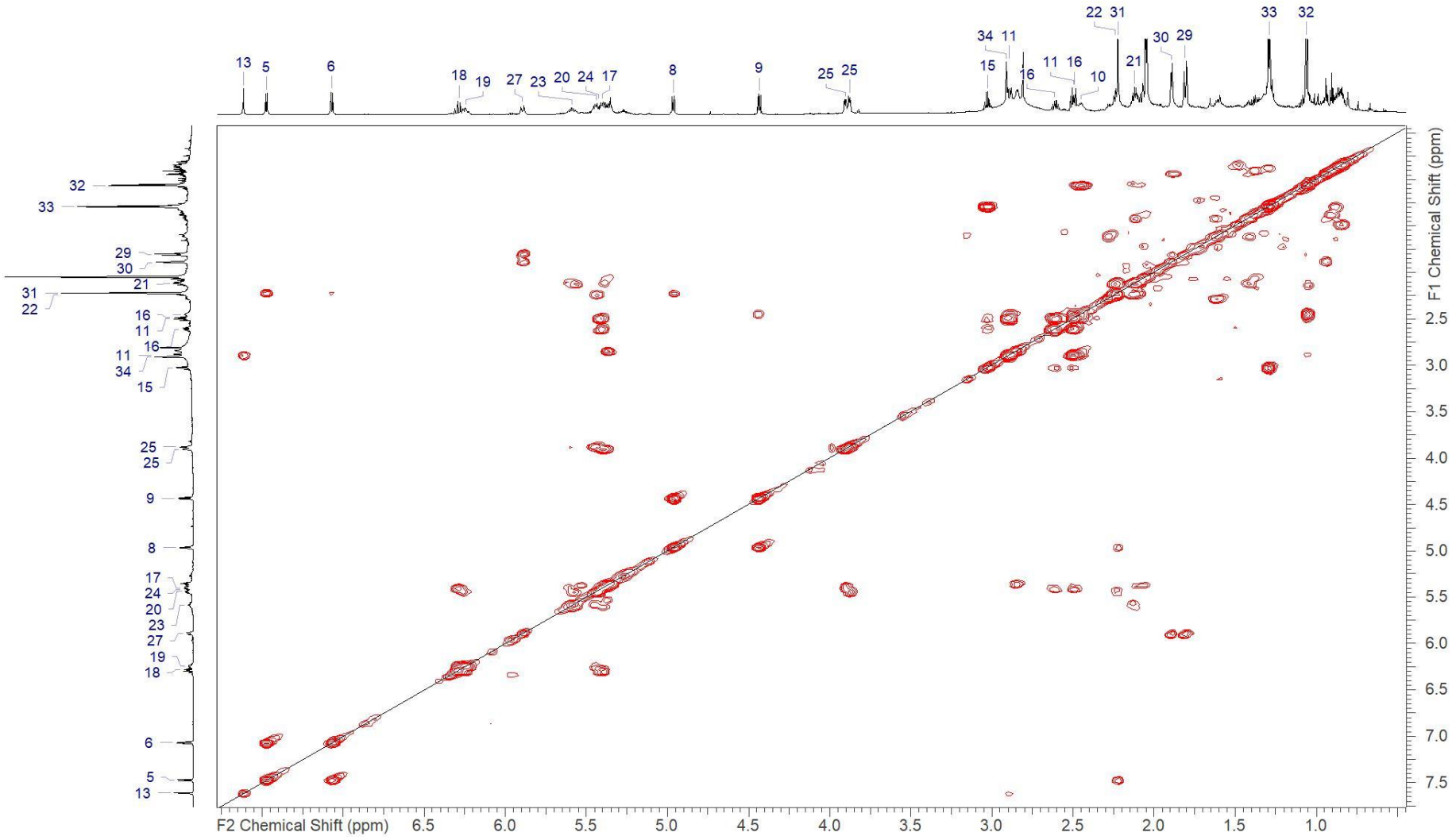
NMR spectra



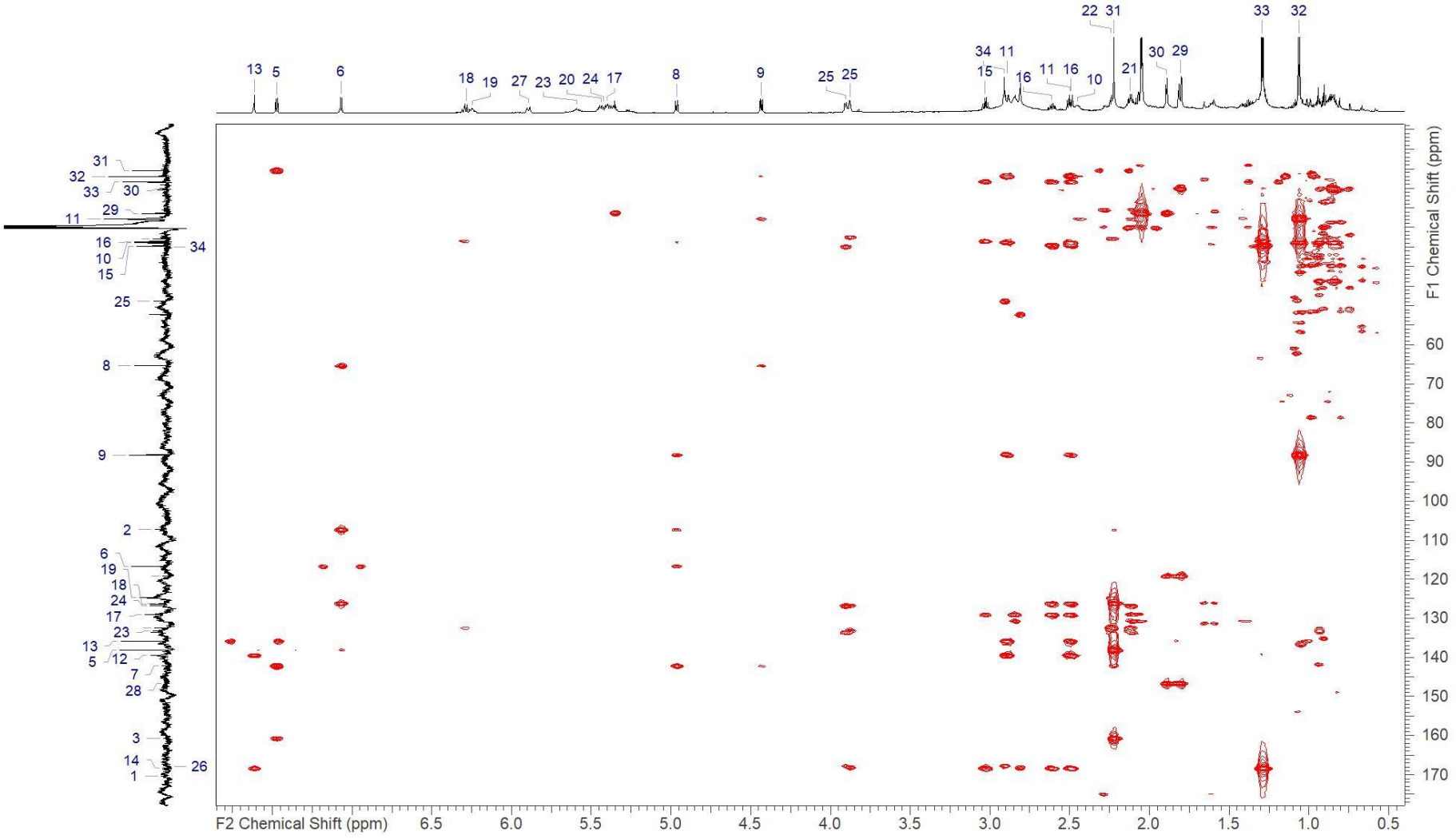
Ajudazol C (2) - ^{1}H NMR recorded in acetone-*d*6 at 700 MHz



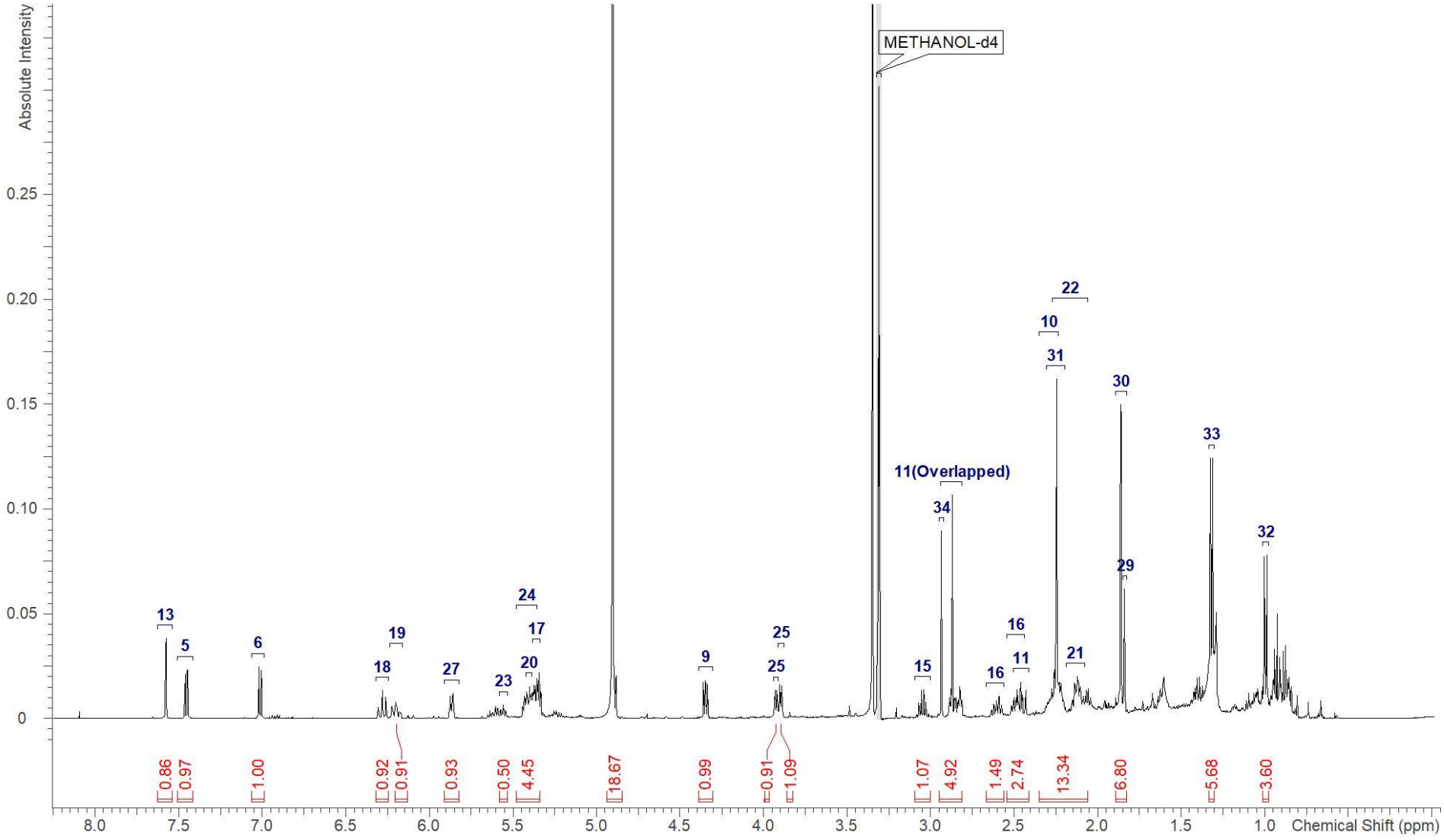
Ajudazol C (2) - HSQC NMR recorded in acetone-*d*6 at 175/700 (F1/F2) MHz



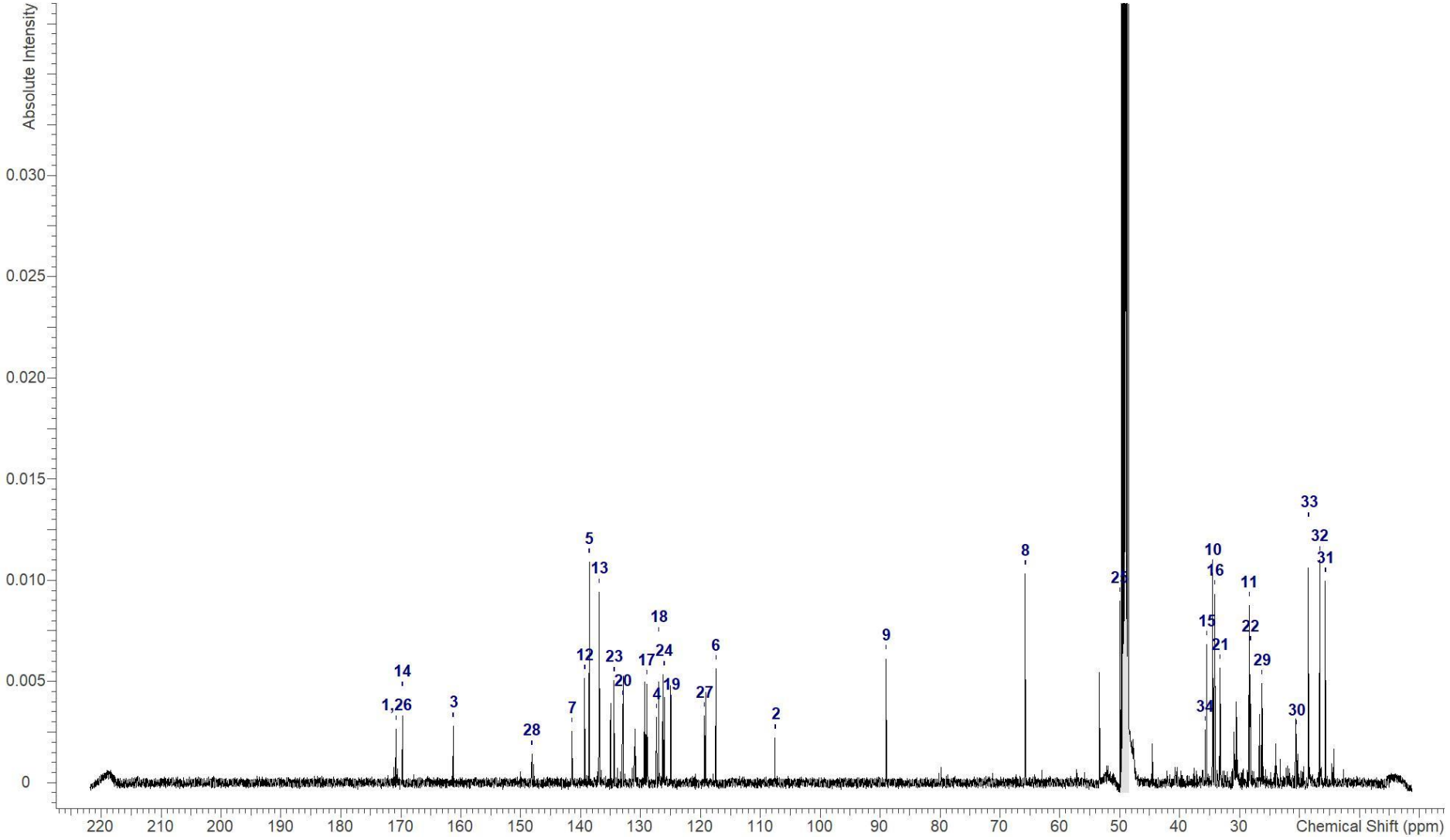
Ajudazol C (2) - COSY NMR recorded in acetone-*d*6 at 700 MHz



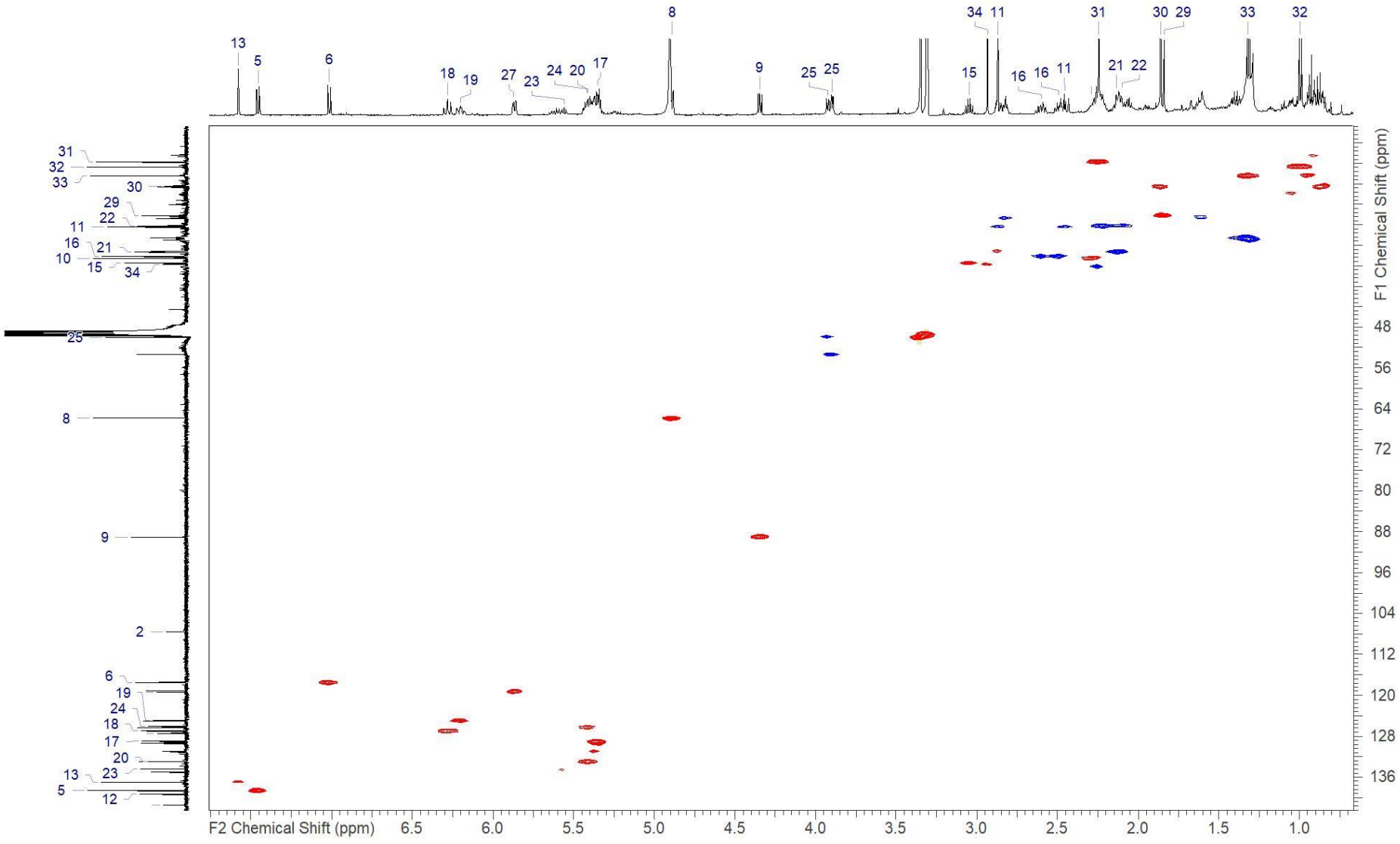
Ajudazol C (2) - HMBC NMR recorded in acetone-*d*6 at 700/175 (F1/F2) MHz



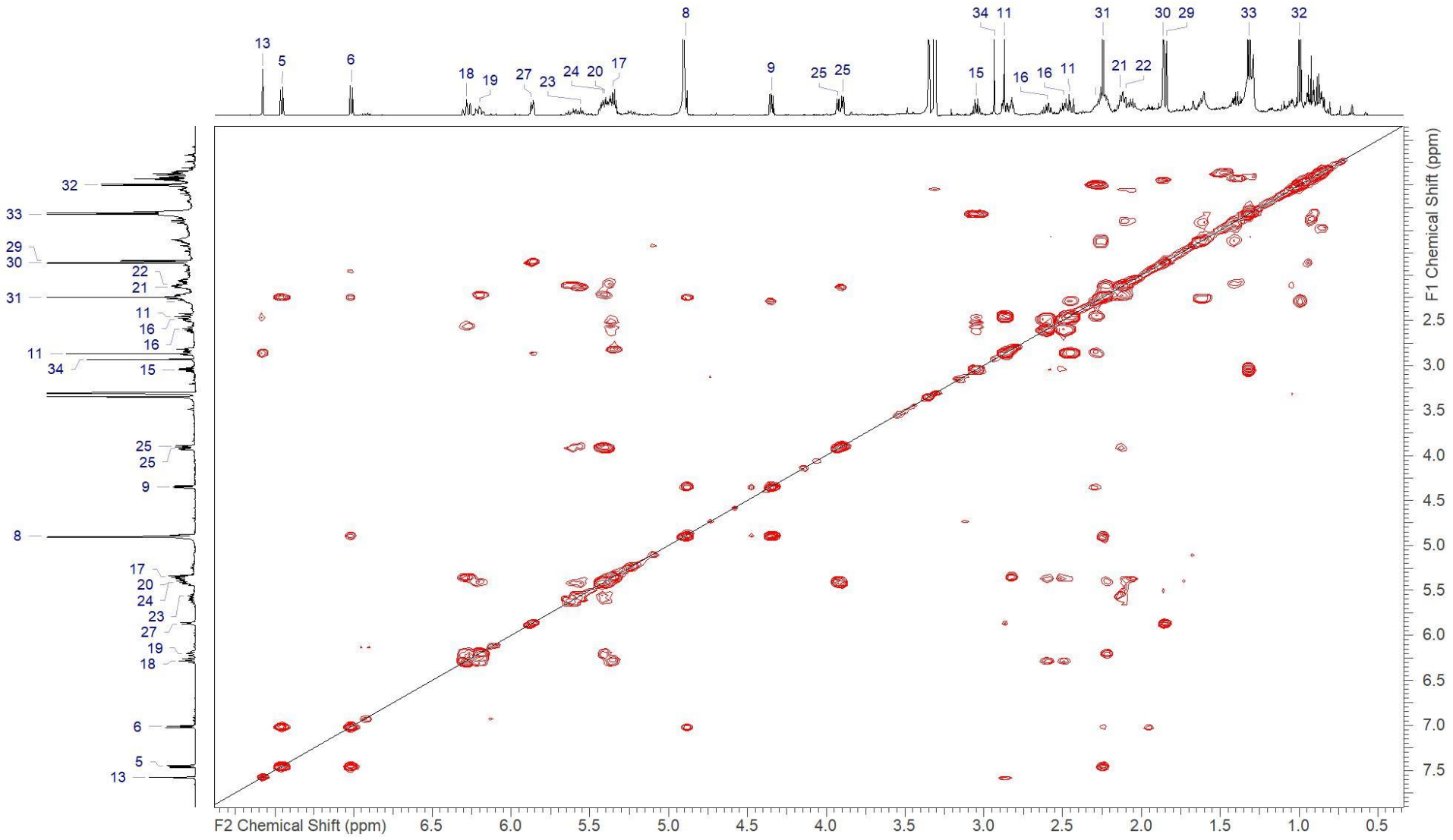
Ajudazol C (2) - ^{1}H NMR recorded in methanol-*d*4 at 500 MHz



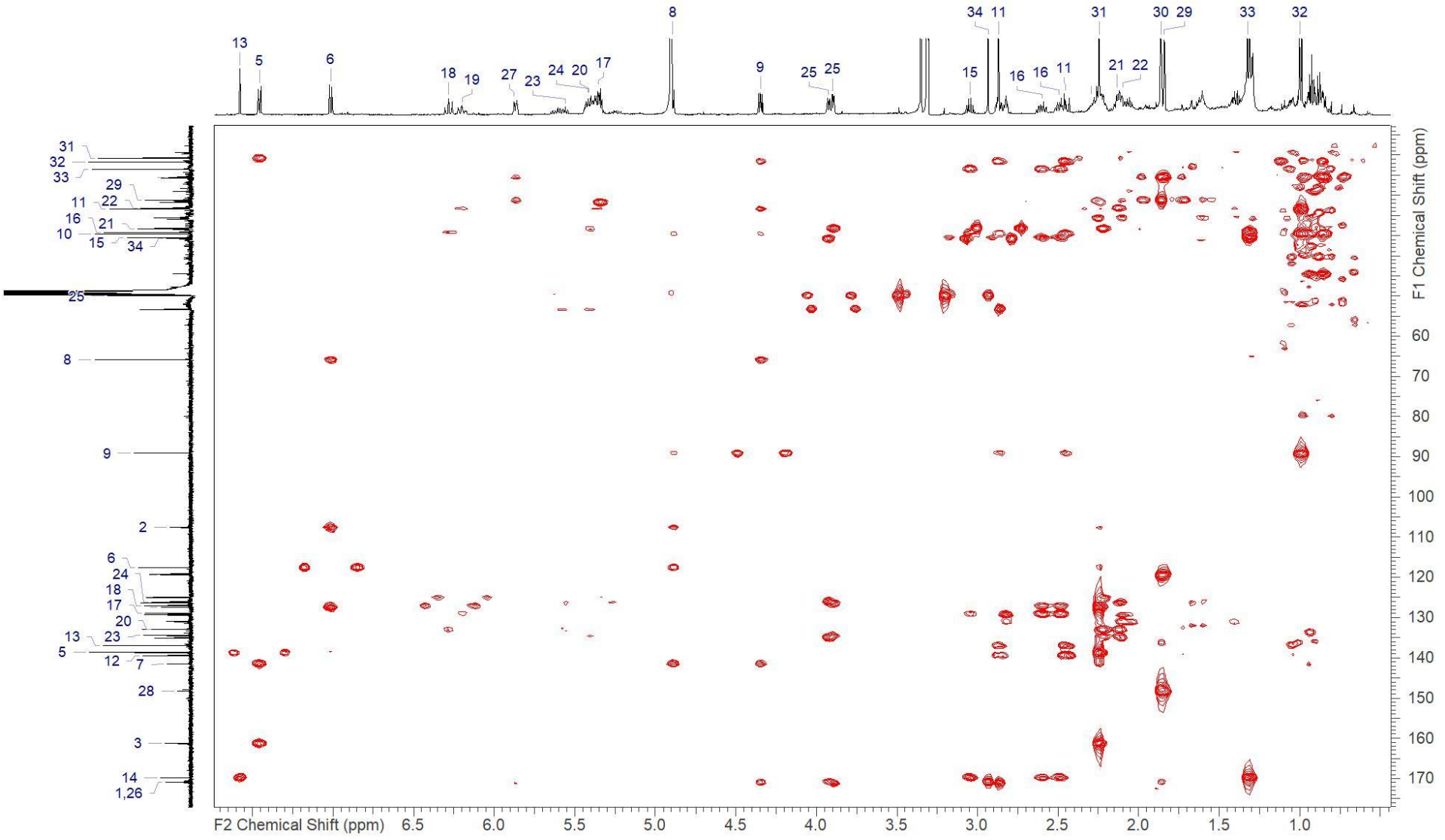
Ajudazol C (2) –^{13}C NMR recorded in methanol-*d*4 at 150 MHz



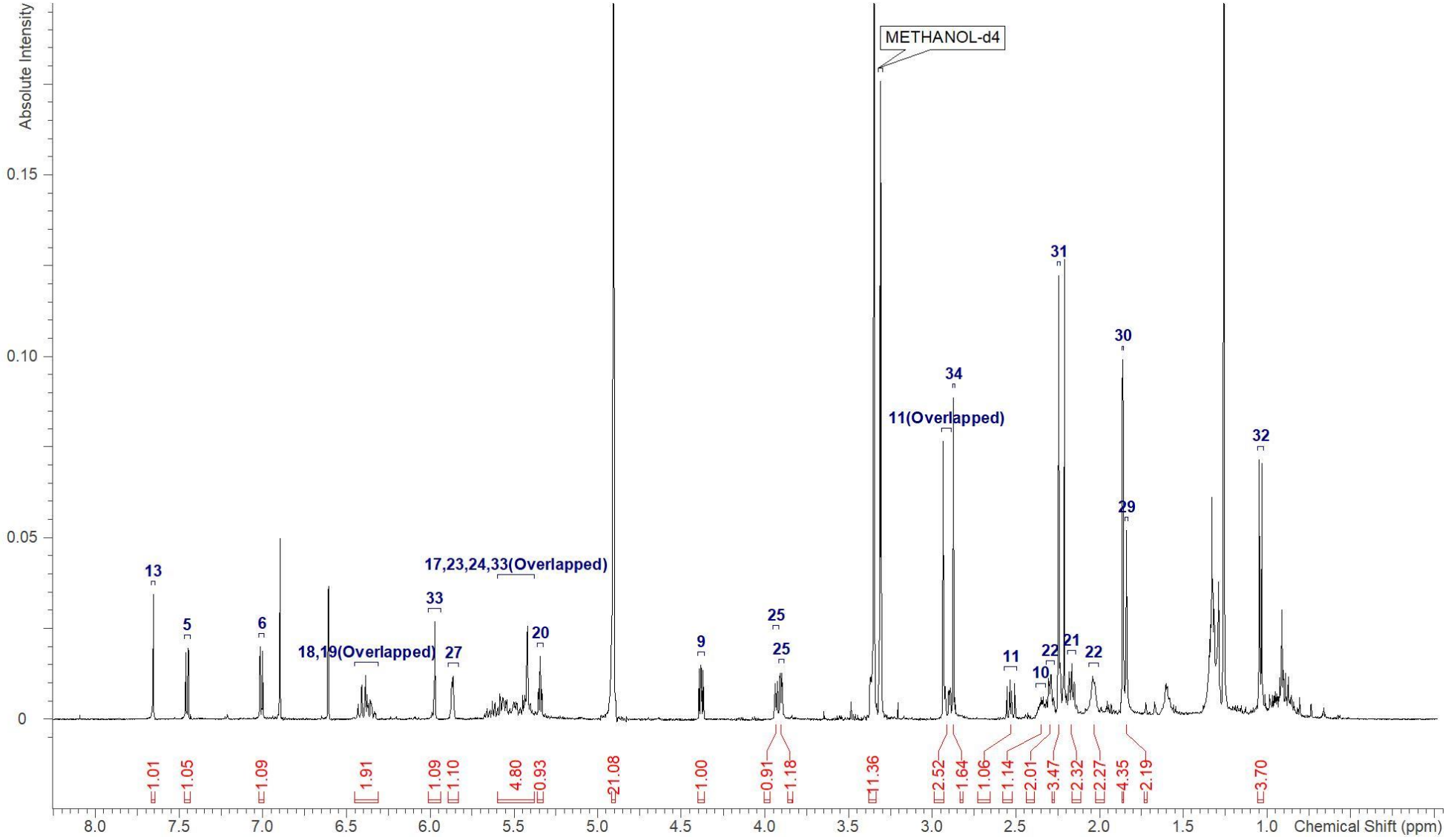
Ajudazol C (2) - HSQC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



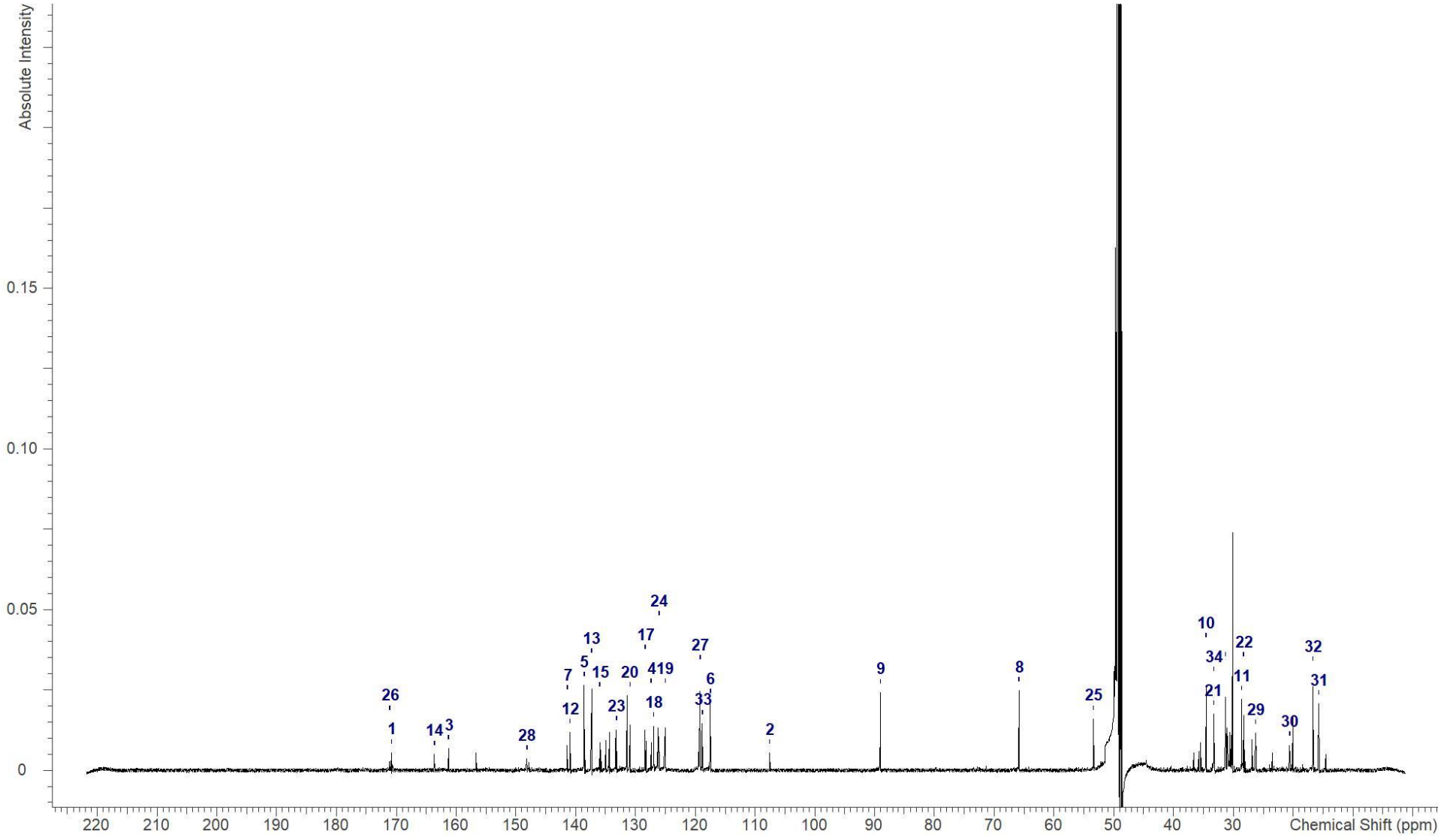
Ajudazol C (2) - COSY NMR recorded in methanol-*d*4 at 500 MHz



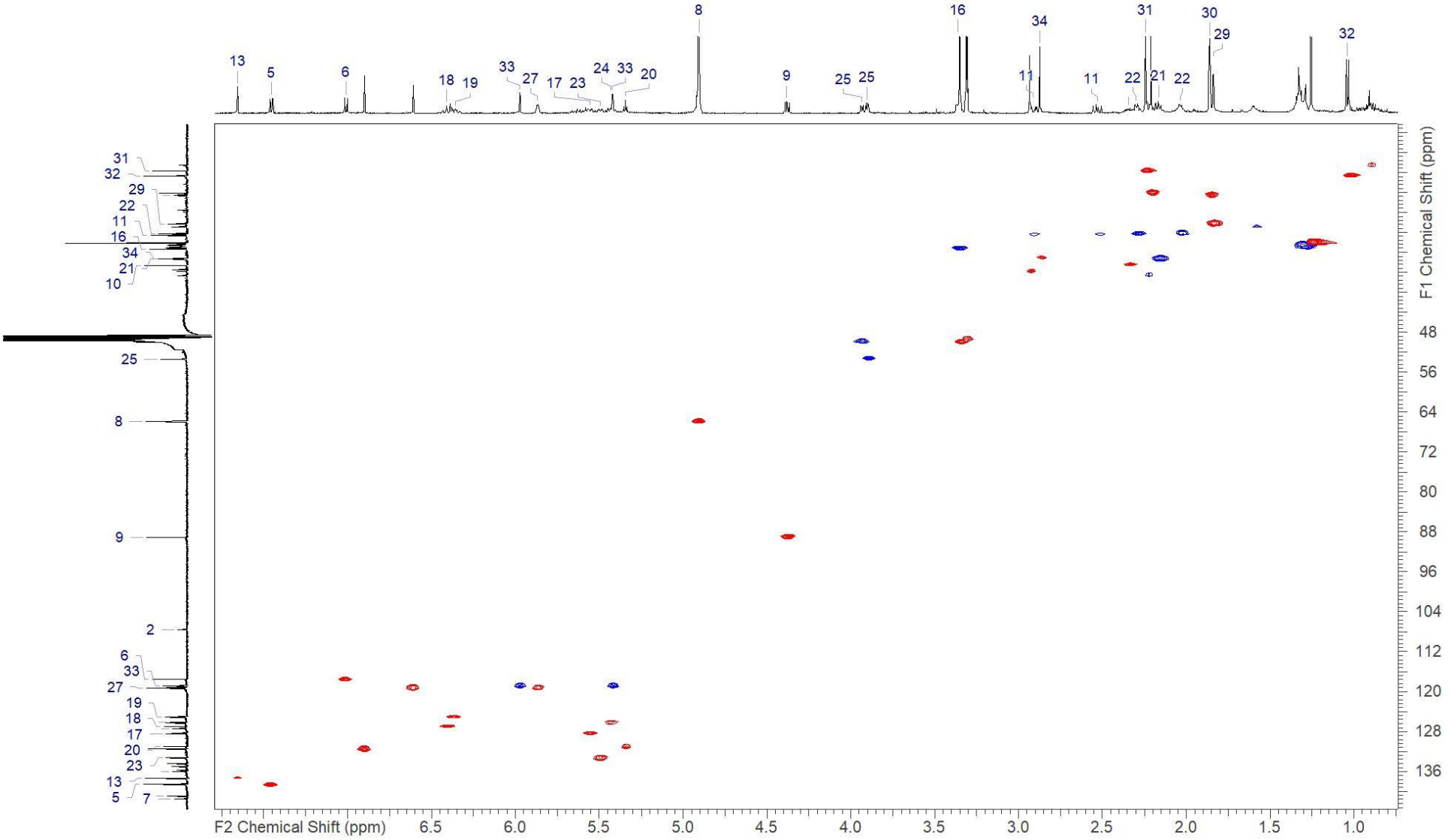
Ajudazol C (2) - HMBC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



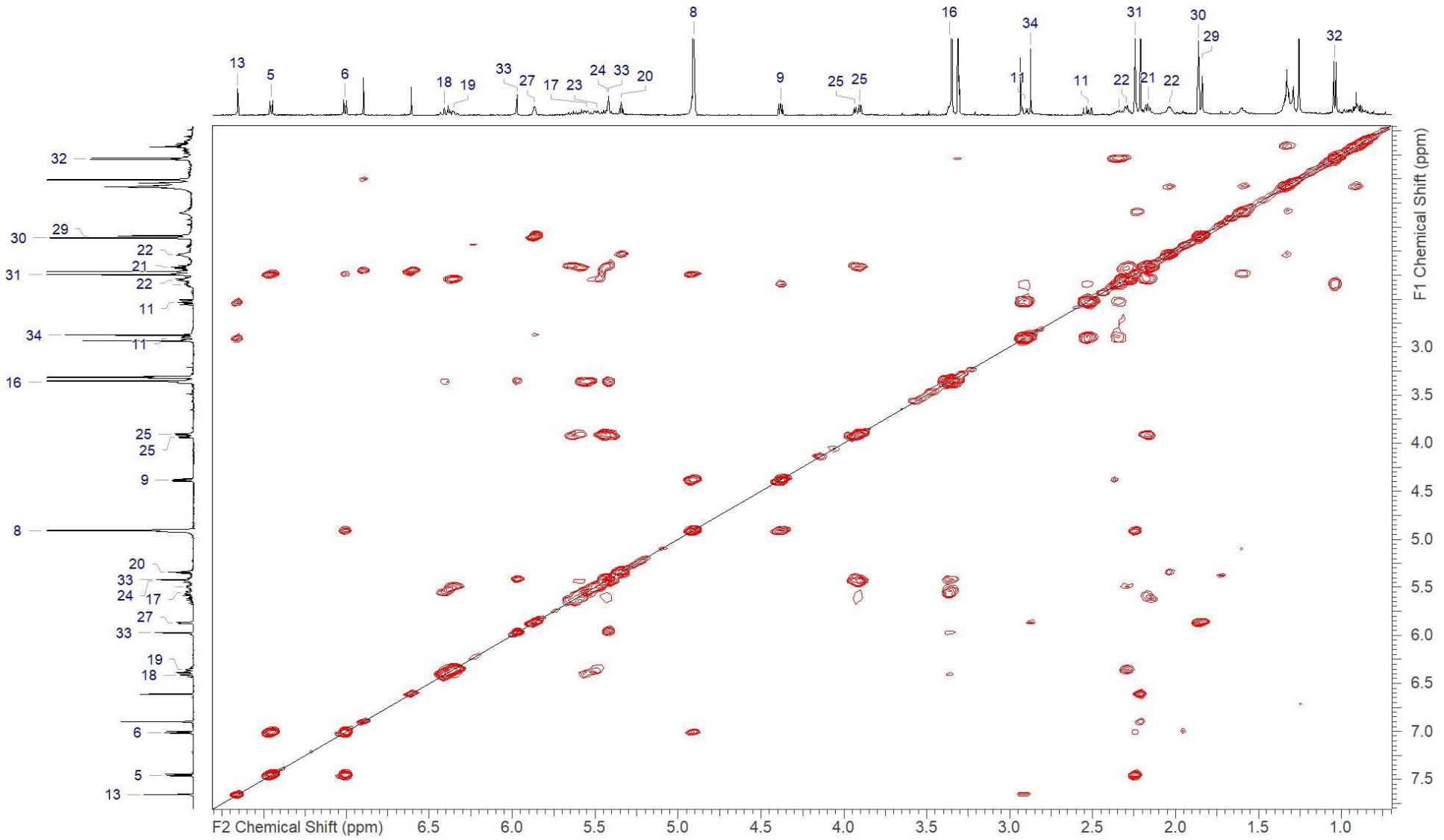
Ajudazol D (3) – ^{1}H NMR recorded in methanol-*d*4 at 500 MHz



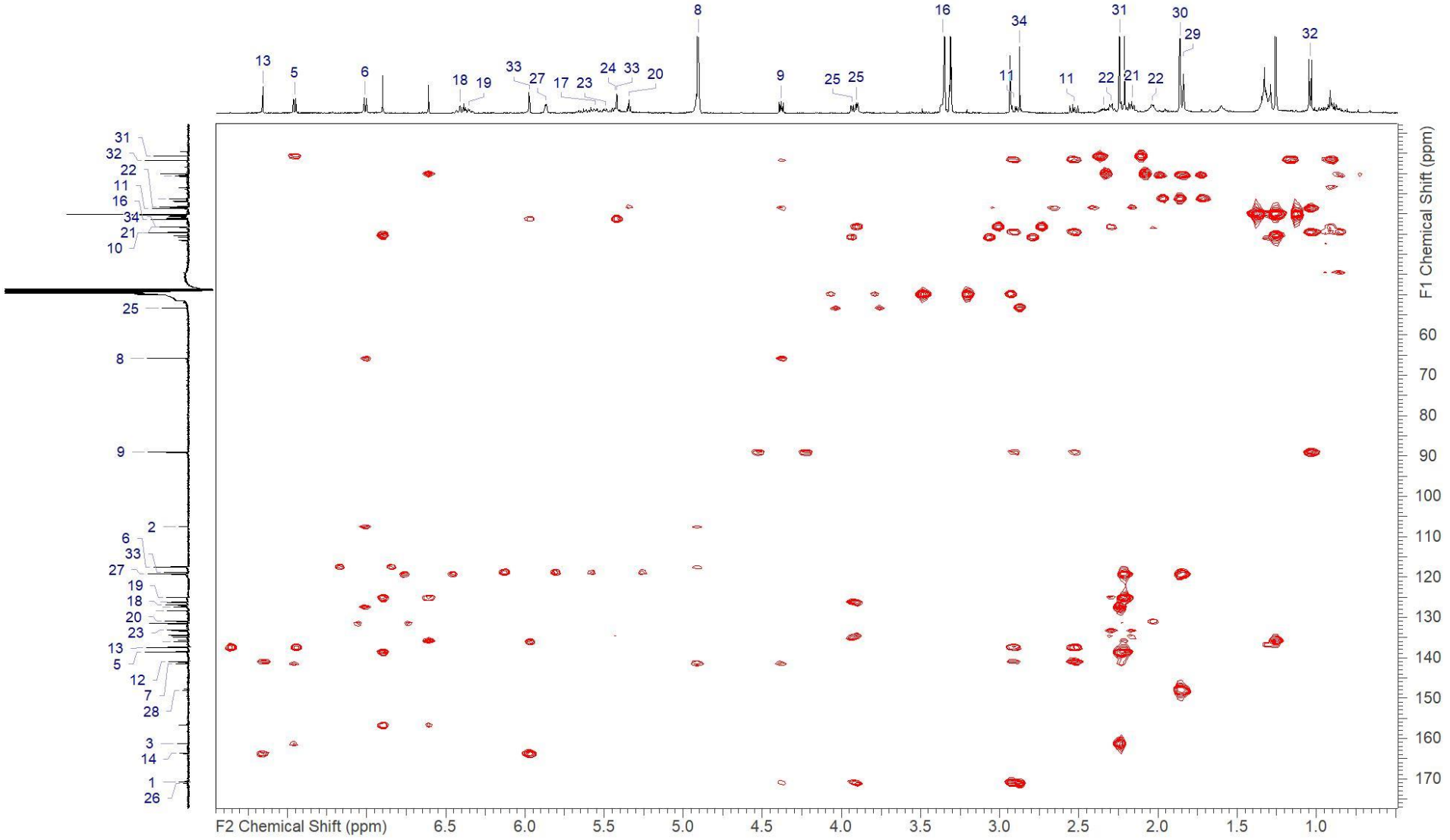
Ajudazol D (3) – ^{13}C NMR recorded in methanol-*d*4 at 150 MHz



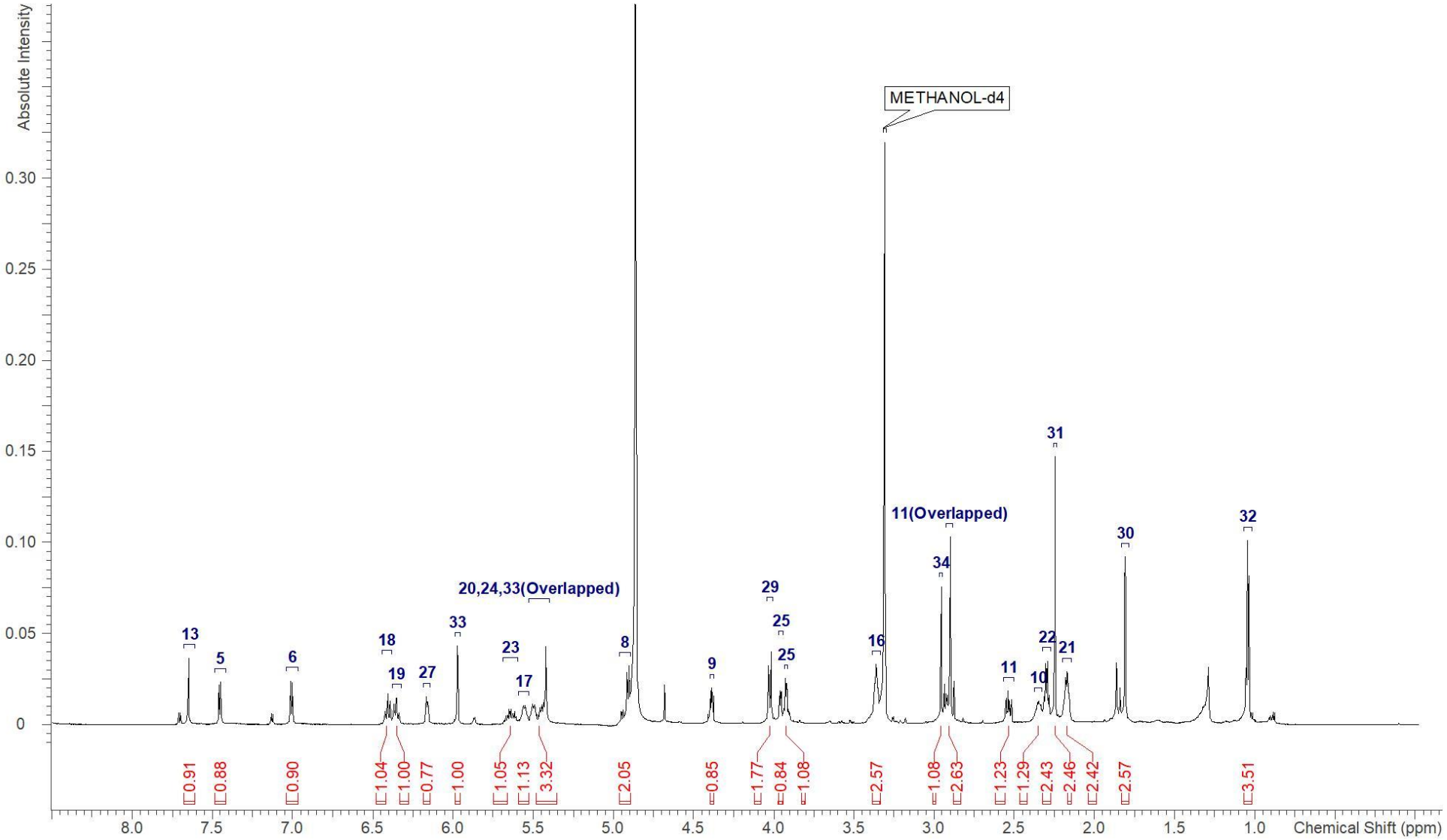
Ajudazol D (3) – HSQC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



