STRATEGIES FOR PROFILING AND DISCOVERY OF ACYLSUGAR SPECIALIZED METABOLITES

By

Steven Michael Hurney

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree  of

Chemistry – Doctor of Philosophy

2018

**ABSTRACT**

STRATEGIES FOR PROFILING AND DISCOVERY OF ACYLSUGAR SPECIALIZED
METABOLITES

By

Steven Michael Hurney

Plant metabolic processes have evolved to produce an enormous array of chemically diverse metabolites
that serve important roles in growth, development and resistance, while offering nutritional, medicinal, and
economic value for humans. In the post-genome era, comprehensive measurements of plant chemistries,
termed plant metabolomics, have experienced rapid growth as a field of research, with a broader goal being
to provide the foundation for uncovering the relationship between the functions of genes, proteins and
enzymes responsible for the biosynthesis of metabolites. However, researchers in this field face formidable
challenges because unlike proteins and oligonucleotides which are constructed from a limited set of
precursors, the building blocks of metabolites are far more varied. As a result, annotation and identification
of novel metabolites remains the greatest obstacle to understanding the mechanisms responsible for
metabolite accumulation and the functional significance of new metabolites.

To address these challenges, the goals of this research have been to discover novel specialized metabolites
(those that are taxonomically restricted and not involved in central metabolism) to facilitate the discovery
of enzymatic processes responsible for assembly of plant chemical defenses, and to develop improved
methods and technologies for annotation, identification and dereplication of metabolite discovery.

Chapters 2 and 3 of this dissertation demonstrate approaches for untargeted metabolite profiling of
specialized metabolites extracted from leaf surface glandular trichomes (hair-like epidermal cells) from two
species of the family Solanaceae, *Salpiglossis sinuata* and *Solanum quitoense*. Liquid
chromatography/mass spectrometry (LC/MS) profiling revealed diverse multiply esterified sugar
metabolites, known as acylsugars, a family of metabolites known for anti-insect activity. Acylsugar
metabolites were purified and their structures were elucidated using one- and two-dimensional (1D and 2D)

nuclear magnetic resonance spectroscopy (NMR). These efforts established structures of 16 (of more than 400) new acylsucrose metabolites extracted from *S. sinuata* and established a novel group of acylated *myo*-inositols and *myo*-inositol glycosides (*N*-acetylglucosaminyl, glucopyranosyl and xylopyranosyl) from *S. quitoense* (9 structures). These results guided the discovery of previously unidentified acylsugar biosynthetic enzymes operating in *S. sinuata* and *S. quitoense* acylsugar biosynthetic pathways and extended our understanding of the evolution of specialized metabolism in the Solanaceae, while providing new analytical approaches for defining acylsugar composition and biodiversity.

Mass spectrometry-based platforms are extremely effective tools for metabolite detection and investigations of metabolomes. However, for many metabolites, mass spectrometry alone does not provide unambiguous metabolite identification. While NMR spectroscopy provides more detailed structural information than MS, some researchers avoid its use because it commonly involves time-consuming metabolite purifications. In addition, as more metabolite structures are determined, there is a growing chance that researchers will purify and identify compounds that are already known but thought to be novel isomers. To aid dereplication of metabolite discovery, a homologous set of *S*-alkyl glutathione (GS-*n*-alkyl) standards featuring normal saturated chain lengths (1-24 carbons) was synthesized and exploited as liquid chromatographic (LC) retention index standards (Chapter 4). These standards encompass a wide reversed phase-LC retention range, are easily ionized by electrospray ionization (ESI) in positive- and negative-ion modes, and show improved capacity for standardizing chromatographic retention. A thorough investigation of the dependence of acylsucrose retention index values using GS-*n*-alkyl standards was performed while altering several important chromatographic experiment parameters, including a comparison of columns, solvent delivery systems, aqueous mobile phase pH, column temperature, LC gradient slope, and organic solvent component. GS-*n*-alkyl standard LC/MS analysis also shows promise for evaluating RP-HPLC column performance, batch-to-batch column reproducibility, and degradation of column performance with use. This research has potential to improve interspecies SM metabolite discovery and dereplication.

# ACKNOWLEDGEMENTS

Before I joined MSU, I began my career as an Analytical Chemist as a Pre-Bachelor Analytical College Co-op in Food Packaging Research at The Dow Chemical Company. I am grateful for the time and effort that Analytical Chemists Thomas S. Lardie and Vickie Langer devoted to training me, as well as their friendship and encouragement. The skills and experience I acquired during our time together was immensely helpful in graduate school and set the stage for me to pursue my PhD in Chemistry. I also thank Dr. Kenneth Kearns (previously Professor of Analytical Chemistry at Saginaw Valley State University), who suggested during my undergraduate program that I should pursue an advanced degree in science and gave me confidence to do so.

Finally, I am fortunate to have the love and encouragement of wonderful family and friends. During my childhood, my parents praised learning and academic achievements. They provided me the confidence and inspiration needed to seek lofty educational goals. This accomplishment would not have been possible without their love and support. This achievement required considerable time and effort, at the expense of time enjoyed with my parents, siblings, family and friends. I sincerely appreciate their patience and repeated encouragement. I am especially grateful for the support and encouragement of my wonderful and loving wife, Dr. Krystyna Kijewska. I could not imagine my life without her. She makes everything that I do worthwhile. And… She fed me, did all the dishes and took care of me for months while I was writing. She deserves all the credit. Love you.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

[M+NH$_4$]$^+$ (*m/z* 836.427) for S7:27:0 showing one major and three minor chromatographic peaks; **(B)** Negative-ion mode chromatogram for [M+formate]$^-$ (*m/z* 863.392) on the same absolute y-axis scale (100% = 1.08 x 10$^4$ ion counts) shows negligible signal for S7:27:0. The structure shown is putative and not solved by NMR, but is proposed based on our knowledge of acyl group selectivity (established by NMR resolved structures of other acylsucroses). The [M+NH$_4$]$^+$ consistent with S7:27:0 and fragment ion *m/z* 373.15 (Function 3 = 25 V) are consistent with an acylsucrose that has a furanose ring substituted by one C5 and three C2 ester groups. ......................................................................................................................51

xix

**Chapter 1: Introduction**

We depend on plants as sources of food, medicine and renewable energy. Plants produce a boundless array of small molecules that have great economic value and play critical roles in plant growth, development and response to stresses [1]. The world's growing population requires more productive/higher yielding crops that show greater tolerance to climate changes and resistance to insects and pathogens [2]. In view of the importance of plants and the chemicals they synthesize, exploration of plant chemistries and genetic factors responsible for their biosynthesis could offer opportunities to genetically improve resistance and/or modify plants to become biosynthetic factories for production of useful natural products [3].

**1.1 Role of plant metabolomics for functional genomics and systems biology**

Metabolites are the small molecule (<1500 Da) output of biochemical processes in living things, yet our understanding of the roles that individual metabolic enzymes play in the grand scheme of metabolism is often unclear. The plant metabolome refers to the complete set of all small molecules produced within an organism and can be defined on all levels of complexity, such as the organism, tissues, cells or cellular compartments [4]. The metabolome is a representation of the chemical phenotype and is the result of expression of genetic code, metabolic networks and molecular activities. In the post-genome era, plant metabolomics has been a rapidly growing field of research [1, 2, 5]. There is a need for advancement of strategies for exploration of plant chemistries, to identify biological processes responsible for assembly of plant metabolites. Functional genomics and systems biology approaches aim to uncover the functions of genes, proteins, enzymes and their relationship to a plant's metabolic phenotype. To achieve this, it is necessary to integrate genomics, transcriptomics, and other 'omics' technologies with downstream metabolomics approaches. Despite great advancements in 'omics' technologies, most metabolites remain undefined. As a result, metabolite annotation and identification continue to present a bottleneck for determining plant metabolic gene functions.

**1.2 The evolution of plant specialized metabolism and natural chemical diversity**

Plants have evolved to synthesize an array of structurally diverse organic compounds including carbohydrates, amino acids, nucleic acids, and lipids to serve essential cellular functions. These central metabolites are produced using fixed carbon derived from $CO_2$, nitrogen-, sulfur-, and phosphorus-containing substances taken up by roots, and other precursors. Selection for increased fitness has resulted in plant lineages synthesizing distinct sets of metabolites that are not directly involved in primary metabolism. These compounds were typically referred to as secondary metabolites, in large part due to their unknown functions and presence is some species, but not in others. However, their presence or absence in different species make it clear that plants have evolved the ability to synthesize these compounds to satisfy specific ecological niches. For example, they serve as attractants for pollinators, and defenses against herbivory and microbes, and they provide communication with other plant and non-plant species [6-8]. In view of their taxonomic restriction and significance, more researchers have begun referring to these compounds as "specialized metabolites" (SMs).

The diversity of specialized metabolites and the observation of novel compounds in each species drive a true need for improved analytical strategies for defining the specialized metabolome in order to discover genes responsible for their biosynthesis [5]. Estimates suggest that perhaps 20% of plant genomes are involved in SM production and the number of SMs made by all plant species has been estimated to be in the hundreds of thousands [9]. However, there are reasons to believe this to be a gross underestimation [1, 8]. It is difficult to estimate the number of SMs in plant species, largely because the activities of proteins and enzymes encoded by the majority of genes in any genome are not well understood [8, 10]. Furthermore, existing analytical tools have yet to provide deep information about plant chemistry because most specialized metabolites have yet to be identified. From metabolite profiling, it is clear that each plant species can only make a small fraction of the total number of SMs found in the plant kingdom [8]. Most SMs are only found in a limited range of species and tissue types and as a result, the genes responsible for their chemistry are not well understood [8].

The evolution of so many different SM is considered the result of gene duplication [11]. Plants are organisms that exhibit extensive polyploidy (whole genome duplication, more than two sets of chromosomes). In fact, extensive genetic evidence suggests that whole genome duplication in angiosperms (flowering plants) has been common [12], providing potential novel opportunities for evolutionary success. Extensive gene redundancy 'shields' polyploids from the deleterious effects of mutations and provides the framework for diversification of gene function by altering redundant copies of important or essential genes [13]. When gene duplication leads to a situation where one copy of the gene retains the same function, and random mutation of the other copy alters gene product function (known as neofunctionalization), the result leads to variability in fitness across a population and serves as the foundation for natural selection. In particular, the presence of a mutant gene that leads to accumulation of SMs in individuals of the same species offers opportunities that reward SM production through advantageous fitness, particularly with regard to resistance to herbivory of pathogen infection. In practical terms, this may involve a mutant biosynthetic enzyme that catalyzes chemistry similar to the original (central metabolic) enzyme but exhibits altered substrate preference and leads to formation of novel SMs.

## 1.3 Challenges of identifying plant specialized metabolites

Plants produce far more metabolites than most other organisms and most have yet to be identified [1]. Identification of plant SMs presents many formidable challenges because unlike proteins and oligonucleotides which are constructed from a limited set of precursors, the building blocks of SMs are far more varied. Solvent extracts of plant tissues often present complex mixtures with hundreds or thousands of metabolites that exhibit large ranges in relative concentrations ($\sim 10^6$). Many SMs are present in a limited range of species and tissue types [8, 14], have numerous isomeric forms and few authentic standards are available for comparison. Analysis of plant SMs often requires multiple extraction, sample preparation and detection techniques because of the wide range of physical properties and chemical stability through extraction protocols. Furthermore, no single technology has yet been demonstrated that can detect, identify, and quantify all SMs from plants. Rather, a combination of analytical approaches is required. Multiple

chromatographic and mass spectrometric techniques are typically employed for comprehensive metabolite profiling. However, too often mass spectrometry alone is not sufficient to identify SMs. *De novo* structural elucidation of SMs usually requires time-consuming purifications for analysis by nuclear magnetic resonance (NMR) spectroscopy. In addition, analytical structural techniques require extensive training and experience. As a result, only a small fraction of the total number of SMs have been fully identified.

**1.4 Introduction to trichomes and acylsugars of the family Solanaceae**

**1.4.1 Trichome structures, functions and chemical compositions**

Specialized cell types are prolific at synthesizing SMs. A large fraction of plant species have glandular epidermal structures called trichomes [15]. These specialized uni- and multicellular epidermal protrusions are found on the surfaces of leaves, stems and other aerial tissues of plants and may differ in size, shape and morphology [15]. They are widely considered the plant's first line of defense, providing both physical and chemical barriers against attacking herbivores [16, 17], as well as guiding pollinators [7], protecting against UV-B radiation [18], leaf temperature stress and water loss [19].

The widespread abundance of glandular trichomes (GTs), found on ~30% of all vascular plants, suggests that diverse evolutionary forces shape specialized metabolism in plants [20]. Plants use GTs to synthesize, store and secrete large amounts of SMs [15, 20, 21]. The storage compartment of GTs is usually located at the tip of a hair-like cellular structure, permitting the plant to amass sticky and/or toxic secretions in a compartment that is remote from remaining plant tissues. GT exudates are harvested easily by scraping them from plant surfaces, gentle physical contact or dipping leaflets in to an organic solvent [22-24], permitting detailed study of their metabolites, as well as genes and proteins responsible for them. In addition, GTs exhibit novel biochemical pathways for specialized metabolism and have been suggested as targets for genetic engineering, with aims including optimization of trichome density, customization of essential oil production, and tuning of biocide activity to enhance crop protection [15, 21].

Humans are unknowingly familiar with GTs and their metabolites, even when we are unaware of their specific chemistries [15]. Trichome-derived SMs contribute to the aroma and flavor properties of many herbs and spices. For instance, members of the plant family Lamiaceae synthesize several SMs in peltate GTs, including terpenoids menthol and limonene in peppermint (*Mentha* × *piperita*) and phenylpropenes eugenol and methylchavicol in basil (*Ocimum basilicum*) [25]. A simple internet search of the word 'trichome' will invariably yield images of trichomes from *Cannabis sativa* of Cannabaceae, which produce and accumulate psychoactive cannabinoids and other volatile terpenoid-derived metabolites that contribute to its strong odor [26]. Another Cannabaceae species, hops (*Humulus lupulus*), produces xanthohumol and related prenylflavonoids that add bitterness and flavor to beer [27]. Trichomes are the source of the important medicinal chemical artemisinin of sweet wormwood (*Artemisia annua*). Artemisinin-based therapies are now standard treatment for malaria and have saved millions of lives [28].

**1.4.2 Acylsugars in glandular trichomes of Solanaceae species**

The family Solanaceae serves as an agronomically important family of flowering plants that includes many species. The family name is derived from its largest documented genus, *Solanum*, which includes important cultivated crops including tomato (*S. lycopersicum*), potato (*S. tuberosum*) and eggplant (*S. melongena*). Species of the Solanaceae display extensive diversity of trichome types and chemistry. Luckwill's 1943 taxonomic survey of tomato and its wild relatives found four major morphological types of GTs: type I consist of a globular and multicellular base with a small and round glandular cell in the trichome tip (6-10 cells and 2-3 mm long); type IV are similar to type I, but shorter (0.2-0.4 mm) and with a flat, unicellular base; type VI are thick and short, composed of two stalk cells and a head made of 4 secretory cells; type VII are very small (0.05 mm) with a head consisting of 4-8 cells [21, 29].

Members of the family Solanaceae synthesize and/or store diverse arrays of SMs in GTs, including terpenes, phenylpropenes, flavonoids, alkaloids and fatty acid derivatives [15, 21]. Species within the genera *Solanum*, *Petunia*, and *Nicotiana* (among others) produce large quantities of structurally diverse sugar polyesters in type I & IV trichomes [21]. These SMs are comprised of a carbohydrate core esterified by one

or more ester groups (Figure 1.1) at hydroxyl positions around the sugar ring. Various reports have described these SMs as sugar esters, *O*-acyl sugars or acylsugars. Documented acylsugars of the Solanaceae are mostly sucrose or glucose sugar groups, consisting of three to five branched and/or straight aliphatic C2-C12 acyl chains (Table 1.1). These SMs play critical roles for physical and/or chemical defense, including insect resistance [30-32], oviposition suppression [33], protection against fungal pathogens [34], and indirect defense against herbivory [35].

Most descriptions of acylsugar composition in GTs report only a few acylsugar metabolites per species (see Table 1 for references). However, it is clear that that this is only the tip of the iceberg, as many species are prolific manufacturers of diverse acylsugars [36], most of which have not been identified. In part this is because acylsugars are complex mixtures that display multiple isomeric forms, structural redundancy of similar functional groups and unique acylation patterns, making structural elucidation formidable [23]. For instance, sucrose-based acylsugars (acylsucroses) demonstrate up to eight hydroxyl positions available for acylation (Figure 1.1). To the best or our knowledge, evidence of 25 different acylsugar acyl groups has been reported in published papers and herein (acyl groups and their acronyms are shown in Figure 1.1). From the known acyl group building blocks alone, the number of possible acylsucrose permutations exceeds 46 billion (Equation 1.1). Only ~100 acylsugars have been identified from different species within Solanaceae (Table 1.1), but explorations of acylsugar diversity are still at an early stage. Since the Solanaceae display capabilities to synthesize precursors, mutations in the enzymes that assemble acylsugars offer the prospect that an enormous number of acylsugars may exist across the Solanaceae, suggesting that diversity of acylsugars has potential to far exceed the range of terpenoid SMs, documented in 1992 to exceed 22,000 [37].

**Equation 1.1.** Number of possible acylsucrose permutations

$$Possible\ acylsucroses = \sum_{r=1}^{8} \frac{25!}{(25-r)!} = 4.6 \times 10^{10}$$
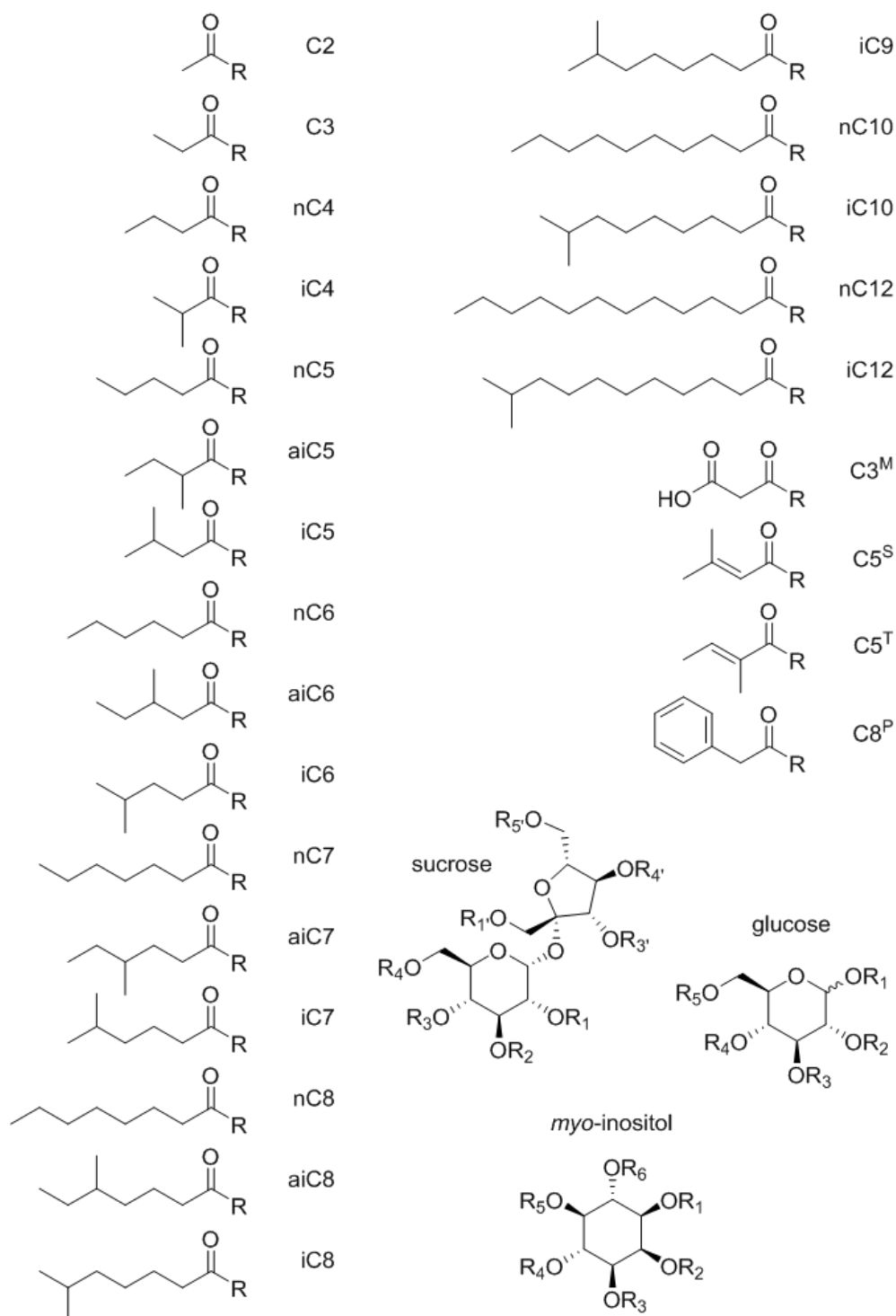
**Figure 1.1.** Acylsugar ester groups and carbohydrates observed within Solanaceae species. C2: acetyl, C3: propionyl, nC4: butyryl, iC4: isobutyryl, nC5: pentanoyl, aiC5: 2-methylbutanoyl, iC5: 3-methylbutanoyl, nC6: hexanoyl, aiC6: 3-methylpentanoyl, iC6: 4-methylpentanoyl, nC7: heptanoyl, aiC7: 4-methylhexanoyl, iC7: 5-methylhexanoyl, nC8: octanoyl, aiC8: 5-methylheptanoyl, iC8: 6-methylheptanoyl, iC9: 7-methyloctanoyl, nC10: decanoyl, iC10: 8-methylnonanoyl, nC12: dodecanoyl, iC12: 10-methylundecanoyl, C3$^M$: malonyl, C5$^S$: senecioyl, C5$^T$: tiglyl, C8$^P$: phenylacetyl.

7

**Table 1.1.** Examples of leaf surface acylsugars from Solanaceae species with identified positions of acylations and their residues

| Species | Carbohydrates | No. of acylsugars identified | No. of acylations | Positions of acylations | Acyl residues (certain acyl residues may occur in multiple positions or show position specificity) | References |
|---|---|---|---|---|---|---|
| **Solanum** | | | | | | |
| *S. lycopersicum* M82 | sucrose | 15 of 24 | 3,4 | 2,3,4,3' | C2, iC4, iC5, aiC5, iC10, aiC11, nC12, iC12 | Ghosh et al. 2013 [38] |
| *S. habrochaites* LA 1392 and LA1362 | sucrose | 24 of 24 | 3,4,5 | 2,3,4,1',3',6' | C2, iC4, iC5, aiC5, iC10, aiC11, nC12, iC12 | Ghosh et al. 2013 [38] |
| *S. habrochaites* LA 1777 | sucrose | 22 of 24 | 3,4,5 | 2,3,4,1',3',6' | C2, iC4, iC5, aiC5, iC10, aiC11, nC12, iC12 | Ghosh et al. 2013 [38] |
| *S. lycopersicum* M82 and IL-3 | sucrose | 2 | 3,4 | 2,3,4,3' | C2, iC5, aiC5 | Schilmiller et al. 2010 [24] |
| *S. lycopersicum* IL11-3 | sucrose | 1 | 3 | 3,4,3' | iC5, iC5, nC12 | Schilmiller et al. 2015 [39] |
| *S. pennellii* LA0716 | glucose, sucrose | 2 | G3, S3 | 2,3,4 | iC4, iC5, aiC5, iC10 | Schilmiller et al. 2016 [40] |
| *S. habrochaites* (previously *Lycopersicon hirsutum*) | sucrose | 1 | 4 | 2,3,4,1' | iC4, aiC5 | King et al. 1990 [41] |
| *S. habrochaites* LA 1777 (previously *Lycopersicon typicum*) | sucrose | 6 | 3,4 | 2,3,4,6,3' | C2, iC4, aiC5, iC5, iC10 | King et al. 1993 [42] |
| *S. berthaultii* | sucrose | 1 | 3 | 3,4,1' | iC4, iC4, nC10 | King et al. 1986 [43] |
| *S. berthaultii* | sucrose | 2 | 4 | 3,4,6,3' | iC4, aiC5, nC10 | King et al. 1987 [44] |
| *S. neocardenasii* | sucrose | 3 | 4 | 2,3,4,3' | C2, iC4, nC6, nC10 | King and Singh 1988 [45] |
| *S. aethiopsicum* | glucose | 3 | 3 | 2,3,4 | C2, iC4, aiC5, iC9, iC10 | King et al. 1988 [46] |
| *S. pennellii* (previously *Lycopersicon pennellii*) | glucose | N/A | 3 | 2,3,4 | iC4, iC5, iC5, aiC5, nC10, iC10 | Burke et al. 1987 [47] |
| *S. pennellii* (previously *Lycopersicon pennellii*) | glucose | 1 | 3 | 2,3,4 | iC4 | Li and Steffens 2000 [48] |
| *S. lanceolatum* | *myo*-inositol glycosides | N/A | 1 | 2 | C12 to C20, even numbered branched or straight chain (glucopyranosyl or xylopyranosyl at position 1) | Herrera-Salgado et al. 2005 [49] |
| *S. quitoense* | *myo*-inositol glycosides | 9 | 3,4 | 1,2,3,4 | nC2, nC10, nC12 (*N*-acetylglucosamine, glucopyranosyl or xylopyranosyl at position 4) | Hurney et al. (Chapter 3) |

(continued on following page)

**Table 1.1.** (continued)

| Species | Carbohydrates | No. of acylsugars identified | No. of acylations | Positions of acylations | Acyl residues (certain acyl residues may occur in multiple positions or show position specificity) | References |
|---|---|---|---|---|---|---|
| **Petunia** | | | | | | |
| *P. integrifolia* | sucrose | 21 of 26 | 4,5 | 2,3,4,6,1' | C3[M], iC4, aiC5, iC5, iC6, iC7, iC8 | Liu et al. 2017 [50] |
| *P. exserta* | sucrose | 26 of 26 | 4,5 | 2,3,4,6,1',6' | C3[M], iC4, aiC5, iC5, iC6, iC7, iC8 | Liu et al. 2017 [50] |
| *P. axillaris* | sucrose | 21 of 26 | 4,5 | 2,3,4,6,1' | C3[M], iC4, aiC5, iC5, iC6, iC7, iC8 | Liu et al. 2017 [50] |
| *P. nyctaginiflora* | sucrose | 2 | 3 | 2,3,4 | iC6, iC7 | Singh et al. 2003 [51] |
| *P. nyctaginiflora* | sucrose | 4 | 3,4 | 2,3,4 | C2, iC6, iC7 | Begum et al. 2004 [52] |
| *P. nyctaginiflora* | sucrose | 2 | 3,4 | 2,3,4,6' | C2, iC7, iC8 | Begum et al. 2005 [53] |
| **Nicotiana** | | | | | | |
| *N. tabacum* TI-165 | sucrose | 1 | 4 | 2,3,4,6 | C2, iC4, aiC5 | Severson et al. 1985 [54] |
| *N. glutinosa* | sucrose | N/A | 3,4 | 2,3,4,6,3' | C2, C3, iC4, aiC5, iC5, nC5, aiC6, iC6, nC6, aiC7, iC7, nC7, iC8, nC8 | Arrendale et al. 1990 [55] |
| *N. acuminata* | glucose, sucrose | N/A | G3,G4,G5,S5 | 1,2,3,4,6,3' | C2, C3, iC4, aiC5, iC6, aiC6 and unsaturated aiC5 | Matsuzaki et al. 1991 [56] |
| *N. pauciflora* | glucose, sucrose | N/A | G3,G4,G5,S5 | 1,2,3,4,6,3' | C2, C3, iC4, aiC5, iC6, aiC6, iC7, aiC7 | Matsuzaki et al. 1991 [56] |
| **Physalis** | | | | | | |
| *P. philadelphica* | sucrose | 5 | 2,3,4 | 2,3,1',3' | iC4, aiC5, nC10 | Zhang et al. 2016 [57] |
| *P. viscosa* | sucrose | 1 | 2 | 3,4 | C5[S], nC12 | Ovenden et al. 2005 [58] |
| **Other Solanaceae** | | | | | | |
| *Datura metel* | glucose | 2 | 2,3 | 1,2,3 | nC6 | |
| *Salpiglossis sinuata* | sucrose | 16 | 4,5,6 | 2,3,4,1',3',6' | C2, C3, iC4, aiC5, aiC6, C5[T], C8[P] | Hurney et al. (Chapter 2) |
| *Salpiglossis sinuata* | glucose, sucrose | 2 | G3,S5 | 2,3,4,6,3' | C2, aiC6 | Castillo et al. 1989 [59] |

\*\*\* Updated table [60]. N/A indicates no structures were identified, however ester substitution positions were assigned from mixtures using NMR spectroscopy.

### 1.4.3 Biosynthesis of acylsugars

Acylation of oxygen-containing functional groups serves as a widespread mechanism for assembly of plant SMs [61]. Until recently, the functions of individual enzymes required for acylsugar assembly were largely undefined. The combination of metabolite profiling and identification [38, 50] with analysis of gene expression and biochemistry has guided the discovery of trichome-expressed enzymes required for acylsucrose assembly within the Solanaceae [39, 62-64]. Acylsugars are synthesized enzymatically by a group of enzymes called BAHD acyltransferases that catalyze the position selective acylation of sugar cores by acyl-CoAs. This family was named according to the first letter of each of the first four biochemically characterized enzymes of this family (BEAT, AHCT, HCBT, and DAT) [65].

In tomato (*S. lycopersicum*), acylsucrose assembly begins with acylation of the pyranose ring of sucrose at position 4 by the enzyme AcylSugar AcylTransferase 1 (Sl-ASAT1) using various acyl-CoAs, which serve as acyl group donors, to make monoacylsucroses [63]. ASAT nomenclature has been proposed that numbers the enzymes in the order they participate in acylsucrose biosynthesis, with ASAT1 being the first. Sl-ASAT2 then adds a second acyl group (iC4, aiC5, iC5, iC6, iC10 or nC12, as defined in Figure 1.1) to position 3 of the Sl-ASAT1 product to make diacylsucrose substrates for Sl-ASAT3 [63]. Subsequent catalysis of acylation by Sl-ASAT3 transfers short (four to five carbon) branched acyl chains to the furanose ring at position 3′ of diacylsucrose acceptors, producing triacylsucroses [39], which can be further acylated by Sl-ASAT4 (previously Sl-AT2) which catalyzes the transfer of a C2 group to position 2 of tri-acylated sucrose acceptors to form tetra-acylated sucroses [62]. Interestingly, Sl-ASAT1 showed substrate promiscuity *in vitro*, acylating sucrose at position 4 using acyl-CoAs with different acyl chain lengths (iC4, iC5, aiC5, nC10 and iC12). However, only iC4 and iC5 acyl groups have been observed at position 4 of acylsucroses from tomato species [38]. Furthermore, ASAT2 was unable to use iC12 monoacylsucroses to produce diacylsucroses, suggesting that longer chain monoacylsucroses represent dead-end products that are recycled by an acylsugar acylhydrolase to generate the starting metabolite sucrose [40]. In such manner,

acylsucroses may be enzymatically remodeled, leading to accumulation of a narrower distribution of acylsucroses that do not include less-stable intermediate forms.

The orders of action of ASAT enzyme orthologs, which are those related by descent from a common ancestor, are not conserved between petunia and tomato [64]. In contrast to tomato acylsugar biosynthesis which begins by acylation at position 4 by Sl-ASAT1, tetraacylsucrose assembly begins in *Petunia axillaris* with acylation of the pyranose ring of sucrose at position 2 by the petunia enzyme Pax-ASAT1 to produce a monoacylsucrose. Pax-ASAT2, which is the ortholog of Sl-ASAT1, does not use sucrose as acyl acceptor, but instead acylates the monoacylsucrose generated by Pax-ASAT1 to add a second acyl group (aiC5) to position 4 to make diacylsucrose Pax-ASAT3 substrates. Pax-ASAT3 catalyzes the transfer of acyl groups from various acyl-CoAs  (aiC5, aiC6, iC6, iC7, iC8) to the pyranose ring at position 3 of diacylsucrose acceptors, producing triacylsucroses which can be further acylated by Pax-ASAT4 which catalyzes the transfer of acyl groups from acyl-CoAs (iC4, aiC5 and iC5) to position 6 of the tri-acylated sucrose acceptor to form a tetra-acylated sucrose. Notably, Pax-ASAT1 and Pax-ASAT4 catalyze the acylation of positions 2 and 6 of the pyranose ring of sucrose respectively, and no orthologs to tomato were discovered. Acylsucroses from *P. axillaris* also showed unusual malonate esters ($C3^M$) on the furanose ring position $1'$ that had not been observed previously [50]. The identity of the ASAT responsible for adding $C3^M$ group remains unknown. These results demonstrate that while the position-selectivity of ASATs is often conserved across orthologs between tomato and petunia, there is evolutionary divergence in the carbohydrate acyl acceptors of ASATs that alters the order of acylsugar assembly in the biosynthetic pathway.

Understanding the biochemical processes used by GTs for the production of acylsugar SMs offers the opportunity exploit these genes in industry and agriculture. The full potential of trichomes as prolific chemical factories has not been fully exploited because our understanding of metabolic pathways and intermediates is still limited. Understanding the genetic and biochemical factors that lead to substrate specificity and enzymatic promiscuity is central to plant improvement efforts [63]. Of particular note is the

observation that small changes in amino acid sequences in the acyltransferases can alter acyl acceptor substrate preferences [63]. As genomic technologies continue to improve, the discovery of acyltransferase functions continues to be limited by our ability to profile and identify acylsugar SMs, particularly with regard to positions of acylations. Thus, there is a need for advanced analytical strategies that can provide more detailed information about identities of pathway intermediates and products.

## 1.5 Analytical approaches for profiling and discovery of acylsugar metabolites

### 1.5.1 Mass spectrometry

Metabolite analysis usually begins with mass spectrometry (MS), because MS allows for detection and quantification of a broad range of metabolites with speed, sensitivity and accuracy. Mass spectrometers separate and detect ionized gas-phase analytes according to their mass-to-charge ratios ($m/z$). Accurate $m/z$ measurements provide information about possible elemental formulas of analytes, which serves as one of the most useful pieces of information for metabolite annotation and discovery. Comprehensive metabolite profiling usually requires multiple MS-based techniques. For instance, gas chromatography (GC) coupled with electron ionization (EI) is typically employed for separation and detection of lower molecular weight, volatile, or non-labile molecules. In contrast, liquid chromatography (LC) techniques coupled with atmospheric pressure chemical ionization (APCI) or electrospray ionization (ESI) are better suited for measuring more polar, labile and non-volatile metabolites without need for derivatization to improve volatility.

### 1.5.1.1 Gas Chromatography/Mass Spectrometry (GC/MS)

GC/MS is the technique of choice for smaller volatile metabolites. The GC separation offers superior chromatographic resolution compared to other chromatographic techniques and was coupled with MS long before other inline separations [66]. GC/MS instruments often employ electron ionization (EI) at 70 eV electron energy to generate radical $M^+$ molecular ions. EI is regarded as a hard ionization technique because it generates fragment rich spectra that may lack unambiguous information about molecular mass. The

principal advantage of EI derives from reproducibility of mass spectra across instrument platforms which facilitates matching of spectra to spectral databases for identification of known metabolites.

Analysis of intact acylsugar metabolites by GC/MS has proven to be impractical because acylsugars are present as non-volatile metabolites that usually have multiple free hydroxyl groups that require conversion to GC amenable derivatives (such as trimethylsilyl (TMS) ethers) for GC separations [23]. In addition, derivatized disaccharides such as acylsucroses fragment extensively under EI conditions, displaying cleavage at the glycosidic linkage [54], but usually without molecular mass information in their EI mass spectra. However, GC/MS has proven to be well-suited for the analysis of total acyl composition of acylsugars by first converting acyl groups to alkyl esters such as ethyl [67] or butyl esters [68] by transesterification, followed by GC/MS analysis. This method is advantageous because most GC separations resolve their branched alkyl ester isomers (e.g. using a (5%-phenyl)-methylpolysiloxane capillary column), and authentic standards can be purchased or easily made. Comparisons of mass spectra and retention times to authentic standards allow for comprehensive acyl group identification. Several reports have employed similar GC/MS approaches to characterize and quantify the number and types of ester groups that make up acylsugars in the Solanaceae, including *Solanum*, *Petunia* and *Nicotiana* species [47, 49, 54-56, 68]. However, these approaches have a disadvantage in that they do not provide information about the number of acyl groups attached to each acylsugar or the exact positions of acylation.

## 1.5.1.2 Ultra-High-Performance Liquid Chromatography/Mass Spectrometry (UHPLC/MS)

The combination of MS with LC separations provides the foundation for chemical analysis of non-volatile metabolites because intact metabolites can usually be analyzed without derivatization. Today's UHPLC columns typically employ high purity silica-based supports with <2 µm particle sizes and operate at high pressures up to 15,000 psi, providing optimal peak capacity, bandwidth, resolution, sensitivity and sample throughput. Some modern columns also utilize superficially porous particles that operate at similar pressures to more traditional HPLC systems (up to ~5,000 psi), while achieving chromatographic resolution similar to columns with smaller fully-porous particles. Reversed-phase hydrophobic stationary phases (C18

and C8) are widely used to separate analytes based on their hydrophobicity, and are often the first approach

analysis of hydrophobic analytes. Phenyl-containing phases such as pentafluorophenyl (F5) bonded phases

are often used because they interact strongly with analytes containing aromatic/unsaturated groups ($\pi$

electron systems) or basic analytes [69, 70]. Hydrophobic interaction liquid chromatography (HILIC) uses

amino, diol, amide, and silica phases and offers greater retention of more polar analytes [71, 72].

The development of atmospheric pressure ionization techniques including atmospheric pressure chemical

ionization (APCI) and electrospray ionization (ESI) enabled improved coupling of LC to MS and was a

major breakthrough for the analysis of intact metabolites [73]. ESI is often the method of choice for

advanced SM profiling because the ESI process does not require analyte evaporation at high temperatures

that may degrade labile ionic compounds [74]. ESI generates ions by applying a high voltage to an LC flow

via a liquid capillary, creating a charged aerosol that is further de-solvated producing gas-phase ions. One

or more LC solvents often contain weak acids/bases to facilitate analyte ionization, one example is

ammonium formate ($NH_4HCOO$). Ionization by ESI is dependent on the acidity, basicity, or ion affinities

of the analyte, and can be performed in either positive- or negative-ion modes. ESI is considered a soft

ionization technique because it generates fewer fragment ions by comparison to ionization techniques such

as EI. Protonated $[M+H]^+$ or deprotonated $[M-H]^-$ molecules are observed for analytes with ionizable

functional groups. Molecules lacking acidic or basic groups may form pseudo-molecular adduct ions

including $[M+NH_4]^+$ and $[M+HCOO]^-$ when ammonium formate is added to the LC mobile phase.

Selective extraction of acylsugars from trichomes is easily achieved by dipping leaflets into an extraction

solvent such as acetonitrile: isopropanol: water (3:3:2, v/v/v; with 0.1 % formic acid to inhibit

rearrangements of acyl group positions) for just 1-2 minutes. This solvent extracts a wide range of

metabolites from GTs with minimal penetration into the surface tissue of leaflets [23].

Recent efforts to perform deep profiling of acylsugars using liquid chromatography have employed C18

superficially porous analytical columns with gradient elution using aqueous ammonium formate (10 mM

adjusted to pH 2.8 with formic acid) and acetonitrile organic component [38, 50]. These methods often

14

feature long-shallow gradients with run times in excess of 100 mins per sample and are used to probe acylsugar complexity. Results of such analyses demonstrate that most acylsugar profiles present multiple isomeric forms for each elemental composition that may not be resolved using more rapid LC gradients.

### 1.5.1.3 Collision Induced Dissociation (CID)

Annotation of acylsugars by LC/MS is facilitated by assigning molecular masses and formulas, masses of individual acyl groups, the number of acylations, and to the extent possible, the sugar core and positions of ester substituents. Because ESI generates mostly intact pseudomolecular ions from acylsugars under gentle ionization conditions, their mass spectra lack information about fragment ion masses, limiting the information available for structure annotation. To circumvent this issue, fragment ion information has been obtained using data-independent nonselective collision induced dissociation (CID) that subjects all ions to collisional activation via collision with a target gas without precursor mass selection [75]. When an ion with high translational energy (up to $\sim$ 100 eV) collides with a target gas such as argon, part of the ion's translational energy is converted to internal vibrational energy which is rapidly (ps-ns) distributed throughout the molecule. Population of excited vibrational states causes a dramatic increase of rates of unimolecular ion dissociation, forming fragment ions with masses that provide information about structure. CID may be applied in a multiplexed parallel set of collision conditions, with steps in potential from 0-200 V requiring only a few milliseconds per function. At low collision potentials, acylsugars produce abundant formate ($[M+HCOO]^-$) and ammonium ($[M+NH_4]^+$) adduct ions when ammonium formate is used in the mobile phase, as well as $[M+Cl]^-$, $[M+NO_3]^-$, $[M+Na]^+$ and $[M+K]^+$ whose masses facilitate assignments of elemental formulas. CID at elevated collision potentials generates fragment ions whose masses are used to identify masses of acyl groups, the number of acylations, the mass of the disaccharide sugar core, and, in the case of acylsucroses, assignments of acyl groups to either the fructose or pyranose rings.

### 1.5.1.4 High Resolution Mass Spectrometry

High resolution mass analyzers, including time-of-flight (ToF) and orbitrap instruments, provide high mass resolving power and accurate mass measurements (*m/z* measurements accurate to $\sim$ 5 parts-per-million or

less) that are not achieved by nominal resolution mass analyzers such as quadrupoles. As the molecular mass of a metabolite increases, the number of possible elemental compositions for a detected *m/z* value within a given tolerance window also increases [76]. A major benefit of high resolution mass spectrometry lies in the unique exact-mass of each element and isotope. The combination of exact-mass and natural isotopic abundance measurements usually reduces the number of possible elemental compositions to only a few possibilities.

The value of high resolution MS is demonstrated in acylsugar profiling. Most acylsugars observed to date in surface extracts of Solanaceae plants consist of homologous series differing by the number of acylations and $CH_2$ units in the acyl groups, observed as increases of +14 Da. However, *Ghosh et al.* demonstrated that isobaric acylsugars (those with different elemental formulas, but with the same nominal (integer) masses) such as a triacylsucrose of formula $C_{30}H_{52}O_{14}$ and tetraacylsucrose $C_{29}H_{48}O_{15}$ both have the same nominal masses of 636 Da [23]. However, their monoisotopic masses differ by 0.036 Da, with the triacylated sucrose having the slightly greater mass. From the molecular mass measurement alone, modern high resolution MS instruments can distinguish the number of acyl groups, as well as how many total carbons are in the acyl groups for acylsugars. Accurate mass measurements are also informative for assigning elemental formulas to fragment ions. For instance, positive-ion mode spectra of acylated sucroses generated at elevated collision energies shows cleavage of the glycosidic bond. From the exact masses of their pyranose and furanose ring fragment ions, the number of acylations on each ring can be determined [23], though fragment ion masses have yet to establish specific positions of acylations because much fragmentation involves neutral losses of acyl groups.

### 1.5.1.5 Tandem Mass Spectrometry (MS/MS)

Even when long and shallow LC gradients are used, LC/MS profiles of most plant extracts present several co-eluting chromatographic peaks, making nonselective CID spectra challenging to interpret because it is not clear which fragment ion is derived from each precursor. Tandem mass spectrometry (MS/MS) offers multiple steps of mass spectrometry selection with CID fragmentation in between. In the two stage MS/MS

16

product ion scan, the first mass analyzer (often a quadrupole) selects a specific precursor ion for fragmentation in the collision cell, filtering away ions of all other m/z values, while the second mass analyzer scans through a range of *m/z* values. For acylsugar analysis, product ion MS/MS spectra may be employed to distinguish coeluting acylsugars that differ in molecular mass by 2 Da or more, such as acylsugars with one or more unsaturated ester groups. Tandem MS/MS spectra may be used to distinguish differences in acyl group carbon lengths and saturations (from those with double bonds or rings), while relative ion abundance in MS/MS experiments may provide useful clues for distinguishing differences among carbohydrate groups that form the foundation of acylsugar metabolites. However, multiple isomeric forms of acylsugars are present in some plant extracts that differ by their acyl group branching or by their sugar substitution positions that are not evident from MS/MS spectra [23].

### 1.5.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

Definitive identification of a metabolite often requires that all of the connections between atoms be catalogued, including their stereochemical configurations. However, mass spectrometry often cannot distinguish structural isomers, leaving many metabolite annotations ambiguous. For example, *myo*-inositol and glucose (Figure 1.1) are isomers of hexose ($C_6H_{12}O_6$), sugars that are not readily differentiated by mass spectrometry alone. NMR spectroscopy, especially the combination of one- and two-dimensional (1D and 2D) techniques, provides more detailed structural information of metabolites, and is considered the 'gold standard' for structural elucidation [5]. NMR also has an added benefit in that it is non-destructive, allowing for further spectroscopic analysis. Unfortunately, one of the major challenges of using NMR is that it is less sensitive compared to MS techniques, often requiring laborious purifications of milligram quantities of analyte. Some NMR spectrometers are equipped with cryoprobe and capillary probe accessories that make it possible to generate spectra for lower (µg) quantities of purified material [77]. Only one of the spectrometers at Michigan State University, the Bruker Avance 900, is equipped with cryoprobe capabilities, and none has capillary probe capabilities.

**1.5.2.1 Extraction and purification of acylsugar metabolites for NMR analysis**

The amount of plant tissue needed to purify 1 mg of a specific metabolite depends on the plant genotype, stage of development, growth conditions, and concentration of the target metabolite in the plant tissue [23]. Several studies demonstrated purification of acylsugars using a myriad of organic extraction solvents and separation techniques (Table 1.1). Recent reports from our laboratories have documented rapid trichome extractions of tomato and petunia species using organic solvents, followed by purification of concentrated trichome extracts using reversed-phase C18 semi-preparative HPLC with fraction collection [38, 50]. For *Petunia* species, *Liu et al.* employed a pre-fractionation step using strong anion exchange (SAX) solid phase extraction (SPE) column for separating a group of anionic malonate ester acylsucroses from neutral acylsucroses that had similar retention behavior using reversed phase HPLC [50]. Subsequent separation using reversed phase HPLC yielded fractions of sufficient quantity and purity for structure elucidation using NMR spectra.

**1.5.2.2 Identification by 1D and 2D NMR spectroscopy**

Most plant specialized metabolites are composed only of hydrogen, carbon, nitrogen and oxygen. Of these, $^1$H and $^{13}$C isotopes are the most important isotopes for NMR structural elucidation. One-dimensional (1D) NMR provides useful chemical shifts (indicators of atomic environment), coupling constants (interactions between atoms) and relative proton quantification for establishing the connectivity of elements/functional groups and their relative stereochemistry in a compound. However, acylsugars are large molecules with many similar functional groups, and their 1D spectra exhibit signals with overlapping resonances, or significant second order couplings, making it challenging to interpret the results. In addition, the low natural abundance of $^{13}$C (1.07% of total carbon) compared to $^1$H (99.99% of total hydrogen) makes for weak $^{13}$C signals when amounts of sample are low, which may limit the amount of information available from NMR in such cases.

To circumvent these issues, structural elucidation of acylsugars is facilitated by a variety 2D NMR spectroscopic techniques [23, 38, 50]. COSY (COrrelation SpectroscopY) is a homonuclear $^1$H-$^1$H

technique for measuring signals that arise from couplings of neighboring protons, usually up to 4 bonds away. For acylsugar analyses, COSY is particularly useful for assigning ring hydrogen connectivity. The heteronuclear HSQC (Heteronuclear Single Quantum Coherence) technique measures short-range couplings arising from one bond interactions of $^1$H-$^{13}$C, revealing which protons are attached to specific carbon atoms. HSQC is also used to distinguish $CH_2$ groups from $CH/CH_3$, which show negative and positive signal respectively. HMBC (Heteronuclear Multiple-Bond Correlation) is a heteronuclear technique similar to HSQC, however, long-distance couplings of $^1$H-$^{13}$C are measured (out to 2-4 bonds). For acylsugars, HMBC is especially useful for confirming ring connectivity, and is a powerful approach for determining the positions of acyl substitutions. Additional 2D NMR techniques are also useful for determining acylsugar atomic connectivity and relative stereoisomerism. For instance, *J*-resolved NMR may be used for discerning $^1$H-$^1$H coupling constants of overlapping $^1$H resonances.  2D-TOCSY (TOtal Correlated SpectroscopY) spectra are useful for revealing the coupling network of sugar ring hydrogens, as all protons on a given ring will have coupling with all other protons, but not with those that are on other sugar rings. Coupled-HSQC spectra yield $^1$H-$^{13}$C coupling constants that can be used to discern relative stereochemistry of glycosylated positions (i.e. *α* versus *β* glycosylation).

**1.5.3 Chromatographic Retention Indexing (RI) for annotation of specialized metabolites**

Our ability to perform rapid detection of compounds in complex extracts using UHPLC/MS far exceeds our ability to generate unambiguous structure information about the overwhelming majority of them. As we discover more SMs, the risk of discovering the same compounds rises. Plants with close genetic relationships often produce some of the same or structurally similar metabolites [78]. Annotation of most SMs is initially driven by mass spectra and retention times measured using LC/MS. However, variation in GC and LC retention times across laboratories, instruments, and experimental protocols compromise retention time comparisons and determination of whether metabolites from two genotypes that exhibit similar masses and chromatographic retention times are the same or different compounds. Greater

confidence in retention time comparisons may avoid replication of metabolite purification and structure elucidation.

**1.5.3.1 GC/MS Kováts retention indexing**

In gas chromatographic separations, Kováts retention indices have long been used to convert retention time into system independent constants, whereby the retention index of a specific compound is normalized to the retention times of adjacent eluting homologous *n*-alkanes [79]. RI values are calculated by linear interpolation between retention standards according to Equation 1.2, where, $t_r$ is the retention time, and **n** and **N** are the number of carbon atoms in the shorter and longer *n*-alkyl standards bracketing the peak of interest respectively. For metabolite profiling, GC/MS indexing is usually used with electron ionization to form mass spectral libraries that contain retention index information [80]. Tables of GC retention indices in conjunction with mass spectrometry can help to identify compounds with known retention indices, often providing additional confidence to metabolite annotations, particularly when mass spectra alone are not particularly discriminating.

**Equation 1.2.** Calculation of RI values.

$$RI = 100 \times \left[ n + (N - n) \frac{t_{r(unknown)} - t_{r(n)}}{t_{r(N)} - t_{r(n)}} \right]$$

**1.5.3.2 HPLC and LC/MS retention indexing approaches**

Even though liquid chromatography is often the preceding step before MS analysis, modified Kováts liquid chromatographic indices are rarely used. In fact, a recent review of open-access spectral databases by Johnson *et al.* recognizes that none of the major open-access spectral databases (e.g. MassBank, METLIN) include retention indexing as an orthogonal measurement parameter for LC/MS/MS data [81]. In part, this is because there is no widely-accepted or unified HPLC RI system for reversed-phase (RP) chromatography [73]. A few reports of RI standards have consisted of homologous series of 2-ketoalkanes, alkyl aryl ketones, and 1-nitroalkanes, as all have common chromophores detected using ultraviolet spectroscopy [82-84]. However, most lack functional groups that facilitate their detection using mass spectrometry. A 2006

dissertation touted a series of linear aliphatic esters of 4-hydroxybenzoic acid which ionized well in negative-ion mode ESI, but a complete set of homologs was not prepared, and the range of retention was limited, as the longest alkyl ester was dodecyl [85]. Because ESI is often the first approach for non-volatile SM discovery, indexing standards that are ionizable using both positive- and negative-ion modes would be preferable. This way, the same standard series could be used in both ionization modes. Indexing standards should have a wide retention range, have molecular masses that are outside of targeted approaches, be inexpensive to produce, non-toxic and non-reactive [73]. By adding a time-indexed dimension to complex LC/ESI/MS metabolomics datasets, similarities and differences in related plant species can be determined rapidly and efficiently. This information has the potential to accelerate discoveries about how plant specialized metabolism has evolved.

**1.6 Summary of research**

In this dissertation, analytical strategies for profiling and discovery of acylsugar specialized metabolites form glandular trichomes of plants of the family Solanaceae are outlined. Untargeted reversed-phase UHPLC/ESI/Q-Tof/MS metabolite profiling in positive- and negative-ion modes of extracts from leaf surface glandular trichomes of two species of the family Solanaceae, *Salpiglossis sinuata* and *Solanum quitoense*, revealed diverse acylsugar metabolites that were different from previously discovered metabolites of this compound class. Acylsugar metabolites were purified by semi-preparative HPLC and their structures were elucidated utilizing $^1$H, $^{13}$C, COSY, *J*-resolved, TOCSY, HSQC, HMBC, NOESY and ROESY NMR spectroscopy. These efforts established structures of new acylsucrose and acylinositol metabolites extracted from *S. sinuata* and *S. quitoense* and provide new analytical approaches for defining acylsugar composition and biodiversity. Furthermore, these results have guided the discovery of previously unidentified acylsugar biosynthetic enzymes operating in acylsugar biosynthetic pathways and have extended our understanding of the evolution of specialized metabolism in the Solanaceae.

As we learn more about the Solanaceae, chances of re-discovering the same compounds in closely related plant species rises. To address the need for a HPLC RI system, a homologous set of *S*-alkyl glutathione

(GS-*n*-alkyl) RI standards featuring normal saturated chain lengths (1-24 carbons) was synthesized. These RI standards encompass a wide RP-UHPLC retention range, are easily ionized by ESI in positive- and negative-ion modes (yielding yielded [M+H]$^+$ and [M-H]$^-$ ions) and show improved capacity for archiving retention data by gradient elution. To test the performance of this retention indexing approach, GS-*n*-alkyl standards were added at 1.0 µM each to an extract of *S. sinuata* containing acylsugars. A thorough investigation of acylsucrose RI dependence by gradient elution using GS-*n*-alkyl standards was performed while altering several important chromatographic experiment variables, including columns, solvent delivery systems, aqueous mobile phase pH, column temperature, LC gradient and organic solvent component. In addition, GS-*n*-alkyl standard LC/MS analysis shows promise for evaluating RP-HPLC column performance, batch-to-batch column reproducibility and column lifetime. This research has the potential to improve interspecies SM metabolite discovery and dereplication. With this work, we hope to encourage more metabolomics researchers to adopt retention indexing systems as additional measurement parameters for open-access databases.

# REFERENCES

# REFERENCES

1.  Saito, K. and F. Matsuda, *Metabolomics for functional genomics, systems biology, and biotechnology.* Annu Rev Plant Biol, 2010. **61**(1): p. 463-89.

2.  Kumar, R., et al., *Metabolomics for Plant Improvement: Status and Prospects.* Front Plant Sci, 2017. **8**(1302): p. 1302.

3.  Verpoorte, R. and J. Memelink, *Engineering secondary metabolite production in plants.* Curr. Opin. Biotechnol., 2002. **13**(2): p. 181-187.

4.  Fiehn, O., *Metabolomics — the link between genotypes and phenotypes*, in *Functional Genomics*, C. Town, Editor. 2002, Springer Netherlands: Dordrecht. p. 155-171.

5.  Last, R.L., A.D. Jones, and Y. Shachar-Hill, *Towards the plant metabolome and beyond.* Nat Rev Mol Cell Biol, 2007. **8**(2): p. 167-74.

6.  Mithofer, A. and W. Boland, *Plant defense against herbivores: chemical aspects*, in *Annual Review of Plant Biology, Vol 63*, S.S. Merchant, Editor. 2012, Annual Reviews: Palo Alto. p. 431-450.

7.  Pichersky, E. and J. Gershenzon, *The formation and function of plant volatiles: perfumes for pollinator attraction and defense.* Curr Opin Plant Biol, 2002. **5**(3): p. 237-43.

8.  Pichersky, E. and E. Lewinsohn, *Convergent evolution in plant specialized metabolism*, in *Annual Review of Plant Biology, Vol 62*, S.S. Merchant, W.R. Briggs, and D. Ort, Editors. 2011, Annual Reviews: Palo Alto. p. 549-566.

9.  Dixon, R.A. and D. Strack, *Phytochemistry meets genome analysis, and beyond.* Phytochemistry, 2003. **62**(6): p. 815-6.

10. Schwab, W., *Metabolome diversity: too few genes, too many metabolites?* Phytochemistry, 2003. **62**(6): p. 837-49.

11. Moghe, G.D. and R.L. Last, *Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism.* Plant Physiol, 2015. **169**(3): p. 1512-23.

12. Van de Peer, Y., S. Maere, and A. Meyer, *The evolutionary significance of ancient genome duplications.* Nat Rev Genet, 2009. **10**(10): p. 725-32.

13. Comai, L., *The advantages and disadvantages of being polyploid.* Nat Rev Genet, 2005. **6**(11): p. 836-46.

14. Higashi, Y. and K. Saito, *Network analysis for gene discovery in plant-specialized metabolism.* Plant Cell Environ, 2013. **36**(9): p. 1597-606.

15. Schilmiller, A.L., R.L. Last, and E. Pichersky, *Harnessing plant trichome biochemistry for the production of useful compounds.* Plant J, 2008. **54**(4): p. 702-11.

16. Levin, D.A., *The Role of Trichomes in Plant Defense.* The Quarterly Review of Biology, 1973. **48**(1, Part 1): p. 3-15.

17.     Wagner, G.J., E. Wang, and R.W. Shepherd, *New approaches for studying and exploiting an old protuberance, the plant trichome.* Ann Bot, 2004. **93**(1): p. 3-11.

18.     Landry, L.G., C.C. Chapple, and R.L. Last, *Arabidopsis mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage.* Plant Physiol, 1995. **109**(4): p. 1159-66.

19.     Ehleringer, J., *Ecology and ecophysiology of leaf pubescence in North American desert plants.* Biology and chemistry of plant trichomes, 1984.

20.     Fahn, A., *Secretory tissues in vascular plants.* New Phytologist, 1988. **108**(3): p. 229-257.

21.     Glas, J.J., et al., *Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores.* Int J Mol Sci, 2012. **13**(12): p. 17077-103.

22.     Li, C., Z. Wang, and A.D. Jones, *Chemical imaging of trichome specialized metabolites using contact printing and laser desorption/ionization mass spectrometry.* Anal Bioanal Chem, 2014. **406**(1): p. 171-82.

23.     Ghosh, B. and A.D. Jones, *Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar esters).* Analytical Methods, 2017. **9**(6): p. 892-905.

24.     Schilmiller, A., et al., *Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines.* Plant J, 2010. **62**(3): p. 391-403.

25.     Gang, D.R., et al., *An investigation of the storage and biosynthesis of phenylpropenes in sweet basil.* Plant Physiol, 2001. **125**(2): p. 539-55.

26.     Turner, C.E., M.A. Elsohly, and E.G. Boeren, *Constituents of Cannabis sativa L. XVII. A review of the natural constituents.* J Nat Prod, 1980. **43**(2): p. 169-234.

27.     Stevens, J.F. and J.E. Page, *Xanthohumol and related prenylflavonoids from hops and beer: to your good health!* Phytochemistry, 2004. **65**(10): p. 1317-30.

28.     *2015 Nobel Prize in Physiology or Medicine*. 2015, The Nobel Assembly at Karolinska Institutet.

29.     Luckwill, L.C., *The genus Lycopersicon. An historical, biological and taxonomic survey of the wild and cultivated tomatoes.* PhD thesis, Department of Botany, University of Aberdeen, UK, 1943.

30.     van den Oever-van den Elsen, F., et al., *Quantitative resistance against Bemisia tabaci in Solanum pennellii: Genetics and metabolomics.* J Integr Plant Biol, 2016. **58**(4): p. 397-412.

31.     Baier, J.E., et al., *Indirect selection of industrial tomato genotypes that are resistant to spider mites (Tetranychus urticae).* Genetics and Molecular Research, 2015. **14**(1): p. 244-252.

32.     Goffreda, J.C., et al., *Aphid deterrence by glucose esters in glandular trichome exudate of the wild tomato, Lycopersicon pennellii.* J Chem Ecol, 1989. **15**(7): p. 2135-47.

33.     Leckie, B.M., et al., *Differential and synergistic functionality of acylsugars in suppressing oviposition by insect herbivores.* PLoS One, 2016. **11**(4): p. e0153345.

34.     Luu, V.T., et al., *O-acyl sugars protect a wild tobacco from both native fungal pathogens and a specialist herbivore.* Plant Physiol, 2017. **174**(1): p. 370-386.

35.     Weinhold, A. and I.T. Baldwin, *Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation.* Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7855-9.

36.     Moghe, G.D., et al., *Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway.* Elife, 2017. **6**: p. e28486.

37.     Connolly, J.D. and R.A. Hill, *Dictionary of terpenoids*. Flavour and Fragrance Journal. 1991, London: Chapman & Hall.

38.     Ghosh, B., T.C. Westbrook, and A.D. Jones, *Comparative structural profiling of trichome specialized metabolites in tomato (Solanum lycopersicum) and S. habrochaites: acylsugar profiles revealed by UHPLC/MS and NMR.* Metabolomics, 2014. **10**(3): p. 496-507.

39.     Schilmiller, A.L., et al., *Functionally divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in Solanum trichomes.* Plant Cell, 2015. **27**(4): p. 1002-17.

40.     Schilmiller, A.L., et al., *Acylsugar Acylhydrolases: Carboxylesterase-Catalyzed Hydrolysis of Acylsugars in Tomato Trichomes.* Plant Physiol, 2016. **170**(3): p. 1331-44.

41.     King, R.R., et al., *Sucrose esters associated with glandular trichomes of wild Lycopersicon species.* Phytochemistry, 1990. **29**(7): p. 2115-2118.

42.     King, R.R., et al., *Characterization of 2,3,4,3'-tetra-O-acylated sucrose esters associated with the glandular trichomes of Lycopersicon-typicum.* Journal of Agricultural and Food Chemistry, 1993. **41**(3): p. 469-473.

43.     King, R.R., et al., *3,4-Di-O-Isobutyryl-6-O-Caprylsucrose - the Major Component of a Novel Sucrose Ester Complex from the Type-B Glandular Trichomes of Solanum-Berthaultii Hawkes (Pi 473340).* Journal of the Chemical Society-Chemical Communications, 1986(14): p. 1078-1079.

44.     King, R.R., R.P. Singh, and L.A. Calhoun, *Isolation and characterization of 3,3',4,6-tetra-O-acylated sucrose esters from the type-B glandular trichomes of Solanum-berthaultii Hawkes (Pi-265857).* Carbohydrate Research, 1987. **166**(1): p. 113-121.

45.     King, R.R. and A.K. Singh, *Elucidation of structures for a unique class of 2,3,4,3'-tetra-O-acylated sucrose esters from the type B glandular trichomes of Solunum neocardenasii Hawkes & HJerting (PI 498129).* Carbohydrate Research, 1988. **173**: p. 235-241.

46.     King, R.R., L.A. Calhoun, and R.P. Singh, *3,4-Di-O- and 2,3,4-tri-O-acylated glucose esters from the glandular trichomes of nontuberous Solanum species.* Phytochemistry, 1988. **27**(12): p. 3765-3768.

47.     Burke, B.A., G. Goldsby, and J.B. Mudd, *Polar epicuticular lipids of Lycopersicon pennellii.* Phytochemistry, 1987. **26**(9): p. 2567-2571.

48.     Li, A.X. and J.C. Steffens, *An acyltransferase catalyzing the formation of diacylglucose is a serine carboxypeptidase-like protein.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6902-6907.

49.     Herrera-Salgado, Y., et al., *Myo-inositol-derived glycolipids with anti-inflammatory activity from Solanum lanceolatum.* J Nat Prod, 2005. **68**(7): p. 1031-6.

50.     Liu, X.X., et al., *Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of Petunia species.* Metabolomics, 2017. **13**(7): p. 85.

51.     Singh, A.P., et al., *Two acyl sucroses from Petunia nyctaginiflora.* Phytochemistry, 2003. **63**(4): p. 485-9.

52.     Begum, A.S., et al., *New pairs of acyl sucroses from Petunia nyctaginiflora Juss.* Journal of the Indian Chemical Society, 2004. **81**(6): p. 495-500.

53.     Begum, A.S., et al., *Two novel acyl sucroses from Petunia nyctaginiflora.* Indian Journal of Chemistry Section B-Organic Chemistry Including Medicinal Chemistry, 2005. **44**(3): p. 648-650.

54.     Severson, R.F., et al., *Isolation and characterization of the sucrose esters of the cuticular waxes of green tobacco leaf.* Journal of Agricultural and Food Chemistry, 1985. **33**(5): p. 870-875.

55.     Arrendale, R.F., et al., *Characterization of the sucrose ester fraction from Nicotiana-glutinosa.* Journal of Agricultural and Food Chemistry, 1990. **38**(1): p. 75-85.

56.     Matsuzaki, T., et al., *Leaf surface glycolipids from Nicotiana-acuminata and Nicotiana-pauciflora.* Agricultural and Biological Chemistry, 1991. **55**(5): p. 1417-1419.

57.     Zhang, C.R., et al., *New antiinflammatory sucrose esters in the natural sticky coating of tomatillo (Physalis philadelphica), an important culinary fruit.* Food Chem, 2016. **196**: p. 726-32.

58.     Ovenden, S.P., et al., *Physaloside A, an acylated sucrose ester from Physalis viscosa.* J Nat Prod, 2005. **68**(2): p. 282-4.

59.     Castillo, M., et al., *Partially acylated glucose and sucrose derivatives from Salpiglossis-sinuata (Solanaceae).* Journal of Chemical Research-S, 1989(12): p. 398-399.

60.     Eckart, E., *Secondary Metabolites Derived from Fatty Acids and Carbohydrates*, in *Solanaceae and Convolvulaceae: Secondary Metabolites: Biosynthesis, Chemotaxonomy, Biological and Economic Significance (A Handbook)*, E. Eich, Editor. 2008, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 525-582.

61.     D'Auria, J.C., *Acyltransferases in plants: a good time to be BAHD.* Curr Opin Plant Biol, 2006. **9**(3): p. 331-40.

62.     Schilmiller, A.L., A.L. Charbonneau, and R.L. Last, *Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16377-82.

63.     Fan, P., et al., *In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network.* Proc Natl Acad Sci U S A, 2016. **113**(2): p. E239-48.

64.     Nadakuduti, S.S., et al., *Characterization of trichome-expressed BAHD acyltransferases in petunia axillaris reveals distinct acylsugar assembly mechanisms within the Solanaceae.* Plant Physiol, 2017. **175**(1): p. 36-50.

65.     St-Pierre, B. and V. De Luca, *Chapter Nine Evolution of acyltransferase genes: Origin and diversification of the BAHD superfamily of acyltransferases involved in secondary metabolism.* Vol. 34. 2000. 285-315.

66.     Gohlke, R.S., *Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography.* Analytical Chemistry, 1959. **31**(4): p. 535-541.

67.     Walters, D.S. and J.C. Steffens, *Branched chain amino acid metabolism in the biosynthesis of Lycopersicon pennellii glucose esters.* Plant Physiol, 1990. **93**(4): p. 1544-51.

68.     Kroumova, A.B., D. Zaitlin, and G.J. Wagner, *Natural variability in acyl moieties of sugar esters produced by certain tobacco and other Solanaceae species.* Phytochemistry, 2016. **130**: p. 218-27.

69.     Bell, D.S., H.M. Cramer, and A.D. Jones, *Rational method development strategies on a fluorinated liquid chromatography stationary phase: mobile phase ion concentration and temperature effects on the separation of ephedrine alkaloids.* J Chromatogr A, 2005. **1095**(1-2): p. 113-8.

70.     Bell, D.S. and A.D. Jones, *Solute attributes and molecular interactions contributing to "U-shape" retention on a fluorinated high-performance liquid chromatography stationary phase.* J Chromatogr A, 2005. **1073**(1-2): p. 99-109.

71.     LCGC-Editors, *HPLC Column Selection*, in *LCGC Europe*. 2013. p. 298.

72.     Jandera, P., *Stationary and mobile phases in hydrophilic interaction chromatography: a review.* Anal Chim Acta, 2011. **692**(1-2): p. 1-25.

73.     Kind, T. and O. Fiehn, *Advances in structure elucidation of small molecules using mass spectrometry.* Bioanal Rev, 2010. **2**(1-4): p. 23-60.

74.     Bruins, A.P., *Mechanistic aspects of electrospray ionization.* Journal of Chromatography A, 1998. **794**(1-2): p. 345-357.

75.     Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry.* J Mass Spectrom, 2004. **39**(10): p. 1091-112.

76.     Xian, F., C.L. Hendrickson, and A.G. Marshall, *High resolution mass spectrometry.* Anal Chem, 2012. **84**(2): p. 708-19.

77.     Molinski, T.F., *Microscale methodology for structure elucidation of natural products.* Curr Opin Biotechnol, 2010. **21**(6): p. 819-26.

78.     Zhao, N., et al., *Studying plant secondary metabolism in the age of genomics.* Critical Reviews in Plant Sciences, 2013. **32**(6): p. 369-382.

79.     Kováts, E., *Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone.* Helvetica Chimica Acta, 1958. **41**(7): p. 1915-1932.

80.     Vinaixa, M., et al., *Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects.* Trac-Trends in Analytical Chemistry, 2016. **78**: p. 23-35.

81.     Johnson, S.R. and B.M. Lange, *Open-access metabolomics databases for natural product research: present capabilities and future potential.* Front Bioeng Biotechnol, 2015. **3**: p. 22.

82.     Baker, J.K. and C.Y. Ma, *Retention index scale for liquid-liquid chromatography.* Journal of Chromatography, 1979. **169**(Feb): p. 107-115.

83. Smith, R.M., *Alkylarylketones as a retention index scale in liquid-chromatography.* Journal of Chromatography, 1982. **236**(2): p. 313-320.

84. Bogusz, M. and R. Aderjan, *Improved standardization in seversed-phase high-performance liquid-chromatography using 1-nitroalkanes as a retention index scale.* Journal of Chromatography, 1988. **435**(1): p. 43-53.

85. Hanley Jr, J.C., *Chromatography/mass spectrometry methodologies for metabolomics, doctoral dissertation.* 2006, The Pennsylvania State University. ProQuest UMI#3229404.

**Chapter 2: LC/MS profiling and NMR structural elucidation of specialized metabolites from _Salpiglossis sinuata_ reveals extensive acylsucrose diversity including unsaturated and aromatic esters**

**2.1 Introduction**

Metabolic processes that increase fitness in plants evolved through mutations and led to structurally diverse specialized metabolites (SMs; also referred to as natural products and secondary metabolites) that are not directly involved in central metabolism. Conservative estimates have suggested that the plant kingdom produces more than 200,000 SMs [1, 2], with many of these being taxonomically restricted to a limited range of species or genotypes.

Accumulation of SMs in specialized cell types or tissues facilitates discovery of candidate biosynthetic genes which are often differentially expressed in these tissues. Glandular trichomes (GTs) serve as convenient model tissues in that they are hair-like epidermal outgrowths (Appendix Figure 2.5), are accessible for sampling and accumulate large quantities of SMs. GTs are found on the surfaces of ~30% of vascular plant species [3] and are prolific chemical factories that synthesize, store and secrete large quantities of diverse SMs. GT exudates are easily sampled, facilitating detailed characterization of their metabolites as well as transcripts that encode for biosynthetic enzymes [4-6]. In addition, GTs exhibit novel biochemical pathways for specialized metabolism and have been suggested as targets for genetic engineering for production of useful natural products and engineering plant resistance against insects and pathogens [6-9].

Members of the family Solanaceae, which includes a large number of agronomically important plants including tomato, potato, pepper, and tobacco, synthesize and/or store diverse arrays of SMs in GTs, including terpenes, flavonoids, alkaloids and fatty acid derivatives [6, 7, 10]. Most investigated Solanaceae Type I/IV GTs accumulate structurally diverse sugar esters, known as acylsugars. Common examples include glucose and sucrose with three or four acylations of C2-C12 branched or straight chain aliphatic esters. These SMs often provide the first line of physical and/or chemical defenses against herbivores [3,

11]. GTs of tobacco (*Nicotiana tabacum*), petunia (*Petunia spp.*), wild potato (*Solanum berthaultii*), cultivated tomato (*Solanum lycopersicum*) and its wild relatives (*S. habrochaites and S. pennellii*) are widely-studied prolific accumulators of acylsugars [12-14]. NMR and GC/MS structural characterization have identified >20 known acylsugar acyl groups in the Solanaceae [15]. From known acyl group building blocks alone, the number of possible acylsucrose permutations that span mono- to complete-acylation of eight sucrose hydroxyl groups extends into the billions. We estimate that thousands (or more) structurally diverse acylsugars exist across the Solanaceae, with only ~100 acylsugar metabolites fully characterized at present.

Acylation serves as a widespread mechanism for assembly of plant SMs [16]. Until recent years, functions of individual enzymes that catalyze acylations in SM biosynthesis remained undefined. Our laboratories' recent combinations of transcript analysis, LC/MS metabolite profiling of *S. lycopersicum M82 x S. pennellii LA0716* ('wild tomato') introgression lines, and *in vitro* biochemistry using recombinant acyltransferases guided the discovery of the entire set of four BAHD acyltransferase (ASAT) enzymes required for assembly of tetraacylsucroses in tomato [17-19]. Members of the broader Solanaceae employ multiple ASATs that convert carbohydrates and acyl-CoAs to complex arrays of acylsucroses, but substrate selectivities of ASATs may not be apparent from DNA sequence similarity alone [17, 19-21]. Functional assays for ASATs from the genera *Solanum*, *Petunia*, and more recently, *Salpiglossis* have shown that small numbers of amino acid substitutions can alter substrate selectivity of ASATs, position-selectivity of acylation of orthologues, and the order in which individual ASATs are used in acylsugar biosynthetic pathways [20-22].

Several reports have characterized the number, topologies, and lengths of ester groups that make up acylsugars in numerous accessions from the Solanaceae, predominantly from the genus Nicotiana [15, 23, 24]. However, outside of the genera *Solanum* and *Petunia* [12-14], comprehensive descriptions of acyl substitution positions and acyl chain selectivity at each position of sugar cores are limited. Given our recent evidence that the main differences in individual ASAT functions derive from varied selectivities of both

31

acyl acceptors and acyl donors [17, 19, 21], annotation of ASAT functions requires definition of which carbohydrate positions are acylated as well as the diversity of acyl groups at each position.

The focus of the current study is to define acylsugar diversity in *Salpiglossis sinuata*, a member of an early emerging lineage in the Solanaceae family, estimated to have diverged ~25 mya relative to the carefully studied tomato clade [25]. Trichome extracts were subjected to UHPLC/MS profiling, purification and structural elucidation using NMR spectroscopy for the purpose of uncovering differences in acylsugar composition and to facilitate discovery of genetic factors responsible for their biosynthesis. In doing so, we have discovered new acylsugars, including those containing unsaturated acyl and aromatic esters; suggesting novel gene functions are at work in acylsugar biosynthesis in *S. sinuata*. This untargeted approach aims to establish the diversity of acyl groups and their position-selective attachment and guided the discovery of biosynthetic enzymes functioning in *S. sinuata* [21].

## 2.2 Materials and methods

### 2.2.1 Plant cultivation and metabolite extraction

*S. sinuata* seeds were obtained from The New York Botanical Gardens (NY, USA), and were grown using Jiffy peat pellets in growth chambers at MSU. For metabolite profiling and purification, the complete aerial tissues of 28 ten-week old *S. sinuata* plants (~0.5 m height; plants were cut at the stems in 10-15 cm segments) were dipped into 1.9 L of acetonitrile: isopropanol (AcN:IPA, v/v, 1:1) for 10 mins. A 10-mL aliquot of this bulk extract solution, with 5.0 µM telmisartan added as internal standard, was used for all metabolite profiling and purification method development. The remaining extract was concentrated for metabolite purification as described below. Additional plant growth and extraction metadata are provided in Appendix Table 2.2.

### 2.2.2 Profiling of acylsugar metabolites using UHPLC/MS and MS/MS

Acylsugar metabolite profiling was performed using a Waters Xevo G2-XS quadrupole time-of-flight mass spectrometer (QTof/MS) equipped with an Ascentis® Express C18 Analytical HPLC column (10 cm × 2.1

mm, 2.7 µm particle size) and operated using CID in positive- and negative-ion modes. Untargeted CID spectra were acquired over $m/z$ 50-1500 in centroid format using four quasi-simultaneous collision potential functions (0, 10, 25, 50 V, each with 0.1 s acquisition times) to generate fragment ions. MS/MS spectra were generated from purified acylsucrose metabolites using a CID ramp from 5-60 V collision potential. Additional metadata regarding the UHPLC/MS and MS/MS methods are provided in Appendix Tables 2.3 and 2.5. MS/MS product ion spectra are provided in Appendix Figures 2.8-2.23.

### 2.2.2.1 Deep profiling of acylsucrose specialized metabolites by LC/ESI+/MS

Deep profiling of acylsucrose metabolites was carried out by first calculating exact molecular masses for permutations of acylsucrose pseudomolecular $m/z$ values: $[M+NH_4]^+$, $[M+Na]^+$, and $[M+K]^+$. NMR results, as well as LC/MS and GC/MS profiling revealed evidence for eight ester groups: C2, C3, iC4, iC5, aiC5, $C5^T$, aiC6, and $C8^P$. Since acylsucroses have up to eight hydroxyl positions available for acylation, all degrees of acylation were considered (mono- to octaacylsucroses) with saturated and unsaturated esters, including acylsucroses with unsaturated ester combinations of up to two $C5^T$ and/or $C8^P$ groups (i.e. acylsucroses with 1, 2, 4, 5 and 8 double bond equivalencies relative to sucrose saturated ester analogues). For each acylsucrose permutation, $[M+NH_4]^+$, $[M+Na]^+$, and $[M+K]^+$ extracted ion chromatograms were generated (in a stacked format within MassLynx software) using a 40 ppm mass window. Acylsucrose retention times and the presence of all three pseudomolecular masses were used as diagnostic criteria. For each chromatographic peak, its corresponding mass spectra were then analyzed manually. Since many acylsucrose metabolites are at low levels and co-elute with more abundant forms, mass spectra were carefully analyzed for mass accuracy, isotope abundancies and centroid satellite peaks to avoid false positives. For instance, an acylsucrose with one unsaturation (containing $C5^T$) has M+2 isotope signal that is consistent with a saturated acylsucrose analogue. Here, examination of only extracted ion chromatograms may lead to an incorrect assignment of two acylsucrose signals instead of one. For example, acylsucroses S5:22:1(2,5,5,5,$5^T$) with $[M+NH_4+2]^+$ has $m/z$ 738.382 versus S5:22:0(2,2,6,6,6) with $[M+NH_4]^+$ has $m/z$

738.391. Also of note, there is a shift towards reduced retention time for acylsucroses containing unsaturations (in the previous example, 52.55 mins versus 56.08 mins respectively).

### 2.2.3 Assessment of acyl group diversity by transesterification and GC/MS

To assess the ester composition of *S. sinuata* acylsugars, a transesterification reaction was performed on a 10 mL aliquot of the bulk extract to convert acyl groups to ethyl esters amenable to GC/MS analysis. Transesterification and GC/MS procedures are described in detail in Appendix Table 2.6. Mass spectra are shown in Appendix Figures 2.25-2.30.

### 2.2.4 Purification of acylsugar metabolites by semi-preparative HPLC

For metabolite purification, approximately 1 L of the *S. sinuata* bulk extract (described in Section 2.1) was concentrated to dryness under vacuum, dissolved in 5 mL of AcN:IPA, and fractionated by repeated injection of 50 µL samples onto a Thermo Scientific Acclaim™ 120 C18 semi-preparative HPLC column (4.6 x 150 mm, 5 µm particle size) with automated fraction collection (Appendix Table 2.8). Acylsugars purified for NMR analysis were isolated with the method described above except the $C8^P$ containing acylsugar, which required secondary purification using a Supelco Ascentis® Express F5 semi-preparative HPLC column (15 cm × 4.6 mm, 2.7 µm particle size). A shortened, but comparable semi-preparative gradient was used for the secondary purification of this acylsugar (Appendix Table 2.8).

### 2.2.5 Analysis of acylsugars by NMR spectroscopy

HPLC fractions of sufficient purity for a single metabolite, as assessed by LC/MS, were combined and concentrated to dryness under $N_2$ gas. Samples were dissolved in 250 or 300 µL of deuterated NMR solvents $CDCl_3$ (99.8 atom % D) and acetonitrile-$d_3$ (99.96 atom % D) and transferred to solvent-matched Shigemi tubes for analysis. ^{1}H, ^{13}C, gCOSY, gHSQC, gHMBC, *J*-resolved ^{1}H, and ROESY NMR experiments were performed using a Bruker Avance 900 spectrometer equipped with a TCI triple resonance probe. All spectra were referenced to non-deuterated solvent signals: $CDCl_3$ ($\delta_H = 7.26$ and $\delta_C = 77.20$ ppm) and acetonitrile-

$d_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm). More experimental conditions and spectra are located in Appendix Tables 2.9-2.27, as well as Appendix Figures 2.31-2.141.

**2.3 Results and Discussion**

**2.3.1 UHPLC/ESI/CID/QTof/MS profiling establishes diversity of *S. sinuata* acylsucroses**

Profiling of metabolites in complex mixtures requires sensitive detection, capacity to resolve similar compounds including isomers, and accurate measurements of molecular and fragment masses that serve as the basis for metabolite annotation. UHPLC coupled with QTof/MS detection using ESI in positive- and negative-ion modes aided in resolving acylsugars in complex extracts. Base peak intensity chromatograms of *S. sinuata* trichome extracts showed hundreds of distinct peaks differing in retention time. A group of compounds eluting between 25-75 mins using a 110-min gradient exhibited retention times, ion masses, and mass defects similar to acylsugars from tomato and its wild relatives [26, 27]. High resolution mass spectra of these peaks were consistent with a complex array of substituted acylsugars. At least 20 of the most abundant acylsugars were observed as peaks in the base peak intensity (BPI) chromatogram with retention times in the range of 45-65 mins (Figure 2.1A). Further evidence supporting their designation as acylsugars was obtained from examining mass spectra at elevated collision energies. Proposed structures of purified metabolites and InChI keys are presented in Appendix Tables 2.12-2.27. As these metabolites were identified without available authentic standards or synthetic confirmation, and their identities are based on NMR and mass spectrometric data alone, their structures should be considered as putative, meeting level 2 criteria of the Metabolomics Standards Initiative guidelines [28].
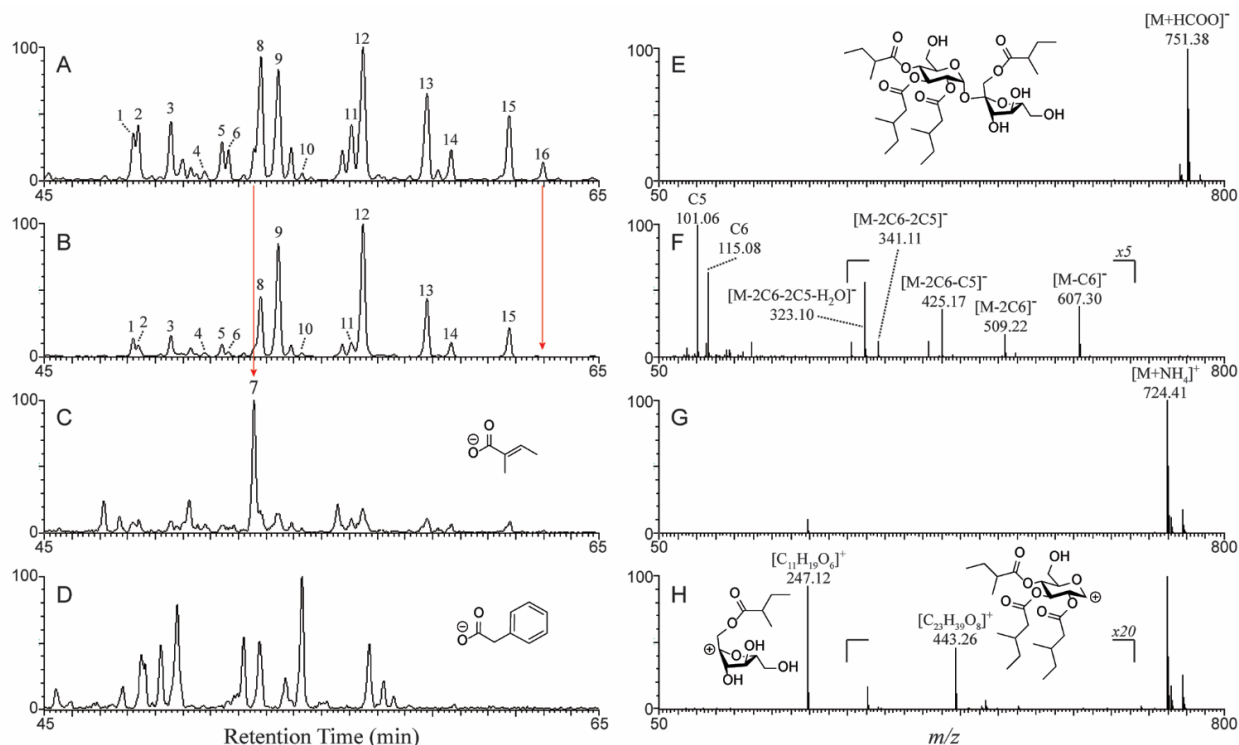
**Figure 2.1.** UHPLC/ESI/CID/QTof/MS metabolite profiles (left) and mass spectra generated from acylsugar #12 (right): **(A)** ESI(-) BPI chromatogram displaying acylsugars (formate adducts) at CID Function 1 = 0 V, purified acylsugars characterized by NMR analysis are annotated by a number indicative of their order of chromatographic elution; **(B)** ESI(-) XIC for C5 and C6 carboxylate fragment anions, $m/z$ 101.06 and 115.08 combined (Function 4 = 50 V), hexaacylsucrose #16 shows significantly weaker signal for C5 and C6 carboxylate anions compared to tetra- and pentaacylsucrose analogues; **(C)** ESI(-) XIC for $C5^T$ carboxylate fragment anion, $m/z$ 99.05 (Function 4 = 50 V), co-elution of acylsucrose #7 with #8 is evident; **(D)** ESI(-) XIC for $C8^P$ carboxylate fragment anion, $m/z$ 135.05 (Function 3 = 25 V), **(E)** ESI(-) low energy (Function 1) mass spectrum of acylsucrose #12, S4:22:0(5,5,6,6), displaying [M+HCOO]⁻ adduct, **(F)** ESI(-) at elevated energy (Function 4), displaying neutral losses consistent with two C6 and two C5 acyl groups and prominent C5 and C6 carboxylate fragments at $m/z$ 101.06 and 115.08; **(G)** ESI(+) low energy (Function 1 = 0 V) mass spectrum of peak #12 displaying [M+NH₄]⁺ adduct, **(H)** ESI(+) at elevated energy (Function 2 = 10 V), displaying cleavage of the glycosidic linkage to reveal a C5-acylated furanose ring fragment $m/z$ 247.12 and a pyranose ring fragment $m/z$ 443.26 acylated with two C6 and one C5 esters.

Annotation of acylsugars by LC/MS was facilitated by assigning molecular masses and formulas, the masses of the acyl groups, the number of acylations, and to the extent possible, the sugar core and positions of substituents. This information was obtained using multiplexed CID at various collision potentials. Low collision potentials produced abundant formate and ammonium adduct ions, [M+HCOO]⁻ and [M+NH₄]⁺, as well as [M+Cl]⁻, [M+NO₃]⁻, [M+Na]⁺ and [M+K]⁺ whose masses facilitate assignments of molecular formulas (MFs). CID at elevated collision potentials generated fragment ions used to identify important differences among acyl groups, the number of acylations, the mass of the disaccharide sugar core, and assignments of acyl groups to either the fructose or pyranose rings of sucrose.

36

To illustrate acylsugar annotation with an example, a major peak in the BPI chromatogram (Figure 2.1A) at $t_r$ = 56.49 min (acylsucrose #12) yielded mass spectra showing [M+HCOO]$^-$ at $m/z$ 751.38 in negative-ion mode (Figure 2.1E) and [M+NH$_4$]$^+$ at $m/z$ 724.41 in positive-ion mode (Figure 2.1G). These ion masses are consistent with an acylsucrose of MF $C_{34}H_{58}O_{15}$ (structure in Figure 2.1E). Because sucrose has eleven oxygen atoms, the four additional oxygens suggest four acylations, as each acylation adds one oxygen atom. Spectra generated at elevated collision energies (Figure 2.1F) show abundant fragment ions at $m/z$ 115.08 and 101.06 indicative of carboxylate (acyl anions) with MFs $C_6H_9O_2^-$ and $C_5H_7O_2^-$. Further confirmation of C5 and C6 esters comes from observation of two neutral mass losses each from [M-H]$^-$ of 84.06 and 98.07 Da ($C_5H_8O$ and $C_6H_{10}O$). Losses of two C6 and two C5 groups leaves $m/z$ 341.11, corresponding to the deprotonated sucrose core. In addition, neutral losses of 18.01 Da consistent with $H_2O$ were regularly observed during fragmentation (one example is $m/z$ 323.10). Positive ion spectra obtained at elevated collision energy (Figure 2.1H) had an abundant fragment ion $m/z$ 247.12, consistent with one C5 acyl chain on the furanose ring. Cleavage producing a less abundant pyranose fragment ion $m/z$ 443.26 suggests substitution of two C6 and one C5 on this ring. This acylsucrose was annotated as S4:22:0(5,5,6,6), using a modified nomenclature where "S" indicates the sugar core is sucrose, the number "4" indicates four ester groups, the number in parentheses describe the number of carbon atoms in each acyl group, the number 22 reflects the total number of carbon atoms across all acyl groups, and the numeral "0" indicates zero rings or double bonds in the acyl groups.

Most acylsugars observed to date in surface extracts of Solanaceae plants consist of homologous series differing by the number of acylations and $CH_2$ units in the acyl groups, observed as increases of +14 Da [5]. In addition, reverse phase HPLC retention times of acylsugars increase with acyl group chain length. However, the predictability of acylsugar masses and retention times obscures deeper exploration of low abundance acylsugars. Acyl groups differing by only one degree of unsaturation (or more) frequently co-elute with saturated analogues and may be missed by manual inspection.

Metabolite profiles of *S. sinuata* exhibited extensive complexity, with frequent chromatographic overlap of acylsugar metabolites obscuring less abundant compounds. To improve chromatographic resolution of lower abundance acylsugar metabolites from more abundant forms, a long and shallow gradient of 110-min duration was employed. The resulting chromatogram resolved peaks with molecular masses that deviate by -2 Da and -8 Da from common saturated acyl analogues. This is illustrated by a compound with $t_r = 52.55$ min (#7, Figure 2.1A and 2.1C), exhibiting [M+HCOO]$^-$ at *m/z* 763.34 and [M+NH$_4$]$^+$ at *m/z* 736.38, consistent with neutral formula $C_{34}H_{54}O_{16}$. This compound exhibits one more degree of unsaturation than a penta-substituted acylsucrose with saturated esters. Negative mode MS/MS product ion spectra of *m/z* 763.34 (Figure 2.2A) show abundant product ions at *m/z* 101.06 and 99.04 indicative of C5 saturated and monounsaturated acyl anions, the latter consistent with MF $C_5H_5O_2^-$. Given the occurrence of tiglyl esters of tropane alkaloids in the Solanaceae [29], we initially annotated this metabolite as a tiglyl (C5$^T$) acylsugar (further confirmed by NMR results in Section 2.3.3). Further support of this conclusion came from discerning combinations of neutral mass losses from [M-H]$^-$ of a C5$^T$ group (82.04 Da, $C_6H_8O$), and additional product ions provided evidence of three saturated C5 acyl groups and one C2 group (neutral loss 42.01 Da, $C_2H_2O$). Positive mode MS/MS product ion spectra of *m/z* 736.38 (Figure 2.2B) showed cleavage of the glycosidic bond, producing abundant fragment ion *m/z* 289.13, consistent with one saturated C5 and one C2 acyl chain on the furanose ring. The pyranose fragment ion at *m/z* 413.22 suggests substitutions of two C5 and one C5$^T$ esters on this ring, and supports annotation as S5:22:1(2,5,5,5,5$^T$).
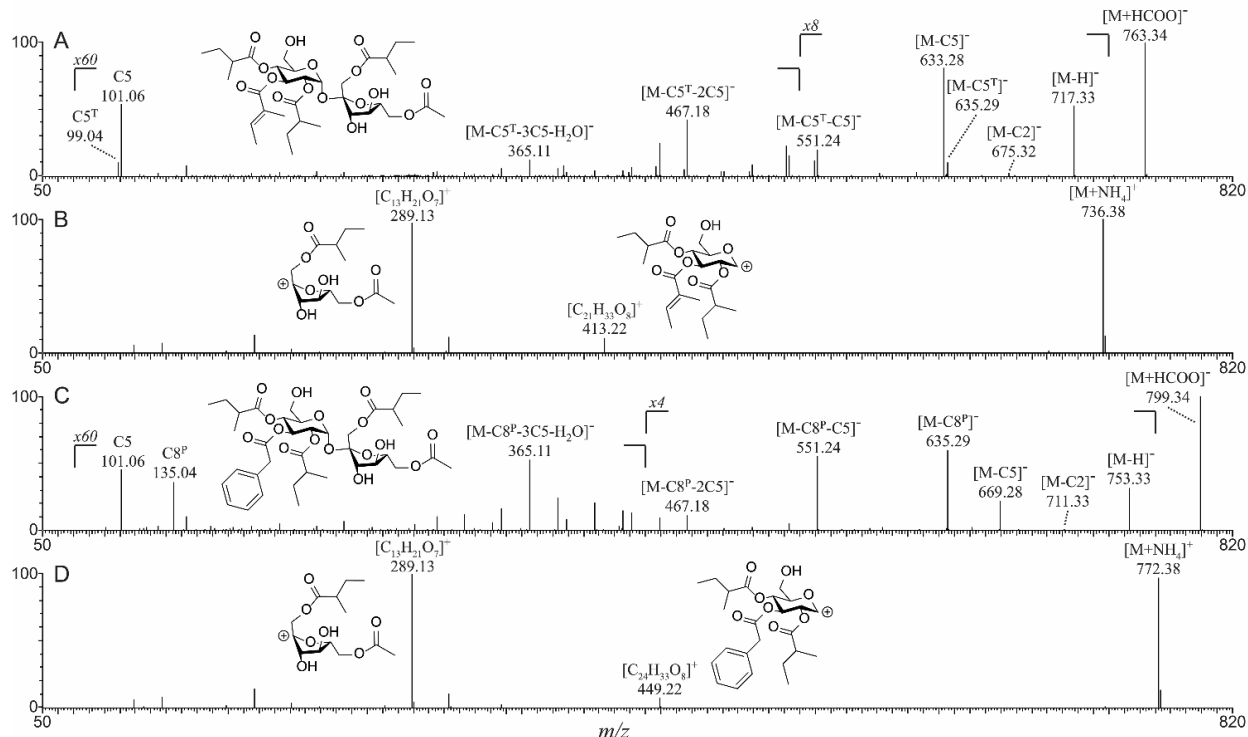
**Figure 2.2.** MS/MS product ion spectra of [M+HCOO]⁻ and [M+NH$_4$]⁺ of acylsugars generated using a linear 5-60 V collision energy ramp with 0.5 s scan time; **(A)** ESI(-) MS/MS spectrum of product ions of $m/z$ 763 ([M+HCOO]⁻) for acylsugar #7 showing evidence for neutral losses of C2, C5$^T$, C5 and H$_2$O, as well as corresponding carboxylate anions for ester groups C5$^T$ and C5 ($m/z$ 99 and 101); **(B)** ESI(+) MS/MS spectrum of product ions of $m/z$ 736 ([M+NH$_4$]⁺) for acylsugar #7 showing cleavage of glycosidic linkage to yield abundant acylated furanose ($m/z$ 289), and less abundant acylated pyranose ($m/z$ 413) product ions; **(C)** ESI(-) MS/MS spectrum of product ions of $m/z$ 799 ([M+HCOO]⁻) of acylsugar #10 showing evidence of neutral losses of C2, C5, C8$^P$ and H$_2$O, as well as corresponding carboxylate anions for ester groups C5 ($m/z$ 101) and C8$^P$ ($m/z$ 135); **(D)** ESI(+) MS/MS spectrum of products of $m/z$ 772 ([M+NH$_4$]⁺) for acylsugar #10 showing cleavage of glycosidic linkage to produce abundant acylated furanose ($m/z$ 289), and less abundant acylated pyranose ($m/z$ 449) fragment ions.

Another acylsugar, t$_r$ = 54.29 min (#10, Figure 2.1A) exhibiting [M+HCOO]⁻ at $m/z$ 799.34 and [M+NH$_4$]⁺ with $m/z$ 772.38 was assigned the formula C$_{37}$H$_{54}$O$_{16}$. In this case, the formula indicates four more degrees of unsaturation than its corresponding penta-substituted acylsucrose with saturated esters. MS/MS product ion spectra in negative mode (Figure 2.2C) showed abundant fragment ions with masses of C5 saturated acyl anions, as well as $m/z$ 135.04, consistent with formula C$_8$H$_7$O$_2$⁻, suggesting a single acyl group with four unsaturations. Given the precedent of phenylacetyl tropane esters in the Solanaceae [29], we annotated this metabolite as a phenylacetyl (C8$^P$) acylsugar (further confirmed by NMR in Section 2.3.3). Support for this assignment derives from neutral losses of 118.04 Da (C$_8$H$_6$O), three saturated C5 and one C2 from [M-H]⁻. The positive mode MS/MS product ion spectrum of $m/z$ 772.38 (Figure 2.2D) shows cleavage of the glycosidic bond which produced an abundant fragment ion $m/z$ 289.13, consistent with one saturated C5

39

and one C2 acyl chain on the furanose ring, and a pyranose fragment ion, $m/z$ 449.22, with substitutions of two C5 and one C8$^P$ esters on this ring. Thus, this acylsugar was annotated S5:25:4(2,5,5,5,8$^P$).

Deep analysis of *S. sinuata* LC/MS profile data using positive-ion mode revealed 127 different MFs consistent with tri- through hepta-substituted acylsucroses (Appendix Table 2.4). All acylsugars appear to have at least one C5 or C6 acyl chain, illustrated in the combined XIC for C5 and C6 acyl anions ($m/z$ 101 and 115) in Figure 2.1B. We found no evidence of saturated long chain esters exceeding C6 carbon chain lengths. Notably, negative-ion mode CID spectra of hexaacylsucroses (e.g. acylsucrose $t_r$ = 62.99 min; #16 in Figure 2.1A), shows weaker fragment ion signals corresponding to C5 and C6 acyl anions (Figure 2.1B) than acylsucroses with fewer acylations under the same CID conditions. Also of note, heptaacylsucroses were not detected in negative ion mode, but were detected at low levels in positive ion mode (Appendix Figure 2.6). The lack of heptaacylsucrose signals in negative-ion mode suggests that at least two non-esterified hydroxyl positions facilitate formation of formate adduct ions. In addition, further inspection of XICs for fragment acyl anions of C5$^T$ (Figure 2.1C) and C8$^P$ (Figure 2.1D) revealed a complex series of acylsugars that includes a diverse array of tiglyl and phenylacetyl esters.

### 2.3.2 GC/MS profiling reveals sugar ester composition

As aliphatic ester groups of four or more carbon atoms exhibit isomeric forms that differ by their branching positions [15], determining the acyl group composition of acylsugars is aided by gas chromatographic resolution of isomers. This experimental approach has been important for distinguishing branching isomers of acyl groups because CID spectra often fail to generate fragment ion information that distinguishes branching positions. In this study, sugar esters were analyzed via transesterification to generate ethyl esters followed by GC/MS analysis. Acyl group identifications were based on comparisons of their mass spectra to the National Institute of Standards and Technology (NIST11) mass spectrum library, and retention times were referenced to an *n*-alkane series to generate retention indices. These results demonstrate a high abundance of aiC5 and aiC6 acyl groups (Appendix Table 2.7). In comparison, minor signals consistent

with iC4, C5$^T$, iC5 and C8$^P$ were also observed. Taken together, analysis of LC/MS and GC/MS profiling suggested eight different ester groups: C2, C3, iC4, iC5, aiC5, C5$^T$, aiC6, and C8$^P$ (Figure 2.3).
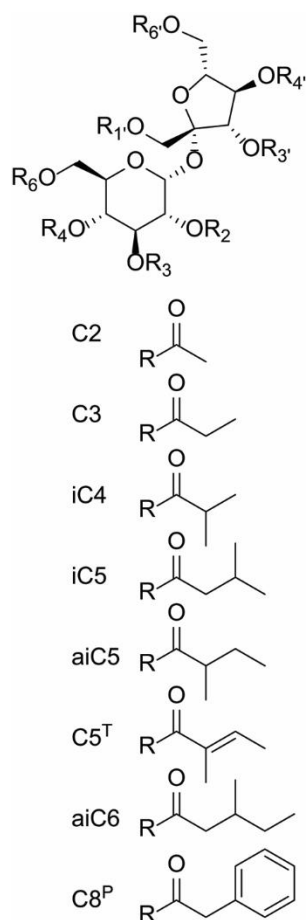


**Figure 2.3.** Hydroxyl positions on sucrose available for acylation, and ester groups observed in *S. sinuata* GT extracts

### 2.3.3 1D and 2D NMR of purified acylsucroses reveals structural diversity

To better understand the diversity of acylsugar metabolites, sixteen abundant acylsugar metabolites (Figure 2.1A and Table 2.1) were purified by HPLC for structure elucidation by NMR spectroscopy, including eight tetra-substituted, seven penta-substituted, two of which contain unsaturated esters, and one hexa-substituted acylsugars. 1D ($^1$H and $^{13}$C) and 2D (gCOSY, gHSQC, gHMBC, *J*-resolved and ROESY) NMR spectroscopic techniques served as the basis for structure elucidation of purified acylsugars. Proton resonances (determined from $^1$H, COSY and *J*-resolved) displayed several informative chemical shift regions (Appendix Table 2.10). Sucrose core hydrogens of acylated positions in the range of 4.0-5.6 ppm

were often apparent due to ~1 ppm downfield shifts compared to hydroxyl containing analogues, which ranged from 3.5-4.6 ppm. Aliphatic H's from acyl groups displayed signals ranging from 0.8-1.9 ppm, while $\alpha$-H's to ester groups were in the range 2.0-2.5 ppm. Similar chemical shift regions and coupling constants were reported for Solanaceae acylsugars with saturated esters [12, 14]. In contrast, characteristic resonances were observed for acylsugars with unsaturated esters. For example, acylsucrose #7 containing a C5$^T$ ester displayed a distinctive multiplet centered at 6.80 ppm for a $\beta$-H alkene, and two overlapping methyl resonances at 1.75 ppm. Acylsucrose #10 containing C8$^P$ ester showed unique multiplets for phenyl H's in the range 7.1-7.4 ppm and a methylene signal centered at 3.54 ppm (Figure 2.4). COSY and ROESY correlations, along with couplings from $J$-resolved spectra, assisted determination of neighboring ring H's on sucrose and atomic connectivity of acyl groups.
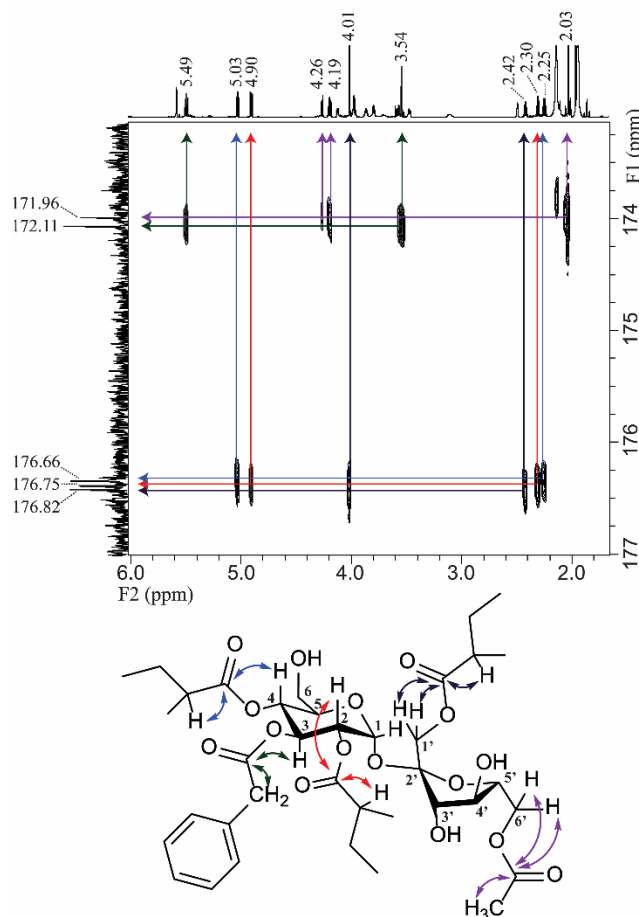
**Figure 2.4.** HMBC spectrum (top) and correlations of acylsucrose #10 (bottom); ^{1}H spectrum is projected on the F2 axis and ^{13}C spectrum is projected on the F1 axis. Cross peaks of sucrose ring H's and α-H's of esters to carbonyl C's of ester groups are represented by colored arrows for clarity.

The ^{13}C, HSQC and HMBC spectra were used to establish atomic connectivity by assigning carbon chemical shifts of sucrose ring (Appendix Table 2.11) and acyl carbons. HSQC measurements provided carbon attachments to protons and suggested the number of protons attached to each carbon. HMBC was used to measure atomic connectivity, usually ranging 2-4 bonds, and was vital for the assignment of positions of attachment of specific acyl groups. Multiple bond correlations of ring H's of sucrose and acyl H's to ester carbons enabled determination of positions and types of acylations. For instance, Figure 2.4 shows an HMBC spectrum of acylsucrose #10, a pentaacylsucrose with five carbonyl carbon resonances (176.75, 172.11, 176.66, 176.82, 171.96 ppm) that show cross peaks for three bond couplings to sucrose ring H's at 4.90, 5.49, 5.03, 4.01 and a pair at 4.26 and 4.19 ppm (positions **2**, **3**, **4**, **1′** and **6′** on sucrose

respectively). Similarly, $\alpha$-H's of these acyl groups centered at 2.25, 3.54, 2.30, 2.42 and 2.03 ppm also show cross peaks for two bond correlations to corresponding carbonyl carbons. Additional correlations from 1D and 2D NMR show these acyl groups are aiC5, C8$^P$, aiC5, aiC5 and C2 esters at these positions. Characteristic $^{13}$C chemical shifts were also observed for acylsugars with unsaturated esters. For the C8$^P$ ester, $^{13}$C resonances at 135.06, 130.86, 129.82 and 128.47 ppm are consistent with phenyl carbons and further confirmed the presence of C8$^P$ ester group. For C5$^T$, distinctive $^{13}$C signals at 139.16 and 127.73 ppm for alkene carbons confirmed the presence of a branched monounsaturated ester such as tiglic or angelic acid (*E* vs. *Z* isomer). The $\beta$-methyl carbon, which showed $^{13}$C chemical shift at 12.08 ppm, agreed with literature values for tiglic acid [30].

Table 2.1 summarizes the structures of NMR resolved acylsucroses purified from *S. sinuata* extracts (further evidence of these structural assignments is provided in Appendix Tables 2.12-2.27 and Appendix Figures 2.31-2.141). Of eight hydroxyl positions on sucrose available for acylation (Figure 2.3), the full range of NMR structures identified six of the eight having at least one saturated or unsaturated esters. The dominant acyl groups in *S. sinuata* acylsucroses are C2, aiC5 and aiC6. All purified acylsugars contain one or more aiC5 or aiC6 groups, and the pyranose ring was always tri-substituted at **2**, **3** and **4** positions with mostly aiC5 and/or aiC6 acyl chains, though some acylsugars are substituted with C3 and iC4 acylations at position **4** of this ring. In addition, position **3** shows dominant aiC6 acyl substitutions, except for those acylsugars that contain unsaturated acylations of C5$^T$ and C8$^P$ at the **3** position. Finally, the furanose ring is substituted with differing degrees and combinations (mono-, di-, tri-) of C2 and/or aiC5 acylations. The **1′** position was acylated with aiC5 or not at all, while the **3′** and **6′** positions exhibited only C2 acylations or were not acylated. Aside from short chain C2 and C3 acyl groups, all other aliphatic esters were branched iC4, aiC5 and aiC6 acyl chains.

**Table 2.1.** Summary of NMR resolved acylsucroses purified from *S. sinuata* extracts and percent peak area of $[M+NH_4]^+$ ion.

| Acylsucrose # | Acylsugar ID | Ret. time (min) | Experimental *m/z* of $[M+NH_4]^+$ | Theoretical *m/z* of $[M+NH_4]^+$ | Analyte Molecular Formula | $R_2$ | $R_3$ | $R_4$ | $R_{1'}$ | $R_{3'}$ | $R_{6'}$ | Peak Area ($\times 10^5$) of $[M+NH_4]^+$ | % of Total Acylsucrose Peak Area |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Tetraacylsucroses* | | | | | | | | | | | | |
| 1 | S4:19:0(3,5,5,6) | 48.21 | 682.3664 | 682.3644 | $C_{31}H_{52}O_{15}$ | aiC5 | aiC6 | C3 | aiC5 | H | H | 3.11 | 2.8% |
| 2 | S4:19:0(2,5,6,6) | 48.39 | 682.3675 | 682.3644 | $C_{31}H_{52}O_{15}$ | aiC6 | aiC6 | aiC5 | H | C2 | H | 3.52 | 3.1% |
| 4 | S4:20:0(4,5,5,6) | 50.78 | 696.3804 | 696.3801 | $C_{32}H_{54}O_{15}$ | aiC5 | aiC6 | iC4 | aiC5 | H | H | 0.99 | 0.9% |
| 5 | S4:20:0(3,5,6,6) | 51.42 | 696.3815 | 696.3801 | $C_{32}H_{54}O_{15}$ | aiC6 | aiC6 | C3 | aiC5 | H | H | 3.02 | 2.7% |
| 6 | S4:20:0(2,6,6,6) | 51.65 | 696.3815 | 696.3801 | $C_{32}H_{54}O_{15}$ | aiC6 | aiC6 | aiC6 | H | C2 | H | 2.09 | 1.9% |
| 9 | S4:21:0(5,5,5,6) | 53.44 | 710.3993 | 710.3957 | $C_{33}H_{56}O_{15}$ | aiC5 | aiC6 | aiC5 | aiC5 | H | H | 8.06 | 7.2% |
| 12 | S4:22:0(5,5,6,6) | 56.49 | 724.4154 | 724.4114 | $C_{34}H_{58}O_{15}$ | aiC6 | aiC6 | aiC5 | aiC5 | H | H | 9.91 | 8.8% |
| 14 | S4:23:0(5,6,6,6) | 59.67 | 738.4266 | 738.4270 | $C_{35}H_{60}O_{15}$ | aiC6 | aiC6 | aiC6 | aiC5 | H | H | 2.45 | 2.2% |
| | *Pentaacylsucroses* | | | | | | | | | | | | |
| 3 | S5:20:0(2,2,5,5,6) | 49.57 | 710.3621 | 710.3594 | $C_{32}H_{52}O_{16}$ | aiC5 | aiC6 | aiC5 | H | C2 | C2 | 5.29 | 4.7% |
| 8 | S5:21:0(2,2,5,6,6) | 52.81 | 724.3783 | 724.3750 | $C_{33}H_{54}O_{16}$ | aiC6 | aiC6 | aiC5 | H | C2 | C2 | 10.62 | 9.5% |
| 11 | S5:22:0(2,2,6,6,6) | 56.08 | 738.3937 | 738.3907 | $C_{34}H_{56}O_{16}$ | aiC6 | aiC6 | aiC6 | H | C2 | C2 | 4.93 | 4.4% |
| 13 | S5:23:0(2,5,5,5,6) | 58.80 | 752.4092 | 752.4063 | $C_{35}H_{58}O_{16}$ | aiC5 | aiC6 | aiC5 | aiC5 | H | C2 | 6.98 | 6.2% |
| 15 | S5:24:0(2,5,5,6,6) | 61.76 | 766.4236 | 766.4220 | $C_{36}H_{60}O_{16}$ | aiC6 | aiC6 | aiC5 | aiC5 | H | C2 | 4.90 | 4.4% |
| | *Pentaacylsucroses w/unsaturated ester* | | | | | | | | | | | | |
| 7 | S5:22:1(2,5,5,5,5$^T$) | 52.55 | 736.3743 | 736.3750 | $C_{34}H_{54}O_{16}$ | aiC5 | C5$^T$ | aiC5 | aiC5 | H | C2 | 2.84 | 2.5% |
| 10 | S5:25:4(2,5,5,5,8$^P$) | 54.29 | 772.3748 | 772.3750 | $C_{37}H_{54}O_{16}$ | aiC5 | C8$^P$ | aiC5 | aiC5 | H | C2 | 0.78 | 0.7% |
| | *Hexaacylsucrose* | | | | | | | | | | | | |
| 16 | S6:25:0(2,2,5,5,5,6) | 62.99 | 794.4166 | 794.4169 | $C_{37}H_{60}O_{17}$ | aiC5 | aiC6 | aiC5 | aiC5 | C2 | C2 | 2.34 | 2.1% |

% Acylsucrose Peak Area with Structural Identification = 64.1%

### 2.3.4 Structure diversity of acylsucroses from *S. sinuata*

Our broader goal is to understand the evolution of genes involved in plant specialized metabolism, focusing on the family Solanaceae. *S. sinuata* was chosen for this investigation because it is further removed genetically from tomato and its close relatives for which extensive acylsugar profiling has been performed. To this end, this report describes the isolation and structure elucidation of 16 acylsugars from *S. sinuata*, relying on NMR spectroscopy to establish cites of attachment of specific acyl groups. Interpretation of these findings relies on comparison to the published acylsugar structures.

### 2.3.4.1 Acyl group diversity in *S. sinuata*

Acyl groups aiC5 and aiC6 were the dominant medium chain ester groups, consistent with isoleucine serving as the major acyl group precursor as has been observed in other members of the Solanaceae [31]. At least 10-fold lower levels of iC4 and iC5 were detected, and fragment ion evidence was found for aliphatic esters up to C6. While C2 esters were abundant, lesser amounts of C3 esters were observed. The observation of acylsugars containing unsaturated $C5^T$ and $C8^P$ in the Solanaceae has not been described before. We postulate that $C5^T$ is derived from isoleucine pathway metabolism and $C8^P$ arises from phenylalanine by chemistry similar to other amino acids. To our knowledge, the only prior report of an unsaturated C5 acylsugar was in tobacco [24], and we are not aware of $C8^P$ in acylsugars from prior reports. Amazingly, more than 400 acylsugars were annotated by LC/MS (Appendix Table 2.4) despite only eight of the > 20 acyl groups known in acylsugars of the Solanaceae [15] being detected. This is slightly more than the > 300 peaks detected using automated processing of LC/MS data using a steeper gradient [21], and represents a more diverse group of acylsucroses than has been reported for any plant, with the 73 acylsugars annotated from *S. habrochaites* being the largest number described to date [32]. In contrast to an earlier report of an acylglucose identified from *S. sinuata* [33], we generated extracted ion chromatograms but found no evidence of acylglucoses, only acylsucroses.

**2.3.4.2 Number of acylations in *S. sinuata* acylsucroses**

The *S. sinuata* acylsucroses are notably different from those produced by cultivated and wild tomato in that the maximum number of acyl groups (7) exceeds the number reported in any accessions in the genera *Solanum*, *Nicotiana*, or *Petunia*. [12, 14]. In contrast, *Petunia* species show high levels of malonylated pentaacylsucroses [14]. Here, we present NMR structures for as many as six acyl groups on sucrose from *S. sinuata*, and deep mining of LC/MS data suggests that *S. sinuata* produces detectable levels of tri- through heptaacylsucroses (Appendix Table 2.4). Penta- and hexaacylsucroses were of high abundance in *S. sinuata* (Appendix Figure 2.7) and their structures led to identifications of *S. sinuata* ASATs in our recent manuscript [21].

**2.3.4.3 Positions of acyl groups in *S. sinuata* acylsucroses**

The diversity in position selectivity of acyl group attachment to the sucrose core is revealed from NMR spectra. *S. sinuata* acylsucroses are distinguished from those accumulated by tomato and its relatives in that C2 acylation in *S. sinuata* was found at the **3′** and **6′** positions, but only at the **2** position in tomato [12, 13]. No other acyl groups were observed at the **3′** or **6′** positions of *S. sinuata* acylsucroses, in contrast to tomato species, which often have iC5, iC10, aiC11 or nC12 at the **3′** position. Our recent paper found evidence that SsASAT5 catalyzes acetylation at the **6′** position as well as positions **6** and **1′** [21].

Although less comprehensive, other reports of positions of acyl groups outside of the tomato clade show varied attachment selectivity. Reported wild potato (*Solanum berthaultii*) acylsucroses contrast with *S. sinuata* in that the former are not known to exhibit acylation at the **2** position, and exhibit acylation by iC4 at the **3′** position and C10 esters at the **6** position [34, 35]. In a previous report, structures of one triacylglucose and one pentaacylsucrose that displayed C2 acylation at the **6** position of sucrose were elucidated from *S. sinuata* [33]. However, we did not observe any acylsugars substituted at this position from our NMR results. Interestingly, XICs of positive mode spectra at elevated collision potentials did show evidence of low abundance heptaacylsucroses with fully substituted furanose ring at the **1′**, **3′**, **4′** and **6′** positions (Appendix Figure 2.6). As with *S. sinuata*, reports of petunia acylsugars described at least four

tetraacylsucroses substituted with C2 acyl groups at the **6′** position [36-38], but a recent manuscript from our laboratory yielded evidence that these may form by decarboxylation of malonate esters at the **6′** position [14]. Petunia species also exhibit iC4, aiC5 and aiC6 esters, in addition to straight chain C6, C7 and iso-branched iC5, iC6, iC7 and iC8 esters at the **2**, **3**, and **4** positions. Tobacco species produce acylsugars that bear resemblance to those found in *S. sinuata* [24, 39]. These acylsucroses exhibit extensive aiC6, aiC5 and iC4 at the **2**, **3** or **4** positions (relative positions are uncertain), including C2 groups at the **3′** position. In addition, acylglucoses from these species were found to contain C3 and unsaturated C5 esters, which are presumed to be $C5^T$ [24].

## 2.4 Conclusions

Identifying novel metabolites from complex profiles continues to present a central challenge in functional genomic investigations of metabolism. In this work, LC/MS profiling and NMR spectroscopy of acylsugar SMs from GTs of *S. sinuata* revealed diverse acylsucrose compositions. In addition to documentation of hundreds of acylsucrose metabolites, documentation of unsaturated ester groups $C5^T$ and $C8^P$ extends our understanding of acylsugar diversity in the Solanaceae, where only saturated forms were previously reported. Acylsucrose structures provided evidence for enzyme mediated and position-selective attachment of ester groups to the sucrose core that differed from reports for other members of the Solanaceae [21], including evidence for as many as seven acylations on a single sucrose core.

Considering the roles that acylsugars have in physical and chemical defense, understanding the mechanisms involved in their formation may provide substantial guidance for engineering acylsugar content of plants. The results herein guided the discovery of four acylsugar biosynthetic enzymes of the *S. sinuata* pathway and extended our understanding of the evolution of specialized metabolism in the Solanaceae. We postulate that promiscuity is an important feature of these SM enzymes, and thus, additional enzymes that acylate with unsaturated acyl-CoA groups may not be necessary.

**APPENDIX**

**Figure 2.5.** Micrographs of *S. sinuata* Type I/IV glandular trichomes located on the surface of a young leaflet.
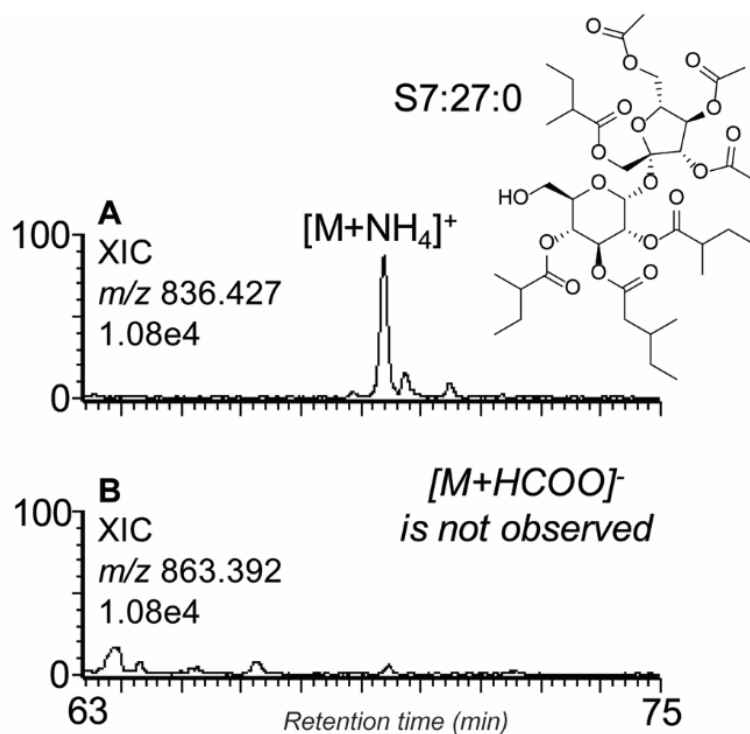
**Figure 2.6.** LC/MS shows evidence for heptaacylsucrose S7:27:0 using positive-ion mode ESI but negligible signal is detected in negative-ion mode. **(A)** Positive-ion mode extracted ion chromatogram for [M+NH4]+ (*m/z* 836.427) for S7:27:0 showing one major and three minor chromatographic peaks; **(B)** Negative-ion mode chromatogram for [M+formate]- (*m/z* 863.392) on the same absolute y-axis scale (100% = 1.08 x 10⁴ ion counts) shows negligible signal for S7:27:0. The structure shown is putative and not solved by NMR, but is proposed based on our knowledge of acyl group selectivity (established by NMR resolved structures of other acylsucroses). The [M+NH4]+ consistent with S7:27:0 and fragment ion *m/z* 373.15 (Function 3 = 25 V) are consistent with an acylsucrose that has a furanose ring substituted by one C5 and three C2 ester groups.

**Figure 2.7**. Sum of extracted ion peak areas of $[M+NH_4]^+$ ions for *S. sinuata* acylsucroses, organized by the number of acyl groups on the sucrose core (e.g. S4, S5, and S6 refer to tetra-, penta-, and hexa-acylsucroses). Separate bars indicate peak area totals for acylsucroses containing all saturated acyl groups (green) and one or two $C5^T$ and $C8^P$ acyl groups (blue), which accounted for ~14% of the total acylsugar peak area.

**Table 2.2.** Plant cultivation and metabolite extraction metadata

| Species | *Salpiglossis sinuata* |
|---|---|
| Genotype | NYBG, *S. sinuata* seeds were obtained from a single plant growing in The New York Botanical Gardens (NY, USA). The second generation of these seeds were used for the experiments in this study. |
| Organ | Aerial tissues |
| Organ specification | Aerial tissues included leaf and stem tissues |
| Cell type | Glandular trichomes |
| Growth location | Growth chamber at MSU |
| Growth support | Seeds were germinated using 500 µM gibberellic acid on Whatman #1 filter paper placed in petri dishes. 3-4 day old seedlings were transferred to Jiffy peat pellets. |
| Light | 300 µE m$^{-2}$ sec $^{-1}$; 16 h light/8 h dark |
| Humidity | 50% relative humidity |
| Temperature | 25C day/12C night |
| Watering regime | Bottom watering as per requirement |
| Nutritional regime | Half strength Hoagland's solution, once a week |
| Plant growth stage | 10 weeks post germination (plants about 0.5 m height) |
| Metabolism quenching method | Aerial tissues were extracted in 1.9 L of acetonitrile: isopropanol (AcN:IPA, v/v, 1:1) for 10 mins in a 2-L beaker with horizontal mixing at 120 rpm. The presence of acylsugars in trichomes were confirmed by dipping leaf and stem tissue into liquid N$_2$, and scraping trichomes into a solution of AcN:IPA. |
| Harvest method | Plants were cut at the stems in 10-15 cm increments |
| Sample storage | The extract was decanted into two 1 L glass Wheaton bottles with Teflon lined caps, and stored in a freezer at -20°C. |

**Table 2.3.** UHPLC/MS metadata

| | |
|---|---|
| Facility Director | Dr. A. Daniel Jones |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Mass Spectrometry and Metabolomics Core |
| Facility Instrument Title | LCMS: G2-XS QTof #1 |
| LC System | Acquity UPLC I-Class Binary Solvent Manager equipped with Acquity Column Manager |
| Manufacturer | Waters |
| Autosampler | 2777C Sample Manager |
| Column | Ascentis Express C18 Analytical HPLC, 10 cm x 2.1mm x 2.7µm |
| Column Manufacturer | Supelco |
| Catalogue Number | 53823-U |
| Serial Number | USRB002977 |
| Packing Lot Number | S11020 |
| Injection Volume | 10 µL |
| Flow Rate | 0.3 mL/min |
| Mobile Phases: | |
| A | 10 mM ammonium formate in water (pH 2.8, adjusted with formic acid) |
| B | Acetonitrile |
| Gradient Profile | Hold 1% B at 0-1 min, linear 1-100% B at 1-100 min, hold 100% B at 100-107 min, linear 100-1% B at 107-108 min, and hold 1% B at 108-110 min |
| Column Oven Temperature | 50 ºC |
| Sample Temperature in autosampler | 10 °C |
| Inlet Method Name | SMH_110min |
| Mass Spectrometer | Xevo G2-XS QTof |
| Manufacturer | Waters |
| Software | MassLynx v4.1 |
| Ionization Source | Electrospray Ionization (ESI) |
| Data Acquisition | Sensitivity Mode |
| Polarity | Positive, Negative |
| Mass Range | *m/z* 50-1500 |
| Data Format | Centroid |
| Capillary Voltage | 2.0 kV |
| Sample Cone | 35 V |
| Source Temperature | 100 °C |
| Source Offset | 80 V |
| Desolvation Temperature | 350 °C |
| Cone Gas Flow | 50 L/hr (ESI+), 0 L/hr (ESI-) |

**Table 2.3.** (continued)

| | |
|---|---|
| Desolvation Gas Flow | 600 L/hr |
| Collision Potential | |
| Function 1 | 0 V |
| Function 2 | 10 V |
| Function 3 | 25 V |
| Function 4 | 50 V |
| Scan Duration | 0.10 s |
| Inter Scan Delay | 0.014 s |
| Collision Cell Pressure | 0.06 mbar (ESI+), 0.05 mbar (ESI-) |
| Lock Spray | Leu-enkephalin |
| Lock mass (*m/z*) | 556.2771 (ESI+), 554.2615 (ESI-) |
| Lock Spray Scan Time | 0.2 s |
| Lock Spray Scan Frequency | 10 s |
| MS Method Files | SMH_CID_pos_110 min_Lock, SMH_CID_neg_110 min_Lock |
| Sample handling | Aerial tissues from 28 plants aged 10 weeks were harvested (plants were cut at the stems in 10-15 cm increments) and extracted in 1.9 L of AcN:IPA for 10 mins in a 2 L beaker with horizontal mixing at 120 rpm. A 10 mL aliquot of this bulk extract solution, with 5.0 µM telmisartan added as internal standard, was used for all metabolite profiling. |
| Sample Storage | -20 °C in Wheaton glass vessel with PTFE lined cap |
| Protocol when analyzing the samples | The instrument was calibrated in ESI+/- modes using 500 µM sodium formate solution. First, samples were analyzed in ESI- mode. After column equilibration, a solvent blank sample (AcN:IPA) was analyzed, followed by *S. sinuata* extract. The two samples were then analyzed in ESI+ mode. |
| Deep profiling data analysis parameters: | |
| XIC Window | 40 ppm |
| Enabled Smoothing | Yes |
| Window Size | 10 scans |
| Number of Smooths | 3 |
| Smoothing Method | Savitzky-Golay |
| ApexTrack Peak Integration | Yes |
| Peak-to-Peak Baseline Noise | Automatic |
| Peak Width at 5% Height (mins) | Automatic |
| Baseline Start Threshold | 1% |
| Baseline End Threshold | 1% |
| Note: some peak integrations were adjusted manually | |

**Table 2.4.** Deep profiling results (metabolites highlighted in bold were identified by NMR spectroscopy)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| *Triacylsucroses* | | | | | | |
| 1 | S3:13:0 | 2 | 584.291 | 27.783 | 5698 | 895 |
| 2 | S3:13:0 | 2 | 584.291 | 27.920 | 5691 | 815 |
| 3 | S3:14:0 | 2 | 598.307 | 31.603 | 15935 | 3040 |
| 4 | S3:14:0 | 2 | 598.307 | 32.578 | 875 | 139 |
| 9 | S3:15:0 | 6 | 612.323 | 33.010 | 7136 | 1331 |
| 10 | S3:15:0 | 6 | 612.323 | 33.494 | 590 | 89 |
| 11 | S3:15:0 | 6 | 612.323 | 33.857 | 1739 | 286 |
| 12 | S3:15:0 | 6 | 612.323 | 35.444 | 431 | 77 |
| 13 | S3:15:0 | 6 | 612.323 | 37.260 | 747 | 108 |
| 14 | S3:15:0 | 6 | 612.323 | 38.682 | 2180 | 320 |
| 22 | S3:16:0 | 4 | 626.338 | 36.398 | 27512 | 4412 |
| 23 | S3:16:0 | 4 | 626.338 | 38.341 | 6103 | 956 |
| 24 | S3:16:0 | 4 | 626.338 | 40.656 | 8684 | 1439 |
| 25 | S3:16:0 | 4 | 626.338 | 41.117 | 7793 | 1509 |
| 39 | S3:17:0 | 4 | 640.354 | 43.697 | 84074 | 13052 |
| 40 | S3:17:0 | 4 | 640.354 | 43.999 | 6488 | 1202 |
| 41 | S3:17:0 | 4 | 640.354 | 44.316 | 4210 | 726 |
| 42 | S3:17:0 | 4 | 640.354 | 44.574 | 944 | 153 |
| 69 | S3:18:0 | 1 | 654.370 | 47.001 | 44022 | 7234 |
| 5 | S3:15:1 | 2 | 610.307 | 32.624 | 375 | 74 |
| 6 | S3:15:1 | 2 | 610.307 | 35.505 | 828 | 127 |
| 28 | S3:17:1 | 1 | 638.338 | 47.622 | 1057 | 175 |
| 51 | S3:18:1 | 1 | 652.354 | 51.018 | 634 | 107 |
| 80 | S3:19:1 | 1 | 666.370 | 52.114 | 1238 | 208 |
| 15 | S3:16:4 | 2 | 618.276 | 29.077 | 1912 | 289 |
| 16 | S3:16:4 | 2 | 618.276 | 29.220 | 2004 | 275 |
| 45 | S3:18:4 | 2 | 646.307 | 33.849 | 2039 | 358 |
| 46 | S3:18:4 | 2 | 646.307 | 38.024 | 708 | 99 |
| 71 | S3:19:4 | 4 | 660.323 | 37.729 | 1278 | 184 |
| 72 | S3:19:4 | 4 | 660.323 | 40.247 | 1003 | 151 |
| 73 | S3:19:4 | 4 | 660.323 | 41.230 | 1012 | 143 |
| 74 | S3:19:4 | 4 | 660.323 | 41.624 | 1022 | 216 |
| 99 | S3:20:4 | 2 | 674.338 | 43.379 | 1815 | 260 |
| 100 | S3:20:4 | 2 | 674.338 | 44.354 | 27660 | 4376 |
| 133 | S3:22:8 | 1 | 694.307 | 41.193 | 2423 | 367 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| *Tetraacylsucroses* | | | | | | |
| 7 | S4:14:0 | 2 | 612.286 | 27.209 | 699 | 126 |
| 8 | S4:14:0 | 2 | 612.286 | 27.481 | 920 | 151 |
| 19 | S4:15:0 | 3 | 626.302 | 31.096 | 2619 | 662 |
| 20 | S4:15:0 | 3 | 626.302 | 33.146 | 6268 | 1342 |
| 21 | S4:15:0 | 3 | 626.302 | 35.959 | 719 | 125 |
| 30 | S4:16:0 | 9 | 640.318 | 34.878 | 526 | 86 |
| 31 | S4:16:0 | 9 | 640.318 | 35.559 | 915 | 169 |
| 32 | S4:16:0 | 9 | 640.318 | 36.723 | 3273 | 582 |
| 33 | S4:16:0 | 9 | 640.318 | 37.131 | 908 | 155 |
| 34 | S4:16:0 | 9 | 640.318 | 38.281 | 511 | 77 |
| 35 | S4:16:0 | 9 | 640.318 | 38.674 | 9308 | 1529 |
| 36 | S4:16:0 | 9 | 640.318 | 39.703 | 11615 | 1863 |
| 37 | S4:16:0 | 9 | 640.318 | 39.998 | 1128 | 175 |
| 38 | S4:16:0 | 9 | 640.318 | 40.142 | 1247 | 196 |
| 56 | S4:17:0 | 13 | 654.333 | 38.205 | 26244 | 4426 |
| 57 | S4:17:0 | 13 | 654.333 | 38.576 | 710 | 152 |
| 58 | S4:17:0 | 13 | 654.333 | 38.977 | 1479 | 269 |
| 59 | S4:17:0 | 13 | 654.333 | 39.461 | 923 | 144 |
| 60 | S4:17:0 | 13 | 654.333 | 40.035 | 725 | 114 |
| 61 | S4:17:0 | 13 | 654.333 | 41.133 | 550 | 69 |
| 62 | S4:17:0 | 13 | 654.333 | 41.836 | 7274 | 1259 |
| 63 | S4:17:0 | 13 | 654.333 | 42.146 | 9669 | 2010 |
| 64 | S4:17:0 | 13 | 654.333 | 42.441 | 8626 | 1791 |
| 65 | S4:17:0 | 13 | 654.333 | 42.691 | 3578 | 402 |
| 66 | S4:17:0 | 13 | 654.333 | 43.031 | 48568 | 9418 |
| 67 | S4:17:0 | 13 | 654.333 | 44.370 | 705 | 133 |
| 68 | S4:17:0 | 13 | 654.333 | 44.800 | 391 | 79 |
| 88 | S4:18:0 | 7 | 668.349 | 41.450 | 17944 | 2885 |
| 89 | S4:18:0 | 7 | 668.349 | 42.932 | 8737 | 1422 |
| 90 | S4:18:0 | 7 | 668.349 | 43.704 | 30739 | 4900 |
| 91 | S4:18:0 | 7 | 668.349 | 45.163 | 431891 | 91237 |
| 92 | S4:18:0 | 7 | 668.349 | 45.678 | 125428 | 39088 |
| 93 | S4:18:0 | 7 | 668.349 | 46.125 | 18917 | 4981 |
| 94 | S4:18:0 | 7 | 668.349 | 47.523 | 6581 | 1122 |
| 118 | S4:19:0 | 7 | 682.364 | 47.001 | 19685 | 3248 |
| 119 | S4:19:0 | 7 | 682.364 | 47.282 | 9253 | 1445 |
| 120 | S4:19:0 | 7 | 682.364 | 47.599 | 17753 | 2456 |
| **121** | **S4:19:0(3,5,5,6)** | **7** | **682.364** | **48.211** | **1752487** | **311272** |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| **122** | **S4:19:0(2,5,6,6)** | **7** | **682.364** | **48.393** | **2052739** | **351859** |
| 123 | S4:19:0 | 7 | 682.364 | 48.869 | 222575 | 47411 |
| 124 | S4:19:0 | 7 | 682.364 | 50.813 | 3345 | 558 |
| 153 | S4:20:0 | 5 | 696.380 | 50.291 | 681323 | 136081 |
| **154** | **S4:20:0(4,5,5,6)** | **5** | **696.380** | **50.783** | **524859** | **98549** |
| **155** | **S4:20:0(3,5,6,6)** | **5** | **696.380** | **51.418** | **1552608** | **301912** |
| **156** | **S4:20:0(2,6,6,6)** | **5** | **696.380** | **51.645** | **1319806** | **209027** |
| 157 | S4:20:0 | 5 | 696.380 | 52.364 | 1771 | 298 |
| **184** | **S4:21:0(5,5,5,6)** | **4** | **710.396** | **53.438** | **3509795** | **806099** |
| 185 | S4:21:0 | 4 | 710.396 | 53.899 | 1343764 | 240461 |
| 186 | S4:21:0 | 4 | 710.396 | 54.263 | 10410 | 1101 |
| 187 | S4:21:0 | 4 | 710.396 | 54.611 | 166994 | 27295 |
| **225** | **S4:22:0(5,5,6,6)** | **3** | **724.411** | **56.494** | **4000121** | **991011** |
| 226 | S4:22:0 | 3 | 724.411 | 57.053 | 276237 | 51593 |
| 227 | S4:22:0 | 3 | 724.411 | 57.401 | 7417 | 968 |
| **259** | **S4:23:0(5,6,6,6)** | **2** | **738.427** | **59.670** | **1282562** | **245010** |
| 260 | S4:23:0 | 2 | 738.427 | 60.124 | 45298 | 9220 |
| 288 | S4:24:0 | 1 | 752.443 | 62.756 | 100656 | 19118 |
| 17 | S4:15:1 | 2 | 624.286 | 31.678 | 1158 | 156 |
| 18 | S4:15:1 | 2 | 624.286 | 31.777 | 1194 | 156 |
| 26 | S4:16:1 | 2 | 638.302 | 34.500 | 327 | 65 |
| 27 | S4:16:1 | 2 | 638.302 | 37.010 | 403 | 68 |
| 47 | S4:17:1 | 4 | 652.318 | 35.725 | 526 | 94 |
| 48 | S4:17:1 | 4 | 652.318 | 37.238 | 7592 | 1267 |
| 49 | S4:17:1 | 4 | 652.318 | 39.211 | 1323 | 237 |
| 50 | S4:17:1 | 4 | 652.318 | 40.437 | 3667 | 548 |
| 77 | S4:18:1 | 3 | 666.333 | 42.366 | 80341 | 13161 |
| 78 | S4:18:1 | 3 | 666.333 | 43.220 | 1236 | 241 |
| 79 | S4:18:1 | 3 | 666.333 | 43.697 | 530 | 83 |
| 105 | S4:19:1 | 7 | 680.349 | 44.475 | 3920 | 772 |
| 106 | S4:19:1 | 7 | 680.349 | 44.733 | 9162 | 1507 |
| 107 | S4:19:1 | 7 | 680.349 | 45.133 | 3149 | 467 |
| 108 | S4:19:1 | 7 | 680.349 | 45.299 | 10693 | 1254 |
| 109 | S4:19:1 | 7 | 680.349 | 45.542 | 33346 | 8161 |
| 110 | S4:19:1 | 7 | 680.349 | 45.852 | 22612 | 3807 |
| 111 | S4:19:1 | 7 | 680.349 | 46.419 | 1297 | 254 |
| 138 | S4:20:1 | 5 | 694.364 | 47.138 | 317379 | 53412 |
| 139 | S4:20:1 | 5 | 694.364 | 47.607 | 5991 | 1034 |
| 140 | S4:20:1 | 5 | 694.364 | 47.856 | 12153 | 2086 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH₄]⁺ m/z XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 141 | S4:20:1 | 5 | 694.364 | 48.326 | 22296 | 4448 |
| 142 | S4:20:1 | 5 | 694.364 | 48.710 | 2359 | 463 |
| 172 | S4:21:1 | 4 | 708.380 | 50.209 | 261538 | 50055 |
| 173 | S4:21:1 | 4 | 708.380 | 50.852 | 101670 | 18843 |
| 174 | S4:21:1 | 4 | 708.380 | 51.282 | 1263 | 200 |
| 175 | S4:21:1 | 4 | 708.380 | 51.570 | 1888 | 304 |
| 208 | S4:22:1 | 4 | 722.396 | 53.037 | 67327 | 11935 |
| 209 | S4:22:1 | 4 | 722.396 | 53.347 | 15275 | 2827 |
| 210 | S4:22:1 | 4 | 722.396 | 53.582 | 9632 | 1497 |
| 211 | S4:22:1 | 4 | 722.396 | 54.051 | 4834 | 1374 |
| 75 | S4:18:2 | 1 | 664.318 | 40.649 | 4108 | 621 |
| 132 | S4:20:2 | 1 | 692.349 | 45.360 | 9900 | 1608 |
| 43 | S4:17:4 | 2 | 646.271 | 28.622 | 403 | 73 |
| 44 | S4:17:4 | 2 | 646.271 | 28.857 | 455 | 82 |
| 70 | S4:18:4 | 1 | 660.286 | 34.182 | 1944 | 474 |
| 95 | S4:19:4 | 4 | 674.302 | 36.314 | 162 | 28 |
| 96 | S4:19:4 | 4 | 674.302 | 36.912 | 256 | 35 |
| 97 | S4:19:4 | 4 | 674.302 | 39.052 | 303 | 45 |
| 98 | S4:19:4 | 4 | 674.302 | 39.635 | 1018 | 134 |
| 125 | S4:20:4 | 6 | 688.318 | 38.712 | 4884 | 817 |
| 126 | S4:20:4 | 6 | 688.318 | 39.733 | 2170 | 353 |
| 127 | S4:20:4 | 6 | 688.318 | 41.381 | 4942 | 758 |
| 128 | S4:20:4 | 6 | 688.318 | 41.957 | 1939 | 481 |
| 129 | S4:20:4 | 6 | 688.318 | 42.131 | 1395 | 194 |
| 130 | S4:20:4 | 6 | 688.318 | 42.743 | 1865 | 519 |
| 160 | S4:21:4 | 8 | 702.333 | 43.129 | 592 | 115 |
| 161 | S4:21:4 | 8 | 702.333 | 43.735 | 698 | 109 |
| 162 | S4:21:4 | 8 | 702.333 | 44.408 | 30423 | 5201 |
| 163 | S4:21:4 | 8 | 702.333 | 44.718 | 115687 | 18449 |
| 164 | S4:21:4 | 8 | 702.333 | 44.937 | 8303 | 1032 |
| 165 | S4:21:4 | 8 | 702.333 | 45.565 | 24427 | 5952 |
| 166 | S4:21:4 | 8 | 702.333 | 45.905 | 5795 | 1020 |
| 167 | S4:21:4 | 8 | 702.333 | 47.901 | 725 | 93 |
| 193 | S4:22:4 | 4 | 716.349 | 46.858 | 7578 | 1697 |
| 194 | S4:22:4 | 4 | 716.349 | 47.478 | 18341 | 3312 |
| 195 | S4:22:4 | 4 | 716.349 | 47.819 | 114431 | 19553 |
| 196 | S4:22:4 | 4 | 716.349 | 48.628 | 99948 | 17053 |
| 237 | S4:23:4 | 4 | 730.364 | 49.209 | 175130 | 31121 |
| 238 | S4:23:4 | 4 | 730.364 | 49.905 | 35714 | 7685 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH4]+ _m/z_ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 239 | S4:23:4 | 4 | 730.364 | 50.246 | 19092 | 3845 |
| 240 | S4:23:4 | 4 | 730.364 | 51.865 | 10483 | 1955 |
| 267 | S4:24:4 | 4 | 744.380 | 52.190 | 294850 | 49202 |
| 268 | S4:24:4 | 4 | 744.380 | 52.742 | 180719 | 33086 |
| 269 | S4:24:4 | 4 | 744.380 | 53.718 | 48407 | 11307 |
| 270 | S4:24:4 | 4 | 744.380 | 54.240 | 19602 | 3512 |
| 297 | S4:25:4 | 4 | 758.396 | 55.177 | 26389 | 4796 |
| 298 | S4:25:4 | 4 | 758.396 | 55.821 | 7229 | 1539 |
| 299 | S4:25:4 | 4 | 758.396 | 56.728 | 159085 | 28758 |
| 300 | S4:25:4 | 4 | 758.396 | 57.265 | 1592 | 293 |
| 330 | S4:26:4 | 1 | 772.411 | 59.776 | 10590 | 2229 |
| 230 | S4:23:5 | 1 | 728.349 | 46.858 | 9117 | 1552 |
| 262 | S4:24:5 | 1 | 742.364 | 50.708 | 3703 | 718 |
| 243 | S4:24:8 | 2 | 736.318 | 44.823 | 2078 | 309 |
| 244 | S4:24:8 | 2 | 736.318 | 45.020 | 4003 | 594 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH₄]⁺ m/z XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| *Pentaacylsucroses* | | | | | | |
| 29 | S5:15:0 | 1 | 640.281 | 33.668 | 251 | 66 |
| 52 | S5:16:0 | 4 | 654.297 | 35.876 | 203 | 45 |
| 53 | S5:16:0 | 4 | 654.297 | 36.587 | 229 | 36 |
| 54 | S5:16:0 | 4 | 654.297 | 37.744 | 372 | 63 |
| 55 | S5:16:0 | 4 | 654.297 | 38.364 | 706 | 93 |
| 81 | S5:17:0 | 7 | 668.312 | 36.882 | 338 | 65 |
| 82 | S5:17:0 | 7 | 668.312 | 37.676 | 329 | 66 |
| 83 | S5:17:0 | 7 | 668.312 | 39.551 | 15768 | 2761 |
| 84 | S5:17:0 | 7 | 668.312 | 40.399 | 7074 | 1275 |
| 85 | S5:17:0 | 7 | 668.312 | 40.633 | 2776 | 473 |
| 86 | S5:17:0 | 7 | 668.312 | 40.951 | 1796 | 384 |
| 87 | S5:17:0 | 7 | 668.312 | 41.186 | 677 | 104 |
| 112 | S5:18:0 | 6 | 682.328 | 42.978 | 5011 | 731 |
| 113 | S5:18:0 | 6 | 682.328 | 43.098 | 5809 | 998 |
| 114 | S5:18:0 | 6 | 682.328 | 43.288 | 3256 | 408 |
| 115 | S5:18:0 | 6 | 682.328 | 43.817 | 152580 | 26965 |
| 116 | S5:18:0 | 6 | 682.328 | 44.060 | 89101 | 15755 |
| 117 | S5:18:0 | 6 | 682.328 | 45.020 | 5186 | 790 |
| 144 | S5:19:0 | 9 | 696.344 | 45.678 | 4394 | 601 |
| 145 | S5:19:0 | 9 | 696.344 | 46.207 | 153189 | 26924 |
| 146 | S5:19:0 | 9 | 696.344 | 46.435 | 40705 | 6064 |
| 147 | S5:19:0 | 9 | 696.344 | 46.714 | 161029 | 30770 |
| 148 | S5:19:0 | 9 | 696.344 | 47.198 | 356435 | 79325 |
| 149 | S5:19:0 | 9 | 696.344 | 47.773 | 7117 | 1350 |
| 150 | S5:19:0 | 9 | 696.344 | 48.174 | 4093 | 758 |
| 151 | S5:19:0 | 9 | 696.344 | 48.544 | 3645 | 597 |
| 152 | S5:19:0 | 9 | 696.344 | 48.915 | 3619 | 709 |
| 176 | S5:20:0 | 8 | 710.359 | 49.150 | 265038 | 47394 |
| **177** | **S5:20:0(2,2,5,5,6)** | **8** | **710.359** | **49.573** | **2410053** | **528518** |
| 178 | S5:20:0 | 8 | 710.359 | 49.989 | 1325229 | 279439 |
| 179 | S5:20:0 | 8 | 710.359 | 50.488 | 449698 | 92180 |
| 180 | S5:20:0 | 8 | 710.359 | 50.791 | 38080 | 9636 |
| 181 | S5:20:0 | 8 | 710.359 | 51.351 | 4877 | 924 |
| 182 | S5:20:0 | 8 | 710.359 | 51.850 | 11982 | 2428 |
| 183 | S5:20:0 | 8 | 710.359 | 52.152 | 13242 | 3344 |
| **216** | **S5:21:0(2,2,5,6,6)** | **9** | **724.375** | **52.810** | **4393749** | **1062497** |
| 217 | S5:21:0 | 9 | 724.375 | 53.226 | 505090 | 120334 |
| 218 | S5:21:0 | 9 | 724.375 | 53.551 | 956454 | 157126 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH₄]⁺ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 219 | S5:21:0 | 9 | 724.375 | 53.756 | 61114 | 7127 |
| 220 | S5:21:0 | 9 | 724.375 | 53.997 | 7332 | 1260 |
| 221 | S5:21:0 | 9 | 724.375 | 54.322 | 2695 | 341 |
| 222 | S5:21:0 | 9 | 724.375 | 54.890 | 2267 | 466 |
| 223 | S5:21:0 | 9 | 724.375 | 55.117 | 2832 | 489 |
| 224 | S5:21:0 | 9 | 724.375 | 55.420 | 5586 | 1000 |
| 252 | S5:22:0 | 7 | 738.391 | 54.905 | 4914 | 795 |
| 253 | S5:22:0 | 7 | 738.391 | 55.110 | 2344 | 295 |
| 254 | S5:22:0 | 7 | 738.391 | 55.752 | 1388972 | 281505 |
| **255** | **S5:22:0(2,2,6,6,6)** | **7** | **738.391** | **56.077** | **2131308** | **492714** |
| 256 | S5:22:0 | 7 | 738.391 | 56.668 | 324870 | 55041 |
| 257 | S5:22:0 | 7 | 738.391 | 57.559 | 34545 | 8165 |
| 258 | S5:22:0 | 7 | 738.391 | 58.081 | 4282 | 932 |
| 281 | S5:23:0 | 7 | 752.406 | 58.173 | 270116 | 47058 |
| **282** | **S5:23:0(2,5,5,5,6)** | **7** | **752.406** | **58.800** | **2927149** | **697683** |
| 283 | S5:23:0 | 7 | 752.406 | 59.201 | 632895 | 113993 |
| 284 | S5:23:0 | 7 | 752.406 | 59.504 | 5069 | 630 |
| 285 | S5:23:0 | 7 | 752.406 | 59.746 | 27467 | 4375 |
| 286 | S5:23:0 | 7 | 752.406 | 60.479 | 9031 | 1506 |
| 287 | S5:23:0 | 7 | 752.406 | 60.699 | 29338 | 5267 |
| 318 | S5:24:0 | 5 | 766.422 | 60.494 | 2909 | 521 |
| 319 | S5:24:0 | 5 | 766.422 | 61.515 | 293690 | 51296 |
| **320** | **S5:24:0(2,5,5,6,6)** | **5** | **766.422** | **61.758** | **2166846** | **489817** |
| 321 | S5:24:0 | 5 | 766.422 | 62.219 | 93848 | 21032 |
| 322 | S5:24:0 | 5 | 766.422 | 63.762 | 3372 | 671 |
| 345 | S5:25:0 | 4 | 780.438 | 64.117 | 9318 | 2095 |
| 346 | S5:25:0 | 4 | 780.438 | 64.504 | 43629 | 6816 |
| 347 | S5:25:0 | 4 | 780.438 | 64.760 | 197258 | 44599 |
| 348 | S5:25:0 | 4 | 780.438 | 65.139 | 4903 | 982 |
| 374 | S5:26:0 | 3 | 794.453 | 66.938 | 40387 | 8410 |
| 375 | S5:26:0 | 3 | 794.453 | 67.347 | 4532 | 1040 |
| 376 | S5:26:0 | 3 | 794.453 | 67.680 | 16341 | 3146 |
| 396 | S5:27:0 | 1 | 808.469 | 69.760 | 41077 | 8377 |
| 405 | S5:28:0 | 2 | 822.485 | 72.339 | 1106 | 168 |
| 406 | S5:28:0 | 2 | 822.485 | 72.535 | 1922 | 419 |
| 76 | S5:17:1 | 1 | 666.297 | 37.744 | 1365 | 214 |
| 101 | S5:18:1 | 4 | 680.312 | 40.611 | 758 | 143 |
| 102 | S5:18:1 | 4 | 680.312 | 41.079 | 1053 | 211 |
| 103 | S5:18:1 | 4 | 680.312 | 41.867 | 3809 | 810 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH₄]⁺ *m/z* XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 104 | S5:18:1 | 4 | 680.312 | 42.192 | 6716 | 1169 |
| 134 | S5:19:1 | 4 | 694.328 | 43.341 | 45422 | 7384 |
| 135 | S5:19:1 | 4 | 694.328 | 43.764 | 2531 | 579 |
| 136 | S5:19:1 | 4 | 694.328 | 44.150 | 1122 | 142 |
| 137 | S5:19:1 | 4 | 694.328 | 44.959 | 4218 | 829 |
| 169 | S5:20:1 | 3 | 708.344 | 46.578 | 59103 | 12792 |
| 170 | S5:20:1 | 3 | 708.344 | 46.926 | 5197 | 715 |
| 171 | S5:20:1 | 3 | 708.344 | 47.720 | 337358 | 63440 |
| 202 | S5:21:1 | 6 | 722.359 | 49.557 | 40983 | 7908 |
| 203 | S5:21:1 | 6 | 722.359 | 49.792 | 30656 | 3756 |
| 204 | S5:21:1 | 6 | 722.359 | 49.928 | 38048 | 6191 |
| 205 | S5:21:1 | 6 | 722.359 | 50.148 | 83474 | 16724 |
| 206 | S5:21:1 | 6 | 722.359 | 50.836 | 41387 | 11896 |
| 207 | S5:21:1 | 6 | 722.359 | 51.660 | 4663 | 788 |
| **246** | **S5:22:1(2,5,5,5,5ᵀ)** | **4** | **736.375** | **52.553** | **1526524** | **283707** |
| 247 | S5:22:1 | 4 | 736.375 | 52.999 | 24941 | 5713 |
| 248 | S5:22:1 | 4 | 736.375 | 53.196 | 39069 | 8149 |
| 249 | S5:22:1 | 4 | 736.375 | 53.498 | 13193 | 4292 |
| 273 | S5:23:1 | 4 | 750.391 | 55.571 | 474547 | 87436 |
| 274 | S5:23:1 | 4 | 750.391 | 56.016 | 89748 | 16109 |
| 275 | S5:23:1 | 4 | 750.391 | 56.289 | 1823 | 207 |
| 276 | S5:23:1 | 4 | 750.391 | 56.592 | 3047 | 395 |
| 306 | S5:24:1 | 5 | 764.406 | 58.339 | 31285 | 6439 |
| 307 | S5:24:1 | 5 | 764.406 | 58.603 | 20318 | 3552 |
| 308 | S5:24:1 | 5 | 764.406 | 58.884 | 1835 | 269 |
| 309 | S5:24:1 | 5 | 764.406 | 59.087 | 2436 | 575 |
| 310 | S5:24:1 | 5 | 764.406 | 59.504 | 1644 | 281 |
| 338 | S5:25:1 | 2 | 778.422 | 60.873 | 9866 | 1825 |
| 339 | S5:25:1 | 2 | 778.422 | 61.259 | 2892 | 658 |
| 368 | S5:26:1 | 2 | 792.438 | 63.785 | 2461 | 511 |
| 369 | S5:26:1 | 2 | 792.438 | 64.163 | 1062 | 189 |
| 131 | S5:19:2 | 1 | 692.312 | 41.911 | 1032 | 178 |
| 168 | S5:20:2 | 1 | 706.328 | 45.769 | 33592 | 5785 |
| 198 | S5:21:2 | 3 | 720.344 | 48.159 | 5478 | 1167 |
| 199 | S5:21:2 | 3 | 720.344 | 48.703 | 757 | 129 |
| 200 | S5:21:2 | 3 | 720.344 | 49.255 | 664 | 92 |
| 241 | S5:22:2 | 1 | 734.359 | 50.526 | 97918 | 17205 |
| 158 | S5:20:4 | 2 | 702.297 | 40.270 | 12737 | 2160 |
| 159 | S5:20:4 | 2 | 702.297 | 40.679 | 516 | 81 |

