GLOBAL METABOLOMICS-BASED IDENTIFICATION OF SMALL-MOLECULE SIGNALS THAT REGULATE DEVELOPMENT IN NEMATODES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Neelanjan Bose

January 2014

GLOBAL METABOLOMICS-BASED IDENTIFICATION OF SMALL-MOLECULE SIGNALS THAT REGULATE DEVELOPMENT IN NEMATODES

Neelanjan Bose, Ph.D.

Cornell University 2014

Small-molecule signaling serves important functions at all levels of organismal organization and requires diverse biosynthetic mechanisms for encoding biological information in chemical structures. Whereas fungi and bacteria have dedicated biosynthetic machinery that enables production of a great diversity of chemical structures, e.g. polyketides and non-ribosomal peptides, most animals are presumed to lack the ability to produce elaborate small-molecule architectures.

Herein, the author describes the integrated use of 2D NMR and high resolution HPLC-MS/MS to aid global metabolomics of complex natural samples. Applications of this methodology to metabolomes of *Caenorhabditis elegans* and *Pristionchus pacificus* showed that these nematodes generate a library of complex signaling molecules, ascarosides and paratosides, via selective assembly of building blocks from several primary metabolic pathways, including an unusual xylopyranose-based nucleoside. These compounds act as interorganismal signals controlling larval development, adult morphology, or function as potent attraction pheromones. These findings further indicate species-specific evolution of chemical signaling in nematodes, with regard to both chemical structures and their biological functions. The library of small molecule signals presented in this dissertation, provide striking examples for combinatorial

generation of structural diversity in nematodes and connect primary metabolism to regulation of development and adult phenotypic plasticity.

Further the author used comparative metabolomics to investigate ascaroside biogenesis in *C. elegans*. Profiling ascarosides in *C. elegans* wild-type and peroxisomal β-oxidation mutant metabolomes via HPLC-MS/MS and 2D NMR clarified the functions of the acyl-CoA-oxidase, ACOX-1, and the β-ketoacyl-CoA thiolase, DAF-22 in ascaroside biogenesis. Following peroxisomal β-oxidation, the ascarosides are selectively derivatized with moieties of varied biogenetic origin that can dramatically affect biological activity.

Finally, using a 2D NMR-based comparative metabolomics approach, the author identified the endogenous ligands of the *C. elegans* nuclear hormone receptor (NHR), DAF-12, a vitamin D and liver X receptor homolog that regulates larval development, fat metabolism, and lifespan. The identified molecules include only one of two previously predicted DAF-12 ligands and feature unusual structural motifs, e.g. a Δ^1-desaturated steroid. These results demonstrate the advantages of comparative metabolomics over traditional candidate-based approaches and provide a blueprint for the identification of ligands for other *C. elegans* and mammalian NHRs.

BIOGRAPHICAL SKETCH

Neelanjan Bose was born in Kolkata, India to Drs. Anjan and Santa Bose. Neelanjan attended the Ramakrishna Mission Vidyalaya, Narendrapur where he realized his passion for chemistry. He then attended the prestigious Presidency College, Kolkata to receive his bachelor's of science (hons.) degree in Chemistry.

Next, Neelanjan took admission in the Master of Science (2 year) program at the Indian Institute of Technology (IIT), Kanpur, India via a national level competitive examination. At IIT, Kanpur he worked in the research group of Prof. Parimal K. Bharadwaj on solvothermal syntheses and X-ray crystallographic study of 3D-metal organic frameworks (3D-MOFs). In the summer of 2007 he worked as a summer intern in the lab of Prof. Soumen Basak, Saha Institute of Nuclear Physics, Kolkata, India. Here he studied the effects of 2,2,2-trifluororoethanol on the structure and aggregation of reduced κ-caesin using transmission electron microscopy (TEM), circular dichroism (CD), and fluorescence spectroscopy. He loved this experience and realized the potential of using chemistry techniques for solving problems in biology.

To pursue his dreams he came all the way to the US to join the Cornell/Rockefeller/Sloan-Kettering Tri-Institutional Training Program in Chemical Biology. He decided to stay in Ithaca and joined Prof. Frank C. Schroeder's research group and has worked on developing new techniques to study small-molecule signaling in nematodes.

Dedicated to

the leading ladies of my life

my mother and my wife

ACKNOWLEDGMENTS

Special Committee Chair and Adviser: Prof. Frank C. Schreoder

Special Committee: Profs. Brian Crane, Siu Sylvia Lee, and Sean F. Brady

Collaborating Principal Investigators: Profs. Ralf J. Sommer, Paul W. Sternberg, Adam Antebi, Patrik J. Hu, Arthur Edison, Judith A. Appleton, and Jonathan Ewbank

Collaborating Researchers: Dr. Stephan H. von Reuss, Dr. Axel Bethke, Joshua J. Yim, Joshua C. Judkins, Parag Mahanti, Yevgeniy Izrayelit, Dr. Akira Ogawa, Dr. Erik J. Ragsdale, Dr. Cameron Weadick, Jan M. Meyer, Dr. Gabriel Markov, Dr. Jagan Srinivasan, Dr. Joshua Wollam, Dr. Kathleen J. Dumas, Dr. OIivier Zugasti

Facility Managers: Dr. Ivan Keresztes, Anthony M. Condo, and Maciej Kukula.

Organizations: Boyce Thompson Institute for Plant Research and Cornell University, Ithaca, NY.

Funding: Tri-Institutional Training Program in Chemical Biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Nematodes as model systems for studying small molecule signaling: Nematodes or roundworms are simple, fully differentiated, multi-cellular organism group that comprise the most speciated phylum in the animal kingdom.[1] Most of the nematode species described are parasitic, many of which infect about 25% of the world's population and significantly impact agricultural crops and animals.[2,3] Several nematode species, both free-living and parasitic, have been established as laboratory model systems for studying various aspect of biology, medicine, and pest control.[4-12] Of these, the most studied and well-established is the free living soil nematode *Caenorhabditis elegans*.

C. elegans, among the first organisms whose genomes were sequenced,[13] is small enough for high-throughput whole-organism screens, yet fully differentiated. Many of signaling pathways elucidated in *C. elegans* show strong analogies or are directly homologous to corresponding pathways in higher animals.[14] This high level of genetic conservation allows ancient features of endocrine pathways to be explored in *C. elegans*. Significantly, *C. elegans* research has also produced interesting insights into the signaling pathways that control metabolism, growth, reproductive maturation, and lifespan regulation, which revealed a deeply intertwined regulatory network that remains, at best, partly understood.[15-24] Considerable knowledge of disease pathways

for Alzheimer's disease,[25] diabetes,[26] and depression[27] has emerged through *C. elegans* research. More recently, other nematode species, most notably the free-living necromenic roundworm *Pristionchus pacificus* have been established as satellite model systems for the study of evolutionary and developmental biology.[4] Satellite model organisms are species that are sufficiently closely related to major model organisms so

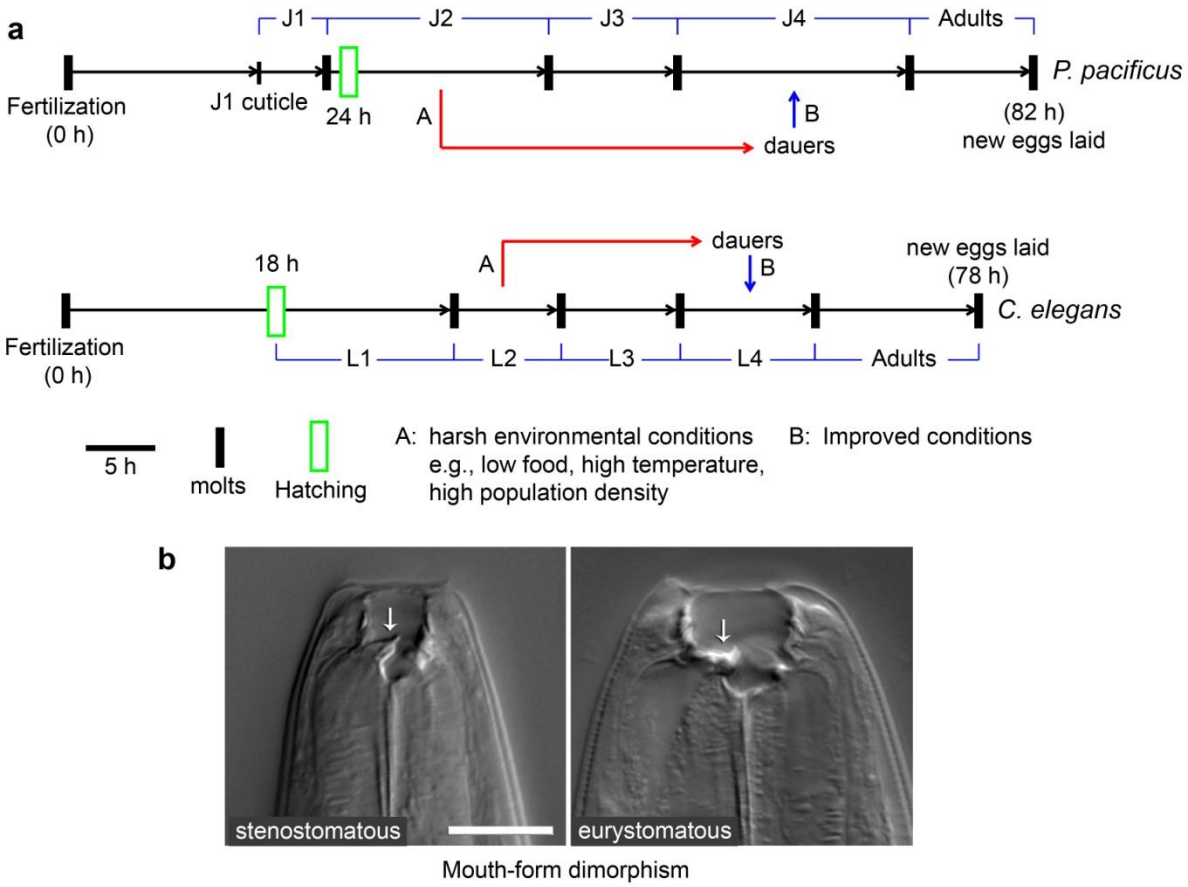


Figure 1.1: (**a**) Life cycle of *C. elegans* and *P. pacificus* at 20 °C. Nematodes propagate through four larval/juvenile stages, called L1/J1 to L4/J4 in *C. elegans* and *P. pacificus*. In contrast to *C. elegans*, for *P. pacificus* the J1 to J2 molt is embryonic and only the J2 stage hatches from the egg. Under harsh environmental conditions a L2/J2 larva progresses into an alternate noon-feeding, seemingly non-aging stage of developmental arrest, caller dauer (figure adopted from Ref. 29). (**b**) Dimorphism of the buccal cavity observed in *P. pacificus*: narrow/stenostomatous (left) and wide/eurystomatous (figure adopted from Ref. 43).

that the genetic regulation of homologous biosynthetic, cellular, and developmental processes can be studied, enabling the identification of the molecular changes that underlie phenotypic and biochemical differences or variation.[28]

In favorable laboratory conditions both *C. elegans* and *P. pacificus* develop from eggs and mature through four larval stages, named L1-L4 (*C. elegans*) or J1-J4 (*P. pacificus*) before molting into reproductive adults (**Figure 1.1a**).[29] However, under unfavorable conditions such as lack of food, high temperatures, and high population density, L2/J2 larva enter a stage of developmental arrest called the dauer (from German 'dauer' for 'enduring', **Figure 1.1a**), instead of developing into normal L3/J3 worms.[30-34] Dauer larvae are non-feeding and can stay alive for many months until conditions improve, much longer than the typical adult lifespan of 2-3 weeks. Dauer development represents a unique example of developmental plasticity in metazoans.

Unlike *C. elegans, P. pacificus* belong to the family Diplogastridae, and like all species of this family, the first embryonic molt (J1 larvae) is not free-living and non-feeding and molt to J2 before they hatch from the egg (**Figure 1.1a**). In the wild, *P. pacificus* forms an unusual association to scarab beetles: residing on the beetles as dauers for the most part and resuming development by feeding on the microbes that infest the beetle carcass. Such a necromenic association with beetles may represent a pre-adaptation to the evolution of true parasitism.[35] In addition to dauer formation, *P. pacificus* further exhibits a dimorphism in mouth development: adult worms can either have a "stenostomatous"/narrow- or a "eurystomatous"/wide-buccal cavity, the latter developing in response to conditions of low food availability (**Figure 1.1b**).[36]

Genomics and proteomics are highly developed for *C. elegans* and developing rapidly in *P. pacificus.*[29] However, comprehensive structural and functional characterization of these model nematode's metabolomes has been explored only to a very limited extent. Recent studies have shown that *C. elegans* utilizes small-molecule architectures, namely ascarosides (**Figure 1.2a**) of unanticipated diversity and complexity and bile acid-like steroids, called dafachronic acids (**Figure 1.2b**) in exocrine and endocrine signaling.[34,37-48] These molecules play a central role in the nematode's life history by regulating developmental timing, lifespan, stress resistance, phenotypic plasticity, and a wide range of social behaviors. Of these, the ascarosides form a modular library that integrates building blocks derived from conserved primary metabolic pathways in which different combinations of "modules" are associated with different signaling contexts.

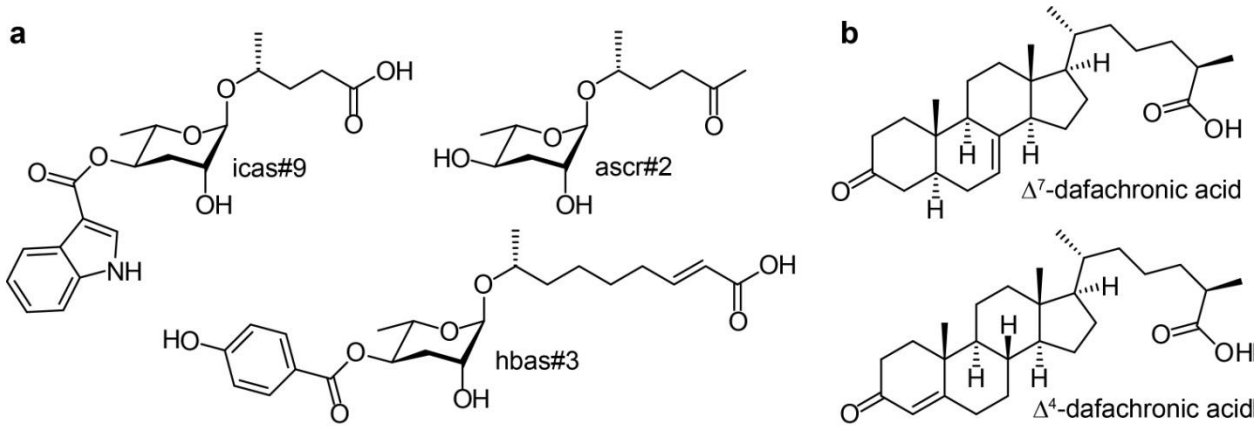


Figure 1.2: Structures of two types of signaling molecules identified from *C. elegans* - (**a**) ascarosides, which form a modular library of small-molecule signals, e.g. ascr#2, icas#9, and hbas#3 and (**b**) bile-acid like steroids, called dafachronic acids (DAs), e.g. Δ^7- and Δ^4-DA.

Research efforts at the Schroeder group suggest that a vast majority of *C. elegans* and *P. pacificus* metabolites remain unidentified. The knowledge gained from their structural elucidation and functional characterization will greatly advance understanding of the biology of these model systems, in particular signaling processes that regulate lifespan and conserved primary metabolism as well as associated disease-relevant pathways in mammals. Thus there exists a clear rationale for developing *C. elegans* and *P. pacificus* as model organisms for small molecule signaling and metabolism.

1.2. Developing a global metabolomics approach for investigating nematode small molecule signaling: Small molecules typically are defined as organic, non-polymeric chemical entities with molecular mass <1000 Daltons. Biogenic small molecules serve a variety of biological functions, such as facilitating communication, chemical defense, and predation; functioning as hormones and second messengers in animals and plants; as well as serving as building blocks for biological macromolecules.[49]

"Metabolomics" is the systematic study for identification and quantitation of biogenic small molecules in mixtures, extracted from a complex biological matrix (such as a cell, tissue, organ or an organism), representing a specific metabolic state.[50] A related field of study, "natural product chemistry" aims at isolation and structural characterization of small molecules with specific biological or otherwise desirable activities from complex natural samples.[50] Even though the two fields have overlapping

objectives - structural and functional characterization of metabolites from complex natural mixtures - there exists a basic difference in experimental design and techniques used. Natural products research traditionally relies on activity guided fractionation to isolate or highly enrich individual compounds; whereas metabolomics utilizes statistical methods to identify a subset of small molecules relevant in a specific biological context.[50] Traditionally, metabolomics primarily uses mass spectrometry-based metabolite profiling of complex mixtures prior to the application of statistical techniques.[51,52] On the other hand, Nuclear Magnetic Resonance (NMR) spectroscopy has been traditionally employed for unambiguous structural assignment of pure or highly enriched individual compounds as in natural products chemistry, and only rarely for broad metabolite profiling.[53-55] Each field of study has its limitations and biases: metabolomics suffers due to the absence of in-depth structural information from mass spectrometry based metabolite profiling; and natural products is biased towards the identification of only stable metabolites that can handle multiple rounds of chromatography used to purify the active compound(s).

Groundbreaking contributions by Schroeder et al. demonstrated that, contrary to the accepted notion, application of 2D NMR spectroscopy on crude biological samples can provide a broad overview of the composition (similar to mass spectrometry), as well as reveal full or partial structural information of the metabolites therein.[56,57] These initial studies were done on crude metabolite extracts from *Myrmicaria* ants utilizing double quantum filtered correlation spectroscopy (dqfCOSY, a 2D NMR spectroscopy experiment) to identify novel oligocyclic alkaloids. The use of 2D NMR spectroscopy also proved extremely useful for the characterization of labile small molecule products

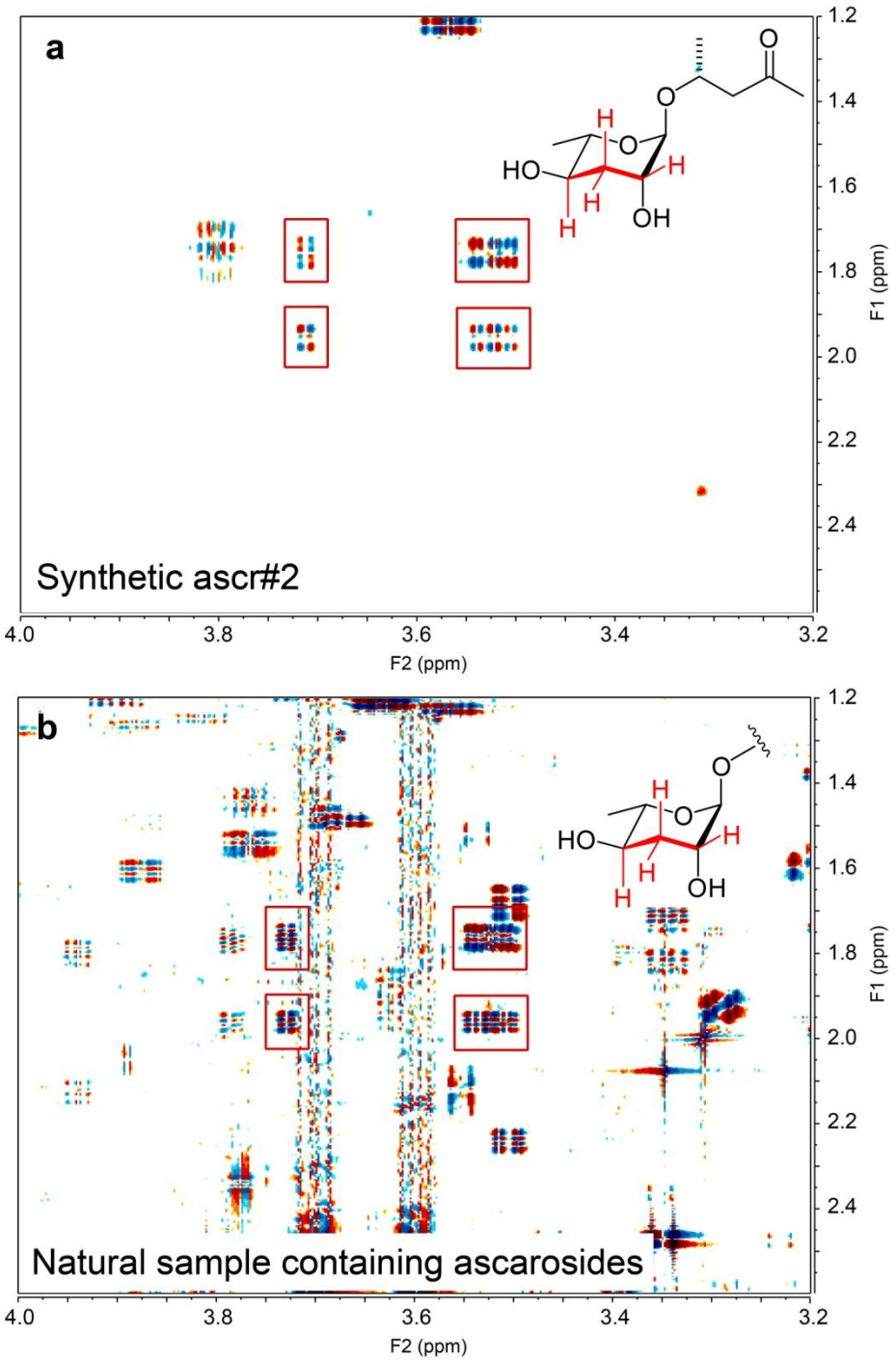


Figure 1.3: Sections of dqfCOSY spectra of (**a**) synthetic ascr#2 and (**b**) a natural sample containing ascarosides. Characteristic crosspeaks arising due to the spin system of an ascarylose ring are boxed red. These crosspeaks can be discerned even from a very complex spectrum of a natural sample containing thousands of compounds by the analysis of the associated chemical shifts and fine-structures.

of the orphan gene cluster *pksX* from *B. subtilis* as it helped to eliminate detrimental chromatography steps.[58] In the subsequent years the Schroeder group at Cornell University has taken the lead in combining ideas from both metabolomics and natural products chemistry to overcome the biases of the two fields for structural and functional characterization of small molecules. The group has pioneered the use of 2D NMR-based comparative metabolomics tools such as: DANS (Differential Analysis by 2D NMR Spectroscopy) which greatly aided the linking of small molecule metabolites with their functions and biogenesis.[29,56,57] Further, mvaDANS (multivariate DANS) was developed that enabled automated processing and comparative computational analysis of 2D NMR spectra.[37]

Central to the use of 2D NMR as a metabolomics tool is the use of the dqfCOSY experiment. dqfCOSY is a high-resolution, high dynamic range, ^{1}H-^{1}H, 2D NMR experiment that displays an array of crosspeaks symmetrically placed across a diagonal. Each crosspeak marks scalar *J*-coupling-based interaction of two different classes of protons with different chemical shift values. In-depth analysis of dqfCOSY crosspeak fine-structures (**Figure 1.3**) provides accurate chemical shifts and scalar *J*-coupling values for the associated protons, which are indispensible for assigning molecular structures. In addition, dqfCOSY spectra have a high dynamic range (~200:1 or more) and are able to capture correlations, i.e. show crosspeaks, for a large number of different molecular species, as typically present in a crude natural small-molecule extract. Furthermore, interconnected sets of crosspeaks (with their associated fine-structures) in a dqfCOSY spectrum represent as unique identifiers (fingerprints) of particular structural features, which are often discernible even from highly complex

mixtures (**Figure 1.3**). In summary, dqfCOSY can provide an unbiased overview of the metabolome's composition and frequently allows detection and partial identification of even very minor metabolites.[59-61]

Inspired by the advancement of 2D NMR-based metabolomics, this dissertation reports the use of a "global metabolomics" approach - an innovative analytical workflow that utilizes a combination of 2D NMR (predominantly dqfCOSY) and high resolution HPLC-MS/MS; activity guided fractionation, and chemical synthesis to achieve broad metabolite profiling of complex biological matrices. This technique was successfully used for the identification, functional characterization, and studying the biogenesis of novel secondary metabolites from the model nematodes *C. elegans* and *P. pacificus*.[37-40,43] This integrative metabolomics approach utilized in this dissertation largely overcomes the biases of the fields of metabolomics and natural products chemistry. However, this methodology suffers to some extent from the limitations of the analytical techniques used and specific aspects of experimental design. To list a few: (1) the primary 2D NMR technique used, dqfCOSY, only provides structural information of small-molecules that exhibit ^{1}H-^{1}H scalar coupling(s), hence molecules with inter-connected quaternary carbons or exchangeable ^{1}Hs will not be observed; (2) the HPLC-MS/MS technique used only record and provide data for molecules that ionize under a specified set of experimental conditions; and (3) the extraction protocols used for the most part may not be suitable for highly polar compounds (e.g. nucleotides).

1.3. Small molecule signaling in nematodes: Small molecule signaling serves important functions at all levels of organismal organization and requires diverse biosynthetic mechanisms for encoding biological information in chemical structures.[62] Whereas microorganisms have well-studied biosynthetic machinery that enables production of a great diversity of structures (e.g. non-ribosomal peptides and polyketides),[63,64] most animals were presumed to lack the ability to produce elaborate small-molecule architectures. However, recent studies showed that *C. elegans,* a fully differentiated multi-cellular metazoan produces a modular library of complex small molecule signals, called ascarosides (**Figure 1.2a**) that regulate nematode development,[33,45,48,65] act as sex pheromone,[34,47] and mediate aggregation behavior.[39,40] In addition, steroid hormones, called dafachronic acids (Δ^4- and Δ^7-DA, **Figure 1.2b**) were identified, which act as ligands for the nuclear hormone receptor DAF-12 that regulates both adult lifespan and larval development in *C. elegans*.[18,65,66]

Genetic dissection of ascarosides and dafachronic signaling in *C. elegans* has revealed a complex interconnected network of signal transduction pathways, reviewed in Ref. 20 (**Figure 1.4**). Cues from the environment in the form of nutrient levels, temperature, and small molecule pheromones are perceived by GPCRs in sensory neurons of the worm.[46,67,68] Signals from the GPCRs are transduced via G-proteins and membrane resident guanylyl cyclase, DAF-11[69] to cGMP-gated ion-channels, TAX-2, -4 on the neuron membrane.[70,71] When conditions are favorable, DAF-7/TGF-β and insulin-like peptides (ILPs) are produced. Binding of ILPs and TGF-β to their cognate cell surface receptors DAF-2 (Insulin/IGF-1 receptor) and DAF-1, -4, respectively, results in nuclear localization of DAF-8,-14/SMADs and retention of phosphorylated DAF-

16/FOXO in the cytoplasm.[20,72-80] These events are believed to directly or indirectly regulate expression of putative hormone synthesis enzymes (e.g., DAF-9/CYP) that produce ligands for the nuclear hormone receptor DAF-12/NHR, a homolog of mammalian vitamin D and Liver-X receptors. Ligand-bound DAF-12 putatively recruits

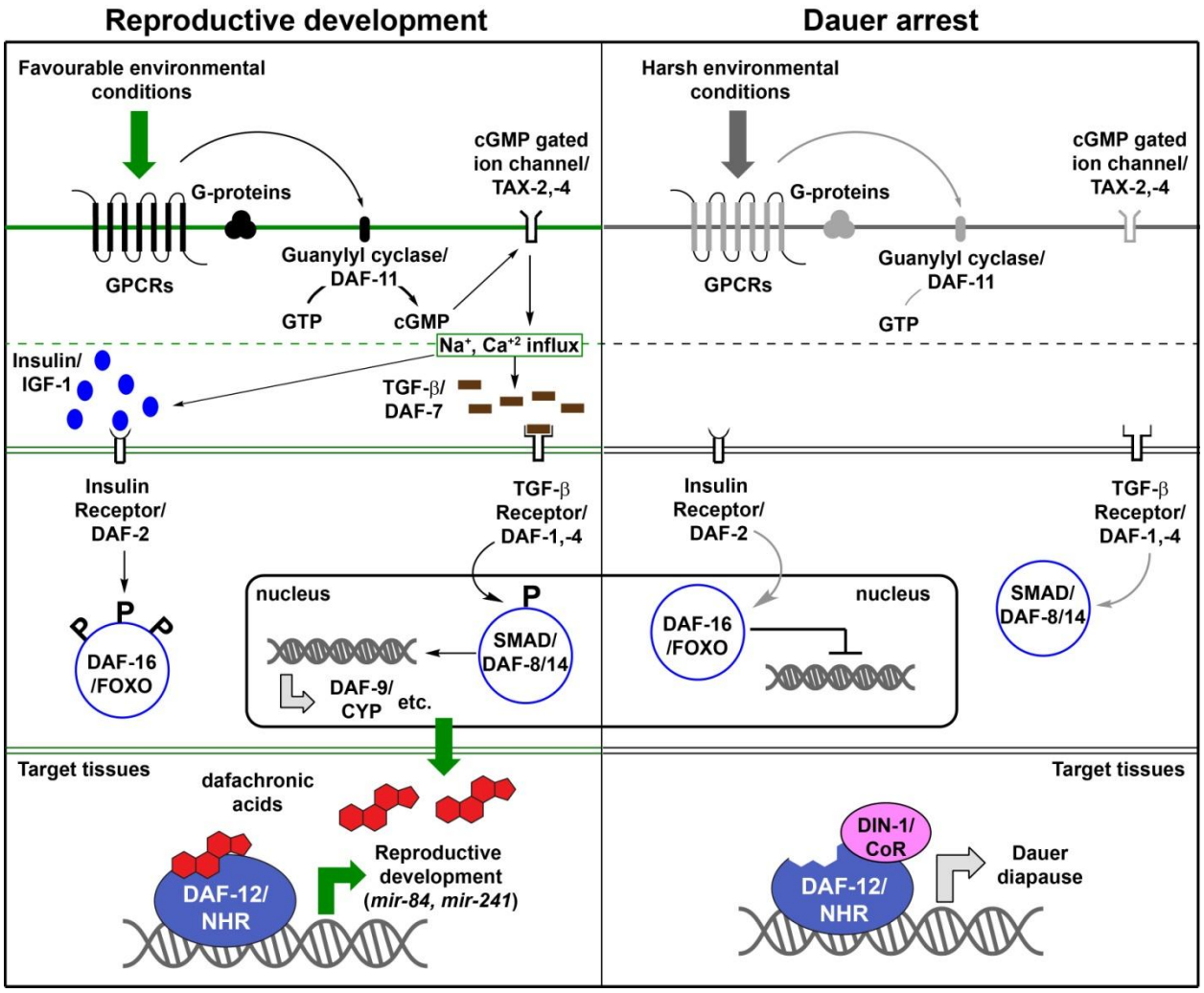


Figure 1.4: Schematic model for small molecule-signaling in *C. elegans*, reviewed in Ref. 20. Environmental signals (e.g. small molecule pheromones) are perceived by GPCRs and transduced via evolutionarily conserved signaling cascades, including the: guanylyl cyclase, TGF-β, and insulin/IGF-1 pathways. When environmental conditions are favorable, upstream signaling results in expression of genes required for the biosynthesis of steroids ligands (dafachronic acids) of DAF-12/NHR. Liganded DAF-12 promotes development, in part via transcription of the *let-7*-family microRNAs *mir-84* and *mir-241*. Under unfavorable conditions, ligand biosynthesis is inhibited, resulting in interaction of unliganded DAF-12 with its corepressor DIN-1 and target gene repression for dauer formation.

yet unidentified coactivators and promotes reproductive development. Upon sensing inclement conditions, Insulin/IGF-1 and TGF-β signaling shut off and DAF-16/FOXO is nuclear localized to promote dauer programs and directly or indirectly inhibit expression of DAF-12-ligand biosynthesis genes. In contrast, unliganded DAF-12 promotes dauer programs through association with its corepressor DIN-1/SHARP.[81-83] **Figure 1.4** is a simple scheme to represent the central players that transduce small molecule signals in nematodes. The pathway does not necessarily work in sequence as represented, can work in parallel, and have independent outputs.

1.4. Dissertation outline: The central theme of this dissertation is the use of global metabolomics of complex nematode-derived mixtures, encompassing orthogonal analytical data sets from 2D NMR and high resolution HPLC-MS/MS. At the time when the author started his graduate research, only a few of examples of signaling molecules were known in *C. elegans* and none were reported from *P. pacificus*. The few that were identified relayed that small molecule signaling is indispensible and associated with many distinct phenotypes in nematodes.[31-34,47,84,85] Hence, these model nematodes stood extremely suitable systems to explore metazoan small molecule signaling by the application of an innovative analytical workflow.

In the subsequent chapters, the author extends the understanding of nematode-derived small molecules that functions from conferring environmental cues to the downstream regulation of NHR activity (**Figure 1.4**). Structural and functional characterization of small molecule signals in nematodes described in this dissertation

provide additional routes to investigate evolutionarily conserved signaling pathways, such as guanylyl cyclase, TGF-β, and Insulin/IGF-1 that are implicated in many grave disease such as: cancer, diabetes, and atherosclerosis.[86-88] Many of the identified signaling molecules feature unusual structural complexities, such as xylopyranose-based nucleosides, uncharacteristic of developed metazoans. However, these molecules are generated via simple combinatorial assembly of building blocks derived from primary metabolism that are essentially available to all organisms. This suggests that higher metazoans may have under-appreciated biosynthetic capabilities and thus provide an opportunity to re-investigate mammalian metabolome using techniques from global metabolomics approach described in this dissertation.

REFERENCES

(1) Coghlan, A. *WormBook : the online review of C. elegans biology* **2005**, 1.

(2) Jasmer, D. P.; Goverse, A.; Smant, G. *Annual review of phytopathology* **2003**, *41*, 245.

(3) Perry, R. N.; Moens, M. **2011**, 3.

(4) Hong, R. L.; Sommer, R. J. *Bioessays.* **2006**, *28*, 651.

(5) Sommer, R. J.; McGaughran, A. *Molecular ecology* **2013**, *22*, 2380.

(6) Opperman, C. H.; Bird, D. M. *Current opinion in plant biology* **1998**, *1*, 342.

(7) Girard, L. R.; Fiedler, T. J.; Harris, T. W.; Carvalho, F.; Antoshechkin, I.; Han, M.; Sternberg, P. W.; Stein, L. D.; Chalfie, M. *Nucleic acids research* **2007**, *35*, D472.

(8) Cox-Paulson, E. A.; Grana, T. M.; Harris, M. A.; Batzli, J. M. *CBE life sciences education* **2012**, *11*, 165.

(9) Shim, Y. H.; Paik, Y. K. *Proteomics* **2010**, *10*, 846.

(10) Gaertner, B. E.; Phillips, P. C. *Genetics research* **2010**, *92*, 331.

(11) Kaletta, T.; Hengartner, M. O. *Nature reviews. Drug discovery* **2006**, *5*, 387.

(12) Lok, J. B. *WormBook : the online review of C. elegans biology* **2007**, 1.

(13) The C. elegans Sequencing Consortium *Science* **1998**, *282*, 2012.

(14) Harris, T. W.; Chen, N.; Cunningham, F.; Tello-Ruiz, M.; Antoshechkin, I.; Bastiani, C.; Bieri, T.; Blasiar, D.; Bradnam, K.; Chan, J.; Chen, C.-K.; Chen, W. J.; Davis, P.; Kenny, E.; Kishore, R.; Lawson, D.; Lee, R.; Muller, H.-M.; Nakamura, C.; Ozersky, P.; Petcherski, A.; Rogers, A.; Sabo, A.; Schwarz, E. M.; Van Auken, K.; Wang, Q.; Durbin, R.; Spieth, J.; Sternberg, P. W.; Stein, L. D. *Nucl. Acids Res.* **2004**, *32*, D411.

(15) Wang, M. C.; O'Rourke, E. J.; Ruvkun, G. *Science* **2008**, *322*, 957.

(16) O'Rourke, E. J.; Kuballa, P.; Xavier, R.; Ruvkun, G. *Genes Dev* **2013**, *27*, 429.

(17) Wollam, J.; Magner, D. B.; Magomedova, L.; Rass, E.; Shen, Y.; Rottiers, V.; Habermann, B.; Cummins, C. L.; Antebi, A. *Plos Biol* **2012**, *10*, e1001305.

(18) Shen, Y.; Wollam, J.; Magner, D.; Karalay, O.; Antebi, A. *Science* **2012**, *338*, 1472.

(19) Lee, S. S.; Schroeder, F. C. *PLoS Biol* **2012**, *10*, e1001307.

(20) Fielenbach, N.; Antebi, A. *Genes & development* **2008**, *22*, 2149.

(21) Lucanic, M.; Held, J. M.; Vantipalli, M. C.; Klang, I. M.; Graham, J. B.; Gibson, B. W.; Lithgow, G. J.; Gill, M. S. *Nature* **2011**, *473*, 226.

(22) Ackerman, D.; Gems, D. *BioEssays* **2012**, *34*, 466.

(23) Kenyon, C. J. *Nature* **2010**, *464*, 504.

(24) Watson, E.; MacNeil, L. T.; Arda, H. E.; Zhu, L. J.; Walhout, A. J. *Cell* **2013**, *153*, 253.

(25) Wittenburg, N.; Eimer, S.; Lakowski, B.; Rohrig, S.; Rudolph, C.; Baumeister, R. *Nature* **2000**, *406*, 306.

(26) Ogg, S.; Paradis, S.; Gottlieb, S.; Patterson, G. I.; Lee, L.; Tissenbaum, H. A.; Ruvkun, G. *Nature* **1997**, *389*, 994.

(27) Ranganathan, R.; Sawin, E. R.; Trent, C.; Horvitz, H. R. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **2001**, *21*, 5871.

(28) Dieterich, C.; Roeseler, W.; Sobetzko, P.; Sommer, R. J. *Nucleic acids research* **2007**, *35*, D498.

(29) Sommer, R. J. *WormBook : the online review of C. elegans biology* **2006**, 1.

(30) Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y. *Proc Natl Acad Sci U S A.* **2009**, *106*, 1875.

(31) Butcher, R. A.; Ragains, J. R.; Kim, E.; Clardy, J. *Proc Natl Acad Sci U S A.* **2008**, *105*, 14288.

(32) Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J. *Nature chemical biology* **2007**, *3*, 420.

(33) Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K. *Nature.* **2005**, *433*, 541.

(34) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A.* **2009**, *106*, 7708.

(35) Rae, R.; Riebesell, M.; Dinkelacker, I.; Wang, Q.; Herrmann, M.; Weller, A. M.; Dieterich, C.; Sommer, R. J. *J Exp Biol* **2008**, *211*, 1927.

(36) von Lieven, A. F.; Sudhaus, W. *J Zool Syst Evol Res* **2000**, *38*, 37.

(37) Izrayelit, Y.; Robinette, S. L.; Bose, N.; von Reuss, S. H.; Schroeder, F. C. *ACS chemical biology* **2013**, *8*, 314.

(38) Artyukhin, A. B.; Yim, J. J.; Srinivasan, J.; Izrayelit, Y.; Bose, N.; von Reuss, S. H.; Jo, Y.; Jordan, J. M.; Baugh, L. R.; Cheong, M.; Sternberg, P. W.; Avery, L.; Schroeder, F. C. *J Biol Chem* **2013**, *288*, 18778.

(39) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc.* **2012**, *134*, 1817.

(40) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol.* **2012**, *10*, e1001237. Epub 2012 Jan 10.

(41) Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W. *Curr Biol* **2012**, *22*, 772.

(42) Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C. *ACS chemical biology* **2012**, *7*, 1321.

(43) Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C. *Angew Chem Int Ed Engl* **2012**, *51*, 12438.

(44) Kaplan, F.; Alborn, H. T.; von Reuss, S. H.; Ajredini, R.; Ali, J. G.; Akyazi, F.; Stelinski, L. L.; Edison, A. S.; Schroeder, F. C.; Teal, P. E. *PloS one* **2012**, *7*, e38735.

(45) Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J. *Nat Chem Biol.* **2007**, *3*, 420.

(46) Park, D.; O'Doherty, I.; Somvanshi, R. K.; Bethke, A.; Schroeder, F. C.; Kumar, U.; Riddle, D. L. *Proc Natl Acad Sci U S A* **2012**, *109*, 9917.

(47) Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E. A.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *Nature* **2008**, *454*, 1115.

(48) Ludewig, A. H.; Izrayelit, Y.; Park, D.; Malik, R. U.; Zimmermann, A.; Mahanti, P.; Fox, B. W.; Bethke, A.; Doering, F.; Riddle, D. L.; Schroeder, F. C. *Proc Natl Acad Sci U S A* **2013**, *110*, 5522.

(49) Meinwald, J. *Journal of Natural Products* **2011**, *74*, 305.

(50) Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc Chem Res* **2011**.

(51) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass spectrometry reviews* **2007**, *26*, 51.

(52) Kuehnbaum, N. L.; Britz-McKibbin, P. *Chemical reviews* **2013**, *113*, 2437.

(53) Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. *Nature reviews. Drug discovery* **2002**, *1*, 153.

(54) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica; the fate of foreign compounds in biological systems* **1999**, *29*, 1181.

(55) Nicholson, J. K.; Wilson, I. D. *Nature reviews. Drug discovery* **2003**, *2*, 668.

(56) Schroder, F.; Franke, S.; Francke, W.; Baumann, H.; Kaib, M.; Pasteels, J. M.; Daloze, D. *Tetrahedron* **1996**, *52*, 13539.

(57) Schroder, F.; Sinnwell, V.; Baumann, H.; Kaib, M. *Chem Commun* **1996**, 2139.

(58) Butcher, R. A.; Schroeder, F. C.; Fischbach, M. A.; Straight, P. D.; Kolter, R.; Walsh, C. T.; Clardy, J. *Proc Natl Acad Sci U S A* **2007**, *104*, 1506.

(59) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A* **2009**, *106*, 7708.

(60) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Biol* **2011**, *15*, 38.

(61) Forseth, R. R.; Fox, E. M.; Chung, D.; Howlett, B. J.; Keller, N. P.; Schroeder, F. C. *J Am Chem Soc* **2011**, *133*, 9678.

(62) Meinwald, J. *J Nat Prod* **2011**, *74*, 305.

(63) Walsh, C. T. *Acc Chem Res.* **2008**, *41*, 4.

(64) Meier, J. L.; Burkart, M. D. *Chem Soc Rev.* **2009**, *38*, 2012.

(65) Motola, D. L.; Cummins, C. L.; Rottiers, V.; Sharma, K. K.; Li, T. T.; Li, Y.; Suino-Powell, K.; Xu, H. E.; Auchus, R. J.; Antebi, A.; Mangelsdorf, D. J. *Cell* **2006**, *124*, 1209.

(66) Antebi, A.; Yeh, W. H.; Tait, D.; Hedgecock, E. M.; Riddle, D. L. *Genes & development* **2000**, *14*, 1512.

(67) McGrath, P. T.; Xu, Y. F.; Ailion, M.; Garrison, J. L.; Butcher, R. A.; Bargmann, C. I. *Nature* **2011**, *477*, 321.

(68) Kim, K.; Sato, K.; Shibuya, M.; Zeiger, D. M.; Butcher, R. A.; Ragains, J. R.; Clardy, J.; Touhara, K.; Sengupta, P. *Science* **2009**, *326*, 994.

(69) Birnby, D. A.; Link, E. M.; Vowels, J. J.; Tian, H.; Colacurcio, P. L.; Thomas, J. H. *Genetics* **2000**, *155*, 85.

(70) Coburn, C. M.; Bargmann, C. I. *Neuron* **1996**, *17*, 695.

(71) Komatsu, H.; Mori, I.; Rhee, J. S.; Akaike, N.; Ohshima, Y. *Neuron* **1996**, *17*, 707.

(72) Ren, P.; Lim, C. S.; Johnsen, R.; Albert, P. S.; Pilgrim, D.; Riddle, D. L. *Science* **1996**, *274*, 1389.

(73) Kimura, K. D.; Tissenbaum, H. A.; Liu, Y.; Ruvkun, G. *Science* **1997**, *277*, 942.

(74) Schroeder, F. C. *ACS chemical biology* **2006**, *1*, 198.

(75) Sommer, R. J.; Ogawa, A. *Curr Biol* **2011**, *21*, R758.

(76) Georgi, L. L.; Albert, P. S.; Riddle, D. L. *Cell* **1990**, *61*, 635.

(77) Estevez, M.; Attisano, L.; Wrana, J. L.; Albert, P. S.; Massague, J.; Riddle, D. L. *Nature* **1993**, *365*, 644.

(78) Inoue, T.; Thomas, J. H. *Developmental biology* **2000**, *217*, 192.

(79) Lin, K.; Dorman, J. B.; Rodan, A.; Kenyon, C. *Science* **1997**, *278*, 1319.

(80) Kenyon, C.; Chang, J.; Gensch, E.; Rudner, A.; Tabtiang, R. *Nature* **1993**, *366*, 461.

(81) Ludewig, A. H.; Kober-Eisermann, C.; Weitzel, C.; Bethke, A.; Neubert, K.; Gerisch, B.; Hutter, H.; Antebi, A. *Genes & development* **2004**, *18*, 2120.

(82) Gerisch, B.; Antebi, A. *Development* **2004**, *131*, 1765.

(83) Mak, H. Y.; Ruvkun, G. *Development* **2004**, *131*, 1777.

(84) Butcher, R. A.; Ragains, J. R.; Clardy, J. *Organic letters* **2009**, *11*, 3100.

(85) Bento, G.; Ogawa, A.; Sommer, R. J. *Nature.* **2010**, *466*, 494.

(86) Gordon, K. J.; Blobe, G. C. *Biochimica et biophysica acta* **2008**, *1782*, 197.

(87) Blobe, G. C.; Schiemann, W. P.; Lodish, H. F. *The New England journal of medicine* **2000**, *342*, 1350.

(88) Moses, A. C. *Endocrine development* **2005**, *9*, 121.

CHAPTER 2

COMPARATIVE METABOLOMICS REVEALS BIOGENESIS OF ASCAROSIDES, A MODULAR LIBRARY OF SMALL MOLECULE SIGNALS IN *C. ELEGANS*

2.1. Introduction: Several different aspects of the life history of the nematode *Caenorhabditis elegans* are regulated by ascarosides, glycosides of the dideoxysugar ascarylose (**Figure 2.1**).[1-6] The ascarosides ascr#1−3 were originally identified as major components of the dauer pheromone, a population-density signal that promotes entry into an alternate larval stage, the non-feeding and highly persistent dauer diapause.[1,2,4,5,7] Entry into the dauer stage is mediated by several highly conserved signaling pathways, including insulin/IGF-1 signaling and transforming growth factor beta (TGF-β) signaling,[8-12] which contributed to interest in the ascarosides' structures, their biosynthesis, and their mode of action. More recent work showed that specific mixtures of ascarosides including the 4-aminobenzoic acid derivative ascr#8 act as strong male-specific attractants,[3,5] whereas ascarosides including a tryptophan-derived indole-3-carboxy moiety function as aggregation signals at femtomolar concentrations.[6]

The biosynthetic pathways that control specific assembly of the ascarylose, lipid side chain, and peripheral building blocks are largely unknown. Earlier work showed that worms carrying a mutation in the gene *daf-22* are defective in the biosynthesis of both the dauer pheromone and male-attracting signals.[3,5,13,14] *daf-22* encodes a protein

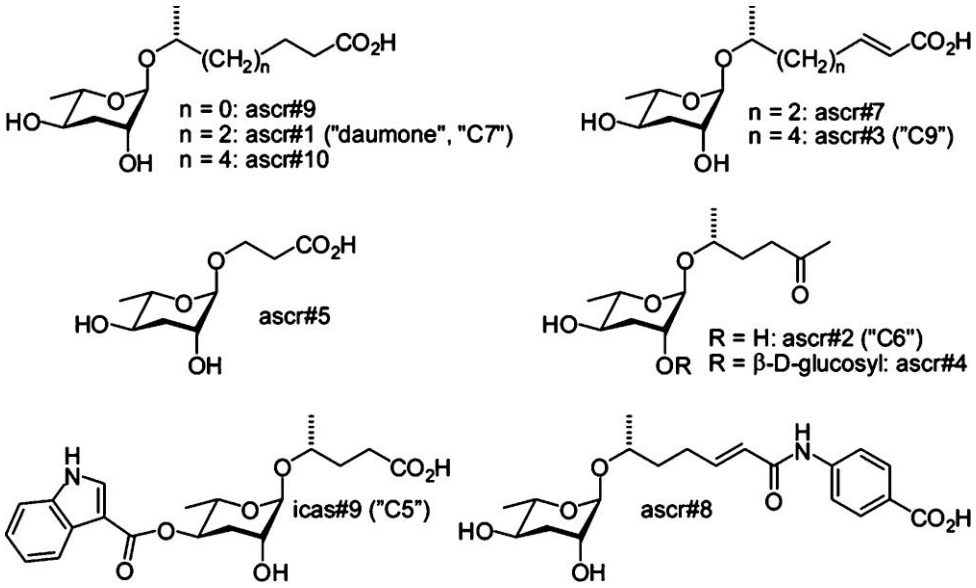


Figure 2.1: Ascarosides that regulate development and behavior in *C. elegans.* ascr#1-4, ascr#5 and icas#9 were identified via activity guided fractionation,[1,2,15] ascr#7 and ascr#8 were identified using differential analysis of 2D NMR spectroscopy (DANS),[5] and ascr#9 and #10 were detected using mass spectrometric techniques.[6]

with strong homology to human sterol carrier protein SCPx and in *C. elegans* functions in peroxisomal β-oxidation of long-chained fatty acids, producing the 3−9-carbon side chains of the ascarosides.[7] Two other components of peroxisomal β-oxidation were shown to participate in ascaroside biosynthesis: the acyl-CoA oxidase ACOX-1[16] and the β-hydroxyacyl-CoA dehydrogenase DHS-28,[7,17] a partial homologue of human multifunctional enzyme type 2 (MFE-2).

Among the studies that aimed at elucidating ascaroside biosynthesis, the publication by Joo et al. (2010),[16] is of particular importance for this dissertation chapter. In this paper, the authors proposed that the enzyme ACOX-1 acts as the acyl-CoA oxidase in peroxisomal β-oxidation pathway for ascaroside biosynthesis in *C. elegans*. Further, relative quantification of ascr#1-3 (for structures see **Figure 2.1**) at

temperatures 20 and 25 ^{o}C of wild-type and *acox-1(ok2257)* mutant, the authors concluded that *acox*-genes likely contribute to the dynamic balance of ascarosides in response to temperature variations. However, for this study the authors worked with a limited set of previously published ascarosides and did not venture into identifying shunt metabolites (if any) that accumulate in the *acox-1* mutant metabolome. Hence, the predicted role of ACOX-1 in ascaroside biogenesis remained putative. Similarly, the exact role of DHS-28[7,17] in ascaroside biosynthesis remained unclear, because the effect of *dhs-28* mutations on ascaroside production was only partially characterized.

The recent discovery of new structural variants with important divergent functions, e.g., icas#3 and icas#9,[6,15] suggested that knowledge of ascaroside structures and functions in *C. elegans* remains incomplete. Results from biological studies further indicate that even small structural differences between ascarosides can be associated with significant functional differences; for example, ascr#3 is a potent male attractant, whereas the structurally similar ascr#7 is nearly inactive in this male attraction assay.[5,18] Correspondingly, several previous studies indicate that ascaroside biosynthesis is tightly controlled by environmental factors such as temperature, nutrient availability, and population density.[7,16,19] Therefore, detailed investigation of ascaroside structures and their biosynthetic pathways is essential for many aspects of the biology of this model organism.

To facilitate sensitive detection and quantitation of the known ascarosides in the metabolomes of different *C. elegans* strains and mutants and aid with the discovery of new ascaroside derivatives, Dr. Stephan H. von Reuss, Schroeder group, developed a robust HPLC-MS/MS-based metabolomics tool.[20] Because of the vast complexity of the

C. elegans metabolome, HPLC-MS analysis of metabolite extracts resulted in extremely crowded chromatograms that are difficult to interpret (**Figure 2.2a**). Investigation of ESI MS/MS fragmentation of a series of synthetic ascarosides showed that with negative-ion electrospray ionization (ESI−), ascarosides give rise to an intense and highly diagnostic product ion at *m/z* 73.02939 [$C_3H_5O_2$] which originates from the ascarylose unit (**Figure 2.2b**).

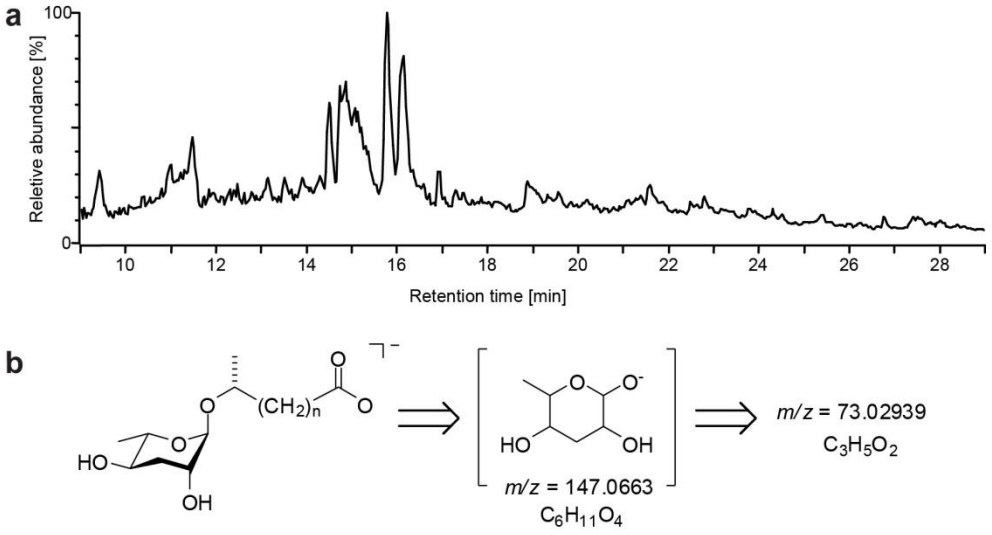


Figure 2.2: (**a**) HPLC-MS total ion current chromatogram of wild-type *C. elegans* excretome (ESI−). (**b**) MS/MS fragmentation of ascarosides.

This detection method proved suitable for all known ascarosides, except for ascr#2 and ascr#4 which do not ionize well under ESI− conditions. Hence a screen for precursor ions of *m/z* 73 was used to detect known as well as yet unidentified ascarosides in the *C. elegans* wild-type and mutant metabolomes.[20]

In this dissertation chapter, the author has utilized Dr. von Reuss's new HPLC-MS/MS- as well as 2D NMR-based metabolomics as tools for ascaroside profiling in *C. elegans*. Application of these methods to wild-type *C. elegans* revealed that the previously described ascarosides are part of a much larger, structurally diverse library of compounds derived from modular combination of building blocks from three different metabolic pathways. Subsequently the author used HPLC-MS/MS and 2D NMR on wild-type, *acox-1,* and *daf-22* metabolomes to interrogate ascaroside biosynthesis and homeostasis.

2.2. Identification of novel ascarosides using HPLC-MS/MS-based metabolomics: For the initial HPLC-MS/MS-based metabolomics screen that Dr. von Reuss developed, liquid culture (**Appendix Section A.1.3**) metabolite extracts from wild-type worms were used, which contain accumulated excreted metabolites from large numbers of worms (the worm "exo-metabolome"). The resulting HPLC-MS/MS chromatograms from wild-type *C. elegans* showed a large number of well-resolved peaks, most of which were found to represent ascarosides, including several families of previously undetected compounds. Dr. von Reuss first confirmed the identities of the known ascarosides using synthetic standards. In addition, the author in collaboration with Dr. von Reuss found that the known saturated ascarosides ascr#1, ascr#9, and ascr#10 are accompanied by substantial quantities of homologues with 6−16-carbon side chains (**Figures 2.3a**, **2.4a, 2.5b**, and **Appendix Table A.1**). Identification of this homologous series was based on high-resolution MS/MS, HPLC retention times, and synthesis of representative members (see **Appendix Section A.1.7** and **Appendix Figure A.1**). The HPLC-MS/MS

screen further revealed that ascarosides with side chains of 5−11 carbons are accompanied by smaller quantities of slightly less polar isomers. These ascaroside isomers are produced more abundantly by *acox-1* mutant worms (**Figures 2.3b**, **2.4b** and **2.5b**, *vide infra*).

Several additional peaks in the wild-type MS/MS chromatograms could not be assigned to any of the known ascaroside classes. For two of these compounds, Dr. von Reuss noticed that they produced high resolution MS/MS product ions at *m/z* 301.1651 [$C_{15}H_{25}O_6$] suggestive of ascr#3 derivatives. The putative ascr#3 derivatives were

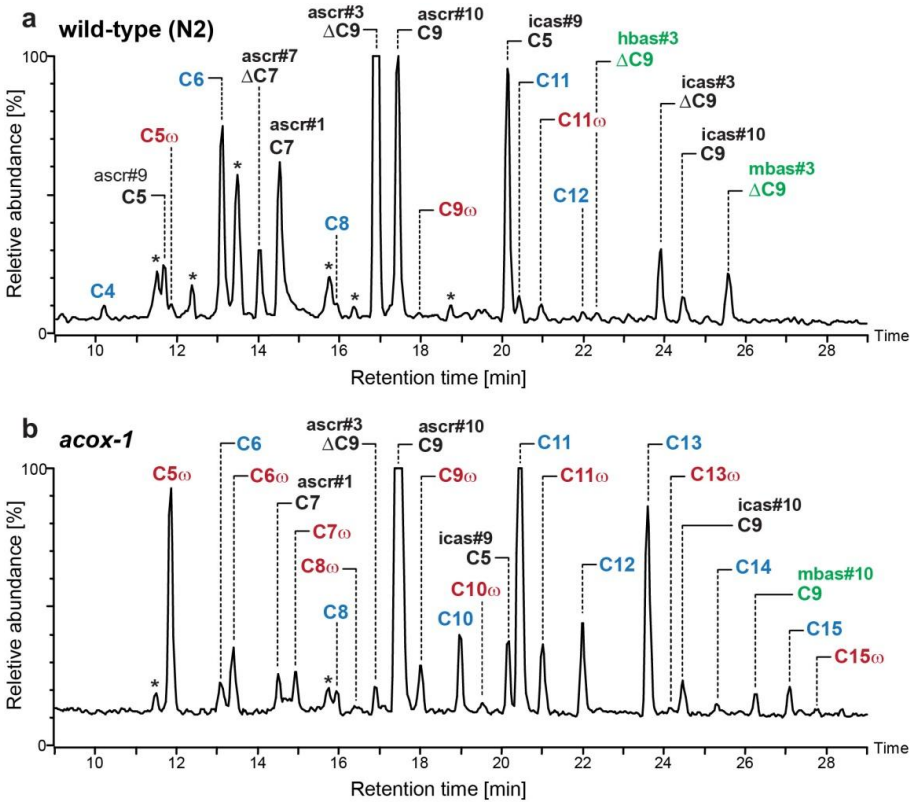


Figure 2.3: HPLC-MS/MS screen (precursors of *m/z* 73) of (**a**) wild-type and (**b**) *acox-1* exo-metabolome reveals known ascarosides (black), new homologues (blue), new ω-oxygenated isomers (red), and new 4'-acylated derivatives (green) (* indicates signals from non-ascarosides). The highly polar ascr#5 elutes at 6.5 min, outside of the retention time range shown.

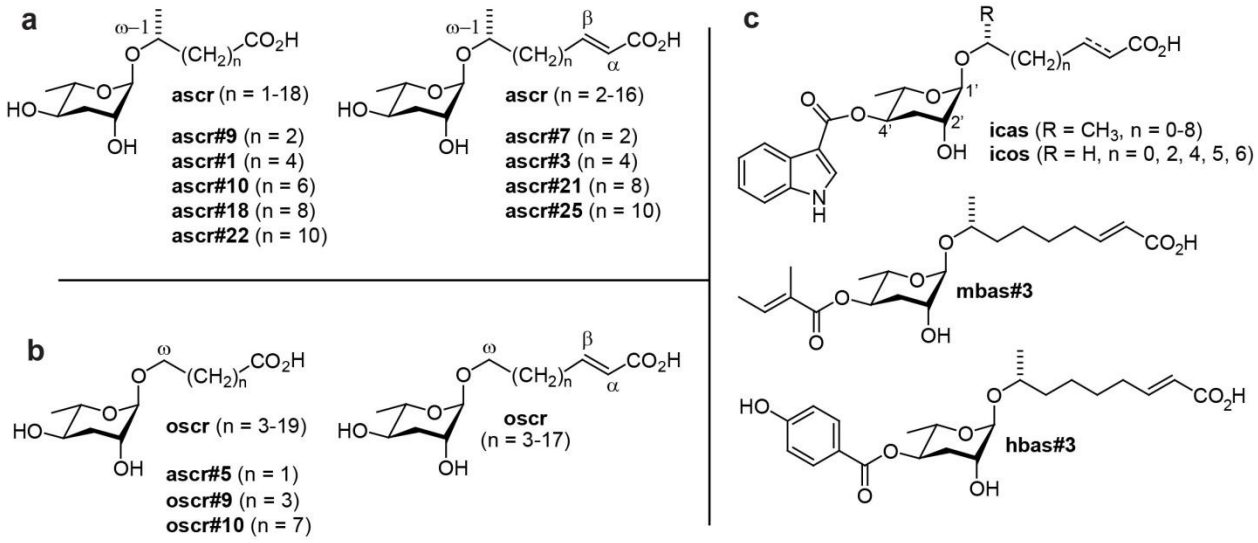


Figure 2.4: Representative ascaroside classes identified in wild-type and *acox-1* mutant worms via HPLC-MS/MS: (**a**) (ω−1)-oxygenated ascarosides, (**b**) ω-oxygenated ascarosides, and (**c**) examples for 4'-acylated derivatives. The stereochemistry of compounds that were not synthesized was proposed as shown on the basis of analogy and HPLC-MS retention times (**Appendix Figure A.1**). See also **Appendix Figure A.2.**

isolated by preparative HPLC and analyzed using 2D NMR spectroscopy (dqfCOSY, see **Chapter 1**, **Section 1.2** for details on dqfCOSY and **Appendix Figures A.3** and **A.4** for dqfCOSY spectra of the two novel compounds) by the author of this dissertation, which confirmed that these compounds are ascr#3-based and further indicated the presence of a 4-hydroxybenzoyl and a (*E*)-2-methyl-2-butenoyl (tigloyl) moieties attached to the 4-position of the ascarylose (**Figure 2.4c**). These structural assignments were corroborated via total synthesis of authentic samples by Dr. von Reuss (reported in Ref. 20). In analogy to the recently reported indole-3-carboxy derivative of ascr#3 ("icas#3"),[6] the 4-hydroxybenzoyl and (*E*)-2-methyl-2-butenoyl derivatives of ascr#3 were named with the small molecule identifiers or SMIDs (www.smid-db.org) "hbas#3" and "mbas#3", respectively. hbas#3 and mbas#3 are the first ascarosides to incorporate

4-hydroxybenzoyl and (*E*)-2-methyl-2-butenoyl moieties, which in analogy to the indole-3-carboxy moiety in icas#3[6] could be derived from amino acid precursors.

Next, synthetic samples of hbas#3 were submitted to Prof. Paul W. Sternberg's research group at Caltech, Pasadena, CA for testing this compound's effects on worm behavior. Results from Dr. Jagan Srinivasan, Sternberg group showed that the compound strongly attracts *C. elegans* at concentrations as low as 10 fM,[20] which exceeds the potency of any previously known *C. elegans* small-molecule signal.[3,6] Low femtomolar activity is unusual for small-molecule signals in animals but matched by some classes of peptide hormones.[21,22]

2.3. Role of ACOX-1, a conserved component of peroxisomal β-oxidation in ascaroside biosynthesis: Next, the author in collaboration with Dr. von Reuss used comparative HPLC-MS/MS to investigate ascaroside biogenesis. Previous studies suggested that the side chains of the ascarosides are derived from peroxisomal β-oxidation of longer-chained precursors and that the acyl-CoA oxidase ACOX-1 participates in the first step of ascaroside sidechain β-oxidation, introducing α,β-unsaturation (**Figure 2.5a**).[16] In vertebrates as well as in Drosophila, the next two steps in peroxisomal β-oxidation, hydration of the double bond and subsequent dehydrogenation to the β-ketoacyl-CoA ester, are catalyzed by one protein, e.g., MFE-2. These two enzymatic functions appear to be separated in *C. elegans* such that the hydratase and dehydrogenase are distinct proteins.[20,23]

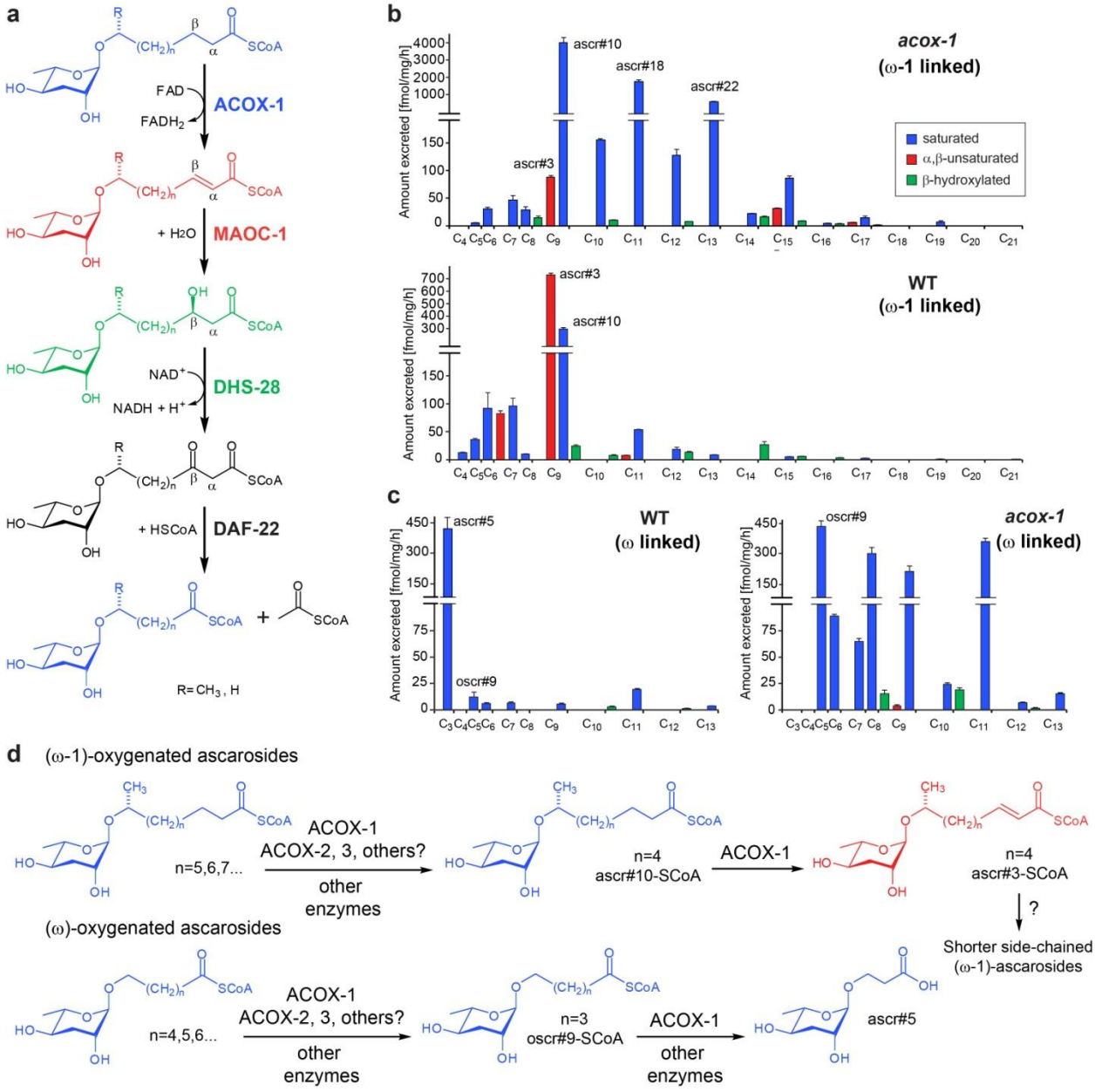


Figure 2.5: (**a**) Proposed roles of peroxisomal β-oxidation enzymes ACOX-1, MAOC-1, DHS-28, and DAF-22 in ascaroside biosynthesis. (**b**) (ω−1)-Oxygenated ascarosides in wild-type and *acox-1* mutant with saturated (blue), α,β-unsaturated (red), and β-hydroxylated (green) side chains. (**c**) ω-Oxygenated ascarosides in wild-type and *acox-1* mutants. (**d**) ACOX-enzymes' involvement in ascaroside biosynthesis: ACOX-1 works with other unidentified ACOX-2, 3, etc. for chain shortening of C_{10} and longer side chained (ω-1)-oxygentaed ascarosides and C_6 or longer side chained (ω)-oxygentaed ascarosides. Further, ACOX-1 is essential for ascr#10-SCoA to ascr#3-SCoA (ω-1, C_9-side chain) as well as oscr#9-SCoA to ascr#5 (ω, C_5-side chain) conversions. Both ascr#3 and ascr#5 are observed abundantly in wild-type samples, but starkly downregulated or absent in *acox-1* mutants (**Figure 2.5b,c**).

Previous work showed that *C. elegans* DHS-28, a protein with homology to the (*R*)-selective β-hydroxyacyl-CoA dehydrogenase domain of human MFE-2, likely participates in converting β-hydroxyacyl-CoA-derivatives into the corresponding β-ketoacyl-CoA intermediates, which are subsequently cleaved by the β-ketoacyl-CoA thiolase DAF-22.[7,17] Using the newly developed MS/MS-based ascaroside screen, Dr. von Reuss re-investigated the ascaroside profiles of *dhs-28(hj8)*, and *daf-22(ok693)* mutant worms to establish their roles as a dehydrogenase and a thiolase in ascaroside biosynthesis (**Figure 2.5a**).[20,24] Additionally, Dr. von Reuss analyzed the exo-metabolomes of several other peroxisomal mutants, including *maoc-1(hj13)*, a peroxisomal 2-enoyl-CoA hydratase,[25] which he showed participates in the hydration step of ascaroside β-oxidation (**Figure 2.5a**).[20]

The author of this dissertation focused on ascaroside-profiling of *acox-1(ok2257)* mutant in *C. elegans* to corroborate its role in ascaroside side-chain β-oxidation. ACOX-1 has previously been proposed to act as the acyl-CoA oxidase in ascaroside biosynthesis.[16] However, the metabolome of the mutant *acox-1* was not properly investigated to validate the putative role of the enzyme. HPLC-MS/MS analysis of the exo-metabolome of *acox-1(ok2257)* mutant worms revealed that the abundance of the α,β-unsaturated ascr#3, the dominating component of wild-type exo-metabolome, was greatly reduced (**Figures 2.3** and **2.5b**). This decrease in ascr#3 and other α,β-unsaturated ascarosides does not appear to be the result of overall down-regulation in ascaroside production, but instead is accompanied by accumulation of a series of saturated ascaroside, e.g. ascr#10, the dihydro-derivative and direct precursor of ascr#3 is 13.6 times more abundant in *acox-1(ok2257)* than in the wild-type exo-

metabolome. The corresponding homologues with 11- and 13-carbon side chains, ascr#18 and ascr#22, are 29 times and 66 times more abundant in *acox-1* than in the wild-type exo-metabolome, respectively (**Figure 2.5b**). The buildup of longer chained saturated ascarosides in the *acox-1(ok2257)* exo-metabolome confirms the importance of ACOX-1 in α,β-dehydrogenation of the ascaroside side chain (**Figure 2.5a**). Because ascaroside biosynthesis is not abolished in *acox-1(ok2257)* worms, it seems likely that other, yet-unidentified ACOX-enzymes contribute to peroxisomal β-oxidation of long chain ascaroside precursors. However, HPLC-MS/MS analysis of exo-metabolome of several other peroxisomal *acox* mutants (see **Appendix Section A.1.2** and **Appendix Figure A.5**) revealed largely wild-type-like ascaroside profiles.

2.4. Shunt metabolites from peroxisomal β-oxidation mutants provide deeper insights into ascaroside biosynthesis: The author further used HPLC-MS/MS analysis and 2D NMR-based metabolomics for identifying additional ascaroside-based shunt metabolites upregulated in *C. elegans acox-1* and *daf-22* mutants. HPLC-MS/MS analysis of *acox-1(ok2257)* further revealed the complete absence of ascr#5, one of the major dauer-inducing ascarosides produced abundantly in wild-type (**Figure 2.5c**). ascr#5 differs from all other previously identified ascarosides in that its side chain is attached to the ascarylose sugar via the terminal carbon ("ω linkage") and not the penultimate carbon ("ω−1 linkage") (**Figure 2.4b**). Instead of ascr#5, the author detected large quantities of a new homologous series of saturated ascarosides in *acox-1(ok2257)*, smaller amounts of which were also present in the wild-type exo-metabolome (**Figures 2.4b** and **2.5c**). The most abundant component of this series of

isomers was isolated via preparative HPLC and identified by dqfCOSY as an ω-linked ascaroside with a C_5 side chain (**Figure 2.4b** and **Appendix Figure A.6**). This suggested that the additional series of compounds observed in *acox-1* represents ω-linked saturated ascarosides (**Table A.2**), which were confirmed by synthesis of C_5 and C_9 ω-linked ascarosides by Dr. von Reuss (reported in Ref. 20). In order to differentiate the ω-linked ascarosides from their previously described (ω−1)-linked isomers, the newly found ω-linked compounds were named with the SMID "oscr", e.g., ω-linked isomers of ascr#9 and ascr#10 were named oscr#9 and oscr#10 (**Figure 2.4b**).

Thus production of ω-linked ascr#5 is abolished in *acox-1(ok2257)* worms, whereas production of longer chain homologues with 5−13-carbon side chains, e.g., oscr#9, is starkly up-regulated (**Figure 2.5c**). These results indicate that β-oxidation in *acox-1(ok2257)* worms is strongly dependent on whether the side chain is (ω−1)- or ω-functionalized (**Figure 2.5d**). Chain shortening of (ω−1)-oxygenated substrates appears to stall at a chain length of 9 carbons as in ascr#10, whereas ω-oxygenated substrates are processed for two additional rounds of β-oxidation to afford large quantities of ω-oxygenated oscr#9 featuring a 5-carbon side chain. This suggests that side-chain oxygenation precedes peroxisomal β-oxidation. In contrast, the time point of ascarylose attachment seems less certain. The absence of any (ω−1)- or ω-hydroxylated fatty acids in the investigated *C. elegans* mutant metabolome samples suggests a biosynthetic model in which very long-chain ascarosides (VLCAs) serve as substrates for peroxisomal β-oxidation; however, the possibility that β-oxidation occurs prior to ascarylose attachment cannot be excluded.

In contrast to wild-type and *acox-1(ok2257)* worms, Dr. von Reuss's data showed that short-chain ($<C_9$) ascarosides were not detected in *maoc-1(hj13)* and *dhs-28(hj8)* worms, which instead accumulate several series of (ω−1)- and ω-oxygenated medium- and long-chain ascarosides ($\geq C_9$).[20] The ascaroside profile of the *maoc-1(hj13)* exo-metabolome[20] was dominated by α,β-unsaturated ascarosides such as ascr#21 (C_{13}) and ascr#25 (C_{15}), supporting the hypothesis that MAOC-1 functions as an enoyl-CoA hydratase in the ascaroside biosynthetic pathway (**Figure 2.5a)**. Dr. von Reuss further showed that the *dhs-28(hj8)* ascaroside profile[20] is dominated by compounds with side chains ranging from C_9–C_{21}. The major components included (ω−1)- and ω-oxygenated β-hydroxyascarosides (SMIDs associated with for (ω−1)- and ω-oxygenated β-hydroxylated ascarosides are bhas and bhos) with odd-numbered side chains from C_{13}–C_{21}, consistent with the proposed biosynthetic role of DHS-28 as a β-hydroxyacyl-CoA dehydrogenase.[7,17] The ascaroside-based shunt metabolites from *maoc-1* and *dhs-28* mutants, smaller amounts of some of which were also detected in wild-type and *acox-1* mutants, originally identified by Dr. von Reuss are presented in **Appendix Figure A.2** and their identification reported in Ref. 20**.**

Preliminary analysis of *daf-22(ok693)* exo-metabolome[20] by Dr. von Reuss revealed the absence of all ascarosides with side chains shorter than 12 carbons, as reported earlier.[5,7,17] Additionally, *daf-22(ok693)* produced small amounts of homologous series of (ω−1)- and ω-oxygenated long-chain ascarosides featuring saturated (ascr and oscr) and β-hydroxylated side chains (bhas and bhos). To investigate the major ascaroside-shunt metabolites in *daf-22(ok693),* the author used

2D NMR-based comparative metabolomics (mvaDANS, **Chapter 1**, **Section 1.2,** developed in collaboration with Yevgeniy Izrayelit, Schroeder research group and Dr. Steven L. Robinette, Imperial College London, London, United Kingdom).[24] This analysis (reported in Ref. 24) revealed that *daf-22(ok693)* exo-metabolome contains large amounts of long-chained ascarosides belonging to the series: (ω-1)-oxygenated 2-ketoascarosides, (ω-1)-oxygenated β-ketoascaroside ethanolamides, (ω-1)-oxygenated α-methyl α,β-unsaturated ascaroside ethanolamides, and (ω-1)-oxygenated α-methyl ascaroside ethanolamides (**Figure 2.6**). The abundance of β-ketoascarosides in *daf-22* confirmed its role as a β-ketoacyl-CoA thiolase in ascaroside biosynthesis. The upregulation of ascaroside ethanolamides in *daf-22* mutants were concurrent with an observed downregulation of N-acyl-ethanolamine (NAE) production,[24] including the most active endocannabinoid in *C. elegans*,[26] eicosapentaenoyl ethanolamide (EPEA), and the ligand of mammalian cannabinoid receptors, arachidonoyl ethanolamide (AEA, or anandamide), suggesting a shift of ethanolamine utilization from the NAEs to VLCAs.[24]

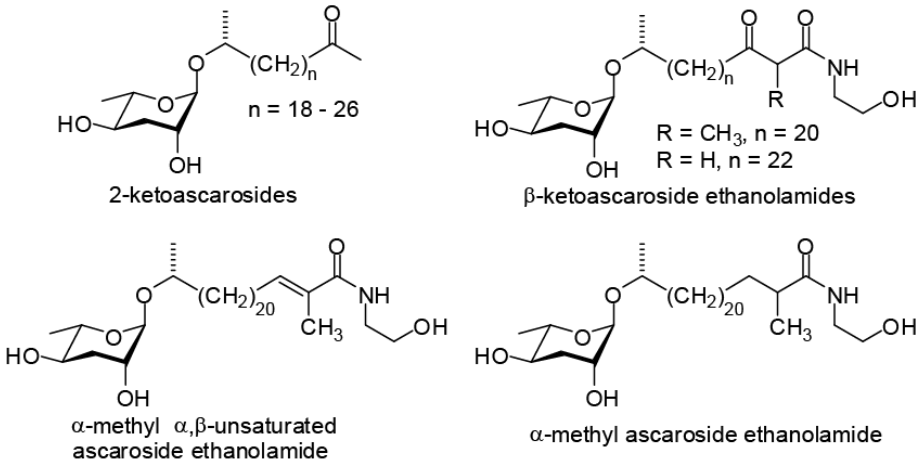


Figure 2.6: *daf-22* mutant-specific long-chained ascarosides identified via 2D NMR-based metabolomics and subsequent HPLC−MS analysis.