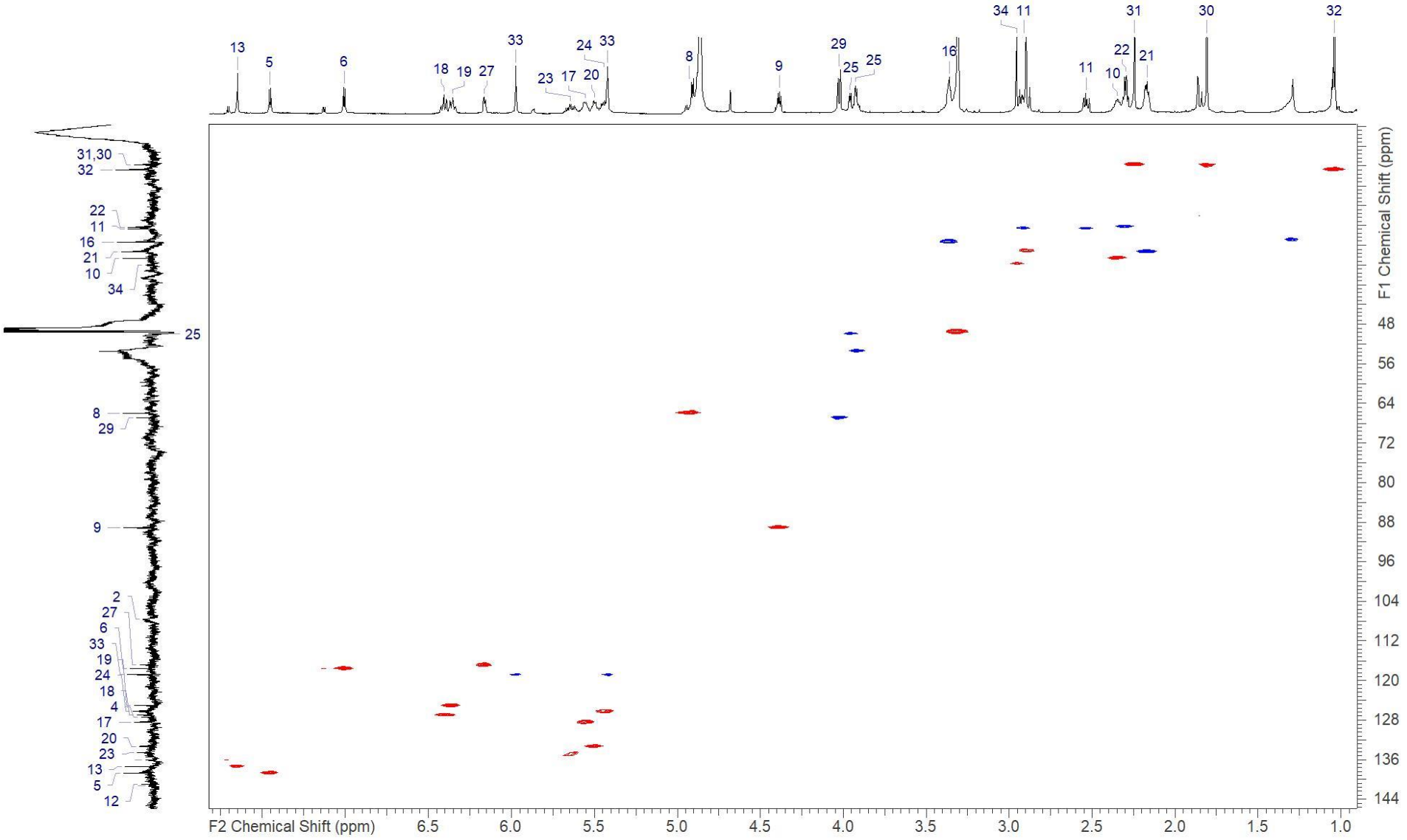
Ajudazol D (3) – COSY NMR recorded in methanol-*d*4 at 500 MHz



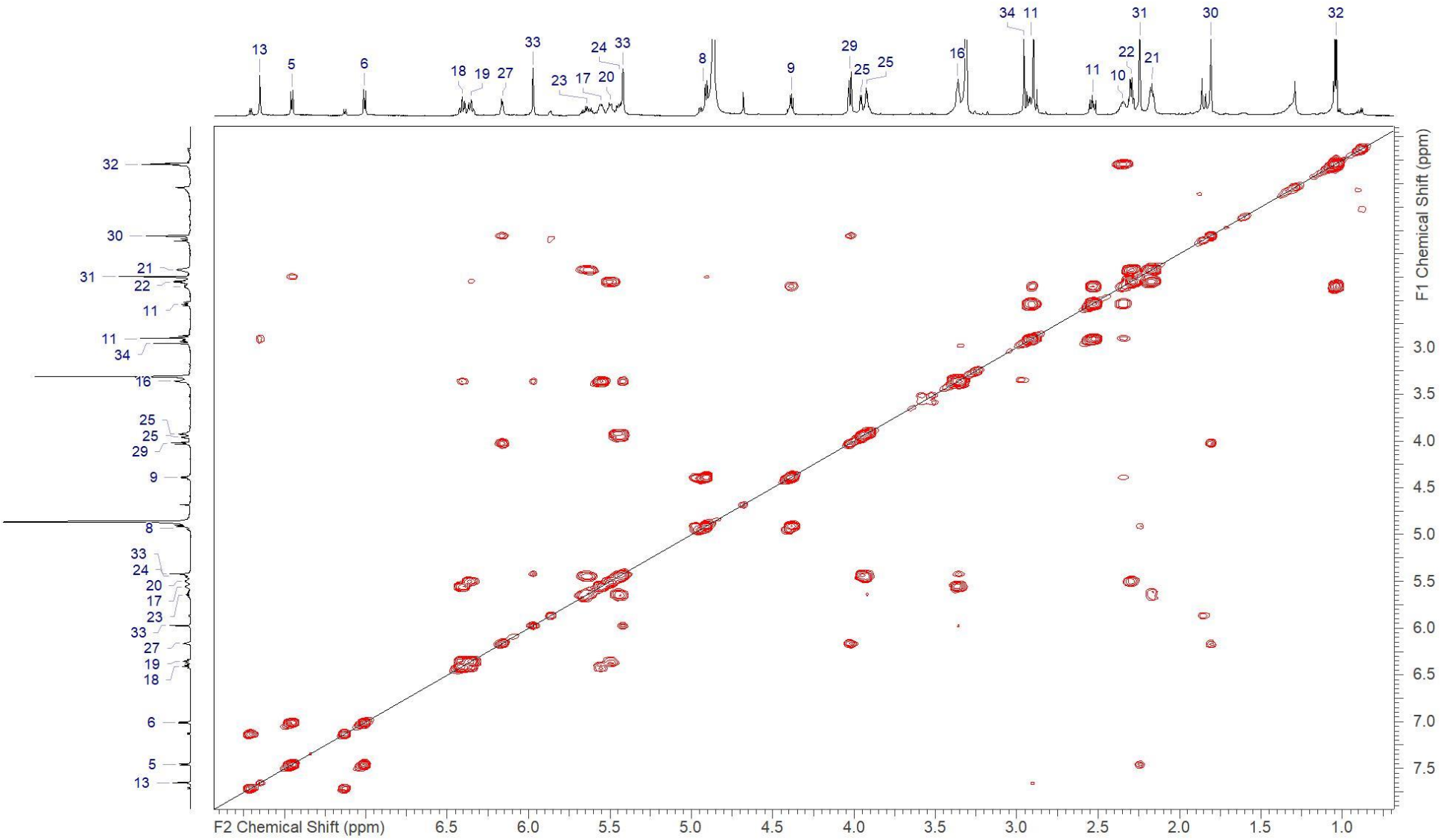
Ajudazol D (3) – HMBC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



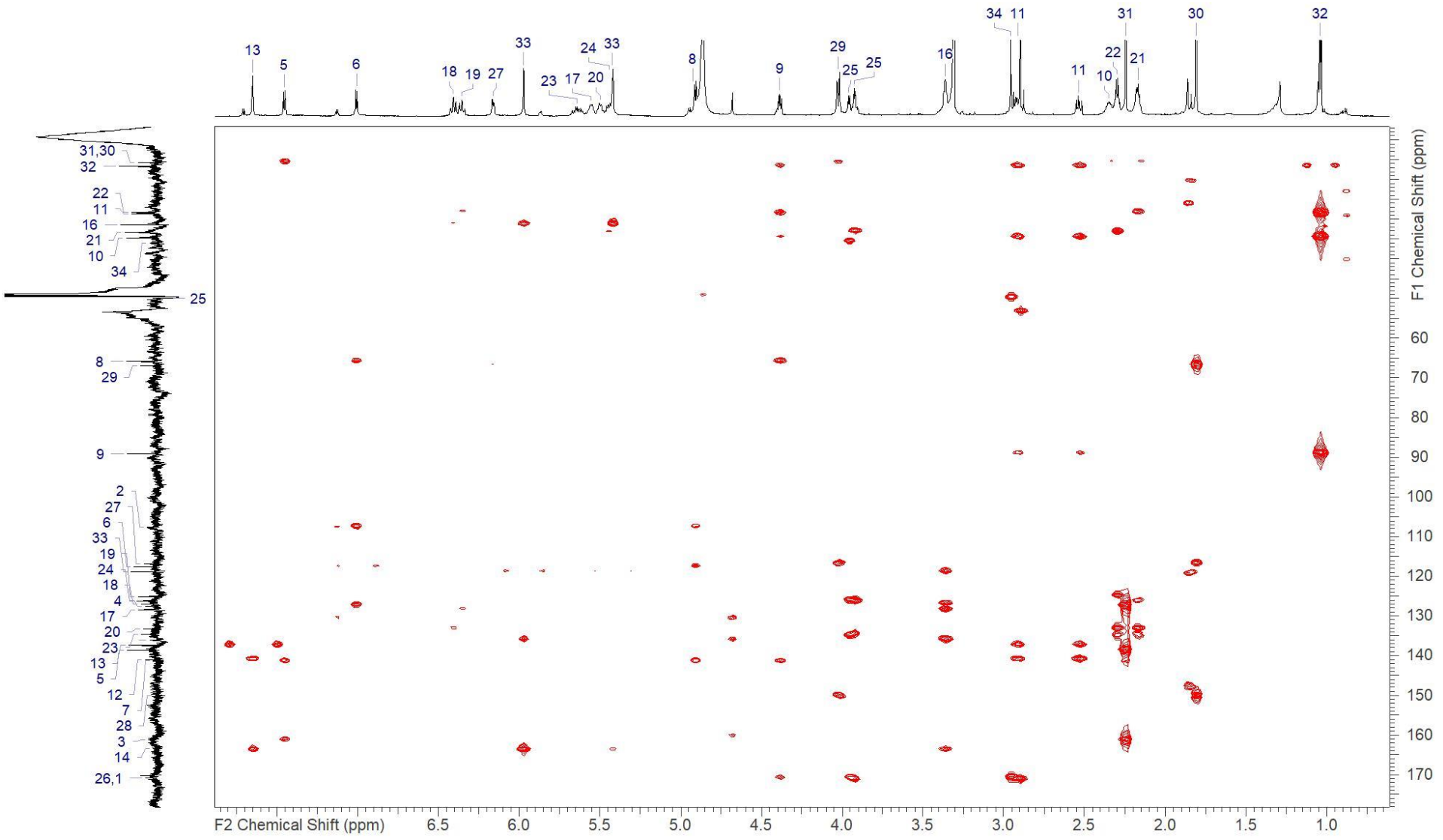
Ajudazol E (4) - ^{1}H NMR recorded in methanol-*d*4 at 700 MHz



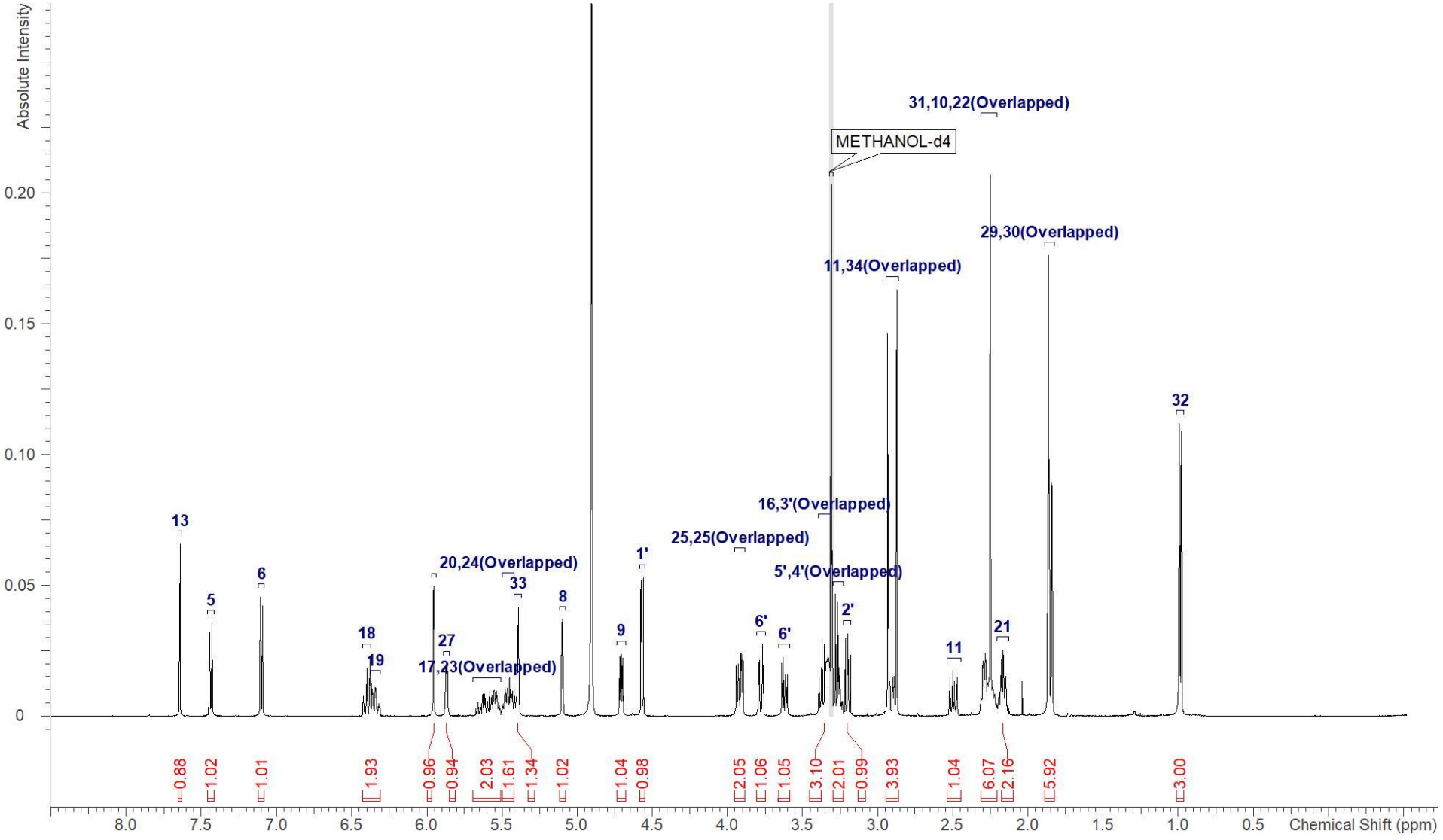
Ajudazol E (4) – HSQC NMR recorded in methanol-*d*4 at 175/700 (F1/F2) MHz



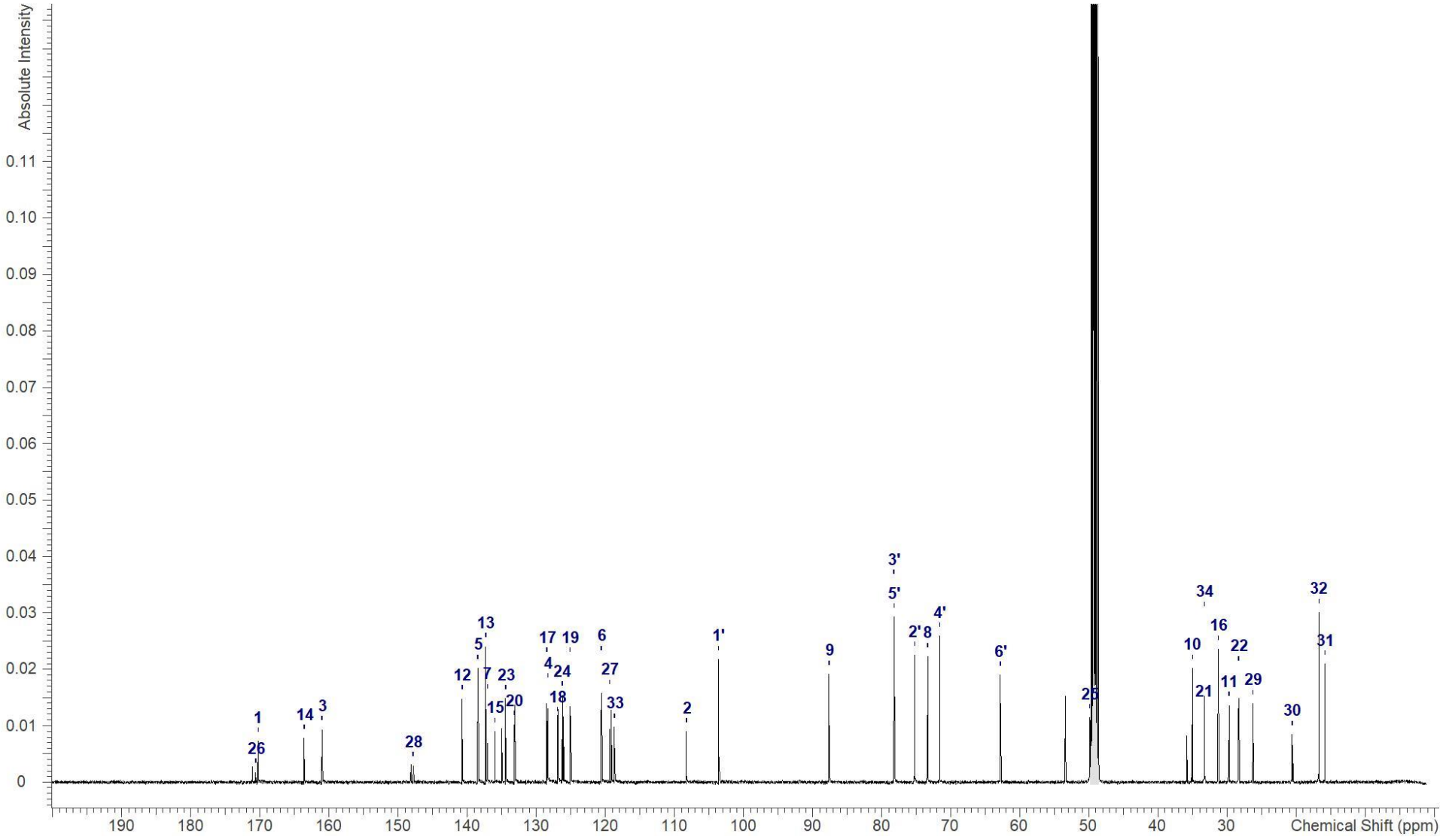
Ajudazol E (4) – COSY NMR recorded in methanol-*d*4 at 700 MHz



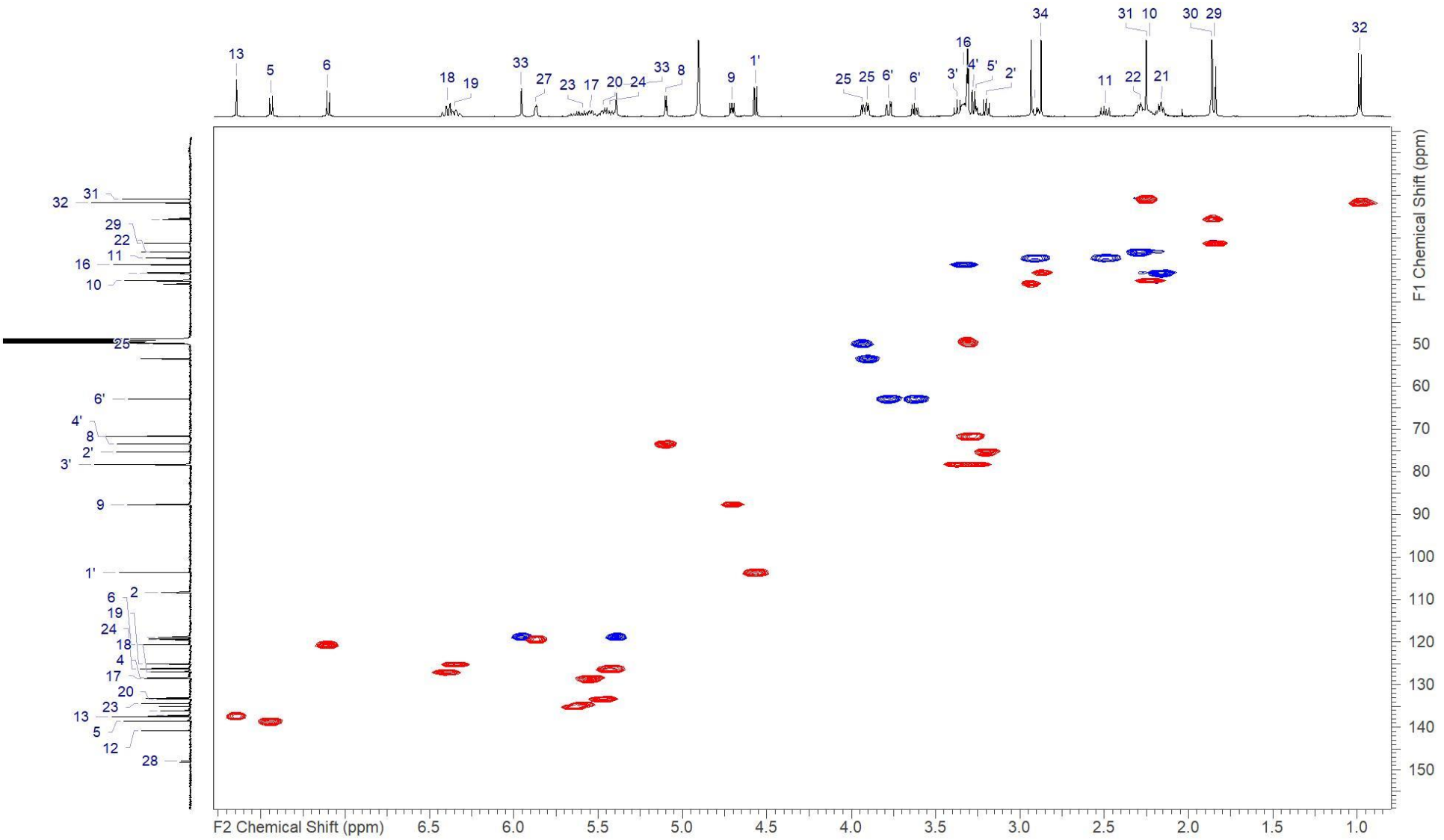
Ajudazol E (4) – HMBC NMR recorded in methanol-*d*4 at 175/700 (F1/F2) MHz



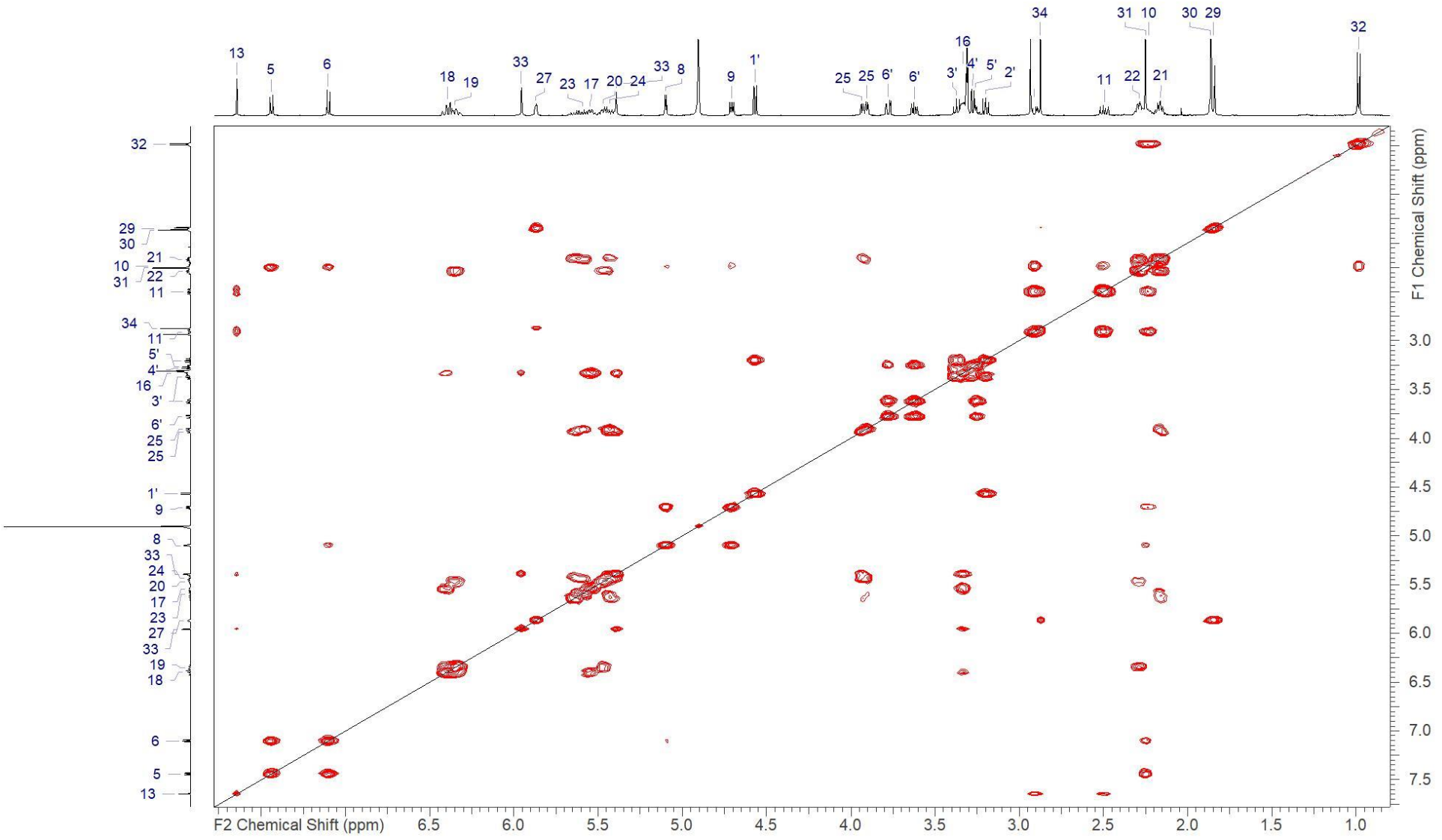
Ajudazol F (5) - ^{1}H NMR recorded in methanol-*d*4 at 500 MHz



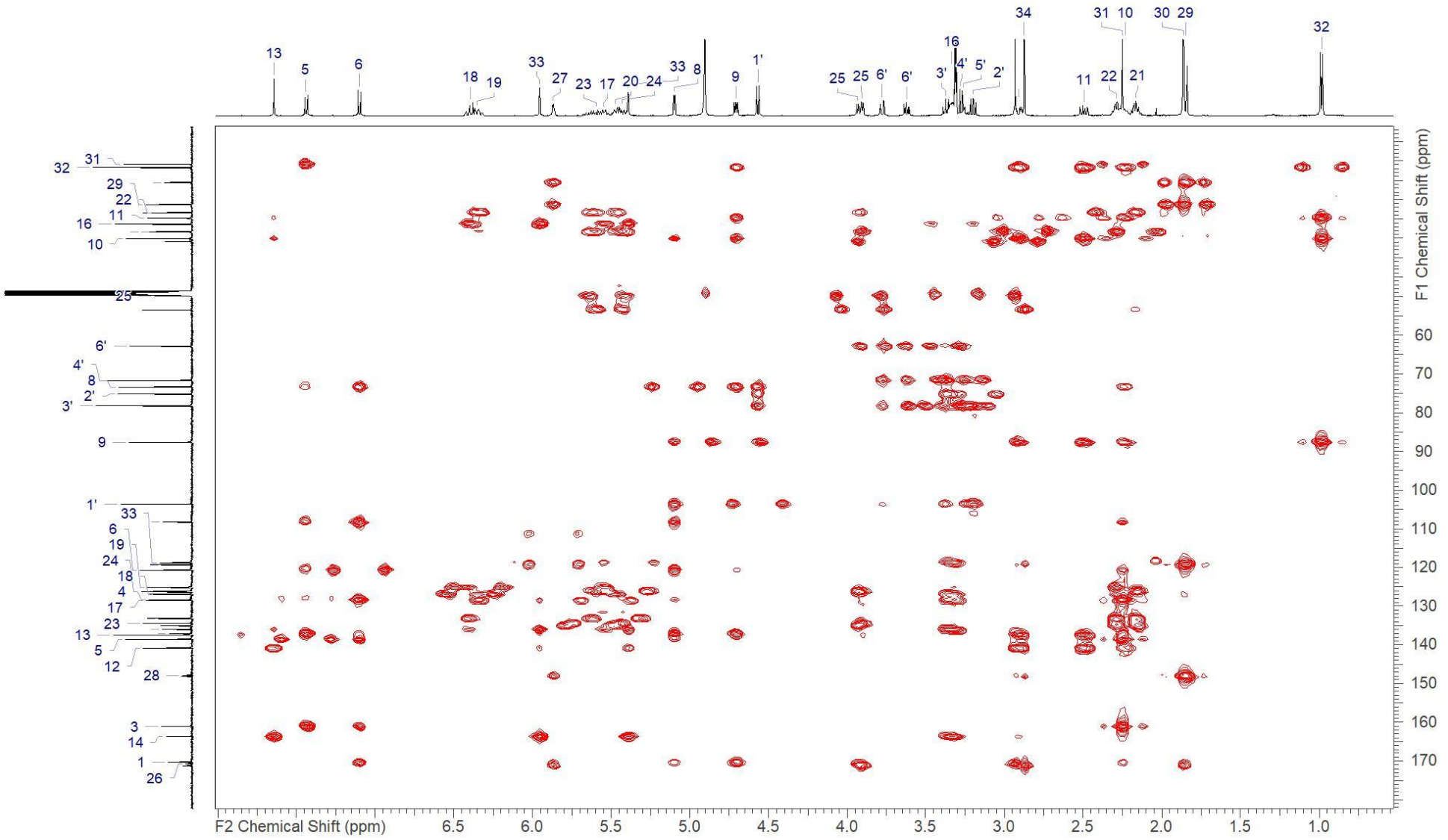
Ajudazol F (5) – ^{13}C NMR recorded in methanol-*d*4 at 150 MHz



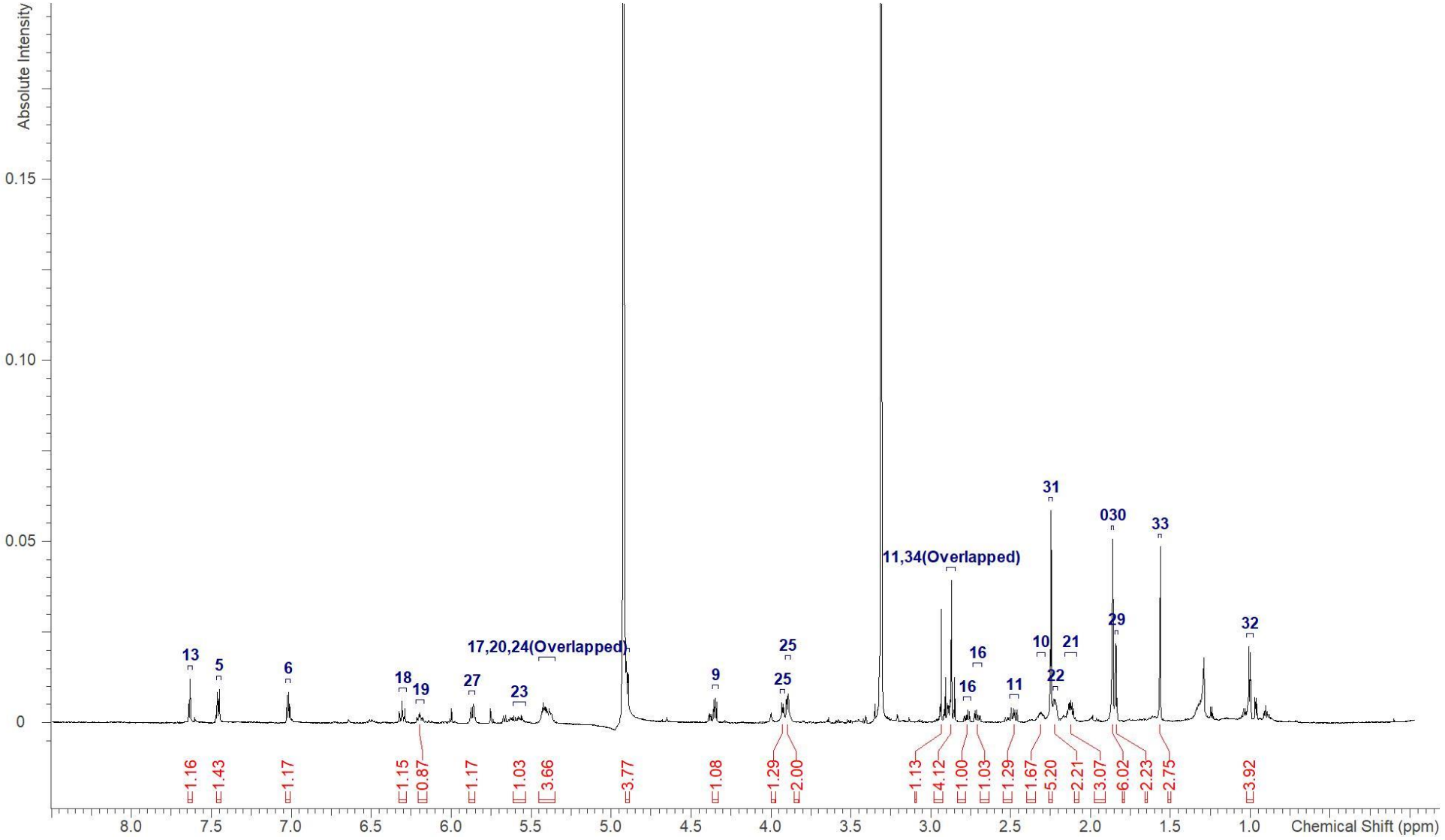
Ajudazol F (5) – HSQC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



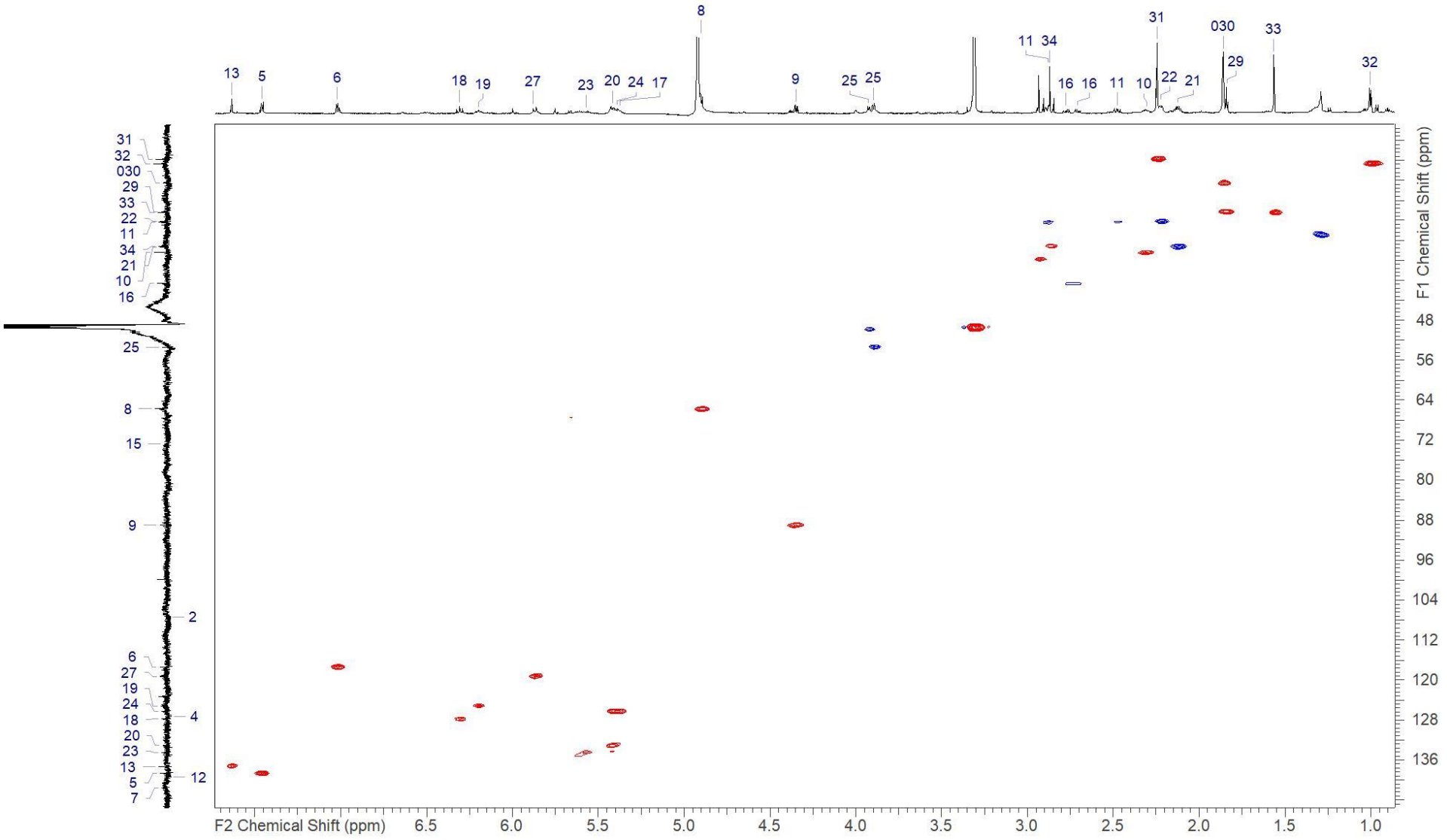
Ajudazol F (5) – COSY NMR recorded in methanol-*d*4 at 500 MHz



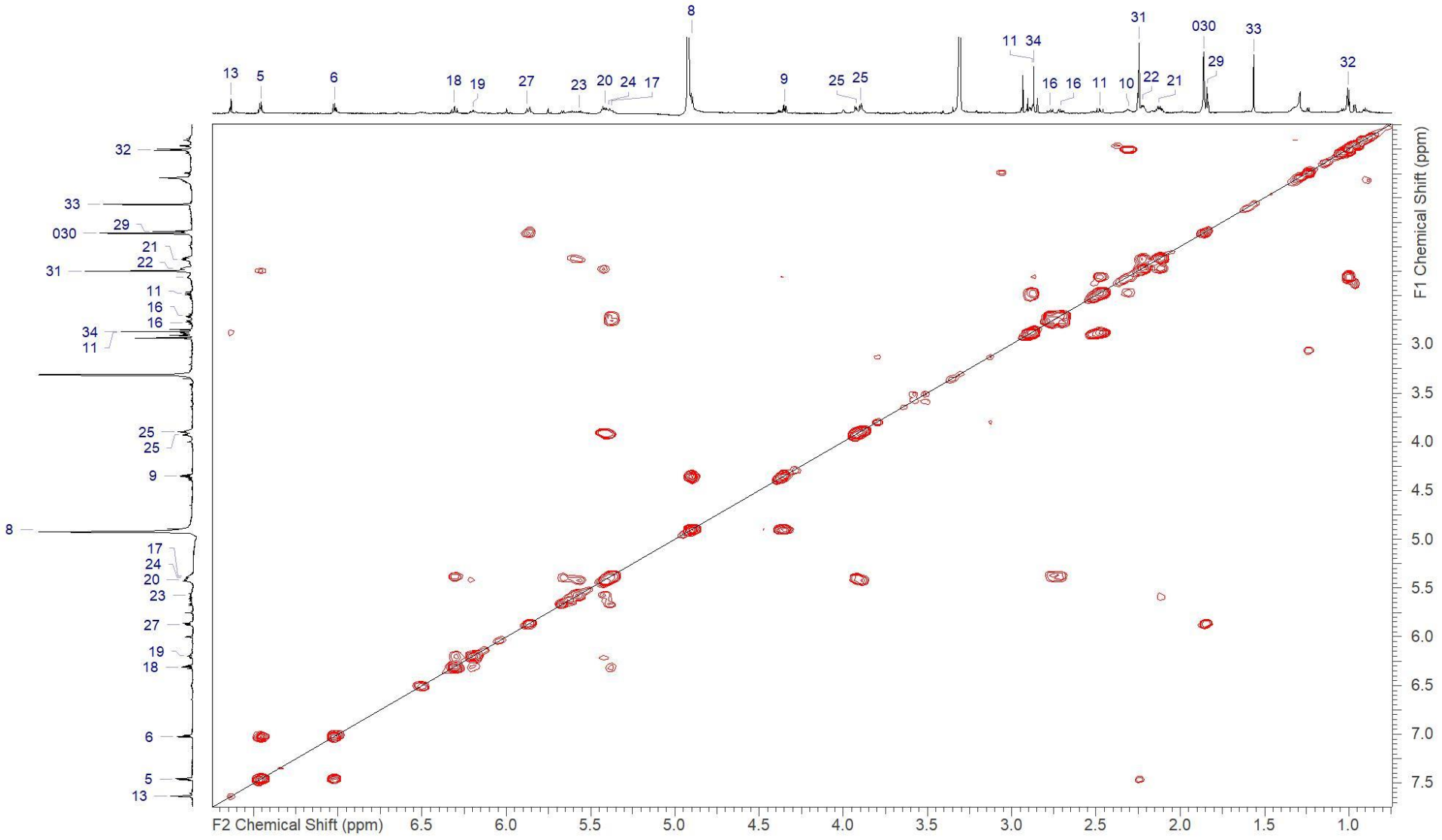
Ajudazol F (5) – HMBC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



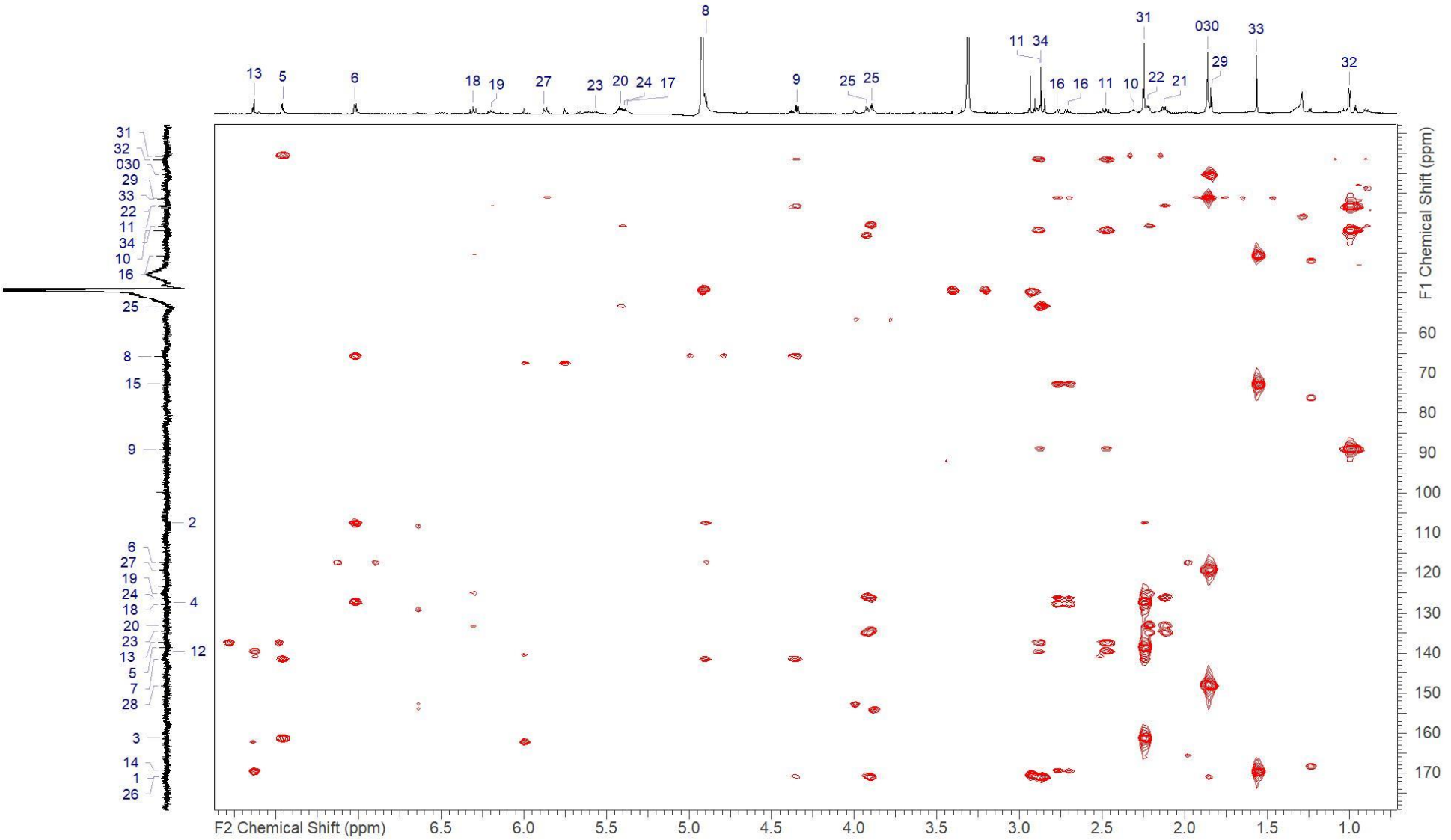
Ajudazol G (6) - ^{1}H NMR recorded in methanol-*d*4 at 700 MHz



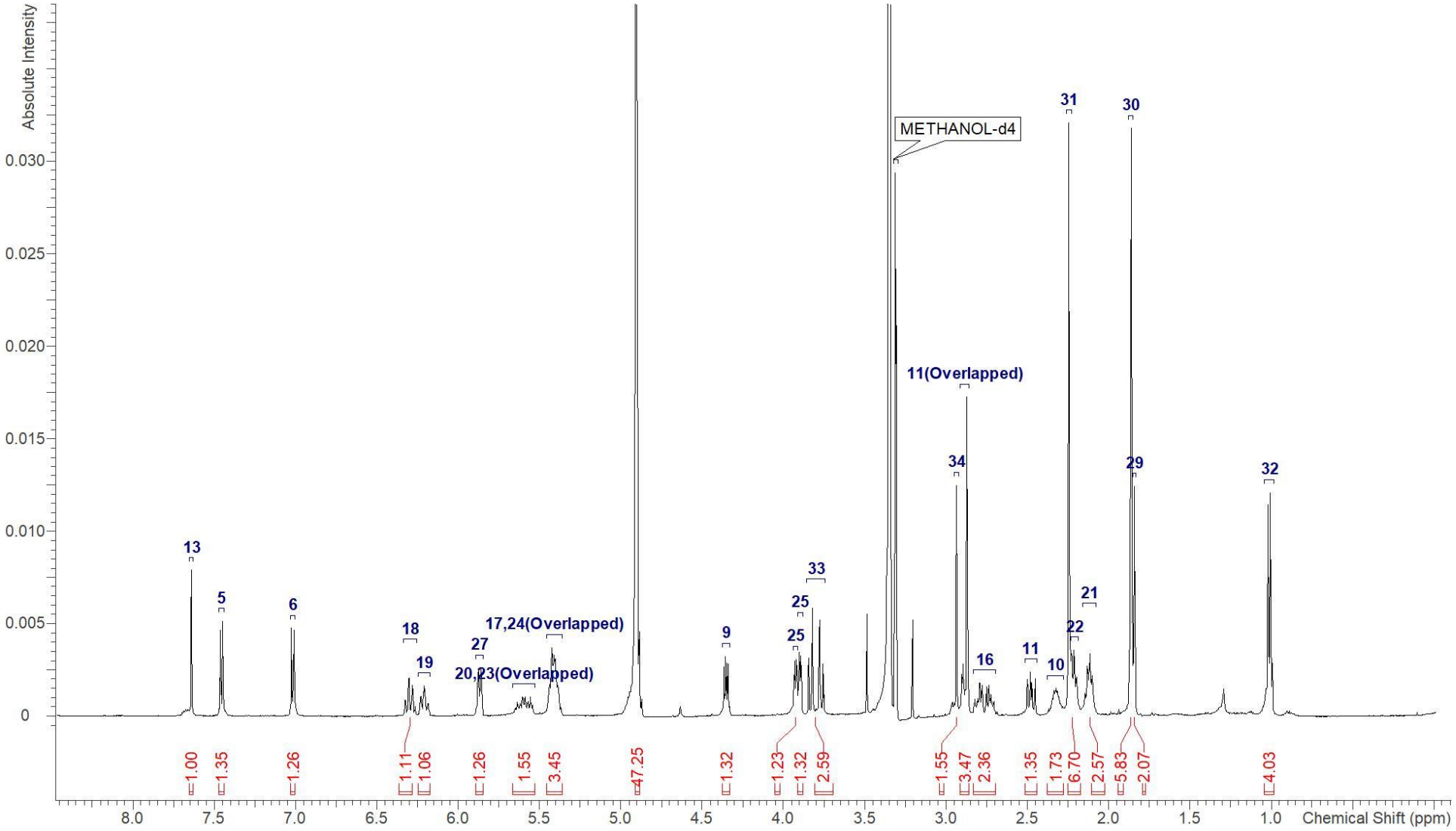
Ajudazol G (6) – HSQC recorded in methanol-*d*4 at 175/700 (F1/F2) MHz