Table 2.4. (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH₄]⁺ m/z XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 188 | S5:21:4 | 5 | 716.312 | 42.804 | 5507 | 978 |
| 189 | S5:21:4 | 5 | 716.312 | 43.213 | 1654 | 280 |
| 190 | S5:21:4 | 5 | 716.312 | 43.507 | 14257 | 2253 |
| 191 | S5:21:4 | 5 | 716.312 | 44.090 | 3708 | 835 |
| 192 | S5:21:4 | 5 | 716.312 | 44.611 | 1273 | 218 |
| 231 | S5:22:4 | 6 | 730.328 | 45.414 | 177030 | 28728 |
| 232 | S5:22:4 | 6 | 730.328 | 45.928 | 32721 | 7624 |
| 233 | S5:22:4 | 6 | 730.328 | 46.299 | 1385 | 226 |
| 234 | S5:22:4 | 6 | 730.328 | 46.699 | 2823 | 395 |
| 235 | S5:22:4 | 6 | 730.328 | 46.934 | 12787 | 2575 |
| 236 | S5:22:4 | 6 | 730.328 | 47.228 | 3645 | 566 |
| 263 | S5:23:4 | 4 | 744.344 | 48.492 | 516258 | 85398 |
| 264 | S5:23:4 | 4 | 744.344 | 48.953 | 10415 | 2311 |
| 265 | S5:23:4 | 4 | 744.344 | 49.785 | 379584 | 82590 |
| 266 | S5:23:4 | 4 | 744.344 | 50.012 | 13056 | 1245 |
| 292 | S5:24:4 | 5 | 758.359 | 51.592 | 80150 | 13944 |
| 293 | S5:24:4 | 5 | 758.359 | 51.955 | 41327 | 8281 |
| 294 | S5:24:4 | 5 | 758.359 | 52.144 | 10803 | 1501 |
| 295 | S5:24:4 | 5 | 758.359 | 52.704 | 58779 | 11180 |
| 296 | S5:24:4 | 5 | 758.359 | 52.871 | 56335 | 8776 |
| 326 | S5:25:4 | 4 | 772.375 | 53.551 | 5425 | 885 |
| **327** | **S5:25:4(2,5,5,5,8ᴾ)** | **4** | **772.375** | **54.293** | **439869** | **77683** |
| 328 | S5:25:4 | 4 | 772.375 | 54.974 | 61766 | 17401 |
| 329 | S5:25:4 | 4 | 772.375 | 55.865 | 2445 | 438 |
| 356 | S5:26:4 | 6 | 786.391 | 56.637 | 7389 | 1084 |
| 357 | S5:26:4 | 6 | 786.391 | 57.234 | 233892 | 43948 |
| 358 | S5:26:4 | 6 | 786.391 | 57.605 | 175200 | 30339 |
| 359 | S5:26:4 | 6 | 786.391 | 57.930 | 3136 | 588 |
| 360 | S5:26:4 | 6 | 786.391 | 58.725 | 2224 | 470 |
| 361 | S5:26:4 | 6 | 786.391 | 58.997 | 2821 | 469 |
| 383 | S5:27:4 | 4 | 800.406 | 60.116 | 6104 | 1177 |
| 384 | S5:27:4 | 4 | 800.406 | 60.502 | 1840 | 551 |
| 385 | S5:27:4 | 4 | 800.406 | 61.485 | 5610 | 1337 |
| 386 | S5:27:4 | 4 | 800.406 | 61.682 | 2248 | 346 |
| 409 | S5:29:4 | 2 | 828.438 | 65.040 | 1503 | 292 |
| 410 | S5:29:4 | 2 | 828.438 | 66.689 | 528 | 113 |
| 417 | S5:30:4 | 1 | 842.453 | 69.381 | 434 | 72 |
| 228 | S5:22:5 | 2 | 728.312 | 43.402 | 1111 | 192 |
| 229 | S5:22:5 | 2 | 728.312 | 43.704 | 1131 | 185 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 261 | S5:23:5 | 1 | 742.328 | 47.312 | 11112 | 3176 |
| 289 | S5:24:5 | 1 | 756.344 | 49.498 | 5625 | 1086 |
| 323 | S5:25:5 | 1 | 770.359 | 51.827 | 58791 | 10975 |
| 242 | S5:23:8 | 1 | 736.281 | 40.920 | 444 | 83 |
| 301 | S5:25:8 | 1 | 764.312 | 45.995 | 13080 | 2449 |
| 331 | S5:26:8 | 3 | 778.328 | 48.884 | 5016 | 839 |
| 332 | S5:26:8 | 3 | 778.328 | 49.490 | 6146 | 1002 |
| 333 | S5:26:8 | 3 | 778.328 | 49.747 | 942 | 128 |
| 362 | S5:27:8 | 1 | 792.344 | 51.592 | 6472 | 1052 |
| 387 | S5:28:8 | 1 | 806.359 | 53.846 | 18840 | 3149 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| *Hexaacylsucroses* | | | | | | |
| 143 | S6:18:0 | 1 | 696.307 | 42.358 | 1664 | 272 |
| 212 | S6:20:0 | 4 | 724.339 | 47.917 | 3776 | 670 |
| 213 | S6:20:0 | 4 | 724.339 | 48.144 | 1247 | 184 |
| 214 | S6:20:0 | 4 | 724.339 | 48.938 | 26325 | 5056 |
| 215 | S6:20:0 | 4 | 724.339 | 49.309 | 2004 | 317 |
| 250 | S6:21:0 | 2 | 738.354 | 51.758 | 6158 | 1234 |
| 251 | S6:21:0 | 2 | 738.354 | 52.198 | 107161 | 19747 |
| 277 | S6:22:0 | 4 | 752.370 | 54.519 | 40410 | 10452 |
| 278 | S6:22:0 | 4 | 752.370 | 54.966 | 45185 | 10571 |
| 279 | S6:22:0 | 4 | 752.370 | 55.465 | 87876 | 16719 |
| 280 | S6:22:0 | 4 | 752.370 | 57.038 | 14385 | 3932 |
| 311 | S6:23:0 | 7 | 766.386 | 57.234 | 16777 | 3453 |
| 312 | S6:23:0 | 7 | 766.386 | 57.651 | 271418 | 69452 |
| 313 | S6:23:0 | 7 | 766.386 | 57.938 | 36258 | 4343 |
| 314 | S6:23:0 | 7 | 766.386 | 58.203 | 38668 | 7533 |
| 315 | S6:23:0 | 7 | 766.386 | 59.201 | 6439 | 1526 |
| 316 | S6:23:0 | 7 | 766.386 | 59.739 | 5688 | 1520 |
| 317 | S6:23:0 | 7 | 766.386 | 60.207 | 3750 | 742 |
| 340 | S6:24:0 | 5 | 780.401 | 59.859 | 67633 | 18369 |
| 341 | S6:24:0 | 5 | 780.401 | 60.396 | 213736 | 45630 |
| 342 | S6:24:0 | 5 | 780.401 | 60.880 | 182116 | 39975 |
| 343 | S6:24:0 | 5 | 780.401 | 62.324 | 61875 | 12977 |
| 344 | S6:24:0 | 5 | 780.401 | 62.809 | 5490 | 1165 |
| **370** | **S6:25:0(2,2,5,5,5,6)** | **4** | **794.417** | **62.990** | **1109307** | **233661** |
| 371 | S6:25:0 | 4 | 794.417 | 63.527 | 212097 | 43297 |
| 372 | S6:25:0 | 4 | 794.417 | 63.875 | 12480 | 2476 |
| 373 | S6:25:0 | 4 | 794.417 | 65.335 | 38955 | 7662 |
| 393 | S6:26:0 | 3 | 808.433 | 66.068 | 608240 | 126909 |
| 394 | S6:26:0 | 3 | 808.433 | 66.416 | 25085 | 6380 |
| 395 | S6:26:0 | 3 | 808.433 | 68.255 | 2640 | 488 |
| 404 | S6:27:0 | 1 | 822.448 | 68.928 | 54212 | 11668 |
| 413 | S6:28:0 | 2 | 836.464 | 70.576 | 520 | 90 |
| 414 | S6:28:0 | 2 | 836.464 | 71.688 | 2224 | 438 |
| 419 | S6:29:0 | 1 | 850.480 | 73.450 | 921 | 168 |
| 201 | S6:20:1 | 1 | 722.323 | 46.261 | 14991 | 2597 |
| 245 | S6:21:1 | 1 | 736.339 | 49.724 | 4238 | 677 |
| 271 | S6:22:1 | 2 | 750.354 | 51.547 | 31806 | 5434 |
| 272 | S6:22:1 | 2 | 750.354 | 51.880 | 4027 | 978 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 302 | S6:23:1 | 4 | 764.370 | 54.270 | 6205 | 1081 |
| 303 | S6:23:1 | 4 | 764.370 | 54.678 | 6907 | 1443 |
| 304 | S6:23:1 | 4 | 764.370 | 55.185 | 9032 | 1813 |
| 305 | S6:23:1 | 4 | 764.370 | 56.282 | 1871 | 333 |
| 334 | S6:24:1 | 4 | 778.386 | 56.924 | 23288 | 4431 |
| 335 | S6:24:1 | 4 | 778.386 | 57.356 | 2716 | 836 |
| 336 | S6:24:1 | 4 | 778.386 | 57.598 | 1876 | 308 |
| 337 | S6:24:1 | 4 | 778.386 | 57.855 | 2291 | 386 |
| 363 | S6:25:1 | 5 | 792.401 | 59.980 | 10714 | 2454 |
| 364 | S6:25:1 | 5 | 792.401 | 60.449 | 9101 | 1957 |
| 365 | S6:25:1 | 5 | 792.401 | 60.691 | 768 | 109 |
| 366 | S6:25:1 | 5 | 792.401 | 60.911 | 746 | 122 |
| 367 | S6:25:1 | 5 | 792.401 | 61.576 | 1390 | 232 |
| 388 | S6:26:1 | 5 | 806.417 | 62.831 | 3577 | 709 |
| 389 | S6:26:1 | 5 | 806.417 | 63.104 | 951 | 160 |
| 390 | S6:26:1 | 5 | 806.417 | 63.422 | 373 | 75 |
| 391 | S6:26:1 | 5 | 806.417 | 64.087 | 485 | 92 |
| 392 | S6:26:1 | 5 | 806.417 | 64.586 | 1819 | 347 |
| 403 | S6:27:1 | 1 | 820.433 | 65.607 | 628 | 107 |
| 197 | S6:20:2 | 1 | 720.307 | 44.362 | 2415 | 397 |
| 290 | S6:23:4 | 2 | 758.323 | 48.326 | 15241 | 2504 |
| 291 | S6:23:4 | 2 | 758.323 | 48.635 | 1308 | 199 |
| 324 | S6:24:4 | 2 | 772.339 | 51.010 | 3317 | 582 |
| 325 | S6:24:4 | 2 | 772.339 | 51.471 | 6050 | 1083 |
| 352 | S6:25:4 | 4 | 786.354 | 53.150 | 8527 | 1542 |
| 353 | S6:25:4 | 4 | 786.354 | 53.605 | 26904 | 4304 |
| 354 | S6:25:4 | 4 | 786.354 | 54.141 | 1616 | 286 |
| 355 | S6:25:4 | 4 | 786.354 | 54.481 | 1585 | 314 |
| 377 | S6:26:4 | 6 | 800.370 | 55.699 | 9964 | 1906 |
| 378 | S6:26:4 | 6 | 800.370 | 56.183 | 5249 | 1231 |
| 379 | S6:26:4 | 6 | 800.370 | 56.372 | 1347 | 154 |
| 380 | S6:26:4 | 6 | 800.370 | 56.675 | 4658 | 779 |
| 381 | S6:26:4 | 6 | 800.370 | 57.129 | 4649 | 806 |
| 382 | S6:26:4 | 6 | 800.370 | 57.590 | 4217 | 732 |
| 397 | S6:27:4 | 6 | 814.386 | 58.165 | 61542 | 10899 |
| 398 | S6:27:4 | 6 | 814.386 | 58.483 | 4300 | 703 |
| 399 | S6:27:4 | 6 | 814.386 | 58.710 | 9016 | 1654 |
| 400 | S6:27:4 | 6 | 814.386 | 59.594 | 3888 | 719 |
| 401 | S6:27:4 | 6 | 814.386 | 60.525 | 3738 | 655 |

**Table 2.4.** (continued)

| Elution Order | Acylsucrose ID | # of Isomers | [M+NH$_4$]$^+$ $m/z$ XIC (±20 ppm) | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|---|---|---|
| 402 | S6:27:4 | 6 | 814.386 | 61.319 | 851 | 154 |
| 407 | S6:28:4 | 2 | 828.401 | 61.175 | 35416 | 6602 |
| 408 | S6:28:4 | 2 | 828.401 | 62.007 | 8272 | 1553 |
| 415 | S6:29:4 | 2 | 842.417 | 63.928 | 381 | 82 |
| 416 | S6:29:4 | 2 | 842.417 | 65.441 | 1370 | 330 |
| 349 | S6:25:5 | 3 | 784.339 | 50.783 | 1059 | 196 |
| 350 | S6:25:5 | 3 | 784.339 | 51.230 | 1101 | 207 |
| 351 | S6:25:5 | 3 | 784.339 | 52.295 | 638 | 110 |
| *Heptaacylsucroses* | | | | | | |
| 411 | S7:27:0 | 2 | 836.427 | 68.806 | 2230 | 436 |
| 412 | S7:27:0 | 2 | 836.427 | 69.276 | 389 | 79 |
| 418 | S7:28:0 | 1 | 850.443 | 71.756 | 1138 | 254 |

**Table 2.5.** LC/MS/MS metadata

| | |
|---|---|
| Facility Director | Dr. A. Daniel Jones |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Mass Spectrometry and Metabolomics Core |
| Facility Instrument Title | LCMS: G2-XS QTof #2 |
| LC System | Acquity UPLC I-Class Binary Solvent Manager equipped with Acquity Column Manager |
| Manufacturer | Waters |
| Autosampler | 2777C Sample Manager |
| Column | Ascentis Express C18 Analytical HPLC, 10 cm x 2.1mm x 2.7 μm |
| Column Manufacturer | Supelco |
| Catalogue Number | 53823-U |
| Serial Number | USRB002977 |
| Packing Lot Number | S11020 |
| Injection Volume | 10 μL |
| Flow Rate | 0.4 mL/min |
| Mobile Phases: | |
| A | 10 mM ammonium formate in water (pH 2.8, adjusted with formic acid) |
| B | Acetonitrile |
| Gradient Profile | Hold 1% B at 0-1 min, linear 1-40% B at 1-2 min, linear 40%-60% B at 2-27 min, linear 60-100% B at 27-28 min, hold 100% B at 28-32 min, linear 100-1% B at 32-32.01 min, and hold 1% B at 32.01-35 min |
| Column Oven Temperature | 50 ºC |
| Sample Temperature in autosampler | 10 °C |
| Inlet Method Name | SMH_C18_Column1_35min_40to60 |
| Mass Spectrometer | Xevo G2-XS QTof |
| Manufacturer | Waters |
| Software | MassLynx v4.1 |
| Ionization Source | Electrospray Ionization (ESI) |
| Data Acquisition | Sensitivity Mode (low mass and high mass resolving quadrupole setting at 15) |
| Polarity | Positive, Negative |
| Mass Range | $m/z$ 50-1000 |
| Data Format | Centroid |
| Capillary Voltage | 3 kV (ESI+), 2 kV (ESI-) |
| Sample Cone | 35 V (ESI+), 40 V (ESI-) |
| Source Temperature | 100 °C |
| Source Offset | 80 V |

**Table 2.5.** (continued)

| | |
|---|---|
| Desolvation Temperature | 350 °C |
| Cone Gas Flow | 25 L/hr |
| Desolvation Gas Flow | 350 L/hr |
| LM Resolution | 15 |
| HM Resolution | 15 |
| Collision Potential Ramp | |
| Start Potential | 5 V |
| End Potential | 60 V |
| Scan Duration | 0.5 s |
| Inter Scan Delay | 0.014 s |
| Collision Cell Pressure | 0.2 mbar |
| MS Method Files | SMH_MSMS_pos_35 min_Mass A, SMH_MSMS_neg_35 min_Mass A |
| Sample handling | Each acylsucrose NMR sample was concentrated to dryness under $N_2$ gas in autosampler vials, reconstituted in 0.50 mL AcN:IPA and diluted ~100x by removing a 5.0 µL aliquot and adding it to 0.50 mL of AcN:IPA solution. |
| Sample Storage | -20 °C in autosampler vials with screw PTFE lined caps |
| Protocol when analyzing the samples | The instrument was calibrated in ESI+/- modes using 500 µM sodium formate solution. First, samples were analyzed in ESI- mode. After column equilibration, a blank sample (AcN:IPA) and a *S. sinuata* bulk extract sample were analyzed in CID mode, followed by MS/MS analysis of the first eight acylsucrose samples in ascending elution order. Again, the blank and a *S. sinuata* bulk extract samples were analyzed in CID mode, followed by MS/MS analysis of the remaining eight acylsucrose samples in ascending elution order. Finally, the blank and *S. sinuata* bulk extract samples were analyzed in CID mode. Subsequently, the sequence was repeated in ESI+ mode. MS/MS product ion spectra were generated using a collision potential ramp, targeting the respective precursor pseudomolecular ion [M+HCOO]$^-$ and [M+NH$_4$]$^+$ for the entirety of the chromatographic runtime (35 mins). |
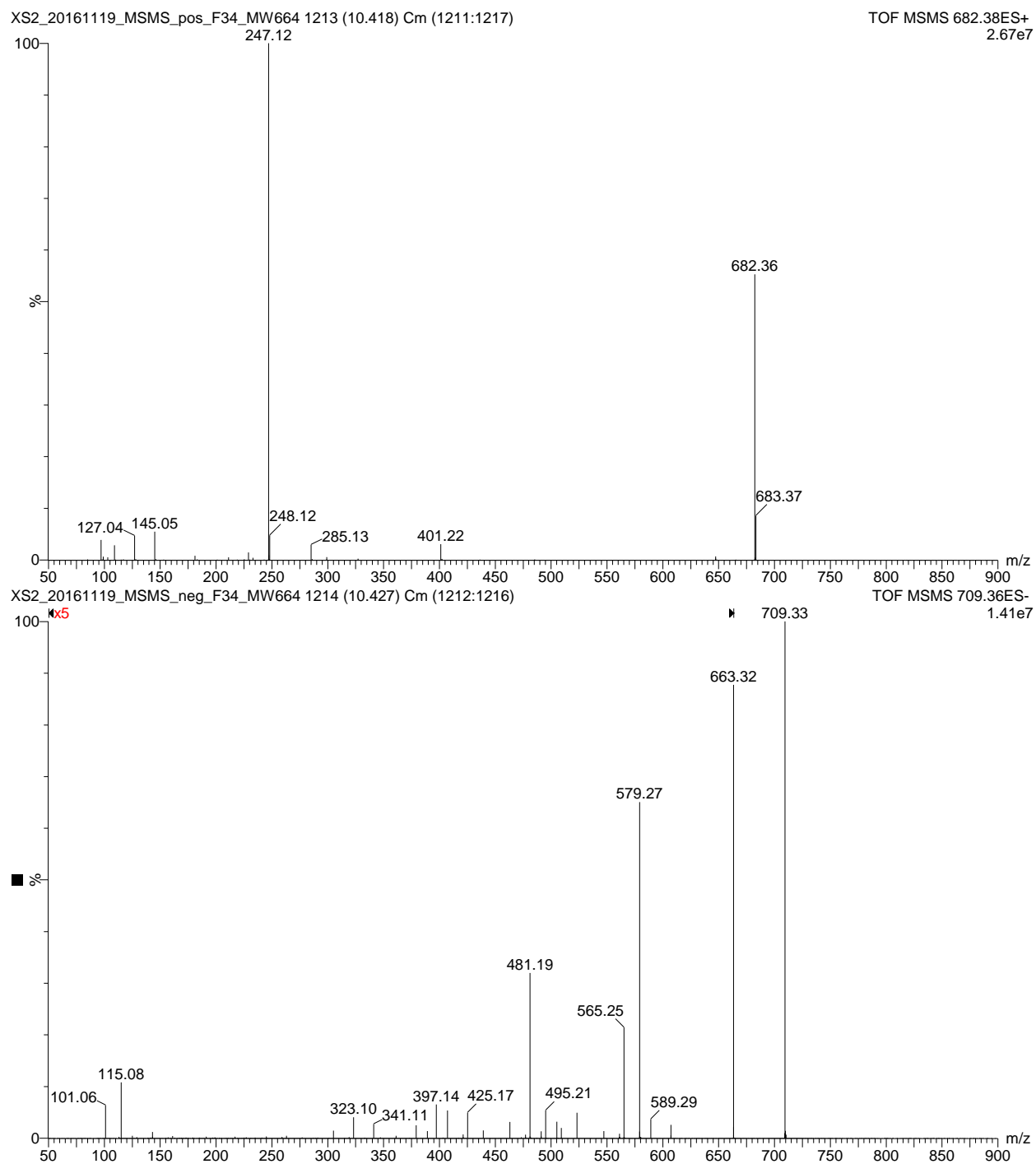
**Figure 2.8.** S4:19:0(3,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 5x over m/z 51-664)
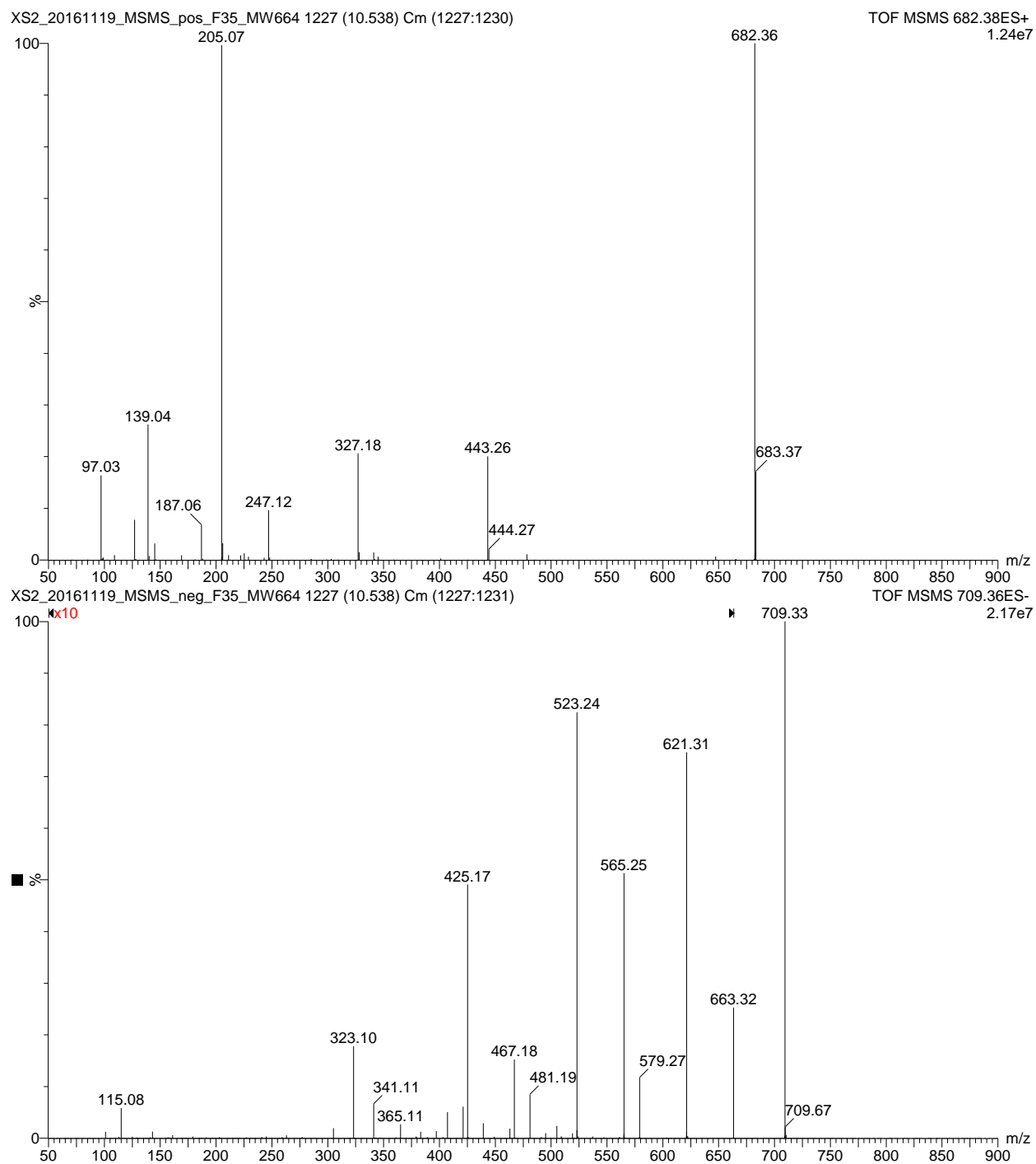
**Figure 2.9.** S4:19:0(2,5,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 10x over *m/z* 51-664)
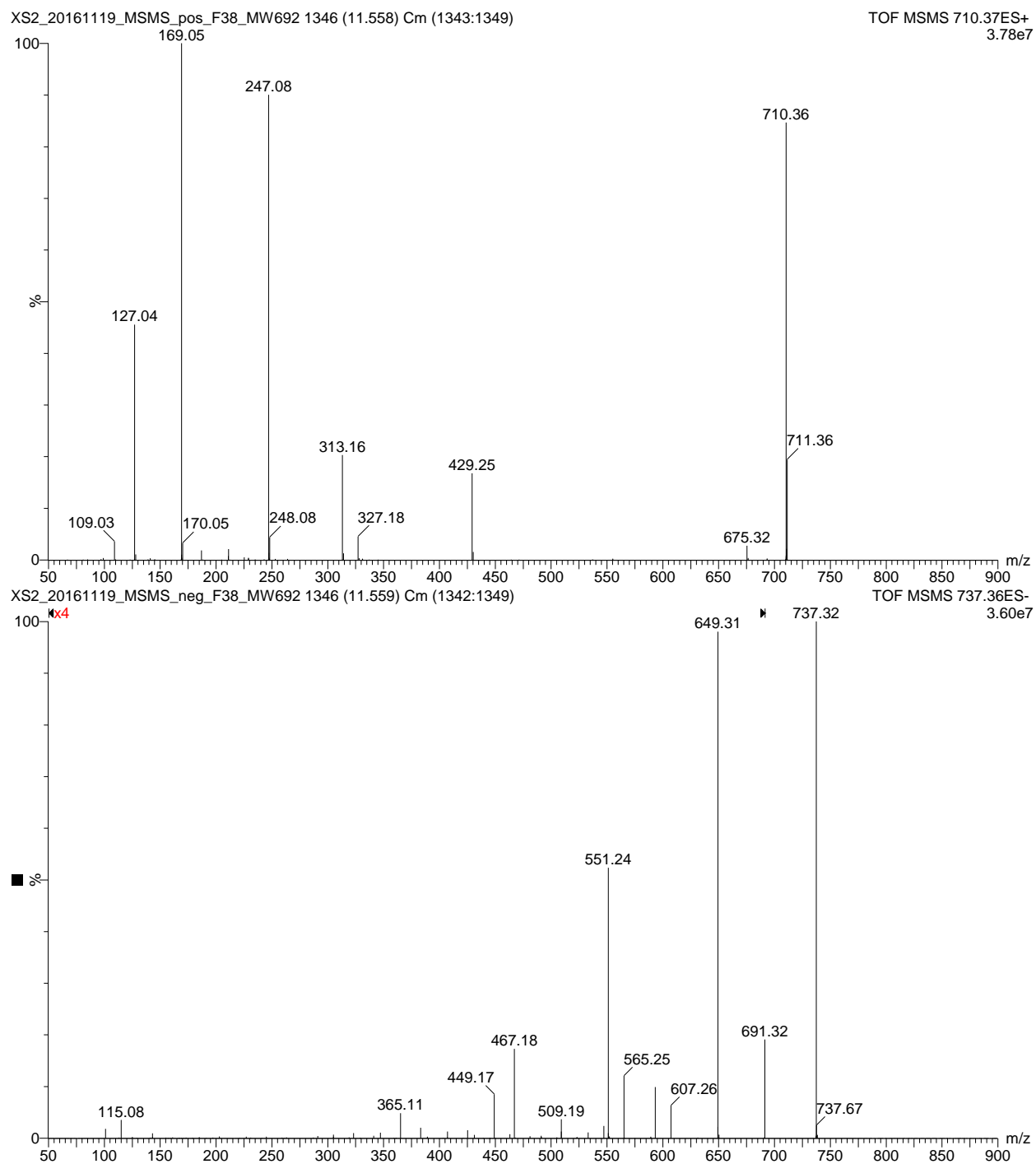
**Figure 2.10.** S5:20:0(2,2,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 4x over *m/z* 51-692)
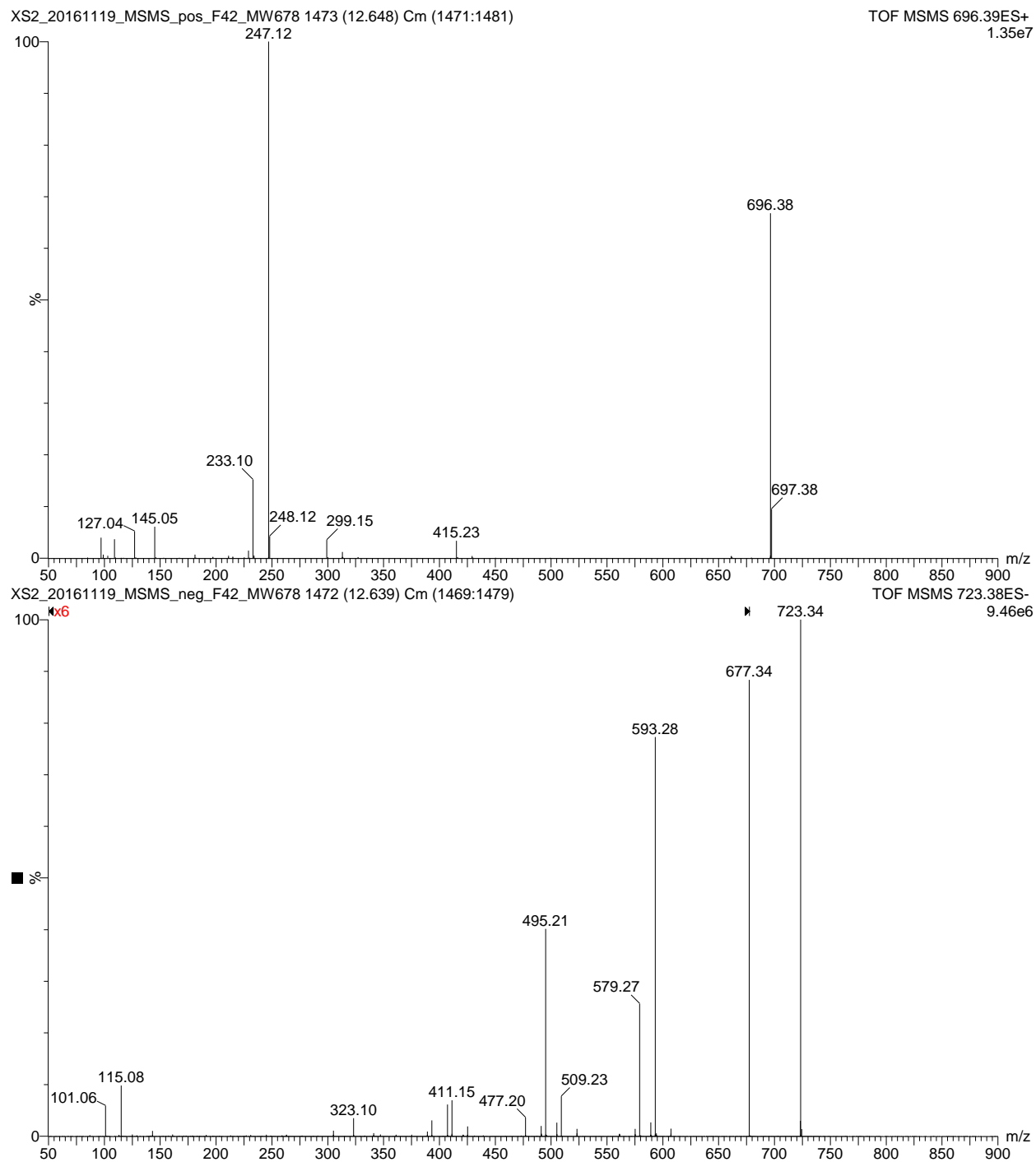
**Figure 2.11.** S4:20:0(4,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 6x over *m/z* 51-678)
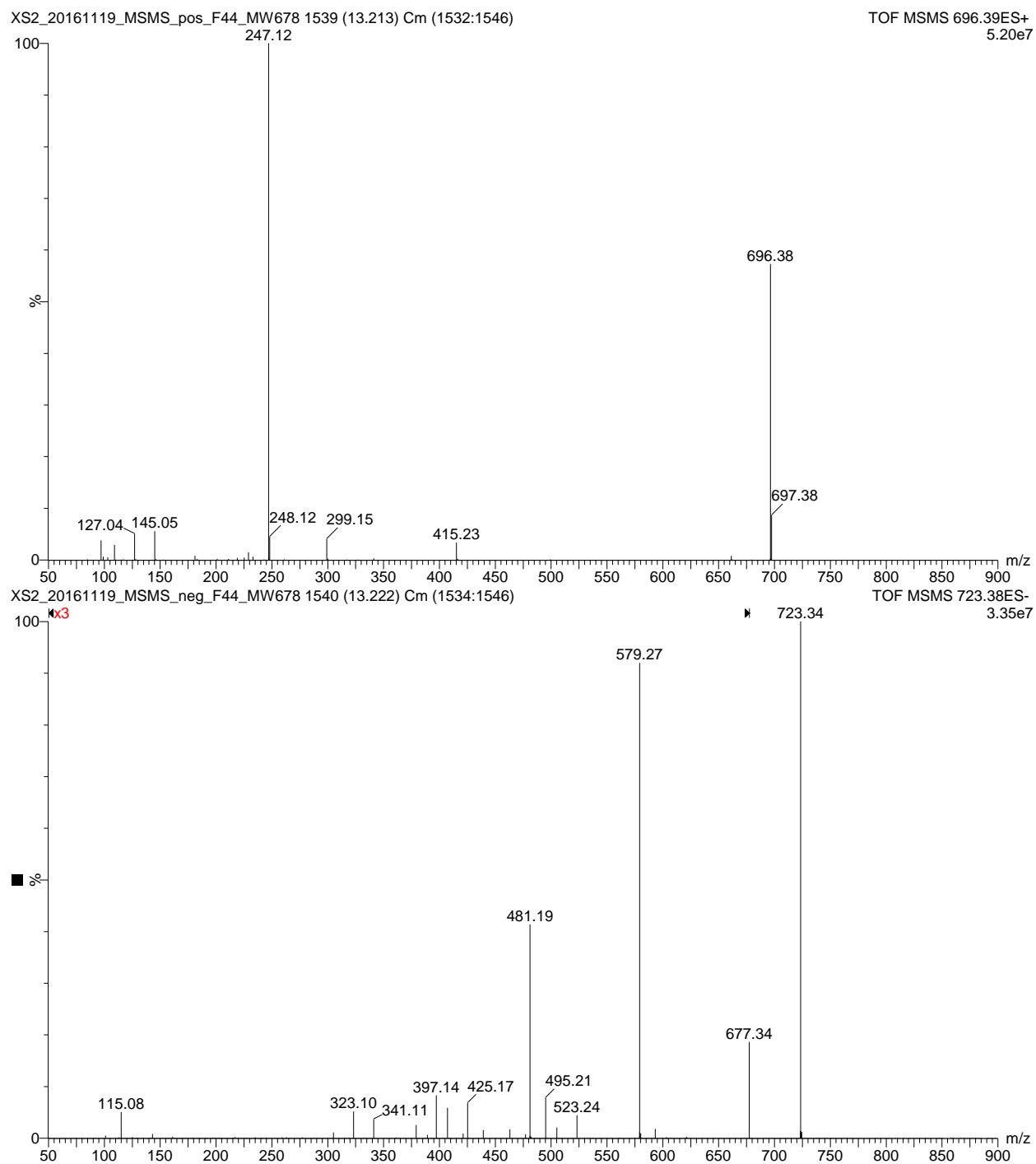
**Figure 2.12.** S4:20:0(3,5,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 3x over *m/z* 51-678)
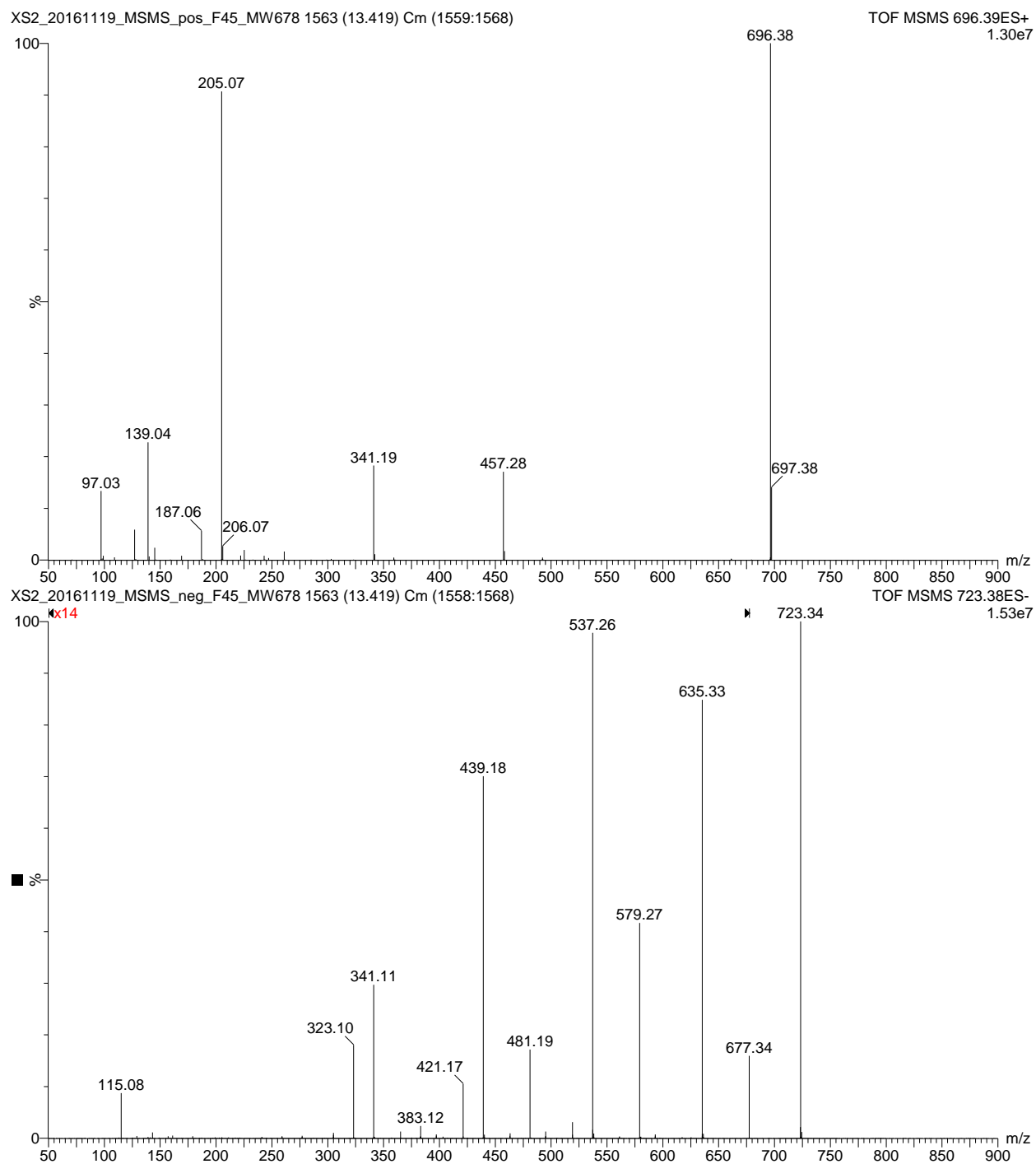
**Figure 2.13.** S4:20:0(2,6,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 14x over *m/z* 51-678)
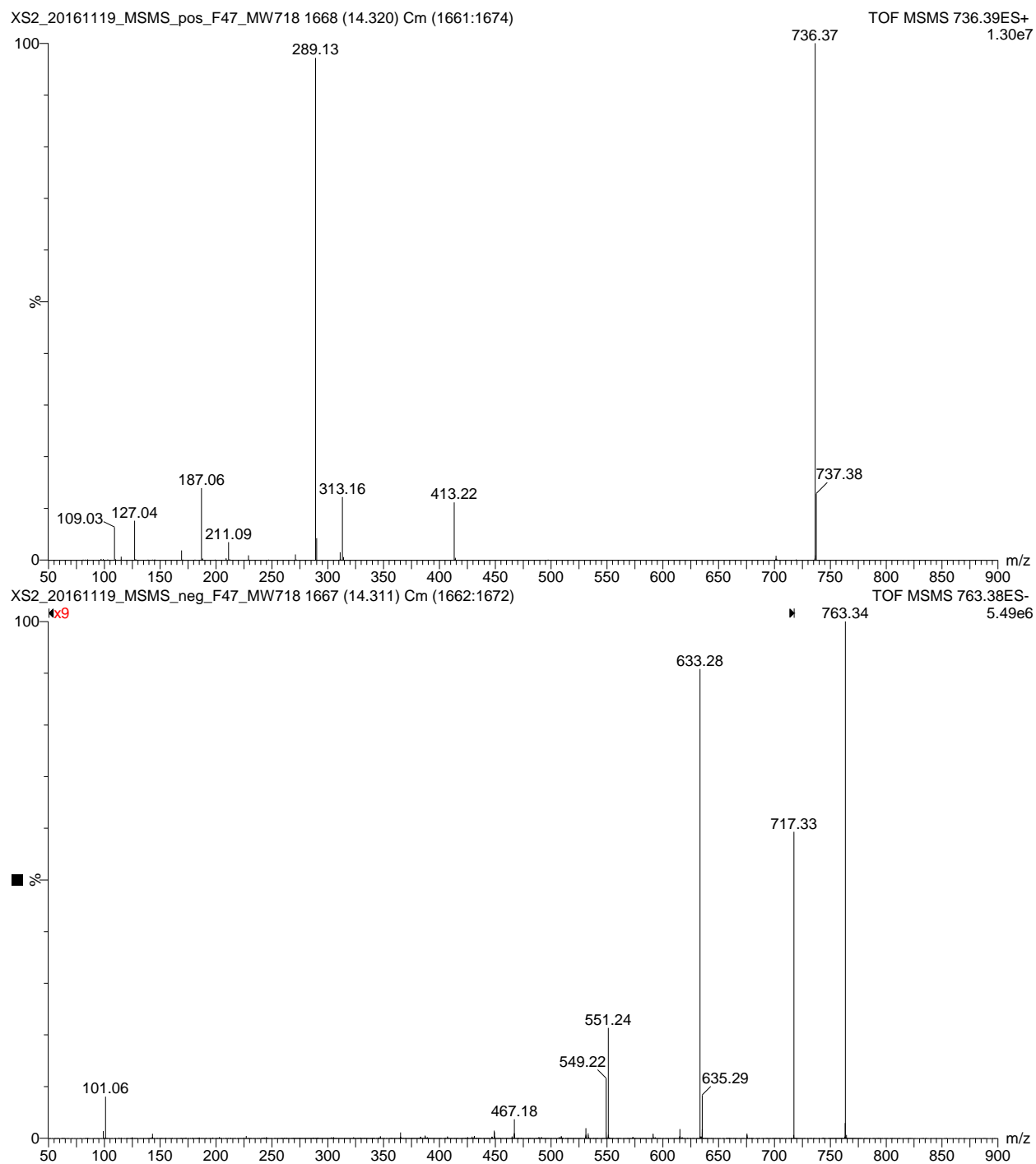
**Figure 2.14.** S5:22:1(2,5,5,5,5$^T$) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 9x over $m/z$ 51-718)
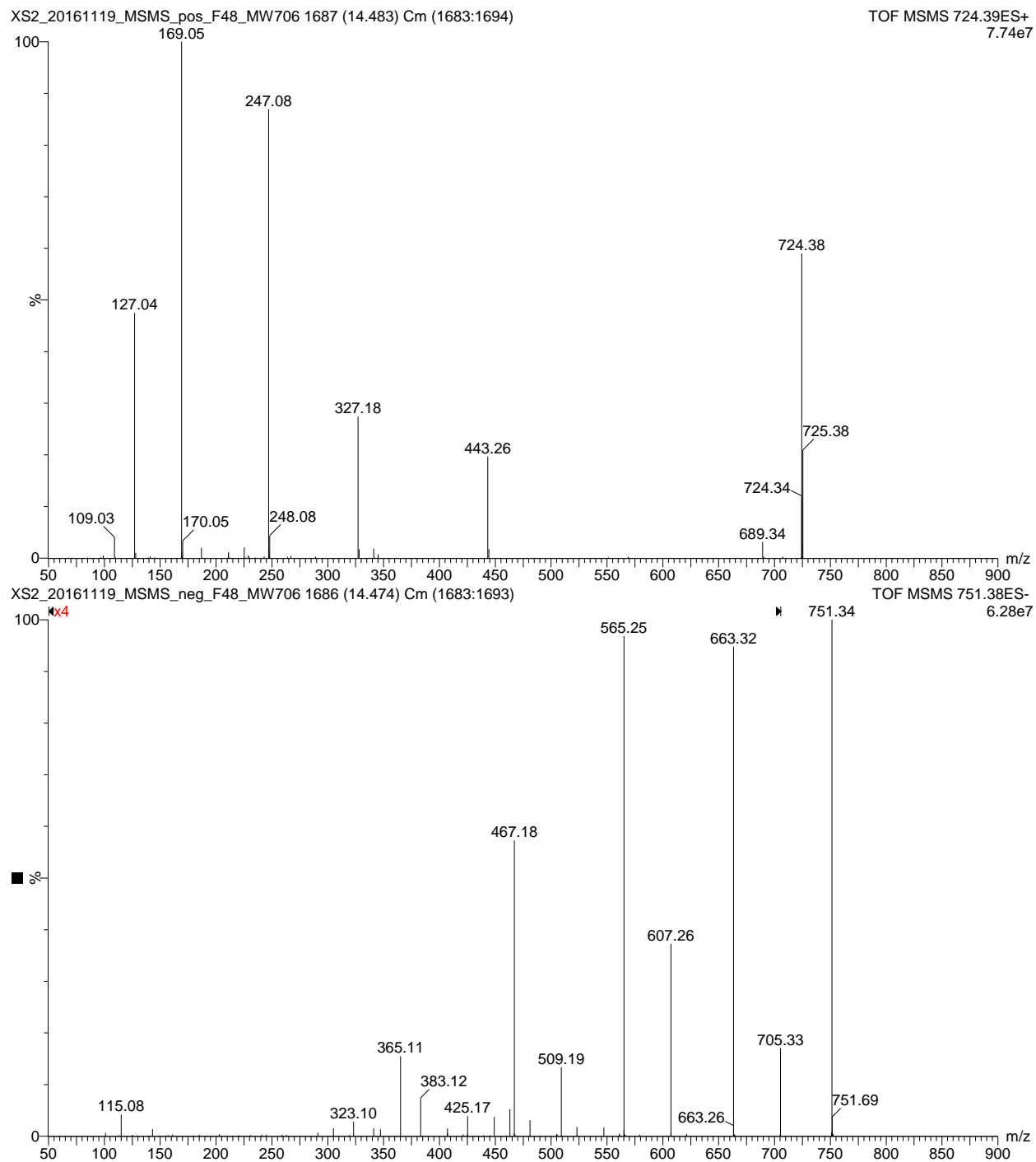
**Figure 2.15.** S5:21:0(2,2,5,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 4x over *m/z* 51-706)
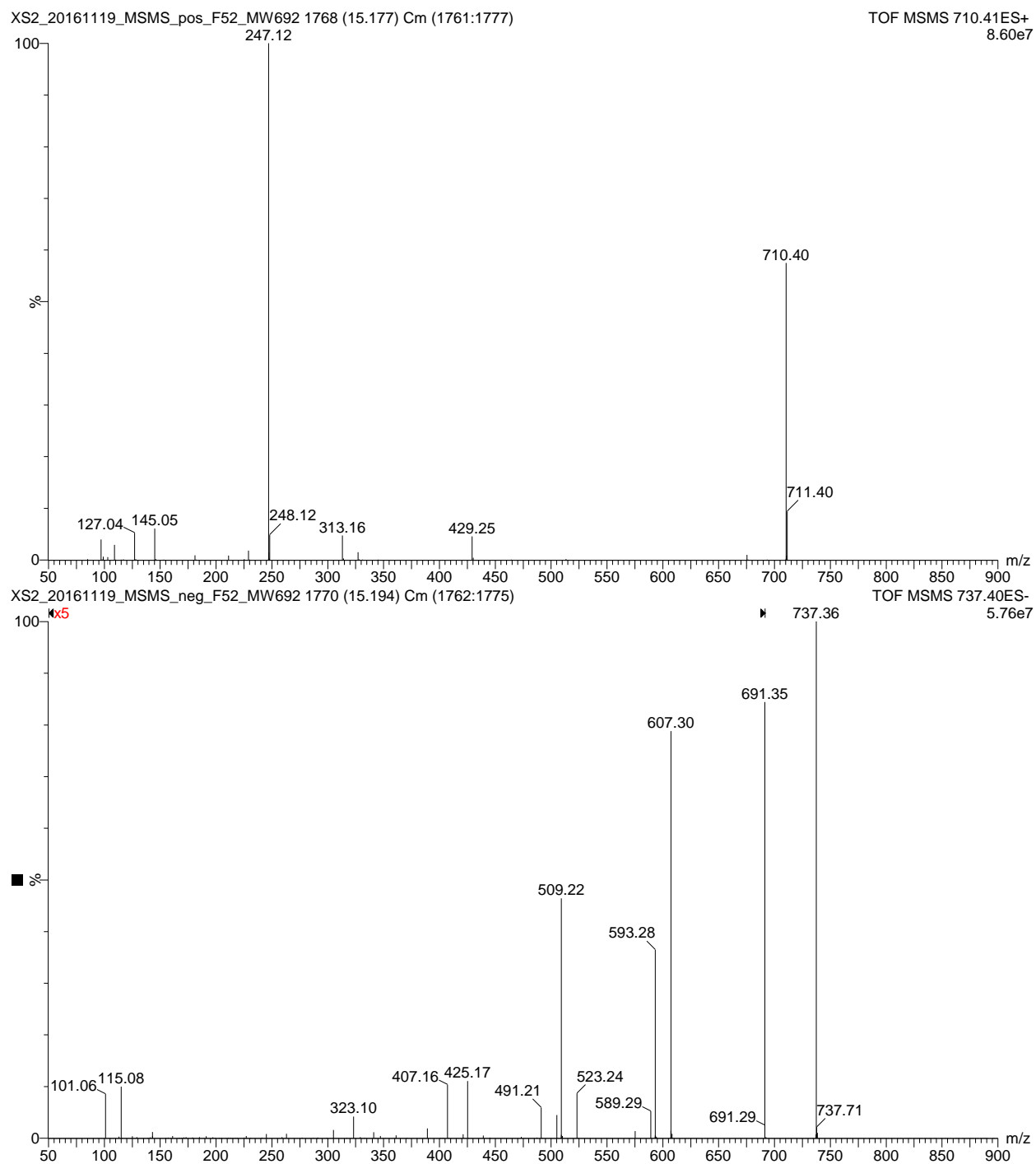
**Figure 2.16.** S4:21:0(5,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 5x over *m/z* 51-692)
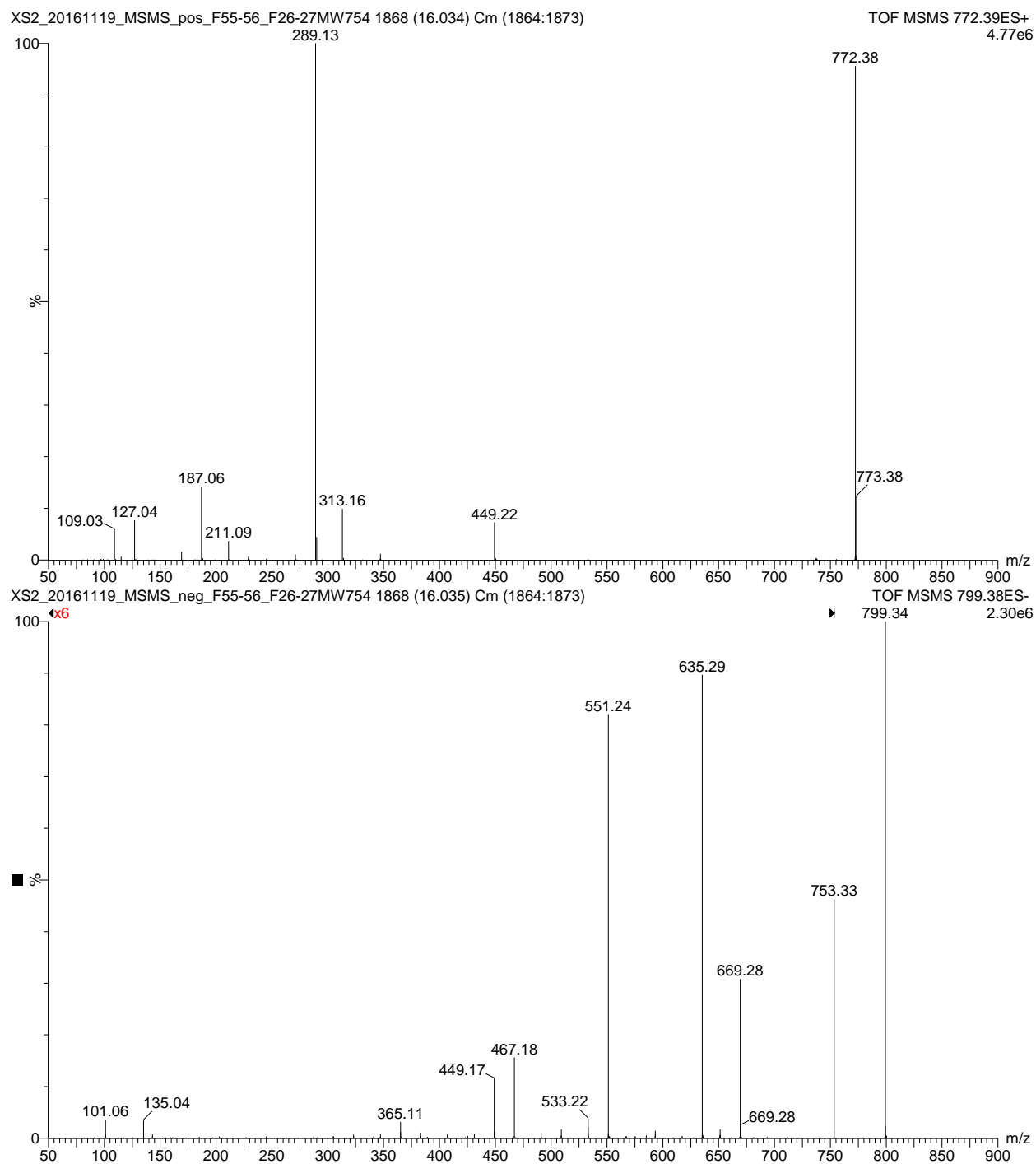
**Figure 2.17.** S5:25:4(2,5,5,5,8$^P$) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 6x over $m/z$ 51-754)
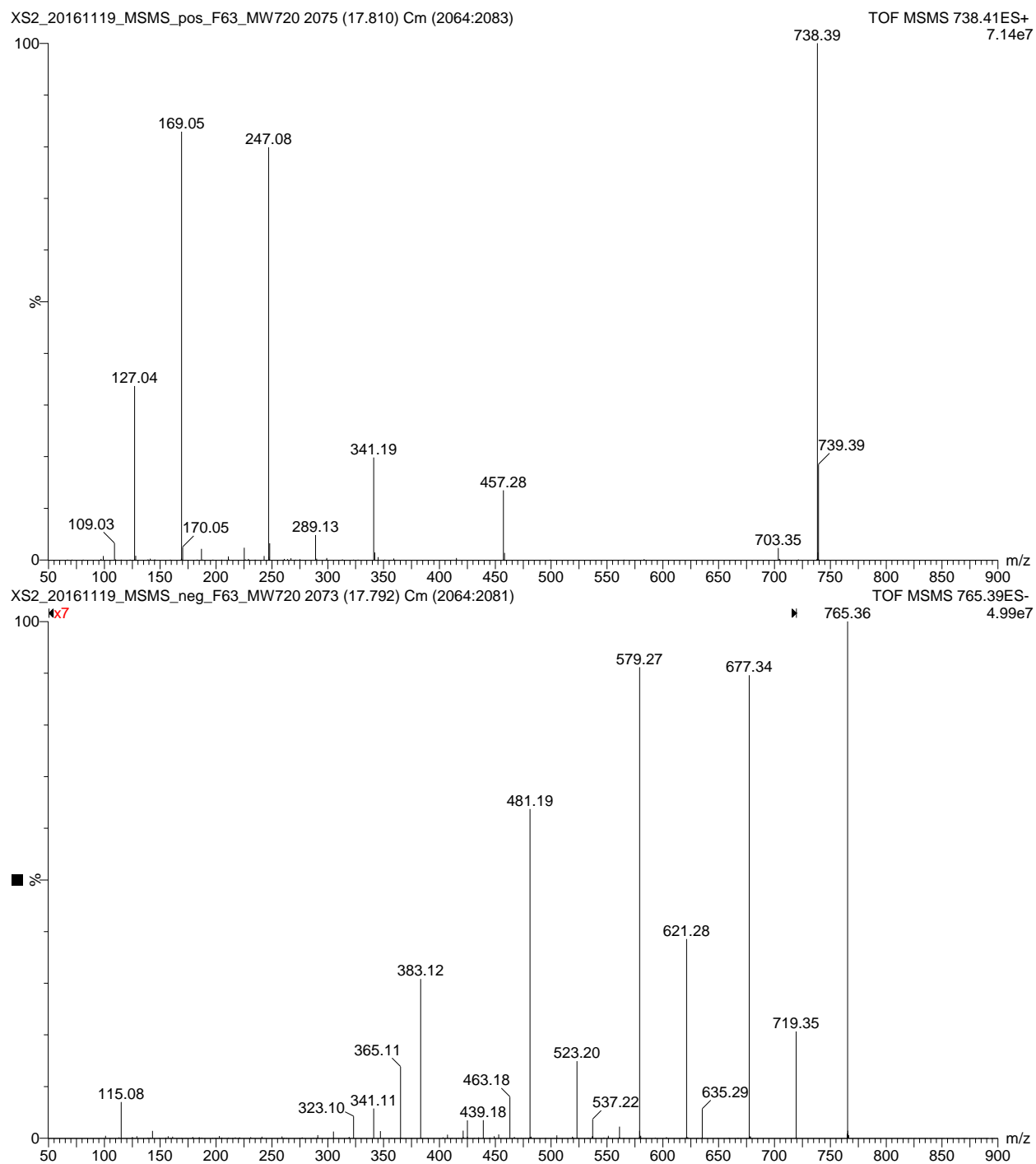
**Figure 2.18.** S5:22:0(2,2,6,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 7x over *m/z* 51-720)
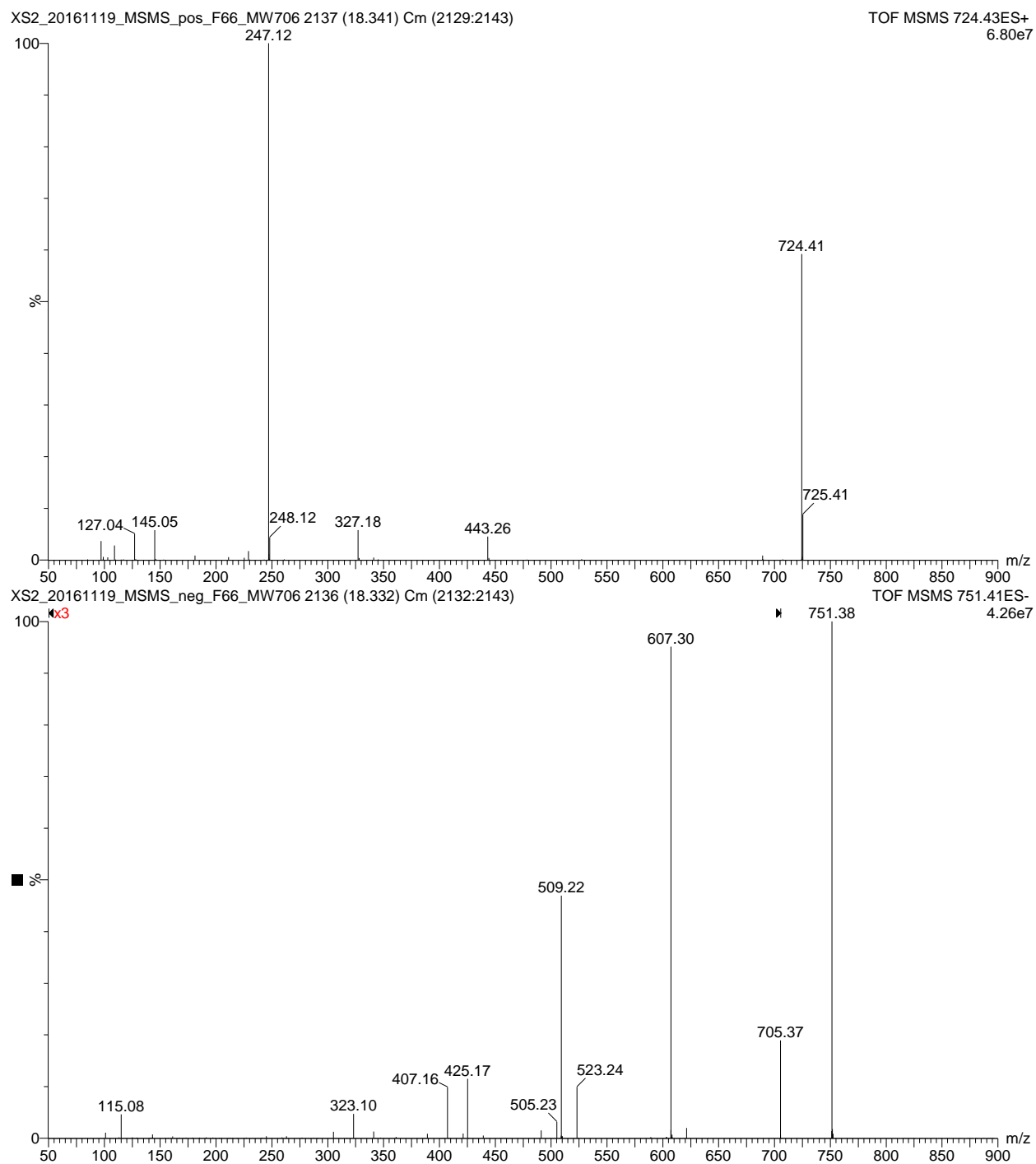
**Figure 2.19.** S4:22:0(5,5,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 3x over *m/z* 51-706)

82
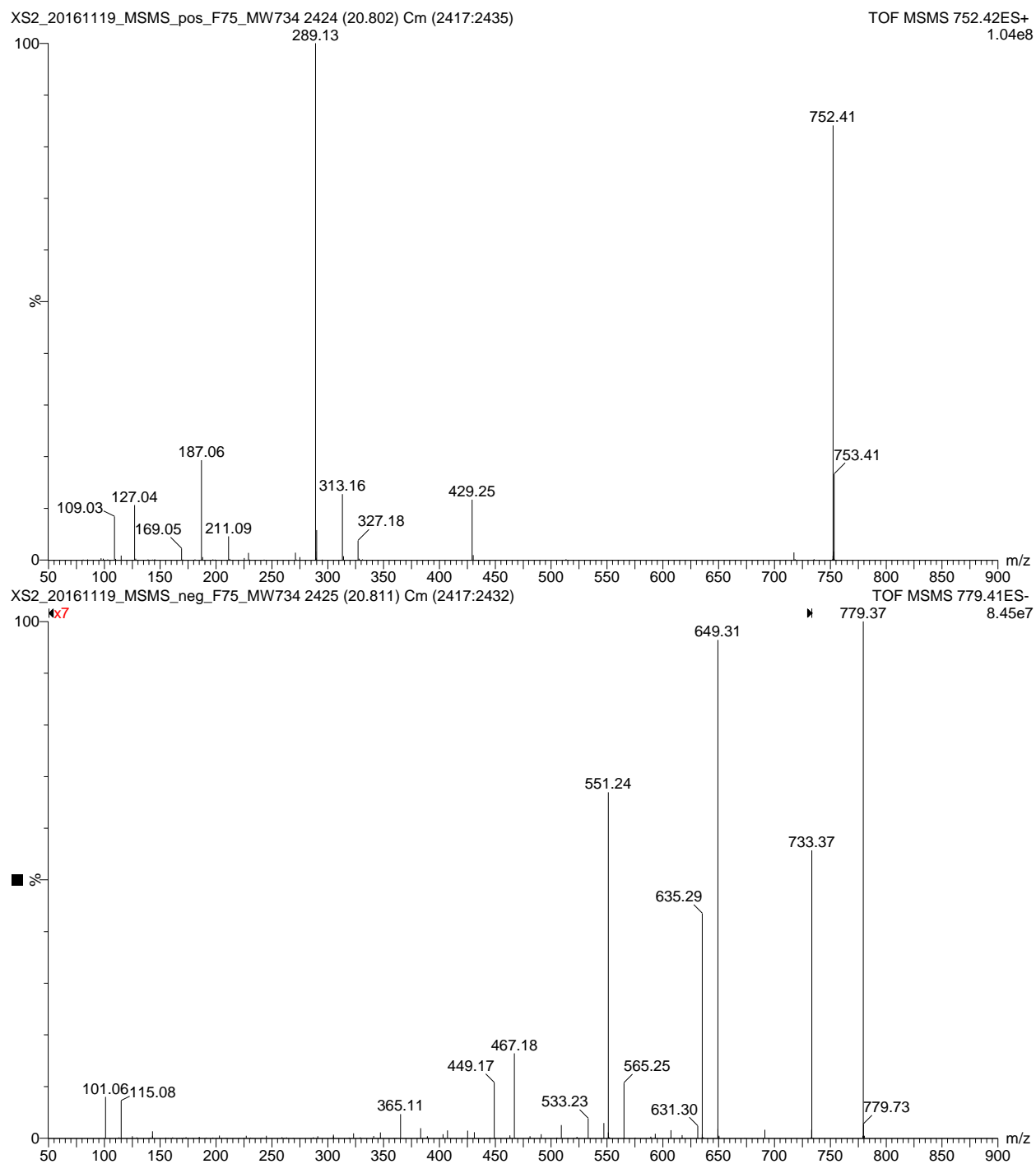
**Figure 2.20.** S5:23:0(2,5,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 7x over *m/z* 51-734)

83
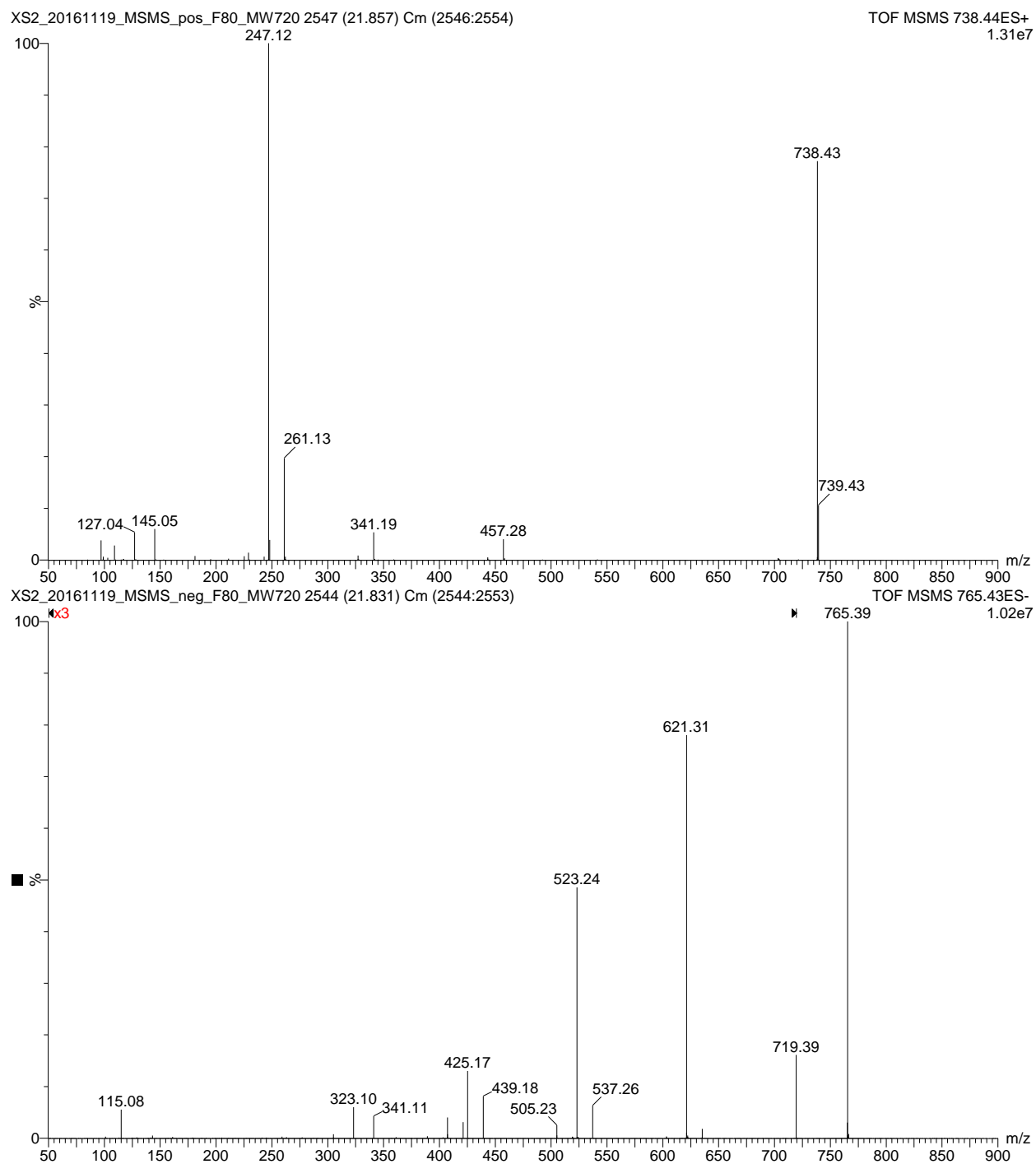
**Figure 2.21.** S4:23:0(5,6,6,6) MS/MS product ion spectra, precursor ion [M+NH₄]⁺ (top) & [M+HCOO]⁻ (bottom, magnified 3x over *m/z* 51-720)
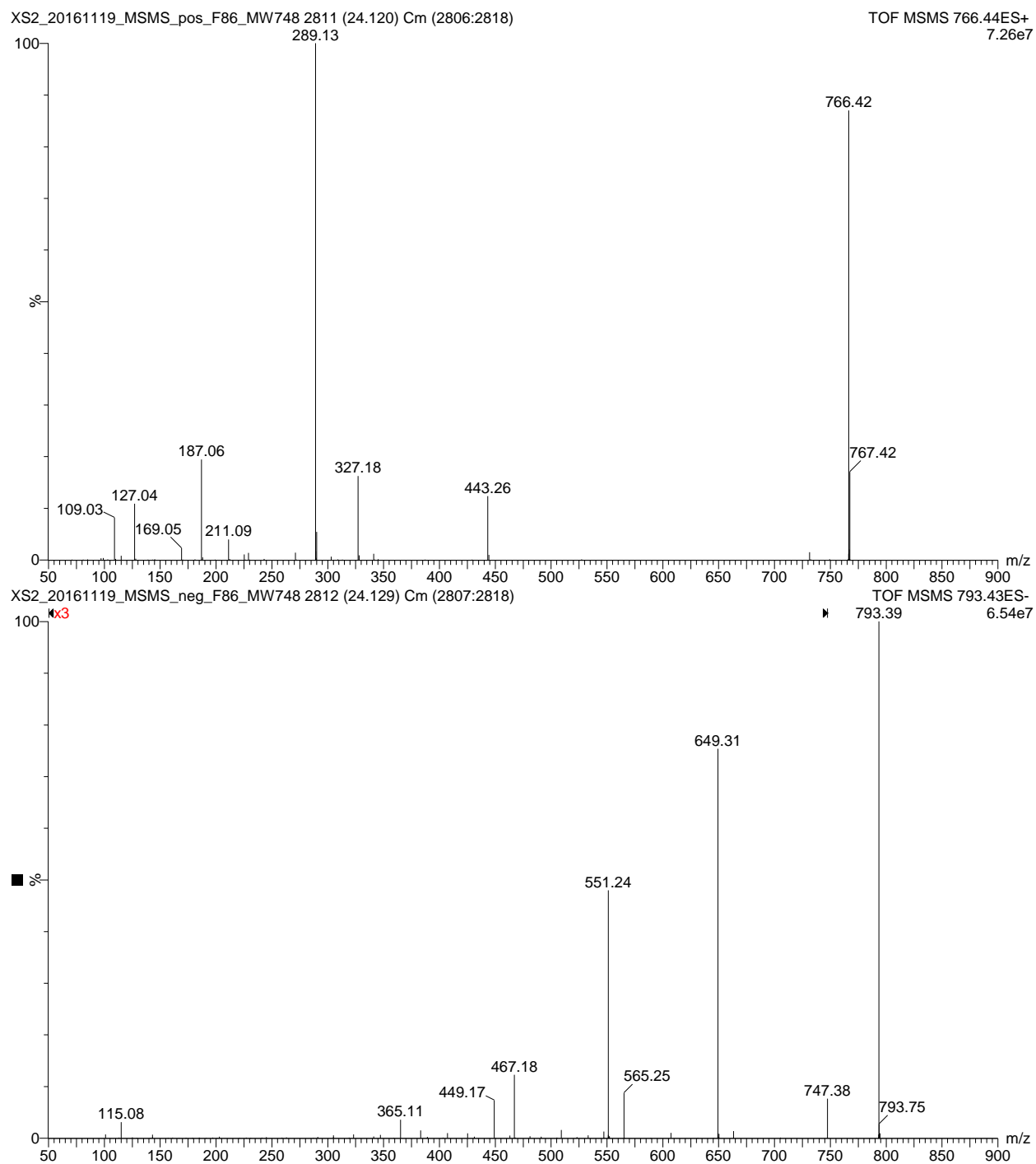
**Figure 2.22.** S5:24:0(2,5,5,6,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 3x over *m/z* 51-748)

85
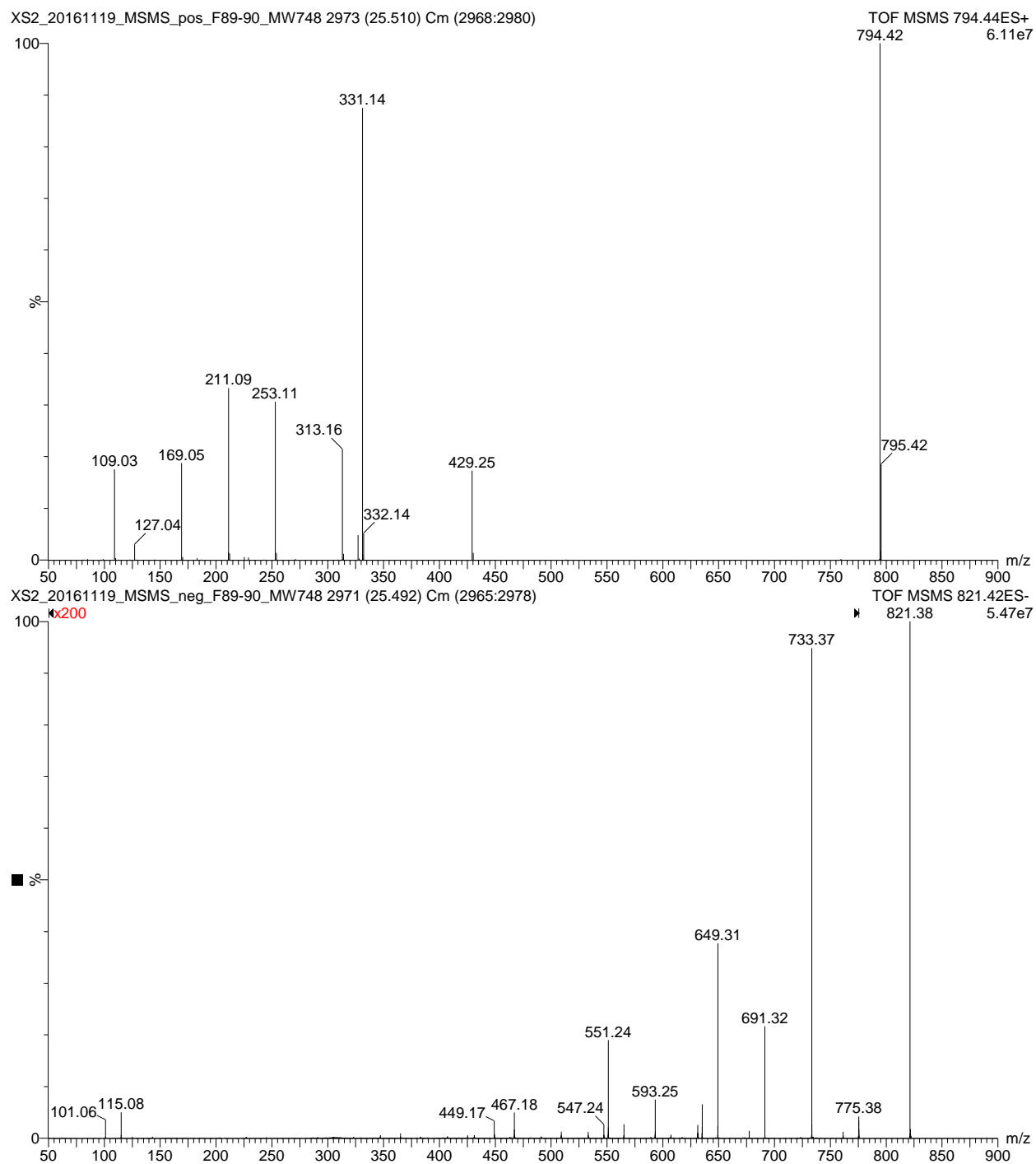
**Figure 2.23.** S6:25:0(2,2,5,5,5,6) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+HCOO]$^-$ (bottom, magnified 200x over *m/z* 51-776). Note; there is a graph header mistake. This sample was accidently named "MW748", but should have been named "MW 776".

**Table 2.6.** GC/MS metadata

| | |
|---|---|
| Facility Director | Dr. A. Daniel Jones |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Mass Spectrometry and Metabolomics Core |
| Facility Instrument Title | GCMS: Waters GCT |
| GC System | Agilent Technologies (model 6890N) |
| Manufacturer | Waters |
| Injector | Agilent Technologies (model V7683B) |
| Column | VF-5ms, 30m x 250µm ID, 0.25µm film thickness (+10m EZ-Guard) |
| Column Manufacturer | Agilent Technologies |
| Part Number | CP9013 |
| Serial Number | 90310154380 |
| Injection Volume | 1.0 µL (splitless mode) |
| Flow Rate | 1.0 mL/min (constant flow) |
| Carrier Gas | Helium |
| Gradient Profile | Initial 36 °C, linear 4 °C/min to 220°C, linear 20 °C/min to 320 °C, and hold at 320 °C for 4 mins |
| Solvent Delay | 2.45-3.35 mins |
| Inlet Temperature | 280 ℃ |
| Purge Time | 0.75 min |
| Purge Pressure | 30.0 kPa |
| Inlet Method Name | SMH_splitless_56min_36-320C |
| Mass Spectrometer | GCT Premier CAB067 |
| Manufacturer | Waters |
| Software | MassLynx v4.1 |
| Ionization Mode | Electron Ionization |
| Mass Range | $m/z$ 50-600 |
| Data Format | Centroid |
| Electron Energy | 70.0 eV |
| Source Temperature | 200 °C |
| Scan Duration | 0.2 s |
| Inter Scan Delay | 0.05 s |
| MS Method Files | SMH_56min |
| Retention Index Series | Supelco C7-C30 saturated alkane standard series diluted to 40 µg/mL in n-hexane |
| Lot Number | XA17133V |
| Sigma Aldrich Product Number | 49451-U |

**Table 2.6.** (continued)

| | |
|---|---|
| Sample handling | For transesterification, 10 mL of extract solution was concentrated to dryness by purging with N2 gas. 300 µL of 21% (v/v) sodium ethoxide solution was added. The reaction proceeded for 40 mins with horizontal mixing at 140 rpm. 400 µL of n-hexane was added, followed by vortex mixing. The hexane layer was removed for analysis by performing three washes with 500 µL of sodium chloride solution. A 50 µL aliquot of the hexane solution was transferred to an autosampler vial with limited volume insert. |
| Sample Storage | -20 °C in autosampler vials with screw PTFE lined caps |
| Protocol when analyzing the samples | The instrument was calibrated in EI+ mode using perfluorotributylamine. First, a hexane blank was analyzed, followed by a process blank (10 mL AcN:IPA concentrated to dryness and taken through the sodium ethoxide reaction and hexane extraction process). Then, the transesterified *S. sinuata* sample was analyzed, followed by the 40 ppm C7-C30 saturated alkanes standard. |
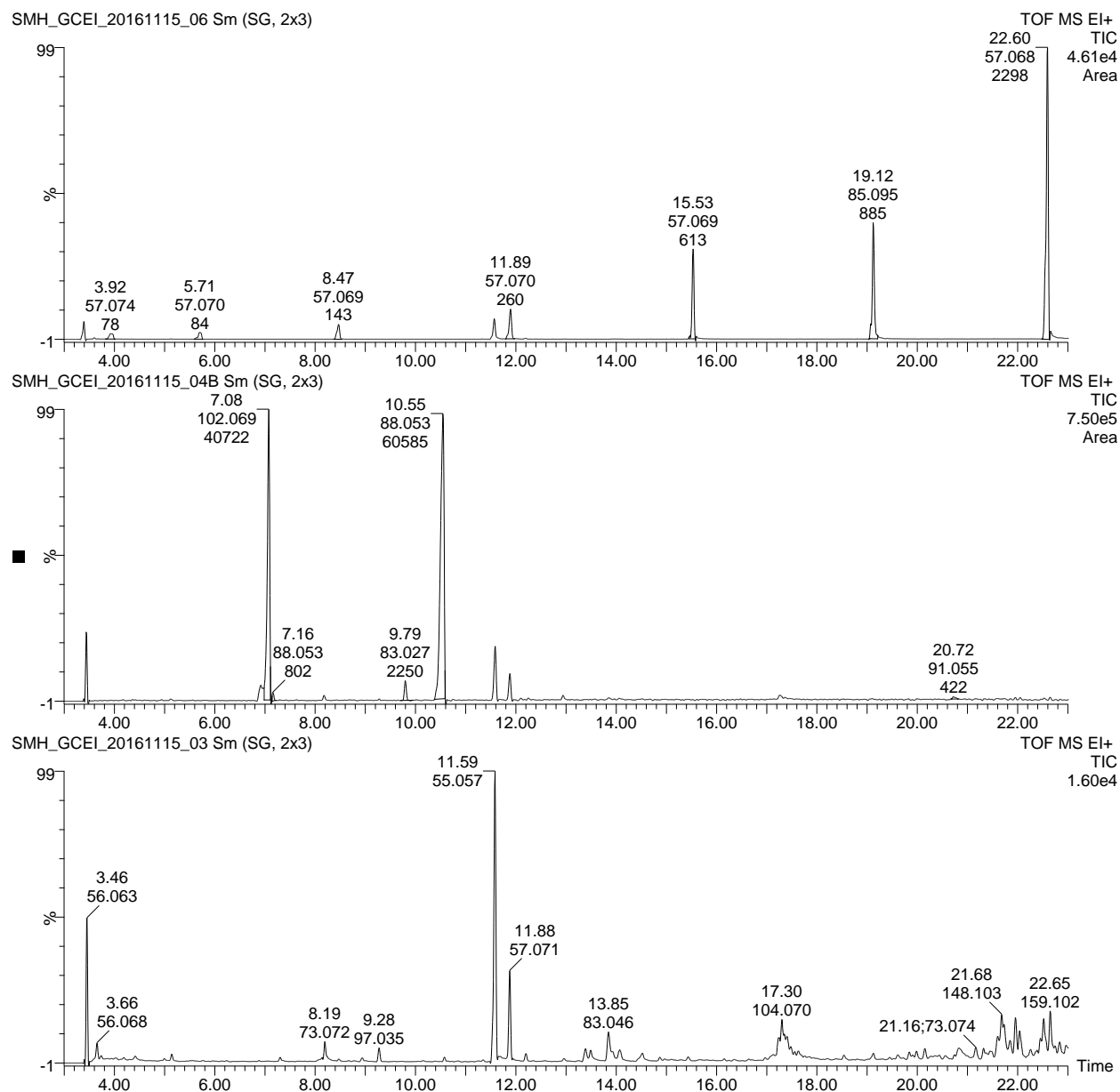
**Figure 2.24.** GC/MS total ion chromatograms (3-23 mins): process blank (bottom), transesterified *S. sinuata* sample (middle), and retention indexing series C7-C13 (top). Peaks are annotated with retention time, base peak *m/z*, and integrated peak area.

**Table 2.7.** GC/MS retention indexing and TIC peak area assessment

| Standard | Ret. Time (min) | Peak Height | Peak Area |
|---|---|---|---|
| heptane | 3.922 | 852 | 78 |
| octane | 5.706 | 1064 | 84 |
| nonane | 8.466 | 2362 | 143 |
| decane | 11.894 | 4711 | 260 |
| undecane | 15.531 | 14196 | 613 |
| dodecane | 19.118 | 18333 | 885 |
| tridecane | 22.597 | 46077 | 2298 |

| Ethyl Ester | | | | Retention index | % of Ethyl Ester TIC Peak Area |
|---|---|---|---|---|---|
| C2 | ND | --- | --- | --- | |
| C3 | ND | --- | --- | --- | |
| iC4 | 4.940 | BLQ | BLQ | 757 | --- |
| aiC5 | 7.075 | 746669 | 40722 | 850 | 38.9% |
| iC5 | 7.159 | 22638 | 802 | 853 | 0.8% |
| C5$^T$ | 9.795 | 50639 | 2250 | 939 | 2.1% |
| aiC6 | 10.546 | 732441 | 60585 | 961 | 57.8% |
| C8$^P$ | 20.722 | 6216 | 422 | 1246 | 0.4% |

**ND** = Not detected

**BLQ** = Below Limit of Quantitation in total ion chromatogram
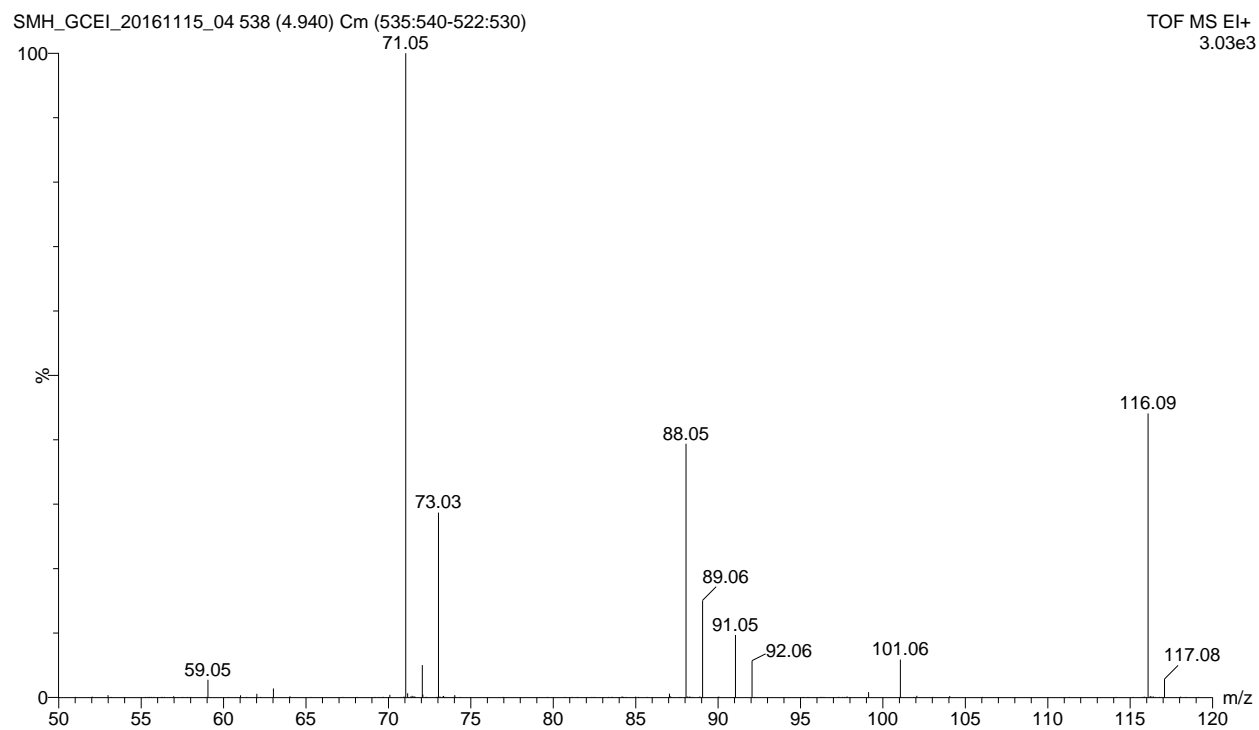
**Figure 2.25.** GC/EI/MS of 2-methylpropanoic acid ethyl ester (iC4)
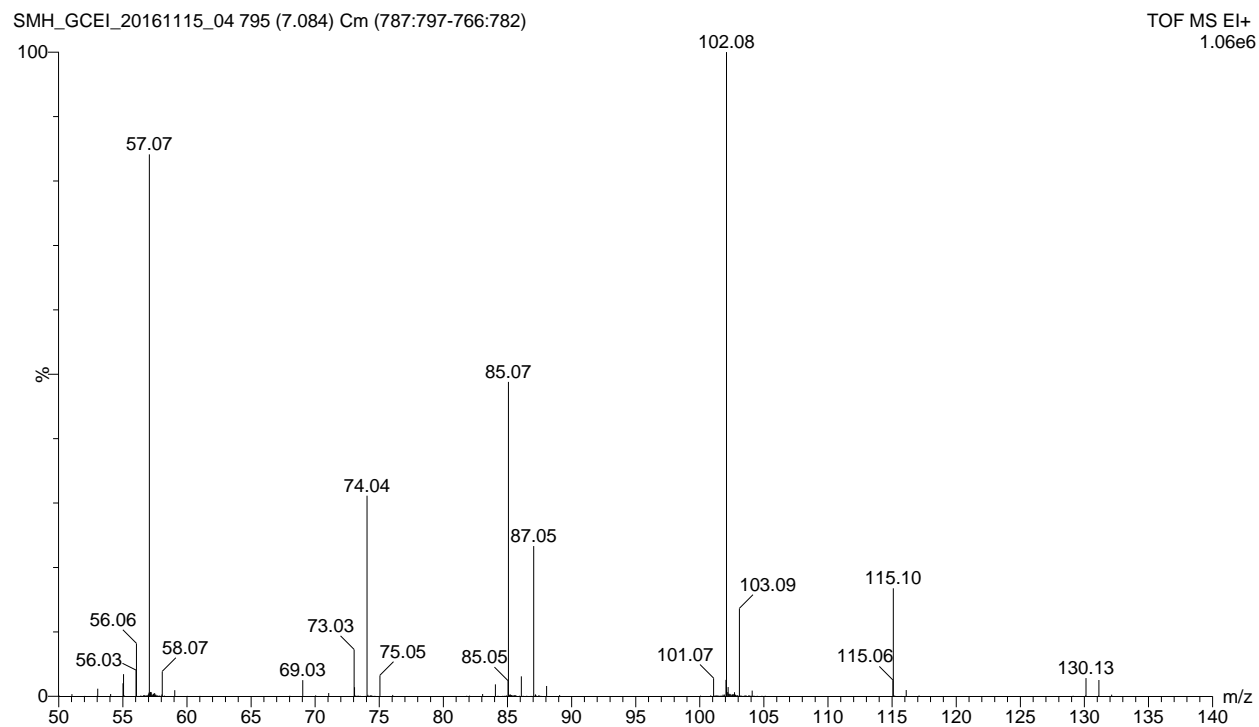
**Figure 2.26.** GC/EI/MS of 2-methylbutanoic acid ethyl ester (aiC5)
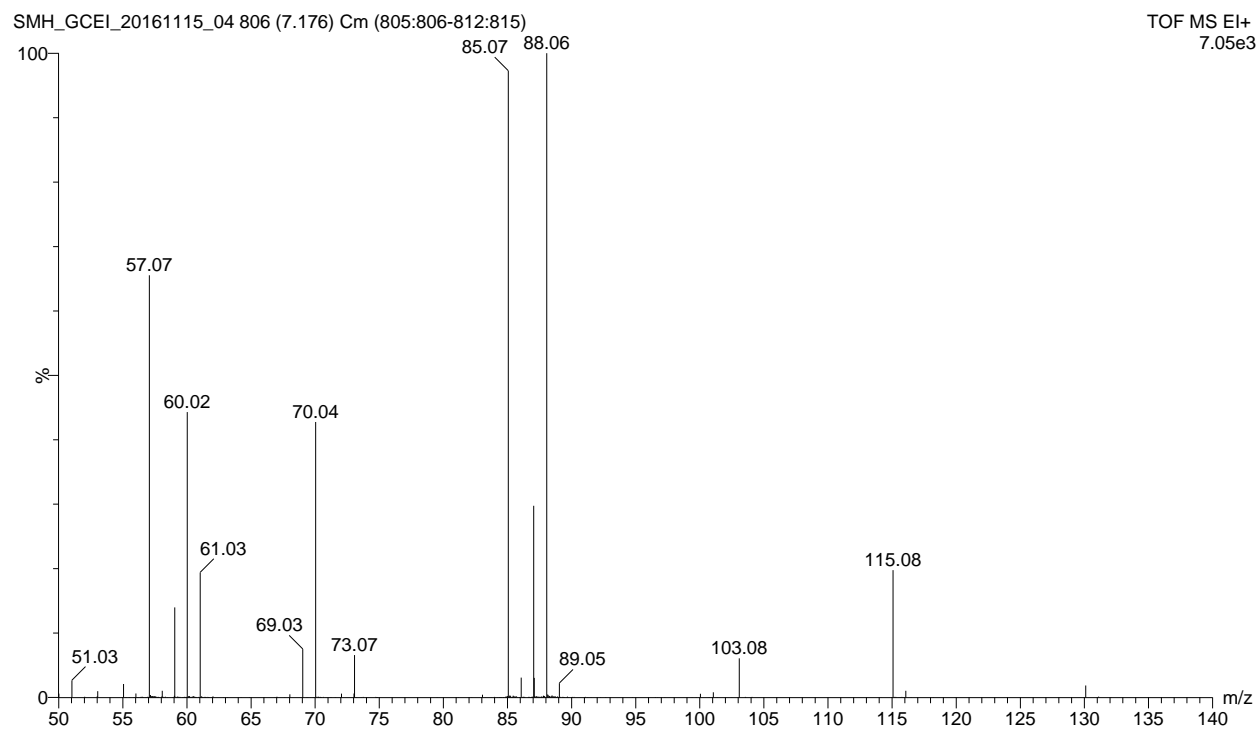
**Figure 2.27.** GC/EI/MS of 3-methylbutanoic acid ethyl ester (iC5)
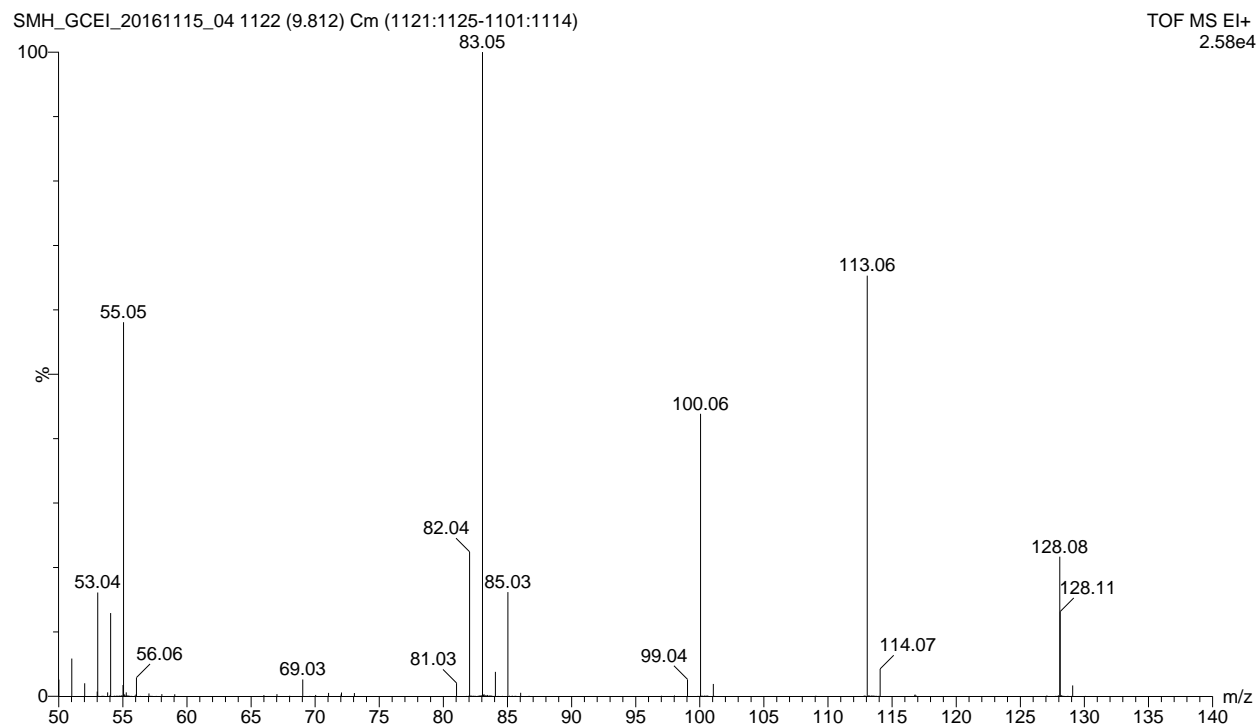
**Figure 2.28.** GC/EI/MS of ethyl tiglate (C5$^T$)
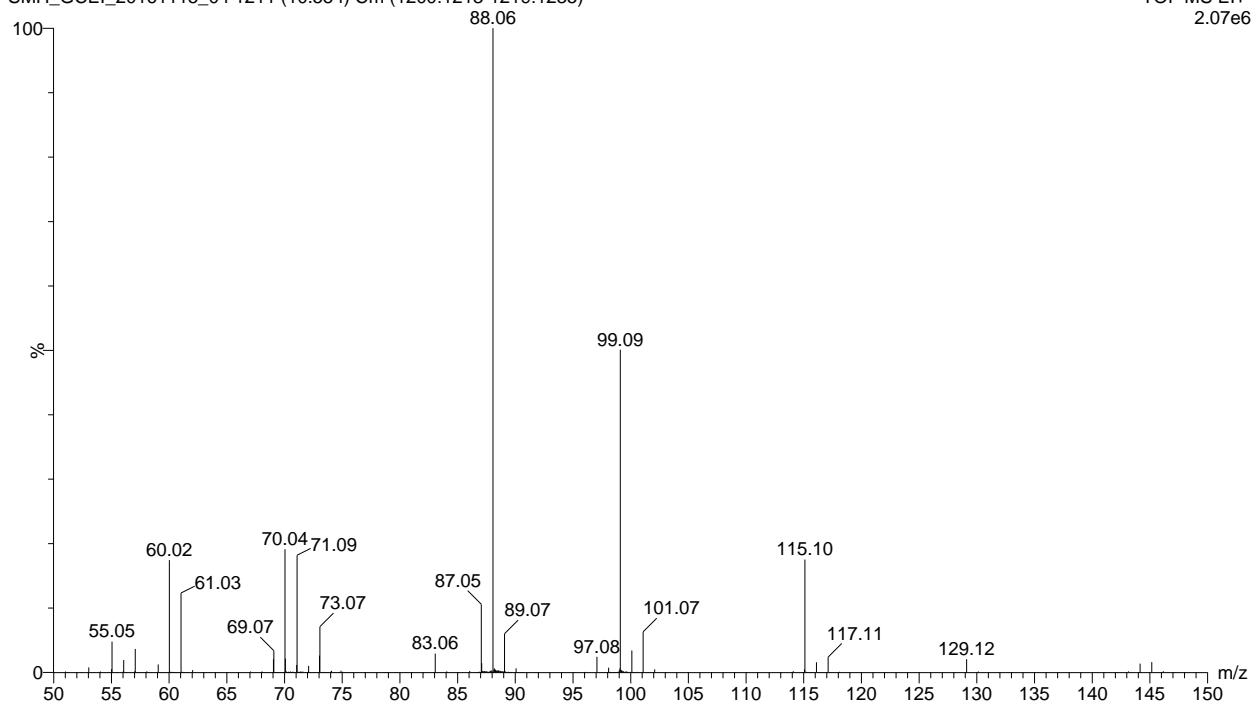
**Figure 2.29.** GC/EI/MS of 3-methylpentanoic acid ethyl ester (aiC6)
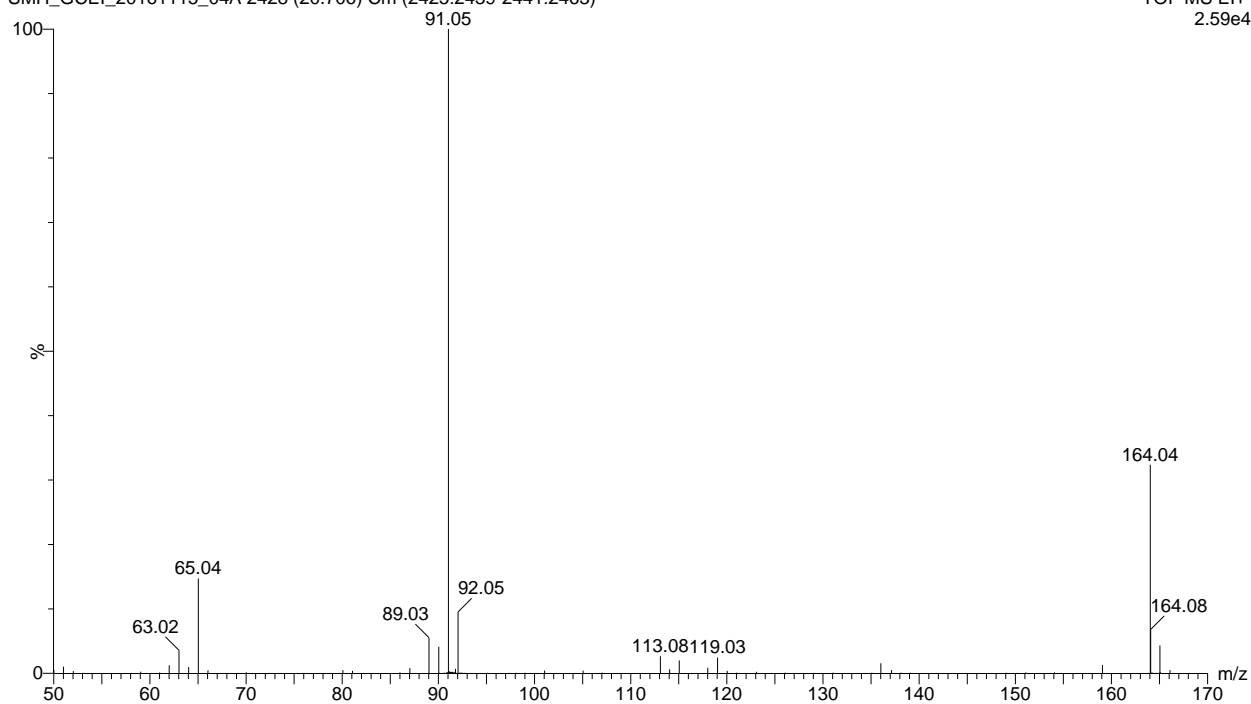
**Figure 2.30.** GC/EI/MS of phenylacetic acid ethyl ester (C8ᴾ)

**Table 2.8.** Purification of acylsucroses by semi-preparative HPLC

| | |
|---|---|
| Analyst | Steven M. Hurney |
| Instrument Location | A. Daniel Jones Laboratory |
| Instrument Title | Waters semi-prep HPLC (Jones Lab) |
| LC System | Waters 2795 Separations Module equipped with LKB Bromma 2211 Superrac Fraction Collector with automated fraction collection |
| Manufacturers | Waters |
| Column | Thermo Scientific Acclaim 120 C18 semi-preparative HPLC column (4.6 x 150 mm, 5 µm particle size) |
| Column Manufacturer | Thermo Scientific |
| Product Number | 059148 |
| Serial Number | 005992 |
| Packing Lot Number | 014-25-017 |
| Injection Volume | 50 µL |
| Flow Rate | 1.5 mL/min |
| Mobile Phases: | |
| A | 0.15% Formic Acid in Water |
| B | Acetonitrile |
| C | dichloromethane: acetone: methanol (v/v/v, 1;1:1) |
| Gradient Profile (solvents A and B only) | Hold 5% B at 0-1 min, linear 5-45% B at 1-15 min, linear 45%-49% B at 15-55 min, linear 49-52% B at 55-70 min, linear 52-65% B at 70-97 min, linear 65-100% B at 97-100 min. |
| Column Wash and Re-equilibration | Hold 100% B at 100-102 min, linear 100% B to 100% C at 102-103 min, hold 100% C at 103-111 min, linear 100% C to 100% B at 111-112 min, linear 100-5% B at 112-113 min, and hold 5% B at 113-120 min. |
| Column Oven Temperature | 50 ℃ |
| Sample Temperature in autosampler | Room Temperature |
| Method Name | SP_56C_70min_AcN_MeOH |
| Sample handling | Approximately 1 L of the *S. sinuata* bulk extract was added in portions of ~100 mL to a 250 mL round bottom flask, and dried via rotary evaporation under vacuum at ~30°C, leaving a green residue. The residue was reconstituted in 5.0 mL AcN:IPA with sonication for 5 mins while manually swirling. The solution was centrifuged by Eppendorf Centrifuge 5480R at $10000 \times g$ for 10 mins. The supernatant was then transferred to LC autosampler vials for semi-preparative HPLC purification. |
| Sample Storage | -20 °C in autosampler vials |
| Protocol when analyzing the samples | One minute fractions were collected in Pyrex glass culture tubes ($18 \times 150$ mm) in eight batches labeled letter A-H. Each batch consisted of six injections each. Fractions were concentrated to dryness under vacuum using a Thermo Savant SPD 131 DDA SpeedVac Concentrator with BOC Edwards XDS Dry Pump. Fractions were labeled by minute, with fractions 33-96 containing the most abundant acylsucroses. |

**Table 2.8.** (continued)

| | |
|---|---|
| Protocol for testing the fractions | In order to test for purity and reproducibility, fractions from each batch were reconstituted in 0.50 mL AcN:IPA, and transferred to LC autosampler vials. 5.0 µL aliquots from each fraction were diluted in autosampler vials containing 0.50 mL of AcN:IPA and analyzed on an LCT Premier mass spectrometer equipped with Shimadzu LC-20AD pumps, Shimadzu SIL-5000 autosampler and a Shimadzu CTO-20A column oven using a 30 min gradient (0.3 mL min$^{-1}$, Ascentis Express C18 Analytical HPLC, 10cm x 2.1mm x 2.7µm) optimized for separating *S. sinuata* acylsugars. The elution program is as follows: hold 1% B at 0-1 min, linear 1-45% B at 1-2 min, linear 45%-65% B at 2-27 min, linear 65-100% B at 27-28 min, hold 100% B at 28-32 min, linear 100-1% B at 32-32.01 min, hold 1% B at 32.01-35 min. HPLC fractions were combined according to metabolite purity and by comparison to a *S. sinuata* bulk extract solution analyses run before and after each group of fractions (i.e. the sample order was blank, *S. sinuata* bulk extract, ~20 fraction samples in ascending elution order, blank, *S. sinuata* bulk extract, and etc.). |
| Purification of S5:25:4(2,5,5,5,8$^P$) | Fractions #55-56 (A-G) were combined and concentrated to dryness under N$_2$ gas, and reconstituted in AcN:IPA and transferred to autosampler vials with limited volume insert. Two 200 µL injections were performed for secondary purification using a Supelco Ascentis® Express F5 semi-preparative HPLC column (15 cm × 4.6 mm, 2.7 µm particle size, Cat#53591-U, BL: S14029, Col: USBK001496). The elution program is as follows (1.0 mL min$^{-1}$): hold 5% B at 0-1 min, linear 1-45% B at 1-15 min, linear 45%-49% B at 15-55 min, linear 49-100% B at 55-56 min, hold 100% B at 56-58 min, linear 100% B to 100% C at 58-59 min, hold at 100% C at 59-69 mins, linear 100% C to 100% B at 69-70 mins, 100-5% B at 70-71 mins, and hold 5% B at 71-80 mins. Fractions 26-27 were combined according to purity assessed by LC/MS. |

**Table 2.9.** Bruker 900 MHz NMR Instrument Metadata

| | |
|---|---|
| Facility Supervisor | Dr. Daniel Holmes |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Max T. Rogers NMR Facility |
| Facility Instrument Title | 900 MHz Bruker |
| Time of acquisition | November 2015 - March 2016 |
| Manufacturer | Bruker |
| Field Frequency Lock | Chloroform-$d_1$ or Acetonitrile-$d_3$ |
| Additional Solute | None |
| Solvent | $CDCl_3$: 250-300 µL and $CD_3CN$: 250 µL |
| Chemical Shift Standard | $CDCl_3$ ($\delta_H$ = 7.26 and $\delta_C$ = 77.20 ppm)<br>$CH_3CN$-$d_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm) |
| Concentration Standard | None |
| Instrument | Bruker Avance 900 MHz NMR |
| Geographic location of instrument | 42.7164, -84.4677 |
| Magnet | 899.13-899.00 MHz |
| Probe | Bruker TCI triple-resonance inverse detection cryoprobe |
| Console | Bruker Avance |
| Acquisition Software | Topspin 2.1.6 |
| Acquisition Parameters: | |
| a) Acquisition parameters file reference | $^1$H:/opt/topspin216/exp/stan/nmr/lists/pp/zg<br><br>$^{13}$C:/opt/topspin216/exp/stan/nmr/lists/pp/zgpg30<br><br>COSY:opt/topspin216/exp/stan/nmr/lists/pp/cosygpmfph<br><br>HSQC:/opt/topspin216/exp/stan/nmr/lists/pp/hsqcedetgpsisp2.2<br><br>HMBC: /opt/topspin216/exp/stan/nmr/lists/pp/hmbcgplpndqf<br><br>*J*-resolved: /opt/topspin216/exp/stan/nmr/lists/pp/jresqf<br><br>ROESY: /opt/topspin216/exp/stan/nmr/lists/pp/roesyetgp |
| b) Sample Details | Shigemi (5 mm) NMR tube, Temperature @ 298 K, No Spinning |
| c) Instrument operation details (recorded for each sample independently and are roughly the same for all samples measured; displayed here is the sample containing S4:19:0(3,5,5,6) as an example) | Radiation frequency (MHz):<br><br>$^1$H: 899.0000263<br><br>$^{13}$C: 226.0536927<br><br>COSY: 899.0000268 (F2), 899.0000194 (F1)<br><br>HSQC: 899.0000277 (F2), 226.0536917 (F1)<br><br>HMBC: 899.0000250 (F2), 226.0536889 (F1)<br><br>*J*-resolved: 899.0000266 (F2), 899.0000000 (F1)<br><br>ROESY: 899.0000271 (F2), 899.0000266 (F1)<br><br>Acquisition nucleus:<br><br>$^1$H: 90° = 7.59 µs, $^{13}$C: 90° = 25.0 µs |

**Table 2.9.** (continued)

| | |
|---|---|
| d) Number of scans (scans, dummy scans) | <sup></sup>1H: 32-64, 0<br><br><sup></sup>13C: 1000-10000, 8<br><br>COSY: 8-16, 16<br><br>HSQC: 12-32, 16<br><br>HMBC: 30-70, 32<br><br>*J*-resolved: 32-64, 4<br><br>ROESY: 12-16, 8 |
| e) Number of data points acquired (F2, F1) | $^1$H: 148144<br><br>$^{13}$C: 65536<br><br>COSY: 2048, 200<br><br>HSQC: 1024, 400<br><br>HMBC: 4096, 480<br><br>*J*-resolved: 2048, 128<br><br>ROESY: 1982, 512 |
| e) Dwell time (µs) | $^1$H: 27.0<br><br>$^{13}$C: 9.225<br><br>COSY: 46.4<br><br>HSQC: 46.4<br><br>HMBC: 46.4<br><br>*J*-resolved: 46.4<br><br>ROESY: 46.4 |
| FID and spectral processing parameters: | |
| Processing software | Topspin 3.5.b.91 pl 7 |
| a) Number of data points in spectrum (F2, F1) | $^1$H: 262144<br><br>$^{13}$C: 131072<br><br>COSY: 4096, 4096<br><br>HSQC: 1024, 1024<br><br>HMBC: 4096, 2048<br><br>*J*-resolved: 4096, 4096<br><br>ROESY: 4096, 2048 |
| b) Window function details | $^1$H: exponential multiply, LB = 0 Hz<br>$^{13}$C: exponential multiply, LB = 2 Hz<br>COSY: QSINE, QSINE; LB = -0.3, -0.3 Hz; GB = 0.3, 0; SSB = 2, 2; TM1 = 0, 1; TM2 = 0, 1<br>HSQC: QSINE, QSINE; SSB = 2, 2; TM1 = 0, 0.1; TM2 = 0, 0.9<br>HMBC: QSINE, SINE<br>*J*-resolved: SINE, SINE; LB = -0.3, -0.3 Hz; GB = 0.3, 0; TM1 = 0, 1; TM2 = 0, 1<br>ROESY: QSINE, QSINE; LB = 1.0, 0.3 Hz; GB = 0.35, 0.1; SSB = 2, 2; TM1 = 0, 0.1; TM2 = 0, 0.9 |

**Table 2.10.** Summary of $^1$H chemical shifts of sucrose core hydrogen atoms. Chemical shifts labeled in bold indicate acyl substitutions are located at those positions. All spectra were referenced to non-deuterated solvent signal of CDCl$_3$ ($\delta_H$ = 7.26 ppm), except for S5:25:4(2,5,5,5,8$^P$) which was referenced to non-deuterated solvent signal of acetonitrile-d$_3$ ($\delta_H$ = 1.94 ppm).

| Acylsugar ID | R$_1$ | R$_2$ | R$_3$ | R$_4$ | R$_5$ | R$_6$ | R$_{1'}$ | R$_{2'}$ | R$_{3'}$ | R$_{4'}$ | R$_{5'}$ | R$_{6'}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Tetraacylsucrose** | | | | | | | | | | | | |
| S4:19:0(3,5,5,6) | 5.84 | **4.81** | **5.55** | **4.93** | 4.16 | 3.64, 3.61 | **4.11, 4.05** | --- | 4.17 | 4.31 | 3.67 | 3.89, 3.71 |
| S4:19:0(2,5,6,6) | 5.64 | **4.87** | **5.53** | **4.92** | 4.14 | 3.61 | 3.66, 3.54 | --- | **5.17** | 4.60 | 3.89 | 3.89, 3.71 |
| S4:20:0(4,5,5,6) | 5.84 | **4.81** | **5.57** | **4.92** | 4.14 | 3.62, 3.59 | **4.12, 4.06** | --- | 4.17 | 4.31 | 3.71 | 3.89, 3.70 |
| S4:20:0(3,5,6,6) | 5.78 | **4.87** | **5.53** | **4.92** | 4.16 | 3.64, 3.60 | **4.08, 4.06** | --- | 4.17 | 4.31 | 3.71 | 3.89, 3.71 |
| S4:20:0(2,6,6,6) | 5.63 | **4.87** | **5.52** | **4.92** | 4.14 | 3.62 | 3.66, 3.54 | --- | **5.16** | 4.61 | 3.88 | 3.89, 3.71 |
| S4:21:0(5,5,5,6) | 5.87 | **4.78** | **5.58** | **4.93** | 4.15 | 3.63, 3.60 | **4.13, 4.06** | --- | 4.17 | 4.31 | 3.71 | 3.88, 3.71 |
| S4:22:0(5,5,6,6) | 5.81 | **4.85** | **5.54** | **4.93** | 4.14 | 3.63, 3.58 | **4.09, 4.07** | --- | 4.17 | 4.31 | 3.69 | 3.88, 3.71 |
| S4:23:0(5,6,6,6) | 5.78 | **4.86** | **5.53** | **4.92** | 4.12 | 3.64 | **4.09, 4.06** | --- | 4.17 | 4.31 | 3.71 | 3.88, 3.70 |
| | | | | | | | | | | | | |
| **Pentaacylsucrose** | | | | | | | | | | | | |
| S5:20:0(2,2,5,5,6) | 5.65 | **4.88** | **5.58** | **4.98** | 4.06 | 3.66, 3.58 | 3.63, 3.55 | --- | **5.22** | 4.36 | 4.07 | **4.41, 4.28** |
| S5:21:0(2,2,5,6,6) | 5.62 | **4.90** | **5.55** | **4.97** | 4.05 | 3.66, 3.58 | 3.63, 3.55 | --- | **5.20** | 4.35 | 4.08 | **4.41, 4.28** |
| S5:22:0(2,2,6,6,6) | 5.62 | **4.90** | **5.54** | **4.98** | 4.04 | 3.67, 3.60 | 3.63, 3.55 | --- | **5.18** | 4.35 | 4.08 | **4.43, 4.28** |
| S5:23:0(2,5,5,5,6) | 5.67 | **4.90** | **5.56** | **4.94** | 4.16 | 3.65, 3.59 | **4.18, 4.15** | --- | 4.14 | 4.11 | 3.87 | **4.39, 4.24** |
| S5:24:0(2,5,5,6,6) | 5.66 | **4.93** | **5.54** | **4.94** | 4.14 | 3.66, 3.59 | **4.16** | --- | 4.14 | 4.11 | 3.90 | **4.42, 4.22** |
| | | | | | | | | | | | | |
| **Pentaacylsucrose w/unsaturated ester** | | | | | | | | | | | | |
| S5:22:1(2,5,5,5,5$^T$) | 5.65 | **5.02** | **5.63** | **5.03** | 4.15 | 3.68, 3.61 | **4.21, 4.17** | --- | 4.15 | 4.11 | 3.90 | **4.42, 4.23** |
| S5:25:4(2,5,5,5,8$^P$) | 5.58 | **4.90** | **5.49** | **5.03** | 4.12 | 3.57, 3.45 | **4.01** | --- | 3.98 | 3.97 | 3.80 | **4.26, 4.19** |
| | | | | | | | | | | | | |
| **Hexaacylsucrose** | | | | | | | | | | | | |
| S6:25:0(2,2,5,5,5,6) | 5.70 | **4.87** | **5.57** | **4.97** | 4.07 | 3.66, 3.59 | **4.14, 4.04** | --- | **5.23** | 4.35 | 4.04 | **4.40, 4.30** |

**Table 2.11.** Summary of $^{13}$C chemical shifts of sucrose core carbon atoms. Chemical shifts labeled in bold indicate acyl substitutions are located at those positions. All spectra were referenced to non-deuterated solvent signal of CDCl$_3$ ($\delta_C$ = 77.20 ppm), except for S5:25:4(2,5,5,5,8$^P$) which was referenced to non-deuterated solvent signal of acetonitrile-d$_3$ ($\delta_C$ = 118.70 ppm).

| Acylsugar ID | R$_1$ | R$_2$ | R$_3$ | R$_4$ | R$_5$ | R$_6$ | R$_{1'}$ | R$_{2'}$ | R$_{3'}$ | R$_{4'}$ | R$_{5'}$ | R$_{6'}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Tetraacylsucrose | | | | | | | | | | | | |
| S4:19:0(3,5,5,6) | 88.96 | **70.87** | **68.88** | **68.82** | 72.05 | 61.55 | **63.88** | 103.61 | 78.40 | 72.66 | 81.10 | 59.71 |
| S4:19:0(2,5,6,6) | 89.75 | **70.53** | **68.86** | **68.62** | 71.94 | 61.86 | 64.63 | 104.15 | **80.13** | 71.54 | 82.38 | 59.86 |
| S4:20:0(4,5,5,6) | 89.03 | **70.92** | **68.72** | **68.61** | 72.09 | 61.49 | **63.93** | 103.66 | 78.49 | 72.89 | 81.16 | 59.74 |
| S4:20:0(3,5,6,6) | 89.14 | **70.56** | **68.98** | **68.82** | 72.19 | 61.57 | **63.63** | 103.45 | 78.45 | 72.79 | 81.33 | 59.77 |
| S4:20:0(2,6,6,6) | 89.78 | **70.52** | **68.87** | **68.71** | 71.99 | 61.91 | 64.59 | 104.16 | **80.31** | 71.52 | 82.32 | 59.76 |
| S4:21:0(5,5,5,6) | 88.79 | **71.11** | **68.76** | **68.58** | 71.99 | 61.51 | **64.07** | 103.74 | 78.43 | 72.56 | 80.98 | 59.71 |
| S4:22:0(5,5,6,6) | 89.00 | **70.81** | **68.82** | **68.55** | 72.15 | 61.52 | **63.80** | 103.58 | 78.39 | 72.74 | 81.22 | 59.78 |
| S4:23:0(5,6,6,6) | 89.15 | **70.69** | **68.83** | **68.66** | 72.21 | 61.55 | **63.69** | 103.52 | 78.43 | 72.93 | 81.35 | 59.81 |
| | | | | | | | | | | | | |
| Pentaacylsucrose | | | | | | | | | | | | |
| S5:20:0(2,2,5,5,6) | 89.52 | **70.55** | **68.9$^a$** | **68.9$^a$** | 71.11 | 61.64 | 64.34 | 104.58 | **79.23** | 73.72 | 80.38 | **64.08** |
| S5:21:0(2,2,5,6,6) | 89.52 | **70.48** | **68.98** | **68.85** | 71.13 | 61.64 | 64.27 | 104.65 | **79.30** | 73.90 | 80.43 | **64.27** |
| S5:22:0(2,2,6,6,6) | 89.58 | **70.49** | **69.0$^b$** | **69.0$^b$** | 71.17 | 61.67 | 64.23 | 104.66 | **79.56** | 74.04 | 80.39 | **64.27** |
| S5:23:0(2,5,5,5,6) | 89.36 | **70.24** | **68.88** | **68.94** | 71.28 | 61.65 | **63.70** | 103.63 | 78.11 | 74.69 | 78.63 | **63.49** |
| S5:24:0(2,5,5,6,6) | 89.51 | **70.22** | **68.93** | **68.90** | 71.31 | 61.64 | **63.49** | 103.57 | 78.16 | 75.03 | 78.80 | **63.69** |
| | | | | | | | | | | | | |
| Pentaacylsucrose w/unsaturated ester | | | | | | | | | | | | |
| S5:22:1(2,5,5,5,5$^T$) | 89.99 | **70.06** | **69.12** | **68.73** | 71.49 | 61.58 | **63.39** | 103.58 | 78.31 | 75.33 | 78.79 | **63.62** |
| S5:25:4(2,5,5,5,8$^P$) | 90.22 | **71.37** | **71.49** | **69.70** | 72.08 | 61.95 | **64.02** | 104.55 | 78.29 | 75.16 | 80.59 | **65.60** |
| | | | | | | | | | | | | |
| Hexaacylsucrose | | | | | | | | | | | | |
| S6:25:0(2,2,5,5,5,6) | 89.41 | **70.37** | **68.87** | **68.97** | 71.15 | 61.67 | **64.10** | 102.98 | **78.95** | 73.45 | 80.22 | **63.76** |

a - Two $^{13}$C signals not resolved in 2D spectra (68.91, 68.93 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (68.96, 68.97 ppm)

**Table 2.12.** S4:19:0(3,5,5,6) Chemical shifts and coupling constants
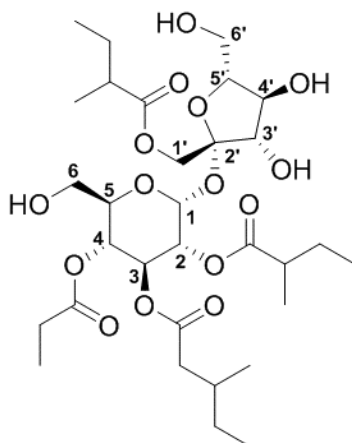


**Molecular Formula:** $C_{31}H_{52}O_{15}$

**110 min Retention Time (ESI+):** 48.21 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{31}H_{56}NO_{15}^+$ ([M+NH$_4$$^+$]): 682.3644, found: 682.3664

**Fraction, Batch:** #34, A-D
**Sample mass for NMR analysis:** 0.8 mg
**NMR Solvent:** CDCl$_3$
**InChi Key:** HGYUMQHIZULVOD-AOJWIRPASA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.84 (d, *J* = 4.0 Hz) | 88.96 |
| **2**(CH) | 4.81 (dd, *J* = 10.4, 4.0 Hz) | 70.87 |
| -1(CO) | | 177.40 |
| -2(CH) | 2.42 (sextet, *J* = 7.0 Hz) | 40.76 |
| -3(CH$_3$) | 1.14 (d, *J* = 7.0 Hz) | 16.16 |
| -4(CH$_2$) | 1.63 (m), 1.45 (m) | 26.80 |
| -5(CH$_3$) | 0.87 (t, *J* = 7.4 Hz) | 11.60 |
| **3**(CH) | 5.55 (dd, *J* = 10.6, 9.2 Hz) | 68.88 |
| -1(CO) | | 172.24 |
| -2(CH$_2$) | 2.25 (dd, *J* = 15.2, 5.8 Hz), 2.03 (dd, *J* = 15.2, 8.3 Hz) | 41.26 |
| -3(CH) | 1.80 (m) | 31.72 |
| -4(CH$_3$) | 0.88 (d, *J* = 6.7 Hz) | 19.30 |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.35 |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.35 |
| **4**(CH) | 4.93 (dd, *J* = 10.7, 9.2 Hz) | 68.82 |
| -1(CO) | | 173.84 |
| -2(CH$_2$) | 2.35 (m), 2.29 (m) | 27.56 |
| -3(CH$_3$) | 1.12 (t, *J* = 7.6 Hz) | 9.08 |
| **5**(CH) | 4.16 (ddd, *J* = 10.4, 6.5, 2.5 Hz) | 72.05 |
| **6**(CH$_2$) | 3.64 (dd, *J* = 12.5, 2.4 Hz), 3.61 (dd, *J* = 12.5, 6.5 Hz) | 61.55 |
| **1'**(CH$_2$) | 4.11 (d, *J* = 11.6 Hz), 4.05 (d, *J* = 11.6 Hz) | 63.88 |
| -1(CO) | | 175.92 |
| -2(CH) | 2.41 (sextet, *J* = 7.0 Hz) | 41.13 |
| -3(CH$_3$) | 1.15 (d, *J* = 7.0 Hz) | 16.73 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH$_3$) | 0.91 (t, *J* = 7.4 Hz) | 11.74 |

**Table 2.12.** (continued)

| | | |
|---|---|---|
| **2'**(C) | | 103.61 |
| **3'**(CH) | 4.17 (d, $J = 8.8$ Hz) | 78.40 |
| **4'**(CH) | 4.31 (t, $J = 8.9$ Hz) | 72.66 |
| **5'**(CH) | 3.67 (dt, $J = 9.0, 2.3$ Hz) | 81.10 |
| **6'**(CH$_2$) | 3.89 (dd, $J = 13.4, 2.4$ Hz), 3.71 (dd, $J = 13.4, 2.3$ Hz) | 59.71 |

**Figure 2.31.** S4:19:0(3,5,5,6) ^{1}H NMR

**Figure 2.32.** S4:19:0(3,5,5,6) $^{13}$C NMR

**Figure 2.33.** S4:19:0(3,5,5,6) $^1$H-$^1$H gCOSY

**Figure 2.34.** S4:19:0(3,5,5,6) gHSQC

**Figure 2.35.** S4:19:0(3,5,5,6) gHMBC

**Figure 2.36.** S4:19:0(3,5,5,6) *J*-resolved

**Figure 2.37.** S4:19:0(3,5,5,6) ROESY

**Table 2.13.** S4:19:0(2,5,6,6) Chemical shifts and coupling constants



**Molecular Formula:** $C_{31}H_{52}O_{15}$

**110 min Retention Time (ESI+):** 48.39 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{31}H_{56}NO_{15}^+$ ([M+NH$_4^+$]): 682.3644, found: 682.3675

**Fraction, Batch:** #35, A-D

**Sample mass for NMR analysis:** 1.3 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** IWACBCFUOAGBIU-QRRKJAOQSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.64 (d, $J$ = 3.7 Hz) | 89.75 |
| **2**(CH) | 4.87 (dd, $J$ = 10.4, 3.7 Hz) | 70.53 |
| -1(CO) | | 172.78 |
| -2(CH$_2$) | 2.32 (dd, $J$ = 15.7, 5.7 Hz), 2.06 (dd, $J$ = 15.7, 8.3 Hz) | 40.99 |
| -3(CH) | 1.82 (m) | 31.58 |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.37 |
| -5(CH$_2$) | 1.34 (m), 1.20 (m) | 29.34 |
| -6(CH$_3$) | 0.88 (t, $J$ = 7.4 Hz) | 11.4[a] |
| **3**(CH) | 5.53 (dd, $J$ = 10.7, 9.1 Hz) | 68.86 |
| -1(CO) | | 172.3[b] |
| -2(CH$_2$) | 2.23 (dd, $J$ = 15.5, 5.6 Hz), 2.02 (dd, $J$ = 15.5, 8.4 Hz) | 41.22 |
| -3(CH) | 1.79 (m) | 31.64 |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.43 |
| -5(CH$_2$) | 1.31 (m), 1.18 (m) | 29.38 |
| -6(CH$_3$) | 0.85 (t, $J$ = 7.4 Hz) | 11.4[a] |
| **4**(CH) | 4.92 (dd, $J$ = 10.7, 9.2 Hz) | 68.62 |
| -1(CO) | | 176.12 |
| -2(CH) | 2.36 (sextet, $J$ = 7.0 Hz) | 41.08 |
| -3(CH$_3$) | 1.10 (d, $J$ = 7.0 Hz) | 16.53 |
| -4(CH$_2$) | 1.67 (m), 1.44 (m) | 26.63 |
| -5(CH$_3$) | 0.90 (t, $J$ = 7.4 Hz) | 11.82 |
| **5**(CH) | 4.14 (ddd, $J$ = 10.2, 5.5, 2.8 Hz) | 71.94 |
| **6**(CH$_2$) | 3.61 (m) | 61.86 |
| **1'**(CH$_2$) | 3.66 (d, $J$ = 12.3 Hz), 3.54 (d, $J$ = 12.3 Hz) | 64.63 |
| **2'**(C) | | 104.15 |

**Table 2.13.** (continued)

| | | |
|---|---|---|
| **3'**(CH) | 5.17 (d, $J$ = 7.8 Hz) | 80.13 |
| -1(CO) | | 172.3[b] |
| -2(CH$_3$) | 2.26 (s) | 21.00 |
| **4'**(CH) | 4.60 (t, $J$ = 7.8 Hz) | 71.54 |
| **5'**(CH) | 3.89 (m) | 82.38 |
| **6'**(CH$_2$) | 3.89 (m), 3.71 (m) | 59.86 |

a - Two $^{13}$C signals not resolved in 2D spectra (11.37, 11.39 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (172.27, 172.29 ppm)

**Figure 2.38.** S4:19:0(2,5,6,6) $^1$H NMR

111

**Figure 2.39.** S4:19:0(2,5,6,6) ^{13}C NMR

112

**Figure 2.40.** S4:19:0(2,5,6,6) $^1$H-$^1$H gCOSY

**Figure 2.41.** S4:19:0(2,5,6,6) gHSQC

**Figure 2.42.** S4:19:0(2,5,6,6) gHMBC

**Figure 2.43.** S4:19:0(2,5,6,6) *J*-resolved

**Figure 2.44.** S4:19:0(2,5,6,6) ROESY

**Table 2.14.** S5:20:0(2,2,5,5,6) Chemical shifts and coupling constants



**Molecular Formula:** $C_{32}H_{52}O_{16}$

**110 min Retention Time (ESI+):** 49.57 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{32}H_{56}NO_{16}^+$ ([M+NH$_4^+$]): 710.3594, found: 710.3621