2.5. Identification of new indole ascarosides: Indole-3-carbonylated ascarosides are much less abundant than the corresponding unfunctionalized ascarosides and have recently been shown to function as highly potent aggregation signals.[6] The author in collaboration with Dr. von Reuss used HPLC-MS/MS to screen wild-type and peroxisomal β-oxidation mutant metabolomes and revealed several new types of indole ascarosides (**Figure 2.4c**, **Appendix Figure A.2**, and **Appendix Tables A.3**, **A.4**). Dr. von Reuss showed using synthetic samples of icas#3, icas#9, and icas#1,[6] that indole ascarosides exhibit a characteristic fragmentation pattern that includes neutral loss of 143 amu [C_9H_5NO] due to cleavage of the indole-3-carbonyl unit, as well as the ascaroside-diagnostic product ion at *m/z* 73.[20] HPLC-MS/MS screening for components with this fragmentation pattern by the author revealed that known (ω−1)-oxygenated isomers icas#1, icas#9, and icas#10 in *acox-1(ok2257)* are accompanied by a series of ω-oxygenated indole ascarosides (SMIDs: icos#1, icos#9, and icos#10), which was confirmed by chemical synthesis of icos#10 as a representative example by Dr. von Reuss (reported in Ref. 20).

2.6. Indole ascaroside biogenesis: Using application experiments with deuterium-labeled tryptophan and axenic *in vitro* cultures, it was previously shown that the indole-3-carbonyl moiety of indole ascarosides originates from L-tryptophan.[6] A similar L-tyrosine or L-phenylalanine origin seems likely for the 4-hydroxybenzoyl moiety of hbas#3, whereas the tigloyl group of mbas#3 could be derived from L-isoleucine.[27] However, it remained unclear at what stage in ascaroside biosynthesis the indole-3-

carbonyl moiety is attached. Comparison of ascaroside and indole ascaroside profiles revealed that indole ascaroside biosynthesis is tightly controlled. For example, the author found that in *acox-1* mutants the ω-ascaroside oscr#9 is over 100 times more abundant than the (ω−1)-ascaroside ascr#9, whereas (ω−1)-indole ascaroside icas#9 is much more prominent than ω-indole ascaroside icos#9 (**Figure 2.7a**). Similarly, icas#3 and icas#9, the dominating indole ascarosides in the wild-type exo-metabolome are produced in roughly equal amounts, whereas ascr#3 is about 20 times more abundant than ascr#9 (**Figure A.7**). This strong dependence of indole ascaroside biosynthesis on side-chain length and (ω−1)- vs ω-oxygenation suggested that these compounds originate from specific attachment of an indole-3-carbonyl unit to the corresponding non-indole ascarosides.

To test whether non-indole ascarosides serve as precursors for indole ascarosides, the author incubated *daf-22(m130)* worms (which are devoid of all short-chain indole and non-indole ascarosides) with a 1:1 mixture of ascr#10 and oscr#10 for 5 days. Subsequent analysis by HPLC-MS showed partial conversion into icas#10 and icos#10 (**Figure 2.7b**), indicating that non-indole ascarosides serve as specific precursors to their corresponding indole derivatives. Moreover, conversion of (ω−1)-ascaroside ascr#10 was preferred over conversion of ω- ascaroside oscr#10, reflecting the ratios of these compounds in wild-type and *acox-1* mutants. Similarly, the author found that *daf-22(m130)* worms convert added ascr#3 into icas#3 (data not shown). Further conversion of indole or non-indole ascarosides into shorter chain derivatives (e.g., ascr#1 or icas#1) was not observed. These results indicate that attachment of an

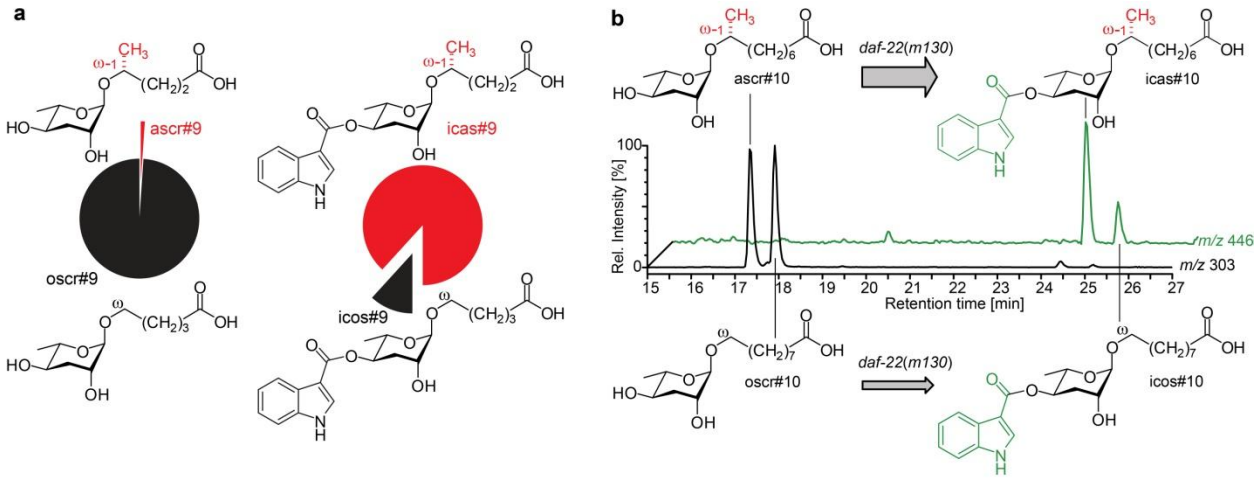


Figure 2.7: Indole ascaroside biosynthesis. (**a**) Relative abundance of (ω–1) and ω-oxygenated C_5-ascarosides ascr#9 and oscr#9 and their corresponding indole ascarosides icas#9 and icos#9 in *acox-1* indicates that indole-3-carbonyl attachment is highly specific. (**b**) HPLC-MS ion traces of *daf-22*(*m130*) excretome following incubation with a 1:1 mixture of ascr#10 and oscr#10, showing a preference for indole-3-carbonyl attachment to the (ω–1)-oxygenated ascr#10.

L-tryptophan-derived indole-3-carbonyl unit represents the final step in indole ascaroside biosynthesis.

2.7. Ascaroside excretion is selective: Despite detailed investigations of ascaroside function, little is known about how ascarosides are released and transported from their site of biosynthesis. In collaboration with Dr. von Reuss, the author compared the ascaroside profiles of the wild-type exo-metabolomes (liquid culture supernatant extracts) and endo-metabolomes (worm pellet extracts) to identify possible non-excreted ascaroside derivatives and to determine quantitative differences. The results indicated that ascaroside profiles of worm pellet extracts differed significantly from those excreted into the media, indicating that ascarosides are differentially released by *C. elegans* (**Figure 2.8**). In the worm pellets, the most abundant ascarosides, for example

ascr#3 in wild-type and ascr#10 in *acox-1* worms, are accompanied by significantly more-polar derivatives, which are absent from the media extracts (**Figure A.8**). MS/MS analysis suggested that these components represent ascaroside *O*-glycoside esters. Isolation of the putative ascr#10 glycoside from *acox-1(ok2257)* worm pellet extracts and subsequent dqfCOSY analysis (**Appendix Figure A.9**) by the author indicated esterification of ascr#10 with the anomeric hydroxyl group of β-glucose (SMID for 1-β-D-glucosyl ascr#10 is glas#10, **Figure 2.8**), which was subsequently confirmed via total synthesis by Dr. von Reuss (reported in Ref. 20). The fact that large quantities of highly polar glas#10 and other ascaroside glucosides (**Table A.5**) are retained in the worm bodies but not excreted suggests that they represent transport or storage forms of the ultimately excreted signaling molecules.

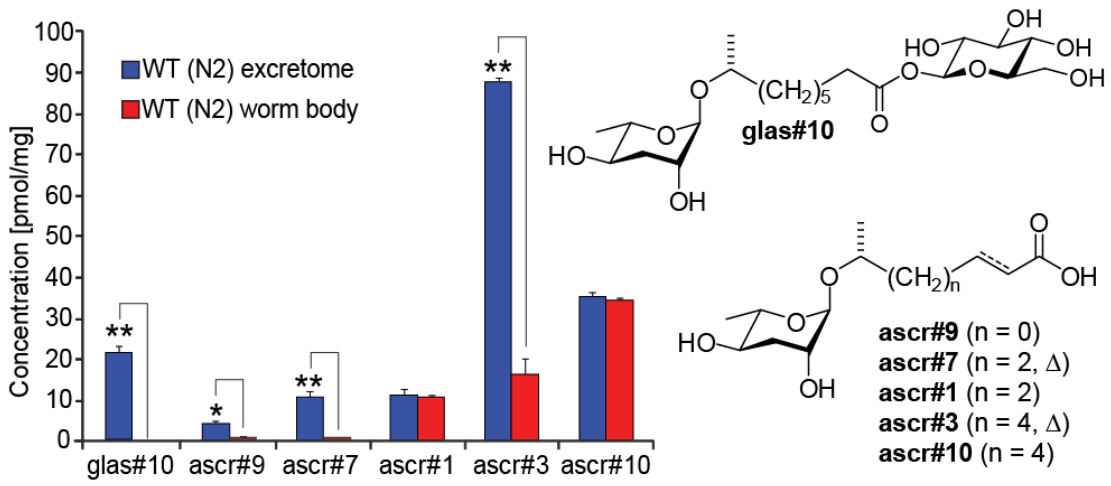


Figure 2.8: Analysis of worm body ascaroside profiles reveals ascaroside glucosides (e.g., glas#10) and indicates preferential excretion of unsaturated ascarosides (Δ indicates components with (*E*)-configured α,β-unsaturated sidechains). Significance, *=p<0.01, **=p<0.001, ***=p<0.001, **Appendix Section A.1.8**.

2.8. Discussion: Ascarosides play important roles for several different aspects of *C. elegans* biology. This functional diversity is paralleled by corresponding structural diversity and a complex biosynthetic pathway. The data presented in this dissertation chapter revealed several novel ascaroside classes, which show unexpected features including ω-oxygenation of the fatty-acid-derived side chains, 4-hydroxybenzoylation or (*E*)-2-methyl-2-butenoylation of the ascarylose unit, and glucosyl esters. Most ascarosides occur as members of homologous series of compounds with (ω−1)- or ω-linked saturated, α,β-unsaturated, or β-hydroxylated side chains ranging from 3 to 21 (occasionally more) carbons. Importantly, only a few members of each series are produced abundantly in wild-type, and incorporation of specific structural features such as modification in position 4 of the ascarylose (e.g., as in indole ascarosides) appears to be tightly controlled. Given their assembly from carbohydrate, lipid, and amino-acid-derived building blocks, the ascarosides appear as a modular library of small-molecule signals that integrate inputs from three basic metabolic pathways (**Figure 2.9a**). The ascarosides then transduce input from these pathways via their diverse signaling functions in *C. elegans*'s behavior and development, including dauer formation, mate attraction, hermaphrodite repulsion, and aggregation.[6,18] Their specific biosyntheses suggest that many of the newly identified ascarosides also contribute to known or as yet undetermined functions in *C. elegans*. More comprehensive biological evaluation will require consideration of synergistic activities (i.e., testing of compound mixtures), which, given the very large number of compounds, may necessitate high-throughput approaches, for example based on microfluidic devices.[28]

The results presented further allowed the author to propose a working model for ascaroside biogenesis (**Figure 2.9b**). The finding that ascarosides are members of several homologous series with side chains up to 21 (and more) carbons suggests their origin from peroxisomal β-oxidation of very long-chain precursors. Previous studies reported the presence of VLCAs with C_{29} and C_{31} side chains in wild-type and *daf-22* mutants, which could represent precursors or intermediates in ascaroside biosynthesis.[5,29] Alternatively, VLCFAs could undergo peroxisomal β-oxidation prior to (ω−1)- or ω-functionalization and subsequent attachment of the ascarylose. The observation that the *acox-1* mutation affects (ω−1)- and ω-oxygenated ascarosides differently suggests that (ω−1)- and ω-functionalization of VLCFA precursors occurs upstream of their breakdown by peroxisomal β-oxidation. Given the large range in side-chain lengths, it seems likely that resulting (ω−1)- and ω-hydroxy VLCFAs are linked to ascarylose prior to entering the β-oxidation pathway, though the presence of a promiscuous ascarosyl transferase cannot be excluded (**Figure 2.9b**).

Chain shortening of VLCA then progresses via repetitive cycles of peroxisomal β-oxidation. Results from HPLC-MS/MS- and 2D NMR-based comparative metabolomics allowed the author to propose precise roles for enzymes participating in the first and the fourth step of the β-oxidation cycle: acyl-CoA oxidase ACOX-1 and β-ketoacyl-CoA thiolase DAF-22.[20,24] HPLC-MS/MS screen by Dr. von Reuss further elucidated the roles of enoyl-CoA hydratase MAOC-1 and β-hydroxyacyl-CoA dehydrogenase DHS-28, for steps two and three for β-oxidation cycle in ascaroside biogenesis.[20] These data show that mutations in *acox-1*, *maoc-1*, *dhs-28,* and *daf-22* result in specific changes of the corresponding ascaroside profiles, in agreement with their proposed functions.

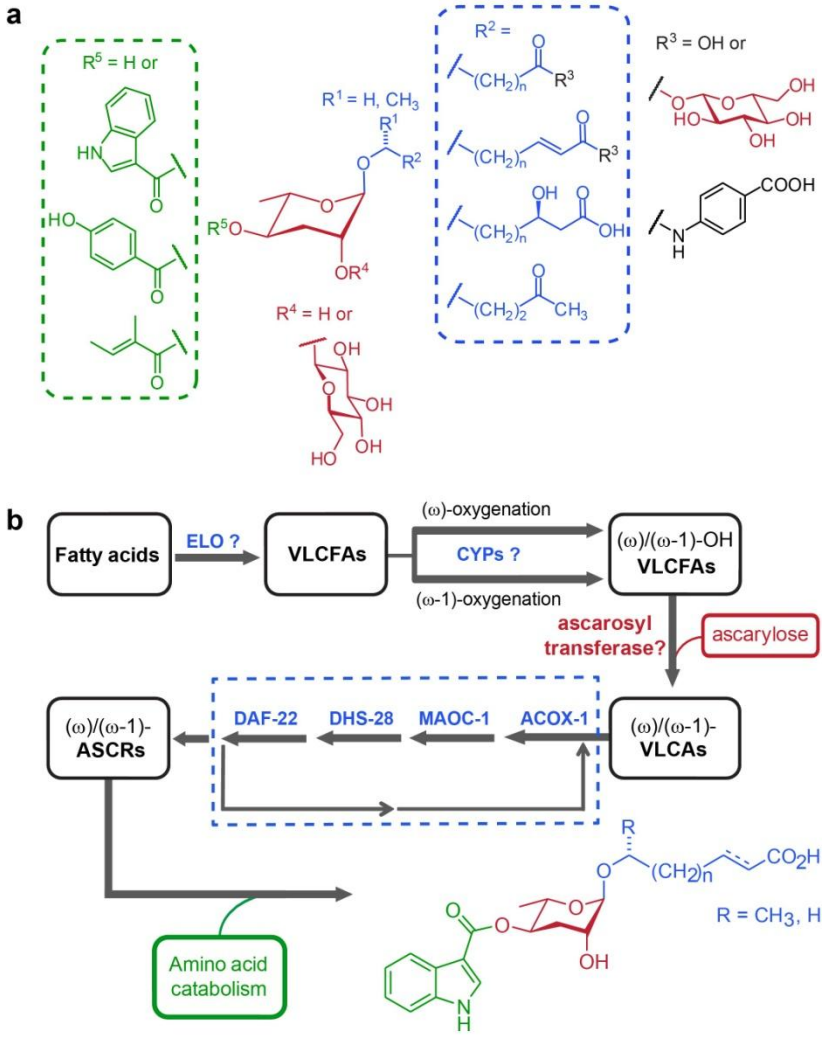


Figure 2.9: (**a**) Modular assembly of ascarosides from amino acid (green), fatty acid (blue), and carbohydrate (red) building blocks. (**b**) Model for ascaroside biogenesis. Chain elongation of fatty acids by putative elongase homologues *elo-1−9*[30] is followed by (ω−1)- or ω-oxygenation of VLCFAs and ascarylose attachment. The resulting VLCAs enter peroxisomal β-oxidation via ACOX-1, MAOC-1, DHS-28, and DAF-22, producing short-chain ascarosides, which are linked to amino-acid-derived moieties and other building blocks.

The acyl-CoA oxidase ACOX-1 has been the subject of a previous study which suggested that mutations in *acox-1* primarily affect the biosynthesis of ascr#2 and ascr#3, but not of ascr#1.[16] However, the author's results indicate that *acox-1(ok2257)* mutants have a reduced ability to process C_9 (ω−1)-functionalized ascarosides, resulting

in diminished production of all shorter-chained ascarosides and buildup of C_9 and longer-chained saturated ascarosides.

In addition to HPLC-MS/MS, 2D NMR-based comparative metabolomics revealed long-chained ascarosyl ethanolamides in *daf-22(ok693)* as an unexpected class of shunt metabolites in *daf-22* mutant worms. The identification of β-keto derivatives and methyl ketones among the identified long-chain ascarosides confirms the role of *daf-22* as a thiolase in ascaroside biosynthesis. The abundance of ascaroside ethanolamides in *daf-22(ok693)* connect nematode and mammalian fatty-acid metabolism to endocannabinoid signaling and phosphatidylethanolamine utilization. In *C. elegans*, reduced NAE levels promote dietary restriction-induced lifespan extension,[26] suggesting that defects in peroxisomal β-oxidation may affect lifespan via changes in NAE levels.

The data reported in this dissertation chapter further shows that attachment of the tryptophan-derived indole-3-carbonyl unit in indole ascarosides likely represents the last step in their biosynthesis, and that this step is highly specific. As attachment of an indole-3-carbonyl group to ascarosides can dramatically alter their biological function, such tight regulation makes sense. For example, indole-3-carbonyl addition to the dauer-inducing and strongly repulsive signal ascr#3 results in the potent hermaphrodite attractant icas#3.[6] Therefore, identification of the enzymes that attach indole-3-carbonyl and other functional groups to the ascarosides will be of great interest. The biosynthesis of ascarylose in *C. elegans* has not been investigated, but detection of ascarosides in axenic *C. elegans* cultures demonstrated that *C. elegans* produce ascarylose endogenously.[6] Ascarylose biosynthesis in bacteria is well understood, and the *C.*

elegans genome includes several homologues of bacterial genes in this pathway, for example, *ascE* from *Yersinia pseudotuberculosis* (**Appendix Figure A.11**),[31] providing potential entry points for the study of ascarylose biosynthesis and its regulation in nematodes. In addition, the oxidases catalyzing (ω−1)- or ω-functionalization of VLCFA precursors remain to be identified.

Finally, the author demonstrates that ascaroside excretion is surprisingly specific. Given the high sensitivity and selectivity of HPLC-MS/MS, the author believes that ascaroside profiling using this method will aid identification of additional genes and environmental factors that participate in ascaroside biosynthesis and homeostasis.

2.9. Author's note: HPLC-MS/MS-based workflow to analyze *C. elegans* wild-type and peroxisomal β-oxidation mutants was developed by Dr. Stephan von Reuss, postdoctoral researcher in Prof. Frank C. Schroeder's group at Cornell University, Ithaca, NY (currently Project Leader, Bioorganic Chemistry Department, Max Planck Institute for Chemical Ecology, Jena, Germany). Dr. von Reuss took the lead in the analysis of the metabolomes of *C. elegans* wild-type and *maoc-1*, *dhs-28*, and *daf-22* mutants via HPLC-MS/MS in collaboration with the author. Dr. von Reuss also performed the total synthesis of representative ascarosides (mbas#3, hbas#3, oscr#9, icos#10, and glas#10) to corroborate the structural assignments of new ascarosides made by the author, detailed reaction conditions and experimental procedures are reported in Ref. 20. The 2D NMR-based comparative metabolomics study (mvaDANS) to identify specific VLCAs in *daf-22* mutants was done in collaboration with Mr.

Yevgeniy Izrayelit, graduate student, Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY and Steven L. Robinette, Imperial College London, London, UK. Biological assays to elucidate the activity of hbas#3 were done by Dr. Jagan Srinivasan, postdoctoral researcher in Prof. Paul W. Sternberg's group at Caltech, Pasadena, CA (currently Assistant Professor, Worcester Polytechnic Institute, Worchester, MA).

REFERENCES

(1) Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J. *Nature chemical biology* **2007**, *3*, 420.

(2) Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K. *Nature.* **2005**, *433*, 541.

(3) Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E. A.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *Nature* **2008**, *454*, 1115.

(4) Butcher, R. A.; Ragains, J. R.; Kim, E.; Clardy, J. *Proc Natl Acad Sci U S A.* **2008**, *105*, 14288.

(5) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A.* **2009**, *106*, 7708.

(6) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol.* **2012**, *10*, e1001237. Epub 2012 Jan 10.

(7) Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y. *Proc Natl Acad Sci U S A.* **2009**, *106*, 1875.

(8) Ren, P.; Lim, C. S.; Johnsen, R.; Albert, P. S.; Pilgrim, D.; Riddle, D. L. *Science* **1996**, *274*, 1389.

(9) Kimura, K. D.; Tissenbaum, H. A.; Liu, Y.; Ruvkun, G. *Science* **1997**, *277*, 942.

(10) Schroeder, F. C. *ACS chemical biology* **2006**, *1*, 198.

(11) Fielenbach, N.; Antebi, A. *Genes & development* **2008**, *22*, 2149.

(12) Sommer, R. J.; Ogawa, A. *Curr Biol* **2011**, *21*, R758.

(13) Golden, J. W.; Riddle, D. L. *Molecular & general genetics : MGG* **1985**, *198*, 534.

(14) White, J. Q.; Nicholas, T. J.; Gritton, J.; Truong, L.; Davidson, E. R.; Jorgensen, E. M. *Curr Biol* **2007**, *17*, 1847.

(15) Butcher, R. A.; Ragains, J. R.; Clardy, J. *Organic letters* **2009**, *11*, 3100.

(16) Joo, H. J.; Kim, K. Y.; Yim, Y. H.; Jin, Y. X.; Kim, H.; Kim, M. Y.; Paik, Y. K. *J Biol Chem* **2010**, *285*, 29319.

(17) Joo, H. J.; Yim, Y. H.; Jeong, P. Y.; Jin, Y. X.; Lee, J. E.; Kim, H.; Jeong, S. K.; Chitwood, D. J.; Paik, Y. K. *Biochem J* **2009**, *422*, 61.

(18) Edison, A. S. *Current opinion in neurobiology* **2009**, *19*, 378.

(19) Kaplan, F.; Srinivasan, J.; Mahanti, P.; Ajredini, R.; Durak, O.; Nimalendran, R.; Sternberg, P. W.; Teal, P. E.; Schroeder, F. C.; Edison, A. S.; Alborn, H. T. *PloS one* **2011**, *6*, e17804.

(20) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc.* **2012**, *134*, 1817.

(21) Rittschof, D.; Cohen, J. H. *Peptides* **2004**, *25*, 1503.

(22) Gozes, I.; Morimoto, B. H.; Tiong, J.; Fox, A.; Sutherland, K.; Dangoor, D.; Holser-Cochav, M.; Vered, K.; Newton, P.; Aisen, P. S.; Matsuoka, Y.; van Dyck, C. H.; Thal, L. *CNS drug reviews* **2005**, *11*, 353.

(23) Haataja, T. J.; Koski, M. K.; Hiltunen, J. K.; Glumoff, T. *Biochem J* **2011**, *435*, 771.

(24) Izrayelit, Y.; Robinette, S. L.; Bose, N.; von Reuss, S. H.; Schroeder, F. C. *ACS chemical biology* **2013**, *8*, 314.

(25) Zhang, S. O.; Box, A. C.; Xu, N.; Le Men, J.; Yu, J.; Guo, F.; Trimble, R.; Mak, H. Y. *Proc Natl Acad Sci U S A* **2010**, *107*, 4640.

(26) Lucanic, M.; Held, J. M.; Vantipalli, M. C.; Klang, I. M.; Graham, J. B.; Gibson, B. W.; Lithgow, G. J.; Gill, M. S. *Nature* **2011**, *473*, 226.

(27) Attygalle, A. B.; Wu, X.; Will, K. W. *Journal of chemical ecology* **2007**, *33*, 963.

(28) Chung, K.; Zhan, M.; Srinivasan, J.; Sternberg, P. W.; Gong, E.; Schroeder, F. C.; Lu, H. *Lab Chip* **2011**, *11*, 3689.

(29) Zagoriy, V.; Matyash, V.; Kurzchalia, T. *Chemistry & biodiversity* **2010**, *7*, 2016.

(30) Agbaga, M. P.; Brush, R. S.; Mandal, M. N.; Henry, K.; Elliott, M. H.; Anderson, R. E. *Proc Natl Acad Sci U S A* **2008**, *105*, 12843.

(31) Thorson, J. S.; Lo, S. F.; Ploux, O.; He, X. M.; Liu, H. W. *J Bacteriol* **1994**, *176*, 5483.

CHAPTER 3

COMPLEX ARCHITECTURES DERIVED FROM MODULAR ASSEMBLY OF PRIMARY METABOLITES REGULATE DEVELOPMENT AND PHENOTYPIC PLASTICITY IN *P. PACIFICUS*

3.1. Introduction: Small molecule signaling serves important functions at all levels of organismal organization and requires diverse biosynthetic mechanisms for encoding biological information in chemical structures. They facilitate inter- and intra-species chemical communication, are used for chemical defense and predation, function as hormones and second messengers in animals and plants, and serve as building blocks for biological macromolecules.[1] In contrast to many groups of microorganisms and plants, whose genomes encode a great variety of "secondary" small-molecule biosynthetic pathways (for example, for polyketides and non-ribosomal peptides),[2,3] most animals are not known to have dedicated biosynthetic pathways to generate structurally complex small molecules. Correspondingly, many unusual metabolites that have been isolated from basal animals (e.g., sponges and bryozoans, among others) and arthropods have turned out to be of microbial origin or acquired through their diet.[4-7]

Studies in the model organism *Caenorhabditis elegans* (see **Chapter 2**) showed that this nematode produces a family of small molecule signals, the ascarosides, e.g. ascr#1-3, and hbas#3 (**Figure 3.1a**), which control multiple aspects of *C. elegans* life history, including larval development, mating, and social behaviors.[8-13] The structures of

these *C. elegans* signaling molecules derive from combination of the dideoxysugar ascarylose with a variety of lipid- and amino acid-metabolism derived moieties, suggesting that nematodes, and perhaps other animals, harbor yet unrecognized biosynthetic capabilities.

As part of a broad 2D NMR-spectroscopic screen of nematode metabolomes, the author analyzed the exo-metabolome (the entirety of all secreted and excreted metabolites) of the necromenic roundworm *Pristionchus pacificus*. Like *C. elegans*, *P. pacificus* is a free-living nematode that has been established as a model organism for the study of developmental and evolutionary biology.[14] *P. pacificus* forms a necromenic association with beetles, which may represent a pre-adaptation to the evolution of true parasitism.[15] *P. pacificus* exhibits two types of phenotypic plasticity that are key to its survival in the wild. Like in many other nematode species, harsh environmental conditions, for example food shortage, trigger developmental arrest as dauer worms, a highly stress-resistant alternate larval stage.[16] *P. pacificus* further exhibits a unique dimorphism in mouth development, representing an example for phenotypic plasticity of morphology in an adult metazoan (**Figure 3.1b**). Adult worms can have either a narrow ("stenostomatous") or a wide and more complex ("eurystomatous") mouth opening, the latter developing in response to conditions of low food availability. The two different mouth forms are associated with different feeding habits: stenostomatous worms are considered to feed primarily on bacteria, whereas the eurystomatous form is suited for predatory behavior toward other nematodes. Previous studies showed that both dauer formation and mouth dimorphism are regulated by population density, suggesting that these two examples of phenotypic plasticity are driven by inter-organismal small-

molecule signaling (**Figure 3.1b**).[17] Interestingly, crude dauer pheromones isolated from multiple *P. pacificus* wild (natural) strains from around the world show markedly different efficacies in dauer formation when cross-tested on each other.[18] This suggests that there is a significant natural variation in small molecule profiles that comprise the dauer pheromone in *P. pacificus*.

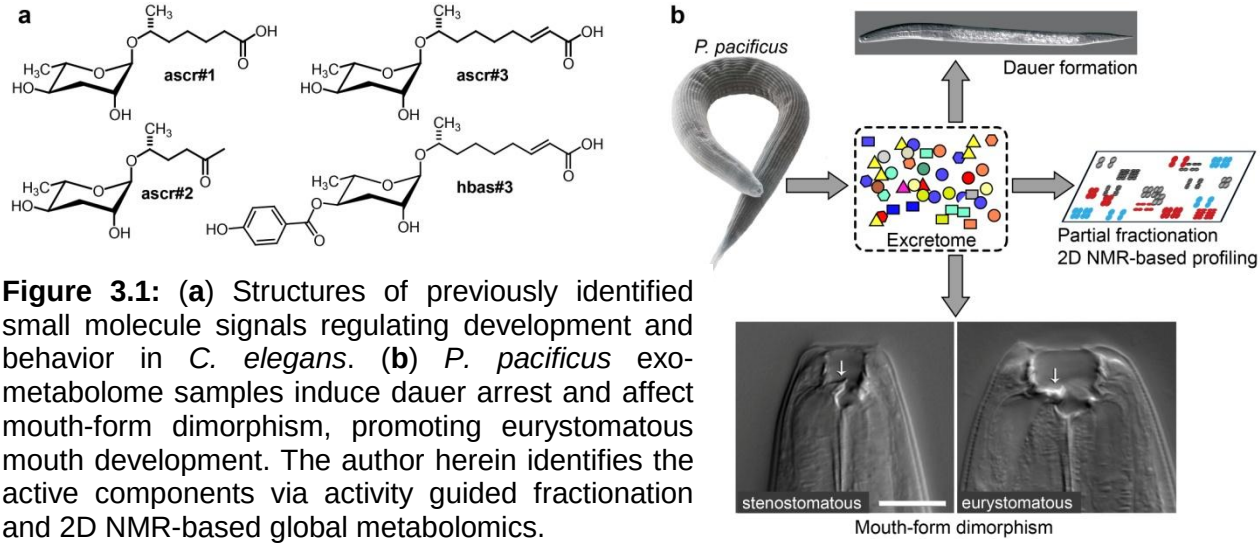


Figure 3.1: (**a**) Structures of previously identified small molecule signals regulating development and behavior in *C. elegans*. (**b**) *P. pacificus* exo-metabolome samples induce dauer arrest and affect mouth-form dimorphism, promoting eurystomatous mouth development. The author herein identifies the active components via activity guided fractionation and 2D NMR-based global metabolomics.

In this dissertation chapter, the author shows that *P. pacificus* generates structurally complex signaling molecules via selective assembly of building blocks from several primary metabolic pathways, including an unusually modified nucleoside based on xylopyranose instead of ribofuranose.[19] These modular compounds act as interorganismal signals controlling larval development, adult morphology, and lifestyle. A subset of compounds induces developmental arrest as highly persistent dauer larvae, whereas a different set of compounds promotes alternate mouth-form development, the body-plan change that enables predatory behavior.[19] These results demonstrate small molecule regulation of adult phenotypic plasticity in a metazoan, which relies on co-

option of conserved endocrine signaling pathways. Furthermore, the author's findings indicate species-specific evolution of chemical signaling in nematodes, with regard to both chemical structures and their biological functions. The *P. pacificus* small-molecule library provides striking examples for combinatorial generation of structural diversity in a metazoan and connect primary metabolism to regulation of development and adult phenotypic plasticity. In addition, comparative HPLC-MS metabolomic analysis of 6 *P. pacificus* natural isolates show both qualitative and quantitative differences in small molecule profiles and in part explain the previously observed natural variation in dauer pheromone production by *P. pacificus*.[18]

3.2. Global metabolomics using 2D NMR and HPLC-MS/MS reveals novel modular small molecules from *P. pacificus*: *P. pacificus* RS2333 (formerly known as PS312)[20] was grown in large 3 L liquid cultures (see **Appendix Section B.1.4**) by Dr. Akira Ogawa, Sommer research group, Max-Plank Institute (MPI) for Developmental Biology, Tuebingen, Germany. Dr. Ogawa subsequently performed a round of preliminary fractionation of the corresponding exo-metabolome (liquid culture supernatant) and generated three fractions I, II, and III (see **Appendix Section B.1.4**). Fraction II showed the most activity in his dauer formation assays and was used by the author for 2D NMR- and HPLC-MS/MS-based metabolomics for the presence of novel molecular architectures in *P. pacificus*.

The author first used dqfCOSY (**Chapter 1**, **Section 1.2**), a 2D NMR experiment on Fraction II to achieve a broad overview of the different structural features that

comprise this active fraction. Analysis of crosspeaks and associated fine-structures of the dqfCOSY spectrum of Fraction II revealed a striking diversity of signals suggestive of primary metabolites, for example, xylose, threonine, adenosine, short-chain fatty acids, succinate, phenylethanolamine, and several dideoxysugar derivatives (**Figure 3.2**). Specifically, the dideoxysugar derivatives fell into two chemically different subgroups, as judged by dqfCOSY crosspeak fine-structure analysis. One group was based on ascarylose, a sugar that is widely produced by nematodes,[21] whereas a second group of compounds appeared to include a related sugar, paratose, which previously had been reported only in bacteria (**Figure 3.2**).[22] Further analysis of the dqfCOSY and corroboration via the analysis of HMBC spectrum of Fraction II indicated that most of primary metabolic building blocks are part of larger assemblies based on the dideoxysugars (**Figures 3.2** and **3.3**).

In addition, the author analyzed the crude *P. pacificus* exo-metabolome extract (**Appendix Section B.1.4**) by HPLC-MS/MS[8] for the presence of previously identified compounds from *C. elegans.* A detailed description of this HPLC-MS/MS procedure is presented in **Chapter 2**, **Section 2.1**. Briefly, in this protocol, crude nematode-derived extracts are screened in negative ionization mode for precursor ions of *m/z* 73, a highly diagnostic fragment ion originating from ascarosides (or paratosides, molecules based on the dideoxysugar paratose, C-2 epimer of the ascarylose sugar, *vide infra*). The results of the HPLC-MS/MS analysis of crude *P. pacificus* exo-metabolome extract revealed the presence of ascr#1, ascr#9, and ascr#12, as previously reported from *P. pacificus* (**Figure 3.4**, and **Appendix Figure B.1**).[21] Further, the HPLC-MS/MS total ion chromatogram (TIC) showed several additional peaks (**Appendix Figure B.1**) that did

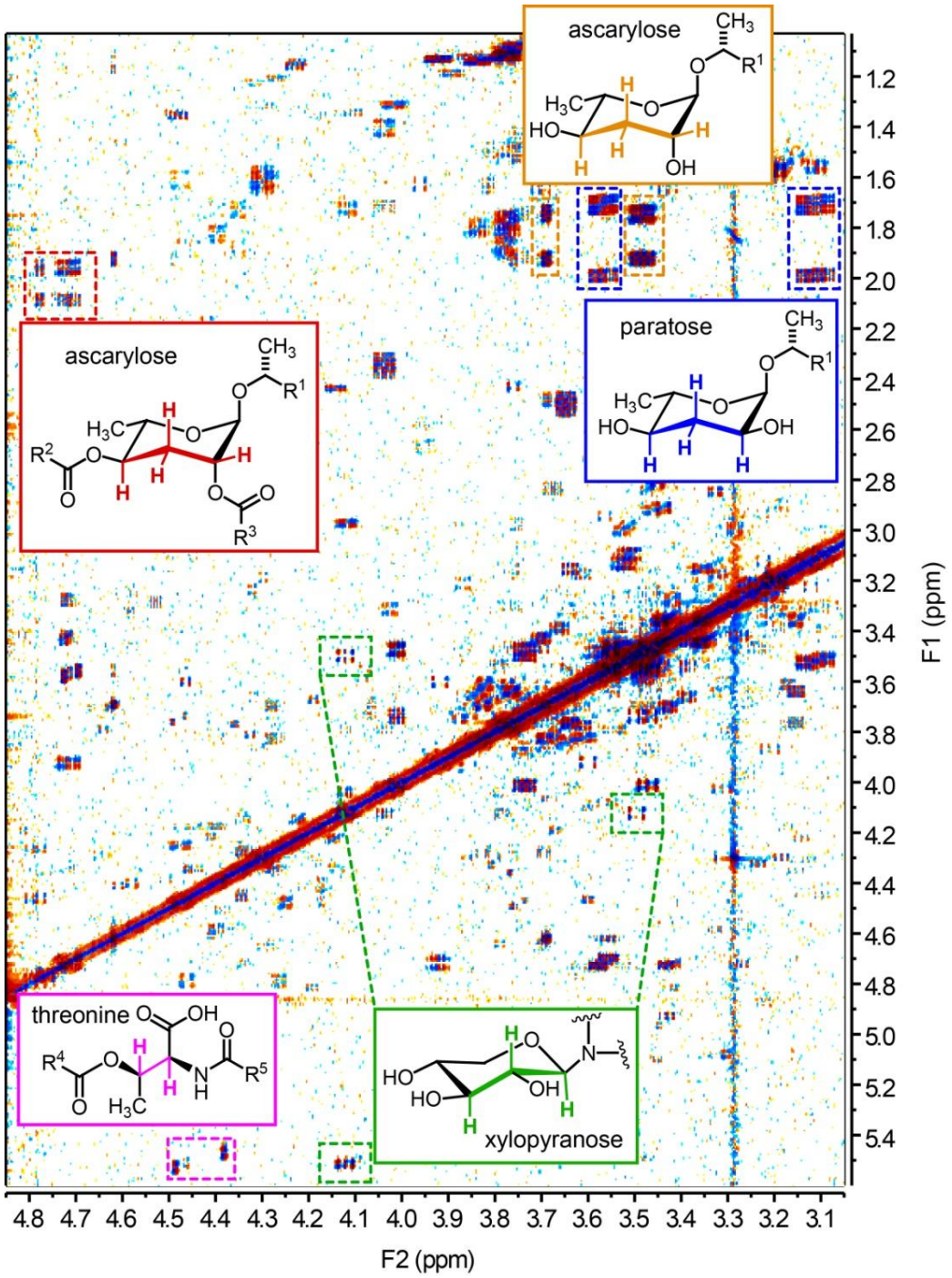


Figure 3.2: Example section of 2D NMR (dqfCOSY) spectrum of *P. pacificus* exo-metabolome Fraction II (see **Appendix Section B.1.4**), revealing a complex metabolite mixture, including known primary metabolites as well as unknown components. Detailed analysis of crosspeak fine structure and additional HMBC spectra (**Figure 3.3**) led to detection of a series of unusual chemical structures based on combinations of ascarylose, paratose, threonine, xylose, and other building blocks.

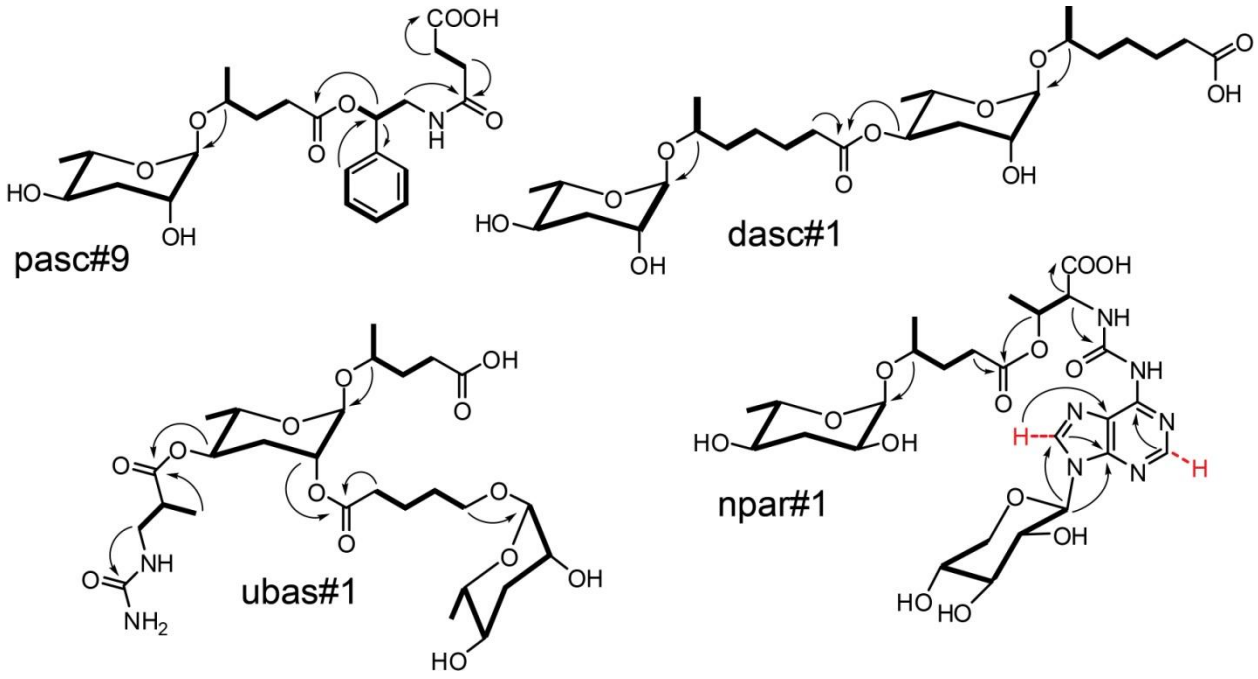


Figure 3.3: NMR spectroscopic structure elucidation of major *P. pacificus* small molecules: pasc#9, npar#1, dasc#1, and ubas#1. The **bold** lines indicate spin systems in dqfCOSY spectra. Curved arrows indicate key HMBC correlations used to assign the structures. Marked protons (**---H**) in **npar#1** are characteristic of N^6-carbamoyl adenosine and observed in ^{1}H, HSQCAD, and HMBC spectra.

not correspond to known compounds and hence suggested the presence of dideoxysugar-based novel small molecules in *P. pacificus.*

The structures of the unknown compounds of interest were enriched via preparative HPLC (**Appendix Section B.1.6**) and their structures assigned based on subsequent NMR (^{1}H, dqfCOSY, and HMBC) and molecular formula as established from high resolution MS analyses (**Figures 3.3, Appendix Figure B.1**, and **Appendix Table B.1**). The most abundant ascaroside derivative, named pasc#9 (see **Appendix Section B.1.1** for nomenclature) was proposed as an *N*-succinyl-1-phenylethanolamide linked to ascarylose by way of a 4-hydroxypentanoic acid chain (**Figures 3.3** and **3.4b**). This compound was accompanied by two *dimeric* ascaroside derivatives, a dimer (dasc#1)

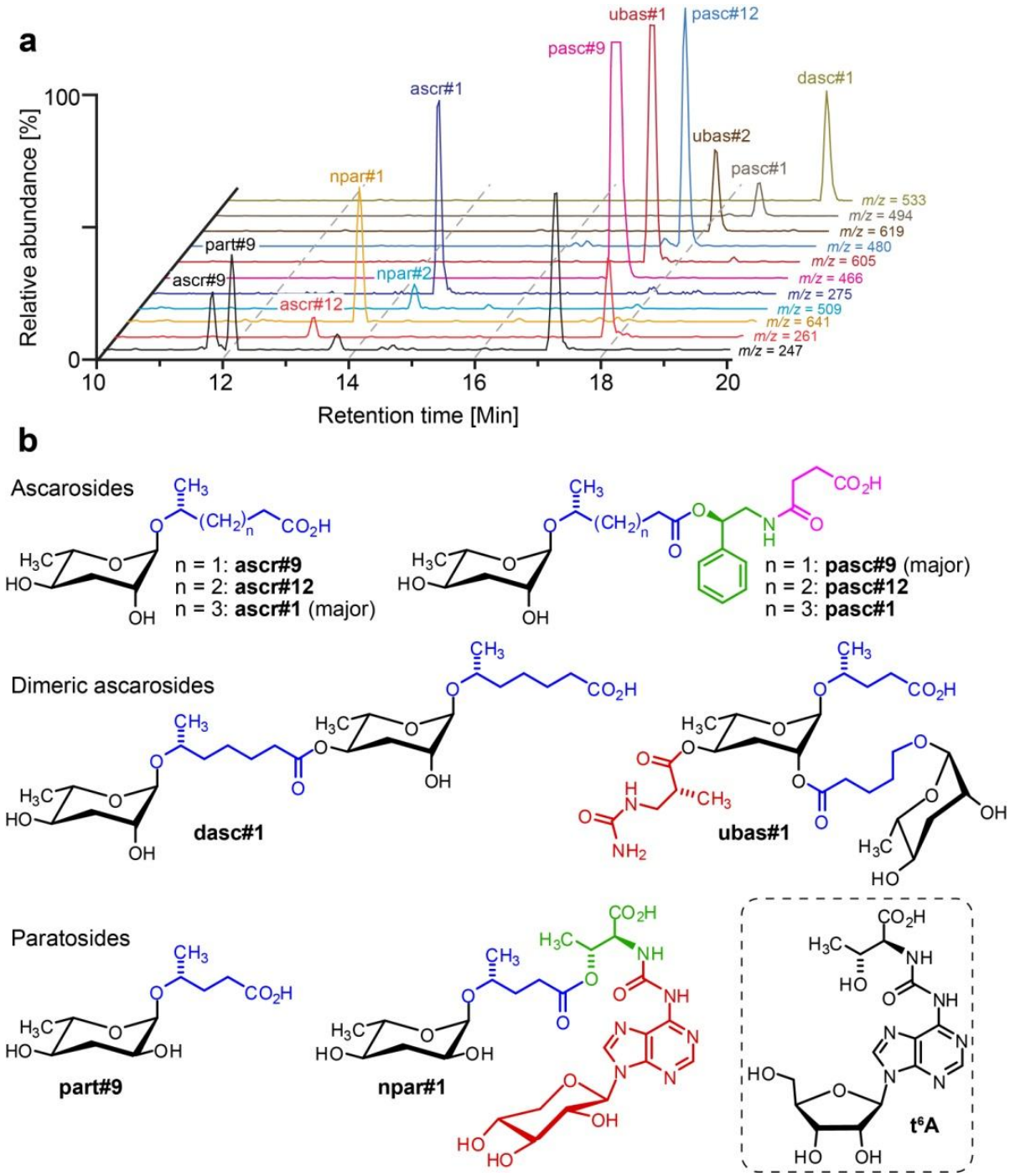
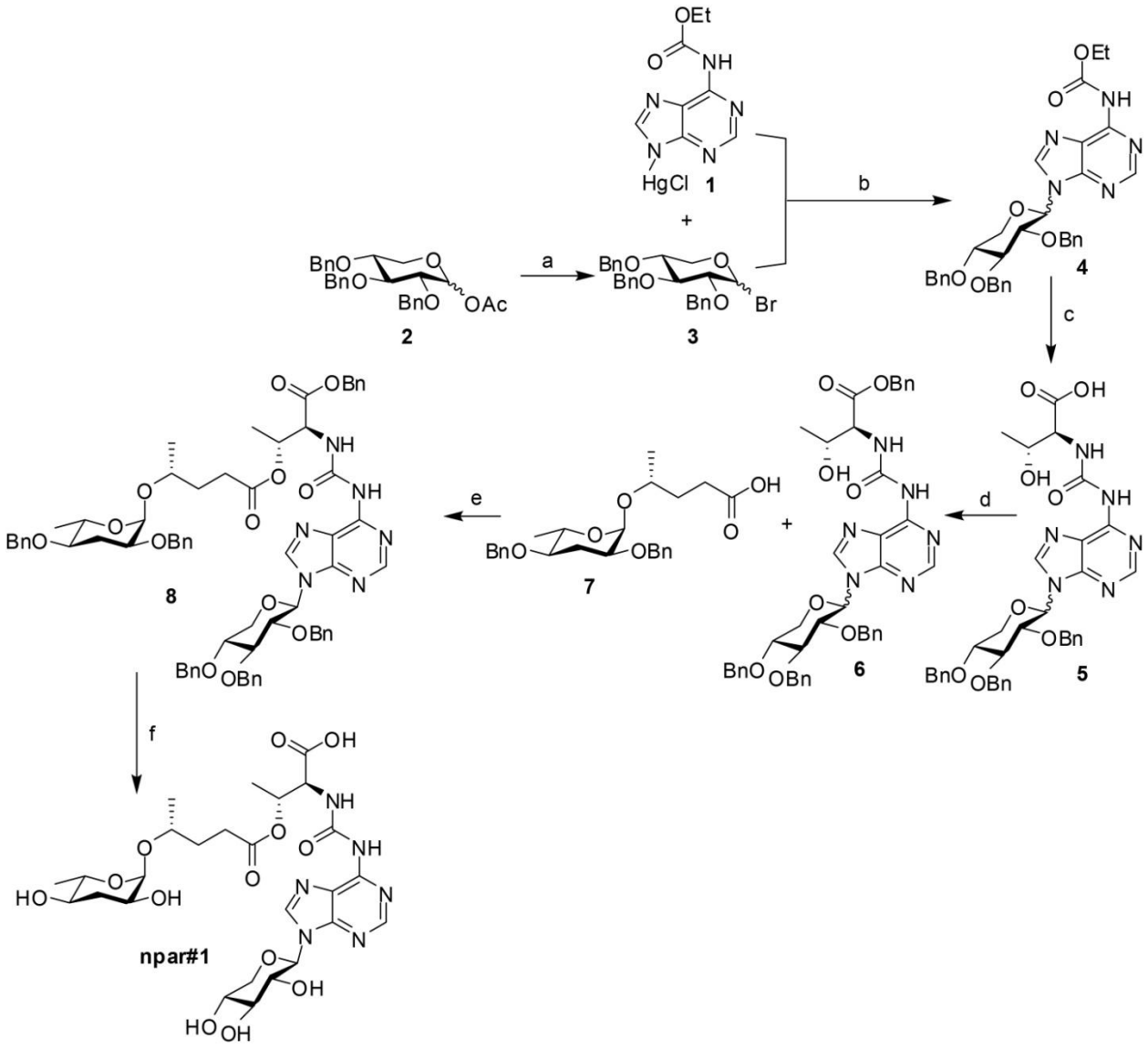


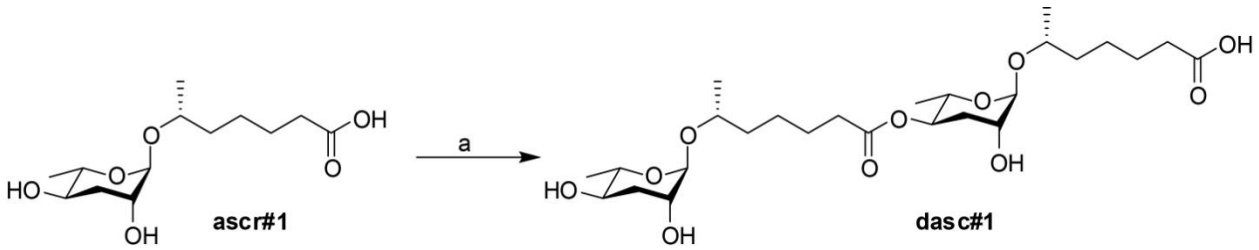
Figure 3.4: (**a**) HPLC-MS analysis showing molecular ion traces for small molecules identified from *P. pacificus*. See **Appendix Figure B.1** for HPLC-MS/MS analysis**.** (**b**) Major components of the *P. pacificus* exo-metabolome derived from assembly of building blocks from fatty acid (blue), carbohydrate (black), amino acid (green), and nucleoside (red) metabolism, as well as TCA cycle-derived succinate (magenta). Also shown is the related hyper-modified tRNA nucleoside, N^6-threonylcarbamoyl adenosine (t6A).

of the known ascr#1 in which one ascr#1 unit is attached to carbon 4 of the other ascr#1 unit, and a second dimer (ubas#1) consisting of ascr#9 to which the (ω)-oxygenated ascaroside oscr#9 is attached at position 2 (**Figures 3.3** and **3.4b**). This second dimer, ubas#1, also has a 3-ureido isobutyrate moiety attached to carbon 4. To the author's knowledge, neither dimeric ascarosides nor ureido isobutyrate-substituted metabolites have previously been reported from nature. These ascarosides were accompanied by two abundant metabolites that included paratose instead of ascarylose, npar#1 and part#9 (**Figures 3.3** and **3.4b**). In npar#1, the paratose moiety was linked to a short lipid side chain, which in turn was connected to threonine. This amino acid was connected further by way of a carbamoyl group to a derivative of the nucleoside adenosine (**Figures 3.3** and **3.4b**). Strikingly, this adenosine was found to include a xylopyranose, and not ribofuranose or deoxyribofuranose as in DNA, RNA, and known nucleoside-based signaling molecules. The accompanying part#9 represents the paratose and side-chain portions of npar#1 (**Figures 3.4b**).

To exclude the possibility that the identified compounds are bacterial metabolites, the author additionally analyzed the metabolome of the *E. coli* OP50 bacteria used as food for *P. pacificus*. None of the identified compounds (pasc#9, dasc#1, ubas#1, npar#1, and part#9) were found to be present in the bacterial metabolome (**Appendix Figure B.2**). Furthermore, the author showed that all identified compounds are also produced in *P. pacificus* cultures fed with *Pseudomonas sp.* instead of *E. coli* as well as in axenic[9,23] (bacteria-free) cultures (**Appendix Figure B.3**).



Scheme 3.1: Overview of synthesis of npar#1. Reagents and conditions: (**a**) TMSBr, DCM, -40 °C to r.t.; (**b**) toluene, reflux; (**c**) L-threonine, pyridine, 107 °C; (**d**) 2-benzyloxy-1-methylpyridinium triflate,[24] Et_3N, $PhCF_3$, 83 °C; (**e**) EDC, DMAP, DCM; (**f**) 10% Pd/C, H_2 (*g*) 10% formic acid in MeOH. Details of synthetic procedures in



Scheme 3.2: Overview of synthesis of dasc#1. Reagents and conditions: (**a**) EDC, DMAP, DMF.

3.3. Total synthesis for the identified metabolites and assignment of their relative and absolute stereochemistries: To confirm these structural assignments, determine the stereochemistry, and explore their biological functions the author in collaboration with Mr. Joshua J. Yim and Dr. Stephan H. von Reuss, Schroeder group, designed total syntheses for each of the proposed structures, taking advantage of their modular nature. The author designed and completed the total syntheses of npar#1 and dasc#1 outlined in **Schemes 3.1** and **3.2** (see **Appendix Sections B.3.1** and **B.3.2** for detailed synthetic procedures). Synthesis of pasc#9, ubas#1, part#9, and several non-natural isomers of pasc#9, npar#1, and part#9 were completed by Mr. Yim and Dr. von Reuss (detailed reaction conditions are reported in Ref. 19). Mr. Yim also provided the dibenzyl-part#9 (**7**, **Scheme 3.1**), a key intermediate for npar#1 synthesis to the author.

Comparison of the ^{1}H NMR spectra of natural pasc#9 with synthetic pasc#9-diastereomers including either a (*R*)-*N*-succinyl-1-phenylethanolamide moiety or a (*S*)-*N*-succinyl-1-phenylethanolamide moiety established the configuration of the *N*-succinyl-1- phenylethanolamide moiety in natural pasc#9 as *R* (**Figures 3.4b** and **3.5**). Similar comparison of dqfCOSY spectra of natural dasc#1 with an authentic synthetic sample allowed the author to confirm the structural and stereochemical assignments for dasc#1 (**Figure 3.4b** and **Appendix Figure B.4**).

Comparison of the HPLC-MS retention times of natural part#9 and two synthetic part#9 diastereomers (D-paratosyl-4*R*-hydroxypentanoic acid and D-paratosyl-4*S*-hydroxypentanoic acid, **Figure 3.6**) showed that natural part#9 is either D-paratosyl-4*S*-hydroxypentanoic acid or its enantiomer L-paratosyl-4*R*-hydroxypentanoic acid. (**Figure 3.6**).

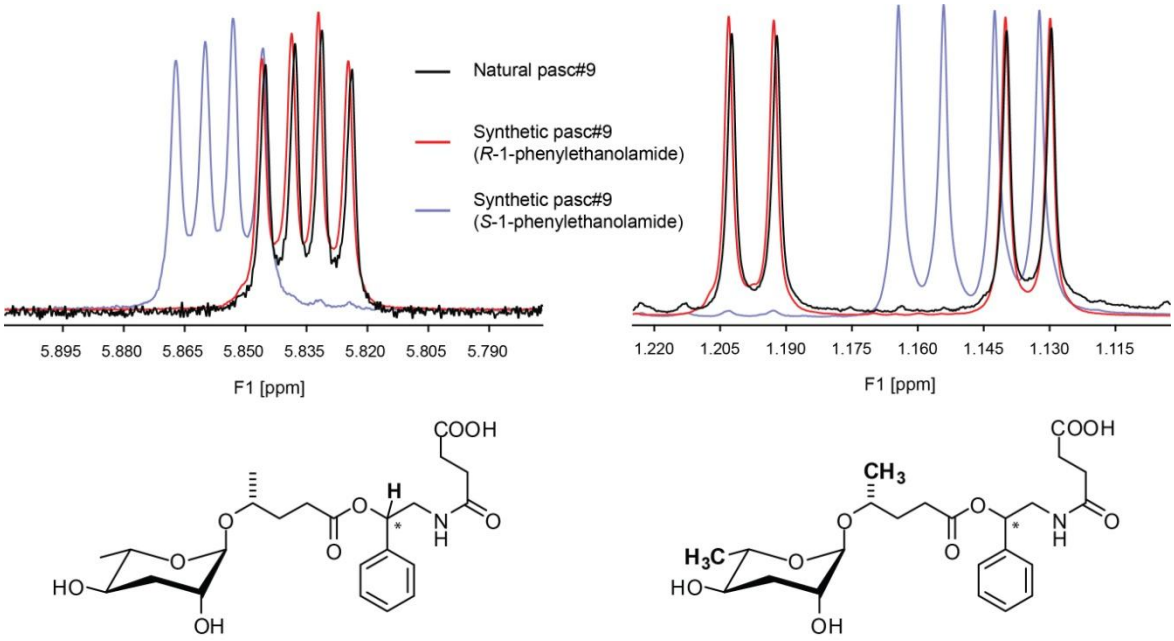


Figure 3.5: Comparison of sections of ^{1}H-NMR spectra of natural pasc#9 (black), synthetic pasc#9 including a (*R*)-*N*succinyl-1-phenylethanolamide moiety (red), and synthetic pasc#9 including a (*S*)-*N*-succinyl-1-phenylethanolamide moiety (blue) shows that natural pasc#9 contains (*R*)-*N*-succinyl-1-phenylethanolamide.

The author initially assumed that part#9 and npar#1 contained D-paratose, which had previously been described from bacteria, whereas L-paratose had not been found in nature. Furthermore, D-paratose is a putative intermediate in the biosynthesis of L-ascarylose,[22] on which all ascarosides in nematodes are based.[8,21,25] However, the HPLC retention time of a synthetic npar#1 diastereomer including D-paratosyl-4*S*-hydroxypentanoic acid did not match the data obtained for natural npar#1 (**Figure 3.7a**). Therefore, the author concluded that npar#1 must be based on L-paratosyl-4*R*-hydroxypentanoic acid, which was confirmed by synthesizing the corresponding npar#1 diastereomer and comparing its ^{1}H NMR and HPLC-UV retention times with those of natural npar#1 (**Figure 3.7b,c**). Using chiral derivatization agents (Mosher's acid chlorides, **Appendix Section B.3.3**),[26] the author further showed that part#9 is also

based on L-paratose (**Appendix Figure B.5**). The sugar L-paratose has not previously been found in nature; however, its occurrence in nematodes might result from epimerization of L-ascarylose at position 2.

Similarly, comparison of the NMR spectra and HPLC retention times allowed the author to corroborate the structure and assign the stereochemistry of ubas#1 (**Figure 3.4b** and **Appendix Figure B.6**). The *R*-configuration of the 3-ureido isobutyrate moiety in ubas#1 is consistent with its likely origin from thymine catabolism.[27]

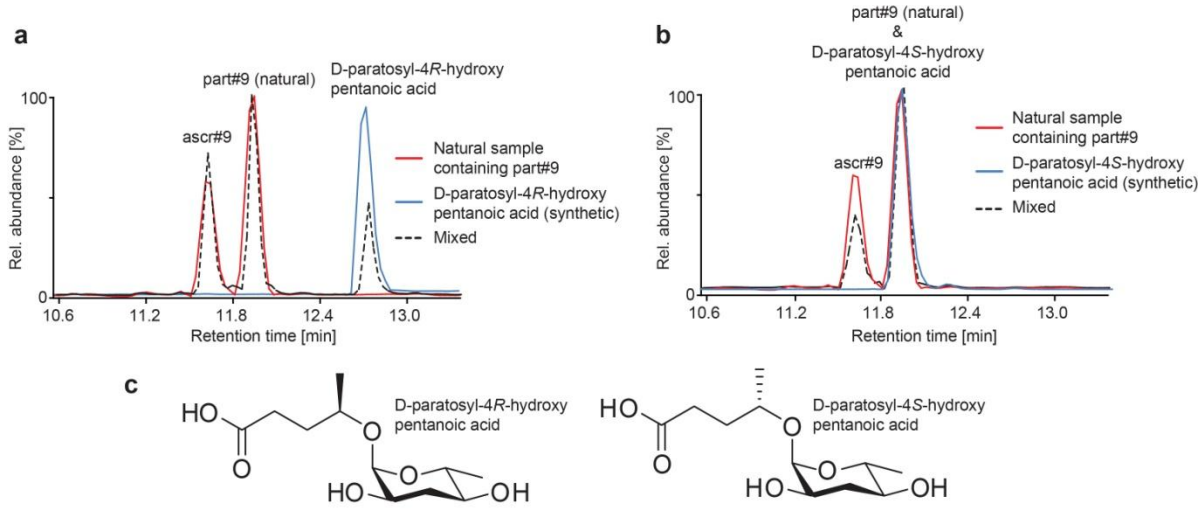


Figure 3.6: Comparison of HPLC-MS retention times (ESI-, ion chromatogram for *m/z* = 247) of natural mixture of ascr#9 and part#9 (red), synthetic samples of part#9 (blue), and a 1:1 mixture of the natural and synthetic sample (dotted black). (**a**) Synthetic D-paratosyl-4*R*-hydroxypentanoic acid. HPLC-retention times do not match natural part#9, indicating that neither D-paratosyl-4*R*-hydroxypentanoic acid nor its enantiomer could be natural part#9. (**b**) Synthetic D-paratosyl-4*S*-hydroxypentanoic acid. HPLC-retention times of D-paratosyl-4*S*-hydroxypentanoic acid match that of natural part#9. This indicates that natural part#9 is either D-paratosyl-4*S*-hydroxypentanoic acid or its enantiomer L-paratosyl-4*R*-hydroxypentanoic acid. (**c**) Structures of D-paratosyl-4*R*-hydroxypentanoic acid and D-paratosyl-4*S*-hydroxypentanoic acid (synthesized by Mr. Yim and reported in Ref. 19).

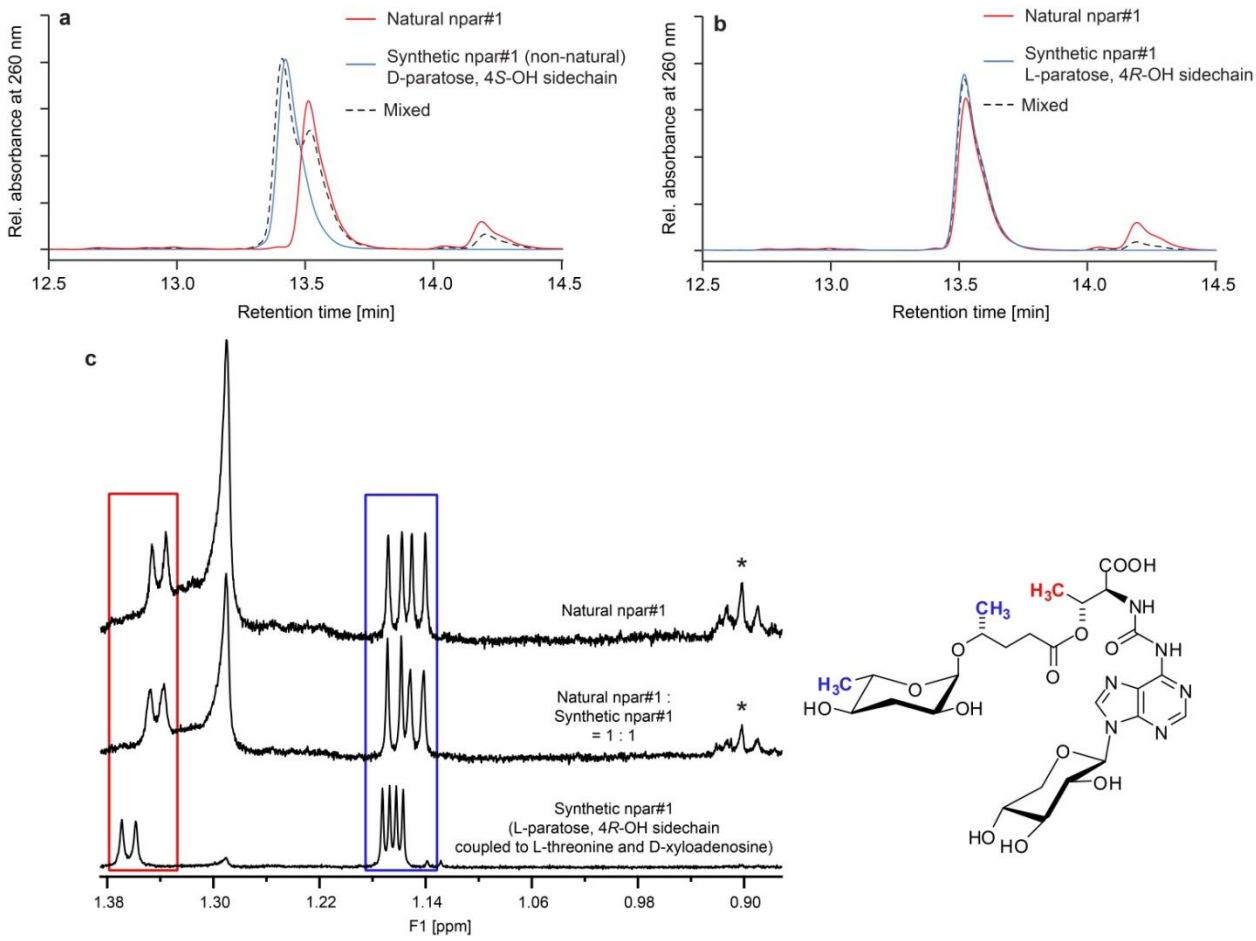


Figure 3.7: Comparison of HPLC-UV (260 nm) retention times of natural sample containing npar#1 (red), synthetic samples of npar#1 (blue), and mixtures of the natural and synthetic samples (dotted black). (**a**) Synthetic npar#1 diastereomer derived from D-paratosyl-4*S*-hydroxypentanoic acid coupled to L-threonine and D-xyloadenosine. HPLC-retention times do not match that of natural npar#1. (**b**) Synthetic npar#1 diastereomer derived from L-paratosyl-4*R*-hydroxypentanoic acid coupled to L-threonine and D-xyloadenosine. HPLC-retention times match that of natural npar#1. (**c**) Comparison of sections of ^{1}H-NMR spectra of synthetic npar#1 derived from L-paratosyl-4*R*-hydroxypentanoic acid coupled to L-threonine and D-xyloadenosine (bottom), natural npar#1 (top), and a 1:1 mixture of the two (middle) shows that changes in pH and concentrations affect the shifts of the three characteristic methyl doublets (indicated by the red and blue boxes in the figure and colored atoms in the accompanying structure). In the mixed sample however, no new peaks show up and the relative intensity of the methyl doublets increases in comparison to unrelated peaks in the natural sample (marked with *). In combination with the HPLC-UV results from (**a**) and (**b**), these findings show that natural npar#1 consists of L-paratosyl-4*R*-hydroxypentanoic acid coupled to L-threonine and D-xyloadenosine.

3.4. Identification of related small molecules using high-resolution HPLC-MS/MS reveals specificity of modular assembly: Next the author asked whether assembly of the identified small molecules from sugar, amino acid, lipid, and nucleoside-derived building blocks is selective. To address this, the author carefully re-analyzed the entire *P. pacificus* exo-metabolome by high-resolution HPLC-MS/MS, quantified the identified compounds using synthetic standards (**Table B.1**), and screened for homologues or alternative combinations of the primary metabolism-derived building blocks in the identified structures (**Section 3.2**). The author found that pasc#9 is accompanied by trace amounts of two homologues including six- and seven-carbon side chains, which were also detected by NMR spectroscopy (**Appendix Figure B.7**). In addition, a small amount of a homologue of ubas#1 as well as a derivative of npar#1 whose MS data indicated loss of the xylose (ubas#2 and npar#2, **Appendix Table B.1**) was detected.

Importantly, the author did not observe any non-specific or seemingly random combinations of building blocks that would suggest a non-enzymatic genesis of the identified compounds.

3.5. Biological activities exhibited by the novel small molecules: Next synthetic samples of the identified compounds were submitted to Prof. Ralf J. Sommer's group for testing their activities in the *P. pacificus* dauer- and mouth-form-dimorphism assays (see **Appendix Sections B.1.7** and **B.1.8**). As expected from previous studies that showed that *C. elegans*-derived extracts are not active in the *P. pacificus* mouth-form dimorphism and dauer assays,[28] ascr#1, a compound abundantly excreted

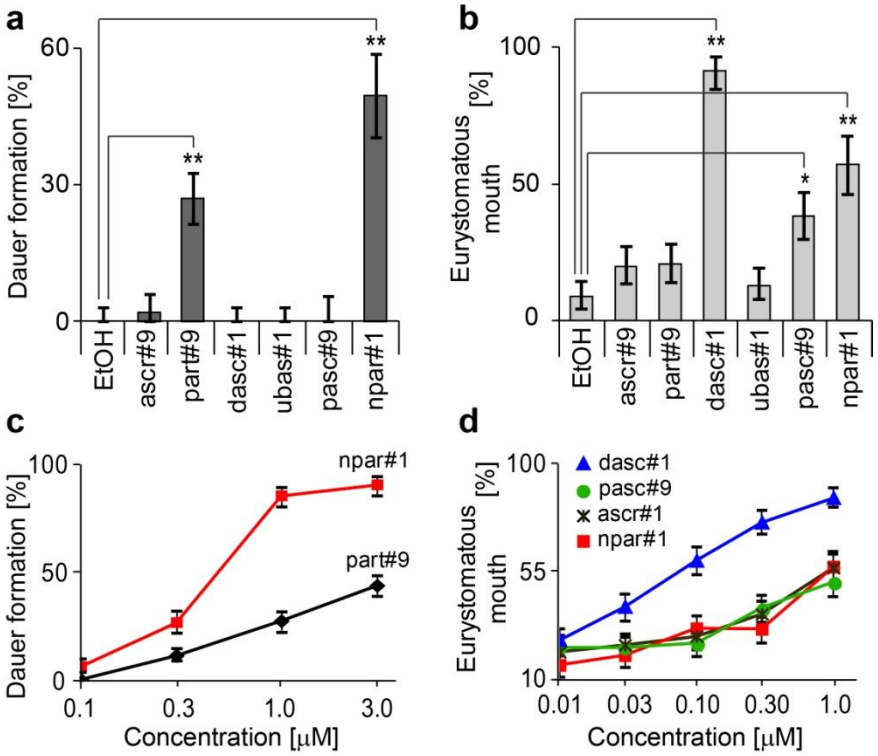


Figure 3.8: Regulation of mouth dimorphism and dauer induction by synthetic samples of identified *P. pacificus* metabolites. All experiments were performed in triplicate for each treatment. (**a**,**b**) Compounds were assayed at 1 μM concentration (Significance, *=p<0.01, **=p<0.001, ***=p<0.001, **Appendix Section B.1.9**). (**c**,**d**) Compounds with significant activity at 1 μM (p<0.01) were subsequently tested at a range of concentrations.

by *C. elegans*,[8,29] was found to have no dauer-inducing activity in *P. pacificus*, even at concentrations higher than what is physiologically observed (20 μM, **Appendix Figure B.8**). In contrast, physiological concentrations of the nucleoside derivative npar#1 strongly induced dauer formation and appear to account for most of the reported dauer-inducing activity in the non-fractionated exo-metabolome (**Figures 3.8a,c**).[18] Additionally, weaker dauer induction than npar#1 was observed with part#9, whereas all other compounds tested were inactive in this assay.

Testing synthetic compounds in the mouth dimorphism assay revealed that the *dimeric* compound dasc#1, which was inactive in the dauer-formation assay, strongly

induces the eurystomatous mouth form (**Figures 3.8b,d**). In addition, weaker induction of the eurystomatous mouth form was observed for high concentrations of pasc#9, ascr#1, and npar#1, whereas ascr#9 and part#9 as well as the dimeric ubas#1 were inactive at physiological concentrations in the wild-type strain tested (**Figure 3.8b** and **Table B.1**).

3.6. Natural variation of small molecule production in *P. pacificus* wild isolates: The dauer stage plays a very important role in *P. pacficus'* survival in the wild. These worms stay as dauers on scarab beetles for the majority of their lifespan, resuming development only when the beetle dies, feeding on the microbes that infest the beetle carcass. Melanie G. Mayer at the Sommer group explored *P. pacificus* dauer entry and exit using a natural variation approach.[18]

Dauer exit study with 8 *P. pacificus* wild isolates from around the world showed that dauer larvae in *P. pacificus* can survive for up to 1 year under experimental conditions (with different mean survivals for the wild isolates), significantly higher than the *C. elegans* wild-type strain (N2, Bristol).[18] In addition, dauer pheromones from 16 *P. pacificus* wild strains were isolated, and subsequently tested for natural variation in pheromone production and sensitivity in cross-reactivity assays. Interestingly, majority of the strains (13 out of 16) produce a pheromone (a blend of small molecules) that is more active on individuals of another genotype.[18] These results were intriguing to the author as they suggested that there exists a significant natural variation in small

molecule production in different *P. pacificus* wild isolates, i.e. each *P. pacificus* natural isolate produces a different blend of small molecules.

To explore natural variation of small molecule production in *P. pacificus*, the author in a proof-of-principle study, screened the metabolomes of 6 *P. pacificus* wild (natural) isolates using HPLC-MS for qualitative and quantitative variations in the newly identified small molecules from *P. pacificus* RS2333.

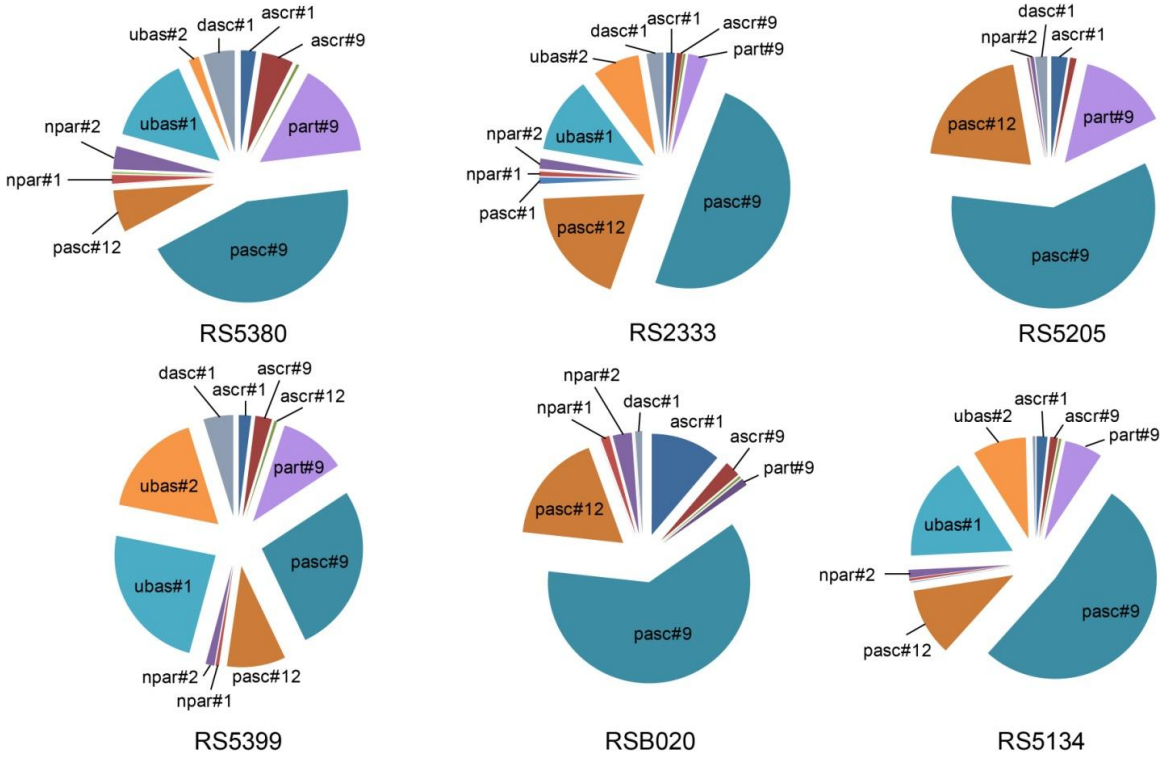


Figure 3.9: Exo-metabolome small molecule profiles of 6 *P. pacificus* wild isolates, represented by the mole percentage of the 12 identified small molecule signals from strain RS2333 (**Figure 3.4b** and **Table B.1**).