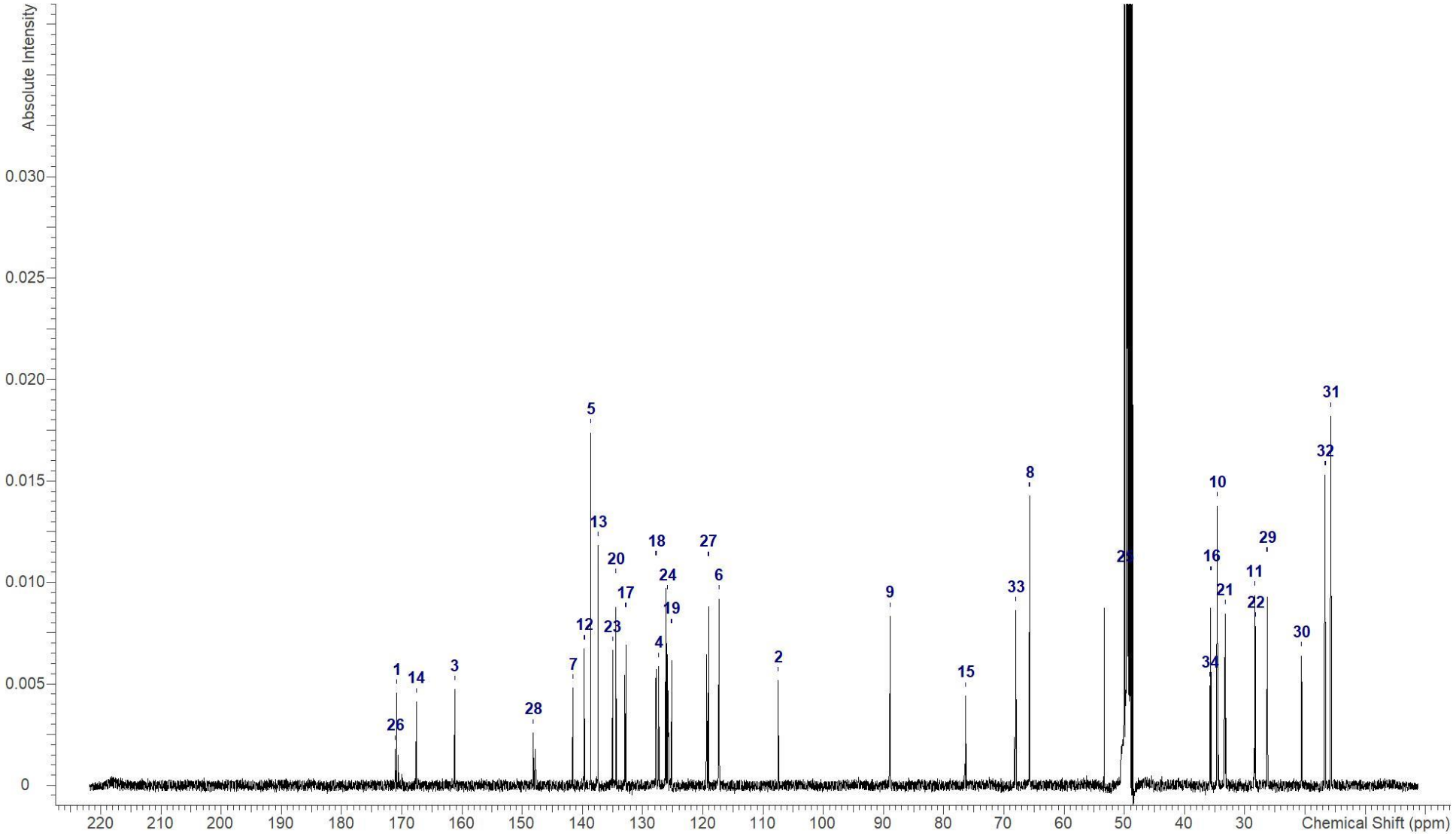
Ajudazol G (6) – COSY recorded in methanol-*d*4 at 700 MHz



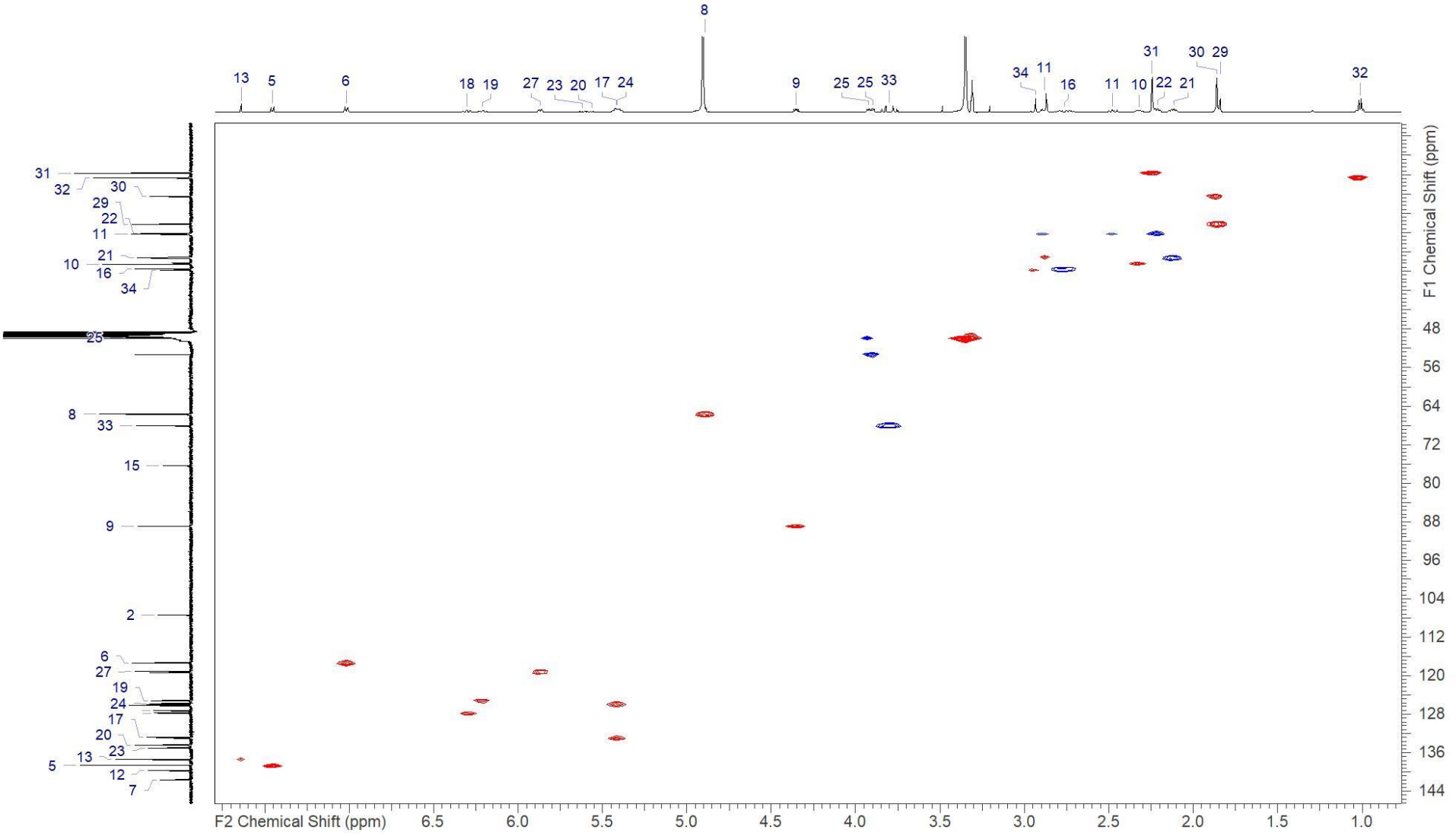
Ajudazol G (6) – HMBC recorded in methanol-*d*4 at 175/700 (F1/F2) MHz



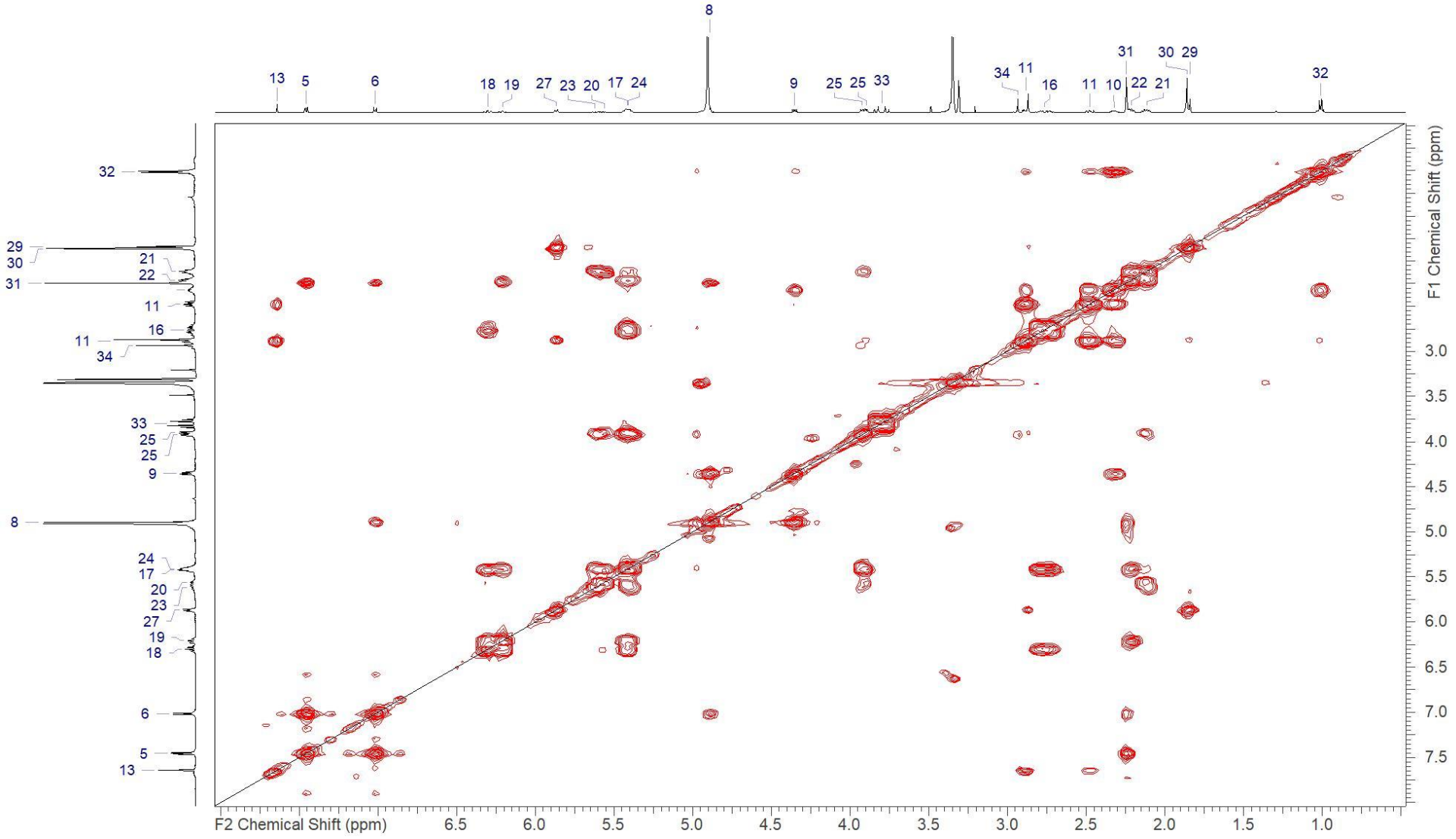
Ajudazol H (7) - ^{1}H NMR recorded in methanol-*d*4 at 500 MHz



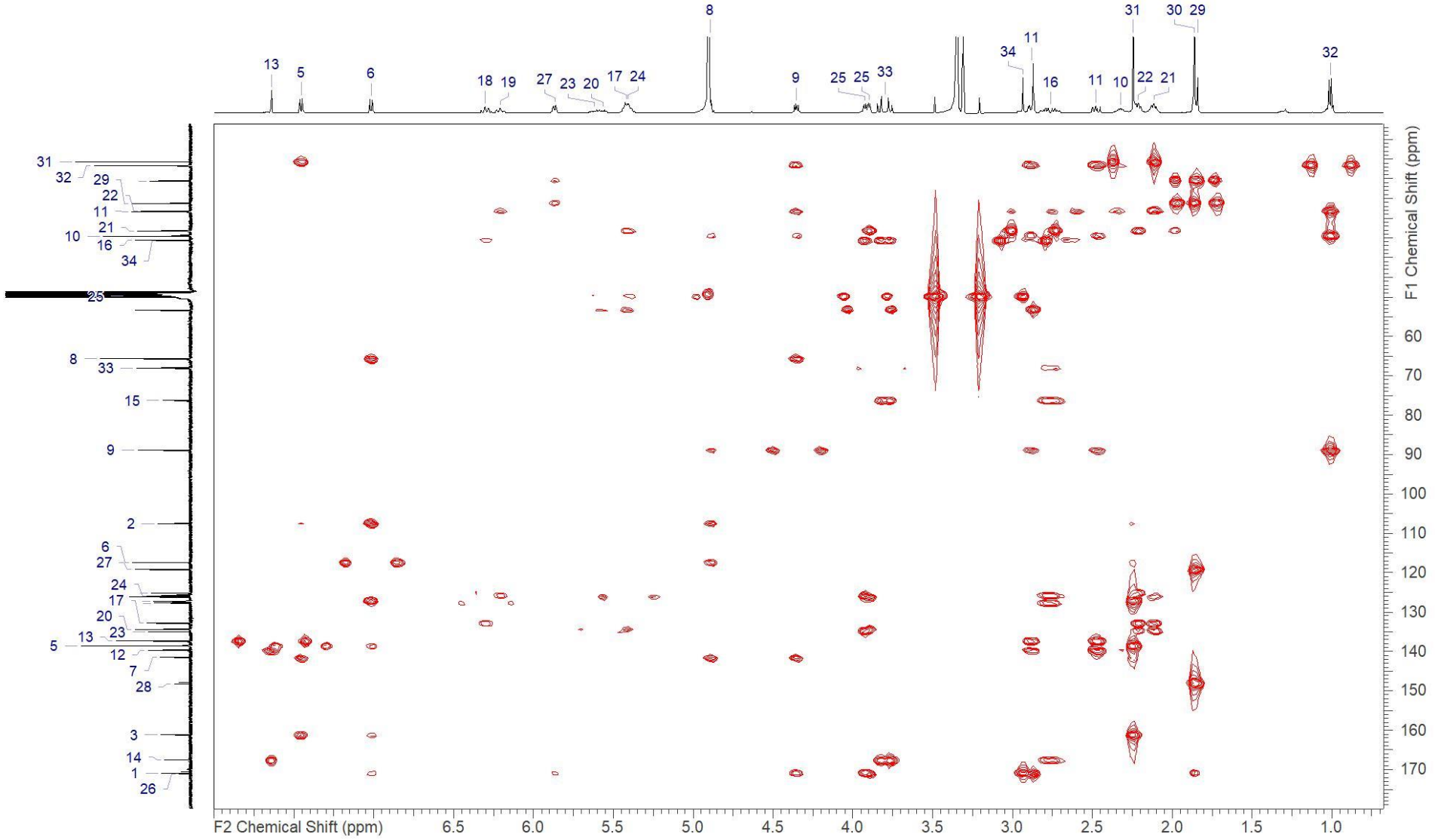
Ajudazol H (7) – ^{13}C NMR recorded in methanol-*d*4 at 150 MHz



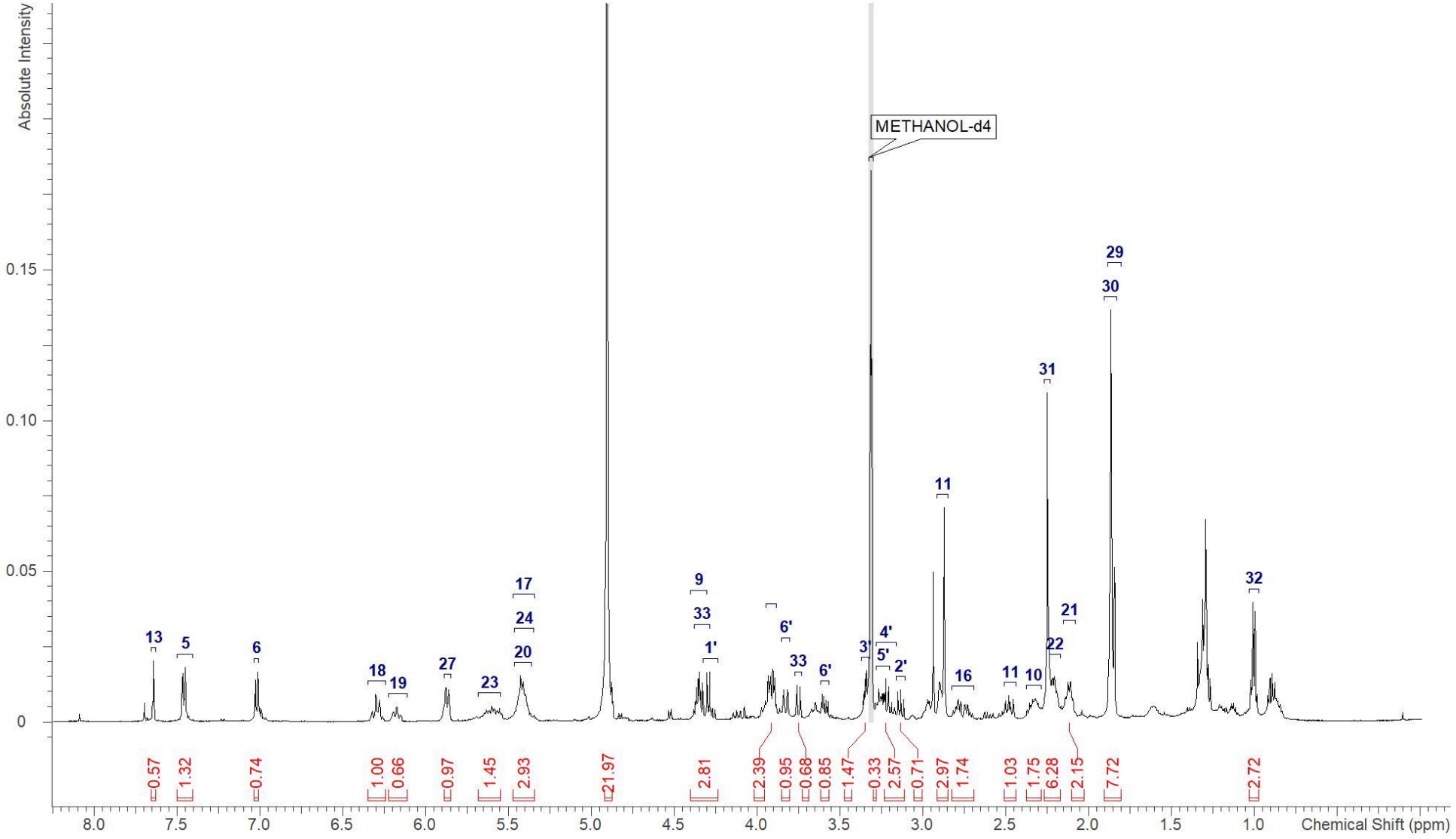
Ajudazol H (7) – HSQC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



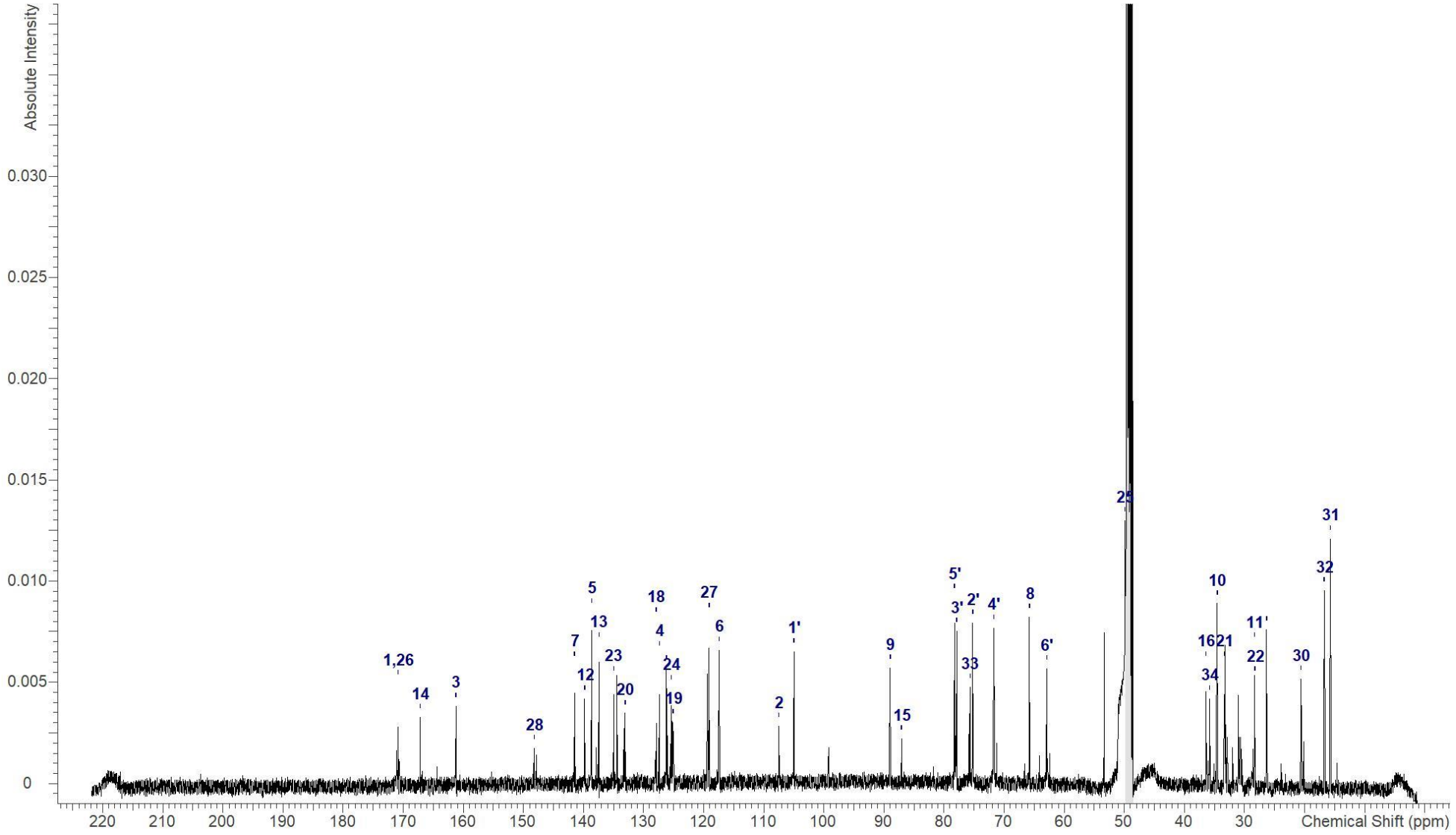
Ajudazol H (7) – COSY NMR recorded in methanol-*d*4 at 500 MHz



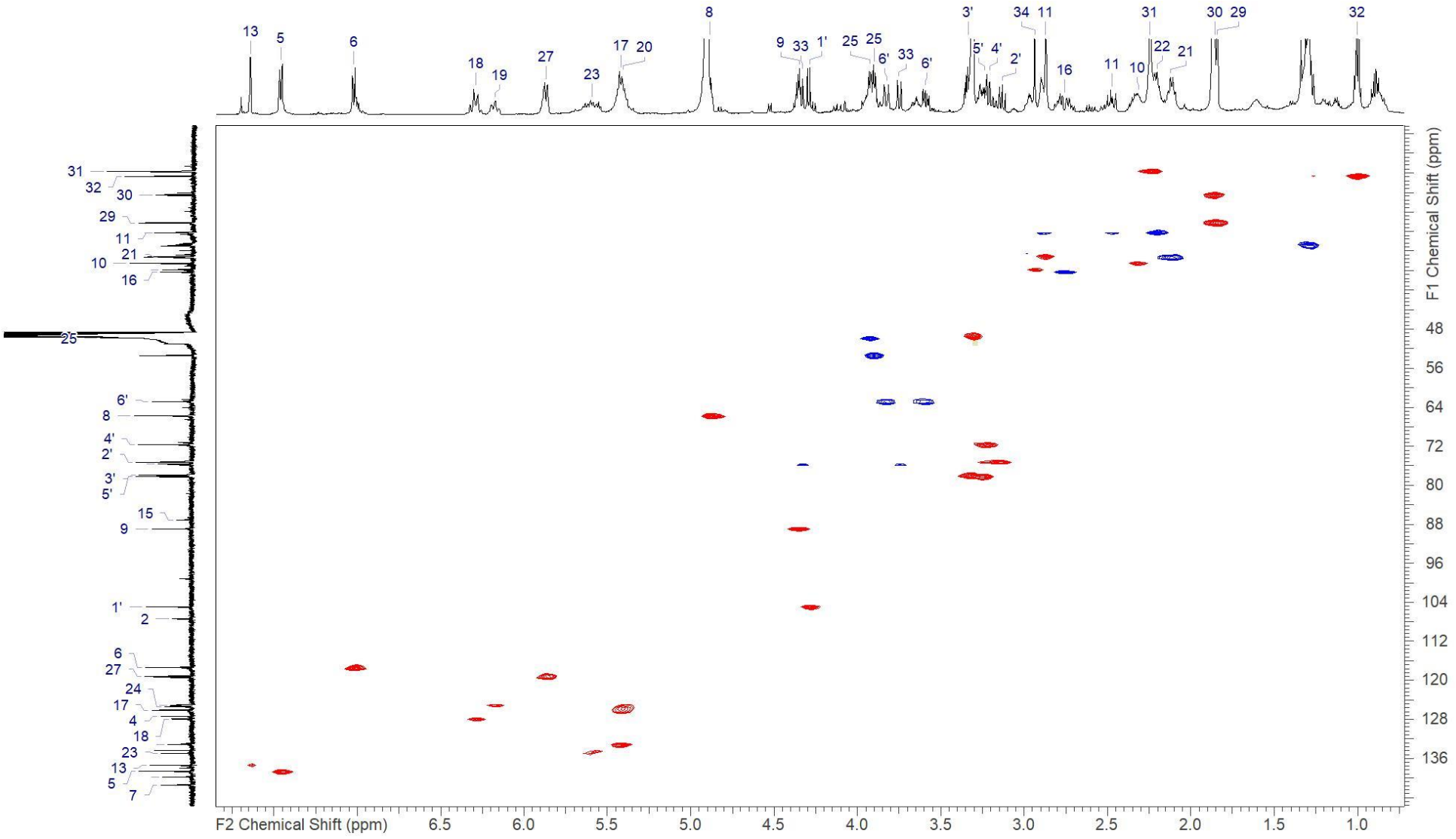
Ajudazol H (7) – HMBC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



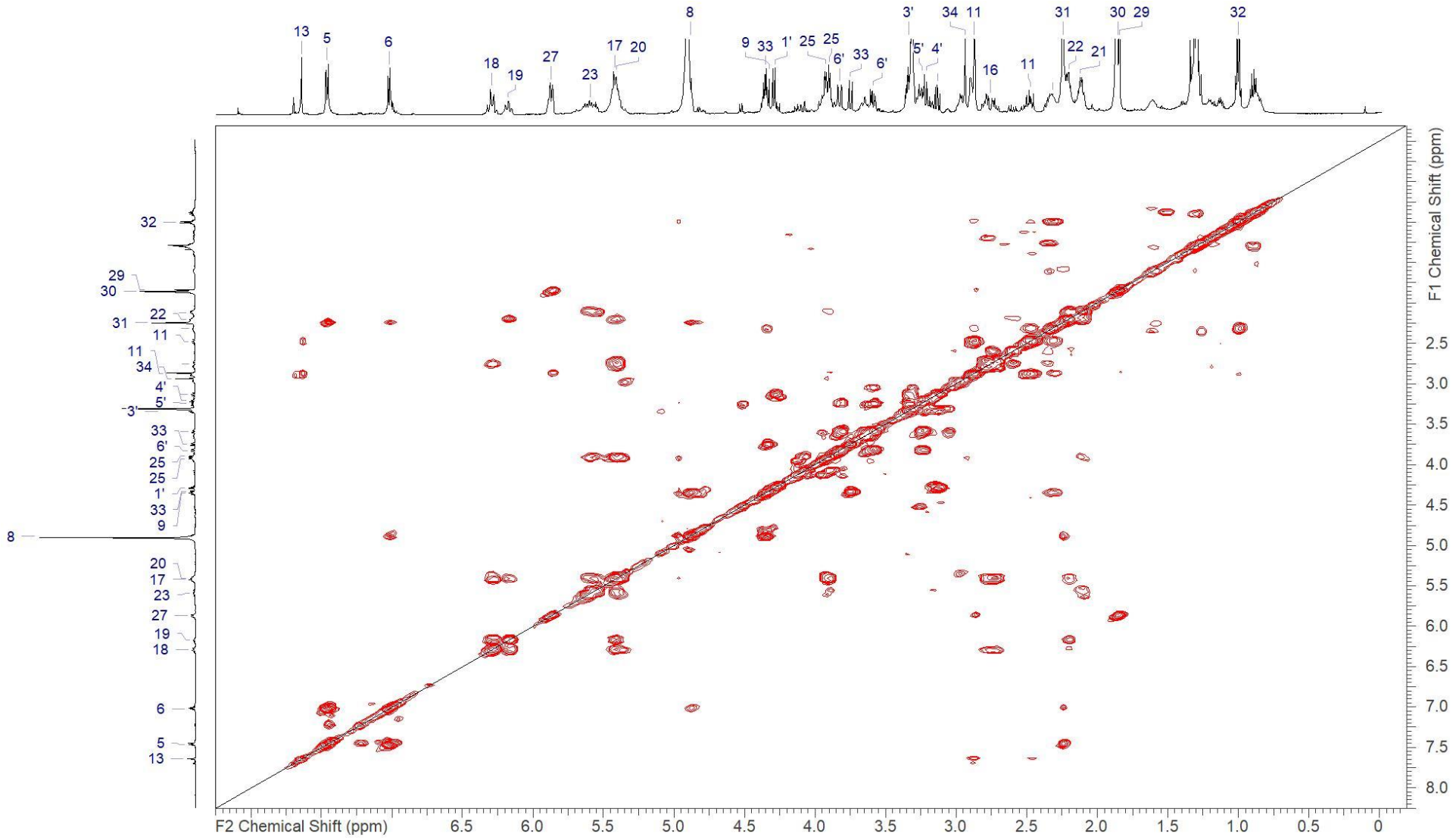
Ajudazol I (8) - ^{1}H NMR recorded in methanol-*d*4 at 500 MHz



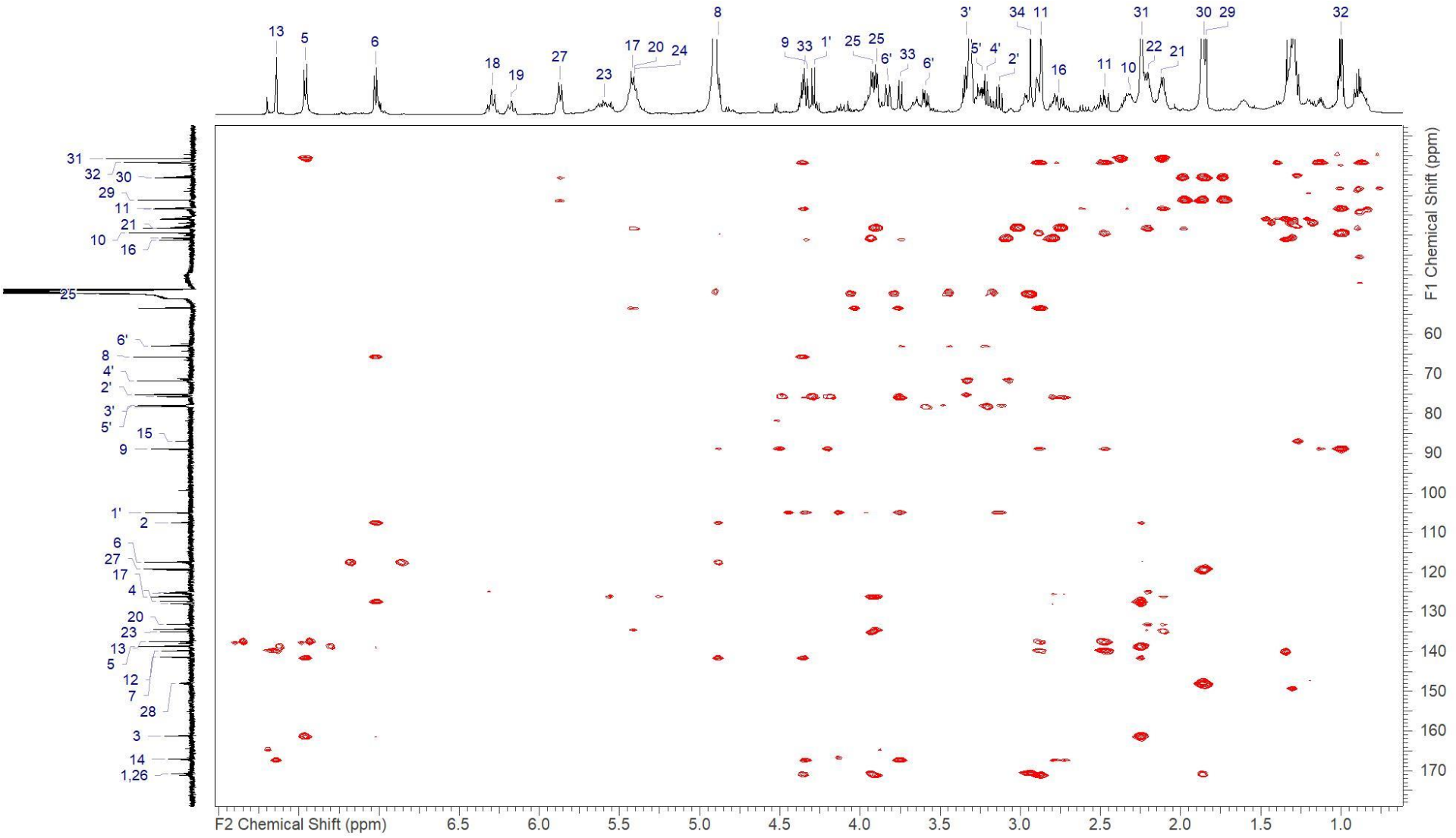
Ajudazol I (8) – ^{13}C NMR recorded in methanol-*d*4 at 150 MHz



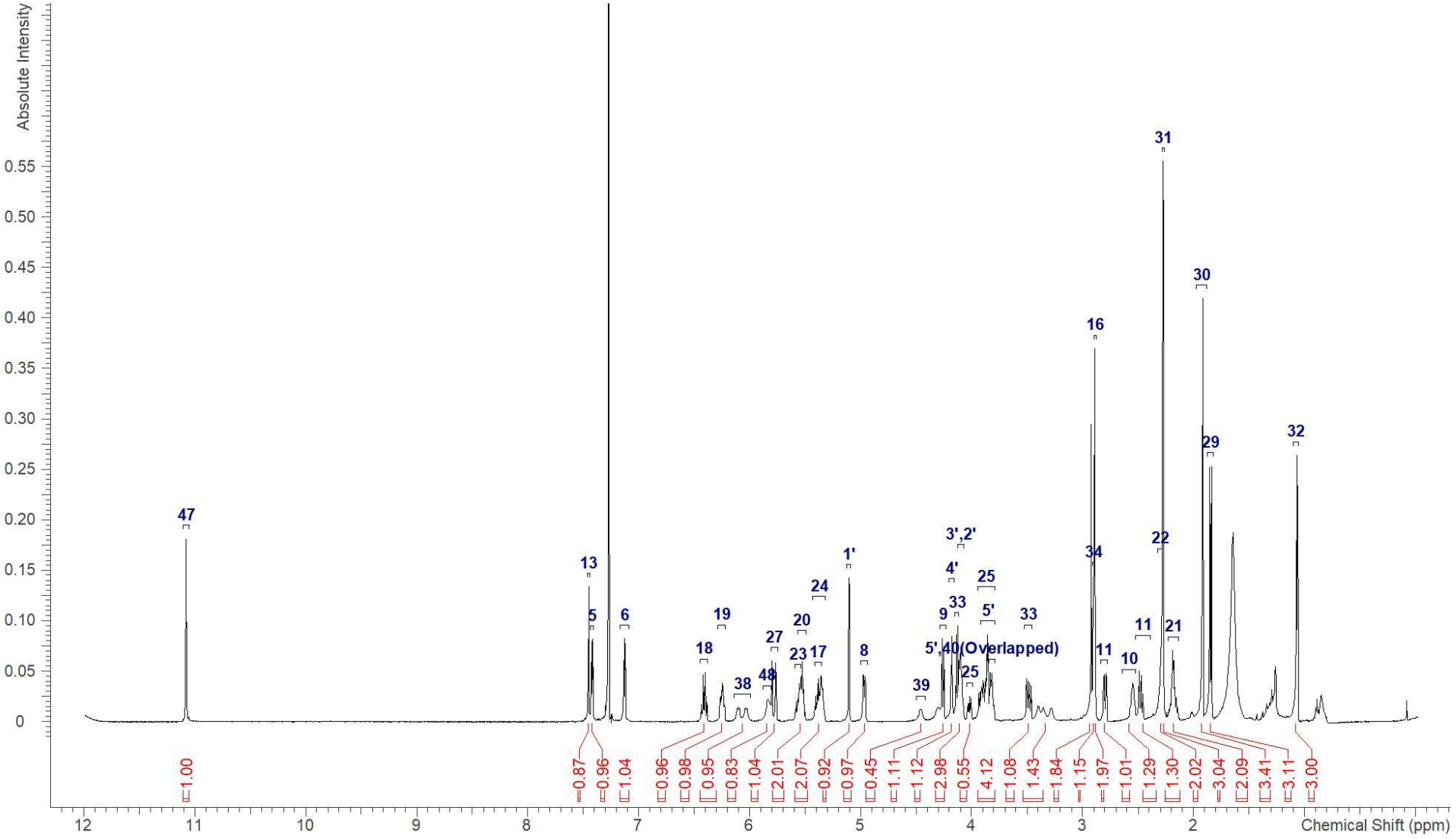
Ajudazol I (8) – HSQC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



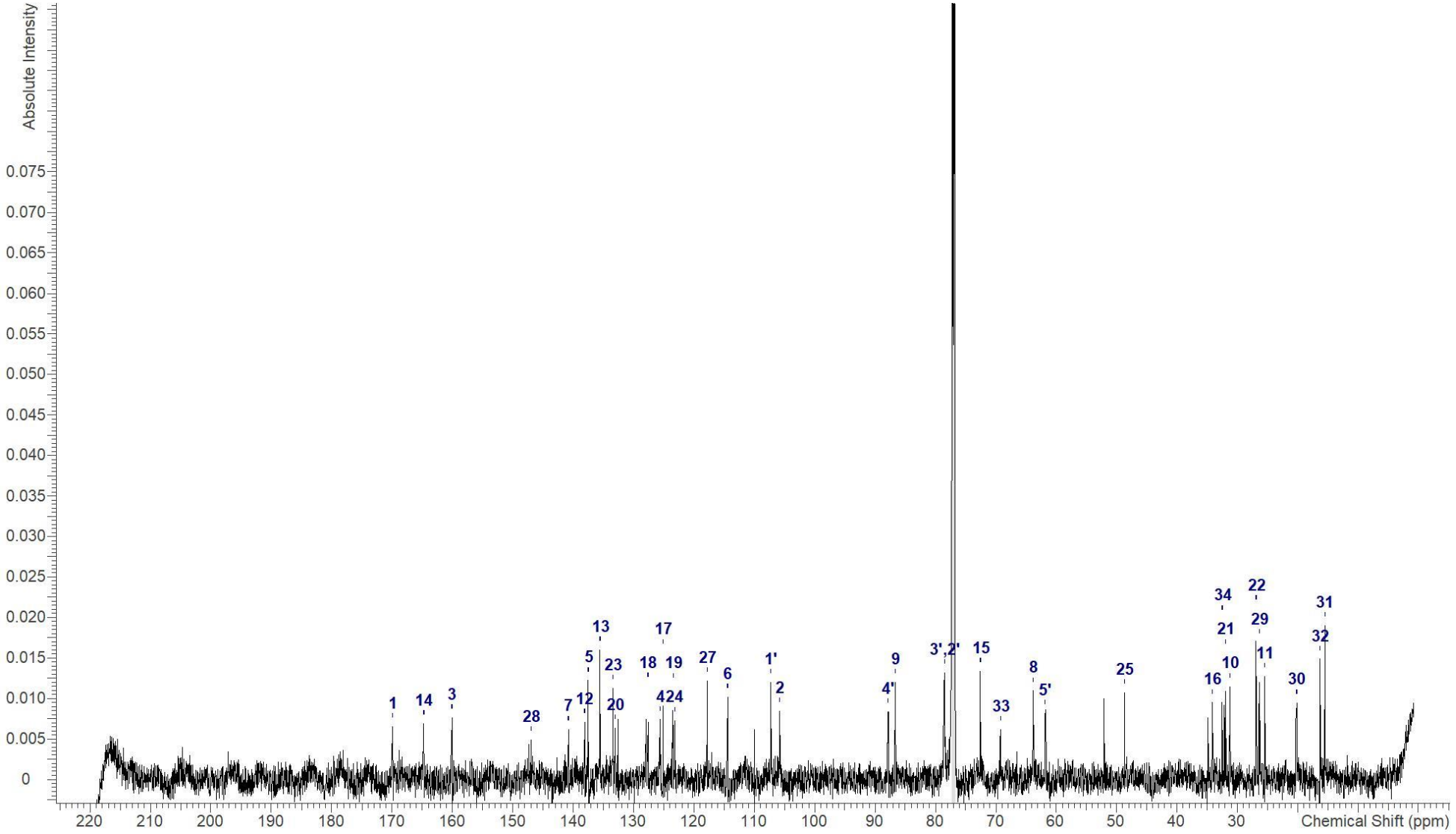
Ajudazol I (8) – COSY NMR recorded in methanol-*d*4 at 500 MHz



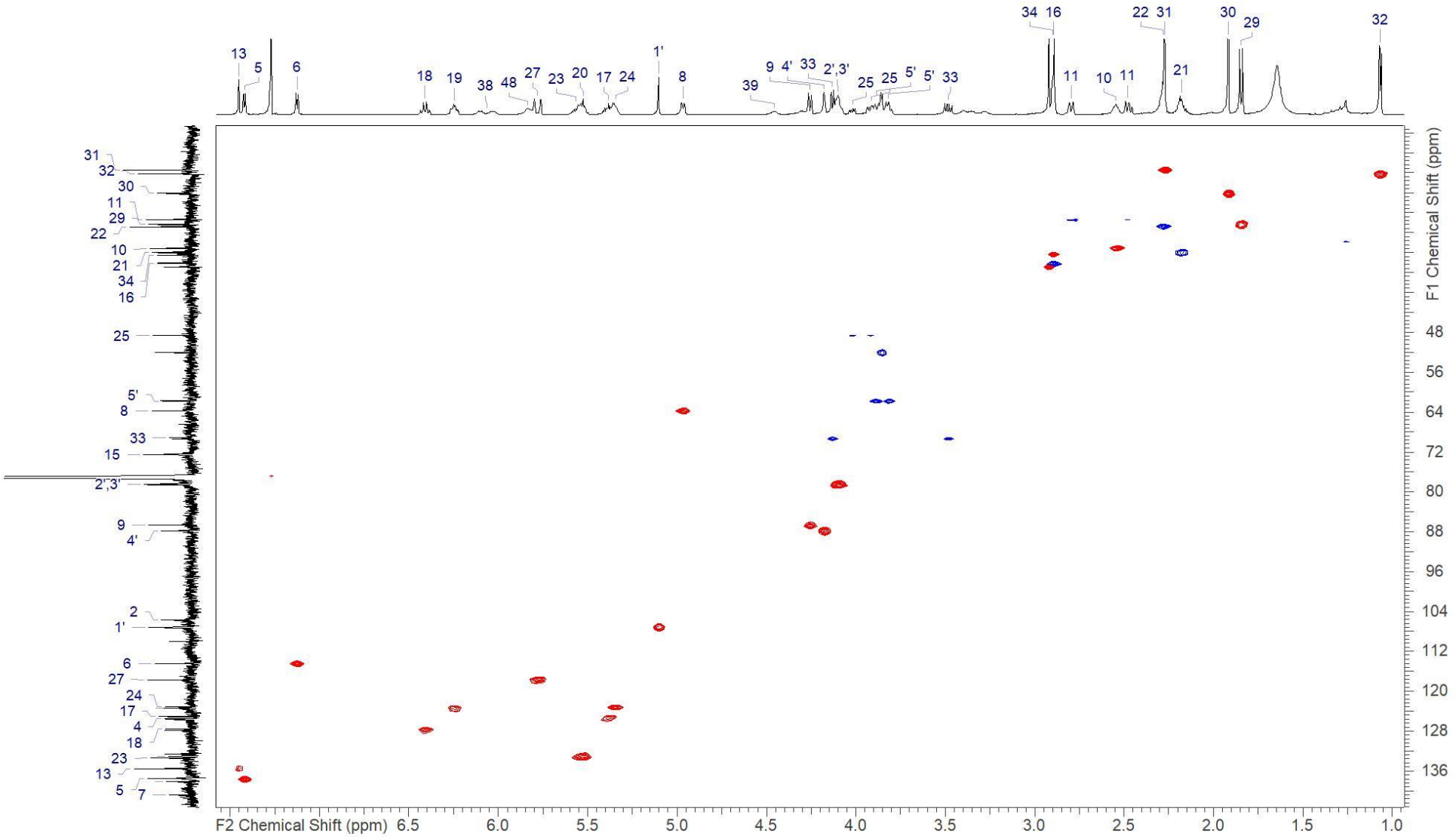
Ajudazol I (8) – HMBC NMR recorded in methanol-*d*4 at 150/500 (F1/F2) MHz



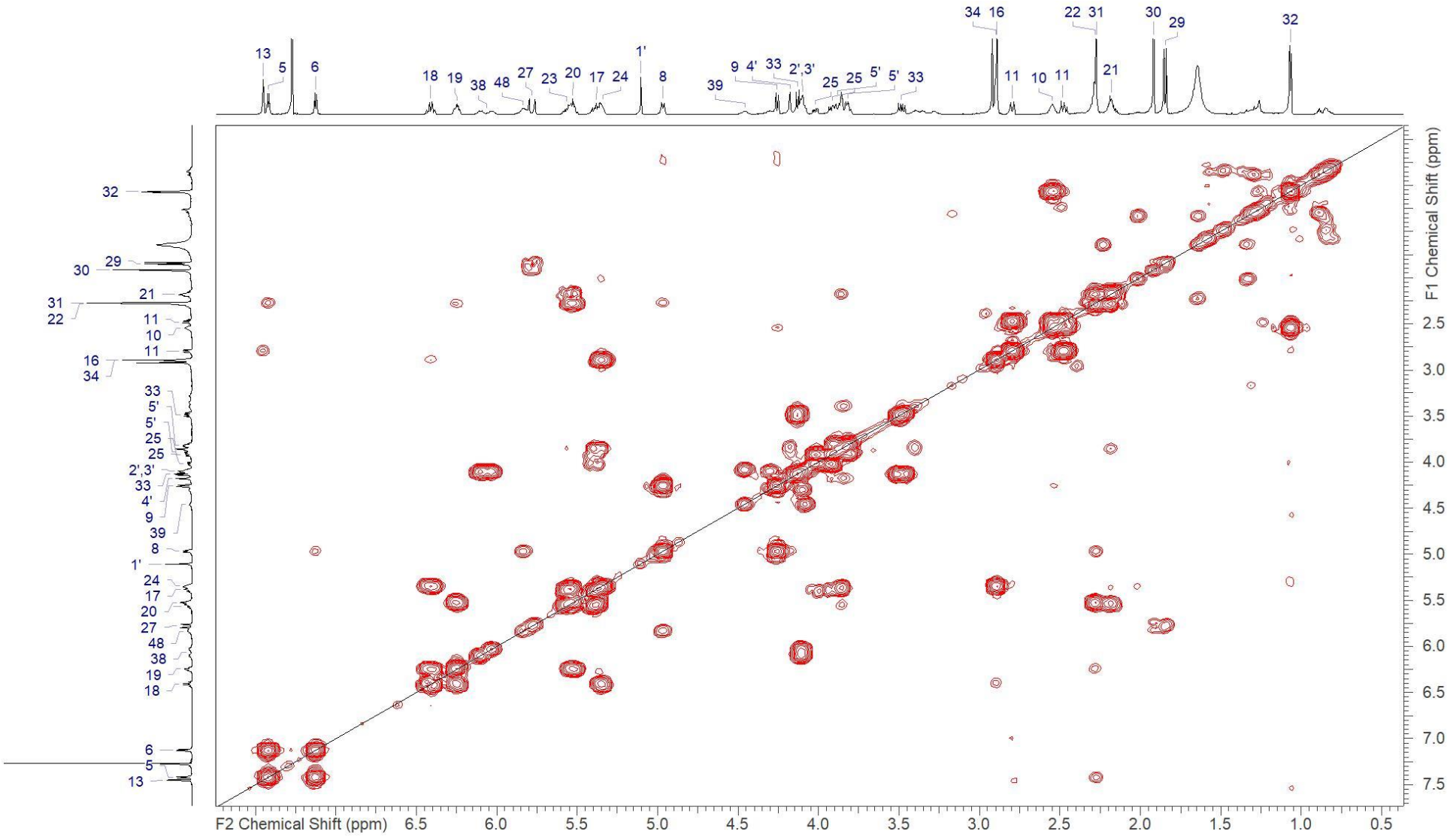
Ajudazol J (9) - ^{1}H NMR recorded in chloroform-*d* at 700 MHz



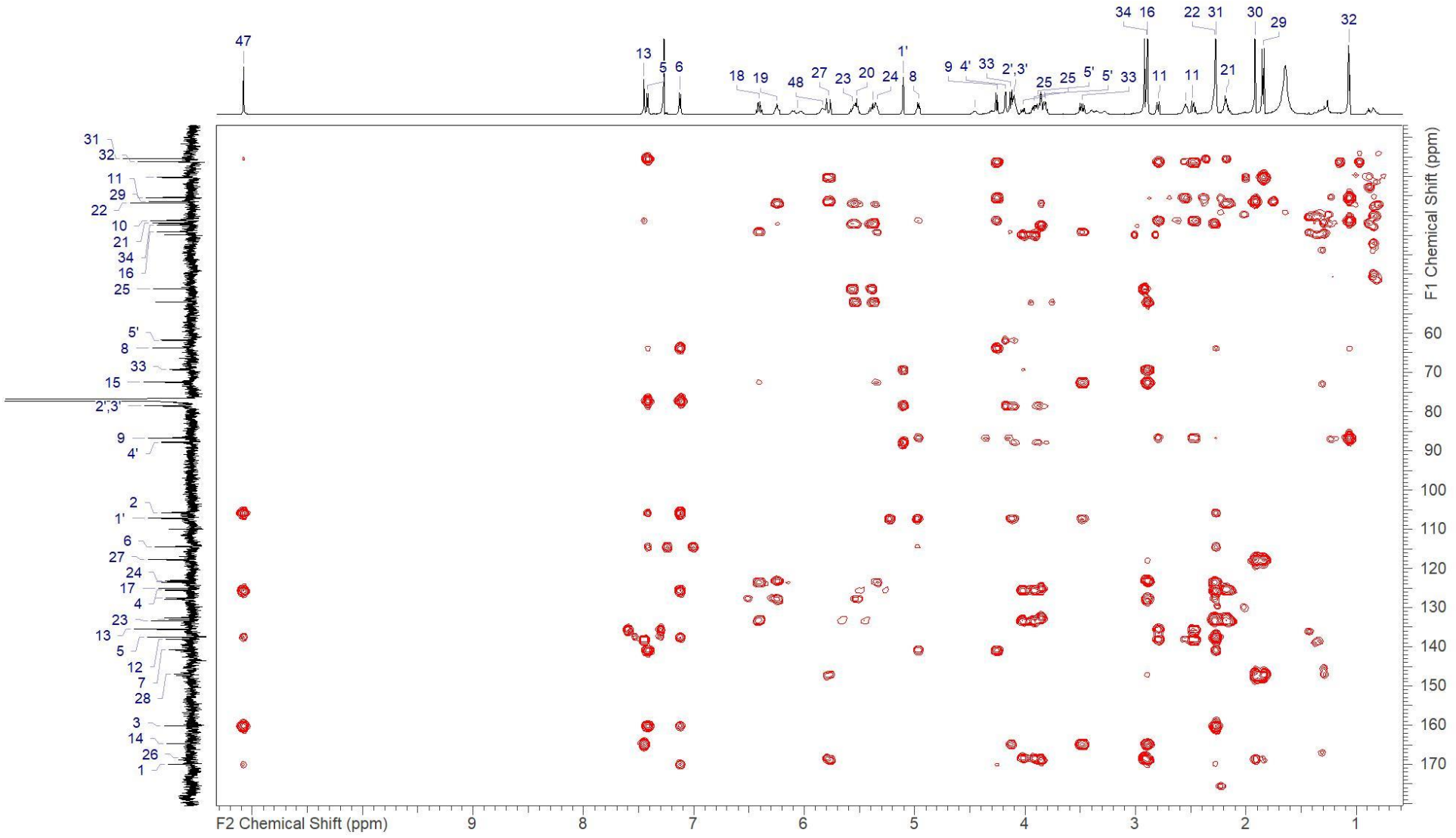
Ajudazol J (9) – ^{13}C NMR recorded in chloroform-*d* at 175 MHz



Ajudazol J (9) – HSQC recorded in chloroform-*d* at 175/700 (F1/F2) MHz



Ajudazol J (9) – COSY recorded in chloroform-*d* at 700 MHz



Ajudazol J (9) – HMBC recorded in chloroform-*d* at 175/700 (F1/F2) MHz

Supplementary references

(1) Essig, S.; Schmalzbauer, B.; Bretzke, S.; Scherer, O.; Koeberle, A.; Werz, O.; Müller, R.; Menche, D. Predictive bioinformatic assignment of methyl-bearing stereocenters, total synthesis and an additional molecular target of ajudazol B. J. Org. Chem. 2016, 81, 1333–1357.

