



Technische Universität München
Fakultät für Chemie
Professur für Biosystemchemie



**Investigations into (cyano-)bacterial secondary metabolite
biosynthesis in heterologous expression systems**

Elke Regina Duell

Dissertation

2019



Technische Universität München
Fakultät für Chemie
Professur für Biosystemchemie



Investigations into (cyano-)bacterial secondary metabolite biosynthesis in heterologous expression systems

Elke Regina Duell

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Hubert A. Gasteiger

Prüfende der Dissertation: 1. Prof. Dr. Tobias A. M. Gulder
2. Prof. Dr. Michael Groll

Die Dissertation wurde am 19.02.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 20.03.2019 angenommen.

Nur eins darf mer nit verjesse:
Et jeht wigger hingerm Horizont,
Hingerm Horizont.
BAP

Für mich
und alle die mich barfuß treffen.

Parts of this thesis have been published in:

- C. Greunke*, **E. R. Duell***, P. M. D’Agostino*, A. Glöckle, K. Lamm, T. A. M. Gulder: Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Met. Eng.* **2018**, *47*, 334–345, DOI 10.1016/j.ymben.2018.03.010
*equally contributing authors
- **E. R. Duell***, P. M. D’Agostino*, N. Shapiro, T. Woyke, T. M. Fuchs, T. A. M. Gulder: Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. *Microb. Cell Fact.* **2019**, *18*, DOI 10.1186/s12934-019-1080-6
*equally contributing authors

This thesis contains content which will be part of the publication:

- **E. R. Duell***, T. M. Milzarek*, T. F. Schäberle, G. M. König, T. A. M. Gulder: Investigations into ambigol biosynthesis. *submitted*
*equally contributing authors

Abstract

Within this thesis, a new technique for natural product biosynthetic gene cluster (BGC) cloning and expression was co-developed and extended by an easy, mRNA-independent method to check for full gene transcription immediately during heterologous expression. Direct Pathway Cloning (DiPaC) allows capturing of small- to mid-size gene clusters straight into vectors utilizable for heterologous expression. Already during this process, the respective DNA sequence can be modified, e.g., by adding, exchanging or deleting genomic regulators such as promoters and terminators or by rearranging the order of individual genes. The applicability of DiPaC to clone challenging secondary metabolite BGCs was demonstrated on three examples.

The cyanobacterial, hexapeptidic anabaenopeptins are protease inhibitors, which are biosynthesized in different variants by several distinct cyanobacteria. *Nostoc punctiforme* PCC 73102 produces the two L-homophenylalanine (L-Hph) and L-homotyrosine (L-Hty) containing variants anabaenopeptin NZ857 and nostamide A encoded on the 29.2 kb *apt* gene cluster. The genes responsible for the homoamino acid biosynthesis are located up- and downstream of the core nonribosomal peptide synthetase (NRPS) unit, which consists of four genes necessary for sequential assembly and cyclization of the final products. Utilizing DiPaC, the *apt* BGC was integrated into two different expression plasmids, which allowed independent expression induction of the precursor synthesizing enzymes and subsequently the NRPS assembly machinery. This strategy led to the first successful production of anabaenopeptin NZ857 and nostamide A in the heterologous expression host *Escherichia coli* BAP1.

Sodorifen is a volatile organic compound (VOC) with so far unknown biological activity, which is produced by various *Serratia plymuthica* strains. *S. plymuthica* WS3236 features a potential 4.6 kb sodorifen producing cluster (*sod*), which was cloned with DiPaC. Its predicted identity was subsequently confirmed by heterologous expression in *E. coli* BL21. At the same time, a transcription reporter system based on the C-terminally attached green fluorescent protein (GFP) was successfully employed. By altering the cultivation conditions, the heterologous sodorifen production rate could be increased up to 26-fold compared to the wild type strain, while the purity of the compound in raw extracts was shown to be above 90%.

Abstract

Ambigols are polychlorinated natural products synthesized by *Fischerella ambigua* 108b that exhibit interesting antibacterial, antifungal and cytotoxic activities. So far, three different derivatives, ambigol A-C, have been described. They are comprised of three dichlorophenol (DCP) units, which are linked via biaryl- and biaryl-ether-bonds and exhibit unusual bond connectivities. The two chlorine atoms on the central phenol unit of ambigol A and B are located in relative *meta* position. In ambigol C the biaryl-ether-bonds connecting the three phenols are found in *meta* position relative to the hydroxy group of the central phenol. By now, there is no enzymatic mechanism known which could explain either the biaryl bonds or the chlorine atoms in these positions. Therefore, the ambigol biosynthesis was investigated on different levels within this work.

The previous bioinformatic identification of the potential 14.2 kb ambigol (*ab*) BGC in *F. ambigua* was updated based on sequence alignments with the latest available genomic data sets. The identity of the *ab* cluster was confirmed by heterologous expression in the cyanobacterial host *Synechococcus elongatus* PCC 7942. This was only possible by extending the method for natural cyanobacterial transformation, as integration of large DNA fragments into the *Synechococcus* genome has not been described so far. By splitting the *ab* cluster in half and subsequently integrating both parts into different chromosomal neutral sites using DiPaC, double mutants expressing at least one ambigol molecule could be generated.

The *ab* cluster harbors two cytochrome P450 enzymes, Ab2 and Ab3, which are believed to introduce all biaryl- and biaryl-ether-bonds present in the ambigols. Their enzymatic reactivities were characterized using multiple approaches. After heterologous expression in *E. coli* BL21 and subsequent purification, *in vitro* assays were conducted with the potential substrate 2,4-dichlorophenol (2,4-DCP). The structures of the reaction products were elucidated using chemically synthesized standards. Ab2 was found to catalyze the formation of mainly one biaryl-ether coupled dimer, whereas Ab3 mainly formed a biaryl-coupled dimer. Both observed products feature the newly formed bond only in the usually permitted *ortho* position relative to the hydroxy group. These results were verified by *in vivo* 2,4-DCP feeding assays using both heterologous *E. coli* and *S. elongatus* strains expressing Ab2 and/or Ab3.

Zusammenfassung

Im Rahmen dieser Arbeit wurde eine neue Technik zur Klonierung und Expression von Naturstoff-Biosynthese-Genclustern (BGCs) mitentwickelt und um eine einfache, mRNA-unabhängige Methode zur Kontrolle auf vollständige Gentranskription direkt während der heterologen Expression erweitert. Direct Pathway Cloning (DiPaC) ermöglicht die Integration kleiner bis mittelgroßer Gencluster direkt in für die heterologe Expression geeignete Vektoren. Die entsprechende DNA-Sequenz kann schon während dieses Prozesses modifiziert werden, beispielsweise durch Addition, Austausch oder Deletion genomischer Regulatoren wie Promotoren und Terminatoren, oder durch das Vertauschen der Reihenfolge einzelner Gene. Die Eignung von DiPaC zur Klonierung anspruchsvoller Sekundärmetabolit-BGCs wurde anhand dreier Beispiele demonstriert.

Die cyanobakteriellen, hexapeptidischen Anabaenopeptide sind Proteaseinhibitoren, welche in verschiedenen Varianten von mehreren Cyanobakterien synthetisiert werden. *Nostoc punctiforme* PCC 73102 stellt die beiden L-Homophenylalanin (L-Hph) und L-Homotyrosin (L-Hty) enthaltenden Varianten Anabaenopeptin NZ857 und Nostamide A her, welche auf dem 29,2 kb großen *apt* Gencluster codiert sind. Die für die Homoamino-säure-Biosynthese verantwortlichen Gene befinden sich up- und downstream der zentralen nichtribosomalen Peptidsynthetase-(NRPS)Einheit, welche aus vier Genen besteht und für den Aufbau und die Zyklisierung des finalen Produkts verantwortlich ist. Unter Verwendung von DiPaC konnte das *apt* BGC aufgeteilt und in zwei verschiedene Expressionsplasmide integriert werden. Dadurch war eine unabhängige Expressionsinduktion der Vorstufen-synthetisierenden Enzyme und anschließend der assemblierenden NRPS-Maschinerie möglich. Diese Strategie führte zur ersten erfolgreichen Produktion von Anabaenopeptin NZ857 und Nostamide A in dem heterologen Expressionsstamm *Escherichia coli* BAP1.

Sodorifen ist eine volatile organische Verbindung (VOC) mit bisher unbekannter biologischer Aktivität, welche von verschiedenen *Serratia plymuthica* Stämmen produziert wird. *S. plymuthica* WS3236 enthält ein potentiell, 4,6 kb großes Sodorifen-Cluster (*sod*), welches mit DiPaC kloniert und anschließend durch heterologe Expression in *E. coli* BL21 verifiziert wurde. Gleichzeitig wurde ein erfolgreiches Transkriptions-Reportersystem entwickelt, welches auf dem C-terminal angefügten, grün fluoreszierenden Protein (GFP)

Zusammenfassung

basiert. Durch Variation der Kultivierungsbedingungen konnte eine 26-fach höhere, heterologe Sodorifenproduktion im Vergleich zum Wildtyp-Stamm erzielt werden, wobei die Reinheit der Substanz im Rohextrakt bei über 90% lag.

Ambigole sind polychlorierte Naturstoffe, welche von *Fischerella ambigua* 108b synthetisiert werden. Sie weisen interessante antibakterielle, antimykotische und zytotoxische Aktivitäten auf. Bisher wurden drei Derivate, Ambigol A-C, beschrieben. Sie bestehen aus drei Dichlorphenol-Einheiten, welche über Biaryl- und Biaryletherbindung verknüpft sind und weisen ungewöhnliche Bindungskonnektivitäten auf. Die beiden Chloratome an der zentralen Phenol-Einheit von Ambigol A und B befinden sich in relativer *meta* Position. Bei Ambigol C liegen die Biaryletherbindungen, welche die drei Phenole miteinander verbinden, in relativer *meta* Position zur Hydroxygruppe des zentralen Phenols vor. Bisher ist keine enzymatische Reaktion bekannt, welche entweder die Biarylbindungen oder die Chloratome in diesen Positionen erklären könnte. Daher wurde die Ambigol-Biosynthese im Rahmen dieser Arbeit auf mehreren Ebenen untersucht.

Die vorangegangene, bioinformatische Identifikation des potentiellen, 14,3 kb großen Ambigol (*ab*) BGCs in *F. ambigua* wurde auf Basis neuester genomischer Datensätze aktualisiert. Die Identität des *ab* Clusters wurde durch heterologe Expression in dem cyanobakteriellen Wirtssystem *Synechococcus elongatus* PCC 7942 bestätigt. Die zur natürlichen Transformation von Cyanobakterien verwendete Methode musste hierfür weiter entwickelt werden, da die Integration derart langer DNA Fragmente in das Genom von *Synechococcus* bisher nicht beschrieben wurde. Unter Verwendung von DiPaC konnte, durch das Halbieren des *ab* Clusters und die aufeinanderfolgende Integration beider Teile in verschiedene neutrale Stellen innerhalb des Chromosoms von *S. elongatus*, erfolgreich eine Doppelmutante erstellt werden, welche mindestens ein Ambigol-Derivat synthetisiert.

Das *ab* Cluster enthält die beiden Cytochrom P450 Enzyme Ab2 und Ab3, von welchen anzunehmen ist, dass sie für die Installation aller Biaryl- und Biaryletherbindungen in den Ambigolen verantwortlich sind. Die enzymatische Reaktivität beider Proteine wurden mithilfe verschiedener Ansätze charakterisiert. Im Anschluss an die heterologe Expression in *E. coli* BL21 und die darauf folgende Reinigung wurden *in vitro* Assays mit dem potentiellen Substrat 2,4-Dichlorphenol (2,4-DCP) durchgeführt. Die Strukturen der Reaktionsprodukte wurden durch Vergleiche mit chemisch synthetisierten Standards aufgeklärt. Ab2 katalysiert die Bildung eines Biarylether-verknüpften Dimers, während Ab3 hauptsächlich ein Biaryl-verbrücktes Dimer bildet. Beide beobachteten Produkte enthalten die neu geknüpften Bindungen in üblicher *ortho*-Position relativ zur Hydroxygruppe. Diese Ergebnisse wurden durch *in vivo* 2,4-DCP-Feedingassays mit Ab2 und/oder Ab3-exprimierenden, heterologen *E. coli* und *S. elongatus* Stämmen bestätigt.

Contents

Abstract	vii
Zusammenfassung	ix
Contents	xi
1 Introduction	1
1.1 Natural products as essential pharmaceutical leads	1
1.2 Volatile natural products	7
1.3 Secondary metabolites of cyanobacterial origin	11
1.3.1 Anabaenopeptins	15
1.3.2 Natural products found in <i>Fischerella ambigua</i>	17
2 Aims of this thesis	21
3 Results and Discussion	23
3.1 Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential	23
3.2 Direct Pathway Cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in <i>E. coli</i>	37
3.3 Investigations into ambigol biosynthesis	50
4 Summary and Outlook	61
4.1 Unlocking the biosynthetic potential of secondary metabolites with DiPaC on the example of anabaenopeptin	61
4.2 Utilizing DiPaC for heterologous expression of the volatile compound sodorifen	62
4.3 Studies on the biosynthesis of the structurally unusual ambigols of cyanobacterial origin	63
Bibliography	67

Contents

List of Abbreviations	79
List of Figures	81
List of Schemes	83
Attachments	85
A.1 Supplemental materials of publications and submitted manuscript	85
A.1.1 Direct Pathway Cloning (DiPaC) to unlock natural product biosyn- thetic potential	85
A.1.2 Direct Pathway Cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in <i>E. coli</i>	118
A.1.3 Investigations into ambigol biosynthesis	139
A.2 Approval letters from publisher	248
Acknowledgements	251

1 Introduction

1.1 Natural products as essential pharmaceutical leads

Natural products - also termed secondary metabolites - are organic compounds synthesized by living organisms such as bacteria, plants and fungi that are not essential for their survival. They continuously evolved over thousands of years to generate a certain advantage for the producer, e.g., by repressing competitors in the same habitat, repelling enemies or enabling ways of intra- and interspecies communication. Originating from standard metabolic intermediates such as acetyl- and propionyl-CoA, mevalonic acid, methylerythritol-4-phosphate and various amino acids, these substrates can undergo numerous modifications, finally leading to an almost infinite number of individual natural products [1, 2]. Our main interest in secondary metabolites arises from their evolution-based optimized biological functions, which can often be repurposed for pharmaceutical interests.

Long before famous Sir Alexander Fleming discovered the first antibiotic penicillin in 1928 by serendipity [3], humankind started to use especially plant-based natural products as a source of medication. We have knowledge about the application of more than 1000 plant-derived substances in Mesopotamia (2600 B.C.), including morphine (**1**)-containing *Papaver somniferum* used as an analgesic and *Cupressus sempervirens* oil for colds, coughing and bronchitis [4–7]. An ancient Egyptian papyrus dating back to circa 1500 B.C. describes over 850 healing plants such as the anti-inflammatory *Aloe vera* harboring aloe emodin (**2**) and aloin (**3**) and the labour stimulating *Ricinus communis* [8–10]. Records of Chinese traditional medicine from around 1000 B.C. are rich sources of natural product containing plants used for medical purposes, including *Artemisia annua* harboring anti-malarian artemisinin (**4**) and the pain-killing tubers of *Corydalis yanhusuo* [11, 12]. The Indian Ayurvedic system evolved simultaneously and describes more than 800 pharmaceutically active plants, e.g. the reserpine (**5**) containing *Rauwolfia serpentina* used to treat snake bites or *Withania somnifera* exhibiting anti-inflammatory properties [4, 6, 13]. A structural overview on the so far mentioned natural products is given in Figure 1.

1 Introduction

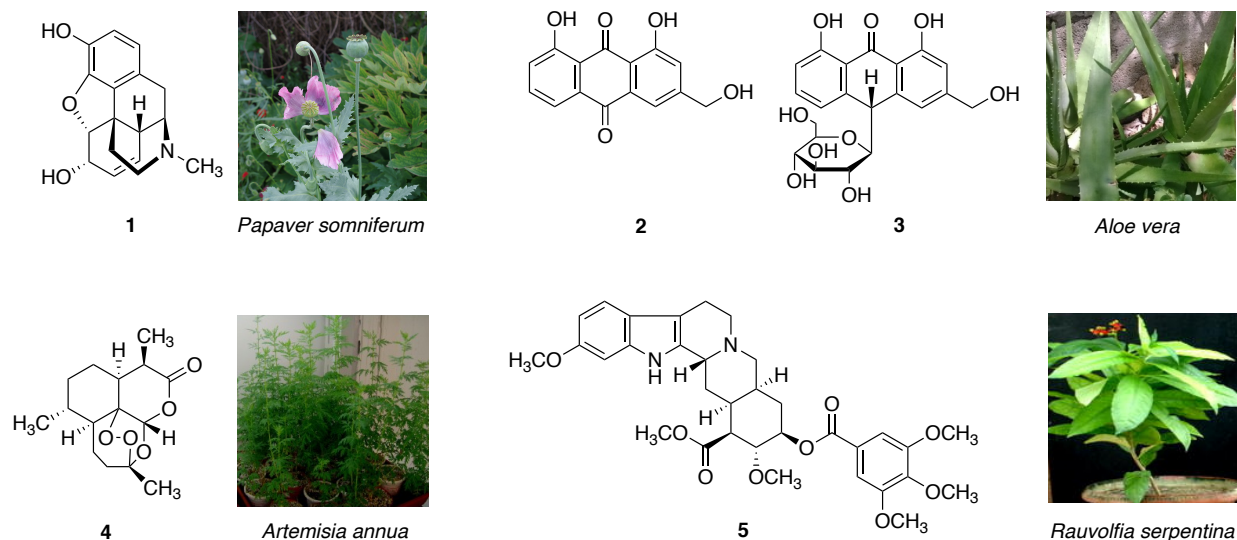


Figure 1. Pharmaceutically active secondary metabolites found in healing plants used by various ancient cultures. Pain-relieving morphine (**1**) from *Papaver somniferum* used in Mesopotamia [4, 7, 14], aloe emodin (**2**) and aloin (**3**) found in *Aloe vera* applied as anti-inflammatory drug in old Egypt [15, 16], anti-malarial artemisinin (**4**) from *Artemisia annua* used in traditional Chinese medicine [12, 17] and reserpine (**5**) contained in *Rauvolfia serpentina* recommended to treat snake bites by the Indian Ayurvedic culture [4, 18].

Historical knowledge of pharmaceutically active plants in the Western world is mainly based on the Greek and Roman civilizations. Great natural scientists and physicians such as the Greeks Theophrastus (300 B.C.), Dioscorides (100 A.D.) and Galen (200 A.D.) as well as famous Roman Pliny the Elder (100 A.D.) gathered and preserved information about the collection, storage and use of medicinal herbs to treat various diseases [19]. Within monasteries in England, France and Germany, this expertise was conserved during the Dark and Middle Ages from the 5th to the 12th century. Beyond that, the Arabs did not only retain this knowledge, but expanded it by including their own know-how, as well as parts of the traditional Chinese and Indian experience, which were so far unknown in the Greece and the Roman empires [6]. In the 10th century, the Arab and the Greco-Roman knowledge met in the south of Italy, which is said to be the start of the famous Salerno school of medicine, a sort of antecedent to our modern universities. The invention of the letterpress by Johannes Gutenberg in the 15th century allowed the effective distribution of this expertise throughout entire Europe, e.g., by books like “The Mainz Herbal” and the “German Herbal”, both written in 1484 by Peter Schöffer, the “Herbarium Vivae Eicones” written by Otto Brunfels in 1530 or the “Kreütter Buch” written in 1546 by Hieronymus Bock [20].

1.1 Natural products as essential pharmaceutical leads

One of the most important milestones in natural product pharmaceutical research in modern times was the isolation of pure **1** (see Figure 1) from the opium poppy plant in 1804 by the German apothecary assistant Friedrich Sertürner [21]. Not only did he discover a new class of drugs, the alkaloids, but his finding also led to a strong intensification of attempts to isolate other biologically active molecules from medicinal plants described so far. In the following, many more pharmaceutically utilizable compounds, mostly alkaloids, were discovered. This includes quinine (**6**) from *Cinchona calisaya* to treat malaria [22], cocaine (**7**) with stimulating and hunger-suppressing effects found in *Erythroxylum coca* [23], and codeine (**8**) present in *Papaver somniferum*, which is a pain reliever as well as a coughing medicine [24] (see Figure 2). The apothecaries who isolated, purified and investigated those substances can be seen as the ancestors of modern pharmaceutical companies, of which H. E. Merck, who worked in Darmstadt, Germany, was in retrospect amongst the most successful ones [25].

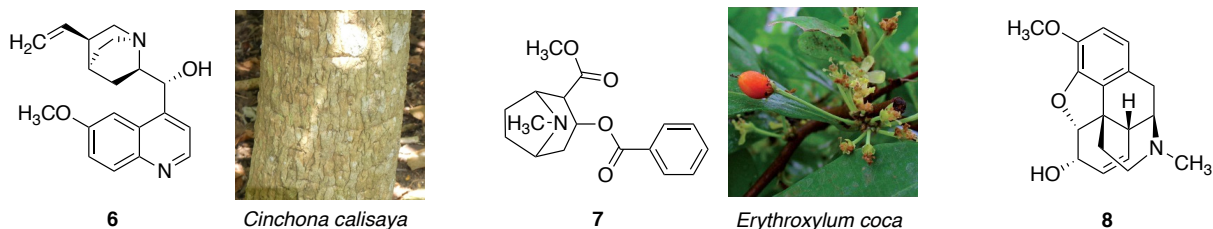


Figure 2. Structures of natural products isolated in their pure form for the first time in the 19th century. Quinine (**6**) found in the bark of *Cinchona calisaya* is active against the malaria-causing *Plasmodium falciparum* [22, 26]. The stimulating substance cocaine (**7**) was isolated from *Erythroxylum coca* [23, 27] and codeine (**8**), used to treat pain and coughing, was extracted from *Papaver somniferum* also harboring **1** [24].

However, with the upcoming field of chemical synthesis evolving more and more, isolation of (new) secondary metabolites was later neglected by efforts focusing on fully synthesizing already known compounds of natural origin in the laboratory. This seemed to be a more promising way to increase production rates while improving the purity and quality at lower costs [20]. In 1853, salicylic acid, one of the oldest and widest known plant secondary metabolite to reduce pain and fever found in *Salix alba* was synthesized for the first time by Charles Gerhardt [28]. The probably most important breakthrough in natural product related pharmaceutical science is owed to Alexander Fleming, who, as mentioned before, discovered the antibiotic penicillin at the beginning of the 20th century. The substance is produced by the filamentous fungus *Penicillium notatum*, which grew as a contamination on a *Staphylococcus* culture of A. Fleming. He discovered that the bacteria could no longer grow in close proximity to the fungus and subsequently, the first

1 Introduction

β -lactam antibiotic penicillin G (**9**) (see Figure 3) was assigned to be the cause of his observation [3]. Within the wartimes of the 1940s, **9** became a broadly used drug saving countless lives, which in 1945 led to awarding of the Nobel Prize in Physiology or Medicine to Alexander Fleming, Ernst Boris Chain and Howard Walter Florey for “The discovery of penicillin and its curative effect in various infectious diseases” [21]. This finding ushered in the “golden age of natural product drug discovery”, which was dominated by intense investigations to identify new biologically active agents from natural sources [29].

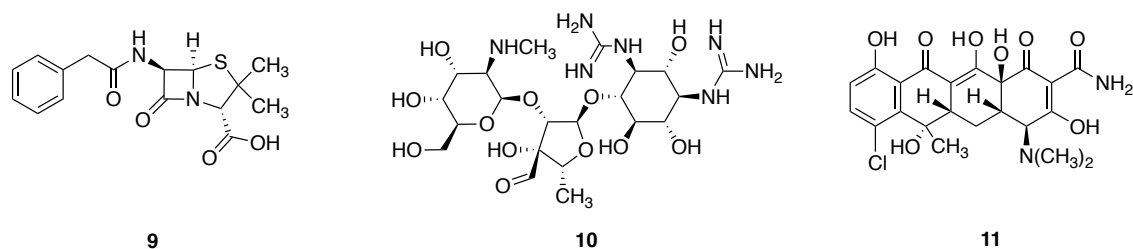


Figure 3. Important antibiotic substances discovered in the early 20th century. Penicillin G (**9**) was isolated from the fungus *Penicillium notatum* in 1928 [3], followed by the discovery of antimicrobial compounds produced by *Streptomyces* spp. such as streptomycin (**10**) from *S. griseus* and chlortetracycline (**11**) from *S. aureofaciens* [30].

In contrast to early times, research now focussed strongly on microorganisms as a new resource for the discovery of useful pharmaceutical substances [6]. In the following years, other important antibiotics such as the aminoglycoside streptomycin (**10**) from *Streptomyces griseus* (winning the Nobel prize in 1952) or chlortetracycline (**11**) from *Streptomyces aureofaciens* were identified (see Figure 3). Natural product research was now mainly fueled by pharmaceutical companies, which tremendously increased their efforts to develop novel drug types especially based on microbial fermentation technologies [30, 31]. The spectrum of identified biologically active secondary metabolites originating from microorganisms quickly grew bigger and bigger as immunosuppressive agents such as cyclosporin (**12**) from the fungus *Tolypocladium inflatum* and rapamycin (**13**) from *Streptomyces hygroscopicus*, cholesterol lowering substances such as mevastatin (**14**) found in *Penicillium citrinum* and lovastatin (**15**) produced by *Aspergillus terreus* as well as the anti-parasitic avermectin (**16**) from *Streptomyces avermitilis* were discovered in the 1970s (see Figure 4) [32].

This successful period of natural product based antimicrobial substance discovery ended in the 1990s and a period that is now called the “innovation gap” or “discovery void” in antibiotic research began [33]. The development of various important pharmaceuticals was still ongoing, e.g., fueled by the discovery of the anti-tumor agent paclitaxel in the bark of *Taxus brevifolia* used to treat various types of cancer [31]. However, in the field

1.1 Natural products as essential pharmaceutical leads

of antibiotics, no major new compound classes were identified. The quick occurrence of (multi)resistant bacteria all over the world became one of the biggest health problems of modern medicine. It was and still is mainly caused by the evolutionary pressure on the bacteria and a massive misuse of antimicrobial substances, combined with the lack of novel, innovative drugs to fight those pathogens.

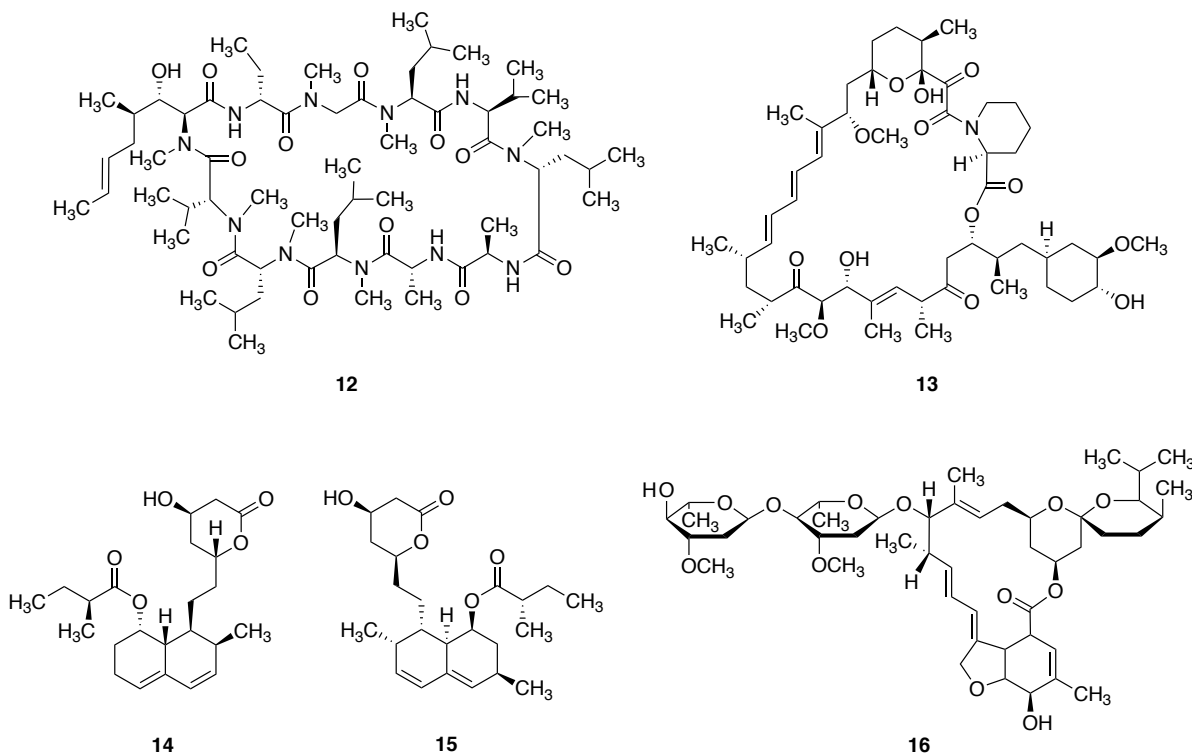


Figure 4. Structures of important natural product pharmaceuticals discovered in the 1970s. Cyclosporin (12) from *Tolypocladium inflatum* and rapamycin (13) from *Streptomyces hygroscopicus* have immunosuppressive activities, mevastatin (14) produced by *Penicillium citrinum* and lovastatin (15) from *Aspergillus terreus* are cholesterol lowering agents and avermectin (16) found in *Streptomyces avermitilis* exhibits anti-parasitic activities [32].

At the same time, many of the big pharmaceutical companies stopped their natural product isolation efforts and shifted their focus to synthetic and combinatorial chemistry. The continuous re-isolation of the same active compounds, the difficulties in obtaining sufficient amounts of biological material for further testing and developing, as well as their challenging structural complexity made the development of promising drug candidates too time consuming and expensive for many companies [20, 21]. Therefore, in the following two decades, classical natural product chemistry has mainly been replaced by endeavors to generate and screen synthetic libraries based on molecular target assays [34]. This strategy provides one big advantage as all substances introduced into activity test-

1 Introduction

ing had previously been successfully synthesized. Thereby, all time and money invested in (unsuccessful) attempts to synthesize a promising secondary metabolite could be circumvented. Despite all these efforts put into the development of so-called new chemical entities (NCEs) as innovative pharmaceutical leads, only one *de novo* chemical drug was successfully released by 2015, which is the kinase inhibitor sorafenib (**17**) (see Figure 5) used for renal and liver cancer treatment. In contrast, about 65% of all small molecule approved drugs currently on the market are based on natural products [35]. During the late 1990s, synthetic chemists realized that the myriads of new substances they had made so far were missing two important features compared to compounds of natural origin [36]. Secondary metabolites do not only possess an enormous structural and chemical diversity so far unmatched by any chemical library [37], but they have also been evolutionarily adapted to their target structures for thousands of years, finally leading to an optimal biological function. The randomly synthesized NCEs, by contrast, only had an incidental chance of interacting with a specific target without affecting any other structure or pathway within the human body.

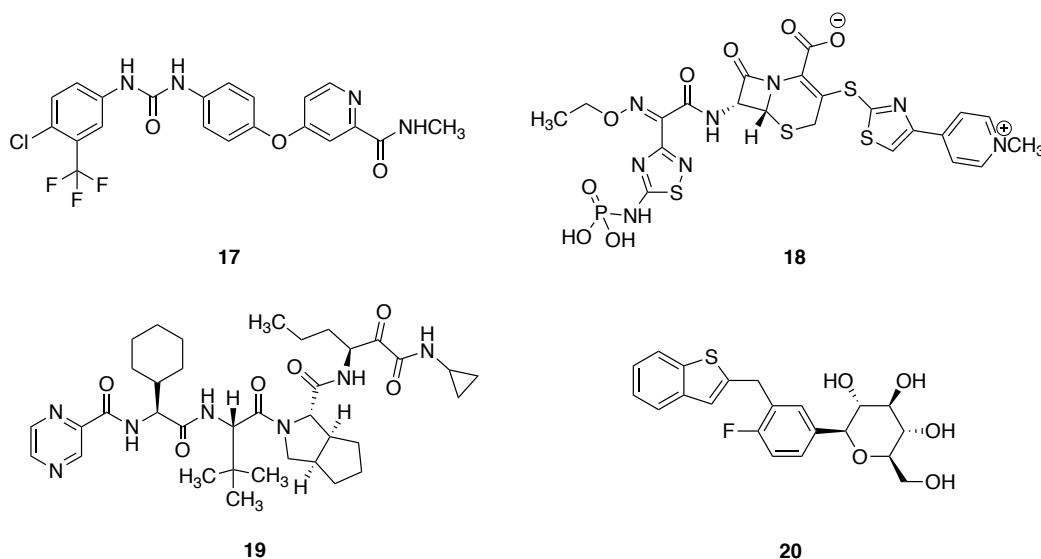


Figure 5. Modern drugs developed and approved within recent years. The anti-cancer drug sorafenib (**17**) is the only approved drug solely originating from *de novo* combinatorial chemistry approaches. Ceftriaxone fosamil (**18**) is an antibiotic natural product derivative and the virostatic telaprevir (**19**) as well as the anti-diabetic ipragliflozin (**20**) are both based on optimization of secondary metabolite core structures [35].

Since the end of the last century, new strategies in drug discovery have evolved, leading to the development of novel pharmaceuticals such as the antibacterial natural product derivative ceftaroline fosamil (**18**), the virostatic compound telaprevir (**19**) inhibiting hepatitis C, and the anti-diabetic ipragliflozin (**20**). The latter both are based on natural

product core structures (see Figure 5). Synthetic approaches now are more closely orientated on structural scaffolds of natural origin and the chemical derivatization of promising natural product lead structures [35]. High throughput *in vivo* and *in vitro* screening methods have been significantly improved to give more reliable results within a short period of time [38]. Additionally, great progress was made in cultivating of so far uncultivable microorganisms [39], in genome sequencing and bioinformatic identification of natural product biosynthetic gene clusters (BGCs) [40] and in the heterologous expression of those clusters [41]. All these advances will continue to significantly contribute to the discovery of new biologically active and pharmaceutically usable compounds of natural origin. It is undisputed that secondary metabolites will continue to be essential key sources for the development of novel, innovative therapeutics [38]. With about two million different species of plants, animals, fungi and microorganisms living on this planet and so far only less than 5% screened for bioactive compounds, an encouraging foresight into the future of drug discovery and development is allowed [42].

1.2 Volatile natural products

Volatile organic compounds (VOCs) are a fascinating group of molecules, which play an important role in the overall organization of life. They are produced by plants [43–47], fungi [48], bacteria [49], animals [50–53] and humans [54] and are typically small lipophilic molecules with low molecular mass (<300 Da), high vapor pressure and low boiling point. Their physical properties enable evaporation and distribution over long distances, thus facilitating intra- as well as inter-species communication. However, the mentioned characteristics also greatly exacerbate the investigation of volatile substances, especially when it comes to the elucidation of long distance interactions [55]. During the last decades, VOC related research mainly focused on induced interactions between plants and insects.

Plants constantly exchange information concerning their surrounding environment via chemical signals (biogenic VOCs) above and below the ground, which can result in activation of specific survival strategies [56]. (Negative) influences, such as diseases caused by fungi, parasitic plants or nematodes and environmental factors like light, oxygen, air pollution, soil composition or agrochemicals lead to the release of characteristic VOCs [57–59]. The first identified plant volatile was ethene (C_2H_4), which can regulate ripening, germination, elongation, senescence and responses against (a)biotic stresses [60, 61]. Due to its importance as a phytohormone, the mechanisms of action have been intensively studied. Ethene binds to a cellular membrane receptor protein, thereby initiating a phos-

1 Introduction

phorylation cascade resulting in changes within the transcriptional regulation of certain genes. To date, no other VOC recognition receptor has been identified in plants [60].

Microorganisms are able to produce and release a plethora of volatile natural products, some of which are well known to humankind as aroma and odor compounds made by bacteria during the production of cheese [62], wine [63] and other fermented goods [64]. Likewise, the unpleasant smells arising from rotting food that naturally warn us from consuming them are usually of bacterial origin. This process is an intuitive example for inter-species, VOC based communication [49]. The bacterial contributions to changes in odor and aroma in the processes of food transformation, e.g., during cheese production, have been thoroughly characterized [65–67]. In contrast, the general capability of bacteria to synthesize VOCs, as well as their complex natural functions and mechanisms of action, have not been investigated very deeply yet. One of the reasons for this are difficulties with assigning a VOC detected in a non-laboratory environment to either the producing or the receiving and reacting organism. Despite that, a number of interesting studies on bacterial volatile natural product chemistry under defined artificial conditions have been published. Some microorganisms can produce a great variety of VOCs simultaneously. Certain *Streptomyces* strains for example are capable of releasing up to 80 different compounds [68]. The purpose of most bacterial volatiles remains elusive so far, although several biologically relevant functions like communication and defense have been suggested [69, 70]. Bacterial VOCs with antifungal activity have for instance presumably evolved in order to restrict fast growing fungi within contested habitats [71–74]. They can also serve as attractants for various animals [75–77] or promote growth [78] and defense of plants [79].

The metabolic origins of volatile bacterial natural products are quite diverse. Hydrocarbons, aliphatic alcohols and ketones are most likely derived from fatty acid precursors. 9-Methyldecan-3-one (**21**) and (*S*)-9-methyldecan-3-ol (**22**) (see Figure 6) are released by the myxobacterium *Myxococcus xanthus*. Both compounds arise from leucine and in total three malonyl-CoA units, one of which harboring an additional methyl group [80]. Buibuilactone (**23**) is not only a scarab beetle pheromone [81], but is also produced by marine α -proteobacteria. The latter additionally synthesize (*Z*)-dodec-6-en-4-olide (**24**) differing from **23** only in the position of the double bond [82]. Interestingly, this molecule is also found in secretions of two antelope species and is produced by the bacterium *Planococcus citreus* living in the interdigital glands of these animals [83].

Many bacterial VOCs identified so far have been shown to be attractants for fruit flies including the nitrogen containing compounds triethylamine (**25**), 1-pyrroline (**26**), tetrahydropyridine (**27**), as well as methylated pyrazines (**28**) (see Figure 6) [75, 77, 84]. Pyrazines with longer side chains can be frequently found in flies, bees and ants and are believed to be of bacterial origin [85]. Methoxylated pyrazines (**29**) are widespread

in Nature and can be detected by humans even in very low concentrations due to their characteristic odor. They are thought to act as warning signals emitted by toxic insects [86], but are also responsible for the musty smell of wine corks [87] and act as attractants to the pineapple beetle *Carpophilus humeralis* [88, 89]. The biosynthetic origin of pyrazine containing VOCs has not yet been fully elucidated.

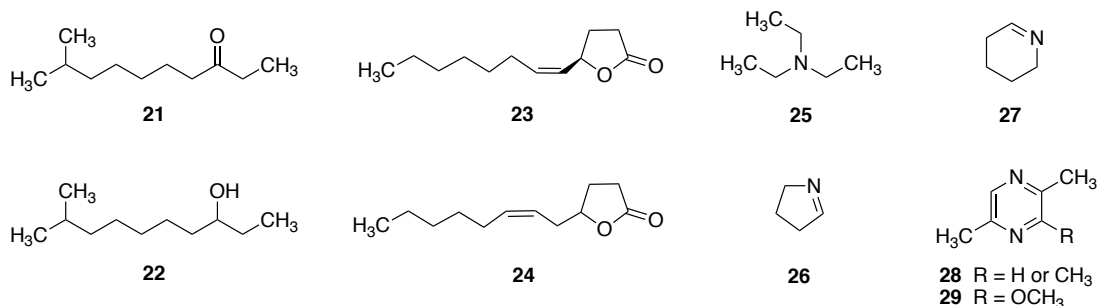


Figure 6. Structures of bacterial volatile organic compounds (VOCs). 9-Methyldecan-3-one (**21**) and (*S*)-9-methyldecan-3-ol (**22**) are synthesized by *Myxococcus xanthus* [80]. Buibuilactine (**23**) and (*Z*)-dodec-6-en-4-olide (**24**) are i.a. produced by marine α -proteobacteria [81, 82]. Nitrogen containing VOCs like triethylamine (**25**), 1-pyrroline (**26**), tetrahydropyridine (**27**) and methylated pyrazines (**28**) are fruit fly attractants [75, 77, 84], whereas methoxylated pyrazines (**29**) are found in various different contexts including warning signals from toxic insects [86].

Indole (**30**) (see Figure 7) is found in *Enterobacter* [90] and *Klebsiella* [91] and is the major volatile compound produced by several *Escherichia coli* strains, wherein it has been shown to regulate biofilm formation [92–94]. Benzothiazole (**31**) is formed by diverse *Streptomyces* [68], myxobacteria [80, 95] and cyanobacteria [96] and inhibits the development of fungal sclerotia, thus probably promoting survival of the producers in contested habitats [71]. Other sulfur containing VOCs contribute to the aroma of fermented foods like the S-methyl thioesters (**32–35**) produced by *Lactococcus* and *Lactobacillus* during cheese ripening processes [62, 97–103]. They are further believed to play a role in cloud formation and climate regulation [104–106]. Most of these molecules are derived from L-methionine catabolic pathways [49, 103] and are continuously released in large quantities by marine (micro)organisms [107, 108].

Terpenoids are biosynthetically derived from the two universal building blocks dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP), both arising either from the mevalonate or the deoxyxylulose phosphate pathway. Typical terpene VOCs show either a monoterpene (C_{10}) or a sesquiterpene (C_{15}) core structure [49]. The most prominent terpenoid volatiles of bacterial origin are geosmin (**36**) and 2-methylisoborneol (**37**) (see Figure 7). Both compounds contribute to the characteristic odor commonly

1 Introduction

associated with musty soil and are produced by a variety of different microorganisms including actinomycetes, myxobacteria and cyanobacteria [109, 110].

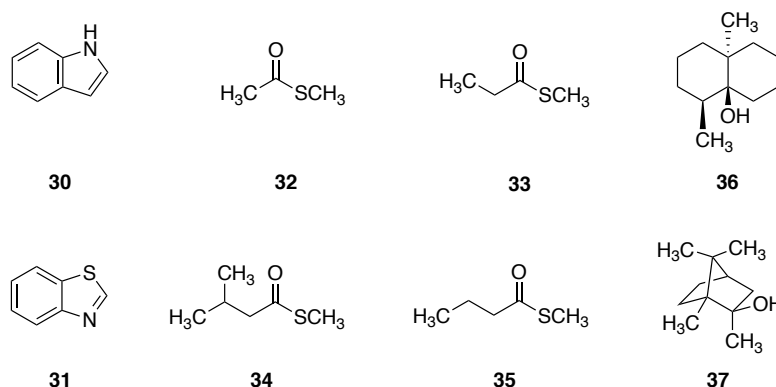
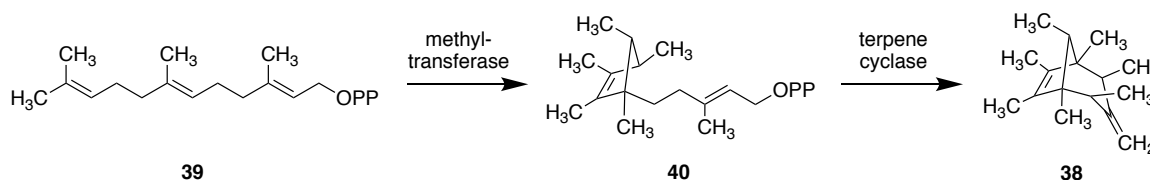


Figure 7. Further VOCs of bacterial origin. Indol (**30**) is the main volatile produced by *E. coli* [92, 93]. Benzothiazole (**31**) is synthesized by various different bacterial genera and has an inhibitory effect on fungal growth [71]. *S*-methyl thioesters (**32-35**) are associated with the ripening of cheese [62, 97–103] as well as with climate regulation processes [104–106]. Geosmin (**36**) and 2-methylisoborneol (**37**) are highly abundant bacterial VOCs that are responsible for the characteristic odor of soil [109, 110].

Sodorifen (**38**) is a methylated sesquiterpene ($C_{16}H_{26}$) with an unusual bicyclo[3.2.1]-octadiene structure (see Scheme 1) [111, 112] that is synthesized by various *Serratia plymuthica* strains. It originates from farnesyl pyrophosphate (FPP) (**39**), which is formed within the deoxyxylulose phosphate pathway. Recent investigations on the involved catalytical mechanisms revealed a surprising cyclization activity of the SAM-dependent methyltransferase present in the sodorifen BGC. Hence, **39** is methylated and initially cyclized by a single enzyme, leading to the intermediate pre-sodorifen (**40**), which is then transformed into **38** by a terpene cyclase [113]. The amount of emitted **38** strongly varies within different *S. plymuthica* strains [114] and can be upregulated upon co-cultivation with the pathogenic fungus *Fusarium culmorum* [115]. The biosynthesis of **38** was further shown to be under control of the carbon catabolite repression system [116]. The biological role and function of **38** still remain elusive.



Scheme 1. Sodorifen biosynthesis. Starting from FPP (**39**), a methyltransferase installs one methyl group and the first cycle to form pre-sodorifen (**40**), which is then cyclized to the final product sodorifen (**38**) by a terpene cyclase [113].

Another important part of VOC-centered research deals with volatiles emitted by the human body. The focus is hereby set to detect abnormalities or changes that may harbor valuable information for the diagnosis of various diseases. Volatile metabolites emitted in body fluids and tissue can be non-invasively identified in the headspace area above the human skin and in exhaled air, as well as in sweat, urine, blood and saliva [117–121].

1.3 Secondary metabolites of cyanobacterial origin

Cyanobacteria are an ancient lineage of photosynthetic prokaryotes that can be found all over the world [122]. They are probably best known as blue-green algae dwelling in freshwater and coastal areas, but are also occurring in deep seas and even in terrestrial habitats [123]. Cyanobacteria have always played an important role in the ecosystem. They were responsible for the evolution of the early atmosphere [124] and still have a significant ecological impact by providing oxygen through photosynthetic carbon fixation, converting nitrogen into organic forms and providing a nutrition source for many other organisms [125]. This phylum of photosynthetic bacteria is well known for the prolific production of a range of highly diverse bioactive natural products including VOCs, potential drug candidates and dangerous toxins [122, 126, 127]. Recently, some cyanobacteria received increasing attention from biotechnologists due to their potential applications in the ecofriendly production of important energy sources such as biodiesel and bioethanol [128].

Secondary metabolites of cyanobacterial origin have been shown to exhibit promising antibacterial [129], anticancerous [130, 131], antiplasmodial [132], antiviral [133] and immunosuppressive activities [134, 135]. The dragonamides, for example, are linear lipopeptides (see Figure 8) isolated from different *Lyngbya* strains. Dragonamide A (**41**) features antiparasitic [136–138] and dragonamide C (**42**) anticarcinogenic [139] properties. Cryptophycin 1 (**43**) is a macrolidic depsipeptide that was first isolated from *Nostoc* sp. ATCC 53789 as an antifungal lead structure by Merck [140]. Undergoing further bioactivity screenings, **43** was found to display potent cytotoxicity against various different types of carcinoma cell lines by inhibition of the microtubule polymerization [141, 142]. To date, more than 25 natural analogs of **43** have been isolated, which differ, e.g., in their chlorination pattern and the absolute configuration of the α -carbon of the tyrosine ring [143–145]. The semi-synthetic analog cryptophycin 52 (**44**) made it to phase II clinical trials, where it was finally rejected due to severe side effects. However, encouraging data from other derivatives exist, which support further evaluation of this compound class [146–148].

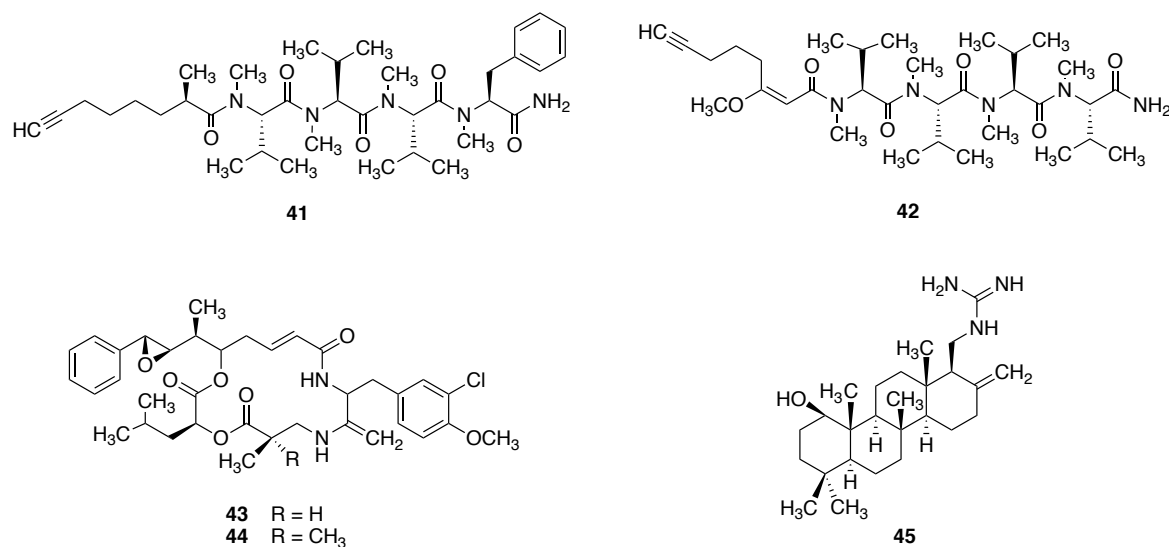


Figure 8. Cyanobacterial natural products with high pharmacological potential. Dragonamide A (**41**) exhibits antiparasitic and dragonamide C (**42**) anticarcinogenic activity. Both were isolated from *Lyngbya* sp. [136–139]. Cryptophycin 1 (**43**) was discovered as antifungal compound from *Nostoc* sp. ATCC 53789, whereupon the semi-synthetic cryptophycin 52 (**44**) was developed for cancer treatment. The antibacterial scytonema (**45**) produced by *Scytonema* spp. is one of the few terpene secondary metabolites isolated from cyanobacteria so far [149].

Terpenoides are a class of natural products that are not commonly found in cyanobacteria [150–152] and scytonema (**45**) is the first identified cyanobacterial scalarane sesquiterpene. It is synthesized by the terrestrial strain *Scytonema* spp. and shows antibiotic activity against various dangerous bacteria including *Bacillus anthracis* and *Mycobacterium tuberculosis*. Its six-membered tetracyclic structure is connected to a guanidine group that is crucial for the observed antimicrobial effect [149].

Besides their undoubtedly high potential for natural product based pharmaceutical research, some cyanobacteria are responsible for one of the most demanding ecological problems of our time. Given certain environmental conditions such as high nutrient loads, low turbidity, warmth and sunlight, exponential growth of cyanobacteria can episodically lead to harmful algal blooms (HABs) (see Figure 9). These blooms do not only affect the color, taste and scent of the particular waterbody, but can also become a serious health risk if the involved cyanobacteria are capable of synthesizing toxic secondary metabolites. Those so-called “cyanotoxins” usually accumulate inside the cells and are released into the water when cells are damaged or die. Therefore, they are often still detectable long after the actual bloom has dissipated [123, 125, 153].

Cyanotoxins have been shown to exhibit hepatotoxic, neurotoxic and dermatotoxic effects on humans and animals [125, 154, 155]. Many of them have the potential to bioaccumulate

in the tissue of plants and animals feeding on cyanobacteria and are thereby transferred within the food chain until they eventually end up in the human body [156–158]. Among the most dangerous cyanotoxins associated with HABs are the cyclic peptide microcystins (MCYSTs) and the alkaloids anatoxin-a (**46**), lyngbyatoxin A (**47**) and saxitoxin (**48**) (see Figure 9) [154, 155].

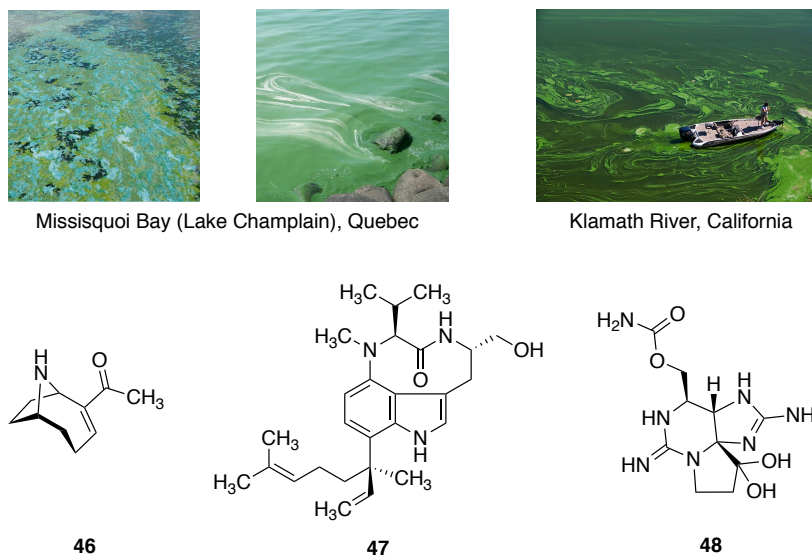


Figure 9. Cyanobacterial blooms are often enriched with toxic natural products. Blooms in the pelagic (left) and littoral area (middle) of Missisquoi Bay at Lake Champlain near Quebec in Canada [159] and in the Klamath River near Copco, California, in the USA (right). Typical cyanotoxins found in hazardous blooms are the neurotoxic anatoxin-a (**46**), the dermatotoxic lyngbyatoxin A (**47**) and the neurotoxic saxitoxin (**48**) [155].

The MCYSTs are by far the most detected freshwater cyanotoxins worldwide and have also been found to be produced by terrestrial cyanobacteria [160–164]. While chronic intoxication with low amounts may lead to liver or colorectal cancer [165–167], contact with larger amounts has been associated with lethal intoxication of wildlife, livestock and humans [165, 168]. MCYSTs are mainly transported into the liver where their inhibition of type 1 and 2A phosphatase causes tissue destructions finally leading to haemorrhagic shocks and death [169–171]. The MCYSTs are cyclic heptapeptides (see Figure 12) synthesized by a hybrid nonribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) pathway [172] with over 100 derivatives described so far [173].

Anatoxin-a (**46**), also known as “very fast death factor”, is a neurotoxic alkaloid produced by various genera of freshwater cyanobacteria, including *Anabaena*, *Microcystis*, *Planktothrix* and *Nostoc* [174–179]. It irreversibly binds and inactivates nicotinic and muscarinic acetylcholine receptors and thereby leads to paralysis and, in severe cases, death via asphyxiation. Anatoxin-a (**46**) was firstly isolated in the 1970s in Canada in

1 Introduction

the course of investigations on unexplainable cattle fatalities [180]. Since then, many more cases of animal intoxications especially affecting canines have been reported [181]. Compared to other cyanotoxins, **46** is a rather small bi-cyclic nonane including a secondary amine and an attached acetyl group (see Figure 9). So far, only three analogs of **46** are known, which feature either a propionyl instead of an acetyl group [182] or a reduction at the C7-C8 bond [180, 183]. The dermatotoxic effect of **47** was first discovered in 1979 when swimmers came in contact with a *Moorea producens* HAB in Hawaii and, as a result, suffered from symptoms of seaweed dermatitis [184]. Further reports even describe severe poisoning and death caused by consumption of contaminated turtle meat [185, 186]. Lyngbyatoxin A (**47**) features an indolactam ring and a linalyl side group attached at C-7. So far, seven natural derivatives of **47** have been isolated [184, 187–190].

Saxitoxin (**48**) and its derivatives are commonly known as “paralytic shellfish toxins” as they are associated with paralytic shellfish poisoning. It is the most toxic natural product of cyanobacterial origin and overall belongs to the deadliest known natural molecules by now [191]. The neurotoxic tricyclic alkaloids (see Figure 9) are synthesized by several different freshwater cyanobacteria species including *Cylindrospermum* and *Planktotrix* [192, 193] as well as by marine eukaryotic dinoflagellates [194–196]. To date, more than 57 naturally occurring derivatives of **48** have been isolated, all exposing different combinations of functional groups including hydroxylation, carbamylation, sulphation or acetylation [197]. The toxic effect of **48** is mediated by its binding to voltage-gated ion channels, which can cause weakness, nausea, paralysis and, in extreme cases, respiratory failure and death [198–200]. Due to its lipophilicity, **48** bioaccumulates in resistant aquatic organisms such as shellfish and thereby leads to intoxications of animals or humans consuming contaminated seafood [201–203]. Its occurrence is considered a global environmental and health hazard.

However, not all toxic cyanobacterial secondary metabolites are associated with HABs. Recently, the biosynthetic origin of polybrominated pyrrole- and phenol-based natural products **49-52** (see Figure 10) found in the marine environment has been elucidated [204]. Polybrominated diphenyl ethers (PBDEs, **49, 50**) were found to be synthesized by sponge-associated *Hormoscilla spongelliae* cyanobacteria [205] and various γ -proteobacteria. The latter are additionally capable of synthesizing brominated bipyrrols, e.g., **51** as well as phenol-pyrrole hybrid molecules like pentabromopseudilin (**52**) [204]. These toxic natural products can be found in all tropic levels of marine life. They have strong bioaccumulative potential and are detectable in plants [158], algae [206–208], animals [156, 209, 210] and even human tissue [211–214]. Their potential to interfere with mammalian hormone mediated signaling pathways and to further inhibit essential enzymatic reactions is responsible for their high toxicity levels [214–218].

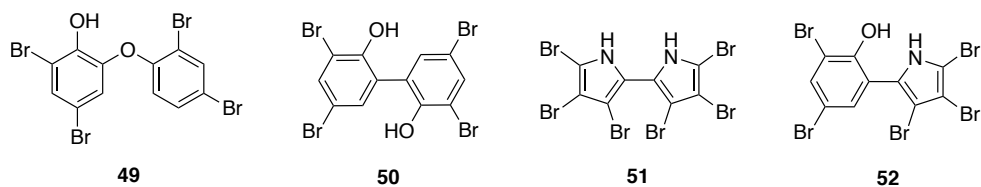


Figure 10. Polybrominated aromatic compounds produced by marine (cyano-)bacteria. Polybrominated diphenyl ethers (**49**, **50**) are synthesized by marine cyanobacteria and γ -proteobacteria. The latter are additionally capable of producing bipyrroles (**51**) and phenol-pyrrole hybrid molecules such as **52** [204, 205].

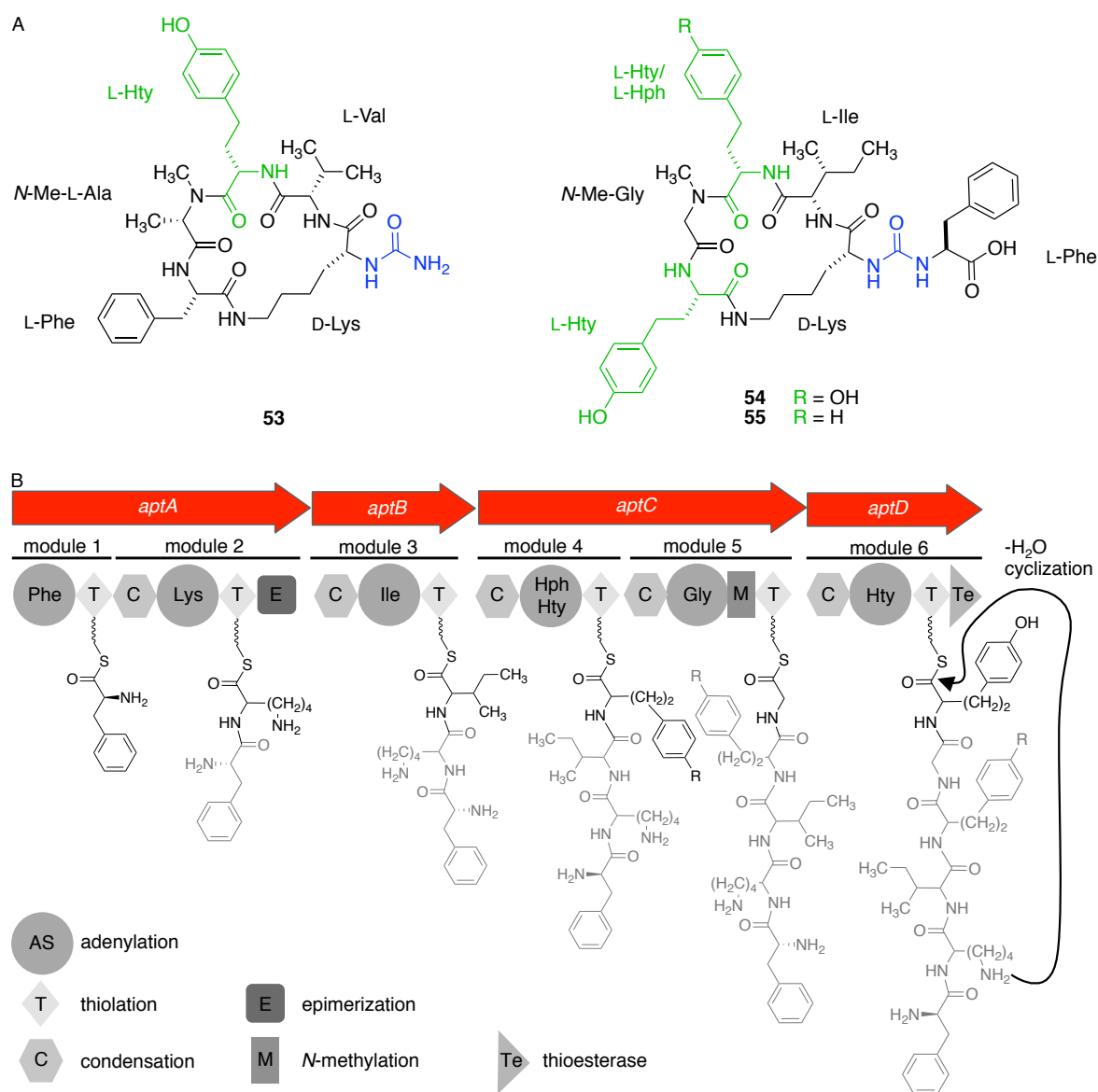
1.3.1 Anabaenopeptins

The anabaenopeptins are a group of hexapeptidic secondary metabolites often found in HABs that are synthesized by several different cyanobacterial genera such as *Planktothrix*, *Microcystis*, *Anabaena* and *Nostoc* [219–222]. Their nomenclature follows irregular principles and is based on the producing organism, the strain source, the place of isolation or their molecular weight [223]. So far, at least 104 different anabaenopeptin variants of cyanobacterial origin are known [224]. Some of them exhibit impressively potent protein inhibitory activities towards phosphatase 1 and 2A [223] and different types of exopeptidases including serine proteases and zinc-containing metallopeptidases (carboxypeptidases) [222, 225–230]. The active anabaenopeptin analogs are highly interesting compounds for the treatment of cardiovascular diseases [231–234] and certain types of cancer [235]. They usually occur in high intracellular concentrations in the producing organisms. Additionally - taking the observed inhibitory effects on eukaryotic enzymes into consideration - it is assumed that the anabaenopeptin synthesis is a chemical defense strategy against predators like chytrid [236] and herbivorous fungi [237] and amoeba [238].

Anabaenopeptins, e.g., **53-55** are nonribosomally synthesized cyclic peptides featuring a characteristic ureido motif that connects a conserved D-lysine from the macrocyclic pentapeptide to an exocyclic amino acid (see Scheme 2A) [223, 239, 240]. The peptide ring is formed by cyclization of the *C*-terminal carboxyl to the primary ϵ -amine of the *N*-terminal lysine. The amino acid composition of the different derivatives varies according to a X_1 -CO-[Lys- X_3 - X_4 -Me X_5 - X_6] motif and includes proteinogenic and non-proteinogenic amino acids [223, 241]. The anabaenopeptins are assembled by a modular NRPS enzyme complex. Every module features substrate specific domains responsible for adenylation, thiolation and, if necessary, epimerization or *N*-methylation of the single amino acids, as well as for the elongation of the sequence by peptide bond formation (compare Scheme 2B). Evolutionary events such as mutation and recombination of the adenylation domains, which activate specific amino acids prior to the peptide assembly, are responsible for the

1 Introduction

great variation observed within the composition of the so far described anabaenopeptins [222, 239, 242].



Interestingly, the origin of the characteristic ureido bond remains elusive even almost 25 years after the structure of the first anabaenopeptin was solved [219]. However, it can

be presumed that its formation must be encoded somewhere in the NRPS modules, as no other suitable genes within the respective BGCs or in close proximity to them could be identified so far [224, 244]. Recently, the truncated anabaenopeptin 679 (**53**), missing the exocyclic amino acid and instead featuring the ureido bond in terminal position (see Scheme 2A) was discovered and shown to still exhibit strong inhibitory activity against carboxypeptidase A [243].

Nostoc punctiforme PCC 73102 produces the two homoamino acid containing derivatives anabaenopeptin NZ857 (**54**) (2x L-Hty) and nostamide A (**55**) (1x L-Hty and 1x L-Hph) (see Scheme 2A) [244, 245]. The biosynthesis of the incorporated non-proteinogenic building blocks was elucidated in 2013 when the three responsible genes *hphA*, *hphB* and *hphCD* were identified up- and downstream of the anabaenopeptin NRPS genes. Based on homologous steps within the leucine biosynthesis, Koketsu *et al.* proposed a mechanism for the transformation of L-phenylalanine to L-Hph. It includes addition of acetyl-CoA, spontaneous hydrolyzation of the thioester and subsequent oxidation, decarboxylation and transamination. Investigations on the specificity of the *hph* system revealed L-tyrosine and fluorinated phenylalanine-derivatives as additional substrates for *hphA-CD*, thus also explaining the appearance of L-Hty in some anabaenopeptin derivatives [246].

1.3.2 Natural products found in *Fischerella ambigua*

Cyanobacteria of subsection V, also called stigonematales, are truly branching, filamentous microorganisms that are capable of forming a variety of morphologically advanced structures such as heterocysts, akinetes and hormogonia (see Figure 11). Compared to other cyanobacterial (sub)sections, they are currently rather underrepresented in studies dealing with the identification and characterization of bioactive natural products.

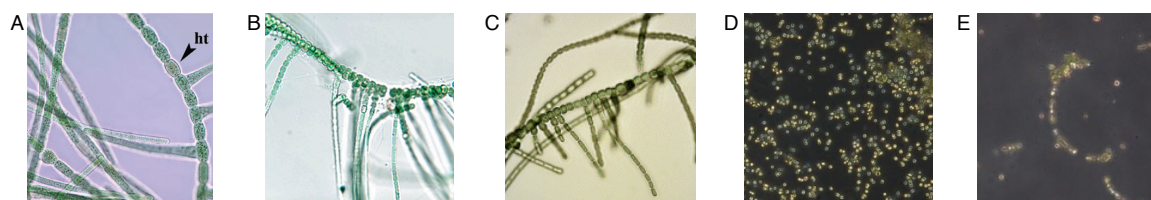


Figure 11. Morphologic characteristics of subsection V cyanobacteria. A) Branched *Stigonema* sp. CCM-UFV036 showing hormogonia (ht) [247], B) branched *Fischerella* 52-1 [123], C) branched *Fischerella thermalis* PCC 7521 [248] and images of D) unicellular and E) filamentous *Fischerella ambigua* 108b cells [249].

Members of this ecologically diverse group can be found in a wide range of terrestrial and freshwater habitats and are capable of producing a variety of interesting metabolites relevant to both human and environmental health. The most prominent representatives of

1 Introduction

those compounds are the hapalindoles and related indole alkaloids, e.g., the ambiguines, fischerindoles and welwitindolinones. The biosynthetic origin of their polycyclic carbon skeleton has been proposed to trace back to condensation of a tryptophan derivative with geranyl pyrophosphate [250–252]. All these substances must be considered pharmacologically as well as toxicologically relevant [123, 253].

The genus *Fischerella* has been intensively investigated due to its production of diverse natural products, including hepatotoxic MCYST derivatives, e.g., microcystin-LR (**56**) [254, 255] and photosynthesis inhibiting fischerellins, e.g., fischerellin A (**57**) [256, 257] (see Figure 12). *F. ambigua* species in particular were shown to produce a variety of biologically active secondary metabolites such as the ambiguine isonitriles, e.g., ambiguine K isonitrile (**58**) which exhibits promising antimicrobial as well as cytotoxic effects [258–262] and the related fischambiguines [261]. The latter can be isolated in non-chlorinated (fischambiguine A **59**) and chlorinated form (fischambiguine B **60**). Only the chlorinated version that also contains an additional epoxy feature instead of a terminal double bond exhibits antimicrobial activity [261].

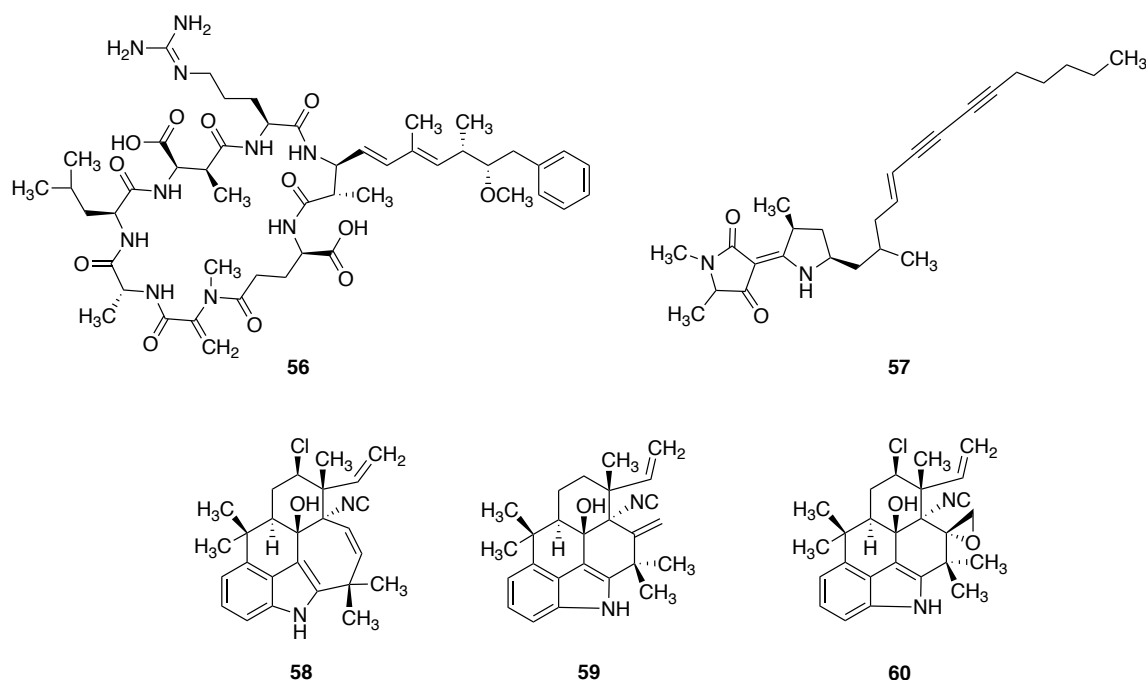


Figure 12. Secondary metabolites isolated from the genus *Fischerella*. Hepatotoxic microcystin-LR (**56**) produced by *Fischerella* sp. strain CENA161 [254], allelochemical fischerellin A (**57**) isolated from *F. muscicola* UTEX 1829 [256] as well as antimicrobial ambiguine K isonitrile (**58**), inactive fischambiguine A (**59**) and antibiotic fischambiguine B (**60**) all synthesized by *F. ambigua* UTEX 1903 [260, 261].

1.3 Secondary metabolites of cyanobacterial origin

The strain *F. ambigua* 108b (= Näg. Gomont, see Figure 11D and E) produces further chlorinated secondary metabolites, one of which is tjipanazole D (**61**) (see Figure 13). It consists of an indolocarbazole scaffold which most likely originates from tryptophan chlorinated at the C5-position of the indole ring [263, 264]. Various biological screenings did not detect any significant activity of **61** [265, 266]. In contrast, the polychlorinated triphenols ambigol A (**62**), B (**63**) and C (**64**), also synthesized by *F. ambigua* 108b, exhibit varying biological activities including antimicrobial, cytotoxic and virostatic effects [263, 265–267]. All ambigols are formally comprised of three dichlorophenol units that are linked by biaryl- and biaryl-ether-bridges, thereby leading to a total of six chlorine atoms present in each derivative.

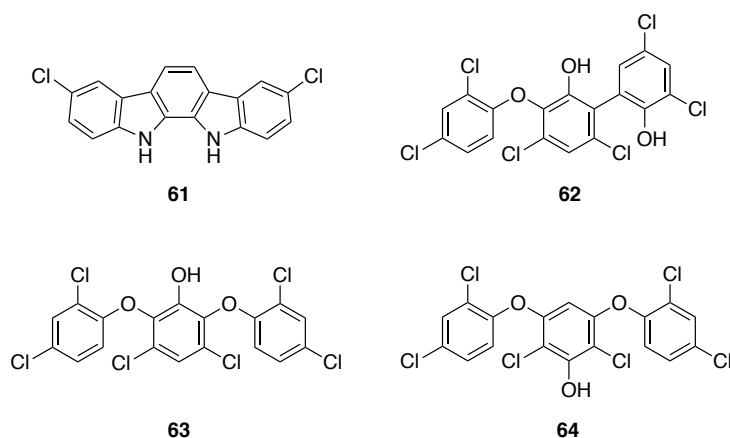
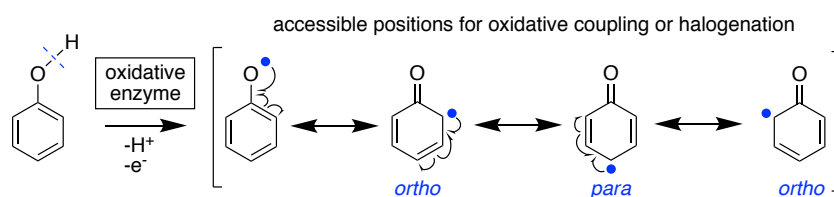


Figure 13. Chlorinated natural products synthesized by *F. ambigua* 108b. Tjipanazole D (**61**) shows no relevant biological activity whereas ambigol A (**62**), B (**63**) and C (**64**) exhibit antimicrobial, cytotoxic and virostatic effects in varying intensities [265, 267].

Ambigol A (**62**) shows strong antibiotic, antifungal and virostatic activity plus moderate cytotoxicity, whereas **64** only exhibits activities against gram positive bacteria and some fungi and **63** shows no significantly promising biological effects [265, 267]. The ambigols are structurally related to the previously mentioned polybrominated diphenyl ethers (PBDEs) (**49**, **50**) and therefore might as well have bioaccumulative potential, possibly including toxic effects on animals and humans. Besides the fact that they feature one additional phenol unit and are chlorinated instead of brominated, **62–64** exhibit a substantial structural difference when compared to the common PBDEs described so far. The halogens attached to the central phenol unit of **62** and **63** and the two biaryl-ether-linkages of **64** are located in relative *meta* position to the hydroxy group, whereas both halogens and biaryl-(ether-)bridges present in the PBDEs are located in *ortho* or *para* position of the biosynthetic 2,4-dihalophenol precursor. It is important to remember that enzymatic formation of biaryl- and biaryl-ether-linkages and halogenation of phenols

1 Introduction

usually follows a common principle. Firstly, an oxygen radical is generated by abstraction of a proton and an electron from the hydroxy group by an oxidative enzyme, e.g., a cytochrome P450 (CYP P450). This radical is mesomerically stabilized and can move into the relative *ortho* and *para* positions of the phenol ring, thus enabling bond formations at these locations, whereas the *meta* position remains inaccessible (see Scheme 3) [268]. It is therefore evident that the catalytic mechanism underlying the ambigol biosynthesis not only significantly differs from the ones described for the PBDEs but must also include some unprecedented enzymatic reactions.



Scheme 3. Accessible reactive positions in radical mediated enzymatic oxidative coupling and halogenation. After enzymatic abstraction of a proton and an electron, the resulting radical is mesomerically stabilized in relative *ortho* and *para* position on the phenol [268].

2 Aims of this thesis

The first part of this work is devoted to the development of a new cloning technique to capture natural product BGCs, the Direct Pathway Cloning (DiPaC). DiPaC is solely based on standard *in vitro* methods such as polymerase chain reaction (PCR) and Gibson Assembly or Sequence and Ligation Independent Cloning (SLIC) and is suited for capturing small- to mid-size gene clusters. The applicability of DiPaC for the cloning of challenging BGCs should be evaluated exemplarily for two model clusters. The 29.2 kb mid-size cyanobacterial anabaenopeptin (*apt*) cluster from *Nostoc punctiforme* PCC 73102 contains eight genes and required rearranging of certain genes prior to heterologous expression. The 4.6kb small sodorifen (*sod*) cluster from *Serratia plymuthica* WS3236 consists of only four genes and produces a VOC. Both clusters should be cloned and heterologously expressed in the host organism *Escherichia coli*. A simple, mRNA-independent reporter system allowing for direct verification of full gene transcription during heterologous expression should additionally be developed.

Elucidating individual steps of the ambigol biosynthesis was the aim of the second part of this work. Three ambigol derivatives **62-64** are produced by the terrestrial cyanobacterium *Fischerella ambigua* 108b. The compounds are comprised of three dichlorophenol units, that are linked via biaryl- and biaryl-ether-bridges. All three ambigols **62-64** exhibit unusual connectivity in *meta* position on the central phenol unit: ambigol A (**62**) and B (**63**) exhibit chlorine atoms and ambigol C (**64**) shows biaryl-ether-bridges in relative 3- and 5-position to the hydroxy group. To date, no enzymatically catalyzed reactions have been described that could explain the biosynthesis of such structures. Therefore, the 14.2 kb ambigol (*ab*) BGC should be bioinformatically characterized and subsequently heterologously expressed in the cyanobacterial host *Synechococcus elongatus* PCC 7942. The natural transformation method for introduction of foreign DNA into *S. elongatus* needed further development, as no technique for the integration of such long sequences into cyanobacterial host systems has been described so far. In addition, the two cytochrome P450 enzymes found within the cluster, Ab2 and Ab3, should be further investigated *in vitro* and *in vivo* to characterize their reactivity, as they are believed to introduce all necessary biaryl-(ether-)bonds present in the structurally diverging ambigols.

3 Results and Discussion

3.1 Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential

Based on: C. Greunke*, E. R. Duell*, P. M. D'Agostino*, A. Glöckle, K. Lamm, T. A. M. Gulder: Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Met. Eng.* **2018**, *47*, 334–345, DOI 10.1016/j.ymben.2018.03.010

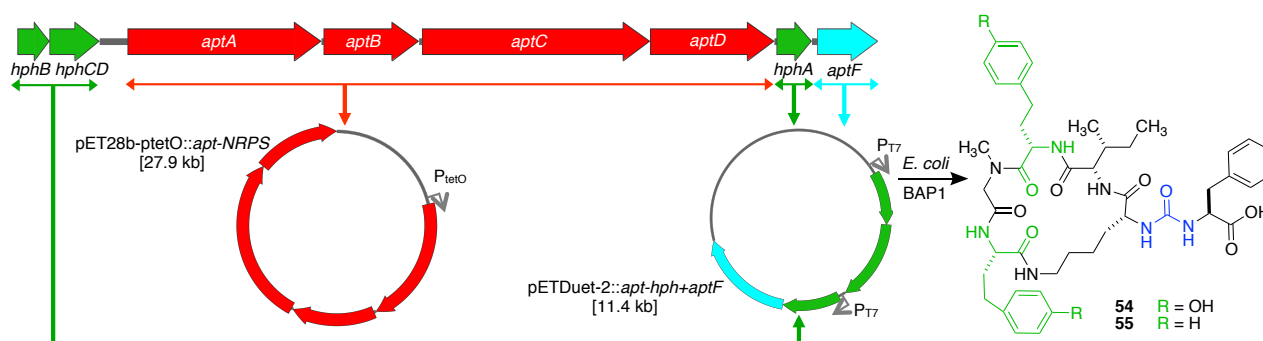
*equally contributing authors

In recent years, natural product research has been revolutionized by cheaper and more efficient whole-genome sequencing techniques combined with the development of bioinformatic tools to identify interesting BGCs *in silico*. To capture clusters of interest, *in vivo* methods have been developed to clone and heterologously express BGCs, including large-insert genomic DNA libraries as well as recombineering [269–274] based methods such as liner-linear homologous recombineering (LLHR) [275], linear-circular homologous recombineering (LCHR) [276], transformation-associated recombination (TAR) [277–280], exonuclease combined recombineering (ExoCET) [281] and Cas9-assisted targeting of chromosome segments (CATCH) [282]. These techniques are supplemented by PCR based *in vitro* methods such as circular polymerase extension cloning (CPEC) [283] and assembly of fragment ends after PCR (AFEAP) [284]. However, they all require specialized methods and can be very time consuming, particularly when the BGC needs further refactoring for successful heterologous expression. To overcome these restrictions, a complimentary *in vitro* cluster capturing method termed DiPaC was developed as part of this thesis. DiPaC is a synthetic biology cloning method for capturing small- to mid-sized BGCs directly into any expression vector. It is based on long-range PCR and *in vitro* DNA assembly techniques such as Gibson assembly or SLIC and allows for cluster refactoring directly during the capturing process, including promoter exchange, terminator deletion or reorganization of open-reading frames.

Cyanobacteria in general are a rich source for interesting bioactive secondary metabolites, but are hardly genetically tractable and often show slow growth under laboratory

3 Results and Discussion

conditions. Utilizing DiPaC, the 29.2 kb anabaenopeptin *apt* gene cluster from *Nostoc punctiforme* PCC 73102 was successfully cloned and heterologously expressed in *E. coli* BAP1 as part of this dissertation. The anabaenopeptins are a family of NRPS-derived cyclic hexapeptidic protease inhibitors produced by a variety of cyanobacteria that all harbor a preserved terminal ureido-bond [226, 285]. The two variants anabaenopeptin NZ857 (**54**) and nostamide A (**55**) produced by *N. punctiforme* PCC 73102 contain the homoamino acids L-Hph and L-Hty, whose biosynthesis is also encoded in the *apt* cluster (see Scheme 4) [244, 246].



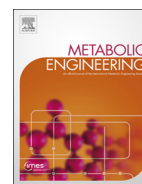
Scheme 4. DiPaC strategy for heterologous production of anabaenopeptin NZ857 (54**) and nostamide A (**55**) from *N. punctiforme* in *E. coli* BAP1.** The homoamino acid synthesis genes *hphA-CD* (green) and the transporter *aptF* (blue) were integrated into a pETDuet plasmid containing two IPTG inducible T7 promoters, whereas the NRPS genes *aptA-D* (red) were cloned into pET28b-ptetO under the control of the tetracycline inducible promoter PtetO. Subsequent activation of first T7 and then PtetO led to the successful heterologous expression of **54** and **55** in *E. coli* BAP1 double transformants.

By utilizing DiPaC, it was possible to amplify a 22 kb fragment of the *apt* BGC containing the four NRPS genes *aptA-D* responsible for anabaenopeptin assembly without mutation. The sequence was subsequently incorporated into a tetracycline inducible expression plasmid. The three genes catalyzing the formation of L-Hph and L-Hty, *hphA*, *hphB* and *hphCD* are located up- and downstream of *aptA-D* and therefore have been amplified and integrated separately into an IPTG inducible expression plasmid also harboring the transporter *aptF* (see Scheme 4). Formation of **54** and **55** in the *E. coli*-based heterologous expression system could be observed after the expression of the precursor synthesis genes and the transporter was induced ahead of the NRPS transcription. With this set of experiments, an easy access to secondary metabolites produced by slow-growing and hard to handle microorganisms such as cyanobacteria could be developed.



Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential

Christian Greunke^{1,2,3,4}, Elke Regina Duell^{1,2,3,4}, Paul Michael D'Agostino^{1,2,3,4}, Anna Glöckle², Katharina Lamm², Tobias Alexander Marius Gulder^{*,2,4}

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85748 Garching bei München, Germany

ARTICLE INFO

Keywords:

Metabolic engineering
Synthetic biology
Heterologous expression
Natural products
Biosynthesis

ABSTRACT

Specialized metabolites from bacteria are an important source of inspiration for drug development. The genes required for the biosynthesis of such metabolites in bacteria are usually organized in so-called biosynthetic gene clusters (BGCs). Using modern bioinformatic tools, the wealth of genomic data can be scanned for such BGCs and the expected products can often structurally be predicted *in silico*. This facilitates the directed discovery of putatively novel bacterial metabolites. However, the production of these molecules often requires genetic manipulation of the BGC for activation or the expression of the pathway in a heterologous host. The latter necessitates the transplantation of the BGC into a suitable expression system. To achieve this goal, powerful cloning strategies based on *in vivo* homologous recombination have recently been developed. This includes LCHR and LLHR in *E. coli* as well as TAR cloning in yeast. Here, we present Direct Pathway Cloning (DiPaC) as an efficient complementary BGC capturing strategy that relies on long-amplicon PCR and *in vitro* DNA assembly. This straightforward approach facilitates full pathway assembly, BGC refactoring and direct transfer into any vector backbone *in vitro*. The broad applicability and efficiency of DiPaC is demonstrated by the discovery of a new phenazine from *Serratia fonticola*, the first heterologous production of anabaenopeptins from *Nostoc punctiforme* and the transfer of the native erythromycin BGC from *Saccharopolyspora erythraea* into *Streptomyces*. Due to its simplicity, we envisage DiPaC to become an essential method for BGC cloning and metabolic pathways construction with significant applications in metabolic engineering, synthetic biology and biotechnology.

1. Introduction

With the discovery of the antibiotic penicillins from filamentous fungi and streptomycin from bacteria, bioprospecting microorganisms came into focus of academic and industrial research in the middle of the 20th century. After a few decades of the resulting golden age of antibiotic discovery, the challenges in natural product isolation and structure elucidation combined with increasingly high rediscovery rates resulted in severe cutbacks of pharmaceutical and agrochemical investments into natural product discovery projects in favor of (automated) synthetic compound libraries. Nevertheless, molecules derived from Nature continue to have significantly higher hit rates in biological screenings (Li and Vederas, 2009). It is thus not surprising that natural products and natural product-derived small molecules are still the most

important source and inspiration of modern chemotherapeutics (Newman and Cragg, 2016).

The field of bacterial natural product chemistry regained significant momentum after the publication of the genome sequences of *Streptomyces coelicolor* (Bentley et al., 2002) and *Streptomyces avermitilis* (Ikeda et al., 2003). While decades of previous intense natural product research on these model strains had only led to the discovery of a handful of specialized metabolites, these contributions revealed the presence of an about 10-fold higher number of biosynthetic gene clusters (BGCs) encoding unknown small molecules. Most of these pathways were simply not active (silent or orphan pathways) under the laboratory conditions used thus far, a phenomenon that holds true for most bacterial natural product producers. This notion has not only led to the directed discovery of a tremendous amount of novel chemistry

* Corresponding author.

E-mail address: tobias.gulder@tum.de (T.A.M. Gulder).¹ Contributed equally to this work.² Developed the DiPaC method and experimental design.³ Performed experiments.⁴ Wrote the manuscript.<https://doi.org/10.1016/j.ymben.2018.03.010>

Received 19 February 2018; Received in revised form 4 March 2018; Accepted 11 March 2018

Available online 13 March 2018

1096-7176/ © 2018 International Metabolic Engineering Society. Published by Elsevier Inc. All rights reserved.

from both these model organisms (Challis, 2014; Ikeda et al., 2014), but also laid the foundation for an entirely new field of novel natural product discovery, i.e., by the bioinformatic screening of genomic data, a process now denominated as genome mining (Bachmann et al., 2014; Ziemert et al., 2016). Reduced costs and improved NGS (next-generation sequencing) techniques (van Dijk et al., 2014) have since led to a massive increase of publicly available bacterial DNA sequence data (Land et al., 2015; Schorn et al., 2016). Genome mining of such data has been facilitated by the development of easy to use online bioinformatic tools, among them antiSMASH (Blin et al., 2017a, 2017b) and PRISM (Skinnider et al., 2015), that enable a fast and automated initial assessment of any genomic data. Ultimately, successful expression of the *in silico* predicted silent/orphan BGCs is necessary before the encoded natural product can be isolated for characterization (Rutledge and Challis, 2015). This can be achieved by pleiotropic methods such as screening of diverse cultivation conditions following the OSMAC (one strain / many compounds) approach (Bode et al., 2002), by using small molecule elicitors (Okada and Seyedsayamdost, 2017; Pettit, 2011), by studying metabolite production in the context of inter species interactions (Adnani et al., 2017), or by pathway specific methods such as the direct genetic manipulation of the BGC of interest in the natural strain or a suitable heterologous host (Rebets et al., 2014; Ren et al., 2017; Zhang, 2016). The latter requires reliable tools for the interception, mobilization and functional expression of the pathway of interest in the host organism of choice (Kim et al., 2015; Luo et al., 2016; Ongley et al., 2013).

Historically, the method of choice for DNA capturing has been the construction of large-insert genomic libraries, for which randomly sheared or partially digested genomic DNA is packaged into either cosmid-/fosmid- or BAC-based (bacterial artificial chromosomes) clone libraries, with each individual clone harboring ca. 35 or > 100 kb, respectively, of the target genome sequence. To reliably cover a typical *Streptomyces* genome, however, approximately 1000–2000 cosmid library clones need to be generated, individually picked, cultured, and tediously screened for the presence of the target BGC. In addition, the entire BGC of interest is rarely located on a single library insert, thus requiring downstream stitching to reassemble the complete pathway on a single construct for heterologous expression. This can be achieved by *in vivo* homologous recombination using the Red/ET system, consisting of Red $\alpha\beta\gamma$ from coliphage λ and RecET from the Rac prophage (Muyrers et al., 1999; Sharan et al., 2009; Yu et al., 2000; Zhang et al., 1998, 2000). This technology was established for the assembly of the myxochromide BGC (43 kb) from two cosmids (Wenzel et al., 2005) and has meanwhile been applied to a large range of other pathways from different organisms (Ongley et al., 2013). Recombination thereby is achieved between linear (target vector) DNA and circular, replicating DNA (e.g., cosmid/fosmid), known as linear plus circular homologous recombination (LCHR). To overcome the dependence on genome libraries and significantly streamline the BGC capturing procedure, a mechanistically different homologous recombination method utilizing the full length RecE and RecT solely using linear DNA, termed linear plus linear homologous recombination (LLHR), was recently developed (Fu et al., 2012). This method has been used to directly intercept pathways of up to ca. 52 kb in size (Duchaud et al., 2003). Larger BGCs can be cloned by capturing several pieces by LLHR, followed by subsequent other recombination steps for full cluster assembly, as recently achieved for the 106 kb salinomycin BGC (Yin et al., 2015). The latter cluster was recently also cloned in one piece following the ExoCET methodology, i.e., by applying CRISPR/Cas9 mediated DNA cleavage *in vitro*, thereby also overcoming the previous dependence on unique restriction sites that had to be present in proximity to the target BGC borders to linearize and digest the DNA to a suitable size prior to homologous recombination (Wang et al., 2017). In addition to LLHR, transformation-associated recombination (TAR), which takes advantage of the natural recombination ability of *Saccharomyces cerevisiae* (Kouprina and Larionov, 2006, 2008), has recently emerged as a tool in

biomolecular natural product research. This methodology can be used for the assembly of complete clusters from clone libraries, as well as for the interception of full pathways directly from genomic DNA (gDNA), e.g., the 56 kb colibactin pathway (Kim et al., 2010), or the so far largest TAR-cloned 73 kb taromycin BGC (Yamanaka et al., 2014).

LLHR and TAR are extremely powerful tools facilitating the mobilization of BGCs for heterologous expression experiments that will continue to have a tremendous impact on modern genome mining and natural product discovery. However, repetitive DNA sequences, as frequently observed in natural product BGCs, can negatively interfere with the desired homologous recombination event. In particular for TAR, the generation of highly transformation-efficient yeast spheroplasts, as well as the re-isolation of the fully assembled expression vector from yeast can be challenging for researchers without much experience in experimental work with yeast. Novel powerful methods such as Cas9-assisted targeting of chromosome segments (CATCH) utilize the CRISPR/Cas9 system to precisely digest gDNA for subsequent ligation into designated matching vectors using Gibson assembly and thereby facilitate capturing of fragments of up to 100 kb (Jiang and Zhu, 2016). This approach necessitates knowledge on the design and application of suitable CRISPR/Cas9 systems. LLHR, TAR and CATCH require highly specialized capturing vectors equipped with sequence homology arms to the target BGC and relatively large quantities of genomic DNA of the target organism for efficient recombination.

To further simplify capturing of BGCs, a broadly applicable, easy to use methodology for the fast cloning and refactoring of natural product biosynthetic pathways into any desirable vector system would thus be an excellent complementary tool for natural product chemistry laboratories. Herein, we describe Direct Pathway Cloning (DiPaC), a new strategy utilizing long-amplicon PCR and homology-based assembly for the *in vitro* construction of vector systems suitable for downstream heterologous expression of BGCs.

2. Materials and methods

2.1. General materials and methods

2.1.1. Bacterial strains and plasmids

Bacterial strains and plasmids used and generated in this study are listed in Tables S2 and S3 in the supporting information. Selection antibiotics were applied as follows: ampicillin sodium salt (Amp) at 100 $\mu\text{g}/\text{mL}$; kanamycin sulfate (Kan) at 50 $\mu\text{g}/\text{mL}$; chloramphenicol (Cam) at 12.5 $\mu\text{g}/\text{mL}$; nalidixic acid (NA) at 25 $\mu\text{g}/\text{mL}$.

2.1.2. DiPaC *in silico* simulations

For the development of cloning strategies and primer design, SnapGene (GSL Biotech; Version 4.0.4) or the Geneious (Kearse et al., 2012) software package (Version 8.1.9) and the NEBuilder assembly web tool (New England Biolabs; <http://nebuilder.neb.com>) were used. Maps of plasmids were constructed using the SnapGene software. The sequenced genomes of *Serratia fonticola* DSM 4576 (phenazines; *ehp* gene cluster), *Nostoc punctiforme* PCC 73102 (anabaenopeptins; *apt* gene cluster) (Rouhiainen et al., 2010) and *Saccharopolyspora erythraea* NRRL 2338/DSM 40517 (erythromycin A; *ery* gene cluster) (Oliynyk et al., 2007) were downloaded from the NCBI database using the accession numbers NZ_CP011254, NC_010628 and NC_009142, respectively.

2.1.3. Primer design and PCR

Primers for Gibson assembly were designed using the NEBuilder assembly tool. Gibson homology arms were composed of at least 18 nt with a calculated melting temperature of $\geq 50^\circ\text{C}$. HiFi DNA assemblies were simulated using SnapGene, in this way excluding multiple binding sites of the primers. Also, to avoid severe secondary hairpin structures and primer dimer formation, evaluations were performed using the bioinformatic tools OligoAnalyzer (Integrated DNA Technologies;

<https://eu.idtdna.com/calc/analyser>) and the OligoEvaluator (Sigma-Aldrich; <http://www.oligoevaluator.com>).

A standard 25 μL PCR reaction batch for long-amplicon cycling reactions consisted of: 1 \times Q5 reaction buffer, 100–200 μM deoxynucleotide triphosphates, 500 nM of forward and reverse primer, HighGC Enhancer (only for *ery* gene cluster), DNA template and 0.01–0.02 U/ μL Q5 High-Fidelity DNA polymerase (NEB). Template DNA amounts were 50 ng for *S. fonticola*, 15 ng for *N. punctiforme* and 100 ng for *S. erythraea*. Cycling was conducted using a T100 Thermal Cycler (Biorad) or a LifeECO Thermal Cycler (Biozym) as follows: 1.) Initial denaturation, 98 °C for 1 min; 2.) Denaturation, 98 °C for 10 s; 3.) Annealing, T_A for 20 s; 4.) Extension, 72 °C for t_{Ext} ; steps 2.) to 4.) were repeated in total for 30 cycles; 5.) Final extension, 72 °C for t_{Ext} , and 6.) End phase, 16 °C. T_A was estimated using NEBs Tm Calculator tool (<http://tmcalculator.neb.com>) for the initial binding parts of the primers and t_{Ext} was approximately determined as 30 s per 1000 bp for low GC organisms or 45 s per 1000 bp for high GC organisms.

Colony screening PCRs were performed using Taq DNA polymerase (NEB). Clones were picked and resuspended in 40 μL of pure water and examined in a 25 μL PCR batch composed as follows: Taq buffer (10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, pH 8.3 at 25 °C), 4% DMSO, 100 μM deoxynucleotide triphosphates, 200 nM of forward and reverse primer, 5 μL DNA template (bacterial suspension in water) and Taq DNA polymerase (0.025 U/ μL , NEB). Optimal cycling parameters were set as follows: 1.) Initial denaturation, 95 °C for 5 min; 2.) Denaturation, 95 °C for 45 s; 3.) Annealing, T_A for 30 s; 4.) Extension, 72 °C for t_{Ext} ; steps 2.) to 4.) were repeated in total 34 times; 5.) Final extension, 72 °C for t_{Ext} , at least 2 min, and 6.) End phase, 16 °C. T_A was estimated using NEBs Tm Calculator tool and a minimal T_A of 47 °C at least. t_{Ext} was calculated as 1 min per 1000 bp.

2.1.4. Cloning reagents

Q5 High-Fidelity DNA polymerase, deoxynucleotide triphosphates, NEBuilder HiFi DNA assembly master mix, various restriction endonucleases, if possible as High-Fidelity (HF) versions, and Antarctic phosphatase (AnP) were purchased from NEB. T4 DNA ligase was purchased from Jena Bioscience. Oligonucleotides were from Sigma Aldrich, synthesis option “deprotected and desalted”, resuspended at a stock concentration of 100 μM in pure water and stored on ice for immediate use or at -20 °C for long term storage. The peqGOLD Plasmid Miniprep Kit 1 (C-Line) from VWR was used to isolate plasmids from *E. coli* LB overnight cultures. For DNA extractions from agarose gels, the Monarch DNA Gel Extraction Kit (NEB) was applied. PCR products or linearized vector backbones were purified using the PCR Purification Kit from Jena Bioscience. Agarose was purchased from Sigma-Aldrich. SERVA DNA Stain Clear G (Serva) was applied according to the manufacturers' instructions in prestained agarose gels. Rotiphorese TAE buffer was used for agarose gel electrophoresis and was purchased from Carl Roth as a 50-fold concentrate.

2.1.5. Setup of HiFi DNA assembly reactions

The HiFi DNA assemblies were composed as recommended by the NEB guidelines: For a two-fragment assembly, at least 0.02 pmol of prepared vector backbone together with a one- to two-fold amount of insert were used in a total reaction batch of 20 μL . For three-fragment assemblies, DNA components were applied in equimolar ratios in the range of 0.03–0.2 pmol. In all cases, the isothermal reaction was performed at 50 °C for one hour.

2.1.6. Sanger sequencing of plasmid samples

Sanger sequencing was performed at GATC Biotech (Konstanz, Germany). Results were retrieved as ab1-type files and were analyzed using Geneious software. Sequence alignments to the related reference sequences were performed using Geneious Alignment with default settings.

2.1.7. Compound analysis by HPLC-MS, MS² and HR-MS

HPLC-ESI-MS and MS² experiments were conducted on an UltiMate 3000 LC System coupled to a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). The chromatographic HPLC separation was carried out on a Hypersil Gold aQ C18 column (150 \times 2.1 mm, 3 μm particle size). The following liquid chromatography methods were applied with water (A) and acetonitrile (B) as the eluents, both supplemented with 0.1% formic acid. For fast extract screening, the gradient was set as follows (method 1): Preconditioning 5% B (2.5 min); 5% B (0 min) \rightarrow 95% B (8 min) \rightarrow 100% B (8.4 min) \rightarrow 100% B (10.8 min) \rightarrow 5% B (11.2 min) \rightarrow 5% B (12 min), at a constant flow rate of 0.7 mL/min. For a better peak separation of components, the gradient was set as follows (method 2): Preconditioning 5% B (4 min); 5% B (0 min) \rightarrow 50% B (20 min) \rightarrow 100% B (20.5 min) \rightarrow 100% B (24.5 min) \rightarrow 5% B (25 min) \rightarrow 5% B (26 min), at a constant flow rate of 0.7 mL/min. Mass detection was performed in positive ionization mode. In the MS² measurements, 35% normalized collision energy (cid) was applied to fragment the mass $m/z = 858.4 \pm 3.0$ Da (anabaenopeptin NZ857), $m/z = 842.4 \pm 3.0$ Da (nostamide A), and $m/z = 734.5 \pm 3.0$ Da with SRM adjusted to fragment $m/z = 576.22 \pm 2.0$ Da (erythromycin A).

HPLC-ESI-HR-MS analysis was performed on a Thermo Finnigan LTQ FT-ICR equipped with a Dionex Ultimate 3000 separation module eluting on a Waters Xbridge C18 column (4.6 \times 100 mm, 3.5 μm particle size). The column temperature was maintained at 30 °C. The gradient was set as follows: Preconditioning 20% B \rightarrow 30% B (2 min), 30% B (7 min); 30% B \rightarrow 100% B (37 min) \rightarrow 100% B (42 min) \rightarrow 100% B \rightarrow 20% B (42.5 min) \rightarrow 20% B (47 min) at a constant flow rate of 1.1 mL/min. Interpretation of all recorded MS data was performed using the Thermo Xcalibur Qual Browser 2.2 SP1.48 software.

2.1.8. Structure elucidation using NMR

NMR spectra were recorded on a Bruker AVHD500 and a Bruker AV500-cryo spectrometer. The chemical shifts δ are listed as parts per million [ppm] and refer to $\delta(\text{TMS}) = 0$. The spectra were calibrated using residual undeuterated solvent as an internal reference ($\delta(\text{acetone}) = 2.05$ ppm for ¹H NMR; $\delta(\text{acetone}) = 29.8$ ppm for ¹³C NMR). The following abbreviations (or combinations thereof) are used to explain the multiplicities: s = singlet, d = doublet, dd = doublet of doublets, t = triplet, m = multiplet, b = broad.

2.2. Cluster specific materials and methods

2.2.1. Phenazines

2.2.1.1. *Bioinformatic analysis.* Genome mining of the *S. fonticola* DSM 4576 genome was performed using an initial preliminary analysis with the AntiSMASH software package (Version 3.0) (Weber et al., 2015). All nucleotide sequences were maintained and visualized within the Geneious software packages (Version 8.1.9).

2.2.1.2. *Bacterial culturing and gDNA isolation.* *S. fonticola* DSM 4576 was purchased from the Leibniz Institute DSMZ. *S. fonticola* was grown in Nutrient Broth (Carl Roth, Germany) at 28 °C and shaking at 200 rpm. A total of 2 mL of cells were collected by centrifugation followed by gDNA extraction using the Bacterial DNA Preparation Kit (Jena Bioscience, Germany) as described in the manufacturers' protocol. The gDNA was stored in TE buffer (1 mM Tris, 0.1 mM EDTA, [pH 8.0]) at -20 °C until use.

2.2.1.3. *DiPaC strategy for ehp.* Linear fragments for DiPaC of both vector and insert were generated using PCR (Section 2.1.3). A 9.7 kb fragment of the *S. fonticola* DSM 4576 *ehp* gene cluster was amplified using the Gibson primer pair gib_ehp-phenF/R (Table S1). These primers targeted the first gene within the *ehp* operon and contained all essential *ehp* core genes, a resistance gene and two species specific genes of unknown function. At the 5' end of the insert specific primer

set, a 30 bp homology region was added consistent with the terminal sequence of the PCR generated vector. The expression vector was generated using the specific primer pair pET28PD1-vecF/R (Table S1). This primer pair consisted of two outward facing specific primers and was positioned to remove the SUMO-tag and other non-essential components of the pET-28b-SUMO vector. Both linear insert and vector amplicons were purified by agarose gel electrophoresis. The purified vector was digested with *DpnI* to ensure all background template DNA was removed. The HiFi DNA assembly reaction was performed as described in Section 2.1.5. A total of 5 μ L of HiFi DNA assembly reaction mixture was chemically transformed into *E. coli* DH5 α . Positive pET-28b-PD1::*ehp* clones selected by screening PCR were confirmed using restriction digest and sequenced by sanger sequencing.

2.2.1.4. Heterologous expression of the *S. fonticola* DSM 4576 *ehp* cluster. The pET-28b-PD1::*ehp* expression vector was chemically transformed into the expression host *E. coli* BAP1 (Pfeifer et al., 2001) and single colonies were picked to provide 15 mL pre-expression cultures. These pre-expression cultures were used to inoculate expression cultures (1% v/v) in 1 L of LB medium supplemented with 50 μ g/mL Kan. Expression cultures were incubated at 30 °C with shaking (200 rpm) until an OD₆₀₀ of 0.6–0.8 was reached. Cultures were then cooled on ice (to ~18 °C), induced with IPTG to a final concentration of 0.25 mM and incubated for 16 h at 20 °C with shaking at 200 rpm.

The post-expression culture was collected by centrifugation at 10,000 \times g for 10 min at room temperature and the pellet washed with 0.9% NaCl. The cell pellet was frozen and stored at –80 °C until extraction. Extraction of the cells was performed using 1:1 MeOH:acetone with sonication for 30 min in a sonicator bath. Lysed cellular debris was removed by centrifugation (10,000 \times g for 15 min at 4 °C). Solvent extracts were desiccated using a rotary vacuum evaporator at 40 °C. Thereafter, the extract was dissolved in HPLC-grade MeOH and filtered through a Millex-GP syringe driven 0.22 μ m PES membrane filter (Millipore, USA) prior to HPLC analysis.

2.2.1.5. Analysis and isolation of phenazines. Methanolic extracts were analyzed by analytical HPLC on a JASCO system consisting of a UV-1575 Intelligent UV/VIS Detector, DG-2080–53 3-Line Degasser, two PU-1580 Intelligent HPLC Pumps, AS-1550 Intelligent Sampler and HG-1580–32 Dynamic Mixer controlled by the Galaxie chromatography software (Version 1.8.6.1) provided by Jasco. A 20–50 μ L sample was injected into a Eurosphere II 100–3 C18 A (150 \times 4.6 mm) column with integrated precolumn manufactured by Knauer. Wavelengths of 220 nm and 360 nm were used for phenazine detection and a PDA UV spectrum of 200–600 nm was obtained for each peak. HPLC chromatographic separation was performed with water (A) and acetonitrile (B) as eluents, both supplemented with 0.05% trifluoroacetic acid. The gradient was set as follows (method 3): Preconditioning 5% B (2 min); 5% B (0 min) \rightarrow 100% B (45 min) \rightarrow 100% B (54 min) \rightarrow 5% B (59 min), at a constant flow rate of 1 mL/min. Peaks of interest from pET-28b-PD1::*ehp* heterologous expression extracts were pre-purified using a LH20 sephadex (Sigma) column with MeOH as the mobile phase. Phenazine molecules were isolated from pre-purified fractions by semi-preparative HPLC for downstream NMR structural elucidation. Semi-preparative HPLC was controlled by a Jasco HPLC system consisting of an UV-1575 Intelligent UV/VIS Detector, two PU-2068 Intelligent prep. Pumps, a MIKA 1000 Dynamic Mixing Chamber (1000 μ L Portmann Instruments AG Biel-Benken), a LC-NetII/ADC, and a Rheodyne injection valve. The system was controlled by the Galaxie software and the eluent system consisted of solvents A (water) and B (acetonitrile), both buffered with 0.05% TFA. A Eurosphere II 100–5 C18 A (250 \times 16 mm) column with precolumn (30 \times 16 mm) provided by Knauer was used as the stationary phase. Compound detection was performed at 360 nm. Molecules were separated using a

flow rate of 7 mL/min and a gradient from 35% B to 40% B over a 50 min period. The column was washed with 100% B for 10 min and further equilibrated at 35% B for 10 min prior to the next injection.

2.2.2. *Anabaenopeptins*

2.2.2.1. Bacterial culturing and gDNA isolation. *N. punctiforme* PCC 73102 was obtained from the Collections des Cyanobactéries, Institut Pasteur, Paris, France. *N. punctiforme* was cultivated in BG-11 medium ([pH 8], Sigma-Aldrich, Germany) at room temperature for two weeks. High molecular weight gDNA was extracted using an optimized method adapted from D'Agostino et al. (2016), and was performed as follows: fresh or frozen cell pellets were washed with 0.9% NaCl once and resuspended in 5 mL lysis buffer (25 mM EDTA, 0.3 M sucrose, 25 mM Tris-HCl [pH 7.5]). The cells were freeze-thawed three times (liquid nitrogen / 50 °C) to assist in cell lysis. Lysozyme (Sigma-Aldrich) was added to a final concentration of 1 mg/mL and RNase (Carl Roth, Germany) to a final concentration of 10 μ g/mL, ensued by an incubation at 37 °C for 60 min. Proteinase K (Amresco, USA) was added to a final concentration of 0.5 mg/mL followed by the addition of SDS to a final concentration of 1% (w/v). The cell mixture was first incubated at 37 °C for 30 min and then at 55 °C for 30 min. NaCl was added to a final concentration of 1.3 M followed by the final concentration of 1% (v/v) CTAB solution (10% (w/v) CTAB (Sigma-Aldrich) in 0.7 M NaCl), with the mixture incubated at 65 °C for 10 min. A total of 1 vol of chloroform: isoamyl alcohol (24:1) was added to the cell lysis solution, mixed by inversion and incubated on ice while shaking for 30 min. The aqueous phase was removed and phenol-chloroform-isoamylalcohol (25:24:1) extraction was performed twice before the addition of 0.6 vol of isopropanol. The DNA cell pellet was washed with ice cold 70% ethanol and finally dissolved in TE buffer (1 mM Tris-HCl, 0.1 mM EDTA, [pH 8.0]). The quality of gDNA was analyzed by gel electrophoresis using 0.7% (w/v) agarose gels and quantified with a P330 NanoPhotometer (Implen, Germany). The gDNA was stored at –20 °C until further use.

2.2.2.2. DiPaC strategy for *apt*. All primers used for the cloning of the *apt* cluster are listed in Table S1, resulting plasmids can be found in Table S3. The genes of the *apt* cluster were amplified using Q5 DNA polymerase as follows: the NRPS part *apta-aptD* in one 22.0 kb fragment, *hphB-hphCD* in one 2.8 kb fragment, *hphA* and *aptF* in separate fragments of 1.2 kb and 2.1 kb, respectively. Optimal annealing temperatures (T_A) were determined experimentally using temperature gradients from 50 °C to 72 °C. Extension times were calculated with 30 s per 1000 bp, adding a plus of 2 min for fragments larger than 10 kb. The *NcoI* linearized and dephosphorylated pET-28b-ptetO vector was used to capture the *apta-D* fragment using HiFi DNA assembly, resulting in the generation of the pET-28b-ptetO::*apt-NRPS* expression construct. *hphB-CD* (*BamHI/NotI*) and *hphA* (*NdeI/XhoI*) were integrated downstream of the two T7 promoters of pETDuet-1 using standard restriction/ligation based cloning techniques, generating pETDuet-1::*apt-hph*. The construct was PCR linearized and *aptF* was added downstream of *hphA* by HiFi DNA assembly (NEB). To ensure stable coexistence of both expression plasmids, the *ColE1* origin of replication of pETDuet-1 was exchanged to the *ColA* originating from pCOLADuet-1, in this way generating pETDuet-2::*apt-hph-aptF*. The assembled constructs were transformed into chemically-competent *E. coli* EPI300-T1 (pET-28b-ptetO construct) or DH5 α (all pETDuet based plasmids) cells. Positive clones were selected by screening PCR using primer pairs binding the vector backbone and the insert and further verified using analytical restriction digests. All constructs were checked for integrity as well as for mutations in the entire insert sequences by sanger sequencing.

2.2.2.3. Heterologous expression of *apt*. Electrocompetent *E. coli* BAP1 were co-transformed with pET-28b-ptetO::*apt-NRPS* and pETDuet-

2::*apt-hph* or pETDuet-2::*apt-hph-aptF* and selected on LB agar plates containing Amp and Kan. Single clones were used to grow overnight precultures in a 37 °C shaker in LB medium containing Amp and Kan. 100 mL expression cultures were inoculated 1:50 into 100 mL production medium (20 g/L glucose, 2 g/L MgSO₄, 16 g/L KH₂PO₄, 14 g/L K₂HPO₄, 2 g/L NH₄SO₄, 1 g/L citric acid, 1 g/L L-tyrosine, 5 g/L casamino acids, 50 mg/L FeSO₄·7H₂O, 10 mg/L thiamine-HCl, 10 mg/L MnSO₄·5H₂O, [pH 7.2]) adapted from Koketsu et al. (2013) and supplemented with Amp and Kan. Cells were grown to an OD₆₀₀ of 0.5 at 37 °C while shaking (200 rpm). After incubation on ice for 15 min, the expression of pETDuet-2::*apt-hph* or pETDuet-2::*apt-hph-aptF* was induced by adding IPTG to a final concentration of 1 mM. The cultures were transferred to 25 °C while shaking (200 rpm) and incubated until an OD₆₀₀ of 0.8 was reached. Expression of pET-28b-ptetO::*apt-NRPS* was induced by adding 0.5 µg/mL tetracycline. Due to the light instability of tetracycline, the flasks were wrapped in aluminum foil and the incubation was continued for 48 h. After harvesting the cells by centrifugation (6000 × g, 10 min, 4 °C), the supernatant was stored at 4 °C and the cell pellet at –80 °C upon extraction. Supernatants were adjusted to pH 4–5 with 37% HCl for extraction and centrifuged (10,000 × g, 10 min, room temperature) to remove occurring precipitate. After extraction with 0.75 vol EtOAc for three times, the organic phases were combined and evaporated in vacuo at 40 °C. The residues were resuspended in 500 µL MeOH and filtered using a PTFE 0.2 µm syringe filter. Pellets were extracted using 20 mL MeOH. After resuspension, the cells were disrupted in a Branson ultrasonic cleaner 3510EMTH (Branson, USA) for 30 min and the insoluble compounds as well as the cell debris were removed by centrifugation (10,000 × g, 10 min, room temperature). The solvent was evaporated in vacuo at 40 °C and the samples were prepared as the supernatant samples above. A standard of anabaenopeptin NZ857 and nostamide A was prepared by extracting *N. punctiforme* cells after storage at –80 °C as described for *E. coli* cells. As a negative control, empty *E. coli* BAP1 cells were grown, induced and extracted as described above, but without antibiotics.

2.2.3. Erythromycin

2.2.3.1. *Bacterial culturing and gDNA isolation.* *S. erythraea* DSM 40517 was obtained from the Leibniz Institute DSMZ, the German collection of microorganisms and cell cultures GmbH. *S. erythraea* was cultivated in GYM liquid medium (0.4% glucose, 0.4% yeast extract, 1.0% malt extract, pH 7.2) at 28 °C while moderate shaking (200 rpm) for seven days. The gDNA was isolated using the Bacterial DNA Preparation Kit (Jena Bioscience) according to the manufacturer's protocol. After hydration, the gDNA was diluted with 5 mM TE buffer (pH 8.5) to a stock concentration of 100 ng/µL and stored at –20 °C until use.

2.2.3.2. *DiPaC strategy for ery.* Q5 polymerase with HighGC Enhancer was used to amplify the *ery* cluster as five sections. The vector backbone of pET-28b-SUMO was initially used as the cloning vehicle and fragments 1–4 were assembled in a sequential manner. At first, pET-28b-SUMO was linearized using *Bam*HI and dephosphorylated. With each insertion of parts *ery*-p1, 2, 3 and 4, a unique *Sna*BI restriction site on the 3'-end was added for backbone linearization (Table S6). This site was included in the reverse primer and thus got attached to the *ery* fragments during the amplification process. *Sna*BI was removed during assembly so that the wild type sequence between the cluster fragments was maintained and the restriction site itself on the 3'-end was kept unique. All the pET-28b-SUMO constructs (Table S3) were screened using suitable screening primer pairs binding at the end of the insert in combination with the backbone primer pET-RP (Table S1). Due to the large size of pET-28b-SUMO::*ery*-p1234, the vector backbone was exchanged to a pCC1FOS based plasmid by LCHR using *E. coli* GB08-red to construct pCC1FOS::*ery*-p1234 (see below, Section 2.2.3.3). For successful LCHR the helper plasmid pCC1FOS::HR_ery-p1234 was first constructed. pCC1FOS::HR_ery-p1234 was built to contain one 500 bp

capturing arm homologous to the beginning of *ery*-p1 and a second 479 bp capturing arm homologous to the end-sequence of *ery*-p4, and was generated in a three fragment HiFi DNA assembly reaction out of *Hind*III-linearized pCC1FOS and the PCR amplified capturing fragments. This construct was *Hind*III-linearized, dephosphorylated and then used for the LCHR assembly. To complete the pCC1FOS::*ery* vector, *ery*-p5 was inserted in a last assembly reaction into pCC1FOS::*ery*-p1234 to generate pCC1FOS::*ery*. *E. coli* EPI300-T1 colonies were screened using vector and insert specific primers. The expression construct for *Streptomyces*, pLK01::*ery*, was constructed by LCHR using pCC1FOS::*ery* as circular template and a compatible linear fragment of pLK01 (6617 bp) generated by PCR using primers pLK01_Lin_For/Rev (use of HighGC Enhancer mandatory) (Kaysser et al., 2012).