The author found that each of the 6 wild isolates release a unique and strain-specific blend of small molecules (**Figure 3.9**). The observation of a different blend of small molecules by various wild isolates potentially explains the previously observed natural variation in dauer pheromone efficacy in *P. pacificus*.[18] Interestingly, while most

compounds were present in all strains, ubas#1 and ubas#2 were not detected in the metabolomes of RS5205 and RSB020 (**Figures 3.9** and **3.10**, **Appendix Figures B.9** and **B.10**). This indicates that certain strains may have evolved to lose the capacity to biosynthesize ubas- and plausibly other compounds.

Alternatively, these strains may produce the apparent deficient compounds only under specific conditions that were not explored in this study, such as elevated temperatures. Also to note is *P. pacificus* is androdioecious, i.e. comprised of males and hermaphrodites. Under conditions of this natural variation study (*P. pacificus* liquid cultures, **Appendix Section B.1.4**), ~99.8% of the worms were hermaphrodites, hence low level production of a specific compound (e.g. ubas#1) only by the male population of a particular strain (e.g. RSB020) will not be observed, given the dynamic range and sensitivity of the techniques used. Further studies, beyond the scope of this dissertation, may reveal the biochemical reasons for the absence of certain compounds in a particular strain.

The author further compared the small molecule profiles of the exo-metabolome (liquid culture supernatant extracts) and endo-metabolomes (worm pellet extracts) for the 6 *P. pacificus* wild isolates to see if there were evidence for differential regulation in small molecule release. The author found that the worm pellet small molecule profiles differ significantly from those excreted into the supernatant. The endo-metabolome profiles were dominated by simple ascarosides, namely ascr#1, ascr#9, and ascr#12, which were only minor components of the exo-metabolome (**Figure 3.10** and **Appendix Figure B.10**). The more complex small molecules such as ubas#1, dasc#1, and npar#1 showed clear preference for being released into the supernatant for most strains

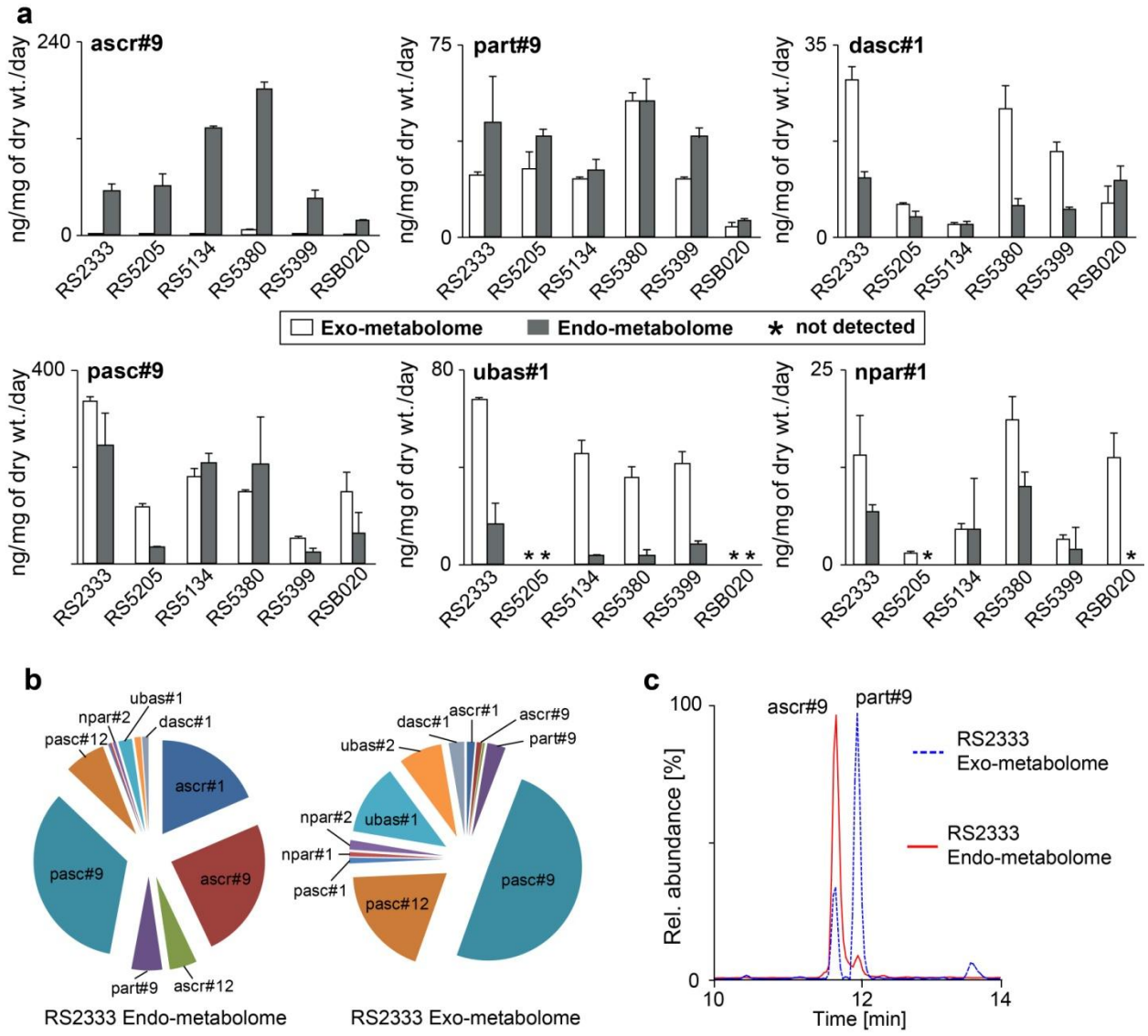


Figure 3.10: (**a**) Absolute comparison of representative small molecules in the exo- and endo-metabolomes of 6 *P. pacificus* wild isolates (also see **Appendix Figure B.10**). (**b**) Relative comparison of the exo- and endo-metabolome small molecule profiles for RS2333, represented by the mole percentage of the identified small molecules (**c**) HPLC-MS ion traces comparing the relative proportions of ascr#9/part#9 in the RS2333 exo- and endo-metabolomes.

(**Figure 3.10** and **Appendix Figure B.10**). Notably, ascr#9 and part#9 (**Figure 3.4b**) are chemically very similar and elute very closely in the HPLC method (**Figure 3.10c**, **Table B.1**), hence if polarities of the two compounds were the only factor influencing compound release, there should not be any detectable difference in their excretion

profiles. However, ascr#9 is almost entirely retained in the worm body, whereas, part#9 shows no preference for being released or retained (**Figure 3.10a**). These data suggest that *P. pacificus* exhibits remarkable control over the release of small molecule signals that is conserved across several natural isolates from across the world.

3.7. Discussion: The identified small molecules in *P. pacificus* appear to integrate specific inputs from the major primary metabolic pathways: fatty acid, carbohydrate, amino acid, and nucleoside metabolism (**Figure 3.11**). Specific assembly of building blocks from these pathways using ester and amide linkages generates the modular molecules identified, which are further distinguished from previously known animal metabolites by the inclusion of L-paratose and xylopyranose-based adenosine. This unprecedented nucleoside is likely derived from metabolism of canonical (ribo)threonylcarbamoyl adenosine (t^6A, **Figures 3.4b** and **3.11**), a highly conserved nucleoside found directly adjacent to the anticodon triplet of a subset of tRNAs.[30] t^6A plays an important role in maintaining translational fidelity; however, it usually accounts for only a very small fraction of tRNAs, and the production of large quantities of xylopyranose derivative in *P. pacificus* is surprising.

Results from biological testing of synthetic samples show that adult phenotypic plasticity and larval development in *P. pacificus* are controlled by distinct yet partially overlapping sets of signaling molecules. Whereas mouth-form dimorphism is primarily regulated by dasc#1, the product of highly specific ascaroside dimerization, dauer

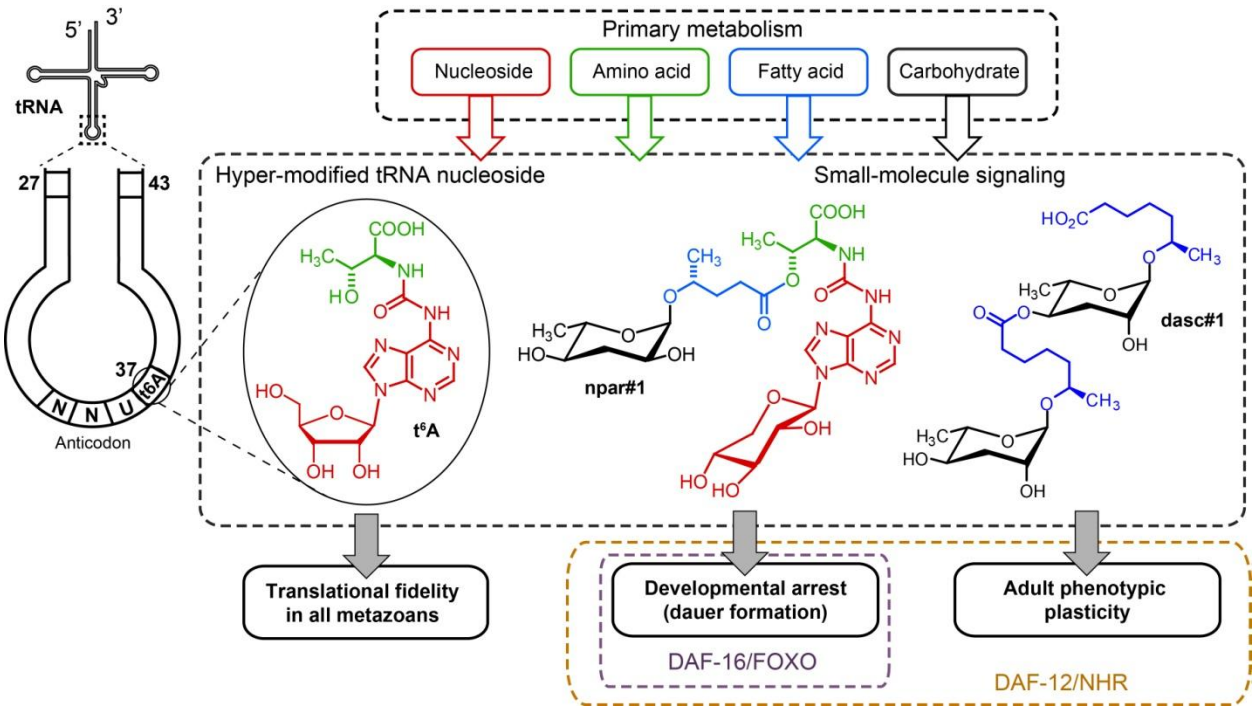


Figure 3.11: Small-molecule signaling in nematodes occupies a central position connecting primary metabolism to evolutionarily conserved transcription factors, including DAF-16/FOXO and the nuclear hormone receptor (NHR) DAF-12, a vitamin D and liver X receptor homologue. Comparison of the structure of npar#1 with t^6A suggests the unusual xylopyranose-based adenosine in npar#1 is probably derived from metabolism of canonical ribofuranose-based adenosine in t^6A. t^6A is a conserved hyper-modified nucleoside found directly adjacent to the anticodon triplet of a subset of tRNAs, plays an important role in maintaining translational fidelity in all metazoans, and putatively is biosynthesized utilizing primary metabolic building blocks.

formation is controlled by a molecule combining a paratoside with an unusual nucleoside. Previous work showed that the signaling molecules controlling phenotypic plasticity in *P. pacificus* act upstream of evolutionarily conserved transcription factors, including DAF-16/FOXO and the nuclear hormone receptor DAF-12 (**Figure 3.11**),[31,32] whereby *daf-12* is required for both dauer induction and mouth-form dimorphism, whereas *daf-16* is required for dauer induction but dispensable for regulation of mouth-form dimorphism.[32] Therefore, the different subsets of small molecules regulating dauer formation and mouth-form dimorphism appear to target different downstream effectors.

Based on the recent identification of several G-protein coupled receptors (GPCRs) of the *C. elegans* dauer pheromone,[33-35] it is probable that GPCRs expressed in specific chemosensory neurons connect npar#1 and dasc#1 with their respective downstream pathways in *P. pacificus*.

The results presented further demonstrate species-specific co-option of small molecule biosynthetic pathways for regulating different aspects of development, as ascr#1 contributes to dauer in *C. elegans*,[29] whereas its dimer dasc#1 regulates adult morphology in *P. pacificus*. Similar to the multifunctional signaling properties of ascarosides in *C. elegans*,[8-13] it is possible that dasc#1, npar#1, and the other identified compounds serve additional signaling functions in *P. pacificus*, for example, in mediating social behaviors and sex-specific attraction.

P. pacificus provides a powerful system for using Next-Generation Sequencing-based population genomics and genome-wide association studies (GWAS) to study small molecule biosynthesis. The genome of 104 strains has already been fully analyzed; another set of 154 complete genome sequences has just been obtained and is currently in the bioinformatics pipeline (Roedelsperger et al., submitted; personal communication with Prof. Ralf J. Sommer). In a proof-of-principle study reported in this dissertation chapter, the analysis of the metabolomes of 6 different *P. pacificus* wild isolates using HPLC-MS revealed that the relative abundances of the monitored compounds differ between strains and that most strains produce all compounds. However, the author also found two strains defective in production of ubas#1 and ubas#2. The ubas- defective strains are not genetically clustered and originate from different geographical locations. Extending the HPLC-MS-based metabolomic analysis

to include all of the wild isolates (~360) may identify additional strains defective in ubas, as well as other small molecule production. Corresponding bioinformatic analysis based on comparison of genome data from the compound-defective strains with that of the compound-producing strains will provide a unique route to identifying candidate genes responsible for small molecule biosynthesis and homeostasis.

In summary, the results reported in this chapter provide direct evidence for small-molecule control of adult phenotypic plasticity in a metazoan that relies on conserved endocrine signaling pathways. Whether the biosynthesis of the identified signaling molecules involves conserved pathways or depends on dedicated enzymes specific to *P. pacificus* (and perhaps other nematodes) is not known. However, the finding that the *P. pacificus* compounds are derived from assembly of modified primary metabolites suggests that their biosynthesis is largely based on conserved biochemical pathways (**Figure 3.11**). Notably, the *P. pacificus* genome contains more than 25,000 predicted genes with many specific gene duplication events among genes encoding primary metabolic enzymes.[36,37] Supporting the involvement of primary metabolism in the biosynthesis of nematode signaling molecules, recent investigations of ascaroside biogenesis in *C. elegans* showed that the lipid-like ascaroside side chains are derived from conserved peroxisomal-β-oxidation (**Chapter 2)**.[8,38] Known signaling molecules and co-factors in higher animals, for example, *S*-adenosyl methionine or phosphatidylinositols, often rely on the combination of building blocks derived from one or two different primary metabolic pathways. The author's results demonstrate that metazoans may extend such strategies to produce signaling molecules of much greater structural complexity, suggesting that detailed spectroscopic re-analysis of

metabolomes from higher animals, including mammals, may also reveal novel types of modular small-molecule signals.

3.8. Author's note: Synthesis of pasc#9, ubas#1, part#9, and several non-natural isomers of pasc#9, npar#1, and part#9 were completed by Joshua J. Yim and Dr. Stephan H. von Reuss, Schroeder research group, Cornell University, Ithaca, NY (reported in Ref. 19). Mr. Yim also provided the dibenzyl-part#9 (**7**, **Scheme 3.1**), a key intermediate for npar#1 synthesis to the author. 1-3 L liquid cultures of *P. pacificus* strains RS2333, RS5205, RS5380, RS5134, RS5399, and RSB080 and the preliminary fractionation of RS2333 exo-metabolome was done by Dr. Akira Ogawa. Biological assays to test for activities of synthetic compounds identified from *P. pacificus* included in this dissertation chapter were performed by Dr. Akira Ogawa (dauer assay) and Dr. Erik J. Ragsdale (mouth-form dimorphism assay). Dr. Ogawa and Dr. Ragsdale are postdoctoral researchers in Prof. Ralf J. Sommer's group at Max Planck Institute for Developmental Biology, Tuebingen, Germany.

REFERENCES

(1) Meinwald, J. *Journal of Natural Products* **2011**, *74*, 305.

(2) Walsh, C. T. *Acc Chem Res.* **2008**, *41*, 4.

(3) Meier, J. L.; Burkart, M. D. *Chem Soc Rev.* **2009**, *38*, 2012.

(4) Piel, J. *Nat Prod Rep* **2009**, *26*, 338.

(5) Piel, J. *Nat Prod Rep* **2004**, *21*, 519.

(6) Dossey, A. T. *Nat Prod Rep.* **2010**, *27*, 1737.

(7) Gronquist, M., Schroeder, F. C., Lew, M. & Hung-Wen, L. *Comprehensive Natural Products II*; Elsevier, 2010.

(8) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc.* **2012**, *134*, 1817.

(9) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol.* **2012**, *10*, e1001237. Epub 2012 Jan 10.

(10) Izrayelit, Y.; Srinivasan, J.; Campbell, S. L.; Jo, Y.; von Reuss, S. H.; Genoff, M. C.; Sternberg, P. W.; Schroeder, F. C. *ACS chemical biology* **2012**, *7*, 1321.

(11) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A.* **2009**, *106*, 7708.

(12) Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E. A.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *Nature* **2008**, *454*, 1115.

(13) Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J. *Nat Chem Biol.* **2007**, *3*, 420.

(14) Hong, R. L.; Sommer, R. J. *Bioessays.* **2006**, *28*, 651.

(15) Rae, R.; Riebesell, M.; Dinkelacker, I.; Wang, Q.; Herrmann, M.; Weller, A. M.; Dieterich, C.; Sommer, R. J. *J Exp Biol* **2008**, *211*, 1927.

(16) Weller, A. M.; Mayer, W. E.; Rae, R.; Sommer, R. J. *J Parasitol.* **2010**, *96*, 525.

(17) Bento, G.; Ogawa, A.; Sommer, R. J. *Nature.* **2010**, *466*, 494.

(18) Mayer, M. G.; Sommer, R. J. *P Roy Soc B-Biol Sci* **2011**, *278*, 2784.

(19) Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C. *Angew Chem Int Ed Engl* **2012**, *51*, 12438.

(20) Sommer, R. J.; Carta, L. K.; Kim, S. Y.; Sternberg, P. W. *Fundamental and Applied Nematology* **1996**, *19*, 511.

(21) Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W. *Curr Biol* **2012**, *22*, 772.

(22) Thorson, J. S.; Lo, S. F.; Ploux, O.; He, X. M.; Liu, H. W. *J Bacteriol* **1994**, *176*, 5483.

(23) Lu, N. C.; Goetsch, K. M. *Nematologica* **1993**, *39*, 303.

(24) Tummatorn, J.; Albiniak, P. A.; Dudley, G. B. *J Org Chem.* **2007**, *72*, 8962.

(25) Bartley, J. P.; Bennett, E. A.; Darben, P. A. *J Nat Prod* **1996**, *59*, 921.

(26) Hoye, T. R.; Jeffrey, C. S.; Shao, F. *Nat Protoc* **2007**, *2*, 2451.

(27) van Kuilenburg, A. B. P.; Stroomer, A. E. M.; van Lenthe, H.; Abeling, N. G. G. M.; van Gennip, A. H. *Biochem J* **2004**, *379*, 119.

(28) Ogawa, A.; Streit, A.; Antebi, A.; Sommer, R. J. *Curr Biol* **2009**, *19*, 67.

(29) Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K. *Nature.* **2005**, *433*, 541.

(30) Deutsch, C.; El Yacoubi, B.; de Crecy-Lagard, V.; Iwata-Reuyl, D. *J Biol Chem* **2012**, *287*, 13666.

(31) Sommer, R. J.; Ogawa, A. *Curr Biol* **2011**, *21*, R758.

(32) Ogawa, A.; Bento, G.; Bartelmes, G.; Dieterich, C.; Sommer, R. J. *Development* **2011**, *138*, 1281.

(33) McGrath, P. T.; Xu, Y. F.; Ailion, M.; Garrison, J. L.; Butcher, R. A.; Bargmann, C. I. *Nature* **2011**, *477*, 321.

(34) Kim, K.; Sato, K.; Shibuya, M.; Zeiger, D. M.; Butcher, R. A.; Ragains, J. R.; Clardy, J.; Touhara, K.; Sengupta, P. *Science* **2009**, *326*, 994.

(35) Park, D.; O'Doherty, I.; Somvanshi, R. K.; Bethke, A.; Schroeder, F. C.; Kumar, U.; Riddle, D. L. *Proc Natl Acad Sci U S A* **2012**, *109*, 9917.

(36) Borchert, N.; Dieterich, C.; Krug, K.; Schutz, W.; Jung, S.; Nordheim, A.; Sommer, R. J.; Macek, B. *Genome Res* **2010**, *20*, 837.

(37) Dieterich, C.; Clifton, S. W.; Schuster, L. N.; Chinwalla, A.; Delehaunty, K.; Dinkelacker, I.; Fulton, L.; Fulton, R.; Godfrey, J.; Minx, P.; Mitreva, M.; Roeseler, W.; Tian, H.; Witte, H.; Yang, S. P.; Wilson, R. K.; Sommer, R. J. *Nat Genet* **2008**, *40*, 1193.

(38) Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y. *Proc Natl Acad Sci U S A.* **2009**, *106*, 1875.

CHAPTER 4

COMPARATIVE METABOLOMICS REVEALS ENDOGENOUS LIGANDS OF DAF-12, A NUCLEAR HORMONE RECEPTOR REGULATING *C. ELEGANS* DEVELOPMENT

4.1. Introduction: Small-molecule ligands of nuclear hormone receptors (NHRs), an evolutionarily conserved family of ligand-activated transcription factors, control diverse aspects of metazoan metabolism, cell differentiation, development, and aging. Precise knowledge of ligand structures and biosynthetic pathways is essential for understanding NHR function,[1,2] because even small differences in ligand structures may result in dramatic changes of transcriptional activity and specificity.[3,4] However, the endogenous ligands of many NHRs have remained poorly characterized, in part because ligands often constitute very minor components of highly complex animal metabolomes.[5] The free living nematode *C. elegans* has 284 NHRs, allows easy genetic manipulation, and can be grown in large quantities, providing an opportunity to investigate structures, biosynthesis, and functions of NHR ligands in a relatively simple model system.[6,7]

In *C. elegans*, DAF-12, a homolog of vertebrate vitamin D (VDR) and liver X receptors (LXR), functions as a ligand-gated switch that regulates both adult lifespan and larval development.[8-11] The biosynthesis of the steroidal ligands of DAF-12 is controlled by a complex endocrine signaling network, of which many components appear to be highly conserved between *C. elegans* and mammals.[9] Perception of environmental stimuli by chemosensory neurons regulates signaling via the conserved

insulin/IGF and TGF-β pathways, which converge on genes implicated in DAF-12-ligand biosynthesis (**Figure 4.1a**, see **Chapter 1**, **Section 1.3** for a more detailed picture of the signal transduction pathway). Under unfavorable conditions such as overcrowding or scarcity of food, ligand biosynthesis is suppressed, and unliganded DAF-12 interacts with its co-repressor DIN-1.[12] The resulting transcriptional repression of DAF-12 target genes causes developmental arrest and entry into a highly stress-resistant larval stage called the dauer diapause.[13-15] In contrast, favorable environmental conditions trigger upregulation of DAF-12-ligand biosynthesis. DAF-12-ligand binding then results in dissociation of the corepressor DIN-1 to allow expression of DAF-12-target genes, promoting rapid developmental progression from larvae to reproductive adults.[9,12] Based on extensive biochemical studies, two bile acid-like steroids named Δ^4- and Δ^7-dafachronic acid were proposed as endogenous ligands of DAF-12 (**Figure 4.1b**).[16] Central to identification of the dafachronic acids (DAs) as DAF-12-ligand candidates were precursor studies in which a variety of 3-keto sterols were identified as substrates for the cytochrome P450, DAF-9, which had been shown to act upstream of DAF-12 in DAF-12-ligand biosynthesis.[16-18] DAF-9 was further shown to act on the sidechain in these cholestenones introducing a terminal carboxyl group (**Figure 4.1b**).[16] In a separate study, a DAF-12-activating isomer of 3β-hydroxy cholest-5-enoic acid was detected in *C. elegans* metabolite extracts.[19] However, given the very low concentrations of the putative DAF-12-ligands in *C. elegans*, isolation and full spectroscopic characterization of these compounds were not pursued.[16] Although none of the structures of the proposed DAF-12-ligands have been confirmed based on comprehensive spectroscopic analysis of *C. elegans*-derived samples, a biosynthesis

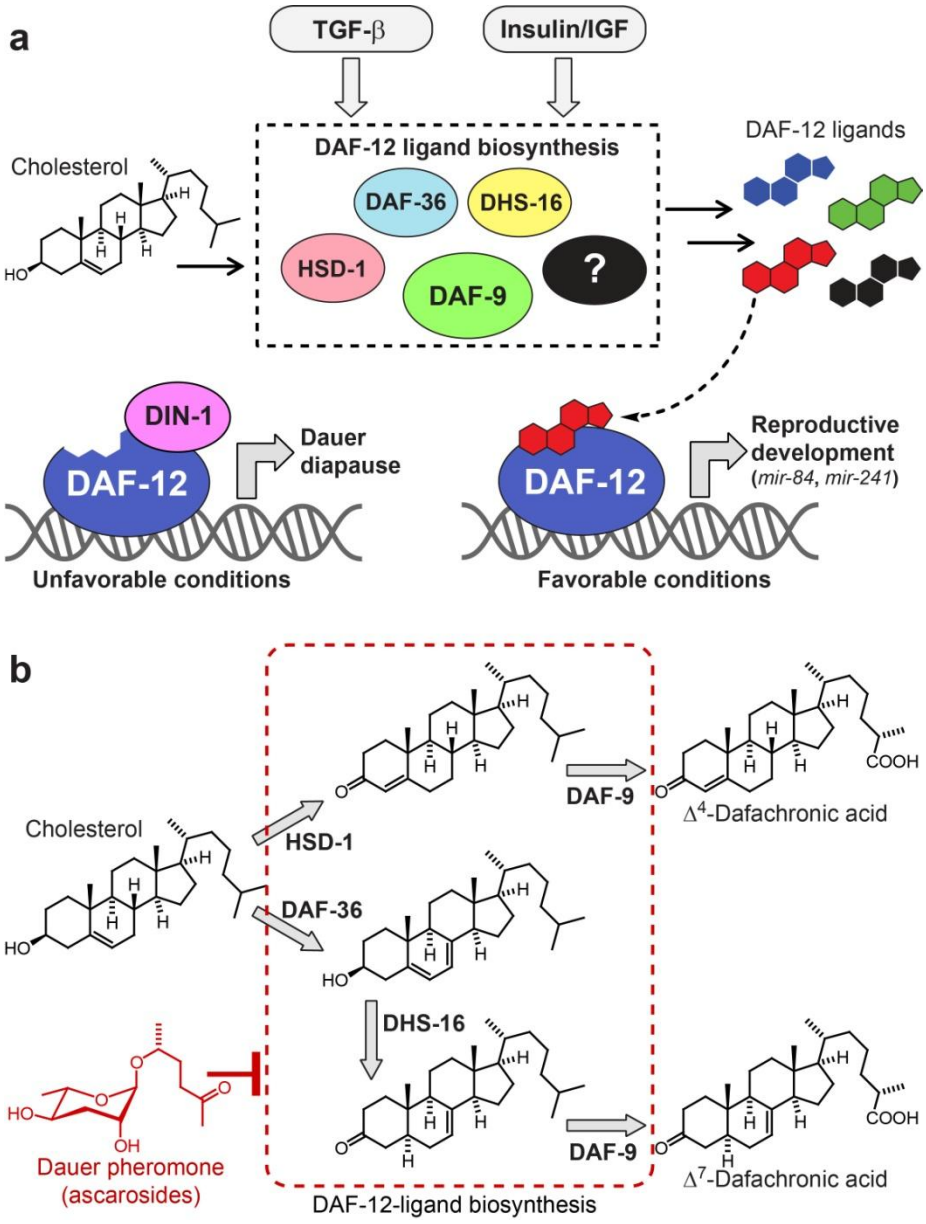


Figure 4.1: Steroidal ligands control *C. elegans* development and lifespan via the nuclear hormone receptor DAF-12. (**a**) Under favorable conditions, signaling via the insulin/IGF and TGF-β pathways drive biosynthesis of steroidal DAF-12-ligands. Liganded DAF-12 promotes development, in part via transcription of the *let-7*-family microRNAs *mir-84* and *mir-241*. Under unfavorable conditions, ligand biosynthesis is inhibited, resulting in interaction of unliganded DAF-12 with its corepressor DIN-1 and target gene repression (see also **Chapter 1**, **Figure 1.4**).[12] (**b**) Previously described DAF-12 ligands and proposed biosynthetic pathway.[20-26] DAF-12 ligand biosynthesis is downregulated in response to dauer pheromone,[15] a blend of ascarosides, e.g. the shown ascr#2 (red).

model has been developed (**Figure 4.1b**), and the biochemical roles of genes proposed to function upstream of DAF-9 in DAF-12-ligand biosynthesis have been studied extensively.[20,22-26] More recent work has shown that the proposed DAF-12-ligands do not explain all DAF-12 associated functions and has suggested a functionally divergent biosynthetic network with the possibility of alternative DAs.[25,26]

In this dissertation chapter, the author describes the identification of the endogenous ligands of DAF-12 using a 2D NMR-based comparative metabolomics approach[27] and subsequent structural confirmation using Selective Ion Monitoring (SIM)-GC/MS. This method revealed steroid molecules with unusual functionalization as novel ligands of DAF-12 in *C. elegans*.

4.2. Customizing a comparative metabolomics approach: The author, in collaboration with Parag Mahanti, Schroeder research group, set out to identify the endogenous ligands of DAF-12 based on direct spectroscopic evidence, in contrast to earlier work that had relied on combining classical genetics and biochemical experiments. For this purpose the author employed a combination of activity-guided fractionation and NMR-based comparative metabolomics via DANS (Differential Analyses by 2D NMR Spectroscopy, **Figure 4.2a**, see also **Chapter 1**, **Section 1.3**). dqfCOSY, a 2D NMR spectroscopy technique (**Chapter 1**, **Section 1.3**) can provide a largely unbiased overview of the metabolome composition, and comparing dqfCOSY spectra of different mutant backgrounds via DANS often permits detection and partial

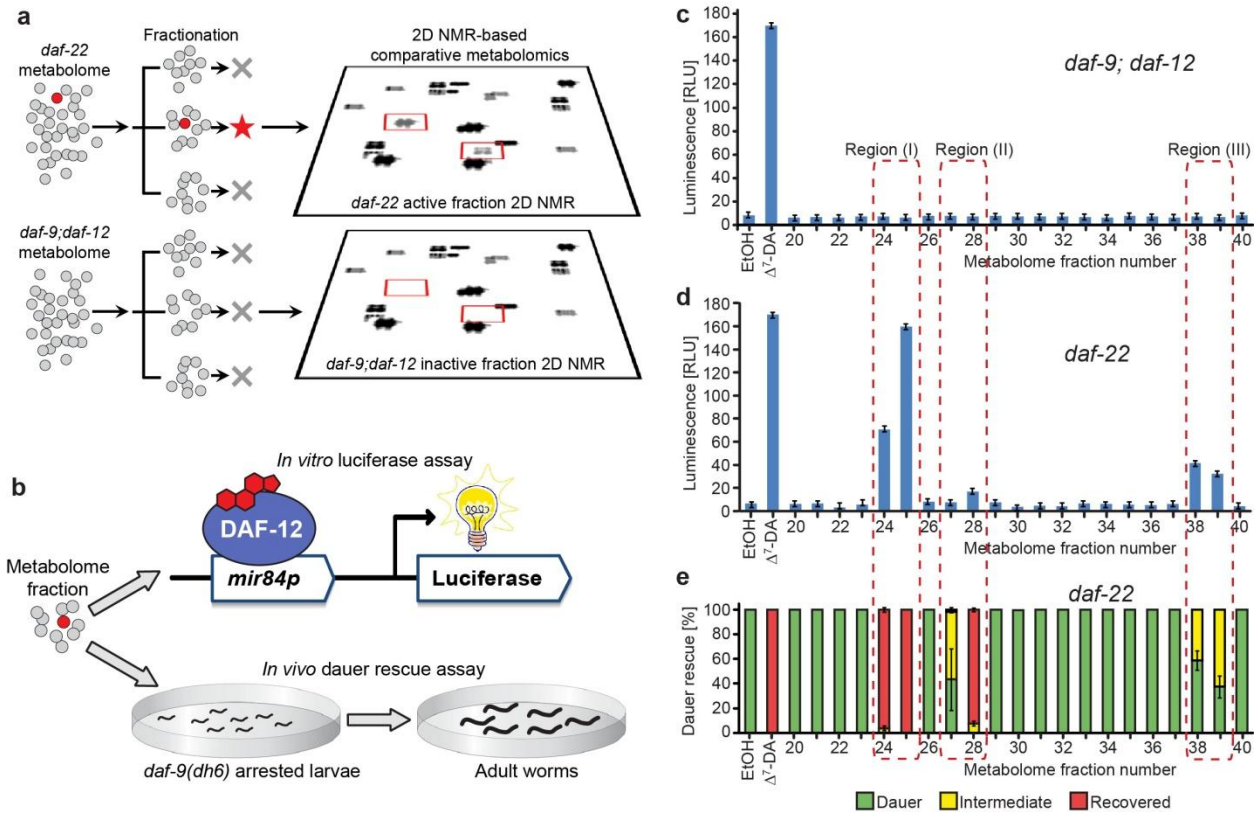


Figure 4.2: Detection of DAF-12-ligands in *C. elegans* mutant metabolomes. (**a**) Fractionation of active, ligand-rich *daf-22* and inactive *daf-9;daf-12* metabolomes is followed by 2D NMR-based comparative metabolomics of active fractions. (**b**) Assessment of DAF-12-ligand content using (1) an *in vitro* luciferase assay in HEK-293T cells transfected with full-length DAF-12 and a *mir84p*-luciferase reporter vector and (2) *in vivo daf-9(dh6)* dauer rescue assays. (**c**) *daf-9;daf-12* metabolome fractions are inactive in the luciferase assay. 100 nM Δ^7-DA is used as a positive control. (**d**) Luciferase assays of *daf-22* metabolome fractions reveal three active regions (see **Appendix Figure C.1** for wild-type data). (**e**) *daf-9(dh6)* dauer rescue assays of *daf-22* metabolome fractions show activity in the same three regions (see **Appendix Figure C.1** for wild-type data). For worm images of scored phenotypes see **Appendix Figure C.1c**.

identification of very minor metabolites such as signaling molecules.[28,29] DANS relies on correlating genetic changes with metabolomic changes for compound identification and thereby reduces the need for extensive fractionation, which frequently results in activity loss or the introduction of artifacts.[29] The author envisioned that this approach could be applied to identify DAF-12-ligands if one compared a *C. elegans* mutant metabolome

lacking DAF-12-ligands with the metabolome of animals that produces DAF-12-ligands abundantly.

For DANS-based ligand identification, the author chose *daf-9;daf-12* double mutants as the ligand-deficient strain and *daf-22* mutants as a ligand-rich reference strain. *daf-9;daf-12* double mutants do not produce DAF-12-ligands, but nonetheless bypass the dauer stage, because lack of DAF-12 prevents the execution of genetic programs required for dauer formation[30] and therefore can be grown in quantities large enough for NMR spectroscopic analyses. *daf-22* mutant worms develop normally to adulthood, but are defective in the biosynthesis of the dauer-inducing ascarosides (**Chapter 2**),[31,32] which was hypothesized may adversely affect DAF-12-ligand production in wild-type liquid cultures (**Figure 4.1b**). Downregulation of DAF-12-ligand biosynthesis by ascarosides is also suggested by the finding that exposure to high concentrations of dauer pheromone abolished expression of DAF-9, one of the key enzymes in the proposed biosynthetic pathway of DAF-12-ligands.[15]

In preparation for DANS analysis, metabolome extracts of *daf-9;daf-12*, wild-type, and *daf-22* mixed-stage liquid cultures were fractionated using an automated, highly reproducible chromatography system (**Figure 4.2a** and **Appendix Sections C.1.4** and **C.1.5**). The resulting parallel sets of metabolome fractions were assessed for DAF-12-ligand content using *in vivo* and *in vitro* bioassays (**Figure 4.2b**). The *in-vivo* assay (**Appendix Section C.1.7**) used *daf-9(dh6)* worms, which are defective in DAF-12-ligand production. In the absence of exogenously added DAF-12-ligand or a suitable precursor, developing *daf-9(dh6)* worms arrest as dauer larvae, because DAF-12 constitutively interacts with its corepressor DIN-1 (**Figure 4.2b**).[12,21,30] The assay scored

the ability of added metabolome fractions to rescue the arrested dauer larvae and promote development to adulthood, providing a measure for the presence of a DAF-12 ligand or suitable precursor in the exogenously added fractions that would dissociate the DAF-12/DIN-1 corepressor complex. The *in vitro* assay (**Appendix Section C.1.8**) measured DAF-12 transcriptional activation of a luciferase reporter in HEK-293T cells that were co-transfected with full-length DAF-12 and the reporter construct.[33] This assay provided a measure for ligand-dependent interaction of DAF-12 with mammalian coactivator(s) endogenous to the cell line. Both of these assays consistently showed activity for three groups of *daf-22* and wild type metabolome fractions (regions I-III in **Figures 4.2c-e**; for wild-type assay data, see **Figure C.1**), suggesting the presence of DAF-12-ligands or precursors, whereas all *daf-9;daf-12* fractions were inactive in both assays, as expected based on previous work.[16,21,30] As anticipated, *daf-22* fractions were significantly more active in both assays than the corresponding wild-type fractions, suggesting much higher production of DAF-12-ligands in *daf-22* mutants (**Figure C.1**, *vide infra*).

4.3. DANS revealed steroid with unexpected Δ^1-desaturation as a novel ligand for DAF-12: dqfCOSY spectra of the most active group of *daf-22* fractions, active region I, revealed the presence of long-chained ascarosides,[28] in addition to a complex mixture of fatty acids, glycerides, other lipids, and several epidioxy sterol derivatives, (**Figures 4.3**, **4.4**, and **Appendix Figure C.2**),[34,35] all of which were also present in similar concentrations in corresponding *daf-9;daf-12* metabolome fractions. However, closer inspection of dqfCOSY spectra of region I revealed several sets of signals that were

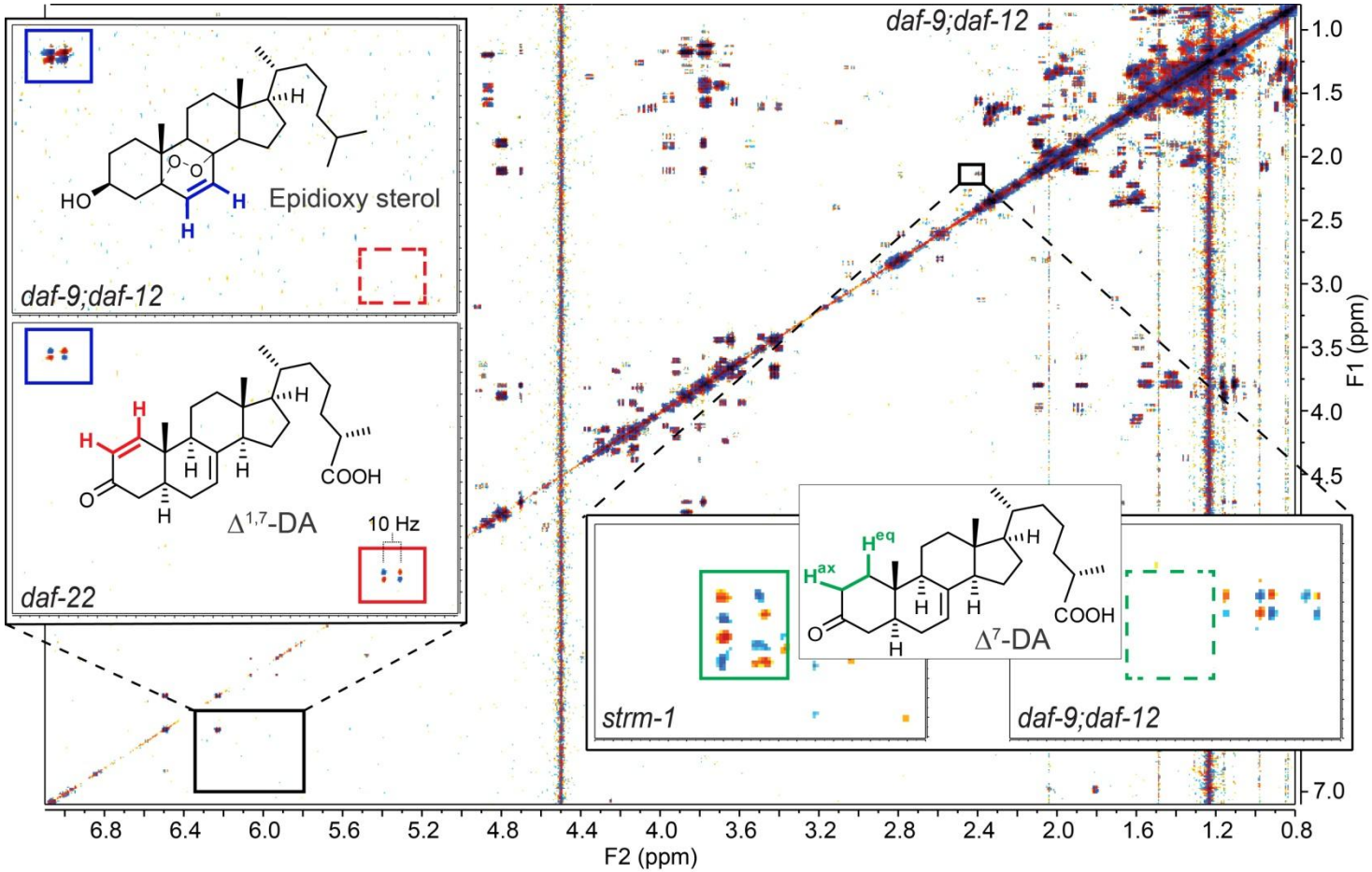


Figure 4.3. Detection of endogenous DAF-12-ligands via DANS. dqfCOSY spectrum of inactive *daf-9;daf-12* metabolome fraction corresponding to active region I (**Figures 4.2c-e**). Enlarged section (upper left) compares *daf-9;daf-12* with corresponding *daf-22* section, showing one of the differential crosspeaks (red) that led to identification of $\Delta^{1,7}$-DA, next to non-differential signals representing a metabolite present in both *daf-9;daf-12* and *daf-22*, an epidioxy sterol (blue). Enlarged section (lower right) shows example crosspeaks from the comparison of the *strm-1* and *daf-9;daf-12* metabolomes, showing signals (green) characteristic for Δ^{7}-DA in the *strm-1* spectrum but not the *daf-9;daf-12* spectrum.

consistently absent in *daf-9;daf-12*, and thus appeared to be *daf-9*-dependent (**Figure 4.3**). Further analysis of these differential signals strongly suggested steroidal structures (**Figure 4.4** and **Appendix Figure C.4**). Because of their extremely low concentrations, detailed characterization of the putative *daf-9*-dependent steroids required additional fractionation via HPLC (**Appendix Section C.1.5**), which resulted in two active samples each containing 1-2% of *daf-9*-dependent components (**Figure 4.4**, **Appendix Figures C.3** and **C.4**). NMR-spectroscopic analysis of the most active fractions showed a distinct set of *daf-9*-dependent signals at 5.9 and 7.0 ppm with a coupling constant of 10 Hz, which strongly suggested the presence of an unusual Δ^1-desaturated 3-keto steroid (**Figures 4.3** and **4.5a**).

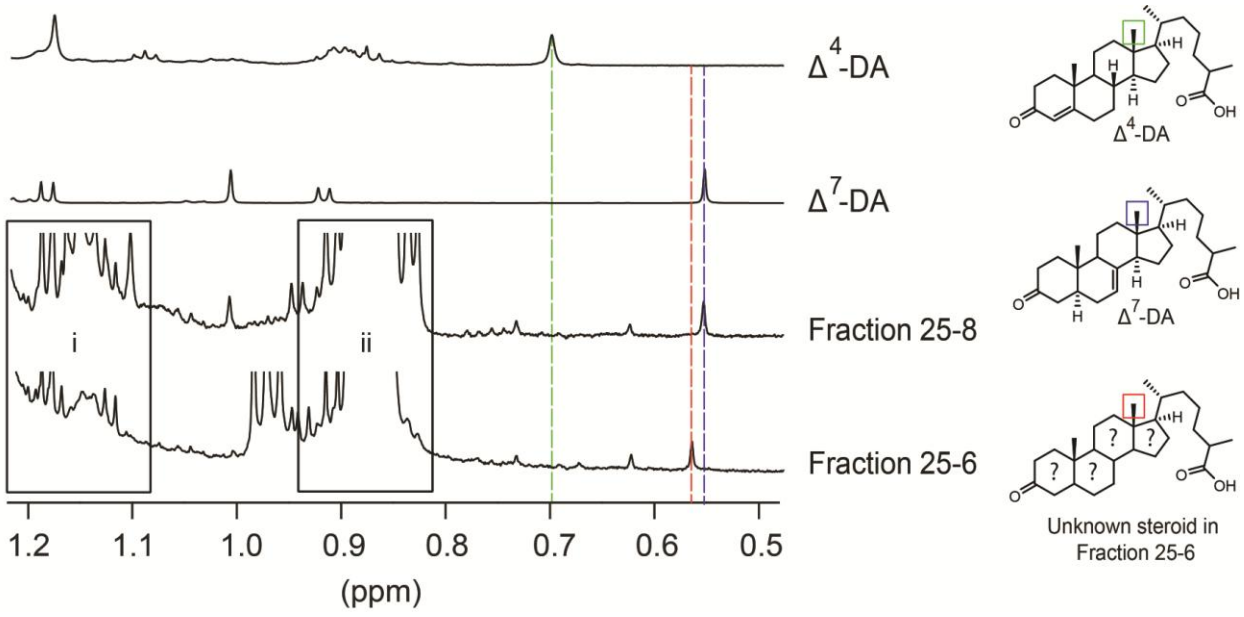


Figure 4.4: ^{1}H NMR Analysis of HPLC-enriched fractions. Comparison of representative sections of ^{1}H-NMR spectra of active HPLC-enriched fractions with the proposed DAF-12 ligands[36] revealed the presence of Δ^7-DA in fraction 25-8, and an unknown steroid ligand in fraction 25-6 that is distinctly different from the previously proposed Δ^7- and Δ^4-DA. ^{1}H-NMR spectra of the natural fractions also show that *C. elegans* derived lipids (i, ii) constitutes the majority of the active fractions.

Comparison with literature data suggested 3-oxocholesta-1,7-dienoic acid ("**$\Delta^{1,7}$-DA**"), which had not previously been described from worms. In addition, DANS comparison of *strm-1* mutants,[37] which had previously been shown to produce elevated levels of DAF-12-ligands revealed the known Δ^7-DA[16] (**Figures 4.3** and **4.5a**). The methyltransferase, STRM-1, regulates DAF-12-ligand levels by converting cholesterol-derived intermediates of ligand biosynthesis into 4-methylated steroids, thereby rendering them unsuitable as ligand precursors.[37] Δ^1-desaturated steroids are rare in nature, and the author is aware of only one other example in animals, the identification of (25*S*)-3-oxocholesta-1,4-dienoic acid from the Indonesian soft coral, *Minabea* sp.[38]

Further, the author in collaboration with Mr. Mahanti analyzed the active region II via similar comparative metabolomics, which led to the identification of an unusual (25*S*)-3α-hydroxy cholest-7-enoic acid ("3α-OH- Δ^7-DA") as a second novel ligand of DAF-12 (**Figure 4.5a**). Details of the structural characterization and biological activity of this compound will be included in the Ph.D dissertation of Mr. Mahanti. Active region III, was represented by a series of metabolome fractions with comparatively low *daf-9(dh6)* dauer rescue activity that were not characterized further. To corroborate the structural assignments of the novel steroidal ligands of DAF-12 and biological testing, robust total syntheses were developed by Mr. Joshua C. Judkins, Schroeder research group, for each of the proposed ligands. The details of the synthetic procedures and experimental conditions will be included in the Ph.D dissertation of Mr. Judkins, the manuscript in question (Judkins, et al.) is in preparation.

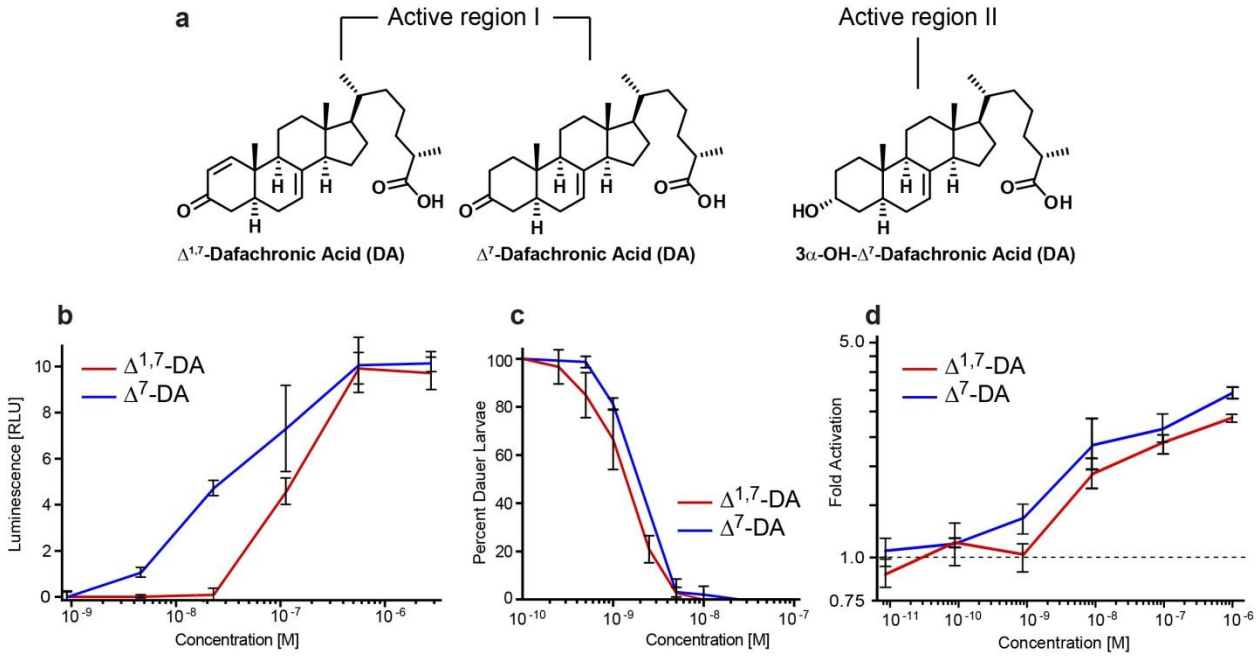


Figure 4.5: Structures and biological activity of endogenous DAF-12 ligands. (**a**) Structures of endogenous DAF-12 ligands in *C. elegans.* (**b**) Assessment of DAF-12 transcriptional activation in HEK-293T cells by the identified endogenous DAF-12 ligands from active region I. Luciferase assays were measured in triplicates, error bars as SD. (**c**) *daf-9(dh6)* dauer rescue with the identified endogenous DAF-12-ligands from active region I at 27 °C. For each data point there were two replicates with 100 animals per replicate, error bars as SD. (**d**) Alphascreen assay for ligand-dependent recruitment of SRC1-4 peptide by DAF-12, showing fold activation of DAF-12 with different ligand candidates over control (ethanol). Data for biological activities of 3α-OH- Δ^7-DA will be included in the Ph.D dissertation of Mr. Mahanti.

Next the author used synthetic samples of $\Delta^{1,7}$- and Δ^7-DA for testing in the *in vivo daf-9(dh6)* dauer rescue and *in vitro* luciferase assays developed for assessing ligand content of metabolome fractions (**Figure 4.2b**) . Results from the bioassays showed that the novel $\Delta^{1,7}$-DA activates DAF-12 in mammalian cells (EC_{50} = 146 nM) and that its potency in the *daf-9(dh6)* dauer rescue assay is similar (EC_{50} = 2 nM) or slightly higher than that of Δ^7-DA (**Figure 4.5b,c** and **Appendix Table C.1**). To test whether the $\Delta^{1,7}$-DA and Δ^7-DA (endogenous DAF-12 ligands from active region I) constitute *bona fide* ligands of DAF-12, Mr. Mahanti measured ligand dependant binding

of DAF-12 with the SRC1-4 peptide containing the nuclear receptor box ("NR box") motif of mammalian coactivator SRC-1 in a cell-free *in vitro* system.[16,39] Detailed procedures for this "Alphascreen assay" will be included in the Ph.D dissertation of Mr. Mahanti; a brief description is included in **Appendix Section C.1.9**. The results showed both Δ^7-DA and $\Delta^{1,7}$-DA effect concentration-dependent recruitment of SRC1-4, with EC_{50} values of 8 nM and 15 nM for Δ^7-DA and $\Delta^{1,7}$-DA, respectively (**Figure 4.5d** and **Appendix Table C.1**). These relative potencies of Δ^7-DA and $\Delta^{1,7}$-DA are similar to relative activities observed in the luciferase assay (**Figure 4.5b**).

4.4. SIM-GC/MS based structural confirmation of $\Delta^{1,7}$- and Δ^7-DAs and absence of the previously proposed Δ^4-DA in *C. elegans* metabolome: To confirm the structures of the endogenous steroid ligands of DAF-12 proposed using 2D NMR-based comparative metabolomics and for their quantification in *C. elegans*-derived fractions, the author developed a highly sensitive Selective Ion Monitoring (SIM)-GC/MS method (see **Appendix Sections C.1.10-C.1.12**). Briefly, for this method, the author first reacted known quantities of synthetic standards of $\Delta^{1,7}$- and Δ^7-DA or 1-10 % of active metabolome fractions with an excess of trimethylsilyldiazomethane to produce the corresponding DA-methyl esters (e.g., $\Delta^{1,7}$-DA-methyl ester, **Figure 4.6c**). The reaction mixtures were then analyzed by SIM-GC/MS where the retention times of characteristic mass spectrometric fragments of volatile DA-methyl ester derivatives' from active natural fractions in region I were matched with those of synthetic samples of $\Delta^{1,7}$- and Δ^7-DA methyl esters (**Figure 4.6a-c** and **Appendix Figure C.4e-i**).

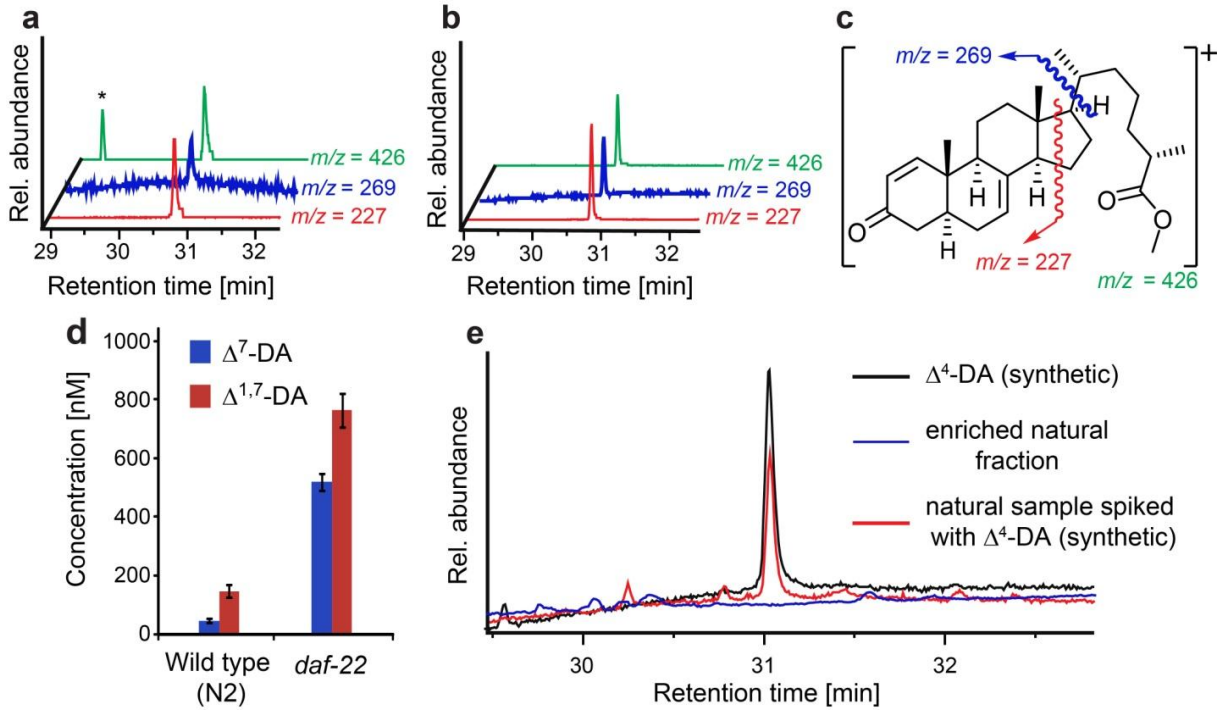


Figure 4.6: SIM-GC/MS-based structural confirmation and quantification of DAF-12 ligands from region I. (**a**) SIM-GC/MS of active *daf-22* metabolome fraction suggesting the presence of $\Delta^{1,7}$-DA. An additional peak (*) eluting at ~29.4 minutes in the ion trace *m/z* = 426 does not match the fragmentation profile consistent with dafachronic acids and hence was not pursued further. (**b**) SIM-GC/MS of synthetic $\Delta^{1,7}$-DA methyl ester confirm retention times and fragmentation patterns. (**c**) Major EI-MS fragments of $\Delta^{1,7}$-DA used in **Figure 4.6a,b**. (**d**) *In vivo* concentrations of Δ^{7}-DA and $\Delta^{1,7}$-DA derived from SIM-GC/MS-based quantification in wild-type (N2) and *daf-22* mutant (error bars, SD) grown in mixed stage liquid cultures. (**e**) SIM-GC/MS total ion chromatograms (TIC) of synthetic Δ^{4}-DA methyl ester, 1% injection of methylated inactive fraction 26 from *daf-22* to which trace amounts of synthetic Δ^{4}-DA (100 ng, ~20-fold less than Δ^{7}-, $\Delta^{1,7}$-DAs in active metabolome fractions) had been added, and 100% injection of a methylated and HPLC-enriched natural fraction matching the LC retention time of synthetic Δ^{4}-DA. The EI-MS fragmentation profile for Δ^{4}-DA methyl ester is reported in **Appendix Figure C.5**

Results from SIM-GC/MS unambiguously confirmed the structures of both $\Delta^{1,7}$- and Δ^{7}-DA as endogenous ligands of DAF-12 in *C. elegans* from region I (**Figure 4.6a**).

Quantification of DAs using SIM-GC/MS by the author (**Appendix Section C.1.12**) showed that $\Delta^{1,7}$-DA is slightly more abundant than Δ^{7}-DA in *daf-22* worms, whereas in wild-type animals $\Delta^{1,7}$-DA is more than twice as abundant as Δ^{7}-DA, with

concentrations (averaged over the worm bodies) of 40 nM and 130 nM for Δ^7-DA and $\Delta^{1,7}$-DA, respectively (**Figure 4.6d**). Based on the specific activities determined for synthetic samples of Δ^7-DA and $\Delta^{1,7}$-DA it appears that these two compounds can account for all of the activity in region I in both wild-type and *daf-22* metabolomes.

Using SIM-GC/MS, the author also checked the *daf-22* mutant and wild-type metabolomes for the presence of the previously reported Δ^4-DA. The author was unable to detect Δ^4-DA in any of the analyzed *C. elegans* metabolome samples, whereas metabolome fractions spiked with trace quantities of synthetic Δ^4-DA confirmed the sensitivity of the author's detection methods (**Figure 4.6e** and **Appendix Figure C.5**). The author then considered the possibility that the growth conditions used may have affected the production of the

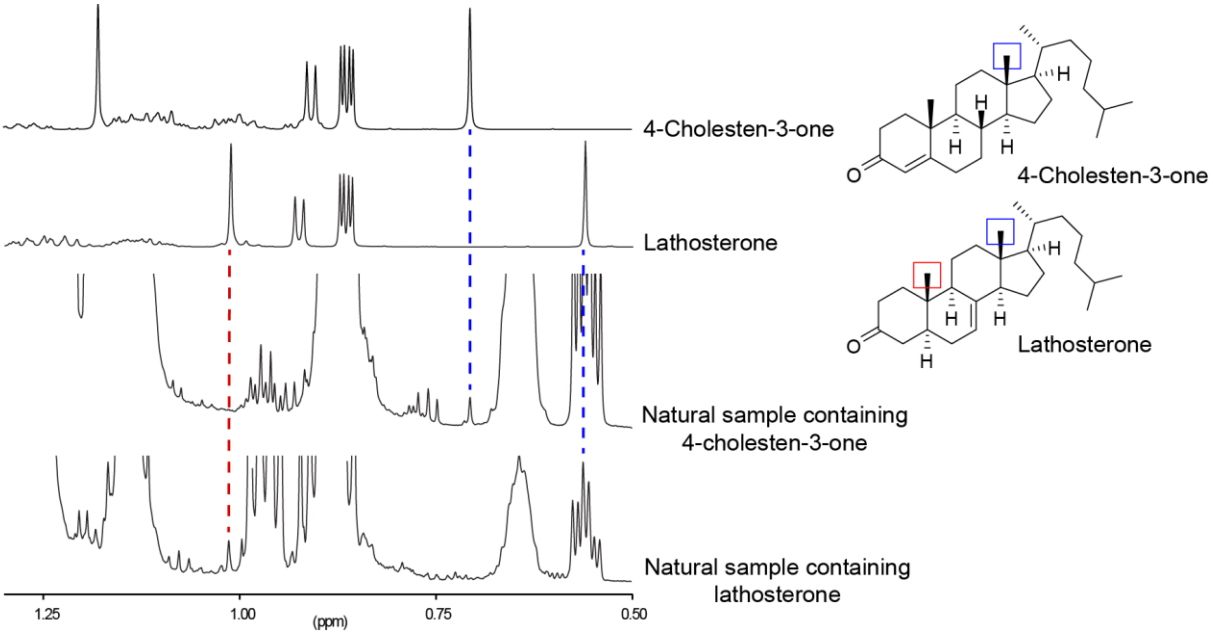


Figure 4.7: Identification of dafachronic acid precursors. Comparison of representative sections of ^{1}H-NMR spectra of enriched fractions from *daf-22* with synthetic 4-cholesten-3-one and lathosterone suggested the presence of the 3-keto-steroids in the natural sample.

putative precursor of Δ^4-DA, 4-cholesten-3-one.[16,25] However, analysis of wild-type metabolome samples by NMR spectroscopy (**Figure 4.7** and **Appendix Figure C.6**) and SIM-GC/MS (**Appendix Figure C.6**) revealed that 4-cholesten-3-one is as abundant as lathosterone, a putative precursor of Δ^7-DA,[16,40] suggesting that absence of Δ^4-DA is not the result of a lack of suitable precursors. Therefore, it appears that Δ^4-DA may not play a significant role as a DAF-12 ligand under the author's culture conditions (**Appendix Section C.1.2**), although it's transient or very low-level production cannot be excluded.

Taken together, these results indicate that $\Delta^{1,7}$-DA and Δ^7-DA are high-affinity ligands of DAF-12 that promote reproductive development and adult longevity and Δ^4-DA is not present at physiologically relevant concentrations. These findings indicate that previous hypotheses about DAF-12-ligand structures and their biosynthetic pathways must be revised.

4.5. Discussion: Identification of the endogenous ligands of NHRs is central to understanding their role as transcriptional regulators in metazoans. Screening of synthetic ligands of the vitamin D receptor and other mammalian NHRs has demonstrated that even small changes in ligand structures can strongly affect gene transcription.[2-4] Yet there are few approaches to comprehensively identify the endogenous NHR ligands from complex animal metabolomes, whose chemical annotations remain largely incomplete. In this dissertation chapter, the author demonstrates the use of comparative metabolomics to identify the endogenous ligands

of DAF-12, a central regulator of development and adult lifespan in *C. elegans* and an important model for ligand-regulated NHRs in higher animals. In contrast to earlier work that, primarily based on screening candidate structures, proposed Δ^4-DA and Δ^7-DA as endogenous DAF-12 ligands, this dissertation chapter revealed $\Delta^{1,7}$-DA as the most abundant ligand in wild-type worms, in addition to smaller amounts of Δ^7-DA and 3α-OH-Δ^7-DA.

Steroids featuring a Δ^1-double bond have been described from very few natural sources,[38] although it is well known that introduction of Δ^1-desaturation in natural 3-keto sterols, e.g. testosterone, can have pronounced effects on biological properties.[41] Analysis of X-ray structures of the DAF-12-ligand binding domain complexed with DAs have demonstrated that small structural changes in the A and B rings of the bound steroid have significant effects on the ligand's affinity to DAF-12,[42,43] suggesting that specific modifications in the steroid A-ring may serve to fine-tune DAF-12 transcriptional regulation. Identification of the enzyme(s) introducing the Δ^1-double bond will play an important role in elucidating functional differences between $\Delta^{1,7}$- and Δ^7-DA and may motivate re-analysis of mammalian metabolomes for the presence of endogenous Δ^1-steroids.

It should be noted that both the transcriptional activation assay in mammalian cell-culture (**Figure 4.5b**) and the Alphascreen assay (**Figure 4.5d**) have limited cogency for judging the relative potency of different DAF-12 ligands *in vivo*, as both assays depend on recruitment of mammalian coactivators such as SRC-1, whereas *in vivo* function of DAF-12 is thought to involve ligand-dependent dissociation of the

endogenous *C. elegans* corepressor DIN-1 followed by binding of yet unidentified co-activators.[12] The identification of multiple endogenous small molecule regulators of DAF-12 in this study will accelerate the pursuit of yet elusive DAF-12 interactors and other components of DAF-12-dependent dauer and lifespan regulation.

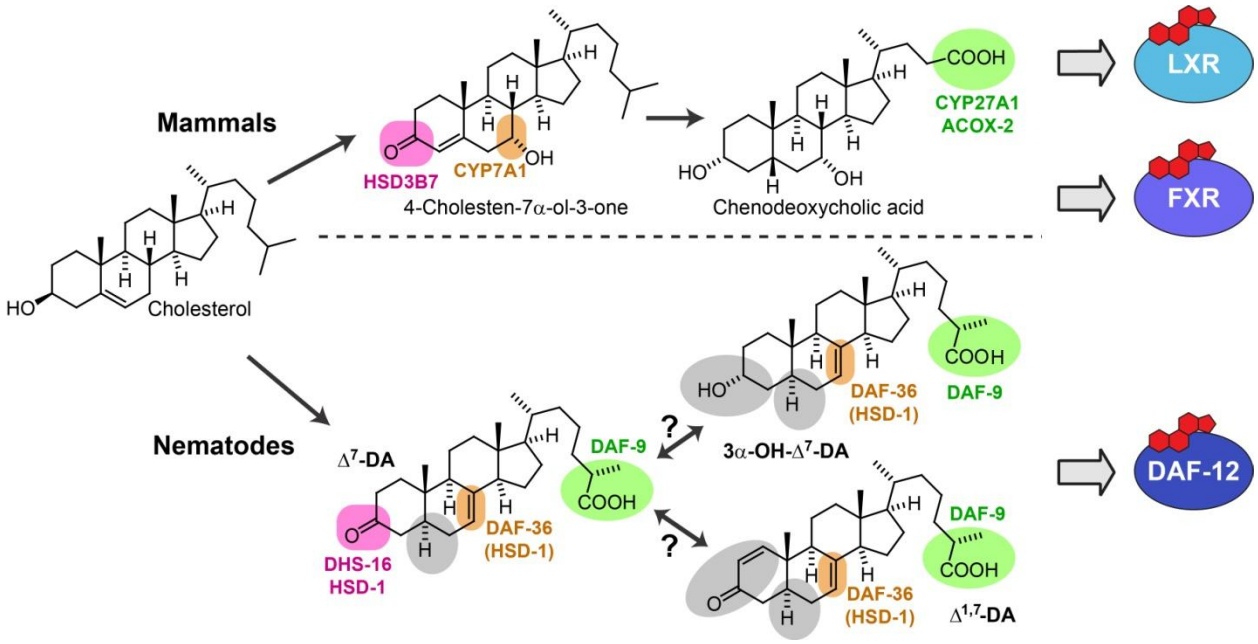


Figure 4.8: Comparison of NHR signaling in nematodes and mammals. In nematodes, oxidation/epimerization in position 3, oxidation in position 7, and side chain oxidation produces ligands of DAF-12. Recent work show that similar modifications of the steroid skeleton in mammals produce bile acids that might serve as ligands of liver X receptor[44] (LXR) and farnesoid X receptor[45] (FXR). Although DAF-9 has often been assumed to act directly upstream of DAF-12, our results suggest that DAF-9 may act on a variety of different substrates, including both 3-keto and 3-hydroxy sterols. A unique color has been associated for the known enzymes in the pathway and the corresponding transformations, whereas enzymes introducing the structural features highlighted in gray are not known.

Whereas it is well established that DAF-12-ligands are ultimately derived from dietary cholesterol, identification of $\Delta^{1,7}$-DA and the absence of Δ^{4}-DA necessitates revision of DAF-12-ligand biosynthesis models (**Figures 4.1b**). Subsequent studies (Mahanti and

Bose et al., *Cell Metabolism*, in revision) revealed DAF-36 as well as HSD-1 participate in the biosynthesis of Δ^7-DA and possibly $\Delta^{1,7}$-DA, whereas previously, HSD-1 had been assumed to function in Δ^4-DA biosynthesis[25] (**Figure 4.8**). In addition we found that mutation of *dhs-16* affects Δ^7-DA much more strongly than $\Delta^{1,7}$-DA production suggesting that, introduction of the 3-keto moiety in $\Delta^{1,7}$-DA may involve a different enzyme (Mahanti and Bose et al., *Cell Metabolism*, in revision). Therefore, it appears that different DAF-12-ligands are produced via partially divergent biosynthetic pathways, which is also supported by the subsequent finding that in *strm-1* mutants Δ^7-DA production is increased to a much greater extent than that of $\Delta^{1,7}$-DA (Mahanti and Bose et al., *Cell Metabolism*, in revision) .

Taken together, the identification of $\Delta^{1,7}$-DA and 3α-OH-Δ^7-DA, indicates that several DAF-12-ligand biosynthetic enzymes remain to be identified. Ultimately, elucidation of DAF-12-ligand biosynthetic pathways will require combining comparative metabolomics with tissue-specific manipulation of candidate genes. This will entail consideration of life-stage specific aspects of ligand functions and biosynthesis, especially with regard to the intriguing role of DAF-12-ligands for adult longevity, which depends on additional signaling from the germline.[46] DAF-12-ligand biosynthetic enzymes link conserved insulin/IGF and TGF-β signaling to transcriptional regulation by DAF-12.[40] Although the enzymes in DA biosynthesis are not strict orthologs of functionally corresponding enzymes in mammalian bile-acid biosynthesis, the striking similarities of steroidal NHR-ligand biosynthetic pathways in nematodes and humans demonstrate the utility of *C. elegans* as a model organism for endocrine signaling in metazoans (**Figure 4.8**). Further functional characterization of the identified DAF-12-

ligands will advance understanding of the roles of ligand-dependent NHRs in organism-wide coordination of metazoan development and aging.

In summary, this dissertation chapter shows that NMR-based comparative metabolomics provide detailed insight into metazoan small molecule signaling pathways, and that this approach can reveal signaling molecules and biosynthetic functions not suspected based on classical genetics and biochemical approaches. Finally, the novel ligands discovered here will yield important insights to combat nematode parasitism, which uses the DAF-12/DA signaling mechanism to regulate emergence from the infective stage.[42,47]

4.6. Author's note: The work presented in this dissertation chapter was done in collaboration with Mr. Parag Mahanti. Synthetic sample of $\Delta^{1,7}$- and Δ^{7}-DA were provided by Mr. Joshua C. Judkins. Mr. Judkins also provided enriched natural samples of 4-cholesten-3-one and lathosterone, putative precursors of Δ^{4}- and Δ^{7}-DAs for analysis using NMR and SIM-GC/MS. Mr. Mahanti and Mr. Judkins are both graduate students in the Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY.

REFERENCES

(1) Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schutz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. *Cell* **1995**, *83*, 835.

(2) Wollam, J.; Antebi, A. *Annual review of biochemistry* **2011**, *80*, 885.

(3) Brown, A. J.; Slatopolsky, E. *Mol Aspects Med* **2008**, *29*, 433.

(4) Singarapu, K. K.; Zhu, J.; Tonelli, M.; Rao, H.; Assadi-Porter, F. M.; Westler, W. M.; DeLuca, H. F.; Markley, J. L. *Biochemistry* **2011**, *50*, 11025.

(5) Schupp, M.; Lazar, M. A. *J Biol Chem* **2010**, *285*, 40409.

(6) Taubert, S.; Ward, J. D.; Yamamoto, K. R. *Mol Cell Endocrinol* **2010**, *334*, 49.

(7) Hulme, S. E.; Whitesides, G. M. *Angew Chem Int Ed Engl* **2011**, *50*, 4774.

(8) Antebi, A.; Yeh, W. H.; Tait, D.; Hedgecock, E. M.; Riddle, D. L. *Genes & development* **2000**, *14*, 1512.

(9) Fielenbach, N.; Antebi, A. *Genes & development* **2008**, *22*, 2149.

(10) Kenyon, C. *Annals of the New York Academy of Sciences* **2010**, *1204*, 156.

(11) Shen, Y.; Wollam, J.; Magner, D.; Karalay, O.; Antebi, A. *Science* **2012**, *338*, 1472.

(12) Ludewig, A. H.; Kober-Eisermann, C.; Weitzel, C.; Bethke, A.; Neubert, K.; Gerisch, B.; Hutter, H.; Antebi, A. *Genes & development* **2004**, *18*, 2120.

(13) Larsen, P. L.; Albert, P. S.; Riddle, D. L. *Genetics* **1995**, *139*, 1567.

(14) Gems, D.; Sutton, A. J.; Sundermeyer, M. L.; Albert, P. S.; King, K. V.; Edgley, M. L.; Larsen, P. L.; Riddle, D. L. *Genetics* **1998**, *150*, 129.

(15) Schaedel, O. N.; Gerisch, B.; Antebi, A.; Sternberg, P. W. *Plos Biol* **2012**, *10*, e1001306.

(16) Motola, D. L.; Cummins, C. L.; Rottiers, V.; Sharma, K. K.; Li, T.; Li, Y.; Suino-Powell, K.; Xu, H. E.; Auchus, R. J.; Antebi, A.; Mangelsdorf, D. J. *Cell* **2006**, *124*, 1209.

(17) Gerisch, B.; Weitzel, C.; Kober-Eisermann, C.; Rottiers, V.; Antebi, A. *Dev Cell* **2001**, *1*, 841.

(18) Jia, K.; Albert, P. S.; Riddle, D. L. *Development* **2002**, *129*, 221.

(19) Held, J. M.; White, M. P.; Fisher, A. L.; Gibson, B. W.; Lithgow, G. J.; Gill, M. S. *Aging Cell* **2006**, *5*, 283.

(20) Rottiers, V.; Motola, D. L.; Gerisch, B.; Cummins, C. L.; Nishiwaki, K.; Mangelsdorf, D. J.; Antebi, A. *Dev Cell* **2006**, *10*, 473.

(21) Gerisch, B.; Rottiers, V.; Li, D.; Motola, D. L.; Cummins, C. L.; Lehrach, H.; Mangelsdorf, D. J.; Antebi, A. *Proc Natl Acad Sci U S A* **2007**, *104*, 5014.

(22) Dumas, K. J.; Guo, C.; Wang, X.; Burkhart, K. B.; Adams, E. J.; Alam, H.; Hu, P. J. *Developmental biology* **2010**, *340*, 605.

(23) Wollam, J.; Magomedova, L.; Magner, D. B.; Shen, Y.; Rottiers, V.; Motola, D. L.; Mangelsdorf, D. J.; Cummins, C. L.; Antebi, A. *Aging Cell* **2011**, *10*, 879.

(24) Yoshiyama-Yanagawa, T.; Enya, S.; Shimada-Niwa, Y.; Yaguchi, S.; Haramoto, Y.; Matsuya, T.; Shiomi, K.; Sasakura, Y.; Takahashi, S.; Asashima, M.; Kataoka, H.; Niwa, R. *J Biol Chem* **2011**, *286*, 25756.

