Discovery of natural products through heterologous expression of biosynthetic gene clusters in *Streptomyces albus*

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

Von

Hui Shuai

Saarbrücken

2020

Tag des Kolloquiums: 14. Dezember 2020

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

Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisor Prof. Andriy Luzhetskyy for the opportunity he granted me to complete my PhD thesis in his research group, for his insightful tutoring and for the fruitful scientific discussions we had.

I express my special gratitude to Prof. Rolf Müller who kindly accepted to be my second supervisor and reviewed my dissertation.

I want to express my gratitude to all of the former and current group members for their help, support and friendly working environment. My gratitude belongs to my mentor Dr. Maksym Myronovskyi, who worked with me since the first day in lab. Especially I would like to thank Birgit Rosenkränzer who facilitated the entry into the new work environment and laboratory work. I owe my thanks also to Dr. Suvd Nadmid, Dr. Constanze Paulus, Dr. Nils Gummerlich, Dr. Josef Zapp and Marc Stierhof for supporting me in NMR structure elucidation.

I would also take this opportunity to express my gratitude to my dear motherland China for provided financial support during my stay in Germany.

Finally, I would like to thank to my friends and my family for their support and encouragements during these years.

Summary

This thesis focused on the identification of new natural products through heterologous expression of their cluster in the *Streptomyces albus* Del14 strain. Genomic library of a rare actinomycete strain *Kutzneria albida* DSM 43870 was used as a source of secondary metabolite clusters. Two clusters of *K. albida* were successfully expressed in *S. albus* Del14. Expression of a cryptic nucleoside gene cluster led to the production of a new member of pyrrolopyrimidine family of natural products which was called huimycin. Expression of the cryptic NRPS cluster led to the production of a new cyclopeptide called cyclohuinilsopeptin A. Both compounds were isolated and their structures were elucidated by NMR. Biosynthetic pathways for the isolated compounds were proposed based on sequence analysis and on the results of gene inactivation studies.

Another focus of this work was deciphering of the biosynthetic pathway leading to the production of isoquinoline alkaloids mansouramycins in *Streptomyces albus*. Mansouramycin biosynthetic cluster was identified by heterologous expression and gene inactivation studies. Results of extensive genetic experiments and feeding studies indicate that the identified biosynthetic pathway leading to the production of isoquinolines mansouramycins is new. In contrast to all known isoquinoline biosynthetic routes, tryptophan instead of tyrosine serves as a main precursor for the production of mansouramycins.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Identifizierung neuer Naturstoffe durch heterologe Expression biosynthetischer Gencluster in dem Stamm *Streptomyces albus* Del14. Eine Genbibliothek eines Actinomyceten, *Kutzneria albida* DSM 43870, wurde als Quelle der Biosynthesegencluster verwendet. Zwei Gencluster konnten erfolgreich in *S. albus* Del14 exprimiert werden. Die Expression eines kryptischen Nucleosid-Genclusters führte zur Produktion eines Naturstoffs aus der Gruppe der Pyrrolopyrimidine, welcher Huimycin genannt wurde. Darüber hinaus führte die Expression eines NRPS Genclusters zur Produktion eines neuen cyclischen Peptids namens Cyclohuinilsopeptin A. Beide Stoffe wurden isoliert und die Struktur wurde mittels NMR aufgeklärt. Vermutungen zur Biosynthese wurden basierend auf der Analyse des jeweiligen Clusters und auf Ergebnissen von Geninaktivierungen erstellt.

Ein weiterer Fokus dieser Arbeit liegt auf der Entschlüsselung der Biosynthese des Isochinolin-Alkaloids Mansouramycin in *Streptomyces albus*. Der Gencluster, welcher für die Biosynthese verantwortlich ist, wurde durch heterologe Expression und Geninaktivierungen identifiziert. Die Ergebnisse umfangreicher genetischer Experimente sowie Fütterungsstudien führten zur Entdeckung des neuen Biosynthesewegs des Isochinolins Mansouramycin. Im Gegensatz zu bekannten Biosynthesewegen von Isochinolinen dient Tryptophan anstelle von Tyrosin als Vorstufe für die Produktion vom Isochinolin Mansouramycin D.

Publications

Shuai, H.; Myronovskyi, M.; Nadmid, S.; Luzhetskyy, A. Identification of a Biosynthetic Gene Cluster Responsible for the Production of a New Pyrrolopyrimidine Natural Product-Huimycin Biomolecules 2020 10(7):E1074.

Shuai, H.; Gummerlich, N.; Zapp, J.; Myronovskyi, M.; Luzhetskyy, A. Identification of a Biosynthetic Gene Cluster Responsible for the Production of a New Cyclopeptide Cyclohuinilsopeptin. To be published.

Shuai, H.; Rosenkränzer, B.; Paulus, C.; Nadmid, S.; Myronovskyi, M.; Luzhetskyy, A. Biosynthesis of Isoquinolines Mansouramycins in *Streptomyces albus* J1074. To be published.

Shuai, H.; Zapp, J.; Myronovskyi, M.; Luzhetskyy, A. New mansouramycin derivatives from *Streptomyces albus*. To be published.

Contents

Chapter 1: Introduction

1 Natural products in drug discovery

Natural products have inspired chemists and pharmacists for millennia. The wide range of diversity of multi-dimensional chemical structures and therapeutic properties made them the most fruitful source of potential drug leads [1]. A lot of indispensable drugs such as penicillin, streptomycin, erythromycin, avermectin, morphine etc. were derived from natural products [2–6]. Since the day of discovery, many of those drugs did not lose their importance and are still widely used. A detailed analysis of new medicines approved by the US Food and Drug Administration (FDA) from 1981 to 2019 reveals that more than one-third of therapeutic agents are derived from or inspired by natural products. More than 50% of the developed small-molecule drugs originated from natural products, semi-synthetic nature products or natural product-derived mimetics[7]. Therefore natural products are of utmost importance as a source of clinical drugs.

In ancient times, most of the bioactive natural products were discovered by the conventional trial-and-error method [8]. In 1929 Alexander Fleming with the discovery of the first antibiotic penicillin demonstrated a new biochemical approach for the discovery of new natural products which revolutionized drug discovery research. The following discovery of streptomycin in 1942 by Selman A Waksman can be regarded as a start of the god age of natural product discovery [2]. During this time, tetracycline, erythromycin and many other types of antibiotics, mainly from actinomycetes, were brought onto the market [9]. Altogether the discovery of antibiotics made a great contribution to the extension of people`s lifespan and has saved millions of lives worldwide [10].

In 1990s a gradual decrease in the efficiency of the traditional drug discovery strategy could be observed. Direct screening of the extracts for desirable bioactivity, followed by compound isolation and characterization was leading increasingly often to rediscovery of already known compounds resulting in general discouragement and cessation of screening programs[11]. The drug discovery pipeline was inevitably almost running dry with no new class of antibiotic being brought to the market found in the last 20 years. In the light of the lack of new antibiotics, the rise of antibiotic resistance has become one of the biggest public health challenges of our time. Nowadays, fighting the threat of multidrug-resistant infections is a public health priority [12].

Considering the urgent need for new drugs to fight the infectious crisis, there has never been a higher necessity and desire for drugs leads than now. Actinobacteria as a prolific source of structurally diverse bioactive metabolites produce two-thirds of all known antibiotics used in the clinic today, a vast array of anticancer compounds, immunosuppressants, anthelmintics and antiviral compounds [13–15]. Advance in genome sequencing and genome mining technologies give evidence that the potential of actinobacteria to produce natural products is far from being exhausted [16]. Direct heterologous expression and refactoring of their biosynthetic gene cluster is a

new approach called to solve the rediscovery problem. Furthermore, evidence exists that exploration of rare underexplored taxa of actinobacteria with modern natural product discovery tools may lead to the discovery of numerous novel bioactive compounds [17–21]. Hence, Actinobacteria can still be a promising source for novel bioactive natural products and have excellent prospects for further exploration.

The main focus of this is the discovery of new natural products via heterologous expression of actinobacterial secondary metabolite clusters in a chassis strain *Streptomyces albus* Del14.

2 Microbial production of selected groups of natural products

2.1 General overview

Microorganisms are talented chemists with the ability to generate tremendously complex and diverse natural products that harbor potent biological activities [22]. Secondary metabolites are biologically active small molecules that are not essential for survival under laboratory conditions but which provide a competitive advantage to the producer microbes [23]. Secondary metabolites are a significant source of natural products that have the potential to be new antibiotics, immune suppressants, and other drugs [24]. In most of the cases, the multienzyme biochemical pathway required for the biosynthesis of natural products are encoded by the genes which are clustered together in bacterial chromosomes [25]. Therefore the identification of a single gene involved in the biosynthesis of a particular secondary metabolite is sufficient for the identification of a whole biosynthetic pathway.

Bacterial secondary metabolites are often classified into the following main groups of compounds: terpenoids and steroids, enzyme cofactors, alkaloids, nucleosides, polyketides and nonribosomal polypeptides [26]. Since alkaloids, nucleosides and nonribosomal polypeptides were objects of this study a short introduction into these classes of compounds is provided below.

2.2 Pyrrolopyrimidine group of nucleoside natural products

Pyrrolopyrimidines is a class of natural products characterized by the presence of 7-deazapurine moiety in their structure [27]. This class of compounds is widely spread among three domains of life: archaea, bacteria, and eukarya [27]. The study on pyrrolopyrimidines began with the discovery of an anti-*Candida* compound, toyocamycin, in 1956, from the culture broth of *Streptomyces toyocaensis* [28]. Soon after the discovery of toyocamycin, other diffusing 7-deazapurines tubercidin and sangivamycin were isolated from the culture filtrates of *Streptomyces tubercidicus* and *Streptomyces rimosus* [29,30]. To date, more than 30 different 7-deazapurines have been isolated from various biological sources, including bacteria, cyanobacteria, red algae, marine sponges, tunicates, etc. [27]. In addition to diffusing pyrrolopyrimidines which act as natural products, pyrrolopyrimidines can also be found as modified bases in tRNA. Quenosine is one of the

most studied, modified nucleosides containing 7-deazapurine moiety [31]. It is present in the wobble position of aspartyl, asparaginyl, histidyl and tyrosyl tRNA, where it likely modifies base-pairing characteristics of tRNA and enhances translational efficiency [32–34]. Distributed ubiquitously among prokaryotes and eukaryotes, quenosine cannot be found in archaebacteria. In archaeal species archeosine, non-canonical 7-deazapurine nucleoside, is incorporated in tRNA [35].

The genes responsible for the production of 7-deazapurines have been identified in the genome of the toyocamycin and sangivamycin producer *Streptomyces rimosus* in 2008 [27,36,37].

2.3 Isoquinoline family of alkaloids

Alkaloids are a large group of substances with more than 12000 members known at present. Alkaloids mostly contain basic nitrogen atoms in their structure which are usually situated in the cyclic group [38]. Alkaloids are mostly wide spread in plants. More than 30 different kinds of alkaloids were isolated only from opium poppy [39]. Actinobacteria as a source of natural products of alkaloid family remain underexplored [40].

Isoquinoline alkaloids are a large and diverse group of heterocyclic aromatic organic compounds with around 2500 defined structures [41]. Isoquinoline is a structural isomer of quinoline. Isoquinoline and quinoline are benzopyridines, which are composed of a benzene ring fused to a pyridine ring [41]. Aromatic amino acid tyrosine serves as a main biosynthetic precursor for the formation of the isoquinoline ring in plant natural compounds [42]. Isoquinoline alkaloids have remarkable biological activities and have demonstrated a critical role in the development of new anticancer drugs [41].

The isoquinoline moiety of alkaloids formed through the cyclization of the Schiff base formed between the dopamine and 4-hydroxyphenylacetaldehyde. Both biosynthetic intermediates are derived from tyrosine. The condensation of dopamine and 4-hydroxyphenylacetaldehyde followed by cyclization, hydroxylation, and methylation gives rise to (S)-reticuline, a central biosynthetic intermediate of all isoquinoline alkaloids [43].

2.4 Biochemistry of PKS and NRPS and PKS/NRPS or NRPS/PKS hybrids

Polyketides and nonribosomal peptides natural products are two major classes of secondary metabolites which constitute a broad class of compounds with various structural features and biological activities [44]. Both are produced by multi-enzymatic, multi-domain mega-synthases polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) encoded in discrete genomic clusters[45].

Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) — are also seen in PKS-NRPS or NRPS-PKS hybrids, seemingly chimeric genes creating a chimeric compound[46].

The products of hybrids are often bioactive, e.g., the mycotoxins cyclopiazonic acid, pyranonigrin, and cytochalasin [47].

Polyketide synthases (PKSs) assemble small acetic acid-type acyl building blocks into polyketides through C–C bonds, and nonribosomal peptide synthetases (NRPSs) assemble amino acids into peptides through amide bonds [48]. Apparently, type I PKSs and NRPSs share a quite similar strategy for the biosynthesis of two distinct classes of secondary metabolites. Both systems use carrier proteins to tether the growing chain [48].

Generally, one module within type I PKS is responsible for the introduction of either a starter or one elongation unit into the nascent polyketide chain (Figure 1). Within one PKS module, three separate core domains can be recognized: an acyltransferase (AT) domain, a ketosynthase (KS) domain and an acyl carrier (ACP) domain [49]. Starter modules of type I PKS consist only of an AT and ACP domains. An elongation module consists of three domains: KS, AT and ACP. The last elongation module contains additionally a thioesterase domain responsible for the release of newly synthesized polyketide chain [49]. The AT domain catalyzes the transfer of the activated acyl-CoA starter or elongation unit onto the phosphopantetheine arm of the ACP to form a thioester conjugate. The KS domain, which is found in elongation (extending) modules, catalyzes a decarboxylative Claisen condensation between two neighboring ACP tethered acyl units. While a

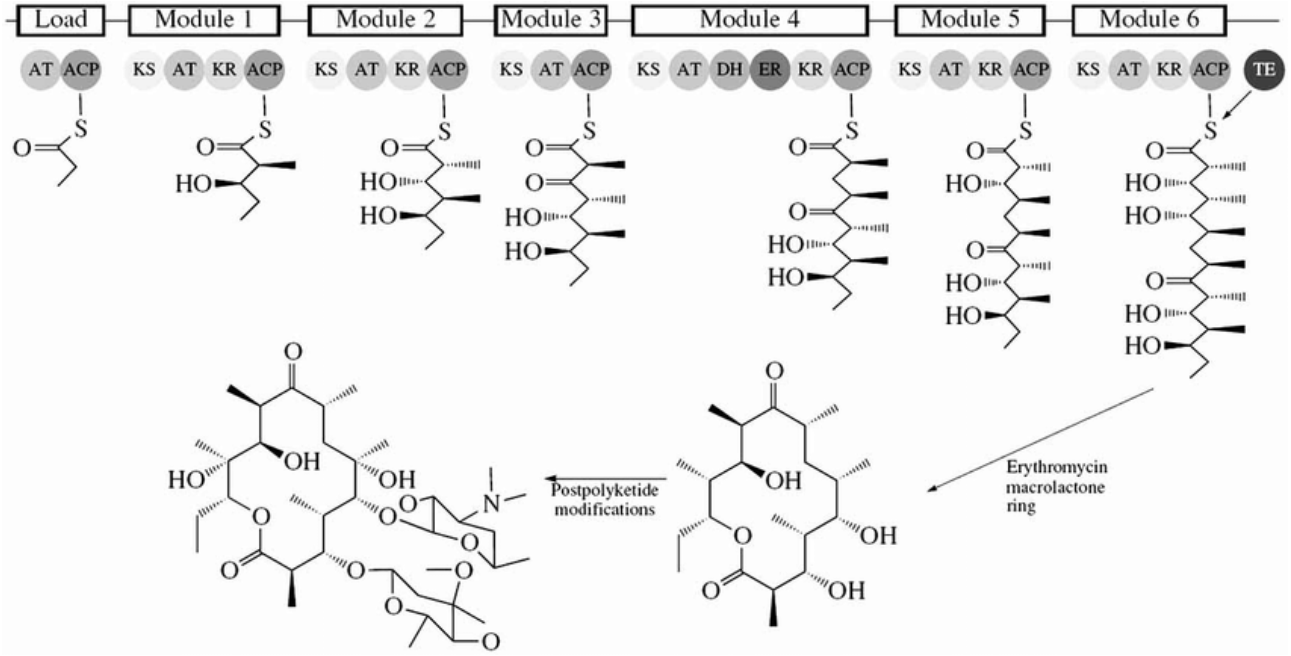

Figure 1. Biosynthesis of erythromycin by type I PKS [50].

variety of acyl-CoA substrates can be utilized as a starter unit, mainly malonyl-CoA or methylmalonyl-CoA is used for elongation of the polyketide chain. In addition to the above-mentioned core domains, the modules of type I PKS may include additional catalytic domains. The keto-reductase (KR) domain reduces the β-keto group to a β-hydroxy group. The latter can be further modified by the dehydratase (DH) domain which removes water and leads to the formation

of α-β-unsaturated alkene. The reduction of the double bond of the nascent polyketide chain is accomplished by the enoyl reductase (ER) domain. An optional methyltransferase domain can further diversify the products of PKS [49]. The elongation reaction can be repeated until the end of the assembly line; Finally, Termination or releasing domain-TE, The TE domain hydrolyzes the completed polyketide chain from the ACP-domain of the previous module [51].

The functional logic of NRPS systems is similar to that of type I PKS (Figure 2). The NRPS multienzyme complex can be dissected into modules which are responsible for the attachment of single amino acid residues. Three different types of core domains can be identified within NRPS modules: an adenylation (A) domain, peptidylcarrier (PCP or T) domain and condensation (C) domain. Initiation modules of NRPS mostly only A and PCP domains, while the elongation domains consist of C, A, and PCP domains [52]. The A domain share the same chemistry with aminoacyl-tRNA synthetases (AARSs) which act in the initial step of ribosome-driven protein synthesis [52]. The A domains specifically activate amino acids and load them onto the phosphopantetheine chain of PCP domain. The main function of the swinging phosphopantetheine group of PCP domain is the transfer of the amino acid monomers or growing peptide chain to various catalytic sites [52]. The C domain catalyzes the formation of an amide bond by transferring the aminoacyl group or the growing peptidyl chain from the donor PCP domain onto the amino group of the aminoacyl group bound to the acceptor PCP domain. The last NRPS module of NRPS contains similarly to type I PKS a thioesterase (TE) domain which releases the linear peptide intermediate by either hydrolysis or internal cyclization [53]. The NRPS module sometimes include epimerase (E), methyltransferase (MT), and oxidase (Ox) domains [52].

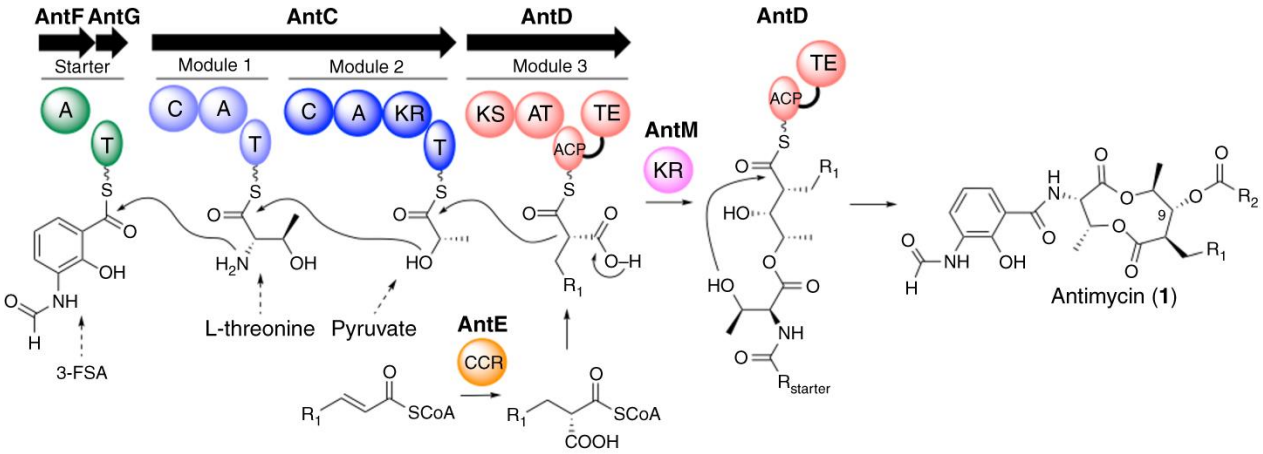

Figure 2. Biosynthesis of antimycin by hybrid PKS-NRPS [54]

In the PKS–NRPS hybrid system, both PKS and NRPS systems are present in the assembly line. The hybrid assembly lines build structurally complex polyketide–amino acid/peptide hybrid molecules that incorporate both acyl and aminoacyl building blocks into their products [55]. Their combined functionalities expand the biological activities of the products by mixing their chemical

properties. Epothilone, rapamycin, and bleomycin are essential examples of such hybrid systems [55].

3 Secondary metabolite production through heterologous expression in Streptomycetes

3.1 General aspects

The expression of a gene or a set of genes from one organism in a different host is known as heterologous expression. Traditionally in genetics of actinomycetes, heterologous gene expression has been used to confirm the clustering of secondary metabolite biosynthetic genes, to analyze natural product biosynthesis, to produce variants of natural products by genetic engineering, and to discover new compounds by screening genomic libraries [56]. Nowadays, heterologous expression has become one of the most promising approaches to access new bioactive natural products [57]. With the advances in genome sequencing technology observed in the early 2000s, numerous uncharacterized biosynthetic gene clusters for natural products were unveiled in microbial genomes [16]. The vast majority of these gene clusters are typically silenced or regarded as cryptic – either the genes are not expressed during laboratory growth or yields of metabolites insufficient for direct isolation and characterization. Targeted heterologous expression of the secondary metabolite clusters in optimized chassis strains allows activation of these pathways and enables access to the encoded compounds [57].

3.2. BAC vectors for cloning and heterologous expression of biosynthetic gene clusters for natural products

Physical cloning of secondary metabolite clusters is a direct prerequisite for their activation and analysis through heterologous expression. Genetic analysis of clusters which are transcriptionally active in their native hosts is also often not possible without heterologous expression due to the refractiveness of these strains to genetic manipulations. Therefore the construction of genomic libraries of secondary metabolite producers is unavoidable in the natural product field.

Secondary metabolite clusters encoded within the genomes of actinobacteria can often reach the length of 100 kb. Cloning of these secondary metabolite clusters requires special vectors capable to support DNA inserts of such size. For the small clusters which often do not exceed 20 kb size construction of large-insert genome libraries is also beneficial. In this way, the size of the constructed libraries can be reduced and the chances to find the library clone with the entire clusters are increased. Both BAC and YAC vectors allow cloning of DNA fragments beyond 100 kb size [58,59]. BAC vector systems have found broader applications in constructing genomic libraries.

The BAC (bacterial artificial chromosome) system is based on the F factor of *Escherichia coli* which is present in a single copy. Cloning of DNA fragments approaching 200 kb in length was

reported directly at the moment of development of BAC technology [60]. Various BAC vectors have been designed for the construction of DNA libraries to facilitate genomic mapping and DNA sequencing efforts [61,62].In natural product research, the BAC libraries are constructed in E. coli and then transferred into Streptomyces hosts by conjugation [63].

3.3 Heterologous host strains

The success of heterologous expression of a secondary metabolite cluster greatly depends on a used heterologous host strain. The ideal heterologous host should be characterized by a clean metabolite background, established genetic manipulation tools and high levels of precursors required for the biosynthesis of target natural product. These features will greatly facilitate the discovery and characterization of the heterologously expressed natural products. The most widely used Streptomyces heterologous host strains for the expression of BGCs are derivatives of *Streptomyces coelicolor*, *Streptomyces lividans*, *Streptomyces albus* and *Streptomyces avermitilis* strains [64]. A cluster free derivative of *Streptomyces albus* – *S. albus Del14* was used in this study as a first choice heterologous host strain [65].

S. albus Del14 is the first cluster-free *Streptomyces albus* chassis strains with a minimized genome of 6.3 Mb and short life cycle – 4 days on solid agar medium. Fifteen biosynthetic gene clusters were deleted in fourteen deletion steps in the genome of *S. albus* Del14. Furthermore, *S. albus Del14* was modified by introducing additional *phi*C31 *att*B sites generating *S. albus* strains B2P1 and B4. This allowed the integration of multiple copies of the BGC into the chromosomes of the constructed strains, leading to higher production yields. The *S. albus Del14* strain shows markedly improved properties as compared to the parent strain such. The strain does not produce own secondary metabolites and has a clean metabolic background. The strain is also characterized by increased production yields and a high success rate of heterologous expression of biosynthetic gene clusters [65]. Many cryptic BGCs have been successfully expressed in the *S. albus.del14* strain, including tunicamycin B2, moenomycin M, 2,8-didemethoxyaranciamycinone, demethoxy aranciamycinone, griseorhodine A and cinnamycin [65].

4 Outline of this work

One of the objectives of this study was to elucidate the complex biosynthetic pathway leading to the production of isoquinoline alkaloids mansouramycins by *Streptomyces albus*. Discovery of new natural products through heterologous expression of biosynthetic gene clusters from a rare actinomycete *Kutzneria albida* was the second aim of this study.

4.1 Mansouramycins

Mansouramycins were first isolated from the marine strain *Streptomyces sp.* Mei37 in 2009 [66]. Until now, the biosynthetic route leading to the production of these compounds has not been reported. Production of mansouramycins A and D was detected in the extracts of the *S. albus* Del14 strain, which was previously developed in our lab. The strain is devoid of all secondary metabolite clusters which could be identified by genome mining software. This indicates that an unknown pathway leading to the production of mansouramycins is encoded in the genome of *S. albus*. The aim of this work was to identify the mansouramycin biosynthetic genes and to decipher mansouramycin biosynthesis.

Production of mansouramycin D by *S. albus* Del14 has been proved by isolation and structure elucidation. Feeding of labeled tyrosine which serves a precursor for plant-derived isoquinolines did not show incorporation into the mansouramycin D structure. Further feeding studies have demonstrated that the isoquinoline core of mansouramycin D is derived from tryptophan. This indicates that an alternative pathway leading to the production of isoquinolines is encoded in the *S. albus* genome. Homology search using the sequence of the genes involved in the degradation of tryptophan led to the identification of the mansouramycin gene cluster. The involvement of the genes in mansouramycin biosynthesis was proved by heterologous expression and gene inactivation analysis. The sequential order in which the products of the genes act in mansouramycin biosynthesis was determined by cocultivation studies with gene inactivation mutants. The results of feeding of various labeled biosynthetic precursors and isolation of biosynthetic intermediates provided some insights into the mansouramycin biosynthetic pathway.

4.2 Huimycin

A putative nucleoside gene cluster was identified in the genome of *K. albida* by genome mining with Antismash software [67]. Expression of the cluster in *S. albus* Del14 led to the identification of a new peak on LC-MS chromatogram. Isolation and structure elucidation of the compound revealed that the product of the expressed cluster is a new member of pyrrolopyrimidine family of natural products which was called huimycin. Toyocamycin is one of the best studied members of pyrrolopyrimidines of microbial origin. Structurally huimycin is closely related to dapiramycins A and B: dapiramycins share the same chromophore with huimycin; sugar moieties of

huimycin and dapiramycins are attached to the amino group at the 2nd position of the 7-deazapurine moiety. The minimal set of genes required for huimycin biosynthesis was identified through a series of gene deletion experiments. Based on the available data for toyocamycin biosynthesis and sequence analysis of huimycin gene cluster a biosynthetic pathway for huimycin was proposed.

4.3 Cyclohuinilsopeptin

Cyclohuinilsopeptin A is a new cyclic peptide obtained after heterologous expression of a cryptic NRPS cluster from *K. albida* in the *S. albus* Del14 strain. The compound was isolated and its structure was elucidated by NMR. Cyclohuinilsopeptin A has unique structural features like 3,4-cyclopropyl-proline and 2-(2,2,3-trimethylcyclopropyl)-glycine. In the course of gene inactivation experiments, a derivative of cyclohuinilsopeptin A was detected in the extracts of a mutant strain. The compound was isolated and its structure was successfully elucidated by NMR. The isolated derivative was named cyclohuinilsopeptin B. 2-(2,2,3-trimethylcyclopropyl)-glycine moiety is substituted with leucine residue in the structure of cyclohuinilsopeptin B. Based on results of gene inactivation studies and sequence analysis some insight into the biosynthesis of cyclohuinilsopeptin B were provided.

5 References

1. Newman, D.J.; Cragg, G.M.; Snader, K.M. The influence of natural products upon drug discovery (Antiquity to late 1999). *Nat. Prod. Rep.* **2000**, *17*, 215–234, doi:10.1039/a902202c.
2. Waksman, S.A. Streptomycin: Background, Isolation, Properties, and Utilization. *Science* **1953**, *118*, 259–266, doi:10.1126/science.118.3062.259.
3. Gaynes, R. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerg. Infect. Dis.* **2017**, *23*, 849–853, doi:10.3201/eid2305.161556.
4. Ottesen, E.A.; Campbell, W. Ivermectin in human medicine. *Journal of Antimicrobial Chemotherapy* **1994**, *34*, 195–203, doi:10.1093/jac/34.2.195.
5. Brook, K.; Bennett, J.; Desai, S.P. The Chemical History of Morphine: An 8000-year Journey, from Resin to de-novo Synthesis. *Journal of Anesthesia History* **2017**, *3*, 50–55, doi:10.1016/j.janh.2017.02.001.
6. Guay, D.R. Macrolide antibiotics in paediatric infectious diseases. *Drugs* **1996**, *51*, 515–536, doi:10.2165/00003495-199651040-00002.
7. 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*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
8. Dias, D.A.; Urban, S.; Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2012**, *2*, 303–336, doi:10.3390/metabo2020303.
9. Gould, K. Antibiotics: from prehistory to the present day. *Journal of Antimicrobial Chemotherapy* **2016**, *71*, 572–575, doi:10.1093/jac/dkv484.
10. Ventola, C.L. The antibiotic resistance crisis: part 1: causes and threats. *P T* **2015**, *40*, 277–283.
11. Aslam, B.; Wang, W.; Arshad, M.I.; Khurshid, M.; Muzammil, S.; Rasool, M.H.; Nisar, M.A.; Alvi, R.F.; Aslam, M.A.; Qamar, M.U.; et al. Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist* **2018**, *11*, 1645–1658, doi:10.2147/IDR.S173867.
12. Piddock, L.J. The crisis of no new antibiotics—what is the way forward? *The Lancet infectious diseases* **2012**, *12*, 249–253.
13. Barka, E.A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G.P. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 1, doi:10.1128/MMBR.00019-15.
14. Genilloud, O. Actinomycetes: still a source of novel antibiotics. *Nat. Prod. Rep.* **2017**, *34*, 1203–1232, doi:10.1039/C7NP00026J.
15. Rebets, Y.; Brötz, E.; Tokovenko, B.; Luzhetskyy, A. Actinomycetes biosynthetic potential: how to bridge in silico and in vivo? *J Ind Microbiol Biotechnol* **2014**, *41*, 387–402, doi:10.1007/s10295-013-1352-9.

16. Bentley, S.D.; Chater, K.F.; Cerdeño-Tárraga, A.-M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* **2002**, *417*, 141–147, doi:10.1038/417141a.
17. Sayed, A.M.; Hassan, M.H.A.; Alhadrami, H.A.; Hassan, H.M.; Goodfellow, M.; Rateb, M.E. Extreme environments: microbiology leading to specialized metabolites. *Journal of Applied Microbiology* **2020**, *128*, 630–657, doi:10.1111/jam.14386.
18. Tiwari, K.; Gupta, R.K. Rare actinomycetes: a potential storehouse for novel antibiotics. *Critical Reviews in Biotechnology* **2012**, *32*, 108–132, doi:10.3109/07388551.2011.562482.
19. Doroghazi, J.R.; Albright, J.C.; Goering, A.W.; Ju, K.-S.; Haines, R.R.; Tchalukov, K.A.; Labeda, D.P.; Kelleher, N.L.; Metcalf, W.W. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology* **2014**, *10*, 963–968, doi:10.1038/nchembio.1659.
20. Doroghazi, J.R.; Metcalf, W.W. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **2013**, *14*, 611, doi:10.1186/1471-2164-14-611.
21. Charlop-Powers, Z.; Pregitzer, C.C.; Lemetre, C.; Ternei, M.A.; Maniko, J.; Hover, B.M.; Calle, P.Y.; McGuire, K.L.; Garbarino, J.; Forgione, H.M.; et al. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc Natl Acad Sci USA* **2016**, *113*, 14811, doi:10.1073/pnas.1615581113.
22. 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*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
23. Craney, A.; Ahmed, S.; Nodwell, J. Towards a new science of secondary metabolism. *J Antibiot* **2013**, *66*, 387–400, doi:10.1038/ja.2013.25.
24. Maplestone, R.A.; Stone, M.J.; Williams, D.H. The evolutionary role of secondary metabolites — a review. *Gene* **1992**, *115*, 151–157, doi:10.1016/0378-1119(92)90553-2.
25. Cimermancic, P.; Medema, M.H.; Claesen, J.; Kurita, K.; Wieland Brown, L.C.; Mavrommatis, K.; Pati, A.; Godfrey, P.A.; Koehrsen, M.; Clardy, J.; et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421, doi:10.1016/j.cell.2014.06.034.
26. McMurry, J.E. *Organic chemistry with biological applications*; Cengage Learning, 2014; ISBN 1-285-84291-X.
27. McCarty, R.M.; Bandarian, V. Biosynthesis of pyrrolopyrimidines. *Bioorganic chemistry* **2012**, *43*, 15, doi:10.1016/j.bioorg.2012.01.001.

28. Nishimura, H.; Katagiri, K.; Sato, K.; Mayama, M.; Shimaoka, N. Toyocamycin, a new anti-candida antibiotics. *J. Antibiot.* **1956**, *9*, 60–62.
29. Anzai, K.; Nakamura, G.; Suzuki, S. A new antibiotic, tubercidin. *J. Antibiot.* **1957**, *10*, 201–204.
30. Rao, K.V. Structure of sangivamycin. *J. Med. Chem.* **1968**, *11*, 939–941, doi:10.1021/jm00311a005.
31. Vinayak, M.; Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci. Rep.* **2009**, *30*, 135–148, doi:10.1042/BSR20090057.
32. Harada, F.; Nishimura, S. Possible anticodon sequences of tRNA His , tRNA Asm , and tRNA Asp from Escherichia coli B. Universal presence of nucleoside Q in the first postion of the anticondons of these transfer ribonucleic acids. *Biochemistry* **1972**, *11*, 301–308, doi:10.1021/bi00752a024.
33. Meier, F.; Suter, B.; Grosjean, H.; Keith, G.; Kubli, E. Queuosine modification of the wobble base in tRNAHis influences “in vivo” decoding properties. *EMBO J.* **1985**, *4*, 823–827.
34. Urbonavicius, J.; Qian, Q.; Durand, J.M.; Hagervall, T.G.; Björk, G.R. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **2001**, *20*, 4863–4873, doi:10.1093/emboj/20.17.4863.
35. Kilpatrick, M.W.; Walker, R.T. The nucleotide sequence of the tRNAMMet from the archaebacterium Thermoplasma acidophilum. *Nucleic Acids Res.* **1981**, *9*, 4387–4390, doi:10.1093/nar/9.17.4387.
36. McCarty, R.; Bandarian, V. Rosetta stone for deciphering deazapurine biosynthesis: pathway for pyrrolopyrimidine nucleosides toyocamycin and sangivamycin. *Chem Biol* **2008**, *15*, 790–798, doi:10.1016/j.chembiol.2008.07.012.
37. McCarty, R.M.; Bandarian, V. Deciphering deazapurine biosynthesis: pathway for pyrrolopyrimidine nucleosides toyocamycin and sangivamycin. *Chem. Biol.* **2008**, *15*, 790–798, doi:10.1016/j.chembiol.2008.07.012.
38. Knolker, H.-J. *The Alkaloids*; Academic Press, 2017; ISBN 0-12-812198-X.
39. Kirby, G.W. Biosynthesis of the morphine alkaloids. *Science* **1967**, *155*, 170–173, doi:10.1126/science.155.3759.170.
40. Subramani, R.; Sipkema, D. Marine Rare Actinomycetes: A Promising Source of Structurally Diverse and Unique Novel Natural Products. *Marine Drugs* **2019**, *17*, doi:10.3390/md17050249.
41. Shamma, M. *The isoquinoline alkaloids chemistry and pharmacology*; Elsevier, 2012; Vol. 25; ISBN 0-323-14450-0.
42. Phillipson, J.D.; Roberts, M.F.; Zenk, M. *The chemistry and biology of isoquinoline alkaloids*; Springer Science & Business Media, 2012; ISBN 3-642-70128-0.

43. Hagel, J.M.; Facchini, P.J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* **2013**, *54*, 647–672, doi:10.1093/pcp/pct020.
44. Dutta, S.; Whicher, J.R.; Hansen, D.A.; Hale, W.A.; Chemler, J.A.; Congdon, G.R.; Narayan, A.R.H.; Håkansson, K.; Sherman, D.H.; Smith, J.L.; et al. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512–517, doi:10.1038/nature13423.
45. Dutta, S.; Whicher, J.R.; Hansen, D.A.; Hale, W.A.; Chemler, J.A.; Congdon, G.R.; Narayan, A.R.H.; Håkansson, K.; Sherman, D.H.; Smith, J.L.; et al. Structure of a modular polyketide synthase. *Nature* **2014**, *510*, 512–517, doi:10.1038/nature13423.
46. Theobald, S.; Vesth, T.C.; Andersen, M.R. Genus level analysis of PKS-NRPS and NRPS-PKS hybrids reveals their origin in Aspergilli. *BMC Genomics* **2019**, *20*, 1–12, doi:10.1186/s12864-019-6114-2.
47. Theobald, S.; Vesth, T.C.; Andersen, M.R. Genus level analysis of PKS-NRPS and NRPS-PKS hybrids reveals their origin in Aspergilli. *BMC Genomics* **2019**, *20*, 1–12, doi:10.1186/s12864-019-6114-2.
48. Miyanaga, A.; Kudo, F.; Eguchi, T. Protein–protein interactions in polyketide synthase–nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* **2018**, *35*, 1185–1209, doi:10.1039/C8NP00022K.
49. Fischbach, M.A.; Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106*, 3468–3496, doi:10.1021/cr0503097.
50. Lagunin, A.; Filimonov, D.; Poroikov, V. Multi-targeted natural products evaluation based on biological activity prediction with PASS. *Curr. Pharm. Des.* **2010**, *16*, 1703–1717, doi:10.2174/138161210791164063.
51. Du, L.; Lou, L. PKS and NRPS release mechanisms. *Natural Product Reports* **2010**, *27*, 255–278, doi:10.1039/B912037H.
52. Walsh, C.T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat Prod Rep* **2016**, *33*, 127–135, doi:10.1039/c5np00035a.
53. Du, L.; Lou, L. PKS and NRPS release mechanisms. *Natural Product Reports* **2010**, *27*, 255–278, doi:10.1039/B912037H.
54. Awakawa, T.; Fujioka, T.; Zhang, L.; Hoshino, S.; Hu, Z.; Hashimoto, J.; Kozone, I.; Ikeda, H.; Shin-Ya, K.; Liu, W.; et al. Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution. *Nat Commun* **2018**, *9*, 3534, doi:10.1038/s41467-018-05877-z.
55. Maria Fisch, K. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Advances* **2013**, *3*, 18228–18247, doi:10.1039/C3RA42661K.
56. Gomez-Escribano, J.P.; Bibb, M.J. Chapter Fourteen - Streptomyces coelicolor as an Expression Host for Heterologous Gene Clusters. In *Methods in Enzymology*; Hopwood, D.A.,

Ed.; Natural Product Biosynthesis by Microorganisms and Plants, Part C; Academic Press, 2012; Vol. 517, pp. 279–300.

57. Gomez-Escribano, J.P.; Bibb, M.J. Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor : from genome mining to manipulation of biosynthetic pathways. *J Ind Microbiol Biotechnol* **2014**, *41*, 425–431, doi:10.1007/s10295-013-1348-5.
58. Murray, A.W.; Szostak, J.W. Construction of artificial chromosomes in yeast. *Nature* **1983**, *305*, 189–193, doi:10.1038/305189a0.
59. O'Connor, M.; Peifer, M.; Bender, W. Construction of large DNA segments in Escherichia coli. *Science* **1989**, *244*, 1307–1312, doi:10.1126/science.2660262.
60. Shizuya, H.; Birren, B.; Kim, U.J.; Mancino, V.; Slepak, T.; Tachiiri, Y.; Simon, M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *PNAS* **1992**, *89*, 8794–8797, doi:10.1073/pnas.89.18.8794.
61. Sosio, M.; Giusino, F.; Cappellano, C.; Bossi, E.; Puglia, A.M.; Donadio, S. Artificial chromosomes for antibiotic-producing actinomycetes. *Nat Biotechnol* **2000**, *18*, 343–345, doi:10.1038/73810.
62. Martinez, A.; Kolvek, S.J.; Yip, C.L.T.; Hopke, J.; Brown, K.A.; MacNeil, I.A.; Osburne, M.S. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Appl. Environ. Microbiol.* **2004**, *70*, 2452–2463, doi:10.1128/AEM.70.4.2452-2463.2004.
63. Nah, H.-J.; Pyeon, H.-R.; Kang, S.-H.; Choi, S.-S.; Kim, E.-S. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Frontiers in Microbiology* **2017**, *8*, doi:10.3389/fmicb.2017.00394.
64. Baltz, R.H. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *J Ind Microbiol Biotechnol* **2010**, *37*, 759–772, doi:10.1007/s10295-010-0730-9.
65. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324, doi:10.1016/j.ymben.2018.09.004.
66. Hawas, U.W.; Shaaban, M.; Shaaban, K.A.; Speitling, M.; Maier, A.; Kelter, G.; Fiebig, H.H.; Meiners, M.; Helmke, E.; Laatsch, H. Mansouramycins A−D, Cytotoxic Isoquinolinequinones from a Marine Streptomycete(1). *J. Nat. Prod.* **2009**, *72*, 2120–2124, doi:10.1021/np900160g.

67. 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.

Chapter 2: Insights into the biosynthesis of isoquinoline quinones alkaloids mansouramycins in *Streptomyces albus* J1074

1 Introduction

Mansouramycins are a family of bioactive isoquinoline quinones originally isolated from the ethyl acetate extracts of the strain *Streptomyces* sp. Mei37 [1]. The family consists of five members: mansouramycins A-D and the 3-methyl-7-(methylamino)-5,8-isoquinolinequinone [1]. From the structural point of view, all mansouramycins share the same 7-(methylamino)-5,8-isoquinolinequinone core with different substituents at positions 3, 4, and 6 of the isoquinolinequinone moiety (Figure 2.1). Mansouramycin A harbors two methyl groups at the positions3 and 4; mansouramycin B has methyl group and chlorine substitutions at the positions 3 and 6, mansouramycin C – methoxycarbonyl group at the position 3 and mansouramycin D – indole group at the position 3 [1].

The biological activity assays with the purified Mansouramycins revealed their broad biological activity. In the agar diffusion tests, mansouramycins showed moderate antimicrobial activity. Mansouramycins inhibit the growth of the gram-positive bacteria *Staphylococcus aureus* and *Bacillus subtilis* and of the gram-negative bacterium *Escherichia coli*. Mansouramycins also exhibit growth inhibition of the green algae *Chlorella vulgaris*, *Chlorella sorokiniana*, and *Scenedesmus subspicatus*. High cytotoxicity of mansouramycins was also demonstrated in monolayer cell proliferation assay with 36 human tumor cell lines, comprising 14 different solid tumor types [1]. This activity is in accordance with previously published results on the activity of isoquinoline quinones against murine lymphoblastoma L5178Y cells [2].

Since the discovery of mansouramycins in 2009, the biosynthetic route and the biosynthetic genes leading to the production of these compounds remain unknown. The biosynthesis of plant benzylisoquinolines which are structurally related to mansouramycins was an object of numerous studies. Despite the great structural diversity of benzylisoquinoline alkaloids, they share the same biosynthetic origin. The central intermediate in the biosynthesis of benzylisoquinolines – (S)-norcoclaurine is formed through the condensation of two tyrosine derivatives, dopamine and 4-hydroxyphenylacetaldehyde [3]. Recently the biosynthetic pathway leading to the production of isoquinoline alkaloids, the fumisoquins, was discovered in fungus *Aspergillus fumigatus* [4]. Here a product of a small NRPS-like *fsqF* gene lacking a condensation domain is responsible for the formation of the isoquinoline core. FsqF catalyzes the condensation of the dehydroalanine unit, which is derived from serine, with tyrosine. The amino group of the tyrosine moiety within the resulting open-chain precursor is methylated through the action of S-adenosyl methionine-dependent N-methyltransferase encoded by *fsqC*. The formation of the isoquinoline ring is catalyzed by the product of fsqB, which is proposed to be a functional analog of plant berberine bridge enzyme [4,5].

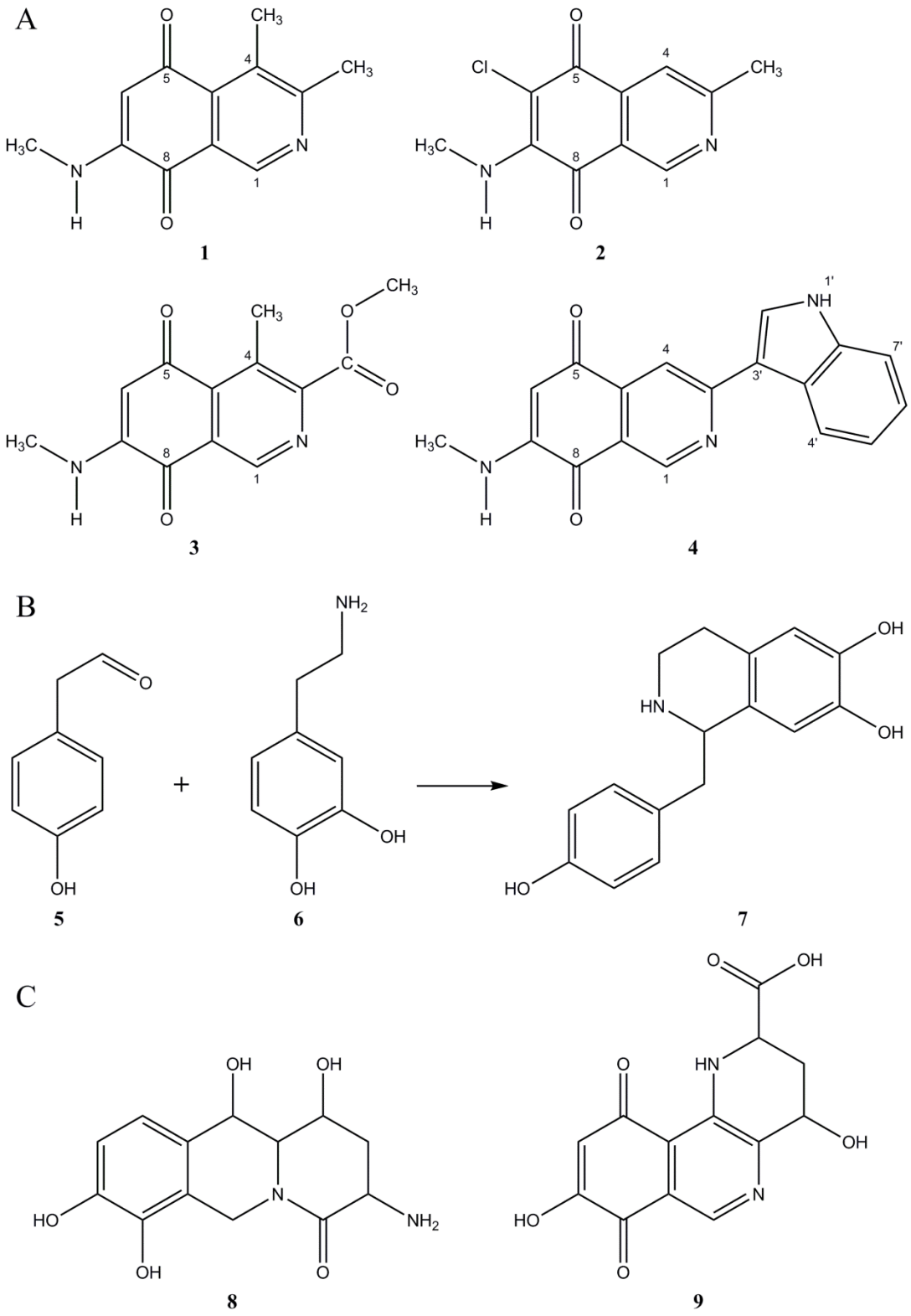

Figure 2.1. Structures of various isoquinoline alkaloids. A – Structures of mansouramycins A-D (1 – mansouramycin A, 2 – mansouramycin B, 3 – mansouramycin C, 4 – mansouramycin D). B – Formation of norcoclaurine (7) through condensation of 4-hydroxyphenylacetaldehyde (5) and dopamine (6). C – Structures of fungal isoquinolines – fumisoquin A and fumisoquin C.

Here we present the identification of the biosynthetic gene cluster leading to the production of mansouramycins in *Streptomyces albus* J1074. Through the series of gene inactivation, cocultivation, and feeding experiments, we demonstrate that in contrast to plants and fungi the isoquinoline core of mansouramycins is derived from tryptophan and that the biosynthetic pathway leading to mansouramycins is unique.

2 Materials and methods

General experimental procedures

All strains, plasmids and BACs used in this work are listed in Tables 2.1 and 2.2. *Escherichia coli* strains were cultured in LB medium [6]. Streptomyces strains were grown on soya flour mannitol agar (MS agar) [7] and in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA). For mansouramycin production, liquid DNPM medium [8] was used. The antibiotics kanamycin, apramycin, hygromycin, ampicillin and nalidixic acid were supplemented when required.

DNA manipulation

Isolation of DNA and all subsequent manipulations were performed according to standard protocols [6]. All of the primers used in this study are listed in Table 2.3.

Strain construction

For the chromosomal deletions in the genome of *S. albus*, a previously constructed genomic BAC library was used [9]. For the deletion of mansouramycin gene cluster, a BAC 3F18 was isolated from the genomic library of *S. albus*. The antibiotic cassette consisting of an apramycin resistance marker origin of conjugal transfer *oriT* and the P-GG and B-CC sites for phiC31 integrase was amplified from the plasmid patt-saac-oriT [9] by PCR with the primers Mans_delF and Mans_delR. The amplified fragment was used for RedET modification of the 3F18 BAC yielding 3F18_Am. In the constructed BAC the genes XNR_4432 – XNR_4443 were substituted with the apramycin resistance cassette. The constructed BAC 3F18_Am was transferred into *S. albus* Del14 [10] by conjugation [7] and the double crossover clones were selected using blue-white screen [11] (due to the presence of the *gus* gene in the BAC backbone). The antibiotic resistance marker and the *oriT* were excised from the chromosome of the constructed strain *S. albus* Del15_Am by expressing the *phi*C31 integrase. For this purpose, the plasmid pUWLint31 [9] containing the gene encoding phiC31 integrase was transferred into the *S. albus* Del15_Am strain by conjugation. The spores of the exconjugants were plated in dilutions on nonselective medium and the clones with the loss of both apramycin and thiostrepton markers were screened. The obtained strain was called *S. albus* Del15. Successful deletion of the mansouramycin gene cluster and excision of the apramycin resistance cassette was proved by PCR with the pair of primers Mans_delchF/Mans_delchR and sequencing.

Deletion of the tryptophan degradation pathway and of the genes encoding alpha and beta chains of tryptophan synthase was performed in a similar manner. Deletion of the tryptophan biosynthetic genes XNR_4841 and XNR_4842 was performed in the *S. albus* Del14 strain. The genes were deleted in the BAC 2B7 using antibiotic cassette amplified with the primers KO_bios_F

Table 2.1 Bacterial strains used in the study.

Strain	Description	Source
Streptomyces albus J1074	Cloning host; producer of mansouramycins	[12]
Streptomyces albus Del14	Cluster-free derivative of *S. albus* J1074	[10]
S. albus biosW	Derivative of *S. albus Del14* with the deletion of the tryptophan biosynthetic pathway	This study
S. albus Del15	Derivative of *S. albus Del14* with the deletion of mansouramycin biosynthetic pathway	This study
S. albus Del15 3F18_int	Derivative of *S. albus* Del15 harboring the BAC 3F18_int	This study
S. albus Del15 3F18_int_KO_0	Derivative of *S. albus* Del15 harboring the BAC 3F18_int_KO_0	This study
S. albus Del15 3F18_int_KO_1	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_1	This study
S. albus Del15 3F18_int_KO_2	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_2	This study
S. albus Del15 3F18_int_KO_3	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_3	This study
S. albus Del15 3F18_int_KO_4	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_4	This study
S. albus Del15 3F18_int_KO_5	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_5	This study
S. albus Del15 3F18_int_KO_6	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_6	This study
S. albus Del15 3F18_int_KO_7	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_7	This study
S. albus Del15 3F18_int_KO_8	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_8	This study
S. albus Del15 3F18_int_KO_9	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_9	This study
S. albus Del15 3F18_int_KO_10	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_10	This study
S. albus Del15 3F18_int_KO_11	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_11	This study
S. albus Del15 3F18_int_KO_12	Derivative of *S. albus Del15* harboring the BAC 3F18_int_KO_12	This study
S. albus degrW	Derivative of *S. albus Del15* with the deletion of tryptophan degradation pathway	This study
S. albus degrW 3F18_int_KO_0	Derivative of *S. albus* degrW with 3F18_int_KO_0	This study
S. albus degrW 3F18_int_KO_1	Derivative of *S. albus* degrW with 3F18_int_KO_1	This study
S. albus degrW 3F18_int_KO_2	Derivative of *S. albus* degrW with 3F18_int_KO_2	This study
S. albus degrW 3F18_int_KO_4	Derivative of *S. albus* degrW with 3F18_int_KO_4	This study
S. albus degrW 3F18_int_KO_10	Derivative of *S. albus* degrW with 3F18_int_KO_10	This study
S. albus degrW 3F18_int_KO_12	Derivative of *S. albus* degrW with 3F18_int_KO_12	This study
LV1-652	*Streptomyces* sp. strain	LV strain collection
LV1-4	*Streptomyces* sp. strain	LV strain collection
LV1-22	*Streptomyces* sp. strain	LV strain collection
LV1-114	*Streptomyces* sp. strain	LV strain collection
LV1-18.2	*Streptomyces* sp. strain	LV strain collection
LV1-208	*Streptomyces* sp. strain	LV strain collection
LV1-209	*Streptomyces* sp. strain	LV strain collection
LV1-213	*Streptomyces* sp. strain	LV strain collection
LV1-652 3F18_int_KO_0	Derivative of LV1-652 harboring 3F18_int_KO_0	This study
LV1-4 3F18_int_KO_0	Derivative of LV1-4 harboring 3F18_int_KO_0	This study
LV1-22 3F18_int_KO_0	Derivative of LV1-22 harboring 3F18_int_KO_0	This study
LV1-114 3F18_int_KO_0	Derivative of LV1-114 harboring 3F18_int_KO_0	This study
LV1-18.2 3F18_int_KO_0	Derivative of LV1-18.2 harboring 3F18_int_KO_0	This study
LV1-208 3F18_int_KO_0	Derivative of LV1-208 harboring 3F18_int_KO_0	This study
LV1-209 3F18_int_KO_0	Derivative of LV1-209 harboring 3F18_int_KO_0	This study
LV1-213 3F18_int_KO_0	Derivative of LV1-213 harboring 3F18_int_KO_0	This study
Escherichia coli GD10B	General coning host	[13]
Escherichia coli ET12567 pUB307	Donor strain for intergeneric conjugation	[14]
Escherichia coli Rosetta 2 DE3	Protein expression host	Novagen

Table 2.2 Plasmids and BACs used in the study.

Name	Description	Source
3F18	BAC; contains genomic fragment of *S. albus* J1074 chromosome with the mansouramycin cluster	[9]
3F18_Am	Derivative of 3F18; contains deletion of mansouramycin gene cluster	This study
3F18_int	Derivative of 3F18 with the phiC31 integration system, apramycin resistance marker and origin of conjugal transfer	This study
3F18_int_KO_0	Derivative of 3F18_int with the deletion of the genes not involved in mansouramycin biosynthesis	This study
3F18_int_KO_1	Derivative of 3F18_int_KO_0 with the deletion of gene 1	This study
3F18_int_KO_2	Derivative of 3F18_int_KO_0 with the deletion of gene 2	This study
3F18_int_KO_3	Derivative of 3F18_int_KO_0 with the deletion of gene 3	This study
3F18_int_KO_4	Derivative of 3F18_int_KO_0 with the deletion of gene 4	This study
3F18_int_KO_5	Derivative of 3F18_int_KO_0 with the deletion of gene 5	This study
3F18_int_KO_6	Derivative of 3F18_int_KO_0 with the deletion of gene 6	This study
3F18_int_KO_7	Derivative of 3F18_int_KO_0 with the deletion of gene 7	This study
3F18_int_KO_8	Derivative of 3F18_int_KO_0 with the deletion of gene 8	This study
3F18_int_KO_9	Derivative of 3F18_int_KO_0 with the deletion of gene 9	This study
3F18_int_KO_10	Derivative of 3F18_int_KO_0 with the deletion of gene 10	This study
3F18_int_KO_11	Derivative of 3F18_int_KO_0 with the deletion of gene 11	This study
3F18_int_KO_12	Derivative of 3F18_int_KO_0 with the deletion of gene 12	This study
2B7	BAC; contains the genes XNR_4841 and XNR_4842 encoding alpha and beta chains of tryptophan synthase	[9]
2B7_aac	Derivative of 2B7 with the deletion of the genes XNR_4841 and XNR_4842	This study
2H13	BAC; contains the genes XNR_3200, XNR_3201 and XNR_3202 encoding tryptophan degradation pathway	[9]
2H13_aac	Derivative of with the deletion of the genes XNR_3200, XNR_3201, XNR_3202	This study
patt-saac-oriT	Plasmid; contains apramycin resistance cassette for gene inactivation	[9]
pUWLint31	Plasmid; contains *phi*C31 integrase gene and thiostrepton resistance marker	[9]
pUCcatint	Plasmid; contains apramycin resistance gene, *ori*T, *phi*C31 integrase and *att*P	[10]
pRT801	Plasmid; BT1 integrative vector for Streptomyces strains; apramycin resistance	[15]
pRT801_cat	Derivative of pRT801 with promoter TS81 and chloramphenicol resistance gene	This study
pRT801_cat_ampery	Derivative of the BT1 integrative vector pRT801. Apramycin resistance gene was substituted with ampicillin-erythromycin resistance cassette. Strong synthetic promoter TS81 was inserted in the polylinker as a part of chloramphenicol resistance cassette	This study
pRT801_proM3	Derivative of pRT801_cat_ampery; contains the gene 3 under control of TS81	This study
pRT801_proM4	Derivative of pRT801_cat_ampery; contains the gene 4 under control of TS81	This study
pRT801_proM5	Derivative of pRT801_cat_ampery; contains the gene 5 under control of TS81	This study
pRT801_proM6	Derivative of pRT801_cat_ampery; contains the gene 6 under control of TS81	This study
pRT801_proM7	Derivative of pRT801_cat_ampery; contains the gene 7 under control of TS81	This study
pRT801_proM8	Derivative of pRT801_cat_ampery; contains the gene 8 under control of TS81	This study
pRT801_proM9	Derivative of pRT801_cat_ampery; contains the gene 9 under control of TS81	This study
pET-28 a (+)	Plasmid; protein expression vector	Sigma-Aldrich
pET-28-8	Derivative pET-28 a (+) with the gene 8 cloned	This study

Table 2.3 Primers used in this study.