2.2.3.3. *Homologous recombination to conduct vector backbone exchanges.* Backbone exchanges were performed by homologous recombination in *E. coli* GB08-red (LCHR) (Fu et al., 2012) with introduction of the desired template plasmids by electroporation. Cells were PCR-verified to ensure successful plasmid uptake and were then used to prepare LB overnight cultures, grown under antibiotic pressure at 37 °C while shaking (200 rpm). 15 mL of LB medium were inoculated in a ratio of 1:100 and cultures were grown to an OD₆₀₀ value of 0.2–0.3. L-arabinose was added to a final concentration of 0.1% to induce the recombinase system. Cultivation was continued until the OD₆₀₀ was between 0.6 and 0.8. Electrocompetent cells were generated by washing with sterile, ice-cold 10% glycerol (three times) with intermediate centrifugation at 5000 × g for 5 min at 4 °C. Cell pellets were resuspended in 300 µL of 10% (v/v) glycerol solution to prepare three aliquots with a volume of 100 µL each and were stored on ice. Cells were then electroporated with the linearized capturing backbone (100–200 ng of DNA). The electroporation was carried out in a 0.2 cm ice-cold electroporation cuvette using a BioRad GenePulser II set to: 200 Ω, 25 µF and 2.5 kV. Directly after pulsing, 700 µL of SOC medium were added to the cells, which subsequently were incubated at 37 °C while shaking (200 rpm) for one hour. The bacterial suspension was streaked on LB agar plates with the corresponding antibiotic of the target backbone and cultivated overnight at 37 °C. Colonies of *E. coli* GB08-red were checked by PCR using suitable primer pairs. The miniprep eluates from *E. coli* GB08-red overnight cultures were taken to transform plasmid storage strains (*E. coli* DH5α, TOP10, EPI300-T1). For LB overnight cultures of *E. coli* EPI300-T1 containing pCC1FOS/pLK01, L-arabinose was added to a final concentration of 0.01% to induce high-copy replication of plasmids. After isolation of the target plasmids, screening PCR and analytical restriction digests were used to validate the sequence of the exchanged plasmids. End-sequencing of vector/insert border regions was used in addition to prove correct integration.

2.2.3.4. *Streptomyces conjugation.* Plasmid pLK01::*ery* was introduced into *S. coelicolor* M1152 and M1154 by triparental intergeneric conjugation with *E. coli* ET12567/pUB307 (Flett et al., 1997). In short, pLK01::*ery* was applied to transform electrocompetent *E. coli* ET12567. PCR-validated clones were selected and used to inoculate LB supplemented with Kan. Similarly, LB cultures of *E. coli* ET12567/pUB307 were grown. Overnight cultures of *E. coli* ET12567 pLK01::*ery* and *E. coli* ET12567/pUB307 were used to prepare LB cultures which were cultivated at 37 °C while shaking (200 rpm). When reaching an OD₆₀₀ value of about 0.5–0.6, bacteria were pelleted by centrifugation (5000 × g, 5 min, 4 °C). The supernatants were discarded and cells washed twice under sterile conditions with LB medium. Pellets were resuspended in 500 µL LB medium and dense cellular suspensions were obtained. Spores of *S. coelicolor* M1152/M1154 were mixed with 50 mM TES buffer pH 8.0 and incubated at 50 °C for ten minutes. Cells were then incubated further at 28 °C until use. For conjugation, *E. coli* ET12567 pLK01::*ery*, *E. coli* ET12567/pUB307 and activated spores

were mixed. After pelleting by centrifugation ($4000 \times g$, 2 min, room temperature), the supernatants were carefully removed and the cells resuspended in the remaining medium. The cellular suspension was streaked onto MS agar plates (2% mannitol, 2% soya flour, 10 mM $MgCl_2$, 60 mM $CaCl_2$, 2% agar) (Wang and Jin, 2014) and cultivated at $30^\circ C$. After 16–20 h, agar plates were overlaid with an aqueous solution containing 1 mg NA and 1 mg Kan. Exconjugates were obtained after four to seven days of cultivation. A secondary selection was performed streaking single colonies on MS agar plates containing selection antibiotics (Kan, NA). Kan-resistant clones were selected, PCR verified and used to inoculate CASO medium (Carl Roth), supplemented with the appropriate antibiotics.

2.2.3.5. Heterologous expression of ery. The expression hosts *S. coelicolor* strains M1152 and M1154 were chosen. $pLK01::ery$ was transferred by conjugation and exconjugates were taken after two passages of antibiotic selection and screened for the presence of a 485 bp fragment of *ery* (Fig. S22). Validated clones were streaked on MS agar plates and cultivated for twelve days at $30^\circ C$. The agar was cut into small pieces and incubated with a solvent mixture consisting of 4:1 EtOAc:MeOH. After a ten min incubation in the sonicating bath, the mixture was filtrated and evaporated in vacuo at $40^\circ C$. The residues were resuspended in 500 μL of MeOH, insoluble materials were removed by centrifugation ($17,000 \times g$, 2 min, $4^\circ C$) using a benchtop centrifuge and filtered (PTFE; 0.2 μm particle filter). The samples were then subjected to HPLC-MS analysis. Commercially available erythromycin (Sigma-Aldrich, catalogue no.: E5389–1 G) was used to validate the retention time.

3. Results and discussion

In the course of our work directed at the cloning, recombinant production and biocatalytic application of biosynthetic enzymes for natural product diversification and total synthesis, e.g., in the field of polycyclic tetramic acid containing macrolactams (PoTeMs) (Antosch et al., 2014; Greunke et al., 2015, 2017), we realized that the PCR amplification of large genes (> 5 kb) utilizing Q5 polymerase (NEB) was highly efficient and reliably error free. Based on this observation, we devised a straightforward strategy for the interception of BGCs by combining long-template PCR amplification with HiFi DNA assembly to efficiently build BGC heterologous expression constructs (Fig. 1). If sufficiently large PCR products with homology arms to the expression vector of choice could be generated reliably error free, this strategy would not only be technically simple, but also provide several other favorable features, such as: (i) it will allow direct amplification of any BGC of interest from even small amounts of DNA, (ii) the vector backbone for heterologous expression can be freely chosen, (iii) the complete assembly of the pathway will occur *in vitro*, allowing late choice of a suitable heterologous host in case of product toxicity, (iv) the BGC can be refactored during the initial cloning process, providing flexibility concerning cloning strategy and expression construct structure, (v) besides the vector of choice, only commercially available, ready-to-use kits and reagents are required. Taken together, such a strategy would nicely complement the powerful LLHR and TAR cloning methods. Particularly for small- to mid-sized BGCs, DiPaC would thus have the potential to significantly streamline DNA capturing.

3.1. DiPaC for the discovery of novel natural products by activation of small orphan BGCs

To first demonstrate the application of DiPaC on small BGCs from medium range GC content genomes, we selected the putative phenazine biosynthetic gene cluster (*ehp*) from *S. fonticola* DSM 4576 (9.5 kb, 53.6% GC). The phenazines are a diverse class of small natural products with a wide range of bioactivities, making the identification of novel analogues highly significant (Guttenberger et al., 2017). Extracts of *S.*

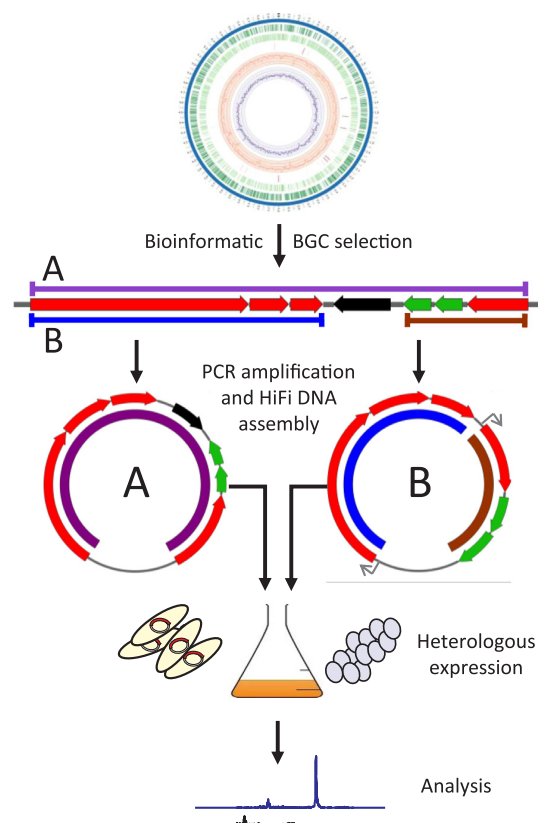


Fig. 1. DiPaC cloning strategy for the identification of novel natural products. Genome mining is used to select a BGC of interest. A cloning strategy is then devised, including selection of an appropriate vector and a suitable expression host. Additionally, based on the source organism and bioinformatic analysis of the BGC of interest, the researcher can choose to amplify the cluster in its native form (A) or to refactor the cluster by altering genetic organization, exchanging promoters, or by removing intergenic regions or predicted regulatory genes (black) via the selected incorporation of multiple PCR fragments (B). If the cluster must be constructed over multiple rounds due to the incorporation of multiple genetic fragments, a unique restriction site can be incorporated into the primer design to enable downstream linearization of the vector backbone prior to the next insertion. Once the vector has been constructed by either (A) or (B), it is transferred into a suitable expression host for compound production.

fonticola DSM 4576 did not appear to contain any phenazine-like molecules indicating the corresponding BGC to be silent under the culture conditions used. Bioinformatic analysis of the *S. fonticola* DSM 4576 *ehp* gene cluster revealed five core phenazine genes, four phenazine auxiliary genes and two species-specific genes (Fig. 2A; Table S4). Additionally, a sequence alignment of the *S. fonticola* DSM 4576 EhpF with the *Pantoea agglomerans* EhpF crystal structure (Bera et al., 2010) revealed a unique binding pocket (Fig. S1). Thus, it was predicted that the silent *S. fonticola* DSM 4576 *ehp* cluster is likely to encode a novel phenazine derivative.

The amplification of both the *ehp* cluster from *S. fonticola* DSM 4576 gDNA and the pET-28b-PD1 backbone from the pET28b-SUMO template yielded strong amplification products. The primers were designed to target a 9.7 kb region harboring *ehpR* (resistance), all five core phenazine biosynthetic genes, three putative phenazine tailoring genes and two species-specific genes, all encoded on a single operon (Fig. 2A). For cloning, 30 bp homology arms were added to the 5'-end of the insert specific primer set. By including vector homology arms onto the insert specific primers, the vector backbone PCR product can be utilized for

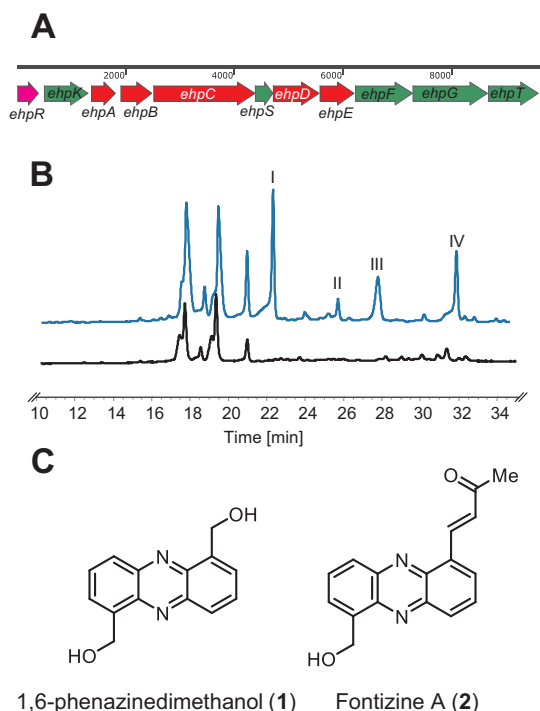


Fig. 2. Overview of the *ehp* gene cluster, heterologous expression of pET-28b-PD1::*ehp* and structure of identified phenazines. (A) The *ehp* gene cluster was identified within the *S. fonticola* DSM 4576 genome via genome mining. The cluster contains eleven ORFs including one resistance gene (pink), five core phenazine genes (red) and five auxiliary genes (green) of which *ehpS* and *ehpT* appear to be strain specific. The *ehp* cluster was cloned as a single linear fragment under the control of a T7 promoter. (B) *E. coli* BAP1 expression cultures were extracted with organic solvent and analyzed by HPLC with a detection wavelength of 360 nm. Cultures expressing pET-28b-PD1::*ehp* (blue) were found to contain four new peaks relative to *E. coli* BAP1 expressing empty pET-28b-SUMO vector (black). (C) Peak I and Peak IV were isolated and structurally elucidated by 2D NMR. Compound 1 was identified as 1,6-phenazinedimethanol while 2 was identified as a novel structure designated as fontizine A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the cloning of multiple distinct insert fragments of other BGCs. The terminal regions of each DNA fragment were designed so the insert would be located in an ideal position for successful expression utilizing the pET-28-PD1 T7 promoter (Fig. S2). Several positive clones were chosen following transformation of the HiFi DNA assembly reaction, confirmed by restriction digest and initial sequencing of the insert terminal regions followed by full sequencing of the complete BGC insert. As desired, the complete phenazine BGC was indeed successfully (and error free) transferred into the heterologous expression construct.

Extracts of *E. coli* BAP1 expression cultures harboring pET-28b-PD1::*ehp* were found to produce several unique peaks compared to empty pET-28b-SUMO expression cultures when analyzed by HPLC with a detection wavelength of 360 nm. This wavelength was used based on known UV spectra of the phenazine core ring system (Saleh et al., 2012; Kadam et al., 2013). A total of four additional peaks were identified in pET-28b-PD1::*ehp* expression cultures compared to the control. Two molecules were identified as major products with a retention time of 22 (I) and 32 min (IV) (Fig. 2B). Two minor constituents with a phenazine-type UV profile were also observed at 25.5 and 27.5 min, but could not be isolated in sufficient amounts for NMR structure determination.

The two major metabolites were isolated by semi-preparative HPLC for downstream high-resolution mass spectrometry and NMR analysis.

Peak I was found to have a HR-ESI-MS of m/z of 241.0971 $[M + H]^+$ corresponding to a chemical formula of $C_{14}H_{12}N_2O_2$. Further 1H NMR analysis revealed the compound to be 1,6-phenazinedimethanol (1) (Fig. 2C; Fig. S4; Table S5), a symmetrical phenazine previously isolated from the marine bacterium *Brevibacterium* sp. KMD 003, but for which no BGC has previously been reported (Choi et al., 2009). A bioinformatic search of publicly available *Brevibacterium* genomes could not identify an identical *ehp* gene cluster to that of *S. fonticola*. ESI-HR-MS analysis of Peak IV (2) showed a m/z of 279.1127 $[M + H]^+$ with a calculated molecular formula of $C_{17}H_{14}N_2O_2$. Mass analysis and 1H NMR indicated the presence of a new phenazine analog now referred to as fontizine A (2). Full structure elucidation by 2D NMR spectroscopy (COSY, HSQC, HMBC, see Figs. S5–S8 and Table S5) firmly established the molecular framework of 2 for this new phenazine. Biosynthetically, this compound might derive from an initial desymmetrization of 1 by selective oxidation of one of the hydroxymethylene side-chains to the corresponding aldehyde. Nucleophilic attack at the aldehyde function by an acetone enolate, e.g., generated by decarboxylation of acetoacetic acid, followed by elimination of water from the intermediate aldol product furnishes fontizine A (2). The final yields of the purified natural products isolated from the *E. coli* expression cultures were approx. 0.6 mg/L and 0.4 mg/L for 1 and 2, respectively. Having established DiPaC for a BGC of approx. 10 kb, we next turned our attention to cloning a larger BGC, concurrently refactoring parts of the BGC directly during the cloning procedure.

3.2. DiPaC for the interception, refactoring and heterologous expression of mid-sized BGCs

To validate the DiPaC strategy for BGCs beyond a size of 10 kb, we selected the anabaenopeptin (*apt*) BGC from the cyanobacterium *Nostoc punctiforme* PCC 73102 (29.2 kb, 42% GC) as a suitable model system (Rouhiainen et al., 2010). The anabaenopeptins are hexapeptide products of a non-ribosomal peptide synthetase (NRPS) composed of six modules on four individual genes. All anabaenopeptins have an unusual terminal ureido bond in common. Slight variations in the amino acid composition of the peptide backbone lead to the different members of this natural product family, with anabaenopeptin NZ857 (3) and nos-tamide A (4) being the major products in *N. punctiforme* (Fig. 3B). Most importantly, these compounds exhibit strong protease inhibitory activity. Consequently, they are highly interesting from a biomedical perspective. However, cultivation of cyanobacteria in the laboratory for compound production can be challenging and the genetic manipulation of filamentous cyanobacteria is very difficult. Additionally, there is often poor expression of native cyanobacterial promoters within *E. coli* heterologous hosts. Thus, we utilized the DiPaC strategy to simultaneously clone and refactor the *apt* BGC for the heterologous production of anabaenopeptins in *E. coli*.

All core *apt* biosynthetic NRPS genes *aptA–D* are situated on a single operon, while the amino acid precursor biosynthetic genes *hphB/hphCD* and *hphA*, responsible for the assembly of L-homotyrosine and L-homophenylalanine found in 3 and 4, are located up- and downstream, respectively, of the large NRPS (Fig. 3A). The 22 kb NRPS consisting of *aptA–D* was PCR amplified as a single product in high yields (Fig. S10), simultaneously adding homology overhangs of 20 bp at the *N*- and 29 bp at the *C*-terminus. Subsequent integration into the pET-28b-ptetO vector possessing a tetracycline inducible *ptetO* promoter was performed by HiFi DNA assembly. After successful transformation into *E. coli*, a positive clone was submitted for sequencing and the complete 22 kb NRPS as well as the integration sites were proven to be free of mutations (Fig. S12). To facilitate efficient activation of the homoamino acid biosynthesis genes independent and ahead of the NRPS assembly line, the corresponding *hph* genes were integrated into the pETDuet-1 vector containing two IPTG inducible T7 promoters. The genes *hphB/hphCD* were PCR amplified and placed under control of the first promoter while *hphA* and *aptF* were inserted behind the second T7

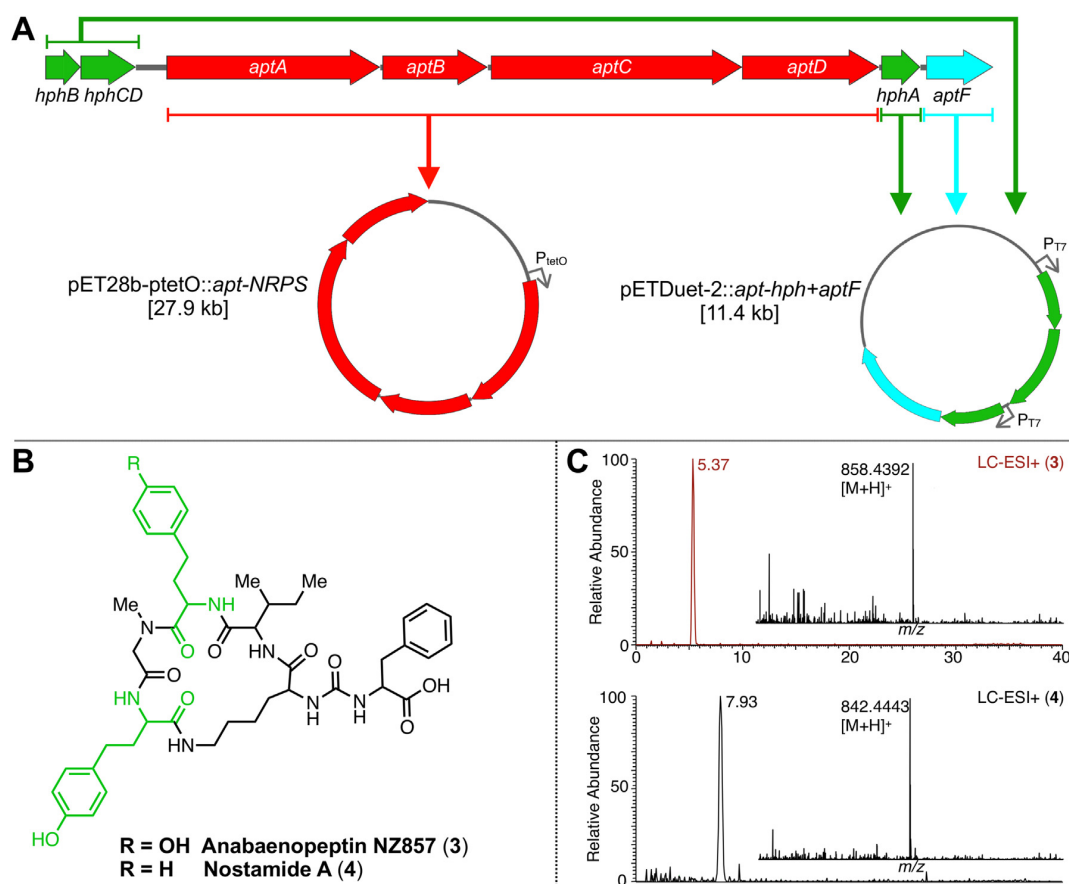


Fig. 3. DiPaC of the *apt* gene cluster of *N. punctiforme*. (A) DiPaC construct design, (B) structures of anabaenopeptin NZ857 (**3**) and nostamide A (**4**). The incorporated homoamino acids L-Hph/L-Hty, highlighted in green, require the biosynthetic genes *hphA*, *hphB-D* (green). (C) HPLC-HR-MS analysis of *E. coli* BAP1 pET-28b-ptetO::*apt-NRPS* pETDuet-2::*apt-hph-aptF* extracts containing **3** and **4** (EICs of **3** and **4** are shown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

promotor. To ensure stable co-expression of both plasmids, the ColE1 origin of replication of pETDuet-1 was replaced by ColA (Fig. S11).

Expression in L-tyrosine rich production medium adapted from Koketsu et al. (2013) was carried out using *E. coli* BAP1 pET28b-ptetO::*apt-NRPS* + pETDuet-2::*apt-hph-aptF*. Transcription of the *hph* genes and *aptF* was induced with IPTG prior to inducing the *apt* NRPS genes with tetracycline. HR-HPLC-MS [ESI⁺] measurements of the pellet and supernatant extracts were compared to *N. punctiforme* extracts and clearly showed the production of anabaenopeptin NZ857 (**3**) (m/z of 858.439 [M + H]⁺) and nostamide A (**4**) (m/z of 842.444 [M + H]⁺) both in the heterologous host as well as in the natural producer (Fig. 3C and Fig. S13). Further HPLC-MS² [HESI⁺] spectra measured for both compounds match with data published by Rouhiainen et al. (2010) (Fig. S14). Based on direct comparison of the heterologous system and the wild-type production strain by HPLC-UV/MS analyses, the heterologous system reached > 100-fold higher production levels (per volume). In addition, the cultivation time was dramatically decreased from 50 days in case of the cyanobacterial producer to 48 h in *E. coli*. Interestingly, the role of the putative ABC transporter *aptF* has not yet been experimentally investigated. However, when *E. coli* BAP1 pET28b-ptetO::*apt-NRPS* + pETDuet-2::*apt-hph* expression cultures lacking *aptF* were used for heterologous expression experiments, products **3** and **4** were no longer detectable, thus emphasizing the importance of this transporter for compound production. This observation is in accordance with previous results on the ABC transporter *mycH* from microcystin biosynthesis (Pearson et al., 2004).

Using DiPaC to capture and refactor a cyanobacterial BGC, we were thus able to successfully express anabaenopeptins in *E. coli*. To our knowledge, this is the first heterologous production of anabaenopeptins.

3.3. DiPaC for the mobilization and heterologous expression of large BGCs

As evident from the phenazine and anabaenopeptin examples above, DiPaC can readily be used for the cloning of small- to mid-sized BGCs of up to approximately 25 kb in size using a single long-amplicon PCR setup. However, the yield of the desired PCR product drops significantly for larger amplicons. This limitation of DiPaC can be eliminated by subdividing larger BGCs into several PCR fragments with their subsequent fusion by HiFi DNA assembly. As a proof of concept for this approach, we selected the erythromycin BGC *ery* (54.6 kb, 72.5% GC) from *S. erythraea* DSM 40517 for DiPaC (Oliynyk et al., 2007). This pathway constitutes the text book example of a multimodular type I polyketide synthase (PKS) assembly line encoding the macrolactone 6-deoxyerythronolide B (6-dEB), which is further elaborated into erythromycin A (**5**) by a series of tailoring reactions, in particular oxygenations, glycosylations and methylation (Fig. S25) catalyzed by enzymes likewise encoded within *ery* (Staunton and Wilkinson, 1997; Staunton and Weissman, 2001).

The *ery* BGC was arbitrarily dissected into five fragments between 9.5 and 12.6 kb in size. Because of the significantly higher success rates of two-fragment HiFi DNA assembly over multiple fragment

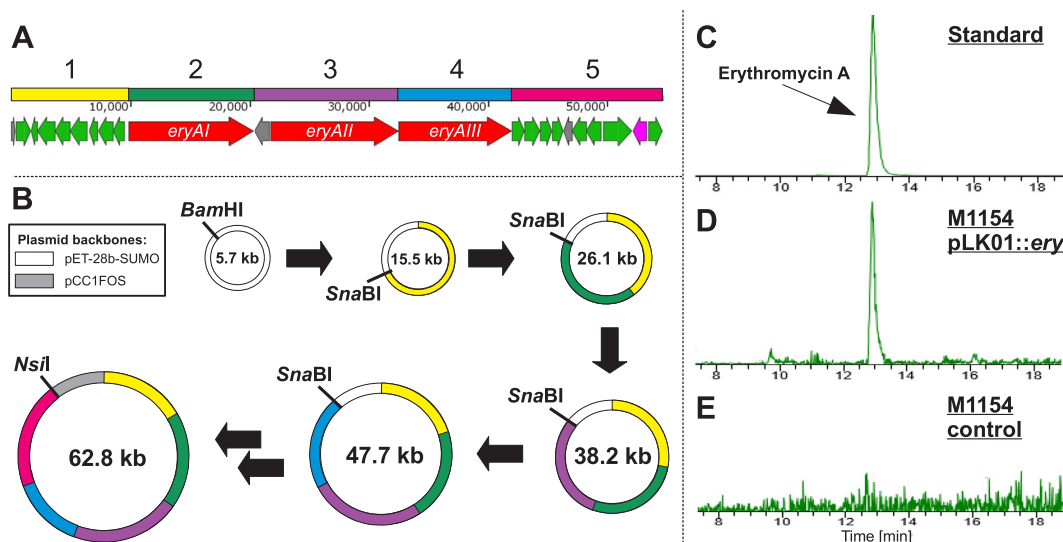


Fig. 4. Cloning and heterologous expression of the erythromycin BGC. (A) Overview of the *ery* gene cluster (54.6 kb) from *S. erythraea* NRRL 2338 and *in silico* fragment partitioning into five pieces of 9.5–12.6 kb (yellow, green, purple, light blue, magenta) which were PCR-amplified. (B) Schematic representation of the DiPaC reconstitution strategy involving unique restriction sites for plasmid linearizations required for fragment insertions. (C) HPLC-MS analysis of an erythromycin A (5) standard ($m/z = 734.46 [M + H]^+$), (D) in comparison with 5 from heterologous expression in *S. coelicolor* M1154 pLK01::ery and an extract derived from (E) a control culture without the expression plasmid (EICs for 5 is shown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

approaches, a sequential DiPaC strategy utilizing the pET-28b-SUMO backbone was devised. This included the introduction of a unique restriction site between the vector homology arms and the BGC cluster sequence during PCR amplification, which allowed linearization of the cloning vector after each round of HiFi DNA assembly to insert further PCR amplified fragments (Fig. 4A and B, Figs. S15–S16). The artificial cutting site is removed during each round of HiFi DNA assembly, leading to seamless cloning of the BGC of interest. The five cluster fragments of *ery* were amplified in high yield using Q5 polymerase (NEB) (Fig. S17) and sequentially cloned into pET-28b-SUMO. To ensure plasmid stability and to test for efficiency of backbone exchange with large DNA fragments using LCHR, the vector backbone was exchanged to pCC1FOS after integration of the fourth fragment. The pCC1FOS derivative pLK01 (Kayser et al., 2012), which possesses an *oriT-traJ* element for conjugation as well as a neomycin/kanamycin resistance cassette for selection, was ultimately chosen to facilitate transfer of the complete *ery* pathway into selected heterologous *S. coelicolor* hosts (Fig. S26). The integrity of all the intermediate and the final expression constructs was confirmed by restriction analyses and end-sequencing (Figs. S18–S21).

Triparental conjugation, facilitated by *E. coli* ET12567 pUB307 together with *E. coli* ET12567 pLK01::ery, was used for stable integration of the *ery* BGC into the genomes of *S. coelicolor* M1152 and M1154 (Flett et al., 1997). Exconjugates were PCR-analyzed for the presence of the *ery* BGC (Fig. S22). Two clones were selected for each strain and cultures grown on mannitol soy agar. The engineered *Streptomyces* produced 5 (m/z of 734.46 $[M + H]^+$), which was readily detected in the organic solvent extracts by HPLC-MS in comparison to a commercially available standard (Fig. 4C, D, E, Fig. S23). In addition, MS²-fragmentation spectra were measured and clearly confirmed the presence of 5, in our case by monitoring the specific ion fragment fingerprint (Fig. S24). While production levels in the recombinant host without optimization of the cultivation parameters were low, detection and unambiguous identification of 5 by HPLC-MS and MS² clearly validated the successful cloning of the native *ery* pathway using DiPaC.

It is important to note that previous studies have successfully produced erythromycin A (5) (Zhang et al., 2010) and its precursors 6-dEB

and EB in heterologous host systems, however, with significantly more complex construct designs and cloning approaches (Fayed et al., 2015; Kumpfmüller et al., 2016; Pfeifer et al., 2001). For the production of 5 in *E. coli*, multiple plasmids, each harboring complementary parts of *ery*, were used. In the most improved situation, this required four individual plasmids covering 20 *ery* biosynthetic genes (Zhang et al., 2010; Jiang et al., 2013). 6-dEB was also produced in *Streptomyces* that were conjugated with at least three plasmids covering the polyketide synthase genes *eryAI/II/III*. Introduction of an additional vector harboring an angolomycin glycosylation gene cassette facilitated production of angolosaminylated 6-dEB and EB (Fayed et al., 2015). Heterologous expression of 6-dEB in *Bacillus subtilis* required several rounds of homologous recombination, three to integrate *eryAI/II/III*, as well as the transfer of an intact copy of *sfp* (Kumpfmüller et al., 2016). In our current work, DiPaC allowed us to quickly access the entire *ery* cluster by a sequential assembly over a series of intermediate plasmids. The integrative vector pLK01::ery was used for the stable integration of *ery* in a single conjugation step, thus being able to successfully produce 5. This study clearly showcases that even large BGCs can readily be targeted by DiPaC. In principle, such a multiple fragment approach should be suitable even for significantly larger BGCs. Using Q5 polymerase on high-GC genomic template DNA, we were able to readily amplify up to 15 kb large DNA fragments, which in case of low-GC template can be increased up to 25 kb.

4. Conclusions

Genome mining for the discovery of novel small molecules from Nature has emerged as a flourishing field in modern biomolecular natural product research (Bachmann et al., 2014; Ziemert et al., 2016). Besides the heroic development of easy and free to use bioinformatic tools, such as antiSMASH (Blin et al., 2017) and PRISM (Skinnider et al., 2015), novel powerful techniques for BGC interception, particularly LCHR, LLHR and TAR (Ongley et al., 2013; Muyrers et al., 1999; Sharan et al., 2009; Yu et al., 2000; Zhang et al., 1998, 2000; Wenzel et al., 2005; Fu et al., 2012; Duchaud et al., 2003; Yin et al., 2015; Kouprina and Larionov, 2006, 2008; Kim et al., 2010; Yamanaka et al.,

2014; Bian et al., 2012; Cobb and Zhao, 2012), are driving forces for the fast progress of the field. These methods facilitate the mobilization, isolation and heterologous expression of BGCs with high efficiency. In addition to their general need for relatively large amounts of gDNA, these techniques have some specific individual requirements, most importantly the use of highly specialized method-specific vectors and hosts. Downstream full assembly of the desired expression construct by *in vivo* homologous recombination followed by its intact isolation can also be challenging for the non-expert, as often the case in natural product discovery laboratories. We thus aimed at the development of a complementary additional tool, DiPaC, which is easy to use and does not require any specialized equipment, vector, or organism.

DiPaC relies on the amplification of the BGC of interest by PCR using state-of-the-art high-fidelity DNA polymerases. While such an approach has generally been neglected because of the expected incorporation of multiple errors by PCR reactions targeting long stretches of DNA, modern polymerases, such as the Q5 polymerase used in this study, seem to perform with perfect accuracy. Indeed, we have not detected a single mutation in any of the up to 22 kb large constructs fully sequenced in the course of this study. Given the risk of undesired recombination events during *in vivo* homologous recombination, e.g., erroneous deletion of parts of a BGC with high sequence homology as found in many PKS and NRPS systems, our PCR-based strategy thus seems to be equally reliable.

Within this work, we demonstrated the power of DiPaC for the cloning and heterologous expression of small- to mid-sized BGCs of up to ca. 25 kb in size, with its possible extension to larger gene clusters by HiFi DNA assembly of multiple large PCR amplicons. As exemplarily shown for the *ehp* BGC described above, DiPaC is extremely efficient for the cloning and heterologous expression of small- to mid-sized BGCs in their native form. In fact, the BGC of interest can be amplified by PCR, assembled into any vector of choice, transferred into and isolated from *E. coli* within 24 h, making the design, synthesis and delivery of the cloning PCR primers the most time-consuming step of the procedure.

As showcased for the anabaenopeptin BGC, DiPaC is also compatible to pathway refactoring directly during the cloning procedure. The common failure of cyanobacterial promoters to be activated in *E. coli* expression hosts and the difficulty to genetically manipulate cyanobacteria has significantly hampered genetically mining these organisms, with only few impressive examples described to date (D'Agostino et al., 2016; Kim et al., 2012; Liu et al., 2017; Ongley et al., 2013). DiPaC allows for a straightforward refactoring of BGCs directly during cloning to overcome these problems, e.g., via the exchange of native (silent) promoters for inducible ones or by removing predicted transcriptional terminators. Traditionally, this would have to occur after capturing the BGC of interest from the gDNA, such as with the expression of the microcystin biosynthetic gene cluster (Liu et al., 2017). Furthermore, the ability to circumvent the addition of an antibiotic selection marker ensures the expression vector can then be altered with ease after construction by already established recombineering techniques such as LCHR.

From a technical perspective, DiPaC solely requires commercially available reagents (High-Fidelity DNA polymerase, HiFi DNA assembly kit) and standard molecular biology equipment. The assembly of the expression construct is completely conducted *in vitro*, thus facilitating its direct transfer into any genetically transformable heterologous host. This can be particularly advantageous for BGCs encoding natural products toxic to the classical *in vivo* homologous recombination organisms. DiPaC is furthermore not restricted to any specialized cloning vector. Instead, any expression vector applicable to a heterologous host of interest (*E. coli*, *Streptomyces*, *Bacillus*, yeast, cyanobacteria) can be directly utilized. The vector of choice can either be linearized by restriction digest and dephosphorylated, or the desired part of the backbone can be amplified by long-range PCR. The homology arms required for HiFi DNA assembly can be attached to the BGC to be inserted during PCR amplification, thereby eliminating tedious procedures for the

installation of homology arms onto the cloning vector template. This also enables the use of any produced, storable linear cloning vector for multiple BGCs of interest. Further, we envisage the future of DiPaC to be applied to more complex samples, such as capturing BGCs directly from metagenomic samples without the need for large-insert genomic library preparation. This would be a particularly well-suited method for metagenomes where the amount of gDNA is limited. With the increased research interest in small- to mid-sized BGCs, including small (iterative) PKS/NRPS systems, e.g., encoding polycyclic tetramic acid containing macrolactams (PoTeMs) (Greunke et al., 2017; Blodgett et al., 2010) or bacterial pyrrolizidine alkaloids (Schimming et al., 2015), ribosomally synthesized and post-translationally modified peptides (RiPPs) (Viehrig et al., 2017; Mo et al., 2017; Chen et al., 2017; Agrawal et al., 2017) or terpenoid pathways (Valiante et al., 2017; Khalid et al., 2017; Janevska et al., 2016), we expect DiPaC to evolve as an invaluable new tool not only for the natural product research community, but also for metabolic (bio)engineering and synthetic biology in general.

Acknowledgments

We thank Dr. Philipp Baer and Prof. Dr. Michael Groll (Chair of Biochemistry, TUM) for providing pET-28b-SUMO, Prof. Brett Neilan and Prof. Dr. Rolf Müller for the pET-28b-ptetO plasmid, Catharina Seel and Prof. Dr. Tanja Gulder (Biomimetic Catalysis, TUM) for HR-MS analysis of the phenazines.

Funding: C.G. and E.R.D. thank the Friedrich Naumann Foundation for Freedom and the Deutsche Bundesstiftung Umwelt (DBU), respectively, for their Ph.D. scholarships. E.R.D. thanks CIPSM-Women for funding. P.M.D. thanks the TUM University Foundation Fellowship for funding. We thank the DFG for generous financial support of the work in our laboratory (Emmy Noether program (GU 1233/1-1) and Center for Integrated Protein Science Munich CIPSM).

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.ymben.2018.03.010>.

References

- Adnani, N., Rajski, S.R., Bugni, T.S., 2017. Symbiosis-inspired approaches to antibiotic discovery. *Nat. Prod. Rep.* 34, 784–814.
- Agrawal, P., Khater, S., Gupta, M., Sain, N., Mohanty, D., 2017. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* 43, W80–W88.
- Antosch, J., Schaefer, F., Gulder, T.A.M., 2014. Heterologous reconstitution of ikarugamycin biosynthesis in *E. coli*. *Angew. Chem. Int. Ed. Engl.* 53, 3011–3014.
- Bachmann, B.O., Van Lanen, S.G., Baltz, R.H., 2014. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* 41, 175–184.
- Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147.
- Bera, A.K., Atanasova, V., Gamage, S., Robinson, H., Parsons, J.F., 2010. Structure of the D-alanylgriseoliteic acid biosynthetic protein EhpF, an atypical member of the ANL superfamily of adenylyating enzymes. *Acta. Crystallogr. D. Biol. Crystallogr.* 66, 664–672.
- Bian, X., Huang, F., Stewart, F.A., Xia, L., Zhang, Y., Müller, R., 2012. Direct cloning, genetic engineering, and heterologous expression of the syringolin biosynthetic gene cluster in *E. coli* through Red/ET recombineering. *ChemBioChem* 13, 1946–1952.
- Blin, K., Kim, H.U., Medema, M.H., Weber, T., 2017a. Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene

- clusters. *Brief. Bioinform* 1–11.
- Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breiting, R., Takano, E., Lee, S.Y., Weber, T., Medema, M.H., 2017b. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45, W36–W41.
- Blodgett, J.A., Oh, D.C., Cao, S., Currie, C.R., Kolter, R., Clardy, J., 2010. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc. Natl. Acad. Sci. USA* 107, 11692–11697.
- Bode, H.B., Bethe, B., Höfs, R., Zecek, A., 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* 3, 619–627.
- Challis, G.L., 2014. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* 41, 219–232.
- Chen, E., Chen, Q., Chen, S., Xu, B., Ju, J., Wang, H., 2017. Mathermycin, a lantibiotic from the marine actinomycete *Marinactinospora thermotolerans* SCSIO 00652. *Appl. Environ. Microbiol.* 83.
- Choi, E.J., Kwon, H.C., Ham, J., Yang, H.O., 2009. 6-Hydroxymethyl-1-phenazine-carboxamide and 1,6-phenazinedimethanol from a marine bacterium, *Brevibacterium* sp. KMD 003, associated with marine purple vase sponge. *J. Antibiot.* 62, 621–624.
- Cobb, R.E., Zhao, H., 2012. Direct cloning of large genomic sequences. *Nat. Biotechnol.* 30, 405–406.
- D'Agostino, P.M., Javalkote, V.S., Mazmouz, R., Pickford, R., Puranik, P.R., Neilan, B.A., 2016a. Comparative profiling and discovery of novel glycosylated mycosporine-like amino acids in two strains of the cyanobacterium *Scytonema cf. crispum*. *Appl. Environ. Microbiol.* 82, 5951–5959.
- D'Agostino, P.M., Song, X., Neilan, B.A., Moffitt, M.C., 2016b. Proteogenomics of a saxitoxin-producing and non-toxic strain of *Anabaena circinalis* (cyanobacteria) in response to extracellular NaCl and phosphate depletion. *Environ. Microbiol.* 18, 461–476.
- Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Givaudan, A., Taourit, S., Bocs, S., Boursaux-Eude, C., Chandler, M., Charles, J.F., Dassa, E., Derosé, R., Derzelle, S., Freyssinet, G., Gaudriault, S., Médigue, C., Lanois, A., Powell, K., Siguier, P., Vincent, R., Wingate, V., Zouine, M., Glaser, P., Boemare, N., Danchin, A., Kunst, F., 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* 21, 1307–1313.
- Fayed, B., Ashford, D.A., Hashem, A.M., Amin, M.A., El Gazayerly, O.N., Gregory, M.A., Smith, M.C., 2015. Multiplexed integrating plasmids for engineering of the erythromycin gene cluster for expression in *Streptomyces* spp. and combinatorial biosynthesis. *Appl. Environ. Microbiol.* 81, 8402–8413.
- Flett, F., Mersinias, V., Smith, C.P., 1997. High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS. Microbiol. Lett.* 155, 223–229.
- Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F., Zhang, Y., 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* 30, 440–446.
- Greunke, C., Antosch, J., Gulder, T.A.M., 2015. Promiscuous hydroxylases for the functionalization of polycyclic tetramate macrolactams – conversion of ikarugamycin to butremycin. *Chem. Commun.* 51, 5334–5336.
- Greunke, C., Glöckle, A., Antosch, J., Gulder, T.A.M., 2017. Biocatalytic Total Synthesis of Ikarugamycin. *Angew. Chem. Int. Ed. Engl.* 56, 4351–4355.
- Guttenberger, N., Blankenfeldt, W., Breinbauer, R., 2017. Recent developments in the isolation, biological function, biosynthesis, and synthesis of phenazine natural products. *Bioorg. Med. Chem.* 25, 6149–6166.
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Omura, S., 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* 21, 526–531.
- Ikeda, H., Kazuo, S.Y., Omura, S., 2014. Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J. Ind. Microbiol. Biotechnol.* 41, 233–250.
- Janevska, S., Arndt, B., Niehaus, E.M., Burkhardt, I., Rösler, S.M., Brock, N.L., Humpf, H.U., Dickschat, J.S., Tudzynski, B., 2016. Glibepyrone biosynthesis in the rice pathogen *Fusarium fujikuroi* is facilitated by a small polyketide synthase gene cluster. *J. Biol. Chem.* 291, 27403–27420.
- Jiang, M., Fang, L., Pfeifer, B.A., 2013. Improved heterologous erythromycin A production through expression plasmid re-design. *Biotechnol. Prog.* 29, 862–869.
- Jiang, W., Zhu, T.F., 2016. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat. Protoc.* 11, 960–975.
- Kadam, M.S., Patil, S.G., Dane, P.R., Pawar, M.K., Chincholkar, S.B., 2013. Methods for purification and characterization of microbial phenazines. In: Chincholkar, S., Thomashow, L. (Eds.), *Microbial Phenazines: Biosynthesis, Agriculture and Health*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 101–140.
- Kaysser, L., Bernhardt, P., Nam, S.J., Loesgen, S., Ruby, J.G., Skewes-Cox, P., Jensen, P.R., Fenical, W., Moore, B.S., 2012. Meroterpenoid A-D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J. Am. Chem. Soc.* 134, 11988–11991.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Khalid, A., Takagi, H., Panthee, S., Muroi, M., Chappell, J., Osada, H., Takahashi, S., 2017. Development of a terpenoid-production platform in *Streptomyces reveromyceticus* SN-593. *ACS Synth. Biol.* 6, 2339–2349.
- Kim, E., Moore, B.S., Yoon, Y.J., 2015. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* 11, 649–659.
- Kim, E.J., Lee, J.H., Choi, H., Pereira, A.R., Ban, Y.H., Yoo, Y.J., Kim, E., Park, J.W., Sherman, D.H., Gerwick, W.H., Yoon, Y.J., 2012. Heterologous production of 4-O-demethylbarbamide, a marine cyanobacterial natural product. *Org. Lett.* 14, 5824–5827.
- Kim, J.H., Feng, Z., Bauer, J.D., Kallifidas, D., Calle, P.Y., Brady, S.F., 2010. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers* 93, 833–844.
- Koketsu, K., Mitsuhashi, S., Tabata, K., 2013. Identification of homophenylalanine biosynthetic genes from the cyanobacterium *Nostoc punctiforme* PCC73102 and application to its microbial production by *Escherichia coli*. *Appl. Environ. Microbiol.* 79, 2201–2208.
- Kouprina, N., Larionov, V., 2006. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat. Rev. Genet.* 7, 805–812.
- Kouprina, N., Larionov, V., 2008. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* 3, 371–377.
- Kumpfmüller, J., Methling, K., Fang, L., Pfeifer, B.A., Lalk, M., Schweder, T., 2016. Production of the polyketide 6-deoxyerythronolide B in the heterologous host *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 100, 1209–1220.
- Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinet, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., Ussery, D.W., 2015. Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics.* 15, 141–161.
- Li, J.W., Vederas, J.C., 2009. Drug discovery and natural products: end of an era or an endless frontier? *Science* 325, 161–165.
- Liu, T., Mazmouz, R., Ongley, S.E., Chau, R., Pickford, R., Woodhouse, J.N., Neilan, B.A., 2017. Directing the heterologous production of specific cyanobacterial toxin variants. *ACS Chem. Biol.* 12, 2021–2029.
- Luo, Y., Enghiad, B., Zhao, H., 2016. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat. Prod. Rep.* 33, 174–182.
- Mo, T., Liu, W.Q., Ji, W., Zhao, J., Chen, T., Ding, W., Yu, S., Zhang, Q., 2017. Biosynthetic insights into linaridin natural products from genome mining and precursor peptide mutagenesis. *ACS Chem. Biol.* 12, 1484–1488.
- Muyrers, J.P., Zhang, Y., Testa, G., Stewart, A.F., 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* 27, 1555–1557.
- Newman, D.J., Cragg, G.M., 2016. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661.
- Okada, B.K., Seyedsayamdost, M.R., 2017. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol. Rev.* 41, 19–33.
- Oliynyk, M., Samborsky, M., Lester, J.B., Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., Leadlay, P.F., 2007. Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* 25, 447–453.
- Ongley, S.E., Bian, X., Neilan, B.A., Müller, R., 2013a. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep.* 30, 1121–1138.
- Ongley, S.E., Bian, X., Zhang, Y., Chau, R., Gerwick, W.H., Müller, R., Neilan, B.A., 2013b. High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. *ACS Chem. Biol.* 8, 1888–1893.
- Pearson, L.A., Hisbergues, M., Borner, T., Dittmann, E., Neilan, B.A., 2004. Inactivation of an ABC transporter gene, *mcvH*, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Appl. Environ. Microbiol.* 70, 6370–6378.
- Pettit, R.K., 2011. Small-molecule elicitation of microbial secondary metabolites. *Microb. Biotechnol.* 4, 471–478.
- Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 291, 1790–1792.
- Rebets, Y., Brotz, E., Tokovenko, B., Luzhetskyy, A., 2014. Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *J. Ind. Microbiol. Biotechnol.* 41, 387–402.
- Ren, H., Wang, B., Zhao, H., 2017. Breaking the silence: new strategies for discovering novel natural products. *Curr. Opin. Biotechnol.* 48, 21–27.
- Rouhiainen, L., Jokela, J., Fewer, D.P., Urman, M., Sivonen, K., 2010. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). *Chem. Biol.* 17, 265–273.
- Rutledge, P.J., Challis, G.L., 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* 13, 509–523.
- Saleh, O., Bonitz, T., Flinspach, K., Kulik, A., Burkard, N., Mühlenweg, A., Vente, A., Polnick, S., Lämmerhofer, M., Gust, B., Fiedler, H.P., Heide, L., 2012. Activation of a silent phenazine biosynthetic gene cluster reveals a novel natural product and a new resistance mechanism against phenazines. *MedChemComm* 3, 1009–1019.
- Schimming, O., Challinor, V.L., Tobias, N.J., Adihou, H., Grün, P., Pöschel, L., Richter, C., Schwalbe, H., Bode, H.B., 2015. Structure, biosynthesis, and occurrence of bacterial pyrrolizidine alkaloids. *Angew. Chem. Int. Ed. Engl.* 54, 12702–12705.
- Schorn, M.A., Alanjary, M.M., Aguinaldo, K., Korobeynikov, A., Podell, S., Patin, N., Linceum, T., Jensen, P.R., Ziemert, N., Moore, B.S., 2016. Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology* 162, 2075–2086.
- Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., Court, D.L., 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* 4, 206–223.
- Skinnider, M.A., Dejong, C.A., Rees, P.N., Johnston, C.W., Li, H., Webster, A.L., Wyatt, M.A., Magarvey, N.A., 2015. Genomes to natural products Prediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* 43, 9645–9662.

- Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* 18, 380–416.
- Staunton, J., Wilkinson, B., 1997. Biosynthesis of erythromycin and papamycin. *Chem. Rev.* 97, 2611–2630.
- Valiante, V., Mattern, D.J., Schuffler, A., Horn, F., Walther, G., Scherlach, K., Petzke, L., Dickhaut, J., Guthke, R., Hertweck, C., Nett, M., Thines, E., Brakhage, A.A., 2017. Discovery of an extended austinoïd biosynthetic pathway in *Aspergillus calidoustus*. *ACS Chem. Biol.* 12, 1227–1234.
- van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. *Trends Genet.* 30, 418–426.
- Viehrig, K., Surup, F., Volz, C., Herrmann, J., Abou Fayad, A., Adam, S., Kohnke, J., Trauner, D., Müller, R., 2017. Structure and biosynthesis of crocagins: polycyclic posttranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem. Int. Ed. Engl.* 56, 7407–7410.
- Wang, H., Li, Z., Jia, R., Yin, J., Li, A., Xia, L., Yin, Y., Muller, R., Fu, J., Stewart, A.F., Zhang, Y., 2017. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* 46, e28.
- Wang, X.K., Jin, J.L., 2014. Crucial factor for increasing the conjugation frequency in *Streptomyces netropsis* SD-07 and other strains. *FEMS Microbiol. Lett.* 357, 99–103.
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Brucoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H., 2015. AntiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243.
- Wenzel, S.C., Gross, F., Zhang, Y., Fu, J., Stewart, A.F., Müller, R., 2005. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via red/ET recombineering. *Chem. Biol.* 12, 349–356.
- Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* 111, 1957–1962.
- Yin, J., Hoffmann, M., Bian, X., Tu, Q., Yan, F., Xia, L., Ding, X., Stewart, A.F., Müller, R., Fu, J., Zhang, Y., 2015. Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2). *Sci. Rep.* 5, 15081.
- Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., Court, D.L., 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 97, 5978–5983.
- Zhang, H., Wang, Y., Wu, J., Skalina, K., Pfeifer, B.A., 2010. Complete biosynthesis of erythromycin A and designed analogs using *E. coli* as a heterologous host. *Chem. Biol.* 17, 1232–1240.
- Zhang, M.M., Wang, Y., Ang, E.L., Zhao, H., 2016. Engineering microbial hosts for production of bacterial natural products. *Nat. Prod. Rep.* 33, 963–987.
- Zhang, Y., Buchholz, F., Muirers, J.P., Stewart, A.F., 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* 20, 123–128.
- Zhang, Y., Muirers, J.P., Testa, G., Stewart, A.F., 2000. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* 18, 1314–1317.
- Ziemert, N., Alanjary, M., Weber, T., 2016. The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* 33, 988–1005.

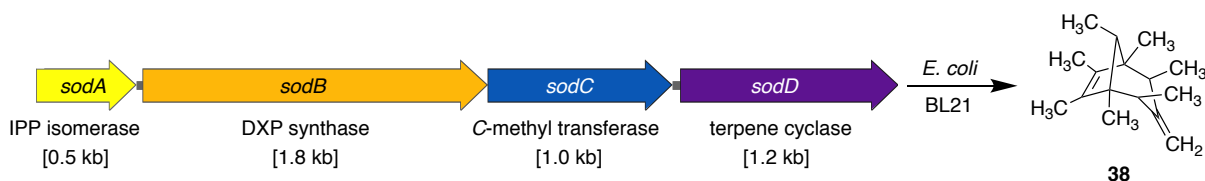
3.2 Direct Pathway Cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*

Based on: **E. R. Duell***, P. M. D'Agostino*, N. Shapiro, T. Woyke, T. M. Fuchs, T. A. M. Gulder: Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. *Microb. Cell Fact.* **2019**, *18*, DOI 10.1186/s12934-019-1080-6

*equally contributing authors

After initial demonstration of the usefulness of DiPaC for quick and simple cloning and expression of cyanobacterial natural products, its applicability for a different but equally challenging field of secondary metabolites, the VOCs, should be demonstrated. These substances are lipophilic molecules with a mass of less than 300 Da, high vapor pressure and low boiling point. Those characteristics make them highly valuable in Nature, as they can play important biological roles such as enabling long-distance communication or conveying cross-kingdom interactions [286–288]. In contrast to that, their physical properties make them hard-to-handle substances in a laboratory environment. By now, a lot of different VOCs have been structurally characterized, but knowledge regarding their biological role and function in the natural environment, as well as potential pharmaceutical applications is lacking behind.

Sodorifen (**38**) is a VOC with an unusual polymethylated hydrocarbon structure containing a bicyclo[3.2.1]octadiene skeleton and is emitted by some *Serratia plymuthica* strains [111, 116]. Despite extensive investigation over the last years, the biological function of **38** still remains elusive. Most *S. plymuthica* strains investigated so far emit dozens of different VOCs at the same time at very low concentrations, which makes the functional characterization of **38** very challenging [116, 289].



Scheme 5. Heterologous expression of the *S. plymuthica* WS3236 *sod* BGC in *E. coli* BL21 leads to efficient sodorifen (**38**) production. The *sod* cluster consists of four genes of which *sodA* and *sodB* are responsible for precursor supply and *sodC* and *sodD* catalyze the methylation and cyclization reactions to form **38**.

3 Results and Discussion



To tackle this problem and to demonstrate the broad applicability as well as the efficiency of DiPaC, the 4.6 kb sodorifen *sod* BGC from *S. plymuthica* WS3236 (see Scheme 5) was cloned and heterologously expressed within this thesis. The *sod* cluster was amplified as one fragment by PCR and integrated into a tetracycline inducible expression plasmid equipped with a *C*-terminal *gfp* transcription reporter gene. Heterologous expression in *E. coli* BL21 led to 26-fold increased emission of **38** compared to production rates in *S. plymuthica* WS3236 cultivated under the same conditions. By varying the media used for heterologous production, 90% pure **38** was detected in the raw extracts without any purification. In contrast to this, *S. plymuthica* WS3236 emitted a variety of different VOCs with **38** by far not being the most abundant one. In summary, within this work a DiPaC-based, quick and easy access for further investigations and characterizations of challenging secondary metabolites could be demonstrated.

RESEARCH

Open Access



Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*

Elke R. Duell^{1†} , Paul M. D'Agostino^{1†} , Nicole Shapiro², Tanja Woyke², Thilo M. Fuchs^{3,4} and Tobias A. M. Gulder^{1,5*} 

Abstract

Background: *Serratia plymuthica* WS3236 was selected for whole genome sequencing based on preliminary genetic and chemical screening indicating the presence of multiple natural product pathways. This led to the identification of a putative sodorifen biosynthetic gene cluster (BGC). The natural product sodorifen is a volatile organic compound (VOC) with an unusual polymethylated hydrocarbon bicyclic structure (C₁₆H₂₆) produced by selected strains of *S. plymuthica*. The BGC encoding sodorifen consists of four genes, two of which (*sodA*, *sodB*) are homologs of genes encoding enzymes of the non-mevalonate pathway and are thought to enhance the amounts of available farnesyl pyrophosphate (FPP), the precursor of sodorifen. Proceeding from FPP, only two enzymes are necessary to produce sodorifen: an S-adenosyl methionine dependent methyltransferase (SodC) with additional cyclisation activity and a terpene-cyclase (SodD). Previous analysis of *S. plymuthica* found sodorifen production titers are generally low and vary significantly among different producer strains. This precludes studies on the still elusive biological function of this structurally and biosynthetically fascinating bacterial terpene.

Results: Sequencing and mining of the *S. plymuthica* WS3236 genome revealed the presence of 38 BGCs according to antiSMASH analysis, including a putative sodorifen BGC. Further genome mining for sodorifen and sodorifen-like BGCs throughout bacteria was performed using SodC and SodD as queries and identified a total of 28 *sod*-like gene clusters. Using direct pathway cloning (DiPaC) we intercepted the 4.6 kb candidate sodorifen BGC from *S. plymuthica* WS3236 (*sodA–D*) and transformed it into *Escherichia coli* BL21. Heterologous expression under the control of the tetracycline inducible *Ptet_O* promoter firmly linked this BGC to sodorifen production. By utilizing this newly established expression system, we increased the production yields by approximately 26-fold when compared to the native producer. In addition, sodorifen was easily isolated in high purity by simple head-space sampling.

Conclusions: Genome mining of all available genomes within the NCBI and JGI IMG databases led to the identification of a wealth of *sod*-like pathways which may be responsible for producing a range of structurally unknown sodorifen analogs. Introduction of the *S. plymuthica* WS3236 sodorifen BGC into the fast-growing heterologous expression host *E. coli* with a very low VOC background led to a significant increase in both sodorifen product yield and purity compared to the native producer. By providing a reliable, high-level production system, this study sets the stage for future investigations of the biological role and function of sodorifen and for functionally unlocking the bioinformatically identified putative *sod*-like pathways.

*Correspondence: tobias.gulder@ch.tum.de

†Elke R. Duell and Paul M. D'Agostino contributed equally to this work

¹ Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany
Full list of author information is available at the end of the article



Keywords: Sodorifen, *Serratia plymuthica*, Terpenes, Genome mining, DiPaC, Heterologous expression

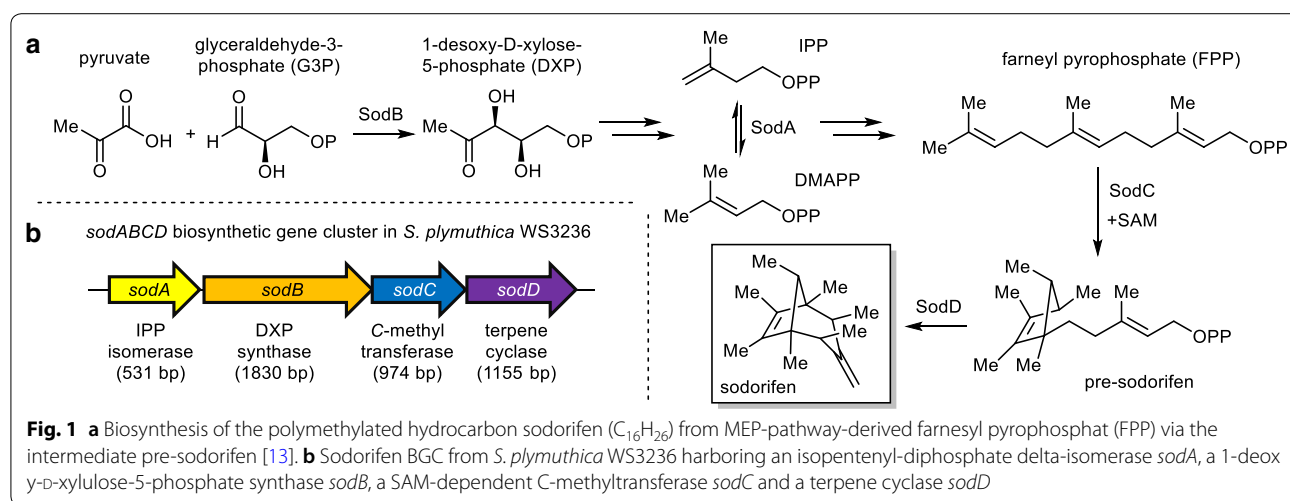
Background

Volatile organic compounds (VOCs) are lipophilic small molecules (<300 Da) that are characterized by their high vapor pressures and low boiling points. They play important biological roles, e.g., as signaling molecules by enabling communication over long-distances and are able to facilitate cross-kingdom interactions [1–3]. Thus, understanding their ecological functions is important, for instance when looking deeper into microbial ecology or plant-rhizosphere interactions. The rhizobacterium *Serratia plymuthica* 4Rx13, for example, is emitting over 100 different VOCs which have been shown to inhibit plant and fungal growth [4–9]. The VOC profile of *S. plymuthica* 4Rx13 contains sodorifen as one major metabolite, a polymethylated hydrocarbon with an unusual bicyclo[3.2.1]octadiene skeleton (Fig. 1a) [10].

Recently, a 4.6 kb sodorifen (*sod*) BGC cluster harboring four individual genes was found to be responsible for the sodorifen production in *S. plymuthica* 4Rx13 (Fig. 1b) [11]. The first two genes, an isopentenyl-diphosphate delta-isomerase (IPP isomerase, *sodA*) and a 1-deoxy-D-xylulose-5-phosphate synthase (DXP synthase, *sodB*) are homologs of genes found in the non-mevalonate pathway (MEP-pathway) and may supply additional farnesyl-pyrophosphat (FPP), the precursor of sodorifen. Notably, knock-out-mutants showed that SodB is not crucial for sodorifen emission, most likely due to the fact that one of the house-keeping enzymes of the MEP-pathway provides sufficient precursor flux [11]. Expression of *sodCD* in *Escherichia coli* additionally harboring the plasmid pMEV with three MEP-pathway associated enzymes led to the production of minute amounts of

sodorifen and its isomers [12]. Recent investigations in vivo and in vitro demonstrated that the SAM-dependent C-methyltransferase SodC is not only responsible for the addition of a methyl group to FPP, but also exhibits a surprising cyclisation activity leading to the phosphorylated version of pre-sodorifen, which is then cyclized by the terpene cyclase SodD to yield the final bicyclic compound sodorifen (Fig. 1a) [13].

Interestingly, sodorifen is produced by multiple members of the rhizobacterium family of *S. plymuthica* [10] suggesting an important ecological role of this specialized metabolite. Most *S. plymuthica* strains investigated so far emit a complex mixture of different VOCs mainly at very low concentrations [4], greatly hampering the efficient production for functional screening of individual compounds such as sodorifen. Sodorifen production can be triggered to some extent by co-cultivating *S. plymuthica* with the fungal pathogen *Fusarium culmorum* [12] and is furthermore regulated by the carbon catabolite repression system [14]. Different isolates of *S. plymuthica* exhibit surprisingly variable amounts of sodorifen emission, ranging from <0.1 to 50% of the total VOC spectrum under identical fermentation conditions [14]. Given the rather low titers of sodorifen in all natural producers, its inherent volatility and the observed complexity of VOC mixtures obtained by fermentation of *S. plymuthica*, studies on the functional role of this intriguing metabolite have so far been impeded. Optimized biotechnological production of sodorifen in a suitable recombinant host with low VOC background has the potential to solve this apparent supply problem thus enabling in-depth functional studies. Such an



approach requires the cloning and successful heterologous, functional expression of the encoding biosynthetic gene cluster. Biosynthetic gene clusters can be captured using several in vivo methods based on recombineering, such as liner-linear homologous recombineering (LLHR) [15] or linear-circular-homologous recombineering (LCHR) [16], exonuclease combined recombination (ExoCET) [17] and Cas9-assisted targeting of chromosome segments (CATCH) [18], all utilizing the Rec/ET system [19], or using the natural recombination capability of *Saccharomyces cerevisiae* for transformation-associated recombination (TAR) cloning [20–22]. In addition to this, there are PCR based in vitro BGC cloning techniques such as circular polymerase extension cloning (CPEC) [23], assembly of fragment ends after PCR (AFEAP) [24] and direct pathway cloning (DiPaC) [25, 26]. DiPaC is characterized by long-amplicon PCR combined with homologous nucleotide overhangs which allows for in vitro DNA assembly via Gibson assembly or sequence- and ligation-independent cloning (SLIC) of the BGC directly into the expression vector of choice. This enables the rapid, efficient and cheap capturing and expression of BGCs of interest in heterologous hosts such as *E. coli* or *Streptomyces* spp. [25, 26]. Within this study we present the application of DiPaC to functionally link a putative sodorifen biosynthetic gene cluster to sodorifen production in a recombinant host system and the optimization of the fermentative strategy to almost exclusive, high-level sodorifen production.

Results

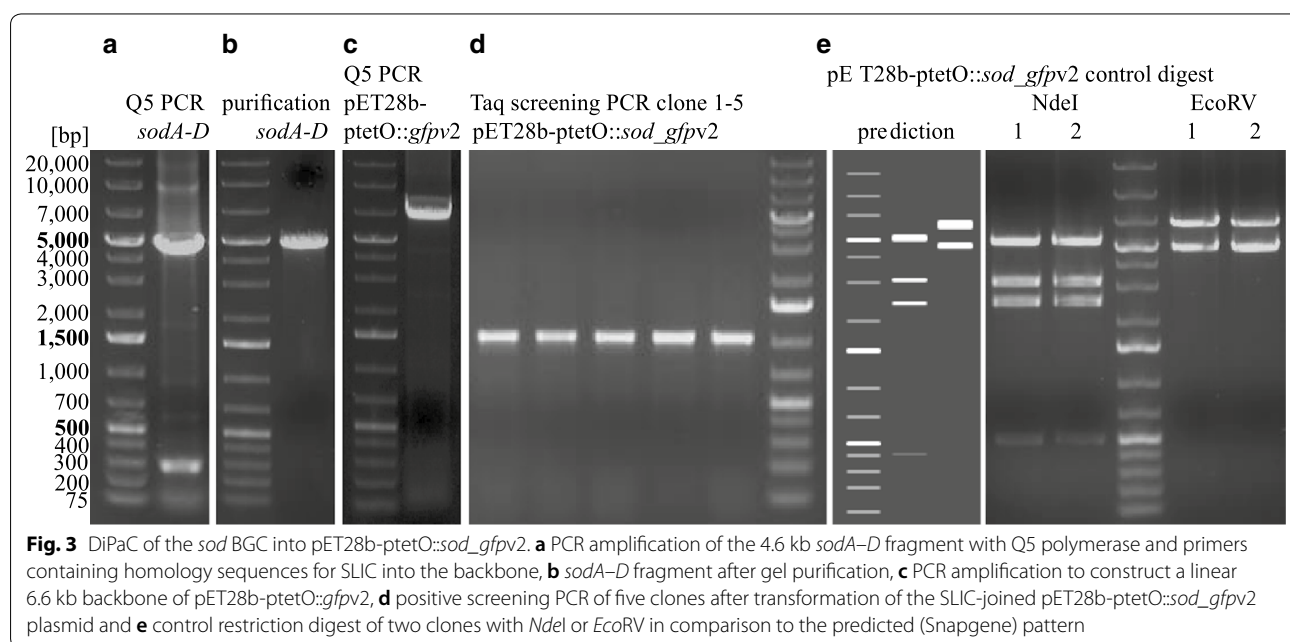
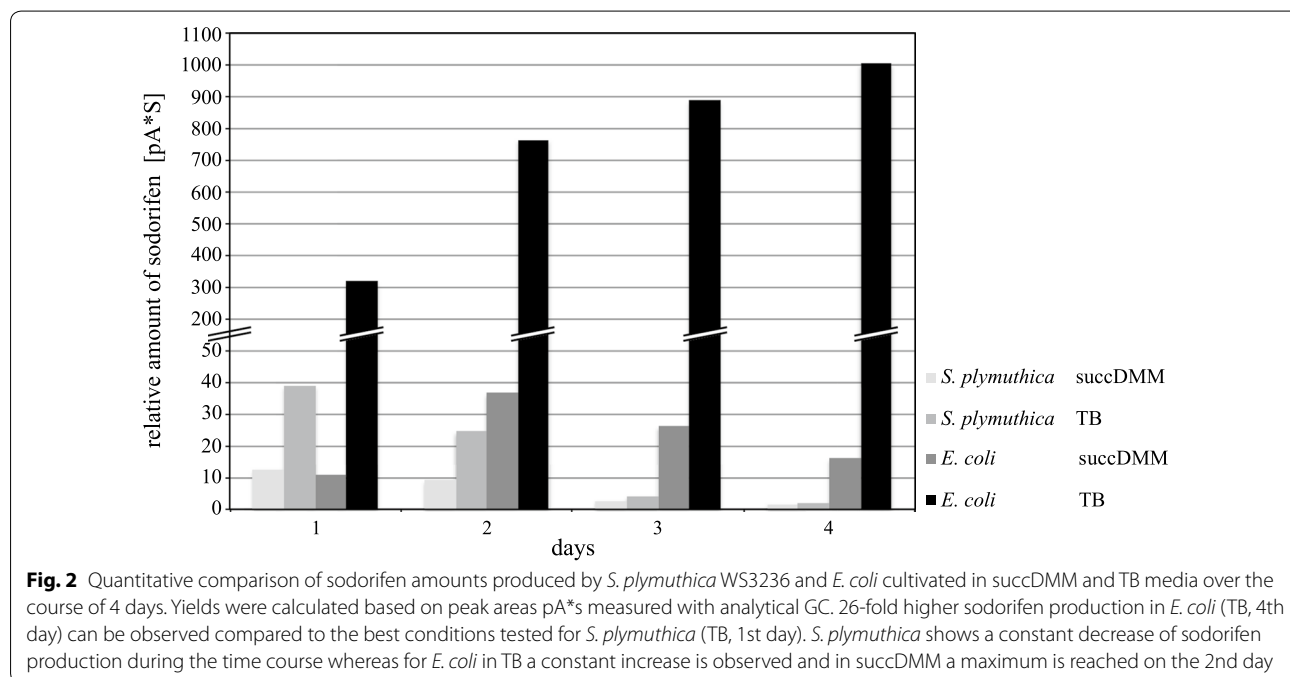
Serratia plymuthica WS3236 was selected for WGS after preliminary genetic and chemical screening (data not shown) that pointed to the presence of natural product pathways. The resulting genome has a GC content of 55.93% and encodes a total of 5107 genes, of which 4915 (96.24%) are protein encoding genes. The genome sequence has been deposited at GenBank SRA (BioProject ID: PRJNA442736) and IMG/MER (Accession: 2773857786) databases. To disclose the secondary metabolite potential of *S. plymuthica* WS3236, an in silico approach utilizing antiSMASH [27] and ClusterFinder [28] algorithms was performed. A total of 38 BGCs were identified (Additional file 1: Table S1), of which two displayed high homology to known BGCs encoding zeamine [29] (MIBiG BGC-ID: BGC0001056_c1) and sodorifen [11, 13] (MIBiG BGC-ID: BGC0001361_c1). A further seven did not show similarity to characterized pathways but belonged to well-known natural product families including non-ribosomal peptide synthase (NRPS; 4 clusters), arylpolyene-siderophore (1 cluster), polyketide synthetase/NRPS (PKS/NRPS; 1 cluster) and thiopeptide (1 cluster) biosynthesis (Additional file 1: Table S1); the

remaining 29 pathways were predicted by ClusterFinder and either belonged to fatty acid, saccharide or putative pathways. Within this work, we decided to further investigate the candidate sodorifen BGC, to (1) firmly validate its small molecule product and (2) establish a reliable high-level recombinant system for sodorifen production that will facilitate future investigations into its biological functions.

To functionally link the *sod* candidate BGC to sodorifen production, *S. plymuthica* WS3236 was cultivated either in TB or succDMM, the latter being reported as the optimal sodorifen production medium for *S. plymuthica* [13]. Sodorifen yields were highest after day 1 and constantly decreased after this time point (Fig. 2). Interestingly, we found that cultivation in TB resulted in slightly higher sodorifen emission, which might solely be a consequence of higher growth rates and cell densities compared to the minimal succDMM medium. As neither TB nor succDMM contain glucose, the carbon catabolite repression system regulating the *sod* BGC, as described by Magnus et al. [14], does not come into effect. Most importantly, all production experiments with the natural producer *S. plymuthica* WS3236 only resulted in small amounts of sodorifen relative to other VOCs produced. As purification of highly volatile organic compounds is generally difficult, we set out to construct a heterologous production system that reliably gives access to larger amounts of sodorifen, most importantly with high purity by simple head-space sampling.

For the interception of the *sod* BGC the DiPaC strategy was applied. An expression plasmid containing the tetracycline inducible promoter *Ptet_O* was chosen for heterologous expression as the use of stronger promoters such as T7 has previously been shown to frequently hamper secondary metabolite production [30]. Therefore the 4.6 kb *sod* BGC was PCR amplified in one piece using primers equipped with homologous overhangs for the vector backbone pET28b-ptetO::*gfpv2* (see Fig. 3a, b). This vector carries a copy of the *gfp* gene downstream of the multiple cloning site thus enabling the validation of transcription/translation of the complete construct by simple fluorescence detection of GFP. After linearizing the vector backbone by PCR (see Fig. 3c), the two fragments were joined using SLIC. Successful integration of the *sod* BGC into pET28b-ptetO::*gfpv2* was verified by colony screening PCR, analytical restriction digest and sequencing of the integration site of the resulting plasmid pET28b-ptetO::*sod_gfpv2* (see Fig. 3 and Additional file 1: Figure S2).

Full transcription/translation of the *sod* BGC under the control of *Ptet_O* was confirmed by detecting green fluorescence in the presence of tetracycline, but not in the controls (see Additional file 1: Figure S3). SDS



gel analysis of the whole cell protein of *S. plymuthica* WS3236, *E. coli* BL21 as well as induced versus uninduced *E. coli* carrying plasmid pET28b-ptetO::*sod_gfpv2* did not reveal any differences in the protein expression pattern (see Additional file 1: Figure S4). This clearly underlines the versatility of the applied fluorescence screening using the GFP reporter protein downstream of the *sod* pathway to prove successful

transcription/translation of the entire construct. GC-MS analysis of all VOCs emitted into the headspace of the heterologous fermentation in all three tested media revealed the accumulation of a molecule with a molecular mass of m/z 218 [M⁺] which clearly showed the typical mass spectrum of sodorifen [10] (see Additional file 1: Figure S5). In addition, low amounts of sodorifen isomers known to also be produced by *S. plymuthica*

were likewise observed [13]. As no potential sodorifen pathway intermediates were detected, apparently all terpenoid precursors available to the sodorifen biosynthetic system were fully converted to the end product by SodC and SodD.

Optimization of the heterologous production of sodorifen was tested in three different media: LB, TB and succDMM. The amounts of emitted VOC products were compared to those of *S. plymuthica* WS3236 under the same growth conditions. Using succDMM as cultivation medium for *E. coli* BL21 pET28b-ptetO::*sod_gfpv2*, the yield of emitted sodorifen reached a maximum on day 2 and was comparable to the production of *S. plymuthica* WS3236 grown in TB (cf. Fig. 2). Performing the heterologous expression in LB medium showed a maximal production on day 2, and increased the yield about tenfold compared to succDMM (see Additional file 1: Table S4). In contrast, *E. coli* BL21 pET28b-ptetO::*sod_gfpv2* cultivation in TB showed a tremendous boost of sodorifen emission which continuously increased with each day and revealed a 26-fold higher production titer on day 4 when compared to the optimal production condition of the original producer (Fig. 2; see Additional file 1:

Table S4 for a complete overview of produced sodorifen amounts).

GC and GC–MS analysis revealed that *S. plymuthica* WS3236 produces a variety of different VOCs in comparable amounts (Fig. 4) as described previously [4]. In contrast to this result, *E. coli* pET28b-ptetO::*sod_gfpv2* exhibits a very clean VOC profile with sodorifen as the major VOC (Fig. 4). Only in fermentations carried out in LB we found a second compound that significantly added to the spectrum of emitted substances and was subsequently identified as indole by NMR analysis, a VOC known to be produced by *E. coli* in high quantities [31, 32]. NMR measurements of the crude extract from the sampled head space of the heterologous *E. coli* fermentation in TB on day 2 showed high purity of sodorifen without any further purification (see Additional file 1: Figure S6). The relative production titers analyzed by GC–MS based on relative peak areas compared to mesitylene as an internal standard with defined concentration were nicely consistent across analyses of replicates (e.g., ranging from 232 to 318 on day one and from 1003 to 1018 at maximum production on day 3 or 4). However, isolated yields of sodorifen derived by head-space sampling

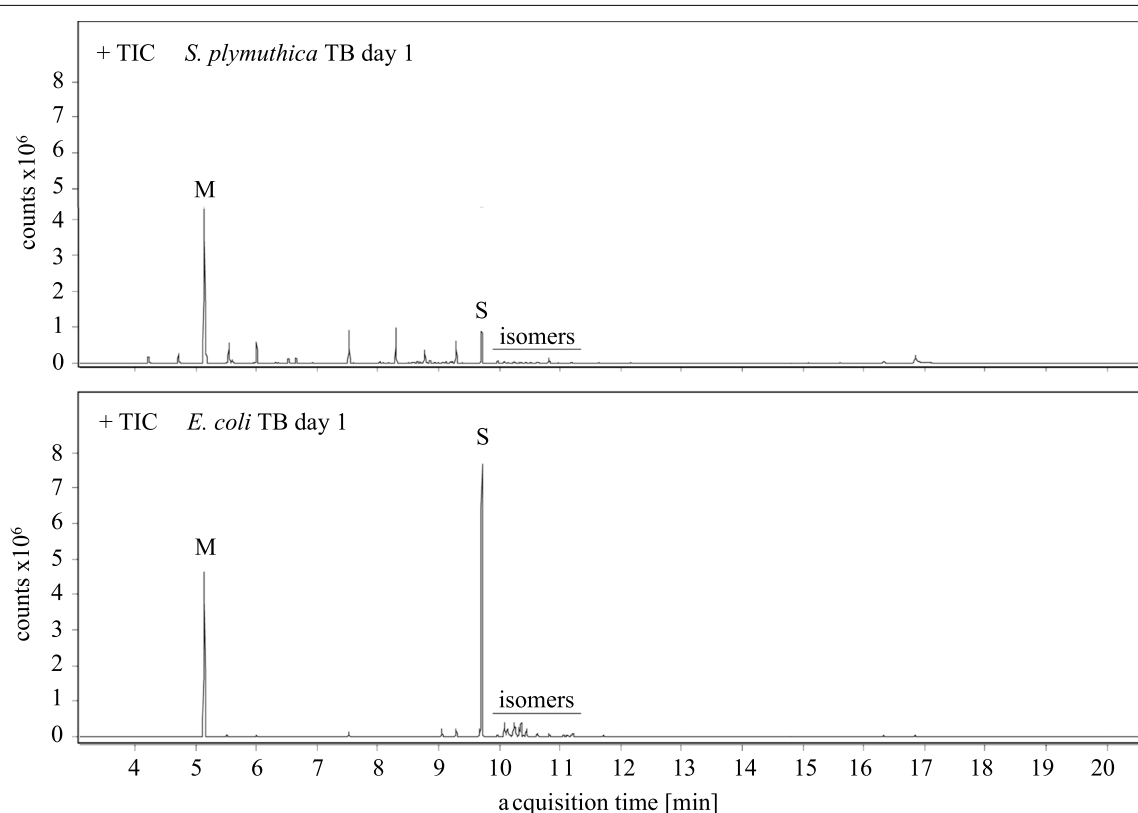


Fig. 4 GC–MS total ion chromatograms of *S. plymuthica* WS3236 (top) and *E. coli* pET28b-ptetO::*sod_gfpv2* (bottom) VOCs produced in TB medium within the first 24 h hours of cultivation. Mesitylene (M) was added as an internal standard with a concentration of 250 $\mu\text{g}/\text{mL}$. The amount of sodorifen (S) produced heterologously by *E. coli* is significantly higher combined with better purity compared to *S. plymuthica*

varied more significantly (e.g., ranging from 6.4 to 7.9 mg within 24 h at maximum production on days 3 or 4). This is a direct consequence of the volatility of sodorifen, requiring careful removal of the organic solvent under reduced pressure (pentane, used for extraction of the target compound from activated charcoal, see “Methods”). The total amount of combined isolated sodorifen head-space extract from TB fermentation over a time-course of 4 days thus ranged from 13.4 up to 25.5 mg. The developed expression system is therefore suitable for the production of sufficient amounts of sodorifen by simple head-space sampling thus establishing a reliable supply of the compound.

Given the conserved nature of the *sod* pathway and the unique rearranged terpene structure that is assembled following the *sod* biosynthetic logic, we aimed at identifying further *sod*-type BGCs by performing a bioinformatic screen of all publicly available microbial genomes within the NCBI and JGI IMG databases using SodC (methyltransferase) and SodD (terpene cyclase) as a query sequence, since these are essential for sodorifen biosynthesis. A total of 28 putative *sod*-like gene clusters were identified in *S. plymuthica*, *Serratia* sp.,

Pseudomonas chlororaphis, *Pseudomonas grimontii*, *Pseudomonas schloroaphis*, *Burkholderia pyrrocinia*, *Burkholderia singularis* and *Streptomyces tsukubensis* (Additional file 1: Table S2). While a *sod*-like gene cluster has previously been reported in *P. chlororaphis* O6 and *Streptomyces tsukubensis* NRRL18488 [11, 12], we report similar genomic regions in several other species here for the first time. Comparison of the gene synteny allowed the categorization of *sod* clusters into eight groups including differences in encoding one or several oxidoreductase-type enzymes, multiple methyltransferases, or multiple copies of terpene cyclase-type enzymes (Fig. 5). Based on our data, *sod*-like clusters are unique only at the inter-species level. *Serratia* sp. FS14 displayed an intriguing phylogenetic distribution, with the FS14-SodD clustering with the *P. chlororaphis* clade (Additional file 1: Figure S1). Of further interest was the identification of three similar terpene cyclase genes forming an operon within the *B. singularis* TSV85 *sod* cluster. This bioinformatic analysis clearly reveals the presence of further *sod*-type BGCs that are interesting targets for functional expression to obtain further unusual, structurally novel terpenoids.

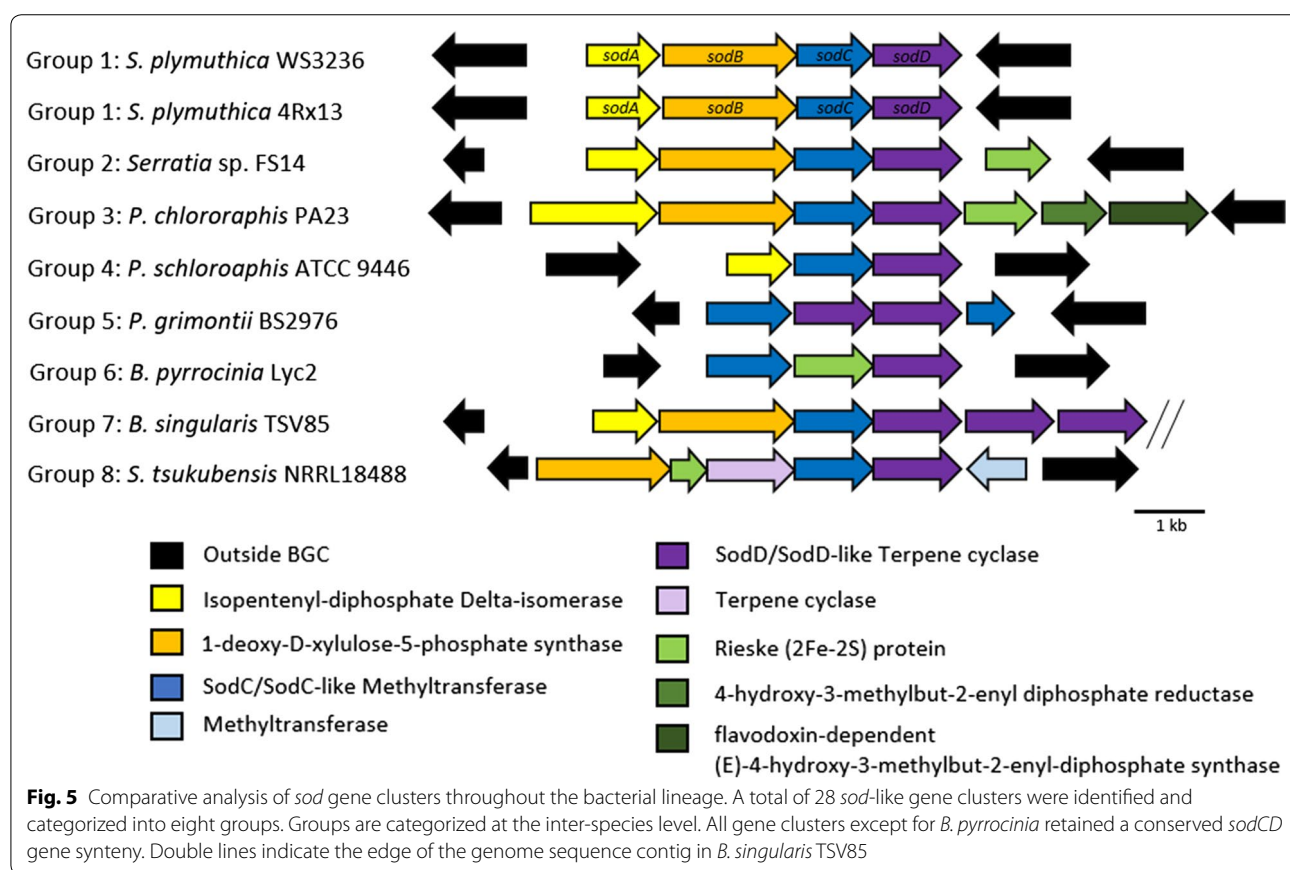


Fig. 5 Comparative analysis of *sod* gene clusters throughout the bacterial lineage. A total of 28 *sod*-like gene clusters were identified and categorized into eight groups. Groups are categorized at the inter-species level. All gene clusters except for *B. pyrrocinia* retained a conserved *sodCD* gene synteny. Double lines indicate the edge of the genome sequence contig in *B. singularis* TSV85

Discussion

Serratia plymuthica is notorious for producing a wide array of VOCs. Sequencing data of the *S. plymuthica* WS3236 genome was consistent with other *S. plymuthica* strains, in particular the well characterized *S. plymuthica* 4Rx13 (formerly *Serratia odorifera* 4Rx13 [4]), which has a similar genome size and number of CDS. In silico analysis of the encoded BGCs within *S. plymuthica* WS3236 revealed 38 in total, with 9 BGCs belonging to well-characterized natural product families. Comparison of BGCs against *S. plymuthica* 4Rx13 revealed only 6 out of the 38 clusters are shared between the two organisms, highlighting the inter-species BGC variability in the genus *Serratia*.

Most *S. plymuthica* strains investigated so far emit a complex mixture of a large number of different VOCs, most of them produced at very low levels [4], making the functional screening of individual compounds such as sodorifen almost impossible. Schmidt et al. [12] recently developed a successful *E. coli* based system harboring the sodorifen methyltransferase and the terpene cyclase as well as an additional mevalonate pathway-expression plasmid encoding a mevalonate kinase, a phosphomevalonate kinase and a mevalonate pyrophosphate decarboxylase to increase the presence of farnesyl pyrophosphate, the precursor of sodorifen, but were only able to produce low amounts of sodorifen [12]. To further improve the yield of heterologously producible sodorifen, we thus decided to use the entire natural sodorifen production gene cassette under the control of an inducible promoter. We selected the tetracycline-induced promoter *Ptet_O*, which has successfully been utilized in the heterologous expression of multiple natural product BGCs [25, 26, 33]. This expression system was also successful in this study and allowed for a constant sodorifen biosynthesis in amounts significantly higher than the natural producer. With this approach, we were able to increase the amount of emitted sodorifen up to 26-fold compared to the optimal production conditions of the original producer *S. plymuthica* WS3236. In addition, we obtained sodorifen in high purity by head-space sampling, thus providing an efficient system for sodorifen production, which sets the stage for further investigations into the biological function of this interesting compound.

Within this work, we furthermore investigated the occurrence of additional candidate sodorifen-type BGCs in bacteria. In silico screening of published genomes led to the identification of 28 *sod* and *sod*-like gene clusters throughout bacteria, which is a significant increase on previously reported numbers [11, 12]. Intriguingly, while *Serratia* sp. FS14 was isolated from plants, similar to other sodorifen producing bacteria, its genome is more closely related to non-plant

associated *Serratia marcescens* [34], but still encodes a *sod* gene cluster. The *sod* cluster of *Serratia* sp. FS14 differed significantly compared to other *Serratia*, with a much lower sequence identity (~63% FS14 vs WS3236 compared to ~96% 4Rx13 vs WS3236) and the presence of a Rieske (2Fe-2S) protein that is similar to those in *P. chlororaphis* (58%) and *B. pyrrocinia* (56%). The phylogenetic clustering of FS14-SodD with homologous proteins from *P. chlororaphis* (Additional file 1: Figure S1) may be the first hint that organisms outside of *S. plymuthica* species are capable of producing sodorifen or sodorifen analogs.

Based on gene synteny analysis, we were able to categorize the *sod*-like clusters into eight groups, which mostly separated at the species level. Differences between these *sod*-like clusters include the presence of one or several oxidoreductase-type enzymes, multiple methyltransferases, or multiple copies of terpene cyclase-type enzymes. This suggests that there are significant modifications to the sodorifen backbone in order to produce a range of analogues, which could be a significant source of novel unusual bacterial terpenes.

Conclusions

Genome sequencing and bioinformatic analysis revealed that there is a significant BGC diversity within different *S. plymuthica* species. Comparison and phylogenetic analysis of SodD of *Serratia* sp. FS14 indicates this pathway shares some homology with *P. chlororaphis*, raising interesting evolutionary questions about sodorifen biosynthesis. Our in silico analysis furthermore revealed the presence of a diversity of *sod*-type BGCs across phylogenetically diverse bacteria. Significant genetic variability (e.g., additional methyltransferases, terpene cyclases and oxidoreductase enzymes) between these pathways indicates that there likely is a plethora of unique and structurally intriguing sodorifen analogs awaiting discovery. Using DiPaC, we quickly integrated the native four gene containing sodorifen cluster *sodABCD* into a vector under the control of one single promoter upstream of *sodA*, thereby mimicking the transcription setting in the original producer. This ensures that the relevant enzymes of the MEP-pathway SodA and SodB are overproduced synchronously with SodC and SodD, which finally transform SAM and FPP into sodorifen. By screening a variety of different culture conditions, sodorifen production was increased up to 26-fold compared to the original producer, accompanied by significantly improved purity in the crude extract. The results of this work pave the way for future studies investigating the biological role and function of sodorifen.

Methods

Bacterial culturing, strains and plasmids

Bacterial strains and plasmids generated in this study are listed in Table 1. *S. plymuthica* WS3236 was obtained from the Weihenstephaner collection of microorganisms (Chair of Microbial Ecology, Technical University of Munich Freising, Germany). *S. plymuthica* WS3236 was cultivated in nutrient broth (Carl Roth, Germany) at 28 °C (liquid cultures) or 30 °C (agar plates). Liquid cultures were incubated while shaking at 200 rpm. Under routine culture conditions unless otherwise specified, *E. coli* strains were grown in lysogeny broth (LB) medium (Carl Roth, Germany) supplemented with 50 µg/mL kanamycin.

Genomic DNA extraction, whole genome sequencing (WGS) and assembly

Extraction of high-molecular weight genomic DNA was performed as described by Micallef [35] and Greunke et al. [25] with few alterations. After culturing, cells were placed at 4 °C for 24 h to stop cell division and DNA replication prior to DNA extraction. After centrifugation (10 min at 10,000×g), the cell pellet was resuspended in 5 mL lysis buffer (25 mM EDTA, 0.3 M sucrose, 25 mM Tris-HCl [pH 7.5]). The cells were freeze-thawed three times (liquid nitrogen/50 °C) to promote cell lysis. Lysozyme (Sigma-Aldrich) was added to a final concentration of 1 mg/mL and RNase (Carl Roth, Germany) to a final concentration of 10 µg/mL followed by an incubation at 37 °C for 60 min. Proteinase K (Amresco, USA) was added to a final concentration of 0.5 mg/mL followed by the addition of SDS to a final concentration of 1% (w/v). The cell mixture was incubated at 37 °C for 30 min followed by 55 °C for 30 min. NaCl was added to a final concentration of 1.3 M, and cetyltrimethylammonium bromide (CTAB) [10% (w/v) CTAB (Sigma-Aldrich) in 0.7 M NaCl] to a final concentration of 1% (v/v) CTAB solution, with the mixture incubated at 65 °C for 10 min. A total of 1 vol of chloroform:isoamyl alcohol (24:1) was

added to the cell lysis solution, mixed by inversion and incubated on ice while shaking for 30 min. The aqueous phase was removed and phenol–chloroform–isoamylalcohol (25:24:1) extraction was performed twice before the addition of 0.6 vol of isopropanol. The DNA cell pellet was washed with ice cold 70% ethanol and finally dissolved in 0.1× TE buffer (1 mM Tris-HCl, 0.1 mM EDTA [pH 8.0]). The quality of gDNA was analyzed by gel electrophoresis using 0.7% (w/v) agarose gels and quantified with a P330 NanoPhotometer (Implen, Germany).

The whole genome of *S. plymuthica* WS3236 was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology [36]. A > 10 kb PacBio SMRTbell library was constructed and sequenced on the PacBio RS2 platform, which generated 78,182 filtered subreads totaling 307,584,001 bp. The raw reads were assembled using HGAP (smrtanalysis/2.3.0 p5, HGAP 3) [37]. The final draft assembly contained 1 contig in 1 scaffold, totaling 5,349,225 bp in size. The input read coverage was 45.2×. The genome sequencing QC report can be found as Additional file 2.

Bioinformatic and phylogenetic analysis

In silico analysis of the BGCs present in the *S. plymuthica* WS3236 genome was performed using AntiSMASH (Version 4) [27]. The presence of the *sod* cluster throughout all available genome sequences against the NCBI and JGI IMG databases (as of October 2018) was searched using the *S. plymuthica* SodC (methyltransferase; locus_tag: Ga0236286_4400) and SodD (terpene cyclase; locus_tag: Ga0236286_4399) as query sequences. BlastP analysis with these query sequences was performed against the NCBI non-redundant and JGI Integrated Microbial Genomes System [38] databases. All hits consisting of adjacent methyltransferase and terpene cyclase genes were categorized as possible *sod*-type clusters. Phylogenetic analysis was performed using the Phylogeny.fr online tool with 'one click' analysis [39, 40]. All

Table 1 Bacterial strains and plasmids used in this study

	Description	Reference or source
Strains		
<i>Escherichia coli</i> DH5a	Host strain for cloning	NEB
<i>Escherichia coli</i> BL21	Heterologous expression strain	NEB
<i>Serratia plymuthica</i> WS3236	Native producer of sodorifen	ZIEL Institute Culture Collection
Plasmids		
pET28b-ptetO- <i>gfpv2</i> (6029 bp)	Tetracycline inducible expression plasmid, ColE1, Kan ^R , <i>gfp</i> reporter gene downstream of promoter	This study
pET28b-ptetO:: <i>sod_gfpv2</i> (11,124 bp)	pET28b-ptetO- <i>gfpv2</i> with <i>sodABCD</i> cloned as single fragment between P _{tetO} and <i>gfp</i>	This study

sequences were stored and visualized using the Geneious Software Package.

Direct pathway cloning of the *sod* cluster

The 4.6 kb *sod* cluster of *S. plymuthica* WS3236 was cloned in one piece into pET28b-ptetO using SLIC-mediated DiPaC [26]. Linearized PCR amplicons of both the vector backbone and the *sodABCD* fragment were generated using PCR with Q5 polymerase (NEB) under standard conditions with 50 ng gDNA per 25 μ L reaction setup (see Additional file 1: Table S3 for primer sequences and chapter 2.1 for PCR setup). At the 5' end of the cluster specific primer pair, 22 bp homology sequences consistent with the terminal region of the PCR generated vector were added. After purification, the fragments were assembled with T4-DNA polymerase and 8 μ L of the reaction mixture was chemically transformed into *E. coli* DH5 α . Positive pET28b-ptetO::*sod_gfpv2* clones were selected by Taq screening PCR with primers binding *sodD* and the T7 terminator (see Additional file 1: Table S3 for primer sequences and chapter 2.2 for PCR setup) and confirmed using restriction digest and Sanger sequencing of the integration sites.

Bacterial fermentation for sodorifen production

All fermentations were performed in 3 L scale using a BIOSTAT A plus fermenter (type 8843812, Sartorius Stedim Biotech) and the volatile compounds were collected at the gas outlet with a filter consisting of 0.5 g pure activated charcoal between two VitraPOR glass filters (porosity 00, ROBU, Germany) in a plastic centrifuge column (Pierce 5 mL, Thermo Scientific, Germany). 3 L cultures were inoculated 1:50 from overnight pre-cultures grown in LB medium and 300 μ L of Antifoam SE-15 (Sigma-Aldrich/Merck, Germany) was added. *S. plymuthica* WS3236 cultures were grown in terrific broth (TB) medium (Carl Roth, Germany) and modified Davis and Mingioli succinate minimal medium (succDMM: 7 g/L $K_2HPO_4 \cdot 3 H_2O$, 3 g/L KH_2PO_4 , 1 g/L $(NH_4)_2SO_4$, 0.5 g/L sodium citrate dihydrate, 6.49 g/L succinate, 0.1 g/L $MgSO_4 \cdot 7 H_2O$, pH 6.2 with NaOH) at 30 $^\circ$ C with 200 rpm stirring. Trapping of volatile compounds started directly after inoculation and the filters were exchanged every 24 h for up to 4 days. 300 μ L of Antifoam SE-15 was again added after 40 h and 72 h.

For the heterologous expression of sodorifen, *E. coli* BL21 cells were chemically transformed with pET28b-ptetO::*sod_gfpv2*. Expression was carried out in LB, TB and succDMM, each supplemented with 50 mg/L kanamycin. After inoculation, cells were grown at 37 $^\circ$ C at 200 rpm for 5 h before the temperature was lowered to 16 $^\circ$ C and expression induced with 708 μ g/L tetracycline. Volatile compound collection was started immediately

after induction and the filters were exchanged every 24 h for up to 4 days. 300 μ L of Antifoam SE-15 was again added after 48 h.

After fermentation, the active charcoal was extracted twice with 2 mL pentane, which was filtered through a 0.45 μ m PTFE syringe filter (Fisher Scientific, Germany) to remove remaining coal pieces before evaporation in vacuo. The residues were weighed and re-suspended in pentane or $CDCl_3$ for further measurements. Mesitylene was used as an internal standard for GC analysis with a concentration of 250 μ g/mL.

GC–(MS) and NMR analysis of bacterial head-space extracts

Analytical gas chromatography was performed at a HP 6890 Series GC (Agilent, stationary phase: HP-5 column, poly-dimethyl/diphenyl-siloxane, 95/5) with a flame ionization detector using the following temperature profile: 60 $^\circ$ C (hold 3 min), then 15 $^\circ$ C/min to 250 $^\circ$ C (hold 5 min). The amounts of mesitylene and sodorifen were determined by peak area integration. Mass spectrometric analysis was performed with electron impact ionization (EI, 70 eV) on a Agilent HP 6890 Series GC–MS (Agilent, stationary phase: HP-5MS column, poly-dimethylsiloxane, 30 m, mass detection: Agilent 5973 Network Mass Selective Detector) using the same temperature profile as for analytical GC. NMR spectra of sodorifen and the *E. coli*-derived indole were directly recorded from head-space samples on a Bruker AVHD500 and a Bruker AV500-cryo spectrometer in $CDCl_3$ (see Additional file 1: Figure S6).

Additional files

Additional file 1. Antismash and Cluster Finder results; organisation of *sod*-like BGCs into cluster types; phylogenetic analysis of SodC and SodD; methods for the cloning and expression of the *sod* BGC; results of the heterologous expression of the *sod* cluster; GC–MS spectrum of sodorifen; NMR spectra of raw head-space samples of *E. coli* harbouring pET28b-ptetO::*sod_gfpv2* expression vector.

Additional file 2. JGI whole-genome sequencing report of *S. plymuthica* WS3236.

Abbreviations

BGC: biosynthetic gene cluster; CTAB: cetyl trimethylammonium bromide; DiPaC: direct pathway cloning; DXP: 1-deoxy-D-xylulose-5-phosphate; EDTA: ethylenediaminetetraacetic acid; *E. coli*: *Escherichia coli*; FPP: farnesyl pyrophosphate; GC: gas chromatography; GC–MS: gas chromatography-mass spectrometry; GFP: green fluorescent protein; IPP: isopentenyl-diphosphate; LB: lysogeny broth; MEP: non-mevalonate pathway; NMR: nuclear magnetic resonance; PCR: polymerase chain reaction; PTFE: polytetrafluoroethylene; SAM: S-adenosyl methionine; *S. plymuthica*: *Serratia plymuthica*; succDMM: modified Davis and Mingioli minimal medium containing 55 mM succinate; TB: terrific broth; TE: Tris-EDTA; Tris: tris(hydroxymethyl)aminomethane; VOC: volatile organic compound; WGS: whole genome sequencing.

Authors' contributions

ERD performed the cloning and expression experiments. PMD performed genomic DNA extraction and bioinformatic analysis. NS and TW performed genome sequencing. TMF contributed to strain selection and provided the *S. plymuthica* WS3236 strain. ERD, PMD and TAMG designed the overall project and wrote the manuscript. All authors edited the manuscript. All authors read and approved the final manuscript.

Author details

¹ Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany. ² Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. ³ ZIEL Institute for Food & Health, Lehrstuhl für Mikrobielle Ökologie, Department biowissenschaftliche Grundlagen, Technical University of Munich, Munich, Germany. ⁴ Friedrich-Loeffler-Institut, Institut für Molekulare Pathogenese, Jena, Germany. ⁵ Chair of Technical Biochemistry, Technische Universität Dresden, Bergstraße 66, 01602 Dresden, Germany.

Acknowledgements

We thank Andreas Tröster, Christoph Brenninger, and Simone Stegbauer (group of Prof. Dr. Thorsten Bach, TU Munich) for technical assistance with the GC and GC-MS analysis.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional information files or are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

The work conducted as part of JGI Proposal #503161 by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. E.R.D. thanks the Deutsche Bundesstiftung Umwelt (DBU) for her PhD scholarship and CIPSM-Women for funding. P.M.D thanks the TUM University Foundation Fellowship and the Marie Skłodowska-Curie Actions Individual Fellowship (Project ID: 745435) for funding. We thank the DFG for generous financial support of the work in our laboratory (Emmy Noether program (GU 1233/1-1) and Center for Integrated Protein Science Munich CIPSM). This work was supported by the German Research Foundation (DFG) and the Technical University of Munich (TUM) in the framework of the Open Access Publishing Program.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 20 November 2018 Accepted: 30 January 2019

Published online: 07 February 2019

References

- Effmert U, Kalderas J, Warnke R, Piechulla B. Volatile mediated interactions between bacteria and fungi in the soil. *J Chem Ecol*. 2012;38:665–703.
- Wenke K, Kai M, Piechulla B. Belowground volatiles facilitate interactions between plant roots and soil organisms. *Planta*. 2010;231:499–506.
- Wenke K, Weise T, Warnke R, Valverde C, Wanke D, Kai M, Piechulla B. Bacterial volatiles mediating information between bacteria and plants. In: Witzany G, Baluška F, editors. *Biocommunication of plants*. Berlin: Springer; 2012. p. 327–47.
- Kai M, Crespo E, Cristescu SM, Harren FJ, Francke W, Piechulla B. *Serratia odorifera*: analysis of volatile emission and biological impact of volatile compounds on *Arabidopsis thaliana*. *Appl Microbiol Biotechnol*. 2010;88:965–76.
- Vespermann A, Kai M, Piechulla B. Rhizobacterial volatiles affect the growth of fungi and *Arabidopsis thaliana*. *Appl Environ Microbiol*. 2007;73:5639–41.
- Kai M, Effmert U, Berg G, Piechulla B. Volatiles of bacterial antagonists inhibit mycelial growth of the plant pathogen *Rhizoctonia solani*. *Arch Microbiol*. 2007;187:351–60.
- Kai M, Vespermann A, Piechulla B. The growth of fungi and *Arabidopsis thaliana* is influenced by bacterial volatiles. *Plant Signal Behav*. 2008;3:482–4.
- Kai M, Piechulla B. Plant growth promotion due to rhizobacterial volatiles—an effect of CO₂? *FEBS Lett*. 2009;583:3473–7.
- Kai M, Piechulla B. Impact of volatiles of the rhizobacteria *Serratia odorifera* on the moss *Physcomitrella patens*. *Plant Signal Behav*. 2010;5:444–6.
- von Reuss SH, Kai M, Piechulla B, Francke W. Octamethylbicyclo[3.2.1]octadienes from the rhizobacterium *Serratia odorifera*. *Angew Chem Int Ed Engl*. 2010;49:2009–10.
- Domik D, Magnus N, Piechulla B. Analysis of a new cluster of genes involved in the synthesis of the unique volatile organic compound sodorifen of *Serratia plymuthica* 4Rx13. *FEMS Microbiol Lett*. 2016;363(14):fnw139.
- Schmidt R, Jager V, Zuhlke D, Wolff C, Bernhardt J, Cankar K, Beekwilder J, Ijcken WV, Sleutels F, Boer W, et al. Fungal volatile compounds induce production of the secondary metabolite Sodorifen in *Serratia plymuthica* PRI-2C. *Sci Rep*. 2017;7:862.
- von Reuss S, Domik D, Lemfack MC, Magnus N, Kai M, Weise T, Piechulla B. Sodorifen biosynthesis in the rhizobacterium *Serratia plymuthica* involves methylation and cyclization of MEP-derived farnesyl pyrophosphate by a SAM-dependent C-methyltransferase. *J Am Chem Soc*. 2018;140:11855–62.
- Magnus N, Weise T, Piechulla B. Carbon catabolite repression regulates the production of the unique volatile sodorifen of *Serratia plymuthica* 4Rx13. *Front Microbiol*. 2017;8:2522.
- Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Müller R, Stewart AF, Zhang Y. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol*. 2012;30:440–6.
- Wang H, Bian X, Xia L, Ding X, Müller R, Zhang Y, Fu J, Stewart AF. Improved seamless mutagenesis by recombineering using ccdB for counterselection. *Nucleic Acids Res*. 2014;42:e37.
- Wang H, Li Z, Jia R, Yin J, Li A, Xia L, Yin Y, Müller R, Fu J, Stewart AF, Zhang Y. ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res*. 2018;46:e28.
- Jiang W, Zhu TF. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc*. 2016;11:960–75.
- Wang H, Li Z, Jia R, Hou Y, Yin J, Bian X, Li A, Müller R, Stewart AF, Fu J, Zhang Y. RecET direct cloning and Redα recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat Protoc*. 2016;11:1175–90.
- Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci USA*. 2014;111:1957–62.
- Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers*. 2010;93:833–44.
- Kouprina N, Larionov V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat Protoc*. 2008;3:371–7.
- Quan J, Tian J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat Protoc*. 2011;6:242–51.
- Zeng F, Zang J, Zhang S, Hao Z, Dong J, Lin Y. AFEAP cloning: a precise and efficient method for large DNA sequence assembly. *BMC Biotechnol*. 2017;17:81.

25. Greunke C, Duell ER, D'Agostino PM, Glöckle A, Lamm K, Gulder TAM. Direct pathway cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab Eng*. 2018;47:334–45.
26. D'Agostino PM, Gulder TAM. Direct pathway cloning combined with sequence- and ligation-independent cloning for fast biosynthetic gene cluster refactoring and heterologous expression. *ACS Synth Biol*. 2018;7:1702–8.
27. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res*. 2017;45:W36–41.
28. Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*. 2014;158:412–21.
29. Masschelein J, Mattheus W, Gao LJ, Moons P, Van Houdt R, Uytterhoeven B, Lamberigts C, Lescrinier E, Rozenski J, Herdewijn P, et al. A PKS/NRPS/FAS hybrid gene cluster from *Serratia plymuthica* RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. *PLoS ONE*. 2013;8:e54143.
30. Ongley SE, Bian X, Zhang Y, Chau R, Gerwick WH, Müller R, Neilan BA. High-titer heterologous production in *E. coli* of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. *ACS Chem Biol*. 2013;8:1888–93.
31. Elgaali H, Hamilton-Kemp TR, Newman MC, Collins RW, Yu K, Archbold DD. Comparison of long-chain alcohols and other volatile compounds emitted from food-borne and related Gram positive and Gram negative bacteria. *J Basic Microbiol*. 2002;42:373–80.
32. Yu K, Hamilton-Kemp TR, Archbold DD, Collins RW, Newman MC. Volatile compounds from *Escherichia coli* O157:H7 and their absorption by strawberry fruit. *J Agric Food Chem*. 2000;48:413–7.
33. Liu T, Mazmouz R, Ongley SE, Chau R, Pickford R, Woodhouse JN, Neilan BA. Directing the heterologous production of specific cyanobacterial toxin variants. *ACS Chem Biol*. 2017;12:2021–9.
34. Li P, Kwok AH, Jiang J, Ran T, Xu D, Wang W, Leung FC. Comparative genome analyses of *Serratia marcescens* FS14 reveals its high antagonistic potential. *PLoS ONE*. 2015;10:e0123061.
35. Micallef ML, Sharma D, Bunn BM, Gerwick L, Viswanathan R, Moffitt MC. Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. *BMC Microbiol*. 2014;14:213.
36. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323:133–8.
37. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:563–9.
38. Chen IMA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 2018;47:D666–77.
39. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol*. 2010;10:8.
40. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 2008;36:W465–9.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

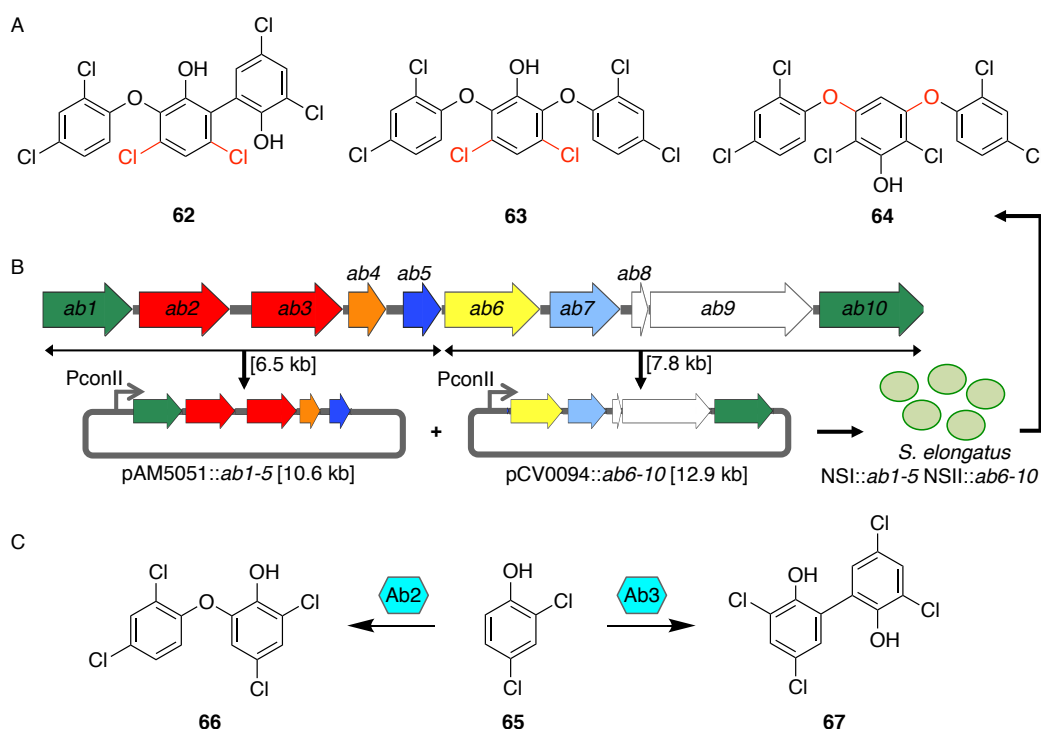


3.3 Investigations into ambigol biosynthesis

Based on: E. R. Duell*, T. M. Milzarek*, T. F. Schäberle, G. M. König, T. A. M. Gulder: Investigations into ambigol biosynthesis. *submitted*

*equally contributing authors

The ambigols (**62-64**) are polychlorinated triphenols isolated from the terrestrial cyanobacterium *Fischerella ambigua* 108b [263, 267]. Several of their properties, including their interesting antimicrobial, antiviral and cytotoxic activities [265, 267], their structural homology to the bioaccumulative and highly toxic marine PBDEs (**49**, **50**) [204, 205, 290] and their unusual bond connectivities (see Scheme 6A) justify their further investigation. The 14.6 kb *ab* BGC putatively responsible for the ambigol biosynthesis has initially been proposed in 2011 [264] and was now bioinformatically and experimentally re-evaluated within this thesis.



Scheme 6. Experimental strategies to investigate the ambigol biosynthesis. A) Ambigol A (**62**) and B (**63**) each contain two chlorine atoms and ambigol C (**64**) two biaryl-ether-bridges in relative *meta* position on the central phenol unit. B) By splitting the 14.6 kb *ab* BGC into two parts, sequential integration of the full cluster into *S. elongatus* was successful and led to the production of one ambigol derivative. C) *In vitro* and *in vivo* assays with 2,4-DCP (**65**) and the two *ab* CYP P450 enzymes revealed Ab2 to catalyze a biaryl-ether (**66**) and Ab3 to form a biaryl (**67**).

3.3 Investigations into ambigol biosynthesis

The cluster was first integrated into *E. coli* BAP1 for heterologous expression, but neither **62-64** nor their putative precursors could be detected within the culture supernatant or the cells. Therefore, an expression system more closely related to the original strain, the cyanobacterium *Synechococcus elongatus* PCC 7942, was chosen for further expression tests. The *ab* cluster was sequentially integrated in two parts into the neutral sites NSI and NSII of the cyanobacterial host chromosome (see Scheme 6B). The first part included *ab1-5* and had a size of 6.5 kb, whereas the second part consisted of *ab6-10* and was 7.8 kb long. Extracts of a five day expression culture, supplied with 100 mg/L 4-hydroxybenzoic acid (4-HBA) per day, contained one possible ambigol intermediate, later identified to be 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**66**), the *ortho* C-O-C bridged 2,4-DCP (**65**) dimer, as well as at least one ambigol derivative, whose structure remains elusive so far. To our knowledge, neither the insertion of (in total) such large DNA sequences into the *S. elongatus* genome, nor the construction of double mutants containing a complex secondary metabolite BGC and its successful expression has previously been reported.

Using **65** as substrate, the catalytic reaction scope of the two cytochrome P450 enzymes present in the *ab* cluster was further investigated *in vitro* and *in vivo*. Ab2 synthesized the biaryl-ether linked **66** coupled in relative *ortho* position, which was already found in the heterologous expression experiments (see Scheme 6C). Ab3 by contrast not only forms the *ortho*-C-C bridged 3,3',5,5'-tetrachloro-(1,1'-biphenyl)-2,2'-diol (**67**), but also two structurally not yet characterized ambigol derivatives missing one chlorine atom each. As the yields of both *in vitro* and *in vivo* assays could not be sufficiently increased to fully elucidate the structures of **66** and **67**, they were resolved by comparison to chemically synthesized standards.

Investigations into Ambigol Biosynthesis

Elke R. Duell,^{a†} Tobias M. Milzarek,^{ab†} Mustafa El Omari,^c Luis J. Linares-Otoya,^{de} Till F. Schäberle,^{de} Gabrielle M. König^c and Tobias A. M. Gulder^{ab*}

The terrestrial cyanobacterium *Fischerella ambigua* 108b produces the three polychlorinated triphenyls ambigol A–C that exhibit interesting antimicrobial, antiviral and cytotoxic activities. They show a strong structural homology to polybrominated diphenylethers synthesized by diverse marine bacteria that are known to be highly toxic and are bioaccumulating in natural food webs. All ambigols display unusual connectivities: Ambigol A and B exhibit chlorination and ambigol C biaryl-ether bonds in relative *meta* position on the central phenol unit, which is flanked by two 2,4-dichlorophenol units in all three compounds. Here we report on the identification of the biosynthetic gene cluster (BGC) responsible for ambigol production in *F. ambigua*. After bioinformatic discovery of a putative 10 genes containing BGC, we subsequently cloned and heterologously expressed the *ab* cluster in *Synechococcus elongatus* PCC 7942 using Direct Pathway Cloning (DiPaC). *In vivo* and *in vitro* characterization of the two cytochrome P450 enzymes present in the *ab* BGC revealed complementary selectivity for either biaryl-ether bond formation for Ab2 or biaryl formation for Ab3, when exposed to the possible common ambigol precursor 2,4-dichlorophenol.

Introduction

Polyhalogenated aromatic compounds can meanwhile be found almost everywhere in Nature and are often associated with negative biological effects on human and animal life, especially as they are known to be highly persistent and bioaccumulative in natural food webs. Many extremely toxic representatives of this class of molecules have an anthropogenic source and their incautious distribution in the environment is clearly caused by humankind. Amongst those substances are the polychlorinated biphenyls (PCBs), formally used, i.e., as industrial coolants and plasticizers, and the polybrominated diphenylethers (PBDEs), utilized as flame retardants. All these compounds are now restricted from commercial use due to their toxicity and aggregation in the environment. Polyhalogenated dibenzo-*p*-dioxines (PHDDs), e.g., 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), which is a side product in the synthesis of herbicides such as 2,4,5-trichlorophenoxyacetic acid (trioxone), is one of the most toxic and environmentally persistent man-made pollutants known today.^{1–3} The structurally related substance

triclosan, which was developed as disinfectant, can still be found in toothpastes, soaps and detergents. However, recent studies suggest it to be very likely toxic for animals and humans,^{4–6} especially as the formation of PHDDs from triclosan via photodegradation has been observed.^{7,8}

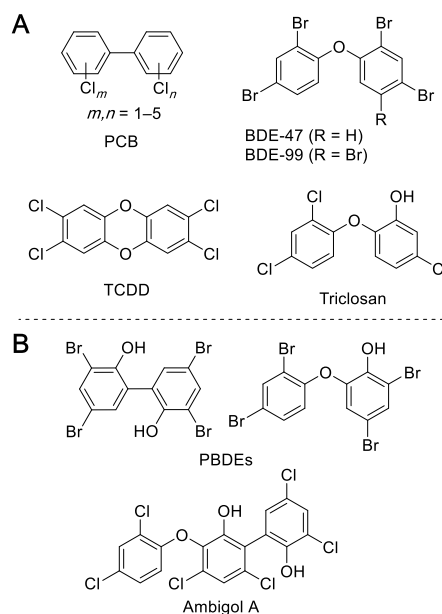


Fig. 1 Structures of selected A) synthetic and B) naturally occurring polyhalogenated aromatic compounds.