(25) Patel, D. S.; Fang, L. L.; Svy, D. K.; Ruvkun, G.; Li, W. *Development* **2008**, *135*, 2239.

(26) Williams, T. W.; Dumas, K. J.; Hu, P. J. *Aging (Albany NY)* **2010**, *2*, 742.

(27) Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc Chem Res* **2011**.

(28) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A* **2009**, *106*, 7708.

(29) Forseth, R. R.; Schroeder, F. C. *Curr Opin Chem Biol* **2011**, *15*, 38.

(30) Gerisch, B.; Antebi, A. *Development* **2004**, *131*, 1765.

(31) Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y. *Proc Natl Acad Sci U S A* **2009**, *106*, 1875.

(32) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc* **2012**, *134*, 1817.

(33) Bethke, A.; Fielenbach, N.; Wang, Z.; Mangelsdorf, D. J.; Antebi, A. *Science* **2009**, *324*, 95.

(34) Gunatilaka, A. A. L.; Gopichand, Y.; Schmitz, F. J.; Djerassi, C. *J Org Chem* **1981**, *46*, 3860.

(35) Sera, Y.; Adachi, K.; Shizuri, Y. *J Nat Prod* **1999**, *62*, 152.

(36) Motola, D. L.; Cummins, C. L.; Rottiers, V.; Sharma, K. K.; Li, T. T.; Li, Y.; Suino-Powell, K.; Xu, H. E.; Auchus, R. J.; Antebi, A.; Mangelsdorf, D. J. *Cell* **2006**, *124*, 1209.

(37) Hannich, J. T.; Entchev, E. V.; Mende, F.; Boytchev, H.; Martin, R.; Zagoriy, V.; Theumer, G.; Riezman, I.; Riezman, H.; Knolker, H. J.; Kurzchalia, T. V. *Dev Cell* **2009**, *16*, 833.

(38) Wang, W.; Lee, J. S.; Nakazawa, T.; Ukai, K.; Mangindaan, R. E.; Wewengkang, D. S.; Rotinsulu, H.; Kobayashi, H.; Tsukamoto, S.; Namikoshi, M. *Steroids* **2009**, *74*, 758.

(39) Fukuyama, M.; Rougvie, A. E.; Rothman, J. H. *Curr Biol* **2006**, *16*, 773.

(40) Wollam, J.; Magner, D. B.; Magomedova, L.; Rass, E.; Shen, Y.; Rottiers, V.; Habermann, B.; Cummins, C. L.; Antebi, A. *Plos Biol* **2012**, *10*, e1001305.

(41) Counsell, R. E.; Klimstra, P. D. *J Med Pharm Chem* **1962**, *91*, 477.

(42) Wang, Z.; Zhou, X. E.; Motola, D. L.; Gao, X.; Suino-Powell, K.; Conneely, A.; Ogata, C.; Sharma, K. K.; Auchus, R. J.; Lok, J. B.; Hawdon, J. M.; Kliewer, S. A.; Xu, H. E.; Mangelsdorf, D. J. *Proc Natl Acad Sci U S A* **2009**, *106*, 9138.

(43) Zhi, X.; Zhou, X. E.; Melcher, K.; Motola, D. L.; Gelmedin, V.; Hawdon, J.; Kliewer, S. A.; Mangelsdorf, D. J.; Xu, H. E. *J Biol Chem* **2012**, *287*, 4894.

(44) Theofilopoulos, S.; Wang, Y.; Kitambi, S. S.; Sacchetti, P.; Sousa, K. M.; Bodin, K.; Kirk, J.; Salto, C.; Gustafsson, M.; Toledo, E. M.; Karu, K.; Gustafsson, J. A.; Steffensen, K. R.; Ernfors, P.; Sjovall, J.; Griffiths, W. J.; Arenas, E. *Nature chemical biology* **2013**, *9*, 126.

(45) Russell, D. W. *Annual review of biochemistry* **2003**, *72*, 137.

(46) Yamawaki, T. M.; Berman, J. R.; Suchanek-Kavipurapu, M.; McCormick, M.; Gaglia, M. M.; Lee, S. J.; Kenyon, C. *Plos Biol* **2010**, *8*.

(47) Ogawa, A.; Streit, A.; Antebi, A.; Sommer, R. J. *Curr Biol* **2009**, *19*, 67.

CHAPTER 5

CONCLUSIONS AND OUTLOOK

5.1. Global metabolomics to investigate nematode and other metazoan metabolomes: In this dissertation, the author has utilized an innovative analytical workflow integrating aspects of 2D NMR, high resolution HPLC-MS/MS; activity guided fractionation, and chemical synthesis to achieve global metabolite profiling of complex nematode-derived samples. Application of this methodology for the investigation of *Caenorhabditis elegans* wild-type and peroxisomal β-oxidation mutant metabolomes (**Chapter 2**) revealed several novel ascarosides,[1] a class of signaling molecules that regulate different aspects of *C. elegans* biology. This functional diversity of ascarosides is paralleled by corresponding structural complexity featuring several unusual structural elements, such as ω-oxygenation of the fatty-acid-derived side chains, 4-hydroxybenzoylation or (*E*)-2-methyl-2-butenoylation of the ascarylose unit, and glucosyl esters. The data presented allowed the author to propose a model for ascaroside biogenesis that is tightly regulated. In this model a modular library of ascarosides are produced integrating components of primary metabolism, e.g. peroxisomal β-oxidation of fatty acids, carbohydrate-, and amino acid-metabolism.

Given the structural complexity and specificity of biogenesis exhibited by the library of ascarosides, it appears that many of the newly identified compounds also

contribute to known or as yet undetermined functions in *C. elegans*. Given the large number of compounds identified, comprehensive functional characterization of the ascarosides will require design of high-throughput assays, for example based on microfluidic devices.[2]

The structural complexity exhibited by the library of ascarosides in *C. elegans* and the unique mode of ascaroside biogenesis suggested that nematodes, and perhaps other animals, harbor yet unrecognized biosynthetic capabilities. To test this idea, the author embarked on global metabolomics of several nematode species including the necromenic roundworm *Pristionchus pacificus* (**Chapter 3**). 2D NMR and high resolution HPLC-MS/MS analyses revealed that *P. pacificus* produces small molecule architectures of unanticipated complexity that regulate the worm's developmental decisions and phenotypic plasticity.[3] The inclusion of *P. pacificus'* small molecules considerably expands the nematode-derived modular library of signaling molecules. This library of compounds is generated via integrating building blocks derived from all conserved primary metabolic pathways, including lipid-, amino acid-, carbohydrate-, and nucleoside-metabolism. Unprecedented structural features in this library include xylopyranose- (as opposed to ribose-) derived nucleosides and moieties derived from neurotransmitter metabolism and TCA cycle. Further 2D NMR and HPLC-MS/MS-based analysis revealed pasa#9 as a minor component of the *P. pacificus* metabolome. This compound is a complex derivative of the reported pasc#9[3] including an additional anthranilate module (**Figure 5.1**). The endogenous function of pasc#9 has not yet been elucidated, suggesting specific inputs from the incorporation of the anthranilate in pasa#9 may serve important yet to be known biochemical functions. The identification of

pasa#9 further extends the scope of modularity in nematode small molecule assembly. Such striking examples for combinatorial generation of nematode small molecules unfold a new paradigm of small molecule assembly in metazoans.

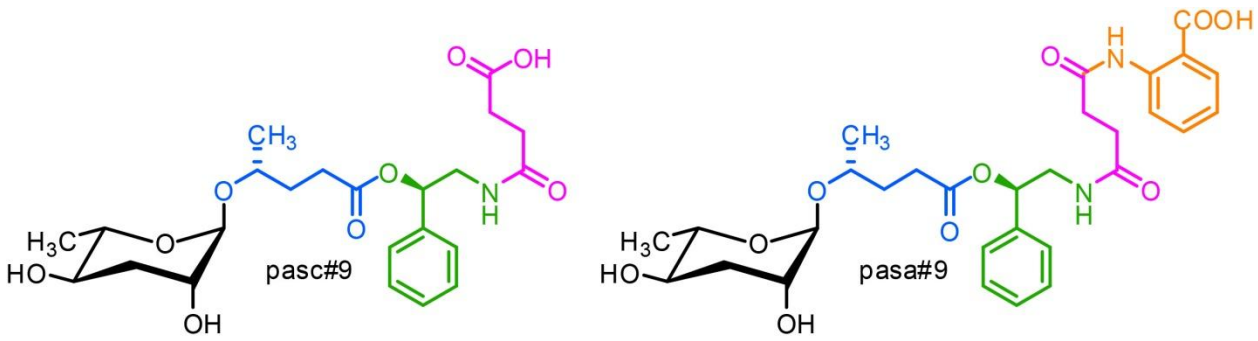


Figure 5.1: Structures of pasc#9 (confirmed) and pasa#9 (putative) derived from assembly of building blocks from carbohydrate- (black), fatty acid- (blue), amino acid- (green), and as well as TCA cycle-derived succinate (magenta) and anthranilate (chrome yellow). pasc#9 is a major component of the *P. pacificus* metabolome, whereas pasa#9 is a very minor component. The endogenous function of pasc#9 has not yet been determined.

Nematodes are arguably the most numerous animals on earth with regard to both numbers of species and individuals. Only ~3% of an estimated 1 million different nematode species have been isolated and described.[4] In-depth analysis of the metabolome of just two model species resulted in hundreds of novel small molecules displaying structural complexities unexpected from fully differentiated multi-cellular metazoans. Traditionally, structural characterization and biological testing of natural products isolated from plants and microbes have been used extensively for designing small molecule-based therapeutics.[5] However, results presented in this dissertation and the diversity of the phylum suggest that nematodes may be a veritable source for complex natural products. Thus, a concerted effort is called for to identify additional nematode-derived small molecules using global metabolomics and subsequent testing in disease models, which may lead to advanced small molecule therapies.

The recognition of molecular patterns or signals associated with specific pathogens (or food sources) is fundamental to ecology and plays a major role in the evolution of symbiotic and predator-prey relationships.[6] Recent studies suggest that ascarosides are not only limited to the model nematodes *C. elegans* and *P. pacificus*, but are widely distributed in the nematode phylum, including both human- and plant-parasitic nematodes.[7] Hence, ascarosides (and perhaps other nematode-derived small molecules such as dafachronic acids) may represent an evolutionarily conserved molecular signature of nematodes that may be perceived by a wide range of organisms that nematodes interact with, including plant and animal hosts as well as microorganisms associated with nematodes. This idea is supported by recent studies that demonstrated that nematophagous fungi perceive ascarosides. In nematophagous fungi,[8] which are natural predators of soil-dwelling nematodes, ascarosides trigger trap formation, an essential morphological change that enables predatory behavior.[9] These results indicate that nematode-derived small molecules may be developed into therapeutics in humans, livestock, and plants to combat nematode parasitism. This provides additional incentives to expand the library of nematode-derived small molecules using global metabolomics on a wider selection of parasitic and free-living nematodes.

The identification of the nematode small molecule-signals reported in this dissertation demonstrates underappreciated biosynthetic capabilities in metazoans. The generation of this library may be dependent on dedicated nematode-specific biosynthetic genes. However, it is important to note that the library is derived from

specific modular assembly of primary metabolites available to nearly all organisms. Also, known signaling molecules and co-factors in higher animals, for example, *S*-adenosyl methionine, phosphatidylinositols, etc. rely on modular assembly of components from multiple conserved primary metabolic pathways. Thus metazoans may extend analogous strategies to produce signaling molecules of much greater structural complexity. This ultimately requires fresh spectroscopic re-analysis of metabolomes from higher animals, including mammals, which may also reveal novel small molecule signals.

5.2. Harnessing comparative metabolomics and genomics, a unique route for the study of small molecule biogenesis: Nematode-derived small molecules are involved in the regulation of several aspects of *C. elegans* and *P. pacificus* biology; hence elucidation of the biosynthesis of molecules reported in this dissertation will be essential for the field.[10-13] *P. pacificus* is a powerful system for using Next-Generation Sequencing-based population genomics and genome-wide association studies (GWAS). The genome of 104 natural isolates has already been fully analyzed; another set of 154 complete genome sequences has just been obtained and is currently in the bioinformatics pipeline.[14] Together, the large collection of *P. pacificus* isolates and the observed genetic diversity make this species an excellent candidate for studying the biosynthesis of modular small molecules via comparative genomics. Proof-of-principle evidence for natural variation in small molecule production is provided in **Chapter 3.** In this study the author used HPLC-MS-based analysis of the metabolomes of 6 *P.*

pacificus wild strains. 2 out these 6 strains (RS5205 and RSB020) were deficient in ubas#1 and ubas#2 biosynthesis (at least under our experimental conditions).

Extending this idea of natural variation, the author conducted additional preliminary studies aiming to analyze the metabolomes of 27 different *P. pacificus* wild isolates and profile the recently identified[3] small molecules. All strains were grown in triplicate, and the exo-metabolome analyzed by HPLC-MS. As before (results for 6 strains analyzed in **Chapter 3**) the author found that each strain is represented by a specific small molecule profile, but most strains produce all of the previously identified compounds. Notably, two strains did not produce the dimeric ascaroside dasc#1, RS5419 and RS5412 (**Figure 5.2a**). In addition six strains were defective in production of ubas#1 and ubas#2 (**Figure 5.2b**). These data are deemed preliminary by the author since the study was based on small 10 mL cultures and the current analytical technique used to profile small-molecules have limited sensitivity and dynamic range. Future research at groups of Prof. Frank C. Schroeder and Prof. Ralf J. Sommer will be directed towards scaling up of the deficient strains as well as analysis via more sensitive MS detection methods such as Selective Ion Monitoring (SIM)-LC/MS to rule out low level production of the apparent deficient compounds.

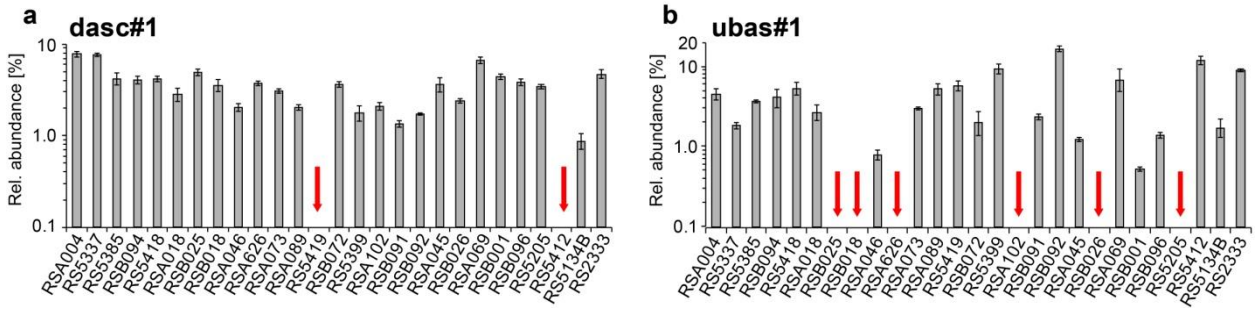


Figure 5.2: Results from preliminary HPLC/MS screen of 27 *P. pacificus* strains. (**a**) Two strains do not produce dascr#1 and (**b**) eight strains are defective in ubas#1 (ubas#2 not shown).

The ubas- and dasc-defective strains are not genetically clustered and originate from different geographical locations, which is advantageous for bioinformatic comparison of dasc- and ubas-producing strains with non-producers. Preliminary bioinformatic analysis based on comparison of the two dasc#1-defective strains with genome data from the dasc#1-producing strains by Dr. Gabriel Markov, Sommer research group, MPI, Tuebingen, Germany identified several hundred single nucleotide polymorphisms (SNPs), which is small compared to the total number of SNP's between strains (> 50,000). Incorporation of data from additional dasc#1-producing and non-producing strains will rapidly decrease this number further and produce a small number

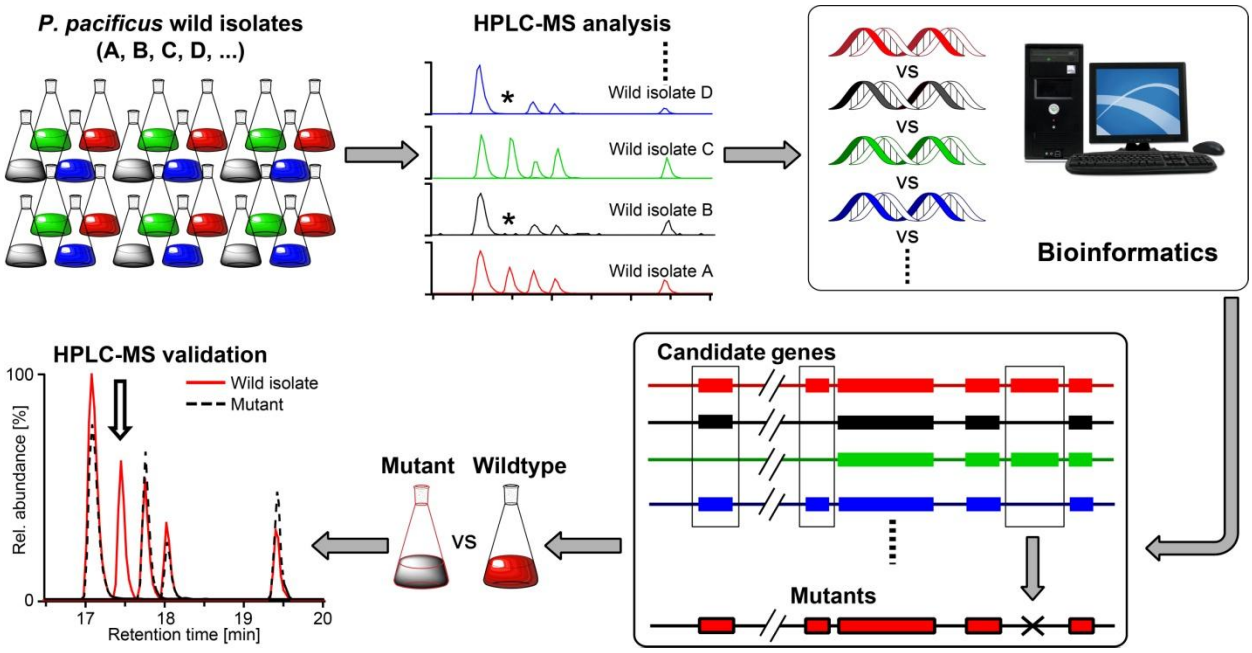


Figure 5.3: Harnessing comparative metabolomics and genomics. A high-throughput workflow for culturing, metabolite extraction, and targeted HPLC-MS analysis for 360 sequenced *P. pacificus* wild isolates. Subsequent bioinformatics and mutant design will identify and establish the biochemical roles of genes that regulate small molecule biogenesis and export. This strategy may uniquely correlate the genomes to the metabolomes of worm strains from around the world.

of protein-coding sequences from which candidates will be selected. The author is optimistic about this approach, since analysis of only 27 strains (about 10% of the total number of genotyped strains) 8 were found to have significant defects.

Based on these preliminary observations, the author has designed (for future research in the Schroeder research group) a high-throughput workflow for culturing, metabolite extraction, and targeted HPLC-MS analysis for the remaining ~300 sequenced *P. pacificus* wild isolates represented in **Figure 5.3**. Subsequent inputs from bioinformatics will provide candidate genes that influence small molecule biogenesis, the biochemical role of which will be confirmed via mutant design and comparative HPLC-MS analysis of the mutant and wild-type metabolomes (**Figure 5.3**). This strategy will uniquely correlate the genomes to the metabolomes of worm strains from around the world and provide a novel route for investigating small molecule biogenesis.

5.3. Comparative metabolomics to provide a blueprint for the identification of ligands for other *C. elegans* and mammalian NHRs: Small-molecule ligands of nuclear hormone receptors (NHRs) govern the transcriptional regulation of metazoan development, cell differentiation and metabolism. Even though the proper structural characterization of endogenous ligands of NHRs is central to understanding their role as transcriptional regulators in metazoans, there are only few approaches to comprehensively identify the endogenous NHR ligands from complex animal metabolomes. The incomplete chemical annotation of metazoan metabolomes is one of the major reasons for several "orphan" (or ligand-less) NHRs in model systems such as

C. elegans as well as in higher organisms. In **Chapters 2** and **3** the author described the development of a global metabolomics workflow that the author is optimistic will significantly impact proper chemical annotation of complex metazoan metabolomes.

Synthetic ligand screen for the vitamin D receptor and other mammalian NHRs has demonstrated that even small changes in ligand structures can strongly affect gene transcription, further bolstering the need for the identification of endogenous NHR ligands.[15-17] Using 2D NMR-based comparative metabolomics, the author identified the endogenous ligands of the *C. elegans* NHR, DAF-12, a vitamin D and liver X receptor homolog that regulates larval development, fat metabolism and lifespan (**Chapter 4**). Studies by the author demonstrated the advantages of comparative metabolomics over traditional candidate-based approaches and provide a blueprint for the identification of ligands for other *C. elegans* and mammalian NHRs.

Interestingly, a steroid with Δ^1 –desaturation was found to be an endogenous ligand for DAF-12. Steroids featuring a Δ^1-double bond have rarely been described in nature,[18] although it has been shown that introduction of Δ^1-unsaturation in natural 3-keto sterols, e.g. testosterone, greatly enhances their biological properties.[19] The enzyme that incorporates this critical Δ^1 –desaturation is not known, identification of which in *C. elegans* may motivate characterization of homologous genes in mammals and may initiate a spectroscopic re-analysis of mammalian metabolomes for the presence of endogenous Δ^1-steroids.

Ultimately, DAF-12-ligand biosynthetic enzymes link evolutionarily conserved insulin/IGF and TGF-β signaling to transcriptional regulation by DAF-12 (**Chapter 1**,

Figure 1.4 and **Chapter 4**, **Figure 4.1**).[20] Structural and functional characterization of small molecule signals in nematodes described in this dissertation provides additional routes to investigate these conserved pathways that are implicated in many grave disease such as: cancer, diabetes, and atherosclerosis.[21-23]. Further functional characterization of the identified DAF-12-ligands will advance understanding of the roles of ligand-dependent NHRs in organism-wide coordination of metazoan development and aging.

REFERENCES

(1) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc.* **2012**, *134*, 1817.

(2) Chung, K.; Zhan, M.; Srinivasan, J.; Sternberg, P. W.; Gong, E.; Schroeder, F. C.; Lu, H. *Lab Chip* **2011**, *11*, 3689.

(3) Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C. *Angew Chem Int Ed Engl* **2012**, *51*, 12438.

(4) Hugot, J. P.; Baujard, P.; Morand, S. *Nematology* **2002**, *3*, 199.

(5) Newman, D. J.; Cragg, G. M. *J Nat Prod* **2012**, *75*, 311.

(6) Boller, T.; Felix, G. *Annual review of plant biology* **2009**, *60*, 379.

(7) Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W. *Curr Biol* **2012**, *22*, 772.

(8) Thorn, R. G.; Moncalvo, J.-M.; Reddy, C. A.; Vilgalys, R. *Mycologia* **2000**, *92*, 241.

(9) Hsueh, Y. P.; Mahanti, P.; Schroeder, F. C.; Sternberg, P. W. *Curr Biol* **2012**.

(10) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol* **2012**, *10*, e1001237.

(11) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc* **2012**, *134*, 1817.

(12) Choe, A.; von Reuss, S. H.; Kogan, D.; Gasser, R. B.; Platzer, E. G.; Schroeder, F. C.; Sternberg, P. W. *Curr Biol* **2012**.

(13) Sommer, R. J.; Ogawa, A. *Curr Biol* **2011**, *21*, R758.

(14) Roedelsperger et al., submitted, personal communication with Prof. Ralf J. Sommer

(15) Brown, A. J.; Slatopolsky, E. *Mol Aspects Med* **2008**, *29*, 433.

(16) Wollam, J.; Antebi, A. *Annual review of biochemistry* **2011**, *80*, 885.

(17) Singarapu, K. K.; Zhu, J.; Tonelli, M.; Rao, H.; Assadi-Porter, F. M.; Westler, W. M.; DeLuca, H. F.; Markley, J. L. *Biochemistry* **2011**, *50*, 11025.

(18) Wang, W.; Lee, J. S.; Nakazawa, T.; Ukai, K.; Mangindaan, R. E.; Wewengkang, D. S.; Rotinsulu, H.; Kobayashi, H.; Tsukamoto, S.; Namikoshi, M. *Steroids* **2009**, *74*, 758.

(19) Counsell, R. E.; Klimstra, P. D. *J Med Pharm Chem* **1962**, *91*, 477.

(20) Wollam, J.; Magner, D. B.; Magomedova, L.; Rass, E.; Shen, Y.; Rottiers, V.; Habermann, B.; Cummins, C. L.; Antebi, A. *PLoS Biol* **2012**, *10*, e1001305.

(21) Gordon, K. J.; Blobe, G. C. *Biochimica et biophysica acta* **2008**, *1782*, 197.

(22) Blobe, G. C.; Schiemann, W. P.; Lodish, H. F. *The New England journal of medicine* **2000**, *342*, 1350.

(23) Moses, A. C. *Endocrine development* **2005**, *9*, 121.

APPENDIX A

COMPARATIVE METABOLOMICS REVEALS BIOGENESIS OF ASCAROSIDES, A MODULAR LIBRARY OF SMALL MOLECULE SIGNALS IN *C. ELEGANS*

A.1. Materials and Methods:

A.1.1. Analytical instrumentation: NMR spectra were recorded on Varian INOVA 600 (600 MHz for ^{1}H, 151 MHz for ^{13}C), INOVA 500 (500 MHz for ^{1}H, 125 MHz for ^{13}C), or INOVA 400 (400 MHz for ^{1}H, 100 MHz for ^{13}C) spectrometers. NMR spectra were processed using Varian VNMR, MestreLabs MestReC and MestRecNova software packages.

Low-resolution HPLC–MS and HPLC–MS/MS was performed using an Agilent 1100 Series HPLC system equipped with a diode array detector and connected to a Quattro II mass spectrometer (Micromass/Waters). High resolution MS/MS was performed using an LTQ Orbitrap Velos Hybrid FT mass spectrometer (Thermo Scientific, Cornell University Life Sciences Core Laboratories Center). High resolution HPLC-MS was performed using a Waters nanoACQUITY UPLC System equipped with a Waters Acquity UPLC HSS C-18 column (2.1 x 100 mm, 1.8 µm particle diameter) connected to a Xevo G2 QTof Mass Spectrometer. MassLynx software was used for MS data acquisition and processing.

HPLC was performed using the Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB-C18 column (9.4 x 250 mm, 5 µm particle diameter) coupled to a Teledyne ISCO Foxy 200 fraction collector.

A.1.2. *C. elegans* strains and general culture methods: *C. elegans* variety Bristol, strain N2 (wild type), *acox-1*(*ok2257*), *daf-22*(*m130*), *daf-22*(*ok693*), F58F9.7(*tm4033*), C48B4.1(*ok2619*), F59F4.1(*ok2119*), and F08A8.3(*tm5192*) were maintained at 20 ^{o}C on NGM agar plates, made with Bacto agar (BD Biosciences) and seeded with *E. coli* OP50 grown overnight.

A.1.3. Preparation of metabolite extracts: Wild-type (N2, Bristol) or *acox-1(ok2257)*, and *daf-22(ok693)* mutant worms from four 10 cm NGM plates were washed using M9-medium into a 100 mL S-medium pre-culture where they were grown for four days at 22 ^{o}C on a rotary shaker at 220 rpm. Concentrated *E. coli* OP50 derived from 1 L of bacterial culture was added as food at days 1 and 3. Subsequently, each pre-culture was divided equally into four 500 mL Erlenmeyer flasks containing 100 mL of S-medium on day 4. These cultures were grown for 5 days at 22 ^{o}C on a rotary shaker and fed with concentrated OP50 derived from 500 mL of bacterial culture every day from day 1 to day 4. On day 5 they were harvested, centrifuged, and the resultant supernatant media and worm pellets were frozen over dry ice-acetone slush and lyophilized separately. The lyophilized materials from the supernatant were extracted with 150 mL of 95% ethanol at room temperature for 16 h. The worm pellets were crushed with ~2 g of granular

NaCl using a mortar pestle and extracted with 100 mL of 100% ethanol at room temperature for 16 h. The resulting suspensions were filtered and the filtrate evaporated *in vacuo* at room temperature, producing media metabolite (the worm "exo-metabolome") extracts and worm pellet metabolite ("endo-metabolomes") extracts.

A.1.4. Ascaroside feeding experiment with *daf-22(m130)*: Ascaroside feeding experiments were performed with the *daf-22(m130)* mutant, which is less sensitive to growth defects due to added ascarosides than the *daf-22(ok693)* mutant (Schroeder, unpublished results). HPLC-MS analysis of *daf-22(m130)* showed similar ascaroside profiles as *daf-22(ok693)*, notably a total lack of short chain ascarosides with chain length less than 12 carbons. Cultures of *daf-22 (m130)* were grown as described before with the addition of 5 µM ascr#3 or a 1:1 mixture of 2.5 µM of each of ascr#10 and oscr#10 per culture on day 1 after pre-culture splitting.

A.1.5. HPLC-MS/MS Sample preparation: Media or worm pellet metabolite extracts were resuspended in ~15 mL methanol, centrifuged and the supernatant collected. The supernatant were then concentrated *in vacuo* at room temperature and resuspended in 1 mL methanol, centrifuged, and 30 µL of this extract was directly injected for LC-MS/MS analysis.

A.1.6. HPLC-MS/MS analysis: A 0.1% acetic acid – acetonitrile solvent gradient was used at a flow rate of 3.6 ml/min, starting with an acetonitrile content of 5% for 5 min

which was increased to 100% over a period of 40 min. Metabolite extracts were analyzed by HPLC-ESI-MS in negative ion mode using a capillary voltage of 3.5 kV and a cone voltage of -40 V. HPLC-MS/MS screening for precursor ions of *m/z* = 73.0 was performed using argon as collision gas at 2.1 mtorr and 30 eV. Ascaroside fragmentation was further analyzed by high-resolution MS/MS using the LTQ Orbitrap. To confirm elemental composition of new compounds, wild-type and mutant metabolome samples and fractions were additionally analyzed by high-resolution HPLC-MS using the Xevo G2 QTof.

A.1.7. Identification and quantification of ascarosides: For the identification of putative ascarosides detected in *C. elegans* wild-type and mutants[1] (ascr, oscr, icas, icos, and glas; see **Figures 2.4** and **Appendix Tables A.1-A.6** for bhas, bhos, ibha, and ibho see **Appendix Figure A.2** and Ref. 1), HPLC-retention times were plotted versus *m/z* (or chain length). Components belonging to a homologous series exhibited almost linear elution profiles (**Appendix Figure A.1**), indicating that components within a series share the same relative stereochemistry. The structure and stereochemistry of the various series were then identified based on (1) isolation of representative examples and 2D NMR analysis (for examples see **Appendix Figures A.3**, **A.4**, **A.6**, and **A.9**), (2) comparison of representative examples with synthetic standards provided by Dr. Stephan H. von Reuss, (3) molecular formula as established from high-resolution MS (**Tables A.1-A.6**), (4) characteristic MS/MS fragmentation, and (5) HPLC-retention times that matched retention time values extrapolated from those of the synthetic samples. The (*E*)-configuration of α,β-unsaturated ascarosides was established by comparison

with ascr#3, (*Z*)-configured ascr#3, and ascr#7, and is also suggested by the stereoselectivity of acyl-CoA-oxidase (ACOX) activity.

Quantification of ascarosides was performed by integration of HPLC-MS signals from the corresponding ion-traces. Ascaroside concentrations were calculated using response factors determined for synthetic standards of ascr#1, #3, #5, #7, #9, #10, oscr#9, #10, icas#3, #9, icos#10, and glas#10 (provided by Dr. Stephan von Reuss and reported in Ref. 1). For most compounds, mass spectrometer response was roughly linear (less than 10% error) for amounts of 1 pmol to 2 nmol per injection. Response factors for ascarosides that were not synthesized were extrapolated based on data observed for the available standards. Generally, strong differences between the response factors of short-chained members of each series (side-chains less than C_7) were observed, whereas differences between response factors of longer-chained homologs were small. Since not all short-chained members of all series were synthesized, the systematic errors of the absolute amounts reported for some short-chain ascarosides could be larger than for longer-chained compounds.

In order to roughly account for culture duration and worm biomass, ascaroside content of the excretome and worm pellet samples are reported in fmol ascarosides produced per hour of culture time per mg of worm pellet dry weight. All quantitative data reported in the Figures were derived from at least two independent biological repeats.

A.1.8. Statistical analysis: Unpaired student's t-tests with Welch's correction were used for comparing ascaroside profiles between wild-type and mutant metabolomes.

A.2. Figures:

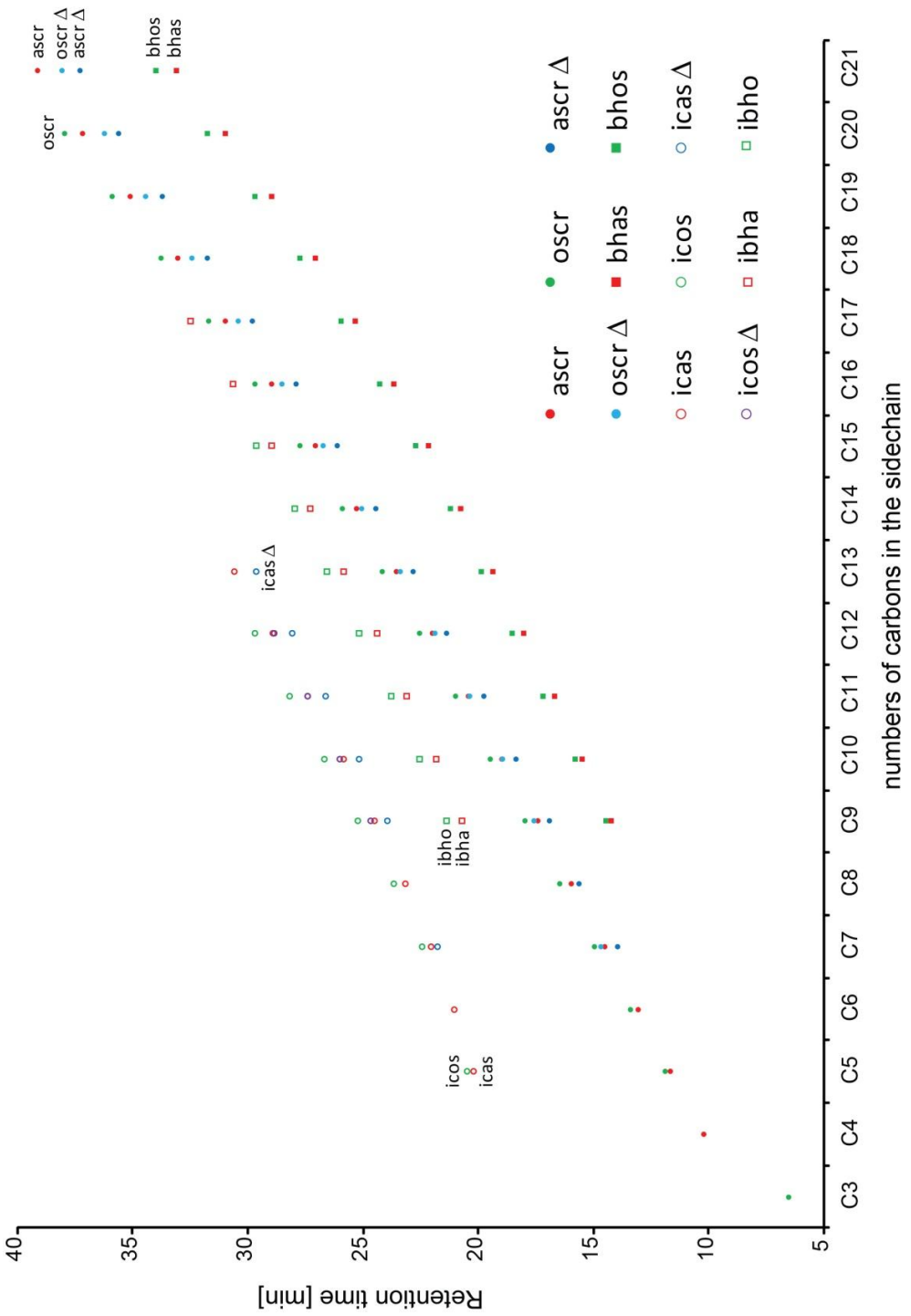


Figure A.1: HPLC elution profiles of ascarosides identified in wild-type and mutant excretome extracts of *C. elegans* (Δ indicates components with (*E*)-configured α,β-unsaturated sidechains). Data represents collaborative efforts of the author of this dissertation and Dr. Stephan H. von Reuss and reported in Ref. 1.

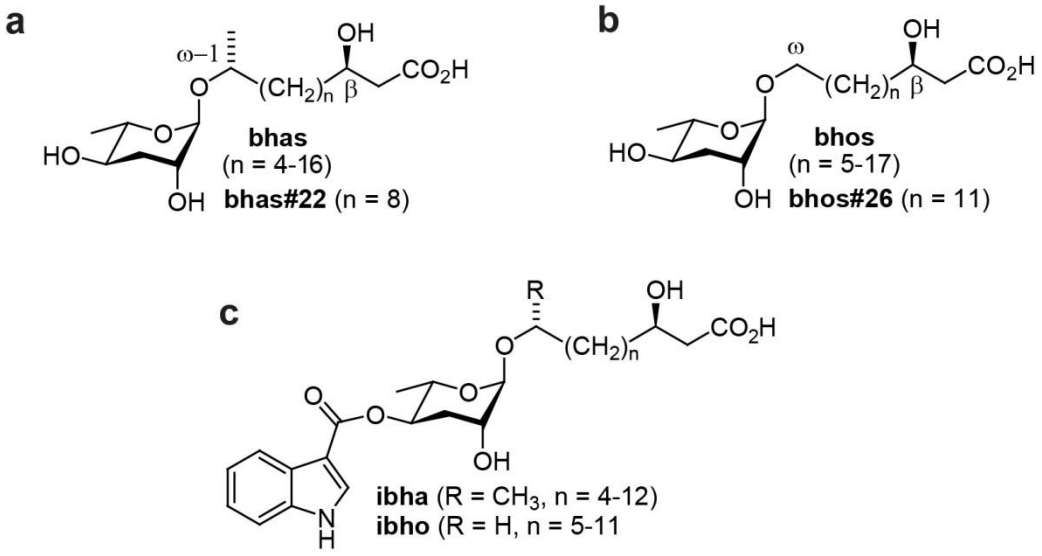


Figure A.2: Representative β-hydroxyascaroside classes[1] identified in wild-type, *acox-1*, *maoc-1*, *dhs-28*, and *daf-22* worms via HPLC-MS/MS: (**a**) (ω−1)-oxygenated β-hydroxyascarosides, (**b**) ω-oxygenated β-hydroxyascarosides, and (**c**) examples for 4'-indole-3-carbonylated β-hydroxyascarosides. See also **Figure 2.4**.

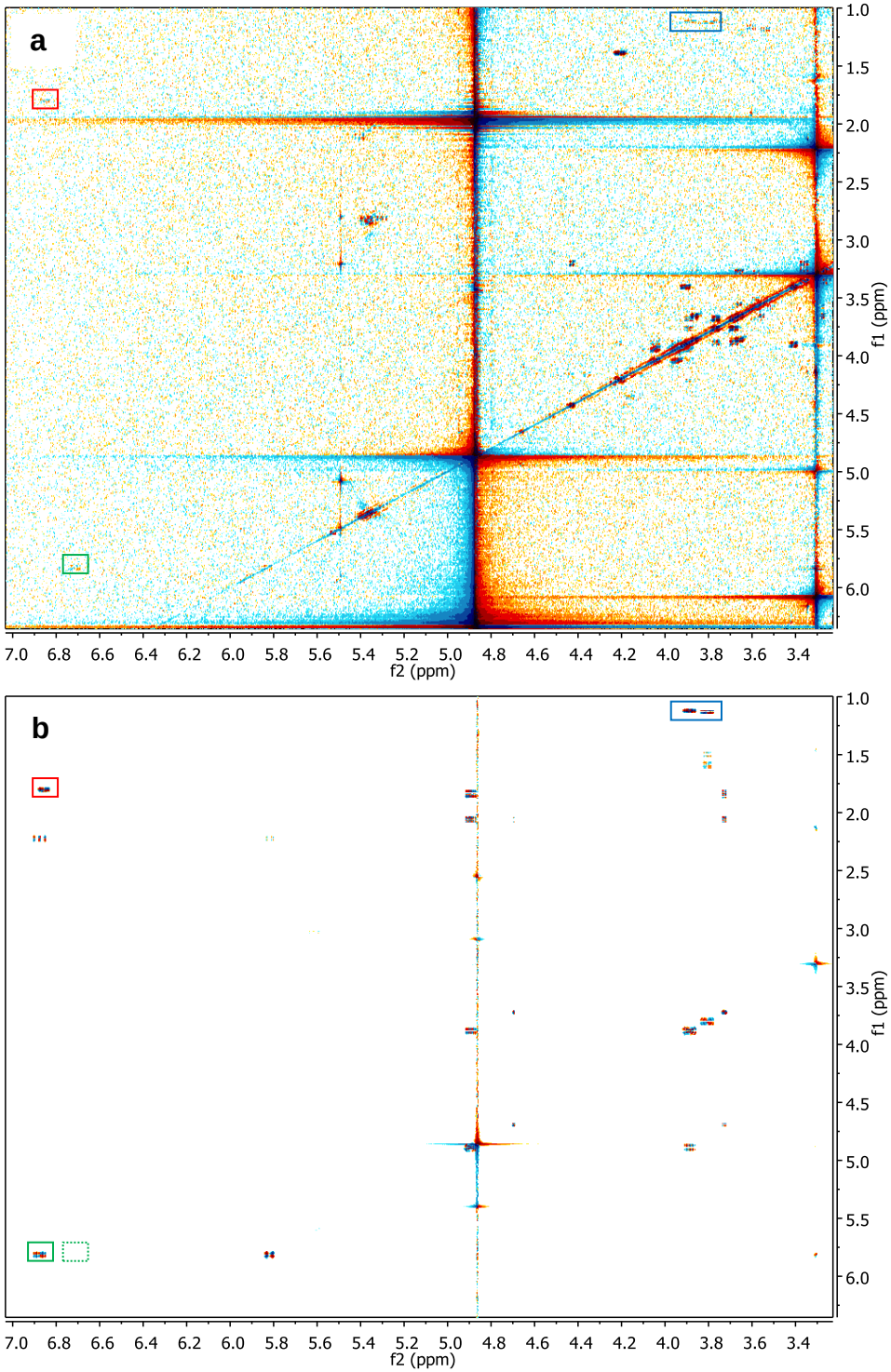


Figure A.3: Sections of dqfCOSY spectra (600 MHz, methanol-d_4) of mbas#3-enriched fraction from wild-type *C .elegans* media extracts (**a**) and synthetic mbas#3 (**b**) showing characteristic signals for methyl groups of the ascarylose ring and the side chain (blue), the allylic methyl group of the tiglate unit (red), and the pH dependant signal for the side chain double bond (green).

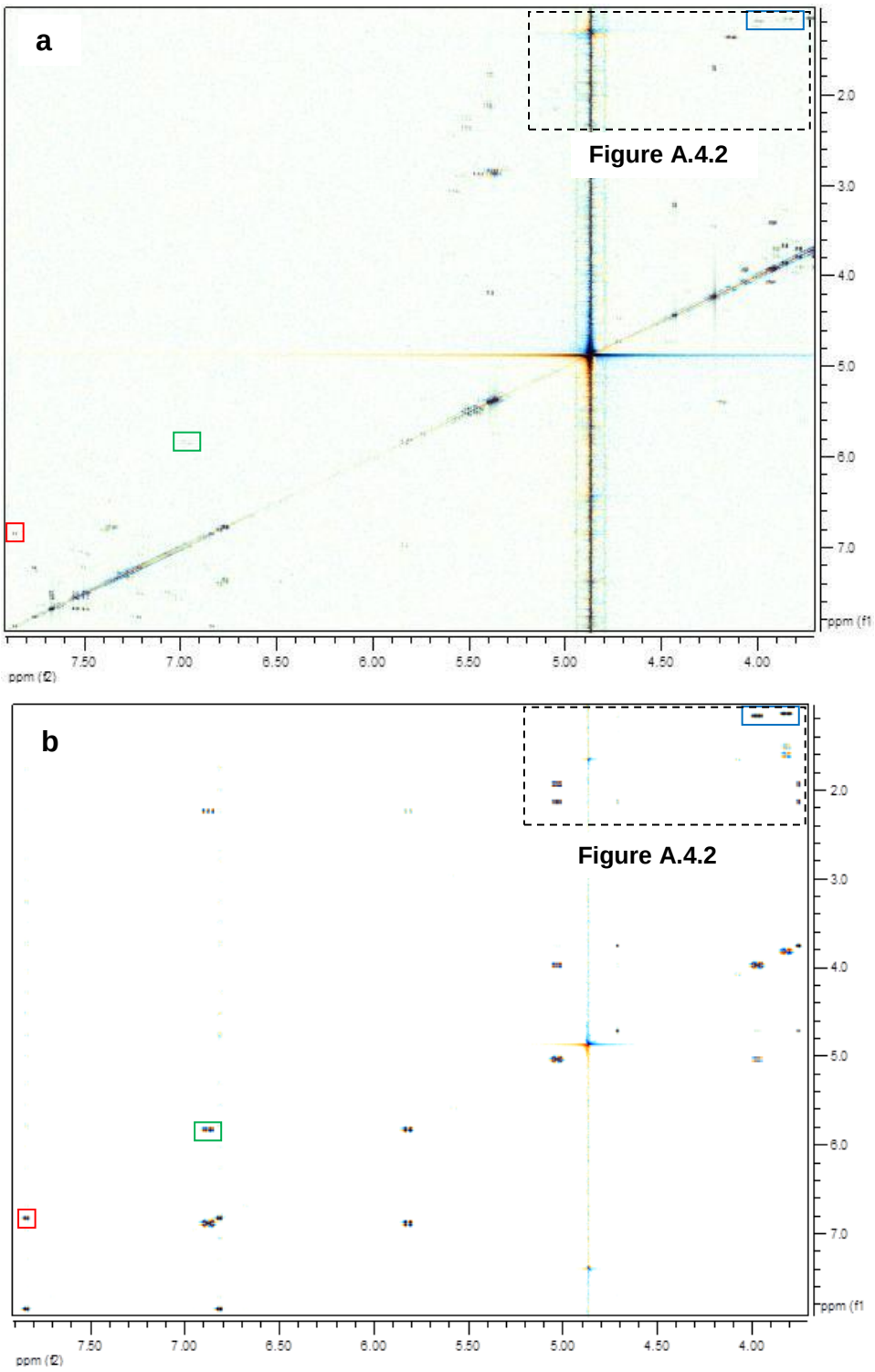


Figure A.4.1. Sections of dqfCOSY spectra (600 MHz, methanol-d_4) of hbas#3-enriched fraction from wild-type *C .elegans* media extracts (**a**) and synthetic hbas#3 (**b**) showing characteristic signals for methyl groups of the ascarylose ring and the side chain (blue), the *para*-substituted 4-hydroxybenzoyl unit (red), and the side chain double bond (green).

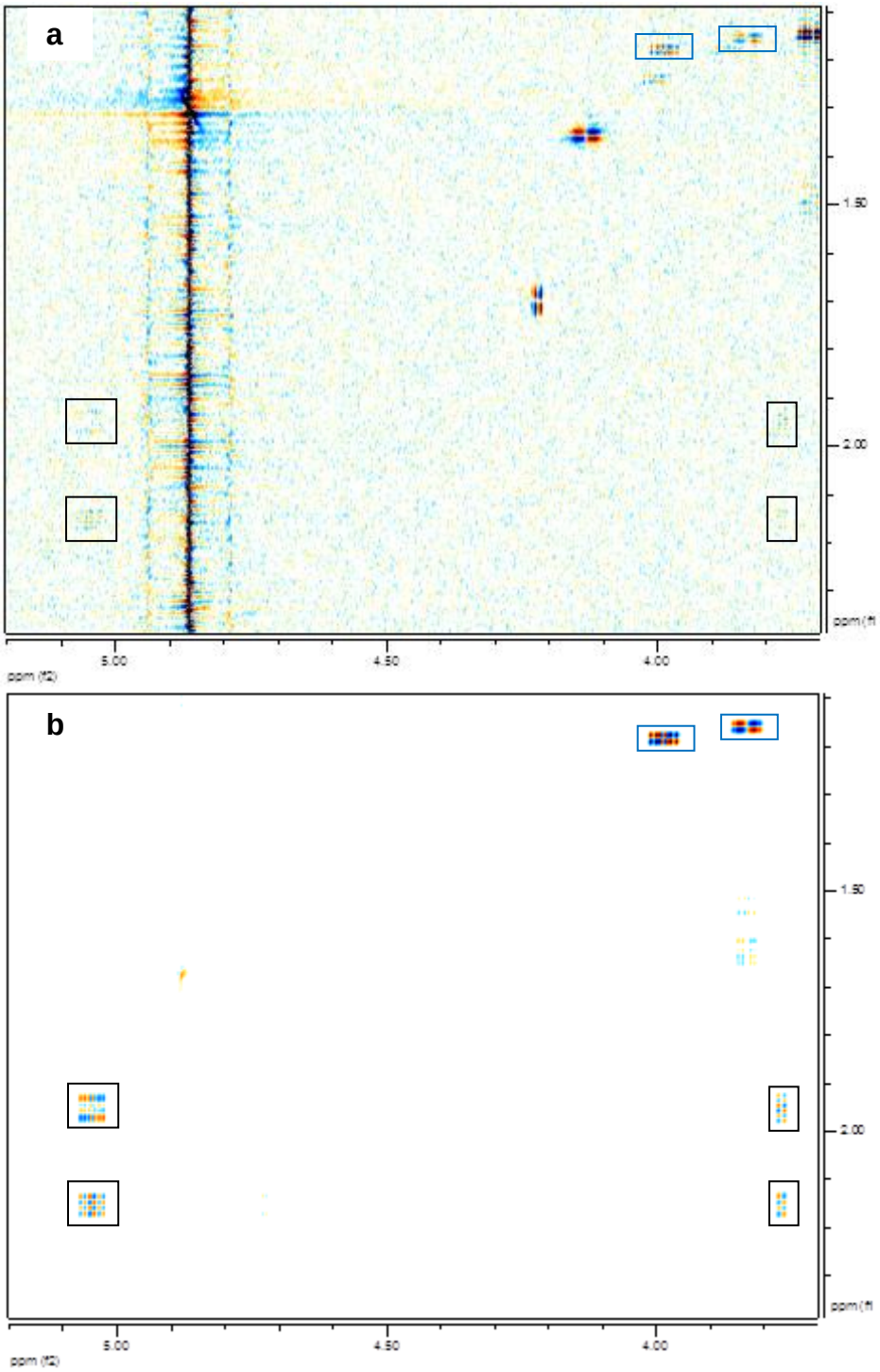


Figure A.4.2: Enlarged sections of dqfCOSY spectra (600 MHz, methanol-d_4) of hbas#3-enriched fraction from wild-type *C .elegans* media extracts (**a**) and synthetic hbas#3 (**b**) showing characteristic signals for methyl groups of the ascarylose ring (blue), and the ascarylose spin system (black).

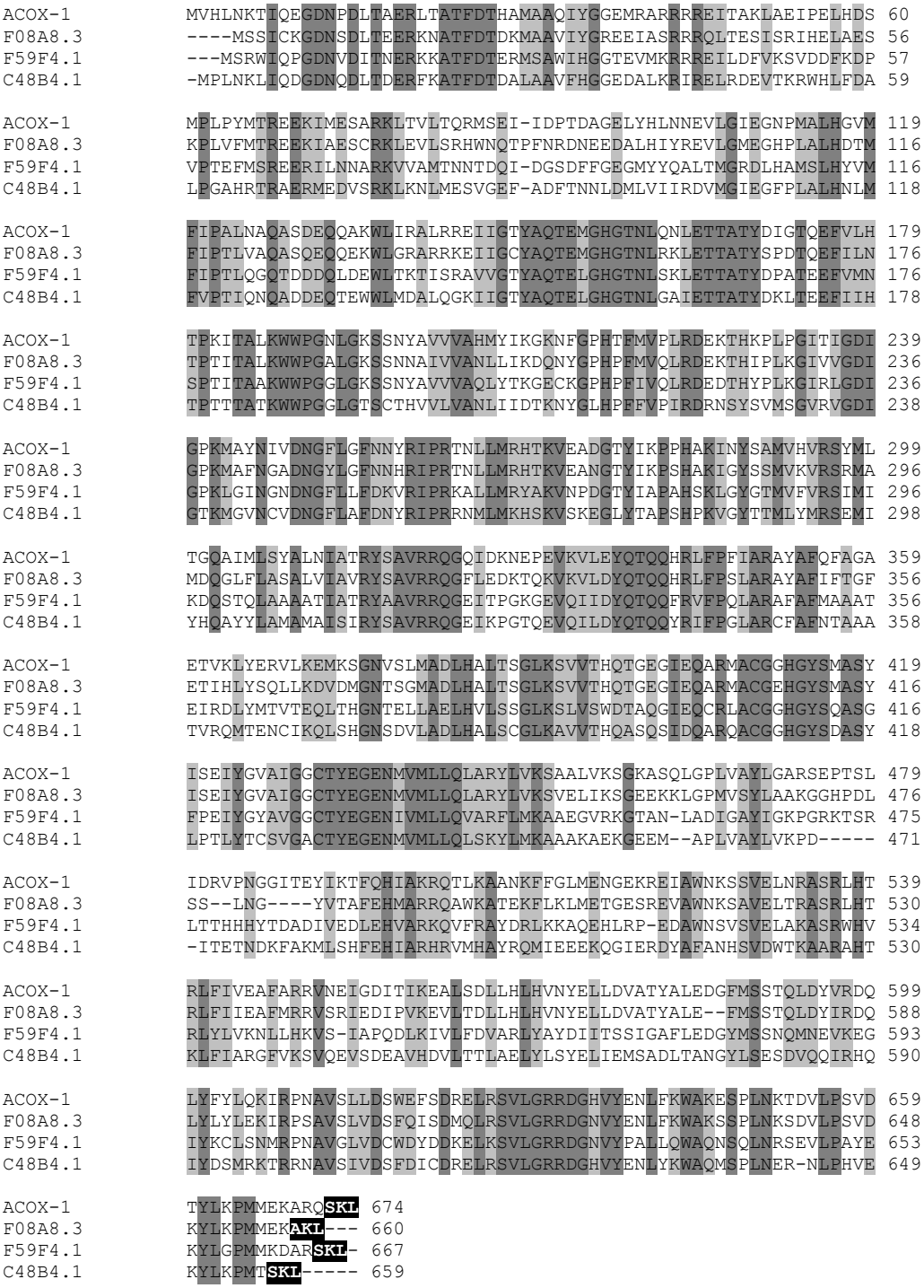


Figure A.5: Alignment of *C. elegans'* ACOX-1 isoform a.1 with other peroxisomal acyl-CoA oxidases was performed using ClustalW. Identical amino acids are marked in grey, similar amino acids are marked in light grey, and the peroxisomal targeting signal is marked in black.

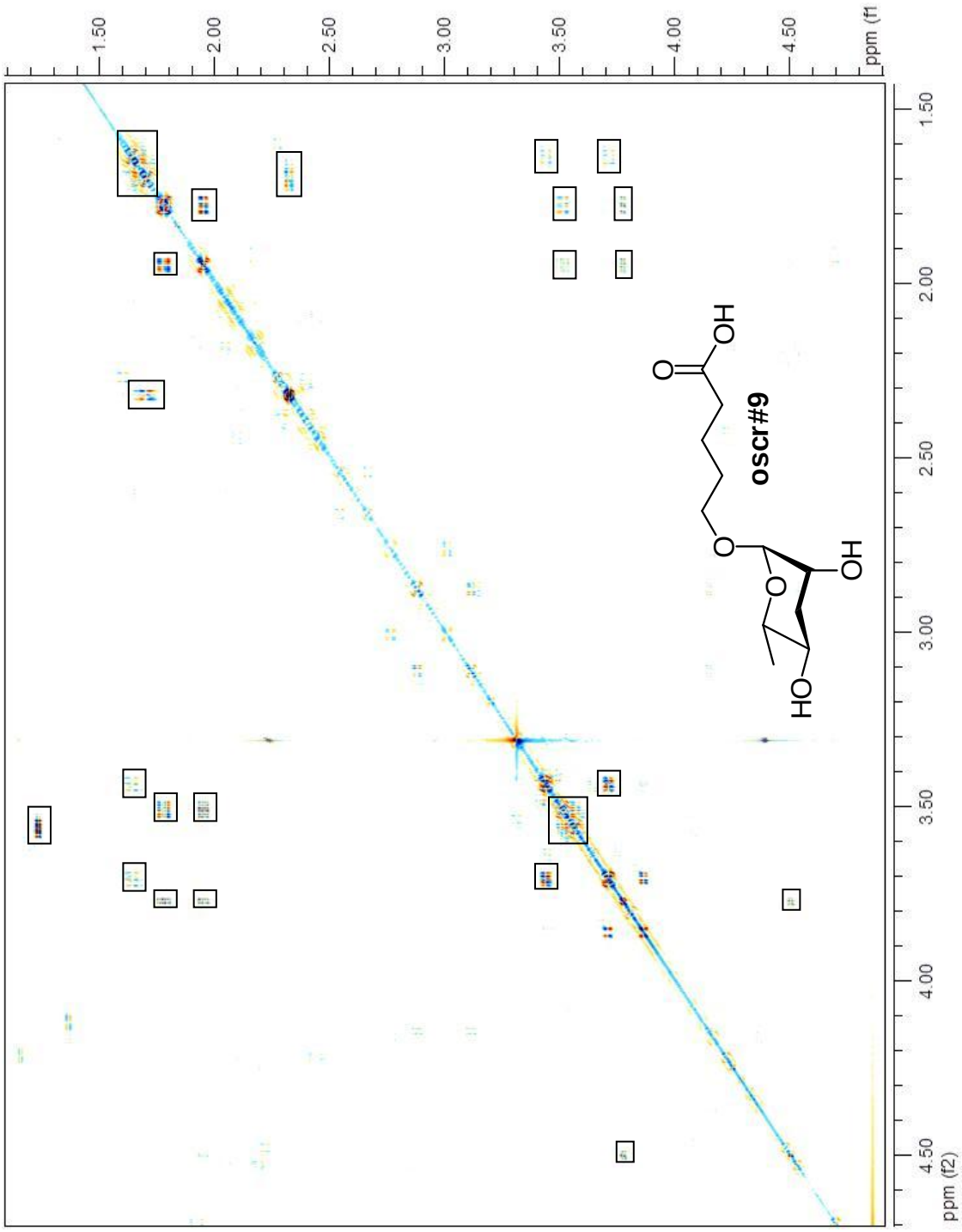


Figure A.6: dqfCOSY spectrum (600 MHz, methanol-d_4) of oscr#9–enriched fraction from *acox-1(ok2257)* media extracts.

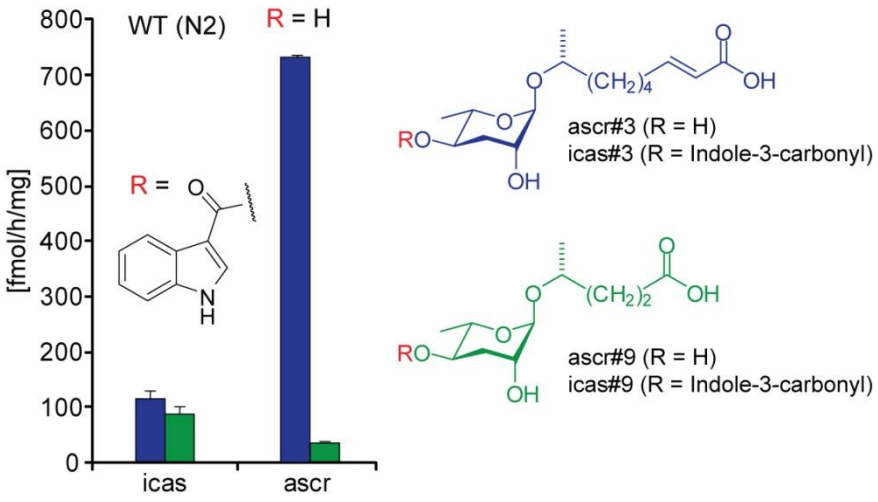


Figure A.7: Relative abundance of ascarosides ascr#3 and ascr#9 and their corresponding indole ascarosides icas#3 and icas#9 in wild-type excretome extracts indicates that indole attachment is highly dependent on side chain length.

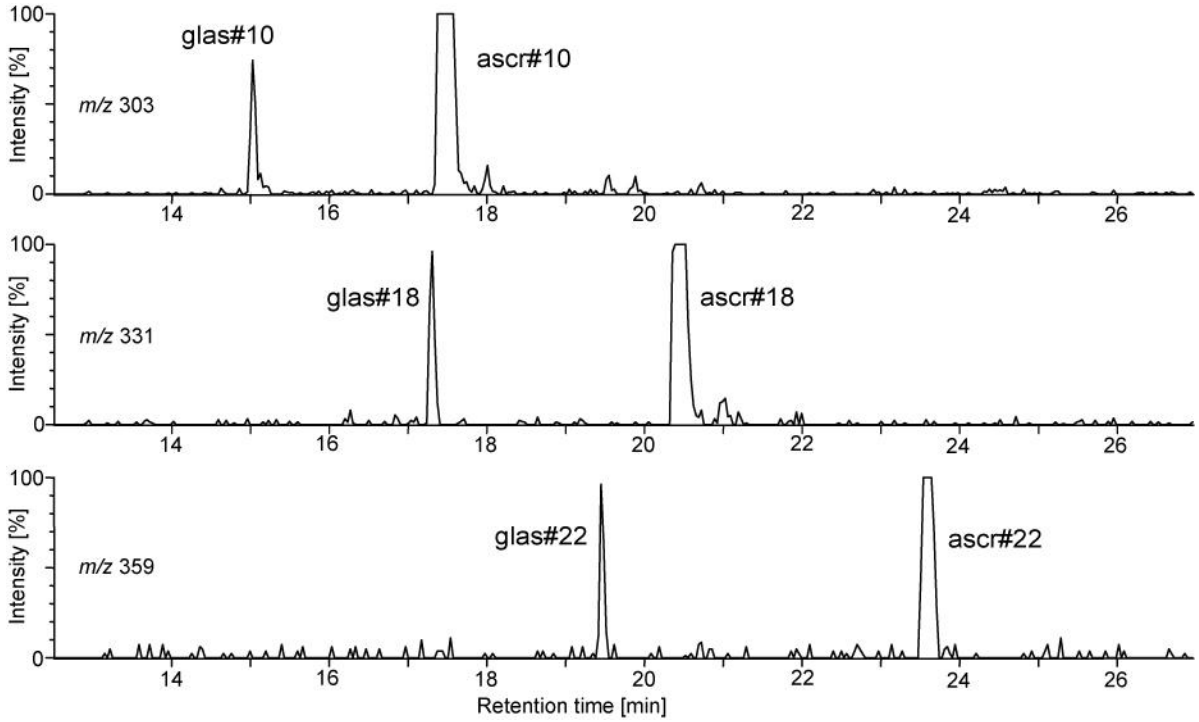


Figure A.8: HPLC-MS/MS chromatograms (precursor ions of *m/z* = 73) of *acox-1(ok2257)* worm body extracts showing glucosyl esters glas#10, glas#18, and glas#22 and the corresponding non-glycosylated ascarosides ascr#10, ascr#18, and ascr#22.

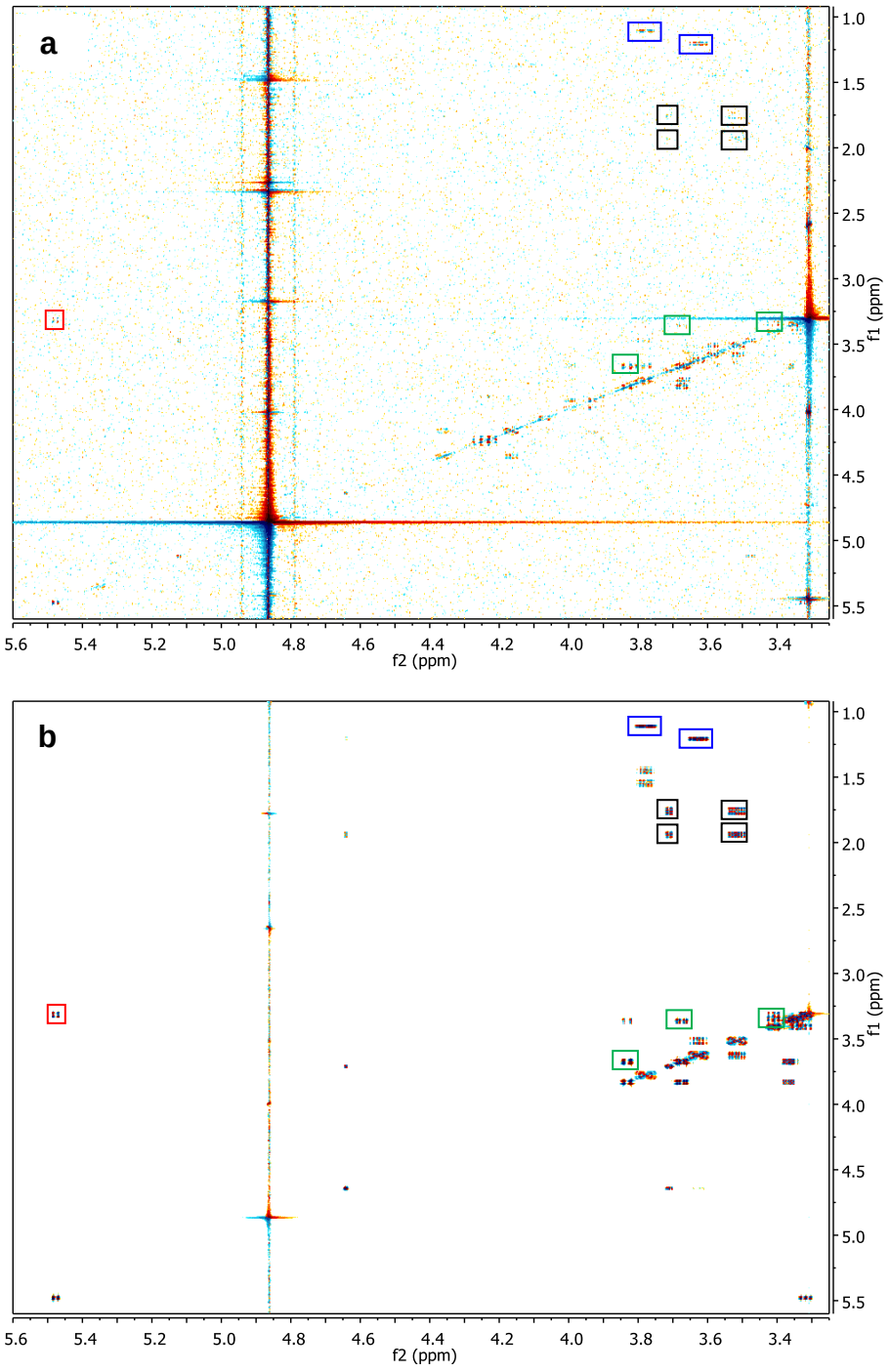


Figure A.9: Sections of dqfCOSY spectra (600 MHz, methanol-d_4) of glas#10–enriched fraction from *acox-1(ok2257)* worm pellet extract fraction (**a**) and synthetic glas#10 (**b**), showing characteristic signals for methyl groups of the ascarylose ring and the side chain (blue), the anomeric hydrogen of the glucose unit (red), the glucose spin system (green), and the ascarylose spin system (black).

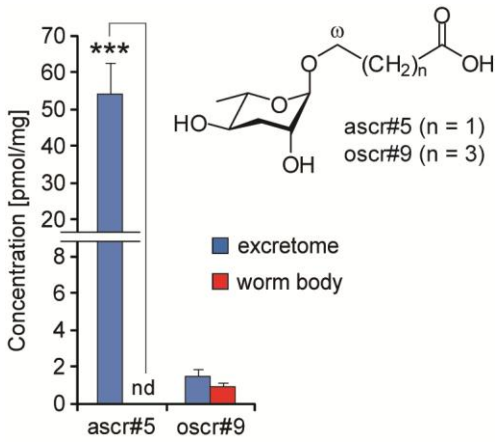


Figure A.10: Differential excretion of (ω)-oxygenated ascarosides by wild-type *C. elegans.* (Significance, *=*p*<0.01, **=*p*<0.001, ***=*p*<0.001, **Appendix Section A.1.8**)

```
ascE            MGVIVPHYLMIFKKLDIEGC---YLIEFNKFIDSRGTFVKTFHSDFFSE-NGIVLDMREE 56
C14F11.6        MSHPTPGKRFQLEKEVIEAIPDLLVIKPKVFPDERGFFSESYNKTEWAEKIGYTEDLQQD 60

ascE            FYSISAKNVIRGMHFQMPPAEHDKLVYCVNGAVLDVILDIRKDSKTYGEYFSIELSYENS 116
C14F11.6        NHSFSHYGVLRGLHTQP---HMGKLVTVVSGEIFDVAVDIRKDSPTYGKWHGVVLNGDNK 117

ascE            LALWVPKGLAHGFLSLADN-SIMFYKTSSVHNVECDSGIK--WNSFGFKWPIDNP---II 170
C14F11.6        HAFWIPAGFLHGFQVLSKEGAHVTYKCSAVYDPKTEFGINPFDEDINVDWPIRDKTVVIV 177

ascE            SEKDNSLCYFDEFDSSF 187
C14F11.6        SERDTQHASFKSL---- 190
```

Figure A.11: Alignment of *Yersinia pseudotuberculosis* CDP-3, 6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase or *asc*E (AAA88702.1) with *C. elegans* homolog C14F11.6 (CCD64543.1) was performed using ClustalW. Identical amino acids are marked in grey and similar amino acids are marked in light grey.

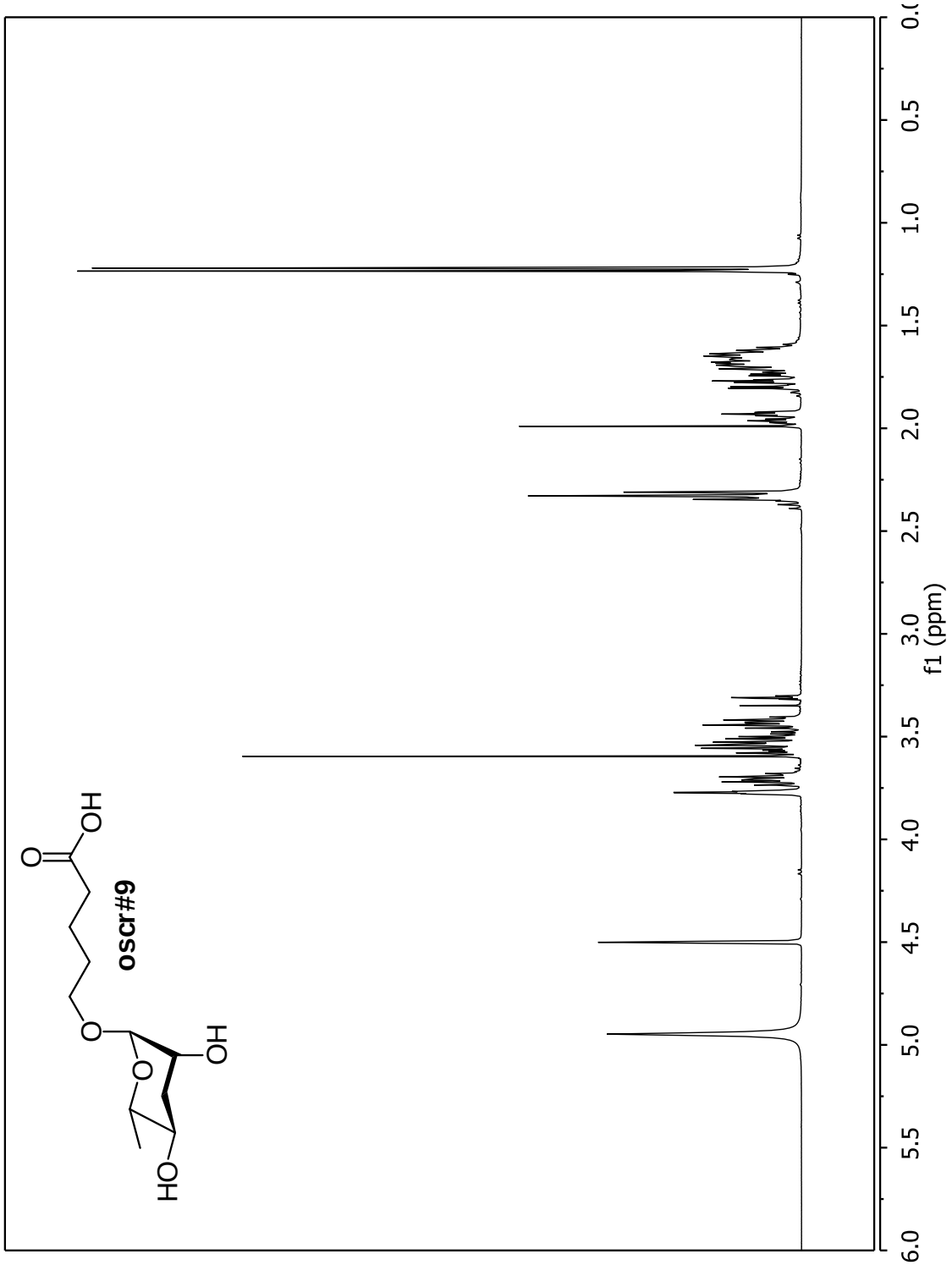


Figure A.12.1: ^{1}H NMR spectrum (400 MHz, methanol-d_4) of 5-(3'*R*,5'*R*-dihydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)pentanoic acid (**oscr#9**).

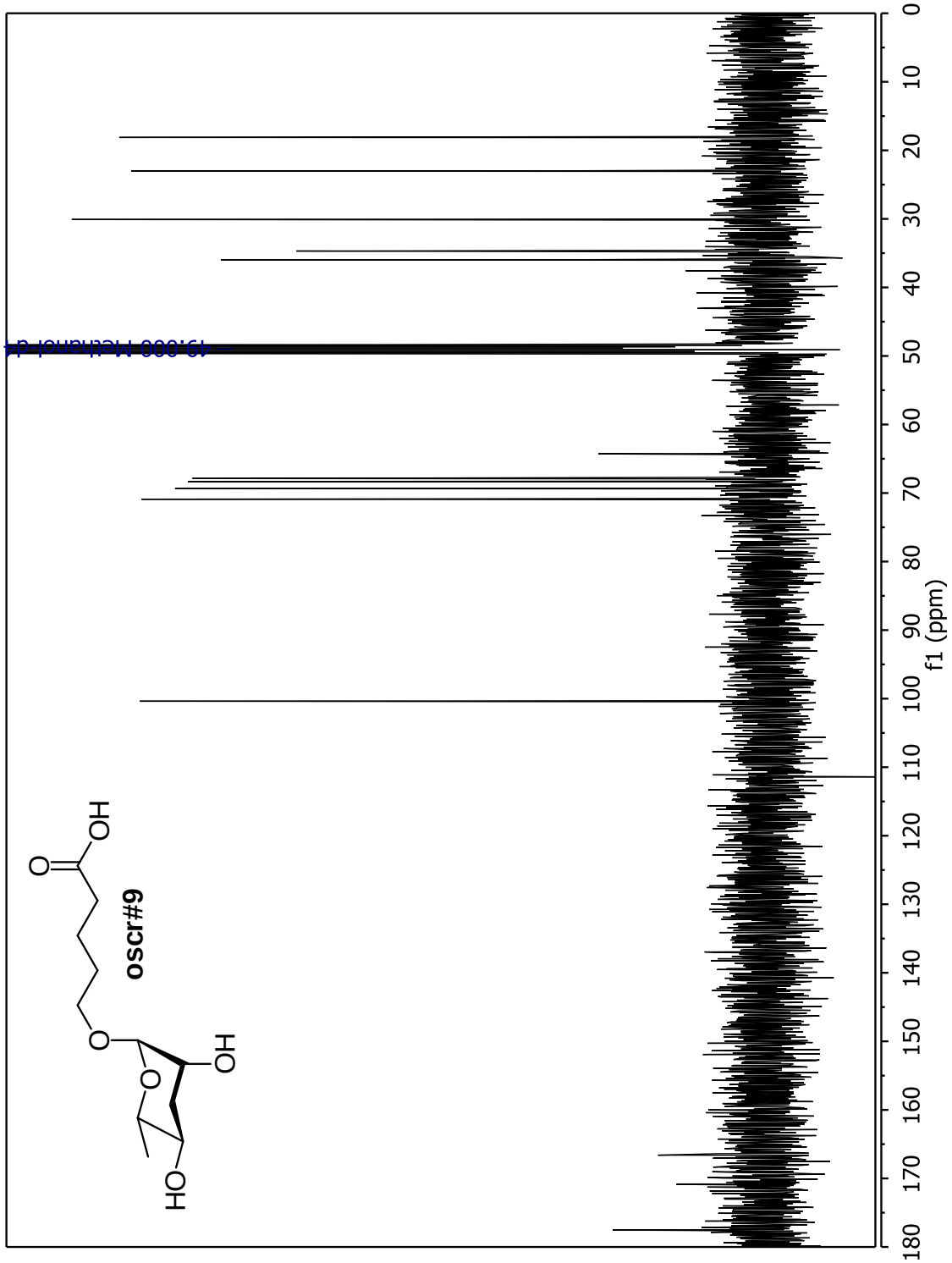


Figure A.12.2: ^{13}C NMR spectrum (100 MHz, methanol-d_4) of 5-(3'*R*,5'*R*-dihydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)pentanoic acid (**oscr#9**).