Chapter 2

Cystopipecotides: discovery, structure elucidation and biosynthesis of myxobacterial peptides featuring a 5-hydroxyl-6-hydroxymethylpipecolic acid building block

Hu Zeng, Joy Birkelbach, Lena Keller, Carsten Volz, Chengzhang Fu and Rolf Müller

Contributions to the presented work

Author's contribution

The author contributed to the conception of the study, designed and performed experiments, evaluated and interpreted the resulting data. The laboratory and *in silico* work regarding gene cluster analysis, isolation of cystopipecotides, as well as the mutagenesis, feeding experiments and *in vitro* reconstruction of the biosynthesis of 5-hydroxyl-6-hydroxymethylpipecolic acid were performed by the author. The biosynthetic pathway was also proposed by the author. Furthermore, the author contributed significantly to conceiving and writing this chapter.

Contribution by co-workers

Joy Birkelbach and Lena Keller contributed by elucidating the structure of cystopipecotides as well as by conceiving, writing and interpreting all the NMR data. Carsten Volz contributed to conception and supervision of this study, contributed to conceiving, writing and editing of this chapter. Chengzhang Fu contributed to supervision of the *in vitro* reaction work. Rolf Müller was responsible for the conception of the project and performed the proofreading of this chapter.

3.1 Abstract

Myxobacteria are an abundant source of natural products, many of them featuring unique structures and biosynthetic pathways. Their biosynthetic potential, even in some well-explored species, is still far from being exhausted. In this chapter, the identification and characterization of a novel family of linear peptides from *Cystobacter* sp. SBCb004, the cystopipecotides, is summarized. These peptides contain an uncommon 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) as a building block. A NRPS-PKS hybrid gene cluster responsible for the biosynthesis of cystopipecotides was identified via retro-biosynthetic analysis. Based on the organization of the gene cluster we suggested that HHMPA is biosynthesized by transketolases as well as several lysine biosynthetic enzymes involved in the α-aminoadipate pathway. Based on this proposal the biosynthesis pathway was reconstructed *in vitro*, revealing a novel route for the biosynthesis of substituted pipecolate in microbes. It starts from the linkage of L-Glutamic acid to the LysW like protein CpcI, followed by sequential phosphorylation, reduction, transfer of the dihydroxyethyl group, cyclization and imine reduction eventually giving rise to HHMPA.

3.2 Introduction

Besides actinomycetes and fungi, myxobacteria have proven to be producers of a large variety of natural products[1,2]. In addition to the unique structural features and mode of actions of their metabolites[3], representatives of the myxobacteria own the largest genomes known from prokaryotes[4]. In these large genomes, numerous putative biosynthetic gene clusters (BGCs) encoding for the assembly lines of unknown natural products can be found, implying that the biosynthetic potential of myxobacteria is by far underexplored. However, many myxobacteria in strain collections are not considered further in screening attempts for novel bioactive compounds. This is due to both, the production of already known bioactive compounds or the lack of significant biological activity in the extracts. Nevertheless, many of such strains own numerous BGCs without any assigned compound.

The secondary metabolites discovered in myxobacteria, are mainly synthesized by polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS) and by the action of mixed NRPS-PKS assembly lines. The non-iterative type I PKS and NRPS are generally multimodular megasynthetases which can make use of various amino acids and carboxylic acids as building blocks, respectively, and each module is normally responsible for only one

extension cycle. Thus, the number of the modules correlate with the number of extension cycles. This one-to-one correspondence of biosynthetic assembly line and metabolic structure is termed the collinearity rule and facilitates the prediction of the product's structure based on the composition of enzymatic domains, and vice versa.

Cystobacter sp. SBCb004 harbors 24 PKS and NRPS gene clusters in its genome. Despite those BGCs encoding the biosynthetic machineries for stigmatellin[5], tubulysin[6] and argyrin[7], no natural product could be assigned to the other 21 clusters. These known metabolites exhibit antimicrobial, antifungal and antitumor activities and make the discovery of novel compounds from SBCb004 via activity-guided screening more difficult than from species without any known metabolites due to a potential interference. In this chapter, we found a novel family of linear peptides named cystopipecotides, by metabolic profiling of extracts obtained from SBCb004. These peptides bear a 5-hydroxyl-6-hydroxymethylpipecolic acid moiety (HHMPA) at their N-terminus and a γ-lactam at the C-terminus. The corresponding BGC was identified to be a NRPS-PKS hybrid gene cluster via insertional mutagenesis. The biosynthetic pathway of HHMPA involving the concerted action of transketolases and a set of lysine biosynthetic enzymes via fragments of the α-aminoadipate pathway, was proposed and reconstructed *in vitro*, exhibiting a novel route to generate substituted pipecolate in bacteria.

3.3 Materials and methods

3.3.1 Bacterial strains, culture conditions and reagents

Escherichia coli DH10B was used for cloning and *E. coli* BL21 (DE3) was used as host to express *cpc* genes. *E. coli* cells were cultivated at 37 °C in LB medium (1 % trypton, 0.5 % yeast extract, 0.5 % NaCl). *Cystobacter* SBCb004 was cultivated at 30 °C and 180 rpm in CYH medium (0.3 % casitone, 0.3 % yeast extract, 0.2 % $CaCl_2 \cdot 2H_2O$, 0.2 % soy meal, 0.2 % glucose, 0.8 % soluble starch, 0.1 % $MgSO_4 \cdot 7H_2O$, 1.19 % HEPES, 8 mg/l Fe-EDTA, adjusted to pH 7.3 with 10 N KOH) for production analysis and compound isolation while in M medium (1 % phytone, 1 % maltose, 0.1 % $CaCl_2 \cdot 2H_2O$, 0.1 % $MgSO_4 \cdot 7H_2O$, 8 mg/l Fe-EDTA, 1.19 % HEPES, adjusted to pH 7.2 with 10 N KOH) for transformation and feeding. 50 µg/ml Kanamycin was added to the culture when necessary.

Restriction endonucleases, Taq DNA polymerase, Phusion DNA polymerase, FastAP, T4 DNA ligase were purchased from Thermo scientific. NEBuilder HiFi DNA assembly Master

Mix was purchased from NEB. All the isotope precursors, unlabeled amino acids, Isopropyl β-D-1-thiogalactopyranoside (IPTG), *N*-Methylhydroxylamine hydrochloride, Fluorenylmethyloxycarbonyl chloride (Fmoc-Cl), 1,4-Dithiothreitol (DTT) and β-Nicotinamide adenine dinucleotide 2'-phosphate reduced tetrasodium salt hydrate (NADPH) were purchased from Sigma-Aldrich. Oligonucleotide primers were synthesized by Sigma-Aldrich and DNA sequencing of PCR products were performed by LGC, Biosearch Technologies.

3.3.2 Gene inactivation in SBCb004

Gene inactivation in SBCb004 was accomplished by homologous recombination of a suitable plasmid backbone in the encoding region. At this, internal DNA fragments of the target genes served as homologous regions to mediate the insertion of the constructed plasmids into the target gene via homologous recombination. Internal DNA regions of the target gene with a size of 800 bp to 1200 bp were amplified from SBCb004 genomic DNA by using corresponding primers 9101-sc-gene-F and 9101-sc-gene-R (Table S1). Subsequently, these DNA fragments were cloned into plasmid pbluekan kindly provided by Dr. Carsten Volz (unpublished). pbluekan encodes a aminoglycoside phosphotransferase (*aph*(3')-II) conferring kanamycin resistance[8] as well as a β-lactam resistance cassette (*bla*) and a *lacZ* gene under the control of the T7A1 promoter[9] (Figure S1). Stop codons were introduced at both ends of the cloned homologous regions upon PCR amplification by the use of mutated oligonucleotide primer. The constructed plasmids (Table S2) were confirmed by Sanger sequencing using primer seq-pblueCN-R. Correct plasmids were introduced into SBCb004 by electroporation following a previously published protocol[6]. The correct transformants were confirmed by PCR reactions with primer CN-check-R which binds the plasmid backbone and primer gene_CK_F (where "gene" stands for the designation of the target gene) which binds the upstream of the target gene.

3.3.3 UPLC-MS/MS measurement

The measurements of crude extracts and pure compounds were performed on a Dionex UltiMate 3000 RSLC system equipped with Waters BEH C18 column (50 × 2.1 mm, 1.7 µm). The flow rate was 600 µl/min and the column temperature was 45 °C. The gradient was as follows: 0 - 0.5 min, 95 % water with 0.1 % formic acid (A) and 5 % acetonitrile with

0.1 % formic acid (B); 0.5 - 18.5 min, 5 - 95 % B; 18.5 - 20.5 min, 95 % B; 20.5 - 21 min, 95 - 5 % B; 21 - 22.5 min, 5 % B. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the Bruker Daltonics maXis 4G hr-qToF mass spectrometer using the Apollo II ESI source. Mass spectra were acquired in centroid mode ranging from 150 - 2500 m/z at a 2 Hz full scan rate. Full scan and MS/MS spectra were acquired at 2 Hz.

3.3.4 Intact protein measurement

Intact protein measurements were performed on a Dionex UltiMate 3000 RSLC system using an Aeris Widepore XB-C8 column (150 × 2.1 mm, 3.6 µm, Phenomenex). The flow rate was 300 µl/min and the column temperature was 45 °C. The separation of 1 µl sample was achieved using the followng gradient: 0 - 0.5 min, 98 % water with 0.1 % formic acid (A) and 2 % acetonitrile with 0.1 % formic acid (B); 0.5 - 10.5 min, 2 - 75 % B; 10.5 - 13.5 min, 75 % B; 13.5 - 14 min, 75 - 2 % B; 14-15.5 min, 2 % B. UV spectra were recorded by a DAD ranging from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the maXis-4 hr-qToF mass spectrometer (Bruker Daltonics) using the Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4,000 V, the dry-gas flow was 5.0 L/min and the nebulizer was set to 1.0 bar. Mass spectra were acquired in positive ionization mode in the range from 600 - 1800 m/z at a scan rate of 2.5 Hz. Protein masses were deconvoluted using the Maximum Entropy algorithm (Copyright 1991-2004 Spectrum Square Associates, Inc.).

3.3.5 Isolation and purification of cystopipecotide A and B

SBCb004 was grown in 20 L CYH medium supplemented with 2 % Amberlite XAD-16 resin at 30 °C for 10 days. Cells and XAD-16 resin were harvested by centrifugation and extracted with 3 × 1.5 L acetone. The extract was partitioned between hexane and methanol (3:1, v/v) and the methanol extract was subsequently partitioned between ethyl acetate and deionized water (3:1, v/v). The water phase extract was separated by semi-preparative HPLC (Waters XSelect CSH Phenyl-Hexyl, 5 µm, 250 × 10 mm; 5 ml/min; UV 220 nm) using a gradient as follows: 0 - 5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 23 % B; 8 - 31 min, 23 % to 23.7 % B; 31 - 33

min, 23.7 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. Two impure fractions collected at 16.2 min and 21.9 min were obtained.

The impure fraction collected at t_R = 21.9 min was further purified on a Waters Xbridge Peptide BEH C18 HPLC column (10 × 250 mm, 5 µm; 5 ml/min; UV 220 nm) using the following gradient: 0 - 5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.4 % B; 31 - 33 min, 30.4 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. The impure fraction was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 ml/min; UV 220 nm) using the following gradient: 0 - 5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.3 % B; 31 - 33 min, 30.3 % to 95 % B; 33 -34 min, 95 % to 5 % B; 34 - 36 min, 5 % B, affording cystopipecotide A (9.8 mg, t_R = 16.5 min).

The impure fraction collected at t_R = 16.2 min was further purified on a Waters Xbridge Peptide BEH C18 HPLC column (10 × 250 mm, 5 µm; 5 ml/min; UV 220 nm) using the following gradient: 0 - 5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 30 % B; 8 - 31 min, 30 % to 30.4 % B; 31 - 33 min, 30.4 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B. The impure fraction was further purified on a Phenomenex Kinetex Biphenyl HPLC column (10 × 250 mm, 5 µm; 5 ml/min; UV 220 nm) using the following gradient: 0 - 5 min, 5 % H_2O with 0.1 % formic acid (A) and 5 % acetonitrile with 0.1 % formic acid (B); 5 - 8 min, 5 % B to 25 % B; 8 - 31 min, 25 % to 26.5 % B; 31 - 33 min, 26.5 % to 95 % B; 33 - 34 min, 95 % to 5 % B; 34 - 36 min, 5 % B, affording cystopipecotide B (4.2 mg, t_R = 17.0 min).

3.3.6 Structure elucidation

UV/Vis data were obtained using the DAD of the Dionex Ultimate 3000 SL system linked to a Bruker maXis 4G UHRqTOF for mass spectrometric analysis.

NMR data were recorded on an UltraShield 500 MHz (1H at 500 MHz, 13C at 125 MHz) or an AVANCE III 700 MHz NMR (1H at 700 MHz, 13C at 175 MHz) equipped with a 5 mm inverse TCI cryoprobe (Bruker, Billerica, MA, USA). Shift values (δ) are calculated in ppm and coupling constants (J) in Hz.

3.3.7 Feeding experiment

The feeding studies of SBCb004 were performed in 50 ml M medium. 11.34 mg D-Glucose-1-^{13}C, 11.30 mg D-Glucose-2-^{13}C, 3.87 mg L-Glutamic-2,3,3,4,4-d_5-acid, 5.10 mg L-Valine-d_8, 2.37 mg D/L-pipecolic acid, 22.50 mg Sodium propionate-1-^{13}C, 7.12 mg L-Tryptophan-$^{15}N_2$ and 3.0 mg L-Lysine-$^{13}C_6$,$^{15}N_2$ hydrochloride were resolved in 1 ml water, respectively, while 1.6 mg L-Leucine-5, 5, 5-d_3 were resolved in 1 ml 1 N HCl and sterilized by filtration. Equal aliquots (250 µl) were added to the culture after 24 h, 36 h, 48 h and 60 h. A control experiment without isotopically labelled precursor was done in parallel. 2 % Amberlite XAD-16 resin was added after 72 h. The culture was harvested after 90 h by centrifugation. The cells/XAD were extracted with 50 ml acetone. The solvents were evaporated and the residue was resolved in MeOH and analyzed by UPLC-MS.

3.3.8 Protein expression and purification

The coding sequences of CpcG, CpcH, CpcI, CpcJ, CpcK, CpcN, CpcO and CpcP were separately amplified from SBCb004 genomic DNA. All the PCR primers and plasmids used are listed in Table S1 and Table S2. CpcI, CpcJ, CpcO and CpcK were cloned into pHisTEV, while CpcG, CpcP and CpcN were cloned into pHisSUMOTEV (a gift from Dr. David Owen, DIAMOND Light Source) and CpcH was cloned into pCDF-1b (Novagen). The resulting expression plasmids were verified by endonuclease digestion and Sanger sequencing before being transformed into *E. coli* BL21 (DE3). CpcG and CpcH were coexpressed. Each transformant was pre-cultured in LB medium supplemented with kanamycin (50 µg/ml, for vecto pHisTEV and pHisSUMOTEV) or spectinomycin (50 µg/ml, for vector pCDF-1b) at 37 °C overnight and then transferred (1 to 100) into fresh LB medium containing appropriate antibiotics. The *E. coli* cells were grown at 37 °C till the optical density (OD_{600}) reached 0.6. The cultures were then transferred to 16 °C followed by adding 0.1 mM IPTG to induce the protein expression. The cells were continued at 16 °C overnight before being harvested by centrifugation.

The cell pellets except those used to co-express CpcG and CpcH, were resuspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 20 mM imidazole, 1 mM DTT, 10 % glycerol). For co-purification of CpcG and CpcH, lysis buffer (20 mM Bis-Tris, pH 6.8, 200 mM NaCl, 20 mM imidazole, 1 mM DTT, 10 % glycerol) was used. The cell suspension was lysed via passing through a cell disrupter (Constant Systems) at 30,000 psi, and cell

debris was removed by centrifugation (23,500 rpm, 4 °C, 30 min). The supernatant was collected and directly loaded onto a 5-ml Ni-NTA Superflow Cartridge (QIAGEN) pre-equilibrated with lysis buffer. The column was washed extensively with 30 column volumes lysis buffer before the protein was eluted with lysis buffer supplemented with 250 mM imidazole. Fractions containing the targeted protein were directly loaded onto a gel filtration column (HiPrep^{TM} 26/100, GE healthcare) pre-equilibrated in gel filtration buffer (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM DTT). The fractions of the highest purity indicated by SDS-PAGE were pooled and concentrated by Amicon Ultra centrifugal filters (Merck Millipore). Concentrated proteins were flash frozen by liquid nitrogen and stored at -80 °C. Protein concentrations were measured by Nanodrop UV-Vis spectrophotometer.

3.3.9 Enzymatic reactions *in vitro*

3.3.9.1 Substrate specificity of CpcI and CpcJ

To determine the native amino acid substrate of CpcI and CpcJ, 20 mM L-Glutamic acid, L-Aspartic acid, L-Arginine or L-Lysine was mixed with 10 mM ATP, 10 mM $MgCl_2$, 10 µM CpcI, 1 µM CpcJ and 100 mM Tris-HCl (pH 8.0) and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

3.3.9.2 Idenfication of CpcI-γ-Glu-phosphate generated by CpcI, CpcJ and CpcO *in vitro*

To determine CpcI-γ-Glu-phosphate, 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 µM CpcI, 1 µM CpcJ, 1 µM CpcO, 100 mM Tris-HCl (pH 8.0) and 160 mM *N*-Methylhydroxylamine hydrochloride were mixed and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

3.3.9.3 Detection of CpcI-γ-Glu-semialdehyde produced by CpcI, CpcJ, CpcO and CpcK *in vitro*

CpcI-γ-Glu-semialdehyde was produced *in vitro* reactions as follows. 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 µM CpcI, 1 µM CpcJ, 1 µM CpcO, 1 µM CpcK, 1mM NADPH and 100 mM Tris-HCl (pH 8.0) were mixed and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

3.3.9.4 Analysis of reactions mediated by CpcI, CpcJ, CpcO, CpcK and CpcG/CpcH

The reactions were conducted with 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 µM CpcI, 1 µM CpcJ, 1 µM CpcO, 1 µM CpcK, 0.3 µM CpcG/CpcH, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM Tris-HCl (pH 8.0) at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

3.3.9.5 Reactions mediated by CpcI, CpcJ, CpcO, CpcK, CpcG/CpcH and CpcP

The reactions were conducted with 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 20 µM CpcI, 1 µM CpcJ, 1 µM CpcO, 1 µM CpcK, 0.3 µM CpcG/CpcH, 0.8 µM CpcP, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM Tris-HCl (pH 8.0) and incubated at 30 °C for 3 h. The reaction product was applied to intact protein measurement.

3.3.9.6 *In vitro* synthesis of HHMPA

HHMPA was prepared by *in vitro* reactions as follows. 20 mM L-Glutamic acid, 10 mM ATP, 10 mM $MgCl_2$, 170 µM CpcI, 60 µM CpcJ, 3 µM CpcO, 3 µM CpcK, 0.3 µM CpcG/CpcH, 1 µM CpcP, 1 µM CpcN, 50 mM D-fructose-6-phosphate (F-6-P), 0.1 mM thiamine pyrophosphate (ThDP), 1mM NADPH and 100 mM PBS (pH 6.5) were mixed and incubated at 30 °C overnight.

Fmoc-Cl derivatization was used to determine the production of HHMPA as follows. The reaction was dried in *vacuo*, and was resuspended in 120 µl 83 % acetonitrile. After centrifugation, 10 µl 0.2 M sodium borate (pH 10.5) and 10 µl 100 mM Fmoc-Cl were added to the supernatant and the mixture was incubated at 50 °C for 30 min followed by LC-MS analysis.

3.3.10 Antimicrobial activity test

All the tested pathogens were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ). For micro dilution assays, overnight cultures of each strain were diluted in the growth medium to approximately 10^6 cfu/ml and added to sterile 96-well plates. The cell suspension was treated with the respective compounds in serial dilution. After incubation on a microplate shaker (750 rpm, 37 °C) for 16 h, the growth inhibition was assessed by visual inspection.

3.3.11 Cytotoxicity assay

Cell lines were obtained from the DSMZ and were cultured under conditions as recommended by the depositor. 6×10^3 cells were seeded in 180 µl complete medium in each well of the 96-well plates and treated with compounds in serial dilution after 2 h of equilibration. Both compounds and the internal solvent control was tested in duplicates. After 5 d of incubation, 20 µl of 5 mg/mL MTT (Thiazolyl blue tetrazolium bromide) in PBS was added per well and it was continued to be at 37 °C for 2 h. The medium was then discarded and cells were washed with 100 µl PBS before adding 100 µl 2-propanol/10 N HCl (250:1) in order to dissolve formazan granules. The absorbance at 570 nm was measured using a microplate reader and the percentage relative to the respective methanol control was used to express cell viability. IC_{50} values were determined by sigmoidal curve fitting.

3.3.12 Sequence of cystopipecotide biosynthetic gene cluster in *Cystobacter* sp. SBCb004

The major sequence of the cystopipecotide BGC and BGC13.1 in SBCb004 were provided by 454 technology[10,11]. The gaps in genes *cpcB, cpcC, cpcQ, cpcL* and *BGC13_L* and the gap between *BGC13_B* and *BGC13_C* were refilled by PCR and Sanger sequencing with the primers 9104-gene-gap-F and 9104-gene-gap-R in Table S1 (“gene” refers to the gene with a gap). The complete sequence will be deposited in GenBank when the manuscript is submitted.

3.4 Results and discussion

3.4.1 Discovery of cystopipecotides based on specific MS^2 features

During the metabolic profiling of SBCb004 extracts, two molecular features exhibiting an m/z of 806.44499 $[M+H]^+$ (t_R = 6.89 min) and an m/z of 792.42984 $[M+H]^+$ (t_R = 6.37 min) attracted our attention as both of them obviously consist of at least four amino acids as indicated by MS^2. Besides, their N- and C-termini were proposed to contain the same non-proteinogenic amino acids (Figure 3.4.1 A). The two molecules were referred to as cystopipecotide A (m/z 806.44499) and B (m/z 792.42984) (see below and Figure 3.4.1 B). The difference between both molecules is that the third amino acid in cystopipecotide A may be leucine or isoleucine while it may be valine in case of cystopipecotide B (Figure 3.4.1 A). As no compound matching to these molecules could be found in either, the Dictionary of Natural Products[12] or our in house database Mxbase[13], these amino acids-containing-metabolites were supposed to be novel and were subsequently isolated. Their structures were elucidated by a combination of HR-MS/MS and NMR experiments. The interpretation of the structure elucidation will be shown in Joy Birkelbach's thesis and the NMR tables and spectra are attached in the supplementary information. Cystopipecotides A and B turned out to be hexapeptides with a HHMPA moiety at the N-terminus and a γ-lactam at the C-terminus. The amino acid sequence could be determined to HHMPA, proline, leucine/valine, proline, valine and tryptophan.

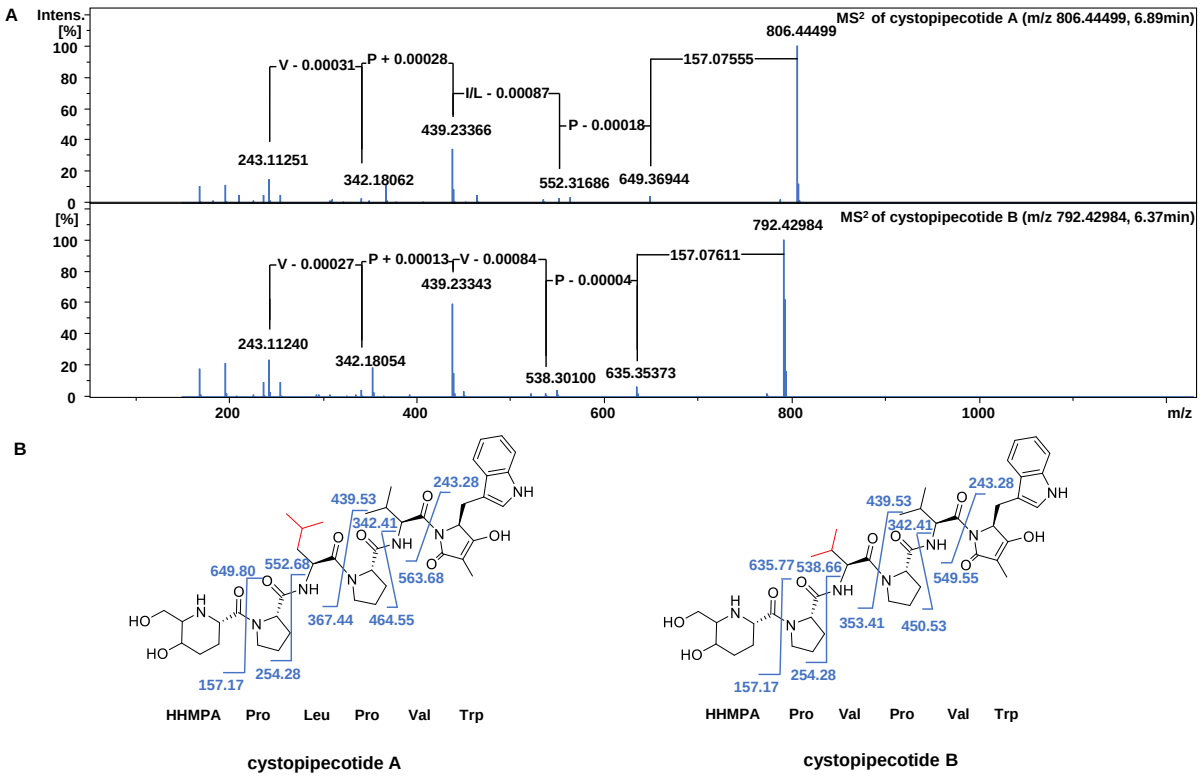


Figure 3.4.1 MS^2 pattern and chemical structures of cystopipecotides A and B. A MS^2 spectra of cystopipecotides A and B. **B** Structure and fragmentation scheme of cystopipecotides A and B in the MS^2 experiments.

3.4.2 Identification of the cystopipecotide biosynthetic gene cluster

The structural similarity between cystopipecotides A and B suggested that they might originate from the same biosynthetic machinery followed by alternative subsequent modifications or using different extender units. To shed light on the origin of the building blocks, L-Leucine-5,5,5-d_3 and L-Valine-d_8 were fed to the SBCb004 wild type strain. Consequently, one leucine and one valine were observed to be incorporated into cystopipecotide A, while no leucine but two valines were incorporated into cystopipecotide B (Figure S1). The results clearly indicate that the structural difference between cystopipecotides A and B originates from the incorporation of different building blocks rather than from alternative modification steps after the common core structure is synthesized.

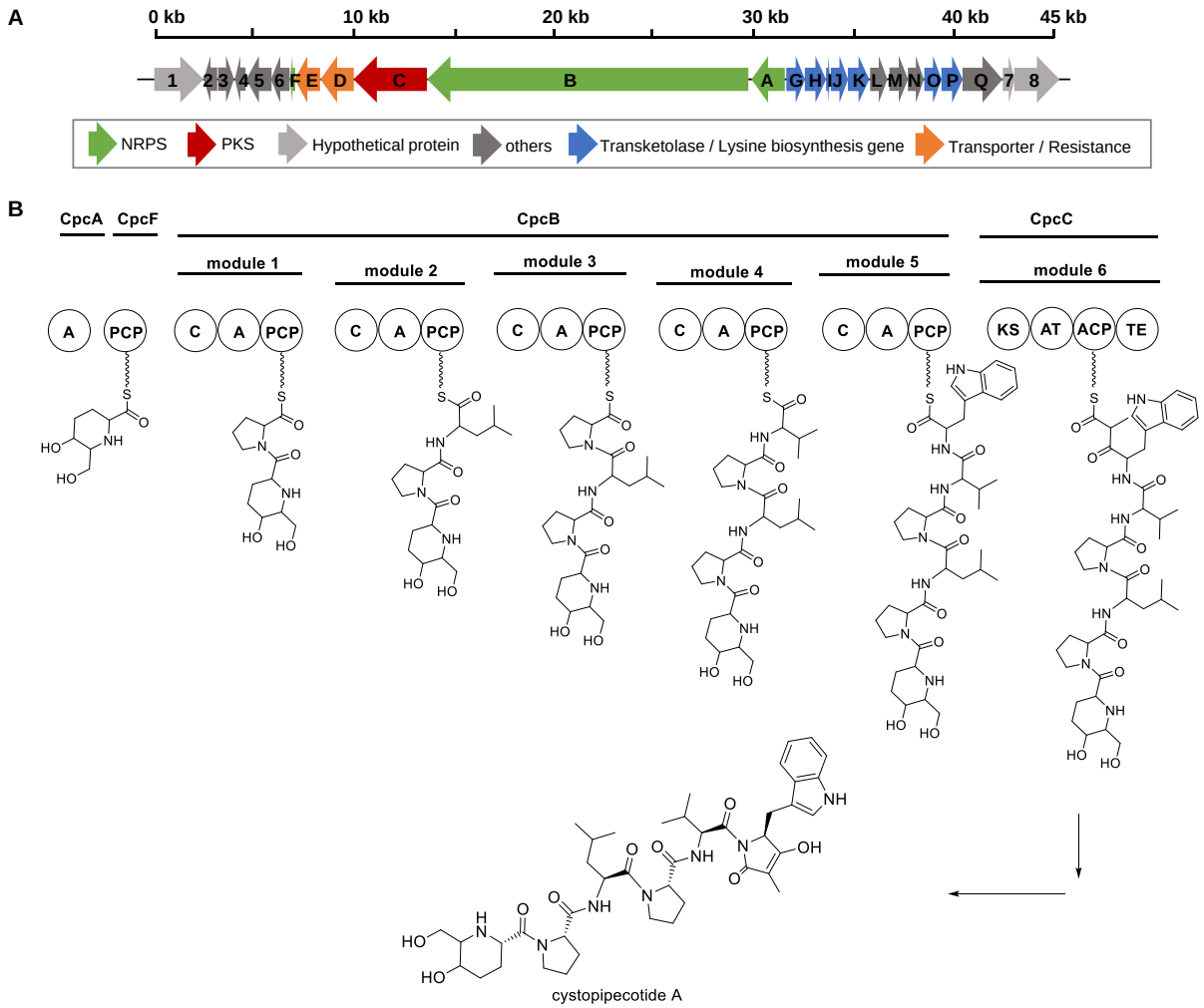


Figure 3.4.2.1 Biosynthetic gene cluster and proposed biosynthetic pathway of the cystopipecotides (cpc). A Genetic organization of the putative biosynthetic gene cluster of cystopipecotides. **B** Putative biosynthetic pathway of cystopipecotide A. A: adenylation. C: condensation. PCP: peptidyl carrier protein. AT: acyl transferase. KS: ketosynthase. ACP: acyl carrier protein. TE: thioesterase.

The C-terminus of cystopipecotides A and B comprises the backbone of a tryptophan and a γ-lactam, which is postulated to be formed by a carboxyl group provided by a C_3 unit and the amine group of tryptophan. Feeding of sodium propionate-1-^{13}C and L-Tryptophan-$^{15}N_2$ led to the enrichment of cystopipecotide A (m/z 806) isotopic peaks m/z 807 and m/z 808, respectively (Figure S2), supporting the proposal of the incorporation of methylmalonyl-CoA and tryptophan. Therefore, the C-terminus is proposed to derive from tryptophan and one methylmalonyl-CoA is assembled to the carbon chain through a Claisen-like condensation. The product is offloaded from the assembly line by lactamization between the terminal carbonyl group and the amine group of tryptophan. The nascent carbonyl group

resulting from the Claisen condensation then spontaneously tautomerizes to an enolate (Figure 3.4.2.1 B).

As no putative precursor peptide in the genome of SBCb004 matches the oligopeptide sequence of P-L/V-P-V-W in cystopipecotide A and B, we proposed that cystopipecotides do not belong to ribosomally synthesized and post-translationally modified peptides, but are generated by a machinery comprising at least five NRPS modules and one PKS module. Thus, a NRPS-PKS hybrid gene cluster encoding one AMP-binding domain (CpcA), five entire NRPS modules (CpcB) and one PKS module (CpcC), was identified in the genome of strain SBCb004 (Figure 3.4.2.1, Table S3). According to our *in silico* analysis, the predicted substrate specificity of A domains in NRPS module 1, 3 and 4 matches amino acids (P, P and V) in the corresponding positions of cystopipecotides (Table S4). AT domain in PKS module (CpcC) may utilize methylmalonyl-CoA (Table S5), which is also consistent to our proposal. *cpcD* and *cpcE* in the downstream of the NRPS-PKS genes putatively encode a transporter and a secretion protein, respectively, which may act during transportation and resistance.

The inactivation of *cpcB* via insertional mutagenesis resulted in the abolishment of the production of both, cystopipecotides A and B. In addition, three peaks exhibiting m/z values of 667.381 (t_R = 6.26 min), m/z of 681.398 (t_R = 6.68 min) and m/z of 766.451 (t_R = 6.73 min) vanished in the HPLC/MS chromatogram of extracts obtained from mutant SBCb004 cpcB$^-$ (Figure 3.4.2.2 A), suggesting that cystopipecotides A and B may not be the only compounds produced by the biosynthetic machinery. These molecules were postulated to be the derivatives of cystopipecotides A and B and were designated as cystopipecotides C-E (Figure 3.4.2.2 B and C). According to the MS/MS analyses, the fifth amino acid valine in cystopipecotides A and B is missing in cystopipecotides C and D (Figure 3.4.2.2 B). The fragment of the C-terminus in cystopipecotides C-E is 26 Dalton lighter than that in A and B (Figure 3.4.2.2 B), which may result from a decarboxylation after the product is liberated from the NRPS-PKS assembly line. Therefore, the C-terminus of cystopipecotides C-E may be a propionyl residue rather than a γ-lactam (Figure 3.4.2.2 C). Based on the analysis of all the possible combinations of amino acids and the varying C-terminus, we proposed the existence of cystopipecotides F-H (Figure 3.4.2.2 C). However, the intensities of the predicted m/z values are too low to be detected with confidence in the chromatogram. As no additional domains like an epimerization domain or a methyltransferase domain exist in any

of the NRPS modules, the amino acids are proposed to remain in the original L-configuration in the final products.

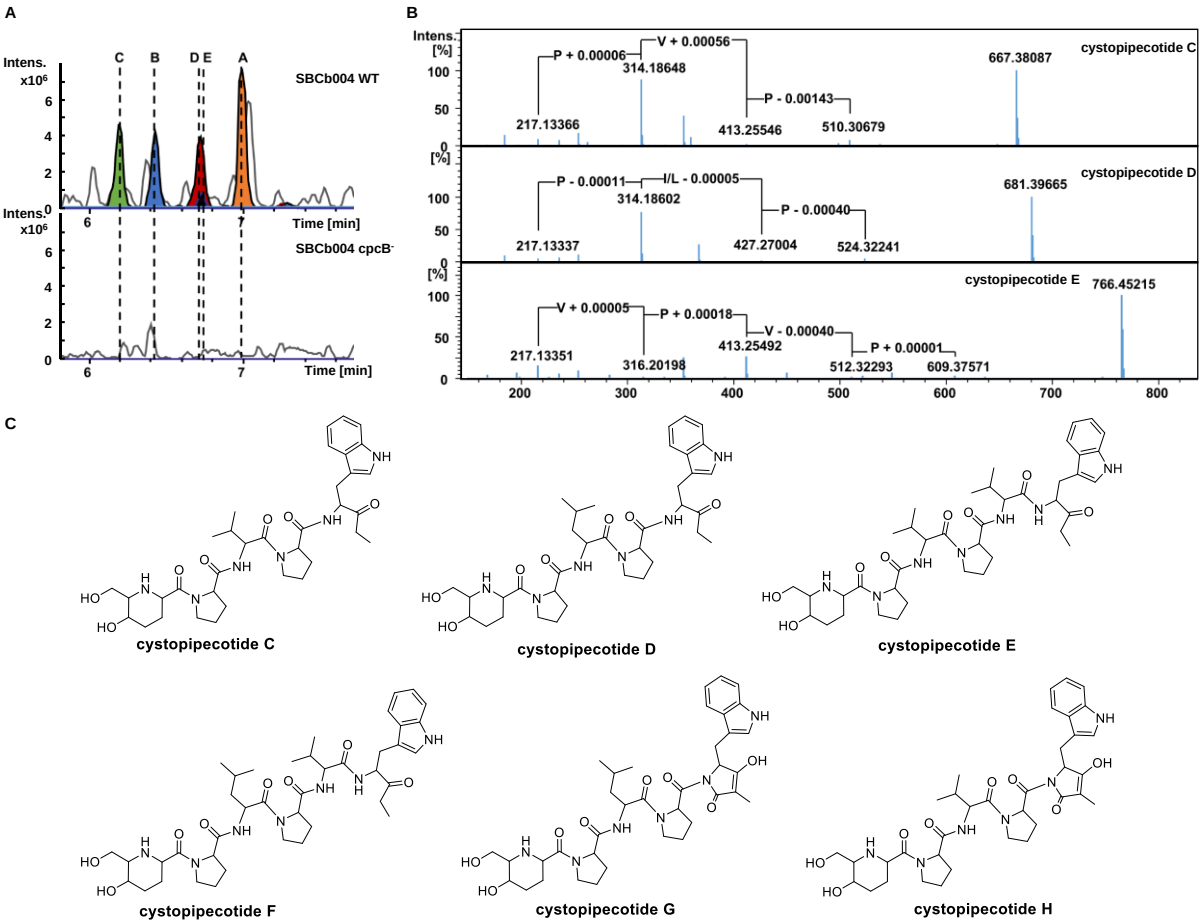


Figure 3.4.2.2 A HPLC-MS analysis of cystopipecotide production in SBCb004 wild type strain (WT) and the mutant SBCb004 cpcB^{-}. The extracted ion chromatograms of cystopipecotides A-E are filled with different colors. **B** MS^2 spectra of cystopipecotides C-E. **C** Predicted structures of cystopipecotides C-H.

Taken together, we propose that the biosynthesis of cystopipecotides A and B is initiated with the formation of HHMPA (see below) which is adenylated by an AMP-binding protein encoded by *cpcA*. The activated HHMPA is then loaded to a dissociated PCP domain, which is probably encoded by the standalone *cpcF* found downstream (Figure 3.4.2.1 A). Subsequently, the C domain of NRPS module 1 within CpcB recognizes the HHMPA-tethered PCP domain and condenses HHMPA to the amino group of PCP-tethered proline, which is delivered by the A domain of NRPS module 1. Then one leucine/valine, one proline,

one valine and one tryptophan are successively added by NRPS modules 2-5 within CpcB, followed by the PKS module (CpcC) mediated incorporation of one methylmalonyl-CoA to the peptide chain via a Claisen condensation. Ultimately, the product is offloaded from the assembly line by γ-lactam formation in cystopipecotides A and B including the spontaneous tautomerization resulting in the enolate (Figure 3.4.2.1 B). In the biosynthesis of cystopipecotides C and D, module 4 of CpcB incorporating valine is presumably skipped. And in the biosynthesis of cystopipecotides C-E, the carboxyl group resulting from the release from the TE using water instead of the intermolecular amide group is decarboxylated after release of the free acid.

3.4.3 Biosynthesis of HHMPA building block

In microbes, pipecolic acid is usually synthesized from lysine[14] (Figure 3.4.3.1 A). It was also reported that the γ-substituted pipecolate could be biosynthesized from *O*-acetyl-L-homoserine and acetoacetate in fungus[15] (Figure 3.4.3.1 B). Besides, there exists a myxobacterial metabolite, chloromyxamide, containing a 6-chloromethyl-5-methoxypipecolic acid (CMPA) building block, the biosynthesis of which originates from L-Glutamic acid[16] (Figure 3.4.3.1 C). However, the pivotal genes involved in the above-mentioned biosynthetic routes could not be found in the cystopipecotide BGC, suggesting an alternative biosynthetic pathway of HHMPA.

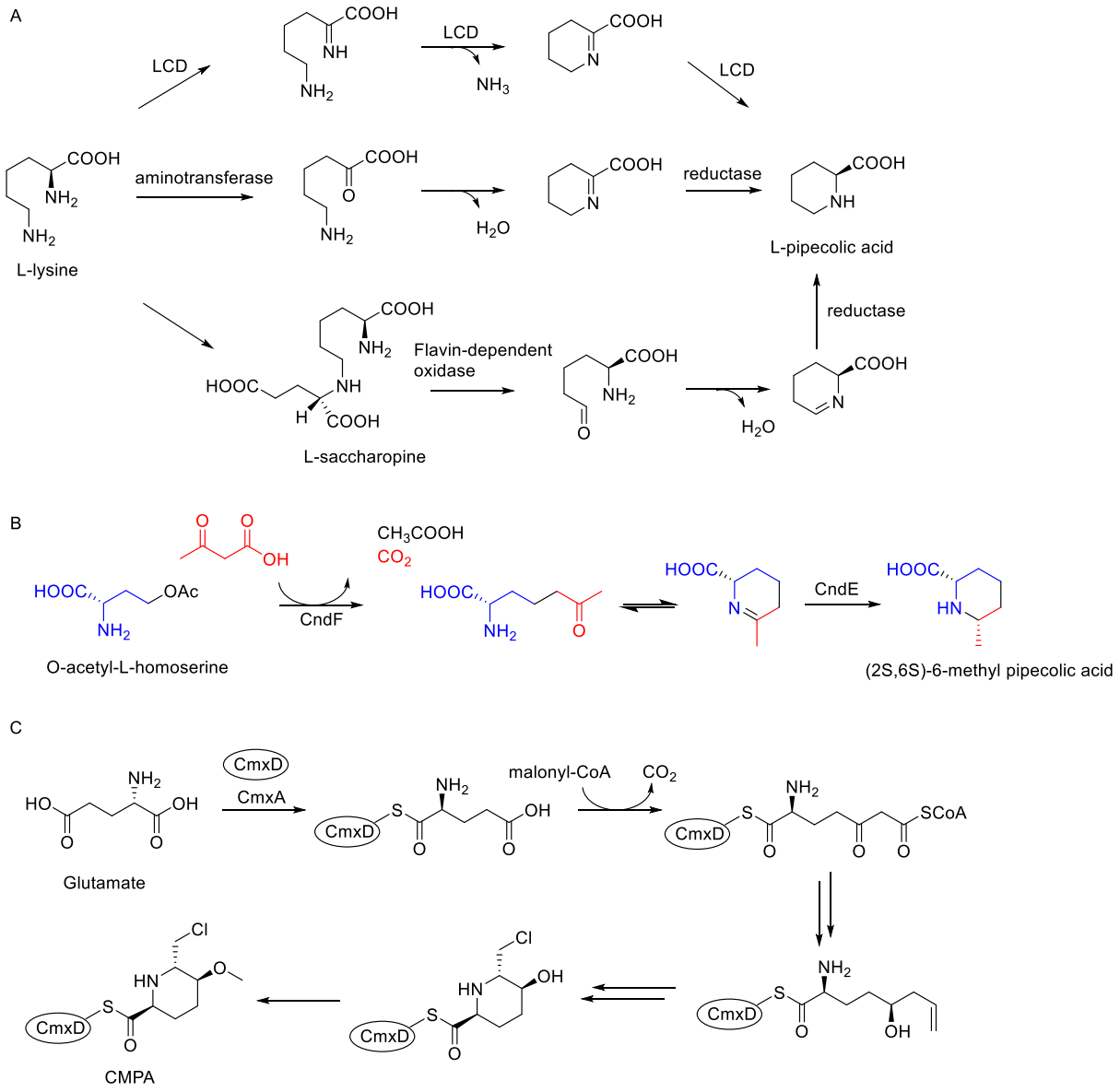


Figure 3.4.3.1 Various derivations of pipecolate biosynthesis in microbe. A Biosynthesis of L-pipecolic acid from L-lysine. LCD: lysine-cyclodeaminase. **B** Biosynthesis of γ-substituted pipecolate from *O*-acetyl-L-homoserine and acetoacetate. **C** Putative biosynthetic pathway of 6-chloromethyl-5-methoxypipecolic acid (CMPA) from L-glutamic acid.