**Fraction, Batch:** #38, A-D
**Sample mass for NMR analysis:** 1.5 mg
**NMR Solvent:** CDCl$_3$
**InChi Key:** DVSXLPKXNQNXQO-CJWUYPCVSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.65 (d, *J* = 3.6 Hz) | 89.52 |
| **2**(CH) | 4.88 (dd, *J* = 10.5, 3.7 Hz) | 70.55 |
| -1(CO) | | 175.86 |
| -2(CH) | 2.36 (sextet, *J* = 7.0 Hz) | 41.10 |
| -3(CH$_3$) | 1.10 (d, *J* = 7.0 Hz) | 16.47 |
| -4(CH$_2$) | 1.67 (m), 1.45 (m) | 26.69 |
| -5(CH$_3$) | 0.90 (t, *J* = 7.4 Hz) | 11.82 |
| **3**(CH) | 5.58 (dd, *J* = 10.7, 9.1 Hz) | 68.9[a] |
| -1(CO) | | 172.29 |
| -2(CH$_2$) | 2.24 (dd, *J* = 15.8, 5.4 Hz), 2.01 (dd, *J* = 15.8, 8.6 Hz) | 41.15 |
| -3(CH) | 1.78 (m) | 31.45 |
| -4(CH$_3$) | 0.88 (d, *J* = 6.7 Hz) | 19.47 |
| -5(CH$_2$) | 1.31 (m), 1.18 (m) | 29.43 |
| -6(CH$_3$) | 0.85 (t, *J* = 7.4 Hz) | 11.39 |
| **4**(CH) | 4.98 (dd, *J* = 10.6, 9.2 Hz) | 68.9[a] |
| -1(CO) | | 176.40 |
| -2(CH) | 2.35 (sextet, *J* = 7.0 Hz) | 40.73 |
| -3(CH$_3$) | 1.11 (d, *J* = 7.0 Hz) | 16.08 |
| -4(CH$_2$) | 1.61 (m), 1.43 (m) | 26.83 |
| -5(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.67 |
| **5**(CH) | 4.06 (ddd, *J* = 10.4, 5.1, 2.1 Hz) | 71.11 |
| **6**(CH$_2$) | 3.66 (dd, *J* = 12.8, 2.5 Hz), 3.58 (dd, *J* = 12.8, 5.1 Hz) | 61.64 |
| **1'**(CH$_2$) | 3.63 (d, *J* = 12.2 Hz), 3.55 (d, *J* = 12.2 Hz) | 64.34 |
| **2'**(C) | | 104.58 |

**Table 2.14.** (continued)

| | | |
|---|---|---|
| **3'**(CH) | 5.22 (d, $J = 7.7$ Hz) | 79.23 |
| -1(CO) | | 172.05 |
| -2(CH$_3$) | 2.22 (s) | 20.99 |
| **4'**(CH) | 4.36 (t, $J = 7.8$ Hz) | 73.72 |
| **5'**(CH) | 4.07 (ddd, $J = 7.8$, 6.1, 3.7 Hz) | 80.38 |
| **6'**(CH$_2$) | 4.41 (dd, $J = 12.1$, 6.0 Hz), 4.28 (dd, $J = 12.2$, 3.5 Hz) | 64.08 |
| -1(CO) | | 171.69 |
| -2(CH$_3$) | 2.13 (s) | 21.01 |

a - Two $^{13}$C signals not resolved in 2D spectra (68.91, 68.93 ppm)

**Figure 2.45.** S5:20:0(2,2,5,5,6) ^{1}H NMR

**Figure 2.46.** S5:20:0(2,2,5,5,6) $^{13}$C NMR

**Figure 2.47.** S5:20:0(2,2,5,5,6) $^1$H-$^1$H gCOSY

**Figure 2.48.** S5:20:0(2,2,5,5,6) gHSQC

**Figure 2.49.** S5:20:0(2,2,5,5,6) gHMBC

**Figure 2.50.** S5:20:0(2,2,5,5,6) *J*-resolved

**Figure 2.51.** S5:20:0(2,2,5,5,6) ROESY

**Table 2.15.** S4:20:0(4,5,5,6) Chemical shifts and coupling constants



**Molecular Formula:** $C_{32}H_{54}O_{15}$

**110 min Retention Time (ESI+):** 50.78 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{32}H_{58}NO_{15}^{+}$ ([M+NH$_4$$^{+}$]): 696.3801, found: 696.3804
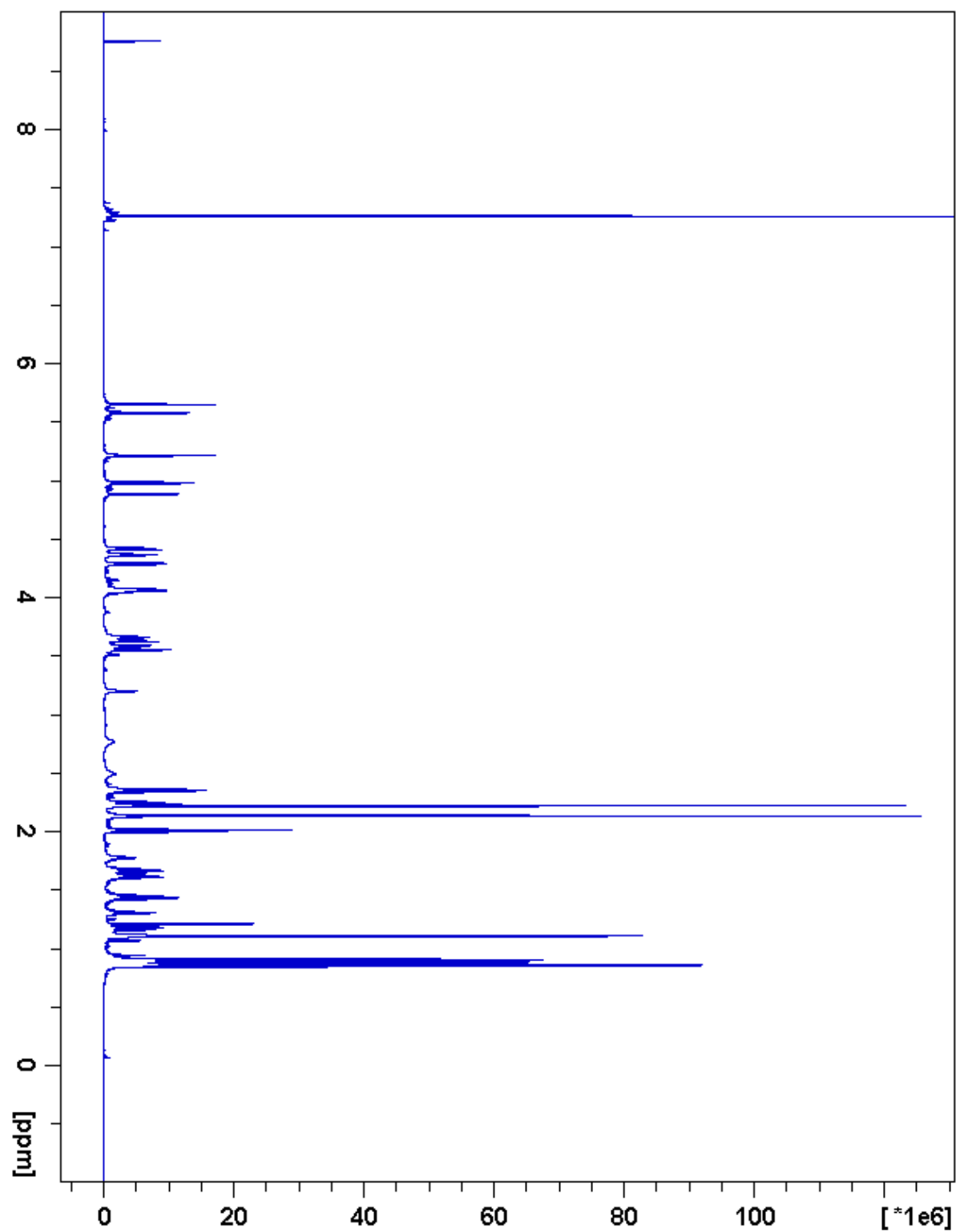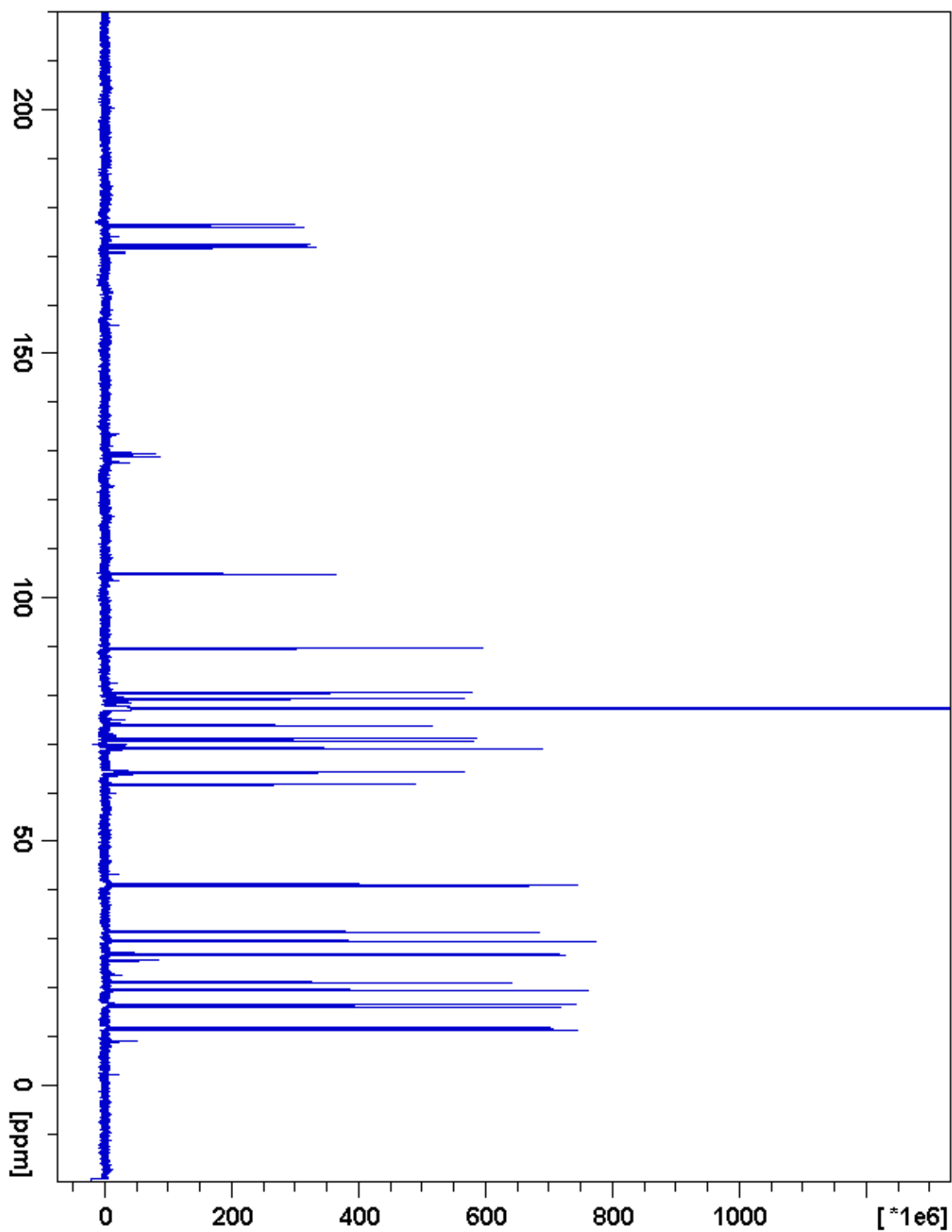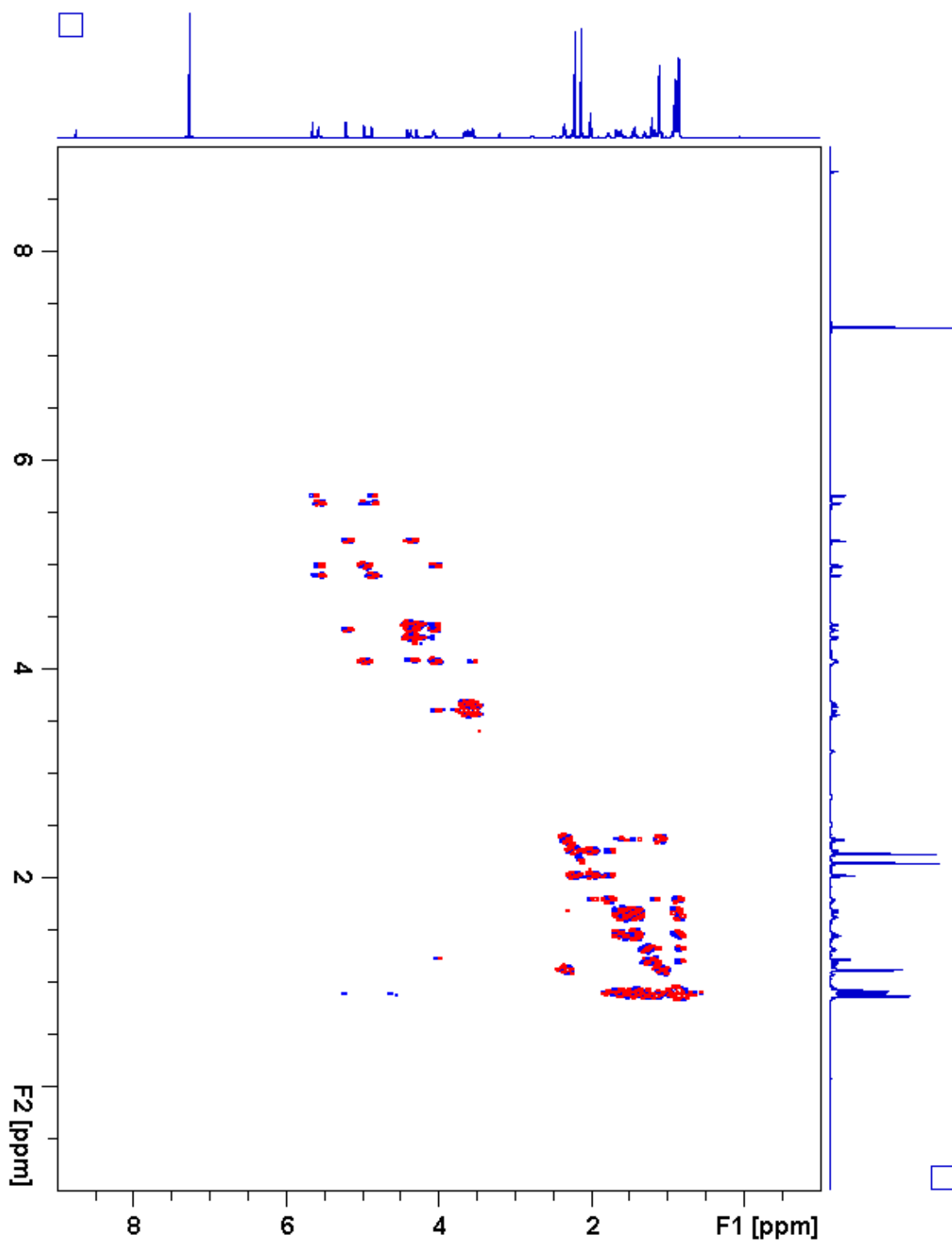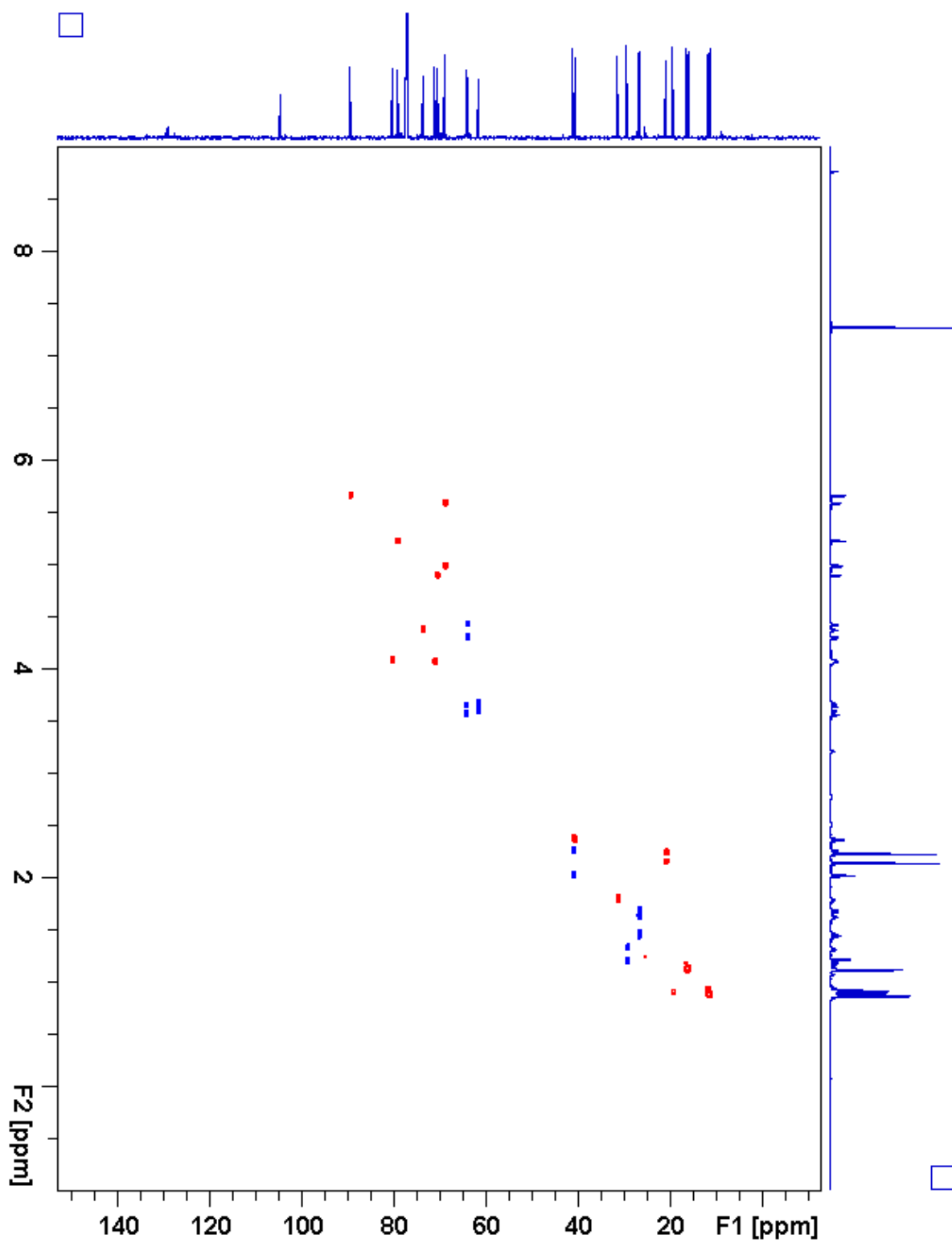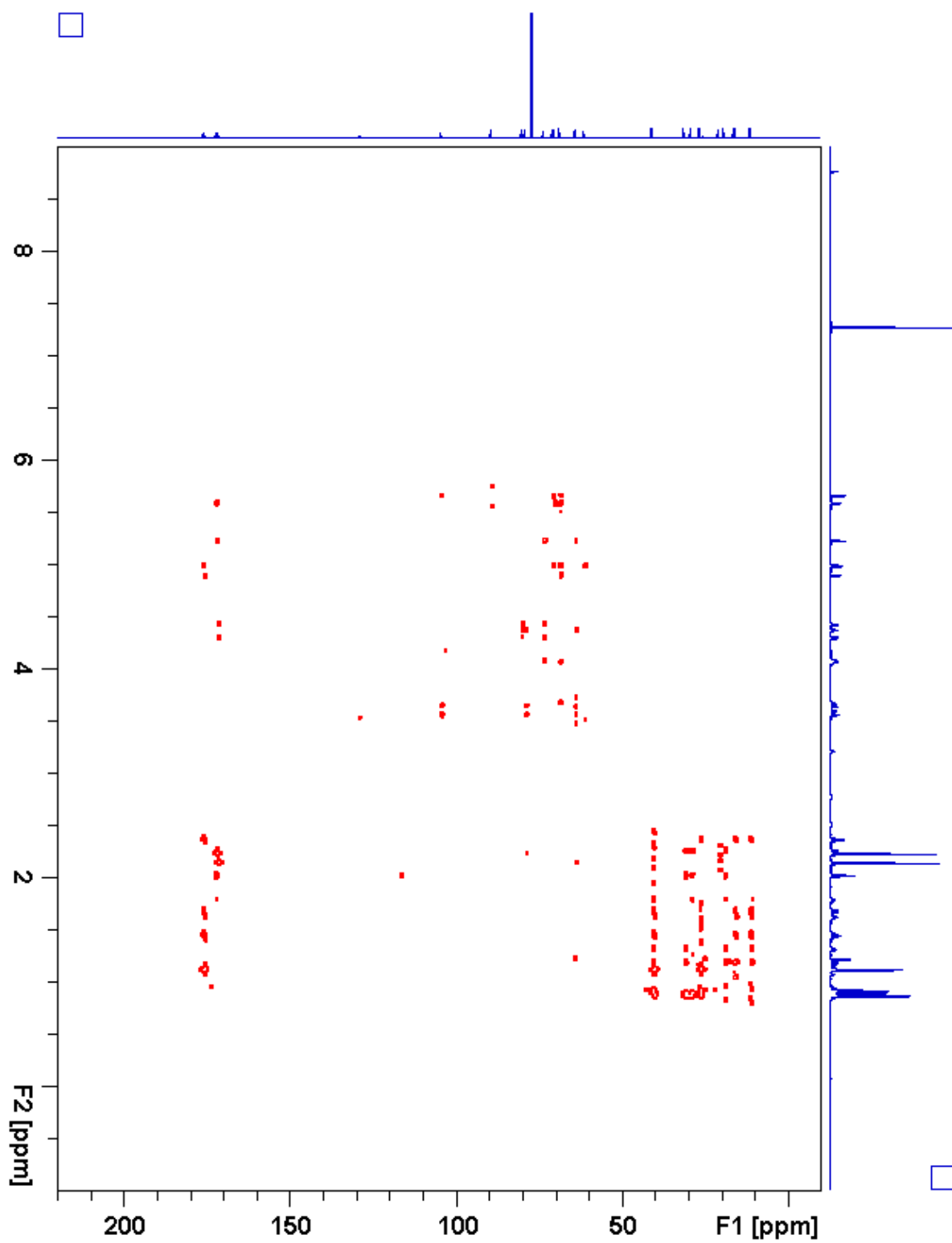
**Fraction, Batch:** #42, A-D

**Sample mass for NMR analysis:** 0.2 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** JSZQPBVIKRWRRQ-IRCRNTFMSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.84 (d, *J* = 4.0 Hz) | 89.03 |
| **2**(CH) | 4.81 (dd, *J* = 10.3, 4.1 Hz) | 70.92 |
| -1(CO) | | 177.36 |
| -2(CH) | 2.41 (sextet, *J* = 7.0 Hz) | 40.76 |
| -3(CH$_3$) | 1.15 (d, *J* = 7.0 Hz) | 16.15 |
| -4(CH$_2$) | 1.64 (m), 1.45 (m) | 26.76 |
| -5(CH$_3$) | 0.87 (t, *J* = 7.4 Hz) | 11.62 |
| **3**(CH) | 5.57 (dd, *J* = 10.6, 9.3 Hz) | 68.72 |
| -1(CO) | | 172.11 |
| -2(CH$_2$) | 2.24 (dd, *J* = 15.6, 5.6 Hz), 2.02 (dd, *J* = 15.6, 8.4 Hz) | 41.1[a] |
| -3(CH) | 1.80 (m) | 31.59 |
| -4(CH$_3$) | 0.89 (d, *J* = 6.7 Hz) | 19.41 |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.38 |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.38 |
| **4**(CH) | 4.92 (dd, *J* = 10.7, 9.3 Hz) | 68.61 |
| -1(CO) | | 176.56 |
| -2(CH) | 2.54 (septet, *J* = 7.0 Hz) | 34.10 |
| -3(CH$_3$) | 1.16 (d, *J* = 7.0 Hz) | 19.08 |
| -4(CH$_3$) | 1.13 (d, *J* = 7.0 Hz) | 18.83 |
| **5**(CH) | 4.14 (ddd, *J* = 10.5, 6.1, 1.9 Hz) | 72.09 |
| **6**(CH$_2$) | 3.62 (dd, *J* = 12.7, 2.1 Hz), 3.59 (dd, *J* = 12.7, 6.1 Hz) | 61.49 |

**Table 2.15.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.12 (d, $J$ = 11.7 Hz), 4.06 (d, $J$ = 11.7 Hz) | 63.93 |
| -1(CO) | | 175.91 |
| -2(CH) | 2.41 (sextet, $J$ = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.16 (d, $J$ = 7.0 Hz) | 16.74 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.89 |
| -5(CH$_3$) | 0.91 (t, $J$ = 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.66 |
| **3'**(CH) | 4.17 (d, $J$ = 8.7 Hz) | 78.49 |
| **4'**(CH) | 4.31 (t, $J$ = 8.8 Hz) | 72.89 |
| **5'**(CH) | 3.71 (dt, $J$ = 9.0, 2.3 Hz) | 81.16 |
| **6'**(CH$_2$) | 3.89 (dd, $J$ = 13.3, 2.4 Hz), 3.70 (dd, $J$ = 13.3, 2.3 Hz) | 59.74 |

a – Two $^{13}$C signals not resolved in 2D spectra (41.13, 41.15 ppm)
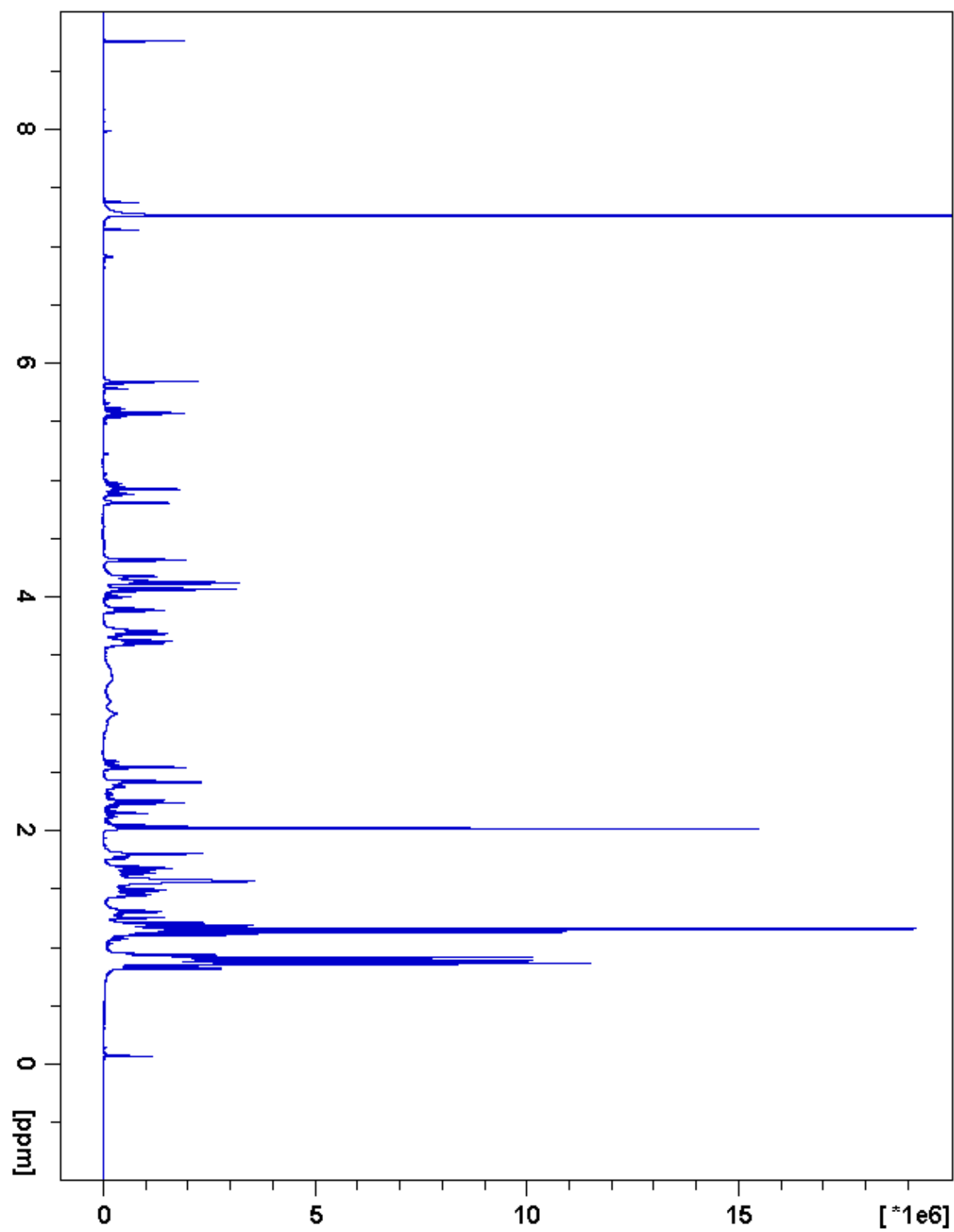
**Figure 2.52.** S4:20:0(4,5,5,6) $^1$H NMR
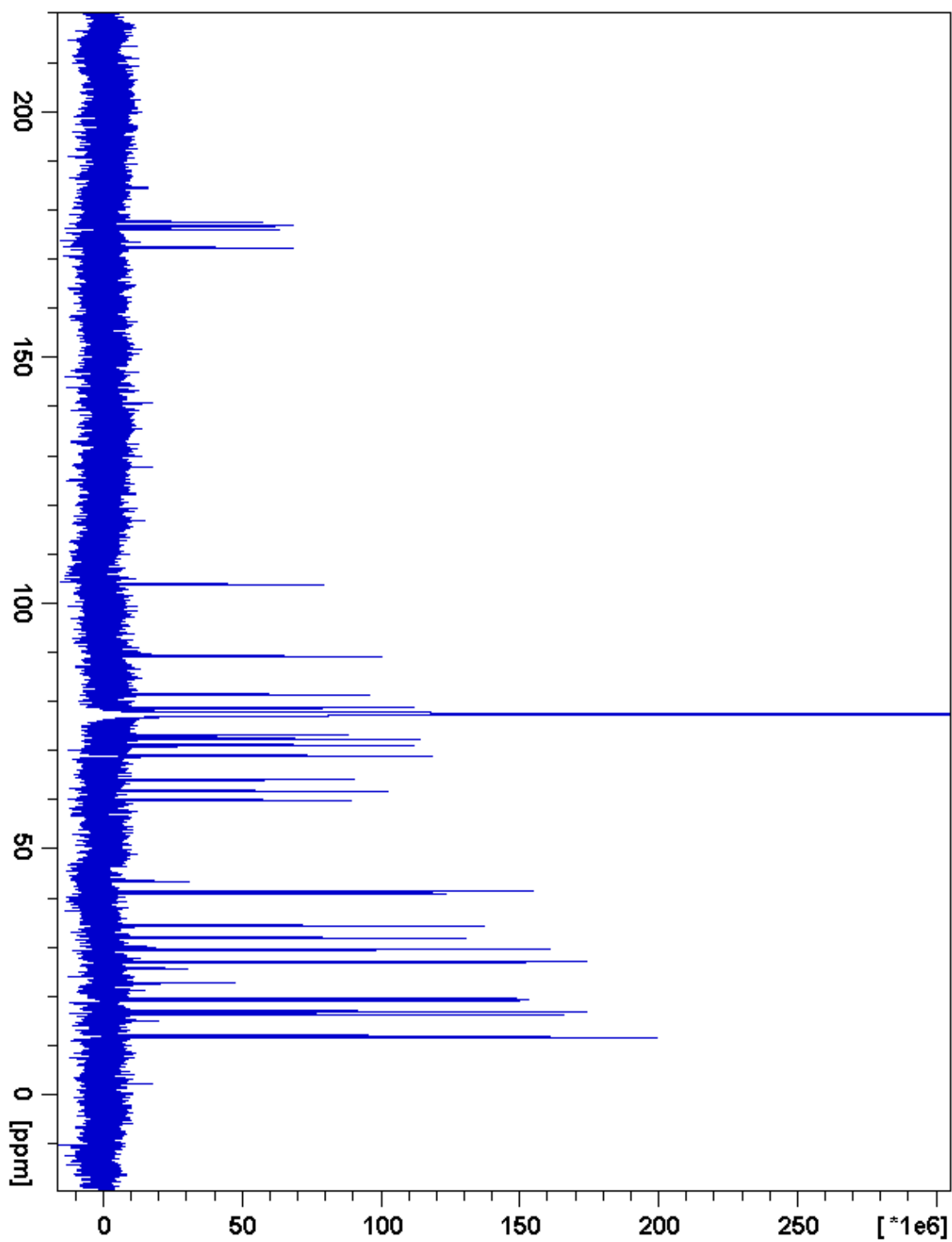
**Figure 2.53.** S4:20:0(4,5,5,6) $^{13}$C NMR
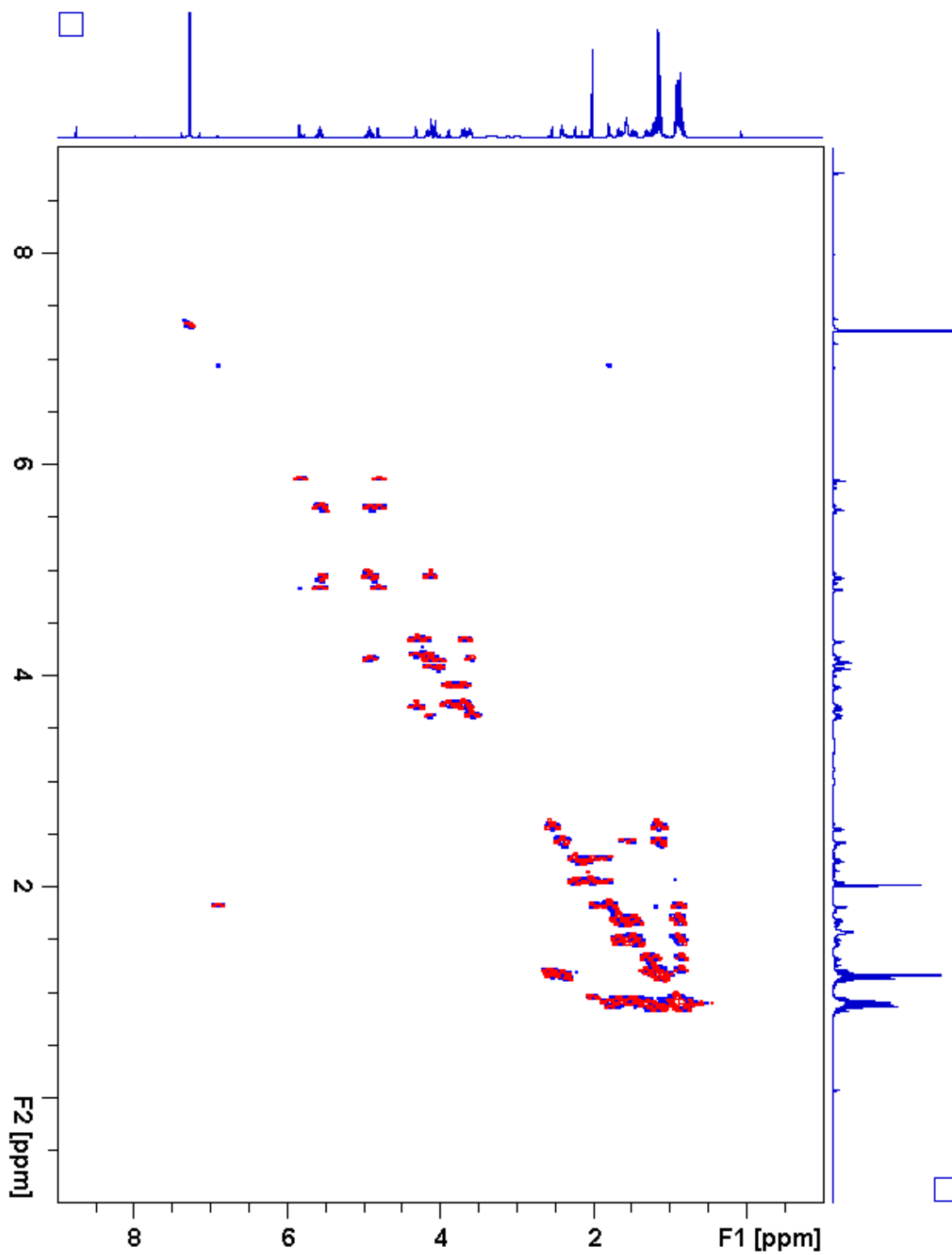
**Figure 2.54.** S4:20:0(4,5,5,6) $^{1}$H-$^{1}$H gCOSY
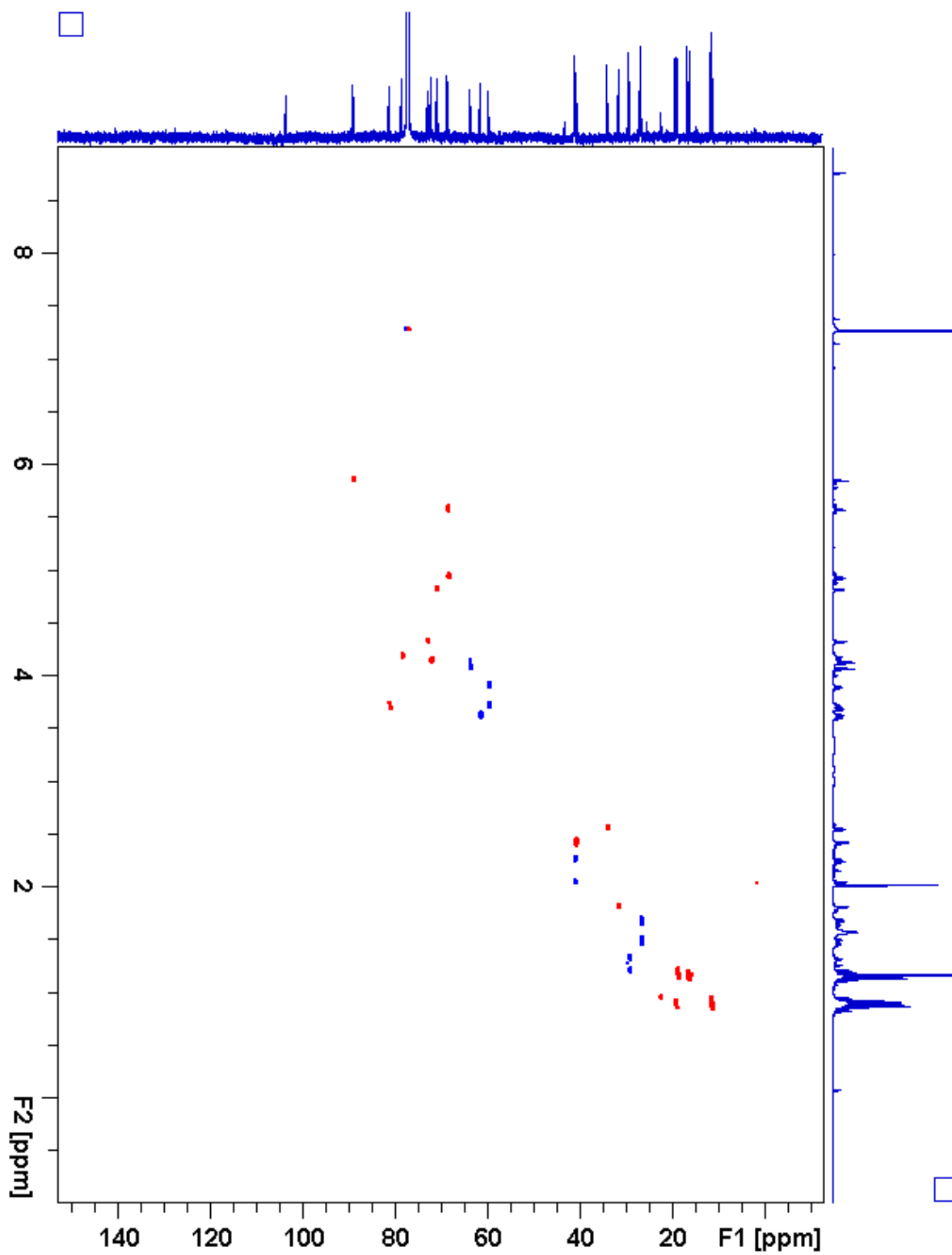
**Figure 2.55.** S4:20:0(4,5,5,6) gHSQC
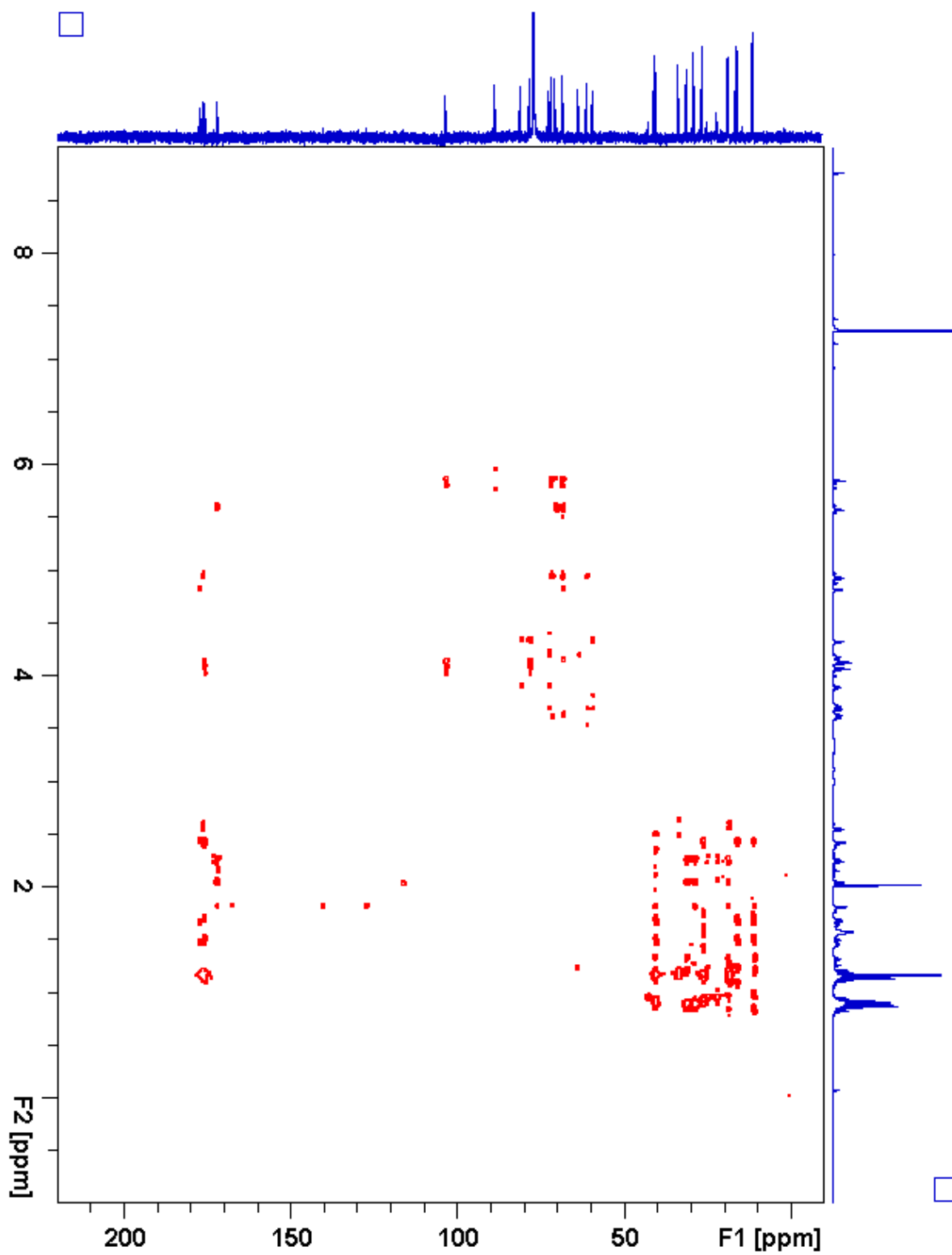
**Figure 2.56.** S4:20:0(4,5,5,6) gHMBC
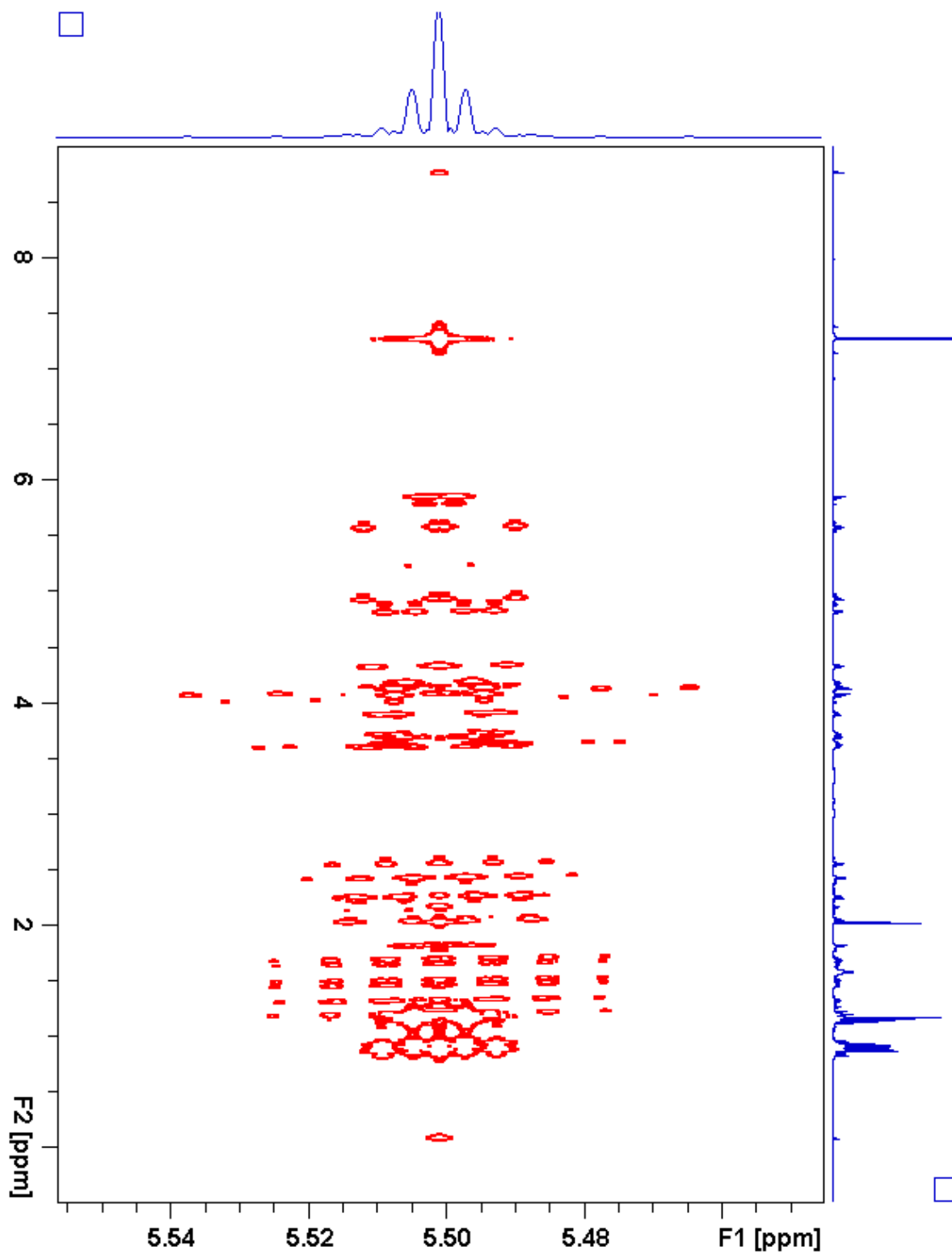
**Figure 2.57.** S4:20:0(4,5,5,6) *J*-resolved
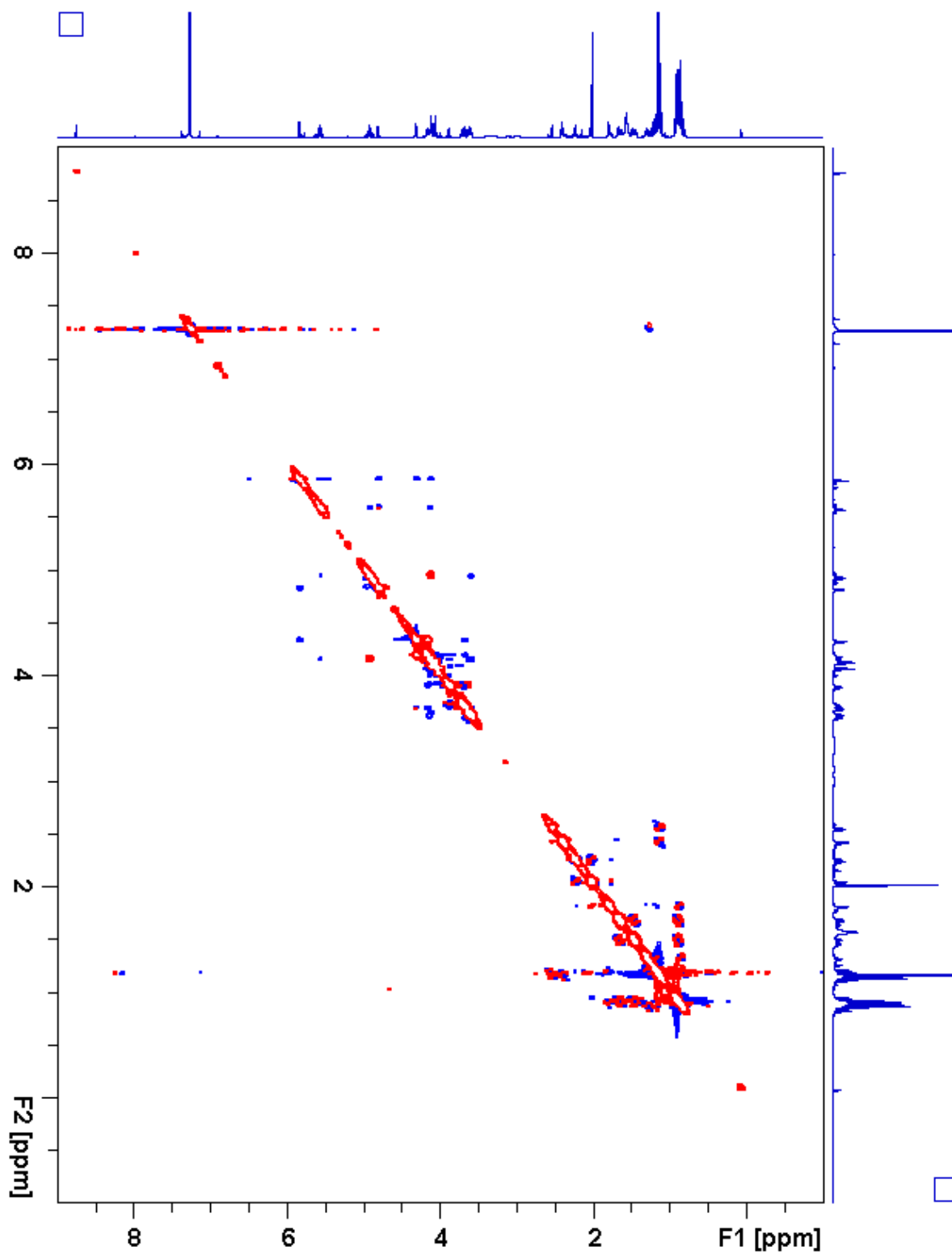
**Figure 2.58.** S4:20:0(4,5,5,6) ROESY

**Table 2.16.** S4:20:0(3,5,6,6) Chemical shifts and coupling constants
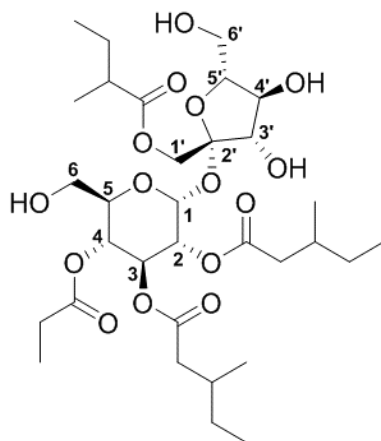


**Molecular Formula:** $C_{32}H_{54}O_{15}$

**110 min Retention Time (ESI+):** 51.42 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{32}H_{58}NO_{15}^{+}$ ([M+NH$_4$$^+$]): 696.3801, found: 696.3815

**Fraction, Batch:** #44, A-D

**Sample mass for NMR analysis:** 0.7 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** BYNYVKQGYIXALZ-ZWCYDAJKSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.78 (d, $J$ = 4.0 Hz) | 89.14 |
| **2**(CH) | 4.87 (dd, $J$ = 10.4, 4.0 Hz) | 70.56 |
| -1(CO) | | 173.62 |
| -2(CH$_2$) | 2.38 (dd, $J$ = 15.5, 5.7 Hz), 2.13 (dd, $J$ = 15.5, 8.3 Hz) | 41.17 |
| -3(CH) | 1.84 (m) | 31.73 |
| -4(CH$_3$) | 0.90 (d, $J$ = 6.7 Hz) | 19.37 |
| -5(CH$_2$) | 1.34 (m), 1.20 (m) | 29.3[a] |
| -6(CH$_3$) | 0.88 (t, $J$ = 7.4 Hz) | 11.4[b] |
| **3**(CH) | 5.53 (dd, $J$ = 10.6, 9.2 Hz) | 68.98 |
| -1(CO) | | 172.33 |
| -2(CH$_2$) | 2.25 (dd, $J$ = 15.0, 5.9 Hz), 2.04 (dd, $J$ = 15.0, 8.3 Hz) | 41.31 |
| -3(CH) | 1.80 (m) | 31.83 |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.28 |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.3[a] |
| -6(CH$_3$) | 0.86 (t, $J$ = 7.4 Hz) | 11.4[b] |
| **4**(CH) | 4.92 (dd, $J$ = 10.7, 9.3 Hz) | 68.82 |
| -1(CO) | | 173.85 |
| -2(CH$_2$) | 2.35 (m), 2.31 (m) | 27.57 |
| -3(CH$_3$) | 1.12 (t, $J$ = 7.6 Hz) | 9.09 |
| **5**(CH) | 4.16 (ddd, $J$ = 10.3, 5.9, 2.6 Hz) | 72.19 |
| **6**(CH$_2$) | 3.64 (dd, $J$ = 12.7, 2.4 Hz), 3.60 (dd, $J$ = 12.7, 6.3 Hz) | 61.57 |

**Table 2.16.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.08 (d, $J =$ 11.6 Hz), 4.06 (d, $J =$ 11.6 Hz) | 63.63 |
| -1(CO) | | 175.90 |
| -2(CH) | 2.41 (sextet, $J =$ 7.0 Hz) | 41.10 |
| -3(CH$_3$) | 1.15 (d, $J =$ 7.0 Hz) | 16.73 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.87 |
| -5(CH$_3$) | 0.91 (t, $J =$ 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.45 |
| **3'**(CH) | 4.17 (d, $J =$ 8.7 Hz) | 78.35 |
| **4'**(CH) | 4.31 (t, $J =$ 8.7 Hz) | 72.79 |
| **5'**(CH) | 3.71 (m) | 81.33 |
| **6'**(CH$_2$) | 3.89 (dd, $J =$ 13.7, 2.8 Hz), 3.71 (m) | 59.77 |

a - Two $^{13}$C signals not resolved in 2D spectra (29.32, 29.33 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (11.35, 11.36 ppm)
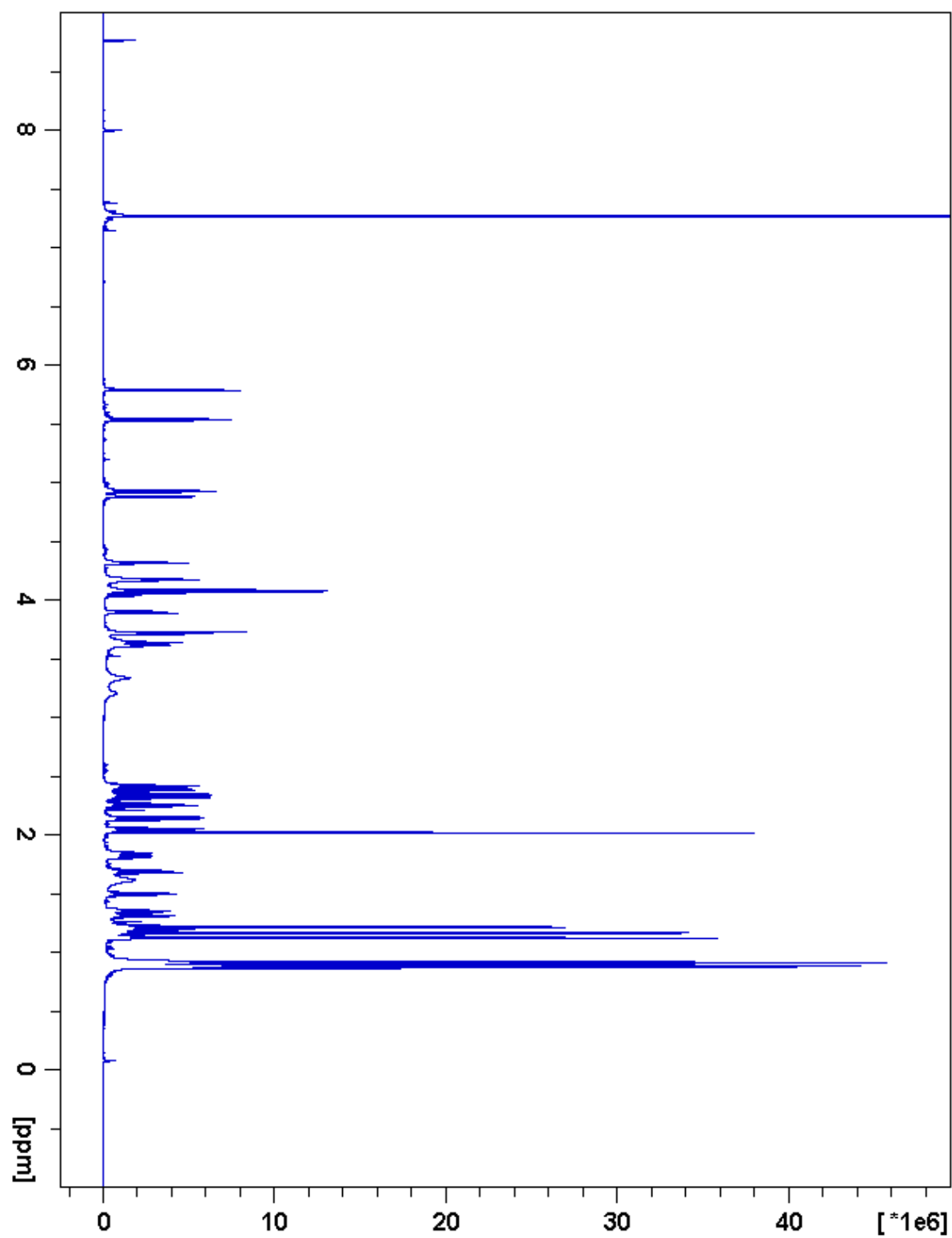
**Figure 2.59.** S4:20:0(3,5,6,6) $^1$H NMR

138
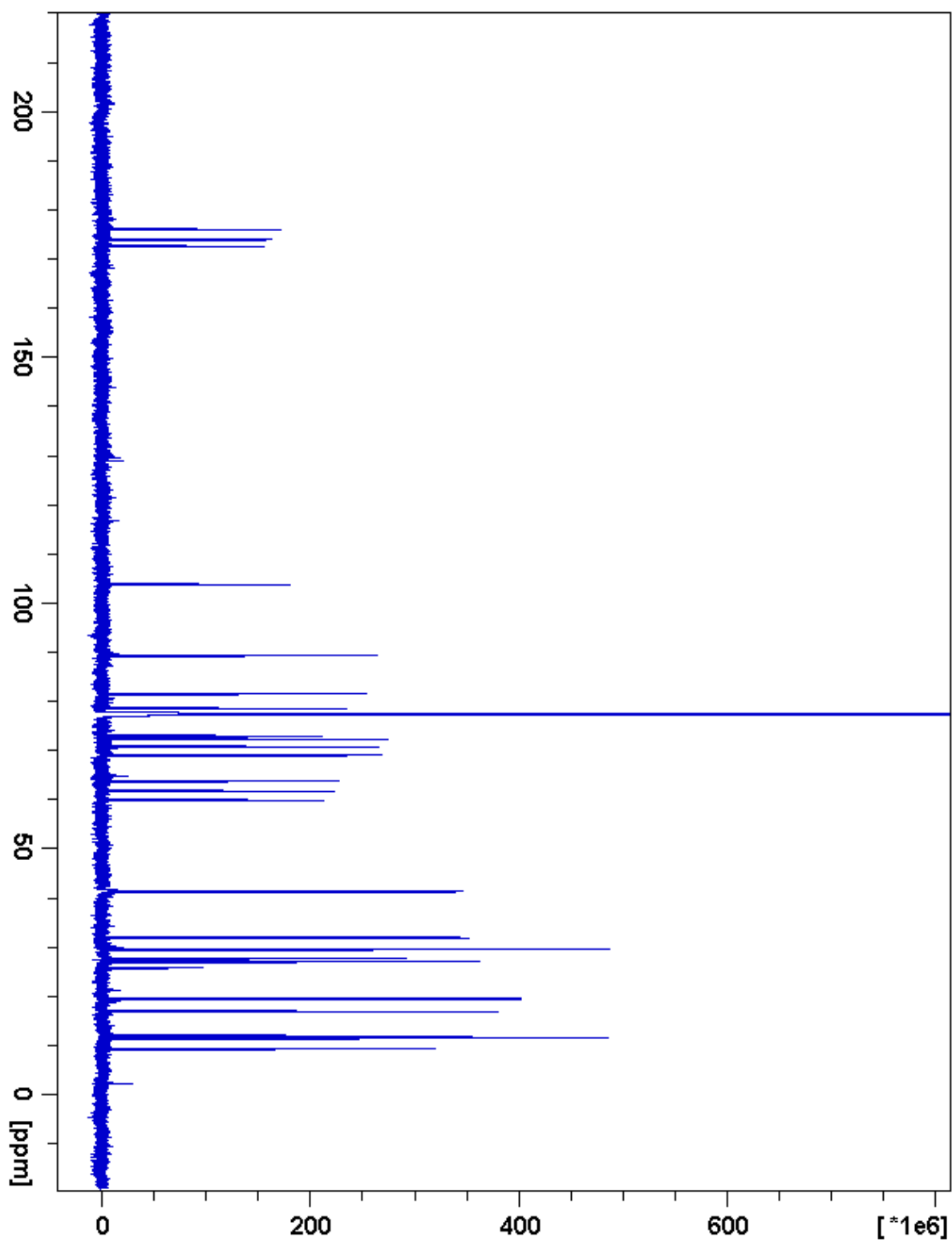
**Figure 2.60.** S4:20:0(3,5,6,6) $^{13}$C NMR
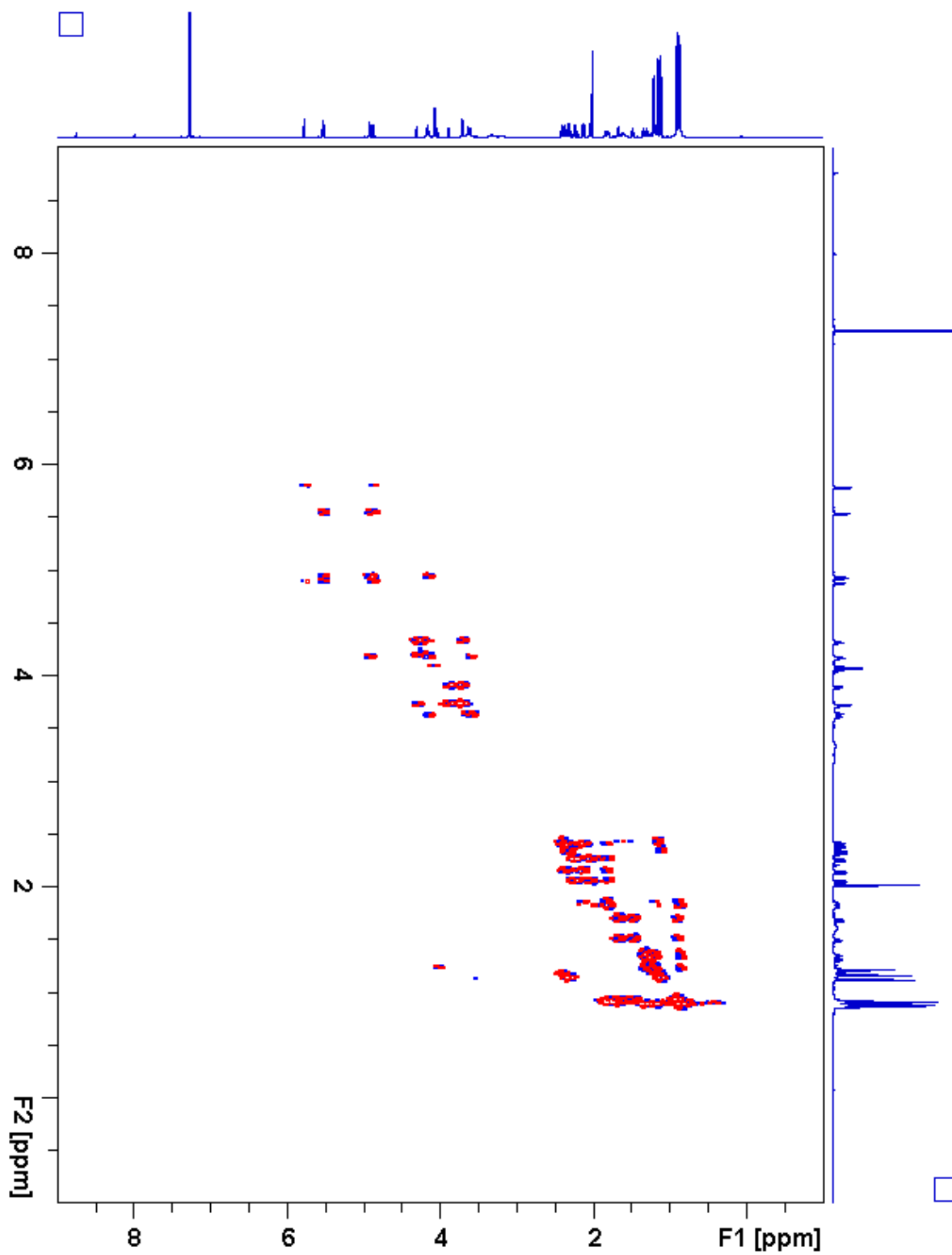
**Figure 2.61.** S4:20:0(3,5,6,6) $^1$H-$^1$H gCOSY
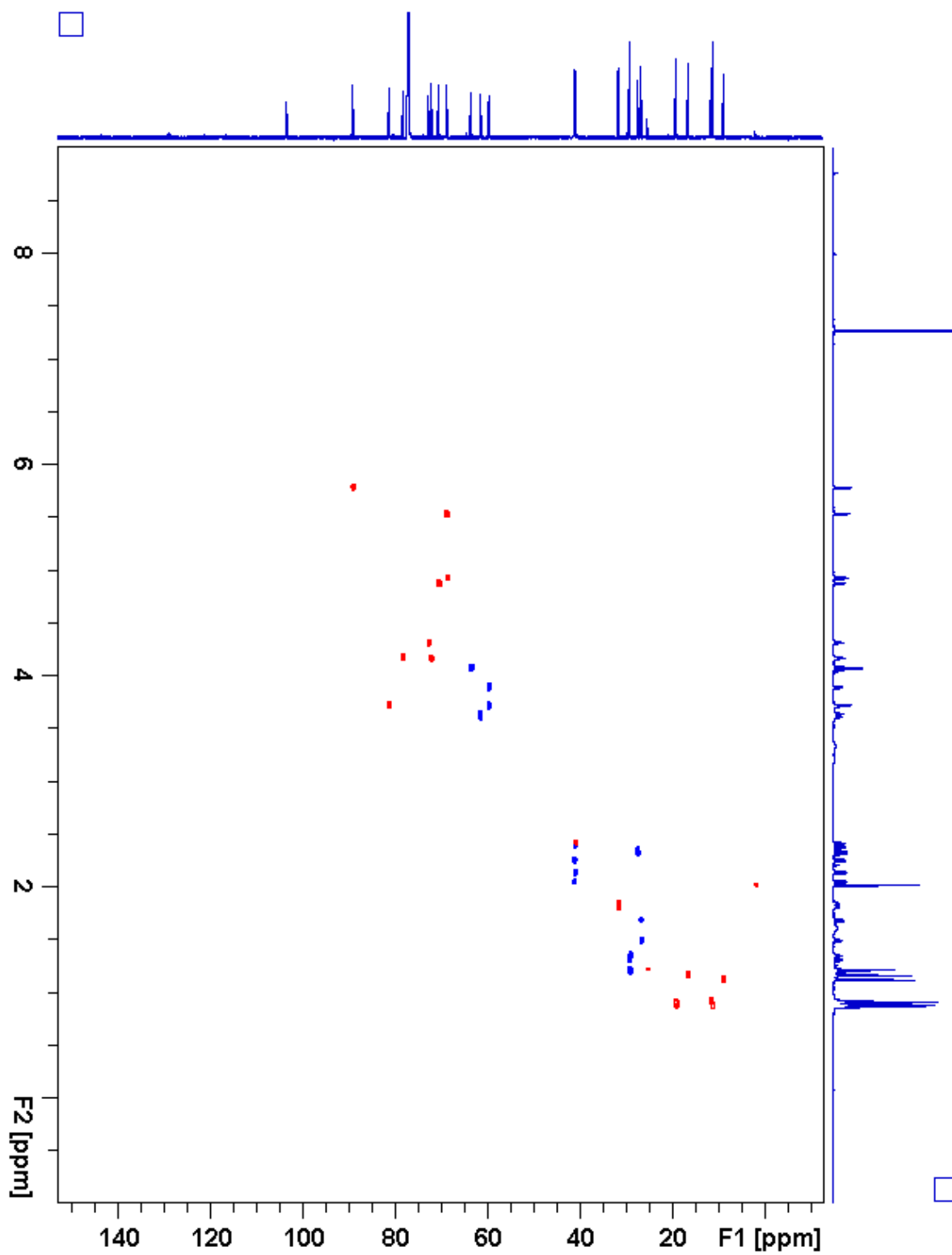
**Figure 2.62.** S4:20:0(3,5,6,6) gHSQC
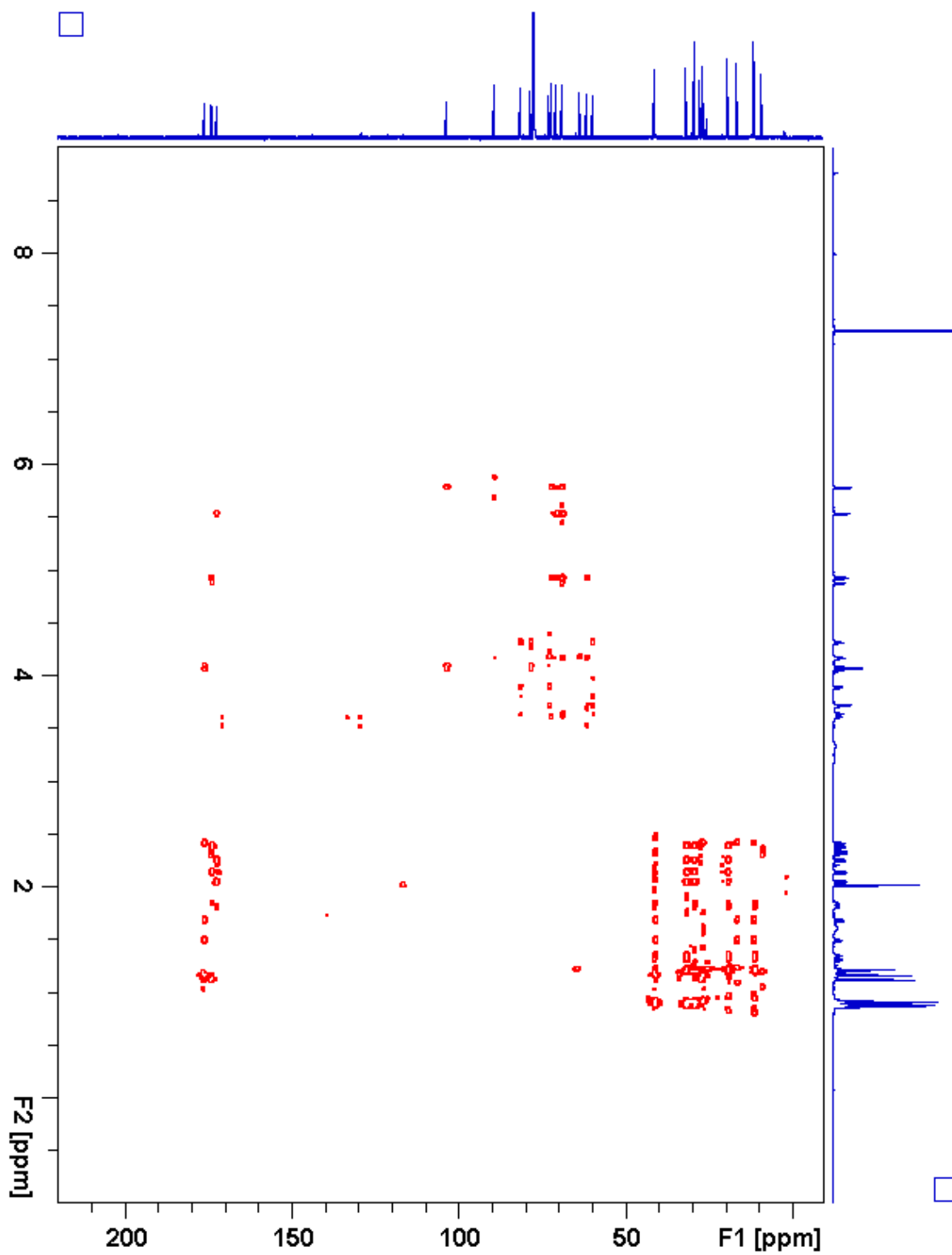
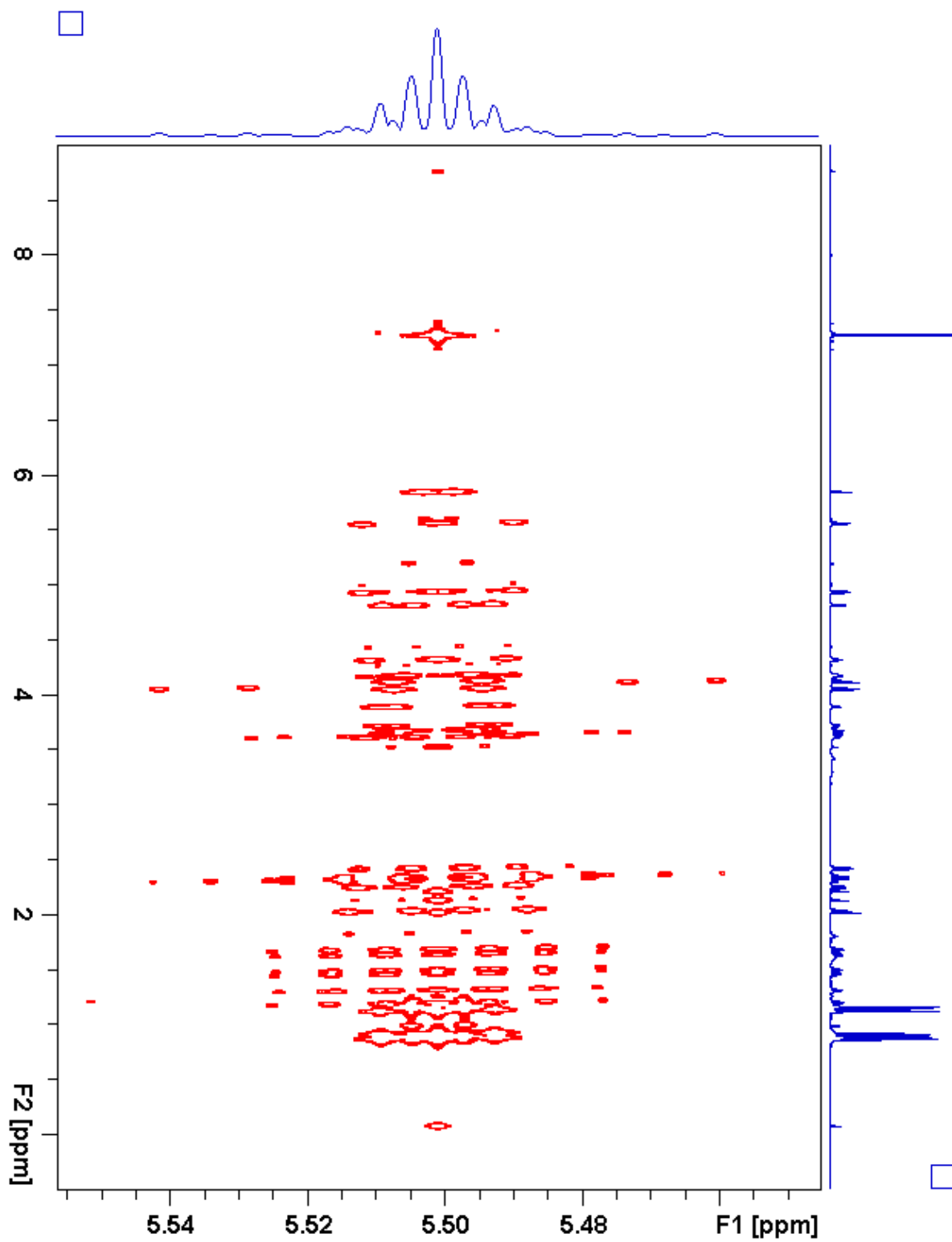**Figure 2.63.** S4:20:0(3,5,6,6) gHMBC
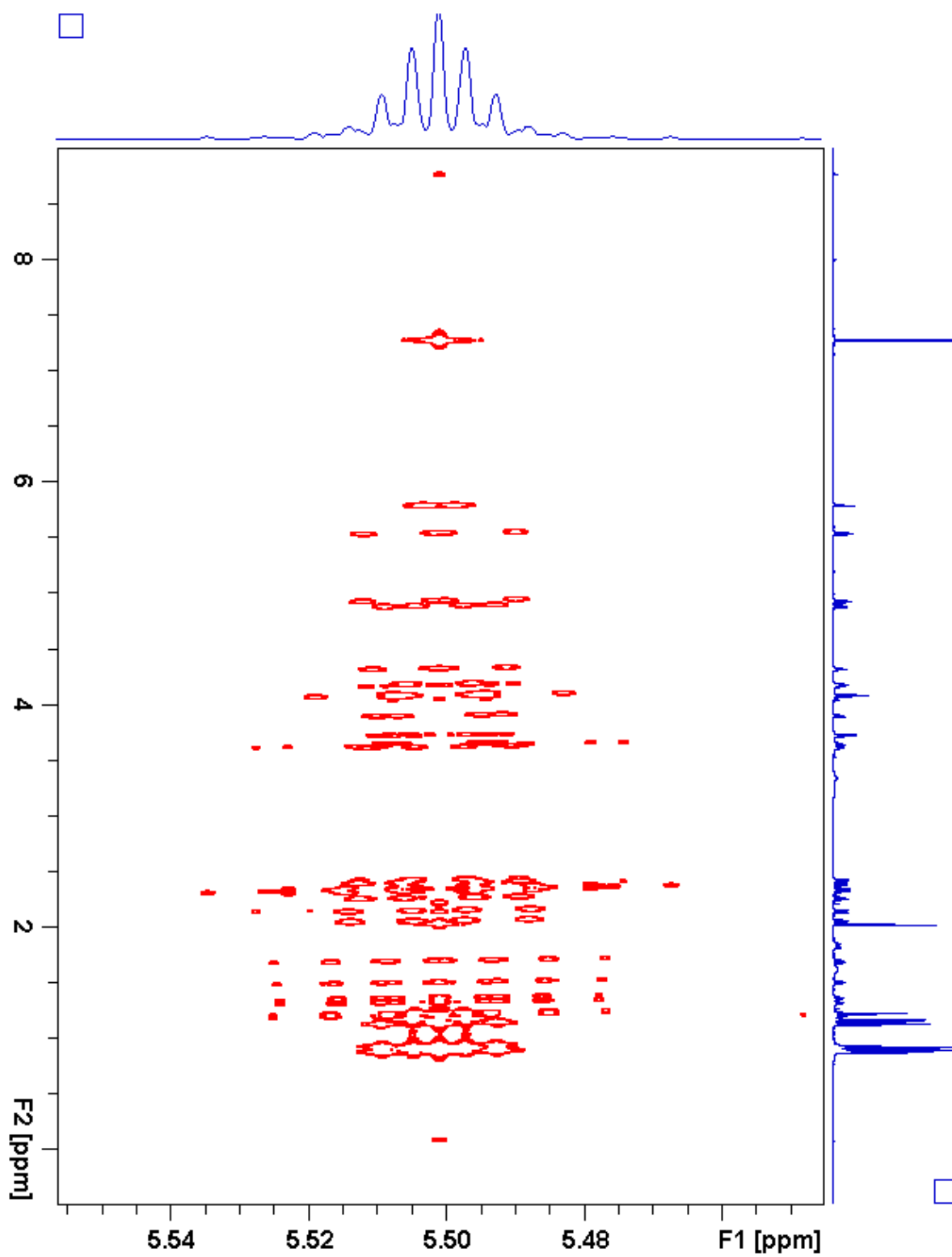
**Figure 2.64.** S4:20:0(3,5,6,6) *J*-resolved

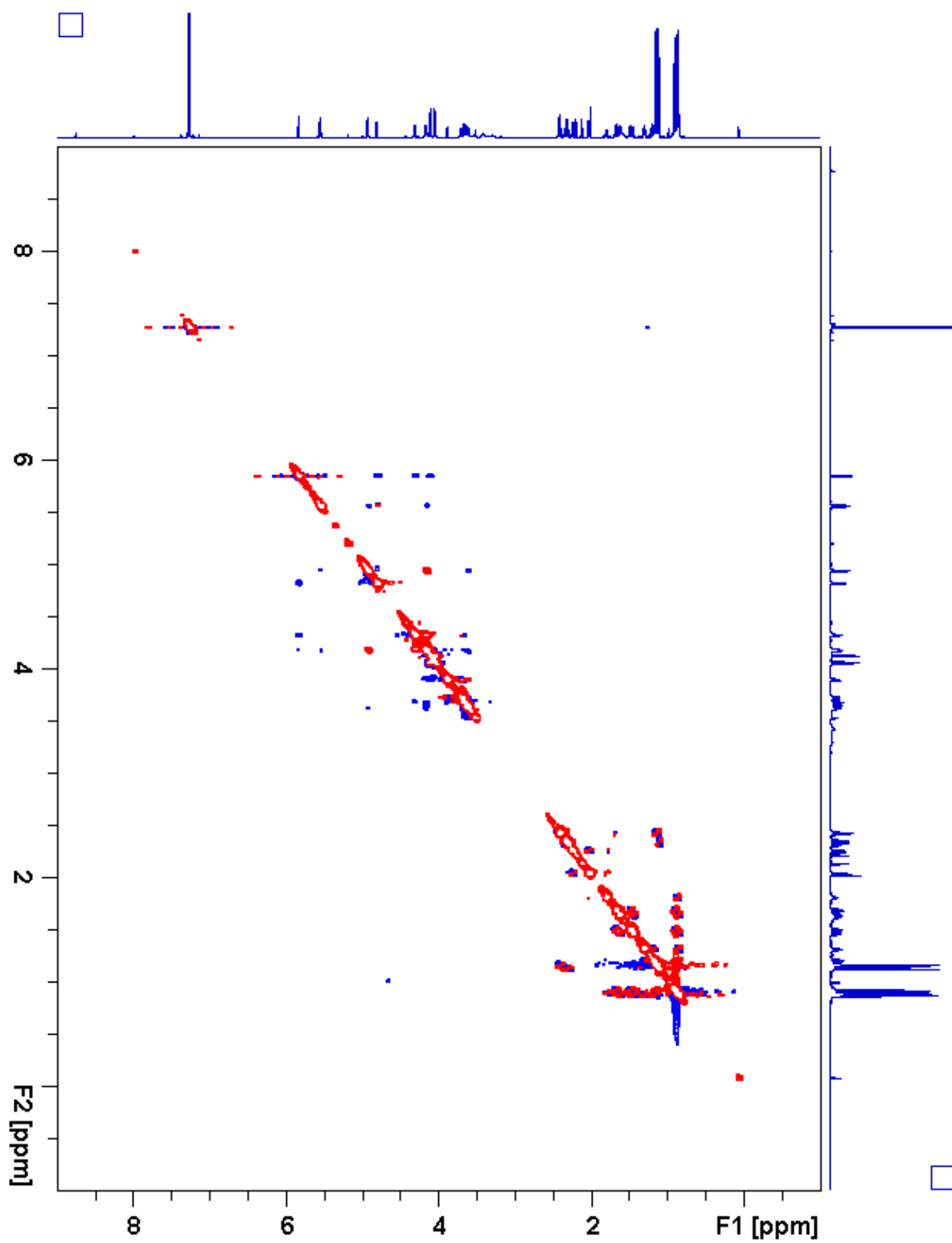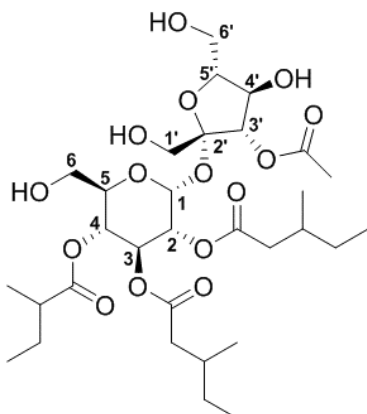**Figure 2.65.** S4:20:0(3,5,6,6) ROESY

**Table 2.17.** S4:20:0(2,6,6,6) Chemical shifts and coupling constants
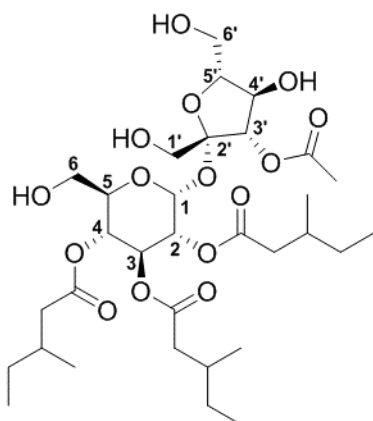


**Molecular Formula:** $C_{32}H_{54}O_{15}$

**110 min Retention Time (ESI+):** 51.65 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{32}H_{58}NO_{15}^{+}$ ([M+NH$_4^+$]): 696.3801, found: 696.3815

**Fraction, Batch:** #45, A-D

**Sample mass for NMR analysis:** 0.5 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** WLTSSQZLQMTYDB-MPGSVAJHSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.63 (d, $J = 3.6$ Hz) | 89.78 |
| **2**(CH) | 4.87 (dd, $J = 10.4, 3.7$ Hz) | 70.52 |
| -1(CO) | | 172.71[a] |
| -2(CH$_2$) | 2.31 (dd, $J = 15.7, 5.8$ Hz), 2.06 (dd, $J = 15.7, 8.3$ Hz) | 41.01 |
| -3(CH) | 1.82 (m) | 31.60 |
| -4(CH$_3$) | 0.89 (d, $J = 6.7$ Hz) | 19.4[b] |
| -5(CH$_2$) | 1.35 (m), 1.20 (m) | 29.34 |
| -6(CH$_3$) | 0.88 (t, $J = 7.4$ Hz) | 11.4[c] |
| **3**(CH) | 5.52 (dd, $J = 10.7, 9.2$ Hz) | 68.87 |
| -1(CO) | | 172.28 |
| -2(CH$_2$) | 2.24 (dd, $J = 15.3, 5.7$ Hz), 2.02 (dd, $J = 15.3, 8.3$ Hz) | 41.32 |
| -3(CH) | 1.79 (m) | 31.73 |
| -4(CH$_3$) | 0.88 (d, $J = 6.7$ Hz) | 19.4[b] |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.4[d] |
| -6(CH$_3$) | 0.86 (t, $J = 7.4$ Hz) | 11.4[c] |
| **4**(CH) | 4.92 (dd, $J = 10.6, 9.2$ Hz) | 68.71 |
| -1(CO) | | 172.71[a] |
| -2(CH$_2$) | 2.29 (dd, $J = 15.3, 5.8$ Hz), 2.11 (dd, $J = 15.3, 8.4$ Hz) | 41.21 |
| -3(CH) | 1.83 (m) | 31.81 |
| -4(CH$_3$) | 0.91 (d, $J = 6.7$ Hz) | 19.4[b] |
| -5(CH$_2$) | 1.32 (m), 1.20 (m) | 29.4[d] |
| -6(CH$_3$) | 0.88 (, $J = 7.4$ Hz) | 11.4[c] |
| **5**(CH) | 4.14 (ddd, $J = 10.3, 5.4, 2.8$ Hz) | 71.99 |
| **6**(CH$_2$) | 3.62 (m) | 61.91 |

**Table 2.17.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 3.66 (d, $J$ = 12.4 Hz), 3.54 (d, $J$ = 12.4 Hz) | 64.59 |
| **2'**(C) | | 104.16 |
| **3'**(CH) | 5.16 (d, $J$ = 7.9 Hz) | 80.31 |
| -1(CO) | | 172.38 |
| -2(CH$_3$) | 2.27 (s) | 21.02 |
| **4'**(CH) | 4.61 (t, $J$ = 8.2 Hz) | 71.52 |
| **5'**(CH) | 3.88 (dt, $J$ = 8.6, 2.4 Hz) | 82.32 |
| **6'**(CH$_2$) | 3.89 (d, $J$ = 13.4, 2.5 Hz), 3.71 (d, $J$ = 13.4, 2.8 Hz) | 59.76 |

a - Two $^{13}$C signals overlapping

b - Three $^{13}$C signals not resolved in 2D spectra (19.36, 19.39, 19.43 ppm)

c - Three $^{13}$C signals not resolved in 2D spectra (11.38, 11.39, 11.42 ppm)

d - Two $^{13}$C signals not resolved in 2D spectra ( 29.40, 29.42 ppm)
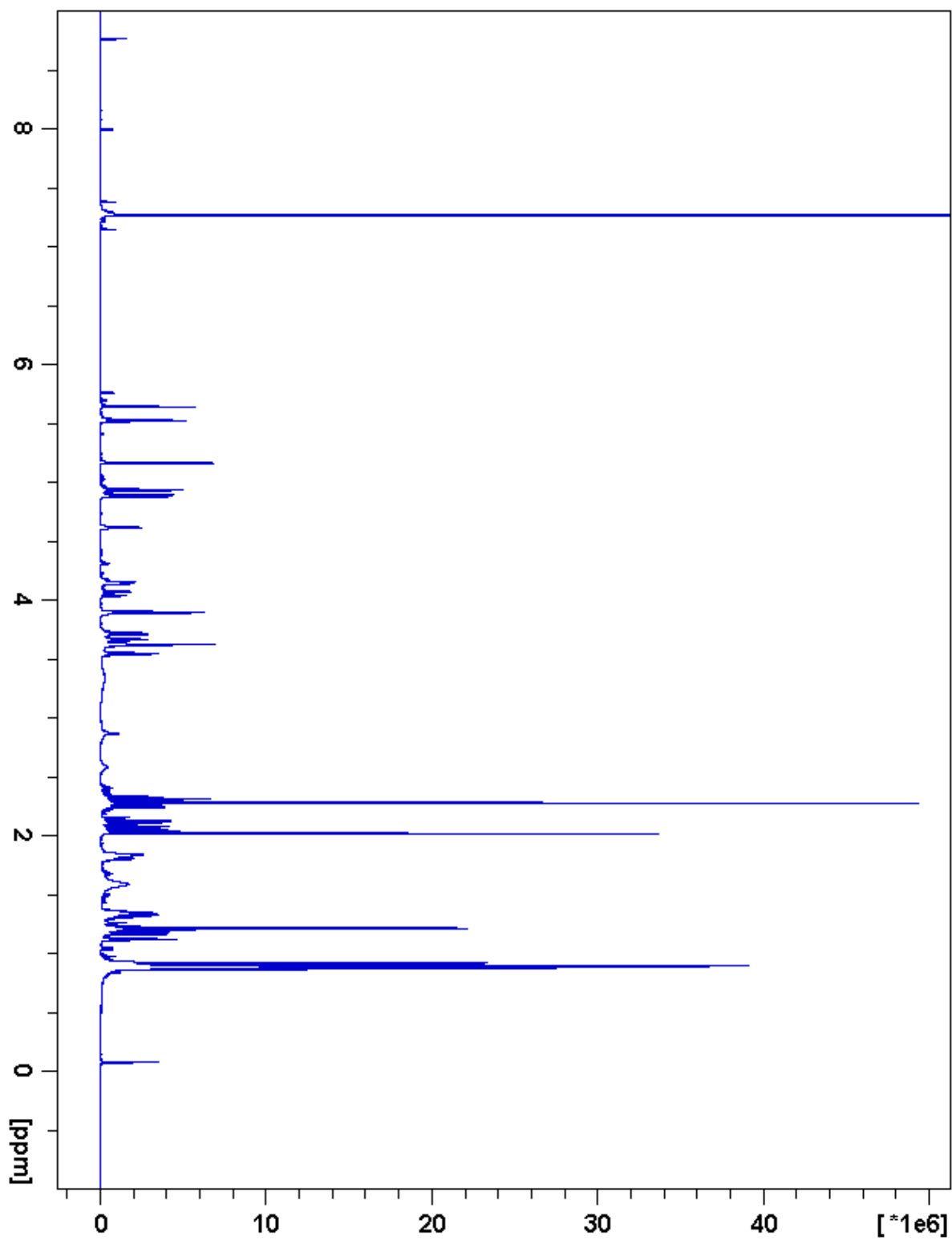
**Figure 2.66.** S5:20:0(2,6,6,6) $^1$H NMR
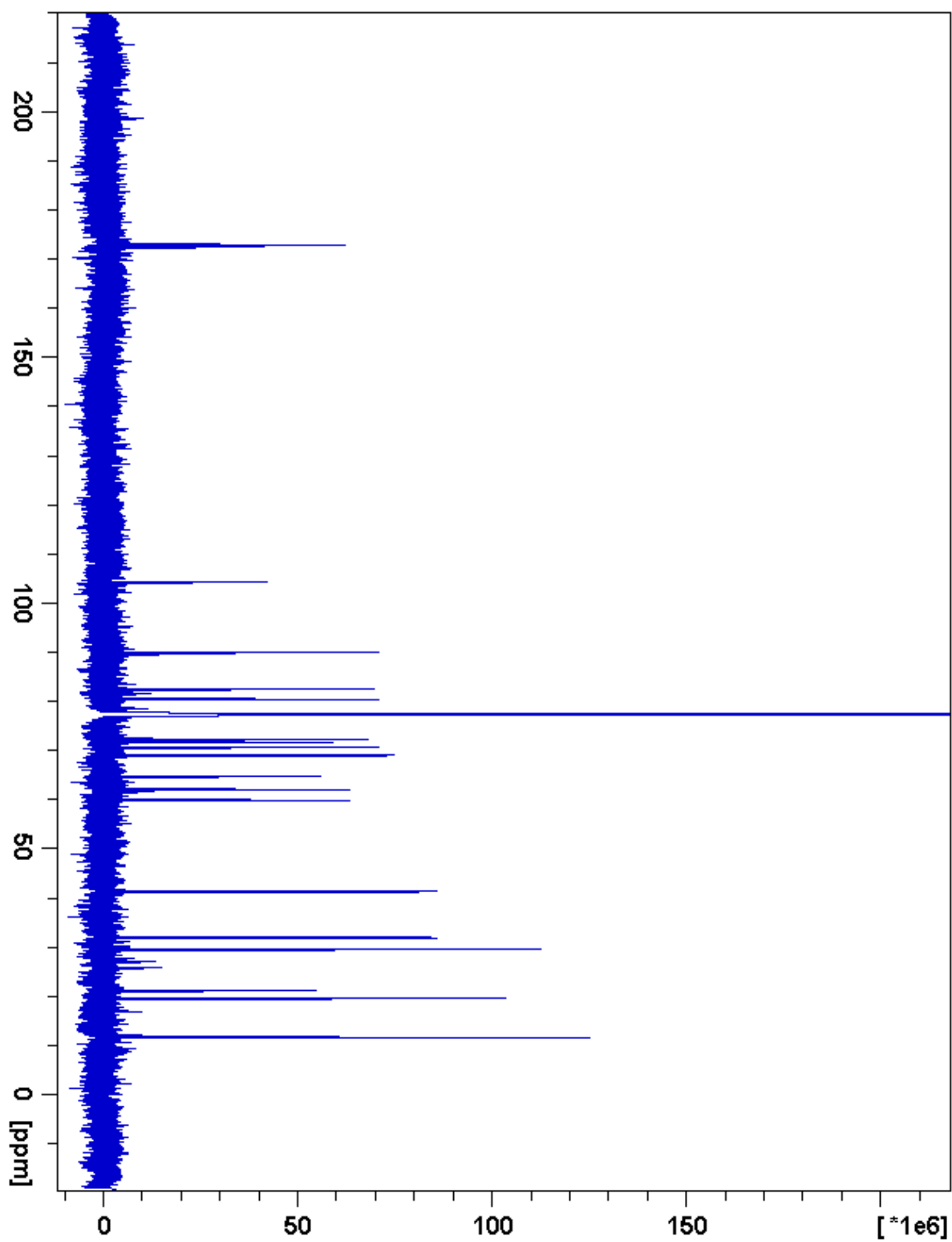
**Figure 2.67.** S5:20:0(2,6,6,6) $^{13}$C NMR

148
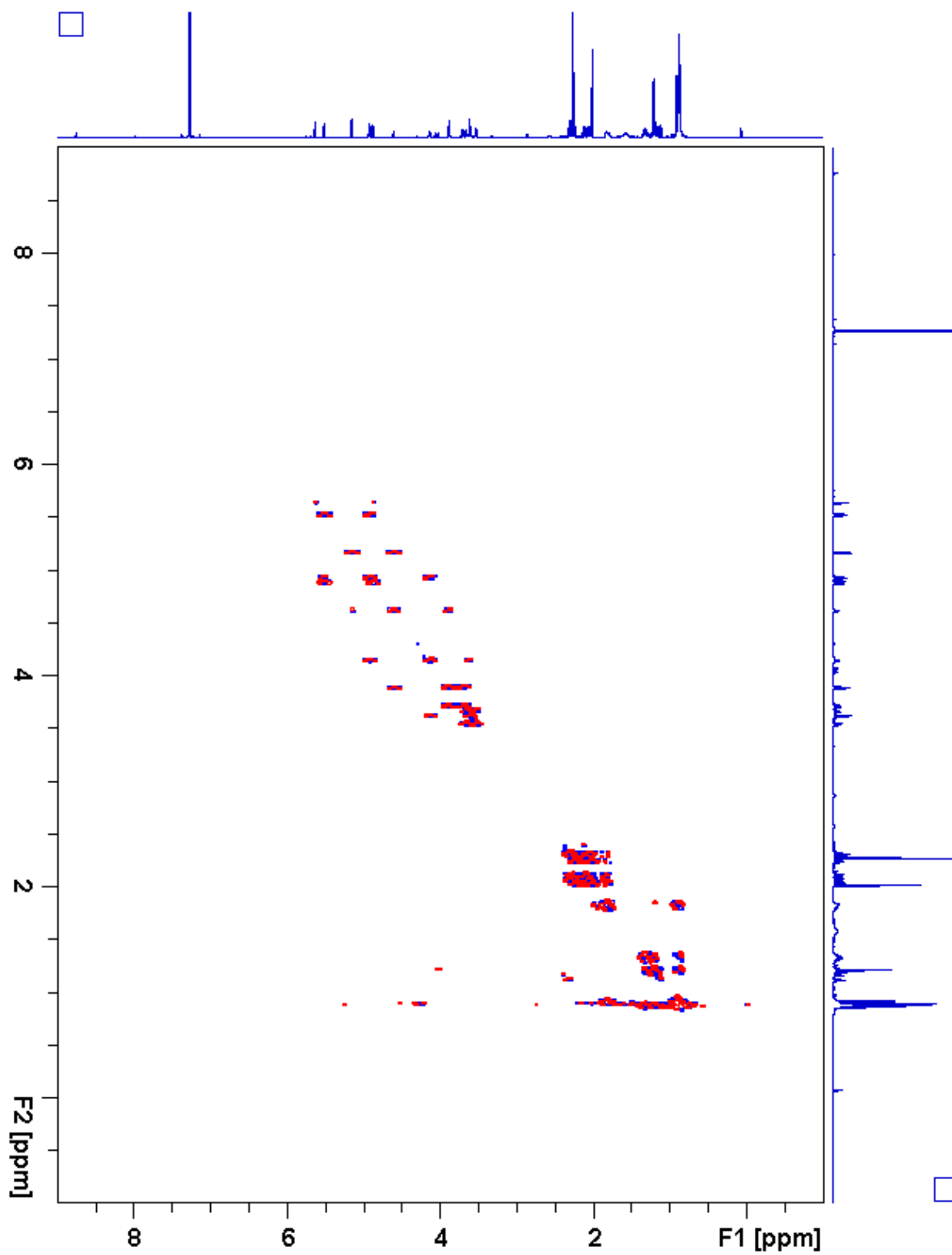
**Figure 2.68.** S5:20:0(2,6,6,6) $^1$H-$^1$H gCOSY
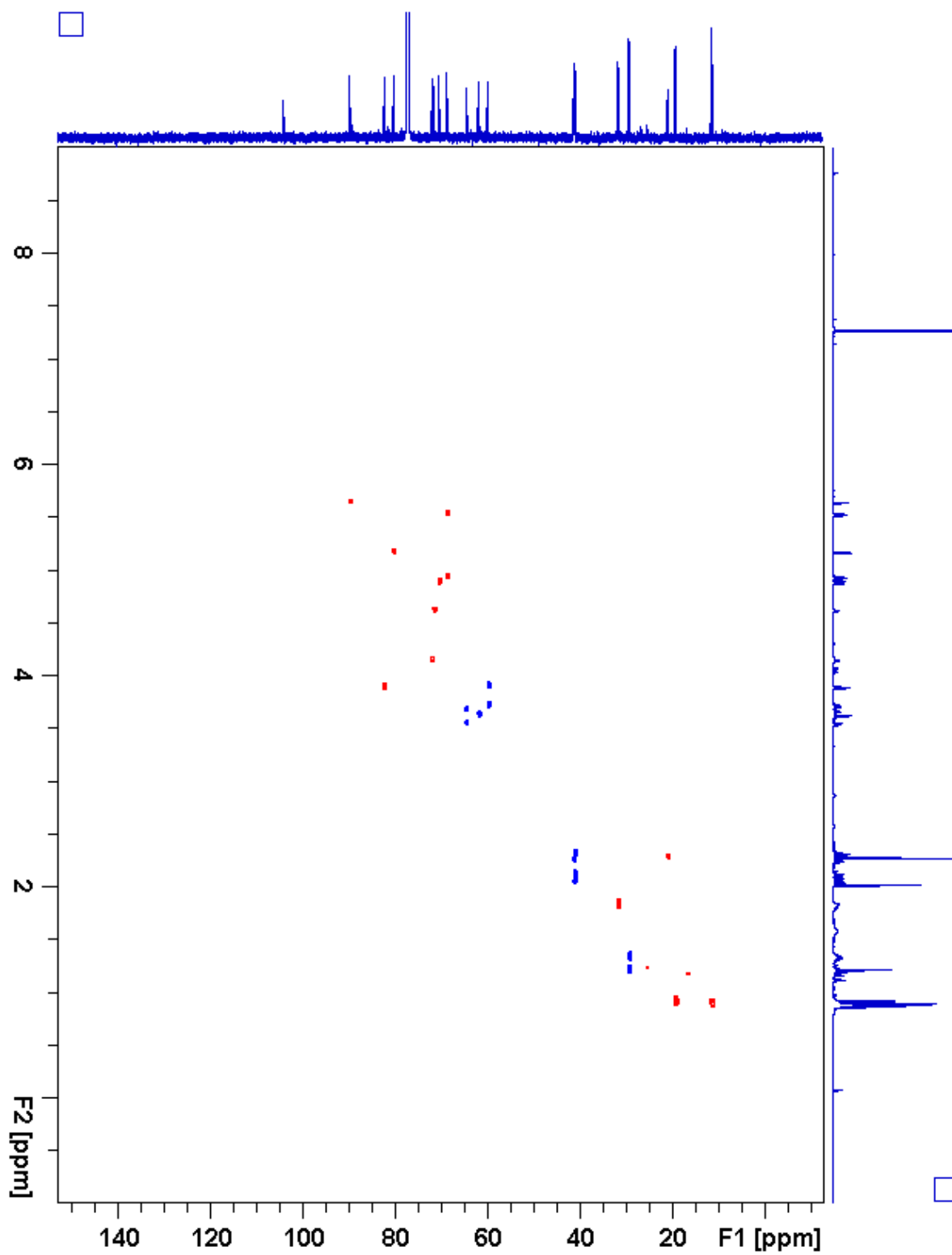
**Figure 2.69.** S5:20:0(2,6,6,6) gHSQC
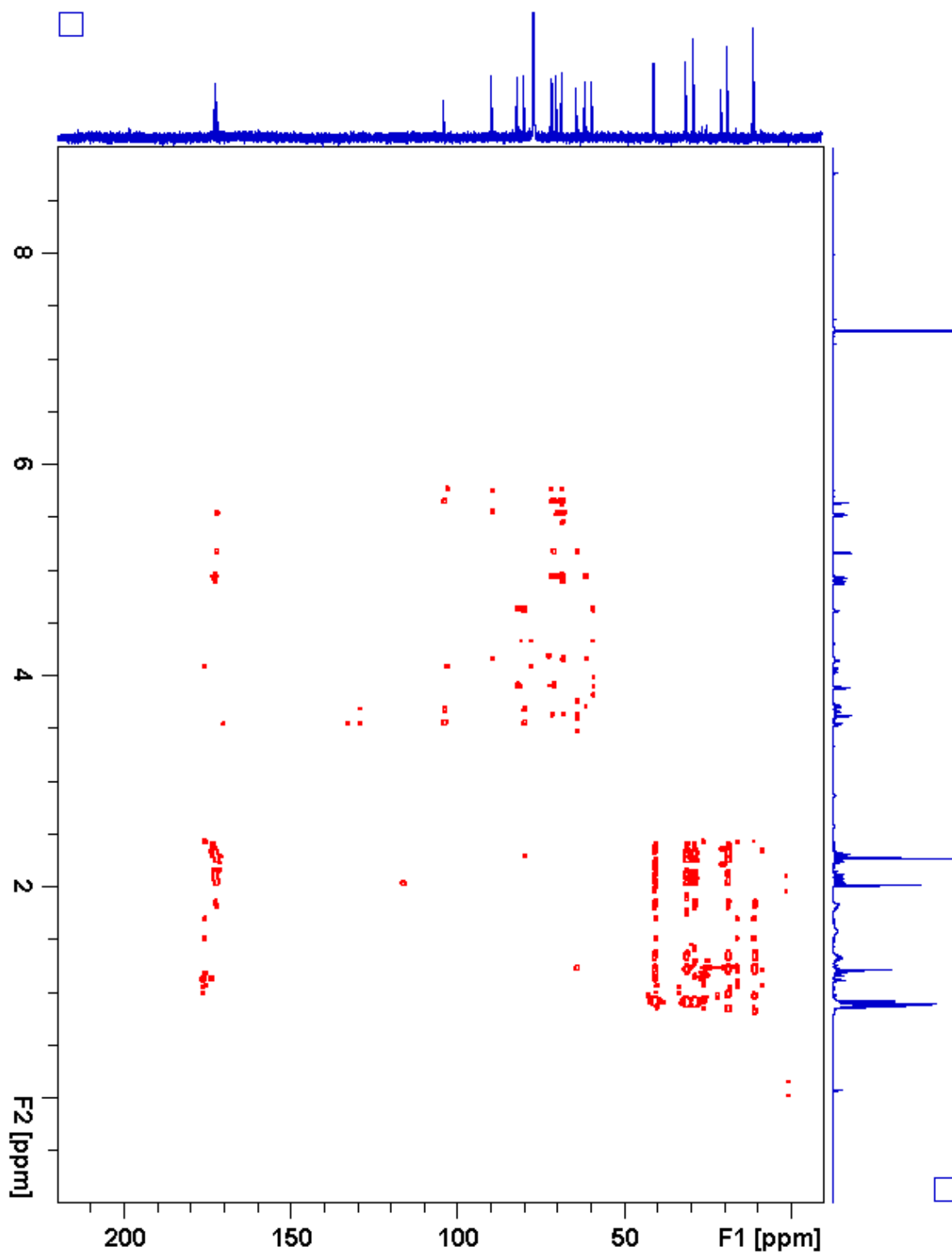
**Figure 2.70.** S5:20:0(2,6,6,6) gHMBC
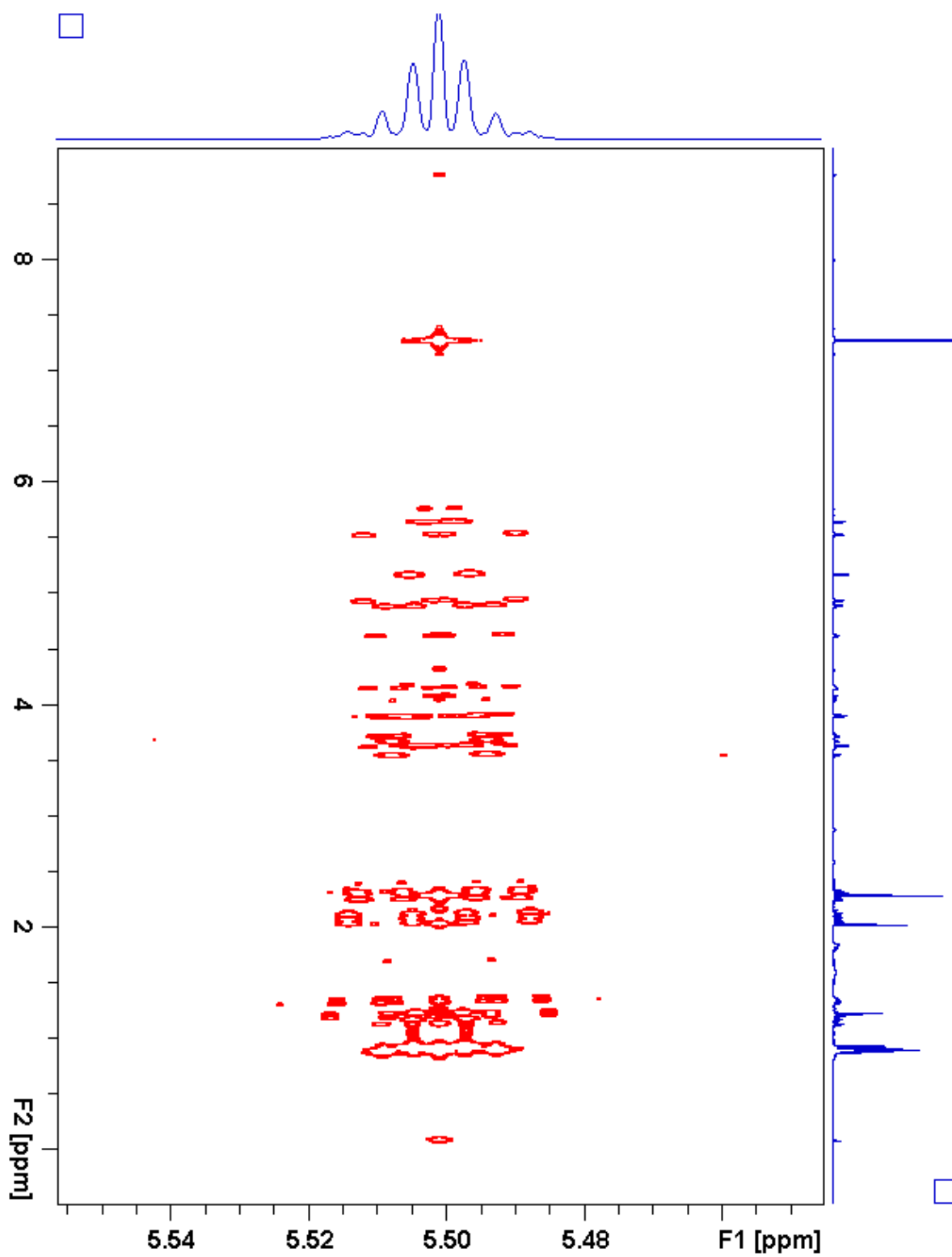
**Figure 2.71.** S5:20:0(2,6,6,6) *J*-resolved
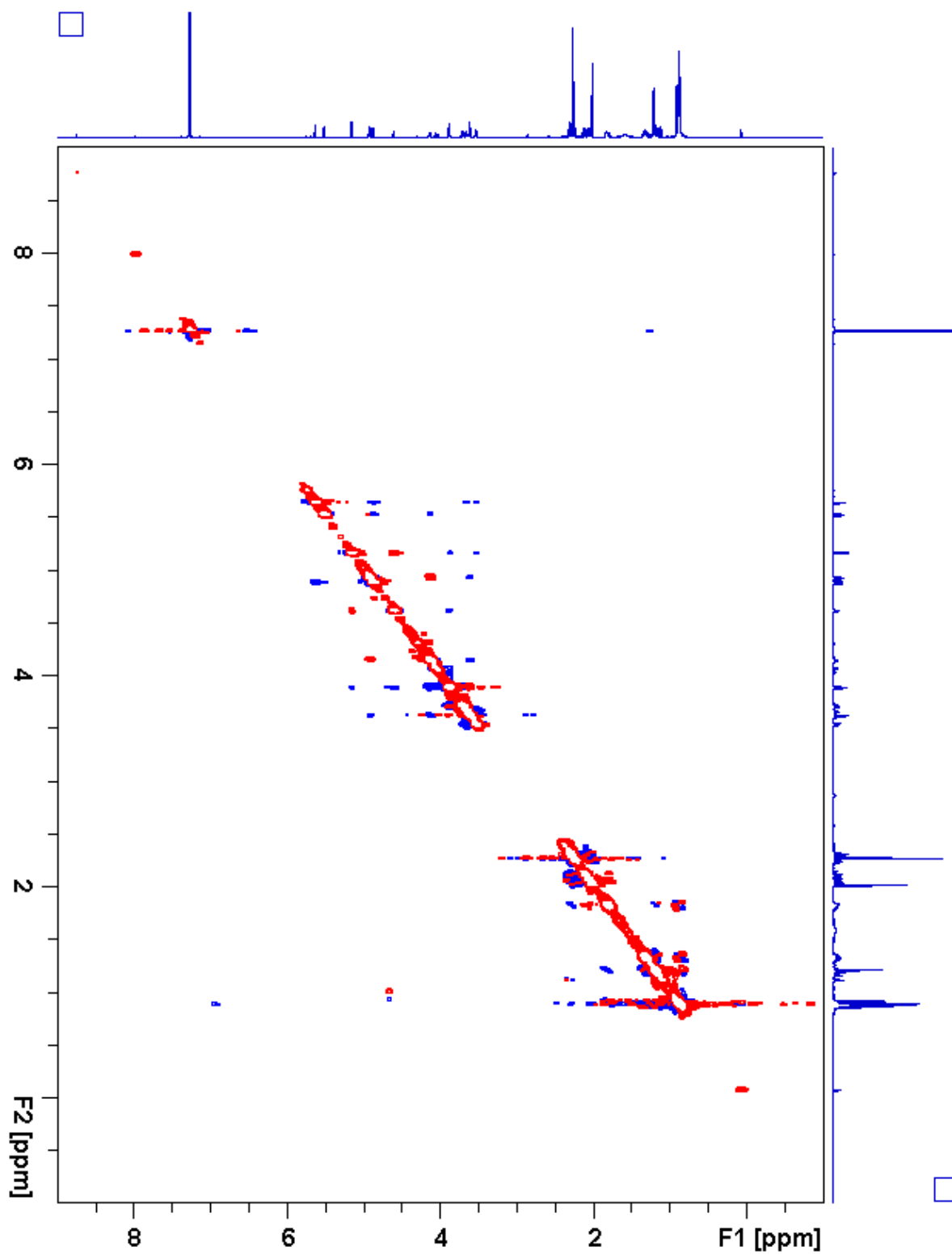
**Figure 2.72.** S5:20:0(2,6,6,6) ROESY

**Table 2.18.** S5:22:1(2,5,5,5,5$^T$) Chemical shifts and coupling constants
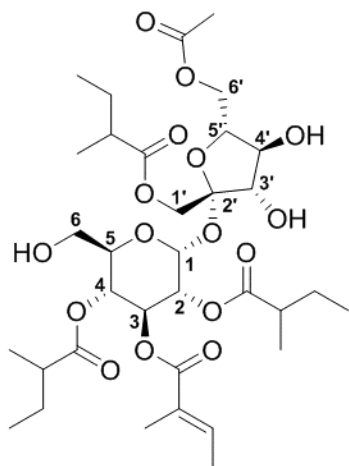


**Molecular Formula:** C$_{34}$H$_{54}$O$_{16}$

**110 min Retention Time (ESI+):** 52.55 mins
**HRMS:** (ESI+) *m/z* calculated for C$_{34}$H$_{58}$NO$_{16}^+$ ([M+NH$_4^+$]): 736.3750, found: 736.3743

**Fraction, Batch:** #47, D

**Sample mass for NMR analysis:** 0.2 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** AHDZAIMCFBHKJA-QZUIWNAASA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.65 (d, *J* = 3.9 Hz) | 89.99 |
| **2**(CH) | 5.02 (dd, *J* = 10.4, 3.9 Hz) | 70.06 |
| -1(CO) | | 175.85 |
| -2(CH) | 2.37 (sextet, *J* = 7.0 Hz) | 40.81 |
| -3(CH$_3$) | 1.08 (d, *J* = 7.0 Hz) | 16.25 |
| -4(CH$_2$) | 1.55 (m), 1.39 (m) | 26.77 |
| -5(CH$_3$) | 0.77 (t, *J* = 7.4 Hz) | 11.38 |
| **3**(CH) | 5.63 (dd, *J* = 10.6, 9.3 Hz) | 69.12 |
| -1(CO) | | 166.85 |
| -2(C) | | 127.73 |
| -3(CH$_3$) | 1.75 (m) | 12.08 |
| -4(CH) | 6.80 (m) | 139.16 |
| -5(CH$_3$) | 1.75 (m) | 14.63 |
| **4**(CH) | 5.03 (dd, *J* = 10.6, 9.3 Hz) | 68.73 |
| -1(CO) | | 176.43 |
| -2(CH) | 2.32 (sextet, *J* = 7.0 Hz) | 41.43 |
| -3(CH$_3$) | 1.04 (d, *J* = 7.0 Hz) | 16.87 |
| -4(CH$_2$) | 1.61 (m), 1.41(m) | 26.62 |
| -5(CH$_3$) | 0.84 (t, *J* = 7.4 Hz) | 11.88 |
| **5**(CH) | 4.15 (m) | 71.49 |
| **6**(CH$_2$) | 3.68 (dd, *J* = 12.7, 3.1 Hz), 3.61 (dd, *J* = 12.7, 5.6 Hz) | 61.58 |

**Table 2.18.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.21 (d, $J$ = 11.8 Hz), 4.17 (d, $J$ = 11.8 Hz) | 63.39 |
| -1(CO) | | 175.96 |
| -2(CH) | 2.41 (sextet, $J$ = 7.0 Hz) | 41.11 |
| -3(CH$_3$) | 1.16 (d, $J$ = 7.0 Hz) | 16.69 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH$_3$) | 0.92 (t, $J$ = 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.58 |
| **3'**(CH) | 4.15 (m) | 78.31 |
| **4'**(CH) | 4.11 (td, J = 8.4, 3.3 Hz) | 75.33 |
| **5'**(CH) | 3.90 (ddd, $J$ = 8.4, 5.3, 3.1 Hz) | 78.79 |
| **6'**(CH$_2$) | 4.42 (dd, $J$ = 12.3, 5.2 Hz), 4.23 (dd, $J$ = 12.3, 3.0 Hz) | 63.62 |
| -1(CO) | | 171.69 |
| -2(CH$_3$) | 2.15 (s) | 21.01 |
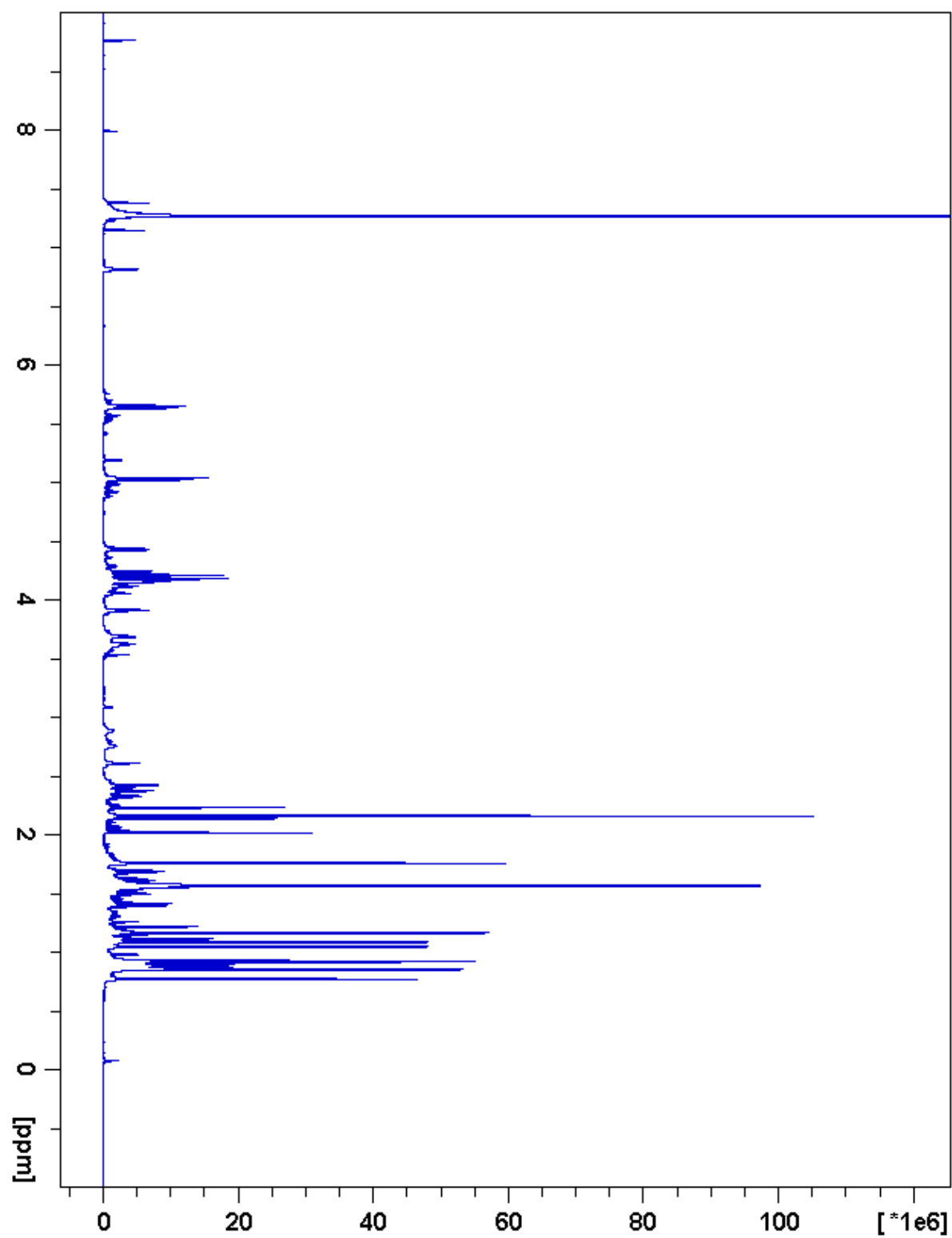
**Figure 2.73.** S5:22:1(2,5,5,5,5$^{\text{T}}$) $^1$H NMR
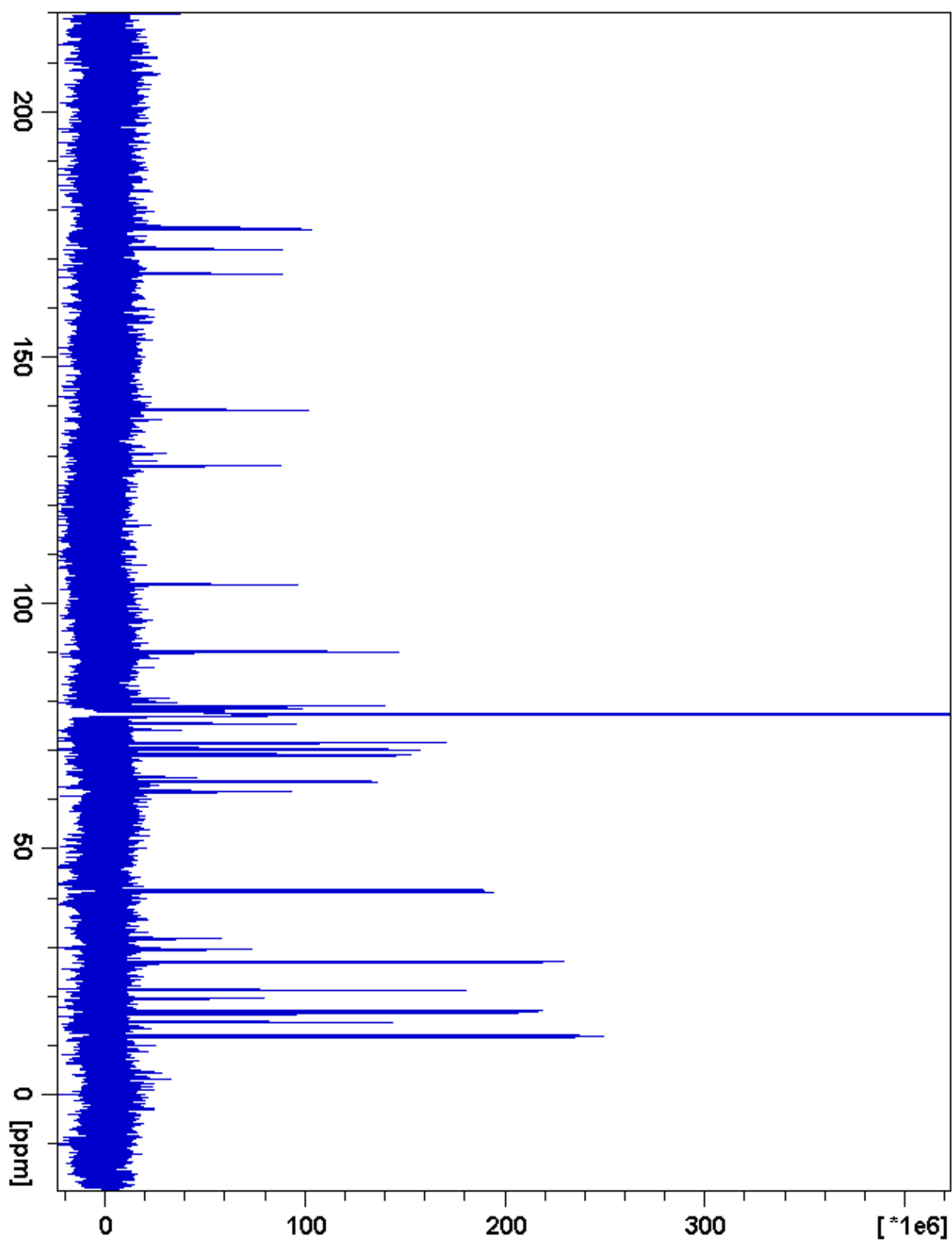
**Figure 2.74.** S5:22:1(2,5,5,5,5$^T$) $^{13}$C NMR
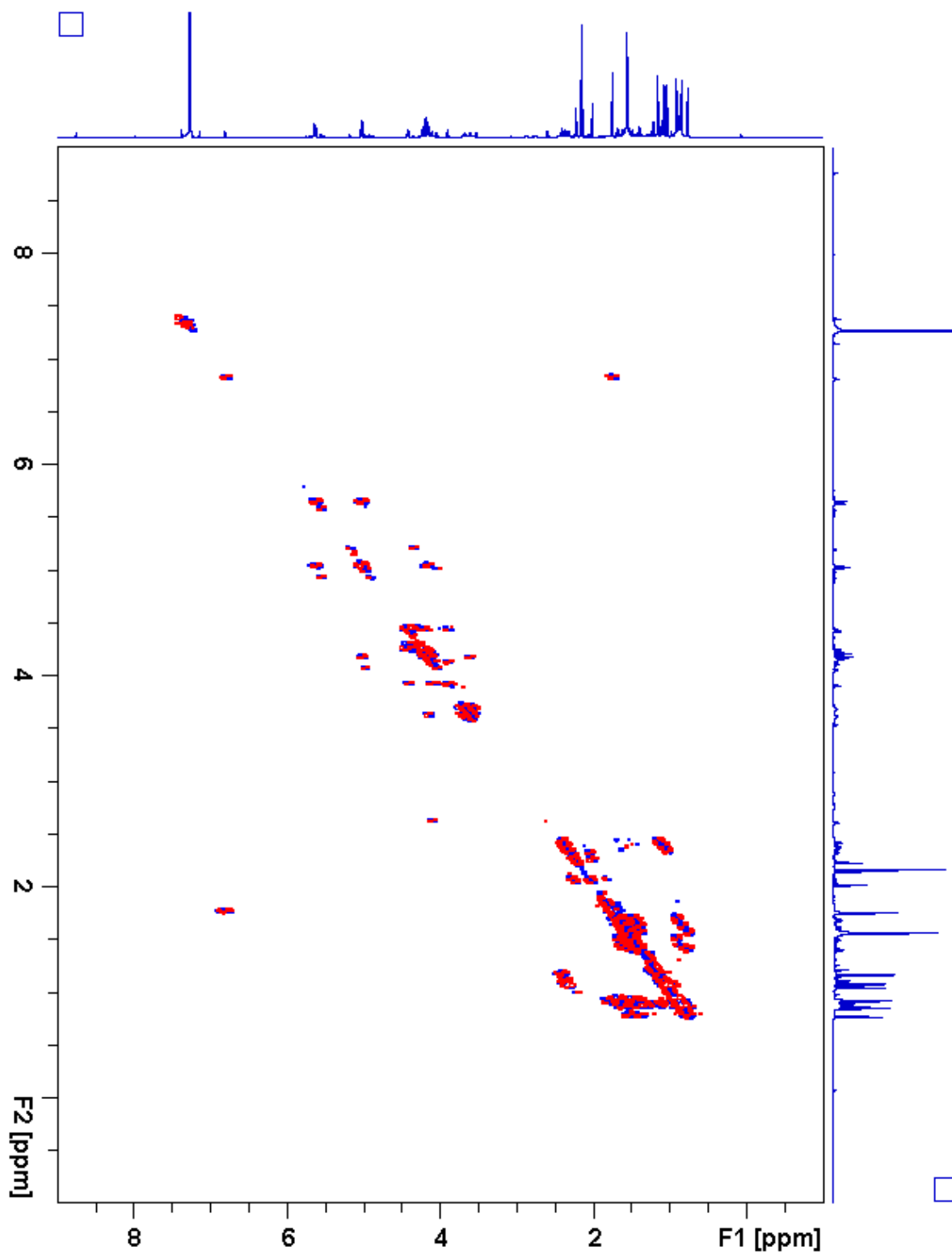
**Figure 2.75.** S5:22:1(2,5,5,5,5$^T$) $^1$H-$^1$H gCOSY
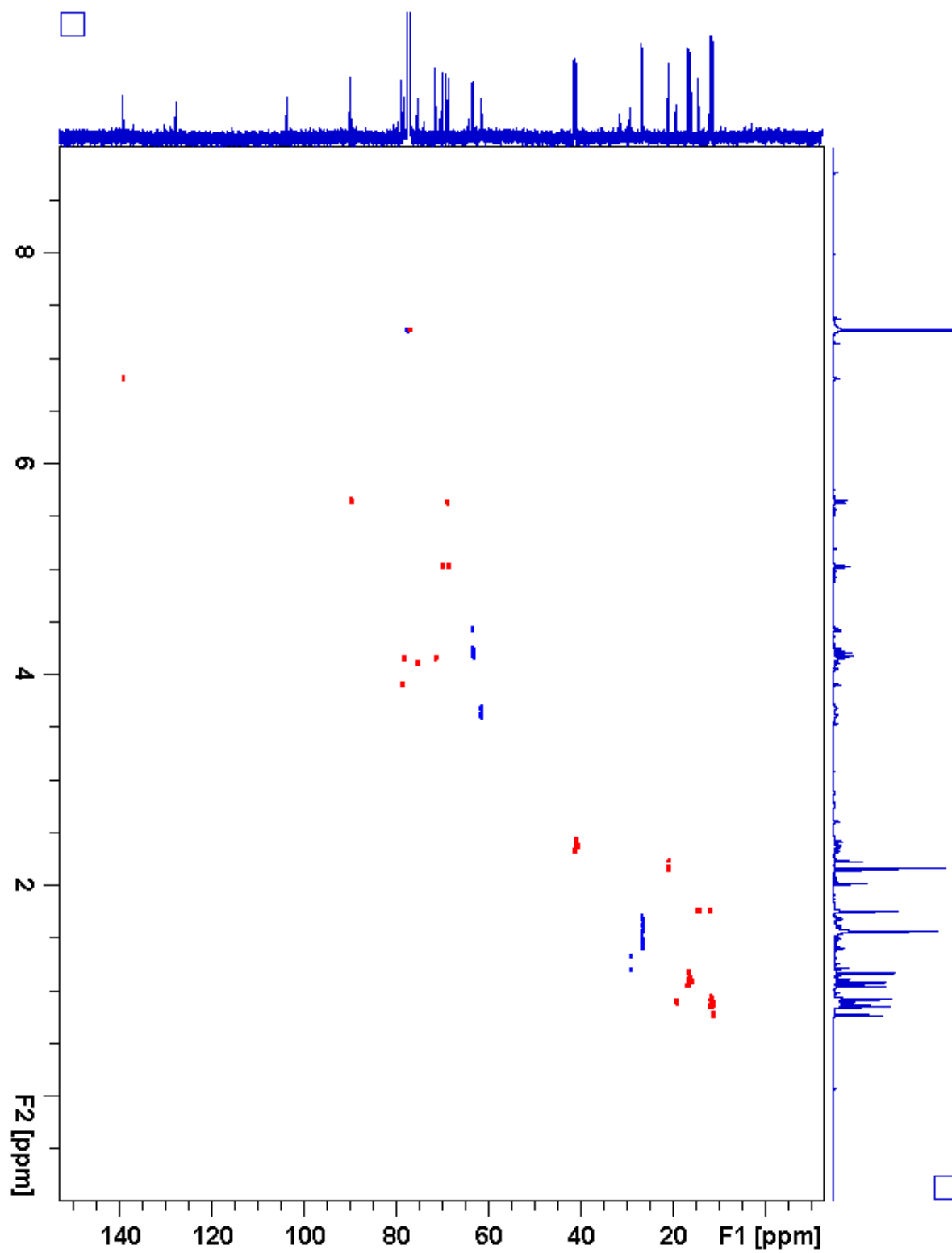
**Figure 2.76.** S5:22:1(2,5,5,5,5$^T$) gHSQC
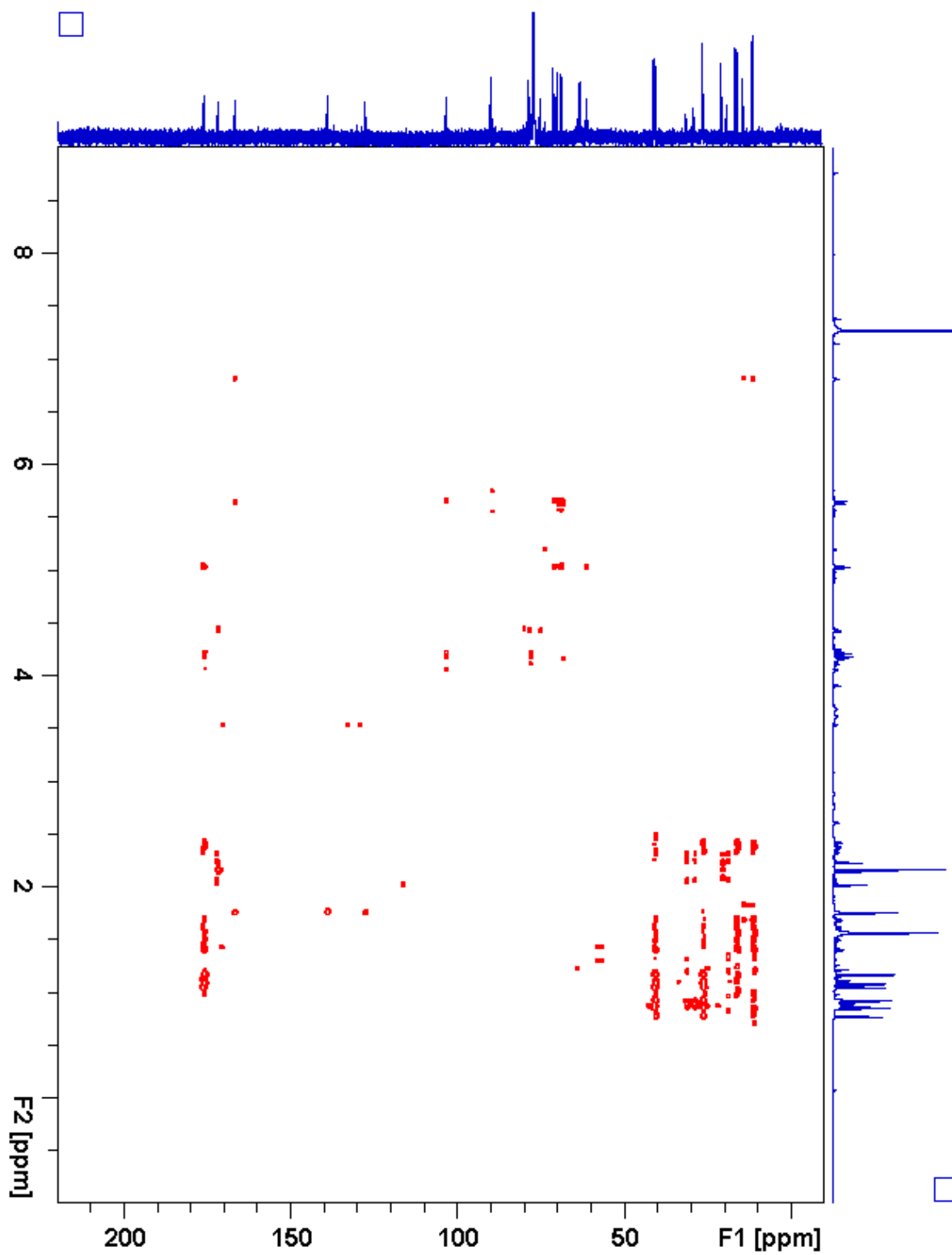
**Figure 2.77.** S5:22:1(2,5,5,5,5$^{T}$) gHMBC
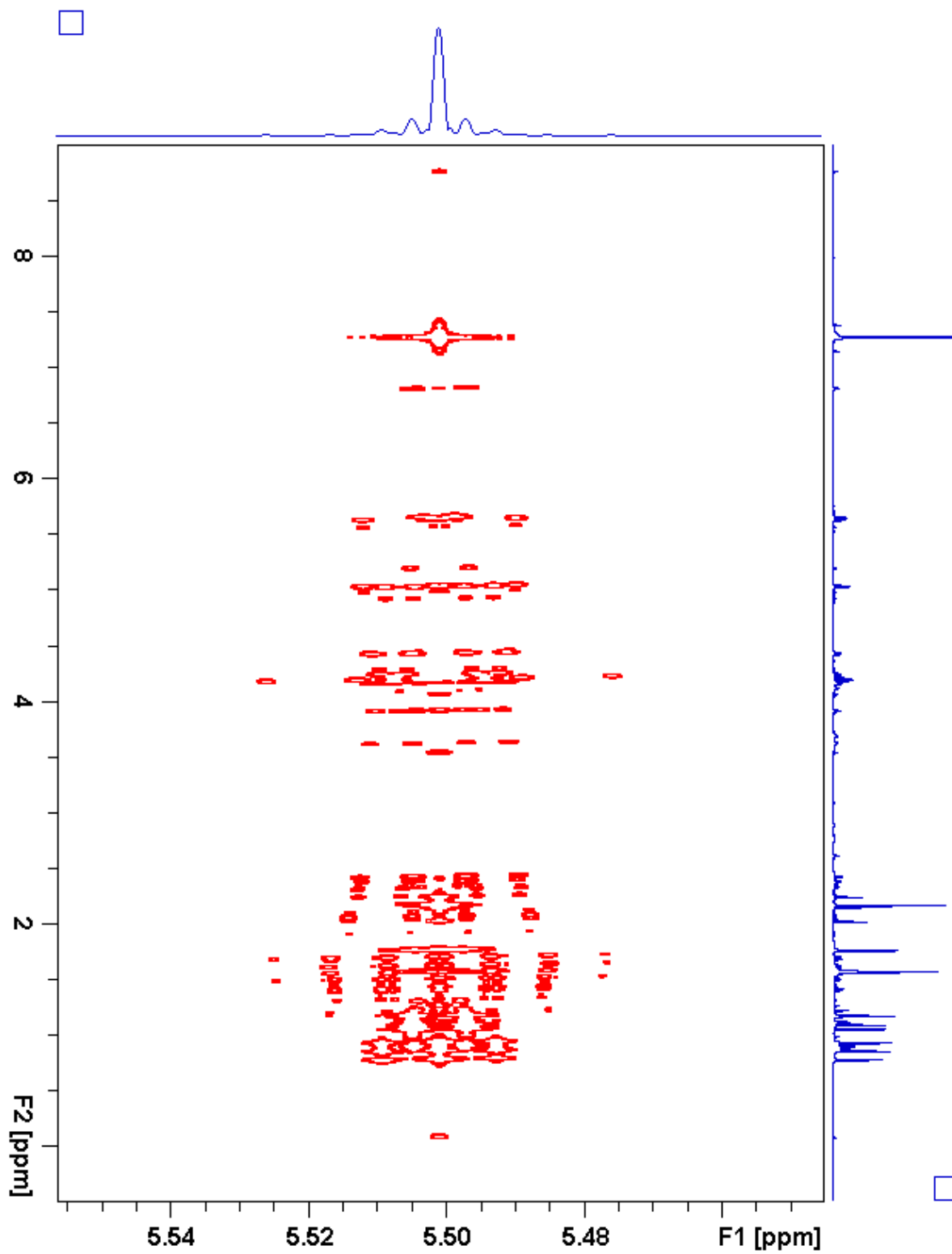
**Figure 2.78.** S5:22:1(2,5,5,5,5$^T$) *J*-resolved

**Table 2.19.** S5:21:0(2,2,5,6,6) Chemical shifts and coupling constants
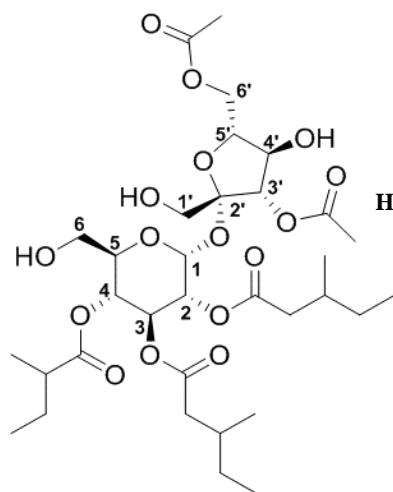


**Molecular Formula:** $C_{33}H_{54}O_{16}$

**110 min Retention Time (ESI+):** 52.81 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{33}H_{58}NO_{16}^+$ ([M+NH$_4^+$]): 724.3750, found: 724.3783