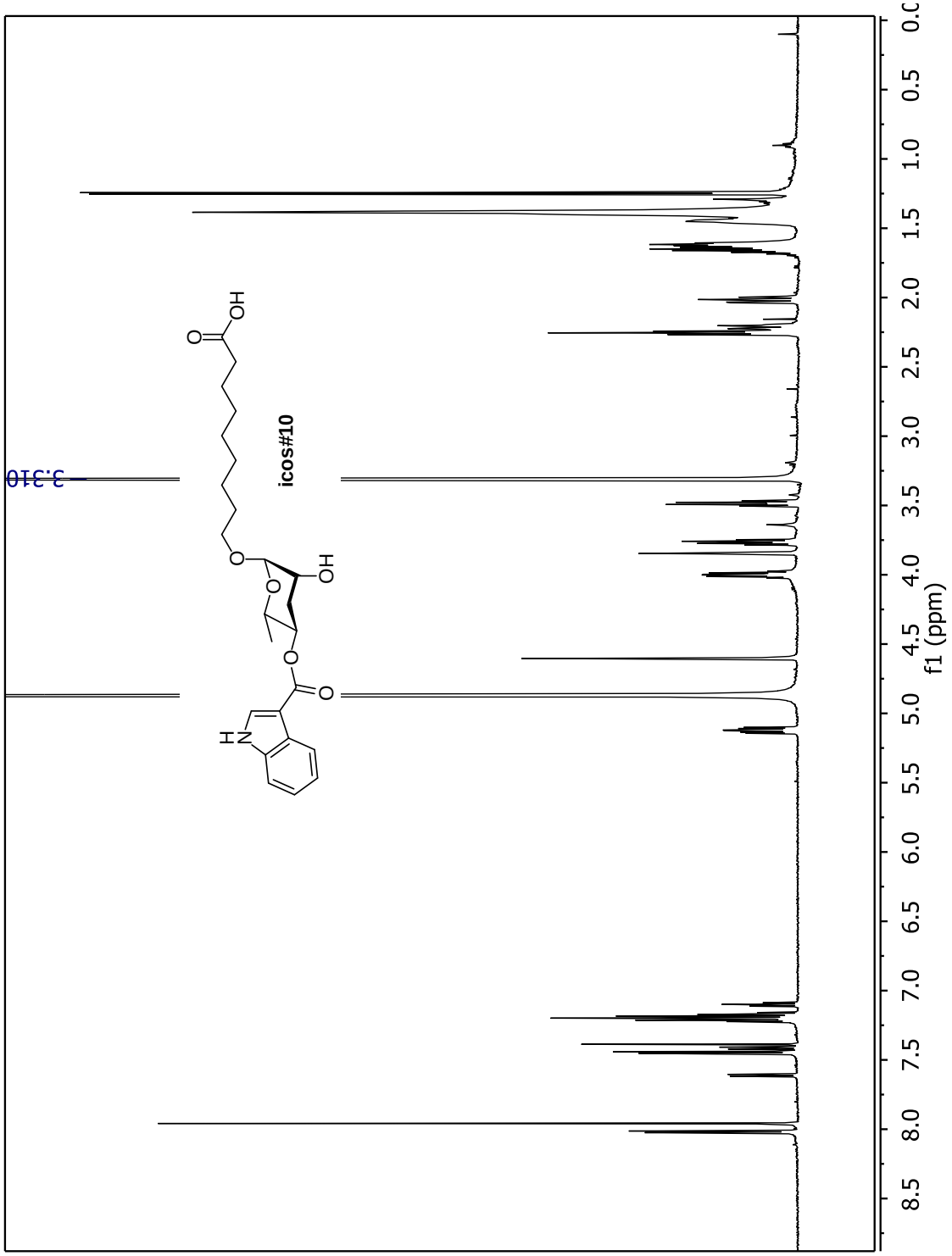


Figure A.13.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of 9-(5'*R*-(*1H*-indole-3-carbonyloxy)-3'*R*-hydroxy-6'*S*-methyl-tetrahydro-(*2H*)-pyran-2'-yloxy)nonanoic acid (**icos#10**).

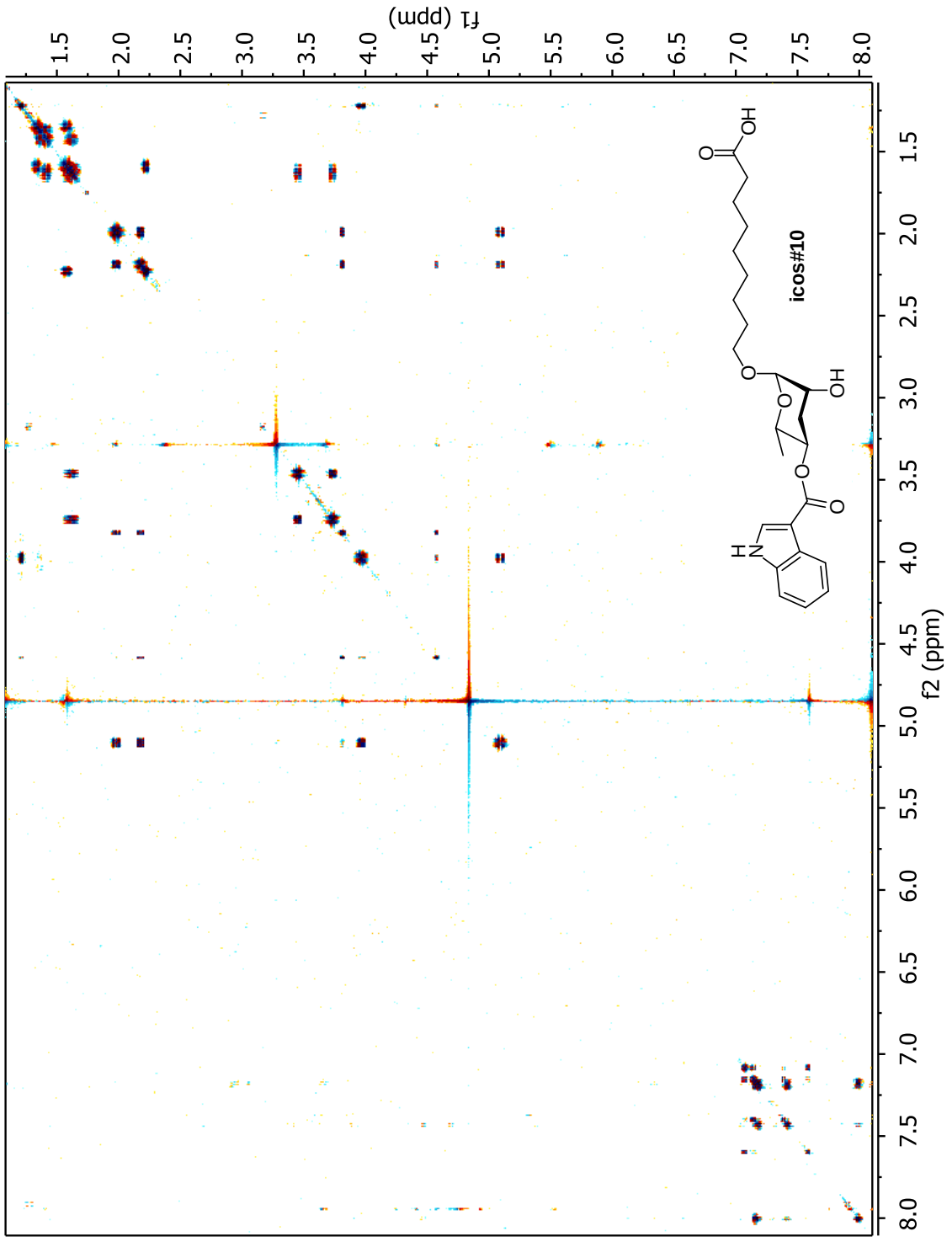


Figure A.13.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of 9-(5'*R*-(*1H*-indole-3-carbonyloxy) -3'*R*-hydroxy-6'*S*-methyl-tetrahydro-(*2H*)-pyran-2'-yloxy)nonanoic acid (**icos#10**).

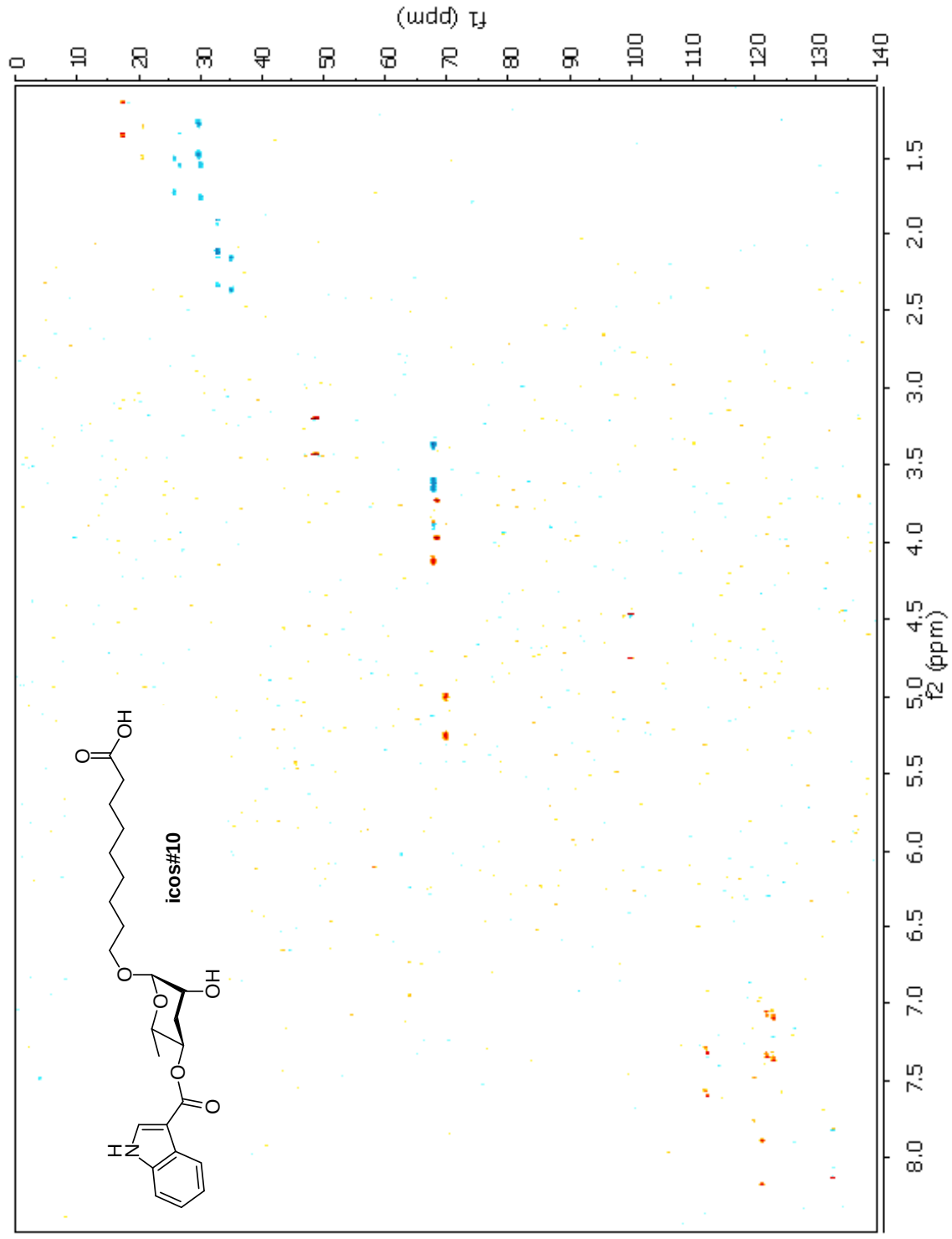


Figure A.13.3: HSQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of 9-(5'*R*-(*1H*-indole-3-carbonyloxy)-3'*R*-hydroxy-6'*S*-methyl-tetrahydro-(*2H*)-pyran-2'-yloxy)nonanoic acid (**icos#10**).

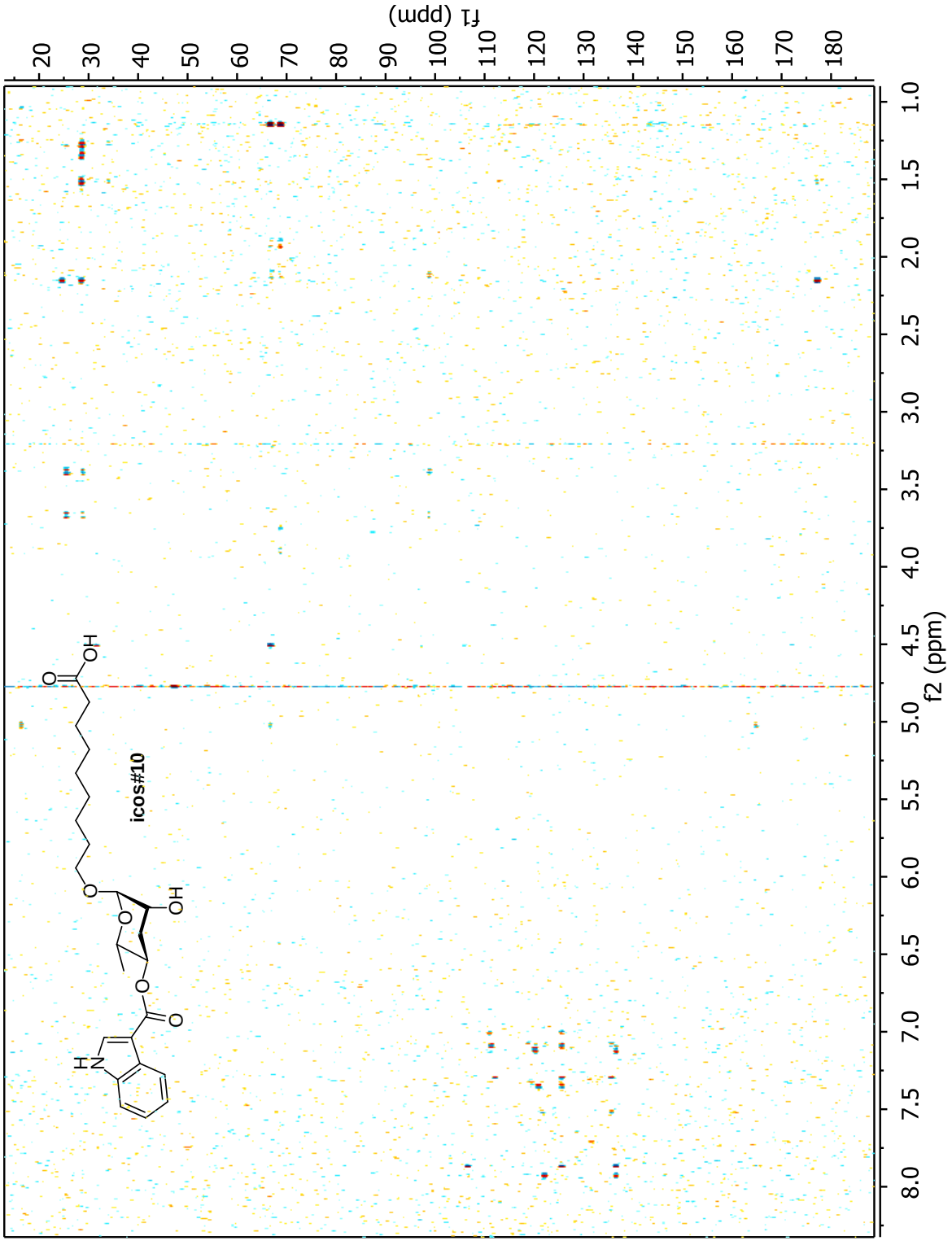


Figure A.13.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of 9-(5'*R*-(*1H*-indole-3-carbonyloxy)-3'*R*-hydroxy-6'*S*-methyl-tetrahydro-(*2H*)-pyran-2'-yloxy)nonanoic acid (**icos#10**).

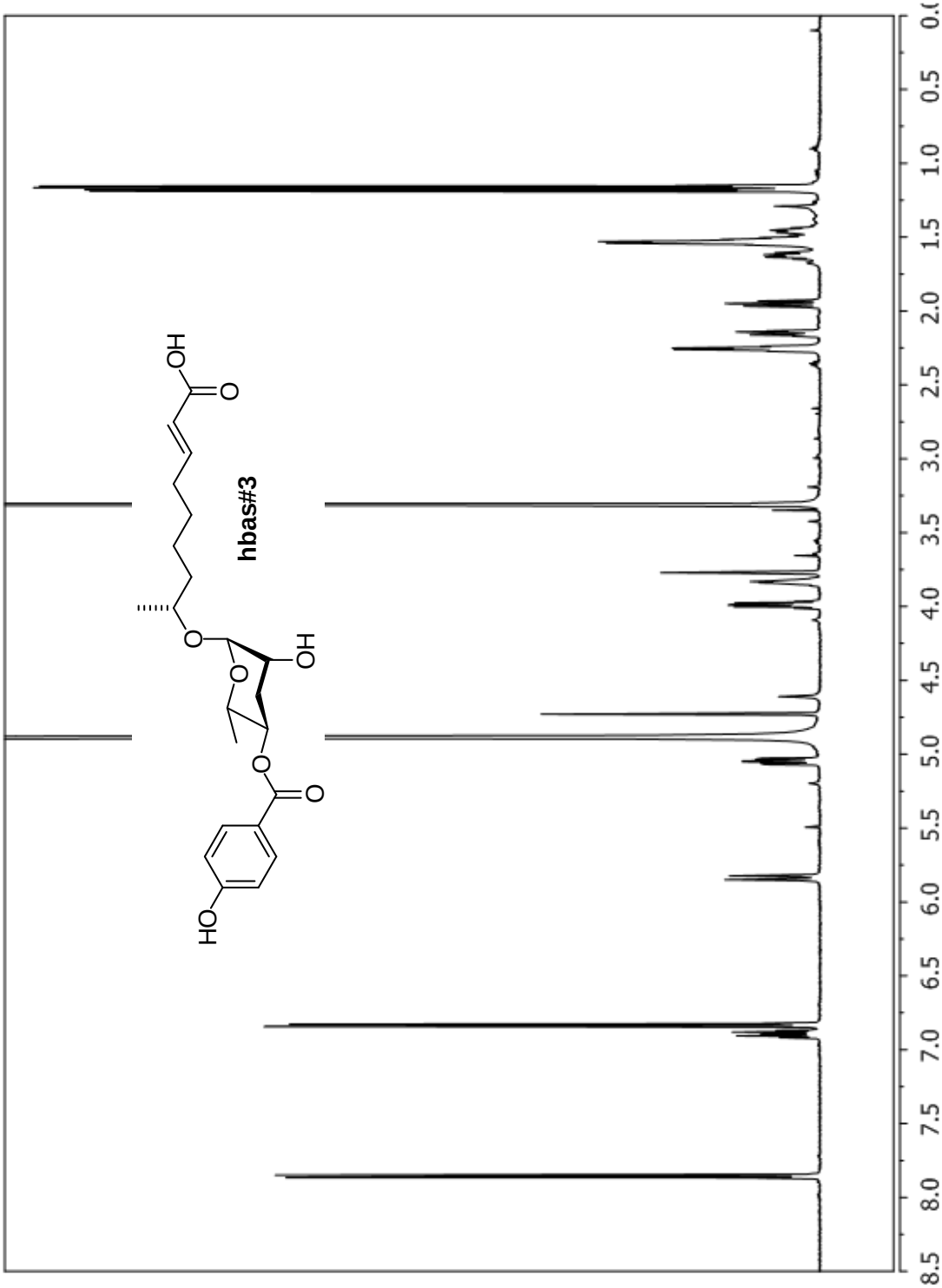


Figure A.14.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (8*R*)-(3'*R*-hydroxy-5'*R*-(4-hydroxybenzoyloxy)-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)non-(2*E*)-enoic acid (**hbas#3**).

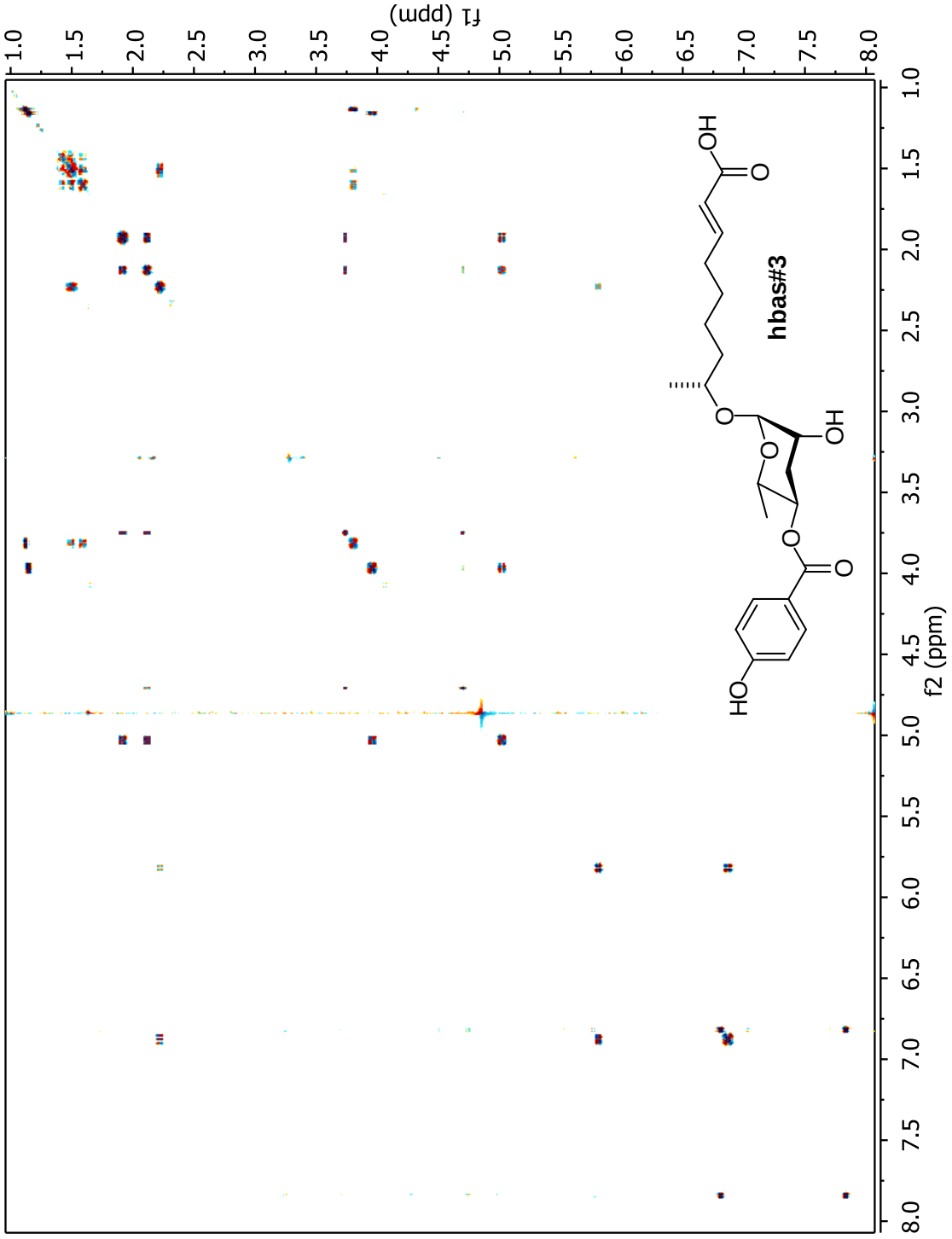


Figure A.14.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of (8*R*)-(3'*R*-hydroxy-5'*R*-(4-hydroxybenzoyloxy)-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)non-(2*E*)-enoic acid (**hbas#3**)

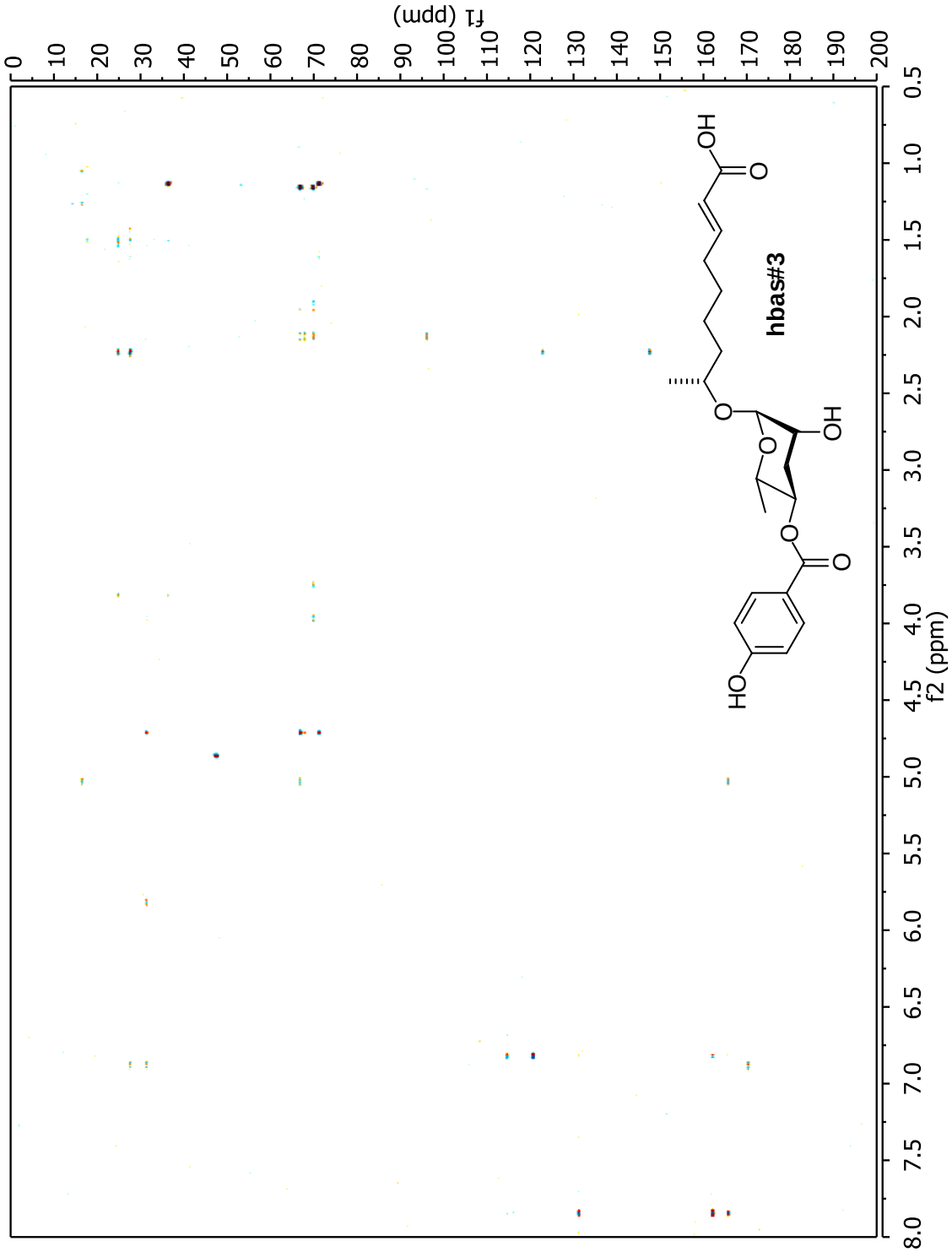


Figure A.14.3: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (8*R*)-(3'*R*-hydroxy-5'*R*-(4-hydroxybenzoyloxy)-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)non-(2*E*)-enoic acid (**hbas#3**).

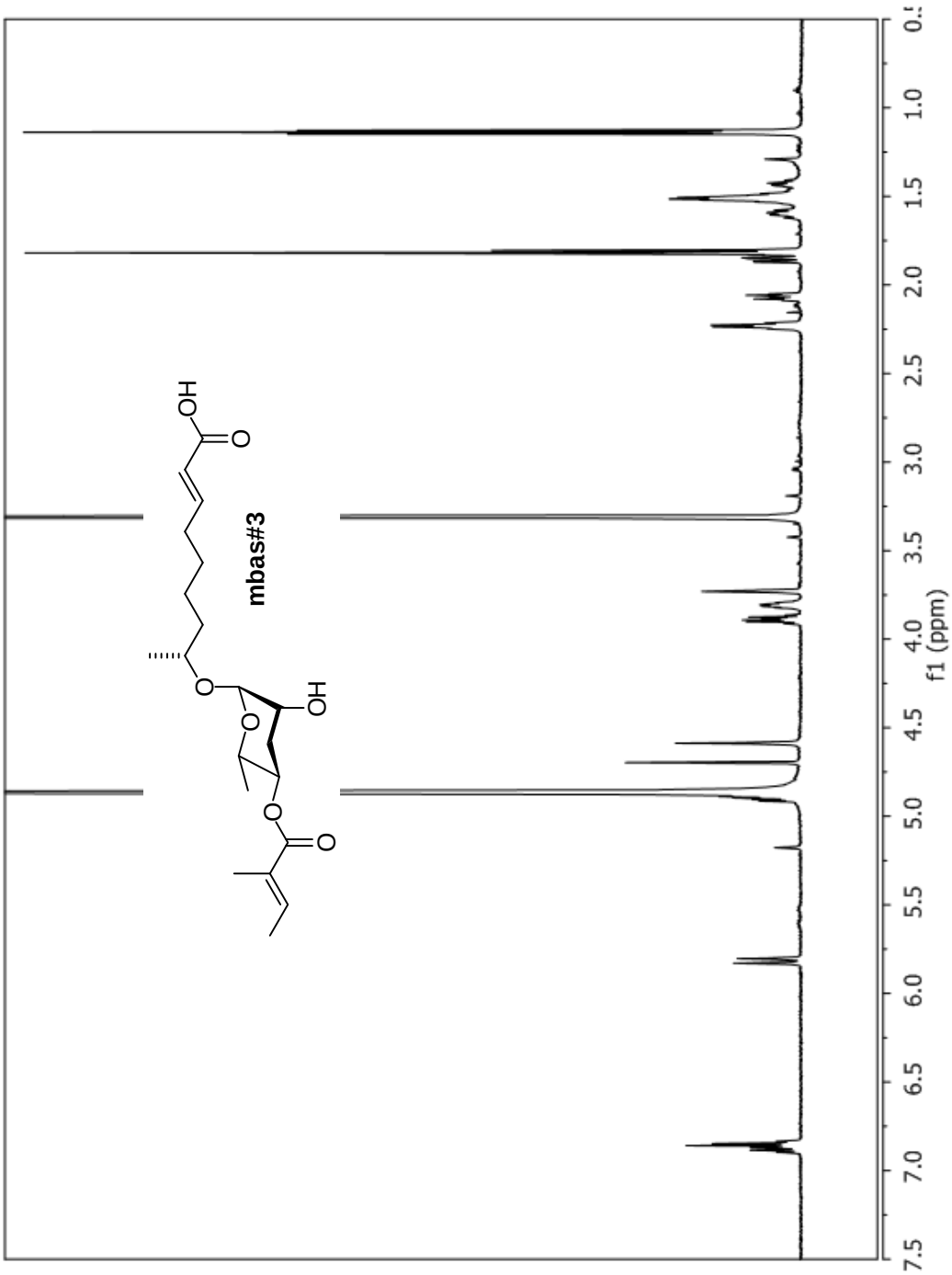


Figure A.15.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (8*R*)-(3'*R*-hydroxy-5'*R*-(*E*)-(2-methylbut-2-enoyloxy)-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)non-(2*E*)-enoic acid (**mbas#3**).

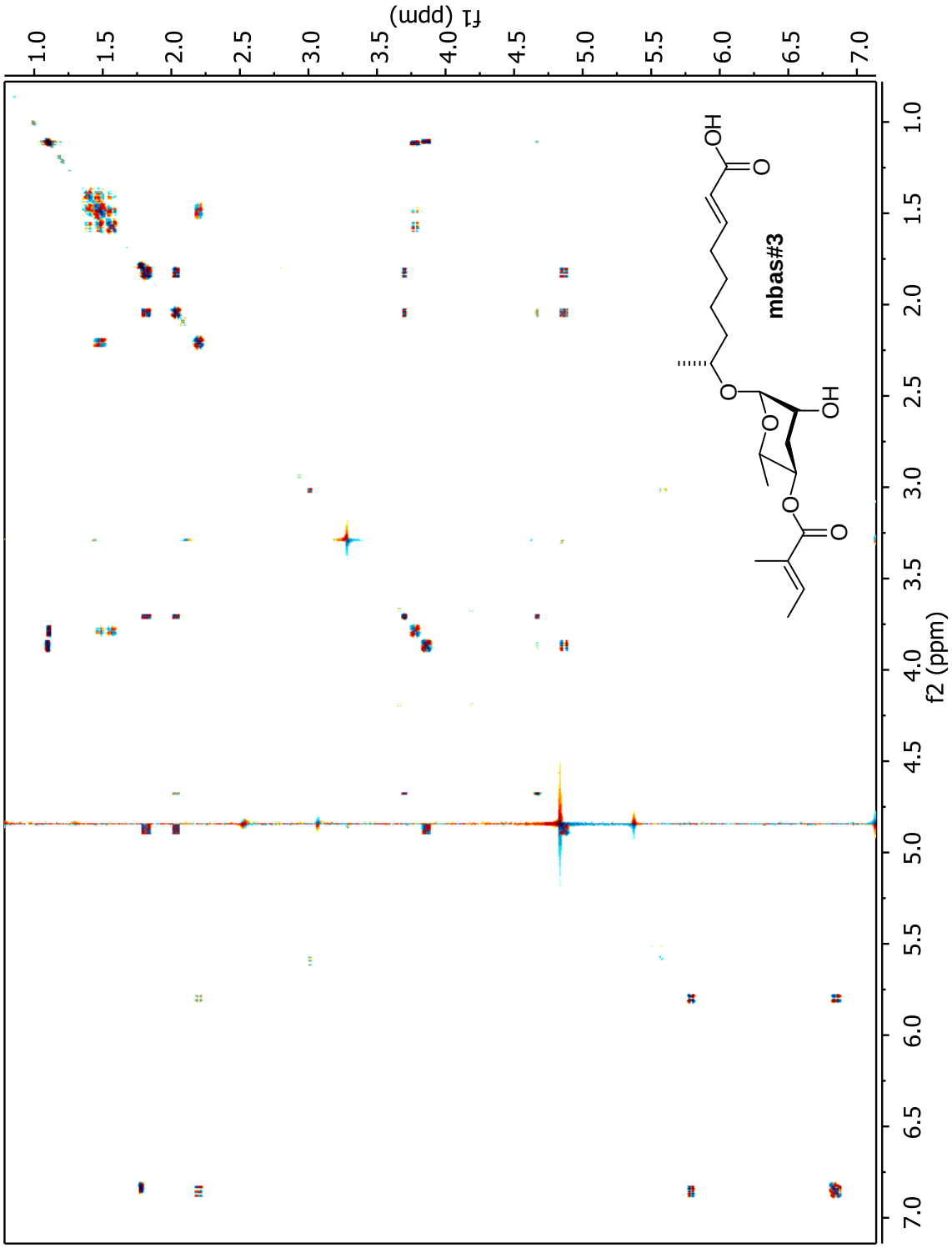


Figure A.15.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of (8*R*)-(3'*R*-hydroxy-5'*R*-(*E*)-(2-methylbut-2-enoyloxy)-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)non-(2*E*)-enoic acid (**mbas#3**).

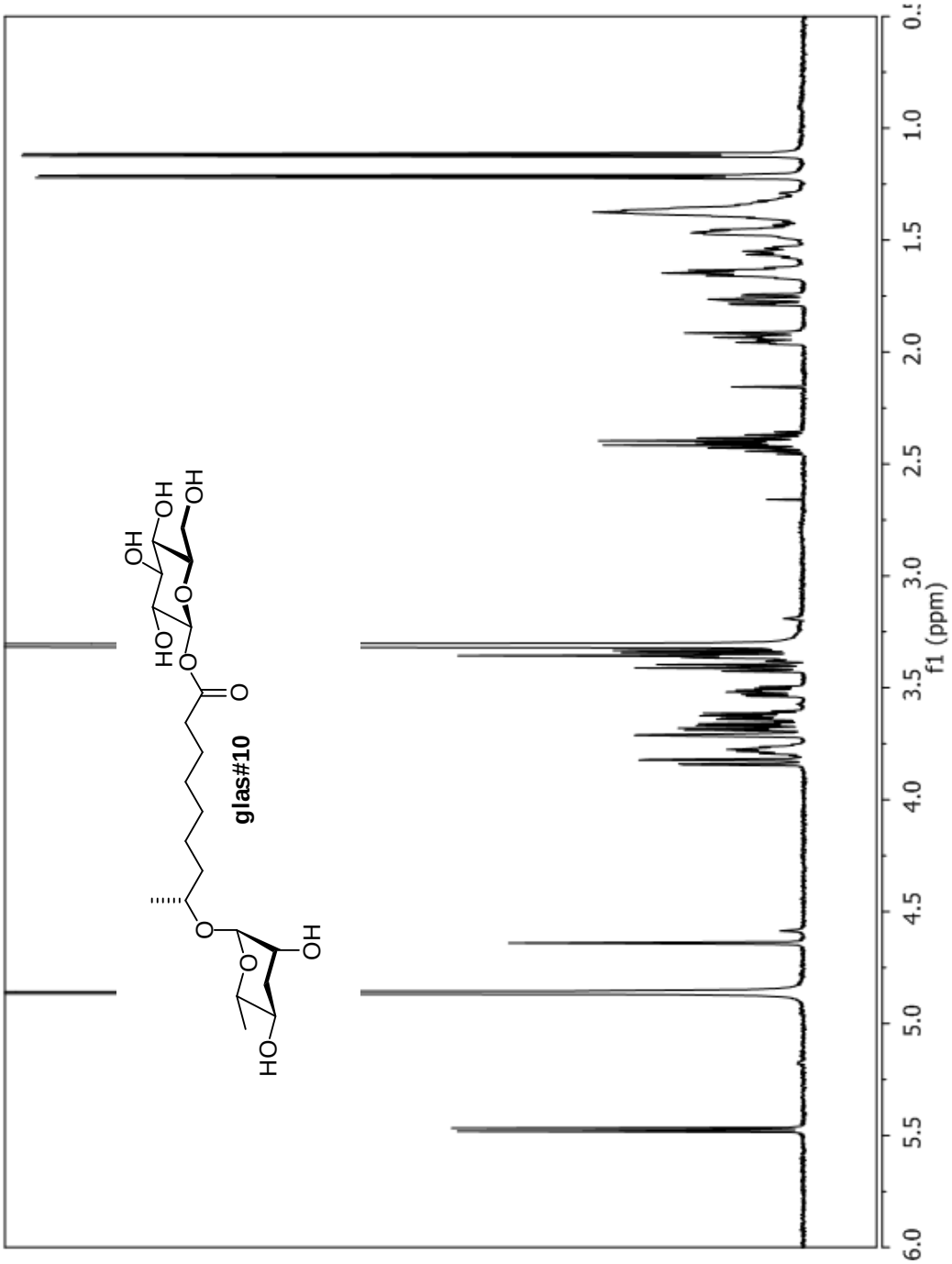


Figure A.16.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of 2-(8*R*)-(3'*R*,5'*R*-di-hydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)nonanoyl-3,4,5-trihydroxy-6-hydroxymethyl-(*2H*)-tetrahydropyran (**glas#10**).

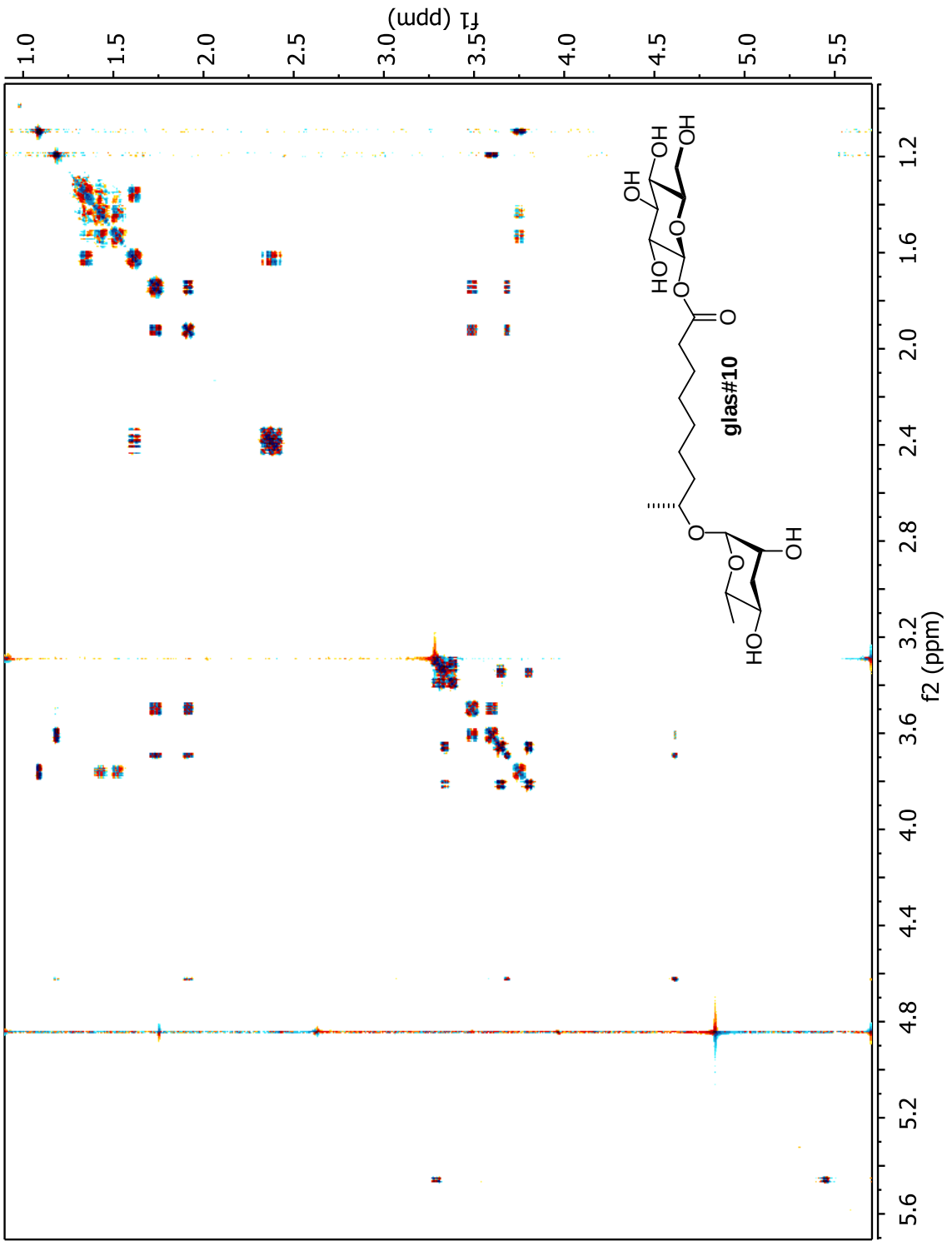


Figure A.16.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of 2-(8*R*)-(3'*R*,5'*R*-di-hydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)nonanoyl-3,4,5-trihydroxy-6-hydroxymethyl-(*2H*)-tetrahydropyran **(glas#10).**

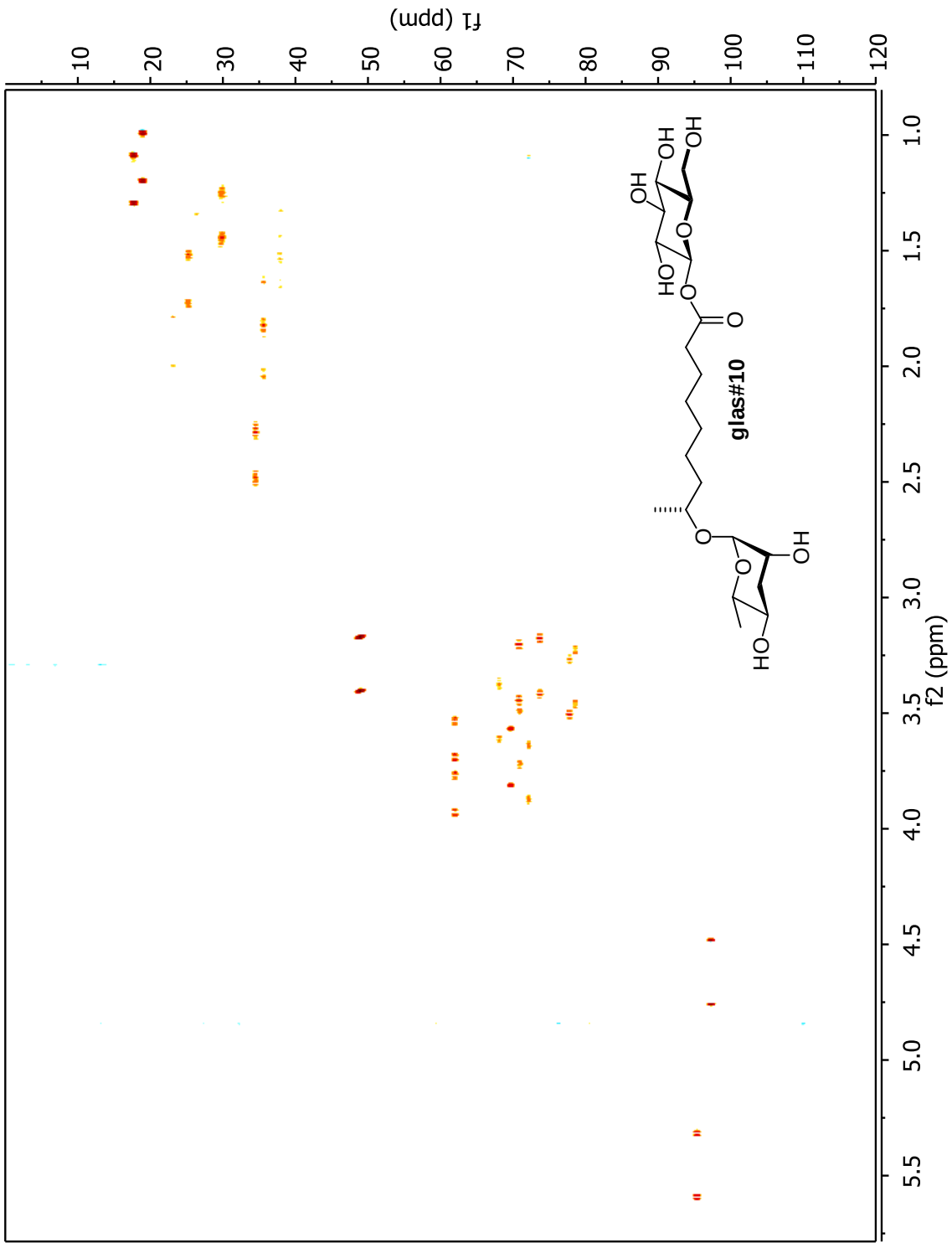


Figure A.16.3: HMQC Spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of 2-(8*R*)-(3'*R*,5'*R*-di-hydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)nonanoyl-3,4,5-trihydroxy-6-hydroxymethyl-(*2H*)-tetrahydropyran (**glas#10**).

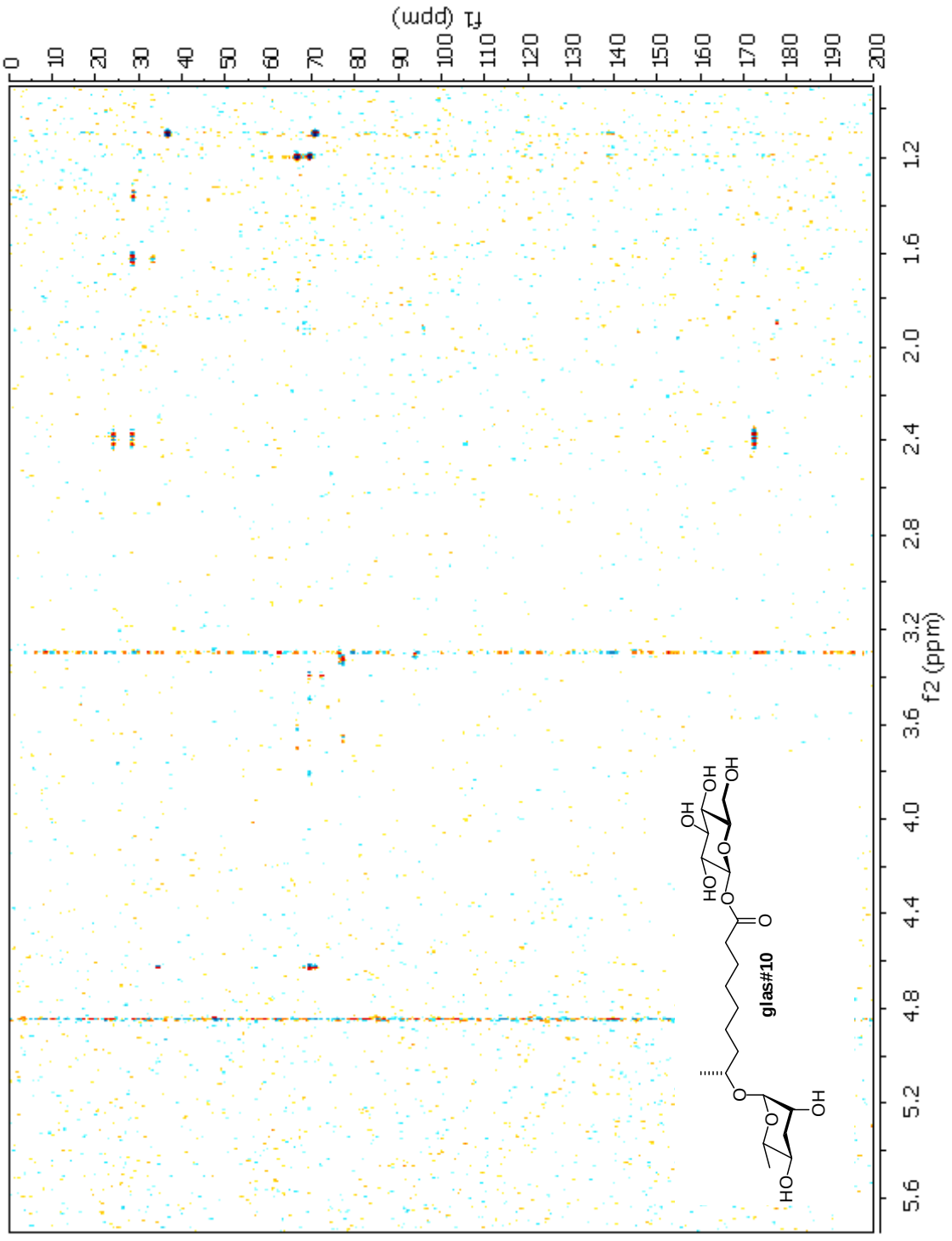
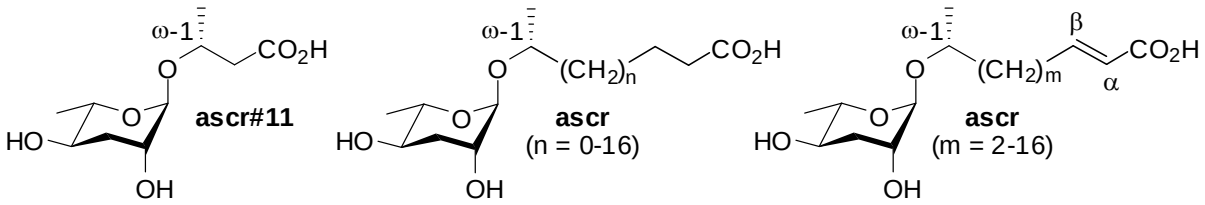


Figure A.16.4: HMBC spectrum (600 MHz, methanol-d_4) of 2-(8*R*)-(3'*R*,5'*R*-di-hydroxy-6'*S*-methyl-(*2H*)-tetrahydropyran-2'-yloxy)nonanoyl-3,4,5-trihydroxy-6-hydroxymethyl-(*2H*)-tetrahydropyran (**glas#10**).

A.3. Tables:

Table A.1: HPLC-ESI-MS data of (ω-1)-oxygenated ascarosides (ascr).

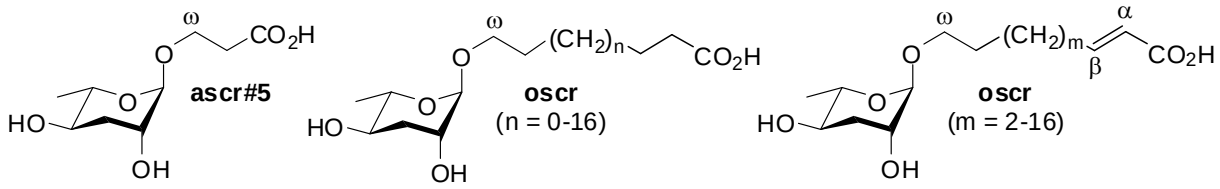


Side chain length (n, m)	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* [M-H]⁻ calculated	*m/z* [M-H]⁻ observed	Retention time [min] ±SD
C_4	ascr#11*	$C_{10}H_{18}O_6$	234.1103	233.1025	233.1031	10.21 ±0.03
C_5, n = 0	ascr#9*[3]	$C_{11}H_{20}O_6$	248.1260	247.1182	247.1189	11.69 ±0.01
C_6, n = 1	ascr#12	$C_{12}H_{22}O_6$	262.1416	261.1338	261.1343	13.09 ±0.02
ΔC_7, m = 2	ascr#7*[4]	$C_{13}H_{22}O_6$	274.1416	273.1339	273.1337	13.96 ±0.07
C_7, n = 2	ascr#1*[5]	$C_{13}H_{24}O_6$	276.1573	275.1495	275.1497	14.52 ±0.04
ΔC_8, m = 3	ascr#13	$C_{14}H_{24}O_6$	288.1573	287.1495	287.1481	15.61 ±0.18
C_8, n = 3	ascr#14	$C_{14}H_{26}O_6$	290.1729	289.1651	289.1647	15.96 ±0.03
ΔC_9, m = 4	ascr#3*[6]	$C_{15}H_{26}O_6$	302.1729	301.1651	301.1652	16.90 ±0.02
C_9, n = 4	ascr#10*[3]	$C_{15}H_{28}O_6$	304.1886	303.1808	303.1800	17.44 ±0.02
ΔC_{10}, m = 5	ascr#15	$C_{16}H_{28}O_6$	316.1886	315.1808	315.1817	18.34 ±0.02
C_{10}, n = 5	ascr#16	$C_{16}H_{30}O_6$	318.2042	317.1964	317.1959	18.98 ±0.05
ΔC_{11}, m = 6	ascr#17	$C_{17}H_{30}O_6$	330.2042	329.1964	329.1957	19.75 ±0.02
C_{11}, n = 6	ascr#18	$C_{17}H_{32}O_6$	332.2199	331.2121	331.2130	20.43 ±0.02
ΔC_{12}, m = 7	ascr#19[7]	$C_{18}H_{32}O_6$	344.2199	343.2121	343.2120	21.36 ±0.03
C_{12}, n = 7	ascr#20	$C_{18}H_{34}O_6$	346.2355	345.2277	345.2278	21.97 ±0.03
ΔC_{13}, m = 8	ascr#21[7]	$C_{19}H_{34}O_6$	358.2355	357.2277	357.2273	22.83 ±0.02
C_{13}, n = 8	ascr#22	$C_{19}H_{36}O_6$	360.2512	359.2434	359.2437	23.58 ±0.02
ΔC_{14}, m = 9	ascr#23[7]	$C_{20}H_{36}O_6$	372.2512	371.2434	371.2444	24.46 ±0.01
C_{14}, n = 9	ascr#24	$C_{20}H_{38}O_6$	374.2668	373.2590	373.2596	25.29 ±0.03

ΔC_{15}, m = 10	ascr#25[7]	$C_{21}H_{38}O_6$	386.2668	385.2590	385.2598	26.15 ±0.02
C_{15}, n = 10	ascr#26	$C_{21}H_{40}O_6$	388.2825	387.2747	387.2743	27.09 ±0.02
ΔC_{16}, m = 11	ascr#27	$C_{22}H_{40}O_6$	400.2825	399.2747	399.2734	27.89 ±0.03
C_{16}, n = 11	ascr#28	$C_{22}H_{42}O_6$	402.2981	401.2903	401.2901	28.97 ±0.04
ΔC_{17}, m = 12	ascr#29	$C_{23}H_{42}O_6$	414.2981	413.2903	413.2891	29.80 ±0.03
C_{17}, n = 12	ascr#30	$C_{23}H_{44}O_6$	416.3138	415.3060	415.3067	30.96 ±0.03
ΔC_{18}, m = 13	ascr#31	$C_{24}H_{44}O_6$	428.3138	427.3060	427.3075	31.78 ±0.03
C_{18}, n = 13	ascr#32	$C_{24}H_{46}O_6$	430.3294	429.3216	429.3221	33.02 ±0.02
ΔC_{19}, m = 14	ascr#33	$C_{25}H_{46}O_6$	442.3294	441.3216	441.3215	33.74 ±0.03
C_{19}, n = 14	ascr#34	$C_{25}H_{48}O_6$	444.3451	443.3373	443.3374	35.12 ±0.08
ΔC_{20}, m = 15	ascr#35	$C_{26}H_{48}O_6$	456.3451	455.3373	455.3371	35.59 ±0.06
C_{20}, n = 15	ascr#36	$C_{26}H_{50}O_6$	458.3607	457.3529	457.3501	37.14 ±0.07
ΔC_{21}, m = 16	ascr#37	$C_{27}H_{50}O_6$	470.3607	469.3529	469.3519	37.71 ±0.13
C_{21}, n = 16	ascr#38	$C_{27}H_{52}O_6$	472.3764	471.3686	471.3697	39.15 ±0.06

* confirmed using synthetic standards.

Table A.2: HPLC-ESI-MS data of (ω)-oxygenated ascarosides (oscr).

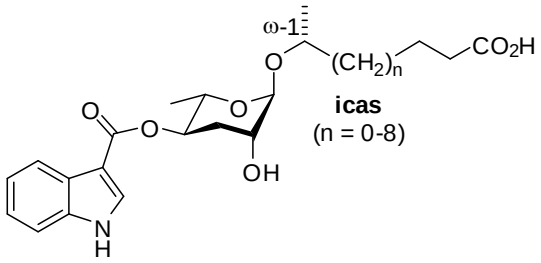


Side chain length (n, m)	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ observed	Retention time [min] ±SD
C_3	ascr#5*[8]	$C_9H_{16}O_6$	220.0947	219.0869	219.0871	6.53 ±0.07
C_5, n = 0	oscr#9*[1]	$C_{11}H_{20}O_6$	248.1260	247.1182	247.1192	11.88 ±0.02
C_6, n = 1	oscr#12	$C_{12}H_{22}O_6$	262.1416	261.1338	261.1345	13.40 ±0.02
ΔC_7, m = 2	oscr#7	$C_{13}H_{22}O_6$	274.1416	273.1339	273.1350	14.67 ±0.02
C_7, n = 2	oscr#1	$C_{13}H_{24}O_6$	276.1573	275.1495	275.1503	14.95 ±0.03

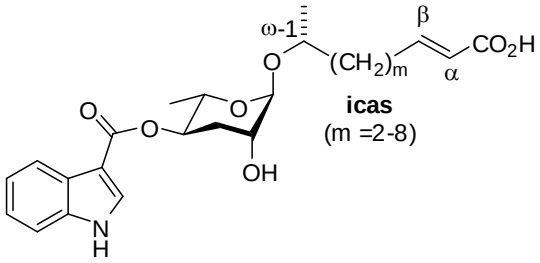
C_8, n = 3	oscr#14	$C_{14}H_{26}O_6$	290.1729	289.1651	289.1672	16.45 ±0.02
ΔC_9, n = 4	oscr#3	$C_{15}H_{26}O_6$	302.1729	301.1651	301.1636	17.58 ±0.03
C_9, n = 4	oscr#10*[1]	$C_{15}H_{28}O_6$	304.1886	303.1808	303.1814	18.00 ±0.02
ΔC_{10}, m = 5	oscr#15	$C_{16}H_{28}O_6$	316.1886	315.1808	315.1816	18.91 ±0.06
C_{10}, n = 5	oscr#16	$C_{16}H_{30}O_6$	318.2042	317.1964	317.1967	19.48 ±0.03
ΔC_{11}, m = 6	oscr#17	$C_{17}H_{30}O_6$	330.2042	329.1964	329.1956	20.39 ±0.01
C_{11}, n = 6	oscr#18	$C_{17}H_{32}O_6$	332.2199	331.2121	331.2124	20.98 ±0.08
ΔC_{12}, m = 7	oscr#19	$C_{18}H_{32}O_6$	344.2199	343.2121	343.2125	21.86 ±0.06
C_{12}, n = 7	oscr#20	$C_{18}H_{34}O_6$	346.2355	345.2277	345.2302	22.54 ±0.03
ΔC_{13}, m = 8	oscr#21	$C_{19}H_{34}O_6$	358.2355	357.2277	357.2271	23.41 ±0.02
C_{13}, n = 8	oscr#22	$C_{19}H_{36}O_6$	360.2512	359.2434	359.2452	24.19 ±0.02
ΔC_{14}, m = 9	oscr#23	$C_{20}H_{36}O_6$	372.2512	371.2434	371.2436	25.04 ±0.03
C_{14}, n = 9	oscr#24	$C_{20}H_{38}O_6$	374.2668	373.2590	373.2589	25.91 ±0.02
ΔC_{15}, m = 10	oscr#25	$C_{21}H_{38}O_6$	386.2668	385.2590	385.2567	26.74 ±0.01
C_{15}, n = 10	oscr#26	$C_{21}H_{40}O_6$	388.2825	387.2747	387.2739	27.73 ±0.02
ΔC_{16}, m = 11	oscr#27	$C_{22}H_{40}O_6$	400.2825	399.2747	399.2728	28.54 ±0.03
C_{16}, n = 11	oscr#28	$C_{22}H_{42}O_6$	402.2981	401.2903	401.2905	29.67 ±0.02
ΔC_{17}, m = 12	oscr#29	$C_{23}H_{42}O_6$	414.2981	413.2903	413.2900	30.42 ±0.02
C_{17}, n = 12	oscr#30	$C_{23}H_{44}O_6$	416.3138	415.3060	415.3080	31.68 ±0.04
ΔC_{18}, m = 13	oscr#31	$C_{24}H_{44}O_6$	428.3138	427.3060	427.3053	32.44 ±0.02
C_{18}, n = 13	oscr#32	$C_{24}H_{46}O_6$	430.3294	429.3216	429.3207	33.78 ±0.02
ΔC_{19}, m = 14	oscr#33	$C_{25}H_{46}O_6$	442.3294	441.3216	441.3218	34.44 ±0.05
C_{19}, n = 14	oscr#34	$C_{25}H_{48}O_6$	444.3451	443.3373	443.3372	35.86 ±0.05
ΔC_{20}, m = 15	oscr#35	$C_{26}H_{48}O_6$	456.3451	455.3373	455.3384	36.23 ±0.05
C_{20}, n = 15	oscr#36	$C_{26}H_{50}O_6$	458.3607	457.3529	457.3545	37.96 ±0.05
ΔC_{21}, m = 16	oscr#37	$C_{27}H_{50}O_6$	470.3607	469.3529	469.3504	38.08 ±0.19
C_{21}, n = 16	oscr#38	$C_{27}H_{52}O_6$	472.3764	471.3686	471.3679	40.19 ±0.13

* confirmed using synthetic standards

Table A.3: HPLC-ESI-MS data of (ω-1)-oxygenated indole ascarosides (icas).



icas
(n = 0-8)

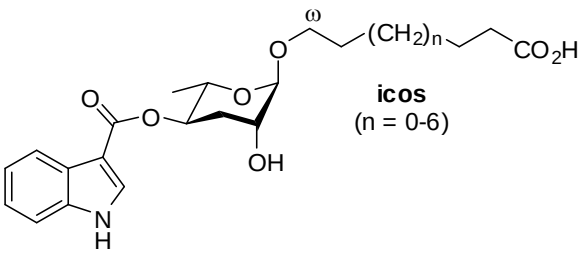


icas
(m =2-8)

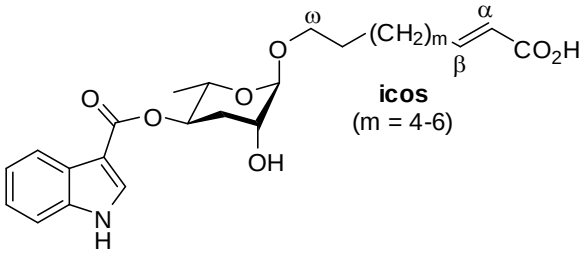
Side chain length (n, m)	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ observed	Retention time [min] ±SD
C_5, n = 0	icas#9*[9]	$C_{20}H_{25}NO_7$	391.1631	390.1553	390.1555	20.18 ±0.02
C_6, n = 1	icas#12	$C_{21}H_{27}NO_7$	405.1788	404.1709	404.1703	21.02 ±0.04
ΔC_7, m = 2	icas#7*[3]	$C_{22}H_{27}NO_7$	417.1788	416.1709	416.1719	21.74 ±0.03
C_7, n = 2	icas#1*[3]	$C_{22}H_{29}NO_7$	419.1944	418.1866	418.1864	22.06 ±0.03
C_8, n = 3	icas#14	$C_{23}H_{31}NO_7$	433.2101	432.2022	432.2013	23.20 ±0.04
ΔC_9, m = 4	icas#3*[3]	$C_{24}H_{31}NO_7$	445.2101	444.2022	444.2029	23.94 ±0.03
C_9, n = 4	icas#10[3]	$C_{24}H_{33}NO_7$	447.2257	446.2179	446.2185	24.55 ±0.03
ΔC_{10}, m = 5	icas#15	$C_{25}H_{33}NO_7$	459.2257	458.2179	458.2198	25.21 ±0.05
C_{10}, n = 5	icas#16	$C_{25}H_{35}NO_7$	461.2414	460.2335	460.2369	25.89 ±0.04
ΔC_{11}, m = 6	icas#17	$C_{26}H_{35}NO_7$	473.2414	472.2335	472.2344	26.68 ±0.04
C_{11}, n = 6	icas#18	$C_{26}H_{37}NO_7$	475.2570	474.2492	474.2494	27.44 ±0.03
ΔC_{12}, m = 7	icas#19	$C_{27}H_{37}NO_7$	487.2570	486.2492	486.2486	28.20 ±0.06
C_{12}, n = 7	icas#20	$C_{27}H_{39}NO_7$	489.27265	488.2648	488.2628	28.98 ±0.03
ΔC_{13}, m = 8	icas#21	$C_{28}H_{39}NO_7$	501.27265	500.2648	500.2640	29.71 ±0.04
C_{13}, n = 8	icas#22	$C_{28}H_{41}NO_7$	503.28830	502.2805	502.2807	30.66 ±0.03

* confirmed using synthetic standards

Table A.4: HPLC-ESI-MS data of (ω)-oxygenated indole ascarosides (icos).



icos
(n = 0-6)

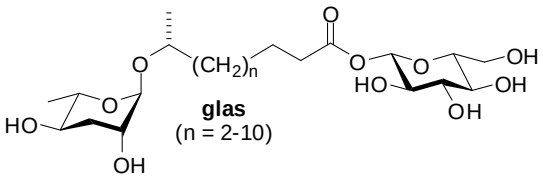


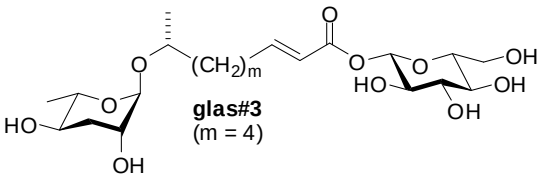
icos
(m = 4-6)

Side chain length (n, m)	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ observed	Retention time [min] ±SD
C_5, n = 0	icos#9	$C_{20}H_{25}NO_7$	391.1631	390.1553	390.1541	20.48 ±0.03
C_7, n = 2	icos#1	$C_{22}H_{29}NO_7$	419.1944	418.1866	418.1858	22.63 ±0.03
ΔC_9, m = 4	icos#3	$C_{24}H_{31}NO_7$	445.2101	444.2022	444.2049	24.66 ±0.04
C_9, n = 4	icos#10*[1]	$C_{24}H_{33}NO_7$	447.2257	446.2179	446.2171	25.29 ±0.03
ΔC_{10}, m = 5	icos#15	$C_{25}H_{33}NO_7$	459.2257	458.2179	458.2170	26.02 ±0.05
C_{10}, n = 5	icos#16	$C_{25}H_{35}NO_7$	461.2414	460.2335	460.2350	26.73 ±0.04
ΔC_{11}, m = 6	icos#17	$C_{26}H_{35}NO_7$	473.2414	472.2335	472.2325	27.45 ±0.03
C_{11}, n = 6	icos#18	$C_{26}H_{37}NO_7$	475.2570	474.2492	474.2490	28.21 ±0.04

* confirmed using synthetic standards

Table A.5: HPLC-ESI-MS data of glucosyl ascaroside esters (glas).

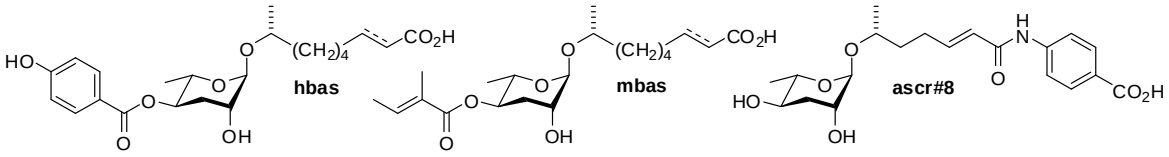




Side chain length (n, m)	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* $[M+Cl]^-$ calculated	*m/z* $[M+Cl]^-$ observed	Retention time [min] ±SD
C_7, n = 2	glas#1	$C_{19}H_{32}O_{11}$	438.2101	473.1795	473.1803	11.99 ±0.04
ΔC_9, m = 4	glas#3	$C_{21}H_{36}O_{11}$	464.2258	499.1952	499.1932	14.64 ±0.03
C_9, n = 4	glas#10*[1]	$C_{21}H_{38}O_{11}$	466.2414	501.2108	501.2112	15.05 ±0.03
C_{10}, n = 5	glas#16	$C_{22}H_{40}O_{11}$	480.2571	515.2265	515.2269	16.19 ±0.05
C_{11}, n = 6	glas#18	$C_{23}H_{42}O_{11}$	494.2727	529.2421	529.2402	17.34 ±0.04
C_{12}, n = 7	glas#20	$C_{24}H_{44}O_{11}$	508.2884	543.2578	543.2551	18.41 ±0.04
C_{13}, n = 8	glas#22	$C_{25}H_{46}O_{11}$	522.3040	557.2734	557.2720	19.49 ±0.05
C_{14}, n = 9	glas#24	$C_{26}H_{48}O_{11}$	536.3197	571.2891	571.2896	20.59 ±0.04
C_{15}, n = 10	glas#26	$C_{27}H_{50}O_{11}$	550.3353	585.3047	585.3095	21.77 ±0.04

* confirmed using synthetic standards

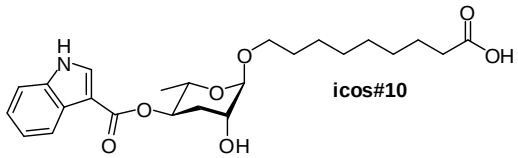
Table A.6: HPLC-ESI-MS data of ascr#8, 4-(4-hydroxybenzoyl)- and 4-(2-(*E*)-methyl-2-butenoyl)-ascarosides (hbas and mbas).



Side chain length	SMID[2]	Molecular formula	Molecular weight [amu]	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ observed	Retention time [min] ±SD
ΔC_7	ascr#8*[4]	$C_{20}H_{27}NO_7$	393.1788	392.1709	392.1712	16.77 ±0.04
ΔC_9	hbas#3*[1]	$C_{22}H_{30}O_8$	422.1941	421.1862	421.1866	22.41 ±0.03
C_9	hbas#10	$C_{22}H_{32}O_8$	424.2097	423.2019	423.2018	22.94 ±0.04
ΔC_9	mbas#3*[1]	$C_{20}H_{32}O_7$	384.2148	383.2070	383.2079	25.66 ±0.04
C_9	mbas#10	$C_{20}H_{34}O_7$	386.2305	385.2226	385.2239	26.38 ±0.04

* confirmed using synthetic standards

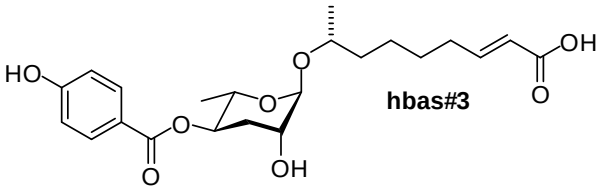
Table A.7: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **icos#10** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_2HOD$) = 49.05 ppm.



icos#10

Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Coupling [Hz]	Key HMBC correlations
1	176.4			
2	35.5	2.26	$J_{2,3}$ = 7.5	C-1, C-3, C-4
3	26.1	1.62		
4-5	30.1	1.38		
6	21.0	1.38		
7	27.1	1.43		
8	30.4	1.65		
9	68.2	3.49, 3.77	$J_{9,9}$ = 9.7, $J_{9,8}$ = 6.5	C-1′
1′	100.2	4.60		C-3′, C-9
2′	68.8	3.85		
3′	33.2	2.02 (*ax*)	$J_{3'ax,3'eq}$ = 13.1, $J_{3'ax,4'}$ = 11.3, $J_{2',3'ax}$ = 3.0	C-4′, C-5′
		2.21 (*eq*)	$J_{3'ax,3'eq}$ = 12.9, $J_{3'eq,4'}$ = 4.0	C-1′, C-4′, C-5′
4′	70.3	5.13	$J_{4',5'}$ = 9.8	C-5′, C-6′, C3″
5′	68.2	4.00	$J_{5',6'}$ = 6.3	
6′	17.8	1.25		C4′, C-5′
2″	133.1	7.96		C-3″, C3a″, C7a″
3″	108.1			
3″-CO	166.0			
3a″	137.9			
4″	121.5	8.02		C5″, C3a″
5″	123.4	7.20		C4″, C3a″
6″	122.3	7.20		C7″, C7a″
7″	112.7	7.45		C6″, C7a″
7a″	127.0			

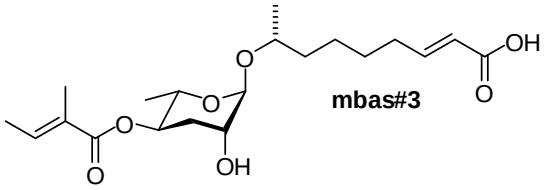
Table A.8*:* ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **hbas#3** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_2HOD$) = 49.05 ppm.



hbas#3

Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Coupling [Hz]	Key HMBC correlations
1	171.7			
2	124.4	5.83	$J_{2,3}$ = 15.6	C-4
3	149.0	6.90	$J_{3,4}$ = 7.0	C-1, C-4, C-5
4	33.1	2.26		C-2, C-3, C5, C-6
5, 6	29.2, 26.5	1.40 – 1.54		C-4, C-6
7	38.1	1.53, 1.62		C-5, C-6, C-8
8	72.7	3.83		C-6, C-7, C-9
9	19.4	1.16	$J_{8,9}$ = 6.1	C-7, C-8
1′	97.5	4.73		C-3′, C-5′, C-8
2′	69.5	3.77		C-4′
3′	33.1	1.95 (*ax*)	$J_{3'ax,3'eq}$ = 12.9, $J_{3'ax,4'}$ = 11.2, $J_{2',3'ax}$ = 2.9	C-4′, C-5′
		2.15 (*eq*)	$J_{2',3'eq}$ = 3.2, $J_{3'eq,4'}$ = 4.7	C-2′, C-4′, C-5′
4′	71.5	5.05	$J_{4',5'}$ = 9.6	C-5′, C-6′, C-7″
5′	68.4	3.98	$J_{5',6'}$ = 6.3	C-4′, C-6′
6′	18.1	1.18		C4′, C-5′
7″-COO	167.2			
1″	122.1			
2″,6″	132.8	7.85	J = 8.9	C-1″,C-3″,5″,C-4″,C-7″
3″,5″	116.2	6.83		C-2″,6″, C-4″
4″	163.6			

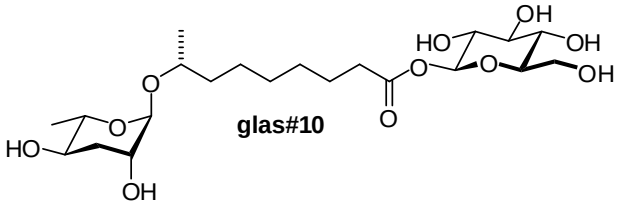
Table A.9: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **mbas#3** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_2HOD$) = 49.05 ppm.



mbas#3

Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Coupling [Hz]	Key HMBC correlations
1	171.7			
2	124.4	5.82	$J_{2,3}$ = 15.6	
3	149.0	6.88	$J_{3,4}$ = 7.0	
4	33.0	2.23		C-5, C-6
5, 6	29.3, 26.4	1.42–1.55		
7	37.9	1.60 1.52		
8	72.4	3.81		
9	18.9	1.14	$J_{8,9}$ = 6.1	C-7, C-8
1′	97.4	4.70		C-3′, C-5′, C-8′
2′	69.3	3.73		
3′	32.8	1.85 (*ax*)	$J_{3'ax,3'eq}$ = 12.9, $J_{3'ax,4'}$ = 11.2, $J_{2',3'ax}$ = 2.9	
		2.07 (*eq*)	$J_{2',3'eq}$ = 3.2, $J_{3'eq,4'}$ = 4.7	
4′	71.4	4.90	$J_{4',5'}$ = 9.6	
5′	68.2	3.89	$J_{5',6'}$ = 6.1	
6′	17.8	1.13		C-4′, C-5′
1″-COO	168.5			
2″	129.2			
3″	138.8	6.85	$J_{3'',4''}$ = 7.0	
4″	14.2	1.81		C-2″, C-3″
5″	11.8	1.82		C-1″, C-2″, C-3″

Table A.10: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **glas#10** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_2HOD$) = 49.05 ppm.



glas#10

Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Coupling [Hz]	Key HMBC correlations
1	173.8			
2	34.7	2.41	$J_{2,2}$ = 15.4, $J_{2,3}$ = 7.8	C-1
3	25.5	1.64		C-1, C-2
4		1.37		
5, 6	29.8, 30.1	1.30-1.50		
7	26.6	1.45 – 1.55		
8	72.3	3.78		
9	19.2	1.12	$J_{8,9}$ = 6.1	C-8, C-7
1′	97.4	4.60		C-9, C-3′, C-5′
2′	69.8	3.71		
3′	35.8	1.77 (*ax*)	$J_{3'ax,3'eq}$ = 13.0, $J_{3'ax,4'}$ = 11.4, $J_{2',3'ax}$ = 2.9	
		1.95 (*eq*)	$J_{2',3'eq}$ = 3.2, $J_{3'eq,4'}$ = 4.7	
4′	68.2	3.52	$J_{4',5'}$ = 9.5	
5′	71.0	3.63	$J_{5',6'}$ = 6.3, $J_{5',4'}$ = 9.3	
6′	18.0	1.22		C-5′, C-4′
1″	95.4	5.47	$J_{1'',2''}$ = 8.1	C-1
2″	73.8	3.32	$J_{2'',3''}$ = 9.1	
3″	77.8	3.41	$J_{3'',4''}$ = 9.7	
4″	70.9	3.35	$J_{4'',5''}$ = 9.7	
5″	78.7	3.36		
6″	62.2	3.67	$J_{6'',6''}$ = 12.0, $J_{6'',5''}$ = 4.8	
		3.83	$J_{6'',6''}$ = 12.0, $J_{6'',5''}$ = 1.9	

REFERENCES

(1) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc* **2012**, *134*, 1817.