In the upstream region of the NRPS and PKS genes *cpcABC*, a conserved operon (*cpcG* to *cpcP*) which is widely distributed in myxobacteria can be found (Figure 3.4.2.1 A, Table S3). CpcG and CpcH share high similarity (88 % and 99 %, respectively, amino acid level) to two-subunit transketolases which can transfer a C_2 keto-containing unit from a 2-ketose to the first carbon atom of an aldose[17]. Proteins CpcI, CpcJ, CpcO, CpcK and CpcP are homologous to the enzymes involved in lysine biosynthesis through the α-aminoadipate

(AAA) pathway and arginine biosynthesis (LysW, LysX, LysZ/ArgB, LysY/ArgC and LysK/ArgE) (Figure 3.4.3.2 A and B, Table S3). What is missing is the homolog to LsyJ/ArgD, which can transfer the amino group from glutamine to the semialdehyde[18]. Thus, the hypothetical product of CpcI, CpcJ, CpcO, CpcK and CpcP may stay as semialdehyde instead of being further transformed to lysine or arginine. AAA pathway enzymes coupled with transketolases were also reported to synthesize the precursor (2S,6R)-diamino-(5R,7)-dihydroxy-heptanoic acid (DADH) in streptomycetes[19,20]. In this case, the whole AAA pathway cassette besides transketolases are present. The transketolases interfere the process of AAA enzymes by formally transferring a dihydroxyethyl group to the semialdehyde. With the presence of the LysJ homologue, transamination occurs at the nascent keto group introduced by the transketolases. When LysW is released later by the hydrolase, the linear product DADH is formed (Figure 3.4.3.2 C).

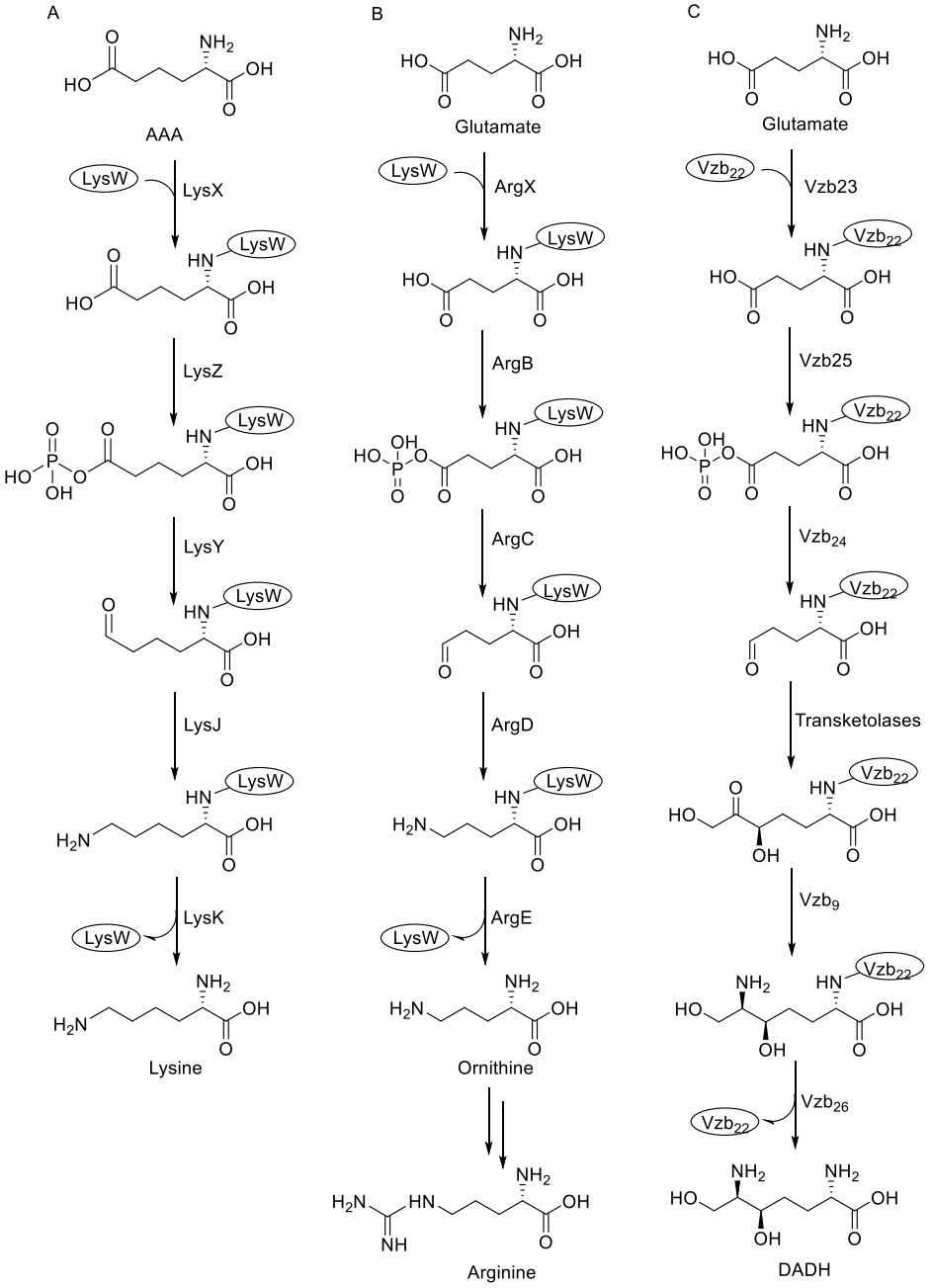


Figure 3.4.3.2 Comparison of LysW mediated biosyntheses. A Lysine biosynthesis via α-aminoadipate (AAA) pathway in *Thermus thermophilus*. **B** Arginine biosynthesis mediated by LysW in *Sulfolobus acidocaldarius*. **C** (2S,6R)-diamino-(5R,7)-dihydroxy-heptanoic acid (DADH) biosynthesis catalyzed by AAA pathway enzymes and transketolases in *Streptomyces* sp. SANK 60404.

Therefore, the biosynthesis of HHMPA was proposed to start from the binding of glutamate to the putative LysW like protein CpcI, catalyzed by the putative RimK family alpha-L-glutamate ligase CpcJ. CpcI may act as a carrier protein and in parallel protects the amino

group of glutamate[18]. The γ-carboxyl group of glutamate is then phosphorylated by the putative LysW-aminoadipate kinase CpcO, followed by a reduction to the semialdehyde catalyzed by the putative N-acetyl-γ-glutamyl-phosphate reductase CpcK. The putative transketolases CpcG and CpcH, in the presence of the essential cofactor thiamin diphosphate (ThDP), might transfer a dihydroxyethyl group from a 2-ketose to the semialdehyde via a nucleophile attack, generating a carbon-carbon bond accompanied by the reduction of the aldehyde to the keto group. Then, LysW-like CpcI releases the amino group hydrolyzed by CpcP. In contrast to the biosynthesis of lysine, arginine and DADH, no transaminase exists in the HHMPA biosynthetic cassette. Thus, the keto group introduced by the transketolases CpcG and CpcH might next attack the free amino group and form the tetrahydropicolinate, which is further saturated by the putative short-chain dehydrogenase/reductase family oxidoreductase CpcN thus yielding HHMPA (Figure 3.4.3.3).

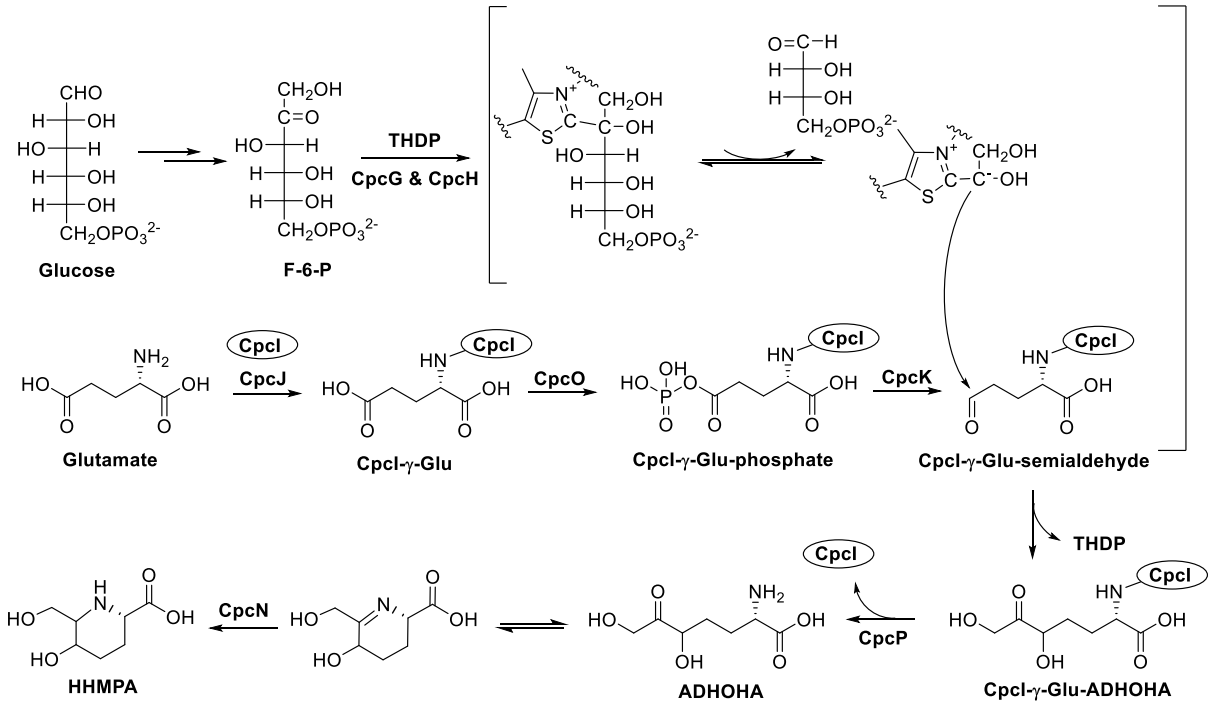


Figure 3.4.3.3 Proposed biosynthesis of 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA). F-6-P: Fructose 6-phosphate. THDP: Thiamin diphosphate. ADHOHA: 2-amino-5,7-dihydroxy- 6-oxoheptanoic acid.

To verify our hypothesis, D-Glucose-1-^{13}C, D-Glucose-2-^{13}C, and L-Glutamic-2,3,3,4,4-d_5-acid as well as L-Lysine-$^{13}C_6$,$^{15}N_2$ hydrochloride and deuterated pipecolic acid were fed to

SBCb004. Intriguingly, only glucoses could be incorporated into cystopipecotides (Figure S2), indicating that the putative transketolases may be involved. Nevertheless, both glucose, lysine and glutamate could participate in various metabolic pathways in myxobacteria. Therefore, the source of HHMPA cannot be evaluated based on feeding experiments only.

In order to investigate whether genes flanking the NRPS-PKS encoding genes *cpcA-C* are involved in the biosynthesis of cystopipecotides as postulated, the insertional mutants of *cpcG*, *cpcK*, *cpcM*, *cpcN*, *cpcP* and *orf6* were obtained, respectively, and their metabolites were checked as well. However, the production of cystopipecotide A in these mutants were not totally abolished like in SBCb004 cpcB$^-$, but were only reduced compared to SBCb004 wild type strain (Figure S3). After interrogation of the SBCb004 genome, we found another BGC, designated as BGC13, containing an additional set of homologs of *cpcA*, *cpcD-F*, *cpcG-P* and *orf4-6*. The difference between BGC13 and BGC of cystopipecotide is that BGC13 does not contain the homolog of the PKS encoding gene *cpcC* besides the NRPS encoding gene in BGC13 may only encode three NRPS modules (Figure 3.4.5, Table S6). To investigate whether BGC13 participates in the biosynthesis of cystopipecotide or other metabolites, the insertional mutants of *BGC13_B* and *BGC13_K* were created. However, no metabolic change were found in the insertional mutant SBCb004 BGC13_B$^-$ and SBCb004 BGC13_K$^-$ (Figure S3), indicating that the NRPS modules in BGC13 may be silent under the applied laboratory condition. As no resistance marker except kanamycin could be used in SBCb004, insertional mutagenesis of *BGC13 G-P* and *orf4-6* in the mutants of their corresponding congeners within cystopipecotide BGC is not achievable. Given in-frame gene deletion in SBCb004 could not be established, it is not possible to investigate whether *cpcG-P* or *BGC13 G-P* are involved in the biosynthesis of HHMPA via *in vivo* knockout. Therefore, experiments *in vitro* to reconstitute the biosynthesis of HHMPA are necessary to obtain more solid evidence.

3.4.4 Reconstitution of HHMPA biosynthesis *in vitro*

To elucidate the biosynthesis of HHMPA, we expressed and purified the recombinant proteins CpcI, CpcJ, CpcO, CpcK, CpcG/CpcH, CpcP and CpcN (Figure S4) and reconstructed the proposed biosynthetic pathway stepwise.

The biosynthesis is supposed to be initiated by the linkage of L-Glutamic acid to LysW like protein CpcI via the amine group of L-Glutamic acid and the γ-carboxyl group of the C-

terminal Glutamic acid in CpcI, catalyzed by CpcJ[21] (Figure 3.4.4 A). As our feeding experiments did not provide any information about the origin of HHMPA, different amino acids including L-Glutamic acid, L-Aspartic acid, L-Arginine and L-Lysine were used to react with CpcI and CpcJ, respectively, to test their substrate specificity. In the intact protein analysis performed by LC-MS, only CpcI tethered Glutamic acid (CpcI-γ-Glu) could be detected (Figure 3.4.4 B), suggesting that L-Glutamic acid is the specific substrate of CpcI and CpcJ.

Next, we added CpcO to potentially phosphorylate the γ-carboxyl group of CpcI-γ-Glu in the presence of ATP (Figure 3.4.4 C). However, the CpcO product CpcI-γ-Glu-phosphate was found to be unstable and easily converted to CpcI-γ-Glu during MS analysis (Figure 3.4.4 D), which is consistent with the report about LysW-γ-AAA-phosphate in lysine biosynthesis[22]. Hence, we added *N*-Methylhydroxylamine to convert the unstable carbonyl phosphate group to the hydroxamate group and observed the expected product CpcI-γ-Glu-hydroxamate (Figure 3.4.4 C and D).

CpcK was then added to the reaction and CpcI-γ-Glu-phosphate was observed to convert to CpcI-γ-Glu-semialdehyde (Figure 3.4.4 E and F). Adjacent to the product (molecular weight: 10271.65 Da), we identified an unexpected peak with the molecular weight of 10253.6372 Da (Figure 3.4.4 F), which may be caused by the instability of the aldehyde group.

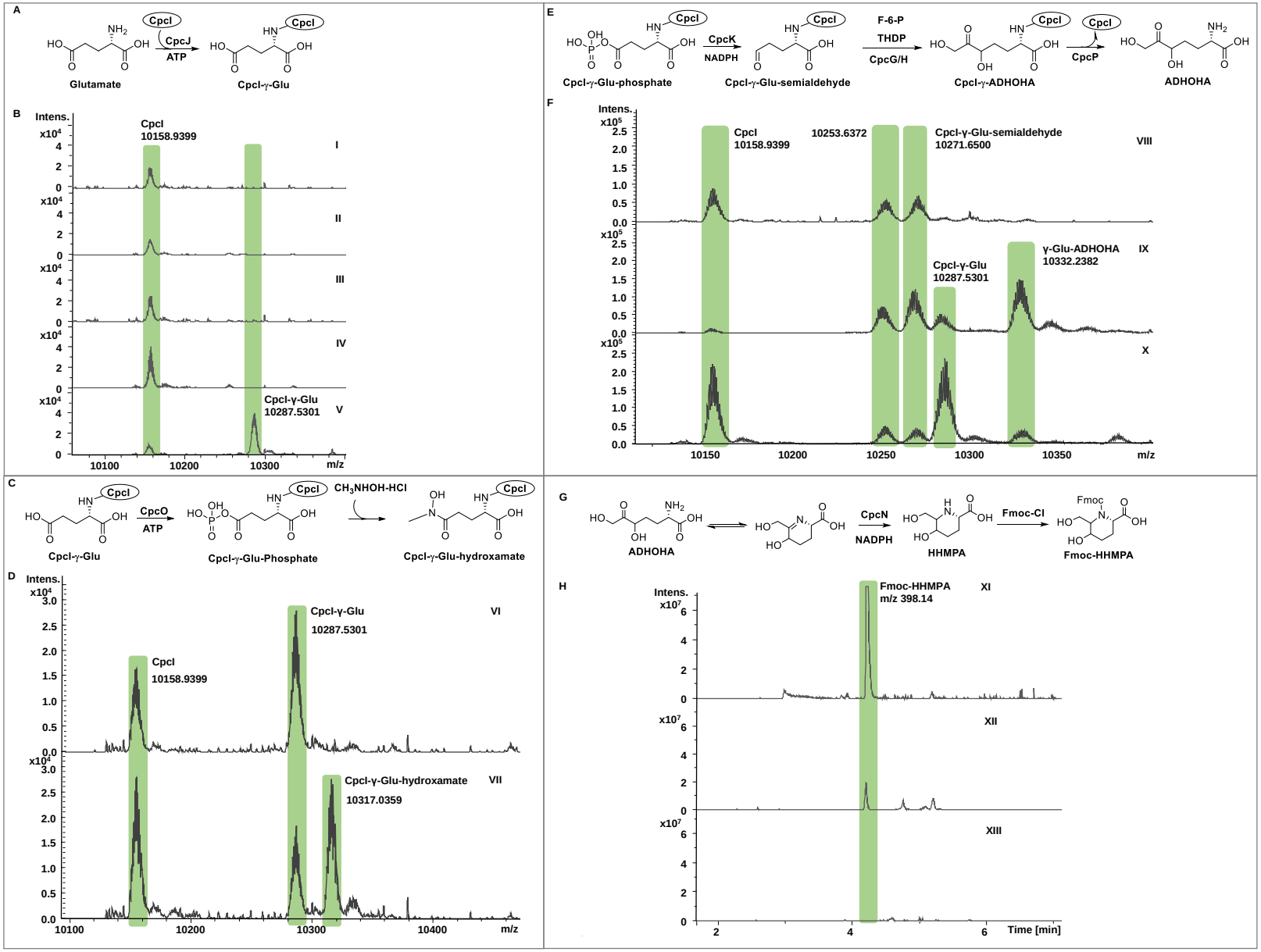

A
Glutamate
CpcJ
ATP
Cpcl-γ-Glu
B
Intens.
Cpcl
10158.9399
Cpcl-γ-Glu
10287.5301
I
II
III
IV
V
m/z
C
Cpcl-γ-Glu
CpcO
ATP
Cpcl-γ-Glu-Phosphate
CH3NHOH-HCl
Cpcl-γ-Glu-hydroxamate
D
Cpcl-γ-Glu
10287.5301
VI
Cpcl-γ-Glu-hydroxamate
10317.0359
VII
E
Cpcl-γ-Glu-phosphate
CpcK
NADPH
Cpcl-γ-Glu-semialdehyde
F-6-P
THDP
CpcG/H
Cpcl-γ-ADHOHA
CpcP
ADHOHA
F
10253.6372
Cpcl-γ-Glu-semialdehyde
10271.6500
VIII
γ-Glu-ADHOHA
10332.2382
IX
X
G
ADHOHA
CpcN
NADPH
HHMPA
Fmoc-Cl
Fmoc-HHMPA
H
Fmoc-HHMPA
m/z 398.14
XI
XII
XIII
Time [min]

Figure 3.4.4 Reconstitution of HHMPA biosynthesis *in vitro*. A Scheme of linkage of Glutamate to CpcI, catalyzed by CpcJ. **B** Deconvoluted mass spectra of CpcI reacted with: Glu (**I**), Arg and CpcJ (**II**), Asp and CpcJ (**III**), Lys and CpcJ (**IV**) and Glu and CpcJ (**V**). The calculated molecular weight (MW) of CpcI, CpcI-γ-Arg, CpcI-γ-Asp, CpcI-γ-Lys and CpcI-γ-Glu are 10159.06 Da, 10315.25 Da, 10274.15 Da, 10287.23 Da and 10288.18 Da, respectively. **C** Scheme of CpcO-mediated reaction and derivatization by *N*-Methylhydroxylamine. **D** Deconvoluted mass spectra of CpcI reacted with Glu, CpcJ and CpcO. **VI**: without *N*-Methylhydroxylamine. **VII**: with *N*-Methylhydroxylamine. The calculated MW of CpcI-γ-Glu-phosphate and CpcI-γ-Glu-hydroxamate are 10368.16 Da and 10317.13 Da. **E** Scheme of reactions successively mediated by CpcK, CpcG/H and CpcP. **F** Deconvoluted mass spectra of CpcI reacted with Glu, CpcJ, CpcO and CpcK (**VIII**), CpcG/H (**IX**), CpcP (**X**). The calculated MW of CpcI-γ-Glu-semialdehyde and CpcI-γ-Glu-ADHOHA (2-amino-5,7-dihydroxy-6-oxoheptanoic acid) are 10272.18 Da and 10332.13 Da. **G** Scheme of CpcN-mediated reaction and derivatization by Fmoc-Cl. **H** Extracted ion chromatogram of Fmoc-HHMPA (m/z 398.14) from: **XI** cystopipecotide A hydrolysis product, **XII** reaction with CpcI, CpcJ, CpcO, CpcK, CpcG/H, CpcP and CpcN, **XIII** reaction withCpcI, CpcJ, CpcO, CpcK, CpcG/H and CpcP.

To investigate the functions of transketolases CpcG/CpcH, a reaction containing CpcI, CpcJ, CpcO, CpcK and CpcG/CpcH was constructed. In this reaction, fructose-6-phosphate was used as the donor of C_2 unit (Figure 3.4.4 E). A peak with the molecular weight of 10332, corresponding to the CpcI linked 2-amino-5,7-dihydroxy-6-oxoheptanoic acid (ADHOHA), was detected (Figure 3.4.4 F), which is consistent with our assumption.

When CpcP was added to the above reaction, the peak corresponding to the CpcI-ADOHOHA decreased while the peaks corresponding to CpcI and CpcI-γ-Glu were found enriched (Figure 3.4.4 F). This result suggests the release of ADHOHA from CpcI, which in turn increased the amount of upstream intermediate.

The released ADHOHA equilibrates with tetrahydropicolinate, which might subsequently be reduced by CpcN (Figure 3.4.4 G). To confirm this assumption, a one pot reaction with all the above-mentioned enzymes was constructed. As the final product HHMPA is a small hydrophilic molecule, which is hard to be detected in reversed LC-MS, Fmoc-Cl was used to derivatize it and facilitate the detection[23]. Cystopipecotide A was hydrolyzed by HCl to release the HHMPA unit as a positive control. In the reaction, a signal with the m/z of 398.14, which corresponds to the derivative Fmoc-HHMPA, was detected in LC-MS (Figure 3.4.4 H). Thus, the whole biosynthetic pathway of HHMPA was successfully reconstituted *in vitro*.

It is noteworthy that CMPA unit from another myxobacterial compound, chloromyxamide, [16], is quite similar to HHMPA and could in theory be transferred from the latter by *O*-methylation at C_5 position and chlorination at C_7 position. The biosynthesis of CMPA was proposed to start from glutamate and the C_2 unit was presumably provided by malonyl-CoA through Claisen condensation[16]. Intriguingly, the proposed operon of HHMPA could be found in the genome of chloromyxamide producing strains *Myxococcus* sp. MCy10608 and *Pyxidicoccus* sp. MCy10649 (Figure 3.4.5, Table S10), while no proposed CMPA producing operon exists in the available genome of SBCb004, implying the possibility of multiple pathways to form similar structures in myxobacteria.

3.4.5 Distribution and diversity of the putative HHMPA biosynthetic operons in myxobacteria

Inspired by the existence of the proposed HHMPA biosynthetic operon in chloromyxamide producing strains, we analyzed all the myxobacterial genomes in our in-house database with the amino acid sequence of CpcG-P and revealed that there are 32 myxobacteria containing the putative biosynthetic HHMPA operon (Table S10). Intriguingly, all these 32 species belong to the suborder of *Cystobacterineae*, whereas 60 *Cystobacterineae* strains, 120 *Sorangiineae* strains and 20 *Nannocystineae* strains in database were checked. This indicates that the putative HHMPA operon may exclusively be distributed in *Cystobacterineae*.

According to the organization of the NRPS-PKS genes adjacent to the HHMPA biosynthetic operon, these BGCs which may generate HHMPA containing compounds could be classified into two major types. Type A refers to the BGCs with PKS module at the downstream of the NRPS modules, suggesting that the C-terminus of the product may form γ-lactam as in cystopipecotides A and B. Type B depicts BGCs containing only NRPS modules. Each type could be further classified into two subtypes based on the position of the loading PCP domain which is putatively responsible for the delivery of HHMPA at the beginning of the biosynthesis. Subtype 1 refers to the case that the loading PCP domain is located adjacent to the loading A domain. While in subtype 2, the loading PCP domain is at the downstream of all the NRPS-PKS genes as in the BGC of cystopipecotide (Figure 3.4.5, Table S10). Among all the 32 species, *Cystobacter* sp. SBCb004 and *Archangium minus* MCy10943 have type A2 and type B2 BGCs simultaneously. All the other strains only exhibit one HHMPA

biosynthetic operon containing BGC. The number of NRPS modules in each BGC is not fixed (Figure 3.4.5, Table S10), thus the size of their products would be variable as well.

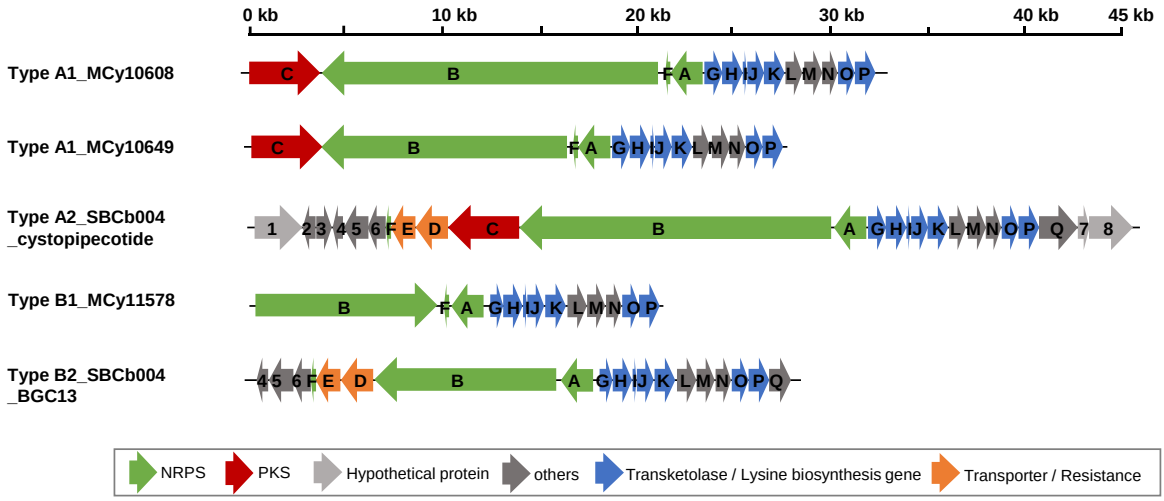


Figure 3.4.5 Organization of representative BGCs containing the putative HHMPA biosynthetic operons.

3.4.6 Bioassay

Cystopipecotide A and B were tested against several bacterial pathogen strains and cancer cell lines. However, neither of them exhibited a significant antimicrobial activity against the tested strains even at the concentration of 64 µg/ml. Only cystopipecotide B was weakly cytotoxic (Table S9). Thus, cystopipecotides may be generated to perform some unknown function, which requires further exploration.

3.4.7 Conclusion

In this study, we uncovered a group of novel peptides containing the non-proteinogenic amino acid HHMPA through metabolomics analysis. The biosynthetic gene cluster of the cystopipecotides was identified by an insertional mutagenesis experiment and a putative biosynthetic pathway was proposed. The biosynthesis of HHMPA, which partially mimics the lysine biosynthesis via the AAA pathway and assembles a C_2 unit by transketolases, was reconstituted *in vitro* and could exhibit an alternative pathway to form substituted pipecolate in microbial natural products.

3.5 References

(1) Wenzel, S. C.; Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. Mol. Biosyst. 2009, 5 (6), 567–574. DOI: 10.1039/b901287g.

(2) Wenzel, S. C.; Müller, R. The biosynthetic potential of myxobacteria and their impact on drug discovery. *Curr. Opin. Drug Discov. Devel.* 2009, *12* (2), 220–230.

(3) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* 2020, *39*, 107480. DOI: 10.1016/j.biotechadv.2019.107480.

(4) Wenzel, S. C.; Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* 2009, *26* (11), 1385–1407. DOI: 10.1039/b817073h.

(5) Beyer, S.; Kunze, B.; Silakowski, B.; Müller, R. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim. Biophys. Acta* 1999, *1445* (2), 185–195.

(6) Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K. J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* 2010, *17* (3), 296–309. DOI: 10.1016/j.chembiol.2010.01.016.

(7) Pogorevc, D.; Popoff, A.; Fayad, A. A.; Wenzel, S. C.; Müller, R. Production profile engineering and precursor directed biosynthesis approaches for production of novel argyrin derivatives, *unpublished results*.

(8) Wright, G. D.; Thompson, P. R. Aminoglycoside phosphotransferases: proteins, structure, and mechanism. *Frontiers in bioscience : a journal and virtual library* 1999, *4*, D9-21.

(9) Deuschle, U.; Kammerer, W.; Gentz, R.; Bujard, H. Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* 1986, *5* (11), 2987–2994.

(10) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, T. P.; Jirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S.

M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005. DOI: 10.1038/nature03959.

(11) Rothberg, J. M.; Leamon, J. H. The development and impact of 454 sequencing. *Nat. Biotechnol.* 2008, *26* (10), 1117–1124. DOI: 10.1038/nbt1485.

(12) *Dictionary of Natural Products*. http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml.

(13) Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* 2014, *31* (6), 768–783. DOI: 10.1039/c3np70127a.

(14) He, M. Pipecolic acid in microbes: Biosynthetic routes and enzymes. *J. Ind. Microbiol. Biotechnol.* 2006, 1–7. DOI: 10.1007/s10295-006-0078-3.

(15) Chen, M.; Liu, C.-T.; Tang, Y. Discovery and Biocatalytic Application of a PLP-Dependent Amino Acid γ-Substitution Enzyme That Catalyzes C-C Bond Formation. *J. Am. Chem. Soc.* 2020, *142* (23), 10506–10515. DOI: 10.1021/jacs.0c03535.

(16) Gorges, J.; Panter, F.; Kjaerulff, L.; Hoffmann, T.; Kazmaier, U.; Müller, R. Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* 2018, *57* (43), 14270–14275. DOI: 10.1002/anie.201808028.

(17) Kochetov, G. A.; Solovjeva, O. N. Structure and functioning mechanism of transketolase. *Biochimica et biophysica acta* 2014, *1844* (9). DOI: 10.1016/j.bbapap.2014.06.003.

(18) Ouchi, T.; Tomita, T.; Horie, A.; Yoshida, A.; Takahashi, K.; Nishida, H.; Lassak, K.; Taka, H.; Mineki, R.; Fujimura, T.; Kosono, S.; Nishiyama, C.; Masui, R.; Kuramitsu, S.; Albers, S. V.; Kuzuyama, T.; Nishiyama, M. Lysine and arginine biosyntheses mediated by a common carrier protein in Sulfolobus. *Nat. Chem. Biol.* 2013, *9* (4), 277–283. DOI: 10.1038/nchembio.120.

(19) Matsuda, K.; Hasebe, F.; Shiwa, Y.; Kanesaki, Y.; Tomita, T.; Yoshikawa, H.; Shin-Ya, K.; Kuzuyama, T.; Nishiyama, M. Genome Mining of Amino Group Carrier Protein-Mediated

Machinery: Discovery and Biosynthetic Characterization of a Natural Product with Unique Hydrazone Unit. *ACS chemical biology* 2017, *12* (1). DOI: 10.1021/acschembio.6b00818.

(20) Hasebe, F.; Matsuda, K.; Shiraishi, T.; Futamura, Y.; Nakano, T.; Tomita, T.; Ishigami, K.; Taka, H.; Mineki, R.; Fujimura, T.; Osada, H.; Kuzuyama, T.; Nishiyama, M. Amino-group carrier-protein-mediated secondary metabolite biosynthesis in Streptomyces. *Nature chemical biology* 2016, *12* (11). DOI: 10.1038/nchembio.2181.

(21) Horie, A.; Tomita, T.; Saiki, A.; Kono, H.; Taka, H.; Mineki, R.; Fujimura, T.; Nishiyama, C.; Kuzuyama, T.; Nishiyama, M. Discovery of proteinaceous N-modification in lysine biosynthesis of Thermus thermophilus. *Nature chemical biology* 2009, *5* (9). DOI: 10.1038/nchembio.198.

(22) Yoshida, A.; Tomita, T.; Fujimura, T.; Nishiyama, C.; Kuzuyama, T.; Nishiyama, M. Structural insight into amino group-carrier protein-mediated lysine biosynthesis: crystal structure of the LysZ·LysW complex from Thermus thermophilus. *The Journal of biological chemistry* 2015, *290* (1), 435–447. DOI: 10.1074/jbc.M114.595983. Published Online: Nov. 12, 2014.

(23) FMOC-Cl as derivatizing agent for the analysis of amino acids and dipeptides by the absolute analysis method.

Chapter 2 Supplementary information

Table S1. Oligonucleotide primers used in this study

Primer	Oligonucleotide 5' to 3' (introduced restriction site)
9101-sc-cpcB-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TCAACCCACCACGATGTCCA
9101-sc-cpcB-R	**GGGCTGGCTTAAAGTCGACTCAATTG**GGTGATGTAAGAGGTGGTG G
9101-sc-cpcG-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAAGTCACGAATTGCCTCC
9101-sc-cpcG-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTAGAACATGGGAACGCCC G
9101-sc-cpcK-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAGACGGGCGGAGAGGTG
9101-sc-cpcK - R	**GGGCTGGCTTAAAGTCGACTCAATTG**CCAGTTCCTGTCGGGATCCA
9101-sc-cpcM-F	**CCAGTCTAGCTATTAATAGGCCTAGG**GTAGGGTTTCTTTCACGCCG
9101-sc-cpcM- R	**GGGCTGGCTTAAAGTCGACTCAATTG**GAGAGCTACTTCTTCCACCC
9101-sc-cpcN-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TAACTGGGACTGGCAGGTA G
9101-sc-cpcN-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TCAGCCATCGACGAGCTG
9101-sc-cpcP-F	**CCAGTCTAGCTATTAATAGGCCTAGG**CTAGTGGATGGTGGAGCAGT
9101-sc-cpcP-R	**GGGCTGGCTTAAAGTCGACTCAATTG**CTAGGCGAGGGTGGGCACC
9101-sc-orf6-F	**CCAGTCTAGCTATTAATAGGCCTAGG**TCACGTCATGTCGCCCAC
9101-sc-orf6-R	**GGGCTGGCTTAAAGTCGACTCAATTG**TAGGAGTTGTCCCATCTCCA G
9101-sc-BGC13_B-F	GCGCCTAGGGCGGACGGGAGCATCTAGTT (MfeI)
9101-sc-BGC13_B-R	CCTCAATTGAGCTGTCAATGAGCGCCAGG (XmaJI)
9101-sc-BGC13_K-F	**CCAGTCTAGCTATTAATAGGCCTAGG**ACCTAGGATCAATTCGCGG
9101-sc-BGC13_K-R	**GGGCTGGCTTAAAGTCGACTCAATTG**ATTGACGACGATCCGGTTTC
cpcB-CK-F	CTCCTGTAGCGCGAACATCA
cpcG-CK-F	ACTTCTCCGCCATCTCGTTC
cpcK -CK-F	GAGGTCAACCACTGTGTCGA
cpcM-CK-F	CACTTCTACGTGGCTGCACT

cpcN-CK-F	CGACCTGGGAAAGAACCACA
cpcP-CK-F	AATGATCCGTTGACGCTCGT
orf6-CK-F	AATATCCCTCGCGCTTACGG
BGC13_B-CK-F	CATCGAGGACAGGTTCACGG
BGC13_K-CK-F	GACAAGCCCGGCTATGATCT
CN-check-R	ACCTCTGACACATGCAGCTC
seq-pblueCN-R	ACCTCTGACACATGCAGCTC
9101-orf4 gap-F	AACAGCTCACCAACCCGTC
9101-orf4 gap-R	GTGGAGCAGGTGATTCAGGA
9101-cpcC gap-F	GTGGATGCCAATCTCCTCGA
9101-cpcC gap-R	CTGTCTTCCTGTTTCCCGGA
9101-cpcB gap1-F	GCTGGGGATCCATCTGCTC
9101-cpcB gap1-R	ACTACGTCGCTCCGGAGT
9101-cpcB gap2-F	GAAGAAGTTGTCGTGCGCAC
9101-cpcB gap2-R	CACGAACTCTTTGAGGCGTG
9101-cpcB gap3-F	CCACCACCTCTTCCATCACC
9101-cpcB gap3-R	GCTGGCGTTGATGCTCAATG
9101-cpcL gap-F	CCCCAGACAGAAATCACCGA
9101-cpcL gap-R	CGCATGTGATTCGGGAGTCT
9101-cpcQ gap-F	GGTTCGTGGAGAAGGAGCTG
9101-cpcQ gap-R	GTAGCAGTCCTGACACTGGC

9101-BGC13_B_C gap-F	CGACTTCCTGTCCACCTTCC
9101 BGC13_B_C gap-R	CACGCTTCGACATCACGTTG
9101-BGC13_L gap-F2	TCGCGACGTTCATCGACTAC
9101-BGC13_L gap-R2	TGAAGCCGACCCAATAGCTG
CpcG-sumohistev-F	**AAAACCTGTATTTTCAGGGC**ACGAGCCATATCCTGATTTC
CpcG-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTATGGGACCTCTCCCATGG
CpcH-CDF-F	**TAATAAGGAGATATACCATG**AGCGAGCCCCAGACGGCGGG
CpcH-CDF-R	**CGAGCTCCCAATTGGGATCC**TTAGTTCCTCCTCGTTGTGA
CpcI-histev-F	**AAAACCTGTATTTTCAGGGC**CAGGTGCAAAAGCAGCTGGA
CpcI-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTATTCGCCCCAGTCTTCCT
CpcJ-histev-F	**AAAACCTGTATTTTCAGGGC**AAGAAGATCGATCTGATCTA
CpcJ-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTATGGATGCACCCCGGCG
CpcK-histev-F	**AAAACCTGTATTTTCAGGGC**ACGCGCGTGGCCGTGCTGGG
CpcK-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTAGATGGGATGCAGGCCGC
CpcN-sumohistev-F	**AAAACCTGTATTTTCAGGGC**GAACTGGGACTGGCAGGTAG
CpcN-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTATCCGAGCACACTCGCTC
CpcO-histev-F	**AAAACCTGTATTTTCAGGGC**AGCGCGTCCTCCATCGTGGT
CpcO-histev-R	**TCGAGTGCGGCCGCAAGCTT**TTACGTGCCCGCATGCGTCG
CpcP-sumohistev-F	**AAAACCTGTATTTTCAGGGC**CGGGCACGTGAGAAGACGAT
CpcP-sumohistev-R	**TCGAGTGCGGCCGCAAGCTT**TTACCCTCCCGACTTCGCGG

The bold letters indicate the overlap sequence to the vector.

Table S2. Plasmids used in this study

Plasmids	**Description/derivation**	**Reference**
pbluekan	*ori (pMB1)*, *bla*, *lacZ*, *neo*	C. Volz, Unpublished
pbluekan-cpcB-HR	pbluekan derivative with *cpcB* homologous region	This study
pbluekan-cpcG-HR	pbluekan derivative with *cpcG* homologous region	This study
pbluekan-cpcK-HR	pbluekan derivative with *cpcK* homologous region	This study
pbluekan-cpcM-HR	pbluekan derivative with *cpcM* homologous region	This study
pbluekan-cpcN-HR	pbluekan derivative with *cpcN* homologous region	This study
pbluekan-cpcP-HR	pbluekan derivative with *cpcP* homologous region	This study
pbluekan-orf6-HR	pbluekan derivative with *orf6* homologous region	This study
pbluekan-BGC13_B-HR	pbluekan derivative with *BGC13_B* homologous region	This study
pbluekan-BGC13_K-HR	pbluekan derivative with *BGC13_K* homologous region	This study
pHisTEV	*ori (pBR322)*, *lac* operator, *neo*, T7 promoter, ribosome binding site, N-HisTag	Novagen
pHisSUMOTEV	*ori (pBR322)*, *lac* operator, *neo*, T7 promoter, ribosome binding site, N-HisTag, SUMO-Tag	Dr. David Owen, unpublished
pCDF-1b	*ori (CloDF13)*, *lac* operator, *spect(variant)*, T7 promoter, ribosome binding site, HisTag	Novagen
pHisSUMOTEV-CpcG	pHisSUMOTEV derivative with CpcG coding sequence	This study
pCDF-CpcH	pCDF-1b derivative with CpcH coding sequence	This study
pHisTEV-CpcI	pHisTEV derivative with CpcI coding sequence	This study

pHisTEV-CpcJ	pHisTEV derivative with CpcJ coding sequence	This study
pHisTEV-CpcK	pHisTEV derivative with CpcK coding sequence	This study
pHisSUMOTEV-CpcN	pHisSUMOTEV derivative with CpcN coding sequence	This study
pHisTEV-CpcO	pHisTEV derivative with CpcO coding sequence	This study
pHisSUMOTEV-CpcP	pHisSUMOTEV derivative with CpcP coding sequence	This study

Table S3. Putative proteins encoded in the cystopipecotide biosynthetic gene cluster

Protein	**Size aa***	**Proposed function of the similar protein**	**Sequence similarity to source**	**Similarity/identity (aa* level)**	**Accession number of the closest homologue**
Orf1	818	penicillin acylase family protein	*Vitiosangium sp*. GDMCC 1.1324	94 / 89 %	WP_108077922.1
Orf2	225	sulfotransferase family protein	*Chloroflexi bacterium* AL-N1	74 / 57 %	NOK60153.1
Orf3	267	3'(2'),5'-bisphosphate nucleotidase CysQ	*Vitiosangium sp*. GDMCC 1.1324	94 / 87 %	WP_108077923.1
Orf4	418	GTP-binding protein	*Pyxidicoccus fallax*	96 / 93 %	WP_169347789.1
Orf5	188	adenylyl-sulfate kinase	*Pyxidicoccus fallax*	94 / 90 %	WP_169347795.1
Orf6	302	sulfate adenylyltransferase subunit CysD	*Vitiosangium sp*. GDMCC 1.1324	98 / 97 %	WP_108075484.1
CpcF	88	acyl carrier protein	*Dyella sp*. L4-6	67 / 44 %	WP_114847001.1

CpcE	417	HlyD family secretion protein	*Vitiosangium sp*. GDMCC 1.1324	94 / 89 %	WP_108077906.1
CpcD	546	DHA2 family efflux MFS transporter permease subunit	*Vitiosangium sp*. GDMCC 1.1324	93 / 86 %	WP_108077907.1
CpcC	1223	PKS (KS_{17-426}, $AT_{523-818}$, $ACP_{913-977}$, $TE_{1002-1217}$)	*Vitiosangium sp*. GDMCC 1.1324	88 / 81 %	WP_108077908.1
CpcB	5360	NRPS (C_{43-479}, $A_{485-1011}$, $PCP_{1022-1086}$, $C_{1107-1541}$, $A_{1545-2069}$, $PCP_{2081-2145}$, $C_{2166-2596}$, $A_{2602-3128}$, $PCP_{3140-3204}$, $C_{3225-3660}$, $A_{3664-4188}$, $PCP_{4198-4262}$, $C_{4283-4719}$, $A_{4723-5247}$, $PCP_{5258-5322}$)	*Archangium violaceum*	79 / 70 %	WP_052519231.1
CpcA	562	AMP-binding protein	*Vitiosangium sp*. GDMCC 1.1324	96 / 92 %	WP_108077857.1
CpcG	316	transketolase	*Pyxidicoccus fallax*	88 / 81 %	WP_169346720.1
CpcH	347	transketolase	*Vitiosangium sp*. GDMCC 1.1324	95 / 92 %	WP_108077855.1
CpcI	69	lysine biosynthesis protein LysW	*Pyxidicoccus fallax*	95 / 87 %	WP_169346718.1
CpcJ	281	RimK family alpha-L-glutamate ligase	*Vitiosangium sp*. GDMCC 1.1324	96 / 94 %	WP_108077854.1
CpcK	347	N-acetyl-gamma-glutamyl-phosphate reductase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077853.1
CpcL	287	phage major capsid protein	*Vitiosangium sp*. GDMCC 1.1324	99 / 97 %	WP_108077852.1

CpcM	310	sulfotransferase	*Pyxidicoccus fallax*	91 / 84 %	WP_169346714.1
CpcN	258	short-chain dehydrogenase/reductase (SDR) family oxidoreductase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077850.1
CpcO	284	[LysW]-aminoadipate kinase	*Vitiosangium sp*. GDMCC 1.1324	95 / 90 %	WP_108077849.1
CpcP	356	M20/M25/M40 family metallo-hydrolase	*Vitiosangium sp*. GDMCC 1.1324	95 / 91 %	WP_108077848.1
CpcQ	91	acyl-CoA dehydrogenase family protein	*Vitiosangium sp*. GDMCC 1.1324	92 / 91 %	WP_108077859.1
Orf7	148	hypothetical protein MEBOL_002563	*Melittangium boletus* DSM 14713	77 / 67 %	ATB29114.1
Orf8	190	hypothetical protein Q664_04790	*Archangium violaceum* Cb vi76	81 / 73 %	KFA94092.1
Orf9	752	PAS domain S-box protein	*Cystobacter fuscus*	78 / 66 %	WP_002623053.1

*aa: amino acid.

Table S4. Substrate specificity analysis of A domains from the cystopipecotide NRPS from C*ystobacter* sp. SBCb004[1]

A domain	**Stachelhaus[1] sequence**	**Neareast Stachelhaus code**	**NRPSPredictor2[2]**	**Consensus**
A_CpcA	ssyyaGlAVK	pip^	hydrophobic-aliphatic*	-
A_CpcB_module1	DVQYIAqVaK	pro	pro	pro

A_CpcB_module2	DAFFLGgTFK	ile	hydrophobic-aliphatic	-
A_CpcB_module3	DVQYIAqVaK	pro	pro	pro
A_CpcB_module4	DALfvGGTFK	val	val	val
A_CpcB_module5	DALvLGLtnK	hpg[#]	ile	-

[1] Predicted substrate specificities were retrieved from outcomes of the antiSMASH 5.0[3] gene cluster analysis.

[*] hydrophobic - aliphatic: gly, ala, val, leu, ile, abu (2-amino-butyric acid), iva (isovaline). [#] hpg: 4-hydoxy-phenyl-glycine. [^] pip: pipecolic acid. -: no consensus.

Table S5. Substrate specificity analysis of AT domain from the cystopipecotide PKS from C*ystobacter* sp. SBCb004[1]

AT domain	ATSignature (Top 3 matches, score)	minova (Top 3 predication, score)
AT_CpcC	Malonyl-CoA (58.3 %), Isobutyryl-CoA (54.2 %), Methylmalonyl-CoA (54.2 %)	Malonyl-CoA (115.7), Methoxymalonyl-CoA (64.3), Methylmalonyl-CoA (62.9)

[1] Predicted substrate specificities were retrieved from outcomes of the antiSMASH 5.0[3] gene cluster analysis.

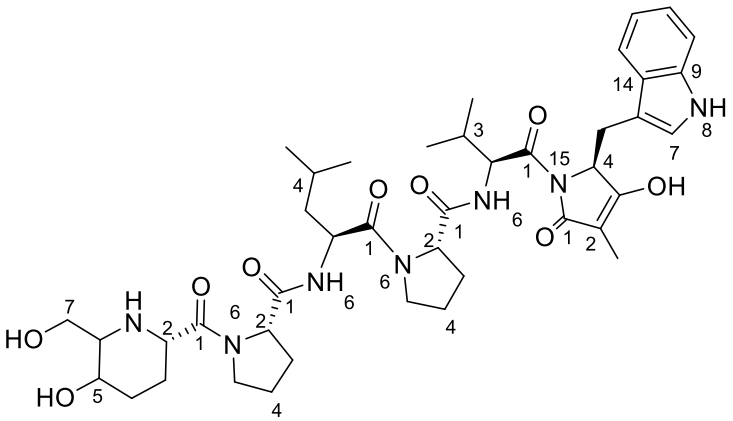
Table S6. Comparison of BGC13 and cystopipecotide BGC (CPC) *in silico*

Protein	Congener in CPC	Putative function	Identity/similarity (aa* level))
BGC13_A	CpcA	AMP-binding protein	88 %/ 94 %
BGC13_B	CpcB	NRPS (C_{-432}, $A_{438-966}$, $PCP_{978-1042}$, $C_{1067-1497}$, $A_{1501-2025}$, $PCP_{2037-2100}$, $C_{2121-2551}$, $A_{2557-3085}$, $PCP_{3099-3163}$, $TE_{3183-3434}$)	78 %/ 85 %
BGC13_D	CpcD	DHA2 family efflux MFS transporter permease subunit	82 %/ 91 %
BGC13_E	CpcE	HlyD family secretion protein	82 %/ 90 %
BGC13_F	CpcF	acyl carrier protein	86 %/ 93 %
BGC13_G	CpcG	transketolase	86 %/ 91 %
BGC13_H	CpcH	transketolase	86 %/ 90 %

BGC13_I	CpcI	lysine biosynthesis protein LysW	89 %/ 95 %
BGC13_J	CpcJ	RimK family alpha-L-glutamate ligase	93 %/ 96 %
BGC13_K	CpcK	N-acetyl-gamma-glutamyl-phosphate reductase	90 %/ 95 %
BGC13_L	CpcL	phage major capsid protein	97 %/ 98 %
BGC13_M	CpcM	sulfotransferase	83 %/ 90 %
BGC13_N	CpcN	short-chain dehydrogenase/reductase (SDR) family oxidoreductase	88 %/ 94 %
BGC13_O	CpcO	[LysW]-aminoadipate kinase	86 %/ 93 %
BGC13_P	CpcP	M20/M25/M40 family metallo-hydrolase	90 %/ 95 %
BGC13_Q	CpcQ	acyl-CoA dehydrogenase family protein	95 %/ 97 %
BGC13_ORF6	ORF6	sulfate adenylyltransferase subunit CysD	96 %/ 98 %
BGC13_ORF5	ORF5	adenylyl-sulfate kinase	89 %/ 95 %
BGC13_ORF4	ORF4	GTP-binding protein	91 %/ 94 %

*aa: amino acid.