**Fraction, Batch:** #48, A-D
**Sample mass for NMR analysis:** 5.0 mg
**NMR Solvent:** CDCl$_3$
**InChi Key:** OBSFKETUAQFCKT-KUQBJEAISA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.62 (d, $J$ = 3.6 Hz) | 89.52 |
| **2**(CH) | 4.90 (dd, $J$ = 10.4, 3.7 Hz) | 70.48 |
| -1(CO) | | 172.3[a] |
| -2(CH$_2$) | 2.30 (dd, $J$ = 15.5, 5.8 Hz), 2.05 (dd, $J$ = 15.5, 8.3 Hz) | 41.02 |
| -3(CH) | 1.82 (m) | 31.62 |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.34 |
| -5(CH$_2$) | 1.33 (m), 1.20 (m) | 29.35 |
| -6(CH$_3$) | 0.87 (t, $J$ = 7.4 Hz) | 11.38[b] |
| **3**(CH) | 5.55 (dd, $J$ = 10.7, 9.2 Hz) | 68.98 |
| -1(CO) | | 172.3[a] |
| -2(CH$_2$) | 2.23 (dd, $J$ = 15.6, 5.5 Hz), 2.02 (dd, $J$ = 15.5, 8.4 Hz) | 41.26 |
| -3(CH) | 1.79 (m) | 31.65 |
| -4(CH$_3$) | 0.87 (d, $J$ = 6.7 Hz) | 19.41 |
| -5(CH$_2$) | 1.30 (m), 1.17 (m) | 29.37 |
| -6(CH$_3$) | 0.85 (t, $J$ = 7.4 Hz) | 11.38[b] |
| **4**(CH) | 4.97 (dd, $J$ = 10.5, 9.2 Hz) | 68.85 |
| -1(CO) | | 176.36 |
| -2(CH) | 2.35 (sextet, $J$ = 7.0 Hz) | 41.13 |
| -3(CH$_3$) | 1.10 (d, $J$ = 7.0 Hz) | 16.53 |
| -4(CH$_2$) | 1.66 (m), 1.44 (m) | 26.67 |
| -5(CH$_3$) | 0.90 (t, $J$ = 7.4 Hz) | 11.83 |
| **5**(CH) | 4.05 (ddd, $J$ = 10.3, 5.2, 2.1 Hz) | 71.13 |
| **6**(CH$_2$) | 3.66 (dd, $J$ = 12.8, 2.3 Hz), 3.58 (dd, $J$ = 12.8, 5.3 Hz) | 61.64 |
| **1'**(CH$_2$) | 3.63 (d, $J$ = 12.3 Hz), 3.55 (d, $J$ = 12.3 Hz) | 64.27[c] |
| **2'**(C) | | 104.65 |

**Table 2.19.** (continued)

| | | |
|---|---|---|
| **3'**(CH) | 5.20 (d, $J$ = 7.6 Hz) | 79.30 |
| -1(CO) | | 172.01 |
| -2(CH$_3$) | 2.22 (s) | 20.99 |
| **4'**(CH) | 4.35 (t, $J$ = 7.6 Hz) | 73.90 |
| **5'**(CH) | 4.08 (ddd, $J$ = 7.6, 6.1, 3.6 Hz) | 80.43 |
| **6'**(CH$_2$) | 4.41 (dd, $J$ = 12.1, 6.2 Hz), 4.28 (dd, $J$ = 12.1, 3.6 Hz) | 64.27[c] |
| -1(CO) | | 171.72 |
| -2(CH$_3$) | 2.12 (s) | 21.01 |

a - Two $^{13}$C signals not resolved in 2D spectra (172.34, 172.35 ppm)

b - Two $^{13}$C signals overlapping

c - Two $^{13}$C signals overlapping
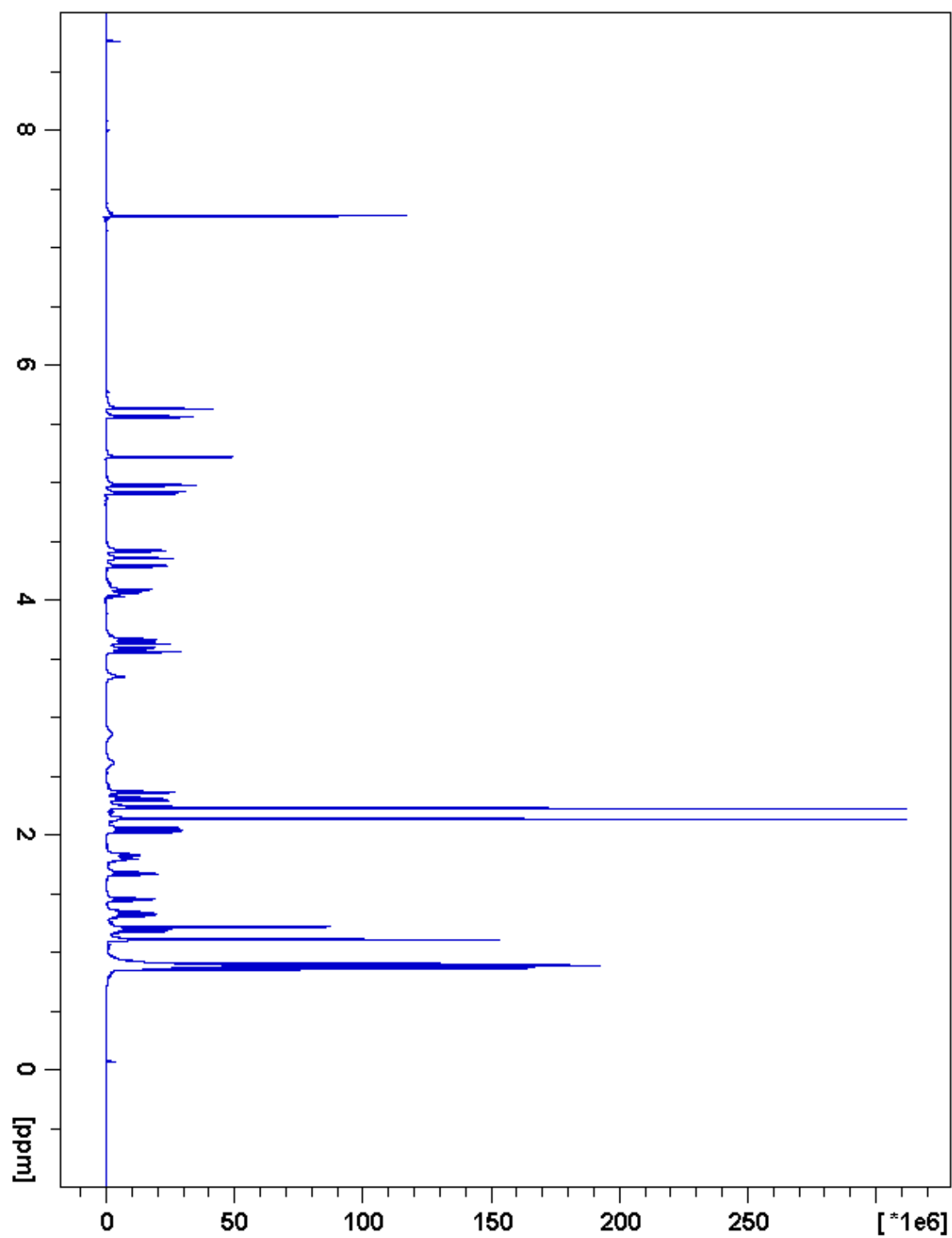
**Figure 2.79.** S5:21:0(2,2,5,6,6) $^1$H NMR
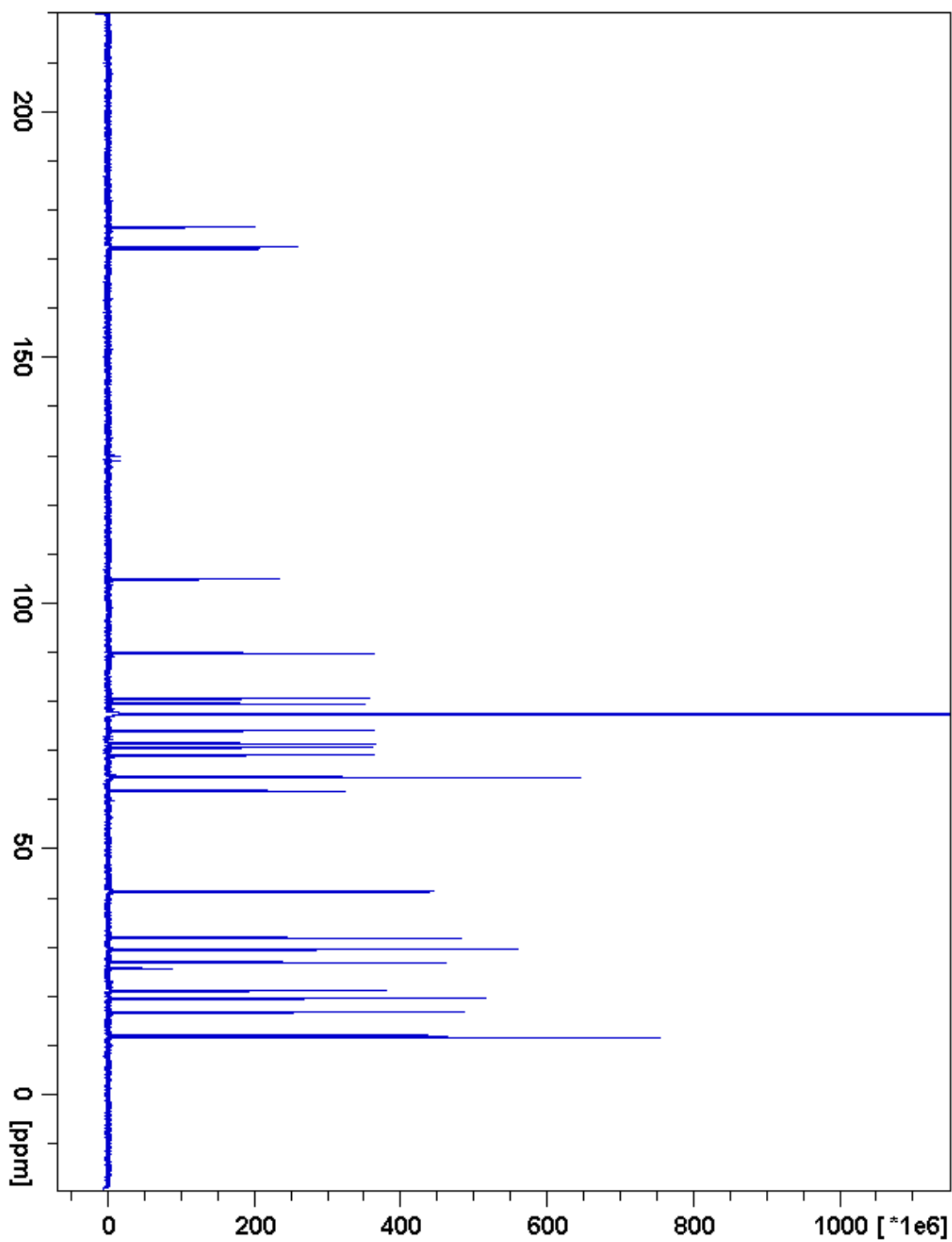
**Figure 2.80.** S5:21:0(2,2,5,6,6) $^{13}$C NMR
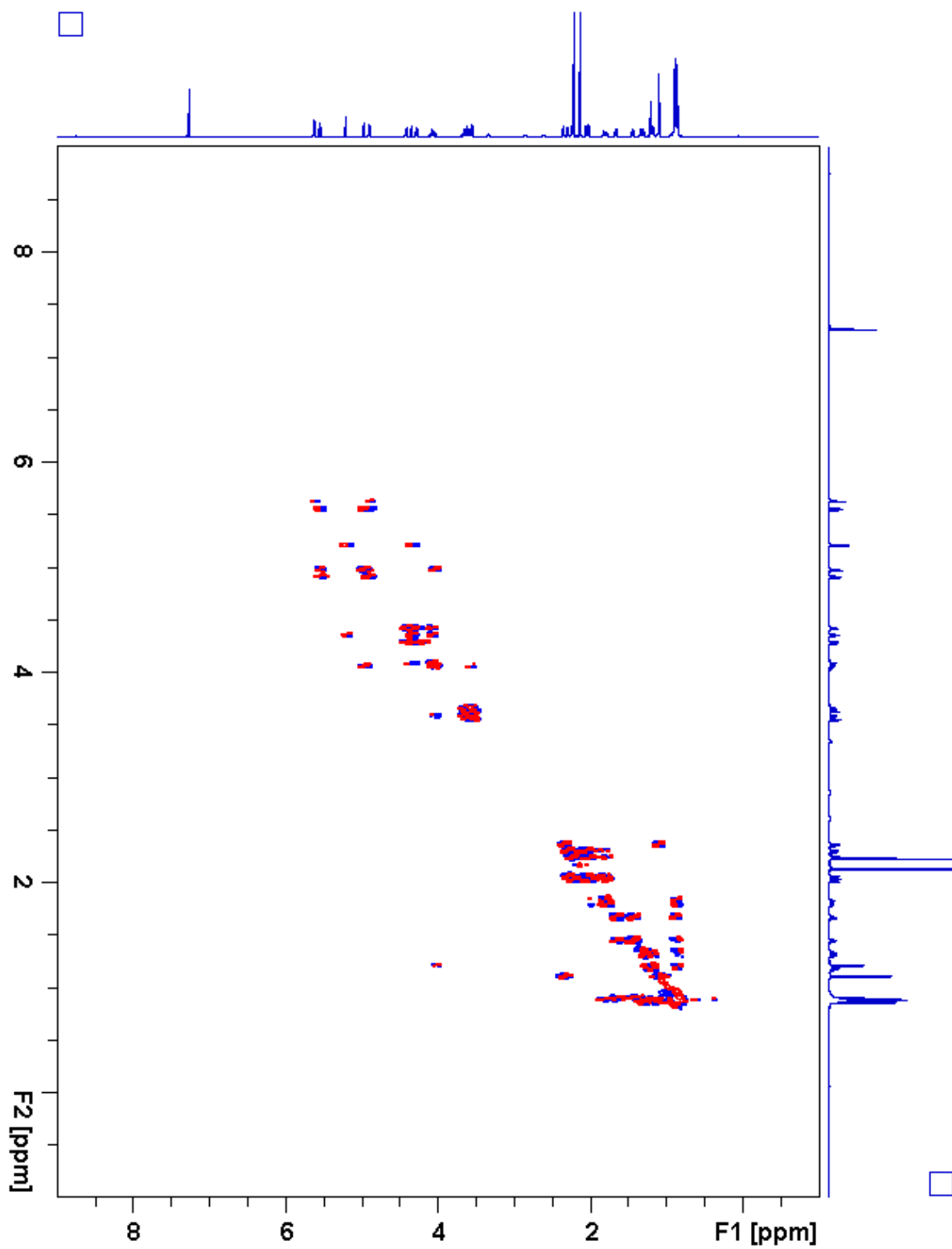
**Figure 2.81.** S5:21:0(2,2,5,6,6) $^1$H-$^1$H gCOSY
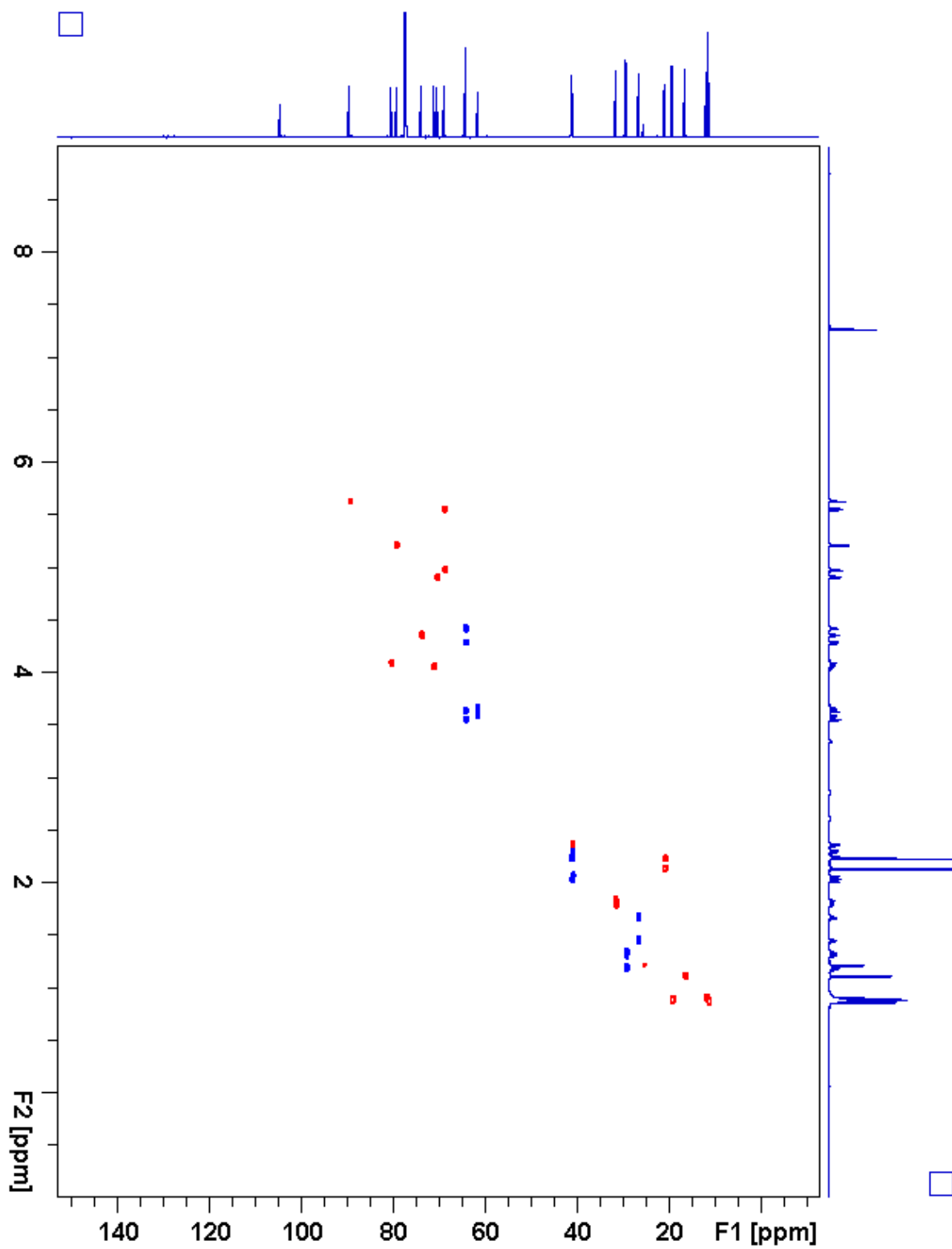
**Figure 2.82.** S5:21:0(2,2,5,6,6) gHSQC
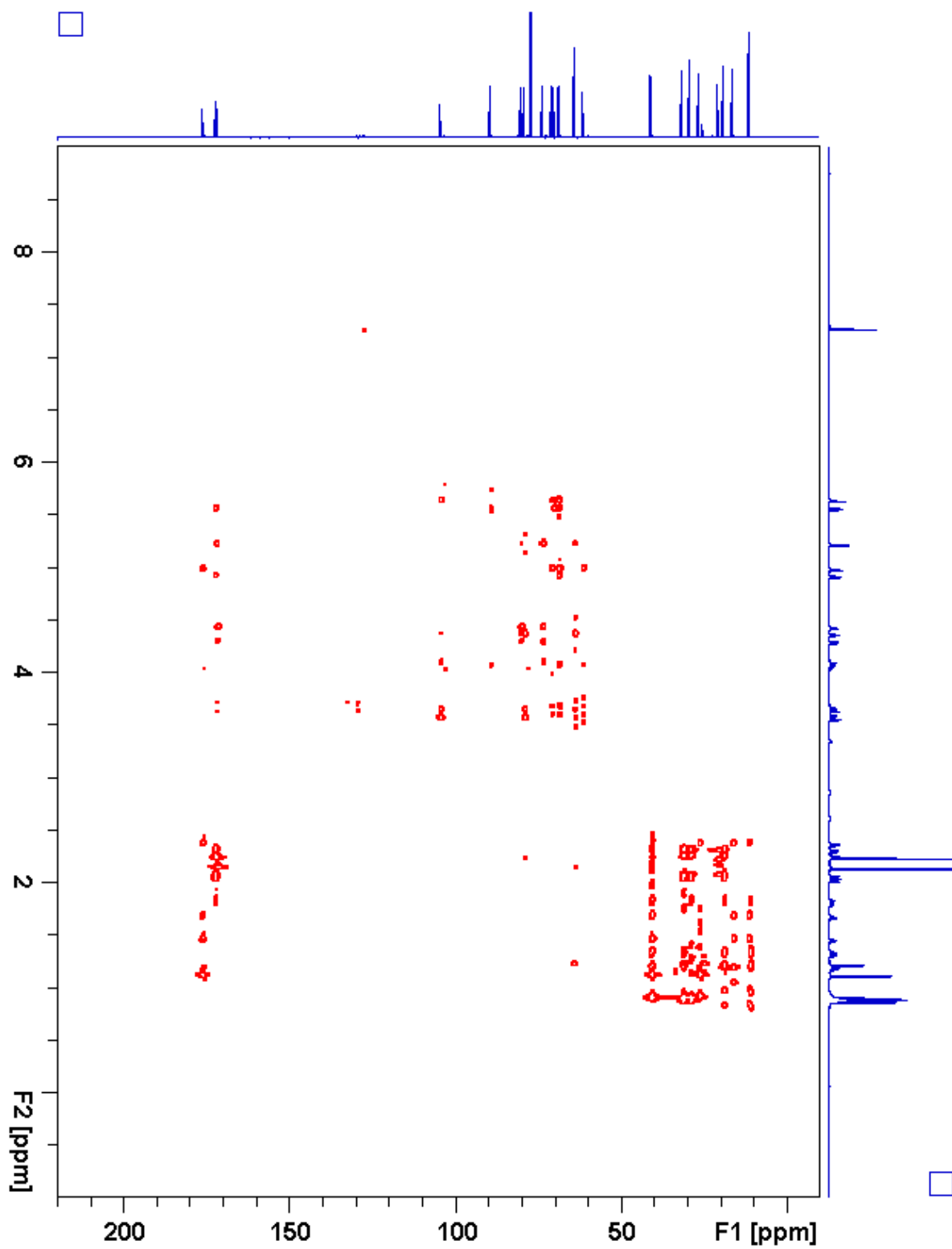
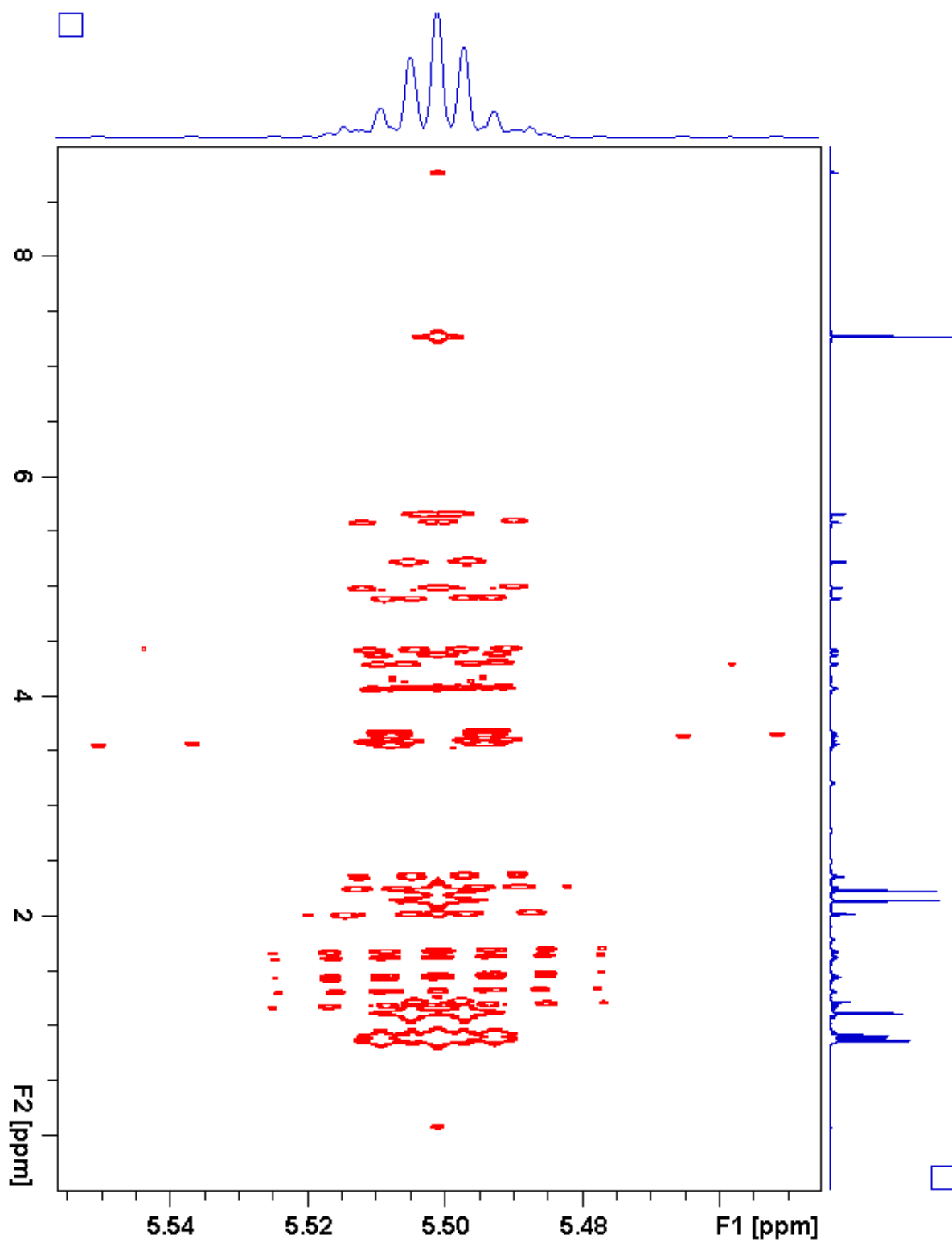**Figure 2.83.** S5:21:0(2,2,5,6,6) gHMBC
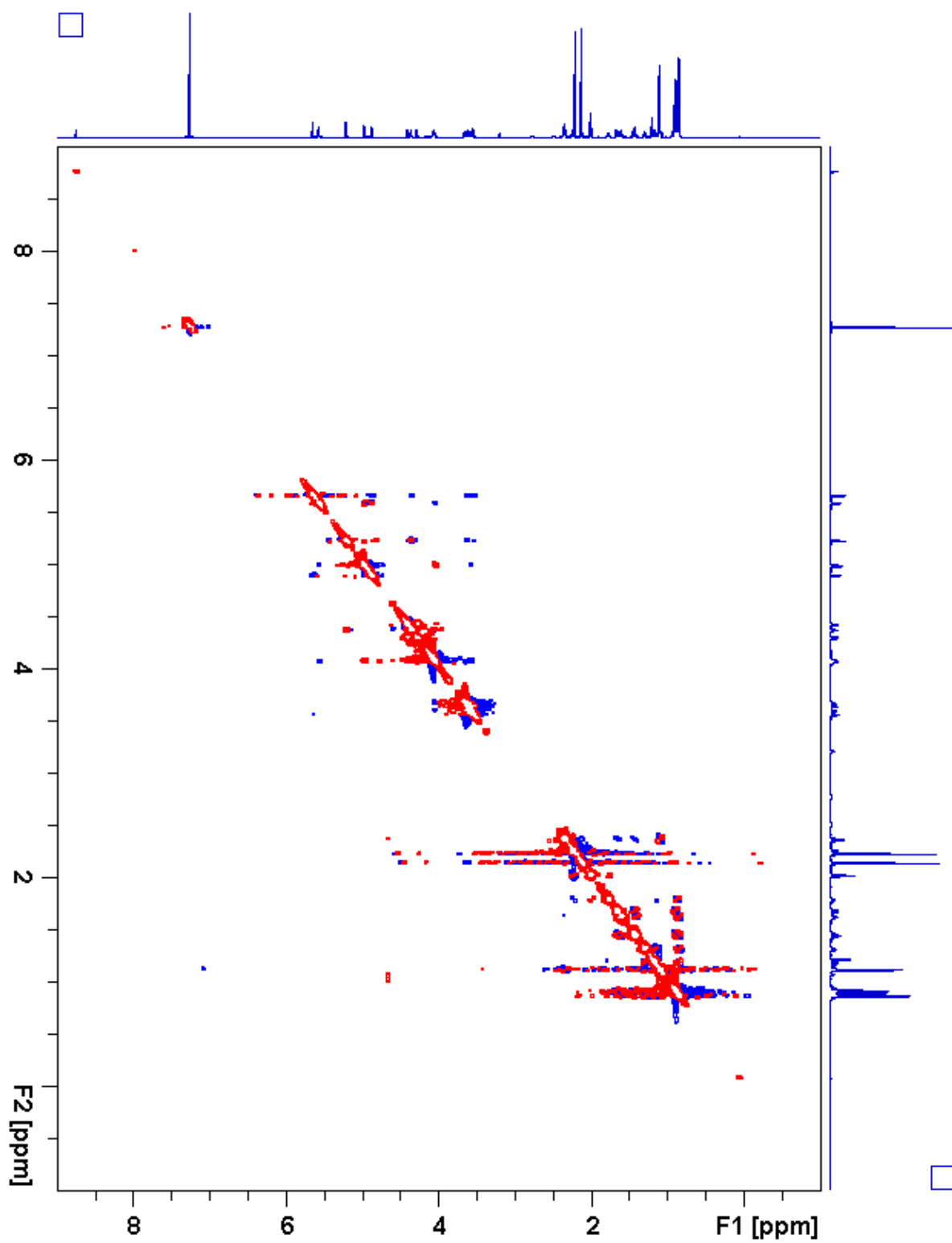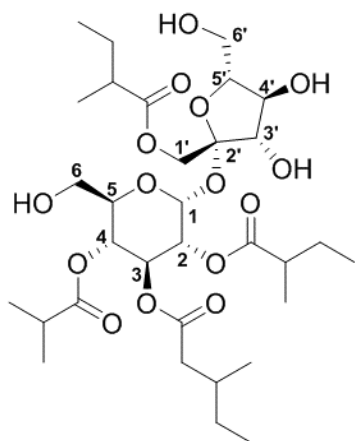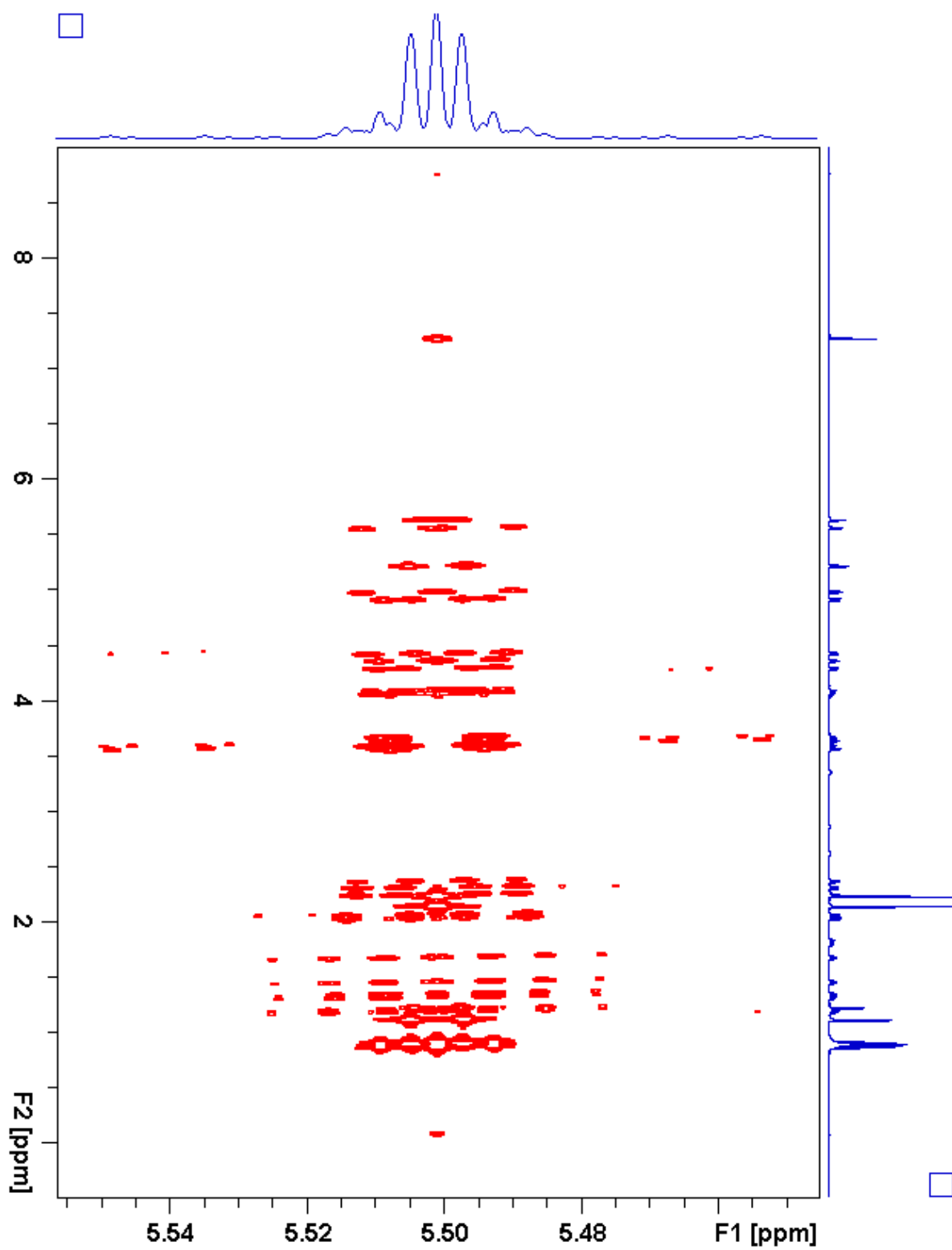
**Figure 2.84.** S5:21:0(2,2,5,6,6) *J*-resolved

**Figure 2.85.** S5:21:0(2,2,5,6,6) ROESY

**Table 2.20.** S4:21:0(5,5,5,6) Chemical shifts and coupling constants
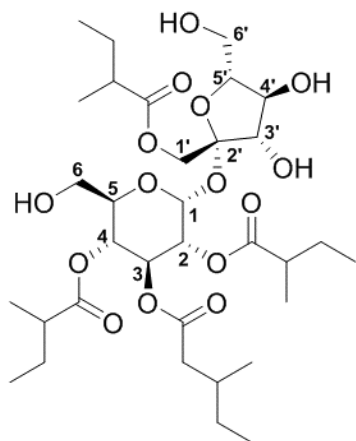


**Molecular Formula:** $C_{33}H_{56}O_{15}$

**110 min Retention Time (ESI+):** 53.44 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{33}H_{60}NO_{15}^+$ ([M+NH$_4^+$]): 710.3957, found: 710.3993

**Fraction, Batch:** #52, A-D

**Sample mass for NMR analysis:** 1.1 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** XBHZETAMEBHMPK-LLOMWPEHSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.87 (d, *J* = 4.0 Hz) | 88.79 |
| **2**(CH) | 4.78 (dd, *J* = 10.3, 4.1 Hz) | 71.11 |
| -1(CO) | | 177.57 |
| -2(CH) | 2.41 (sextet, *J* = 7.0 Hz) | 40.78 |
| -3(CH$_3$) | 1.14 (d, *J* = 7.0 Hz) | 16.21 |
| -4(CH$_2$) | 1.64 (m), 1.45 (m) | 26.73 |
| -5(CH$_3$) | 0.87 (t, *J* = 7.4 Hz) | 11.64 |
| **3**(CH) | 5.58 (dd, *J* = 10.6, 9.3 Hz) | 68.76 |
| -1(CO) | | 172.15 |
| -2(CH$_2$) | 2.25 (dd, *J* = 15.8, 5.4 Hz), 2.02 (dd, *J* = 15.8, 8.5 Hz) | 41.1[a] |
| -3(CH) | 1.79 (m) | 31.53 |
| -4(CH$_3$) | 0.89 (d, *J* = 6.7 Hz) | 19.45 |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.40 |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.39 |
| **4**(CH) | 4.93 (dd, *J* = 10.7, 9.3 Hz) | 68.58 |
| -1(CO) | | 176.15 |
| -2(CH) | 2.36 (sextet, *J* = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.10 (d, *J* = 7.0 Hz) | 16.42 |
| -4(CH$_2$) | 1.66 (m), 1.44 (m) | 26.64 |
| -5(CH$_3$) | 0.90 (t, *J* = 7.4 Hz) | 11.82 |
| **5**(CH) | 4.15 (ddd, *J* = 10.6, 6.2, 2.5 Hz) | 71.99 |
| **6**(CH$_2$) | 3.63 (dd, *J* = 12.4, 2.4 Hz), 3.60 (dd, *J* = 12.4, 6.5 Hz) | 61.51 |

**Table 2.20.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.13 (d, $J$ = 11.6 Hz), 4.06 (d, $J$ = 11.6 Hz) | 64.07 |
| -1(CO) | | 175.92 |
| -2(CH) | 2.41 (sextet, $J$ = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.15 (d, $J$ = 7.0 Hz) | 16.73 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.89 |
| -5(CH$_3$) | 0.91 (t, $J$ = 7.4 Hz) | 11.74 |
| **2'**(C) | | 103.74 |
| **3'**(CH) | 4.17 (d, $J$ = 8.9 Hz) | 78.43 |
| **4'**(CH) | 4.31 (t, $J$ = 8.9 Hz) | 72.56 |
| **5'**(CH) | 3.71 (dt, $J$ = 9.2, 2.3 Hz) | 80.98 |
| **6'**(CH$_2$) | 3.88 (dd, $J$ = 13.4, 2.4 Hz), 3.71 (dd, $J$ = 13.4, 2.3 Hz) | 59.71 |

a - Three $^{13}$C signals not resolved in 2D spectra (41.08, 41.10, 41.13 ppm)
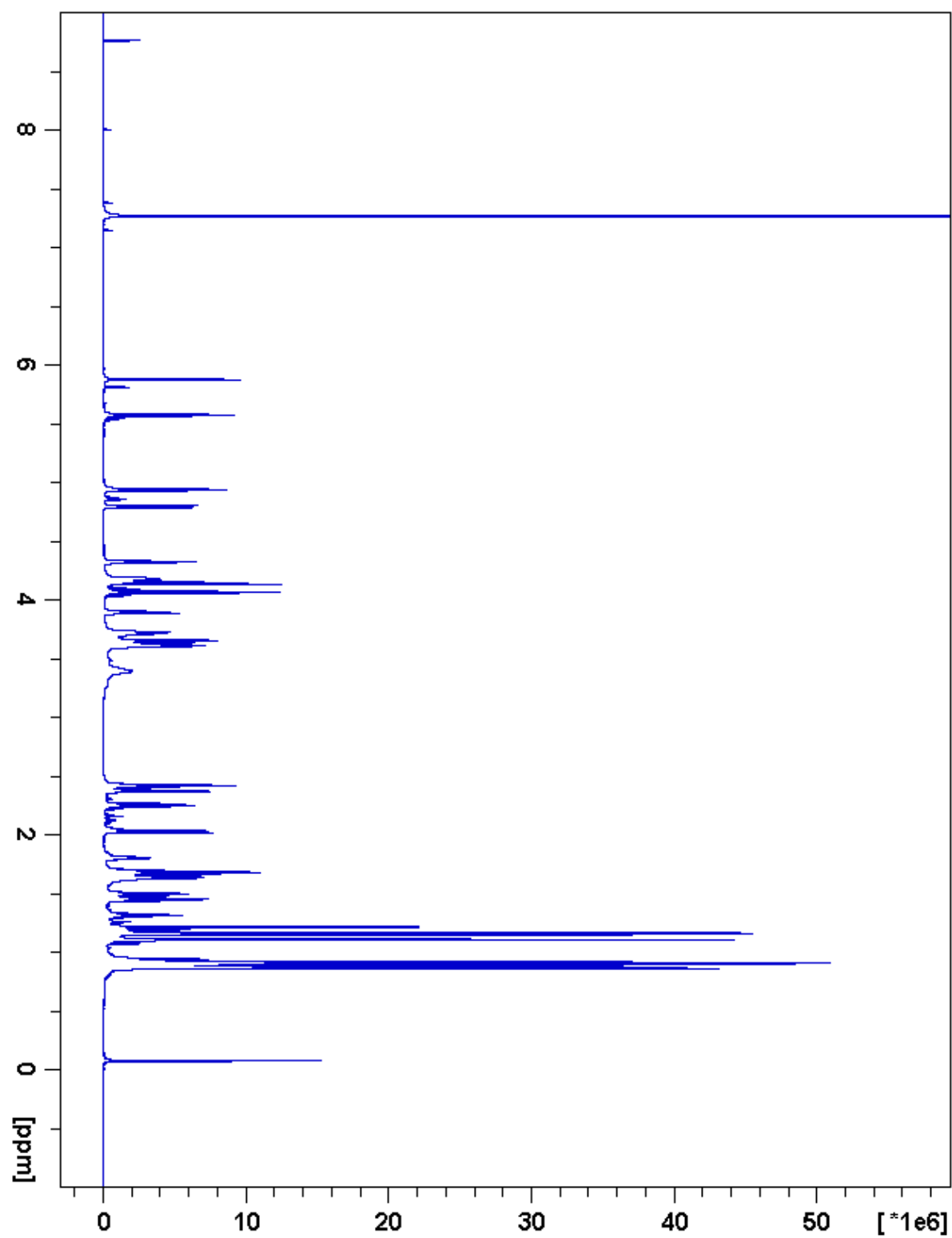
**Figure 2.86.** S4:21:0(5,5,5,6) $^1$H NMR
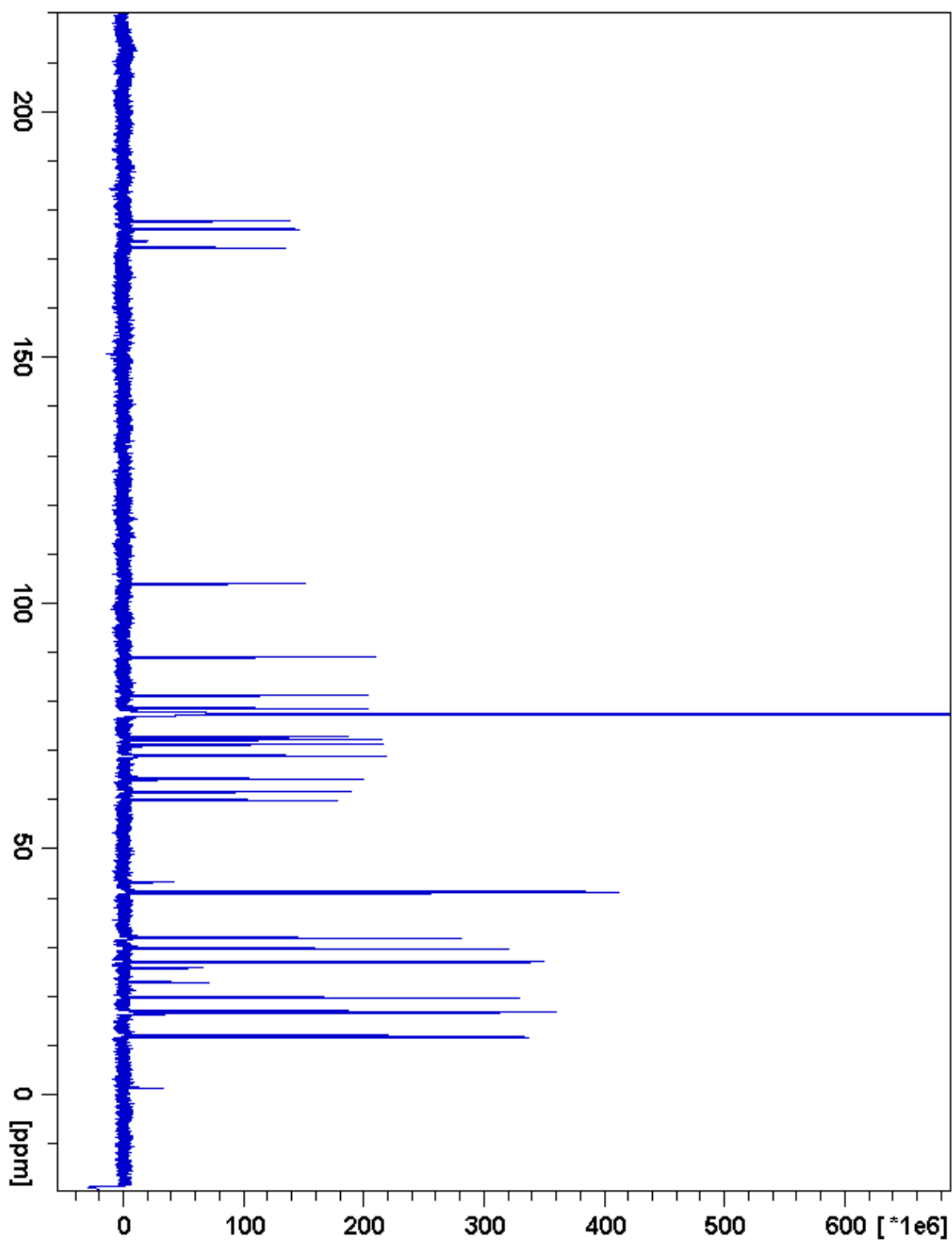
**Figure 2.87.** S4:21:0(5,5,5,6) $^{13}$C NMR

174
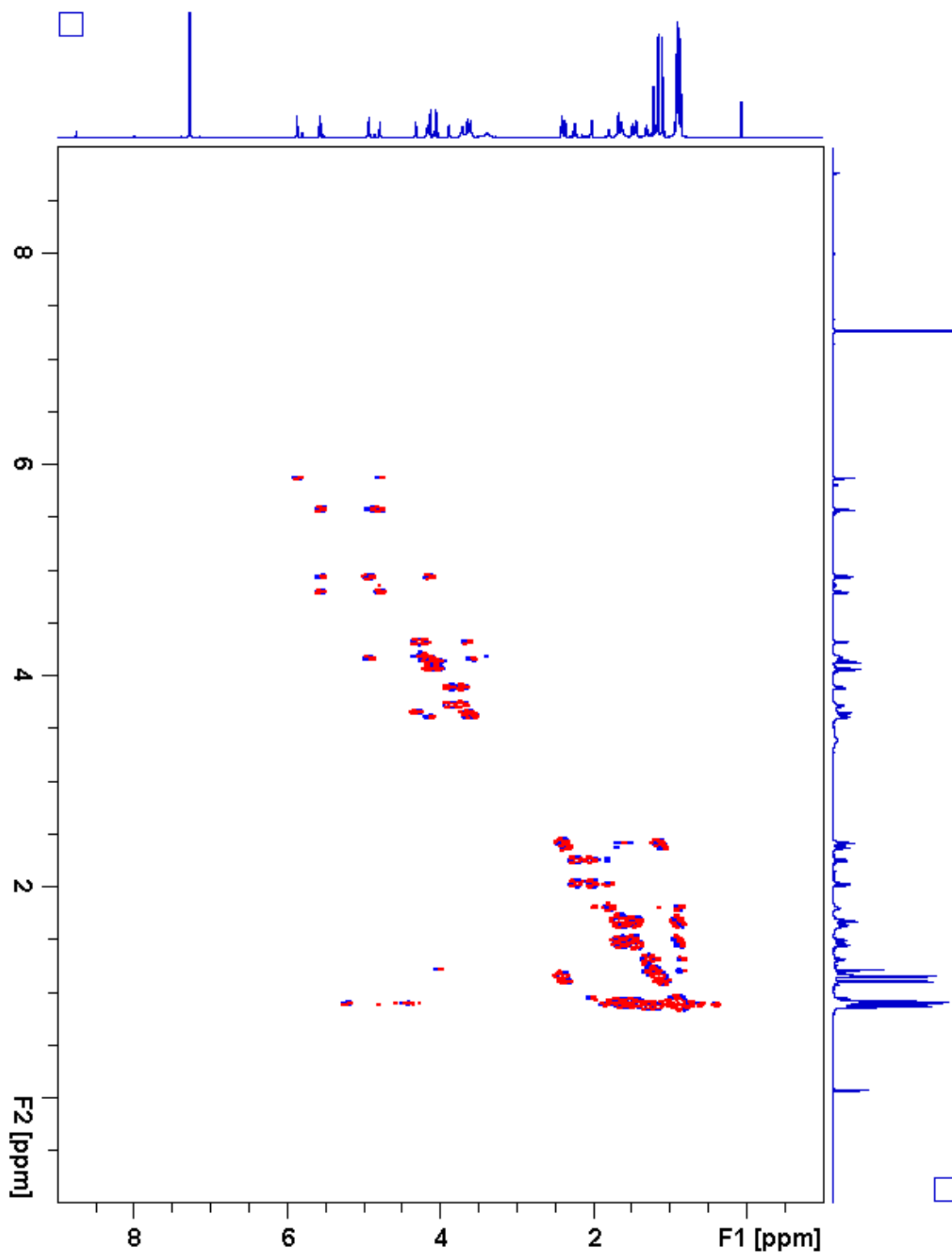
**Figure 2.88.** S4:21:0(5,5,5,6) $^1$H-$^1$H gCOSY
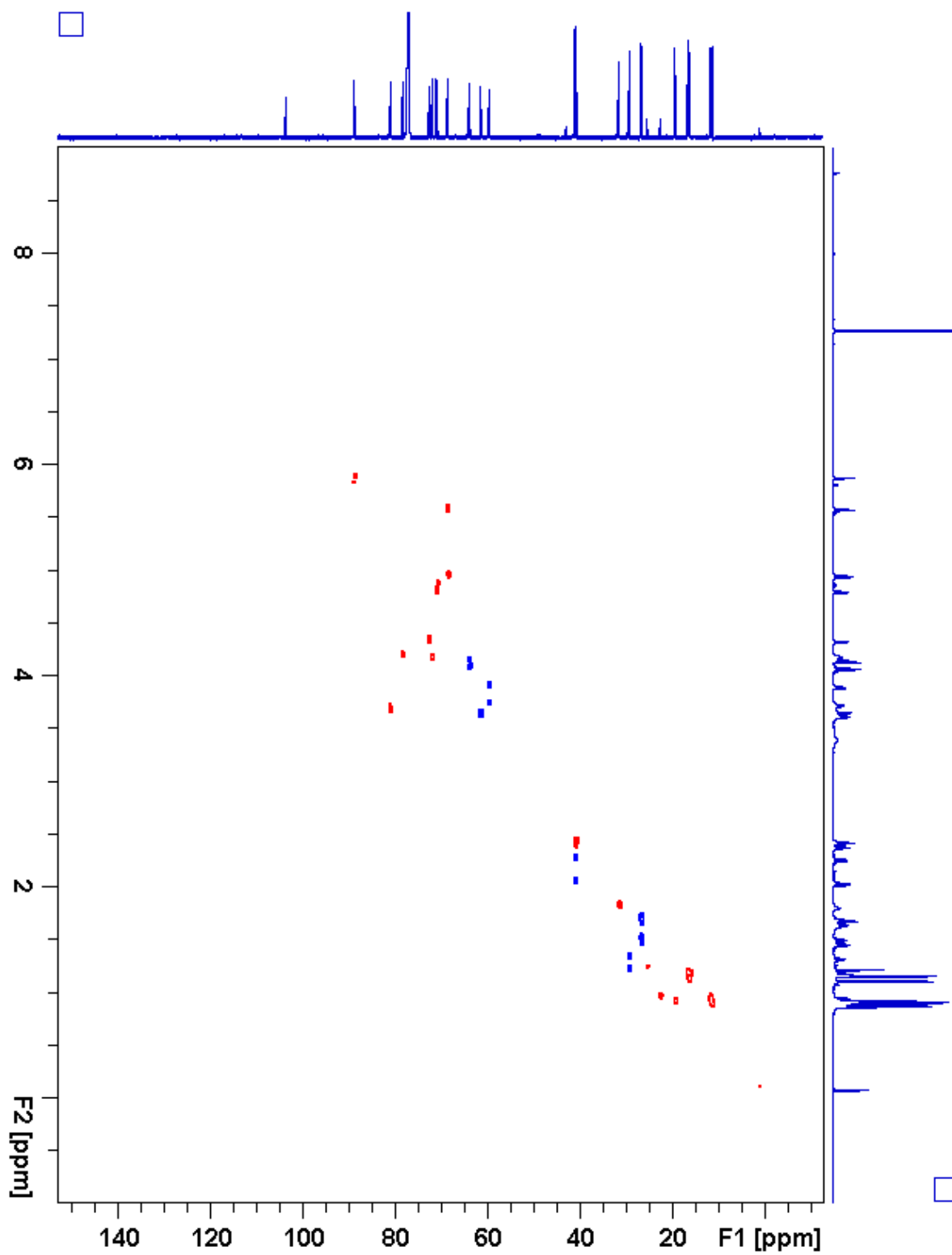
**Figure 2.89.** S4:21:0(5,5,5,6) gHSQC
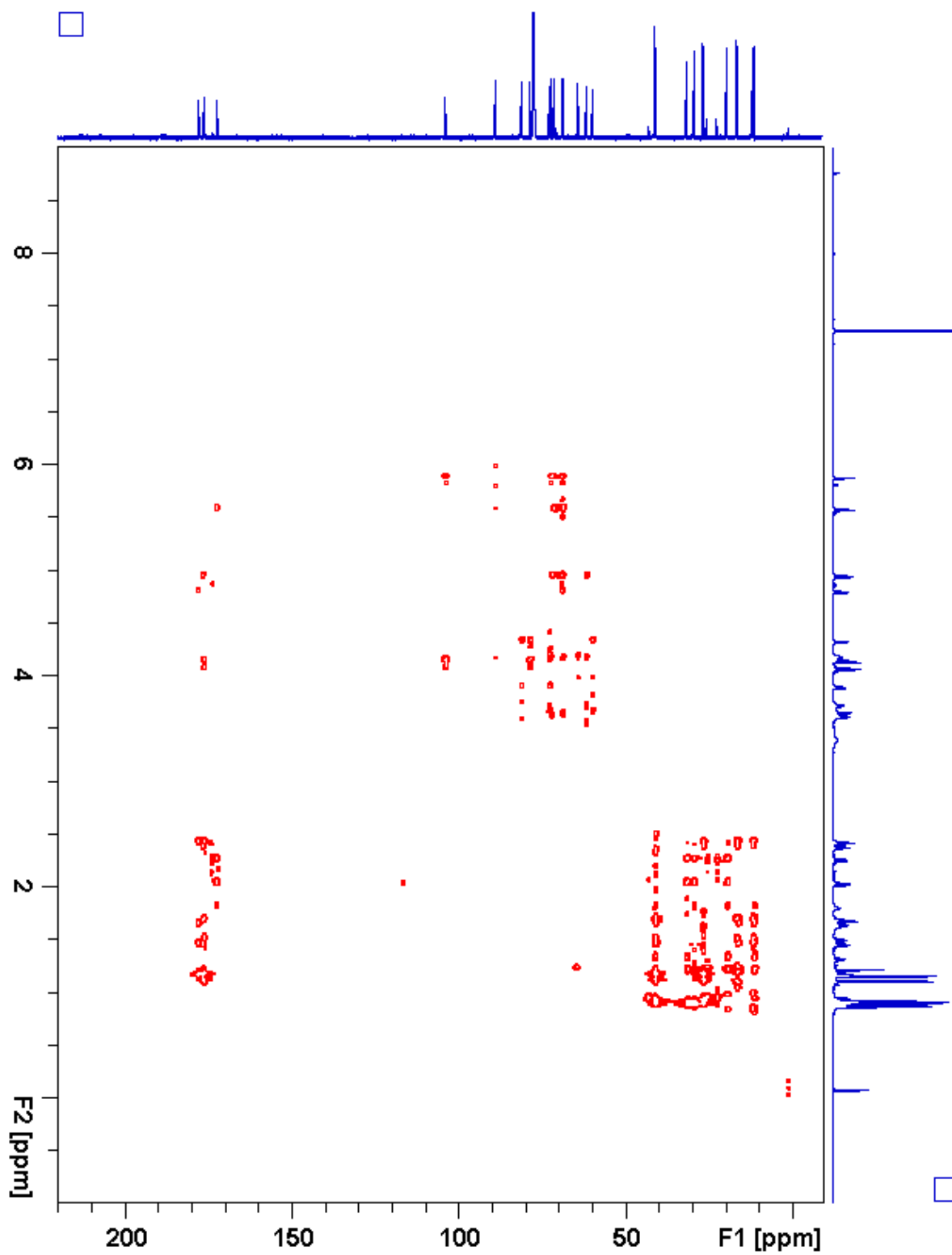
**Figure 2.90.** S4:21:0(5,5,5,6) gHMBC
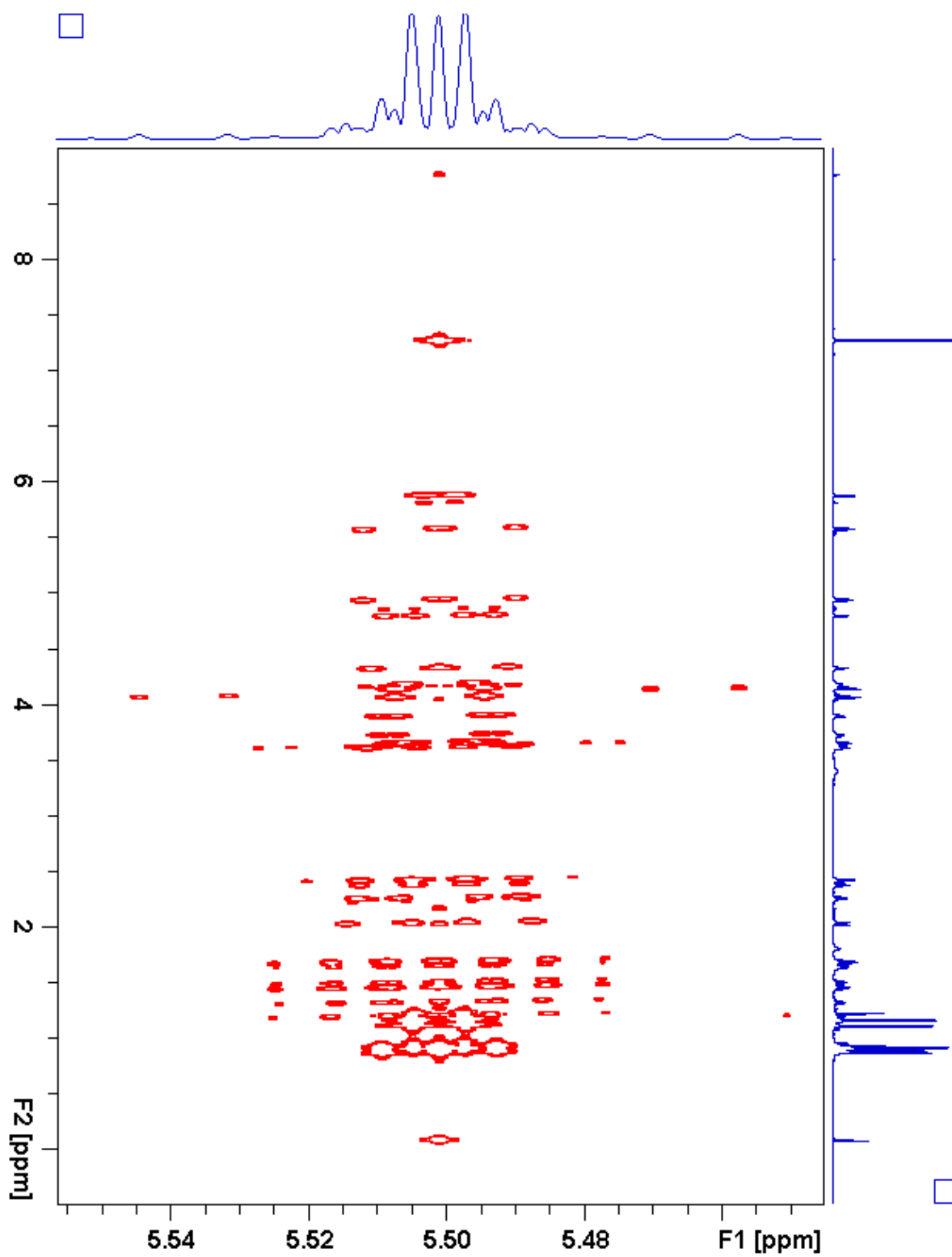
**Figure 2.91.** S4:21:0(5,5,5,6) *J*-resolved
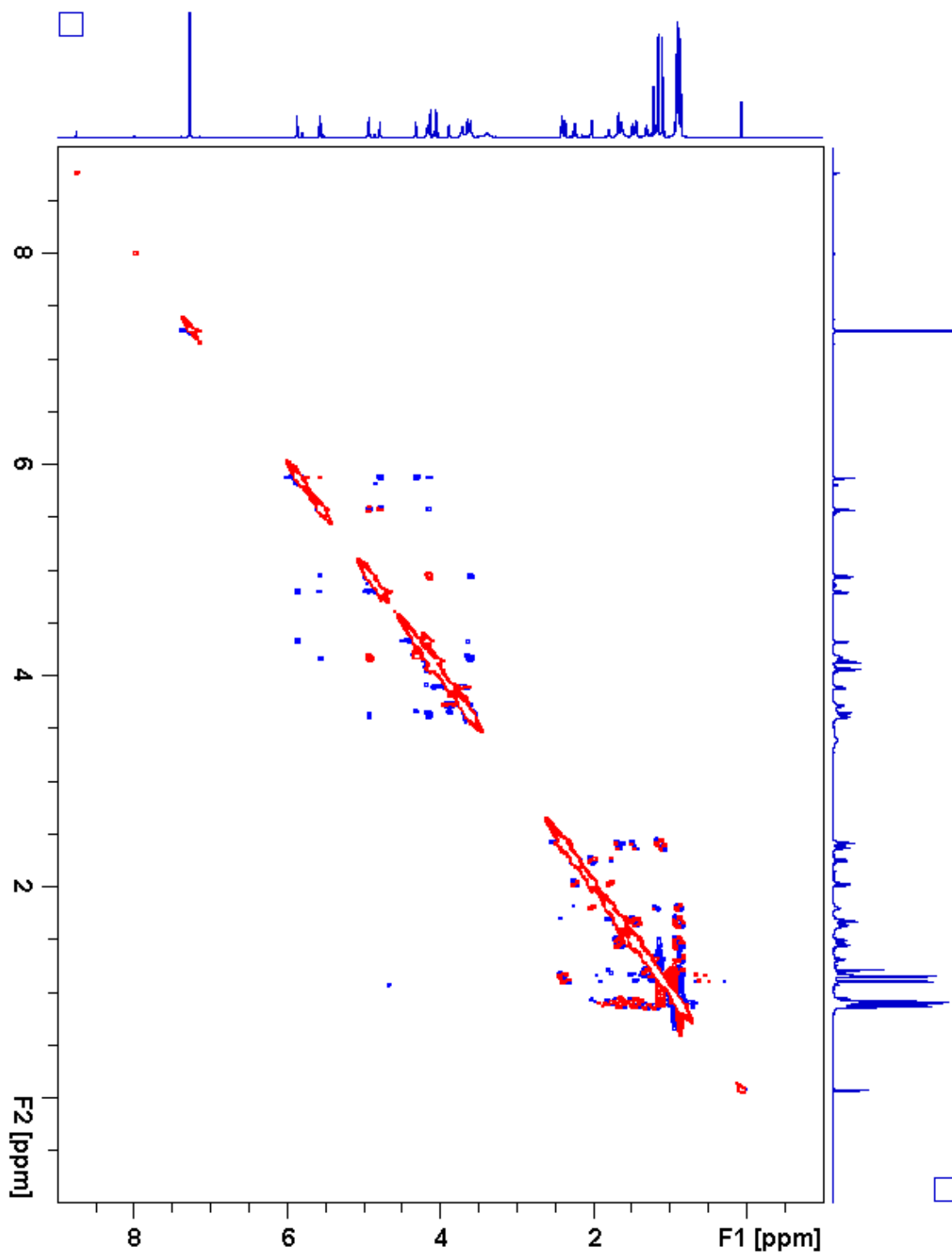
**Figure 2.92.** S4:21:0(5,5,5,6) ROESY

**Table 2.21.** S5:25:4(2,5,5,5,8$^P$) Chemical shifts and coupling constants
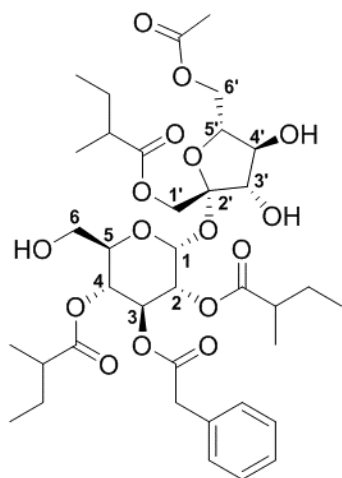


**Molecular Formula:** $C_{37}H_{54}O_{16}$

**110 min Retention Time (ESI+):** 54.29 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{37}H_{58}NO_{16}^+$ ([M+NH$_4^+$]): 772.3750, found: 772.3748

**Fraction, Batch:** #55-56, A-G & 2nd purification  w/F5 column #26-27 combined

**Sample mass for NMR analysis:** < 0.1 mg

**NMR Solvent:** d$_3$-acetonitrile

**InChi Key:** ZOZCRBKJAIYBPH-IJNFVYORSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.58 (d, *J* = 3.9 Hz) | 90.22 |
| **2**(CH) | 4.90 (dd, *J* = 10.4, 3.9 Hz) | 71.37 |
| -1(CO) | | 176.75 |
| -2(CH) | 2.30 (sextet, *J* = 7.0 Hz) | 41.74[a] |
| -3(CH$_3$) | 1.03 (d, *J* = 7.0 Hz) | 16.54 |
| -4(CH$_2$) | 1.53 (m), 1.35 (m) | 27.71 |
| -5(CH$_3$) | 0.83 (t, *J* = 7.4 Hz) | 12.23 |
| **3**(CH) | 5.49 (dd, *J* = 10.7, 9.2 Hz) | 71.49 |
| -1(CO) | | 172.11 |
| -2(CH$_2$) | 3.55 (d, *J* = 16.0 Hz), 3.53 (d, *J* = 16.0 Hz) | 41.73[a] |
| -3(C) | | 135.06 |
| -4(2 x CH) | 7.19 (m) | 130.86 |
| -5(2 x CH) | 7.30 (m) | 129.82 |
| -6(CH) | 7.26 (m) | 128.47 |
| **4**(CH) | 5.03 (dd, *J* = 10.6, 9.2 Hz) | 69.70 |
| -1(CO) | | 176.66 |
| -2(CH) | 2.25 (sextet, *J* = 7.0 Hz) | 42.06[b] |
| -3(CH$_3$) | 0.98 (d, *J* = 7.0 Hz) | 17.19 |
| -4(CH$_2$) | 1.57 (m), 1.37 (m) | 27.40 |
| -5(CH$_3$) | 0.85 (t, *J* = 7.4 Hz) | 12.26 |
| **5**(CH) | 4.12 (*J* = 10.3, 5.7, 2.6 Hz) | 72.08 |
| **6**(CH$_2$) | 3.57 (dd, *J* = 12.2, 2.9 Hz), 3.45 (dd, *J* = 12.2, 5.7 Hz) | 61.95 |

**Table 2.21.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.01 (s) | 64.02 |
| -1(CO) | | 176.82 |
| -2(CH) | 2.42 (sextet, $J$ = 7.0 Hz) | 42.06[b] |
| -3(CH$_3$) | 1.11 (d, $J$ = 7.0 Hz) | 17.25 |
| -4(CH$_2$) | 1.63 (m), 1.48 (m) | 27.88 |
| -5(CH$_3$) | 0.89 (t, $J$ = 7.4 Hz) | 12.21 |
| **2'**(C) | | 104.55 |
| **3'**(CH) | 3.98 (m) | 78.29 |
| **4'**(CH) | 3.97 (t, $J$ = 8.9 Hz) | 75.16 |
| **5'**(CH) | 3.80 (ddd, $J$ = 8.9, 6.7, 2.6 Hz) | 80.59 |
| **6'**(CH$_2$) | 4.26 (dd, $J$ = 12.2, 2.7 Hz), 4.19 (dd, $J$ = 12.2, 7.0 Hz) | 65.60 |
| -1(CO) | | 171.96 |
| -2(CH$_3$) | 2.03 (s) | 21.39 |

a - Two $^{13}$C signals not resolved in 2D spectra (41.73, 41.74 ppm)

b - Two $^{13}$C signals overlapping
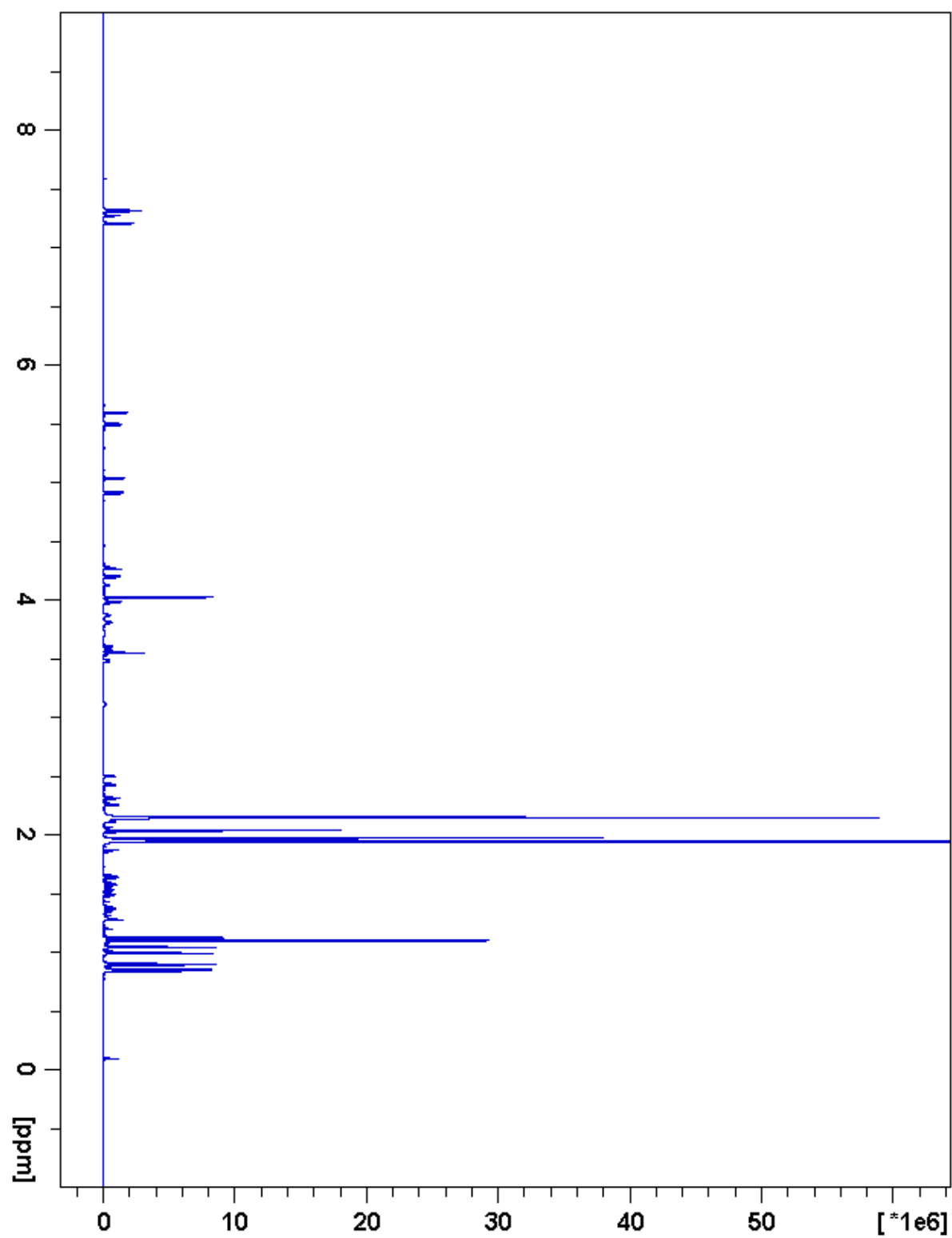
**Figure 2.93.** S5:25:4(2,5,5,5,8[P]) ^{1}H NMR
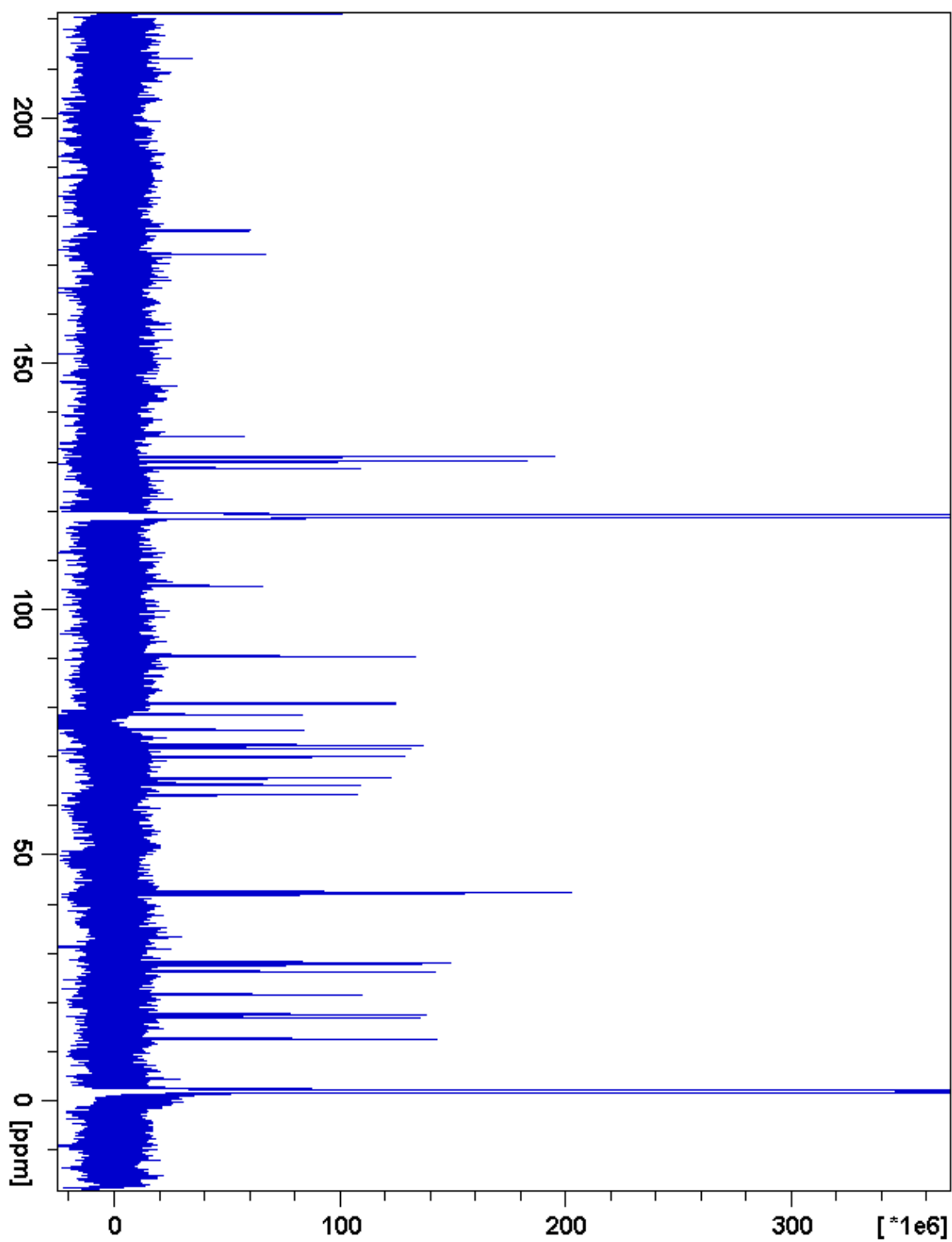
**Figure 2.94.** S5:25:4(2,5,5,5,8$^P$) $^{13}$C NMR
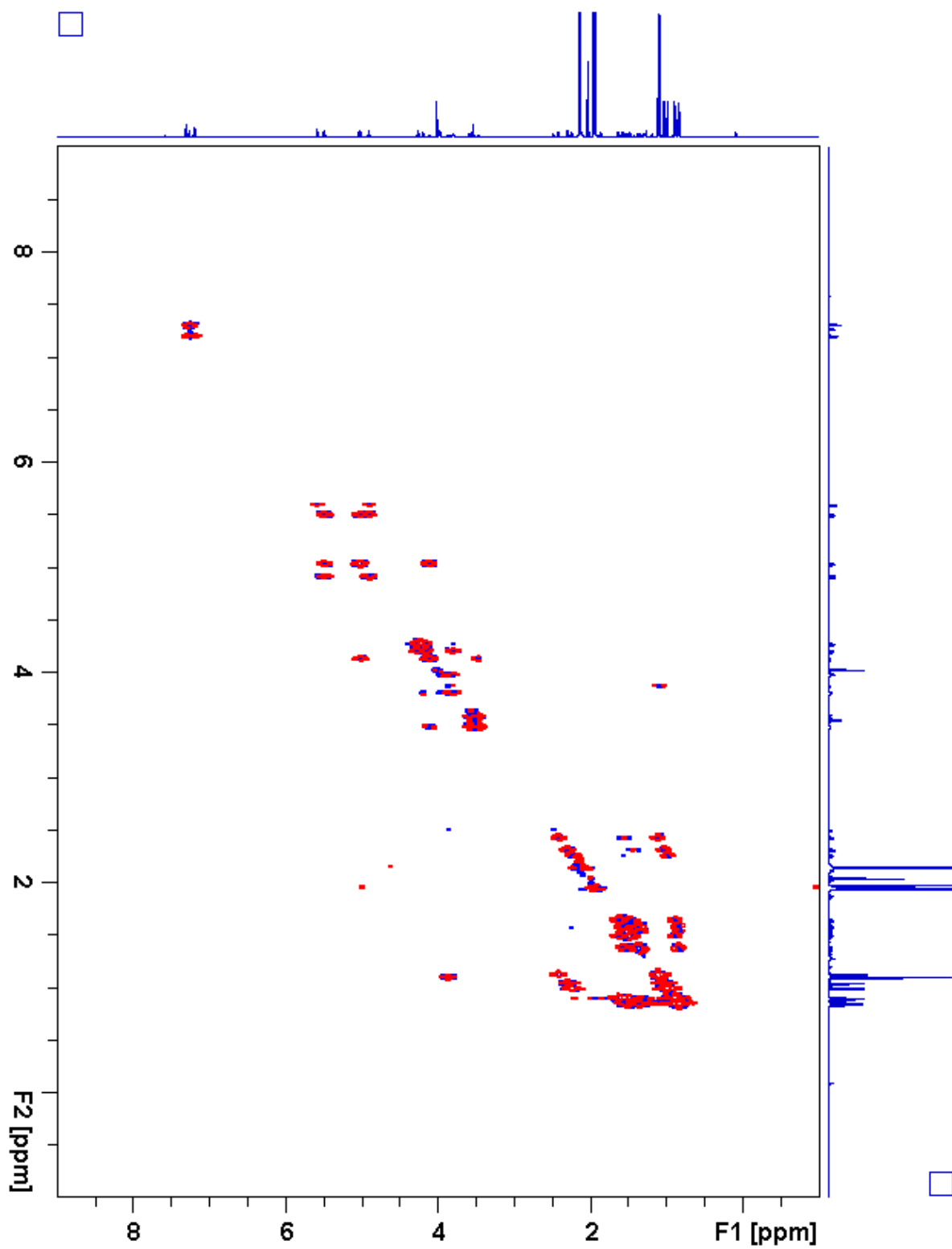
**Figure 2.95.** S5:25:4(2,5,5,5,8$^P$) $^1$H-$^1$H gCOSY
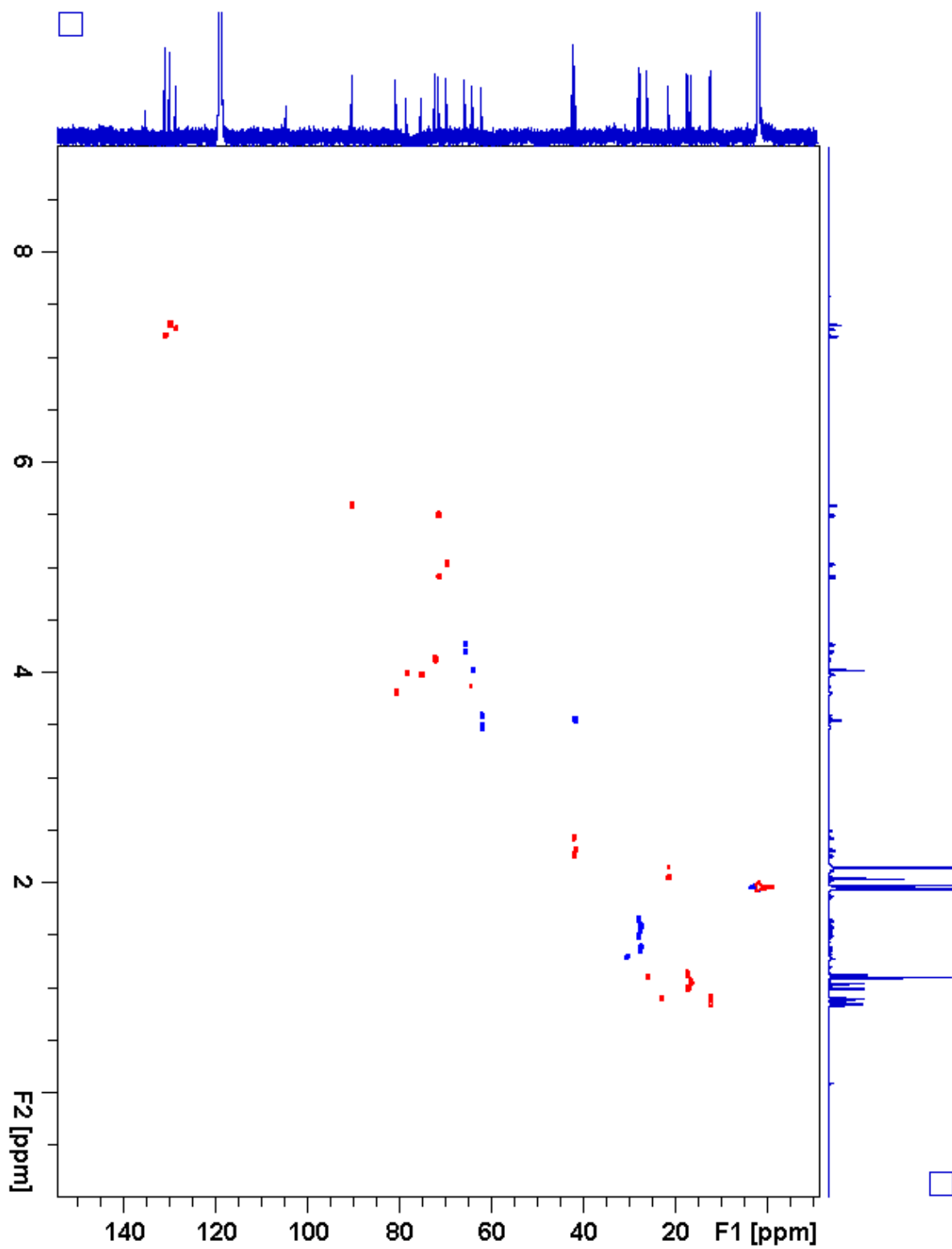
**Figure 2.96.** S5:25:4(2,5,5,5,8$^P$) gHSQC
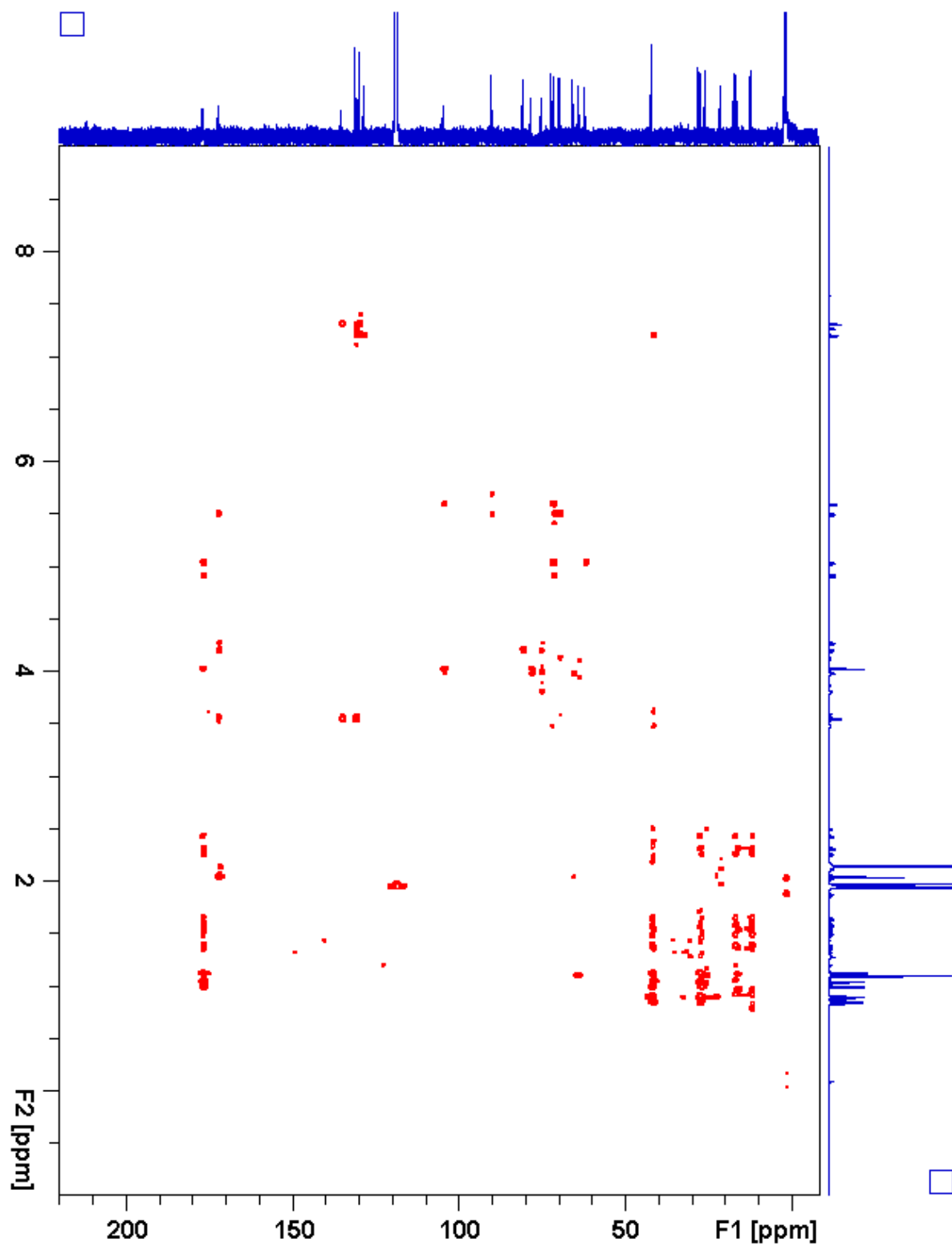
**Figure 2.97.** S5:25:4(2,5,5,5,8$^P$) gHMBC
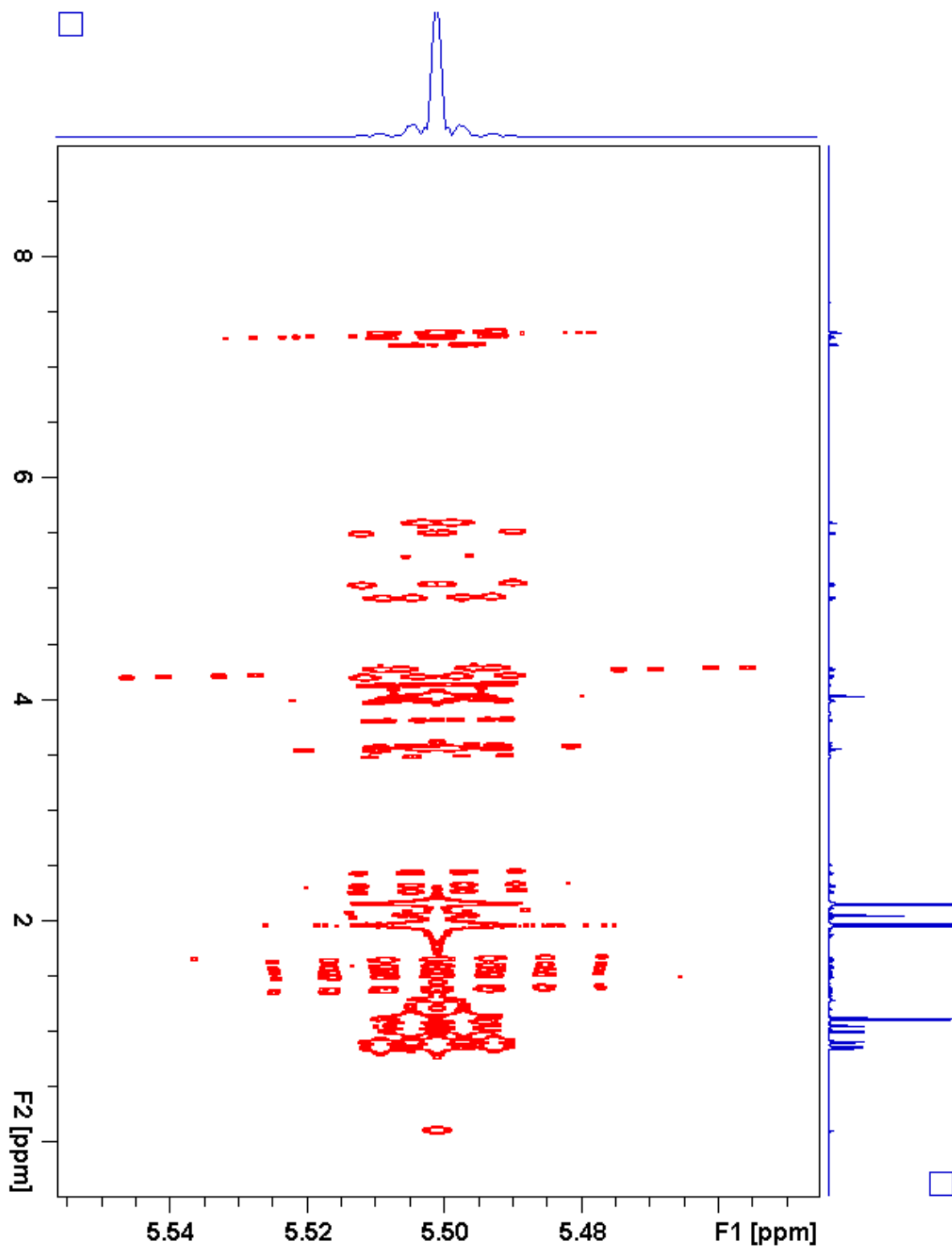
**Figure 2.98.** S5:25:4(2,5,5,5,8$^P$) *J*-resolved
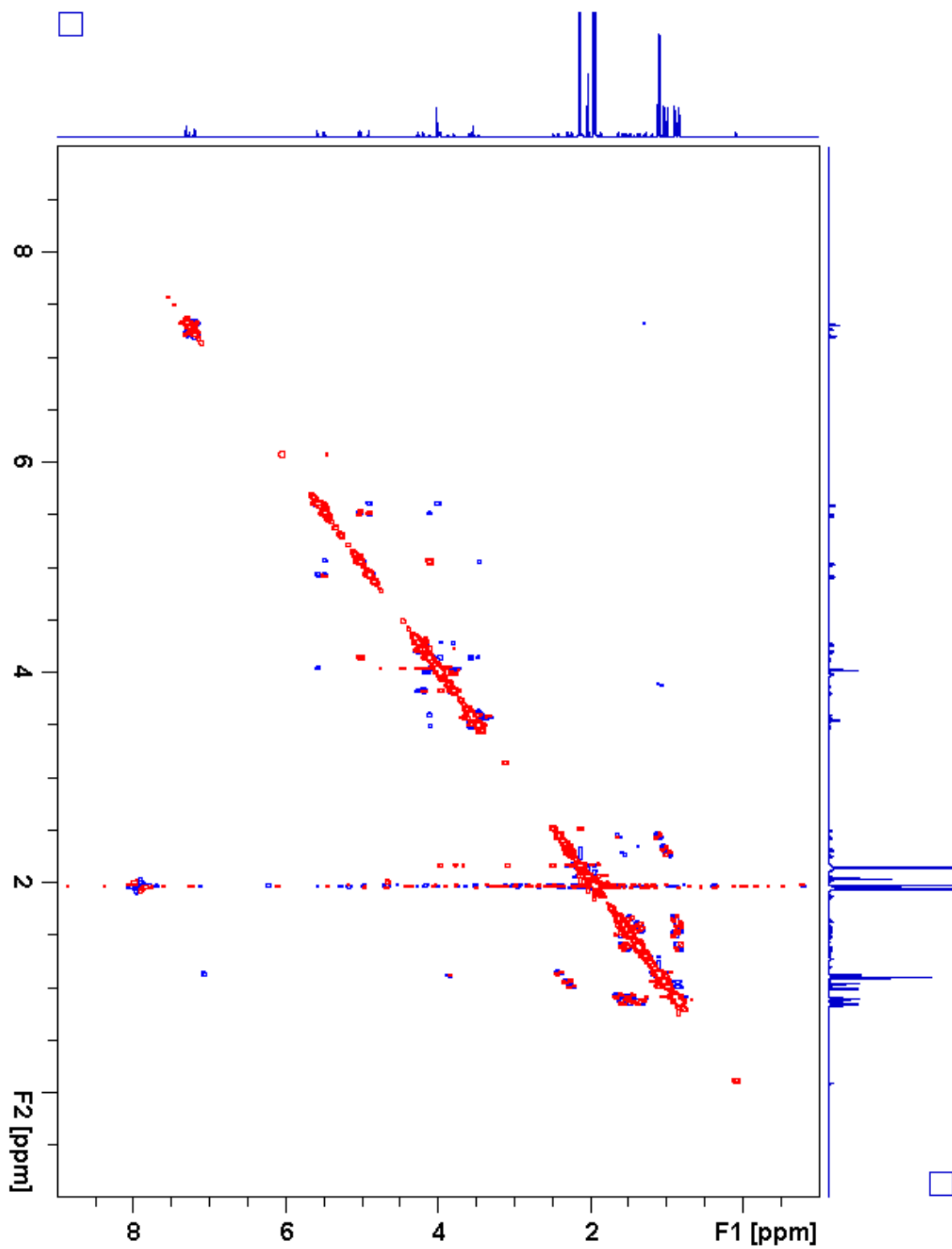
**Figure 2.99.** S5:25:4(2,5,5,5,8$^P$) ROESY

**Table 2.22.** S5:22:0(2,2,6,6,6) Chemical shifts and coupling constants
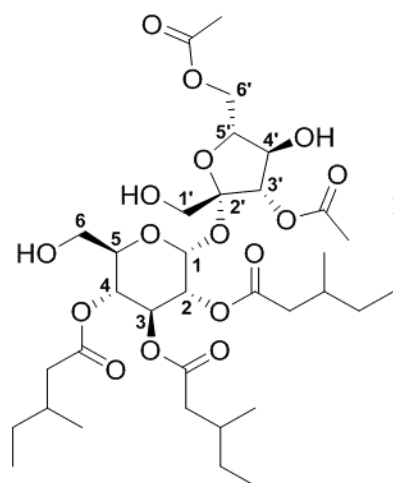


**Molecular Formula:** $C_{34}H_{56}O_{16}$

**110 min Retention Time (ESI+):** 56.08 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{34}H_{60}NO_{16}^+$ ([M+NH$_4^+$]): 738.3907, found: 738.3937

**Fraction, Batch:** #63, A-D
**Sample mass for NMR analysis:** 1.0 mg
**NMR Solvent:** CDCl$_3$
**InChi Key:** VVXKFHPDDJUGBH-NHDNHSBXSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.62 (d, $J$ = 3.6 Hz) | 89.58 |
| **2**(CH) | 4.90 (dd, $J$ = 10.5, 3.6 Hz) | 70.49 |
| -1(CO) | | 172.31 |
| -2(CH$_2$) | 2.30 (dd, $J$ = 15.5, 5.8 Hz), 2.05 (dd, $J$ = 15.4, 8.3 Hz) | 41.04 |
| -3(CH) | 1.80 (m) | 31.7[a] |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.4[b] |
| -5(CH$_2$) | 1.33 (m), 1.20 (m) | 29.4[c] |
| -6(CH$_3$) | 0.87 (t, $J$ = 7.4 Hz) | 11.4[d] |
| **3**(CH) | 5.54 (dd, $J$ = 10.7, 9.3 Hz) | 69.0[e] |
| -1(CO) | | 172.35 |
| -2(CH$_2$) | 2.24 (dd, $J$ = 15.3, 5.7 Hz), 2.03 (dd, $J$ = 15.3, 8.4 Hz) | 41.35 |
| -3(CH) | 1.82 (m) | 31.7[a] |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.4[b] |
| -5(CH$_2$) | 1.33 (m), 1.20 (m) | 29.4[c] |
| -6(CH$_3$) | 0.87 (, $J$ = 7.4 Hz) | 11.4[d] |
| **4**(CH) | 4.98 (dd, $J$ = 10.5, 9.3 Hz) | 69.0[e] |
| -1(CO) | | 172.96 |
| -2(CH$_2$) | 2.29 (dd, $J$ = 15.3, 5.8 Hz), 2.10 (dd, $J$ = 15.3, 8.4 Hz) | 41.27 |
| -3(CH) | 1.83 (m) | 31.7[a] |
| -4(CH$_3$) | 0.91 (d, $J$ = 6.7 Hz) | 19.4[b] |
| -5(CH$_2$) | 1.33 (m), 1.20 (m) | 29.4[c] |
| -6(CH$_3$) | 0.87 (, $J$ = 7.4 Hz) | 11.4[d] |
| **5**(CH) | 4.04 (ddd, $J$ = 10.2, 5.1, 2.2 Hz) | 71.17 |
| **6**(CH$_2$) | 3.67 (dd, $J$ = 12.8, 2.6 Hz), 3.60 (dd, $J$ = 12.8, 5.1 Hz) | 61.67 |

189

**Table 2.22.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 3.63 (d, $J$ = 12.4 Hz), 3.55 (d, $J$ = 12.4 Hz) | 64.23 |
| **2'**(C) | | 104.66 |
| **3'**(CH) | 5.18 (d, $J$ = 7.6 Hz) | 79.56 |
| -1(CO) | | 172.15 |
| -2(CH$_3$) | 2.22 (s) | 20.0[f] |
| **4'**(CH) | 4.35 (t, $J$ = 7.6 Hz) | 74.04 |
| **5'**(CH) | 4.08 (ddd, $J$ = 7.7, 6.2, 3.6 Hz) | 80.39 |
| **6'**(CH$_2$) | 4.43 (dd, $J$ = 12.1, 6.2 Hz), 4.28 (dd, $J$ = 12.1, 3.6 Hz) | 64.27 |
| -1(CO) | | 171.68 |
| -2(CH$_3$) | 2.13 (s) | 20.0[f] |

a - Three ^{13}C signals not resolved in 2D spectra (31.64, 31.74, 31.85 ppm)

b - Three ^{13}C signals not resolved in 2D spectra (19.34, 19.38, 19.42 ppm)

c - Three ^{13}C signals not resolved in 2D spectra (29.35, 29.40, 29.42 ppm)

d - Three ^{13}C signals not resolved in 2D spectra (11.39, 11.39, 11.42 ppm)

e - Two ^{13}C signals not resolved in 2D spectra (68.96, 68.97 ppm)

f - Two ^{13}C signals not resolved in 2D spectra (21.01, 21.02 ppm)
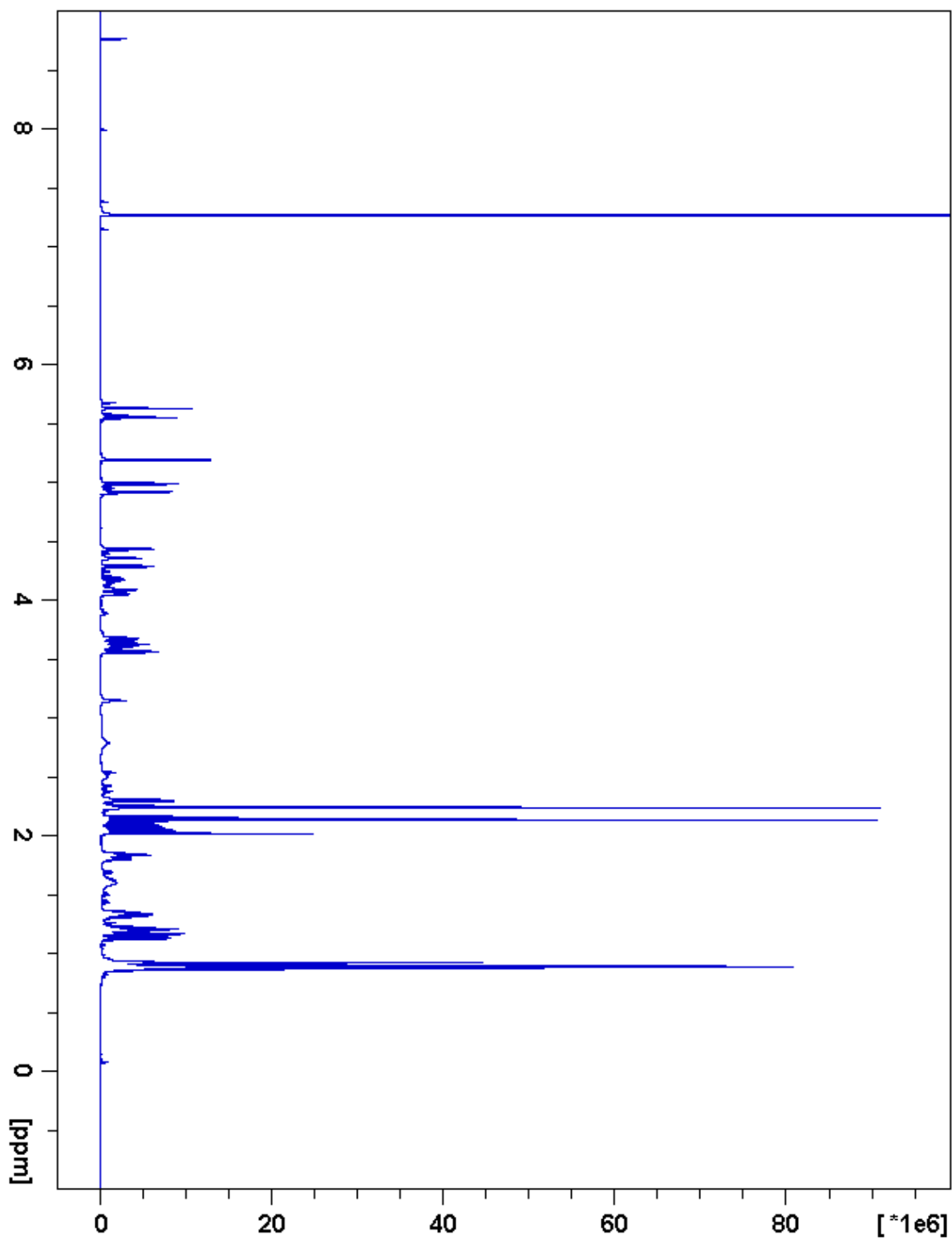
**Figure 2.100.** S5:22:0(2,2,6,6,6) ^{1}H NMR
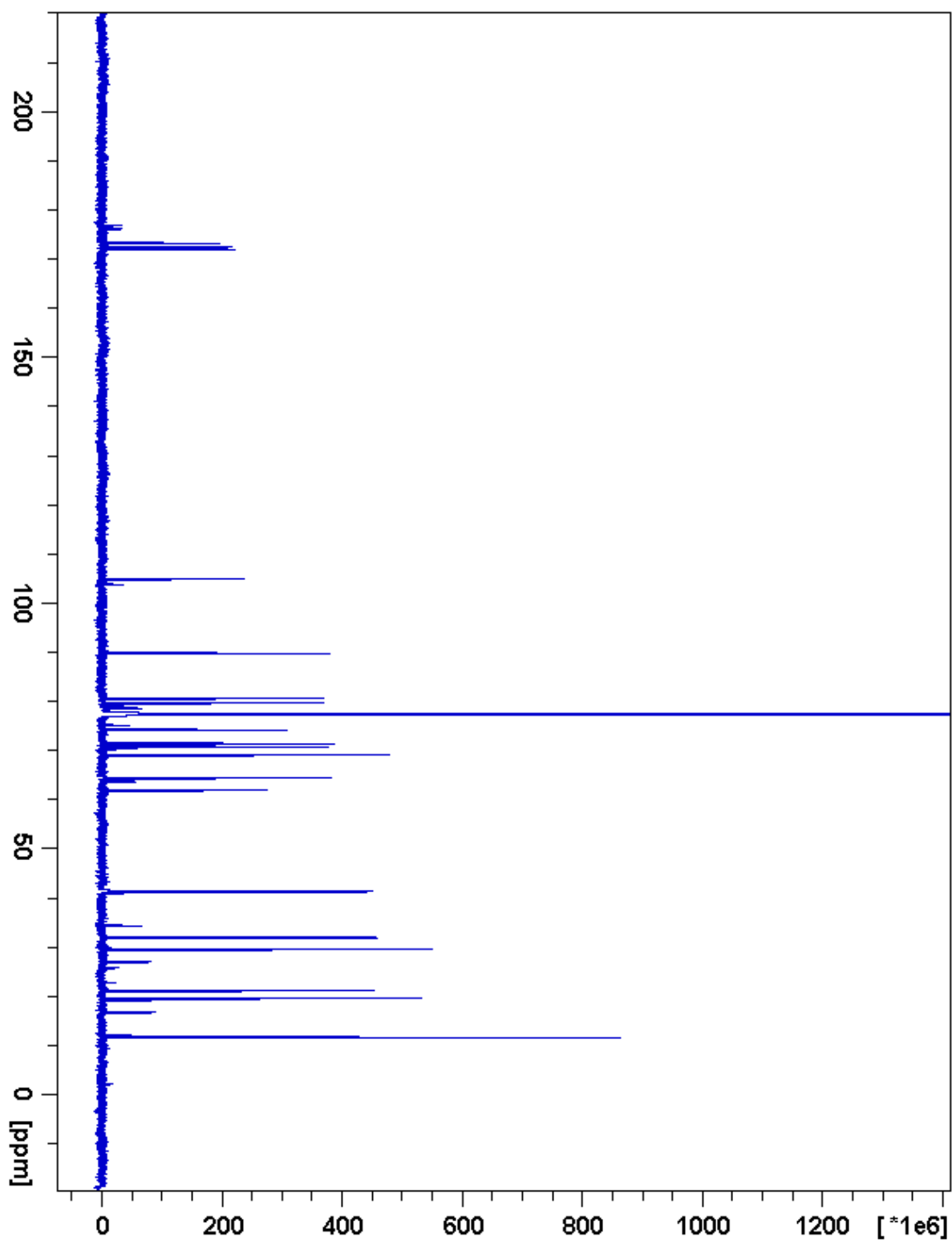
**Figure 2.101.** S5:22:0(2,2,6,6,6) $^{13}$C NMR
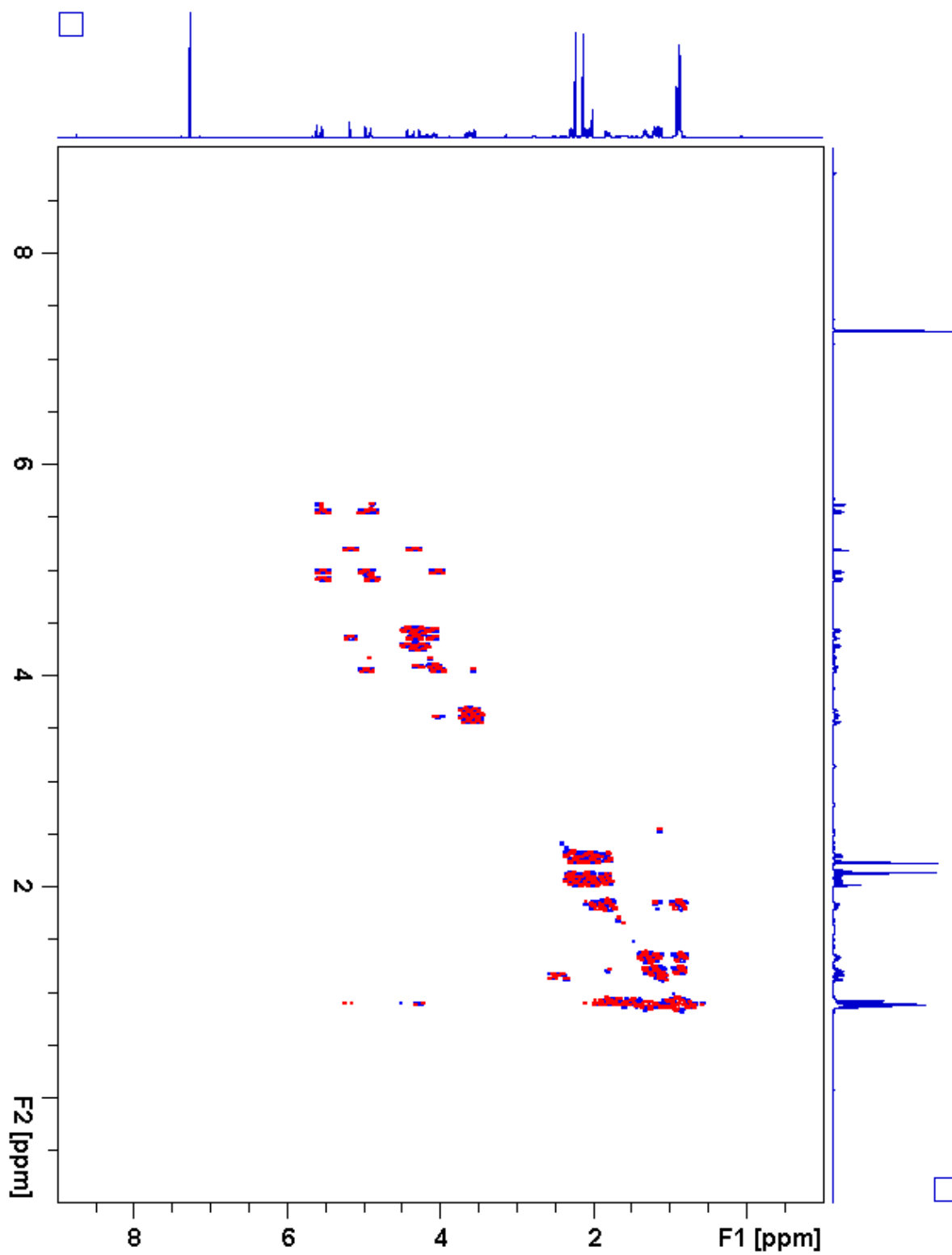
**Figure 2.102.** S5:22:0(2,2,6,6,6) $^1$H-$^1$H gCOSY
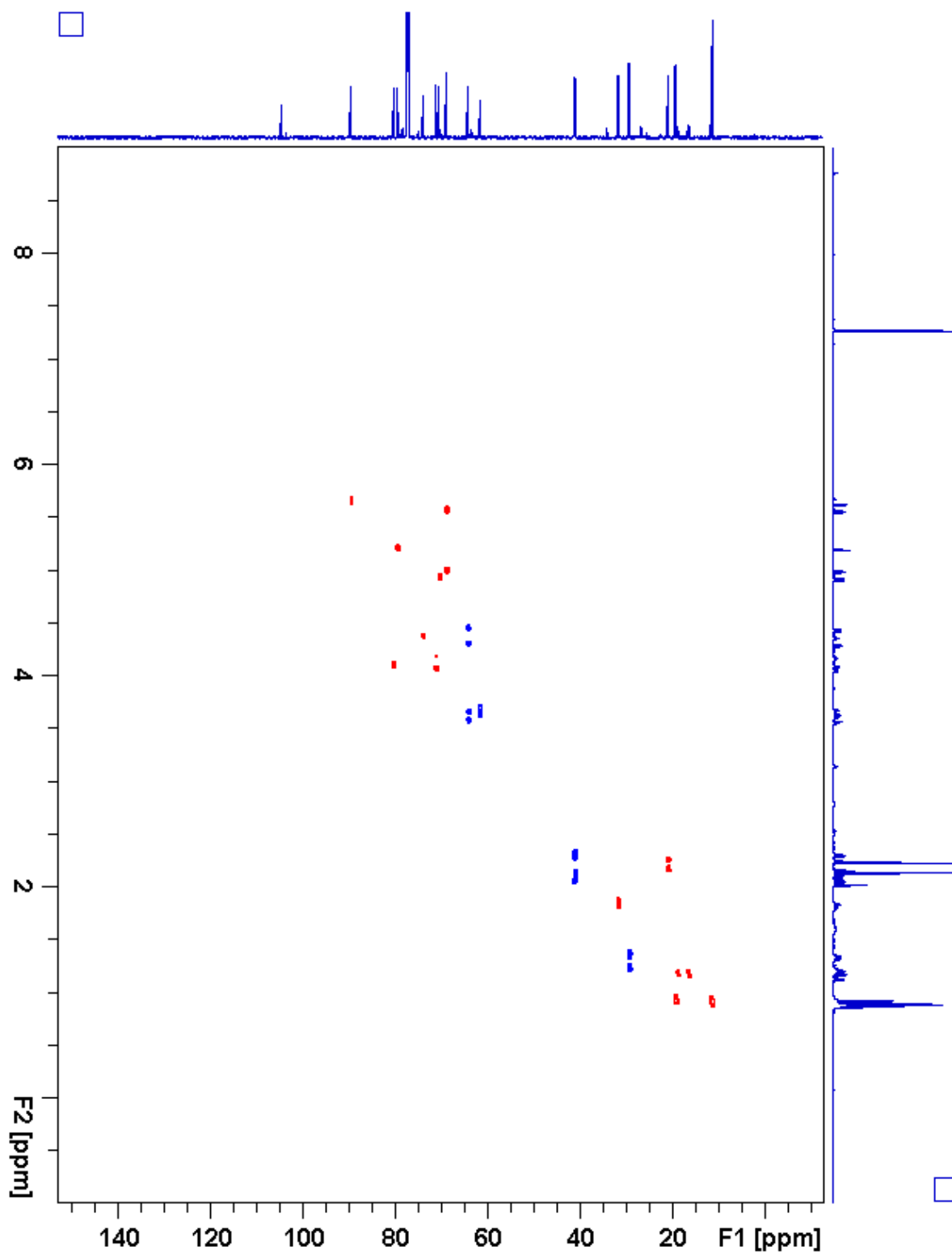
**Figure 2.103.** S5:22:0(2,2,6,6,6) gHSQC
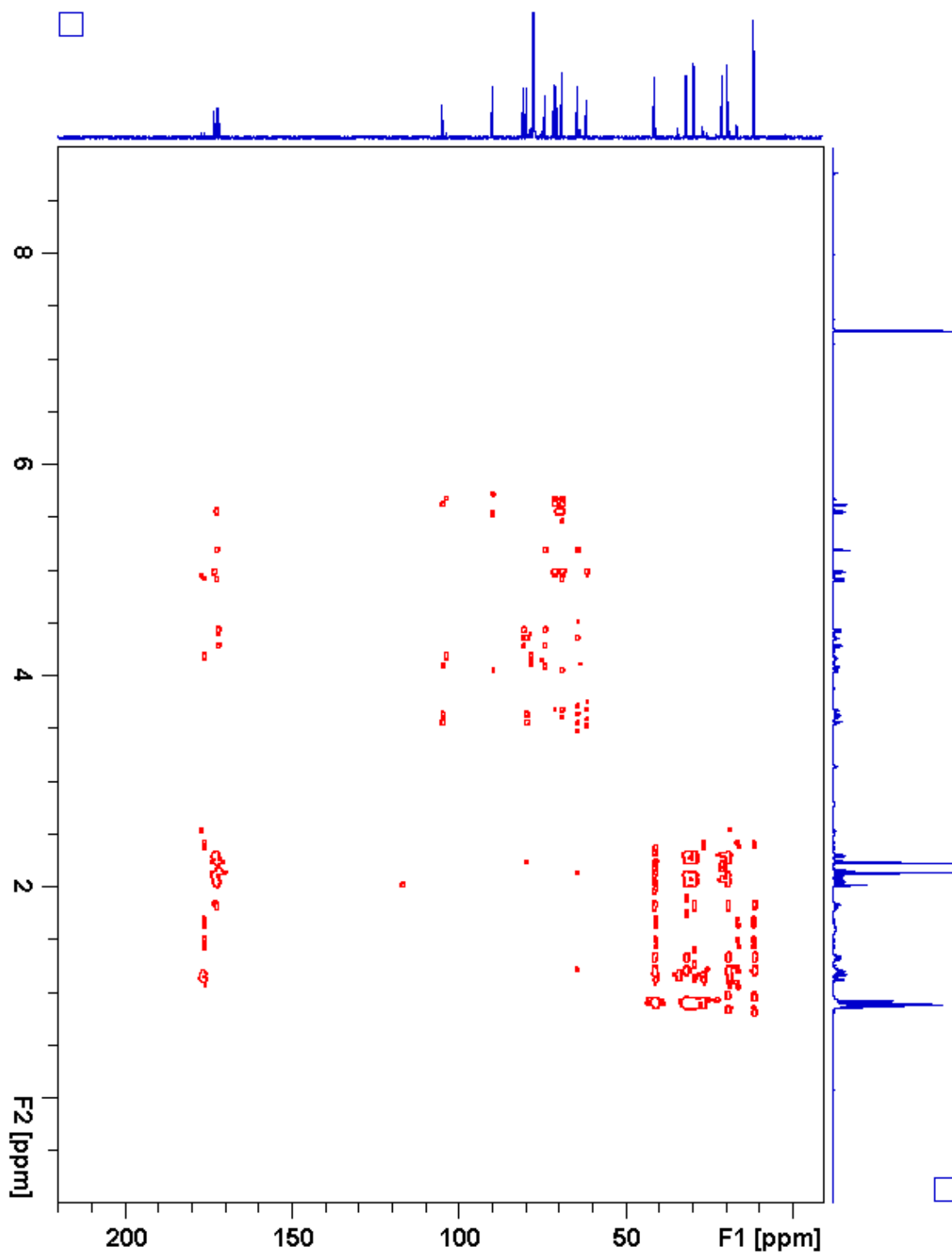
**Figure 2.104.** S5:22:0(2,2,6,6,6) gHMBC
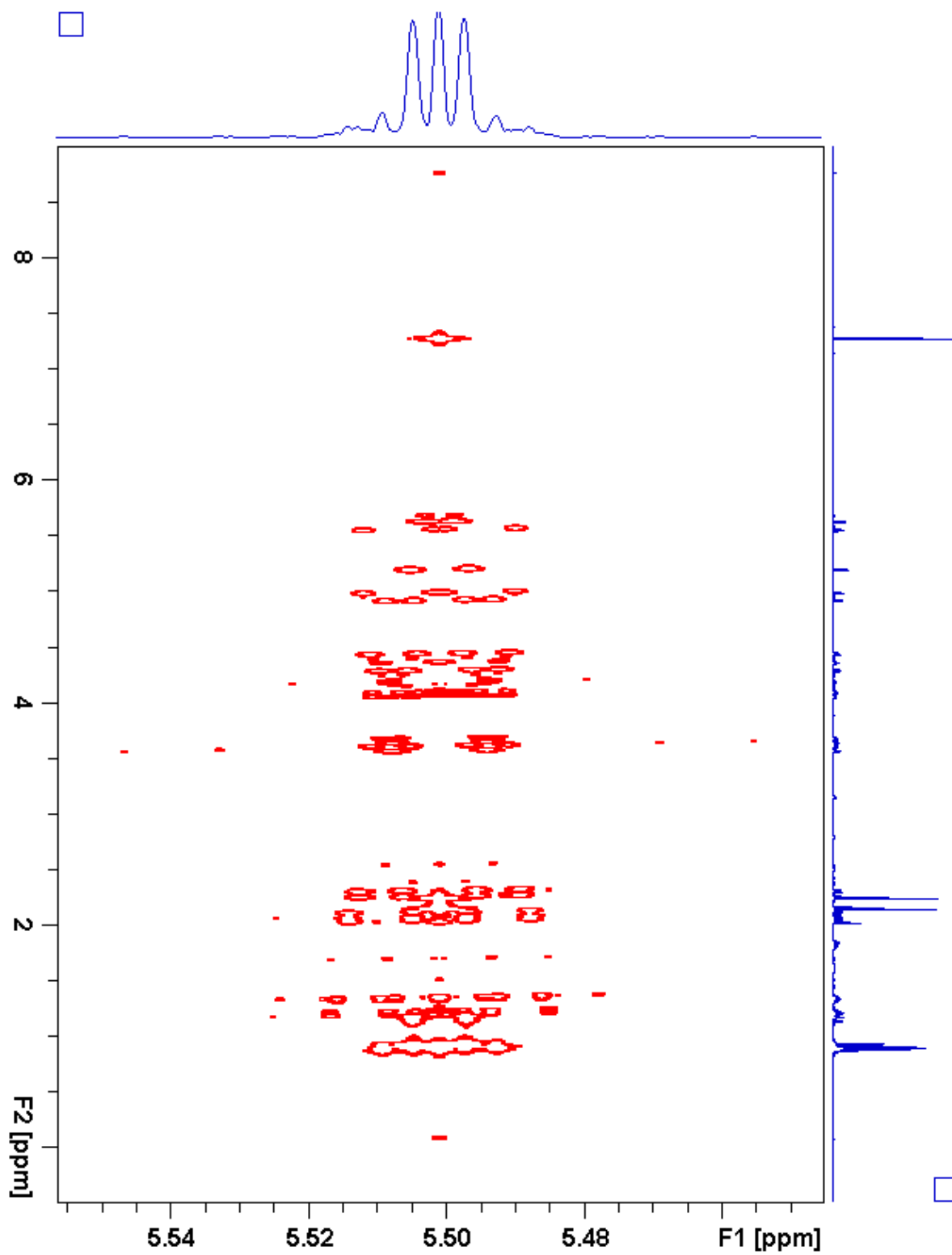
**Figure 2.105.** S5:22:0(2,2,6,6,6) *J*-resolved
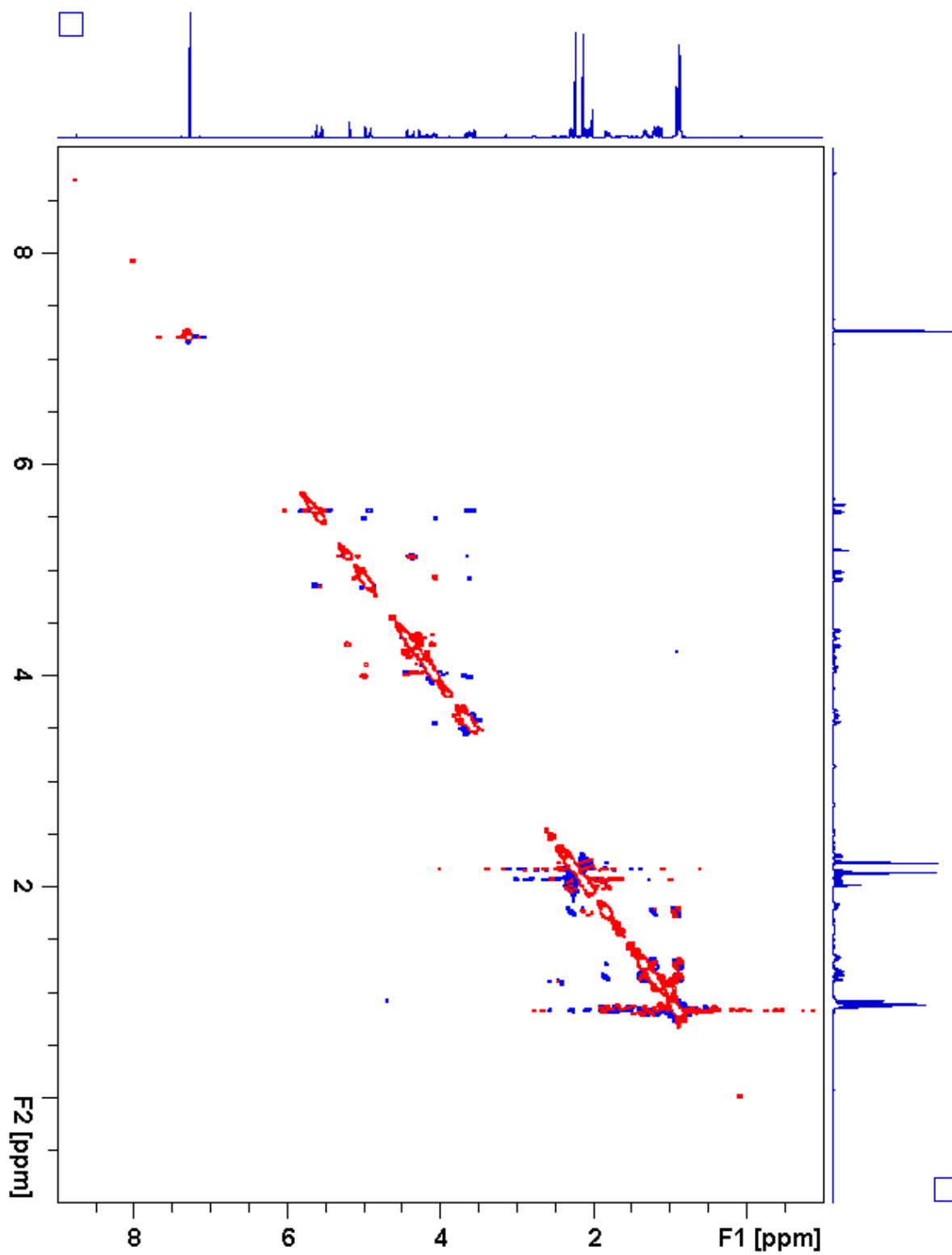
**Figure 2.106.** S5:22:0(2,2,6,6,6) ROESY

**Table 2.23.** S4:22:0(5,5,6,6) Chemical shifts and coupling constants
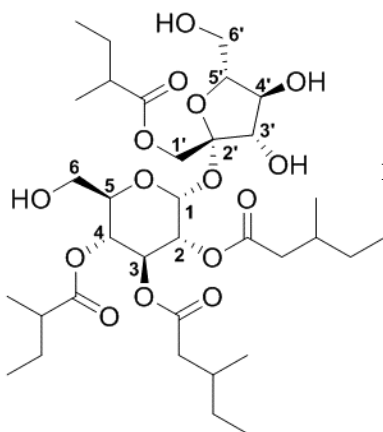


**Molecular Formula:** $C_{34}H_{58}O_{15}$

**110 min Retention Time (ESI+):** 56.49 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{34}H_{62}NO_{15}^+$ ([M+NH$_4^+$]): 724.4114, found: 724.4154

**Fraction, Batch:** #66, A-B

**Sample mass for NMR analysis:** 1.3 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** CACGEXBARUEDGB-FJNLHFDOSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.81 (d, *J* = 4.0 Hz) | 89.00 |
| **2**(CH) | 4.85 (dd, *J* = 10.4, 4.0 Hz) | 70.81 |
| -1(CO) | | 173.79 |
| -2(CH$_2$) | 2.39 (dd, *J* = 15.7, 5.6 Hz), 2.12 (dd, *J* = 15.7, 8.4 Hz) | 41.1[a] |
| -3(CH) | 1.83 (m) | 31.65 |
| -4(CH$_3$) | 0.88 (d, *J* = 6.7 Hz) | 19.4[b] |
| -5(CH$_2$) | 1.35 (m), 1.21 (m) | 29.32 |
| -6(CH$_3$) | 0.87 (t, *J* = 7.4 Hz) | 11.36 |
| **3**(CH) | 5.54 (dd, *J* = 10.6, 9.2 Hz) | 68.82 |
| -1(CO) | | 172.23 |
| -2(CH$_2$) | 2.24 (dd, *J* = 15.5, 5.6 Hz), 2.03 (dd, *J* = 15.5, 8.4 Hz) | 41.1[a] |
| -3(CH) | 1.80 (m) | 31.72 |
| -4(CH$_3$) | 0.88 (d, *J* = 6.7 Hz) | 19.4[b] |
| -5(CH$_2$) | 1.31 (m), 1.19 (m) | 29.37 |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.39 |
| **4**(CH) | 4.93 (dd, *J* = 10.7, 9.3 Hz) | 68.55 |
| -1(CO) | | 176.18 |
| -2(CH) | 2.37 (sextet, *J* = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.10 (d, *J* = 7.0 Hz) | 16.48 |
| -4(CH$_2$) | 1.67 (m), 1.44 (m) | 26.63 |
| -5(CH$_3$) | 0.89 (t, *J* = 7.4 Hz) | 11.83 |
| **5**(CH) | 4.14 (ddd, *J* = 10.3, 6.3, 1.9 Hz) | 72.15 |
| **6**(CH$_2$) | 3.63 (d, *J* = 12.7, 2.5 Hz), 3.58 (d, *J* = 12.7, 6.3 Hz) | 61.52 |

**Table 2.23.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.09 (d, $J$ = 11.7 Hz), 4.07 (d, $J$ = 11.7 Hz) | 63.80 |
| -1(CO) | | 175.90 |
| -2(CH) | 2.41 (sextet, $J$ = 7.0 Hz) | 41.1$^a$ |
| -3(CH$_3$) | 1.15 (d, $J$ = 7.0 Hz) | 16.74 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH$_3$) | 0.91 (t, $J$ = 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.58 |
| **3'**(CH) | 4.17 (d, $J$ = 8.8 Hz) | 78.39 |
| **4'**(CH) | 4.31 (t, $J$ = 8.8 Hz) | 72.74 |
| **5'**(CH) | 3.69 (dt, $J$ = 8.9, 2.3 Hz) | 81.22 |
| **6'**(CH$_2$) | 3.88 (dd, $J$ = 13.3, 2.4 Hz), 3.71 (dd, $J$ = 13.3, 2.3 Hz) | 59.78 |

a - Four $^{13}$C signals not resolved in 2D spectra (41.11, 41.11, 41.14, 41.19 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (19.40, 19.41 ppm)
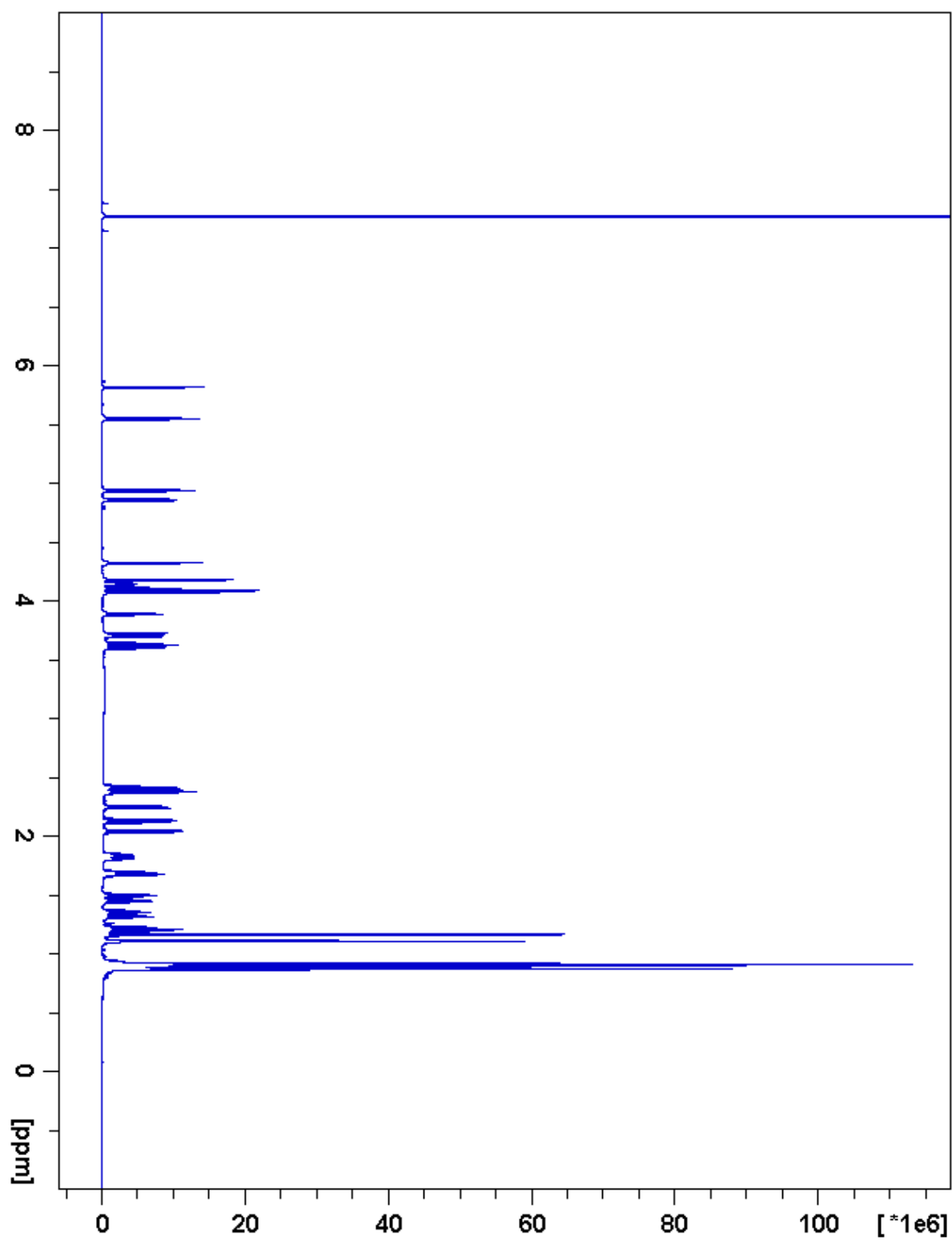
**Figure 2.107.** S4:22:0(5,5,6,6) ^{1}H NMR
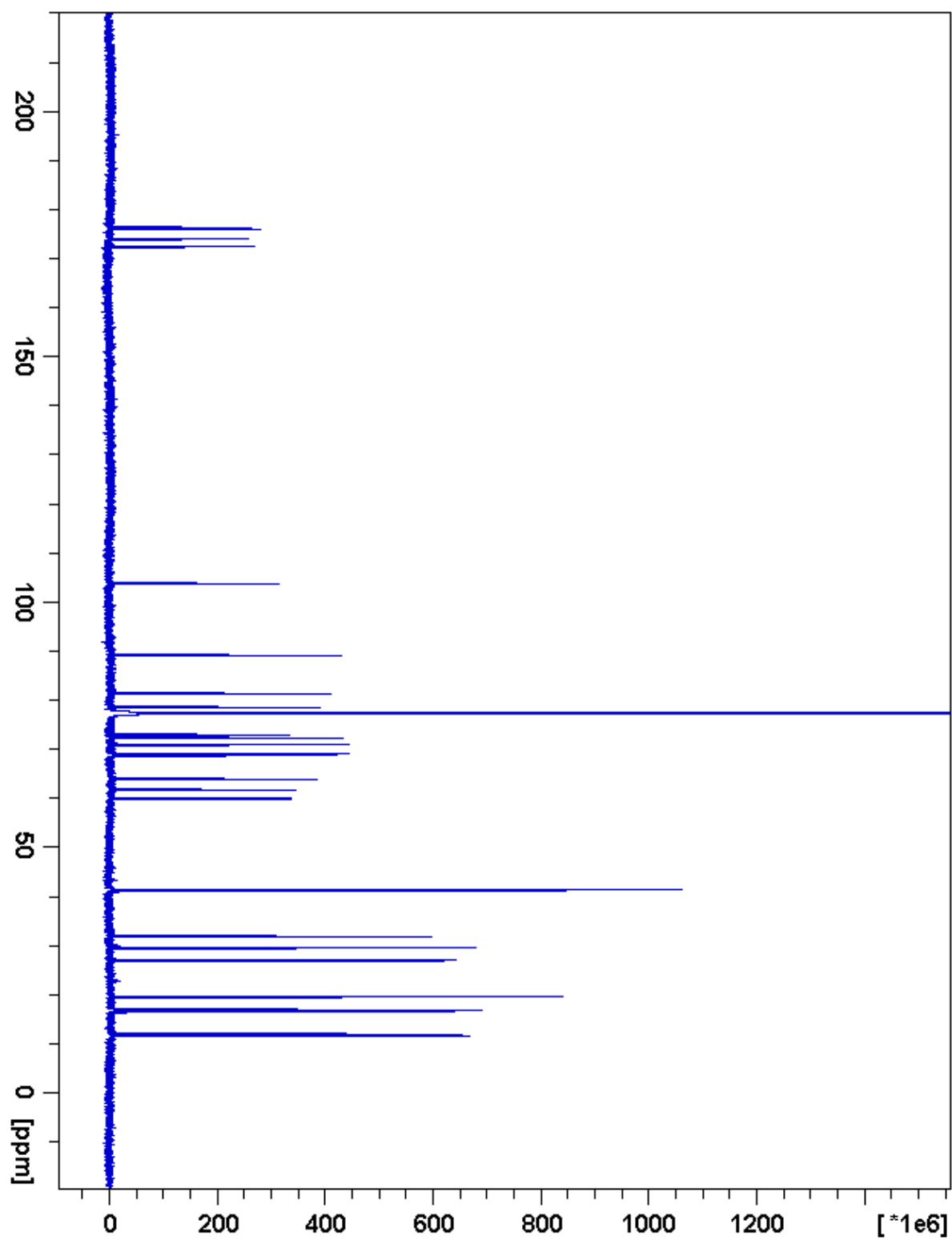
**Figure 2.108.** S4:22:0(5,5,6,6) $^{13}$C NMR
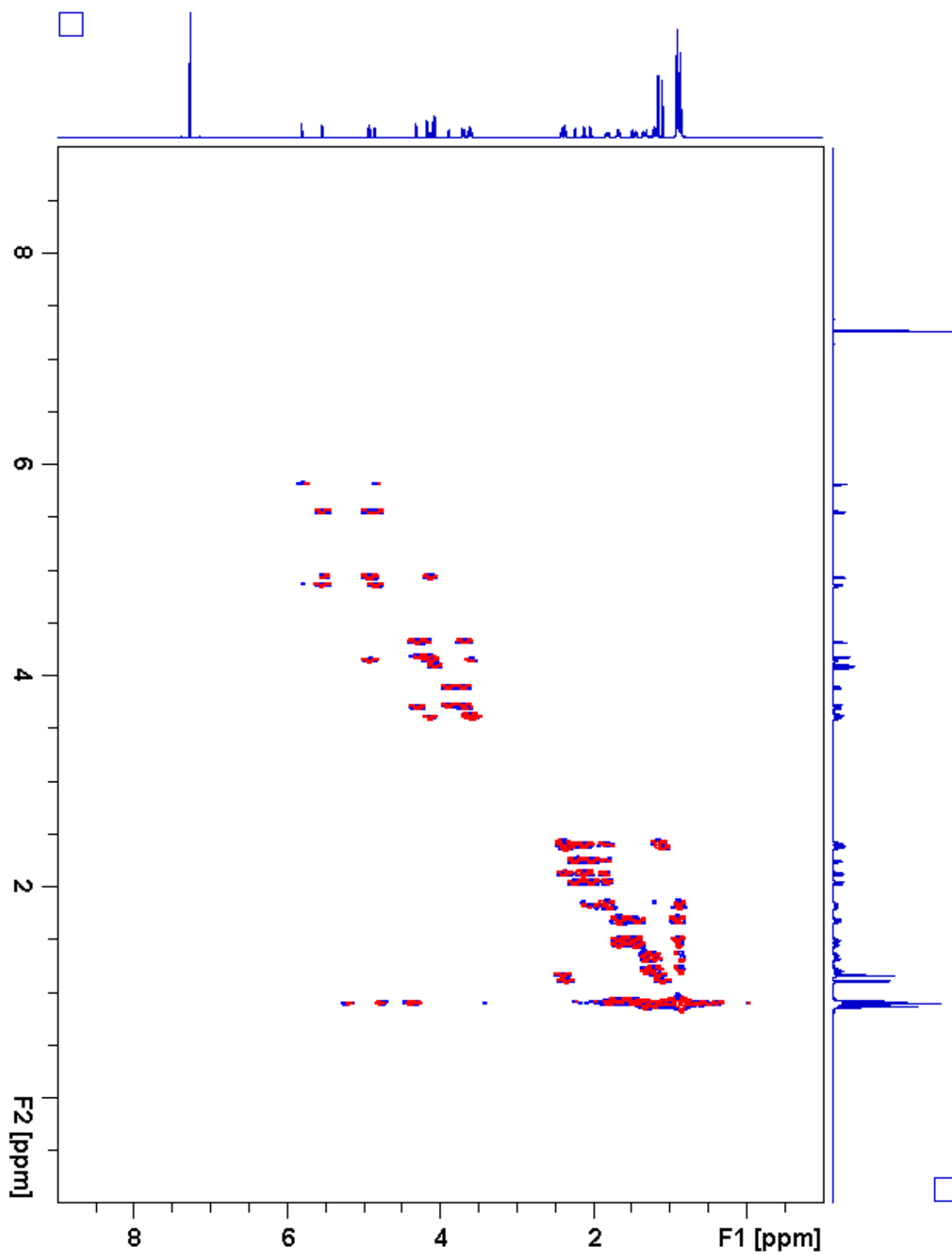
**Figure 2.109.** S4:22:0(5,5,6,6) $^1$H-$^1$H gCOSY
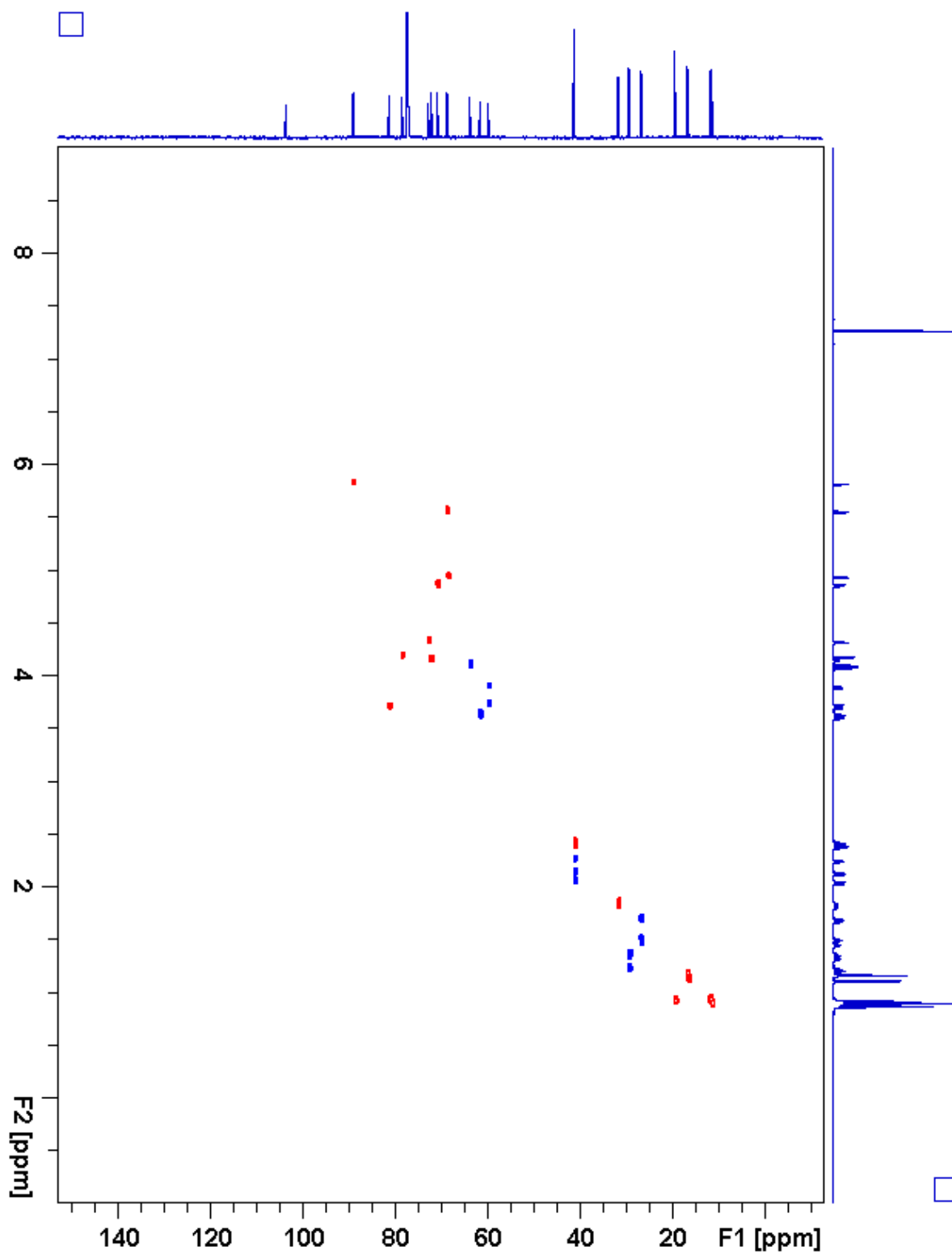
**Figure 2.110.** S4:22:0(5,5,6,6) gHSQC
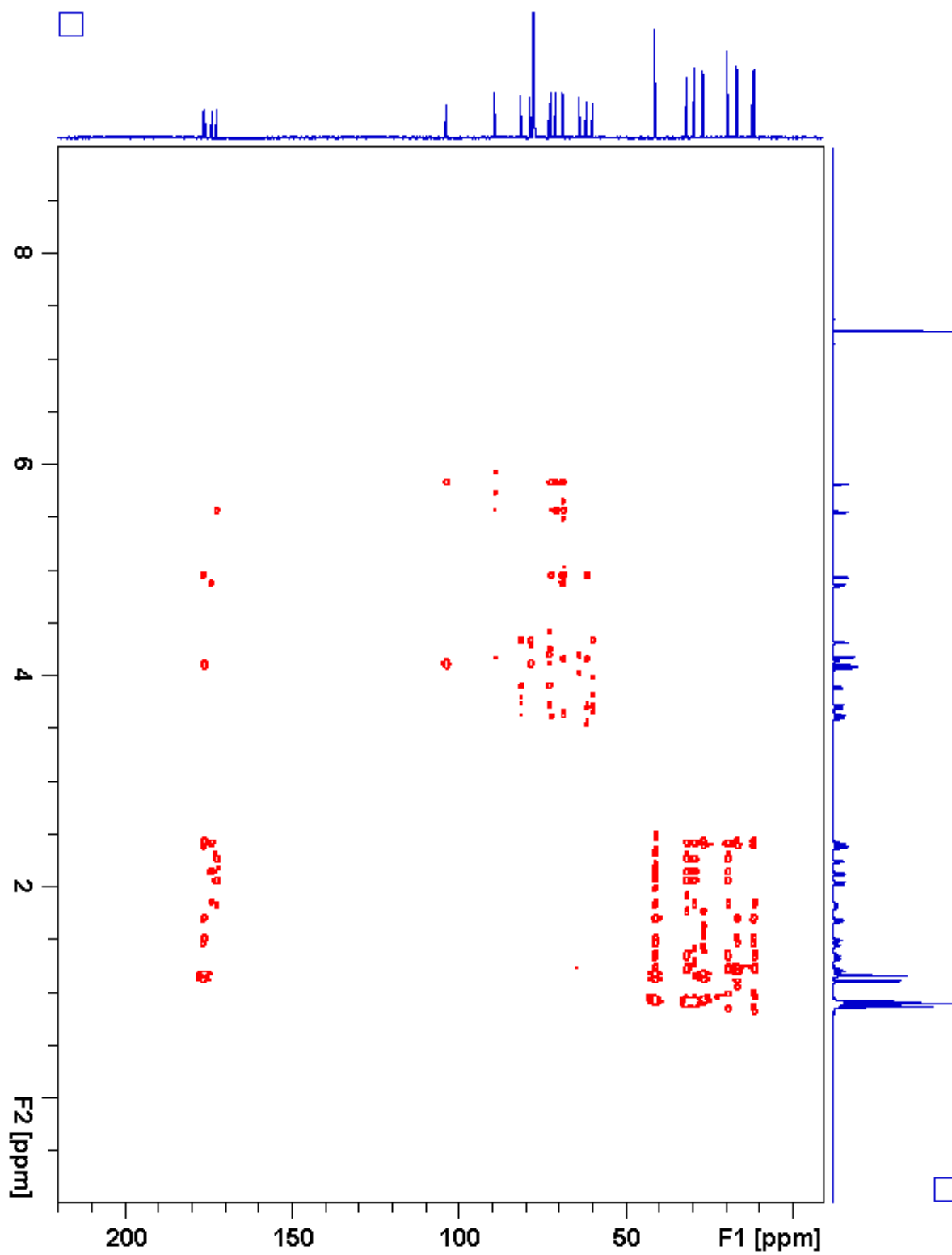
**Figure 2.111.** S4:22:0(5,5,6,6) gHMBC
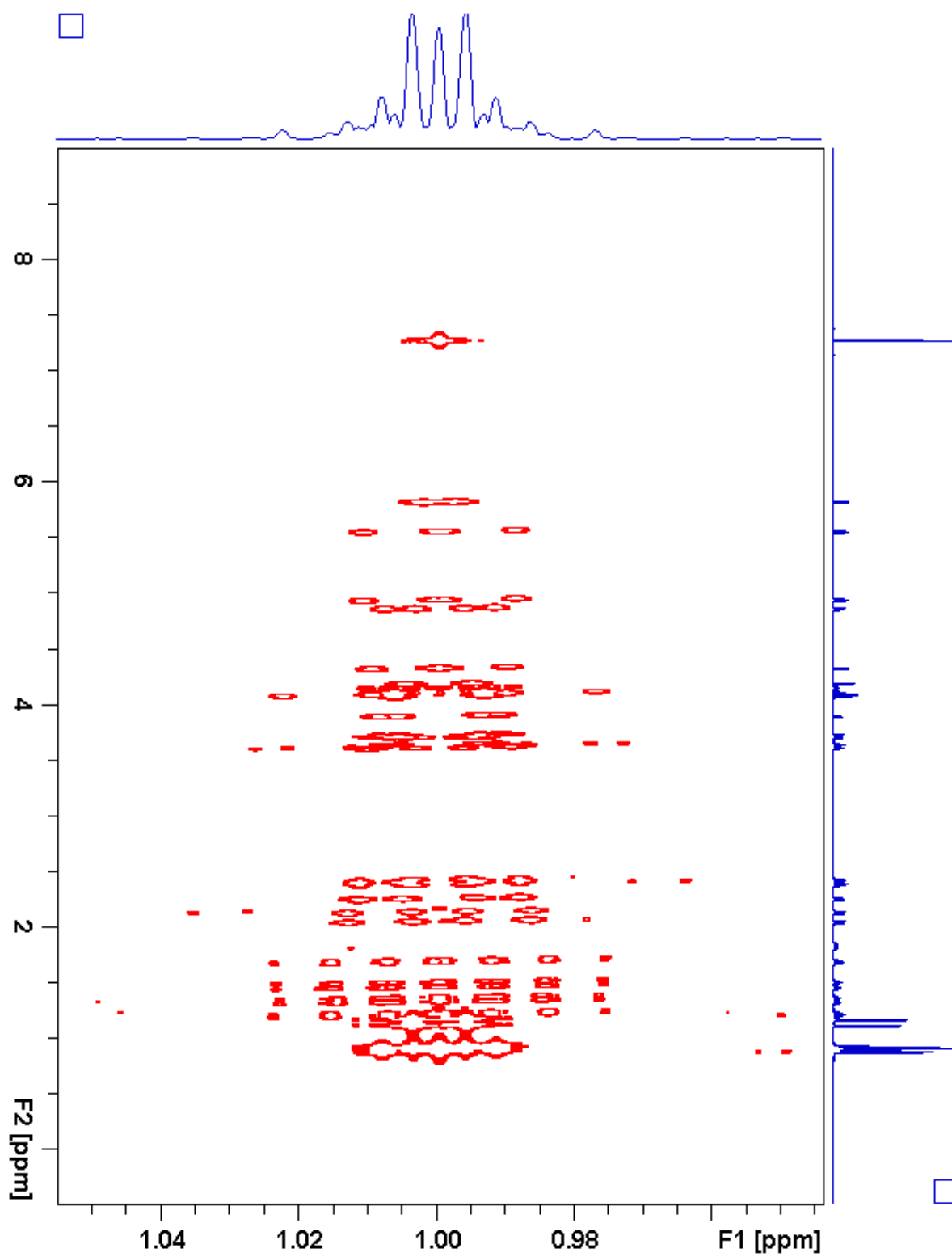
**Figure 2.112.** S4:22:0(5,5,6,6) *J*-resolved
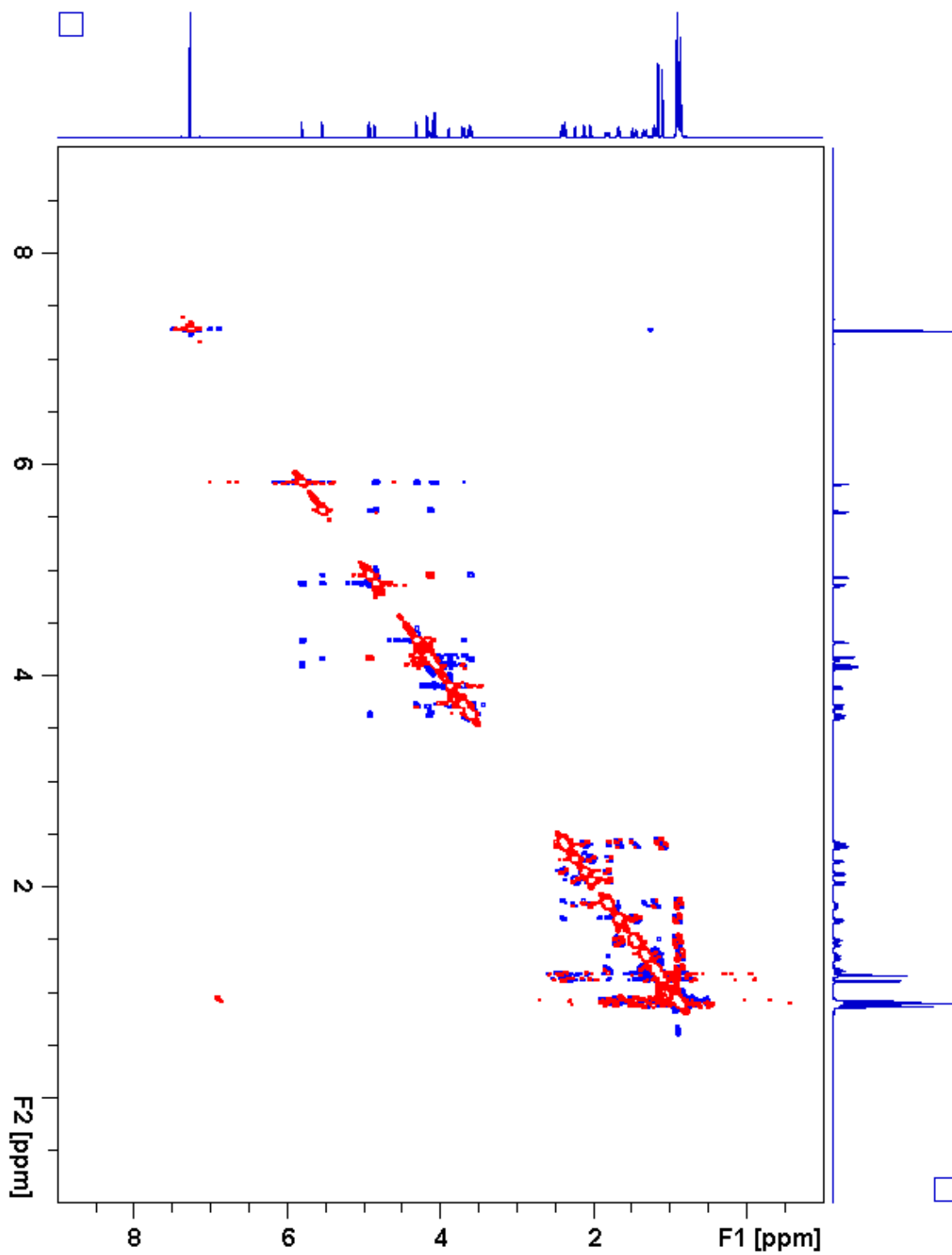
**Figure 2.113.** S4:22:0(5,5,6,6) ROESY