However, not all harmful polyhalogenated aromatic substances are of anthropogenic origin. A wide range of PBDEs has been

^a Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany. E-mail: tobias.gulder@ch.tum.de

^b Chair of Technical Biochemistry, Technische Universität Dresden, Bergstraße 66, 01062 Dresden, Germany. E-mail: tobias.gulder@tu-dresden.de

^c Institute for Pharmaceutical Biology, University of Bonn, Nußallee 6, 53115 Bonn, Germany.

^d Institute for Insect Biotechnology, Justus Liebig University of Giessen, Heinrich-Buff-Ring 26–32, 35392 Giessen, Germany.

^e Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Giessen, Germany.

[†] These authors contributed equally to this work.

[†] Electronic Supplementary Information (ESI) available: Containing all experimental data. See DOI: 10.1039/x0xx00000x

isolated from marine invertebrates^{9–11} and only recently it was shown that they are produced by a variety of marine bacteria such as γ -proteobacteria of the genus *Pseudoalteromonas* and the cyanobacterial *Hormoscilla spongelliae* spp.^{12,13} These naturally occurring PBDEs, including their hydroxylated (OH-BDEs) and dioxine-like derivatives, can be found in all trophic levels of marine life and are accumulating in the food chain via marine plants¹⁴ and algae^{15–17} and are ultimately detectable in marine animals^{18–20} and even human tissue.^{21–24} Similar to their man-made relatives, they exhibit toxic effects on mammalian hormone mediated signalling pathways and essential enzymatic reactions.^{24–28} The biosynthesis of these PBDEs of natural origin has been extensively studied by pioneering work of the Moore laboratory over the last years, leading to the discovery of the encoding biosynthetic gene cluster (BGC) *bmp*, which can be found in varying genetic compositions in all bacterial strains associated with PBDE synthesis.^{12,13} The biosynthesis of PBDEs utilizes the primary metabolite chorismic acid (**1**), which is converted to 4-hydroxybenzoic acid (**2**) by the chorismate lyase Bmp6 (cf. Fig. 4).^{29,30} A so far unprecedented decarboxylative phenol bromination mechanism, catalyzed by the halogenase Bmp5, uses **2** as a substrate to form 2,4-dibromophenol (2,4-DBP).¹² The cytochrome P450 enzyme Bmp7 is then able to couple two 2,4-DBP units to form a variety of different dimeric PBDEs.^{12,13,30} These biaryl- and biaryl-ether-coupling reactions between aromatic rings follow the common principle of phenol oxidative coupling, where an initial hydroxyl radical is formed by abstraction of one electron and one proton (Fig. 2A).³¹ Due to mesomeric delocalization of the free electron on the benzene ring in relative *ortho* and *para* position to the oxygen functionality, a large diversity of C–C and C–O–C bond containing coupling products can be biosynthesized.³²

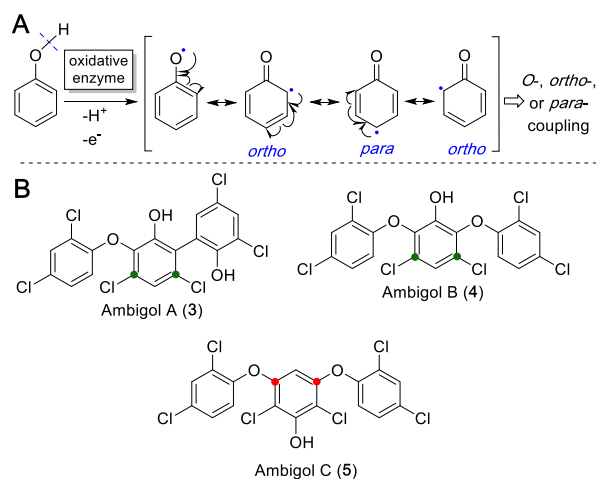


Fig. 2 A) The radical mechanism underlying oxidative coupling of phenols allows reactions only at the phenol oxygen or in relative *ortho* and *para* position to this activating group. The identical positions are activated for nucleophilic halogenation by flavin-dependent halogenases. B) Ambigol A (**3**) and B (**4**) show chlorination at *meta* position (green dots), ambigol C (**5**) contains two biaryl-ether-bonds in *meta* position (red dots).

Not only marine but also terrestrial cyanobacteria are known to produce polyhalogenated phenols. Due to its abundance in soil, chlorine is generally favoured over bromine for incorporation into halogenated natural products of terrestrial origin. In 1993

and in 2005, the three polychlorinated ambigols A–C (**3–5**, see Fig. 1B and 2B) were discovered from the soil-dwelling filamentous cyanobacterium *Fischerella ambigua* 108b. Ambigol A (**3**) exhibits strong antimicrobial, antifungal and virostatic activities and moderate cytotoxicity whereas ambigol C (**5**) exclusively shows good activity against gram positive bacteria and fungi and ambigol B (**4**) generally displays only weak activity in biological assays.^{33–35} At first glance, the ambigols **3–5** show a strong structural similarity to the marine PBDEs as they also contain two halogen atoms per phenol unit and are linked either via biaryl- or biaryl-ether-bonds. Unlike the PBDEs, however, **3–5** incorporate a third dihalophenol unit. In addition, either the chlorine substituents (in **3** and **4**) or the biaryl ether structural motif (in **5**) are not located at the typical relative *ortho* or *para* position to the activating phenol function in the central aromatic portion, but rather situated in *meta* position (Fig. 2B). Owing to the interesting biological activities of the ambigol natural product family and the described structural peculiarities pointing at unusual biosynthetic transformations, we investigated the biosynthesis of the ambigol natural product family within this work.

Results and Discussion

Bioinformatic analysis of the *ab* BGC

To identify the ambigol BGC, the natural producer *Fischerella ambigua* 108b was submitted to 454 genome sequencing (see Supporting Information). This allowed for the identification of a cosmid clone harboring a 14.3 kb BGC containing 10 genes (Table 1), among them a number of genes predictable to be present in a putative ambigol biosynthetic pathway *ab* (Fig. 3). Besides genes encoding halogenating enzymes, additional similarities to the gene clusters encoding the structurally related PBDEs were likely to be identified in the ambigol cluster. Indeed, *ab* contains a chorismate lyase *ab5* that is highly homologous to *bmp6* found in *H. spongelliae*.¹³ In analogy to the *bmp* pathway, *ab5* thus likely recruits building blocks from the pool of primary metabolites for ambigol biosynthesis by formation of 4-HBA (**2**) from chorismic acid (**1**). Furthermore, *ab* contains a copy of a 3-deoxy-7-phosphoheptulonate (DAHP) synthase (*ab7*), an enzyme catalyzing the first dedicated step in the shikimic acid biosynthetic pathway towards **1** (see Fig. 4).

The entire *ab* BGC is framed by two genes encoding FAD-dependent halogenases, *ab1* and *ab10*. As in PBDE biosynthesis, **2** has to be chlorinated twice to give 2,4-dichlorophenol (**6**) as monomeric building block found in all ambigol structures. Formation of **2** in case of PBDE biosynthesis was shown to be catalyzed by a single halogenase, Bmp5, which not only performs an *ortho* bromination, but also a highly unusual second decarboxylative halogenation, thus replacing the acid function in *para* position to the phenol in **2** by a bromine substituent.¹² In contrast, the *ab* BGC contains the two FAD-dependent halogenases *ab1* and *ab10*. The underlying mechanistically different halogenation reactions towards **6** therefore might require two separate halogenases in the case of ambigol biosynthesis. Interestingly, neither *ab1* nor *ab10* show a particularly strong homology to *bmp5*. Alternatively,

one of the halogenases might be involved in the assembly of 2,6-dichlorophenol (**7**) or 3,5-dichlorophenol (**8**), as potential further required central ambigol biosynthetic building blocks (Fig. 4). Three additional genes located in *ab*, i.e., *ab6*, *ab8*, and *ab9*, resemble elements of PKS/NPRS biosynthesis for precursor

function (cf. Fig. 2A). The *bmp* BGC in γ -proteobacteria only contains a single cytochrome P450 encoding gene, *bmp7*. The corresponding protein catalyzes the required biaryl bond formations between two units of 2,4-dibromophenol.¹³ In contrast, the *ab* cluster harbours two genes encoding the

Table 1 Genetic composition of the *ab* BGC.

Gene	Size (bp)	Proposed function	Closest xBLAST hit	Identity (protein)	GenBank acc. number
<i>ab1</i>	1452	Precursor chlorination	NAD(P)/FAD-dependent oxidoreductase [<i>Scytonema hofmannii</i>]	281/486 (58%)	WP_017744817.1
<i>ab2</i>	1464	Biaryl coupling	Cytochrome P450 [<i>Hormoscilla spongelliae</i> GUM020]	337/485 (69%)	AQU14168.1
<i>ab3</i>	1476	Biaryl coupling	Cytochrome P450 [<i>Hormoscilla spongelliae</i> GUM020]	301/474 (64%)	AQU14168.1
<i>ab4</i>	615	Unknown	Cyclase [<i>Nostoc punctiforme</i> NIES-2108]	123/180 (68%)	RCJ31428.1
<i>ab5</i>	624	Synthesis of 1 from 2	Chorismate lyase [<i>Hormoscilla spongelliae</i> GUM096]	106/177 (60%)	AQU14179.1
<i>ab6</i>	1536	Acyl activation	Acyl-CoA synthetase [filamentous cyanobacterium CCP3]	410/511 (80%)	WP_106922238.1
<i>ab7</i>	1131	Synthesis of DAHP	3-Deoxy-7-phosphoheptulonate synthase [<i>Nostoc minutum</i> NIES-26]	296/374 (79%)	RCJ32495.1
<i>ab8</i>	273	PKS (ACP)	Acyl carrier protein [<i>Chroocloeocystis siderophyllia</i>]	49/81 (60%)	WP_073551326.1
<i>ab9</i>	2631	NRPS (C-PCP-TE)	non-ribosomal peptide synthetase [<i>Moorea producens</i>]	519/870 (60%)	WP_008188544.1
<i>ab10</i>	1725	Precursor chlorination	FAD-dependent oxidoreductase [<i>Scytonema</i> sp. HK-05]	430/573 (75%)	WP_073630058.1

activation (*ab6*) and covalent tethering (*ab8*, *ab9*). The corresponding proteins might thus likewise be involved in enzymatic halogenation, as PCP-tethering is known to be required by a number of FAD-dependent halogenases.^{36–39}

However, despite considerable attempts to obtain soluble halogenases *Ab1* and *Ab10* (data not shown), we did not yet

successfully obtain soluble halogenases *Ab1* and *Ab10* (data not shown), we did not yet closely related CYP P450 homologs *Ab2* and *Ab3*. It was thus tempting to speculate that one of these enzymes catalyzes the usual *ortho/para* site selective biaryl coupling reactions, while the second homolog might be responsible for a formal coupling in *meta* position. The latter might be feasible when following a hypothesized biaryl coupling mechanism involving a 1,2-hydroxyl shift in the substrate during the radical coupling

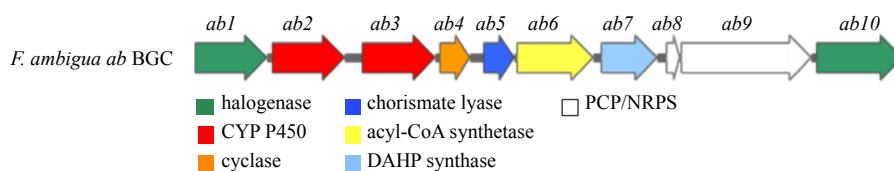


Fig. 3 Schematic illustration of the genetic organization within the *ab* BGC containing ten genes.

succeed in recombinant production of these enzymes in a soluble form and thus could not further probe their functions. In case of all PBDEs found so far, cross-coupling of the phenolic building blocks typically occurs in the mechanistically permitted *ortho* and *para* positions relative to an activating phenol

procedure (see box in Fig. 4), additionally having the charm of only requiring 2,4-dichlorophenol (**6**) as the sole biosynthetic precursor for the biosynthesis of **3** and **4**. We thus set out to firmly link the putative ambigol BGC to ambigol production and

to characterize the function of the cytochrome P450 enzymes Ab2 and Ab3.

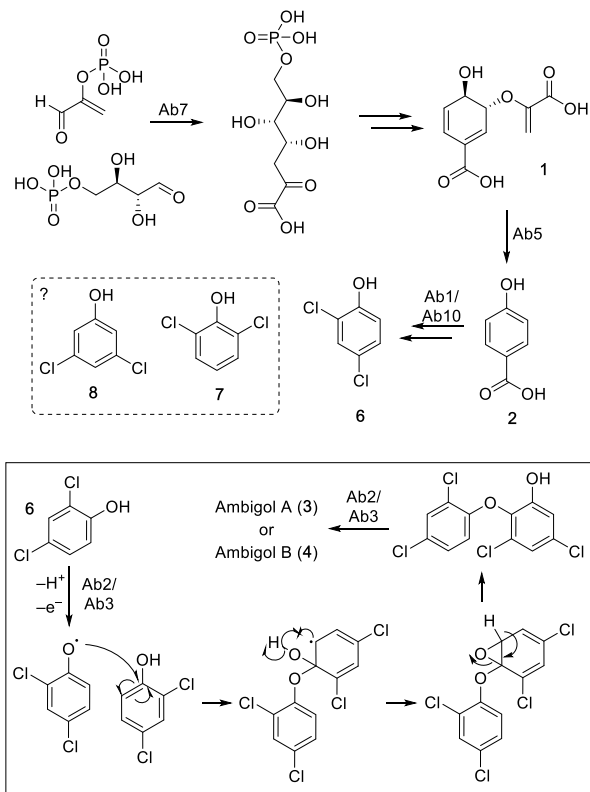


Fig. 4 Biosynthetic route to the ambigol precursor **6**. Box: hypothetical mechanism to explain the observed *meta*-halogenation pattern in **3** and **4** resulting from a biaryl cross-coupling reaction that involves a 1,2-hydroxyl shift.

Cloning and expression of the *ab* BGC in *E. coli* and *S. elongatus*

To unambiguously link the putative ambigol BGC *ab* to ambigol biosynthesis, we first aimed at recombinant production of the target molecules. Mutation free cloning of the whole *ab* cluster into pET-28b-SUMO as well as subsequent addition of a *gfp* gene downstream of the *ab* BGC as reporter for successful transcription of the entire pathway⁴⁰ was achieved using our DiPaC strategy.⁴¹⁻⁴³ Initial expression tests with the resulting construct *E. coli* BAP1 harbouring pET-28b-SUMO::*ab-gfp* did not show production of any new compounds when compared to empty cells as controls, nor of Gfp-mediated fluorescence. The T7 promoter was therefore exchanged by ptetO, as this promoter is known to efficiently upregulate larger transcripts.⁴⁴ While the resulting expression system pET28b-ptetO::*ab-gfp* facilitated the expression of Gfp as indicated by significant fluorescence of the recombinant *E. coli* BAP1 host cells (see Supporting Information Fig. S5), thus clearly also confirming complete transcription of all *ab* genes, again no production of ambigols or potential intermediates was observable. The reason for this is likely the inability of *E. coli* to translate and/or fold one or more enzymes of the *ab* BGC into their catalytically active forms. Related problems were observed by Moore *et al.*

when expressing selected genes of the *bmp* BGC of cyanobacterial origin,¹³ while the *bmp* genes from γ -proteobacteria, by contrast, were expressible in *E. coli*.^{12,30,39,45} We thus focused on the recombinant production of ambigols in a phylogenetically more closely related species when compared to *F. ambigua*, i.e., the cyanobacterial heterologous host *Synechococcus elongatus* PCC 7942^{46,47} previously utilized for the production of PBDEs¹³ and industrially important compounds such as butyrate⁴⁸, 2-deoxy-scyllo-inose⁴⁹ and propanediol.^{49,50} Due to the size of the *ab* BGC and the known difficulties in introducing large DNA fragments into *S. elongatus* by natural transformation, we decided to split the *ab* BGC in two halves of approx. 7–8 kb and to sequentially integrate these parts into two independent genomic sites of *S. elongatus* PCC 7942.

Mutation free integration of *ab1–5* into pAM5051⁴⁷ and *ab6–10* into pCV0094¹³ as well as the integration of *ab6–10* into the chromosomal site NSII of *S. elongatus* by natural transformation was successful (see Supporting Information Fig. S3). Generation of double transformants additionally carrying *ab1–5* in all copies of NSI was subsequently verified by PCR (see Supporting Information Fig. S3). To the best of our knowledge, this is the first introduction of a >10 kb BGC into the *S. elongatus* host system combined with the successful heterologous production of the encoded metabolites: production of one dichlorophenol-dimer with m/z 323.1 [M–H][–], later identified to be 2,4-dichloro-6-(2,4-dichlorophenoxy)-phenol (see Fig. 5 and Supporting Information Fig. S7), and one trimeric molecule with m/z 482.7 [M–H][–] corresponding to the ambigol molecular mass was observed. In addition, two dichlorophenol-trimers each lacking

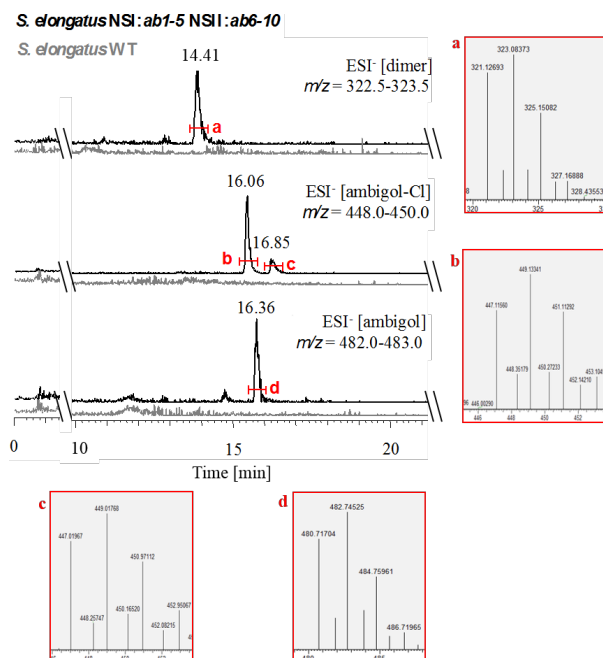


Fig. 5 LC-MS analysis of pellet extracts of *S. elongatus* WT (grey) and *S. elongatus* carrying *ab1–5* in NSI and *ab6–10* in NSII (black). The production of one DCP-dimer, two trimers–Cl and one ambigol trimer showing the characteristic Cl-isotopic pattern (see a–d) can only be observed in the presence of the *ab* cluster.

one Cl-atom with m/z 449.1 $[M-H]^-$ were produced. As no dichlorophenol monomer was detectable, we assume that this intermediate is rapidly consumed by the downstream enzymes to give the observed dimeric and trimeric products. Each of the observed products displayed the expected isotopic pattern of multi-chlorinated substances. The corresponding signals were not present in wildtype *S. elongatus* control experiments, thus firmly linking ambigol production to the identified *ab* BGC. Unfortunately, no stable production system could be established, likely due to the metabolic stress exerted on the recombinant *S. elongatus* host system resulting from integration of two genetic elements requiring cultivation of the highly sensitive host system in the presence of multiple antibiotics. In general, the double mutant *S. elongatus* cells took significantly longer to grow compared to wild type or single mutant cells and were visibly less vital.

Cloning, expression and purification of the CYP P450 enzymes Ab2 and Ab3

We next turned our attention to the *in vivo* and *in vitro* characterization of the enzymes Ab2 and Ab3 putatively involved in biaryl formation. For the *in vivo* experiments, we

again utilized *S. elongatus* as well as *E. coli* BL21 as host organisms. Mutation free integration of the native sequences of *ab2* and/or *ab3* into pCV0092 was confirmed by sequencing. Successful natural transformation of *S. elongatus* with these constructs was verified by screening PCR (see Supporting Information Fig. S4). As this heterologous system is not capable of producing high amounts of proteins after induction, the transformed cells were only used for *in vivo* feeding experiments. For *in vitro* work, *E. coli* codon optimized versions of *ab2* and *ab3* were successfully integrated into the pMal expression plasmid containing an *N*-terminal MBP-tag. Heterologous overexpression in *E. coli* BL21 and purification via MBP-tag based affinity purification led to soluble protein, resulting in a slightly yellow enzyme solution (see Supporting Information Fig. S6). Before characterizing the enzymatic transformation catalyzed by these CYPs, we set out to synthesize all feasible C–C and C–O–C coupling dimers as analytical standards to allow for an unambiguous structural assignment, as very similar chromatographic and spectroscopic properties were to be expected for these structurally strongly related compounds.

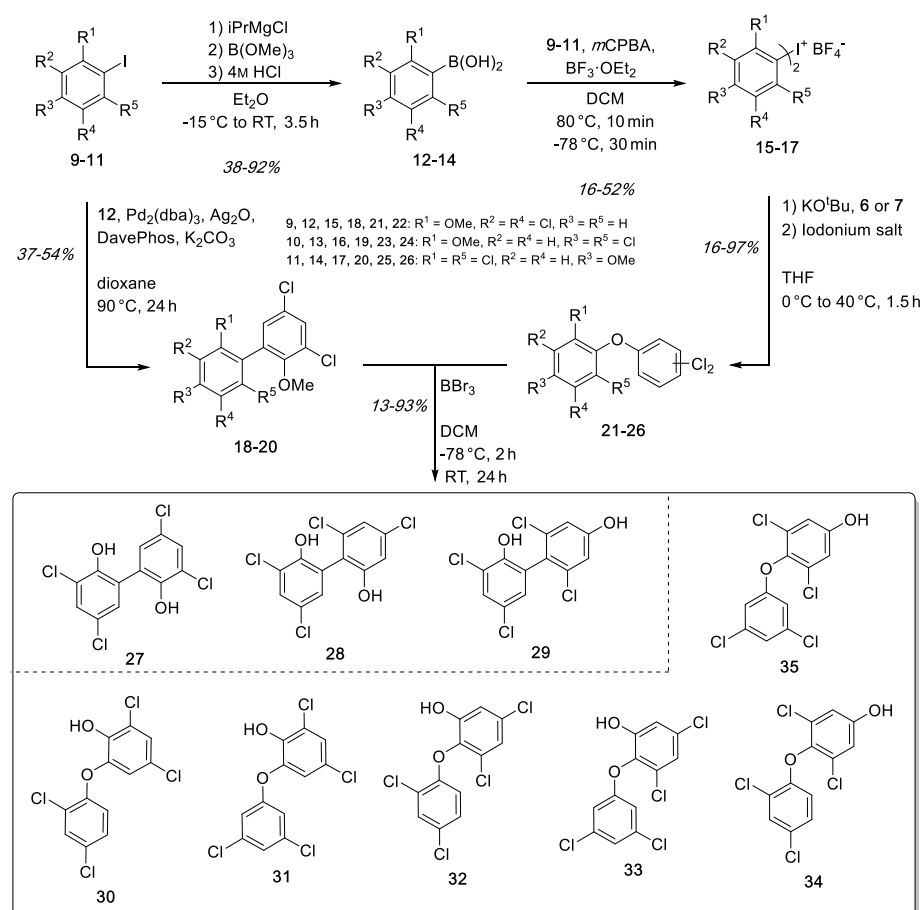


Fig. 6 Developed synthetic route for the formation of C–C coupled dimers 27–29 and of C–O–C coupled dimers 30–35.

Chemical synthesis of standards and functional characterization of the CYP P450 enzymes *in vitro* and *in vivo*

The synthesis of the biaryls **27–29** and biaryl-ethers **30–35** was performed by an improved palladium-catalyzed Suzuki cross-coupling reaction as well as a nucleophilic aromatic substitution (S_NAr) using hypervalent and symmetric iodine(III) salts (Fig. 6). The basic building blocks, 2,4-dichlorophenol (**6**) and 3,5-dichlorophenol (**8**) were transformed into methoxy-protected aryl iodides **9–11** (see Supporting Information), which were used to synthesize the respective boronic acids **12–14** through metal-halogen exchange in 38–92% yield. For the Suzuki coupling, these aryl iodides **9–11** and (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**) were combined to give the protected biaryls **18–20** in 37–54%. The symmetric iodonium salts were synthesized using the same aryl iodides **9–11** and the corresponding boronic acids **12–14** under oxidative conditions.⁵¹ The subsequent conversion in the S_NAr reaction led to the methoxy-protected biaryl-ethers **21–26** in yields of 16–97%.⁵² In order to obtain the final phenolic dimers **27–35**, the methoxy protecting groups in **21–26** were cleaved off using boron tribromide. The deprotection gave the expected products in 13–93% yield. It should be noted that the total yield of all executed reactions is highly depending on the sterical hindrance of the used building blocks. In summary, we have developed a short and fast synthetic route for the formation of both, C–C and C–O–C dimers, using functionalized aromatic building blocks.

LC-MS analysis of both *in vivo* (feeding experiments) and *in vitro* assays of Ab2 with the substrate (**6**) showed the predominant formation of one coupling product with m/z 323.0 [M–H]. NMR analysis of the isolated product suggested the production of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**) (Fig. 7). Comparison of HPLC retention times and ¹H-NMR spectra of biochemical produced dimer with the synthetic standards indeed confirmed the identity of **30** (see Supporting Information Fig. S7, S12 and S13). A formation of trimeric products was not observed.

Ab3 catalyzed the formation of one major product with identical mass but different retention time when compared to the Ab2 product in both *in vitro* and *in vivo* experiments (Fig. 7). Comparison with the synthetic standards revealed the dimer to be 3,3',5,5'-tetrachloro-[1,1'-biphenyl]-2,2'-diol (**27**) resulting from a biaryl-coupling of two units of **6** in relative *ortho* position to the phenol functions (see Supporting Information Fig. S9 and S14). Particularly in *in vitro* assays, Ab2 and Ab3 were found to also catalyze the formation of small amounts of the dimer characteristic of the respective other CYP P450 enzyme. In addition to these products, small amounts of two DCP trimers at m/z 449.1 [M–H][–] and m/z 449.0 [M–H][–], thus lacking one chlorine substituent, were likewise produced both *in vitro* and *in vivo* in the presence of Ab3 (see Supporting Information Fig. S11). The combination of both Ab2 and Ab3 in one assay did not lead to the formation of any additional, new products. *In vivo* and *in vitro* assays with 2,4-DBP instead of **6** showed that both Ab2 and Ab3 can accept and couple this substrate as well (data

not shown), following the same site selectivity. On the basis of these experimental results, we identified two cytochrome P450 enzymes that complementarily perform either biaryl- or biaryl-ether bond formation.

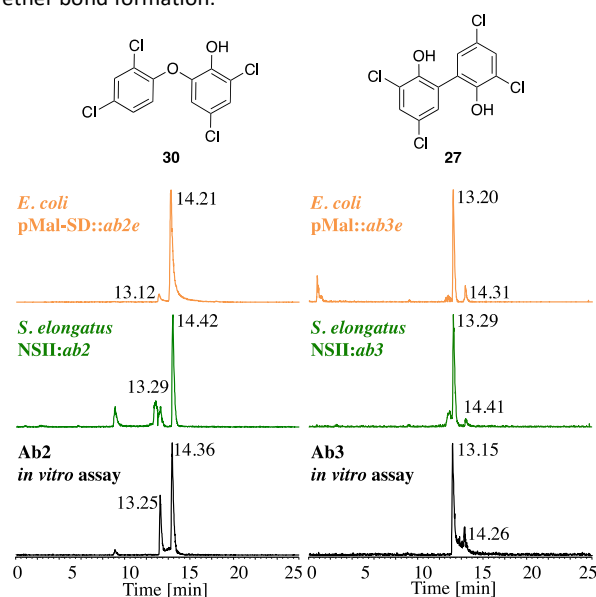


Fig. 7 *In vivo* and *in vitro* catalytic activity of Ab2 and Ab3 when incubated with **6**. Ab2 catalyzed the formation of mainly the C–O–C coupled dimer **30** whereas Ab3 produced the C–C coupled dimer **27**. EIC (ESI[–]) at m/z = 322.5–323.5.

Conclusions

Within this work, we set out to identify the biosynthetic machinery for the production of the polyhalogenated polyphenols ambigol A–C (**3–5**). A putative ambigol BGC *ab* was identified by sequencing of the natural producer of **3–5**, *F. ambigua* 108b. Using the model cyanobacterial heterologous expression system *S. elongatus*, we firmly linked ambigol production to the identified *ab* BGC containing ten genes. To our knowledge, this study is the first example for a successful construction of *S. elongatus* double mutants by integrating in total more than 14 kb foreign DNA into the two neutral sites NSI and NSII.

To shed light on the biaryl coupling biochemistry involved in ambigol biosynthesis, the function of the two cytochrome P450 enzymes, Ab2 and Ab3, encoded in *ab* were probed both *in vivo* (in *E. coli* and *S. elongatus*) and *in vitro* using purified recombinant protein produced in *E. coli*. All these experiments consistently revealed Ab2 to selectively form the *O,ortho*-coupled biaryl ether **30**, whereas Ab3 produces the *ortho,ortho*-linked, C–C-coupled biaryl **27**. The identity of these products was unambiguously established by comparison with all feasible coupling products that were prepared by synthesis. Given the structural simplicity of the joint precursor molecule 2,4-dichlorophenol (**6**), the exceptional site selectivity of Ab2 and Ab3 is highly remarkable and warrants future studies on the underlying structural basis of this selectivity in these biaryl

coupling enzymes. The origin of the unusual substitution pattern of the central building blocks in all ambigols, with relative *meta*-biaryl coupling or -halogenation, however, remains elusive. A key to explaining this phenomenon could be the mono-dehalogenated side-products observed within our full pathway reconstitution as well as in the *in vitro* experiments with Ab3. Due to the so far very low production titers of these compounds, their structural assignment required as a basis to interrogate their potential role in ambigol biosynthesis was not yet possible. Additional work targeting these compounds to further complement the overall picture of ambigol biosynthesis is currently conducted in our laboratories.

Acknowledgements

We thank Prof. Dr. James W. Golden and Dr. Arnaud Taton for sharing their *S. elongatus* PCC 7942 based expression system including the plasmids pAM5054 and pCV0094 and for their helpful advice. We furthermore thank Stefan Ernst and Prof. Dr. Aymelt Itzen for providing the pMal expression plasmid. We thank the group of Prof. Dr. Stephan A. Sieber (TU Munich) for measuring HRMS data. E.R.D. thanks the Deutsche Bundesstiftung Umwelt (DBU) for her Ph.D. scholarship and CIPSM-Women for funding. T.M.M. thanks the Stiftung der Deutschen Wirtschaft (sdw) for funding. We thank the Max Buchner Research Foundation and the DFG for generous financial support of the work in our laboratory (Emmy Noether program (GU 1233/1-1) and Center for Integrated Protein Science Munich CIPSM).

Notes and references

- R. A. Hites, *Environ. Sci. Technol.*, 2011, **45**, 16–20.
- M. Van den Berg, L. Birnbaum, A. T. Bosveld, B. Brunstorm, P. Cook, M. Feeley, J. P. Giesy, A. Hanberg, R. Hasegawa, S. W. Kennedy, T. Kubiak, J. C. Larsen, F. X. van Leeuwen, A. K. Liem, C. Nolt, R. E. Peterson, L. Poellinger, S. Safe, D. Schrenk, D. Tillitt, M. Tysklind, M. Younes, F. Waern and T. Zacharewski, *Environ. Health Perspect.*, 1998, **106**, 775–792.
- M. Van den Berg, L. S. Birnbaum, M. Denison, M. De Vito, W. Farland, M. Feeley, H. Fiedler, H. Hakansson, A. Hanberg, L. Haws, M. Rose, S. Safe, D. Schrenk, C. Tohyama, A. Tritscher, J. Tuomisto, M. Tysklind, N. Walker and R. E. Peterson, *Toxicol. Sci.*, 2006, **93**, 223–241.
- J. L. Wu, F. Ji, H. Zhang, C. Hu, M. H. Wong, D. Hu and Z. Cai, *J. Hazard. Mater.*, 2018, **367**, 128–136.
- H. Yang, W. Wang, K. A. Romano, M. Gu, K. Z. Sanidad, D. Kim, J. Yang, B. Schmidt, D. Panigraphy, R. Pei, D. A. Martin, E. I. Ozay, Y. Wang, M. Song, B. W. Bollinh, H. Xiao, L. M. Minter, G. Y. Yang, Z. Liu, F. E. Rey and G. Zhang, *Sci. Transl. Med.*, 2018, **10**.
- M. F. Yueh, K. Taniguchi, S. Chen, R. M. Evans, B. D. Hammock, M. Karin and R. H. Tokes, *Proc. Natl. Acad. Sci.*, 2014, **111**, 17200–17205.
- R. U. Halden, *Environ. Sci. Technol.*, 2014, **48**, 3603–3611.
- R. U. Halden, A. E. Lindeman, A. E. Aiello, D. Andrews, W. A. Arnold, P. Fair, R. E. Fuoco, L. A. Geer, P. I. Johnson, R. Lohmann, K. McNeill, V. P. Sacks, T. Schettler, R. Weber, R. T. Zoeller and A. Blum, *Environ. Health Perspect.*, 2017, **125**, 64501.
- M. D. Unson, N. D. Holland and D. J. Faulkner, *Mar. Bio.*, 1994, **119**, 1–11.
- X. Fu, F. J. Schmitz, M. Govindan, S. A. Abbas, K. M. Hanson, P. A. Horton, P. Crews, M. Laney and R. C. Schatzman, *J. Nat. Prod.*, 1995, **58**, 1384–1391.
- L. Calcul, R. Chow, A. G. Oliver, K. Tenney, K. N. White, A. W. Wood, C. Fiorilla and P. Crews, *J. Nat. Prod.*, 2009, **72**, 443–449.
- V. Agarwal, A. A. El Gamal, K. Yamanaka, D. Poth, R. D. Kersten, M. Schorn, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2014, **10**, 640–647.
- V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2017, **13**, 537–543.
- S. Gaul, P. Bending, D. Olbrich, N. Reosenfelder, P. Ruff, C. Gaus, J. F. Mueller and W. Vetter, *Mar. Pollut. Bull.*, 2011, **62**, 2463–2468.
- M. Kuniyoshi, K. Yamada and T. Higa, *Experientia*, 1985, **41**, 523–524.
- K. Makoto, K. Tomoyuki, N. Yoshikatsu and U. Daisuke, *Chem. Lett.*, 2005, **34**, 1272–1273.
- A. Malmcarn, Y. Zebuhr, L. Kautsky, K. Bergman and L. Asplund, *Chemosphere*, 2008, **72**, 910–916.
- W. Vetter, E. Stoll, M. J. Garson, S. J. Fahey, C. Gaus and J. F. Müller, *Environ. Toxicol. Chem.*, 2002, **21**, 2014–2019.
- E. L. Teuten, L. Xu and C. M. Reddy, *Science*, 2005, **307**, 917–920.
- N. J. Shaul, N. G. Dodder, L. I. Aluwihare, S. A. Mackintosh, K. A. Maruya, S. J. Chivers, K. Danil, D. W. Weller and E. Hoh, *Environ. Sci. Technol.*, 2015, **49**, 1328–1338.
- Y. Wan, K. Choi, S. Kim, K. Ji, H. Chang, S. Wiseman, P. D. Jones, J. S. Khim, S. Park, J. Park, M. H. Lam and J. P. Giesy, *Environ. Sci. Technol.*, 2010, **44**, 5233–5239.
- H. S. Wand, Z. J. Chen, K. L. Ho, L. C. Ge, J. Du, M. H. Lam, J. P. Giesy, M. H. Wong and C. K. Wong, *Environ. Int.*, 2012, **47**, 66–72.
- A. Chen, J. S. Park, L. Linderholm, A. Rhee, M. Petra, E. A. DeFranco, K. N. Dietrich and S. M. Ho, *Environ. Sci. Technol.*, 2013, **47**, 3902–3908.
- S. B. Wiseman, Y. Wan, H. Chang, X. Zhang, M. Hecker, P. D. Jones and J. P. Giesy, *Mar. Pollut. Bull.*, 2011, **63**, 179–188.
- E. N. Segreaves, R. R. Shah, N. L. Segreaves, T. A. Johnson, S. Whitman, J. K. Sui, V. A. Kenyon, R. H. Cichewicz, P. Crews and T. R. Holman, *J. Med. Chem.*, 2004, **47**, 4060–4065.
- R. F. Canton, D. E. Scholten, G. Marsh, P. C. de Jong and M. van den Berg, *Toxicol. Appl. Pharmacol.*, 2008, **227**, 68–75.
- J. A. de la Fuente, S. Manzanaro, M. J. Marin, T. G. de Quesada, I. Reymundo, S. M. Luengo and F. Gago, *J. Med. Chem.*, 2003, **46**, 5208–5221.
- J. Legradi, A. K. Dahlberg, P. Cenijn, G. Marsh, L. Asplund, A. Bergman and J. Legler, *Environ. Sci. Technol.*, 2014, **48**, 14703–14711.
- U. Hanefeld, H. G. Floss and H. Laatsch, *J. Org. Chem.*, 1994, **59**, 3604–3608.
- V. Agarwal and B. S. Moore, *ACS Chem. Biol.*, 2014, **9**, 1980–1984.
- H. Aldemir, R. Richarz and T. A. M. Gulder, *Angew. Chem. Int. Ed.*, 2014, **53**, 8286–8293.
- V. Agarwal, J. Li, I. Rahman, M. Borgen, L. I. Aluwihare, J. S. Biggs, V. J. Paul and B. S. Moore, *Environ. Sci. Technol.*, 2015, **49**, 1339–1346.
- B. S. Falch, G. M. König, A. D. Wright, O. Sticher, H. Ruegger and G. Bernardinelli, *J. Org. Chem.*, 1993, **58**, 6570–6575.
- B. S. Falch, G. M. König, A. D. Wright, O. Sticher, C. K. Angerhofer, J. M. Pezzuto and H. Bachmann, *Planta Med.*, 1995, **61**, 321–328.

- 35 A. D. Wright, O. Pappendorf and G. M. König, *J. Nat. Prod.*, 2005, **68**, 459–461.
- 36 P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh, *Proc. Natl. Acad. Sci.*, 2005, **102**, 13843–13848.
- 37 S. Lin, S. G. Van Lanen and B. Shen, *J. Am. Chem. Soc.*, 2007, **129**, 12432–12438.
- 38 T. Kittilä, C. Knittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K. H. van Pee, R. D. Süßmuth, W. Wohlleben, M. J. Cryle and E. Stegmann, *Chem. Sci.*, 2017, **8**, 5992–6004.
- 39 A. El Gamal, V. Agarwal, S. Diethelm, I. Rahman, M. A. Schorn, J. M. Sneed, G. V. Louie, K. L. Whalen, T. J. Mincer, J. P. Noel, V. J. Paul, B. S. Moore, *Proc. Natl. Acad. Sci.*, 2016, **113**, 3797–3802.
- 40 E. R. Duell, P. M. D’Agostino, N. Shapiro, T. Woyke, T. M. Fuchs and T. A. M. Gulder, *Microb. Cell Fact.*, 2019, **18**, 32.
- 41 C. Greunke, E. R. Duell, P. M. D’Agostino, A. Glöckle, K. Lamm and T. A. M. Gulder, *Metab. Eng.*, 2018, **47**, 334–345.
- 42 P. M. D’Agostino and T. A. M. Gulder, *ACS Synth. Biol.*, 2018, **7**, 1702–1708.
- 43 E. R. Duell, P. D’Agostino, N. Shapiro, T. Woyke, T. M. Fuchs, T. A. M. Gulder, *Microb. Cell. Fact.*, 2019, available online: doi.org/10.1188/s12934-019-1080-6.
- 44 S. E. Ongley, X. Bian, Y. Zhang, R. Chau, W. H. Gerwick, R. Müller and B. A. Neilan, *ACS Chem. Biol.*, 2013, **8**, 1888–1893.
- 45 A. El Gamal, V. Agarwal, I. Rahman, B. S. Moore, *J. Am. Chem. Soc.*, 2016, **138**, 13167–13170.
- 46 E. M. Clerico, J. L. Ditty and S. S. Golden, *Humana Press (Totowa, NJ)*, 2007, pp 155–171.
- 47 A. T. Ma, C. M. Schmitdt and J. W. Golden, *Appl. Environ. Microbiol.*, 2014, **80**, 6704–6713.
- 48 M. J. Lai and E. I. Lan, *Biotechnol. Bioeng.*, 2019, 1–11.
- 49 S. Watanabe, H. Ozawa, H. Kato, K. Nimura-Matsune, T. Hirayama, F. Kudo, T. Eguchi, K. Kakinuma and H. Yoshikawa, *Biosci. Biotechnol. Biochem.*, 2018, **82**, 161–165.
- 50 H. Li and J. C. Liao, *Microb. Cell. Fact.*, 2013, **12**, 4.
- 51 M. Bielawski, D. Aili and B. Olofsson, *J. Org. Chem.*, 2008, **73**, 4602–4607.
- 52 N. Jalalian, E. E. Ishikawa, L. F. Silva and B. Olofsson, *Org. Lett.*, 2011, **13**, 1552–1555.

4 Summary and Outlook

4.1 Unlocking the biosynthetic potential of secondary metabolites with DiPaC on the example of anabaenopeptin

As mentioned earlier, with the enormous progress in the fields of genomic sequencing and bioinformatic analysis, the bottleneck of natural product research has shifted to the (heterologous) production of interesting compounds in sufficient amounts. Determining the bioactivity and subsequently evaluating the pharmaceutical potential of newly discovered secondary metabolites requires cheap and quick ways to generate the substance of choice in good yields. The herein presented work contributes to the development of the DiPaC method, which represents one approach to tackle the current challenges in this scientific field.

Using only basic laboratory equipment and techniques as well as solely *in vitro* experiments throughout the entire cloning process, we were able to create a universal method for the cloning of small- to mid-sized BGCs. As a consequence, the process is not hindered by a challenging genomic organization of the cluster or potential toxicity of the produced compound. DiPaC is based on simple PCR amplification of the genes of interest with up to at least 22 kb in a single reaction. This allows not only the simultaneous addition of homologous overhangs for subsequent vector integration but also enables on-line refactoring of the cluster. Addition or exchange of promoters, deletion of terminators and changing the order of certain genes is possible directly during the capturing process. By choosing the right vector to capture the BGC, a variety of different expression hosts and conditions can be addressed.

The anabaenopeptins NZ857 (**54**) and nostamide A (**55**) from *N. punctiforme* PCC 73102 were chosen as exemplary compounds to test the DiPaC method on a mid-size cyanobacterial gene cluster with a presumed benefit through gene rearrangement. The hexapeptidic protease inhibitors contain two unnatural amino acids necessary to be synthesized prior to the NRPS-based assembly of the final product. The decision to re-

4 Summary and Outlook

assemble the anabaenopeptin BGC for heterologous production in *E. coli* was based on the total cluster size of almost 30 kb. A single promoter at the beginning of the cluster might not be sufficient to guarantee complete transcription. Furthermore, the three genes necessary for precursor formation are located up- and downstream of the central four genes-comprising NRPS unit and hence could not be activated separately from the assembly machinery, which depends on full substrate availability. By choosing an IPTG inducible plasmid to host the unnatural amino acid synthesizing enzymes and integrating the NRPS genes into a tetracycline inducible vector, the two individual processes of anabaenopeptin biosynthesis could be triggered independently. Thereby, the presence of sufficient amounts of all necessary precursors could be ensured before the assembly of the final natural products was initiated. This strategy led to the successful production of both **54** and **55** in *E. coli* and is, to our knowledge, the first example for heterologous anabaenopeptin expression.

By applying DiPaC to clone and heterologously express a mid-size cyanobacterial BGC with a challenging gene arrangement, its broad versatility and ease of use was demonstrated. Cyanobacteria harbor a great number of natural products with high biological potential, but difficulties in cultivation and manipulation of this bacterial phylum has so far strongly hampered the exploitation of these compounds. This work aims to simplify further investigations into this very interesting part of secondary metabolite research by opening an alternative access to cyanobacterial natural product biochemistry.

4.2 Utilizing DiPaC for heterologous expression of the volatile compound sodorifen

Another class of challenging secondary metabolites are volatile organic compounds (VOCs). Not only their demanding physical properties but also their ubiquitous and yet not in the least understood interactions within the bacterial, plant and animal kingdoms raise many questions. To study their bioactivity in various contexts, it is required but difficult to get hold of sufficient substance amounts. For this reason we chose to further test the DiPaC method on the sodorifen BGC of the rhizobacterium *S. plymuthica* WS3236. Although this VOC has been extensively studied for almost ten years now, its biological role and activity remains elusive.

Harboring only four genes, the *sod* cluster is rather small and does not require any re-assembly or other manipulations. However, based on the volatility of this substance, a way to verify full gene transcription without directly interacting with the experimental setup was desirable. Therefore, a reporter system based on the green fluorescent protein (GFP)

4.3 Studies on the biosynthesis of the structurally unusual ambigols of cyanobacterial origin

was introduced downstream of the last *sod* gene, thus signaling complete read-out of the cluster after induction via increasing green fluorescence exhibited by the expressing cells. Using a fermenter with controllable aeration, an activated charcoal filter was installed on the air outlet of the setup. The reusable filter could be exchanged regularly and bound substances were easily elutable with organic solvents. This construction allowed to monitor and compare the sodorifen (**38**) production of both natural producer and heterologous host over a time course and under different cultivation conditions. In summary, expression in *E. coli* using TB-medium gave more than 25 times higher **38** yields in comparison to *S. plymuthica* cultivated under literature known optimal production conditions. Even more important, the heterologously produced natural product raw extracts were shown to be over 90% pure by NMR analysis, which greatly simplifies any downstream purification process. This is crucial especially for VOCs, as every additional purification step can decrease the yield dramatically and should therefore be avoided if possible. Nonetheless, reliable statements on the biological activity of a secondary metabolite can only be made when the pure compound is available for testing.

With this experiment, the applicability of DiPaC to create access to challenging volatile natural products was demonstrated, thereby paving the way for new scientific approaches towards unlocking the full potential of this fascinating group of substances. Regarding **38** in particular, an efficient method for high yield production with good purity was developed. The biological role of this secondary metabolite remains to be elucidated. Utilizing the herein described expression system, we want to tackle the problem of lacking enough substance for comprehensively addressing this question. Furthermore, during our research we found additional examples of gene clusters related to the *sod* BGC from *S. plymuthica* in several other species such as *Pseudomonas* and *Burkholderia*. Based on gene synteny, these identified clusters can be divided into eight distinct groups which harbor further genes, lack the first one or two genes of the *sod* cluster, or show (multiple) gene duplications. New derivatives of **38** can be expected to be encoded by those clusters and the expression as well as the structural and functional characterization of these so far unknown substances appears within reach of the presently developed method.

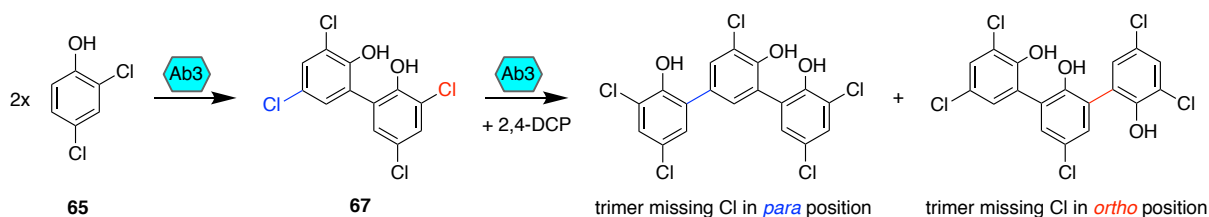
4.3 Studies on the biosynthesis of the structurally unusual ambigols of cyanobacterial origin

After the bioinformatic prediction of a BGC putatively responsible for the production of ambigol A-C (**62-64**) in *F. ambigua* 108b, the heterologous expression of the candidate cluster in *S. elongatus* led to the observation of at least one ambigol derivative. To our

4 Summary and Outlook

knowledge, this experiment represents the first example for the integration of such large DNA sequences into a cyanobacterial heterologous host system, as well as the first creation of *S. elongatus* double mutants harbouring and expressing a complex natural product BGC. Due to significantly decreased viability of the transformed cells, it was so far not possible to prepare enough material to elucidate the structure of the isolated ambigol derivative. Different approaches are conceivable to characterize the produced compound in the future. The heterologous expression could be repeated with freshly transformed mutants under optimized cultivation conditions. Alternatively, a co-cultivation of single mutants harboring either *ab1-5* or *ab6-10* in the neutral chromosomal site II could potentially lead to an improved production of **62-64**. Synthesized precursors might thereby get exchanged between the two different heterologous *S. elongatus* strains, while the reduced antibiotic stress should improve the general viability of the cells. Finally, all possible ambigol derivatives could be synthesized chemically to provide standards for the identification of the heterologously produced compound.

The latter method was already successfully applied for the identification of the two 2,4-DCP dimers **66** and **67** produced *in vitro* and *in vivo* by the CYP P450 enzymes Ab2 and Ab3. Ab2 established a biaryl-ether linkage in *ortho* position to give **66**, whereas Ab3 forms **67** with a biaryl-bond at the respective place as well as two **65** trimers of yet unknown structure, each missing one chlorine. This observation can be explained by the fact that all available *ortho* and *para* positions of **67** are blocked by either the biaryl-bond or one of the four present chlorines. If the reaction mechanism of Ab3 follows the standard, mesomerically stabilized, radical scheme, a third DCP unit can only be attached to **67** if either of the chlorines in 2- or 4-position relative to the hydroxy group is removed prior to or during further bond formation (see Scheme 7).



Scheme 7. Putative trimeric Ab3 reaction products. After abstracting one chlorine in *para* (blue) or *ortho* (red) position from the dimeric intermediate **67**, a third 2,4-DCP (**65**) unit can be added by establishing a second biaryl bond.

A full structural characterization of the observed compounds again requires comparison with authentic standards, as the yields of both *in vitro* and *in vivo* assays are rather low and extensive optimization trials did not lead to significant improvements. Based on the strategies developed for the chemical preparation of **66** and **67**, further synthesis routes

4.3 Studies on the biosynthesis of the structurally unusual ambigols of cyanobacterial origin

to obtain **62-64** as well as their chlorine-missing derivatives need to be developed. Additionally, the origin of the unusual bond connectivities present in **62-64** still needs to be elucidated. The observation of one ambigol derivative in the heterologous *ab* expression system indicates that all enzymes necessary for the biosynthesis have in fact been identified. However, neither the catalytic role of the potential cyclase *ab4*, nor of the unusual NRPS *ab9* or of the additional peptide carrier protein (PCP) domain *ab8* could be predicted so far. Moreover, there is only one AMP ligase *ab6* present in the BGC. This enzyme is necessary for precursor activation prior to PCP-loading and usually exhibits strong substrate specificity. As no amino acid precursors or peptide bonds are found in **62-64**, it is very likely that other intermediates need to be attached to *ab8* and the PCP domain of *ab9* in order to enable further reactions to happen. These might include chlorination of 4-HBA to yield DCP and/or coupling of (different) DCP monomers to form **62-64** or their putative dimeric precursors.

Formally, **62** and **63** consist of two 2,4-DCP (**65**) units and one central 3,5-DCP unit, whereby the transformation of the putatively universal precursor **65** into 3,5-DCP could in theory occur prior or during the first or second coupling reaction. Ambigol C (**64**) in contrast consists of two **65** units connected via one central 2,6-DCP unit which contains the two biaryl-ether linkages in formally prohibited *meta* position. This indicates towards two separate and so far unprecedented mechanisms encoded within the *ab* BGC to enable the biosynthesis of all to date reported ambigols found in *F. ambigua* 108b. Elucidation of the underlying enzymatic mechanisms will not only increase our knowledge about enzymatic biaryl- and biaryl-ether-couplings, but might also contribute to the development of novel, combined chemo-enzymatic approaches towards the synthesis of challenging biaryl substrates.

Bibliography

- [1] P. M. Dewick, *Medicinal natural products: a biosynthetic approach*, John Wiley and Sons, **2002**.
- [2] A. L. Demain, A. Fang, *Adv. Biochem. Eng. Biotechnol.* **2000**, 69, 1–39.
- [3] A. Fleming, *Br. J. Exp. Pathol.* **1929**, 10, 226.
- [4] G. M. Cragg, D. J. Newman, *Pharm. Biol.* **2001**, 39 Suppl 1, 8–17, DOI 10.1076/phbi.39.s1.8.0009.
- [5] K. Batanouny, Guide to Medicinal Plants in North Africa. Centre for Mediterranean Cooperation, International Union for Conservation of Nature and Natural Resources: *Anacyclus pyrethrum* L, tech. rep., ISBN 2-8317-0893-1, p 35-37, **2005**.
- [6] G. M. Cragg, D. J. Newman, *Biochim. Biophys. Acta* **2013**, 1830, 3670–95, DOI 10.1016/j.bbagen.2013.02.008.
- [7] C. C. Presley, C. W. Lindsley, *ACS Chem. Neurosci.* **2018**, 9, 2503–2518, DOI 10.1021/acscemneuro.8b00459.
- [8] J. Sumner, *The natural history of medicinal plants*, Timber press, **2000**.
- [9] J. K. Borchardt, *Drug News Perspect.* **2002**, 15, 187–192, DOI 10.1358/dnp.2002.15.3.840015.
- [10] A. J. Kelly, J. Kavanagh, J. Thomas, *Cochrane Database Syst. Rev.* **2013**, CD003099, DOI 10.1002/14651858.CD003099.pub2.
- [11] K. C. Huang, *The pharmacology of Chinese herbs*, Routledge, **1998**.
- [12] H. Yuan, Q. Ma, L. Ye, G. Piao, *Molecules* **2016**, 21, DOI 10.3390/molecules21050559.
- [13] S. Dubey, M. Kallubai, A. Sarkar, R. Subramanyam, *PLoS One* **2018**, 13, e0200053, DOI 10.1371/journal.pone.0200053.
- [14] N. E. Thomford, D. A. Senthebane, A. Rowe, D. Munro, P. Seele, A. Maroyi, K. Dzobo, *Int. J. Mol. Sci.* **2018**, 19, DOI 10.3390/ijms19061578.
- [15] M. Y. Park, H. J. Kwon, M. K. Sung, *Biosci. Biotechnol. Biochem.* **2009**, 73, 828–32, DOI 10.1271/bbb.80714.
- [16] U. Taukoorah, M. F. Mahomoodally, *Adv. Pharmacol. Sci.* **2016**, 2016, 3720850, DOI 10.1155/2016/3720850.
- [17] M. Li, F. Jiang, X. Yu, Z. Miao, *Biomed. Res. Int.* **2015**, 2015, 504932, DOI 10.1155/2015/504932.
- [18] S. Mehrotra, M. K. Goel, L. U. Rahman, A. K. Kukreja, *PCTOC* **2013**, 114, 31–38, DOI 10.1007/s11240-013-0302-6.
- [19] A. G. Atanasov, B. Waltenberger, E. M. Pferschy-Wenzig, T. Linder, C. Wawrosch, P. Uhrin, V. Temml, L. Wang, S. Schwaiger, E. H. Heiss, J. M. Rollinger, D. Schuster, J. M. Breuss, V. Bochkov, M. D. Mihovilovic, B. Kopp, R. Bauer, V. M. Dirsch, H. Stuppner, *Biotechnol. Adv.* **2015**, 33, 1582–1614, DOI 10.1016/j.biotechadv.2015.08.001.
- [20] S. Bernardini, A. Tiezzi, V. Laghezza Masci, E. Ovidi, *Nat. Prod. Res.* **2018**, 32, 1926–1950, DOI 10.1080/14786419.2017.1356838.
- [21] D. A. Dias, S. Urban, U. Roessner, *Metabolites* **2012**, 2, 303–36, DOI 10.3390/metabo2020303.

Bibliography

- [22] J. Achan, A. O. Talisuna, A. Erhart, A. Yeka, J. K. Tibenderana, F. N. Baliraine, P. J. Rosenthal, U. D'Alessandro, *Malar. J.* **2011**, *10*, 144, DOI 10.1186/1475-2875-10-144.
- [23] F. Gaedcke, *Arch. Pharm. Chem. Life Sci.* **1855**, *132*, 141–150, DOI 10.1002/ardp.18551320208.
- [24] S. Hosztafi, *Pharmazie* **1997**, *52*, 546–50.
- [25] H. Kaiser, *Z. Rheumatol.* **2008**, *67*, 252–62, DOI 10.1007/s00393-008-0257-x.
- [26] S. Eyal, *Toxins (Basel)* **2018**, *10*, DOI 10.3390/toxins10120491.
- [27] R. Hajar, *Heart Views* **2016**, *17*, 42–8, DOI 10.4103/1995-705X.182651.
- [28] S. Norn, H. Permin, P. R. Kruse, E. Kruse, *Dan. Medicinhist. Arbog.* **2009**, *37*, 79–98.
- [29] J. Davies, *ASM News* **1999**, *65*, 304–310.
- [30] D. D. Baker, M. Chu, U. Oza, V. Rajgarhia, *Nat. Prod. Res.* **2007**, *24*, 1225–44, DOI 10.1039/b602241n.
- [31] B. Shen, *Cell* **2015**, *163*, 1297–300, DOI 10.1016/j.cell.2015.11.031.
- [32] A. D. Buss, R. D. Waigh in *Burger's medicinal chemistry and drug discovery. Principles and practice. Vol 1*, (Ed.: M. E. Wolff), John Wiley and Sons, **1995**, pp. 983–1033.
- [33] M. A. Fischbach, C. T. Walsh, *Science* **2009**, *325*, 1089–93, DOI 10.1126/science.1176667.
- [34] I. Ojima, *J. Med. Chem.* **2008**, *51*, 2587–8, DOI 10.1021/jm701291u.
- [35] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2016**, *79*, 629–61, DOI 10.1021/acs.jnatprod.5b01055.
- [36] D. J. Newman, *J. Med. Chem.* **2008**, *51*, 2589–99, DOI 10.1021/jm0704090.
- [37] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2012**, *75*, 311–35, DOI 10.1021/np200906s.
- [38] S. Mathur, C. Hoskins, *Biomed. Rep.* **2017**, *6*, 612–614, DOI 10.3892/br.2017.909.
- [39] L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schäberle, D. E. Hughes, S. Epstein, M. Jones, L. Lazarides, V. A. Steadman, D. R. Cohen, C. R. Felix, K. A. Fetterman, W. P. Millett, A. G. Nitti, A. M. Zullo, C. Chen, K. Lewis, *Nature* **2015**, *517*, 455–9, DOI 10.1038/nature14098.
- [40] N. Ziemert, M. Alanjary, T. Weber, *Nat. Prod. Res.* **2016**, *33*, 988–1005, DOI 10.1039/c6np00025h.
- [41] L. Huo, J. J. Hug, C. Fu, X. Bian, Y. Zhang, R. Müller, *Nat. Prod. Res.* **2019**, DOI 10.1039/c8np00091c.
- [42] B. David, J.-L. Wolfender, D. A. Dias, *Phytochem. Rev.* **2015**, *14*, 299–315, DOI 10.1007/s11101-014-9367-z.
- [43] S. T. Lund, J. Bohlmann, *Science* **2006**, *311*, 804–805, DOI 10.1126/science.1118962.
- [44] R. Kaiser, *Science* **2006**, *311*, 806–807, DOI 10.1126/science.1119499.
- [45] E. Pichersky, J. Noel, N. Dudareva, *Science* **2006**, *311*, 808–811, DOI 10.1126/science.1118510.
- [46] I. T. Baldwin, R. Halitschke, A. Paschold, C. von Dahl, C. Preston, *Science* **2006**, *311*, 812–815, DOI 10.1126/science.1118446.
- [47] S. A. Goff, H. J. Klee, *Science* **2006**, *311*, 815–9, DOI 10.1126/science.1112614.
- [48] J. Schnürer, J. Olsson, T. Börjesson, *Fungal Genet. Biol.* **1999**, *27*, 209–17, DOI 10.1006/fgbi.1999.1139.
- [49] S. Schulz, J. S. Dickschat, *Nat. Prod. Rep.* **2007**, *24*, 814, DOI 10.1039/b507392h.
- [50] C. A. Pfeiffer, R. E. Johnston, *Physiol. Behav.* **1994**, *55*, 129–38, DOI 10.1016/0031-9384(94)90020-5.

- [51] F. P. Drijfhout, A. T. Groot, M. A. Posthumus, T. A. van Beek, A. de Groot, *Chemoecology* **2002**, *12*, 113–118, DOI 10.1007/s00049-002-8334-0.
- [52] M. D. Papke, S. E. Riechert, S. Schulz, *Anim. Behav.* **2001**, *61*, 877–886, DOI 10.1006/anbe.2000.1675.
- [53] S. Schulz, J. Fuhlendorff, J. L. M. Steidle, J. Collatz, J.-T. Franz, *ChemBioChem* **2004**, *5*, 1500–1507, DOI 10.1002/cbic.200400110.
- [54] A. M. Curran, S. I. Rabin, P. A. Prada, K. G. Furton, *J. Chem. Ecol.* **2005**, *31*, 1607–19, DOI 10.1007/s10886-005-5801-4.
- [55] K. Schulz-Bohm, L. Martin-Sanchez, P. Garbeva, *Front. Microbiol.* **2017**, *8*, 2484, DOI 10.3389/fmicb.2017.02484.
- [56] S. Giannoukos, A. Agapiou, B. Brkic, S. Taylor, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2019**, *1105*, 136–147, DOI 10.1016/j.jchromb.2018.12.015.
- [57] M. Heil, W. G. Land, *Front. Plant. Sci.* **2014**, *5*, 578, DOI 10.3389/fpls.2014.00578.
- [58] H. A. Tahir, Q. Gu, H. Wu, W. Raza, A. Hanif, L. Wu, M. V. Colman, X. Gao, *Front. Microbiol.* **2017**, *8*, 171, DOI 10.3389/fmicb.2017.00171.
- [59] V. Cordovez, L. Mommer, K. Moisan, D. Lucas-Barbosa, R. Pierik, R. Mumm, V. J. Carrion, J. M. Raaijmakers, *Front. Plant. Sci.* **2017**, *8*, 1262, DOI 10.3389/fpls.2017.01262.
- [60] K. Matsui, *Curr. Opin. Plant. Biol.* **2016**, *32*, 24–30, DOI 10.1016/j.pbi.2016.05.005.
- [61] M. C. Mescher, C. M. De Moraes, *J. Exp. Bot.* **2015**, *66*, 425–33, DOI 10.1093/jxb/eru414.
- [62] B. Weimer, K. Seefeldt, B. Dias, *Antonie Van Leeuwenhoek* **1999**, *76*, 247–61, DOI 10.1023/A:1002050625344.
- [63] L. Pripis-Nicolau, G. de Revel, A. Bertrand, A. Lonvaud-Funel, *J. Appl. Microbiol.* **2004**, *96*, 1176–84, DOI 10.1111/j.1365-2672.2004.02257.x.
- [64] H. C. Beck, A. M. Hansen, F. R. Lauritsen, *Enzyme Microb. Technol.* **2002**, *31*, 94–101, DOI 10.1016/s0141-0229(02)00067-4.
- [65] J. Crouzet, *Functionality of Food Phytochemicals*, Plenum Press, New York, **1997**.
- [66] P. Morales, I. Feliu, E. Fernandez-Garcia, M. Nunez, *J. Food. Prot.* **2004**, *67*, 567–73, DOI 10.4315/0362-028X-67.3.567.
- [67] O. van Mastriigt, D. Gallegos Tejeda, M. N. Kristensen, T. Abee, E. J. Smid, *Microb. Cell Fact.* **2018**, *17*, 104, DOI 10.1186/s12934-018-0950-7.
- [68] J. S. Dickschat, T. Martens, T. Brinkhoff, M. Simon, S. Schulz, *Chem. Biodivers.* **2005**, *2*, 837–65, DOI 10.1002/cbdv.200590062.
- [69] R. Schmidt, V. Cordovez, W. de Boer, J. Raaijmakers, P. Garbeva, *ISME J.* **2015**, *9*, 2329–35, DOI 10.1038/ismej.2015.42.
- [70] O. Tyc, V. C. L. de Jager, M. van den Berg, S. Gerards, T. K. S. Janssens, N. Zaagman, M. Kai, A. Svatos, H. Zweers, C. Hordijk, H. Besselink, W. de Boer, P. Garbeva, *Microb. Biotechnol.* **2017**, *10*, 910–925, DOI 10.1111/1751-7915.12735.
- [71] W. G. D. Fernando, R. Ramarathnam, A. S. Krishnamoorthy, S. C. Savchuk, *Soil Biol. Biochem.* **2005**, *37*, 955–964, DOI 10.1016/j.soilbio.2004.10.021.
- [72] A. Bruce, D. Stewart, S. Verrall, R. E. Wheatley, *Int. Biodeterior. Biodegradation* **2003**, *51*, 101–108, DOI 10.1016/s0964-8305(02)00088-4.
- [73] B. Chaurasia, A. Pandey, L. M. Palni, P. Trivedi, B. Kumar, N. Colvin, *Microbiol. Res.* **2005**, *160*, 75–81, DOI 10.1016/j.micres.2004.09.013.
- [74] E. Barbieri, A. M. Gioacchini, A. Zambonelli, L. Bertini, V. Stocchi, *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3411–5, DOI 10.1002/rcm.2209.

Bibliography

- [75] D. C. Robacker, A. J. Martinez, J. A. Garcia, R. J. Bartelt, *Fla. Entomol.* **1998**, *81*, 497, DOI 10.2307/3495948.
- [76] L. Tosi, C. Sola, *Ethology* **2010**, *95*, 177–185, DOI 10.1111/j.1439-0310.1993.tb00468.x.
- [77] D. C. Robacker, R. A. Flath, *J. Chem. Ecol.* **1995**, *21*, 1861–74, DOI 10.1007/BF02033682.
- [78] C. M. Ryu, M. A. Farag, C. H. Hu, M. S. Reddy, H. X. Wei, P. W. Pare, J. W. Kloepper, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4927–32, DOI 10.1073/pnas.0730845100.
- [79] C. M. Ryu, M. A. Farag, C. H. Hu, M. S. Reddy, J. W. Kloepper, P. W. Pare, *Plant. Physiol.* **2004**, *134*, 1017–26, DOI 10.1104/pp.103.026583.
- [80] J. S. Dickschat, S. C. Wenzel, H. B. Bode, R. Müller, S. Schulz, *Chembiochem* **2004**, *5*, 778–87, DOI 10.1002/cbic.200300813.
- [81] W. S. Leal, M. Sawada, M. Hasegawa, *J. Chem. Ecol.* **1993**, *19*, 1303–1313, DOI 10.1007/bf00984877.
- [82] J. S. Dickschat, I. Wagner-Döbler, S. Schulz, *J. Chem. Ecol.* **2005**, *31*, 925–947, DOI 10.1007/s10886-005-3553-9.
- [83] B. V. Burger, A. E. Nell, H. S. C. Spies, M. Le Roux, R. C. Bigalke, P. A. J. Brand, *J. Chem. Ecol.* **1999**, *25*, 2057–2084, DOI 10.1023/a:1021036823079.
- [84] D. C. Robacker, C. R. Lauzon, *J. Chem. Ecol.* **2002**, *28*, 1549–63, DOI 10.1023/A:1019920328062.
- [85] E. D. Morgan, R. R. Do Nascimento, S. J. Keegans, J. Billen, *J. Chem. Ecol.* **1999**, *25*, 1395–1409, DOI 10.1023/a:1020987028163.
- [86] B. P. Moore, W. V. Brown, M. Rothschild, *Chemoecology* **1990**, *1*, 43–51, DOI 10.1007/bf01325227.
- [87] R. F. Simpson, D. L. Capone, M. A. Sefton, *J. Agric. Food Chem.* **2004**, *52*, 5425–30, DOI 10.1021/jf049484z.
- [88] J. S. Dickschat, H. Reichenbach, I. Wagner-Döbler, S. Schulz, *Eur. J. Org. Chem.* **2005**, *2005*, 4141–4153, DOI 10.1002/ejoc.200500280.
- [89] B. W. Zilkowski, R. J. Bartelt, *J. Chem. Ecol.* **1999**, *25*, 1759–1770, DOI 10.1023/a:1020921530406.
- [90] D. C. Robacker, C. R. Lauzon, X. He, *J. Chem. Ecol.* **2004**, *30*, 1329–1347, DOI 10.1023/b:joec.0000037743.98703.43.
- [91] E. Thibout, J. F. Guillot, S. Ferary, P. Limouzin, J. Auger, *Experientia* **1995**, *51*, 1073–1075, DOI 10.1007/bf01946919.
- [92] H. Elgaali, T. R. Hamilton-Kemp, M. C. Newman, R. W. Collins, K. Yu, D. D. Archbold, *J. Basic Microbiol.* **2002**, *42*, 373–80, DOI 10.1002/1521-4028(200212)42:6<373::AID-JOBM373>3.0.CO;2-4.
- [93] K. Yu, T. R. Hamilton-Kemp, D. D. Archbold, R. W. Collins, M. C. Newman, *J. Agric. Food Chem.* **2000**, *48*, 413–417, DOI 10.1021/jf990576b.
- [94] P. D. Martino, R. Fursy, L. Bret, B. Sundararaju, R. S. Phillips, *Can. J. Microbiol.* **2003**, *49*, 443–9, DOI 10.1139/w03-056.
- [95] S. Schulz, J. Fuhlendorff, H. Reichenbach, *Tetrahedron* **2004**, *60*, 3863–3872, DOI 10.1016/j.tet.2004.03.005.
- [96] M. R. Tellez, K. K. Schrader, M. Kobaisy, *J. Agric. Food Chem.* **2001**, *49*, 5989–5992, DOI 10.1021/jf010722p.
- [97] G. Urbach, *Int. Dairy J.* **1995**, *5*, 877–903, DOI 10.1016/0958-6946(95)00037-2.
- [98] G. Smit, B. A. Smit, W. J. Engels, *FEMS Microbiol. Rev.* **2005**, *29*, 591–610, DOI 10.1016/j.femsre.2005.04.002.
- [99] P. Bonnarme, C. Lapadatescu, M. Yvon, H. E. Spinnler, *J. Dairy Res.* **2001**, *68*, 663–74, DOI 10.1017/S002202990100509X.

- [100] B. Dias, B. Weimer, *Appl. Environ. Microbiol.* **1998**, *64*, 3320–6.
- [101] P. Bonnarme, L. Psoni, H. E. Spinnler, *Appl. Environ. Microbiol.* **2000**, *66*, 5514–5517, DOI 10.1128/aem.66.12.5514-5517.2000.
- [102] K. Arfi, F. Amarita, H. E. Spinnler, P. Bonnarme, *J. Biotechnol.* **2003**, *105*, 245–253, DOI 10.1016/j.jbiotec.2003.07.003.
- [103] S. Gao, E. S. Mooberry, J. L. Steele, *Appl. Environ. Microbiol.* **1998**, *64*, 4670–5.
- [104] T. S. Bates, R. J. Charlson, R. H. Gammon, *Nature* **1987**, *329*, 319–321, DOI 10.1038/329319a0.
- [105] R. J. Charlson, J. E. Lovelock, M. O. Andreae, S. G. Warren, *Nature* **1987**, *326*, 655–661, DOI 10.1038/326655a0.
- [106] M. O. Andreae, P. Crutzen, *Science* **1997**, *276*, 1052–1058, DOI 10.1126/science.276.5315.1052.
- [107] M. O. Andreae, *Mar. Chem.* **1990**, *30*, 1–29, DOI 10.1016/0304-4203(90)90059-1.
- [108] T. Groene, *J. Mar. Syst.* **1995**, *6*, 191–209, DOI 10.1016/0924-7963(94)00023-5.
- [109] N. N. Gerber, *J. Chem. Ecol.* **1977**, *3*, 475–482, DOI 10.1007/bf00988190.
- [110] P. Darriet, S. Lamy, S. La Guerche, M. Pons, D. Dubourdieu, D. Blancard, P. Steliopoulos, A. Mosandl, *Eur. Food Res. Technol.* **2001**, *213*, 122–125, DOI 10.1007/s002170100346.
- [111] S. H. von Reuss, M. Kai, B. Piechulla, W. Francke, *Angew. Chem. Int. Ed.* **2010**, *49*, 2009–10, DOI 10.1002/anie.200905680.
- [112] D. Domik, A. Thürmer, T. Weise, W. Brandt, R. Daniel, B. Piechulla, *Front. Microbiol.* **2016**, *7*, 737, DOI 10.3389/fmicb.2016.00737.
- [113] S. von Reuss, D. Domik, M. C. Lemfack, N. Magnus, M. Kai, T. Weise, B. Piechulla, *J. Am. Chem. Soc.* **2018**, *140*, 11855–11862, DOI 10.1021/jacs.8b08510.
- [114] T. Weise, A. Thürmer, S. Brady, M. Kai, R. Daniel, G. Gottschalk, B. Piechulla, *FEMS Microbiol. Lett.* **2014**, *352*, 45–53, DOI 10.1111/1574-6968.12359.
- [115] R. Schmidt, V. Jager, D. Zühlke, C. Wolff, J. Bernhardt, K. Cankar, J. Beekwilder, W. V. Ijcken, F. Sleutels, W. Boer, K. Riedel, P. Garbeva, *Sci. Rep.* **2017**, *7*, 862, DOI 10.1038/s41598-017-00893-3.
- [116] N. Magnus, T. Weise, B. Piechulla, *Front. Microbiol.* **2017**, *8*, 2522, DOI 10.3389/fmicb.2017.02522.
- [117] M. Phillips, R. N. Cataneo, A. Chaturvedi, P. D. Kaplan, M. Libardoni, M. Mundada, U. Patel, X. Zhang, *PLoS One* **2013**, *8*, e75274, DOI 10.1371/journal.pone.0075274.
- [118] G. A. Mills, V. Walker, *J. Chromatogr. B Biomed. Sci. Appl.* **2001**, *753*, 259–268, DOI 10.1016/S0378-4347(00)00554-5.
- [119] P. Mochalski, J. King, A. Kupferthaler, K. Unterkofler, H. Hinterhuber, A. Amann, *J. Breath Res.* **2011**, *5*, 046010, DOI 10.1088/1752-7155/5/4/046010.
- [120] H. Al-Kateb, B. de Lacy Costello, N. Ratcliffe, *J. Breath Res.* **2013**, *7*, 036004, DOI 10.1088/1752-7155/7/3/036004.
- [121] N. Shehada, G. Bronstrup, K. Funke, S. Christiansen, M. Leja, H. Haick, *Nano Lett.* **2015**, *15*, 1288–95, DOI 10.1021/nl504482t.
- [122] S. Shah, N. Akhter, B. Auckloo, I. Khan, Y. Lu, K. Wang, B. Wu, Y.-W. Guo, *Mar. Drugs* **2017**, *15*, 354, DOI 10.3390/md15110354.
- [123] K. Walton, J. P. Berry, *Mar. Drugs* **2016**, *14*, DOI 10.3390/md14040073.
- [124] M. R. Badger, *J. Exp. Bot.* **2003**, *54*, 609–622, DOI 10.1093/jxb/erg076.
- [125] R. Wood, *Environment. International* **2016**, *91*, 276–282, DOI 10.1016/j.envint.2016.02.026.

Bibliography

- [126] G. N. A. Senhorinho, G. M. Ross, J. A. Scott, *Phycologia* **2015**, *54*, 271–282, DOI 10.2216/14-092.1.
- [127] S. Mandal, J. Rath, *Extremophilic Cyanobacteria For Novel Drug Development*, Springer, **2015**.
- [128] R. Singh, P. Parihar, M. Singh, A. Bajguz, J. Kumar, S. Singh, V. P. Singh, S. M. Prasad, *Front. Microbiol.* **2017**, *8*, DOI 10.3389/fmicb.2017.00515.
- [129] R.-. B. Volk, F. H. Furkert, *Microbiol. Res.* **2006**, *161*, 180–186, DOI 10.1016/j.micres.2005.08.005.
- [130] W. H. Gerwick, M. A. Roberts, P. J. Proteau, J.-L. Chen, *J. Appl. Phycol.* **1994**, *6*, 143–149, DOI 10.1007/bf02186068.
- [131] N. A. El Semary, M. Fouda, *Asian Pac. J. Trop. Biomed.* **2015**, *5*, 992–995, DOI 10.1016/j.apjtb.2015.09.004.
- [132] O. Papendorf, G. M. König, A. D. Wright, *Phytochem.* **1998**, *49*, 2383–2386, DOI 10.1016/S0031-9422(98)00440-3.
- [133] G. M. L. Patterson, L. K. Larsen, R. E. Moore, *J. Appl. Phycol.* **1994**, *6*, 151–157, DOI 10.1007/bf02186069.
- [134] F. E. Koehn, R. E. Longley, J. K. Reed, *J. Nat. Prod.* **1992**, *55*, 613–619, DOI 10.1021/np50083a009.
- [135] S. Vijayakumar, M. Menakha, *Acute Med.* **2015**, *5*, 15–23, DOI 10.1016/j.jacme.2015.02.004.
- [136] J. I. Jiménez, P. J. Scheuer, *J. Nat. Prod.* **2001**, *64*, 200–203, DOI 10.1021/np000462q.
- [137] K. L. McPhail, J. Correa, R. G. Linington, J. González, E. Ortega-Barria, T. L. Capson, W. H. Gerwick, *J. Nat. Prod.* **2007**, *70*, 984–988, DOI 10.1021/np0700772.
- [138] M. J. Baluñas, R. G. Linington, K. Tidgewell, A. M. Fenner, L.-D. Urena, G. D. Togna, D. E. Kyle, W. H. Gerwick, *J. Nat. Prod.* **2010**, *73*, 60–66, DOI 10.1021/np900622m.
- [139] S. P. Gunasekera, C. Ross, V. J. Paul, S. Matthew, H. Luesch, *J. Nat. Prod.* **2008**, *71*, 887–890, DOI 10.1021/np0706769.
- [140] R. E. Schwartz, C. F. Hirsch, D. F. Sesin, J. E. Flor, M. Chartrain, R. E. Fromtling, G. H. Harris, M. J. Salvatore, J. M. Liesch, K. Yudin, *J. Ind. Microbiol. Biotechnol.* **1990**, *5*, 113–123, DOI 10.1007/BF01573860.
- [141] C. D. Smith, X. Zhang, S. L. Mooberry, G. M. Patterson, R. E. Moore, *Cancer Res.* **1994**, *54*, 3779–84.
- [142] C. D. Smith, X. Zhang, *J. Biol. Chem.* **1996**, *271*, 6192–8, DOI 10.1074/jbc.271.17.9906.
- [143] S. Chaganty, T. Golakoti, C. Heltzel, R. E. Moore, W. Y. Yoshida, *J. Nat. Prod.* **2004**, *67*, 1403–6, DOI 10.1021/np0499665.
- [144] G. Trimurtulu, I. Ohtani, G. M. L. Patterson, R. E. Moore, T. H. Corbett, F. A. Valeriote, L. Demchik, *J. Am. Chem. Soc.* **1994**, *116*, 4729–4737, DOI 10.1021/ja00090a020.
- [145] T. Golakoti, J. Ogino, C. E. Heltzel, T. Le Husebo, C. M. Jensen, L. K. Larsen, G. M. L. Patterson, R. E. Moore, S. L. Mooberry, T. H. Corbett, F. A. Valeriote, *J. Am. Chem. Soc.* **2002**, *117*, 12030–12049, DOI 10.1021/ja00154a002.
- [146] M. J. Edelman, D. R. Gandara, P. Hausner, V. Israel, D. Thornton, J. DeSanto, L. A. Doyle, *Lung Cancer* **2003**, *39*, 197–9, DOI 10.1016/S0169-5002(02)00511-1.
- [147] J. Liang, R. E. Moore, E. D. Moher, J. E. Munroe, R. S. Al-awar, D. A. Hay, D. L. Varie, T. Y. Zhang, J. A. Aikins, M. J. Martinelli, C. Shih, J. E. Ray, L. L. Gibson, V. Vasudevan, L. Polin, K. White, J. Kushner, C. Simpson, S. Pugh, T. H. Corbett, *Invest. New Drugs* **2005**, *23*, 213–24, DOI 10.1007/s10637-005-6729-9.
- [148] G. D’Agostino, J. del Campo, B. Mellado, M. A. Izquierdo, T. Minarik, L. Cirri, L. Marini, J. L. Perez-Gracia, G. Scambia, *Int. J. Gynecol. Cancer* **2006**, *16*, 71–6, DOI 10.1111/j.1525-1438.2006.00276.x.

- [149] S. Mo, A. Kronic, S. D. Pegan, S. G. Franzblau, J. Orjala, *J. Nat. Prod.* **2009**, *72*, 2043–5, DOI 10.1021/np900288x.
- [150] B. Jaki, J. Orjala, O. Sticher, *J. Nat. Prod.* **1999**, *62*, 502–3, DOI 10.1021/np980444x.
- [151] B. Jaki, J. Heilmann, O. Sticher, *J. Nat. Prod.* **2000**, *63*, 1283–5, DOI 10.1021/np000033s.
- [152] B. Jaki, J. Orjala, J. Heilmann, A. Linden, B. Vogler, O. Sticher, *J. Nat. Prod.* **2000**, *63*, 339–43, DOI 10.1021/np9903090.
- [153] K. Rankin, K. Alroy, R. Kudela, S. Oates, M. Murray, M. Miller, *Toxins* **2013**, *5*, 1051–1063, DOI 10.3390/toxins5061051.
- [154] A. Bownik, *Toxin Rev.* **2010**, *29*, 99–114, DOI 10.3109/15569543.2010.516464.
- [155] A. Cullen, L. A. Pearson, R. Mazmouz, T. Liu, A. H. Soeriyadi, S. E. Ongley, B. A. Neilan, *Nat. Prod. Rep.* **2019**, DOI 10.1039/c8np00063h.
- [156] W. Vetter, E. Stoll, M. J. Garson, S. J. Fahey, C. Gaus, J. F. Müller, *Environ. Toxicol. Chem.* **2002**, *21*, 2014–9, DOI 10.1002/etc.5620211002.
- [157] M. Gantar, Z. Svircev, *J. Phycol.* **2008**, *44*, 260–8, DOI 10.1111/j.1529-8817.2008.00469.x.
- [158] S. Gaul, P. Bendig, D. Olbrich, N. Rosenfelder, P. Ruff, C. Gaus, J. F. Mueller, W. Vetter, *Mar. Pollut. Bull.* **2011**, *62*, 2463–8, DOI 10.1016/j.marpolbul.2011.08.022.
- [159] N. Fortin, V. Munoz-Ramos, D. Bird, B. Lévesque, L. Whyte, C. Greer, *Life* **2015**, *5*, 1346–1380, DOI 10.3390/life5021346.
- [160] M. F. Watanabe, S. Oishi, K.-I. Harada, K. Matsuura, H. Kawai, M. Suzuki, *Toxicon* **1988**, *26*, 1017–1025, DOI 10.1016/0041-0101(88)90200-0.
- [161] K. Sivonen, M. Namikoshi, W. R. Evans, B. V. Gromov, W. W. Carmichael, K. L. Rinehart, *Toxicon* **1992**, *30*, 1481–1485, DOI 10.1016/0041-0101(92)90524-9.
- [162] I. Oksanen, J. Jokela, D. P. Fewer, M. Wahlsten, J. Rikkinen, K. Sivonen, *Appl. Environ. Microbiol.* **2004**, *70*, 5756–5763, DOI 10.1128/aem.70.10.5756-5763.2004.
- [163] U. Kaasalainen, J. Jokela, D. P. Fewer, K. Sivonen, J. Rikkinen, *Mol. Plant Microbe Interact.* **2009**, *22*, 695–702, DOI 10.1094/mpmi-22-6-0695.
- [164] U. Kaasalainen, D. P. Fewer, J. Jokela, M. Wahlsten, K. Sivonen, J. Rikkinen, *PNAS* **2012**, *109*, 5886–5891, DOI 10.1073/pnas.1200279109.
- [165] E. M. Jochimsen, W. W. Carmichael, J. An, D. M. Cardo, S. T. Cookson, C. E. M. Holmes, M. B. Antunes, D. A. de Melo Filho, T. M. Lyra, V. S. T. Barreto, S. M. F. O. Azevedo, W. R. Jarvis, *N. Engl. J. Med.* **1998**, *338*, 873–878, DOI 10.1056/nejm199803263381304.
- [166] Y. Ueno, S. Nagata, T. Tsutsumi, A. Hasegawa, M. F. Watanabe, H. D. Park, G. C. Chen, G. Chen, S. Z. Yu, *Carcinogenesis* **1996**, *17*, 1317–21, DOI 10.1093/carcin/17.6.1317.
- [167] C.-C. Ruan, Y.-H. Chen, Z.-Q. Zhang, *World J. Gastroenterol.* **2011**, *3*, 47, DOI 10.3748/wjg.v3.i1.47.
- [168] R. Nishiwaki-Matsushima, T. Ohta, S. Nishiwaki, M. Suganuma, K. Kohyama, T. Ishikawa, W. W. Carmichael, H. Fujiki, *J. Cancer Res. Clin. Oncol.* **1992**, *118*, 420–4, DOI 10.1007/BF01629424.
- [169] R. M. Dawson, *Toxicon* **1998**, *36*, 953–962, DOI 10.1016/s0041-0101(97)00102-5.
- [170] J. E. Eriksson, D. Toivola, J. A. O. Meriluoto, H. Karaki, Y. G. Han, D. Hartshorne, *Biochem. Biophys. Res. Commun.* **1990**, *173*, 1347–1353, DOI 10.1016/s0006-291x(05)80936-2.
- [171] F. Tencalla, D. Dietrich, *Toxicon* **1997**, *35*, 583–595, DOI 10.1016/s0041-0101(96)00153-5.
- [172] M. Kaebnick, E. Dittmann, T. Börner, B. A. Neilan, *Appl. Environ. Microbiol.* **2002**, *68*, 449–55, DOI 10.1128/AEM.68.2.449455.2002.
- [173] M. A. Avila, T. H. J. Niedermeyer, A. Daily, M. Swiatecka-Hagenbruch, J. A. Moscow, *PLoS One* **2014**, *9*, e91476, DOI 10.1371/journal.pone.0091476.

Bibliography

- [174] K. Sivonen, K. Himberg, R. Luukkainen, S. I. Niemelä, G. K. Poon, G. A. Codd, *Environ. Toxicol. Water Qual.* **1989**, *4*, 339–352, DOI 10.1002/tox.2540040310.
- [175] K.-I. Harada, Y. Kimura, K. Ogawa, M. Suzuki, A. M. Dahlem, V. R. Beasley, W. W. Carmichael, *Toxicon* **1989**, *27*, 1289–1296, DOI 10.1016/0041-0101(89)90060-3.
- [176] A. Rantala-Ylinen, S. Känä, H. Wang, L. Rouhiainen, M. Wahlsten, E. Rizzi, K. Berg, M. Gugger, K. Sivonen, *Appl. Environ. Microbiol.* **2011**, *77*, 7271–7278, DOI 10.1128/aem.06022-11.
- [177] H. D. Park, M. F. Watanabe, K. Harda, H. Nagai, M. Suzuki, M. Watanabe, H. Hayashi, *Nat. Toxins* **1993**, *1*, 353–60, DOI 10.1002/nt.2620010606.
- [178] E. Viaggiu, S. Melchiorre, F. Volpi, A. Di Corcia, R. Mancini, L. Garibaldi, G. Crichigno, M. Bruno, *Environ. Toxicol.* **2004**, *19*, 191–7, DOI 10.1002/tox.20011.
- [179] A. Ghassempour, N. M. Najafi, A. Mehdinia, S. S. H. Davarani, M. Fallahi, M. Nakhshab, *J. of Chrom. A* **2005**, *1078*, 120–127, DOI 10.1016/j.chroma.2005.04.053.
- [180] R. A. Smith, D. Lewis, *Vet. Hum. Toxicol.* **1987**, *29*, 153–4.
- [181] L. Backer, J. Landsberg, M. Miller, K. Keel, T. Taylor, *Toxins* **2013**, *5*, 1597–1628, DOI 10.3390/toxins5091597.
- [182] S. Wonnacott, K. L. Swanson, E. X. Albuquerque, N. J. S. Huby, P. Thompson, T. Gallagher, *Biochem. Pharmacol.* **1992**, *43*, 419–423, DOI 10.1016/0006-2952(92)90558-z.
- [183] K. J. James, A. Furey, I. R. Sherlock, M. A. Stack, M. Twohig, F. B. Caudwell, O. M. Skulberg, *J. Chromatogr. A* **1998**, *798*, 147–57, DOI 10.1016/S0021-9673(97)01207-7.
- [184] J. Cardellina, F. Marner, R. Moore, *Science* **1979**, *204*, 193–195, DOI 10.1126/science.107586.
- [185] T. Yasumoto, *Toxicon* **1998**, *36*, 1515–1518, DOI 10.1016/s0041-0101(98)00142-1.
- [186] E. Ito, M. Satake, T. Yasumoto, *Toxicon* **2002**, *40*, 551–556, DOI 10.1016/s0041-0101(01)00251-3.
- [187] N. Aimi, H. Odaka, S. Sakai, H. Fujiki, M. Suganuma, R. E. Moore, G. M. Patterson, *J. Nat. Prod.* **1990**, *53*, 1593–6, DOI 10.1021/np50072a035.
- [188] W. Jiang, W. Zhou, H. Uchida, M. Kikumori, K. Irie, R. Watanabe, T. Suzuki, B. Sakamoto, M. Kamio, H. Nagai, *Mar. Drugs* **2014**, *12*, 2748–2759, DOI 10.3390/md12052748.
- [189] W. Jiang, S. Tan, Y. Hanaki, K. Irie, H. Uchida, R. Watanabe, T. Suzuki, B. Sakamoto, M. Kamio, H. Nagai, *Mar. Drugs* **2014**, *12*, 5788–5800, DOI 10.3390/md12125788.
- [190] D. T. A. Youssef, L. A. Shaala, G. A. Mohamed, S. R. M. Ibrahim, Z. M. Banjar, J. M. Badr, K. L. McPhail, A. L. Risinger, S. L. Mooberry, *Nat. Prod. Res.* **2014**, *29*, 703–709, DOI 10.1080/14786419.2014.982647.
- [191] K. Cusick, G. Sayler, *Mar. Drugs* **2013**, *11*, 991–1018, DOI 10.3390/md11040991.
- [192] H. L. F. Borges, L. H. Z. Branco, M. D. Martins, C. S. Lima, P. T. Barbosa, G. A. S. T. Lira, M. C. Bittencourt-Oliveira, R. J. R. Molica, *Harmful Algae* **2015**, *43*, 46–57, DOI 10.1016/j.hal.2015.01.003.
- [193] F. Pomati, S. Sacchi, C. Rossetti, S. Giovannardi, H. Onodera, Y. Oshima, B. A. Neilan, *J. Phycol.* **2000**, *36*, 553–562, DOI 10.1046/j.1529-8817.2000.99181.x.
- [194] K. A. Lefebvre, *Mar. Drugs* **2008**, *6*, 103–116, DOI 10.3390/md6020103.
- [195] Y. Oshima, S. I. Blackburn, G. M. Hallegraef, *Mar. Biol.* **1993**, *116*, 471–476, DOI 10.1007/bf00350064.
- [196] G. Usup, D. M. Kulis, D. M. Anderson, *Nat. Toxins* **1994**, *2*, 254–62, DOI 10.1002/nt.2620020503.
- [197] M. Wiese, P. M. D'Agostino, T. K. Mihali, M. C. Moffitt, B. A. Neilan, *Mar. Drugs* **2010**, *8*, 2185–211, DOI 10.3390/md8072185.

- [198] H. Terlau, S. H. Heinemann, W. Stühmer, M. Pusch, F. Conti, K. Imoto, S. Numa, *FEBS Lett.* **1991**, *293*, 93–96, DOI 10.1016/0014-5793(91)81159-6.
- [199] J. Wang, J. J. Salata, P. B. Bennett, *J. Gen. Physiol.* **2003**, *121*, 583–598, DOI 10.1085/jgp.200308812.
- [200] Z. Su, M. Sheets, H. Ishida, F. Li, W. H. Barry, *J. Pharmacol. Exp. Ther.* **2004**, *308*, 324–9, DOI 10.1124/jpet.103.056564.
- [201] A. D. Turner, M. Dhanji-Rapkova, K. Dean, S. Milligan, M. Hamilton, J. Thomas, C. Poole, J. Haycock, J. Spelman-Marriott, A. Watson, K. Hughes, B. Marr, A. Dixon, L. Coates, *Toxins (Basel)* **2018**, *10*, DOI 10.3390/toxins10030094.
- [202] S. M. Etheridge, *Toxicon* **2010**, *56*, 108–22, DOI 10.1016/j.toxicon.2009.12.013.
- [203] J. H. Landsberg, *Rev. Fish. Sci. Aquac.* **2002**, *10*, 113–390, DOI 10.1080/20026491051695.
- [204] V. Agarwal, A. A. El Gamal, K. Yamanaka, D. Poth, R. D. Kersten, M. Schorn, E. E. Allen, B. S. Moore, *Nat. Chem. Biol.* **2014**, *10*, 640–7, DOI 10.1038/nchembio.1564.
- [205] V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen, B. S. Moore, *Nat. Chem. Biol.* **2017**, *13*, 537–543, DOI 10.1038/nchembio.2330.
- [206] M. Kuniyoshi, K. Yamada, T. Higa, *Experientia* **1985**, *41*, 523–524, DOI 10.1007/bf01966182.
- [207] K. Makoto, K. Tomoyuki, N. Yoshikatsu, U. Daisuke, *Chem. Lett.* **2005**, *34*, 1272–1273, DOI 10.1246/c1.2005.1272.
- [208] A. Malmvarn, Y. Zebuhr, L. Kautsky, K. Bergman, L. Asplund, *Chemosphere* **2008**, *72*, 910–6, DOI 10.1016/j.chemosphere.2008.03.036.
- [209] E. L. Teuten, L. Xu, C. M. Reddy, *Science* **2005**, *307*, 917–20, DOI 10.1126/science.1106882.
- [210] N. J. Shaul, N. G. Dodder, L. I. Aluwihare, S. A. Mackintosh, K. A. Maruya, S. J. Chivers, K. Danil, D. W. Weller, E. Hoh, *Environ. Sci. Technol.* **2015**, *49*, 1328–38, DOI 10.1021/es505156q.
- [211] Y. Wan, K. Choi, S. Kim, K. Ji, H. Chang, S. Wiseman, P. D. Jones, J. S. Khim, S. Park, J. Park, M. H. Lam, J. P. Giesy, *Environ. Sci. Technol.* **2010**, *44*, 5233–9, DOI 10.1021/es1002764.
- [212] H. S. Wang, Z. J. Chen, K. L. Ho, L. C. Ge, J. Du, M. H. Lam, J. P. Giesy, M. H. Wong, C. K. Wong, *Environ. Int.* **2012**, *47*, 66–72, DOI 10.1016/j.envint.2012.06.004.
- [213] A. Chen, J. S. Park, L. Linderholm, A. Rhee, M. Petreas, E. A. DeFranco, K. N. Dietrich, S. M. Ho, *Environ. Sci. Technol.* **2013**, *47*, 3902–8, DOI 10.1021/es3046839.
- [214] S. B. Wiseman, Y. Wan, H. Chang, X. Zhang, M. Hecker, P. D. Jones, J. P. Giesy, *Mar. Pollut. Bull.* **2011**, *63*, 179–88, DOI 10.1016/j.marpolbul.2011.02.008.
- [215] E. N. Segraves, R. R. Shah, N. L. Segraves, T. A. Johnson, S. Whitman, J. K. Sui, V. A. Kenyon, R. H. Cichewicz, P. Crews, T. R. Holman, *J. Med. Chem.* **2004**, *47*, 4060–5, DOI 10.1021/jm049872s.
- [216] R. F. Canton, D. E. Scholten, G. Marsh, P. C. de Jong, M. van den Berg, *Toxicol. Appl. Pharmacol.* **2008**, *227*, 68–75, DOI 10.1016/j.taap.2007.09.025.
- [217] J. A. de la Fuente, S. Manzanaro, M. J. Martin, T. G. de Quesada, I. Reymundo, S. M. Luengo, F. Gago, *J. Med. Chem.* **2003**, *46*, 5208–21, DOI 10.1021/jm030957n.
- [218] J. Legradi, A. K. Dahlberg, P. Ceniijn, G. Marsh, L. Asplund, A. Bergman, J. Legler, *Environ. Sci. Technol.* **2014**, *48*, 14703–11, DOI 10.1021/es5039744.
- [219] K.-i. Harada, K. Fujii, T. Shimada, M. Suzuki, H. Sano, K. Adachi, W. W. Carmichael, *Tetrahedron Lett.* **1995**, *36*, 1511–1514, DOI 10.1016/0040-4039(95)00073-1.
- [220] D. E. Williams, M. Craig, C. F. B. Holmes, R. J. Andersen, *J. Nat. Prod.* **1996**, *59*, 570–575, DOI 10.1021/np9601081.

Bibliography

- [221] J. Fastner, M. Erhard, H. von Döhren, *Appl. Environ. Microbiol.* **2001**, *67*, 5069–5076, DOI 10.1128/aem.67.11.5069-5076.2001.
- [222] E. Entfellner, M. Frei, G. Christiansen, L. Deng, J. Blom, R. Kurmayer, *Front. Microbiol.* **2017**, *8*, 219, DOI 10.3389/fmicb.2017.00219.
- [223] L. Spooft, A. Blaszczyk, J. Meriluoto, M. Ceglowska, H. Mazur-Marzec, *Mar. Drugs* **2015**, *14*, 8, DOI 10.3390/md14010008.
- [224] S. T. Lima, D. O. Alvarenga, A. Etchegaray, D. P. Fewer, J. Jokela, A. M. Varani, M. Sanz, F. A. Dörr, E. Pinto, K. Sivonen, M. F. Fiore, *ACS Chem. Biol.* **2017**, *12*, 769–778, DOI 10.1021/acscchembio.6b00948.
- [225] S. Kodani, S. Suzuki, K. Ishida, M. Murakami, *FEMS Microbiol. Lett.* **1999**, *178*, 343–348, DOI 10.1111/j.1574-6968.1999.tb08697.x.
- [226] M. Murakami, S. Suzuki, Y. Itou, S. Kodani, K. Ishida, *J. Nat. Prod.* **2000**, *63*, 1280–2, DOI 10.1021/np000120k.
- [227] S. Repka, M. Koivula, V. Harjunpa, L. Rouhiainen, K. Sivonen, *Appl. Environ. Microbiol.* **2004**, *70*, 4551–4560, DOI 10.1128/aem.70.8.4551-4560.2004.
- [228] A. Bubik, B. Sedmak, M. Novinec, B. Lenarčič, T. T. Lah, *Biol. Chem.* **2008**, *389*, DOI 10.1515/bc.2008.153.
- [229] E. Zafrir-Ilan, S. Carmeli, *Tetrahedron* **2010**, *66*, 9194–9202, DOI 10.1016/j.tet.2010.09.067.
- [230] N. Halland, M. Brönstrup, J. Czech, W. Czechtizky, A. Evers, M. Follmann, M. Kohlmann, M. Schiell, M. Kurz, H. A. Schreuder, C. Kallus, *J. Med. Chem.* **2015**, *58*, 4839–44, DOI 10.1021/jm501840b.
- [231] A. Redlitz, A. K. Tan, D. L. Eaton, E. F. Plow, *J. Clin. Invest.* **1995**, *96*, 2534–2538, DOI 10.1172/jci118315.
- [232] L. Bajzar, *Arter. Thromb. Vasc. Biol.* **2000**, *20*, 2511–2518, DOI 10.1161/01.atv.20.12.2511.
- [233] T. H. Kim, J. S. Park, S. S. An, H. Kang, *J. Surg. Res.* **2015**, *193*, 560–566, DOI 10.1016/j.jss.2014.07.056.
- [234] H. Schreuder, A. Liesum, P. Lönze, H. Stump, H. Hoffmann, M. Schiell, M. Kurz, L. Toti, A. Bauer, C. Kallus, C. Klemke-Jahn, J. Czech, D. Kramer, H. Enke, T. H. J. Niedermeyer, V. Morrison, V. Kumar, M. Brönstrup, *Sci. Rep.* **2016**, *6*, DOI 10.1038/srep32958.
- [235] C. Blanco-Aparicio, M. A. Molina, E. Fernández-Salas, M. L. Frazier, J. M. Mas, E. Querol, F. X. Avilés, R. de Llorens, *J. Biol. Chem.* **1998**, *273*, 12370–12377, DOI 10.1074/jbc.273.20.12370.
- [236] T. Rohrlack, G. Christiansen, R. Kurmayer, *Appl. Environ. Microbiol.* **2013**, *79*, 2642–2647, DOI 10.1128/aem.03499-12.
- [237] J. F. Blom, H. I. Baumann, G. A. Codd, F. Jüttner, *Arch. Hydrobiol.* **2006**, *167*, 547–559, DOI 10.1127/0003-9136/2006/0167-0547.
- [238] S. Dirren, M. M. Salcher, J. F. Blom, M. Schweikert, T. Posch, *Protist* **2014**, *165*, 745–758, DOI 10.1016/j.protis.2014.08.004.
- [239] G. Christiansen, B. Philmus, T. Hemscheidt, R. Kurmayer, *J. Bacteriol.* **2011**, *193*, 3822–3831, DOI 10.1128/jb.00360-11.
- [240] J. C. Kehr, D. Gatte Picchi, E. Dittmann, *Beilstein J. Org. Chem.* **2011**, *7*, 1622–35, DOI 10.3762/bjoc.7.191.
- [241] M. Welker, H. Von Döhren, *FEMS Microbiol. Rev.* **2006**, *30*, 530–563, DOI 10.1111/j.1574-6976.2006.00022.x.
- [242] H. Kaljunen, S. H. H. Schiefelbein, D. Stummer, S. Kozak, R. Meijers, G. Christiansen, A. Rentmeister, *Angew. Chem. Int. Ed.* **2015**, *54*, 8833–8836, DOI 10.1002/anie.201503275.

- [243] H. Harms, K. L. Kurita, L. Pan, P. G. Wahome, H. He, A. D. Kinghorn, G. T. Carter, R. G. Linington, *Bioorg. Med. Chem. Lett.* **2016**, *26*, 4960–4965, DOI 10.1016/j.bmcl.2016.09.008.
- [244] L. Rouhiainen, J. Jokela, D. P. Fewer, M. Urmann, K. Sivonen, *Chem. Biol.* **2010**, *17*, 265–73, DOI 10.1016/j.chembiol.2010.01.017.
- [245] O. Grach-Pogrebinsky, S. Carmeli, *Tetrahedron* **2008**, *64*, 10233–10238, DOI 10.1016/j.tet.2008.08.015.
- [246] K. Koketsu, S. Mitsuhashi, K. Tabata, *Appl. Environ. Microbiol.* **2013**, *79*, 2201–8, DOI 10.1128/AEM.03596-12.
- [247] A. A. Esteves-Ferreira, J. H. F. Cavalcanti, M. Vaz, L. V. Alvarenga, A. Nunes-Nesi, W. L. Araujo, *Genet. Mol. Biol.* **2017**, *40*, 261–275, DOI 10.1590/1678-4685-GMB-2016-0050.
- [248] R. Koch, A. Kupczok, K. Stucken, J. Ilhan, K. Hammerschmidt, T. Dagan, *BMC Evol. Biol.* **2017**, *17*, DOI 10.1186/s12862-017-1053-5.
- [249] C. Wagner, PhD thesis, **2008**.
- [250] R. E. Moore, C. Cheuk, G. M. L. Patterson, *J. Am. Chem. Soc.* **1984**, *106*, 6456–6457, DOI 10.1021/ja00333a079.
- [251] J. M. Richter, Y. Ishihara, T. Masuda, B. W. Whitefield, T. Llamas, A. Pohjakallio, P. S. Baran, *J. Am. Chem. Soc.* **2008**, *130*, 17938–17954, DOI 10.1021/ja806981k.
- [252] K. Stratmann, R. E. Moore, R. Bonjouklian, J. B. Deeter, G. M. L. Patterson, S. Shaffer, C. D. Smith, T. A. Smitka, *J. Am. Chem. Soc.* **1994**, *116*, 9935–9942, DOI 10.1021/ja00101a015.
- [253] M. L. Micallef, P. M. D'Agostino, D. Sharma, R. Viswanathan, M. C. Moffitt, *BMC Genomics* **2015**, *16*, 669, DOI 10.1186/s12864-015-1855-z.
- [254] M. F. Fiore, D. B. Genuario, C. S. da Silva, T. K. Shishido, L. A. Moraes, R. Cantusio Neto, M. E. Silva-Stenico, *Toxicon* **2009**, *53*, 754–61, DOI 10.1016/j.toxicon.2009.02.010.
- [255] S. Cires, C. Alvarez-Roa, S. A. Wood, J. Puddick, V. Loza, K. Heimann, *Toxicon* **2014**, *88*, 62–6, DOI 10.1016/j.toxicon.2014.06.010.
- [256] L. Hagmann, F. Jüttner, *Tetrahedron Lett.* **1996**, *37*, 6539–6542, DOI 10.1016/0040-4039(96)01445-1.
- [257] U. Papke, E. M. Gross, W. Francke, *Tetrahedron Lett.* **1997**, *38*, 379–382, DOI 10.1016/s0040-4039(96)02284-8.
- [258] T. A. Smitka, R. Bonjouklian, L. Doolin, N. D. Jones, J. B. Deeter, W. Y. Yoshida, M. R. Prinsep, R. E. Moore, G. M. L. Patterson, *J. Org. Chem.* **1992**, *57*, 857–861, DOI 10.1021/jo00029a014.
- [259] A. Raveh, S. Carmeli, *J. Nat. Prod.* **2007**, *70*, 196–201, DOI 10.1021/np060495r.
- [260] S. Mo, A. Kronic, G. Chlipala, J. Orjala, *J. Nat. Prod.* **2009**, *72*, 894–899, DOI 10.1021/np800751j.
- [261] S. Mo, A. Kronic, B. D. Santarsiero, S. G. Franzblau, J. Orjala, *Phytochem.* **2010**, *71*, 2116–2123, DOI 10.1016/j.phytochem.2010.09.004.
- [262] U. M. Acuna, J. Zi, J. Orjala, E. J. Carcache de Blanco, *Int. J. Cancer Res. (Tortola)* **2015**, *49*, 1655–1662.
- [263] B. S. Falch, G. M. König, A. D. Wright, O. Sticher, H. Ruegger, G. Bernardinelli, *J. Org. Chem.* **1993**, *58*, 6570–6575, DOI 10.1021/jo00076a013.
- [264] M. Omari, PhD thesis, **2011**.
- [265] B. S. Falch, G. M. König, A. D. Wright, O. Sticher, C. K. Angerhofer, J. M. Pezzuto, H. Bachmann, *Planta. Med.* **1995**, *61*, 321–8, DOI 10.1055/s-2006-958092.
- [266] A. D. Wright, O. Papendorf, G. M. König, A. Oberemm, *Chemosphere* **2006**, *65*, 604–8, DOI 10.1016/j.chemosphere.2006.02.004.

Bibliography

- [267] A. D. Wright, O. Papendorf, G. M. König, *J. Nat. Prod.* **2005**, *68*, 459–61, DOI 10.1021/np049640w.
- [268] H. Aldemir, R. Richarz, T. Gulder, *Angew. Chem. Int. Ed.* **2014**, *53*, 8286–8293, DOI 10.1002/anie.201401075.
- [269] Y. Zhang, F. Buchholz, J. P. Muyrers, A. F. Stewart, *Nat. Genet.* **1998**, *20*, 123–8, DOI 10.1038/2417.
- [270] J. P. Muyrers, Y. Zhang, G. Testa, A. F. Stewart, *Nucleic Acids Res.* **1999**, *27*, 1555–7, DOI 10.1093/nar/27.6.1555.
- [271] D. Yu, H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, D. L. Court, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5978–83, DOI 10.1073/pnas.100127597.
- [272] Y. Zhang, J. P. Muyrers, G. Testa, A. F. Stewart, *Nat. Biotechnol.* **2000**, *18*, 1314–7, DOI 10.1038/82449.
- [273] S. K. Sharan, L. C. Thomason, S. G. Kuznetsov, D. L. Court, *Nat. Protoc.* **2009**, *4*, 206–23, DOI 10.1038/nprot.2008.227.
- [274] H. Wang, Z. Li, R. Jia, Y. Hou, J. Yin, X. Bian, A. Li, R. Müller, A. F. Stewart, J. Fu, Y. Zhang, *Nat. Protoc.* **2016**, *11*, 1175–90, DOI 10.1038/nprot.2016.054.
- [275] J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Müller, A. F. Stewart, Y. Zhang, *Nat. Biotechnol.* **2012**, *30*, 440–6, DOI 10.1038/nbt.2183.
- [276] H. Wang, X. Bian, L. Xia, X. Ding, R. Müller, Y. Zhang, J. Fu, A. F. Stewart, *Nucleic Acids Res.* **2014**, *42*, e37, DOI 10.1093/nar/gkt1339.
- [277] N. Kouprina, V. Larionov, *Nat. Rev. Genet.* **2006**, *7*, 805–12, DOI 10.1038/nrg1943.
- [278] N. Kouprina, V. Larionov, *Nat. Protoc.* **2008**, *3*, 371–7, DOI 10.1038/nprot.2008.5.
- [279] J. H. Kim, Z. Feng, J. D. Bauer, D. Kallifidas, P. Y. Calle, S. F. Brady, *Biopolymers* **2010**, *93*, 833–44, DOI 10.1002/bip.21450.
- [280] K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein, B. S. Moore, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1957–62, DOI 10.1073/pnas.1319584111.
- [281] H. Wang, Z. Li, R. Jia, J. Yin, A. Li, L. Xia, Y. Yin, R. Müller, J. Fu, A. F. Stewart, Y. Zhang, *Nucleic Acids Res.* **2017**, *46*, e28, DOI 10.1093/nar/gkx1296.
- [282] W. Jiang, T. F. Zhu, *Nat. Protoc.* **2016**, *11*, 960–75, DOI 10.1038/nprot.2016.055.
- [283] J. Quan, J. Tian, *Nat. Protoc.* **2011**, *6*, 242–51, DOI 10.1038/nprot.2010.181.
- [284] F. Zeng, J. Zang, S. Zhang, Z. Hao, J. Dong, Y. Lin, *BMC Biotechnol.* **2017**, *17*, 81, DOI 10.1186/s12896-017-0394-x.
- [285] Y. Itou, S. Suzuki, K. Ishida, M. Murakami, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1243–6, DOI 10.1016/S0960-894X(99)00191-2.
- [286] K. Wenke, M. Kai, B. Piechulla, *Planta. Med.* **2010**, *231*, 499–506, DOI 10.1007/s00425-009-1076-2.
- [287] U. Effmert, J. Kalderas, R. Warnke, B. Piechulla, *J. Chem. Ecol.* **2012**, *38*, 665–703, DOI 10.1007/s10886-012-0135-5.
- [288] K. Wenke, T. Weise, R. Warnke, C. Valverde, D. Wanke, M. Kai, B. Piechulla in *Biocommunication of Plants*, (Eds.: G. Witzany, F. Baluška), Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 327–347.
- [289] M. Kai, E. Crespo, S. M. Cristescu, F. J. Harren, W. Francke, B. Piechulla, *Appl. Microbiol. Biotechnol.* **2010**, *88*, 965–76, DOI 10.1007/s00253-010-2810-1.
- [290] V. Agarwal, B. S. Moore, *ACS Chem. Biol.* **2014**, *9*, 1980–4, DOI 10.1021/cb5004338.

List of Abbreviations

4-HBA	4-hydroxybenzoic acid
AFEAP	assembly of fragment ends after PCR
BGC	biosynthetic gene cluster
CATCH	Cas9-assisted targeting of chromosome segments
CoA	coenzyme A
CPEC	circular polymerase extension cloning
CYP P450	cytochrome P450
DCP	dichlorophenol
DiPaC	Direct Pathway Cloning
DMAPP	dimethylallyl pyrophosphate
ExoCET	exonuclease combined recombineering
FPP	farnesyl pyrophosphate
GFP	green fluorescent protein
HAB	harmful algal bloom
IPP	isopentenyl pyrophosphate
LCHR	linear-circular homologous recombineering
L-Hph	L-homophenylalanine
L-Hty	L-homotyrosine
LLHR	liner-linear homologous recombineering
MCYST	microcystin
NCE	new chemical entity

List of Abbreviations

NRPS	nonribosomal peptide synthetase
PBDE	polybrominated diphenyl ether
PCP	peptide carrier protein
PCR	polymerase chain reaction
PKS	polyketide synthase
SAM	<i>S</i> -adenosyl methionine
SLIC	Sequence and Ligation Independent Cloning
TAR	transformation-associated recombination
VOC	volatile organic compound

List of Figures

1	Pharmaceutically active secondary metabolites found in healing plants used by various ancient cultures.	2
2	Structures of natural products isolated in their pure form for the first time in the 19th century.	3
3	Important antibiotic substances discovered in the early 20th century. . . .	4
4	Structures of important natural product pharmaceuticals discovered in the 1970s.	5
5	Modern drugs developed and approved within recent years.	6
6	Structures of bacterial volatile organic compounds (VOCs).	9
7	Further VOCs of bacterial origin.	10
8	Cyanobacterial natural products with high pharmacological potential. . . .	12
9	Cyanobacterial blooms are often enriched with toxic natural products. . . .	13
10	Polybrominated aromatic compounds produced by marine (cyano-)bacteria.	15
11	Morphologic characteristics of subsection V cyanobacteria.	17
12	Secondary metabolites isolated from the genus <i>Fischerella</i>	18
13	Chlorinated natural products synthesized by <i>F. ambigua</i> 108b.	19

List of Schemes

1	Sodorifen biosynthesis.	10
2	Selected anabaenopeptins containing the characteristic ureido motif (blue) and non-proteinogenic homoamino acids (green) and their schematic NRPS assembly.	16
3	Accessible reactive positions in radical mediated enzymatic oxidative coupling and halogenation.	20
4	DiPaC strategy for heterologous production of anabaenopeptin NZ857 (54) and nostamide A (55) from <i>N. punctiforme</i> in <i>E. coli</i> BAP1.	24
5	Heterologous expression of the <i>S. plymuthica</i> WS3236 <i>sod</i> BGC in <i>E. coli</i> BL21 leads to efficient sodorifen (38) production.	37
6	Experimental strategies to investigate the ambigol biosynthesis.	50
7	Putative trimeric Ab3 reaction products.	64

Attachments

A.1 Supplemental materials of publications and submitted manuscript

A.1.1 Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential

The supplemental information is related to the following publication which is included in chapter 3.1:

- C. Greunke*, **E. R. Duell***, P. M. D'Agostino*, A. Glöckle, K. Lamm, T. A. M. Gulder: Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Met. Eng.* **2018**, *47*, 334–345, DOI 10.1016/j.ymben.2018.03.010
*equally contributing authors

Direct Pathway Cloning (DiPaC) to Access Natural Product Biosynthetic Gene Clusters

Christian Greunke,^a Elke R. Duell,^a Paul M. D'Agostino,^a Anna Glöckle, Katharina Lamm and Tobias A. M. Gulder^b

Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technische Universität München, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

^aThese authors contributed equally. ^bCorresponding author; E-mail: tobias.gulder@ch.tum.de.

Supporting Information

Table of contents

	Page
1. Materials and Methods	2
1.1 Table of primers	2
1.2 Table of strains	6
1.3 Table of plasmids	7
2. Results	8
2.1 Phenazines	8
2.2 Anabaenopeptins	17
2.3 Erythromycin	22
4. References	32

1. Material and Methods

1.1. Table of primers

Table S1. List of oligonucleotides used for cloning and screening procedures.