Table S7. NMR spectroscopic data for cystopipecotide A (1)[a]



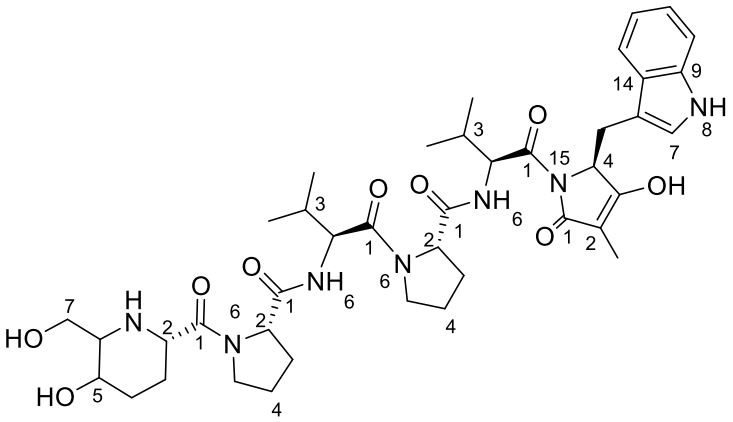
	Pos	δ_X[b], X type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]	N-HMBC[f]
HHMPA	1	169.2, C	-	-	3, 2, 2(Pro-1)	-

	2	51.5, CH	3.86, br d (4.74)	3	3, 4	-
	3a	23.3, CH2	1.87, m	3, 2	4, 2, 5	-
	3b	23.3, CH2	1.65, m	4, 3, 2	5, 2, 1	-
	4a	27.7, CH2	1.52, m	4, 5	3, 6, 2	-
	4b	27.7, CH2	1.74, m	4, 3	3, 6, 2	-
	5	64.8, CH	3.40, m	4, 6	3, 7	-
	6	59.2, CH	3.19, m	5, 7	4	-
	7a	60.3, CH2	3.64, m	6, 7	5	-
	7b	60.3, CH2	3.49, m	7	5	-
	8	40.3, NH	n.a.	-	-	7, 3
Pro-1	1	171.3, C	-	-	4, 2, 7(Leu)	-
	2	58.9, CH	4.37, m	4	2(HHMPA)	-
	3a	28.9, CH2	2.03, u	4, 2	5, 2, 4, 1	-
	3b	28.9, CH2	1.86, m	4, 2	5, 2, 4	-
	4 [g]	24.6, CH2	1.76, m	2	5, 2, 1	-
	5a [g]	46.7, CH2	3.62, m	5	4, 2	-
	5b [g]	46.7, CH2	3.46, m	5	4, 2	-
	6 [g]	129.2, N	-	-	-	2, 4, 3
Leu	1 [g]	170.8, C	-	-	3, 2, 2(Pro-2)	-
	2	48.6, CH	4.53, m	3, 7	3, 7	-
	3a	40.4, CH2	1.52, m	2	6, 5, 4, 2, 7, 1	-
	3b	40.4, CH2	1.45, br d (3.97)	2	6, 5, 4, 2, 7, 1	-
	4	24.0, CH	1.72, br d (6.28)	6, 5	6, 5, 3, 2	-
	5	23.3, CH2	0.90, m	4	6, 5, 3	-
	6	21.6, CH3	0.89, m	4	3, 4	-
	7	117.5, NH	8.09, br d (7.69)	2	3, 2	2, 3
Pro-2	1 [g]	170.8, C	-	-	2	-
	2	59.2, CH	4.55, m		3, 1	
	3	28.6, CH2	2.01, m	2	4, 1	
	4 [g]	24.6, CH2	1.76, m	2	5, 2, 1(Leu)	
	5a [g]	46.7, CH2	3.62, m	5	4, 2	
	5b [g]	46.7, CH2	3.46, m	5	4, 2	
	6 [g]	129.2, N	-	-	-	3, 4
Val	1	169.7, C	-	-	3	-
	2	55.1, CH	5.66, br s	3, 6	4, 5	-

	3	30.8, CH	1.98, m	4, 5, 2	4, 5, 1	-
	4	16.7, CH3	0.75, m	3	5, 3, 2	-
	5	19.7, CH3	0.78, m	3	4, 3, 2	-
	6	110.4, NH	7.61, br s	2	3	3, 3(Pro-2)
MMTrp	1	172.4, C	-	-	2-Me, 4	-
	2	94.1, C	-	-	2-Me	-
	2-Me	6.3, CH3	1.20, s	-	2, 1, 3	-
	3	182.1, C	-	-	2-Me, 5, 4	-
	4	60.0, CH	4.20, br s	5	5, 7	-
	5a	24.4, CH2	3.50, m	5, 4	4, 7	-
	5b	24.4, CH2	3.22, m	5, 7	4, 7	-
	6	108.2, C	-	-	5, 4, 7, 13, 8	-
	7	123.8, CH	6.80, m	8	5, 8, 4, 6, 9	-
	8	n.a., NH	10.66, br s	7	6, 7, 14	-
	9	135.5, C	-	-	7, 12, 13, 8	-
	10	110.9, CH	7.20, m	12	11, 13, 14	-
	11	117.9, CH	6.82, m	12, 13	12, 10, 13, 14	-
	12	120.2, CH	6.94, t (7.50, 7.50)	11, 10	11, 13, 14, 9	-
	13	118.8, CH	7.44, br d (7.95)	11	11, 12, 6, 11, 14, 9	-
	14	128.4, C	-	-	5, 11, 12, 10, 13, 8	-
	15	163.7, N	-	-	-	5

[a]Recorded in DMSO-d_6. [b]Acquired at 175 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 39.51 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 2.45 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbon. [f]Nitrogen showing HMBC correlation to indicated proton. [g]Carbons/ protons showing the same signal.

Table S8. NMR spectroscopic data for cystopipecotide B (2)[a]



	Pos	δ_X[b], X type	δ_H[c], (*J* in Hz)	COSY[d]	HMBC[e]
HHMPA	1	171.9, C	-	-	2
	2	51.1, CH	3.61, m	3	3, 4, 6
	3a	24.2, CH2	1.75, m	3, 2, 4	4
	3b	24.2, CH2	1.48, m	4, 3, 2	4
	4a	28.6, CH2	1.64, m	3, 5	6, 2
	4b	28.6, CH2	1.23, m	5, 3	6, 2
	5	65.5, CH	3.28, m	4	6, 7
	6	58.3, CH	2.84, m	2, 7	4, 7, 2
	7a	61.8, CH2	3.56, m	7	5, 6
	7b	61.8, CH2	3.35, m	7	5, 6
	8	n.a., NH	n.a.	-	-
Pro-1	1	171.0, C	-	-	3, 2(Val-1), 6(Val-1)
	2	59.3, CH	4.50, br d (6.92)	3, 4	-
	3a	28.6, CH2	1.99, m	4, 2	1
	3b	28.6, CH2	1.88, m	4, 2	4
	4[g]	24.3, CH2	1.87, m	2, 5, 3	3
	5	46.4, CH2	3.48, m	4	4
	6	n.a., N	-	-	-
Val-1	1[g]	169.3, C	-	-	3, 2
	2	55.2, CH	4.35, m	3, 6	3, 4
	3	40.4, CH2	1.52, m	2, 4, 5	2, 4, 5
	4	17.9, CH3	0.90, m	3	5, 3, 2
	5	19.2, CH3	0.94, m	3	2, 3, 4

	6	n.a., NH	7.81, br d (8.33)	2	3, 1(Pro-1)
Pro-2	1 [g]	169.3, C	-	-	3(Val-2)
	2	58.9, CH	4.41, m	3	3, 4
	3a	28.3, CH2	1.97, m	3, 2	4, 5
	3b	28.3, CH2	1.81, m	3, 2, 4	4, 5
	4 [g]	24.3, CH2	1.87, m	3, 5	3, 2
	5	46.8, CH2	3.62, m	4	4, 3
	6	n.a., N	-	-	-
Val-2	1	169.7, C	-	-	3
	2	57.4, CH	5.54, m	6	-
	3	31.1, CH	2.00, m	4, 5	4, 5
	4	19.5, CH3	0.78, br d (6.41)	3	5, 3
	5	16.7, CH3	0.75, br d (6.92)	3	4, 3
	6	n.a., NH	8.11, m	2	-
MMTrp	1	172.3, C	-	-	2-Me, 4
	2	89.8, C	-	-	2-Me
	2-Me	6.4, CH3	1.17, m	4	-
	3	186.9, C	-	-	2-Me, 5, 4
	4	60.8, CH	3.92, m	2-Me, 5	5
	5a	24.5, CH2	3.39, m	5, 4, 7	4
	5b	24.5, CH2	3.15, br d (13.46)	5, 7, 4	4
	6	108.8, C	-	-	5, 4, 7, 13, 8
	7	123.2, CH	6.80, m	5, 8	5, 8
	8	n.a., NH	10.53, br s	7, 13	6, 7, 14, 9
	9	135.2, C	-	-	7, 12, 13, 8
	10	110.3, CH	7.19, d (8.08)	12	11, 13, 14
	11	117.2, CH	6.80, m	12, 13	12, 10
	12	119.7, CH	6.92, t (7.50, 7.50)	11, 10	11, 13, 9
	13	118.8, CH	7.48, br d (7.95)	11, 8	10, 12, 6, 14, 9
	14	127.9, C	-	-	5, 7, 10, 13, 8
	15	n.a., N	-	-	-

[a]Recorded in DMSO-d_6. [b]Acquired at 175 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 39.51 ppm). [c]Acquired at 700 MHz, adjusted to the solvent signal of DMSO-d_6 (δ_H 2.45 ppm). [d]Proton showing COSY correlation to indicated protons. [e]Proton showing HMBC correlation to indicated carbons. [f]Carbons/ protons showing the same signal.

Table S9. Minimum inhibitory concentrations (MIC, µg/ml) value and half maximal inhibitory concentration (IC_{50}, µg/ml) of cystopipecotides A and B

Test organisms	cystopipecotide A	cystopipecotide B
Staphylococcus aureus	≥ 64 µg/ml	≥ 64 µg/ml
Mycobacterium smegmatis	≥ 64 µg/ml	≥ 64 µg/ml
Bacillus subtilis	≥ 64 µg/ml	≥ 64 µg/ml
Citrobacter freundii	≥ 64 µg/ml	≥ 64 µg/ml
Acinetobacter baumannii	≥ 64 µg/ml	≥ 64 µg/ml
Candida albicans	≥ 64 µg/ml	≥ 64 µg/ml
***Escherichia coli* acrB**	≥ 64 µg/ml	≥ 64 µg/ml
***Escherichia coli* DSM1116**	≥ 64 µg/ml	≥ 64 µg/ml
Cryptococcus neoformans	≥ 64 µg/ml	≥ 64 µg/ml
Pseudomonas aeruginose	≥ 64 µg/ml	≥ 64 µg/ml
Pichia anomala	≥ 64 µg/ml	≥ 64 µg/ml
HepG2	> 37 µg/ml	2.58 µg/ml
HCT-116	36.11 µg/ml	10.53 µg/ml
KB3.1	18.44 µg/ml	5.36 µg/ml

Table S10. Distribution and diversity of putative HHMPA biosynthetic operon containing myxobacteria

Strain	Organization of NRPS-PKS	Position of PCP_L domain	Type
Myxococcus fulvus MCy8288	6 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcus sp. MCy10608	6 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy10984	5 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy12228	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10653	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy12733	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1

Cystobacterineae strain MCy12716	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10649	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Cystobacterineae strain MCy10585	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Coallococcus coralloides MCy10705	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus coralloides MCy12446	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy11561	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy8332	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcaceae strain MCy12473	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Corallococcus sp. MCy9049	4 NRPS modules, 1 PKS module	Adjacent to A_L domain	A1
Myxococcus sp. MCy11578	3 NRPS modules	Adjacent to A_L domain	B1
Cystobacterineae strain MCy10644	3 NRPS modules	Adjacent to A_L domain	B1
Cystobacterineae strain MCy9003	3 NRPS modules	Adjacent to A_L domain	B1
Pyxidicoccus sp. MCy8408	1 NRPS modules	Adjacent to A_L domain	B1
Myxococcus fulvus MCy9270	incomplete NRPS sequence	Adjacent to A_L domain	unknown
Myxococcus fulvus MCy9280	incomplete NRPS sequence	Adjacent to A_L domain	unknown
Cystobacter sp. SBCb004	5 NRPS modules, 1 PKS module	downstream	A2

Vitiosangium cumulatum MCy10943	5 NRPS modules, 1 PKS module	downstream	A2
Archangium sp. MCy8383	4 NRPS modules	downstream	B2
Cystobacter sp. SBCb004	3 NRPS modules	downstream	B2
Vitiosangium cumulatum MCy10943	3 NRPS modules	downstream	B2
Cystobacterineae strain MCy10597	3 NRPS modules	downstream	B2
Corallococcus coralloides MCy6431	3 NRPS modules	downstream	B2
Cystobacterineae strain MCy5730	incomplete NRPS sequence	downstream	unknown
Pyxidicoccus fallax MCy8396	incomplete NRPS sequence	downstream	unknown
Cystobacterineae strain MCy8401	incomplete NRPS sequence	downstream	unknown
Cystobacter armeniaca MCy5465	incomplete NRPS sequence	downstream	unknown
Cystobacterineae strain MCy8369	incomplete NRPS sequence	unknown	unknown
Archangium minus MCy1093	incomplete NRPS sequence	unknown	unknown

A_L: loading A domain. PCP_L: loading PCP domain.

Supplementary figures

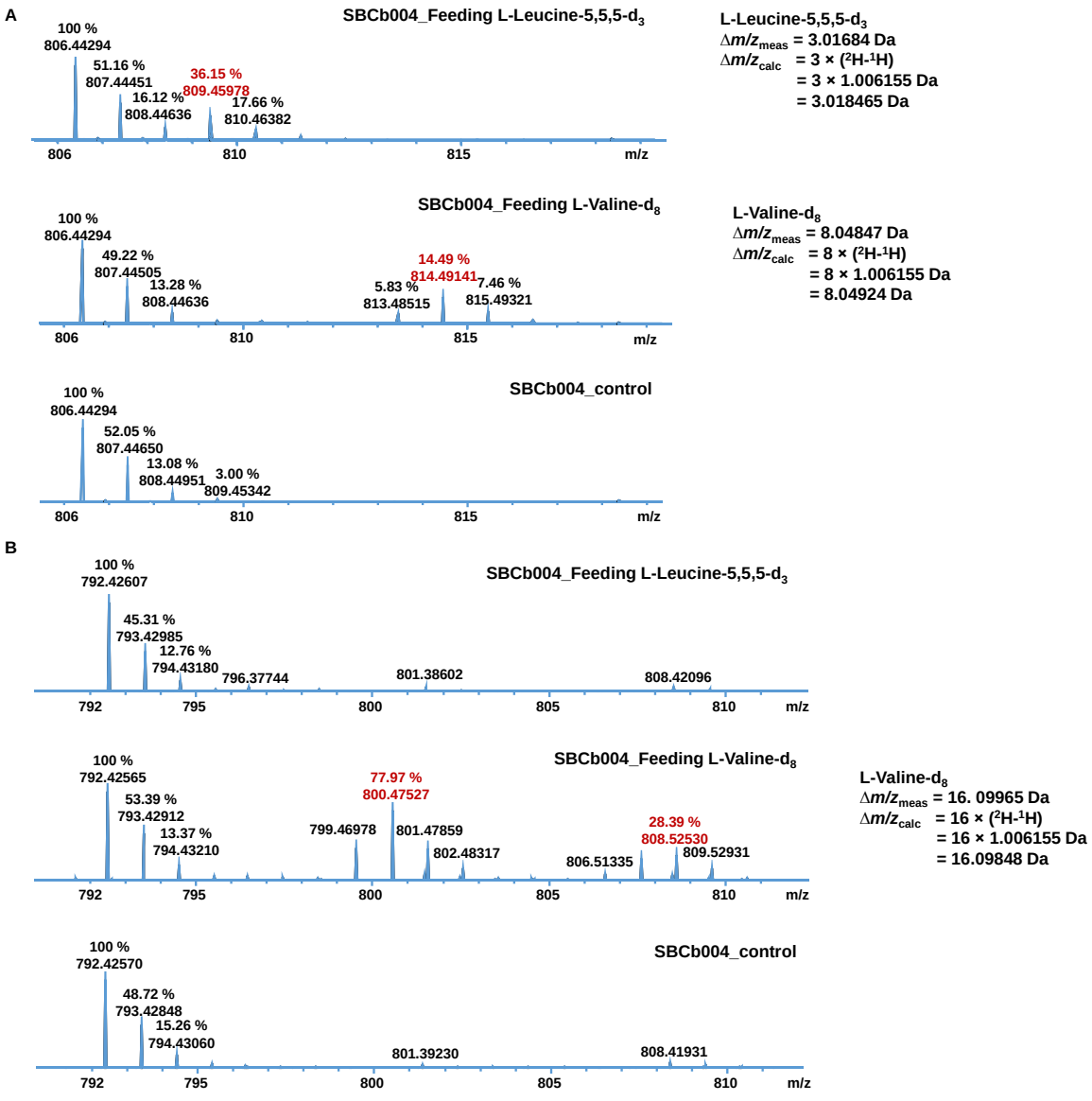


Figure S1. Feeding studies with L-Leucine-5,5,5-*d*$_3$ and L-Valine-*d*$_8$. **A** Mass spectra of cystopipecotide A (806 m/z) with L-Leucine-5,5,5-*d*$_3$ and L-Valine-*d*$_8$ fed to the culture. B Mass spectra of cystopipecotide B (792 m/z) with L-Leucine-5,5,5-*d*$_3$ and L-Valine-*d*$_8$ fed to the culture.

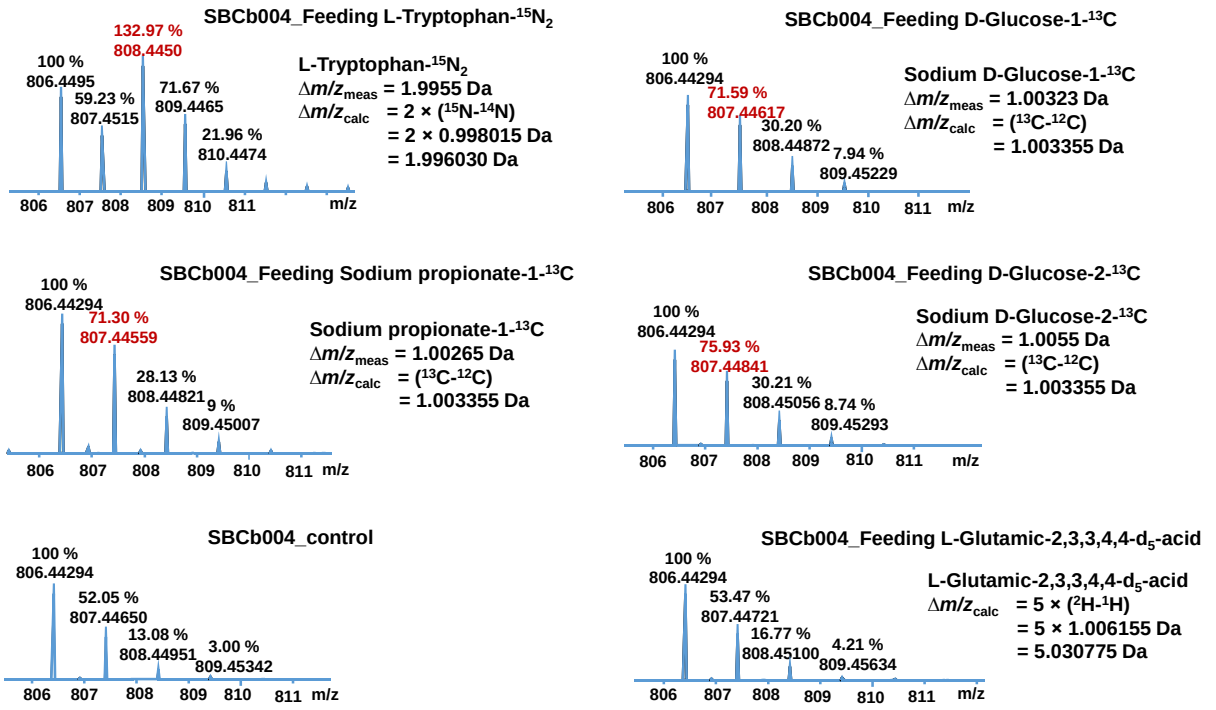


Figure S2. Mass spectra of cystopipecotide A (806 m/z) with L-Tryptophan-$^{15}N_2$, D-Glucose-1-^{13}C, D-Glucose-2-^{13}C, Sodium propionate-1-^{13}C and L-Glutamic-2,3,3,4,4-d_5-acid fed to the culture.

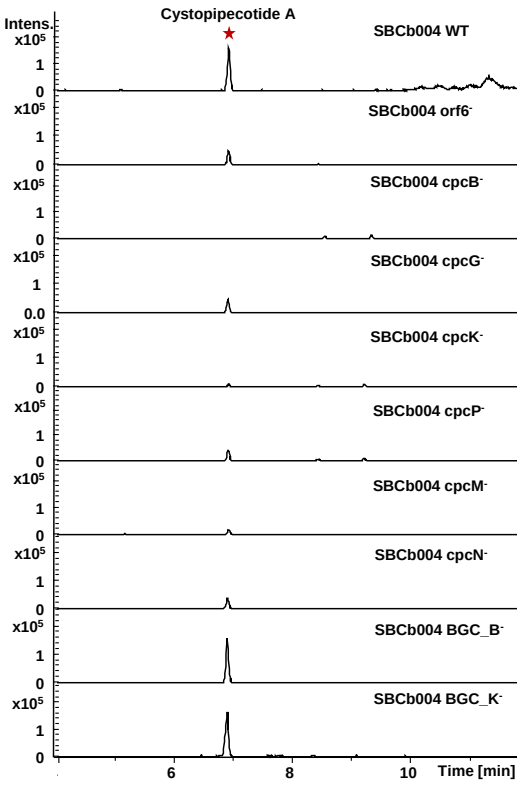


Figure S3. The extracted ion chromatograms of cystopipecotide A in SBCb004 wild type strain (WT) and the mutants SBCb004 orf6$^-$, SBCb004 cpcB$^-$, SBCb004 cpcG$^-$, SBCb004 cpcK$^-$, SBCb004 cpcP$^-$, SBCb004 cpcM$^-$, SBCb004 cpcN$^-$, SBCb004 BGC13_B$^-$ and SBCb004 BGC13_K$^-$.

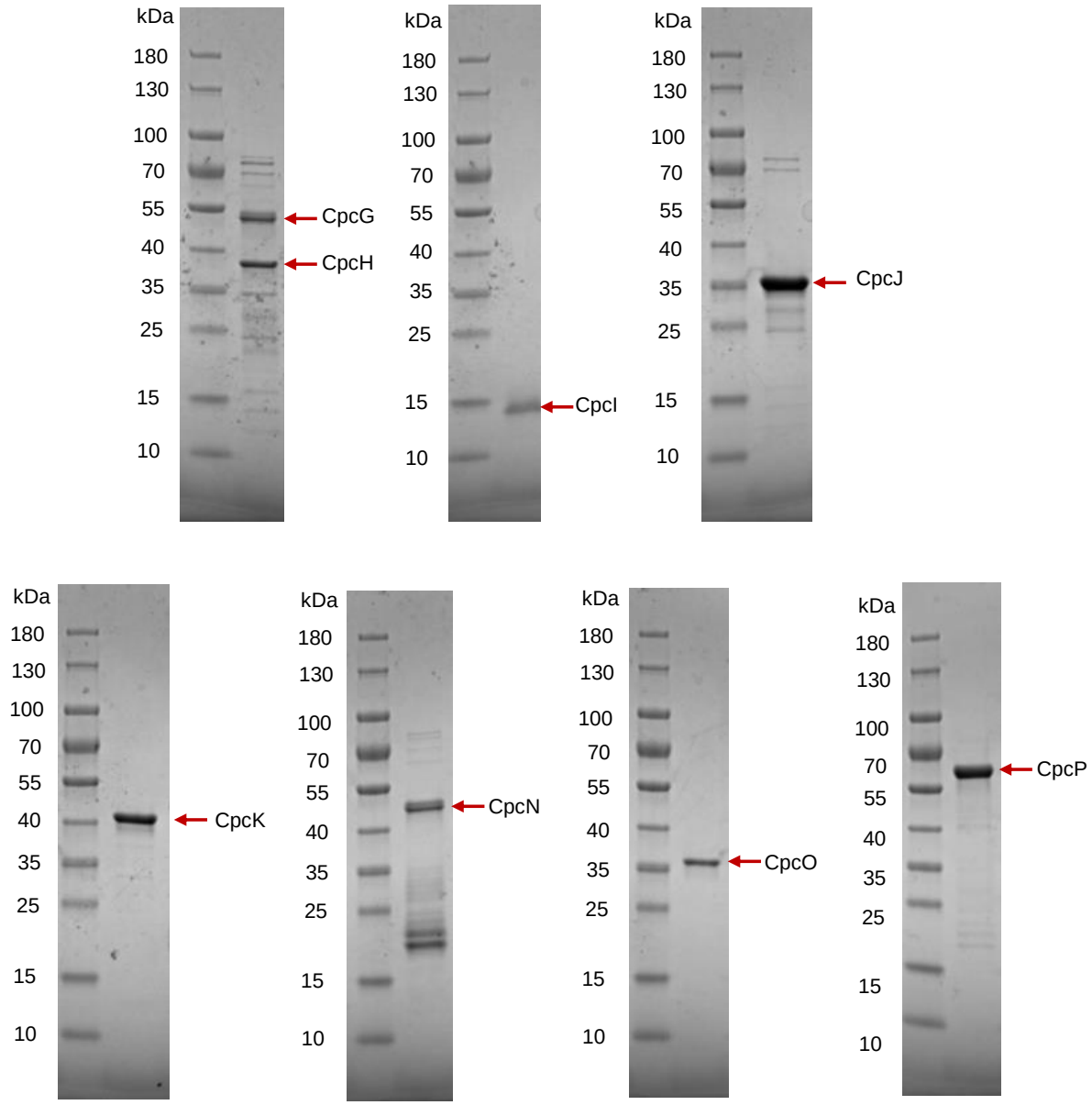
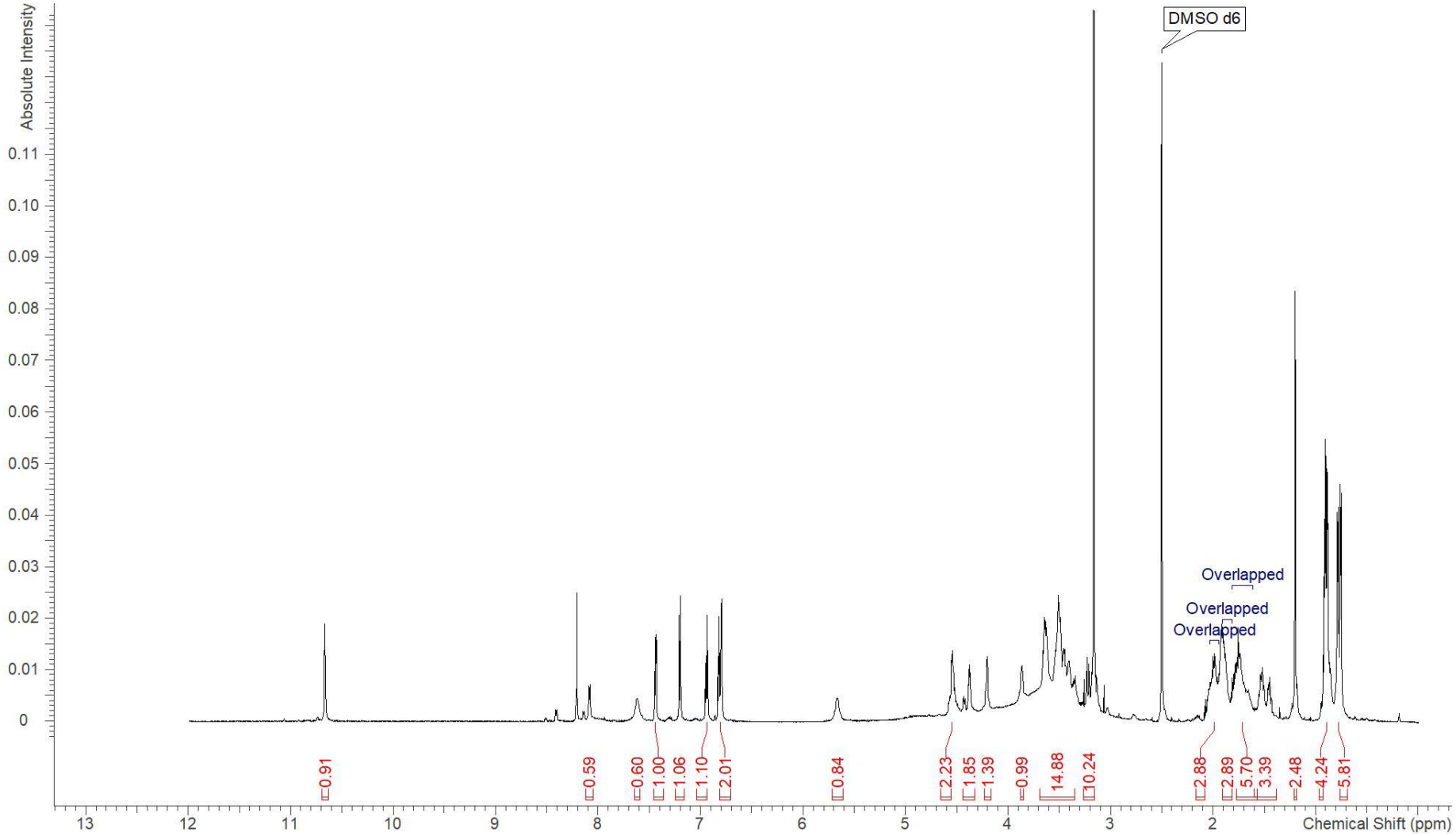
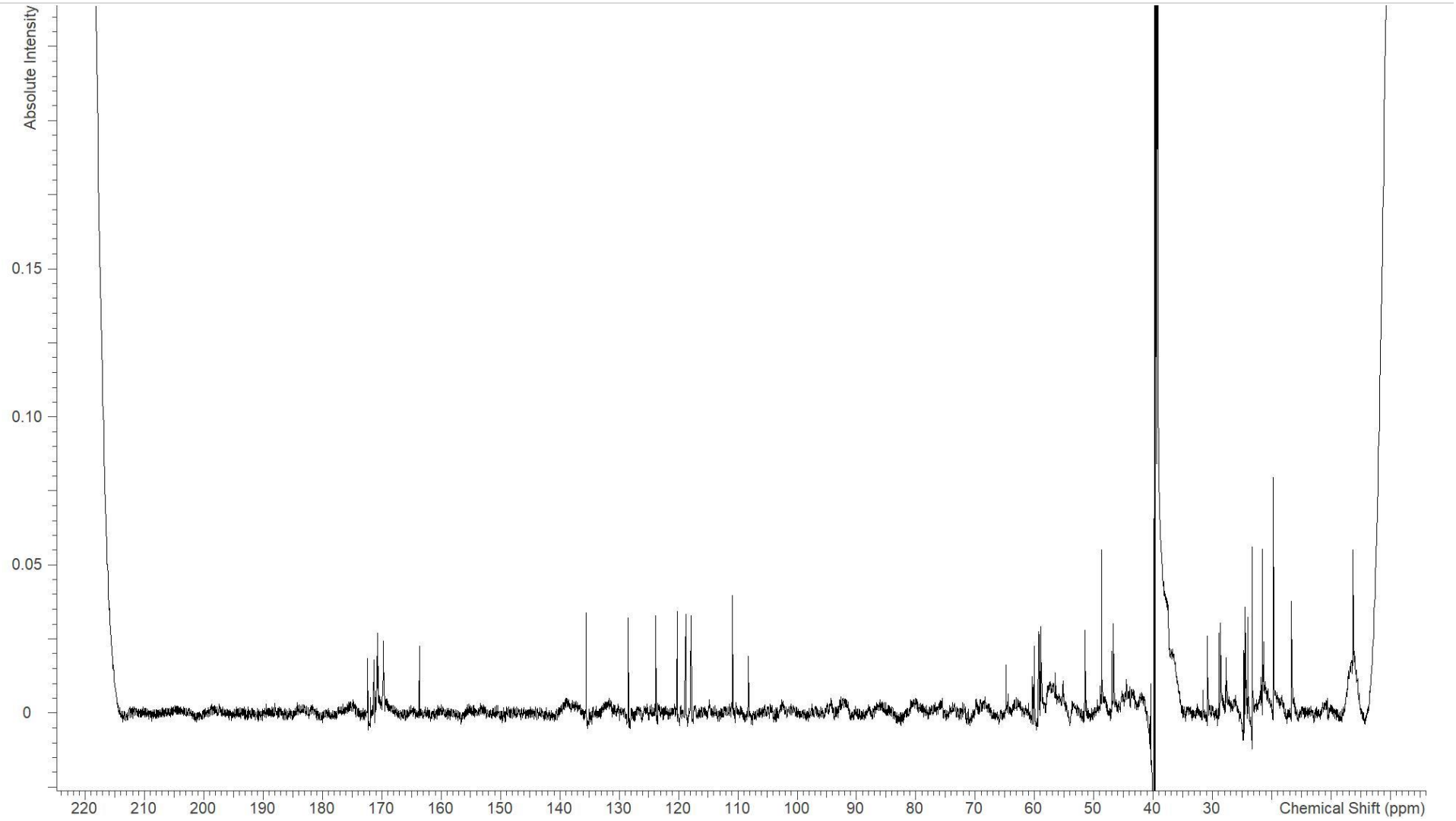


Figure S4. SDS-PAGE analysis of CpcG/H, CpcI, CpcJ, CpcK, CpcN, CpcO and CpcP.

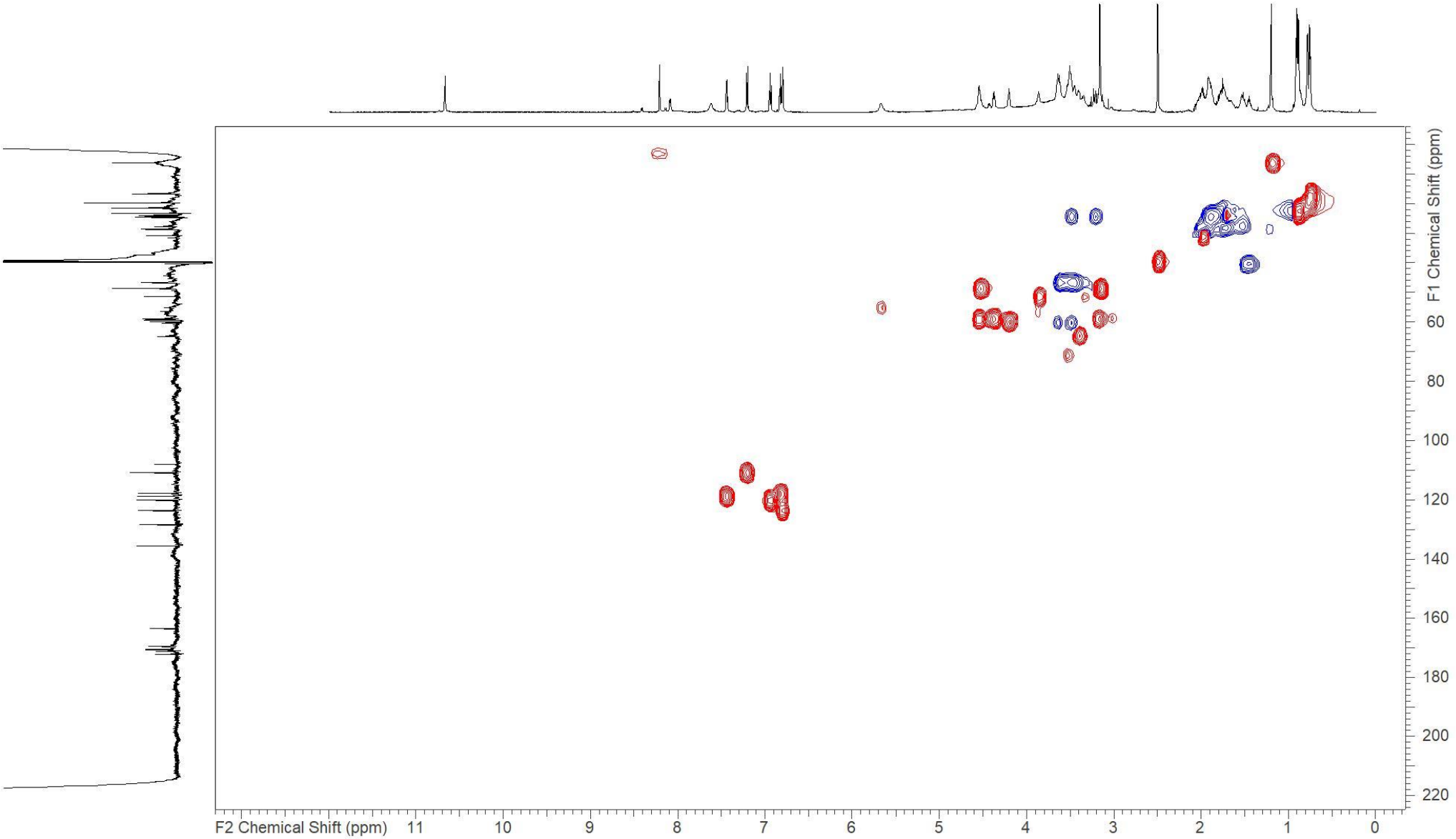
NMR spectra



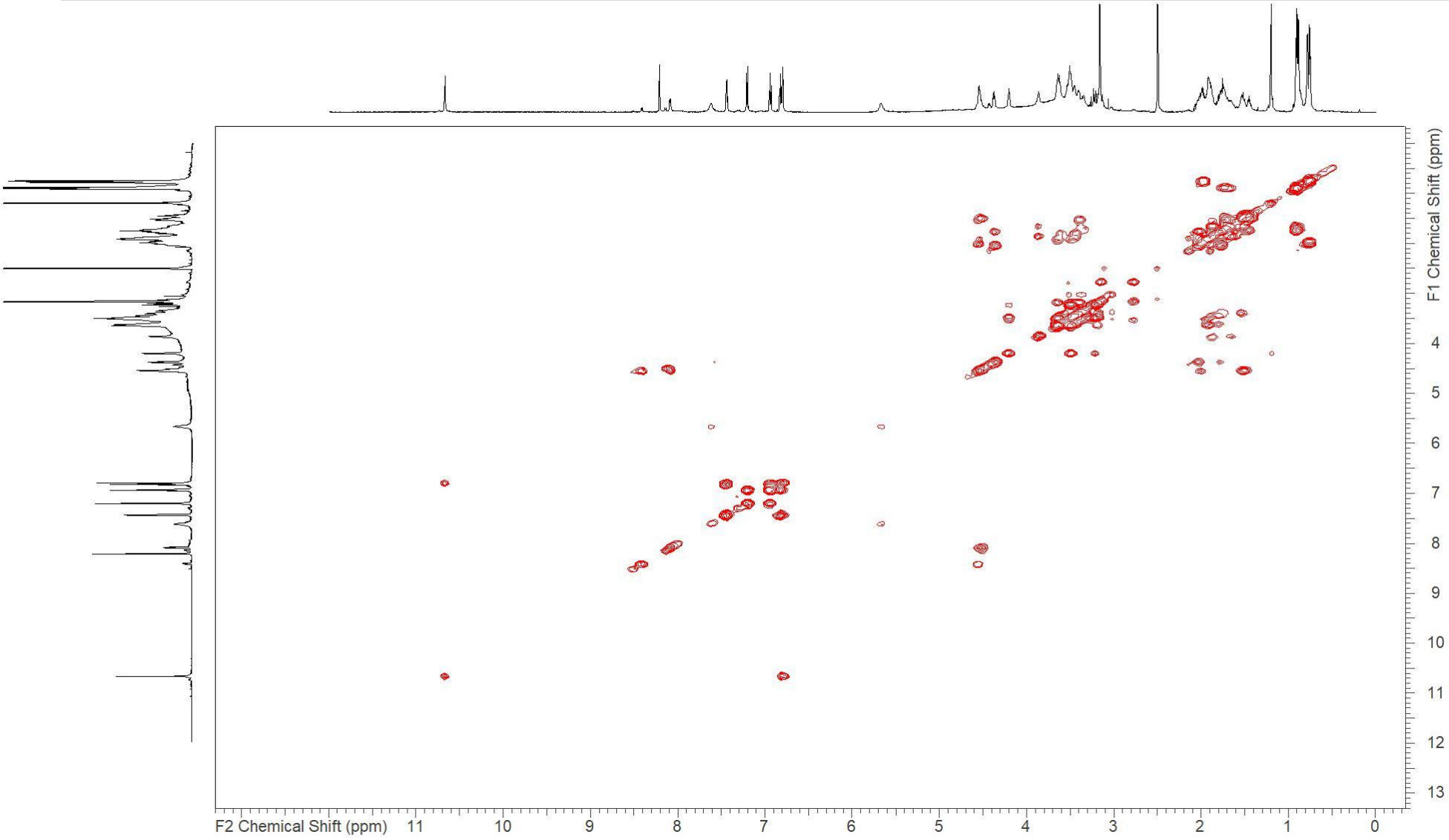
Compound 806 (1) – ^{1}H NMR recorded in DMSO-*d*6 at 700 MHz



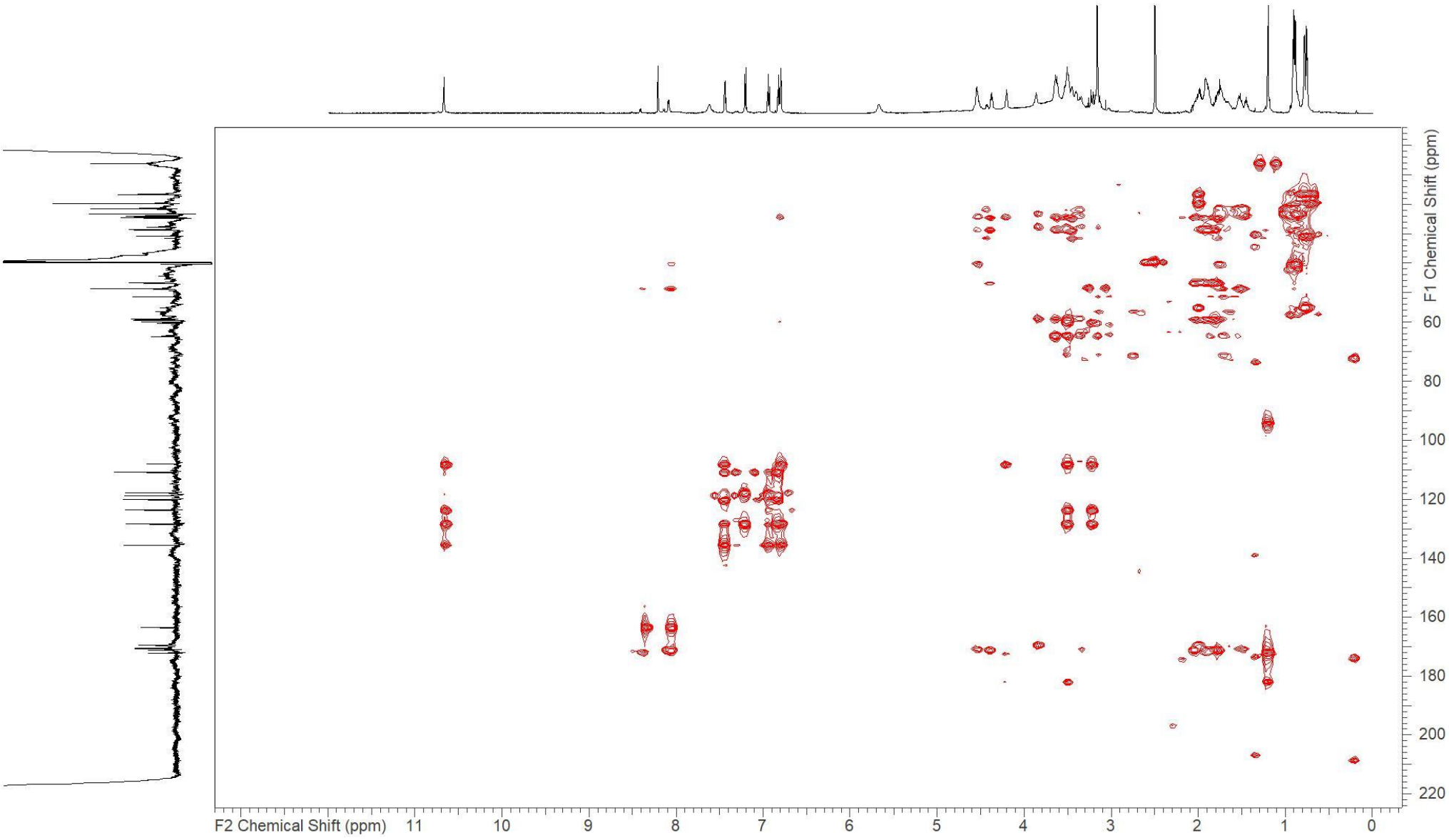
Compound 806 (1) – ^{13}C NMR recorded in DMSO-*d*6 at 175 MHz



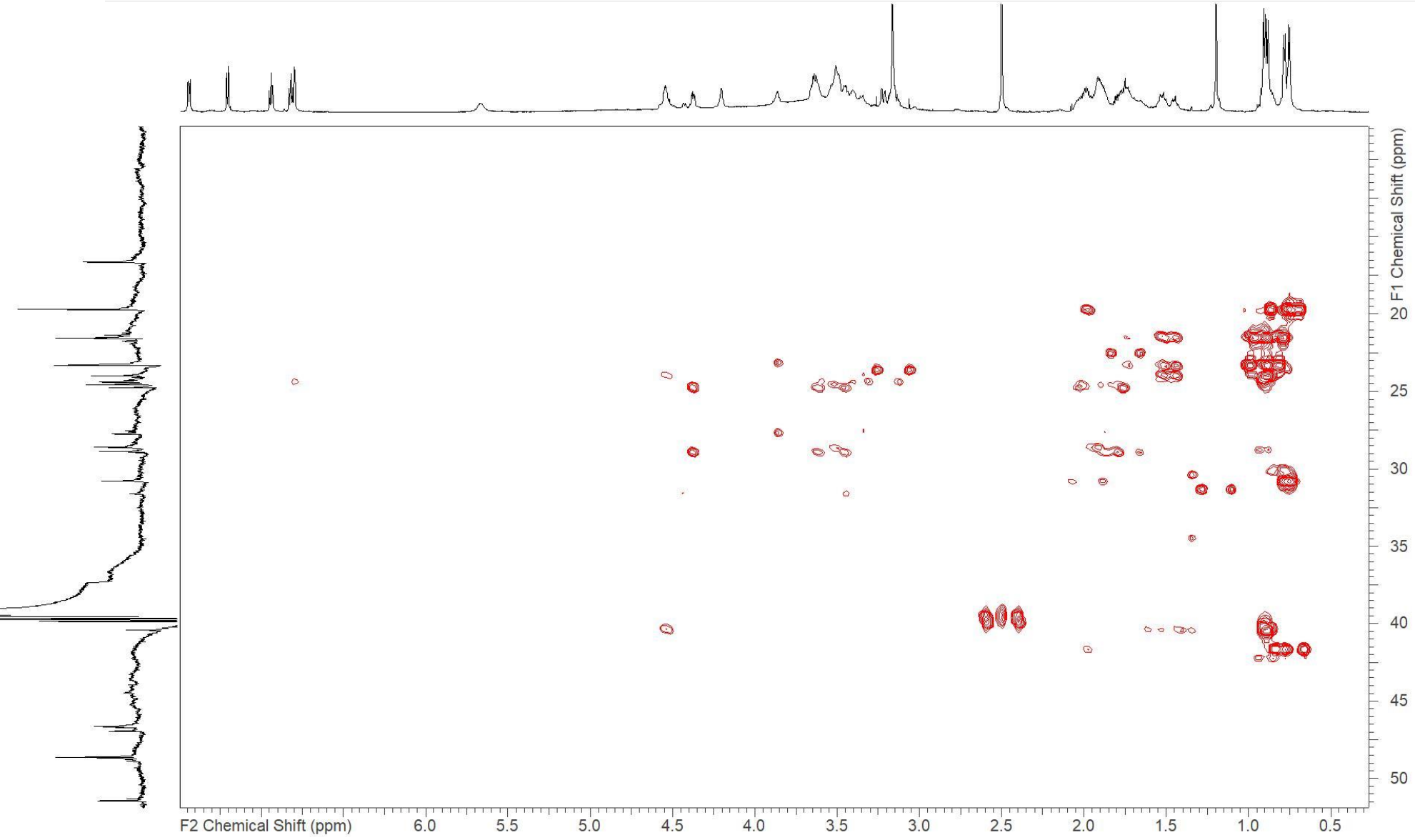
Compound 806 (1) – HSQC NMR recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