Primer	Sequence
Mans_delF	GCCGTGGGTGACGATCAGCCGGTCGACCGGGCGGAAGCGGCCGAGCAGGCTTCCGGGGATCCGTCGACCC
Mans_delR	CCGGGGAGACCCCCCGGGCGAAAGGTTGCAGTGCGCTCAGGGGGATGTTCTGTAGGCTGGAGCTGCTTCG
KO_0F	GGGGCCGGCCACTCCCAGCGCCACTCCGGCCCGGTCCCGTACGCGTTCATGCGATCGCCGTCAGGTGGCACTTTTCG
KO_0R	GTACGTTACTCCCGAATCCGGCGCAATCCCTGGATACCCCCCCCGCGCCCGCGATCGCTTACCAATGCTTAATCAGTG
KO_1F	GTGATCAACGCTGTCACCAGACGCTATGTCGACCGGTGCGCCGAGGCGGGTTTAAACCGTCAGGTGGCACTTTTCG
KO_1R	TCAGTCGGCGGTGAAGACGCAGCTGCTGCGCGTGCCCGGGTTGGAGAAGTTTAAACTTACCAATGCTTAATCAGTG
KO_2F	ATGTCCGAGCACGCCCCCGACGAGGAAGACCGCCCCCACATAGTCGTCAGTTTAAACCG

	TCAGGTGGCACTTTTCG
KO_2R	TCATCCGGTCACCTCCACGGGTGCGGTGTGCGCGGCGGAGAGCGCGGTGTTTAAACTTA CCAATGCTTAATCAGTG
KO_3F	ATGAAGGACTACGCCGAACAGATCCTCGACGGCGAGGGCGAGAACGACTGTTTAAACC GTCAGGTGGCACTTTTCG
KO_3R	TCACAGTCCCTCCCGTACCGGCTGGTCGCCCATCGGTCCGGTGGCGGTGTTTAAACTTAC CAATGCTTAATCAGTG
KO_4F	GTGAGCACCCTCGCCGCCGGGCCGCGCCCCTCCGGCACGCTGAGCGCCGGTTTAAACCG TCAGGTGGCACTTTTCG
KO_4R	TCAGCGGGACCTGTACTCACGCAGCGCCTTACAGAGCGAGGCGATGTCGTTTAAACTTA CCAATGCTTAATCAGTG
KO_5F	ATGTTCACAGCCGAACTCGCCTCCTGGGCCCGCGAACCGGACCCCGGAAGTTTAAACCG TCAGGTGGCACTTTTCG
KO_5R	TCATCGTGTCCTCACGGTCCGGTCGGCCGCGGGCGCGGCGTGGTGGAAGTTTAAACTTA CCAATGCTTAATCAGTG
KO_6F	ATGACCGCAGTGGACGCACCCGTGGGGATCGTCGGAGCCGGGCCCGTCGGTTTAAACC GTCAGGTGGCACTTTTCG
KO_6R	TCATGAATGGGTTCCTCTCAGCAGGGGGGCGGCCACGCGGGCGGCGTCGTTTAAACTTA CCAATGCTTAATCAGTG
KO_7F	ATGAGCGCCCAACCACCCCCGTACGCATGGCTCAACGGCGGAGTCGTCCGTTTAAACCG TCAGGTGGCACTTTTCG
KO_7R	TCATGCGGCACCGCCCGTCCGGGCCGGGACCGGGGTGCACCACTCGGGGTTTAAACTTA CCAATGCTTAATCAGTG
KO_8F	ATGAGCCTGCTGACCGACCTGGTGACGGCGGTGCGCACCGGCGGCGTCCGTTTAAACCG TCAGGTGGCACTTTTCG
KO_8R	TCAGGCTCCTTCGGGGGCGGTCTCGACCAGGGCGAGCACCCGGCTGGGGTTTAAACTTA CCAATGCTTAATCAGTG
KO_9F	ATGCGGGACAGCACGGGTTTCGACCGCGCCAGCGACCTGCTGCGGCGGCGTTTAAACCG TCAGGTGGCACTTTTCG
KO_9R	TCACTCCGCCGTGAGCTCCCACATCTGAAACTCGATCACCATCGTGTCGTTTAAACTTAC CAATGCTTAATCAGTG
KO_10F	GTGCCCCAGAACGACGACGCCCGCGACCGCACCCTCTCCCTCCTGGACGGTTTAAACCG TCAGGTGGCACTTTTCG
KO_10R	TCAGGCGCCGGGGACCCAGCCGTAGCTGTGGTCGATGGCGAACCGGACGTTTAAACTTA CCAATGCTTAATCAGTG
KO_11F	ATGCCCCGCCTGCACGTCGTCATAGCCTCCACCCGCCCCGGCCGCACCGGTTTAAACCG TCAGGTGGCACTTTTCG
KO_11R	TCAGACTGCGGGCTCGCGCAGCGGGCGCAGCGCCTCGGCGAACCGGGCGTTTAAACTTA CCAATGCTTAATCAGTG
KO_12F	GTGCAGCAAGCCCACATCATCCTGGTCGGAAGTGGCGGACAGCCCTACCGTTTAAACCG TCAGGTGGCACTTTTCG
KO_12R	TCATCGGCCGGGCTCCCCGGTGAGCGGGACGACCTCGCTGGTGACCTCGTTTAAACTTA CCAATGCTTAATCAGTG
KO_degr_F	ATGTCGCACGACCTCGGTGCCACGGAAGCCGGTTCGGACACCCCGAATCTTGTAGGCTG GAGCTGCTTCG
KO_degr_R	TCACCAGGCCAGTTGGGCTATCTCCTCAGCGGCCAGGGCGCAGGCGGCCGTTCCGGGGA TCCGTCGACCC
KO_bios_F	ATGCCCAGCGAGTACTTCCTTCCCGACCCCGAGGGGCTCAGCCCCGACGCTGTAGGCTG GAGCTGCTTCG
KO_bios_R	TCAGCCGGTGCGGCGGACGCCTTCGGCCAGCTCGGCGGCGAGCGCGCGGATTCCGGGG ATCCGTCGACCC
Mans_delchF	AATGGCCAAAGTGCAAATAC
Mans_delchR	GACGAAGAAGTACGGGATG
KO_0chF	AAGACGTGGTCGAAGTGGG
KO_0chR	GATCTCCCGGCGGAAAGAC
KO_1chF	GCTCCCGAACGATATCCCC
KO_1chR	CATCAGCGGTCCCTTCAGG
KO_2chF	CGCCGTCATCTCCTTCTCC
KO_2chR	CGACTGGTGGACGATCTGG
KO_3chF	GACGCCTACTACGAGCACG
KO_3chR	CATCGCCTCCAGCATCTCG
KO_4chF	CATCATCGGCCACCAGTCC
KO_4chR	GGGTGAGTTCCTGGAGCAG
KO_5chF	CCACATCGGTCTGCTCGAC
KO_5chR	TGAGGCAGGCTTTGGAACC

KO_6chF	GGCACCCACTACCGGCTCG
KO_6chR	GTCCTGGACGCGGAAGACC
KO_7chF	CCGGTGTGGTGGCTGGTCC
KO_7chR	GCGTCGTAGTGGCTGATCA
KO_8chF	CACCATGTCCGAGCTGCTG
KO_8chR	CAGCCAGGTGGTGAACAGG
KO_9chF	GCCACTGGTACTTCCACGG
KO_9chR	GCCGTCCTTCTTCAGGGTG
KO_10chF	AATCGGGACAGAAGCTCACG
KO_10chR	GATGCCGGTGAACCAGTCG
KO_11chF	CACGGTCATGGTCGAGGAC
KO_11chR	CTGGAGGTGGGGCTGTTGC
KO_12chF	TCGCGATCCACTTCATCGG
KO_12chR	GCCGTAGAAGAGCGCGACG
KO_degr_chF	GATGATGCGAGGTGAGCCC
KO_degr_chR	CCACTCCTTCGCTCAACCC
KO_bios_chF	AGGTCAACGGCATGCGATG
KO_bios_chR	ACGGTTCGTCTCACTCACG
8_petF	AAAACTAGCAGCCTGCTGACCGACCTGGT
8_petR	AAAACTCGAGTCAGGCTCCTTCGGGGGCGG
Cat_promF	AAAAGGATCCCAGGTCGGCTGGTTGGCTGACAAAGTGCGTAGGATGACTGTGAGCTCC ATATGTGATCGGCACGTAAGAGGTT
Cat_promR	AAAAGATATCTTACGCCCCGCCCTGCCACT
AmperyF	AGCAAAAGGGGATGATAAGTTTATCACCACCGACTATTTGCAACAGTGCCTGTAGGCTG GAGCTGCTTCG
AmperyR	AGCAATCAGCGCGACCTTGCCCCTCCAACGTCATCTCGTTCTCCGCTCATCTACCGCTGC CCGGGTCCGC
proM3F	AAAAGAGCTCCGCGCACACCGCACCCGTGG
proM3R	AAAAGATATCTCACAGTCCCTCCCGTACCG
proM4F	AAAAGAGCTCGATGGGCGACCAGCCGGTAC
proM4R	AAAAGATATCTCAGCGGGACCTGTACTCAC
proM5F	AAAAGAGCTCCACTCCCCACCCACACCCGT
proM5R	AAAAGATATCTCATCGTGTCCTCACGGTCC
proM6F	AAAAGAGCTCGCCCGCGGCCGACCGGACCG
proM6R	AAAAGATATCTCATGAATGGGTTCCTCTCA
proM7F	AAAAGAGCTCGGCCGCCCCCCTGCTGAGAG
proM7R	AAAAGATATCTCATGCGGCACCGCCCGTCC
proM8F	AAAAGAGCTCCCCGGTCCCGGCCCGGACGG
proM8R	AAAAGATATCTCAGGCTCCTTCGGGGGCGG
proM9F	AAAACATATGGGTCGAGACCGCCCCCGAAG
proM9R	AAAAGATATCTCACTCCGCCGTGAGCTCCC

and KO_bios_R. The constructed BAC 2B7_aac was used for the deletion of the genes in the chromosome of *S. albus* Del14. The constructed tryptophan auxotrophic strain was called *S. albus* biosW. The successful deletion and marker excision was proved by PCR with the pair of primers KO_bios_chF/KO_bios_chR and sequencing.

Deletion of tryptophan degradation pathway was performed in the *S. albus* Del15 strain. The antibiotic resistance cassette was amplified with the pair of primers KO_degr_F/KO_degr_R and used for substitution of the genes XNR_3200, XNR_3201 and XNR_3202 in the BAC 2H13. The resulting BAC 2H13_aac has been successfully used for the deletion of the respective genes in the chromosome of *S. albus* Del15 strain. The constructed strain was called *S. albus* degrW. The successful deletion and marker excision was proved by PCR with the pair of primers KO_degr_chF/KO_degr_chR and sequencing.

For mansouramycin biosynthetic studies the construction of BAC variants harboring mansouramycin cluster with single genes been inactivated was necessary. The original BAC 3F18

does not contain antibiotic markers for selection in streptomycetes and also cannot be propagated in streptomycetes. A cassette with apramycin resistance gene *ori*T sequence, *phi*C31 integrase gene and *phi*C31 *att*P sequence was cut from the plasmid pUCcatint [10] and introduced into the backbone of 3F18 by RedET. The constructed BAC was named 3F18_int.

The chromosomal fragment cloned in 3F18 contains DNA regions not involved in mansouramycin biosynthesis. In order to minimize the size of the BAC, the genes not involved in mansouramycin biosynthesis were deleted in 3F18_int. For this purpose, the ampicillin resistance marker from the plasmid pUC19 was amplified with the primers KO_0F and KO_0R. The obtained PCR product was used for RedET modification of 3F18_int yielding 3F18_int_KO_0_Ap. The resistance marker was removed by cutting with SfaAI and self-ligating the BAC. The final construct was named 3F18_int_KO_0. The construction of the BAC was proved by PCR with the pair of primers KO_0chF/KO_0chR and sequencing. The constructed BAC 3F18_int_KO_0 contains the entire mansouramycin biosynthetic cluster and contains genetic traits for transfer and propagation in streptomyces strains.

Based on 3F18_int_KO_0 its variants with the inactivated genes 1 – 12 were constructed. For the inactivation of the genes 1 – 12 the ampicillin resistance gene was amplified from the pUC19 plasmid with the pairs of primers KO_1chF/KO_1chR, KO_2chF/KO_2chR, KO_3chF/KO_3chR, KO_4chF/KO_4chR, KO_5chF/KO_5chR, KO_6chF/KO_6chR, KO_7chF/KO_7chR, KO_8chF/KO_8chR, KO_9chF/KO_9chR, KO_10chF/KO_10chR, KO_11chF/KO_11chR and KO_12chF/KO_12chR respectively. The PCR products were used for RedET modification of 3F18_int_KO_0 yielding 3F18_int_KO_1_Ap, 3F18_int_KO_2_Ap, 3F18_int_KO_3_Ap, 3F18_int_KO_4_Ap, 3F18_int_KO_5_Ap, 3F18_int_KO_6_Ap, 3F18_int_KO_7-Ap, 3F18_int_KO_8_Ap, 3F18_int_KO_9_Ap, 3F18_int_KO_10_Ap, 3F18_int_KO_11_Ap and 3F18_int_KO_12_Ap. The ampicillin marker from the obtained BACs was removed by cutting them with MssI with subsequent self-ligation. The final constructs were named 3F18_int_KO_1, 3F18_int_KO_2, 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8, 3F18_int_KO_9, 3F18_int_KO_10, 3F18_int_KO_11 and 3F18_int_KO_12. Successful RedET modification and marker excision were proved by PCR and sequencing using the following pairs of primers: KO_1chF/KO_1chR, KO_2chF/KO_2chR, KO_3chF/KO_3chR, KO_4chF/KO_4chR, KO_5chF/KO_5chR, KO_6chF/KO_6chR, KO_7chF/KO_7chR, KO_8chF/KO_8chR, KO_9chF/KO_9chR, KO_10chF/KO_10chR, KO_11chF/KO_11chR and KO_12chF/KO_12chR.

The constructed BACs 3F18_int_KO_0, 3F18_int_KO_1, 3F18_int_KO_2, 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8, 3F18_int_KO_9, 3F18_int_KO_10, 3F18_int_KO_11 and 3F18_int_KO_12 have been transferred

into the strains *S. albus* Del15, *S. albus* degrW, LV1-652, LV1-4, LV1-22, LV1-114, LV1-18.2, LV1-208, LV1-209 and LV1-213 by intergeneric conjugation [7].

For complementation of gene inactivation, a series of plasmids with individual mansouramycin genes set under the control of strong promoter were constructed. The constructs were created based on the BT1 integrative vector pRT801 [15]. To clone the strong synthetic promoter TS81 the chloramphenicol resistance gene from the backbone of the BAC 3F18 was amplified with the primers Cat_promF and Cat_promR. The promoter sequence was inserted into the primer Cat_promF. The amplified fragment was cut with BamHI/EcoRV and cloned into the respective sites of pRT801 yielding pRT801_cat. The pRT801 vector contains apramycin resistance gene which is used as a selection marker in all BAC variants containing mansouramycin cluster with inactivated genes. The apramycin resistance gene in the backbone of pRT801_cat was substituted with the resistance cassette harboring ampicillin and erythromycin resistance genes [16]. The cassette was amplified from the plasmid pAmpery [16] using the primers AmperyF and AmperyR. The amplified fragment was used for RedET modification of pRT801_cat yielding pRT801_cat_ampery. The obtained plasmid was used as a vector for cloning of the genes 3 – 9 for complementation studies.

To complement the inactivation of the genes 3 – 9 they were amplified from the BAC 3F18 using the pairs of primers proM3F/proM3R, proM4F/proM4R, proM5F/proM5R, proM6F/proM6R, proM7F/proM7R, proM8F/proM8R, proM9F/proM9R. All amplified fragments, with the exception of the fragment amplified with the primers proM9F/proM9R, were digested with SacI/EcoRV and cloned into respective sites of pRT801_cat_ampery yielding pRT801_proM3, pRT801_proM4, pRT801_proM5, pRT801_proM6, pRT801_proM7 and pRT801_proM8. The fragment amplified with the primers proM9F/proM9R was digested with NdeI/EcoRV and cloned into the respective sites of pRT801_cat_ampery yielding pRT801_proM9. For complementation purposes the constructed plasmids were transferred by conjugation into the *S. albus* degrW strain harboring following BACs 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8 and 3F18_int_KO_9 respectively.

For the overexpression of the product of the gene 8 it was PCR amplified from the BAC 3F18 with the primers 8_petF and 8_petR. The amplified fragment was digested with BfaI/XhoI and cloned into pET-28 a (+) digested with NdeI/XhoI. The obtained plasmid was named pET-28-8. The plasmid was transferred into *Escherichia coli* Rosetta 2 DE3 by transformation.

Feeding studies

6-hydroxyindoline-2,3-dione was purchased from Enamine. L-tryptophan ^{13}C11 and anthranilic acid-^{15}N were purchased from Euroisotope. L-tryptophan-1-^{13}C, DL-tryptophan-2-^{13}C, L-tryptophan-^{15}N2 and anthranilic acid-(phenyl-^{13}C6) were purchased from Sigma-Aldrich.

For feeding purposes, tryptophan with different labeling patterns, labeled anthranilic acid, putative mansouramycin biosynthetic intermediates, etc., were dissolved in water at concentration 5 mg/mL. The seed culture of *S. albus* strains was prepared by inoculating fresh spore material into 25 ml of DNPM medium in 250 mL baffled flask and cultivating in an Infors Multitron shaker at 28°C 180rpm for 24 hours. 1 ml of the seed culture was transferred into 25 ml of DNPM medium in 250 mL baffled flask and cultivated for 96 hours under the same conditions. The compounds were fed to the cultures by adding 125 µL of the stock solution at the time of inoculation, after 24, 48 and 72 hours of cultivation. After 96h hours of cultivation, the mycelial part was separated from the liquid fraction by centrifugation. The supernatant was extracted with the equal amount of ethyl acetate. The solvent was dried under nitrogen flow and the dry rest was dissolved in 1 mL of methanol. The obtained extract was used for the LC-MS analysis.

Analysis of secondary metabolite production

For mass determination, a Bruker Amazon Speed and a Thermo LTQ Orbitrap XL mass spectrometer were used. Both machines were coupled to UPLC Thermo Dionex Ultimate3000 RS. Analytes were separated on a Waters ACQUITY BEH C18 Column (1.7 µm, 2.1mm x 100 mm) with water +0.1% formic acid and acetonitrile +0.1% formic acid as the mobile phase. Purification of compounds was performed in three steps. The crude extract was fractionated by normal phase chromatography on a prepacked silica cartridge (Biotage, Uppsala, Sweden) using hexane, dichloromethane, ethyl acetate, and methanol as the mobile phase. Fractions containing compounds to be purified were detected by LC-MS analysis. They were pooled together, evaporated, and dissolved in methanol. The prepurified compounds were further separated by size-exclusion chromatography on an LH 20 Sephadex column (Sigma-Aldrich, USA) using methanol as the solvent. The final purification step was performed using reversed-phase chromatography. Waters autopurification system equipped with Macherey Nagel Nucleodur HTec C18 column (5 µm, 21mm x 250 mm) was used. The mobile phase was blended from water +0.1% formic acid and methanol +0.1% formic acid.

Structural elucidation of compounds

NMR spectra were acquired on a Bruker Ascend 700 MHz NMR spectrometer and on Bruker Avance 500 spectrometer. As solvent deuterated MeOD4 and DMSO-d6 were used. HSQC, HMBC, 1H-1H COSY spectra were recorded using standard pulse programs.

Protein expression and purification

For protein overexpression, the *E. coli* Rosetta 2 DE3 pET-28-8 strain was cultivated in 300mL LB medium supplemented with chloramphenicol 25µg/mL and kanamycin 50 µg/mL at

37°C. After the culture reached the OD_{600}=0.6 it was cooled down on ice, supplemented with IPTG at final concentration of 0.4mM and cultivated at 18°C for 18 hours. The cellular part was separated from the supernatant by centrifugation. The pellet (2.7 g) was resuspended in 19mL of the 20 mM sodium phosphate, 50mM sodium chloride buffer (pH7.5). 76 mg of lysozyme (Carl Roth) and 5 µL of benzonase solution (25-29 U/µl) (Millipore) was added into the cell suspension. The suspension was incubated at room temperature for 5 minutes. The cell debris was separated from the lysate by centrifugation and filtering through a filter with 0.2 µm pore size. The protein was purified using Äkta prime plus chromatography system and 1 mL HiTrap Talon Crude column (GE Healthcare). 20 mM sodium phosphate, 50mM sodium chloride (pH7.5) was used as a loading buffer. Protein was eluted with the buffer containing 20 mM sodium phosphate, 50mM sodium chloride and 150 mM imidazole (pH7.5). The fractions containing purified protein were pooled together. Imidazole was removed using size exclusion chromatography on HiPrep 26/10 Desalting column (GE Healthcare) and buffer containing 20 mM sodium phosphate and 50mM sodium chloride. The fractions containing the purified protein were pooled together and the volume was adjusted to 15 mL.

***In vitro* activity assay**

10 mg of 6-hydroxyindoline-2,3-dione was added into 15 mL protein solution as powder. The reaction mixture was incubated at 37 °C for 30 minutes and then frozen in liquid nitrogen. The reaction mixture was lyophilized and the dry rest was dissolved in methanol and dried under vacuum. The obtained material was used for structure elucidation studies.

3 Results and discussion

3.1 Detection of mansouramycin production in extracts of *Streptomyces albus*

Streptomyces albus J1074 is a widely-used host strain for heterologous expression of various secondary metabolite gene clusters. During the construction of a cluster-free derivative of *S. albus* J1074, *S. albus* Del14, a peak with a high-resolution mass corresponding to mansouramycin D was detected in the butanol extracts of the strain (Figure 2.2). At the time of this discovery production of mansouramycins was not reported for *S. albus* strain. Originally Mansouramycin D was discovered in the extracts of *Streptomyces* sp. Mei37 together with other members of the mansouramycin family: mansouramycin A, B, and C [1]. A more precise analysis of the *S. albus* Del14 extracts revealed the presence of the peak with high-resolution mass corresponding to mansouramycin A (Figure 2.2). Mansouramycins B and C could not be detected in the extracts of the strain. To prove

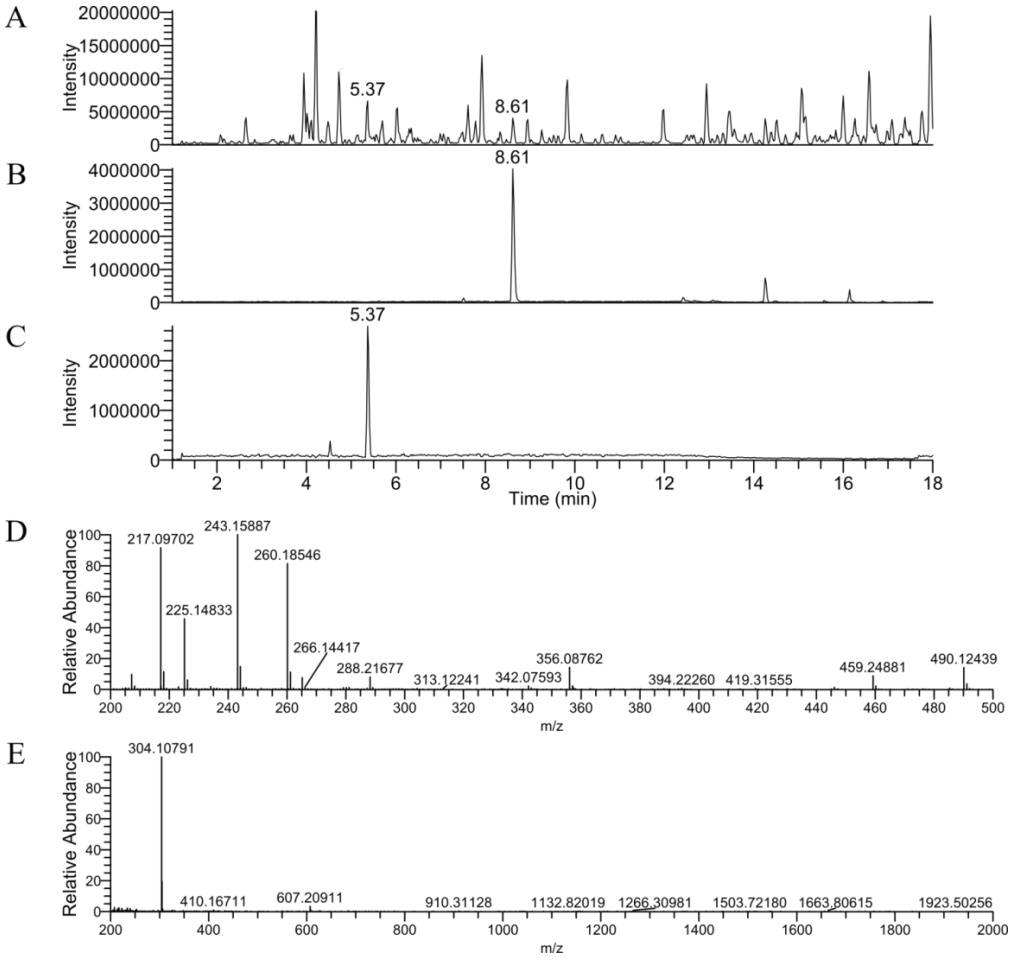

Figure 2.2. Identification of mansouramycins in ethyl acetate extract of *S. albus* Del14 using high-resolution HPLC-MS. A – Base peak chromatogram of ethyl acetate extract of *S. albus* Del14 culture cultivated in the DNPM medium. The peaks at 5.37 min and 8.61 min correspond to mansouramycins A and D respectively. B – Extracted ion chromatogram (304 ± 0.5 Da). The peak corresponding to mansouramycin D can be seen at 8.61 min. C – Extracted ion chromatogram (217 ± 0.5 Da). The peak corresponding to mansouramycin A can be seen at 5.37 min. D – the mass spectrum of the mansouramycin A peak. A molecular ion with m/z value of 217.09 Da corresponds to mansouramycin A. E – the mass spectrum of the mansouramycin D peak. A molecular ion with m/z value of 304.10 Da corresponds to mansouramycin D.

the production of mansouramycins by *S. albus* J1074 it was decided to purify the compound with the high-resolution mass corresponding to mansouramycin D for structure elucidation studies.

Concerning the small intensity of the peak and the overall complexity of the HPLC-MS profile of the *S. albus* J1074 strain, the purification of the compound was performed from the extracts of *S. albus* Del14 strain which is characterized by simplified secondary metabolite profile. For this purpose, the strain was cultivated in 10L of the DNPM medium and extracted with butanol. The putative mansouramycin D was purified from the extract using size exclusion and reverse-phase chromatography. The purified compound was used for structure elucidation purposes. NMR analysis was performed by Dr. Suvd Nadmid.

Analysis of 1H NMR and HSQC spectra of the compound exhibited protons corresponding to methyl doublet at δ 2.82 (δC 28.6) and the amino group as a broadened quartet at δ 7.88, suggesting a presence of N-methyl group. Furthermore, detailed analysis of olefinic protons observed at δ 5.71 (δC 99.4), 8.18 (δC 113.0) and 9.11 (δC 147.6) established 7-methylamino-isoquinolinequinone part of mansouramycin D. The rest of the signals corresponding to second broad NH signal at δ 12.02 which has a COSY correlation to a singlet at δ 8.53 along with four aromatic protons have built indole system. Altogether, the structure of mansouramycin D was concluded and was in concordance with previously reported NMR data [1] (Table 2.4).

The identity of the purified compound to mansouramycin D implies indirectly that the second peak, preliminary assigned during dereplication studies to mansouramycin A, with high probability corresponds to mansouramycin A.

Table 2.4. NMR data (500 MHz, DMSO-d6) for mansouramycin D

position	**δ_H, mult (*J*, Hz)**	**δ_C**	**HMBC**	**COSY**
1	9.11, d (*0.7*)	147.6	4, 8, 9, 10	
3		161.6		
4	8.18, d (*0.7*)	113.0	3, 3‘, 5, 9, 10	
5		179.7		
6	5.71, d (*0.4*)	99.4	5, 7, 8, 9	
7		149.9		
7-NH	7.88, q (*5.0*)		N-Me, 6, 8	N-Me
N-Me	2.82, d (*5.0*)	28.6	6, 7	7-NH
8		179.7		
9		139.7		
10		120.3		
1‘-NH	12.02, br s			2‘
2‘	8.53, br s	129.8	3, 3‘, 8‘, 9‘	1‘-NH
3‘		115.2		
4‘	8.52, dd (*1.9; 6.0*)	121.7	3‘, 5‘	5‘
5‘	7.22, m	122.2	7‘, 8‘	4‘, 6‘
6‘	7.20, m	120.8	7‘, 9‘	5‘, 7‘
7‘	7.48, m	111.9	6‘, 8‘, 9‘	6‘
8‘		137.8		
9‘		124.5		

3.2 Analysis of mansouramycin biosynthetic origins through feeding studies. L-tryptophan is a main precursor of mansouramycin D

Structurally mansouramycin D belongs to the family of isoquinoline quinones. The isoquinoline alkaloids of plant origin are synthesized from two derivatives of amino acid tyrosine –

dopamine and 4-hydroxyphenylacetaldehyde. The fungal isoquinoline alkaloids, the fumisoquins, are also derived from tyrosine and additionally from dehydroalanine. Considering possible precursor role of tyrosine in the biosynthesis of mansouramycin D, C^{13}-labelled L tyrosine (L-Tyrosine-^{13}C9) was fed to the culture of *S. albus* Del14. The fed culture and the control, non-fed culture of *S. albus* Del14 were extracted with butanol. No incorporation of labeled L-Tyrosine into mansouramycin A and D could be detected in the course of HPLC-MS analysis of the obtained extracts. The presence of the indole moiety in the structure of mansouramycin D implies the possible precursor role of tryptophan in its biosynthesis. In order to prove this C^{13}-labelled L-tryptophan (^{13}C11) was fed to the culture of *S. albus*. HPLC-MS analysis of the extracts from the fed and non-fed cultures *S. albus* cultures revealed the incorporation of the labelled L-tryptophan into the structure of mansouramycin D (Figure 2.2 A and B). In the extract of the non-fed *S. albus* culture, mansouramycin D can be detected as $[M + H]^+$ ion with m/z value of 304. After feeding with L-tryptophan (^{13}C11) additional ions could be observed in the mass spectrum of

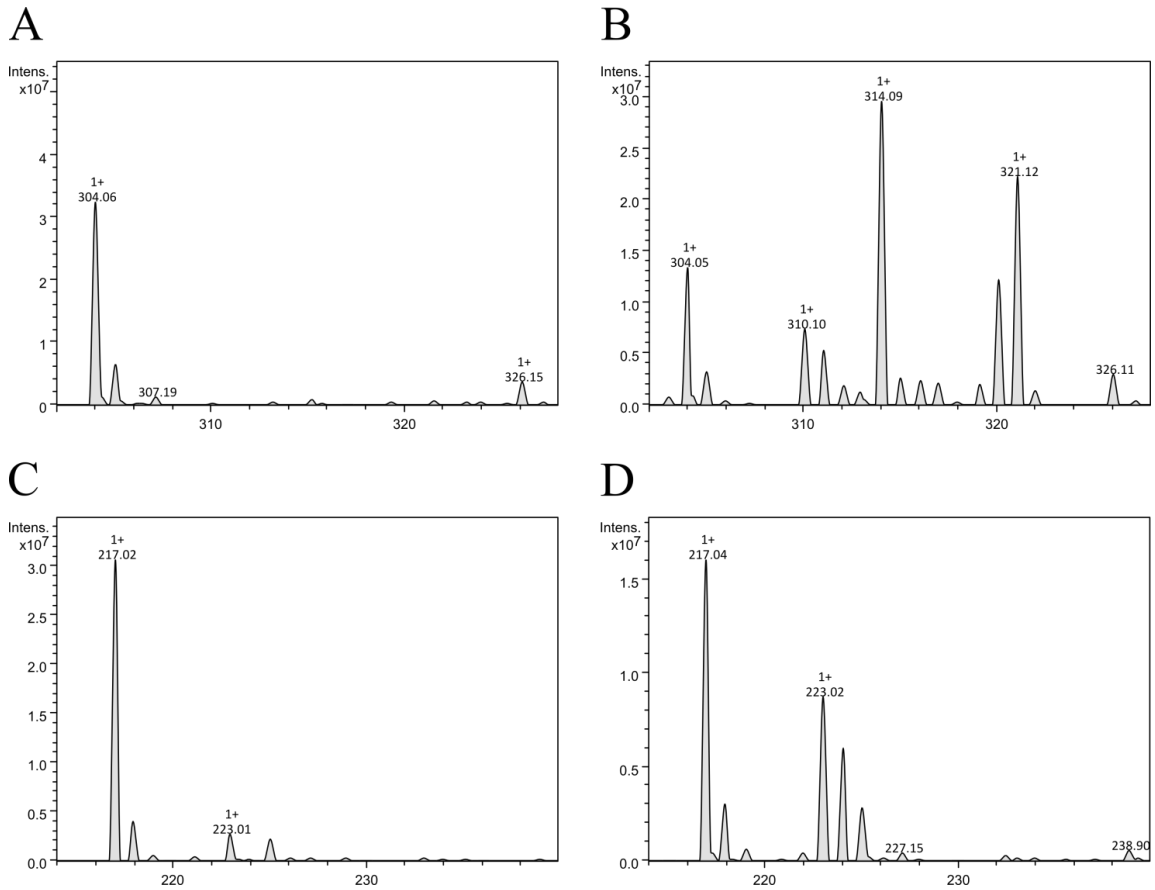

Figure 2.2. Incorporation of tryptophan ^{13}C11 into mansouramycins A and D. A – mass spectrum of mansouramycin D (m/z 304 $[M + H]^+$) extracted from the non-fed culture of *S. albus*. B – mass spectrum of mansouramycin D after feeding with tryptophan ^{13}C11. Mass shifts +6, +7, +10, +16 and +17 can be observed. C – mass spectrum of mansouramycin A (m/z 217 $[M + H]^+$) extracted from the non-fed culture of *S. albus*. D – mass spectrum of mansouramycin A after feeding with tryptophan ^{13}C11. Mass shifts +6 and +7 can be observed.

mansouramycin D peak. The additional ions with m/z values of 310, 311, 314, 320 and 321 correspond to the mass shifts of +6, +7, +10, +16 and +17. This unambiguously demonstrates that L-tryptophan serves as precursor for the biosynthesis of mansouramycin D. The simultaneous

presence of the ions with m/z values of 310 and 311 and of the ions with m/z values of 320 and 321 is intriguing. The intensities of the ions m/z 311 and m/z 321 are too high to expect that they originate from isotopic distribution of the ions with the m/z values of 310 and 320. The presence of the mass shifts +6, +7, +10, +16 and +17 implies that C6, C7 and C10 fragments of L-tryptophan can be incorporated into the structure of mansouramycin D. C6 and C10 fragments can be simultaneously incorporated into mansouramycin D yielding mass shift +16. Similarly, simultaneous incorporation of C7 and C10 fragments of L-tryptophan is possible and yields mass shift +17. Simultaneous incorporation of C6 and C7 fragments does not occur since the mass shift of +13 cannot be observed in the mass spectrum of the labeled mansouramycin D. The incorporation of either C6 or C7 L-tryptophan unit into mansouramycin D implies that either two pathways exist which lead to the production of mansouramycin from both C6 and C7 tryptophan fragments or two separate pathways exist which lead from L-tryptophan to the same C7 biosynthetic precursor of mansouramycin D. In one of the pathways seven carbons of L-tryptophan are sustained while in the other pathway – only six, and the seventh originates from the source other than tryptophan.

Initially, it was assumed that only indole moiety in the structure of mansouramycin D is derived from L-tryptophan. The results of the feeding studies however demonstrate that 17 carbons of 18 carbons present in the structure of mansouramycin D are derived from L-tryptophan. Taking into account that only 11 carbons are present in the structure of L-tryptophan, two structurally different mansouramycin precursors should be synthesized from tryptophan to allow simultaneous incorporation of 17 carbons. One possible precursor inherits 10 carbons from tryptophan and probably involves C8 indole moiety. The second mansouramycin precursor inherits either 6 or 7 carbons from L-tryptophan and is possibly responsible for the formation of the quinone moiety in the structure of mansouramycin D.

The observed incorporation pattern of labeled carbons from l-tryptophan into the structure of mansouramycin D is quite complex. In order to prove that 17 carbon atoms from 18 found in the structure of mansouramycin D are derived from L-tryptophan, it was decided to purify labeled mansouramycin D for NMR studies. For this purpose, the feeding of the *S .albus* Del14 culture with C^{13}-labelled L-tryptophan (^{13}C11) was performed in 5L volume. 2 mg of labeled mansouramycin D were purified. NMR analysis of the purified compound was carried out by Dr. Suvd Nadmid.

When comparing the ^{13}C-NMR data of carbon labeled mansouramycin D with the one of the unlabeled compound, signal intensity and signal pattern of the carbon atoms incorporated from ^{13}C isotopically labeled precursor have been obviously altered. ^{13}C-NMR data of mansouramycin D purified from fully carbon labeled tryptophan (L-Trp $^{13}C_{11}$) fed culture have shown signal splitting on all carbon signals except for 7-*N*-Me group, indicating incorporation of isotopically labeled carbons from the fed precursor on these carbons. (Figure 2.3). This confirms the incorporation of 17

carbon atoms from L-tryptophan into the structure of mansouramycin D (Figure 2.4). Both isoquinoline quinone and indole moieties of mansouramycin D were uniformly labeled. No labeling of the N-methyl group of mansouramycin D could be detected indicating that the N-methyl group is not derived from L-tryptophan.

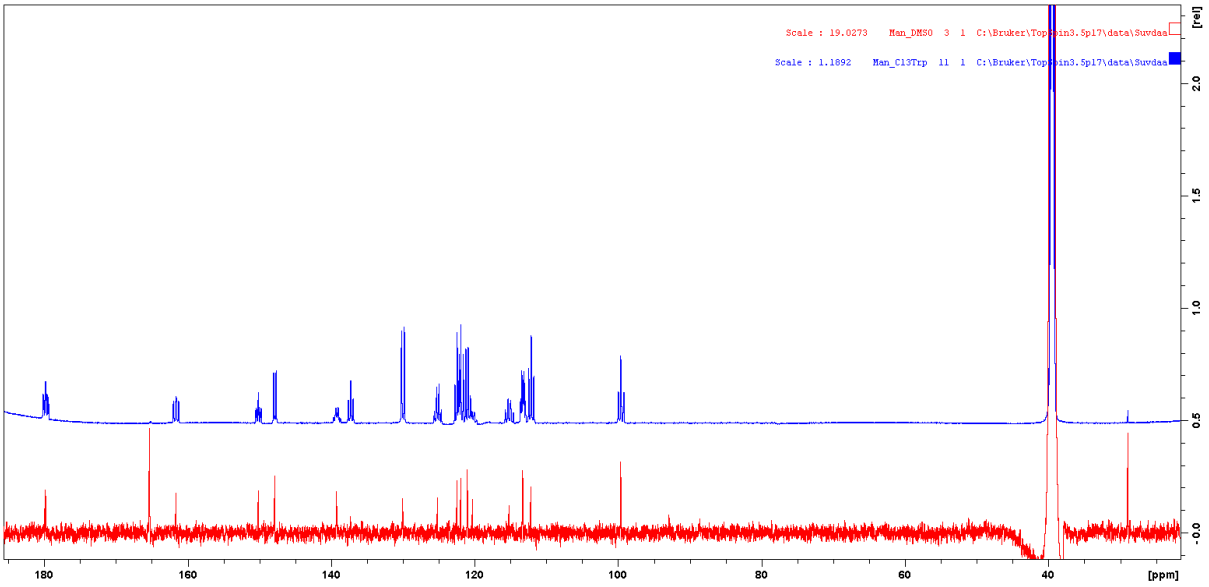

Figure 2.3. ^{13}C-NMR data of mansouramycin D purified from fully carbon labeled tryptophan (L-Trp ^{13}C11) fed culture of *S. albus* Del14. Blue – labeled compound, red – original compound

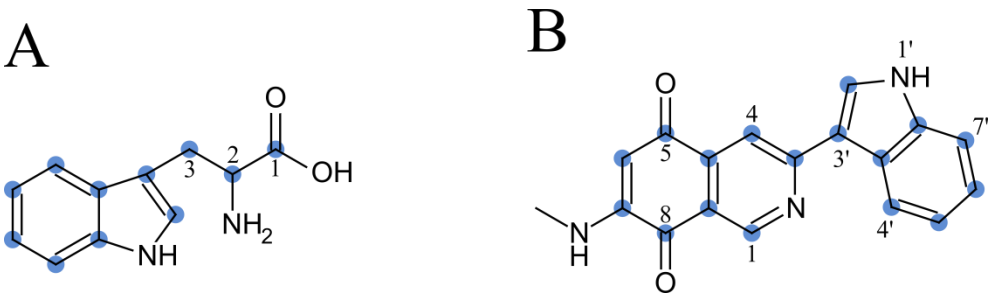

Figure 2.4. Labeling patterns of the L-tryptophan 11C^{13} (A) used for feeding studies and of the mansouramycin D (B) which was purified after feeding of L-tryptophan 11C^{13}.

Production of mansouramycin D by *S. albus* was demonstrated by both high-resolution HPLC-MS and NMR-analysis. In addition to mansouramycin D, a peak presumably corresponding to mansouramycin A was detected in HPLC-MS traces of *S. albus* extracts. Having demonstrated incorporation of carbon atoms of L-tryptophan into the structure of mansouramycin D, it was interesting to analyze the incorporation pattern of L-tryptophan carbon atoms into the structure of mansouramycin A. Analysis of the HPLC-MS traces of *S. albus* extracts after feeding with L-tryptophan ^{13}C11 revealed its incorporation into mansouramycin A (Figure 2.2 C and D). The mass shifts +6 and +7 can be clearly seen in the mass spectrum of mansouramycin A peak after the feeding. Such shifts cannot be observed in the mansouramycin A spectra from the control, non-fed culture of *S. albus*. The mass shifts +6 and +7 indicate that the C6 and C7 fragments of L-

tryptophan are incorporated into mansouramycin A. Similarly to mansouramycin D, simultaneous incorporation of C6 and C7 fragments does not occur into mansouramycin A since the mass shift of +13 cannot be observed in the mass spectrum of the labeled compound. In contrast to mansouramycin D a mass shift +10 cannot be observed in the mass spectrum of labeled mansouramycin A. This can be interpreted through structural differences between mansouramycins A and D. Mansouramycin A harbors two methyl groups at the positions 3 and 4 while mansouramycin D harbors an indole group at the position 3. It was proposed that two structurally different, L-tryptophan-derived precursors C10 and C7/C6 are necessary for the formation of the mansouramycin D ring structure. C7/C6 fragment of L-tryptophan is responsible for the formation of the quinone ring while the C10 fragment involves the indole ring. The absence of the +10 shift in the mass spectrum of the labeled mansouramycin A is in accordance with the absence of the indole group in the structure of mansouramycin A. It was proposed that in case of mansouramycin A only one C7/C6 L-tryptophan fragment plays a precursor role in mansouramycin A biosynthesis. Instead of C10 fragment of L-tryptophan a C4 unit of unknown origin is used to build pyridine ring in mansouramycin A.

In order to determine which carbons in the structure of mansouramycin A are inherited from L-tryptophan an attempt to purify labeled mansouramycin A has been performed. Unfortunately, the amount and purity of the obtained compound did not suffice to perform NMR studies.

The results of the feeding studies with labeled amino acids provide evidence that the ring structure of the isoquinoline quinone mansouramycin D is completely derived from L-tryptophan. In the case of mansouramycin A, its ring structure is partially derived from L-tryptophan. Therefore mansouramycins A and D are the first example of isoquinolines which are derived from L-tryptophan instead of L-tyrosine.

3.3 Feeding studies: incorporation of the side chain of L-tryptophan in mansouramycin D

Incorporation of C10 and C7 fragments of L-tryptophan in mansouramycin D was demonstrated by the feeding studies with the L-tryptophan ^{13}C11. Incorporation of 10 carbon atoms of L-tryptophan from 11 available indicates that one carbon atom of L-tryptophan is not used in the biosynthesis of mansouramycin D. It is logical to assume that the incorporated C10 fragment encompasses the indole ring and two additional carbon atoms of the side chain of L-tryptophan. In order to find out which of the carbon atoms of the tryptophan side chain is not incorporated in mansouramycin D L-Tryptophan C^{13}-labelled in the first position (carboxylic group) and the mixture of D- and L-tryptophan C^{13}-labelled in the second position (DL-tryptophan-2-^{13}C) were fed to *S. albus* Del14 culture. HPLC-MS analysis of the culture extracts revealed that the carbon atom in the first position of L-tryptophan is not incorporated in mansouramycin D. Incorporation of the carbon atom at the second position of L-tryptophan in mansouramycin D was unambiguously

demonstrated by the analysis of the MS-spectrum of the mansouramycin D extracted from the *S. albus* culture fed with the DL-Tryptophan-2-^{13}C (Figure 2.5). A clear mass-shift +1 can be observed in the corresponding mass spectrum.

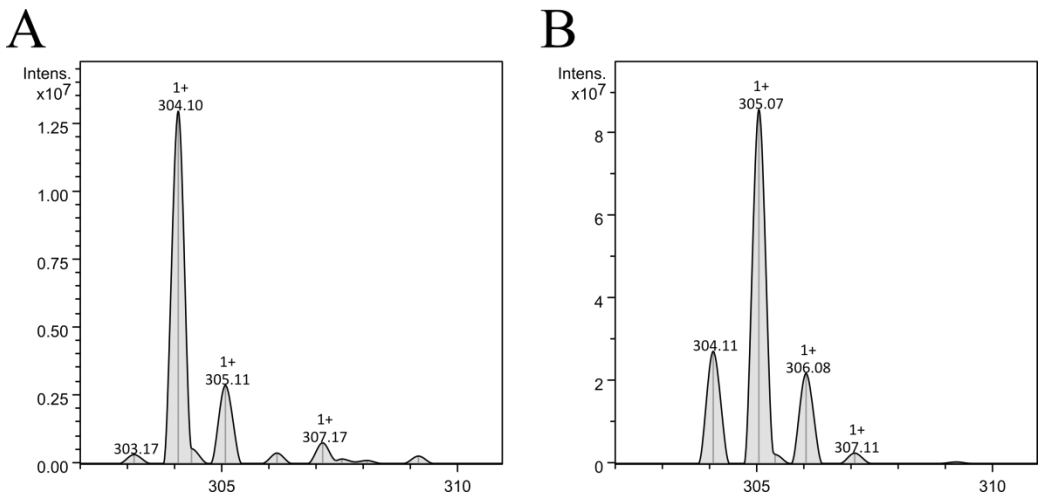

Figure 2.5. Mass spectra of mansouramycin D extracted from the non-fed culture of S .albus Del14 (A) and of mansouramycin D after feeding with DL-Tryptophan-2-^{13}C.

To determine at which position of mansouramycin D the labeled carbon atom of DL-Tryptophan-2-^{13}C is incorporated, NMR studies were carried out. For this purpose, the DL-Tryptophan-2-^{13}C was fed to 5L culture of *S. albus* Del14 and the labeled mansouramycin D was purified. NMR studies of the purified compound were performed by Dr. Suvd Nadmid. After feeding with L-Trp-2-13C enriched signal only at position C-3 could be observed in ^{13}C-NMR spectrum of labeled mansouramycin D (Figure 2.6). This indicates that the carbon atom at the position 2 of tryptophane is incorporated at the position 3 of mansouramycin D (Figure 2.7). Positioning of the labeled carbon at the position 3 of mansouramycin D also indicates a structural rearrangement during biosynthesis

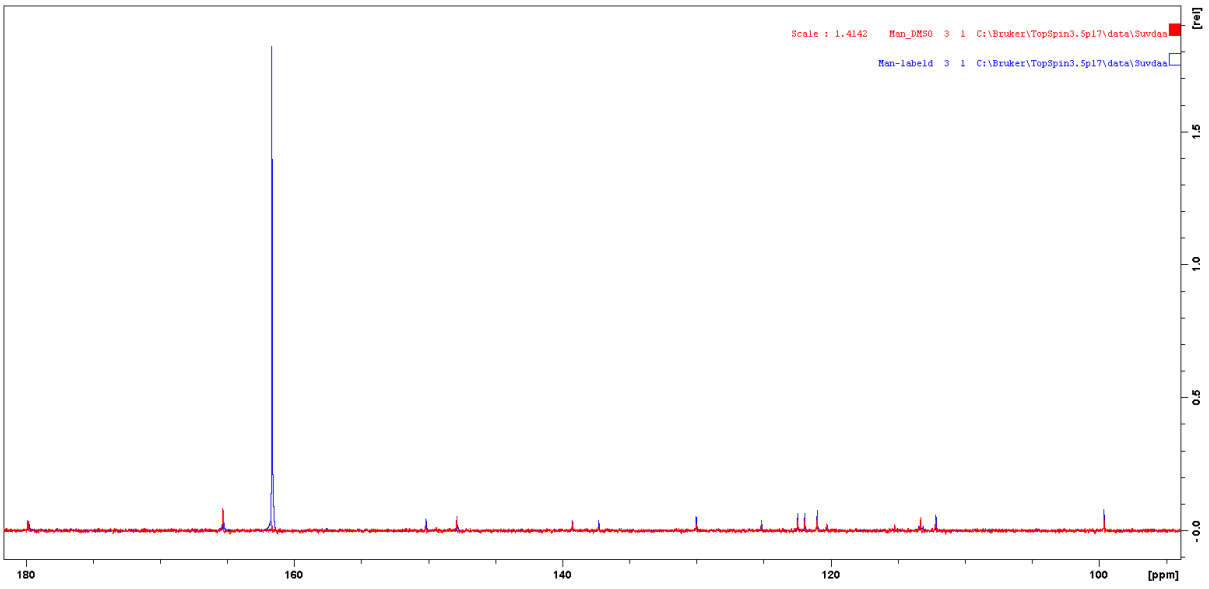

Figure 2.6. ^{13}C-NMR data of mansouramycin D purified from *S. albus* Del14 culture fed with DL-Tryptophan-2-^{13}C. Blue – labeled compound, red – original compound.

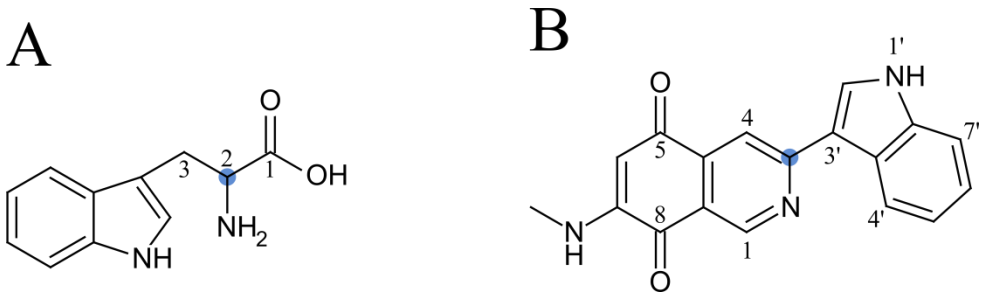

Figure 2.7. Labeling patterns of the DL-tryptophan-2-^{13}C (A) used for feeding studies and of the mansouramycin D (B) which was purified after feeding of DL-Tryptophan-2-^{13}C.

of its C10 tryptophan-derived precursor. In tryptophan, indole moiety is attached to the carbon in the position 3 of the side chain while in the structure of mansouranycin D indole moiety is bound to the carbon atom at the position 3 which was initially found at the position 2 of Tryptophan. From the structure of mansouramycin D we propose that the carbon atom originally found in tryptophan at the position 3 is located at the position 4 of mansouramycin D. Nevertheless the results of the feeding and NMR studies clearly indicate that during the mansouramycin D biosynthesis the indole moiety is transferred from the carbon at the position 3 of L-tryptophan to the atom in the position 2. A similar shift of the indole group was previously reported to occur in the biosynthesis of violacein [17,18].

3.4 Feeding studies: incorporation of nitrogen atoms of L-tryptophan in mansouramycins A and D

Previous feeding studies with C^{13} labeled L-tryptophan revealed that 17 of 18 carbon atoms of mansouramycin D are derived from L-tryptophan. Three nitrogen atoms can be found in the structure of mansouramycin D: one in the indole moiety, one in the isoquinolinequinone moiety and one in the methylamine moiety attached at the position 7 to isoquinolinequinone ring (Figure 2.1, Figure 2.7). Only two nitrogen atoms can be found in the structure of mansouramycin A as the methyl group instead of indole moiety is found at the position 3 (Figure 2.1). The origin of nitrogen atoms in the structures of mansouramycins is unknown. Taking into account that the ring structure of mansouramycin D is completely derived from L-tryptophan is logical to assume that some of the nitrogen atoms of the compound are also derived from L-tryptophan. In order to prove this, N^{15}-labelled L-tryptophan (L-Trp ^{15}N2) (Figure 2.8) was fed to the culture of *S. albus* Del14. HPLC-MS analysis of the mansouramycins extracted from the fed and control non-fed cultures revealed incorporation of nitrogen atoms into mansouramycins (Figure 2.9). Mass shifts +1 and +2 can be seen in the mass spectrum of mansouramycin D from the fed culture (Figure 2.8 A and B). In the mass spectrum of the mansouramycin A only mass shift +1 can be seen (Figure 2.8 C and D). The presence of the mass shifts +1 and +2 and the absence of the mass shift +3 in the mass spectrum of

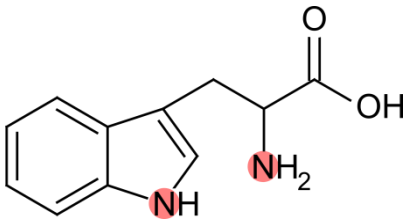

Figure 2.8. Labeling patterns of the L-tryptophan $^{15}N2$ used for feeding studies

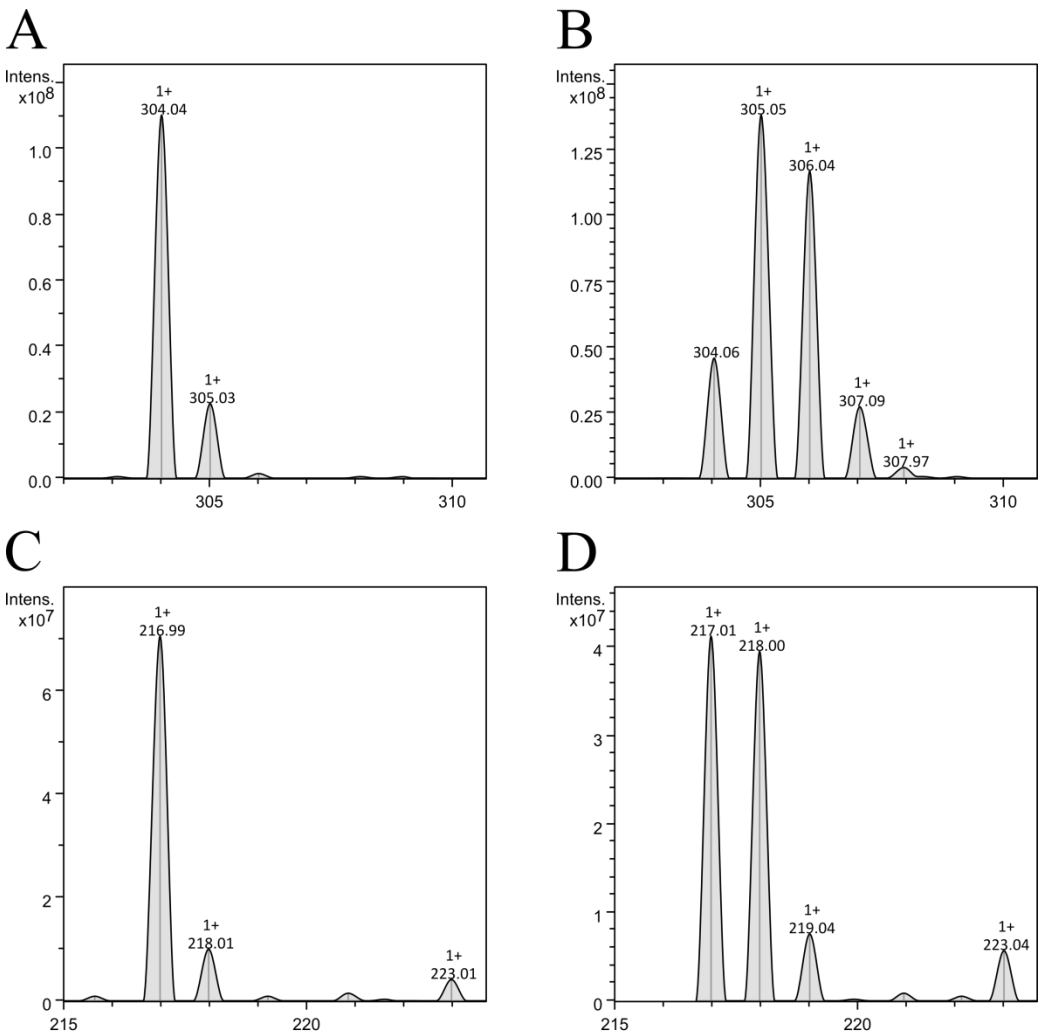

Figure 2.9. Mass spectra of mansouramycins extracted from the non-fed culture of *S .albus* Del14 and of after feeding with L-Tryptophan-$^{15}N2$. A – mass spectrum of mansouramycin D without feeding; B – mass spectrum of mansouramycin D after feeding; C – mass spectrum of mansouramycin A without feeding; D – mass spectrum of mansouramycin A after feeding.

mansouramycin D indicates that two nitrogen atoms are independently incorporated in mansouramycin D. The mass shift +2 corresponds to two independent incorporation events. Feeding studies with C^{13}-labelled tryptophan indicated that mansouramycin D precursors C7/6 and C10 are derived from L-tryptophan. The results of feeding of L-tryptophan $^{15}N2$ indicate that each of the precursors (C7/6 and C10) contains one nitrogen atom of L-tryptophan. This assumption is in accordance with the results of HPLC-MS analysis of mansouramycin A extracted from the *S. albus* Del14 culture after feeding with L-tryptophan $^{15}N2$. Only mass shift +1 can be detected in the mass

spectrum of the labelled mansouramycin A confirming that only C7/6 tryptophan derived precursor takes part in the formation of mansouramycin A. Taking into account the length of the side chain of L-tryptophan (C3) and the size of the putative mansouramycin precursors (C7/6 and C10) we assume that the nitrogen atom of the indole ring is incorporated while the nitrogen of the side chain is not.

3.5 Feeding studies: incorporation anthranilic acid in mansouramycins A and D

Both C7/C6 and C10 mansouramycin biosynthetic precursors are derived from L-tryptophan. The molecule of L-tryptophan contains 11 carbon atoms: eight within the indole ring and additional three in the side chain. In order to synthesize the C7/C6 biosynthetic precursor, the backbone of the indole ring should be cleaved. Such cleavage of the indole ring takes place during catabolism of tryptophan. The tryptophan-to-anthranilate degradation pathway is widespread among microorganisms [19]. The tryptophan degradation pathway leading to the production of anthranilate was described for *Streptomyces coelicolor* [20]. Three genes SCO3644, SCO3645, and SCO3646 encode for a kynurenine formamidase (KFA), a kynureninase (KYN), and a tryptophan 2,3-dioxygenase (TDO) respectively. TDO converts tryptophan in N-formyl-kynurenine (Figure 2.10). Subsequently, KFA removes the formyl group producing L-kynurenine. The latter is hydrolyzed by KYN producing anthranilate [20]. The three gene tryptophan degradation pathway is conservative among streptomycetes and can also be found in *S. albus* genome. The product of tryptophan

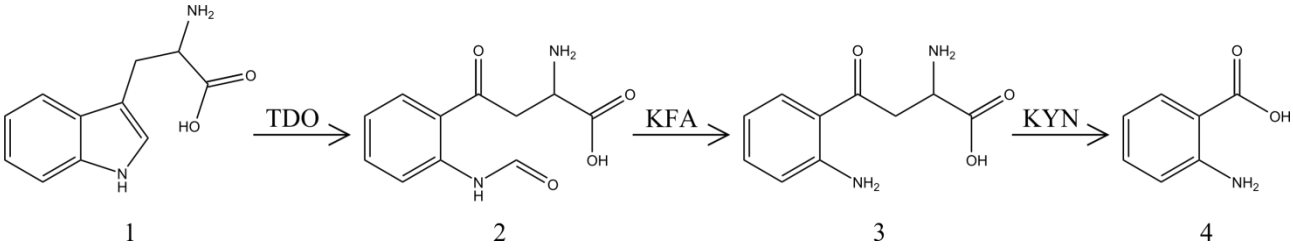

Figure 2.10. Tryphophane catabolism pathway in *S. coelicolor*. 1 – L-tryptophan, 2 - N-formyl-kynurenine, 3 – L-kynurenine, 4 – anthranilic acid.

degradation, anthranilic acid, might be used as a precursor in the biosynthesis of mansouramycins. Furthermore, anthranilic acid contains 7 carbon atoms, which could explain +7 shift in the mass spectrum of mansouramycins after feeding with fully labeled L-tryptophan ^{13}C11.

In order to prove the precursor role of anthranilic acid in the biosynthesis of tryptophan, labeled anthranilic acid (phenyl-^{13}C6) was fed to *S. albus* Del14. HPLC-MS analysis of the resulting extracts demonstrated incorporation of the anthranilic acid into mansouramycins. The mass shift +6 can be seen in the mass spectra of both mansouramycin A and D implying incorporation of the phenyl ring of the labeled anthranilic acid (phenyl-13C6) (Figure 2.11). In the mass spectrum of

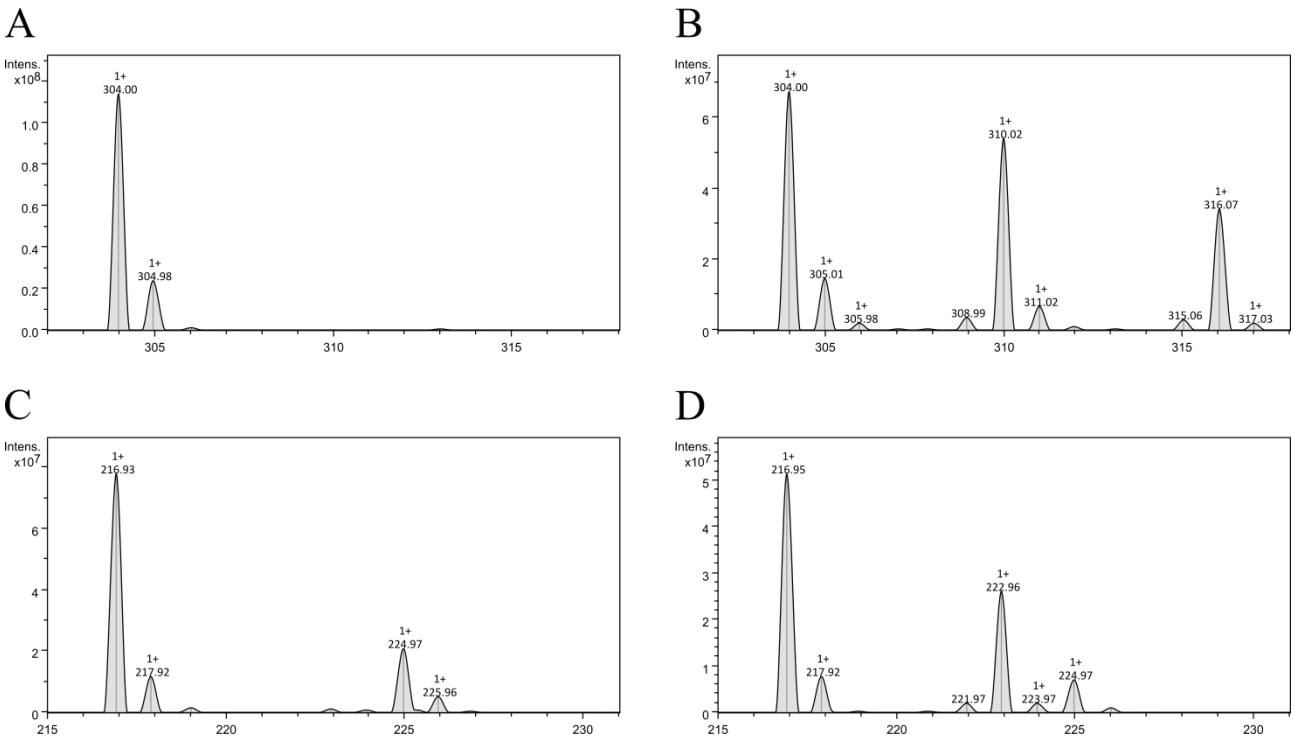

Figure 2.11. Incorporation of anthranilic acid (phenyl-^{13}C6) in mansouramycins. A and C – mass spectra of mansouramycins D and A extracted from the control, non-fed culture of *S. albus* Del14. B and D – mass spectra of mansouramycins D and A from the fed culture of *S. albus* Del14

mansouramycin D additional mass shift +12 can be observed. This mass shift can be only explained if two phenyl rings derived from labeled anthranilic acid are simultaneously incorporated into mansouramycin D. This is however against the initial assumption that the anthranilic acid corresponds to the C7/6 mansouramycin precursor. Simultaneous incorporation of two anthranilic acids in mansouramycin D is only possible if anthranilic acid is incorporated in both C7/6 and C10 mansouramycin precursor. From the previous study, the C10 precursor was assumed to consist of indole ring of tryptophan and two carbon fragment of its side chain. One of the possible ways for anthranilic acid to be involved into C10 mansouramycin precursor is the *denovo* biosynthetic pathway of tryptophan from anthranilic acid (Figure 2.12) [21]. The biosynthesis of L-tryptophan from anthranilic acid questions the role of anthranilic acid as a direct precursor of tryptophan. The possibility exists that anthranilic acid is incorporated into mansouramycins indirectly through biosynthesis of L-tryptophan, from which C7/6 and C10 units are derived. This will however mean that anthranilic acid is not a direct precursor of mansouramycins A and D. Furthermore degradation of tryptophan to anthranilic acid and *denovo* biosynthesis of tryptophan can explain the simultaneous existence of the mass shifts +7 and +6 as well as of mass shifts +17 and +16 in the mass spectrum of mansouramycin D after feeding with L-tryptophan-^{13}C11. During the degradation of the fully-labeled L-tryptophan to anthranilic acid only 7 labeled carbon atoms remain (Figure 2.10). During following *denovo* biosynthesis of L-tryptophan one of the labeled carbons is removed due to decarboxylation during conversion of carboxyphenylamino-deoxyribulose-5-phosphate into

indole-3-glycerol phosphate (Figure 2.12) [21]. Therefore after one cycle of degradation and *denovo* biosynthesis only six labeled carbons remain in the structure of L-tryptophan.