**Table 2.24.** S5:23:0(2,5,5,5,6) Chemical shifts and coupling constants
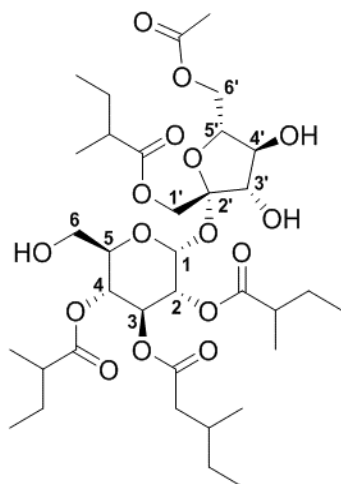


**Molecular Formula:** $C_{35}H_{58}O_{16}$

**110 min Retention Time (ESI+):** 58.80 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{35}H_{62}NO_{16}^{+}$ ([M+NH$_4^+$]): 752.4063, found: 752.4092

**Fraction, Batch:** #75, A-D

**Sample mass for NMR analysis:** 1.5 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** RTBHWDIHLBPIRS-YGWKQMRCSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.67 (d, $J$ = 3.9 Hz) | 89.36 |
| **2**(CH) | 4.90 (dd, $J$ = 10.4, 3.9 Hz) | 70.24 |
| -1(CO) | | 175.83 |
| -2(CH) | 2.37 (sextet, $J$ = 7.0 Hz) | 40.80 |
| -3(CH$_3$) | 1.12 (d, $J$ = 7.0 Hz) | 16.31 |
| -4(CH$_2$) | 1.62 (m), 1.42 (m) | 26.7[b] |
| -5(CH$_3$) | 0.86 (t, $J$ = 7.4 Hz) | 11.70 |
| **3**(CH) | 5.56 (dd, $J$ = 10.7, 9.2 Hz) | 68.88 |
| -1(CO) | | 172.29 |
| -2(CH$_2$) | 2.24 (dd, $J$ = 15.8, 5.4 Hz), 2.01 (dd, $J$ = 15.8, 8.6 Hz) | 41.1[a] |
| -3(CH) | 1.78 (m) | 31.51 |
| -4(CH$_3$) | 0.88 (d, $J$ = 6.7 Hz) | 19.44 |
| -5(CH$_2$) | 1.30 (m), 1.18 (m) | 29.40 |
| -6(CH$_3$) | 0.85 (t, $J$ = 7.4 Hz) | 11.38 |
| **4**(CH) | 4.94 (dd, $J$ = 10.7, 9.2 Hz) | 68.94 |
| -1(CO) | | 176.33 |
| -2(CH) | 2.34 (sextet, $J$ = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.10 (d, $J$ = 7.0 Hz) | 16.47 |
| -4(CH$_2$) | 1.67 (m), 1.44 (m) | 26.7[b] |
| -5(CH$_3$) | 0.90 (t, $J$ = 7.4 Hz) | 11.83 |
| **5**(CH) | 4.16 (m) | 71.28 |
| **6**(CH$_2$) | 3.65 (dd, $J$ = 12.5, 2.4 Hz), 3.59 (dd, $J$ = 12.5, 6.2 Hz) | 61.65 |

**Table 2.24.** (continued)

| | | |
|---|---|---|
| **1'**(CH₂) | 4.18 (d, *J* = 11.7 Hz), 4.15 (d, *J* = 11.7 Hz) | 63.70 |
| -1(CO) | | 176.01 |
| -2(CH) | 2.41 (sextet, *J* = 7.0 Hz) | 41.1ᵃ |
| -3(CH₃) | 1.16 (d, *J* = 7.0 Hz) | 16.69 |
| -4(CH₂) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH₃) | 0.92 (t, *J* = 7.4 Hz) | 11.74 |
| **2'**(C) | | 103.63 |
| **3'**(CH) | 4.14 (d, *J* = 8.8 Hz) | 78.11 |
| **4'**(CH) | 4.11 (t, *J* = 8.6 Hz) | 74.69 |
| **5'**(CH) | 3.87 (ddd, *J* = 8.3, 5.2, 2.8 Hz) | 78.63 |
| **6'**(CH₂) | 4.39 (dd, *J* = 12.4, 5.2 Hz), 4.24 (dd, *J* = 12.4, 2.9 Hz) | 63.49 |
| -1(CO) | | 171.74 |
| -2(CH₃) | 2.15 (s) | 20.99 |

a - Three $^{13}$C signals not resolved in 2D spectra (41.12, 41.12, 41.13 ppm)

a - Two $^{13}$C signals not resolved in 2D spectra (26.65, 26.66 ppm)
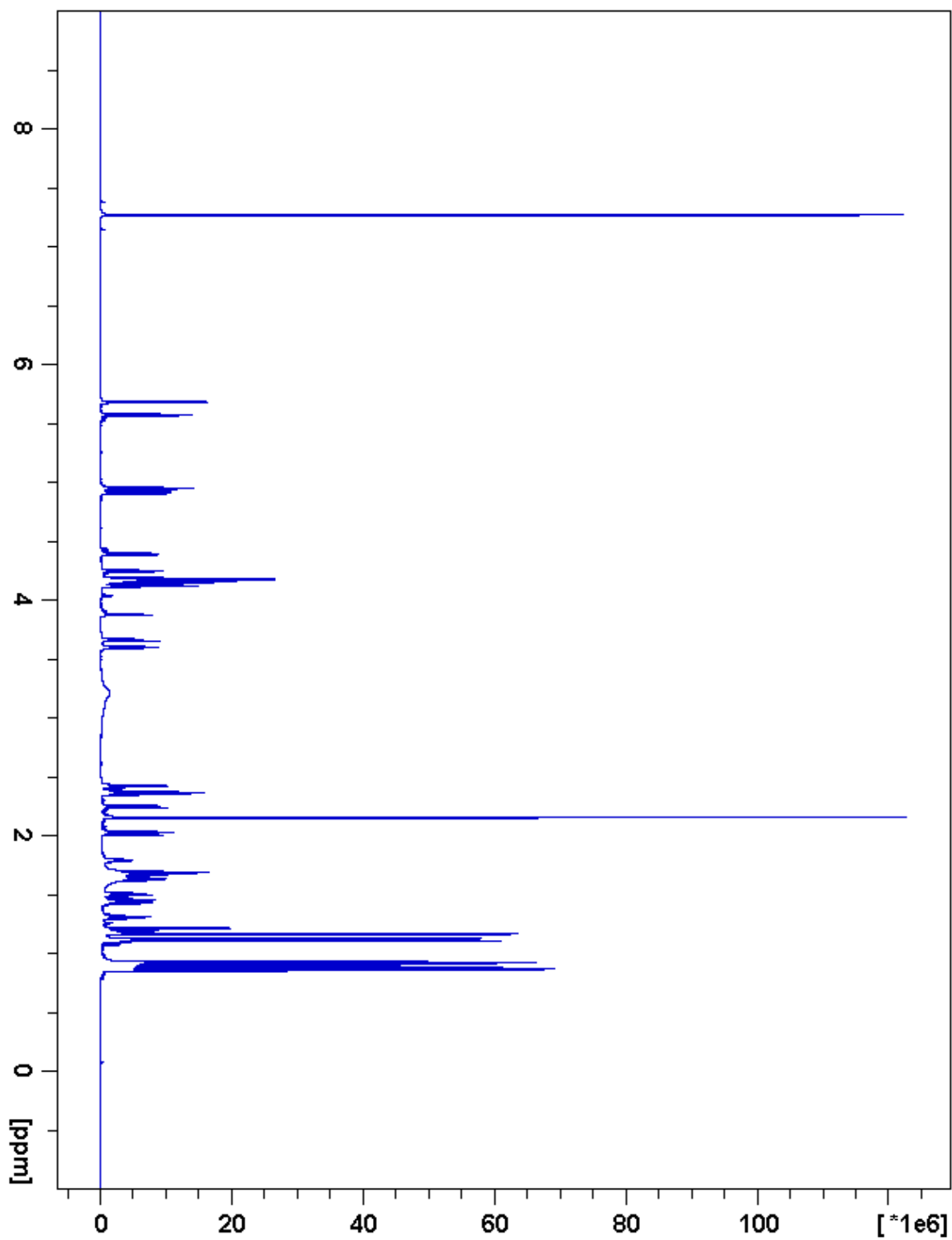
**Figure 2.114.** S5:23:0(2,5,5,5,6) ^{1}H NMR
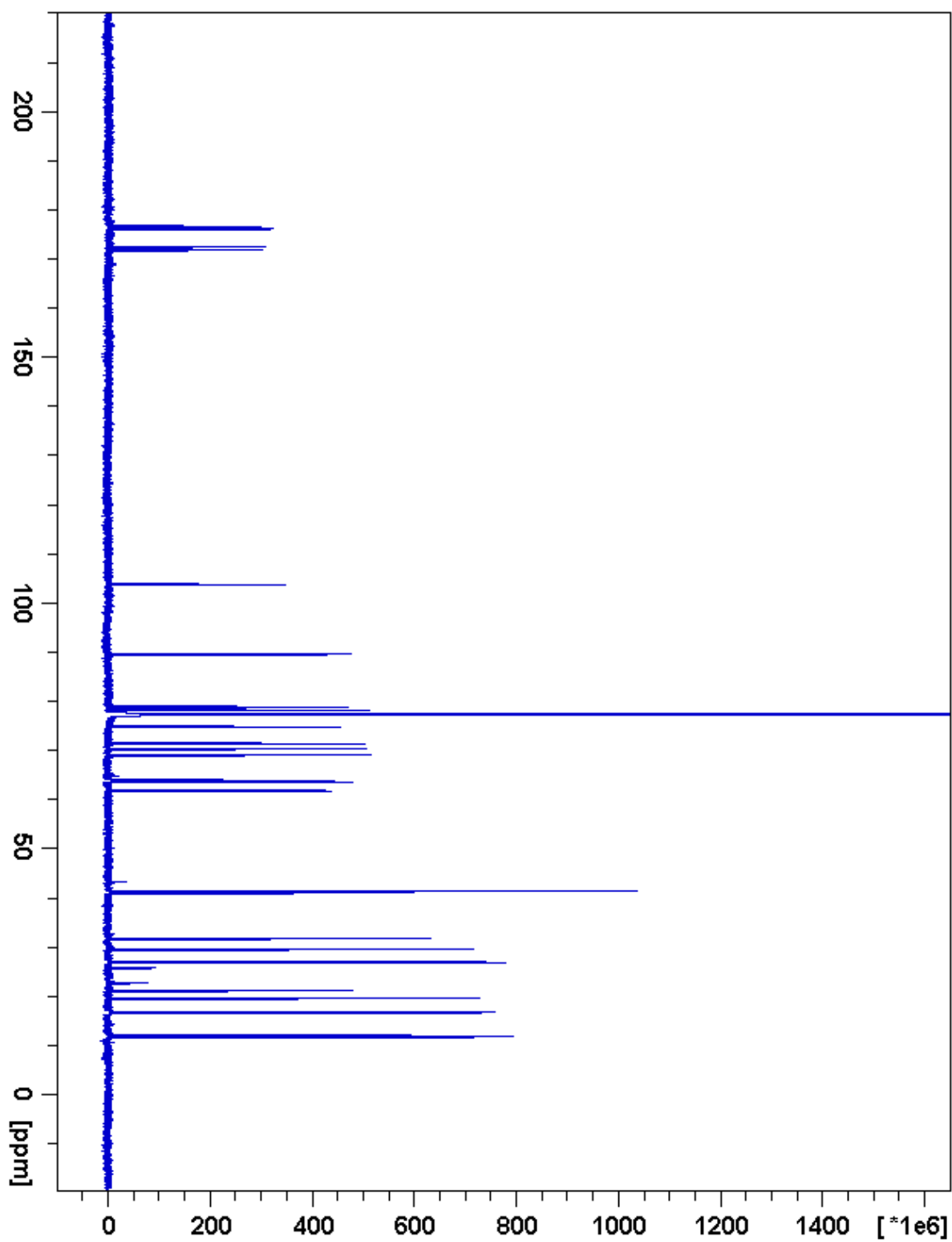
**Figure 2.115.** S5:23:0(2,5,5,5,6) ^{13}C NMR
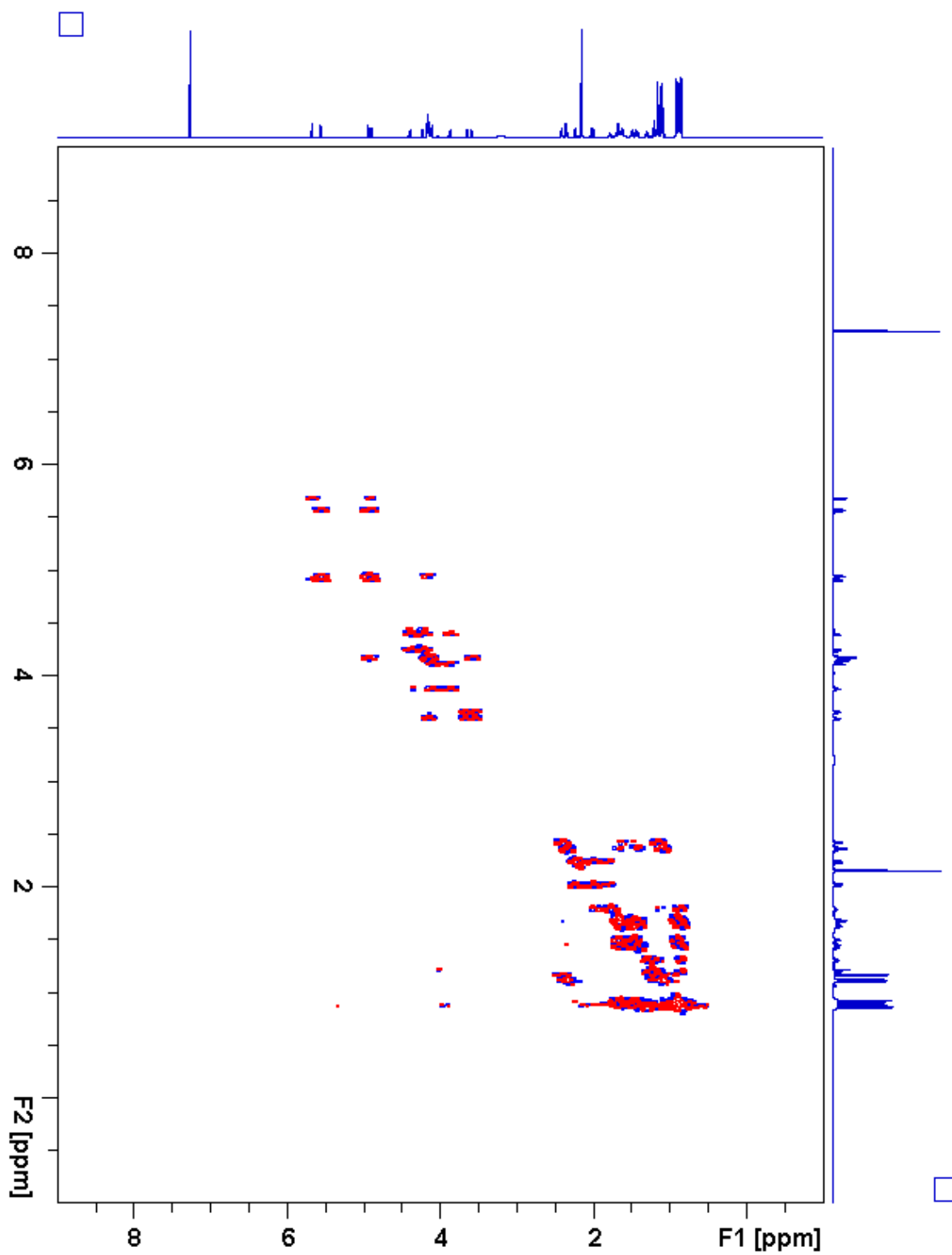
**Figure 2.116.** S5:23:0(2,5,5,5,6) $^1$H-$^1$H gCOSY
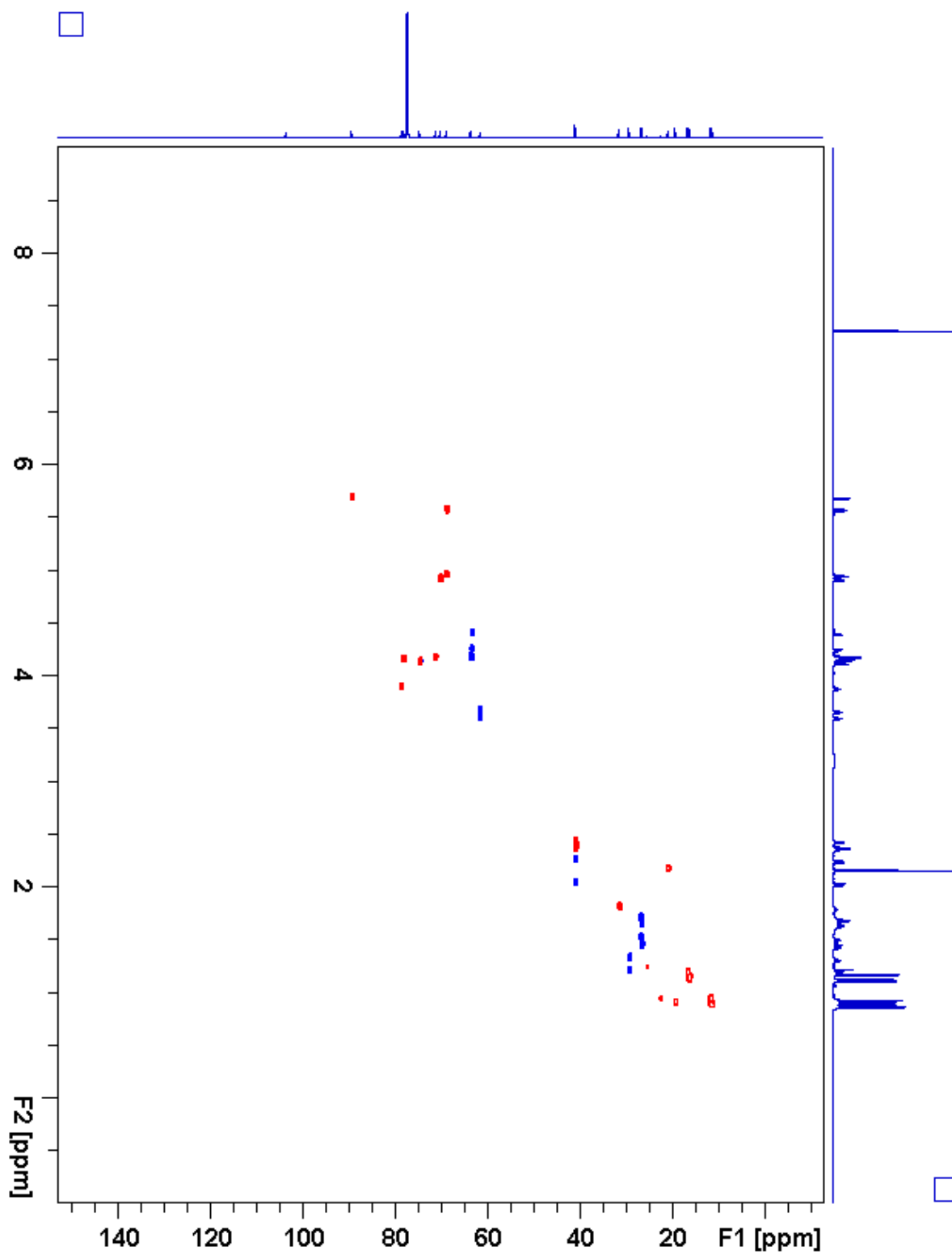
**Figure 2.117.** S5:23:0(2,5,5,5,6) gHSQC
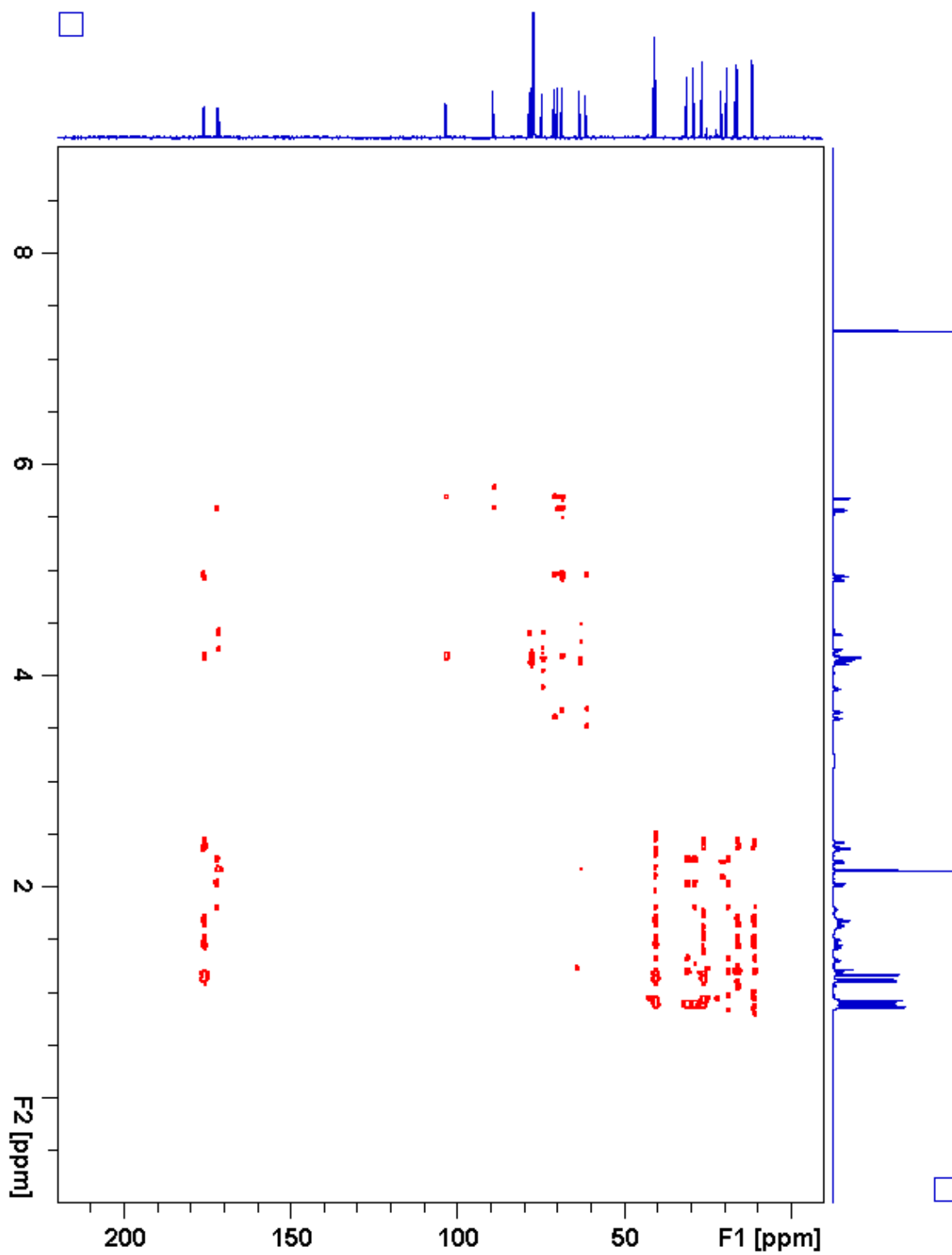
**Figure 2.118.** S5:23:0(2,5,5,5,6) gHMBC
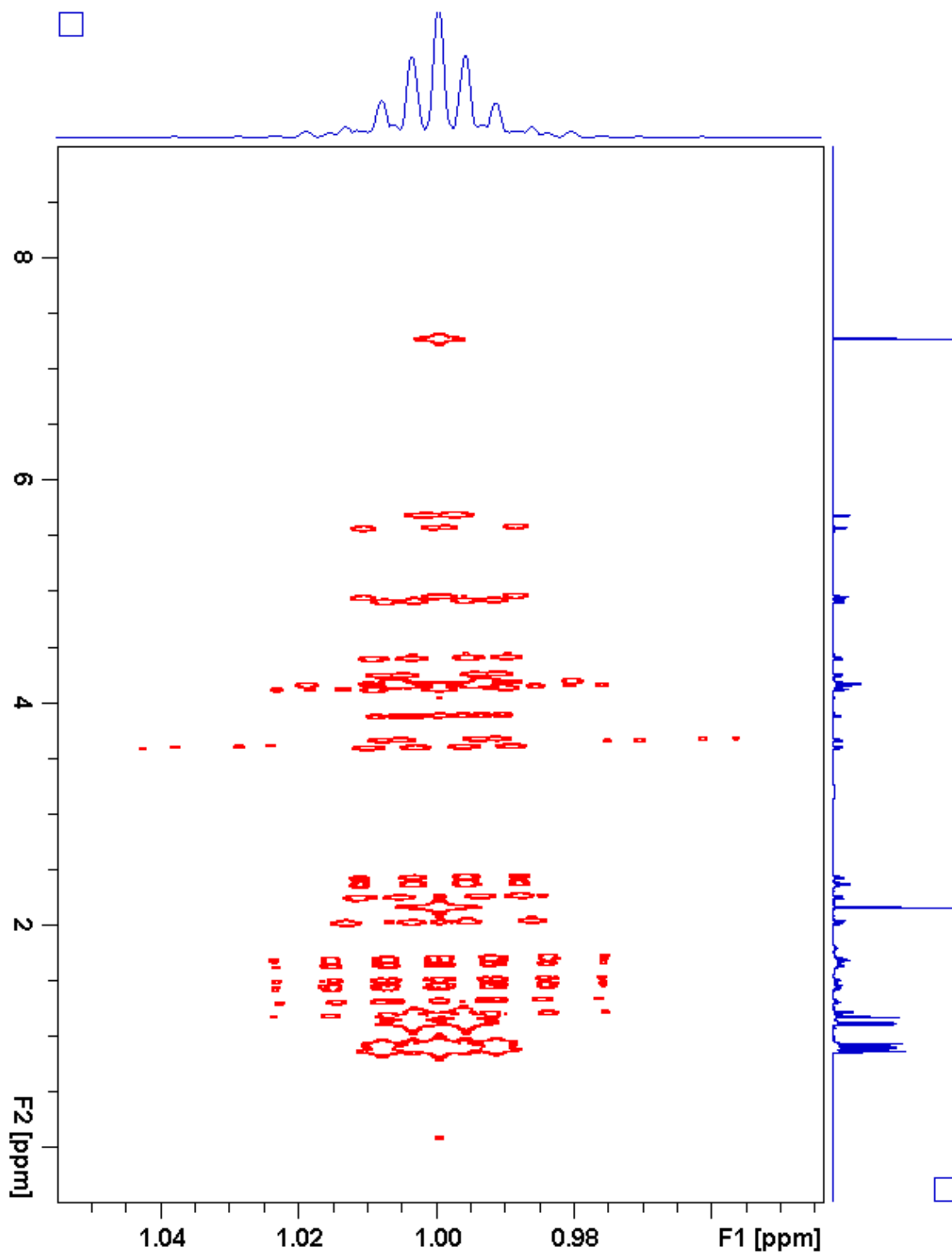
**Figure 2.119.** S5:23:0(2,5,5,5,6) *J*-resolved

214
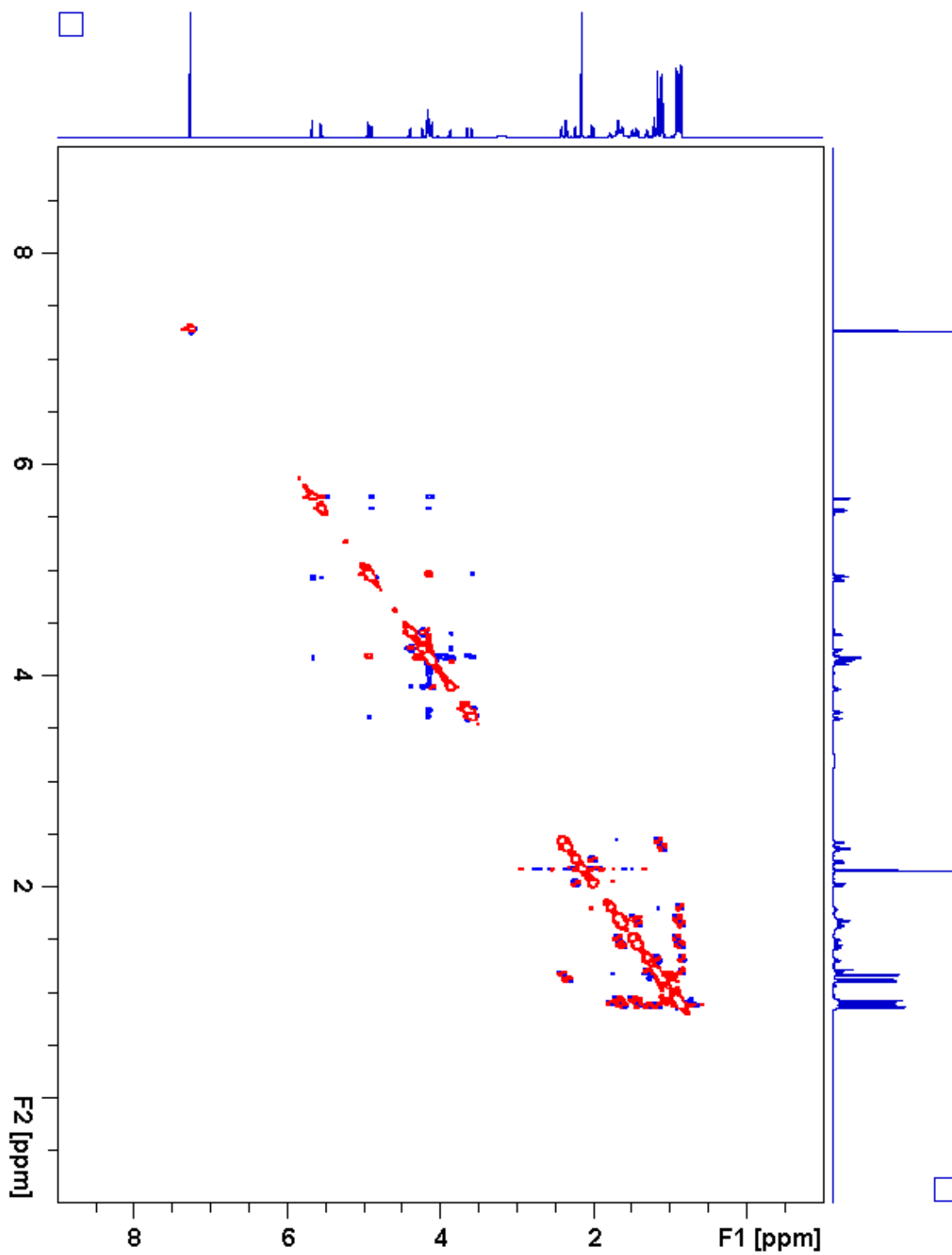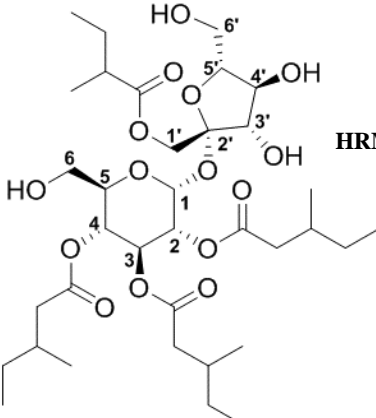
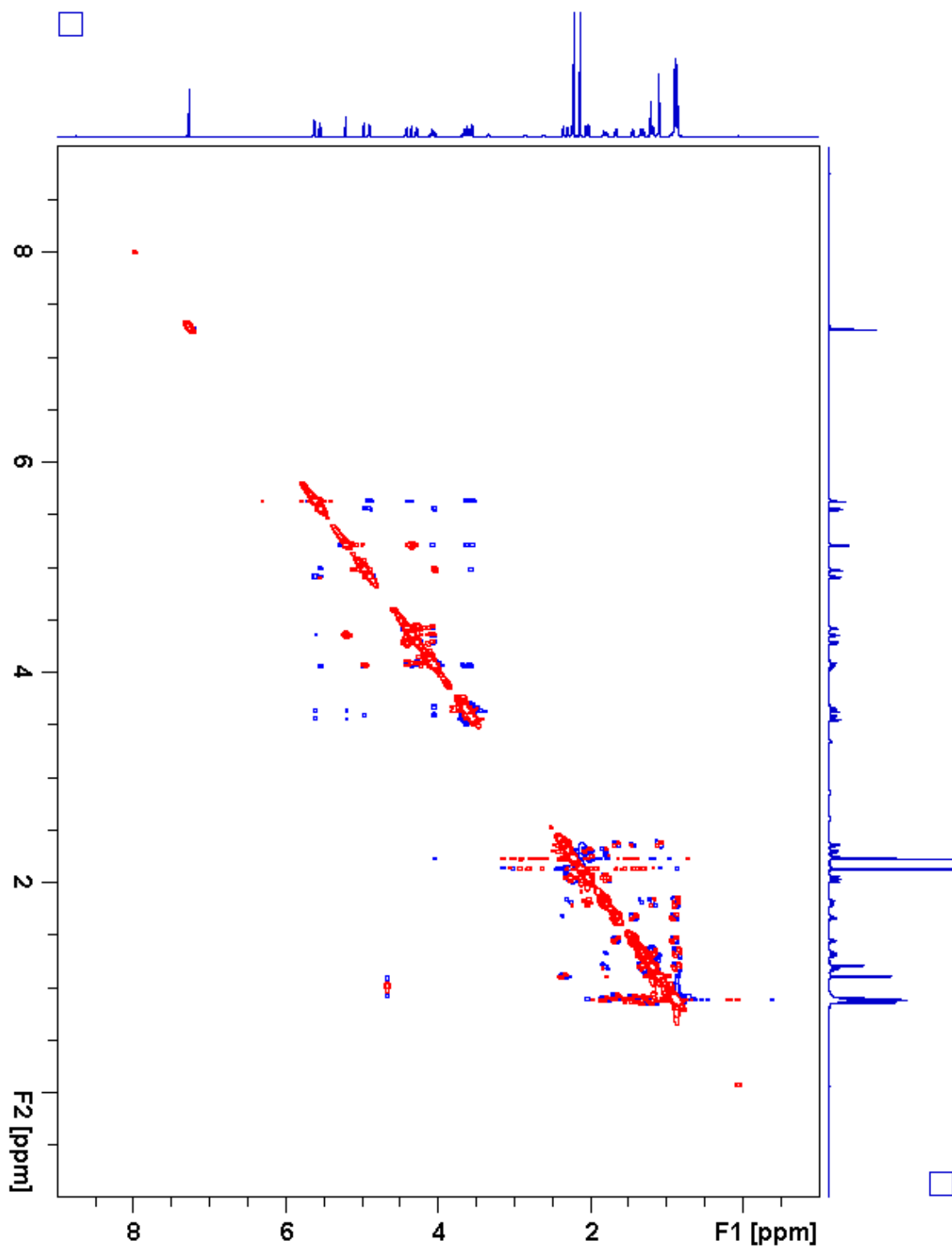**Figure 2.120.** S5:23:0(2,5,5,5,6) ROESY

**Table 2.25.** S4:23:0(5,6,6,6) Chemical shifts and coupling constants



**Molecular Formula:** $C_{35}H_{60}O_{15}$

**110 min Retention Time (ESI+):** 59.67 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{35}H_{64}NO_{15}^+$ ([M+NH$_4^+$]): 738.4270, found: 738.4266

**Fraction, Batch:** #80, A-D
**Sample mass for NMR analysis:** 0.4 mg
**NMR Solvent:** CDCl$_3$
**InChi Key:** HTYLQRCONXFYCT-ZVUFDNGLSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.78 (d, *J* = 3.9 Hz) | 89.15 |
| **2**(CH) | 4.86 (dd, *J* = 10.3, 4.0 Hz) | 70.69 |
| -1(CO) | | 173.65 |
| -2(CH$_2$) | 2.39 (dd, *J* = 15.7, 5.6 Hz), 2.12 (dd, *J* = 15.7, 8.4 Hz) | 41.2[a] |
| -3(CH) | 1.83 (m) | 31.7[b] |
| -4(CH$_3$) | 0.90 (d, *J* = 6.7 Hz) | 19.4[c] |
| -5(CH$_2$) | 1.33 (m), 1.21 (m) | 29.4[d] |
| -6(CH$_3$) | 0.87 (t, *J* = 7.4 Hz) | 11.4[e] |
| **3**(CH) | 5.53 (dd, *J* = 10.6, 9.2 Hz) | 68.83 |
| -1(CO) | | 172.22 |
| -2(CH$_2$) | 2.25 (dd, *J* = 15.3, 5.7 Hz), 2.04 (dd, *J* = 15.3, 8.4 Hz) | 41.2[a] |
| -3(CH) | 1.83 (m) | 31.7[b] |
| -4(CH$_3$) | 0.90 (d, *J* = 6.7 Hz) | 19.4[c] |
| -5(CH$_2$) | 1.33 (m), 1.21 (m) | 29.4[d] |
| -6(CH$_3$) | 0.87 (, *J* = 7.4 Hz) | 11.4[e] |
| **4**(CH) | 4.92 (dd, *J* = 10.7, 9.3 Hz) | 68.66 |
| -1(CO) | | 172.82 |
| -2(CH$_2$) | 2.29 (dd, *J* = 15.3, 5.8 Hz), 2.11 (dd, *J* = 15.3, 8.4 Hz) | 41.2[a] |
| -3(CH) | 1.83 (m) | 31.7[b] |
| -4(CH$_3$) | 0.90 (d, *J* = 6.7 Hz) | 19.4[c] |
| -5(CH$_2$) | 1.33 (m), 1.21 (m) | 29.4[d] |
| -6(CH$_3$) | 0.87 (, *J* = 7.4 Hz) | 11.4[e] |
| **5**(CH) | 4.12 (ddd, *J* = 10.4, 5.8, 2.0 Hz) | 72.21 |
| **6**(CH$_2$) | 3.64 (dd, *J* = 12.8, 2.2 Hz), 3.61 (dd, *J* = 12.6, 6.1 Hz) | 61.55 |

**Table 2.25.** (continued)

| | | |
|---|---|---|
| **1'**(CH₂) | 4.09 (d, *J* = 11.7 Hz), 4.06 (d, *J* = 11.7 Hz) | 63.69 |
| -1(CO) | | 175.89 |
| -2(CH) | 2.41 (sextet, *J* = 7.0 Hz) | 41.2ᵃ |
| -3(CH₃) | 1.16 (d, *J* = 7.0 Hz) | 16.74 |
| -4(CH₂) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH₃) | 0.92 (t, *J* = 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.52 |
| **3'**(CH) | 4.17 (d, *J* = 8.7 Hz) | 78.43 |
| **4'**(CH) | 4.31 (t, *J* = 8.7 Hz) | 72.93 |
| **5'**(CH) | 3.71 (m) | 81.35 |
| **6'**(CH₂) | 3.88 (dd, *J* = 14.2, 3.3 Hz), 3.70 (m) | 59.81 |

a - Four $^{13}$C signals not resolved in 2D spectra (41.10, 41.14, 41.23, 41.26 ppm)

b - Three $^{13}$C signals not resolved in 2D spectra (31.66, 31.78, 31.82 ppm)

c - Three $^{13}$C signals not resolved in 2D spectra (19.37, 19.39, 19.43 ppm)

d - Three $^{13}$C signals not resolved in 2D spectra (29.33, 29.39, 29.42 ppm)

e - Three $^{13}$C signals not resolved in 2D spectra (11.36, 11.39, 11.41 ppm)
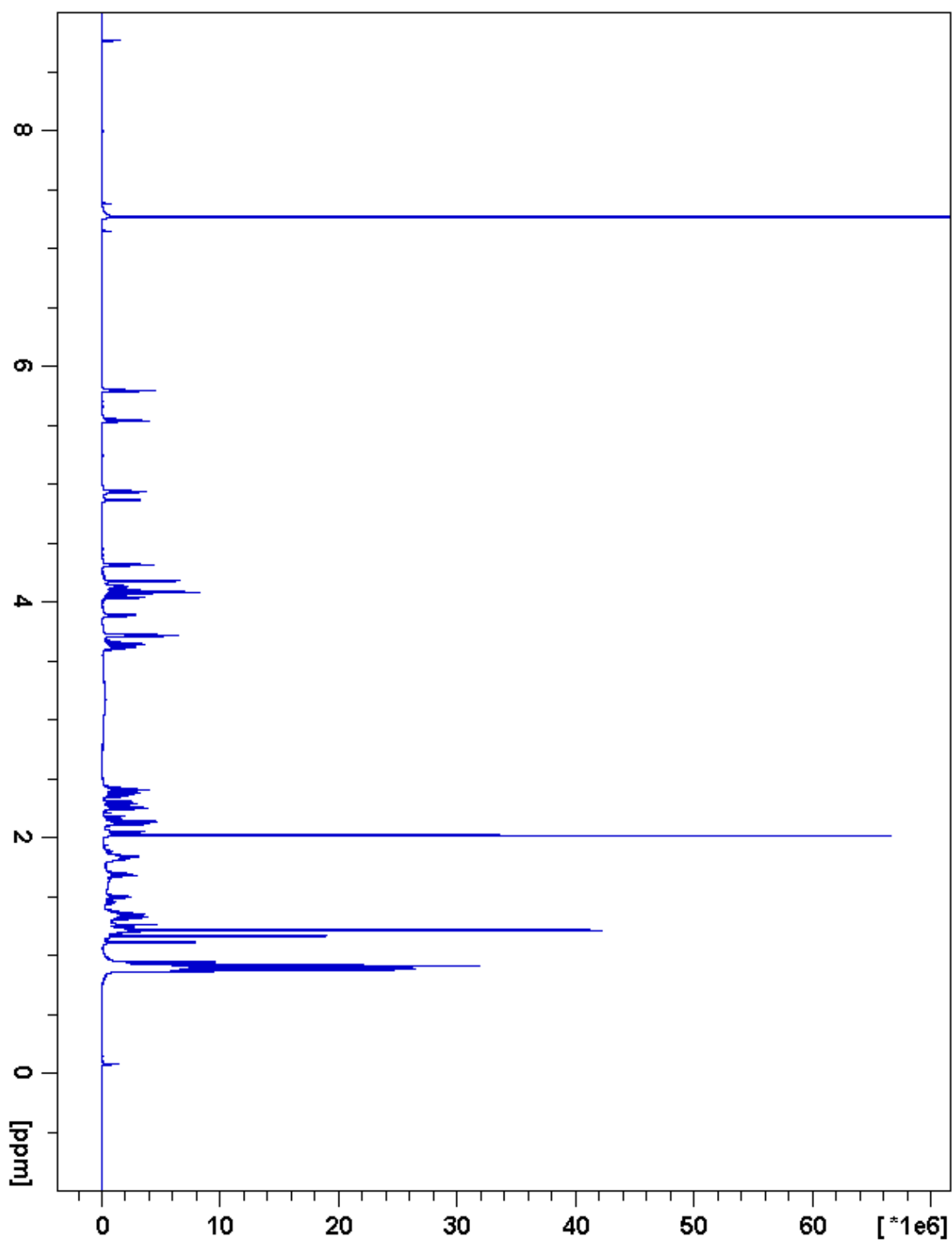
**Figure 2.121.** S4:23:0(5,6,6,6) ^{1}H NMR
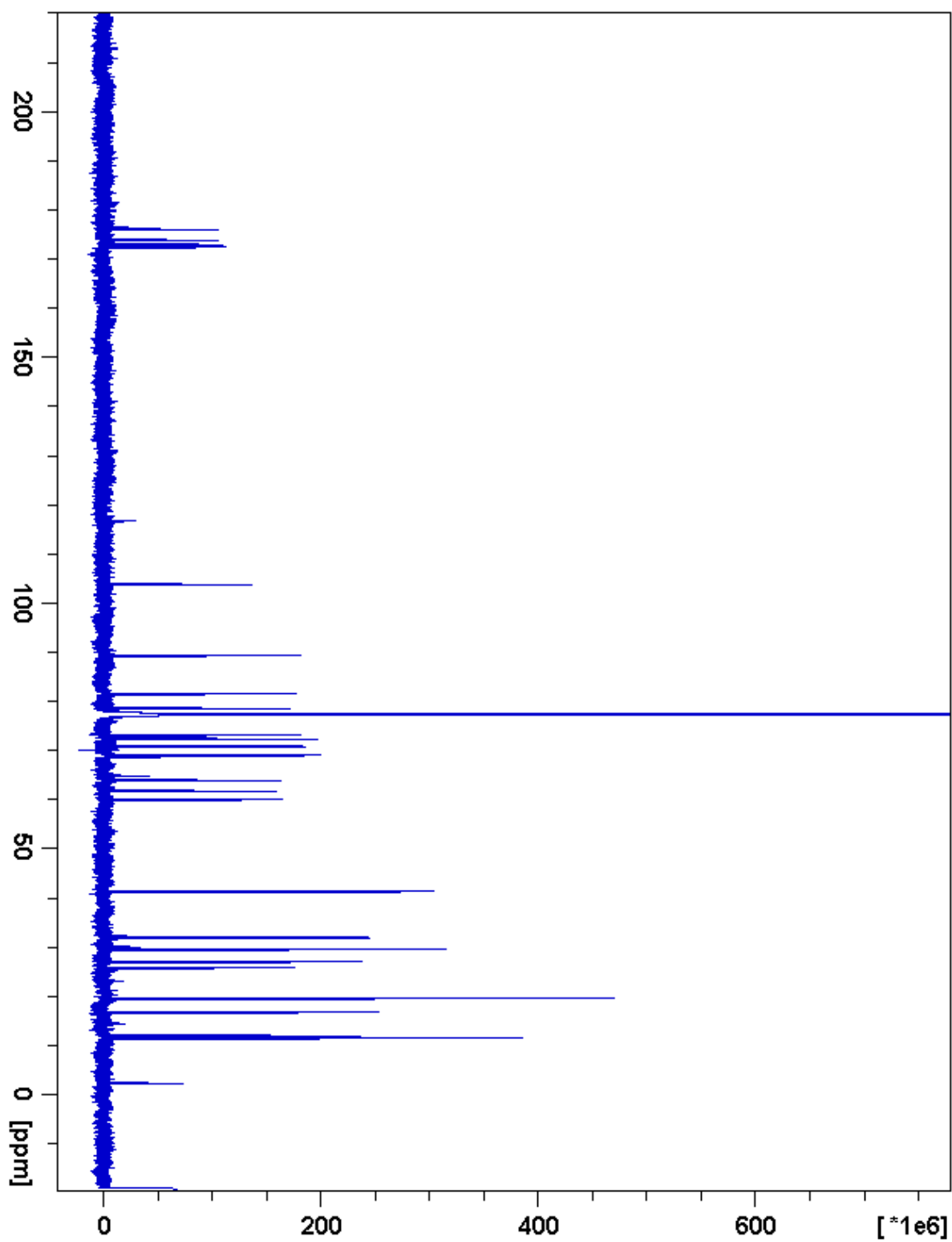
**Figure 2.122.** S4:23:0(5,6,6,6) $^{13}$C NMR
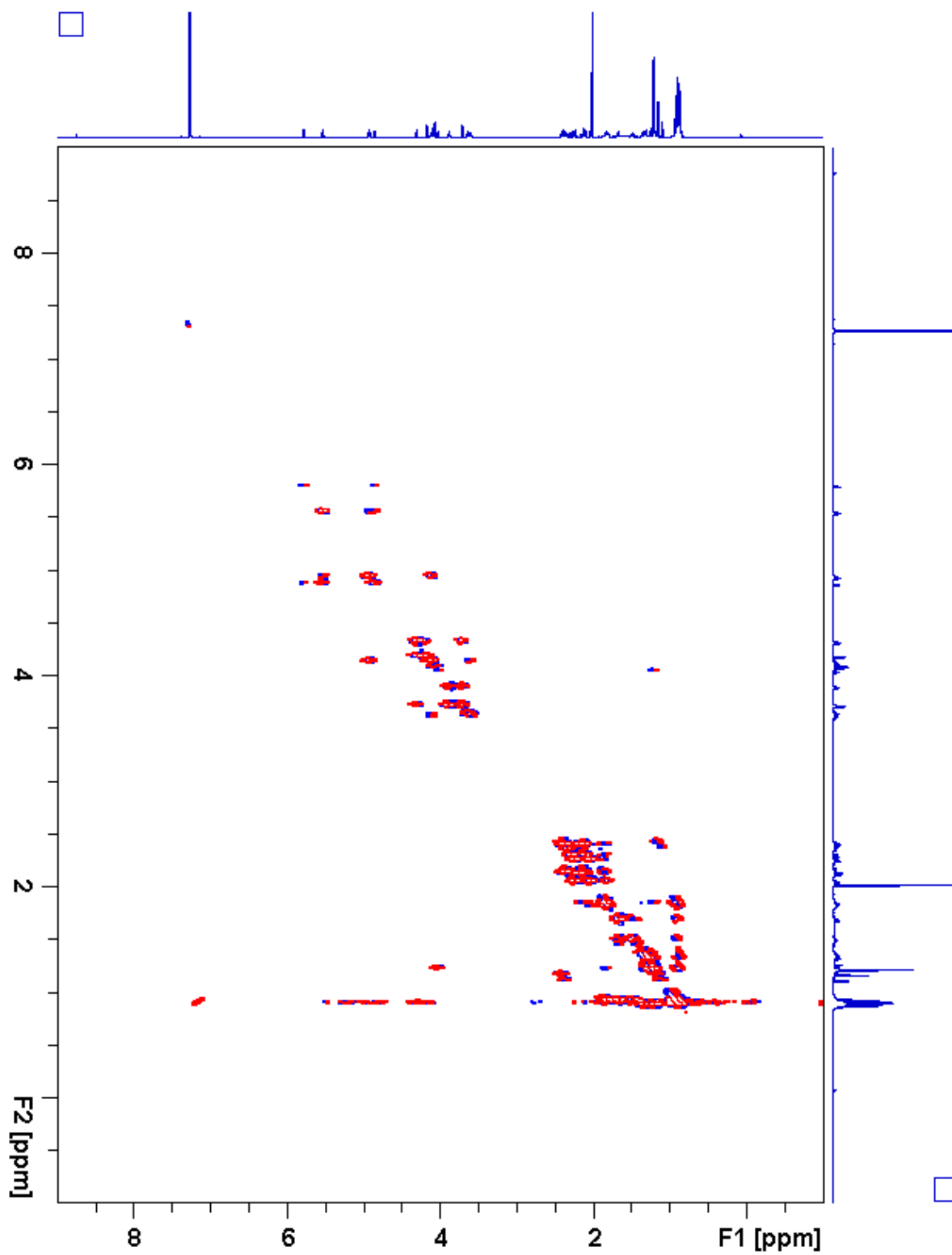
**Figure 2.123.** S4:23:0(5,6,6,6) $^1$H-$^1$H gCOSY
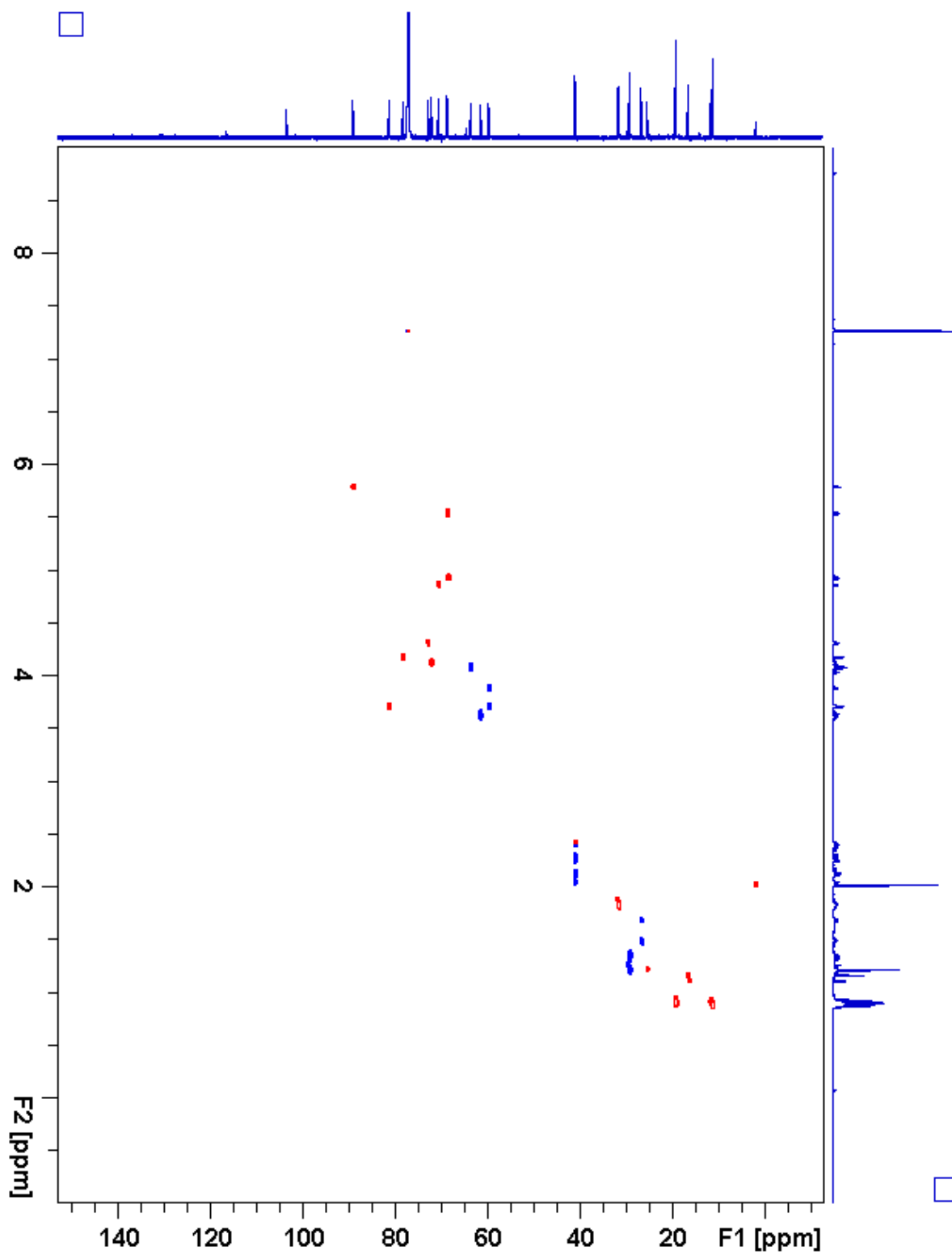
**Figure 2.124.** S4:23:0(5,6,6,6) gHSQC
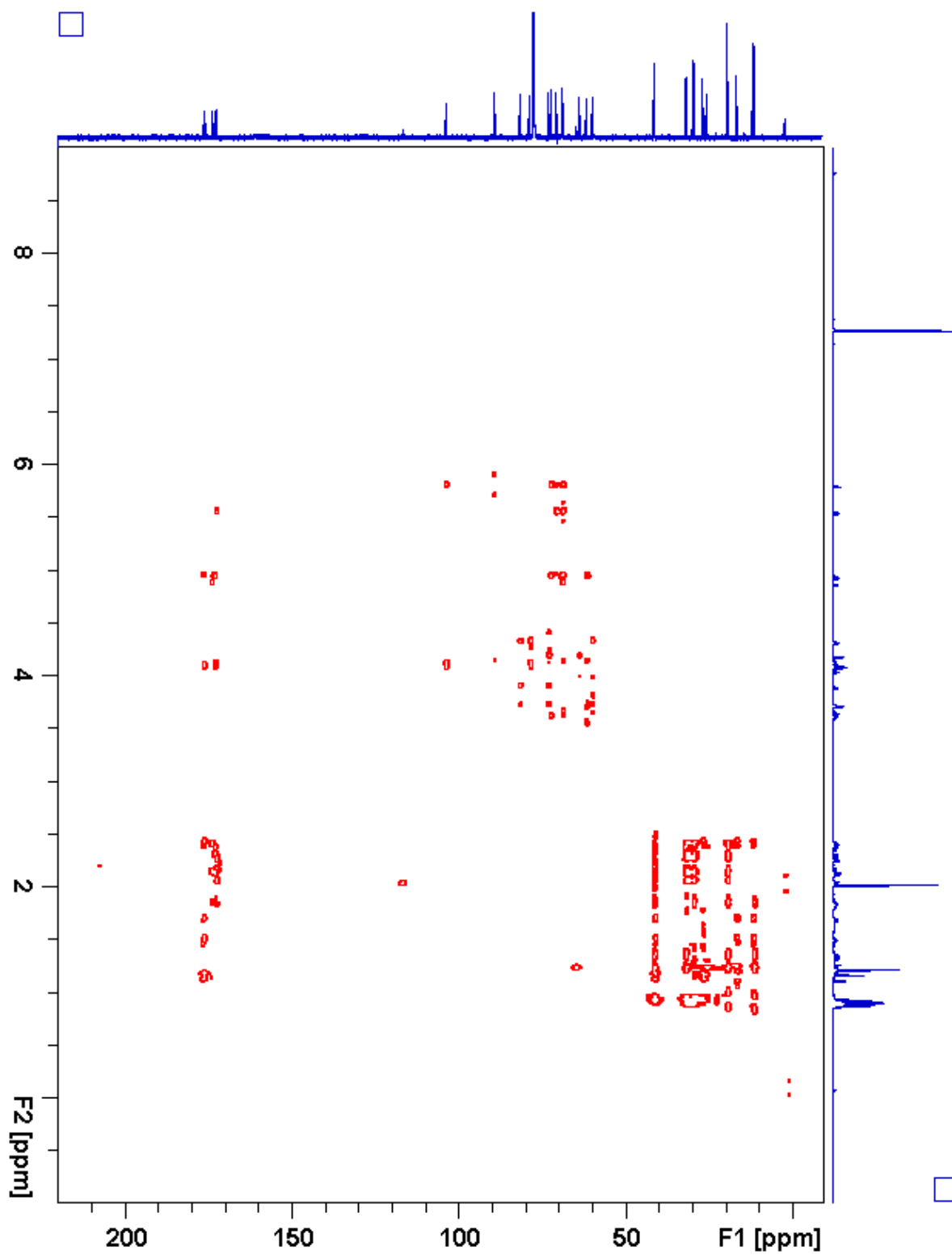
**Figure 2.125.** S4:23:0(5,6,6,6) gHMBC
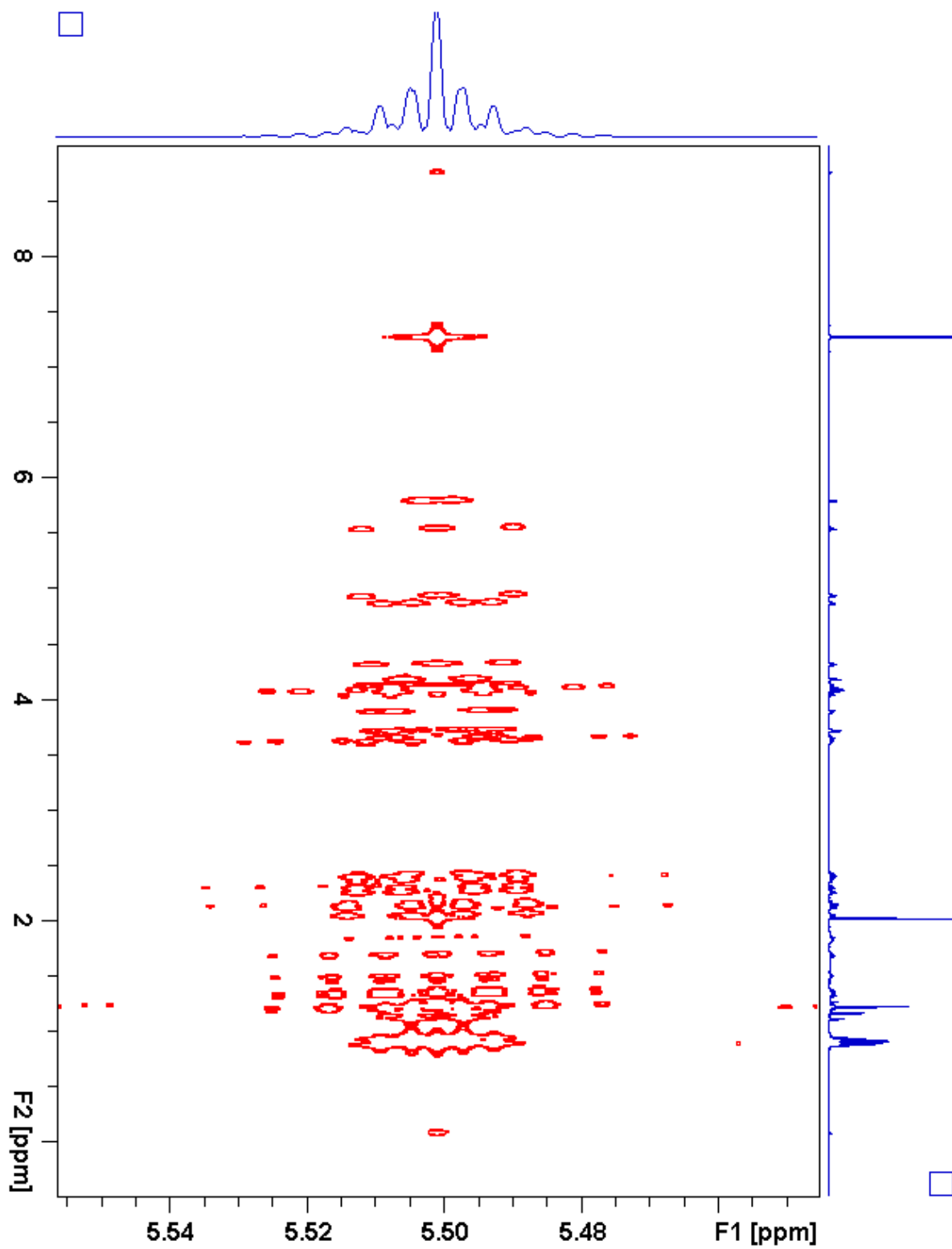
**Figure 2.126.** S4:23:0(5,6,6,6) *J*-resolved
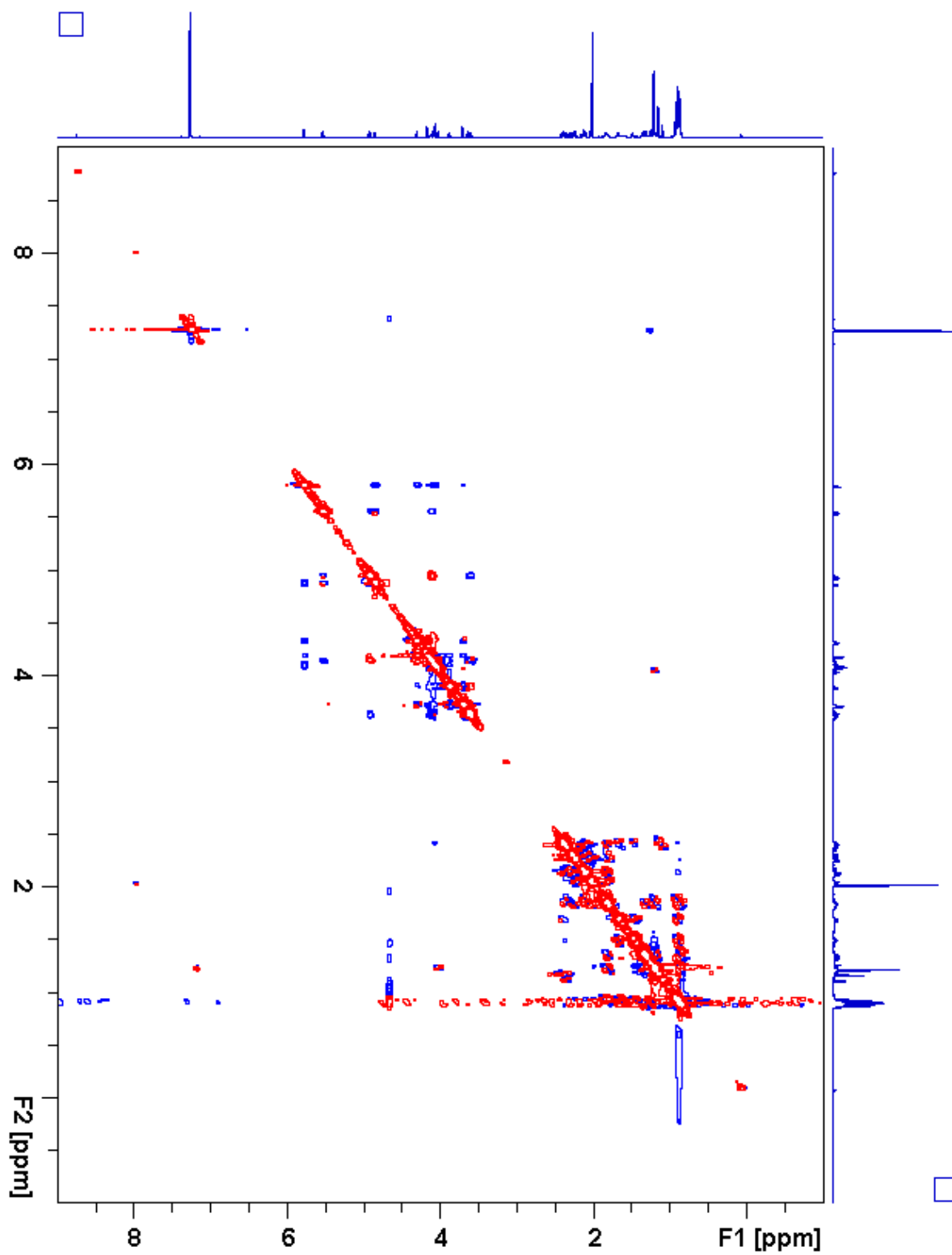
**Figure 2.127.** S4:23:0(5,6,6,6) ROESY

**Table 2.26.** S5:24:0(2,5,5,6,6) Chemical shifts and coupling constants
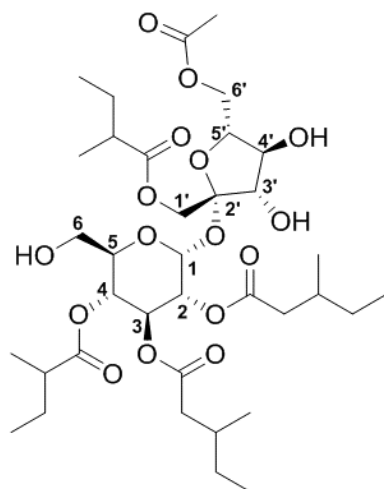


**Molecular Formula:** $C_{36}H_{60}O_{16}$

**110 min Retention Time (ESI+):** 61.76 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{36}H_{64}NO_{16}^+$ ([M+NH$_4^+$]): 766.4220, found: 766.4236

**Fraction, Batch:** #86, A-D

**Sample mass for NMR analysis:** 1.1 mg

**NMR Solvent:** CDCl$_3$

**InChi Key:** XPWKILWQFLWJSO-QBISVFFXSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 5.66 (d, *J* = 3.8 Hz) | 89.51 |
| **2**(CH) | 4.93 (dd, *J* = 10.5, 3.9 Hz) | 70.22 |
| -1(CO) | | 172.38 |
| -2(CH$_2$) | 2.34 (dd, *J* = 15.4, 5.8 Hz), 2.08 (dd, *J* = 15.4, 8.4 Hz) | 41.07 |
| -3(CH) | 1.83 (m) | 31.68 |
| -4(CH$_3$) | 0.89 (d, *J* = 6.7 Hz) | 19.33 |
| -5(CH$_2$) | 1.34 (m), 1.20 (m) | 29.36[a] |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.36 |
| **3**(CH) | 5.54 (dd, *J* = 10.6, 9.2 Hz) | 68.93 |
| -1(CO) | | 172.33 |
| -2(CH$_2$) | 2.23 (dd, *J* = 15.4, 5.6 Hz), 2.03 (dd, *J* = 15.4, 8.4 Hz) | 41.23 |
| -3(CH) | 1.80 (m) | 31.72 |
| -4(CH$_3$) | 0.88 (d, *J* = 6.7 Hz) | 19.39 |
| -5(CH$_2$) | 1.31 (m), 1.18 (m) | 29.36[a] |
| -6(CH$_3$) | 0.86 (t, *J* = 7.4 Hz) | 11.38 |
| **4**(CH) | 4.94 (dd, *J* = 10.6, 9.3 Hz) | 68.90 |
| -1(CO) | | 176.35 |
| -2(CH) | 2.35 (sextet, *J* = 7.0 Hz) | 41.18 |
| -3(CH$_3$) | 1.10 (d, *J* = 7.0 Hz) | 16.53 |
| -4(CH$_2$) | 1.67 (m), 1.44 (m) | 26.65 |
| -5(CH$_3$) | 0.90 (t, *J* = 7.4 Hz) | 11.85 |
| **5**(CH) | 4.14 (m) | 71.31 |
| **6**(CH$_2$) | 3.66 (dd, *J* = 12.6, 2.8 Hz), 3.59 (dd, *J* = 12.6, 6.0 Hz) | 61.64 |

**Table 2.26.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.16 (d, $J$ = 11.6 Hz), 4.16 (d, $J$ = 11.6 Hz) | 63.49 |
| -1(CO) | | 175.95 |
| -2(CH) | 2.41 (sextet, $J$ = 7.0 Hz) | 41.10 |
| -3(CH$_3$) | 1.16 (d, $J$ = 7.0 Hz) | 16.71 |
| -4(CH$_2$) | 1.68 (m), 1.49 (m) | 26.88 |
| -5(CH$_3$) | 0.92 (t, $J$ = 7.4 Hz) | 11.75 |
| **2'**(C) | | 103.57 |
| **3'**(CH) | 4.14 (m) | 78.16 |
| **4'**(CH) | 4.11 (t, $J$ = 8.4 Hz) | 75.03 |
| **5'**(CH) | 3.90 (ddd, $J$ = 8.4, 5.5, 3.0 Hz) | 78.80 |
| **6'**(CH$_2$) | 4.42 (dd, $J$ = 12.3, 5.3 Hz), 4.22 (dd, $J$ = 12.3, 3.0 Hz) | 63.69 |
| -1(CO) | | 171.75 |
| -2(CH$_3$) | 2.15 (s) | 21.00 |

a - Two $^{13}$C signals overlapping
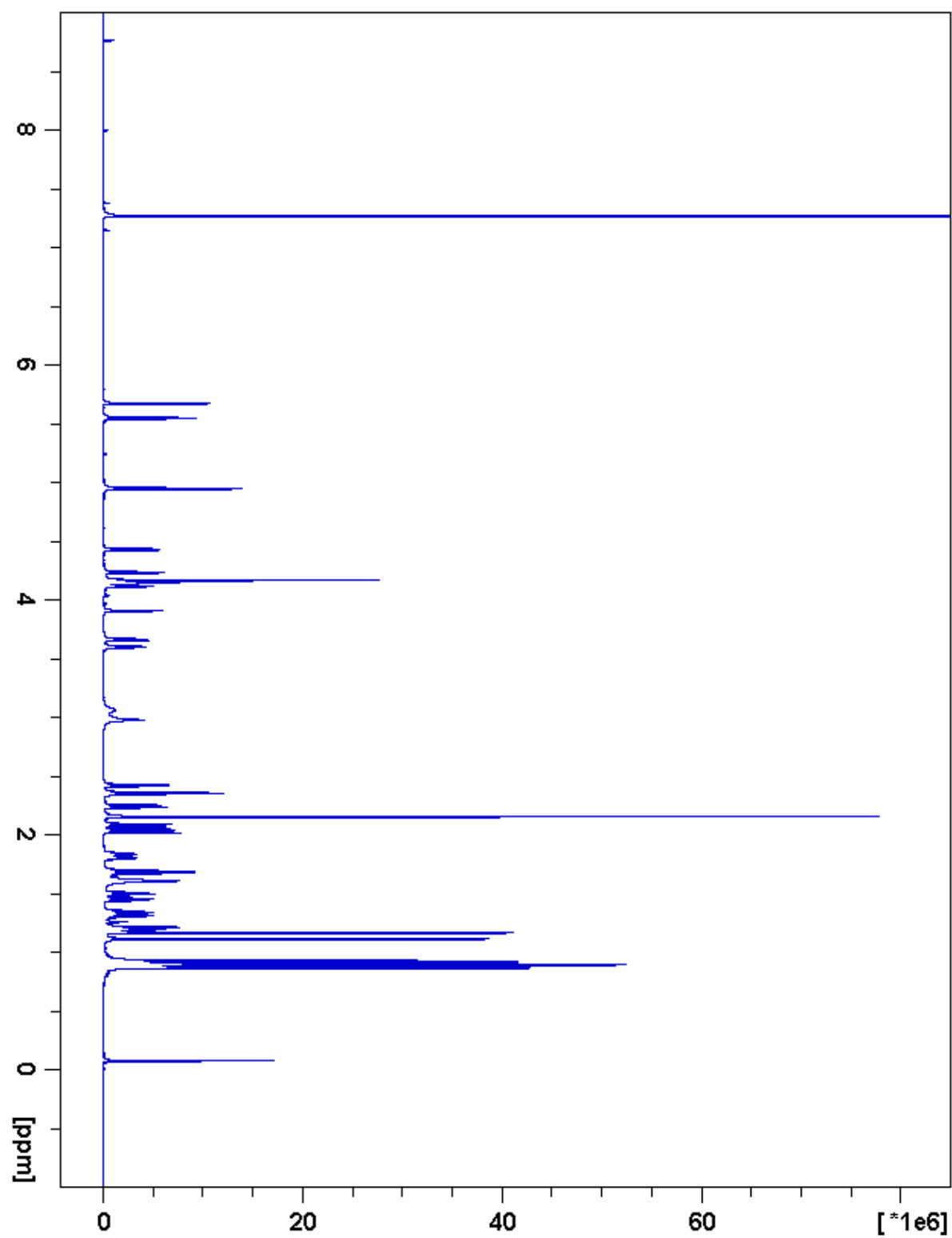
**Figure 2.128.** S5:24:0(2,5,5,6,6) ^{1}H NMR
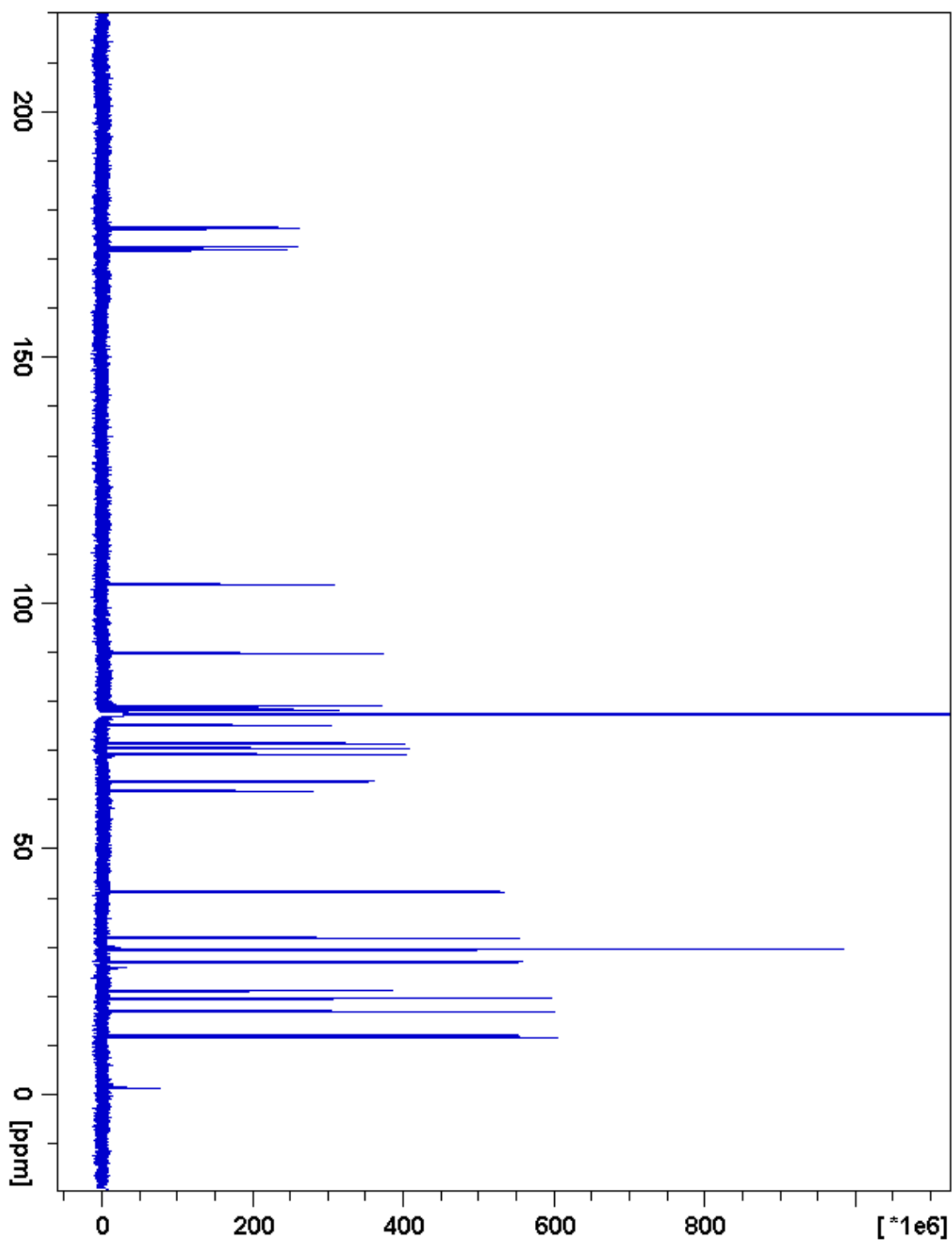
**Figure 2.129.** S5:24:0(2,5,5,6,6) $^{13}$C NMR
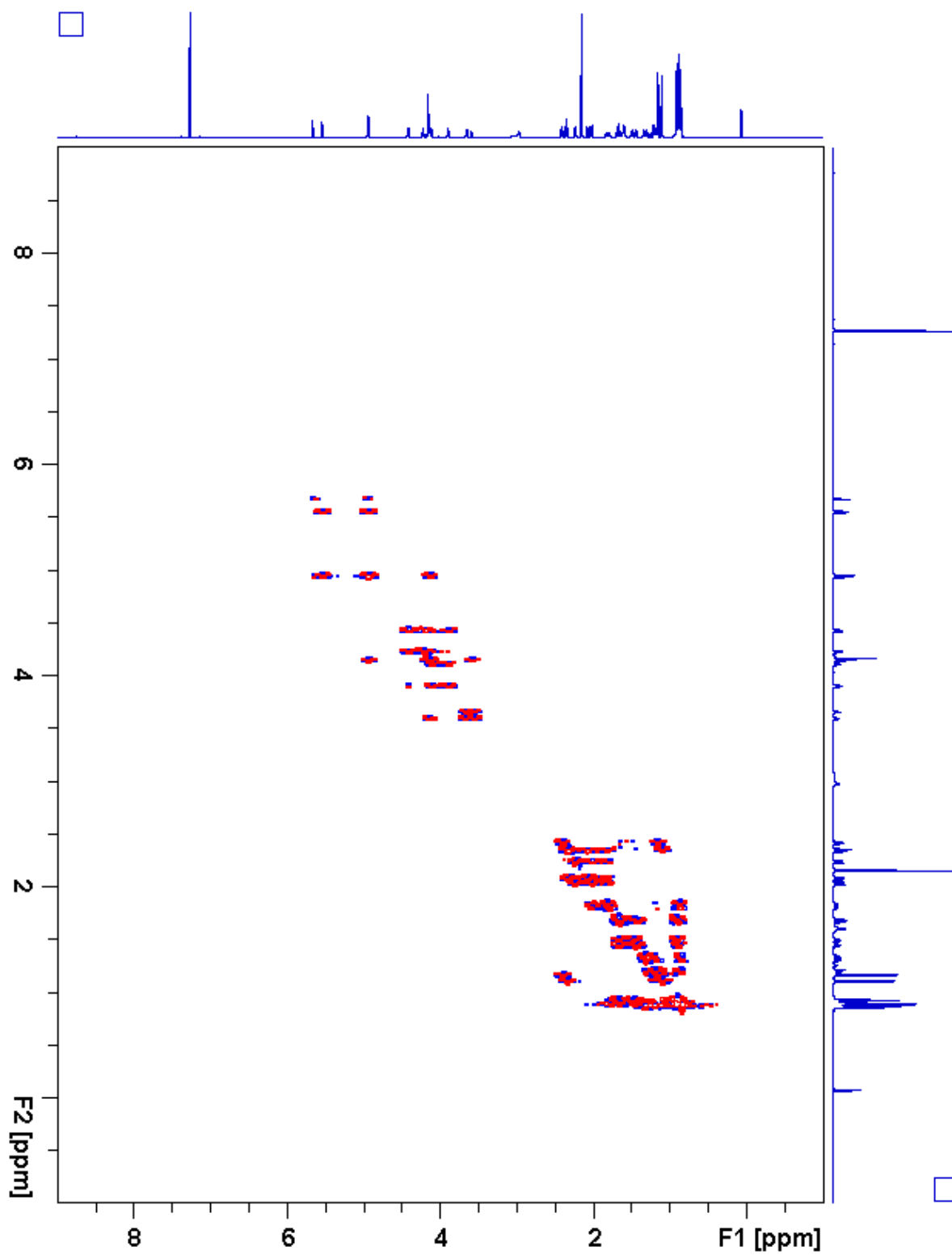
**Figure 2.130.** S5:24:0(2,5,5,6,6) $^1$H-$^1$H gCOSY
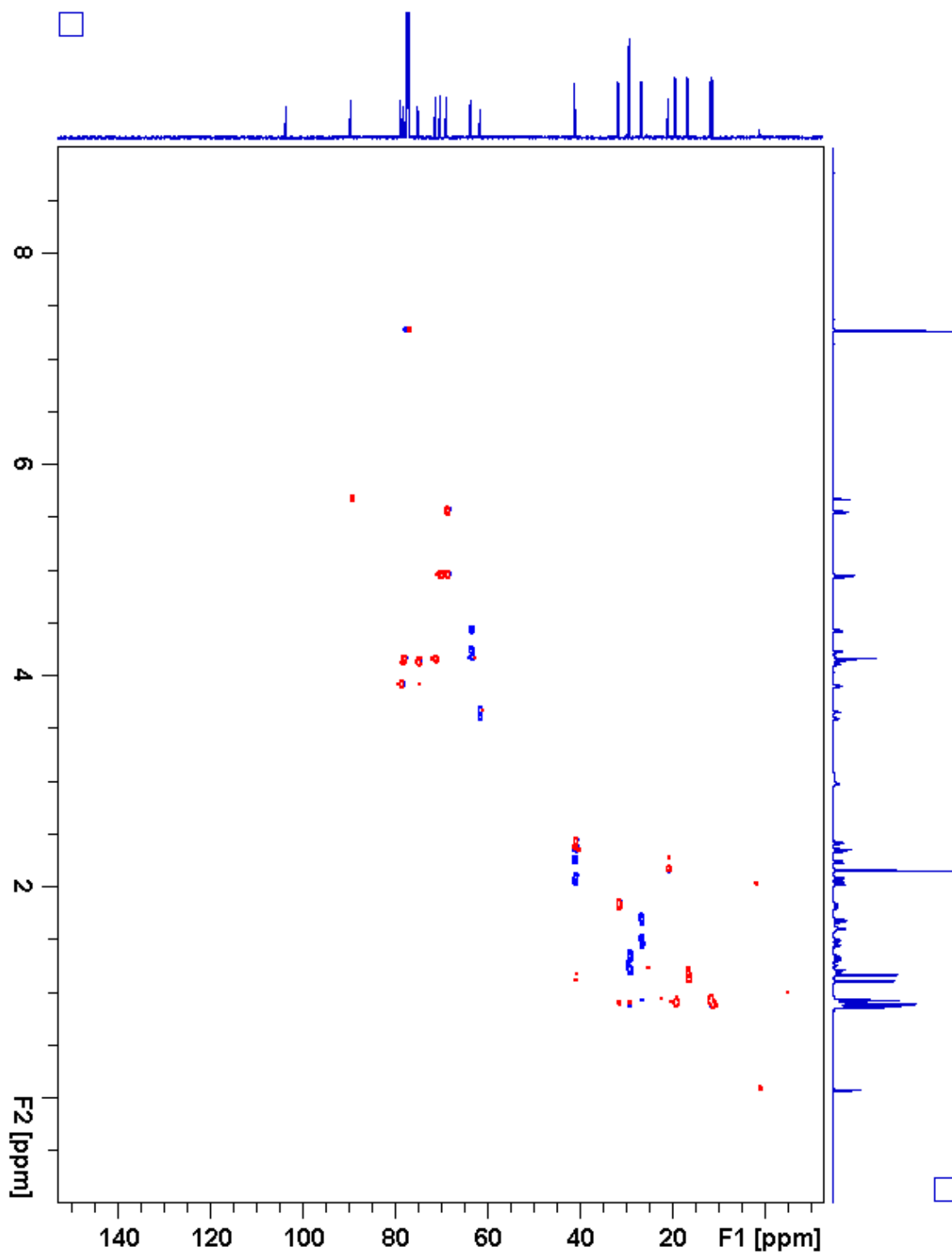
**Figure 2.131.** S5:24:0(2,5,5,6,6) gHSQC
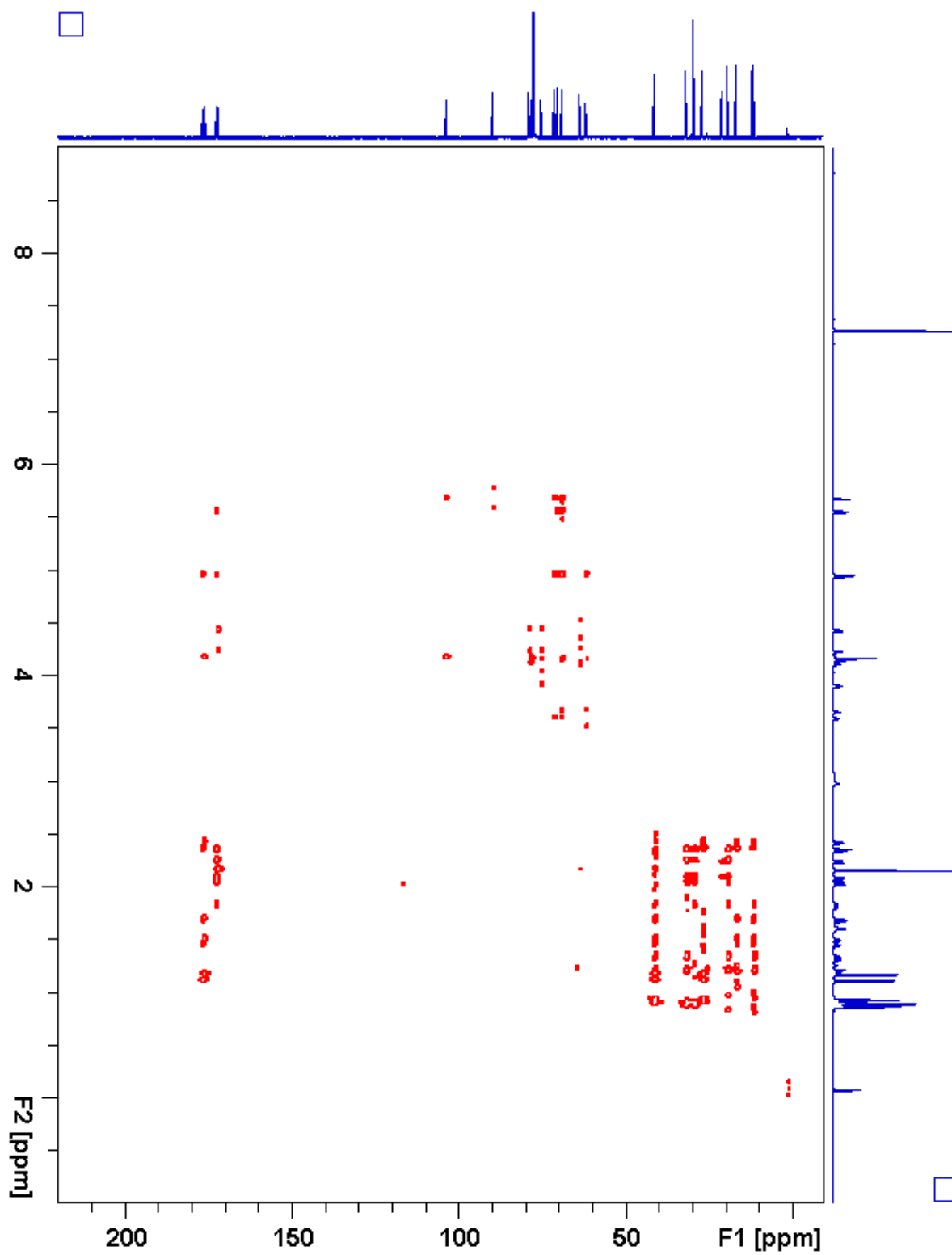
**Figure 2.132.** S5:24:0(2,5,5,6,6) gHMBC
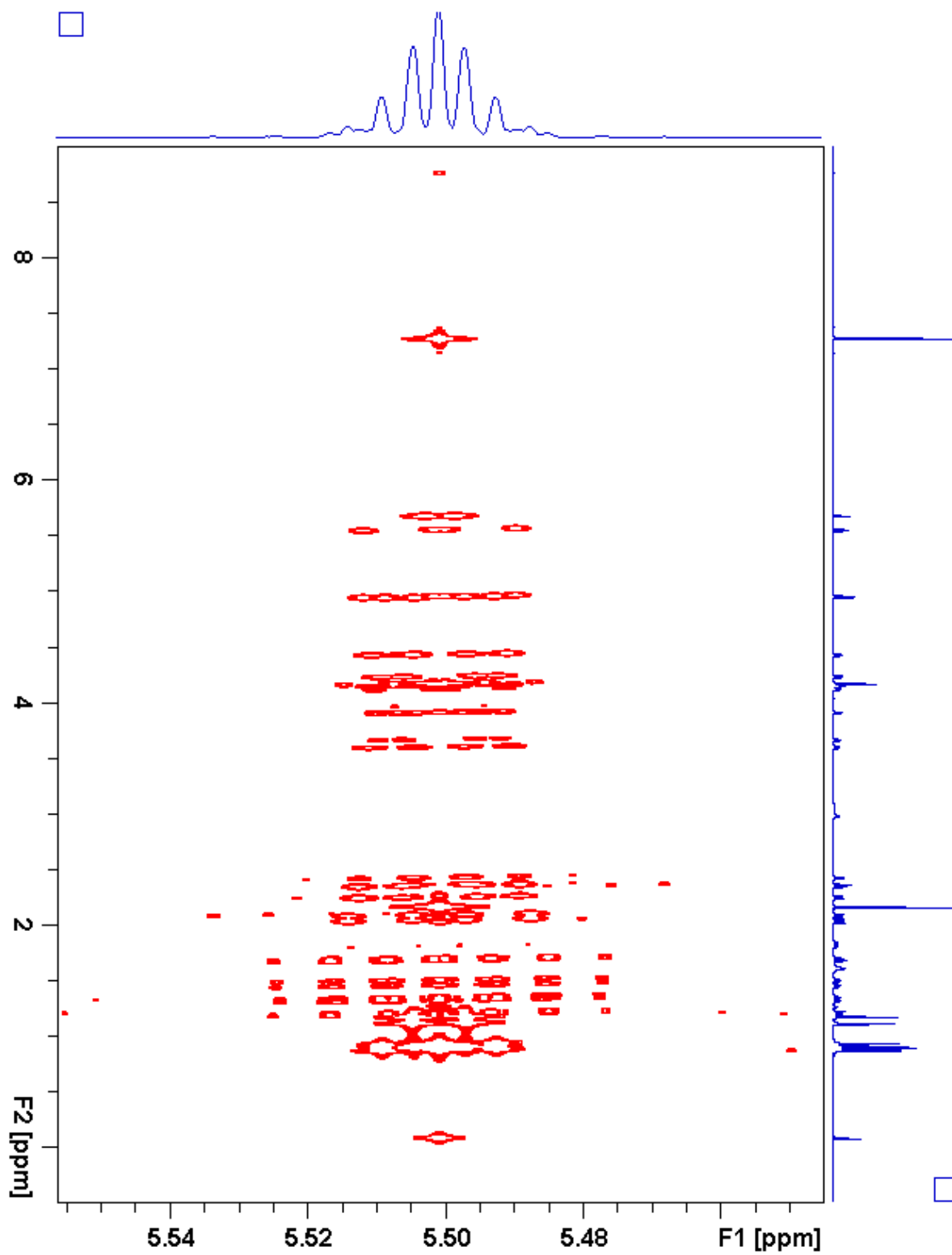
**Figure 2.133.** S5:24:0(2,5,5,6,6) *J*-resolved
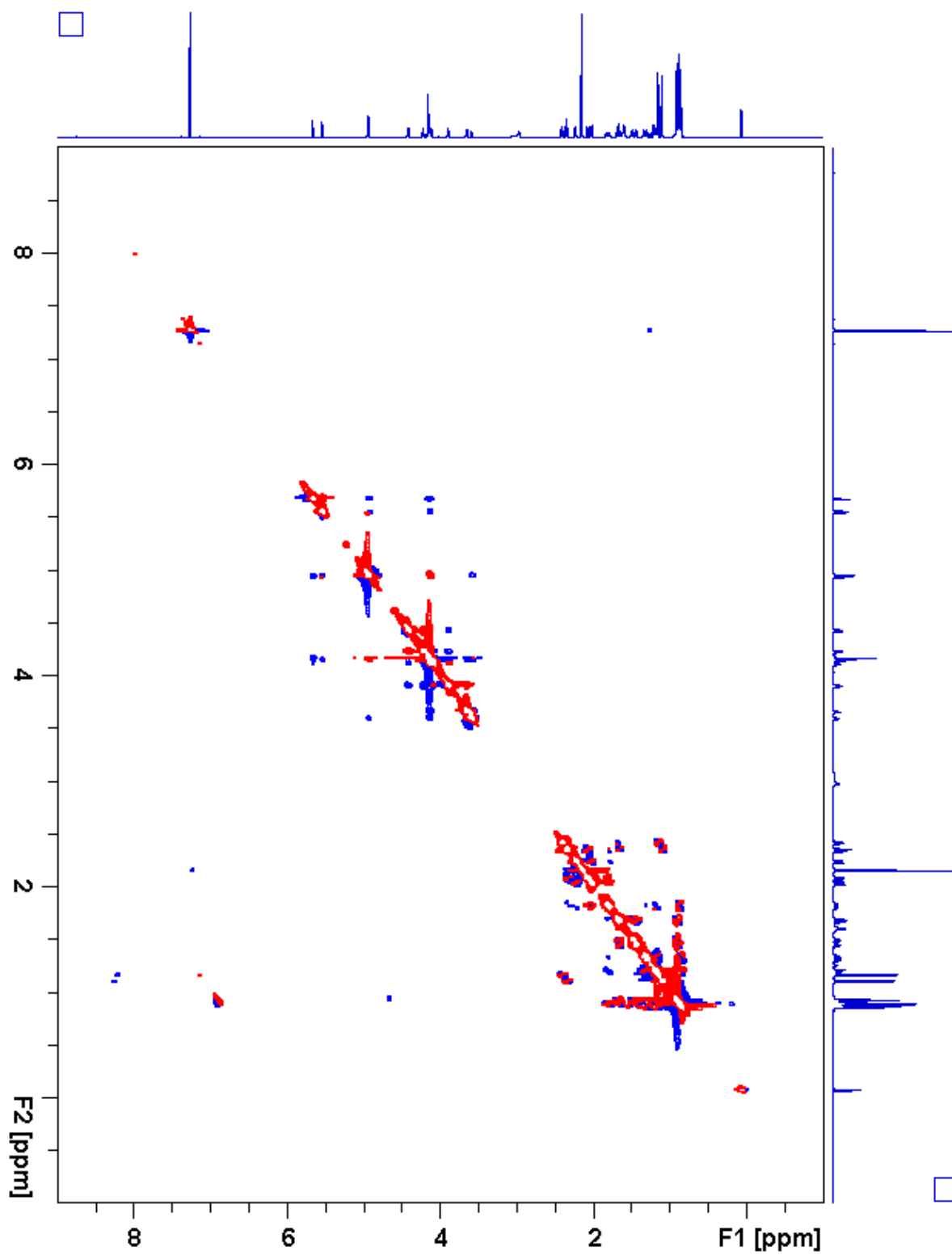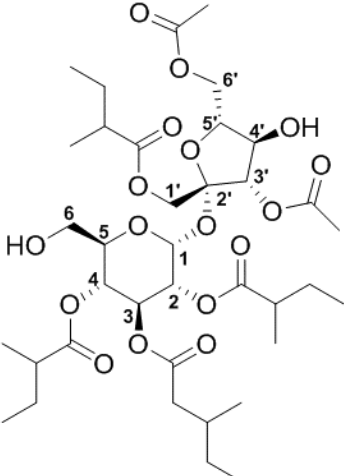
**Figure 2.134.** S5:24:0(2,5,5,6,6) ROESY

**Table 2.27.** S6:25:0(2,2,5,5,5,6) Chemical shifts and coupling constants



**Molecular Formula:** $C_{37}H_{60}O_{17}$

**110 min Retention Time (ESI+):** 62.99 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{37}H_{64}NO_{17}^+$ ([M+NH4+]): 794.4169, found: 794.4166

**Fraction, Batch:** #89-90, A-G

**Sample mass for NMR analysis:** 0.7 mg

**NMR Solvent:** CDCl3

**InChi Key:** RRWBGWSVUMNBBH-NACAQSIGSA-N

| Carbon # (group) | $^1H$ (ppm) | $^{13}C$ (ppm) |
|---|---|---|
| **1**(CH) | 5.70 (d, *J* = 3.6 Hz) | 89.41 |
| **2**(CH) | 4.87 (dd, *J* = 10.4, 3.7 Hz) | 70.37 |
| -1(CO) | | 175.94 |
| -2(CH) | 2.40 (sextet, *J* = 7.0 Hz) | 40.71 |
| -3(CH3) | 1.12 (d, *J* = 7.0 Hz) | 16.11 |
| -4(CH2) | 1.63 (m), 1.44 (m) | 26.77 |
| -5(CH3) | 0.86 (t, *J* = 7.4 Hz) | 11.70 |
| **3**(CH) | 5.57 (dd, *J* = 10.7, 9.2 Hz) | 68.87 |
| -1(CO) | | 172.18 |
| -2(CH2) | 2.24 (dd, *J* = 15.8, 5.4 Hz), 2.01 (dd, *J* = 15.8, 8.6 Hz) | 41.1[a] |
| -3(CH) | 1.78 (m) | 31.47 |
| -4(CH3) | 0.88 (d, *J* = 6.7 Hz) | 19.46 |
| -5(CH2) | 1.30 (m), 1.18 (m) | 29.42 |
| -6(CH3) | 0.85 (t, *J* = 7.4 Hz) | 11.39 |
| **4**(CH) | 4.97 (dd, *J* = 10.3, 9.2 Hz) | 68.97 |
| -1(CO) | | 176.48 |
| -2(CH) | 2.36 (sextet, *J* = 7.0 Hz) | 41.1[a] |
| -3(CH3) | 1.10 (d, *J* = 7.0 Hz) | 16.48 |
| -4(CH2) | 1.66 (m), 1.45 (m) | 26.70 |
| -5(CH3) | 0.90 (t, *J* = 7.4 Hz) | 11.83 |
| **5**(CH) | 4.07 (ddd, *J* = 10.2, 5.1, 2.2 Hz) | 71.15 |
| **6**(CH2) | 3.66 (dd, *J* = 12.8, 2.3 Hz), 3.59 (dd, *J* = 12.8, 5.1 Hz) | 61.67 |

**Table 2.27.** (continued)

| | | |
|---|---|---|
| **1'**(CH$_2$) | 4.14 (d, $J$ = 11.8 Hz), 4.04 (d, $J$ = 11.8 Hz) | 64.10 |
| -1(CO) | | 175.81 |
| -2(CH) | 2.44 (sextet, $J$ = 7.0 Hz) | 41.1[a] |
| -3(CH$_3$) | 1.17 (d, $J$ = 7.0 Hz) | 16.66 |
| -4(CH$_2$) | 1.70 (m), 1.51 (m) | 26.90 |
| -5(CH$_3$) | 0.93 (t, $J$ = 7.4 Hz) | 11.65 |
| **2'**(C) | | 102.98 |
| **3'**(CH) | 5.23 (d, $J$ = 8.1 Hz) | 78.95 |
| -1(CO) | | 171.46 |
| -2(CH$_3$) | 2.20 (s) | 20.91 |
| **4'**(CH) | 4.35 (t, $J$ = 8.1 Hz) | 73.45 |
| **5'**(CH) | 4.04 (ddd, $J$ = 8.2, 5.7, 3.6 Hz) | 80.22 |
| **6'**(CH$_2$) | 4.40 (dd, $J$ = 12.2, 5.6 Hz), 4.30 (dd, $J$ = 12.2, 3.5 Hz) | 63.76 |
| -1(CO) | | 171.65 |
| -2(CH$_3$) | 2.13 (s) | 20.99 |

a - Three $^{13}$C signals not resolved in 2D spectra (41.10, 41.12, 41.14 ppm)
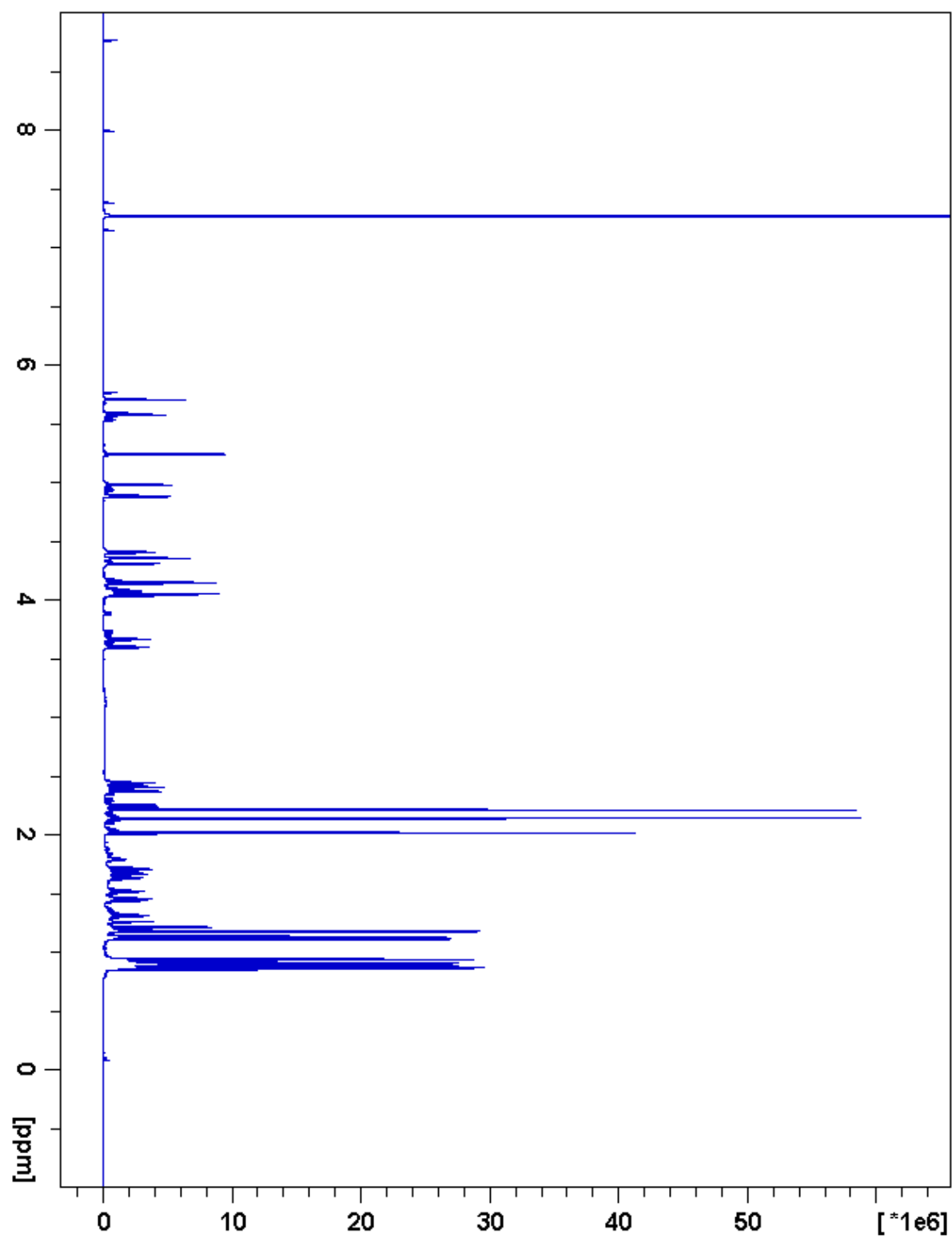
**Figure 2.135.** S6:25:0(2,2,5,5,6) $^1$H NMR
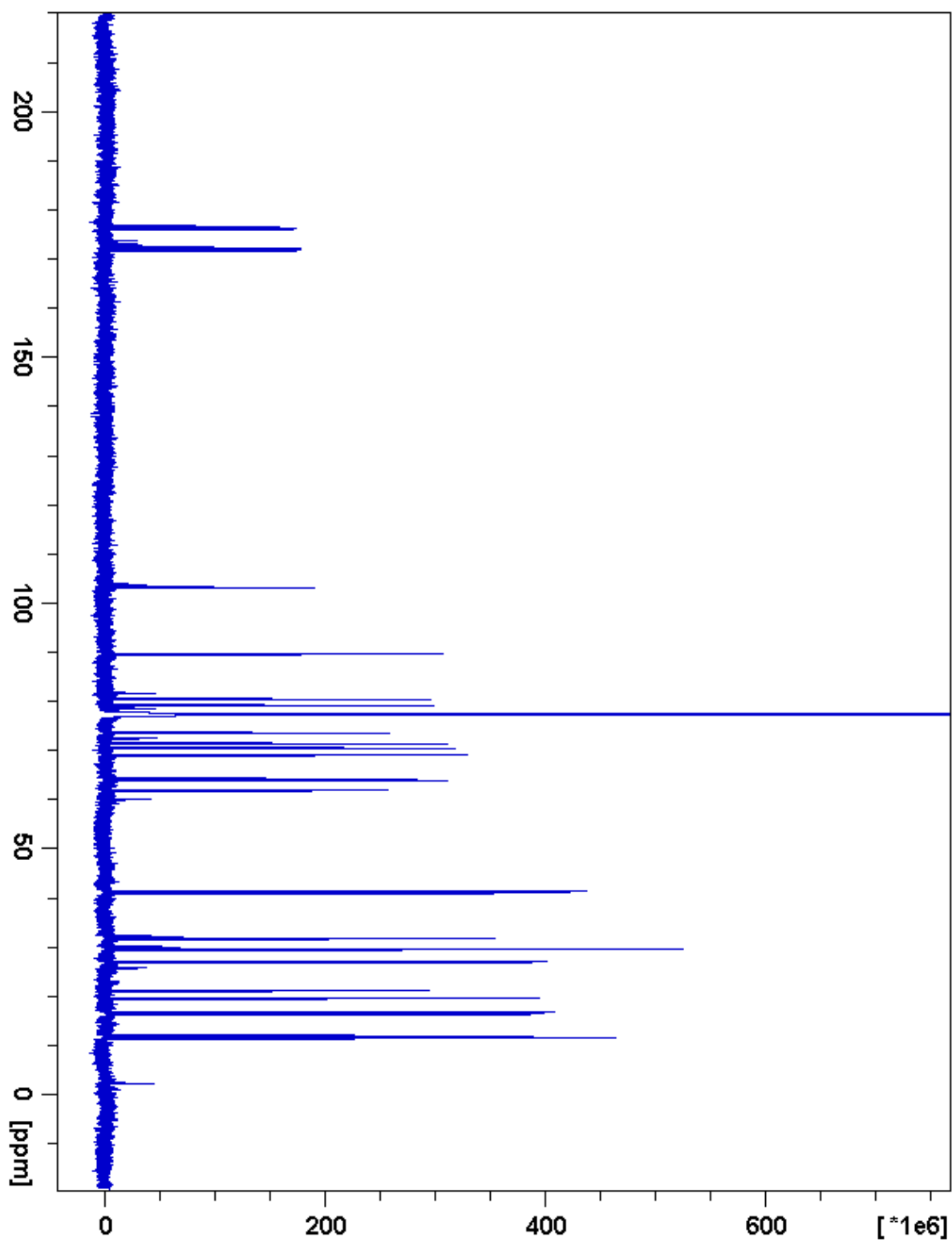
**Figure 2.136.** S6:25:0(2,2,5,5,6) $^{13}$C NMR
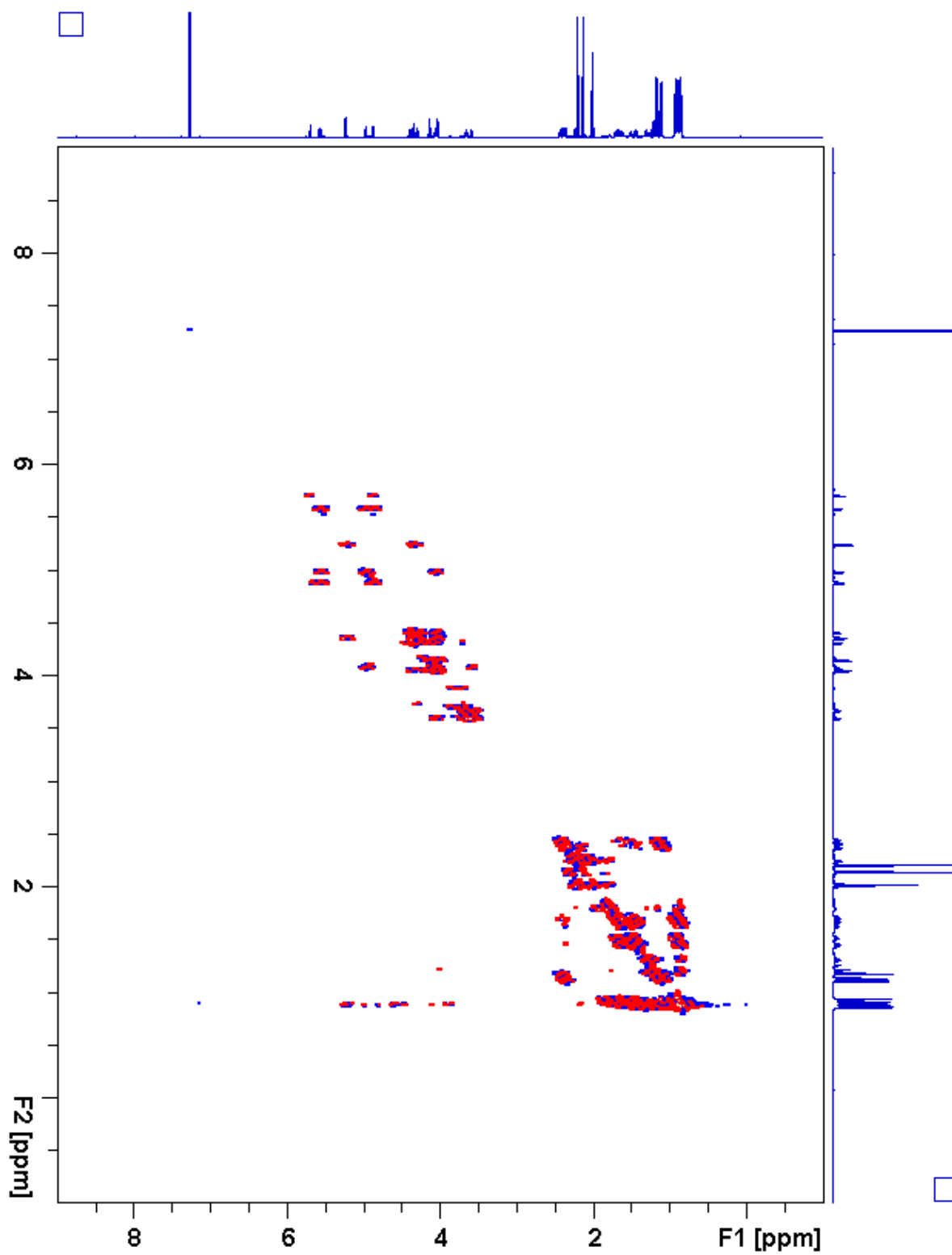
**Figure 2.137.** S6:25:0(2,2,5,5,5,6) $^1$H-$^1$H gCOSY
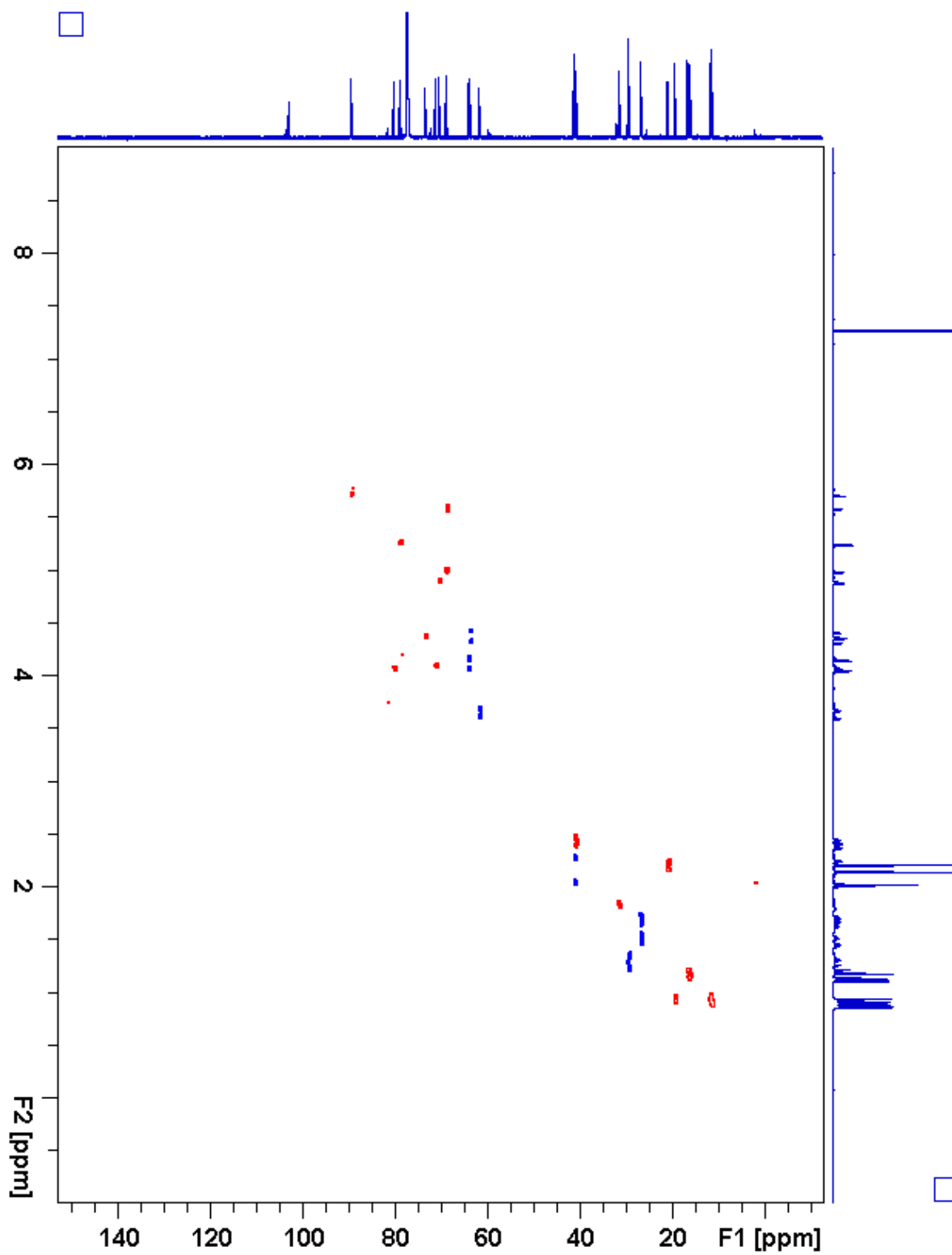
**Figure 2.138.** S6:25:0(2,2,5,5,5,6) gHSQC
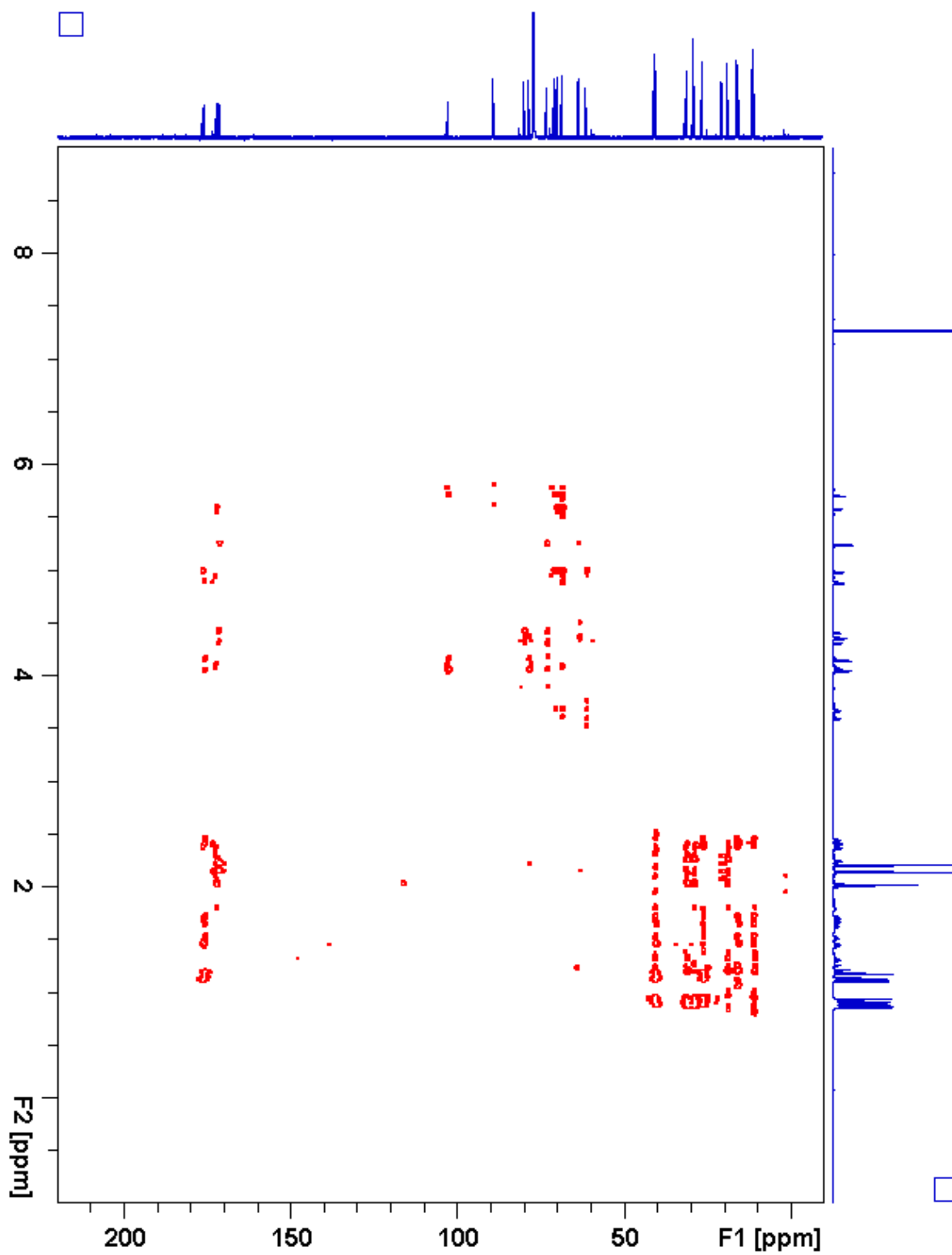
**Figure 2.139.** S6:25:0(2,2,5,5,5,6) gHMBC
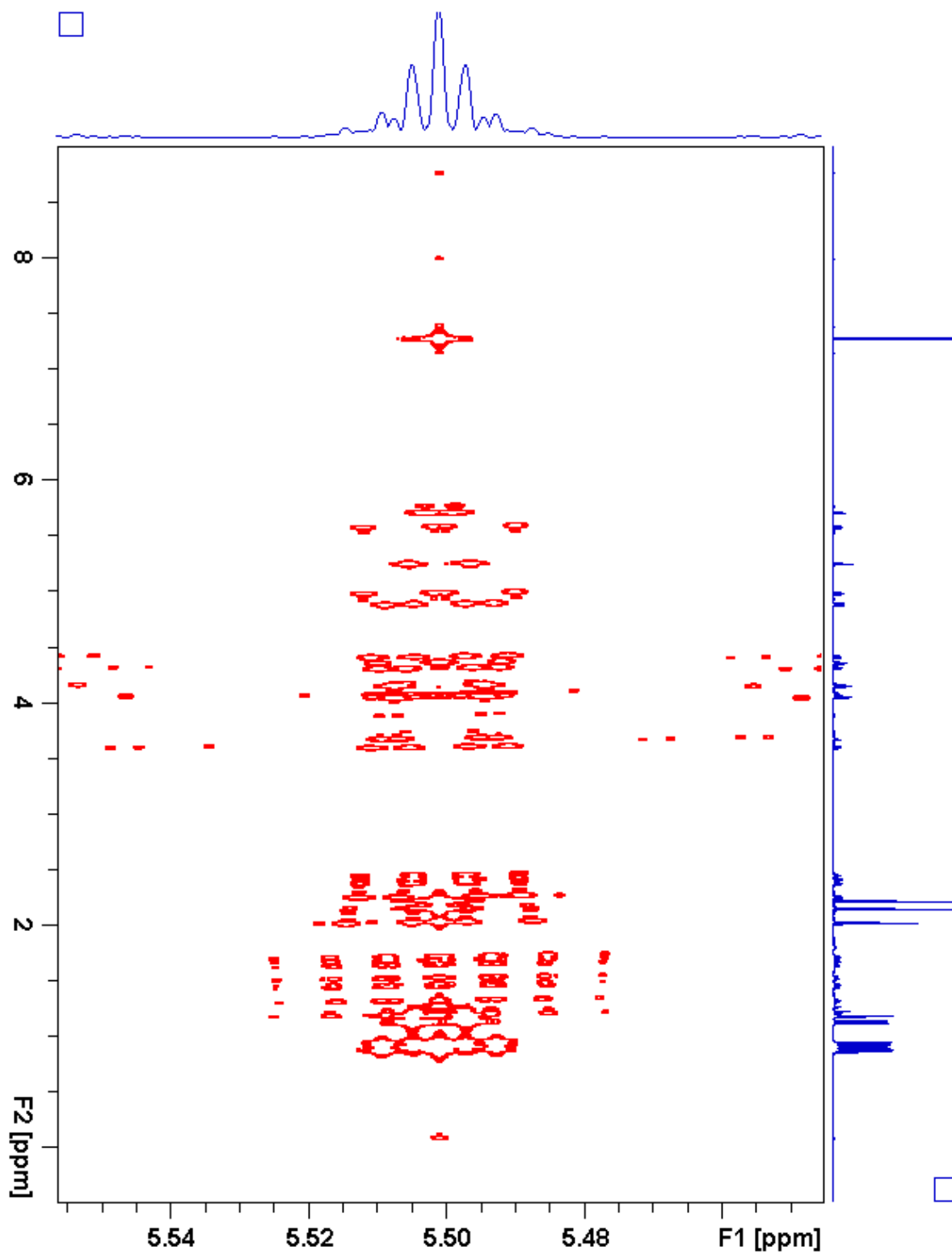
**Figure 2.140.** S6:25:0(2,2,5,5,6) *J*-resolved
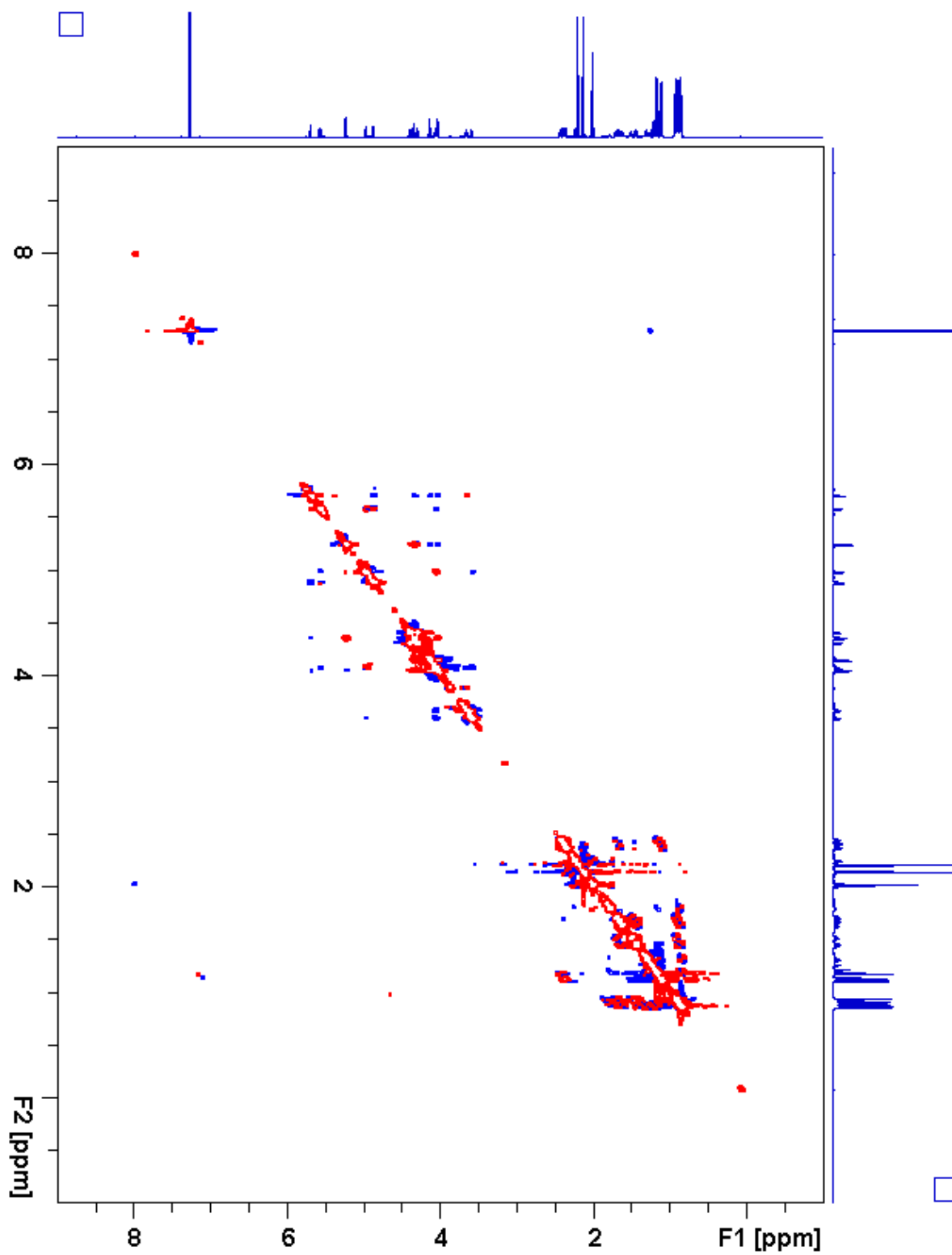
**Figure 2.141.** S6:25:0(2,2,5,5,5,6) ROESY

**REFERENCES**

**REFERENCES**

1.  Dixon, R.A. and D. Strack, *Phytochemistry meets genome analysis, and beyond.* Phytochemistry, 2003. **62**(6): p. 815-6.

2.  Pichersky, E. and E. Lewinsohn, *Convergent evolution in plant specialized metabolism*, in *Annual Review of Plant Biology, Vol 62*, S.S. Merchant, W.R. Briggs, and D. Ort, Editors. 2011, Annual Reviews: Palo Alto. p. 549-566.

3.  Weinhold, A. and I.T. Baldwin, *Trichome-derived O-acyl sugars are a first meal for caterpillars that tags them for predation.* Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7855-9.

4.  Balcke, G.U., et al., *Multi-omics of tomato glandular trichomes reveals distinct features of central carbon metabolism supporting high productivity of specialized metabolites.* Plant Cell, 2017. **29**(5): p. 960-983.

5.  Ghosh, B. and A.D. Jones, *Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar esters).* Anal. Methods, 2017.

6.  Schilmiller, A.L., R.L. Last, and E. Pichersky, *Harnessing plant trichome biochemistry for the production of useful compounds.* Plant J, 2008. **54**(4): p. 702-11.

7.  Glas, J.J., et al., *Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores.* Int J Mol Sci, 2012. **13**(12): p. 17077-103.

8.  Kortbeek, R.W.J., et al., *Engineering of tomato glandular trichomes for the production of specialized metabolites*, in *Synthetic Biology and Metabolic Engineering in Plants and Microbes, Pt B: Metabolism in Plants*, S.E. O'Connor, Editor. 2016, Elsevier Academic Press Inc: San Diego. p. 305-331.

9.  Tissier, A., *Glandular trichomes: what comes after expressed sequence tags?* Plant J, 2012. **70**(1): p. 51-68.

10. Simmons, A.T. and G.M. Gurr, *Trichomes of Lycopersicon species and their hybrids: effects on pests and natural enemies.* Agricultural and Forest Entomology, 2005. **7**(4): p. 265-276.

11. Alba, J.M., M. Montserrat, and R. Fernandez-Munoz, *Resistance to the two-spotted spider mite (Tetranychus urticae) by acylsucroses of wild tomato (Solanum pimpinellifolium) trichomes studied in a recombinant inbred line population.* Exp Appl Acarol, 2009. **47**(1): p. 35-47.

12. Ghosh, B., T.C. Westbrook, and A.D. Jones, *Comparative structural profiling of trichome specialized metabolites in tomato (Solanum lycopersicum) and S. habrochaites: acylsugar profiles revealed by UHPLC/MS and NMR.* Metabolomics, 2014. **10**(3): p. 496-507.

13. King, R.R., et al., *Characterization of 2,3,4,3'-tetra-O-acylated sucrose esters associated with the glandular trichomes of Lycopersicon-typicum.* Journal of Agricultural and Food Chemistry, 1993. **41**(3): p. 469-473.

14. Liu, X.X., et al., *Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of Petunia species.* Metabolomics, 2017. **13**(7): p. 85.

15.    Kroumova, A.B., D. Zaitlin, and G.J. Wagner, *Natural variability in acyl moieties of sugar esters produced by certain tobacco and other Solanaceae species.* Phytochemistry, 2016. **130**: p. 218-27.

16.    D'Auria, J.C., *Acyltransferases in plants: a good time to be BAHD.* Curr Opin Plant Biol, 2006. **9**(3): p. 331-40.

17.    Fan, P., et al., *In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network.* Proc Natl Acad Sci U S A, 2016. **113**(2): p. E239-48.

18.    Schilmiller, A.L., A.L. Charbonneau, and R.L. Last, *Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16377-82.

19.    Schilmiller, A.L., et al., *Functionally divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in Solanum trichomes.* Plant Cell, 2015. **27**(4): p. 1002-17.

20.    Nadakuduti, S.S., et al., *Characterization of trichome-expressed BAHD acyltransferases in petunia axillaris reveals distinct acylsugar assembly mechanisms within the Solanaceae.* Plant Physiol, 2017. **175**(1): p. 36-50.

21.    Moghe, G.D., et al., *Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway.* Elife, 2017. **6**.

22.    Fan, P., et al., *Evolution of a flipped pathway creates metabolic innovation in tomato trichomes through BAHD enzyme promiscuity.* Nat Commun, 2017. **8**(1): p. 2080.

23.    Arrendale, R.F., et al., *Characterization of the sucrose ester fraction from Nicotiana-glutinosa.* Journal of Agricultural and Food Chemistry, 1990. **38**(1): p. 75-85.

24.    Matsuzaki, T., et al., *Leaf surface glycolipids from Nicotiana-acuminata and Nicotiana-pauciflora.* Agricultural and Biological Chemistry, 1991. **55**(5): p. 1417-1419.

25.    Sarkinen, T., et al., *A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree.* BMC Evol Biol, 2013. **13**: p. 214.

26.    Ghosh, B. and A.D. Jones, *Dependence of negative-mode electrospray ionization response factors on mobile phase composition and molecular structure for newly-authenticated neutral acylsucrose metabolites.* Analyst, 2015. **140**(19): p. 6522-31.

27.    Ekanayaka, E.A., M.D. Celiz, and A.D. Jones, *Relative mass defect filtering of mass spectra: a path to discovery of plant specialized metabolites.* Plant Physiol, 2015. **167**(4): p. 1221-32.