Oligonucleotide	Sequence (5' → 3') [†]	Description
Vector primers		
T7 primer	TAATACGACTCACTATAGGG	Colony screening and sequencing of pET28PD1::ehp and pETDuet-1::apt-hphB-CD
pRSET-RPnew	GGGTTATGCTAGTTATTGC	Colony screening and sequencing of pET28PD1::ehp, pET-28b-ptetO::apt-NRPS and all pETDuet-1::apt variants
pET-RP	CTAGTTATTGCTCAGCGG	Colony screening and sequencing of (pET-28b backbone)
lac_prom_pSET152	CTTCCGGCTCGTATGTTG	Colony screening and sequencing of (pCCIFOS backbone)
screen_ptetF	TGGCGTGTTAAATCACTTTAC	Colony screening and sequencing of pET-28b-ptetO::apt-NRPS
bom_seq_rev_1	CCGCACAGATGCGTAAG	Sequencing of pETDuet-2::apt-hph-aptF
bla_seq_rev	CGCCAGTTAATAGTTTGGC	Colony screening of pETDuet-1/2::apt-hph-aptF
ColE1_seq_rev	GCAGAGCGAGGTATGTAG	Colony screening of pETDuet-1/2::apt-hph-aptF
Cloning of ehp		
gib_ehp-phenF	ttttgtttaactttaagaaggagataaccATGTTCAAAACCTAAC AGTGTATT	Amplification of ehp cluster with 30 bp Gibson assembly homology sequence
gib_ehp-phenR	caataaatatcgttatcctccatgtccaaaTGTCGTTATGAGATTG TCGAG	Amplification of ehp cluster with 30 bp Gibson assembly homology sequence
pET28PD1-vectF	ATTTGGACATGGAGGATAACG	Amplification of pET-28b-SUMO backbone to yield pET28PD1 for subsequent Gibson assembly
pET28PD1-vectR	GGTATATCTCCTTCTTAAAGTTAAACAAAA	Amplification of pET-28b-SUMO backbone to yield pET28PD1 for subsequent Gibson assembly
screen_phen1_CtermF	CGCAGGGAATAAAAAGAGATG	Colony screening and sequencing of pET-28b-PD1::ehp
screen_phen1_NtermR	ATGAAAAATCGGTCCTGCTAC	Colony screening and sequencing of pET-28b-PD1::ehp
seq_PhenSCDF1	CAGAAAAATGATGCAAGGC	Sequencing of pET-28b-PD1::ehp
seq_phzBF2	GAGTAAAAAGCAACTGGC	Sequencing of pET-28b-PD1::ehp
seq_phzDF3	CCTATTCAGAGATATCGAAG	Sequencing of pET-28b-PD1::ehp
seq_phzEF4	CGATGGTGCTAGATGAAGAG	Sequencing of pET-28b-PD1::ehp
seq_phzEF5	TAGTCGTAGCCGCAAG	Sequencing of pET-28b-PD1::ehp

seq_phzFF9
 seq_phzFF10
 seq_phen1R7
 seq_phen1R6
 seq_phen1R4

GATCTAGAGCCGGAAACC
 CTCTAGCGCTTCACTTAG
 TAGGAATCAGCGCCAAC
 ACAGTCACTCCGCAACC
 TTCTCCGGCAGTTGTTG

Sequencing of pET-28b-PD1::ehp
 Sequencing of pET-28b-PD1::ehp
 Sequencing of pET-28b-PD1::ehp
 Sequencing of pET-28b-PD1::ehp
 Sequencing of pET-28b-PD1::ehp

Cloning of *apt*

pTet_GA_apt-NRPS_for

pTet_GA_apt-NRPS_rev

aptA_screen_rev_1

aptD_screen_for_1

aptA_seq_for_1

aptA_seq_rev_1

aptA_seq_for_2

aptA_seq_for_3

aptA_seq_for_4

aptA_seq_for_5

aptA_seq_for_6

aptA_seq_for_7

aptB_seq_for_1

aptB_seq_rev_1

aptB_seq_rev_2

aptB_seq_rev_3

aptB_seq_rev_4

aptC_seq_for_1

aptC_seq_for_2

aptC_seq_for_3

aptC_seq_for_4

aptC_seq_for_5

aptC_seq_for_6

aptC_seq_rev_1

aptC_seq_rev_2

aptC_seq_rev_3

gatagagaaggatcgaccATGGAAGAAAACATATCAATTAACATC

ctgtgatgatgatgatggctgctcatttctgtcctgggt
 TTC

TCGACGCAAAATGCCTAC

ATTCAATGAAAATCGTTTTGC

CCGGAGAAGCTTGTAATCC

GACCCGCTATGGTAGGG

CAGCTTCCTTTGCTCAAG

GAAAATCCACGTTTCGGG

ATTAGCGAAGTTCAGTCCC

CCAGCGATGAATTTTGGC

GCAAGTCAGCATCCACG

GCTTTGCCACTAGACTATCC

GTTGCTTTGCATCCTCTC

TTTTTTGGCGGAGTTCAGC

GGGAGCAACATCCAGTTC

CTGTACATCCGGAAGGGC

CCTGCTAGCAATGTTTTGG

TTGGCCGATCAAACTCATC

ATCGCTTTCATCATCCC

AAACACGGGGTCACTTG

GCTTAGATGCCTACGCTC

CCCAGGCTAGTAGCTTATG

AGAACGCAGTCTGAGCC

GGGTAGTTCCTGATTCCTG

CTTGCATTTGTCAGTTGCG

TTTACGCAACGCCAGTTC

Amplification of *aptA-D* for insertion into pET-28b-ptetO

Amplification of *aptA-D* for insertion into pET-28b-ptetO

Colony screening of pET-28b-ptetO::*apt-NRPS*

Colony screening of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

Sequencing of pET-28b-ptetO::*apt-NRPS*

aptC_seq_rev_4	GCTTCGGGTTCTATGCC	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptC_seq_rev_5	AGACTTCAAAAGTGTCCCG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptC_seq_rev_6	TCCTTGTTGTTGATGGTAGG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptD_seq_rev_1	TCAGGGAGTGCCTAAG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptD_seq_rev_2	GGTTGACCAATCGGTGG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptD_seq_rev_3	GGTTTTTGGCAACTTGGG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
aptD_seq_rev_4	GTTGCCAGTTTGGCGTAG	Sequencing of pET-28b-ptetO:: <i>apt-NRPS</i>
apt-hphB_BamHI_for	ataggatccgATGGAAATCTTTAAGTAAGGCATC	Amplification of <i>hphB-CD</i> for restriction based cloning
apt-hphCD_NotI_rev	atatcgggccgCTTAATCAAAAGCGATCGCTAAATTTTC	Amplification of <i>hphB-CD</i> for restriction based cloning
hphB_screen&seq_rev_1	TACCGTGAATGGCCCTCG	Colony screening and sequencing of pETDuet-1:: <i>apt-hphB-CD</i>
hphB_seq_for_1	CGCCTTACCTCATTTGC	Sequencing of pETDuet-1:: <i>apt-hphB-CD</i>
hphCD_seq_for_1	GAATCCTGTAGTTGATGCG	Sequencing of pETDuet-1:: <i>apt-hphB-CD</i>
apt-hphA_NdeI_for	atatacatATGGAAACCCCTTCCTATC	Amplification of <i>hphA</i> for restriction based cloning
apt-hphA_PmeI_XhoI_rev	atatactcgaggtttaaacTTATTTAAACACCCCATTTTTC	Amplification of <i>hphA</i> for restriction based cloning
hphA_screen_for_1	CTTCGGTAAATGGAGTGG	Colony screening of pETDuet-1:: <i>apt-hph</i>
hphA_seq_rev_1	CCAAACGGTATCACAGAGC	Sequencing of pETDuet-1:: <i>apt-hph</i>
hph-bb_GA_apt-aptF_for	tggggtgttaaaataagttTATGCAAGCCAAAACCGTTC	Amplification of <i>aptF</i> for insertion into pETDuet-1:: <i>apt-hph</i>
hph-bb_GA_apt-aptF_rev	ctttaccagactcgaggtttTTAAGCTCTTACCATTTCAAC	Amplification of <i>aptF</i> for insertion into pETDuet-1:: <i>apt-hph</i>
hphA_screen&seq_for_1	CTTCGGTAAATGGAGTGG	Colony screening and sequencing of pETDuet-1:: <i>apt-hph-aptF</i>
aptF_seq_for_1	GTGCTGGAAAATGGCAAC	Sequencing of pETDuet-1:: <i>apt-hph-aptF</i>
pETDuet-1_GA_ColE1_for	CCTTTGATCTTTTCTACGGGGTC	Amplification of pETDuet-1 for insertion of ColA
pETDuet-1_GA_ColE1_rev	ACGCCAGCAACGGCGCCT	Amplification of pETDuet-1 for insertion of ColA
pETDuet-1_GA_oriColA_for	aaagcccgcttgctggcgtTGGTGTCCGGGAAATCCCGTAAAG	Amplification of ColA for insertion into pETDuet-1
pETDuet-1_GA_oriColA_rev	cccgtagaaaaagatcaaaaggAAACGTCCTAGAAAGATGCCAG	Amplification of ColA for insertion into pETDuet-1
Cloning of ery		
screen_ery_for	GTCATCGAAGGCATGACGAC	Screening primer for <i>ery</i>
screen_ery_rev	GTTCTACCGCAACGAGAACC	Screening primer for <i>ery</i>
ery-p1_for	ctcacagagaacagattggtggatccgTGGTCTTCGACGGGGGAGC	Amplification of <i>ery-p1</i> for insertion into pET-28b-SUMO
ery-p1_rev	TggcgccgcttactagtgtaTCGATCGAGATGTGCACG	Amplification of <i>ery-p1</i> for insertion into pET-28b-SUMO
screen_ery-p1_for	TGATCCCATTCACCCGGAG	Colony screening and sequencing of

ery-p2_for	tgcgTgcacatctcgatcgaTTTCCCGATGCGGGTGTGC	Amplification of <i>ery</i> -p2 for insertion into pET-28b-SUMO:: <i>ery</i> -p1
ery-p2_rev	actggcggccgttactagtgtagcgtacgtaGCAATCCCGCCCTTTAG	Amplification of <i>ery</i> -p2 for insertion into pET-28b-SUMO:: <i>ery</i> -p1
screen_ery-p2_for	CTTCGCCGAGTCTACAC	Colony screening and sequencing of
ery-p3_for	CCCTAAAGGGCGGGATTGC	Amplification of <i>ery</i> -p3 for insertion into pET-28b-SUMO:: <i>ery</i> -p12
ery-p3_rev	actggcggccgttactagtgtagcgtactacgtaCTACAGGTCCTCTCCCGCC	Amplification of <i>ery</i> -p3 for insertion into pET-28b-SUMO:: <i>ery</i> -p12
screen_ery-p3_for	TGGACGAGCTGGAGAAG	Colony screening and sequencing of
ery-p4_for	gcgggggagagagacctgtagATGAGCGGTGACAACGGC	Amplification of <i>ery</i> -p4 for insertion into pET-28b-SUMO:: <i>ery</i> -p123
ery-p4_rev	actggcggccgttactagtgtagcgtatCATGAATTCCTCTCCGCC	Amplification of <i>ery</i> -p4 for insertion into pET-28b-SUMO:: <i>ery</i> -p123
screen_ery-p4_for	CCTGATCGACGTCTACCC	Colony screening and sequencing of
SUMO_ery-p5_for	ctcacagagaacagattggtggatccCGACCACTTCAACGATGGTG	Amplification of <i>ery</i> -p5 for subcloning into pET-28b-SUMO
SUMO_ery-p5_rev	tggcggccgttactagtgTCAAGCCCCAGCCTTGAGG	Amplification of <i>ery</i> -p5 for subcloning into pET-28b-SUMO
screen_ery-p5_for	CCTTACAGCTTCTACCCGG	Colony screening and sequencing of
ery-p1234_HR_LP_pCC1FOS_for	gagtgcacctgcaggcatgcGTGGTCTTCGACGGGGAGCT	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_LEFT
ery-p1234_HR_LP_rev	atgaccgcataaagcttCAGCCAGTCGAGGAGTGC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_LEFT
ery-p1234_HR_RP_for	cgactggctgaagcttATGCGGTCAATCAGGACAC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_RIGHT
ery-p1234_HR_RP_pCC1FOS_rev	ggtgagactatagaataactcatgcatTCATGAATTCCTCCGCC	Preparation of a LCHR-compatible helper plasmid of pCC1FOS, introduction of CAP_site_RIGHT
ery-p5_for	tgggcggagggaattcatgaCCACCGACCGATCCGGCC	Amplification of <i>ery</i> -p5 for insertion into pCC1FOS:: <i>ery</i> -p1234
pCC1FOS_ery-p5_rev	ggtgagactatagaataactcatgcatTCAAGCCCCAGCCTTGAGG	Amplification of <i>ery</i> -p5 for insertion into pCC1FOS:: <i>ery</i> -p1234
pLK01_lin_for	ACCAGCATGGATAAAGGC	Preparation of a linearized fragment of pLK01 used for LCHR
pLK01_lin_rev	TGAGTATTCCTATAGTCTCACCTAAATAG	Preparation of a linearized fragment of pLK01 used for LCHR

Underlined sequences indicate restriction sites. Specific nucleotide targets indicated in uppercase. Gibson homology arm sequences are indicated in lowercase.

1.2. Table of strains

Table S2. Strains used in this study.

Strains	Description	Reference / Source
<i>E. coli</i>		
DH5 α	Host strain for cloning	NEB
TOP10	Host strain for cloning	Invitrogen
EPI300-T1	Host strain for cloning	Epicentre
GB08-red	Host strain for LCHR	(1, 2)
ET12567	Donor strain for conjugation	(3)
ET12567/pUB307	Helper strain for triparental conjugation	(4)
BAP1	Heterologous expression strain	(5)
<i>Streptomyces</i>		
<i>Streptomyces coelicolor</i> M1152	Host strain for heterologous expression	(6)
<i>Streptomyces coelicolor</i> M1154	Host strain for heterologous expression	(6)
Others		
<i>Serratia fonticola</i> DSM 4576	Natural producer of fontizine A	DSMZ
<i>Nostoc punctiforme</i> PCC 73102	Natural producer of anabaenopeptin NZ 857 and nostamide A	Institut Pasteur
<i>Saccharopolyspora erythraea</i> DSM 40517	Natural producer of erythromycin A	DSMZ

1.3. Table of plasmids

Table S3. Plasmids used and constructed in this study.

Plasmids	Description	Reference / Source
pET-28b-SUMO	Modified derivative of pET-28b used for assembly operations, Kan ^R	(7)
pET-28b-ptetO	tetracyclin inducible expression plasmid, ColE1, Kan ^R	Prof. Neilan / Prof. Müller
pETDuet-1	IPTG inducible expression plasmid with two T7-promoters, ColE1, Amp ^R	Novagen
pCOLADuet-1	IPTG inducible expression plasmid with two T7-promoters, ColA, Kan ^R	Novagen
pCC1FOS pLK01	Cloning vector used for assembly operations, Cam ^R <i>Streptomyces</i> integrative vector used for the expression of <i>ery</i> ; Kan ^R	Epicentre (8)
pET-28b-PD1:: <i>ehp</i>	Complete phenazine cluster <i>ehp</i> of <i>S. fonticola</i>	This study
pET-28b-ptetO:: <i>apt-NRPS</i>	<i>aptA-F</i> of the <i>apt</i> cluster of <i>N. punctiforme</i> , ColE1, Kan ^R	This study
pETDuet-2:: <i>apt-hph</i>	<i>hphA-CD</i> of the <i>apt</i> cluster of <i>N. punctiforme</i> , ColA, Amp ^R	This study
pETDuet-2:: <i>apt-hph-aptF</i>	<i>hphA-CD</i> and <i>aptF</i> of the <i>apt</i> cluster of <i>N. punctiforme</i> , ColA, Amp ^R	This study
dpET-28b-SUMO:: <i>ery</i> -p1	Part 1 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pET-28b-SUMO:: <i>ery</i> -p12	Parts 1 and 2 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pET-28b-SUMO:: <i>ery</i> -p123	Parts 1, 2 and 3 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pET-28b-SUMO:: <i>ery</i> -p1234	Parts 1, 2, 3 and 4 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pET-28b-SUMO:: <i>ery</i> -p5	Part 5 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pCC1FOS:: <i>HR_ery</i> -p1234	Helper plasmid used for backbone exchange by LCHR; needs to be linearized by <i>HindIII</i>	This study
pCC1FOS:: <i>ery</i> -p1234	Parts 1, 2, 3 and 4 of the <i>ery</i> cluster of <i>S. erythraea</i>	This study
pCC1FOS:: <i>ery</i>	Complete <i>ery</i> cluster of <i>S. erythraea</i>	This study
pLK01:: <i>ery</i>	Complete <i>ery</i> cluster of <i>S. erythraea</i>	This study

2. Results

2.1. Phenazines

Bioinformatic analysis

The *P. agglomerans* EhpF crystal structure was solved with bound phenazine-1,6-dicarboxylic acid (PDC) as substrate, indicating the residues involved in substrate binding. The majority of substrate binding residues are not conserved between EhpF from *P. agglomerans* and *S. fonticola* (red squares, Figure S1). Thus, we hypothesize that *S. fonticola* EhpF does not utilize PDC as a substrate but as an yet unidentified molecule that may be the product of EhpS. Conserved ATP-binding ‘P-loop’ and conserved residues are blue and green, respectively.

Table S4. Bioinformatic analysis of the *Serratia fonticola* DSM 4576 phenazine gene cluster (*ehp*) used for heterologous expression.

Enzyme	Size (aa)	Putative function
EhpR	128	Phenazine resistance protein
EhpK	270	Short chain dehydrogenase/reductase
EhpA	147	Core phenazine biosynthesis (<i>phzB</i>)
EhpB	209	Core phenazine biosynthesis (<i>phzD</i>)
EhpC	622	Core phenazine biosynthesis (<i>phzE</i>)
EhpS	117	Hypothetical protein
EhpD	281	Core phenazine biosynthesis (<i>phzF</i>)
EhpE	211	Core phenazine biosynthesis (<i>phzG</i>)
EhpF	353	AMP-binding
EhpG	459	Acyl-CoA reductase
EhpT	308	Acyltransferase

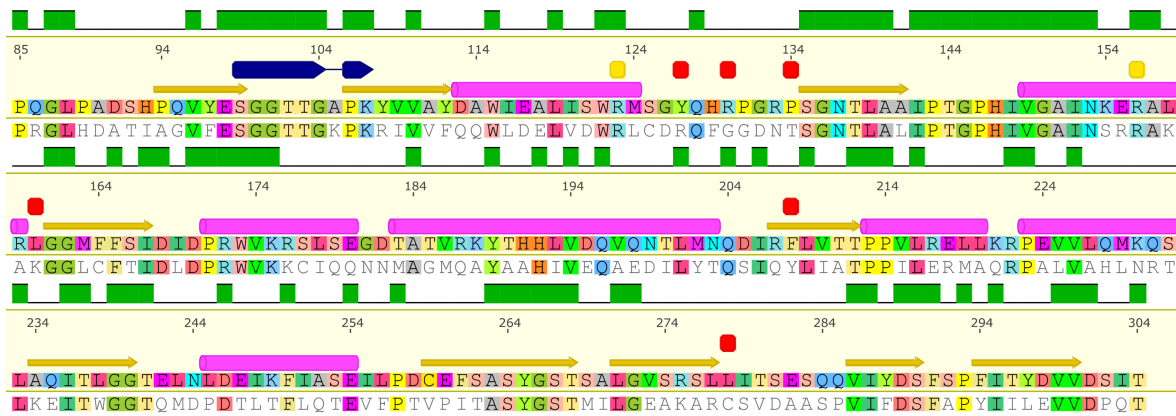


Figure S1. Alignments of EhpF (PDB ID: 3L2K) from *P. agglomerans* (upper) and *S. fonticola* (lower). Sections highlighted in the figure include: conserved amino acid residues (green), ATP-binding P-loop (blue), α -helices (yellow arrows), β -sheet (pink), conserved binding site residues (yellow box), mismatches in substrate binding site (red boxes).

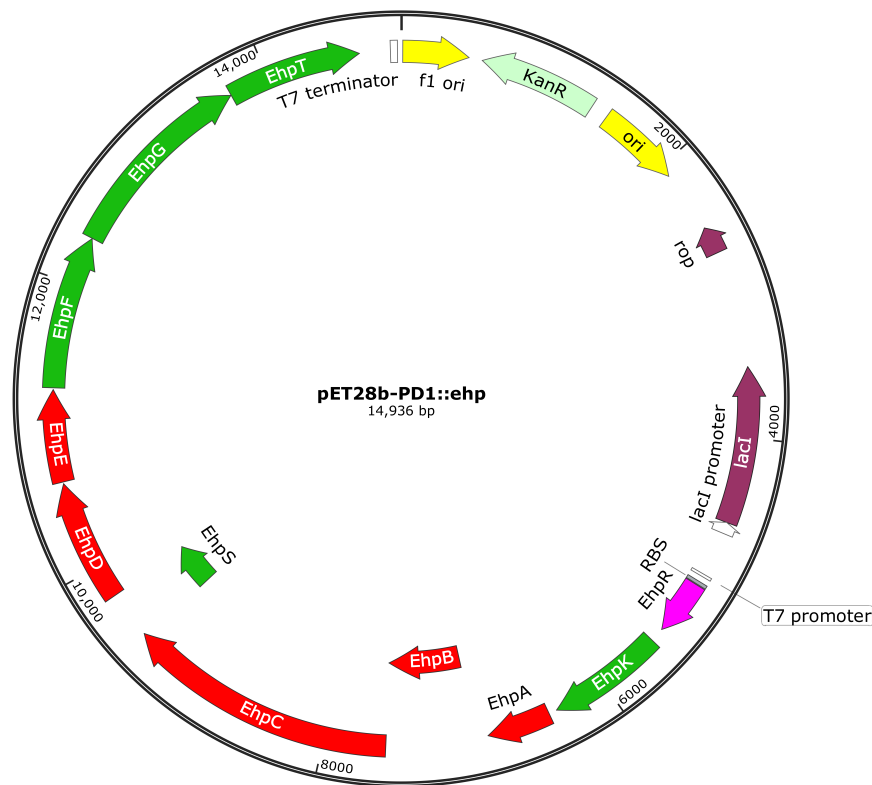


Figure S2. Vector map of the pET-28b-PD1::*ehp* expression plasmid (14,936 bp).

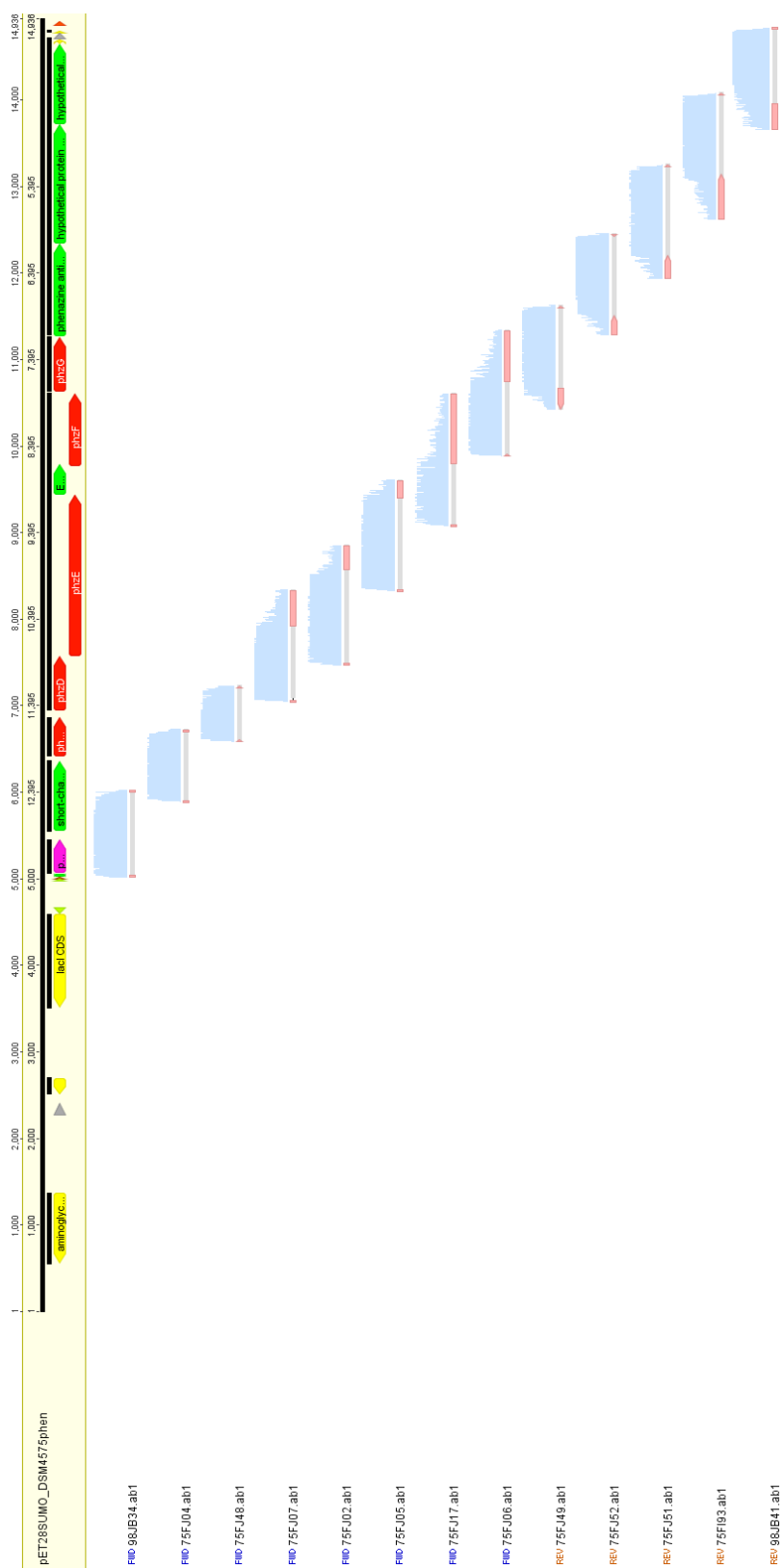


Figure S3. Schematic representation of the sequencing results for the insert *ehp* into the pET-28b-PD1 vector.

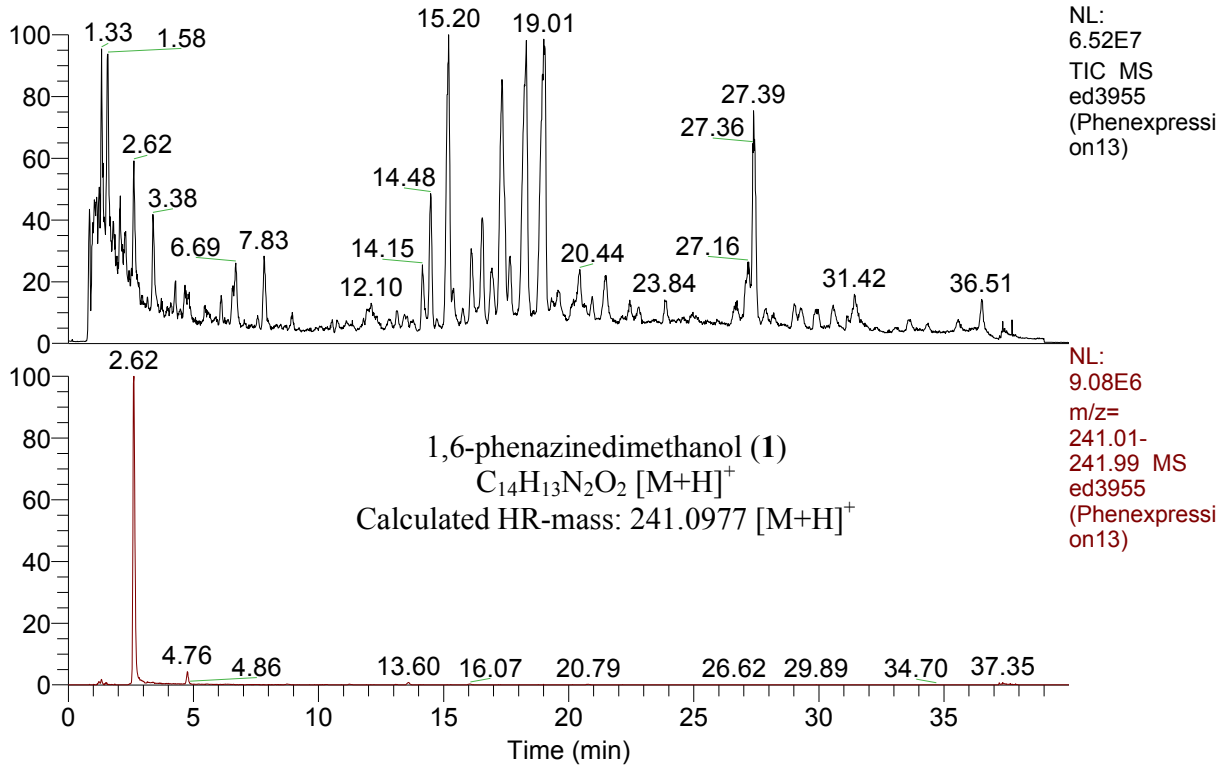
A

C:\Users\...ed3955 (Phenexpression13)
Duell/Gulder: 30-100%B30min; 10µL Inj.

13/04/2017 10:31:29 PM

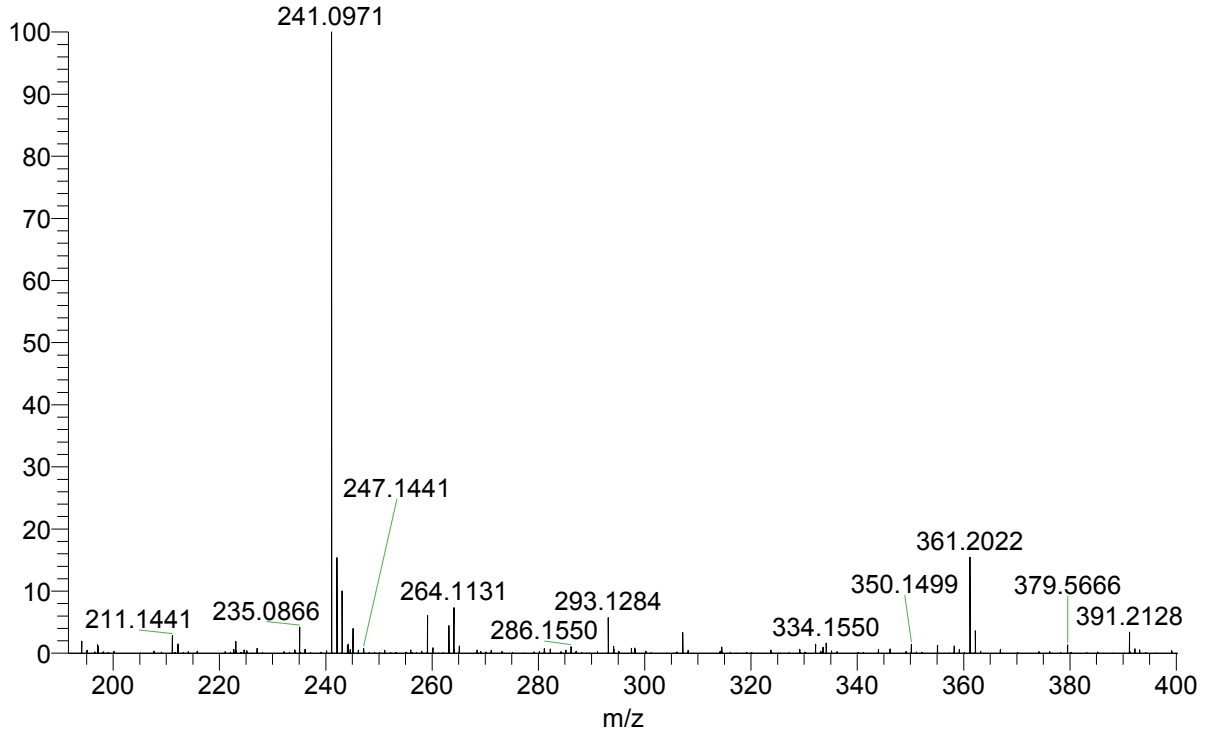
ED170403_1

RT: 0.00 - 39.99



ed3955 (Phenexpression13) #141-166 RT: 2.42-2.80 AV: 26 NL: 2.04E6

T: FTMS + p ESI Full ms [190.00-1500.00]



B

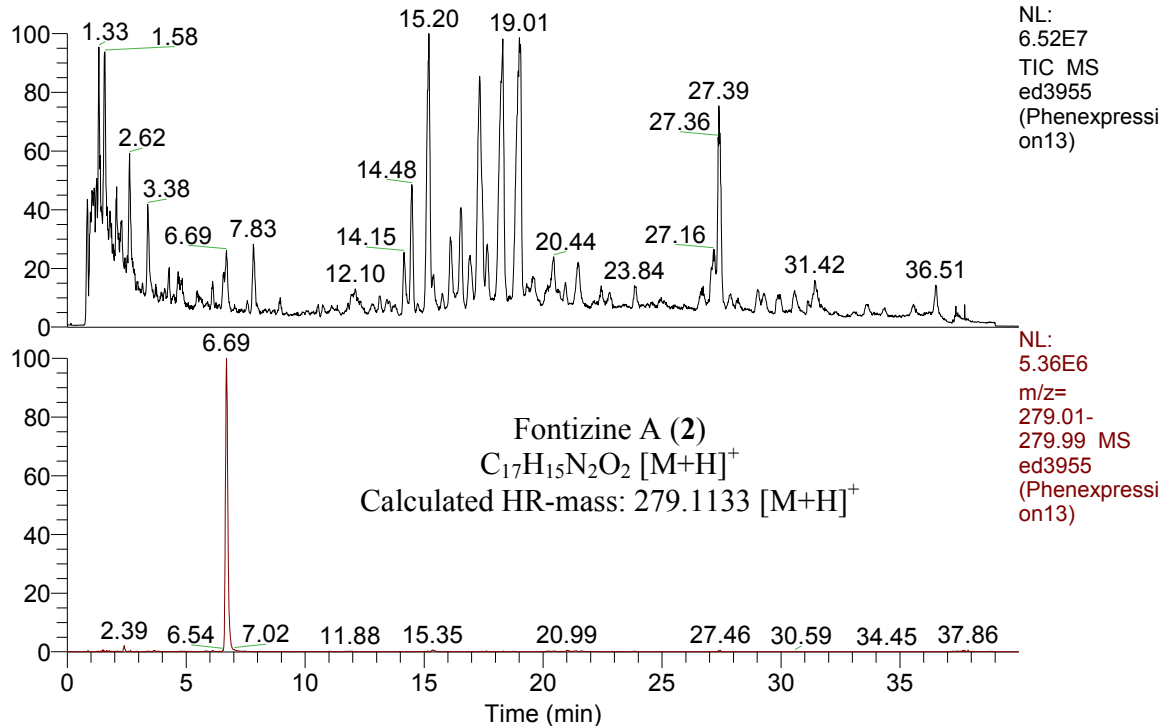
C:\Users\...ed3955 (Phenexpression13)

13/04/2017 10:31:29 PM

ED170403_1

Duell/Gulder: 30-100%B30min; 10µL Inj.

RT: 0.00 - 39.99



ed3955 (Phenexpression13) #403-425 RT: 6.49-6.83 AV: 23 NL: 1.61E6

T: FTMS + p ESI Full ms [190.00-1500.00]

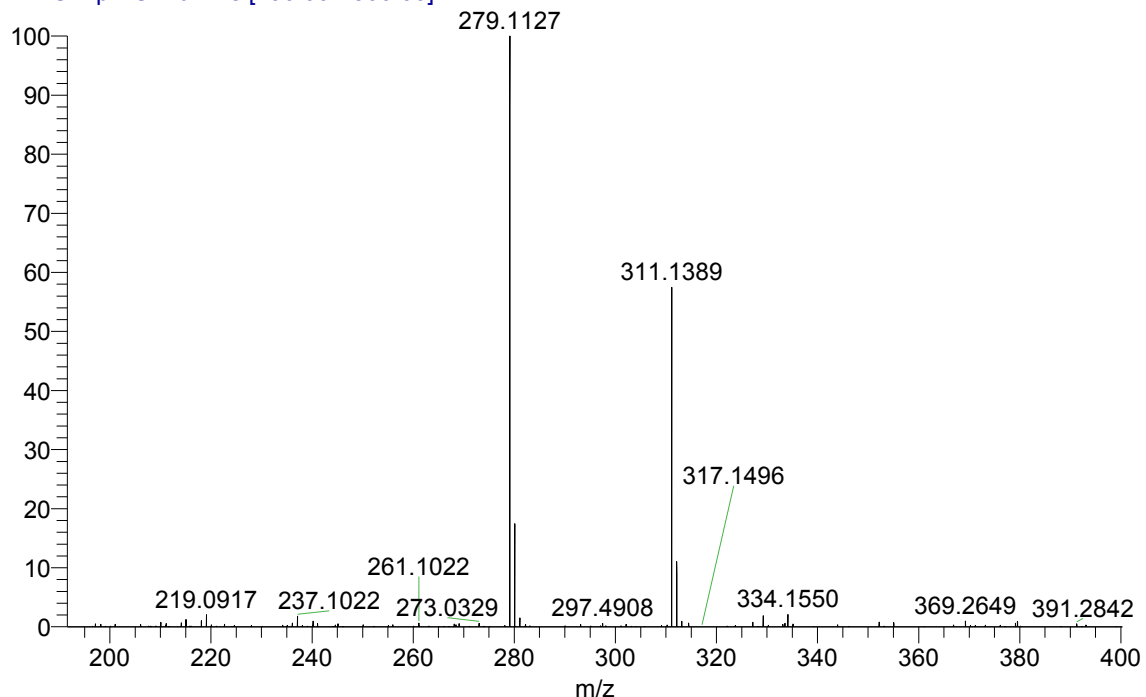


Figure S4. LC-MS results of *E. coli* BAP1 expression cultures harboring pET-28b-PD1::*ehp*. The molecules 1,6-phenazinedimethanol and fontizine A were detected with a m/z of 241.0971 and 279.1127, respectively (A and B).

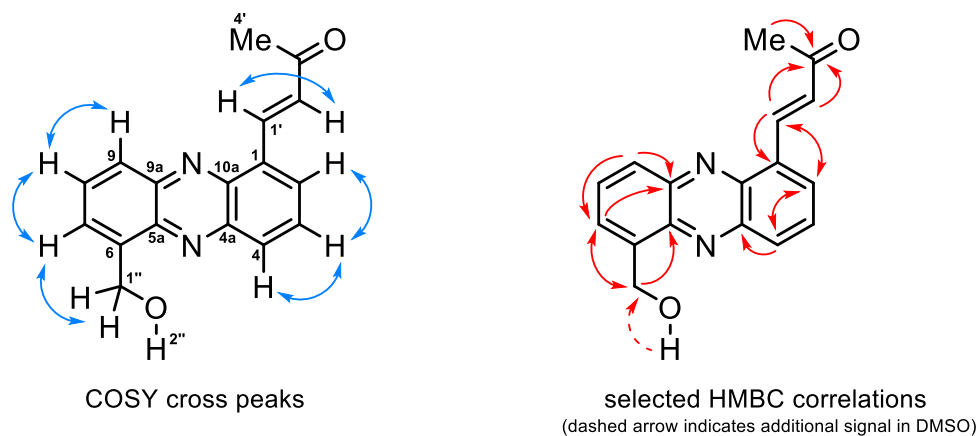


Figure S5. 2D NMR correlations of fontizine A (**2**). Key signals for the structural elucidation of Phenazine X were performed using ^1H - ^1H COSY (red), ^1H - ^{13}C HSQC and ^1H - ^{13}C HMBC (blue) NMR experiments.

Table S5. ^1H spectral data of 1,6-phenazinedimethanol (**1**) and ^1H and ^{13}C NMR spectral data of fontizine A (**2**).

Signal	1,6-phenazinedimethanol in MeOD		fontizine A in acetone- D_6				
	^1H [ppm]	Integral, Mult.	^1H [ppm]	^{13}C [ppm]	Integral, Mult.	COSY	HMBC
1				142.4			
2	7.99	1H, dd, 6.8, 1.3 Hz	8.41	129.7	1H, d, 7.0 Hz	3	1', 4, 4a
3	7.92	1H, dd, 8.7, 6.8Hz	8.00*	131.2	1H, signal overlap with 8	2, 4	**
4	8.20	1H, dd, 8.8, 1.3 Hz	8.33	132.6	1H, dd, 8.7, 1.3 Hz	3	2, 4a
4a				142.1			
5a				141.9			
6				143.5			
7	7.99	1H, dd, 6.8, 1.3 Hz	8.07	128.2	1H, dd, 6.8, 1.4 Hz	8, 1''	9, 9a, 1''
8	7.92	1H, dd, 8.7, 6.8Hz	7.99*	132.1	2H, signal overlap with 3	7, 9	**
9	8.20	1H, dd, 8.8, 1.3 Hz	8.23	129.1	1H, dd, 8.7, 1.3 Hz	8	7, 9a
9a				143.0			
10a				134.3			
1'			8.90	138.0	1H, d, 16.6 Hz	2'	1, 2, 3'
2'			7.32	130.5	1H, d, 16.6 Hz	1'	10a, 3'
3'				198.3			
4'			2.48	27.9	3H, s	-	2', 3', [1']
1''	5.45	2H, s	5.44	61.0	2H, bs	7	7, 5a
2''	***		5.34	OH	1H, t, 5.4 Hz		1'' (in DMSO)

*: assignment might be inverse. **: correlations not unambiguously assignable due to signal overlap/noise. ***: Not observable in MeOD.

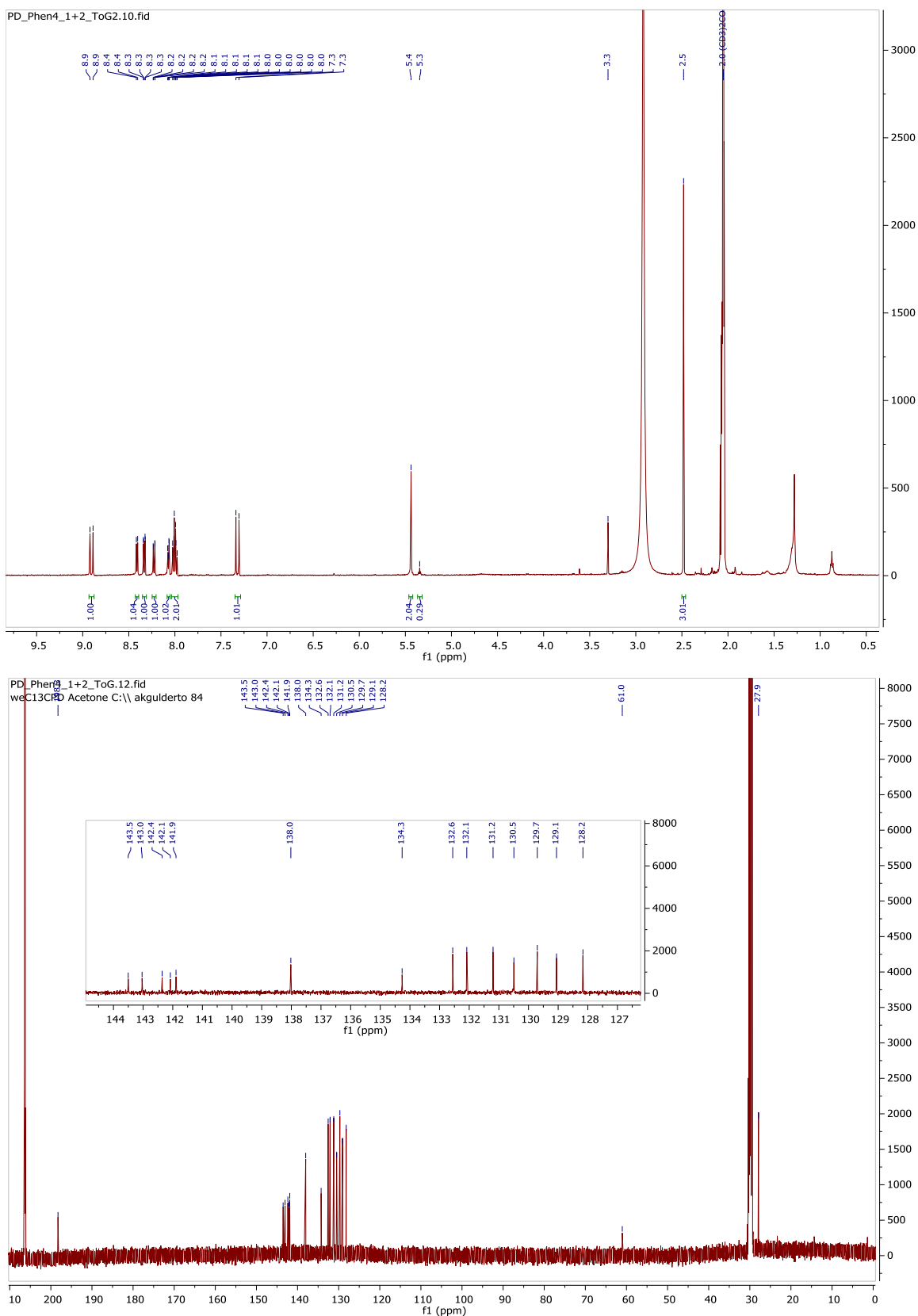


Figure S6. ^1H and ^{13}C NMR spectra of fontizine A (**2**) in acetone- d_6 .

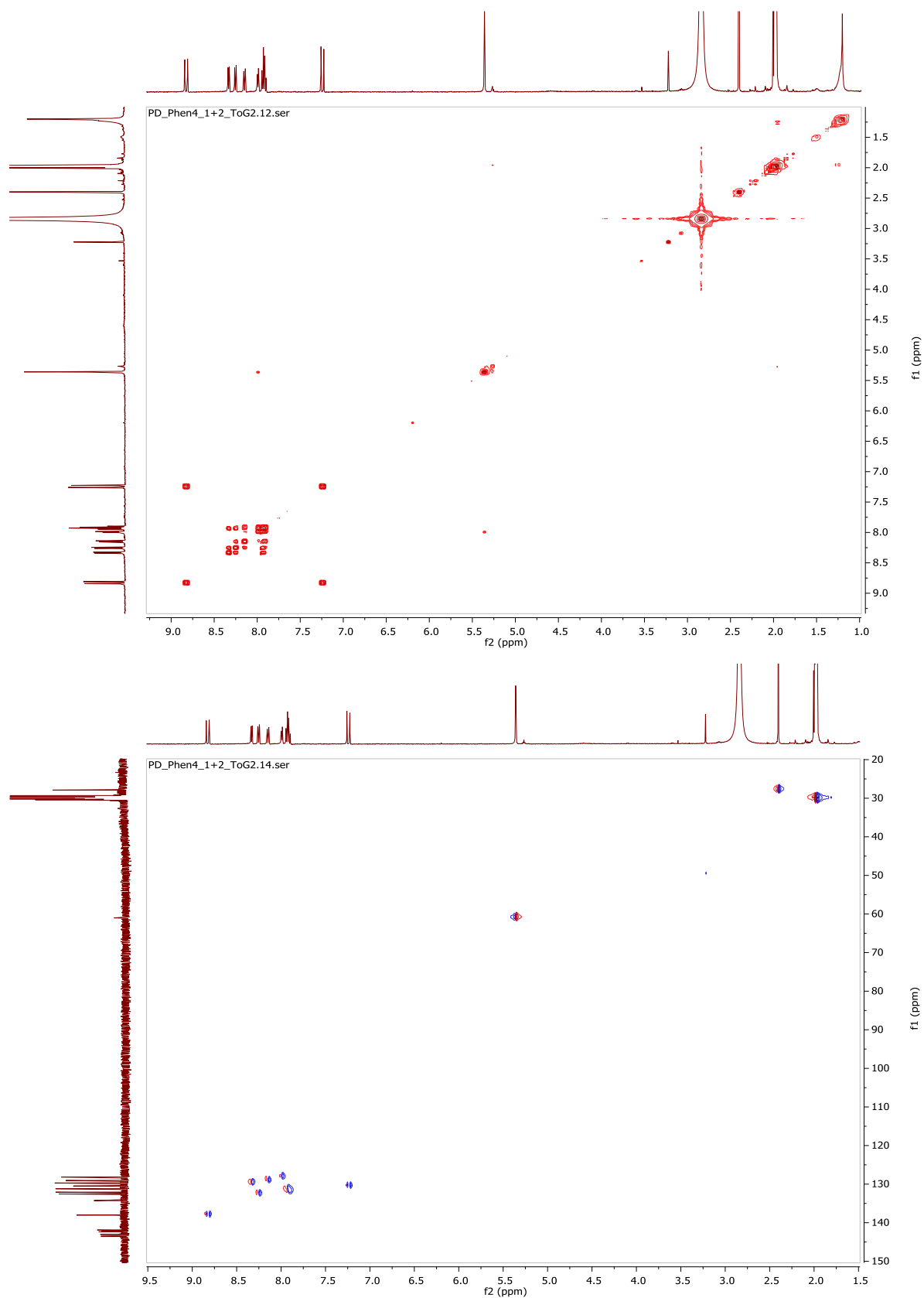


Figure S7. 2D COSY and HSQC spectra of fontizine A (**2**) in acetone- d_6 .

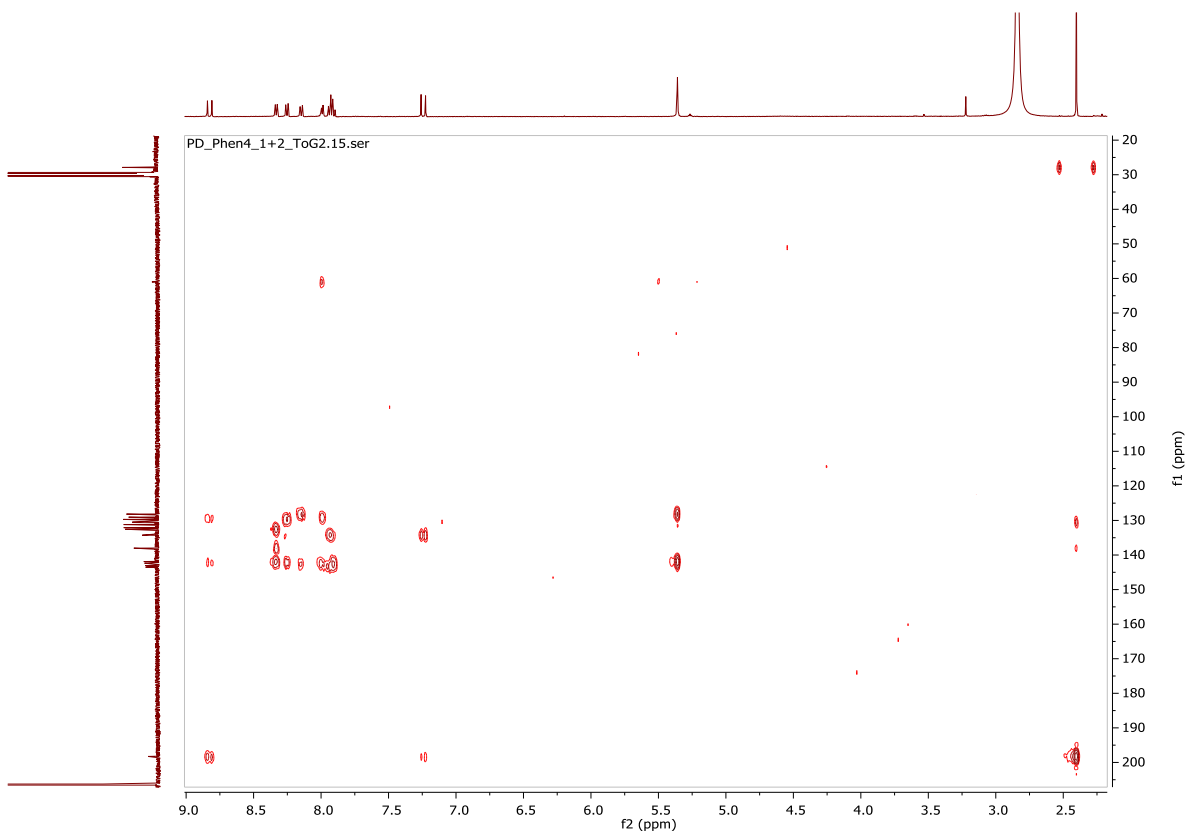


Figure S8. 2D HMBC spectrum of fontizine A (**2**) in acetone- d_6 .

2.2. Results for Anabaenopeptins

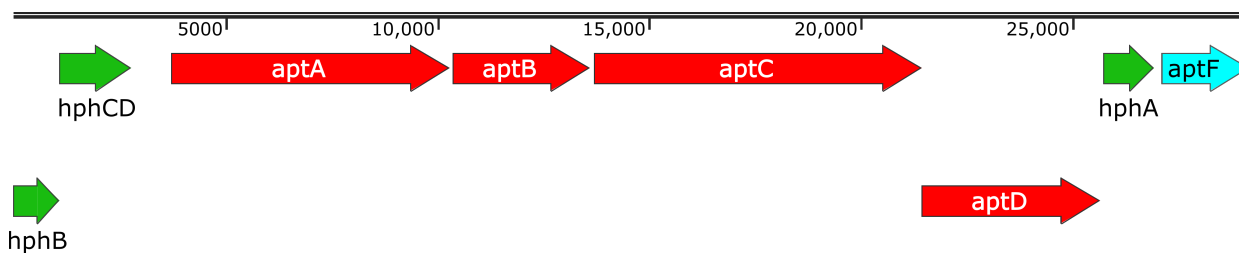


Figure S9. Organization of the *apt* gene cluster (29.2 kb). The NRPS genes *aptA-D* (red) are flanked by the homoamino acid producing genes *hphA-CD* (green) and the transporter *aptF* (cyan).

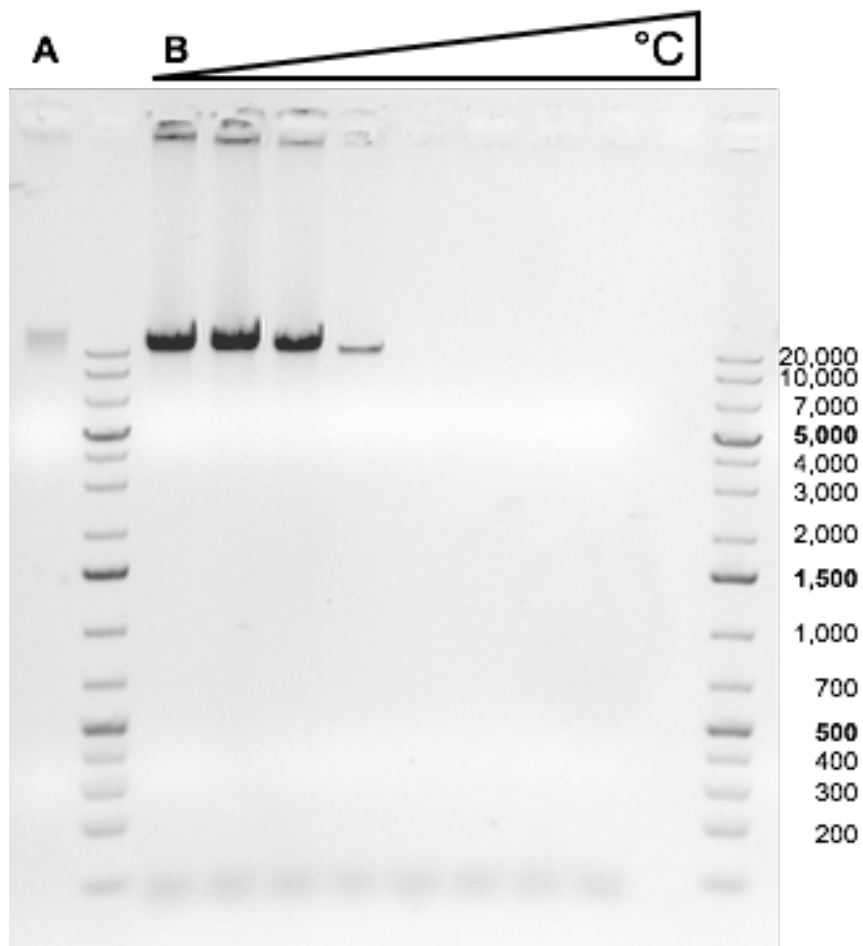
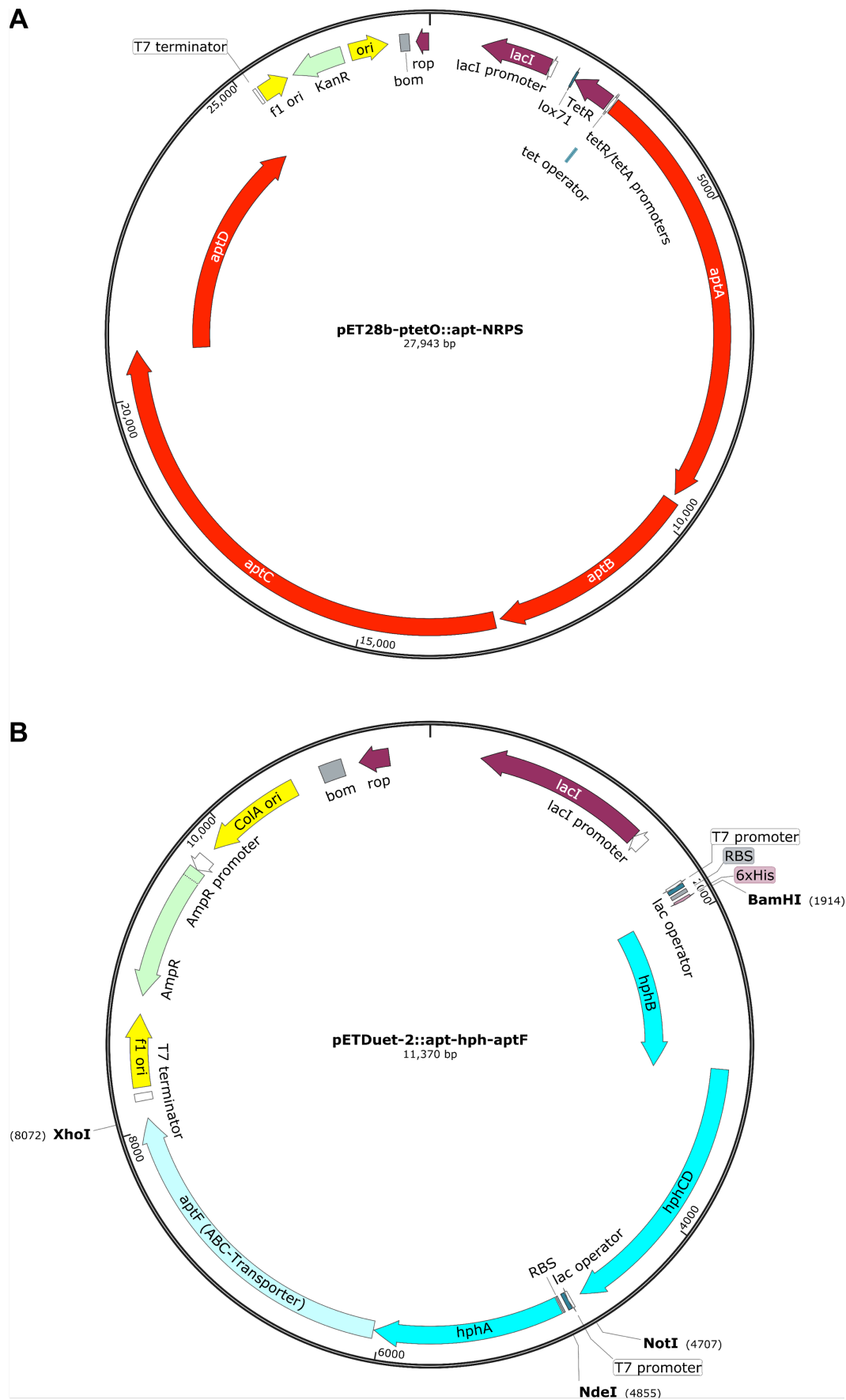


Figure S10. Agarose gel electrophoresis of the 22.0 kb *aptA-D* fragment amplified by Q5 polymerase gradient PCR. (A) *N. punctiforme* gDNA (150 ng), (B) temperature gradient PCR results (54.4-72.0 °C).



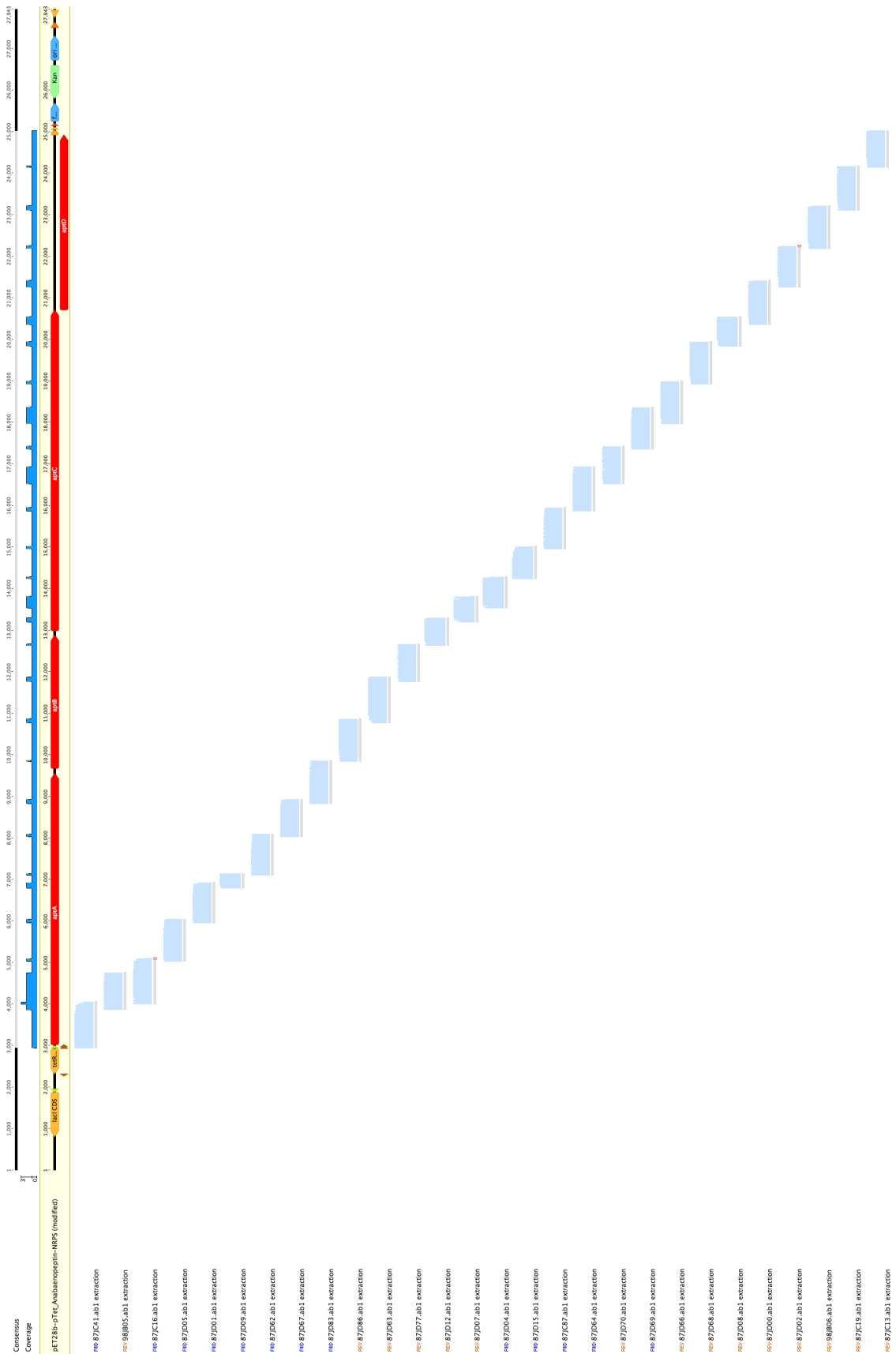


Figure S12. Schematic representation of the sequencing results for the insert *apt-NRPS* of pET-28b-ptetO::*apt-NRPS*. Lines 2 to 29 graphically show the perfect agreement of the individual overlapping sequencing runs with the original sequence of *apt-NRPS* (top line, yellow).

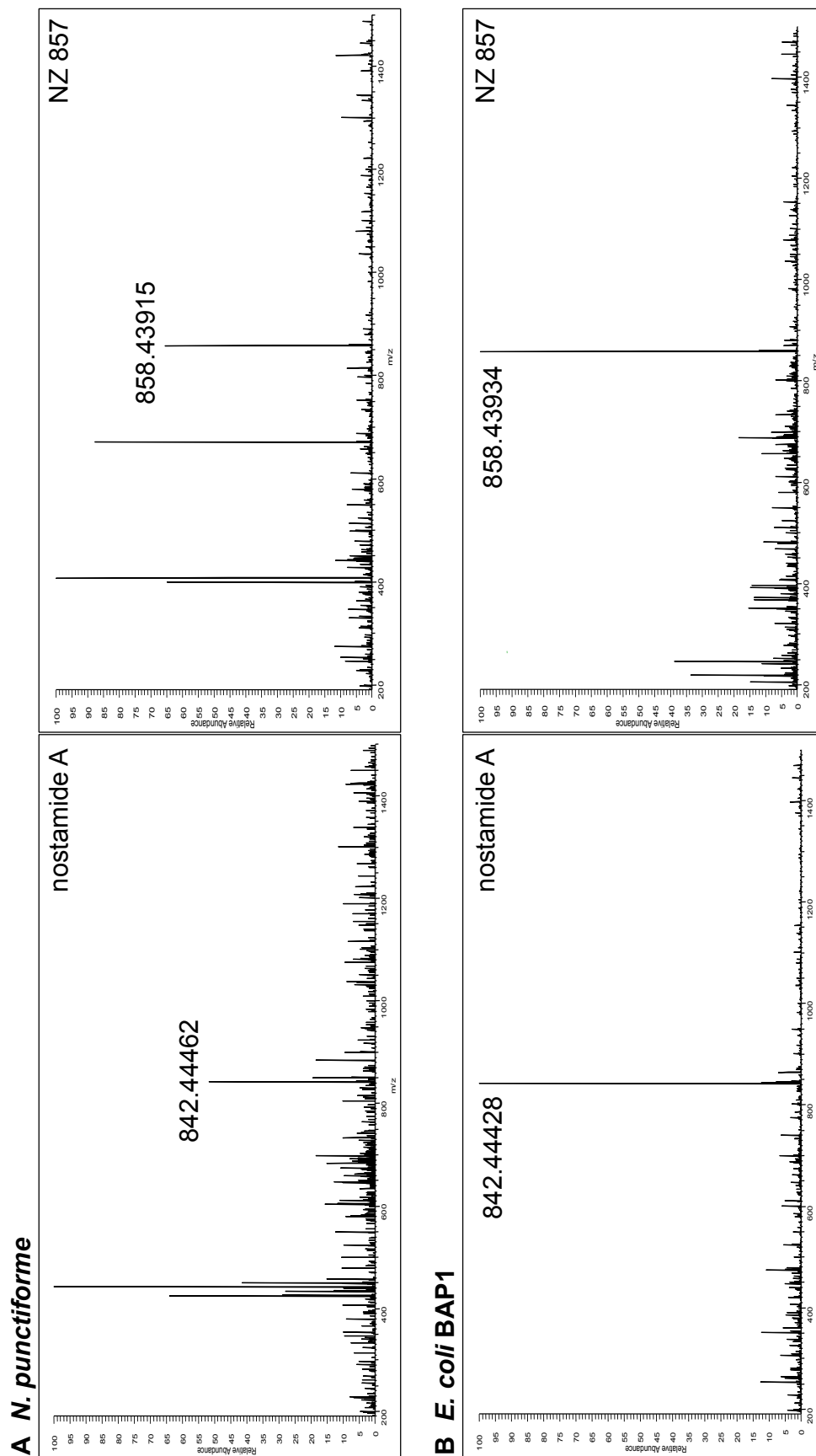


Figure S13. HR-HPLC-MS data of the (A) *N. punctiforme* pellet extraction and (B) *E. coli* BAP1 pET-28b-petO::apt-NRPS + pETDuet-2::apt-hph-aptF supernatant extraction.

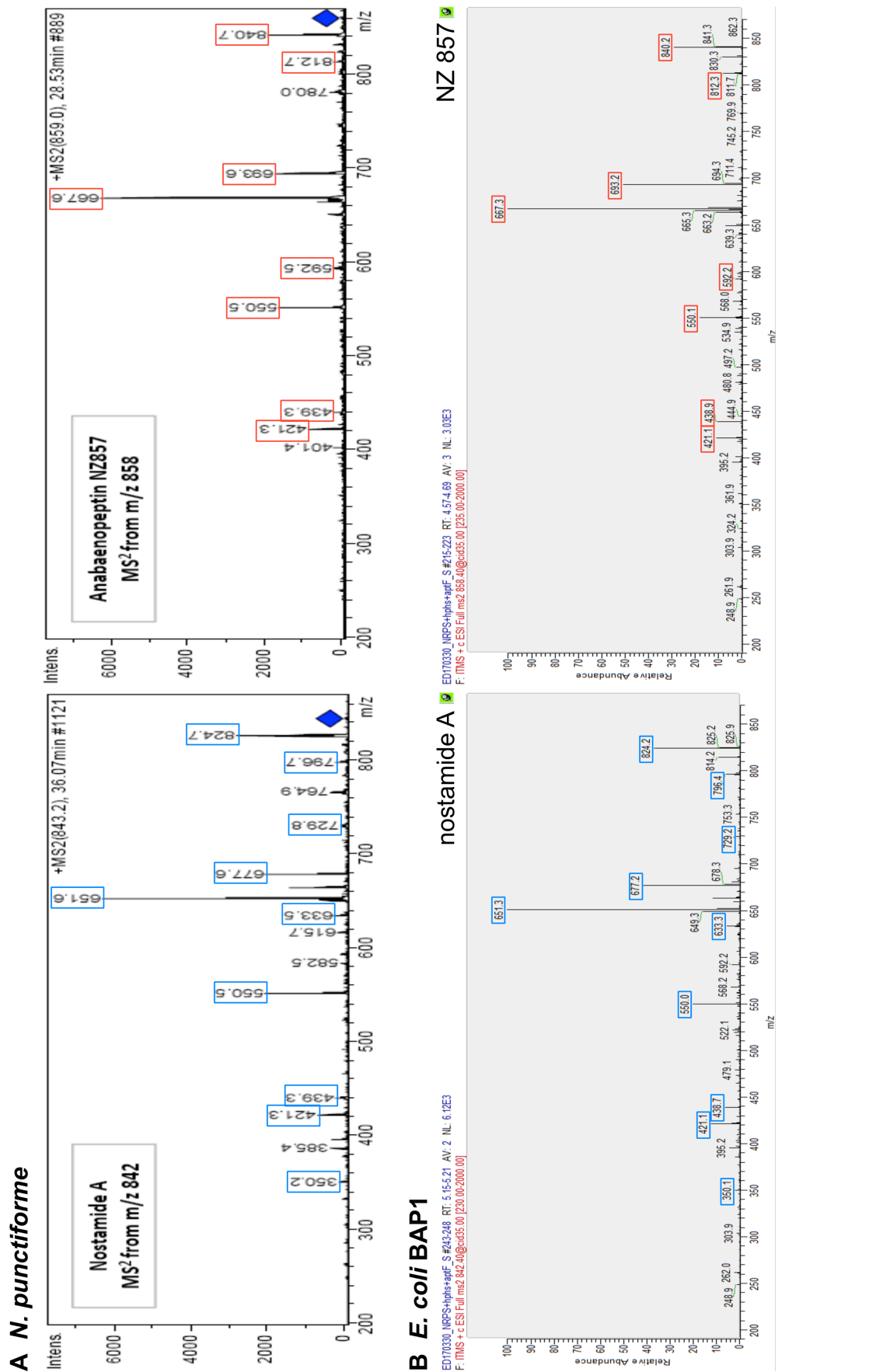


Figure S14. HPLC-MS² data of (B) the *E. coli* BAP1 pET-28b-ptetO::*apt-NRPS* + pETDuet-2::*apt-hph-aptF* supernatant extraction, in comparison with data from Rouhianinen *et al.* (A) (9).

2.3. Results for Erythromycin

Table S6. Restriction site strategy used for vector backbone linearizations to sequentially assemble *ery*.

Plasmid name	Restriction site	Description
pET-28b-SUMO (5650 bp)	<i>Bam</i> HI	Modified derivative of pET-28b used to insert <i>ery</i> -p1 (9.8 kb 70% GC) and subcloning of <i>ery</i> -p5 (12,7 kb 71% GC)
pET-28b-SUMO:: <i>ery</i> -p1 (15,522 bp)	<i>Sna</i> BI	Intermediate fragment to insert <i>ery</i> -p2 (10.6 kb 74% GC)
pET-28b-SUMO:: <i>ery</i> -p12 (26,099 bp)	<i>Sna</i> BI	Intermediate fragment to insert <i>ery</i> -p3 (12.1 kb 73% GC)
pET-28b-SUMO:: <i>ery</i> -p123 (38,170 bp)	<i>Sna</i> BI	Intermediate fragment to insert <i>ery</i> -p4 (9.5 kb 74% GC)
pET-28b-SUMO:: <i>ery</i> -p1234 (47,686 bp)	<i>Sna</i> BI	Insertion of <i>ery</i> -p5 (12.6 kb 71%), failed due to stability issues
pCC1FOS:: <i>ery</i> -p1234 (50,168 bp)	<i>Nsi</i> I	Insertion of <i>ery</i> -p5 (12.6 kb 71%) yielding complete <i>ery</i>

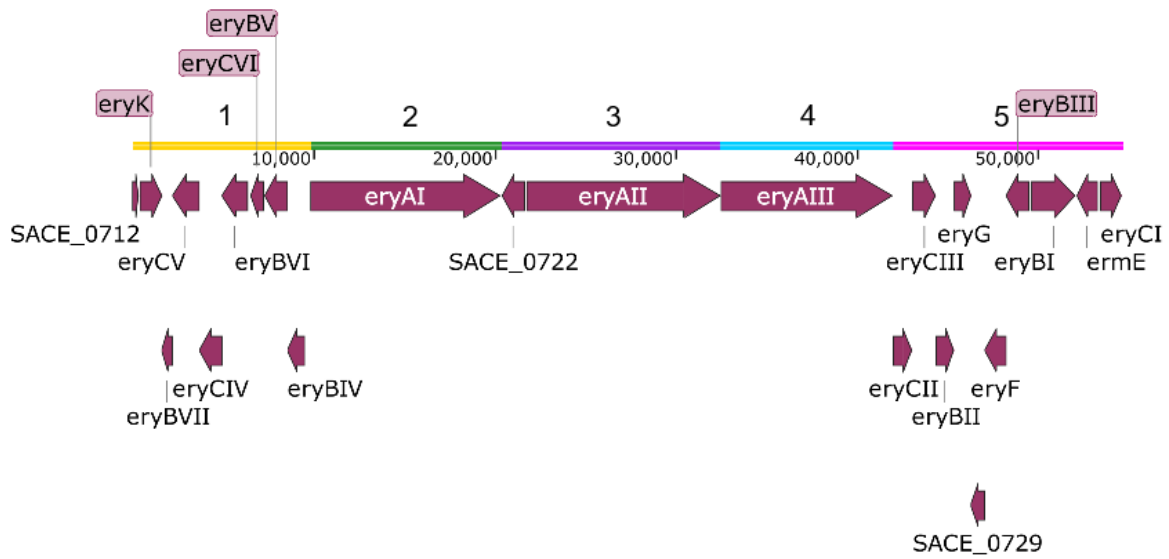
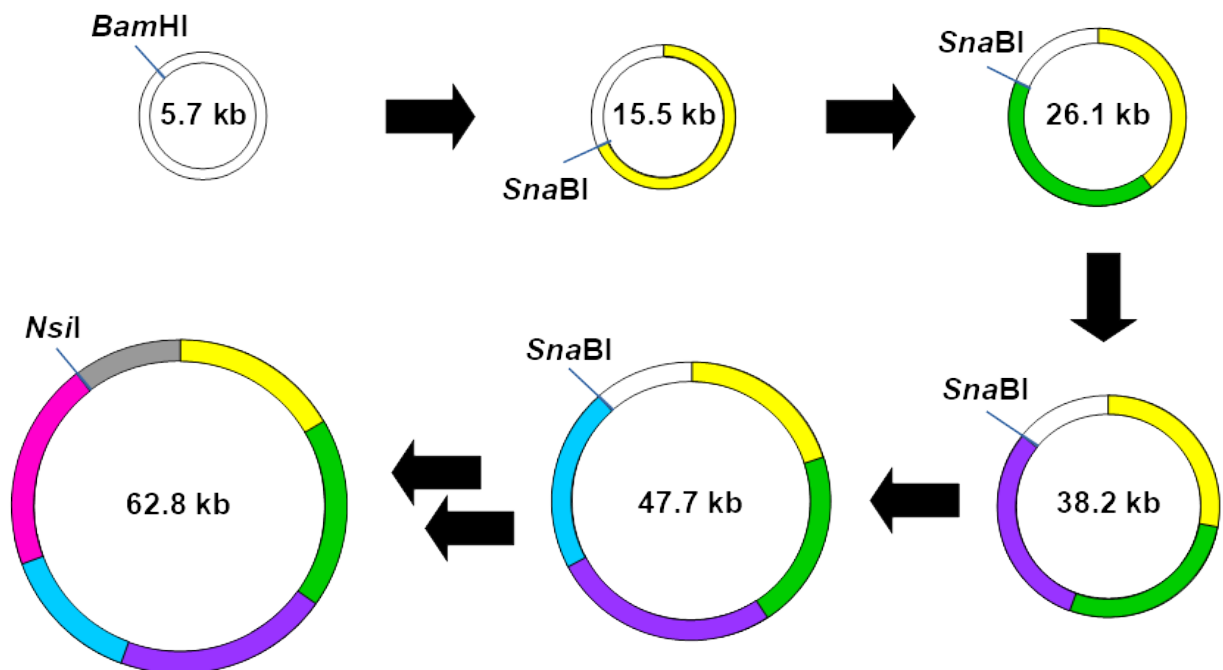
A**B**

Figure S15. Partitioning of the *ery* cluster (54.6 kb, 72.5% GC) for amplification and assembly. (A) The cluster was divided into five parts, *ery*-p1 (9.8 kb, 70% GC, yellow), *ery*-p2 (10.6 kb, 74% GC, green), *ery*-p3 (12.1 kb, 73% GC, purple), *ery*-p4 (9.5 kb, 74% GC, light blue) and *ery*-p5 (12.6 kb, 71% GC, magenta) which were amplified by PCR. (B) Sequential assembly strategy to reconstitute complete *ery*. Vector backbones were pET-28b-SUMO (white) and pCC1FOS (gray).

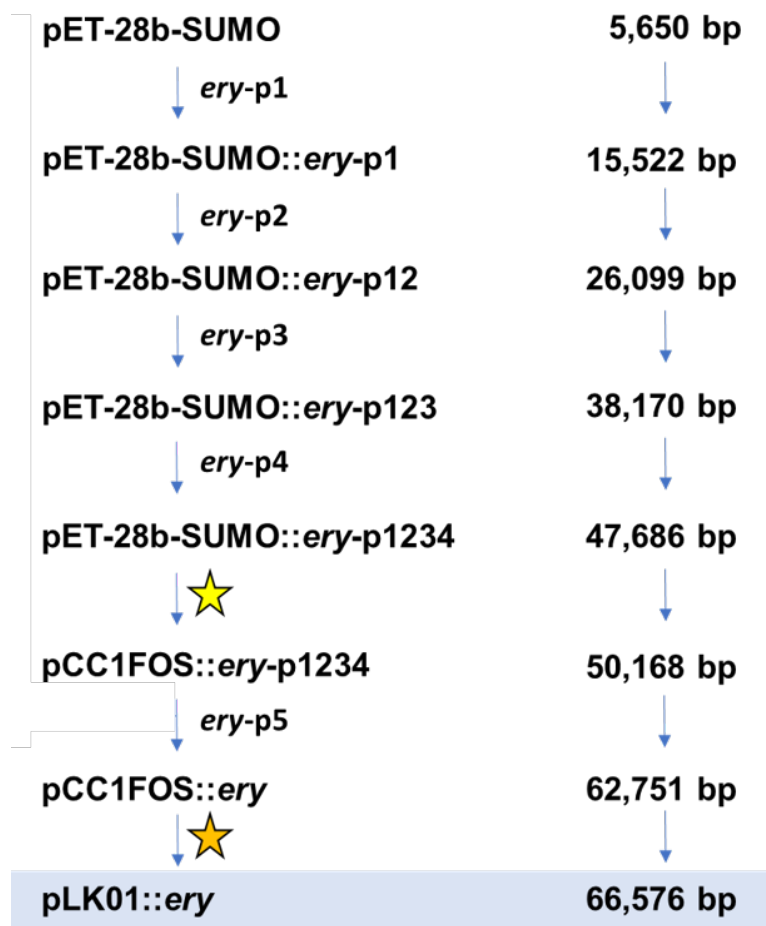


Figure S16. Sequential assembly strategy to access *ery* (54.6 kb). The BGC from *S. erythraea* DSM 40517 was split into five parts (*ery*-p1 to *ery*-p5) which were amplified by PCR. pET-28b-SUMO was the vehicle to assemble the clusters parts *ery*-p1 to -4. A backbone exchange for pCC1FOS was conducted by homologous recombination (LCHR) before integration of *ery*-p5. LCHR was applied to transfer *ery* into pLK01 for heterologous expression in *Streptomyces* (orange star).

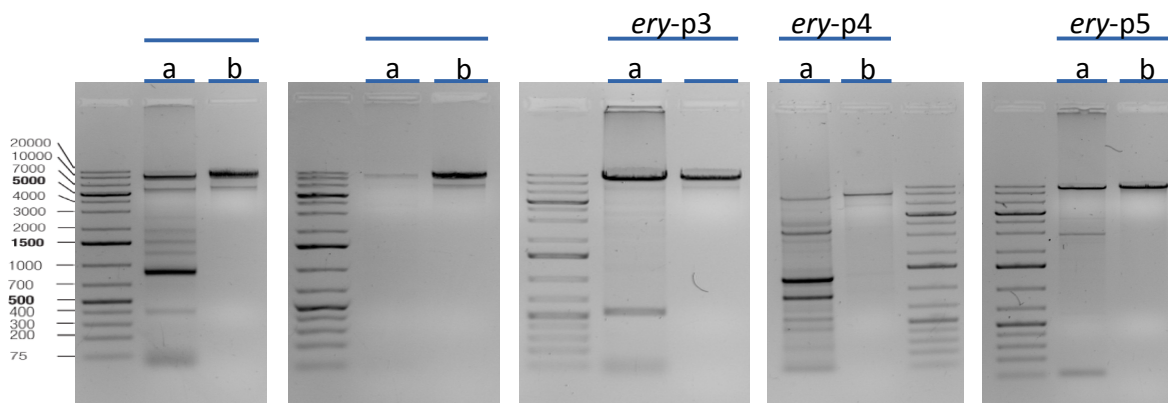


Figure S17. Agarose gel electrophoresis of all five erythromycin PCR fragments a) amplified by Q5 polymerase from gDNA and b) product purified by gel extraction (*ery*-p1 9.8 kb, *ery*-p2 10.6 kb, *ery*-p3 12.1 kb, *ery*-p4 9.5 kb and *ery*-p5 12.6 kb).

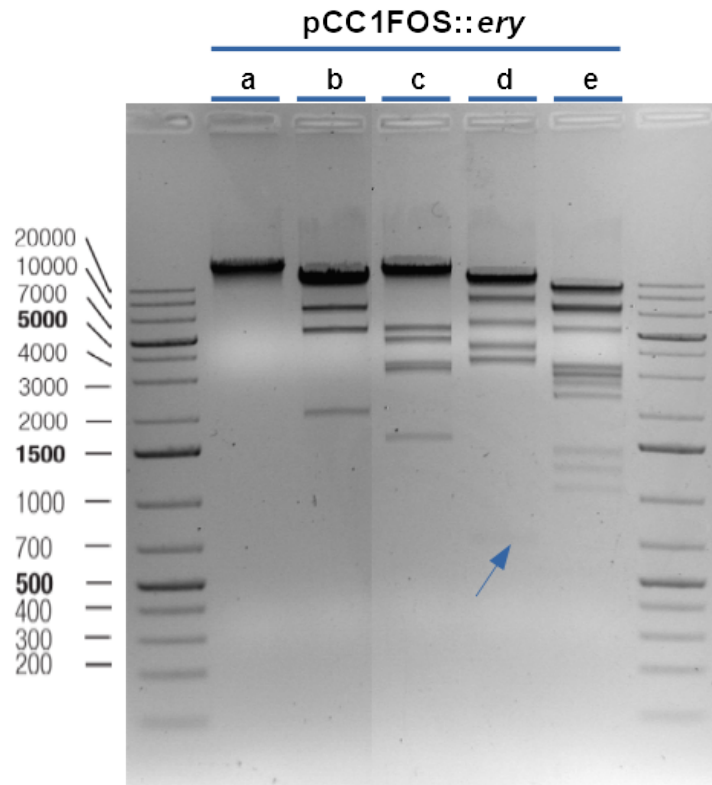


Figure S18. Agarose gel electrophoresis to analyze restriction enzyme digests of pCC1FOS::*ery*: (a) *Nsi*I-HF (62,751 bp), (b) *Bmt*I + *Xba*I (25,763; 22,514; 7317; 5318; 1839 bp), (c) *Hind*III-HF + *Sbf*I-HF (45,645; 5226; 4501; 3087; 2873; 1419 bp), (d) *Psp*XI (18,099; 16,391; 8107; 5456; 3922; 3840; 3183; 3135; 618 bp) and (e) *Not*I-HF (12,166; 11,072; 7498; 6956; 5226; 3129; 3075; 2808; 2785; 2503; 2175; 1293; 1119; 946 bp).

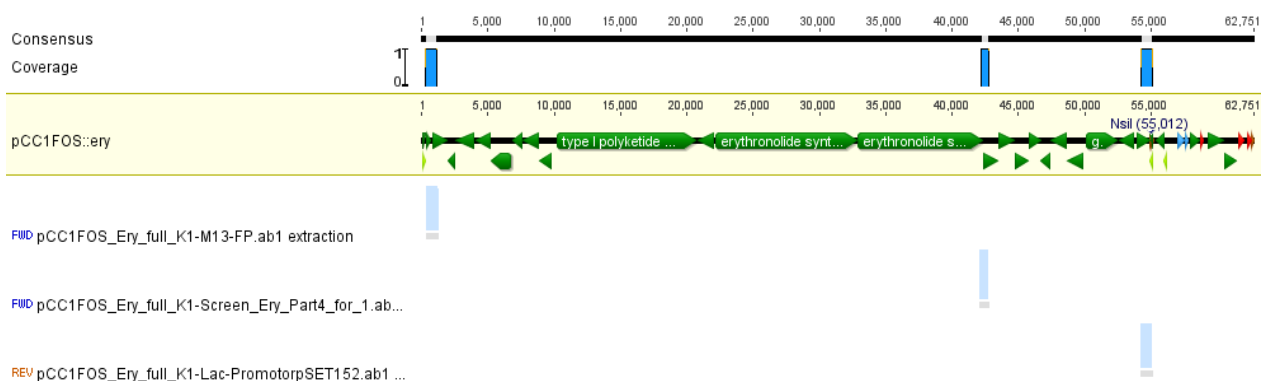


Figure S19. End-sequencing of pCC1FOS::*ery* to confirm correct integration of *ery-p5*. The reference sequence (top line, yellow) was aligned with three sequencing traces covering the overlap of pCC1FOS to the beginning of *ery* (line 2) and *ery-p5* (lines 3, 4).

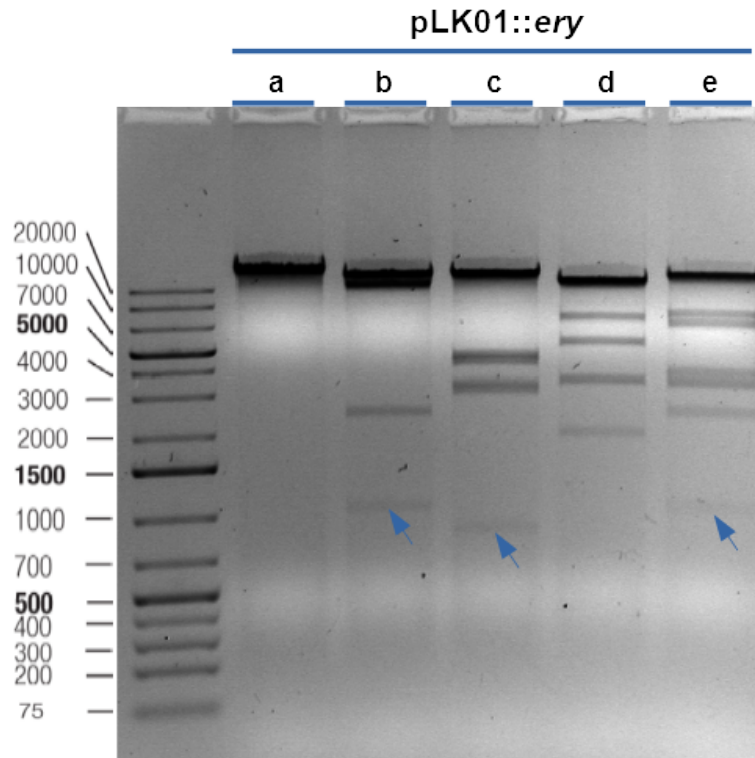


Figure S20. Agarose gel electrophoresis to analyze restriction enzyme digests of pLK01::ery: (a) *NsiI* (66,576 bp), (b) *HindIII* (45,645; 17,676; 2286; 969 bp), (c) *SbfI* (51,194; 4501; 4119; 3087; 2873; 802 bp), (d) *BmiI* + *XbaI* (26,318; 22,514; 7317; 5318; 3270; 1839 bp) and (e) *NdeI* + *HindIII* (42,101; 7672; 6776; 3544; 3228; 2286; 969 bp).

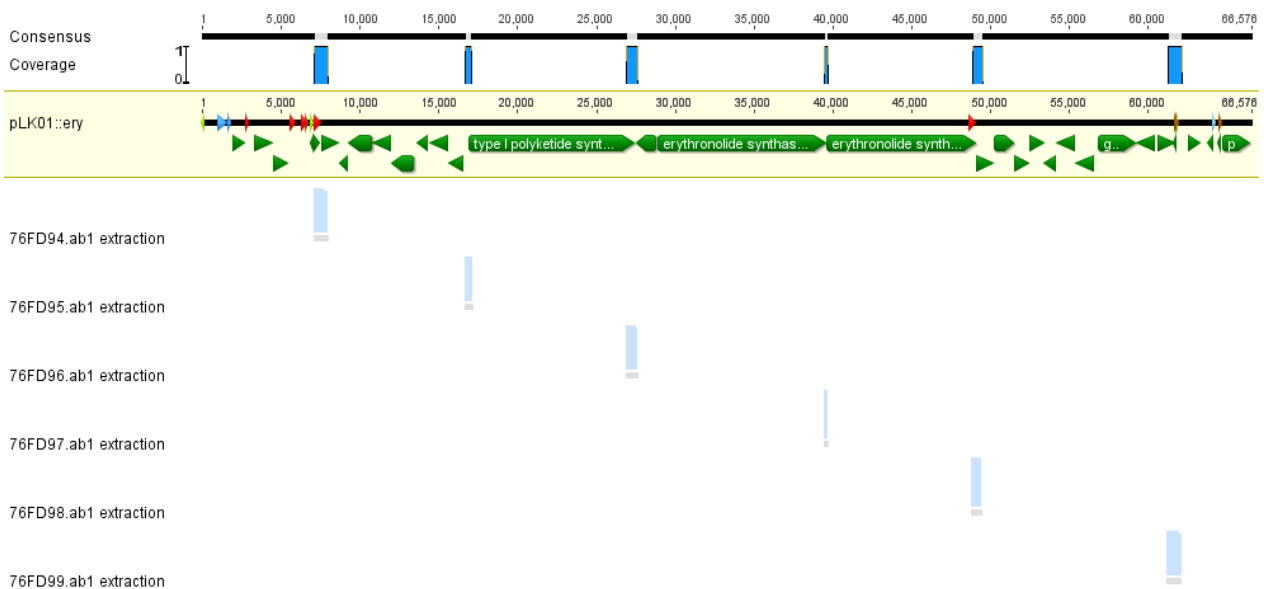


Figure S21. End-sequencing of pLK01::ery to confirm the presence of *ery*. The reference sequence (top line, yellow) was aligned with six sequencing traces covering the overlap of the pLK01 backbone to *ery*, and the overlapping regions of the assembled fragments *ery*-p1 to -p5 (lines 2 to 7).

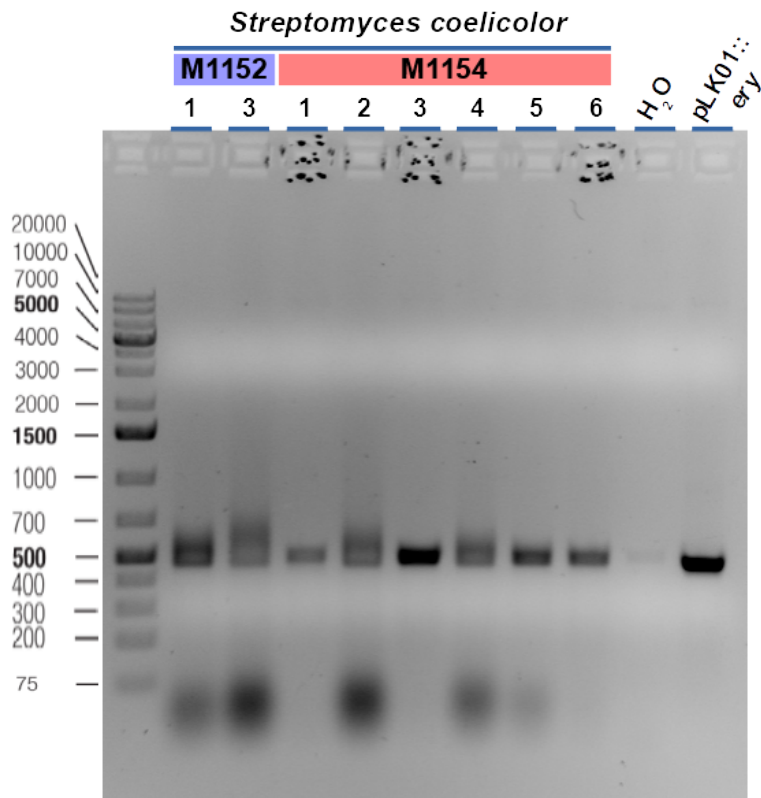


Figure S22. Agarose gel electrophoresis to analyze PCR screenings of CASO precultures of *S. coelicolor* for the presence of *ery*. An amplicon of expected size (485 bp) could be found in all cases, validating a successful conjugational transfer. As controls, reactions with water and plasmid template (pLK01::*ery*) were included.

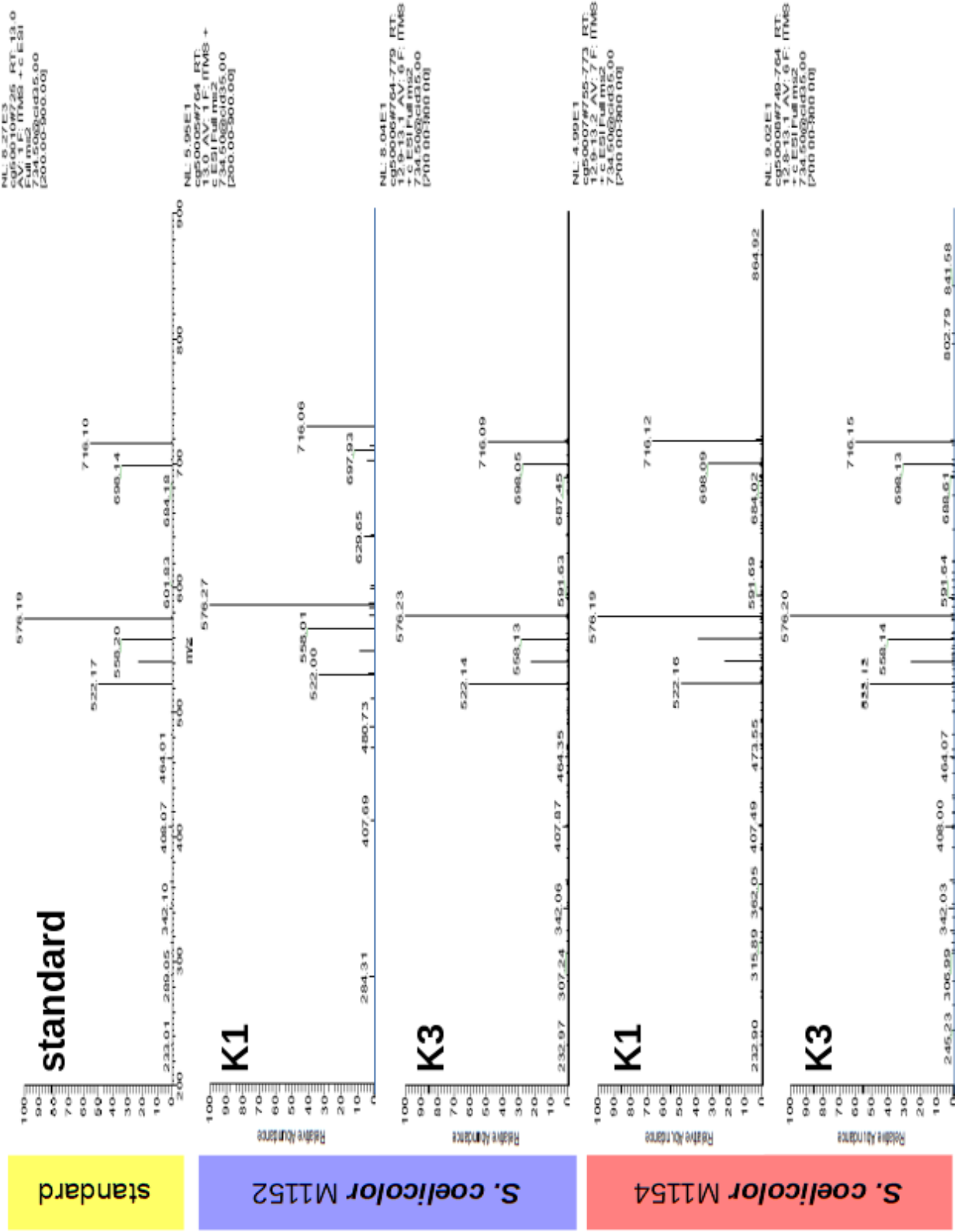


Figure S24. MS² spectra of erythromycin A produced by *S. coelicolor* compared to a commercial standard.

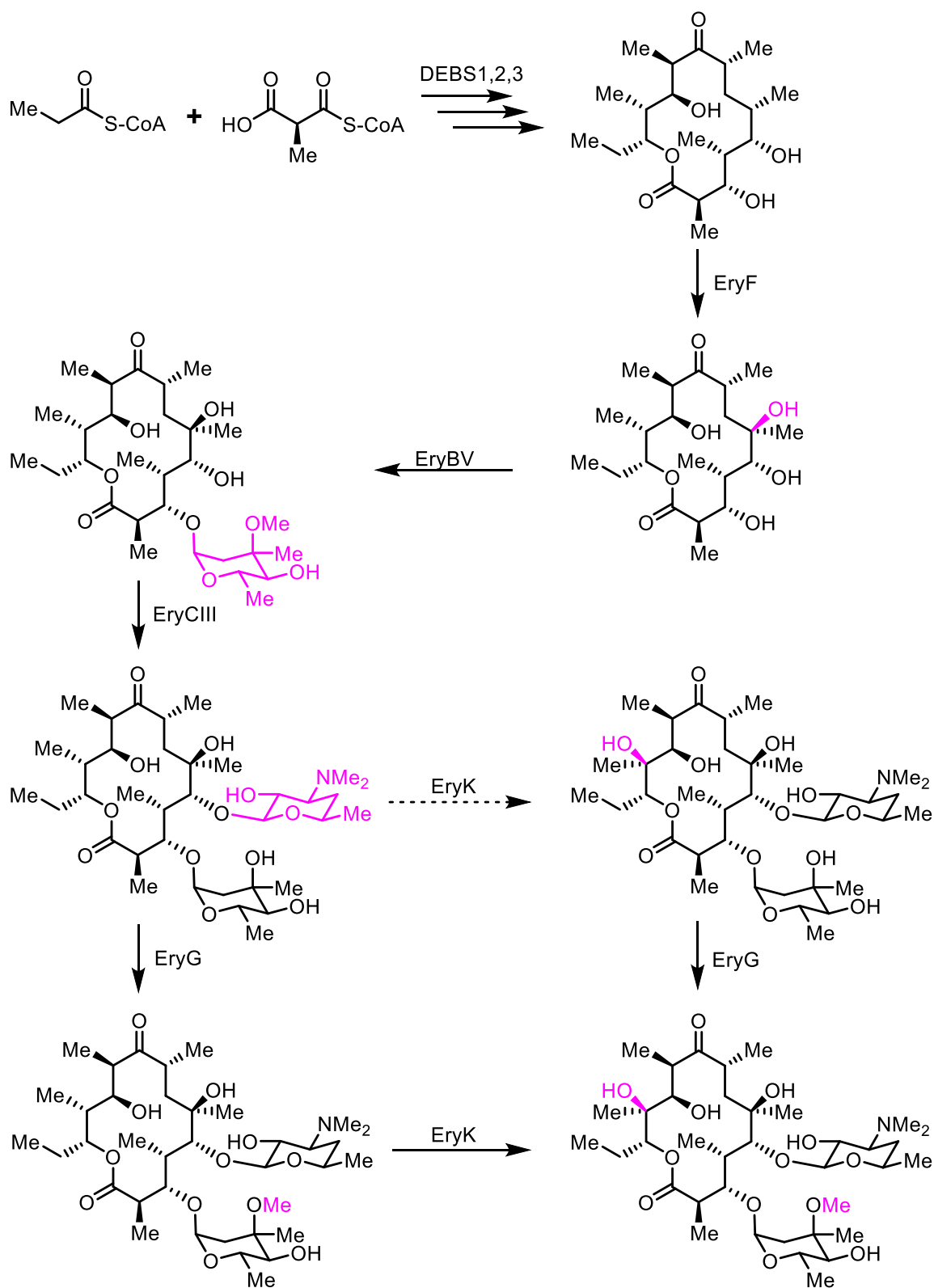


Figure S25. Schematic representation of erythromycin A biosynthesis, according to (10). Post-PKS modifying reactions are highlighted in pink.

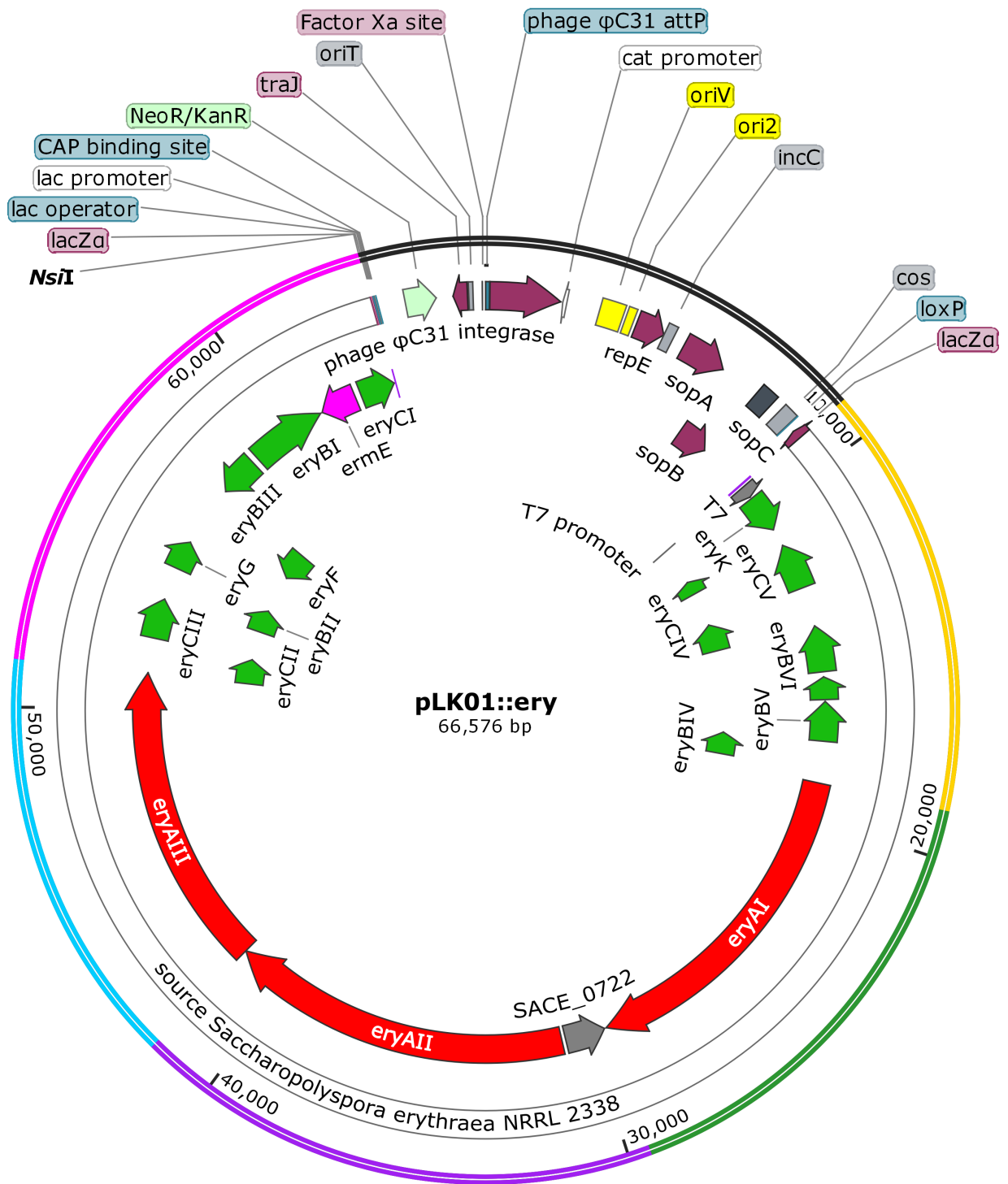


Figure S26. Plasmid map of pLK01::ery (66,576 bp) used for heterologous expression in *S. coelicolor*.

3. References

1. Fu J, *et al.* (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* 30(5):440-446.
2. Wang H, *et al.* (2016) RecET direct cloning and Redalphabeta recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat Protoc* 11(7):1175-1190.
3. MacNeil DJ, *et al.* (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* 111(1):61-68.
4. Flett F, Mersinias V, & Smith CP (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol Lett* 155(2):223-229.
5. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, & Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 291(5509):1790-1792.
6. Gomez-Escribano JP & Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb Biotechnol* 4(2):207-215.
7. Baer P, *et al.* (2014) Induced-fit mechanism in class I terpene cyclases. *Angew Chem Int Ed Engl* 53(29):7652-7656.
8. Kaysser L, *et al.* (2012) Merochlorins A-D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *J Am Chem Soc* 134(29):11988-11991.
9. Rouhiainen L, Jokela J, Fewer DP, Urmann M, & Sivonen K (2010) Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). *Chem Biol* 17(3):265-273.
10. Staunton J & Wilkinson B (1997) Biosynthesis of Erythromycin and Rapamycin. *Chem Rev* 97(7):2611-2630.

A.1.2 Direct Pathway Cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*

The supplemental information is related to the following publication which is included in chapter 3.2:

- **E. R. Duell***, P. M. D'Agostino*, N. Shapiro, T. Woyke, T. M. Fuchs, T. A. M. Gulder: Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. *Microb. Cell Fact.* **2019**, *18*, DOI 10.1186/s12934-019-1080-6

*equally contributing authors

Direct Pathway Cloning of the Sodorifen Biosynthetic Gene

Cluster and Recombinant Generation of its Product in *E. coli*

Elke R. Duell^{1†}, Paul M. D'Agostino^{1†}, Nicole Shapiro², Tanja Woyke², Thilo M. Fuchs^{3,4} and Tobias A. M. Gulder^{1,5*}

¹Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany.

²Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

³ZIEL Institute for Food & Health, Lehrstuhl für Mikrobielle Ökologie, Department biowissenschaftliche Grundlagen, Technical University of Munich, Munich, Germany

⁴Friedrich-Loeffler-Institut, Institut für Molekulare Pathogenese, Jena, Germany

⁵Technische Universität Dresden, Chair of Technical Biochemistry, Bergstraße 66, 01602 Dresden, Germany.

† Contributed equally

*Correspondence: tobias.gulder@ch.tum.de

SUPPORTING INFORMATION

Table of contents

1 Antismash and Cluster Finder results	2
2 Direct Pathway Cloning of the <i>sod</i> cluster	6
2.1 Q5 Polymerase PCR setup	6
2.2 Taq Polymerase PCR setup	6
3 Expression of the <i>sod</i> cluster	7

1 Antismash and Cluster Finder results

Cluster	Category	Start	End	Description	Link
Cluster 1	Nrps	152180	221285	Ravidomycin_biosynthetic_gene_cluster (5% of genes show similarity)	BGC0000263_c1
Cluster 2	Cf_putative	277500	283014	-	-
Cluster 3	Cf_putative	343838	351083	-	-
Cluster 4	Cf_putative	452077	458216	-	-
Cluster 5	Cf_fatty_acid	530237	551451	-	-
Cluster 6	Nrps	585530	638119	Turnerbactin_biosynthetic_gene_cluster (30% of genes show similarity)	BGC0000451_c1
Cluster 7	Cf_fatty_acid	1101356	1122603	-	-
Cluster 8	Cf_putative	1225164	1234949	PM100117/_PM100118_biosynthetic_gene_cluster (21% of genes show similarity)	BGC0001359_c1
Cluster 9	Otherks-T1pks-Pufa-Nrps	1309016	1391220	Zeamine_biosynthetic_gene_cluster (95% of genes show similarity)	BGC0001056_c1
Cluster 10	Cf_putative	1517998	1528737	-	-
Cluster 11	Nrps	1638611	1682552	Lipopolysaccharide_biosynthetic_gene_cluster (16% of genes show similarity)	BGC0000772_c1
Cluster 12	Cf_saccharide	2079656	2111872	Lipopolysaccharide_biosynthetic_gene_cluster (27% of genes show similarity)	BGC0000776_c1
Cluster 13	Cf_putative	2307169	2314735	-	-
Cluster 14	Cf_putative	2360781	2372066	O&K-antigen_biosynthetic_gene_cluster (3% of genes show similarity)	BGC0000780_c1
Cluster 15	Cf_putative	2400546	2408494	-	-
Cluster 16	Cf_putative	2768674	2773791	-	-
Cluster 17	Arylpolyene-Siderophore	3060557	3126399	APE_Ec_biosynthetic_gene_cluster (73% of genes show similarity)	BGC0000836_c1
Cluster 18	Nrps	3174907	3226855	Vulnibactin_biosynthetic_gene_cluster (12% of genes show similarity)	BGC0000460_c1
Cluster 19	Cf_putative	3389902	3400844	-	-
Cluster 20	T1pks-Nrps	3402732	3450570	Rishirilide_B_biosynthetic_gene_cluster (7% of genes show similarity)	BGC0001179_c1
Cluster 21	Cf_putative	3501638	3513935	-	-
Cluster 22	Cf_putative	3592222	3597199	-	-
Cluster 23	Cf_saccharide	3890501	3917908	Stewartan_biosynthetic_gene_cluster (57% of genes show similarity)	BGC0000763_c1

Cluster 24	Cf_saccharide	3920935	3942988	O-antigen_biosynthetic_gene_cluster (30% of genes show similarity)	BGC0000787_c1
Cluster 25	Cf_putative	3988487	3998306	-	-
Cluster 26	Thiopeptide	4020188	4046677	O-antigen_biosynthetic_gene_cluster (14% of genes show similarity)	BGC0000781_c1
Cluster 27	Cf_fatty_acid	4085148	4106422	Taxllaid_biosynthetic_gene_cluster (33% of genes show similarity)	BGC0001133_c1
Cluster 28	Cf_putative	4113644	4134661	Prodigiosin_biosynthetic_gene_cluste r (82% of genes show similarity)	BGC0000258_c1
Cluster 29	Cf_putative	4144920	4160325	-	-
Cluster 30	Cf_fatty_acid	4219983	4240936	-	-
Cluster 31	Cf_putative	4388170	4393085	-	-
Cluster 32	Cf_putative	4415768	4422464	-	-
Cluster 33	Terpene	4586723	4607877	Sodorifen_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0001361_c1
Cluster 34	Cf_putative	4660496	4673055	-	-
Cluster 35	Cf_putative	4742629	4766296	-	-
Cluster 36	Cf_putative	4840226	4854533	-	-
Cluster 37	Cf_putative	5133490	5142678	Polysaccharide_B_biosynthetic_gene_ cluster (6% of genes show similarity)	BGC0001411_c1
Cluster 38	Cf_fatty_acid- Cf_saccharide	5297309	5338267	Tallysomicin_biosynthetic_gene_clus ter (5% of genes show similarity)	BGC0001048_c1

Table S2: Organisms and accession number of organisms harboring <i>sod</i>-like BGCs.		
Organism	Genome Accession	Gene Cluster Group
<i>Serratia plymuthica</i> 4Rx13	CP006250	1
<i>Serratia plymuthica</i> A30	AMSV01000037	1
<i>Serratia plymuthica</i> AS9	CP002773	1
<i>Serratia plymuthica</i> NBRC 102599	NZ_BCTU01000010	1
<i>Serratia plymuthica</i> PRI-2C	NZ_CP015613	1
<i>Serratia plymuthica</i> RVH1	NZ_ARWD01000001	1
<i>Serratia plymuthica</i> S13	CP006566	1
<i>Serratia plymuthica</i> 3Re4-18	NZ_CP012097	1
<i>Serratia plymuthica</i> 3Rp8	NZ_CP012096	1
<i>Serratia plymuthica</i> 4Rx5	NZ_PESE01000001	1
<i>Serratia plymuthica</i> HRO-C48	NZ_LTDN01000056	1
<i>Serratia plymuthica</i> strain V4	CP007439	1
<i>Serratia plymuthica</i> WS3236	2773857786 (JGI)	1
<i>Serratia</i> sp. AS12	CP002774	1
<i>Serratia</i> sp. AS13	CP002775	1
<i>Serratia</i> sp. FS14	CP005927	2
<i>Pseudomonas chlororaphis</i> O6	NZ_CM001490	3
<i>Pseudomonas chlororaphis</i> ATCC 13985	NZ_LT629738	3
<i>Pseudomonas chlororaphis</i> PA23	NZ_CP008696	3
<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> NBRC 3521	NZ_BBQB01000002	3
<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> CD	NZ_LHVB01000012	3
<i>Pseudomonas chlororaphis</i> subsp. <i>aureofaciens</i> LMG 1245	LHVA01000023	3
<i>Pseudomonas chlororaphis</i> subsp. <i>chlororaphis</i> GP72	NZ_AHAY01000123	3
<i>Pseudomonas schloroaphis</i> ATCC 9446	NBAT01000014	4
<i>Pseudomonas grimontii</i> BS2976	FNKM01000002	5
<i>Burkholderia pyrrocinia</i> Lyc2	JPWP01000010	6
<i>Burkholderia singularis</i> TSV85	LOWA01000011	7
<i>Streptomyces tsukubensis</i> NRRL18488	AJSZ01000838	8

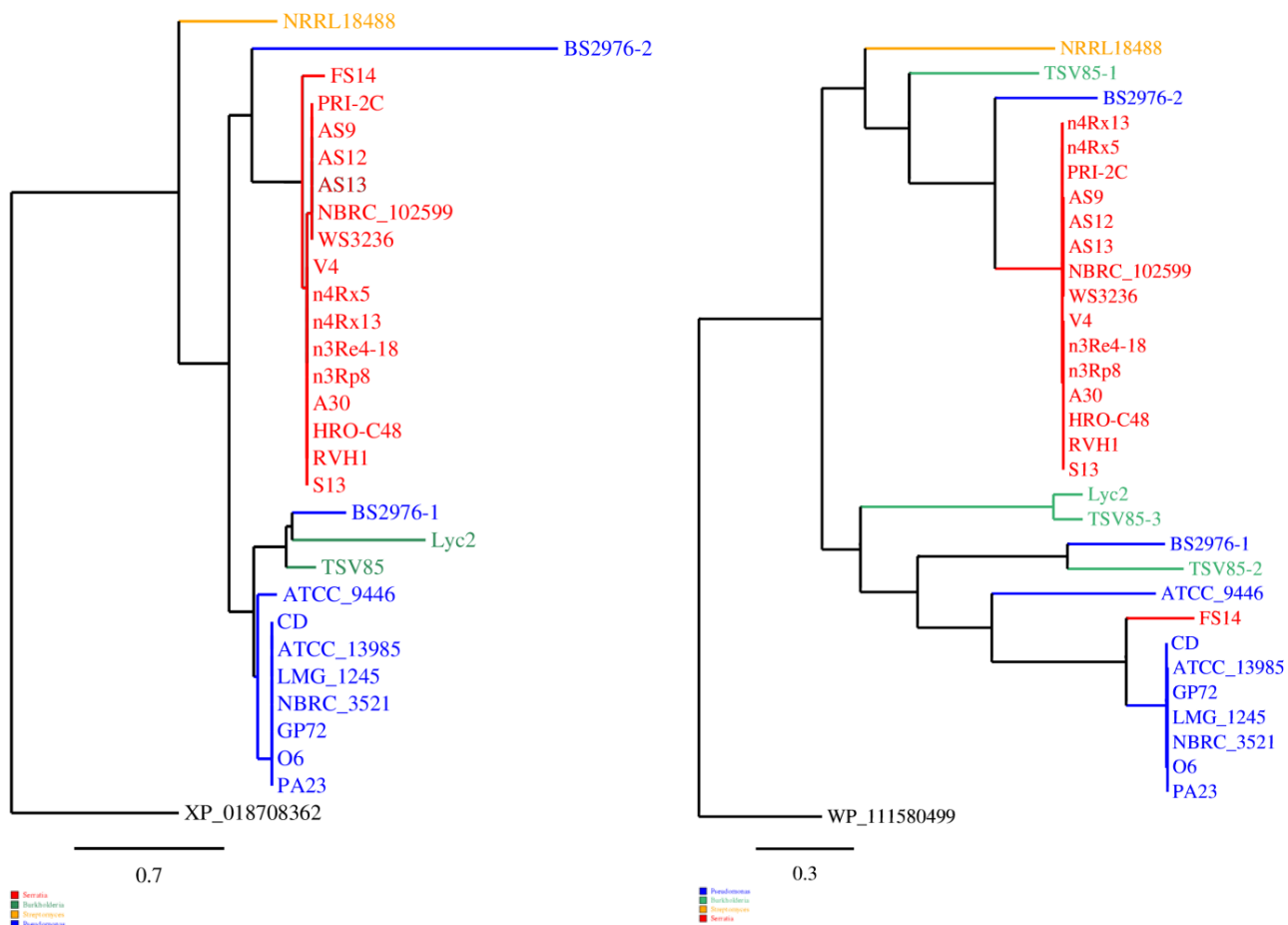


Figure S1: Phylogenetic analysis of SodC (A) and SodD (B). Organism titles can be linked to species list in Table S2. Outgroups are in black and included XP_018708362 (SodC tree) and WP_111580499 (SodD tree). Branches and strain names are colored according to genus: *Serratia*, red; *Pseudomonas*, blue; *Burkholderia*, green; *Streptomyces*, orange.