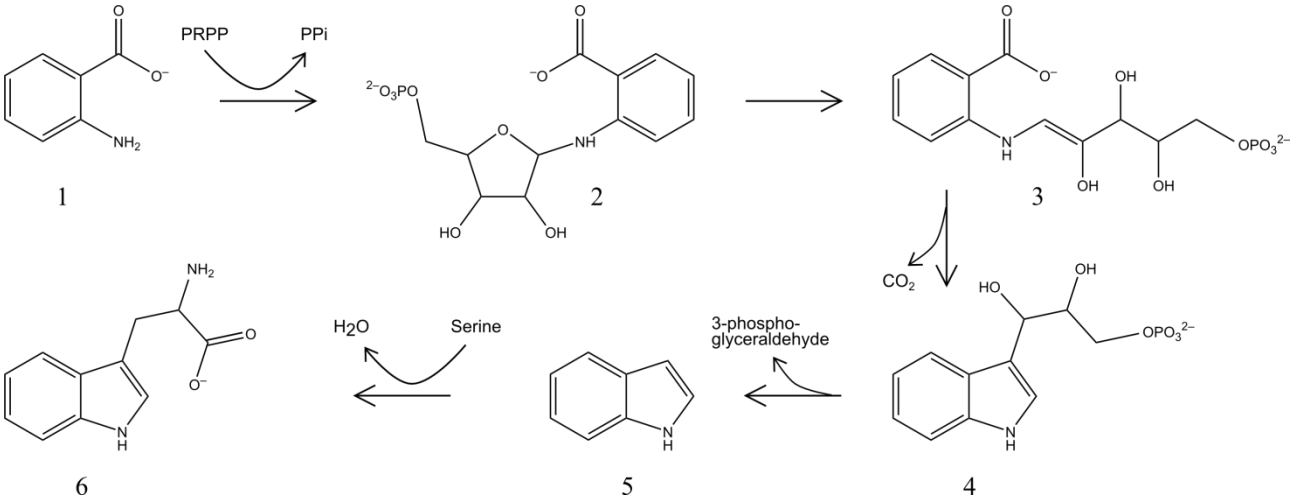

Figure 2.12. Tryptophan biosynthetic pathway. 1 – anthranilic acid, 2 – phosphoribosyl anthranilate, 3 – carboxyphenylamino-deoxyribulose-5-phosphate, 4 – indole-3-glycerol phosphate, 5 – indole, 6 – tryptophan.

3.6 Inactivation of tryptophan biosynthetic pathways in *S. albus*. Anthranilic acid is not a direct precursor for mansouramycin biosynthesis

Simultaneous incorporation of two phenyl groups of anthranilic acid in mansouramycin D implies that anthranilic acid can be an indirect precursor of mansouramycins and is incorporated in mansouramycin through biosynthesis of a direct precursor – L-tryptophan. In order to clarify if anthranilic acid is directly incorporated in mansouramycins or not, a mutant of *S. albus* Del14 auxotrophic for L-tryptophan was constructed. For this purpose, the genes XNR_4841 and XNR_4842 encoding beta and alpha chains of tryptophan synthase were deleted in the chromosome of *S. albus* Del14 strain. The constructed strain was named *S. albus* biosW. The alpha subunit of tryptophan synthase is responsible for the aldol cleavage of indole-3-glycerol phosphate to indole and glyceraldehyde 3-phosphate, while the beta subunit is responsible for the synthesis of L-tryptophan from indole and L-serine. The strain *S. albus* biosW was unable to grow on a minimal agar medium without supplementation with L-tryptophan. On the complete MS agar medium, the strain did not show any differences in growth compared to the *S. albus* Del14 strain. However, when cultivated in the production liquid medium DNPM no mansouramycin production could be detected. The production of mansouramycins could be restored by additional supplememtation with L-tryptophan. Therefore the mansouramycin production seems to require higher intracellular tryptophan concentrations than those required for survival and normal growth.

The labeled anthranilic acid (phenyl-13C6) was fed together with non-labeled L-tryptophan to the cultures of *S. albus* Del14 and *S. albus* biosW. The non-fed culture of *S. albus* Del14 was used as a control. The mansouramycin production was analyzed by HPLC-MS (Figure 2.13). In the

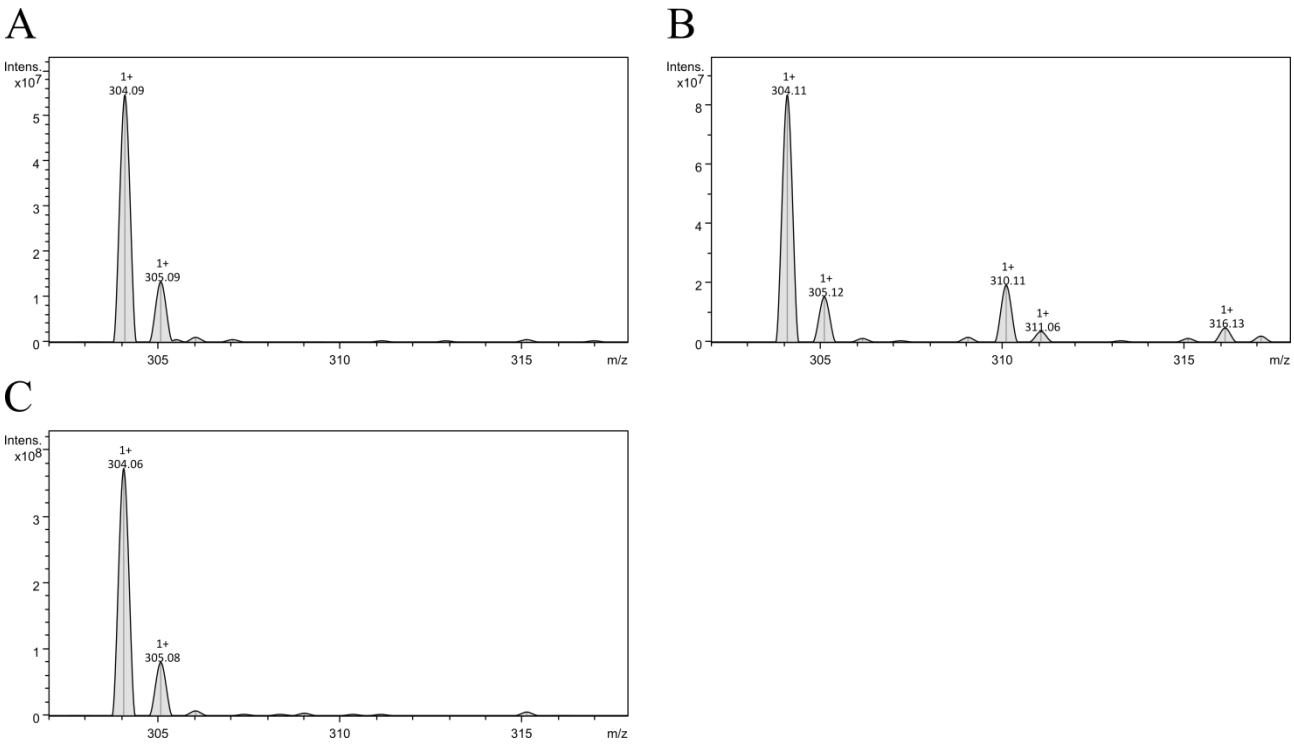

Figure 2.13. Feeding of the labeled anthranilic acid (phenyl-13C6). A – mass spectrum of mansouramycin D extracted from the non-fed culture of *S. albus* Del14. B – mass spectrum of mansouramycin D extracted from the culture of *S. albus* Del14 fed with labeled anthranilic acid and non-labeled L-tryptophan. C – mass spectrum of mansouramycin D extracted from the fed culture of *S. albus* biosW.

mass spectrum of mansouramycin D extracted from the fed culture of *S. albus* Del14 signals m/z = 304, m/z = 310 and m/z = 316 can be seen (Figure 2.13 B). The latter two signals correspond to the previously observed mass shifts +6 and +12 and demonstrate simultaneous incorporation of two phenyl groups of anthranilic acid in mansouramycin D. In the mass spectrum of the mansouramycin D extracted from the fed culture of *S. albus* biosW (Figure 2.13 C) only signal m/z = 304 can be seen which corresponds to the non-labeled mansouramycin D. This result demonstrate that the inactivation of the pathway leading to tryptophan biosynthesis from anthranilic acid prevents incorporation of anthranilic acid in mansouramycin D. Therefore anthranilic acid is not a direct precursor of mansouramycins. It is incorporated in the structure of mansouramycins through L-tryptophan.

3.7 Feeding of the ^{15}N-labelled anthranilic acid. Nitrogen atom of the methylamino group at the position 7 of mansouramycins is derived from L-tryptophan while the atom at the position 2 of mansouramycins is not

Feeding studies with 15N-labelled L-tryptophan showed that two of three nitrogen atoms present in mansouramycin D and one of two nitrogen atoms present in mansouramycin A are derived from L-tryptophan. These results indicate that only one of two nitrogen atoms of the 3-methyl-7-(methylamino)-5,8-isoquinolinequinone core of mansouramycins is tryptophan-derived. The relatively high price of the labeled L-tryptophan ^{15}N2 did not allow performing feeding experiment in a preparative scale for purification of labeled mansouramycins for subsequent NMR

studies. In order to find out which of the two nitrogen atoms, one from the methylamino group at the position 7 or the atom at the position 2 of mansouramycins is derived from L-tryptophan ^{15}N-labelled anthranilic acid was used for feeding experiments. Previous results give evidence that the anthranilic acid is incorporated in mansouramycins indirectly, through the formation of L-tryptophan. During the *denovo* biosynthesis of tryptophan, the nitrogen atom of the indole ring is derived from anthranilic acid (Figure 2.12). Therefore we expected that through the feeding of the ^{15}N-labelled anthranilic acid to the culture of *S. albus* Del14 labeled L-tryptophan will be formed which will be used as a precursor for mansouramycin biosynthesis. NMR-analysis of the mansouramycins labeled in such an indirect way will allow determining which nitrogen atoms of mansouramycins correspond to the nitrogen atom of indole ring of L-tryptophan.

^{15}N-labeled anthranilic acid was fed to the culture of *S. albus* Del14. HPLC-MS analysis of the extracts revealed the incorporation of the labeled nitrogen (Figure 2.14). In the mass spectrum of mansouramycin D mass shifts +1 and +2 can be observed which correspond to incorporation of one and two labeled nitrogen atoms respectively. In the mass spectrum of the labeled mansouramycin A, only the mass shift +1 can be observed what corresponds to the incorporation of one labeled nitrogen atom. Labeled mansouramycin D was purified and used for NMR-analysis. NMR studies were performed by Dr. Constanze Paulus.

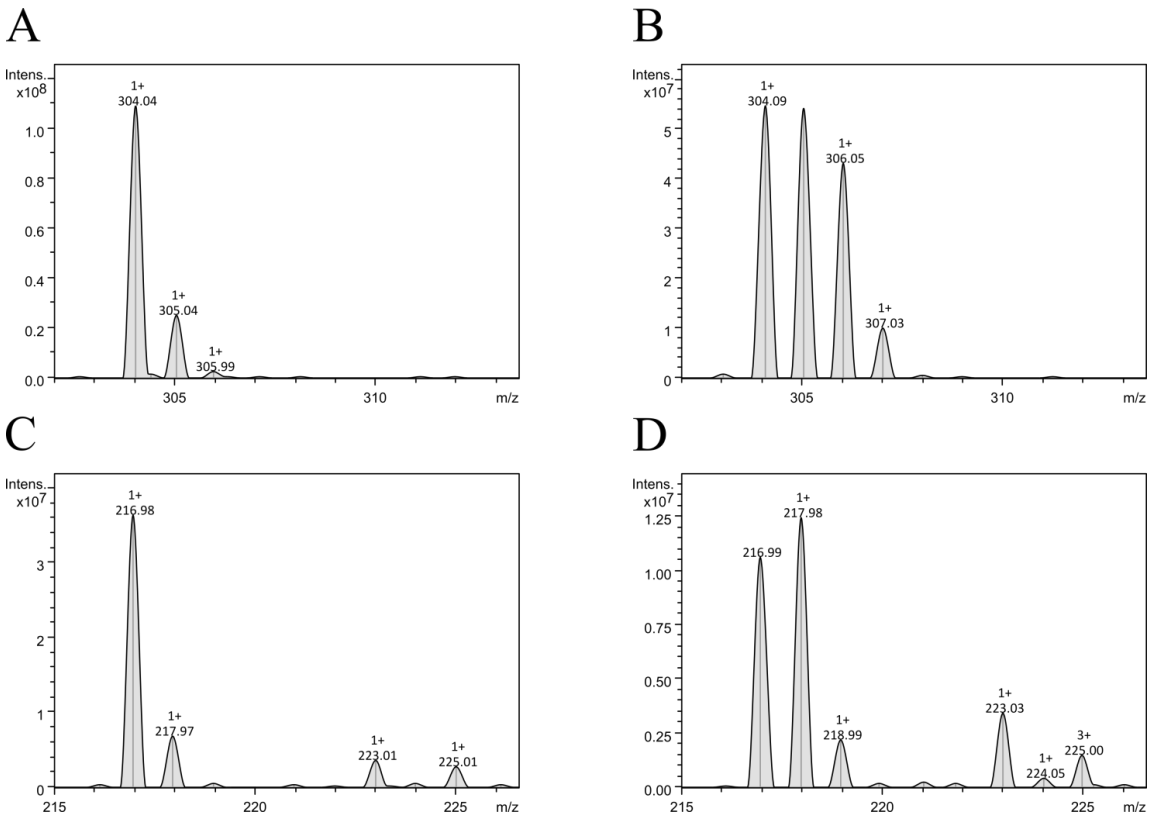

Figure 2.14. Incorporaztion of ^{15}N-labelled anthranilic acid. A – mass spectrum of mansouramycin D extracted from the non-fed culture of *S. albus* Del14. B – mass spectrum of mansouramycin D after feeding with ^{15}N-labelled anthranilic acid. C – mass spectrum of mansouramycin A extracted from the non-fed culture of *S. albus* Del14. D – mass spectrum of mansouramycin A after feeding with ^{15}N-labelled anthranilic acid.

Analysis of the ^{15}N-HMBC spectral data exhibited two correlations with strong intensity from proton at δ_H 8.53 to N-3' and from δ_H 2.82 to N-8 indicating those two nitrogens were enriched during feeding experiment (Figure 2.15, Table 2.5). As opposed to this, a signal for the nitrogen at position 2 is only visible in a very low intensity, suggesting that this atom is not labeled. The results of this analysis demonstrate that the nitrogen atom of the methylamine group at the position 7 of mansouramycin D and the nitrogen atom in the indole moiety are derived from anthranilic acid (Figure 2.16).

Performed feeding studies allow making following conclusions. 1 – L-tryptophan is a main biosynthetic precursor of mansouramycin D. 2 – All carbon atoms of the mansouramycin D except of the carbon atom of methylamino group at position 7 are derived from L-tryptophan. 3 – During biosynthesis of mansouramycin D two tryptophan derived precursors C7 and C10 are formed. 4 – C10 precursor retains an indole ring and two atoms of the side chain of L-tryptophan. 5 – During formation of the C10 precursor indole group is shifted from carbon atom 3 of the side chain of L-

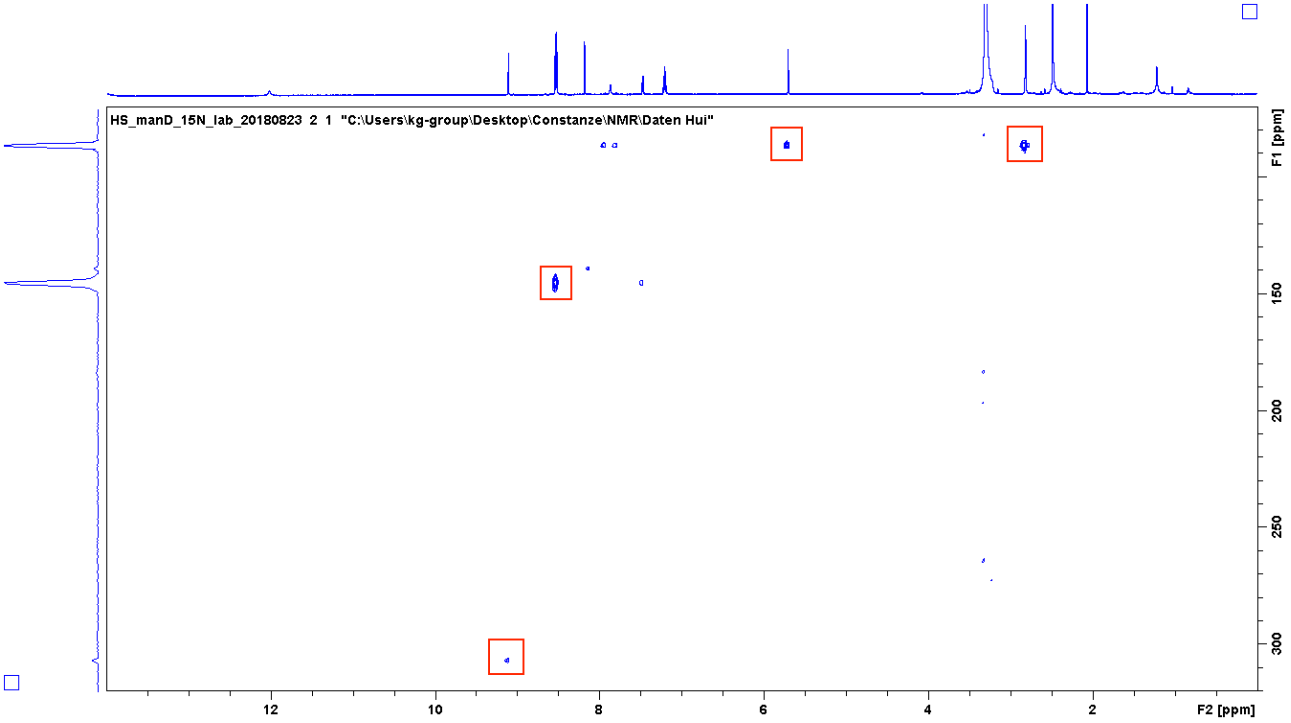

Figure2.15. ^{15}N-NMR data of mansouramycin D purified from *S. albus* Del14 culture fed with 15N-labelled anthranilic acid.

Table 2.5 NMR data (700 MHz, dmso-d_6) for ^{15}N labeled Mansouramycin D

Position	δ_H (in ppm)	δ_N (in ppm)	N-HMBC
1	9.11		2
2	-	306.4	-
6	5.71	-	8
8	7.88	86.4	-
N-Me	2.82		8
2'	8.53	-	3'
3'	12.02	145.5	-
7'	7.48	-	3'

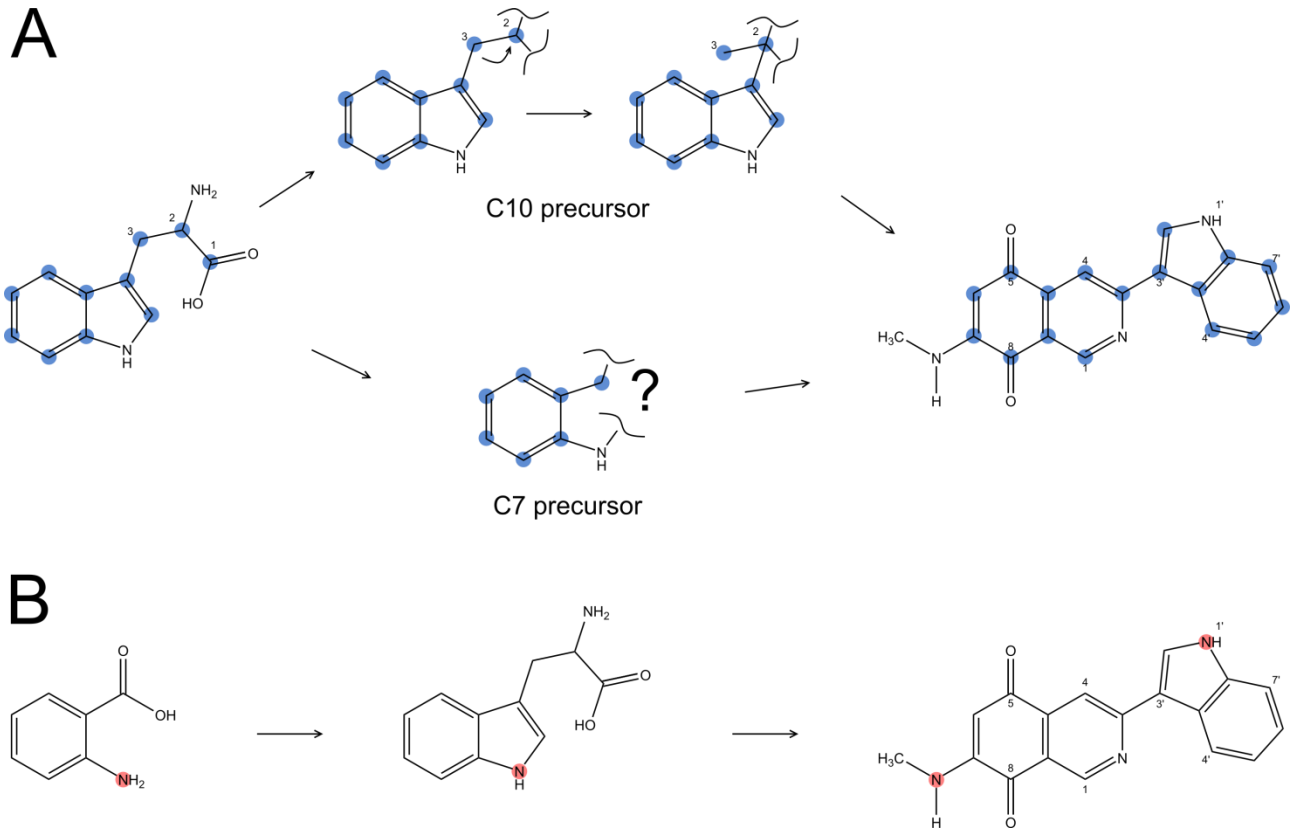

Figure 2.16. Scheme of mansouramycin D biosynthesis. A – 17 carbon atoms of mansouramycin D are derived from L-tryptophan. During formation of the C10 precursor the indole group is shifted from the carbon 3 to carbon 2. B – nitrogen atoms of the indole ring of tryptophan are inherited by mansouramycin D.

tryptophan to carbon at position 2 (Figure 2.16 A). 6 – Only Nitrogen atoms of the indole ring of L-tryptophan are retained in the structure of mansouramycin D (Figure 2.16 B). 7 – From the three nitrogen atoms present in the structure of mansouramycin D only the nitrogen atom of the isoquinoline ring is not tryptophan-derived.

3.8 Identification of the mansouramycin D cluster

Mansouramycins are produced by both *S. albus* J1074 and its cluster free derivative *S. albus* Del14. All secondary metabolite gene clusters which were detected by Antismash genome mining software [22] and which were dispensable under laboratory cultivation conditions were deleted in the chromosome of the *S. albus* Del14 strain [23]. The production of mansouramycins by *S. albus* Del14 implies that the respective biosynthetic cluster cannot be detected by the genome mining software. Results of the feeding experiments demonstrate that the mansouramycin D is for the most part derived from L-tryptophan. According to the proposed biosynthetic scheme for the formation of the C7 mansouramycin precursor the side chain of the L-tryptophan has to be removed and the indole ring cleaved. These catalytic steps also occur during catabolism of L-tryptophan (Figure 2.10). Therefore it was assumed that the pathway leading to the formation of the C7 mansouramycin precursor might have emerged from the conventional tryptophan degradation pathway. Such a possible evolutionary relationship would mean that the enzymes involved in the

degradation of tryptophan and biosynthesis of C7 precursor share substantial homology on the protein level. In order to prove this assumption, the peptide products of *S. coelicolor* genes SCO3644, SCO3645, and SCO3646 encoding for a kynurenine formamidase, a kynureninase, and a tryptophan 2,3-dioxygenase were used for BLAST similarity search against *S. albus* chromosome. This analysis revealed two regions in the chromosome of *S. albus* containing homologs of the above-mentioned genes. One of the regions contains the genes XNR_3202, XNR_3201 and XNR_3200 which share high levels of homology to the *S. coelicolor* genes SCO3644, SCO3645, and SCO3646. The genes XNR_3202, XNR_3201 and XNR_3200 were proposed to constitute the tryptophan degradation pathway in *S. albus*. The second chromosomal region of *S. albus* identified during BLAST search contains genes showing low homology rates to the genes of *S. coelicolor*. The XNR_4435 annotated as a putative aminotransferase showed homology to SCO_3645 encoding kynureninase. Two genes within the identified region, XNR_4434 and XNR_4436, showed low homology to SCO_3646 which encodes tryptophan 2,3-dioxygenase. The genes XNR_4434 and XNR_4435 are a part of a three gene operon comprising genes XNR_4433 – XNR_4435 (Figure 2.17). According to annotation, the gene XNR_4436 is the first gene of a five gene operon comprising genes XNR_4436 – XNR_4440 (Figure 2.17). Such transcriptional organization of the identified chromosomal region implies that not only the genes XNR_4434, XNR_4435 and XNR_4436 might be involved in tryptophan-related biochemical pathway but also other genes in the mentioned operons. Interestingly no homolog of the kynurenine formamidase gene SCO_3644 could be detected in the identified chromosomal region of *S. albus* using homology search. Nevertheless, a gene XNR_4438 within the identified region encodes a putative kynurenine formamidase. The genes XNR_4438 and SCO_3644 show no homology on the protein level.

Sequence analysis of the identified chromosomal region of *S. albus* revealed a set of 13 genes (XNR_4432 – XNR_4444) which might be involved in the biosynthesis of mansouramycins by *S. albus* strains (Figure 2.17) (Table 2.6). In order to prove if the genes XNR_4432 – XNR_4444 encode for mansouramycin biosynthesis, an attempt to delete these genes was undertaken. A single BAC 3F18 covering the region putatively involved in mansouramycin biosynthesis was identified in the ordered genomic library of *S. albus* J1074. The BAC 3F18 contains the fragment of *S. albus* chromosome which comprises the genes XNR_4379 – XNR_4445. Of the gene XNR_4445 only the 790 bp 3'-terminal fragment was cloned in 3F18. Initially the genes XNR_4432 – XNR_4444 were planned to be deleted. Due to very short DNA fragment available in 3F18 downstream of

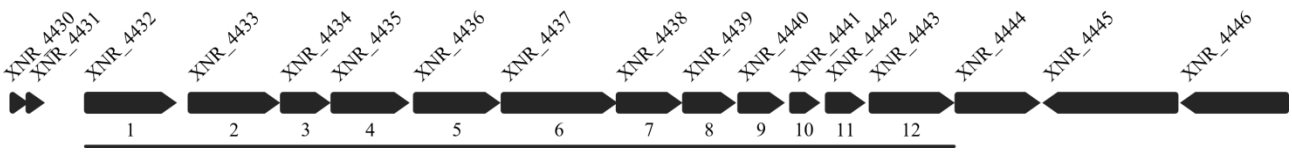

Figure 2.17. Mansouramycin biosynthetic gene cluster with flanking regions. The region marked with the black bar was deleted in *S. albus* Del15.

Table 2.6. Genes putatively involved in mansouramycin biosynthesis

Locus tag		Annotation
XNR_4429		Gas vesicle synthesis protein
XNR_4430		Gas vesicle synthesis protein
XNR_4431		Gas vesicle protein K
XNR_4432	1	Hypothetical protein
XNR_4433	2	Biotin carboxylase / ATP-grasp domain protein
XNR_4434	3	Tryptophan 2,3-dioxygenase
XNR_4435	4	Aminotransferase / Kynureninase
XNR_4436	5	Tryptophan 2,3-dioxygenase
XNR_4437	6	FAD-dependent oxidoreductase / Pentachlorphenol monooxygenase
XNR_4438	7	Branched amino acid aminotransferase / Transaminase
XNR_4439	8	Cyclase / Kynurenine formamidase
XNR_4440	9	Methyltransferase
XNR_4441	10	Pyridoxamine 5'-phosphate oxidase-related protein
XNR_4442	11	Flavin reductase
XNR_4443	12	Phosphoribosylglycinamide synthetase
XNR_4444		Major facilitator superfamily protein
XNR_4445		Amylo-alpha-1, 6-glucosidase
XNR_4446		Hypothetical protein

XNR_4444, only the genes XNR_4432 – XNR_4443 were deleted in 3F18 using RedET yielding 3F18_Am. The XNR_4444 and the terminal fragment of XNR_4445 were left in 3F18_Am for homology recombination purpose. The constructed BAC 3F18_Am was transferred in *S. albus* Del14 by conjugation. The double crossover mutant with the deleted region XNR_4432 – XNR_4443 was selected using blue-white screen. The apramycin resistance marker from the obtained mutant was removed using IMES system. The constructed marker-free deletion mutant was named *S. albus* Del15.

The mansouramycin production of the constructed *S. albus* Del15 strain was assayed using HPLC-MS. For this purpose, *S. albus* Del15 and *S. albus* Del14 were cultivated in the production medium DNPM and the culture supernatant was extracted with ethyl acetate. The masses corresponding to mansouramycin D and mansouramycin A could not be observed in the extract of the *S. albus* Del15 strain while being present in the extract of the *S. albus* Del14 strain (Figure 2.18). This clearly indicates that the DNA region deleted in the *S. albus* Del15 strain either encodes the entire or partial mansouramycin pathway or contains some regulatory genes influencing mansouramycin production in *S. albus*. Taking into account that no putative regulatory genes were present in the deleted region, the most plausible explanation for mansouramycin-null phenotype of *S. albus* Del15 strain is the presence of mansouramycin biosynthetic genes within the deleted region.

The more precise analysis of the HPLC-MS traces revealed that several peaks in addition to those of mansouramycins A and D disappeared in the extract of *S. albus* Del15 strain (Figure 2.18). These unidentified peaks may correspond to other mansouramycin derivatives which are produced by *S. albus* strain. The newly identified peaks correspond to the compounds with the m/z values of 356.1, 432.1, 490.1 and 371.1.

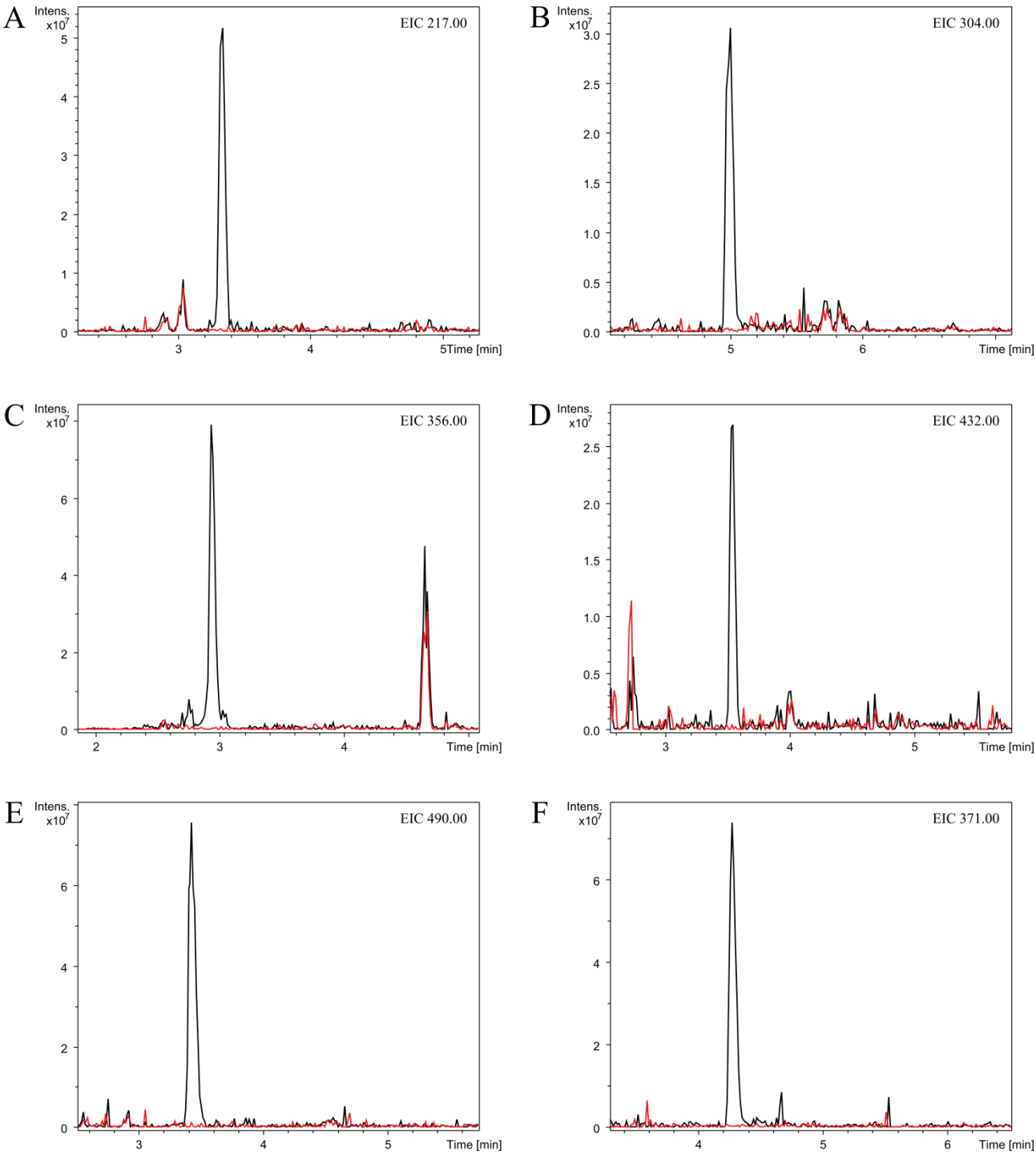

Figure 2.18. HPLC-MS analysis of mansouramycin production by *S. albus* Del14 (black traces) and *S. albus* Del15 (red traces) strains. Extracted ion chromatograms corresponding to individual mansouramycins are shown. A – extracted ion chromatogram 217 ± 0.5 Da showing production of mansouramycin A. A – extracted ion chromatogram 304 ± 0.5 Da showing production of mansouramycin D. C, D, E and F extracted ion chromatograms demonstrating production of putative novel mansouramycins with the m/z values of 356, 432, 490 and 371 respectively.

To find out if the identified and deleted DNA fragment covering the genes XNR_4379 – XNR_4445 encodes the complete mansouramycin gene cluster an attempt to express the DNA fragment in heterologous hosts was undertaken. The native BAC 3F18 which contains the above-mentioned genes cannot be directly transferred into streptomycetes by conjugation and cannot persist in the cells. To enable transfer and stable integration of the BAC in *Streptomyces* cells a cassette containing *phi*C31 integrase gene, *phi*C31 *att*P site, apramycin resistance marker and origin of conjugal transfer *oriT* was inserted into chloramphenicol resistance gene of the BAC backbone

of the 3F18 BAC. This modification was performed by RedET. The resulting BAC was named 3F18_int.

The BAC 3F18 as well as its derivative 3F18_int contain the DNA fragment covering genes XNR_4379 – XNR_4445. This fragment is much bigger than the DNA fragment putatively involved in mansouramycin biosynthesis (XNR_4432 – XNR_4444). In order to shorten the DNA insert cloned in the BAC 3F18_int the genes XNR_4382 – XNR_4431 were substituted in 3F18_int with ampicillin resistance cassette using RedET. The resistance gene was removed by cutting it out and self-ligating the BAC. The Final BAC was named 3F18_int_KO_0. This BAC contains the genes XNR_4379 – XNR_4381, the putative mansouramycin biosynthetic cluster – XNR_4432 – XNR_4444 and the 3'-terminal fragment of XNR_4445.

The constructed BAC 3F18_int_KO_0 was transferred in heterologous hosts *S. lividans* TK24 and *S. coelicolor* M145. The resulting strains were tested for their ability to produce mansouramycins A and D. HPLC-MS analysis of their extracts did not detect production of mansouramycins. This means that either the mansouramycin biosynthetic pathway is not active in the used heterologous strains or the expressed fragment does not encode mansouramycin biosynthetic pathway or the pathway is incomplete. In order to find heterologous host strains which will support the expression of the putative mansouramycin cluster, 3F18_int_KO_0 BAC was transferred into eight non-conventional *Streptomyces* heterologous host strains which were successfully used for pathway expressions in other research projects: LV1-4, LV1-22, LV1-114, LV1-18.2, LV1-208, LV1-209, LV1-213 and LV1-652. The resulting exconjugants were tested for their ability to produce mansouramycins by extracting their culture broths and HPLC-MS analysis. From all tested strains, only the LV1-652 strain successfully produced mansouramycins after expression of 3F18_int_KO_0 BAC (Figure 2.19). No mansouramycin production could be detected in the extracts of LV1-652 without the BAC (Figure 2.19). No homologs of the genes XNR_4432 – XNR_4444 which constitute the putative mansouramycin biosynthetic cluster were identified in the genome sequence of LV1-652. These results provide evidence that the DNA fragment covering the genes XNR_4432 – XNR_4444 encodes the biosynthetic pathway leading to mansouramycin production.

Introduction of the BAC 3F18_int_KO_0 in the strain *S. albus* Del15 with the deleted mansouramycin cluster completely restored mansouramycin production. This means that the BAC 3F18_int_KO_0 can be successfully used for complementation studies in *S. albus* Del15 strain.

3.9 Inactivation of individual mansouramycin biosynthetic genes

Heterologous expression of the BAC 3F18_int_KO_0 has demonstrated that the genes XNR_4432 – XNR_4444 encode the production of mansouramycins. To identify the minimal

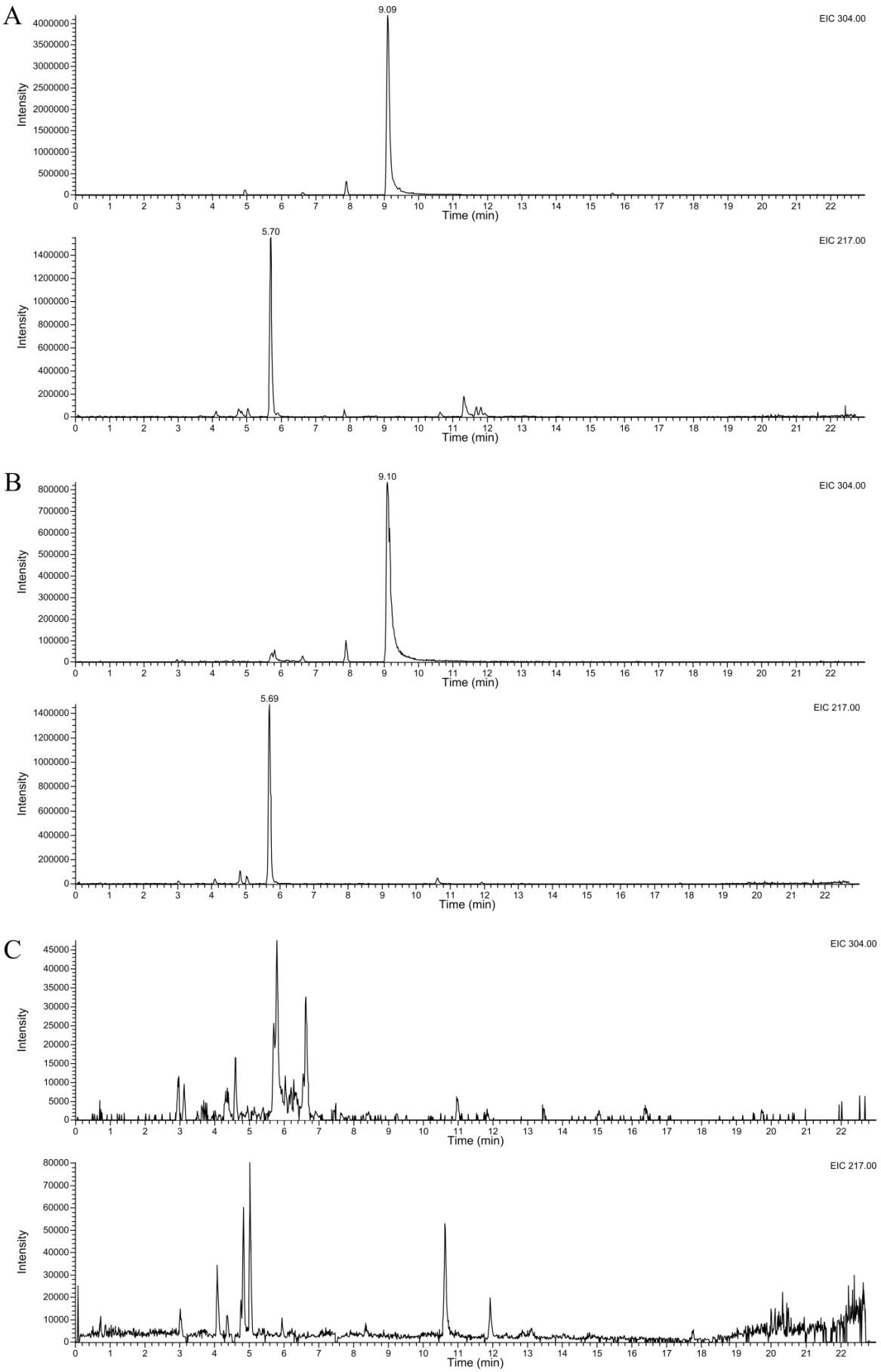

Figure 2.19 HPLC-MS analysis of mansouramycin production in heterologous host LV1-652. A – extracted ion chromatograms 304 ± 0.5 Da and 217 ± 0.5 Da showing the production of mansouramycins D and A by *S. albus* Del14. B – extracted ion chromatograms 304 ± 0.5 Da and 217 ± 0.5 Da showing production of mansouramycins D and A by the heterologous host strain LV1-652 harboring the BAC 3F18_int_KO_0. C – extracted ion chromatograms 304 ± 0.5 Da and 217 ± 0.5 Da showing that the heterologous host strain LV1-652 does not produce mansouramycins without the BAC 3F18_int_KO_0.

mansouramycin biosynthetic gene cluster gene inactivation studies were performed. The genes XNR_4432 – XNR_4443 were individually inactivated in the BAC 3F18_int_KO_0 through RedET recombineering. For the sake of simplicity, the 12 genes XNR_4432 – XNR_4443 the 29 which were object of inactivation studies will be designated in the text according to their sequence number (1 to 12) (Figure 2.17). The internal parts of the genes 1 – 12 were first substituted with the ampicillin resistance marker which was later removed yielding the BACs 3F18_int_KO_1, 3F18_int_KO_2, 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8, 3F18_int_KO_9, 3F18_int_KO_10, 3F18_int_KO_11 and 3F18_int_KO_12. To avoid possible polar effect on downstream genes during inactivation of the genes, especially those situated in operons, the gene deletions were performed in the way that after marker excision the 5'- and 3'-terminal regions of the deleted gene, both encoding 16 amino acid residues, were fused together retaining original reading frame. No additional stop codons were introduced between the fused regions.

The constructed BACs together with the BAC 3F18_int_KO_0 were individually introduced into the *S. albus* Del15 strain with the deleted mansouramycin pathway by conjugation. The ability of the obtained exconjugant strains to produce mansouramycins was assayed by cultivating the strains in the production medium DNPM, extracting their culture broth and analyzing the extracts by HPLC-MS. From the 12 analyzed genes, only inactivation of genes 3, 5, 6, 7, 8 and 9 impaired production of mansouramycins in *S. albus* strain (Table 2.7). These genes encode two putative tryptophan 2,3-dioxygenases, putative FAD-dependent oxidoreductase, branched amino acid aminotransferase, kynurenine formamidase and methyltransferase. The crucial role of the putative tryptophan 2,3-dioxygenase genes 3 and 5 in the biosynthesis of mansouramycin is expected from the previous results which have demonstrated that L-tryptophan serves as a main biosynthetic precursor for mansouramycin D (Figure 2.16). The essential role of methyltransferase gene 9 in biosynthesis of mansouramycins is also expected, as the carbon atom of the methylamino group of

Table 2.7. Effect of gene inactivation on mansouramycin production in *S. albus* Del15 strain

Inactivated gene		Putative function of the inactivated gene	Production of mansouramycin	
			A	D
1	XNR_4432	Hypothetical protein	+	+
2	XNR_4433	Biotin carboxylase / ATP-grasp domain protein	+	+
3	XNR_4434	Tryptophan 2,3-dioxygenase	–	–
4	XNR_4435	Aminotransferase / Kynureninase	+	+
5	XNR_4436	Tryptophan 2,3-dioxygenase	–	–
6	XNR_4437	FAD-dependent oxidoreductase / Pentachlorphenol monooxygenase	–	–
7	XNR_4438	Branched amino acid aminotransferase / Transaminase	–	–
8	XNR_4439	Cyclase / Kynurenine formamidase	–	–
9	XNR_4440	Methyltransferase	–	–
10	XNR_4441	Pyridoxamine 5'-phosphate oxidase-related protein	+	+
11	XNR_4442	Flavin reductase	+	+
12	XNR_4443	Phosphoribosylglycinamide synthetase	+	+

mansouramycin D is the only carbon which is not derived from L-tryptophan (Figure 2.16) and is likely introduced through methyltransferase reaction. Taking into account that the genes involved in the same metabolic pathway are often clustered together in prokaryotes, the genes between the gene 3 encoding tryptophan 2,3-dioxygenase and the gene 9 encoding methyltransferase were supposed to be essential for mansouramycin biosynthesis. However inactivation of the gene 4 encoding putative aminotransferase / kynureninase did not affect mansouramycin production in the *S. albus* Del15 strain. This implies that either the gene 4 is not involved in mansouramycin production or is crosscomplemented by the other gene in the genome of *S. albus* Del15.

To find out if the gene 4 encoding putative aminotransferase / kynureninase is not involved in mansouramycin biosynthesis or is crosscomplemented by another gene in *S. albus* genome a sequence analysis of *S. albus* genome was performed. No other gene with high sequence homology on protein level to gene 4 could be identified in the genome of *S. albus*. The kynureninase is an enzyme taking part in catabolism of L-tryptophan. In the section 3.8 was mentioned that the conservative tryptophan catabolism pathway is present in the *S. albus* strain. The genes XNR_3202, XNR_3201 and XNR_3200 encode a kynurenine formamidase, a kynureninase, and a tryptophan 2,3-dioxygenase. The product of the gene XNR_3201 shows no sequence homology to the product of the gene 4, despite both encode according to annotation a putative kynureninase. In order to check if the gene XNR_3201 can crosscomplement the effect of gene 4 deletion the whole tryptophane degradation pathway (the genes XNR_3202, XNR_3201 and XNR_3200) was deleted in the *S. albus* Del15 strain. For this purpose, a BAC covering the region to be deleted was selected from the genomic library of *S. albus* strain. The genes were substituted with the apramycin resistance marker using RedET. The constructed BAC was used for gene inactivation in the *S. albus* Del15 strain. The resistance marker was excised from the obtained knockout strain using IMES. The constructed strain was named *S. albus* degrW. The constructed strain *S. albus* degrW carries both deletion of the mansouramycin biosynthetic pathway and of the tryptophan degradation pathway. The BAC 3F18_int_KO_0 which carries the intact mansouramycin pathway as well as its derivatives 3F18_int_KO_1, 3F18_int_KO_2, 3F18_int_KO_4, 3F18_int_KO_10 and 3F18_int_KO_12 with the inactivated genes 1, 2, 4, 10 and 12 were transferred into *S. albus* degrW. The obtained exconjugants strains were assayed for their ability to produce mansouramycins. The strains were cultivated in production medium, culture supernatant was extracted and HPLC-MS analysis of the extracts was performed. Production of mansouramycin could be readily observed in the exconjugant strain with the BAC 3F18_int_KO_0 harboring the intact mansouramycin cluster. Production of mansouramacins was also not affected in the strains with the BACs with the inactivated genes 1, 2, 10 and 12 (Table 2.8). No production of mansouramycins could be detected by the strain harboring the BAC BAC 3F18_int_KO_4 with the inactivated gene 4, which encodes putative aminotransferase / kynureninase (Table 2.8). This result indicates that the gene 4 is

essential for mansouramycin production; however its knockout can be crosscomplemented by one of the genes of the tryptophan degradation pathway.

Table 2.8. Effect of gene inactivation on mansouramycin production in *S. albus* degrW strain

Inactivated gene		Putative function of the inactivated gene	Production of mansouramycin	
			A	D
1	XNR_4432	Hypothetical protein	+	+
2	XNR_4433	Biotin carboxylase / ATP-grasp domain protein	+	+
4	XNR_4435	Aminotransferase / Kynureninase	–	–
10	XNR_4441	Pyridoxamine 5'-phosphate oxidase-related protein	+	+
12	XNR_4443	Phosphoribosylglycinamide synthetase	+	+

The inactivation of the genes 1, 2, 10, 11 and 12 did not affect mansouramycin production in *S. albus* strains. The effect of these gene knockouts on mansouramycins was checked in the heterologous host LV1-652. The BAC 3F18_int_KO_0 which carries the intact mansouramycin pathway as well as its derivatives 3F18_int_KO_1, BAC 3F18_int_KO_2, BAC 3F18_int_KO_10 and BAC 3F18_int_KO_12 with the inactivated genes 1, 2, 10 and 12 were introduced into LV1-652. The obtained exconjugants strains were tested for their ability to produce mansouramycins. The results of the HPLC-MS analysis were identical to those obtained with the *S. albus* strains. The inactivation of the genes 1, 2, 10 and 12 did not lead to cessation of mansouramycin biosynthesis (Table 2.9). However it is worth mentioning that the reduction of the mansouramycins production level was observed after the inactivation of genes 1, 2, 11 and 12. Our results do not allow excluding the possibility that some of these genes take part in mansouramycin biosynthesis and that their inactivation is masked by crosscomplementation effect of the other enzyme in the primary metabolism.

Table 2.9. Effect of gene inactivation on mansouramycin production in the LV1-652 strain

Inactivated gene		Putative function of the inactivated gene	Production of mansouramycin	
			A	D
1	XNR_4432	Hypothetical protein	+	+
2	XNR_4433	Biotin carboxylase / ATP-grasp domain protein	+	+
10	XNR_4441	Pyridoxamine 5'-phosphate oxidase-related protein	+	+
11	XNR_4442	Flavin reductase	+	+
12	XNR_4443	Phosphoribosylglycinamide synthetase	+	+

The inactivation of the genes 3 – 9 abolished mansouramycin production in both natural producer *S. albus* and in heterologous host LV1-652. Despite every care was taken to exclude any polar effect on downstream genes while performing gene inactivation, to exclude the possibility of such polar effect, complementation studies have to be performed. To complement deletions of the genes 3 – 9 a set of complementation plasmids has been constructed. The phage BT1-based integrative vector pRT801 was used for this purpose. The strong synthetic promoter TS81 [24] together with the chloramphenicol resistance gene was inserted into the polylinker of pRT801 yielding pRT801_cat. The apramycin resistance gene in the backbone of pRT801_cat was

substituted with ampicillin-erythromycin resistance cassette through RedET. The constructed vector was called pRT801_cat_ampery. This vector was used for cloning of the individual mansouramycin biosynthetic genes for complementation studies. For this purpose the genes 3 – 9 were PCR amplified and cloned into the pRT801_cat_ampery on the place of chloramphenicol resistance gene yielding pRT801_proM3, pRT801_proM4, pRT801_proM5, pRT801_proM6, pRT801_proM7, pRT801_proM8 and pRT801_proM9. The cloned genes in the constructed vectors were placed under transcriptional control of the strong synthetic promoter TS81. The constructed plasmids pRT801_proM3, pRT801_proM4, pRT801_proM5, pRT801_proM6, pRT801_proM7, pRT801_proM8 and pRT801_proM9 were introduced into *S. albus* degrW strain harboring following BACs 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8 and 3F18_int_KO_9 respectively. The strain *S. albus* degrW is a derivative of *S. albus* Del14 strain with the deletions of mansouramycin cluster and the tryptophan degradation pathway. Mansouramycin production of the obtained strains was analyzed. In all cases, the introduced complementation construct successfully complemented respective gene deletion. This result indicates that during the inactivation of the mansouramycin biosynthetic genes, no polar effect was generated. The genes 3 – 9 are essential for mansouramycin production.

3.10 Determination of sequence of mansouramycin biosynthetic reactions through crosscomplementation studies

The results of gene inactivation studies unambiguously demonstrated the involvement of the genes 3 – 9 in the biosynthesis of mansouramycins A and D. From the annotation of these genes it is not obvious how the biosynthesis of mansouramycin functions. To determine the order in which the products of the genes 3 – 9 act during biosynthesis of mansouramycins the crosscomplementation assay was performed. It was assumed that after inactivation of a particular biosynthetic gene the obtained mutant will accumulate the biosynthetic precursor which is a substrate of the missing enzyme (Figure 2.20 A). This biosynthetic intermediate can restore production of the final product when provided to the mutants with the biosynthetic defects on the earlier steps and cannot restore the production of the final compound when fed to the mutants with the defects on the later biosynthetic steps. The crosscomplementation assay was performed on plates with the production medium DNPM-agar. The *S. albus* degrW strains harboring following BACs 3F18_int_KO_3, 3F18_int_KO_4, 3F18_int_KO_5, 3F18_int_KO_6, 3F18_int_KO_7, 3F18_int_KO_8 and 3F18_int_KO_9 were used. Two different mutants were simultaneously streaked on a single plate as a lines under 90 degree angle without intimate contact between the cultures (Figure 2.20 B). The intermediate exchange and chemical crosscomplementation was taking place at the region where the analyzed strains were growing close to each other. In order to determine in which of the strains the mansouramycin production was restored, the agar blocks

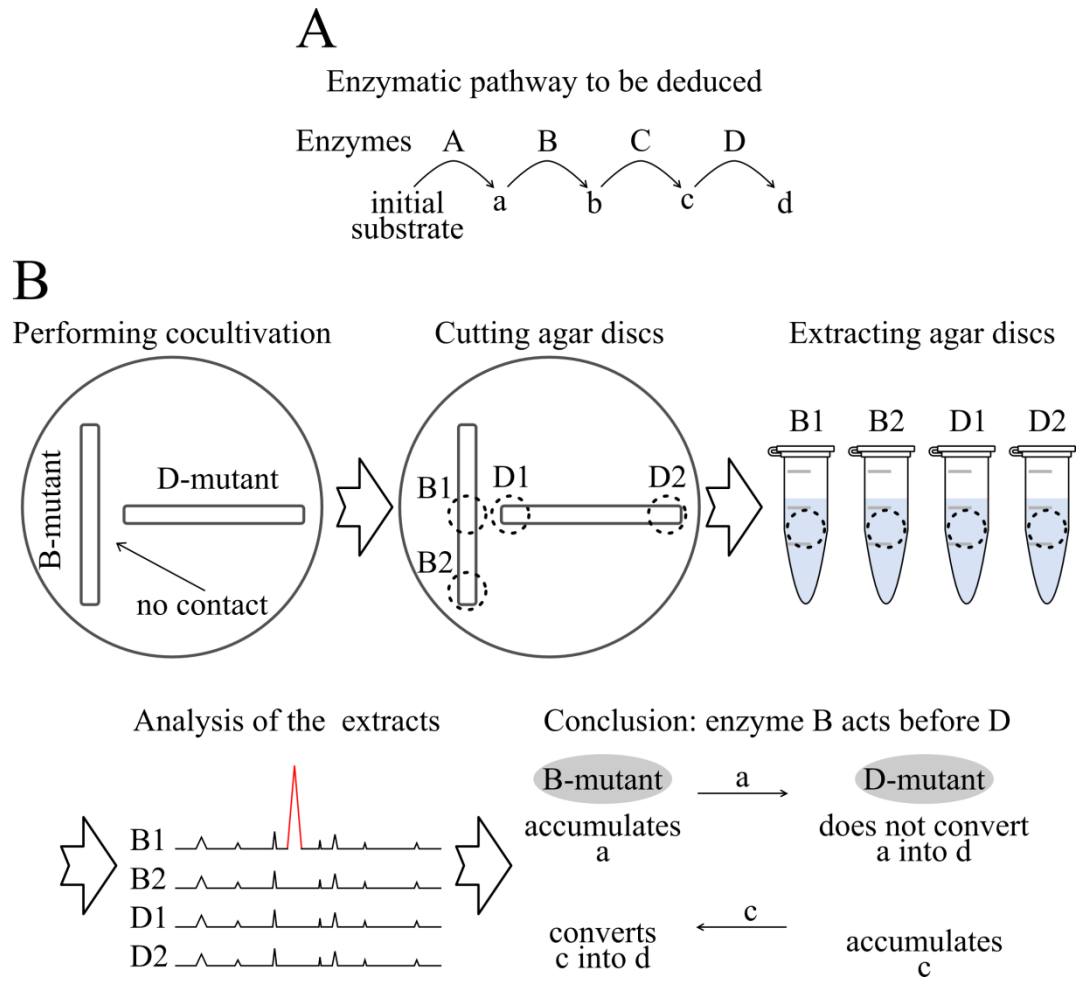

Figure 2.20. Scheme of the performed cocultivation studies to deduce the sequential order of enzymatic reactions in mansouramycin biosynthesis. A – schematic pathway consisting of four enzymatic steps. The enzymes A, B, C, and D catalyze conversion of the intitial substrate into the final product d through biosynthetic intermediates a, b, and c. B – experimatal procedure to determine the crosscomplementation pairs of mutants.

carrying the mycelium were cut from this region (agar blocks B1 and D1 in Figure 2.20 B). As a control agar blocks were cut from the distantly located regions where no crosscomplementation could occur between the strains (agar blocks B2 and D2 in Figure 2.20 B). The agar blocks were extracted with 500 µl ethylacetate and the extracts were analyzed by HPLC-MS. The strain with the restored mansouramycin production was regarded to be defective on an earlier biosynthetic step than the strain which did not produce mansouramycin during cocultivation on a plate. In this manner, the *S. albus* degrW strains harboring mansouramycin cluster with the inactivated genes 3 – 9 were analyzed in all possible combinations (21 pairs). The results of this analysis are summarized in the table 2.10.

The results of the crosscomplementation study indicate that the many of the analyzed mutants produce diffusible mansouramycin intermediates which can be converted into the final compound. No mutant was detected to to restore the mansouramycin production in the *S. albus* degrW strain harboring inactivation of the gene 5 (Figure 2.21). This implies that the enzyme encoded by the gene 5 catalyze the first step in the mansouramycin biosynthesis. The strains with the inactivated mansouramycin biosynthetic genes 4, 6, 7, and 8 restored mansouramycin production in the strain

Table 2.10. Results of the crosscomplementation studies with *S. albus* degrW strains harboring mansouramycin cluster with the inactivated genes 3 – 9

Analyzed mutants	Mutant with the restored production	Conclusion
3 – 4	3 (very low level)	3 → 4
3 – 5	5 (very low level)	5 → 3
3 – 6	3	3 → 6
3 – 7	3 (very low level)	3 → 7
3 – 8	3	3 → 8
3 – 9	no complementation	–
4 – 5	5 (very high level)	5 → 4
4 – 6	4 (good level)	4 → 6
4 – 7	4 (very low level)	4 → 7
4 – 8	4 (good level)	4 → 8
4 – 9	4 (very low level)	4 → 9
5 – 6	5 (very high level)	5 → 6
5 – 7	5 (very low level)	5 → 7
5 – 8	5 (very high level)	5 → 8
5 – 9	5 (very low level)	5 → 9
6 – 7	no complementation	–
6 – 8	8 (low level)	8 → 6
6 – 9	no complementation	–
7 – 8	8 (low level)	8 → 7
7 – 9	no complementation	–
8 – 9	8(low level)	8 → 9

with the inactivation of the gene 3 (Figure 2.21). This implies that the product of the gene 3 acts in the biosynthetic pathway prior the products of the genes 4, 6, 7, and 8. The mutant strain with the inactivated gene 3 restored mansouramycin production in the strain 5 what means that the product of 3 acts after the product of the gene 5 in the pathway. The product of the gene 4 acts after the product of 3 but before the products of the genes 6, 7, and 8 since the latter restore production of mansouramycin in the mutant with the inactivated gene 4 (Figure 2.21). The product of the gene 8 is likely to act after the product of the gene 4 but before the enzymes encoded by the genes 6 and 7 (Figure 2.21). None of the mutants restored the mansouramycin production in the mutants harboring inactivation of the genes 6 and 7. Nevertheless we suppose that the enzyme encoded by 7 acts after the product of 6. According to obtained results, the mutant 6 accumulates large ammounts of mansouramycin biosynthetic intermediates which are readily converted to mansouramycin by other strains (mutants 3, 4, and 5) (Figure 2.21). However, the mutant 6 did not restore the production of mansouramycin in the mutant 7 implying that the enzyme encoded by 7 acts after the product of 6. The mutant 7 also did not restore the mansouramycin production in the mutant 6. Generally the mutant with the inactivated gene 7 has proven itself as problematic in crosscomplementation studies. The mutant with the inactivated gene 7 restored the production only in the mutant 8, and with very low efficiency. This implies possible issues with the mansouramycin intermediate acuumulating after the inactivation of the gene 7. The intermediate might be unstable or it penetration in the cells is not effective. The participation at the late biosynthetic steps and issues with the accumulated intermediate can explain why corssomplenmentation studies with the mutant 7 were not effective.