28.    Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics, 2007. **3**(3): p. 211-221.

29.    Robins, R.J., et al., *Esterification reactions in the biosynthesis of tropane alkaloids in transformed root cultures.* Plant Cell Tissue and Organ Culture, 1994. **38**(2-3): p. 241-247.

30.    Josephnathan, P., J.R. Wesener, and H. Gunther, *A two-dimensional NMR-study of angelic and tiglic acid.* Organic Magnetic Resonance, 1984. **22**(3): p. 190-191.

31. Kroumova, A.B. and G.J. Wagner, *Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants.* Planta, 2003. **216**(6): p. 1013-21.

32. Treutler, H., et al., *Discovering regulated metabolite families in untargeted metabolomics studies.* Anal Chem, 2016. **88**(16): p. 8082-90.

33. Castillo, M., et al., *Partially acylated glucose and sucrose derivatives from Salpiglossis-sinuata (Solanaceae).* Journal of Chemical Research-S, 1989(12): p. 398-399.

34. King, R.R., et al., *3,4-di-O-isobutyryl-6-O-caprylsucrose - the major component of a novel sucrose ester complex from the type-B glandular trichomes of Solanum-berthaultii Hawkes (Pi 473340).* Journal of the Chemical Society-Chemical Communications, 1986(14): p. 1078-1079.

35. King, R.R., R.P. Singh, and L.A. Calhoun, *Isolation and characterization of 3,3',4,6-tetra-O-acylated sucrose esters from the type-B glandular trichomes of Solanum-berthaultii Hawkes (Pi-265857).* Carbohydrate Research, 1987. **166**(1): p. 113-121.

36. Begum, A.S., et al., *New pairs of acyl sucroses from Petunia nyctaginiflora Juss.* Journal of the Indian Chemical Society, 2004. **81**(6): p. 495-500.

37. Begum, A.S., et al., *Two novel acyl sucroses from Petunia nyctaginiflora.* Indian Journal of Chemistry Section B-Organic Chemistry Including Medicinal Chemistry, 2005. **44**(3): p. 648-650.

38. Singh, A.P., et al., *Two acyl sucroses from Petunia nyctaginiflora.* Phytochemistry, 2003. **63**(4): p. 485-9.

39. Severson, R.F., et al., *Isolation and characterization of the sucrose esters of the cuticular waxes of green tobacco leaf.* Journal of Agricultural and Food Chemistry, 1985. **33**(5): p. 870-875.

**Chapter 3: Unexpected diversity in acylsugar metabolites: acylinositols from *Solanum quitoense***

**3.1 Introduction**

Plants across the family Solanaceae have proven to be prolific at synthesizing diverse arrays of acylsugar specialized metabolites (SMs) in epidermal cells known as glandular trichomes (GTs) [1]. Documented acylsugars in the genus *Solanum* include glucose and sucrose decorated with branched and straight chain aliphatic ester groups [2, 3], malonyl esters in *Petunia* [4], and acylsucroses esterified by saturated and unsaturated ester groups including the new report of phenylacetyl and tiglyl esters in *Salpiglossis sinuata* (Chapter 2 of this dissertation). Recent efforts to elucidate the biochemical pathways responsible for the production of acylsugars have identified a class of BAHD acyltransferase enzymes expressed in GTs of Solanaceae species that use CoA thioesters and sucrose substrates to catalyze multi-step position-selective acylation of sucrose substrates [5-8]. Of particular note is the observation that small changes in amino acid sequences in the acyltransferases can alter acyl acceptor substrate preferences.

Variations in acyltransferase substrate preference raises the potential that mutations in BAHD acyltransferase sequences might allow these enzymes to catalyze acylations of alternative carbohydrate cores. To explore acylsugar chemistry in the Solanaceae more broadly, the focus of this study has been to define acylsugar diversity from GTs of *Solanum quitoense* (also known as naranjilla or lulo) (images in Appendix Figure 3.7), a species that is genetically removed from oft-studied acylsugar producers of the family Solanaceae, such as wild tomato (*Solanum pennellii* and *Solanum habrochaites*), tobacco (*Nicotiana tabacum*), and *Petunia* species [2-4, 9, 10].

*S. quitoense* is a little-known species native to South America that is widely appreciated for the pleasant aroma and intense flavor of its fruits [11, 12], and has been suggested as a promising source for new food flavors. Investigations of *S. quitoense* chemistries have targeted flavor and aroma producing specialized metabolites from fruits. *Forero et al.* found that odor-active volatile compounds (Z)-3-hexenal, ethyl butanoate, 3-sulphanylhexyl acetate, and ethyl hexanoate were key aroma compounds [11]. Processing fruit

(juicing or drying) has been used to extend post-harvest shelf life, but spermidine derivatives formed during processing contributed to undesirable bitterness [13]. A series of $C_{13}$-norisoprenoid glucoconjugates were isolated from homogenized leaves of *S. quitoense*, and are believed to contribute to flavor and aroma after they are transported to the fruit and deglycosylated [14]. Further studies of volatile compounds from glycosidic precursors isolated from fruit peelings act as precursors for monoterpenes (linalool, hotrienol, nerol oxide and linalool oxides), which contribute to the fruit flavor [14]. At this stage, little is known about other specialized metabolites that are produced outside the fruit, particularly what plant chemical defenses may be employed in leaf surfaces of *S. quitoense* and how they may differ from previous examinations of species of the Solanaceae.

In this investigation, the specialized metabolites of *S. quitoense* aerial tissues were explored by subjecting extracts of *S. quitoense* leaves and stems to untargeted metabolite profiling using UHPLC/MS. The findings present evidence of acylsugars (believed to be GT derived) with fewer elemental formulas and fewer isomers than prolific acylsugar accumulators in the genus *Solanum* [2, 15], and these displayed unusual molecular masses and fragment ion spectra. To uncover the novelty in acylsugar composition from *S. quitoense* and facilitate discovery of genetic factors responsible for their biosynthesis, several acylsugar metabolites were purified and their structures identified using NMR spectroscopy. In doing so, acylsugars with unexpected carbohydrate chemistries were discovered, which include a group of acylated *myo*-inositols and *myo*-inositol glycosides (*N*-acetylglucosaminyl, glucopyranosyl and xylopyranosyl), suggesting novel enzyme substrate selectivities are at work in acylsugar biosynthesis in *S. quitoense*. Where previous reports about acylsugars from the family Solanaceae have identified novel ester groups and differences among acyl group substitution positions on glucose- and sucrose-based acylsugars, the current findings indicate that structures of core carbohydrates in acylsugars are more variable than is widely appreciated.

## 3.2 Materials and Methods

### 3.2.1 Plant Cultivation and metabolite extraction

*S. quitoense* seeds were obtained from The New York Botanical Gardens (NY, USA), and were grown using Jiffy peat pellets in growth chambers at MSU. Additional plant growth and extraction metadata are provided in Appendix Table 3.2. For metabolite profiling and purification, aerial tissues of 20 ten-week-old *S. quitoense* plants (~0.25 m height) were extracted in 1.9 L of acetonitrile: isopropanol (AcN:IPA, v/v, 1:1) for 10 mins (plants were cut at the stems and stem junctions). For metabolite profiling, a 100 µL aliquot of this bulk extract solution was diluted in 800 µL AcN:IPA containing 0.1 µM telmisartan added as internal standard and 100 µL of aqueous 0.046 M formic acid (to inhibit rearrangements of acyl group positions). A previous study from our laboratories showed that gentle extraction of young leaflets of *S. quitoense* revealed acylsugar metabolites attributed to trichomes [1]. For comparison of chromatographic retention times and mass spectra, a *Solanum pennellii* LA0716 leaf dip extract was prepared by removing two leaflets from an older plant in a growth chamber (age was unknown) and extracting at 25˚C with AcN:IPA for ~2 mins.

### 3.2.2 Profiling of acylsugar metabolites using UHPLC/MS and MS/MS

Metabolite profiling was performed using a Waters Xevo G2-XS quadrupole time-of-flight mass spectrometer (QTof/MS) equipped with an Ascentis Express C18 Analytical HPLC column (10 cm × 2.1 mm, 2.7 µm particle size) and operated using CID in positive- and negative-ion modes. Untargeted CID spectra were acquired over *m/z* 50-1500 using MS$^E$ [16] in continuum mode (6 V and 10-60 V collision potential ramp to generate fragment ions, each with 0.2 s acquisition times). Additional metadata regarding the UHPLC/MS methods are provided in Appendix Table 3.3. Deep profiling of metabolites was performed using Waters Progenesis QI software platform. Additional metadata regarding Progenesis QI settings and strategy are provided in Appendix Table 3.4. Acylsugar annotations from negative- and positive-ion mode LC/MS data are provided in Appendix Tables 3.5 and 3.6.

MS/MS product ion spectra were generated using a CID ramp from 5-60 V collision potential (0.5 s per spectrum). Acylsugar precursor adducts [M+HCOO]- and [M+NH$_4$]+ were selected for MS/MS analysis in positive and negative-ion modes respectively (MS/MS precursor mass and time windows are shown in Appendix Tables 3.7 and 3.8), with low mass and high mass resolving quadrupole resolution settings at 20 (Waters system specific parameters) and sensitivity mode data acquisition. MS/MS product ion spectra of [M-H]- and [M+H]+ were generated by altering cone voltages to produce those precursor ions. Additional source parameters and LC/MS metadata are located in Appendix Table 3.9. MS/MS product ion spectra of acylinositol metabolites identified by NMR analysis are located in Appendix Figures 3.8-3.25.

Annotation of *S. quitoense* acylsugars by LC/MS was carried out from ESI mass spectra generated in positive- and negative-ion modes by assigning molecular mass and formula, the masses of acyl groups and the number of acylations, and by assigning the elemental formulas (EF) of the carbohydrate cores and relative ring substitutions when pertinent. This information was assembled using untargeted MS$^E$ acquisition with two quasi-simultaneous collision potential functions (6 V and 10-60 V collision potential ramp, 0.2 s acquisition times). Abundant low energy ammonium [M+NH$_4$]+ and formate [M+HCOO]- adduct ions, as well as sodium [M+Na]+ and chlorine [M+Cl]- ions assisted assignments of EFs. Collision induced dissociation (CID) at elevated potentials was used to generate useful fragment ion data for assigning acyl group carbon lengths, the number of acylations, the mass of the carbohydrate core and the relative ring substitution of acyl groups. Further evidence supporting structural assignment was performed by examination of high energy CID spectra followed by metabolite purification and NMR analysis.

### 3.2.3 Purification of acylsugar metabolites by semi-preparative HPLC

For metabolite purification, approximately 1 L of the *S. quitoense* bulk extract (described in Section 3.2.1) was concentrated to dryness under vacuum, dissolved in 5 mL of AcN:IPA, and fractionated by repeated injection of 200-µL samples onto a Thermo Scientific Acclaim 120 C18 HPLC column (4.6 x 150 mm, 5 µm particles) with automated fraction collection (Supplemental Materials). Additional methods and metadata are provided in Appendix Table 3.10.

### 3.2.4 Analysis of acylsugars by NMR spectroscopy

HPLC fractions of sufficient purity of a single metabolite, as assessed by LC/MS, were combined and concentrated to dryness under a flow of $N_2$ gas. Samples were dissolved in acetonitrile-$d_3$ (99.96 atom % D) and transferred to solvent-matched Shigemi tubes or Kontes tubes for NMR analysis. ^{1}H, ^{13}C, gCOSY, gHSQC, coupled-gHSQC, gHMBC, *J*-resolved ^{1}H, TOCSY and ROESY NMR experiments were performed using a Bruker Avance 900 MHz spectrometer equipped with a TCI triple resonance probe or an Agilent DDR2 500 MHz spectrometer equipped with OneNMR probe (with Protune accessory for hands-off tuning). 1D-TOCSY transfer experiments were performed using a Varian Inova  600 MHz spectrometer equipped with a Nalorac 5 mm PFG switchable probe (pretuned for ^{1}H and ^{13}C). All spectra were referenced to non-deuterated solvent signals: acetonitrile-$d_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm). NMR spectra were processed using TopSpin 3.5pl7 or MestReNova 12.0.0 software. Additional NMR metadata are located in Appendix Tables 3.11-3.13. NMR spectra, chemical shifts and coupling constants are located in Appendix Figures 3.26-3.95 and Tables 3.14-3.28.

1D (^{1}H, ^{13}C and TOCSY) and 2D (gCOSY, gHSQC, gHMBC, *J*-resolved, TOCSY and ROESY) NMR spectroscopic techniques served as the basis for structure elucidation of nine purified *S. quitoense* acylsugars. Because these metabolites were identified without authentic standards or synthetic confirmation, their structures should be considered putative, meeting level 2 criteria of the Metabolomics Standards Initiative guidelines [17].

### 3.3 Results and Discussion

### 3.3.1 UHPLC/ESI/CID/QTof/MS profiling and NMR structural elucidation establishes diversity of *S. quitoense* acylinositols

Profiling of *S. quitoense* metabolites using UHPLC/QTof MS revealed a group of *S. quitoense* metabolites eluting between 35-86 mins using a 110-min gradient that exhibited molecular masses, relative mass defects and fragment ion spectra consistent with acylsugar SMs [4, 18, 19]. Compared to previous reports of acylsugar profiles in *Solanum habrochaites* accessions [2] and three *Petunia* species [4], LC/MS

chromatograms of *S. quitoense* exhibited fewer peaks annotated as acylsugars. As shown in Figure 3.1, at least 9 abundant acylhexoses and acyl disaccharides exhibited retention times in the range of 57-81 mins and are evident in the base peak intensity (BPI) chromatogram.
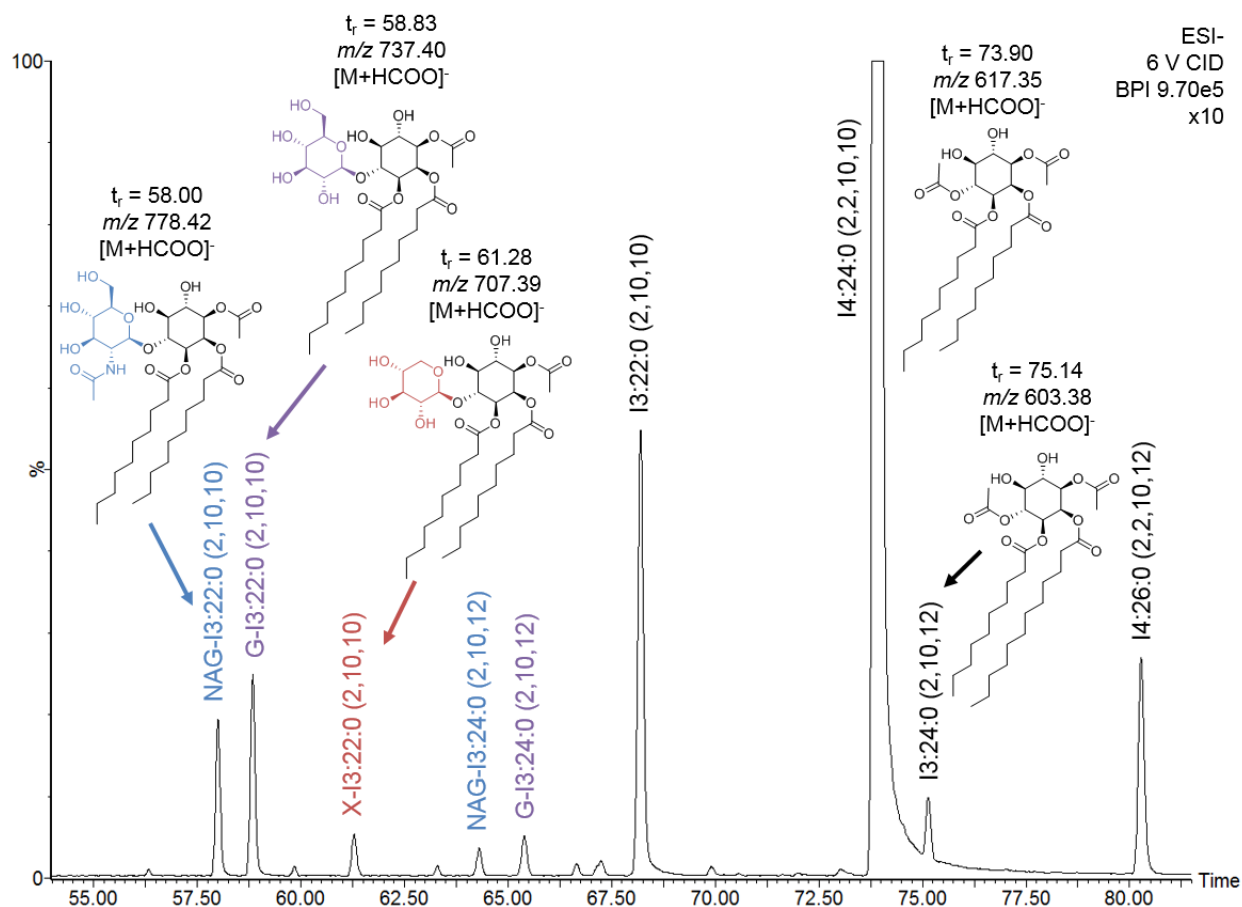


**Figure 3.1.** UHPLC/ESI/MS base peak intensity chromatogram of *S. quitoense* extract generated in negative-ion mode. Annotated acylsugars (formate adducts) detected using CID Function 1 = 6 V (10x magnification). Abbreviations of sugar groups are as follows: *myo*-inositol (I), *N*-acetylglucosaminyl (NAG), glucopyranosyl (G), and xylopyranosyl (X). One example for abbreviation nomenclature is I:24:0(2,2,10,10), where "I" indicates the carbohydrate core is myo-inositol, the number "4" indicates four ester groups, the number "24" reflects the total number of carbon atoms across all acyl groups, the numeral "0" indicates zero rings or double bonds in the acyl groups, and the numbers in parentheses describe the number of carbon atoms in each acyl group.

### 3.3.2 LC/MS profiling and NMR structural elucidation establishes acylated *myo*-inositols

Profiling and annotation of *S. quitoense* acylsugars by LC/MS was carried out by generating ESI mass spectra in positive- and negative-ion modes. To illustrate this, the most abundant peak in the base peak intensity (BPI) chromatogram (Figure 3.1) at $t_r$ = 73.90 mins yielded mass spectra with ions annotated as [M+HCOO]$^-$ at *m/z* 617.35 in negative-ion mode and [M+NH$_4$]$^+$ at *m/z* 590.39 in positive-ion mode (Figure

3.2A-B). These ion masses are consistent with an acylated hexose of EF $C_{30}H_{52}O_{10}$. Because hexoses have six oxygen atoms, the four additional oxygens suggested four acylations, as each acylation adds one oxygen atom. Negative-ion spectra generated at elevated collision potentials showed a fragment ion at *m/z* 171.14, consistent with C10 carboxylate (acyl anion) of EF $C_{10}H_{19}O_2^-$ (Figure 3.2A). Acylated hexoses from *S. quitoense* shared molecular masses with calculated masses of acylated glucoses. Information obtained from negative-ion mode CID spectra alone usually facilitates assignments of the number of carbon atoms of each ester group [18]. However, their negative-ion mode CID spectra generated using a collision potential ramp from 10-60 V exhibited lower abundances of fragment ions than observed for acylglucoses or acylsucroses (Figure 3.2A). Positive-ion mode spectra obtained at elevated collision energy showed abundant fragment ions that were informative for annotating the number acylations and the number of carbons in each ester group, whereas this information was not readily obtained from negative-ion mode spectra. Further confirmation of C10 carboxylic acid esters comes from observation of the *m/z* 155.14 acylium fragment ion with EF $C_{10}H_{19}O^+$ and neutral mass losses from $[M+H]^+$ of 172.15 Da (decanoic acid: $C_{10}H_{20}O_2$) yielding *m/z* 401.22 (Figure 3.2B). Neutral loss of 60.02 Da (acetic acid: $C_2H_4O_2$) from $[M+H]^+$ generated in-source yielded *m/z* 513.34, consistent with an acetate (C2) ester. The neutral losses of C2 or C10 groups from $[M-C10-C2]^+$ at *m/z* 341.20 revealed mono-esterified fragment masses consistent with $[M-C10-2C2]^+$ at *m/z* 281.17 and $[M-2C10-C2]^+$ at *m/z* 169.05 respectively. Thus, the acylated hexose was assigned two C2 and two C10 groups. Using only LC/MS-based analyses, mass spectra of *S. quitoense* acylated hexoses shared resemblance to previously discovered acylglucose derivatives [20, 21]. This is because the hexose core $(C_6H_{12}O_6)$ shares the molecular formula of glucose and all other carbons, hydrogen and oxygen atoms were accounted for by acyl group assignments.
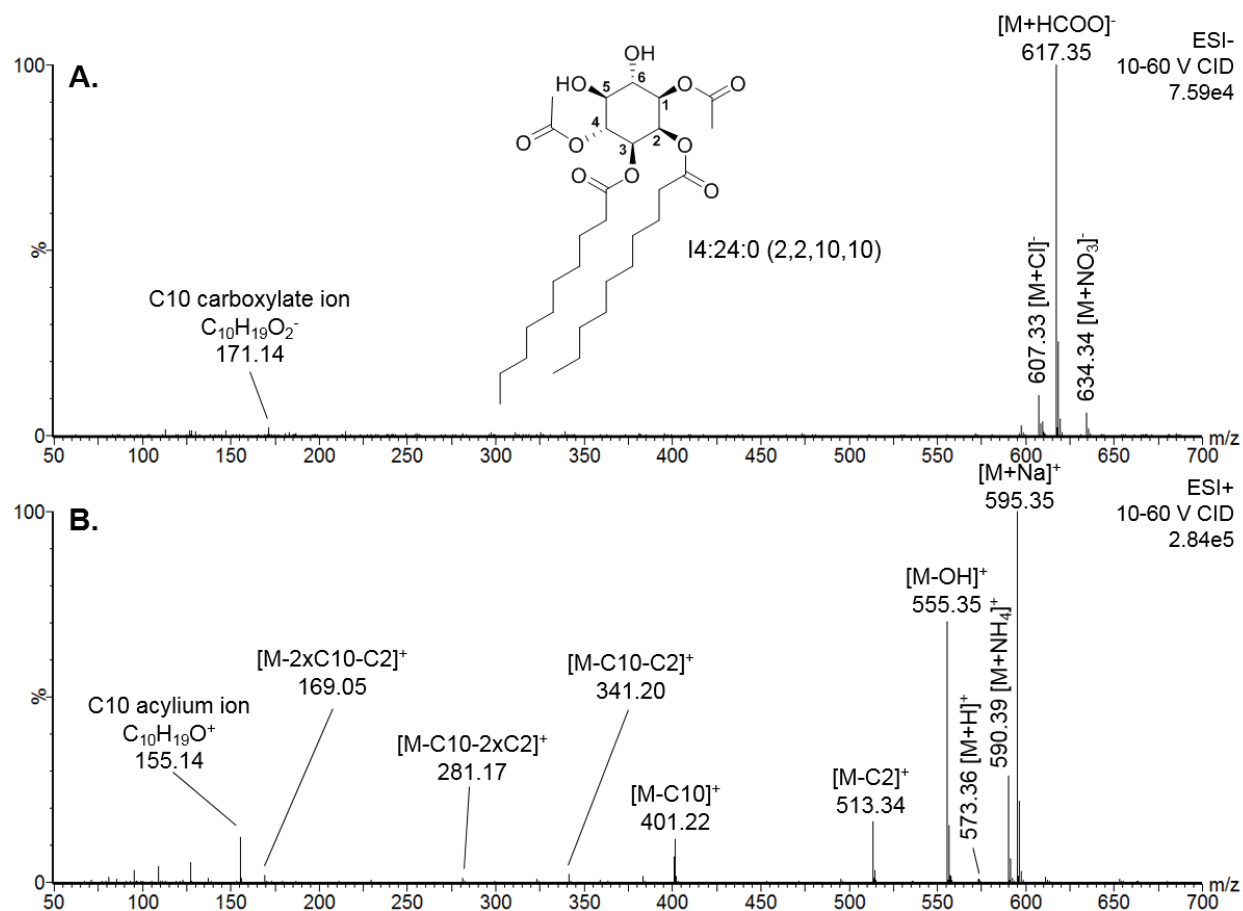
**Figure 3.2.** CID mass spectra of acylsugar at $t_r$ = 73.90 mins (later annotated I4:24:0(2,2,10,10) according to NMR results) using 10-60 V MS$^E$ ramp (0.2 s acquisition times). **(A)** Negative ion mode. **(B)** Positive ion mode.

While LC/MS-based techniques provide annotations of acylsugar metabolites in terms of the number of acyl groups and the acyl group chain lengths, these analyses yield limited information about the atomic connectivity and stereochemistry of atoms that define each carbohydrate group. A surprising finding from structural characterization of purified *S. quitoense* acylsugars by NMR spectroscopy revealed chemical shifts that differed from acylglucoses [20, 21]. For acylsugar at $t_r$ = 73.90 min, the absence of a characteristic downfield doublet for the anomeric (*α* or *β*) ring hydrogen was inconsistent with glucoses [20]. $^1$H spectra of hexose core hydrogens were consistent with the conclusions that acylated hexose positions exhibited chemical shifts in the range of $\delta_H$ 4.7-5.5 ppm that were attributed to ~1 ppm downfield shifts compared to non-esterified analogues, which ranged from $\delta_H$ 3.3-3.8 ppm (ring hydrogen chemical shifts are summarized in Appendix Table 3.14). In addition, HSQC spectra (used to determine the number of protons attached to

each carbon) were absent of methylene ($CH_2$) ring hydrogens and suggested a cyclitol ring system, further confirmed by 2D-TOCSY, which showed a complete set of $^1H$-$^1H$ spin-spin coupled ring hydrogens. Ring $^1H$-$^1H$ couplings determined from $^1H$ and *J*-resolved spectra (and confirmed by COSY spectra) showed decisive coupling constants for discerning the relative stereochemistry of hexose ring hydrogens. A hydrogen labeled **H-2** was shifted furthest downfield and served as the starting point for tracing the connectivity and relative stereochemistry of the cyclitol ring system (Figure 3.2). An equatorial **H-2** signal at 5.49 ppm showed couplings to axial proton **H-1** at 4.81 ppm and **H-3** at 4.97 ppm ($J_{1,2} = J_{2,3} = 3.0$ Hz), both of which were shifted downfield to suggest esters at these positions. The **H-1** signal showed coupling to axial **H-6** at 3.77 ppm ($J_{1,6} = 10.2$ Hz) which lacked an ester group, while **H-3** signal showed coupling to axial **H-4** at 5.21 ppm ($J_{3,4} = 10.0$ Hz) at the site of another ester group. Finally, **H-5** signal at 3.50 ppm (position not esterified) was consistent with axial couplings to **H-4** and **H-6** ($J_{4,5} = J_{5,6} = 9.5$ Hz), confirming a six-membered cyclitol ring. Thus, from NMR results the hexose core was assigned as a *myo*-inositol tetraester. Assignments of acyl group branching and attachment positions of acylinositols are discussed in Section 3.3.6.

Based on the MS and NMR results, the acylinositol whose mass spectrum is shown in Figure 3.2A was annotated as I4:24:0(2,2,10,10), using abbreviation nomenclature where "I" indicates the carbohydrate core is *myo*-inositol, the number "4" indicates four ester groups, the number "24" reflects the total number of carbon atoms across all acyl groups, the numeral "0" indicates zero rings or double bonds in the acyl groups, and the numbers in parentheses describe the number of carbon atoms in each acyl group. Further evidence of this structural assignment is provided in Appendix Table 3.24 and Appendix Figures 3.70-3.77.

The basis for the clockwise carbon numbering system used for assignment of *myo*-inositols in this study traces from the biosynthetic conversion of D-glucose 6-phosphate to 1L-*myo*-inositol-1-phosphate [22]. Due to the plane of symmetry of the *myo*-inositol ring system about the C2 and C5 positions, an enantiomeric assignment is possible by substituting the C1 and C3 positions (1D-*myo*-inositol), followed

by counterclockwise numbering. Due to the scarcity of purified material (most < 3 mg), absolute stereochemical configurations were not determined.

### 3.3.3 Negative-ion mode MS/MS spectra of acylated *myo*-inositols differ from acylglucoses

Mass spectra of *myo*-inositol-based acylsugars generated at elevated collision potentials in negative-ion mode showed few fragment ions in the middle of the spectrum corresponding to neutral losses of acyl groups, unlike previously reported acylglucose analogues [20, 21]. To further explore their differences, Figure 3.3A-B compares MS/MS product ion spectra of [M+HCOO]⁻ for tri-acylated inositol I3:24:0(2,10,12) from *S. quitoense* and triacylglucose G3:19:0(4,5,10) from *Solanum pennellii* LA0716 (structure in Figure 3.3B and 3.3D provided by Dr. Banibrata Ghosh and Dr. Xiaoxiao Liu). Both spectra display carboxylate anions for C10 ester groups at *m/z* 171, however acylglucoses differ from acylinositols in that the former yield prominent product ions derived from acyl group neutral losses, including displacement of the C10 acid (neutral loss of $C_{10}H_{20}O_2$, 172.15 Da) by formate (CHOO, 45.00 Da) to the glucose ring (*m/z* 361.15, [M-C10+HCOO]⁻). This suggests that proximity of charge to ester groups promotes generation of either carboxylate anions from ester groups or neutral losses of ketenes, and is more significant in acylglucoses than acylinositols. Another acylglucose ion showed the neutral loss of $H_2O$ (18.01 Da) and C10 (154.14 Da) as a neutral ketene (*m/z* 315.14, [M-C10-H₂O]⁻). A neutral ketene loss is expected to result from removal of an alpha proton (α-H) of an ester group via an alkoxide anion. The intermediate negative ion can displace the oxygen where the acyl group is attached, liberating a neutral ketene group such as C10. The observation of a fragment ion at *m/z* 315.14 suggests the ion is further stabilized by neutral loss of $H_2O$, though concerted elimination of the neutral carboxylic acid cannot be ruled out. In the case of acylated *myo*-inositol, it is likely that the energy barrier for an alkoxide ion to reach across the molecule to remove an ester group alpha proton is much greater than in the case of glucose (where the charge can migrate to one of the primary alcohol oxygens that can reach across the rings with a lower activation energy barrier). Interestingly, the MS/MS product ion spectrum of [M-H]⁻ of acylinositol (Figure 3.3C) shows similar relative ion abundances for C10, C12 (*m/z* 199.17) and [M-H]⁻ by comparison

to MS/MS product ions of [M+HCOO]⁻ (Figure 3.3A), suggesting the pathway for generating carboxylate ions goes through neutral loss of formic acid, formation of alkoxide ion, followed by liberation of carboxylate anion (perhaps via displacement by an alkoxide ion on an adjacent carbon). In contrast, the MS/MS product ion spectrum of [M-H]⁻ of acylglucose (Figure 3.3D) exhibits different abundances of neutral losses relative to carboxylate formation when compared to MS/MS product ions from [M+HCOO]⁻ in that lesser abundant neutral loss fragment ions are observed in the center of the spectrum (Figure 3.3B). These results suggest there are multiple fragmentation pathways for ion activation of acylglucoses, the formate adduct plays a role in this process by driving charge localization, and the acylsugar core structure influences CID fragmentation in ways that may discriminate different carbohydrate cores in acylsugars.
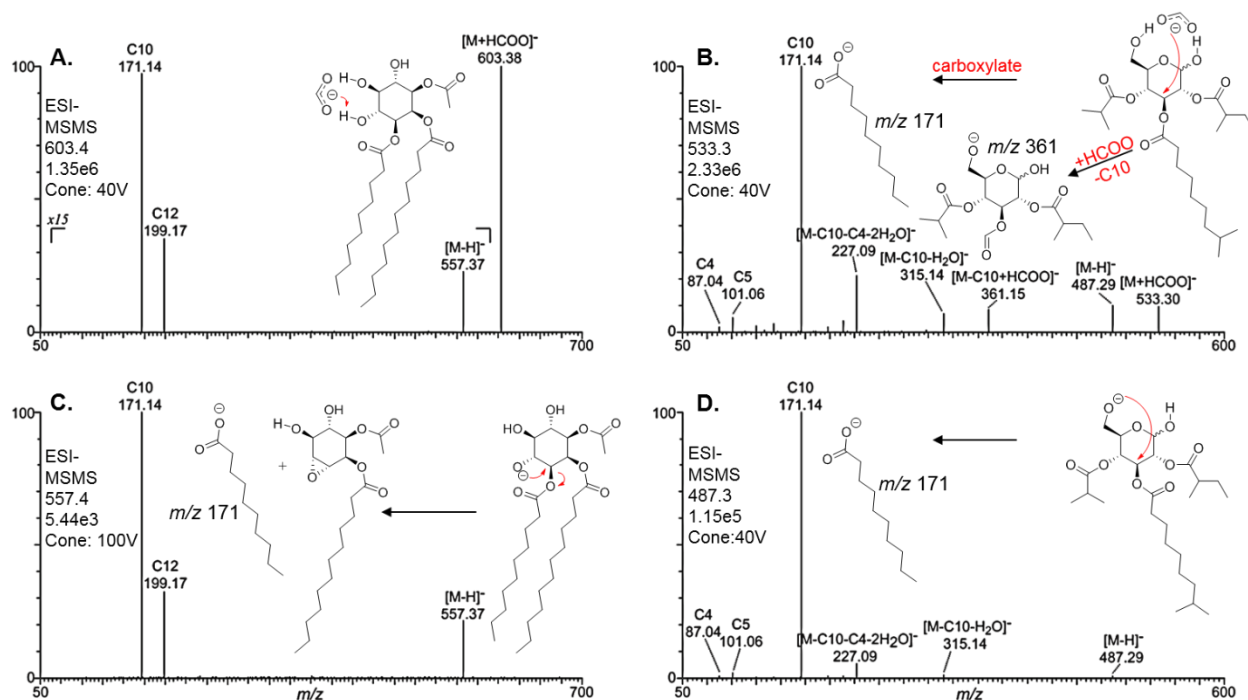


**Figure 3.3.** Negative ion mode MS/MS product ion spectra of tri-acylated *myo*-inositol I3:24:0(2,10,12) from *S. quitoense* and triacylated glucose G3:19:0(4,5,10) from *Solanum pennellii* LA0716 (all spectra generated with a linear 5-60 V MSᴱ collision energy ramp with 0.5 s acquisition time) **(A)** products of I3:24:0(2,10,12) [M+HCOO]⁻ (spectrum magnified 15x over the range *m/z* 51-593) **(B)** products of G3:19:0(4,5,10) [M+HCOO]⁻. **(C)** products of I3:24:0(2,10,12) [M-H]⁻. **(D)** products of G3:19:0(4,5,10) [M-H]⁻.

**3.3.4 LC/MS profiling and NMR structural elucidation establishes acylated *myo*-inositol glycosides**

**3.3.4.1 Discovery of 4-O-*N*-acetylglucosaminyl (NAG) acylated *myo*-inositols**

Although acylated *myo*-inositols were of high abundance in the *S. quitoense* extracts, a group of less abundant *myo*-inositol glycosides were also observed. Documented neutral acylsugar metabolites within the Solanaceae are present as derivatives of carbon, hydrogen and oxygen [2-4]. Therefore, a peak in the *S. quitoense* extract chromatogram at $t_r = 58.00$ min (NAG-I3:22:0(2,10,10) in Figure 3.1) that yielded mass spectra showing $[M+HCOO]^-$ at *m/z* 778.42 (Figure 3.4A) and $[M+H]^+$ at *m/z* 734.43 (Figure 3.4B) stood out because its even ion masses are consistent with a nitrogen-containing acylated *myo*-inositol glycoside of EF $C_{36}H_{63}NO_{14}$. Unlike acylinositol monosaccharide mass spectra, glycosylated inositol forms showed abundant fragment ions in negative-ion CID mass spectra. Negative mode MS/MS product ion spectra of *m/z* 778.42 (Figure 3.4A) show abundant neutral mass losses from $[M-H]^-$ (*m/z* 732.42) such as loss of C10 ketene (loss of 154 Da, *m/z* 578.27) and a product ion resulting from an additional loss of C2 ketene (*m/z* 536.27). In addition, losses of 18.01 Da consistent with $H_2O$ were regularly observed, and are attributed to charge-remote losses of carboxylic acids rather than ketenes. Losses of two C10 and one C2 ketene groups leaves the product ion at *m/z* 382.13, corresponding to the deprotonated disaccharide core with EF $C_{14}H_{24}NO_{11}^-$. Positive-ion mode MS/MS product ion spectra of $[M+NH_4]^+$ at *m/z* 751.46 (Figure 3.4B) showed cleavage of the glyosidic linkage, producing abundant product ion *m/z* 204.09, consistent with an *N*-acetylglucosamine (NAG) group ($C_8H_{14}NO_5^+$). A less abundant fragment ion at *m/z* 513.34 ($C_{28}H_{49}O_8^+$) suggests substitutions of two C10 and one C2 ester groups on the *myo*-inositol ring and no esters on the NAG ring. However, the presence of a NAG glycoside required further confirmation by NMR analysis.
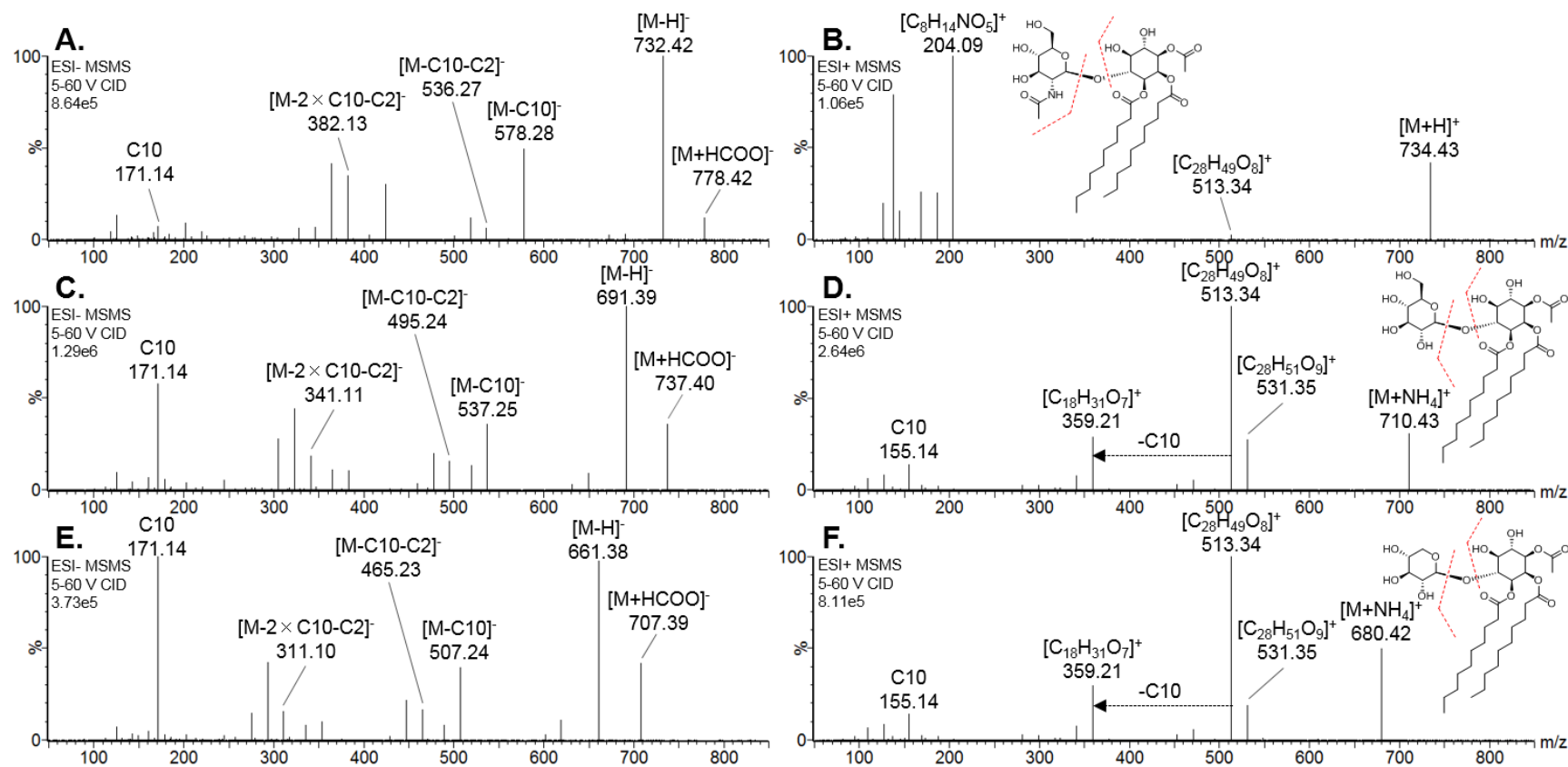
**Figure 3.4.** MS/MS product ion spectra of [M+HCOO]⁻ and [M+NH₄]⁺ of glycosylated acylinositols in an extract of *S. quitoense* using a linear 5-60 V collision energy ramp with 0.5 s acquisition time. **(A)** ESI(-) MS/MS product ion spectrum for *m/z* 778 ([M+HCOO]⁻) for acylsugar NAG-I3:22:0(2,10,10). **(B)** ESI(+) MS/MS product ion spectrum for *m/z* 751 ([M+NH₄]⁺) for acylsugar NAG-I3:22:0(2,10,10). **(C)** ESI(-) MS/MS product ion spectrum for *m/z* 737 ([M+HCOO]⁻) for acylsugar G-I3:22:0(2,10,10). **(D)** ESI(+) MS/MS product ion spectrum for *m/z* 710 ([M+NH₄]⁺) for acylsugar G-I3:22:0(2,10,10). **(E)** ESI(-) MS/MS product ion spectrum for *m/z* 707 ([M+HCOO]⁻) for acylsugar X-I3:22:0(2,10,10). **(F)** ESI(+) MS/MS product ion spectrum for *m/z* 680 ([M+NH₄]⁺) for acylsugar X-I3:22:0(2,10,10).

259

1D and 2D NMR spectroscopic techniques served as the basis for structure elucidation of glycosides of acylinositols (disaccharide chemical structures and numbering are located in Figure 3.5). An anomeric doublet at **H-1′** observed at 4.3-4.5 ppm served as the starting point. The $\beta$ orientation of glycosides was confirmed by $^1$H-$^1$H couplings (d, $J_{1',2'}$ = 7.5-8.3 Hz) and $^1$H-$^{13}$C couplings by coupled-HSQC or from HMBC breakthrough signal ($^{1'}J_{CH}$ = 161-163 Hz) [23]. In all instances, HMBC spectra showed three bond coupling from the anomeric **H-1′** to carbon resonances of *myo*-inositol at position **4**, which ranged from 81.5-83.2 ppm (ring carbon chemical shifts are summarized in Appendix Table 3.15).
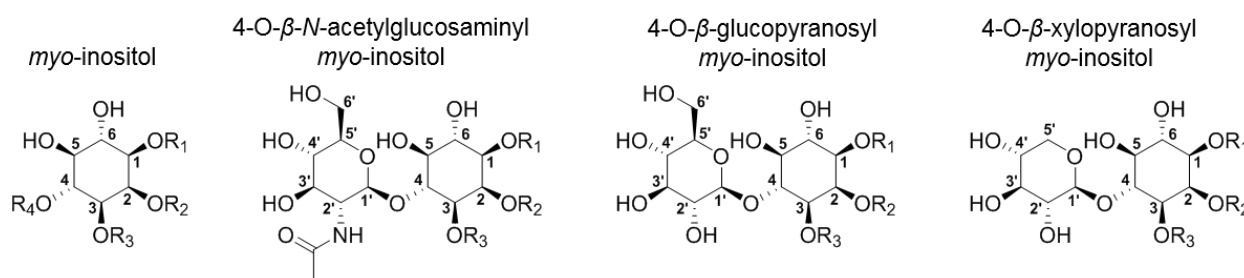


**Figure 3.5.** Chemical structures of core glycosylated *myo*-inositol metabolites from *S. quitoense* determined by NMR spectroscopy. $R_1$, $R_2$, $R_3$ = acylation at that position. $R_4$ = H or acylation for *myo-inositol* monosaccharide. Acylations are listed in Table 3.1.

For acylsugar $t_r$ = 58.00 min (NAG-I3:22:0(2,10,10) in Figure 3.1), the axial **H-1′** signal observed at 4.50 ppm (d, $J$ = 8.4 Hz) exhibited coupling to axial **H-2′** at 3.49 (m) by COSY. HMBC at **H-2′** showed three bond correlation to an acetamido carbonyl carbon signal at 171.55 ppm. Similarly, a **-CH₃** singlet of acetamide centered at 1.82 ppm showed two bond correlations to the carbonyl carbon. Characteristic proton resonances observed for an acetamido **-N-H** at 6.39 ppm and carbon resonance **C-2′** at 57.5 ppm were also telling of an acetylglucosaminyl group. HMBC at **H-2′** showed two bond coupling to **C-1′** at 102.40 ppm and **C-3′** at 75.63 ppm. The remainder of the ring system could be traced by $^1$H, COSY and *J*-resolved from chemical shifts and coupling constants. Axial **H-3′** signal observed at 3.44 ppm (t, $J$ = 9.1 Hz) had coupling to **H-2′** and axial **H-4′** at 3.30 ppm (dd, $J$ = 10.0, 8.1 Hz), **H-4′** had coupling to axial **H-5′** at 3.33 ppm (ddd, $J$ = 10.2, 5.8, 2.8 Hz), and **H-5′** had coupling to diastereotopic methylene hydrogens at **H₂-6′** at 3.78 ppm (dd, $J$ = 11.8, 2.9 Hz) and 3.61 ppm (dd, $J$ = 11.8, 5.8 Hz). Thus, the 4-O-$\beta$-N-acetylglucosaminyl acylated *myo*-inositols core structure was confirmed, and supports the annotation as NAG-I3:22:0(2,10,10).

Assignment of acyl group branching and attachment positions of acylinositols are discussed in Section 3.3.6. Further evidence of this structural assignment is provided in Appendix Table 3.16 and Appendix Figures 3.26-3.32.

**3.3.4.2 Discovery of 4-*O*-glucopyranosyl (G) acylated *myo*-inositols**

Another glycosylated acylinositol at $t_r$ = 58.83 min (G-I3:22:0(2,10,10) in Figure 3.1), exhibiting [M+HCOO]$^-$ at *m/z* 737.40 and [M+NH$_4$]$^+$ at *m/z* 710.43, was assigned the formula $C_{34}H_{60}O_{14}$. Negative mode MS/MS product ion spectra showed neutral mass losses from [M-H]$^-$ of two C10 and one C2 ketene ester groups leaving *m/z* 341.11 (Figure 3.4C), corresponding to EF $C_{12}H_{21}O_{11}$, consistent with previously reported sucrose-based acylsugars. However, initial results would prove misleading. Positive-ion mode MS/MS product ion spectra of [M+NH$_4$]$^+$ at *m/z* 710.43 (Figure 3.4D) showed product ions attributed to cleavage of the glyosidic linkage, producing abundant product ions *m/z* 531.35 and 513.34, consistent with EFs $C_{28}H_{51}O_9{}^+$ and $C_{28}H_{49}O_8{}^+$. Those ions are consistent with neutral mass losses of an anhydrohexose (162.05 Da, $C_6H_{10}O_5$) and hexose (180.06 Da, $C_6H_{12}O_6$) from [M+H]$^+$, and indicate that all the acyl groups (two C10 and one C2) are on the *myo*-inositol ring. From the molecular mass and fragment ion spectra, this acylsugar was consistent with an acylsucrose. However, structural elucidation by NMR spectroscopy showed that this sugar core consisted of a glycoside of *myo*-inositol.

For $t_r$ = 58.83 min (G-I3:22:0(2,10,10) in Figure 3.1), the axial **H-1′** signal observed at 4.33 ppm (d, *J* = 7.9 Hz) had coupling to axial **H-2′** at 3.11 (m) by COSY. HMBC at **H-2′** showed two bond coupling to **C-1′** at 104.69 ppm and **C-3′** at 74.36 ppm. The remainder of the ring system could be traced by $^1$H, COSY and *J*-resolved from chemical shifts and coupling constants. Axial **H-3′** signal observed at 3.27 ppm (m) had coupling to **H-2′** and axial **H-4′** at 3.27 ppm (m), **H-4′** had coupling to axial **H-5′** at 3.33 ppm (ddd, *J* = 9.8, 5.8, 2.7 Hz), and **H-5′** had coupling to diastereotopic methylene hydrogens at **H$_2$-6′** at 3.77 ppm (dd, *J* = 11.9, 2.8 Hz) and 3.61 ppm (dd, *J* = 11.9, 5.9 Hz). Thus, the 4-O-*β*-glucopyranosyl (G) acylated *myo*-inositols core structure was confirmed, and supports the annotation as G-I3:22:0(2,10,10). Assignment of

acyl group branching and attachment positions of acylinositols are discussed in Section 3.3.6. Further evidence of this structural assignment is provided in Appendix Table 3.17 and Appendix Figures 3.33-3.39.

### 3.3.4.3 Discovery of 4-*O*-xylopyranosyl (X) acylated *myo*-inositols

A third type of glycosylated acylinositol, $t_r$ = 61.28 min (X-I3:22:0(2,10,10) in Figure 3.1) exhibiting [M+HCOO]$^-$ at *m/z* 707.39 and [M+NH$_4$]$^+$ at *m/z* 680.42 was assigned the formula $C_{33}H_{58}O_{13}$. Product ion MS/MS spectra were unlike most documented sucrose-based acylsugars from the Solanaceae. Negative mode MS/MS product ion spectra showed neutral mass losses of two C10 and one C2 ester groups as ketenes from [M-H]$^-$, this time leaving *m/z* 311.10 (Figure 3.4E), corresponding to EF $C_{11}H_{19}O_{10}^-$, consistent with a pentosyl *myo*-inositol. Positive-ion mode MS/MS product ion spectra of [M+NH$_4$]$^+$ at *m/z* 680.42 (Figure 3.4F) mirrored G-I3:22:0(2,10,10) (Figure 3.4D), showing cleavage of the glyosidic linkage, again producing abundant fragment ions *m/z* 531.35 and 513.34, consistent with EFs $C_{28}H_{51}O_9^+$ and $C_{28}H_{49}O_8^+$. Neutral mass losses of ammonia plus either 132.04 and 150.05 Da from [M+H]$^+$ were consistent with a pentose group ($C_5H_8O_4$ and $C_5H_{10}O_5$) observed during fragmentation. These results again suggest substitutions of two C10 and one C2 ester groups on the *myo*-inositol ring.

For $t_r$ = 61.28 min (X-I3:22:0(2,10,10) in Figure 3.1), the ring system could be traced by $^1$H, COSY and *J*-resolved from chemical shifts and coupling constants. The axial **H-1′** signal observed at 4.28 ppm (d, *J* = 7.5 Hz) had couplings to axial **H-2′** at 3.11 (dd, *J* = 9.1, 7.4 Hz). Axial **H-3′** signal observed at 3.27 ppm (t, *J* = 8.8 Hz) had coupling to **H-2′** and axial **H-4′** at 3.47 ppm (ddd, *J* = 10.3, 8.4, 5.4 Hz), **H-4′** had coupling to diastereotopic methylene ring hydrogens at **H$_2$-5′** at 3.90 ppm (dd, *J* = 11.4, 5.3 Hz) and at 3.21 ppm (dd, *J* = 11.6, 10.0 Hz). Thus, the 4-O-*β*-xylopyranosyl (X) acylated *myo*-inositols core structure was confirmed, and supports the annotation as X-I3:22:0(2,10,10). Assignment of acyl group branching and attachment positions of acylinositols are discussed in Section 3.3.6. Further evidence of this structural assignment is provided in Appendix Table 3.18 and Appendix Figures 3.40-3.46.

### 3.3.5 Deep profiling of acylinositols by LC/MS

Deep profiling of *S. quitoense* acylsugars using UHPLC/MS in negative- and positive-ion modes revealed 29 different EFs (Appendix Tables 3.5 and 3.6) consistent with acylsugars, and all are annotated as acylated *myo*-inositols and *myo*-inositol glycosides (*N*-acetylglucosaminyl, glucopyranosyl and xylopyranosyl). Notably, acylinositols were not observed in tissues of mature leaflets. The isomeric diversity of *S. quitoense* acylinositols was comparatively less than other Solanaceae species, with only 36 chromatographic peaks annotated using both ionization modes. Most acylinositols and their glycosides were detected in both ionization modes and their molecular masses were consistent with combinations of two medium-length (C10 or C12) ester groups, and varying numbers of C2 ester groups. To our knowledge, di-substituted medium-chain acyl groups (>8 carbons) were not previously detected during profiling of Solanaceae species [2, 4].

The acylinositol peak area with structural identification (nine structures in Table 3.1) was 98.2% by negative-ion mode and 87.6% by positive-ion mode (Appendix Tables 3.5 and 3.6). Positive-ion mode mass spectra suggest that glycosylated *myo*-inositols were only esterified on the inositol moiety and were comprised of tri- and tetra-esters. Acylated *myo*-inositol monosaccharide EFs were consistent with di-esters through fully esterified hexa-esters. For example, an annotation at $t_r = 83.68$ with $m/z$ 674.41 had $[M+NH_4]^+$ consistent with EF $C_{34}H_{56}O_{12}$ a hexa-ester of *myo*-inositol consistent with I6:28:0(2,2,2,2,10,10) (Appendix Table 3.6). However, evidence for penta- and hexa-acylated *myo*-inositol monosaccharides was provided only in positive-ion mode (likely because positive adducts such as ammonium have affinity for ester groups). We interpret the lack of detection of penta- and hexa-acylated *myo*-inositol monosaccharides in negative-ion mode to suggest that at least two non-esterified hydroxyl positions are necessary for formation of $[M+HCOO]^-$ adduct anions, presumably because anions have low affinity for ester groups relative to hydroxyl groups [18].

**3.3.6 1D and 2D NMR of purified acylinositols reveals acylation positions**

Using only LC/MS, *S. quitoense myo*-inositol and glucopyranosyl *myo*-inositol acylsugars resembled previously discovered glucose and sucrose derivatives, in that they had molecular masses consistent with those acylsugars, while *N*-acetylglucosaminyl and xylopyranosyl *myo*-inositol acylsugars had ambiguous core structures that were inconsistent with molecular masses of glucose or sucrose. Because the long-term goal of this work is to understand the diversity of acylsugars from the Solanaceae and their biosynthesis, assignment of the carbohydrates as well as the acyl group branching and attachment positions assists the discovery of acylsugar acyltransferase enzymes responsible for their biosynthesis. In view of this, a total of nine abundant acylinositols were purified by semi-preparative HPLC for structural elucidation by NMR spectroscopy, including four acylated *myo*-inositols, and two 4-O-*β*-*N*-acetylglucosaminyl, two 4-O-*β*-glucopyranosyl and one 4-O-*β*-xylopyranosyl acylated *myo*-inositol glycosides (Figure 3.5).

HMBC was used to measure atomic connectivity, usually ranging 2-4 bonds, and was vital for the assignment of acyl positions on the *myo*-inositol ring. Multiple bond correlations of ring H's of *myo*-inositol and acyl H's to ester carbons enabled determination of positions and types of acylations. For instance, the **H-2** position of I:24:0(2,2,10,10) showed three bond correlation to a carbonyl carbon resonance at 174.11 ppm. Similarly, **2-*α*-CH$_2$** and **2-*β*-CH$_2$** of this acyl group centered at 2.38 and 1.62 ppm showed two and three bond correlations to the carbonyl carbon (resolved acyl hydrogens were annotated with a number prefix to indicate their positions in Figure 3.6A). From the combination of mass spectra and absence of -CH acyl aliphatic units by HSQC we can deduce that the aliphatic ester group at position **2** is a normal "nC10" ester. A second nC10 acyl group assigned to position **3** by HMBC correlations displayed ***α*-CH$_2$** and ***β*-CH$_2$** signals centered at 2.18 and 1.48 ppm, overlapping **-CH$_2$** (polymethylene) signals from both nC10 acyl groups in the range 1.2-1.4 ppm, and **-CH$_3$** triplet signal at ~0.9 ppm. Finally, aliphatic H's from C2 acyl groups displayed **-CH$_3$** singlet signals at ~2.0 ppm and were assigned to positions **1** and **4** by HMBC correlations consistent with those ring hydrogen positions. The structure of I:24:0(2,2,10,10) is summarized in Table 3.1. A similar approach was used for identifying ester branching and acylation position

assignments for each acylinositol identified in this study. However, four acylinositols identified nC10 and nC12 ester groups at positions **2** and **3** of *myo*-inositol that were not differentiated with this approach. Therefore, an alternative 1D-TOCSY NMR method was used to characterize those structures.

Positional assignment of acylinositols esterified with nC10 and nC12 ester groups by NMR presents a daunting analytical challenge, as the spectra of these groups are very similar and $^1$H resonances overlap. Figure 3.6A-B compares the $^1$H NMR spectra of I4:24:0(2,2,10,10) and I4:26:0(2,2,10,12) (Figure 3.1, $t_r$ = 80.27 min) acylinositols, with the only appreciable difference being the shape of a non-resolved polymethylene peak in the range of 1.2-1.4 ppm (due to a mixture of unresolved resonances from both the nC10 and nC12 ester groups). To circumvent this issue, a series of 1D-TOCSY transfer experiments were employed. Sachleben *et. al.* (2014) showed that a 1D-TOCSY technique could be used to identify aliphatic chain length by excitation and transfer magnetization from a resolved methylene resonance, and that the time-dependence of this transfer to another resolved site, such as a terminal methyl group, depends on the number of aliphatic carbons separating the two sites [24]. For this study, the resolved methylene hydrogen at positions **2-*β*-CH$_2$** (1.62 ppm) and **3-*β*-CH$_2$** (1.48 ppm) of acyl groups of I4:24:0(2,2,10,10) and I4:26:0(2,2,10,12) acylinositols (Figure 3.6A-B) were selectively excited at their respective methylene resonances (excitation bandwidth of ~53 Hz) and $^1$H NMR collected over a series of mixing times (0 ms, then every 20 ms for 100-300 ms). As the mixing time was increased, the transfer of magnetization to the terminal -**CH$_3$** unit with signal at ~0.9 ppm showed increasing intensity (Appendix Figures 3.92-3.95). The -**CH$_3$** signal (S) at each mixing time was integrated and normalized to the integral of the excitation peak at **2-*β*-CH$_2$** or **3-*β*-CH$_2$** at 0 ms mixing time (S$_o$).Thus, S$_o$ served as the reference for each of the series of TOCSY transfer experiments (Appendix Tables 3.25-3.28). Figure 3.6C-D illustrates the 1D-TOCSY transfer curves for each of the four experiments. Because I4:24:0(2,2,10,10) must have two nC10 ester groups (nC10 and nC12 assigned from HSQC results), it was used for comparison to I4:26:0(2,2,10,12). Figure 3.6C shows that the nC10 esters at positions **2** and **3** of have similar transfer curves. In contrast, acylinositol I4:26:0(2,2,10,12) in Figure 3.6D shows that the amount of magnetization transferred is less

for the acylation at position **2** as the mixing time increases. Thus, the nC12 ester groups was assigned to position **2**. Notably, we interpret the differences in $S/S_o$ to be a result of signal-to-noise when measuring different sample quantities (48.6 µM versus 2.5 µM). The remaining nC10 and nC12 containing acylinositols and acylinositol glycosides were assigned by inference from this group of measurements. Table 3.1 summarizes the structures of NMR resolved acylinositols purified from *S. quitoense* extracts (further evidence of these structural assignments is provided in Appendix Tables 3.14-3.28 and Appendix Figures 3.26-3.95).
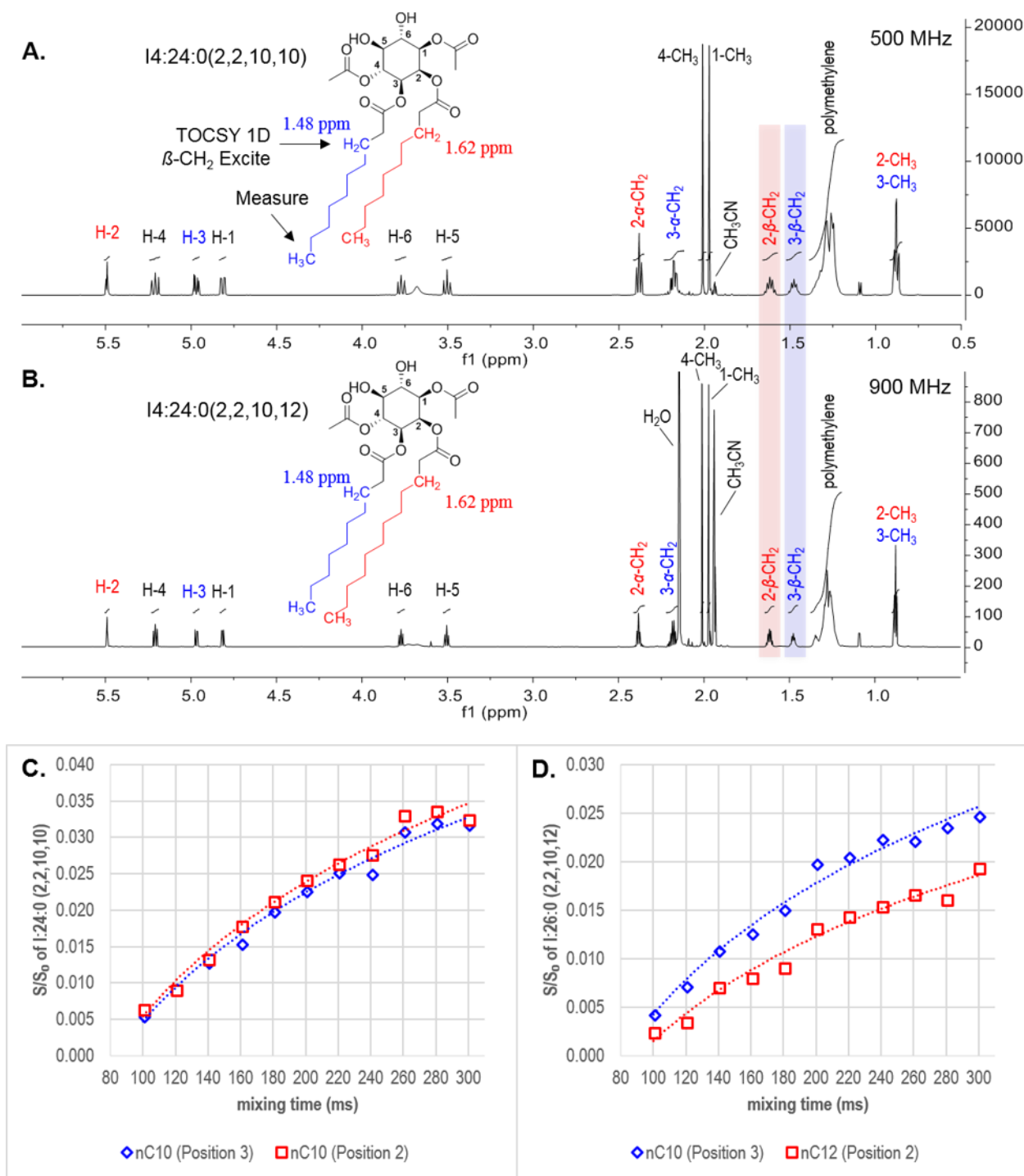
**Figure 3.6.** ^{1}H NMR spectra of **(A)** I4:24:0(2,2,10,10) (highlighted regions are $\beta$-CH$_2$ positions that were selectively excited for 1D-TOCSY transfer experiments, those spectra are located in Appendix Figures 3.92-3.95) and **(B)** I4:26:0(2,2,10,12) acylinositols. 1D-TOCSY transfer curves for excitation of 2-$\beta$-CH$_2$ and 3-$\beta$-CH$_2$ acyl groups of **(C)** I4:24:0(2,2,10,10) and **(D)** I4:26:0(2,2,10,12) acylinositols.

**Table 3.1.** Summary of NMR resolved acylinositols purified from *S. quitoense* extracts and percent peak area of ions by negative ion mode.

| Acylsugar ID | Retention time (min) | Experimental $m/z$[a] | Adducts Detected[b] | Analyte Molecular Formula | Mass Error (ppm) | R$_1$ | R$_2$ | R$_3$ | R$_4$ | Sample Peak Area ($\times 10^4$) | % of Total Acylsugar Peak Area[c] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Acyl *myo*-inositols | | | | | | | | | | | |
| I3:22:0(2,10,10) | 68.18 | 575.3435 | M+Cl, **M+HCOO**, M+NO$_3$ | C$_{28}$H$_{50}$O$_9$ | -0.82 | C2 | nC10 | nC10 | H | 4.42 | 6.3% |
| I4:24:0(2,2,10,10) | 73.90 | 617.3539 | M+Cl, **M+HCOO**, M+NO$_3$ | C$_{30}$H$_{52}$O$_{10}$ | -0.62 | C2 | nC10 | nC10 | C2 | 52.7 | 75.1% |
| I3:24:0(2,10,12) | 75.14 | 603.3745 | M+Cl, **M+HCOO** | C$_{30}$H$_{54}$O$_9$ | 0.71 | C2 | nC12 | nC10 | H | 1.07 | 1.5% |
| I4:26:0(2,2,10,12) | 80.27 | 645.3857 | M+Cl, **M+HCOO**, M+NO$_3$ | C$_{32}$H$_{56}$O$_{10}$ | 0.29 | C2 | nC12 | nC10 | C2 | 2.95 | 4.2% |
| Acyl *β-N*-acetylglucosaminyl *myo*-inositols | | | | | | | | | | | |
| NAG-I3:22:0(2,10,10) | 58.00 | 778.4241 | M-H, M+Cl, **M+HCOO**, M+NO$_3$ | C$_{36}$H$_{63}$NO$_{14}$ | 0.20 | C2 | nC10 | nC10 | NAG | 2.84 | 4.1% |
| NAG-I3:24:0(2,10,12) | 64.28 | 806.4550 | M-H, M+Cl, **M+HCOO** | C$_{38}$H$_{67}$NO$_{14}$ | 0.56 | C2 | nC12 | nC10 | NAG | 0.69 | 1.0% |
| Acyl *β*-glucopyranosyl *myo*-inositol | | | | | | | | | | | |
| G-I3:22:0(2,10,10) | 58.83 | 737.3965 | M-H, M+Cl, **M+HCOO**, M+NO$_3$ | C$_{34}$H$_{60}$O$_{14}$ | -1.07 | C2 | nC10 | nC10 | G | 2.77 | 3.9% |
| G-I3:24:0(2,10,12) | 65.37 | 765.4278 | **M-H, M+HCOO** | C$_{36}$H$_{64}$O$_{14}$ | -0.17 | C2 | nC12 | nC10 | G | 0.78 | 1.1% |
| *β*-xylopyranosyl *myo*-inositol | | | | | | | | | | | |
| X-I3:22:0(2,10,10) | 61.28 | 707.3864 | M-H, M+Cl, **M+HCOO**, M+NO$_3$ | C$_{33}$H$_{58}$O$_{13}$ | 0.69 | C2 | nC10 | nC10 | X | 0.68 | 1.0% |

% Acylinositol Peak Area with Structural Identification = **98.2%**

**a.** Monoisotopic *m/z* value with greatest ion abundance.
**b.** Bolded adduct ion with greatest ion abundance.
**c.** Sample peak area values were calculated by combining integrations of ions detected

### 3.3.3 Structural diversity of acylinositols from *S. quitoense*

Where *S. quitoense* was comparatively lacking in total number of acylsugars, it showed unexpected structural diversity regarding sugar core structures. The presence of acylated *myo*-inositols and their glycosides widens our understanding of acylsugar chemical diversity. To our knowledge, the only prior report of acylated *myo*-inositols within the family Solanaceae was in *Solanum lanceolatum* [25], which showed acylated glucopyranosyl and xylopyranosyl *myo*-inositols, named lanceolitols. In contrast to *S. quitoense,* reported lanceolitols were glycosylated at position 1 (or 3 depending on the nomenclature system that is used) of the *myo*-inositol ring and had only one acylation at position 2 (even numbered C12 to C20 branched or straight chain aliphatic ester groups). *S. quitoense* acylinositols showed limited structural variability in acylation patterns compared to other acylsugar analogues [2-4]. Purified acylinositols were esterified at position 4 by C2 groups or glycosylated by *N*-acetylglucosaminyl, glucopyranosyl or xylopyranosyl groups. Only C2 esters were attached at position 1, whereas nC10 or nC12 esters were observed at position 2 and nC10 at position 3.

### 3.4 Conclusions

*S. quitoense* had fewer abundant acylated SMs, but showed unexpected structural diversity with regard to carbohydrate chemistries. Previous reports have mostly identified glucose and acylsucrose, whereas a group of acylated *myo*-inositols and *myo*-inositol glycosides (*N*-acetylglucosaminyl, glucopyranosyl and xylopyranosyl) were discovered in *S. quitoense* (9 structures). Acylsugars containing two or more long-chain acyl groups (>8 carbons) were not previously detected during profiling of Solanaceae species. Likewise, nitrogen-containing acylsugars are telling of the surprising chemical diversity observed within this species. Using only LC/MS-based techniques, many *S. quitoense* acylinositols resemble previously discovered sucrose and glucose derivatives, while other acylsugars had ambiguous core structures that were inconsistent with molecular masses of glucose and sucrose. This result demonstrates the important marriage between mass spectrometry and NMR spectroscopy for structural characterization of unknown SMs, which is particularly important for distinguishing isomeric acylsugars.

269

Investigations of BAHD acylsugar acyltransferases have shown that glandular trichome-specific enzymes catalyze acylation across a host of available acyl-CoAs and sugar substrates. It is unlikely that *S. quitoense* does not produce sucrose. Rather, one hypothesis suggests that sucrose is not readily available in the same cellular or subcellular compartments as Acyl-CoAs, inositols or glycoinositols, and acyltransferase enzymes. Therefore, acyltransferase enzymes have evolved to synthesize acylsugars via other methods. For acyltransferases, small changes in amino acid sequences can alter acyl acceptor substrate preferences [1, 6, 8]. A more likely scenario is that acyltransferase enzymes in *S. quitoense* have different substrate specificities and cannot acylate sucrose. The limited number of acylinositols from *S. quitoense* and the apparent limited promiscuity of acylation positions suggest that *S. quitoense* acyltransferase enzymes are more selective for acyl group acceptor substrates. Exploration of candidate biosynthetic genes operating within *S. quitoense* will be needed and are under way to ascertain this hypothesis.

**APPENDIX**
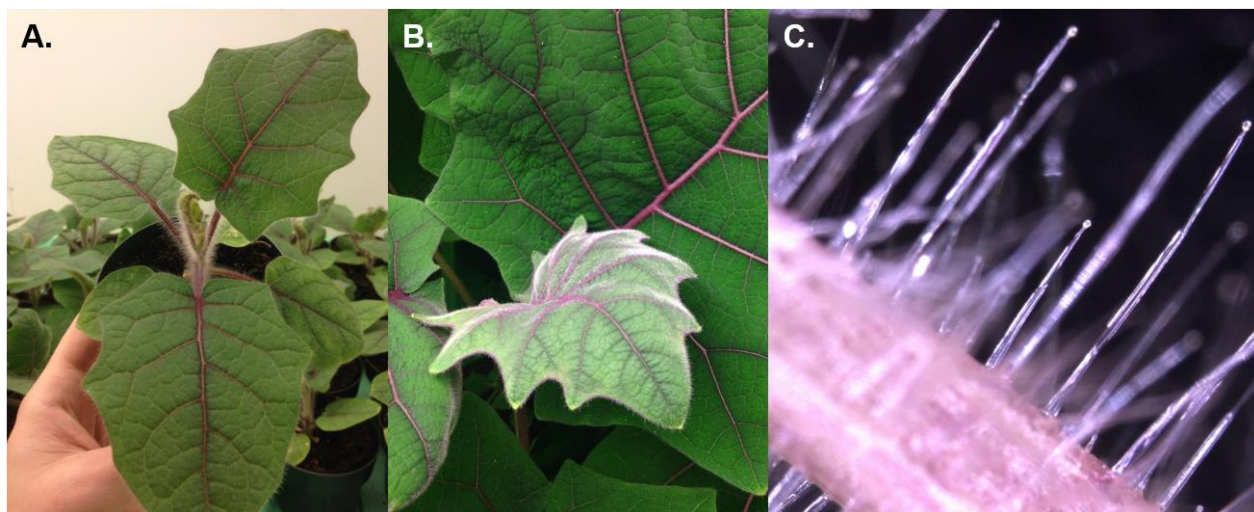
**Figure 3.7.** *S. quitoense* images. **(A)** Picture at ~5 weeks **(B)** Young leaflet. **(C)** Type I-like trichomes [26, 27] on petiole of a young leaflet (approximately 3-6 mm in length).

**Table 3.2.** Plant cultivation and metabolite extraction metadata

| Species | *Solanum quitoense* |
|---|---|
| Genotype | NYBG |
| Organ | Aerial tissues |
| Organ specification | Aerial tissues included leaf and stem tissues |
| Cell type | Glandular trichomes |
| Growth location | Growth chamber at MSU |
| Growth support | Seeds were germinated using MilliQ deionized water on Whatman #1 filter paper placed in petri dishes. One-week old seedlings were transferred to Jiffy peat pellets. |
| Light | 300 µE $m^{-2}$ sec $^{-1}$; 17 h light/7 h dark |
| Humidity | 50% relative humidity |
| Temperature | 25˚C day/12˚C night |
| Watering regime | Bottom watering as per requirement |
| Nutritional regime | Half strength Hoagland's solution, once a week |
| Date(s) of plant establishment | 5/28/2015 |
| Plant growth stage | 10 weeks post germination (plants about 0.25 m height), plants extracted on 8/6/2015 |
| Metabolism quenching method | Aerial tissues were extracted in 1.9 L of acetonitrile: isopropanol (AcN:IPA, v/v, 1:1) for 10 mins in two 1-L Erlenmeyer flasks with horizontal mixing at 120 rpm |
| Harvest method | Plants were cut at the stems; larger leaflets were cut at nodes to reduce bulk volume. |
| Sample storage | The extract was decanted into two 1 L glass Wheaton bottles with Teflon lined caps and stored in a freezer at     -20°C. ~1 L of solution was concentrated to dryness via rotary evaporation under vacuum in a 1 L round bottom flask (4 × 250 mL), leaving a dark green residue. The residue was dissolved in 5 mL of AcN:IPA with sonication, transferred to 15 mL polypropylene centrifuge tubes, and centrifuged by Eppendorf Centrifuge 5480R at 10000 × g for 10 mins, the supernatant was transferred to 1 mL autosampler vials and stored at -20°C prior to purification by semi-preparative HPLC. |

**Table 3.3.** UHPLC/MS metadata

| | |
|---|---|
| Facility Director | Dr. A. Daniel Jones |
| Analyst | Steven M. Hurney |
| Instrument Location | A. Daniel Jones Laboratory |
| Facility Instrument Title | LCMS: Jones Lab G2-XS QTof |
| LC System | Shimadzu LC-20AD pumps and Shimadzu CTO-20A column oven. |
| Manufacturers | Shimadzu / Waters |
| Autosampler | Shimadzu SIL-5000 autosampler |
| Column | Ascentis Express C18 Analytical HPLC, 10 cm x 2.1mm x 2.7 µm |
| Column Manufacturer | Supelco |
| Catalogue Number | 53823-U |
| Serial Number | USRB004698 |
| Packing Lot Number | S17138 |
| Injection Volume | 10 µL (5 µL sample loop) |
| Flow Rate | 0.3 mL/min |
| Mobile Phases: | |
| A | 10 mM ammonium formate in water (pH 2.8, adjusted with formic acid) |
| B | Acetonitrile |
| Gradient Profile | Hold 1% B at 0-1 min, linear 1-100% B at 1-100 min, hold 100% B at 100-105 min, linear 100-1% B at 105-106 min, and hold 1% B at 106-110 min |
| Column Oven Temperature | 50 ℃ |
| Sample Temperature in autosampler | Room Temperature |
| Inlet Method Name | SMH_110min_C18_1to100_linear_1to100mins_AcN_Pump C_50C |
| Mass Spectrometer | Xevo G2-XS QTof |
| Manufacturer | Waters |
| Software | MassLynx v4.1 |
| Ionization Source | Electrospray Ionization (ESI) |
| Data Acquisition | Sensitivity Mode |
| Polarity | Positive, Negative |
| Mass Range | $m/z$ 50-1500 |
| Data Format | Continuum |
| Capillary Voltage | 3.06 kV (ESI+), 2.5 kV (ESI-) |
| Sample Cone | 36 V (ESI+), 40 V (ESI-) |
| Source Temperature | 100 °C |
| Source Offset | 80 V |
| Desolvation Temperature | 300 °C |
| Cone Gas Flow | 50 L/hr |

Table 3.3 (continued)

| Desolvation Gas Flow | 600 L/hr |
|---|---|
| Collision Potential | |
| Function 1 | 6.0 V |
| Function 2 | 10 to 60 V |
| Scan Duration | 0.20 s |
| Inter Scan Delay | 0.014 s |
| Collision Cell Pressure | 0.085 mbar (ESI+), 0.083 mbar (ESI-) |
| Lock Spray | Leu-enkephalin |
| Lock mass (*m/z*) | 556.2771 (ESI+), 554.2615 (ESI-) |
| Lock Spray Scan Time | 0.2 s |
| Lock Spray Scan Frequency | 10 s |
| MS Method Files | SMH_MSe_110 min_10to60V_ST02s_ESI+, SMH_MSe_110 min_10to60V_ST02s_ESI- |
| Sample handling | Aerial tissues from 20 plants aged 10 weeks were harvested (plants were cut at the stems and stem junctions) and extracted in 1.9 L of AcN:IPA for 10 mins in two 1 L Erlenmeyer flasks with horizontal mixing at 120 rpm. For metabolite profiling, the bulk extract solution was diluted to 1/10th concentration, by adding 100 µL to 800 µL of AcN:IPA (containing 0.1 µM telmisartan internal standard) and adding 100 µL of 0.046 M formic acid (pH ~4.0, inhibits of acylinositol rearrangement) |
| Sample Storage | -20 °C in Wheaton glass vessel with PTFE lined cap |
| Protocol when analyzing the samples | The instrument was calibrated in ESI+/- modes using 500 µM sodium formate solution. First, the column was equilibrated by analyzing two blanks using a 22 min gradient. A blank and *S. quitoense* samples were analyzed in ESI- mode, followed by blank and sample in ESI+ mode. |

**Table 3.4.** Progenesis QI metadata

| Peak Picking Parameters: | |
|---|---|
| Sensitivity | Automatic, Value = 5 |
| Ignore Ions after | 90 mins |
| Adducts | ESI(+): M+H-H$_2$O, M+H, M+NH$_4$, M+Na<br>ESI (-): M-H, M+Cl, M+HCOO, M+NO$_3$ |
| Elemental Composition Calculation Parameters: | |
| Composition | Hydrogen (10-100), Carbon (6-100), Nitrogen (0-1), Oxygen (7-30) |
| Precursor tolerance | 20 ppm |
| Isotope similarity | 80% |
| Tag filters (used to exclude markers): | |
| Chromatographic Peak Width | < 0.05 min |
| Sample > Blank | Integrals of sample peak must be > 3x blank |
| *m/z* range | ESI(+): 240 < *m/z* < 850<br>ESI(-): 267 < *m/z* < 877 |
| No Elemental Composition | Peaks that could not be assigned a molecular formula using the designated elemental composition calculation parameters were excluded |
| Assessment of acylinositol peak annotations | To further assess peak annotations, the *m/z* values of acylinositol homologues were calculated. For example, *myo*-inositol (C$_6$H$_{12}$O$_6$, 180.0634 Da) with C2 acyl group (C$_8$H$_{14}$O$_7$, 222.0740 Da) would have [M+NH$_4$]$^+$ = 240.1078 and [M+HCOO]$^-$ = 267.0722. The homologues of the following acylinositols and their adducts were considered: 1) *myo*-inositol with C2 (as indicated); 2) *N*-acetylglucosaminyl myo-inositol with C2 (C$_{16}$H$_{27}$NO$_{12}$, [M+H]$^+$ = 426.1606 and [M+HCOO]$^-$ = 425.1533); 3) glucopyranosyl *myo*-inositol with C2 (C$_{14}$H$_{24}$O$_{12}$, [M+NH$_4$]$^+$ = 402.1606 and [M+HCOO]$^-$ = 429.1250); and 4) xylopyranosyl *myo*-inositol with C2 (C$_{13}$H$_{22}$O$_{11}$, [M+NH$_4$]$^+$ = 372.1500 and [M+HCOO]$^-$ = 399.1144). Annotations that could not be assigned as homologues with *m/z* values of +14.0456 Da (consistent with -CH$_2$- saturated acyl groups) were excluded. Annotations not consistent with homologues of multiply esterified acylinositols were also excluded. For example, diacylated *myo*-inositol with two C2 groups would have formula C$_{10}$H$_{16}$O$_8$ and [M+NH$_4$]$^+$ = 282.1183 and [M+HCOO]$^-$ = 309.0827). Permutations of *m/z* values for all acylinositol and acylinositol glycosides with one to fully esterified hydroxyl groups were considered. Those that did not have *m/z* values consistent with these permutations were excluded. The results are shown in Appendix Tables 3.5 and 3.6. |

**Table 3.5.** *S. quitoense* acylinositol deep profiling results generated by ESI- mode using Waters Progenesis QI software.

| Peak #[a] | Metabolite Annotation[b] | Retention time (min) | m/z[c] | Adducts Detected[d] | Formula | Mass Error (ppm) | Isotope Similarity (%) | Sample Peak Area[e] | Relative Abundance |
|---|---|---|---|---|---|---|---|---|---|
| 1 | I2:12:0 | 35.06 | 421.2070 | **M+HCOO** | $C_{18}H_{32}O_8$ | -2.47 | 93.74 | 353 | 0.05% |
| 2 | G-I3:16:0 | 39.37 | 653.3015 | **M+HCOO** | $C_{28}H_{48}O_{14}$ | -1.85 | 90.48 | 134 | 0.02% |
| 3 | G-I4:18:0 | 41.72 | 695.3129 | **M+HCOO** | $C_{30}H_{50}O_{15}$ | -0.39 | 90.11 | 231 | 0.03% |
| 5* | **NAG-I3:22:0(2,10,10)** | 58.00 | 778.4241 | M-H, M+Cl, **M+HCOO**, M+NO₃ | $C_{36}H_{63}NO_{14}$ | 0.20 | 99.16 | 28418 | 4.05% |
| 6* | **G-I3:22:0(2,10,10)** | 58.83 | 737.3965 | M-H, M+Cl, **M+HCOO**, M+NO₃ | $C_{34}H_{60}O_{14}$ | -1.07 | 99.13 | 27655 | 3.94% |
| 8* | **X-I3:22:0(2,10,10)** | 61.28 | 707.3864 | M-H, M+Cl, **M+HCOO**, M+NO₃ | $C_{33}H_{58}O_{13}$ | 0.69 | 97.99 | 6823 | 0.97% |
| 11* | **NAG-I3:24:0(2,10,12)** | 64.28 | 806.4550 | M-H, M+Cl, **M+HCOO** | $C_{38}H_{67}NO_{14}$ | 0.56 | 98.01 | 6909 | 0.99% |
| 12* | **G-I3:24:0(2,10,12)** | 65.37 | 765.4278 | M-H, **M+HCOO** | $C_{36}H_{64}O_{14}$ | -0.17 | 98.86 | 7774 | 1.11% |
| 13* | X-I4:24:0 | 66.78 | 749.3986 | **M+HCOO** | $C_{35}H_{60}O_{14}$ | 3.01 | 89.59 | 171 | 0.02% |
| 14* | I4:22:0 | 67.13 | 589.3220 | **M+HCOO** | $C_{28}H_{48}O_{10}$ | -1.69 | 94.79 | 1510 | 0.22% |
| 15 | I3:22:0 | 67.25 | 575.3429 | **M+HCOO** | $C_{28}H_{50}O_9$ | -1.41 | 98.21 | 2141 | 0.31% |
| 16* | X-I3:24:0 | 68.07 | 735.4174 | **M+HCOO** | $C_{35}H_{62}O_{13}$ | 0.27 | 96.13 | 1801 | 0.26% |
| 17* | **I3:22:0(2,10,10)** | 68.18 | 575.3435 | M+Cl, **M+HCOO**, M+NO₃ | $C_{28}H_{50}O_9$ | -0.82 | 98.62 | 44175 | 6.30% |
| 18* | I4:23:0 | 70.58 | 603.3375 | **M+HCOO** | $C_{29}H_{50}O_{10}$ | -1.93 | 92.22 | 426 | 0.06% |
| 19* | NAG-I3:26:0 | 70.89 | 834.4857 | **M+HCOO** | $C_{40}H_{71}NO_{14}$ | 0.03 | 89.12 | 491 | 0.07% |
| 20* | I4:24:0 | 72.00 | 617.3534 | **M+HCOO** | $C_{30}H_{52}O_{10}$ | -1.42 | 92.50 | 636 | 0.09% |
| 21* | G-I3:26:0 | 72.18 | 793.4592 | **M+HCOO** | $C_{38}H_{68}O_{14}$ | 0.14 | 89.87 | 632 | 0.09% |
| 22 | I4:24:0 | 73.23 | 617.3538 | **M+HCOO** | $C_{30}H_{52}O_{10}$ | -0.78 | 88.29 | 192 | 0.03% |
| 24* | **I4:24:0(2,2,10,10)** | 73.90 | 617.3539 | M+Cl, **M+HCOO**, M+NO₃ | $C_{30}H_{52}O_{10}$ | -0.62 | 97.25 | 526844 | 75.1% |
| 26 | X-I3:26:0 | 74.94 | 763.4491 | **M+HCOO** | $C_{37}H_{66}O_{13}$ | 0.74 | 82.12 | 56 | 0.01% |
| 27* | **I3:24:0(2,10,12)** | 75.14 | 603.3745 | M+Cl, **M+HCOO** | $C_{30}H_{54}O_9$ | 0.71 | 98.69 | 10658 | 1.52% |
| 28* | I4:25:0 | 77.16 | 631.3700 | **M+HCOO** | $C_{31}H_{54}O_{10}$ | 0.25 | 90.95 | 402 | 0.06% |
| 31* | I5:26:0 | 79.56 | 659.3603 | **M+HCOO** | $C_{32}H_{54}O_{11}$ | -7.35 | 92.00 | 160 | 0.02% |
| 32* | **I4:26:0(2,2,10,12)** | 80.27 | 645.3857 | M+Cl, **M+HCOO**, M+NO₃ | $C_{32}H_{56}O_{10}$ | 0.29 | 98.52 | 29461 | 4.20% |

**Table 3.5.** (continued)

| Peak #[a] | Metabolite Annotation[b] | Retention time (min) | m/z[c] | Adducts Detected[d] | Formula | Mass Error (ppm) | Isotope Similarity (%) | Sample Peak Area[e] | Relative Abundance |
|---|---|---|---|---|---|---|---|---|---|
| 34 | I3:26:0 | 81.73 | 631.4050 | **M+HCOO** | $C_{32}H_{58}O_9$ | -2.26 | 93.38 | 1031 | 0.15% |
| 36* | I4:28:0 | 86.13 | 673.4126 | **M+HCOO** | $C_{34}H_{60}O_{10}$ | -6.72 | 92.28 | 2105 | 0.30% |

**a.** Peak number indicates elution order of *S. quitoense* acylinositol annotation. Peak numbers with asterisks (*) were annotated in both ESI- and ESI+ ion modes. Acronyms: *myo*-inositol (**I**), glucopyranosyl (**G**), *N*-acetylglucosaminyl (**NAG**) and xylopyranosyl (**X**).

**b.** Metabolite annotations in bold were identified by NMR analysis.

**c.** Monoisotopic *m/z* value with greatest ion abundance.

**d.** Bolded adduct ion had greatest ion abundance.

**e.** Sample peak area values were calculated by combining ions detected.

**Table 3.6.** *S. quitoense* acylinositol deep profiling results generated by ESI+ mode using Waters Progenesis QI software.

| Peak #[a] | Metabolite Annotation[b] | Retention time (min) | m/z[c] | Adducts Detected[d] | Formula | Mass Error (ppm) | Isotope Similarity (%) | Sample Peak Area[e] | Relative Abundance |
|---|---|---|---|---|---|---|---|---|---|
| 4 | I5:22:0 | 54.92 | 576.3355 | **M+NH₄,** M+Na | $C_{28}H_{46}O_{11}$ | -4.10 | 94.05 | 966 | 0.09% |
| 5* | **NAG-I3:22:0(2,10,10)** | 57.98 | 734.4296 | **M+H**, M+Na | $C_{36}H_{63}NO_{14}$ | -3.48 | 98.44 | 17785 | 1.73% |
| 6* | **G-I3:22:0(2,10,10)** | 58.82 | 710.4291 | **M+NH₄**, M+Na | $C_{34}H_{60}O_{14}$ | -4.33 | 99.47 | 21817 | 2.13% |
| 7 | I5:20:0 | 59.49 | 548.3062 | M+H, **M+NH₄**, M+Na | $C_{26}H_{42}O_{11}$ | 2.65 | 93.16 | 2976 | 0.29% |
| 8* | **X-I3:22:0(2,10,10)** | 61.25 | 680.4176 | **M+NH₄**, M+Na | $C_{33}H_{58}O_{13}$ | -5.95 | 97.78 | 5908 | 0.58% |
| 9 | X-I3:22:0 | 61.98 | 680.4191 | **M+NH₄** | $C_{33}H_{58}O_{13}$ | -3.65 | 91.59 | 228 | 0.02% |
| 10 | I5:20:0 | 62.23 | 562.3216 | **M+NH₄** | $C_{27}H_{44}O_{11}$ | -1.17 | 92.10 | 215 | 0.02% |
| 11* | **NAG-I3:24:0(2,10,12)** | 64.27 | 762.4595 | **M+H** | $C_{38}H_{67}NO_{14}$ | -5.22 | 97.05 | 4279 | 0.42% |
| 12* | **G-I3:24:0(2,10,12)** | 65.37 | 738.4597 | **M+NH₄**, M+Na | $C_{36}H_{64}O_{14}$ | -5.20 | 97.53 | 6441 | 0.63% |
| 13* | X-I4:24:0 | 66.76 | 722.4296 | **M+NH₄** | $C_{35}H_{60}O_{14}$ | -3.62 | 88.63 | 223 | 0.02% |
| 14* | I4:22:0 | 67.14 | 562.3550 | **M+NH₄**, M+Na | $C_{28}H_{48}O_{10}$ | -6.54 | 91.99 | 3471 | 0.34% |
| 16* | X-I3:24:0 | 68.08 | 708.4477 | **M+NH₄** | $C_{35}H_{62}O_{13}$ | -7.52 | 95.98 | 1823 | 0.18% |
| 17* | **I3:22:0(2,10,10)** | 68.16 | 548.3777 | M+H-H₂O, M+H, **M+NH₄**, M+Na | $C_{28}H_{50}O_9$ | -7.45 | 97.64 | 43496 | 4.24% |
| 18* | I4:23:0 | 70.58 | 576.3701 | **M+NH₄**, M+Na | $C_{29}H_{50}O_{10}$ | -7.46 | 88.66 | 1272 | 0.12% |
| 19* | NAG-I3:26:0 | 70.85 | 790.4894 | M+H | $C_{40}H_{71}NO_{14}$ | -6.69 | 94.31 | 425 | 0.04% |
| 20* | I4:24:0 | 71.96 | 595.3428 | M+H-H₂O, **M+NH₄**, M+Na | $C_{30}H_{52}O_{10}$ | -12.63 | 88.37 | 6854 | 0.67% |
| 21* | G-I3:26:0 | 72.11 | 766.4872 | **M+NH₄** | $C_{38}H_{68}O_{14}$ | -10.06 | 92.91 | 484 | 0.05% |
| 23 | I4:25:0 | 73.85 | 604.3987 | **M+NH₄** | $C_{31}H_{54}O_{10}$ | -11.70 | 82.08 | 2692 | 0.26% |
| 24* | **I4:24:0(2,2,10,10)** | 73.89 | 590.3892 | M+H-H₂O, M+H, **M+NH₄**, M+Na | $C_{30}H_{52}O_{10}$ | -2.92 | 95.16 | 799389 | 77.9% |
| 25 | I4:26:0 | 73.89 | 618.4181 | **M+NH₄** | $C_{32}H_{56}O_{10}$ | -5.09 | 95.75 | 12437 | 1.21% |
| 27* | **I3:24:0(2,10,12)** | 75.13 | 576.4077 | M+H-H2O, **M+NH₄**, M+Na | $C_{30}H_{54}O_9$ | -12.80 | 91.61 | 11086 | 1.08% |
| 28* | I4:25:0 | 77.08 | 604.3985 | **M+NH₄** | $C_{31}H_{54}O_{10}$ | -11.94 | 93.16 | 818 | 0.08% |
| 29 | I5:26:0 | 78.24 | 632.3978 | M+H-H₂O, M+H, **M+NH₄**, M+Na | $C_{32}H_{54}O_{11}$ | -4.32 | 94.93 | 18652 | 1.82% |
| 30 | I5:28:0 | 78.26 | 660.4273 | **M+NH₄** | $C_{34}H_{58}O_{11}$ | -6.91 | 80.54 | 930 | 0.09% |
| 31* | I5:26:0 | 79.51 | 632.3971 | M+H-H₂O, **M+NH₄**, M+Na | $C_{32}H_{54}O_{11}$ | -5.49 | 84.09 | 6393 | 0.62% |

**Table 3.6.** (continued)

| Peak #[a] | Metabolite Annotation[b] | Retention time (min) | m/z[c] | Adducts Detected[d] | Formula | Mass Error (ppm) | Isotope Similarity (%) | Sample Peak Area[e] | Relative Abundance |
|---|---|---|---|---|---|---|---|---|---|
| 32* | **I4:26:0(2,2,10,12)** | 80.26 | 618.4181 | M+H-H$_2$O, M+H, **M+NH$_4$**, M+Na | C$_{32}$H$_{56}$O$_{10}$ | -5.10 | 96.20 | 46869 | 4.57% |
| 33 | I4:28:0 | 80.29 | 646.4453 | **M+NH$_4$** | C$_{34}$H$_{60}$O$_{10}$ | -11.44 | 82.95 | 968 | 0.09% |
| 35 | I6:28:0 | 83.68 | 674.4079 | **M+NH$_4$**, M+Na | C$_{34}$H$_{56}$O$_{12}$ | -4.80 | 90.70 | 4490 | 0.44% |
| 36* | I4:28:0 | 86.07 | 646.4481 | **M+NH$_4$** | C$_{34}$H$_{60}$O$_{10}$ | -7.02 | 86.52 | 3192 | 0.31% |

**a.** Peak number indicates elution order of *S. quitoense* acylinositol annotation. Peak numbers with asterisks (*) were annotated in both ESI- and ESI+ ion modes. Acronyms: *myo*-inositol (**I**), glucopyranosyl (**G**), *N*-acetylglucosaminyl (**NAG**) and xylopyranosyl (**X**).

**b.** Metabolite annotations in bold were identified by NMR analysis.

**c.** Monoisotopic *m/z* value with greatest ion abundance.

**d.** Bolded adduct ion had greatest ion abundance  .

**e.** Sample peak area values were calculated by combining ions detected.

**Table 3.7.** MS/MS precursor masses and time windows

| Acylinositol ID | [M+HCOO]⁻ | [M+NH₄]⁺ | Start (min) | Stop (min) | Cone (V) |
|---|---|---|---|---|---|
| NAG-I3:22:0(2,10,10) | 778.423 | 751.459 | 55.00 | 57.00 | 40 |
| G-I3:22:0(2,10,10) | 737.397 | 710.432 | 57.01 | 58.50 | 40 |
| X-I3:22:0(2,10,10) | 707.386 | 680.422 | 58.51 | 61.00 | 40 |
| NAG-I3:24:0(2,10,12) | 806.454 | 779.490 | 61.01 | 63.25 | 40 |
| G-I3:24:0(2,10,12) | 765.428 | 738.463 | 63.26 | 65.00 | 40 |
| I3:22:0(2,10,10) | 575.344 | 548.379 | 65.01 | 68.00 | 40 |
| I4:24:0(2,2,10,10) | 617.354 | 590.390 | 70.00 | 73.00 | 40 |
| I3:24:0(2,10,12) | 603.375 | 576.411 | 73.01 | 75.00 | 40 |
| I4:26:0(2,2,10,12) | 645.386 | 618.421 | 78.00 | 80.00 | 40 |

**Table 3.8.** MS/MS precursor masses, time windows and cone voltages

| Acylinositol ID | [M-H]⁻ | [M+H]⁺ | Start (min) | Stop (min) | [M-H]⁻ Cone (V) | [M+H]⁺ Cone (V) |
|---|---|---|---|---|---|---|
| NAG-I3:22:0(2,10,10) | 732.418 | 734.432 | 55.00 | 57.00 | 120 | 40 |
| G-I3:22:0(2,10,10) | 691.391 | 693.406 | 57.01 | 58.50 | 120 | 80 |
| X-I3:22:0(2,10,10) | 661.380 | 663.395 | 58.51 | 61.00 | 120 | 80 |
| NAG-I3:24:0(2,10,12) | 760.449 | 762.464 | 61.01 | 63.25 | 140 | 60 |
| G-I3:24:0(2,10,12) | 719.422 | 721.437 | 63.26 | 65.00 | 120 | 80 |
| I3:22:0(2,10,10) | 529.338 | 531.353 | 65.01 | 68.00 | 120 | 120 |
| I4:24:0(2,2,10,10) | 571.349 | 573.363 | 70.00 | 73.00 | 100 | 120 |
| I3:24:0(2,10,12) | 557.370 | 559.384 | 73.01 | 75.00 | 100 | 120 |
| I4:26:0(2,2,10,12) | 599.380 | 601.395 | 78.00 | 80.00 | 100 | 120 |

**Table 3.9.** LC/MS/MS metadata

| | |
|---|---|
| Facility Director | Dr. A. Daniel Jones |
| Analyst | Steven M. Hurney |
| Instrument Location | A. Daniel Jones Laboratory |
| Facility Instrument Title | LCMS: Jones Lab G2-XS QTof |
| LC System | Shimadzu LC-20AD pumps and Shimadzu CTO-20A column oven. |
| Manufacturers | Shimadzu / Waters |
| Autosampler | Shimadzu SIL-5000 autosampler |
| Column | Ascentis Express C18 Analytical HPLC, 10 cm x 2.1mm x 2.7 µm |
| Column Manufacturer | Supelco |
| Catalogue Number | 53823-U |
| Serial Number | USRB003996 |
| Packing Lot Number | S15013 |
| Injection Volume | 10 µL |
| Flow Rate | 0.3 mL/min |
| Mobile Phases: | |
| A | 10 mM ammonium formate in water (pH 2.8, adjusted with formic acid) |
| B | Acetonitrile |
| Gradient Profile | Hold 1% B at 0-1 min, 1-100% B at 1-100 min, 100% B at 100-105 min, 100-1% B at 105-106 min, and 1% B at 106-110 min. |
| Column Oven Temperature | 50 ℃ |
| Sample Temperature in autosampler | Room Temperature |
| Inlet Method Name | SMH_110min_C18_1to100_linear_1to100mins_Pump C |
| Mass Spectrometer | Xevo G2-XS QTof |
| Manufacturer | Waters |
| Software | MassLynx v4.1 |
| Ionization Source | Electrospray Ionization (ESI) |
| Data Acquisition | Sensitivity Mode |
| Polarity | Positive, Negative |
| Mass Range | $m/z$ 50-1000 |
| Data Format | Centroid |
| Capillary Voltage | 3 kV (ESI+), 3 kV (ESI-) |
| Sample Cone | See Appendix Tables 7 and 8 |
| Source Temperature | 100 °C |
| Source Offset | 80 V |
| Desolvation Temperature | 350 °C |
| Cone Gas Flow | 50 L/hr (ESI-), 20 L/hr (ESI+) |
| Desolvation Gas Flow | 600 L/hr (ESI-), 500 L/hr (ESI+) |
| LM Resolution | 20 |
| HM Resolution | 20 |

**Table 3.9.** (continued)

| Collision Potential Ramp | |
|---|---|
| Start Potential | 5 V |
| End Potential | 60 V |
| Scan Duration | 0.5 s |
| Inter Scan Delay | 0.014 s |
| Collision Cell Pressure | 0.1 mbar |
| MS Method Files | SMH_MSMS_5-60V_S quitoense_ESI-, SMH_MSMS_5-60V_S quitoense_ESI+, SMH_MSMS_5-60V_S quitoense_ESI-_M-H, SMH_MSMS_5-60V_S quitoense_ESI+_M+H, |
| Sample handling | *S. quitoense* bulk extract solution as described in Appendix Table 2 |
| Sample Storage | -20 °C in Wheaton vessel with PTFE lined cap |
| Protocol when analyzing the samples | The instrument was calibrated in ESI+/- modes using 500 µM sodium formate solution for all analyses. After column equilibration, an *S. quitoense* bulk extract sample was analyzed by MS/MS using precursor adducts [M+HCOO]$^-$ and [M+NH$_4$]$^+$ according to Appendix Table 7. An *S. quitoense* bulk extract sample was analyzed in CID mode with increasing cone potentials (40, 60, 80, 100, 120, and 140 V). Cone voltages shown in Appendix Table 8 were selected according to maximum [M-H]$^-$ and [M+H]$^+$ ion formation. MS/MS product ion spectra were generated using a collision potential ramp, targeting the respective precursor pseudomolecular ion. |
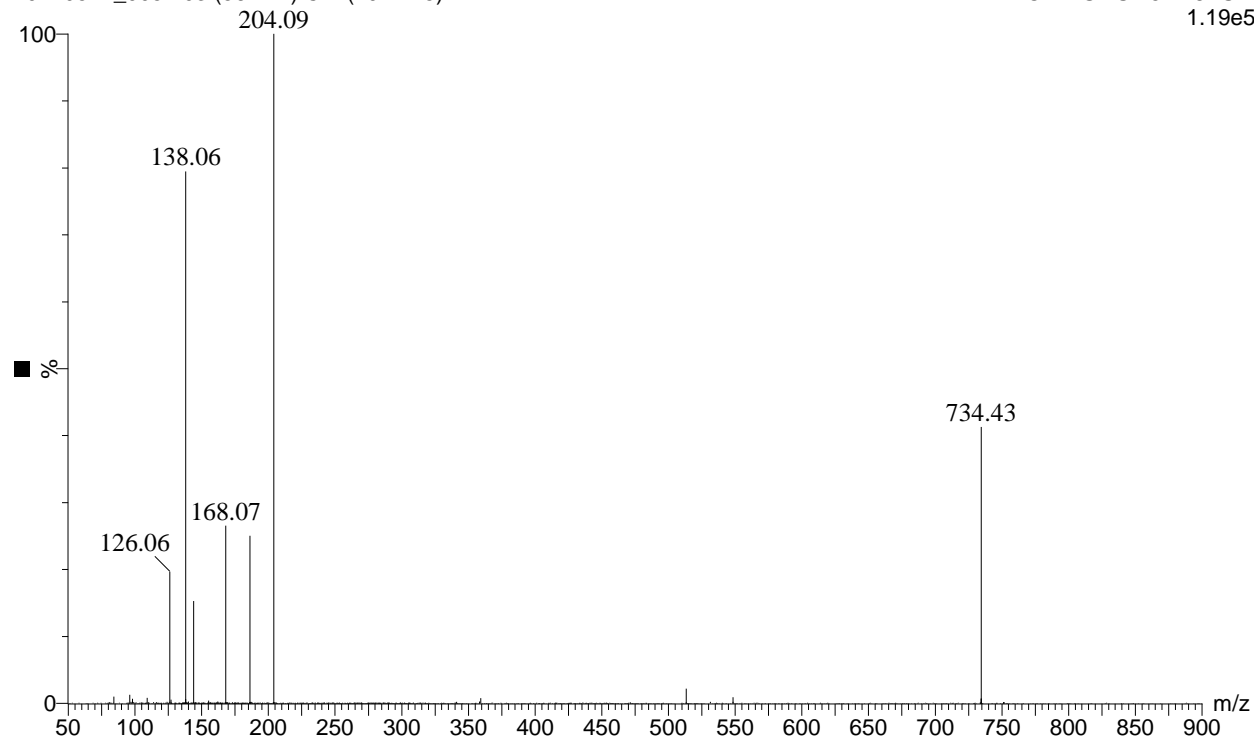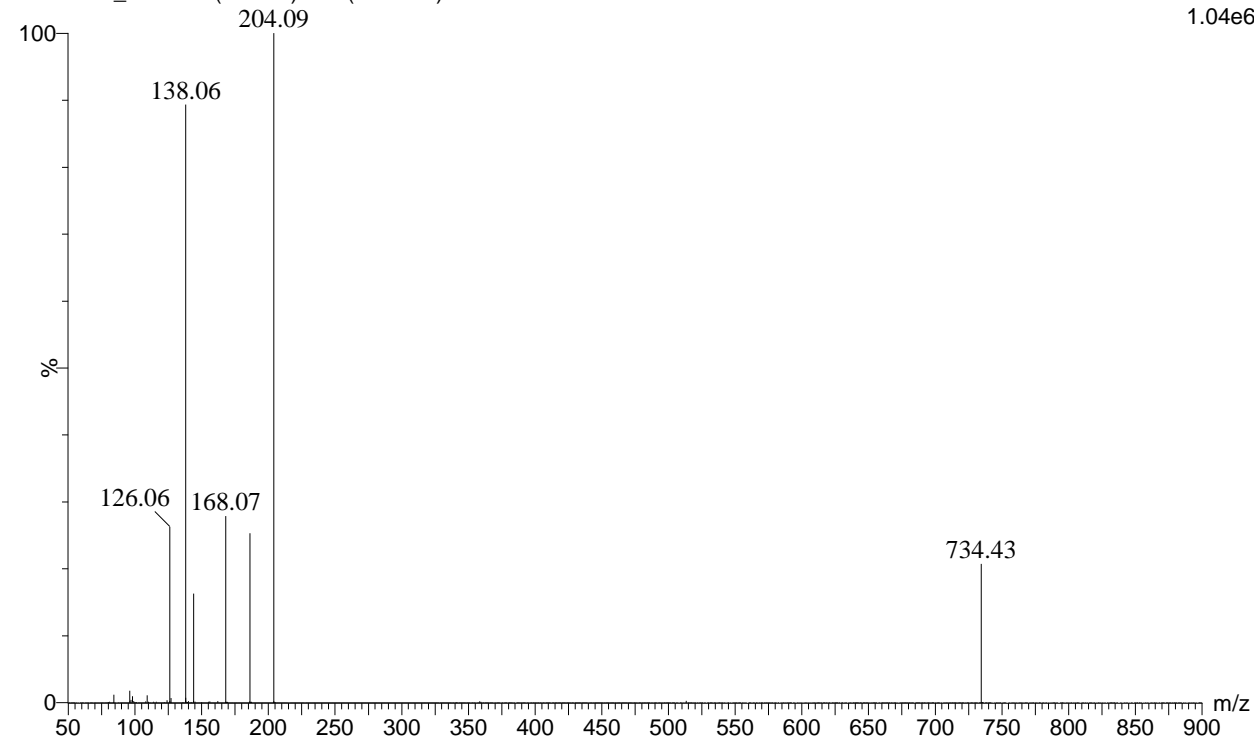
**Figure 3.8.** NAG-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)
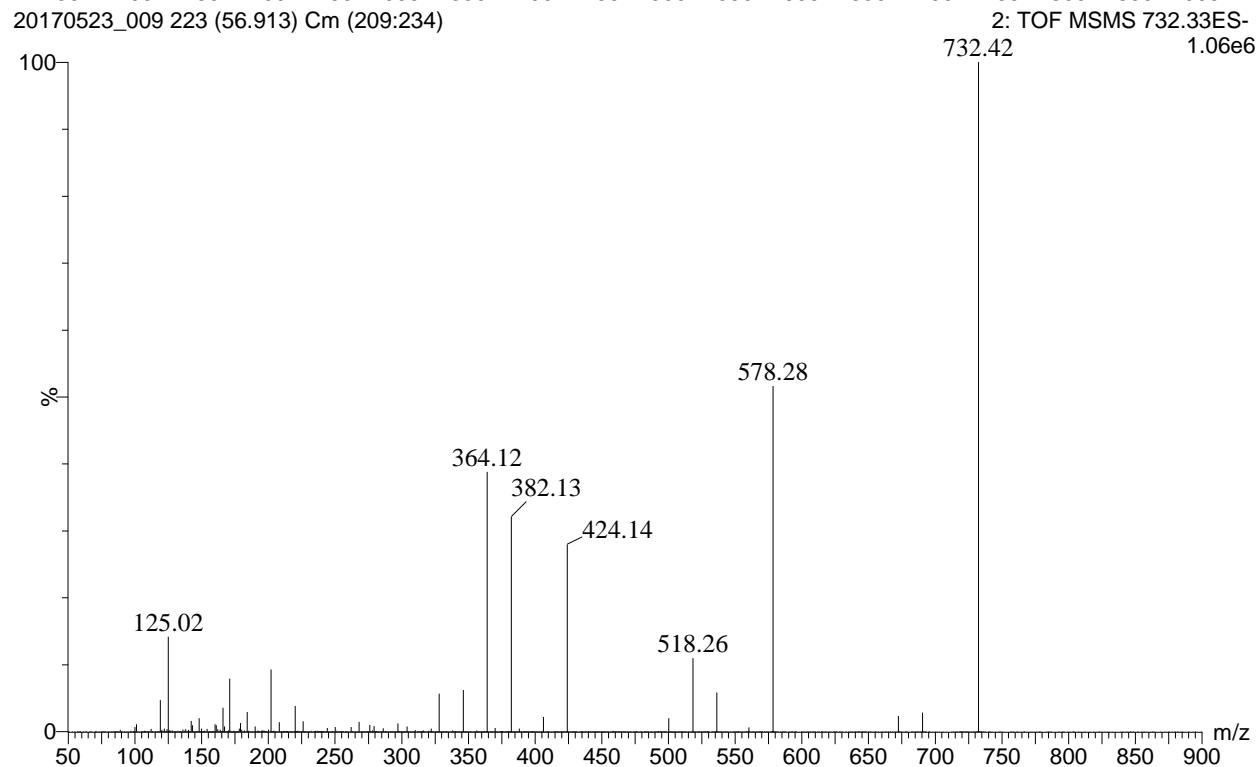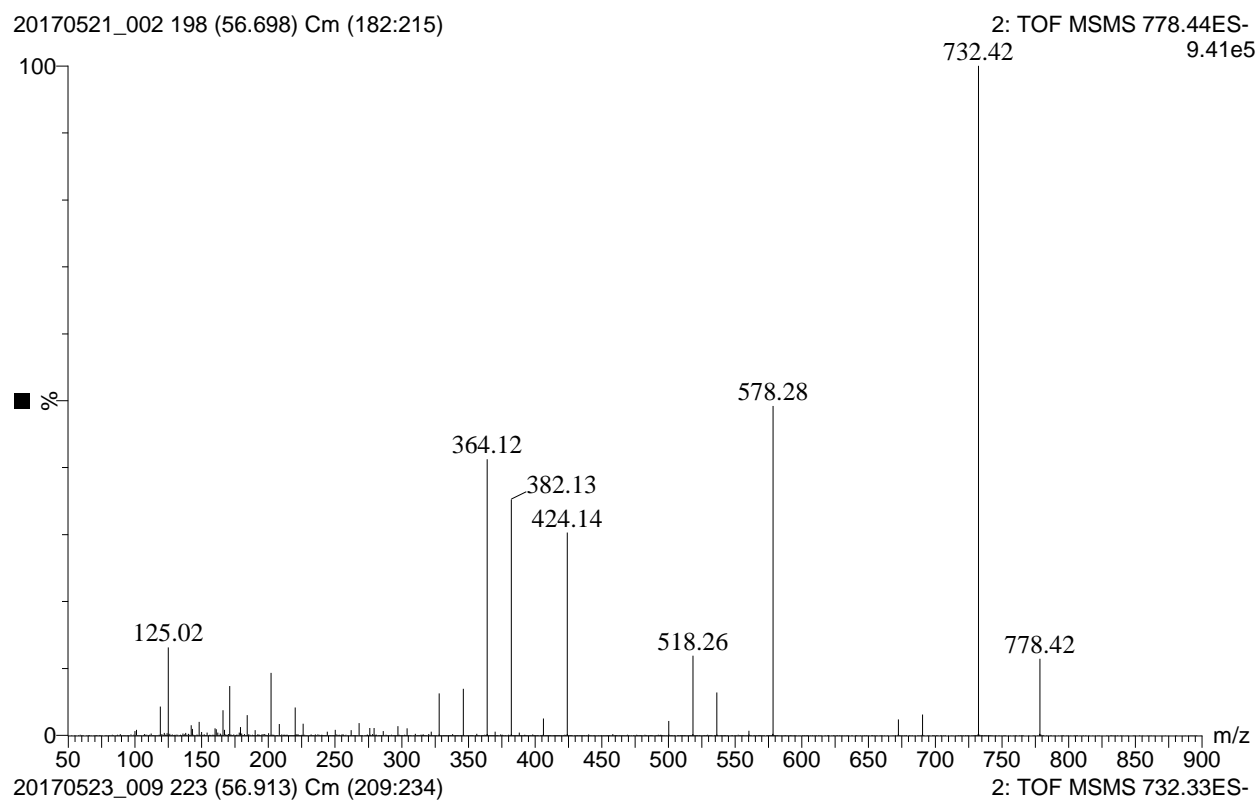
**Figure 3.9.** NAG-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top) & [M-H]⁻ (bottom)
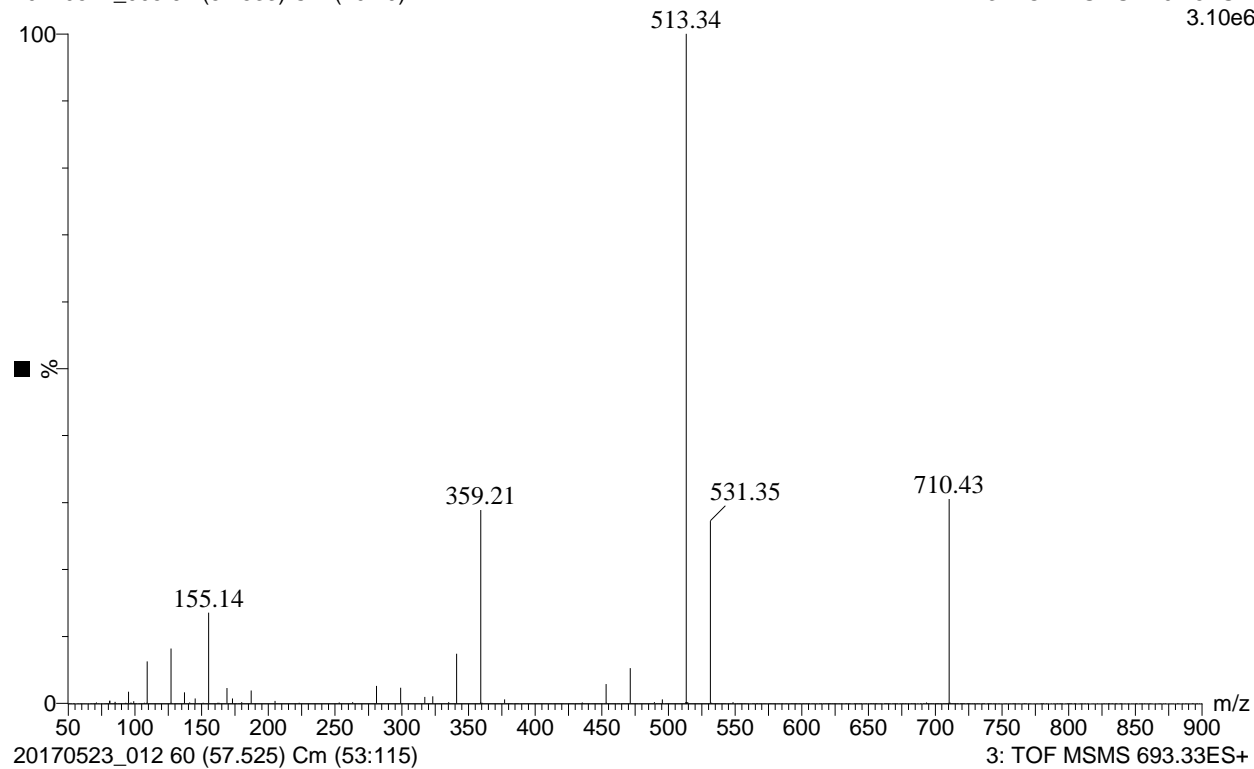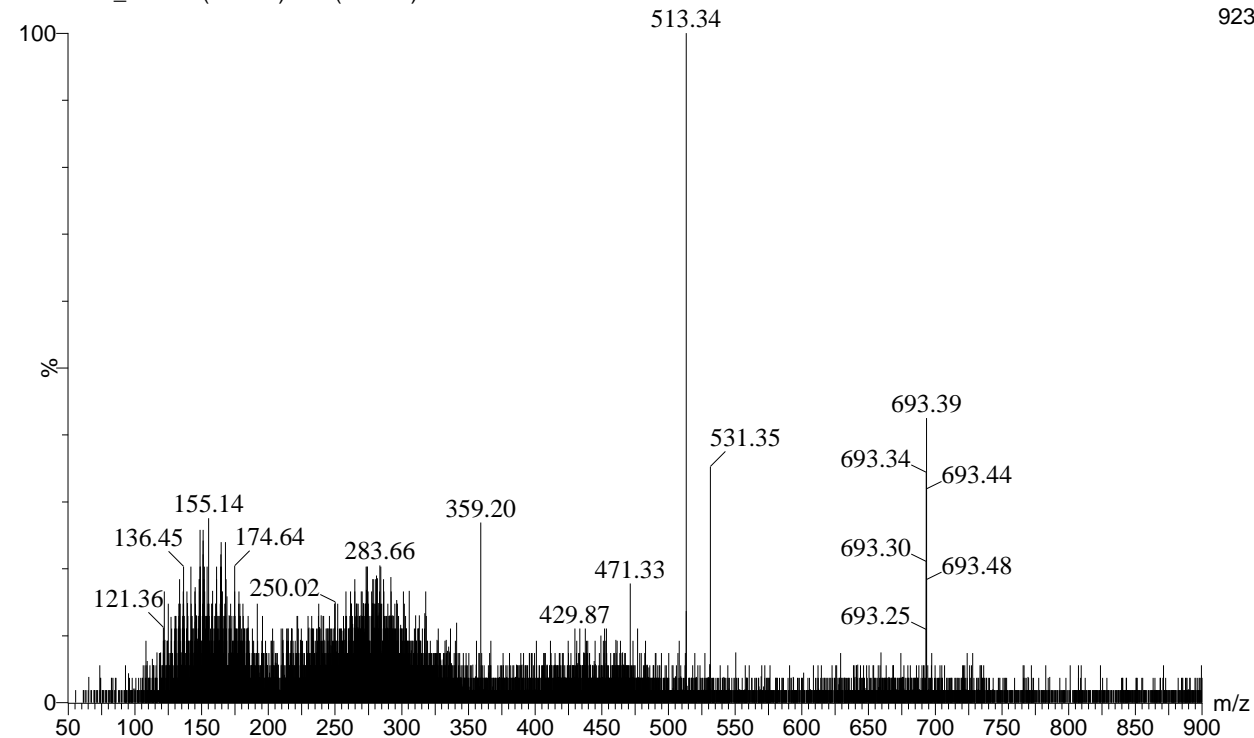
**Figure 3.10.** G-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)
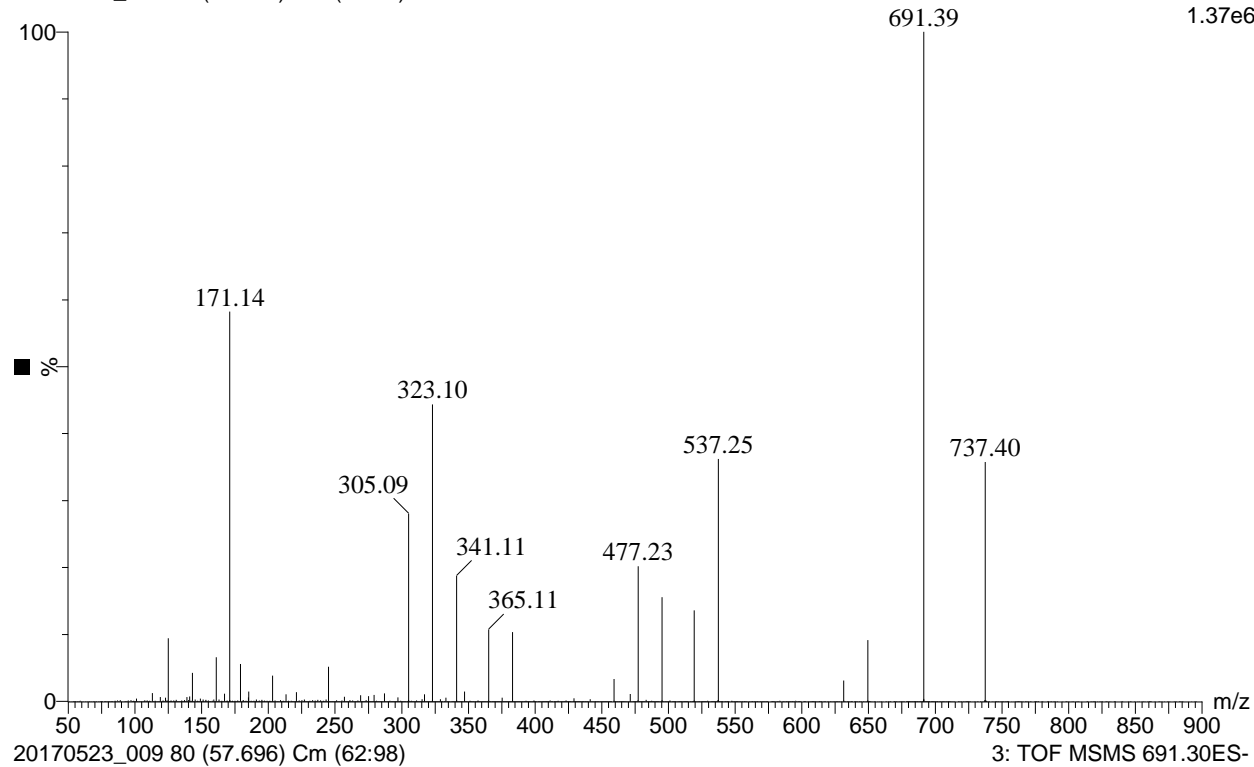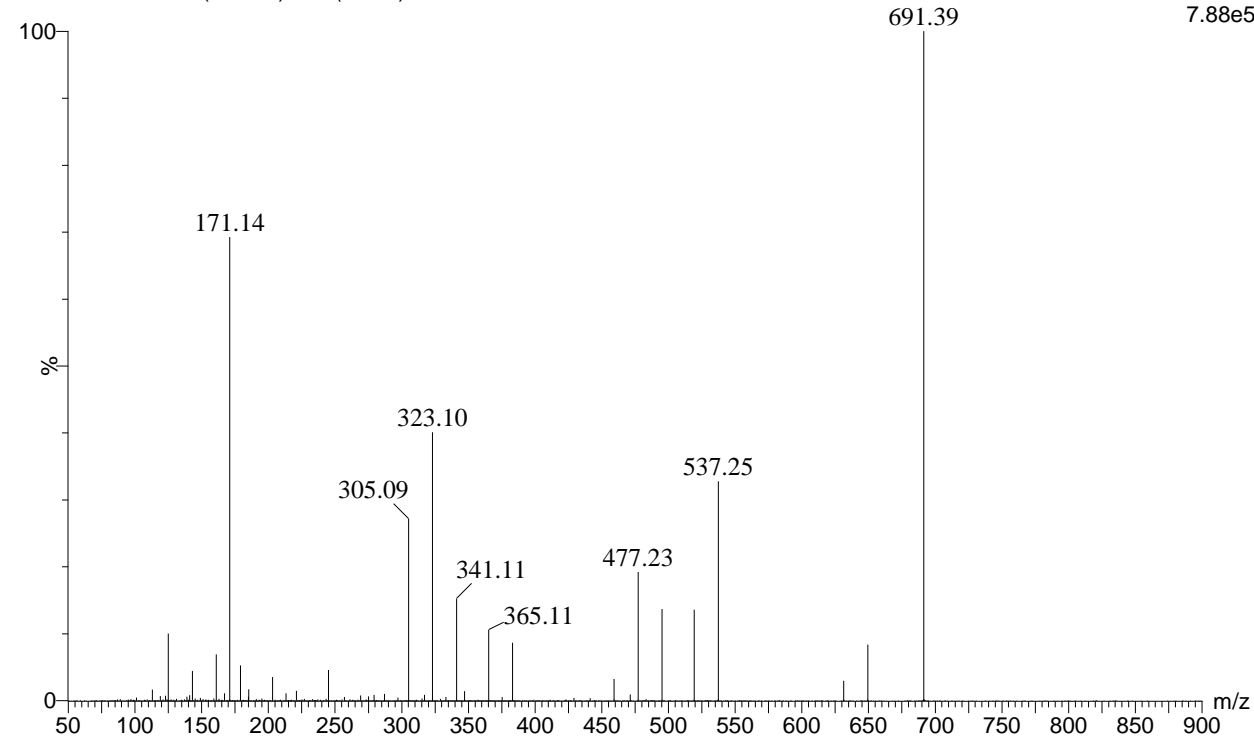
**Figure 3.11.** G-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top) & [M-H]⁻ (bottom)

287
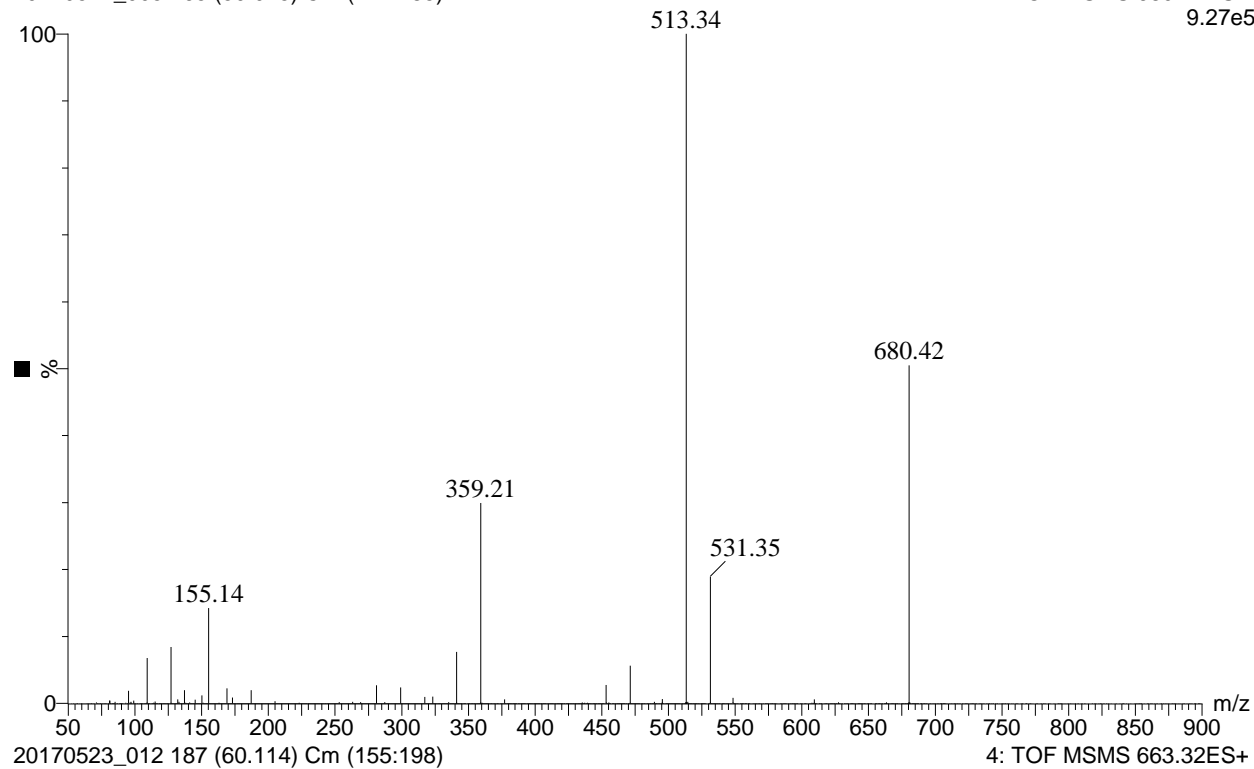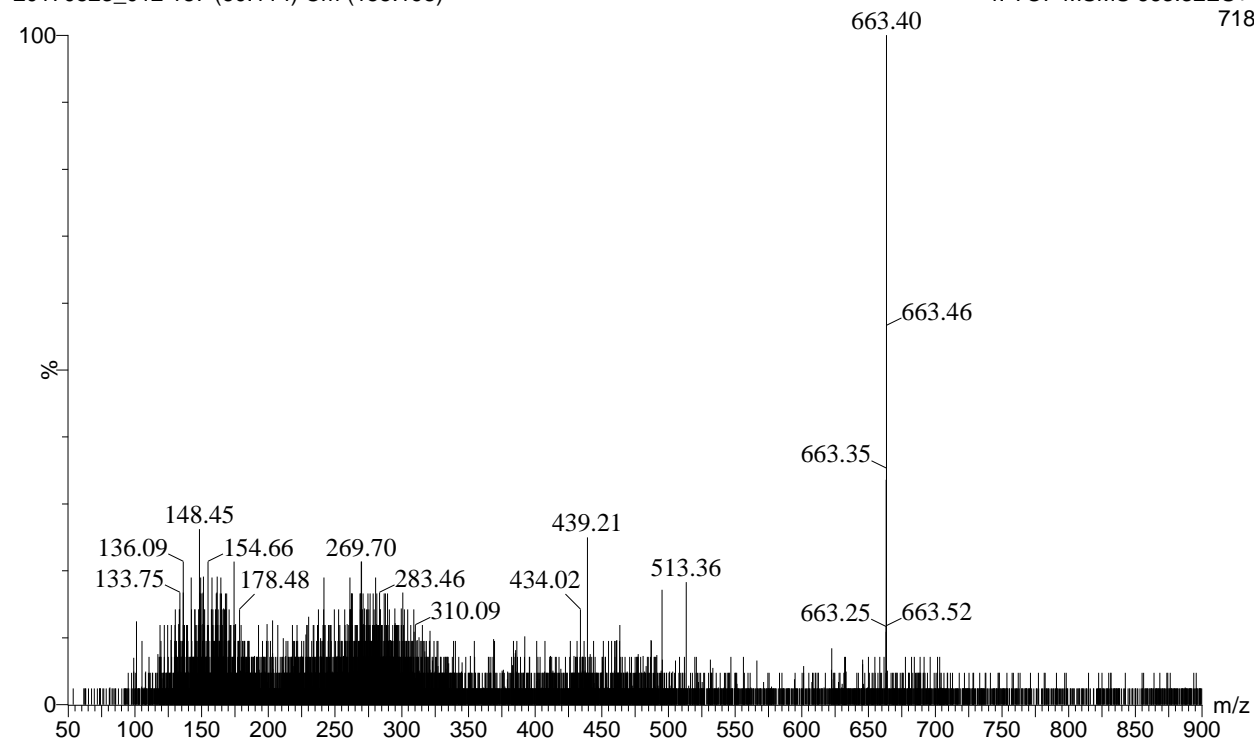
**Figure 3.12.** X-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)
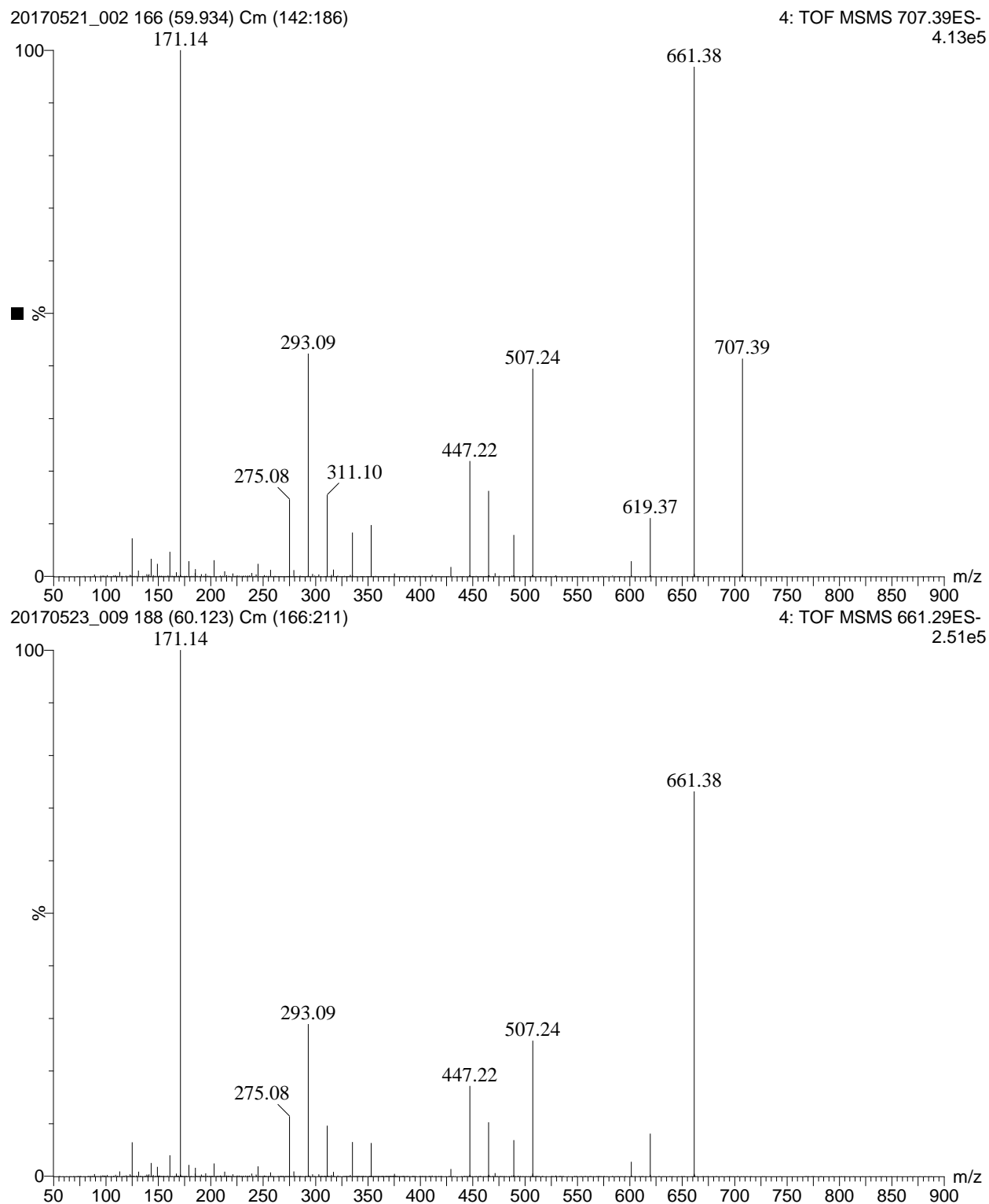
**Figure 3.13.** X-I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top) & [M-H]⁻ (bottom)
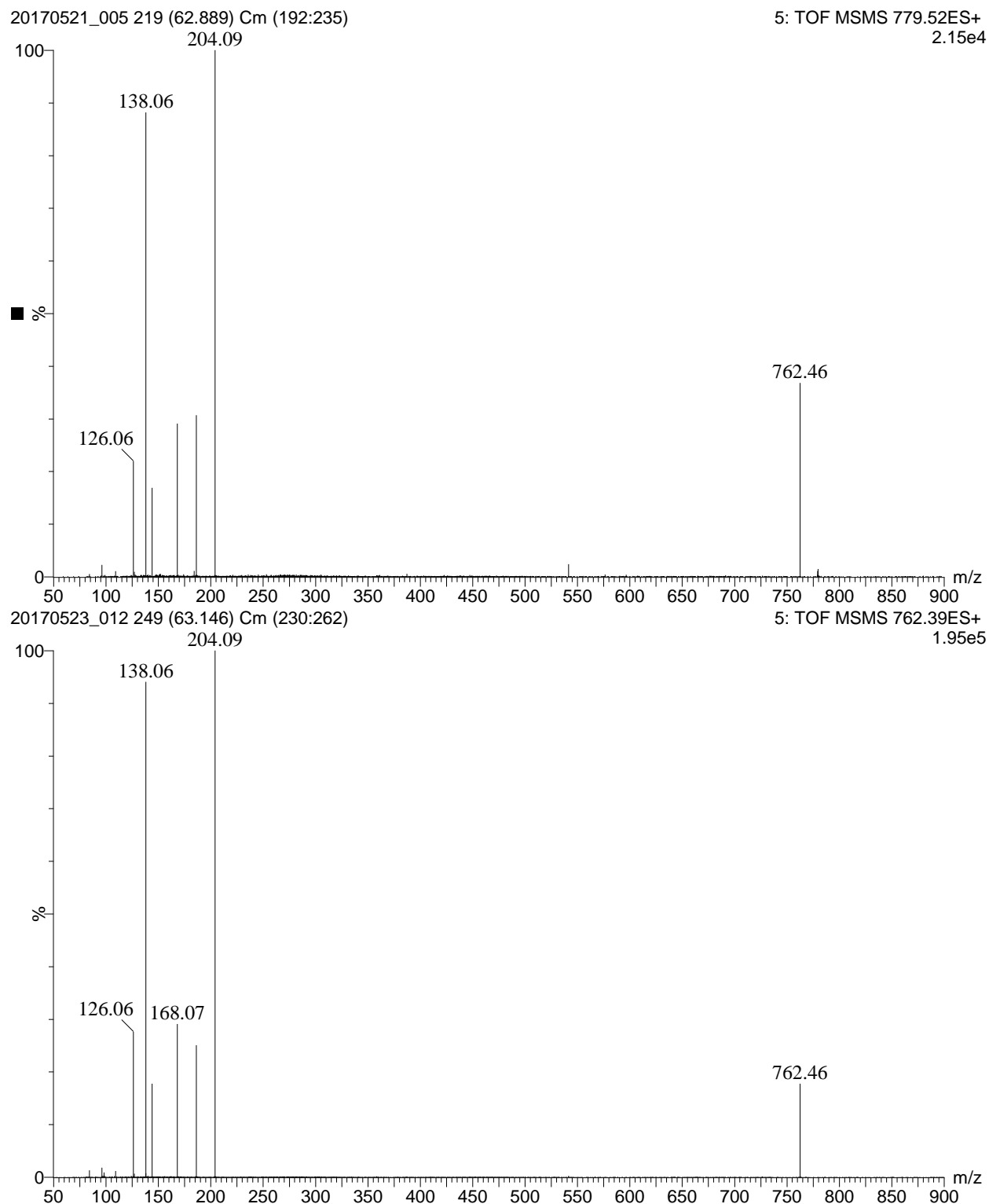
**Figure 3.14.** NAG-I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)
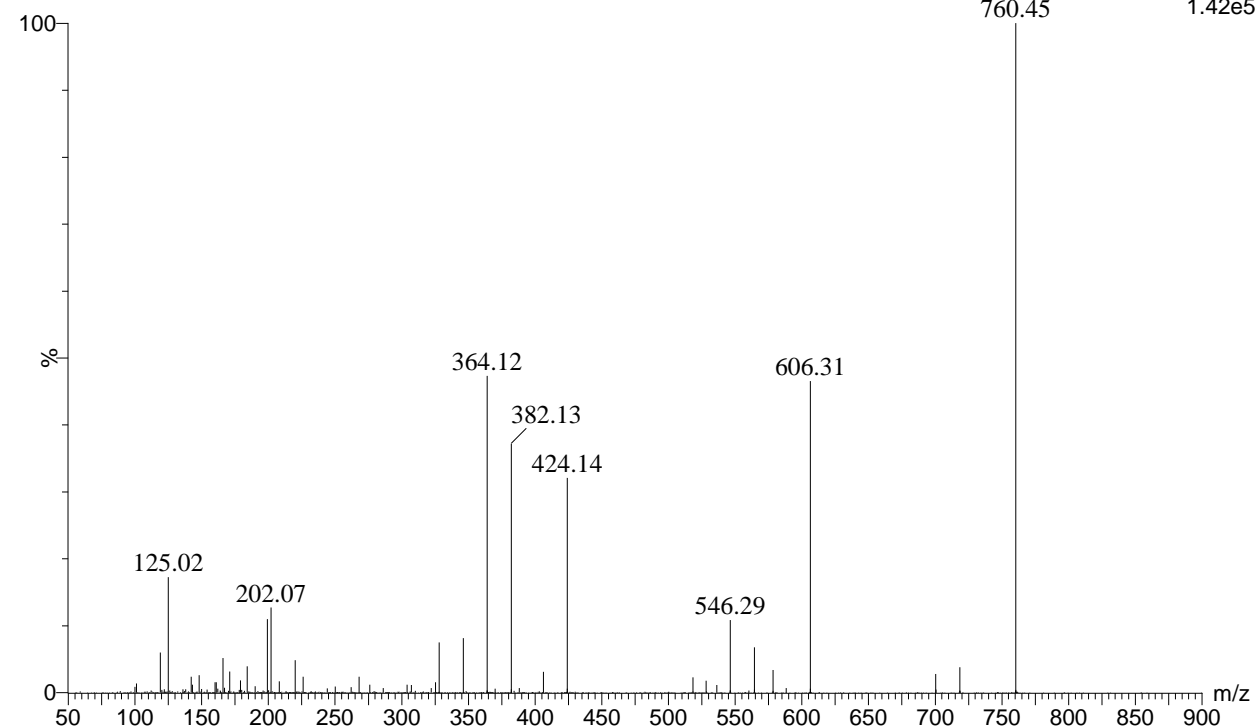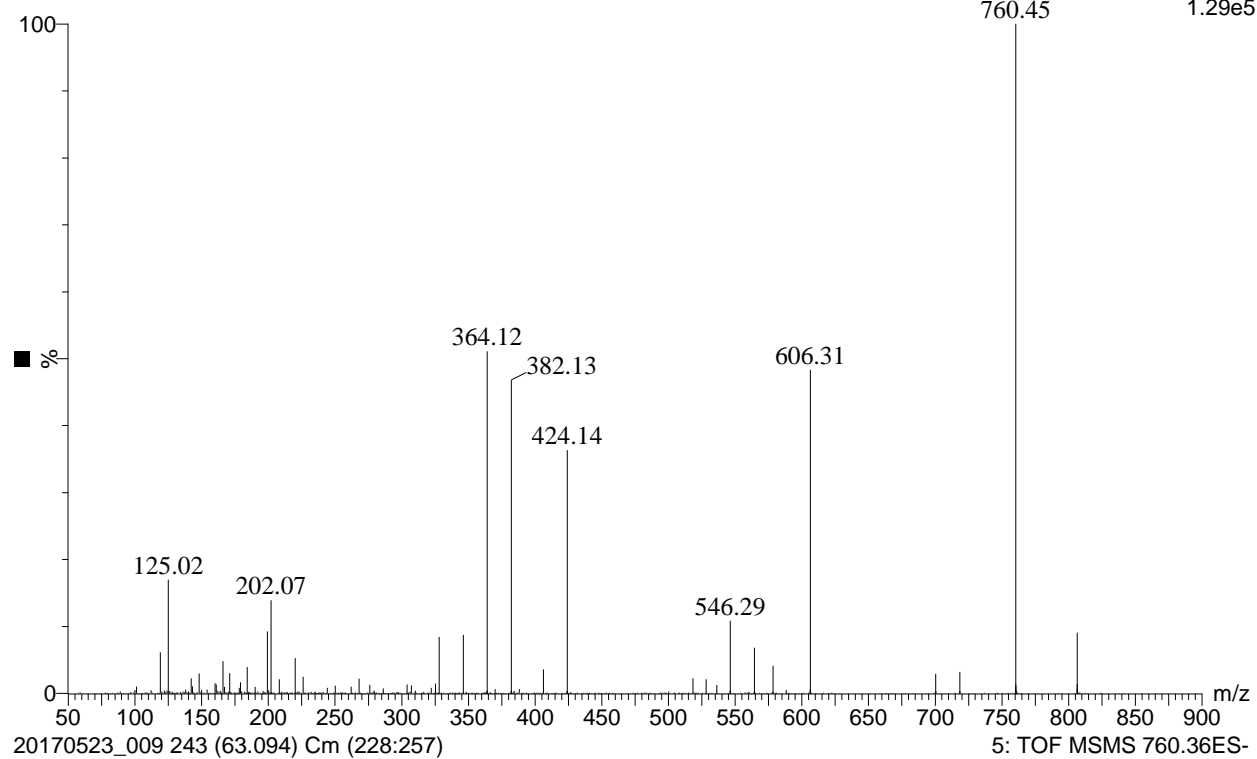
**Figure 3.15.** NAG-I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top) & [M-H]⁻ (bottom)
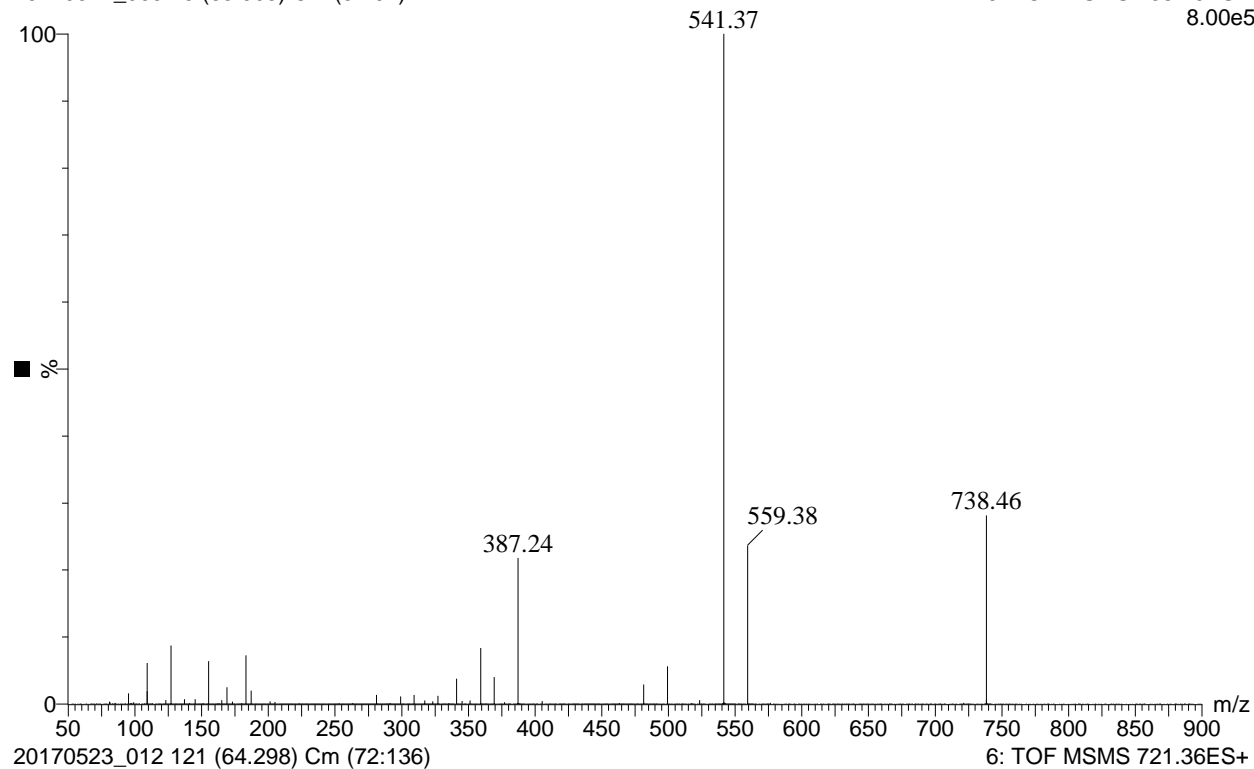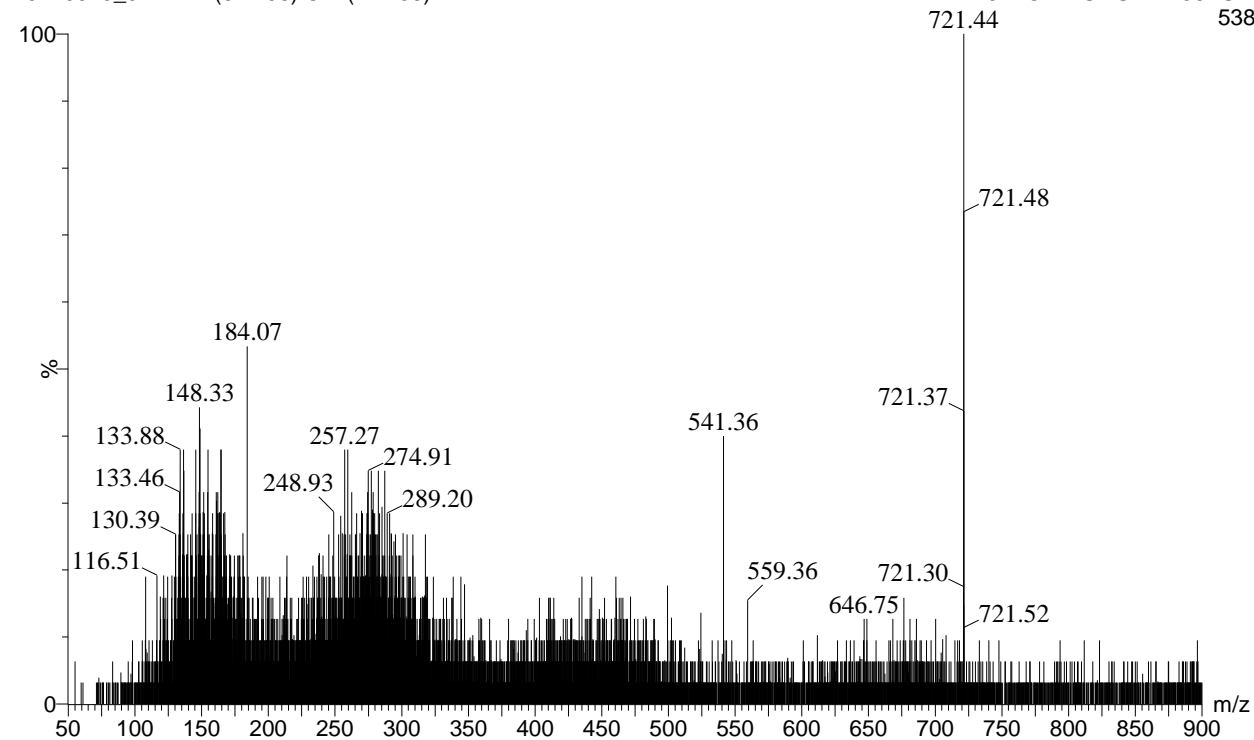
**Figure 3.16.** G-I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+NH₄]⁺ (top) & [M+H]⁺ (bottom)
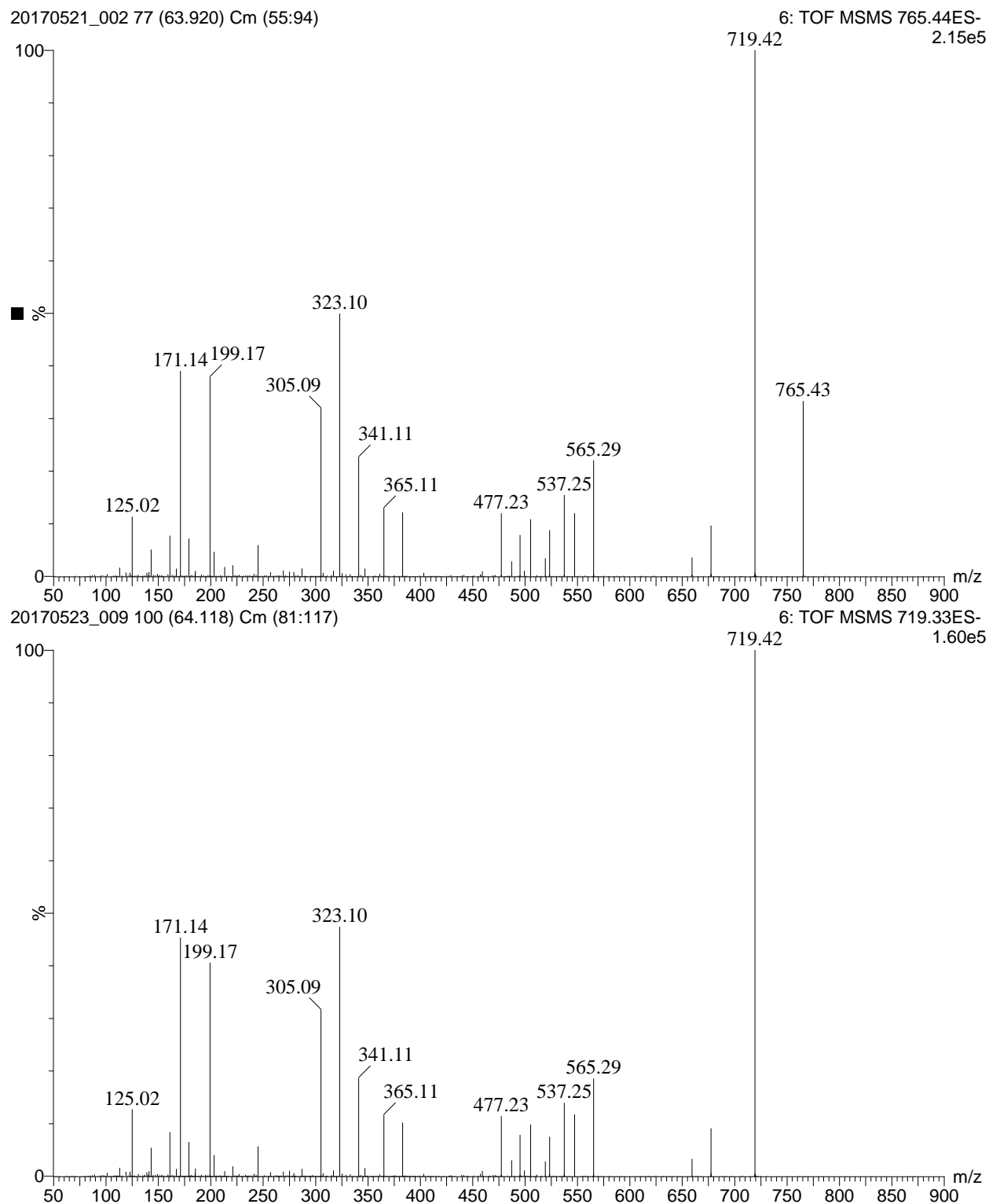
**Figure 3.17.** G-I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top) & [M-H]⁻ (bottom)