(2) www.smid-db.org

(3) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol.* **2012**, *10*, e1001237. Epub 2012 Jan 10.

(4) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc Natl Acad Sci U S A.* **2009**, *106*, 7708.

(5) Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K. *Nature.* **2005**, *433*, 541.

(6) Butcher, R. A.; Fujita, M.; Schroeder, F. C.; Clardy, J. *Nat Chem Biol.* **2007**, *3*, 420.

(7) Butcher, R. A.; Ragains, J. R.; Li, W.; Ruvkun, G.; Clardy, J.; Mak, H. Y. *Proc Natl Acad Sci U S A.* **2009**, *106*, 1875.

(8) Butcher, R. A.; Ragains, J. R.; Kim, E.; Clardy, J. *Proc Natl Acad Sci U S A.* **2008**, *105*, 14288.

(9) Butcher, R. A.; Ragains, J. R.; Clardy, J. *Organic letters* **2009**, *11*, 3100.

APPENDIX B

COMPLEX ARCHITECTURES DERIVED FROM MODULAR ASSEMBLY OF PRIMARY METABOLITES REGULATE DEVELOPMENT AND PHENOTYPIC PLASTICITY IN *P. PACIFICUS*

B.1. Materials and methods:

B.1.1. *Pristionchus pacificus* metabolite naming: All newly identified compounds are named with four letter **"**SMID"s (Small Molecule IDentifiers), e.g. "icas#3" or "ascr#10" or "npar#1". The SMID database (www.smid-db.org) is an electronic resource maintained by Frank C. Schroeder and Lukas Mueller at the Boyce Thompson Institute in collaboration with Paul Sternberg and WormBase (www.wormbase.org). This database catalogues newly identified nematode small molecules, assigns a unique four-letter SMID (a searchable, gene-style Small Molecule IDentifier), and for each compound includes a list of other names and abbreviations used in the literature. In this dissertation chapter,[1] the author introduces the following new four-letter SMIDs: pasc (**p**henylethanolamide **asc**aroside), ubas (3-**u**reido iso**b**utyrate **as**caroside), dasc (**d**imeric **asc**aroside), part (**par**a**t**oside), and npar (**n**ucleoside-based **par**atoside).

B.1.2. Analytical instrumentation: NMR spectra were recorded on a Varian INOVA-600 (600 MHz for ^{1}H, 151 MHz for ^{13}C), INOVA-500 (500 MHz for ^{1}H and 125 MHz for ^{13}C), and INOVA-400 (400 MHz for ^{1}H, 100 MHz for ^{13}C) instruments. HPLC-MS,

MS/MS, and single-ion monitoring (SIM-LCMS) was performed using an Agilent 1100 Series HPLC system equipped with a diode array detector and connected to a Quattro II spectrometer (Micromass/Waters). High resolution mass spectra were acquired using a Xevo G2 QTOF mass spectrometer. Flash chromatography was performed using a Teledyne ISCO CombiFlash system. HPLC fractionation was performed using an Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB-C18 column (9.4 x 250 mm, 5 µm particle diameter) coupled to a Teledyne ISCO Foxy 200 fraction collector.

B.1.3. *P. pacificus* strains and culture conditions: The following *P. pacificus* strains were used for this study: (1) exo-metabolome preparation: RS2333, (2) dauer formation assay: RS5134, (3) mouth-form dimorphism assay: RSB020, (4) exo- and endo-metabolome preparation for natural variation of small molecule production: RS2333, RS5205, RS5134, RS5380, RS5399, and RSB020.

Plates and liquid cultures of worms were prepared as described previously.[2] For axenic cultures, *P. pacificus* (RS2333) gravid adults from ten 10 cm plates were washed with M9 buffer and treated with alkaline hypochlorite solution to isolate eggs.[3] Isolated eggs were washed thoroughly with M9 buffer and allowed to hatch in fresh sterile M9 for 24 h. The resulting synchronized *J2* larvae were transferred to the modified chemically defined growth medium described earlier[4,5] and allowed to grow for 26 days at 20 °C and 80 rpm. After 26 days the population consisted of mostly of gravid adults and large numbers of *J2* larvae.

B.1.4. Preparation of metabolome extracts and preliminary fractionation: 3 L culture of *P. pacificus* strain RS2333 was filtered and centrifuged at 10,000 rpm for 10 min to separate the supernatant and worm pellets. The culture supernatant was applied to a C18 column (Chromabond, Macherey Nagel), which was followed by elution with 50% MeOH in H_2O. This step removed strongly lipophilic components (e.g. triglycerides, long-chain fatty acids) but did not reduce bioactivity. The eluate was evaporated and resuspended with mixture of chloroform and methanol (2:1). The sample was applied to a SiOH column (Chromabond, Macherey Nagel) equilibrated with chloroform/methanol (2:1). The column was washed with chloroform/methanol (2:1, Fraction I), chloroform/methanol (1:5, Fraction II), and chloroform/methanol/water (6:10:1, Fraction III). Fraction II showed the most activity in subsequent dauer formation assays (see **Appendix Section B.1.7**) and was used for 2D NMR spectroscopic analysis.

Several additional 1 L-batches of culture supernatant were prepared to obtain larger quantities of the compounds detected by 2D NMR and studying natural variation of small molecule production in *P. pacificus* (see **Appendix Section B.1.3**). These cultures were harvested, centrifuged, and the resultant supernatant media and worm pellets were frozen over dry ice-acetone slush and lyophilized separately. The lyophilized materials from the supernatant were extracted with 300 mL of 95% ethanol at room temperature for 16 h. The worm pellets were crushed with ~8 g of granular NaCl using a mortar pestle and extracted with 150 mL of 100% ethanol at room temperature for 16 h. The resulting suspensions were filtered and the filtrate evaporated *in vacuo* at room temperature, producing media metabolite (the worm "exo-metabolome") extracts and worm pellet metabolite ("endo-metabolome") extracts.

For high-resolution HPLC-MS analysis, 100 mL sample of unfractionated culture supernatant were lyophilized to a fine powder, which was subsequently extracted with 50 mL of 95% ethanol for 16 h. The extract was concentrated in vacuo, resuspended in 150 µL of methanol, filtered, and used for HPLC-MS. For bacterial control experiments, 1 L of *E. coli* OP50 bacteria culture grown overnight was lyophilized and extracted as described above.

For the analysis of *P. pacificus* axenic cultures, the culture was centrifuged at the end of the 26-day incubation period, and the supernatant was lyophilized and extracted with 50 mL of methanol. To remove the large amounts of glucose contained in the axenic medium, the extract was loaded onto 8 g of ethyl acetate-washed Celite® and filtered over a RediSep *Rf* GOLD 30 g HP C18 reverse-phase column using a water-methanol solvent gradient, starting with 15 min of 98% water, followed by a linear increase of methanol content up to 100% at 60 min. The first 300 mL of eluate contained mostly glucose and were discarded. The remainder of the eluate was concentrated *in vacuo*. The resulting extract was resuspended in 100 µL methanol, filtered, and analyzed by selective ion monitoring (SIM)-LCMS.

B.1.5. 2D NMR spectroscopic analyses: Non-gradient phase-cycled dqfCOSY spectra were acquired using the following parameters: 0.8 s acquisition time, 400-600 complex increments, 8-32 scans per increment. dqfCOSY spectra were zero-filled to 8k × 4k and a cosine bell-shaped window function was applied in both dimensions before Fourier transformation. Gradient and non-gradient HSQCAD and HMBC spectra were acquired using 0.25 s acquisition time and 256-500 complex increments. NMR spectra were

processed using Varian VNMR, MestreLabs' MestReC, and MestRecNova software packages.

B.1.6. HPLC protocol, HPLC-MS/MS, and SIM-LC-MS analyses: HPLC-MS was performed using an Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB-C18 column (9.4 x 250 mm, 5 µm particle diameter) connected to a Quattro II spectrometer (Micromass/Waters) using a 10:1 split. A 0.1% acetic acid-acetonitrile solvent gradient was used at a flow rate of 3.6 mL/min, starting with an acetonitrile content of 5% for 5 min which was increased to 100% over a period of 40 min.

Exo-metabolome fractions and samples for studying natural variation in small molecule production were analyzed by HPLC-ESI-MS in negative and positive ion modes using a capillary voltage of 4.0 kV and a cone voltage of -40 V and +20 V respectively. HPLC-MS/MS screening for precursor ions of *m/z* = 73.0 (negative mode) performed using argon as collision gas at 2.1 mtorr and 40 eV. The HPLC protocol mentioned in this section is also used for enrichment of minor components of the *P. pacificus* exo-metabolome and isolation of synthetic compounds.

For the analysis of exo-metabolome samples from *P. pacificus* axenic cultures, the spectrometer was operated in selective ion monitoring (SIM) mode, and the following ions were selectively observed: *m/z* = 247 (ascr#9, part#9), 466 (pasc#9), 533 (dasc#1), 605 (ubas#1), and 641 (npar#1).

B.1.7. Dauer formation assay: Dauer formation assays were performed by Dr. Akira Ogawa, Sommer group as described previously[2] using heat or kanamycin-killed *E. coli*

OP50 and *P. pacificus* strain RS5134. Briefly, synthetic compounds were dissolved in ethanol (0.5 mM, stock solution) and combined with water to make a 100 μL solution and subsequently added to 3 mL NGM agar without peptone (3, 1, 0.3, 0.1 μM final concentrations). The dauer formation assay was conducted in triplicate for each compound and concentration. 60-100 worms were screened for each condition.

B.1.8. Mouth-form dimorphism assay: Mouth-form dimorphism assays were performed by Dr. Erik J. Ragsdale, Sommer group using *P. pacificus* strain RSB020. The synthetic compounds dissolved in ethanol (0.5 mM) were diluted with water to make 100 μL solution and subsequently added to 3 mL NGM-agar (1 μM final concentration). The assay plates were seeded with 50 μL OP50 culture in LB medium and incubated overnight at 20 °C to allow bacterial growth. Each replicate included the progeny of two mothers, which were picked as adult hermaphrodites of a consistent age (carrying 4-6 eggs) and which were all from the same *P. pacificus* culture plate. Following placement of mothers on assay plates, plates were kept at 20 °C for six days, such that the entire broods were adults at the time of screening. A random sample of 50 hermaphrodite progeny was screened per plate. All animals were screened by differential interference contrast (DIC) microscopy on a Zeiss Axioskop. The following discrete characters were used to discriminate eurystomatous from stenostomatous individuals, respectively: (1) claw-shaped vs. flint-shaped (i.e. dorsoventrally symmetrical) dorsal tooth; (2) presence vs. absence of a subventral tooth; (3) strongly vs. weakly sclerotized stomatal walls. No intermediate mouth-forms were observed. Experiments were conducted in triplicate for each treatment.

In assays of responses to compounds of several concentrations (1, 0.3, 0.1, 0.03, 0.01 μM final concentrations), experiments were performed as described above, with the following modifications. To allow greater resolution of responses to lower concentrations, 60 randomly screened individuals in each of five replicates per concentration per compound were assayed. All concentration-curve experiments were performed in parallel using mothers picked randomly from the same two source populations. To prepare large numbers of individual mothers for these experiments, source populations were established from virgin hermaphrodites to constrain the presence of males and were conditioned to well-fed and ambient conditions for at least one generation before mothers were picked for the assays.

B.1.9. Statistical analyses: Error bars represent a 95% confidence interval in **Figure 3.8a-d** calculated using a binomial test on the total count data. All experiments were conducted in triplicate (or in five replicates for mouth-form concentration-curve assays) for each treatment. Significant differences (*p<0.01 and **p<0.001) between each chemical treatment and the control (EtOH) treatment in **Figure 3.8a,b** were determined using Fisher's exact test in the program R.

B.2. Figures:

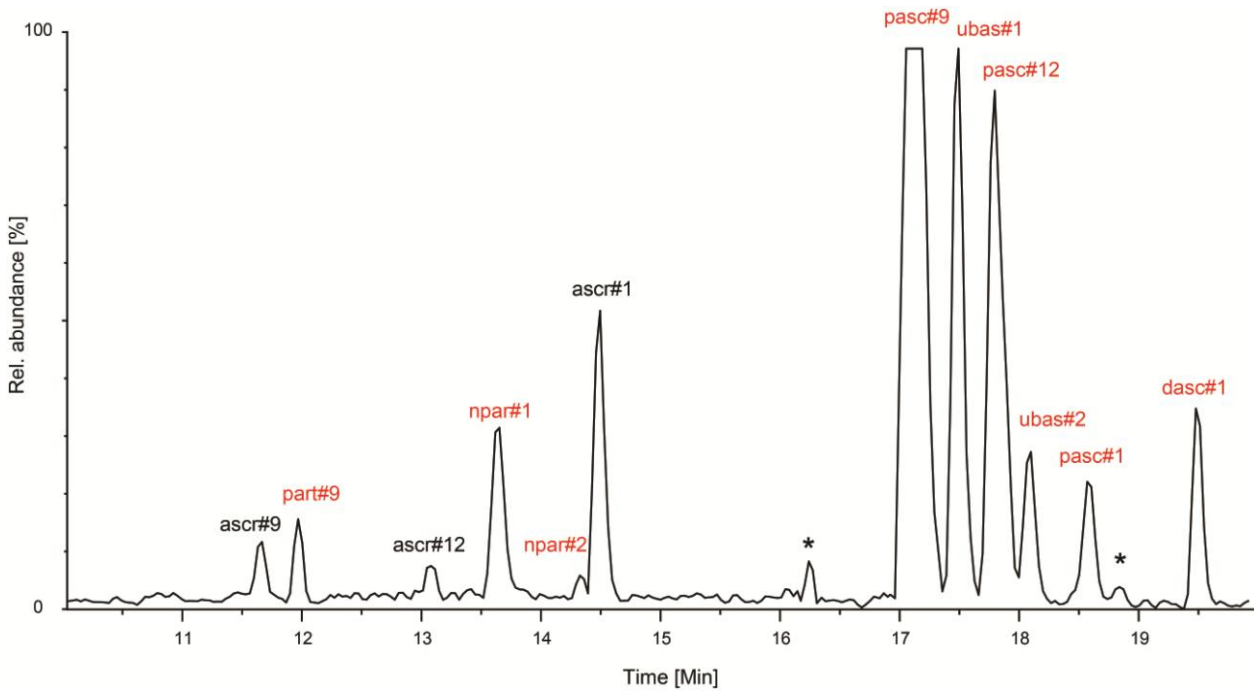


Figure B.1: HPLC-MS/MS screen (precursors of *m/z* 73, **Chapter 2**, **Section 2.1**) of *P. pacificus* RS2333 exo-metabolome reveals known ascarosides (black), and several novel dideoxysugar-based compounds (red, ascarosides: pasc#9, pasc#12, pasc#1, ubas#1, ubas#2, and dasc#1; paratosides: part#9, npar#1, and npar#2) (* indicates signals from non-ascarosides). See also **Figures 3.3**, **3.4**, and **Appendix Table B.1**. This indicates that paratosides also produce the diagnostic fragment ion *m/z* 73.

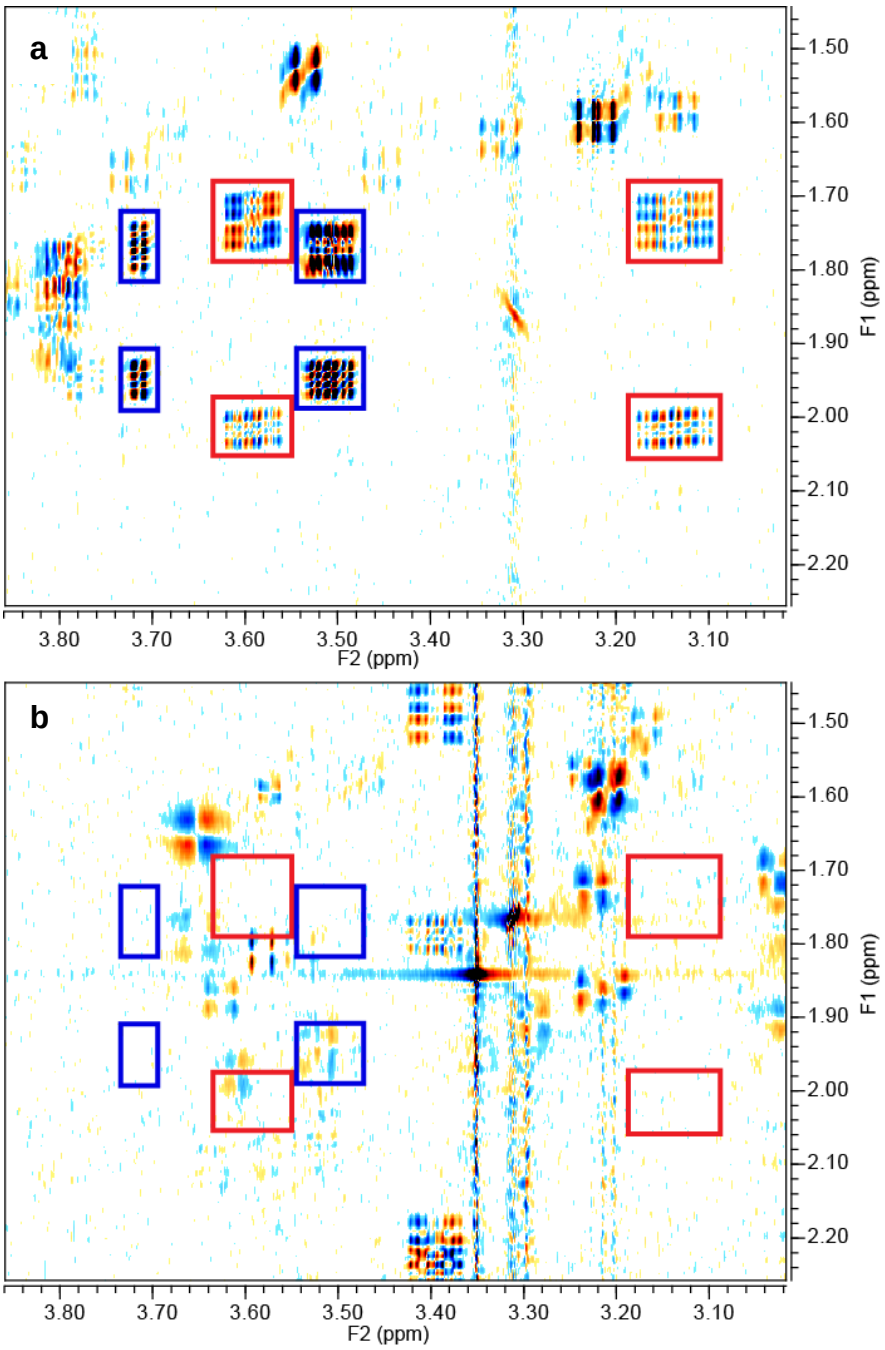


Figure B.2: Small molecule architectures identified from *P. pacificus* exo-metabolome are not of bacterial origin. Sections of dqfCOSY spectra (600 MHz, methanol-*d4*) of (**a**) *P. pacificus* exo-metabolome extract and (**b**) *E. coli* OP50 metabolome extract. Characteristic crosspeaks for ascarosides are boxed blue and that of paratosides are boxed red. Comparison of the two spectra indicates that the complex small molecules identified from *P. pacificus* exo-metabolome are not of bacterial origin. Correspondingly, HPLC-MS analyses of bacterial extracts did not show any of the peaks detected in *P. pacificus* exo-metabolome samples.

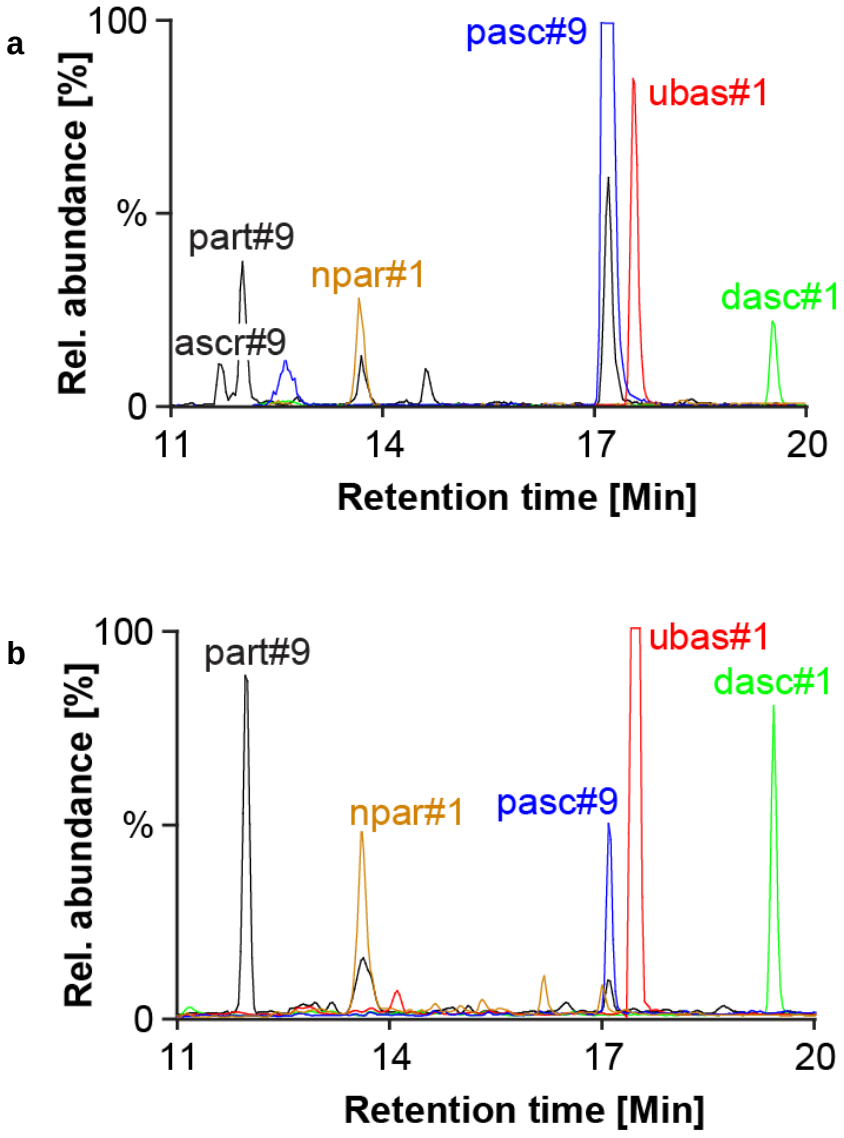


Figure B.3: (**a**) HPLC-MS analysis of exo-metabolome extract from *P. pacificus* cultures fed with *Pseudomonas sp.* and (**b**) SIM-LC-MS analysis of exo-metabolome extract from *P. pacificus* axenic cultures (see **Appendix B**, **Section B.1.6**).

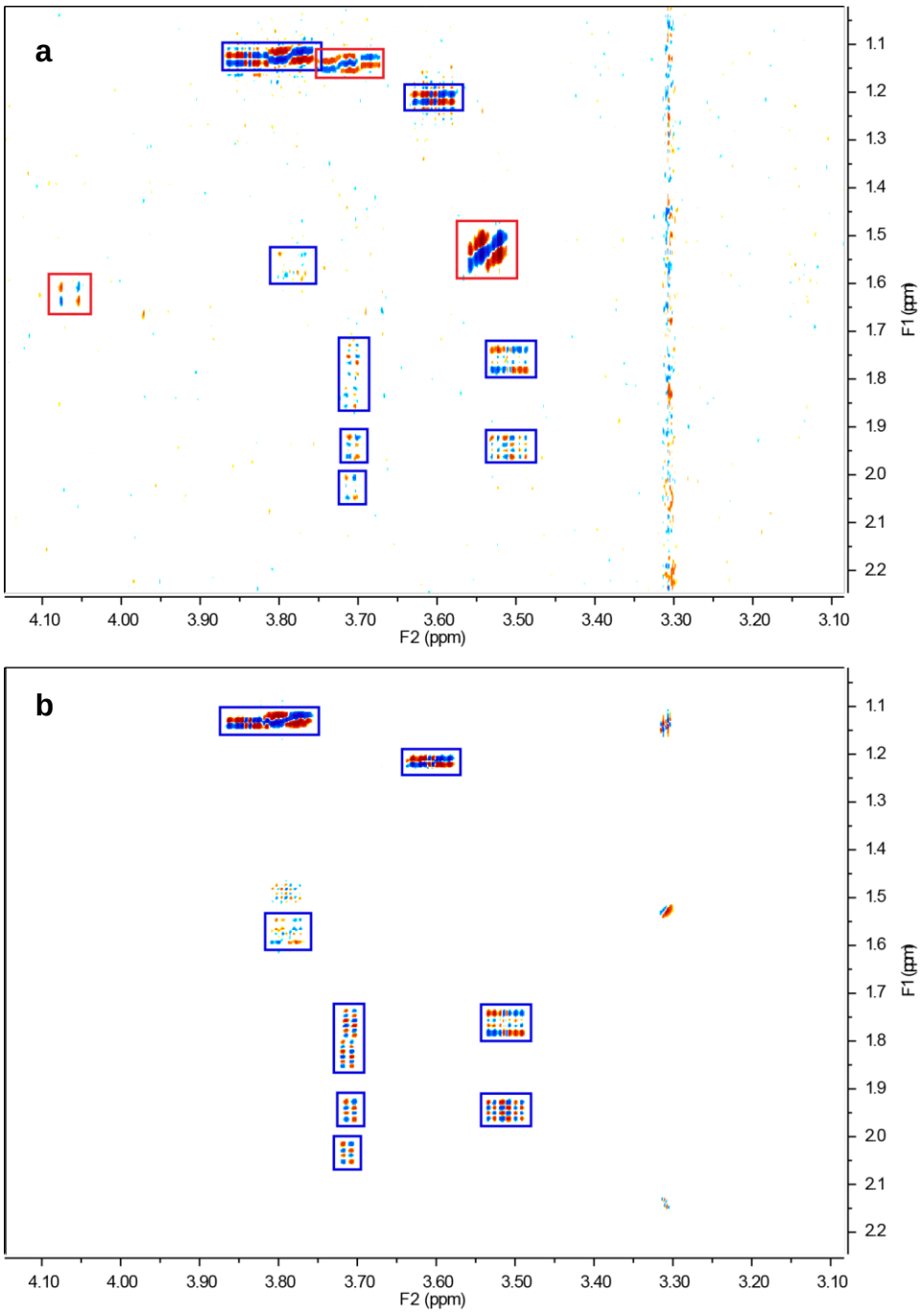


Figure B.4: Sections of dqfCOSY spectra (600 MHz, methanol-*d4*) confirming presence of dasc#1 in *P. pacificus* exo-metabolome. (**a**) HPLC-enriched *P. pacificus* exo-metabolome extract fraction containing dasc#1. (**b**) Synthetic sample of dasc#1. Characteristic crosspeaks for dasc#1 are boxed blue whereas unrelated crosspeaks from other metabolites present in the natural sample are boxed red. The precise match of crosspeaks between the natural and synthetic sample proves dasc#1 structural and stereochemical assignments.

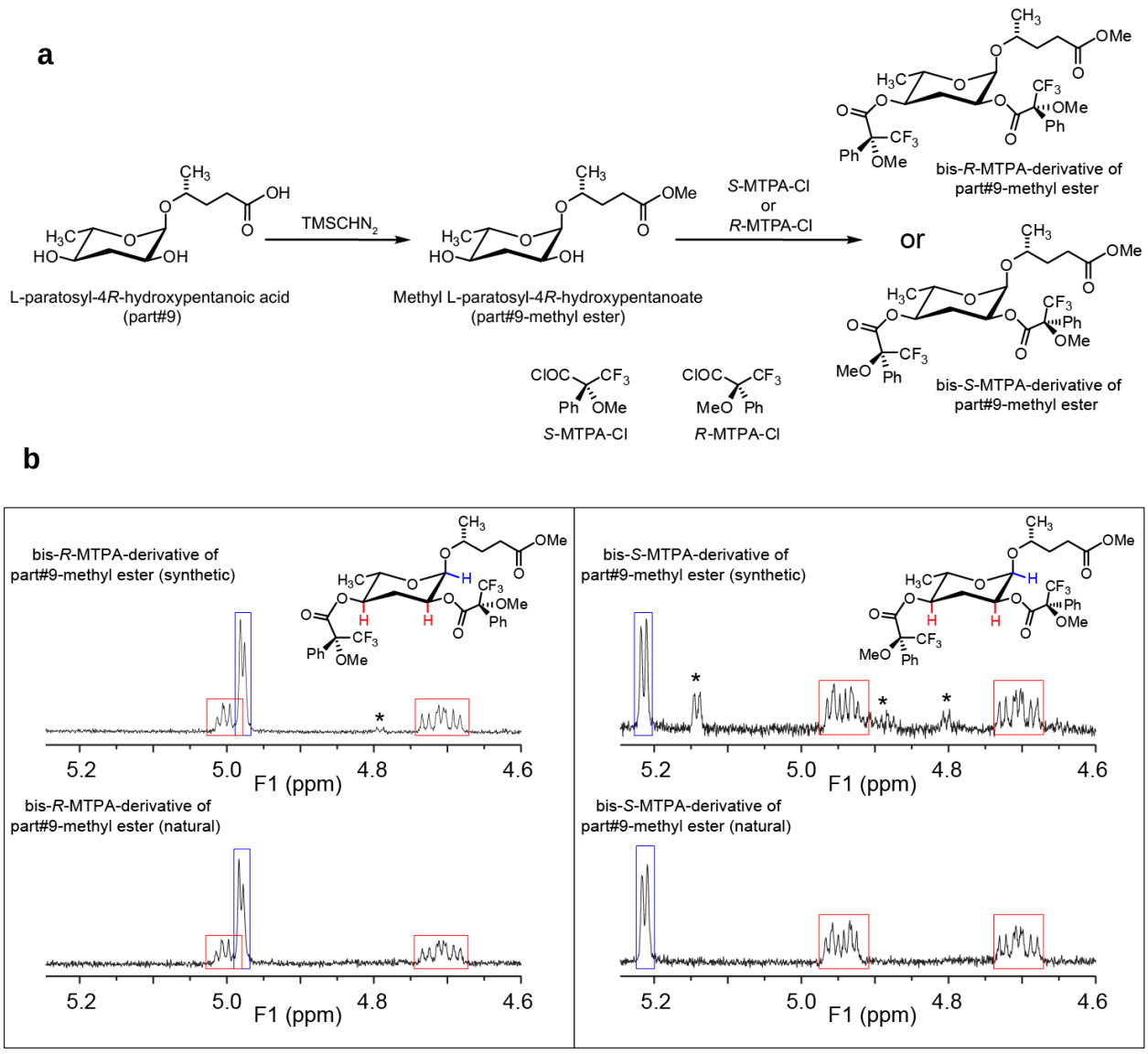


Figure B.5: Determination of absolute configuration of part#9. (**a**) Conversion of synthetic L-paratosyl-4*R*-hydroxypentanoic acid and isolated natural part#9 into the corresponding methyl esters, which were reacted with *S*- and *R*-α-methoxy-α-trifluoromethylphenylacetyl chlorides (Mosher's acid chlorides, *S*- and *R*-MTPA-Cl) to form the diasteromeric di-esters[1] following previously published reaction protocols.[6] (**b**) Comparison of ^{1}H-NMR spectra ($CDCl_3$, 600 MHz) of the derivatization products of natural and syntehtic part#9 etablish natural part#9 as L-paratosyl-4*R*-hydroxypentanoic acid. (*)s indicate peaks due to side products resulting from incomplete reaction of starting materials. For detailed reactions conditions see **Appendix Section B.3.3**.

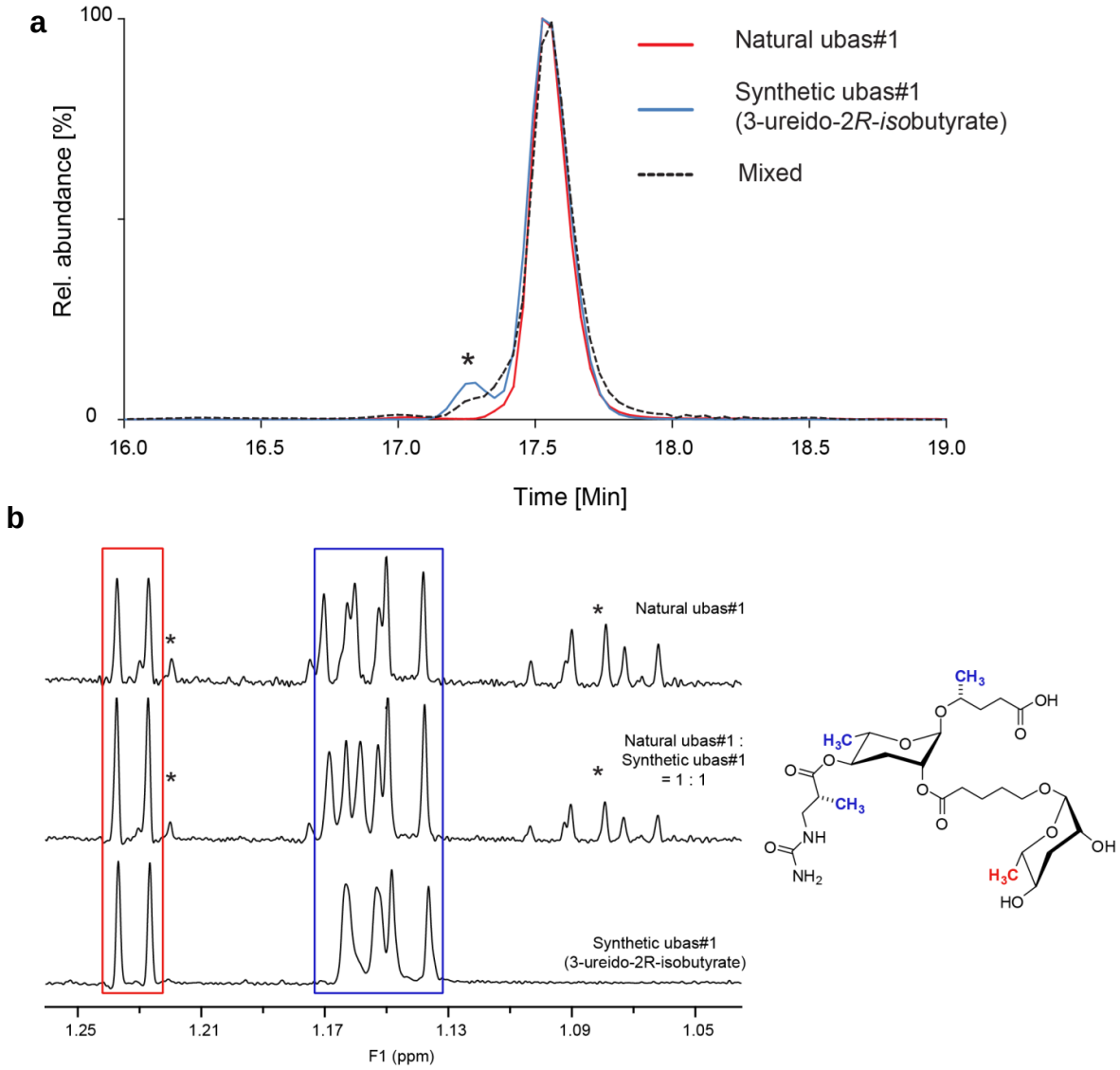


Figure B.6: Determination of stereochemistry of ubas#1. (**a**) Comparison of HPLC-MS retention times (ESI-, ion chromatogram for *m/z* = 605) of natural ubas#1 (red), a synthetic mixture of ubas#1 diastereomers containing the 3-ureido-2*R*-*iso*butyrate and 3-ureido-2*S*-*iso*butyrate in a ~95:5 ratio (blue), and a mixture of the natural and synthetic samples (dotted black). The HPLC-retention time of synthetic (3-ureido-2*R*-*iso*butyrate)-derived ubas#1 matches that of natural ubas#1 and is distinctly different from the (3-ureido-2*S*-*iso*butyrate)-derived ubas#1 diastereomer (marked *) indicating that natural ubas#1 includes a 3-ureido-2*R*-*iso*butyrate moiety. (**b**) Comparison of sections of ^{1}H-NMR spectra of synthetic (3-ureido-2*R*-*iso*butyrate)-derived ubas#1 (bottom), a natural sample containing ubas#1 (top), and a 1:1 mixture of these two samples (middle) shows that the relative intensity of the four characteristic methyl doublets (indicated by the red and blue boxes in the accompanying structure) increases upon adding synthetic ubas#1 to the natural sample (unrelated peaks in the natural sample are marked *). This confirms that natural ubas#1 contains 3-ureido-2*R*-*iso*butyrate, and not 3-ureido-2*S*-*iso*butyrate. Differences in pH and concentrations between the natural and synthetic samples slightly affect chemical shifts of the methyl doublets, resulting in small changes of chemical shift values upon mixing of natural and synthetic sample.

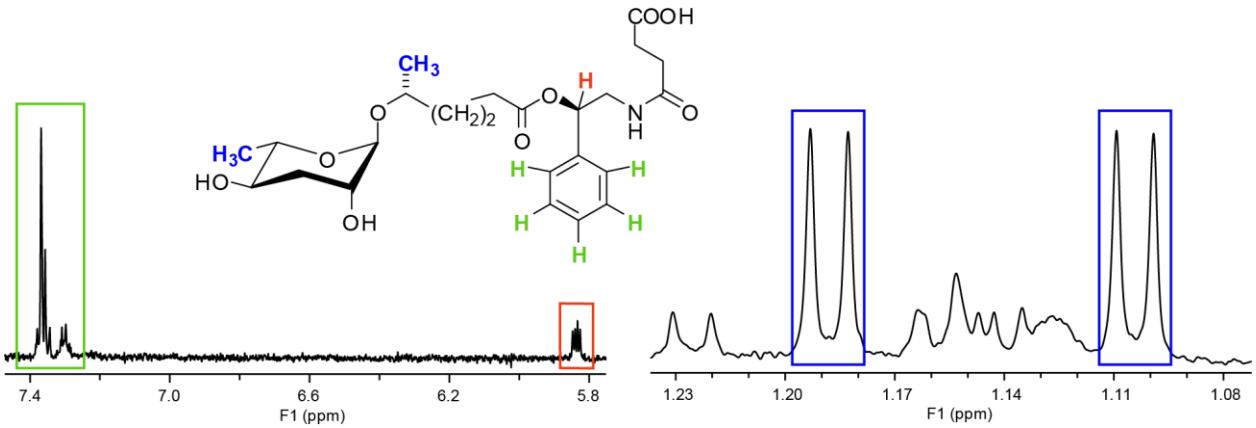


Figure B.7: Characteristic ^{1}H NMR signals for pasc#12 in HPLC-enriched *P. pacificus* exo-metabolome extract fraction.

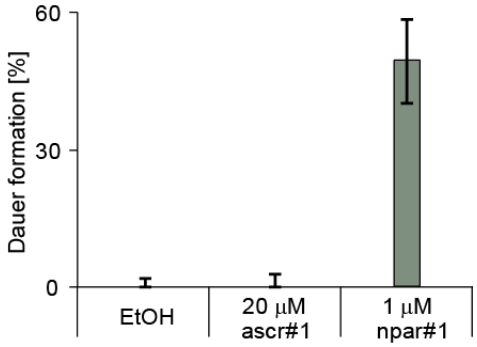


Figure B.8: ascr#1 is not active in dauer formation assays. ascr#1 does not induce dauer formation in *P. pacificus*, even at very high concentrations (20 µM). npar#1 (1 µM) was used as a positive control for *P. pacificus* dauer formation.

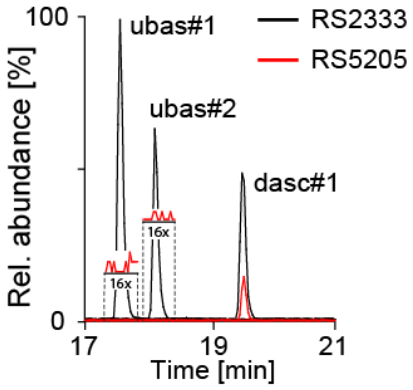


Figure B.9: RS5205 is incapable of biosynthesizing ubas#1 and ubas#2. HPLC-MS ion traces for ubas#1, ubas#2, and dasc#1 comparing the exo-metabolomes of RS2333 and RS5205 shows that RS5205 is incapable of producing ubas-compounds under conditions tested, whereas dasc#1 production is reduced as compared to RS2333.

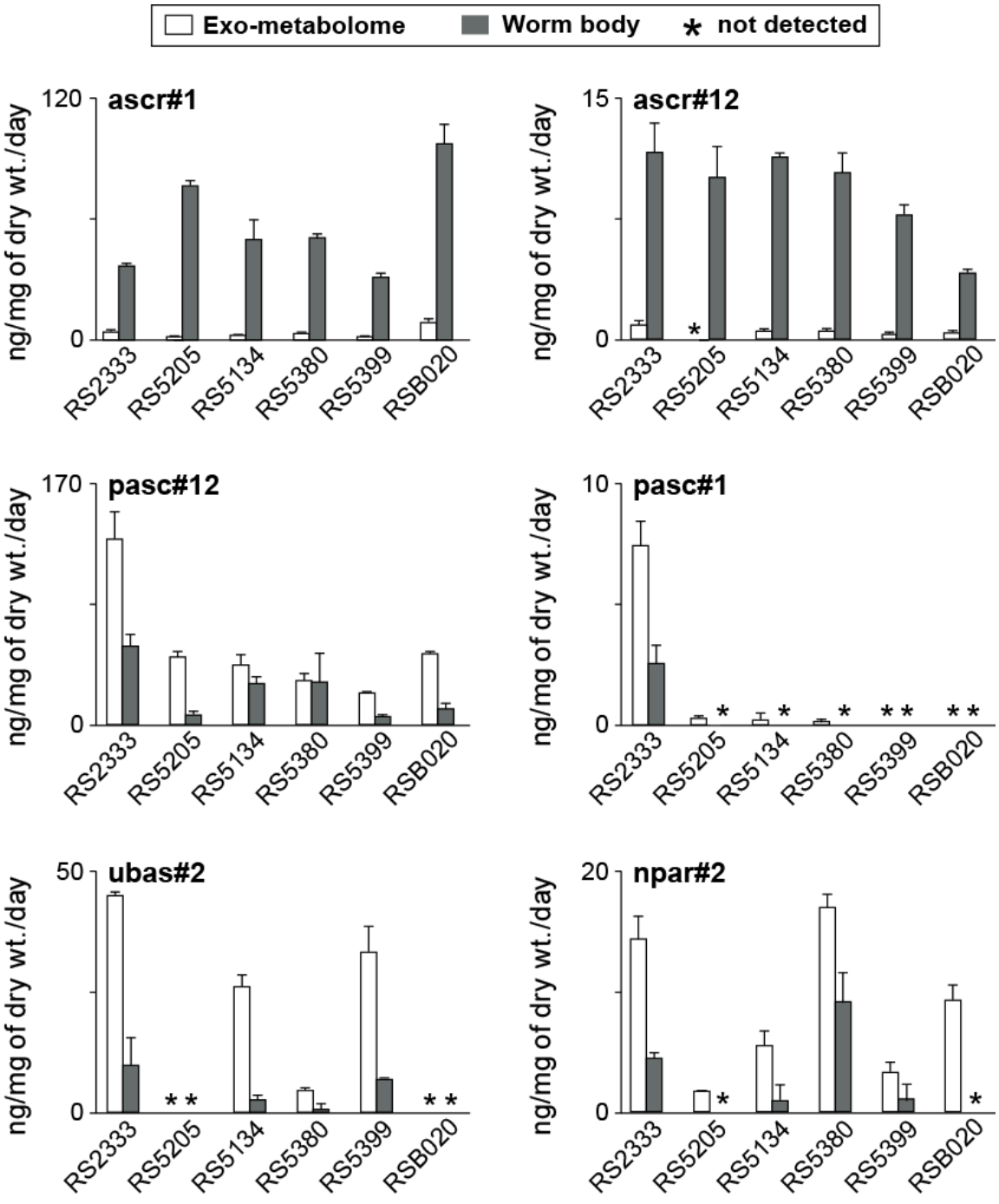


Figure B.10: Absolute comparison of representative small molecules in the exo- and endo-metabolomes of 6 *P. pacificus* wild isolates (also see **Figure 3.10a**).

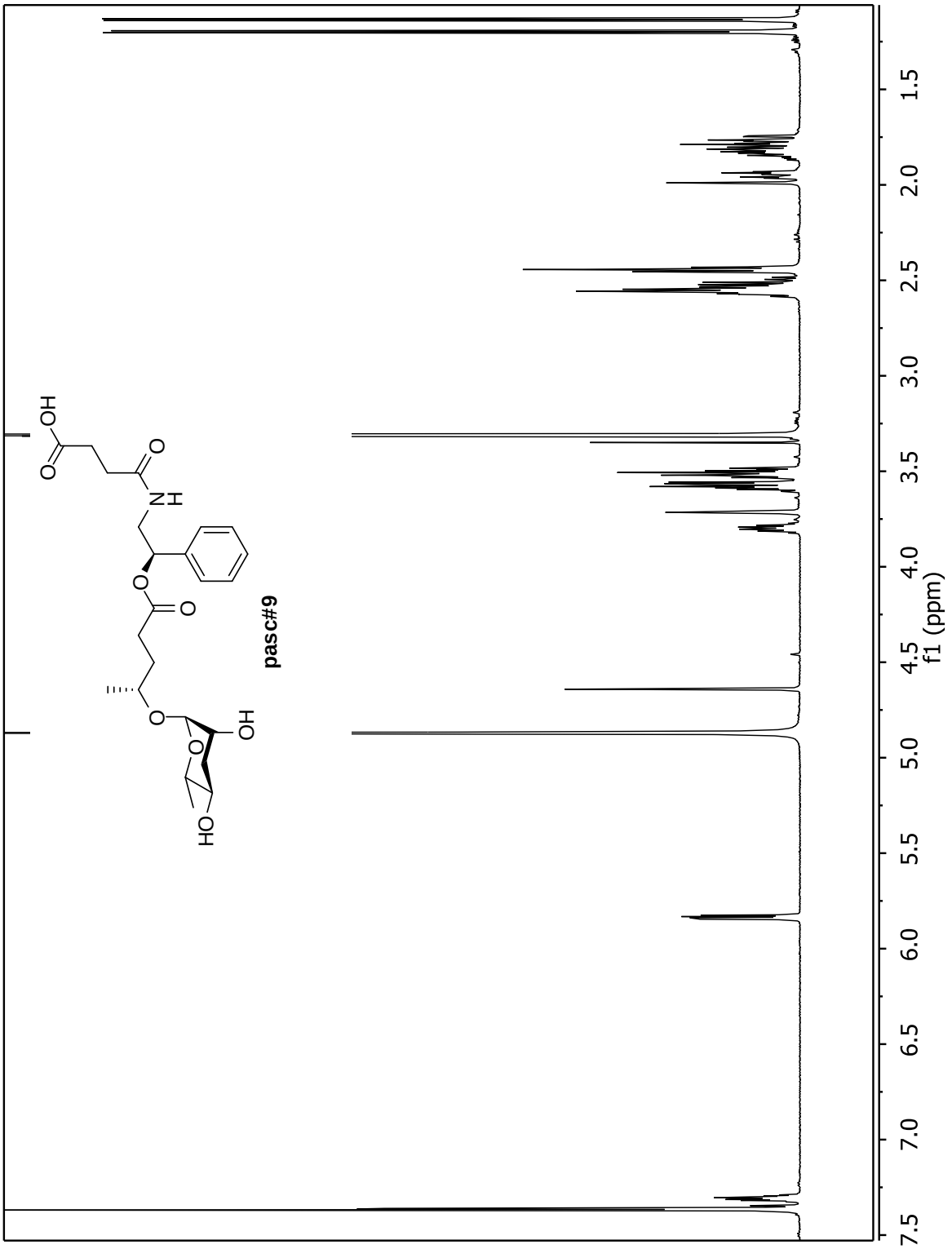


Figure B.11.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of 4-(((*R*)-2-(((*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-phenylethyl)amino)-4-oxo butanoic acid (**pasc#9**).

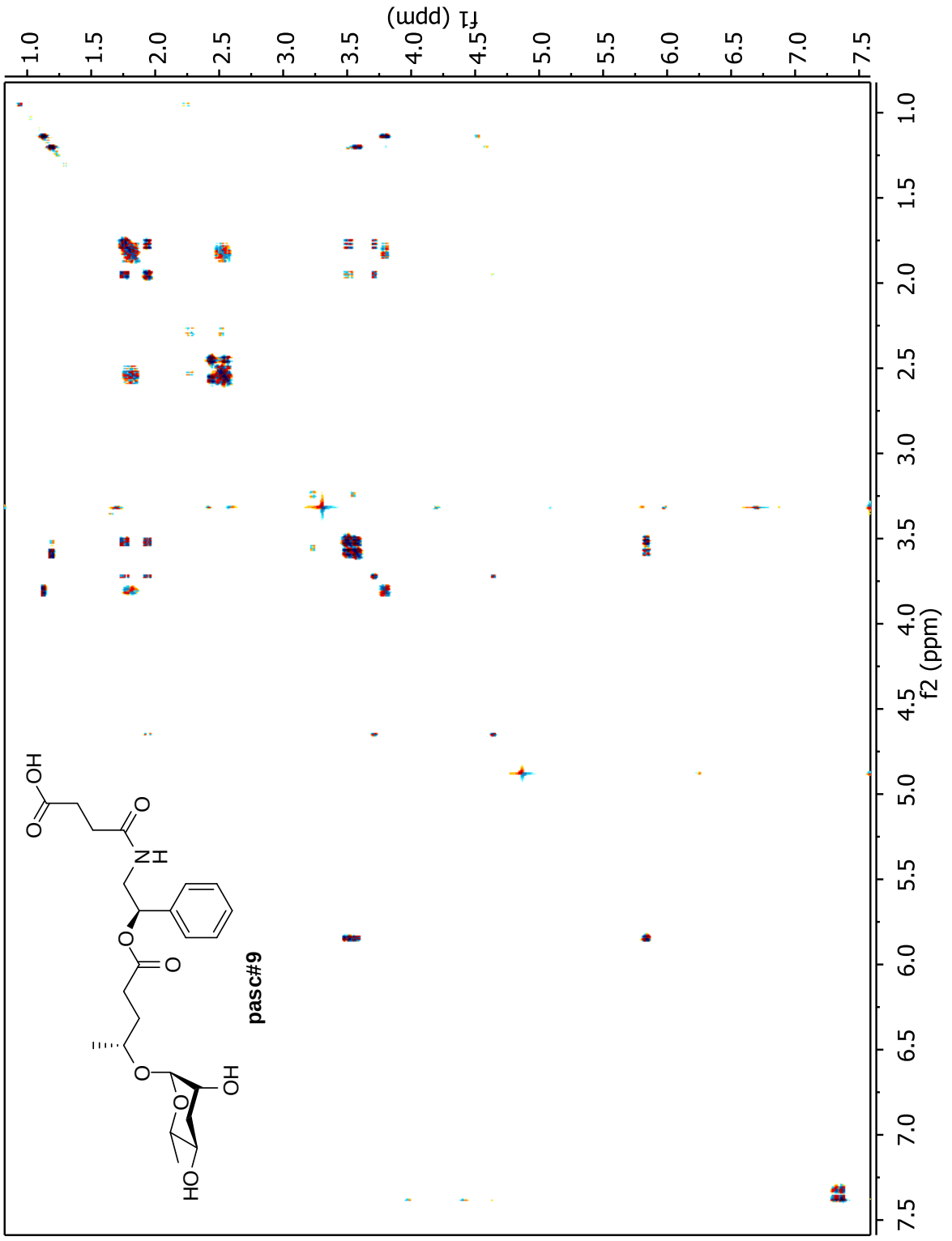


Figure B.11.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of 4-(((*R*)-2-(((*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-phenylethyl)amino)-4-oxo butanoic acid (**pasc#9**).

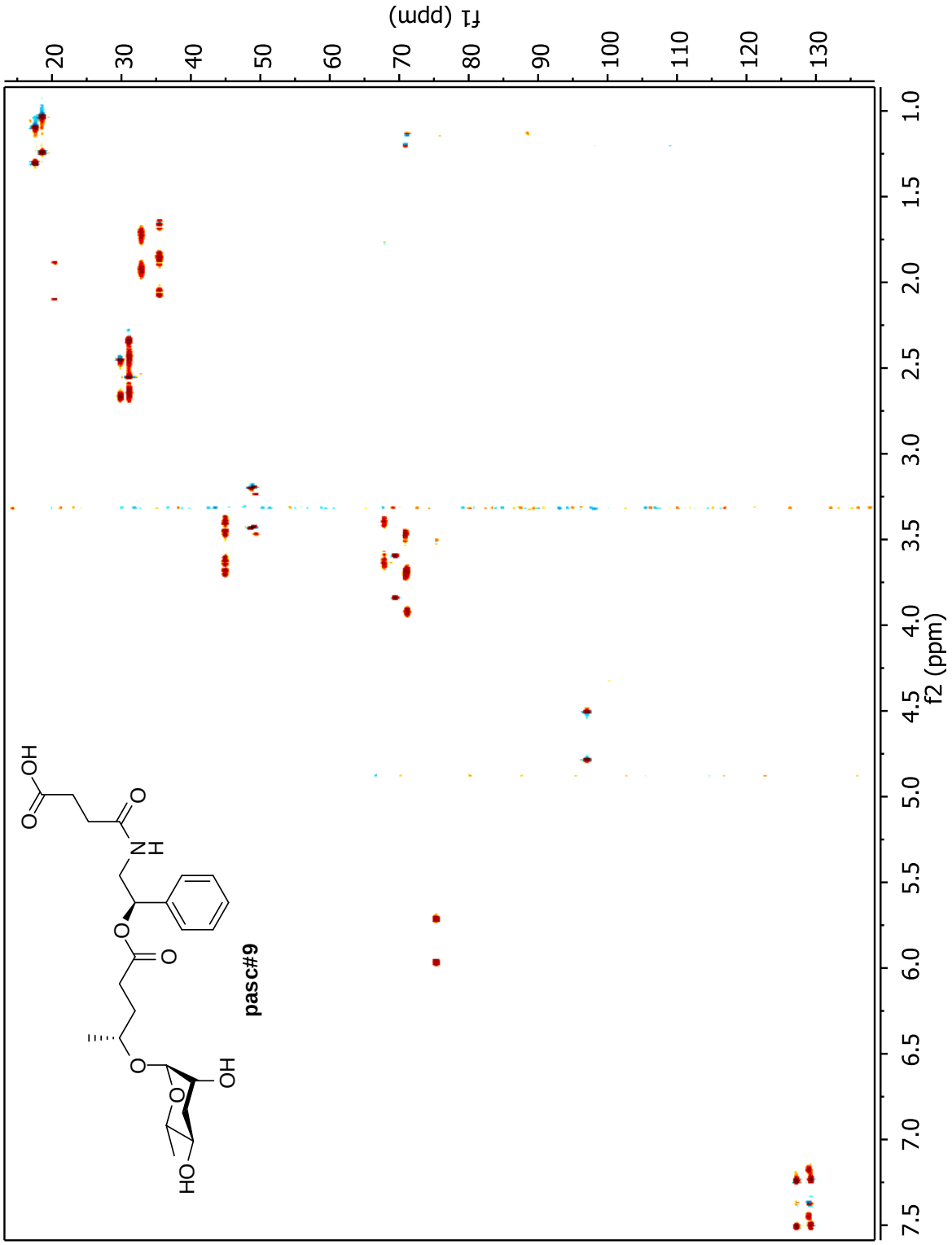


Figure B.11.3: HMQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of 4-(((*R*)-2-(((*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-phenylethyl)amino)-4-oxobutanoic acid (**pasc#9**).

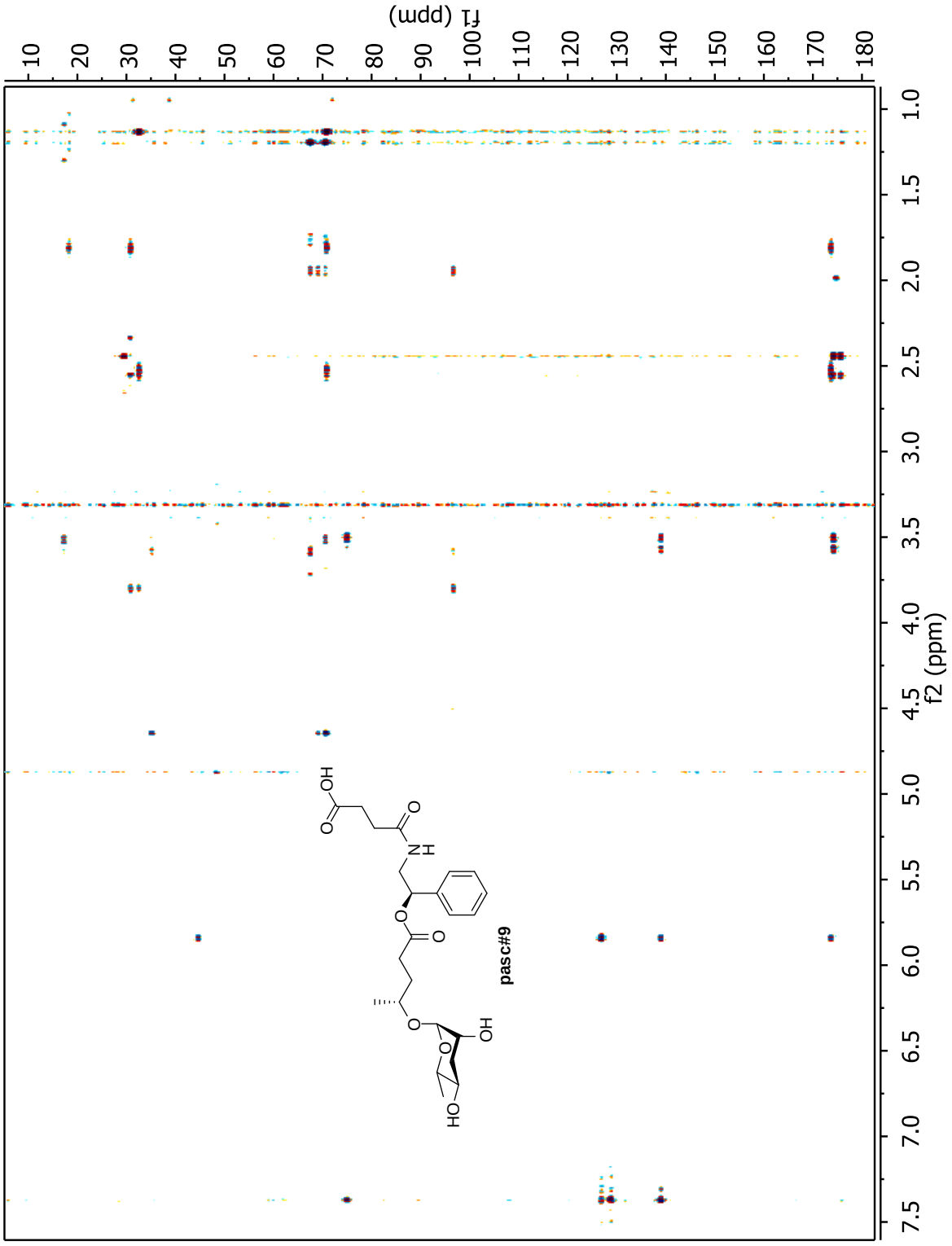


Figure B.11.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of 4-(((*R*)-2-(((*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy) pentanoyl)oxy)-2-phenylethyl)amino)-4-oxobutanoic acid (**pasc#9**).

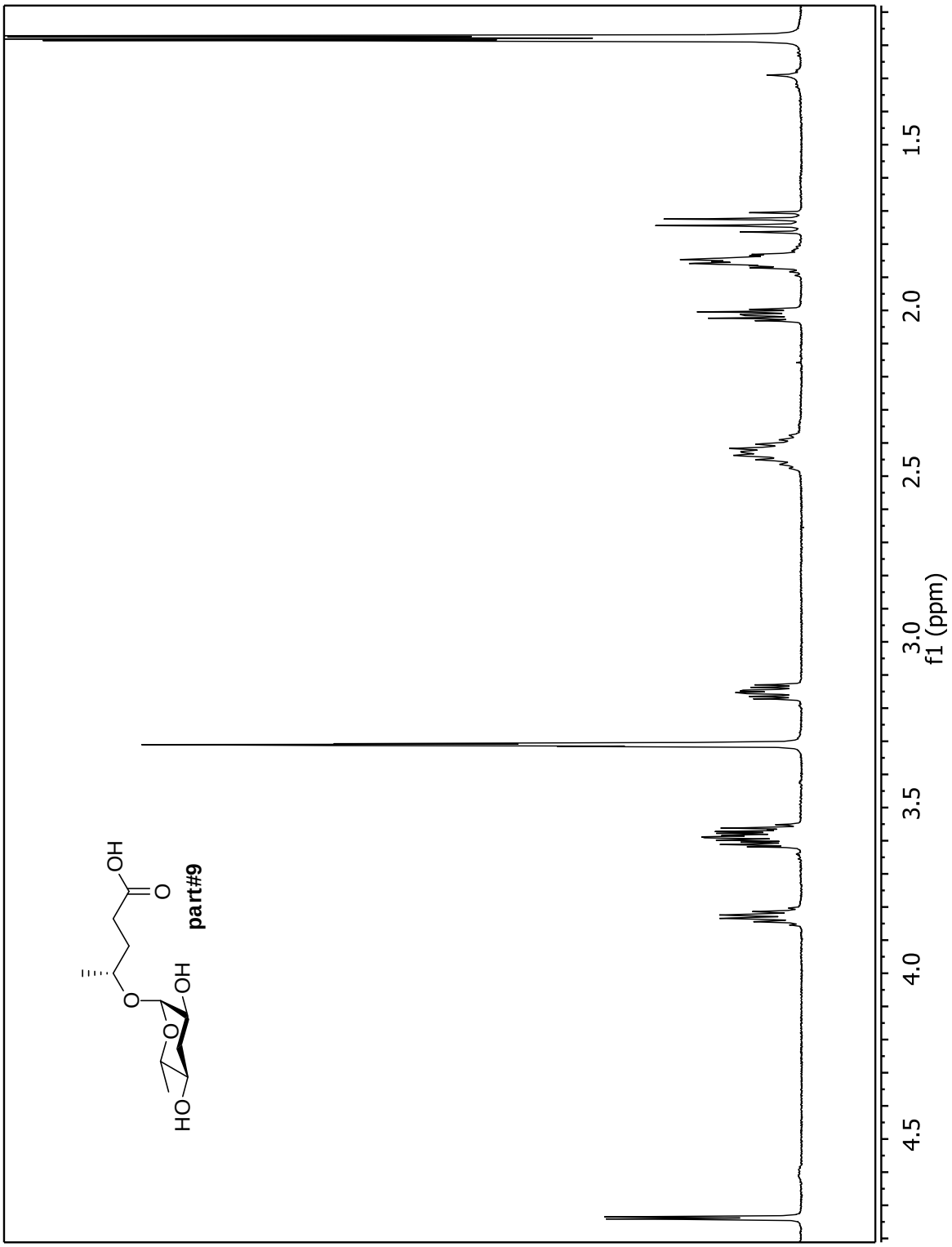


Figure B.12.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy -6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**part#9**).

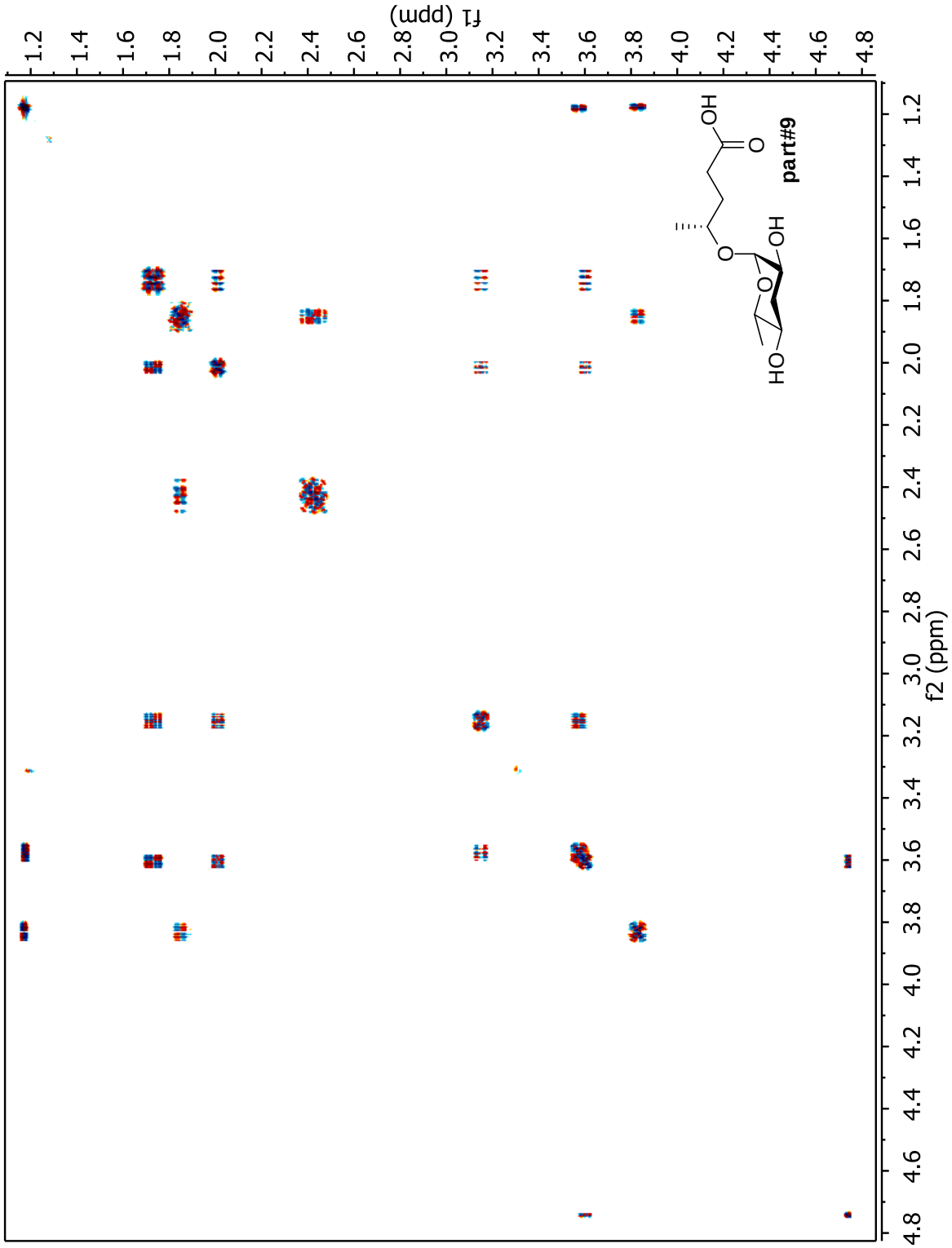


Figure B.12.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of (*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**part#9**).

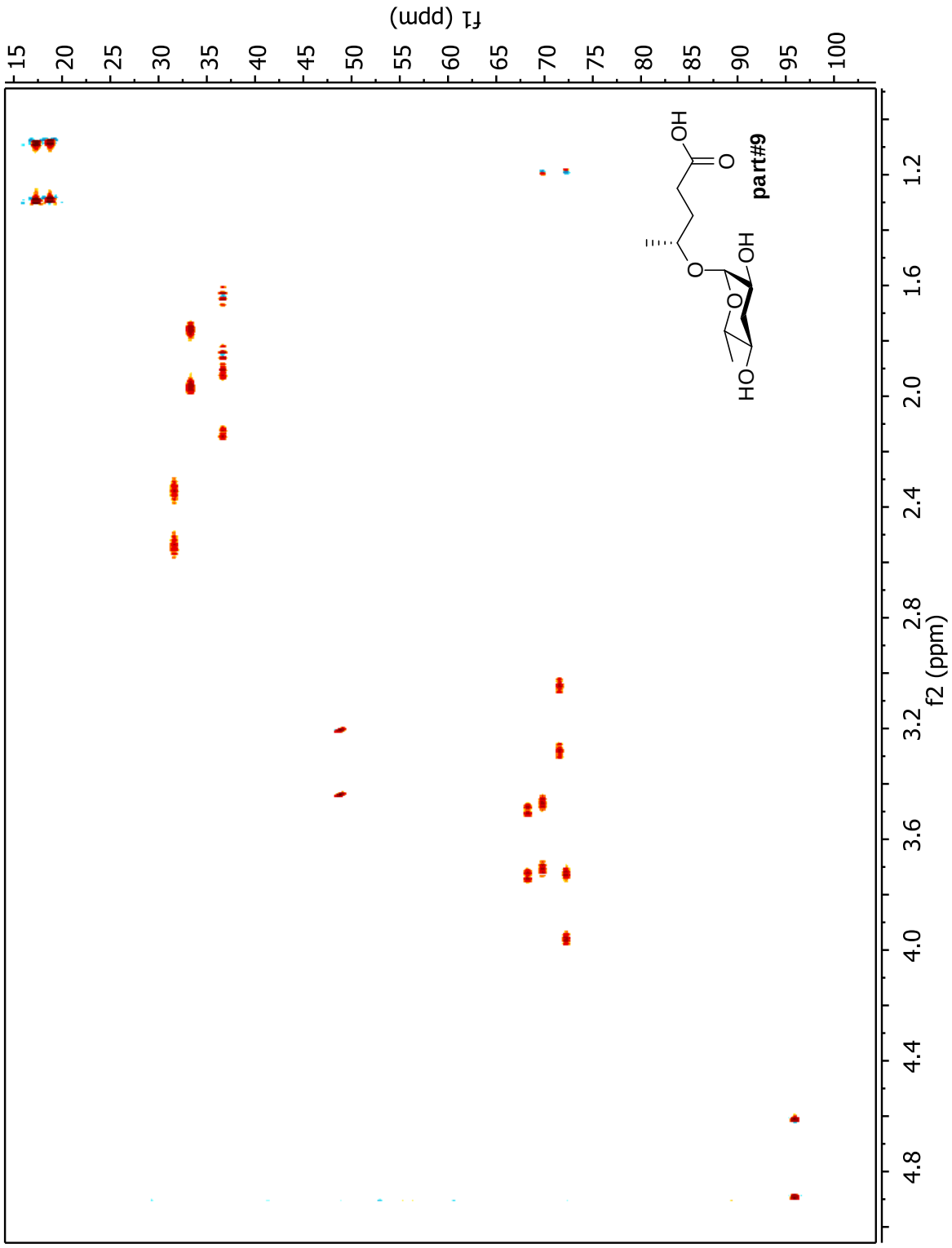


Figure B.12.3: HMQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**part#9**).

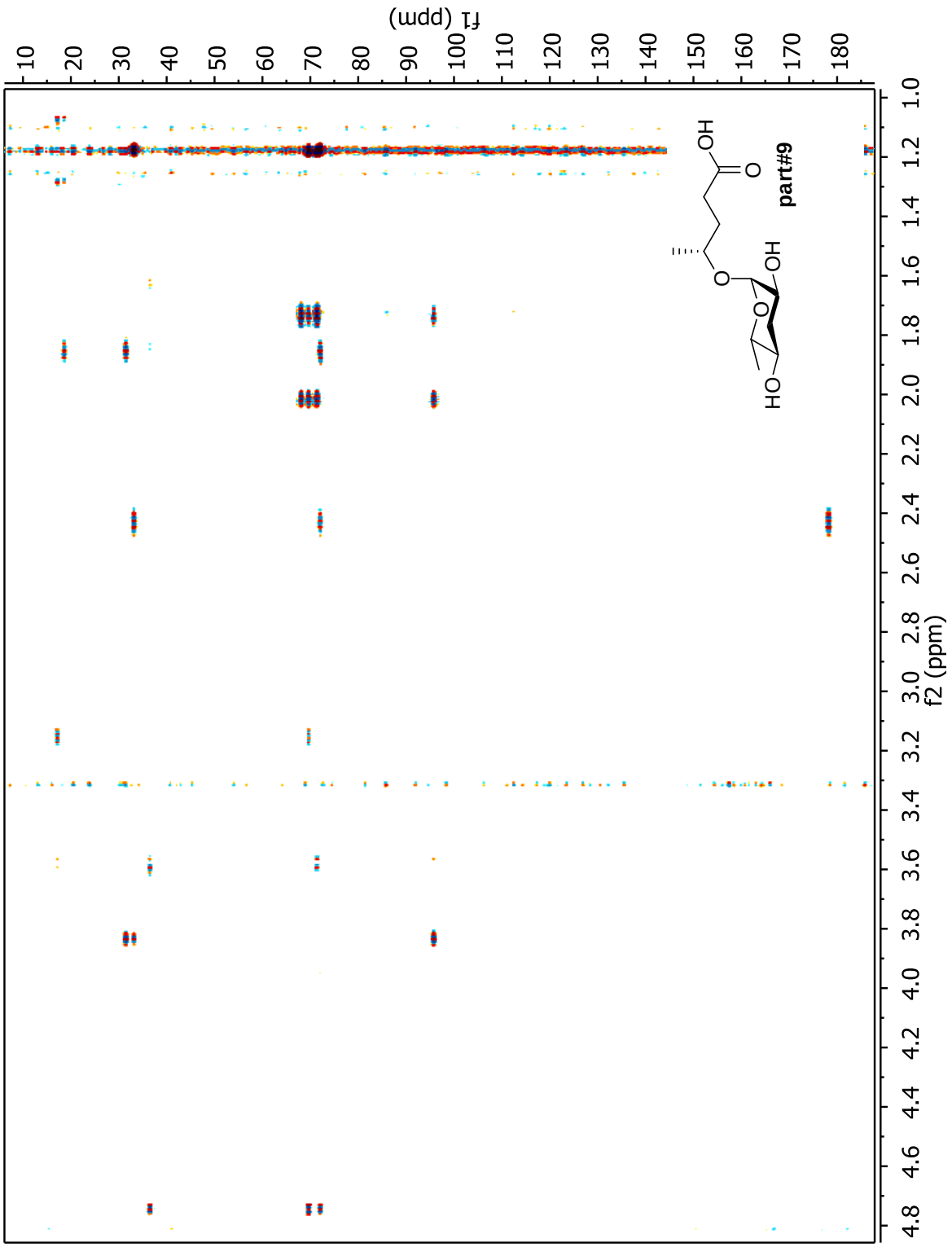


Figure B.12.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**part#9**).