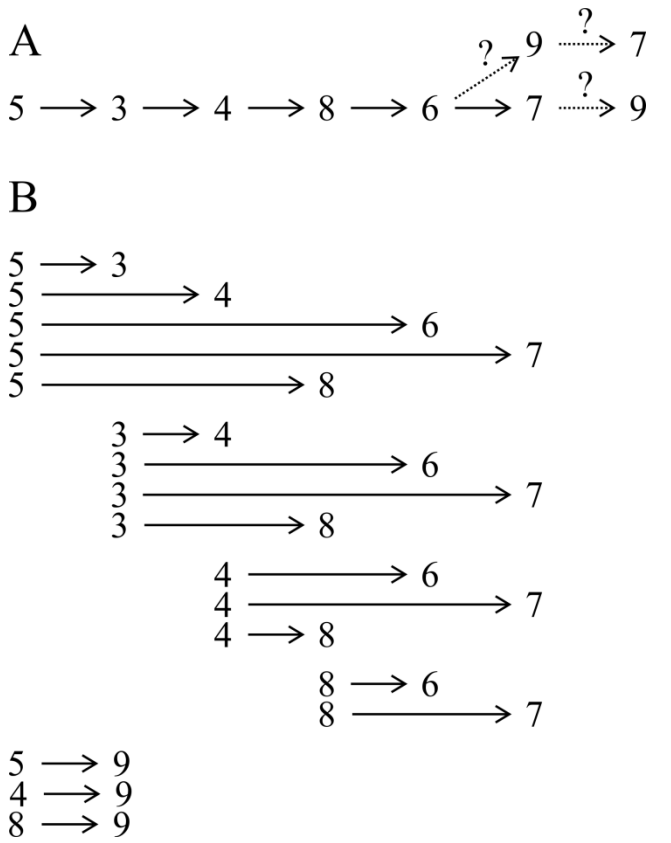

Figure 2.21. Determination of the order in which the enzymes act in mansouramycin biosynthesis. A – the participation order of enzymes 3 – 9 in mansouramycin biosynthesis. B – individual crosscomplememtation groups determined in cocultivation studies.

The exact determination of the position at which the enzyme encoded by the gene 9 acts in the mansouramycin biosynthesis is difficult. The mutant 9 restored mansouramycin production in the mutants 5, 4, and 8 but with low efficiency (Figure 2.21). For the same reasons like in the case of gene 7 we assume that the product of 9 acts after the product of 6. From the available amount of data, it is impossible to deduce in which order the enzymes 9 or 7 act in the mansouramycin biosynthesis. The mutants 9 and 7 do not crosscomplement each other. Low-efficiency restoration of mansouramycin production in the mutant 5, 4, and 8 does not allow assuming that the mutant 9 accumulates high levels of convertible intermediates which could be converted by the mutant 7 in case it the enzyme 7 acts before enzyme 9.

3.11 Identification, isolation and structure elucidation of mansouramycin biosynthetic intermediates

Successful crosscomplementation indicates that at least some of the mansouramycin mutants strains accumulate biosynthetic precursors that can be converted into mansouramycins. According to available data (Table 2.10), the strains with the disrupted genes 6 and 8 effectively restored the mansouramycin production in other biosynthetic mutants. This indicates indirectly that these strains might accumulate large quantities of mansouramycin biosynthetic intermediates. HPLC-MS chromatograms of the extracts from the *S. albus* degrW strains harboring mansouramycin cluster

with the inactivated genes 3 – 9 were analyzed for the presence of the peaks which might correspond to mansouramycin precursors. As a control strain the *S. albus* degrW 3F18_int_KO_0 strain with the intact mansouramycin pathway was used. In the extracts originating from the strains with the inactivated genes 3, 6, and 8, several peaks of reasonable intensity were identified which were absent in the extracts of the control strain. The peaks with the monoisotopic masses of 147.9 Da $[M+H]^+$ and 163.9 Da $[M+H]^+$ and the retention times of 1.0 min and 2.4 min were observed in the extracts of the strains with the inactivated genes 6 and 8 (Figure 2.22 A and D). The production

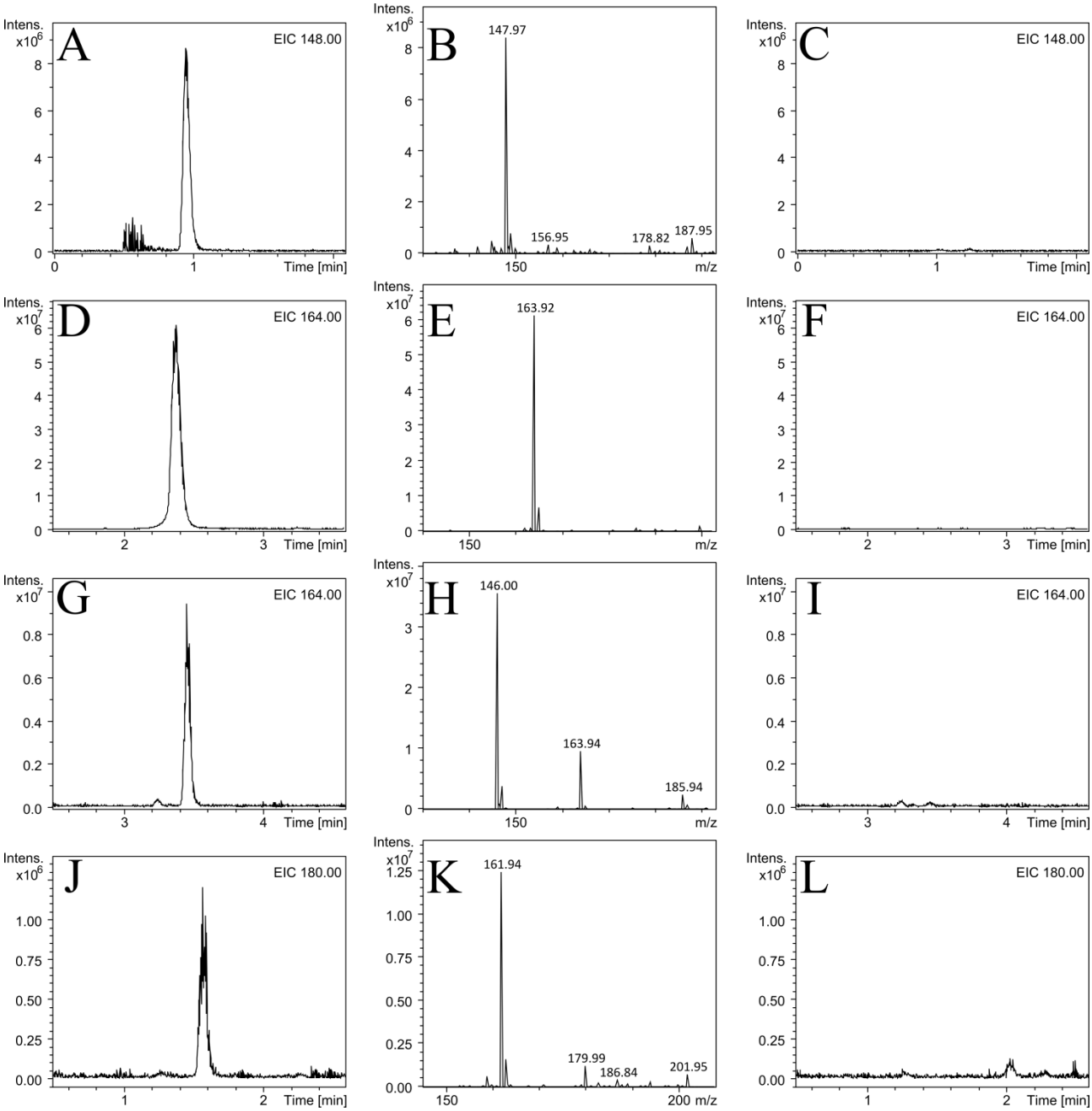

Figure 2.22. Identification of mansouramycin biosynthetic intermediates. A and C – extracted ion chromatogram (148 ± 0.5 Da) of the extracts of the *S. albus degrW* strains harboring the BACs 3F18_int_KO_8 and 3F18_int_KO_0 respectively. B – mass spectrum of the peak shown in B. D and F – extracted ion chromatogram (164 ± 0.5 Da) of the extracts of the *S. albus degrW* strains harboring the BACs 3F18_int_KO_8 and 3F18_int_KO_0 respectively. E – mass spectrum of the peak shown in D. G and I – extracted ion chromatogram (164 ± 0.5 Da) of the extracts of the *S. albus degrW* strains harboring the BACs 3F18_int_KO_3 and 3F18_int_KO_0 respectively. H – mass spectrum of the peak shown in G. J and L – extracted ion chromatogram (180 ± 0.5 Da) of the extracts of the *S. albus degrW* strains harboring the BACs 3F18_int_KO_6 and 3F18_int_KO_0 respectively. K – mass spectrum of the peak shown in J.

of both compounds was much higher in the strain with the gene 8 inactivation than in the strain with the gene 6 inactivation. The compound with the monoisotopic mass of 163.9 Da $[M+H]^+$ and the retention time of 3.5 min was detected in the extract of the strain with the inactivated gene 3 (Figure 2.22 G). Low amounts of this compound could also be observed in the extract of the mutant with the inactivated gene 6. The peak with the monoisotopic mass of 180.0 Da $[M+H]^+$ and the retention tine of 1.6 min was detected only in the extracts of the strain with the inactivated gene 6 (Figure 2.22 J).

The identified compounds were extracted from the culture broth of the mutant strains which showed the highest production yields and purified for structure elucidation by NMR. The structure elucidation of the isolated compounds was performed by Dr. Constanze Paulus.

The compound with the m/z of 163.9 Da (Figure 2.22 D) isolated from the culture of *S. albus degrW* 3F18_int_KO_8 was elucidated as 6-hydroxyindoline-2,3-dione (Exact mass 163.0269 Da) (Figure 2.23 1). Analysis of ^{1}H, HSQC and HMBC NMR data revealed high similarity with the

Table 2.11. NMR data (500 MHz, dmso-d_6) for 6-hydroxyindoline-2,3-dione

Position	δ_H, mult (*J*, Hz)	δ_C	HMBC
1	-	170.8	-
2	-	181.6	-
2a	-	112.7	-
3	7.34, d (8.3)	127.6	5-2-1
4	6.31, dd (8.3, 2.2)	110.5	6
5	-	155.4	-
6	6.18, d (2.2)	99.2	5-2
6a	-	143.9	-
7	NH - 10.72, s	-	-

^{13}C data were taken from HMBC

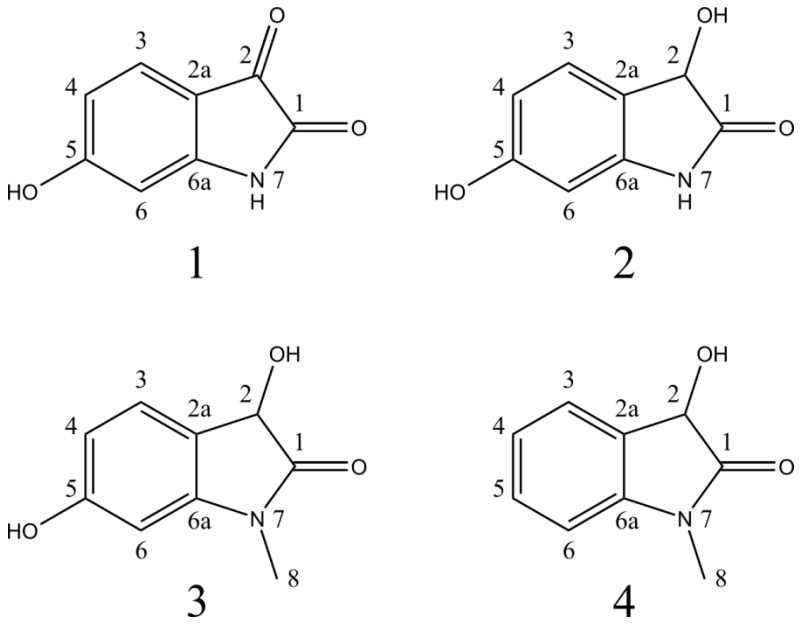

Figure 2.23. Structures of putative mansouramycin biosynthetic intermediates. 1 – 6-hydroxyindoline-2,3-dione isolated from *S. albus degrW* 3F18_int_KO_8 (Exact mass 163.0269 Da). 2 – 3,6-dihydroxyindolin-2-one isolated from *S. albus degrW* 3F18_int_KO_8 (Exact mass 165.0426 Da). 3 – 3,6-dihydroxy-1-methylindolin-2-one isolated from *S. albus degrW* 3F18_int_KO_6 (Exact mass 179.0582 Da). 4 – 3-hydroxy-1-methylindolin-2-one isolated from *S. albus degrW* 3F18_int_KO_5 (Exact mass 163.0633 Da).

known natural product isatin. However, the missing proton signal at C-5 and a chemical shift of δ_C 155.4 indicates substitution with a hydroxyl group at this position (Table 2.11).

The compound with the m/z of 147.9 Da (Figure 2.22 A) isolated from the culture of *S. albus degrW* 3F18_int_KO_8 was elucidated as 3,6-dihydroxyindoline-2-one (Exact mass 165.0426 Da) (Figure 2.23 2). The exact mass of the 3,6-dihydroxyindoline-2-one is 165.0426 Da and differs from the one determined by HPLC-MS analysis. The 18 Da diference can be explained by dehydratation during ionization processs. Analysis of ^{1}H, HSQC and HMBC NMR data resulted in the same indoline core structure as was found for 6-hydroxyindoline-2,3-dione. However, a proton signal at δ_H 4.67 and δ_C 69.3 suggests the presence of a hydroxyl group at position 2 instead of the former ketogroup (Table 2.12).

The compound with the m/z of 180.0 Da (Figure 2.22 J) isolated from the culture of *S. albus degrW* 3F18_int_KO_6 was elucidated as 3,6-dihydroxy-1-methylindolin-2-one (Exact mass 179.0582 Da) (Figure 2.23 3). The NMR spectroscopic data reveals high similarity with compound 3,6-dihydroxyindoline-2-one. Although the missing proton signal at δ_H 10.07 indicates a substitution at the nitrogen. A signal at δ_H 3.02 and δ_C 25.6 was found suggesting a methyl group bound to the nitrogen (Table 2.13).

Table 2.12. NMR data (500 MHz, dmso-d_6) for 3,6-dihydroxyindoline-2-one

Position	δ_H, mult (*J*, Hz)	δ_C	HMBC
1	-	178.9	-
2	4.68, s	69.1	4-2a-3-6a-1
2a	-	119.9	-
3	7.04, d (8.3)	126.2	2-6-6a-5
4	6.32, dd (7.9, 2.2)	108.3	6-2a-5
5	-	158.8	-
6	6.24, d (2.2)	98.1	4-2a-6a-5
6a	-	143.9	-
7	NH - 10.07, s	-	-

^{13}C data were taken from HMBC

Table 2.13. NMR data (500 MHz, dmso-d_6) for 3,6-dihydroxy-1-methylindolin-2-one

Position	δ_H, mult (*J*, Hz)	δ_C	HMBC
1	-	176.3	-
2	4.76, s	68.4	2a-3-6a-1
2a	-	118.3	-
3	7.09, dd (7.9, 0.9)	125.3	2-6-4-6a-5
4	6.41, dd (7.9, 2.1)	108.1	6-2a-5
5	OH – 9.65, s	158.7	-
6	6.36, d (2.1)	96.8	5-6a-2a-4
6a	-	145.0	-
7	-	-	-
8	3.02, s	25.6	6a-1

^{13}C data were taken from HMBC

The compound with the m/z of 163.9 Da (Figure 2.22 G) isolated from the culture of *S. albus degrW* 3F18_int_KO_3 was elucidated as 3-hydroxy-1-methylindolin-2-one (Exact mass 163.0633

Da) (Figure 2.23 4). Analysis of ^{1}H, HSQC and HMBC NMR data reveals that the compound is highly similar to 3,6-dihydroxy-1-methylindolin-2-one. However, an additional aromatic signal at δ_H 7.32 which shows HMBS correlations to C-2a and C-6a, indicates the missing hydroxyl group at C-5 (Table 2.14).

Table 2.14. NMR data (500 MHz, dmso-d_6) for 3-hydroxy-1-methylindolin-2-one

Position	δ_H, mult (*J*, Hz)	δ_C	HMBC
1	-	176.3	-
2	4.76, s	68.4	2a-3-6a-1
2a	-	118.3	-
3	7.09, dd (7.9, 0.9)	125.3	2-6-4-6a-5
4	6.41, dd (7.9, 2.1)	108.1	6-2a-5
5	OH – 9.65, s	158.7	-
6	6.36, d (2.1)	96.8	5-6a-2a-4
6a	-	145.0	-
7	-	-	-
8	3.02, s	25.6	6a-1

^{13}C data were taken from HMBC

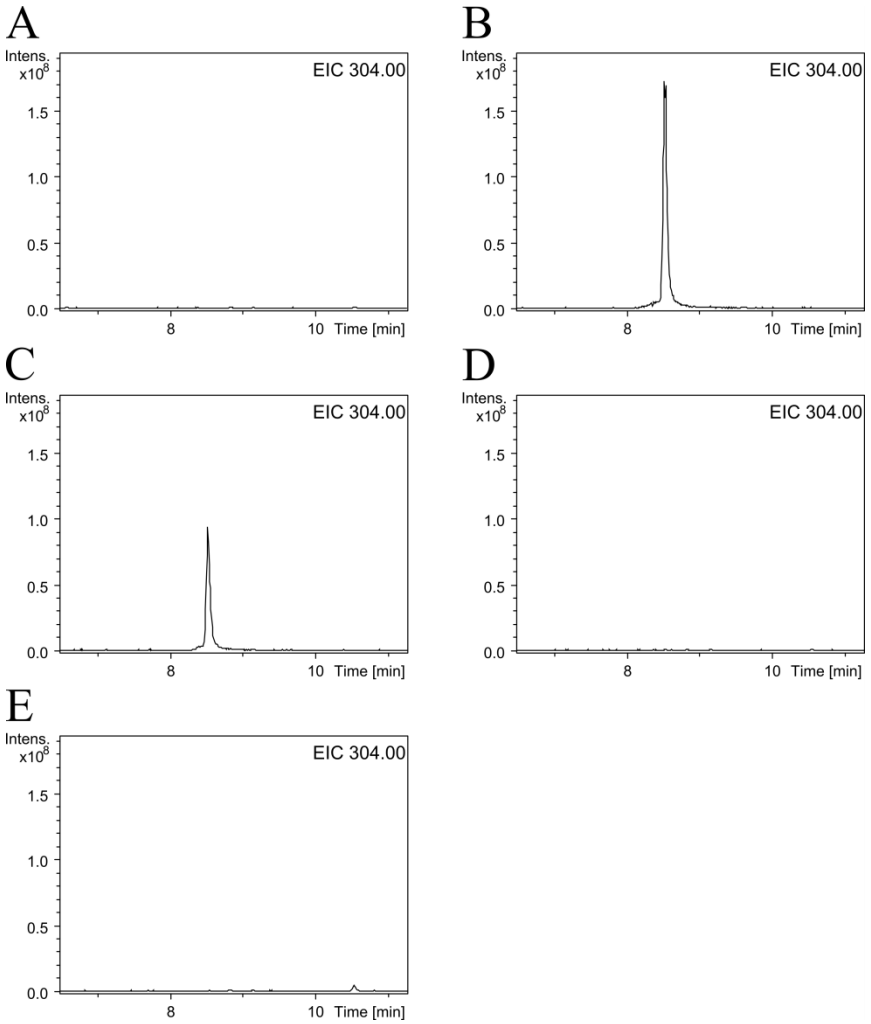

Figure 2.24. Feeding of putative mansouramycin intermediates to the strain *S. albus degrW* 3F18_int_KO_5 blocked on early stage of mansouramycin biosynthesis. Extracted ion chromatograms 304 ± 0.5 Da are shown. A – chromatogram of ethyl acetate extract of *S. albus degrW* 3F18_int_KO_5 culture. B – chromatogram of ethyl acetate extract of *S. albus degrW* 3F18_int_KO_5 culture fed with 3,6-dihydroxyindolin-2-one. The production of mansouramycin D can be seen. C – chromatogram of ethyl acetate extract of *S. albus degrW* 3F18_int_KO_5 culture fed with 6-hydroxyindoline-2,3-dione. The production of mansouramycin D can be seen. D – chromatogram of ethyl acetate extract of *S. albus degrW* 3F18_int_KO_5 culture fed with 3-hydroxy-1-methylindolin-2-one. No production of mansouramycin D can be seen. E – chromatogram of ethyl acetate extract of *S. albus degrW* 3F18_int_KO_5 culture fed with 3,6-dihydroxy-1-methylindolin-2-one. No production of mansouramycin D can be seen.

To prove that the isolated compounds are mansouramycin biosynthetic intermediates and not shunt products they were fed to the culture of the strain *S. albus degrW* 3F18_int_KO_5 which is blocked at early stage of mansouramycin biosynthesis (Figure 2.21). The feeding of 6-hydroxyindoline-2,3-dione and 2 – 3,6-dihydroxyindolin-2-one restored mansouramycin production in *S. albus degrW* 3F18_int_KO_5 (Figure 2.24). 3,6-dihydroxy-1-methylindolin-2-one and 3-hydroxy-1-methylindolin-2-one were not converted to mansouramycins by the strain (Figure 2.24). These results indicate that from all isolated compounds only 6-hydroxyindoline-2,3-dione and 2 – 3,6-dihydroxyindolin-2-one are mansouramycin biosynthetic intermediates. 3,6-dihydroxy-1-methylindolin-2-one and 3-hydroxy-1-methylindolin-2-one are most likely shunt products.

Successful conversion of 6-hydroxyindoline-2,3-dione and 2 – 3,6-dihydroxyindolin-2-one to mansouramycin D by the strain *S. albus degrW* 3F18_int_KO_5 proved the compounds to be mansouramycin biosynthetic intermediates. These compounds could be observed in the culture broth of the strains with inactivated genes 8 and 6. This implies that the enzymes encoded by the genes 5, 3 and 4 which act during the biosynthesis of mansouramycins before the enzymes encoded by the genes 8 and 6 (Figure 2.21) are responsible for the formation of 6-hydroxyindoline-2,3-dione and 2 – 3,6-dihydroxyindolin-2-one. To prove this, the feeding of these compounds should have been performed into all biosynthetic mutants. The amounts of isolated compounds were too low to perform additional feeding experiments. The 6-hydroxyindoline-2,3-dione was purchased from Enamine company. No supplier of 3,6-dihydroxyindolin-2-one was found. The purchased 6-hydroxyindoline-2,3-dione was fed to the cultures of *S. albus degrW* harboring mansouramycin cluster with the inactivated genes 3 – 9. As was expected, the feeding of the compound restored the production of mansouramycins only in the strains with the inactivated genes 5, 3, and 4 (Figure 2.25). This indicates that these genes encode enzymes responsible for the formation of the 6-hydroxyindoline-2,3-dione. The deficiency of these enzymes can be complemented by the external feeding of the compound for whose formation these enzymes are responsible. Feeding of 6-hydroxyindoline-2,3-dione did not restore mansouramycin production in *S. albus degrW* strains harboring inactivation of the genes 8, 6, 7 and 9 (Figure 2.25). This indicates that the genes 8, 6, 7 and 9 encode enzymes which convert 6-hydroxyindoline-2,3-dione into mansouramycin. The feeding of 6-hydroxyindoline-2,3-dione confirms the participation order of mansouramycin biosynthetic enzymes determined in crosscomplementation studies (Figure 2.21).

3.12 *In vitro* characterization of the putative kynurenine formamidase encoded by the gene 8

The results of isolation and feeding of putative mansouramycin biosynthetic intermediates indicate that the genes 5, 3 and 4 encode enzymes necessary for the formation of 6-hydroxyindoline-2,3-dione. Mansouramycin D is synthesized from the 6-hydroxyindoline-2,3-dione and L-tryptophane through the action of enzymes encoded by the genes 6, 7, 8 and 9. According to

the determined participation order of mansouramycin biosynthetic enzymes, the putative kynurenine formamidase encoded by the gene 8 catalyzes the first step in the conversion of 6-hydroxyindoline-2,3-dione to mansouramycin D. In order to characterize the product of the gene 8 biochemically it was overexpressed in *E. coli*. For this purpose, the open reading frame of the gene 8 was cloned into the expression vector pET-28 a (+) yielding pET-28-8. The constructed plasmid

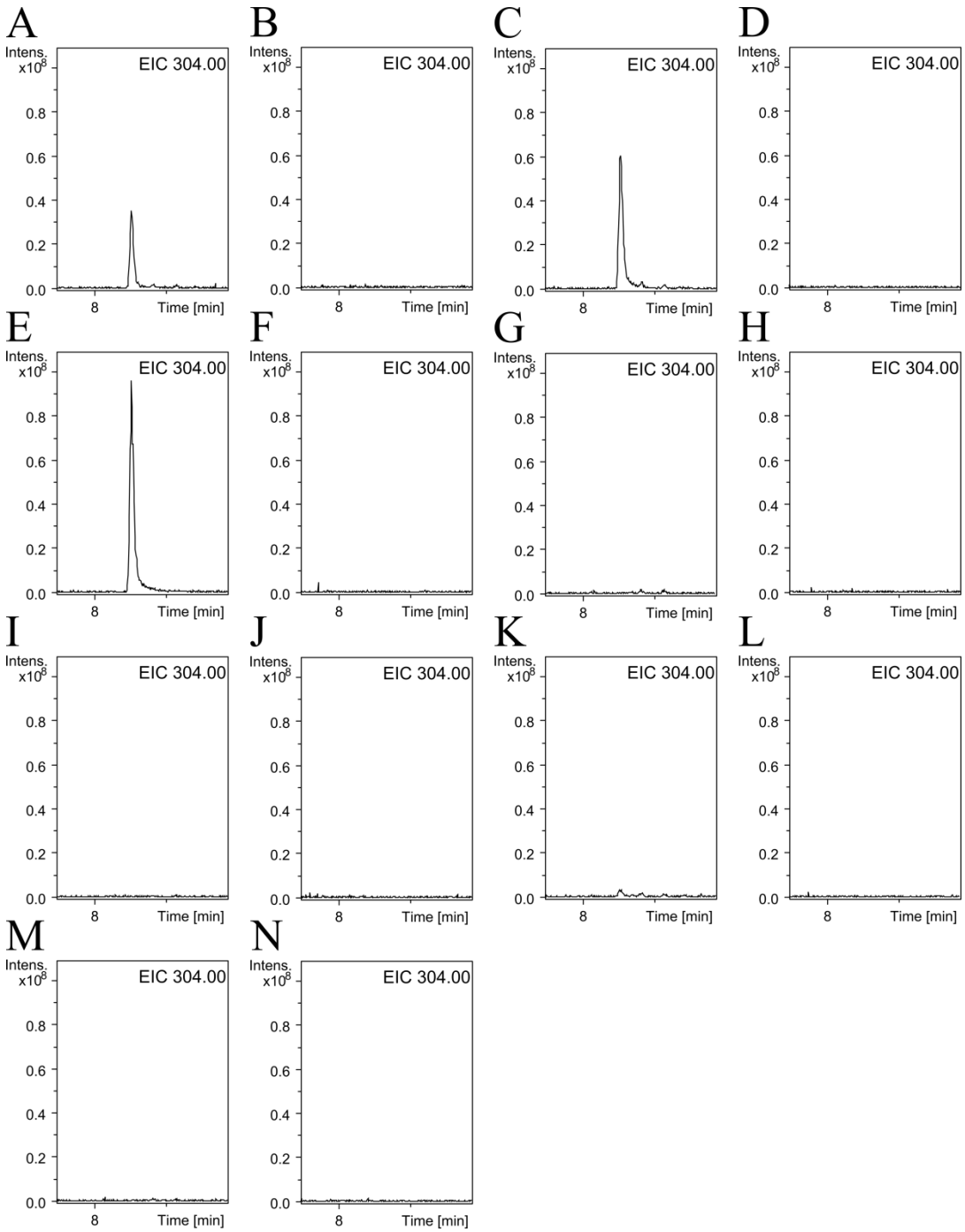

Figure 2.25. Feeding of 6-hydroxyindoline-2,3-dione to *S. albus degrW* harboring mansouramycin cluster with the inactivated genes 3 – 9. Extracted ion chromatograms 304 ± 0.5 Da are shown. A and B – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_3 culture after feeding and without feeding respectively. C and D – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_4 culture after feeding and without feeding respectively. E and F – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_5 culture after feeding and without feeding respectively. G and H – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_6 culture after feeding and without feeding respectively. I and J – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_7 culture after feeding and without feeding respectively. K and L – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_8 culture after feeding and without feeding respectively. M and N – chromatograms of ethyl acetate extracts of *S. albus degrW* 3F18_int_KO_9 culture after feeding and without feeding respectively.

contains a recombinant gene which encodes the studied the putative kynurenine formamidase N-terminally fused to a histidine tag. The pET-28-8 plasmid was transformed into the expression strain *E. coli* Rosetta 2 DE3. The expression of the protein was induced by adding IPTG. The analysis of the cell lysate using SDS-PAGE clearly demonstrated that the recombinant enzyme is overexpressed as a soluble protein in Rosetta 2 DE3 cells (Figure 2.26 A). The recombinant protein was purified using immobilized metal affinity chromatography and used for activity assay with 6-hydroxyindoline-2,3-dione (Figure 2.26 B). In the presence of the purified enzyme 6-hydroxyindoline-2,3-dione rapidly lost its typical orange color. LC-MS analysis of the reaction mixture revealed a new peak with the retention time of 0.7 min which absent samples containing reaction buffer and either the substrate or the enzyme (Figure 2.27 C). The structure of the reaction product was solved by NMR. For this purpose the reaction with the recombinant enzyme and 6-hydroxyindoline-2,3-dione was performed in preparative scale. The reaction mixture was lyophilized and the compounds were extracted with methanol. The extract was used directly for structure elucidation. NMR studies were performed by Marc Stierhof.

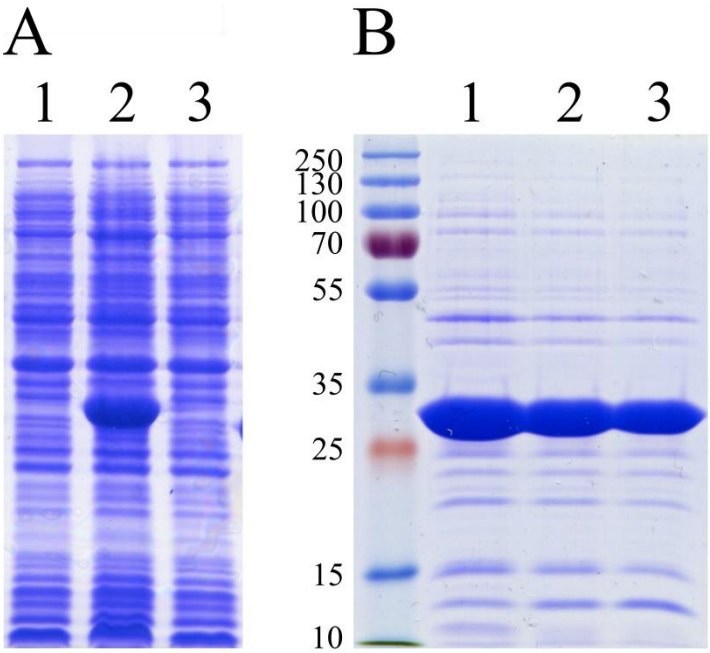

Figure 2.26. Expression of the putative kynurenine formamidase encoded by the gene 8 in *E. coli* Rosetta 2 DE3. A – SDS-PAGE analysis of soluble protein extracts of *E. coli* Rosetta 2 DE3 (1), *E. coli* Rosetta 2 DE3 pET-28-8 induced by adding IPTG (2), *E. coli* Rosetta 2 DE3 pET-28-8 without induction. Overexpression of the recombinant protein can be seen in the sample 2. B – SDS-PAGE analysis of protein fractions (1, 2, and 3) containing recombinant protein. The protein was purified by immobilized metal affinity chromatography. Molecular weight of proteins of the protein ladder in kDA is given as numbers.

Analysis of the reaction mixture revealed the presence of three compounds: 6-hydroxyindoline-2,3-dione, 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid and of 2-amino-4-hydroxybenzoic acid (Figures 2.28 – 2.32; Tables 2.15 – 2.17). 6-hydroxyindoline-2,3-dione, which was used as a substrate in the *in vitro* assay, and 2-amino-4-hydroxybenzoic acid were present in the mixture in small amounts. Analysis of ^{1}H and ^{13}C NMR revealed 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid (Figure 2.28 2, Table 2.16) as the main product of the enzyme reaction. The lactam

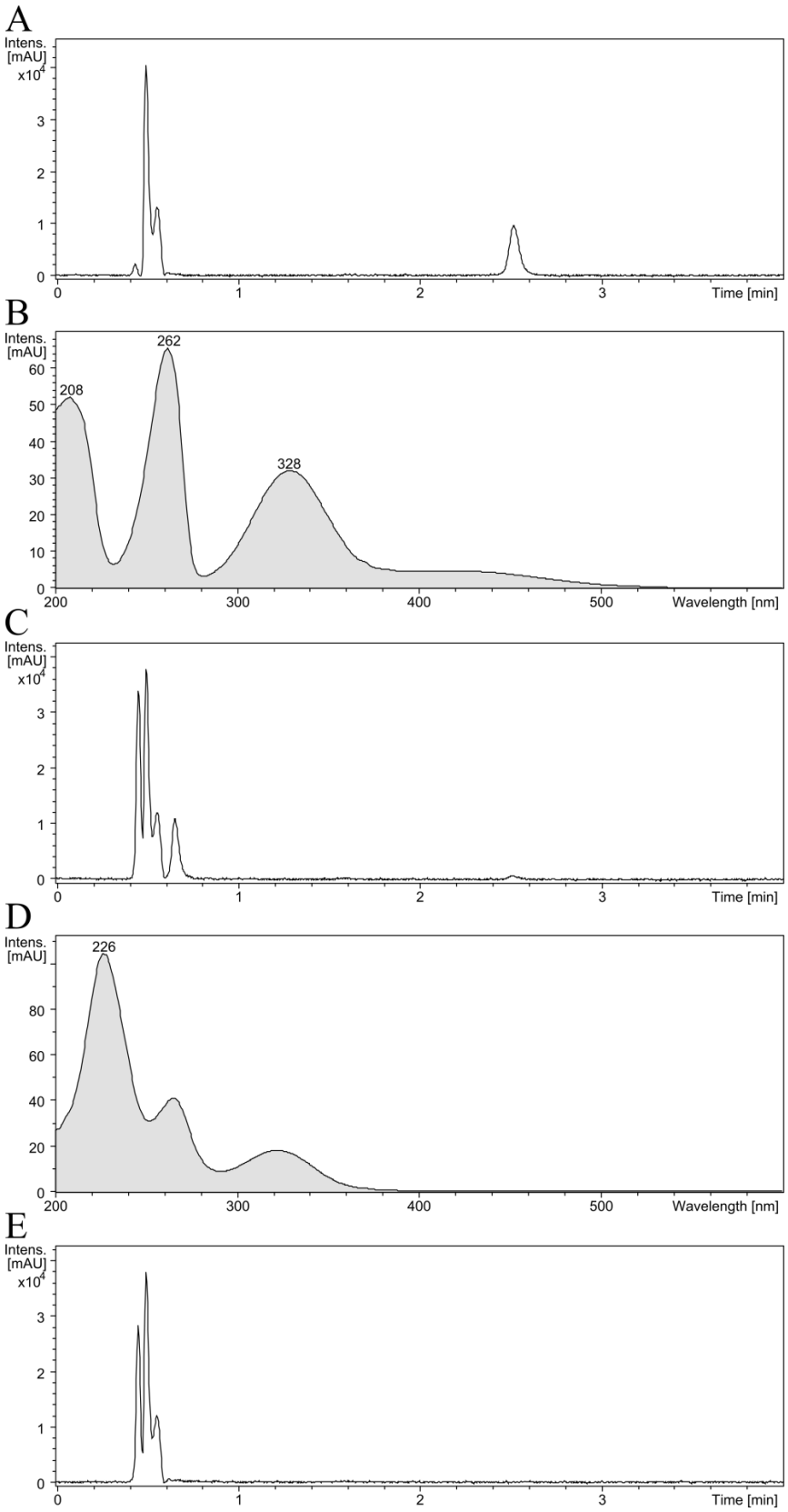

Figure 2.27. In vitro hydrolysis of 6-hydroxyindoline-2,3-dione catalyzed by the product of the gene 8. A – PDA chromatogram of 6-hydroxyindoline-2,3-dione in the reaction buffer. The peak with the RT of 2.5 min corresponds to the compound. B – absorption spectrum of 6-hydroxyindoline-2,3-dione. C – PDA chromatogram of the reaction products resulting from coincubatuon of 6-hydroxyindoline-2,3-dione with the product of the gene 8. The peak with the RT of 0.7 min corresponds to the product of the enzymatic reaction. D – absorption spectrum of the product of the enzymatic reaction. E – PDA chromatogram of the reaction buffer with added enzyme but without the substrate – 6-hydroxyindoline-2,3-dione.

ring opening was confirmed by N-HMBC showing the conversion of the amide nitrogen at δ 205 ppm into an amine at δ 145 ppm. 2-amino-4-hydroxybenzoic acid (Figure 2.28 3) present in minute amounts is regarded as a product of spontaneous decarboxylation of 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid.

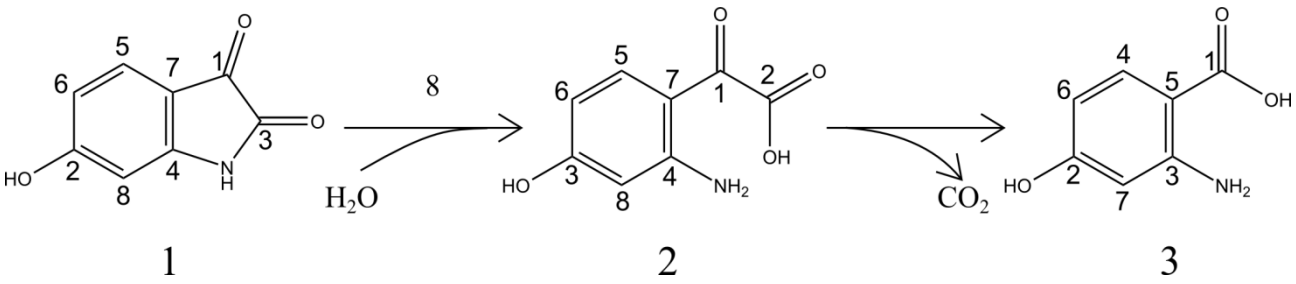

Figure 2.28. Conversion of 6-hydroxyindoline-2,3-dione (1) into 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid (2) catalyzed by the product of the gene 8. The product of reaction undergoes spontaneous decarboxylation to form 2-amino-4-hydroxybenzoic acid (3). Numbering of the atoms refers to the structure elucidation data.

Table 2.15. NMR data (500 MHz, dmso-d_6) for 3,6-dihydroxyindoline-2-one

Nr	δ(^{13}C) [ppm]	δ(^{1}H) [ppm], mult(J)
1	180,8	-
2	168,5	-
3	161,2	-
4	153,9	-
5	127,6	7.34, d (9.1)
6	110,7	6.43, ovl.
7	109,3	-
8	99,5	6.43, ovl.

Table 2.16. NMR data (500 MHz, dmso-d_6) for 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid

Nr	δ(^{13}C) [ppm]	δ(^{1}H) [ppm], mult(J)
1	196,8	-
2	170,5	-
3	162,7	-
4	154,3	-
5	136,1	7.33, d (8.7)
6	107,4	-
7	104,5	5.99, dd (8.7, 2.3)
8	100,2	6.08, d (2.3)

Table 2.17. NMR data (500 MHz, dmso-d_6) for 2-amino-4-hydroxybenzoic acid

Nr	δ(^{13}C) [ppm]	δ(^{1}H) [ppm], mult(J)
1	170,4	-
2	162,9	-
3	154,1	-
4	135,9	7.41, d (8.8)
5	108,2	-
6	103,5	6.07, ovl.
7	98,7	6.20, d (2.5)

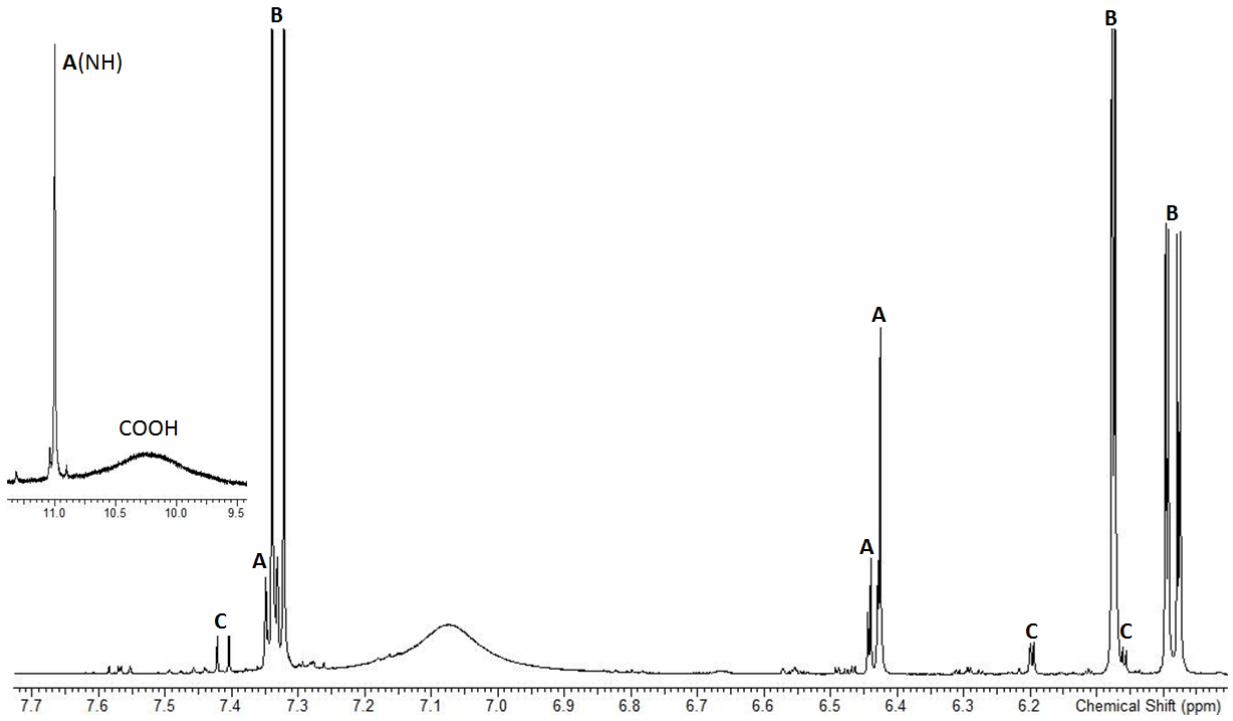

Figure 2.29. ^{1}H NMR spectrum of a mixture consisting of 6-hydroxyindoline-2,3-dione (A), 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid (B) and 2-amino-4-hydroxybenzoic acid (C).

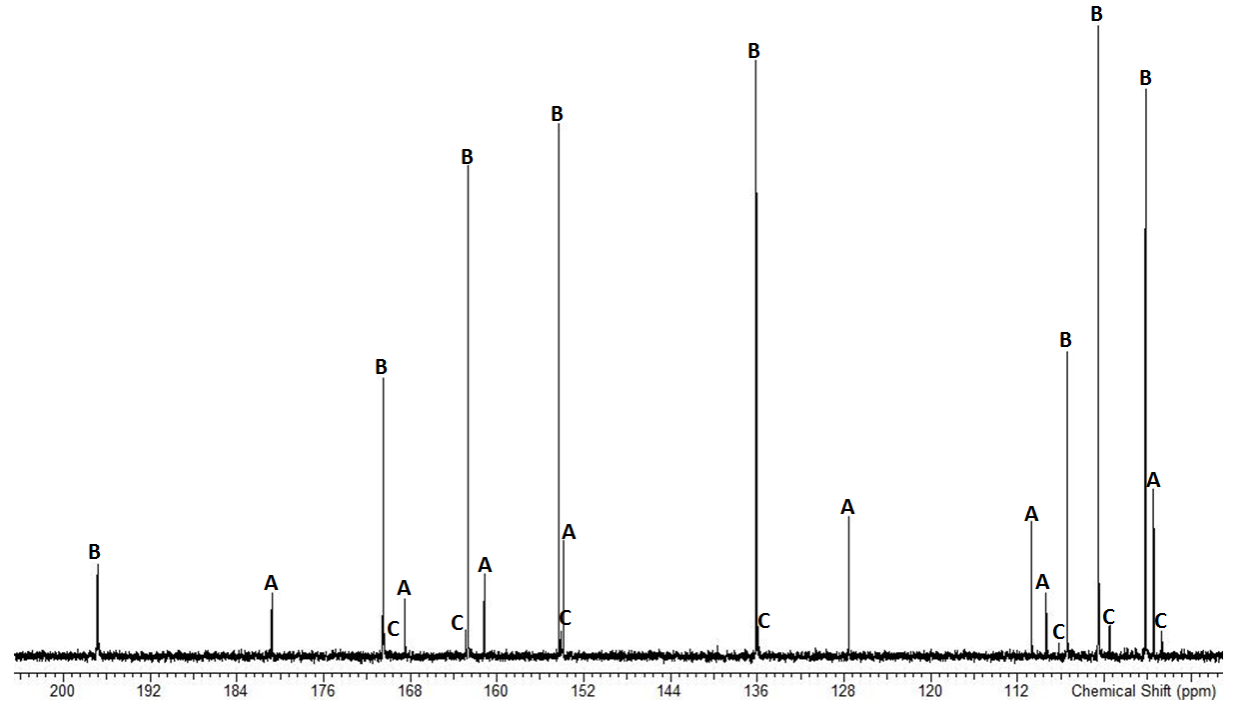

Figure 2.30. ^{13}C NMR spectrum of a mixture consisting of 6-hydroxyindoline-2,3-dione (A), 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid (B) and 2-amino-4-hydroxybenzoic acid (C).

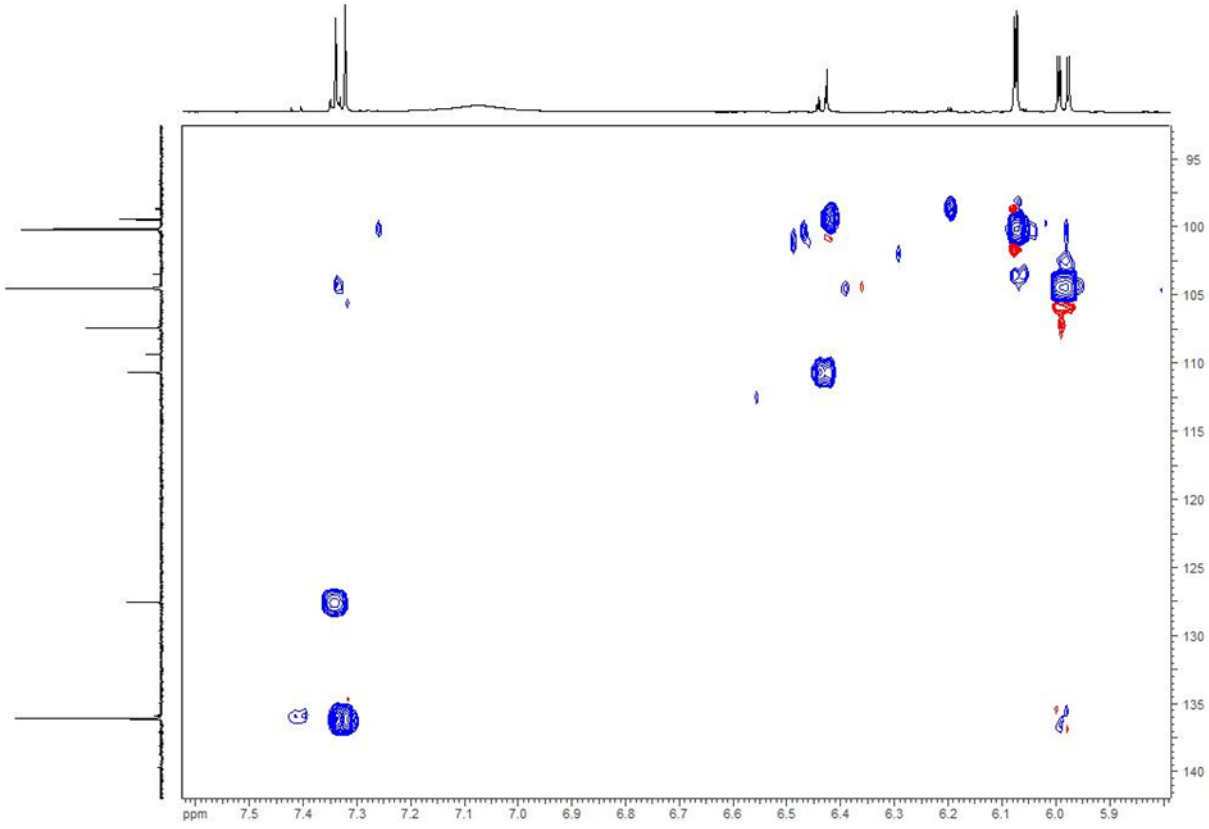

Figure 2.31. HSQC spectrum of a mixture consisting of 6-hydroxyindoline-2,3-dione, 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid and 2-amino-4-hydroxybenzoic acid.

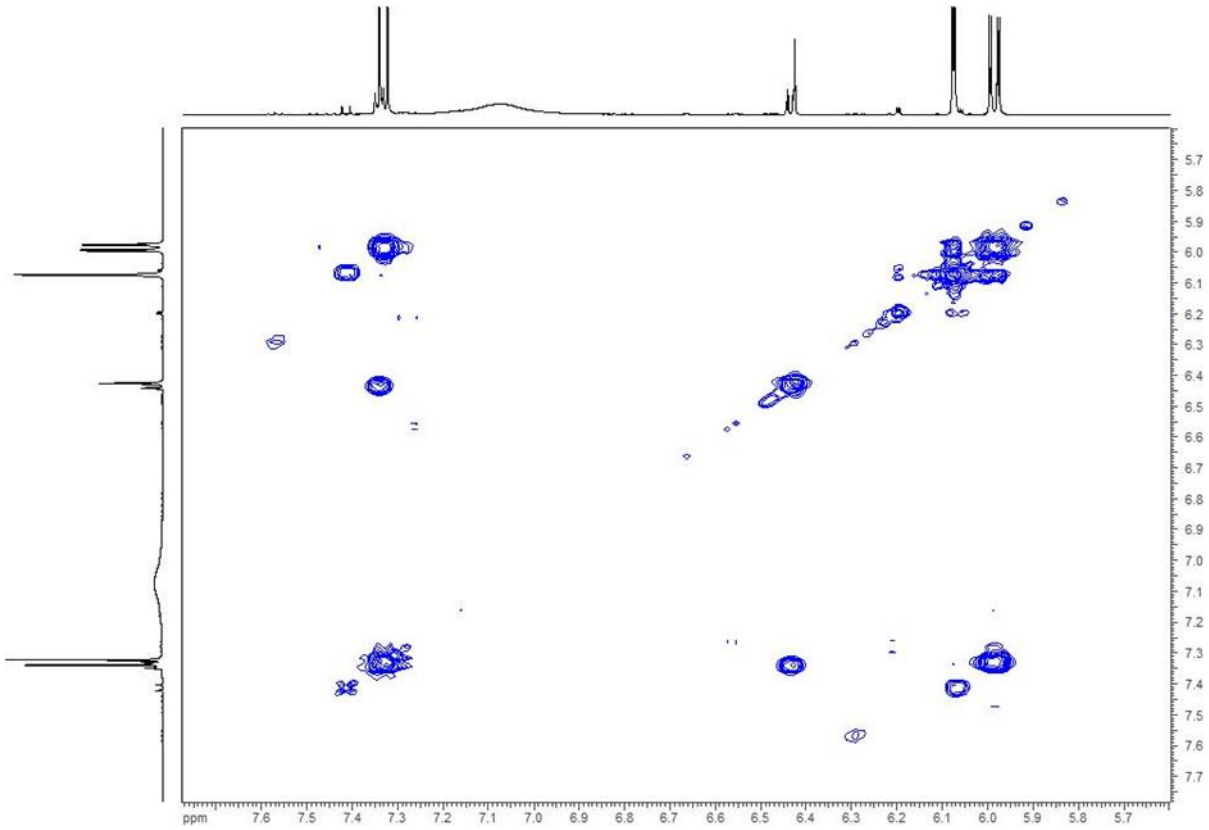

Figure 2.32. COSY spectrum of a mixture consisting of 6-hydroxyindoline-2,3-dione, 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid and 2-amino-4-hydroxybenzoic acid.

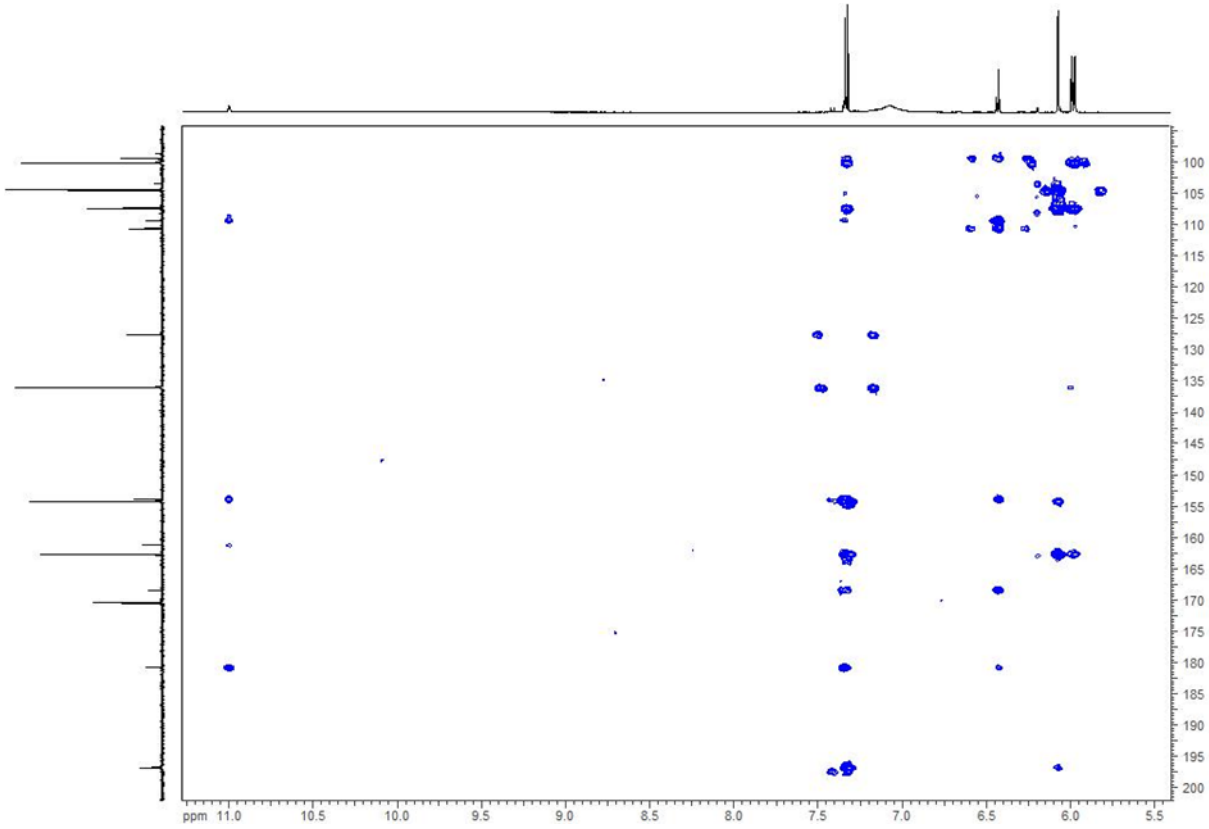

Figure 2.33. HMBC spectrum of a mixture consisting of 6-hydroxyindoline-2,3-dione, 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid and 2-amino-4-hydroxybenzoic acid.

3.13 Proposed scheme of mansouramycin D biosynthesis

The results of the feeding studies provided evidence that L-tryptophan is the main biosynthetic precursor of mansouramycin D. All carbon atoms of the isoquinoline and of the indole rings are inherited from tryptophan; only the carbon atom of the N-methyl group originates from the different precursor. The incorporation pattern of carbon atoms of tryptophan into mansouramycin D indicates that two tryptophan molecules are degraded to C7 and to C10 units which are then united in the structure of the mansouramycin D (Figure 2.16). During the formation of the C7 unit the side chain of the tryptophan is removed, the pyrrole ring is opened and one of the carbons originating from the pyrrole ring is removed. The C7 unit corresponds to a benzene ring with carbon and nitrogen substituents. Feeding of the ^{15}N15-labelled anthranilic acid indicates that the nitrogen atom of the pyrrole ring of tryptophan corresponds to the nitrogen atom of the methylated amino group at the position 7 of mansouramycin D (Figure 2.16). The carbon substituent of the benzene ring of C7 unit corresponds to one of the carbons of the pyridine ring in the mansouramycin D structure.

The results of feeding with L-tryptophan ^{13}C-labelled in the first position (carboxyl group) and with ^{15}N2-L-tryptophan indicate that the carboxyl group and the amino group of tryptophan are removed during the biosynthesis of the C10 unit (Figure 2.16). Furthermore, feeding with ^{13}C-labelled tryptophan in the second position demonstrated that the indole group of the C10 precursor is shifted from the carbon at the position 3 of the side chain to the carbon at the position 2. The two

carbon atoms of the side chain of the C10 unit are involved in the formation of the pyridine ring in the mansouramycin D structure (Figure 2.16).

Gene inactivation experiments revealed mansouramycin gene cluster consisting of 7 genes (named 3 - 9). The participation order of the genes in mansouramycin biosynthesis was determined by crosscomplementation studies as follows: 5→3→4→8→6→7→9. The sequence order of the last two biosynthetic steps is not reliable. Two mansouramycin biosynthetic intermediates were isolated from the *S. albus* mutants with the inactivation of individual mansouramycin biosynthetic genes: 6-hydroxyindoline-2,3-dione and 3,6-dihydroxyindolin-2-one. Analysis of secondary metabolite production profiles of individual biosynthetic mutants as well as feeding of isolated biosynthetic intermediates demonstrated that the genes 5, 3 and 4 encode the production of 6-hydroxyindoline-2,3-dione and 3,6-dihydroxyindolin-2-one. Obviously the isolated mansouramycin biosynthetic intermediates are the precursors of the C7 unit. This was further confirmed in the *in vitro* assay with the product of the gene 8. The enzyme annotated as kynurenine formamidase efficiently hydrolyzed the pyrrolidin ring of 6-hydroxyindoline-2,3-dione what is in accordance with the proposed biosynthetic route of the C7 unit.

Based on the obtained data, a biosynthetic scheme of mansouramycin D production was proposed (Figure 2.34). The products of the genes 5 and 3 which are both annotated as tryptophan-2,3-dioxygenases are proposed to oxidize both the indole ring and the side chain f the L-tryptophan. The side chain of the oxidized product is removed through the action of the enzyme annotated as a putative kynureninase (encoded by 4) yielding 6-hydroxyindoline-2,3-dione and L-serine. Production of 6-hydroxyindoline-2,3-dione was experimentally proved in the culture broth of the strains with the inactivated genes 8 and 6. The feeding of 6-hydroxyindoline-2,3-dione to various biosynthetic mutants implies that the genes 8, 6, 7 and 9 are responsible for its conversion in mansouramycin D. According to the *in vitro* assay the pyrrolidin ring of 6-hydroxyindoline-2,3-dione is hydrolyzed by the product of the gene 8 which encodes putative kynurenine formamidase. 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid is formed as a product of this reaction. According to the determined sequence of the mansouramycin biosynthetic reactions, the product of the gene 6 encoding putative FAD-dependent oxidoreductase catalyzes the next biosynthetic step. We propose that the product of 6 catalyzes the conversion of 2-(2-amino-4-hydroxyphenyl)-2-oxoacetic acid into 5-amino-3,6-dioxocyclohexa-1,4-diene-1-carboxylic acid by oxidizing the benzene ring, decarboxylation and the rearrangement of the ring substituents. The carboxylic group of the compound is likely to be aminated by the product of the gene 7 encoding a putative aminotransferase. The obtained 5-amino-3,6-dioxocyclohexa-1,4-diene-1-carboxamide might correspond to the C7 unit which coupled with the C10 unit gives rise to mansouramycin D.

The biosynthesis of the C10 unit remains elusive. We propose that the amino group of the tryptophan is removed by oxidative deamination catalyzed by one of the enzymes of the

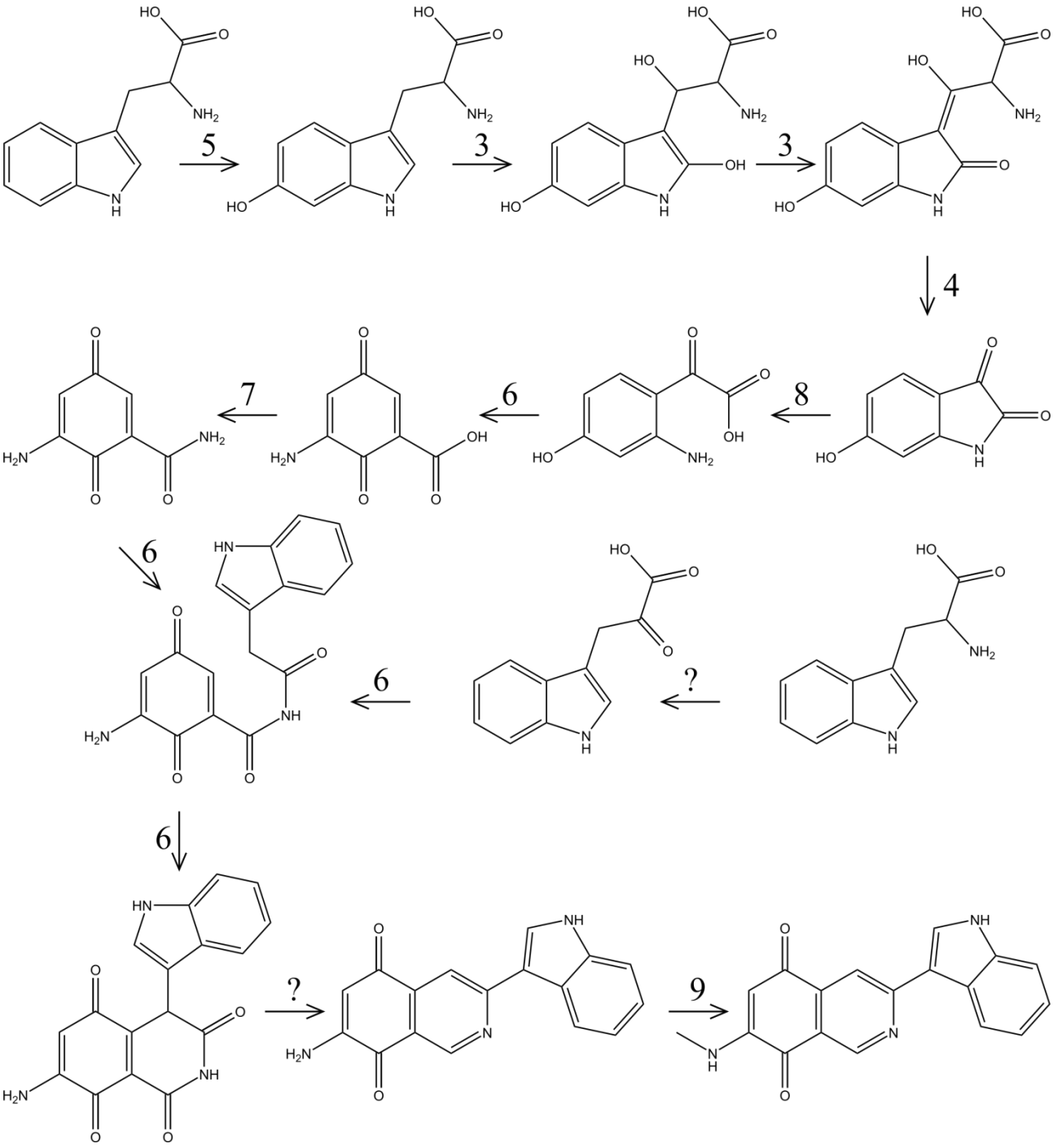

Figure 2.34. Proposed scheme of mansouramycin D biosynthesis. The numbers 1 – 9 correspond to enzymes encoded by mansouramycin biosynthetic genes 1 – 9.

mansouramycin cluster. The product of 6 catalyzes the decarboxylation of the obtained indole-3-pyruvic, its coupling to the C7 unit and cyclization of the product. It is not clear how the aromatization of the pyridine ring and the rearrangement of the indole moiety occur. The final biosynthetic step is proposed to be methylation of the amino group bound to the isoquinoline moiety. The product of the gene 9 annotated as a putative methyltransferase might be responsible for this modification.