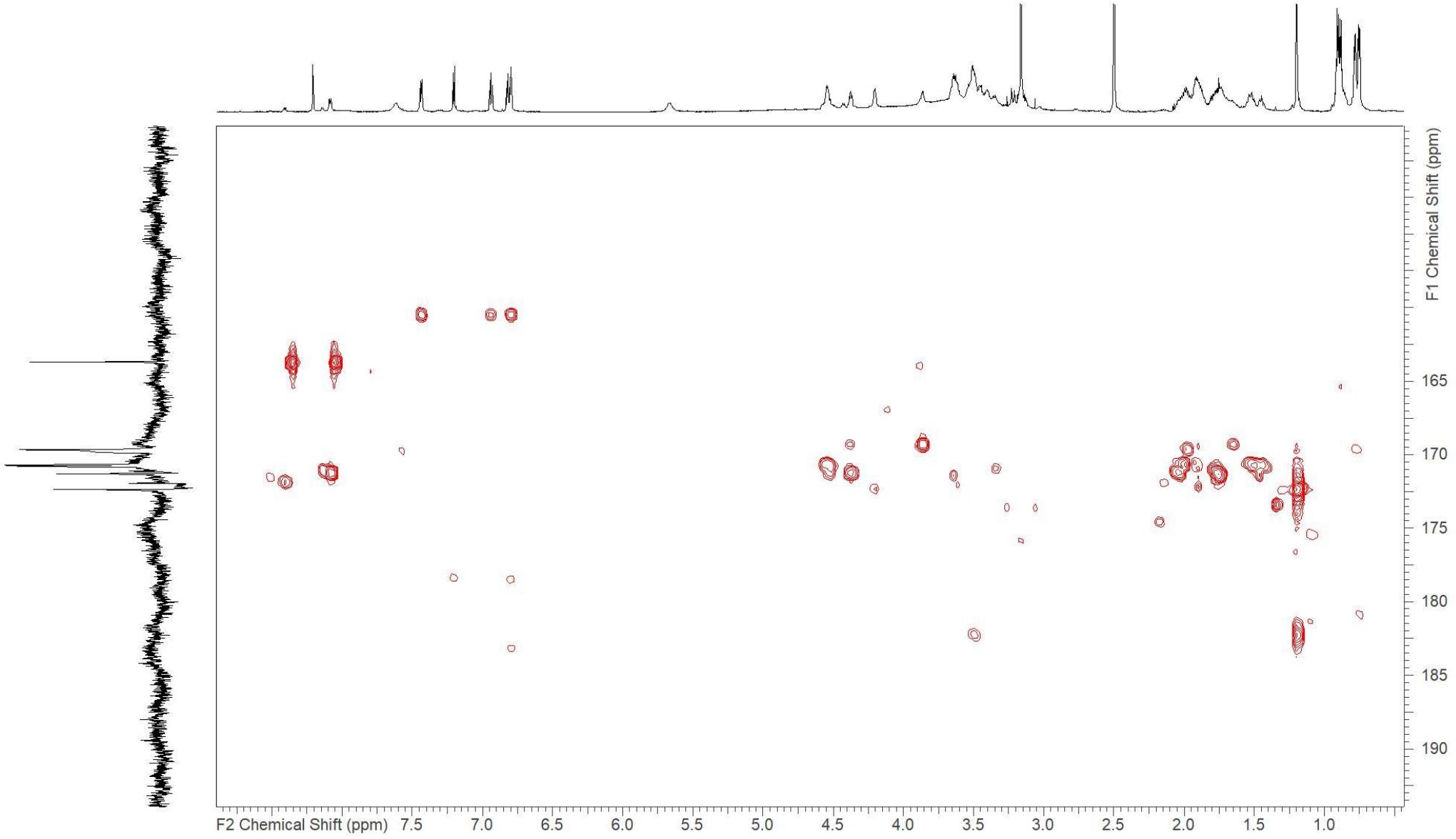
Compound 806 (1) – COSY NMR recorded in DMSO-*d*6 at 700 MHz



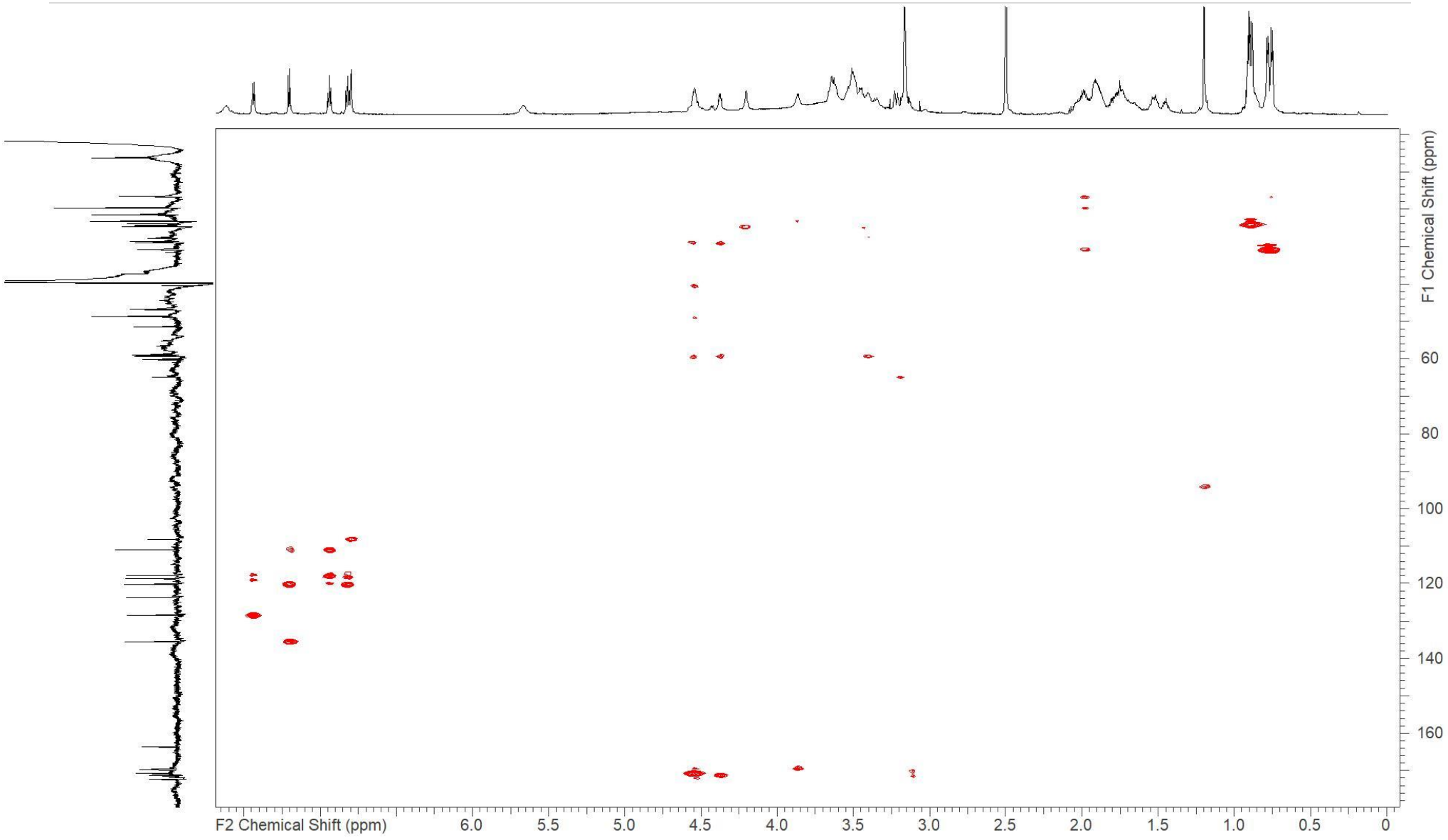
Compound 806 (1) – HMBC NMR recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



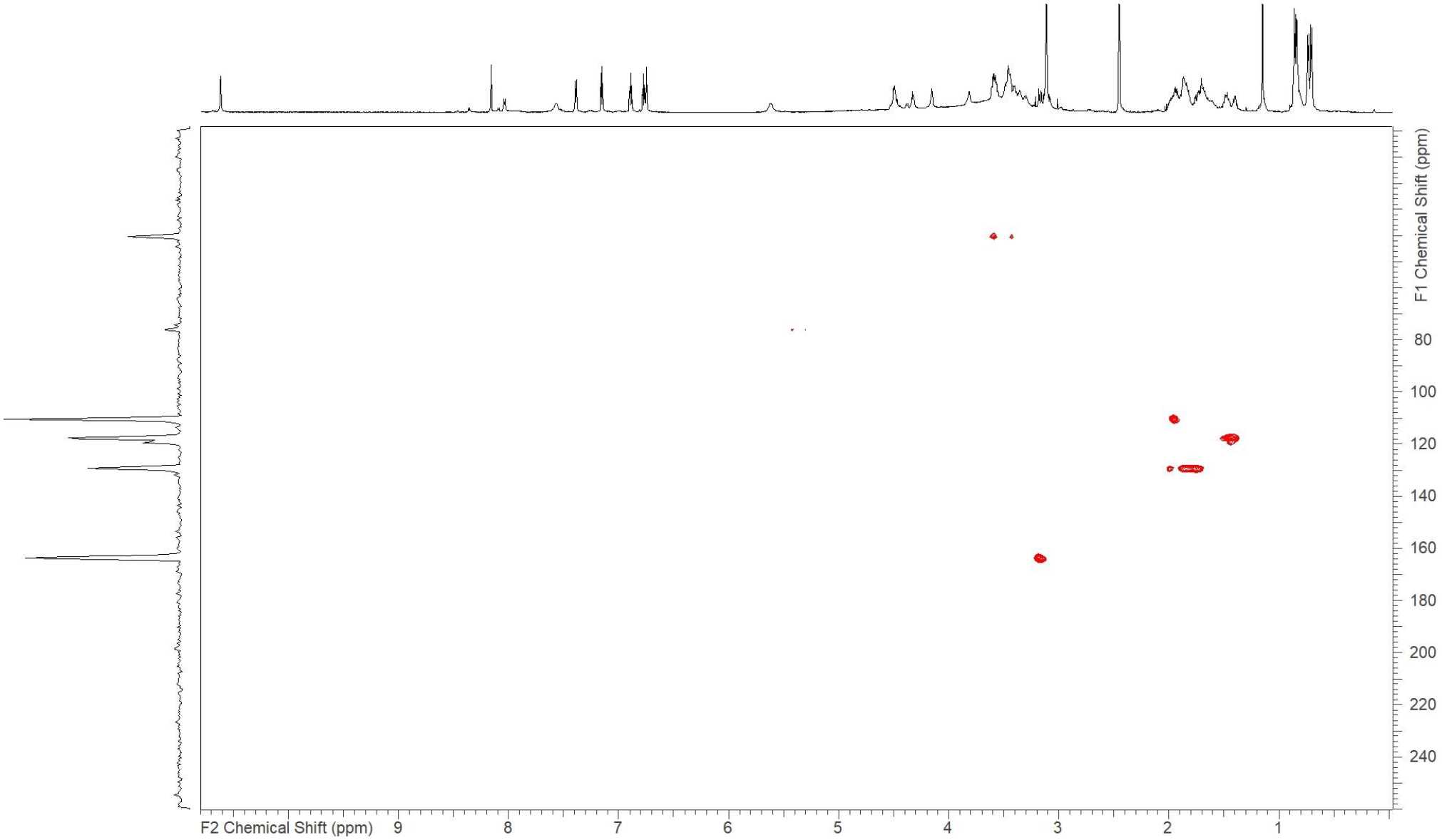
Compound 806 (1) – selective HMBC NMR (F1 0-42 ppm) recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



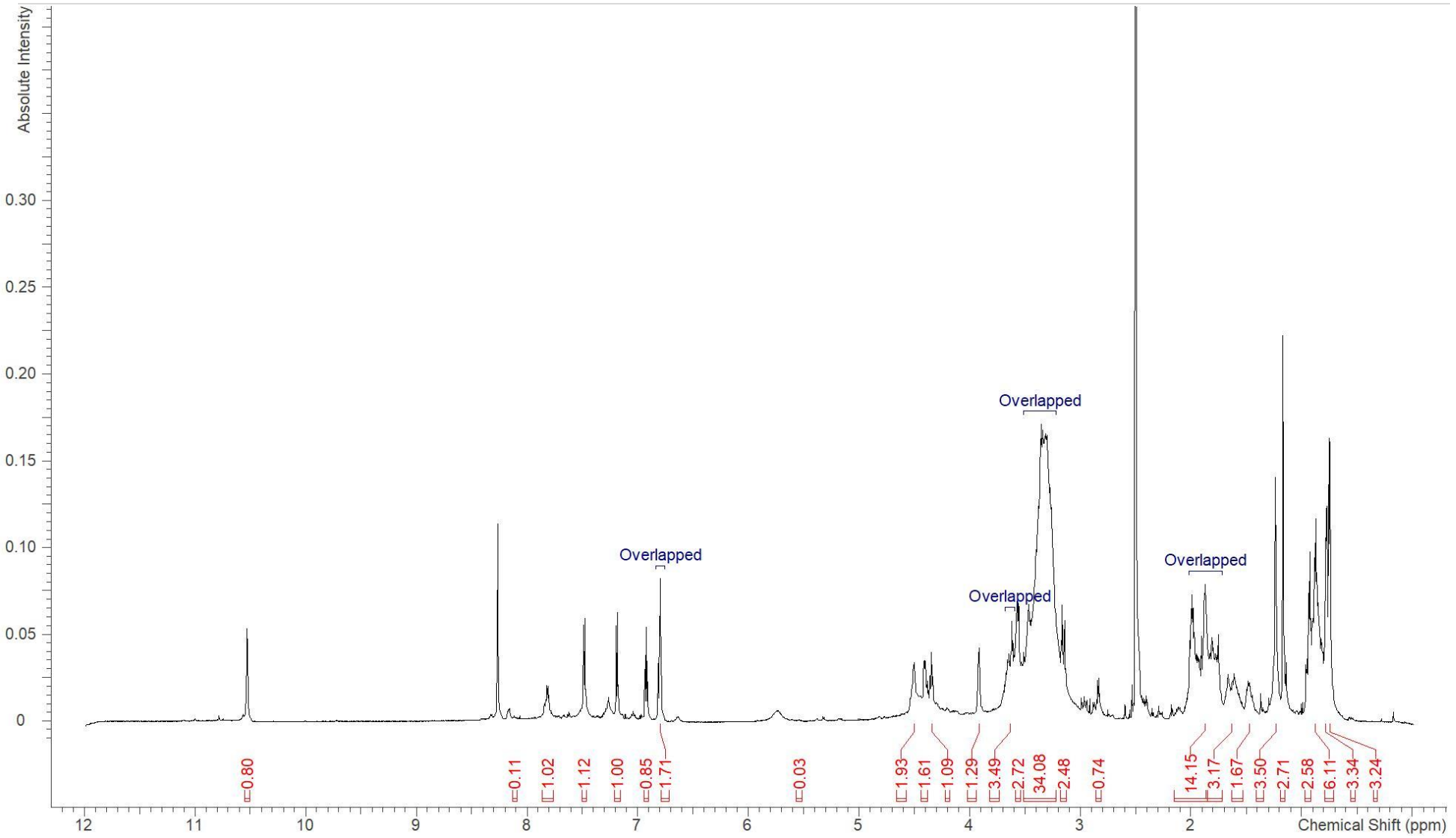
Compound 806 (1) – selective HMBC NMR (F1 160-190 ppm) recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



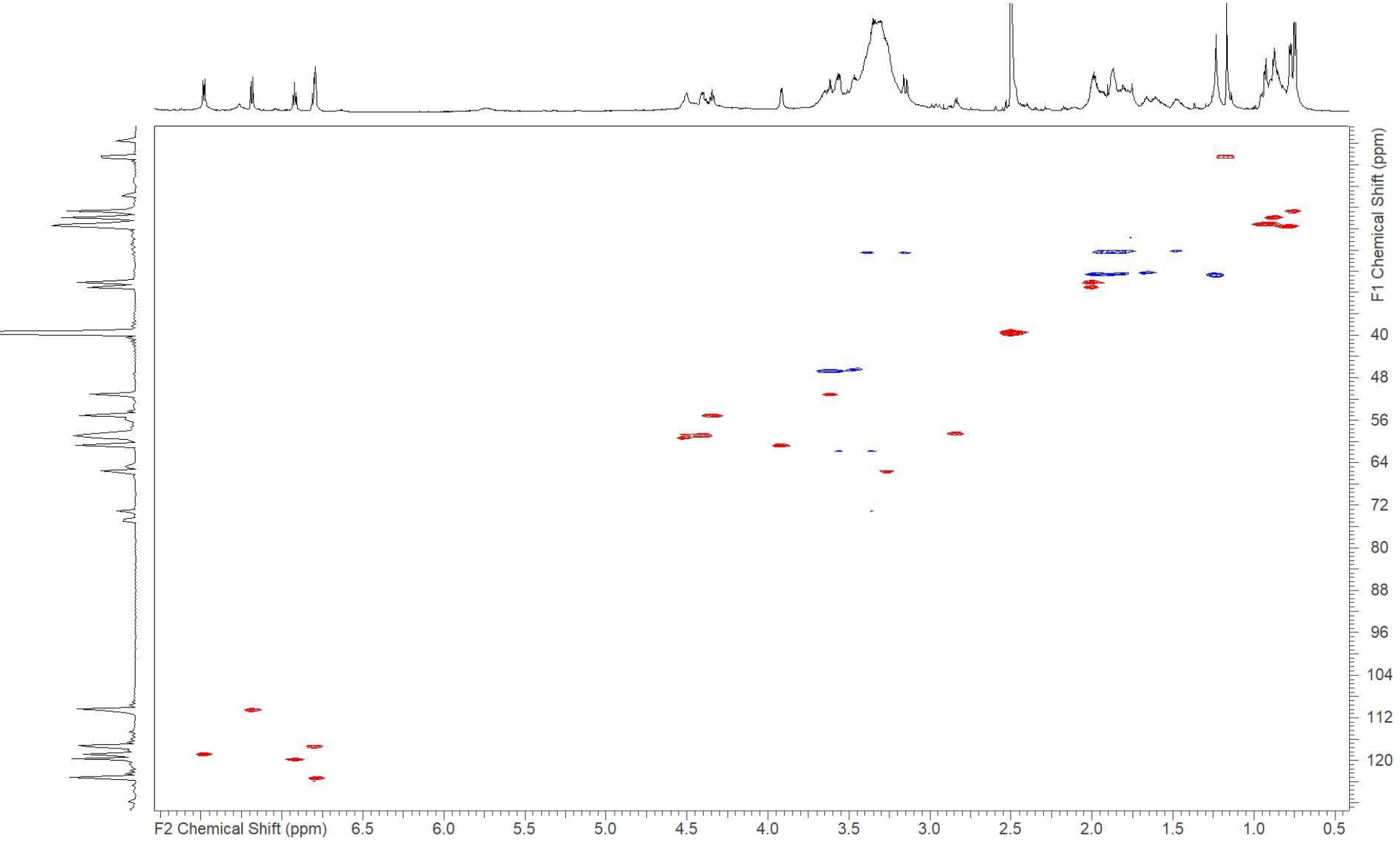
Compound 806 (1) – 1,1 ADEQUATE NMR recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



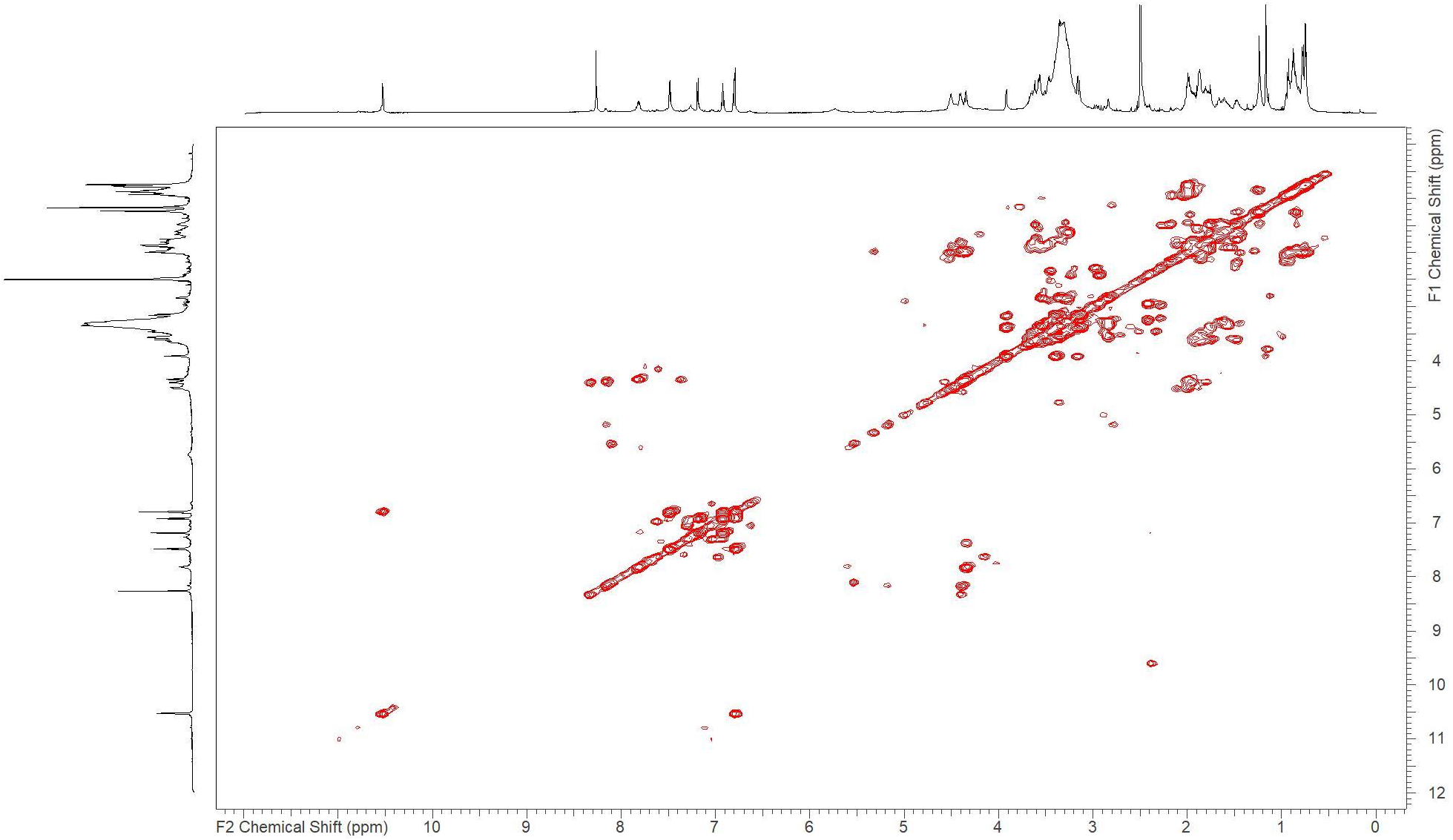
Compound 806 (1) – N-HMBC NMR recorded in DMSO-*d*6 at 71/700 (F1/F2) MHz



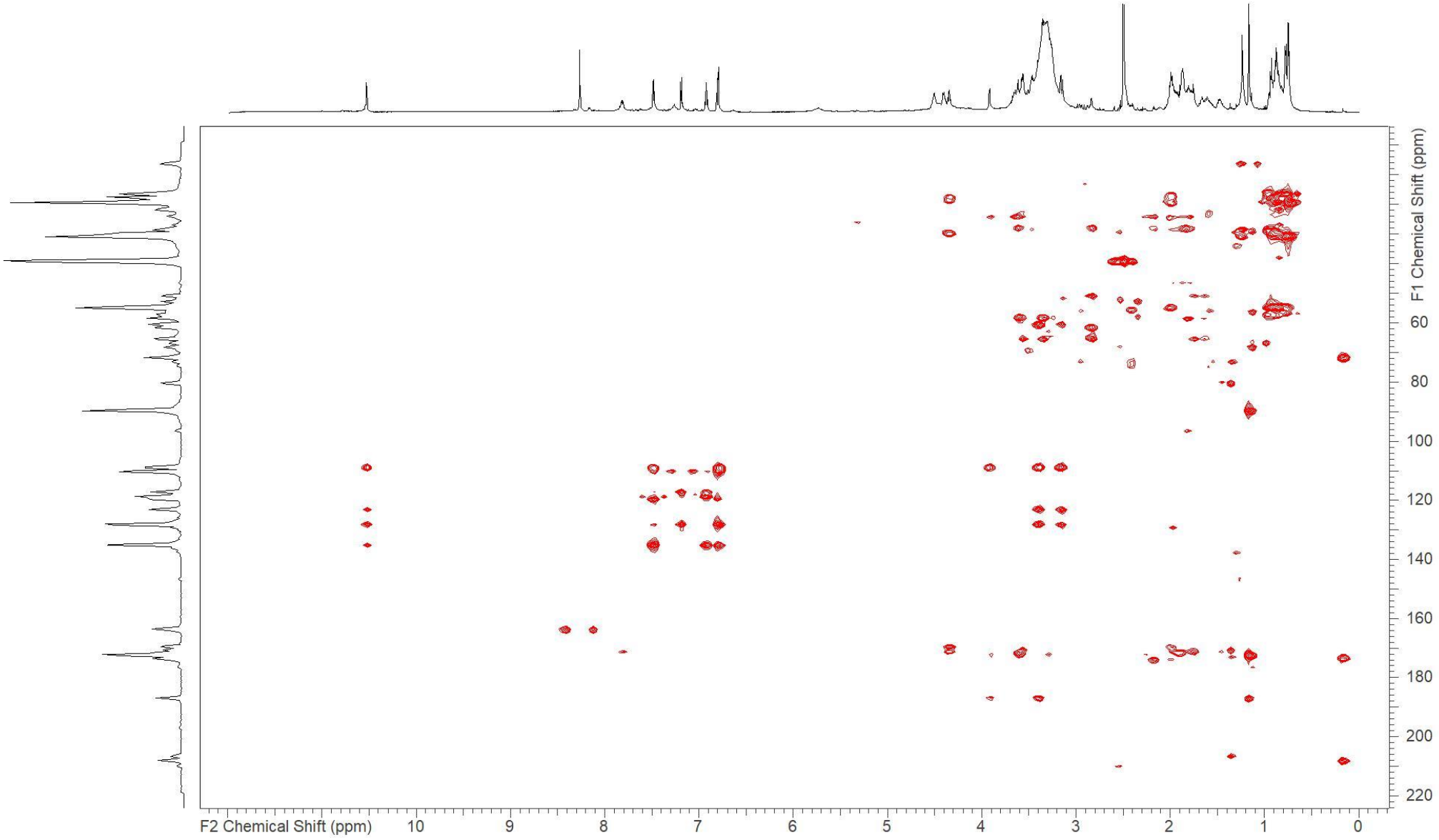
Compound 792 (2) – ^{1}H NMR recorded in DMSO-*d*6 at 700 MHz



Compound 792 (2) – HSQC NMR recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz



Compound 792 (2) – COSY NMR recorded in DMSO-*d*6 at 700 MHz



Compound 792 (2) – HMBC NMR recorded in DMSO-*d*6 at 175/700 (F1/F2) MHz

Supplementary references

(1) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 1999, *6*, 493–505.

(2) Röttig, M.; Medema, M. H.; Blin, K.; Weber, T.; Rausch, C.; Kohlbacher, O. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* 2011, *39*, W362-W367. DOI: 10.1093/nar/gkr323.

(3) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 2019 (47), W81-W87. DOI: 10.1093/nar/gkz310.

4. Discussion

4.1 Unknown secondary metabolites produced by biosynthetic gene clusters (BGCs) in SBCb004: overlooked compounds or cryptic, non-functional gene clusters?

Up to date, PKS, NRPS and PKS-NRPS-derived compounds still seem to be the major secondary metabolites discovered in myxobacteria[1]. For instance, the newly discovered ajudazol derivatives and cystopipecotides described in this work as well as the known metabolites tubulysin, stigmatellin, argyrin and alkylresocinol produced by the same strain SBCb004, are all generated by these three types of biosynthetic machineries (Table 4.1). Moreover, 21 out of the 41 annotated BGCs without any known product in strain SBCb004 can be assigned to these three classes using antiSMASH 5[2]. Other BGCs with unknown product in SBCb004 can mainly be assigned to Ribosomally synthesized and post-translationally modified peptides (RiPPs) family, terpene, aminoglycoside/aminocyclitol and phenazine (Figure 4.1.1).

Table 4.1 Biosynthetic gene clusters (BGCs) with known compound assigned in SBCb004

Known compound	BGC type
Ajudazol C-J	PKS-NRPS
Alkylresocinol	PKS
Argyrin	NRPS
Carotenoid	Terpene
Cystopipecotide	PKS-NRPS
Geosmin	Terpene
Stigmatellin	PKS
Tubulysin	PKS-NRPS

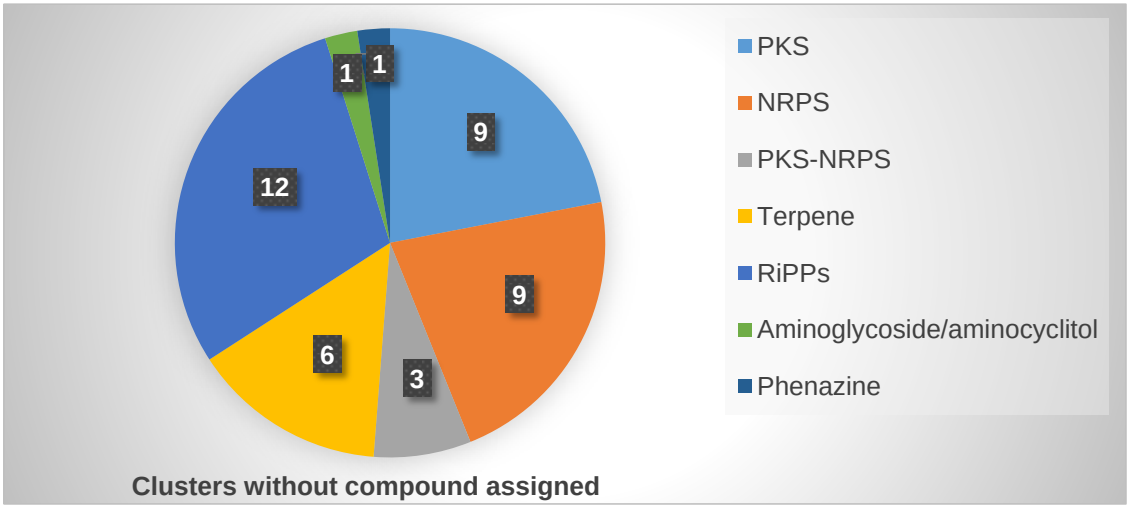


Figure 4.1.1 Statistic overview of biosynthetic gene clusters without assigned secondary metabolite in SBCb004. PKS: polyketide synthase. NRPS: non-ribosomal peptide synthase. PKS-NRPS: hybrid polyketide synthase-non-ribosomal peptide synthetase. RiPPs: Ribosomally synthesized and post-translationally modified peptides.

4.1.1 Terpene and RiPPs BGCs in SBCb004

Terpenes are a large class of natural products. The most common representatives in myxobacteria are geosmin[3] and carotenoids as their BGCs are widely distributed in the known myxobacterial genomes. As terpenes are usually volatile compounds, this might be an explanation why no terpene was discovered in our Liquid Chromatography-Mass Spectrometry-based screening. Using an alternative approach involving Gas Chromatography/Mass Spectrometry (GC/MS) would be more appropriate to screen for such kind of compounds. This also implies, that a combined use of various different sampling, measurement and isolation techniques may rather be suitable to obtain a comprehensive overview of produced compounds and therefore help discovering microbial natural products.

RiPPs seem to be the most neglected group of secondary metabolites in myxobacteria. Compared to the tremendous number of discoveries in other species in the recent years, only cittilin[4] and crocagin[5] have been described in myxobacteria. The 12 predicted RiPPs BGCs in SBCb004 comprise three of the lanthipeptide-type, three clusters for a linear azol(in)e-containing peptide (LAP), one microviridin-like cluster and six bacteriocin clusters. The predicted LAP, bacteriocin and microviridin-like BGCs in SBCb004 only contain one or two enzymes which are characteristic for these types of clusters like the YcaO-family proteins. Precursor peptides and other post-translational modification (PTM) enzymes seem to be

absent, which is in contrast to the common architecture of a RiPPs BGC. It is possible that these predicted BGCs only encode parts of the known and usual functions and the rest of the biosynthetic machinery is encoded elsewhere in the genome. Another explanation could be that these clusters are just partial segments of the entire cluster and the BGCs predicted by the antiSMASH algorithm are simply incomplete due to a lack of a comprehensively sequenced genome. In case of the three predicted lanthipeptide BGCs, both the precursor peptide and a homologue of the key PTM enzyme LanM being responsible for dehydration and cyclization are present. However, no molecular feature being absent in the targeted gene inactivation mutants was observed in the chromatograms of the extracts when compared to the wild type strain (data not shown). Whether these RiPPs BGCs are "silent" clusters by means of a missing expression under the applied culture conditions or whether the amount of product is simply too low to be detected will have to be addressed in further investigations. Unlike other RiPPs classes including lanthipeptides, thiopeptides and LAP which usually contain dozens of amino acids with multiple PTMs, the myxobacterial RiPPs cittilins and crocagins are rather small in size[6] (Figure 4.1.2). This may indicate that the RiPPs putatively produced by myxobacteria might belong to not yet characterized families of RiPPs so that the antiSMASH algorithm is not able to properly identify those kind of clusters. In order to identify such RiPPs in myxobacteria, a combination of a metabolomics strategy and an activity-based screening rather than a genomic-driven approach would be favorable.

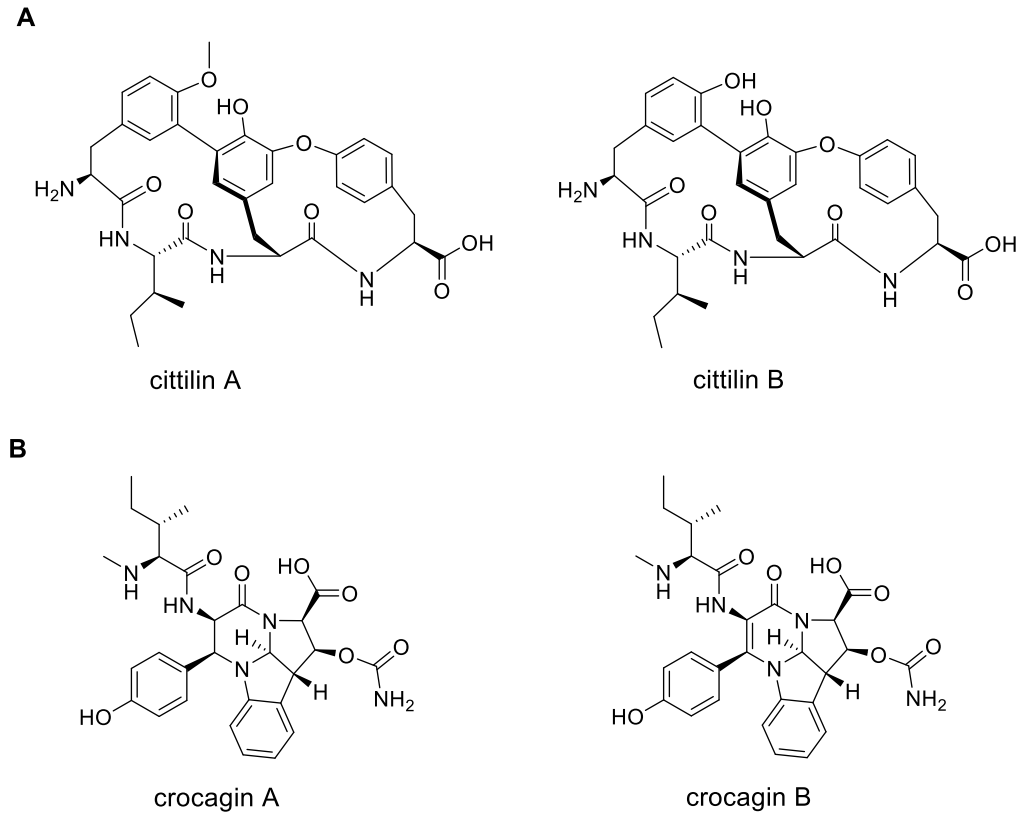


Figure 4.1.2 Chemical structures of myxobacterial RiPPs: cittilins (A) and crocagins (B).

4.1.2 Additional PKS and NRPS BGCs in SBCb004

Prompted by the expression of all the PKS or NRPS BGCs shown in Table 4.1 under current laboratory conditions, another 21 predicted PKS or NRPS BGCs in SBCb004 were inactivated through insertional mutagenesis as well. However, no secondary metabolite could be correlated to those BGCs like in the case of the novel ajudazol derivatives and the cystopipecotides (data not shown). Other strategies including heterologous expression, alternative culture conditions and genetic manipulations in SBCb004 to activate the expression of these clusters seem to be necessary. Further understanding of the biochemistry and biosynthesis of natural products may also be helpful to prioritizing these numerous BGCs.

Not least because enormous unsuccessful efforts to identify the products of the remaining BGCs not only in strain SBCb004, the debate about whether those BGCs without assigned metabolites are “silent” by means of being not or just poorly expressed under these conditions or they only depict truncated remains of formerly functional clusters will go on. It is hard to come to the conclusion especially when some necessary genes, for example,

those encoding loading modules in PKS and NRPS clusters are missing. According to some "textbook principle", these BGCs would be supposed to be incomplete and therefore non-functional. Nevertheless, a genome and all the encoded functions should not only be regarded in a two-dimensional way. The related genes may have been spread to different locations in the genome during evolution and such "standalone enzymes" encoded elsewhere or the interplay between two BGCs may be needed to produce the respective secondary metabolite. In addition, metabolic intermediates originating from primary metabolism or from other biosynthetic pathways could also be utilized, resulting in the incorporation of non-canonical building units. It would be also contradictory for the hosts especially for those harbor the enormous genetic information within their genomes at tremendous energy cost if those BGCs are really silent.

4.2 Non-canonical building blocks in PKS and NRPS

The incorporation of non-canonical building blocks in natural products mediated by PKS and NRPS contribute to the diversity of known secondary metabolites. The use of non-canonical building blocks by the biosynthetic machineries of myxobacteria has already been described in quite a number of cases. For example, isovaleryl-CoA (IV-CoA), acts as the precursor of numerous myxobacterial secondary metabolites including myxothiazol[7,8] and aurafuron[9]. Chloromyxamid contains 6-chloromethyl-5-methyloxypipecolic acid (CMPA) as a building block[10], while cystobactamid consist of multiple aromatic moieties[11]. Both kind of compounds characterized in this work, the novel derivatives of ajudazol and the cystopipecotides, contain non-canonical units (Figure 4.2.1). These uncommon building blocks either derive from metabolic intermediates of the primary metabolism or are synthesized by accessory enzymes encoded by genes flanking the core PKS and NRPS genes within the BGC. Although the real biological or physiological functions of many secondary metabolites remained elusive to date, what might also be a consequence of a lacking knowledge about structure-function relationship, it is obvious that these uncommon units introduce a high degree of structural diversity.

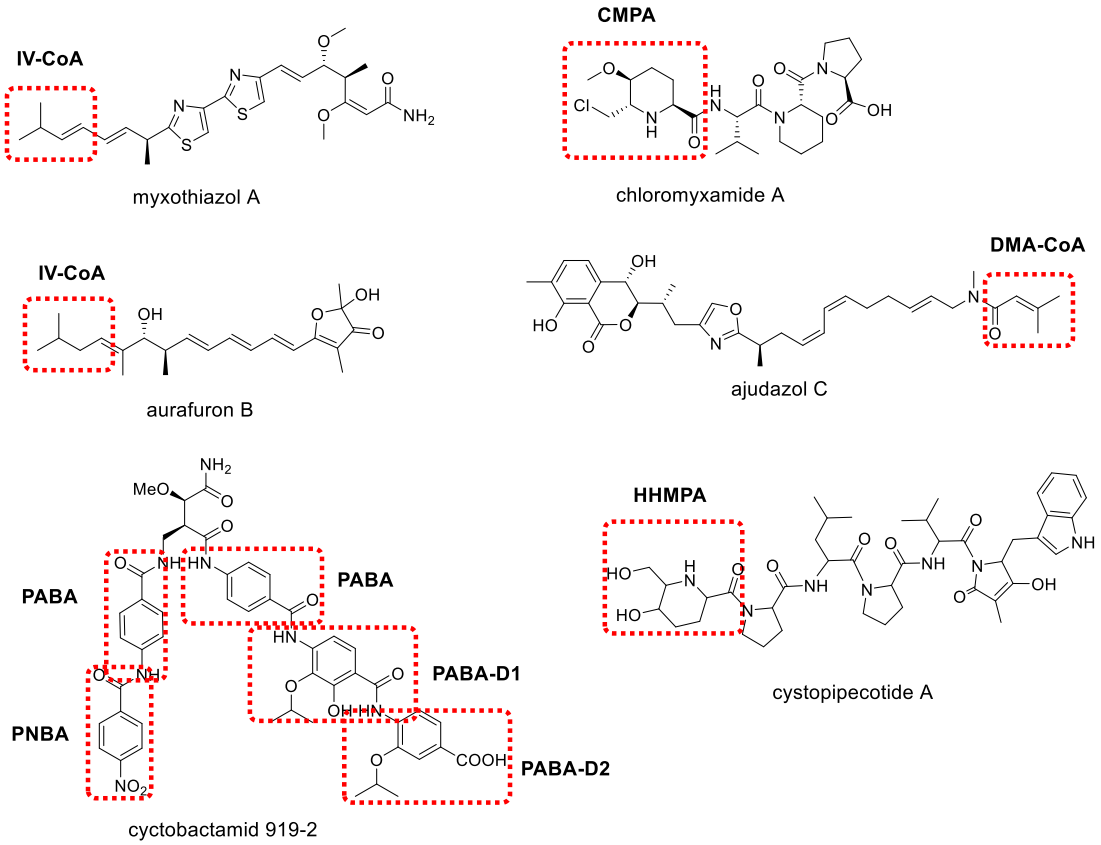


Figure 4.2.1 Examples of myxobacterial metabolites with non-canonical building blocks. The uncommon building blocks in each compound are marked by red boxes. CMPA: 6-chloromethyl-5-methoxypipecolic acid. DMA-CoA: 3,3-dimethylacrylyl-CoA. HHMPA: 5-hydroxyl-6-hydroxymethylpipecolic acid. IV-CoA: isovaleryl-CoA. PABA: *para*-aminobenzoic acid. PABA-D1 in cystobactamid 919-2: 4-amino-2-hydroxy-3-isopropoxybenzoic acid. PABA-D2 in cystobactamid 919-2: 4-amino-3-isopropoxybenzoic acid. PNBA: *para*-nitrobenzoic acid.

4.2.1 3,3-dimethylacrylyl-CoA as a building block in ajudazols C-J

When the biosynthesis of ajudazols A and B in *Chondromyces crocatus* Cm c5 is initiated, one malonyl-CoA and one acetyl-CoA are condensed followed by O-methylation to synthesize the 3-methoxy-butenoic acid at the N-terminus[12]. These two steps are both accomplished by AjuK. In contrast, the biosynthetic machinery of ajudazols C-J in SBCb004 does not contain the homolog of AjuK but utilizes 3,3-dimethylacrylyl-CoA (DMA-CoA), resulting in the formation of the 3-methyl-butenoic acid at the N-terminus. The incorporation of DMA-CoA as a starter unit could be regarded as a consequence of the lack of *ajuK* in SBCb004. Hence, the ajudazol BGC in SBCb004 is proposed to be related to the BGC present in *Chondromyces crocatus* Cm c5 during evolution. It is possible that the ajudazol BGC in strain Cm c5 is the primary version while *ajuK* in SBCb004 was lost during

evolution. However, the acylated 3-methyl-butenoic acid which is structurally similar to the product of AjuK, 3-methoxy-butenoic acid, is recognized by the following AjuL and thereby fed to the assembly line, yielding the new ajudazol derivatives. It is also plausible the other way around, namely that *ajuK* in strain Cm c5 was involved in the machinery later. It is tempting to speculate whether AjuL in SBCb004 are able to accept structurally similar analogs of DMA-CoA regardless of their abundance *in vivo* and whether AjuL in strain Cm c5 could accommodate DMA-CoA especially when *ajuK* would be inactivated and enough DMA-CoA would be supplied.

DMA-CoA is the metabolic intermediate of leucine catabolism as well as IV-CoA which is involved in the biosynthesis of myxothiazol and aurafuron. In myxobacteria, an alternative pathway leading to the *de novo* synthesis of DMA-CoA and IV-CoA has been described. It involves the dehydration of 3-hydroxy-3-methylglutaryl CoA, followed by decarboxylation and reduction, successively forming DMA-CoA and IV-CoA[13] (Figure 4.2.2). In other bacteria, degradation or formation of multiple compounds including β-valine, acyclic monoterpene citronellol and mycrene, *tert*-amyl-ethyl ether, methyl *tert*-butyl ether, *tert*-amyl-methyl ether and 3-methyl-3-pentanol is achieved including an DMA-CoA intermediate[14] (Figure 4.2.2), indicating that DMA-CoA is not only an important substrate of PKS or NRPS machineries but also a critical intermediate to link the metabolism of different metabolites.

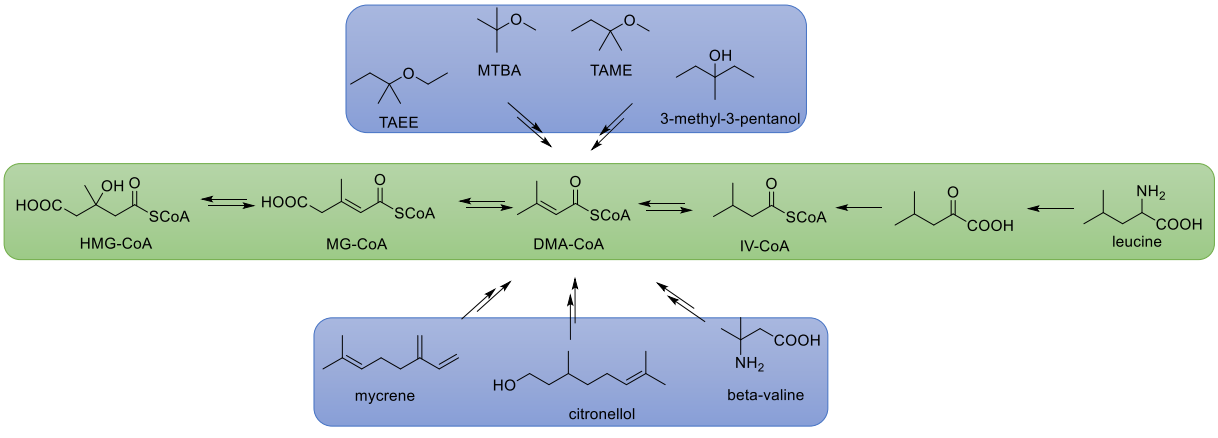


Figure 4.2.2 Metabolic pathway of 3,3-dimethylacrylyl-CoA (DMA-CoA) and isovaleryl-CoA (IV-CoA). The pathway in myxobacteria is highlighted in light green. Compounds being metabolized via DMA-CoA in other bacteria are highlighted in light blue. HMG-CoA: 3-hydroxy-3-methylglutaryl CoA. MG-CoA: 3-methylglutaconyl CoA. TAEE: *tert*-amyl-ethyl ether. MTBA: methyl *tert*-butyl ether. TAME: *tert*-amyl-methyl ether.

4.2.2 Substituted pipecolic acid as a building block in cystopipecotides

Pipecolic acid and its derivatives are prevalent building blocks in natural products and can be synthesized via multiple routes. One important microbial natural product containing pipecolic acid is rapamycin, in the biosynthesis of which L-lysine is deaminated, followed by reduction to form pipecolate (Figure 4.2.3 A). Both the L-lysine deamination and the imine reduction are catalyzed by a single cyclodeaminase LCD[15]. In plants, the α-amino group of L-lysine is first transferred to an oxoacid. Thereby, ε-amino-α-ketocaproic acid (KAC) is produced. KAC is subsequently cyclized and forms the ketimine 1,2-dehydrapipecolic acid. The latter is then reduced to give L-pipecolic acid[16] (Figure 4.2.3 B). In fungi, a C_{ε}-semialdehyde can be generated from saccharopine and subsequently form a 1,6-dehydrapipecolic acid. The cyclization is non-enzymatic catalyzed and the product is then reduced to L-pipecolic acid[17] (Figure 4.2.3 C). As saccharopine could also derive from lysine, these three pathways are designated as lysine-originating routes.