2 Direct Pathway Cloning of the *sod* cluster

Table S3: List of primers.		
Name	Sequence (5' → 3')	Description
C-GFP_for_1	CATGGTTAGCAAAGGTGAAG	amplification of the pET28b-ptetO_ <i>gfpv2</i> backbone
spec-ptet-R	GGTCGATCCTCTTCTCTATC	
SP-sod::ptet_for_1	gtgatagagaagaggatcgaccATGCTGAT CCTCGTTGATGACAAAG	amplification of <i>sod</i> cluster with 22 bp homology sequences
SP-sod::C-GFP_rev_1	TTCTTCACCTTTGCTAACCATGtt agcgccgcccgtactg	
sodD_screen_for_1	GAAGTATTAAATGGACCGC	colony screening of pET28b-ptetO:: <i>sod_gfpv2</i>
pRSET-RPnew	GGGTTATGCTAGTTATGTC	
screen_ptet_F	TGCGCTGTTAATCACTTTAC	sequencing pET28b-ptetO:: <i>sod_gfpv2</i>
screen_GFP_R	TTACCGTTGGTCGCATCACC	

2.1 Q5 Polymerase PCR setup

A standard 25 µL PCR reaction batch for long-amplicon cycling reactions consisted of: 1x Q5 reaction buffer, 200 µM deoxynucleotide triphosphates, 500 nM of forward and reverse primer, 50 ng gDNA template and 0.01 U/µL Q5 High-Fidelity DNA polymerase (NEB) were mixed. Cycling was conducted using a T100 Thermal Cycler (Biorad) as follows: 1.) Initial denaturation, 98 °C for 1 min; 2.) Denaturation, 98 °C for 10 sec; 3.) Annealing, 72 °C for 20 sec; 4.) Extension, 72 °C for 2 min 30 sec; steps 2.) to 4.) were repeated in total for 30 cycles; 5.) Final extension, 72 °C for 5 min, and 6.) End phase, 16 °C.

2.2 Taq Polymerase PCR setup

Colony screening PCRs were performed using *Taq* DNA polymerase (NEB). Clones were picked and resuspended in 50 µL of pure water and examined in a 25 µL PCR batch composed as follows: *Taq* buffer (10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, pH 8.3 at 25 °C), 4% DMSO, 100 µM deoxynucleotide triphosphates, 200 nM of forward and reverse primer, 5 µL DNA template (bacterial suspension in water) and *Taq* DNA polymerase (0.025 U/µL, NEB) were mixed. Optimal cycling parameters were set as follows: 1.) Initial denaturation, 95 °C for 5 min; 2.) Denaturation, 95 °C for 45 sec; 3.) Annealing, 48 °C for 30 sec; 4.) Extension, 72 °C for 65 sec; steps 2.) to 4.) were repeated in total 34 times; 5.) Final extension, 72 °C for 5 min, and 6.) End phase, 16 °C.

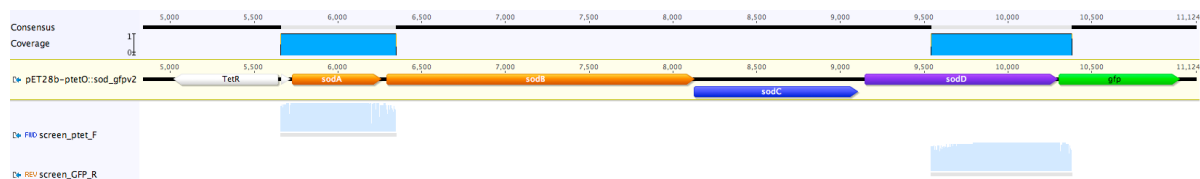


Figure S2: Sequencing of *sodA-D* integration into pET28b-ptetO::*gfpv2*.

3 Expression of the *sod* cluster

strain	medium	day	peak area pA*s	
			mesitylene ¹	sodorifene
<i>S. plymuthica</i> WS3236	succDMM	1	257.9	12.6
		2	264.0	9.1
		3	267.4	2.7
		4	266.7	1.6
	TB	1	272.9	38.7
		2	272.4	24.4
		3	261.7	4.0
		4	263.4	1.7
<i>E. coli</i> BL21 pET28b-ptetO:: <i>sod_gfpv2</i>	succDMM	1	267.0	10.8
		2	261.6	36.5
		3	267.8	26.1
		4	262.8	15.9
	TB	1	254.4	317.7
		2	262.6	765.2
		3	265.9	887.9
		4	267.9	1003.0
	LB	1	263.9	180.0
		2	268.9	252.2
		3	268.7	215.3
		4	264.6	105.5

¹Mesitylene was added as internal standard with a concentration of 250 µg/mL.

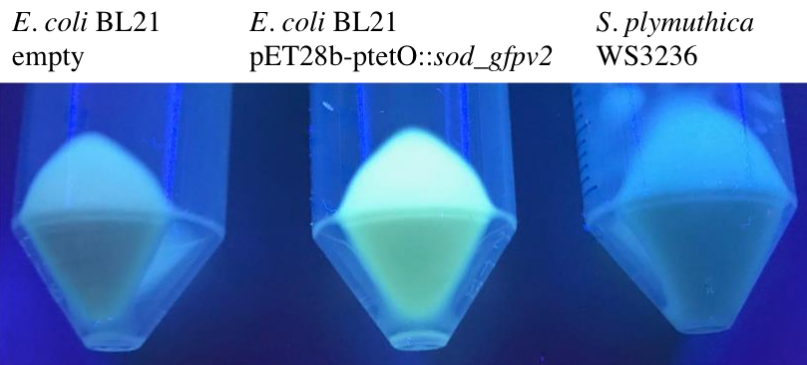


Figure S3: Cell pellets harvested after 24 h cultivation in TB medium. Only the induced heterologous expression strain *E. coli* BL21 pET28b-ptetO::*sod_gfpv2* shows fluorescence caused by the expression of the GFP marker placed downstream of the *sod* genes.

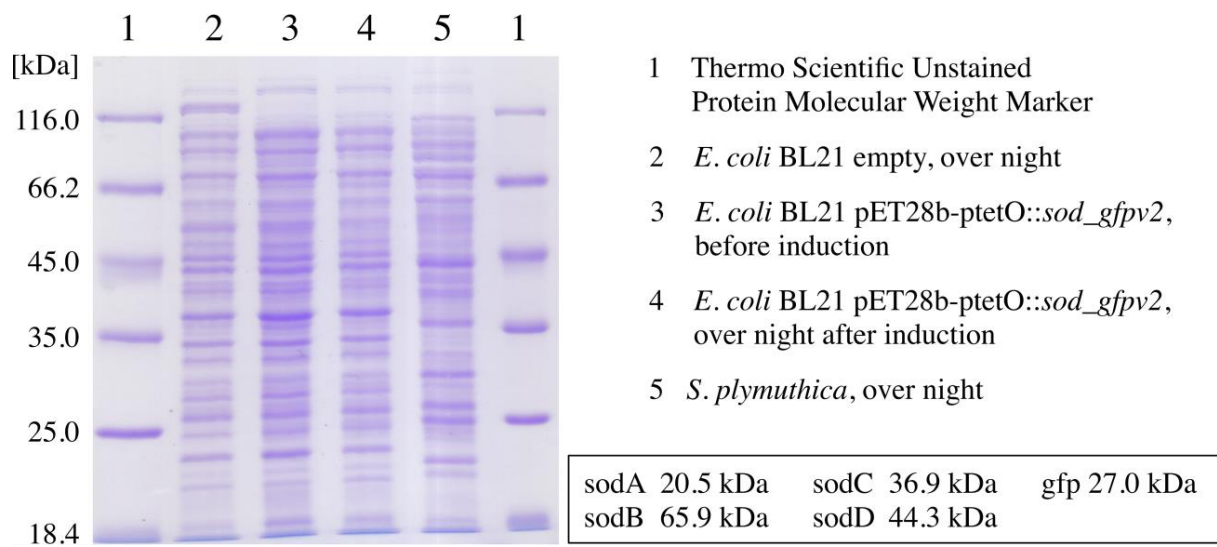


Figure S4: SDS-PAGE analysis of *E. coli* BL21 pET28b-ptetO::*sod_gfpv2* cells before (lane 3) and after induction (lane 4) in comparison to empty *E. coli* BL21 (lane 2) and *S. plymuthica* (lane 5) cells. Due to the usage of the *Ptet_o* promoter, no significant protein overexpression can be detected.

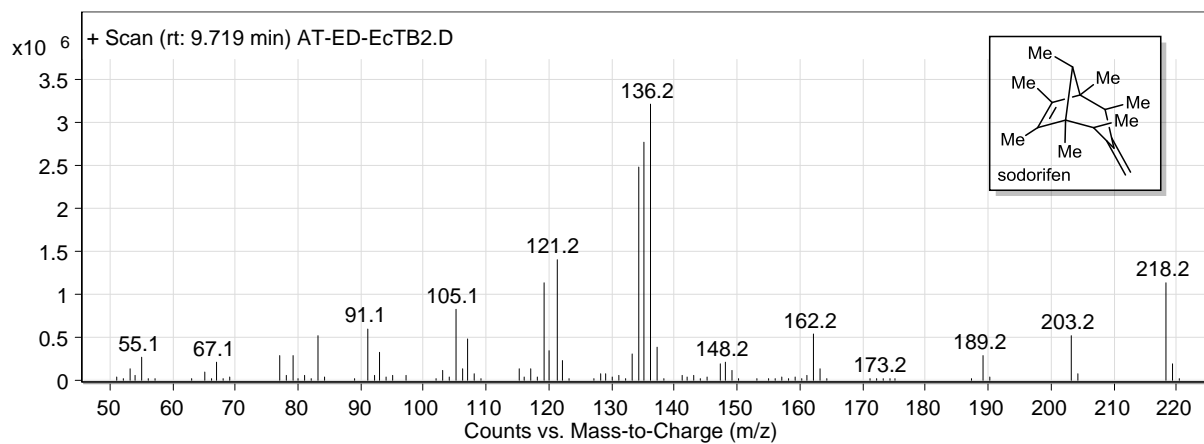


Figure S5: GC-MS spectrum of sodorifen ($m/z = 218.20$).

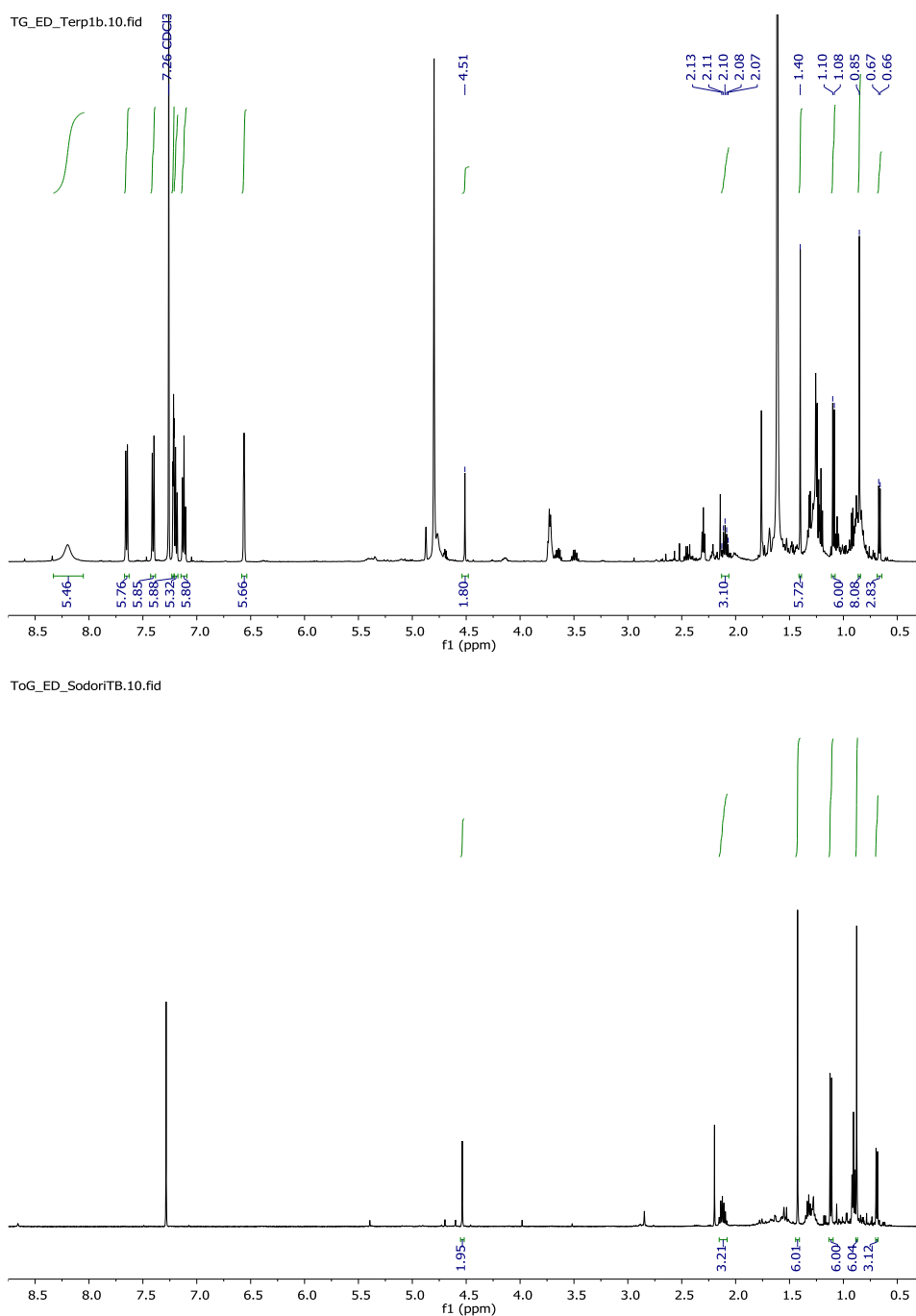


Figure S6: NMR spectra of raw head-space samples of *E. coli* pET28b-ptetO::*sod_gfpv2* grown in LB (top) and TB (bottom) medium. In the LB-derived top spectrum, the ¹H signals of the *E. coli* VOC indole (7 H signals from 8.25 to 6.50 ppm) indicate an about 10-fold excess of indole in the sample when compared to sodorifen (all other integrated ¹H signals; relative amounts calculated based on, e.g., the two protons of the exo-methylene unit at 4.5 ppm). The TB-derived bottom spectrum shows almost exclusive formation of sodorifen in high purity.

1. Project Information

Program	Microbial/CSP 2017
Sequencing Project ID	1176531
Sequencing Project Name	Serratia plymuthica WS3236

2. Read Statistics

Metric	Raw Reads	Filtered Subreads	Error Corrected Reads
Reads	421,168	78,182	11,725
Bases	1,001,968,894	307,584,001	72,490,082
Average Read Length	2,379.0 ± 3,528.4	3,934.2 ± 3,514.5	6,182.5 ± 4,218.7
Reads >5kbp	60,760	19,569	6,913
Bases, reads >5 kbp	591,145,838	180,537,522	64,177,899
Avg Read Length, reads >5kbp	9,729.2 ± 3,814.6	9,225.7 ± 2,985.7	9,283.7 ± 2,378.6

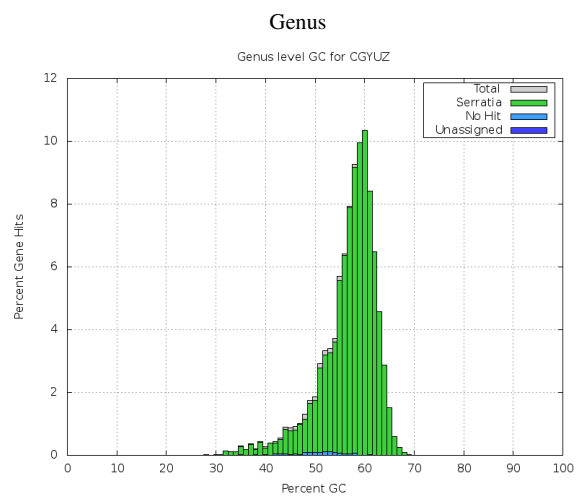
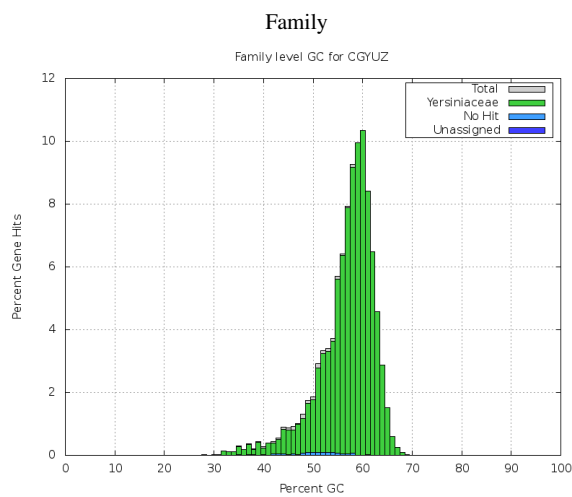
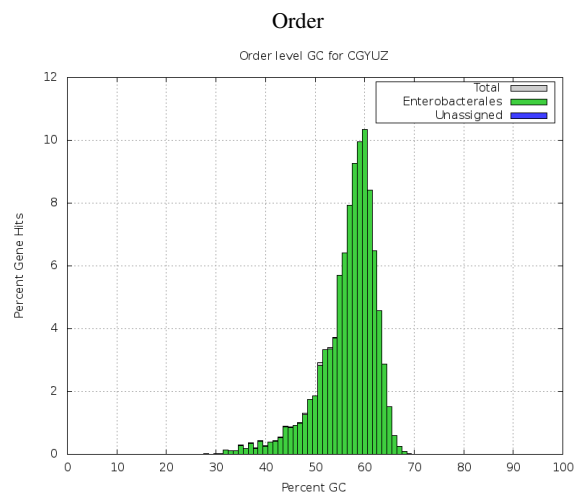
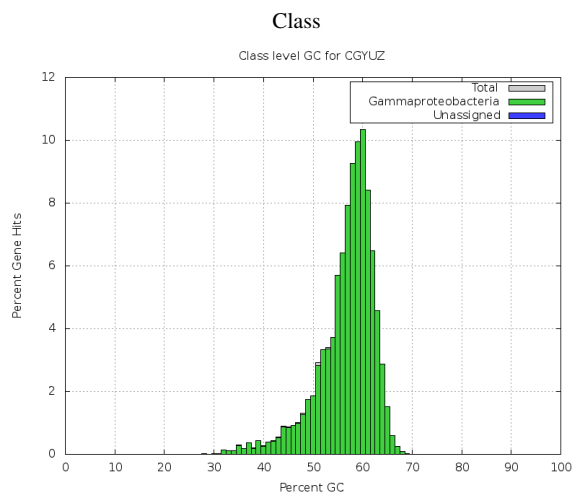
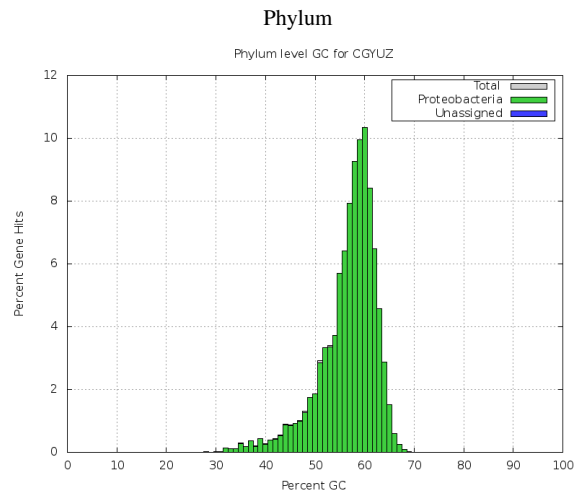
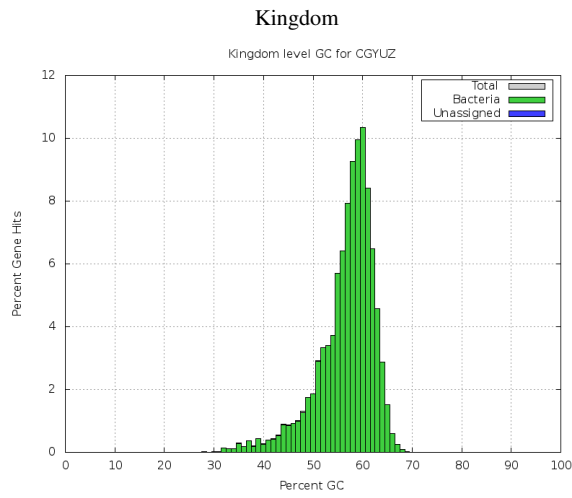
3. Assembly Statistics

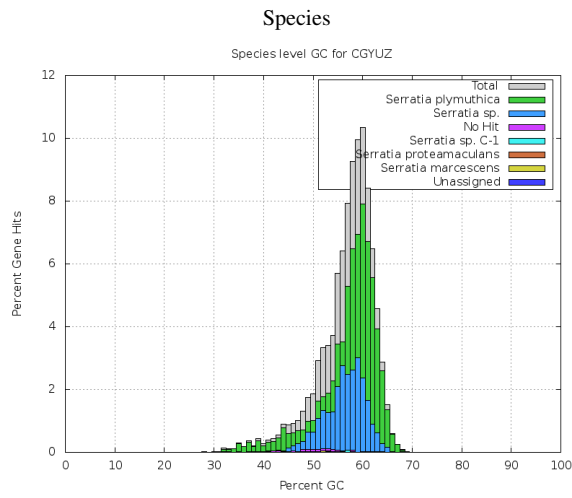
The filtered subreads were assembled using PacBio's HGAP assembler.
 Assembly version: smrtanalysis/2.3.0.p5, HGAP 3

Scaffold total	1
Contig total	1
Scaffold sequence length (bp)	5,349,225
Contig sequence length (bp)	5,349,225
Scaffold N/L50	1/5349225
Contig N/L50	1/5349225
Largest Contig (bp)	5,349,225
Number of scaffolds >50 kb	1
Percent of genome in scaffolds >50 kb	100.00%
Percent of reads assembled	83.32%

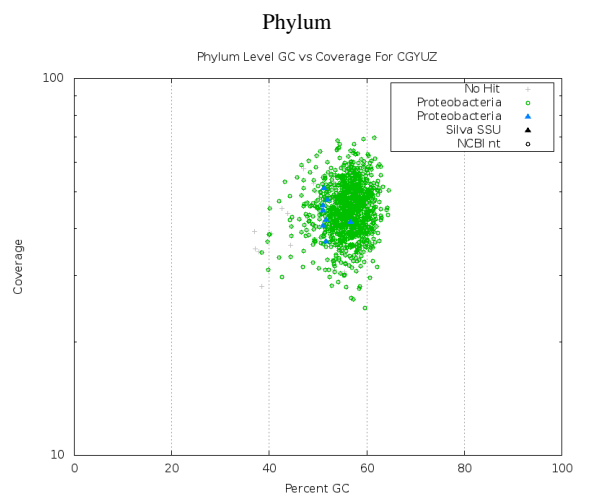
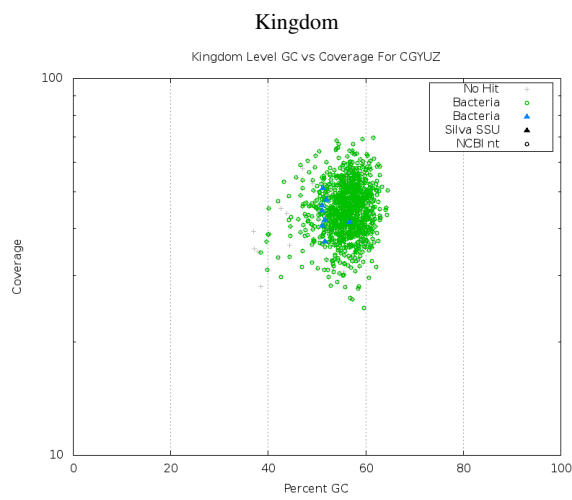
4. Assembly QC Results

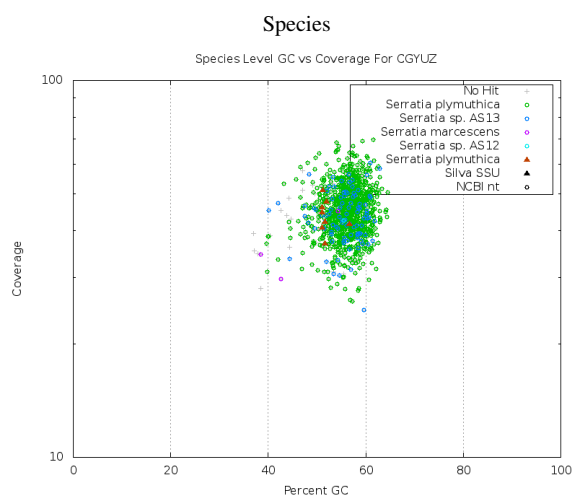
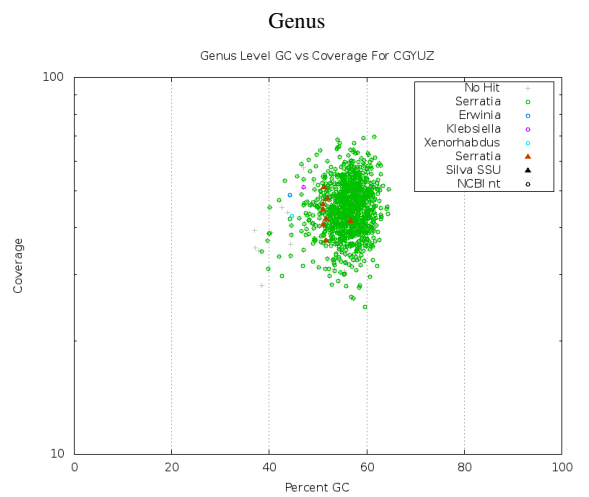
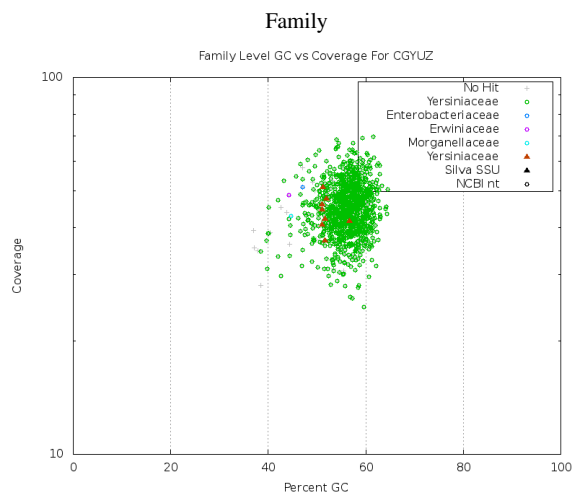
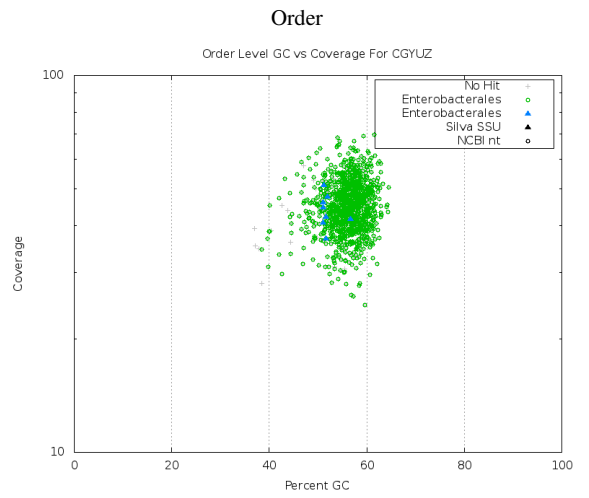
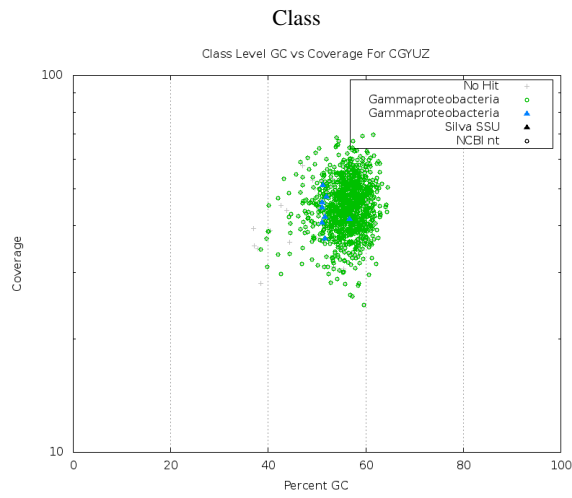
Prodigal was used to predict cds on each scaffold and the output protein sequences were aligned to NCBI nr using LAST. Taxonomic information was extracted from the alignments and used to color-code scaffold GC content histograms.



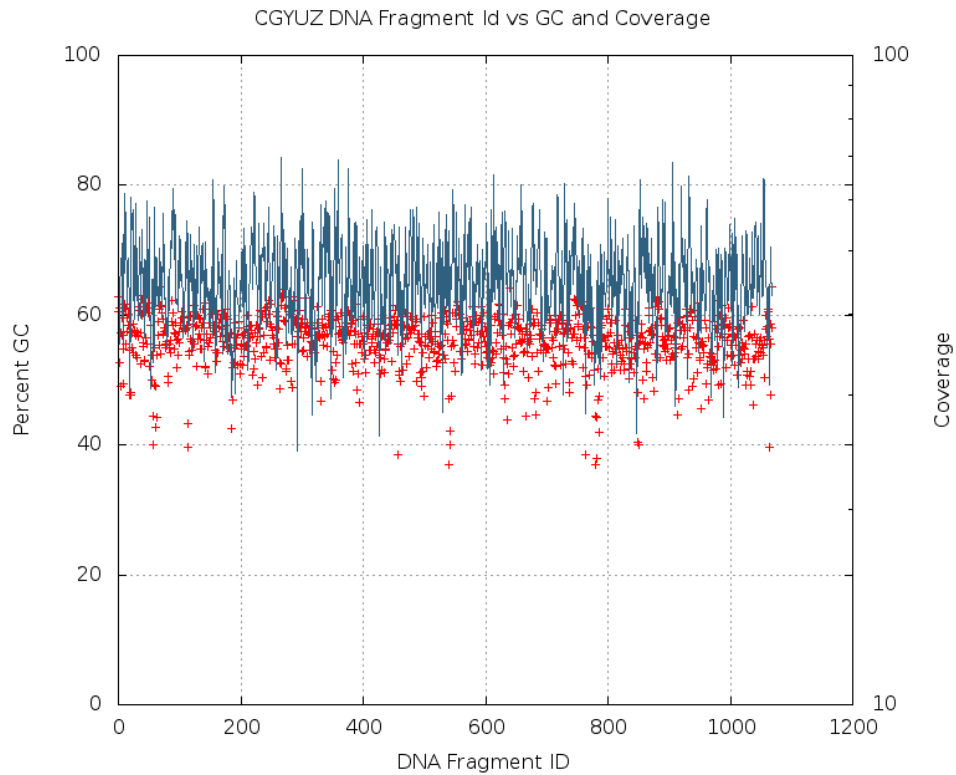


GC versus coverage of assembled scaffolds, overlaid with Silva SSU gene hits and NCBI nt megablast hits shown for different taxonomic levels.

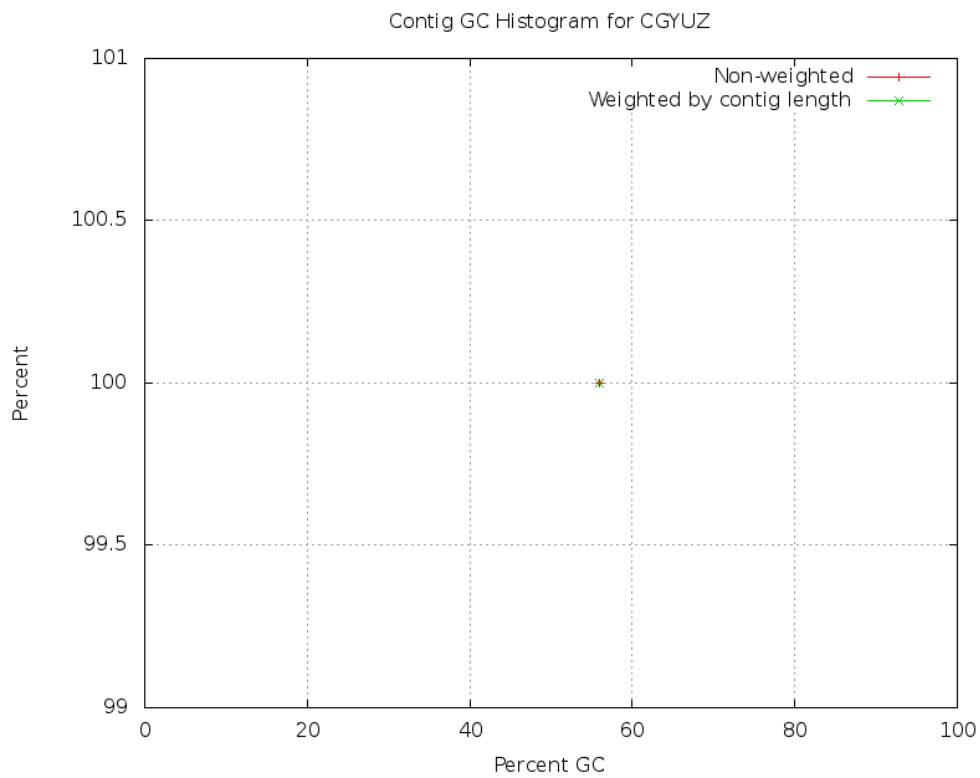




Coverage and GC information. Scaffolds were shredded into non-overlapping 5 kbp fragments and the GC content of each shred was plotted as a data point, colored by scaffold id. Coverage was calculated by mapping the fragment library to the final assembly and plotted as connected points.



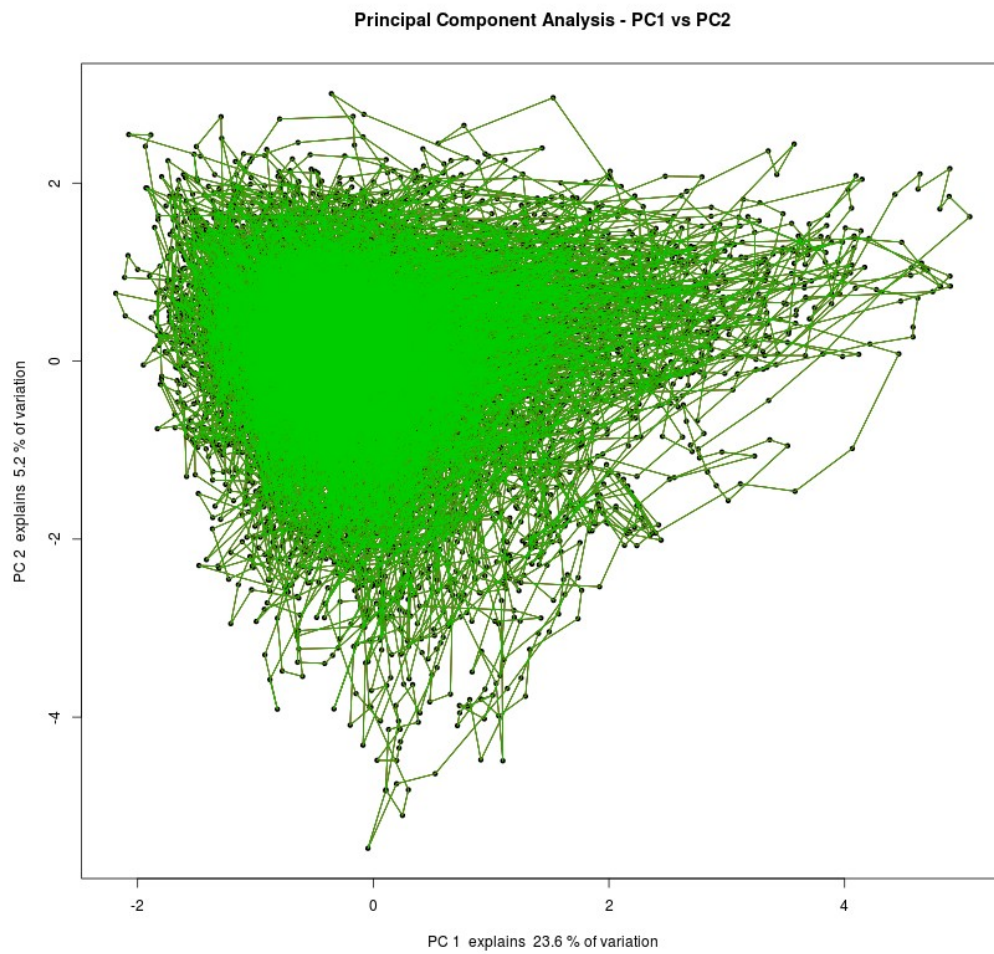
GC histogram of the scaffolds, including scaffold length weighted distribution.



List of the top scaffold megablast hits against 16s ribosomal genes using the Silva SSU database.

Organism: N/ABacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Yersiniaceae;Serratia;Serratia ply-
 muthica;
 Contig Name: CGYUZ_unitig_0|arrow
 Align Length: 1,522 bp
 Percent Id: 100.00%

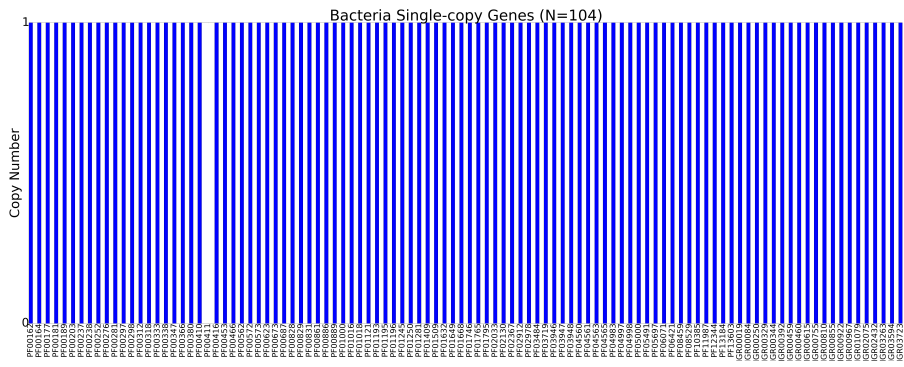
Tetramer frequencies are calculated over 5kb sliding windows of all scaffolds, followed by principal component analysis. Plots of the first two principal components are colored by scaffold.



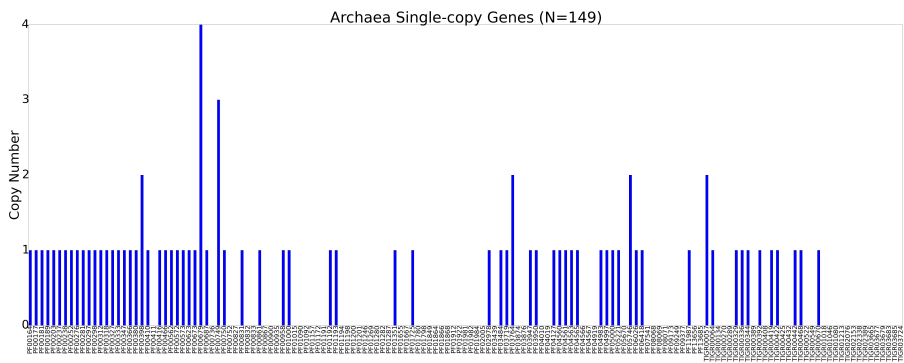
Estimated genome recovery derived from analysis of universal single-copy genes detected in final assembly using CheckM.

HMM	Found Genes	Total Genes	Percent Recovered
Archaea	70	149	46.98%
Bacteria	103	104	99.04%
Lineage Workflow	544	546	99.63%

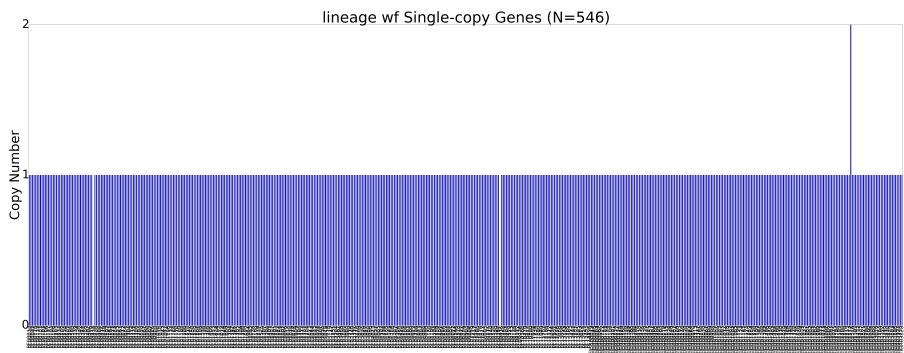
Bacteria Single-copy Gene Histogram

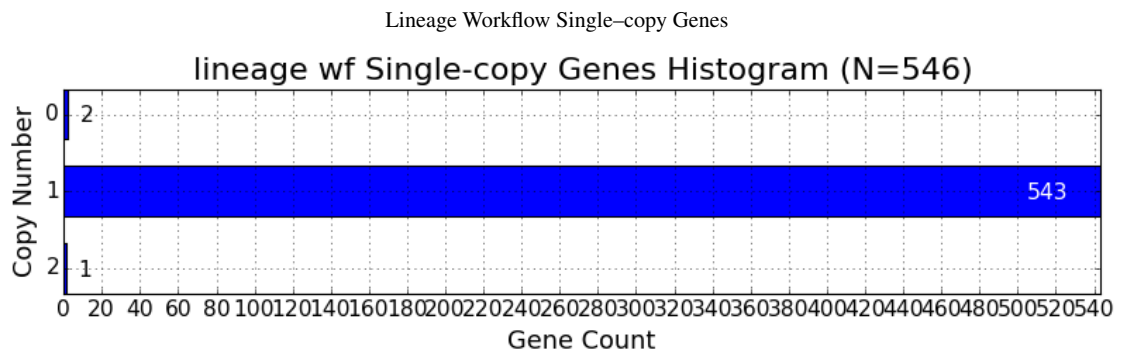
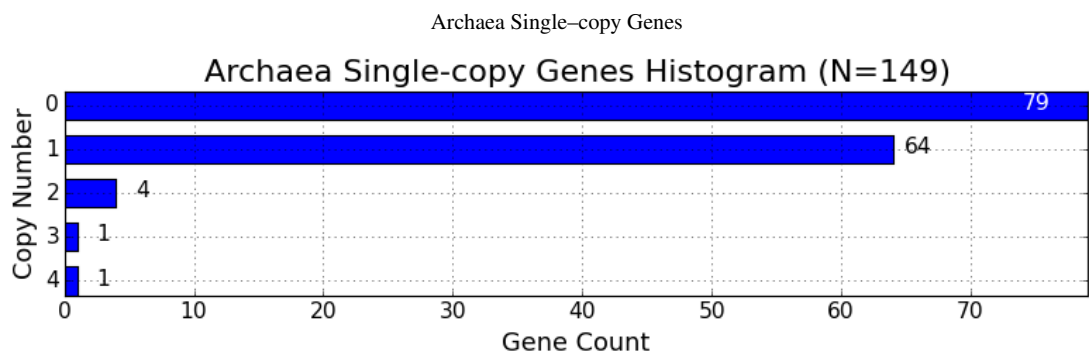
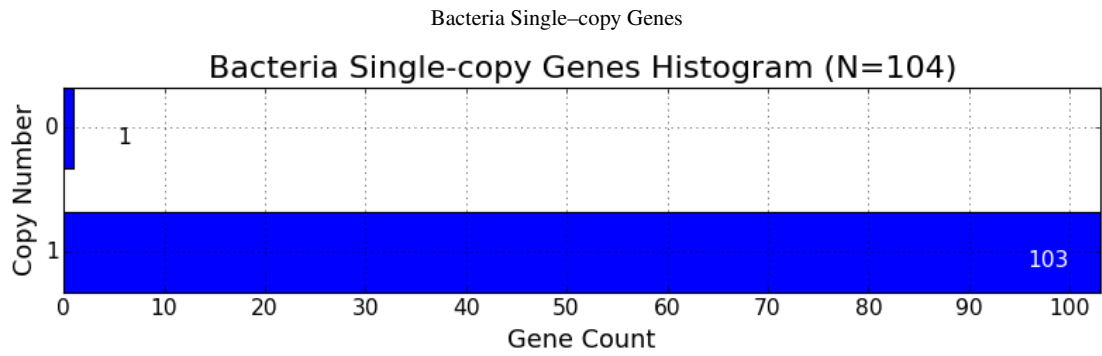


Archaea Single-copy Gene Histogram



Lineage Workflow Single-copy Gene Histogram





5. Sequence Data Availability

The sequence fasta files can be downloaded from our JGI portal website.
<http://www.jgi.doe.gov/genome-projects>

6. Methods

Isolate Improved Draft

Genome Sequencing and Assembly

The draft genome of *Serratia plymuthica* was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology [1]. A >10kbp Pacbio SMRTbell™ library was constructed and sequenced on the PacBio RS2 platform, which generated 78,182 filtered subreads totaling 307,584,001 bp. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov>. The raw reads were assembled using HGAP (smrtanalysis/2.3.0.p5, HGAP 3) [2]. The final draft assembly contained 1 contig in 1 scaffold, totaling 5,349,225 bp in size. The input read coverage was 45.2X.

1. Eid John, et al. RealDNA Sequencing from Single Polymerase Molecules. Science 2008
2. Chin C, et al. Nonhybrid, finished microbial genome assemblies from longread SMRT sequencing data. Nat Methods 2013

DOE Auspice Statement for Publication

The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231.

The data was generated for JGI Proposal #503161.

A.1.3 Investigations into ambigol biosynthesis

The supplemental information is related to the following manuscript which is included in chapter 3.3:

- **E. R. Duell***, T. M. Milzarek*, T. F. Schäberle, G. M. König, T. A. M. Gulder:
Investigations into ambigol biosynthesis. *submitted*

*equally contributing authors

Investigations into Ambigol Biosynthesis

Elke R. Duell,^a Tobias M. Milzarek,^{ab} Mustafa El Omari,^c Luis J. Linares-Otoya,^{de} Till F. Schäberle,^{de} Gabrielle M. König^c and Tobias A. M. Gulder^{ab*}

^a Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748, Germany. E-mail: tobias.gulder@ch.tum.de

^b Chair of Technical Biochemistry, Technische Universität Dresden, Bergstraße 66, 01062 Dresden, Germany. E-mail: tobias.gulder@tu-dresden.de

^c Institute for Pharmaceutical Biology, University of Bonn, Nußallee 6, 53115 Bonn, Germany.

^d Institute for Insect Biotechnology, Justus Liebig University of Giessen, Heinrich-Buff-Ring 26–32, 35392 Giessen, Germany.

^e Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Giessen, Germany.

Supporting Information

Contents

1. General Methods	S2
1.1 Biochemical Methods.....	S3
1.2 Chemical Methods.....	S5
2. Cloning of the Ambigol BGC and the Two Cytochrome P450 Monooxygenases ab2 and ab3	S6
2.1 List of Strains, Primers and Plasmids.....	S6
2.2 Cloning Methods.....	S9
2.3 Plasmid Maps.....	S10
2.4 Agarose Gels of Q5 Screening PCRs of Transformed <i>S. elongatus</i> Cells.....	S11
2.5 DNA Sequences.....	S13
3. Heterologous Expression of the ab BGC in E. coli BAP1	S18
4. SDS PAGE Gels of Purified Ab2, Ab3, Fpr and PetF used for in vitro Assays	S18
5. Experimental Procedures	S19
5.1 Synthesis of Methoxy Protected Aryl Iodides (9–11).....	S19
5.2 General Procedure for the Synthesis of Arylboronic Acids (12–14).....	S20
5.3 General Procedure for the Synthesis of Diaryliodonium Tetrafluoroborates (15–17).....	S21
5.4 General Procedure for the Suzuki Coupling (18–20).....	S22
5.5 General Procedure for the Biaryl Ether Formation (21–26).....	S24
5.6 General Procedure for the Methoxy Deprotection (27–34).....	S26
6. LC-MS Data	S30
7. ¹H-NMR and ¹³C-NMR Spectra	S33
8. UV-absorbance Measurements	S102
9. References	S108

1. General Methods

1.1 Biochemical Methods

Cultivation: *E. coli* cells were grown in liquid culture in LB medium (Carl Roth, Germany) at 37 °C under 200 rpm shaking. *S. elongatus* liquid cultures were grown in BG11 medium (Sigma-Aldrich/Merck, Germany) at 30 °C with 150 rpm shaking and constant 300 LUX illumination from a LED lamp (Fluval A3972 AURA Fresh and Plant Nano LED, 7500 K, Canada). Cells on plates were kept at room temperature with constant 250 LUX illumination growing on BG11 containing 1.5% agarose (Sigma-Aldrich/Merck). Chemically competent *E. coli* DH5 α cells were used as a host strain for all cloning steps. Enzymes used for PCR and cloning were purchased from NEB (Germany). Antibiotics were used at the following concentrations: ampicillin 100 mg/L (Carl Roth, Germany), kanamycin 50 mg/L for *E. coli* cells and 5 mg/L for *S. elongatus* cells (Carl Roth), spectinomycin 2 mg/L (Merck/Sigma-Aldrich, Germany) and streptomycin 2 mg/L (Merck/Sigma-Aldrich).

DNA Isolation, Sequencing and Bioinformatic Analysis of the *ab* BGC: To isolate genomic DNA for subsequent sequencing, first an axenic culture of the *Fischerella ambigua* had to be generated, since a *Pseudomonas* strain was associated with it. Therefore, a spheric cyanobacterial colony was dispersed in 500 μ L water and the resulting cell suspension was then transferred into 5 mL LB medium supplemented with ampicillin (final concentration 500 μ g/mL). The culture was incubated overnight at 37 °C while shaking at 180 rpm. Cyanobacterial cells were recovered by centrifugation, resuspended in 100 mL BG11 medium and grown at 25 °C with shaking at 120 rpm under constant illumination. Cells were harvested after 2–3 weeks of growth, washed three times with sterile water and resuspended in 20 mL SET buffer (75 mM NaCl, 25 mM EDTA, 10 mM Tris-HCl). Using a Potter homogenisator, filaments were broken up and cells separated. SDS 0.5%, proteinase K 500 μ g/mL and lysozyme 2.5 mg/mL were added and the suspension was incubated at 55 °C for 2 h. Subsequently, an extraction with one volume of phenol:chloroform:isoamyl alcohol (25:24:1) was performed for 30 min. This step was repeated until no precipitated proteins were visible between the aqueous and the organic phase. The resulting supernatant was purified using genomic tips, which were supplied with the Blood & Cell Culture DNA Mini Kit (Qiagen, Germany) based on the manufacturer's instruction manual. The resulting DNA was submitted to 454 sequencing applying the Roche GS FLX Titanium sequencer (GATC Biotech, Germany). Open reading frames (ORFs) were predicted with the Geneious 8.0.4 software (Biomatters Limited, New Zealand) and bioinformatically analysed using the NCBI/blastx alignment tool (U.S. National Library of Medicine, USA).¹

Direct Pathway Cloning of the *ab* BGC: The 14.3 kb *ab1-10* BGC was cloned in one piece into pET-28b-SUMO using Gibson Assembly mediated DiPaC.² The linearized *ab* PCR amplicon was generated using PCR with Q5 polymerase with 15 ng gDNA per 25 μ L reaction setup and integrated into the *Bam*HI (NEB) linearised and dephosphorylated vector backbone. The *gfp* reporter gene was amplified by PCR with Q5 polymerase using 100 ng plasmid DNA per 25 μ L reaction setup and integrated into the *Not*I linearized and dephosphorylated pET-28b-SUMO::*ab* plasmid. This expression constructed was transformed into pET-28b-ptetO::*ab-gfp* by exchanging the promoter region with Gibson assembly. Positive clones were selected by Taq screening PCR and confirmed using restriction digest. Mutation free amplification and assembly was verified by Sanger sequencing (GATC Biotech). For expression of the *ab* BGC in the cyanobacterial host *S. elongatus* PCC 7942, the cluster was split in

two parts of roughly equal size with respect to ORF boundaries and integrated into the two neutral sites NSI and NSII of the bacterial chromosome. Therefore, the 6.4 kb *ab1-5* fragment was amplified from plasmid DNA and equipped with homology sequence overhangs by Q5 polymerase PCR and integrated into the *EcoRI* linearised and dephosphorylated pAM5051 vector by Gibson Assembly mediated DiPaC. Positive *E. coli* DH5 α clones were selected by Taq screening PCR and confirmed using restriction digest Sanger sequencing of the integration sites. The 7.8 kb *ab6-10* fragment was integrated into pCV0094 in a similar manner. *S. elongatus* cells were naturally transformed with 0.5 μ g pCV0094::*ab5-7* as described previously.³ After selection of positive clones on BG11_{Kan} plates (BG11 Broth (Sigma-Aldrich/Merck), Trace Metal Mix A5 with Co (Sigma-Aldrich/Merck), 1.5% (w/v) agarose (Sigma-Aldrich/Merck), 1 mM Na₂S₂O₃ (Grüssing, Germany) and verifying complete integration of *ab6-10* into NSII on all genome copies by Q5 PCR on liquid cultures³, the natural transformation and selection procedure was repeated with pAM5051::*ab1-5* resulting in double mutant *S. elongatus* cells harboring *ab1-5* in the chromosomal integration site NSI and *ab6-10* in NSII. Double mutants were grown in BG11 liquid medium or on BG11 agar supplemented with Kan, Spec and Strep.

Heterologous Expression of the *ab* BGC in *E. coli* BAP1 and *S. elongatus* PCC 7942: pET-28b-SUMO::*ab* and pET-28b-SUMO::*ab-gfp* were transformed into chemically competent *E. coli* BAP1 cells. Main cultures were inoculated 1:50 into LB_{Kan} medium (containing 10 g/L NaCl, Carl Roth) from overnight precultures grown in LB_{Kan} and grown at 37 °C 200 rpm until an OD₆₀₀ value of around 0.6. At this point, 20 mg/L 5-aminolevulinic acid (Carl Roth) were added to support formation of the heme cofactor and its incorporation into the target CYP P450 enzymes. Protein expression was induced at OD₆₀₀ around 0.8-1.0 by the addition of 1 mM IPTG (Acros Organics, USA) in case of the pET28b-SUMO based plasmids or 0.5 mg/L tetracycline (Fluka, Switzerland) for pET-28b-ptetO::*ab* plasmid and cells were incubated at 16 °C 200 rpm for 20 h to 5 days with or without the daily addition of 100 mg/L 4-HBA (**2**) (Carl Roth). After harvesting the cells, the pellets were once washed with 0.9% NaCl. Culture supernatant and cell pellets were either directly extracted or stored at -20 °C.

S. elongatus cells harbouring *ab1-5* in NSII and *ab6-10* in NSI were grown in BG11 medium supplemented with Kan, Spec, Strep and 1 g/L NaCl until an OD₇₅₀ of 1.0-1.3. Protein expression was induced with 2 mM theophylline (200 mM stock in DMSO, Acros Organics) and 100 mg/L **2** was added daily for 5 days. After harvesting the cells, the pellets were once washed with 0.9% NaCl. Culture supernatant and cell pellets were either directly extracted or stored at -20 °C.

Culture supernatants were acidified with HCl to a pH of 3 and extracted twice with 1 vol EtOAc. Organic extracts were combined and evaporated *in vacuo*. The residues were resuspended in MeOH or ACN, filtered through a 0.45 μ m PTFE syringe filter (Fisher Scientific, Germany) and analysed by HPLC-MS. Cell pellets were resuspended in 2:1 DCM:MeOH and sonicated for 30 min in a Branson ultrasonic bath 3510 E-MTH (Branson, USA). After centrifugation at 8.000 g for 10 min at 4 °C, the supernatant was evaporated *in vacuo* and the residues treated as described above.

Cloning and Heterologous Expression of Ab2 and Ab3 in *E. coli*: Codon optimised gene sequences of *ab2* and *ab3* (*ab2e* and *ab3e*) for heterologous overexpression in *E. coli* were ordered at BaseClear B.V. (The Netherlands). *Ab2* was integrated into the pMal expression plasmid by Gibson assembly, thereby deleting the ATG start codon and directly attaching the sequence to the N-terminal MBP-tag. *Ab3* was integrated into pMal by restriction cloning using *Bam*HI and *Xho*I and ligation, thus preserving the 102 bp linker sequence between the MBP-tag and the MCS containing a TEV cleavage site.

pMal-SD::*ab2e* and pMal-SD::*ab3e* were transformed into chemically competent *E. coli* BL21 cells. Expression cultures were inoculated 1:50 into TB_{Amp} medium (Carl Roth) from overnight precultures grown in LB_{Amp} medium. Protein expression was otherwise performed as described for the pET-28b-SUMO::*ab-gfp* construct in *E. coli* BAP1 cells but no **2** was added and cells were always harvested after 20 h.

Purification of Ab2 and Ab3 from *E. coli* Heterologous Expression: Cells were resuspended in buffer A (50 mM Tris (Carl Roth) [pH 7.3], 5% (v/v) glycerol (Fisher Scientific) and sonicated nine times for 10 s with 80% amplitude with a Bandelin SONOPULS Homogenisator HD 2070 MS 72 (Bandelin electronics, Germany). Cell debris was removed by centrifugation at 18.000 g at 4 °C for 30 min. Protein purification was carried out at 4 °C with an Äkta Pure 25 M system (GE Healthcare Life Science, Germany) and a MBPtrap HP 5 ml (GE Healthcare) using the following protocol: loading at 4 mL/min, washing with 100 mL buffer A at 4 mL/min, elution with 3.5 mL buffer B (buffer A containing 10 mM maltose, Carl Roth) at 2 mL/min. Protein concentration was determined using an Implen Nanophotometer P330 5382 (Implen, Germany) and the following parameters: Ab2: 97537.67 Da, $\epsilon = 126420$; Ab3: 101088.93 Da, $\epsilon = 124930$.

Ab2 and Ab3 *in vitro* Enzymatic Assays: Purified enzymes were used for enzymatic conversion of 2,4-DCP (**6**) without further concentration. Assays consisted of 20-60 μ M Ab2 and/or Ab3, 3 mM 2,4-DCP (**6**) (300 mM stock in DMSO, Sigma-Aldrich), 4 mM NADPH (100 mM stock in H₂O, Carl Roth), 20 μ M spinach ferredoxin (PetF) and 10 μ M ferredoxin reductase (Fpr). Assays were conducted in 250 μ L scale in 1.5 mL microreaction tubes under air permeable conditions and shaking with 400 rpm at 30 °C for 15 h. After acidifying the mixture with HCl to a pH of 3, organic compounds were extracted with 1 vol EtOAc. After evaporation, the remaining solid was solubilised in MeOH or ACN, filtered through a 0.45 μ m PTFE syringe filter and analysed by HPLC-MS.

Cloning of Ab2 and Ab3 in *S. elongatus* PCC 7942: For chromosomal integration and expression of *ab2* and *ab3* in *S. elongatus*, the native gene sequences were integrated into the *Eco*RI linearized pCV0094 plasmid using Gibson assembly. After verifying the successful cloning by Taq screening PCR, analytical restriction digest and Sanger sequencing, *ab2*, *ab3* or both genes joined together with a 4 bp linker sequence were integrated into the NSII site of the cyanobacterial chromosome by natural transformation. Successful and complete integration into all genome copies was verified by Q5 screening PCR (see Supporting Information Figure S4).³

Ab2 and Ab3 *in vivo* 2,4-DCP Feeding Assays in *E. coli* and *S. elongatus*: To check for Ab2 and Ab3 induced 2,4-DCP (**6**) conversion *in vivo*, *E. coli* BL21 cells carrying pMal-SD::*ab2* or pMal::*ab3* were grown and induced alone or together as described above for heterologous protein production, but were additionally fed with 50 μ M 2,4-DCP (**6**) (300 mM stock in DMSO) and cultivated for 1-3 days. *S. elongatus* cells carrying either *ab2*, *ab3* or both genes were grown to an OD₇₅₀ value of 0.3-0.4 before induction with 2 mM theophylline (200 mM stock in DMSO) and feeding of 50 μ M 2,4-DCP (**6**). Cells were harvested after 48 h. Culture supernatant and cell pellets of both *E. coli* and *S. elongatus* cells were extracted as described above. As negative controls, *E. coli* BL21 and *S. elongatus* wild type cells were cultivated and extracted under the same conditions.

HPLC-MS Analysis of *in vitro* and *in vivo* Extracts: Samples were analyzed using an UltiMate 3000 LC System system coupled to a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). Interpretation of the recorded data was performed using the Thermo Xcalibur Qual Browser 2.2 SP1.48 software. The chromatographic HPLC separation was carried out on a Hypersil Gold AQ C18 (150 x 2.1 mm, 3.0 μm particle size) HPLC column (Thermo Fisher) with the following gradient with water (A) and acetonitrile (B) as the eluents, both buffered with 0.1% formic acid, was applied: 5% B (0 min) \rightarrow 5% B (4.0 min) \rightarrow 95% B (24.0 min) \rightarrow 100% B (24.5 min) \rightarrow 5% B (25.0 min) \rightarrow 5% B (26.0 min). The flow rate was kept constant at 0.7 mL/min. Alternatively, a Eurosphere II 100-3 C18 A (150 x 4.6 mm, 3 μm particle size) HPLC column (Knauer, Germany) was used with the following gradient with water (A) and acetonitrile (B) as the eluents, both buffered with 0.1% formic acid, was applied: 5% B (0 min) \rightarrow 5% B (5.0 min) \rightarrow 30% B (10.0 min) \rightarrow 30% B (40.0 min) \rightarrow 40% B (45.0 min) \rightarrow 100% B (45.5 min) \rightarrow 100% B (50.5 min) \rightarrow 5% B (51.0 min) \rightarrow 5% B (52.0 min). The flow rate was kept constant at 1.0 mL/min.

Preparative HPLC-based Purification of Ab2 *in vitro* and *in vivo* Assay Products: For structural identification of the products formed by Ab2, large scale *in vitro* and *in vivo* assays were purified by preparative HPLC after extraction using a Eurospher II 100-5 C18 A (250 x 16 mm, 5 μm particle size) HPLC column with precolumn (30 x 16 mm) (Knauer) on a Jasco HPLC system consisting of an UV-1575 Intelligent UV/VIS Detector, two PU-2068 Intelligent Preparative Pumps, a MIKA 1000 Dynamic Mixing Chamber (1000 μl , Portmann Instruments AG, Switzerland), a LC-NetII/ADC and a Rheodyne injection valve. The following gradient with water (A) and acetonitrile (B) as eluents, both buffered with 0.05% TFA, was applied: 5% B (0 min) \rightarrow 5% B (2 min) \rightarrow 100% B (45 min) \rightarrow 100% B (54 min) \rightarrow 5% B (55 min) \rightarrow 5% B (60 min). The flow rate was kept constant at 10 mL/min.

1.2 Chemical Methods

Reagents: Solvents for HPLC and MS analysis such as acetonitrile and methanol were purchased from Fisher Scientific and VWR in a purity of over 99% (HPLC-grade). Water was purified using a TKA GenPure water treatment system and deionized. Dry solvents such as diethyl ether, dichloromethane, 1,4-dioxane, tetrahydrofuran and toluene for procedures under inert atmosphere were prepared by distillation and drying over molecular sieve (3 \AA or 4 \AA). Commercial materials and other solvents were purchased at the highest commercial quality from the providers Acros Organics, Alfa Aesar, Carbolution, Carl Roth, Merck, Sigma Aldrich, VWR, TCI Chemicals and Thermo Fisher Scientific.

NMR: ^1H and ^{13}C Nuclear Magnetic Resonance Spectra (NMR) were recorded on Bruker AV-HD400 and AV-HD300 spectrometers at 298 K. The chemical shifts are given in δ -values (ppm) and are calibrated on the residual peak of the deuterated solvent (CDCl_3 : $\delta_{\text{H}} = 7.26$ ppm, $\delta_{\text{C}} = 77.0$ ppm; $\text{DMSO-}d_6$: $\delta_{\text{H}} = 2.50$ ppm, $\delta_{\text{C}} = 39.5$ ppm). The coupling constants J are given in Hertz [Hz]. Following abbreviations were used for the allocation of signal multiplicities: bs - broad signal, s - singlet, d - doublet, dd - doublet of doublets, t - triplet.

MS: Elektrospray-Ionisation Mass spectra (ESI-MS) were recorded on an Advion expression^L CMS system using a single-quadrupole mass analyzer, a Peak Scientific N118LA nitrogen generator, an Edwards RV12 high vacuum pump and a Jasco PU-1580 Intelligent HPLC Pump, or a LCQ Fleet ion trap system (Thermo Scientific), which was combined with an UltiMate 300 HPLC system. For High

resolution mass spectrometry (HRMS) a Thermo LTQ FT Ultra mass spectrometer was used, and analyses of the recorded spectra were again performed using *Thermo Xcalibur Qual Browser 2.2 SP1.48 Software*.

Chromatography: Thin-layer chromatography (TLC) was performed on precoated plates of silica gel F₂₅₄ (Merck) with UV detection at 254 and 365 nm. Column chromatography was performed on silica gel 60 Geduran[®] Si 60 (40-60 μm) (Merck). High Performance Liquid Chromatograms (HPLC) were recorded on a computer-controlled Jasco system including a UV-1575 Intelligent UV/VIS Detector, DG-2080-53 3-Line Degaser, two PU-1580 Intelligent HPLC Pumps, AS-1550 Intelligent Sampler, HG-1580-32 Dynamic Mixer. A Eurospher II 100-3 C18 A (150 × 4.6 mm) column with integrated precolumn manufactured by Knauer was used. The eluent system consisted of A = H₂O + 0.05% TFA, B = MeCN + 0.05% TFA. The analytical method used the following elution gradient: 0-2 min (5% B), 2-25 min (linear increase to 95% B), 25-28 min (95% B), 28-31 min (5% B) with a flow of 1.0 mL/min. For medium pressure liquid chromatography (MPLC) the Reveleris[®] X2 MPLC system (Grace) was used together with Reveleris[®] Reverse Phase (RP) C18 columns (Grace) using UV-detection at 220 nm, 254 nm, and 280 nm. Air- and moisture-sensitive reactions were performed under argon atmosphere (4.6 = 99.996%) using a Schlenk line. Before application, the flasks were repeatedly evacuated (external heating) and refilled with argon.

2. Cloning of the Ambigol BGC and the Two Cytochrome P450 Monooxygenases *ab2* and *ab3*

2.1 List of Strains, Primers and Plasmids

Table 1 List of strains.

Name	Description	Reference /Source
<i>E. coli</i> DH5α	F ⁻ φ80 <i>lacZ</i> ΔM15 Δ(<i>lacZYA-argF</i>)U169 <i>recA1 endA1 hsdR17</i> (r _K ⁻ , m _K ⁺) <i>phoA supE44 λ⁻ thi-1 gyrA96 relA1</i> , host strain for cloning	NEB
<i>E. coli</i> BL21	F ⁻ <i>ompT hsdS_B</i> (r _B ⁻ , m _B ⁻) <i>gal dcm</i> (DE3), heterologous expression strain for <i>ab2/ab3</i> expression and <i>in vivo</i> feeding assays	NEB
<i>E. coli</i> BAP1	F ⁻ <i>ompT hsdS_B</i> (r _B ⁻ , m _B ⁻) <i>gal dcm</i> (DE3) Δ <i>prpRBCD::T7prom-sfp,T7prom-prpE</i> , heterologous expression strain for <i>ab</i> BGC expression	4
<i>S. elongatus</i> PCC 7942	Host strain for <i>ab</i> BGC recombinant expression	5

Table 2 List of primers.

Name	Sequence (5' → 3')	Description
GA_ab:T7_for	CTCACAGAGAACAGATTGGTGGATCCa tgtctaatttacaaaatctac	amplification of the ambigol BGC with homology sequences (capitals) to the pET-28b-SUMO plasmid
GA_ab:T7_rev	TGGCGGCCGTTACTAGTGtcaagcagtaa ttaaactcg	
GA_ab2e:pMal-SD_for	CCCTGAAAGACGCGCAGACTctgtaccag gaggttgcatc	amplification of the codon optimized <i>ab2</i> gene with homology sequences (capitals) to the pMal plasmid without spacer sequence between MBP-tag and MCS
GA_ab2e:pMal-SD_rev	CAAGCTTGAATTCCTCGAGGGTttagaa gctcggtttgacgac	
GA_ab2::pCV0094_for	GCTAAGGAGGCAACAAGatggttatcaa gaagtagcc	amplification of the <i>ab2</i> gene with homology sequences (capitals) to the pCV0094 plasmid including a C-terminal <i>PmeI</i> restriction site (underlined)
GA_ab2::pCV0094_rev	CGAGGTCGAGACGGCTCATG <u>gtttaaact</u> aaaagctaggtttacaacaac	
ab3e_BamHI_for	ggac <u>ggatcc</u> atggtctctcagc	amplification of the codon optimized <i>ab3</i> gene <i>BamHI</i> with restriction site (underlined)
ab3e_XhoI_rev	atat <u>ctc</u> gagttaaaagcgcggtttcg	amplification of the codon optimized <i>ab3</i> gene <i>XhoI</i> with restriction site (underlined)
GA_ab3::pCV0094_for	GCTAAGGAGGCAACAAGatggtttcacia caatcagc	amplification of the <i>ab3</i> gene with homology sequences (capitals) to the pCV0094 plasmid including a C-terminal <i>PmeI</i> restriction site (underlined)
GA_ab3::pCV0094_rev	CGAGGTCGAGACGGCTCATG <u>gtttaaact</u> tagaagcgaggtttgctataactttatag	
GA_gfp::T7-ab_for	CAGATATCCATCACACTGGCatggttagc aaagtggaag	amplification of the <i>gfp</i> gene with homology sequences (capitals) to the pET-28b-SUMO:: <i>ab</i> plasmid
GA_gfp::T7-ab_rev	CAGTGGTGGTGGTGGTGGTCTCGAG TGcttagctgcctttatacagttc	
Ab1_for	atgtctaatttacaaaatctacaaggtattagtt g	amplification of the pET-28b-SUMO:: <i>ab_gfp</i> plasmid without the T7 promoter, lac operator, 6xHis, thrombin site and SUMO-tag
pET28b_rev	gagatctcgatcctctacgc	
GA_ptetO::ab-GFP_for	GCGTAGAGGATCGAGATCTCAATTC	amplification of the lox71 site and tetracycline inducible promoter from pET28b-ptetO plasmid with homology sequences to the pET28b-SUMO:: <i>ab_gfp</i> plasmid
GA_ptetO::ab-GFP_rev	GATTTTGGTAAATTAGACATggtcgatcct cttctctatc	
GA_pMal-SD_for	accctcgaggaattcaagc	amplification of the pMal plasmid without the 105 bp spacer sequence between MBP-tag and MCS
GA_pMal-SD_rev	agtctgcgcgtctttcag	
T7 SUMO_for pET-RP pRSET-RPnew PaadA_for aphI_rev	taatacgactcactataggg ttggacatggaggataacg ctagttattgctcagcgg gggttatgctagtattgctc tgccttcacccgtttccac attccgtcagccagtttag	sequencing and screening primers binding plasmid backbones and facing the MCS
S7942NSI-LA_for	cagatcaatgcccggtggttg	screening primers binding the homologous recombination site NSI in the chromosome of <i>S. elongatus</i> PCC7942
S7942NSI-RA_rev	gaaaagctcaagcgaaggg	

S7942NSII-LA_for	ctccagtaaagtcttcgccgtaac	screening primers binding the homologous recombination site NSII in the chromosome of <i>S. elongatus</i> PCC7942
S7942NSII-RA_rev	ttggtgctgttcagctctggatgc	
ab1_screen_rev	ggcgcaacttgataactg	sequencing and screening primers binding inside the <i>ab</i> BGC
ab1_seq_for	cctttgctttcaactggag	
ab2_seq_for	ccacagtactagaagctg	
ab2_seq_rev	agaacacccccaccgtaac	
ab3_seq_for	caagctcaacaatggtgc	
ab3_seq_rev	tgaagtgacattgaggtgg	
ab3_seq_rev_2	caaactgcgtcaacgagc	
ab4_seq_for	gagaaggaaccagaaagc	
ab5_screen&seq_for	acttgctcattggctcac	
ab6_screen_rev	agaacaactccatttaggcg	
ab6_seq_rev	agctttgacgagttcacc	
ab7_seq_for	aaggattattgacagcacg	
ab7_seq_rev	accaccaccacgtaaaatg	
ab8_seq_for	gagttattgcccgatattgg	
ab9_seq_for	tacctacagaccgtgcc	
ab10_seq_rev	cctatcaagaacgacaactgag	
ab10_seq-rev_2	acgatcaagttccaccatc	
ab10_screen_for	agaagacacaatgccagag	

Table 3 List of plasmids.

Name	Description	Reference /Source
pET-28b-SUMO	IPTG inducible expression plasmid (T7 promoter), <i>N</i> - and <i>C</i> -terminal 6xHis-tag, thrombin site, SUMO-tag, ColE1, Kan ^R	6
pET-28b-ptetO	Tetracycline inducible expression plasmid (ptetO promoter), <i>N</i> - and <i>C</i> -terminal 6xHis-tag, thrombin site, ColE1, Kan ^R	7
pMal	IPTG inducible expression plasmid (T7 promoter), 6xHis-tag, MBP-tag, TEV site, ColE1, Amp ^R	8
pAM5051	Theophylline inducible expression plasmid (PconII promoter), S7942NSI recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Spec/Strep ^R	9
pCV0094	Theophylline inducible expression plasmid (PconII promoter), S7942NSII recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Tet ^R , Kan ^R	5
pET-28b-SUMO:: <i>ab</i>	IPTG inducible expression plasmid (T7 promoter), <i>N</i> - and <i>C</i> -terminal 6xHis-tag, thrombin site, SUMO-tag, ColE1, Kan ^R harboring the <i>ab</i> BGC	This study
pET-28b-SUMO:: <i>ab-gfp</i>	IPTG inducible expression plasmid (T7 promoter), <i>N</i> - and <i>C</i> -terminal 6xHis-tag, thrombin site, SUMO-tag, ColE1, Kan ^R harboring the <i>ab</i> BGC and the <i>gfp</i> gene	This study
pET-28b-ptetO:: <i>ab-gfp</i>	Tetracycline inducible expression plasmid (ptetO promoter), <i>N</i> - and <i>C</i> -terminal 6xHis-tag, thrombin site, ColE1, Kan ^R harboring the <i>ab</i> BGC and the <i>gfp</i> gene	This study

pAM5051::ab1-5	Theophylline inducible expression plasmid (PconII promoter), S7942NSI recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Spec/Strep ^R harboring the genes <i>ab1-5</i> including the respective intergenic regions in the <i>ab</i> BGC	This study
pCV0094::ab6-10	Theophylline inducible expression plasmid (PconII promoter), S7942NSII recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Tet ^R , Kan ^R harboring the genes <i>ab6-10</i> including the respective intergenic regions in the <i>ab</i> BGC	This study
pMal-SD::ab2e	IPTG inducible expression plasmid (T7 promoter), 6xHis-tag, MBP-tag, ColE1, Amp ^R harboring the codon optimized <i>ab2</i> gene directly attached to the MBP-tag	This study
pMal::ab3e	IPTG inducible expression plasmid (T7 promoter), 6xHis-tag, MBP-tag, TEV site, ColE1, Amp ^R harboring the codon optimized <i>ab3</i> gene	This study
pCV0094::ab2	Theophylline inducible expression plasmid (PconII promoter), S7942NSII recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Tet ^R , Kan ^R harboring the <i>ab2</i> gene	This study
pCV0094::ab3	Theophylline inducible expression plasmid (PconII promoter), S7942NSII recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Tet ^R , Kan ^R harboring the <i>ab3</i> gene	This study
pCV0094::ab2-3	Theophylline inducible expression plasmid (PconII promoter), S7942NSII recombination site for the <i>Synechococcus elongatus</i> PCC7942 chromosome, ColE1, Tet ^R , Kan ^R harboring the <i>ab2</i> and the <i>ab3</i> gene joint by a 4 bp linker sequence	This study

2.2 Cloning Methods

Q5 Polymerase PCR Setup: A standard 25 µL PCR reaction batch for long-amplicon cycling reactions consisted of: 1x Q5 reaction buffer, 200-400 µM deoxynucleotide triphosphates, 500 nM of forward and reverse primer, 15 ng gDNA or 100 ng plasmid DNA template and 0.01-0.02 U/µL Q5 High-Fidelity DNA polymerase (NEB). Cycling was conducted using a T100 Thermal Cycler (Biorad) as follows: 1.) Initial denaturation, 98 °C for 1 min; 2.) Denaturation, 98 °C for 10 sec; 3.) Annealing, t_a °C for 20 sec; 4.) Extension, 72 °C for 30-45 sec/kb; steps 2.) to 4.) were repeated in total for 30 cycles; 5.) Final extension, 72 °C for 60 sec/kb, and 6.) End phase, 16 °C. Primer annealing temperatures were determined using the NEB Tm Calculator online tool (<http://tmcalculator.neb.com>).

Taq Polymerase PCR Setup: Colony screening PCRs were performed using *Taq* DNA polymerase (NEB). Clones were picked and resuspended in 50 µL of pure water and examined in a 25 µL PCR batch composed as follows: *Taq* buffer (10 mM Tris-HCl, 1.5 mM MgCl₂, 50 mM KCl, pH 8.3 at 25 °C), 4% DMSO, 100 µM deoxynucleotide triphosphates, 200 nM of forward and reverse primer, 5 µL DNA template (bacterial suspension in water) and *Taq* DNA polymerase (0.025 U/µL, NEB). Cycling was conducted using a T100 Thermal Cycler (Biorad) as follows: 1.) Initial denaturation, 95 °C for 5 min; 2.) Denaturation, 95 °C for 45 sec; 3.) Annealing, t_a °C for 30 sec; 4.) Extension, 72 °C for 60 sec/kb; steps 2.) to 4.) were repeated in total 34 times; 5.) Final extension, 72 °C for 120 sec/kb, and 6.) End phase, 16 °C. Primer annealing temperatures were determined using the NEB Tm Calculator online tool.

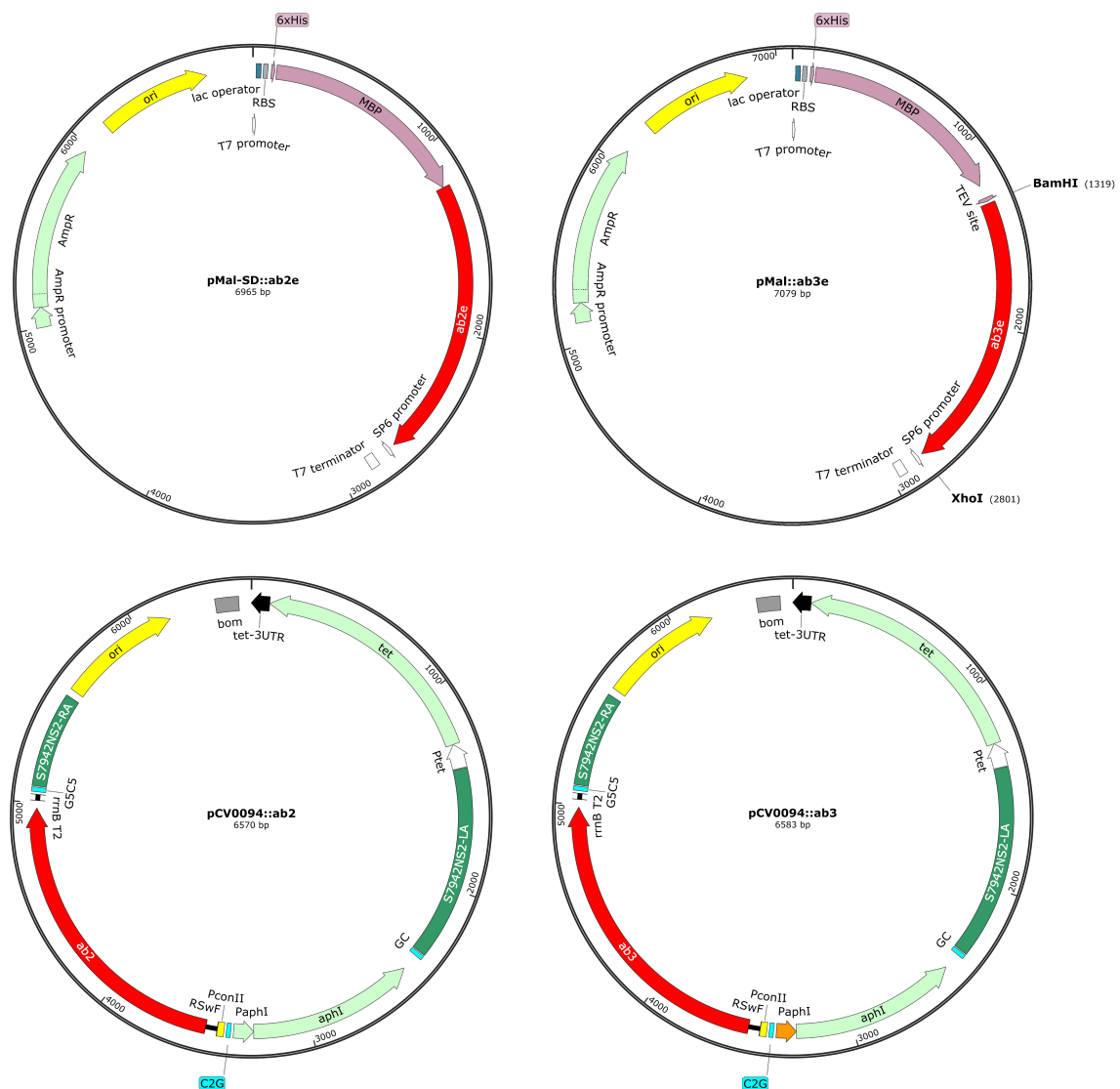


Fig. S2 Plasmids used for heterologous expression of *ab2* and *ab3* in *E. coli* and *S. elongatus*.

2.4 Agarose Gels of Q5 Screening PCRs of Transformed *S. elongatus* Cells

Integration of *ab1-5* into NSI was confirmed by PCR using primer pair PaadA_for/*ab1_screen_rev* expecting 563 bp to verify upstream integration, *ab5_screen&seq_for_1/S7942NSI-RA_rev* expecting 581 bp to verify downstream integration and *S7942NSI-LA_for/S7942NSI-RA_rev* expecting 8140 bp for complete integration of *ab1-5* and 1713 bp for the empty NSI site. Integration of *ab6-10* into NSII was confirmed by PCR using primer pair *aphI_rev/ab6_screen_rev_1* expecting 701 bp to verify upstream integration, *ab10_screen_for/S7942NSII-RA_rev* expecting 447 bp to verify downstream integration and *S7942NSII-LA_for/S7942NSII-RA_rev* expecting 9417 bp for complete integration of *ab6-10* and 1660 bp for the empty NSII site. Sequenced plasmid DNA of pAM5051::*ab1-5* and pCV0094::*ab6-10*

was used as positive control, water and the empty plasmids pAM5051 and pCV0094 were used as negative controls for the PCR.

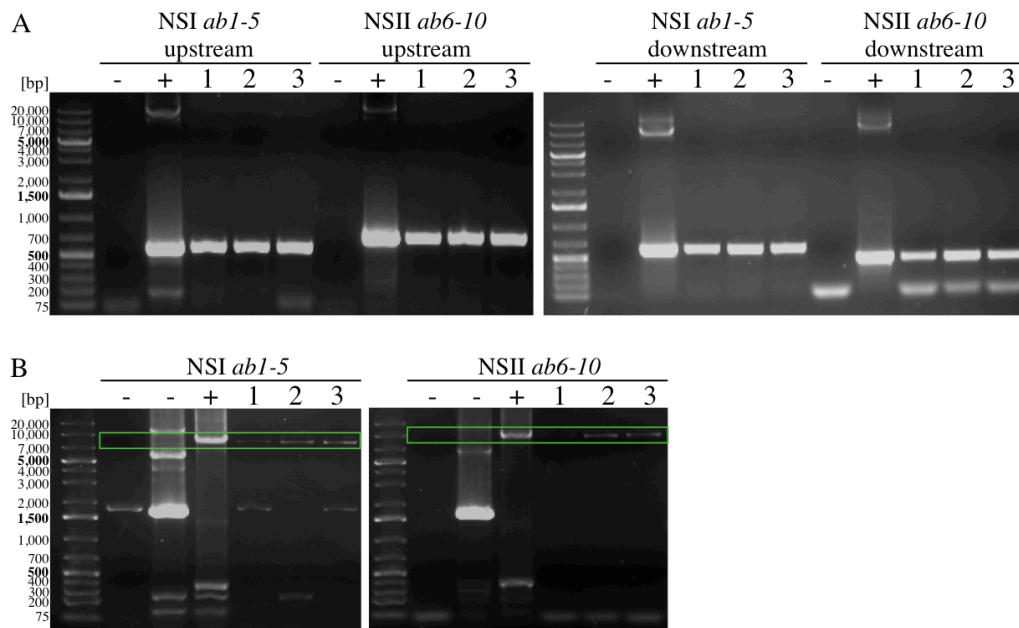


Fig. S3 Agarose gels of Q5 screening PCRs on three *S. elongatus* clones harbouring *ab1-5* in NSI and *ab6-10* in NSII. Verification of A) upstream (left) and downstream (right) integration and of B) complete integration of the respective *ab* cluster parts.

Integration of the cytochrome P450 genes *ab2* and/or *ab3* into NSII was confirmed by PCR using primer pair S7942NSII-LA_for/S7942NSII-RA_rev expecting 3124 bp for *ab2*, 3137 bp for *ab3*, 4605 bp for both *ab2* and *ab3* in succession. Water was used as negative control and sequenced pCV0094::*ab2* plasmid DNA as positive control for the PCR.

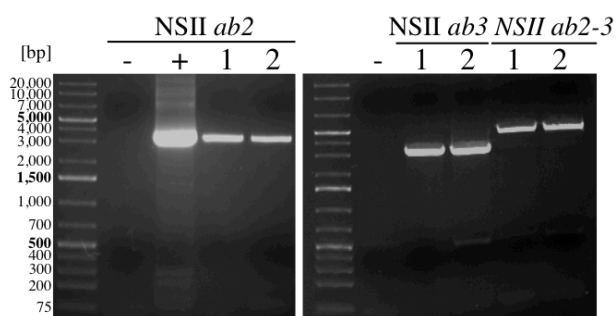


Fig. S4 Agarose gels of Q5 screening PCRs on two *S. elongatus* clones harboring either *ab2*, *ab3* or *ab2-3* in NSII to verify complete integration.

2.5 DNA Sequences

Sequence of the ambigol BGC of *F. ambigua* (Näg.) Gomont 108b, ORFs indicated by color.

ATGTCTAATTTACCAAAATCTACAAAGGTATTAGTTGTTGGGGGTGGACCTGCTGGAACCACTGCTGCTACTCTAT
TAGCTCGAGAAGGTTTTGATATAACGCTGCTTGAAAGAGAAGTCTTCCAAGATACCACATTGGAGAATCTTTAC
TGCCTTCATCTCTAAAAGTTCTAGACTTGCTAGGTGTCAGAGATAAAATAGATGCCCATGGTTTTCAATACAAACC
GGGTGGTCACTATCACTGGGGAGATGAGCATTGGGATTTAAATTTCTCAGATTTATCAGGTAACATTACACACAGT
TATCAAGTGCGCCGTGACGAGTTTGATAAATTACTTTTAGACCATGCCAAAAGCCAGGGAGTGAAGTTTTGAC
GGAATAGGAGTCTCTAGCTTGTCTTTGAGAATGAGCGACCAAAGAGTGCAATTTGGTCTCAAATAATGACAAA
AATCACACAGGAGAGATCTCCTTCGACTTTCTGATTGATGCCACTGGTCTGTTATGGTTTATGGCTAATCATCACTT
AAAGAATAGAGAATATCATGATGATTTCAAATGTAGCTATCTGGGGTATTGGAAGAACGCTGACAGGCTGGA
CAACGGCAGAGAGGGAGCTATTATTATTGAATCTCTCAAAGACGGTTGGTTATGGGGAATTCCACTGCACGATGG
CACAATAAGTGTAGGTCTAGTCGTCCACAAAACCATCTATAAGGAAAAACGCTCTAAGAGCCTGAAAGATATTTAT
TTGGAGGGCATTGCTGAAAGCTTAGACCTCAAGCGTTACTGGAGCCAGGAGAACTTGCATCCGAAGTGCCTTC
TGAGCAGGACTACTCTTACGCAGCTGACTTTTTGCCGGTCAAGTTACTTTATGATTGGAGATGCTGCTTGCTTC
TTAGACCTTTGCTTTCACTGGAGTACACCTTGCTACTTTTAGTGGACTTTTGAGCGCCGCAAGTTTAGCAAGT
GTCATACGCAACCATATTACTGAAGAGCAGGCAATTTCTTTCTTGAAAGAACTTATAACAAGCGTACTTGCCTT
TGATGGCAATGGTTTCAGCCTTCTACGAAAACAGTAAGAAGGAATCTATTTTTGGCAGGCTCAGCAACTGACTA
AAACTCGCCAGAGCAACGAAGATAAAGAAAAATTACACCAAATGTTTTGAATGTGGTCTCTGGTATGGAAGATA
TGTCAGACGCTGAGGAAAACAGAAAGAGCTATTCTAGAGTTATCAGAGCGTTACGGGAAAATTGGTCTCTT
CGTCACAAGCAAACAGCAAATGATTAGATCAGACAGAAGAGGAGAAGTTACGTGCTAGTAATCAGTTTGTAG
TCGGTTAAATGGTCTGTTCTCATTATCAAAGAATCTGCTGTTGAAGGACTGTATATCGTTACAACACCACAGCTAG
GTTTGGTTCAGGTGAATTAAGCAACATAGCAGCAATGGTATGTGGGATCTAGTGCCACGTACCTCAGATGAAA
ATCGTTTTGTTGAGAATAGTTGCCAAATTCATTTAAAGTCGCTATGTTGTATCAAGAAGTAGCCTCCACAGC
GTTCAACAAGATTGCTCCTCCAGGTCCGCCTGTGTACCCTTTGTTGACATGTTGCCCTGCTTAGGTAAGCACTTA
CATTTAGCATTGAATCAACTAGCCAAAAGTACGGTAATATTTTTCAAATTCGTGTGGGTGCCAAAACGTTAGTTG
TACTTAATGGGCTTGAGACTATCAAAGAAGCTTTAGTGAAACAGCCAGATAGTTTTAACGCTAGGGCAGATTTTG
ACATTTATCAGCAACCGCCTCAAGCTCAATTTTTAGAGTTGAAGAGTGGTGAGTCTGGAGAAAACACCATAACA
TTTTAGGTCAAGCTATGCATACCTTTGTGGTAGGCAAGCCAGACATGCTTGAAAGTTGGGCACTAGAAGAAGCA
GCAGACTTGGCAAATATTTCTCAAATTTAGCGGTCAACCTTTGACCCTGATTTGTATATGCCTCTCGCCACCTT
AAGTTTTATGCAAAGACTCATCTTTGACAAAAGAGGCGCTATTAATAATCCTGAAGAGGATCATGAATTTGTTGC
AAGTGCTTATACTTTAAAGCATATCCACAGTACTAGAAGCTGTTAGATTAGAGTACATACAAAATTTGGCAG
CCAATATTTGACTTTCTCGCTGGAAATCATTACGCAACTTTCTCAAGAGTTTGGTTGCACTAGAAAAGCTATGTTTC
AAAAATGTAGCGCAGCATCAAGAGTCCTTCGATCCGAGAATTTGCGGGACATTACAGATGCACTTTTGAAAG
CTAGCAGCGAGCTCACTGAGTCAGACCGAAAACCACTTCAATTTGAGCGAAAATGACATTGTCAATGGATCTTTGA
TGCAGCTCGCGGGGGCAGGTGCGGGACTTGCAGATTTTATGTTACGGTGGGGTGTCTTTACATGATGACCTAC
CCAGCTATTCAAGCCGAAATTCATAAAGAACTTGATGAAGTCGTTGGCAGACAACAACAACCATGCTTAGAGCAT
CGTGGCAAATTGCCATTCACGGAAGCTTGCATCCACGAAATCTTCGCCACTCTTCTACTACTATGCCCCCAT
TACTTATGCCACAACCACAGACGTGACCTTGAGGATTACTTTATCCCAAATACGCCCTTGTGATCAATTACT
ATAGTCTGACTCGTGATCAGCGCTATTGGGAAGAACCCGAGCAATTTAACCTTATCGGTTTTTGGATGAAAACG
GCAAATGAGAAAAACCTTCTTGATAAGTTCTATCCATTTGGAATGGGATCGCGTGGTGCATAGGAGAATATTT
AGGACGTCTTTAGTCTTTACCTTTTTACGAACCTCATGCATAAATGCAAATTTGAAAAGGTACCTGGAGAGAA
ATTGAGTTTTGAATCTATTCCAGGAGCCTTCATAATCCAGAAAAATATAGAGTTGTTGTA AACCTAGCTTTTAGG
TTCCAGGATTTGAATTTGTGAGTAGGTGGCTAAAGATCATTAAAGTAGTCAATGAACTACCACATCACAATTCGCGAGA
AGCCAGA ACTTAAAAAACTTTCTTCAACCATATCCGGACTCTGATGAGCTATTTGTTGTTTGATAGTTAACAACAT
TTAATCGTGTTTATACACTCAGAGTCAGTCTCTTGAAACCTAACAAATATGACCTGACTTTTTGAATTTAGAATTC
TACATCTGAGTAGAGACAACAGAAATTTATAAATTTATCAAAGTTAAGACCAAGTTTTTACATCCTCACAGAAAA
AAAGTTTTGTTGTCATAGCTGCCTCAATTTATCTGAAAATAGCCATGGTTTCAACAACATCAGCTTCCACAGA

GTAAACACGCTTAACAAGATATCTCTCCAGGACCGCCTGCGTTACCCTTTGTCGGCATGTTGCCCTTTTAGGA
AAGCACTTACATTTAGCATTGAATCAACTGGCCAAAAAGTACGGTAATATTATCAAATTCGTGTGGGTGCCAGAA
CGTTAGTTGACTTAATGGGCTTGAGACTATCAAAGAAGCTTTAGTGAAACAGCCAGATAGTTTTAACGCTAGGG
CAGATTTTGACCTTTATCAGCAACCACCTCAATGTCACTTCATGGAGCAGAAAAAGTGGTGAGTCTTGAGAAAAAC
ACCGCACCATTATAGGTCAAGTCATGCATACCTTTGTGGCAAGCAGGTCAGGCACCCTTGAAAAGTTGGGCACTAG
AAGAAGCCGCAGACTTGGCAAATATCTTTGTCAACTCTAGCGGTCAAGCTTTTGACCCTAATTTATATCTGCCTCTA
GCTACCTTAAGTTTTATACAAAGATTGATTTTTGGCAAAAAGAGGCAATCTTGATGATCCTGAAAAGGATACTGATT
TTGTTACAACCTGCTCATAGTATGGGCAAGCTCAACAATGGTGCCAAAATCTTACAAAGTTAGTGCTTCTGCCAAC
AATTTGGCGACCAATATTAATGATCTCTATTTGGAAATCATTACGAGGTTTTGTCAAAGCGGCAGACGCAGCCGAA
GGCTATCTTATAAAAAATGTAGAGCAGCATCAAAGTCTTCGATCCGGAAAAATTTGCGGGACATTACAGATGCA
CTTTAAAAGCTAGCAGCGAGCTCACTGAGTCAGACCGAAATAACCTTGGGTTGAGCGAAAATGACATTGTAA
CGGCTCGTTGACGCAGTTTGTGGGGGCTGGTACAGAGCTCCCAGTCTTATGTTACGCTGGGCTTTGCTTTATAT
GATTACTTACCCAGCTATTCAGGCTGAAATTCATAAAGAAGCTTGATGAAGTAGTTGGCAGACAACAACAACCATG
CTCAGAGCATCGTGGCAAATTACCATTACAGAAGCTTGCATCCACGAAGTCTCCGCCACTCTTCTGCTACTACT
ACACCTGCCTTCATCTACGCTACTACCACTGACGTTACTTTGGACGGTACTTTATCCCCAAAATACACCCTTGCT
TGTCGATTATTATAGTCTGACTCGTGATGAACGTTATTGGGAAAAACCCGAGCAATTAACCCATATCGGTTTTGG
ATGAAAACGGCAAATAAGAAAGAATCTTCTTGATAAGTTTCATCCATTTGGAATCGGCTCGCGCAGGTGTATAG
GAGAGTACATAGGACGTCTTCTCATCTTTACCTTCTCACTCACCTTATGCATAAATGCAAATTTGAAAAGGTACCT
GGTGAGAAATTGAGTCTTGATCCTCAACCAGCGATAATACTCCCTCCACAGAACTATAAAGTTATAGCAAAAACCTC
GCTTCTAAGCTCTAGAGTTTGAATTGCGAGTTAAAAGGATGAGCATACTCTAAGGTTTTGGTATTCAATTCTCAT
TGACAGGAGTTACAGATGAGCAAAAATACGACCTCAACCAGAAATCAGATTTCTCAGGTGCTACTGATAGCTTTA
AAACCGAGAAGGAACCAGAAAAGCTTAAATCCGCCACTAACCAGCTATTCTACAAGATGTGGAAATAAAAATTG
AAAAGCTTGAGGGACGGCAGCGGAGGATCTTTGCCAAAATCCAGATTCCTACCATTAGAGCAAGTCTGGCAA
GTTCTTACAGACTATGAAGCTTTTGCTAAGTTCATGCCTAATATGACACAAAGCCGACGATTGGAACATTCAACCG
CAAGTATTTGTGTCGAACAAGTTCGCACCAAAAGTTTTATGGGCATGAAATTCCTCGCCCGTTCAGTTTTTGATGT
GGAAGAAAAGTTCCACACGAGATTCATTACCAGTTAATTGAAGGAGACTTCAAAGCTTGCTCTGGCTACTGGC
GATTAGAACCTTGAACCTCATCAGATGAAAAAGCAGGAGTTGACCTTATCTATAACTTTTTGATTTGCCAAAGCC
TATCTTTCCAATGCCGCTAGTTGAAAATATTTAAAGCCATGATATACCAGTGGGTATATTAGCGATTCTGCAGCGAG
TGGAAGAGTTATTAGTTCAAATAACTTACTAGGATTTACGCACTGTACAAATGACTCATTACGCATTTAATAAGA
TTTGATAGTTTCAGATTGTCGGCGTGCCAAAATCAATCACCTAAAATAGATAGTTCATCACTCCTGACAGTAGA
AAAATCTCATTTCACTTTAATACCTGAAAATTACGGCGCCACAATTGATTCTAGCTGTCCATGCGTAAATCCTATTAG
AGAAAGAGACGTTTTCTCAGAAAAGTAAACGAATTAATTATTAAGGATAGTATTTAATTATGATTCATCTCCAAAAG
GAATCTTTATCAACTGTCGACTTGCAAGGATGAAGCCAGTTTACCAATTTCTATCCTGACCAATAATTCACAGGAGT
CTCCAAGTCGCAATCCTATAGATCCATCTACTTTGAGTACCTTTCAAAGGATTTTACTGACGACAAATGGAACAGT
CACAGATATTTAGAGTATTACGCATTTGAACAGATCCGAGTTGTGAAATTAGCGGAGCAACTGTCTCATTGGCT
CACGAAATTCATGATGGAATTGAAAGAAGGAACAGAAGTTCTTGTTAGGAAGATTCTACTCAGGGCAAAT
TAGCAGAAAAAATCTTTTATGCCGACTCTATTATTGTGCCTGAAAGGCTAGACGAAAGATTTAGAAAAGCCCTT
TTGGAAACTAAGATGCCAATTGGTAAGCTCTGGTTTGTAGTTAAGGGTAGAAACATTTAAGGAAGTTTTAGATACT
AGTAAAGAAGTTGCGGGAAACTTAGCTGACTATTTTACAGATTCAGCCAGATGACAACATACTTTCTCGCACCTATC
GTGTCATTAATAATCGTAAACCTGTCATGATGATCACTGAAAAGTTCCAGAAAATCTATTATCTGAAATGCTCTTAA
GTTGTTTGTCTATGAGGTGATTTTAGTGCTCACACTTGAGGGAGTTTTGAGGTACTCGTATGTTGACCCAATTAT
TCACTGAGGTGCTATCGAATTATCCTGAAAAGACAGCAATTGTCTATGATCAGACTAAAATCAGCTATCAAACGCT
GTACTCTCAAATCAAAGTTTTAGTCAAGGATTAGGTTCAATTGGTATTGATCAAGGAGATTGTGTTGCACTTCTG
TTGCCTAATTGTCCGAATTTGTCATCAGCTTTTACGCGATCGCTCGCCTAAATGGAGTTGTTCTGCCGCTTAATCA
TTTGTTTAAGGCAGAAGAAGTAAGTCACTACCTGAATGATAGTATGTCGAAAGCTATCATTACTGATTCTCAACGT
GCCGATATATGCAAGAAGATCATCTTAATTTAGGCAAGAAGATAGAGTTAATTGTCGTAGATCAAGCTCCACCTC
CAGCAAAATATTTCTACGACTTAATTCTGCCAAACAGTACAGAAATTCATGAAAGTGTATTACCTTACGAAGGTAAT
GTACTCTACCAGTATTCGCTGGTTCTACAGGACGACCAAGAGGGTGTCTAGAACTCAGAAGAATCTGTATCAT

GAGGCGAGAAATTTACAGAAAGCGTTAAAGTTACACCATCAGACAACATATTGTGTACTGTACCCTGTACCATG
CTCATGGGTTAGGTAAGTGTACTAGCTGCTACTTGCAATGGTTGACGCTGGTGATCCTAGAGCAATCTATACA
AAATGGTGTGTCGGTGGAGGTGCCATTTGTCTTTAAATGTCCAAGAATATTGGAAGTATTAAACAGAGAAAAT
AAGCATTTTTCTGGTGTGCCTTACATTTTTAACTCTCTAGCAGAACTCCTGTCAATATTCAAGCAGATTTATCCA
CACTGAAACTGTGCATCTCTGCAGGAACTTTTTAGGAAAAGATGTTTTCAATAAGTTTTCTCCAACGGTTTTGGAG
TACCGATTTCGACAGCTTTACGGTTGCACAGAAGCAGGCGCTATGTGCATTAATCTAGATGAAAATCCCGAACAAA
CTTGGGATTCAGTCGGTACTCCTCTAAAAAATGTGGGAATCAAATCATTAAATGAACAAGGTTCATGAGTTATCTGT
AGGACAACTGGTGAGATACTCATCAAAAGTCAAGCTCTTACCAATGGGTATGACAACATACCCGATCTTAACCA
ACAAGCATTTAAAGAAGGGACATTTTTTACAGGTGATTTAGGAAAAGTGTGATGAAGCAGGTGCTCTTTATATCAC
TGGTAGAAAAGAAAATATTGATTGATACAGGGGGTGCAGGTTGATCCGATAGAGATAGAGGACATTCTGAATAC
ACACCCACAAGTCAAAGAAGCTGTTGTTGTTGGTGTAAAGGTGCTCATGCTGGTGAAGTCTGCAAGCTGTGA
TTGTTCTTAAGGAACAGAACAGTGTGATGAACAAAAATTTCTCGTACTGTAAAGAAGTTCATGAGAAATCA
AAGTTCCAAAGATTATAGAATCCGTAACGAAATTCCTAAGAGTCTTTGGGAAAATCTTGCGAAAAGCTTTGG
TTGTGTGATTGAGAATTTAAATCAGAAATCTCATGATTTTACATATTTTTTGGTTCGAATTTACATAAAGCTAGCTAT
TAAGAAAAAAGTACTGTTAAAGTCAAAGAATTCAGTATTGCTAGATCAAATCTGACAACTATTCCACACTAACTC
TACTGAGTGTATGGACATTAACAAGATTCTCGATACTAACATTAATCCTTAAGCGTTTTAATGGCTCCAATGCAG
ATGAAGGAGCAACTGCCATAACTCCAGTTGCTACAGAACTGTATTACGAGGTAGACAAGCGGTCAAAGAGAT
TCTAGATGGAAAAGATTCTAGAAAATCATTATTGTTGGTCTTTGTTCTATCCATGATGTGAAAGCTACTTTGGAAT
ATGCAGAAAAACTGAAGACTCTTGACAGACAAAGTCCAGGATAAACTGCTTATTCTGATGAGAGTTTACTTTGAAA
AACCTAGAACAACAATAGGATGGAAAGGGTTGATCAATGACCCAGACTTAGATGATTCTTTAATATTGAGAAAAG
GATTATTGACAGCACGTAATCTGCTGATCAATATTGCTGAATTAGGATTACCATCTGCTACAGAAGCTCTAGATCCT
GTAACACCTCAATATATCTCTGATTTGATTTGTTGGGCTGCAATTGGAGCTCGCACTATTGAATCTCAAACCCATCG
TGAGATGGCAAGCGGGTTATCTATGCCAGTAGGCTTTAAAAATGGCACTGACGGTAATATTCAAGTTGCTCTAGAT
GCCATTCAATCATCTAGAAAATCCCATCATTTTTTAGGGATTGACCAAATAGGTCAAATCAGTATCTTTCAAACAAA
GGGAAATGTCTACGGTCAATCATTTTTACGTGGTGGTGGTGGTCAACCTAATTTGATGCAGCTACAGTGGCTTGG
GTAGAAAAGAAGTTAGAAAATCTCAAGTTGCCTAAGAGAATTGTCATTGACTGTAGCCATGGTAATTCCTATAAAA
ATCATCAGCTACAGACTGCTGTCTTAAACAATGTTCTCCAACAAATTACAGATGGCAATCAGTCCATGATTGGCATG
ATGCTTGAGTCAATCTGTACGAAGGAAATCAGAAGATTCTAGCGATCTGAACCAAGTAAAGTACGGTGTATCG
GTGACAGACAAATGTATCGGTTGGGAAGAAACAGAAAGAAATATTCTATCTGCCATGAAAGGTTAAGTGCAGAT
AGAAATGTTATGTTACATACTTGTTGAATGGTTGATCTGGAACGCTGTCCGCACTTAATGGCAACGAAAGGAT
AGATTTTTCTCGCATTAAATGAATTAGGGCATGAAGCTTTTTAATATGAGCCGAAATCTTTGTCCAGACGTCACATTT
CTCCGTCTTACACTCAAAAAATTGTGAATTCCTACTTCATCAGCAATGCCGTTTTTCTAGTTGCATAATTAGAATCA
ATAGACCACTTTGAAGAGTAGCAATTTAGTTGTTTTGTTGGGAAAATATGTTGTTGAAGAACTGAAGCAA
AAAGCTAAAAATATTGTTTCAGAGTTATTGCCGATATTGGTAATGAAGAGATCACAGATGATACTGACATCTTTAG
TCTTGGGTTAGACTCAATCAATGCAATGACGCTCGTTGCAAATTTGCAAGATACATTTGATATTGATTTAGAAAACA
AACGAGTTTAACTTTGAAAATTTCAAACATCGAAATATCATAAAATGATTAGGAGAAAGAAAGGTTTATAAA
TTAATCAAATGTCAATTTAGATAAATGATATCTAGTATGCTACTAGAAAACGTTGCGCGATCGCAATCAAACATTC
ATCCAGCTTCAGGATTGCAACCACAAAACAGACTAAATCATCAGAAGAAGCTCAGAAAAGACCTTTGGGAGGCA
ATTAATACTGTCATTAGTTTGCAGAATCAAGCTCCGCCCTTAGTGTCAGTTTACGTCAAGGTAATATCCCTCTGTC
TTTTTCTCAAGAACGGTTGTGGTTTTTGGAGCAGTTAGAACCAGCAGATCCTCTGCTTACAATATGCCTTCGGCT
TTTCGGATTACAGGGGCATTAATGTGTCTGTGCTACAGCAGAGCCTCAACGAAATTCACGCCCATGAAGCC
TTGCGAACTACTTTGCAATTTAGAGAAGGCAAATCAGTCCAGGTTATTCACCCTGCACTGACCTTGAATTTACCAA
TCATTGAGCTACAAAATATCTCTTACAGCAACAACATATTAACAATGCAAGTTGATCAGGGAGGAAGTTCAAC
GACCCTTGATTTGAGCCAACTGCCACTCTACGAGCTACGTTGTTGCGTCTCAGTGAAAATGAACATTTACTGCT
CCTGAGTGTTCATCATATTGTGATTGATTTTTGGTCTAAAGGGATTTGTTTCAGGAATTGTCAGTACTTTATGAAG
CCTTCTCGACAGGCAAACCTTCTCCTTTTCTGAACTACCCATACAATATGCTGATTTGTCAGTCTGGCAGCGCCA
ATGGCTCAAGGGAGAGTTTTTAGAGGTGCTCTAAACTACTGGAAGCAGCAGCTTGATAGCAATCTTAGTGAGC
TACATCTACCTACAGACCGTGCCCGATCTATGTTACAAACTCGTGATGGTGCTAACCAAAAGTTAGTGCTATCAA

GGAATTAACAAAGGAACTCAAAGCGTTGAGTCGCCAAGAGGGTACGACTCTCTTTGTAGTACTACTGGCTGCTTT
TAAGGTGTTGTTGATCGCTACACCGAGCAGGATGACATCTTTGTATGTAGTCCCATCGCCAACCGCAATCGCAA
GAAGTTAAAGGACTCATTGGTTACTTTGTCAACTTGCTTATCTTACGTAAGCCTCTCAGACAACCCTAGCTTTG
GGAATTGCTGGGTCCGGTACGCAAGGTCCTTACAGGAGGCTATGCTTACCAAGACCTGCCAGTACAGCAGTTGG
TCAAATCTCTCAATTTGCTTCAAACCTCTTTGTCTCGGGTCATGTTGGTCTGCAAAAATACAGCCATACACAGCTT
AATTTACCTGGTTAACGGTCAGGAGTGTGGATATTGAGGGTGGAAACAGCCGATTTTGATTTGTATCTTTACGTGC
TCGAAGAAGGTAGTACTGACTGCAACCTTAAAGTATAATACCGATTTATTGACGATTGACAATCGTCCAAC
GCTGAACCACTCCAGACTGTTCTGGAGAACATTGCTGTCGACTCAGGGCAGTCTATCCCCTATTGCTGCCTTA
AGCACAGCCGAACAACAGCAGTTAACAGATAAAAGGCTAGAACAATCAAGCTTGAAGCCAGAAGGAGTTTATG
TAGCGCCTCGGAATCCTTTGGAACCTCAGCTCACACAGATTTGGTCTCAGGTTTTGGGTATTAGTCTGTCCGGT
TGAAGGATAACTTTTTGAACTCGGGGGAGAGTCGCTGTTAGCAATGTCTCTGTTTGTAAAATTGAGAAGATAT
TTGGTAAAACCTCTCCCCTAACGACTTTACTTCAAGCACCGACGGTGGAGCAATTCGCTCAGCTTCTCACTCAGG
ACGCAAATTCGGTGTCTTGGTCTTATTAGTGCCATTCAACCCAGTGGGACTAAACCACCTTTATTTGCATACAT
GGTCAACAGGGCAACGTTCTCAATTTCCGAAATTGTCTCAATATCTGGGTTGAGATCAGCCTTTCTATGGATTGC
AAGCTAAAGGGCTTGACGGAAAAGAGCTTCCACTTTCCGTATAGAAGATATCGCAACACACTACATTCAGGAAA
TACGTACTTCAACCAGAGGGACCTTACTTTTAGCAGGTAACCAATGGGGGTACAATAGCCTTCGAGATGG
CTCAGCAATTACATAAACAGGGTCAAAAAGTGGCACTGTTGGTTATGTTTGACACTTTTGGTCTAGATTGTTCCC
ACGGCTATCGTTGAGACGACAGCATTACTGGGCGTACCTTTACAACCTGGCATATCTAAATCCTTCTTAATGAAG
TGAATGAGCTTTGCCAAAGGAGGTTAAAGGAAATGATCAGCAGACTTTATTTGAGTCTGGGTCGTCCTTACCCC
AAAATCTCCGTGACGAATTGGTTGCCGAGGCTAATATGCAAGCCAAAATAGGGTATCAAGCGCAAGTCTATCCAG
GTCGTGTAACCCTATTGAGAGCTAGCCAACCAGCTTTATTTCCAAAGTTGATTTGCCTACATCTGAAGATTGGTAC
AACCGAAATCCTGAGCACGGCTGGAGTGAGGTCGTAGGCGGAGGCTTAGAAATCCACGATGTGCCGGGGCATC
ATTTCTCCATCTTTGAAGAACCTCATGTGCAAGTCTTAGCCGAAAAATTAAGGCTTGTGTTAGATGAGGCACAAAC
TAAATATTAACAGAATACAGCAGGATTGGTCACAATTTAGACATCTAAAACCTTTTTCTCCAGTTGCAAATCATC
ATTTCTCCGTGAAAGCGTTAAGAAGAAGGTTCAATATGAAGAATATTTATGACGTTGCTATTTGCGGTTCTGGTTT
AGCTGGGTTAACTCTGGCAGCAGACTAAAGTTAAAATGCCTGATATCTCAGTTGTCGTTCTTGATAGGCTAGCC
CGTCTTTACCAGAAGCTGGCTTCAAAGTGGGAGAATCATCTGTTGAAGTGGGTGCTTTTTATCTCGCTCATATTG
TGCAGTTAGAAGATTATCTGAAAAGCAACACCTTCATAAACTTGGGCTTCGTTATTTCTGGGTGACACAAAAG
GTCCTTTACAAAAGACCTGAAATTGGGCTTTCCAAATATCATTTCCTAACTCTTATCAATTGGATAGAGGTAAG
TTAGAGAATGATTTGCGCTCCATTAATACTGAAGCGGGTGTAGAGTTGCTAGAGGGTGTGTTAGTTAAGGATATTG
AGCTCGGTGACCCACAACAGCTGCACCAAATATCTACTCAGGAAAATAATAAAGCAACTCAAGCCATTCAAG
CTCGTTGGGTGGTTGATTCTATGGGTTATCGTCGTTTTCTGCAAAGAAAACCTGGTTTAGCTAAACCAAGA
CCCAATTTAGTGCTGTGTGGTTTCGGGTAGAAGTCTGTTTTGACGTGAGCGATTTGTTCTTAGCACAGAAATAG
AGTGGCATGAGCGAGTTCTCACAATAATCGCTATTACTCTACAAATCATTATGTGGCGAGGGCTATTGGGTCTG
GCTCATTCTTTATCTACTGGATATACAAGTATTGGGATTGTGACCAATGAAGAGATTATCCTTTTGGGACATATC
ATACCTATGAAAAAGCATTCAGTGGTTAGAAAAACATGAGCCTGTAGTGGCATTCCACTTAAAAAGCAACCCAC
CAGTTGACTTTATGAAAATACCTCAGTACAGTTACTCATCTAATCAGGTATTTTCGATCAATCGTTGGGCTTGTGTA
GGAGTAGCTGGTGTATTTGCCGATCCTTTCTATTACCCGGTACGGATTTGATTGGCTTCGAAAACCTTAAATCA
CTCAGATGGTGAACTTGATCGTGAAAACAAGCTGACTCCGAAATAGTCAATGAGGCTAATCGCTTTCTGATCA
CATAACGAGAGTGAACGTCAAATATTCACAACGCTTATCTCTGTTTCGGCAATGAAACGGTTATGGTGATGAA
ATACATTTGGGATGTCTTATCTGCATGGCATTAGCGCCCAATGATGTTAATCCCTATTTCTGATTAGACAA
GAGAGCCAAAGTTGCAAAGGGACTGGACAATTTTCTATTGGCACAACGGATGAATCAATTATTTAGAGATTG
GGCAGTTAGTGCAGCGACGGACATCGTTTGTGTTTTATCGATTATTTGCAAATTCCTTTTGTAAAGAAATTGCGT
GCTCGTAATTTAAAACCAACAAAACAGAGCAGGAACTCATTGATGATCATCTAGCTAGCATAAACTTTTTGAA
GAATTAGCTCAGGTGATTTCTTACTCGCATTAGAAGACACAATGCCAGAGAAATCAGCCGATTTCCATCACCAG
TATGGTTAAATGCTTGGGTTGTAGTTTAGATGACAAGCGATGGGAGATTGATGGACTGTTTCGCCAACTTCAA
AACCTCGGATTTACGTCCAATGATGGAACAACCTTTGGCAGAATATCACTTCCGTGCAGCCGATAGAGATTGGA
GTTTAATTAAGTCTTGA

Ab2 codon optimised sequence for expression in *E. coli*.