4 References

1. Hawas, U.W.; Shaaban, M.; Shaaban, K.A.; Speitling, M.; Maier, A.; Kelter, G.; Fiebig, H.H.; Meiners, M.; Helmke, E.; Laatsch, H. Mansouramycins A-D, cytotoxic isoquinolinequinones from a marine streptomycete. *J. Nat. Prod.* **2009**, *72*, 2120–2124, doi:10.1021/np900160g.
2. Take, Y.; Oogose, K.; Kubo, T.; Inouye, Y.; Nakamura, S.; Kitahara, Y.; Kubo, A. Comparative study on biological activities of heterocyclic quinones and streptonigrin. *J. Antibiot.* **1987**, *40*, 679–684, doi:10.7164/antibiotics.40.679.
3. Hagel, J.M.; Facchini, P.J. Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* **2013**, *54*, 647–672, doi:10.1093/pcp/pct020.
4. Baccile, J.A.; Spraker, J.E.; Le, H.H.; Brandenburger, E.; Gomez, C.; Bok, J.W.; Macheleidt, J.; Brakhage, A.A.; Hoffmeister, D.; Keller, N.P.; et al. Plant-like biosynthesis of isoquinoline alkaloids in Aspergillus fumigatus. *Nat. Chem. Biol.* **2016**, *12*, 419–424, doi:10.1038/nchembio.2061.
5. Kutchan, T.M.; Dittrich, H. Characterization and mechanism of the berberine bridge enzyme, a covalently flavinylated oxidase of benzophenanthridine alkaloid biosynthesis in plants. *J. Biol. Chem.* **1995**, *270*, 24475–24481, doi:10.1074/jbc.270.41.24475.
6. Green, M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual (Fourth Edition)*; 4th Ed.; Cold Spring Harbor Lab. Press, Plainview, NY, 2012;
7. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics*; John Innes Foundation, Norwich, England, 2000;
8. Bilyk, O.; Sekurova, O.N.; Zotchev, S.B.; Luzhetskyy, A. Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. *PLoS ONE* **2016**, *11*, e0158682, doi:10.1371/journal.pone.0158682.
9. Myronovskyi, M.; Rosenkränzer, B.; Luzhetskyy, A. Iterative marker excision system. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 4557–4570, doi:10.1007/s00253-014-5523-z.
10. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324, doi:10.1016/j.ymben.2018.09.004.
11. Myronovskyi, M.; Welle, E.; Fedorenko, V.; Luzhetskyy, A. β-Glucuronidase as a Sensitive and Versatile Reporter in Actinomycetes ▿. *Appl Environ Microbiol* **2011**, *77*, 5370–5383, doi:10.1128/AEM.00434-11.
12. Chater, K.F.; Wilde, L.C. Streptomyces albus G mutants defective in the SalGI restriction-modification system. *J. Gen. Microbiol.* **1980**, *116*, 323–334, doi:10.1099/00221287-116-2-323.

13. Grant, S.G.; Jessee, J.; Bloom, F.R.; Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4645–4649, doi:10.1073/pnas.87.12.4645.
14. Flett, F.; Mersinias, V.; Smith, C.P. High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155*, 223–229, doi:10.1111/j.1574-6968.1997.tb13882.x.
15. Gregory, M.A.; Till, R.; Smith, M.C.M. Integration Site for Streptomyces Phage φBT1 and Development of Site-Specific Integrating Vectors. *J Bacteriol* **2003**, *185*, 5320–5323, doi:10.1128/JB.185.17.5320-5323.2003.
16. Lopatniuk, M.; Myronovskyi, M.; Nottebrock, A.; Busche, T.; Kalinowski, J.; Ostash, B.; Fedorenko, V.; Luzhetskyy, A. Effect of "ribosome engineering" on the transcription level and production of S. albus indigenous secondary metabolites. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 7097–7110, doi:10.1007/s00253-019-10005-y.
17. Hoshino, T.; Kondo, T.; Uchiyama, T.; Ogasawara, N. Biosynthesis of Violacein: a Novel Rearrangement in Tryptophan Metabolism with a 1, 2-Shift of the Indole Ring. *Agricultural and Biological Chemistry* **1987**, *51*, 965–968, doi:10.1271/bbb1961.51.965.
18. Momen, A.Z.; Hoshino, T. Biosynthesis of violacein: intact incorporation of the tryptophan molecule on the oxindole side, with intramolecular rearrangement of the indole ring on the 5-hydroxyindole side. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 539–549, doi:10.1271/bbb.64.539.
19. Kurnasov, O.; Goral, V.; Colabroy, K.; Gerdes, S.; Anantha, S.; Osterman, A.; Begley, T.P. NAD biosynthesis: identification of the tryptophan to quinolinate pathway in bacteria. *Chem. Biol.* **2003**, *10*, 1195–1204, doi:10.1016/j.chembiol.2003.11.011.
20. Zummo, F.P.; Marineo, S.; Pace, A.; Civiletti, F.; Giardina, A.; Puglia, A.M. Tryptophan catabolism via kynurenine production in Streptomyces coelicolor: identification of three genes coding for the enzymes of tryptophan to anthranilate pathway. *Appl. Microbiol. Biotechnol.* **2012**, *94*, 719–728, doi:10.1007/s00253-011-3833-y.
21. Radwanski, E.R.; Last, R.L. Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. *Plant Cell* **1995**, *7*, 921–934, doi:10.1105/tpc.7.7.921.
22. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. - PubMed - NCBI Available online: https://www.ncbi.nlm.nih.gov/pubmed/25948579 (accessed on Mar 26, 2020).
23. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. - PubMed - NCBI Available online: https://www.ncbi.nlm.nih.gov/pubmed/30196100 (accessed on Mar 26, 2020).

24. Siegl, T.; Tokovenko, B.; Myronovskyi, M.; Luzhetskyy, A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab. Eng.* **2013**, *19*, 98–106, doi:10.1016/j.ymben.2013.07.006.

Chapter 3: Identification of a biosynthetic gene cluster responsible for the production of a new pyrrolopyrimidine natural product – huimycin

1 Introduction

Pyrrolopyrimidines are a class of natural products characterized by the presence of a 7-deazapurine moiety in their structures [1]. This class of compounds is widely spread among three domains of life: archaea, bacteria, and eukarya [1]. The study of pyrrolopyrimidines began in 1956 with the discovery of an anti-Candida compound, toyocamycin, in the culture broth of *Streptomyces toyocaensis* [2]. Soon after the discovery of toyocamycin, other secreted 7-deazapurine compounds, tubercidin and sangivamycin, were isolated from the culture filtrates of *Streptomyces tubercidicus* and *Streptomyces rimosus* [3,4]. To date, more than 30 different 7-deazapurines have been isolated from various biological sources, including bacteria, cyanobacteria, red algae, marine sponges, and tunicates [1]. In addition to secreted pyrrolopyrimidines, which act as natural products, pyrrolopyrimidines can also be found as modified bases in tRNA. Queuosine is one of the most studied modified nucleosides that contains a 7-deazapurine moiety [5]. It is present in the wobble position of aspartyl, asparaginyl, histidyl and tyrosyl tRNA, where it likely modifies base-pairing characteristics of tRNA and enhances translational efficiency [6–8]. Distributed ubiquitously among prokaryotes and eukaryotes, queuosine cannot be found in archaebacteria. In archaeal species, archeosine, noncanonical 7-deazapurine nucleoside, is incorporated into tRNA [9].

Secreted pyrrolopyrimidines often possess antibacterial, antifungal, anticancer, antiviral and anti-inflammatory activities [1,10–12]. Their structural similarity to purine bases allows 7-deazapurines to interfere with various cellular processes involving nucleotides and nucleosides [1]. For instance, toyocamycin and tubercidin have been shown to be substrates of mammalian adenosine kinase [13]. In the phosphorylated form, toyocamycin is recognized by mammalian DNA and RNA polymerases and can become incorporated into DNA and RNA [14]. Although pyrrolopyrimidines are not substrates for amino acid aminoacylation, toyocamycin and sangivamycin inhibit this process [15]. When incorporated into the acceptor stem of tRNA, pyrrolopyrimidine inhibits tRNA aminoacylation [15]. Sangivamycin has been recently demonstrated to inhibit protein kinases [16,17]. A broad range of biological activities characteristic of pyrrolopyrimidines makes them promising potential drug leads for the pharmaceutical industry. Recent FDA approval of ribociclib and baricitinib, which both containing a 7-deazapurine moiety, for the treatment of breast cancer and rheumatoid arthritis further emphasizes the pharmaceutical importance of the pyrrolopyrimidine class of compounds [18,19].

In this study, we report the identification and heterologous expression of a biosynthetic gene cluster of *Kutzneria albida* DSM 43870 that encodes a new member of the pyrrolopyrimidine family of natural products, huimycin. The compound was isolated, and its structure was elucidated using extensive 2D NMR. Through a series of DNA deletion experiments, the minimal cluster

required for huimycin biosynthesis was determined. Based on the cluster analysis, we also propose the biosynthetic route that produces huimycin.

2 Materials and Methods

Strains and BACs

All of the strains used in this study are listed in Table 3.1. The BACs and plasmids are listed in Table 3.2. *Escherichia coli* strains were cultured in LB medium [20]. For sporulation and conjugation *Streptomyces* strains were cultivated on soya flour mannitol agar (MS agar) [21] and in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA). For secondary metabolite production liquid DNPM medium was used [22]. The antibiotics ampicillin, kanamycin, apramycin, hygromycin and nalidixic acid were supplemented when required.

Table 3.1. Bacterial strains used in this work.

Strain	Description	Reference or source
Kutzneria albida DSM 43870	The wild type strain; the source of the huimycin luster	[23]
Streptomyces albus Del14	The heterologous host strain; cluster-free derivative of the *S. albus* J1074	[24]
S. albus 2I16	Derivative of *S. albus* Del14 harboring the 2I16 BAC	This work
S. albus 2I16_LS	Derivative of *S. albus* Del14 harboring the 2I16_LS BAC	This work
S. albus 2I16_RS	Derivative of *S. albus* Del14 harboring the 2I16_RS BAC	This work
S. albus 2I16_4069	Derivative of *S. albus* Del14 harboring the 2I16_4069 BAC	This work
S. albus 2I16_4074	Derivative of *S. albus* Del14 harboring the 2I16_4074 BAC	This work
S. albus 2I16_4075	Derivative of *S. albus* Del14 harboring the 2I16_4075 BAC	This work
S. albus 2I16_4076	Derivative of *S. albus* Del14 harboring the 2I16_4076 BAC	This work
Escherichia coli ET12567 pUB307	Donor strain for intergeneric conjugation	[25]
Escherichia coli DH10β	General cloning strain	[26]

Table 3.2. Plasmids and BACs used in this work.

BAC	**Description**	**Reference or source**
2I16	The Bac containing 95 kb chromosomal fragment of *K. albida*; contains huimycin gene cluster	[23]
2I16_LS	The derivative of 2I16 with the deletion of 20 kb DNA fragment upstream the huiA gene	This work
2I16_RS	The derivative of 2I16 with the deletion of 36 kb DNA fragment downstream the KALB_4076 gene	This work
2I16_4069	The derivative of 2I16 with the deletion of KALB_4069 (*huiC*) gene	This work
2I16_4074	The derivative of 2I16 with the deletion of KALB_4074 gene	This work
2I16_4075	The derivative of 2I16 with the deletion of KALB_4075 gene	This work
2I16_4076	The derivative of 2I16 with the deletion of KALB_4076 gene	This work
pACS-hyg	The plasmid containing hygromycin resistance gene	[27]
pUC19	General cloning vector	Thermo Scientific

DNA manipulation

Isolation of DNA and all subsequent manipulations were performed according to standard protocols [20]. BAC extraction from the genomic library of *K. albida* was performed using the

BACMAX™ DNA purification kit (Lucigen, Middleton, WI, USA). Restriction endonucleases were used according to manufacturer's recommendations (New England Biolabs, Ipswich, MA, USA). Primers used in this study are listed in Table 3.3.

Table 3.3. Primers used in this study.

Primer	Sequence
LS_F	CGGTTCCGTCCTGTCACGAAGTGCGCCAGTTGGTAGAGGGCCGCGTGCGCAATACTTGAC ATATCACTGT
LS_R	TACGAGCCCAGCTCACTGAGGCTGGTCTCGAGTGTGTTGGGCGCGTCAGTTCTAGATCAG GCGCCGGGGGCGGTGT
RS_F	ACGCAGAGGCAGCCCGACGCGCCAGCCCACGGTCGTAGCAAGGAGCCAGGCGTCAGGTG GCACTTTTCG
RS_R	TGATGCAGCAGGGCACTGGGCAGGACGTAACTGGTTGACCGCAGGATCGCGATATCTTAC CAATGCTTAATCAGTG
4069_F	ATGAGCATGGGAACATCCCGGACAGCGGTGACTCCTCTCGACGCCGAGGGTTTAAACCGT CAGGTGGCACTTTTCG
4069_R	CACAGCACCGAAGCCGTCGACGTGTGGGGGCGGGCGATCAGGACGGCGCCGTTTAAACT TACCAATGCTTAATCAGTG
4074_F	ATGATCGTTCTCGGACTGATCGGACGGCCCGACGTCCCGCTCTGCCATGGTTTAAACCGT CAGGTGGCACTTTTCG
4074_R	CAGGAGCGGCTCACCGCGTACGGCGGCATGACCAGCAGGTCGACGTCCGGGTTTAAACTT ACCAATGCTTAATCAGTG
4075_F	ATGGGGAAGAGCTACGAGCGGATAGACGGCAGACTGCGTGCCTTCATCGGTTTAAACCG TCAGGTGGCACTTTTCG
4075_R	TATTCAGCGCGCGGAAGCGGTGGCAGCGGCAACGGCAACCCCGGCAGGCCGTTTAAACT TACCAATGCTTAATCAGTG
4076_F	ATGGTCGACAACATGCGGAGGTGTGATGAGATACCCGATTCCACGATTGGTTTAAACCGT CAGGTGGCACTTTTCG
4076_R	CAGGTCGGTGAGGTCGCTGTGGGGATTTCGGCGATACGCCAAACGCCGATGTTTAAACTT ACCAATGCTTAATCAGTG
LS_chF	GTTCTCCCTTCCACCAGCC
LS_chR	TGTTCTTCAGGACGCGGAC
RS_chF	TGATTTCGTCGCGGTGGAA
RS_chR	GCCGCATACGACAGGGAAT
4069_chF	AACCTGCGGAACCTGCTAC
4069_chR	AACTTGCTGAGTCCCGCTC
4074_chF	CGAACAGTTGTGGTGTGCG
4074_chR	GGTCTTGACAATTGCTCCGG
4075_chF	CGATCCCCTTGGCATGTGA
4075_chR	CGCTGCTGGATGAGATCGT
4076_chF	AACCAACAAGGGGCTGTCC
4076_chR	AGTTCCACCGCGACGAAAT

To determine the minimal huimycin biosynthetic cluster the BACs 2I16_LS, 2I16_RS, 2I16_4069, 2I16_4074, 2I16_4075 and 2I16_4076 were constructed. In the BAC 2I16_LS 20 kb DNA fragment upstream the *huiA* gene was deleted. For this purpose, the hygromycin gene was amplified from the pACS-hyg plasmid with the LS-F/LS-R pair of primers. The obtained PCR fragment was used for the Red-ET modification of the BAC 2I16 [28]. The constructed BAC 2I16 was checked by PCR with the LS_chF/LS_chR pair of primers with subsequent sequencing of the obtained PCR product.

The BACs 2I16_RS, 2I16_4069, 2I16_4074, 2I16_4075 and 2I16_4076 were constructed in a similar manner. Here ampicillin resistance marker was used for recombineering purposes instead of hygromycin marker. The marker was amplified from the pUC19 plasmid with the pairs of primers RS_F/RS_R, 4069_F/4069_R, 4074_F/4074_R, 4075_F/4075_R, 4076_F/4076_R. The obtained PCR products were utilized for the construction of the abovementioned BACs using Red-ET modification. The construction of the BACs was confirmed using PCR with the pairs of primers RS_chF/RS_chR, 4069_chF/4069_chR, 4074_chF/4074_chR, 4075_chF/4075_chR, 4076_chF/4076_chR and sequencing.

Metabolite extraction and analysis

Streptomyces albus strains containing 2I16 BAC and its derivatives were grown in 15 mL of TSB medium for 24 hours. 1 mL of the seed culture was used for inoculation of 100 ml of DNPM medium. The cultures were cultivated for 7 days at 28 °C. Huimycin was extracted with the equal amount of butanol from the culture supernatant, evaporated, and dissolved in methanol.

For mass determination, a Bruker Amazon Speed and a Thermo LTQ Orbitrap XL mass spectrometer were used. Both machines were coupled to UPLC Thermo Dionex Ultimate3000 RS. Analytes were separated on a Waters ACQUITY BEH C18 Column (1.7 μm, 2.1mm x 100 mm) with water +0.1% formic acid and acetonitrile +0.1% formic acid as the mobile phase.

Huimycin isolation and NMR data acquisition

Streptomyces albus 2I16 was cultivated into 20 L of DNPM medium at 28 °C for 7 days. The mycelial part was separated by centrifugation. The supernatant was extracted once with the equal amount of butanol. The solvent was evaporated under vacuum in a rotary evaporator. The huimycin was purified using size-exclusion and reverse-phase chromatography. Size-exclusion chromatography was performed using Sephadex LH-20 (GE Healthcare, USA) and methanol as a solvent. The HPLC separation was performed on semipreparative HPLC (Dionex UltiMate 3000, Thermo Fisher Scientific, USA) equipped with a C18 column (Synergi 10 μm, 250 × 10 mm; Phenomenex, Aschaffenburg, Germany). Water +0.1% formic acid and acetonitrile +0.1% formic acid were used as the mobile phase.

NMR spectra were recorded in meod4 at 500 MHz on a Bruker Avance 500 spectrometer (Bruker, BioSpin GmbH, Rheinstetten, Germany) equipped with 5mm TXI cryoprobe. HSQC, HMBC, $^{1}H-^{1}H$ COSY, and 2D TOCSY experiments were acquired using standard pulse program. CNST13 of HMBC were set as 2,3JC-H = 2,8 and 10 Hz.

Genome mining and bioinformatics analysis

The *K. albida* genome was screened for secondary metabolite biosynthetic gene clusters using the antiSMASH online tool [29]. Geneious R9 (Biomatters Ltd) software package was used for DNA sequence analysis.

3 Results and discussion

3.1 Identification of the huimycin gene cluster through its heterologous expression in *Streptomyces albus* Del14

Recently, we reported the complete genome sequence of *Kutzneria albida* DSM 43870 (GenBank accession number NZ_CP007155), which is a representative of a rarely observed genus in the *Pseudonocardiaceae* family [23]. Forty-six putative clusters encoding secondary metabolites were identified in the genome of this strain [23]. To enable analysis of these metabolites, a genomic library of *K. albida* was constructed using an integrative BAC vector. In the course of systematic activation of cryptic secondary metabolite clusters from *K. albida*, a cluster annotated by the antiSMASH genome mining software as "nucleoside biosynthetic cluster" was expressed in the heterologous host strain *S. albus* Del14. For this purpose, a BAC 2I16 vector containing the cluster was isolated from the constructed genomic library and transferred into the chassis strain *S. albus* Del14 by conjugation. The obtained exconjugant strain *S. albus* 2I16 and the corresponding control strain without the BAC *S. albus* Del14 were fermented in the production medium DNPM. The culture filtrate of the strains was extracted with butanol, and the extracts were analyzed by LC-MS. This analysis revealed a new peak in the extract of *S. albus* 2I16 as a result of the cluster expression (Figure 3.1). Subsequent analysis of the extract using high resolution LC-MS revealed that the identified peak corresponded to a compound with an $[M+H]^+$ of 393.15 m/z (Figure 3.1). The extract of the control strain *S. albus* Del14 did not contain the identified ion. A search in a natural product database for the identified high resolution mass did not generate any matches, implying that the identified compound might be new.

3.2 Isolation and structure elucidation of the huimycin

To obtain structural information about the potentially new compound obtained from the heterologous expression of the nucleoside gene cluster, we set out to purify it. For this purpose, the *S. albus* 2I16 strain with the expressed nucleoside gene cluster was inoculated into 20 L of DNPM medium, and the culture broth was extracted with butanol. The compound was separated from contaminants in the extract using size-exclusion and reverse-phase chromatography. A total of 8.2 mg of the compound was isolated during the purification process and was used for subsequent structure elucidation purposes (Figures 3.2 – 3.9). The structure elucidation of the compound was performed by Dr. Suvd Nadmid.

The combination of the ^{1}H NMR spectrum with the HSQC data indicated an isolated aromatic signal, one anomeric methine, one methoxy signal together with 3 oxygenated methines and one methylene signal as well as a methyl group. The TOCSY and COSY spectra suggested the presence of a sugar moiety, which was further determined to be 1-amino 2-deoxy-glucose by means of analyzing HMBC cross peaks. HMBC correlations from H-2' and H-9' to C-8' further revealed the

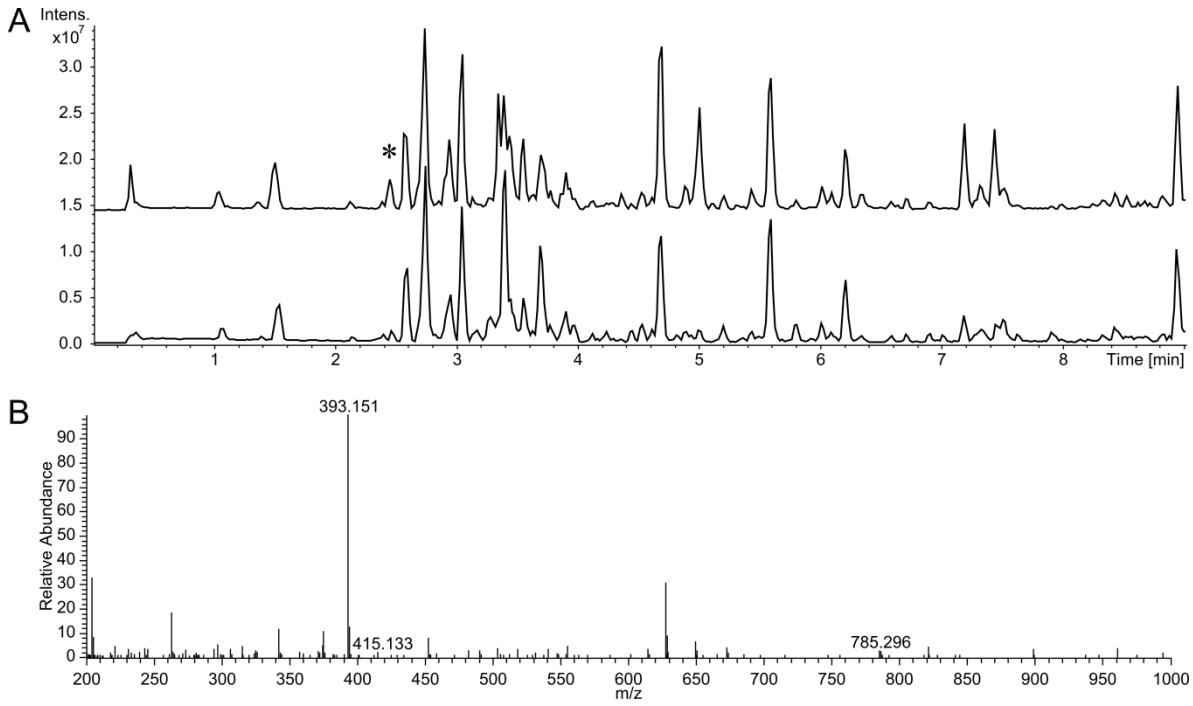

Figure 3.1. LC-MS detection of huimycin. (A) Base peak chromatograms of crude extracts of *S. albus* 2I16 harboring the huimycin gene cluster (upper chromatogram) and of the control strain *S. albus* Del14 (lower chromatogramm). The new peak found in the *S. albus* 2I16 extract is indicated with an asterix. (B) High resolution mass spectrum of the new peak corresponding to huimycin. The molecular ion peaks at m/z 393.151, 415.133 and 785.296 correspond to huimycin [M+H]+, its sodium adduct [M+Na]+ and its dimer $[2M+H]^+$.

substitution of an acetate on the amino group of the sugar (Table 3.4). Large coupling constant observed for the anomeric proton of the glucose moiety at 5.28 (d, J = 9.6 Hz) suggested its β-orientation. Further ROESY correlations were observed between H1'/H3'/H5' as well as H2'/H4' of the glucose moiety.

The constitution of an aglycon core was determined to be purine, which possessed substitutions on C-2, C-6 and C-7. Several HMBC experiments with different long-range coupling constants (CNST 13) were used to link the substitutions to the correct positions. The HMBC spectrum acquired with CNST13=2 Hz showed a more intense cross peak from the H-8 to a carbon at δ 160.3 than what was observed in the HMBC with 10 Hz data, suggesting that carbon at δ 160.3 (C-2) is the furthest carbon from H-8 and shows a 5 bond correlation. Furthermore, according to HMBC, the alpha proton of the sugar moiety showed correlation with C-2, indicating its linkage. The methoxy group was found to be linked on C-6 because of the HMBC correlation (in CNST13=2 Hz data) of methyl protons to C-5 (at δ 98.7 ppm), showing a 4 bond heteronuclear correlation. Due to the biosynthetic similarity of this molecule to toyocamycin and their NMR data similarity, the cyano-group was attached on C-7. This was supported by the HMBC correlation from H-8 to carbon at δ 116.3 (C-10) and the chemical shift of C-7 at δ 84.5 ppm.

Finally, the planar structure was deduced, as shown in Figure 3.10, which is in concordance with the HR-ESI-MS data (meas.d 393.1509, calc.d 393.1517, Δppm 2.0). The isolated compound was named huimycin and was identified as a novel natural compound.

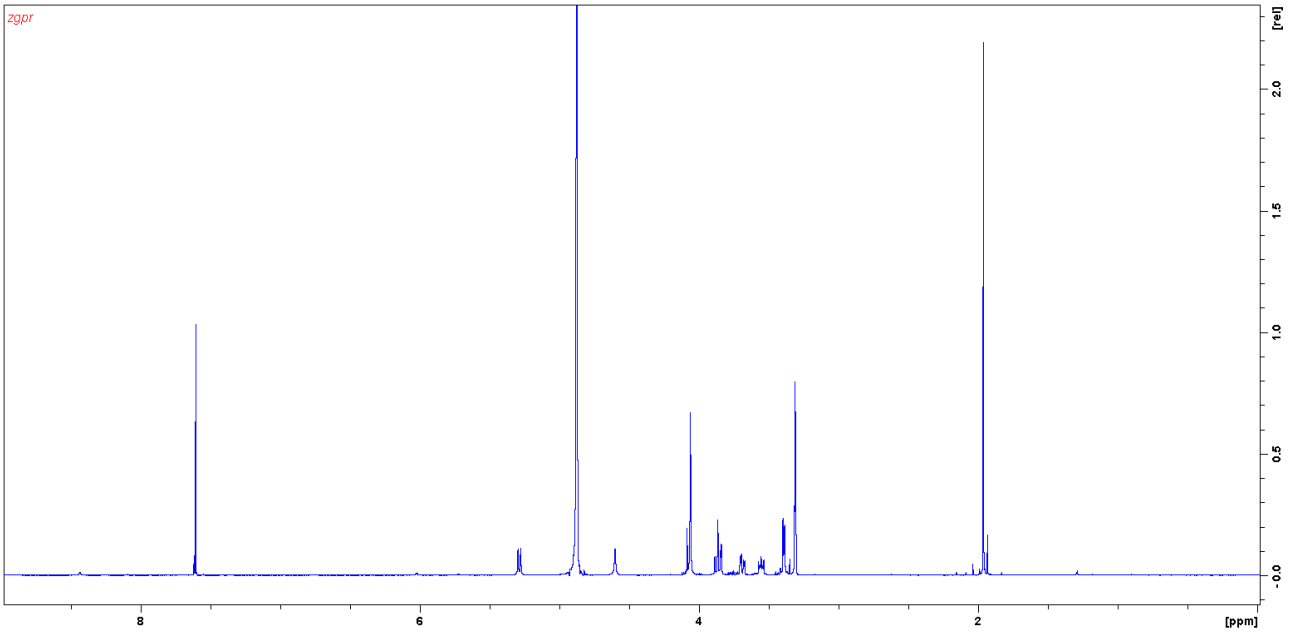

Figure 3.2. ^{1}H NMR (500 MHz, $MeOD_4$) spectrum of huimycin 1.

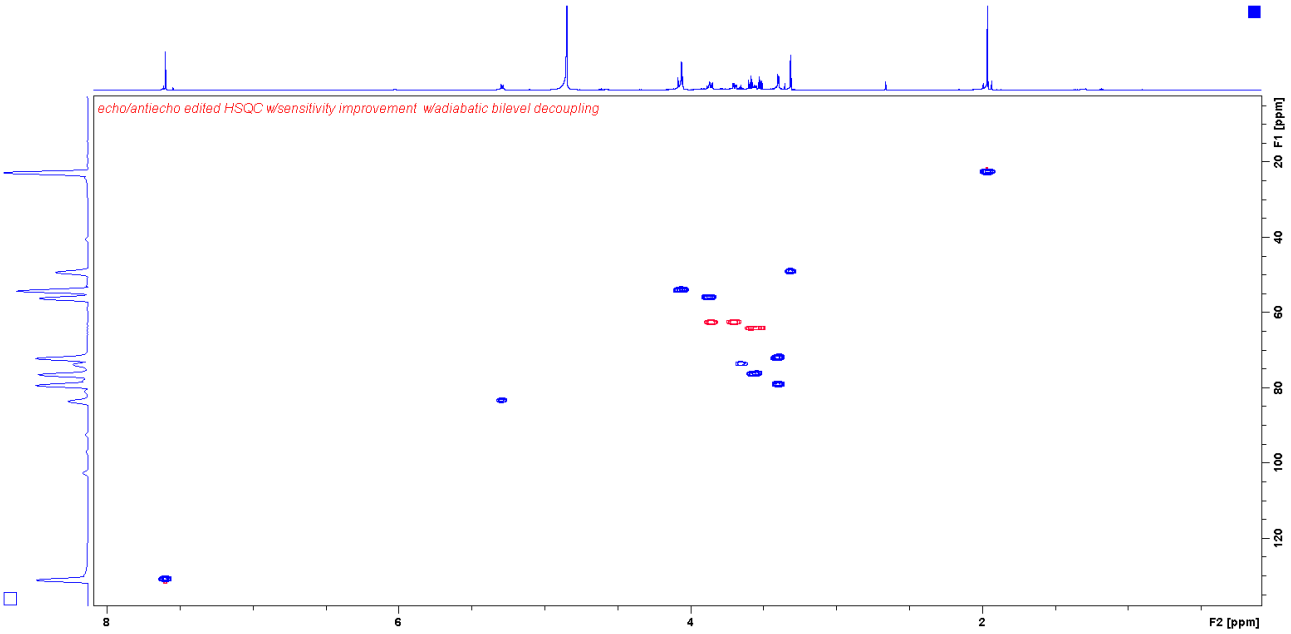

Figure 3.3. HSQC (500 MHz, $MeOD_4$) spectrum of huimycin 1.

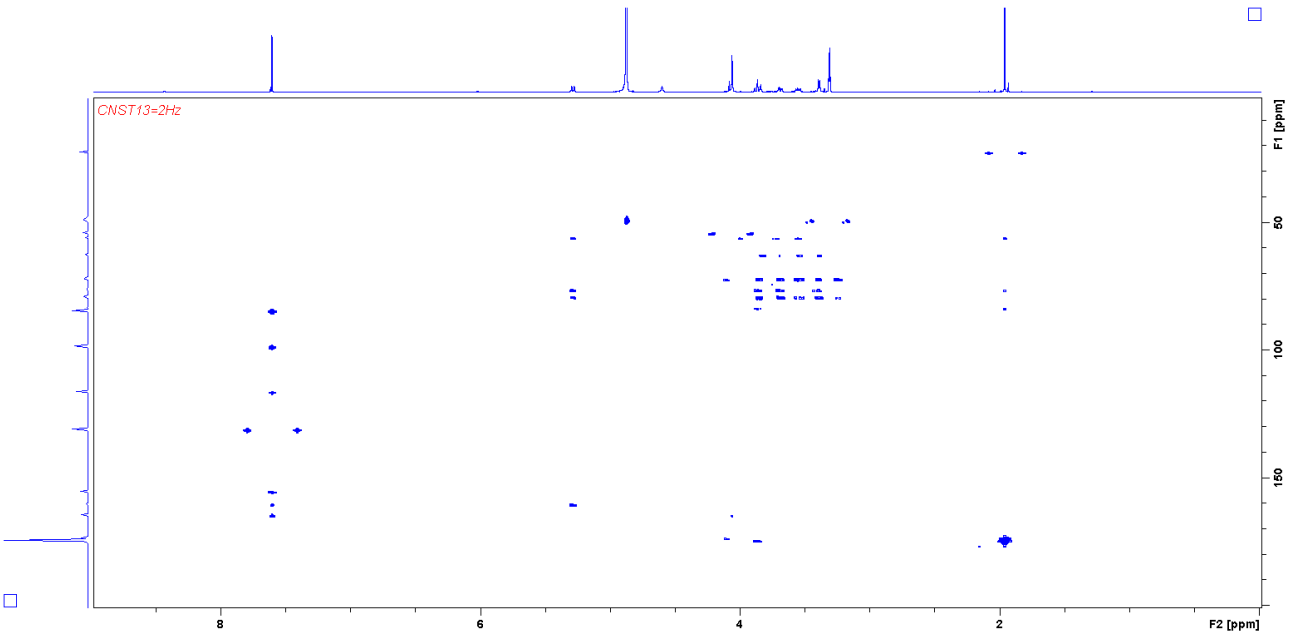

Figure 3.4. HMBC (500 MHz, $MeOD_4$) spectrum (CNST 13=2 Hz) of huimycin 1.

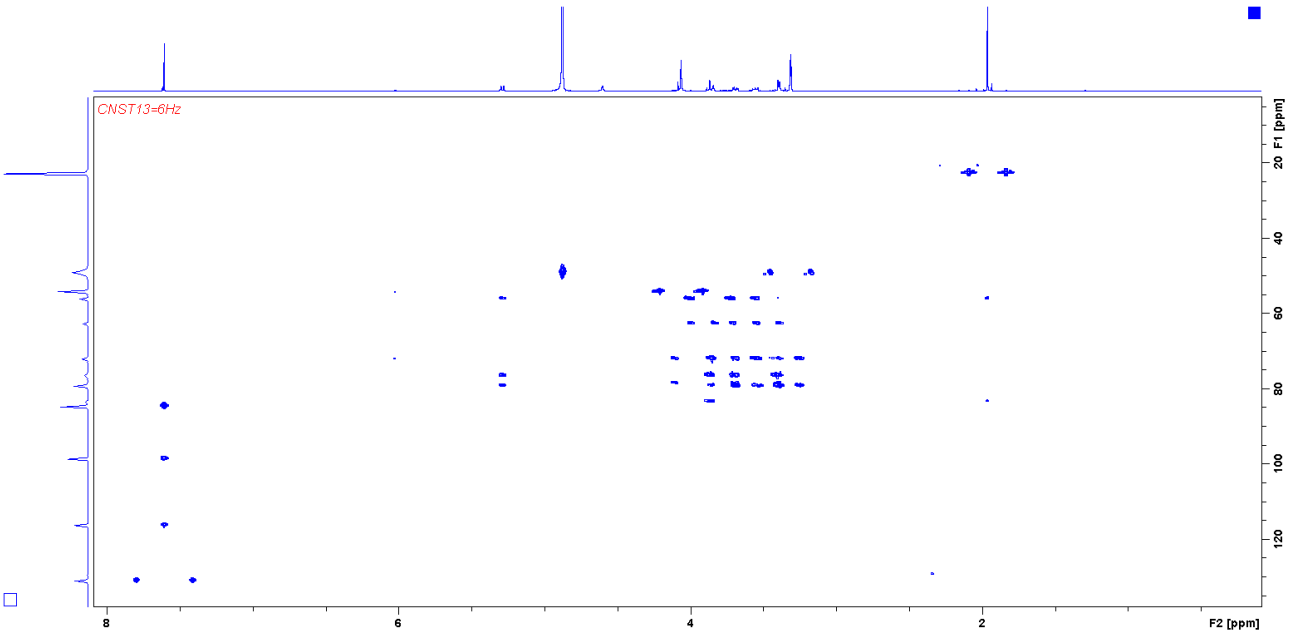

Figure 3.5. HMBC (500 MHz, $MeOD_4$) spectrum (CNST 13=6 Hz) of huimycin 1.

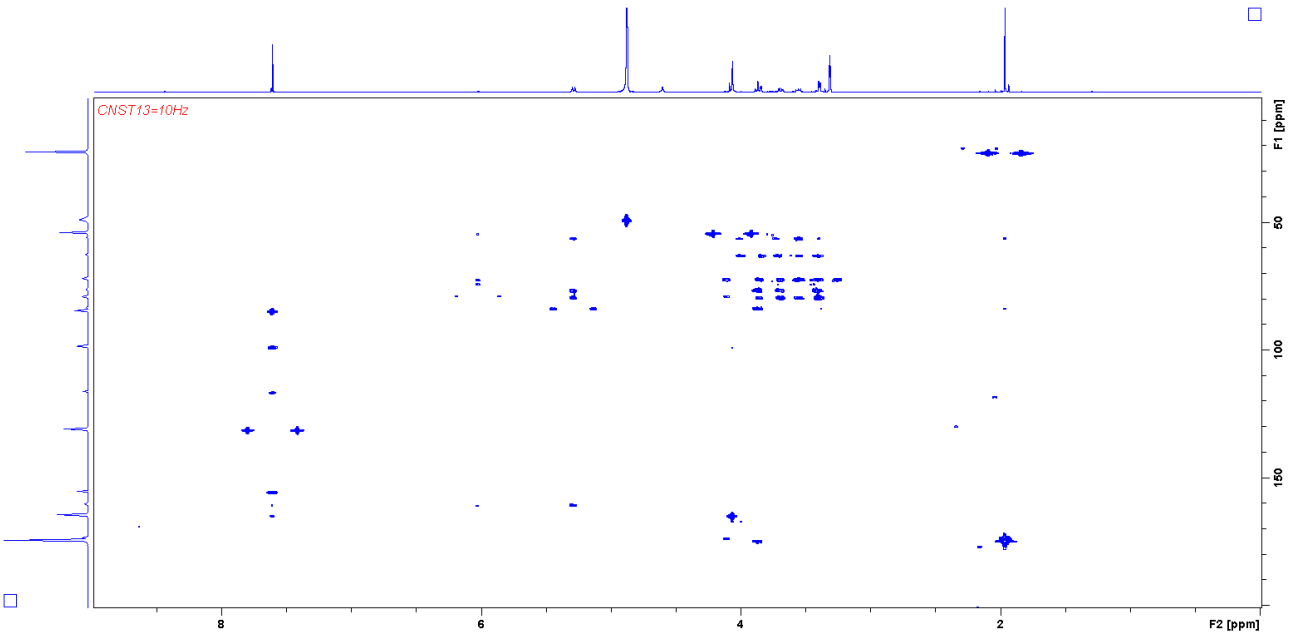

Figure 3.6. HMBC (500 MHz, $MeOD_4$) spectrum (CNST 13=10 Hz) of huimycin 1.

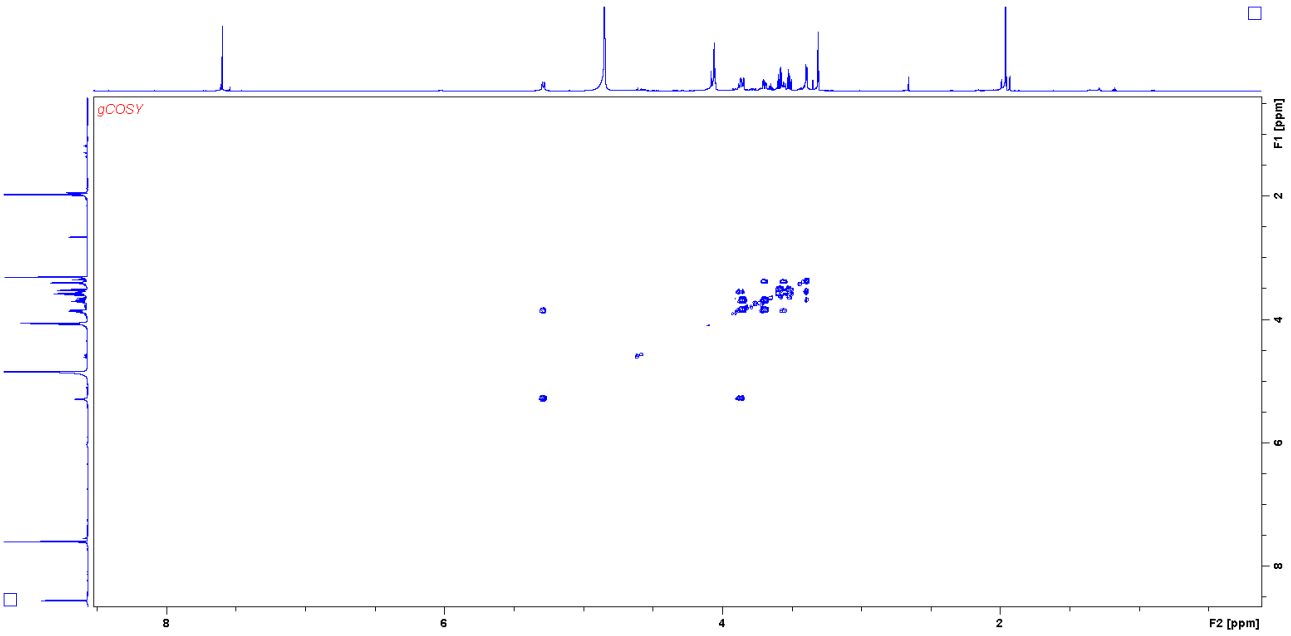

Figure 3.7. ^{1}H-^{1}H COSY (500 MHz, $MeOD_4$) spectrum of huimycin 1.

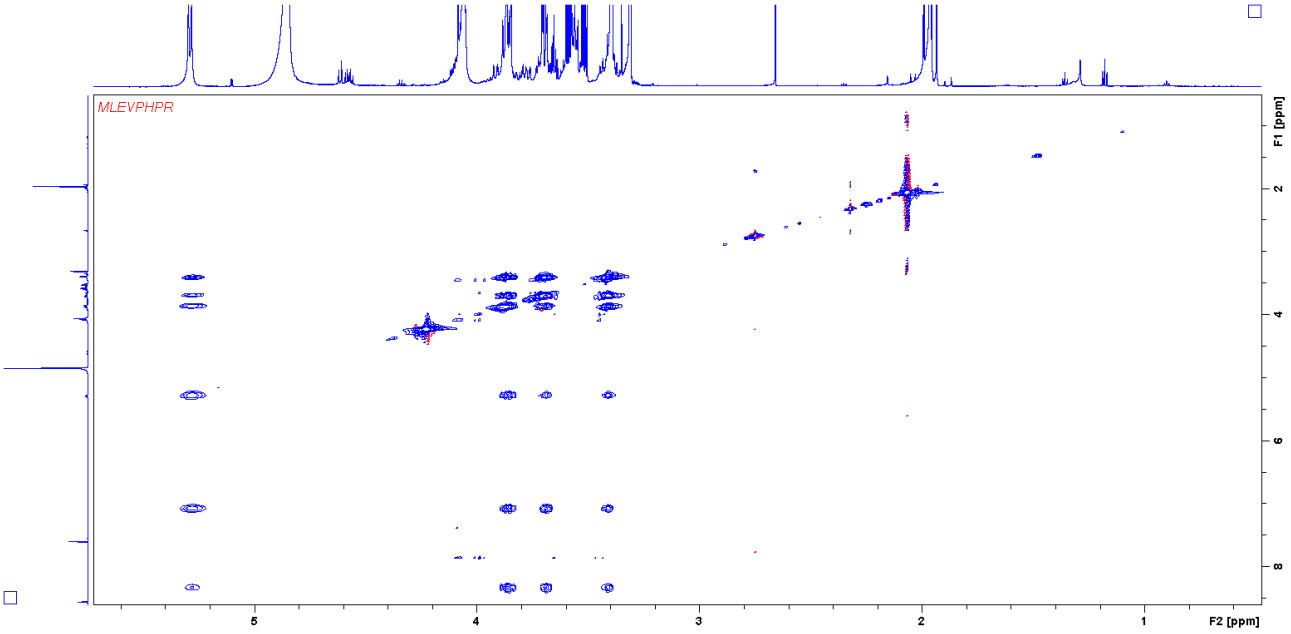

Figure 3.8. TOSCY (500 MHz, $MeOD_4$) spectrum of huimycin 1.

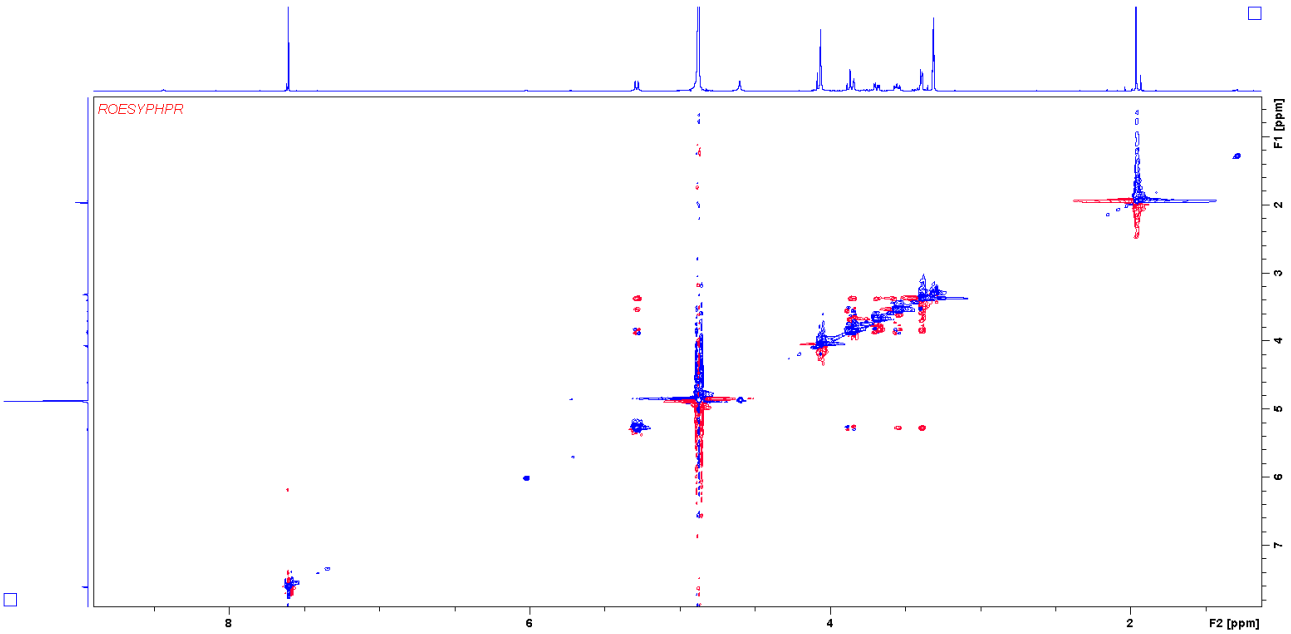

Figure 3.9. ROESY (500 MHz, $MeOD_4$) spectrum of huimycin 1.

3.3 Determination of the minimal huimycin gene cluster

The presence of a 7-deazapurine moiety, typical of a pyrrolopyrimidine class of nucleoside antibiotics [1], in the structure of the isolated huimycin indicates the participation of the expressed nucleoside cluster in its production. The huimycin gene cluster was expressed in the heterologous

Table 3.4. NMR spectroscopic data of huimycin.

Position	δ_C [a]	δ_H mult. (J, Hz) [b]	HMBCc
2	160.3		
4	155.4		
5	98.7		
6	164.5		
7	84.5		
8	131.0	7.60, s	2[e], 4, 5, 6[e], 7, 10
10	116.3		
11	54.2	4.05, s	5[e], 6
1'	83.6	5.28, d 9.6	2, 2', 3', 5'
2'	56.1	3.87 [d]	1', 3', 8'
3'	76.3	3.55, m	2', 4'
4'	71.9	3.38 [d]	5', 6'
5'	79.2	3.39 [d]	
6' a	62.6	3.69, m	4'
6' b		3.85 [d]	
8'	174.5		
9'	22.7	1.96, s	8'

[a] acquired at 125 MHz, referenced to solvent signal meod4 at δ 49.15 ppm.
[b] acquired at 500 MHz, referenced to solvent signal meod4 at δ 3.31 ppm.
[c] proton showing HMBC correlation to indicated carbons.
[d] overlapped signals.
[e] observed in HMBC with CNST13=2Hz.

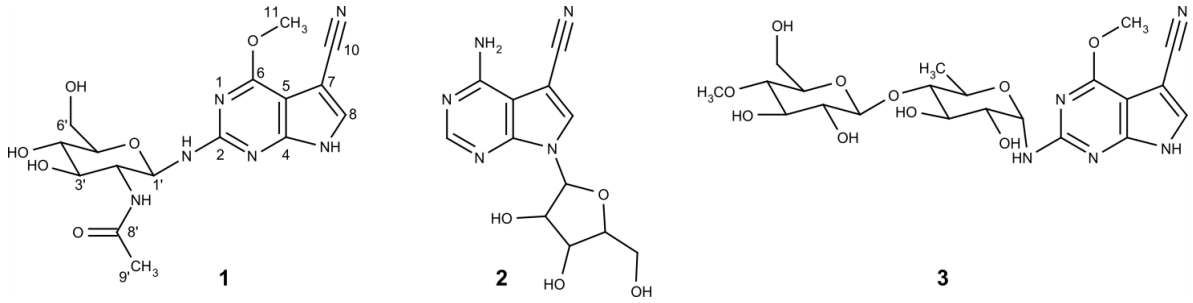

Figure 3.10. The structures of huimycin (1), toyocamycin (2) and dapiramycin A (3).

host strain as a part of the large 95-kb chromosomal fragment in BAC 2I16. To determine the minimal set of genes required for huimycin biosynthesis, a sequence analysis and a series of gene deletion experiments were performed.

The sequence analysis of the DNA fragment cloned in BAC 2I16 leading to nucleoside production revealed the presence of seven open reading frames, *huiA* – *huiG* (locus tags KALB_4067 – KALB_4073), which are highly likely to be involved in huimycin biosynthesis (Figure 3.11, Table 3.5). Five of these genes shared homology at the protein level with the genes within the biosynthetic cluster of toyocamycin – the parent compound of the pyrrolopyrimidine class of antibiotics (Figure 3.10, Table 3.5) [30]. The *huiA* gene, encoding a putative transcriptional

regulator, shares homology with the regulatory gene *toyA* [30]. The genes *huiB*, *huiD*, *huiE* and *huiF* encode putative 7-cyano-7-deazaguanine synthase, 6-carboxytetrahydropterin synthase, 7-carboxy-7-deazaguanine synthase and GTP cyclohydrolase I, and they share homology with the toyocamycin biosynthetic genes *toyM*, *toyB*, *toyC* and *toyD*, respectively (Table 3.5) [30]. The structural similarities between huimycin and toyocamycin also imply similar biosynthetic routes leading to the production of antibiotics. The genes *huiC* and *huiG* do not have counterparts in the toyocamycin gene cluster; they encode putative SAM-dependent methyltransferase and glycosyltransferase, respectively (Table 3.5). The genes *huiC* and *huiG* are likely responsible for the structural differences between huimycin and toyocamycin.

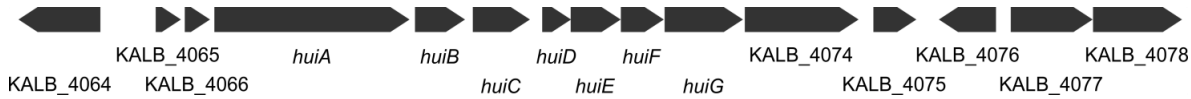

Figure 3.11. The chromosomal fragment of *K. albida* containing the huimycin biosynthetic gene cluster.

Table 3.5. Proposed functions of the genes in the DNA fragment containing huimycin gene.

Gene	**Proposed function**	**Homolog in Toy pathway 3**
KALB_4064	enhanced intracellular survival protein	
KALB_4065	hypothetical protein	
KALB_4066	hypothetical protein	
huiA; KALB_4067	pathway-specific regulator	*toyA*
huiB; KALB_4068	7-cyano-7-deazaguanine synthase	*toyM*
huiC; KALB_4069	SAM-dependent methyltransferase	
huiD; KALB_4070	6-carboxytetrahydropterin synthase	*toyB*
huiE; KALB_4071	7-carboxy-7-deazaguanine synthase	*toyC*
huiF; KALB_4072	GTP cyclohydrolase I	*toyD*
huiG; KALB_4073	glycosyltransferase	
KALB_4074	carbamoyltransferase	
KALB_4075	pyridoxamine 5'-phosphate oxidase	
KALB_4076	SAM-dependent methyltransferase	
KALB_4077	glycosyltransferases	
KALB_4078	glycosyltransferases	

The gene *huiA* encodes a transcriptional regulator and was assumed to constitute the 5' outer border of the huimycin cluster (Figure 3.11). The counterpart of the gene *huiA* within the toyocamycin biosynthetic pathway, *toyA*, also constitutes the first gene of the cluster [30]. The deletion of the *toyA* gene completely abolishes toyocamycin production [31]. The first three genes in the region upstream of the *huiA* gene, KALB_4064, KALB_4065 and KALB_4066 (Figure 3.11), encode an enhanced intracellular survival protein and two hypothetical proteins, respectively. No function in nucleoside biosynthesis could be assigned to these three genes. This further

strengthens the assumption that the *huiA* gene constitutes the 5' boundary of the cluster. To corroborate this assumption empirically, the 20 kb DNA fragment upstream of the *huiA* gene was deleted within the 2I16 BAC through RedET recombineering. The obtained construct, 2I16_LS, resulted in huimycin production when it was transferred into an *S. albus* strain. This clearly indicated that the genes in the upstream region of the *huiA* are not required for huimycin biosynthesis. Thus, it is likely that the *huiA* gene is at the 5' end of the huimycin gene cluster.

Prediction of the 3' boundary of the huimycin cluster was not obvious from the gene annotation. The last gene showing homology to toyocamycin biosynthetic genes is *huiF* (Table 3.5). This gene is followed by the gene *huiG*, which encodes a glycosyltransferase that might participate in nucleoside biosynthesis. The genes KALB_4074, KALB_4075, KALB_4076, KALB_4077, and KALB_4078 in the downstream region of the *huiG* encode a putative carbamoyltransferase, pyridoxamine 5'-phosphate oxidase family protein, SAM-dependent methyltransferase, and two glycosyltransferases, respectively. As is evident in the huimycin structure (Figure 3.10), glycosyltransferase and methyltransferase activities are required to produce huimycin. The required methyltransferase may be encoded either by *huiC* or by KALB_4076. It is also not obvious which of the three glycosyltransferase genes, *huiG*, KALB_4077 or KALB_4078, participates in huimycin biosynthesis. To clarify this, the 36 kb DNA fragment downstream of the KALB_4076 gene was deleted in the 2I16 BAC. The obtained BAC 2I16_RS led to huimycin production when introduced into *S. albus*. This clearly indicates that the genes in the downstream region of KALB_4076, including the two glycosyltransferase genes KALB_4077 and KALB_4078, do not participate in huimycin biosynthesis. The possible participation of the genes KALB_4074, KALB_4075, and KALB_4076 in huimycin biosynthesis was assessed by their inactivation in the 2I16 BAC. The obtained BACs (2I16_4074, 2I16_4075 and 2I16_4076) contained the deletions of the genes KALB_4074, KALB_4075 and KALB_4076, respectively, and they resulted in huimycin production when introduced in *S. albus*. This shows that none of the inactivated genes is essential for huimycin production. Thus, it is likely that the *huiG* gene is at the 3' end of the huimycin gene cluster.

Through DNA deletion studies, the genes located outside the *huiA* – *huiG* fragment have been shown not to participate in huimycin biosynthesis. To demonstrate the importance of the *huiA* – *huiG* fragment for nucleoside production, the *huiC* gene, encoding the methyltransferase, was deleted from the 2I16 BAC. The deletion of the *huiC* gene completely abolished the production of the huimycin, showing its involvement in the production of this nucleoside.

3.4 Biosynthesis of huimycin

Seven genes *huiA* – *huiG* have been shown to be sufficient for huimycin production. The *huiA* gene encodes a putative transcriptional regulator and is likely to be involved in the regulation of

expression of the huimycin biosynthetic gene cluster. The homolog of *huiA*, the *toyA* gene, encodes a pathway-specific regulator of the toyocamycin gene cluster [30]. The six genes *huiB* – *huiG* encode structural enzymes that catalyze huimycin biosynthetic steps. Four of these genes, *huiB*, *huiD*, *huiE* and *huiF*, share sequence homology on the peptide level with the toyocamycin biosynthetic genes *toyM*, *toyB*, *toyC* and *toyD*, respectively. The genes *huiC* and *huiG* have no homologs in the toyocamycin cluster and encode a putative methyltransferase and glycosyltransferase, respectively. The considerable sequence similarity between huimycin and toyocamycin gene clusters implies the similarities in the biosynthetic routes leading to the production of the compounds.

Similar to toyocamycin, GTP is regarded as a main precursor for huimycin production [1,30]. The first reaction in huimycin biosynthesis is the conversion of GTP into 7,8-dihydroneopterin triphosphate (H2NTP) (Figure 3.12). This reaction is catalyzed by the product of the gene *huiF*, which encodes a putative GTP cyclohydrolase I. The product of the *huiF* shares 66% identity with the product of the *toyD* gene, which also catalyzes the first step of the toyocamycin biosynthesis [1,30]. The second step in the pathway is catalyzed by 6-carboxytetrahydropterin synthase encoded by *huiD*, which converts H2NTP into 6-carboxy-5,6,7,8-tetrahydropterin (CPH4) (Figure 3.12). The product of the *huiD* shares 65% identity with the product of the *toyB*, which is responsible for the similar reaction in the toyocamycin biosynthesis [1,30]. The third biosynthetic step is catalyzed by the product of the *toyC* homolog, *huiE*, which encodes the 7-carboxy-7-deazaguanine synthase. This enzyme converts CPH4 into 7-carboxy-7-deazaguanine (CDG) [1,30]. The products of *toyC* and *huiE* share 50% identity. The last step common for both huimycin and toyocamycin biosynthesis is the conversion of the CDG into 7-cyano-7-deazaguanine (PreQ0) through the action of 7-cyano-7-deazaguanine synthase (Figure 3.12). This enzyme is encoded by the *toyM* homolog, *huiB*. On the protein level, these genes share 72% identity. We propose that the last two enzymatic steps required for the conversion of PreQ0 into huimycin are the methylation of 7-cyano-7-deazaguanine and the attachment of the N-acetylglucosamine moiety (Figure 3.12). The

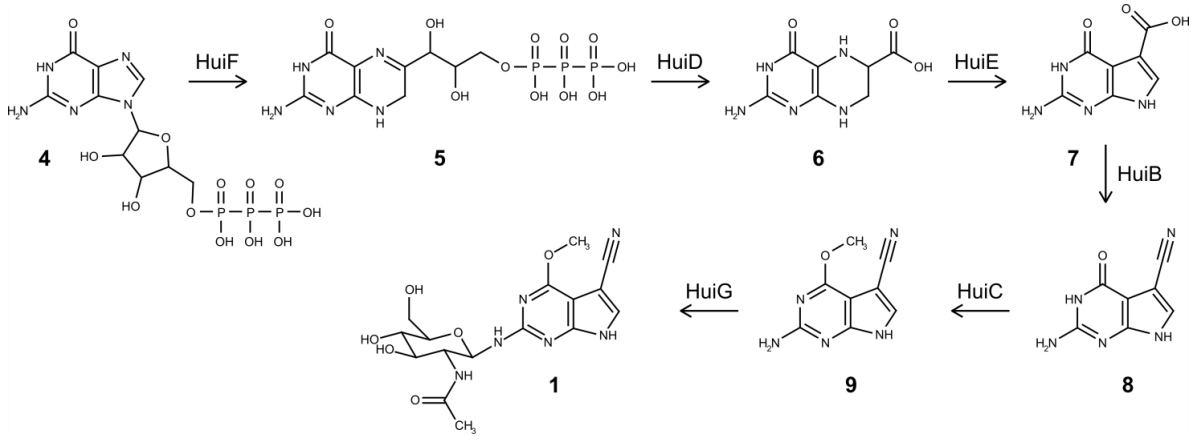

Figure 3.12. Proposed biosynthetic pathway of huimycin. 4 – GTP, 5 – 7,8-dihydroneopterin triphosphate, 6 – 6-carboxy-5,6,7,8-tetrahydropterin, 7 – 7-carboxy-7-deazaguanine, 8 – 7-cyano-7-deazaguanine, 9 – 2-amino-6-methoxy-7-cyano-7-deazapurine, 1 – huimycin.

methylation reaction is likely to be catalyzed by the product of the gene *huiC*, which encodes a SAM-dependent methyltransferase. The attachment of the N-acetylglucosamine to the 2-amino-6-methoxy-7-cyano-7-deazapurine is catalyzed by the glycosyltransferase encoded by the last gene in the huimycin gene cluster – *huiG*. The order in which the last two reactions take place in the biosynthesis is not known. Huimycin is thus far the only known member of the pyrrolopyrimidine class of compounds with an N-acetylglucosamine moiety in its structure. Additionally, the position of the N-glycosidic bond distinguishes huimycin from most of the 7-deazapurine nucleosides. Only in the structures of dapiramycin A and B, sugar moieties are also attached to the amino group at the 2nd position of the 7-deazapurine chromophore (Figure 3.10) [32,33]. Furthermore, dapiramycin A and B share the same chromophore as huimycin. Since the dapiramycin biosynthetic gene cluster has not yet been discovered, it is not possible to study whether the structural similarities between huimycin and the dapiramycins are also reflected on the DNA sequence level.