293
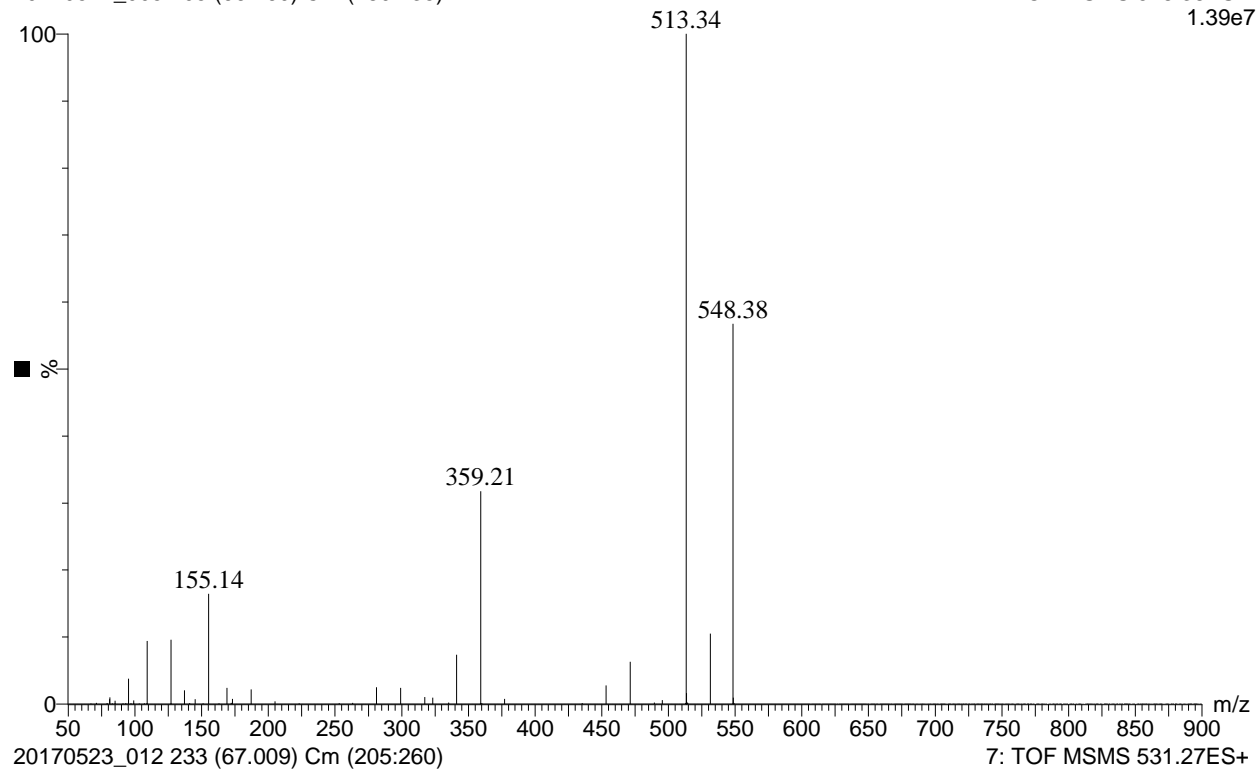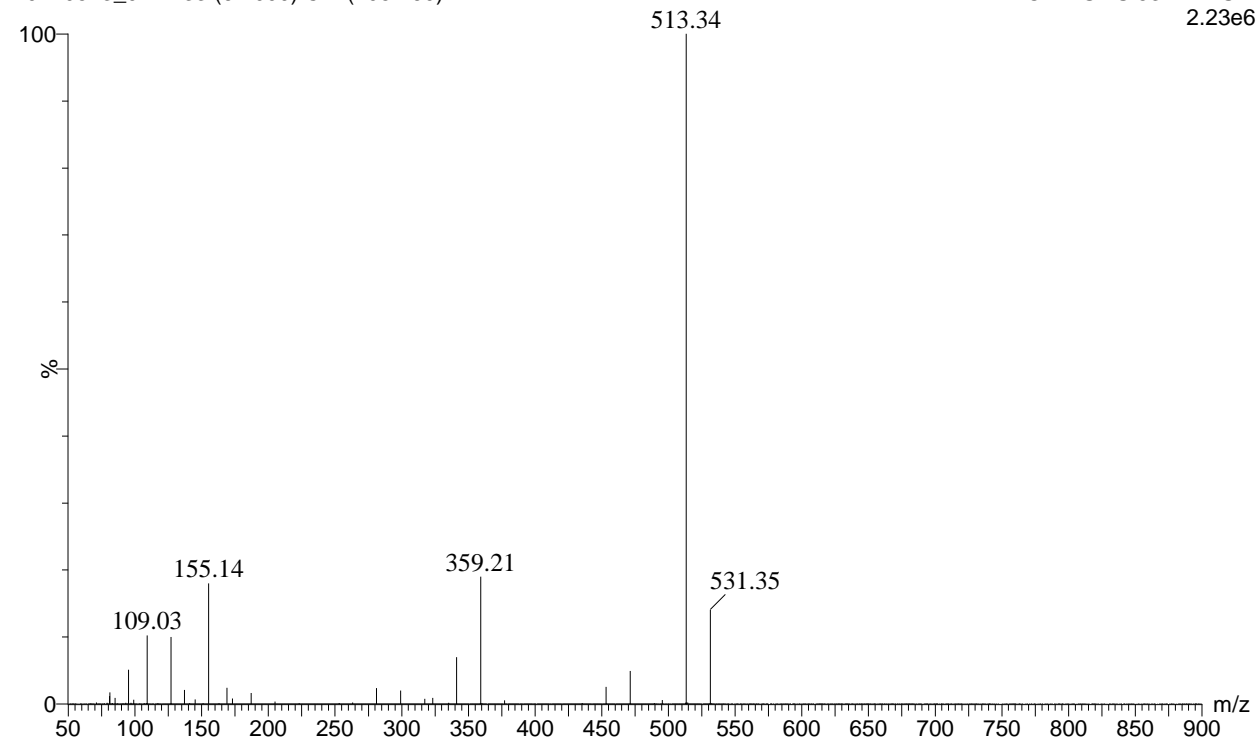
**Figure 3.18.** I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)

100   x10                                                ▸  575.34

171.14

■ %

529.34

0
     50   100  150  200  250  300  350  400  450  500  550  600  650  700  750  800  850  900   m/z

100        171.14

%

529.34

0
     50   100  150  200  250  300  350  400  450  500  550  600  650  700  750  800  850  900   m/z
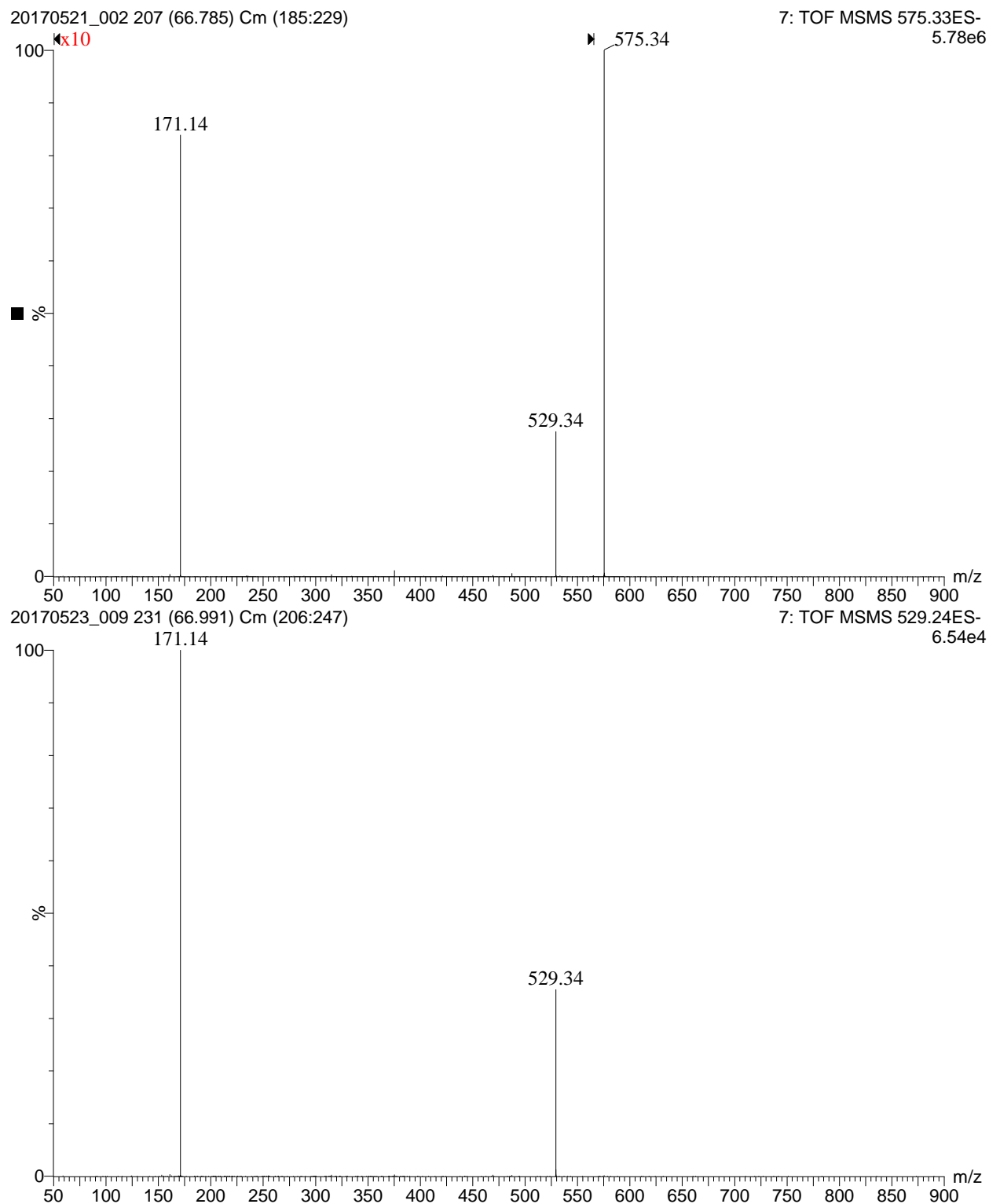
**Figure 3.19.** I3:22:0(2,10,10) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top, magnified 10x over *m/z* 51-566) & [M-H]⁻ (bottom)
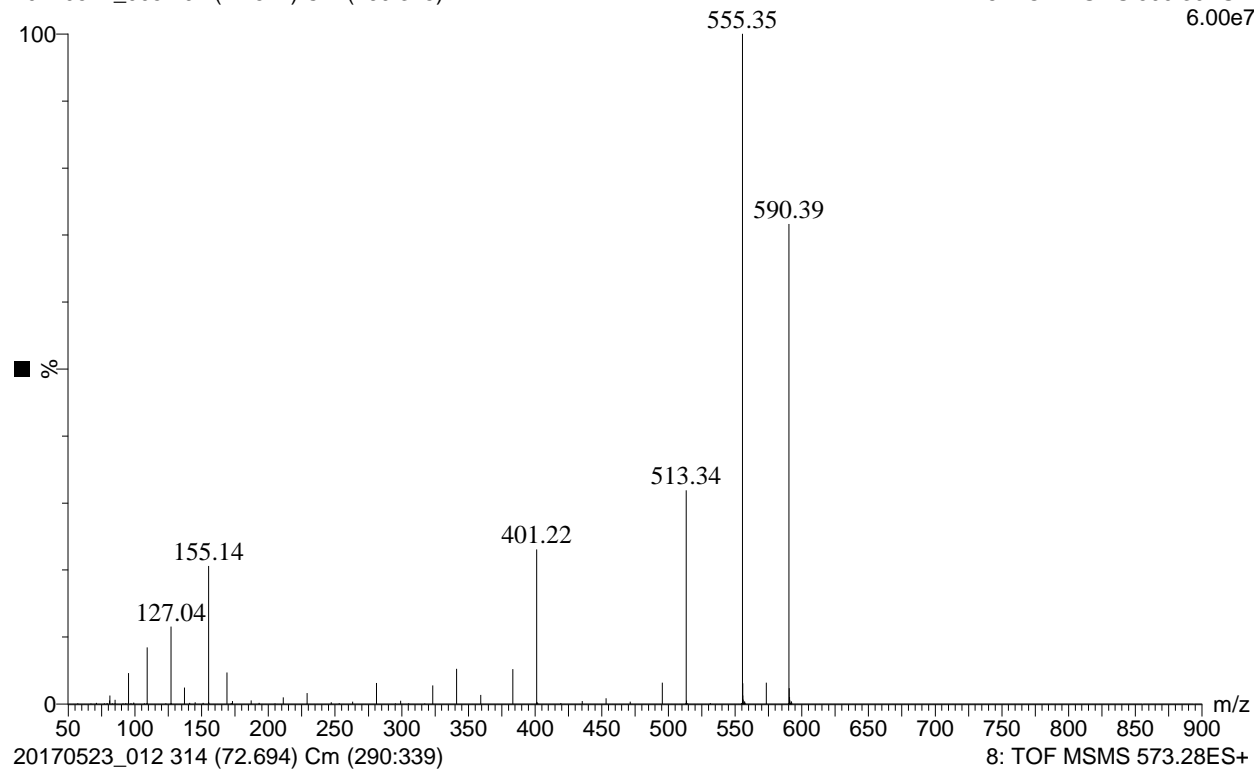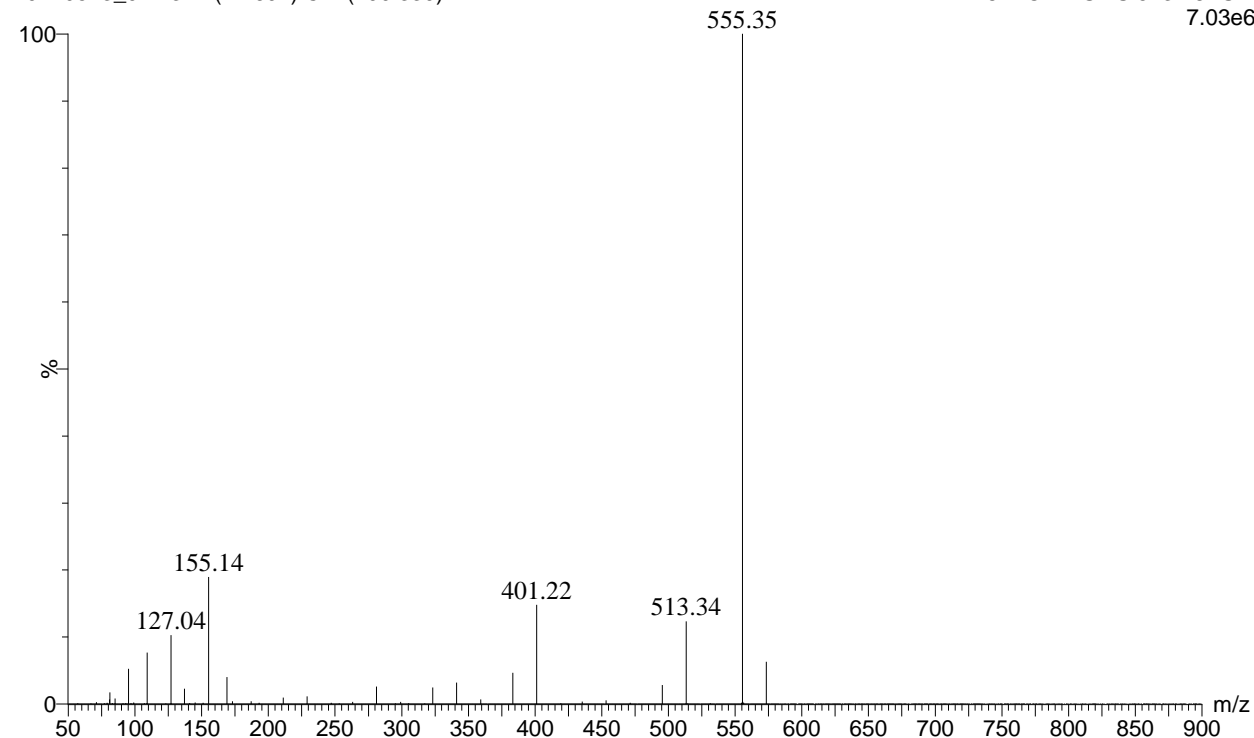
**Figure 3.20.** I4:24:0(2,2,10,10) MS/MS product ion spectra, precursor ion [M+NH4]+ (top) & [M+H]+ (bottom)
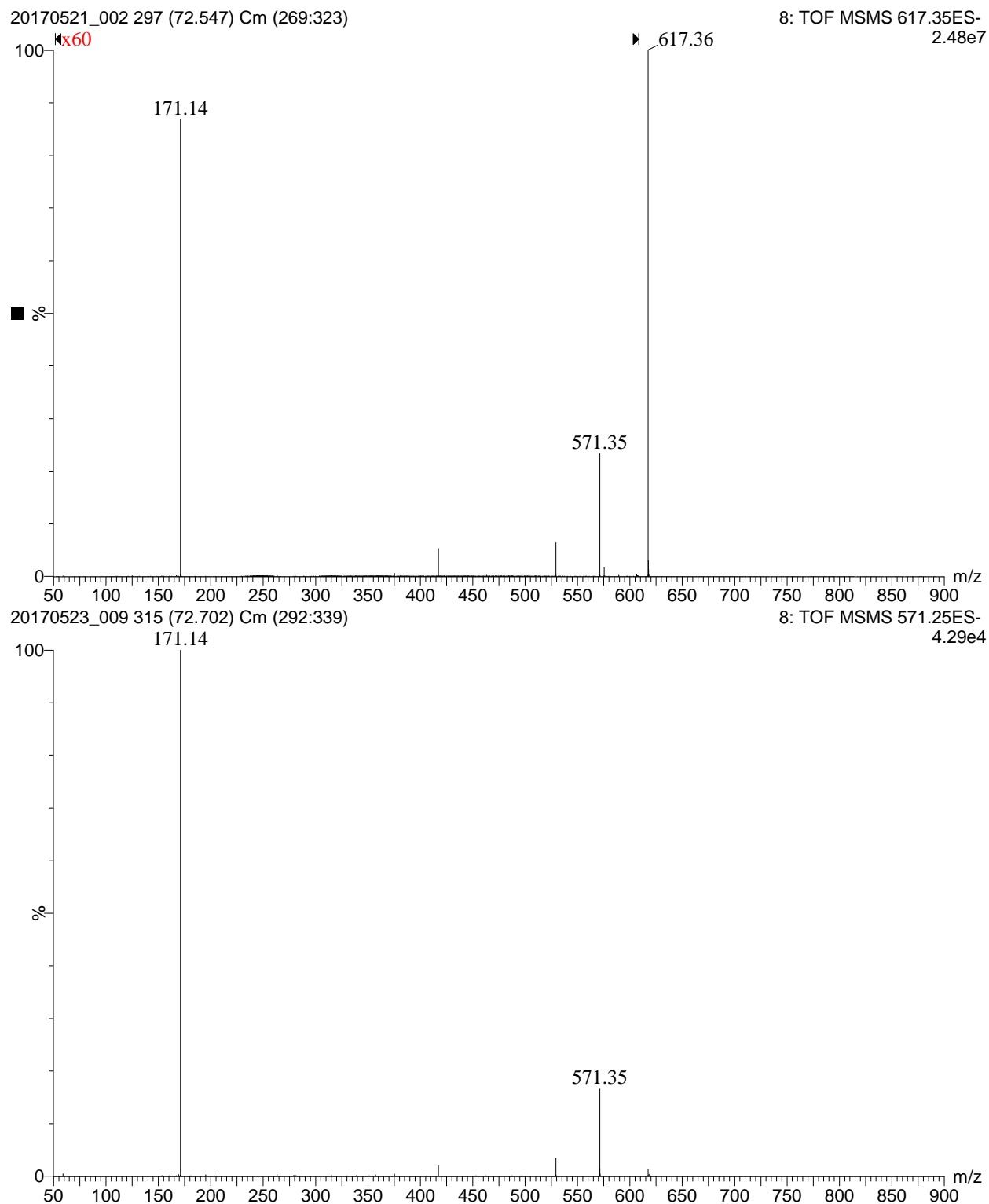
**Figure 3.21.** I4:24:0(2,2,10,10) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top, magnified 60x over *m/z* 51-608) & [M-H]⁻ (bottom)

100 ┤                                                    541.37

                                                         576.41

■ %

                                              387.24
          127.04              359.20

0 ┤
   50  100  150  200  250  300  350  400  450  500  550  600  650  700  750  800  850  900   m/z

100 ┤                                                    541.37

%

                                                              559.38
          127.04                    387.24

0 ┤
   50  100  150  200  250  300  350  400  450  500  550  600  650  700  750  800  850  900   m/z
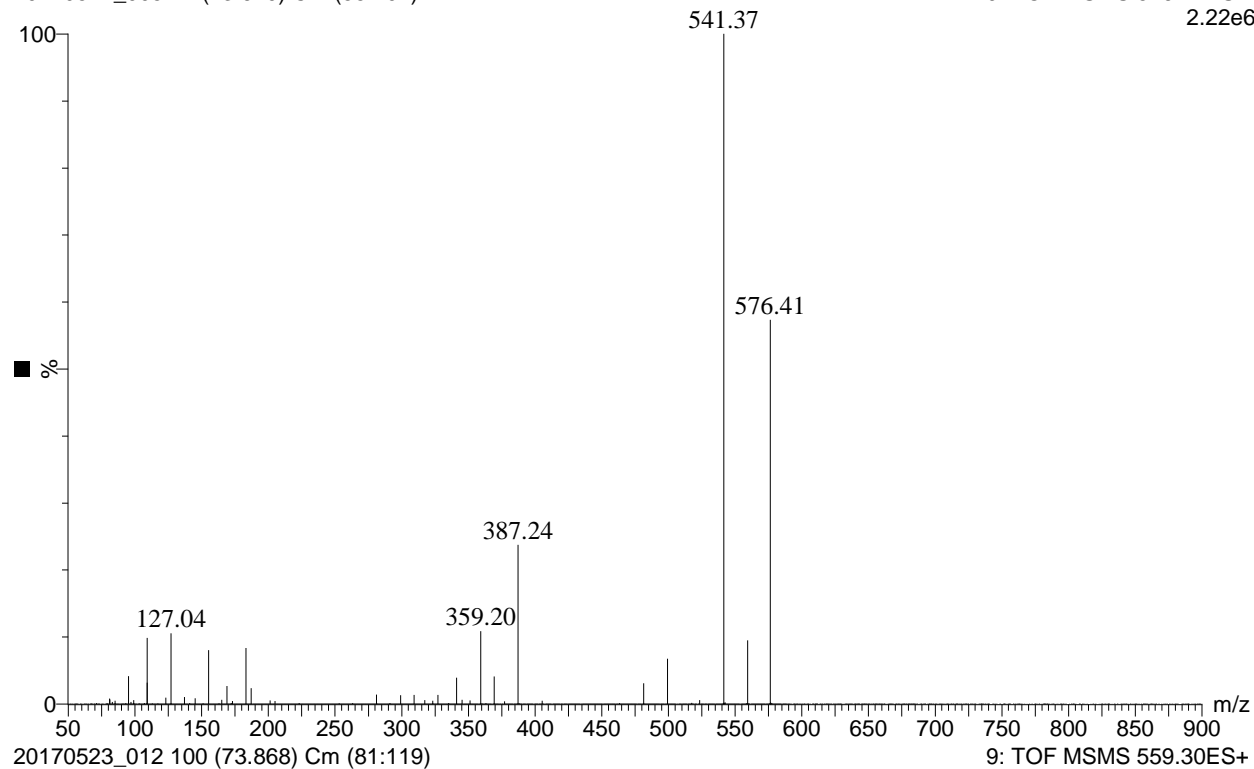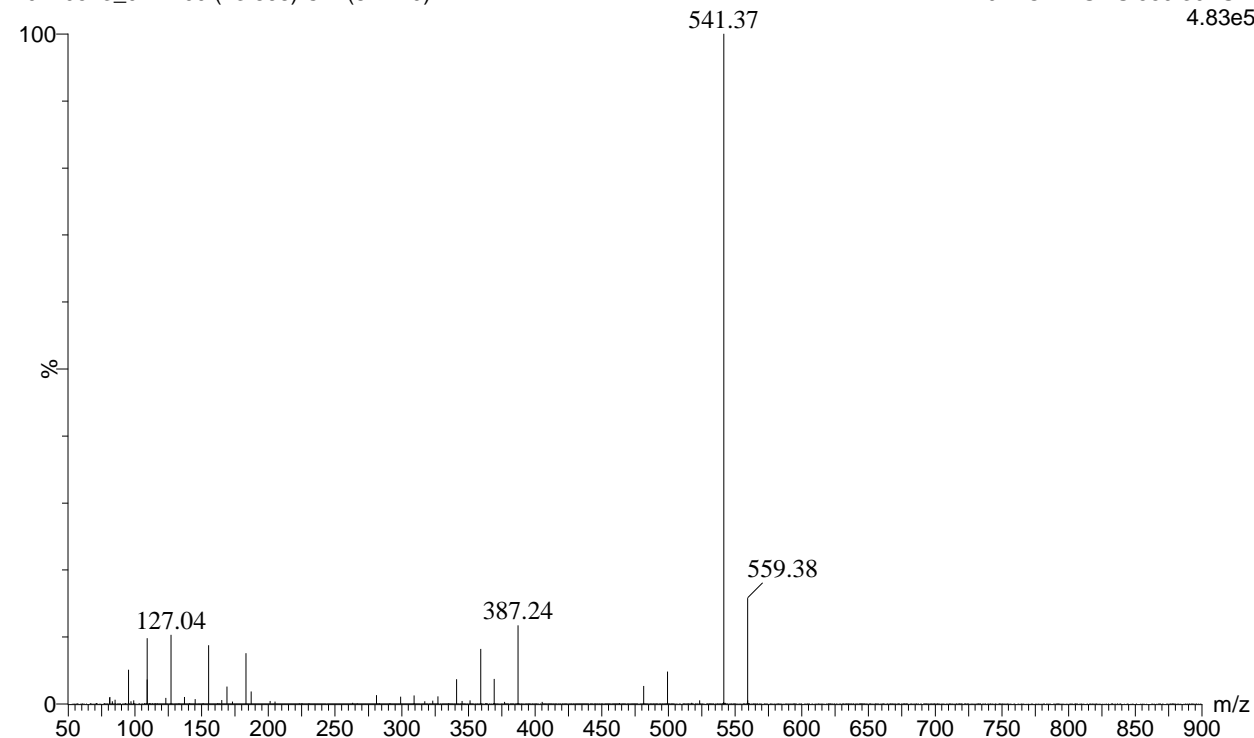
**Figure 3.22.** I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+NH$_4$]$^+$ (top) & [M+H]$^+$ (bottom)

298
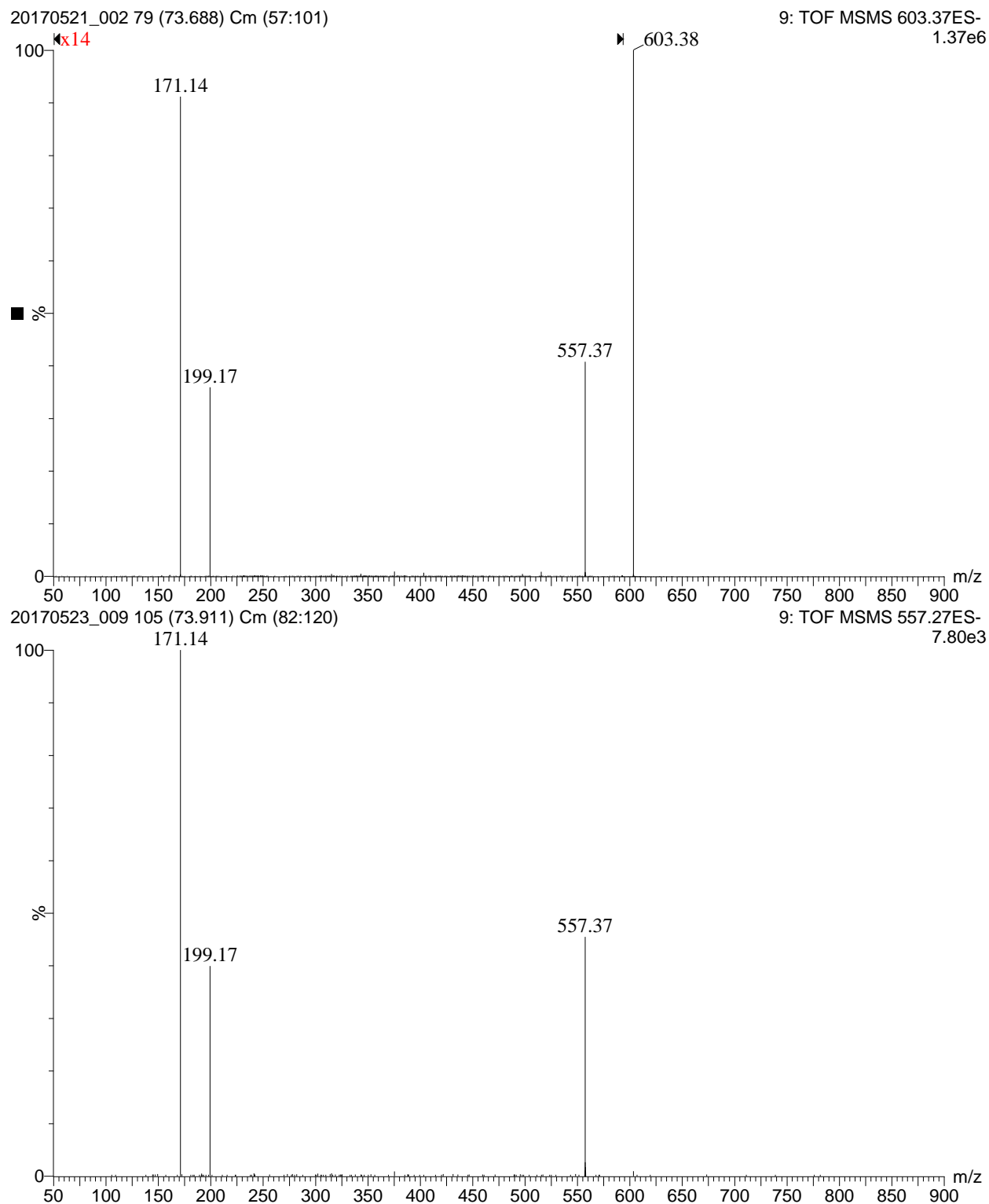
**Figure 3.23.** I3:24:0(2,10,12) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top, magnified 14x over *m/z* 51-594) & [M-H]⁻ (bottom)
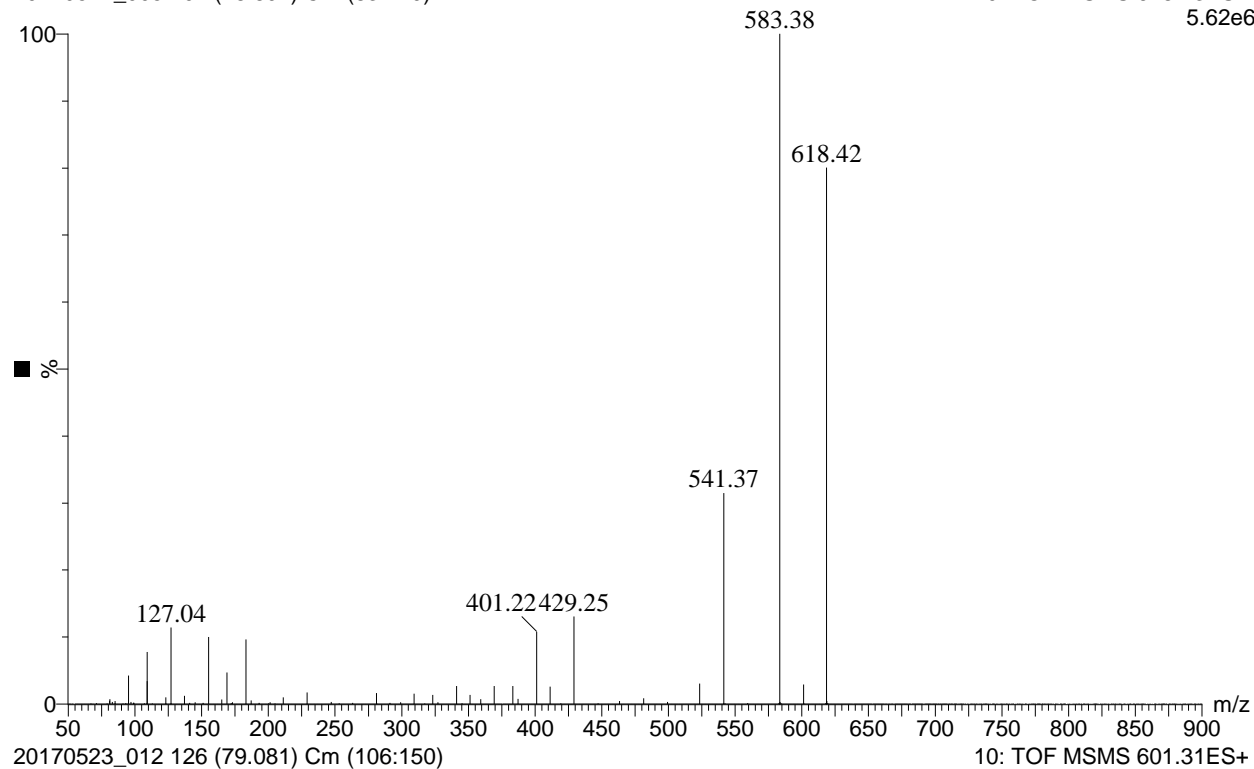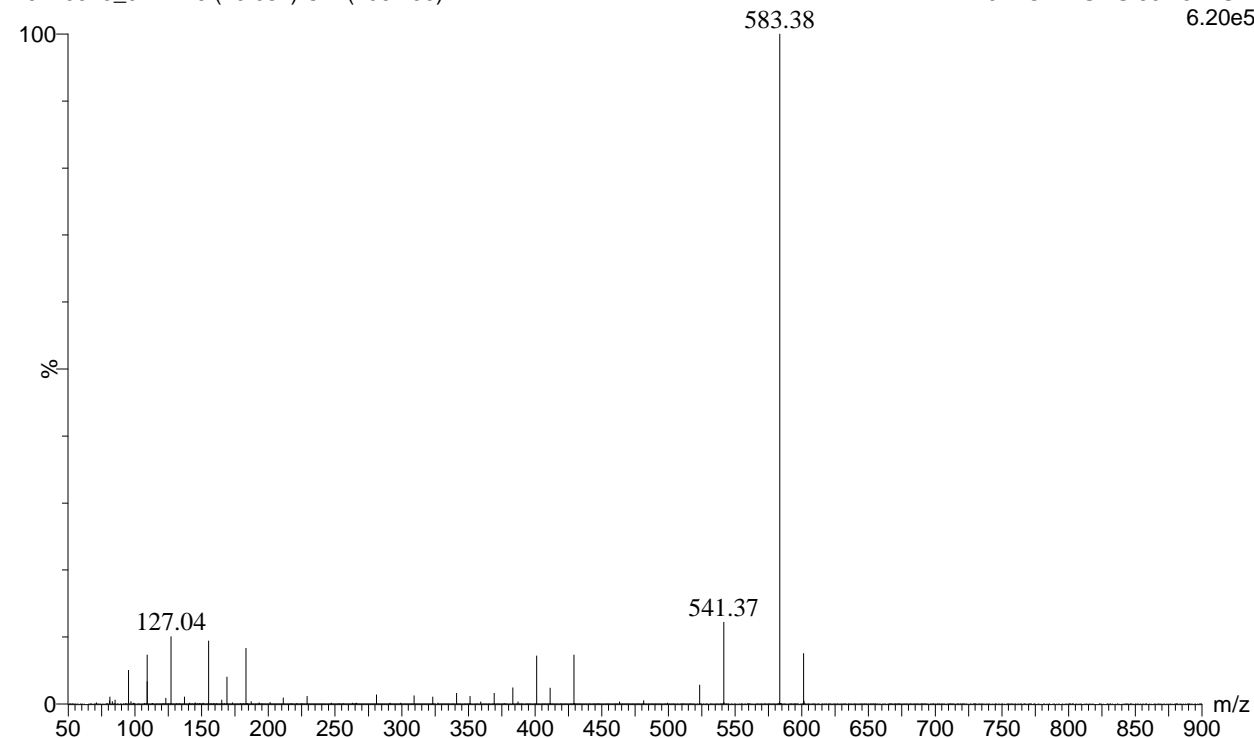
**Figure 3.24.** I4:26:0(2,2,10,12) MS/MS product ion spectra, precursor ion $[M+NH_4]^+$ (top) & $[M+H]^+$ (bottom)

300
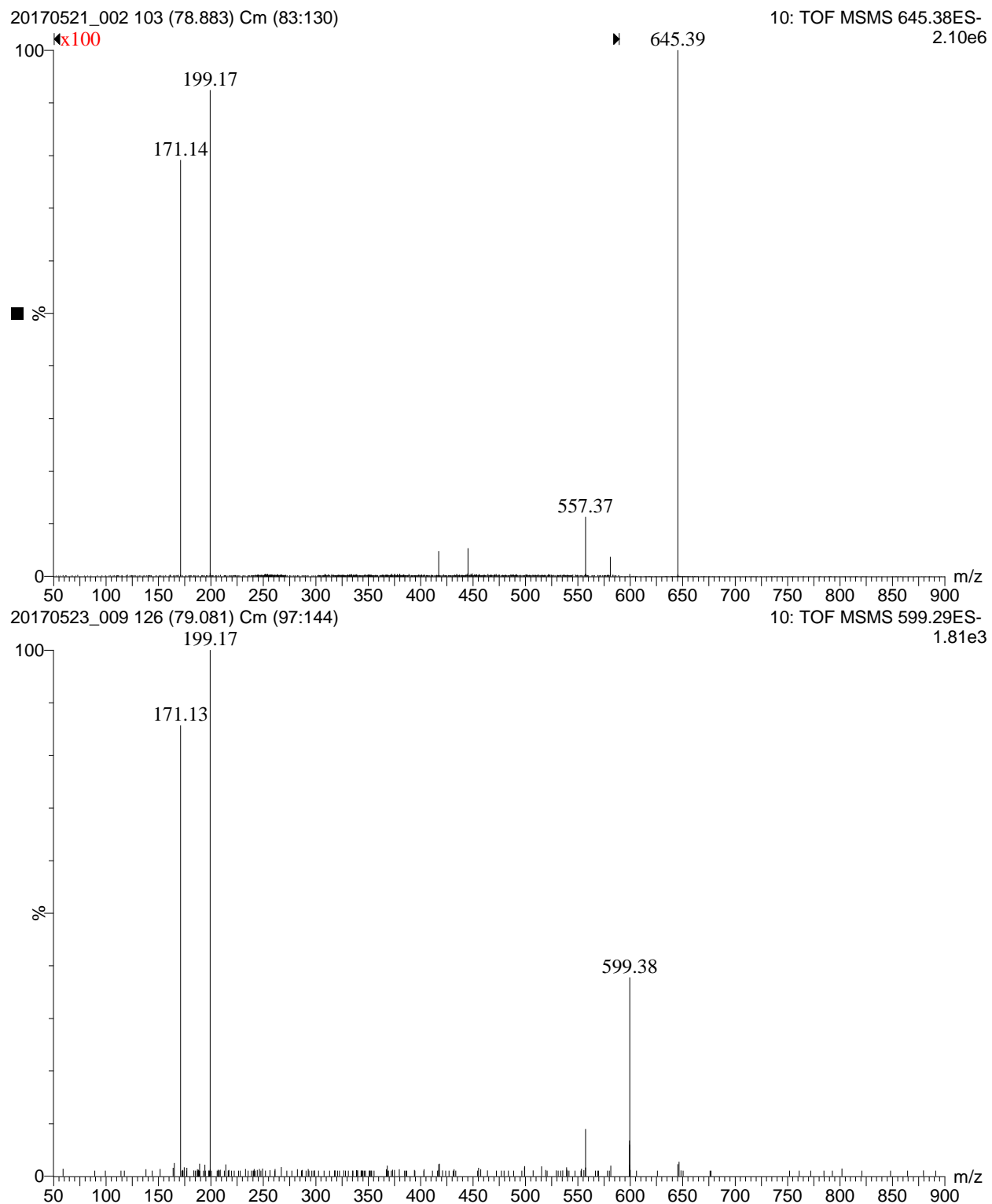
**Figure 3.25.** I4:26:0(2,2,10,12) MS/MS product ion spectra, precursor ion [M+HCOO]⁻ (top, magnified 100x over *m/z* 51-590) & [M-H]⁻ (bottom)

**Table 3.10.** Purification of acylinositols by semi-preparative HPLC

| | |
|---|---|
| Analyst | Steven M. Hurney |
| Instrument Location | A. Daniel Jones Laboratory |
| Instrument Title | Waters semi-prep HPLC (Jones Lab) |
| LC System | Waters 2795 Separations Module equipped with LKB Bromma 2211 Superrac Fraction Collector with automated fraction collection |
| Manufacturers | Waters |
| Column | Thermo Scientific Acclaim 120 C18 semi-preparative HPLC column (4.6 x 150 mm, 5 µm particle size) |
| Column Manufacturer | Thermo Scientific |
| Product Number | 059148 |
| Serial Number | 005992 |
| Packing Lot Number | 014-25-017 |
| Injection Volume | 200 µL |
| Flow Rate | 1.5 mL/min |
| Mobile Phases: | |
| A | 0.15% Formic Acid in Water |
| B | Acetonitrile |
| C | Methanol |
| Gradient Profile (solvents A and B only) | Hold 1% B at 0-1 min, linear 1-56% B at 1-2 min, linear 56%-68% B at 2-40 min, linear 68-90% B at 40-53 min, linear 90-100% B at 53-54 min (followed by a Column Wash). |
| Column Wash and Re-equilibration | Linear 100% B to 100% C at 54 to 55 min, hold 100% C at 55-65 min, linear 100% C to 100% B at 65-66 min, Linear 100% B to 1% B (99% A) at 66-67, hold 1% B at 67-70 min. |
| Column Oven Temperature | 50 ℃ |
| Sample Temperature in autosampler | Room Temperature |
| Method Name | SP_56C_70min_AcN_MeOH |
| Sample handling | Approximately 1 L of the *S. quitoense* bulk extract was added in portions of ~250 mL to a 1000 mL round bottom flask, and dried via rotary evaporation under vacuum at ~30°C, leaving a green residue. The residue was reconstituted in 5.0 mL AcN:IPA with sonication for 5 mins while manually swirling. The solution was centrifuged by Eppendorf Centrifuge 5480R at $10000 \times g$ for 10 mins. The supernatant was then transferred to LC autosampler vials for semi-preparative HPLC purification. |
| Sample Storage | -20 °C in autosampler vials |
| Protocol when analyzing the samples | One minute fractions were collected in Pyrex glass culture tubes ($18 \times 150$ mm) in five batches labeled letter A-E. Each batch consisted of six injections each. Fractions were concentrated to dryness under vacuum using a Thermo Savant SPD 131 DDA SpeedVac Concentrator with BOC Edwards XDS Dry Pump. Fractions were labeled by minute, with fractions 15-55 containing the most abundant acylinositols. |

**Table 3.10.** (continued)

| | |
|---|---|
| Protocol for testing the fractions | In order to test for purity and reproducibility, fractions from each batch were reconstituted in 0.50 mL AcN:IPA, and transferred to LC autosampler vials. 5.0 µL aliquots from each fraction were diluted in autosampler vials containing 0.50 mL of AcN:IPA and analyzed on an LCT Premier mass spectrometer equipped with Shimadzu LC-20AD pumps, Shimadzu SIL-5000 autosampler and a Shimadzu CTO-20A column oven using a 30 min gradient (0.3 mL min$^{-1}$, Ascentis Express C18 Analytical HPLC, 10cm x 2.1mm x 2.7µm) optimized for separating *S. quitoense* acylsugars. The elution program is as follows: hold 1% B at 0-1 min, linear 1-55% B at 1-2 min, linear 55%-85% B at 2-22 min, linear 85-100% B at 22-22.01 min, hold 100% B at 22.01-27 min, linear 100-1% B at 27-27.01 min, hold 1% B at 27.01-30 min. HPLC fractions were combined according to metabolite purity and by comparison to a *S. quitoense* bulk extract solution analyses run before and after each group of fractions (i.e. the sample order was blank, *S. quitoense* bulk extract, ~20 fraction samples in ascending elution order, blank, *S. quitoense* bulk extract, and etc.). |

**Table 3.11.** Bruker 900 MHz NMR Instrument Metadata

| | |
|---|---|
| Facility Supervisor | Dr. Daniel Holmes |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Max T. Rogers NMR Facility |
| Facility Instrument Title | 900 MHz Bruker |
| Time of acquisition | October 2015 - February 2016 |
| Manufacturer | Bruker |
| Field Frequency Lock | Acetonitrile-d$_3$ |
| Additional Solute | None |
| Solvent | 300 µL CD$_3$CN |
| Chemical Shift Standard | CH$_3$CN-d$_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm) |
| Concentration Standard | None |
| Instrument | Bruker Avance 900 MHz NMR |
| Geographic location of instrument | 42.7164, -84.4677 |
| Magnet | 899.13-899.00 MHz |
| Probe | Bruker TCI triple-resonance inverse detection cryoprobe |
| Console | Bruker Avance |
| Acquisition Software | Topspin 2.1.6 |
| Acquisition Parameters: | |
| a) Acquisition parameters file reference | $^1$H:/opt/topspin216/exp/stan/nmr/lists/pp/zg <br><br> $^{13}$C:/opt/topspin216/exp/stan/nmr/lists/pp/zgpg30 <br><br> COSY:opt/topspin216/exp/stan/nmr/lists/pp/cosygpmfph <br><br> HSQC:/opt/topspin216/exp/stan/nmr/lists/pp/hsqcedetgpsisp2.2 <br><br> coupledHSQC:/opt/topspin216/exp/stan/nmr/lists/pp/hsqcedetgpsisp2.2nd <br><br> HMBC: /opt/topspin216/exp/stan/nmr/lists/pp/hmbcgplpndqf <br><br> *J*-resolved: /opt/topspin216/exp/stan/nmr/lists/pp/jresqf <br><br> ROESY: /opt/topspin216/exp/stan/nmr/lists/pp/roesyetgp |
| b) Sample Details | Shigemi (5 mm) NMR tube, Temperature @ 298 K, No Spinning |
| c) Instrument operation details (recorded for each sample independently and are roughly the same for all samples measured; displayed here is the sample containing NAG-I3:22:0(2,10,10) as an example) | Radiation frequency (MHz): <br><br> $^1$H: 899.1300264 <br><br> $^{13}$C: 226.0861132 <br><br> COSY: 899.1300240 (F2), 899.1300049 (F1) <br><br> HSQC: 899.1300246 (F2), 226.0861186 (F1) <br><br> HMBC: 899.1300260 (F2), 226.0860984 (F1) <br><br> *J*-resolved: 899.1300262 (F2), 899.1340467 (F1) <br><br> ROESY: 899.1300266 (F2), 899.1300267 (F1) <br><br> Acquisition nucleus: <br><br> $^1$H: 90° = 7.25 µs, $^{13}$C: 90° = 20.0 µs |

**Table 3.11.** (continued)

| | |
|---|---|
| d) Number of scans (scans, dummy scans) | ¹H: 32, 0<br><br>¹³C: 2000-4000, 8<br><br>COSY: 4-8, 16<br><br>HSQC: 8-16, 16<br><br>Coupled-HSQC: 12-16, 16<br><br>HMBC: 20-40, 32<br><br>*J*-resolved: 16-32, 4<br><br>ROESY: 8-16, 8 |
| e) Number of data points acquired (F2, F1) | ¹H: 148144<br><br>¹³C: 65536<br><br>COSY: 2048, 200<br><br>HSQC: 1024, 400<br><br>Coupled-HSQC: 4308, 400<br><br>HMBC: 4096, 480<br><br>*J*-resolved: 2048, 128<br><br>ROESY: 1982, 512 |
| e) Dwell time (µs) | ¹H: 27.0<br><br>¹³C: 9.225<br><br>COSY: 46.4<br><br>HSQC: 46.4<br><br>Coupled-HSQC: 46.4<br><br>HMBC: 46.4<br><br>*J*-resolved: 46.4<br><br>ROESY: 46.4 |
| FID and spectral processing parameters: | |
| Processing software | Topspin 3.5 pl 7 |
| a) Number of data points in spectrum (F2, F1) | ¹H: 262144<br><br>¹³C: 131072<br><br>COSY: 4096, 4096<br><br>HSQC: 1024, 1024<br><br>Coupled-HSQC: 8192, 1024<br><br>HMBC: 4096, 2048<br><br>*J*-resolved: 4096, 4096<br><br>ROESY: 4096, 2048 |

**Table 3.11.** (continued)

| | |
|---|---|
| b) Window function details | $^1$H: exponential multiply, LB = 0 Hz<br><br>$^{13}$C: exponential multiply, LB = 2 Hz<br><br>COSY: QSINE, QSINE; LB = -0.3, -0.3 Hz; GB = 0.3, 0; SSB = 2, 2; TM1 = 0, 1; TM2 = 0, 1<br><br>HSQC: QSINE, QSINE; SSB = 2, 2; TM1 = 0, 0.1; TM2 = 0, 0.9<br><br>Coupled-HSQC: QSINE, QSINE; SSB = 2, 2; TM1 = 0, 0.1; TM2 = 0, 0.9<br><br>HMBC: QSINE, SINE<br><br>*J*-resolved: SINE, SINE; LB = -0.3, -0.3 Hz; GB = 0.3, 0; TM1 = 0, 1; TM2 = 0, 1<br><br>ROESY: QSINE, QSINE; LB = 1.0, 0.3 Hz; GB = 0.35, 0.1; SSB = 2, 2; TM1 = 0, 0.1; TM2 = 0, 0.9 |

**Table 3.12.** Agilent DDR2 500 MHz NMR Instrument Metadata

| | |
|---|---|
| Facility Supervisor | Dr. Daniel Holmes |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Max T. Rogers NMR Facility |
| Facility Instrument Titles | Ahriman and Ormuzd |
| Time of acquisition | October 2015 |
| Manufacturer | Agilent |
| Field Frequency Lock | Acetonitrile-d$_3$ |
| Additional Solute | None |
| Solvent | CD$_3$CN: 500 μL |
| Chemical Shift Standard | CH$_3$CN-d$_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm) |
| Concentration Standard | None |
| Instrument | Agilent DDR2 500 MHz with 7600AS 96 autosamplers |
| Geographic location of instrument | 42.7288, -84.4745 |
| Magnet | 499.91 MHz |
| Probe | OneNMR Probe with Protune accessory for hands-off tuning |
| Acquisition Software | VnmrJ 3.2A |
| Acquisition Parameters: | |
| a) Acquisition parameters file reference | $^1$H: VnmrJ/Experiment Selector/Common/PROTON<br><br>$^{13}$C: VnmrJ/Experiment Selector/Common/CARBON<br><br>COSY: VnmrJ/Experiment Selector/Common/(HH)gCOSY<br><br>HSQC: VnmrJ/Experiment Selector/Common/(HC)gHSQCAD<br><br>HMBC: VnmrJ/Experiment Selector/Common/(HC)gHMBCAD<br><br>*J*-resolved: VnmrJ/Experiment Selector/Liquid/JSpectra/ HOMO2DJ<br><br>TOCSY: VnmrJ/Experiment Selector/Liquid/TOCSY<br><br>ROESY: VnmrJ/ Experiment Selector/Common/ROESY |
| b) Sample Details | Kontes NMR tube, 8 in, Temperature @ 298 K, No Spinning |
| c) Instrument operation details | Radiation frequency (MHz):<br>$^1$H: 499.907<br><br>$^{13}$C: 125.713<br><br>COSY: 499.700 (F2), 499.700 (F1)<br>HSQC: 499.701 (F2), 125.661 (F1)<br>HMBC: 499.701 (F2), 125.662 (F1)<br>*J*-resolved: 499.906 (F2), 499.906 (F1)<br>TOCSY: 499.906 (F2), 499.906 (F1)<br>ROESY: 499.906 (F2), 499.906 (F1)<br>$^1$H: 90° = 7.80 μs, $^{13}$C: 90° = 9.50 μs |

**Table 3.12.** (continued)

| | |
|---|---|
| d) Number of scans | $^1$H: 512<br><br>$^{13}$C: 4800<br><br>COSY: 4<br><br>HSQC: 4<br><br>HMBC: 8<br><br>*J*-resolved: 8<br><br>TOCSY: 8<br><br>ROESY: 16 |
| e) Number of data points acquired (F2, F1) | $^1$H: 16384<br><br>$^{13}$C: 65536<br><br>COSY: 864, 800<br><br>HSQC: 1500, 512<br><br>HMBC: 1500, 512<br><br>*J*-resolved: 3584, 400<br><br>TOCSY: 860, 512<br><br>ROESY: 860, 400 |
| e) Acquisition time (s) | $^1$H: 1.6384<br><br>$^{13}$C: 1.2321<br><br>COSY: 0.15<br><br>HSQC: 0.15<br><br>HMBC: 0.15<br><br>*J*-resolved: 0.6250<br><br>TOCSY: 0.15<br><br>ROESY: 0.15 |
| FID and spectral processing parameters: | |
| Processing software | MestReNova v12.0.0-20080 |
| a) Number of data points in spectrum (F2, F1) | $^1$H: 131072<br><br>$^{13}$C: 131072<br><br>COSY: 1024, 1024<br><br>HSQC: 2048, 2048<br><br>HMBC: 2048, 2048<br><br>*J*-resolved: 4096, 4096<br><br>TOCSY: 1024, 1024<br><br>ROESY: 1024, 1024 |

**Table 3.12.** (continued)

| | |
|---|---|
| b) Window function details | [1]H: none<br><br>[13]C: exponential, LB = 0.5 Hz<br><br>COSY: F1 Sine Square II, 0%, 50%; F2 Sine Square II, 0%, 50%<br><br>HSQC: F1 Gaussian 56.43; F2 Gaussian 7.65<br><br>HMBC: F1 Gaussian 67.70; F2 Sine Square II, 0%, 50%<br><br>*J*-resolved: F1 Sine Square II, 0%, 50%; F2 Gaussian 1.84, Sine Bell II -0.3%, 50%; First Point 0.51<br><br>TOCSY: F1 Gaussian 17.99; F2 Gaussian 7.65; First Point 0.50<br><br>ROESY: F1 Gaussian 8.24; F2 Gaussian 7.65; First Point 0.50 |

**Table 3.13.** Varian Inova 600 MHz NMR Instrument Metadata

| | |
|---|---|
| Facility Supervisor | Dr. Daniel Holmes |
| Analyst | Steven M. Hurney |
| Instrument Location | MSU Max T. Rogers NMR Facility |
| Facility Instrument Titles | Sobek |
| Time of acquisition | March 2016 |
| Manufacturer | Varian |
| Field Frequency Lock | Acetonitrile-d$_3$ |
| Additional Solute | None |
| Solvent | CD$_3$CN: 300 μL |
| Chemical Shift Standard | CH$_3$CN-d$_3$ ($\delta_H$ = 1.94 and $\delta_C$ = 118.70 ppm) |
| Concentration Standard | None |
| Instrument | Varian Inova 600 MHz |
| Geographic location of instrument | 42.7288, -84.4745 |
| Magnet | 599.77 MHz |
| Probe | Nalorac 5mm PFG switchable probe pretuned for 1H, 13C |
| Acquisition Software | CentOS 5.6 with VnmrJ 3.2A |
| Acquisition Parameters: | |
| a) Acquisition parameters file reference | 1D-TOCSY: VnmrJ/Experiment Selector/Liquid/TOCSY1D |
| b) Sample Details | Shigemi (5 mm) NMR tube, Temperature @ 298 K, No Spinning |
| c) Instrument operation details | Radiation frequency (MHz):<br><br>$^1$H: 599.77<br><br>$^1$H: 90° = 16.15 μs<br>Selective band center: 3.89 ppm; width: 61.6 Hz<br>Selective band center: 1.49 ppm; width: 53.6 Hz<br>Selective band center: 1.63 ppm; width: 53.2 Hz |
| d) Number of scans | $^1$H: 552 |
| e) Number of data points acquired | $^1$H: 32764 |
| e) Acquisition time (s) | $^1$H: 1.7074 |
| FID and spectral processing parameters: | |
| Processing software | MestReNova v12.0.0-20080 |
| a) Number of data points in spectrum | $^1$H: 131072 |
| b) Window function details | $^1$H: none |

**Table 3.14.** Summary of $^1$H chemical shifts of inositol core hydrogen atoms. Chemical shifts labeled in bold indicate acyl substitutions are located at those positions. All spectra were referenced to non-deuterated solvent signal of acetonitrile-d$_3$ ($\delta_H$ = 1.94 ppm).

| Acylinositol ID | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_{1'}$ | $R_{2'}$ | $R_{3'}$ | $R_{4'}$ | $R_{5'}$ | $R_{6'}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAG-I3:22:0(2,10,10) | **4.78** | **5.41** | **4.93** | 3.80 | 3.44 | 3.75 | 4.50 | 3.49 | 3.44 | 3.30 | 3.33 | 3.78, 3.61 |
| G-I3:22:0(2,10,10) | **4.82** | **5.47** | **4.96** | 3.78 | 3.44 | 3.77 | 4.33 | 3.11 | 3.27 | 3.27 | 3.33 | 3.77, 3.61 |
| X-I3:22:0(2,10,10) | **4.80** | **5.47** | **4.94** | 3.79 | 3.40 | 3.75 | 4.28 | 3.11 | 3.27 | 3.47 | 3.90, 3.21 | --- |
| NAG-I3:24:0(2,10,12) | **4.78** | **5.41** | **4.93** | 3.80 | 3.44 | 3.75 | 4.50 | 3.49 | 3.43 | 3.29 | 3.33 | 3.78, 3.61 |
| G-I3:24:0(2,10,12) | **4.82** | **5.47** | **4.96** | 3.78 | 3.44 | 3.76 | 4.33 | 3.10 | 3.27 | 3.27 | 3.32 | 3.77, 3.59 |
| I3:22:0(2,10,10) | **4.76** | **5.44** | **4.78** | 3.68 or 3.69 | 3.30 | 3.68 or 3.69 | --- | --- | --- | --- | --- | --- |
| I4:24:0(2,2,10,10) | **4.81** | **5.49** | **4.97** | **5.21** | 3.50 | 3.77 | --- | --- | --- | --- | --- | --- |
| I3:24:0(2,10,12) | **4.75** | **5.44** | **4.77** | 3.68 | 3.30 | 3.68 | --- | --- | --- | --- | --- | --- |
| I4:26:0(2,2,10,12) | **4.81** | **5.49** | **4.97** | **5.21** | 3.50 | 3.77 | --- | --- | --- | --- | --- | --- |



myo-inositol

4-O-β-N-acetylglucosaminyl myo-inositol

4-O-β-glucopyranosyl myo-inositol

4-O-β-xylopyranosyl myo-inositol

311
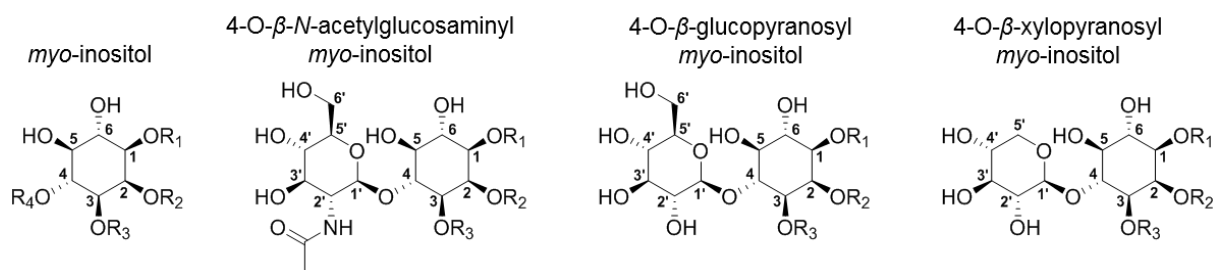
**Table 3.15.** Summary of $^{13}C$ chemical shifts of inositol core carbon atoms. Chemical shifts labeled in bold indicate acyl substitutions are located at those positions. All spectra were referenced to non-deuterated solvent signal of acetonitrile-d$_3$ ($\delta_C$ = 118.70 ppm).

| Acylinositol ID | R$_1$ | R$_2$ | R$_3$ | R$_4$ | R$_5$ | R$_6$ | R$_{1'}$ | R$_{2'}$ | R$_{3'}$ | R$_{4'}$ | R$_{5'}$ | R$_{6'}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAG-I3:22:0(2,10,10) | **72.25** | **69.43** | **71.33** | 81.50 | 73.91 | 71.81 | 102.40 | 57.46 | 75.63 | 72.06 | 77.72 | 62.78 |
| G-I3:22:0(2,10,10) | **72.07** | **69.19** | **70.79** | 82.86 | 73.70 | 71.51 | 104.69 | 74.36 | 77.40 | 71.05 | 77.58 | 62.42 |
| X-I3:22:0(2,10,10) | **72.43** | **69.53** | **71.34** | 81.82 | 73.79 | 71.90 | 105.53 | 74.37 | 77.61 | 70.75 | 66.91 | --- |
| NAG-I3:24:0(2,10,12) | **72.26** | **69.44** | **71.32** | 81.57 | 73.91 | 71.83 | 102.38 | 57.47 | 75.65 | 72.08 | 77.73 | 62.82 |
| G-I3:24:0(2,10,12) | **72.38** | **69.50** | **71.09** | 83.19 | 74.00 | 71.83 | 104.98 | 74.69 | 77.72 | 71.37 | 77.91 | 62.74 |
| I3:22:0(2,10,10) | **72.84** | **69.61** | **72.53** | 71.80 or 71.75 | 75.69 | 71.80 or 71.75 | --- | --- | --- | --- | --- | --- |
| I4:24:0(2,2,10,10) | **72.47** | **69.42** | **70.36** | **72.66** | 73.50 | 71.86 | --- | --- | --- | --- | --- | --- |
| I3:24:0(2,10,12) | **72.84** | **69.61** | **72.53** | 71.80 or 71.75 | 75.70 | 71.80 or 71.75 | --- | --- | --- | --- | --- | --- |
| I4:26:0(2,2,10,12) | **72.48** | **69.42** | **70.36** | **72.67** | 73.52 | 71.87 | --- | --- | --- | --- | --- | --- |



*myo*-inositol

4-O-*β*-*N*-acetylglucosaminyl *myo*-inositol

4-O-*β*-glucopyranosyl *myo*-inositol

4-O-*β*-xylopyranosyl *myo*-inositol
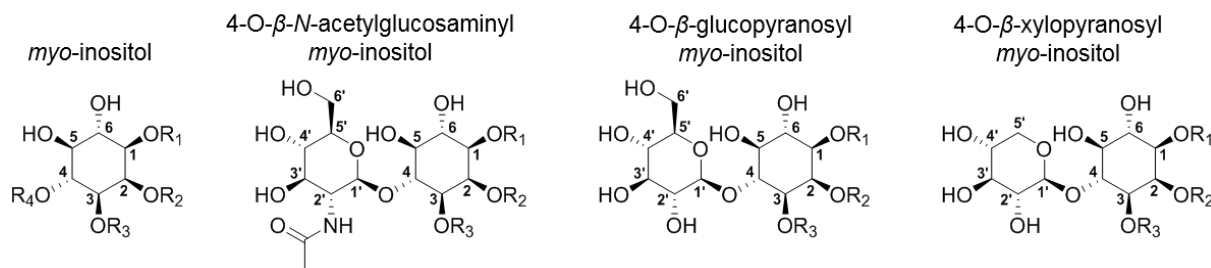
**Table 3.16.** NAG-I3:22:0(2,10,10) Chemical shifts and coupling constants
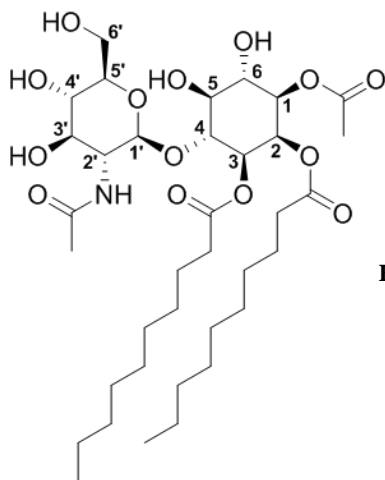


**Molecular Formula:** $C_{36}H_{63}NO_{14}$

**110 min Retention Time (ESI-):** 58.00 mins

**HRMS:** (ESI-) *m/z* calculated for $C_{37}H_{64}NO_{16}^-$
([M+HCOO]$^-$): 778.4231, found: 778.4241
**110 min Retention Time (ESI+):** 57.98 mins

**HRMS:** (ESI+) *m/z* calculated for $C_{36}H_{64}NO_{14}^+$ ([M+H]$^+$):
734.4321, found: 734.4296

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #16
**Sample mass for NMR analysis:** 2.3 mg
**NMR Solvent:** $D_3CN$

**InChi Key:** CSNOFVDMFKBZIV-SHTLPGHTSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.78 (dd, *J* = 10.3, 2.9 Hz) | 72.25 |
| -1(CO) | | 171.45 |
| -2(CH$_3$) | 1.96 | 21.36 |
| **2**(CH) | 5.41 (t, *J* = 2.9 Hz) | 69.43 |
| -1(CO) | | 174.1[a] |
| -2(CH$_2$) | 2.37 (m) | 35.0[b] |
| -3(CH$_2$) | 1.61 (p, *J* = 7.3 Hz) | 26.26 |
| -4 to 9(CH$_2$) | 1.36-1.25 (m) | 30.7-30.1[c], 33.00[d], 23.78[e] |
| -10(CH$_3$) | 0.89 (t, *J* = 7.2 Hz) | 14.78[f] |
| **3**(CH) | 4.93 (dd, *J* = 10.2, 3.0 Hz) | 71.33 |
| -1(CO) | | 174.1[a] |
| -2(CH$_2$) | 2.33 (m), 2.27 (m) | 35.0[b] |
| -3(CH$_2$) | 1.54 (m), 1.50 (m) | 25.90 |
| -4 to 9(CH$_2$) | 1.36-1.25 (m) | 30.6-30.1[c], 33.00[d], 23.78[e] |
| -10(CH$_3$) | 0.88 (t, *J* = 7.2 Hz) | 14.78[f] |
| **4**(CH) | 3.80 (dd, *J* = 10.4, 8.8 Hz) | 81.50 |
| **5**(CH) | 3.44 (t, *J* = 9.1 Hz) | 73.91 |
| **6**(CH) | 3.75 (dd, *J* = 10.5, 8.8 Hz) | 71.81 |
| **1'**(CH) | 4.50 (d, *J* = 8.3 Hz) | 102.40 (H-C, *J* = 163 Hz)[g] |
| **2'**(CH) | 3.49 (m) | 57.46 |
| NH | 6.39 (d, *J* = 8.4 Hz) | --- |
| -1(CO) | | 171.55 |
| -2(CH$_3$) | 1.82 | 23.82 |

**Table 3.16.** (continued)

| | | |
|---|---|---|
| **3'**(CH) | 3.44 (t, $J$ = 9.1 Hz) | 75.63 |
| **4'**(CH) | 3.30 (dd, $J$ = 10.0, 8.1 Hz) | 72.06 |
| **5'**(CH) | 3.33 (ddd, $J$ = 10.2, 5.8, 2.8 Hz) | 77.72 |
| **6'**(CH$_2$) | 3.78 (dd, $J$ = 11.8, 2.9 Hz), 3.61 (dd, $J$ = 11.8, 5.8 Hz) | 62.78 |

a - Two $^{13}$C signals not resolved in 2D spectra (174.14, 174.12 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (35.05, 34.99 ppm)

c - $^{13}$C signals for CH$_2$ carbon positions 4 to 7 (30.63, 30.56, 30.47, 30.45, 30.44, 30.44, 30.12, 30.08 ppm)

d - Overlapping $^{13}$C signals for CH$_2$ carbon position 8

e - Overlapping $^{13}$C signals for CH$_2$ carbon position 9

f - Overlapping $^{13}$C signals for CH$_3$ carbon position 10

g - $^{1}J_{CH}$ determined from HMBC breakthrough signal
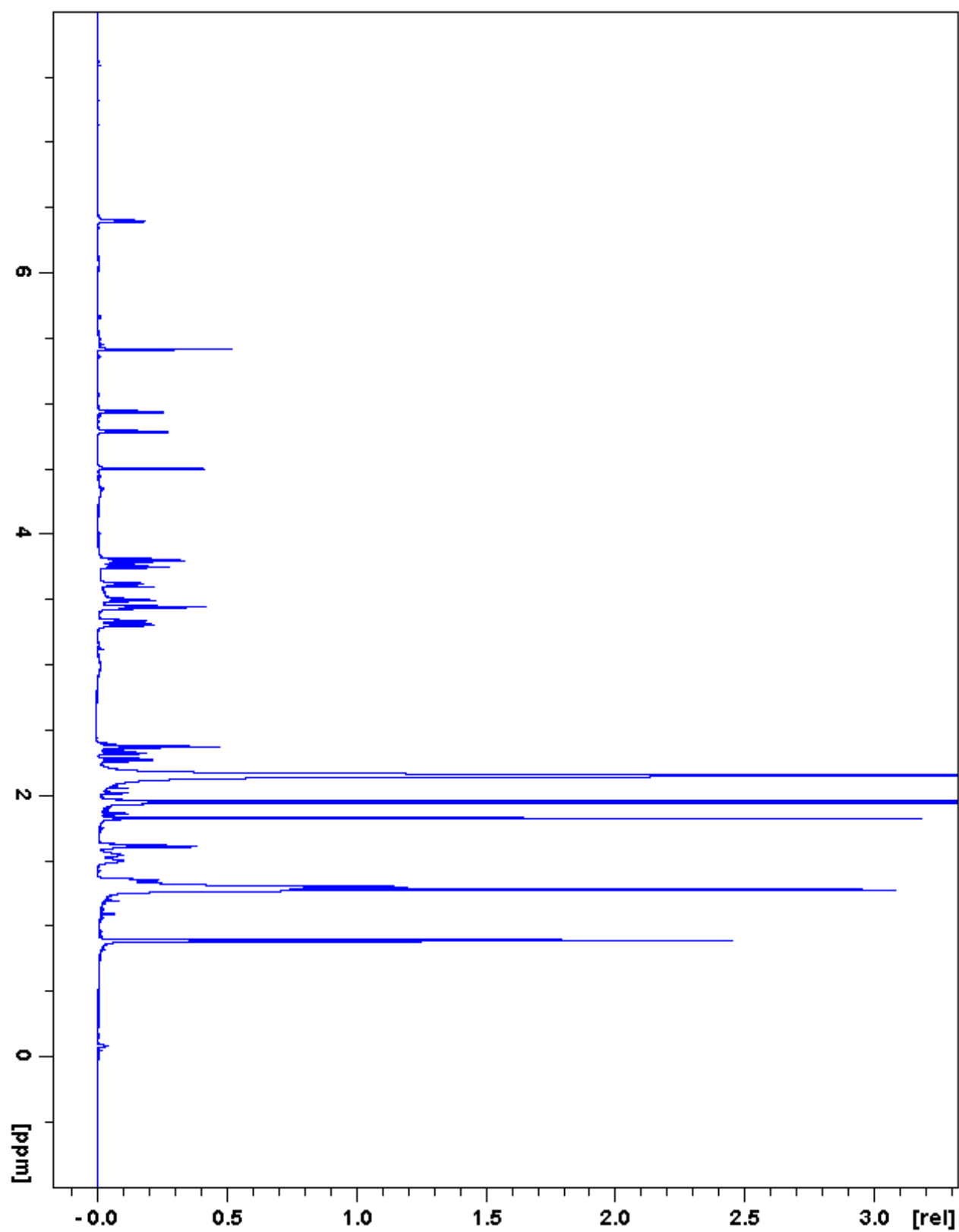
**Figure 3.26.** NAG-I3:22:0(2,10,10) $^1$H NMR
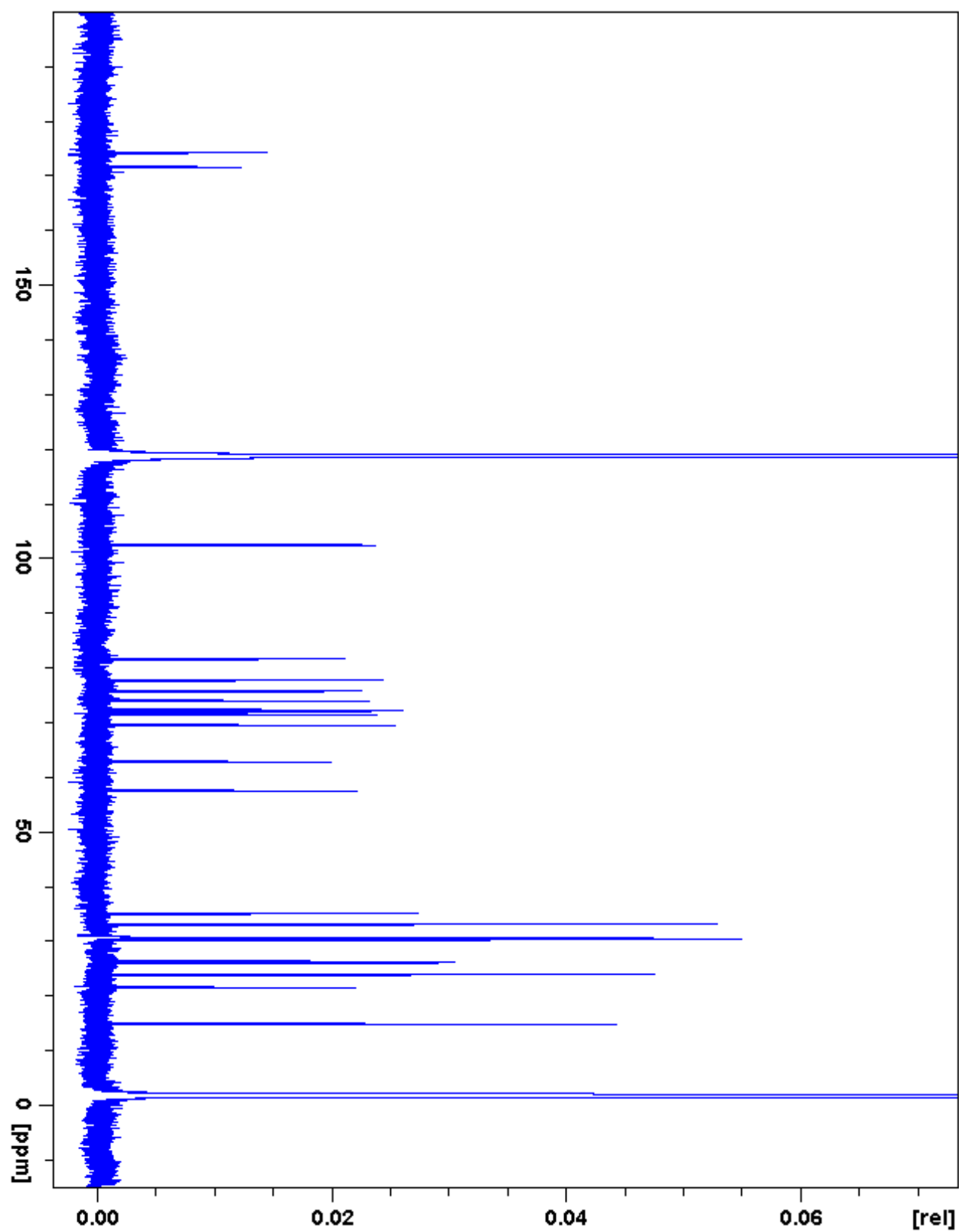
**Figure 3.27**. NAG-I3:22:0(2,10,10) $^{13}C$ NMR
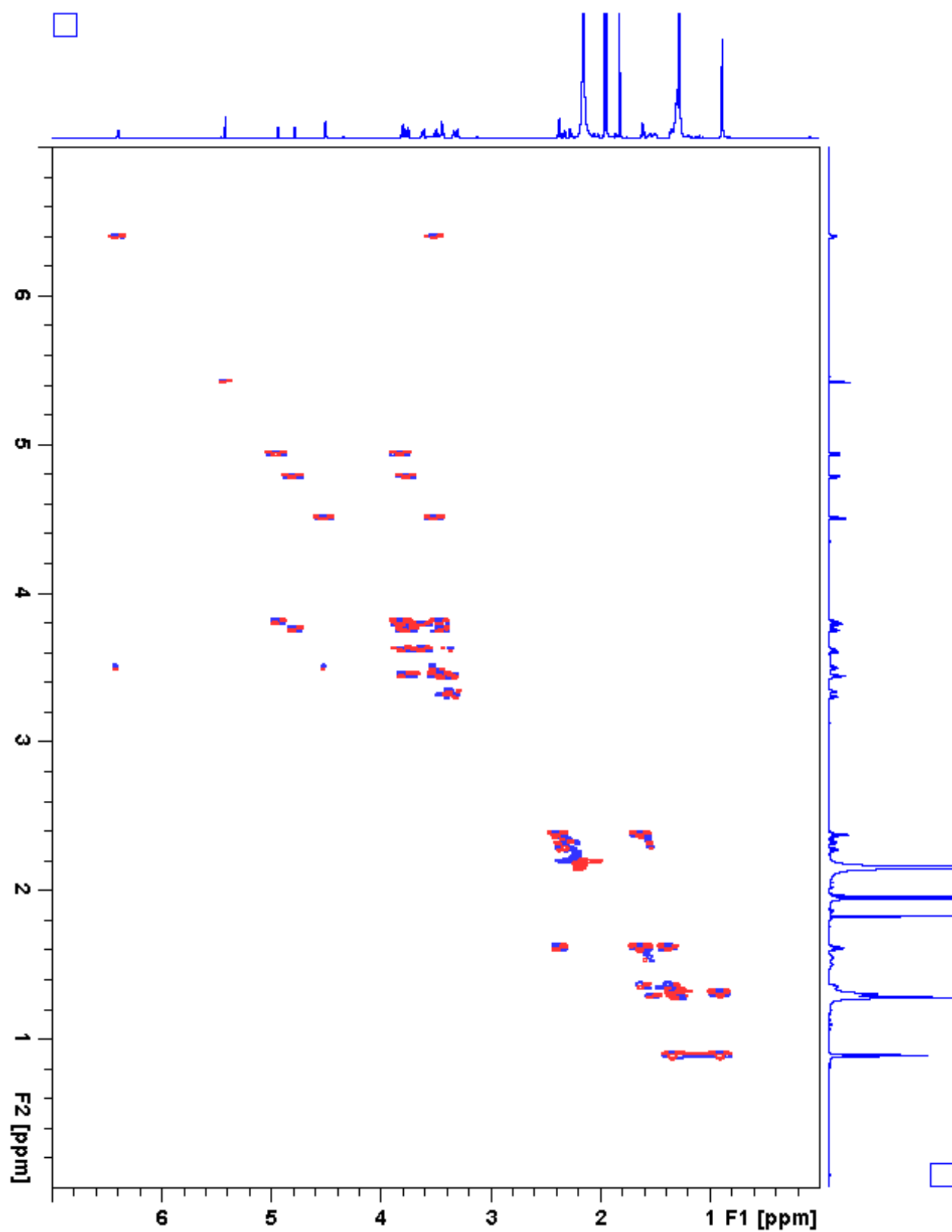
**Figure 3.28.** NAG-I3:22:0(2,10,10) $^1$H-$^1$H gCOSY
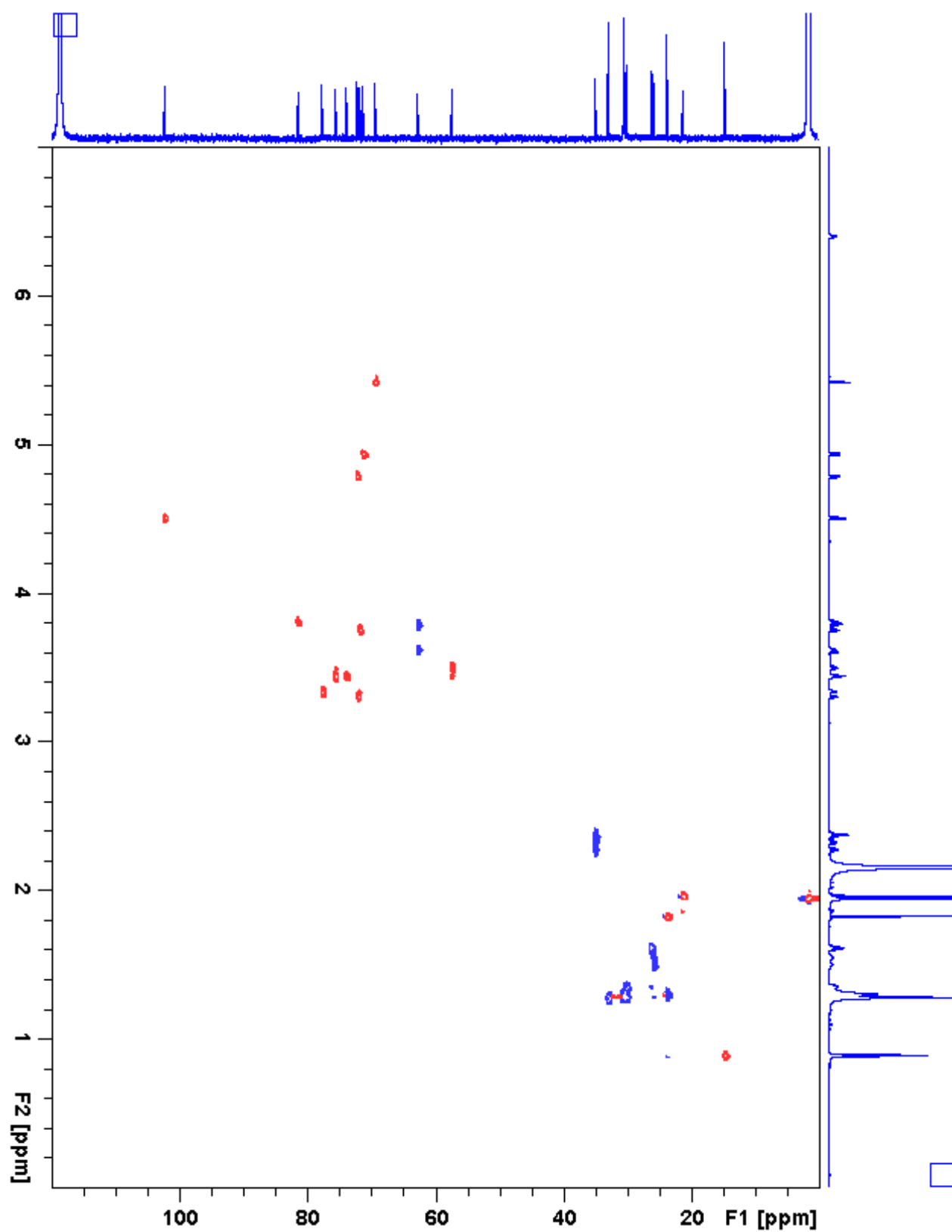
**Figure 3.29.** NAG-I3:22:0(2,10,10) gHSQC

318
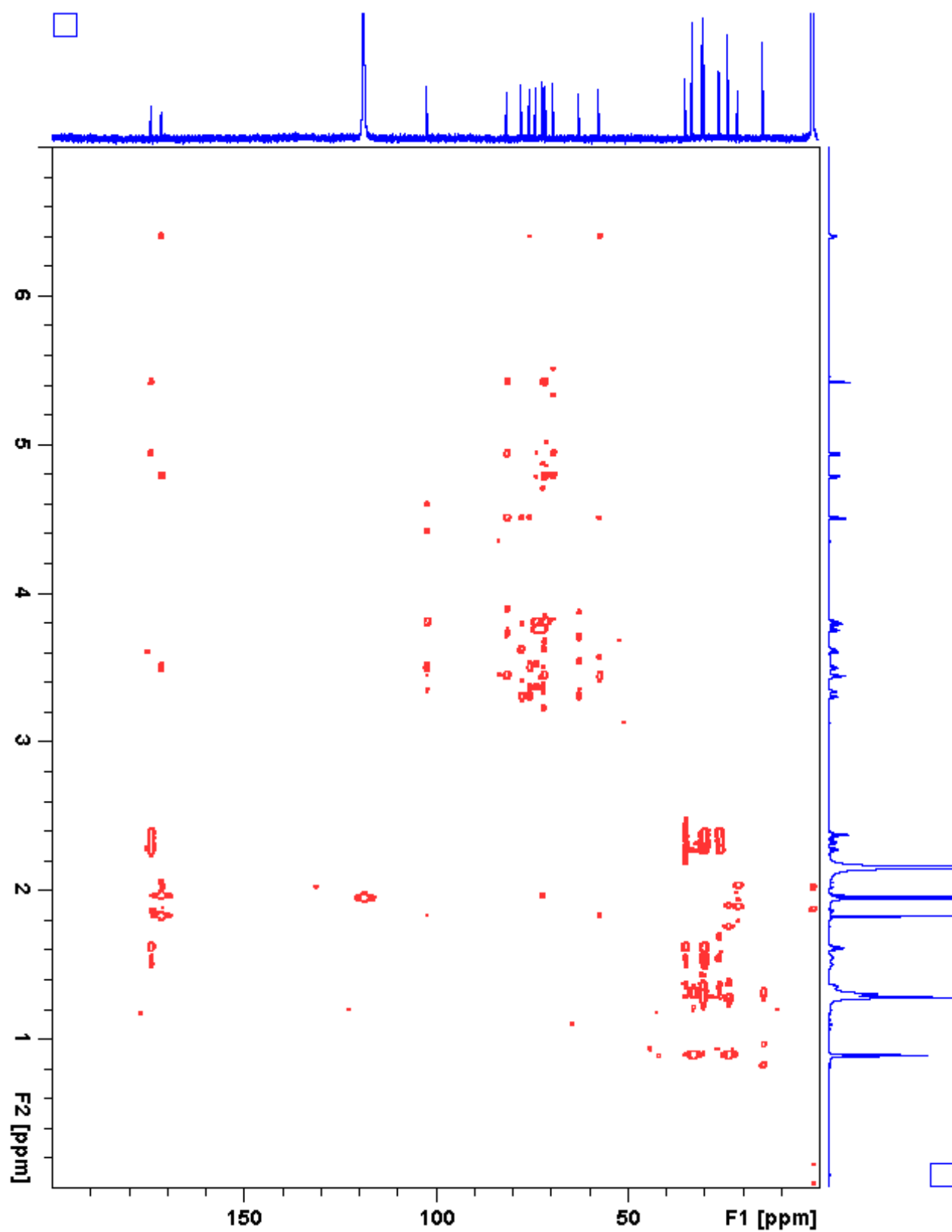
**Figure 3.30.** NAG-I3:22:0(2,10,10) gHMBC
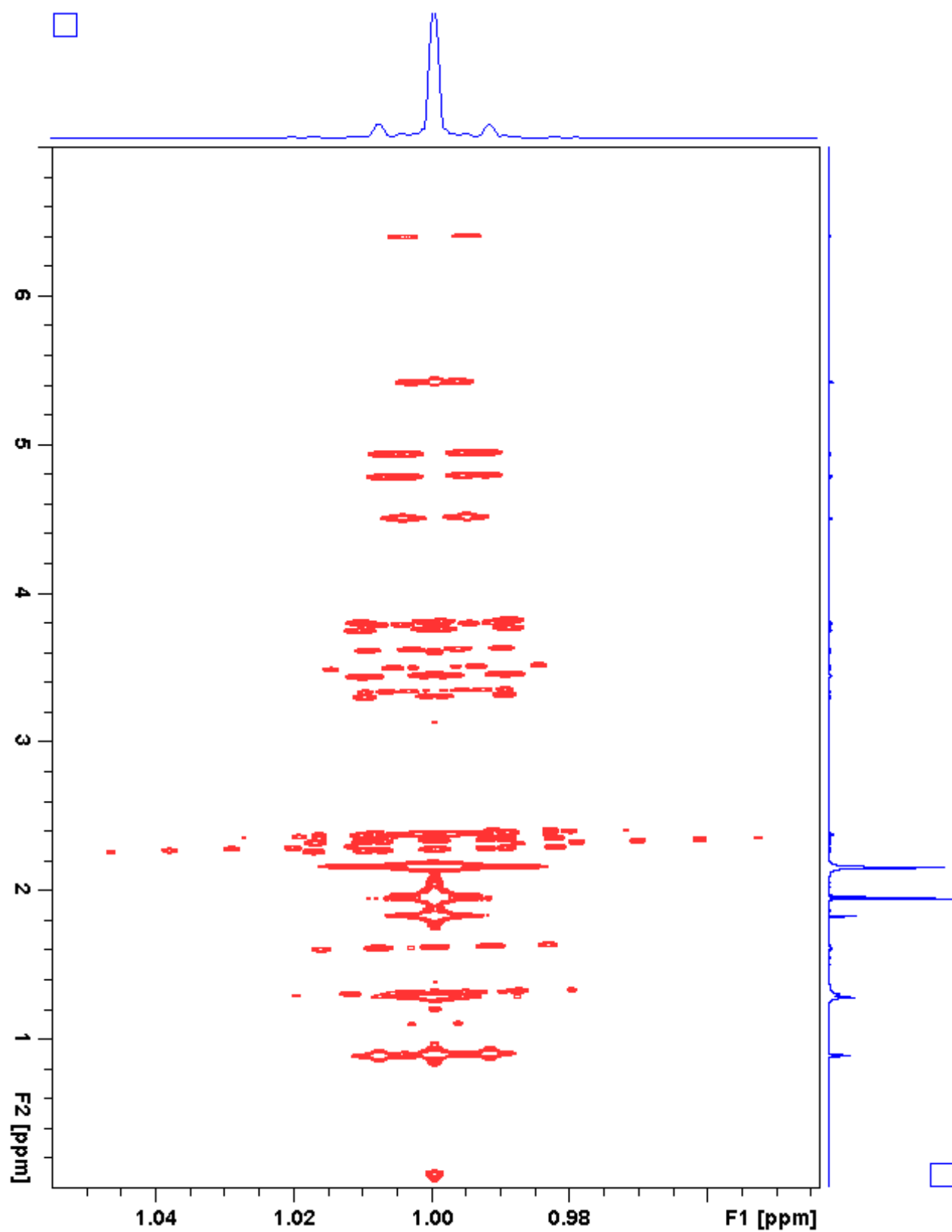
**Figure 3.31.** NAG-I3:22:0(2,10,10) *J*-resolved

**Figure 3.32.** NAG-I3:22:0(2,10,10) ROESY

**Table 3.17.** G-I3:22:0(2,10,10) Chemical shifts and coupling constants
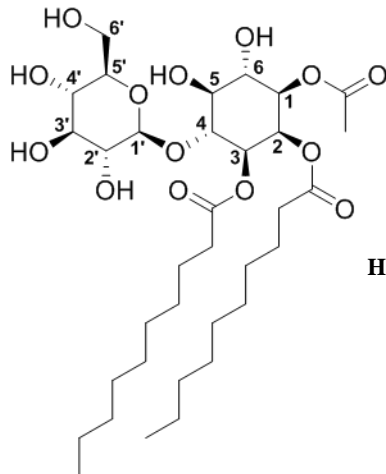


**Molecular Formula:** $C_{34}H_{60}O_{14}$

**110 min Retention Time (ESI-):** 58.83 mins
**HRMS:** (ESI-) *m/z* calculated for $C_{35}H_{61}O_{16}^-$
([M+HCOO]⁻): 737.3965, found: 737.3965

**110 min Retention Time (ESI+):** 58.82 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{34}H_{64}NO_{14}^+$ ([M+NH4]⁺):
710.4321, found: 710.4291

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #18
**Sample mass for NMR analysis:** 3.3 mg
**NMR Solvent:** $D_3CN$
**InChi Key:** LCZLGZXAZPZJHW-QWUQPGGFSA-N

| Carbon # (group) | ¹H (ppm) | ¹³C (ppm) |
|---|---|---|
| **1**(CH) | 4.82 (dd, *J* = 10.3, 2.9 Hz) | 72.07 |
| -1(CO) | | 171.18 |
| -2(CH₃) | 1.97 | 21.06 |
| **2**(CH) | 5.47 (t, *J* = 2.9 Hz) | 69.19 |
| -1(CO) | | 173.79 |
| -2(CH₂) | 2.37 (t, *J* = 7.2 Hz) | 34.71[a] |
| -3(CH₂) | 1.61 (p, *J* = 7.3 Hz) | 25.99 |
| -4 to 9(CH₂) | 1.37-1.24 (m) | 30.3-29.8[b], 32.71[c], 23.47[d] |
| -10(CH₃) | 0.89 (t, *J* = 7.2 Hz) | 14.48[e] |
| **3**(CH) | 4.96 (dd, *J* = 10.2, 3.0 Hz) | 70.79 |
| -1(CO) | | 173.85 |
| -2(CH₂) | 2.24 (m) | 34.71[a] |
| -3(CH₂) | 1.52 (m) | 25.38 |
| -4 to 9(CH₂) | 1.37-1.24 (m) | 30.3-29.8[b], 32.71[c], 23.47[d] |
| -10(CH₃) | 0.88 (t, *J* = 7.2 Hz) | 14.48[e] |
| **4**(CH) | 3.78 (dd, *J* = 10.5, 8.8 Hz) | 82.86 |
| **5**(CH) | 3.44 (t, *J* = 9.1 Hz) | 73.70 |
| **6**(CH) | 3.77 (dd, *J* = 10.6, 8.8 Hz) | 71.51 |
| **1'**(CH) | 4.33 (d, *J* = 7.9 Hz) | 104.69 (H-C, *J* = 162 Hz)[f] |
| **2'**(CH) | 3.11 (m) | 74.36 |
| **3'**(CH) | 3.27 (m) | 77.40 |
| **4'**(CH) | 3.27 (m) | 71.05 |
| **5'**(CH) | 3.33 (ddd, *J* = 9.8, 5.8, 2.7 Hz) | 77.58 |

**Table 3.17.** (continued)

| 6'(CH$_2$) | 3.77 (dd, $J = 11.9$, 2.8 Hz), 3.61 (dd, $J = 11.9$, 5.9 Hz) | 62.42 |
|---|---|---|

a - Overlapping $^{13}$C signals

b - $^{13}$C signals for CH$_2$ carbon positions 4 to 7 (30.34, 30.27, 30.16, 30.16, 30.14, 30.13, 29.82, 29.76 ppm)

c - Overlapping $^{13}$C signals for CH$_2$ carbon position 8

d - Overlapping $^{13}$C signals for CH$_2$ carbon position 9

e - Overlapping $^{13}$C signals for CH$_3$ carbon position 10

f - $^1J_{CH}$ determined from HMBC breakthrough signal

**Figure 3.33.** G-I3:22:0(2,10,10) ^{1}H NMR
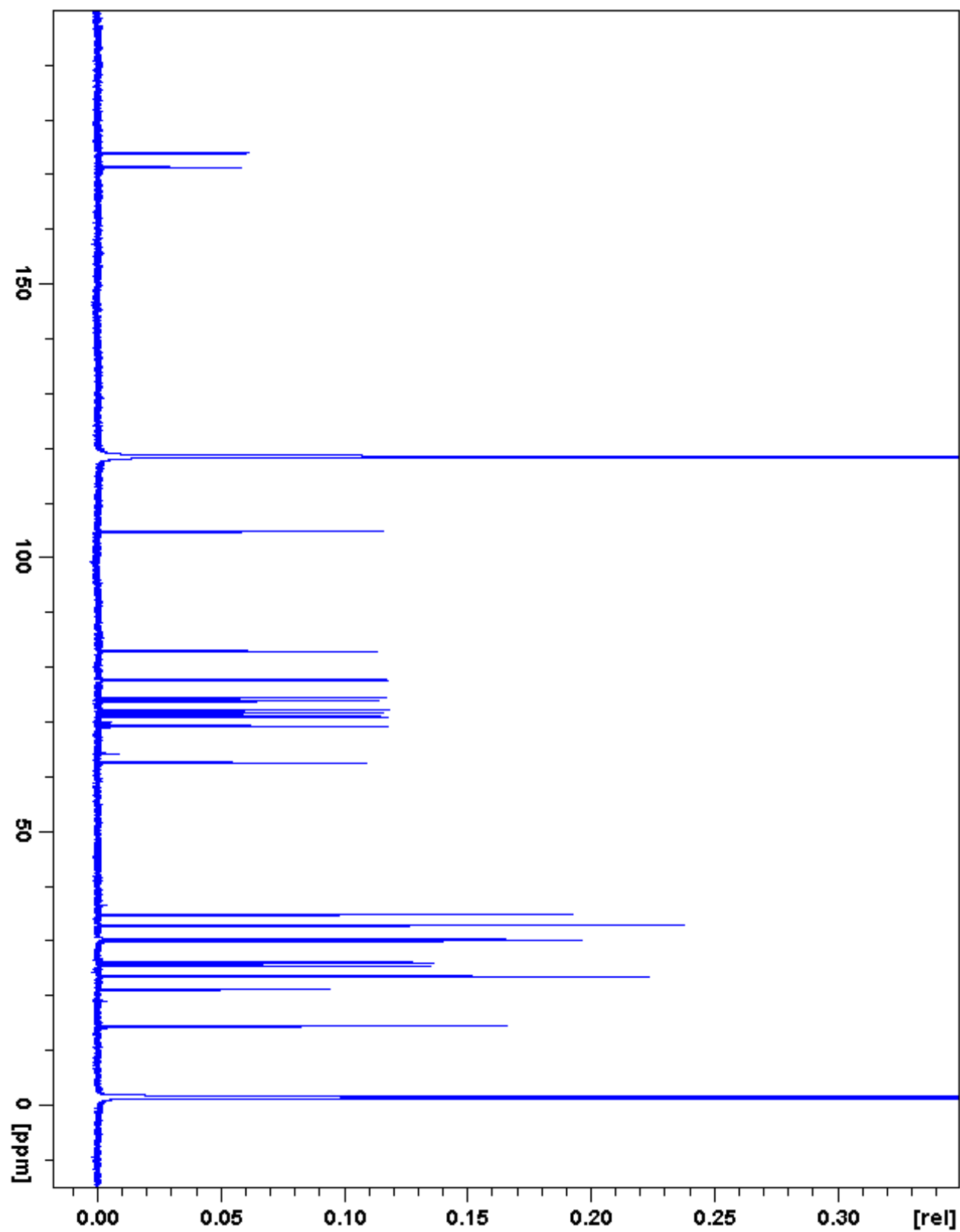
**Figure 3.34.** G-I3:22:0(2,10,10) ^{13}C NMR
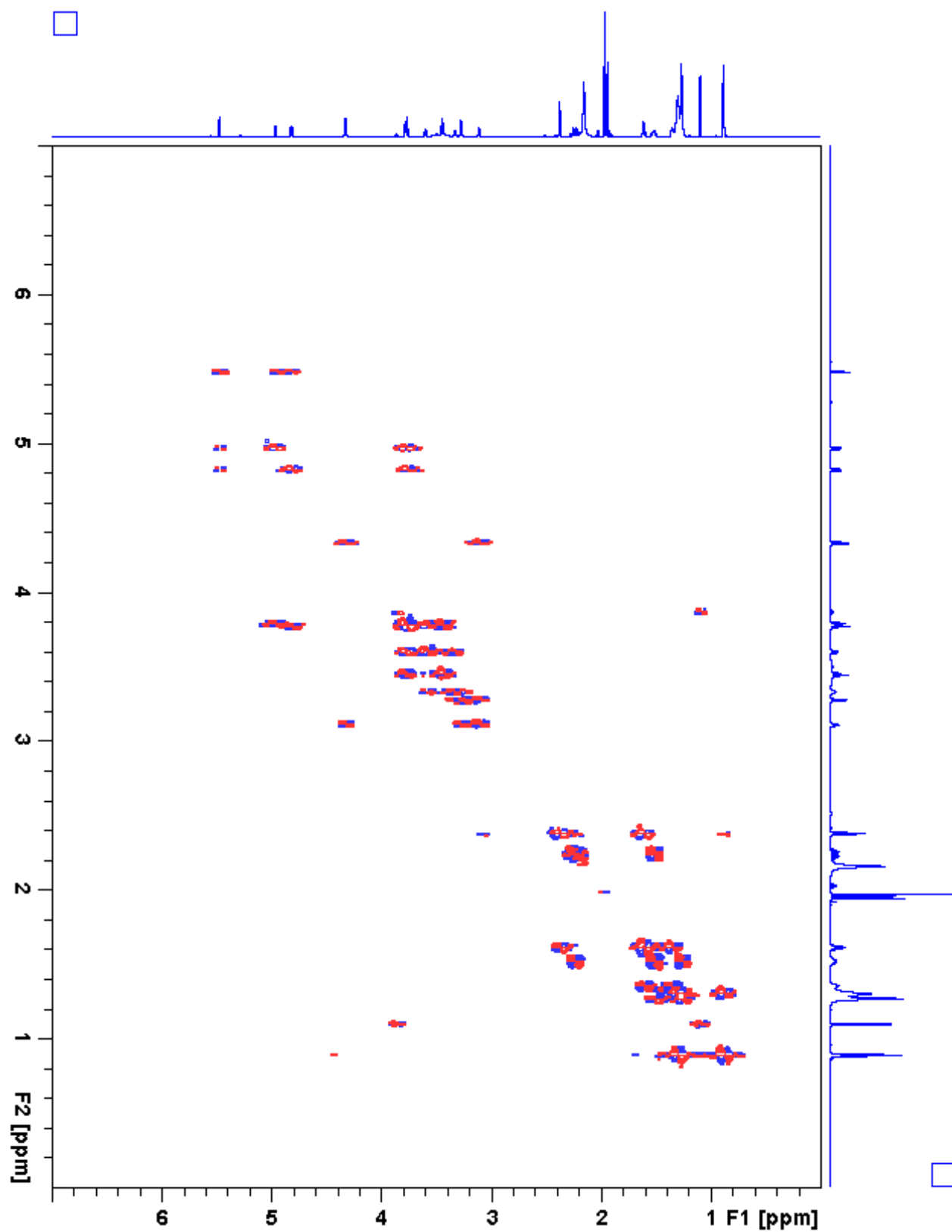
**Figure 3.35.** G-I3:22:0(2,10,10) $^1$H-$^1$H gCOSY

**Figure 3.36**. G-I3:22:0(2,10,10) gHSQC

327
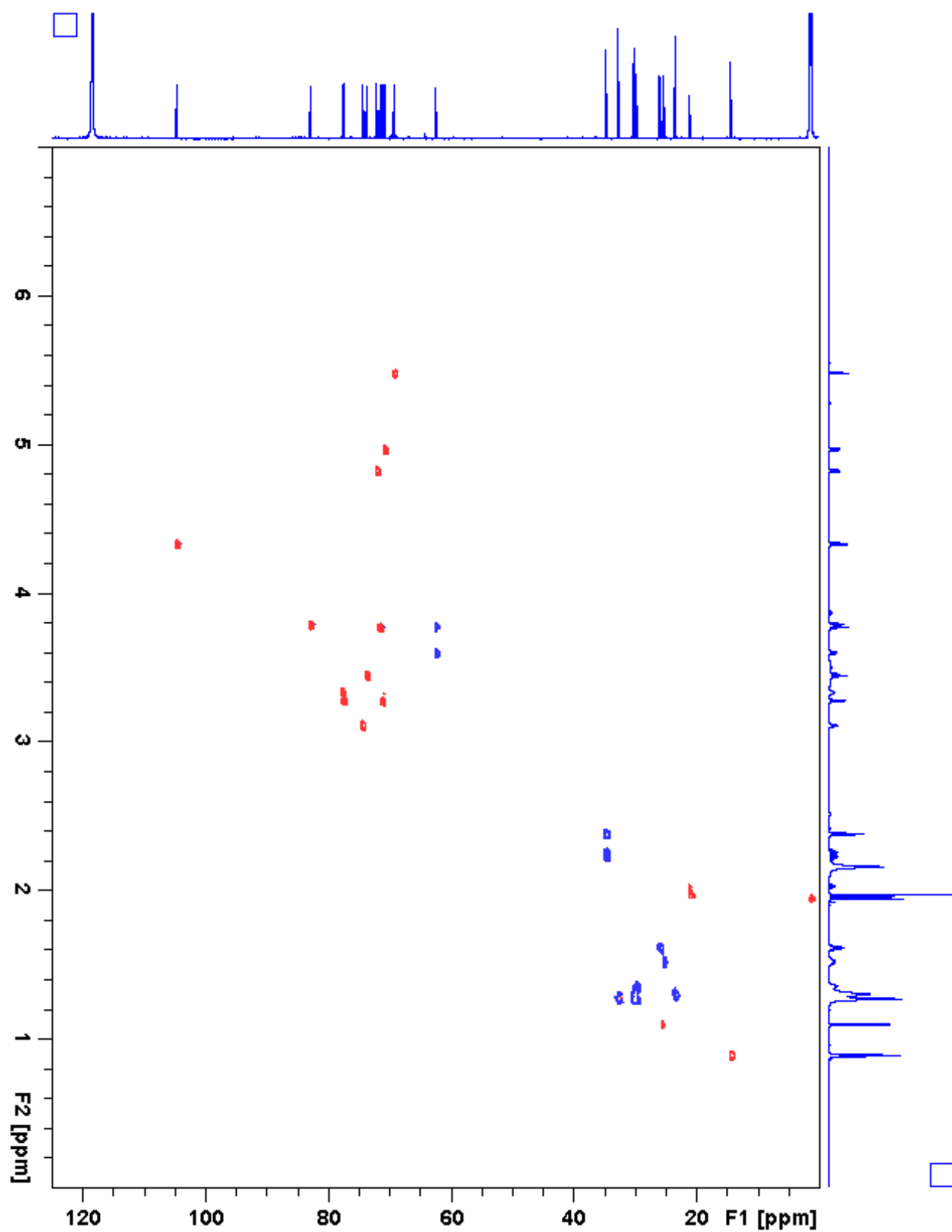
**Figure 3.37**. G-I3:22:0(2,10,10) gHMBC
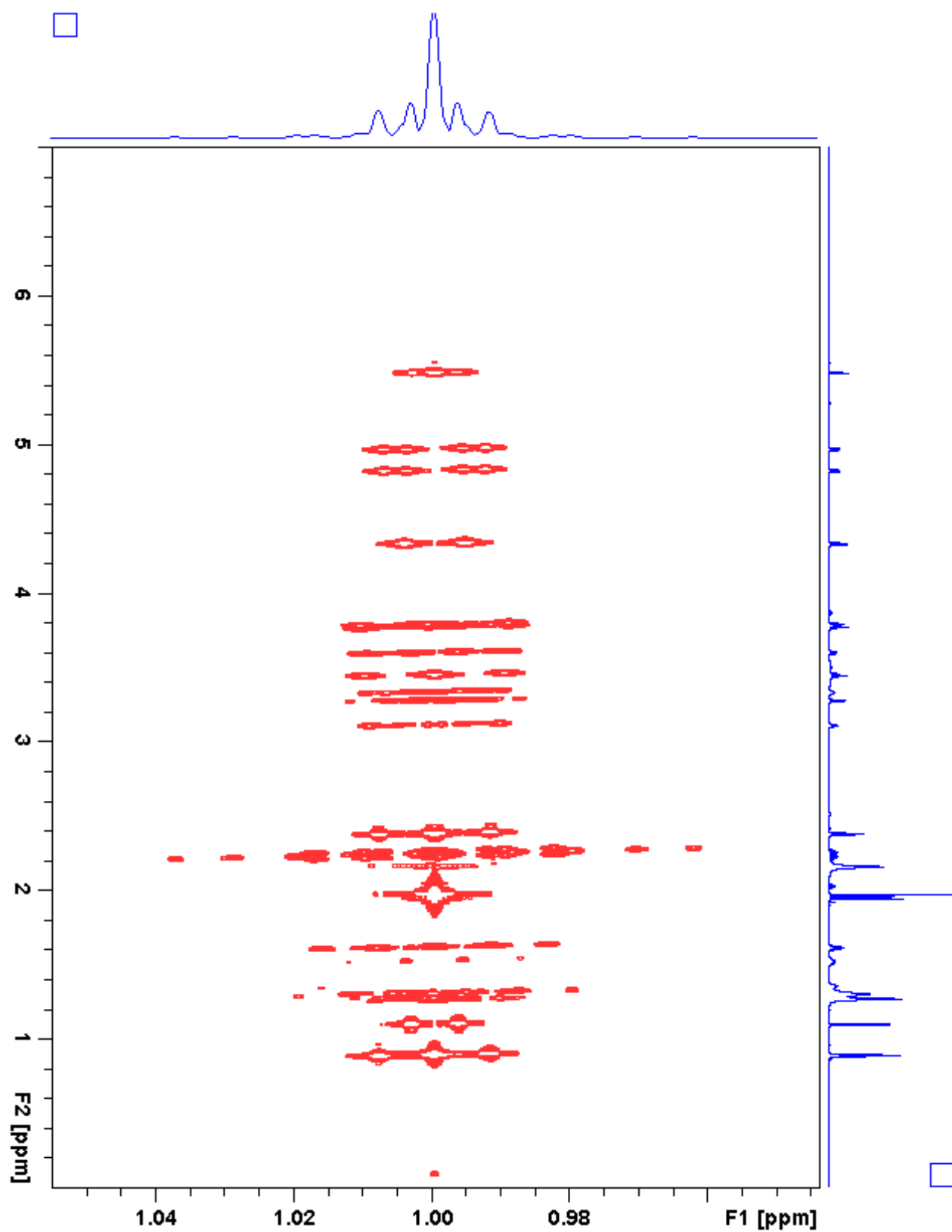
**Figure 3.38**. G-I3:22:0(2,10,10) *J*-resolved

**Figure 3.39**. G-I3:22:0(2,10,10) ROESY

**Table 3.18.** X-I3:22:0(2,10,10) Chemical shifts and coupling constants
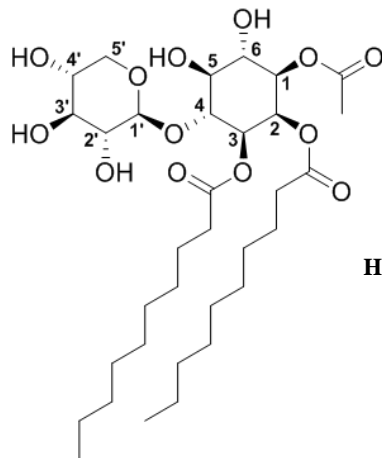


**Molecular Formula:** $C_{33}H_{58}O_{13}$

**110 min Retention Time (ESI-):** 61.28 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{34}H_{59}O_{15}^-$
([M+HCOO]$^-$): 707.3859, found: 707.3864

**110 min Retention Time (ESI+):** 61.25 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{33}H_{62}NO_{13}^+$ ([M+NH$_4$]$^+$):
680.4216, found: 680.4176

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #21
**Sample mass for NMR analysis:** 1.3 mg
**NMR Solvent:** $D_3CN$
**InChi Key:** LSJLPMMRXZQJNM-OTPYBMEMSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.80 (dd, $J$ = 10.2, 2.9 Hz) | 72.43 |
| -1(CO) | | 171.48 |
| -2(CH$_3$) | 1.96 | 21.37 |
| **2**(CH) | 5.47 (t, $J$ = 2.9 Hz) | 69.53 |
| -1(CO) | | 174.08 |
| -2(CH$_2$) | 2.37 (t, $J$ = 7.2 Hz) | 35.0[a] |
| -3(CH$_2$) | 1.61 (p, $J$ = 7.3 Hz) | 26.32 |
| -4 to 9(CH$_2$) | 1.37-1.23 (m) | 30.7-30.1[b], 33.02[c], 23.79[d] |
| -10(CH$_3$) | 0.89 (t, $J$ = 7.2 Hz) | 14.79[e] |
| **3**(CH) | 4.94 (dd, $J$ = 10.3, 3.0 Hz) | 71.34 |
| -1(CO) | | 174.02 |
| -2(CH$_2$) | 2.24 (m) | 35.0[a] |
| -3(CH$_2$) | 1.52 (m) | 25.69 |
| -4 to 9(CH$_2$) | 1.37-1.23 (m) | 30.7-30.1[b], 33.02[c], 23.79[d] |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.79[e] |
| **4**(CH) | 3.79 (dd, $J$ = 10.5, 8.9 Hz) | 81.82 |
| **5**(CH) | 3.40 (t, $J$ = 9.2 Hz) | 73.79 |
| **6**(CH) | 3.75 (dd, $J$ = 10.5, 8.9 Hz) | 71.90 |
| **1'**(CH) | 4.28 (d, $J$ = 7.5 Hz) | 105.53 (H-C, $J$ = 163 Hz)[f] |
| **2'**(CH) | 3.11 (dd, $J$ = 9.1, 7.4 Hz) | 74.37 |
| **3'**(CH) | 3.27 (t, $J$ = 8.8 Hz) | 77.61 |
| **4'**(CH) | 3.47 (ddd, $J$ = 10.3, 8.4, 5.4 Hz) | 70.75 |

**Table 3.18.** (continued)

| 5'(CH) | 3.90 (dd, $J = 11.4$, 5.3 Hz), 3.21 (dd, $J = 11.6$, 10.0 Hz) | 66.91 |
|---|---|---|

a - Two $^{13}C$ signals not resolved in 2D spectra (35.04, 35.01 ppm)

b - $^{13}C$ signals for $CH_2$ carbon positions 4 to 7 (30.66, 30.58, 30.48, 30.47, 30.45, 30.45, 30.14, 30.06 ppm)

c - Overlapping $^{13}C$ signals for $CH_2$ carbon position 8

d - Overlapping $^{13}C$ signals for $CH_2$ carbon position 9

e - Overlapping $^{13}C$ signals for $CH_3$ carbon position 10

f - $^{1}J_{CH}$ determined from HMBC breakthrough signal

**Figure 3.40.** X-I3:22:0(2,10,10) ^{1}H NMR

**Figure 3.41.** X-I3:22:0(2,10,10) ^{13}C NMR

**Figure 3.42.** X-I3:22:0(2,10,10) $^1$H-$^1$H gCOSY

**Figure 3.43.** X-I3:22:0(2,10,10) gHSQC

**Figure 3.44.** X-I3:22:0(2,10,10) gHMBC
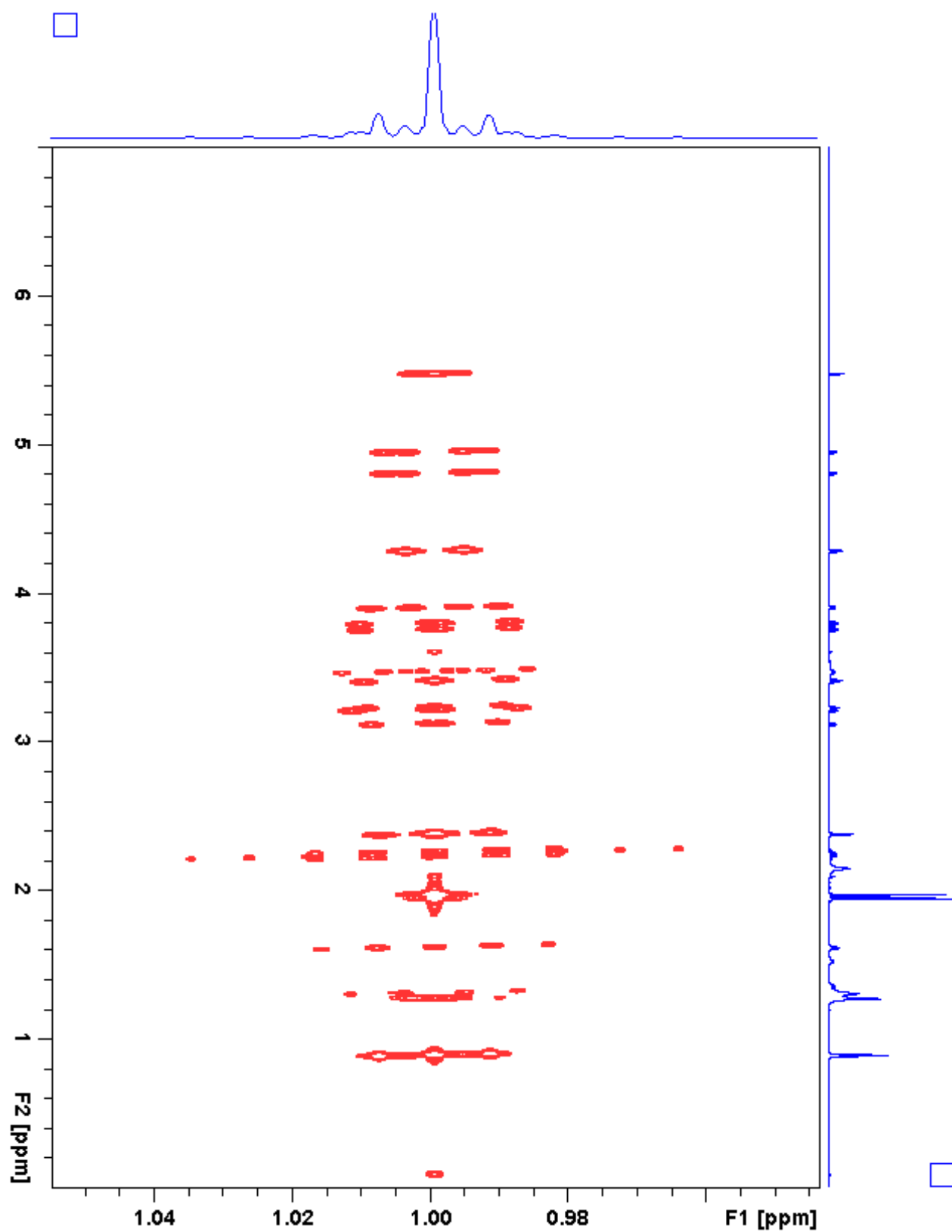
**Figure 3.45.** X-I3:22:0(2,10,10) *J*-resolved

**Figure 3.46.** X-I3:22:0(2,10,10) ROESY

**Table 3.19.** NAG-I3:24:0(2,10,12) Chemical shifts and coupling constants
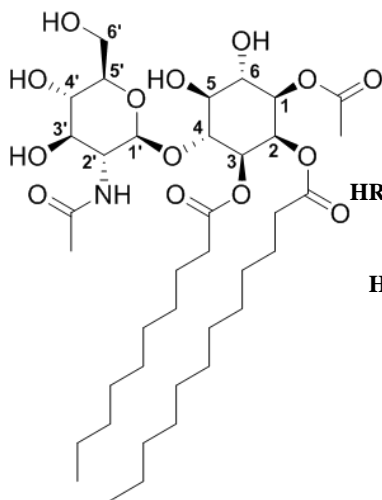


**Molecular Formula:** $C_{38}H_{67}NO_{14}$

**110 min Retention Time (ESI-):** 64.28 mins
**HRMS:** (ESI-) *m/z* calculated for $C_{39}H_{68}NO_{16}^-$ ([M+HCOO]⁻):
806.4544, found: 806.4550

**110 min Retention Time (ESI+):** 64.27 mins
**HRMS:** (ESI+) *m/z* calculated for $C_{38}H_{68}NO_{14}^+$ ([M+H]⁺):
762.4634, found: 762.4595

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #27
**Sample mass for NMR analysis:** 1.0 mg
**NMR Solvent:** $D_3CN$
**InChi Key:** BOJNLBZVYDCKSX-VYHRCMGRSA-N

| Carbon # (group) | ¹H (ppm) | ¹³C (ppm) |
|---|---|---|
| **1**(CH) | 4.78 (dd, *J* = 10.3, 2.9 Hz) | 72.26 |
| -1(CO) | | 171.45 |
| -2(CH₃) | 1.96 | 21.36 |
| **2**(CH) | 5.41 (t, *J* = 2.9 Hz) | 69.44 |
| -1(CO) | | 174.1[a] |
| -2(CH₂) | 2.37 (m) | 35.0[b] |
| -3(CH₂) | 1.61 (p, *J* = 7.3 Hz) | 26.25 |
| -4 to 11(CH₂) | 1.36-1.23 (m) | 30.8-30.1[c], 33.0[d], 23.79[e] |
| -12(CH₃) | 0.88 (t, *J* = 7.2 Hz) | 14.78[f] |
| **3**(CH) | 4.93 (dd, *J* = 10.2, 3.0 Hz) | 71.32 |
| -1(CO) | | 174.1[a] |
| -2(CH₂) | 2.33 (m), 2.27 (m) | 35.0[b] |
| -3(CH₂) | 1.54 (m), 1.50 (m) | 25.91 |
| -4 to 9(CH₂) | 1.36-1.23 (m) | 30.8-30.1[c], 33.0[d], 23.79[e] |
| -10(CH₃) | 0.88 (t, *J* = 7.2 Hz) | 14.78[f] |
| **4**(CH) | 3.80 (dd, *J* = 10.4, 8.8 Hz) | 81.57 |
| **5**(CH) | 3.44 (t, *J* = 9.1 Hz) | 73.91 |
| **6**(CH) | 3.75 (dd, *J* = 10.5, 8.9 Hz) | 71.83 |
| **1'**(CH) | 4.50 (d, *J* = 8.3 Hz) | 102.38 (H-C, *J* = 162.2 Hz)[g] |
| **2'**(CH) | 3.49 (m) | 57.47 |
| NH | 6.38 (d, *J* = 8.3 Hz) | --- |
| -1(CO) | | 171.51 |
| -2(CH₃) | 1.82 | 23.83 |
| **3'**(CH) | 3.43 (dd, *J* = 10.2, 8.5 Hz) | 75.65 |
| **4'**(CH) | 3.29 (dd, *J* = 10.0, 8.2 Hz) | 72.08 |

340

**Table 3.19.** (continued)

| 5'(CH) | 3.33 (ddd, $J$ = 10.0, 5.9, 2.7 Hz) | 77.73 |
|---|---|---|
| 6'(CH$_2$) | 3.78 (dd, $J$ = 11.9, 2.8 Hz), 3.61 (dd, $J$ = 11.9, 5.8 Hz) | 62.82 |

a - Two $^{13}$C signals not resolved in 2D spectra (174.14, 174.13 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (35.07, 34.99 ppm)

c - $^{13}$C signals for CH$_2$ carbon positions 4 to 9, 4 to 7 (30.81, 30.75, 30.68, 30.57, 30.48, 30.45, 30.45, 30.45, 30.13, 30.09 ppm)

d - Two $^{13}$C signals for CH$_2$ carbon position 10 or 8 (33.03, 33.02 ppm)

e - Overlapping $^{13}$C signals for CH$_2$ carbon position 11 or 9

f - Overlapping $^{13}$C signals for CH$_3$ carbon position 12 or 10

g - $^1J_{CH}$ determined from coupled-HSQC measurement
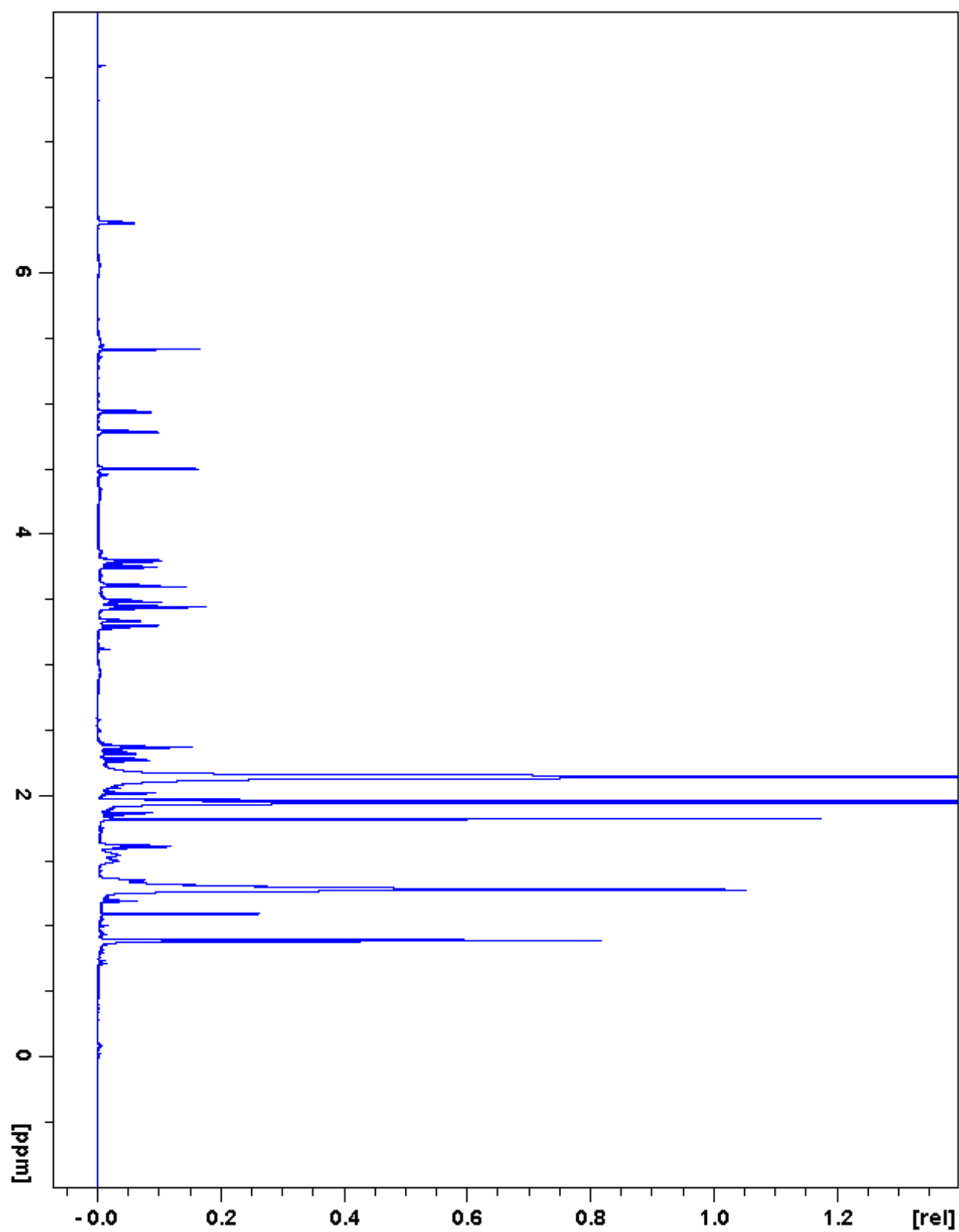
**Figure 3.47.** NAG-I3:24:0(2,10,12) ^{1}H NMR
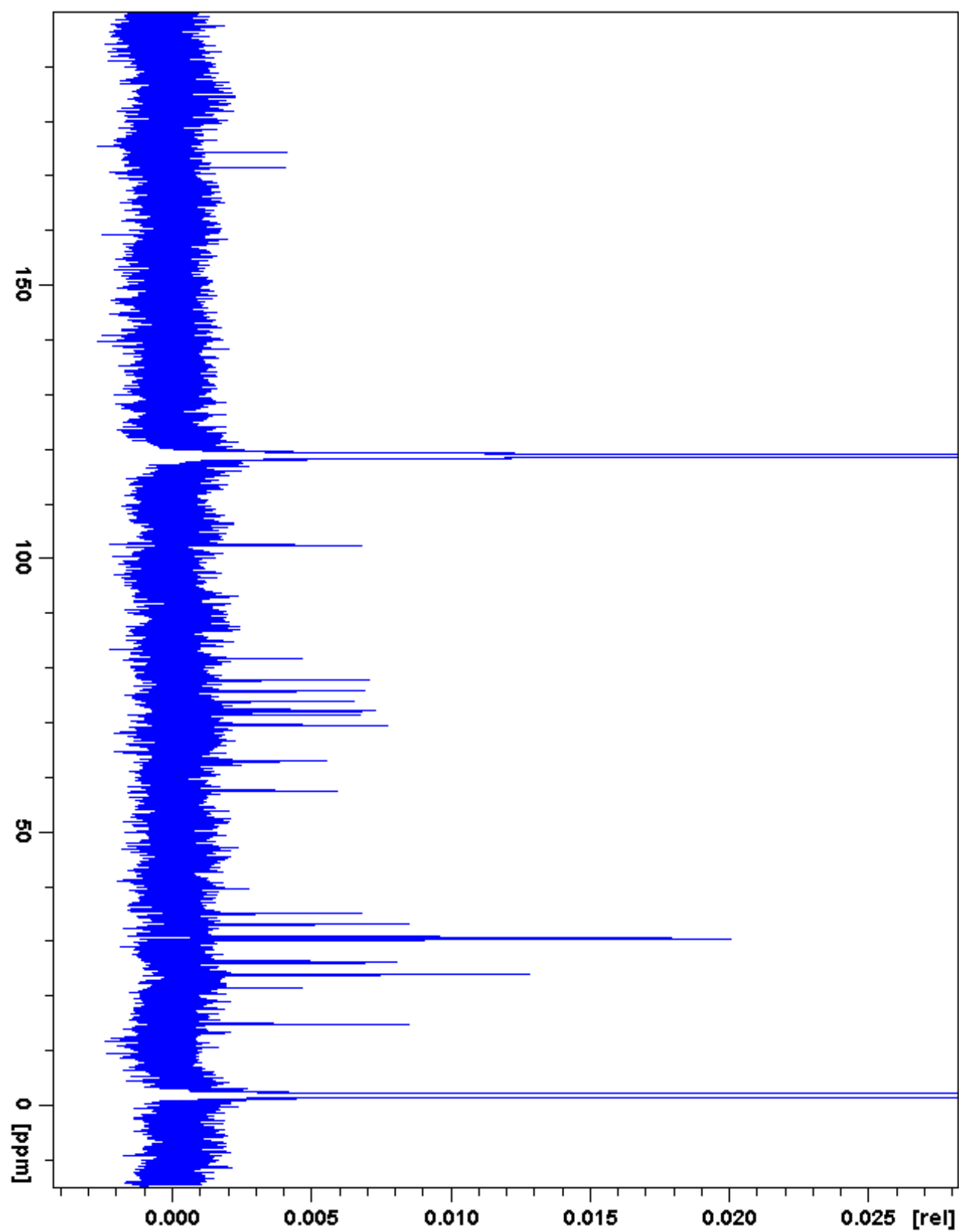
**Figure 3.48.** NAG-I3:24:0(2,10,12) $^{13}$C NMR
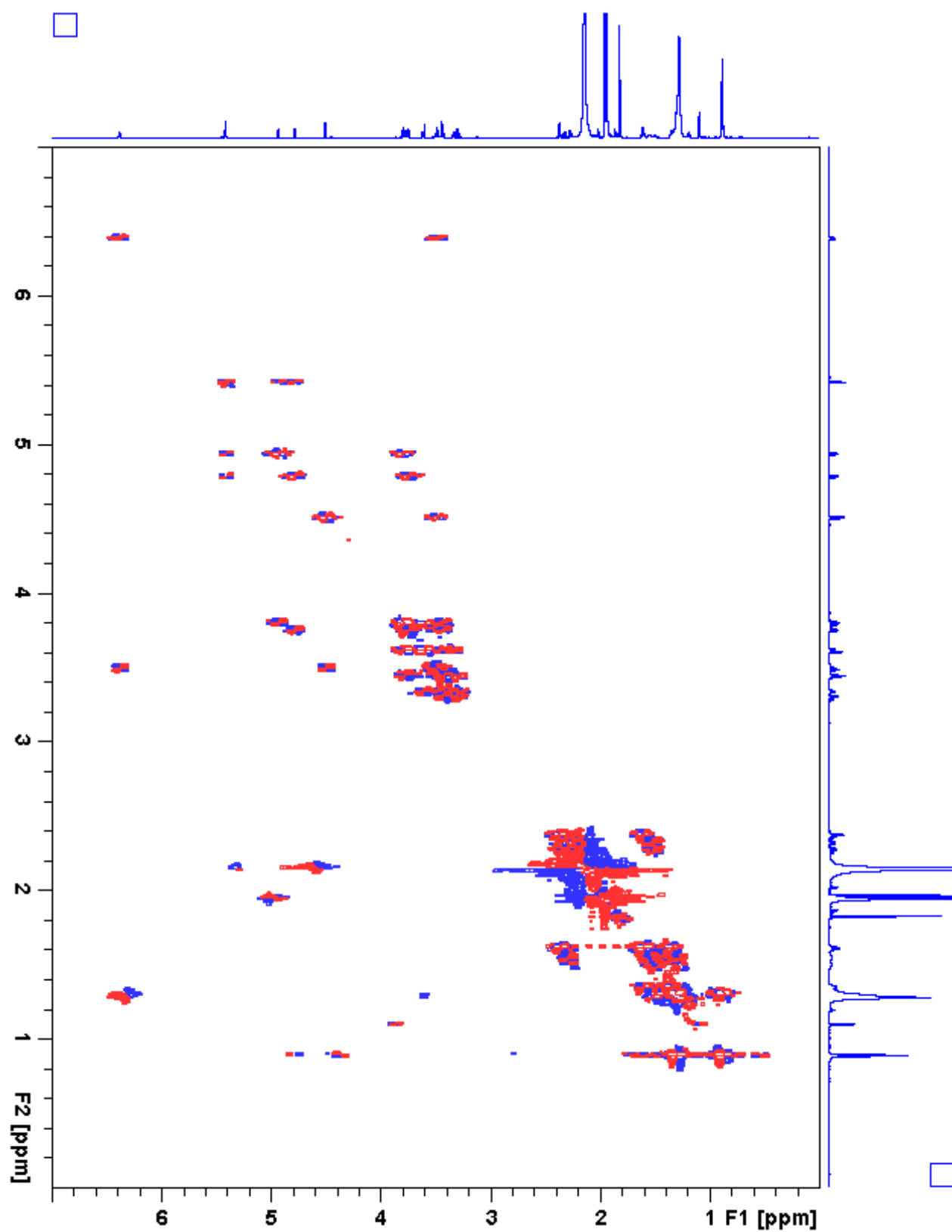
**Figure 3.49.** NAG-I3:24:0(2,10,12) $^{1}$H-$^{1}$H gCOSY
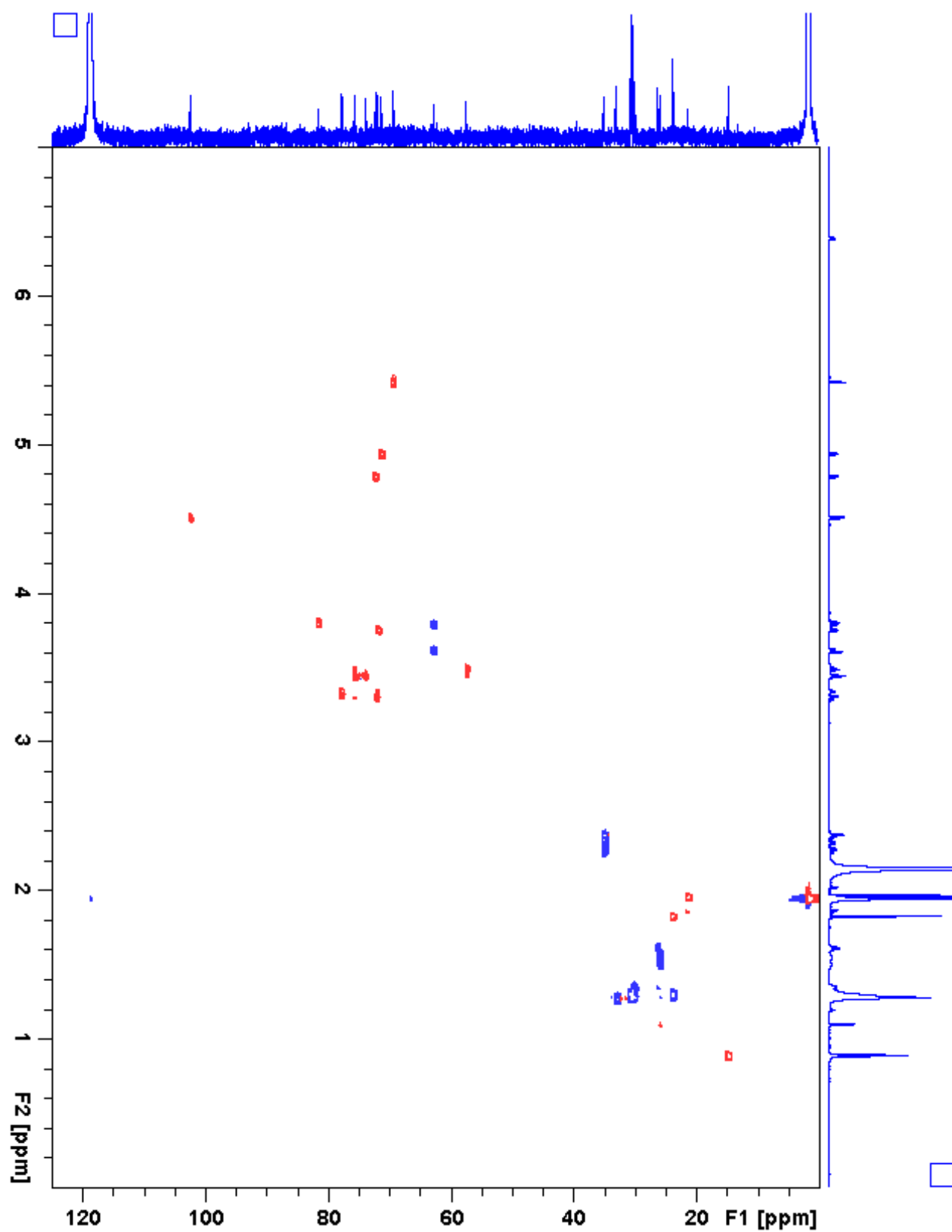
**Figure 3.50.** NAG-I3:24:0(2,10,12) gHSQC
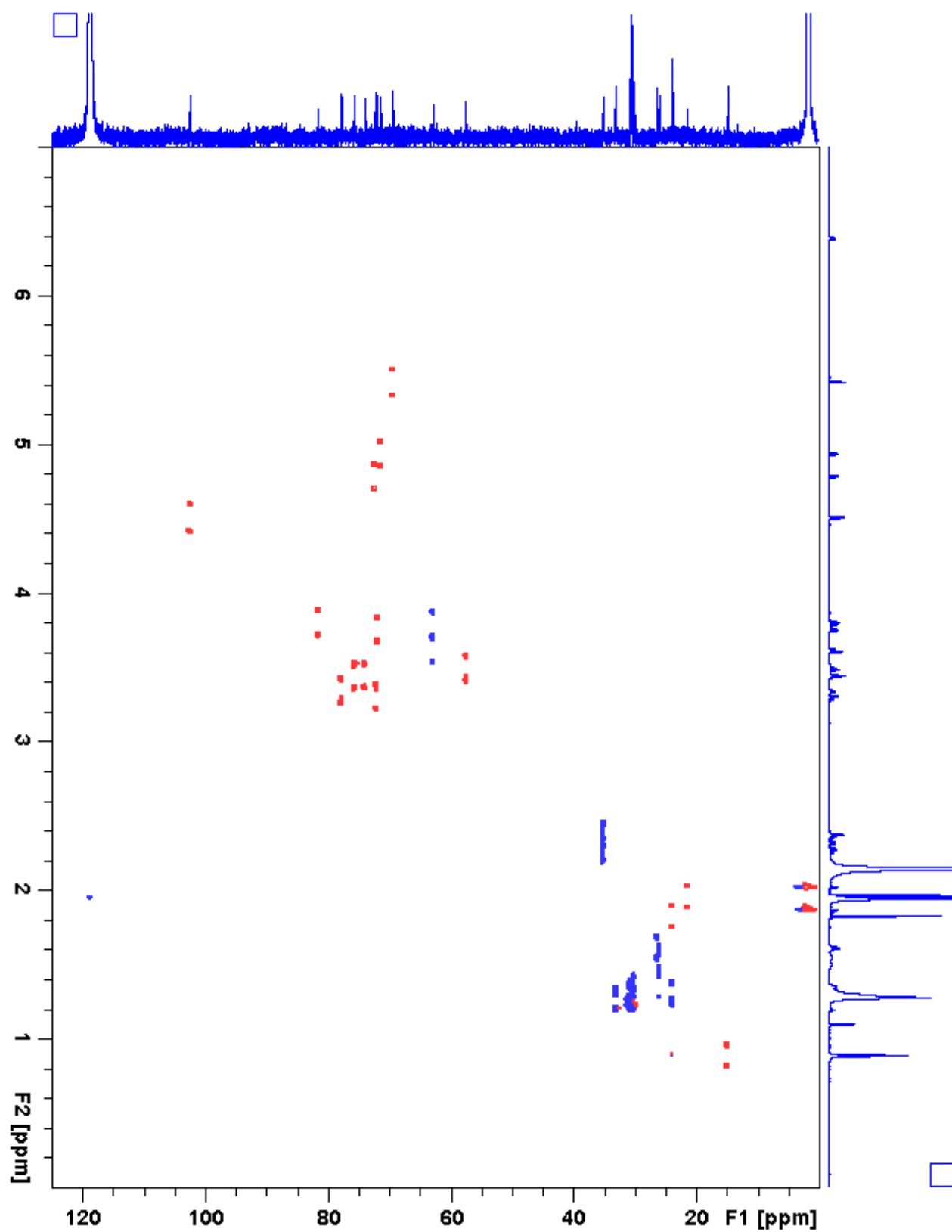
**Figure 3.51.** NAG-I3:24:0(2,10,12) coupled-gHSQC