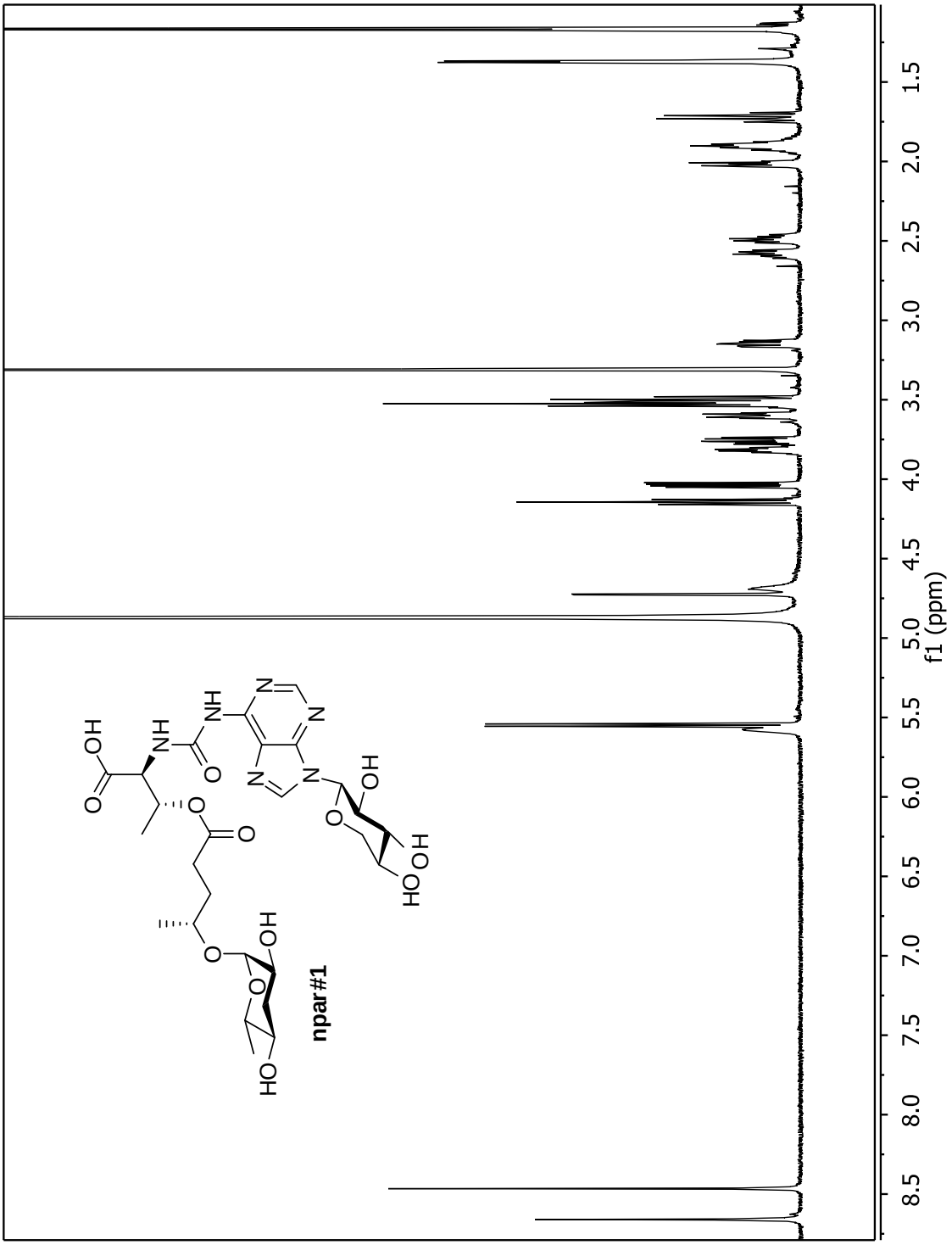


Figure B.13.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido) butanoic acid (**npar#1**).

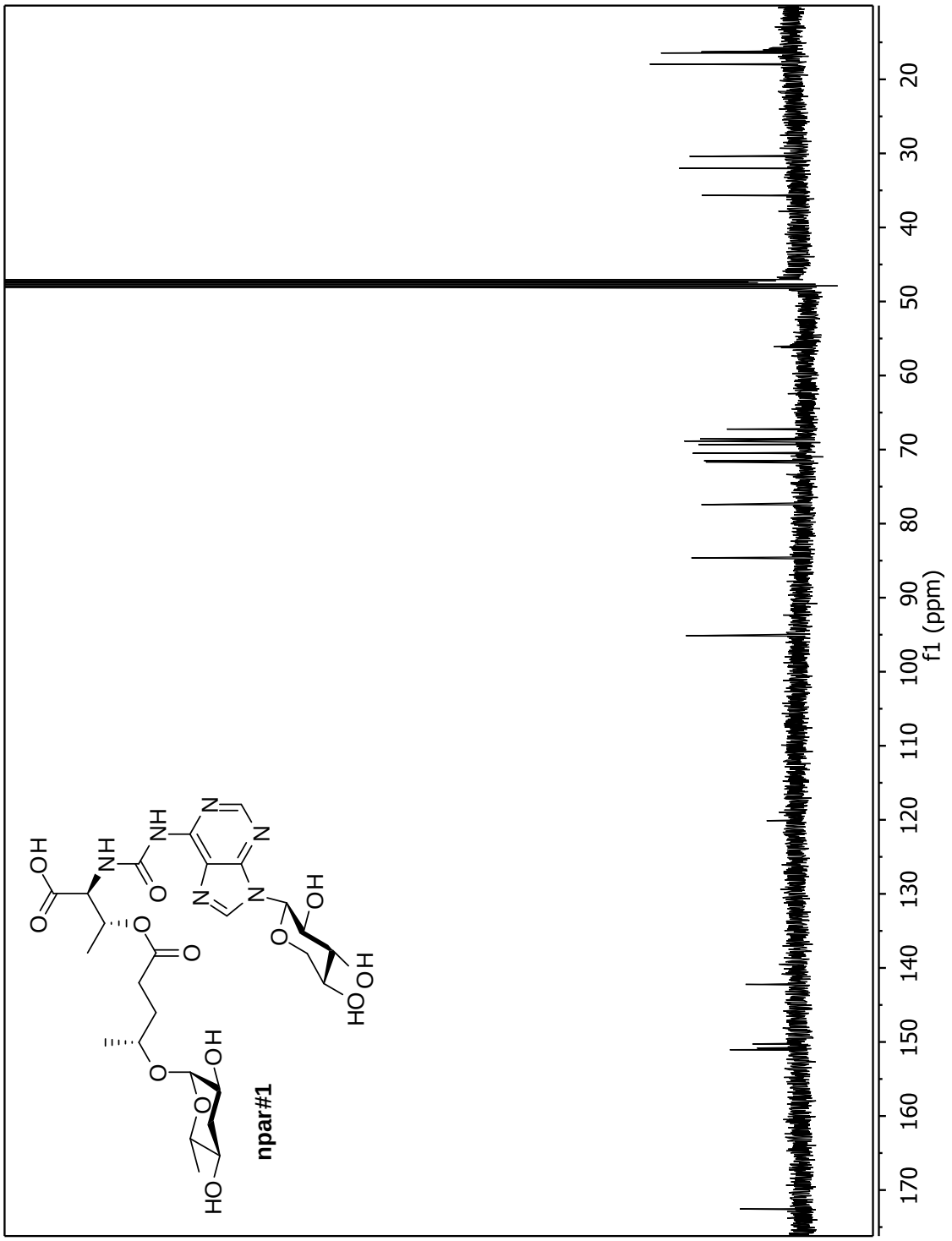


Figure B.13.2: ^{13}C NMR spectrum (151 MHz, methanol-d_4) of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido) butanoic acid (**npar#1**).

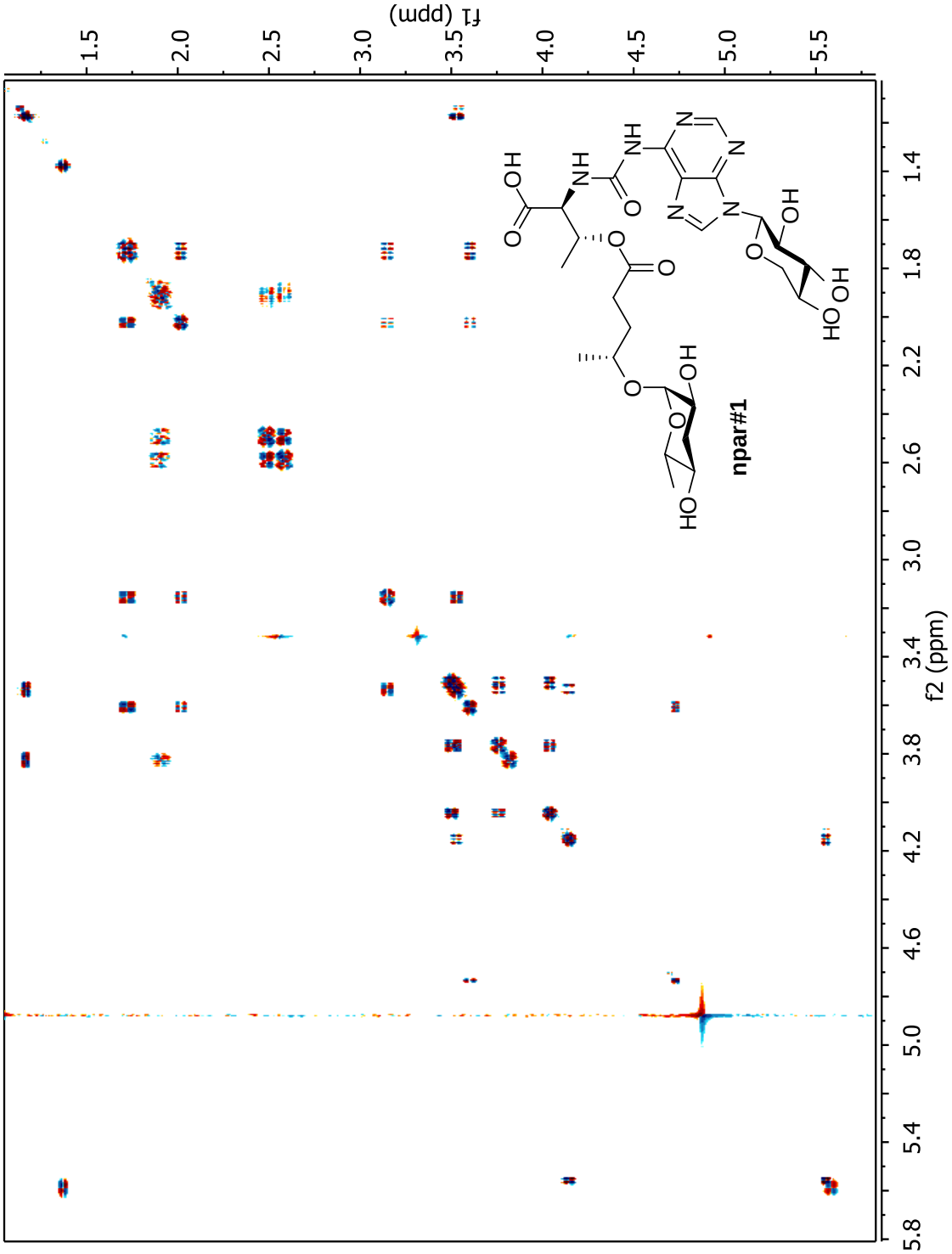


Figure B.13.3: dqfCOSY spectrum (600 MHz, methanol-d_4) of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido) butanoic acid (**npar#1**).

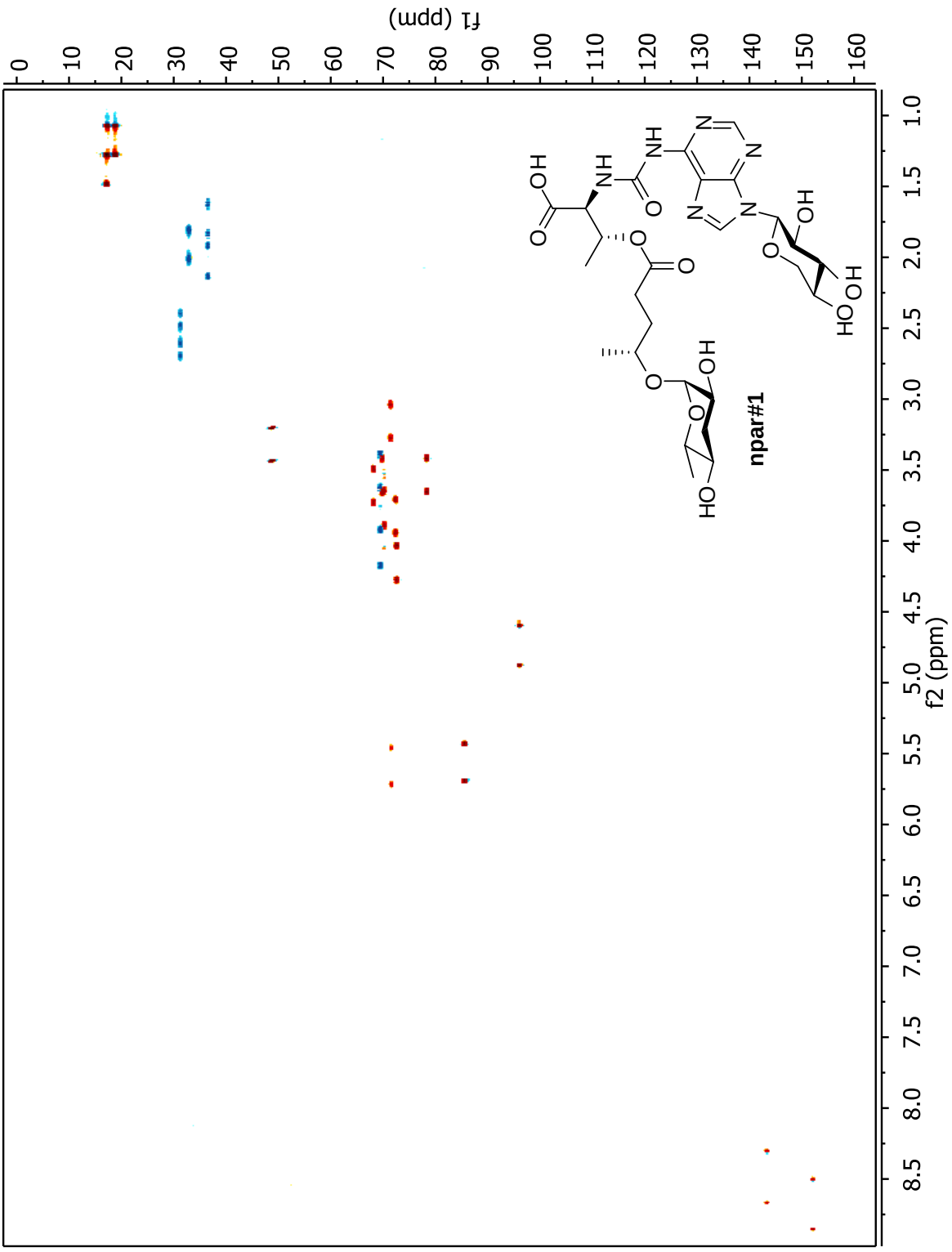


Figure B.13.4: HSQCAD spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido) butanoic acid (**npar#1**).

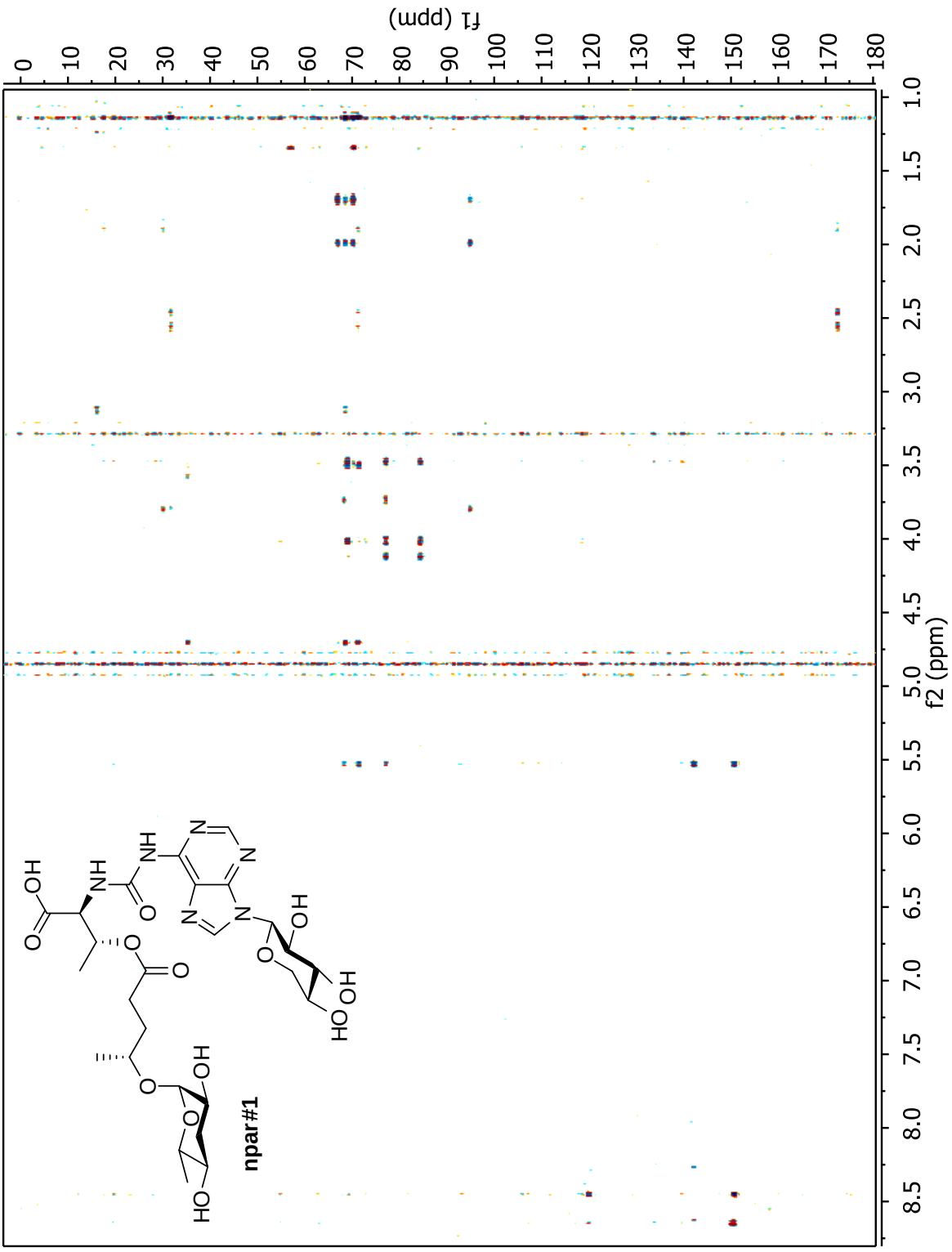


Figure B.13.5: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido) butanoic acid (**npar#1**).

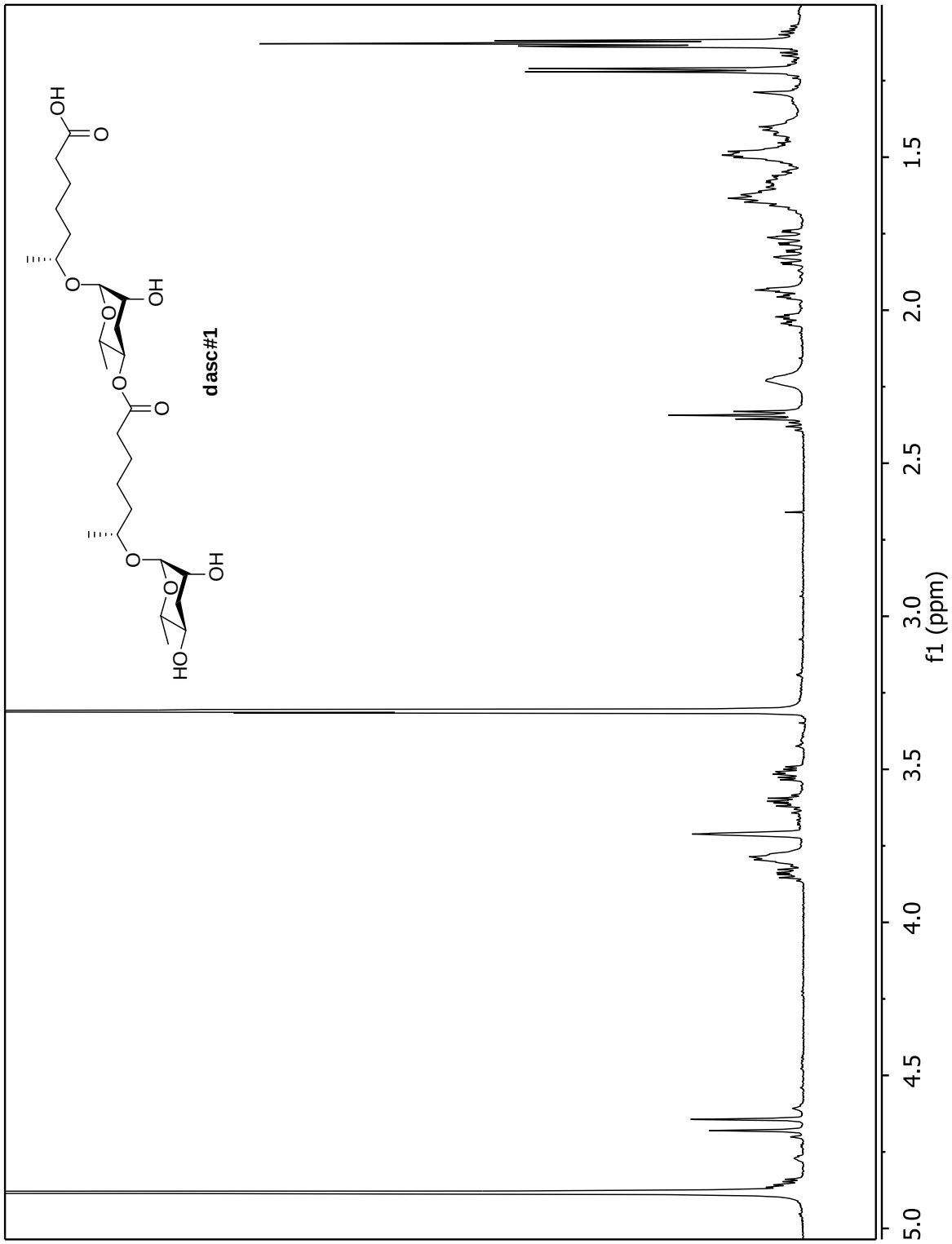


Figure B.14.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-5-(((*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoyl)oxy)-3-hydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoic acid (**dasc#1**).

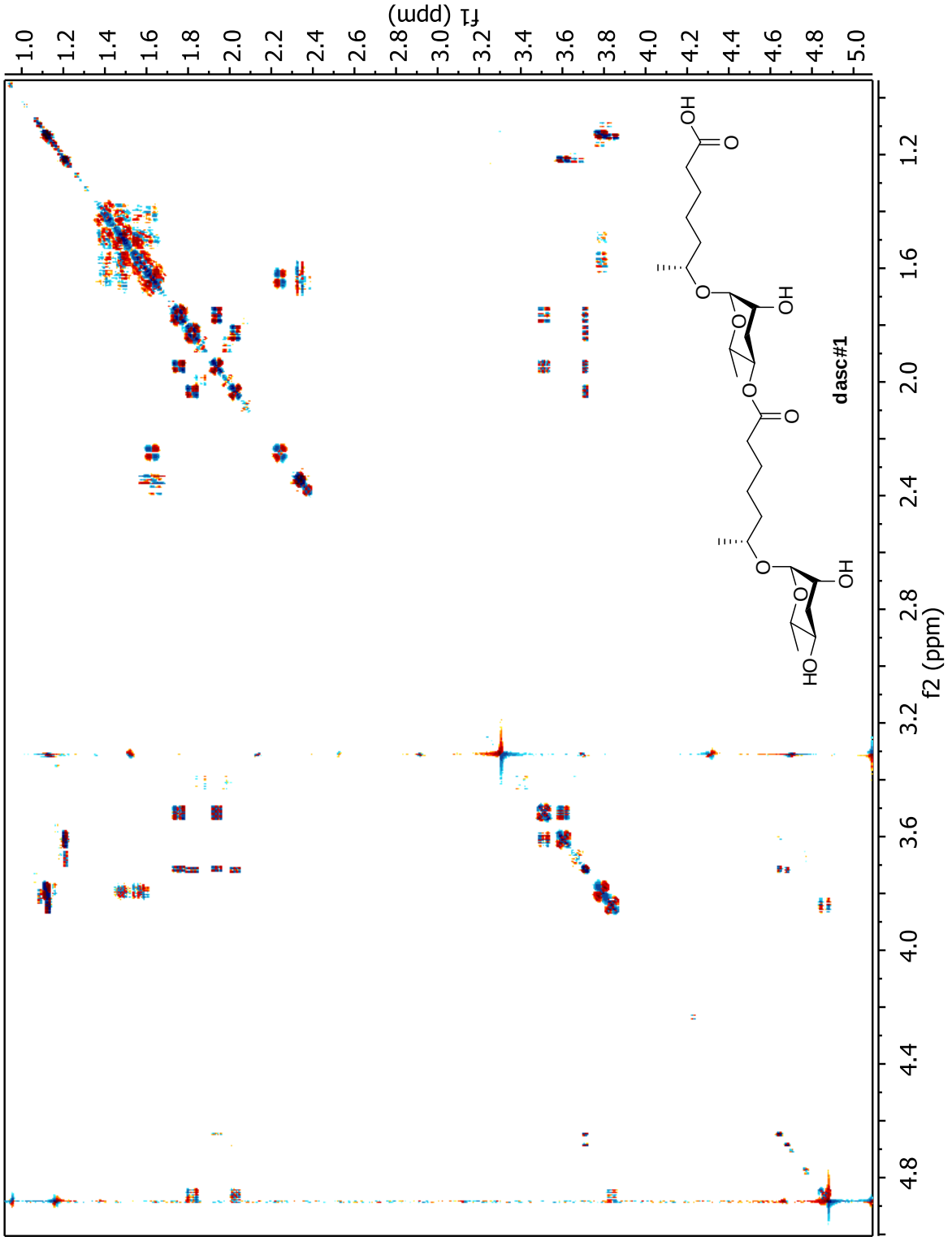


Figure B.14.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of (*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-5-(((*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoyl)oxy)-3-hydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoic acid (**dasc#1**).

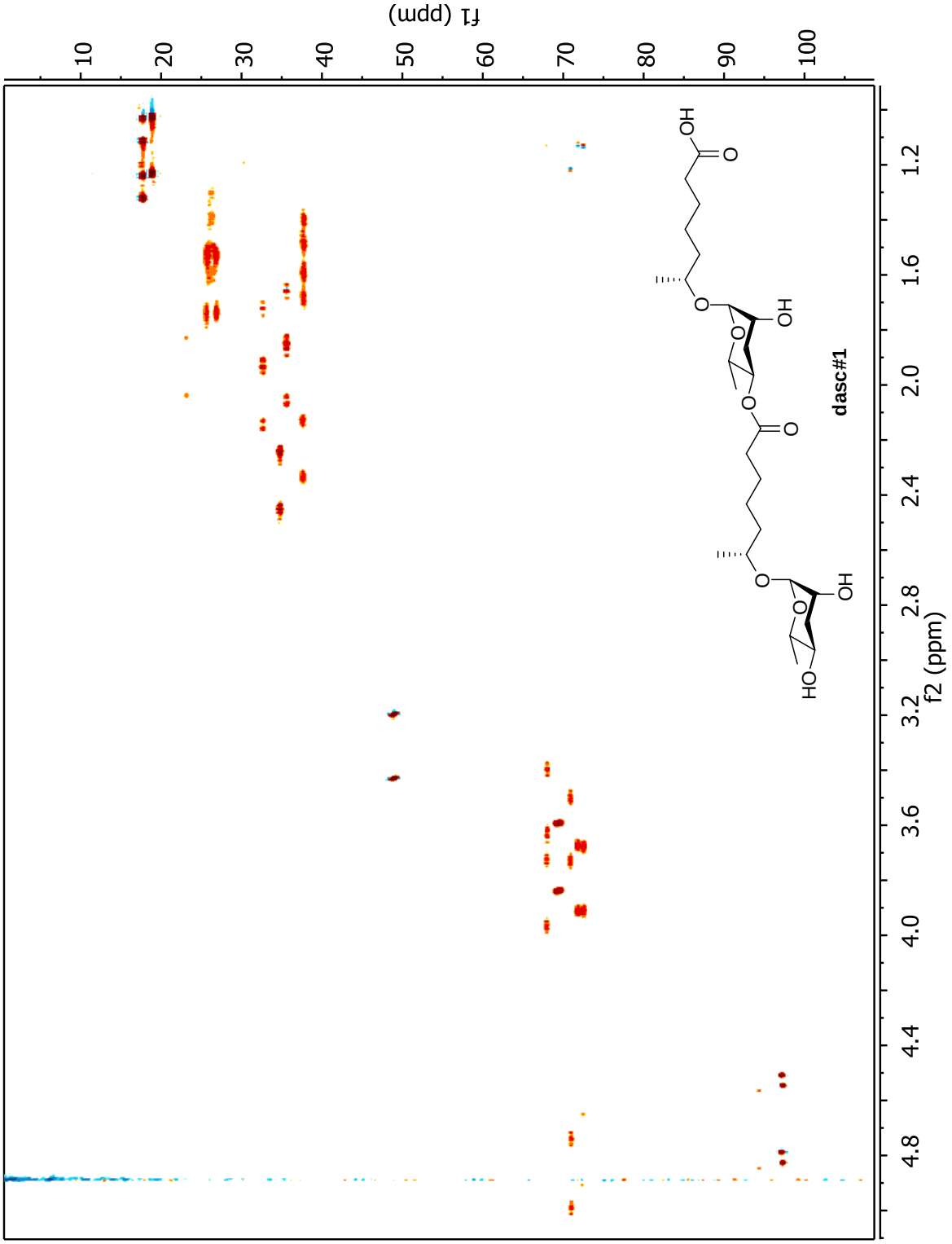


Figure B.14.3: HMQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-5-(((*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoyl)oxy)-3-hydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoic acid (**dasc#1**).

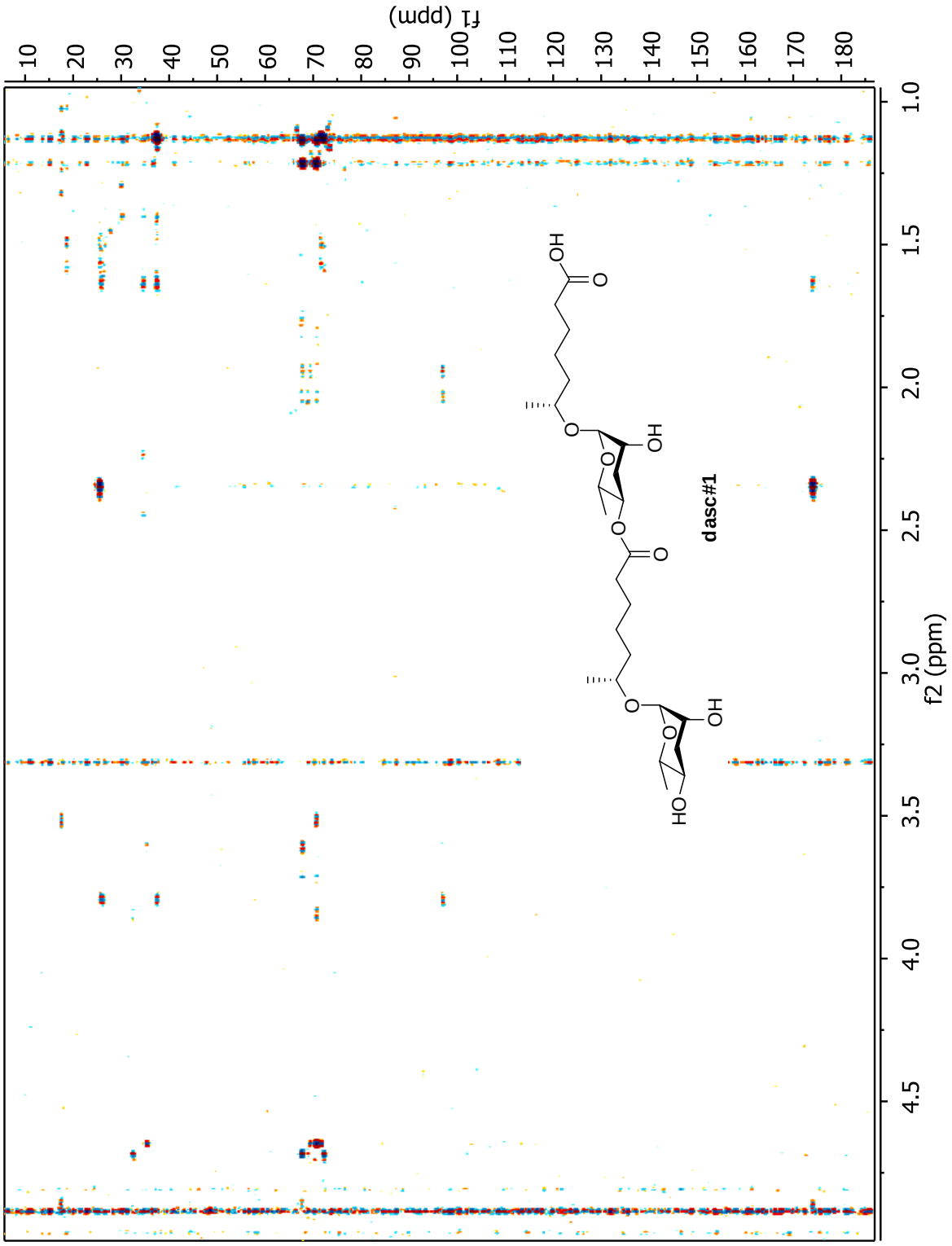


Figure B.14.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-5-(((*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl) oxy)heptanoyl) oxy)-3-hydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoic acid (**dasc#1**).

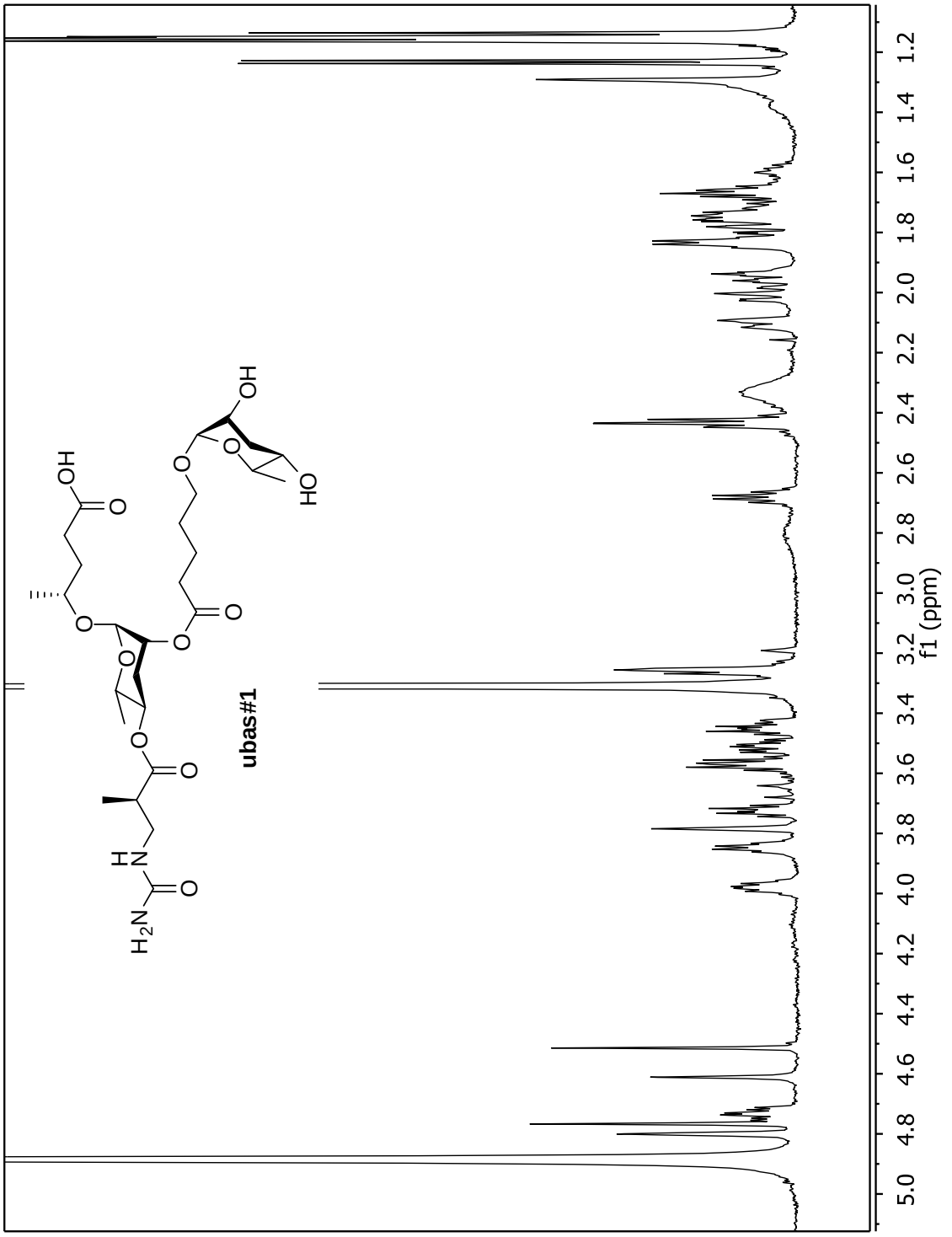


Figure B.15.1: ^{1}H NMR spectrum (600 MHz, methanol-d_4) of (*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3-((5-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-6-methyl-5-(((*R*)-2-methyl-3-ureidopropanoyl)oxy)tetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**ubas#1**).

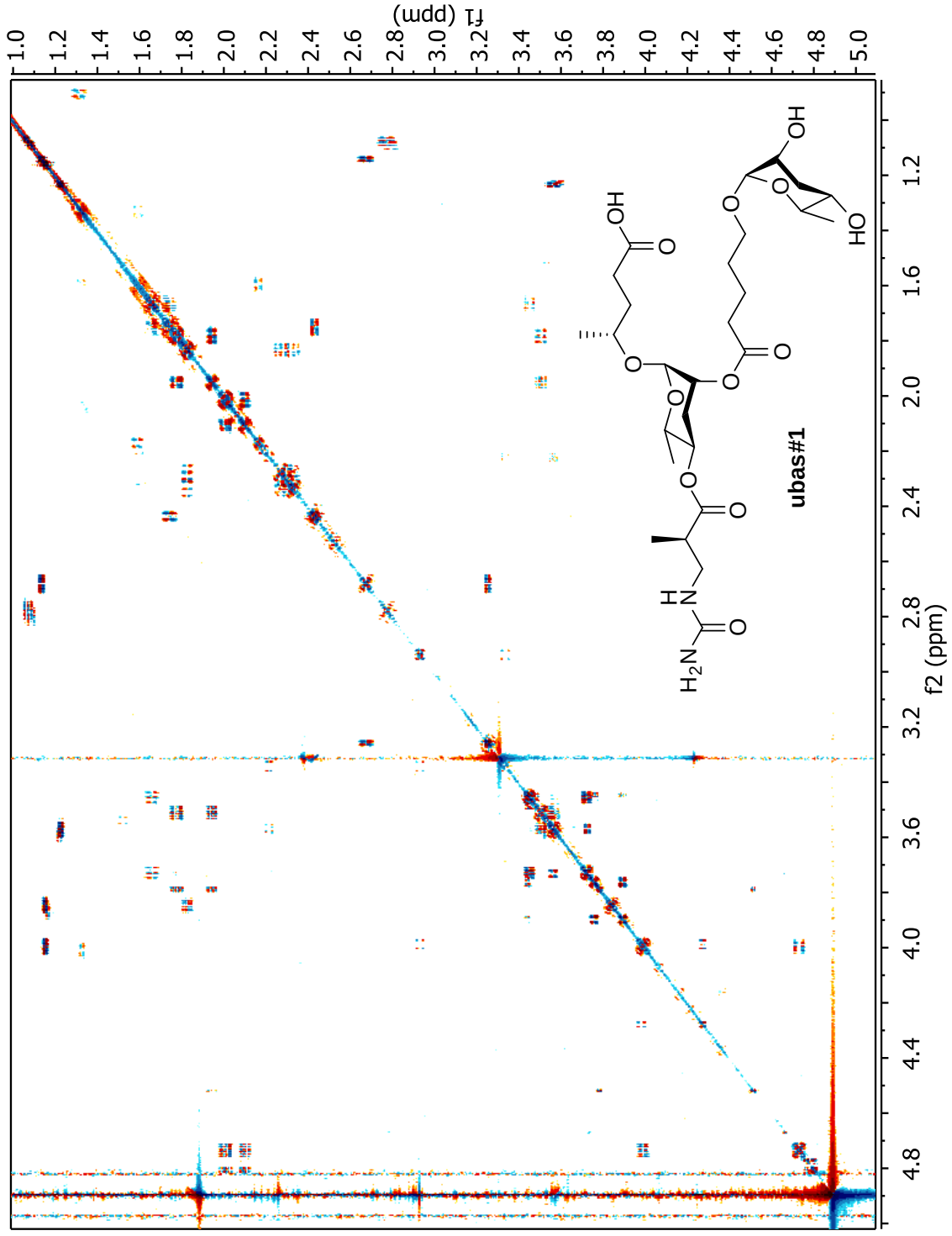


Figure B.15.2: dqfCOSY spectrum (600 MHz, methanol-d_4) of (*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3-((5-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl) oxy)-6-methyl-5-(((*R*)-2-methyl-3-ureidopropanoyl)oxy)tetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**ubas#1**).

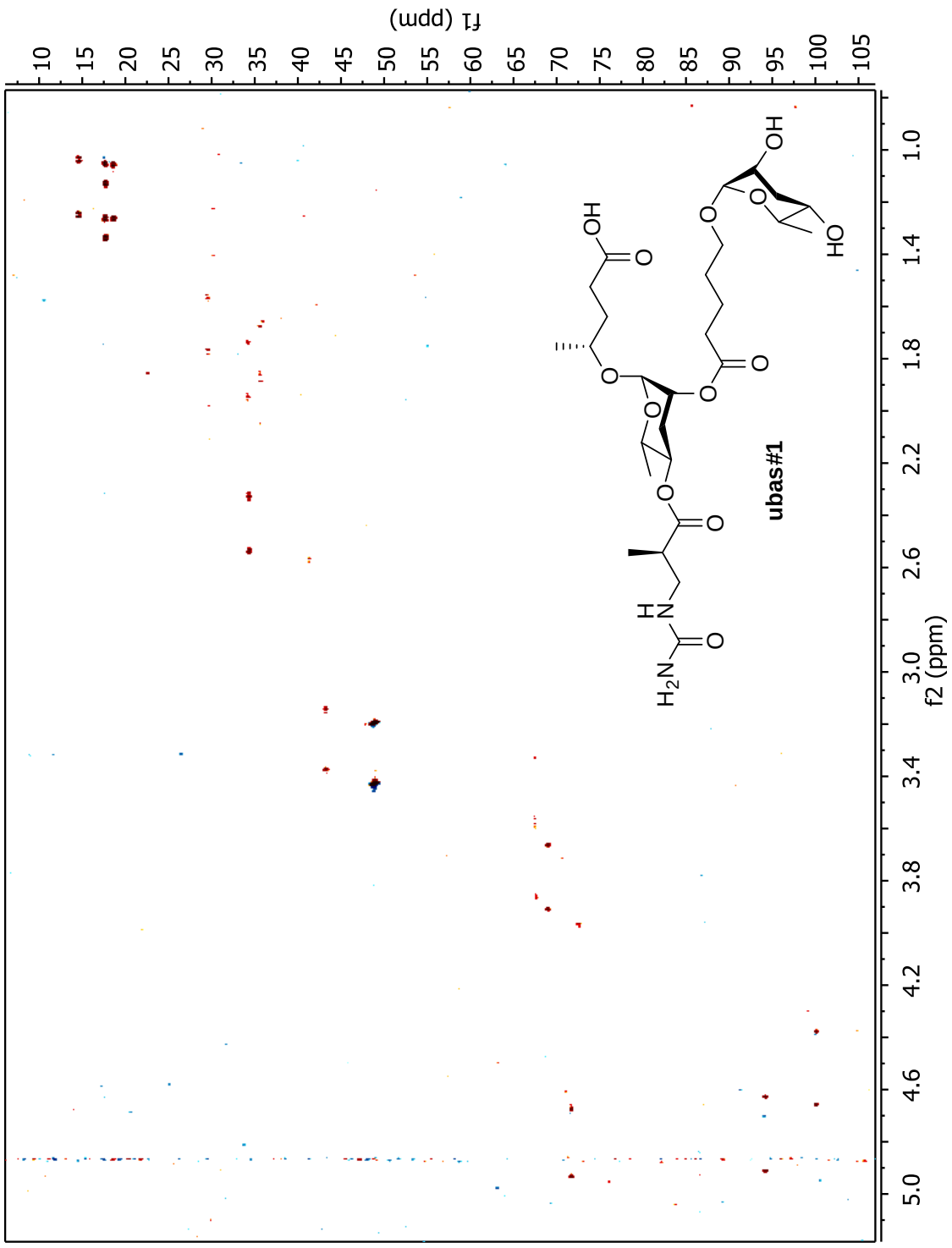


Figure B.15.3: HMQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3-((5-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy) pentanoyl)oxy)-6-methyl-5-(((*R*)-2-methyl-3-ureidopropanoyl)oxy)tetrahydro-2*H*-pyran-2-yl) oxy)pentanoic acid (**ubas#1**).

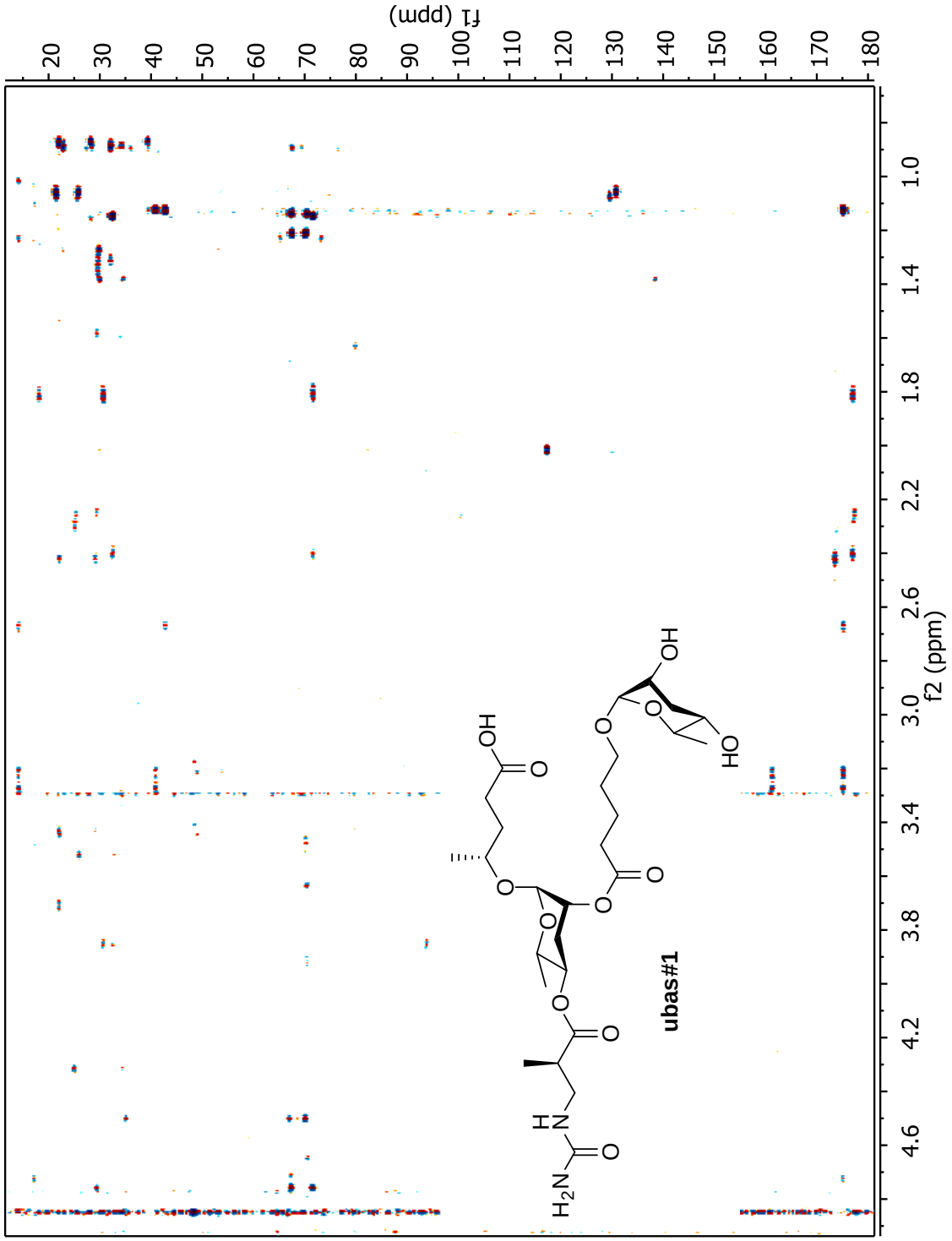
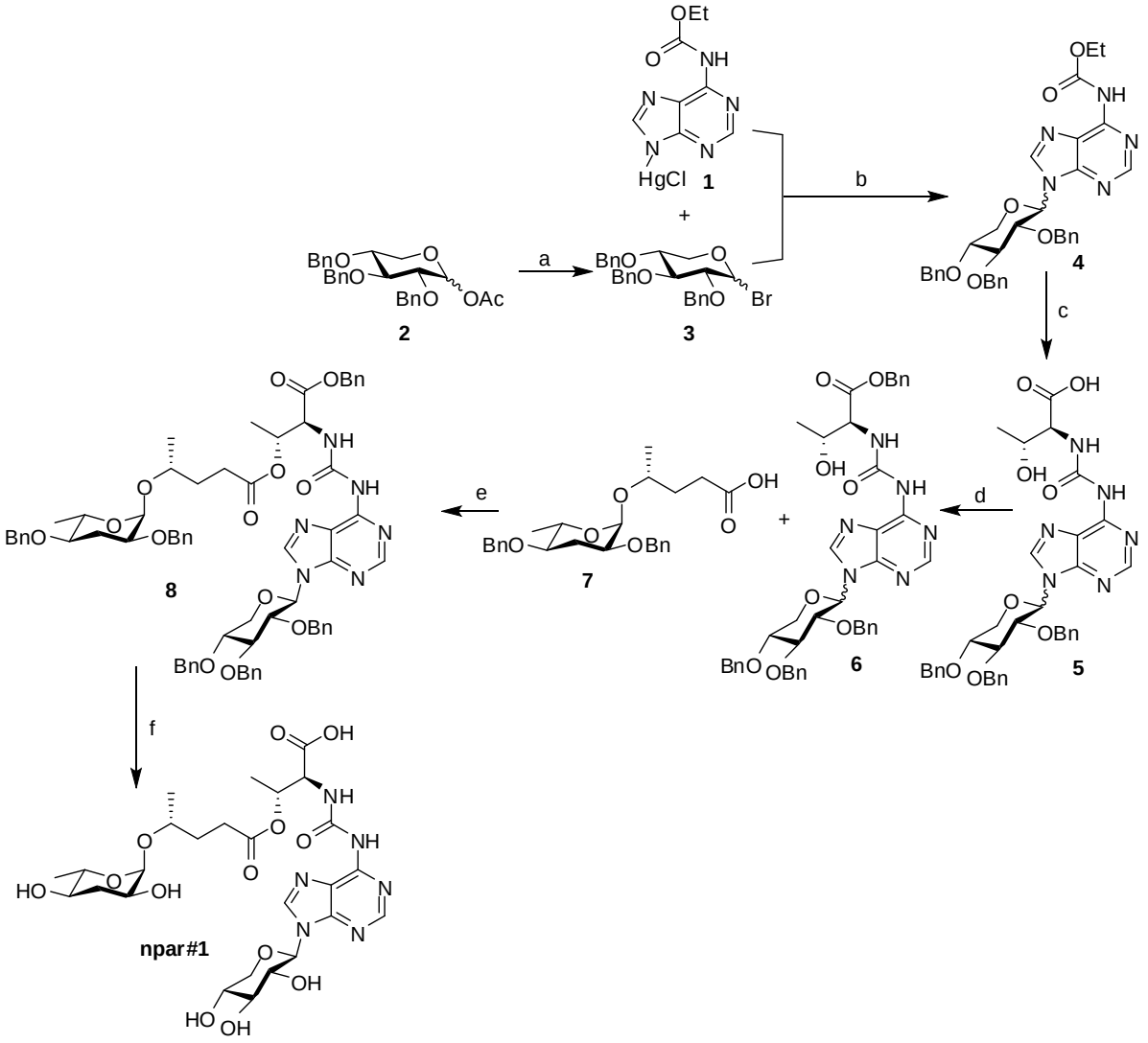


Figure B.15.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, methanol-d_4) of (*R*)-4-(((2*R*,3*R*,5*R*,6*S*)-3-((5-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy) pentanoyl)oxy)-6-methyl-5-(((*R*)-2-methyl-3-ureidopropanoyl)oxy)tetrahydro-2*H*-pyran-2-yl)oxy)pentanoic acid (**ubas#1**).

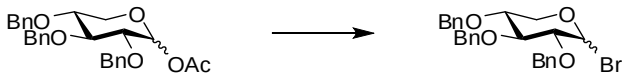
B.3. Chemical synthesis:

B.3.1. Synthesis of npar#1:



Scheme B.1: Overview of synthesis of npar#1. Reagents and conditions: (**a**) TMSBr, DCM, -40 °C to r.t.; (**b**) toluene, reflux; (**c**) L-threonine, pyridine, 107 °C; (**d**) 2-benzyloxy-1-methylpyridinium triflate[7], Et_3N, $PhCF_3$, 83 °C; (**e**) EDC, DMAP, DCM; (**f**) 10% Pd/C, H_2 (*g*) 10% formic acid in MeOH.

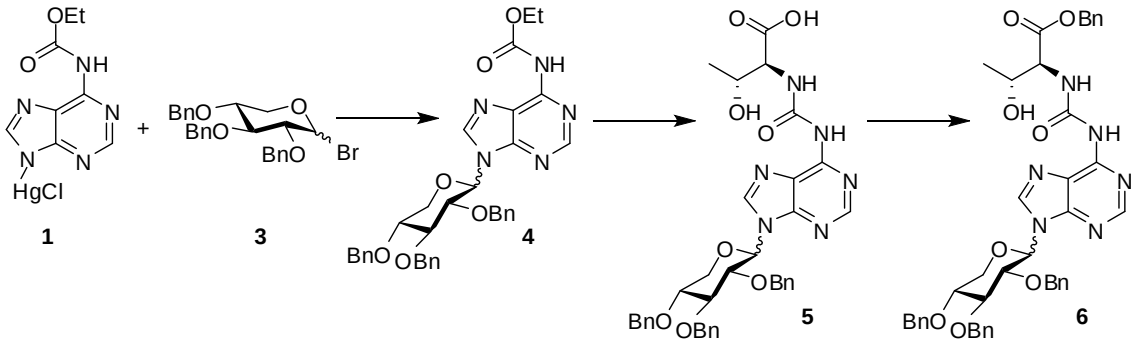
B.3.1.1. Synthesis of (3*R*,4*S*,5*R*)-3,4,5-tris(benzyloxy)-2-bromotetrahydro-2*H*-pyran (3)



2 **3**

To a solution of **2**[8] (550 mg, 1.19 mmol) in dry dichloromethane (1.5 mL) cooled to -40 °C was added trimethylsilyl bromide (3.2 mL, 24.2 mmol) dropwise with constant stirring. The reaction mixture was then allowed to warm up to r.t. and stirred for 45 minutes to afford **3**. The excess reagent and solvent was removed *in vacuo*. The product decomposed in contact to moisture and hence was not characterized further and used for the next step directly.

B.3.1.2. Synthesis of (2*S*,3*R*)-benzyl 3-hydroxy-2-(3-(9-((3*R*,4*S*,5*R*)-3,4,5-tris(benzyloxy)tetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido)butanoate (6):



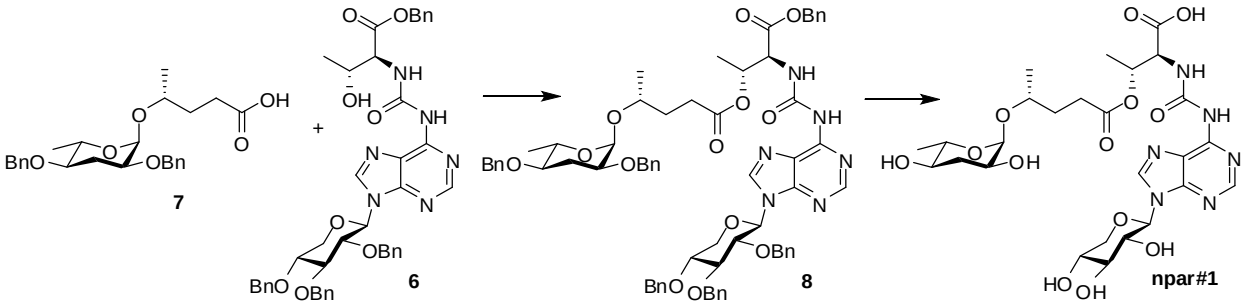
1 **3** **4** **5** **6**

1[9] (500 mg, 1.12 mmol) was dried thoroughly *in vacuo* and added to a solution of **3** in 12 mL dry toluene. The reaction mixture was refluxed for 2.5 h. Toluene was evaporated to reduce the volume to ~3 mL *in vacuo* and 3 mL of petroleum ether was added to it. The

resulting brown suspension was filtered and the precipitate washed with warm chloroform (3 x 10 mL). The filtrate and the washings were combined and washed with 10 mL 30% aq. KI solution, 10 mL water, dried over Na_2SO_4, and concentrated *in vacuo.* Flash column chromatography on silica using a gradient of 0-20% methanol in dichloromethane afforded **4** (160 mg, 262 µmol, 22% over two steps, mixture of α and β anomers in ratio ~ 2:3) as a pale yellow oil. **4** was reacted with L-threonine and worked up following conditions reported for the corresponding 2,3,5-tri-*O*-acetylribofuranoside-derivative.[9] Flash column chromatography on silica using a gradient of 0-30% methanol in dichloromethane containing 0.25% acetic acid afforded **5** (117 mg, 172 µmol, 66%, mixture of α and β-anomers in ratio ~ 2:3) as a yellow solid. A mixture of **5** (72 mg, 106 µmol) and 15 µl triethylamine (210 µmol) in 300 µl trifluoromethylbenzene was treated with 2-benzyloxy-1-methylpyridinium triflate[7] (71 mg, 210 µmol) and stirred at 83 °C for 15 h. The products were partitioned between 2 mL ethyl acetate and 2 mL water, the organic phase washed with 1 mL water, 1 mL brine, dried over Na_2SO_4, and concentrated *in vacuo*. Flash column chromatography on silica using a gradient of 0-20% isopropanol in dichloromethane afforded **6** (8.1 mg, 10 µmol, 10%, β-anomer) as a yellow solid. ^{1}H NMR (500 MHz, methanol-d_4): δ (ppm) 8.39 (s, 1H), 8.27 (s, 1H), 7.43-7.25 (m, 15H), 6.99-6.95 (m, 1H), 6.91-6.86 (m, 2H), 6.67-6.62 (m, 2H), 5.53 (d, *J* = 8.9 Hz, 1H), 5.25 (d, *J* = 12.7 Hz, 1H), 5.23 (d, *J* = 12.7 Hz, 1H), 5.01 (d, *J* = 11.2 Hz, 1H), 4.88 (d, *J* = 11.2 Hz, 1H), 4.74 (s, 2H), 4.62-4.56 (m, 2H), 4.48 (dq, *J* = 6.4 Hz, 2.5 Hz, 1H), 4.26-4.15 (m, 3H), 3.94-3.88 (m, 1H), 3.84-3.79 (m, 1H), 3.53-3.46 (m, 1H), 1.31 (d, *J* = 6.5 Hz, 3H). ^{13}C NMR (125 MHz, methanol-d_4): δ (ppm) 172.2, 156.6, 152.1, 151.5, 151.4, 143.7, 139.9, 139.7, 138.4, 137.2, 129.65, 129.53, 129.46, 129.37,

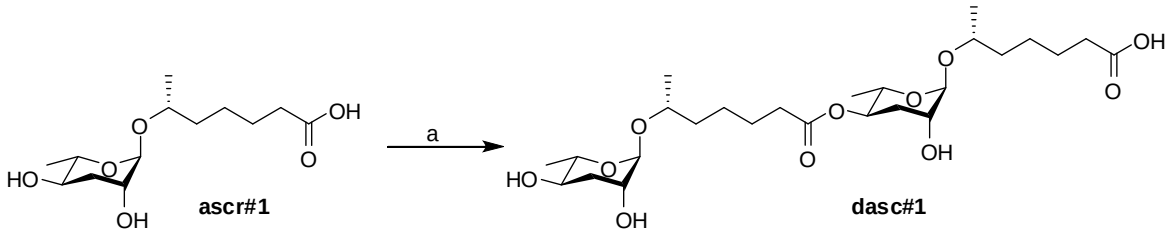
129.24, 129.13, 129.08, 129.06, 128.90, 128.88, 128.74, 128.71, 121.5, 86.5, 85.6, 79.6, 78.8, 76.5, 75.7, 74.1, 68.5, 68.1, 67.7, 60.6, 20.7.

B.3.1.3. Synthesis of (2*S*,3*R*)-3-(((*R*)-4-(((2*R*,3*S*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)pentanoyl)oxy)-2-(3-(9-((2*R*,3*R*,4*S*,5*R*)-3,4,5-trihydroxytetrahydro-2*H*-pyran-2-yl)-9*H*-purin-6-yl)ureido)butanoic acid (npar#1):



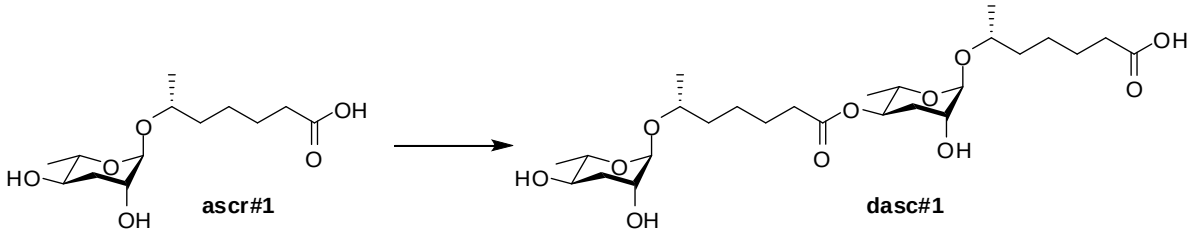
A solution of **7** (4 mg, 9 µmol, provided by Joshua J. Yim) in 450 µL dry dichloromethane was treated with 4-dimethylaminopyridine (2.5 mg, 20 µmol) and EDC hydrochloride (4 mg, 21 µmol). After stirring for 15 minutes, **6** (7.3 mg, 9 µmol) in 300 µL dry dichloromethane was added to the mixture. After stirring for 12 h, the reaction was concentrated *in vacuo*. Flash column chromatography on silica using a gradient of 0-15% isopropanol in dichloromethane afforded **8** (5.5 mg, 4.6 µmol, 51%). A solution of Pd/C (7 mg, 10%, *w/w*) in 500 µL of methanol containing 10% formic acid was first flushed with argon for 5 minutes and subsequently with a moderate flow of H_2 gas. To this stirring solution was added a solution **8** (5.5 mg, 4.6 µmol) in 500 µL methanol. After 4 h, the reaction was filtered over a pad of silica and concentrated *in vacuo*. HPLC purification afforded **npar#1** (1.1 mg, 1.7 µmol, 37%) as a colorless oil. α_D^{20} = -14.5 (*c.* 0.11, methanol). For NMR spectroscopic data, see **Table B.4**.

B.3.2. Synthesis of dasc#1:



Scheme B.2: Overview of synthesis of dasc#1. Reagents and conditions: (**a**) EDC, DMAP, DMF.

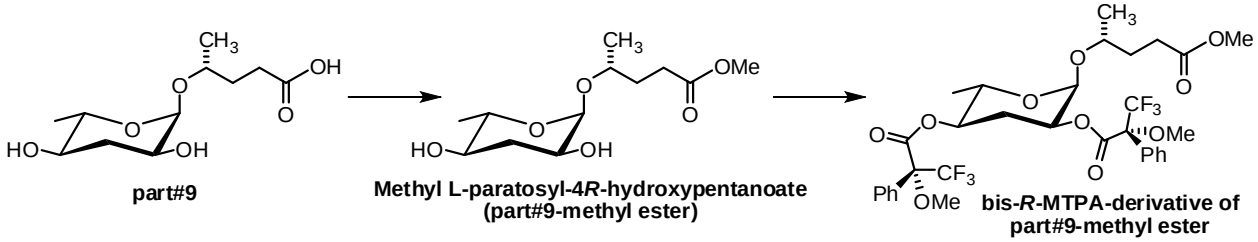
Synthesis of (*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-5-(((*R*)-6-(((2*R*,3*R*,5*R*,6*S*)-3,5-dihydroxy-6-methyltetrahydro-2*H*-pyran-2-yl)oxy)heptanoyl)oxy)-3-hydroxy-6-methyl tetrahydro-2*H*-pyran-2-yl)oxy)heptanoic acid (dasc#1):



A solution of **ascr#1** (15 mg, 54 µmol) in 15 mL dry DMF was added to a solution of 4-dimethylaminopyridine (13.5 mg, 110.7 µmol) and EDC hydrochloride (11 mg, 57.3 µmol) in 7 mL dry DMF. The reaction was monitored by ESI^- MS and was quenched with few drops of glacial acetic acid and concentrated *in vacuo* when polymer peaks (*m/z* = 791 etc.) were observed in significant quantities. Flash column chromatography on silica using a gradient of 0-30% methanol in dichloromethane containing 0.25% and further HPLC purification (see Methods) of crude product mixture afforded **dasc#1** (1.1 mg, 2.1 µmol, 7.8 %) as a colorless oil. α_D^{20} = -115.0 (*c.* 0.11, methanol). For NMR spectroscopic data, see **Table B.5**.

B.3.3. Synthesis of bis-*R*- and bis-*S*- MTPA-derivative of part#9 methyl ester:

B.3.3.1. Synthesis of bis-*R*-MTPA-derivative of part#9 methyl ester:



A solution of **part#9** (240 µg, 0.97 µmol) in a 1:1 mixture (*v/v*) of methanol and toluene (200 µL) was treated with 2.0 M (trimethylsilyl)diazomethane solution (70 µL) in diethyl ether. After stirring for 30 minutes excess reagent was destroyed by addition of acetic acid and the solution concentrated *in vacuo* to yield **part#9-methyl ester,** which was used without further purification. A solution of **part#9-methyl ester** (110 µg, 0.42 µmol) in $CDCl_3$ (300 µL) and dry pyridine (3 µL, 37.5 µmol) was stirred with 4-dimethylaminopyridine (1.4 mg, 11.5 µmol) for 5 min under argon atmosphere and then treated with (*S*)-(+)-α-methoxy-α-trifluoromethylphenylacetyl chloride[6] (*S*-MTPA-Cl) (7 µL, 36.4 µmol) and allowed to stir at r.t. After 8 h, the crude reaction mixture was diluted with $CDCl_3$ (300 µL) and directly placed in an NMR tube for ^{1}H-NMR analysis.

B.3.3.2. Synthesis of bis-*S*-MTPA-derivative of part#9-methyl ester. bis-*S*-MTPA-derivative of part#9-methyl ester was prepared following analogous reaction conditions from part#9-methyl ester and using *R*-(-)-α-methoxy-α-trifluoromethylphenylacetyl chloride[6] (*R*-MTPA-Cl).

B.4. Tables:

Table B.1: HPLC-ESI-MS and concentration estimation data for small molecule signals identified from *P. pacificus* RS2333

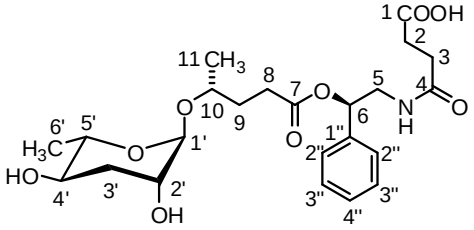
SMID	Molecular formula	LC retention time [Min] ±SD	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ observed**	Estimated concentrations in culture supernatant (μM)***
ascr#9*	$C_{11}H_{20}O_6$	11.68 ± 0.02	247.1187	247.1185	0.3-0.6
ascr#12*	$C_{12}H_{22}O_6$	13.10 ± 0.01	261.1344	261.1305	0.1-0.3
ascr#1*	$C_{13}H_{24}O_6$	14.51 ± 0.02	275.1500	275.1494	0.4-0.8
pasc#9*	$C_{23}H_{33}NO_9$	17.10 ± 0.04	466.2083	466.2089	1.0-2.0
pasc#12	$C_{24}H_{35}NO_9$	17.81 ± 0.03	480.2239	480.2231	0.2-0.5
pasc#1	$C_{25}H_{37}NO_9$	18.59 ± 0.05	494.2396	494.2390	0.05-0.10
ubas#1*	$C_{27}H_{46}N_2O_{13}$	17.51 ± 0.06	605.2927	605.2928	0.2-0.4
ubas#2	$C_{28}H_{48}N_2O_{13}$	18.11 ± 0.05	619.3084	619.3067	0.1-0.2
dasc#1*	$C_{26}H_{46}O_{11}$	19.50 ± 0.01	533.2967	533.2951	0.2-0.5
part#9*	$C_{11}H_{20}O_6$	11.99 ± 0.02	247.1187	247.1181	0.5-1.0
npar#1*	$C_{26}H_{38}N_6O_{13}$	13.64 ± 0.05	641.2424	641.2443	0.5-1.0
npar#2	$C_{21}H_{30}N_6O_9$	14.34 ± 0.07	509.2002	509.1999	0.05-0.10

*Confirmed using synthetic standards.[1,5,10]

**HRMS data was obtained from *P. pacificus* RS2333 exo-metabolome extract analysis.

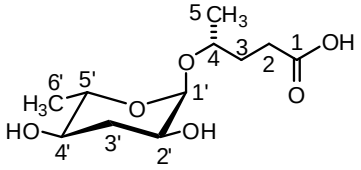
***Quantifications were based on intetegration of HPLC-MS signals from the corresponding ion-traces. Concentrations were calculated using response factors determined for synthetic standards. Concentrations for minor compounds that were not synthesized were based on extrapolation of available standards of closely related structures. A range of concentrations are reported as observed for multiple biological repeats.

Table B.2: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **pasc#9** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_3OD$) = 49.00 ppm.



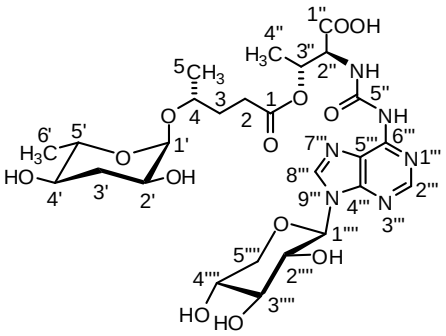
Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Couplings (Hz)	Key HMBC correlations
1	175.8	---		
2	31.3	2.44	$J_{2,3}$ = 7.2	C-1
3	30.1	2.56		C-4
4	174.4	---		
5	45.1	5a = 3.51 5b = 3.58	$J_{5a,5b}$ = 13.7, $J_{5a,6}$ = 8.5 $J_{5b,6}$ = 4.4	C-4
6	75.4	5.84		C-7, C-1″
7	173.8	---		
8	31.3	2.54		C-7
9	33.0	1.81		
10	71.3	3.80	$J_{10,11}$ = 6.6	C-1′
11	18.8	1.14		
1′	97.0	4.64	$J_{1',2'}$ = 2.3	C-2′, C-3′, C-5′
2′	69.4	3.72	$J_{2',3'(ax)}$ = 6.6, $J_{2',3'(eq)}$ = 6.6	
3′	35.6	1.70 (ax) 1.95 (eq)	$J_{3'(ax),3'(eq)}$ = 13.0, $J_{3'(ax),4'}$ = 11.8, $J_{3'(eq),4'}$ = 4.9	
4′	67.8	3.51	$J_{4',5'}$ = 11.8	
5′	71.0	3.58	$J_{5',6'(eq)}$ = 6.4	
6′	17.8	1.20		
1″	139.4	---		
2″	127.3	7.29 – 7.37		C-1″, C-6
3″	129.1	7.29 – 7.37		
4″	128.9	7.29 – 7.37		

Table B.3: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **part#9** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_3OD$) = 49.00 ppm.



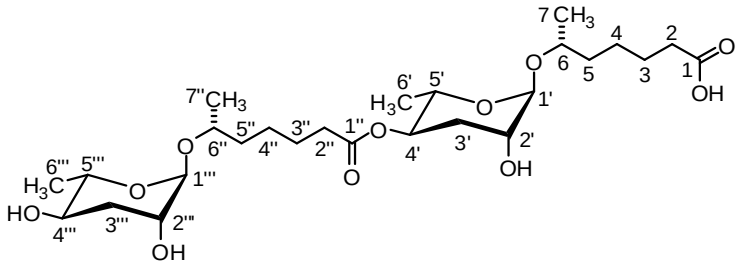
Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Couplings (Hz)	Key HMBC correlations
1	178.3	---		
2	31.8	2.43	$J_{2,3}$ = 7.1	C-1, C-3
3	33.5	1.85		
4	72.4	3.83	$J_{4,5}$ = 6.3	C-1′
5	19.0	1.179		
1′	95.9	4.74	$J_{1',2'}$ = 3.9	C-2′, C-3′, C-5′
2′	68.4	3.60	$J_{2',3'(ax)}$ = 12.1; $J_{2',3'(eq)}$ = 5.5	
3′	36.8	1.74 (ax) 2.02 (eq)	$J_{3'(eq),3'(ax)}$ = 12.3; $J_{3'(eq),4'}$ = 4.7 $J_{3'(ax),4'}$ = 10.9	
4′	71.6	3.15	$J_{4',5'}$ = 9.4	
5′	69.9	3.58	$J_{5',6'(eq)}$ = 6.1	
6′	17.5	1.18		C-4, C-5

Table B.4: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **npar#1** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_3OD$) = 49.00 ppm.



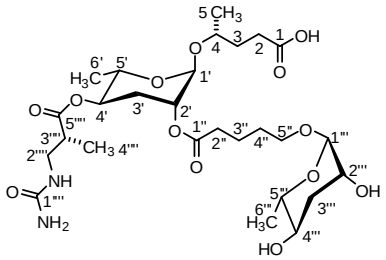
Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Couplings (Hz)	Key HMBC correlations
1	173.6	---		
2	31.5	2.59, 2.49		C-1
3	33.1	1.84-1.97		
4	72.6	3.82	$J_{4,5}$ = 6.2	C-1′
5	19.0	1.17		
1′	96.1	4.73	$J_{1',2'}$ = 3.9	C-2′, C-3′, C-5′
2′	68.3	3.61	$J_{2',3'\,(ax)}$ = 12.1; $J_{2',3'\,(eq)}$ = 5.8	
3′	36.7	1.73 (ax) 2.02 (eq)	$J_{3'\,(eq),3'\,(ax)}$ = 12.4; $J_{3'(eq),4'}$ = 4.6 $J_{3'\,(ax),4'}$ = 11.7	
4′	71.5	3.15	$J_{4',5'}$ = 9.7	
5′	69.9	3.53	$J_{5',6'\,(eq)}$ =6.3	
6′	17.6	1.173		C-5′
1″	174.7	---		
2″	58.4	4.69	$J_{2'',3''}$ = 6.3	C-1″, C-3″, C-5″
3″	71.7	5.58	$J_{3'',4''}$ = 6.8	
4″	17.4	1.38		
5″	154.7	---		
2‴	152.1	8.66		C-6‴
4‴	151.9	---		
5‴	121.2	---		
6‴	151.6	---		
8‴	143.3	8.47		C-4‴, C-5‴
1⁗	85.7	5.56	$J_{1'''',2''''}$ = 9	C-4‴, C-8‴, C-2⁗, C-3⁗
2⁗	72.8	4.15	$J_{2'''',3''''}$ = 9.2	
3⁗	78.6	3.53	$J_{3'''',4''''}$ = 9.2	
4⁗	70.4	3.76	$J_{4'''',5''''(ax)}$ = 10.4, $J_{4'''',5''''(eq)}$ = 5.5	
5⁗	69.7	3.50 (ax) 4.04 (eq)	$J_{5''''(ax),5''''(eq)}$ = 11.5	

Table B.5: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **dasc#1** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_3OD$) = 49.00 ppm.



Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Couplings (Hz)	Key HMBC correlations
1	177.0	---		
2	37.8	2.25	$J_{2,3}$ = 7.7	C-1
3	25.9 or 26.2	1.64		
4	26.4 or 27.1	1.36-1.68		
5	37.9	1.45-1.62		
6	71.9 or 72.7	3.79	$J_{6,7}$ = 6.3	C-1′
7	19.0	1.13		
1′	97.4	4.68	$J_{1',2'}$ = 2.0	C-2′, C-3′, C-5′
2′	69.4	3.72	$J_{2',3'(ax)}$ = 6.3; $J_{2',3'(eq)}$ = 5.9	
3′	32.8	1.83 (ax) 2.04 (eq)	$J_{3'(eq),3'(ax)}$ = 13.5; $J_{3'(eq),4'}$ = 5.0 $J_{3'(ax),4'}$ = 11.7	
4′	71.1	4.86	$J_{4',5'}$ = 10.1	C-1″
5′	68.0	3.84	$J_{5',6'(eq)}$ = 6.4	
6′	17.9	1.14		C-4′, C-5′
1″	174.4	---		
2″	34.9	2.35	$J_{2'',3''}$ = 7.5	C-1″
3″	25.9 or 26.2	1.64		
4″	26.4 or 27.1	1.36-1.68		
5″	37.9	1.45-1.62		
6″	71.9 or 72.7	3.79	$J_{6'',7''}$ = 6.3	C-1‴
7″	19.0	1.13		
1‴	97.2	4.64	$J_{1''',2'''}$ = 2.4	C-2‴, C-3‴, C-5‴
2‴	69.4	3.71	$J_{2''',3'''(ax)}$ = 6.2; $J_{2''',3'''(eq)}$ = 6.0	
3‴	35.7	1.76 (ax) 1.95 (eq)	$J_{3'''(eq),3''' (ax)}$ = 13.4; $J_{3'''(eq),4'''}$ = 4.7 $J_{3'''(ax),4'''}$ = 11.8	
4‴	68.1	3.51	$J_{4''',5'''}$ = 9.4	
5‴	70.9	3.61	$J_{5''',6'''(eq)}$ =6.4	
6‴	17.9	1.22		C-4‴, C-5‴

Table B.6: ^{1}H (600 MHz), ^{13}C (151 MHz), and HMBC NMR spectroscopic data for **ubas#1** in methanol-d_4. Chemical shifts were referenced to ($CD_2\underline{H}OD$) = 3.31 ppm and ($\underline{C}D_3OD$) = 49.00 ppm.