Here we describe the identification and successful heterologous expression of the gene cluster responsible for the production of huimycin, a new member of the pyrrolopyrimidine family of nucleoside natural products. Huimycin features that are unique for nucleosides include an N-acetylglucosamine moiety attached to the 2-amino-6-methoxy-7-cyano-7-deazapurine core. The minimal set of huimycin biosynthetic genes was identified through a series of gene deletion experiments. The expression of the huimycin gene cluster in *S. albus* reported here is the first example of successful heterologous expression of a secondary metabolite pathway from the rare actinomycetal genus *Kutzneria*.

4 References

1. McCarty, R.M.; Bandarian, V. Biosynthesis of pyrrolopyrimidines. *Bioorg. Chem.* **2012**, *43*, 15–25, doi:10.1016/j.bioorg.2012.01.001.
2. Nishimura, H.; Katagiri, K.; Sato, K.; Mayama, M.; Shimaoka, N. Toyocamycin, a new anti-candida antibiotics. *J. Antibiot.* **1956**, *9*, 60–62.
3. Anzai, K.; Nakamura, G.; Suzuki, S. A new antibiotic, tubercidin. *J. Antibiot.* **1957**, *10*, 201–204.
4. Rao, K.V. Structure of sangivamycin. *J. Med. Chem.* **1968**, *11*, 939–941, doi:10.1021/jm00311a005.
5. Vinayak, M.; Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci. Rep.* **2009**, *30*, 135–148, doi:10.1042/BSR20090057.
6. Harada, F.; Nishimura, S. Possible anticodon sequences of tRNA His , tRNA Asm , and tRNA Asp from Escherichia coli B. Universal presence of nucleoside Q in the first postion of the anticondons of these transfer ribonucleic acids. *Biochemistry* **1972**, *11*, 301–308, doi:10.1021/bi00752a024.
7. Meier, F.; Suter, B.; Grosjean, H.; Keith, G.; Kubli, E. Queuosine modification of the wobble base in tRNAHis influences “in vivo” decoding properties. *EMBO J.* **1985**, *4*, 823–827.
8. Urbonavicius, J.; Qian, Q.; Durand, J.M.; Hagervall, T.G.; Björk, G.R. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J.* **2001**, *20*, 4863–4873, doi:10.1093/emboj/20.17.4863.
9. Kilpatrick, M.W.; Walker, R.T. The nucleotide sequence of the tRNAMMet from the archaebacterium Thermoplasma acidophilum. *Nucleic Acids Res.* **1981**, *9*, 4387–4390, doi:10.1093/nar/9.17.4387.
10. Pathania, S.; Rawal, R.K. Pyrrolopyrimidines: An update on recent advancements in their medicinal attributes. *Eur J Med Chem* **2018**, *157*, 503–526, doi:10.1016/j.ejmech.2018.08.023.
11. Saneyoshi, M.; Tokuzen, R.; Fukuoka, F. Antitumor Activities and Structural Relationship of Tubercidine, Toyocamycin, and Their Derivatives. *GANN Japanese Journal of Cancer Research* **1965**, *56*, 219–222, doi:10.20772/cancersci1959.56.2_219.
12. Acs, G.; Reich, E.; Mori, M. Biological and biochemical properties of the analogue antibiotic tubercidin. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *52*, 493–501, doi:10.1073/pnas.52.2.493.
13. Lindberg, B.; Klenow, H.; Hansen, K. Some properties of partially purified mammalian adenosine kinase. *J. Biol. Chem.* **1967**, *242*, 350–356.
14. Suhadolnik, R.J.; Uematsu, T.; Uematsu, H. Toyocamycin: phosphorylation and incorporation into RNA and DNA and the biochemical properties of the triphosphate. *Biochim. Biophys. Acta* **1967**, *149*, 41–49, doi:10.1016/0005-2787(67)90689-2.

15. Uretsky, S.C.; Acs, G.; Reich, E.; Mori, M.; Altwerger, L. Pyrrolopyrimidine nucleotides and protein synthesis. *J. Biol. Chem.* **1968**, *243*, 306–312.
16. Loomis, C.R.; Bell, R.M. Sangivamycin, a nucleoside analogue, is a potent inhibitor of protein kinase C. *J. Biol. Chem.* **1988**, *263*, 1682–1692.
17. Osada, H.; Sonoda, T.; Tsunoda, K.; Isono, K. A new biological role of sangivamycin; inhibition of protein kinases. *J. Antibiot.* **1989**, *42*, 102–106, doi:10.7164/antibiotics.42.102.
18. Roskoski, R. Properties of FDA-approved small molecule protein kinase inhibitors. *Pharmacol. Res.* **2019**, *144*, 19–50, doi:10.1016/j.phrs.2019.03.006.
19. Mogul, A.; Corsi, K.; McAuliffe, L. Baricitinib: The Second FDA-Approved JAK Inhibitor for the Treatment of Rheumatoid Arthritis. *Ann Pharmacother* **2019**, *53*, 947–953, doi:10.1177/1060028019839650.
20. Green, M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual (Fourth Edition)*; 4th Ed.; Cold Spring Harbor Lab. Press, Plainview, NY, 2012;
21. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics*; John Innes Foundation, Norwich, England, 2000;
22. Rodríguez Estévez, M.; Myronovskyi, M.; Gummerlich, N.; Nadmid, S.; Luzhetskyy, A. Heterologous Expression of the Nybomycin Gene Cluster from the Marine Strain Streptomyces albus subsp. chlorinus NRRL B-24108. *Mar Drugs* **2018**, *16*, doi:10.3390/md16110435.
23. Rebets, Y.; Tokovenko, B.; Lushchyk, I.; Rückert, C.; Zaburannyi, N.; Bechthold, A.; Kalinowski, J.; Luzhetskyy, A. Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin Kutzneria albida DSM 43870T, a representative of minor genus of Pseudonocardiaceae. *BMC Genomics* **2014**, *15*, 885, doi:10.1186/1471-2164-15-885.
24. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324, doi:10.1016/j.ymben.2018.09.004.
25. Flett, F.; Mersinias, V.; Smith, C.P. High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155*, 223–229, doi:10.1111/j.1574-6968.1997.tb13882.x.
26. Grant, S.G.; Jessee, J.; Bloom, F.R.; Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4645–4649, doi:10.1073/pnas.87.12.4645.
27. Myronovskyi, M.; Brötz, E.; Rosenkränzer, B.; Manderscheid, N.; Tokovenko, B.; Rebets, Y.; Luzhetskyy, A. Generation of new compounds through unbalanced transcription of

landomycin A cluster. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 9175–9186, doi:10.1007/s00253-016-7721-3.

28. Muyrers, J.P.P.; Zhang, Y.; Benes, V.; Testa, G.; Rientjes, J.M.J.; Stewart, A.F. ET recombination: DNA engineering using homologous recombination in E. coli. *Methods Mol. Biol.* **2004**, *256*, 107–121, doi:10.1385/1-59259-753-X:107.
29. 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.
30. McCarty, R.; Bandarian, V. Rosetta stone for deciphering deazapurine biosynthesis: pathway for pyrrolopyrimidine nucleosides toyocamycin and sangivamycin. *Chem Biol* **2008**, *15*, 790–798, doi:10.1016/j.chembiol.2008.07.012.
31. Xu, J.; Song, Z.; Xu, X.; Ma, Z.; Bechthold, A.; Yu, X. ToyA, a positive pathway-specific regulator for toyocamycin biosynthesis in Streptomyces diastatochromogenes 1628. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 7071–7084, doi:10.1007/s00253-019-09959-w.
32. Shomura, T.; Nishizawa, N.; Iwata, M.; Yoshida, J.; Ito, M.; Amano, S.; Koyama, M.; Kojima, M.; Inouye, S. Studies on a new nucleoside antibiotic, dapiramicin. I. Producing organism, assay method and fermentation. *J. Antibiot.* **1983**, *36*, 1300–1304, doi:10.7164/antibiotics.36.1300.
33. Nishizawa, N.; Kondo, Y.; Koyama, M.; Omoto, S.; Iwata, M.; Tsuruoka, T.; Inouye, S. Studies on a new nucleoside antibiotic, dapiramicin. II. Isolation, physico-chemical and biological characterization. *J. Antibiot.* **1984**, *37*, 1–5, doi:10.7164/antibiotics.37.1.

Chapter 4: Identification of a new cyclopeptide cyclohuinilsopeptin A through heterologous expression of NRPS cluster of *Kutzneria albida* in *Streptomyces albus* Del14

1 Introduction

Actinobacteria are a valuable source of natural products of great structural diversity [1]. The broad range of their biological activities allows their utilization as antibacterial, anticancer and immunosuppressive drugs as well as crop protection agents. Few examples of commercially successful natural products isolated from actinobacteria include the immunosuppressant rapamycin, the antibiotics teicoplanin, vancomycin and daptomycin, the insecticide spinosad, the anthelmintic and insecticide drug avermectin [2–7]. Besides the interesting biological activities and potential use in medicine and agriculture, the structural diversity of natural products is of great fundamental interest. Studying the biosynthetic routes leading to the production of natural products allows the identification of novel enzymatic activities and expands the chemical space of naturally occurring compounds [8]. The discovery of structurally new natural products with novel biological activities still remains of both applied and fundamental interest.

Traditionally the discovery of natural products was performed by screening directly the extracts of the culture broth of various bacterial strains. This led to the discovery of a great variety of new compounds. However, with the lapse of time the traditional discovery approach lost its efficiency due to constantly increasing rate of rediscovery of already known compounds [9]. To circumvent the rediscovery problem, new natural products are currently often discovered using expression in a heterologous expression approach [10]. The genome sequence of natural product producers is first analyzed by genome mining software and structurally intersecting gene clusters encoding potentially new compounds are identified. The identified clusters are cloned and expressed in one or several heterologous hosts specially designed for natural product discovery purposes [11]. The host strains are usually devoid of indigenous pathways encoding biosynthesis of natural products and are additionally optimized for the high-yield production of heterologously expressed compounds. The heterologous expression approach allows identification of new compounds which are produced in low amounts, simplifies their purification procedure due to reduced host's metabolic background and enables the implementation of bioactivity screens with crude extracts [10].

Here we present the identification of a new cyclopeptide cyclohuinilsopeptin A encoded by a NRPS pathway of *Kutzneria albida* DSM 43870, which is a representative of a rarely observed genus of the *Pseudonocardiaceae* family [12]. The NRPS cluster encoding cyclohuinilsopeptin A was identified in the genome sequence of *K. albida* by genome mining and heterologously expressed in the heterologous strain *Streptomyces albus* Del14. The compound was isolated and structure elucidation was performed. Some insights into the biosynthesis of cyclohuinilsopeptin A are also presented here.

2 Materials and methods

General experimental procedures

All strains, plasmids and BACs used in this work are listed in Tables 4.1 and 4.2. *Escherichia coli* strains were cultured in LB medium [13]. Streptomyces strains were grown on soya flour mannitol agar (MS agar) [14] and in liquid tryptic soy broth (TSB; Sigma-Aldrich, St. Louis, MO, USA). For cyclohuinilsopeptin production, liquid DNPM medium [15] was used. The antibiotics kanamycin, apramycin, hygromycin, ampicillin and nalidixic acid were supplemented when required.

Table 4.1 Bacterial strains used in the study.

Strain	Description	Source
Kutzneria albida DSM 43870	The wild type strain; the source of the cyclohuinilsopeptin A cluster	[12]
Streptomyces albus Del14	The heterologous host strain; cluster-free derivative of the *S. albus* J1074	[16]
S. albus 2D19	Derivative of *S. albus* Del14 harboring the 2D19 BAC	This work
S. albus 2D19_5522	Derivative of *S. albus* Del14 harboring the 2D19_5522 BAC	This work
S. albus 2D19_5521	Derivative of *S. albus* Del14 harboring the 2D19_5521 BAC	This work
S. albus 2D19_5520	Derivative of *S. albus* Del14 harboring the 2D19_5520 BAC	This work
S. albus 2D19_5519	Derivative of *S. albus* Del14 harboring the 2D19_5519 BAC	This work
S. albus 2D19_5518	Derivative of *S. albus* Del14 harboring the 2D19_5518 BAC	This work
S. albus 2D19_5516	Derivative of *S. albus* Del14 harboring the 2D19_5516 BAC	This work
S. albus 2D19_5515	Derivative of *S. albus* Del14 harboring the 2D19_5515 BAC	This work
S. albus 2D19_5514	Derivative of *S. albus* Del14 harboring the 2D19_5514 BAC	This work
S. albus 2D19_5512	Derivative of *S. albus* Del14 harboring the 2D19_5512 BAC	This work
S. albus 2D19_5508	Derivative of *S. albus* Del14 harboring the 2D19_5508 BAC	This work
Escherichia coli ET12567 pUB307	Donor strain for intergeneric conjugation	[17]
Escherichia coli DH10β	General cloning strain	[18]

Table 4.2 Plasmids and BACs used in the study.

Name	Description	Source
2D19	The Bac containing 95 kb chromosomal fragment of *K. albida*; contains cyclohuinilsopeptin A gene cluster	[12]
2D19_5522	The derivative of 2D19 with the deletion of KALB_5522 gene	This work
2D19_5521	The derivative of 2D19 with the deletion of KALB_5521 gene	This work
2D19_5520	The derivative of 2D19 with the deletion of KALB_5520 gene	This work
2D19_5519	The derivative of 2D19 with the deletion of KALB_5519 gene	This work
2D19_5518	The derivative of 2D19 with the deletion of KALB_5518 gene	This work
2D19_5516	The derivative of 2D19 with the deletion of KALB_5516 gene	This work
2D19_5515	The derivative of 2D19 with the deletion of KALB_5515 gene	This work
2D19_5514	The derivative of 2D19 with the deletion of KALB_5514 gene	This work
2D19_5512	The derivative of 2D19 with the deletion of KALB_5512 gene	This work
2D19_5508	The derivative of 2D19 with the deletion of KALB_5508 gene	This work
patt-shyg	The plasmid containing hygromycin resistance cassette	[19]

DNA manipulation

Isolation of DNA and all subsequent manipulations were performed according to standard protocols [13]. All of the primers used in this study are listed in Table 4.3.

BAC extraction from the genomic library of *K. albida* was performed using the BACMAX™ DNA purification kit (Lucigen, Middleton, WI, USA). Restriction endonucleases were used according to manufacturer's recommendations (New England Biolabs, Ipswich, MA, USA). Primers used in this study are listed in Table 4.3.

Table 4.3 Primers used in this study.

Primer	Sequence
KALB_5522_F	TCATCGGGCGGCCGACGATTGGCGGAGCAGTTTCTTGTCGGGCTTGCCGTTCCGGGGA TCCGTCGACCC
KALB_5522_R	GTGGTGAGCGCCAACTACATGCACAGGGTTTTGGACTTGCTCGGTCGGCTGTAGGCTG GAGCTGCTTCG
KALB_5521_F	ATGGATTCCCTGCTCCAGCGGCGCTACCTGGACGATGTCGTCGAAACCGTTCCGGGGA TCCGTCGACCC
KALB_5521_R	TCAGTTGTGGGGAGCTTCATCGAAGTTCCGTCTGCCTGGTTCGCAAATGTGTAGGCTG GAGCTGCTTCG
KALB_5520_F	TCAGAAGTAGGACTCCAGGGCGCCGCTGCGTCGCACGTCCAGCGTGGCGTTCCGGGG ATCCGTCGACCC
KALB_5520_R	ATGGCCGGGCAGCCGTTCCCGGCGGAGCAGGTCGCGGCCGCGGCCGCTGTGTAGGCT GGAGCTGCTTCG
KALB_5519_F	GAACGTCCCCCCTCGTCCGGGCGTGGTGCAACAGCGGCCGGCGAGCGTGTTCCGGGG ATCCGTCGACCC
KALB_5519_R	ATGCGCGTCTGGAGGGTCTCGCCGTGGGACGGCCTGTTGCTGGCCTACATGTAGGCTG GAGCTGCTTCG
KALB_5518_F	TCAGGTCCACCAGAGCAGCTCACCGCGGTCGTTCACCCCGGTGAAGCAGTTCCGGGGA TCCGTCGACCC
KALB_5518_R	CGGTACTGCGCCACGGATTGTCGGCAAGTCGTTGCTGGACAACGCATTCTGTAGGCTG GAGCTGCTTCG
KALB_5516_F	GTGAGAAACCTCGACGAGGCCGCCGTCCGGGAGCTGCTCCCGATGAGCGTTCCGGGG ATCCGTCGACCC
KALB_5516_R	TCAACCCAGACTAGACTTATGTGGCGGATTTCCGGCGCACGACACTAGGTGTAGGCTG GAGCTGCTTCG
KALB_5515_F	TGGCACGGGACATCTTCACCGACGACCACCGCGCGTTCCGGGAGATCGTTTCCGGGGA TCCGTCGACCC
KALB_5515_R	TCACAGCGGAAGGGACTGACCGATGATCTCCTTCATGATCTCGGTGGTGTGTAGGCTG GAGCTGCTTCG
KALB_5514_F	CCAAGCACGTCACCTTCGTCGAGCTTTCGGTCTACGACAACACAGTCCCTTCCGGGGA TCCGTCGACCC
KALB_5514_R	CACAGTGTGCTCAGCCCCATCGGCTGACAGGTCGGCAGCAGGGTGCGCATGTAGGCTG GAGCTGCTTCG
KALB_5512_F	TCAGCGGAGTTTGGCGGCCAGGATCGGCCCGATCTGGGCCAGTGACTCGTTCCGGGGA TCCGTCGACCC
KALB_5512_R	GTGCTGAGGGAGCAACCGCCGGCCCCGGACCTGGTCACGATGTTCCAGGTGTAGGCTG GAGCTGCTTCG
KALB_5508_F	CTCGGCGTGGCAGCGGGCGACCGGGGCCGCCACTATCGTGGACCCGAGGTTCCGGGG ATCCGTCGACCC
KALB_5508_R	ACCTGGCCGCGTACTGCCCGAACACCGTGCTGGACCCCGAAACCGGTAATGTAGGCTG GAGCTGCTTCG
5522_ch_F	CGATGAAGCTCCCCACAACT
5522_ch_R	CTCTTCGACTGCGTCCTGTT
5521_ch_F	ATCCTGTCCGAAATGGCTGG
5521_ch_R	TGTAATCCAAGCATGCCGGT
5520_ch_F	GCGACACGACGATGATGTTG
5520_ch_R	AGCCATTTCGGACAGGATGG
5519_ch_F	TCGGTGGTCTTGTCGATGTG
5519_ch_R	GCTGTACCTGACCAGCAAGT
5518_ch_F	AGTCACTCGCCCGATTTTGT
5518_ch_R	GTGGACCCTGTTCTGCTACC
5516_ch_F	GGCACCACCGAGATCATGAA
5516_ch_R	CGATCACCTCTTCGACCTGG
5515_ch_F	ACTGCCACGTGATGATCCAG
5515_ch_R	ATCGCCTTCAACCCGAACAT
5514_ch_F	TTGCCCGACTACATGGTTCC
5514_ch_R	GTGGTCGTCGGTGAAGATGT
5512_ch_F	CACGACAGGAGCATGTCGTA
5512_ch_R	GAGTGGCACTCGTCTACCTG
5508_ch_F	TGCCTTGACACCGACTCTTC
5508_ch_R	CCGTGGAGACGTCTTGTTCA

To study if the genes KALB_5508, KALB_5512, KALB_5514, KALB_5515, KALB_5516, KALB_5518, KALB_5519, KALB_5520, KALB_5521, KALB_5522 are involved in biosynthesis of cyclohuinilsopeptin, they were inactivated individually in the BAC 2D19. For inactivation of the gene KALB_5508 the hygromycin gene was amplified from the pACS-hyg plasmid with the KALB_5508_F/KALB_5508_R pair of primers. The obtained PCR fragment was used for the Red-ET modification of the BAC 2D19 [20]. The constructed BAC 2D19_5508 was checked by PCR with the 5508_ch_F/5508_ch_R pair of primers with subsequent sequencing of the obtained PCR product. The genes KALB_5512, KALB_5514, KALB_5515, KALB_5516, KALB_5518, KALB_5519, KALB_5520, KALB_5521, and KALB_5522 were inactivated in the same manner. The pairs of primers KALB_5512_F/KALB_5512_R, KALB_5514_F/KALB_5514_R, KALB_5515_F/KALB_5515_R, KALB_5516_F/KALB_5516_R, KALB_5518_F/KALB_5518_R, KALB_5519_F/KALB_5519_R, KALB_5520_F/KALB_5520_R, KALB_5521_F/KALB_5521_R and KALB_5522_F/KALB_5522_R were used for the amplification of the hygromycin resistance cassette. The pairs of primers 5512_ch_F/5512_ch_R, 5514_ch_F/5514_ch_R, 5515_ch_F/5515_ch_R, 5516_ch_F/5516_ch_R, 5518_ch_F/5518_ch_R, 5519_ch_F/5519_ch_R, 5520_ch_F/5520_ch_R, 5521_ch_F/5521_ch_R, and 5522_ch_F/5522_ch_R were used to prove successful construction of the BACs 2D19_5512, 2D19_5514, 2D19_5515, 2D19_5516, 2D19_5518, 2D19_5519, 2D19_5520, 2D19_5521, and 2D19_5522 respectively.

The BAC 2D19 and its derivatives were transferred into *S. albus* Del14 by conjugation [21].

Metabolite extraction and analysis

Streptomyces albus strains containing 2D19 BAC and its derivatives were grown in 15 mL of TSB medium for 24 hours. 1 mL of the seed culture was used for inoculation of 100 ml of DNPM medium. The cultures were cultivated for 7 days at 28 °C. Cyclohuinilsopeptin was extracted with the equal amount of butanol from the culture supernatant, evaporated, and dissolved in methanol.

For mass determination a Bruker Amazon Speed and a Thermo LTQ Orbitrap XL mass spectrometer were used. Both machines were coupled to UPLC Thermo Dionex Ultimate3000 RS. Analytes were separated on a Waters ACQUITY BEH C18 Column (1.7 μm, 2.1mm x 100 mm) with water +0.1% formic acid and acetonitrile +0.1% formic acid as the mobile phase.

Cyclohuinilsopeptin isolation and NMR data acquisition

Streptomyces albus 2D19 was cultivated into 20 L of DNPM medium at 28 °C for 7 days. The mycelial part was separated by centrifugation. The supernatant was extracted once with the equal amount of butanol. The solvent was evaporated under vacuum in a rotary evaporator. The cyclohuinilsopeptin was purified using normalphase, size-exclusion and reverse phase chromatography. Normal phase chromatography was performed on a prepacked silica cartridge

(Biotage, Uppsala, Sweden) using hexane, dichloromethane, ethyl acetate, and methanol as the mobile phase. Biotage Isolera One system was used for separation. Size-exclusion chromatography was performed using Sephadex LH-20 (GE Healthcare, USA) and methanol as a solvent. The HPLC separation was performed on semi-preparative HPLC (Dionex UltiMate 3000, Thermo Fisher Scientific, USA) equipped with a C18 column (Synergi 10 μm, 250 × 10 mm; Phenomenex, Aschaffenburg, Germany). Water +0.1% formic acid and acetonitrile +0.1% formic acid were used as the mobile phase.

The structure elucidation was done using Nuclear Magnetic Resonance (NMR) spectroscopy. Purified compounds were solved in 600 μL deuterated solvent (*DMSO-d_6* with a drop of d-TFA and D_2O or with a drop of TFA and H_2O) and measured in a 5 mm tube. NMR data (^{1}H-NMR, HH-COSY, TOCSY, HSQC-TOCSY, HMBC HSQC, ^{13}C-NMR, ^{15}N-HSQC) were acquired either on a Bruker Ascend 700 spectrometer equipped with a 5 mm TXA Cryoprobe or a Bruker Avance 500 spectrometer equipped with a 5mm BBO Probe at 300K (Bruker, BioSpin, GmbH, Rheinstetten, Germany). The data was analyzed using Bruker's TopSpin 3.5a software.

Cyclohuinilsopeptin A **(1)**: **^{1}H-NMR** (*DMSO-d_6+d-TFA+D_2O;* 700 MHz) δH 7.29 (1H, m, H-12), 7.19 (1H, m, H-13), 7.25 (1H, m, H-14), 4.69 (1H, d, J=5.95 Hz, H-15), 3.68 (1H, m, H-16), 4.03 (1H, dd, J=10.2 Hz, H-17), 3.6 (2H, m, H-18), 4.15 (H, m, H-19), 4.19 (1H, m, H-20), 4.89 (1H, m, H-21), 3.14 (1H, m, H-22), 2.57 (2H, m, H-23), 3.40 (2H, m, H-24), 2.99 (2H, m, H-25), 0.79 (1H, m, H-26), 2.09 (2H, m, H-27), 1.74 (2H, m, H-28), 1.37 (2H, m, H-29), 1.85 (2H, m, H-30), 2.14 (1H, m, H-31), 0.36 (1H, dt, J=6.5 Hz, 5.6 Hz, H-32), 1.25 (2H, m, H-33), 1.08 (3H, d, J=0.77 Hz, H-34), 1.03 (3H, d, J=0.77 Hz, H-35), 1.71 (1H, m, H-37), 1.14 (3H, d, J=6.67 Hz, H-38), 0.99 (3H, d, J=6.42 Hz, H-39), 0.58 (2H, m, H-40), 8.27.

^{1}H-NMR (*DMSO-d_6+TFA+H_2O;* 700 MHz) δH (1H, m, NH-1), 8.24 (1H, m, NH-2), 8.13 (1H, d, J=7.71 Hz, NH-3), 7.66 (1H, d, J=6.71 Hz, NH-4), 8.11 (1H, m, NH-5), 7.74 (1H, t, J=5.55 Hz, NH-6), 7.00(2H, d, br, NH-7), 7.44 (1H, t, J=7.44 Hz, NH-8).

^{13}C-NMR (*DMSO-d_6+d-TFA+D_2O;* 175 MHz) δC 174.35 (C, C-1), 173.64 (C, C-2), 171.90 (C, C-3), 171.63 (C, C-4), 171.03 (C, C-5), 170.77 (C, C-6), 170.54 (C, C-7), 169.85 (C, C-8), 169.62 (C, C-9), 157.30 (C, C-10) 142.99 (C, C-11), 128.81 (CH, C-12), 127.34 (CH, C-13), 126.13 (CH, C-14), 62.62 (CH, C-15), 56.6 (CH, C-16), 53.8 (CH, C-17), 52.40 (CH2, C-18), 51.97 (CH, C-19), 51.05 (CH, C-20), 50.75 (CH, C-21), 46.51 (CH, C-22), 42.12 (CH2, C-23), 41.98 (CH2, C-24), 38.94 (CH2, C-25), 32.74 (CH, C-26), 31.49 (CH2, C-27), 29.92 (CH2, C-28), 28.67 (CH2, C-29), 27.34 (CH2, C-30), 23.9 (CH, C-31), 23.47 (CH, C-32), 23.40 (CH2, C-33), 22.43 (CH3, C-34), 21.25 (CH3, C-35), 21.10 (C, C-36), 15.83 (CH, C-37), 15.20 (CH3, C-38), 13.40 (CH3, C-39), 10.82 (CH2, C-40).

^{15}N-NMR (*DMSO-d_6+TFA+H_2O;* 50 MHz) δN 122.68 (NH, N-1), 121.05 (NH, N-2), 120.20 (NH, N-3), 118.51 (NH, N-4), 117.97 (NH, N-5), 117.24 (NH, N-6), 109.1 (NH2, N-7), 77.60 (NH, N-8). HRESMS: *m/z*=880.4676 $[M+H]^+$ (calcd for $C_{42}H_{62}N_{11}O_{10}^+$ 880.4676).

3 Results and discussion

3.1 Identification and heterologous expression of cyclohuinilsopeptin A gene cluster

One of our research foci is the identification of new natural products through heterologous expression of undescribed biosynthetic gene clusters. Within the framework of this project, the genome of *Kutzneria albida* DSM 43870 was sequenced previously and the genomic library the strain vector was constructed on an integrative shuttle BAC vector [22]. During the systematic heterologous expression of the clusters of *K. albida* an uncharacterized nonribosomal peptide-synthetase (NRPS) cluster was expressed in *Streptomyces albus* Del14 strain. For this purpose, the BAC 2D19 covering the cluster was selected from the genomic library of *K. albida* and transferred by conjugation into the heterologous host strain. The obtained recombinant *S. albus* 2D19 as well as the control strain without the cosmid *S. albus* Del14 were cultivated in the DNPM medium and the culture supernatants were extracted with butanol. LC-MS analysis of the extracts revealed the peak with an m/z value of 880.5 (Figure 4.1). High-resolution LC-MS analysis of the obtained peak

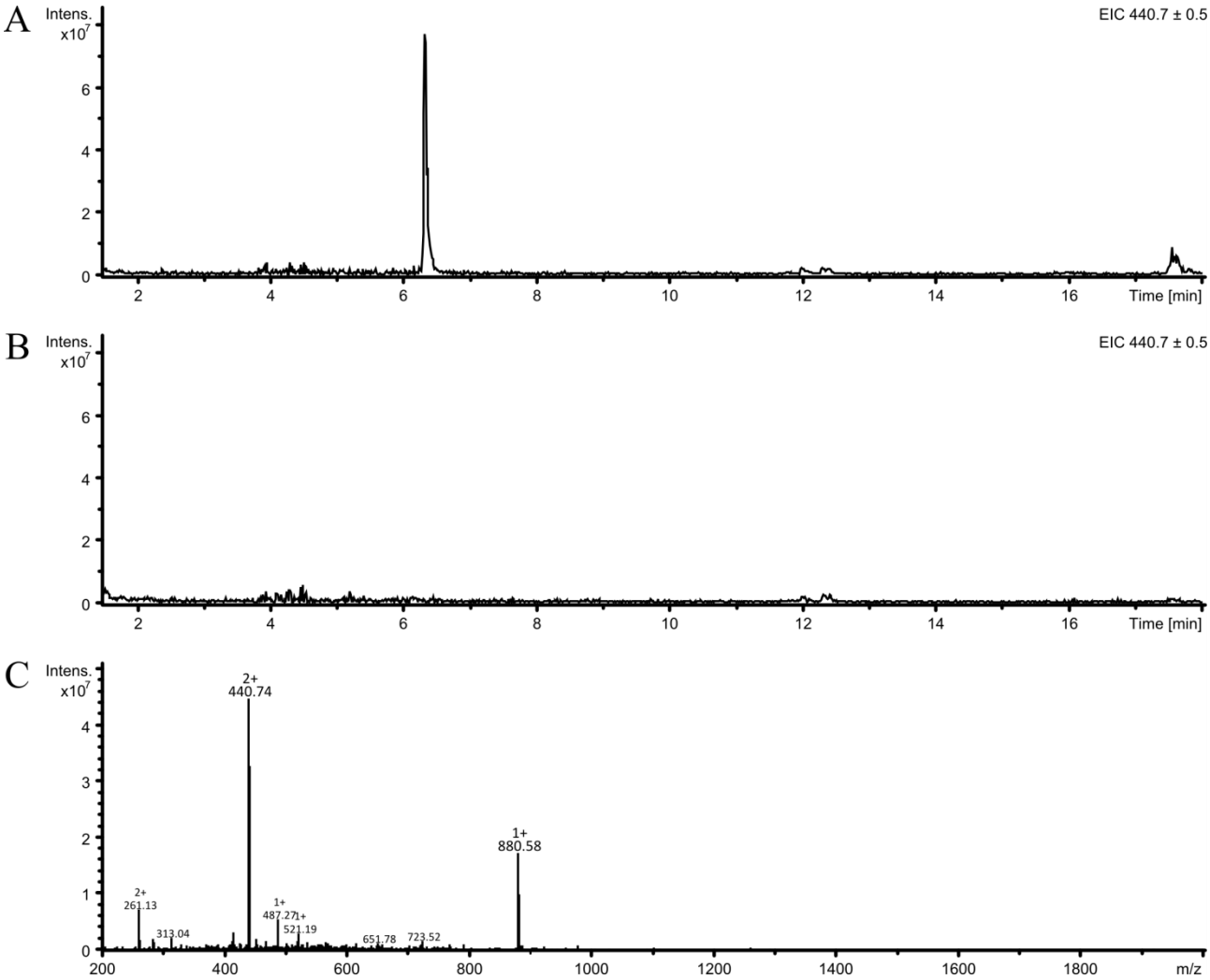

Figure 4.1. LC-MS detection of cyclohuinilsopeptin. (A) Extracted ion chromatogram 440.7 ± 0.5 of crude extract of *S. albus* 2D19 harboring the NRPS gene cluster. The peak corresponding to cyclohuinilsopeptin A has a retention time of 6.3 min. (B) Extracted ion chromatogram 440.7 ± 0.5 of crude extract of the control strain *S. albus* Del14 without the cluster. No peak at 6.3 min can be observed. (C) Mass spectrum of the peak corresponding to cyclohuinilsopeptin A (upper chromatogram; RT 6.3 min). The molecular ion peaks at m/z 880.58 and 440.74 correspond to singly charged cyclohuinilsopeptin A $[M+H]^+$ and doubly charged cyclohuinilsopeptin A $[2M+H]^+$.

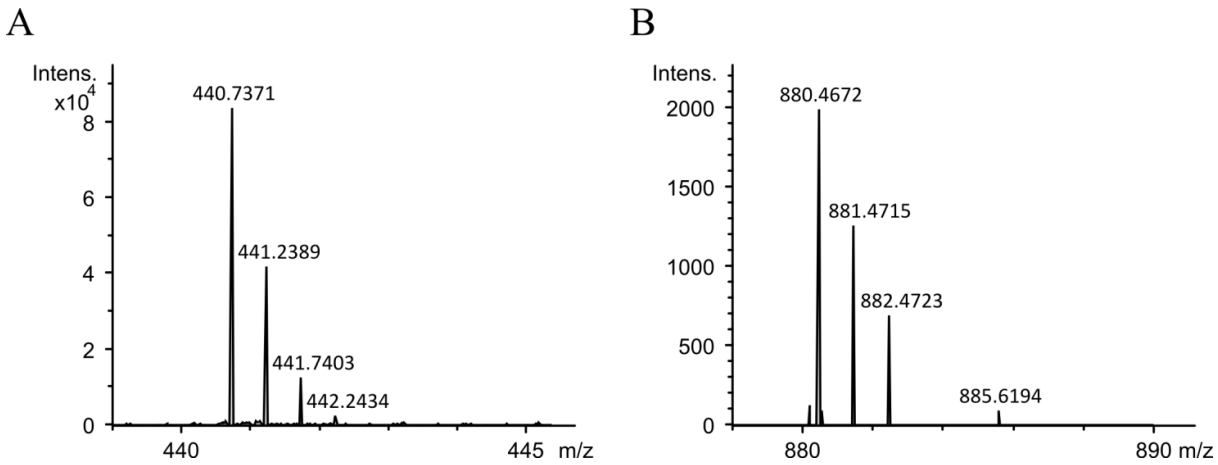

Figure 4.2. High resolution mass spectrum of the peak corresponding to cyclohuinilsopeptin. (A) Mass range 439 – 446 is shown. (B) Mass range 878 – 892 is shown.

(Figure 4.2) was used to determine the molecular formula of the obtained compound as $C_{42}H_{61}N_{11}O_{10}$. The obtained high-resolution monoisotopic mass of the compound was used for surveys in databases of natural products. These surveys did not generate any matches implying possible novelty of the identified compound.

3.2 Isolation and structure elucidation of cyclohuinilsopeptin

Considering the possible novelty of the identified compound, we set out to purify it and to solve its structure. For this purpose, the producing strain *S. albus* 2D19 was cultivated in 10 L of the production medium DNPM. The culture supernatant was extracted with an equal amount of butanol. 4 mg of the compound was purified using normal phase, size exclusion and reverse phase chromatography. The purified compound was used for structure elucidation by NMR. Structure elucidation was performed by Dr. Nils Gummerlich and Dr. Josef Zapp. The isolated compound was named cyclohuinilsopeptin A and was identified as a novel compound.

The structure of cyclohuinilsopeptin A (**1**) (Figure 4.3) was elucidated using data obtained by high-resolution mass spectrometry (**1:** *m/z* = 880.4672 $[M+H]^+$) (Figure 4.2) and nuclear magnetic resonance (NMR) spectroscopy (^{1}H-NMR, ^{13}C-NMR, HH-COSY, HSQC, HMBC, HSQC-TOCSY, ROESY, ^{15}N-HSQC). Detailed information can be found in Table 4.4 and Figures 4.4 – 4.21. As can be seen in Figure 4.3 **1** consists of the seven following subunits: **I** – β-phenylalanine; **II** – 3-guanidino-alanine; **III** – 3,4-cyclopropyl-proline; **IV** – 2-(2,2,3-trimethylcyclopropyl)-glycine; **V** – lysine; **VI** 2-methylmalonic acid and **VII** – glutamine. While utilization of the HSQC-TOCSY experiment drastically simplified the assignment of proton bound carbons to each subunit, the HMBC experiment provided detailed information about the corresponding neighbours of each subunit.

I: The elucidation of the β-phenylalanine is based on all available aromatic signals, as it is the only aromatic subunit of **1**. The β position of the amino group became evident through strong corresponding HMBC signals from proton H-21 to C-14 and from H-23 to C-11. The HMBC

signals from H-20 to C-6 and from H-21 to C-3 provided information about the corresponding C- and N- neighbours (C: **II**; N: **V**).

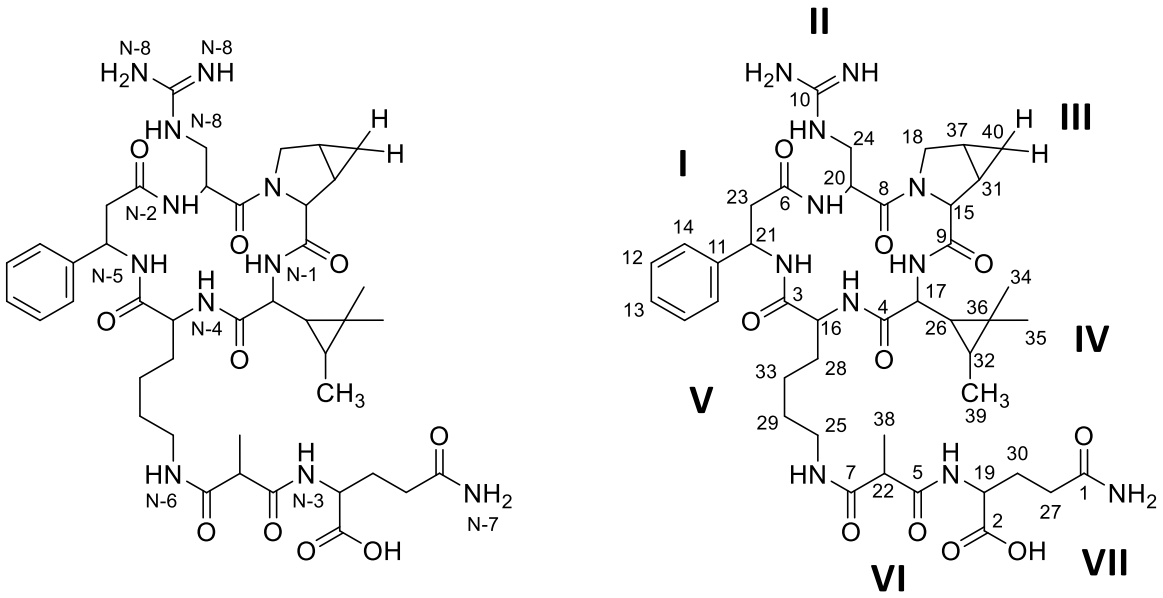

Figure 4.3. Structure of Cyclohuinilsopeptin A with numbered atoms according to the NMR data. **I** – β-phenylalanine; **II** – 3-guanidino-alanine; **III** – 3,4-cyclopropyl-proline; **IV** – 2-(2,2,3-trimethylcyclopropyl)-glycine; **V** – lysine; **VI** 2-methylmalonic acid and **VII** – glutamine.

Table 4.4 NMR spectroscopic data for Cyclohuinilsopeptins A (**1**) (DMSO-d_6, d-TFA, D_2O) and B (**2**).

Cyclohuinilsopeptin A **(1)**				Cyclohuinilsopeptin B **(2)**			
Pos.	δ_C	δ_H(*J* in Hz)	HMBC	Pos	δ_C	δ_H(*J* in Hz)	HMBC
1	174.35 C	-	-	1	175.16 C	-	-
2	173.64 C	-	-	2	174.09 C	-	-
3	171.90 C	-	-	3	172.96 C	-	-
4	171.63 C	-	-	4	172.45 C	-	-
5	171.03 C	-	-	5	171.67 C	-	-
6	170.77 C	-	-	6	171.02 C	-	-
7	170.54 C	-	-	7	170.98 C	-	-
8	169.85 C	-	-	8	170.90 C	-	-
9	169.62 C	-	-	9	170.21 C	-	-
10	157.30 C	-	-	10	157.53 C	-	-
11	142.99 C	-	-	11	143.08 C	-	-
12	128.81 CH	7.29 m (2H)	11, 13, 14	12,12´	129.18 CH	7.29 m (2H)	
13	127.34 CH	7.19 m	11, 12, 14	13	127.78 CH	7.19 m	
14	126.13 CH	7.25 m (2H)	12, 13, 21	14,14´	126.46 CH	7.23 m (2H)	
15	62.62 CH	4.69 d (5.95)	8, 9, 18, 31, 40	15	63.27 CH	4.68 d (6.0)	
16	56.6 CH	3.68 m	3, 4, 28, 33	16	57.84 CH	3.57 t (7.5)	
17	53.8 CH	4.03 dd (10.20)	4, 9, 26, 32, 36	17	52.64 CH2	3.69 dd (12.0, 5.7) 3.46 d (12.0)	
18	52.40 CH2	3.75 - 3.44 m	8, 15, 31, 40	18	52.30 CH	4.14 dd (9.0, 5.0)	
19	51.97 CH	4.15 m	2, 5, 27, 30	19	51.50 CH	4.48 dd (11.0, 3.5)	
20	51.05 CH	4.19 m	6, 8, 24	20	51.24 CH	4.30 t (6.0)	
21	50.75 CH	4.89 m	3, 6, 11, 14, 23	21	51.09 CH	4.92 dd (12.0, 4.5)	
22	46.51 CH	3.14 m	5, 7, 38	22	47.18 CH	3.14 q (7.0)	
23	42.12 CH2	2.67 - 2.47 m	6, 11, 21	23	42.57 CH2	2.61 dd (14.5, 4.5) 2.49 m	

Pos.	δ_N	δ_H(*J* in Hz)	HMBC	Pos.	δ_N	δ_H(*J* in Hz)	HMBC
24	41.98 CH2	3.43 - 3.37 m	8, 10, 20	24	42.20 CH2	3.42 dd (14.5, 6.5) 3.34 dd (14.5, 5.5)	
25	38.94 CH2	2.99 m	7, 29,33	25	40.74 CH2	1.73 m 1.51 m	
26	32.74 CH	0.79 m	4, 17, 32, 36, 34, 35, 39	26	39.2* CH2	2.97 m (2H)	
27	31.49 CH2	2.09 m	1, 19, 30	27	31.76 CH2	2.12 t (7.5, 2H)	
28	29.92 CH2	1.82 - 1.67 m	16, 29, 33	28	29.95 CH2	1.76 m 1.68 m	
29	28.67 CH2	1.37 m	25, 28, 33	29	28.95 CH2	1.34 m (2H)	
30	27.34 CH2	1.94 - 1.76 m	1, 2, 27, 19	30	27.26 CH2	1.95 m 1.76 m	
31	23.9 CH	2.14 m	15, 37, 40	31	25.32 CH	1.51 m	
32	23.47 CH	0.36 dt (6.5. 5.6)	17, 26, 34, 35	32	23.86 CH3	0.89 d (6.9, 3H)	
33	23.40 CH2	1.30 - 1.20 m	16,28,29,25	33	23.79 CH2	1.20 m 1.09 m	
34	22.43 CH3	1.08 d (0.77)	26, 32, 35, 36, 39	34	23.65 CH	2.12 t (7.7)	
35	21.25 CH3	1.03 d (0.77)	26, 32, 34, 36	35	21.51 CH3	0.83 d (6.0, 3H)	
36	21.10 C	-	-	36	16.15 CH	1.70 m	
37	15.83 CH	1.71 m	-	37	15.44 CH3	1.16 d (7.0, 3H)	
38	15.20 CH3	1.14 d (6.67)	5, 7, 22	38	11.23 CH2	0.71 m 0.36 q (4.0)	
39	13.40 CH3	0.99 d (6.42)	26, 34, 35, 36				
40	10.82 CH2	0.75 - 0.41 m	15, 18, 31, 40				
N1	122.68 NH	8.27 m	9	N1	120.3 NH	8.21 (und 8.12)	
N2	121.05 NH	8.24 m	6	N2	119.7 NH	8.11 m	
N3	120.20 NH	8.17 - 8.09 d (7.71)	5	N3	119.0 NH	8.26 t (6.0)	
N4	118.51 NH	7.66 d (6.71)	4	N4	117.5 NH	7.83 dd (9.5, 3.5)	
N5	117.97 NH	8.11 m	3	N5	117.1 NH	8.35 d (6.5)	
N6	117.24 NH	7.78 - 7.69 t (5.55)	7	N6	117.2 NH	7.81 m	
N7	109.1 NH2	7.30 - 6.71 d (br)	-	N7	116.7 NH	7.70	
N8	77.60 NH	7.44 t (6.04)	-	N8	108.8 NH2	7.30 d 6.79 d	
				N9	77.7 NH	7.39 t	
				N10	22.8 NH/NH2	7.12 s (und 7.01?)	

II: The HSQC-TOCSY of 3-guanidino-alanine showed only the α and β atoms of the corresponding amino acid. Nevertheless, H-24 showed a clear correlation towards C-10 of the guanidinium moiety, with the very characteristic chemical shift of 157 ppm. Furthermore, the observed triplet at 7.44 ppm in the ^{1}H-NMR showed a strong peak at 77.6 ppm in the ^{15}N-HSQC, drastically differing from observed amide nitrogen peaks. The HMBC signals from H-18 to C-8 and from H-20 to C-6 provided information about the corresponding C- and N- neighbours (C: **III**; N: **I**).

III: The continuous TOCSY coupling network and the high proton density of the 3,4-cyclopropyl-proline simplified the structure elucidation of this subunit. Especially the drastically upfield shifted H-40 (0.75-0.41 ppm) and its corresponding carbon (10.82 ppm) confirmed the presence of a cyclopropane moiety. The HMBC signals from H-17 to C-9 and from H-18 to C-8 provided information about the corresponding C- and N- neighbours (C: **II**; N: **IV**).

IV: The highly unusual amino acid 2-(2,2,3-trimethylcyclopropyl)-glycine consists of a glycine subunit and a three times methylated cyclopropane. The three methyl groups are clearly visible in the aliphatic region of the spectrum. While H-39 is a clear split doublet with a coupling constant of 6.67 Hz, H-34 and H-35 share a similar coupling constant of 0.77 Hz. Furthermore, all expected HMCB and HSQC-TOCSY correlations were observed. The HMBC signals from H-16 to C-4 and from H-17 to C-9 provided information about the corresponding C- and N- neighbours (C: **V**; N: **III**).

V: Lysine is one of two proteinogenic amino acids in the described structure. Using HMBC and HSQC-TOCSY, all correlations and chemical shifts were as expected. The correlations H-21 to C-3 and H-16 to C4 confirmed the C- and N- neighbours (C: **I**; N: **IV**). Furthermore, a correlation from H-25 to C-7 proved a bond between the terminal nitrogen of lysine subunit and the 2-methylmalonic acid subunit.

VI: 2-methylmalonic acid is the only subunit of **1** which is not an amino acid. The correlations between H-25 to C-7, H-19 to C-5 and H-38 to C-22, C-5 and C-7 confirm the structure as well as the two neighbour units **V** and **VII**.

VII: Glutamine is the only terminal subunit of **1**, as we were not able to observe a correlation from H-27 to a neighbouring subunit. **VI** was confirmed as N-terminal neighbour due to a correlation between H-19 and C-5. Otherwise, all observed correlations and chemical shifts were in expected ranges.

To confirm the proposed structure, a ^{15}N-HSQC was measured. All expected nitrogen atoms were visible and in expected ranges. Furthermore, it is worth mentioning that **1** appears to have conformational isomers, as a peak splitting of all carbons of units **V**, **VI** and **VII** was clearly visible in the ^{13}C-NMR. This indicates two stable conformational isomers caused by two positional changes of the side chain formed by the subunits **V**, **VI** and **VII**.

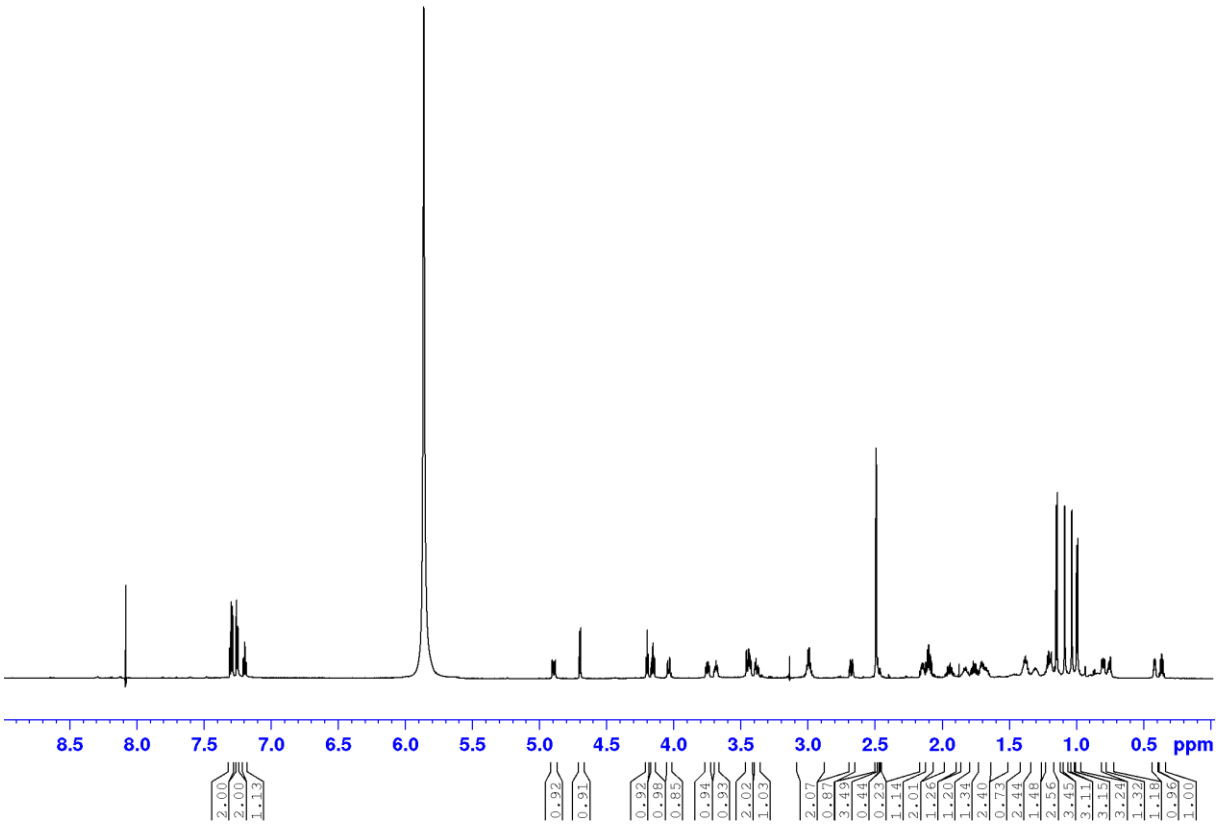

Figure 4.4. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6, d-TFA, D_2O*) of **1**; complete Spectrum.

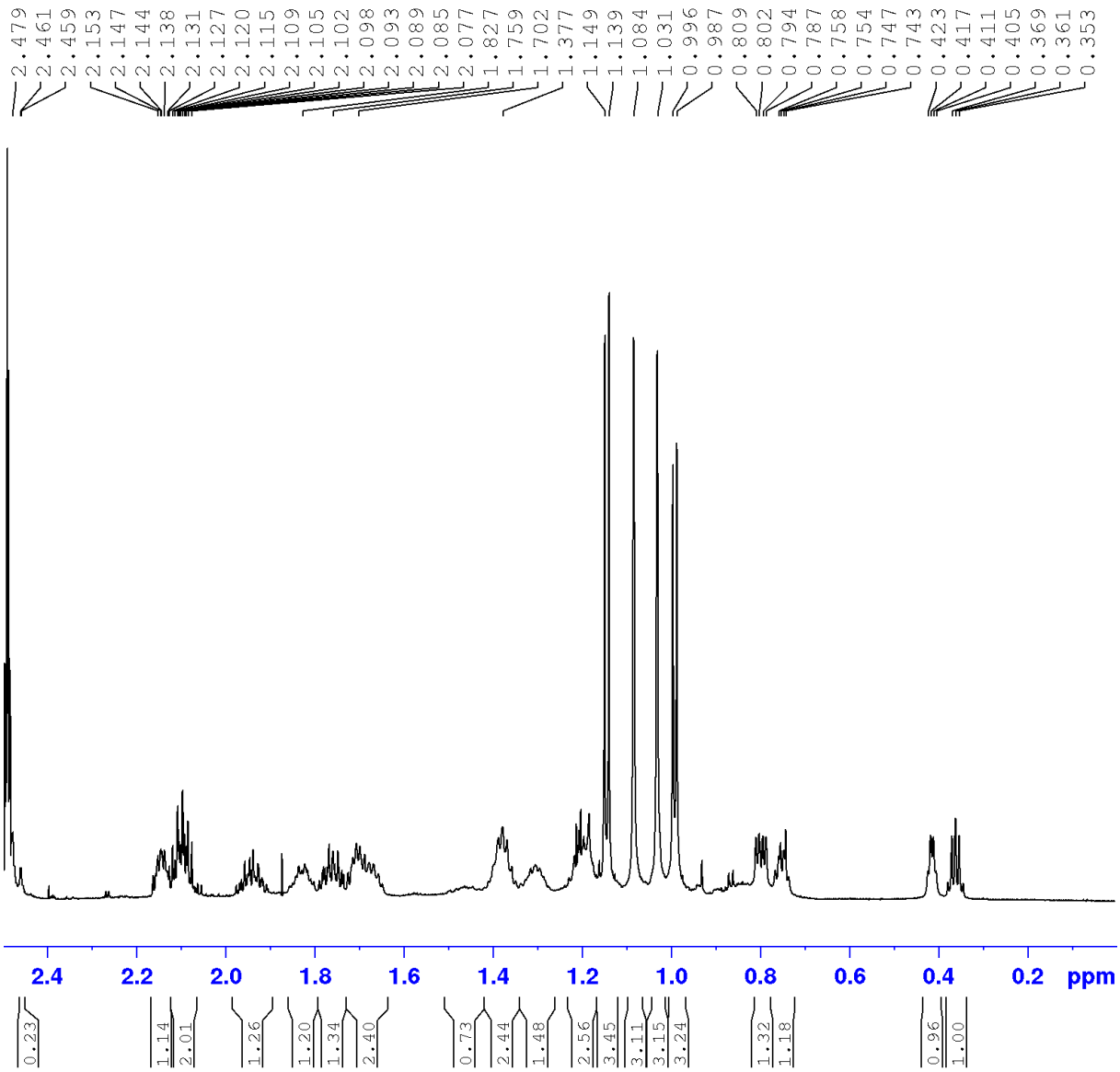

Figure 4.5. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**; zoom from 2.5 ppm to 0ppm.

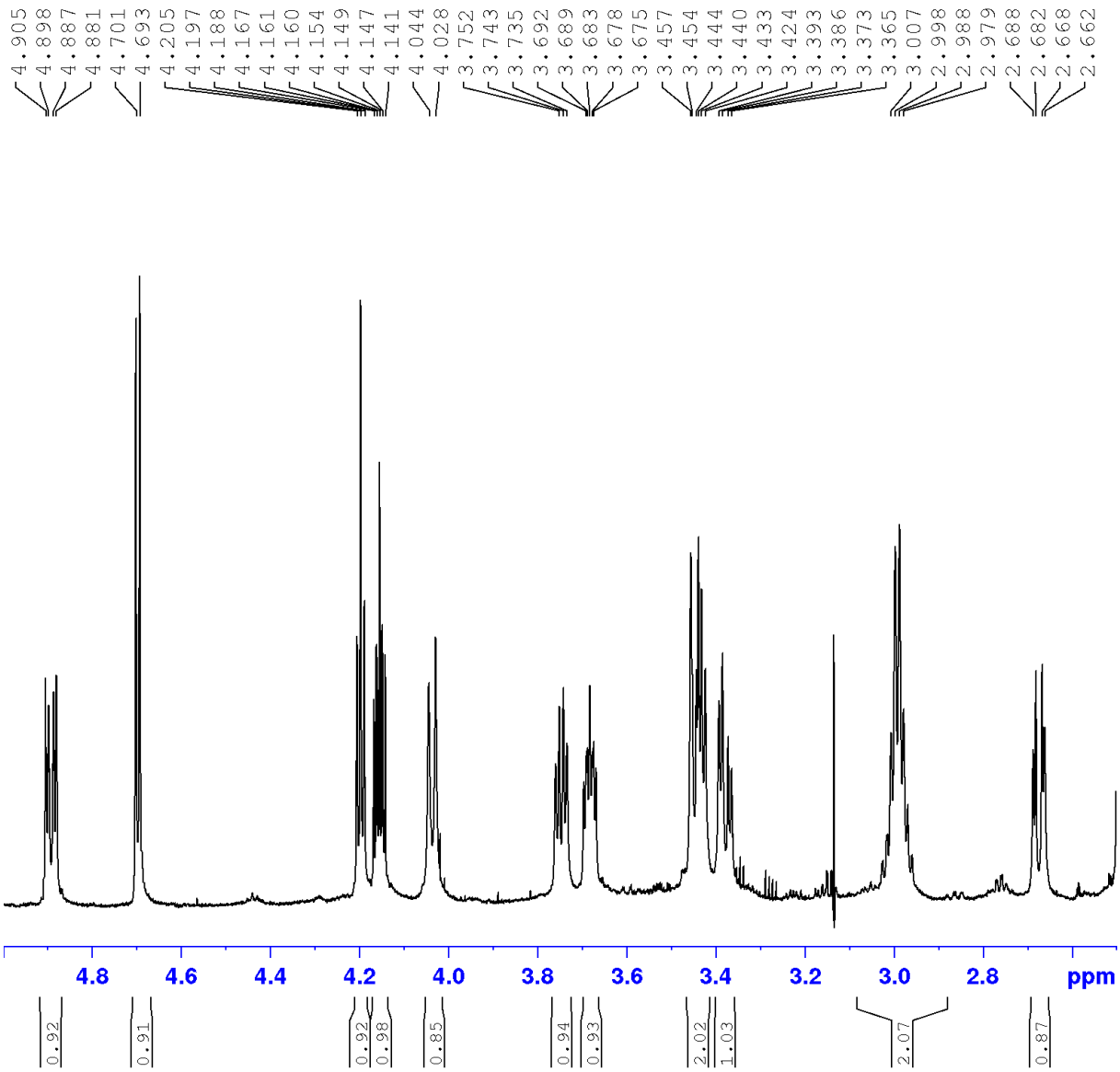

Figure 4.6. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**; zoom from 5 ppm to 2.5 ppm.

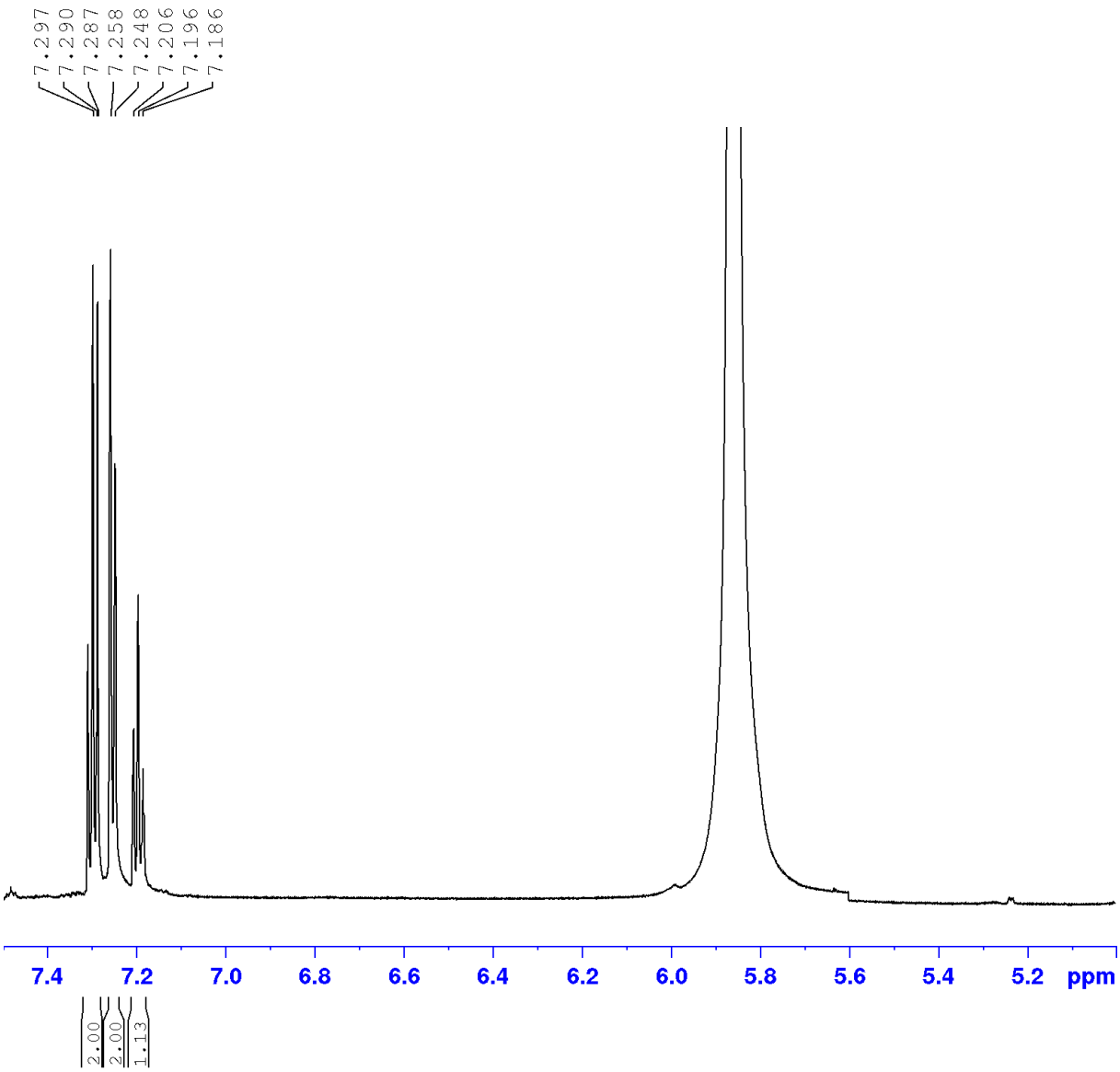

Figure 4.7. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**; zoom from 7.5 ppm to 5 ppm.