Recently, a novel pathway to form γ-alkylated pipecolate was reported in fungi. This route starts from *O*-acetyl-L-homoserine, the *O*-acetyl group of which is eliminated by a pyridoxal phosphate (PLP)-dependent enzyme (CndF). In the meanwhile, a nucleophilic attack by acetoacetate occurs at the C_{γ} and the intermediate forms a cyclized imine automatically. The latter is then reduced by the imine reductase CndE to give a γ-substituted pipecolate[18] (Figure 4.2.3 D). Later on, the biosynthesis of Fusaric acid was revealed, within which a cyclic imine intermediate is generated from aldehyde and acetoacetate, catalyzed by a PLP-dependent enzyme Fub7. The imine product could be intercepted through chemical reduction using $NaBH_3CN$ to synthesize 5-substituted pipecolic acid[19] (Figure 4.2.3 E).

Substituted pipecolic acids were also found previously in the myxobacterial metabolites. One important case is tubulysin which harbors a D-*N*-methyl pipecolic acid building block (D-Mep). The biosynthesis of D-Mep starts from the formation of L-pipecolic acid generated from L-lysine, catalyzed by lysine cyclodeaminase TubZ. L-pipecolic acid is then loaded to NRPS module TubB within which *N*-methylation occurs. The L-Mep was speculated to be converted to its D-isomer downstream of the first chain extension[20] (Figure 4.2.3 F). Another case reported recently is chloromyxamid. In this case the biosynthesis of 6-chloromethyl-5-methoxypipecolic acid (CMPA) was proposed to start with the linkage of glutamate to a LysW-like protein CmxD. Subsequently, the carbon chain extends through a Claisen like condensation, followed by a series of reduction, dehydration, chlorination and a final

cyclization. Thereby, 6-chloromethyl-5-hydroxypipecolic acid is formed, which is further converted into CMPA by *O*-methylation[10] (Figure 4.2.3 G). Intriguingly, LysW-like protein usually links to the glutamate through the α-amino group of glutamate and the γ-carboxyl group of C-terminal glutamate in LysW. However, CmxD was proposed to connect to the substrate through thioester bond formed between the activated carboxyl group of the glutamate and the 4′-phosphopantetheine attached to CmxD like in PKS and NRPS.

In the present work, it was shown that cystopipecotides feature an uncommon 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) building block, which is structurally similar to CMPA. The biosynthesis of HHMPA is also proposed to derive from glutamate. However, it seems to be achieved by a unique pathway with α-aminoadipate (AAA) pathway enzymes involved (Figure 4.2.5 A) and the chain is extended by the transfer of a C-2 unit mediated by transketolases (Figure 4.2.5 C.

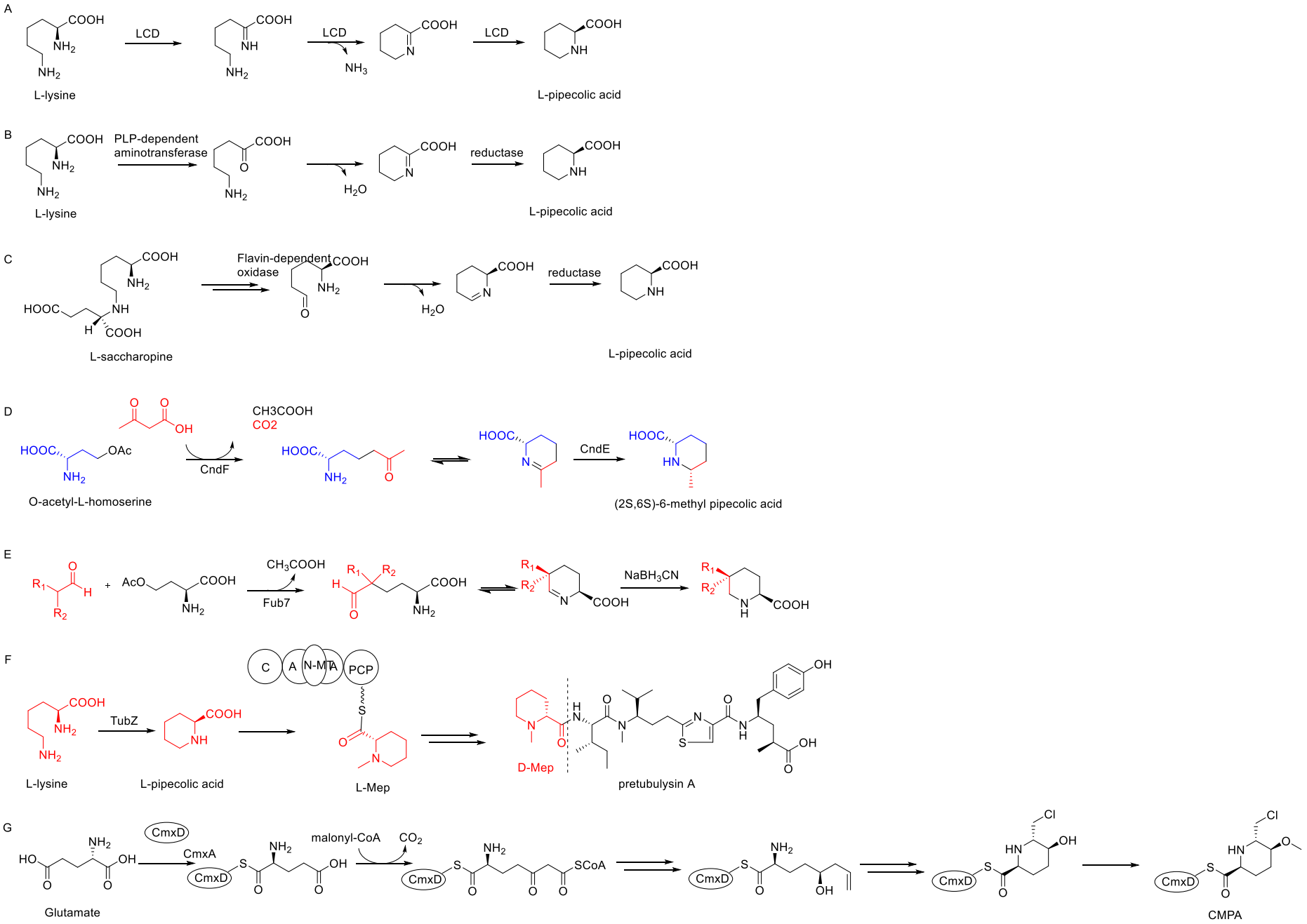

A
LCD
LCD
LCD
L-lysine
L-pipecolic acid
B
PLP-dependent aminotransferase
reductase
L-lysine
L-pipecolic acid
C
Flavin-dependent oxidase
reductase
L-saccharopine
L-pipecolic acid
D
CH3COOH
CO2
CndF
CndE
O-acetyl-L-homoserine
(2S,6S)-6-methyl pipecolic acid
E
Fub7
NaBH3CN
F
TubZ
C
A
N-MTA
PCP
L-lysine
L-pipecolic acid
L-Mep
D-Mep
pretubulysin A
G
CmxD
CmxA
malonyl-CoA
CO2
Glutamate
CMPA

Figure 4.2.3 Overview of biosynthetic and chemoenzymatic synthetic routes towards pipecolic acid and its derivatives. A Lysine-cyclodeaminase (LCD) catalyzed pathway in rapamycin biosynthesis. **B** Pyridoxal phosphate (PLP)-dependent biosynthetic pathway in plants. **C** L-saccharopine pathway in fungi. **D** Biosynthesis of γ-substituted pipecolate catalyzed by a PLP-dependent γ-substitution under the action of CndF and an imine reductase CndE. **E** Chemoenzymatic synthesis of 5-substituted L-pipecolic acid. **F** Biosynthesis of D-*N*-methyl pipecolic acid (D-Mep) in tubulysin. R_1, R_2: alkyl. **G** Putative biosynthetic pathway of 6-chloromethyl-5-methoxypipecolic acid (CMPA) in myxobacteria.

Transketolases are able to transfer a C-2 keto-containing unit from a ketose to the first carbon of an aldose[21]. Moreover, the C-2 unit can also be provided to NRPS or PKS assembly lines and thus participate in the biosynthesis of relevant pharmaceutical compounds. For example, in the biosynthesis of the antitumor compounds naphthyridinomycin and quinocarcin (Figure 4.2.4 A), a hydroxyethyl unit originating from a ketose is transferred to an ACP and acts as an extender unit for the NRPS assembly line[22]. In the biosynthesis of ripostatin, transketolases RipF and RipG mediate the decarboxylation of phenylpyruvate. The newly formed phenylacetyl group is then transferred to the ACP domain within RipH and acts as the starter unit for the ripostatin PKS biosynthesis pathway[23] (Figure 4.2.4 B). Furthermore, transketolases can participate in the biosynthesis of natural products directly. In the biosynthesis of late rifamycin derivatives, a two-component transketolase transfers a dihydroxyethyl group from a 2-ketose to the C_4 carbonyl carbon of rifamycin-S. Followed by bond rearrangements and re-aromatization, the ketone in rifamycin-S is transformed into the ester in rifamycin-L[24] (Figure 4.2.4 C).

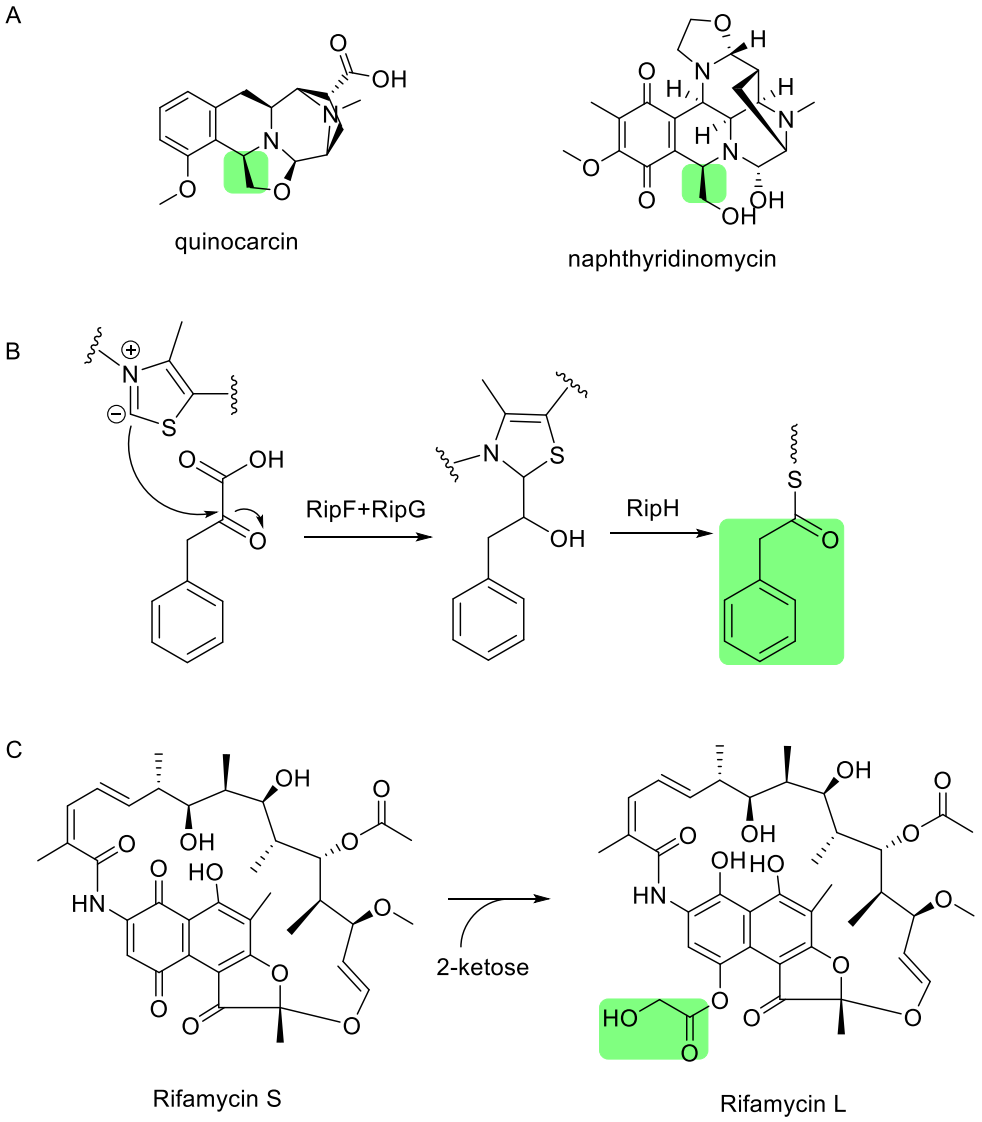


Figure 4.2.4 Examples of natural products with transketolases involved in their biosynthesis. The parts of the molecules deriving from the outcome of transketolases are highlighted with green boxes. **A** C-2 unit acting as an extender unit for the NRPS in case of naphthyridinomycin and quinocarcin biosynthesis. **B** The decarboxylation of phenylpyruvate mediated by transketolases RipF and RipG in ripostatin biosynthesis. **C** C-2 unit forming an ester in case of rapamycin L biosynthesis.

Homologs of lysine biosynthetic enzymes, together with transketolases, were also reported to generate another precursors in natural products[25,26]. In *Streptomyces* sp. SANK 60404, the whole set of the AAA pathway (Figure 4.2.5 A) enzymes has been shown to synthesize (2S,6R)-diamino-(5R,7)-dihydroxy-heptanoic acid (DADH) with the action of auxiliary transketolases. The latter enzymes interfere the AAA pathway and transfer a C-2 unit from a sugar to the semialdehyde. The aldehyde is reduced to a hydroxyl group and the following transamination occurs on the newly generated keto group. DADH is ultimately formed after

the tethered LysW is released[26] (Figure 4.2.5 B). In case of the biosynthesis of HHMPA in strain SBCb004, the resultant keto-group of the action of transketolases and the α-amine undergo cyclization after the release from LysW as the transaminase LysJ is absent. Similar to the last step of pipecolic acid biosynthesis outgoing from saccharopine, the formed imine is then reduced to yield HHMPA (Figure 4.2.5 C). To our knowledge, this is the first report of pipecolic acid biosynthesis in a concerted action of transketolases and enzymes of the AAA pathway. It is noteworthy that the biosynthetic machinery of DADH formation which comprises the whole set of the AAA pathway enzymes is widely distributed in actinomycetes[26] while it is the "truncated" biosynthetic machinery of HHMPA synthesis extensively distributed in myxobacteria, indicating two distinct evolutionary routes. However, it remains elusive whether the incorporation of the HHMPA moiety adds some exclusive biological function to the cystopipecotides. Nevertheless, an alternative route for the biosynthesis of substituted pipecolic acid in bacteria could be identified, which also enriches the toolbox for the synthesis of customized compounds via strategies of synthetic biology in the future.

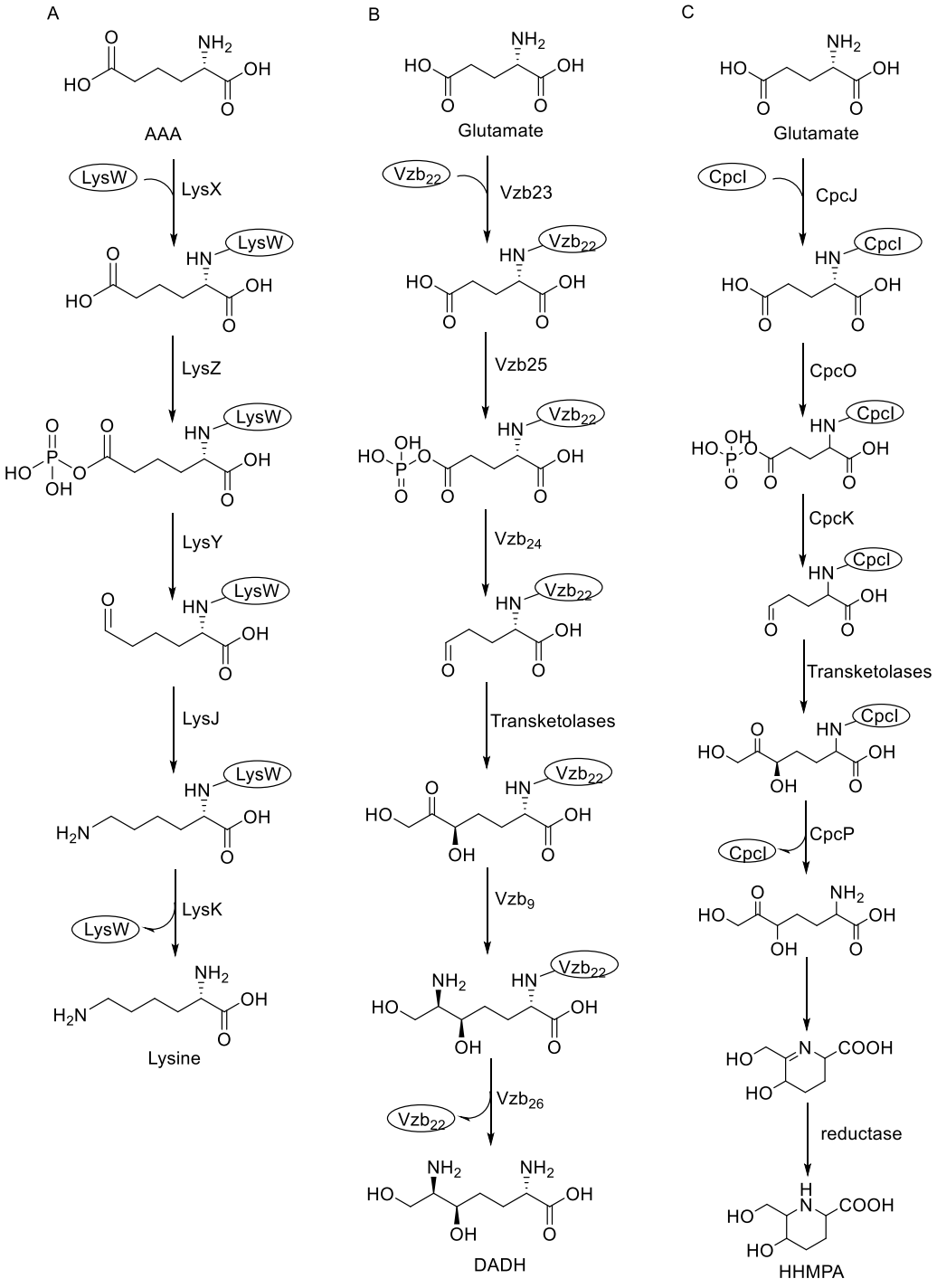


Figure 4.2.5 Comparison of biosynthetic pathways involving α-aminoadipate (AAA) pathway enzymes. A Lysine biosynthesis via AAA in *Thermus thermophiles* and *Sulfolobus acidocaldarius*. **B** (2S,6R)-diamino-(5R,7)-dihydroxy-heptanoic acid (DADH) biosynthesis in *Streptomyces* sp. SANK 60404. **C** 5-hydroxyl-6-hydroxymethylpipecolic acid (HHMPA) biosynthesis in *Cystobacter* sp. SBCb004.

4.3 Post-assembly-line modifications of myxobacterial secondary metabolites

Besides the incorporation of diverse building blocks into the secondary metabolites under the action of the PKS/NRPS assembly lines, tailoring modifications of the resulting core structures by the activity of enzymes such as oxidoreductases, group transferases, halogenases and cyclases also contribute to the structural diversity of natural products. Secondary metabolites of myxobacterial origin usually exhibit various post-assembly-line modifications performed by abovementioned enzymes. For instance, the selective methylation, oxidation and O-methylation at the indole ring in argyrin biosynthesis give rise to the production of several derivatives[27] (Figure 4.3 A). The glycoside antibiotics disciformycins undergo acylation and glycosylation after the core structure is offloaded from the PKS assembly line[28] (Figure 4.3 B). Another example is pretubulysin A, which is first oxidized in two-steps followed by acylations to form tubulysin A[20] (Figure 4.3 C).

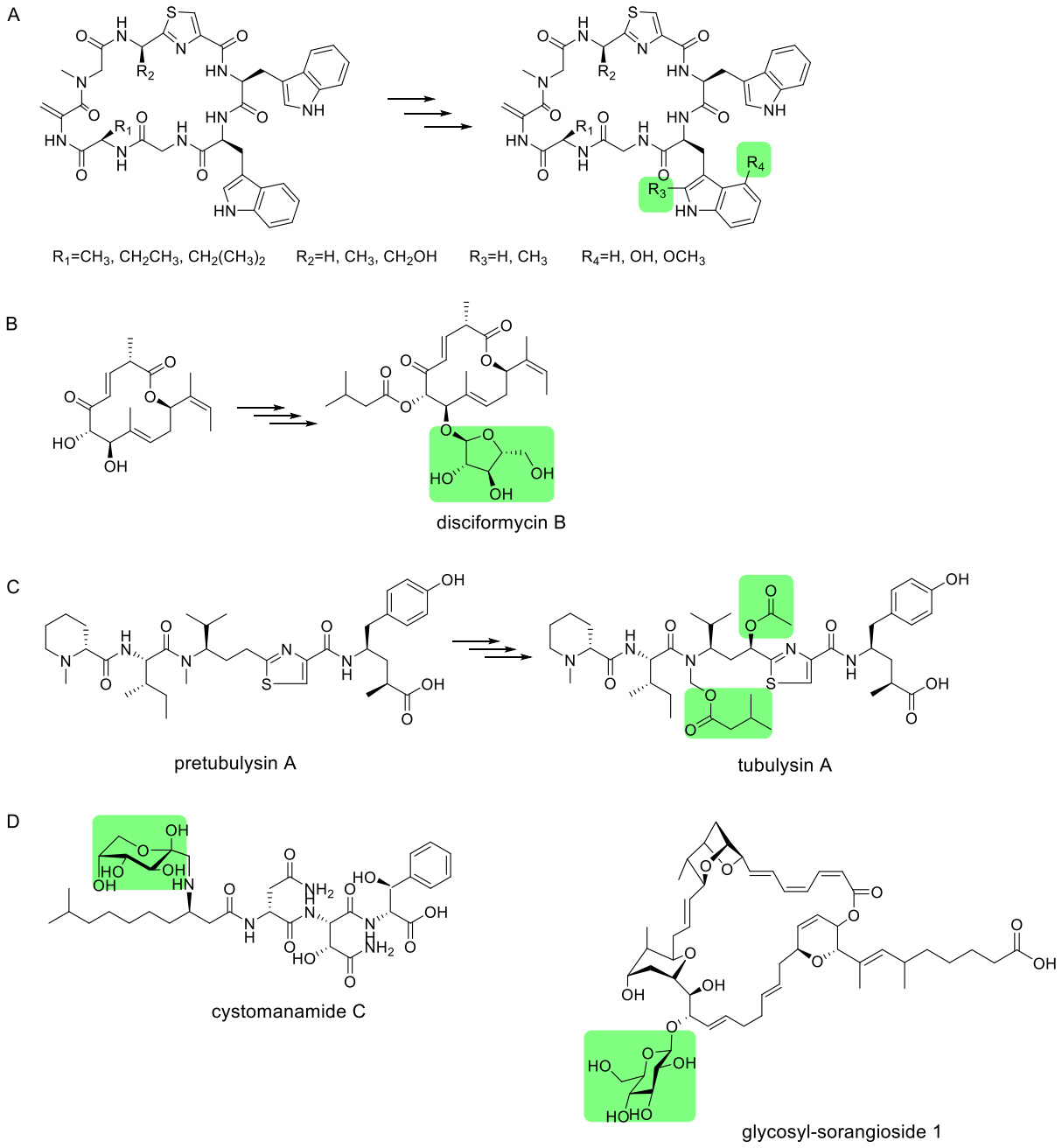


Figure 4.3 Examples of tailoring modifications of myxobacterial secondary metabolites. A Methylation, oxidation and O-methylation in the biosynthesis of argyrin. **B** Acylation and glycosylation in the biosynthesis of disciformycin B. **C** Oxidations and acylations in the biosynthesis of tubulysin A. **D** Cystomanamide C and glycosyl-sorangioside 1. Parts originating from post-assembly line modifications are labelled with green boxes.

Generally, the genes encoding such tailoring enzymes responsible for post-assembly-line modifications are clustered together with other accessory genes, such as regulator and transporter genes, and are flanking the core PKS/NRPS genes. However, exceptions from this rule seem to occur quite often in myxobacterial BGCs. For example, the acylase and

glycosylase participating in disciformycin biosynthesis are not encoded in the BGC and have not been identified yet[28]. The glycosyltransferase involved in the biosynthesis of cystomanamide (Figure 4.3 D) is also missing in the known BGC[29]. In addition, the tailoring enzymes being responsible for the abovementioned oxidations and acetylations of pretubulysins could also not be found in the respective BGC[20].

Another example could be shown by the present work. As described in chapter 1, deshydroxyajudazol C released from the PKS-NRPS assembly line, can be processed by multiple known and unknown enzymes to yield various derivatives exhibiting different patterns of hydroxylation, saturation and glycosylation. The hydroxylation at C_8 and the desaturation at C_{15}-C_{33} in ajudazol C are equal to those in case of ajudazol A and B. Two cytochrome P450 enzymes could be identified in the genome of SBCb004 by screening for genes homologous to *ajuI* and *ajuJ* which are encoded in the ajudazol BGC of strain Cm c5. A homologue of *ajuJ* is encoded in the cluster in SBCb004 and could be shown to be responsible for the hydroxylation at C_8 and the homolog of *ajuI* could be identified elsewhere in the genome of SBCb004. However, the enzymes catalyzing the hydroxylation at C_{15}, C_{29} and C_{33} are missing in the BGC in SBCb004. It remains elusive whether the modifications at C_{15}, C_{29} and C_{33} are a result of the activity of three different enzymes or just due to a wide substrate specificity of one enzyme. As enzymes being able to perform a transfer of a pentose are rarely identified and characterized, only the glycosyltransferase being responsible for the transfer of the hexose in ajudazol biosynthesis was identified by using the amino acid sequence of SorF, the glucose transferase involved in sorangicins as a query[30] (Figure 4.3 D). Even though SorF shows promiscuous donor substrate selectivity *in vitro*, its homologue AjuM in SBCb004 is not able to transfer the pentose. As oxidoreductase family enzymes and glycosyltransferase family enzymes consist of numerous subclasses and are frequently involved in numerous branches of the metabolism[31–33], without further efforts, the unknown enzymes being responsible for the hydroxylations at C_{15}, C_{29} and C_{33} and the transfer of pentose will remain elusive. The genetic locus and identification of the gene encoding AjuM suggests an alternative strategy to circumvent the absence of a glycosyltransferase or other tailoring enzymes while expressing a BGC in a heterologous host. The biosynthetic machinery could be complemented in the heterologous host with simultaneous production of a homologous enzyme with confirmed catalytic function. The distribution of genes encoding glycosyltransferases and oxidoreductases in the genome of SBCb004 involved in ajudazol C–J biosynthesis also supports our hypothesis that these genes work synergistically regardless of the linear distance to each other. It is still not clear if the genome-wide

distribution of these tailoring genes involved in the biosynthesis of ajudazol is caused by the insertion of other DNA fragments or rearrangement of the genome during evolution or it is the byproduct of some discrete enzymes with plastic substrate selectivity.

Modifications such as oxidation and glycosylation, are often critical for the biological activity and stability of the mature natural product[34]. For instance, sugar moieties are often required for target-compound interaction or have influence on pharmacokinetic characteristics[35]. This is demonstrated by the desosamine moiety of erythromycin A which is indispensable for the antibiotic activity due to its interaction with ribosomal RNA[36]. Nevertheless, glycosylation may also work as a self-resistance mechanism to convert an active compound into its inactive form. The glycosylated sorangiosides, which lose their biological activity, is a pertinent example in myxobacteria[30]. In the present work, the glycosylated ajudazols do not show better cytotoxicity than their non-glycosylated congeners, which leads to the hypothesis that the sugar moieties here work as detoxification or transportation. Since no glycosylated ajudazol derivatives were observed in the extracts of strain Cm c5 (the producer of ajudazol A and B), a distinct self-resistant mechanism in Cm c5 would be suggested. Glycosylation of ajudazols may also contribute to another unknown biological function of the ajudazols which differs from the antimicrobial and cytotoxic activities tested in this work.

Noteworthy, post-assembly-line tailoring reactions are not indispensable. For example, myxothiazol A does not need any tailoring reactions after the release from the assembly line for its biological function[37]. In this work, all the structural characteristics featured by cystopipecotides are conferred by the incorporated building blocks and no tailoring reactions were observed during the biosynthesis of the cystopipecotides. The lack of any observed antibacterial activity of the cystopipecotides hints at a different biological function of this new compound family which is currently unknown.

4.4 Genomics and metabolomics in natural product research — still promising?

The history of natural products discovery is accompanied by the exploration and development of a plethora of methods and techniques. Bioactivity-based, metabolomic-driven and genomic-driven methods are nowadays most prevalent strategies pursued in the laboratory, albeit each approach has its own limitations and could only explore partially the diversity of microbial natural products.

Since most bioactivity-guided experiments need higher quantities of compound, secondary metabolites in low amount are prone to be ignored. Moreover, the number of tested strains in one laboratory is limited so that some metabolites with more specific functions would be inevitably overlooked. Therefore, newly identified species would be easily exhausted and loses its priority if only specific biological activities are pursued. For example, ajudazols C–J and cystopipecotides discovered in this work would be ignored following specific bioactivity-guided screening. Normally, the specific biological functions and activities of some natural products which are discovered due to their unique structural features are often revealed later.

Molecular networking and principal component analysis (PCA), these two commonly used metabolomics strategies, indeed provide mega information about the metabolites of the strain under different growth conditions. However, the alteration of growth conditions can have a pleiotropic influence on both primary and secondary metabolism, differentiating valuable secondary metabolites from tremendous molecular features indicated by molecular networking and PCA would be time-consuming and empirical. Besides, the number of metabolites picked up for MS/MS at a certain retention time is objectively limited by the mass spectrometer whereas the metabolites elute at the same retention time may exceed the capacity of mass spectrometer. Therefore, some potential metabolites under the background of others with abundant intensities may be neglected in the network. For example, the molecular network of SBCb004 metabolites was also constructed (data not shown here) but cystopipecotides whose mass spectral signals overlap with other metabolites were overlooked by the software and were not included in the ultimate constructed network.

Generally, the species which belongs to distant phylogenetic clans, feature significant bioactivity or show unique morphological or physiological characteristics are prioritized for whole genome sequencing as these features indicate unexplored realm of resource and the potential of unprecedented secondary metabolites. Thus, those BGCs hidden in the genome of universal species are prone to be neglected. Moreover, the genetic organization and localization of BGCs within the genome can be ambiguous rather than following textbook's principles, which impedes the prediction of the underlying biosynthetic pathway and genetic manipulation. The expression of BGCs in the native host can be influenced by multiple factors such as the absence of precursor, the inhibition of transcriptional factors and the competition of other BGCs[38]. Transcriptomics and proteomics information would be helpful to know the expression state of each BGC regardless of the financial cost. Besides, the manipulation of BGCs is restricted by genetic and molecular biological tools in each

individual species. The genetic mobilization of huge BGCs ranging from 20-200 kb which might also feature multiple repetitive sequences remains challenging and work-intensive[39] despite the wide application of BGC cloning tools such as direct cloning and transformation-associated recombination cloning[40,41].

It recently becomes a trend to utilize multiple approaches to bridge the gap between the genome and the connected secondary metabolites[42,43]. Once the newly isolated microbial species has been identified, the new isolate would be grown under different conditions based on the principle of "One strain-many compounds (OSMAC)", followed by bioassay and metabolites profiling analysis. With the help of global or in-house natural product databases, known metabolites can be identified and excluded from further investigations. Only when the exhibited activity is not connected to a known metabolite, activity-guided fractionation and compound isolation is pursued further. In the meanwhile, based on the Global Natural Product Social (GNPS) platform, a molecular network of metabolites can be constructed and structures with the same molecular features are clustered in distinct groups. Thus, the derivatives or analogs of known structures can be predicted easily with the information of mass spectra. As PCA can be used as an analytical method to figure out how the production of metabolites is influenced through environmental or genetic alterations, newly occurring mass spectral signals in the respective crude extract induced by specific growth conditions could be sensitively detected. Despite bioactivity-based and metabolomic-based strategies, genomic analyses also provide useful information. With genomic information at hand, the known secondary metabolites with identified BGCs in the extracts can be dereplicated. Those unknown BGCs annotated by bioinformatics tools such as antiSMASH present huge potential. Various approaches including heterologous expression and promoter replacement could be used to uncover the associated natural products of these cryptic or silent BGCs (Figure 4.4).

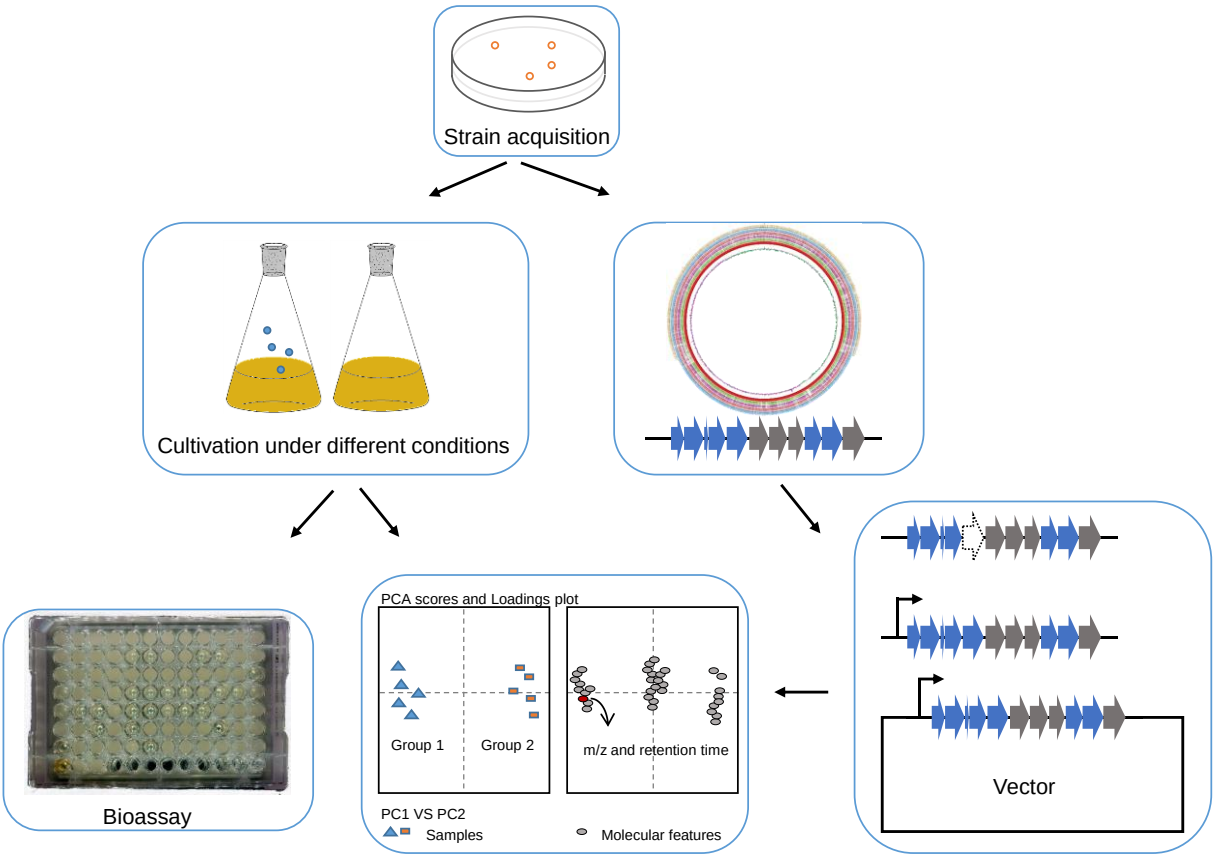


Figure 4.4 Workflow to discover novel secondary metabolites through different strategies. The newly isolated microbial species is cultivated under different conditions followed by bioassay and metabolomics analyses. In parallel, with the genomic information at hand, it is possible to connect the biosynthetic gene cluster to the corresponding class of natural product via optional genetic alterations and modifications.

4.5 Conclusion and outlook

This work payed specific attention to the previously investigated species SBCb004[20,27] and interactively used genomics and metabolomics approaches to unveil the "hidden" secondary metabolites. The discovery of ajudazols C–J in SBCb004 was initiated through the identification of a PKS-NRPS BGC which is highly similar to the known ajudazol BGC from *Chondromyces crocatus* Cm c5, followed by dereplication of ajudazols A and B, and metabolites interrogation of the wild type strain. The BGC were successfully connected to the identified natural products by targeted genetic inactivation experiments and PCA. The discovery of cystopipecotides was prompted by the structural features provided by the metabolic profiling. The BGC was later identified based on the composition of the structures and the collinearity rule of NRPS and PKS.

Furthermore, the biosynthesis of the distinct structural features in both compound families were investigated. The novel ajudazols were revealed to exhibit a DMA-CoA as the initial building block and undergo varying degrees of hydroxylation and different glycosylation patterns. Two P450 enzymes responsible for the hydroxylation and desaturation and one glycosyltransferase processing the transfer of hexose were proposed based on inactivation experiments. A promiscuous network of the post-assembly-line-modifications was postulated as well. Cystopipecotides present a non-proteinogenic amino acid HHMPA as a building block as well as a γ-lactam at C-terminus. Feeding experiments and the analyses of the BGC organization showed that the C-terminal γ-lactam results from the lactamization of C-3 unit with tryptophan. The interrogation of the BGC uncovered an unprecedented route to synthesize substituted pipecolate through glutamic acid and ketose in microbe.

However, the cystopipecotides only showed weak cytotoxicity. Their biological functions still remain ambiguous. Although we revealed the wide distribution of HHMPA biosynthetic operon in *cystobacterineae*, whether the BGCs containing those operons are expressed under laboratory conditions or whether their potential products exhibit another biological functions are still not clear. These all need further investigation and evaluation in the future.

4.6 References

(1) Schäberle, T. F.; Lohr, F.; Schmitz, A.; König, G. M. Antibiotics from myxobacteria. *Nat. Prod. Rep.* 2014, *31* (7), 953–972. DOI: 10.1039/c4np00011k.

(2) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 2019 (47), W81-W87. DOI: 10.1093/nar/gkz310.

(3) Dickschat, J. S.; Bode, H. B.; Mahmud, T.; Müller, R.; Schulz, S. A novel type of geosmin biosynthesis in myxobacteria. *J. Org. Chem.* 2005, *70* (13), 5174–5182. DOI: 10.1021/jo050449g.

(4) Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from Myxococcus xanthus. *ACS Chem. Biol.* 2020, *15* (8), 2221–2231. DOI: 10.1021/acschembio.0c00430.

(5) Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Kohnke, J.; Trauner, D.; Müller, R. Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.* 2017 (56), 1–5. DOI: 10.1002/anie.201612640.

(6) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D. D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G. G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. E.; Tagg, J. R.; Tang, G. L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* 2013, *30* (1), 108–160. DOI: 10.1039/C2NP20085F.

(7) Silakowski, B.; Schairer, H. U.; Ehret, H.; Kunze, B.; Weinig, S.; Nordsiek, G.; Brandt, P.; Blöcker, H.; Höfle, G.; Beyer, S.; Müller, R. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* 1999, *274* (52), 37391–37399. DOI: 10.1074/jbc.274.52.37391.

(8) Gerth, K.; Irschik, H.; Reichenbach, H.; Trowitzsch, W. Myxothiazol, an antibiotic from *Myxococcus fulvus* (myxobacterales). I. Cultivation, isolation, physico-chemical and biological properties. *J. Antibiot.* 1980, *33* (12), 1474–1479.

(9) Kunze, B.; Reichenbach, H.; Müller, R.; Höfle, G. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (myxobacteria). *J. Antibiot.* 2005, *58* (4), 244–251.

(10) Gorges, J.; Panter, F.; Kjaerulff, L.; Hoffmann, T.; Kazmaier, U.; Müller, R. Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* 2018, *57* (43), 14270–14275. DOI: 10.1002/anie.201808028.

(11) Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K. I.; Hüttel, S.; Harmrolfs, K.; Stadler, M.; Müller, R. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* 2014, *53* (52), 14605–14609. DOI: 10.1002/anie.201409964.

(12) Buntin, K.; Rachid, S.; Scharfe, M.; Blöcker, H.; Weissman, K. J.; Müller, R. Production of the antifungal isochromanone ajudazols A and B in *Chondromyces crocatus* Cm c5: biosynthetic machinery and cytochrome P450 modifications. *Angew. Chem. Int. Ed. Engl.* 2008, *47* (24), 4595–4599. DOI: 10.1002/anie.200705569.

(13) Li, Y.; Luxenburger, E.; Müller, R. An alternative isovaleryl CoA biosynthetic pathway involving a previously unknown 3-methylglutaconyl CoA decarboxylase. *Angew. Chem. Int. Ed. Engl.* 2012, *52* (4), 1304–1308. DOI: 10.1002/anie.201207984.

(14) Alma L. Díaz-Pérez; César Díaz-Pérez; Jesús Campos-García. Bacterial l -leucine catabolism as a source of secondary metabolites. *Rev Environ Sci Biotechnol* 2016, *15* (1), 1–29. DOI: 10.1007/s11157-015-9385-3.

(15) Gatto, G. J.; Boyne, M. T.; Kelleher, N. L.; Walsh, C. T. Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J. Am. Chem. Soc.* 2006, *128* (11), 3838–3847. DOI: 10.1021/ja0587603.

(16) Hartmann, M.; Kim, D.; Bernsdorff, F.; Ajami-Rashidi, Z.; Scholten, N.; Schreiber, S.; Zeier, T.; Schuck, S.; Reichel-Deland, V.; Zeier, J. Biochemical Principles and Functional Aspects of Pipecolic Acid Biosynthesis in Plant Immunity. *Plant physiology* 2017, *174* (1), 124–153. DOI: 10.1104/pp.17.00222.

(17) Wickwire, B. M.; Harris, C. M.; Harris, T. M.; Broquist, H. P. Pipecolic acid biosynthesis in Rhizoctonia leguminicola. I. The lysine saccharopine, delta 1-piperideine-6-carboxylic acid pathway. *Journal of Biological Chemistry* 1990, *265* (25), 14742–14747.

(18) Chen, M.; Liu, C.-T.; Tang, Y. Discovery and Biocatalytic Application of a PLP-Dependent Amino Acid γ-Substitution Enzyme That Catalyzes C-C Bond Formation. *J. Am. Chem. Soc.* 2020, *142* (23), 10506–10515. DOI: 10.1021/jacs.0c03535.

(19) Hai, Y.; Chen, M.; Huang, A.; Tang, Y. Biosynthesis of Mycotoxin Fusaric Acid and Application of a PLP-Dependent Enzyme for Chemoenzymatic Synthesis of Substituted l-Pipecolic Acids. *J. Am. Chem. Soc.* 2020, *142* (46), 19668–19677. DOI: 10.1021/jacs.0c09352. Published Online: Nov. 6, 2020.

(20) Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K. J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* 2010, *17* (3), 296–309. DOI: 10.1016/j.chembiol.2010.01.016.

(21) Kochetov, G. A.; Solovjeva, O. N. Structure and functioning mechanism of transketolase. *Biochimica et biophysica acta* 2014, *1844* (9). DOI: 10.1016/j.bbapap.2014.06.003.

(22) Peng, C.; Pu, J.-Y.; Song, L.-Q.; Jian, X.-H.; Tang, M.-C.; Tang, G.-L. Hijacking a hydroxyethyl unit from a central metabolic ketose into a nonribosomal peptide assembly line. *PNAS* 2012, *109* (22), 8540–8545. DOI: 10.1073/pnas.1204232109.

(23) Fu, C.; Auerbach, D.; Li, Y.; Scheid, U.; Luxenburger, E.; Garcia, R.; Irschik, H.; Müller, R. Solving the puzzle of the one-carbon loss in ripostatin biosynthesis. *Angew. Chem. Int. Ed. Engl.* 2017, *56* (8), 2192–2197. DOI: 10.1002/anie.201609950.

(24) Qi, F.; Lei, C.; Li, F.; Zhang, X.; Wang, J.; Zhang, W.; Fan, Z.; Li, W.; Tang, G.-L.; Xiao, Y.; Zhao, G.; Li, S. Deciphering the late steps of rifamycin biosynthesis. *Nature communications* 2018, *9* (1), 2342. DOI: 10.1038/s41467-018-04772-x.

(25) Matsuda, K.; Hasebe, F.; Shiwa, Y.; Kanesaki, Y.; Tomita, T.; Yoshikawa, H.; Shin-Ya, K.; Kuzuyama, T.; Nishiyama, M. Genome Mining of Amino Group Carrier Protein-Mediated Machinery: Discovery and Biosynthetic Characterization of a Natural Product with Unique Hydrazone Unit. *ACS chemical biology* 2017, *12* (1). DOI: 10.1021/acschembio.6b00818.

(26) Hasebe, F.; Matsuda, K.; Shiraishi, T.; Futamura, Y.; Nakano, T.; Tomita, T.; Ishigami, K.; Taka, H.; Mineki, R.; Fujimura, T.; Osada, H.; Kuzuyama, T.; Nishiyama, M. Amino-group carrier-protein-mediated secondary metabolite biosynthesis in Streptomyces. *Nature chemical biology* 2016, *12* (11). DOI: 10.1038/nchembio.2181.

(27) Pogorevc, D.; Tang, Y.; Hoffmann, M.; Zipf, G.; Bernauer, H. S.; Popoff, A.; Steinmetz, H.; Wenzel, S. C. Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* 2019, *8* (5), 1121–1133. DOI: 10.1021/acssynbio.9b00023.

(28) Surup, F.; Viehrig, K.; Mohr, K. I.; Herrmann, J.; Jansen, R.; Müller, R. Disciformycins A and B: 12-membered macrolide glycoside antibiotics from the myxobacterium Pyxidicoccus fallax active against multiresistant staphylococci. *Angew. Chem.* 2014, *53* (49), 13588–13591. DOI: 10.1002/anie.201406973.

(29) Etzbach, L.; Plaza, A.; Garcia, R.; Baumann, S.; Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Org. Lett.* 2014, *16* (9), 2414–2417. DOI: 10.1021/ol500779s.

(30) Kopp, M.; Rupprath, C.; Irschik, H.; Bechthold, A.; Elling, L.; Muller, R. SorF: a glycosyltransferase with promiscuous donor substrate specificity in vitro. *ChemBioChem* 2007, *8* (7), 813–819. DOI: 10.1002/cbic.200700024.

(31) Thibodeaux, C. J.; Melançon, C. E.; Liu, H.-w. Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angewandte Chemie (International ed. in English)* 2008, *47* (51), 9814–9859. DOI: 10.1002/anie.200801204.

(32) Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annual review of biochemistry* 2008, *77*, 521–555. DOI: 10.1146/annurev.biochem.76.061005.092322.

(33) Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function. *Nat. Prod. Rep.* 2017, *34* (9), 1141–1172. DOI: 10.1039/c7np00034k.

(34) Rix, U.; Fischer, C.; Remsing, L. L.; Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* 2002, *19* (5), 542–580. DOI: 10.1039/b103920m.

(35) Weymouth-Wilson, A. C. The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep.* 1997, *14* (2), 99–110.

(36) Schlünzen, F.; Zarivach, R.; Harms, J.; Bashan, A.; Tocilj, A.; Albrecht, R.; Yonath, A.; Franceschi, F. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* 2001, *413* (6858), 814–821. DOI: 10.1038/35101544.

(37) Perlova, O.; Fu, J.; Kuhlmann, S.; Krug, D.; Stewart, A. F.; Zhang, Y.; Muller, R. Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in Myxococcus xanthus. *Appl. Environ. Microbiol.* 2006, *72* (12), 7485–7494. DOI: 10.1128/AEM.01503-06.

(38) Hug, J. J.; Müller, R. Host Development for Heterologous Expression and Biosynthetic Studies of Myxobacterial Natural Products: *Comprehensive Natural Products III: Chemistry and Biology, Chapter 14818* 2020, *In press.* DOI: 10.1016/B978-0-12-409547-2.14818-8.

(39) Fayed, B.; Ashford, D. A.; Hashem, A. M.; Amin, M. A.; El Gazayerly, O. N.; Gregory, M. A.; Smith, M. C. M. Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in Streptomyces spp. and combinatorial biosynthesis. *Applied and environmental microbiology* 2015, *81* (24), 8402–8413. DOI: 10.1128/AEM.02403-15.

(40) Zhang, J. J.; Yamanaka, K.; Tang, X.; Moore, B. S. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods in enzymology* 2019, *621*, 87–110. DOI: 10.1016/bs.mie.2019.02.026.

(41) Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Muller, R.; Stewart, A. F.; Fu, J.; Zhang, Y. RecET direct cloning and Redαβ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* 2016, *11* (7), 1175–1190. DOI: 10.1038/nprot.2016.054.

(42) Wenzel, S. C.; Müller, R. The impact of genomics on the exploitation of the myxobacterial secondary metabolome. *Nat. Prod. Rep.* 2009, *26* (11), 1385–1407. DOI: 10.1039/b817073h.

(43) Krug, D.; Müller, R. Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* 2014, *31* (6), 768–783. DOI: 10.1039/c3np70127a.