ATGCTGTACCAGGAGTTGCATCTACTGCATTCAACAAGATCGCTCCTCCAGGTCCTCCTGACTGCCTTTTCGTTG
ATATGCTGCCATGTCTGGGTAAGCACCTGCACCTGGCACTGAATCAACTGGCAAAGAAGTACGGTAACATCTTCC
AGATCCGTGTGGGTGCAAAGACCCTGGTTGACTGAACGGTCTGGAACTATCAAGGAGGCTCTGGTAAAAACAG
CCGGACTCTTTCAACGCTCGTGCAGACTTCGACATCTACCAGCAGCCACCACAGGCTCAATTCCTGGAAGTAAA
TCCGGTGAGTCTTGGCGTAAACACCACAACATCCTGGGTCAGGCTATGCACACTTTCGTGGTAGGTAAACCGGAC
ATGCTGGAGTCTTGGGCACTGGAAGAAGCAGCTGATCTGGCTAACATCTTCTCAAATTCAGCGTCCAGCCGTT
GACCCAGACCTGTACATGCCGCTGGCTACTCTGTCTTTCATGCAGCGTCTGATTTTCGACAAACGTGGTGTATCA
AAAACCCGGAGGAGGACCACGAGTTCGTGCGTCTGCGTATACTCTGAAACACATCCCAGTGTACTGGAGGCC
GTTCTGTGGAGTATATCCGAAAATCTGGCAGCCGATTTCCGTCTGTCTCGTTGGAAATCTCTGCGTAACTTCT
GAAAAGCCTGGTGGCGCTGGAATCTTACGTGTCTAAAAACGTGGCGCAGCACCAGGAATCCTTCGACCCGGAA
AACCTGCGTGATATTACCGACGCGTGTGAAAGCCAGCAGCGAAGTACTGAAATCCGATCGTAATAACCTGCAT
CTGAGCGAAAACGACATCGTGAACGGTCCCTGATGCAGCTGGCCGGTGCAGGCTGGCGGCTGGCGAGCTTTA
TGCTGCGTTGGGGCGTTCTGTATATGATGACCTATCCGGCGATTCAAGCCGAAATTCATAAAGAAGTGGATGAAG
TTGTTGGCCGTCACAGCAGCCGTGTCTGGAACATCGTGGCAAACGCGTTTACCGAAGCGTGCATTCATGAA
ATTTTTCTGCACTCCAGCATTACCACCATGCCGCCGATTACCTATGCGACCACGACGGATGTTACCCTGGAAGATT
ATTTTATCCCGCAGAATACGCCGCTGCTGATCAATTACTACTCCCTGACCCGCGATCAGCGCTACTGGGAAGAACC
GGAACAGTTTAAACCCGTACCGCTTCTGGATGAAAACGGCAAACGCGCAAAAACCTGCTGGATAAATTCTACCC
GTTCCGCATGGGCTCCCGCCGCTGCATCGGCCAATACCTGGGCCGCTGTGTTTTTACCTTTTTTACCAACCT
GATGCACAAATGCAAATTCGAAAAGTGCCGGGCGAAAAACTGTCCTTCAATCCATCCCAGGCGCGTTCATCAT
CCCGGAAAATACCGCGTTGTCGTCAAACCGAGCTTCTAA

Ab3 codon optimised sequence for expression in *E. coli*.

ATGGTCTCTCAGCAGTCTGCTTCTACCGAACTGAACACCCTGAACAAGATCAGCCCTCCAGGTCCTCCTGCACTG
CCTTTTGTCCGATGCTGCCATTCTGGGTAAGCATCTGCATCTGGCACTGAACCAGCTGGCAAAGAAGTACGGT
AACATCTACCAGATCCGTGTGGGCGCACGTAATCTGGTCTGACTGAACGGTCTGGAACTATCAAGGAGGCTCTG
GTAAAGCAGCCAGACTCTTTCAACGCTCGTGCAGACTTCGACCTGTACCAGCAGCCACCACAGTGTCAATTCATG
GAGCAGAAATCCGGTGTGCTGGCGTAAACACCGTACCATCATCGGTGAGTGCACACCTTCGTGGCATCT
CGTTCTGGTACTCTGGAAAGCTGGGCACTGGAAGAAGCAGCTGACCTGGCTAATATCTTCGTTAACTCCAGCGG
CCAAGCGTTCGACCCGAATCTGTACCTGCCGCTGGTACTCTGTCTTTTATCCAACGCTGTGATCTTCGGCAAACGT
GGTAACCTGGACGACCCGAAAAAGACACTGACTTCTGTGACCACTGCTCATAGCATGGGTAAACTGAACAACGG
CGCCAAAACCTGACCAAACCTGGTGTCTGCTGCCGACTATCTGGCGTCCGATTCTGATGATCAGCATCTGGAAATC
CCTGCGTGGTTTTCTGTAAGCGGCTGACGCTGCCGAAGGTTATCTGATCAAAAACGTTGAGCAGCACCAGAAAT
CCTTCGACCCGGAGAACCTGCGTGATATTACCGATGCGCTGCTGAAAGCCAGCAGCGAAGTACTGAAATCCGATC
GTAACAACCTGGGTCTGAGCGAAAACGATATTGTTAACGGTCCCTGACGAGTTCGTTGGCGCCGGTACCGAA
CTGCCGCTCTGATGCTGCGTTGGGCCCTGCTGTATATGATTACTTACCCGGCGATTACAGGCGGAAATTCACAAAG
AACTGGATGAGGTAGTTGGCCGTCAGCAACAACCGTGTCTGAAACACCGTGGTAAACTGCCGTTTACCGAAGCG
TGCATTCACGAAGTTTTCCGTCACCTTCTGCGACCACGACGCCGGCGTTTATTACGCGACGACCACCGATGTTA
CCCTGGATGGTTATTTTATCCCGCAGAATACCCGCTGCTGTTGATTACTCCCTGACCCGCGATGAACGCTAT
TGGGAAAACCGGAACAGTTCAACCCGTACCGTCTCTGGATGAAAACGGCAAACGCGCAAAAACCTGCTGG
ATAAATTCACCCGTTCCGCATCGCTCCCGCCGCTGCATTGGCGAATACATTGGCCGCTGCTGATCTTTACCTT
TTTTACCCACCTGATGCACAAATGCAAATTCGAGAAAGTACCGGGCGAAAAACTGAGCCTGGACCCGAGCCGG
CGATCATCTGCCGCCGAGAATTACAAAGTTATCGCGAAACCGCGCTTTTAA

3. Heterologous Expression of the *ab* BGC in *E. coli* BAP1

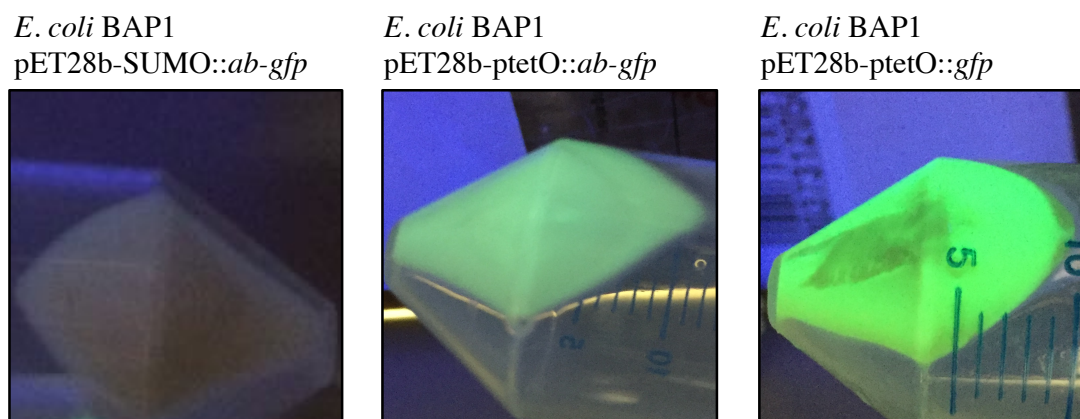


Fig. S5 Cell pellets harvested 24 h after induction of expression. No fluorescence is detected for *E. coli* BAP1 pET28b-SUMO::*ab-gfp*, therefore indicating incomplete gene transcription. In contrast, *E. coli* BAP1 pET28b-ptetO::*ab-gfp* and *E. coli* BAP1 pET28b-ptetO::*gfp* (positive control) show bright green fluorescence indicating full gene transcription.

4. SDS PAGE Gels of Purified Ab2, Ab3, Fpr and PetF used for *in vitro* Assays

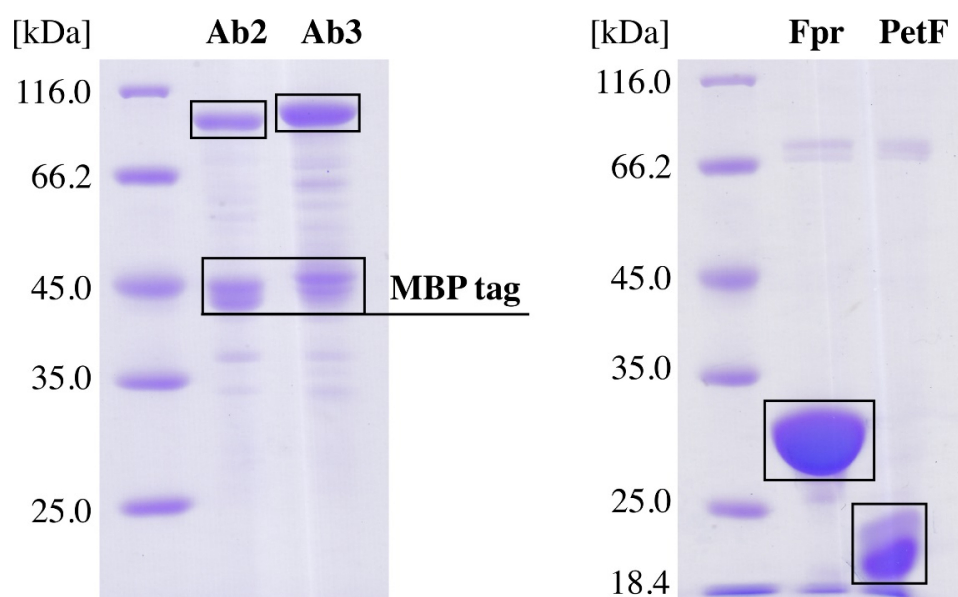
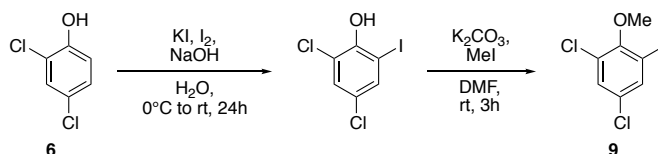


Figure S6 SDS-PAGE analysis of the purified enzymes Ab2 (97.5 kDa), Ab3 (101.1 kDa), Fpr (30.2 kDa) and PetF (12.9 kDa, occurs as dimer) used for *in vitro* assays.

5. Experimental Procedures

5.1 Synthesis of Methoxy Protected Aryl Iodides (9–11)

1,5-Dichloro-3-iodo-2-methoxybenzene (9)



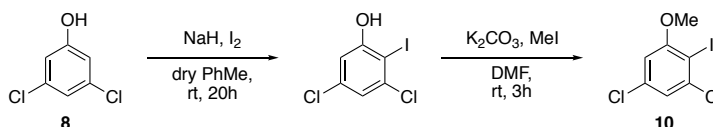
2,4-Dichloro-6-iodophenol was prepared according to a published procedure. Analytical data were in agreement with literature values.¹⁰

$^1\text{H-NMR}$ (400 MHz, CDCl_3): δ = 7.61 (d, J = 2.4 Hz, 1 H), 7.34 (d, J = 2.4 Hz, 1 H), 5.90 (bs, 1 H). **$^{13}\text{C-NMR}$** (100 MHz, CDCl_3): δ = 150.1, 137.1, 129.4, 126.5, 119.3, 83.4. **MS** (ESI⁻): m/z = 308.9 [$\text{M}+\text{Na}-2\text{H}$]⁻, 324.9 [$\text{M}+\text{K}-2\text{H}$]⁻.

1,5-Dichloro-3-iodo-2-methoxybenzene (9) was prepared according to a published procedure. Analytical data were in agreement with literature values.¹⁰

$^1\text{H-NMR}$ (300 MHz, CDCl_3): δ = 7.66 (d, J = 2.5 Hz, 1 H), 7.37 (d, J = 2.4 Hz, 1 H), 3.85 (s, 3 H). **$^{13}\text{C-NMR}$** (75 MHz, CDCl_3): δ = 154.9, 137.3, 130.7 (two signals), 128.1, 92.7, 60.9. **MS** (ESI⁻): m/z = 322.9 [$\text{M}+\text{Na}-2\text{H}$]⁻.

1,5-Dichloro-2-iodo-3-methoxybenzene (10)



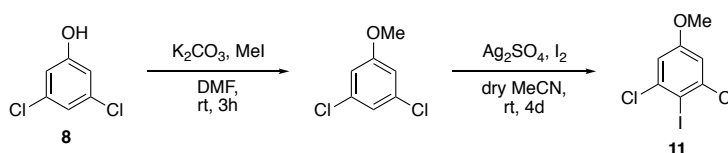
3,5-Dichloro-2-iodophenol was prepared according to a published procedure. Analytical data were in agreement with literature values.¹¹

$^1\text{H-NMR}$ (400 MHz, CDCl_3): δ = 7.08 (d, J = 2.3 Hz, 1 H), 6.91 (d, J = 2.3 Hz, 1 H), 5.62 (bs, 1 H). **$^{13}\text{C-NMR}$** (100 MHz, CDCl_3): δ = 157.0, 139.2, 136.1, 121.7, 113.5, 89.1. **MS** (ESI⁻): m/z = 309.0 [$\text{M}+\text{Na}-2\text{H}$]⁻, 324.9 [$\text{M}+\text{K}-2\text{H}$]⁻.

3,5-Dichloro-2-iodophenol (7.41 g, 25.6 mmol, 1.0 eq.) was dissolved in DMF (26.0 mL, 1.00 mL/mmol) and potassium carbonate (4.01 g, 30.8 mmol, 1.2 eq.) was added. After dropwise addition of methyl iodide (4.37 g, 1.92 mL, 30.8 mmol, 1.2 eq.), the reaction solution was stirred for 3 h at room temperature. After completion, the reaction was diluted with water and Et_2O . The aqueous layer was extracted with Et_2O (2 \times). The combined organic layers were washed with 1N HCl (6 \times), brine, dried over Na_2SO_4 and filtered. The solvent was removed to give 1,5-dichloro-2-iodo-3-methoxybenzene (10, 7.44 g, 24.6 mmol, 96%) as a white solid. Analytical data were in agreement with literature values.¹²

$^1\text{H-NMR}$ (400 MHz, CDCl_3): δ = 7.13 (d, J = 2.1 Hz, 1 H), 6.68 (d, J = 2.1 Hz, 1 H), 3.89 (s, 3 H). **$^{13}\text{C-NMR}$** (100 MHz, CDCl_3): δ = 160.3, 140.4, 135.7, 121.7, 109.6, 89.2, 57.2. **MS** (ESI⁺): m/z = 303.0 [$\text{M}+\text{H}$]⁺.

1,3-Dichloro-2-iodo-5-methoxybenzene (**11**)



3,5-Dichlorophenol (**8**, 4.00 g, 24.5 mmol, eq.) was dissolved in DMF (24.5 mL, 1.00 mL/mmol) and potassium carbonate (3.83 g, 29.4 mmol, 1.2 eq.) was added. After dropwise addition of methyl iodide (4.17 g, 1.83 mL, 29.4 mmol, 1.2 eq.), the reaction solution was stirred for 3 h at room temperature. After completion, the reaction was diluted with water and Et₂O. The aqueous layer was extracted with Et₂O (2×). The combined organic layers were washed with 1N HCl (6×), brine, dried over Na₂SO₄ and filtered. The solvent was removed to give 1,3-dichloro-5-methoxybenzene (**11**, 4.14 g, 23.4 mmol, 95%) as a white solid. Analytical data were in agreement with literature values.¹³

¹H-NMR (300 MHz, CDCl₃): δ = 6.95 (t, J = 1.8 Hz, 1 H), 6.79 (d, J = 1.8 Hz, 2 H), 3.79 (s, 3 H). ¹³C-NMR (75 MHz, CDCl₃): δ = 160.8, 135.5, 121.1, 113.3, 55.8. MS (ESI⁺): m/z = 176.8 [M+H]⁺.

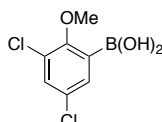
1,3-Dichloro-2-iodo-5-methoxybenzene was prepared according to a published procedure. Analytical data were in agreement with literature values.¹⁴

¹H-NMR (400 MHz, CDCl₃): δ = 6.95 (s, 2 H), 3.79 (s, 3 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 160.3, 140.8, 114.0, 92.3, 56.0. MS (ESI⁺): m/z = 271.0 [M-MeOH+H]⁺. MS (ESI⁻): m/z = 269.0 [M-MeOH-H]⁻, 322.8 [M+Na-2H]⁻.

5.2 General Procedure for the Synthesis of Arylboronic Acids (**12–14**)

The reaction was performed under dry conditions. Methoxy protected aryl iodide (1.0 eq.) was dissolved in dry diethyl ether (5.00 mL/mmol) and cooled down to -15 °C. 2M ¹PrMgCl solution in Et₂O (1.5 eq.) was added dropwise to the reaction solution, which was stirred at -15 °C for 1 h. Freshly distilled trimethyl borate (2.0 eq.) was added rapidly. The reaction was stirred at -10 °C to 0 °C for 2 h before being treated with 4M HCl solution (1.00 mL/mmol). After another 30 min at room temperature, the reaction solution was extracted with Et₂O (3x). The combined organic layers were washed with water, brine, dried over Na₂SO₄ and filtered. The solvent was removed, and the obtained residue was washed with hexane to give the corresponding arylboronic acid as a white solid.

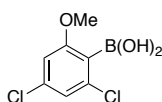
(3,5-Dichloro-2-methoxyphenyl)boronic acid (**12**)



Synthesised from 1,5-dichloro-3-iodo-2-methoxybenzene (**9**, 5.00 g, 16.5 mmol) in 92% yield (3.36 g, 15.2 mmol) as a white solid.

¹H-NMR (300 MHz, DMSO-d₆): δ = 8.40 (bs, 2 H), 7.59 (d, J = 2.6 Hz, 1 H), 7.35 (d, J = 2.6 Hz, 1 H), 3.78 (s, 3 H). ¹³C-NMR (75 MHz, DMSO-d₆): δ = 156.9, 132.1, 129.9, 127.6, 126.7, 61.1. C-B signal not observed due to quadrupolar relaxation. MS (ESI⁻): m/z = 219.0 [M-H]⁻. HRMS (ESI⁻) m/z calcd for C₇H₆BCl₂O₃ [M-H]⁻: 218.9793, found: 218.9793.

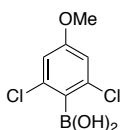
(2,4-Dichloro-6-methoxyphenyl)boronic acid (**13**)



Synthesised from 1,5-dichloro-2-iodo-3-methoxybenzene (**10**, 1.00 g, 3.30 mmol) in 74% yield (538 mg, 2.44 mmol) as a white solid.

$^1\text{H-NMR}$ (300 MHz, DMSO- d_6): δ = 8.30 (bs, 2 H), 7.06 (d, J = 1.6 Hz, 1 H), 6.98 (d, J = 1.6 Hz, 1 H), 3.76 (s, 3 H). $^{13}\text{C-NMR}$ (75 MHz, DMSO- d_6): δ = 162.2, 135.9, 133.8, 119.9, 109.4, 56.0. C–B signal not observed due to quadrupolar relaxation. **MS** (ESI $^-$): m/z = 218.9 [M–H] $^-$. **HRMS** (ESI $^-$) m/z calcd for $\text{C}_7\text{H}_6\text{Cl}_2\text{O}_3$ [M–H] $^-$: 218.9793, found: 218.9794.

(2,6-Dichloro-4-methoxyphenyl)boronic acid (**14**)



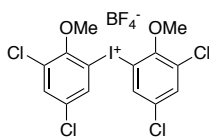
Synthesised from 1,3-dichloro-2-iodo-5-methoxybenzene (**11**, 3.00 g, 9.90 mmol) in 38% yield (843 mg, 3.82 mmol) as a white solid.

$^1\text{H-NMR}$ (300 MHz, DMSO- d_6): δ = 8.20 (bs, 2 H), 6.96 (s, 2 H), 3.78 (s, 3 H). $^{13}\text{C-NMR}$ (75 MHz, DMSO- d_6): δ = 160.2, 136.4, 112.9, 56.0. C–B signal not observed due to quadrupolar relaxation. **MS** (ESI $^-$): m/z = 218.9 [M–H] $^-$. **HRMS** (ESI $^-$) m/z calcd for $\text{C}_7\text{H}_6\text{Cl}_2\text{O}_3$ [M–H] $^-$: 218.9793, found: 218.9794.

5.3 General Procedure for the Synthesis of Diaryliodonium Tetrafluoroborates (**15–17**)

The symmetrical iodonium salts were prepared according to a published procedure.¹⁵ *meta*-Chloroperbenzoic acid (77% active oxidant, 1.1 eq.) was added to a Schlenk tube and dissolved in dry DCM (2.00 mL/mmol). The iodated aryl species (1.0 eq.) was added and the reaction mixture was heated up to 80 °C. After 10 min of stirring at 80 °C, the mixture was cooled down to –78 °C. A 0 °C cold solution of $\text{BF}_3 \cdot \text{OEt}_2$ (48% BF_3 , 2.2 eq.) and arylboronic acid (1.0 eq.) in dry DCM (2.00 mL/mmol) was transferred into the reaction mixture. The obtained yellow solution was stirred for 30 min at –78 °C and warmed to room temperature. The crude reaction solution was purified by flash column chromatography (DCM to DCM/MeOH = 20:1). The product fraction was concentrated and to the obtained residue was added diethyl ether to induce precipitation of the iodonium salt. The solid was washed with diethyl ether and dried under reduced pressure to give the corresponding iodonium salt as a white to beige solid.

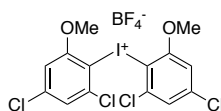
Bis(3,5-dichloro-2-methoxyphenyl)iodonium tetrafluoroborate (**15**)



Synthesised from 1,5-dichloro-3-iodo-2-methoxybenzene (**9**, 908 mg, 3.00 mmol, 1.0 eq.) and (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**, 663 mg, 3.00 mmol, 1.0 eq.) in 52% yield (876 mg, 1.55 mmol) as a white solid.

¹H-NMR (300 MHz, DMSO-*d*₆): δ = 8.55 (d, *J* = 2.4 Hz, 2 H), 8.05 (d, *J* = 2.4 Hz, 2 H), 4.00 (s, 6 H). **¹³C-NMR** (75 MHz, DMSO-*d*₆): δ = 153.3, 135.2, 135.1, 130.4, 127.4, 115.2, 62.5. **MS** (ESI+): *m/z* = 476.8 [M-BF₄]⁺. **HRMS** (ESI+) *m/z* calcd for C₁₄H₁₀Cl₄IO₂ [M-BF₄]⁺: 476.8474, found: 476.8476.

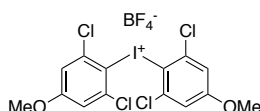
Bis(2,4-dichloro-6-methoxyphenyl)iodonium tetrafluoroborate (**16**)



Synthesised from 1,5-dichloro-2-iodo-3-methoxybenzene (**10**, 454 mg, 1.50 mmol, 1.0 eq.) and (2,4-dichloro-6-methoxyphenyl)boronic acid (**13**, 332 mg, 1.50 mmol, 1.0 eq.) in 24% yield (200 mg, 0.35 mmol) as a white solid.

¹H-NMR (400 MHz, DMSO-*d*₆): δ = 7.61 (d, *J* = 2.0 Hz, 2 H), 7.37 (d, *J* = 2.0 Hz, 2 H), 3.92 (s, 6 H). **¹³C-NMR** (100 MHz, DMSO-*d*₆): δ = 159.3, 140.0, 138.8, 121.9, 112.2, 108.8, 64.9. **MS** (ESI+): *m/z* = 476.7 [M-BF₄]⁺. **HRMS** (ESI+) *m/z* calcd for C₁₄H₁₀Cl₄IO₂ [M-BF₄]⁺: 476.8474, found: 476.8475.

Bis(2,6-dichloro-4-methoxyphenyl)iodonium tetrafluoroborate (**17**)



Synthesised from 1,3-dichloro-2-iodo-5-methoxybenzene (**11**, 490 mg, 1.62 mmol, 1.0 eq.) and (2,6-dichloro-4-methoxyphenyl)boronic acid (**14**, 357 mg, 1.62 mmol, 1.0 eq.) in 16% yield (143 mg, 0.25 mmol) as a white solid.

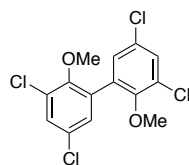
¹H-NMR (400 MHz, DMSO-*d*₆): δ = 7.44 (s, 4 H), 3.88 (s, 6 H). **¹³C-NMR** (100 MHz, DMSO-*d*₆): δ = 163.4, 140.0, 115.4, 56.9. **MS** (ESI+): *m/z* = 476.7 [M-BF₄]⁺. **HRMS** (ESI+) *m/z* calcd for C₁₄H₁₀Cl₄IO₂ [M-BF₄]⁺: 476.8474, found: 476.8474.

5.4 General Procedure for the Suzuki Coupling (**18–20**)

The reaction was performed under dry conditions. Pd₂(dba)₃ (5 mol%) and DavePhos (15 mol%) were added in a Schlenk tube equipped with a stir bar. The Schlenk tube was evacuated and then refilled with argon. Dry 1,4-dioxane (5.00 mL/mmol) was added and three more evacuation-refill cycles (degassing) were performed. The catalyst mixture was heated up to 90 °C using a preheated oil bath. After 20 min, aryl iodide (1.0 eq.), boronic acid (2.0 eq.), potassium carbonate (2.0 eq.) and silver(I)

oxide (0.5 eq.) were added and another three evacuation-refill cycles were performed. The reaction mixture was stirred at 90 °C for 24 h before being diluted with EtOAc at room temperature. The reaction solution was filtered, evaporated and the obtained residue was purified by MPLC to give the corresponding methoxy protected C–C dimers as yellow oils.

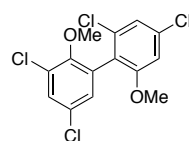
3,3',5,5'-Tetrachloro-2,2'-dimethoxy-1,1'-biphenyl (**18**)



Synthesised from 1,5-dichloro-3-iodo-2-methoxybenzene (**9**, 300 mg, 0.99 mmol, 1.0 eq.) and (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**, 437 g, 1.98 mmol, 2.0 eq.) in 37% yield (132 mg, 0.37 mmol) as a yellow oil. The substrate 1,5-dichloro-3-iodo-2-methoxybenzene (84.0 mg, 0.28 mmol) was re-isolated in 28% yield, based on which the desired product was obtained in 52% isolated yield.

¹H-NMR (300 MHz, CDCl₃): δ = 7.44 (d, *J* = 2.6 Hz, 2 H), 7.19 (d, *J* = 2.6 Hz, 2 H), 3.61 (s, 6 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 152.7, 133.1, 130.4, 129.5, 129.2 (two signals), 61.2. MS (ESI⁻): *m/z* = 385.1 [M+Cl]⁻.

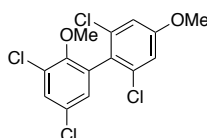
2',3,4',5-Tetrachloro-2,6'-dimethoxy-1,1'-biphenyl (**19**)



Synthesised from 1,5-dichloro-2-iodo-3-methoxybenzene (**10**, 300 mg, 0.99 mmol, 1.0 eq.) and (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**, 437 mg, 1.98 mmol, 2.0 eq.) in 45% yield (158 mg, 0.45 mmol) as a yellow oil. The substrate 1,5-dichloro-2-iodo-3-methoxybenzene (**10**, 41.0 mg, 0.14 mmol) was re-isolated in 14% yield, based on which the desired product was obtained in 53% isolated yield.

¹H-NMR (300 MHz, CDCl₃): δ = 7.42 (d, *J* = 2.6 Hz, 1 H), 7.13 (d, *J* = 1.9 Hz, 1 H), 7.01 (d, *J* = 2.6 Hz, 1 H), 6.88 (d, *J* = 1.9 Hz, 1 H), 3.75 (s, 3 H), 3.57 (s, 3 H). ¹³C-NMR (75 MHz, CDCl₃): δ = 158.5, 153.2, 135.4, 135.2, 131.6, 130.2 (two signals), 129.1, 128.9, 123.8, 121.7, 110.4, 61.0, 56.5. MS (ESI⁻): *m/z* = 385.1 [M+Cl]⁻.

2',3,5,6'-Tetrachloro-2,4'-dimethoxy-1,1'-biphenyl (**20**)



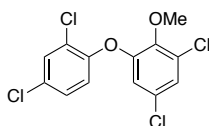
Synthesised from 1,3-dichloro-2-iodo-5-methoxybenzene (**11**, 300 mg, 0.99 mmol, 1.0 eq.) and (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**, 437 mg, 1.98 mmol, 2.0 eq.) in 54% yield (188 mg, 0.53 mmol) as a yellow oil.

¹H-NMR (400 MHz, CDCl₃): δ = 7.45 (d, *J* = 2.5 Hz, 1 H), 7.04 (d, *J* = 2.5 Hz, 1 H), 6.98 (s, 2 H), 3.85 (s, 3 H), 3.60 (s, 3 H). **¹³C-NMR** (75 MHz, CDCl₃): δ = 160.0, 153.1, 135.6, 133.5, 130.5, 130.2, 129.2, 129.1, 127.0, 114.2, 61.0, 56.0. **MS** (ESI⁻): *m/z* = 385.0 [M+Cl]⁻.

5.5 General Procedure for the Biaryl Ether Formation (21–26)

The diphenyl ethers were prepared according to a published procedure.¹⁶ KO^tBu (1.1 eq.) was dissolved in dry THF (4.5 mL/mmol) and cooled down to 0 °C. Dichlorophenol (1.0 eq.) was added and the reaction solution was stirred for 15 min at 0 °C. After addition of the diaryliodonium salt (1.2 eq.), the reaction mixture was stirred at 40 °C for 1 h. H₂O was slowly added at 0 °C and the organic phase was separated. The aqueous layer was extracted with Et₂O (2×). The combined organic layers were washed with brine, dried over Na₂SO₄ and filtered. The solvent was removed and the obtained residue was purified by MPLC to give the corresponding biaryl ether.

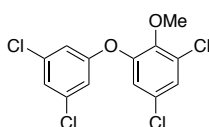
1,5-Dichloro-3-(2,4-dichlorophenoxy)-2-methoxybenzene (21)



Synthesised from 2,4-dichlorophenol (**6**, 35.9 g, 0.22 mmol, 1.0 eq.) and bis(3,5-dichloro-2-methoxyphenyl)iodonium tetrafluoroborate (**15**, 150 mg, 0.27 mmol, 1.2 eq.) in 97% yield (72.0 mg, 0.21 mmol) as a yellow oil.

¹H-NMR (400 MHz, CDCl₃): δ = 7.44 (d, *J* = 2.5 Hz, 1 H), 7.10 (d, *J* = 2.3 Hz, 1 H), 7.05 (dd, *J* = 8.8, 2.5 Hz, 1 H), 6.90 (d, *J* = 2.3 Hz, 1 H), 6.41 (d, *J* = 8.8 Hz, 1 H), 3.78 (s, 3 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 153.8, 151.9, 138.7, 131.8, 130.4, 129.6, 127.6 (two signals), 123.5, 122.1, 115.2, 112.3, 56.7. **MS** (ESI⁺): *m/z* = 337.0 [M+H]⁺.

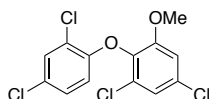
1,5-Dichloro-3-(3,5-dichlorophenoxy)-2-methoxybenzene (22)



Synthesised from 3,5-dichlorophenol (**8**, 96.0 mg, 0.59 mmol, 1.0 eq.) and bis(3,5-dichloro-2-methoxyphenyl)iodonium tetrafluoroborate (**15**, 400 mg, 0.71 mmol, 1.2 eq.) in 82% yield (163 mg, 0.48 mmol) as a yellow oil.

¹H-NMR (400 MHz, CDCl₃): δ = 7.27 (d, *J* = 2.5 Hz, 1 H), 7.11 (t, *J* = 1.8 Hz, 1 H), 6.94 (d, *J* = 2.5 Hz, 1 H), 6.84 (d, *J* = 1.8 Hz, 2 H), 3.82 (s, 3 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 158.1, 149.2, 147.5, 136.0, 130.2, 129.5, 126.9, 124.0, 121.0, 116.2, 61.4. **MS** (ESI⁺): *m/z* = 336.9 [M+H]⁺.

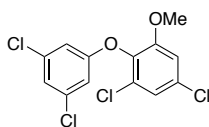
1,5-Dichloro-2-(2,4-dichlorophenoxy)-3-methoxybenzene (23)



Synthesised from 2,4-dichlorophenol (**6**, 96.0 g, 0.59 mmol, 1.0 eq.) and bis(2,4-dichloro-6-methoxyphenyl)iodonium tetrafluoroborate (**16**, 400 mg, 0.71 mmol, 1.2 eq.) in 48% yield (96.0 mg, 0.28 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.44 (d, *J* = 2.5 Hz, 1 H), 7.10 (d, *J* = 2.3 Hz, 1 H), 7.05 (dd, *J* = 8.8, 2.5 Hz, 1 H), 6.90 (d, *J* = 2.3 Hz, 1 H), 6.41 (d, *J* = 8.8 Hz, 1 H), 3.78 (s, 3 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 153.8, 151.9, 138.7, 131.8, 130.4, 129.6, 127.6 (two signals), 123.5, 122.1, 115.2, 112.3, 56.7. MS (ESI+): *m/z* = 336.9 [M+H]⁺.

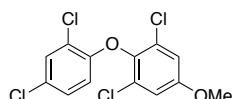
1,5-Dichloro-2-(3,5-dichlorophenoxy)-3-methoxybenzene (24)



Synthesised from 3,5-dichlorophenol (**8**, 96.2 g, 0.59 mmol, 1.0 eq.) and bis(2,4-dichloro-6-methoxyphenyl)iodonium tetrafluoroborate (**16**, 400 mg, 0.71 mmol, 1.2 eq.) in 29% yield (56.2 mg, 0.17 mmol) as a white solid.

¹H-NMR (300 MHz, CDCl₃): δ = 7.10 (d, *J* = 2.3 Hz, 1 H), 7.04 (t, *J* = 1.8 Hz, 1 H), 6.92 (d, *J* = 2.3 Hz, 1 H), 6.72 (d, *J* = 1.8 Hz, 2 H), 3.80 (s, 3 H). ¹³C-NMR (75 MHz, CDCl₃): δ = 158.4, 153.8, 138.0, 135.7, 132.0, 129.7, 123.0, 122.1, 114.2, 112.3, 56.7. MS (ESI+): *m/z* = 336.9 [M+H]⁺.

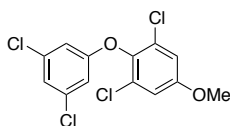
1,3-Dichloro-2-(2,4-dichlorophenoxy)-5-methoxybenzene (25)



Synthesised from 2,4-dichlorophenol (**6**, 48.4 mg, 0.30 mmol, 1.0 eq.) and bis(2,6-dichloro-4-methoxyphenyl)iodonium tetrafluoroborate (**17**, 185 mg, 0.33 mmol, 1.1 eq.) in 43% yield (43.0 mg, 0.13 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.46 (d, *J* = 2.5 Hz, 1 H), 7.07 (dd, *J* = 8.8, 2.5 Hz, 1 H), 6.95 (s, 2 H), 6.39 (d, *J* = 8.8 Hz, 1 H), 3.82 (s, 3 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 157.4, 151.6, 140.5, 130.6, 129.9, 127.8, 127.6, 123.5, 122.1, 115.0, 56.2. MS (ESI+): *m/z* = 305.1 [M-MeOH+H]⁺, 336.9 [M+H]⁺.

1,3-Dichloro-2-(3,5-dichlorophenoxy)-5-methoxybenzene (26)



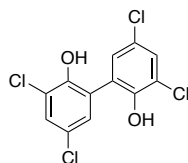
Synthesised from 2,4-dichlorophenol (**8**, 48.4 mg, 0.30 mmol, 1.0 eq.) and bis(2,6-dichloro-4-methoxyphenyl)iodonium tetrafluoroborate (**17**, 185 mg, 0.33 mmol, 1.1 eq.) in 16% yield (16.0 mg, 0.05 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.05 (t, *J* = 1.7 Hz, 1 H), 6.95 (s, 2 H), 6.72 (d, *J* = 1.7 Hz, 2 H), 3.83 (s, 3 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 158.2, 157.5, 139.9, 135.8, 129.9, 123.1, 115.1, 114.2, 56.2. MS (ESI+): *m/z* = 305.1 [M-MeOH+H]⁺, 337.1 [M+H]⁺, 375.1 [M+K]⁺, 381.1 [M+2Na-H]⁺.

5.6 General Procedure for the Methoxy Deprotection (27–34)

The reaction was performed under dry conditions. The corresponding methoxy protected dimer was dissolved in dry DCM (10.0 mL/mmol) and cooled down to -78 °C. 1M Boron tribromide solution (5.0 eq. for biaryl, 2.5 eq. for biaryl ether) was added dropwise and the reaction solution was stirred for 2 h at -78 °C and for 24 h at room temperature. The reaction was quenched with water at 0 °C and the organic phase was separated. The aqueous layer was extracted with DCM (2×). The combined organic layers were washed with brine, dried over Na₂SO₄ and filtered. The solvent was removed and the obtained residue was purified by MPLC to give the corresponding deprotected phenols.

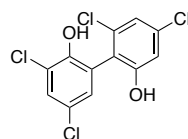
3,3',5,5'-Tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (27)



Synthesised from 3,3',5,5'-tetrachloro-2,2'-dimethoxy-1,1'-biphenyl (**18**, 125 mg, 0.36 mmol) in 29% yield (33.0 mg, 0.10 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.41 (d, *J* = 2.5 Hz, 2 H), 7.19 (d, *J* = 2.5 Hz, 2 H), 5.84 (bs, 2 H). ¹³C-NMR (100 MHz, CDCl₃): δ = 147.6, 130.1, 129.1, 125.9, 125.6, 121.9. ¹H-NMR (300 MHz, DMSO-*d*₆): δ = 9.60 (bs, 2 H), 7.53 (d, *J* = 2.6 Hz, 2 H), 7.17 (d, *J* = 2.6 Hz, 2 H). ¹³C-NMR (75 MHz, DMSO-*d*₆): δ = 149.9, 129.5, 128.6, 127.7, 122.7, 122.1. MS (ESI-): *m/z* = 320.8 [M-H]⁻. HRMS (ESI-) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8174. HPLC: *t*_R = 15.9 min (λ_{max} = 208, 296 nm).

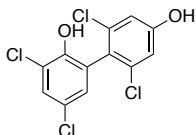
2',3,4',5-Tetrachloro-2,6'-dihydroxy-1,1'-biphenyl (28)



Synthesised from 2',3,4',5-tetrachloro-2,6'-dimethoxy-1,1'-biphenyl (**19**, 153 mg, 0.43 mmol) in 65% yield (91.0 g, 0.28 mmol) as a white solid.

¹H-NMR (300 MHz, CDCl₃): δ = 7.47 (d, *J* = 2.5 Hz, 1 H), 7.134 (d, *J* = 2.6 Hz, 1 H), 7.126 (d, *J* = 2.1 Hz, 1 H), 6.96 (d, *J* = 2.0 Hz, 1 H), 5.68 (bs, 1 H), 5.09 (bs, 1 H). **¹³C-NMR** (75 MHz, CDCl₃): δ = 154.7, 148.4, 135.8, 135.2, 130.8, 130.0, 126.2, 122.2, 121.9, 121.8, 120.6, 115.5. **MS** (ESI⁻): *m/z* = 320.9 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8174. **HPLC**: *t_R* = 14.2 min (λ_{max} = 204, 292 nm).

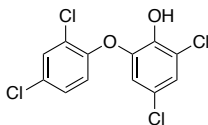
2',3,5,6'-Tetrachloro-2,4'-dihydroxy-1,1'-biphenyl (29)



Synthesised from 2',3,5,6'-tetrachloro-2,4'-dimethoxy-1,1'-biphenyl (**20**, 105 mg, 0.30 mmol) in 93% yield (90.0 mg, 0.28 mmol) as a white solid.

¹H-NMR (300 MHz, CDCl₃): δ = 7.41 (d, *J* = 2.5 Hz, 1 H), 7.05 (d, *J* = 2.5 Hz, 1 H), 6.93 (s, 2 H), 5.52 (bs, 2 H). **¹³C-NMR** (75 MHz, CDCl₃): δ = 156.4, 148.1, 135.9, 130.4, 129.0, 126.3, 125.9, 125.3, 121.1, 115.7. **MS** (ESI⁻): *m/z* = 320.8 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8169. **HPLC**: *t_R* = 13.8 min (λ_{max} = 204, 288 nm).

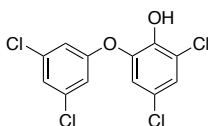
2,4-Dichloro-6-(2,4-dichlorophenoxy)phenol (30)



Synthesised from 1,5-dichloro-3-(2,4-dichlorophenoxy)-2-methoxybenzene (**21**, 149 mg, 0.44 mmol) in 81% yield (116 mg, 0.36 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.51 (d, *J* = 2.5 Hz, 1 H), 7.28 (dd, *J* = 6.6, 2.1 Hz, 1 H), 7.15 (d, *J* = 2.4 Hz, 2 H), 7.00 (d, *J* = 8.7 Hz, 1 H), 6.63 (d, *J* = 2.4 Hz, 1 H), 5.82 (bs, 1 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 149.9, 144.8, 142.1, 131.0, 128.6, 126.9, 125.1, 124.8, 121.8, 116.3. **¹H-NMR** (300 MHz, DMSO-*d*₆): δ = 10.37 (bs, 1 H), 7.76 (d, *J* = 2.5 Hz, 1 H), 7.40 (dd, *J* = 8.8, 2.6 Hz, 1 H), 7.37 (d, *J* = 2.5 Hz, 1 H), 7.02 (d, *J* = 8.8 Hz, 1 H), 6.91 (d, *J* = 2.5 Hz, 1 H). **¹³C-NMR** (75 MHz, DMSO-*d*₆): δ = 150.9, 145.0, 144.1, 130.1, 128.7, 128.3, 125.0, 124.7, 122.7 (two signals), 120.5, 118.1. **MS** (ESI⁻): *m/z* = 320.8 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8175. **HPLC**: *t_R* = 21.4 min (λ_{max} = 204, 284 nm).

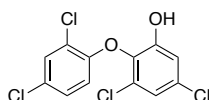
2,4-Dichloro-6-(3,5-dichlorophenoxy)phenol (31)



Synthesised from 1,5-dichloro-3-(3,5-dichlorophenoxy)-2-methoxybenzene (**22**, 138 mg, 0.41 mmol) in 82% yield (108 mg, 0.33 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.23 (d, *J* = 2.4 Hz, 1 H), 7.15 (t, *J* = 1.8 Hz, 1 H), 6.90 (d, *J* = 2.5 Hz, 1 H), 6.89 (d, *J* = 1.8 Hz, 2 H), 5.64 (bs, 1 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 157.6, 143.3, 143.2, 136.1, 125.9, 125.4, 124.5, 122.2, 199.7, 116.6. **MS** (ESI⁻): *m/z* = 320.8 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8169. HPLC: *t_R* = 25.6 min (λ_{max} = 204, 284 nm).

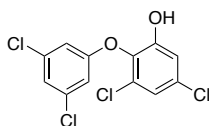
3,5-Dichloro-2-(2,4-dichlorophenoxy)phenol (**32**)



Synthesised from 1,5-dichloro-2-(2,4-dichlorophenoxy)-3-methoxybenzene (**23**, 83.0 mg, 0.25 mmol) in 93% yield (74.0 mg, 0.23 mmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.47 (d, *J* = 2.5 Hz, 1 H), 7.11 (dd, *J* = 8.8, 2.5 Hz, 1 H), 7.06-7.00 (m, 2 H), 6.54 (d, *J* = 8.8 Hz, 1 H), 5.60 (bs, 1 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 151.0, 150.3, 137.1, 132.2, 130.8, 128.9, 128.3, 128.1, 123.8, 122.4, 116.2, 115.8. **MS** (ESI⁻): *m/z* = 320.8 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8170. **HPLC**: *t_R* = 23.3 min (λ_{max} = 204, 284 nm).

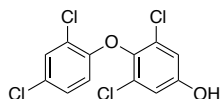
3,5-Dichloro-2-(3,5-dichlorophenoxy)phenol (**33**)



Synthesised from 1,5-dichloro-2-(3,5-dichlorophenoxy)-3-methoxybenzene (**24**, 31.3 mg, 92.6 μmol) in 69% yield (20.7 mg, 63.9 μmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.10 (t, *J* = 1.7 Hz, 1 H), 7.05 (d, *J* = 2.4 Hz, 1 H), 7.02 (d, *J* = 2.4 Hz, 1 H), 6.79 (d, *J* = 1.7 Hz, 2 H), 5.54 (bs, 1 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 157.5, 150.3, 136.3, 136.2, 132.4, 128.6, 124.0, 122.5, 116.3, 114.4. **MS** (ESI⁻): *m/z* = 320.9 [M-H]⁻. **HRMS** (ESI⁻) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9050; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8168.

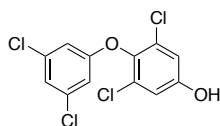
3,5-Dichloro-4-(2,4-dichlorophenoxy)phenol (**34**)



Synthesised from 1,3-dichloro-2-(2,4-dichlorophenoxy)-5-methoxybenzene (**25**, 43.0 mg, 127 μmol) in 51% yield (21.0 mg, 64.8 μmol) as a white solid.

¹H-NMR (400 MHz, CDCl₃): δ = 7.46 (d, *J* = 2.5 Hz, 1 H), 7.07 (dd, *J* = 8.8, 2.5 Hz, 1 H), 6.91 (s, 2 H), 6.40 (d, *J* = 8.8 Hz, 1 H), 5.42 (bs, 1 H). **¹³C-NMR** (100 MHz, CDCl₃): δ = 153.5, 151.5, 140.7, 130.6, 130.0, 127.8, 127.7, 123.5, 116.5, 115.1. **MS** (ESI⁻): *m/z* = 320.7 [M-H]⁻, 642.8 [2M-H]⁻. **HRMS** (ESI) *m/z* calcd for C₁₂H₅Cl₄O₂ [M-H]⁻: 320.9049, found: 320.9051; calcd for C₂₄H₁₁Cl₈O₄ [2M-H]⁻: 642.8171, found: 642.8165.

3,5-Dichloro-4-(3,5-dichlorophenoxy)phenol (**35**)



Synthesised from 1,3-dichloro-2-(3,5-dichlorophenoxy)-5-methoxybenzene (**26**, 16.0 mg, 47.0 μmol) in 13% yield (2.00 mg, 6.17 μmol) as a white solid.

$^1\text{H-NMR}$ (400 MHz, CDCl_3): δ = 7.05 (t, J = 1.7 Hz, 1 H), 6.92 (s, 2 H), 6.72 (d, J = 1.7 Hz, 2 H), 5.15 (bs, 1 H). **$^{13}\text{C-NMR}$** (100 MHz, CDCl_3): δ = 158.1, 153.6, 140.2, 135.8, 130.0, 123.1, 116.6, 114.2. **MS** (ESI $^-$): m/z = 320.9 $[\text{M-H}]^-$, 356.8 $[\text{M+Cl}]^-$, 642.7 $[2\text{M-H}]^-$. **HRMS** (ESI $^-$) m/z calcd for $\text{C}_{12}\text{H}_5\text{Cl}_4\text{O}_2$ $[\text{M-H}]^-$: 320.9049, found: 320.9049; calcd for $\text{C}_{24}\text{H}_{11}\text{Cl}_8\text{O}_4$ $[2\text{M-H}]^-$: 642.8171, found: 642.8161.

6. LC-MS Data

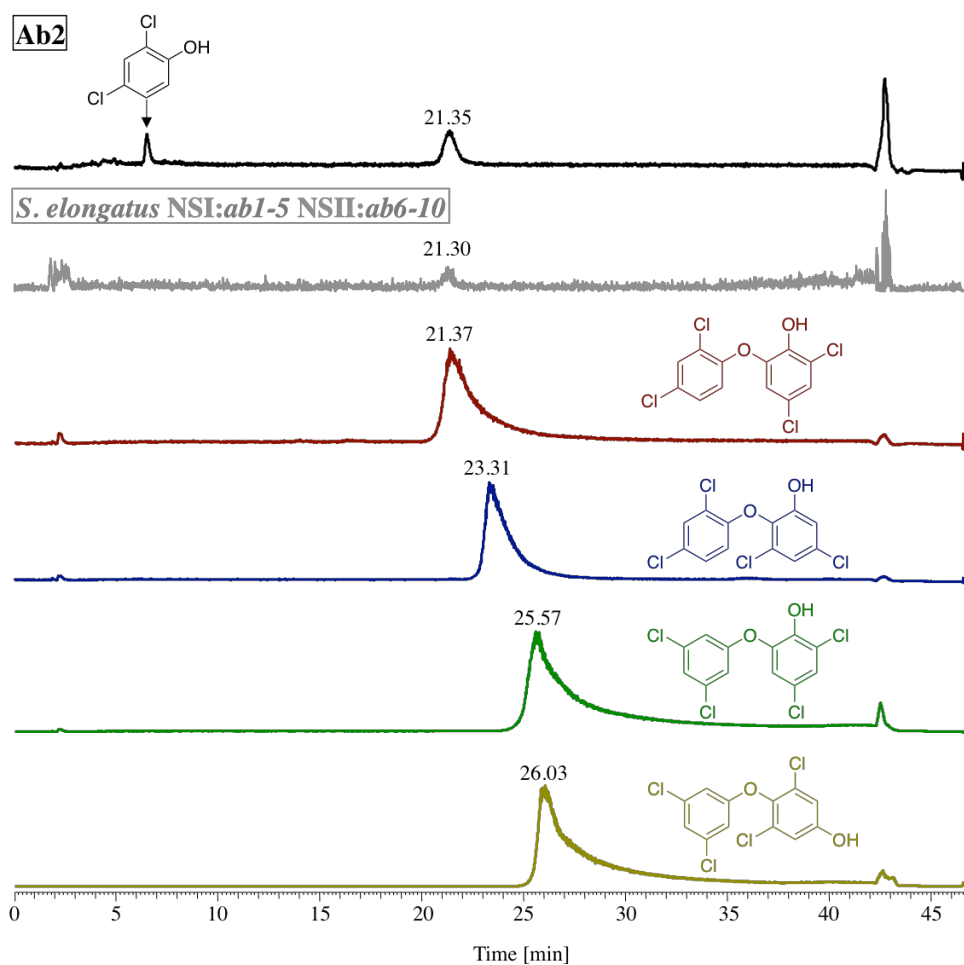


Fig. S7 Comparison of retention times of synthesised dichlorophenol-biaryl-ether-dimers with Ab2 2,4-DCP *in vivo* assay (black) and *S. elongatus* NSI:ab1-5 NSII:ab6-10 expression (grey).

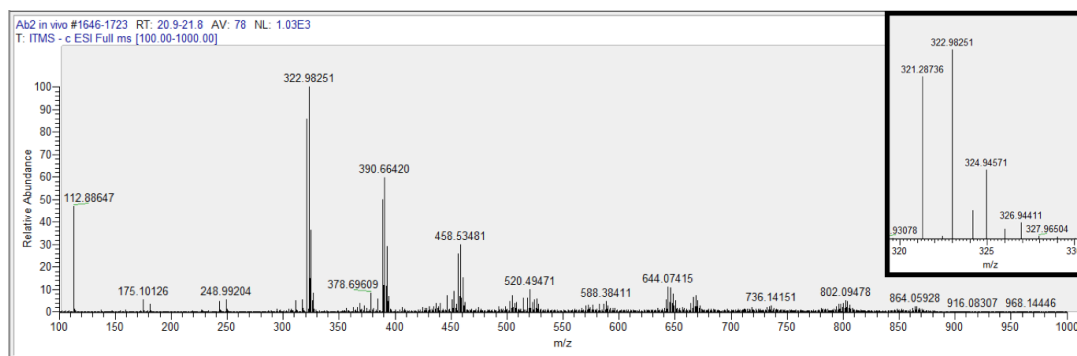


Fig. S8 ESI- spectrum of the 2,4-DCP dimer produced by Ab2, including the characteristic isotopic pattern for a tetrachlorinated substance.

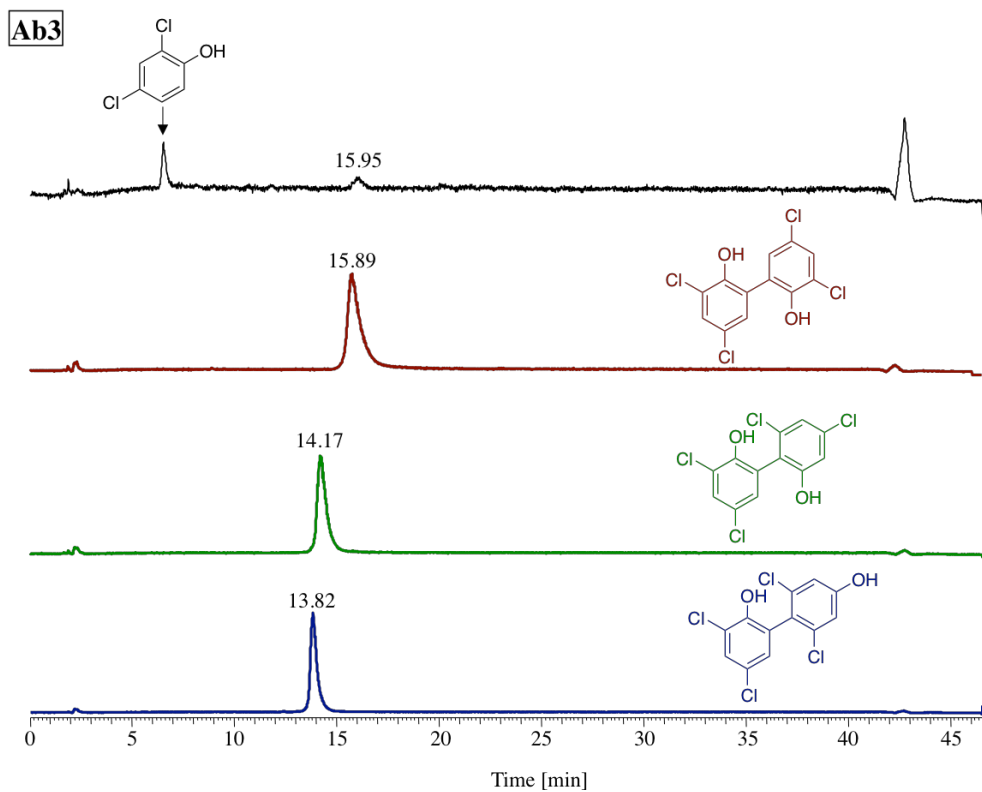


Fig. S9 Comparison of retention times of synthesised dichlorophenol-biaryl-dimers with Ab3 2,4-DCP *in vitro* assay.

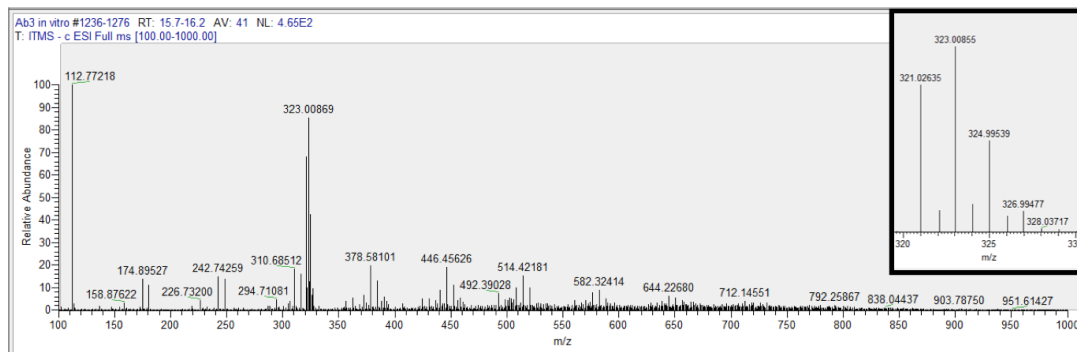


Fig. S10 ESI- spectrum of the 2,4-DCP dimer produced by Ab3, including the characteristic isotopic pattern for a tetrachlorinated substance.

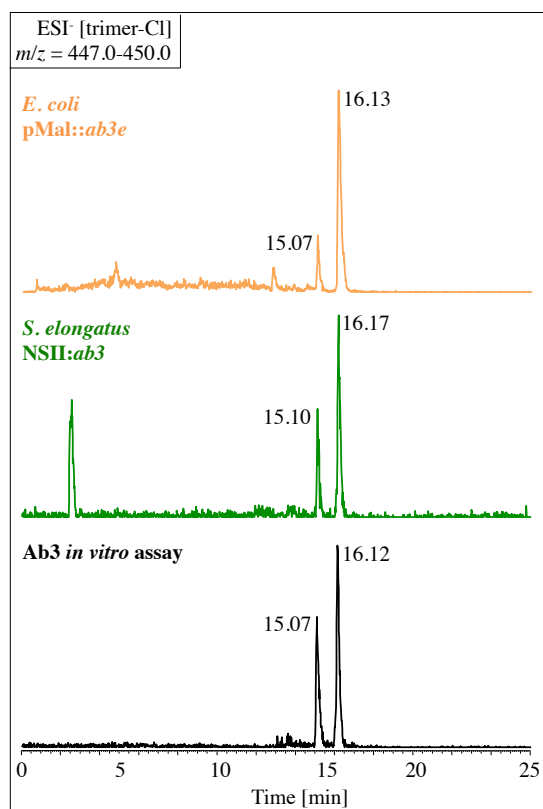


Fig. S11 HPLC-MS analysis (EIC⁻ $m/z = 447.0-450.0$) of *in vivo* and *in vitro* catalytic activity of Ab3 when incubated with 2,4-DCP. The formation of two DCP-trimers lacking one chlorine can be observed.

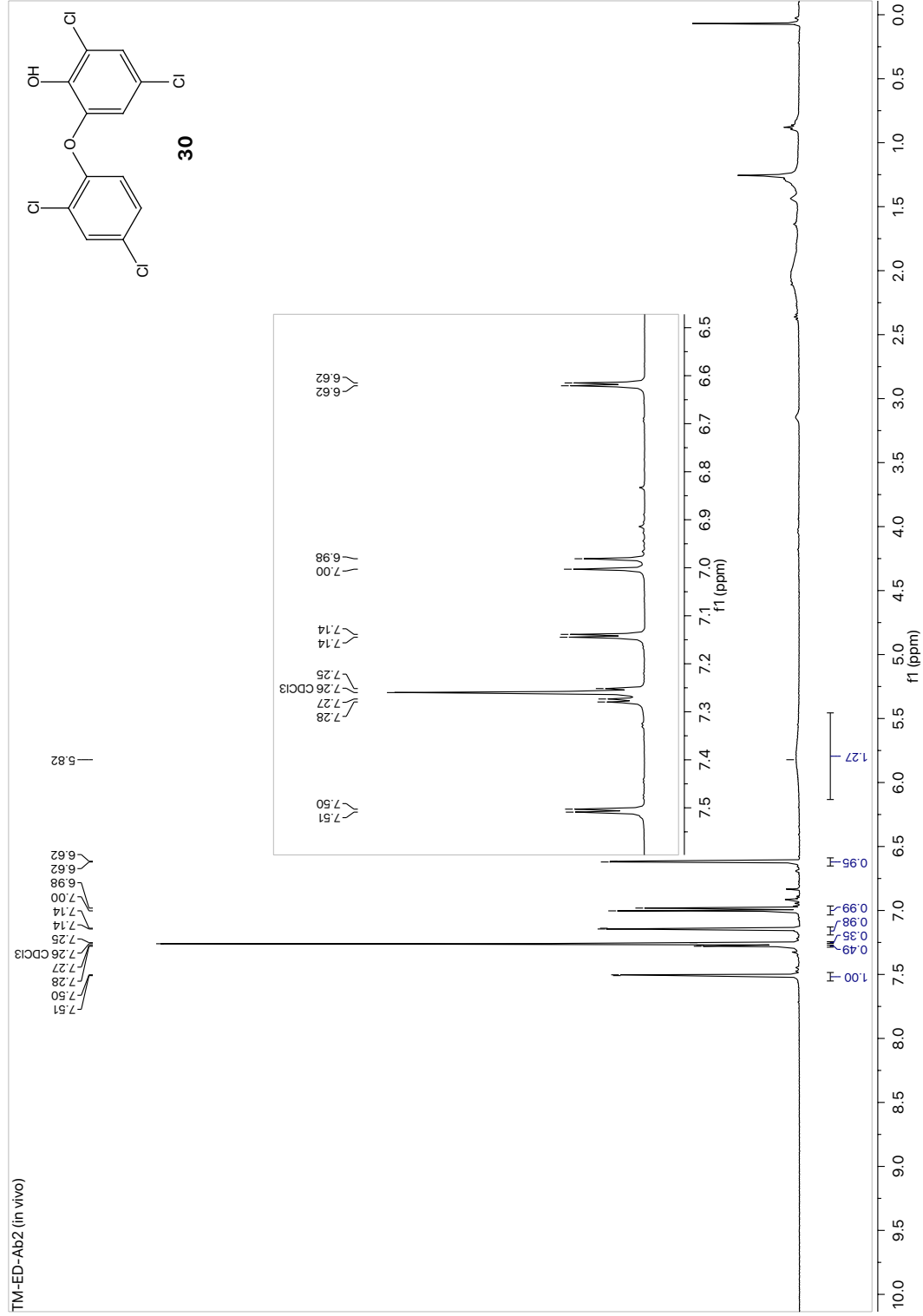
7. $^1\text{H-NMR}$ and $^{13}\text{C-NMR}$ Spectra

Fig. S12 $^1\text{H-NMR}$ spectrum of the major Ab2 2,4-DCP dimeric coupling product extracted and purified from *E. coli* BL21 pMal-SD::ab2e.

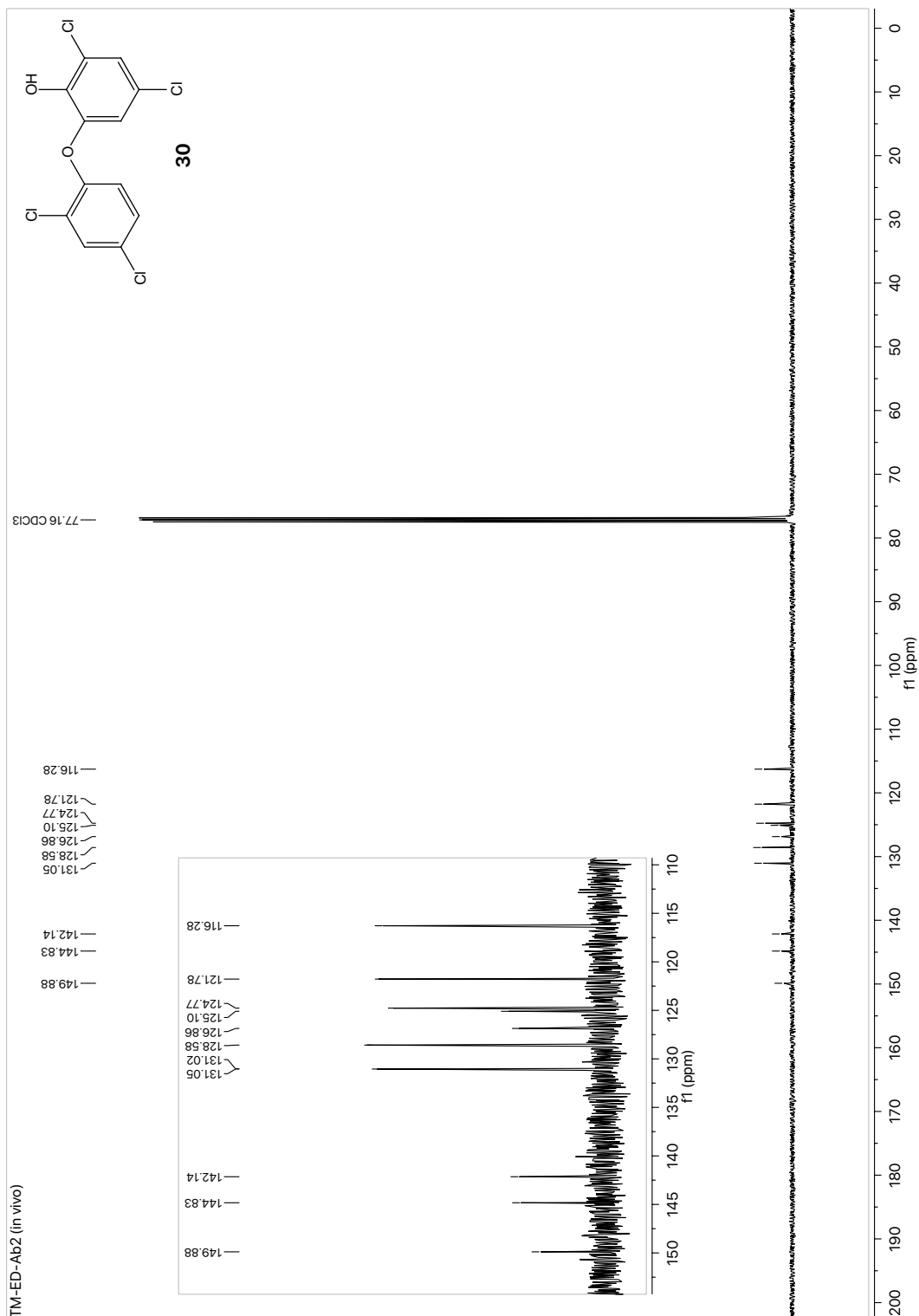


Fig. S13 ¹³C-NMR spectrum of the major Ab2 2,4-DCP dimeric coupling product extracted and purified from *E. coli* BL21 pMal-SD:*ab2e*.

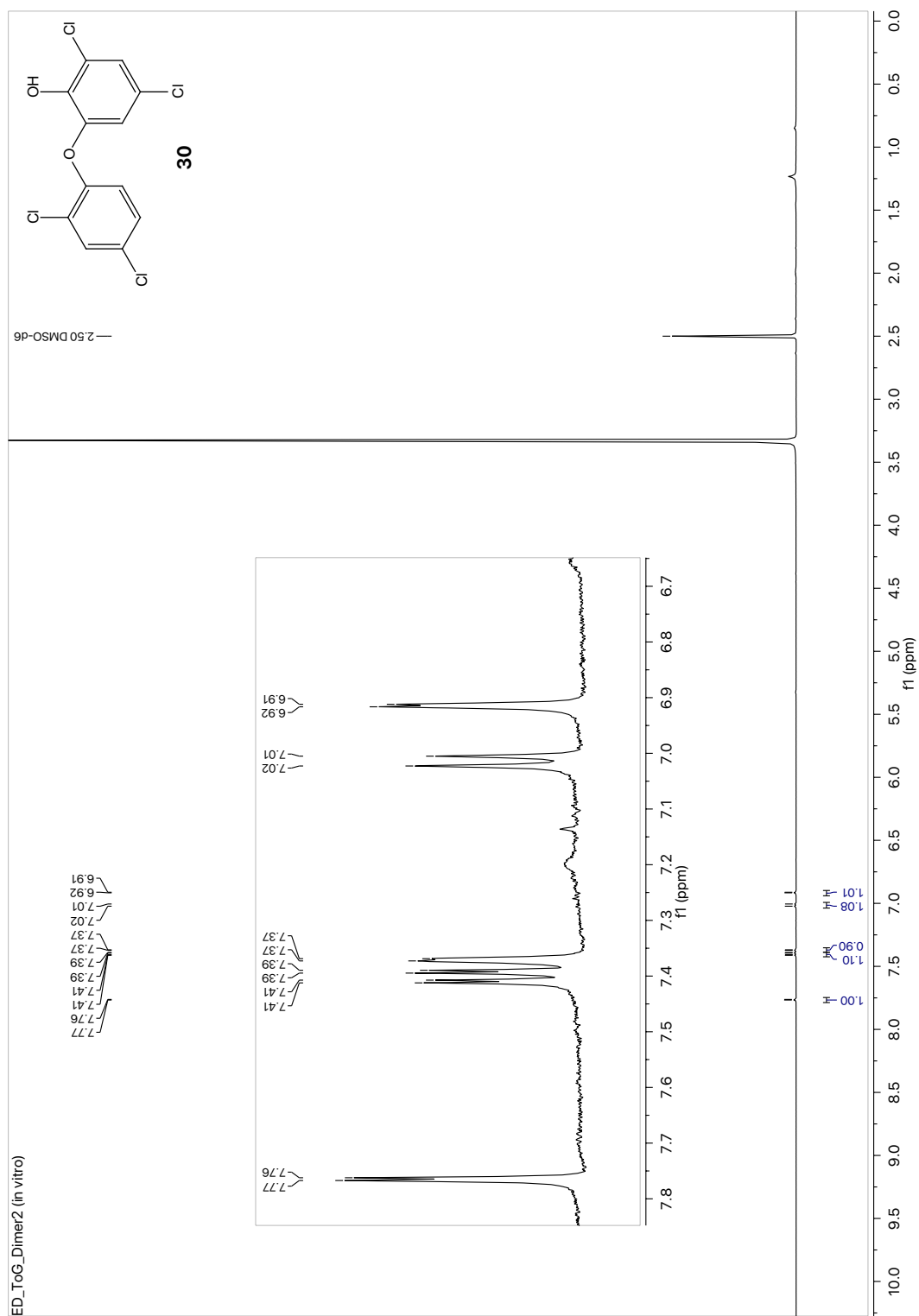


Fig. S14 ¹H-NMR spectrum of the major Ab2 2,4-DCP dimeric coupling product isolated from a large scale *in vitro* assay.

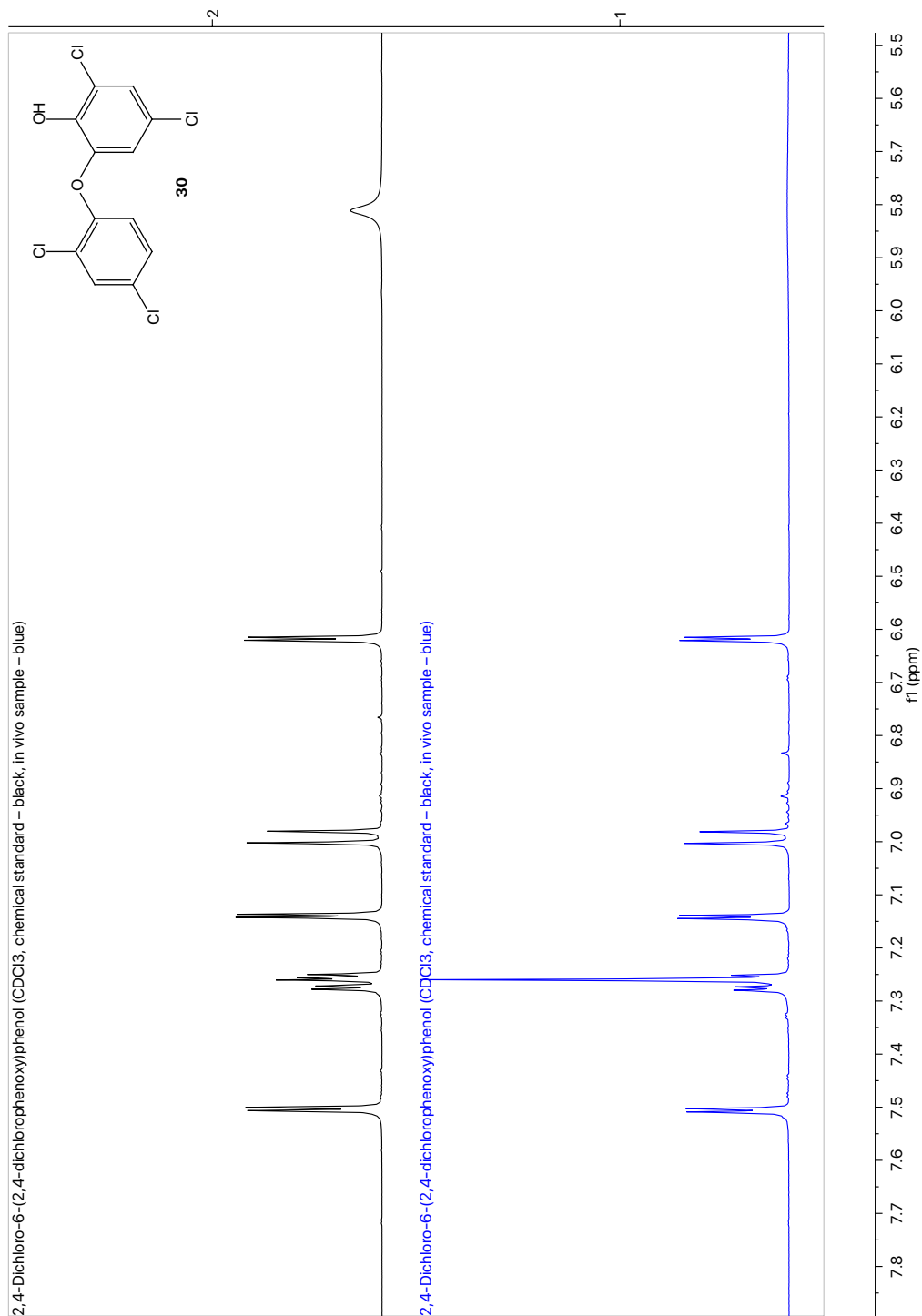


Fig. S15 Comparison of ¹H NMR spectra of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**) produced by synthesis (black) and *in vivo* (blue). Measured in CDCl₃.

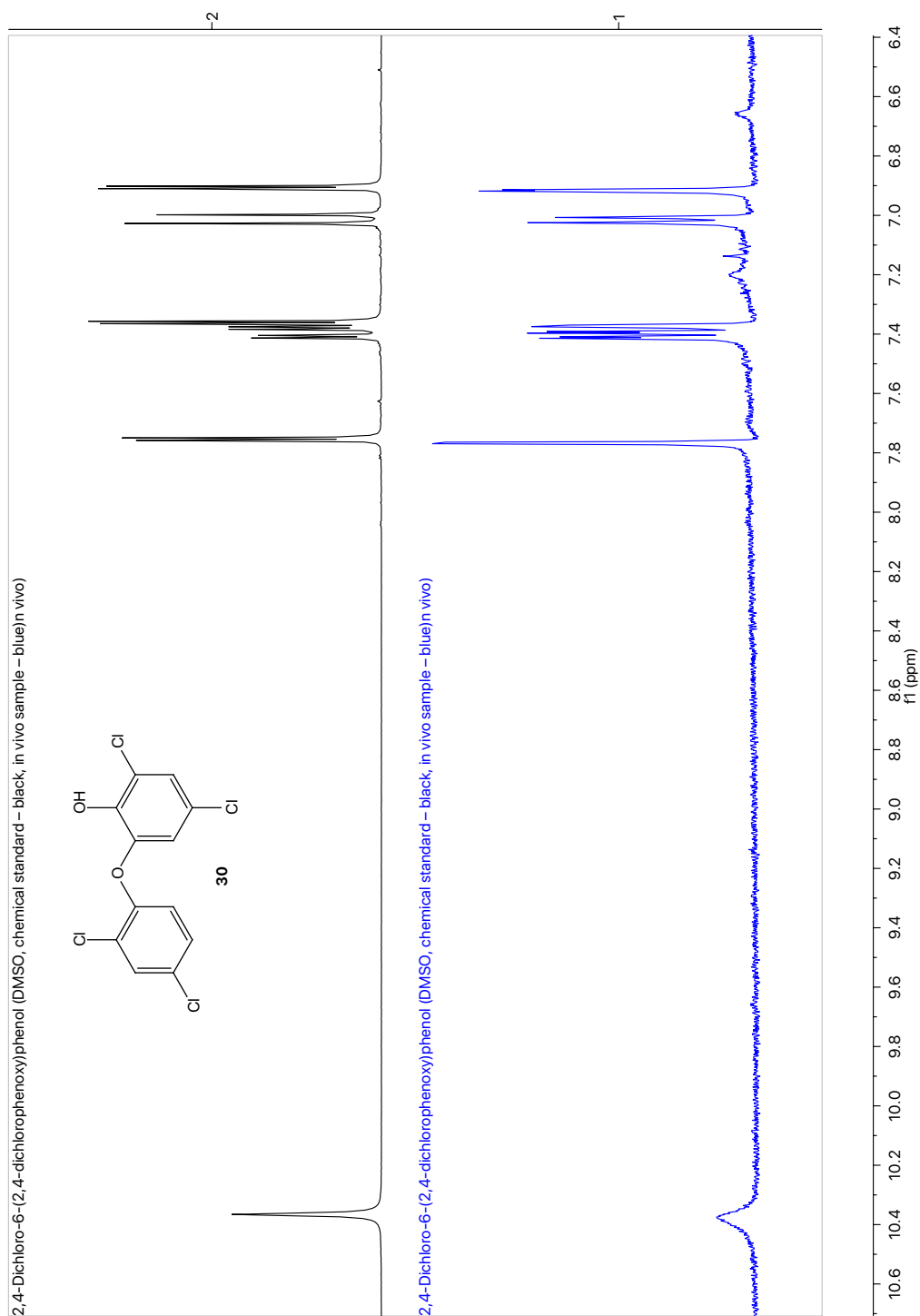


Fig. S16 Comparison of ^1H NMR spectra of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**) produced by synthesis (black) and *in vivo* (blue). Measured in DMSO-d_6 .

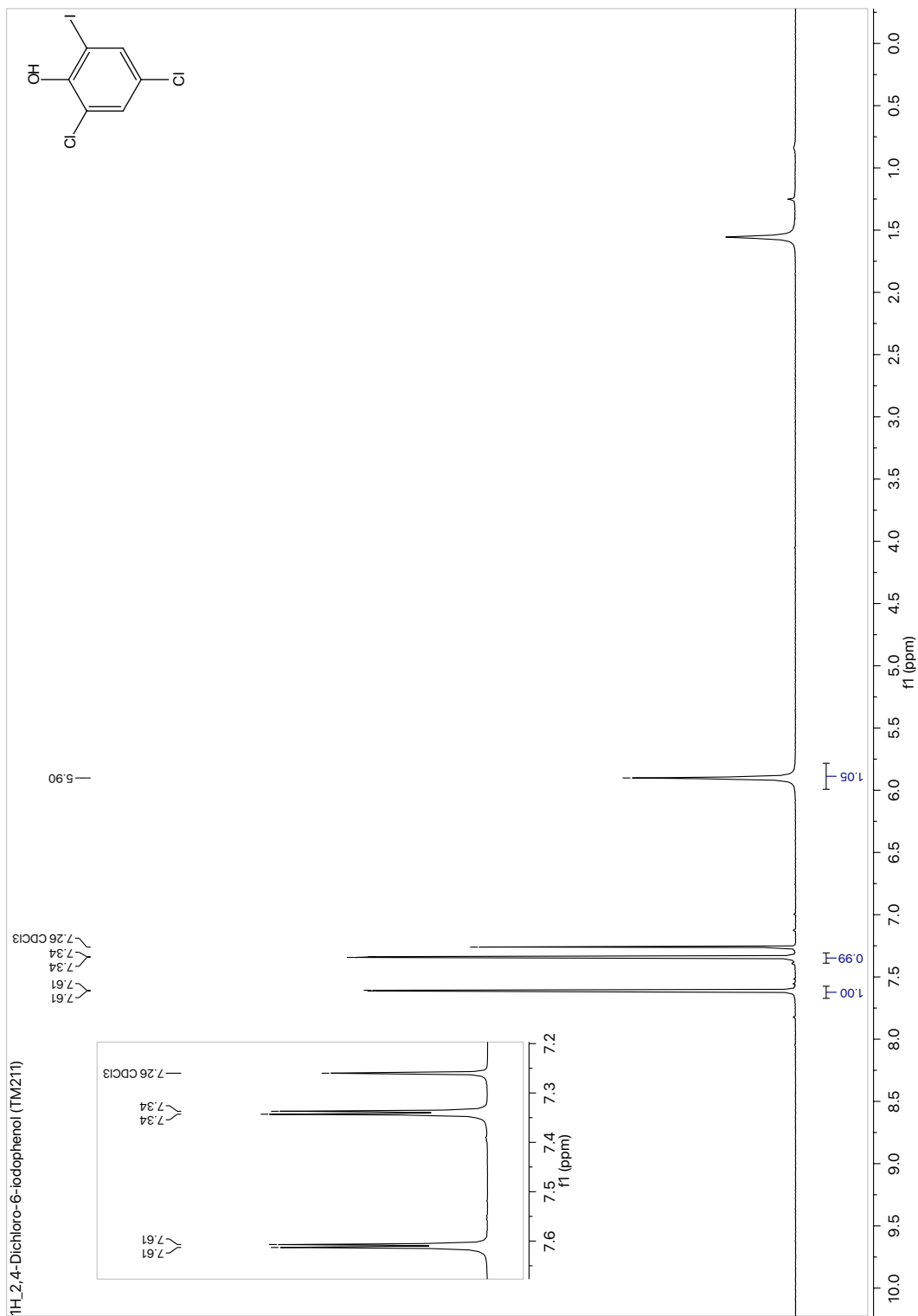


Fig. S17 ¹H-NMR spectrum of 2,4-dichloro-6-iodophenol.

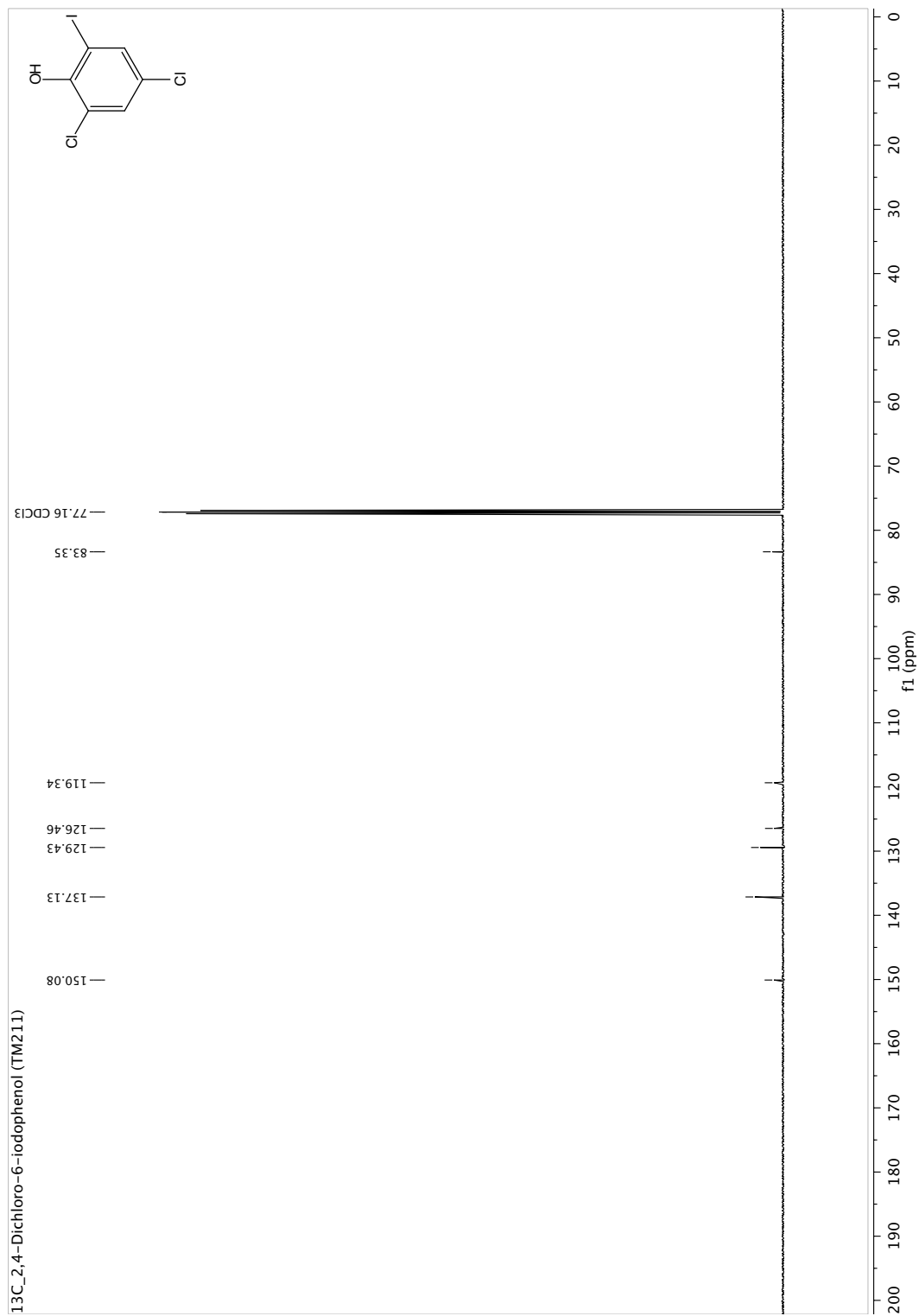


Fig. S18 ¹³C-NMR spectrum of 2,4-dichloro-6-iodophenol.

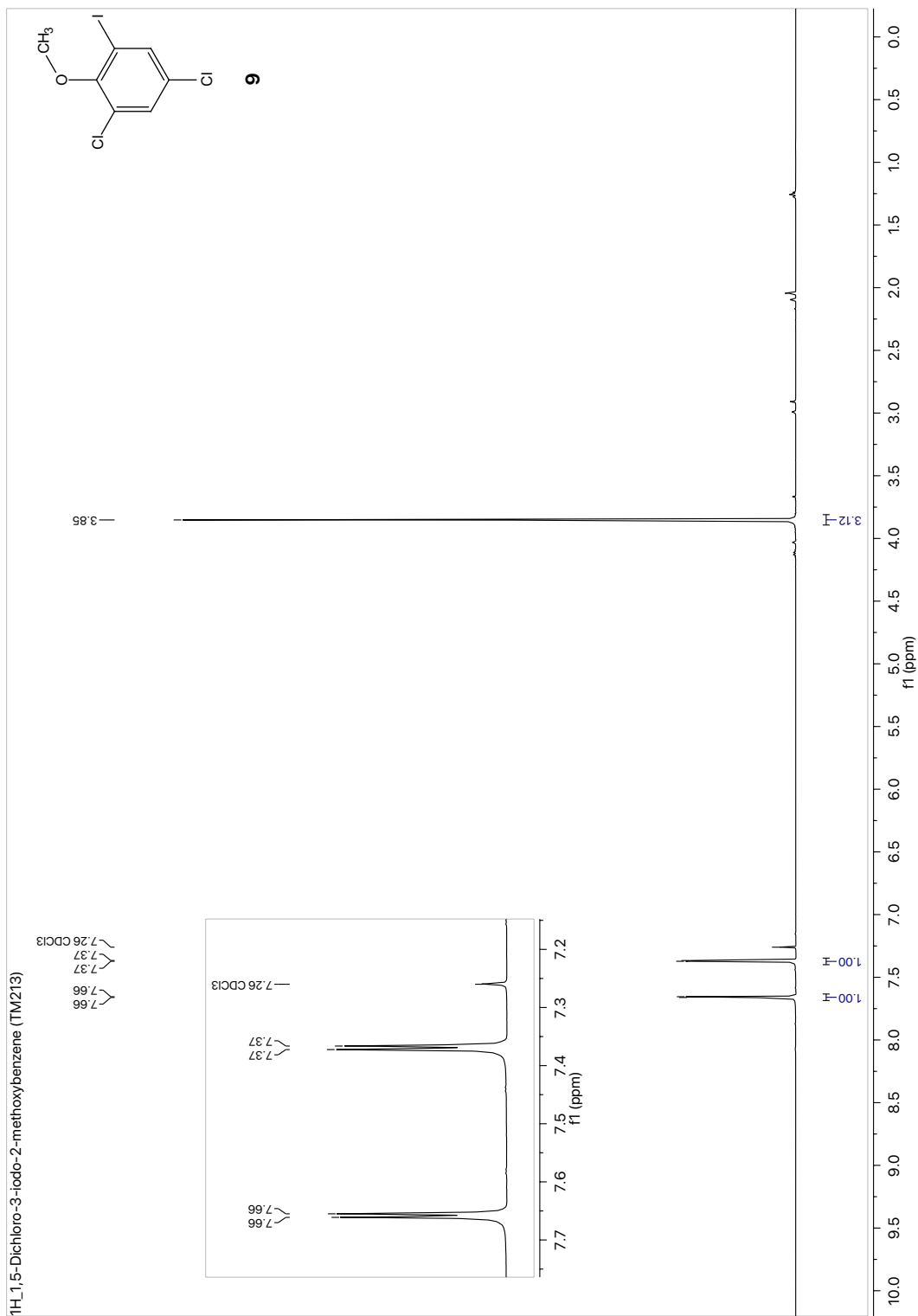


Fig. 19 ¹H-NMR spectrum of 1,5-dichloro-3-iodo-2-methoxybenzene (9).

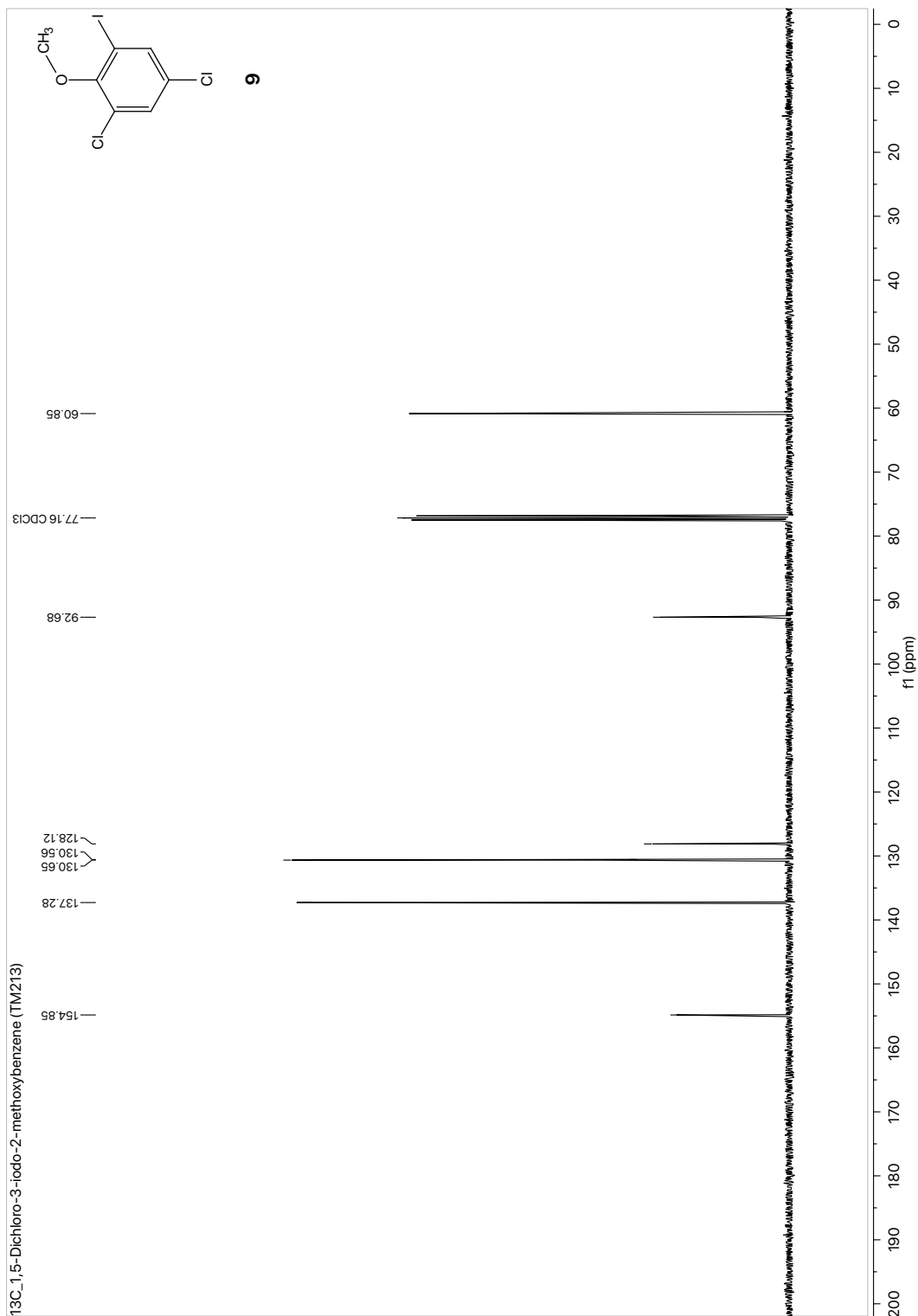


Fig. S20 ¹³C-NMR spectrum of 1,5-dichloro-3-iodo-2-methoxybenzene (**9**).

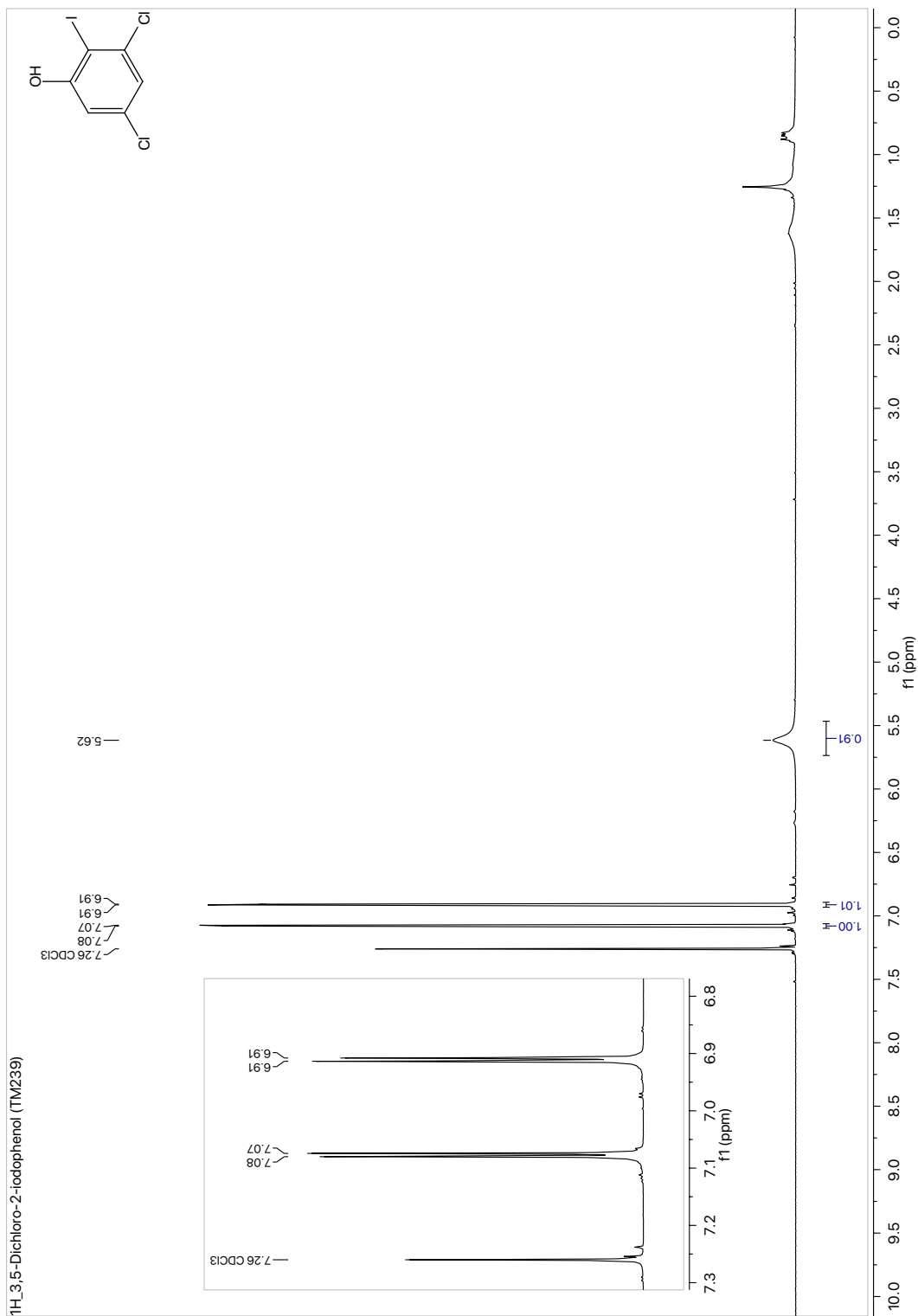


Fig. S21 ¹H-NMR spectrum of 3,5-dichloro-2-iodophenol.

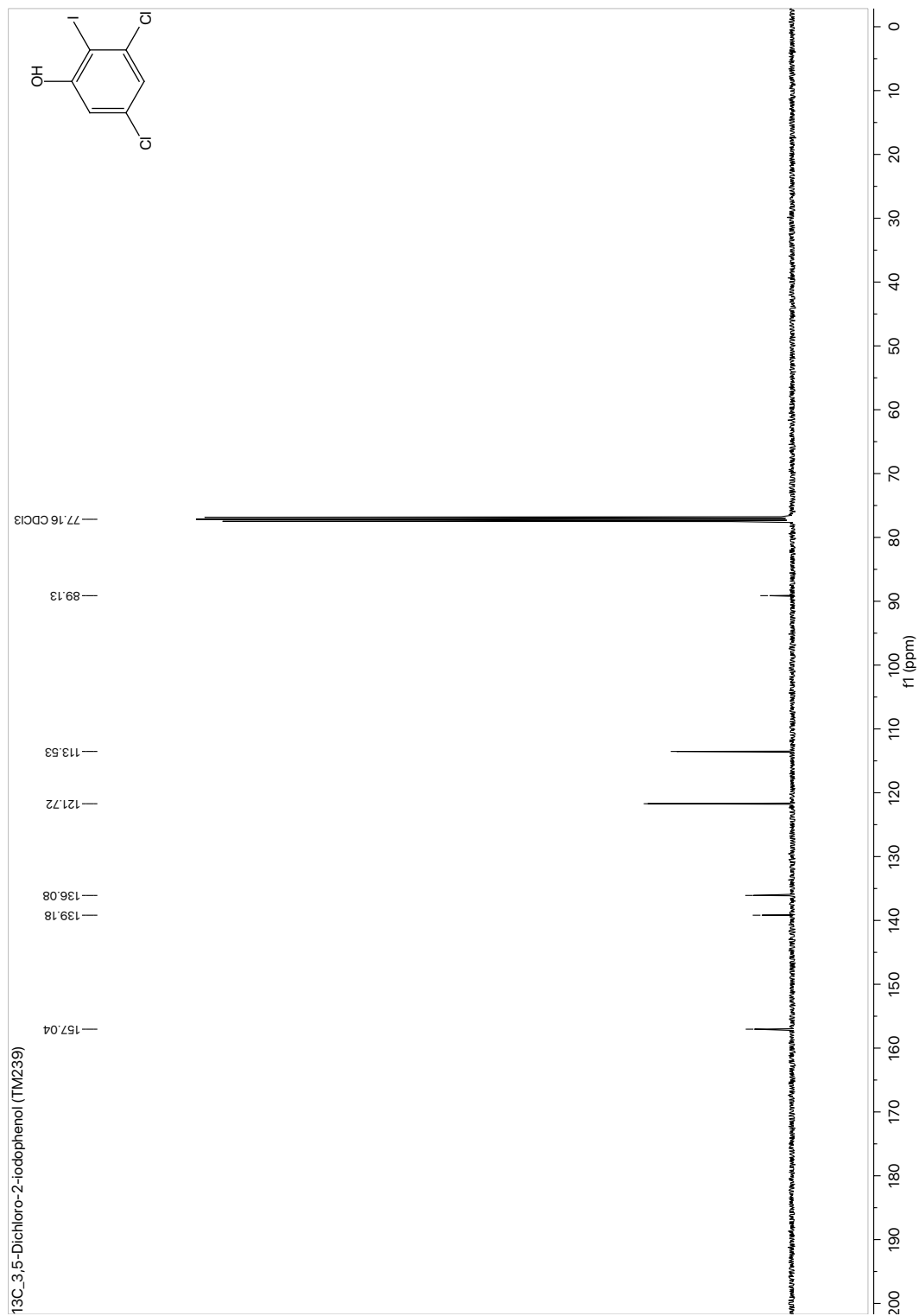


Fig. S22 ¹³C-NMR spectrum of 3,5-dichloro-2-iodophenol.

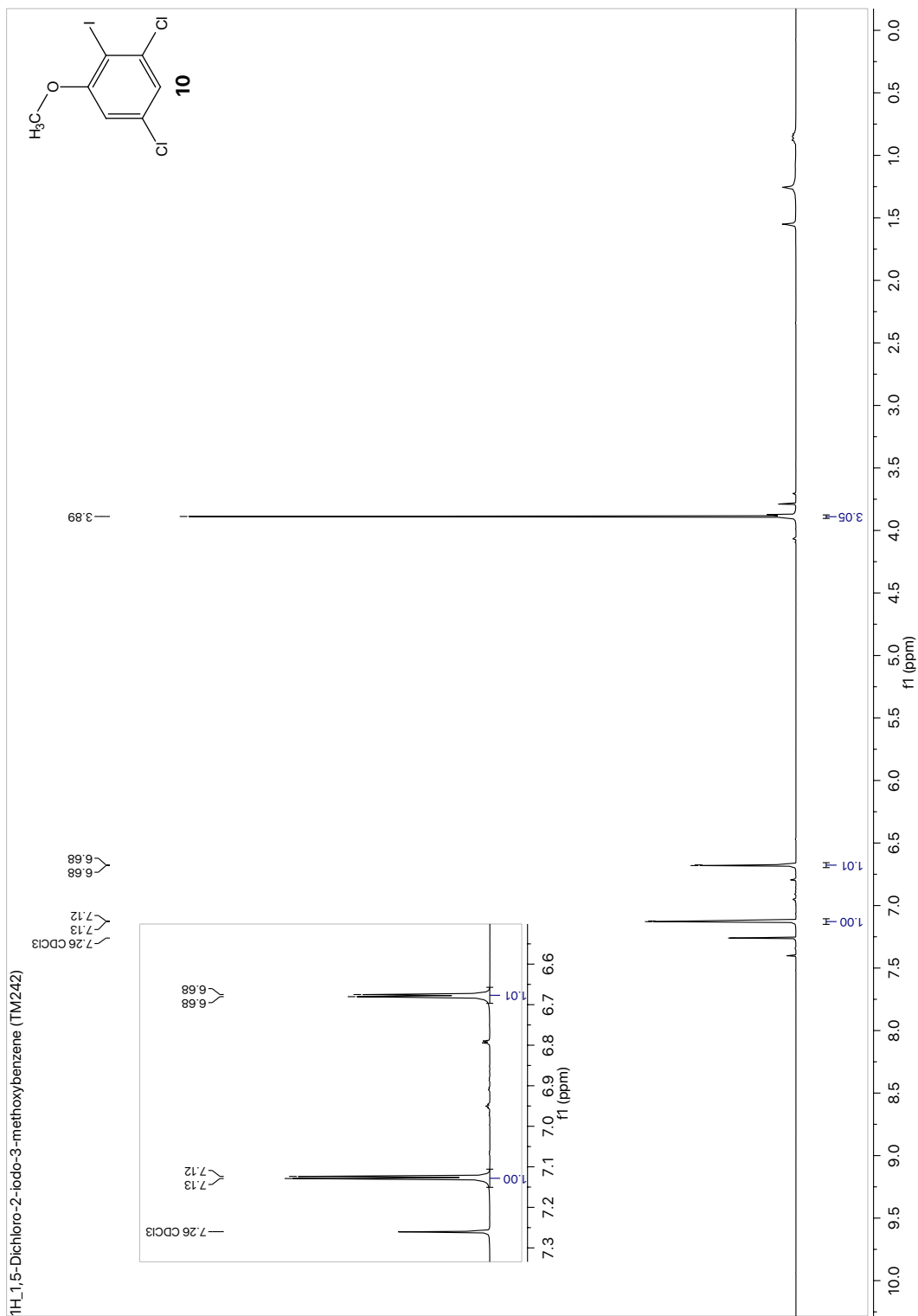


Fig. S23 ¹H-NMR spectrum of 1,5-dichloro-2-iodo-3-methoxybenzene (**10**).

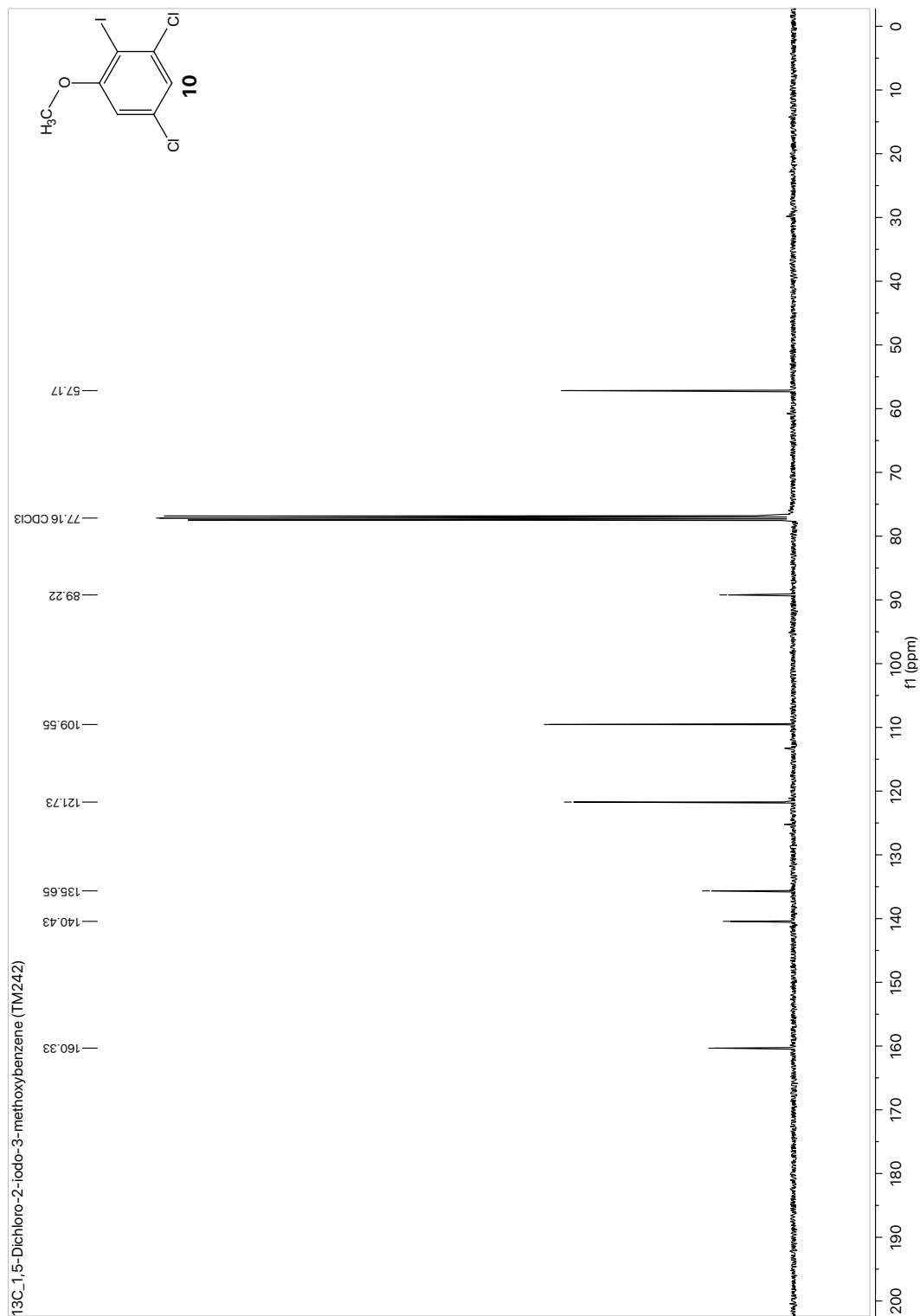


Fig. S24 ¹³C-NMR spectrum of 1,5-dichloro-2-iodo-3-methoxybenzene (**10**).

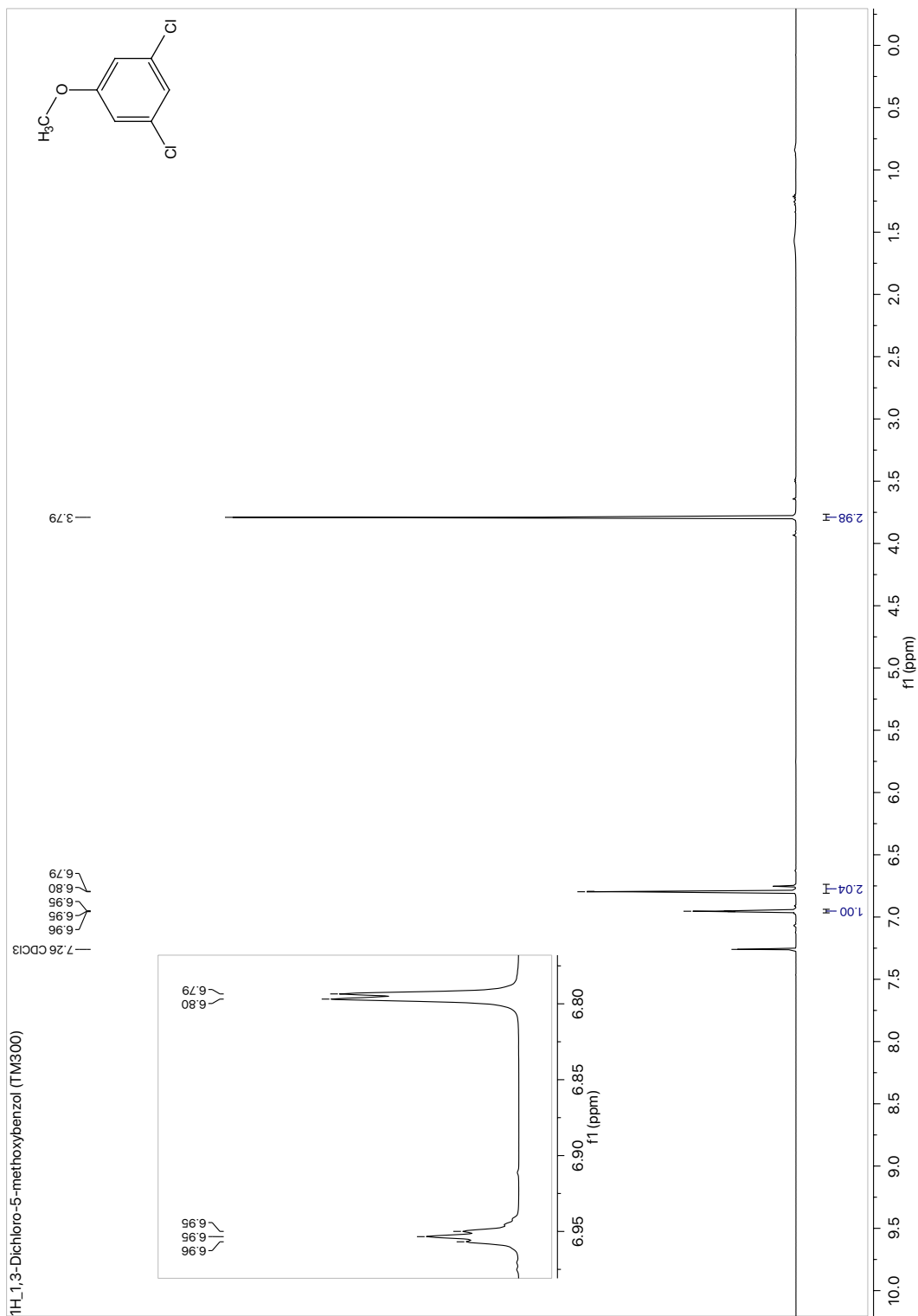


Fig. S25 ¹H-NMR spectrum of 1,3-dichloro-5-methoxybenzene.

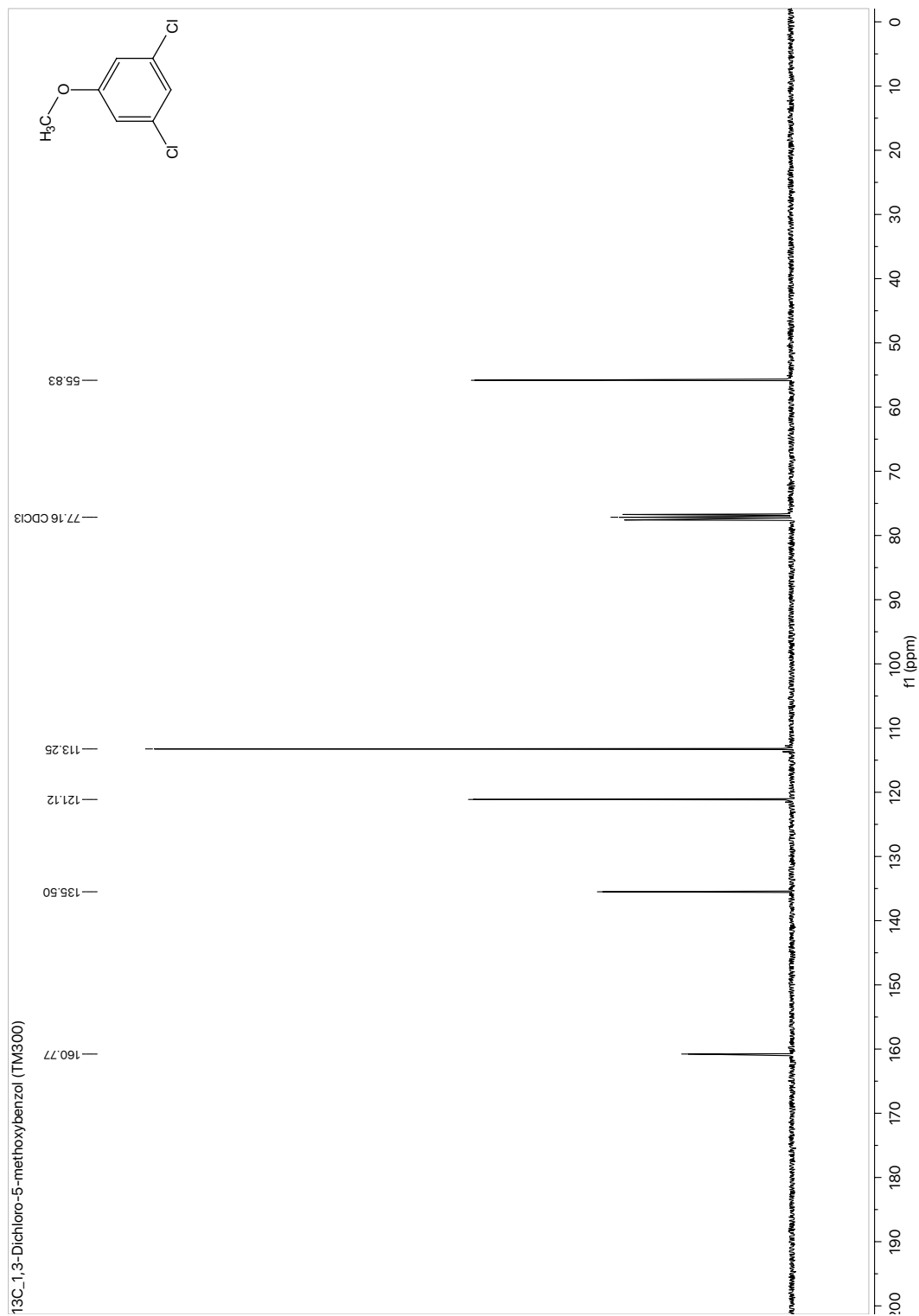


Fig. 26 ¹³C-NMR spectrum of 1,3-dichloro-5-methoxybenzene.