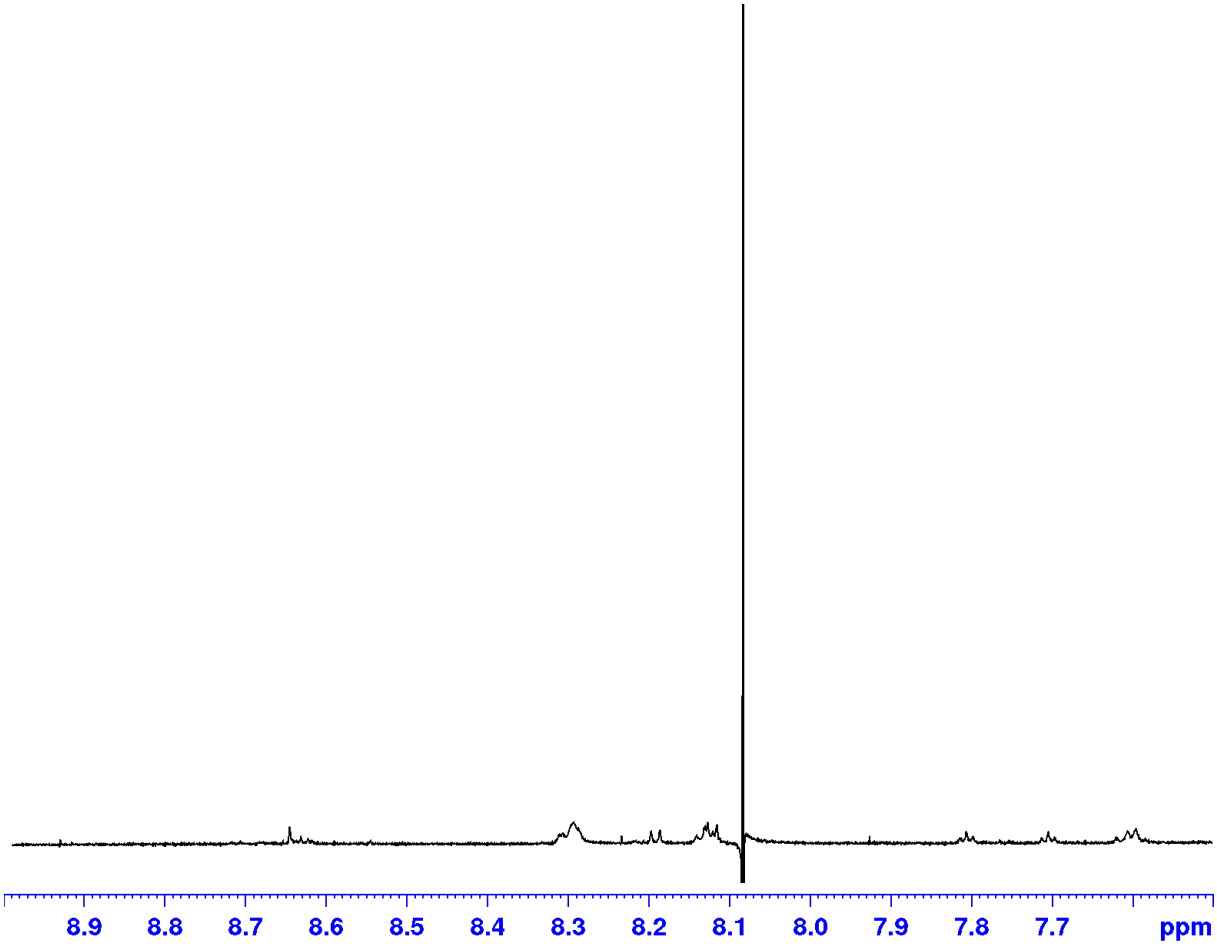

Figure 4.8. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6, d-TFA, D_2O*) of **1**; zoom from 9 ppm to 7.5 ppm.

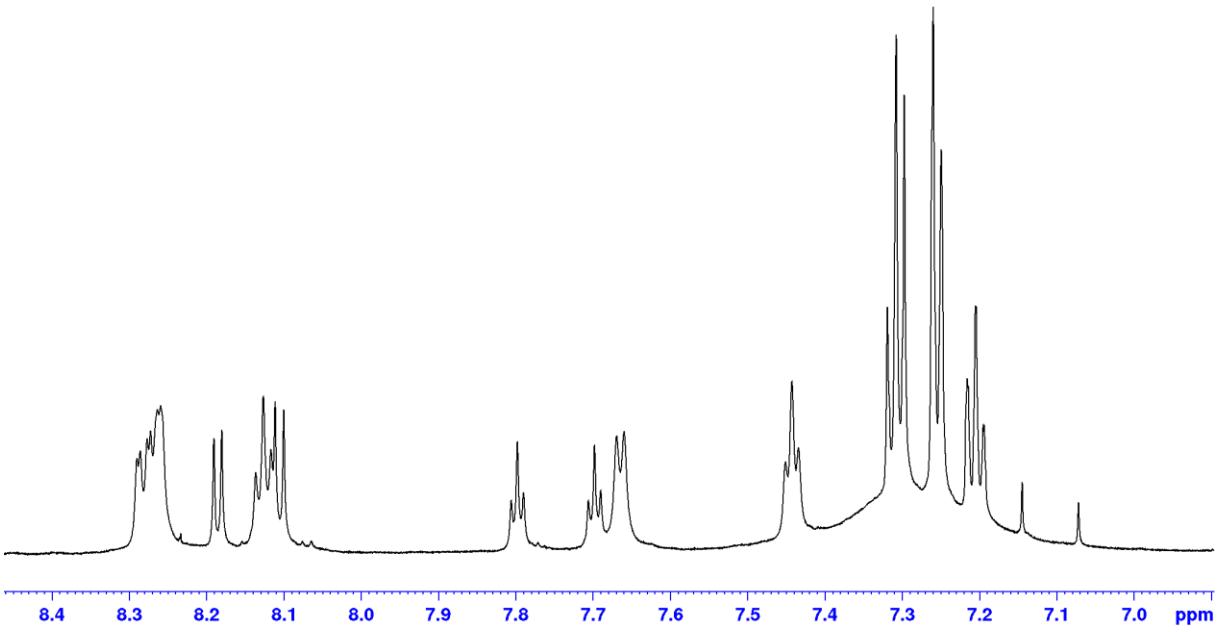

Figure 4.9. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6, TFA, H_2O*) of **1**; zoom from 8.5 ppm to 6.9 ppm.

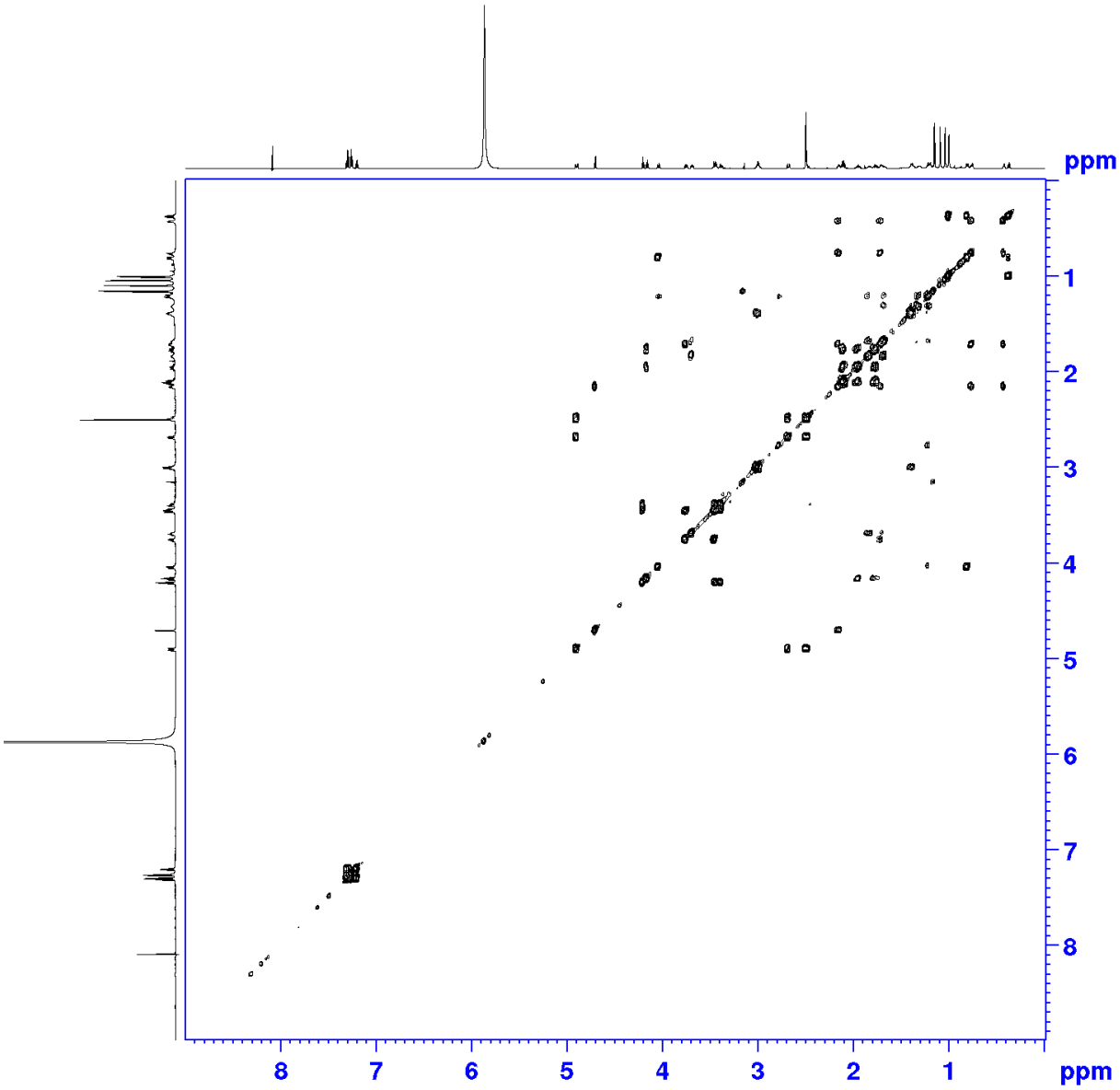

Figure 4.10. ^{1}H-^{1}H- COSY spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, complete spectrum.

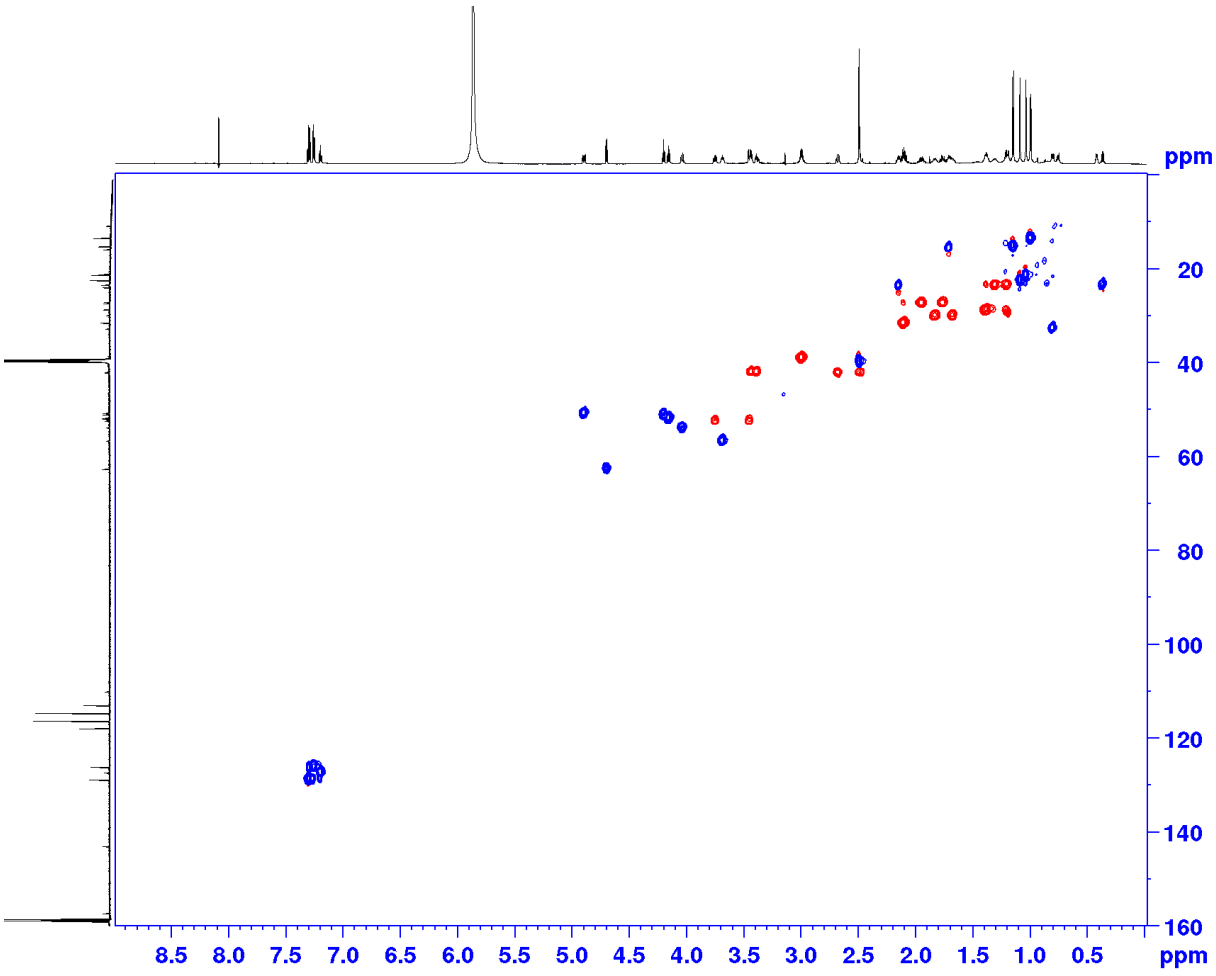

Figure 4.11. HSQC-spectrum (700 MHz, 175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, complete spectrum.

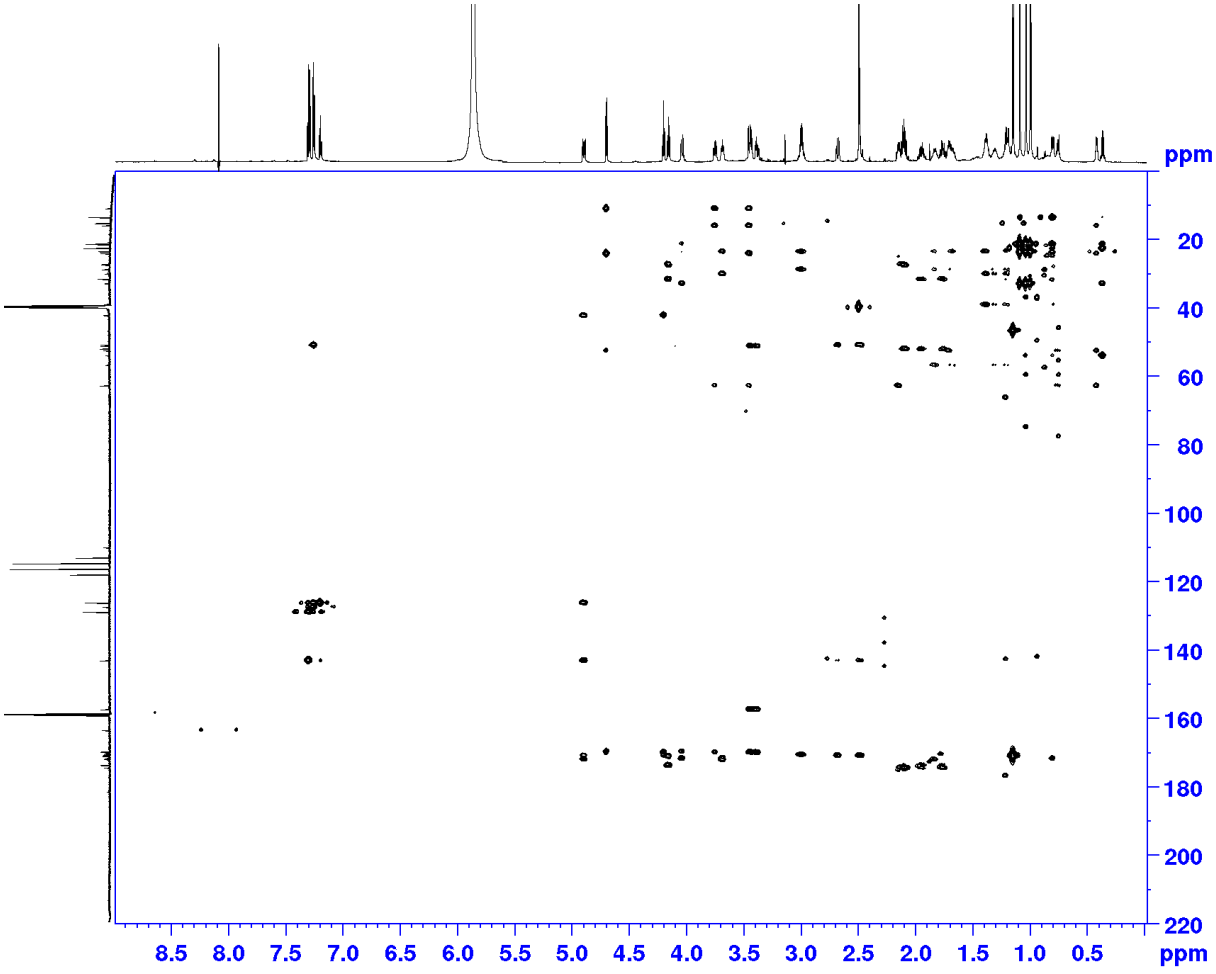

Figure 4.12. HMBC-spectrum (700 MHz, 175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, complete spectrum.

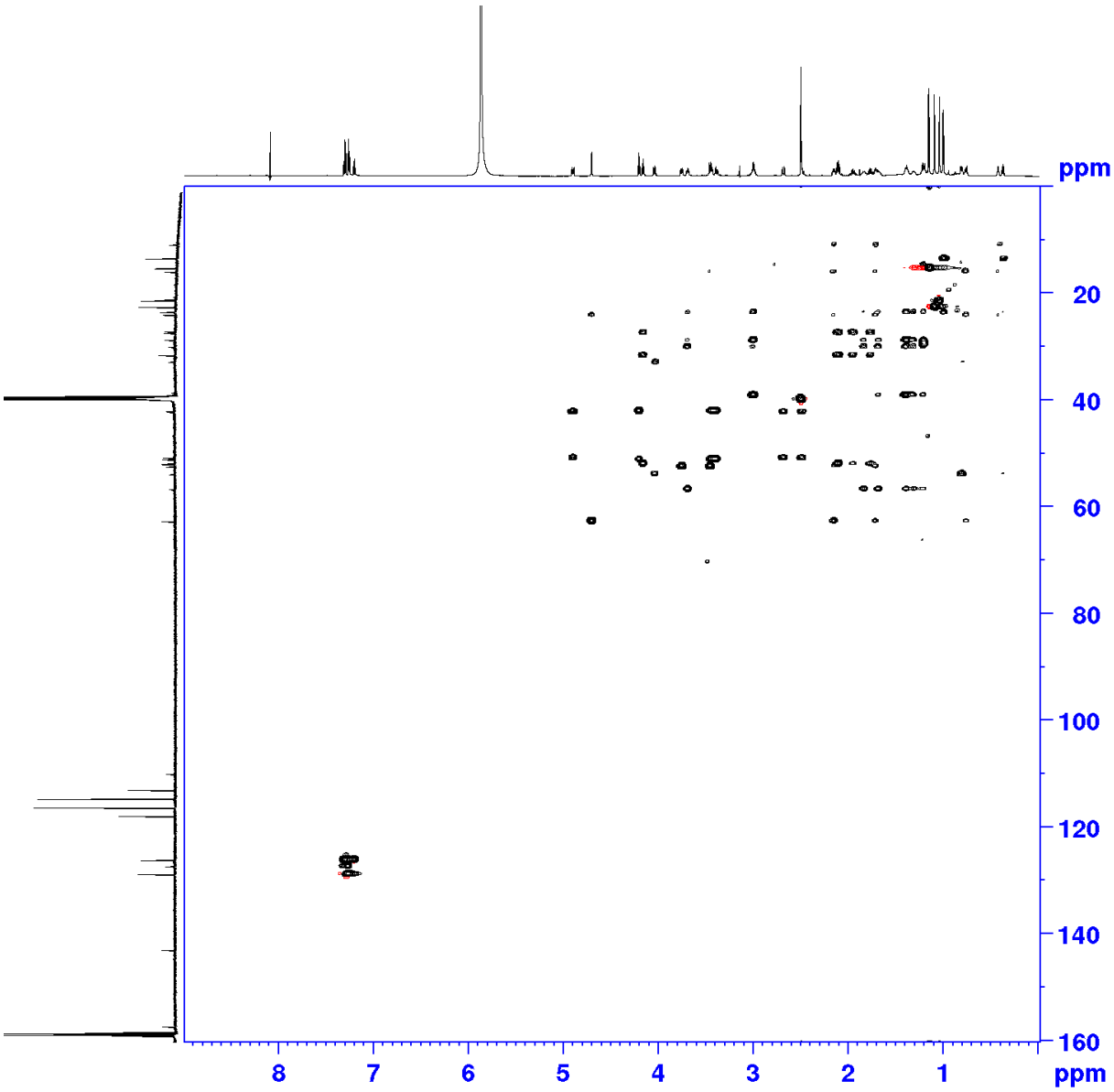

Figure 4.13. HSQC-TOCSY-spectrum (700 MHz, 175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, complete spectrum.

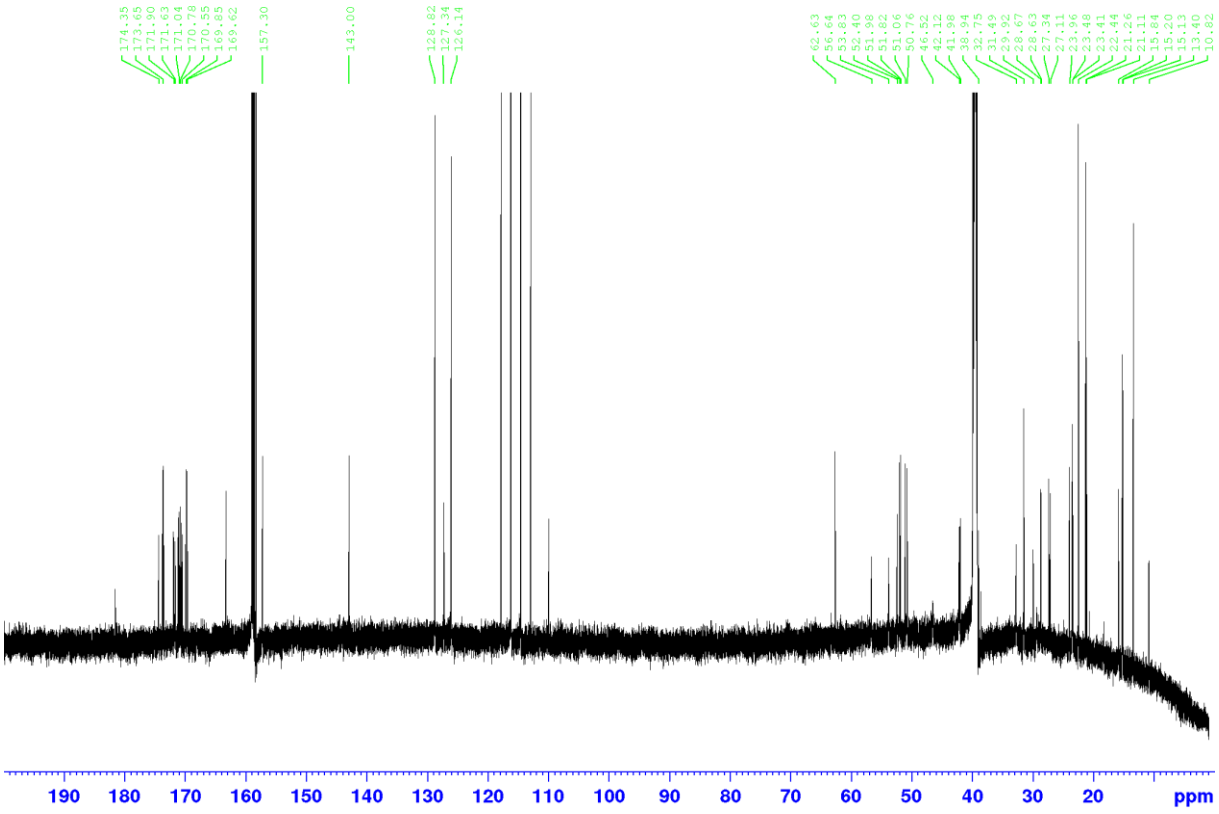

Figure 4.14. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, complete spectrum.

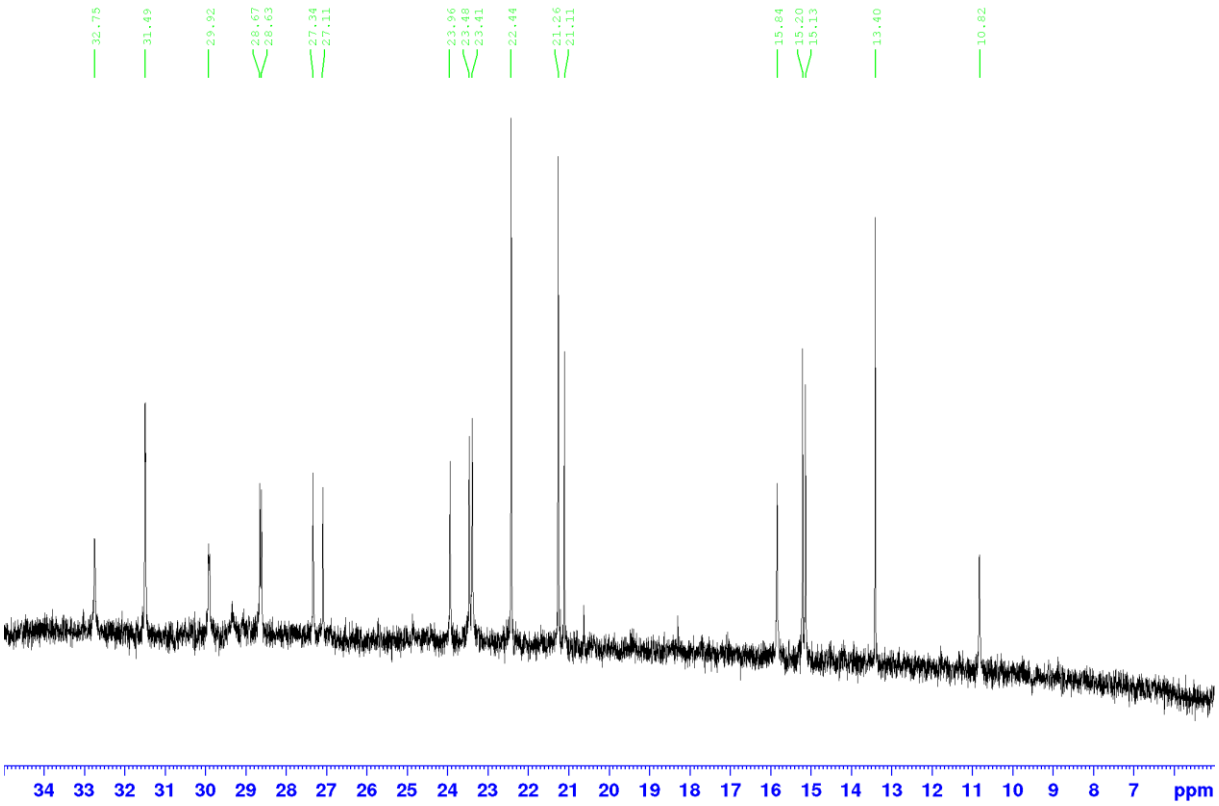

Figure 4.15. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 35 ppm to 0 ppm.

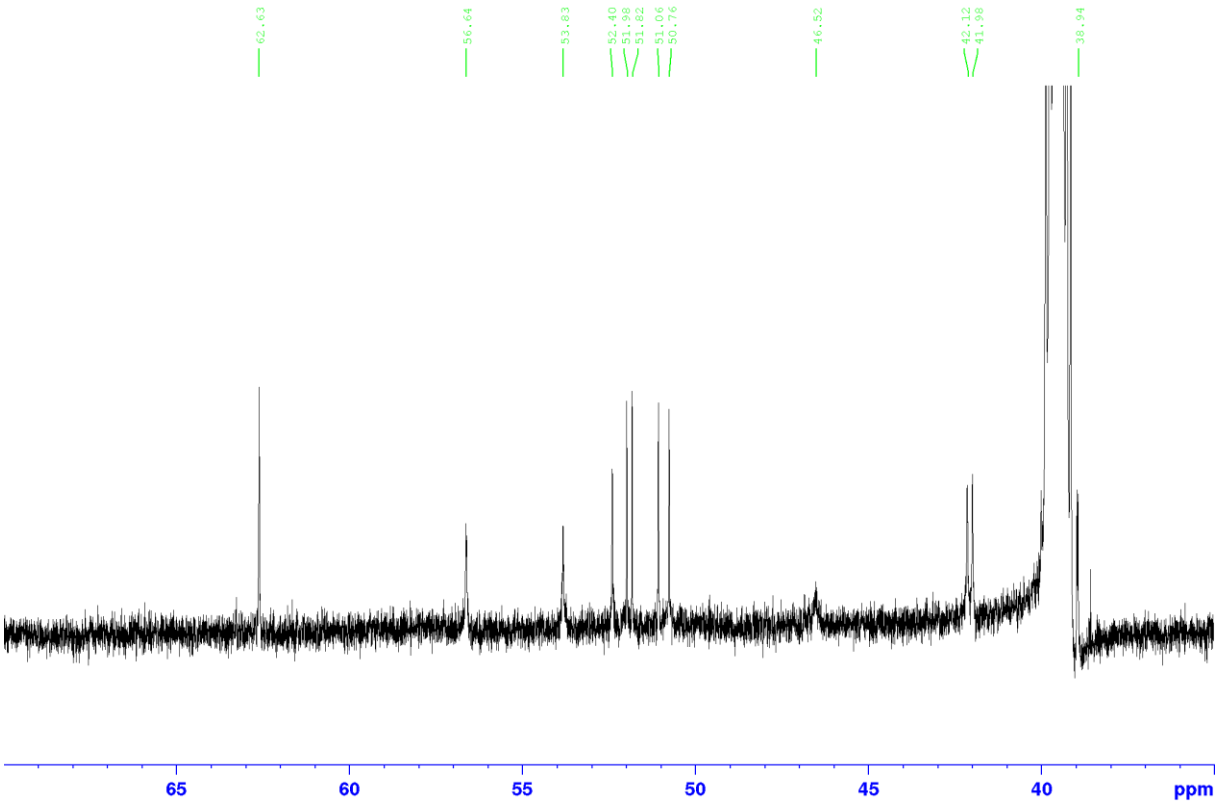

Figure 4.16. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 70 ppm to 35 ppm.

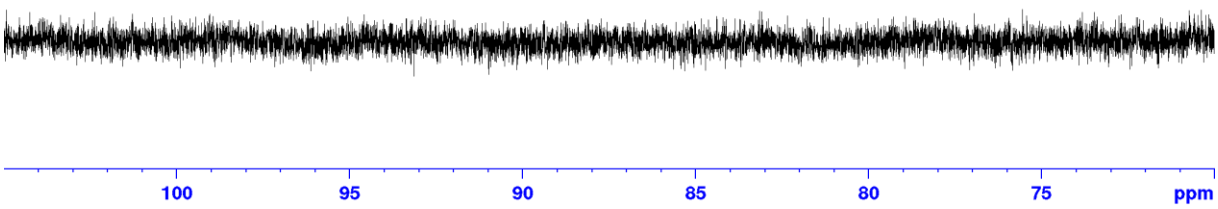

Figure 4.17. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 105 ppm to 70 ppm.

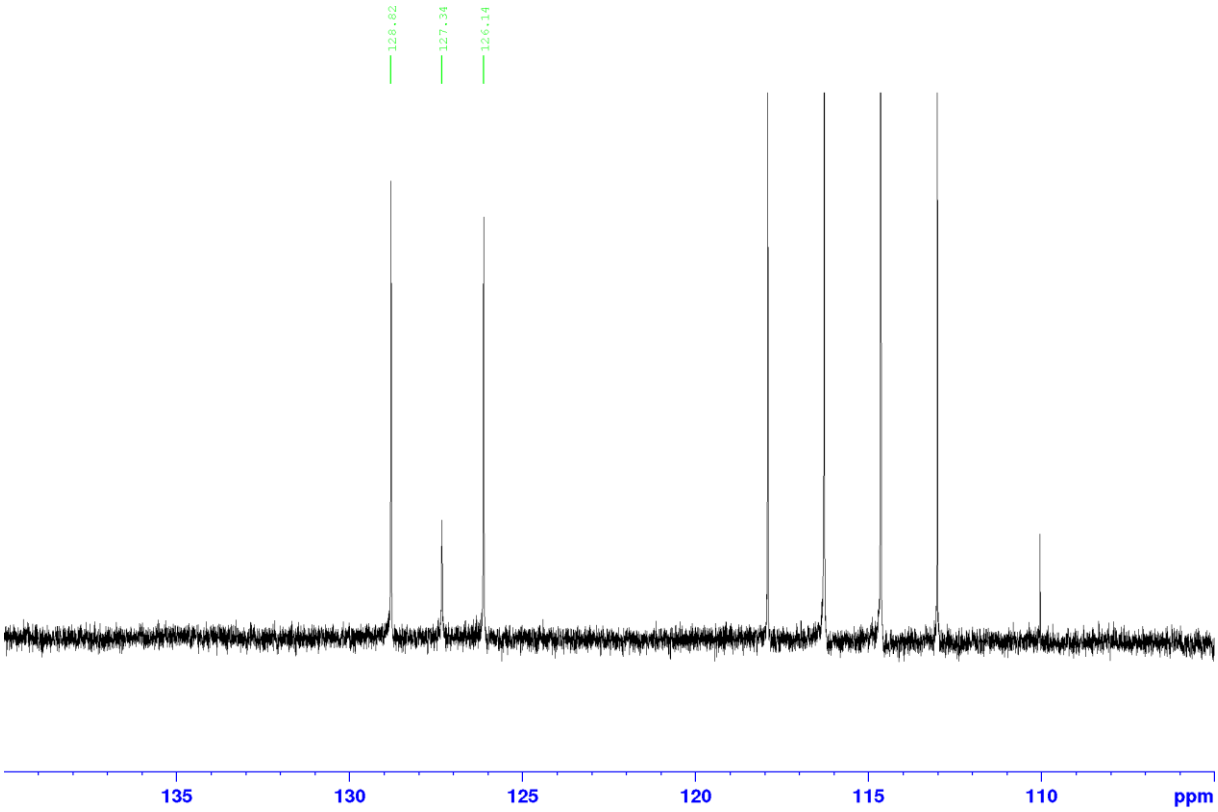

Figure 4.18. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 140 ppm to 105 ppm.

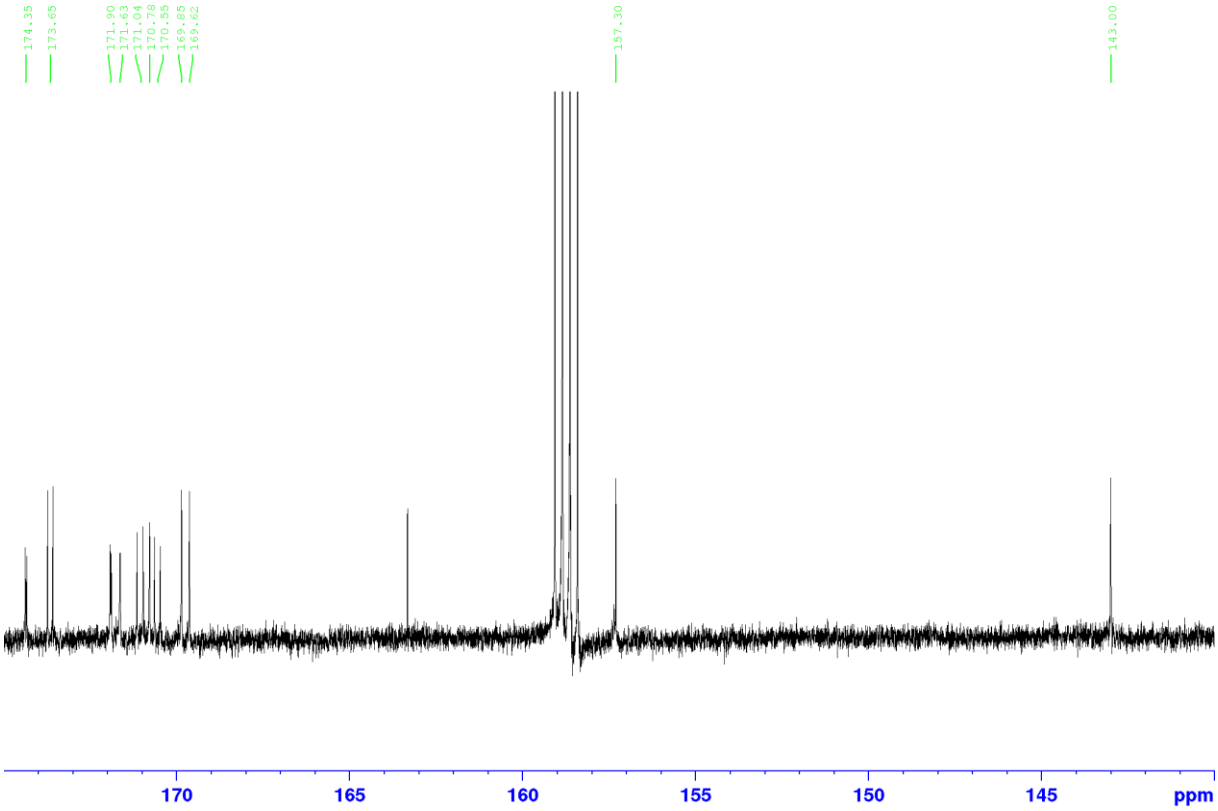

Figure 4.19. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 175 ppm to 140 ppm.

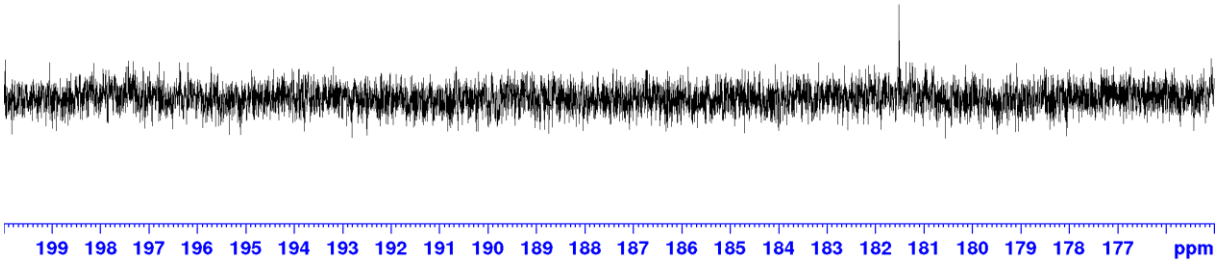

Figure 4.20. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **1**, zoom from 200 ppm to 175 ppm.

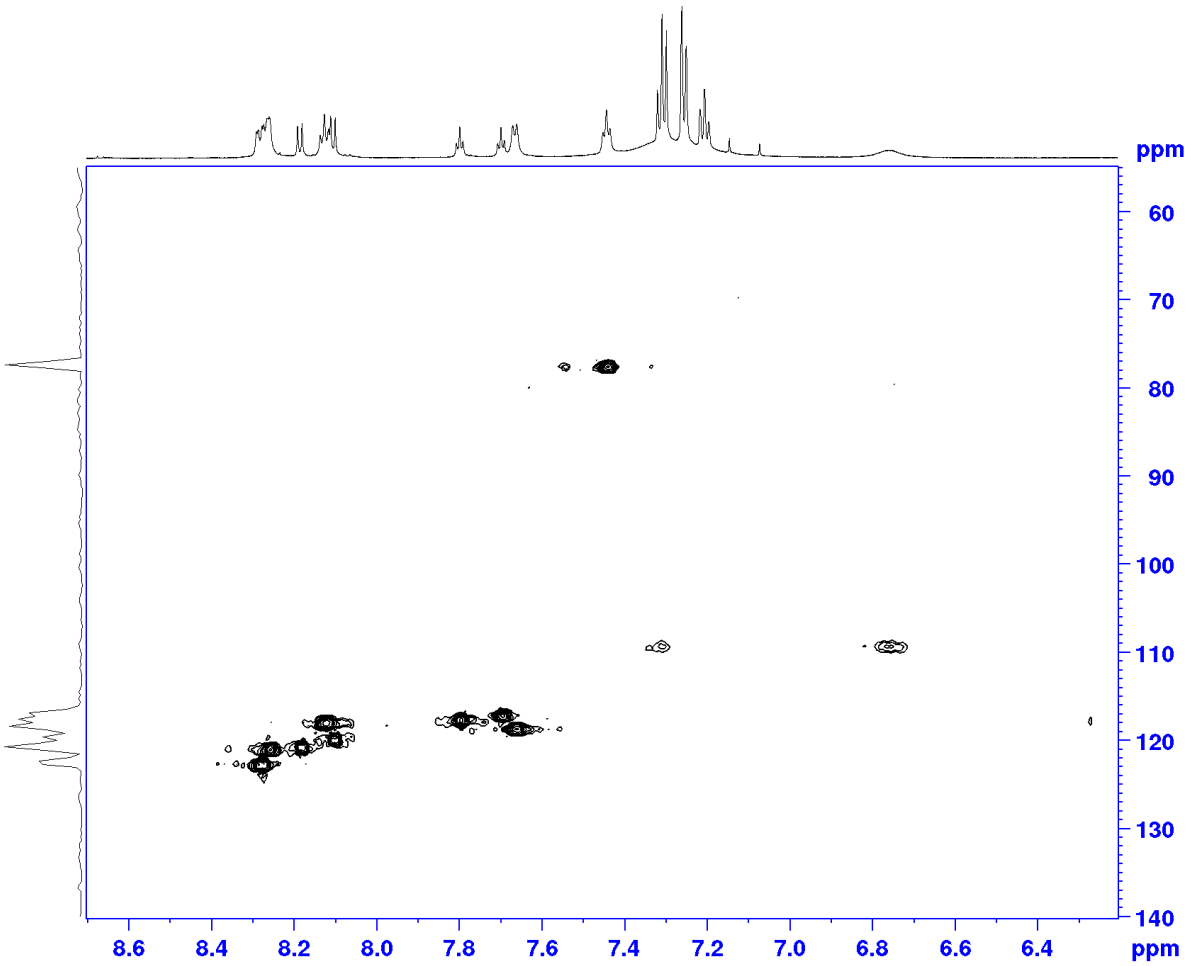

Figure 4.21. ^{15}N-HSQC (50 MHz, *DMSO-d_6*, *TFA*, *H_2O*) of **1**, complete spectrum.

3.3 Identification of the cyclohuinilsopeptin A gene cluster through sequence analysis

The BAC 2D19 which leads to cyclohuinilsopeptin A production in *S. albus* contains a 95 kb fragment of *K. albida* chromosome. 72 open reading frames, KALB_5484 to KALB_5553, were annotated within this region (Figure 4.22). The region comprising the genes KALB_5507 - KALB_5522 was predicted to constitute the cyclohuinilsopeptin A biosynthetic gene cluster using the results of sequence analysis with Antismash and BLASTP software (Figure 4.22, Table 4.5). The peptide nature of the isolated cyclohuinilsopeptin A implies the essential role of the genes KALB_5510, KALB_5512, KALB_5513, which encode NRPS enzymes, in the production of the compound. The NRPS encoding genes KALB_5512 and KALB_5510 are organized in a single operon with the genes KALB_5511, KALB_5509 and KALB_5508 implying the possible involvement of the latter in cyclohuinilsopeptin A production. This assumption is strengthened by the fact that that the genes KALB_5509 and KALB_5508 encode putative aminomutase and TrpB-like enzyme which are involved amino acid metabolism. The gene KALB_5507 encodes a putative transcriptional regulator and is assumed to constitute the 5′ outer border of the cyclohuinilsopeptin A cluster. The five genes upstream of KALB_5507 encode a putative transposase and hypothetical genes of unknown function (Table 4.5) and are not considered to be involved in biosynthesis of cyclohuinilsopeptin.

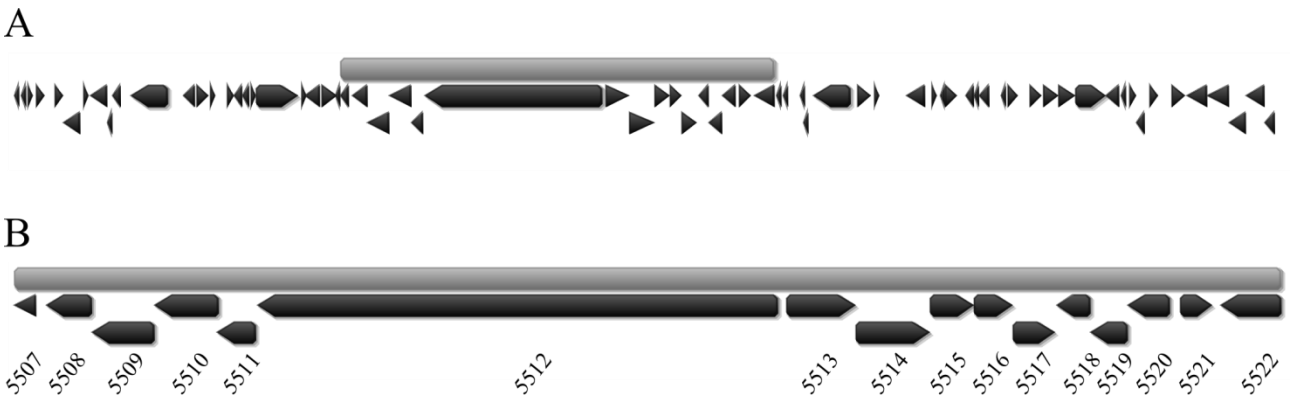

Figure 4.22. Fragment of *K. albida* chromosome containing cyclohuinilsopeptin A gene cluster. A – DNA fragment cloned in the BAC 2D19. B – predicted cyclohuinilsopeptin A gene cluster Genes KALB_5507 - KALB_5522 are shown with numbers 5507 - 5522.

Table 4.5 Predicted products of the genes within cyclohuinilsopeptin A gene cluster and its flanking regions.

Gene	Predicted function	Gene	Predicted function
KALB_5502	tetratricopeptide repeat protein	KALB_5515	acyl-CoA dehydrogenase
KALB_5503	hypothetical protein	KALB_5516	ornithine cyclodeaminase
KALB_5504	hypothetical protein	KALB_5517	transposase
KALB_5505	DUF2029 domain-containing protein	KALB_5518	hypothetical protein
KALB_5506	transposase	KALB_5519	fatty acid desaturase
KALB_5507	transcriptional regulator	KALB_5520	amidinotransferase
KALB_5508	TrpB-like enzyme	KALB_5521	GNAT family N-acetyltransferase
KALB_5509	aminomutase	KALB_5522	fatty acid CoA ligase
KALB_5510	NRPS	KALB_5523	transposase
KALB_5511	ketoacyl-ACP synthase III	KALB_5524	hypothetical protein
KALB_5512	NRPS	KALB_5525	hypothetical protein
KALB_5513	NRPS	KALB_5526	hypothetical protein
KALB_5514	radical SAM protein	KALB_5527	hypothetical protein

The genes KALB_5513 – KALB_5517 constitute a single operon and presumably belong to the cyclohuinilsopeptin A gene cluster. As mentioned before, the gene KALB_5513 encodes a putative NRPS, while the genes KALB_5514, KALB_5515, KALB_5516 and KALB_5517 encode putative radical SAM protein, acyl-CoA dehydrogenase, ornithine cyclodeaminase and transposase respectively. The putative radical SAM enzyme might participate in the formation of trimethylcyclopropyl moiety of the compound (Figure 4.3). The putative ornithine deaminase might be involved in the biosynthesis of cyclopropyl-proline residue of the pentapeptide cycle of cyclohuinilsopeptin. No particular role in the biosynthesis of the compound could be attributed to the gene KALB_5517 which encodes a putative transposase. Nevertheless, the gene KALB_5517 is considered as a part of the cluster, as it is the last gene in the KALB_5513 – KALB_5517 operon.

The genes KALB_5513 – KALB_5517 are followed by a three-gene operon – KALB_5520-KALB_5518 (Figure 4.22). These genes encode putative amidinotransferase, fatty acid desaturase and a hypothetical protein (Table 4.5). The putative amidinotransferase might be involved in the formation of 3-guanidino-alanine in the structure of cyclohuinilsopeptin A by transferring the amidino group from L-arginine to the β-amino group of aminoalanine. For this reason, the KALB_5520-KALB_5518 operon is considered necessary for cyclohuinilsopeptin A production.

The genes KALB_5521 and KALB_5522 encode putative GNAT family N-acetyltransferase and fatty acid CoA ligase. The products of these genes might activate methylmalonate and catalyze its incorporation into the side chain of cyclohuinilsopeptin.

The KALB_5522 gene is regarded as a 3′ outer border of the cyclohuinilsopeptin A gene cluster. The five genes downstream of KALB_5522 encode a putative transposase and four hypothetical proteins (Figure 4.22, Table 4.5). No function in the biosynthesis of cyclohuinilsopeptin A could be assigned to these genes.

3.4 Validation of the identified cyclohuinilsopeptin A gene cluster through gene inactivation

The genes KALB_5507 - KALB_5522 were predicted to constitute the cyclohuinilsopeptin A gene cluster. In order to validate the participation of these genes in cyclohuinilsopeptin A biosynthesis the genes KALB_5508, KALB_5512, KALB_5514, KALB_5515, KALB_5516, KALB_5518, KALB_5519, KALB_5520, KALB_5521, KALB_5522 were inactivated in the BAC 2D19 which was used for heterologous production of cyclohuinilsopeptin A in *S. albus*. Individual genes were substituted with the hygromycin resistance cassette using RedET approach. The constructed BACs 2D19_5508, 2D19_5512, 2D19_5514, 2D19_5515, 2D19_5516, 2D19_5518, 2D19_5519, 2D19_5520, 2D19_5521, and 2D19_5522 with the deletions of the above mentioned genes were transferred into the heterologous host strain *S. albus* Del14 by conjugation. The obtained exconjugants strains as well as the control strain *S. albus* 2D19 were checked for the production of cyclohuinilsopeptin A by cultivating in production medium DNPM, extraction and

LC-MS analysis. No cyclohuinilsopeptin A could be detected in the extracts originating from the strains harboring the BACs with the inactivation of the genes KALB_5508, KALB_5512, KALB_5514, KALB_5515, KALB_5516, KALB_5518, KALB_5519 and KALB_5521 (Figure 4.23) indicating that these genes are involved in cyclohuinilsopeptin A biosynthesis.

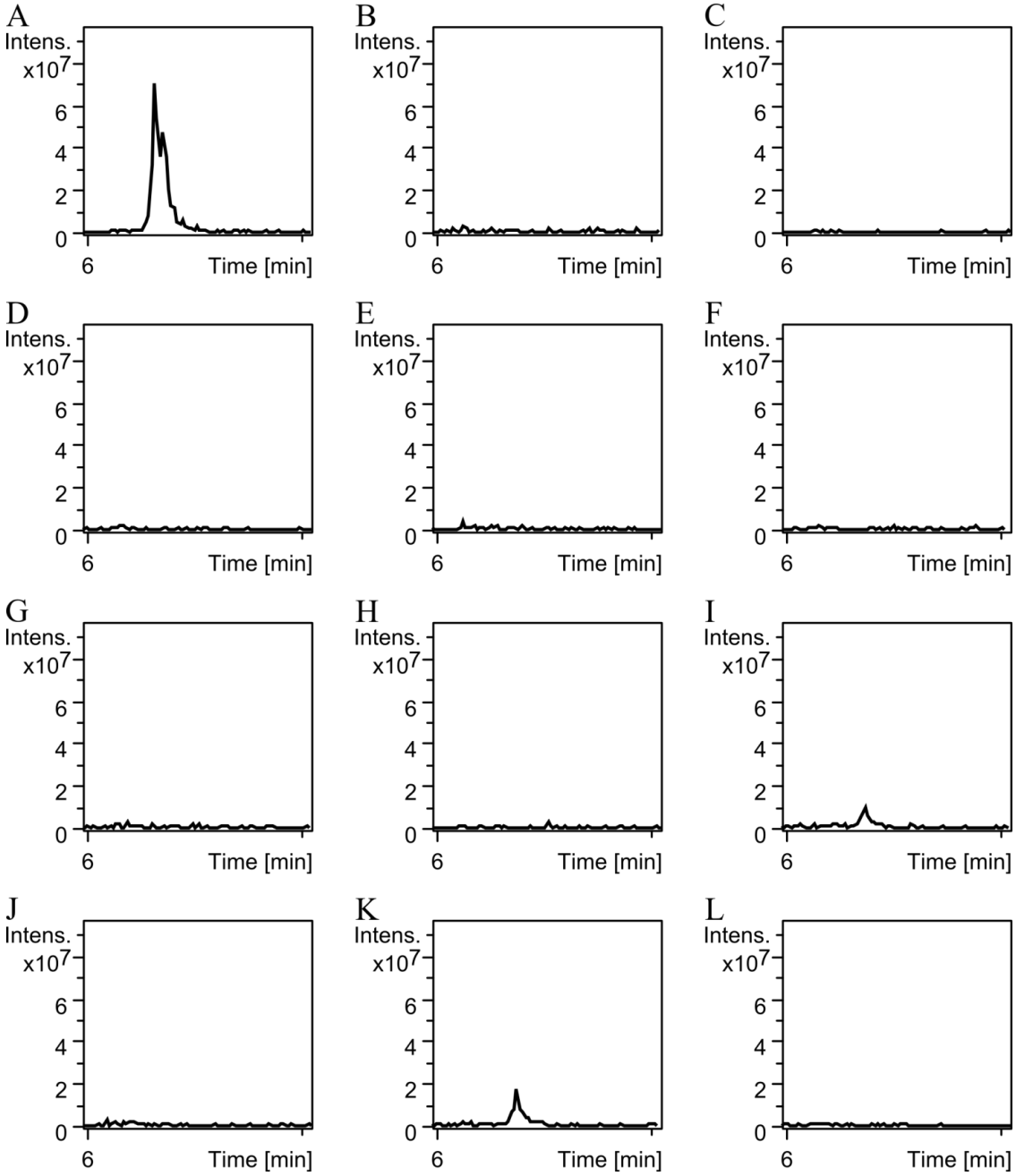

Figure 4.23. LC-MS analysis of cyclohuinilsopeptin production by *S. albus* strains harboring derivatives of 2D19 with the deletions of individual genes. Extracted chromatograms 880.5 ± 0.5 are shown. The peak at retention time 6.3 min corresponds to cyclohuinilsopeptin. A – extract of *S. albus* 2D19. B – extract of *S. albus* 2D19_5508. C – extract of *S. albus* 2D19_5512. D – extract of *S. albus* 2D19_5514. E – extract of *S. albus* 2D19_5515. F – extract of *S. albus* 2D19_5516. G – extract of *S. albus* 2D19_5518. H – extract of *S. albus* 2D19_5519. I – extract of *S. albus* 2D19_5520. J – extract of *S. albus* 2D19_5521. K – extract of *S. albus* 2D19_5522. L – extract of *S. albus* Del14.

Traces of cyclohuinilsopeptin A could be detected in the extracts originating from the strains harboring 2D19 derivatives with the inactivation of the genes KALB_5520 and KALB_5522 which encode putative amidinotransferase and fatty acid CoA ligase (Figure 4.23). It is not possible to conclude whether the genes KALB_5520 and KALB_5522 are involved in cyclohuinilsopeptin A

biosynthesis or not. The possibility cannot be excluded that the inactivation of these genes is crosscomplemented by genes encoded in the genome of the heterologous host strain.

The extracts of the obtained strains with inactivation of individual genes were screened for the presence of peaks which might correspond to the cyclohuinilsopeptin A derivatives or precursors. In all but one strain no new peaks corresponding to possible cyclohuinilsopeptin A intermediates were detected. Solely in the extract of the *S. albus* 2D19_5514 strain with the inactivation of the KALB_5514 gene encoding a putative radical SAM protein a new peak with an m/z value of 854.5 was identified (Figure 4.24). This peak could not be identified in the extract of the negative control strain *S. albus* Del14 and was present only in trace amounts in the extract of the positive control strain S. *albus* 2D19 harboring the intact cyclohuinilsopeptin A gene cluster (Figure 4.24). This implies that the identified compound might correspond to the biosynthetic intermediate of cyclohuinilsopeptin.

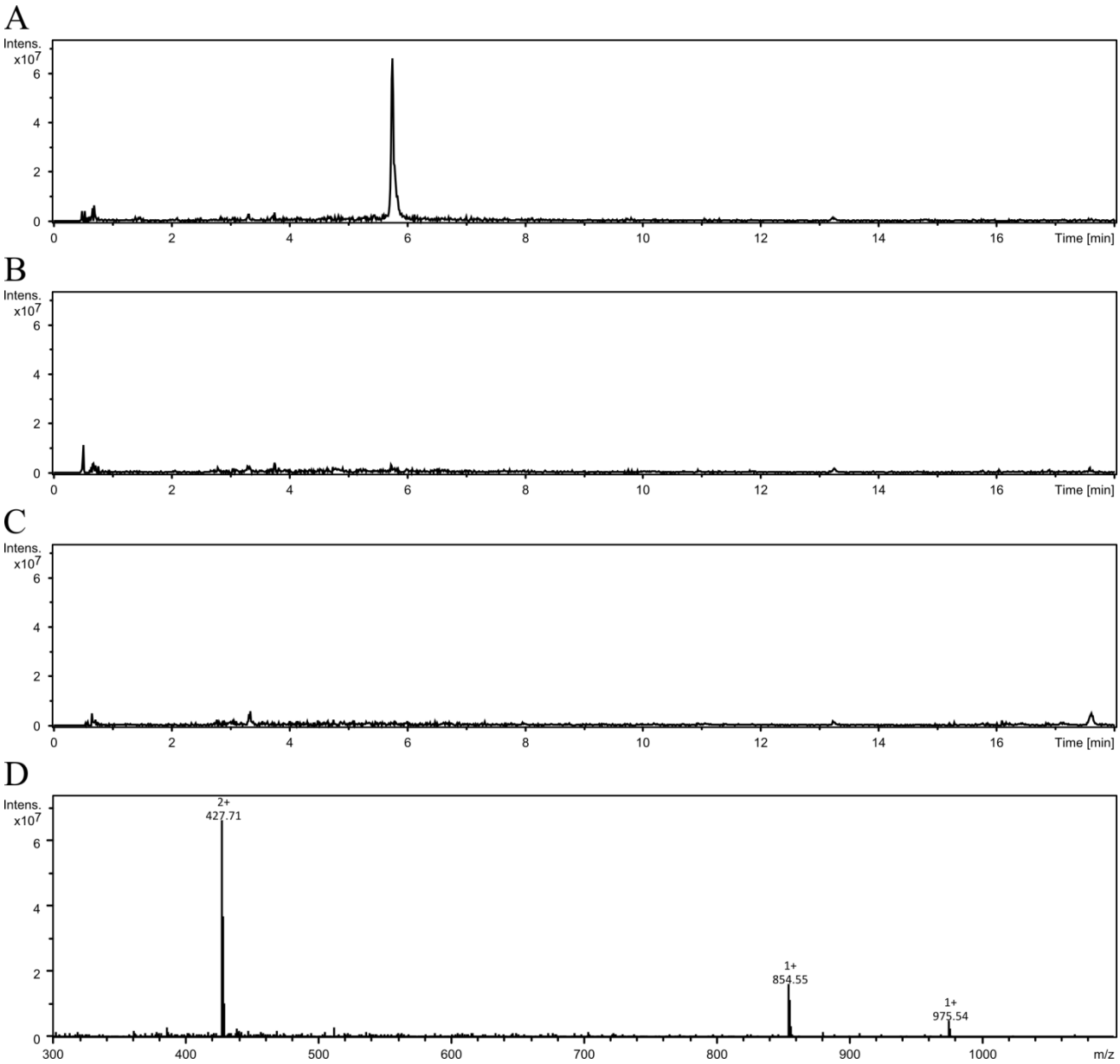

Figure 4.24. LC-MS detection of cyclohuinilsopeptin B. A, B and C – extracted ion chromatograms 854.5 ± 0.5 of the extracts from the cultures of *S. albus* 2D19_5514, *S. albus* 2D19 and *S. albus* Del14 respectively. Cyclohuinilsopeptin can be seen as a peak at 5.8 min. D – mass chromatogram of the peak corresponding to cyclohuinilsopeptin B.