Position	δ ^{13}C [ppm]	δ ^{1}H [ppm]	J_{HH} Couplings (Hz)	Key HMBC correlations
1	174.6	---		
2	35.0	2.31	$J_{2,3}$ = 7.4	C-1
3	34.3	1.83	$J_{3,4}$ = 6.2	
4	72.7	3.84	$J_{4,5}$ = 6.4	C-1′
5	18.9	1.16		
1′	94.2	4.77	$J_{1',2'}$ = 2.4	C-2′, C-3′, C-5′
2′	71.8	4.80	$J_{2',3'(ax)}$ = 6.4; $J_{2',3'(eq)}$ = 5.8	C-1″
3′	30.7	2.01(ax) 2.10(eq)	$J_{3'(eq),3'(ax)}$ = 13.7; $J_{3'(eq),4'}$ = 4.2 $J_{3'(ax),4'}$ = 11.7	
4′	71.1	4.73	$J_{4',5'}$ = 9.2	C-5′′′′
5′	67.9	3.99	$J_{5',6'(eq)}$ = 6.4	
6′	17.8	1.16		
1″	173.9	---		
2″	34.5	2.43	$J_{2'',3''}$ = 7.0	C-1″
3″	22.7	1.75		
4″	29.6	1.67	$J_{4'',5''}$ = 6.2	
5″	67.6	5a″ = 3.45 5b″ = 3.76		C-1‴
1‴	100.1	4.51	$J_{1''',2'''}$ = 2.6	C-2‴, C-3‴, C-5‴
2‴	69.1	3.79	$J_{2''',3'''(ax)}$ = 6.1, $J_{2''',3''' (eq)}$ = 5.9	
3‴	35.7	1.78 (ax) 1.95 (eq)	$J_{3'''(eq),3'''(ax)}$ = 13.1, $J_{3'''(eq),4'''}$ = 4.3 $J_{3'''(ax),4'''}$ = 11.2	
4‴	68.1	3.51	$J_{4''',5'''}$ = 9.5	
5‴	70.5	3.57	$J_{5''',6'''(eq)}$ = 6.4	
6‴	17.9	1.23		
1′′′′	161.7	---		
2′′′′	43.4	3.26	$J_{2'''',3''''}$ = 6.8	C-1′′′′
3′′′′	41.5	2.68	$J_{3'''',4''''}$ = 6.9	C-5′′′′
4′′′′	14.7	1.14		C-5′′′′
5′′′′	175.6			

REFERENCES

(1) Bose, N.; Ogawa, A.; von Reuss, S. H.; Yim, J. J.; Ragsdale, E. J.; Sommer, R. J.; Schroeder, F. C. *Angew Chem Int Ed Engl* **2012**, *51*, 12438.

(2) Ogawa, A.; Streit, A.; Antebi, A.; Sommer, R. J. *Curr Biol* **2009**, *19*, 67.

(3) Stiernagle, T. *WormBook : the online review of C. elegans biology* **2006**, 1.

(4) Lu, N. C.; Goetsch, K. M. *Nematologica* **1993**, *39*, 303.

(5) Srinivasan, J.; von Reuss, S. H.; Bose, N.; Zaslaver, A.; Mahanti, P.; Ho, M. C.; O'Doherty, O. G.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *PLoS Biol.* **2012**, *10*, e1001237. Epub 2012 Jan 10.

(6) Hoye, T. R.; Jeffrey, C. S.; Shao, F. *Nat Protoc* **2007**, *2*, 2451.

(7) Tummatorn, J.; Albiniak, P. A.; Dudley, G. B. *J Org Chem.* **2007**, *72*, 8962.

(8) Lucero, C. G.; Woerpel, K. A. *J Org Chem* **2006**, *71*, 2641.

(9) Chheda, G. B.; Hong, C. I. *J Med Chem* **1971**, *14*, 748.

(10) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J Am Chem Soc.* **2012**, *134*, 1817.

APPENDIX C

COMPARATIVE METABOLOMICS REVEALS ENDOGENOUS LIGANDS OF DAF-12, A NUCLEAR HORMONE RECEPTOR REGULATING *C. ELEGANS* DEVELOPMENT

C.1. Materials and methods:

C.1.1. *C. elegans* strains and maintenance: Nematode stocks were maintained on Nematode Growth Medium (NGM) plates made with Bacto agar (BD Biosciences) and seeded with bacteria (*E. coli* strain OP50) at 20 °C (http://www.wormbook.org/). *C. elegans* strains with the following genotypes were used: wild-type (N2, Bristol), *daf-22(m130), daf-9(dh6), daf-9(dh6);daf-12(rh411rh61),* and *strm-1(tm1781)*. Worms were grown at 20 °C for at least two generations under replete growth conditions prior to growing in liquid cultures.

C.1.2. Liquid cultures: Worms from four 10 cm NGM agar plates were washed using M9-medium into a 100 mL S-complete medium pre-culture where they were grown for four days at 22 °C on a rotary shaker. Concentrated bacteria derived from 1 L of *E. coli* OP50 culture was added as food on days one and three. Subsequently, the pre-culture was divided equally into sixteen 500 mL Erlenmeyer flasks containing 100 mL of S-complete medium on day 4, which was then grown for an additional 5 days at 22 °C on a rotary shaker and fed with concentrated bacteria *ad lib*. The cultures were harvested

on day 5 and centrifuged to separate the supernatant media ("exo-metabolome") and worm pellets ("endo-metabolome"). At harvest, liquid cultures contained approximately 80% L1-L3 worms. The worm pellets were stored at -20 °C until used for further analyses.

C.1.3. Preparation of endo-metabolome extracts: The frozen worm pellets were added to pre-chilled (-78 °C) 200 mL of methanol in a Waring laboratory blender and blended until no chunks remained. Methanol was evaporated *in vacuo* at 20 °C and the residue resuspended in 300 mL of water. The resulting suspension was then frozen using a dry ice-acetone bath and lyophilized. The lyophilized residue was crushed to a fine powder using a mortar and pestle over 8 g granular sodium chloride. The powder was then extracted twice with 250 mL of 9:1 ethyl acetate:ethanol mixture over 12 h. The resulting yellow-brown suspension was filtered and the filtrate evaporated *in vacuo* at room temperature to produce the "endo-metabolome" (worm pellet metabolome) extract and was used for chromatographic separations.

C.1.4. Fractionation protocol for *C. elegans* endo-metabolome extracts: To 8 g of Celite, prewashed with ethyl acetate, was added a solution of worm pellet metabolome/endo-metabolome extract from 16 cultures. After evaporation of the solvent, the Celite® was dry-loaded into an empty 25 g RediSep *Rf* loading cartridge. Fractionation was performed using a Teledyne ISCO CombiFlash system over a RediSep*Rf* GOLD 40 g HP Silica Column using a normal phase dichloromethane-

methanol solvent system, starting with 6 min of 100% dichloromethane, followed by a linear increase of methanol content up to 10% at 24 min, followed by another linear increase of methanol content up to 40% at 40 min, followed by another linear increase of methanol content up to 95% at 45 min which was then continued to 48 min. 70 fractions (~20 mL each) generated from the combiflash run were individually evaporated *in vacuo* and prepared for bio-assays and analyses by NMR spectroscopy (^{1}H NMR, dqfCOSY), HPLC, and SIM-GC/MS.

For the quantification of lathosterone and 4-cholesten-3-one in *daf-22(m130)* worms, fractions 4-10 from the set of 70 fractions were combined and added to 4 g of Celite that had been prewashed with ethyl acetate. After evaporation of the solvent, the Celite was dry-loaded into an empty 12 g RediSep *Rf* loading cartridge. Fractionation was performed using the CombiFlash system and a RediSep *Rf* GOLD 24 g HP silica column using a hexanes-ethyl acetate solvent system, starting with 4.8 min of 100% hexanes, followed by a linear increase of ethyl acetate content up to 40% at 43.2 min to generate 130 fractions (~12 mL each). A synthetic mixture of lathosterone and 4-cholesten-3-one was subsequently run following an identifcal CombiFlash protocol. A range of fractions from the natural sample run that corresponded to the elution profiles of the synthetic 3-keto steroids were individually evaporated *in vacuo* and analyzed by NMR spectroscopy (^{1}H NMR, dqfCOSY) and GC/MS for quantification of lathosterone and 4-cholesten-3-one.

C.1.5. HPLC enrichment protocol: Metabolome fractions of interest were evaporated *in vacuo*, resuspended in 250 μL of methanol and submitted to HPLC, using an Agilent 1100 Series HPLC system equipped with an Agilent Eclipse XDB C-18 column (25 cm x 9.4 mm, 5 μm particle diameter). A 0.1% acetic acid in water (aqueous) – 9:1 acetonitrile:methanol (organic) solvent system was used, starting with 70% organic solvent for 3 min, which was increased linearly to 100% over a period of 20 min and continued at 100% organic solvent for 2 min. One-minute fractions were collected using a Teledyne ISCO Foxy 200 X-Y Fraction Collector connected to the HPLC from 12 to 36 min. Collected fractions were individually evaporated *in vacuo* for further analysis by GC-MS or NMR spectroscopy.

C.1.6. NMR spectroscopic instrumentation and analysis: NMR spectra were recorded on a Varian INOVA 600 MHz NMR spectrometer (600 MHz for 1H, 151 MHz for ^{13}C) equipped with an HCN indirect detection probe, and a Varian INOVA 500 MHz NMR spectrometer (500 MHz for ^{1}H, 125 MHz for ^{13}C) equipped with an DBG broadband probe. Each spectrum was manually phased, baseline corrected and calibrated to solvent peaks ($CHCl_3$ singlet at 7.26 ppm; CHD_2OD quintet at 3.31 ppm). Non-gradient phase-cycled dqfCOSY spectra were acquired using the following parameters: 0.8 s acquisition time, 500-900 complex increments, 16-64 scans per increment. Obtained dqfCOSY spectra were zero-filled to 8k-16k × 4k and a cosine bell-shaped window function was applied in both dimensions before Fourier transformation. NMR spectra were processed using Varian VNMR, MestreLabs' MestReC, and MestRecNova software packages. Dynamic range of the resulting spectra ranged from

300:1 to 500:1. For example, coupling constants could be determined for characteristic steroidal crosspeaks from dqfCOSY spectra containing as little as 5 μg $\Delta^{1,7}$-DA in a 2.5 mg enriched metabolome fraction.

C.1.7. *daf-9(dh6)* dauer rescue assay:

C.1.7.1. Plate-based assay: Metabolome fractions were resuspended in ethanol, mixed with 40 μL of 5 x concentrated OP50 bacteria (from an overnight culture in LB media) and plated on 3 cm plates containing 3 mL NGM agar without added cholesterol. For rescue, ~100 eggs from a 4-8 hour egg lay were transferred onto the bacterial lawn, and scored for dauer arrest at 27 °C after 60 h. For rescue experiments with synthetic steroids (0.1 nM – 500 nM tested), 10 μL compounds in ethanol (or ethanol alone) were mixed with 40 μL 5X concentrated OP50 bacteria and plated. Final concentrations include the total volume of agar (3 mL). 100 nM Δ^7-DA was used as positive control.

C.1.7.2. Liquid culture-based assay: Individual metabolome fractions were dried *in vacuo* and resuspended in 500 μL EtOH. 20-25 gravid *daf-9(dh6)* adults were picked onto three 6 cm NGM agar plates seeded with OP50 containing 25 μL of 10 μM Δ^7-DA each and allowed to grow for three days. On the third day each of the plates was washed with M9 and treated with alkaline hypochlorite solution to obtain eggs. Isolated eggs were allowed to hatch overnight in S-complete media without food. 100 μl of the resulting synchronized L1 suspension was added to 400 μL of HB101-seeded S-complete media and 5 μL of ethanol, ethanol solution of metabolome fraction, or ethanol

solution of synthetic ligands (1.25 nM – 200 nM tested) per well of a 12 well plate. Wells were examined after 48 and 72 h and scored for dauers, recovered animals and intermediate worms with molting defects and/or mig phenotypes. For active metabolome fractions, additional assays using smaller amounts were conducted.

Data obtained from both plate based and liquid culture assays are comparable. Figures in dissertation **Chapter 4** and **Appendix C** are based on results from the plate based assay.

C.1.8. Luciferase assay for DAF-12 transcriptional activation: Luciferase assays to determine transcriptional activation of DAF-12 were performed as described earlier.[1] Briefly, HEK-293T cells were seeded and transfected in 96-well plates with (per well) 30 ng transcription factor vector, 30 ng of GFP expression vector, 30 ng of luciferase reporter, and 5 ng β-galactosidase expression vector using the calcium phosphate precipitate method. Ethanol or ethanol solutions of ligands (synthetic DAs, 1 nM – 3125 nM tested and metabolome fractions) were added 8 h after transfection and the luciferase and β-galactosidase activities were measured by a Synergy 2 Biotek LC Luminometer, 16 h after compound addition. 100 nM Δ^7-DA was used as positive control. Data was processed using GEN5 software. Individual fractions were dried *in vacuo* and resuspended in 500 μL EtOH. 1 μL per 100 μL of media solution was added to each well.

C.1.9. Alphascreen assay for direct binding of DAF-12 ligand candidates: Direct binding of ligand candidates to DAF-12 was assessed by measuring ligand dependent interaction between a heterologously expressed DAF-12 construct and the mammalian coactivator peptide, SRC1-4, in Alphascreen assays (Perkin Elmer). An N-terminally GST-6xHis-tagged DAF-12 construct (amino acids 281-753, which includes most of the hinge region and the entire ligand binding domain, LBD) was purified from BL21 (DE3) cells (Sigma) using a GT sepharose column followed by a size exclusion column. Purified DAF-12 (25 nM final concentration) and biotinylated SRC1-4 (synthetic biotin-QKPTSGPQTPQAQQKSLLQQLLTE obtained from Anaspec, used at 50 nM final concentration), were incubated separately in plastic tubes for 20 min with Ni^{2+} chelate acceptor beads (for DAF-12) or streptavidin donor beads (for SRC1-4, 25 μg/mL final concentration, Perkin Elmer, Cat. No - 6760619C). They were then combined with subsequent addition of ligand candidates or ethanol (stored in glass vials) to obtain a final volume of 20 μL, and incubated for 60 min in white, low volume, 384 well Optiplates (Perkin Elmer). The plates were then read using a Synergy 2 Biotek LC luminometer using the manufacturer's Alphascreen detection protocol. All solutions were made using assay buffer containing 25 mM HEPES (pH 7.4) and 100 mM NaCl. Cholesterol and heptadecanoic acid, assayed at 1 μM concentration as additional negative controls, gave no signal (**Appendix Figure C.7b**). Addition of CHAPS (40 μM) did not significantly affect assay results, whereas addition of 0.01% BSA reduced signal. All incubations were carried out at room temperature in the dark. **Figure 4.5d** shows combined data derived from at least independent experiments (each N = 4 or higher) run on three or more different days.

C.1.10. GC/MS instrumentation and sample preparation: GC/MS analysis was carried out with an Agilent Technologies 6890N Network GC system with a DB-5MS+DG column (25 μm, 30 m x 0.25 mm) operating in split-less mode, connected to a JEOL JMS-GCmateII mass spectrometer.

Methylation of 3-keto-DAs: synthetic standards or 1-10 % of active metabolome fractions were evaporated *in vacuo* and resuspended in toluene:methanol (500 μL, 3:2), followed by the drop-wise addition of trimethylsilyldiazomethane (120 μL, 2 M solution in Et_2O) with stirring. The reaction was stirred at room temperature for 30 min, quenched with acetic acid, evaporated *in vacuo,* and resuspended in dichloromethane (30-200 μL). Of this solution, 1-5 μL were injected per GC run.

C.1.11. GC/MS methods:

C.1.11.1. GC conditions: Injector was kept at 240 °C and 1 mL/min helium flow was maintained. Initial column temperature was at 120 °C for 1.4 min, then increased to 320 °C at a rate of 7 °C/min, and maintained at 320 °C for 10 min.

C.1.11.2. MS conditions: Electron impact ionization (EI) at 70 eV was used. For synthetic samples (high concentrations, >100 ng/μL), MS was first operated in scanning mode for a mass range of *m/z* 35-500 for 3-keto-DA methyl ester and 3-keto-steroids, to select for the most abundant fragment ions for each compound. For low concentrations (<5 ng/μL) of synthetic samples and endo-metabolome fractions, MS was operated in Selective Ion Monitoring (SIM) mode and the following ions were selectively observed:

m/z = 428, 271, 229 (Δ^7-DA); 426, 269, 227 ($\Delta^{1,7}$-DA); 428, 305, 229, 124 (Δ^4-DA); and 384, 261, 229 (4-cholesten-3-one).

C.1.12. Quantification of DAs from endo-metabolome fractions via SIM-GC/MS: GC/MS data was analyzed using Shrader Analytical and Consulting Laboratories, Inc.'s TSSPro 3.0 software package. Quantification of DAs was performed by integration of GC/MS peaks from the following ion traces: *m/z* = 428 (Δ^7-DA), 426 ($\Delta^{1,7}$-DA), and 428 (Δ^4-DA). Dafachronic acid concentrations were calculated using response factors determined from synthetic standards. Mass spectrometer response was roughly linear (<5% error) for amounts of 10 pg to 5 ng per GC run per compound injection.

C.2. Figures:

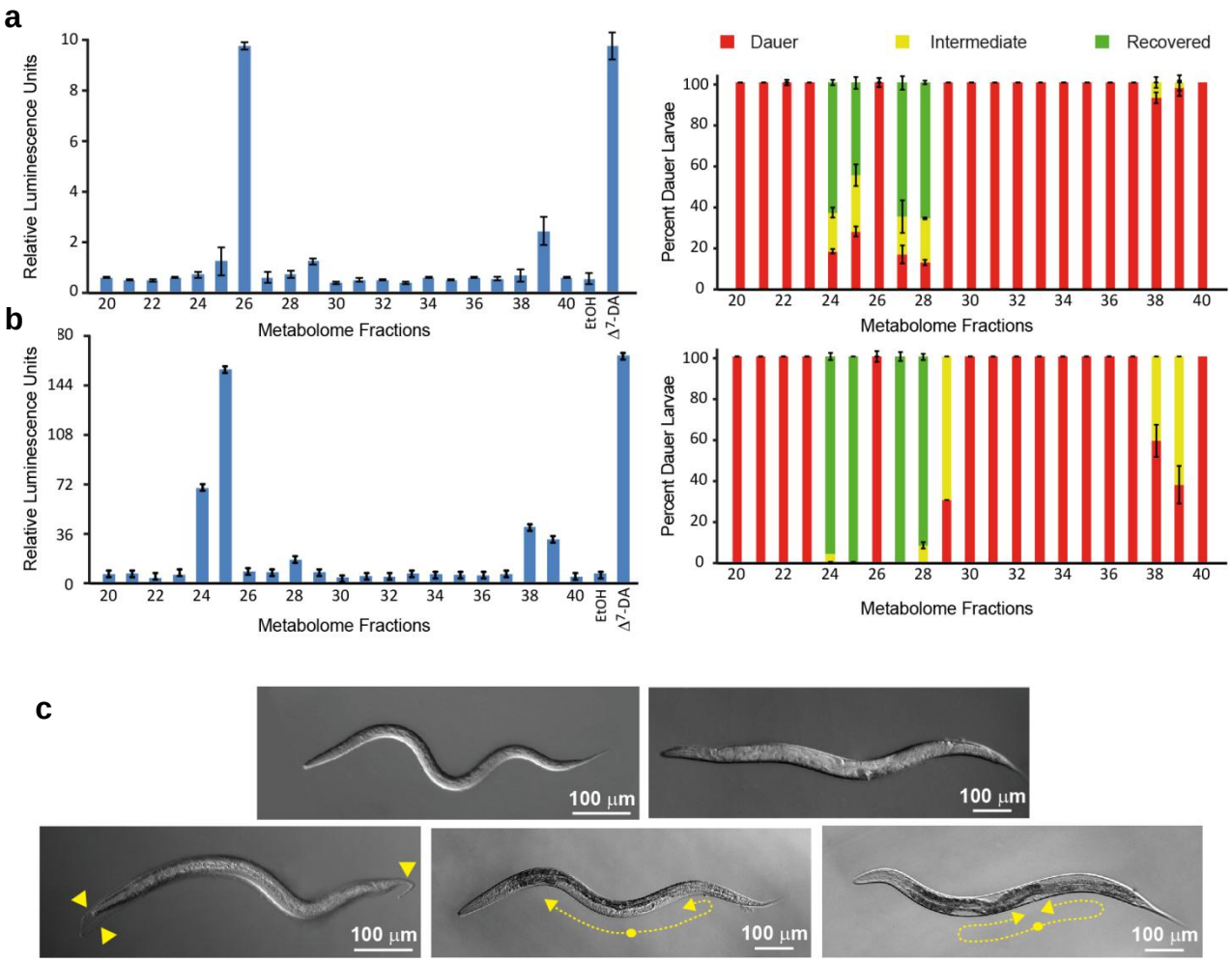


Figure C.1: Bioactivity of metabolome fractions. Assessment of DAF-12 transcriptional activation in the *in vitro* luciferase assay in HEK-293T cells (left) and dauer rescue activity in *daf-9(dh6)* worms (right) with endo-metabolome fractions obtained from (**a**) wild type (**b**) *daf-22* animals. These data show that *daf-9(dh6)* dauer rescue activity is generally higher in *daf-22* mutant metabolome fractions. (**c**) DIC images (courtesy, Dr. Joshua J. Wollam, Antebi research group, Max-Plank Institute for Biology of Ageing, Cologne, Germany) of *daf-9(dh6)* worms showing different phenotypes scored in the dauer rescue assay. Dauer (top left), adult recovered (top right), intermediate phenotypes showing molting defects (bottom left, arrows point to incompletely shed cuticle from previous molt), fully recovered late L4 (bottom center), and intermediate phenotypes showing incomplete gonad migration (bottom right). Yellow lines indicate gonadal shape in shown mig phenotypes.

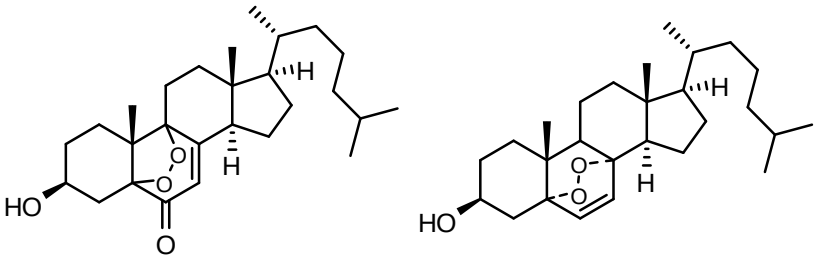


Figure C.2: *daf-9* independent sterols in metabolome fractions. Structures of major epidioxy sterol derivatives that elute with Δ^7 and $\Delta^{1,7}$-DAs in region I active fractions but can also be observed in corresponding inactive *daf-9;daf-12* fractions.[2,3]

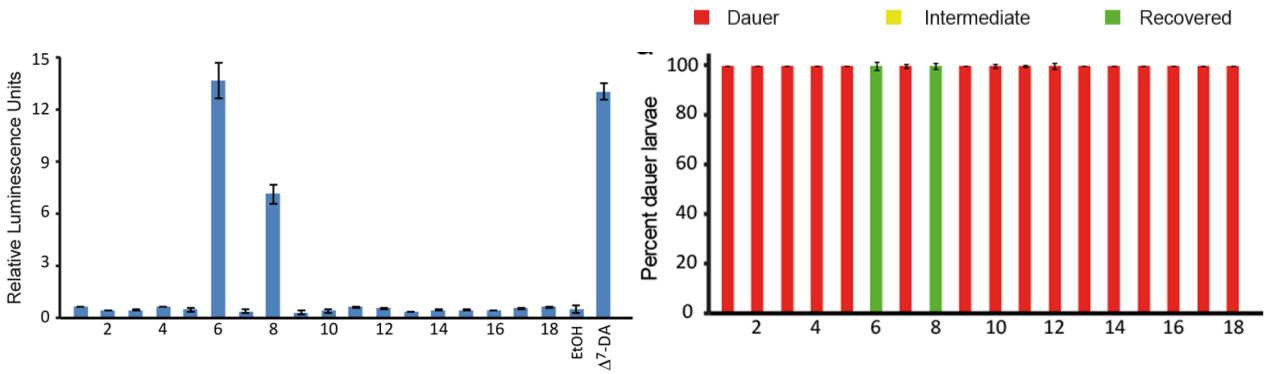


Figure C.3: Bioassays on HPLC-enriched fractions. Assessment of DAF-12 transcriptional activation in the *in vitro* luciferase assay in HEK-293T cells (left) and dauer rescue activity in *daf-9(dh6)* worms (right) with HPLC-enriched fractions 25-6 and 25-8 obtained from *daf-22* active metabolome fraction 25.

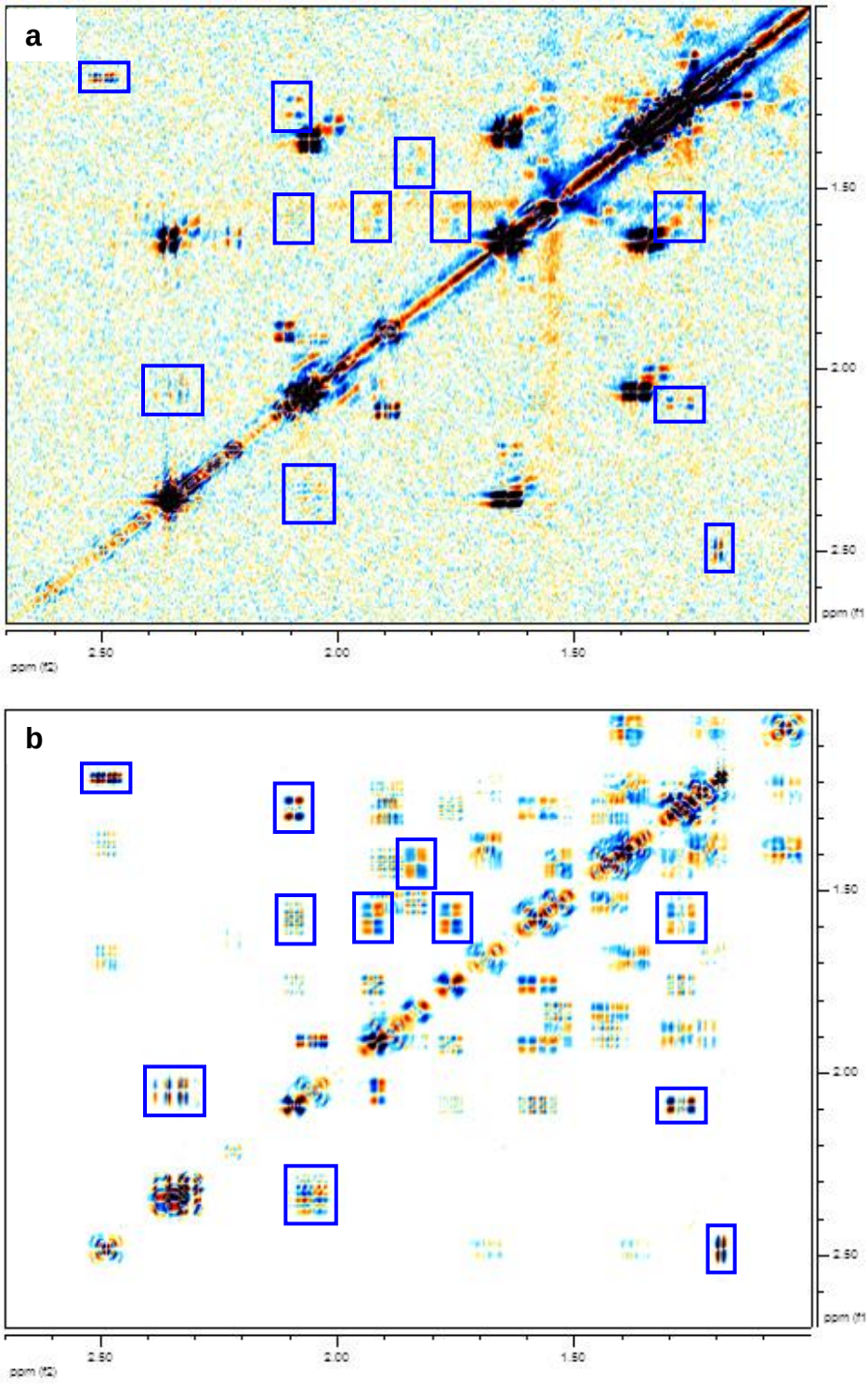


Figure C.4: Identification of endogenous *daf-9* dependant metabolites in active region I. Sections of dqfCOSY spectra (600 MHz, $CDCl_3$) used for identification of $\Delta^{1,7}$-DA in *C. elegans*. (**a**) Natural $\Delta^{1,7}$-DA in HPLC-enriched fraction 25-6 (**Figure 4.4**) from *daf-22* mutant. (**b**) Synthetic $\Delta^{1,7}$-DA. Characteristic crosspeaks are boxed blue.

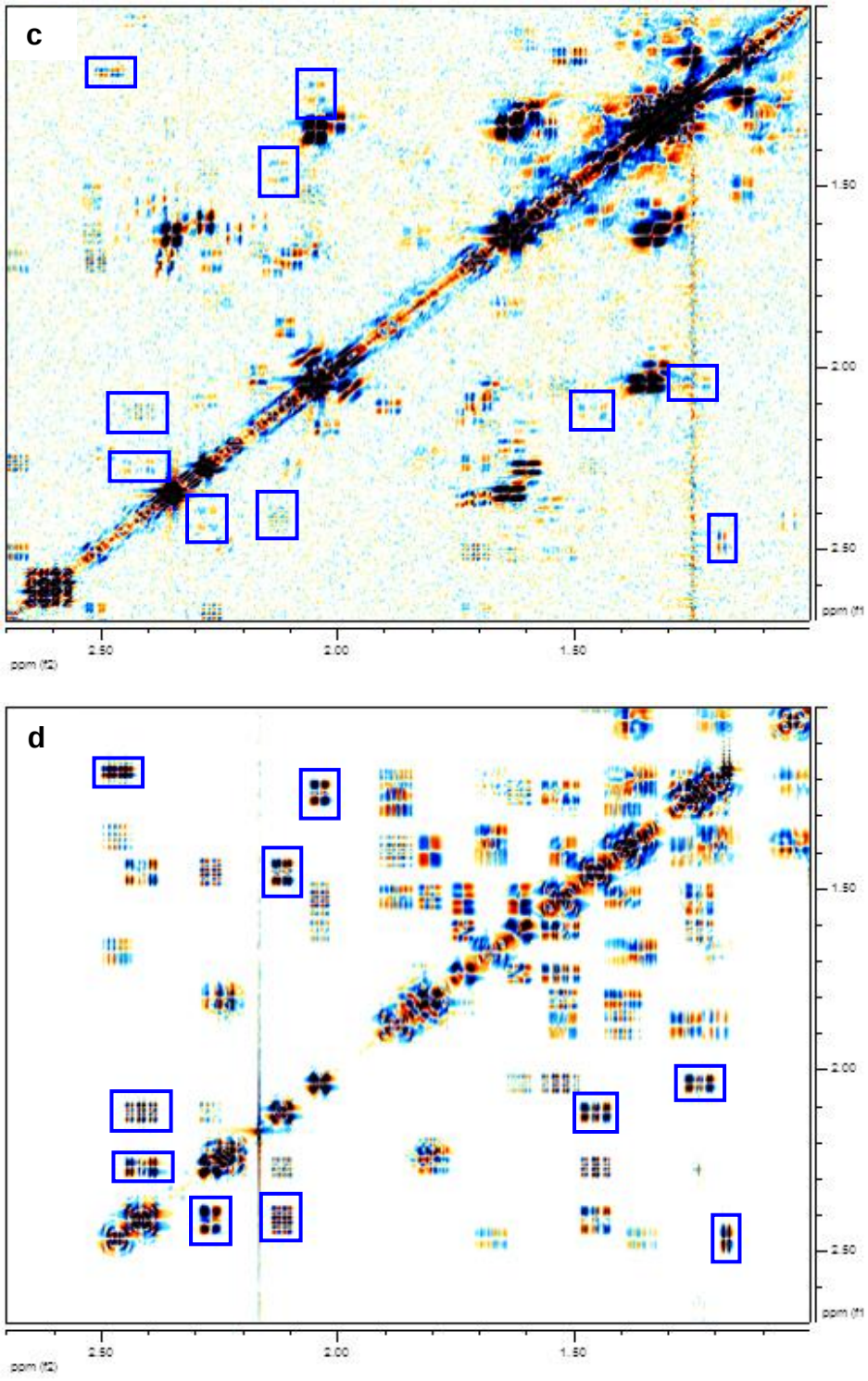


Figure C.4: Identification of endogenous *daf-9* dependant metabolites in active region I. Sections of dqfCOSY spectra (600 MHz, $CDCl_3$) confirming presence of Δ^7-DA in *C. elegans*. (**c**) Natural Δ^7-DA in HPLC-enriched fraction 25-8 from *daf-22* mutant (**Figure 4.4**). (**d**) Synthetic Δ^7-DA. Characteristic crosspeaks are boxed blue.

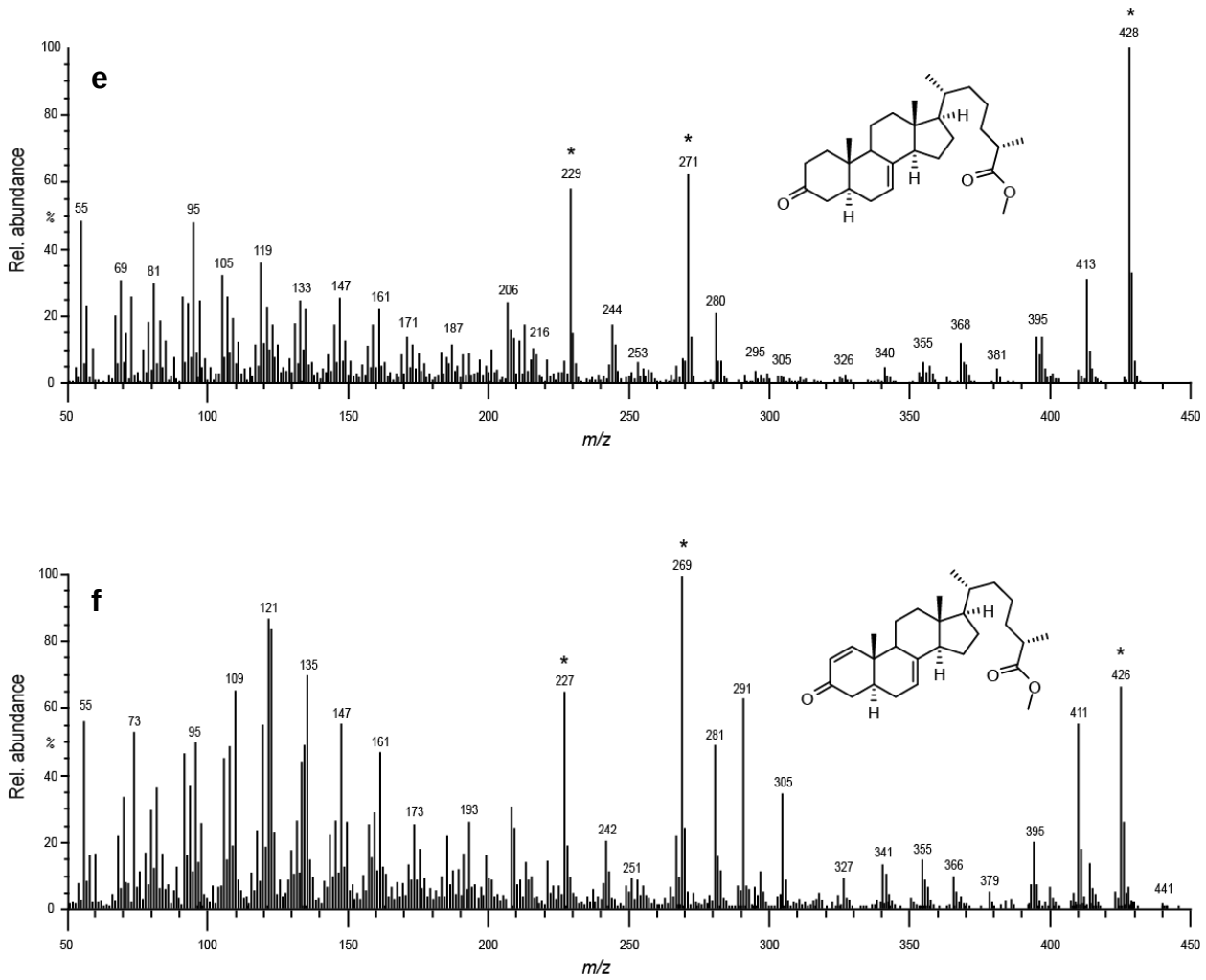


Figure C.4: Identification of endogenous *daf-9* dependant metabolites in active region I. EI-MS for methyl esters of (**e**) Δ^7-DA and (**f**) $\Delta^{1,7}$-DA. (*) Indicates the ion fragments subsequently used for SIM-GC/MS analyses of these compounds in endo-metabolome fractions.

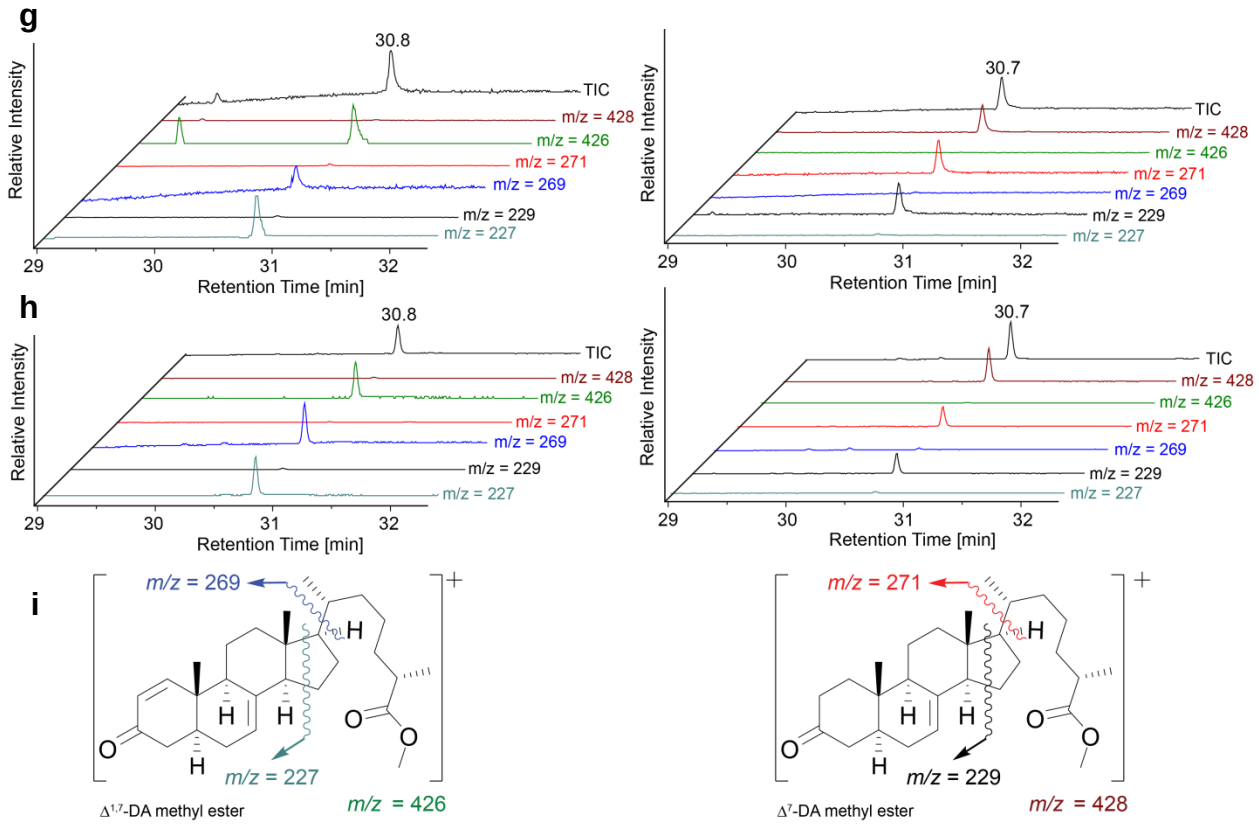


Figure C.4: Identification of endogenous *daf-9* dependant metabolites in active region I. SIM-GC/MS analyses showing characteristic fragment ion traces for methyl esters of (**g**) natural $\Delta^{1,7}$-DA (left), natural Δ^{7}-DA (right) in HPLC-enriched fractions 25-6 and 25-8 respectively from *daf-22,* (**h**) synthetic $\Delta^{1,7}$-DA (left), and synthetic Δ^{7}-DA (right). (**i**) Characteristic EI-MS fragments of $\Delta^{1,7}$-DA methyl ester (left) and Δ^{7}-DA methyl ester (right) shown in **Appendix Figures C.4g,h**.

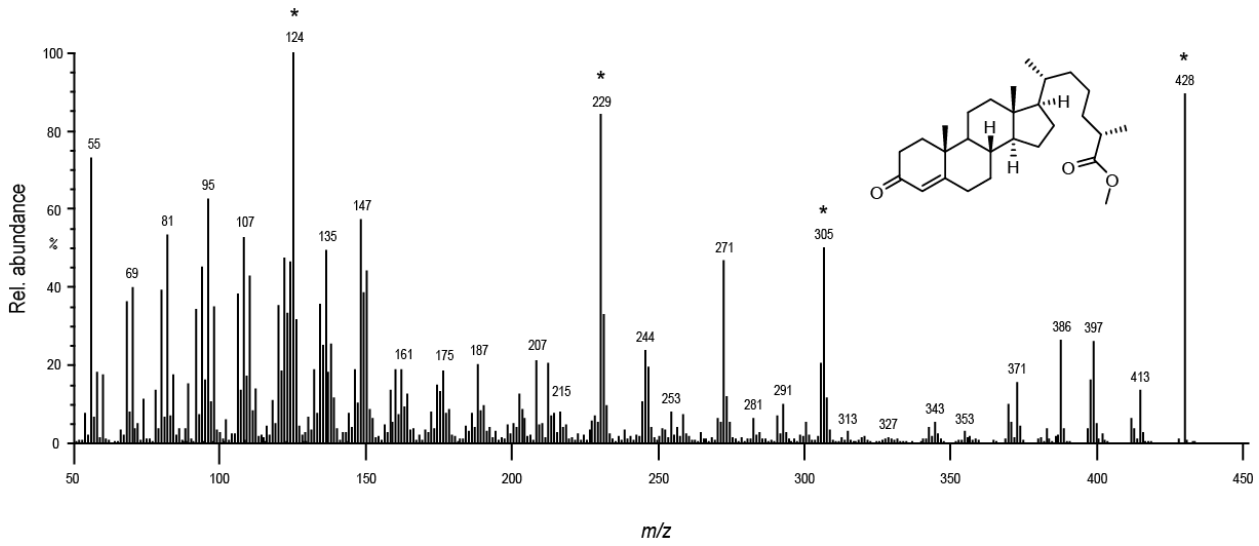


Figure C.5: EI-MS for Δ^4-DA methyl ester. (*) Indicates the ion fragments subsequently used for SIM-GC/MS analysis of this compound in natural samples. See also **Figure 4.6e**.

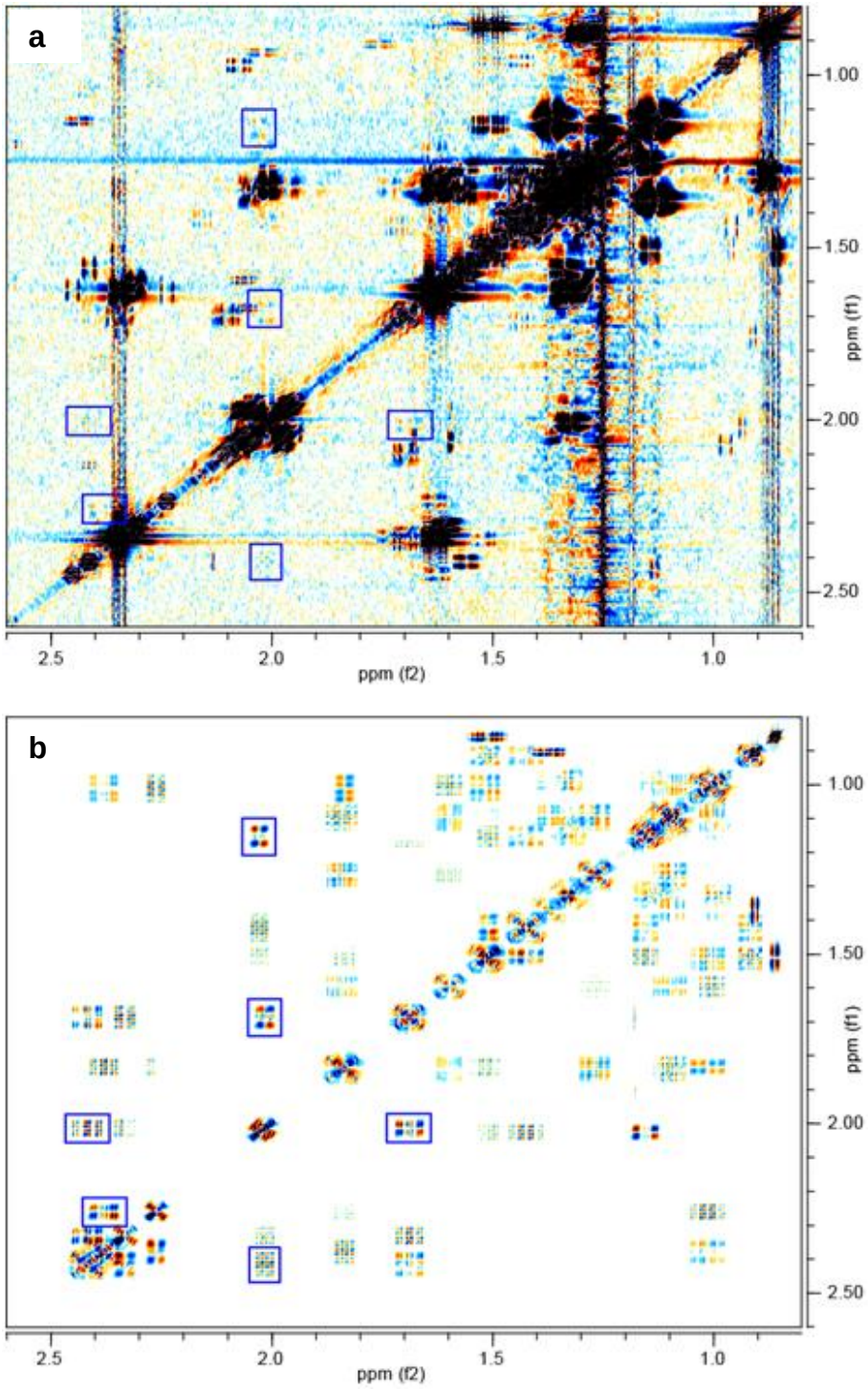


Figure C.6: Identification of dafachronic acid precursors. Sections of dqfCOSY spectra (600 MHz, $CDCl_3$) confirming presence of 4-cholesten-3-one in *C. elegans*. (**a**) 4-Cholesten-3-one in enriched natural fraction from *daf-22*. (**b**) Synthetic 4-cholesten-3-one. Characteristic crosspeaks are boxed blue.

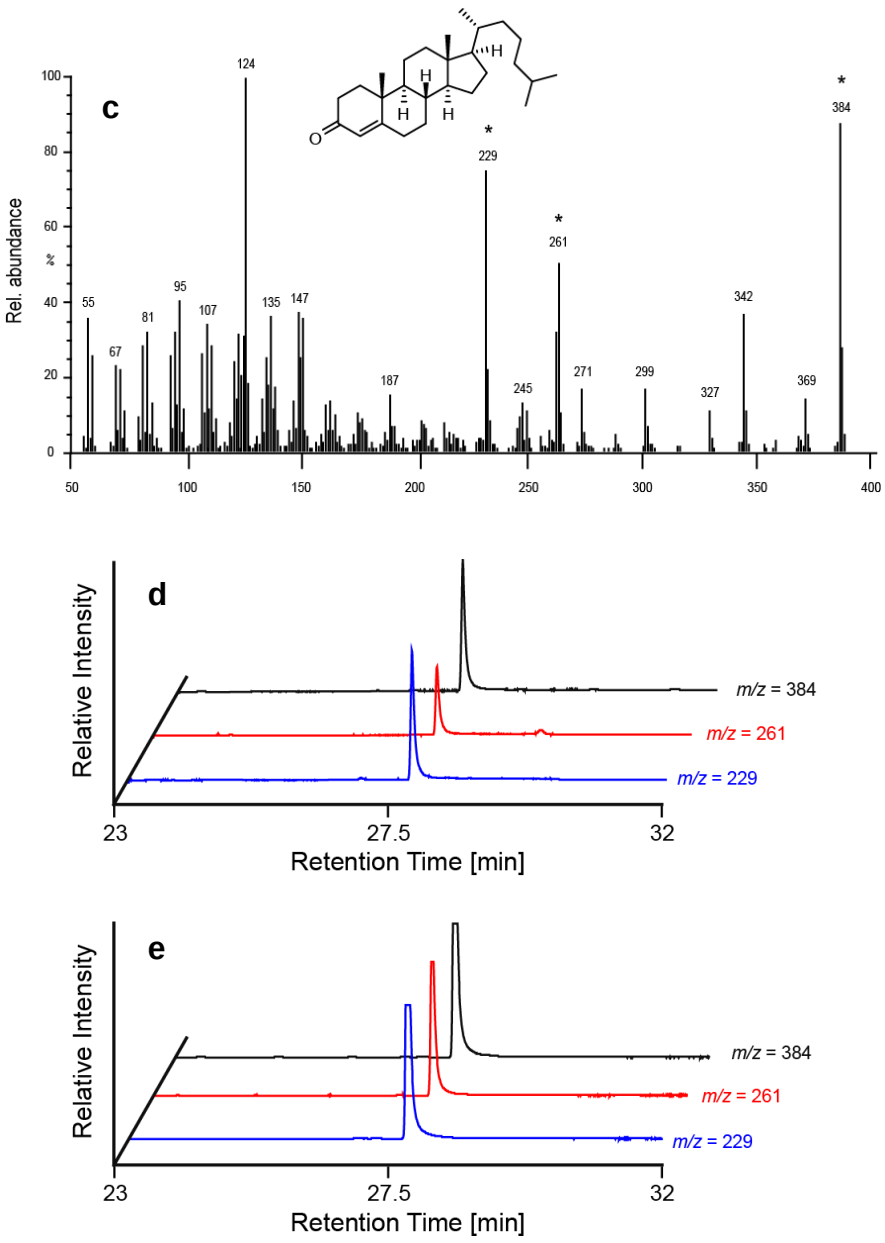


Figure C.6: Identification of dafachronic acid precursors. (**c**) EI-MS for 4-cholesten-3-one. (*) Indicates the ion fragments subsequently used for SIM-GC/MS analysis of this compound in natural samples. (**d**) SIM-GC/MS analysis showing characteristic fragment ion traces for natural fraction from *daf-22*, confirming the presence of 4-cholesten-3-one. (**e**) SIM-GC/MS analysis of a sample of synthetic 4-cholesten-3-one.

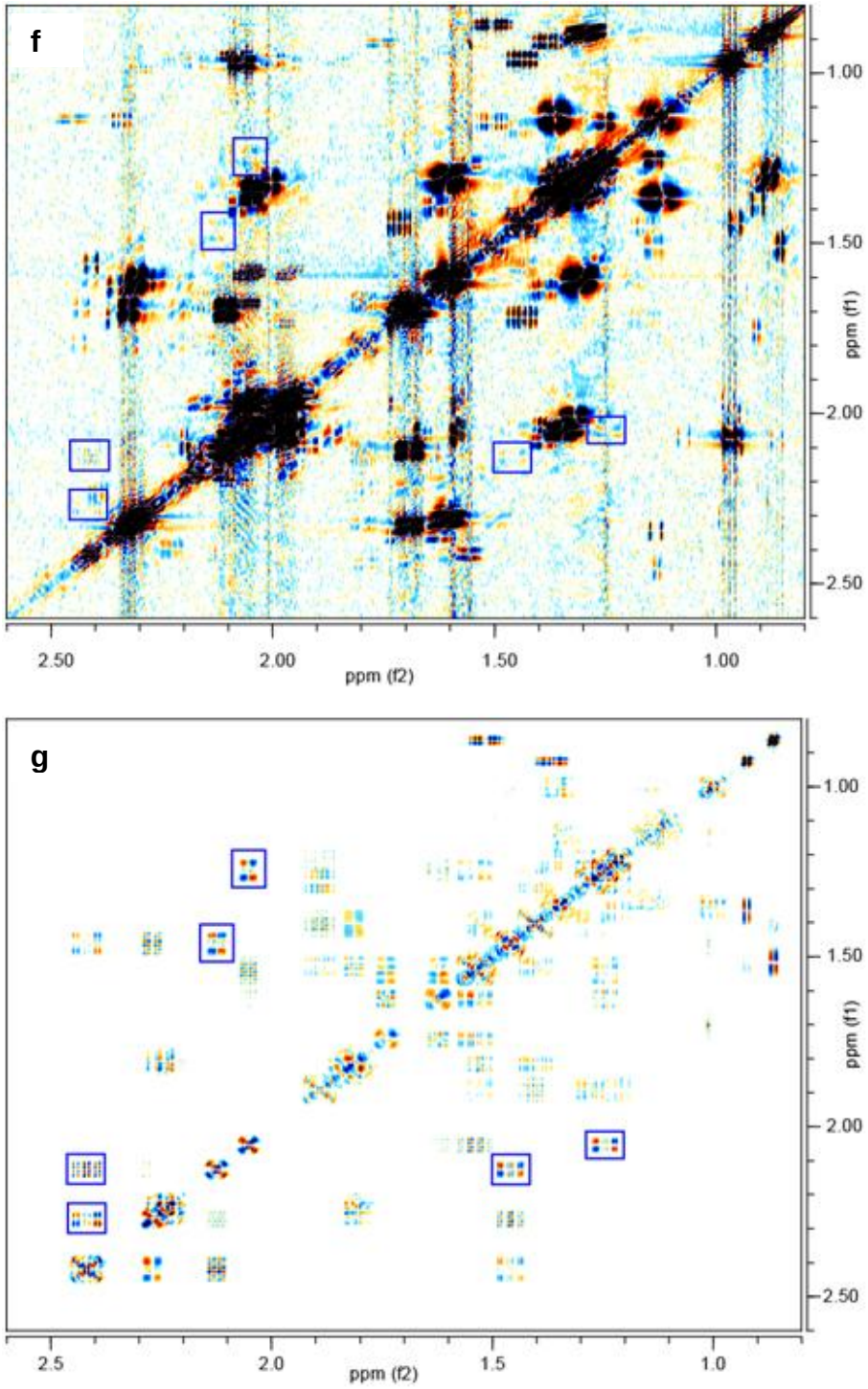


Figure C.6: Identification of dafachronic acid precursors. Sections of dqfCOSY spectra (600 MHz, $CDCl_3$) confirming presence of lathosterone in *C. elegans*. (**f**) Lathosterone in enriched natural fraction from *daf-22*. (**g**) Synthetic lathosterone. Characteristic crosspeaks are boxed blue.

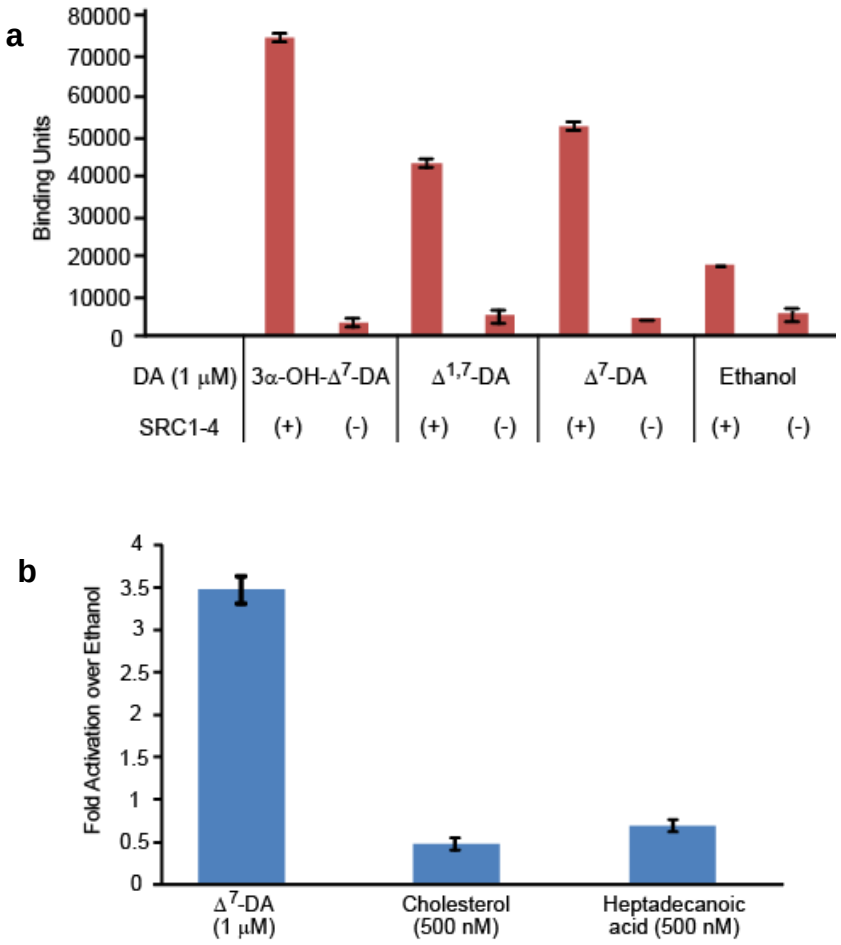


Figure C.7: Alphascreen assays with synthetic DAs. (**a**) DAF-12-ligand candidates produce SRC1-4 dependent signal in the Alphascreen assay. Shown are data obtained from one day of experiments (N = 4). **Figure 4.5d** shows combined data derived from at least independent experiments (each N = 4 or higher) run on at least three different days. (**b**) Cholesterol and heptadecanoic acid do not produce signal in the Alphascreen assay.

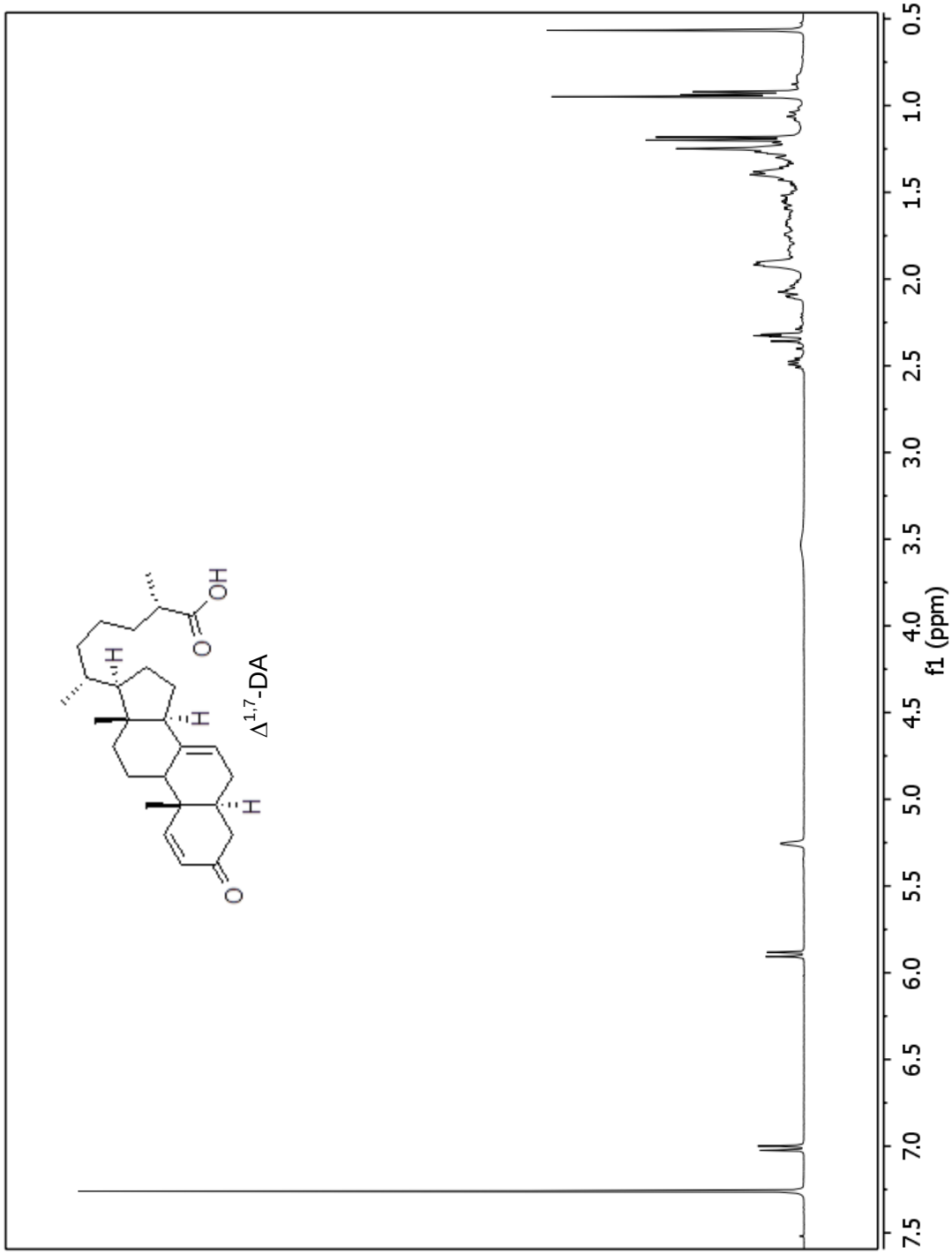


Figure C.8.1: ^{1}H NMR spectrum (600 MHz, $CDCl_3$) of (25*S*)-$\Delta^{1,7}$-Dafachronic Acid.

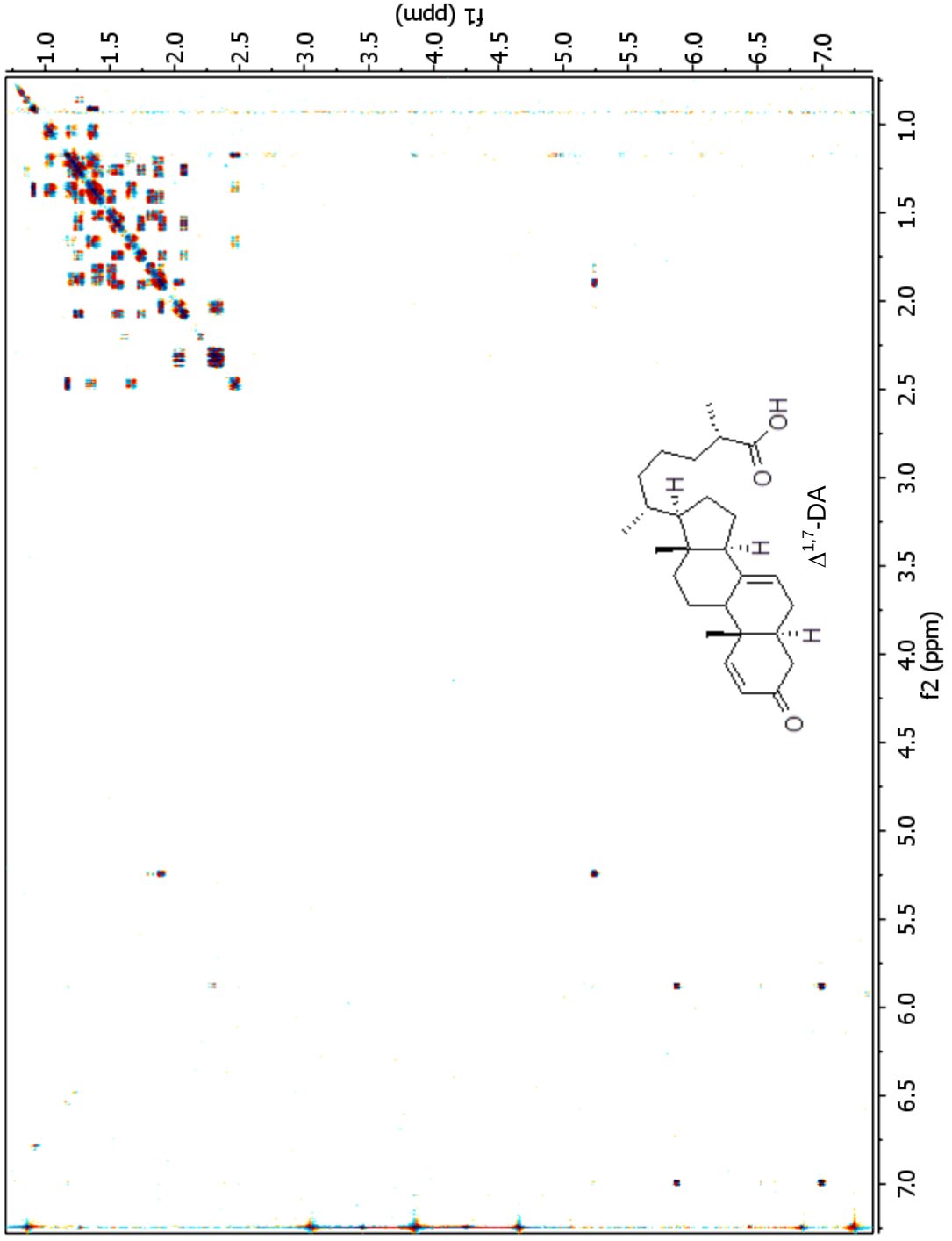


Figure C.8.2: dqfCOSY spectrum (600 MHz, $CDCl_3$) of (25*S*)-$\Delta^{1,7}$-Dafachronic Acid.

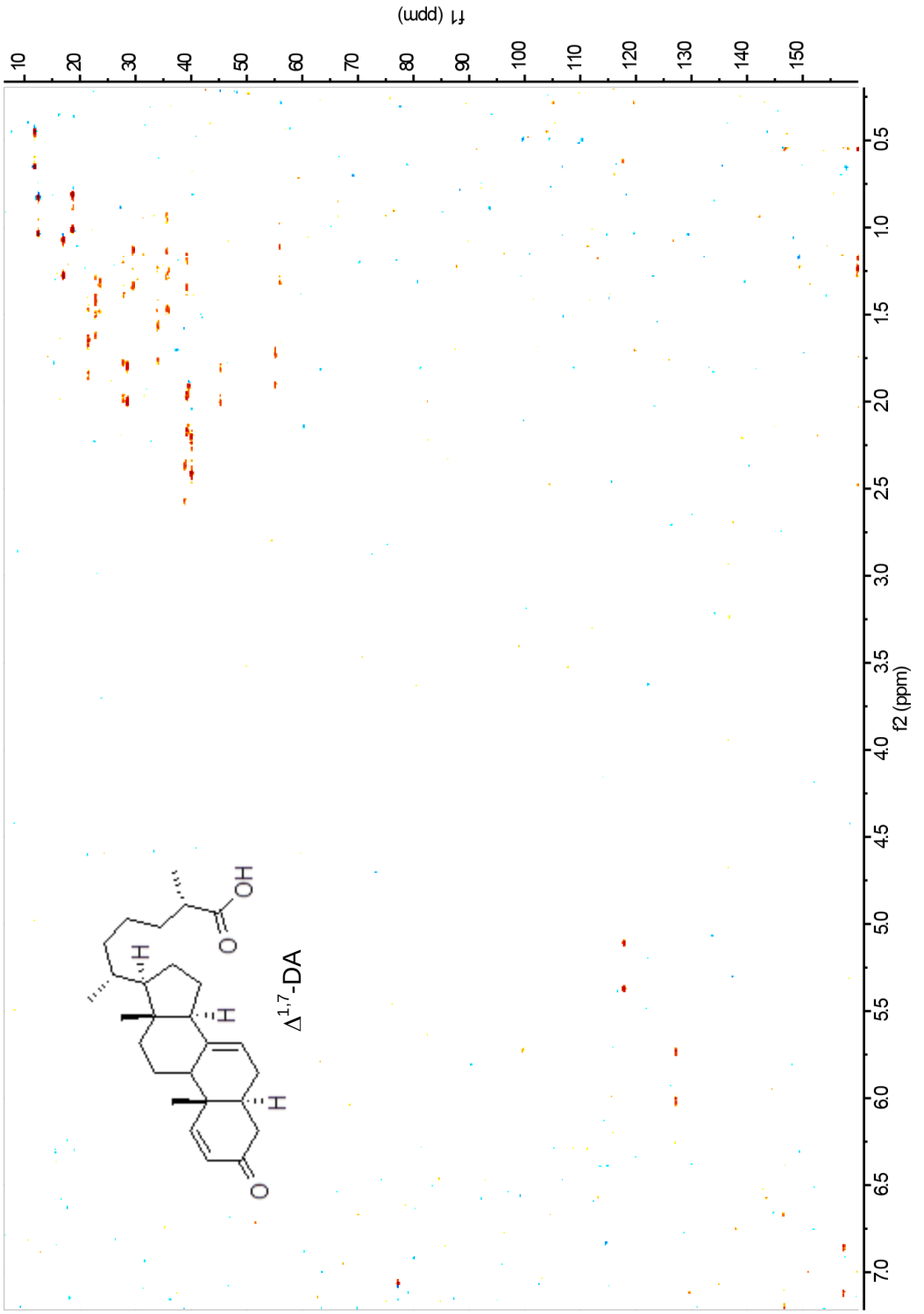


Figure C.8.3: HMQC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, $CDCl_3$) of (25*S*)-$\Delta^{1,7}$-Dafachronic Acid.

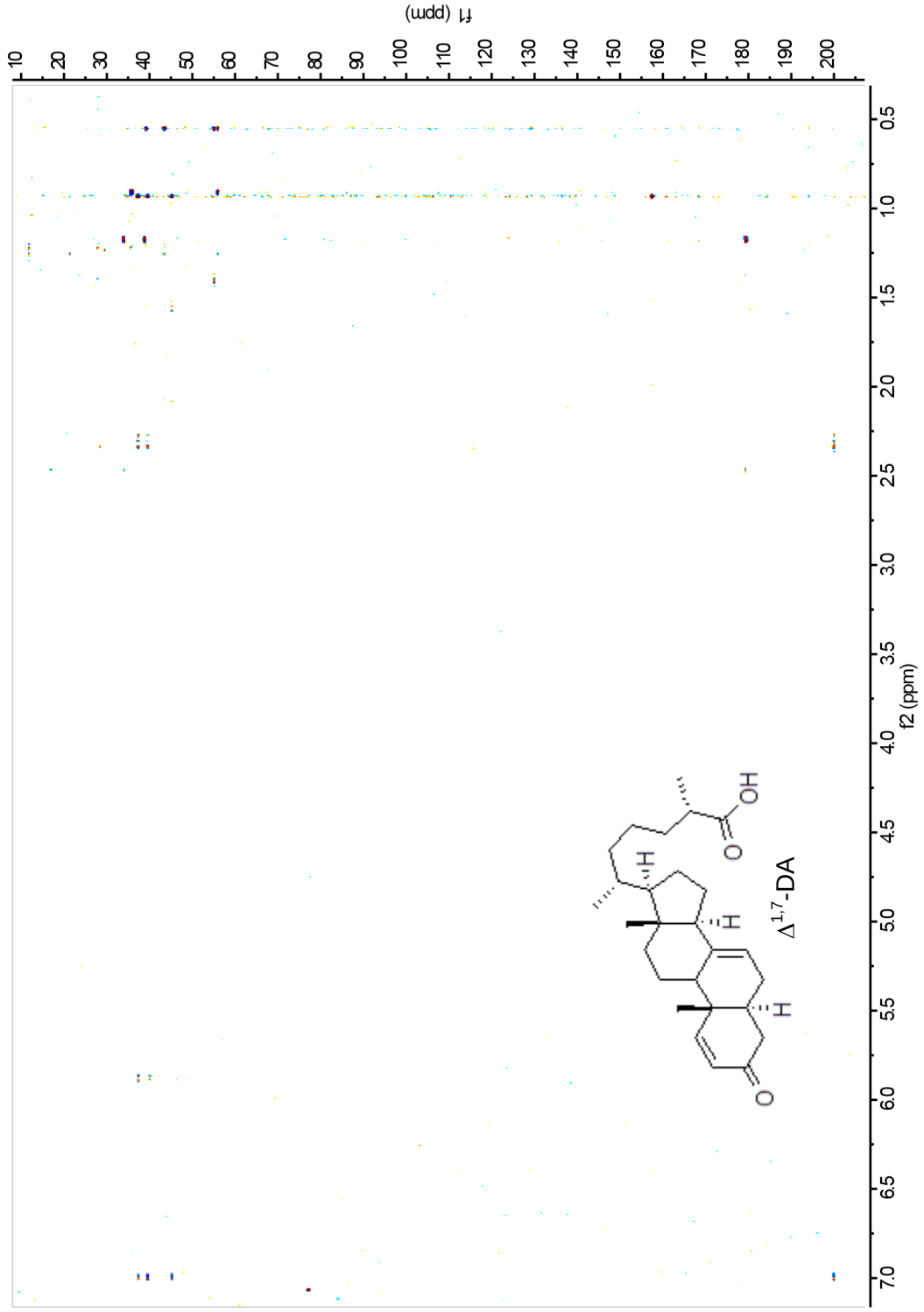


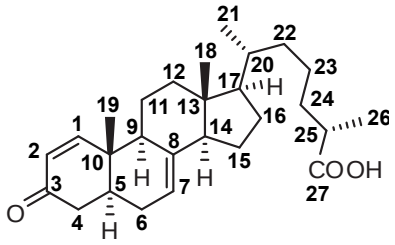
Figure C.8.4: HMBC spectrum (600 MHz for ^{1}H, 151 MHz for ^{13}C, $CDCl_3$) of (25*S*)-$\Delta^{1,7}$-Dafachronic Acid.

C.3. Tables:

Table C.1: Table showing EC_{50} values of synthetic DAs in luciferase, Alphascreen, and *daf-9(dh6)* dauer rescue assays and the HPLC retention times of synthetic DAs.

Compound	EC_{50} (nM) for luciferase assay (*in vitro*)	EC_{50} (nM) for Alphascreen assay (*in vitro*)	EC_{50} (nM) for *daf-9(dh6)* dauer rescue assay (*in vivo*)	HPLC retention time ranges (min)
$\Delta^{1,7}$-DA	146	15	2	20.5-20.9
Δ^{7}-DA	39	8	3	20.9-21.5

Table C.2: ^{1}H (600 MHz), ^{13}C (151 MHz), and important HMBC NMR spectroscopic data for $\Delta^{1,7}$-DA in $CDCl_3$. Chemical shifts were referenced to ($C\underline{H}Cl_3$) = 7.26 ppm and ($\underline{C}DCl_3$) = 77.16 ppm.



$\Delta^{1,7}$-Dafachronic acid

Carbon No.	δ (ppm)	Proton No.	δ (ppm)	J_{HH} (Hz)	Key HMBC Correlations
1	157.3	1-H	7.01	$J_{1,2}$=10	C-10, C-9, C-5, C-3
2	127.1	2-H	5.89		C-10, C-4
3	199.8	----	----	----	----
4	40.1	4-H_α	2.34		C-3, C-10
		4-H_β	2.34		C-3, C-10
5	39.6	5-H	2.05		
6	28.7	6-H_α	1.89-1.94		C-8
		6-H_β	1.89-1.94		
7	117.8	7-H	5.25		
8	138.8	----	----	----	----
9	45.2	9-H	1.93		

10	39.6	----	----	----	----
11	21.5	11-H_{α}	1.76		
		11-H_{β}	1.58		
12	39.2	12-H_{α}	1.27		
		12-H_{β}	2.09		
13	43.5	----	----	----	----
14	55.1	14-H	1.83		
15	23.0	15-H_{α}	1.53		
		15-H_{β}	1.42		
16	27.8	16-H_{α}	1.89		
		16-H_{β}	1.28		
17	55.9	17-H	1.82		
18	11.9	18-H	0.57	----	C-12, C-13, C-14, C-17
19	12.6	19-H	0.95	----	C-1, C-10, C-5, C-9
20	36.1	20-H	1.38		
21	18.7	21-H	0.93	$J_{21,22}$=6.4	
22	35.7	22-H_{α}	1.05		
		22-H_{β}	1.39		
23	23.9	23-H_{α}	1.21		
		23-H_{β}	1.39		
24	34.0	24-H_{α}	1.37		
		24-H_{β}	1.68		
25	38.9	25-H	2.48	$J_{25,26}$=6.9	C-23, C-24, C-27
26	17.0	26-H	1.19		C-25, C-27, C-24
27	179.3	----	----	----	----

REFERENCES

(1) Bethke, A.; Fielenbach, N.; Wang, Z.; Mangelsdorf, D. J.; Antebi, A. *Science* **2009**, *324*, 95.

(2) Gunatilaka, A. A. L.; Gopichand, Y.; Schmitz, F. J.; Djerassi, C. *J Org Chem* **1981**, *46*, 3860.

(3) Yaoita, Y.; Amemiya, K.; Ohnuma, H.; Furumura, K.; Masaki, A.; Matsuki, T.; Kikuchi, M. *Chem Pharm Bull* **1998**, *46*, 944.