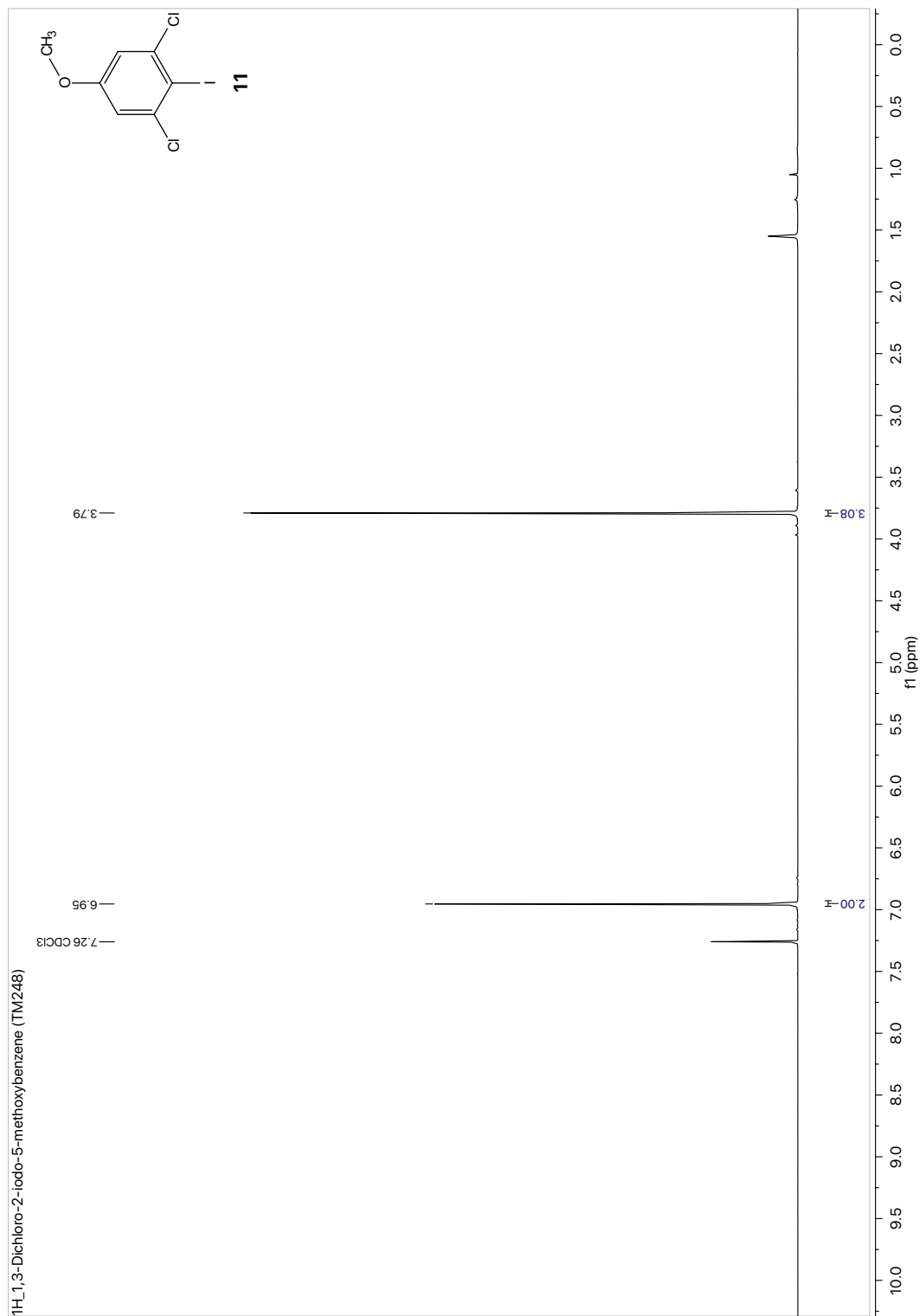


Fig. 27 ¹H-NMR spectrum of 1,3-dichloro-2-iodo-5-methoxybenzene (**11**).

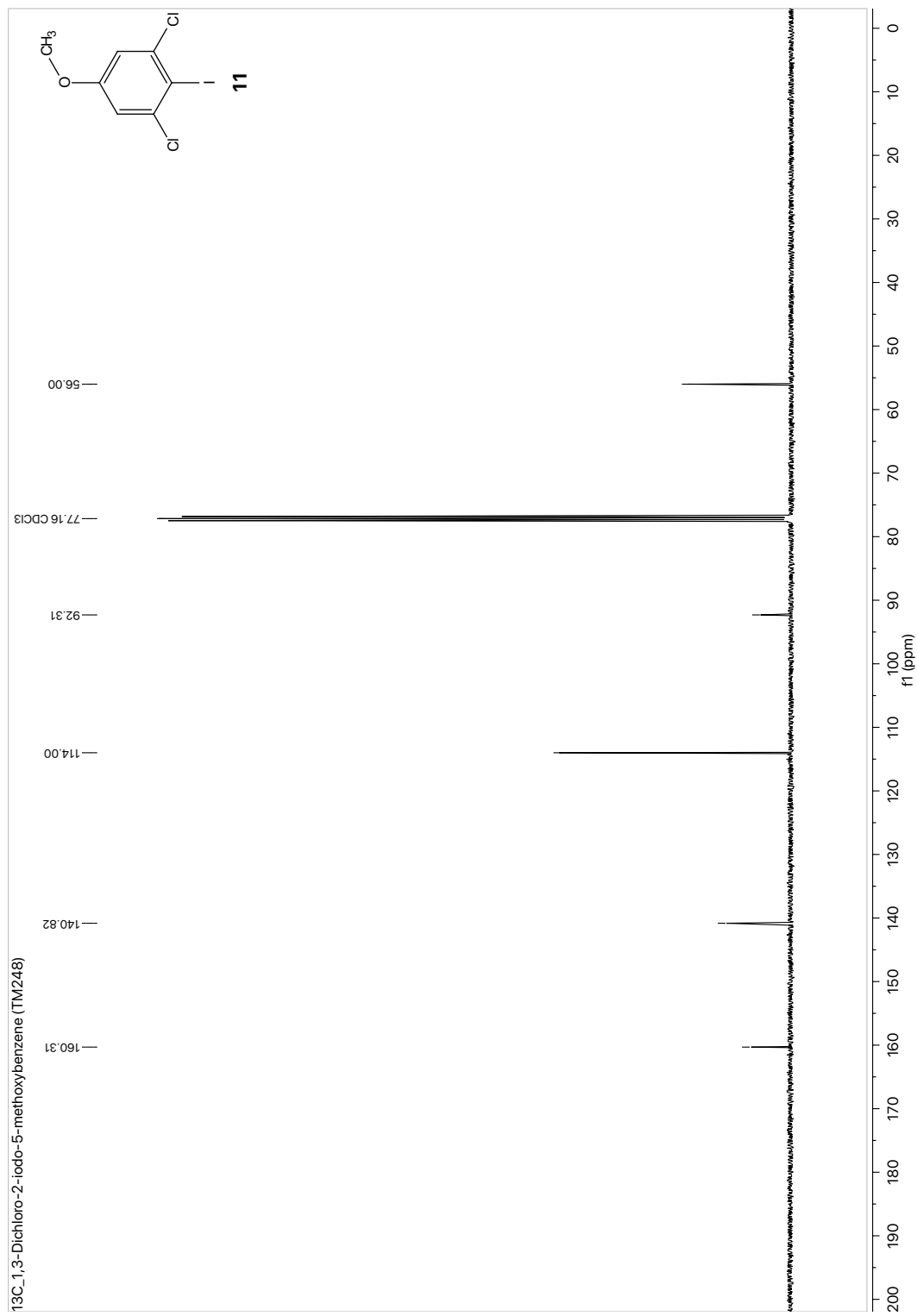


Fig. S28 ¹³C-NMR spectrum of 1,3-dichloro-2-iodo-5-methoxybenzene (**11**).

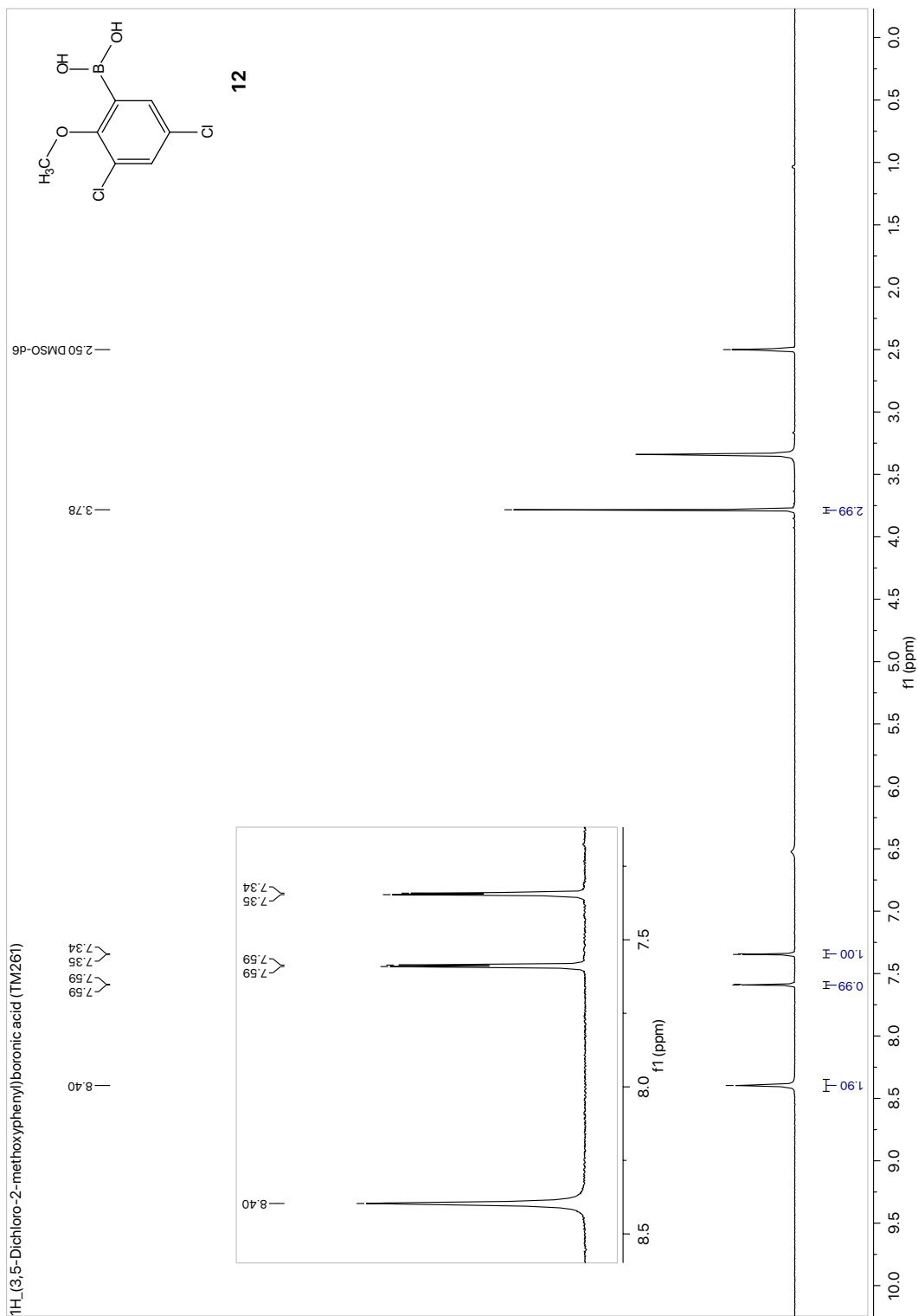


Fig. S29 ¹H-NMR spectrum of (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**).

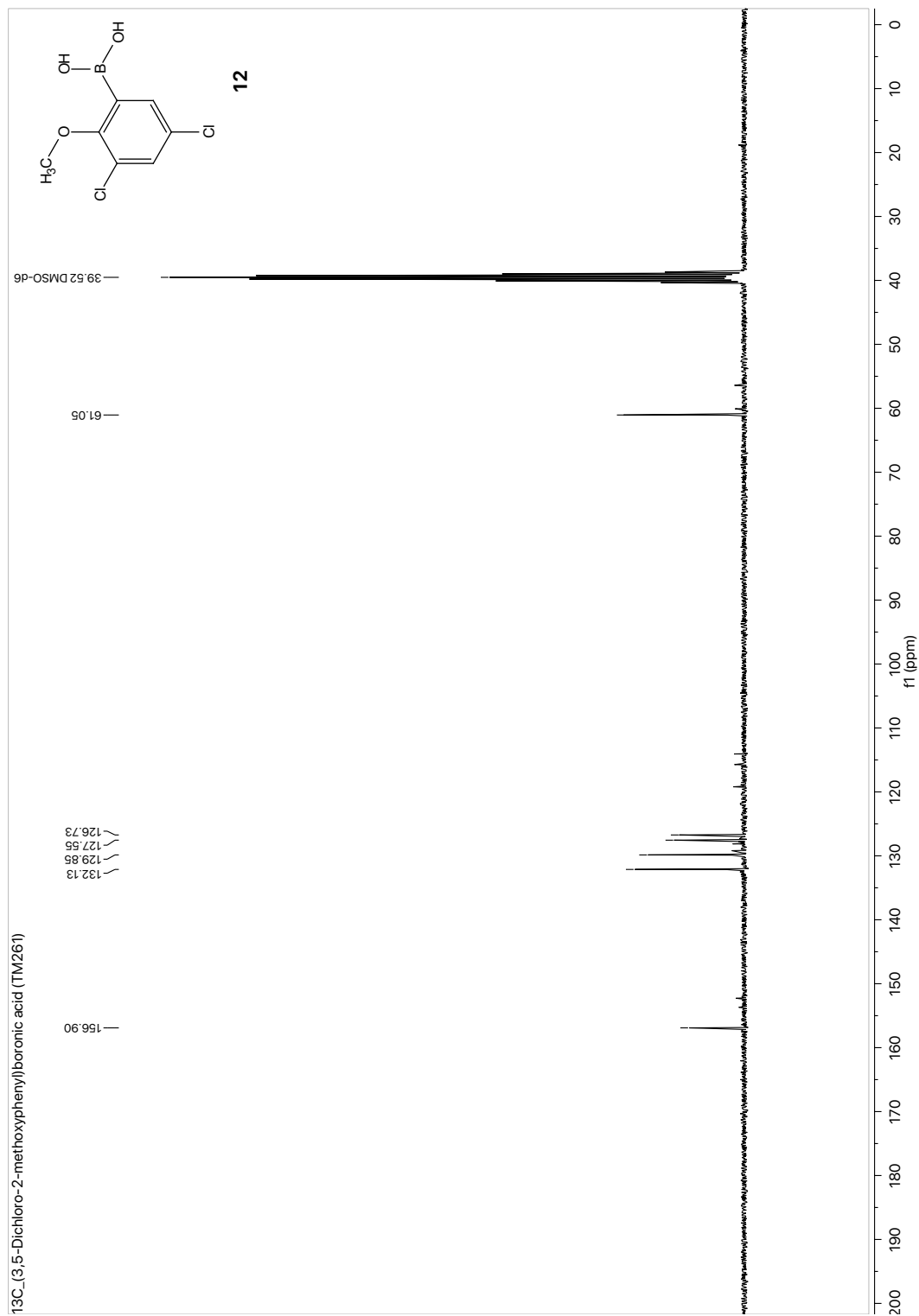


Fig. S30 ¹³C-NMR spectrum of (3,5-dichloro-2-methoxyphenyl)boronic acid (**12**).

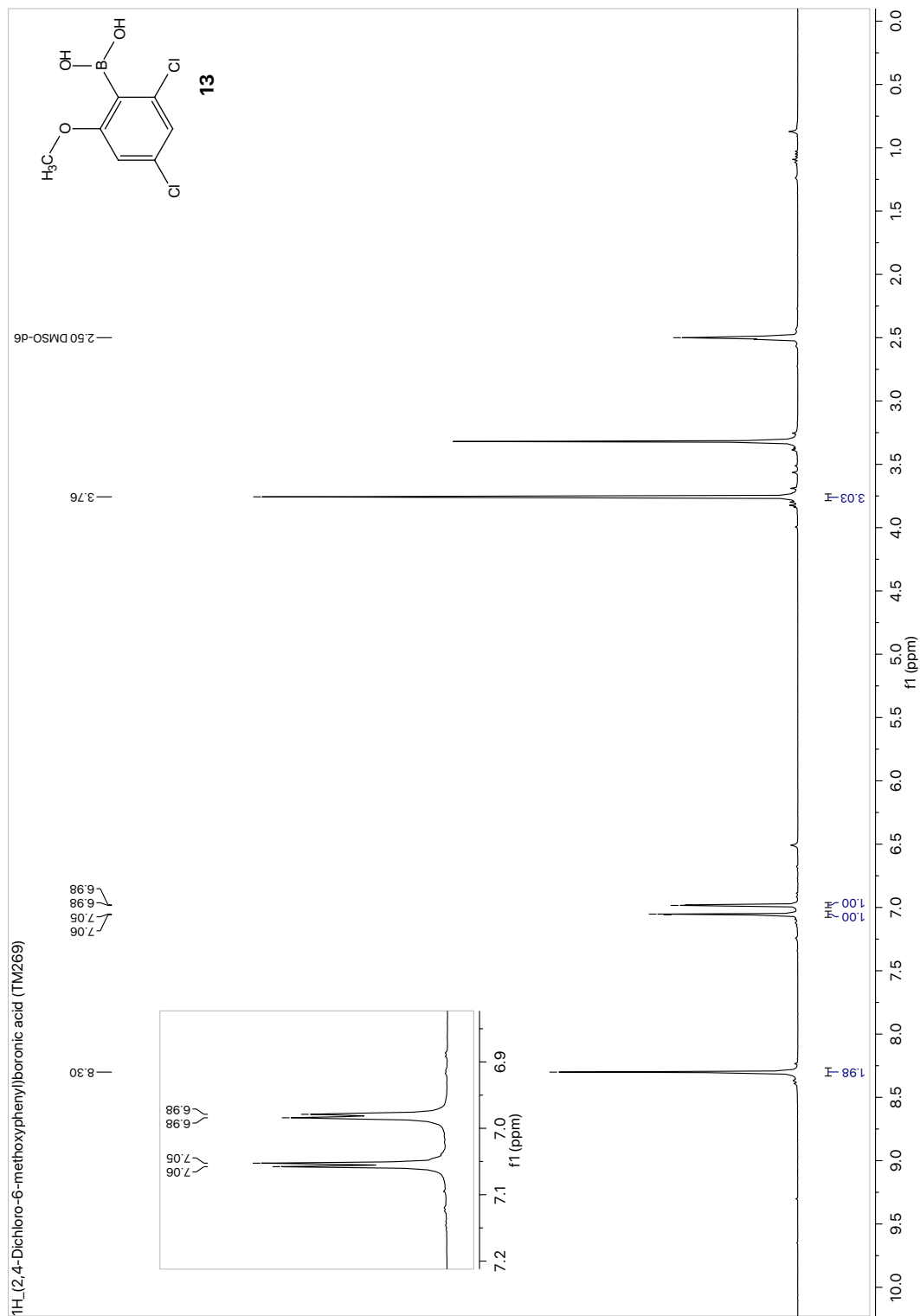


Fig. S31 ¹H-NMR spectrum of (2,4-dichloro-6-methoxyphenyl)boronic acid (**13**).

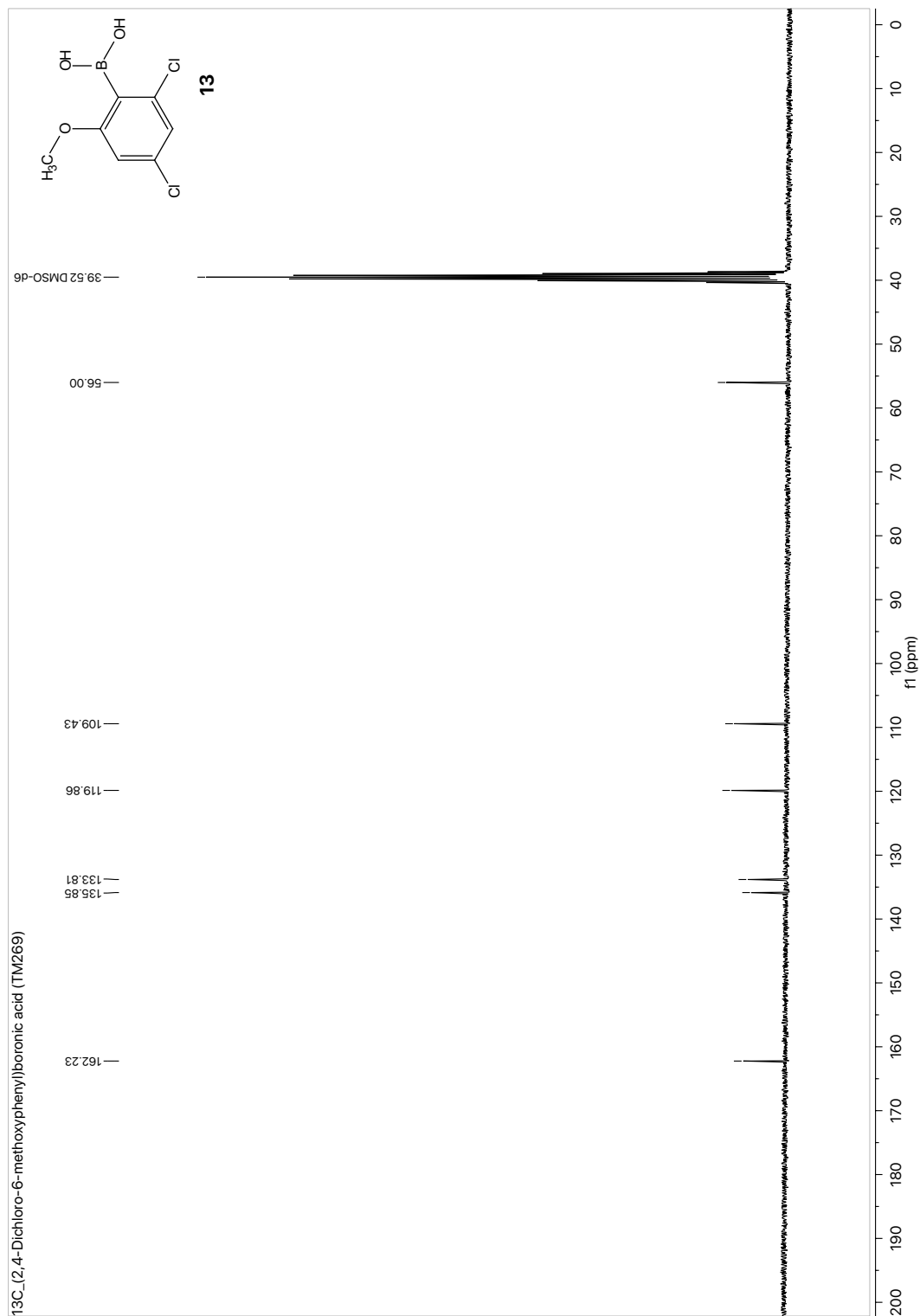


Fig. S32 ¹³C-NMR spectrum of (2,4-dichloro-6-methoxyphenyl)boronic acid (**13**).

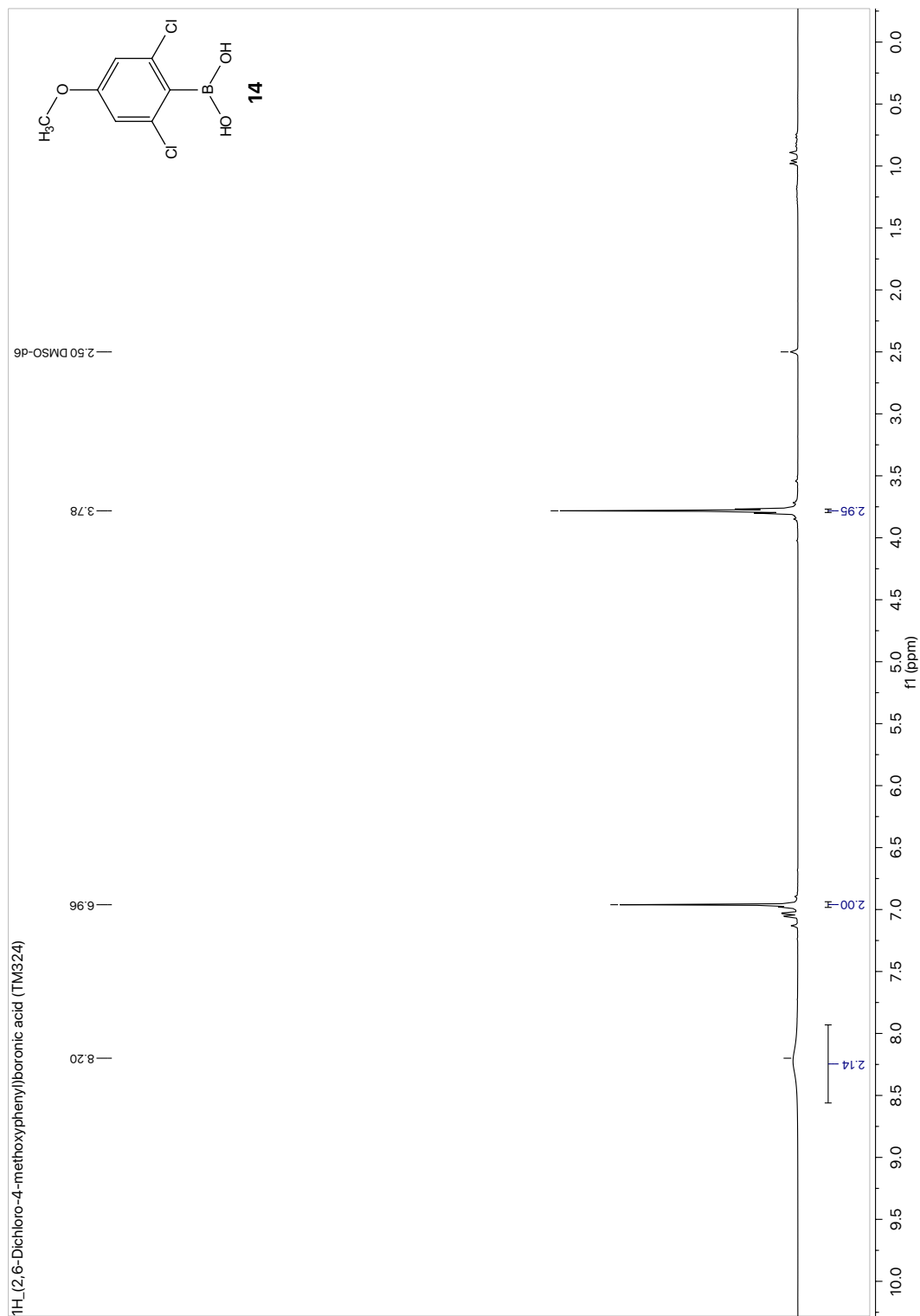


Fig. S33 ¹H-NMR spectrum of (2,6-dichloro-4-methoxyphenyl)boronic acid (**14**).

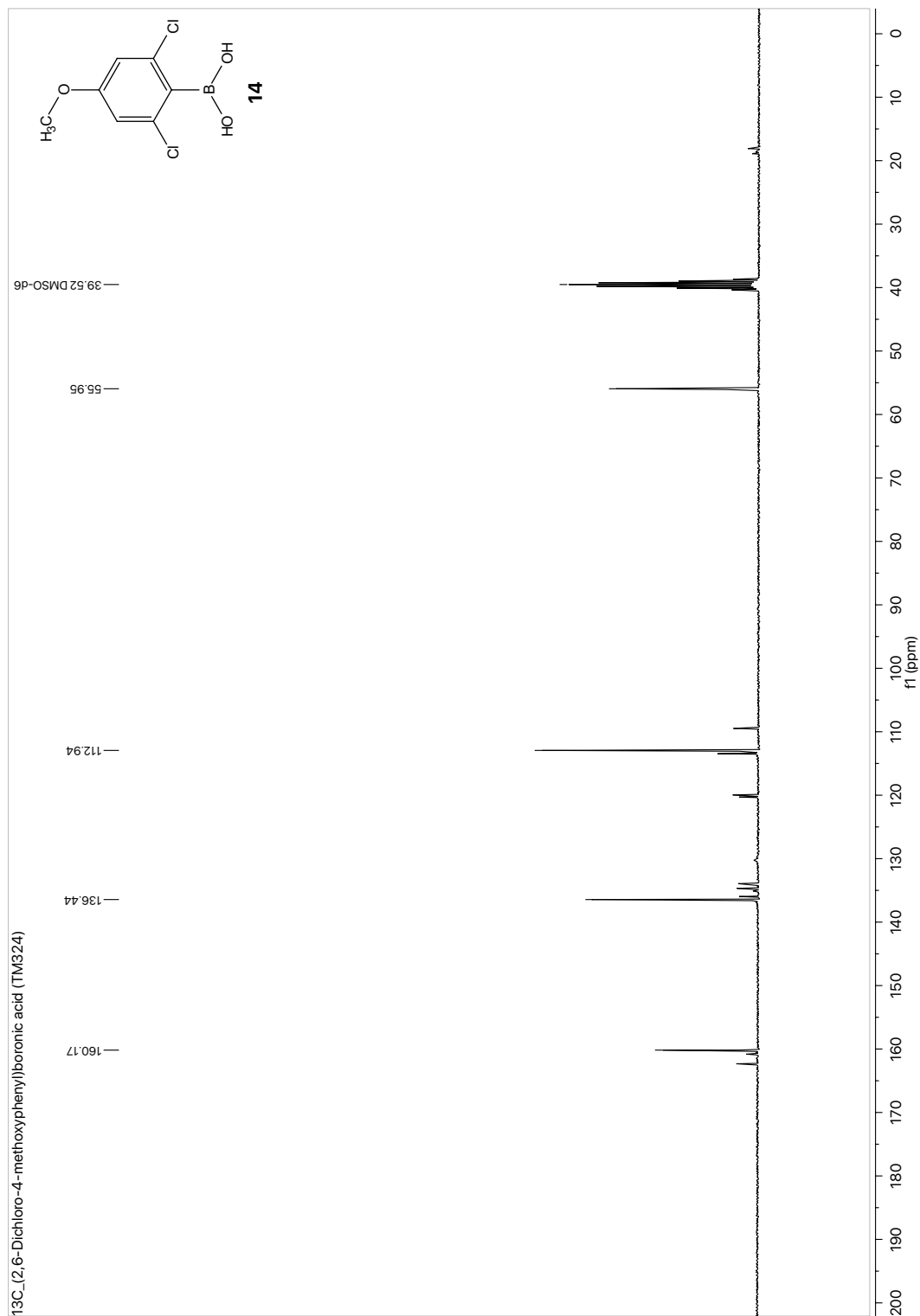


Fig. S34 ¹³C-NMR spectrum of (2,6-dichloro-4-methoxyphenyl)boronic acid (**14**).

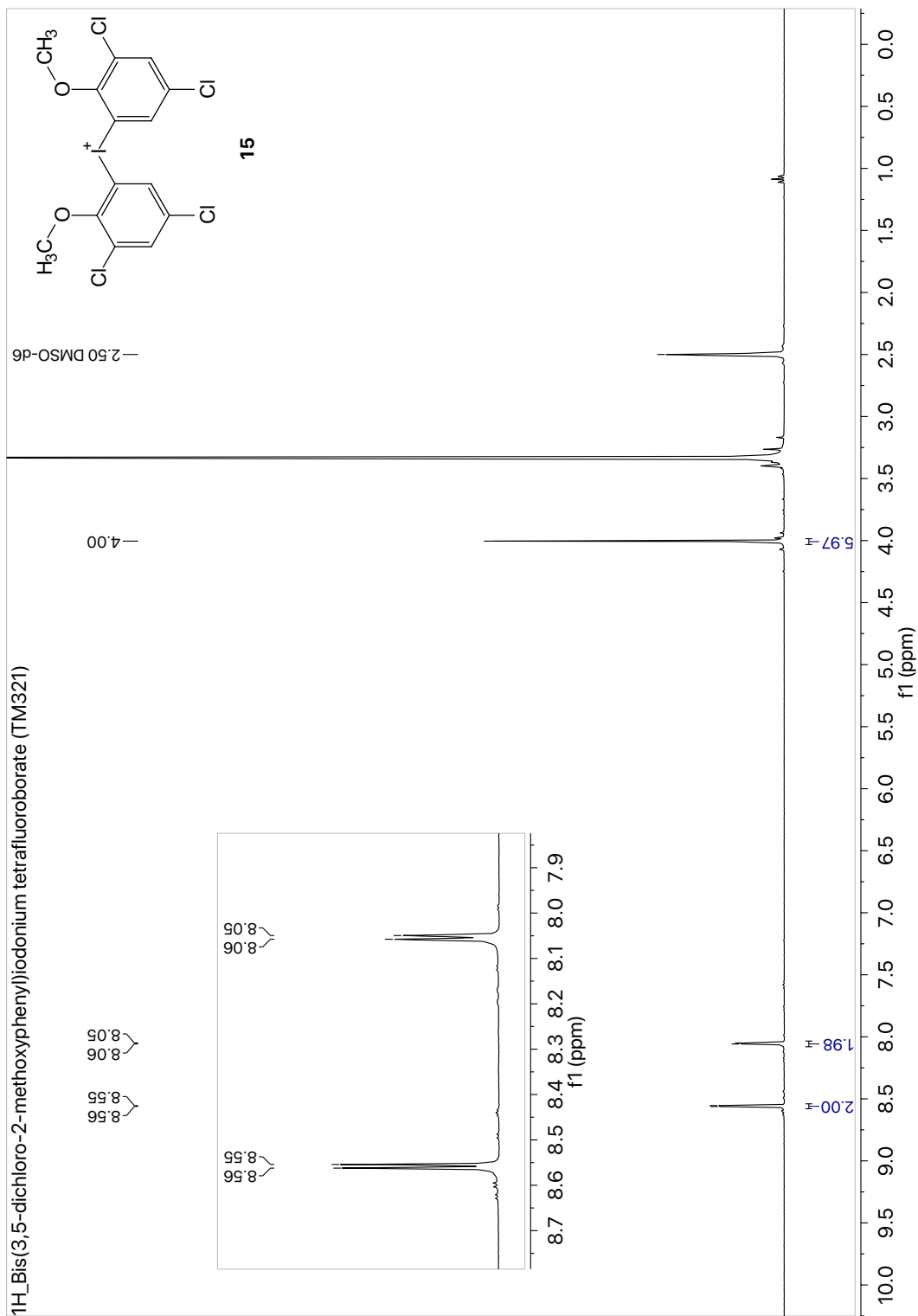


Fig. S35 ¹H-NMR spectrum of bis(3,5-dichloro-2-methoxyphenyl)iodonium tetrafluoroborate (**15**).

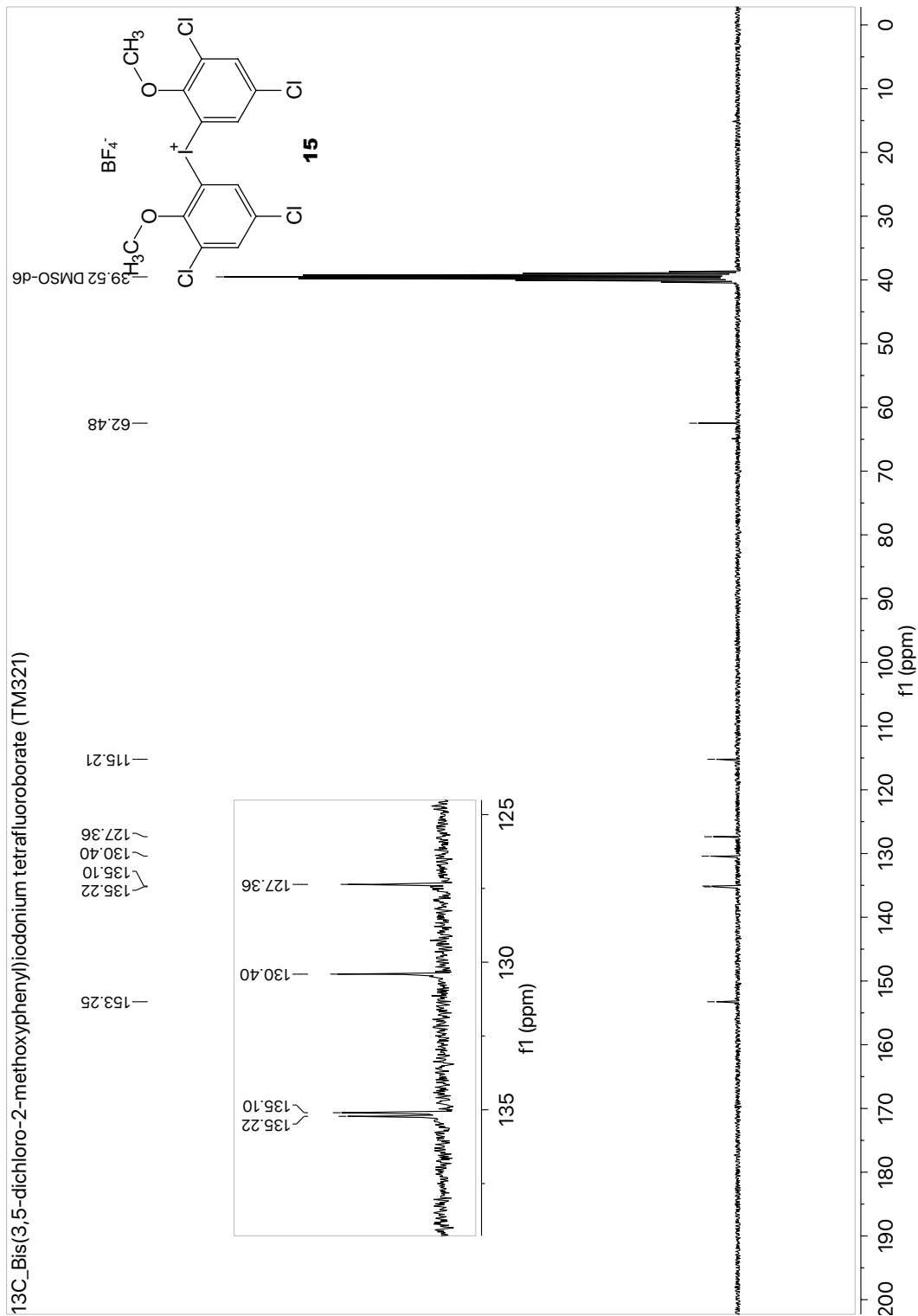


Fig. S36 ¹³C-NMR spectrum of bis(3,5-dichloro-2-methoxyphenyl)iodonium tetrafluoroborate (**15**).

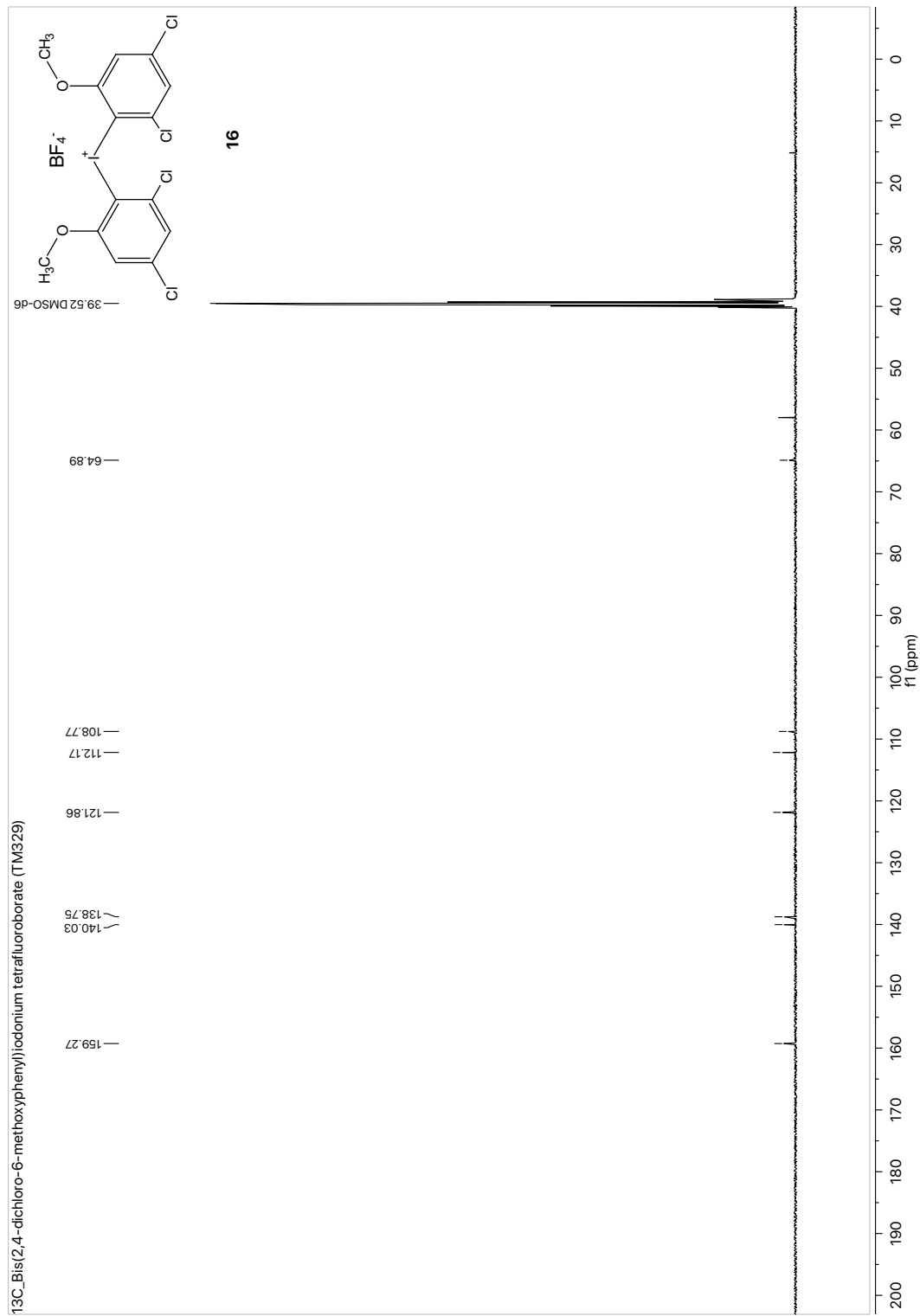


Fig. S38 ¹³C-NMR spectrum of bis(2,4-dichloro-6-methoxyphenyl)iodonium tetrafluoroborate (**16**).

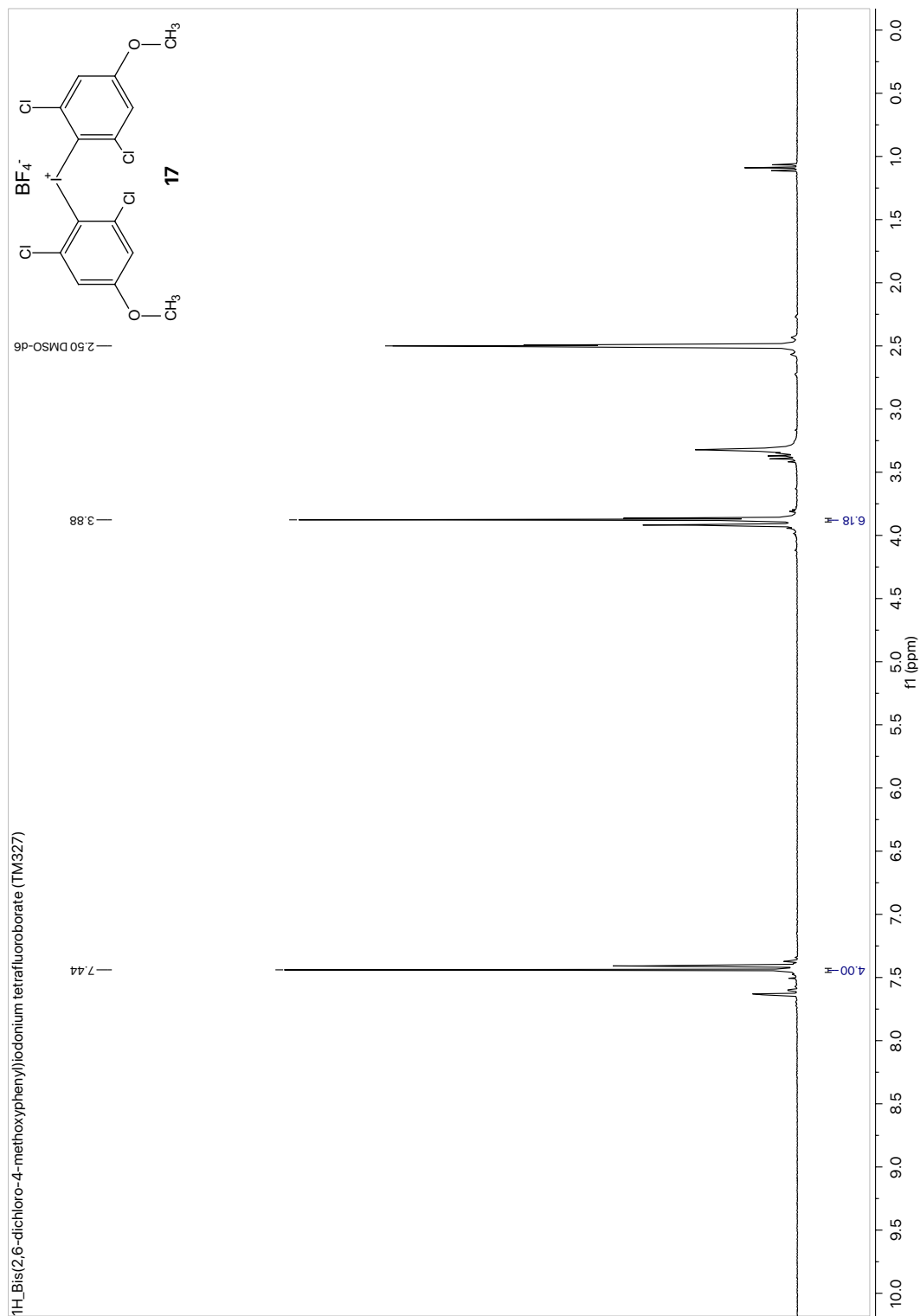


Fig. S39 ¹H-NMR spectrum of bis(2,6-dichloro-4-methoxyphenyl)iodonium tetrafluoroborate (**17**).

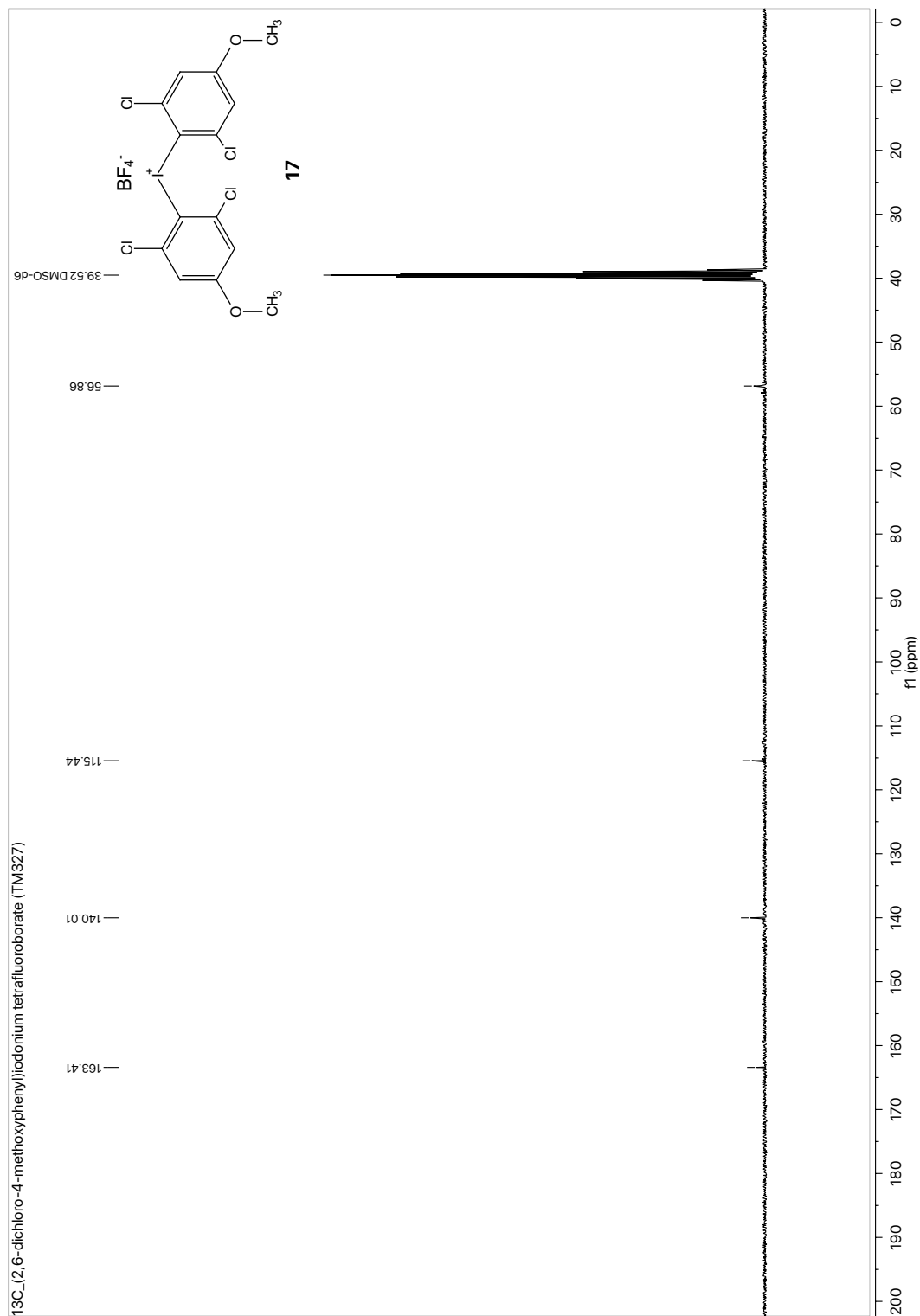


Fig. S40 ¹³C-NMR spectrum of bis(2,6-dichloro-4-methoxyphenyl)iodonium tetrafluoroborate (**17**).

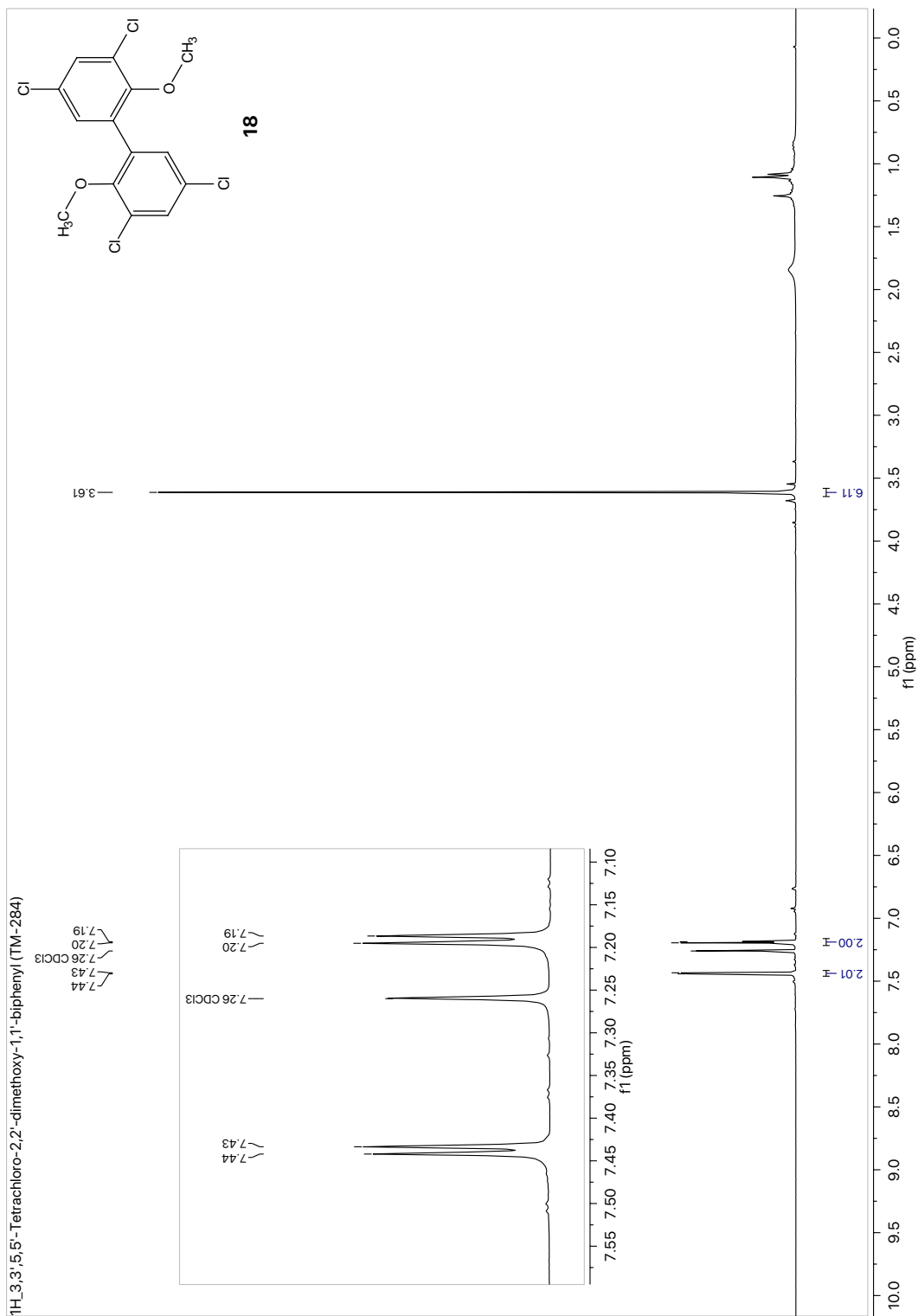


Fig. S41 ¹H-NMR spectrum of 3,3',5,5'-tetrachloro-2,2'-dimethoxy-1,1'-biphenyl (**18**).

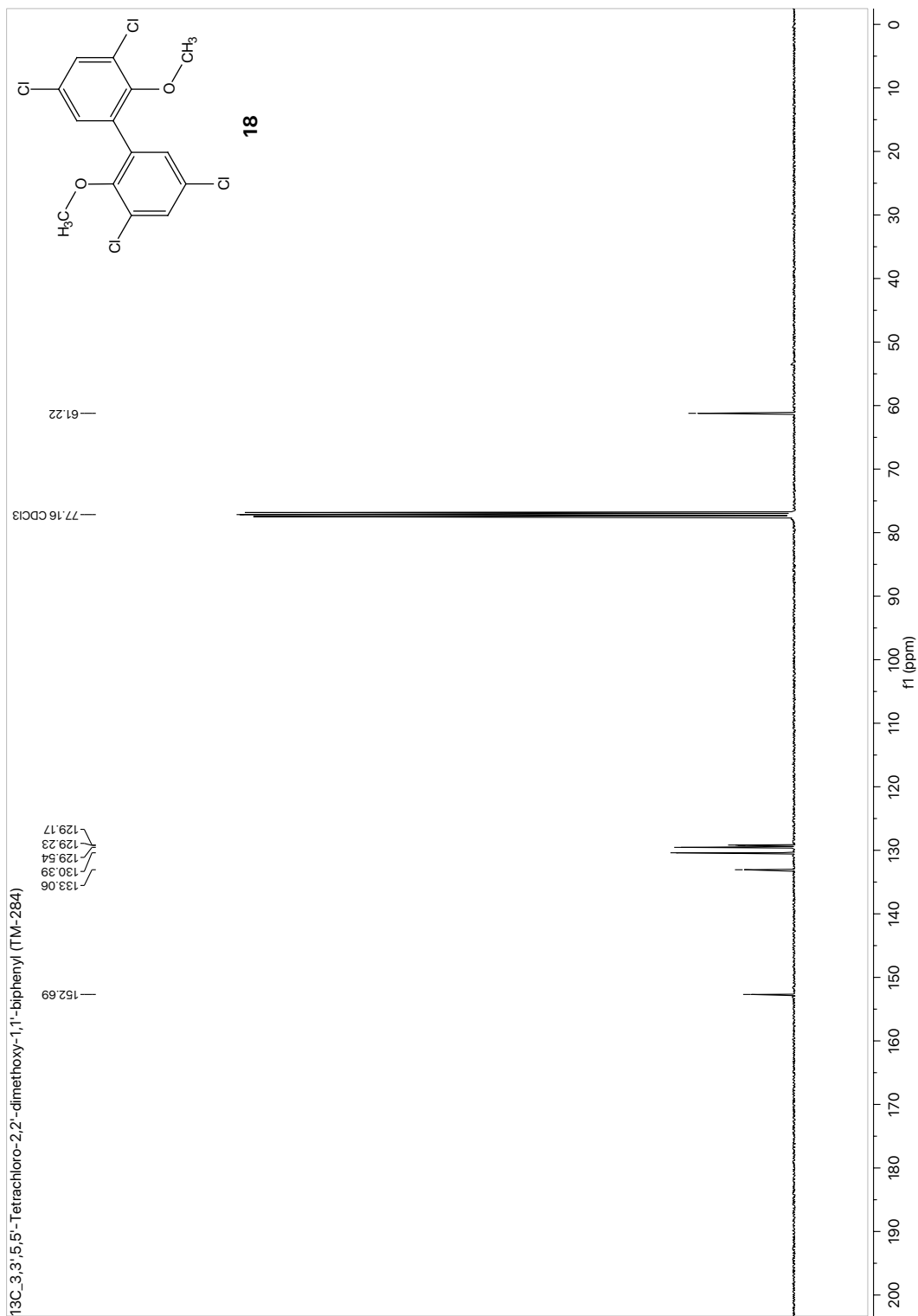


Fig. S42 ¹³C-NMR spectrum of 3,3',5,5'-tetrachloro-2,2'-dimethoxy-1,1'-biphenyl (**18**).

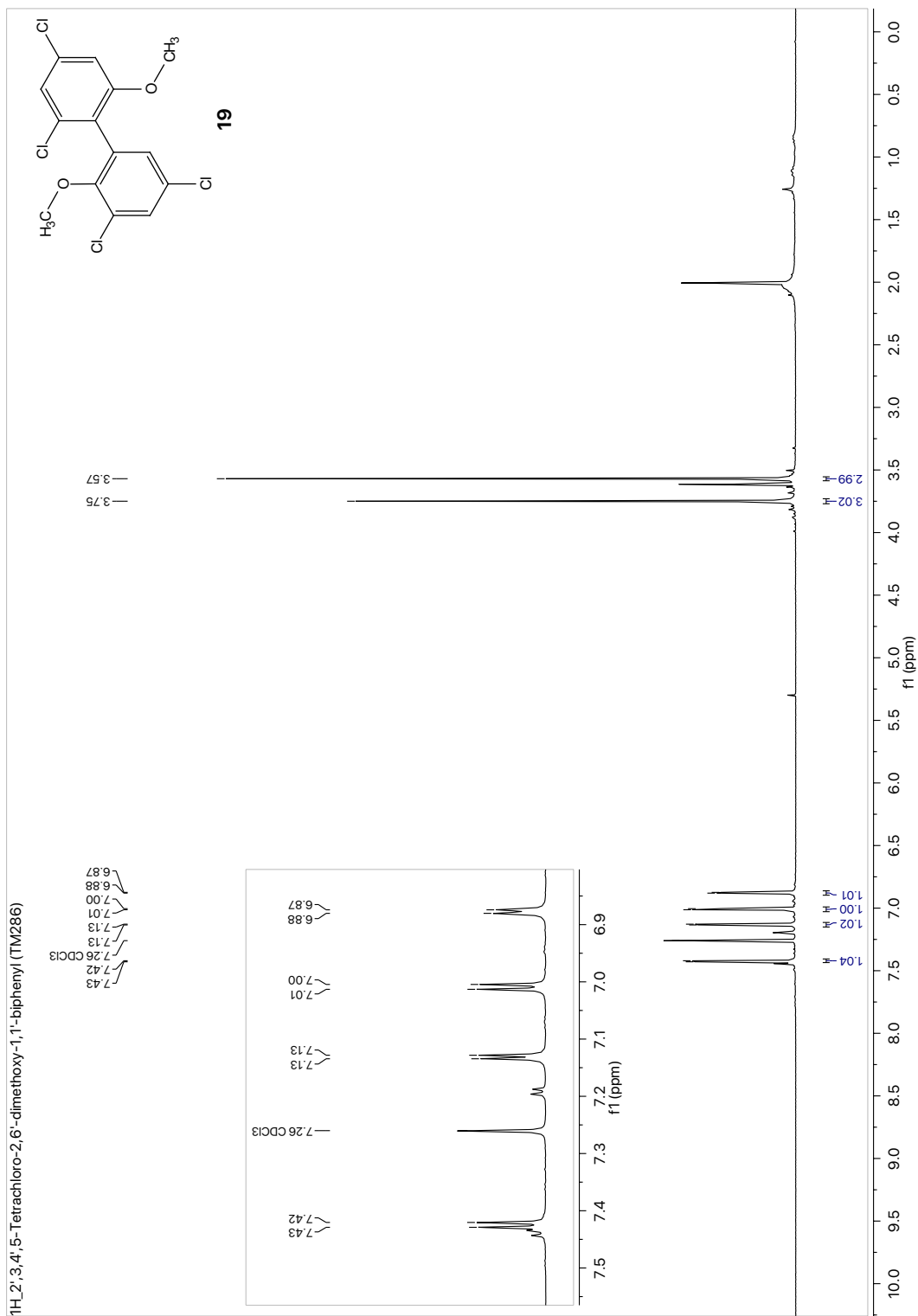


Fig. S43 ¹H-NMR spectrum of 2',3,4',5-tetrachloro-2,6'-dimethoxy-1,1'-biphenyl (19).

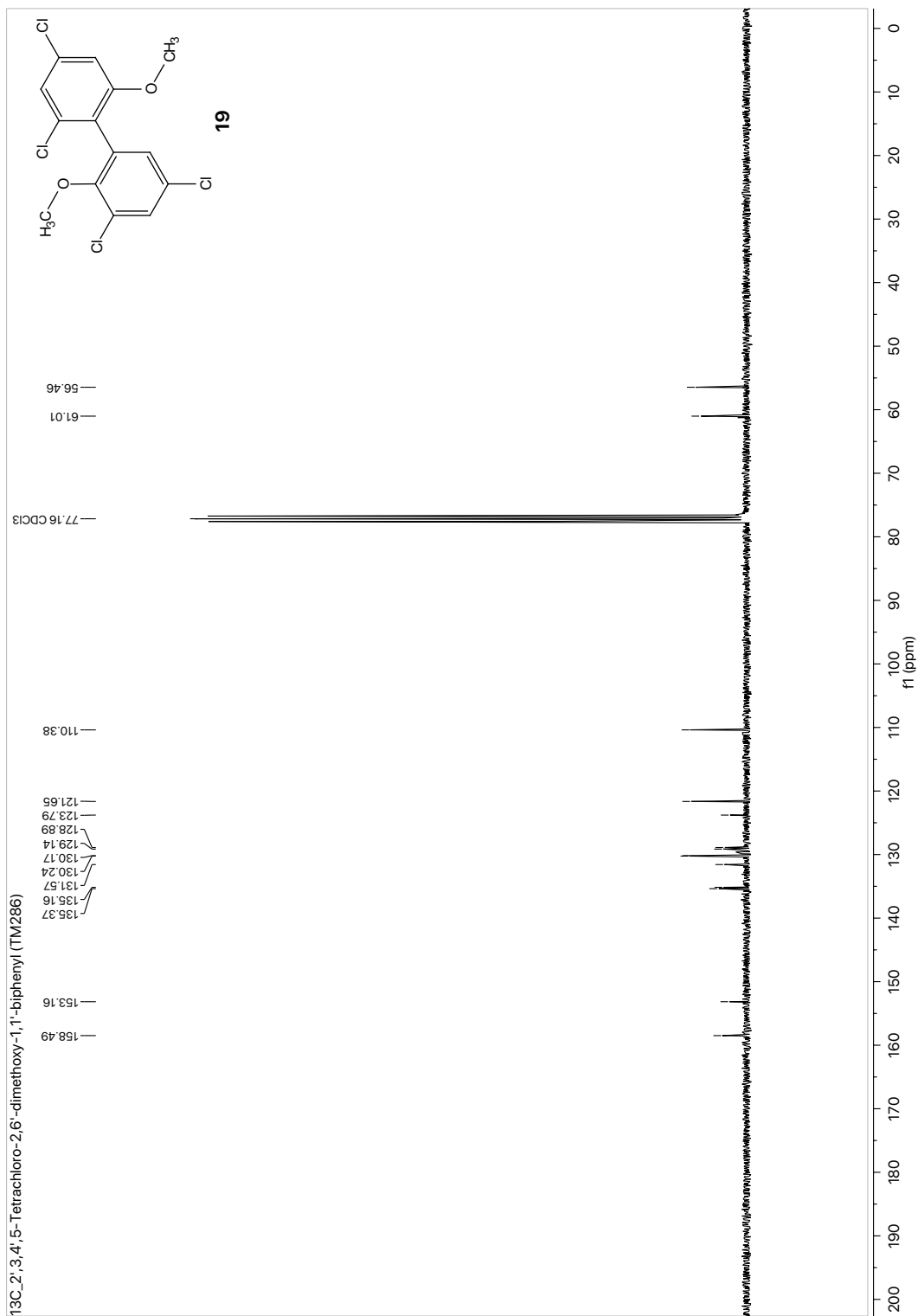


Fig. S44 ¹³C-NMR spectrum of 2',3,4',5-tetrachloro-2,6'-dimethoxy-1,1'-biphenyl (**19**).

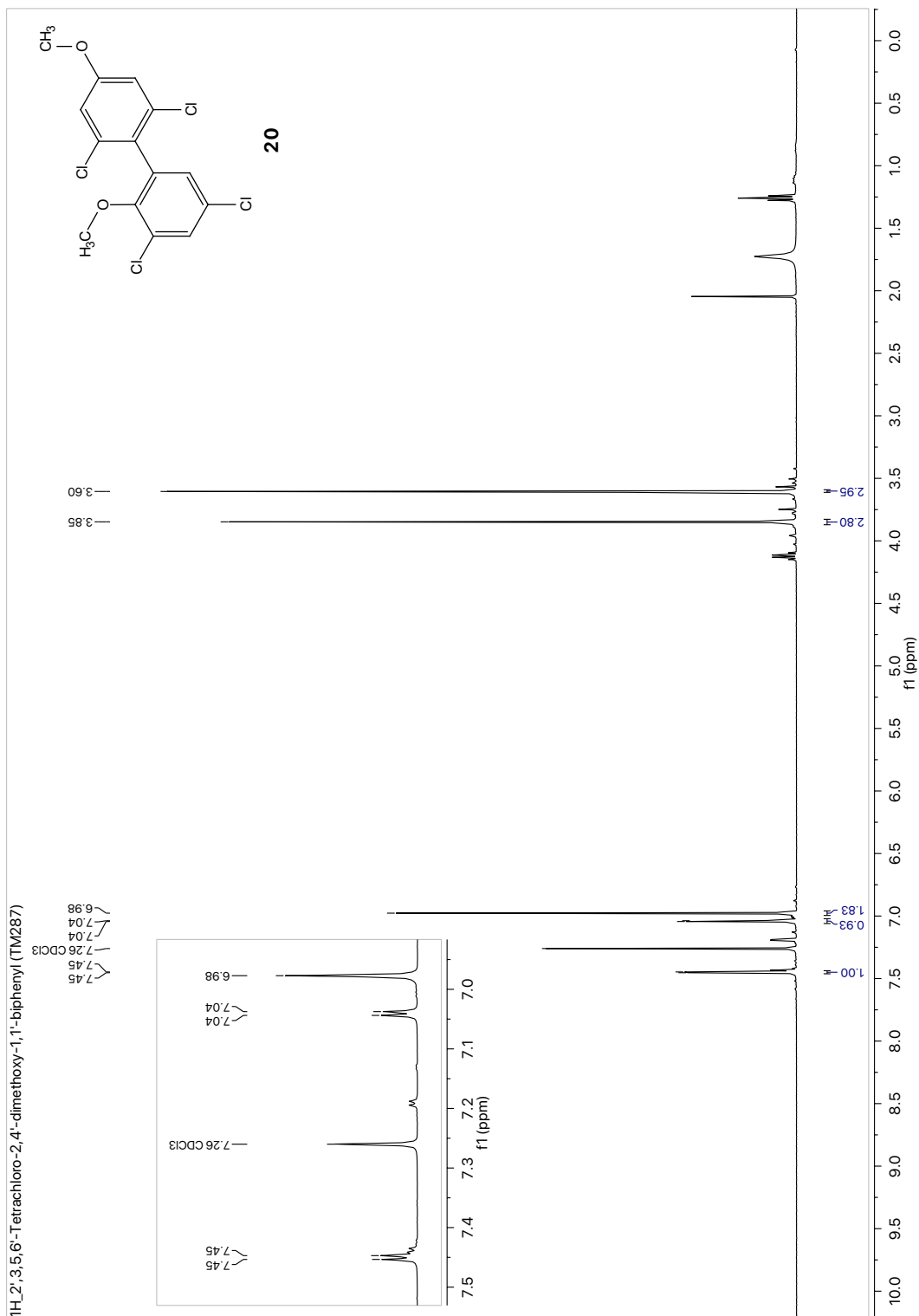


Fig. S45 ¹H-NMR spectrum of 2',3,5,6'-tetrachloro-2,4'-dimethoxy-1,1'-biphenyl (**20**).

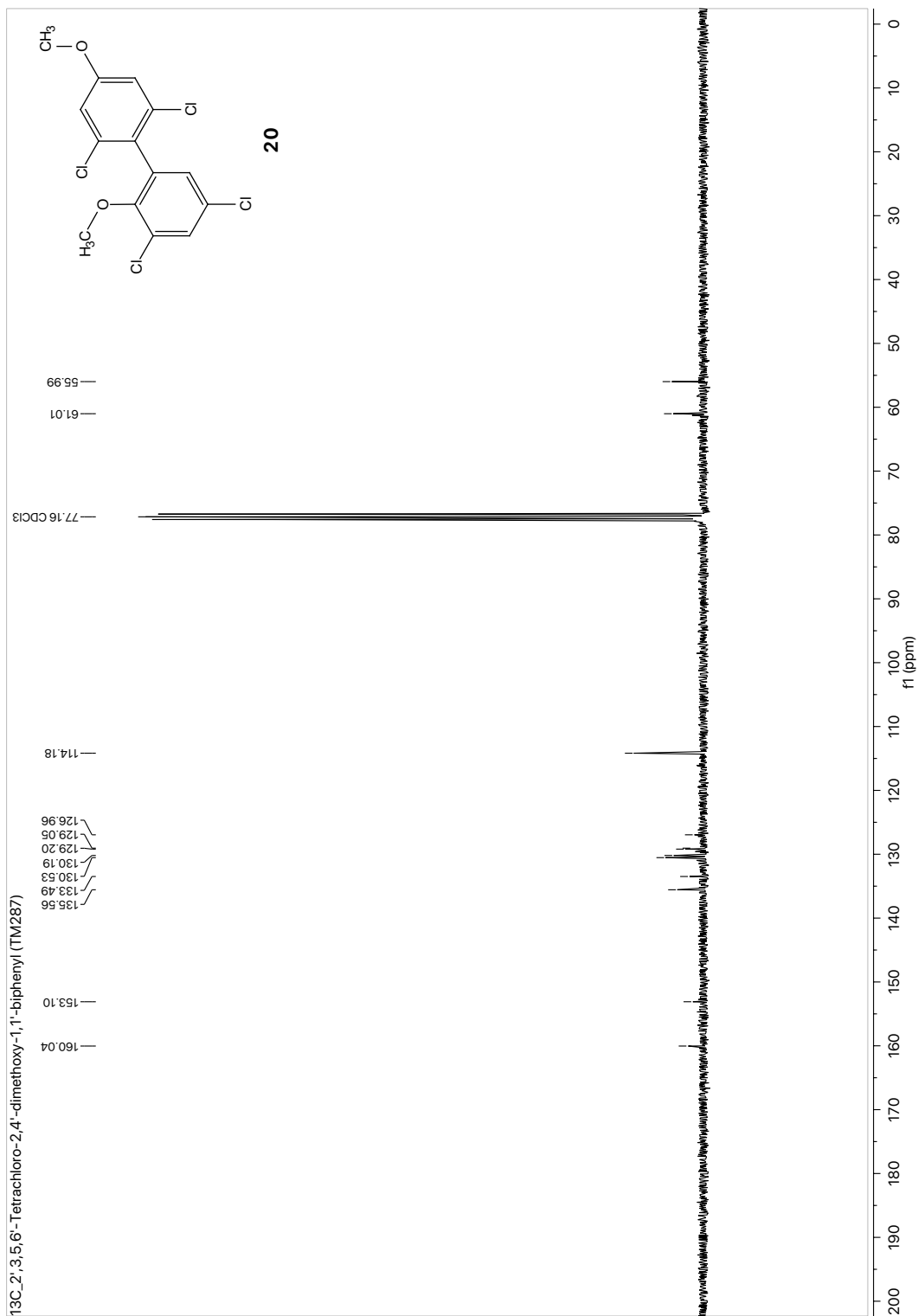


Fig. S46 ¹³C-NMR spectrum of 2',3,5,6'-tetrachloro-2,4'-dimethoxy-1,1'-biphenyl (**20**).

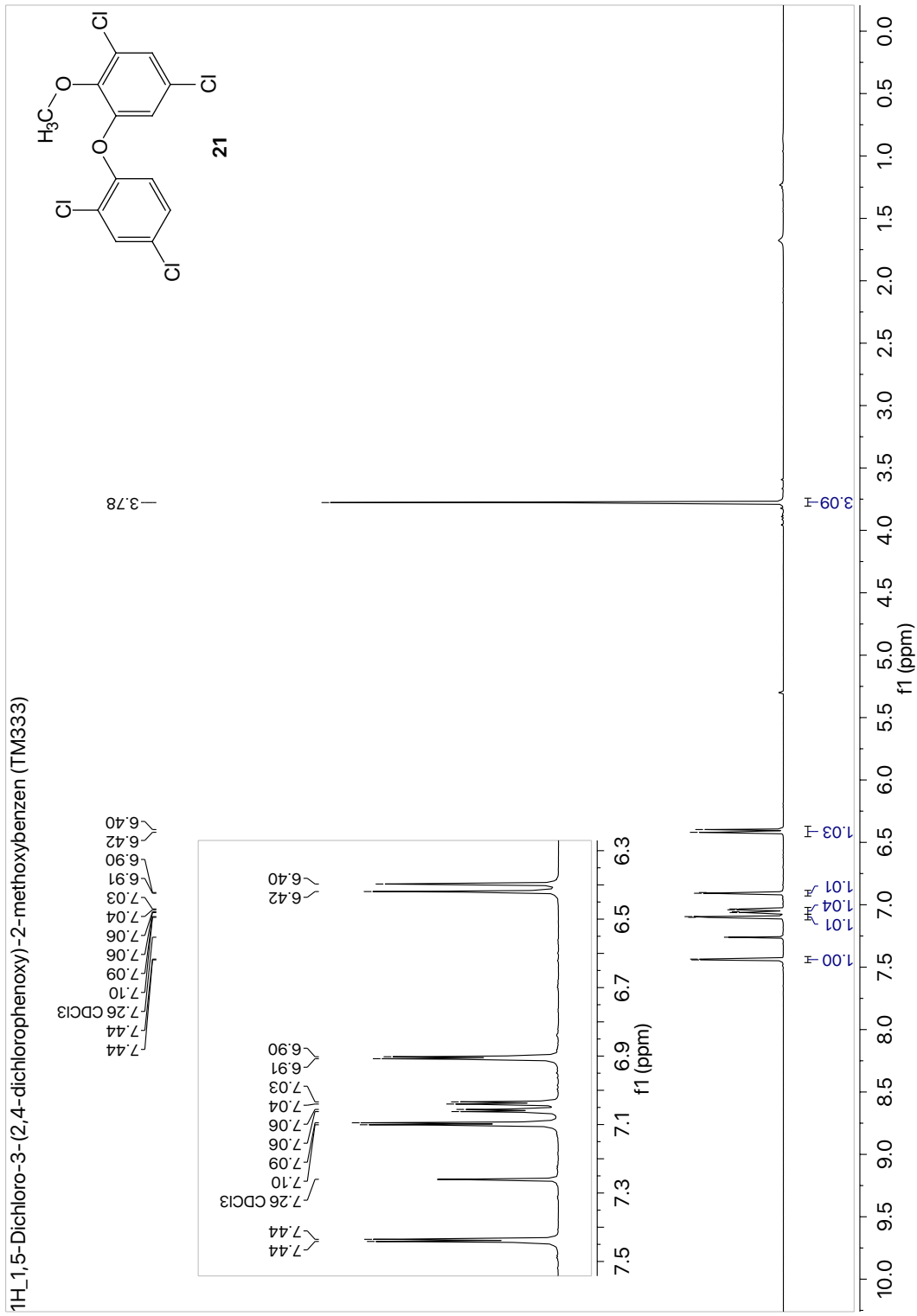


Fig. S47 ¹H-NMR spectrum of 1,5-dichloro-3-(2,4-dichlorophenoxy)-2-methoxybenzene (**21**).

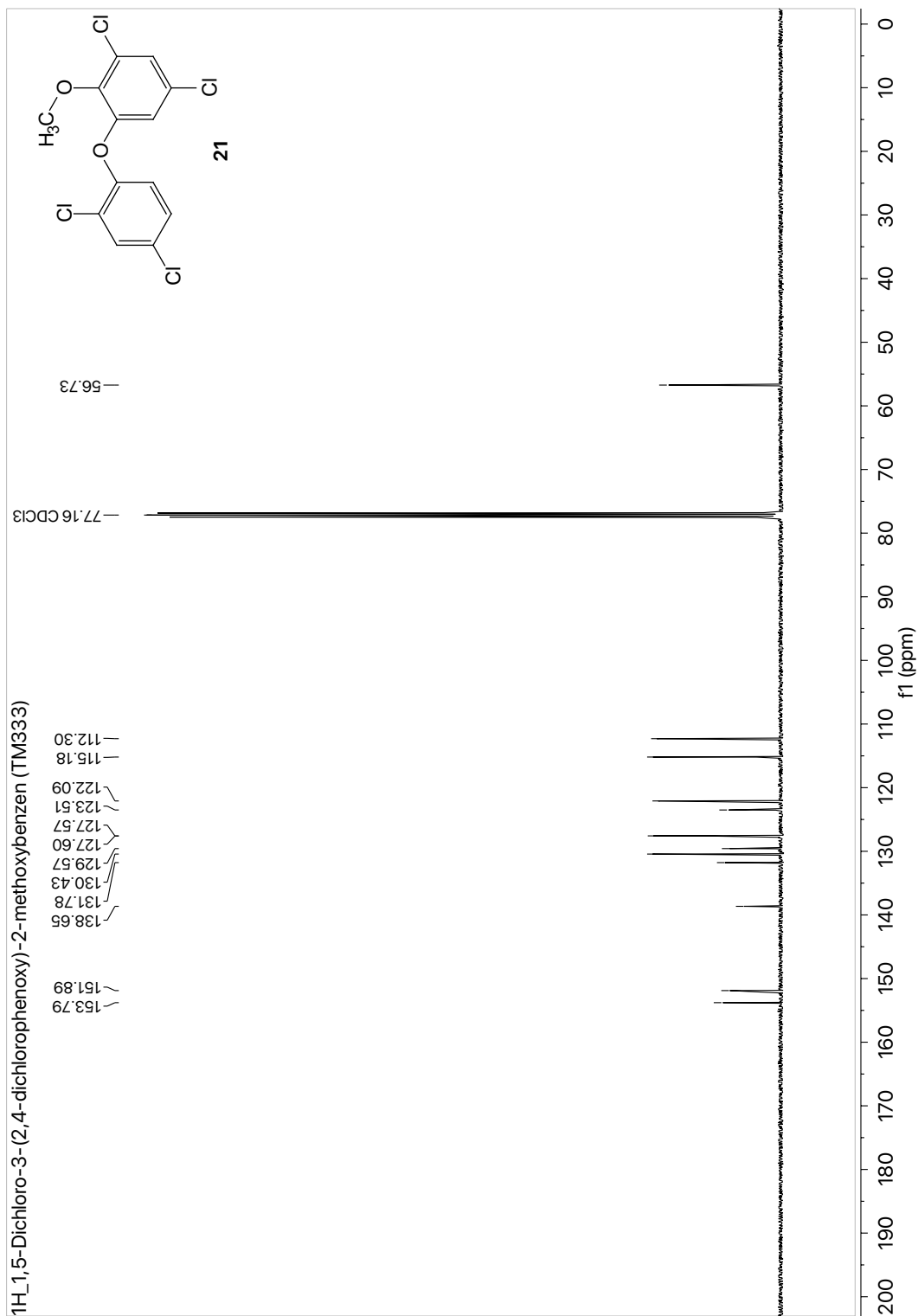


Fig. S48 ¹³C-NMR spectrum of 1,5-dichloro-3-(2,4-dichlorophenoxy)-2-methoxybenzene (**21**).

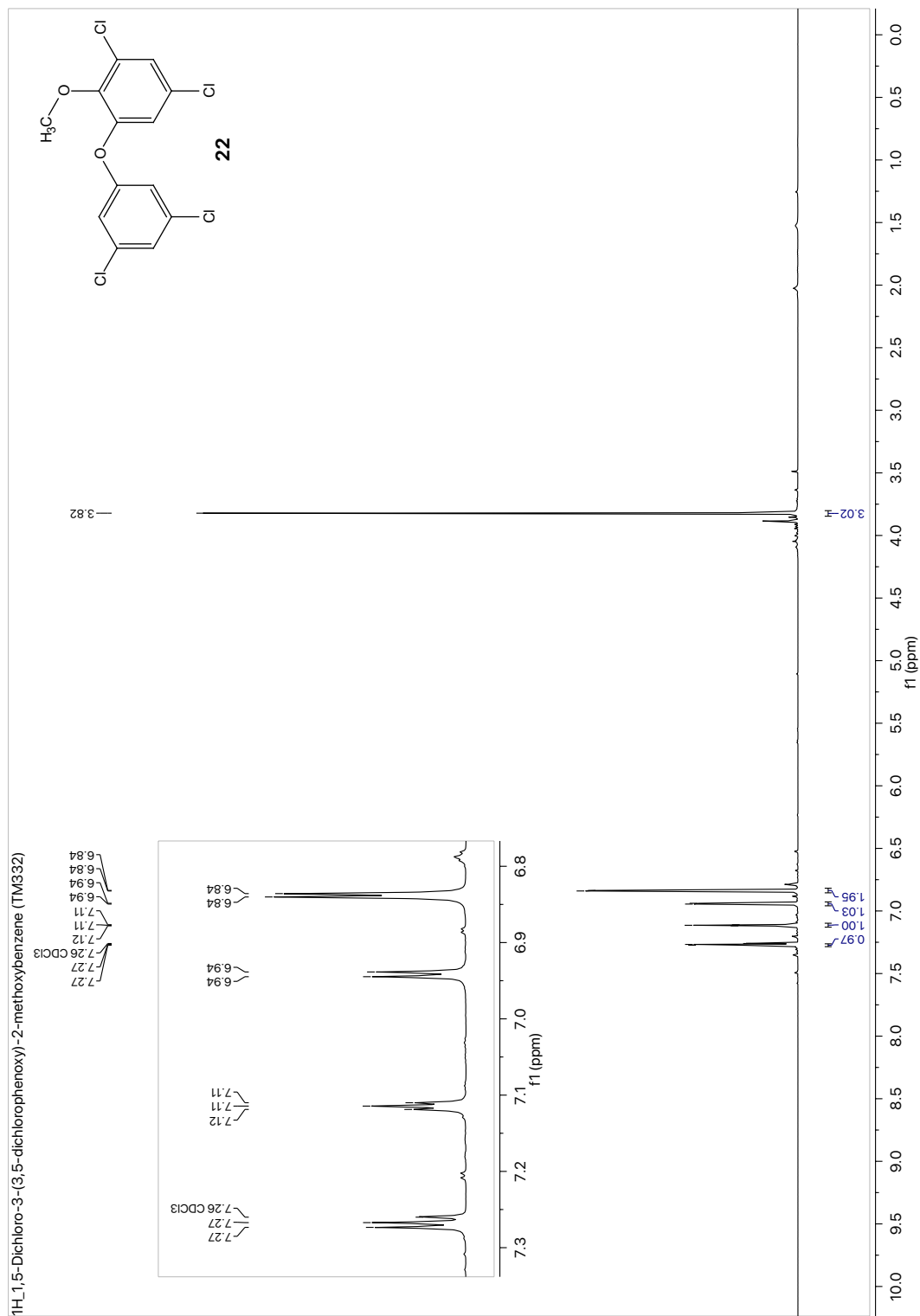


Fig. S49 ¹H-NMR spectrum of 1,5-dichloro-3-(3,5-dichlorophenoxy)-2-methoxybenzene (**22**).

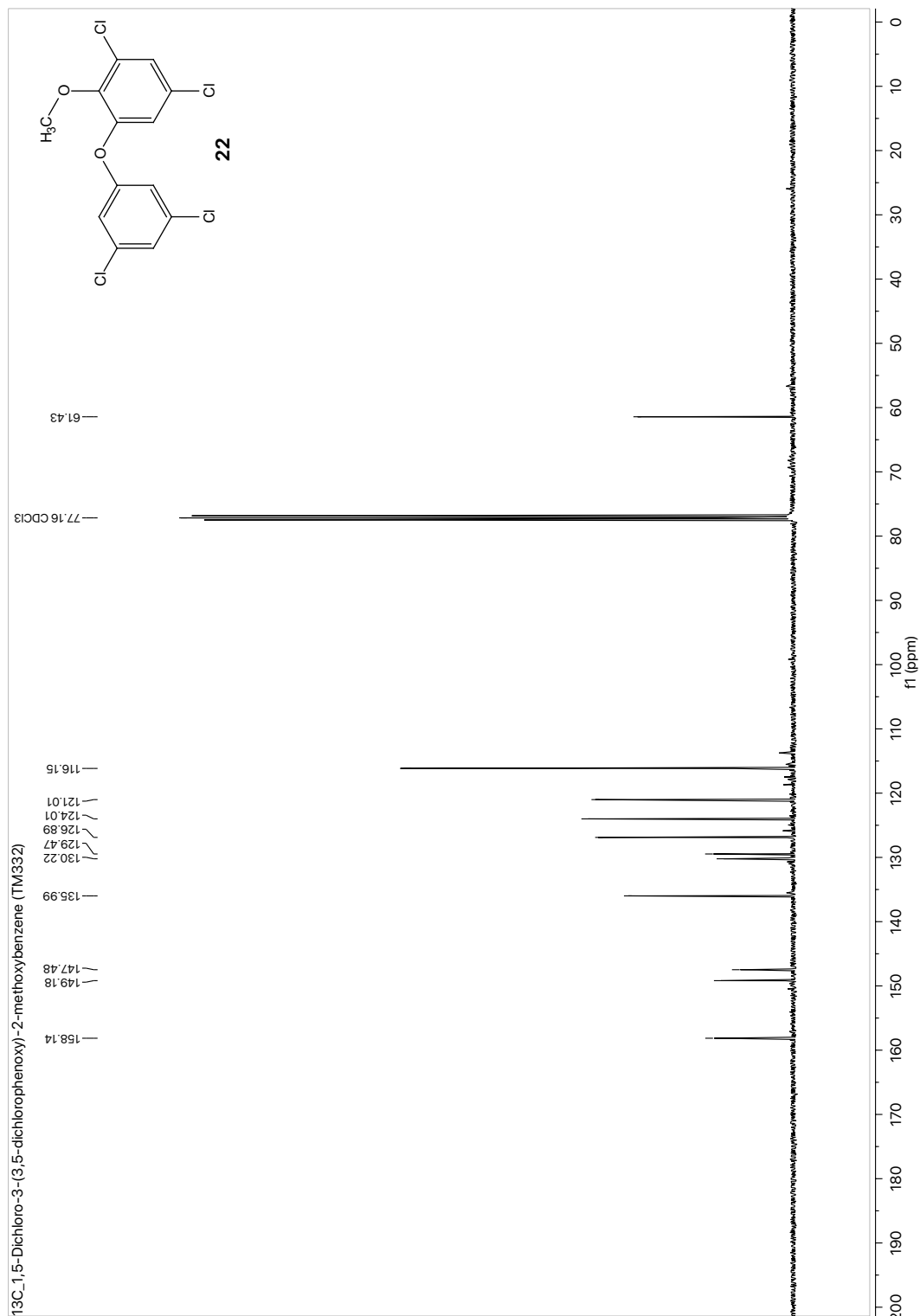


Fig. S50 ¹³C-NMR spectrum of 1,5-dichloro-3-(3,5-dichlorophenoxy)-2-methoxybenzene (**22**).

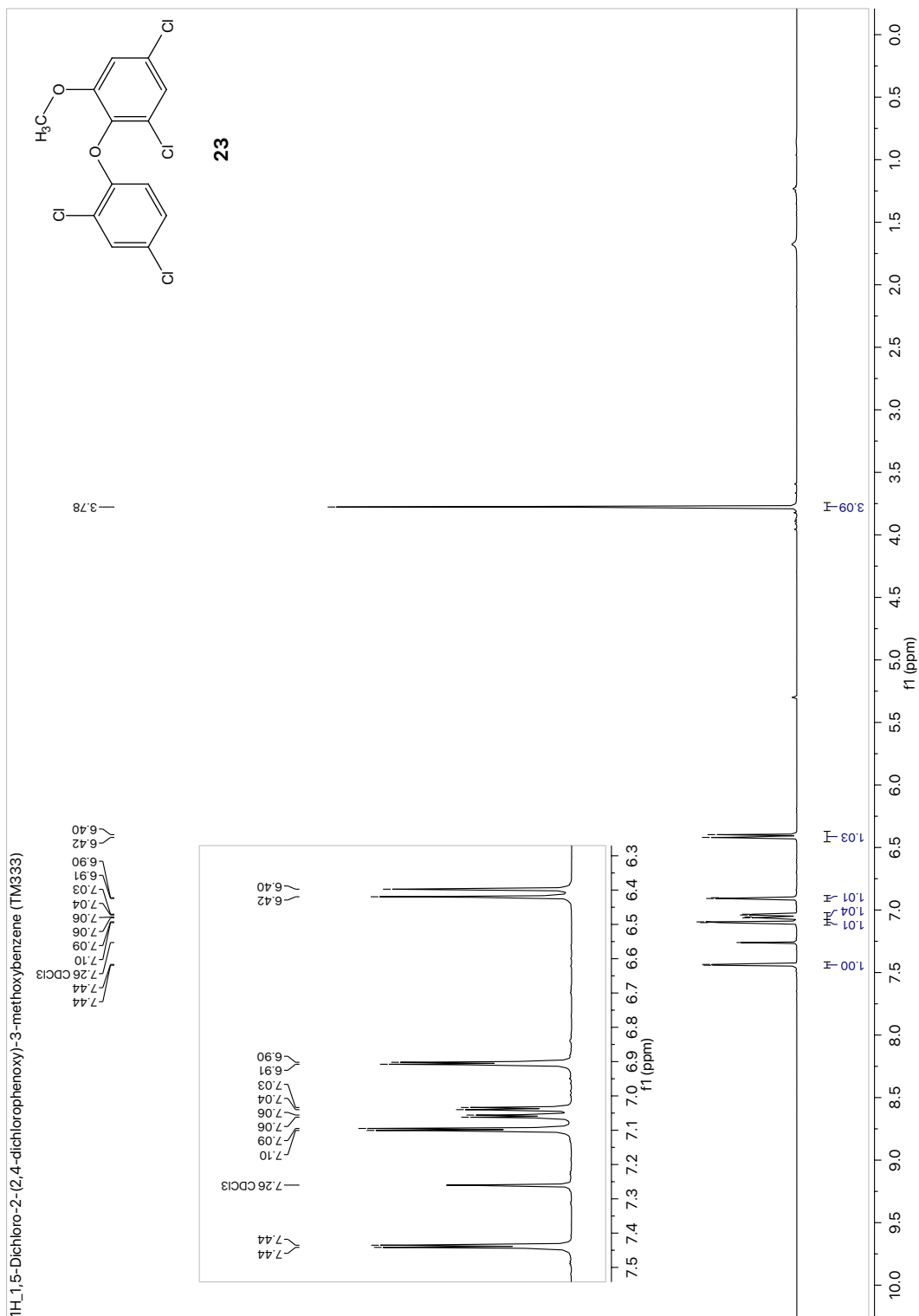


Fig. S51 ¹H-NMR spectrum of 1,5-dichloro-2-(2,4-dichlorophenoxy)-3-methoxybenzene (**23**).

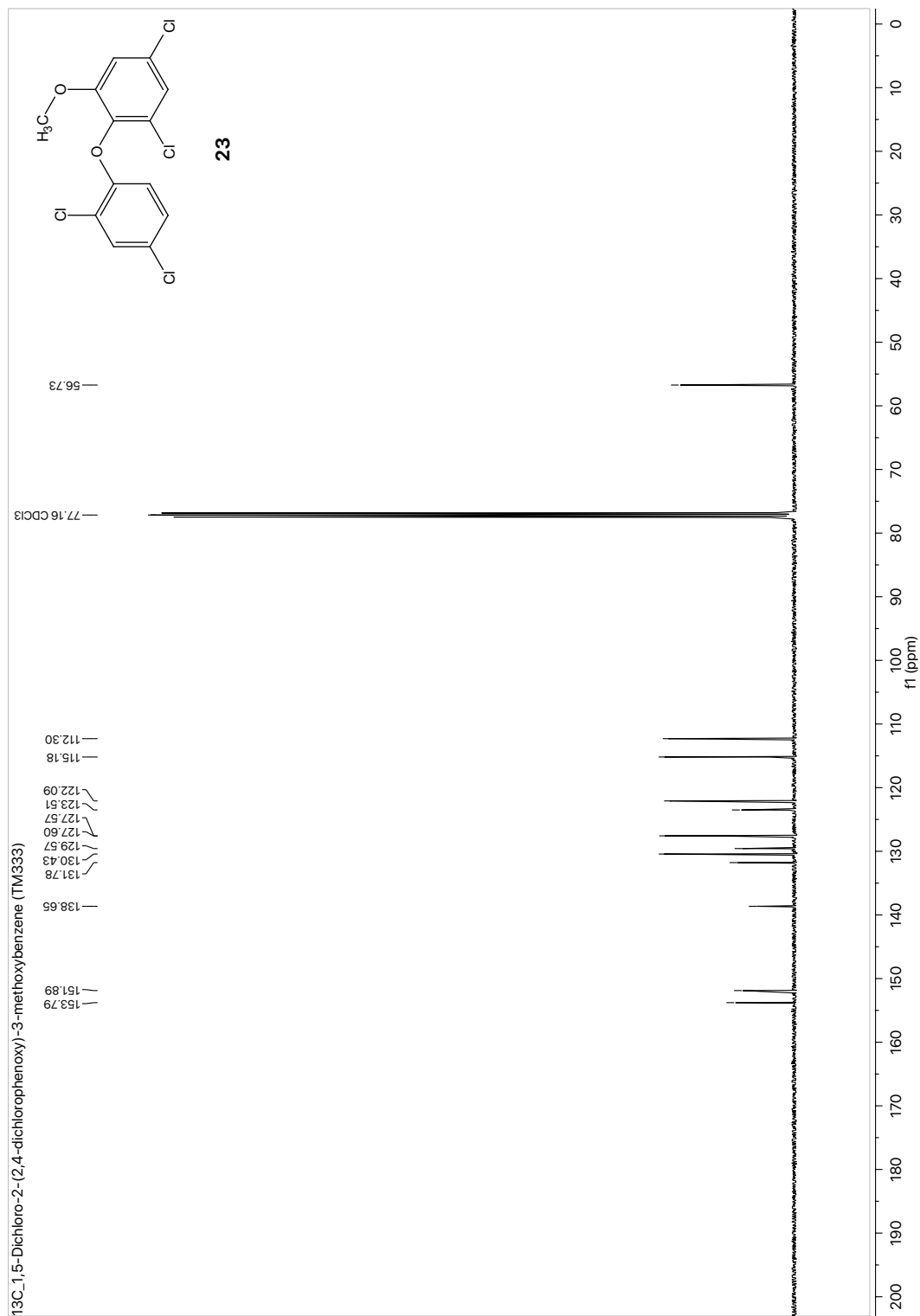


Fig. S52 ¹³C-NMR spectrum of 1,5-dichloro-2-(2,4-dichlorophenoxy)-3-methoxybenzene (**23**).

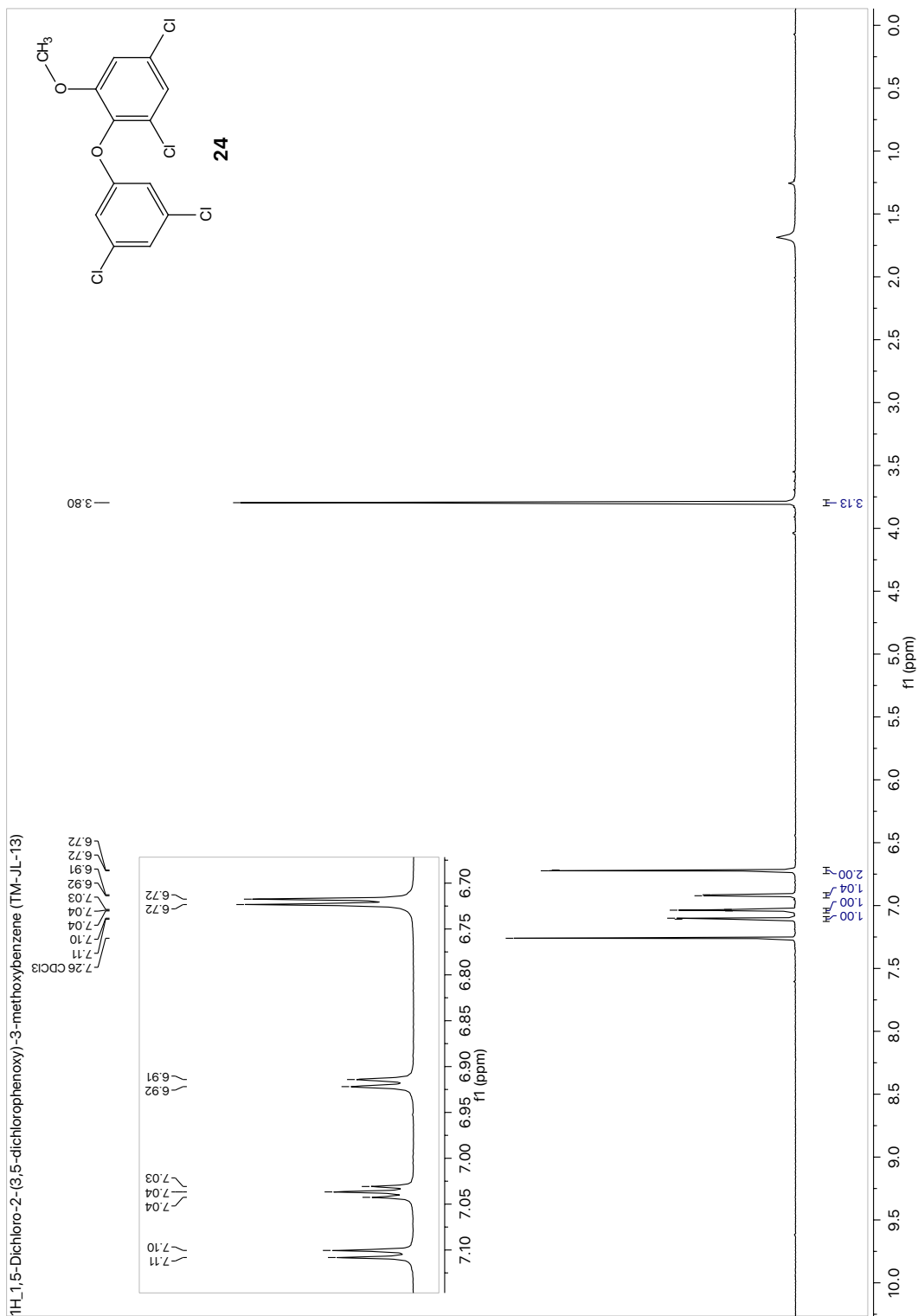


Fig. S53 ¹H-NMR spectrum of 1,5-dichloro-2-(3,5-dichlorophenoxy)-3-methoxybenzene (**24**).

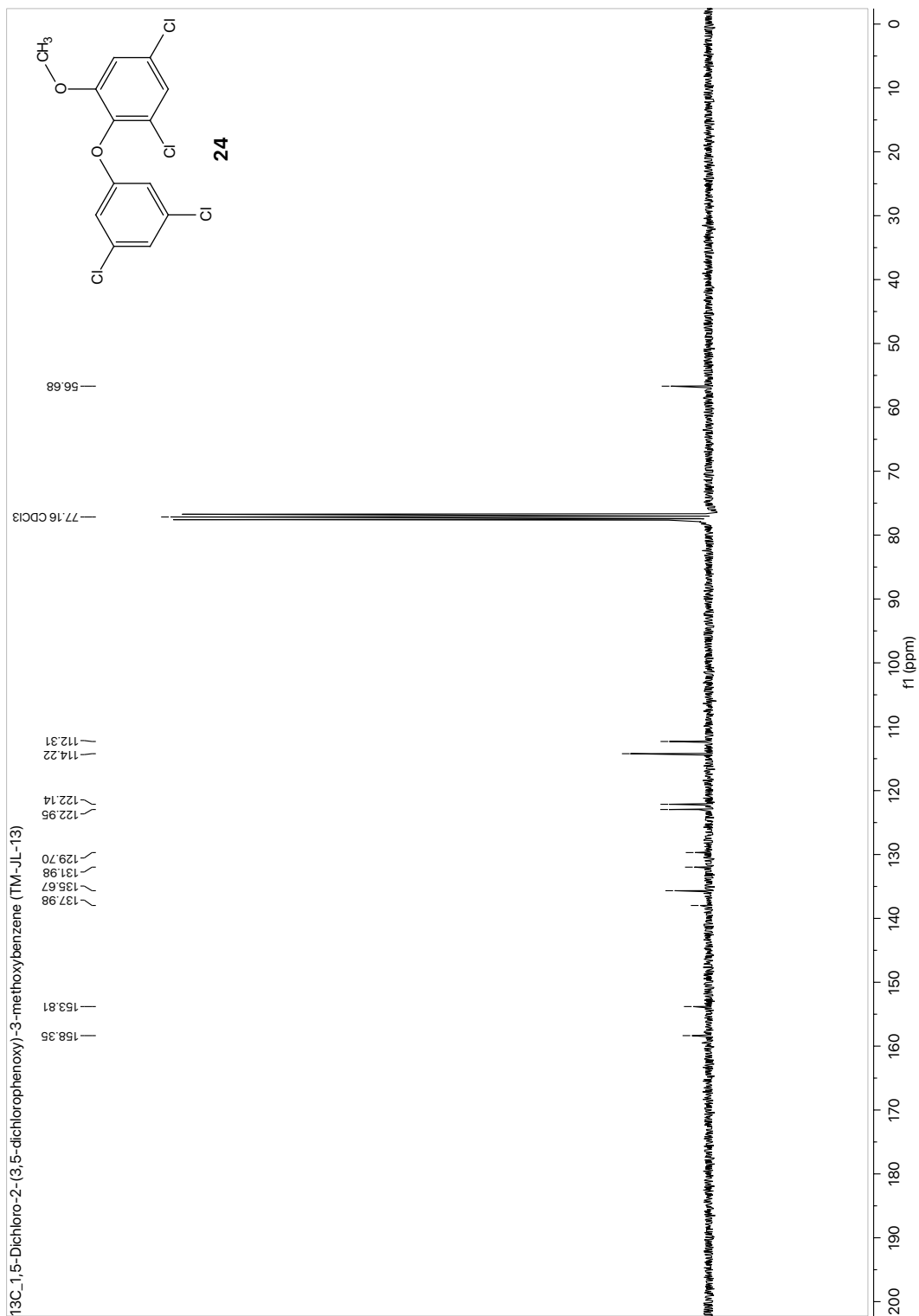


Fig. S54 ¹³C-NMR spectrum of 1,5-dichloro-2-(3,5-dichlorophenoxy)-3-methoxybenzene (**24**).

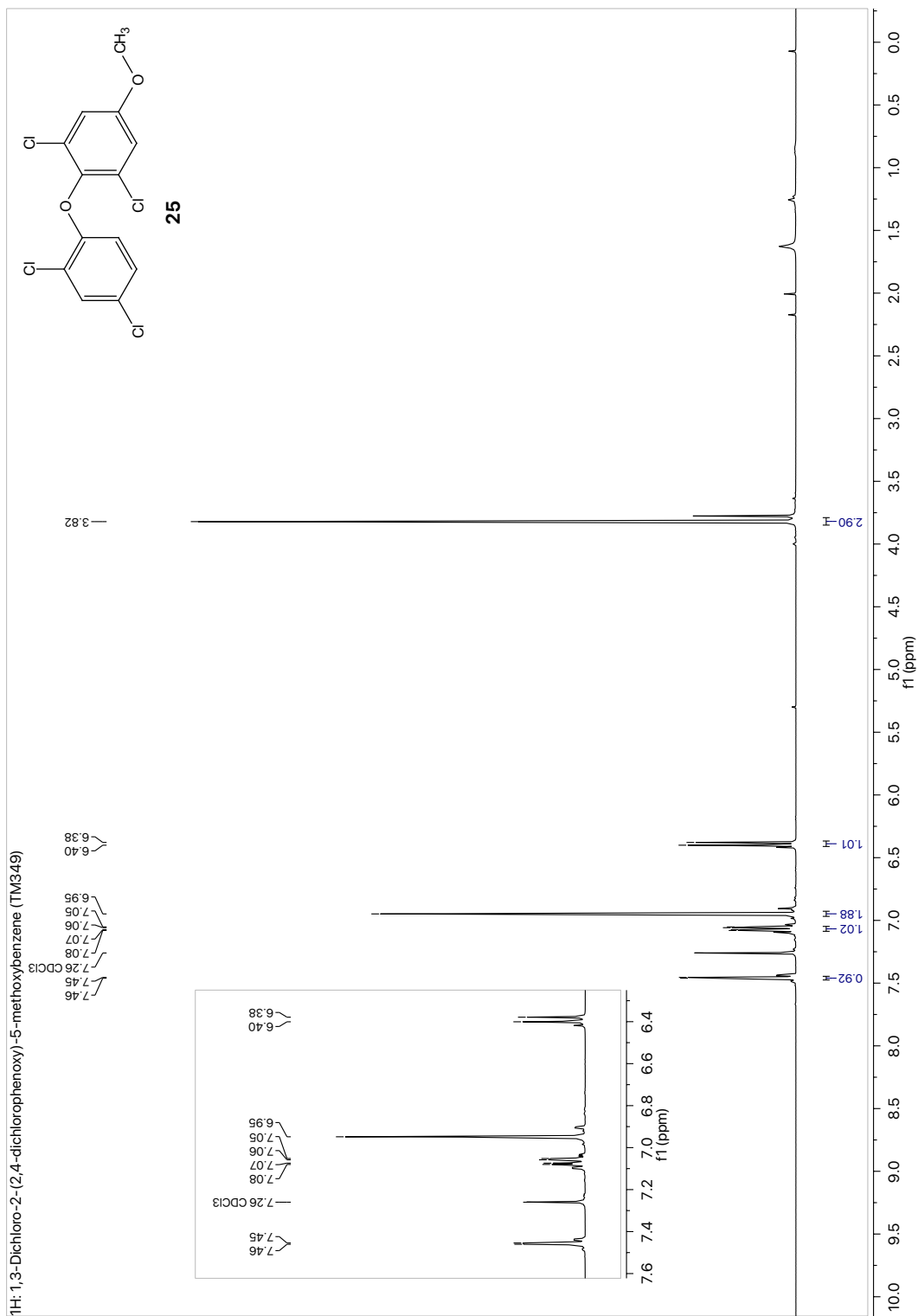


Fig. S55 ¹H-NMR spectrum of 1,3-dichloro-2-(2,4-dichlorophenoxy)-5-methoxybenzene (**25**).

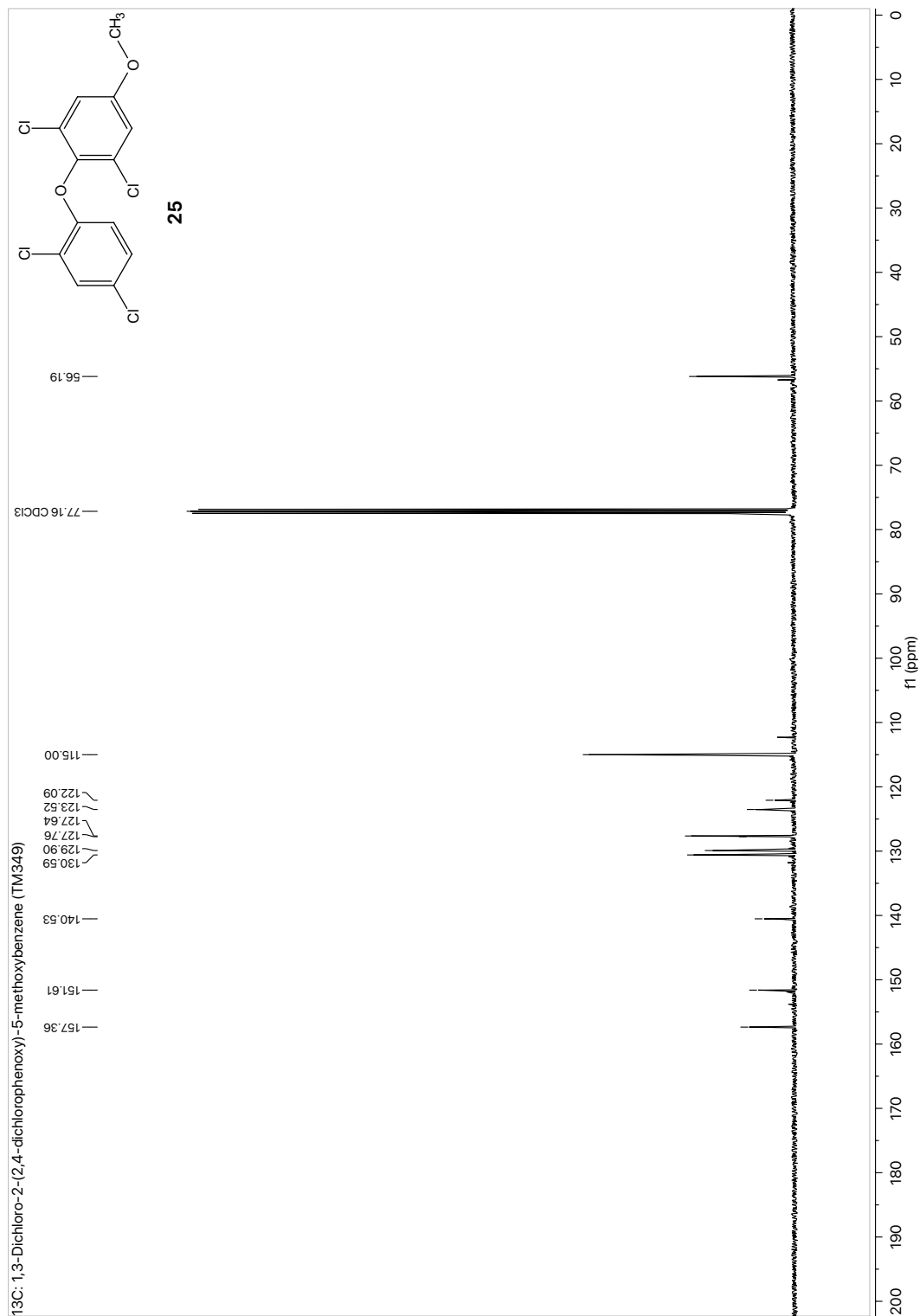


Fig. S56 ¹³C-NMR spectrum of 1,3-dichloro-2-(2,4-dichlorophenoxy)-5-methoxybenzene (**25**).

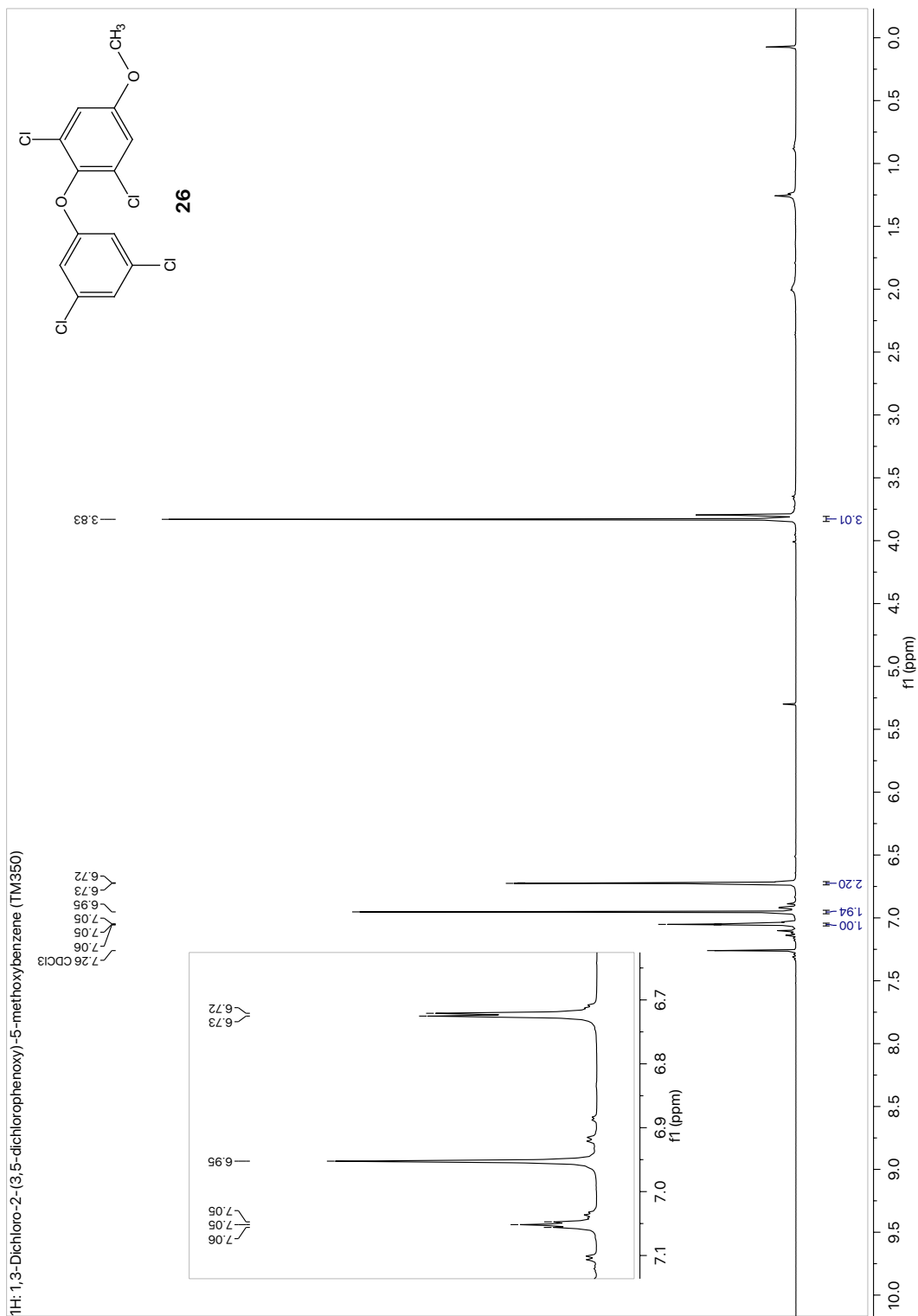


Fig. S57 ¹H-NMR spectrum of 1,3-dichloro-2-(3,5-dichlorophenoxy)-5-methoxybenzene (**26**).

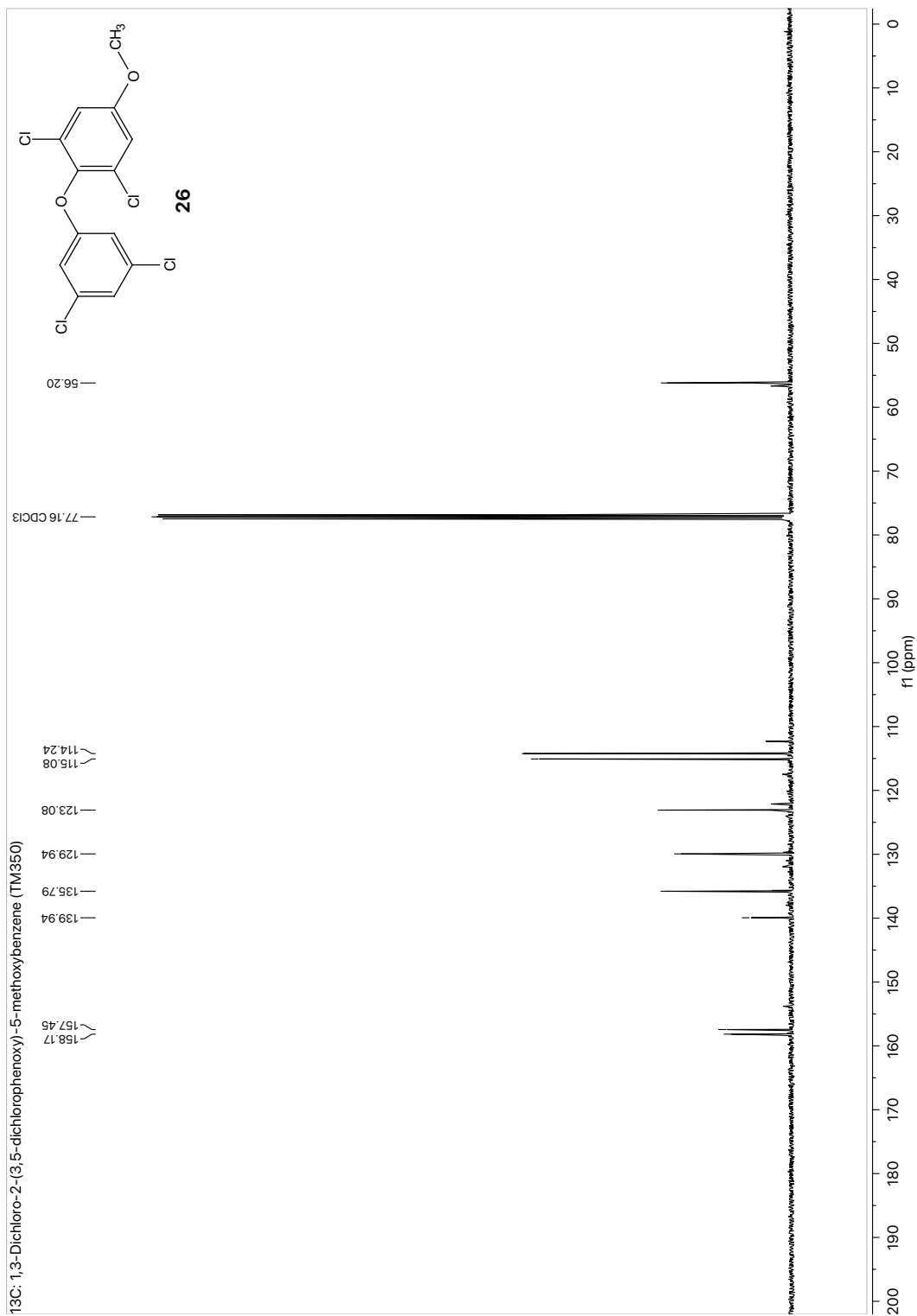


Fig. S58 ¹³C-NMR spectrum of 1,3-dichloro-2-(3,5-dichlorophenoxy)-5-methoxybenzene (**26**).