The compound with the m/z value of 854.5 was isolated from the culture broth of the *S. albus* 2D19_5514 strain for structure elucidation purposes. The NMR analysis of the isolated compound was performed by Dr. Josef Zapp.

The NMR analysis confirmed the assumption that the isolated compound with high probability it is a biosynthetic precursor of cyclohuinilsopeptin A. The isolated compound was named cyclohuinilsopeptin B. The structures of cyclohuinilsopeptin B (2) and cyclohuinilsopeptin A share significant similarity. The structure of cyclohuinilsopeptin B consists of seven subunits of which only subunit IV, lysine, was different from the subunit IV of cyclohuinilsopeptin A (Figures 4.3 and 4.25). Detailed information can be found in table 4.4 and figures 4.26 – 4.39.

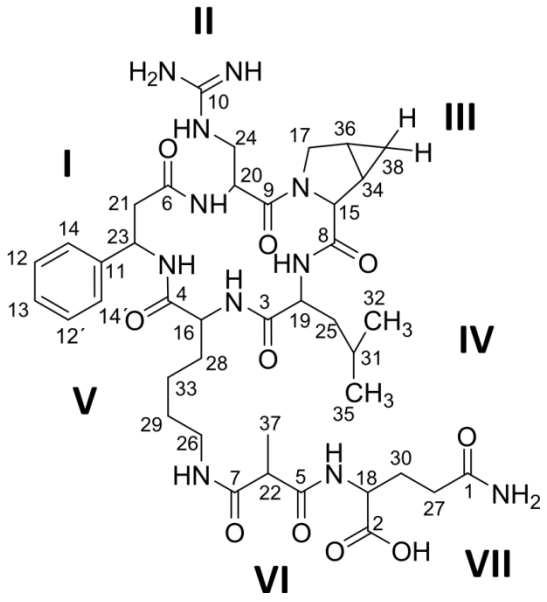

Figure 4.25. Structure of cyclohuinilsopeptin B with numbered atoms according to the NMR Data. I – β-phenylalanine; II – 3-guanidino-alanine; III – 3,4-cyclopropyl-proline; IV – leucine; V – lysine; VI – 2-methylmalonic acid and VII – glutamine.

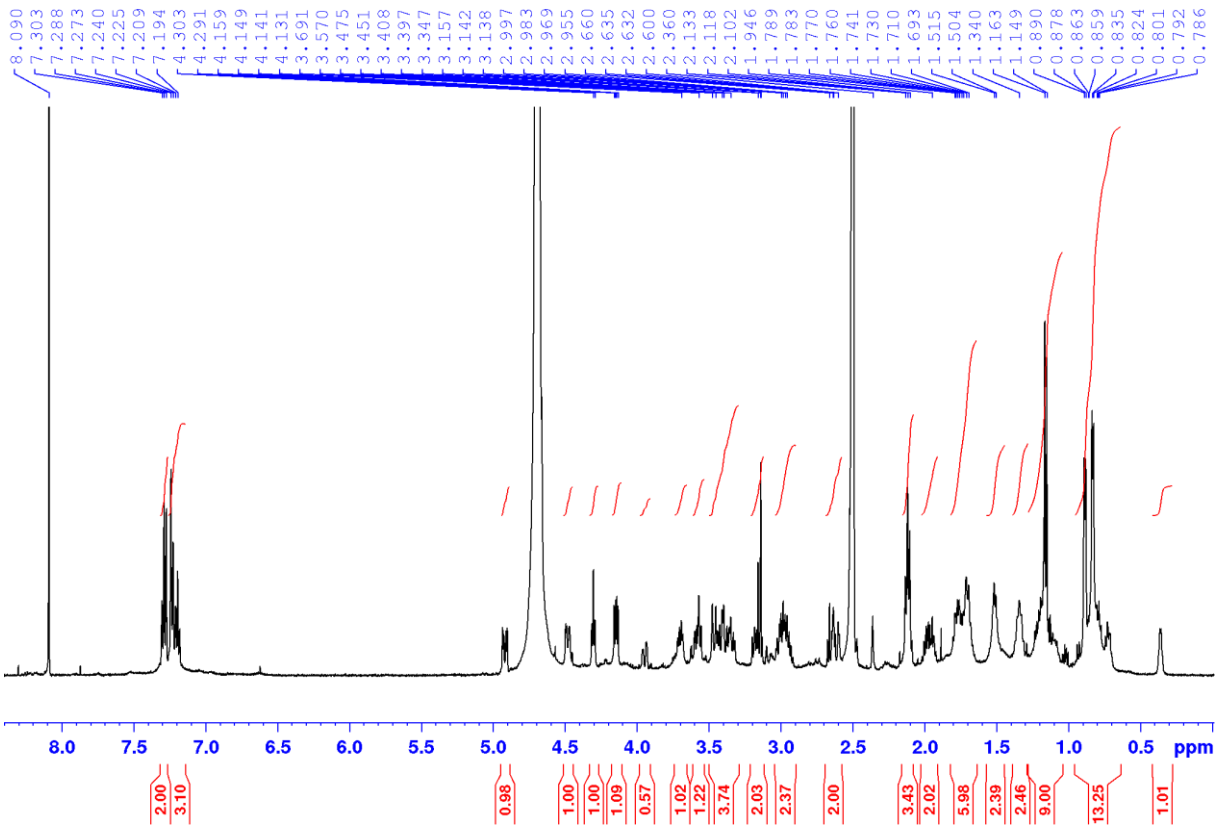

Figure 4.26. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of cyclohuinilsopeptin B **(2)**; complete spectrum.

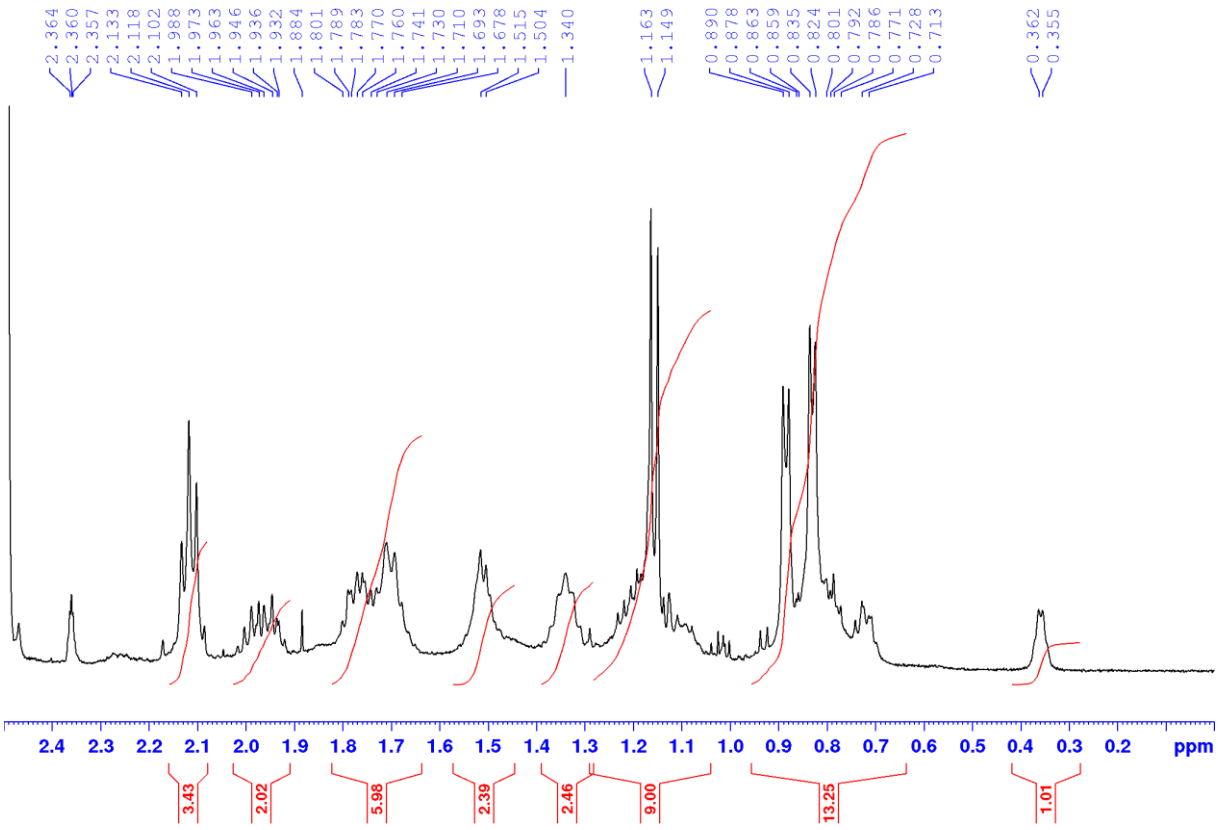

Figure 4.27. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**; zoom from 2.5 ppm to 0ppm.

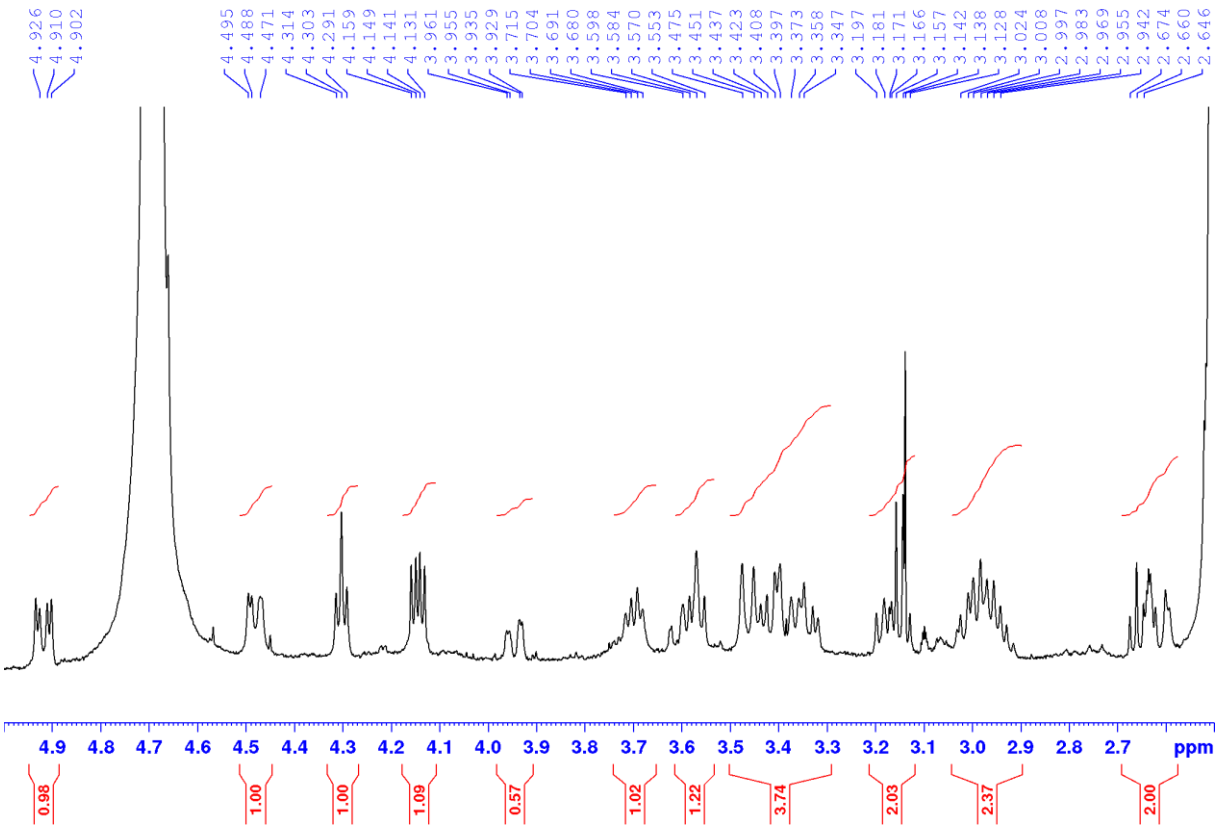

Figure 4.28. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**; zoom from 5 ppm to 2.5 ppm.

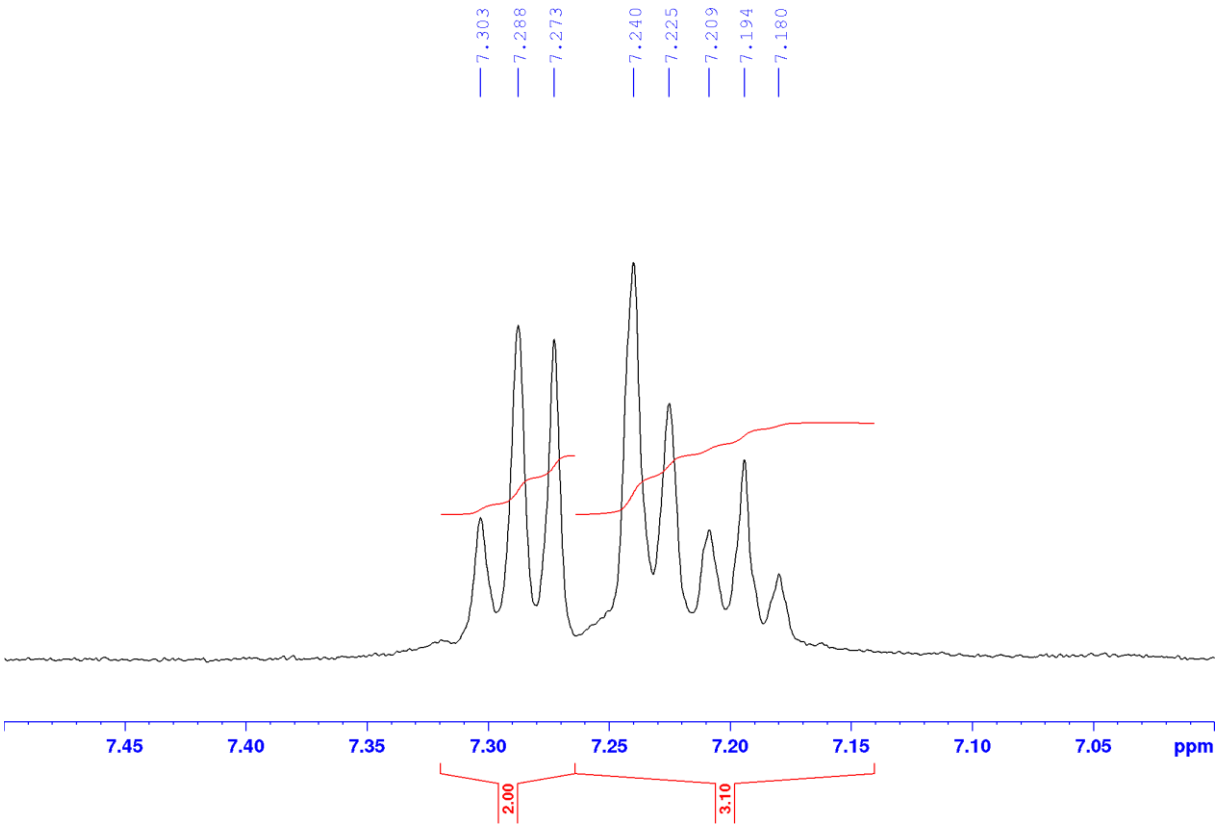

Figure 4.29. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**; zoom from 7.5 ppm to 7 ppm.

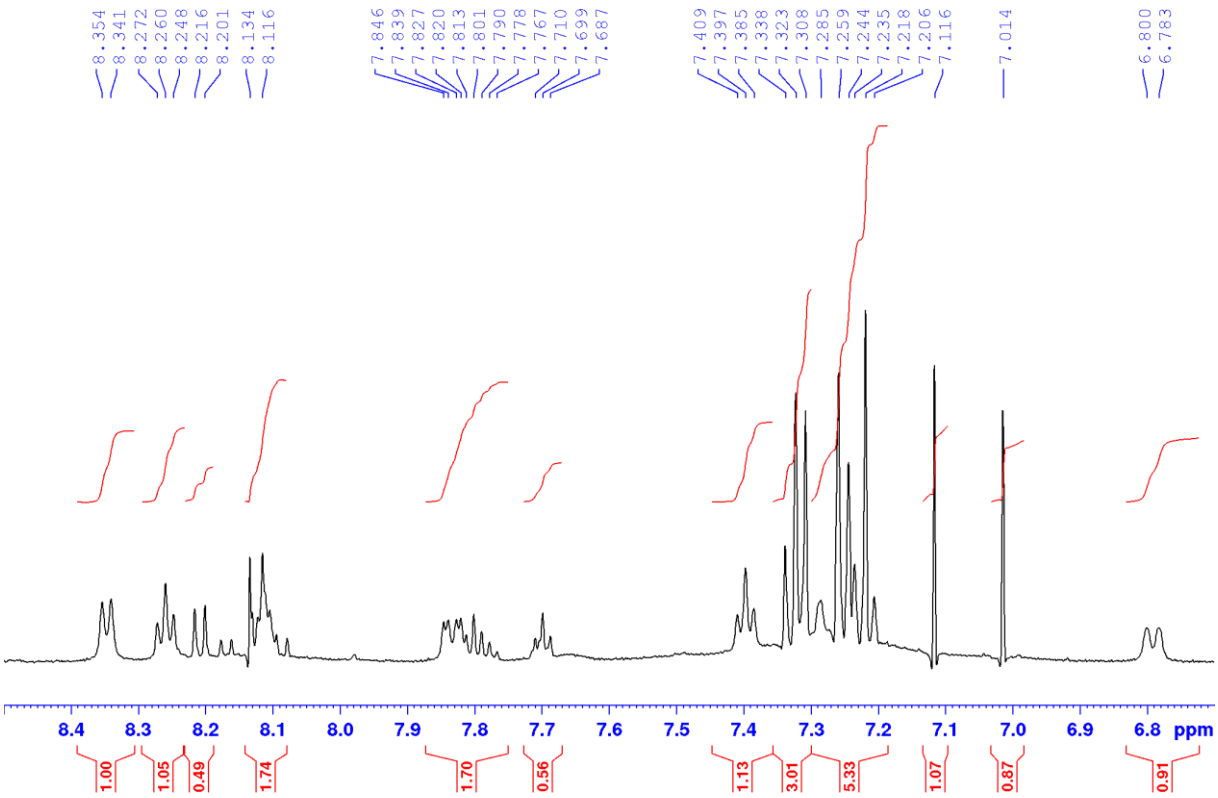

Figure 4.30. ^{1}H-NMR spectrum (700 MHz, *DMSO-d_6*, *TFA*, *H_2O*) of **2**; zoom from 8.5 ppm to 6.9 ppm.

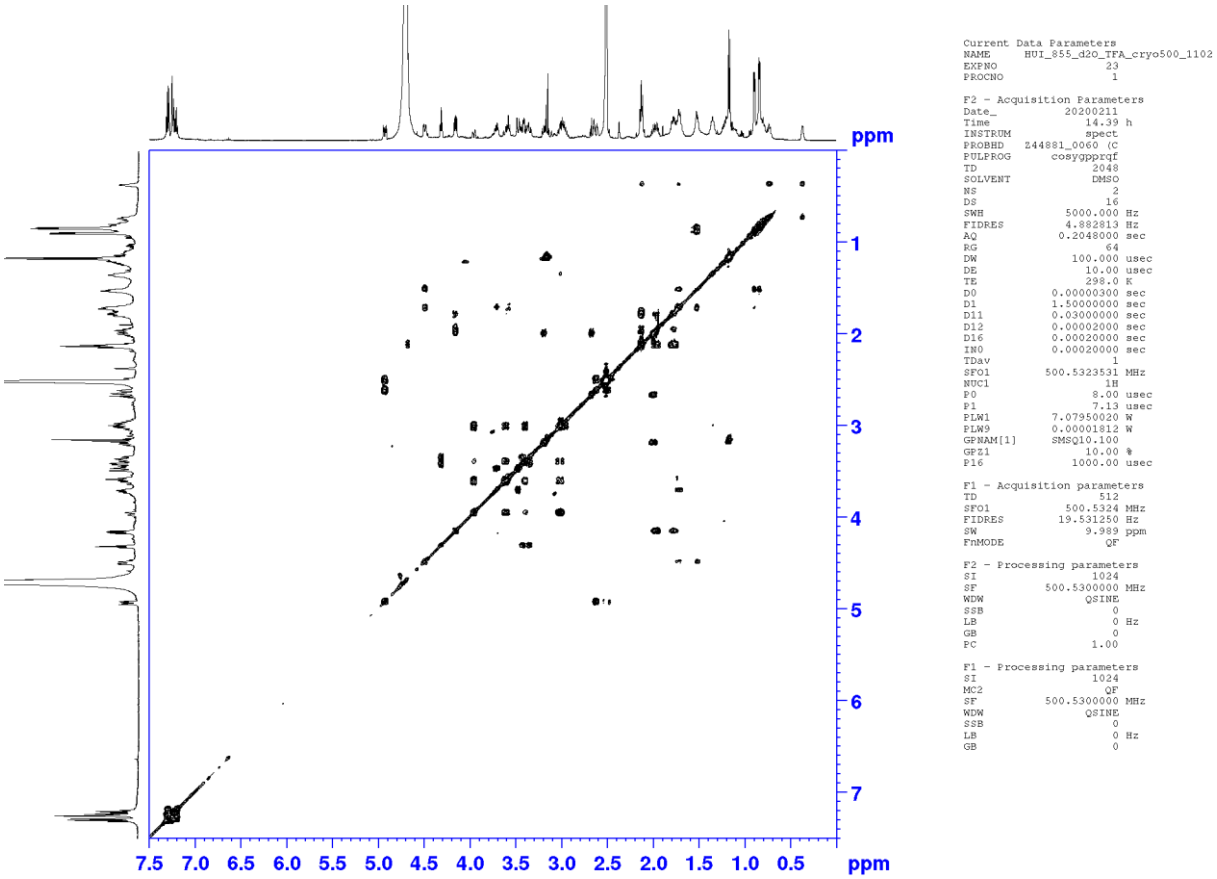

Figure 4.31. ^{1}H-^{1}H- COSY spectrum (700 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, complete spectrum.

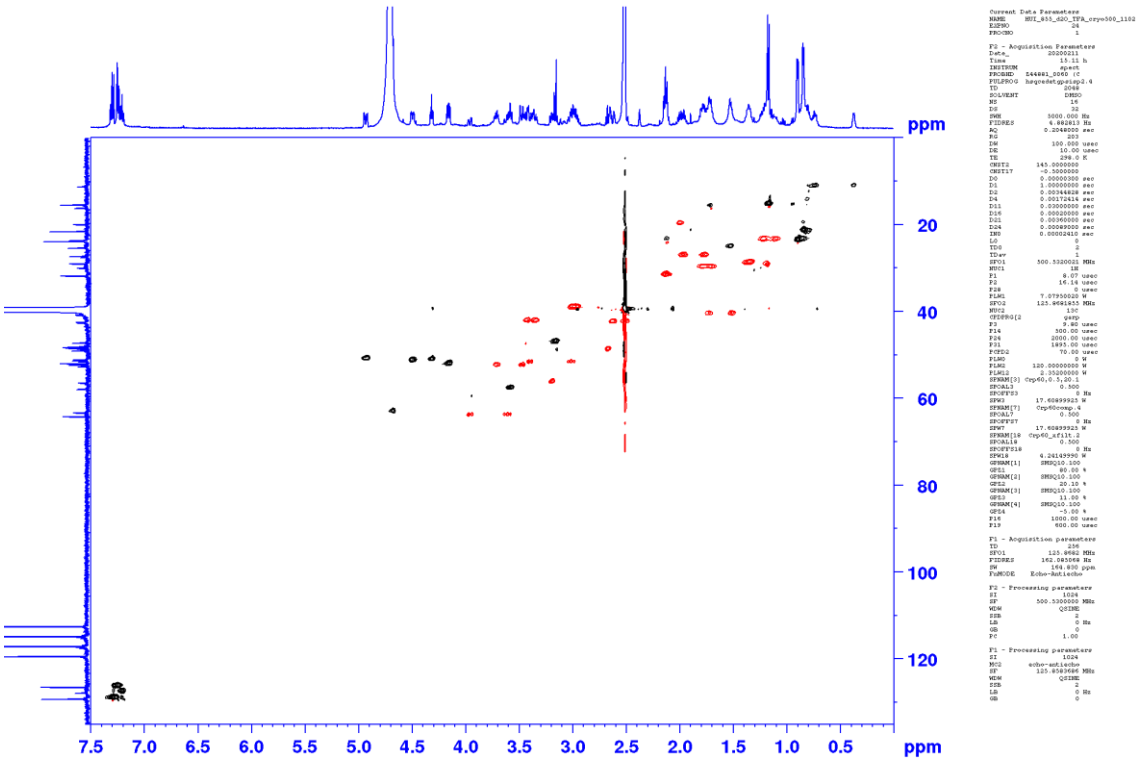

Figure 4.32. HSQC-spectrum (700 MHz, 175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, complete spectrum.

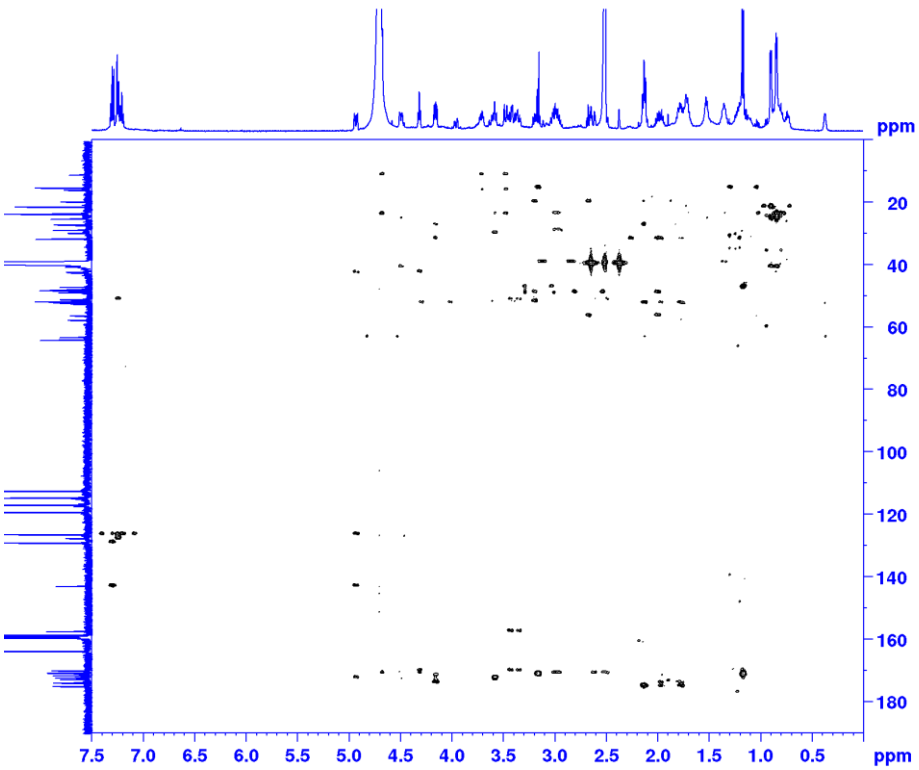

Figure 4.33. HMBC-spectrum (700 MHz, 175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, complete spectrum.

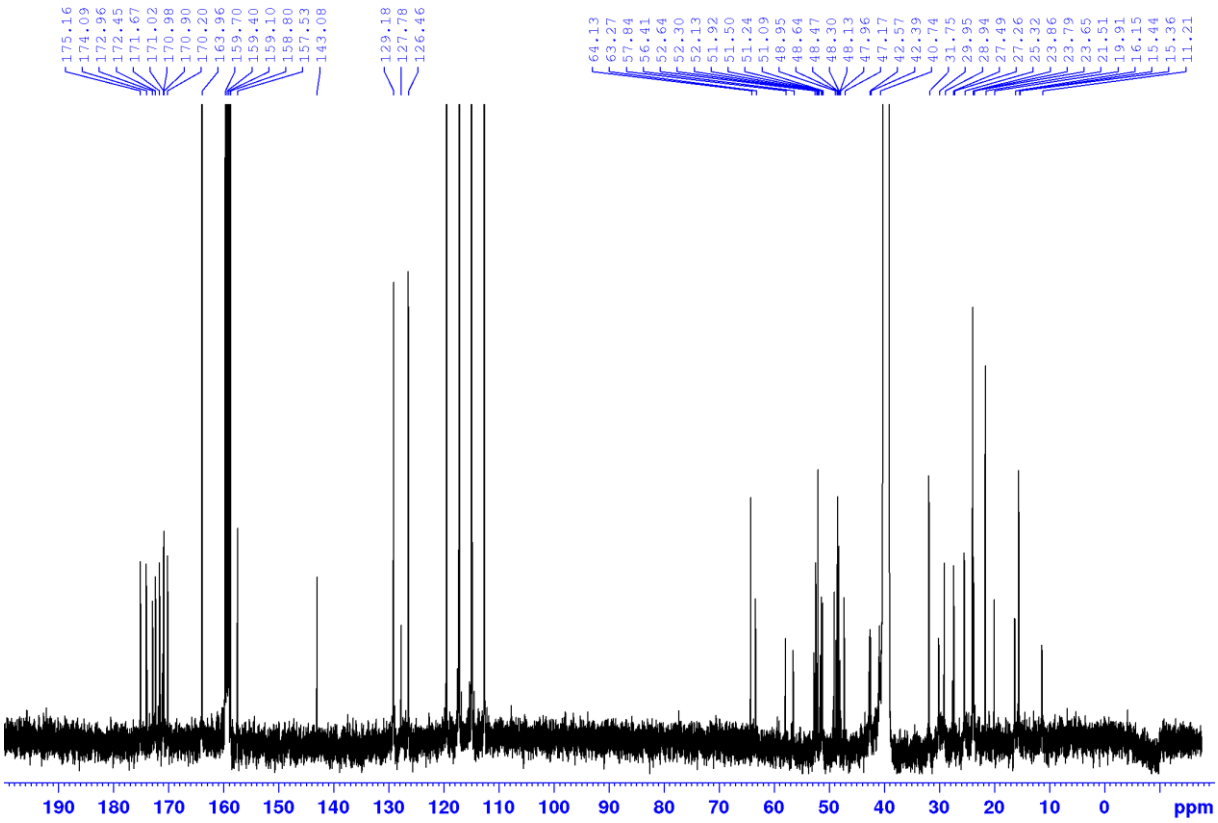

Figure 4.34. ^{13}C-NMR spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, complete spectrum.

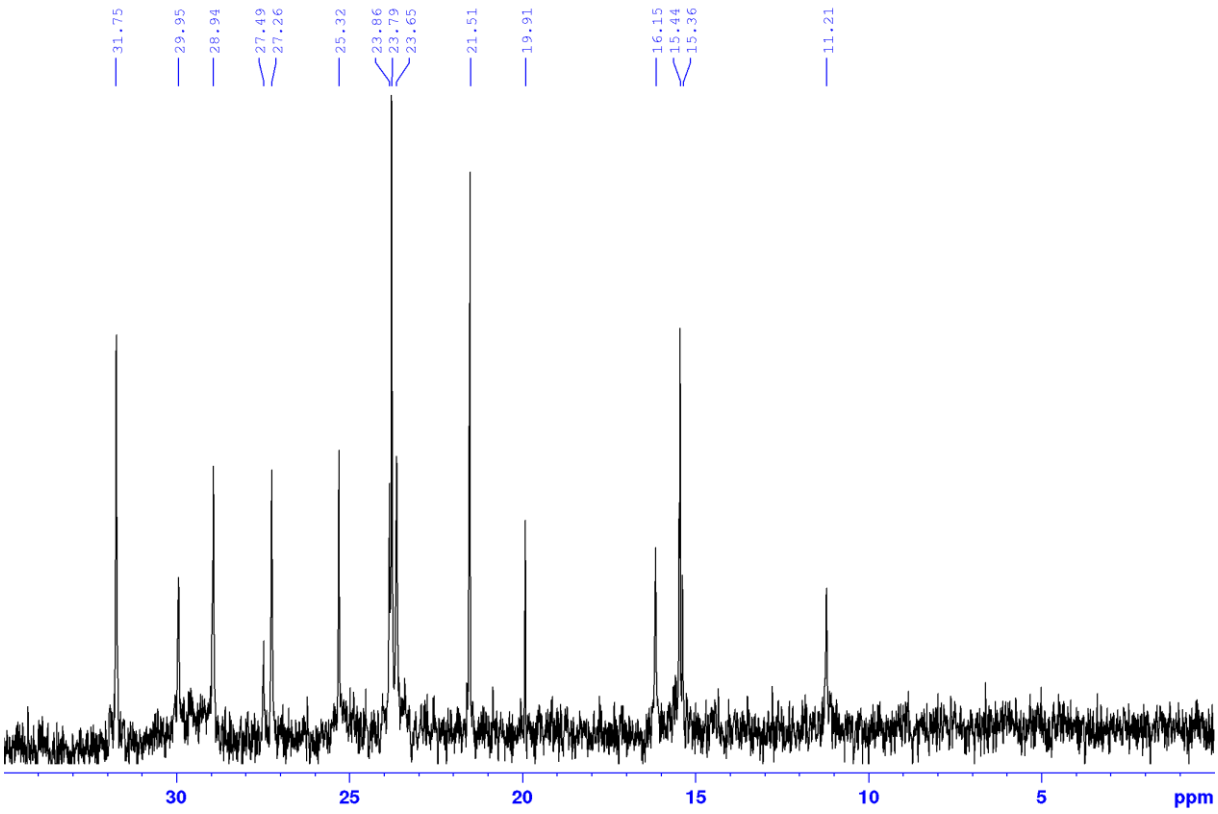

Figure 4.35. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, zoom from 35 ppm to 0 ppm.

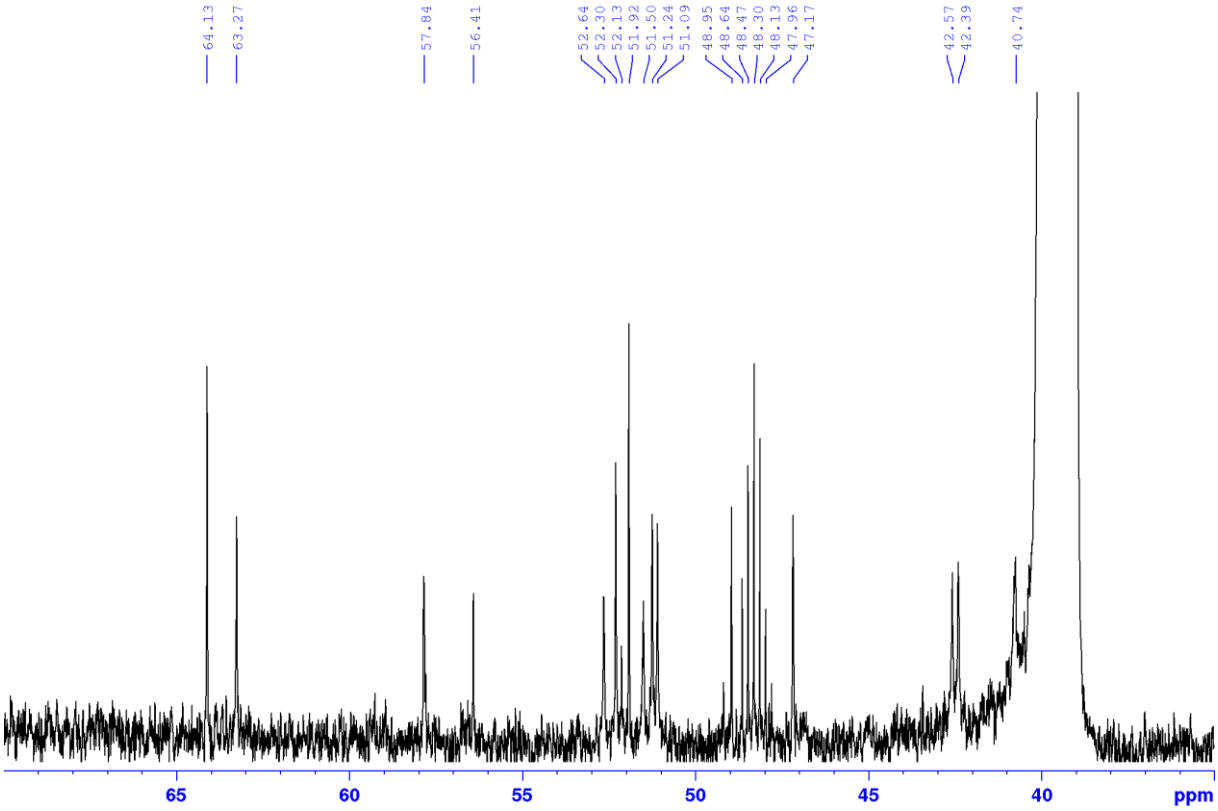

Figure 4.36. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, zoom from 70 ppm to 35 ppm.

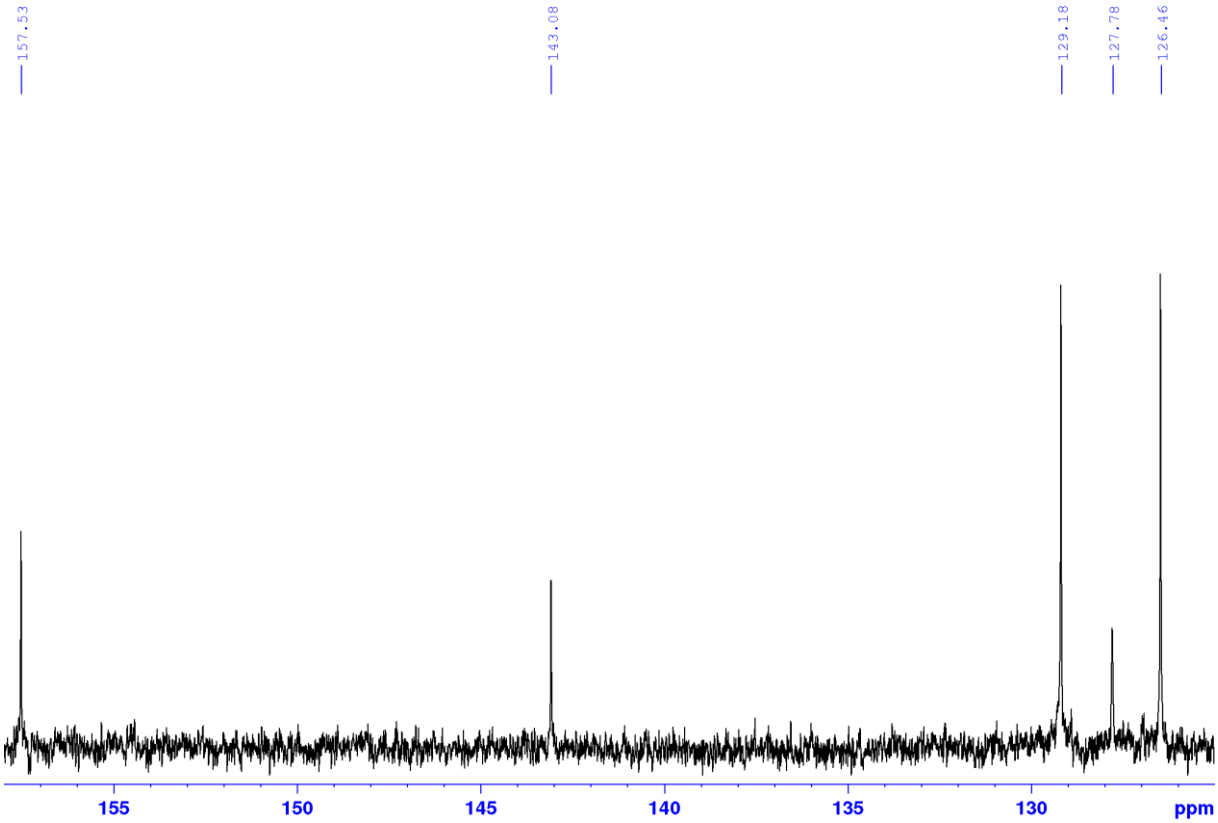

Figure 4.37. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, zoom from 158 ppm to 125 ppm.

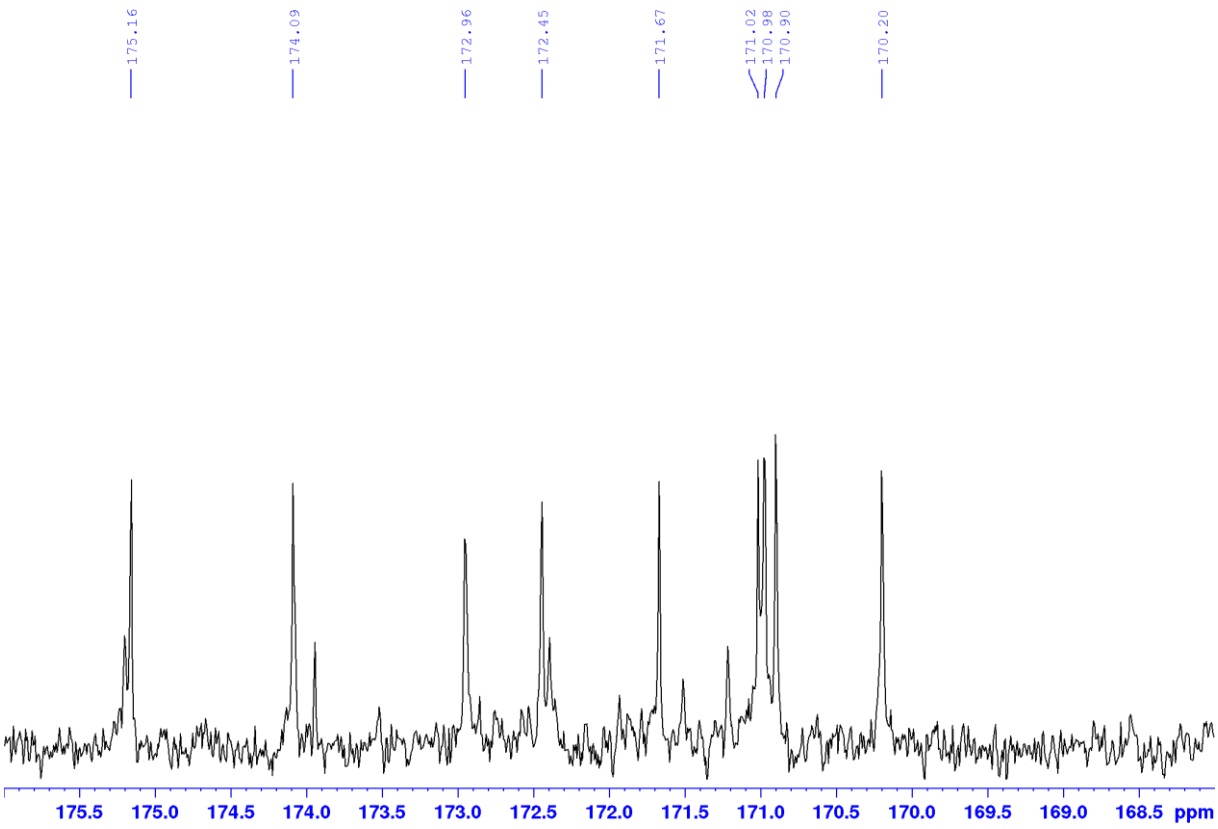

Figure 4.38. ^{13}C-spectrum (175 MHz, *DMSO-d_6*, *d-TFA*, *D_2O*) of **2**, zoom from 176 ppm to 168 ppm.

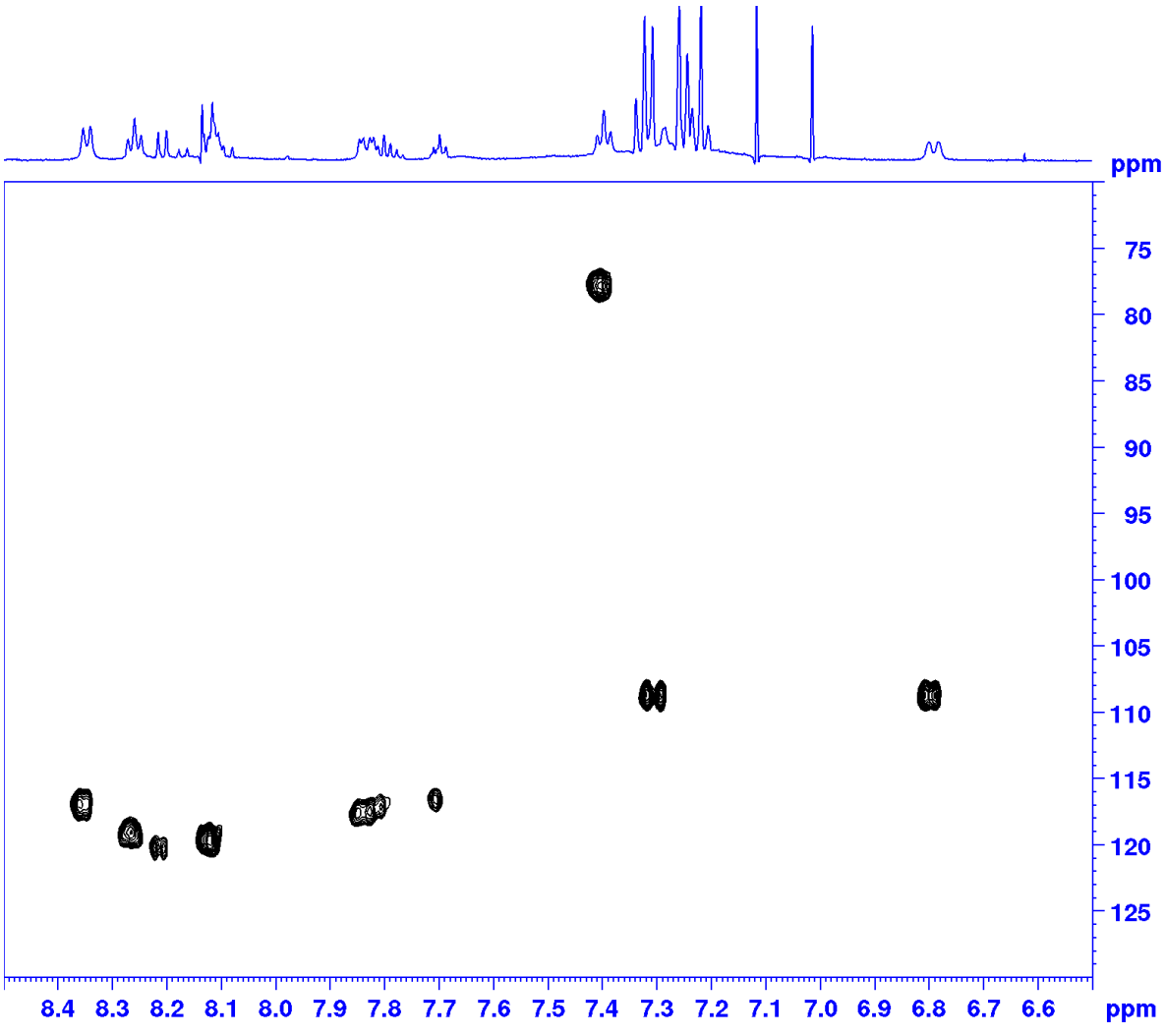

Figure 4.39. ^{15}N-HSQC (50 MHz, *DMSO-d_6*, *TFA*, *H_2O*) of **2**, complete spectrum.

3.5 Insights into biosynthesis of cyclohuinilsopeptin A

The current state of the obtained experimental and sequence analysis data does not allow to predict the entire biosynthetic route leading to the production of the cyclohuinilsopeptin A. The NRPS encoded by the gene KALB_5512 is presumable responsible for the formation of the cyclic pentapeptide moiety of cyclohuinilsopeptin A. Analysis of this NRPS with the Antismash [23] and the Non-Ribosomal Peptide Synthase Substrate Predictor (NRPSsp) [24] revealed the presence of five modules within its sequence (Figure 4.40). Amino acid substrate specificity of the adenylation domains within the identified modules could either not be predicted by the used software or the prediction results were unreliable. The substrate specificity of the adenylation domain of the first module of was predicted by the NRPSsp software to be phenylalanine which is present in the structure of the pentapeptide precursor of cyclohuinilsopeptin A. L-phenylalanine is regarded as a first aminoacid residue of the pentapeptide NRPS product.

The second module of the NRPS lacks the adenylation domain (Figure 4.40). We propose that the aminoalanine residue is transferred to the thiolation domain of the second module by trans adenylation mechanism. Trans adenylation during the biosynthesis of NRPS products was reported for viomycin, andrimid, syringomycin, etc [25–27]. Two genes, KALB_5510 and KALB_5513, both encoding a small NRPS consisting of an adenylation and a thiolation domain are present within the cyclohuinildopeptin gene cluster. We propose that one of these products loads the aminoalanine residue onto the thiolation domain of the second module.

The specificity of the adenylation domain of the third NPRS module was predicted by the NRPSsp software as N-hydroxy-ornithine. We propose that the third NRPS module is responsible for the attachment of the ornithine residue to aminoalanine.

The fourth and the fifth modules of the NRPS encoded by KALB_5512 presumably attach leucine and lysine residues respectively. The specificity of the adenylation domain of the third module was predicted by Antismash as proline and by NRPSsp – as pipecolic acid. Nevertheless, we suggest that the leucine is attached by the third NRPS module. The structure of cyclohuinilsopeptin B obtained after the inactivation of KALB_5514 demonstrates the incorporation of leucine residue in the respective position of the cyclic pentapeptide precursor. No clear substrate specificity prediction could be obtained for the adenylation domain of the last NRPS module. According to the structures of cyclohuinilsopeptins A and B, attachment of the lysine residue should be catalyzed by the fifth module.

The fifth NRPS module contains a thioesterase domain. We propose that this domain catalyzes the hydrolysis of the pentapeptide chain from the NRPS which after cyclization gives rise to the cyclic pentapeptide precursor of cyclohuinilsopeptin A.

The cyclic pentapeptide precursor of cyclohuinilsopeptin A has to undergo a series of post NRPS modifications to give rise to the final molecule. We propose that the ornithine residue of the

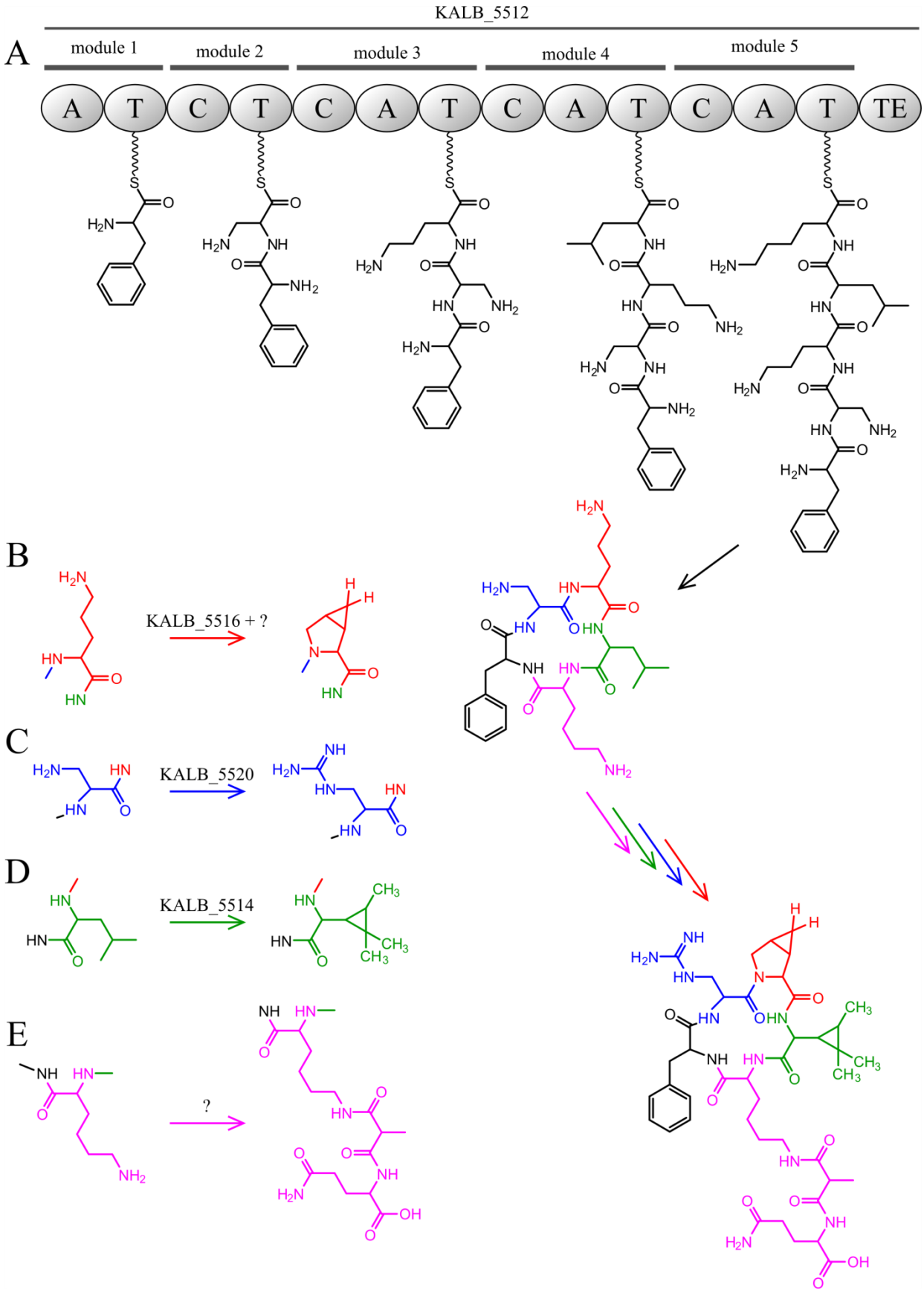

Figure 4.40. Proposed biosynthetic scheme of cyclohuinilsopeptin A biosynthesis. A – biosynthesis of cyclic pentapeptide precursor of cyclohuinilsopeptin by the NRPS encoded by KALB_5512. B – conversion of ornithine residue of the cyclohuinilsopeptin A precursor to 3,4-cyclopropyl-proline by a putative ornithine cyclodeaminase encoded by KALB_5516 and an unidentified enzyme or enzymes. C – conversion of the aminoalanine residue to 3-guanidino-alanine by the amidinotransferase encoded by KALB_5520. D – conversion of the leucine residue to 2-(2,2,3-trimethylcyclopropyl)-glycine by the action of radical SAM protein encoded by KALB_5514. E – attachment of methylmalonyl and glutamyl residues to the lysine residue of the cyclohuinilsopeptin A precursor by unidentified enzymes encoded by the genes within the biosynthetic gene cluster.

precursor is converted to proline through the action of a putative ornithine cyclodeaminase encoded by KALB_5516 (Figure 4.40 B). This enzyme catalyzes the cleavage of the terminal amino group of ornithine and its cyclization to proline [28]. How the proline residue is further converted to 3,4-cyclopropyl-proline remains elusive.

The conversion of the aminoalanine is likely to be catalyzed by the product of KALB_5520 which encodes a putative amidinotransferase (Figure 4.40 C). The enzyme is proposed to transfer amidino group from L-arginine to the amino group in the third position in the similar manner as glycine amidinotransferase acts [29].

The conversion of the leucine residue to 2-(2,2,3-trimethylcyclopropyl)-glycine is likely to be catalyzed by the putative radical SAM protein encoded by KALB_5514 (Figure 4.40 D). Inactivation of the KALB_5514 gene led to the production of cyclohuinilsopeptin B. The produced precursor contains leucine residue instead of 2-(2,2,3-trimethylcyclopropyl)-glycine found in the structure of the final product – cyclohuinilsopeptin A.

Methylmalonyl and glutamyl residues are attached to the lysine in the structure of cyclohuinilsopeptin A. The mechanisms of these reactions as well as the responsible enzymes remain unknown. Additional experiments like feeding of labeled precursors, inactivation gene of additional genes etc, are required to prove and to complete the proposed biosynthetic scheme.

4 References

1. Katz, L.; Baltz, R.H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176, doi:10.1007/s10295-015-1723-5.
2. Vézina, C.; Kudelski, A.; Sehgal, S.N. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J. Antibiot.* **1975**, *28*, 721–726, doi:10.7164/antibiotics.28.721.
3. Bardone, M.R.; Paternoster, M.; Coronelli, C. Teichomycins, new antibiotics from Actinoplanes teichomyceticus nov. sp. II. Extraction and chemical characterization. *J. Antibiot.* **1978**, *31*, 170–177, doi:10.7164/antibiotics.31.170.
4. Mccormick, M.H.; Mcguire, J.M.; Pittenger, G.E.; Pittenger, R.C.; Stark, W.M. Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot Annu* **1955**, *3*, 606–611.
5. Eliopoulos, G.M.; Willey, S.; Reiszner, E.; Spitzer, P.G.; Caputo, G.; Moellering, R.C. In vitro and in vivo activity of LY 146032, a new cyclic lipopeptide antibiotic. *Antimicrob. Agents Chemother.* **1986**, *30*, 532–535, doi:10.1128/aac.30.4.532.
6. Kirst, H.A.; Michel, K.H.; Martin, J.W.; Creemer, L.C.; Chio, E.H.; Yao, R.C.; Nakatsukasa, W.M.; Boeck, L.D.; Occolowitz, J.L.; Paschal, J.W.; et al. A83543A-D, unique fermentation-derived tetracyclic macrolides. *Tetrahedron Letters* **1991**, *32*, 4839–4842, doi:10.1016/S0040-4039(00)93474-9.
7. Burg, R.W.; Miller, B.M.; Baker, E.E.; Birnbaum, J.; Currie, S.A.; Hartman, R.; Kong, Y.L.; Monaghan, R.L.; Olson, G.; Putter, I.; et al. Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. *Antimicrob. Agents Chemother.* **1979**, *15*, 361–367, doi:10.1128/aac.15.3.361.
8. Comprehensive Natural Products Chemistry | ScienceDirect Available online: https://www.sciencedirect.com/referencework/9780080912837/comprehensive-natural-products-chemistry (accessed on Jun 26, 2020).
9. Baltz, R.H. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 281–299, doi:10.1007/s10295-018-2115-4.
10. Myronovskyi, M.; Luzhetskyy, A. Heterologous production of small molecules in the optimized Streptomyces hosts. *Nat Prod Rep* **2019**, *36*, 1281–1294, doi:10.1039/c9np00023b.
11. Baltz, R.H. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* **2010**, *37*, 759–772, doi:10.1007/s10295-010-0730-9.
12. Rebets, Y.; Tokovenko, B.; Lushchyk, I.; Rückert, C.; Zaburannyi, N.; Bechthold, A.; Kalinowski, J.; Luzhetskyy, A. Complete genome sequence of producer of the glycopeptide

antibiotic Aculeximycin Kutzneria albida DSM 43870T, a representative of minor genus of Pseudonocardiaceae. *BMC Genomics* **2014**, *15*, 885, doi:10.1186/1471-2164-15-885.

13. Green, M.R.; Sambrook, J. *Molecular Cloning: A Laboratory Manual (Fourth Edition)*; 4th Ed.; Cold Spring Harbor Lab. Press, Plainview, NY, 2012;
14. Kieser, T.; Bibb, M.J.; Buttner, M.J.; Chater, K.F.; Hopwood, D.A. *Practical Streptomyces Genetics*; John Innes Foundation, Norwich, England, 2000;
15. Bilyk, O.; Sekurova, O.N.; Zotchev, S.B.; Luzhetskyy, A. Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. *PLoS ONE* **2016**, *11*, e0158682, doi:10.1371/journal.pone.0158682.
16. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metab. Eng.* **2018**, *49*, 316–324, doi:10.1016/j.ymben.2018.09.004.
17. Flett, F.; Mersinias, V.; Smith, C.P. High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.* **1997**, *155*, 223–229, doi:10.1111/j.1574-6968.1997.tb13882.x.
18. Grant, S.G.; Jessee, J.; Bloom, F.R.; Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4645–4649, doi:10.1073/pnas.87.12.4645.
19. Myronovskyi, M.; Rosenkränzer, B.; Luzhetskyy, A. Iterative marker excision system. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 4557–4570, doi:10.1007/s00253-014-5523-z.
20. Muyrers, J.P.P.; Zhang, Y.; Benes, V.; Testa, G.; Rientjes, J.M.J.; Stewart, A.F. ET recombination: DNA engineering using homologous recombination in E. coli. *Methods Mol. Biol.* **2004**, *256*, 107–121, doi:10.1385/1-59259-753-X:107.
21. Mazodier, P.; Petter, R.; Thompson, C. Intergeneric conjugation between Escherichia coli and Streptomyces species. *J. Bacteriol.* **1989**, *171*, 3583–3585, doi:10.1128/jb.171.6.3583-3585.1989.
22. Rebets, Y.; Tokovenko, B.; Lushchyk, I.; Rückert, C.; Zaburannyi, N.; Bechthold, A.; Kalinowski, J.; Luzhetskyy, A. Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin Kutzneria albida DSM 43870T, a representative of minor genus of Pseudonocardiaceae. *BMC Genomics* **2014**, *15*, 885, doi:10.1186/1471-2164-15-885.
23. 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.

24. Prieto, C.; García-Estrada, C.; Lorenzana, D.; Martín, J.F. NRPSsp: non-ribosomal peptide synthase substrate predictor. *Bioinformatics* **2012**, *28*, 426–427, doi:10.1093/bioinformatics/btr659.
25. Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D.C.; Grandi, G. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **1998**, *273*, 32857–32863, doi:10.1074/jbc.273.49.32857.
26. Magarvey, N.A.; Fortin, P.D.; Thomas, P.M.; Kelleher, N.L.; Walsh, C.T. Gatekeeping versus promiscuity in the early stages of the andrimid biosynthetic assembly line. *ACS Chem. Biol.* **2008**, *3*, 542–554, doi:10.1021/cb800085g.
27. Thomas, M.G.; Chan, Y.A.; Ozanick, S.G. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob. Agents Chemother.* **2003**, *47*, 2823–2830, doi:10.1128/aac.47.9.2823-2830.2003.
28. Costilow, R.N.; Laycock, L. Ornithine cyclase (deaminating). Purification of a protein that converts ornithine to proline and definition of the optimal assay conditions. *J. Biol. Chem.* **1971**, *246*, 6655–6660.
29. Humm, A.; Fritsche, E.; Mann, K.; Göhl, M.; Huber, R. Recombinant expression and isolation of human L-arginine:glycine amidinotransferase and identification of its active-site cysteine residue. *Biochem. J.* **1997**, *322 (Pt 3)*, 771–776, doi:10.1042/bj3220771.