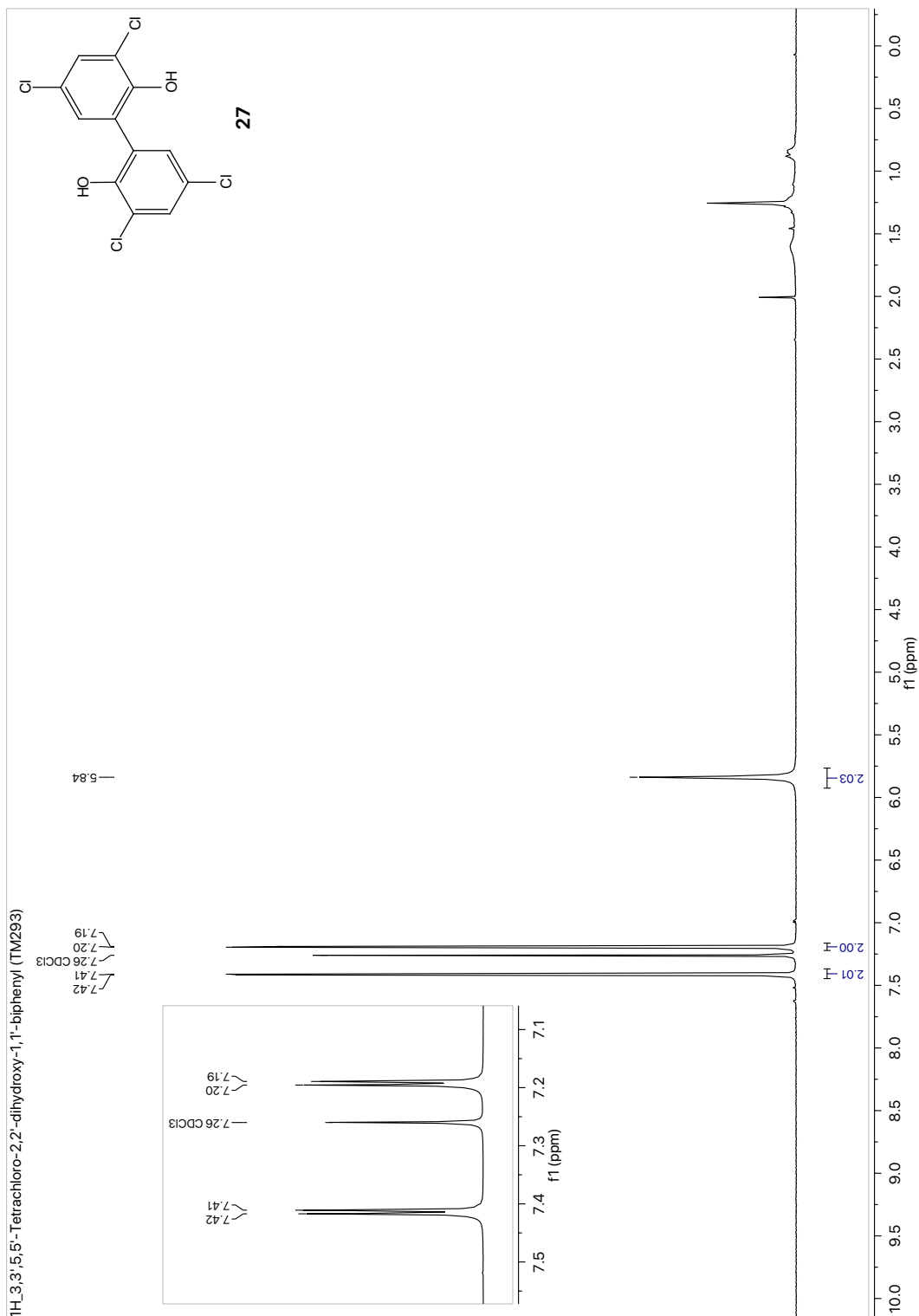


Fig. S59 ¹H-NMR spectrum of 3,3',5,5'-tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (**27**). Measured in CDCl₃.

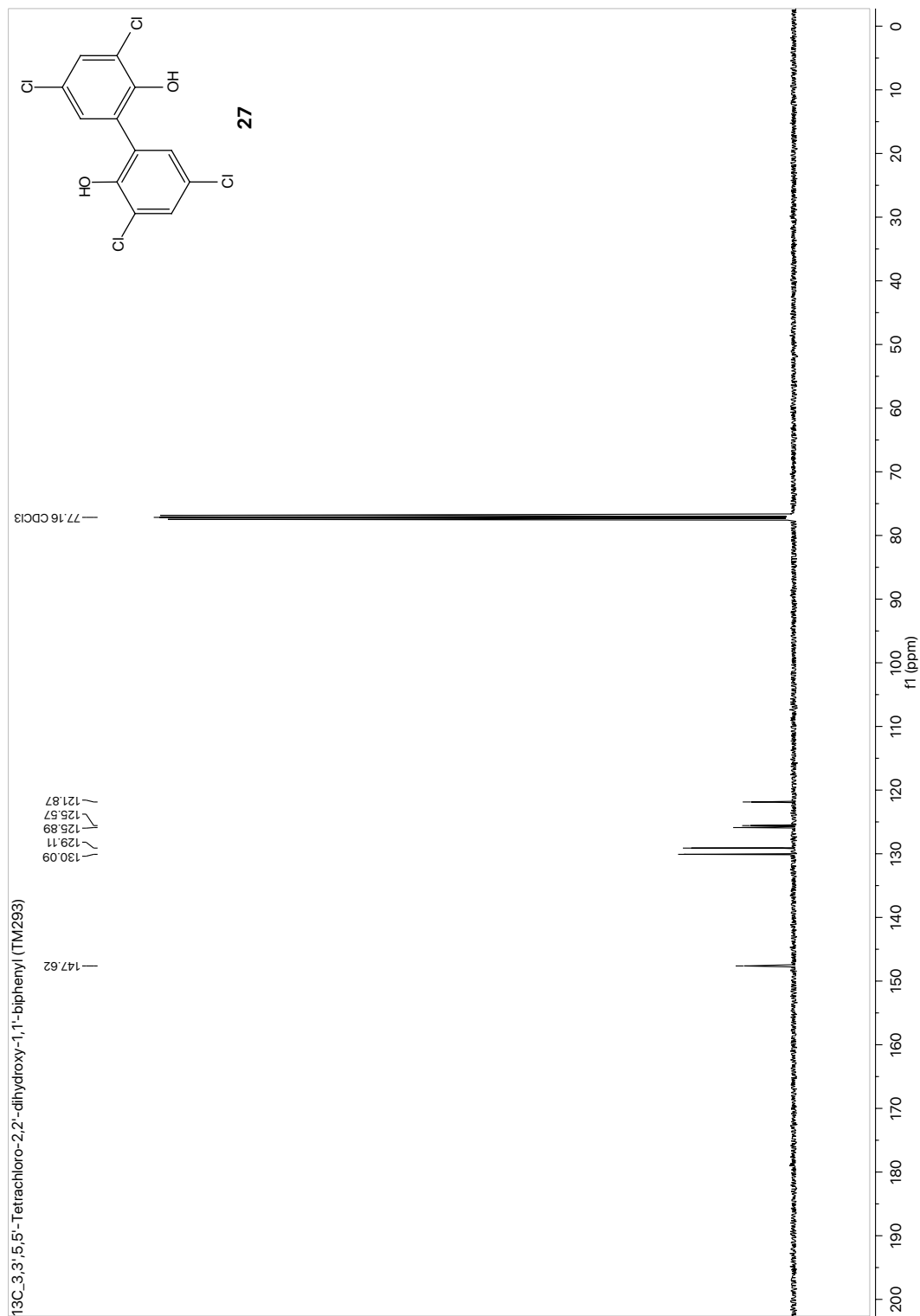


Fig. S60 ¹³C-NMR spectrum of 3,3',5,5'-tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (**27**). Measured in CDCl₃.

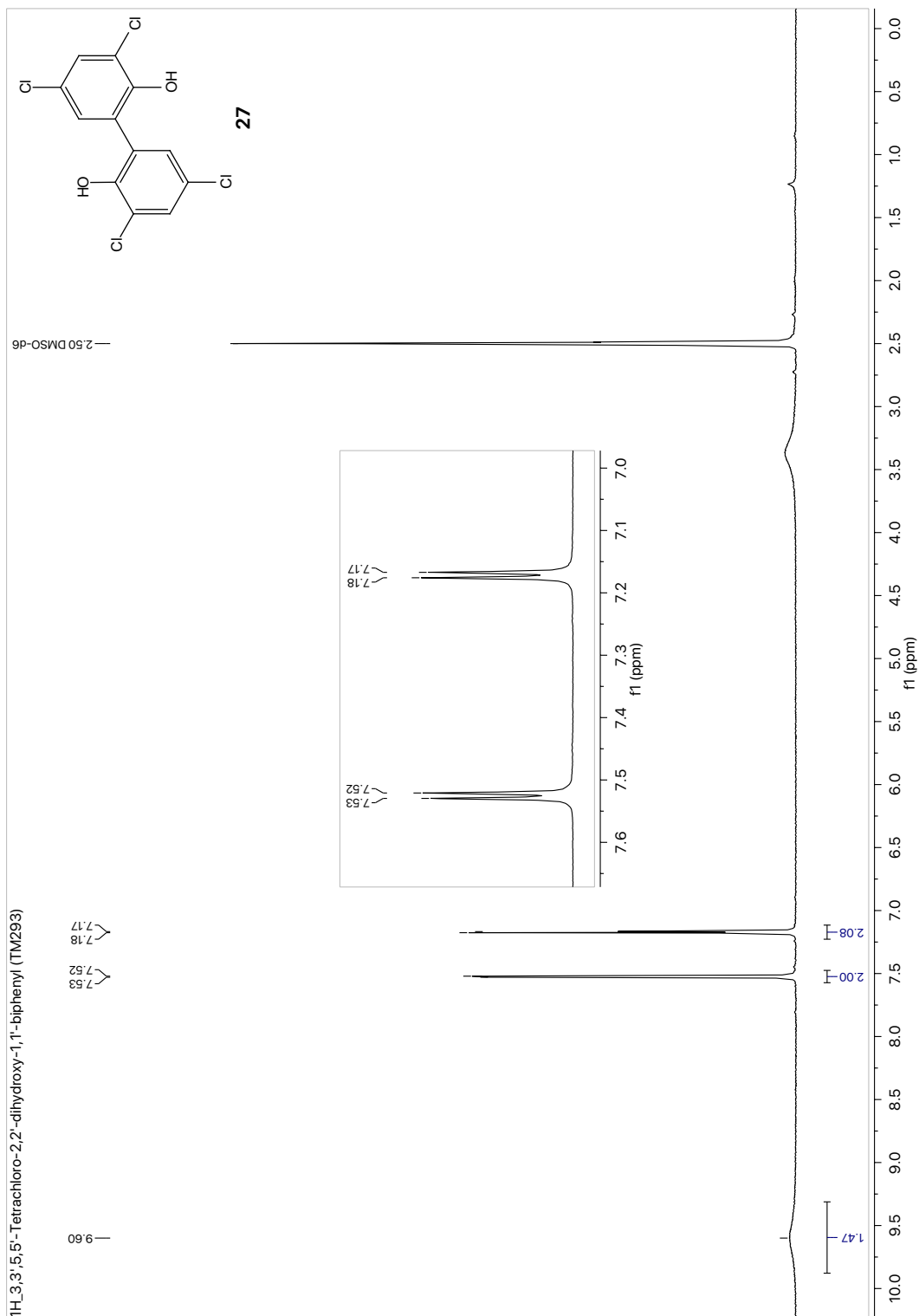


Fig. S61 ¹H-NMR spectrum of 3,3',5,5'-tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (27). Measured in DMSO-d₆.

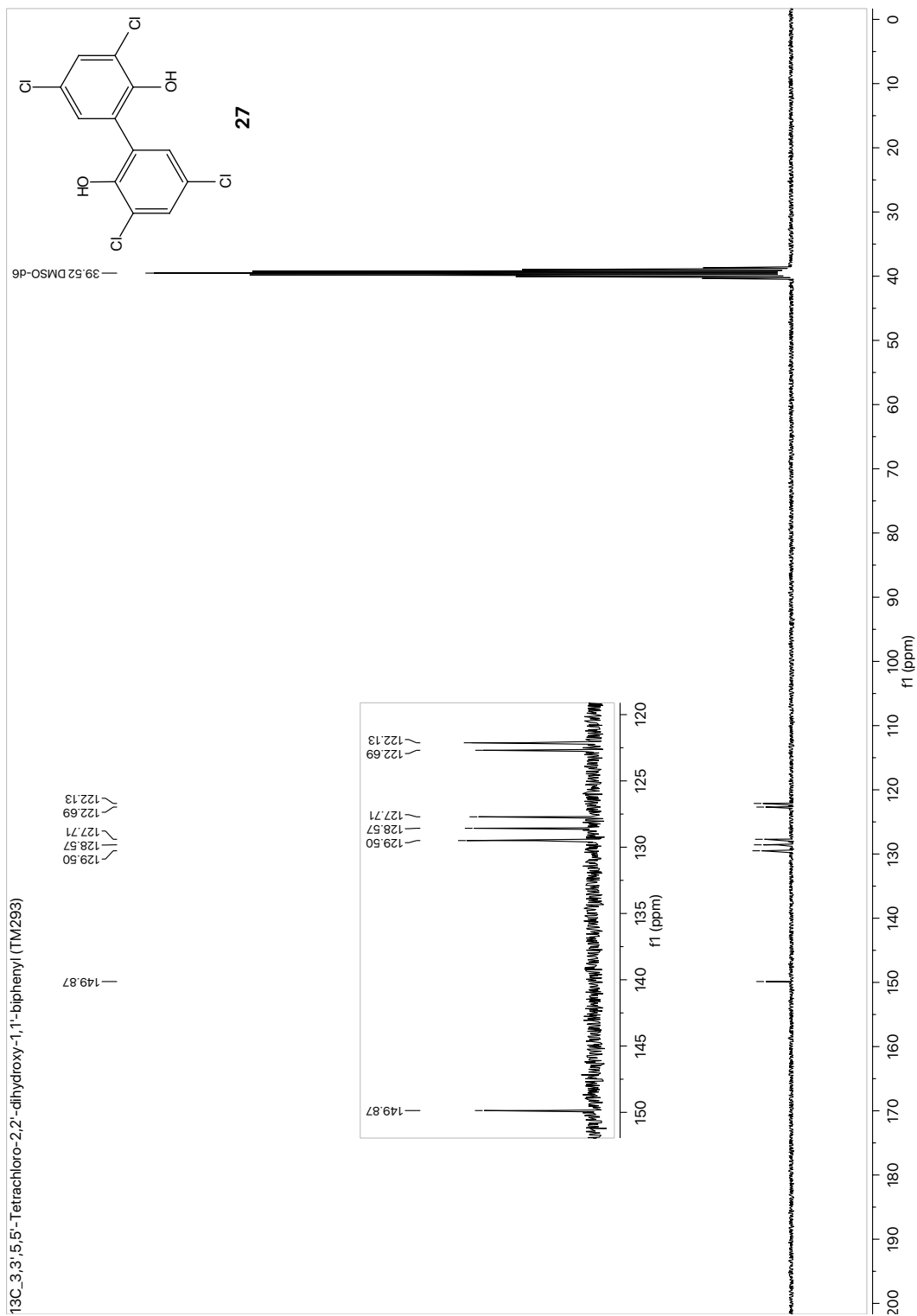


Fig. S62 ¹³C-NMR spectrum of 3',3',5',5'-tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (**27**). Measured in DMSO-d₆.

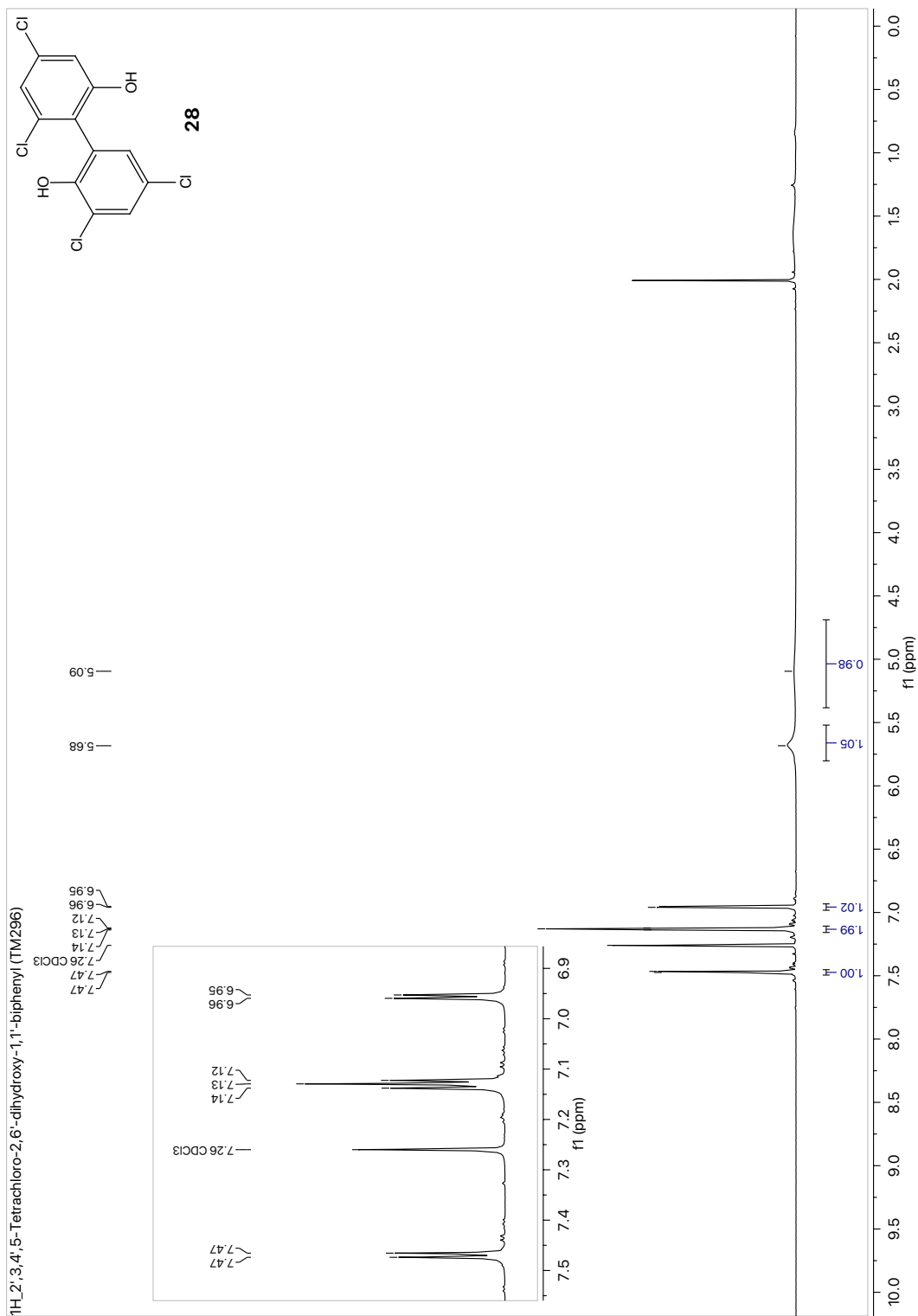


Fig. S63 ¹H-NMR spectrum of 2,3,4,5-tetrachloro-2,6'-dihydroxy-1,1'-biphenyl (**28**).

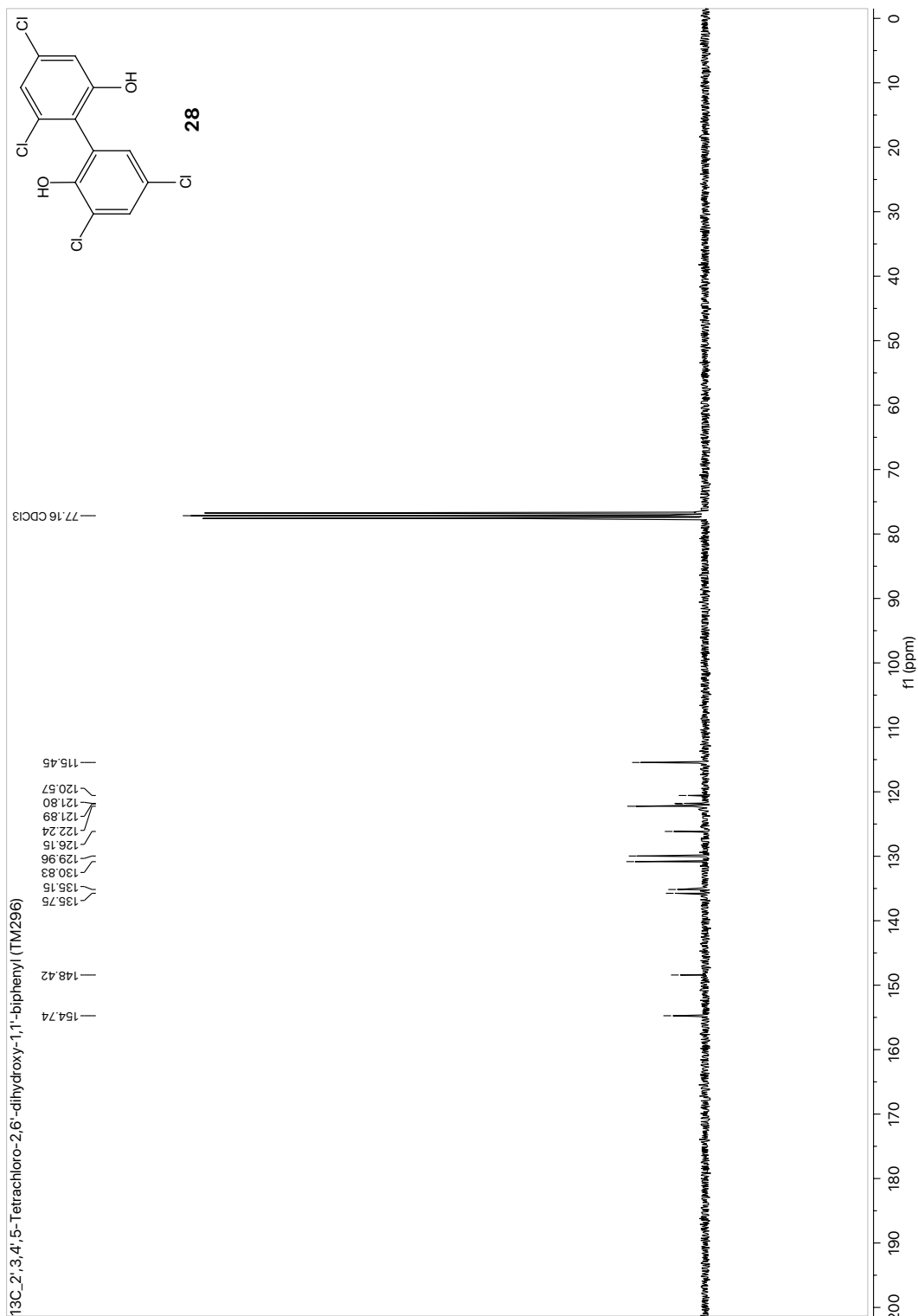


Fig. S64 ¹³C-NMR spectrum of 2',3,4',5-tetrachloro-2,6'-dihydroxy-1,1'-biphenyl (**28**).

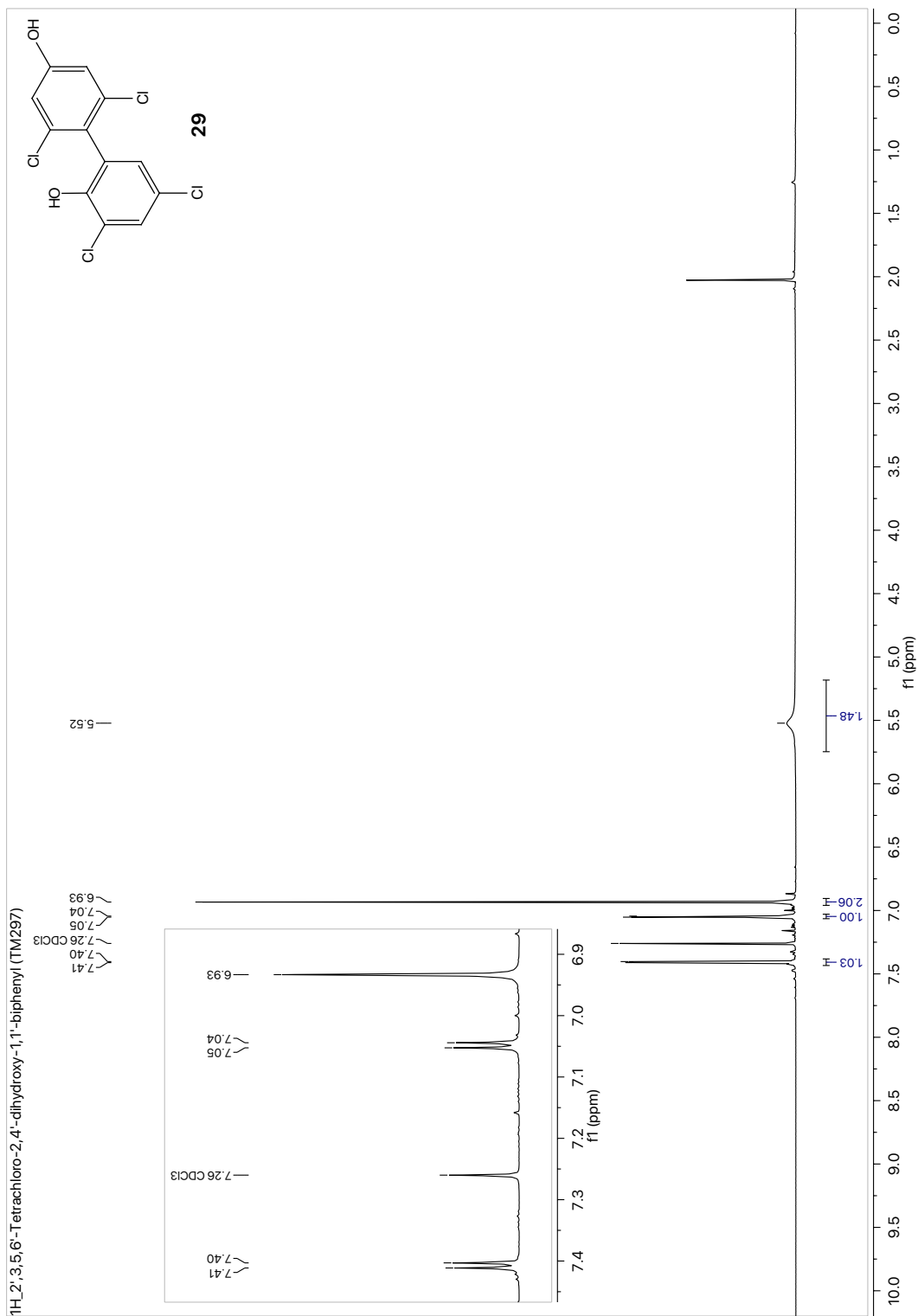


Fig. S65 ¹H-NMR spectrum of 2',3,5,6'-tetrachloro-2,4'-dihydroxy-1,1'-biphenyl (**29**).

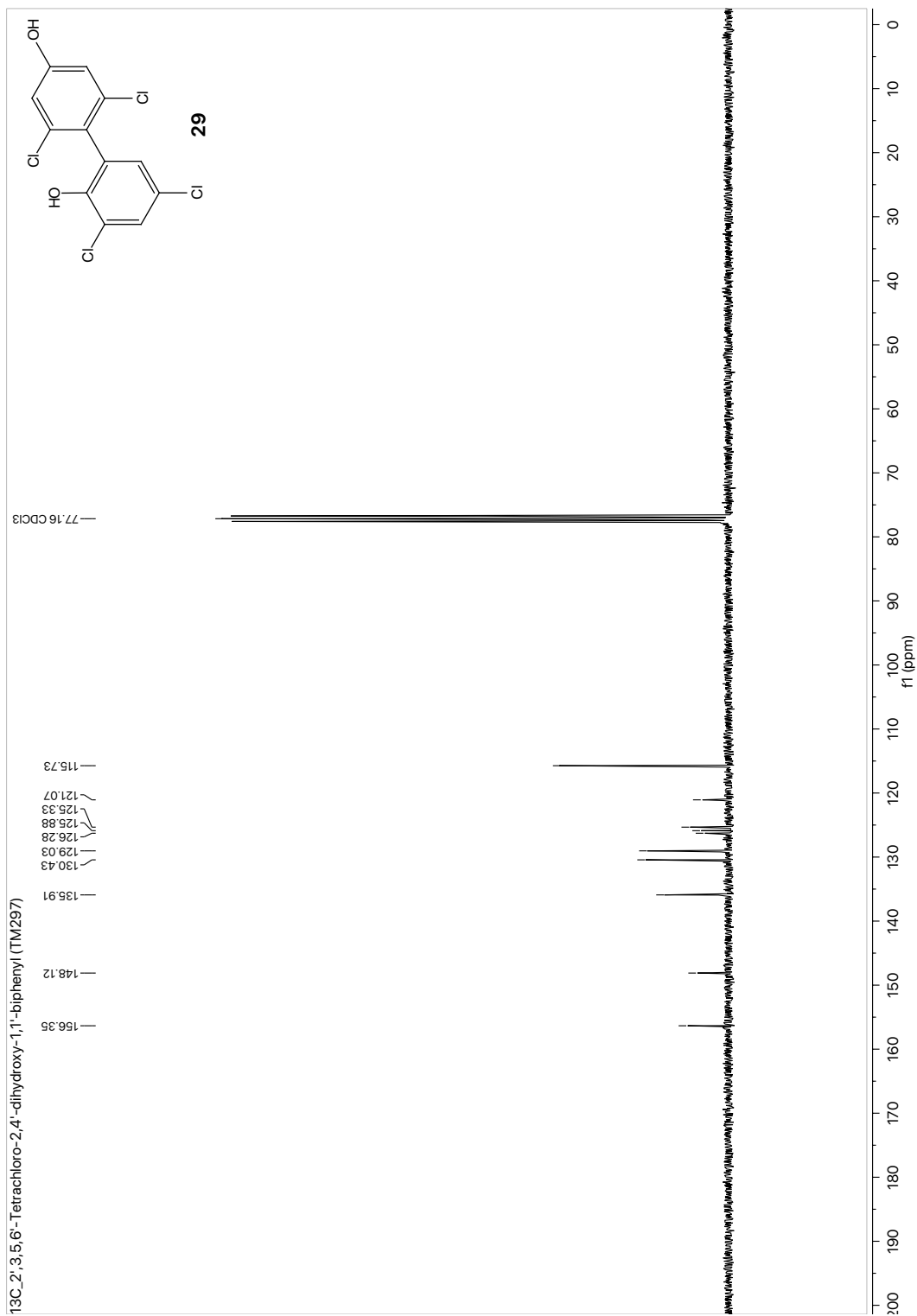


Fig. S66 ¹³C-NMR spectrum of 2',3,5,6'-tetrachloro-2,4'-dihydroxy-1,1'-biphenyl (**29**).

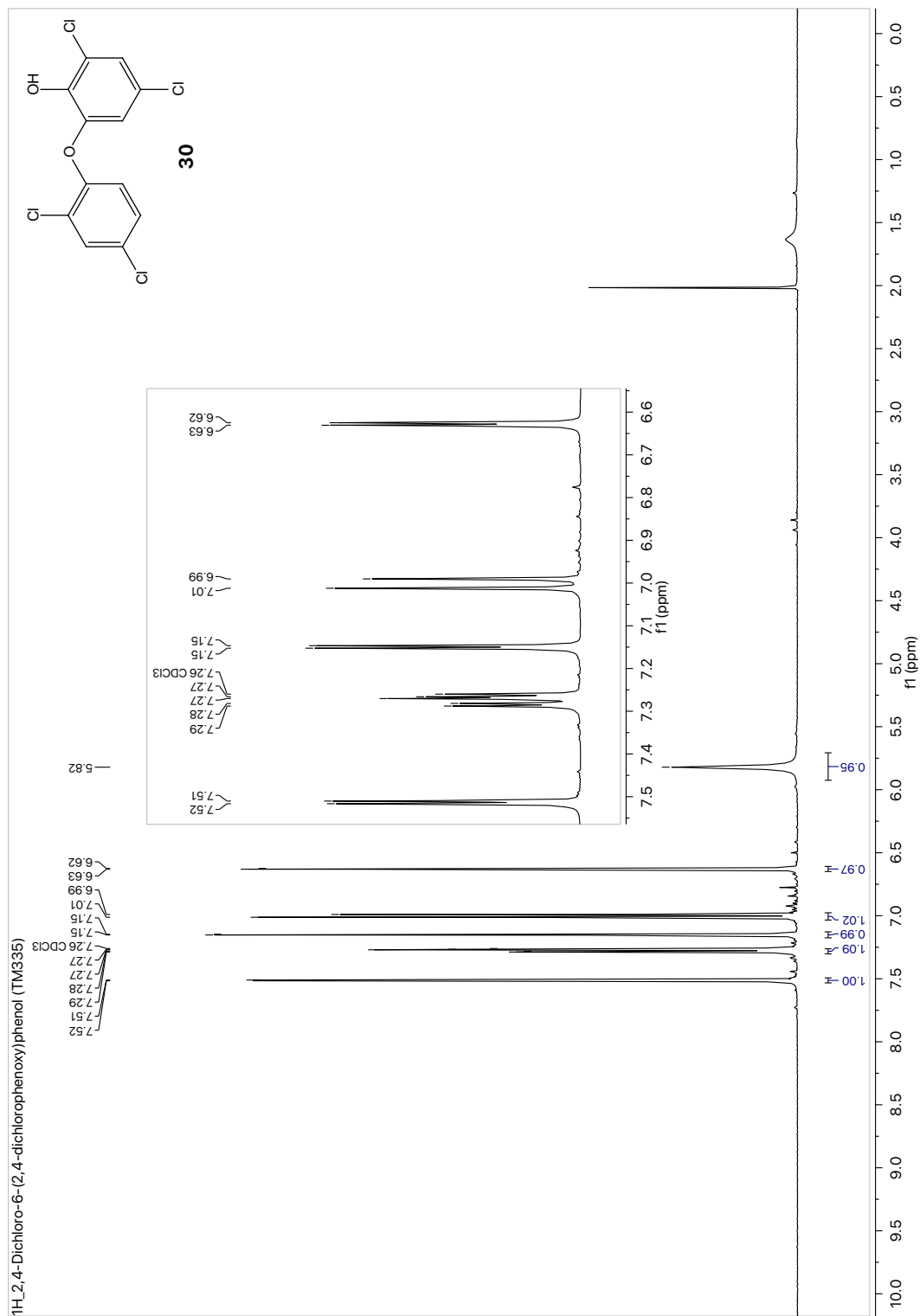


Fig. S67 ¹H-NMR spectrum of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**). Measured in CDCl₃.

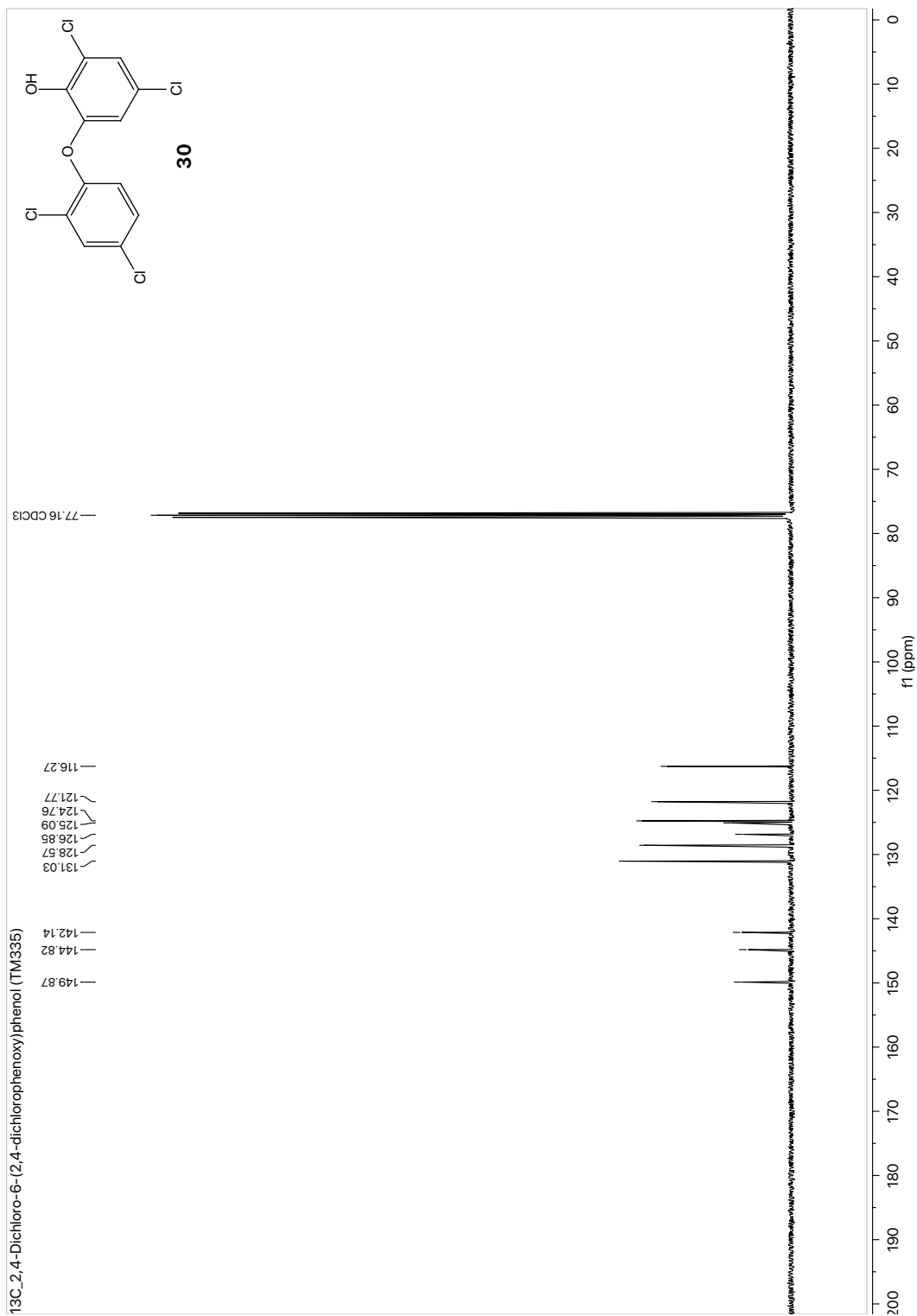


Fig. S68 ¹³C-NMR spectrum of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (30). Measured in CDCl₃.

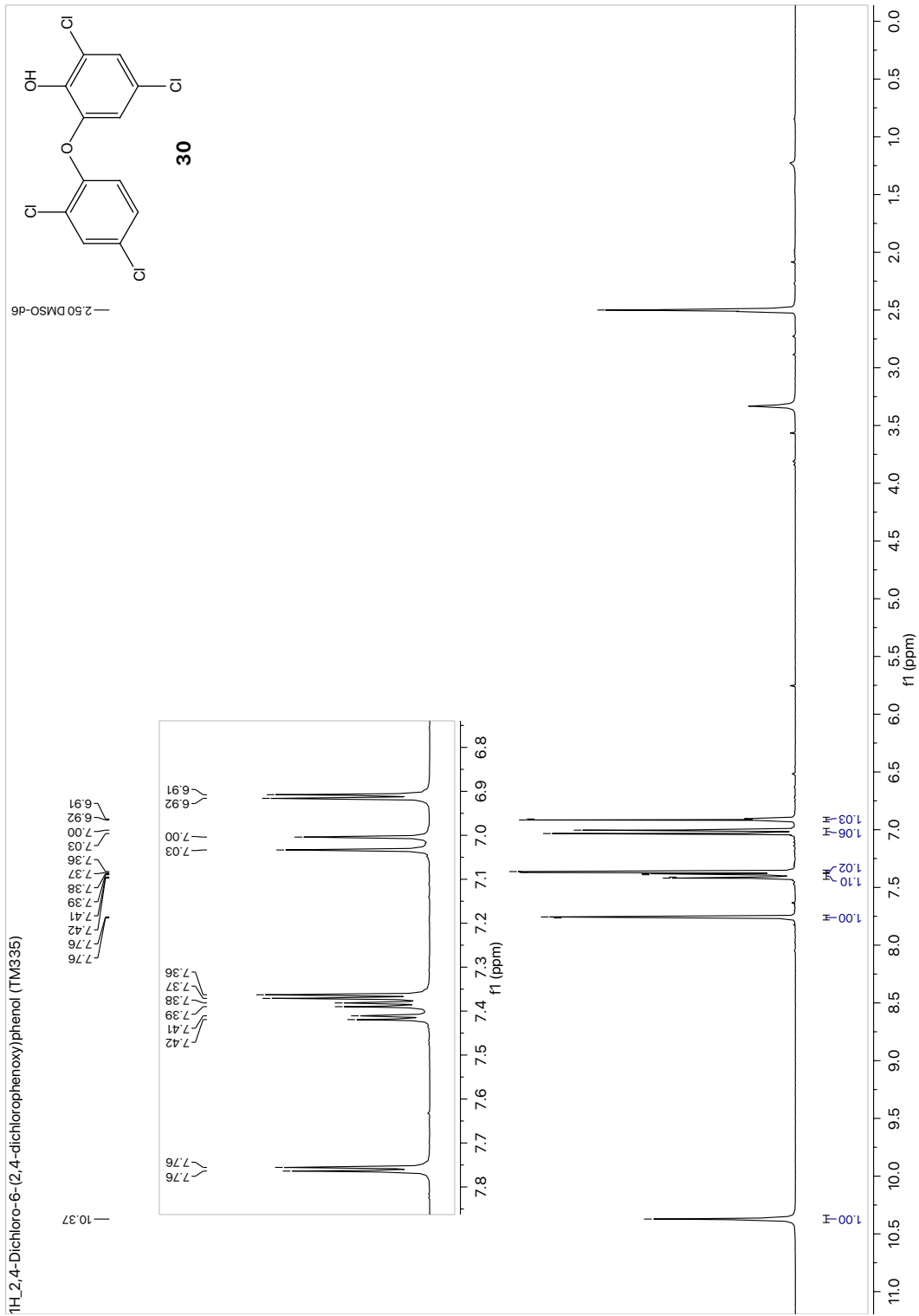


Fig. S69 ¹H-NMR spectrum of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**). Measured in DMSO-d₆.

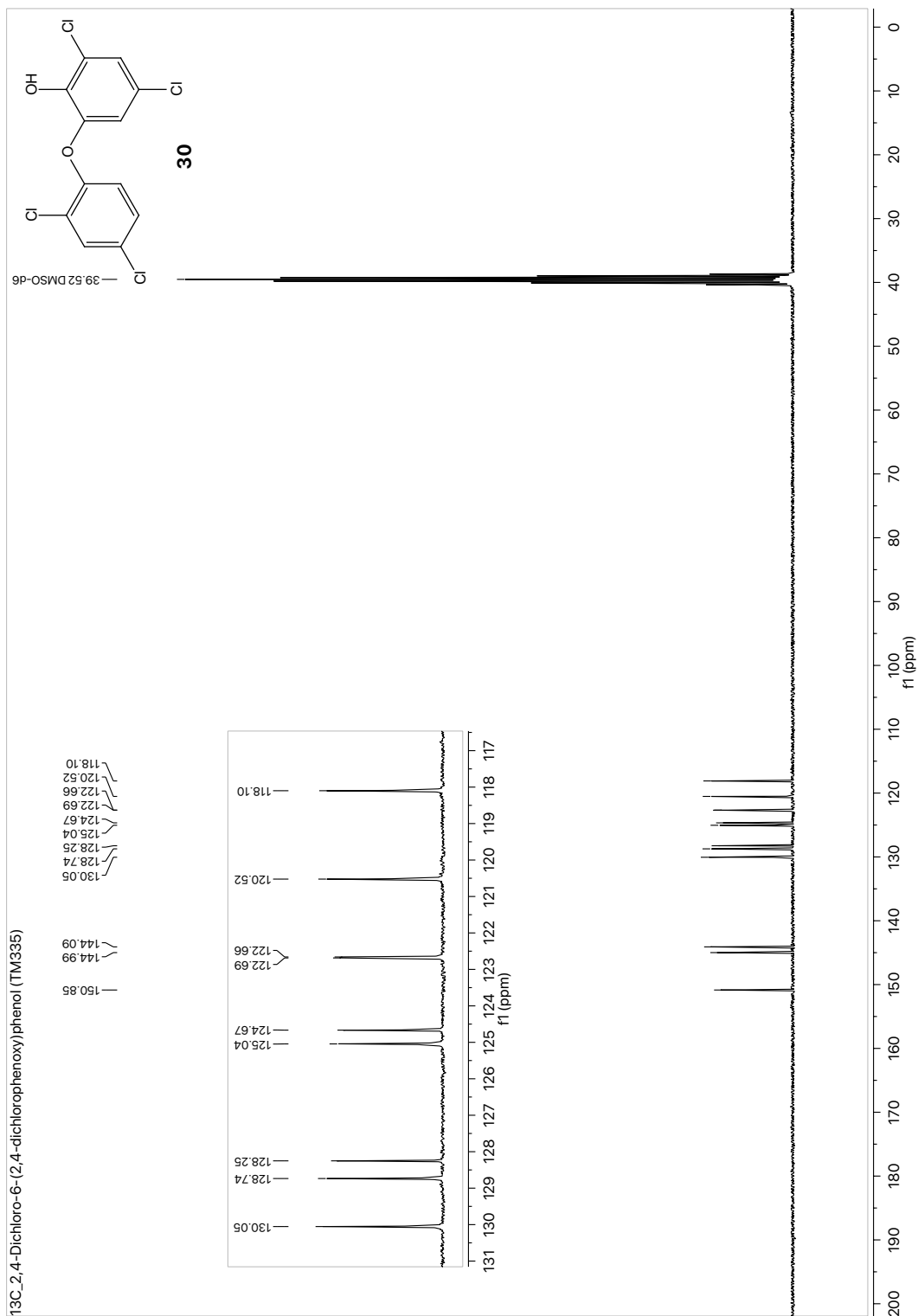


Fig. S70 ¹³C-NMR spectrum of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (**30**). Measured in DMSO-d₆.

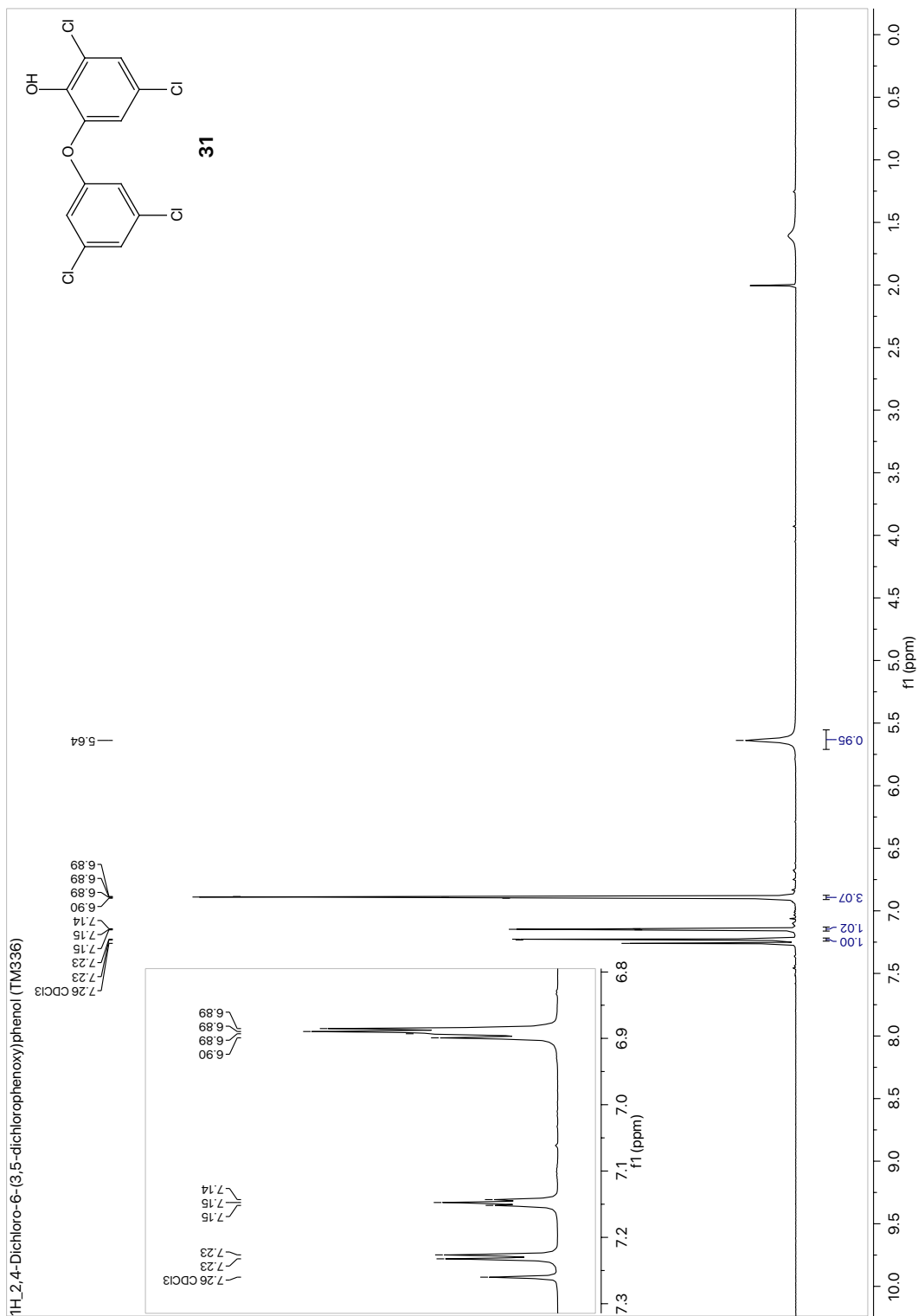


Fig. S71 ¹H-NMR spectrum of 2,4-dichloro-6-(3,5-dichlorophenoxy)phenol (**31**).

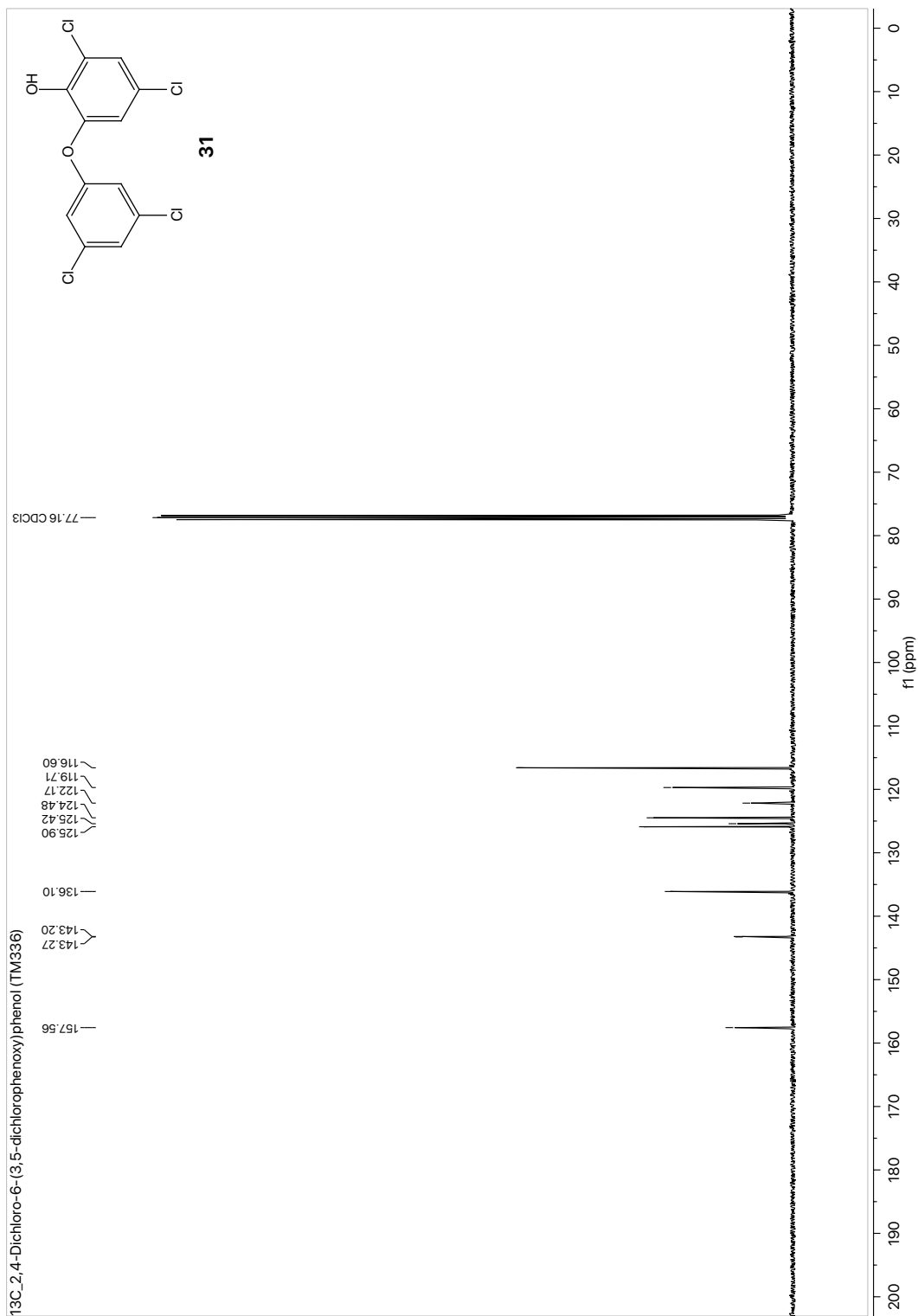


Fig. S72 ¹³C-NMR spectrum of 2,4-dichloro-6-(3,5-dichlorophenoxy)phenol (**31**).

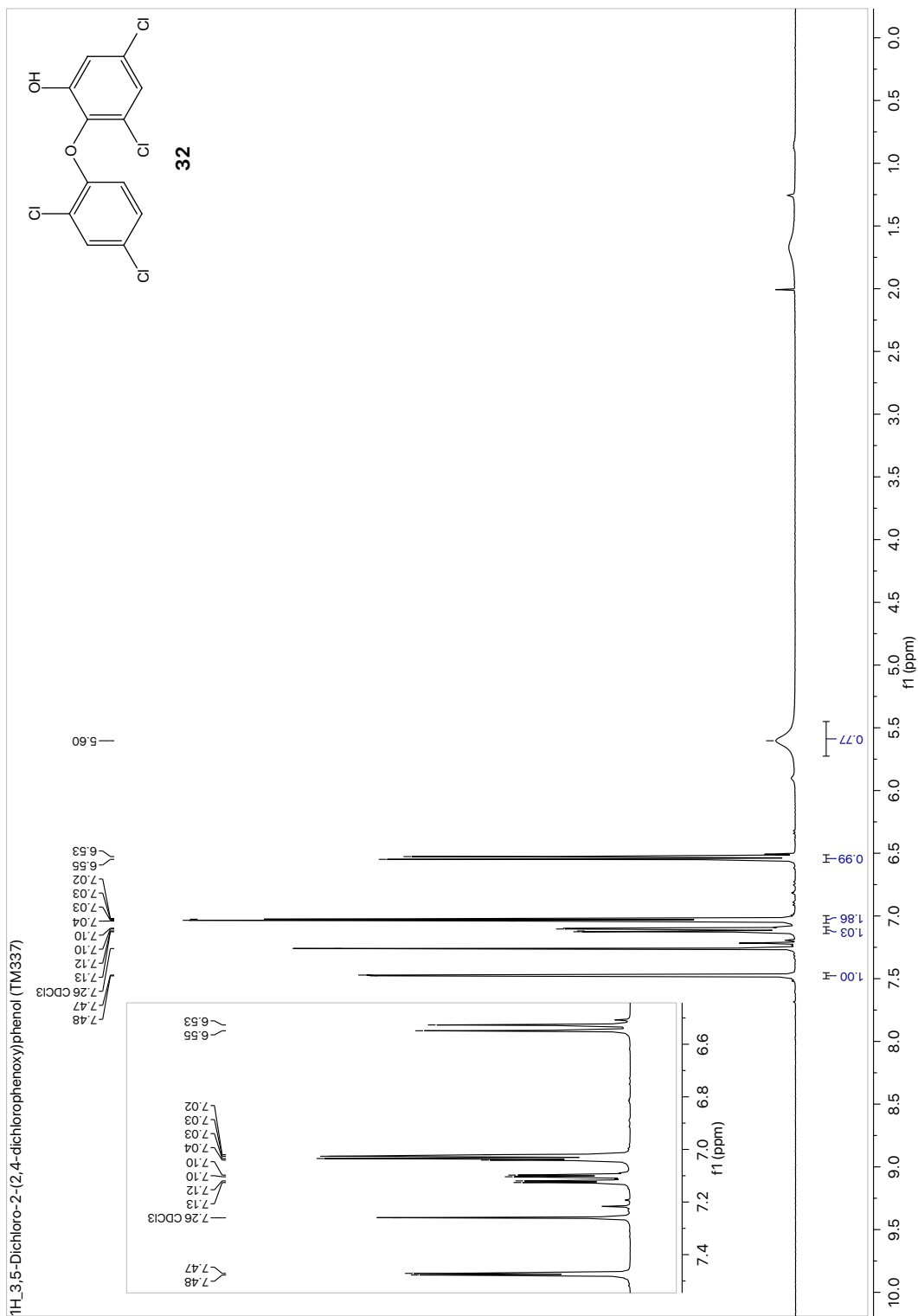


Fig. S73 ¹H-NMR spectrum of 3,5-dichloro-2-(2,4-dichlorophenoxy)phenol (**32**).

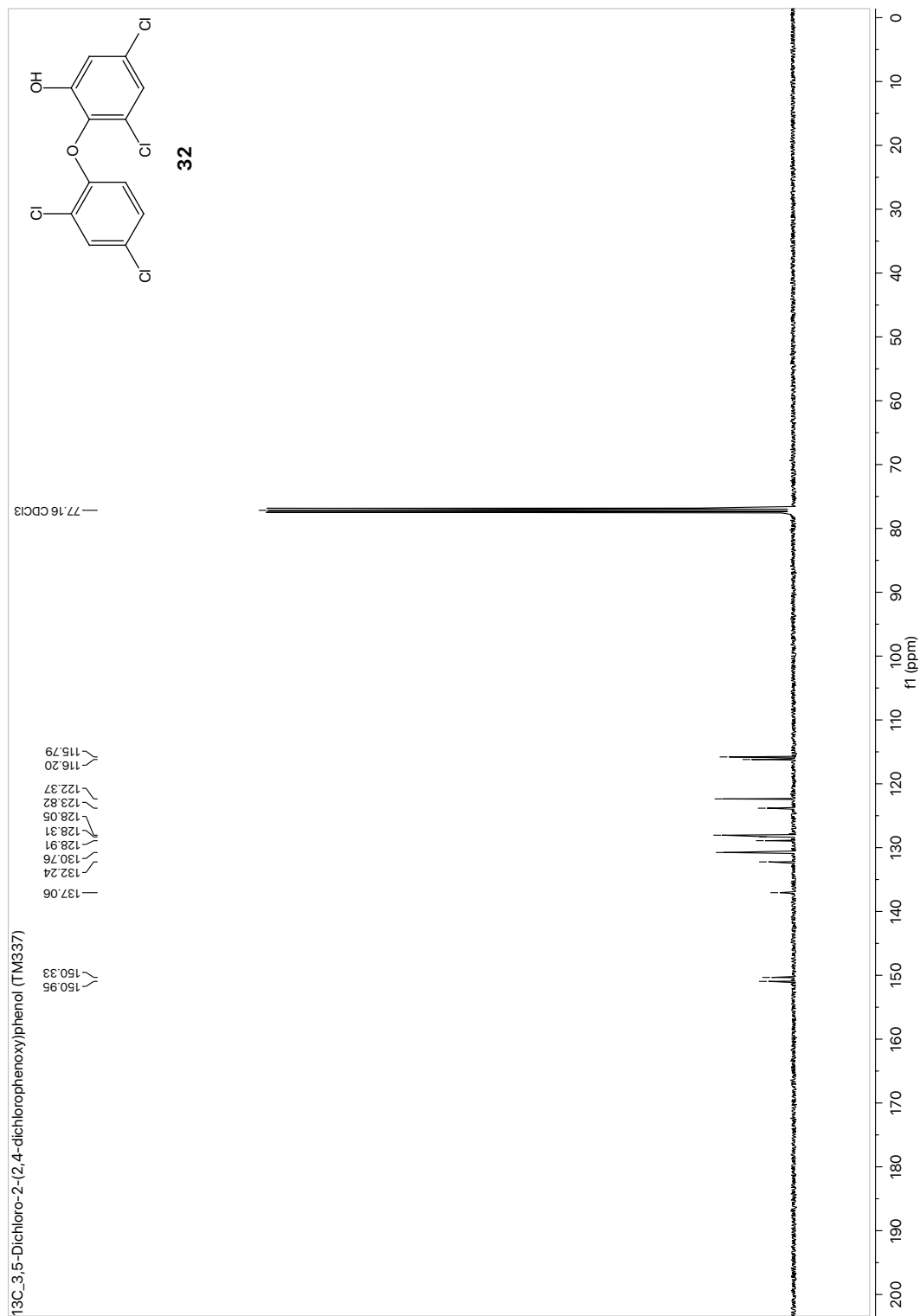


Fig. S74 ¹³C-NMR spectrum of 3,5-dichloro-2-(2,4-dichlorophenoxy)phenol (**32**).

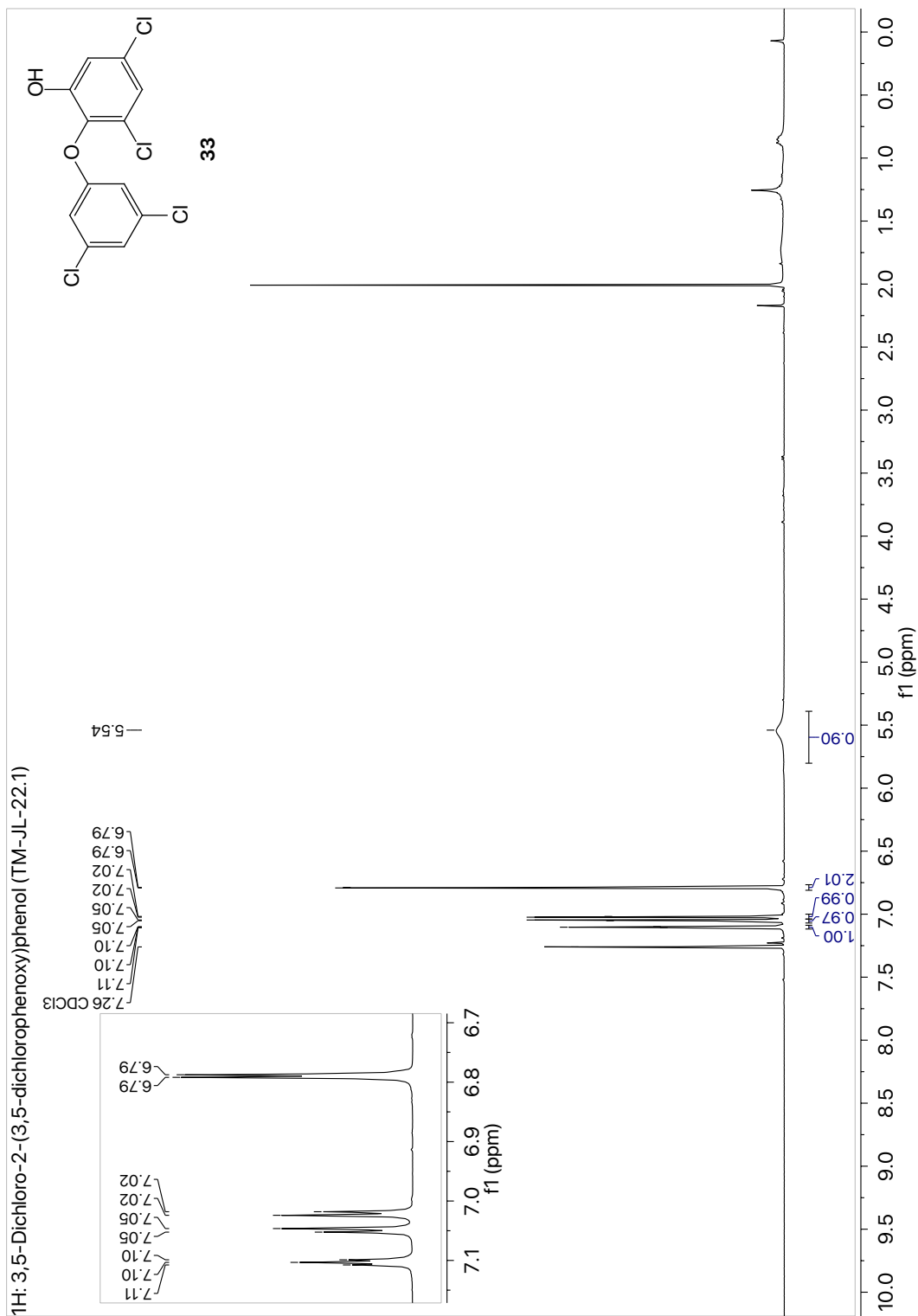


Fig. S75 ¹H-NMR spectrum of 3,5-dichloro-2-(3,5-dichlorophenoxy)phenol (**33**).

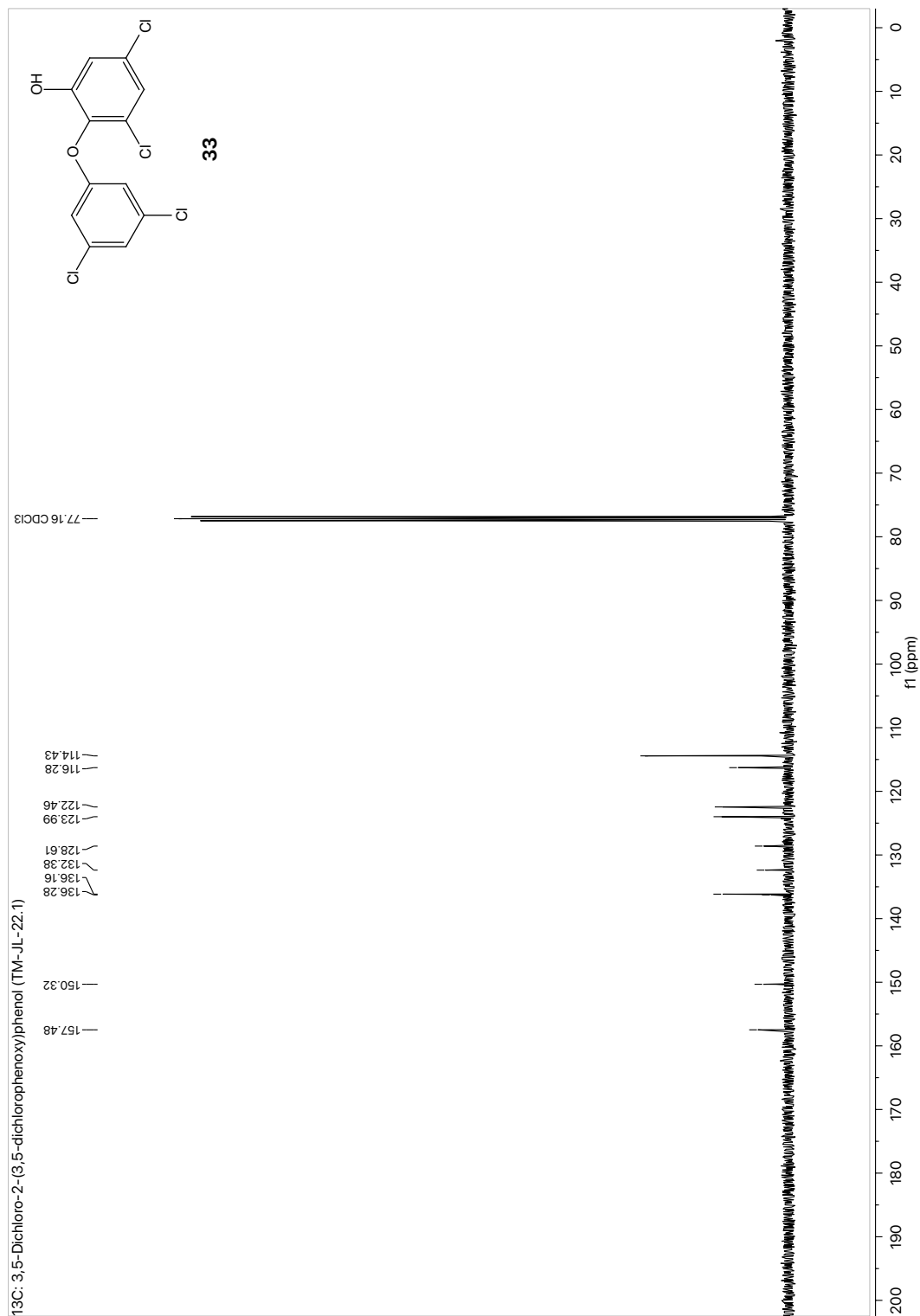


Fig. S76 ¹³C-NMR spectrum of 3,5-dichloro-2-(3,5-dichlorophenoxy)phenol (**33**).

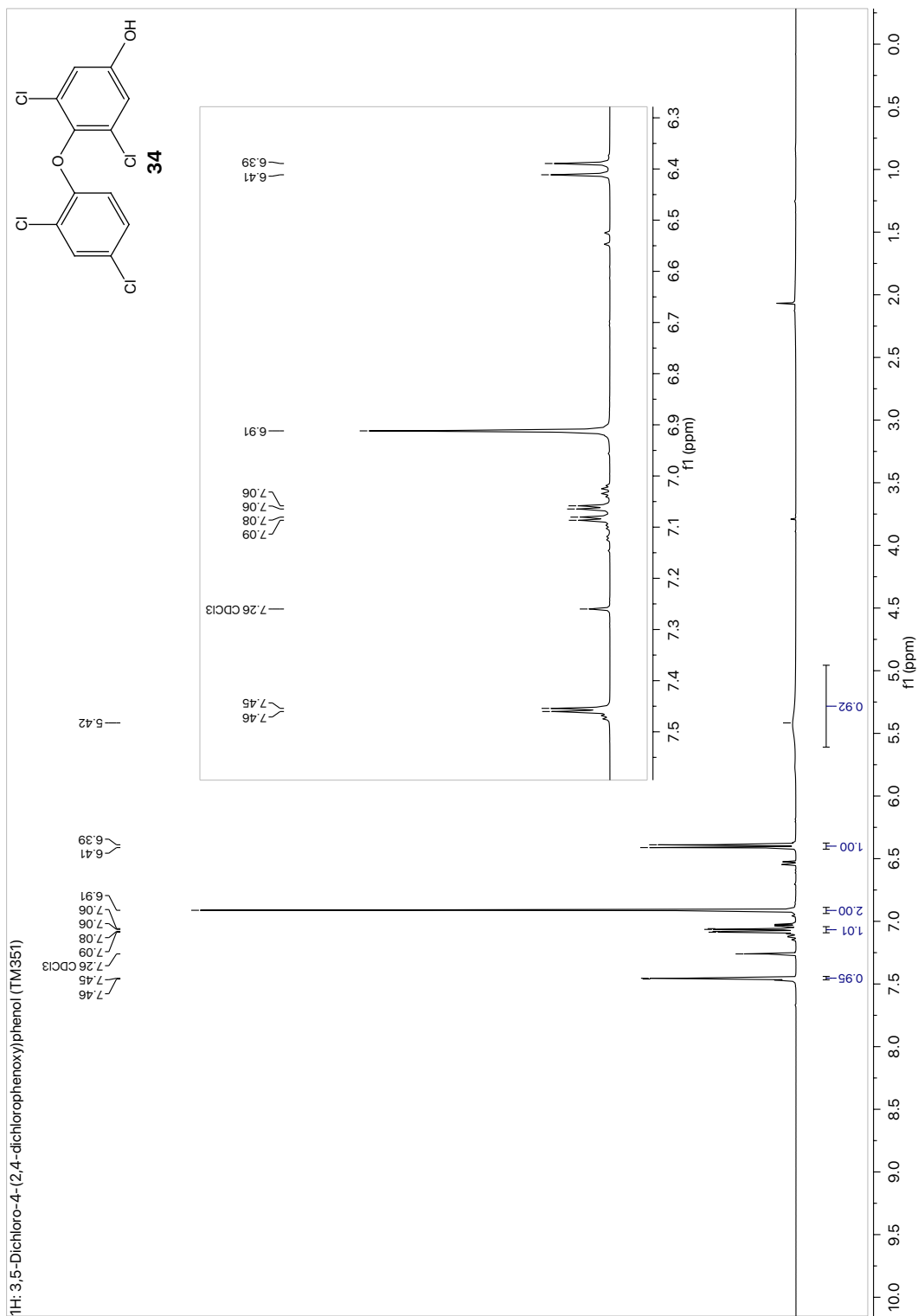


Fig. S77 ¹H-NMR spectrum of 3,5-dichloro-4-(2,4-dichlorophenoxy)phenol (**34**).

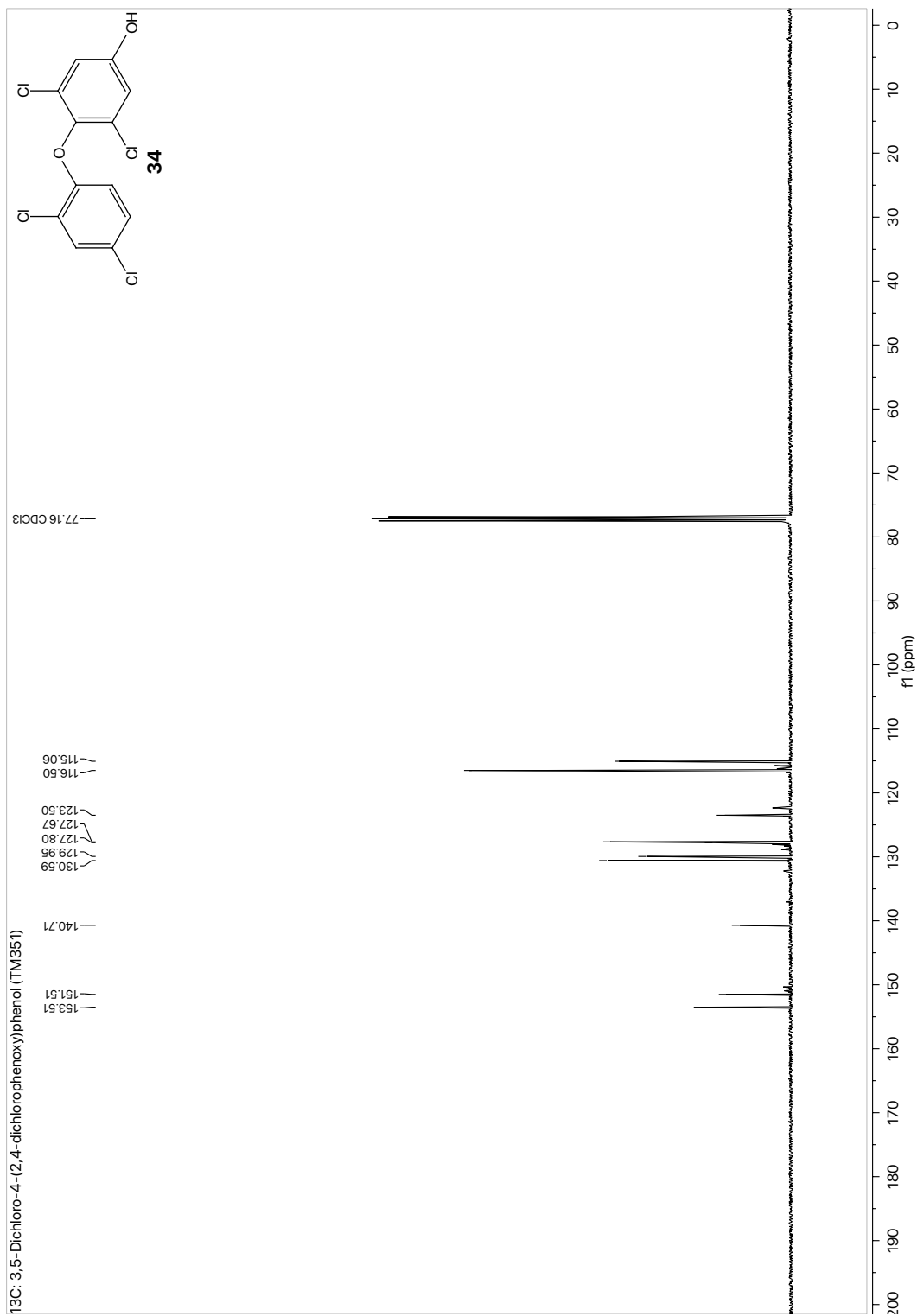


Fig. S78 ¹³C-NMR spectrum of 3,5-dichloro-4-(2,4-dichlorophenoxy)phenol (**34**).

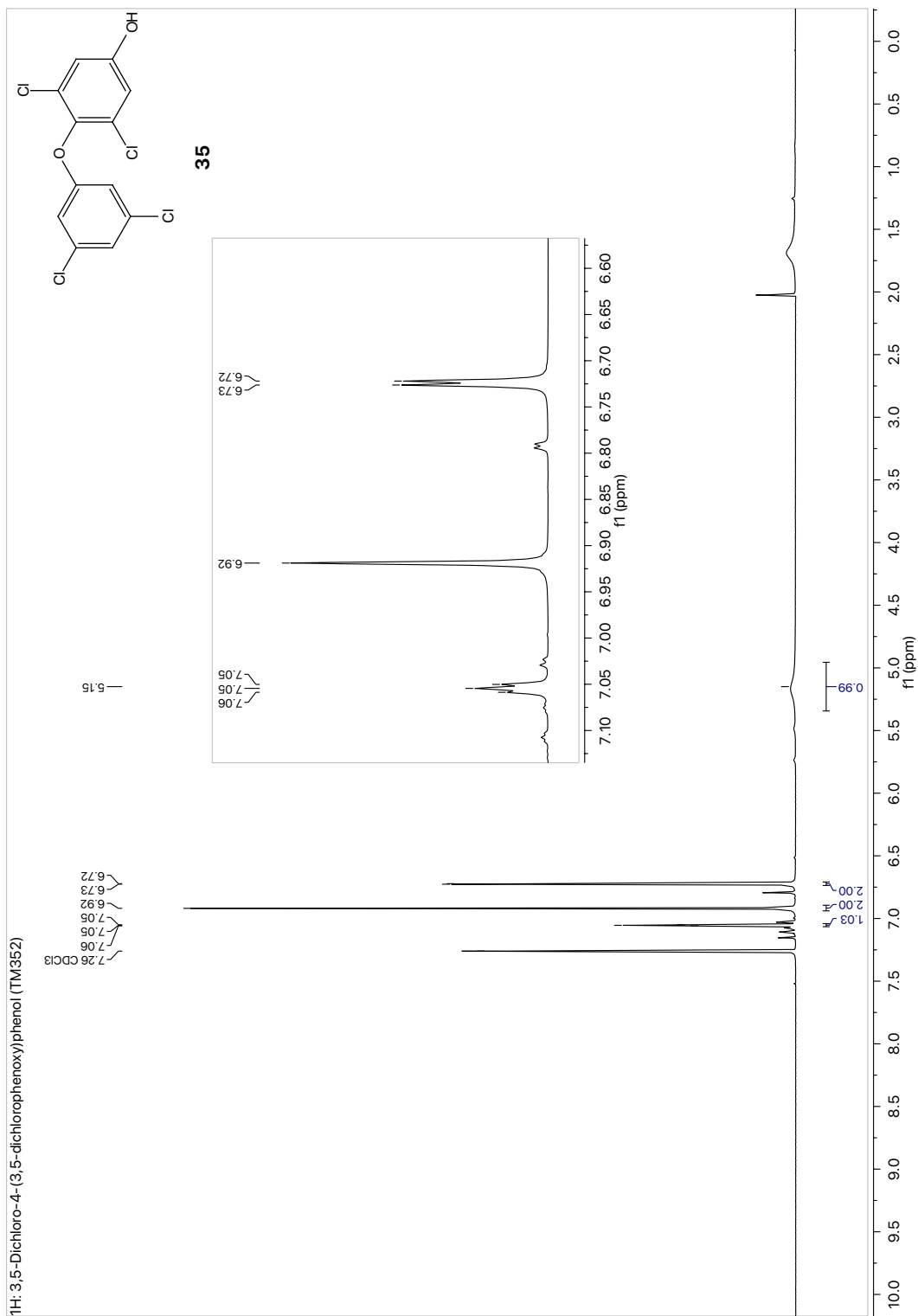


Fig. S79 ¹H-NMR spectrum of 3,5-dichloro-4-(3,5-dichlorophenoxy)phenol (**35**).

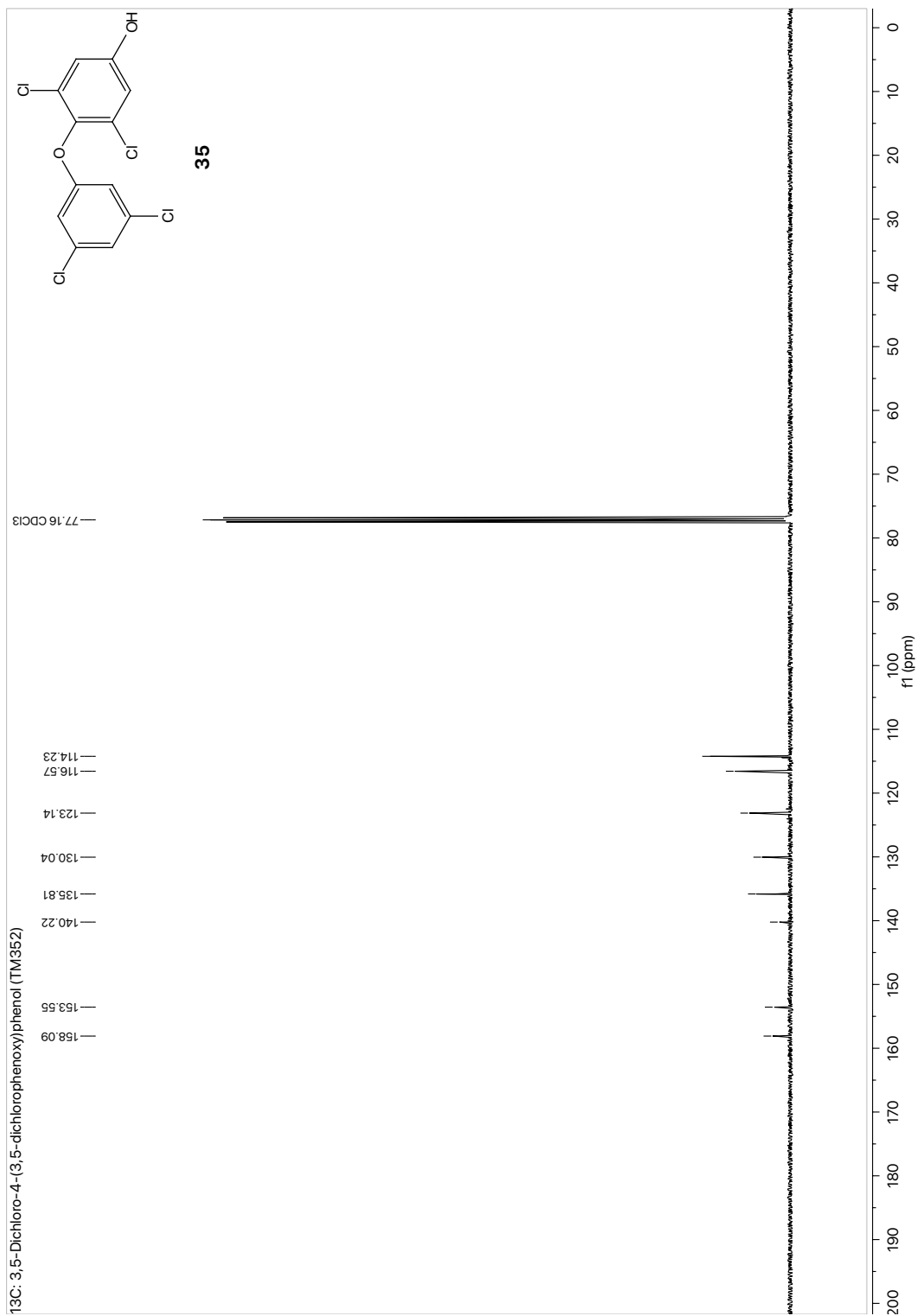


Fig. S80 ¹³C-NMR spectrum of 3,5-dichloro-4-(3,5-dichlorophenoxy)phenol (**35**).

8. UV-absorbance Measurements

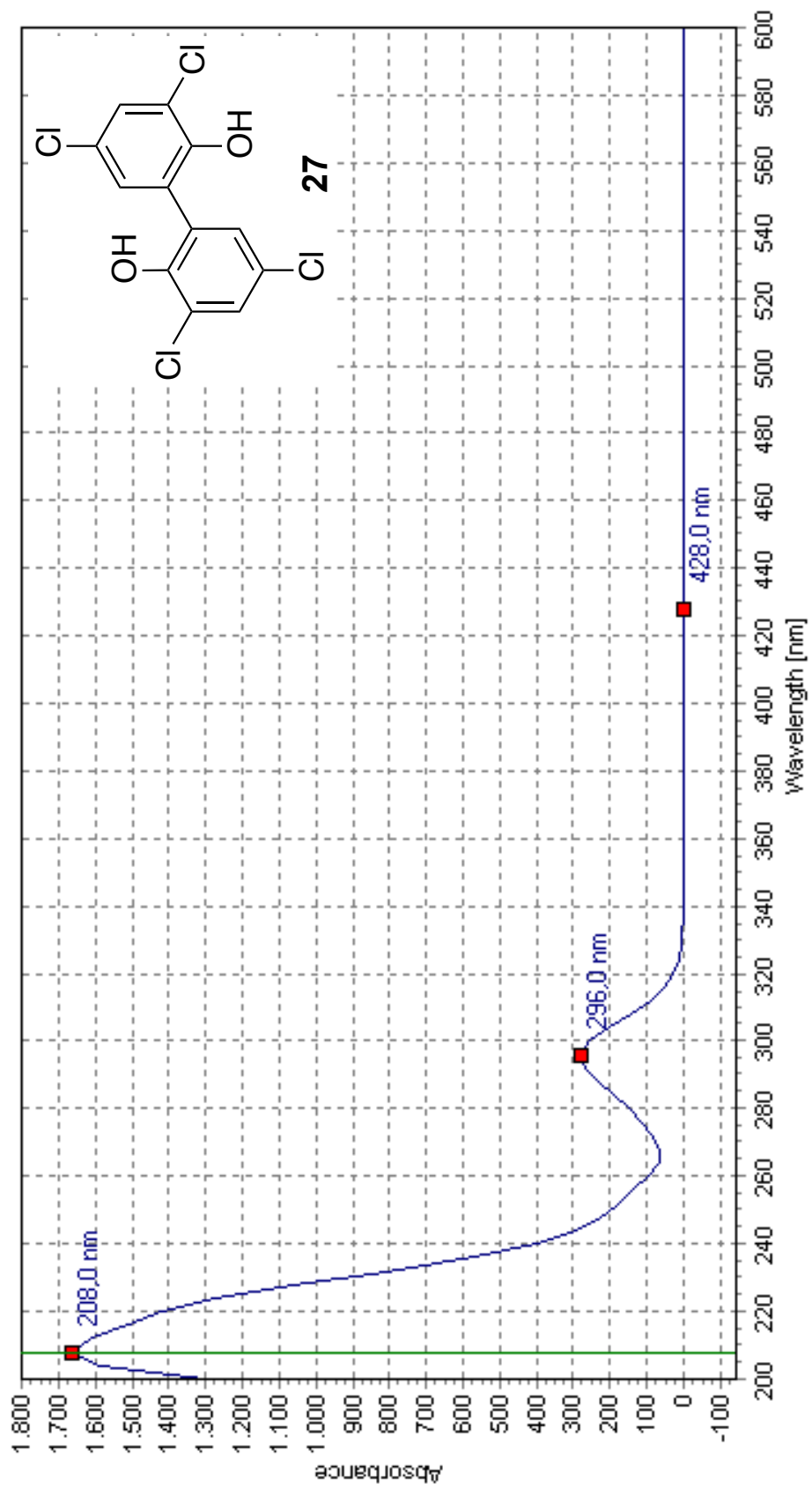


Fig. S81 UV-absorbance measurement of 3,3',5,5'-tetrachloro-2,2'-dihydroxy-1,1'-biphenyl (27).

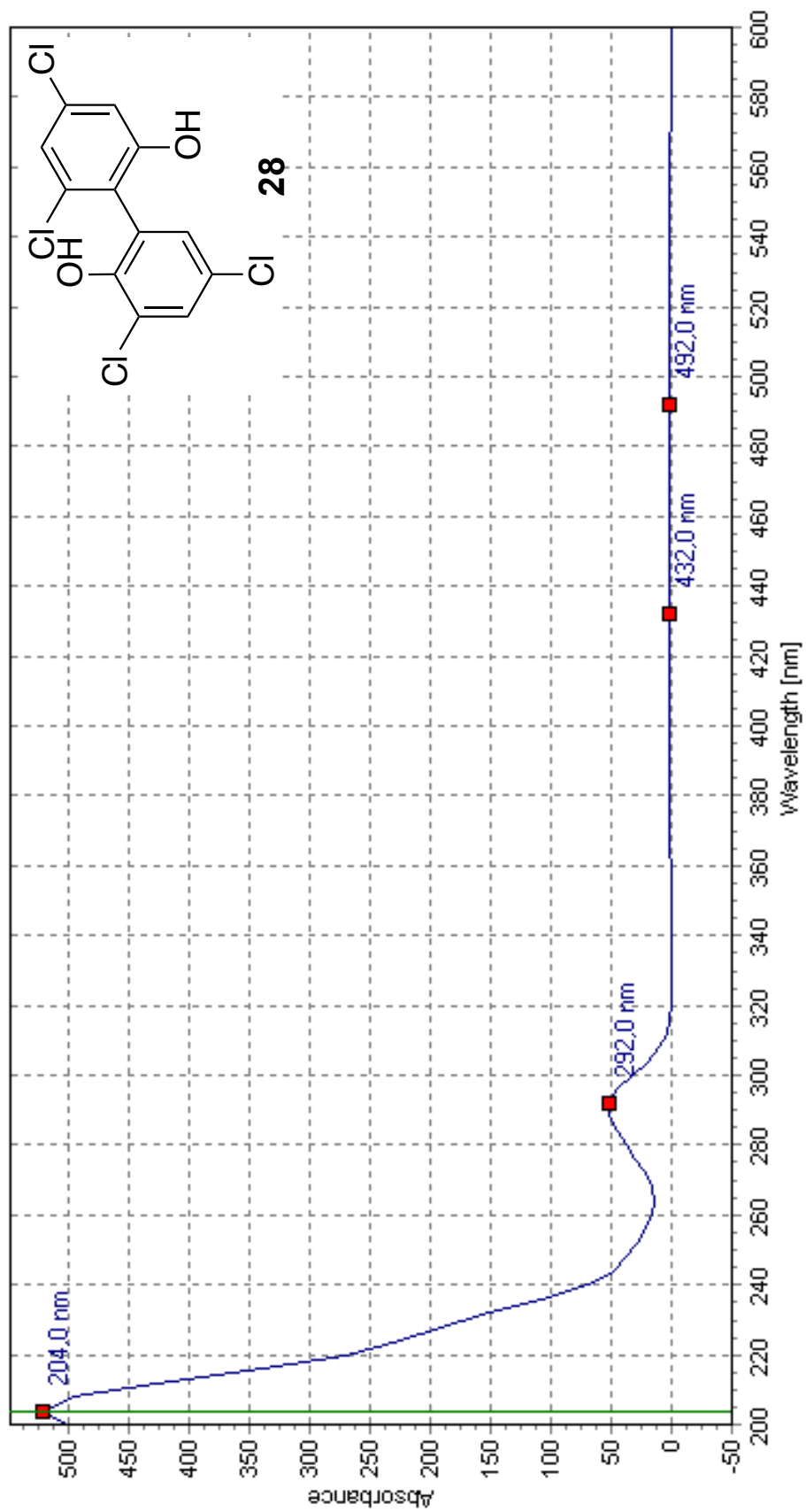


Fig. S82 UV-absorbance measurement of 2,3,4',5-tetrachloro-2,6'-dihydroxy-1,1'-biphenyl (**28**).

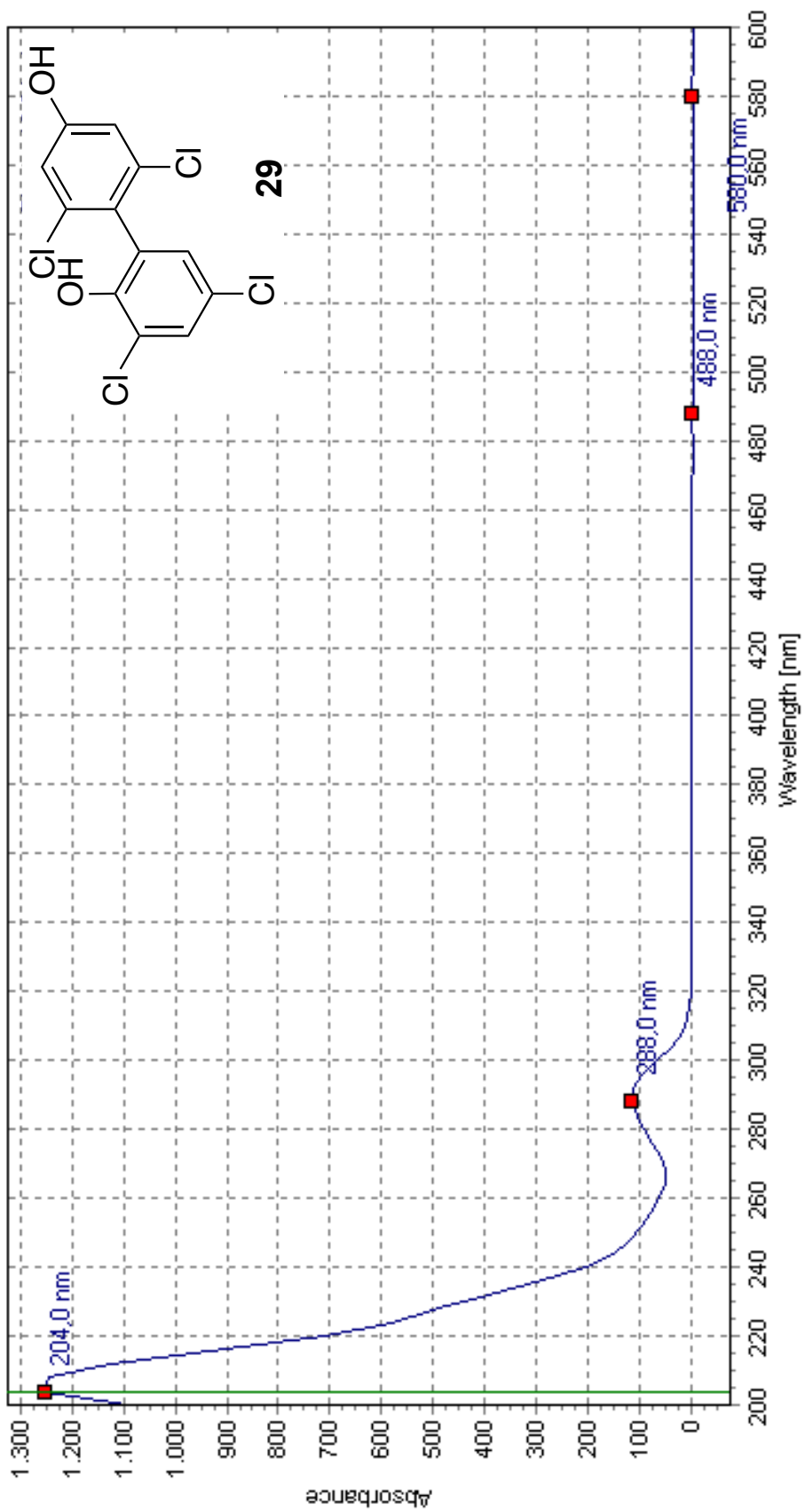


Fig. S83 UV-absorbance measurement of 2',3,5,6'-tetrachloro-2,4'-dihydroxy-1,1'-biphenyl (29).

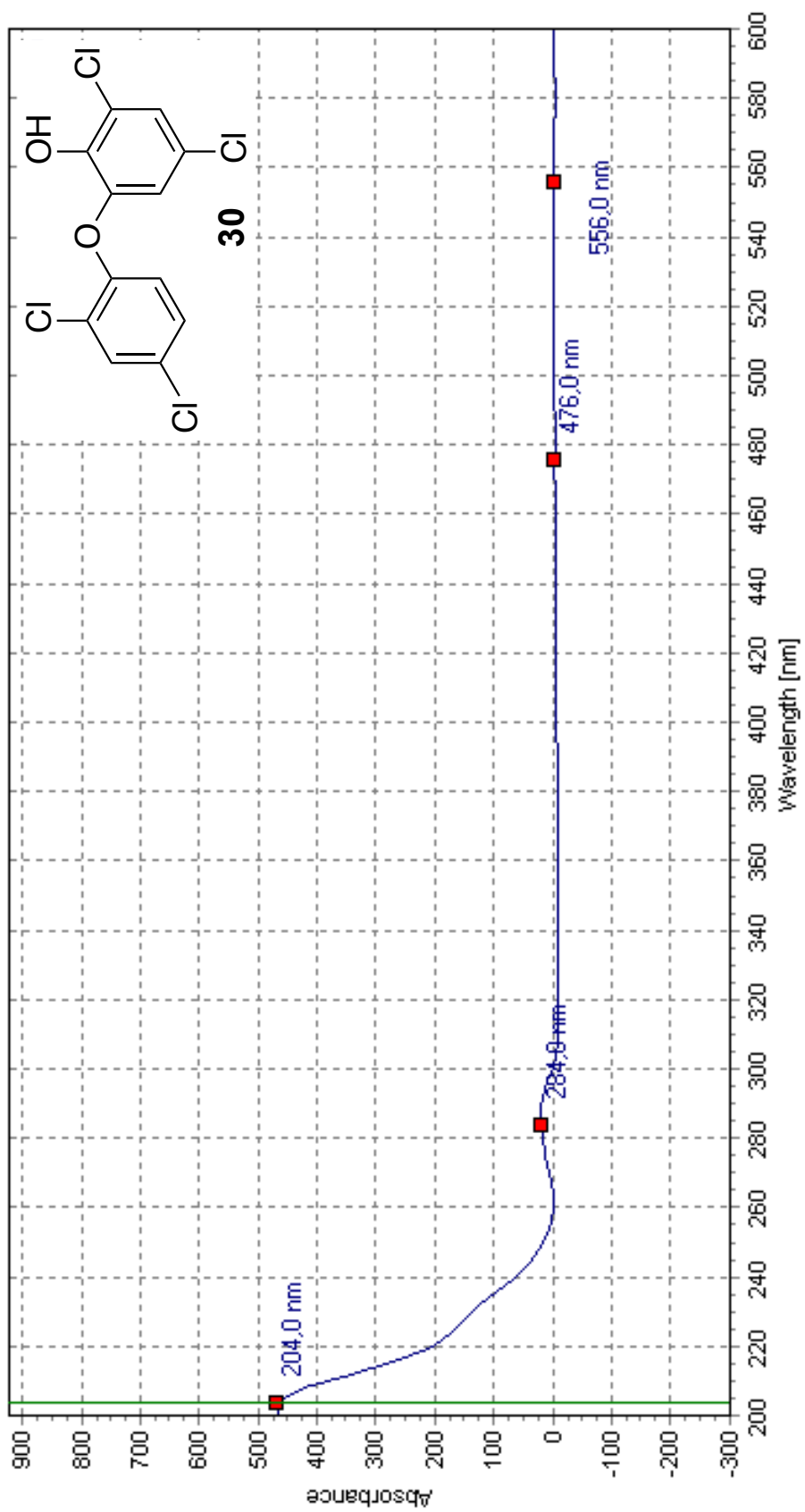


Fig. S84 UV-absorbance measurement of 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (30).

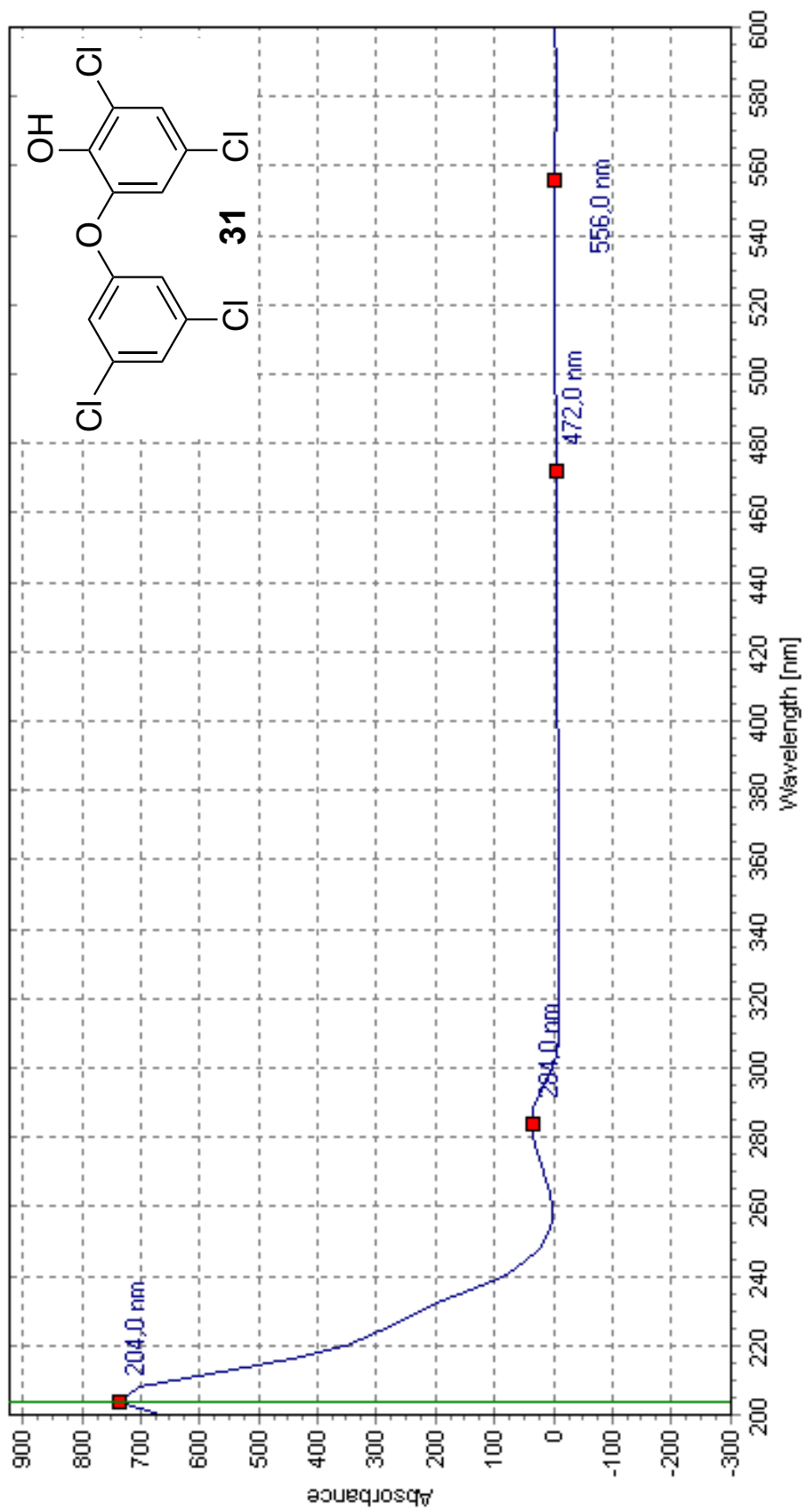


Fig. S85 UV-absorbance measurement of 2,4-dichloro-6-(3,5-dichlorophenoxy)phenol (31).

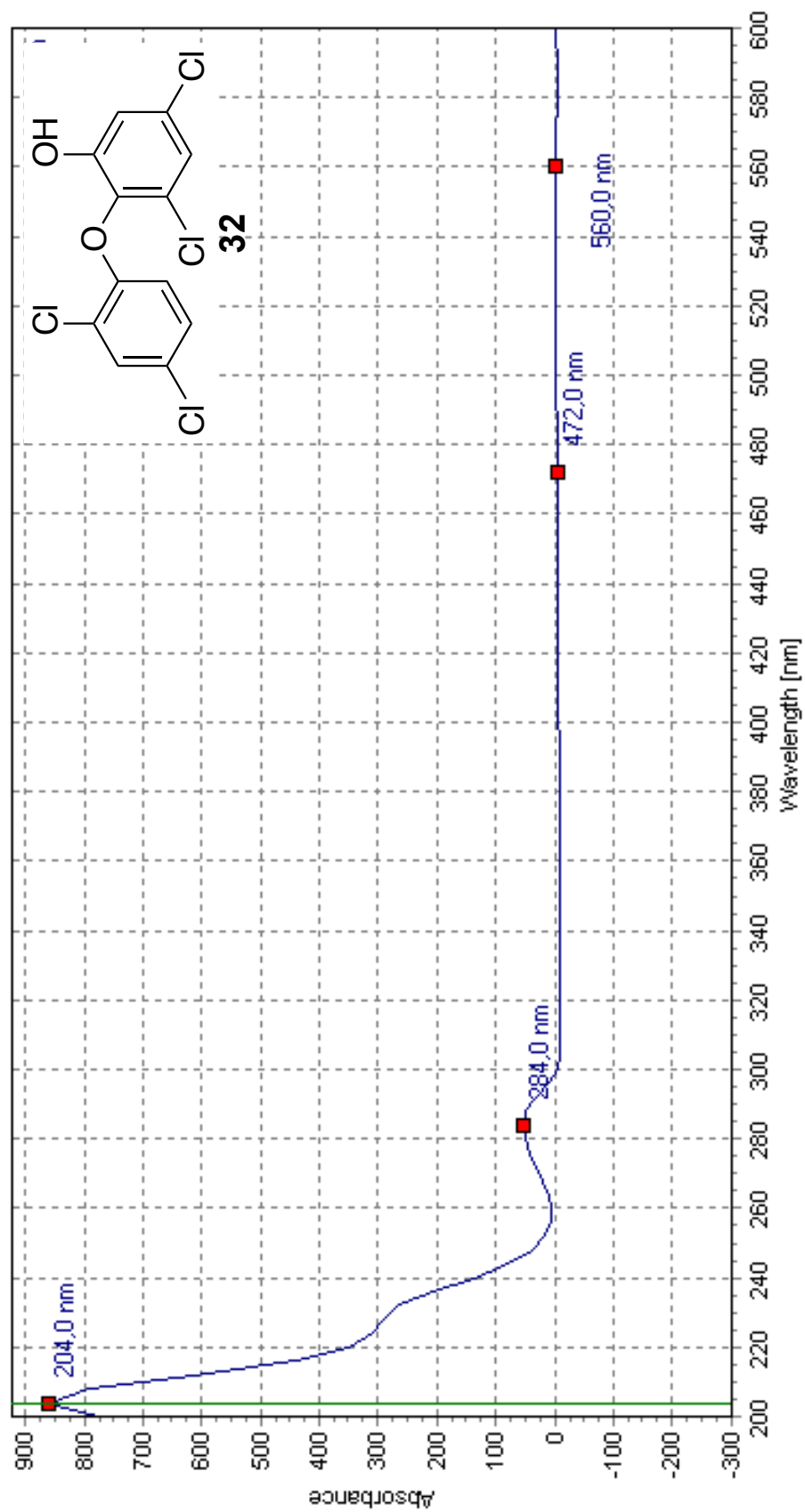


Fig. S86 UV-absorbance measurement of 3,5-dichloro-2-(2,4-dichlorophenoxy)phenol (**32**).

9. References

- 1 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
- 2 C. Greunke, E. R. Duell, P. M. D'Agostino, A. Glöckle, K. Lamm and T. A. M. Gulder, *Metab. Eng.*, 2018, **47**, 334–345.
- 3 V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2017, **13**, 537–543.
- 4 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790–1792.
- 5 V. Agarwal, J. M. Blanton, S. Podell, A. Taton, M. A. Schorn, J. Busch, Z. Lin, E. W. Schmidt, P. R. Jensen, V. J. Paul, J. S. Biggs, J. W. Golden, E. E. Allen and B. S. Moore, *Nat. Chem. Biol.*, 2017, **13**, 537–543.
- 6 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, J. S. Dickschat, *Angew. Chem. Int. Ed.*, 2014, **53**, 7652–7656.
- 7 S. E. Onley, X. Bian, Y. Zhang, R. Chau, W. H. Gerwick, R. Muller and B. A. Neilan, *ACS Chem. Biol.*, 2013, **8**, 1888–1893.
- 8 R. Wachtel, B. Brauning, S. L. Mader, F. Ecker, V. R. I. Kaila, M. Groll and A. Itzen, *Nat. Commun.*, 2018, **9**, 44.
- 9 A. T. Ma, C. M. Schmidt and J. W. Golden, *Appl. Environ. Microbiol.*, 2014, **80**, 6704–6713.
- 10 R. Francke, G. Schnakenburg and S. R. Waldvogel, *Eur. J. Org. Chem.*, 2010, **12**, 2357–2362.
- 11 L. Zehnder, M. Bennett, J. Meng, B. Huang, S. Ninkovic, F. Wang and P.-P Kung, *J. Med. Chem.*, 2011, **54**, 3368–3385.
- 12 S. N. Joshi, S. M. Vyas, H. Wu, M. W. Duffel, S. Parkin and H.-J. Lehmler, *Tetrahedron*, 2011, **67**, 7461–7469.
- 13 L. Tietze, C. Vock, I. Krimmelbein and L. Nacke, *Synthesis*, 2009, **12**, 2040–2060.
- 14 A. M. Bender, N. W. Griggs, C. Gao, T. J. Trask, J. R. Traynor and H. I. Mosberg, *ACS Med. Chem. Lett.*, 2015, **6**, 1199–1203.
- 15 M. Bielawski, D. Aili and B. Olofsson, *J. Org. Chem.*, 2008, **73**, 4602–4607.
- 16 N. Jalalian, E. E. Ishikawa, L. F. Silva & B. Olofsson, *Org. Lett.*, 2011, **13**, 1552–1555.

A.2 Approval letter from publisher

Approval letter from publisher of the publication which is included in chapter 3.1:

- C. Greunke*, E. R. Duell*, P. M. D'Agostino*, A. Glöckle, K. Lamm, T. A. M. Gulder: Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Met. Eng.* **2018**, 47, 334–345, DOI 10.1016/j.ymben.2018.03.010
*equally contributing authors

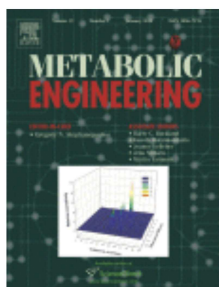


RightsLink®

Home

Account
Info

Help



Title: Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential
Author: Christian Greunke, Elke Regina Duell, Paul Michael D'Agostino, Anna Glöckle, Katharina Lamm, Tobias Alexander Marius Gulder

Publication: Metabolic Engineering

Publisher: Elsevier

Date: May 2018

© 2018 International Metabolic Engineering Society.
Published by Elsevier Inc. All rights reserved.

Logged in as:
Elke Duell
Technical University of Munich
Account #:
3001271763

LOGOUT

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

Copyright © 2018 Copyright Clearance Center, Inc. All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

An explicit approval letter from the publisher of the publication included in chapter 3.2 is not necessary as the following work was published with open access:

- **E. R. Duell***, P. M. D'Agostino*, N. Shapiro, T. Woyke, T. M. Fuchs, T. A. M. Gulder: Direct pathway cloning of the sodorifen biosynthetic gene cluster and recombinant generation of its product in *E. coli*. *Microb. Cell Fact.* **2019**, 18, DOI 10.1186/s12934-019-1080-6

*equally contributing authors

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Acknowledgements

This work would not have been possible without several people whom I deeply want to thank hereinafter. I am very grateful to Professor Tobias Gulder for giving me the opportunity to join his work in such an interesting, important and innovative field of research. During the last four and a half years I enjoyed the productive and liberal atmosphere he created and especially his friendly, encouraging and positive way of guidance.

I also want to thank Professor Hubert Gasteiger for agreeing to chair my oral examination and Professor Michael Groll who is not only the second examiner for this thesis, but also gave me the opportunity to do my master's thesis in his group. The experience I made during this time, especially while working with Dr. Eva Huber, are mainly responsible for my decision to start my doctoral research.

The Deutsche Bundesstiftung Umwelt generously funded my studies for three years for which I am very thankful. They not only continued to trust in me and my work but also allowed me to discover a variety of interesting ideas and gather a great many new experiences. What I learned during my time with the DBU will definitely continue to inspire my future way of life.

Furthermore I would like to thank all former and current members of the T.A.M. Gulder group with whom I shared helpful and constructive as well as casual off-topic discussions and many more enjoyable moments. My deepest thank goes to Dr. Christian Greunke who not only helped me to feel welcome in this group of sworn involuntary Rhineland refugees, but also always was willing to teach me new ways of thinking and doing and who discussed uncountable failed experimental attempts with me to maybe find a solution in the end. I am proud to have played a role in the development of the DiPaC method which was initiated by him. I want to thank Paul D'Agostino for teaching me everything I needed to know for my work with cyanobacteria and with whom I worked together on the DiPaC and the sodorifen project. Michaela Geissler supported my work on ambigols during her master's thesis in our group for which I am very thankful. Beyond that, Tobias Milzarek passionately devoted the first year of his PhD studies to shed light on the ongoing during the ambigol biosynthesis by developing a synthetic access to the molecules I have been biochemically chasing for some years now. I am incredible grateful for his help, without

which I could not have come so far. Furthermore I am also very thankful to my later bench and desk neighbor Katharina Lamm, with whom I shared so many positive and negative thoughts and on whose help I could always trust.

During my time at this university, I met some very warmhearted and kind people, whose friendship I do not want to miss anymore. I am very thankful for Susanne Meier whom I already met during my undergraduate studies and with whom I could share so much more than just biochemistry ever since. I am definitely looking forward to all of the joint dinners and other doings still to come. Additionally I am grateful to Marie-Kristin von Wrisberg and Susanne Mayer, who, both in their own kind and unique way, made the last years not only endurable but special and enjoyable for me.

Beyond the crazy scientific world I could furthermore always trust on true friends who were there when I needed them. I have known Magdalena Kick for almost 25 years now and I hope our friendship will last long enough to let my vision of two wrinkled old ladies happily chatting on a bench somewhere in nowhere come true. Sabrina Englbrecht is definitely the person who has visited me by far the most since I moved to Neufahrn to start my studies in 2009 and I think we share a lot of wonderful memories, to which I really want to add many more in the future. I am furthermore much obliged to all of my friends, including the members of the "Anonymous Heinz Erhardt Fanklub", who always stood by my side in various ways and never failed to cheer me up. I want to thank all of you for your amicable support during the past years.

Still, all of this would not have been possible without the incredibly strong and loving help from my family. My parents Regina and Jürgen raised me to always believe in myself and my capabilities and who did not get tired to lift my spirits again and again and empowered me time after time until I finally made it to this point on my own. I am deeply grateful for your support in so many countless ways! Furthermore I want to thank my brother Richard for all the good times we had together in the last 27 years. I am also very grateful to my whole family - the one I was born into as well as the one I was gifted with by marriage - for what all of you have done for me to make this possible.

With this I am at a point where words alone can no longer describe my thankfulness and gratefulness to the one person without whom I surely could not have made it this far. You are the one who is stuck with me for more than 12 years now, who has to put up with me and my moody ways each and every day, who has seen me at my best and at my worst and who still decided to stay with me. Max, I thank you so much for always being there for me, even in times where you yourself did not know where up and down was. I cannot wait to start the next chapter of our life together and I hope our intertwined story never ends!