**Figure 3.52.** NAG-I3:24:0(2,10,12) gHMBC

**Figure 3.53.** NAG-I3:24:0(2,10,12) *J*-resolved

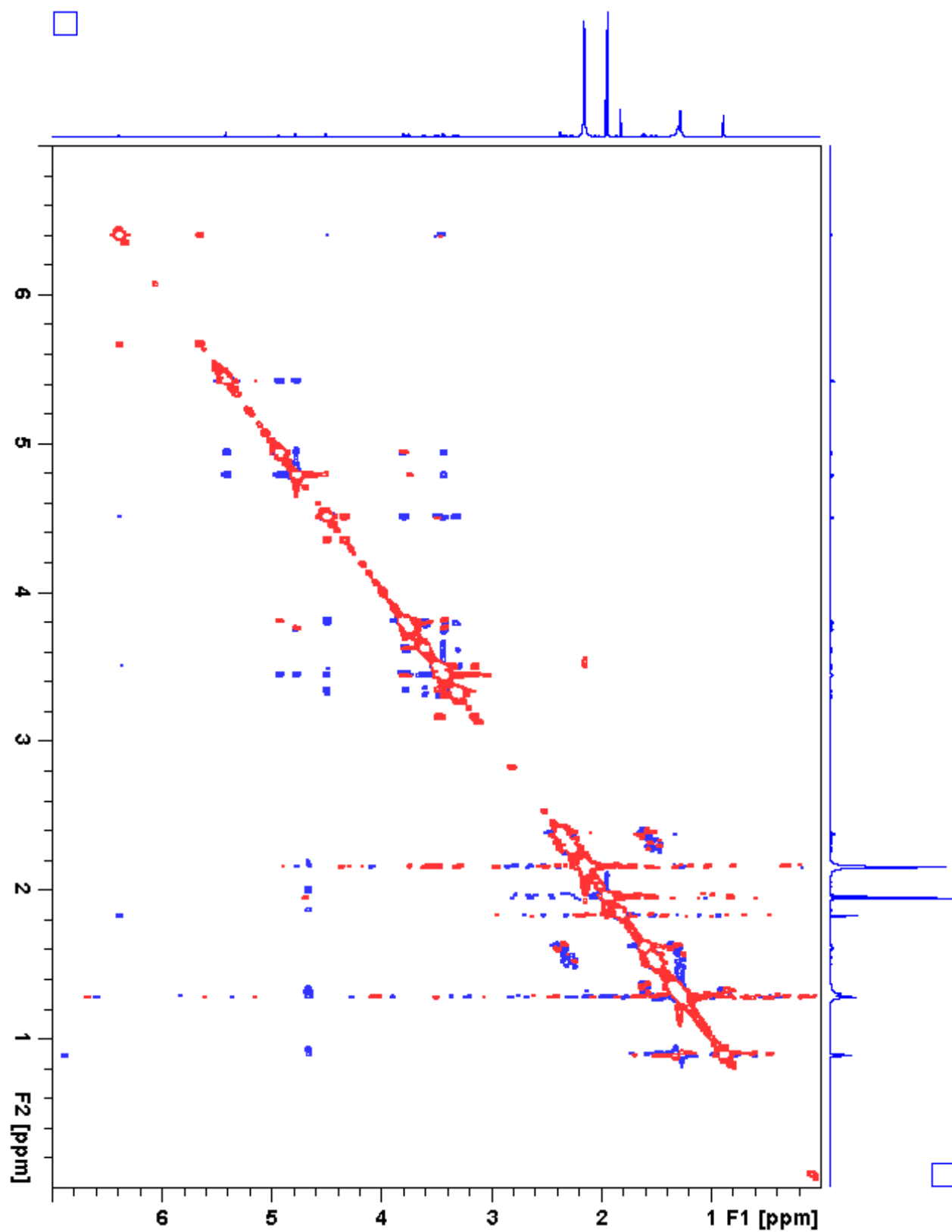**Figure 3.54**. NAG-I3:24:0(2,10,12) ROESY

**Table 3.20.** G-I3:24:0(2,10,12) Chemical shifts and coupling constants
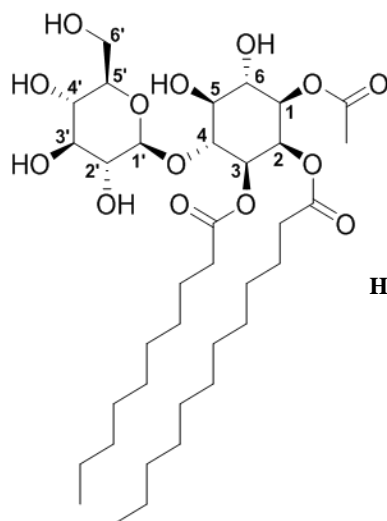


**Molecular Formula:** $C_{36}H_{64}O_{14}$

**110 min Retention Time (ESI-):** 65.37 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{37}H_{65}O_{16}^-$
([M+HCOO]$^-$): 765.4278, found: 765.4278

**110 min Retention Time (ESI+):** 65.37 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{36}H_{68}NO_{14}^+$ ([M+H]$^+$):
738.4634, found: 738.4597

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #29
**Sample mass for NMR analysis:** 1.0 mg
**NMR Solvent:** $D_3CN$
**InChi Key:** CSZWVWXZOGWQNT-FZFLAQSVSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.82 (dd, $J$ = 10.3, 2.9 Hz) | 72.38 |
| -1(CO) |  | 171.48 |
| -2(CH$_3$) | 1.97 | 21.37 |
| **2**(CH) | 5.47 (t, $J$ = 2.9 Hz) | 69.50 |
| -1(CO) |  | 174.10 |
| -2(CH$_2$) | 2.37 (t, $J$ = 7.3 Hz) | 35.03[a] |
| -3(CH$_2$) | 1.61 (p, $J$ = 7.3 Hz) | 26.30 |
| -4 to 11(CH$_2$) | 1.36-1.23 (m) | 30.8-30.1[b], 33.03[c], 23.79[d] |
| -12(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.78[e] |
| **3**(CH) | 4.96 (dd, $J$ = 10.3, 3.0 Hz) | 71.09 |
| -1(CO) |  | 174.16 |
| -2(CH$_2$) | 2.24 (m) | 35.03[a] |
| -3(CH$_2$) | 1.52 (m) | 25.69 |
| -4 to 9(CH$_2$) | 1.36-1.23 (m) | 30.8-30.1[b], 33.03[c], 23.79[d] |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.78[e] |
| **4**(CH) | 3.78 (dd, $J$ = 10.5, 8.8 Hz) | 83.19 |
| **5**(CH) | 3.44 (t, $J$ = 9.1 Hz) | 74.00 |
| **6**(CH) | 3.76 (dd, $J$ = 10.6, 8.8 Hz) | 71.83 |
| **1'**(CH) | 4.33 (d, $J$ = 7.8 Hz) | 104.98 (H-C, $J$ = 161.4 Hz)[f] |
| **2'**(CH) | 3.10 (m) | 74.69 |
| **3'**(CH) | 3.27 (m) | 77.72 |
| **4'**(CH) | 3.27 (m) | 71.37 |
| **5'**(CH) | 3.32 (ddd, $J$ = 9.8, 5.9, 2.5 Hz) | 77.91 |

**Table 3.20.** (continued)

| | | |
|---|---|---|
| **6'**(CH$_2$) | 3.77 (dd, $J$ = 11.9, 2.6 Hz), 3.59 (dd, $J$ = 11.9, 5.9 Hz) | 62.74 |

a - Overlapping $^{13}$C signals

b - $^{13}$C signals for CH$_2$ carbon positions 4 to 9, 4 to 7 (30.81, 30.76, 30.70, 30.59, 30.48, 30.46, 30.46, 30.45, 30.14, 30.07 ppm)

c - Overlapping $^{13}$C signals for CH$_2$ carbon position 10 or 8

d - Overlapping $^{13}$C signals for CH$_2$ carbon position 11 or 9

e - Overlapping $^{13}$C signals for CH$_3$ carbon position 12 or 10

f - $^1J_{CH}$ determined from coupled-HSQC measurement
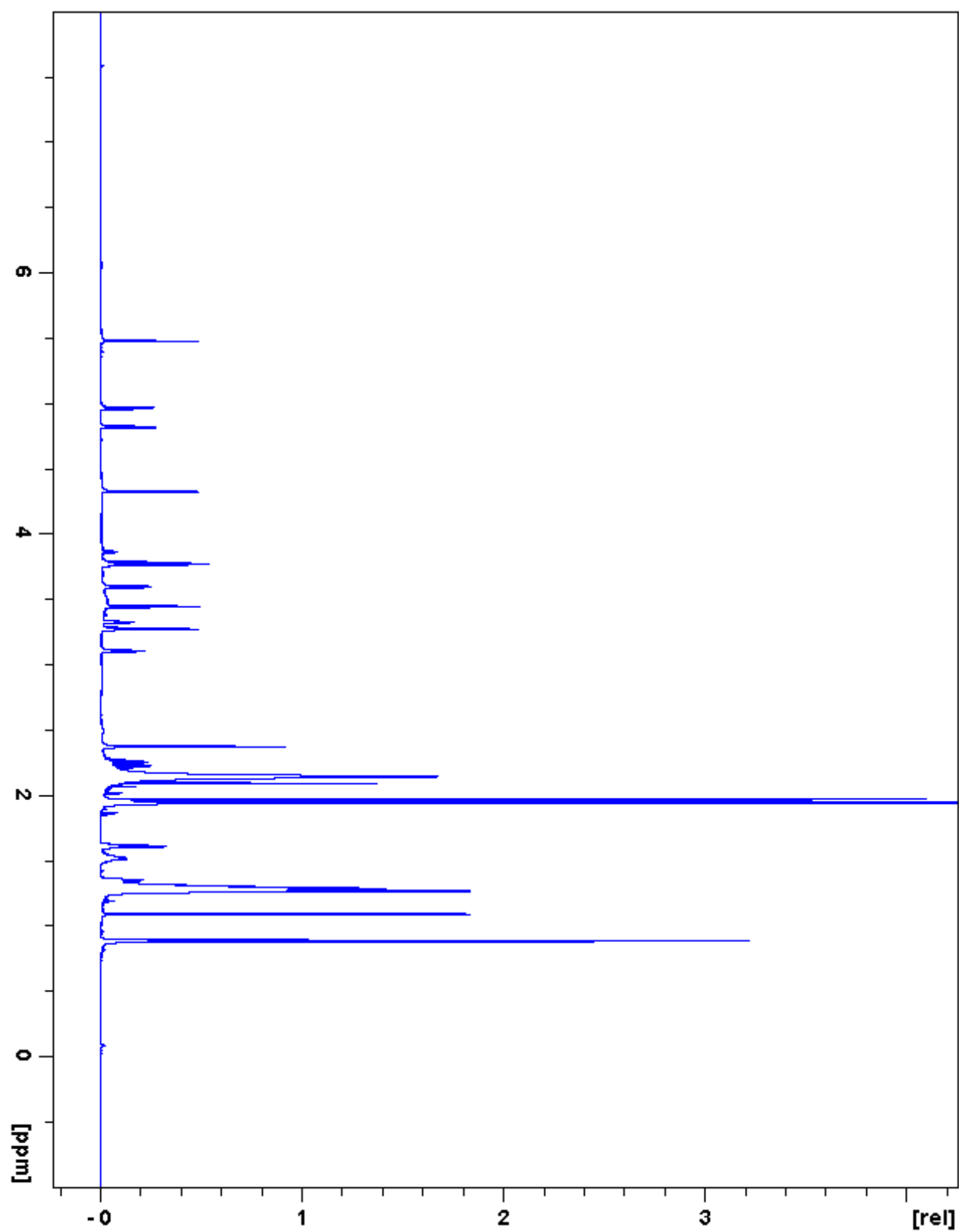
**Figure 3.55.** G-I3:24:0(2,10,12) ^{1}H NMR
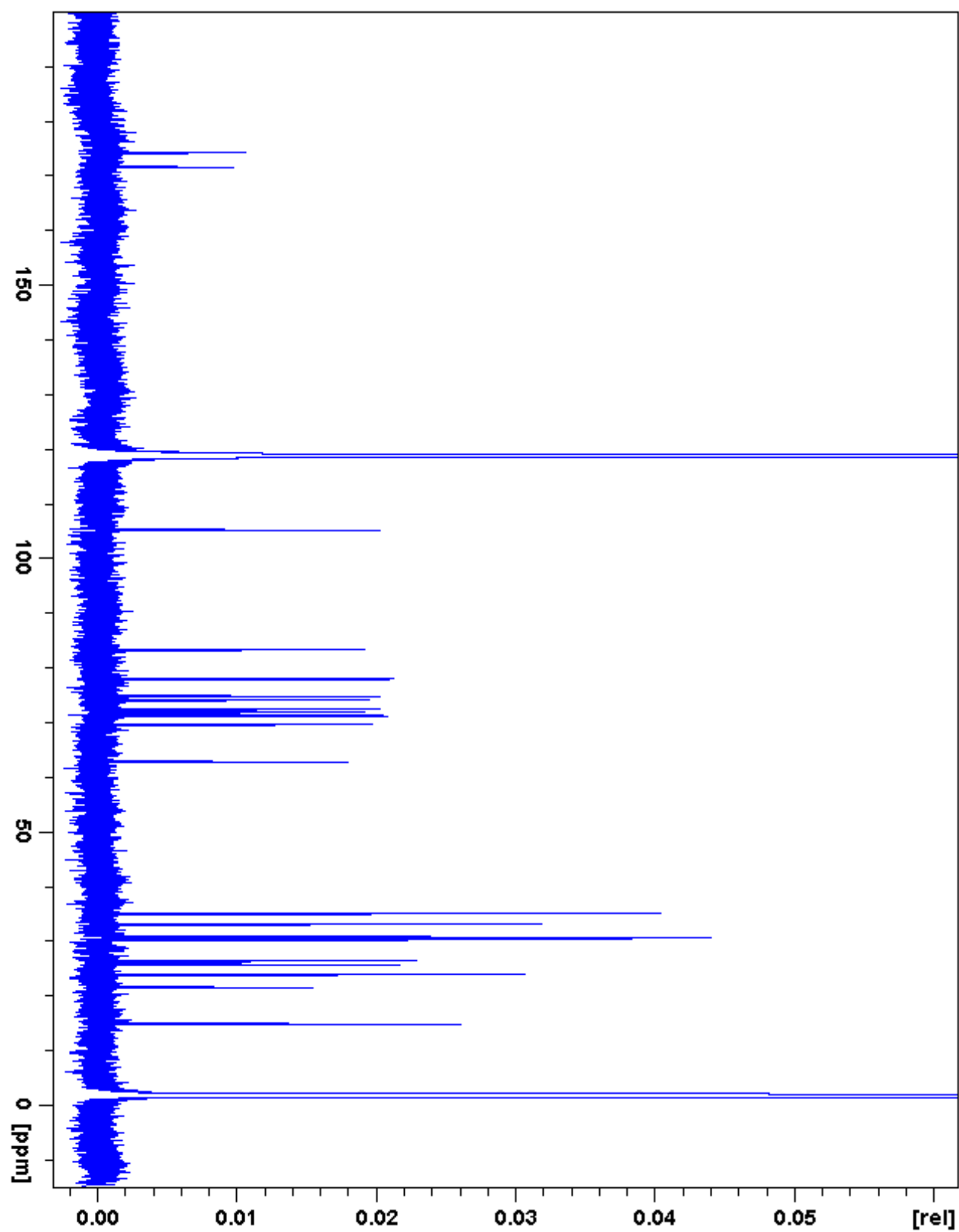
**Figure 3.56.** G-I3:24:0(2,10,12) ^{13}C NMR
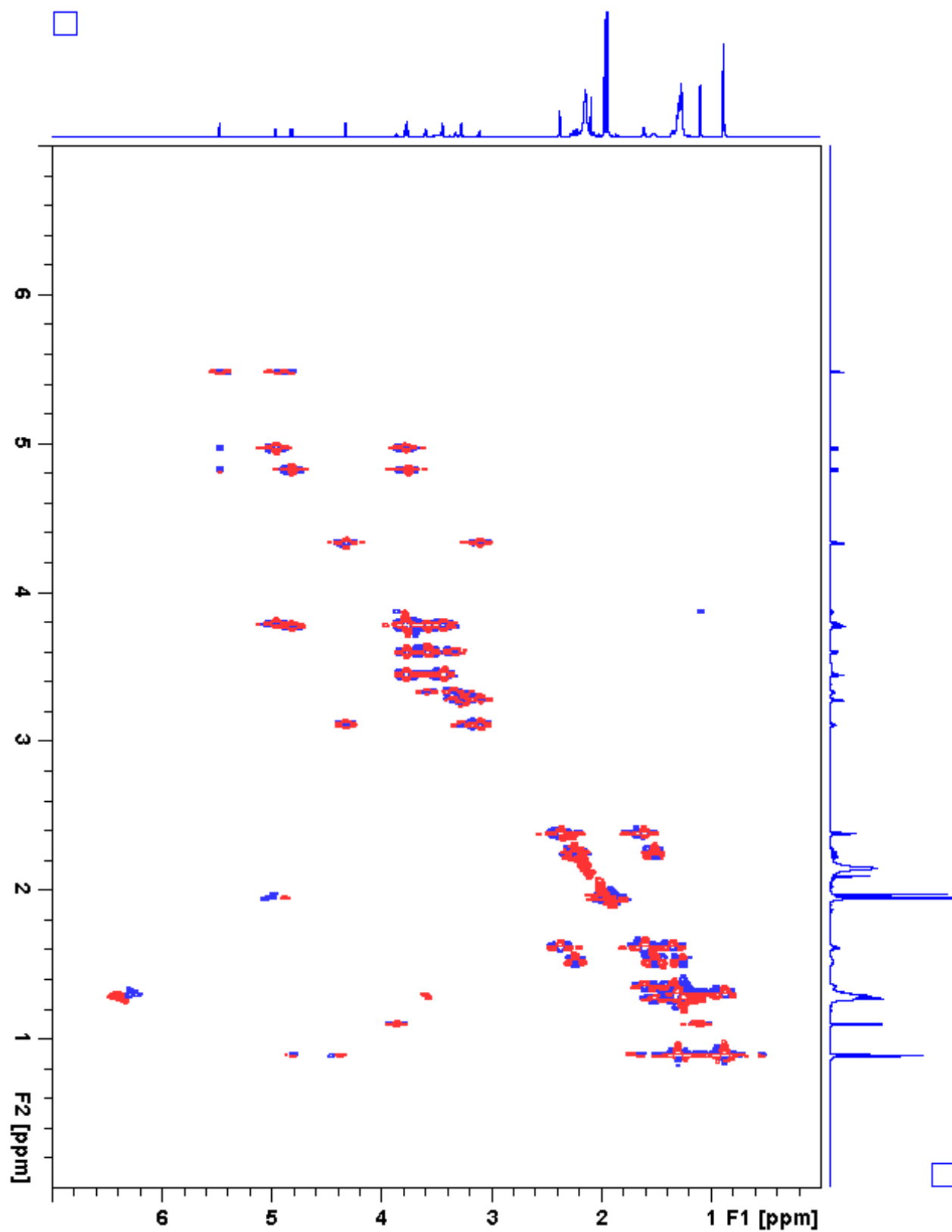
**Figure 3.57.** G-I3:24:0(2,10,12) $^1$H-$^1$H gCOSY
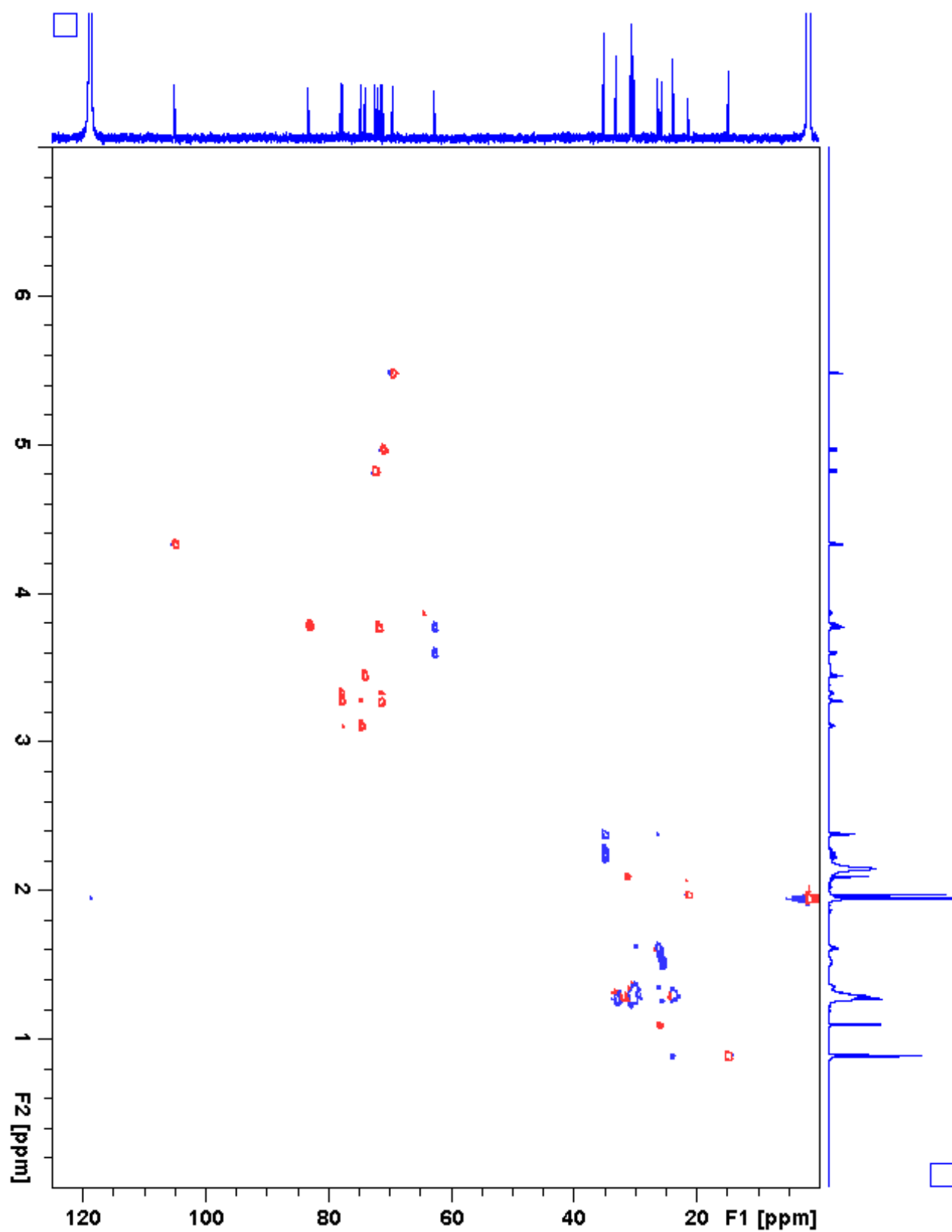
**Figure 3.58.** G-I3:24:0(2,10,12) gHSQC
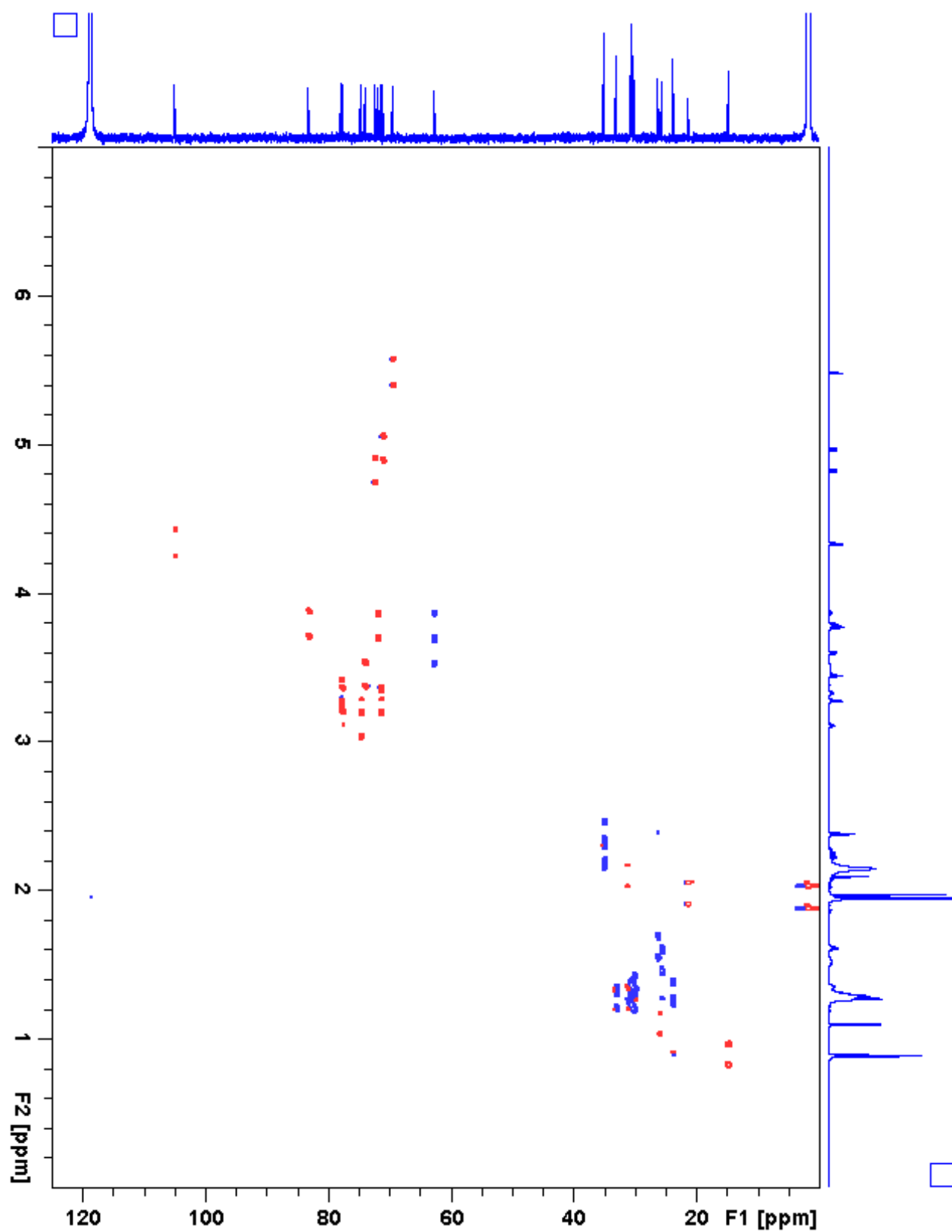
**Figure 3.59.** G-I3:24:0(2,10,12) coupled-gHSQC

356
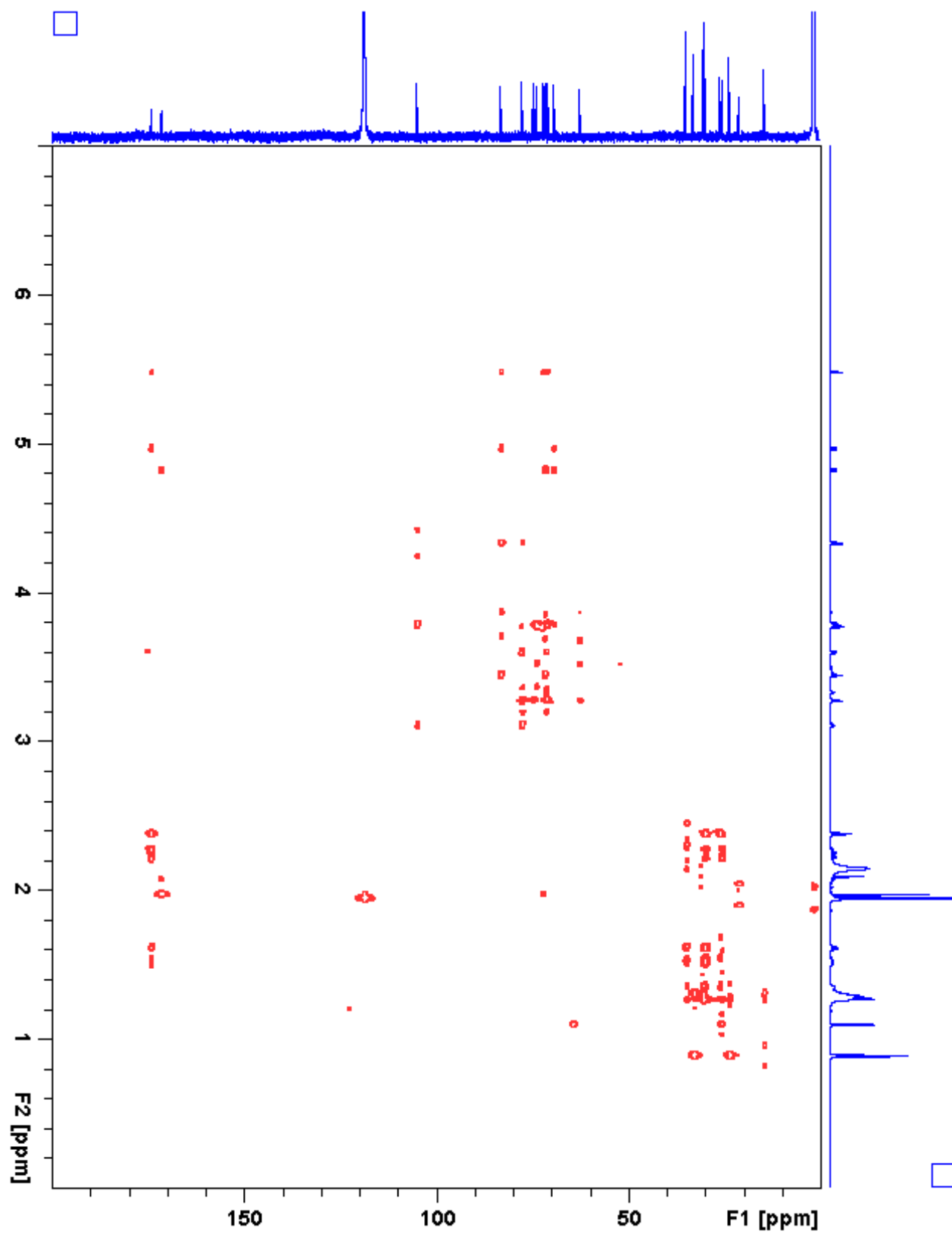
**Figure 3.60.** G-I3:24:0(2,10,12) gHMBC

**Figure 3.61.** G-I3:24:0(2,10,12) *J*-resolved
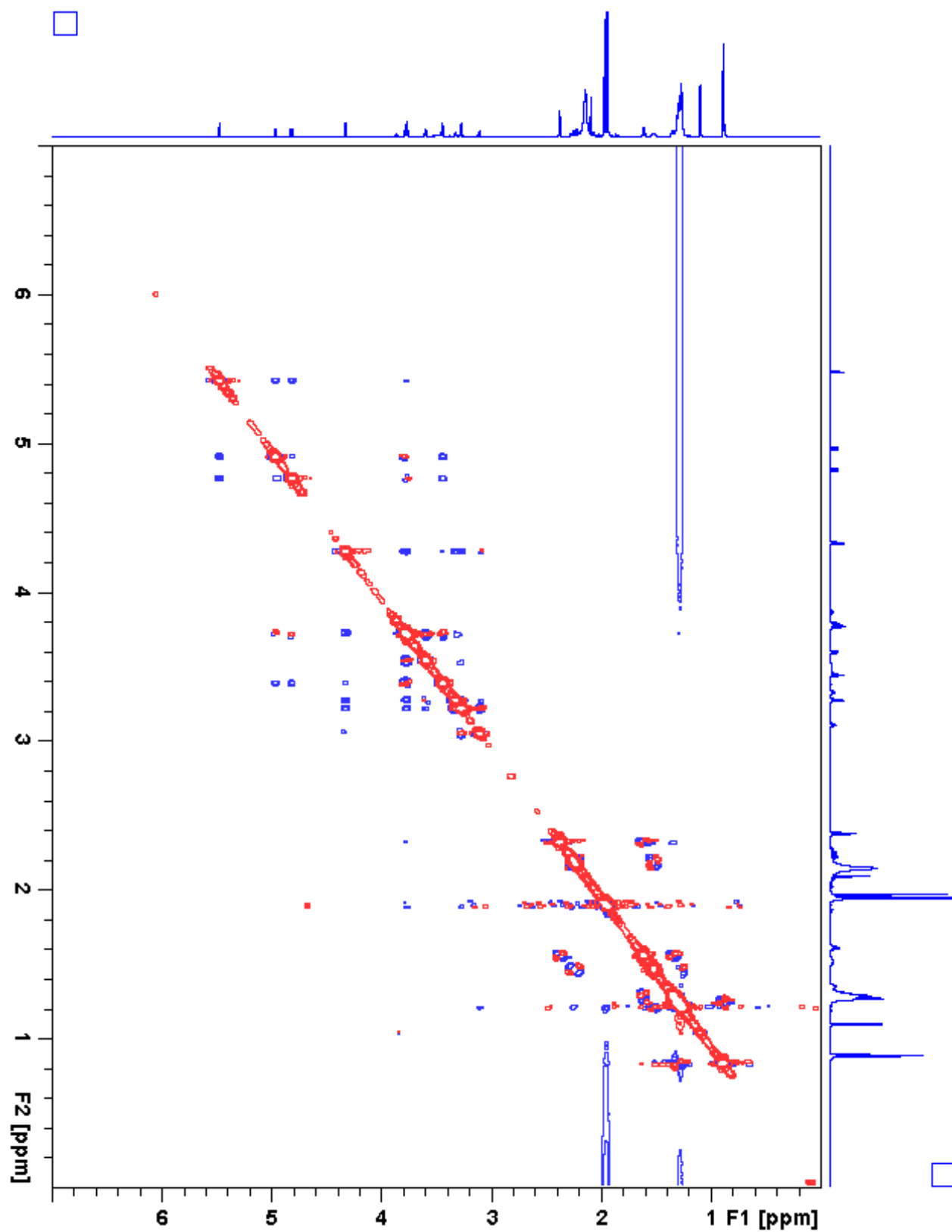
**Figure 3.62.** G-I3:24:0(2,10,12) ROESY

**Table 3.21.** I3:22:0(2,10,10) Chemical shifts and coupling constants
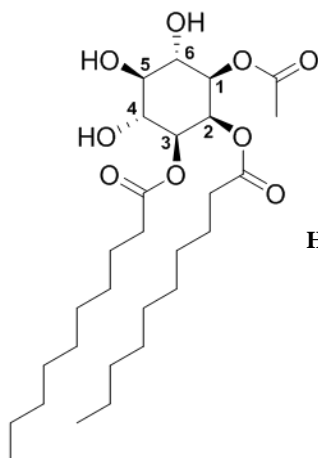


**Molecular Formula:** $C_{28}H_{50}O_9$

**110 min Retention Time (ESI-):** 68.18 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{29}H_{51}O_{11}^{-}$
([M+HCOO]⁻): 575.3437, found: 575.3435

**110 min Retention Time (ESI+):** 68.16 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{28}H_{54}NO_9^{+}$ ([M+NH₄]⁺):
548.3793, found: 548.3777

**Instrument:** Bruker Avance 900 MHz NMR

**Fraction:** #34

**Sample mass for NMR analysis:** 3.6 mg

**NMR Solvent:** D₃CN

**InChi Key:** ROFZSOBDHCZVAN-WBKISLEQSA-N

| Carbon # (group) | ¹H (ppm) | ¹³C (ppm) |
|---|---|---|
| **1**(CH) | 4.76 (dd, $J$ = 10.2, 3.0 Hz) | 72.84 |
| -1(CO) | | 171.49 |
| -2(CH₃) | 1.96 | 21.39 |
| **2**(CH) | 5.44 (t, $J$ = 2.9 Hz) | 69.61 |
| -1(CO) | | 174.1[a] |
| -2(CH₂) | 2.37 (t, $J$ = 7.3 Hz) | 35.0[b] |
| -3(CH₂) | 1.60 (p, $J$ = 7.3 Hz) | 26.30 |
| -4 to 9(CH₂) | 1.36-1.23 (m) | 30.6-30.1[c], 33.01[d], 23.78[e] |
| -10(CH₃) | 0.88 (t, $J$ = 7.2 Hz) | 14.78[f] |
| **3**(CH) | 4.78 (dd, $J$ = 10.2, 3.0 Hz) | 72.53 |
| -1(CO) | | 174.1[a] |
| -2(CH₂) | 2.24 (m) | 35.0[b] |
| -3(CH₂) | 1.53 (m) | 25.95 |
| -4 to 9(CH₂) | 1.36-1.23 (m) | 30.6-30.1[c], 33.01[d], 23.78[e] |
| -10(CH₃) | 0.88 (t, $J$ = 7.2 Hz) | 14.78[f] |
| **4**(CH) | 3.68 or 3.69 (dd, $J$ = 10.5, 9.0 Hz) | 71.8[g] |
| **5**(CH) | 3.30 (t, $J$ = 9.2 Hz) | 75.69 |
| **6**(CH) | 3.68 or 3.69 (dd, J = 10.5, 9.0 Hz) | 71.8[g] |

a - Two ¹³C signals not resolved in 2D spectra (174.15, 174.13 ppm)

b - Two ¹³C signals not resolved in 2D spectra (35.08, 35.04 ppm)

c - ¹³C signals for CH₂ carbon positions 4 to 7 (30.61, 30.56, 30.46, 30.43, 30.42, 30.41, 30.09, 30.05 ppm)

d - Overlapping ¹³C signals for CH₂ carbon position 8

e - Overlapping ¹³C signals for CH₂ carbon position 9

f - Overlapping ¹³C signals for CH₃ carbon position 10

g - Two ¹³C signals not resolved in 2D spectra (71.80, 71.75 ppm)

**Figure 3.63.** I3:22:0(2,10,10) ¹H NMR

361
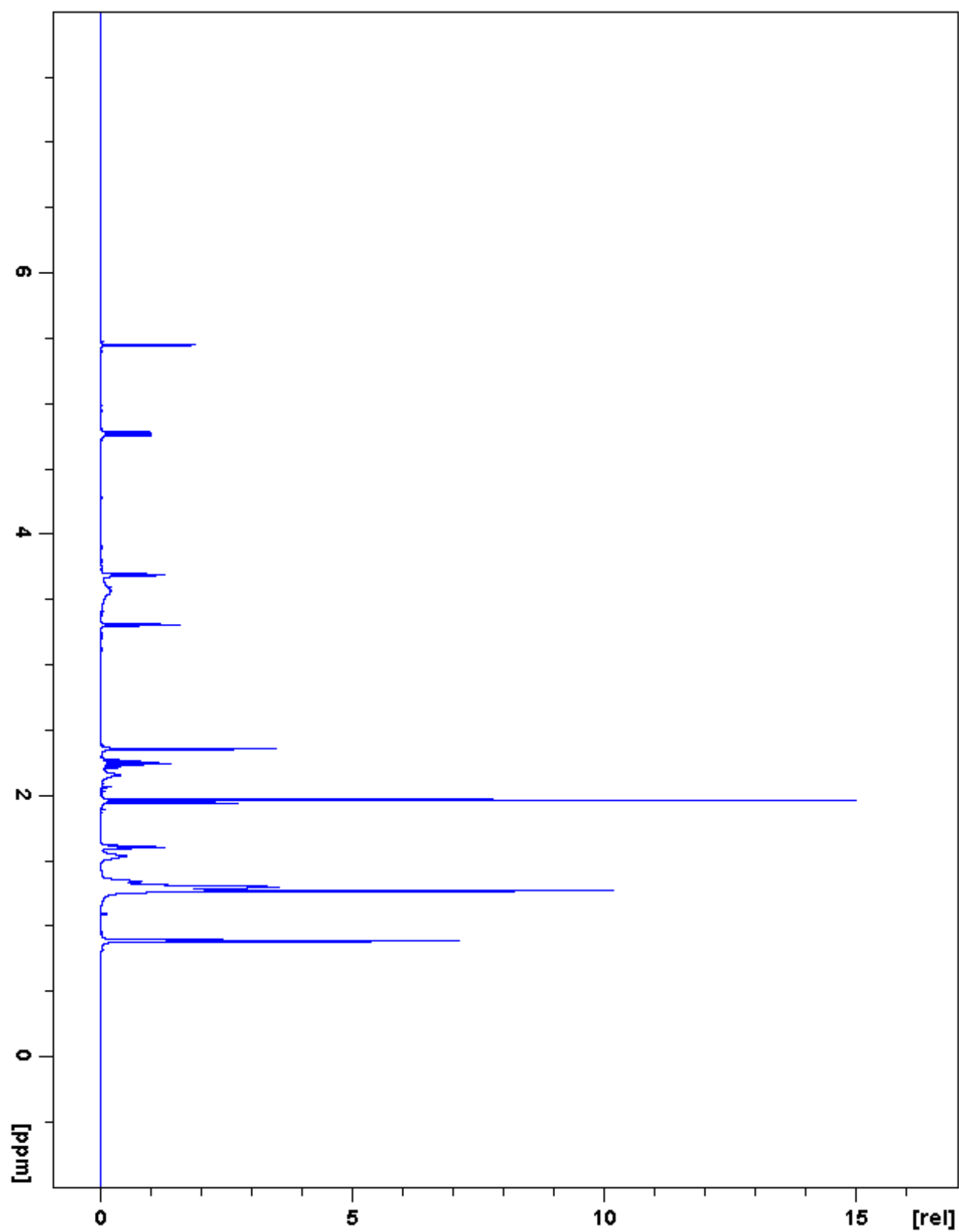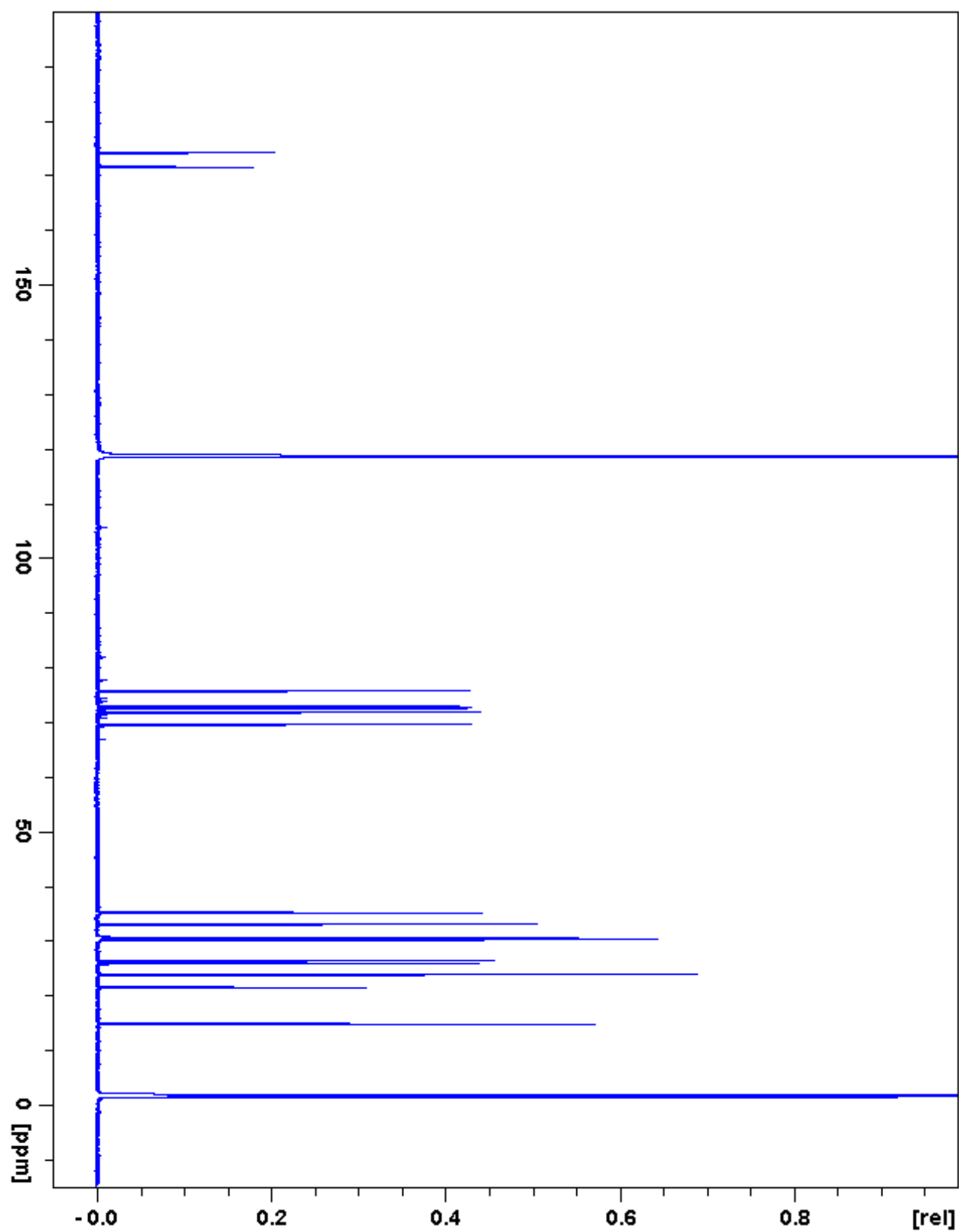
**Figure 3.64.** I3:22:0(2,10,10) $^{13}C$ NMR

362
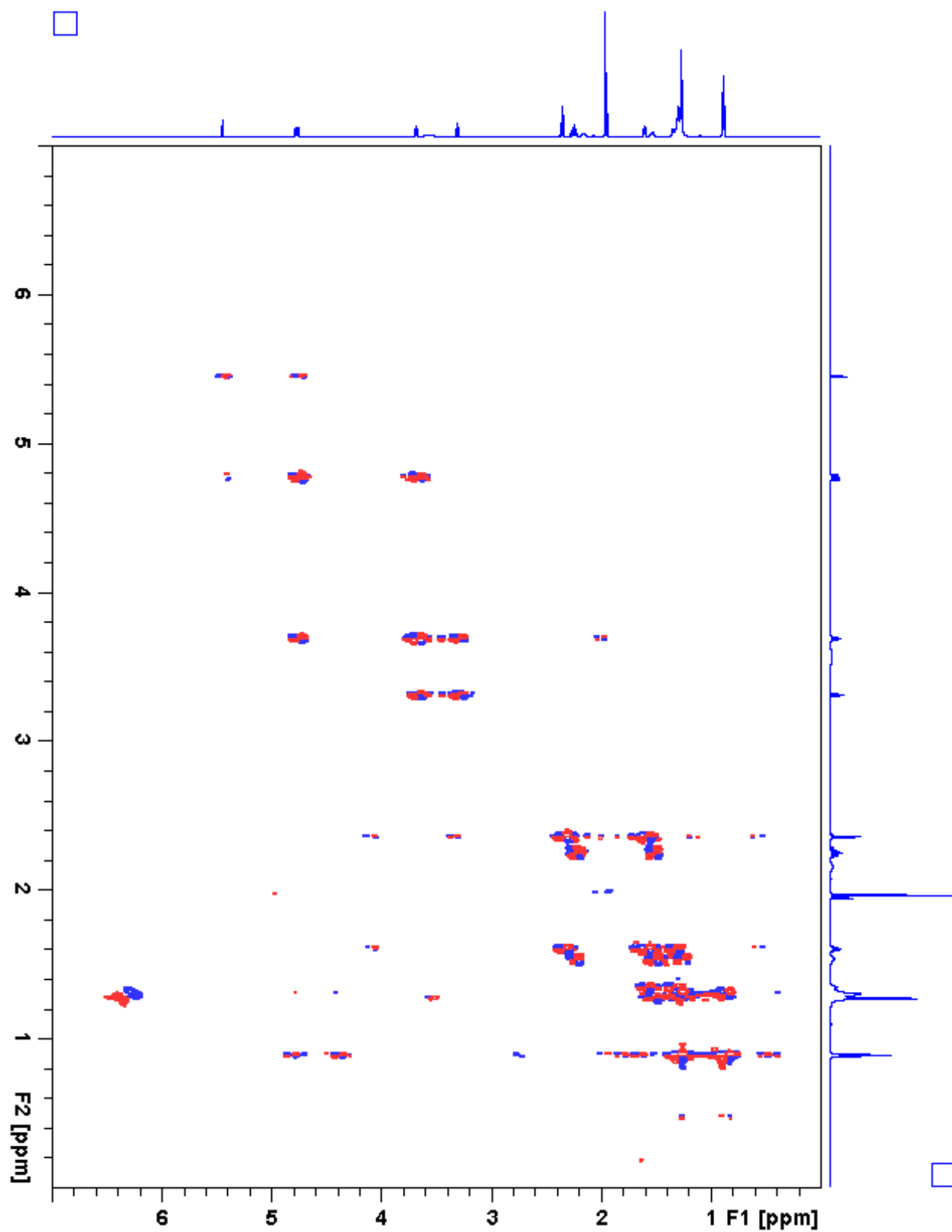
**Figure 3.65.** I3:22:0(2,10,10) $^1$H-$^1$H gCOSY
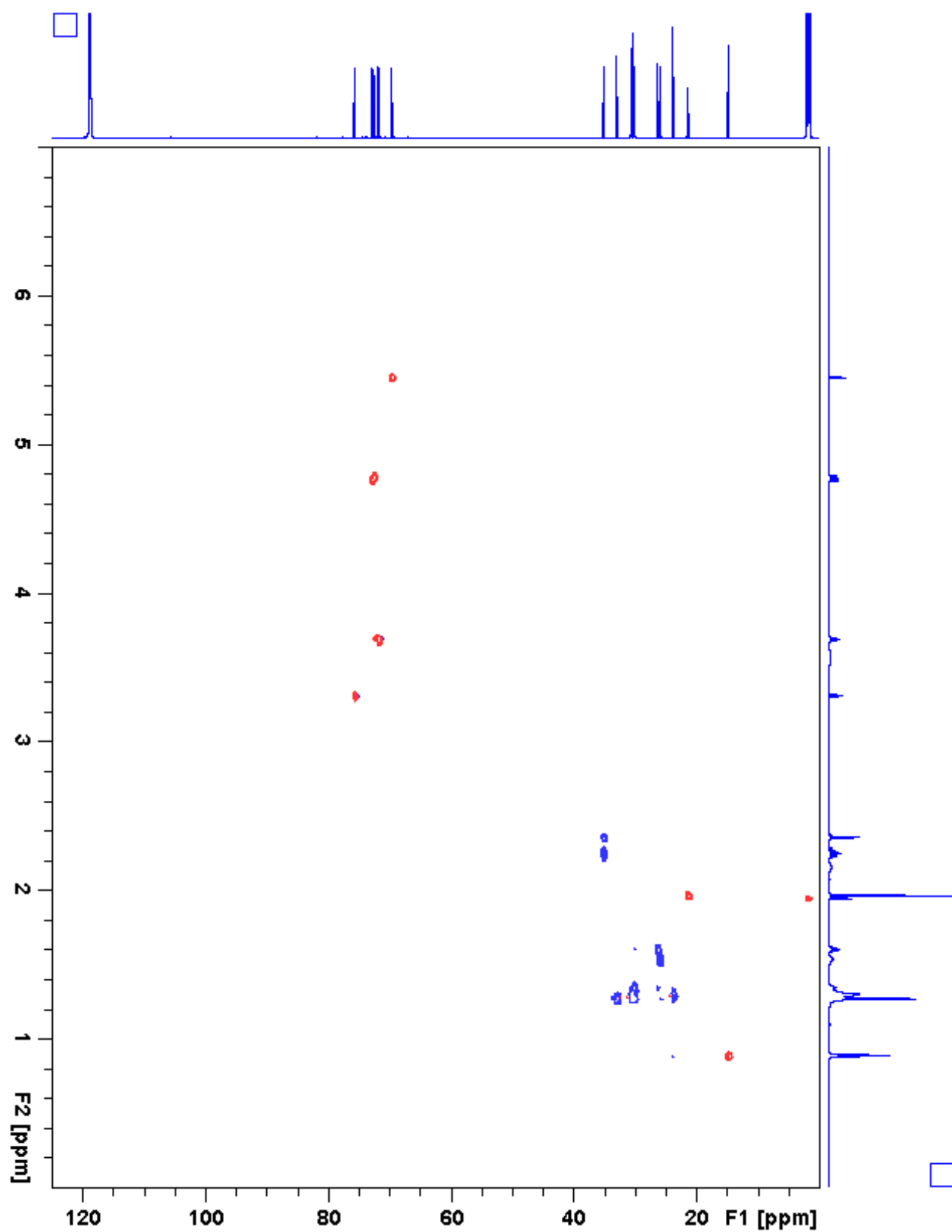
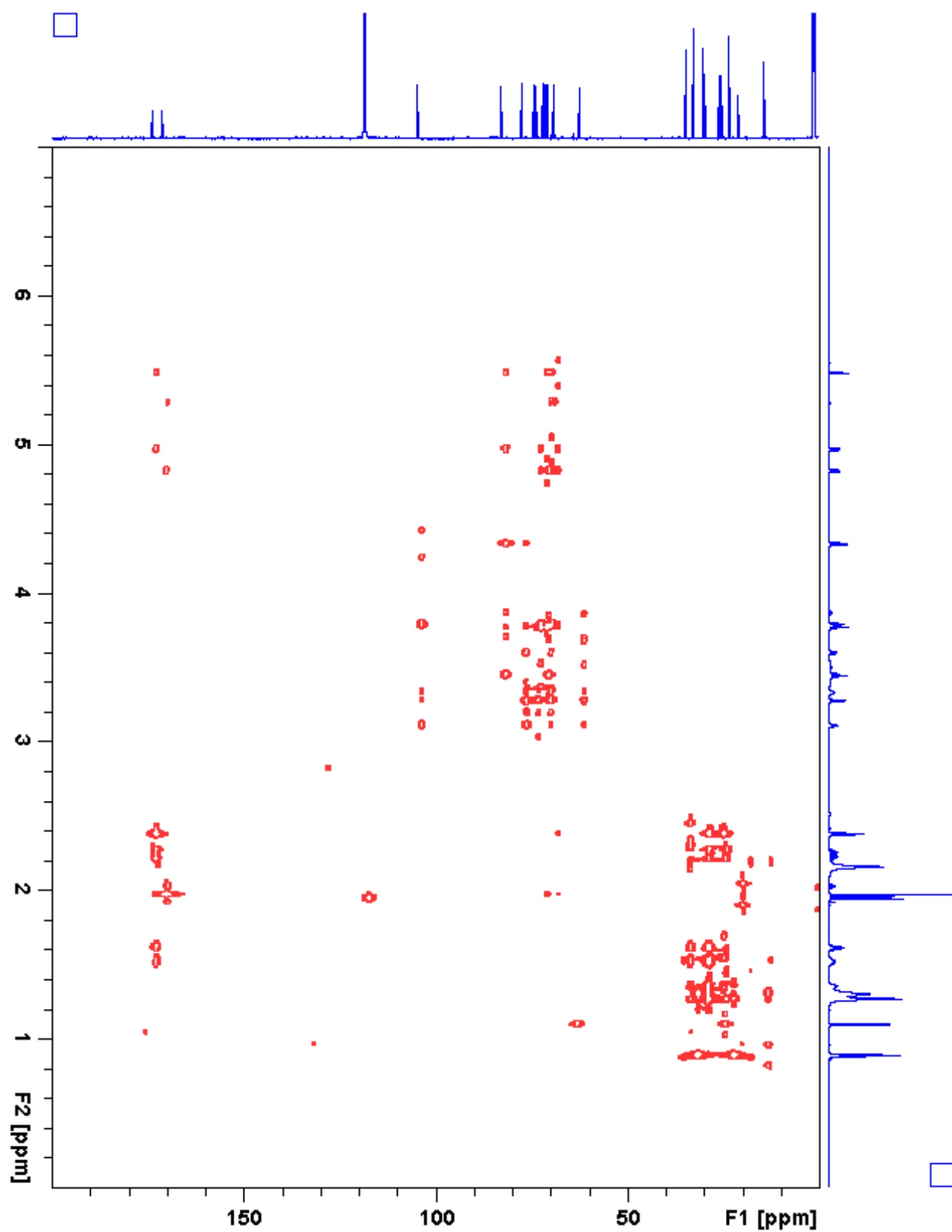**Figure 3.66**. I3:22:0(2,10,10) gHSQC
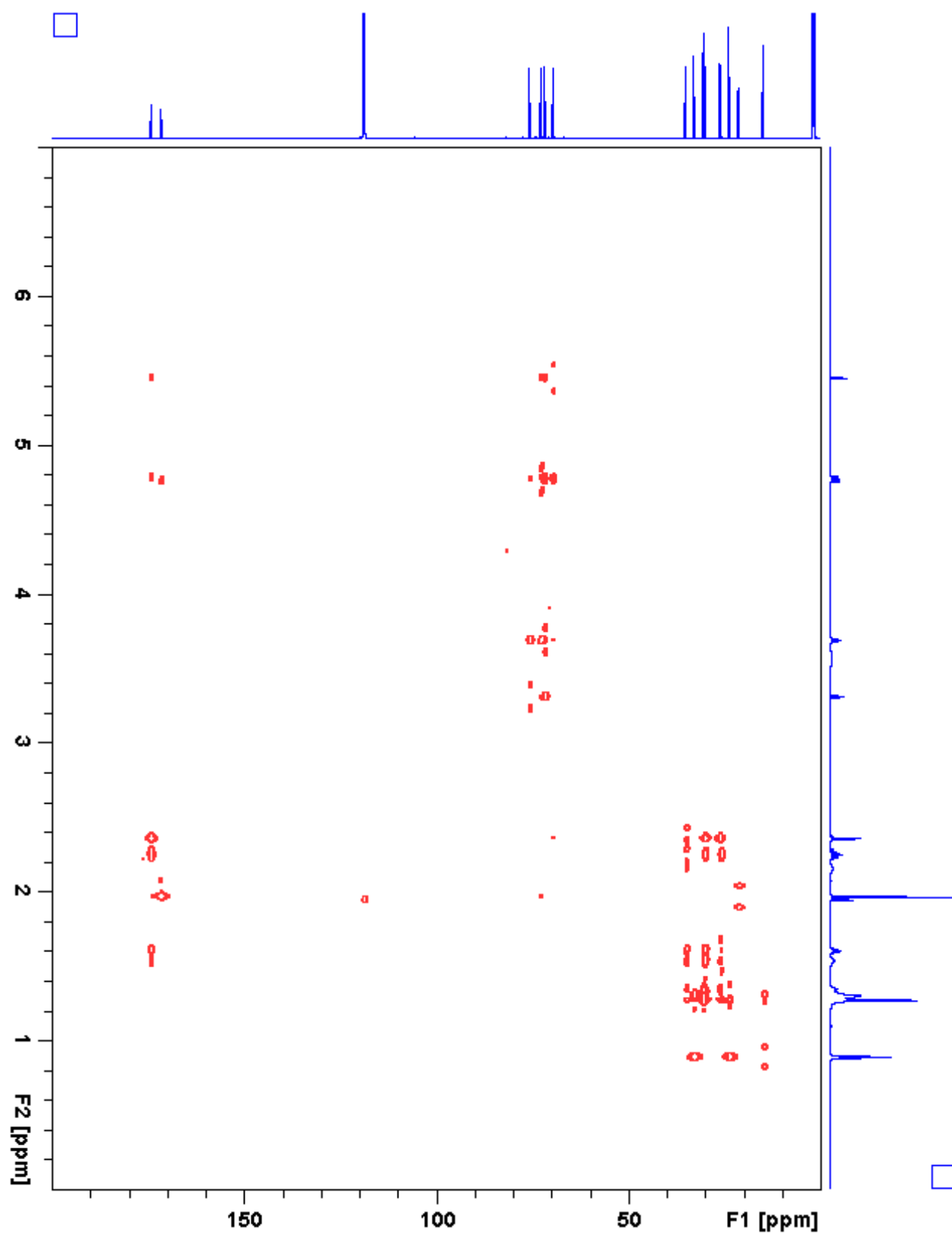
**Figure 3.67.** I3:22:0(2,10,10) gHMBC
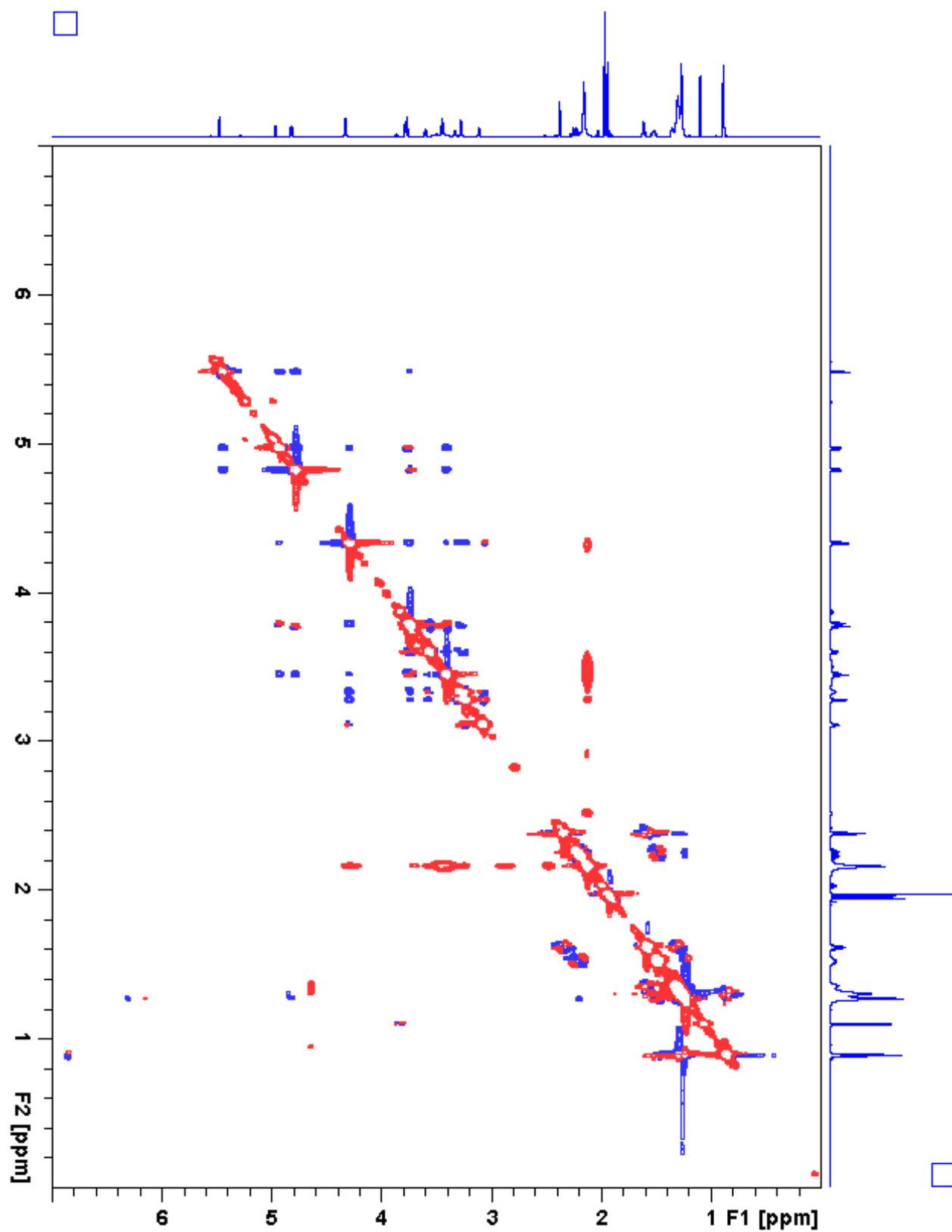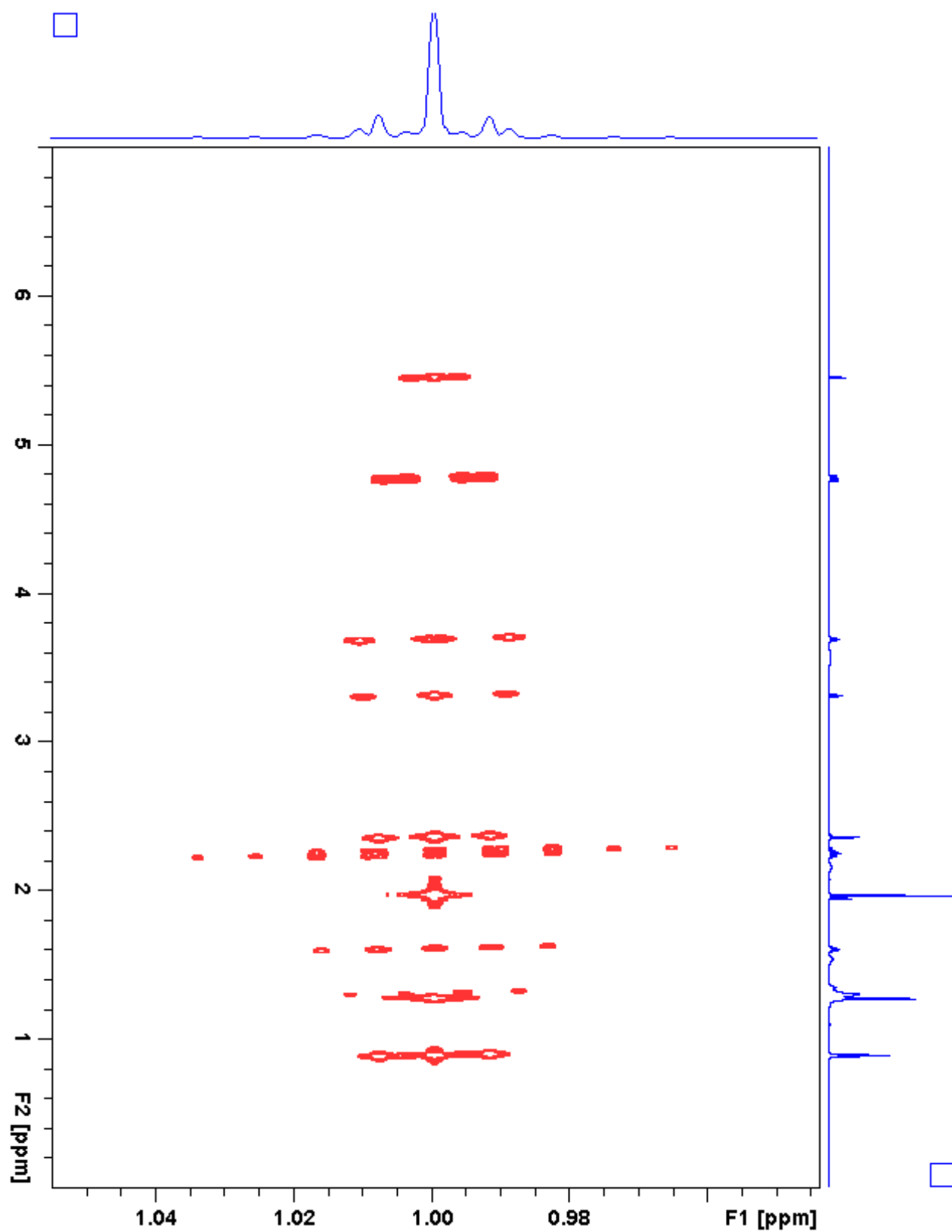
**Figure 3.68.** I3:22:0(2,10,10) *J*-resolved
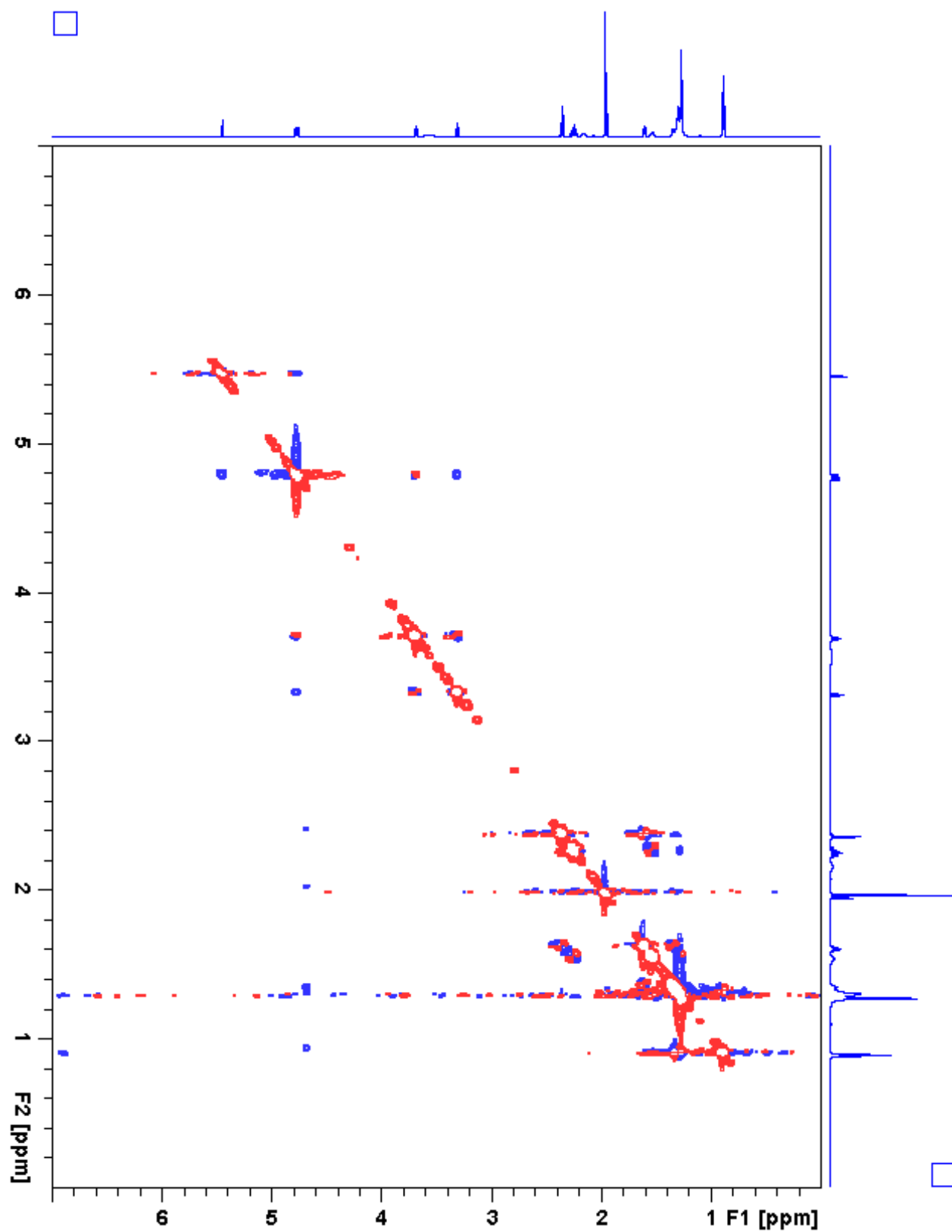
**Figure 3.69.** I3:22:0(2,10,10) ROESY

**Table 3.22.** I4:24:0(2,2,10,10) Chemical shifts and coupling constants
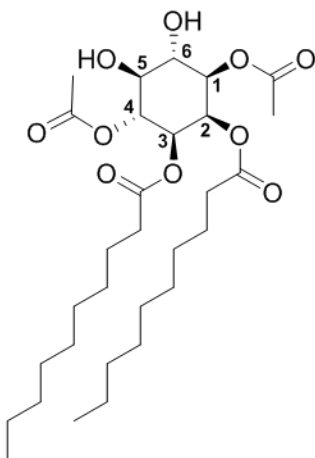


**Molecular Formula:** $C_{30}H_{52}O_{10}$

**110 min Retention Time (ESI-):** 73.90 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{31}H_{53}O_{12}^-$
([M+HCOO]$^-$): 617.3543, found: 617.3539

**110 min Retention Time (ESI+):** 73.89 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{30}H_{56}NO_{10}^+$ ([M+NH$_4$]$^+$):
590.3899, found: 590.3892

**Instrument:** Agilent DDR2 500 MHz NMR

**Fraction:** #44

**Sample mass for NMR analysis:** 27.8 mg

**NMR Solvent:** D$_3$CN

**InChi Key:** QSRBSPJUNLXRLY-XBSCKGQLSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.81 (dd, $J$ = 10.2, 2.9 Hz) | 72.47 |
| -1(CO) | | 171.47 |
| -2(CH$_3$) | 1.97 | 21.37 |
| **2**(CH) | 5.49 (t, $J$ = 3.0 Hz) | 69.42 |
| -1(CO) | | 174.11 |
| -2(CH$_2$) | 2.38 (t, $J$ = 7.2 Hz) | 35.07 |
| -3(CH$_2$) | 1.62 (p, $J$ = 7.2 Hz) | 26.36 |
| -4 to 9(CH$_2$) | 1.37-1.21 (m) | 30.6-30.1$^a$, 33.0$^b$, 23.8$^c$ |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.8$^f$ |
| **3**(CH) | 4.97 (dd, $J$ = 10.6, 3.0 Hz) | 70.36 |
| -1(CO) | | 173.80 |
| -2(CH$_2$) | 2.18 (m) | 35.00 |
| -3(CH$_2$) | 1.48 (p, $J$ = 7.3 Hz) | 25.89 |
| -4 to 9(CH$_2$) | 1.37-1.21 (m) | 30.6-30.1$^a$, 33.0$^b$, 23.8$^c$ |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.8$^f$ |
| **4**(CH) | 5.21 (t, $J$ = 10.0 Hz) | 72.66 |
| -1(CO) | | 171.14 |
| -2(CH$_3$) | 2.01 | 21.54 |
| **5**(CH) | 3.50 (t, $J$ = 9.5 Hz) | 73.50 |
| **6**(CH) | 3.77 (t, J = 9.7 Hz) | 71.86 |

a - $^{13}$C signals for CH$_2$ carbon positions 4 to 9, 4 to 7 (30.60, 30.52, 30.46, 30.30.46, 30.41, 30.36, 30.06, 30.05 ppm)

b - Two $^{13}$C signals for CH$_2$ carbon position 10 or 8 (33.03, 33.00 ppm)

c - Two $^{13}$C signals for CH$_2$ carbon position 11 or 9 (23.78, 23.77 ppm)

d - Two $^{13}$C signals for CH$_3$ carbon position 12 or 10 (14.79, 14.78 ppm)
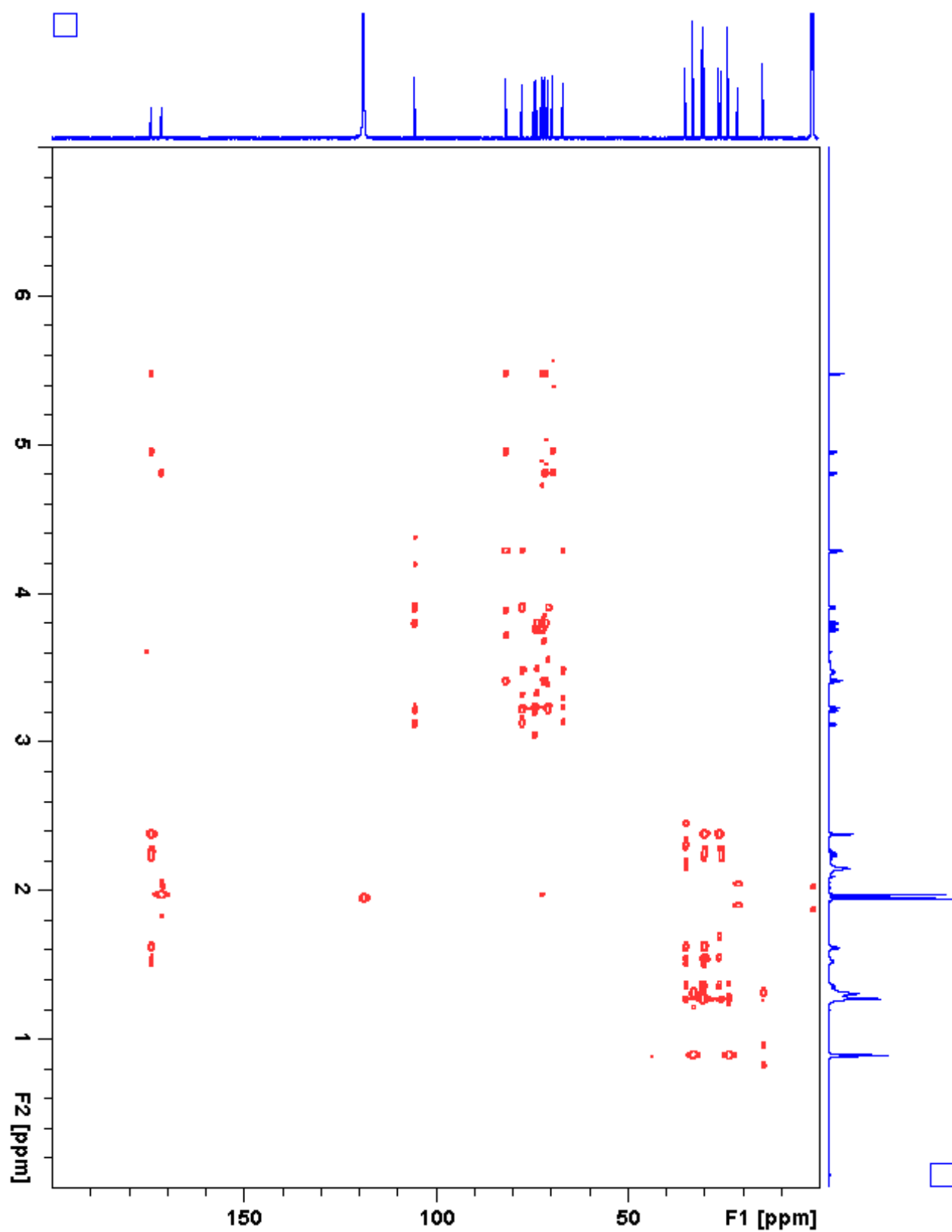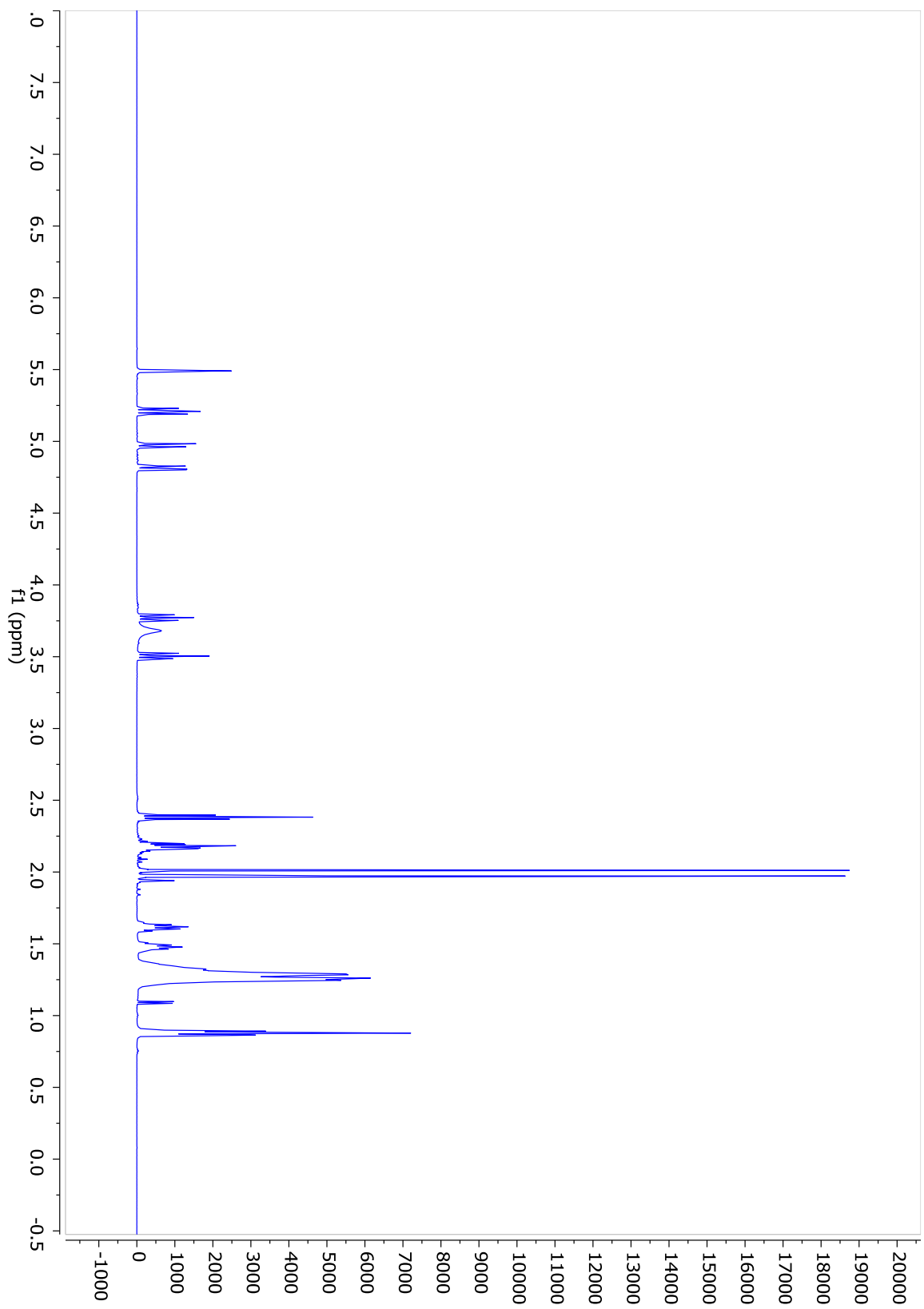
**Figure 3.70.** I4:24:0(2,2,10,10) ¹H NMR
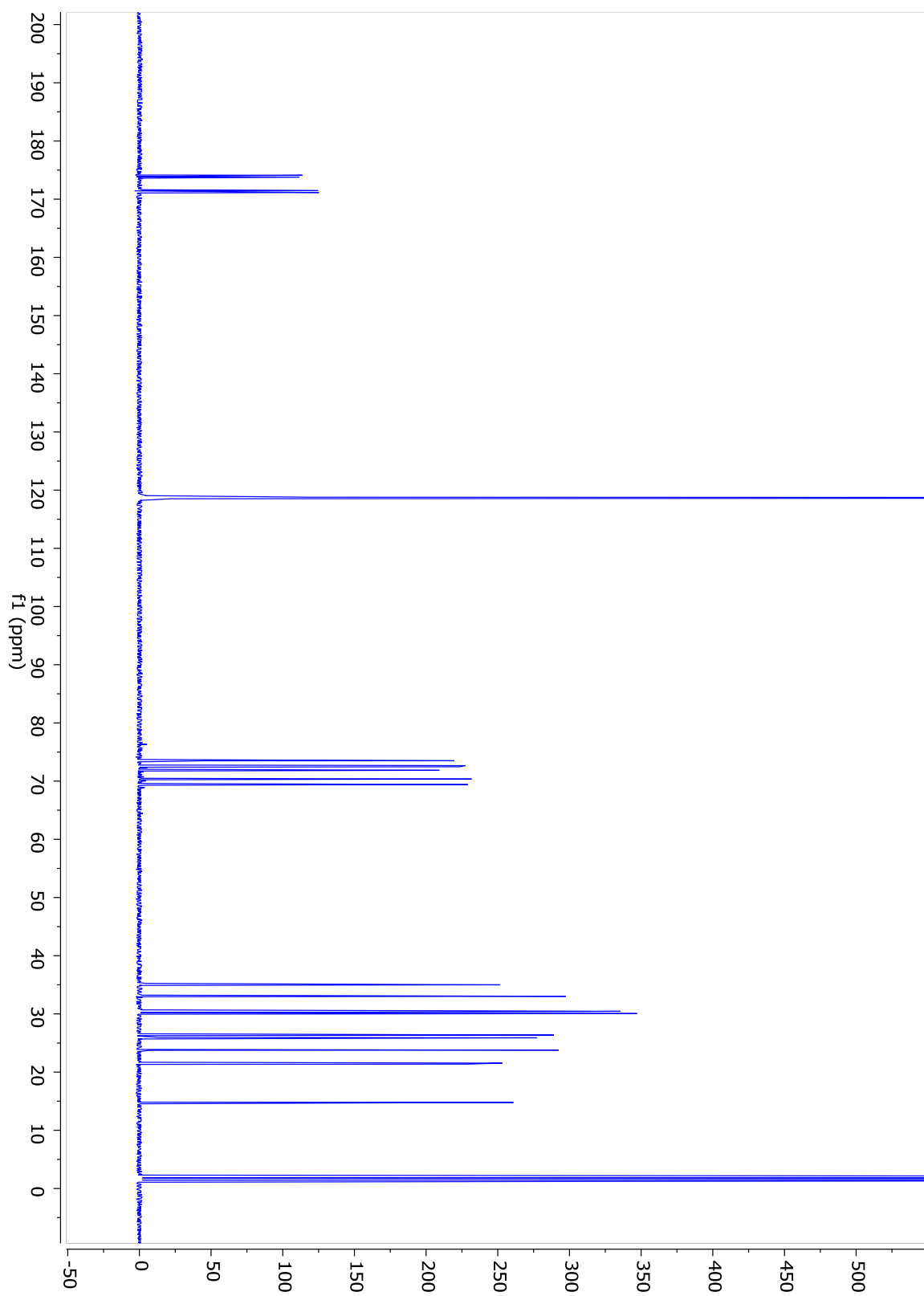
**Figure 3.71.** I4:24:0(2,2,10,10) $^{13}$C NMR
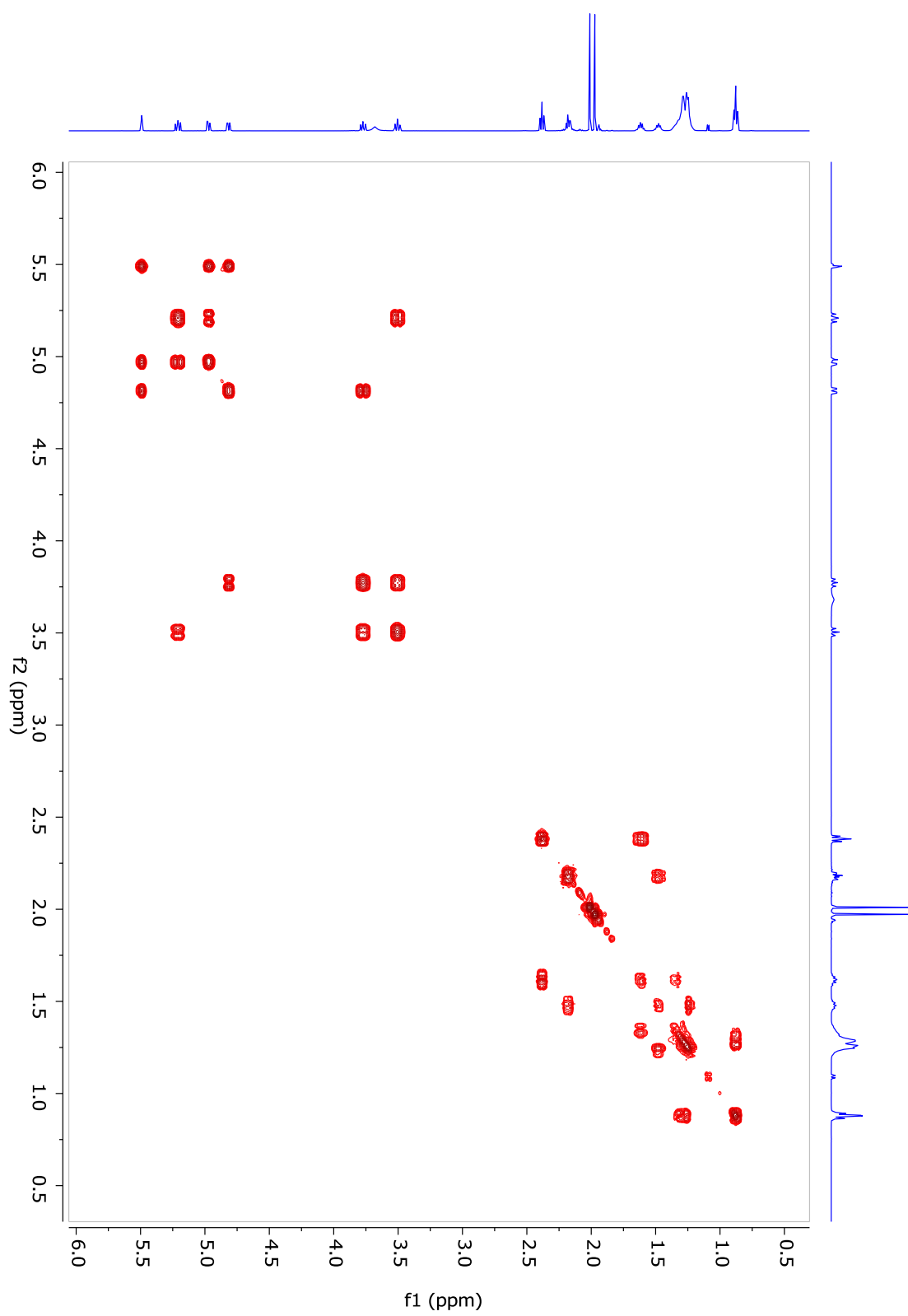
**Figure 3.72.** I4:24:0(2,2,10,10) $^1$H-$^1$H gCOSY

371
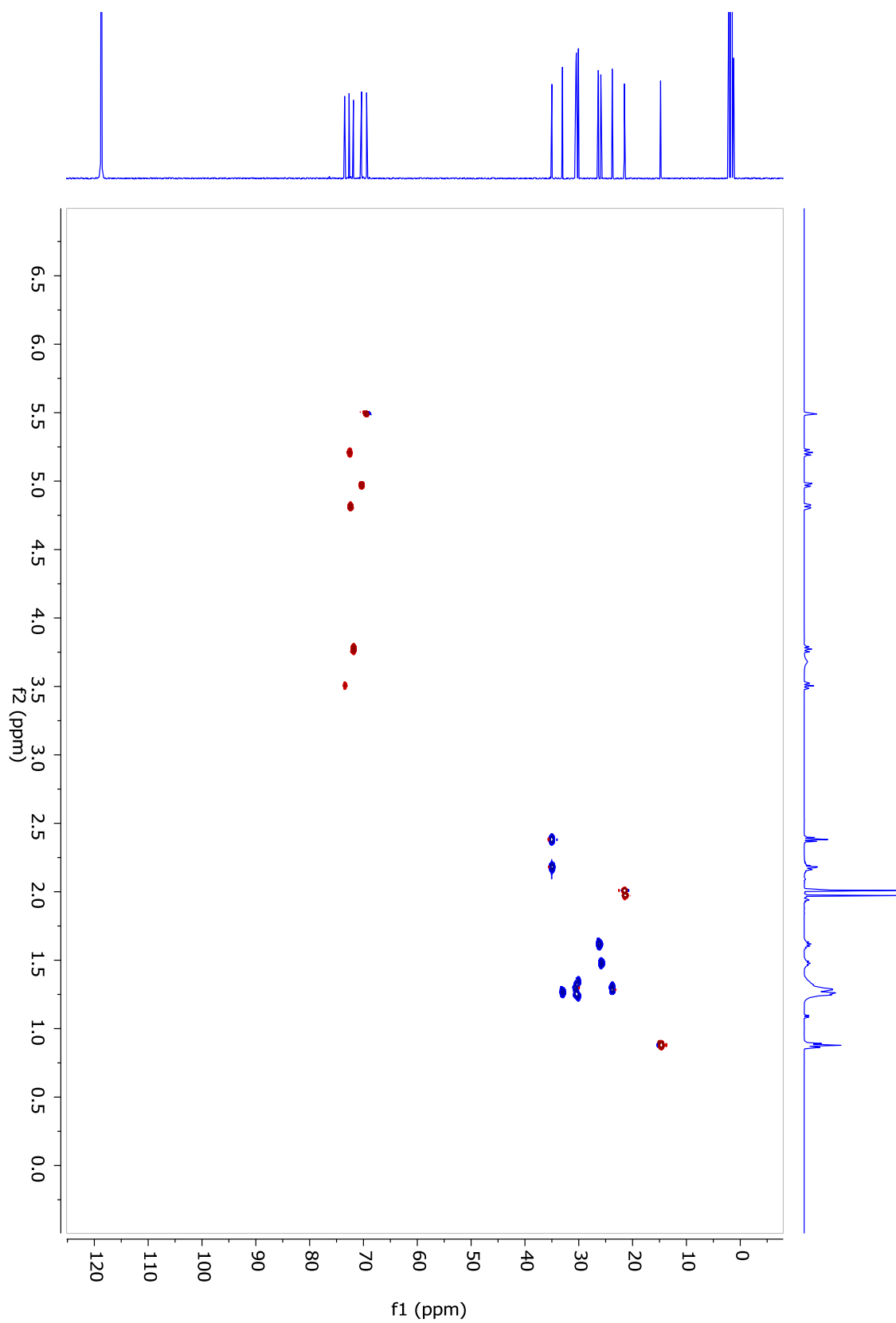
**Figure 3.73.** I4:24:0(2,2,10,10) gHSQC
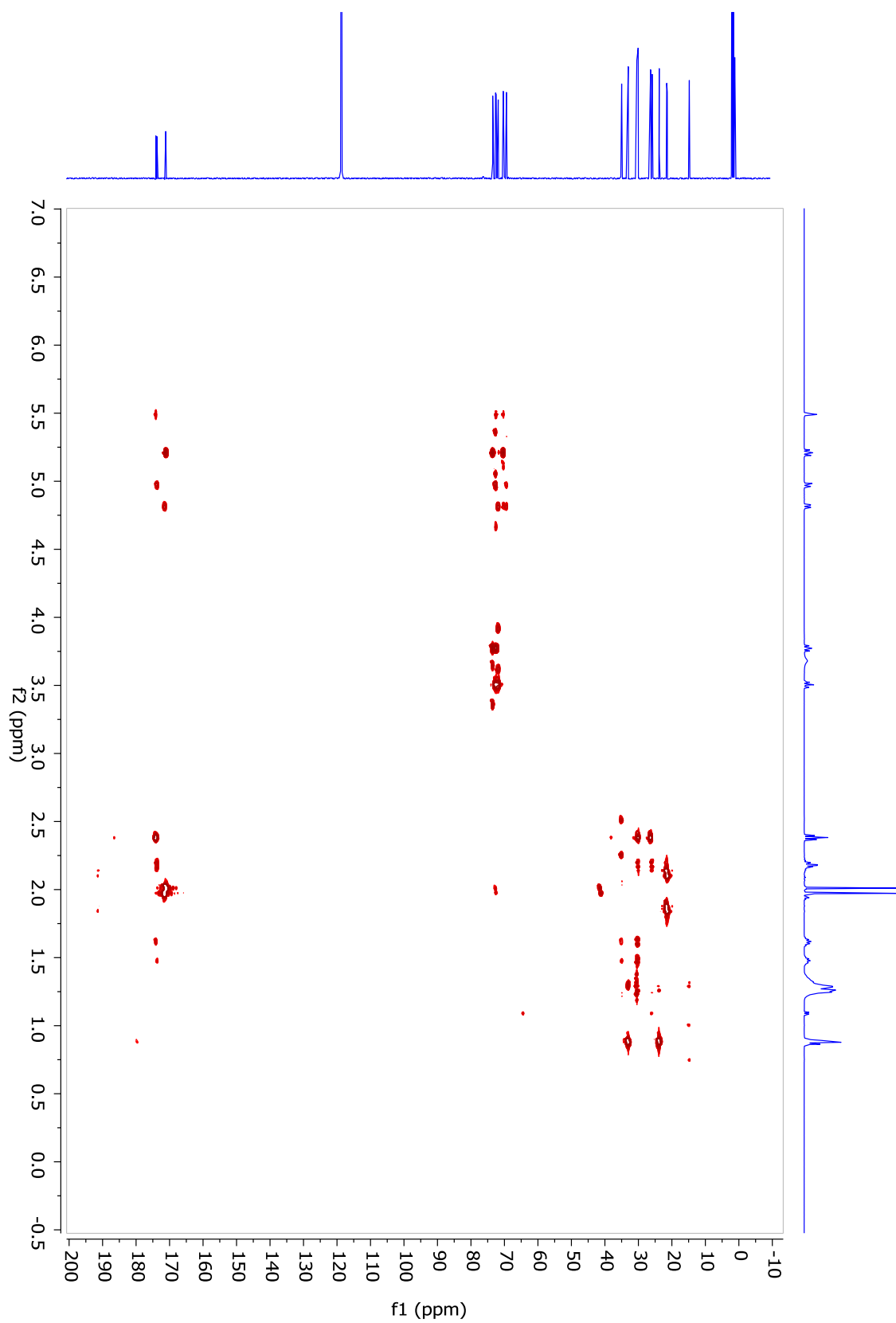
**Figure 3.74.** I4:24:0(2,2,10,10) gHMBC

373
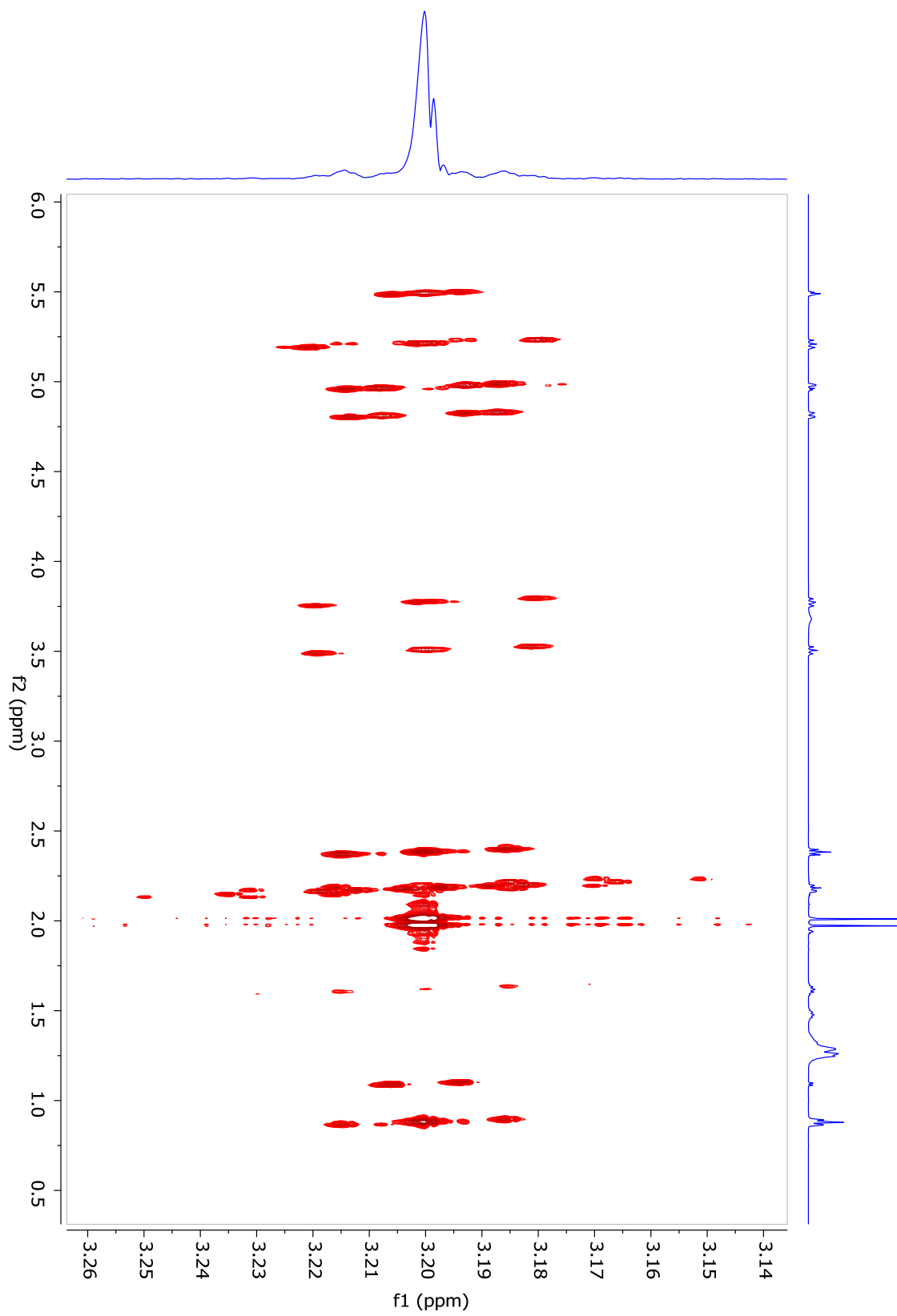
**Figure 3.75.** I4:24:0(2,2,10,10) *J*-resolved
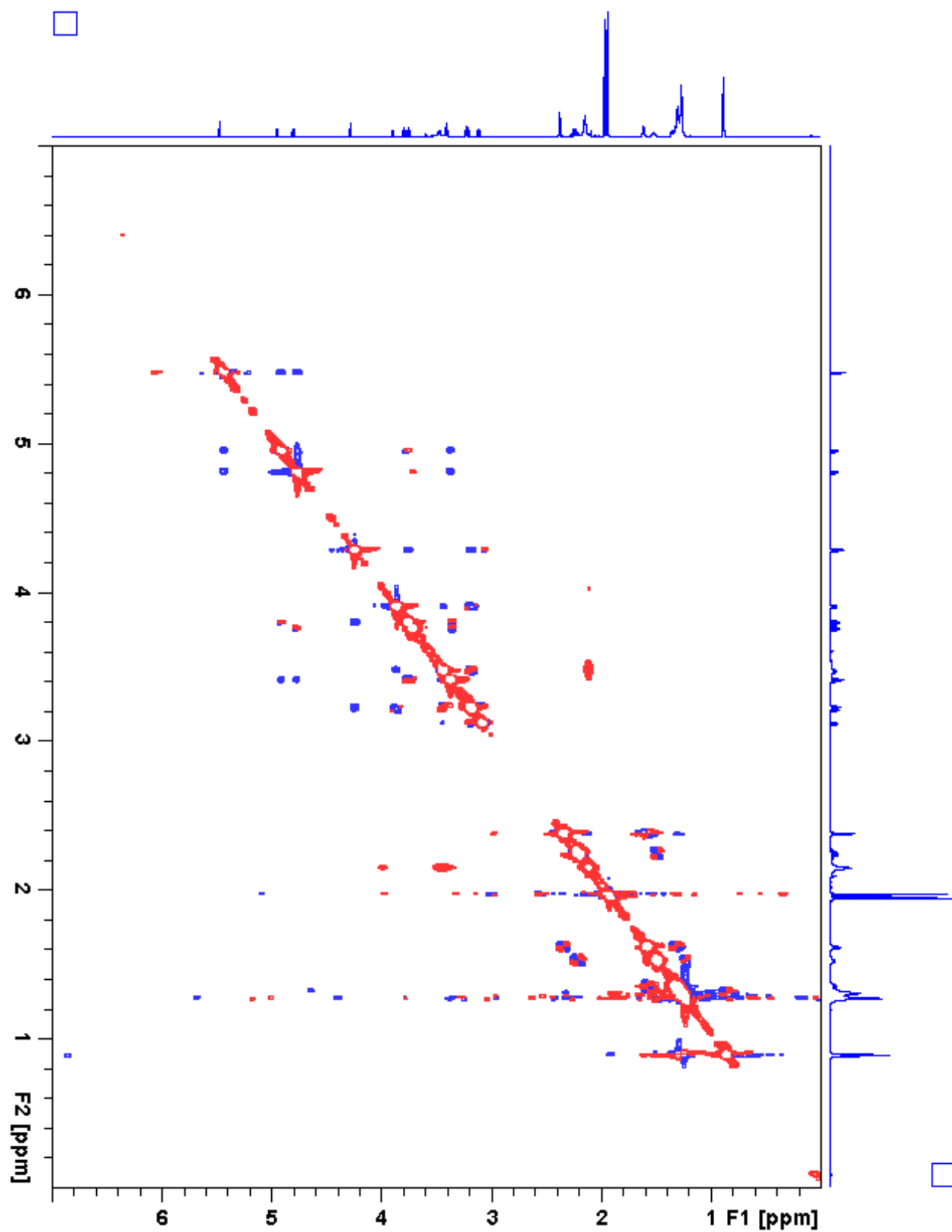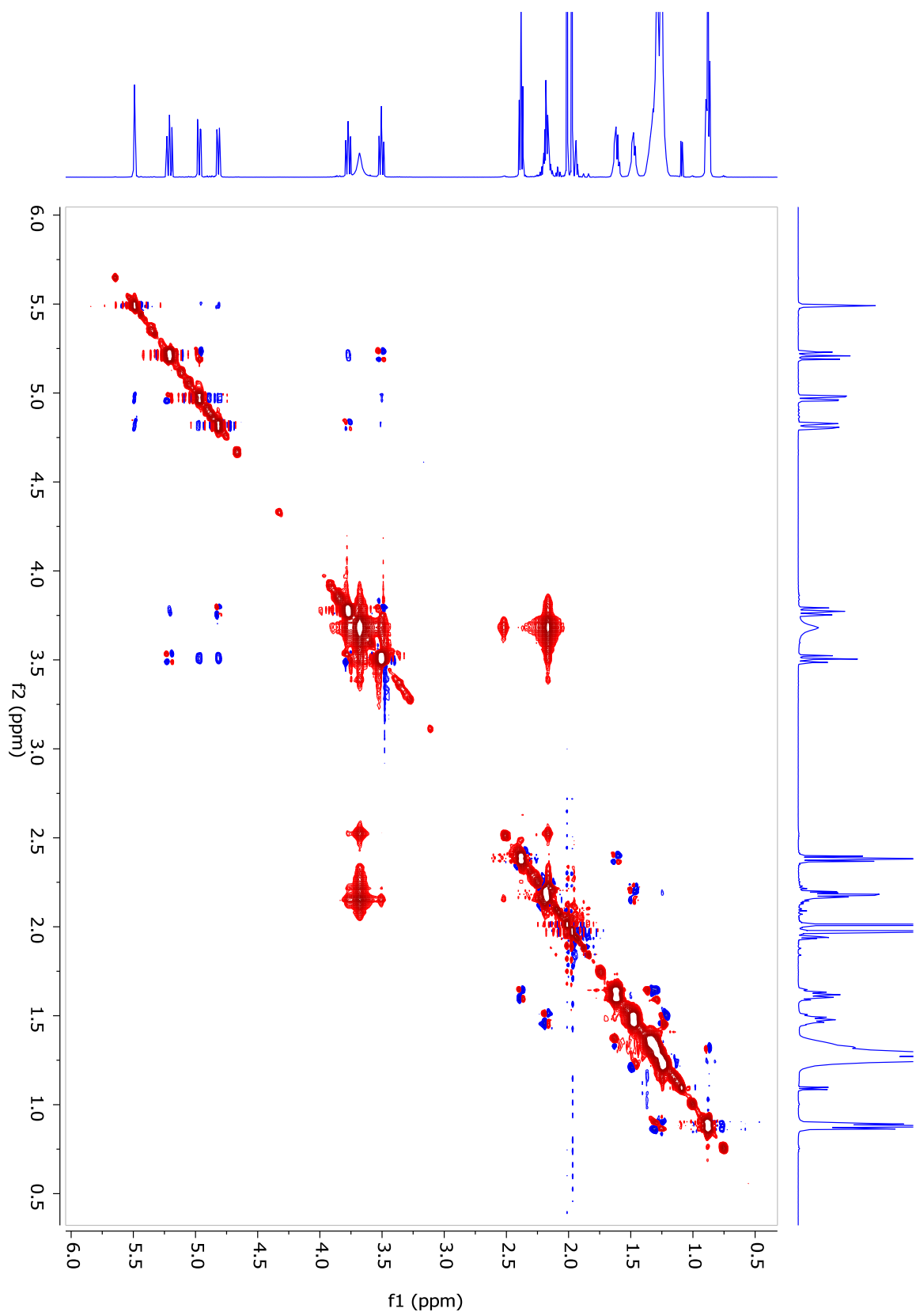
**Figure 3.76.** I4:24:0(2,2,10,10) ROESY

375
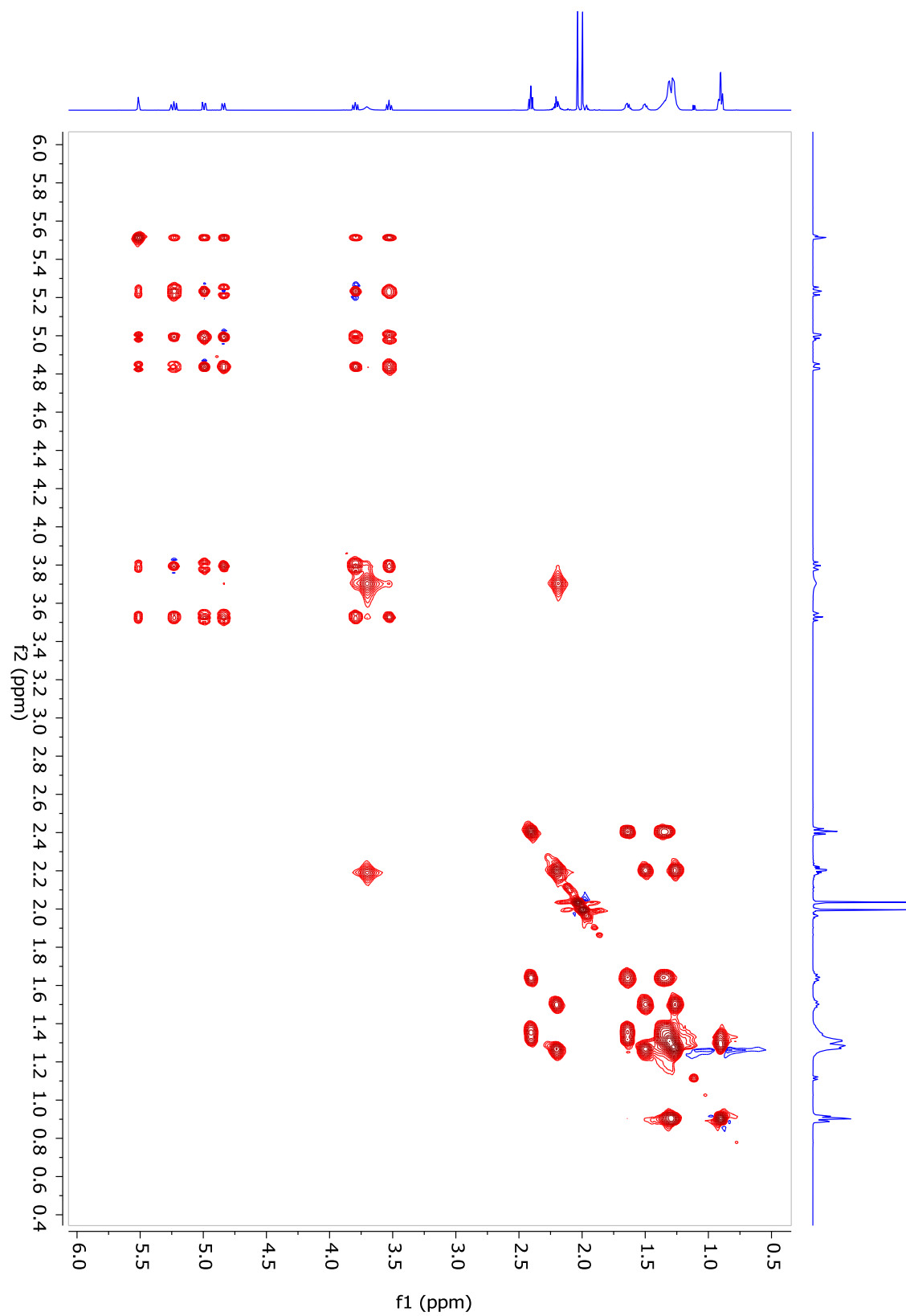
**Figure 3.77.** I4:24:0(2,2,10,10) TOCSY

**Table 3.23.** I3:24:0(2,10,12) Chemical shifts and coupling constants
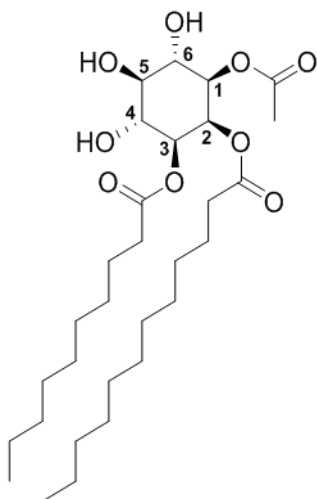


**Molecular Formula:** $C_{30}H_{54}O_9$

**110 min Retention Time (ESI-):** 75.14 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{31}H_{55}O_{11}^-$
([M+HCOO]$^-$): 603.3750, found: 603.3745

**110 min Retention Time (ESI+):** 75.13 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{30}H_{58}NO_9^+$ ([M+NH$_4$]$^+$):
576.4106, found: 576.4077

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #48
**Sample mass for NMR analysis:** 1.1 mg
**NMR Solvent:** D$_3$CN
**InChi Key:** NCEAPMSZCADWMM-XBSCKGQLSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.75 (dd, $J$ = 10.2, 3.0 Hz) | 72.84 |
| -1(CO) | | 171.49 |
| -2(CH$_3$) | 1.96 | 21.38 |
| **2**(CH) | 5.44 (t, $J$ = 2.9 Hz) | 69.61 |
| -1(CO) | | 174.1[a] |
| -2(CH$_2$) | 2.35 (t, $J$ = 7.3 Hz) | 35.1[b] |
| -3(CH$_2$) | 1.60 (p, $J$ = 7.3 Hz) | 26.30 |
| -4 to 11(CH$_2$) | 1.36-1.23 (m) | 30.8-30.1[c], 33.0[d], 23.78[e] |
| -12(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.77[f] |
| **3**(CH) | 4.77 (dd, $J$ = 10.2, 3.0 Hz) | 72.53 |
| -1(CO) | | 174.1[a] |
| -2(CH$_2$) | 2.24 (m) | 35.1[b] |
| -3(CH$_2$) | 1.53 (m) | 25.95 |
| -4 to 9(CH$_2$) | 1.36-1.23 (m) | 30.8-30.1[c], 33.0[d], 23.78[e] |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.77[f] |
| **4**(CH) | 3.68 (dd, $J$ = 10.5, 9.0 Hz) | 71.8[g] |
| **5**(CH) | 3.30 (t, $J$ = 9.2 Hz) | 75.70 |
| **6**(CH) | 3.68 (dd, J = 10.5, 9.0 Hz) | 71.8[g] |

a - Two $^{13}$C signals not resolved in 2D spectra (174.15, 174.13 ppm)

b - Two $^{13}$C signals not resolved in 2D spectra (35.09, 35.04 ppm)

c - $^{13}$C signals for CH$_2$ carbon positions 4 to 9, 4 to 7 (30.78, 30.74, 30.65, 30.56, 30.47, 30.42, 30.42, 30.42, 30.10, 30.05 ppm)

d - Two $^{13}$C signals for CH$_2$ carbon position 10 or 8 (33.03, 30.01 ppm)

e - Overlapping $^{13}$C signals for CH$_2$ carbon position 11 or 9

f - Overlapping $^{13}$C signals for CH$_3$ carbon position 12 or 10

g - Two $^{13}$C signals not resolved in 2D spectra (71.80, 71.75 ppm)
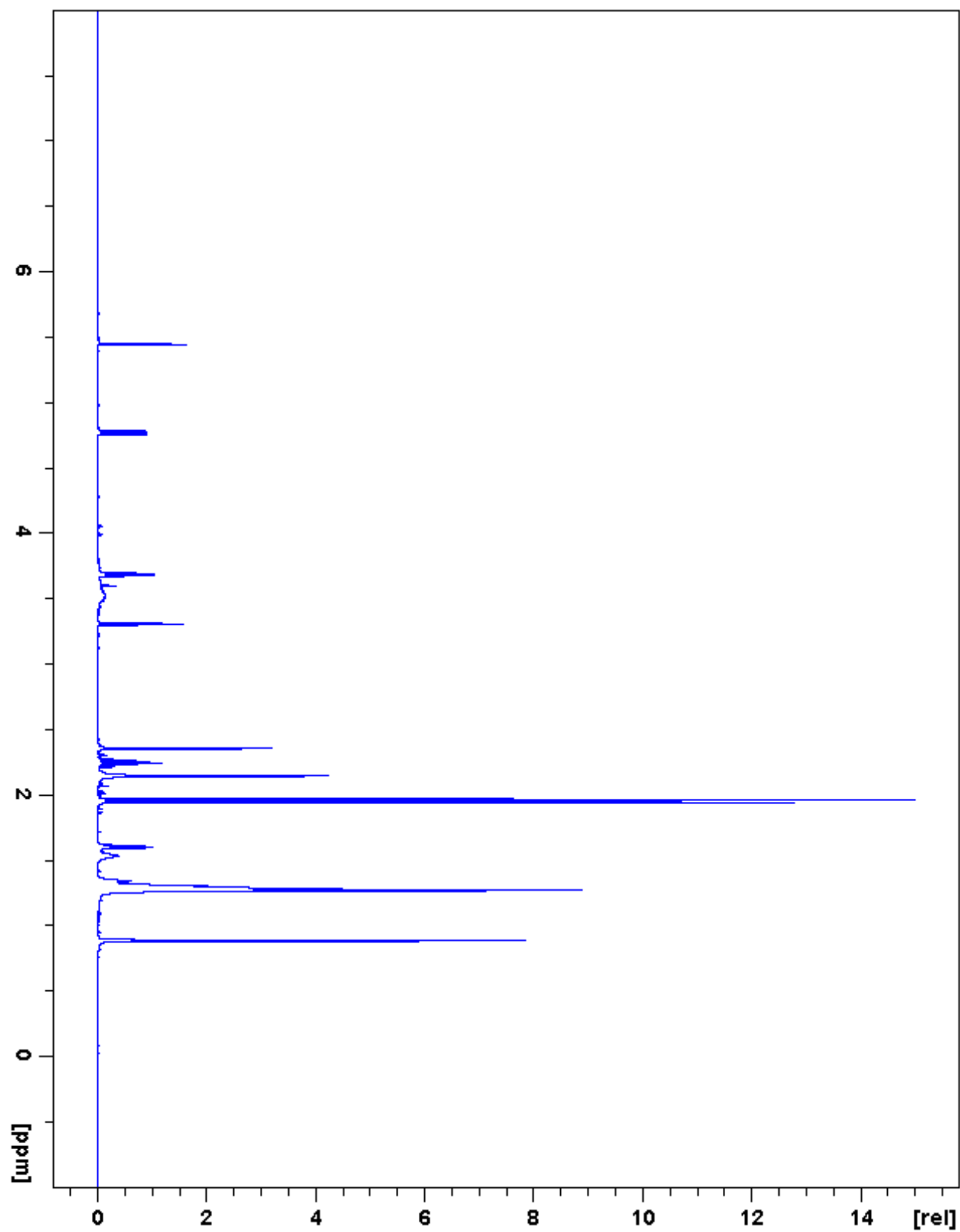
**Figure 3.78.** I3:24:0(2,10,12) ¹H NMR

378
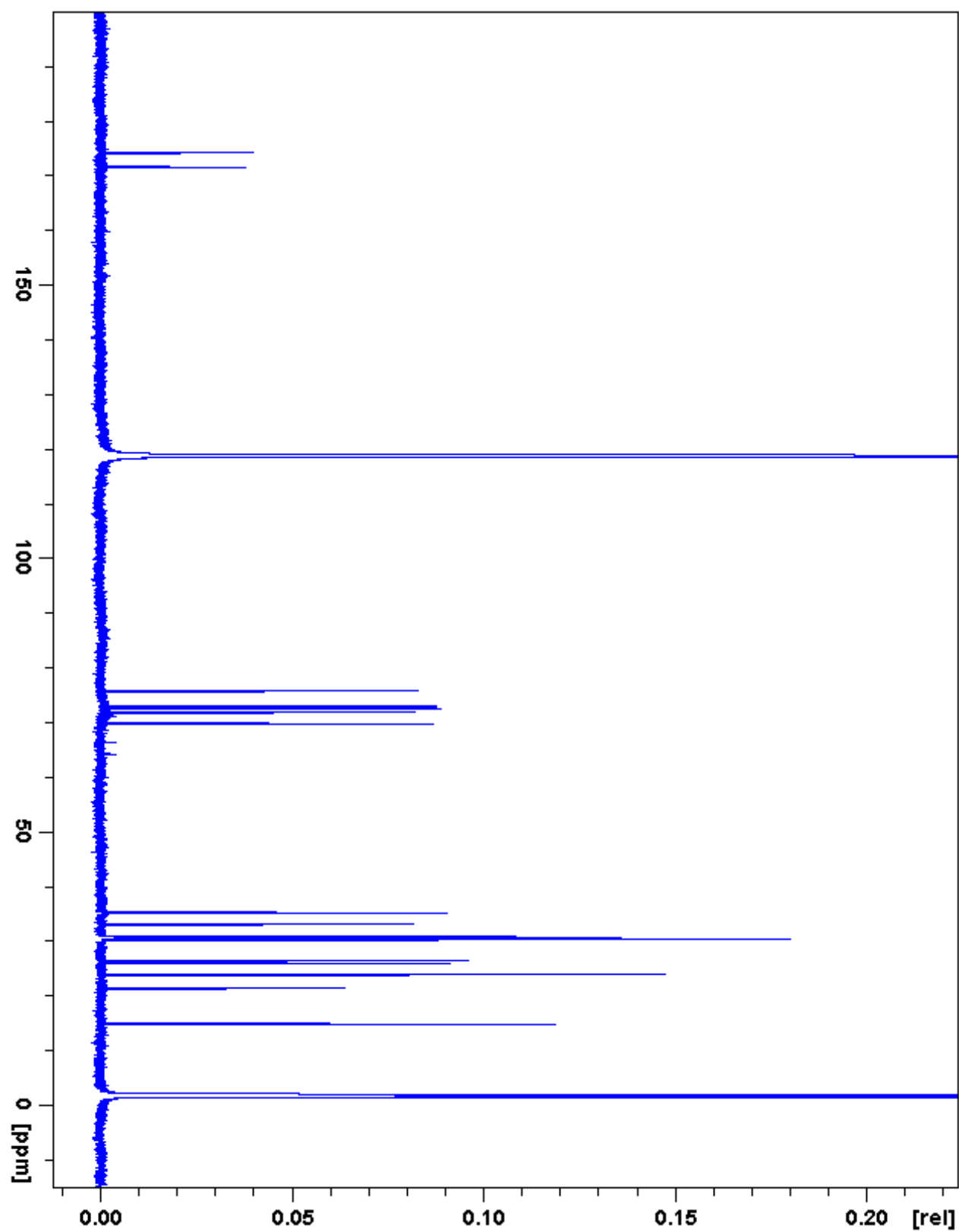
**Figure 3.79.** I3:24:0(2,10,12) $^{13}$C NMR
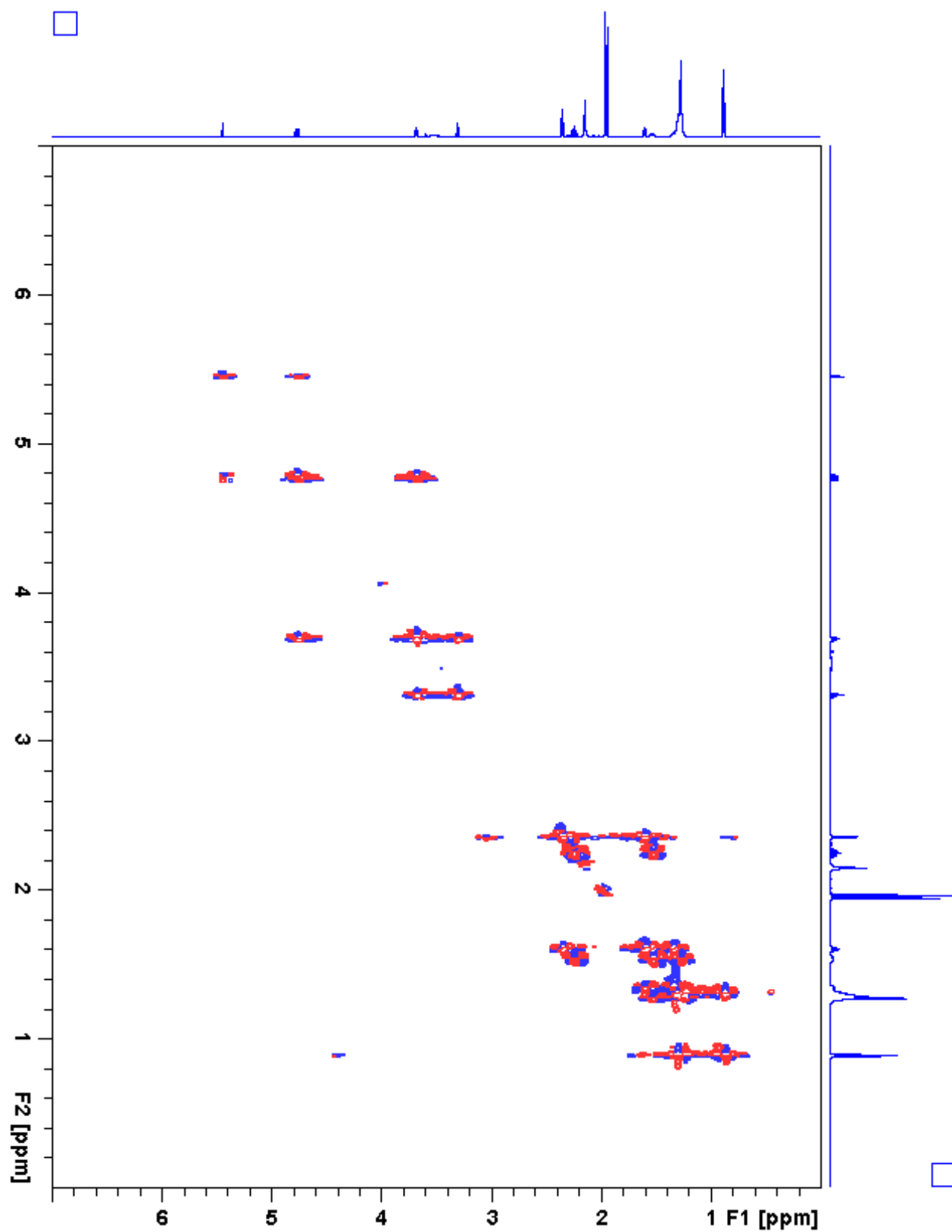
**Figure 3.80.** I3:24:0(2,10,12) $^1$H-$^1$H gCOSY
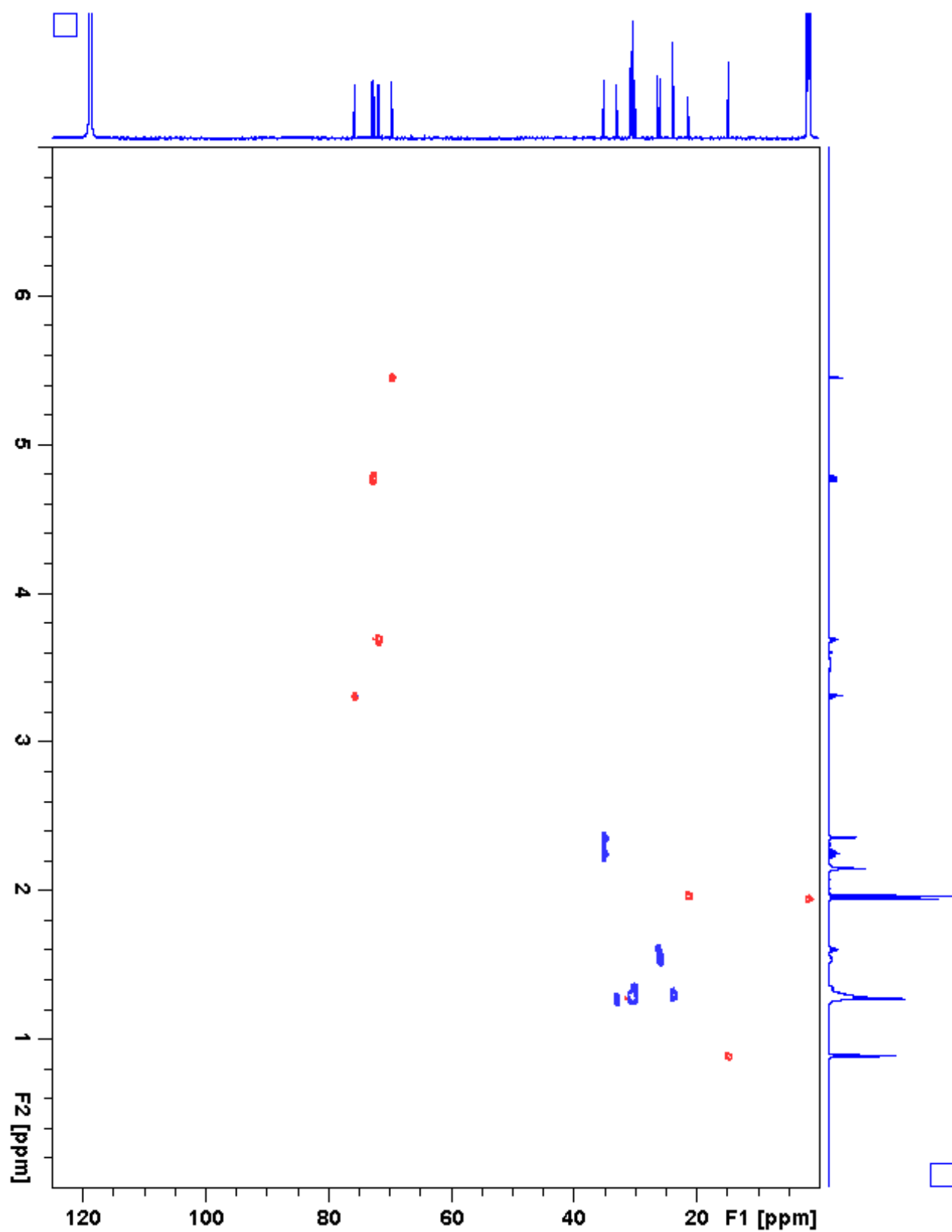
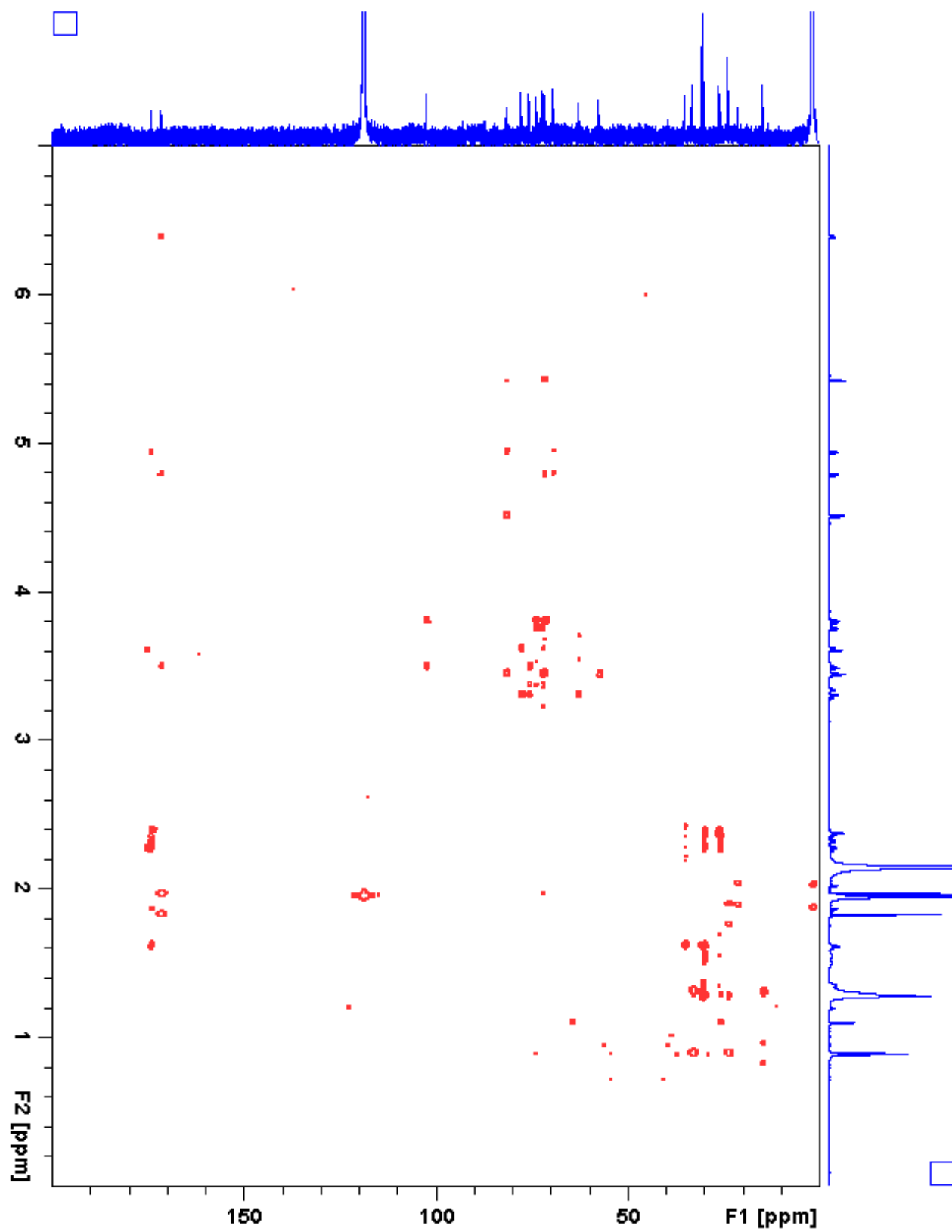**Figure 3.81.** I3:24:0(2,10,12) gHSQC
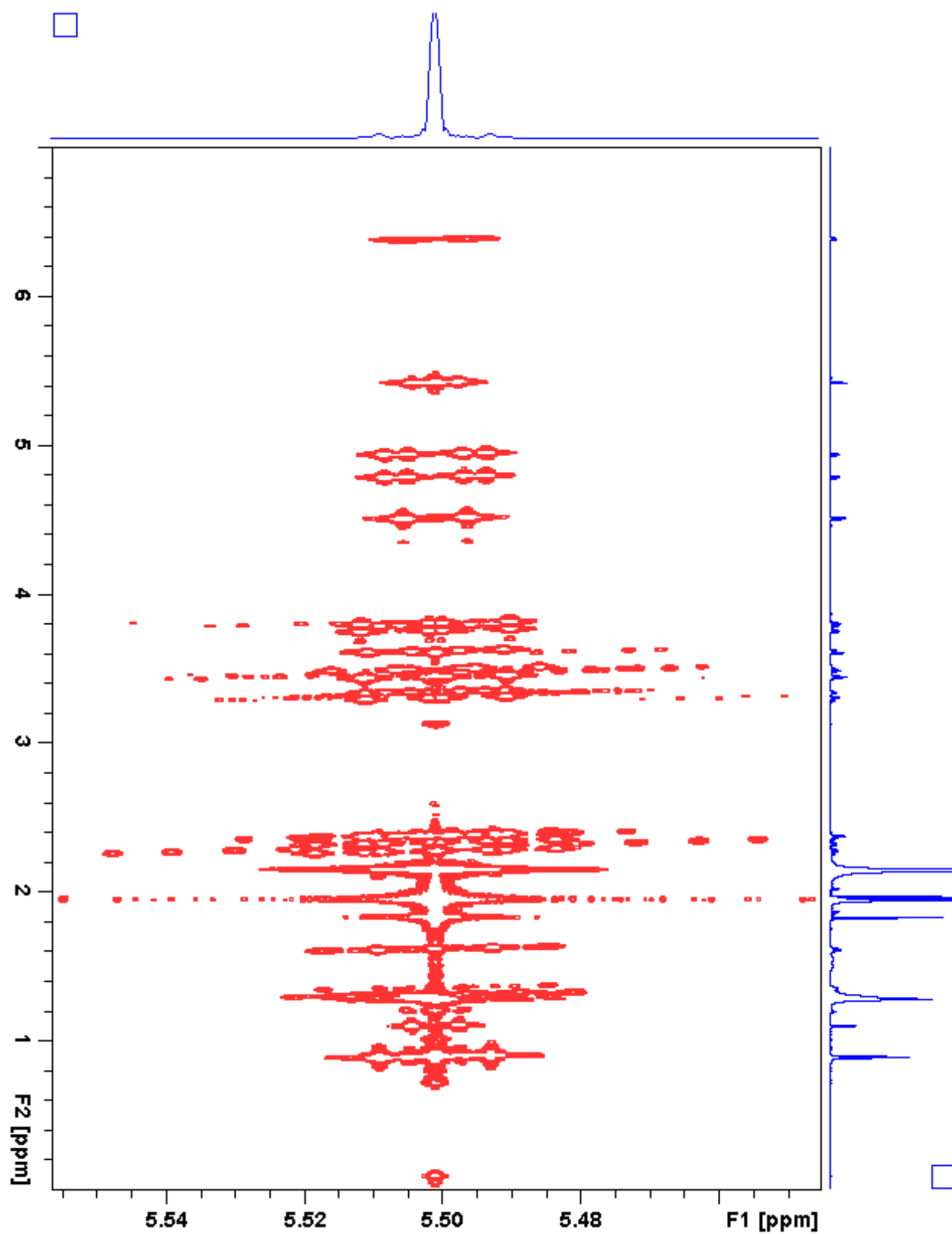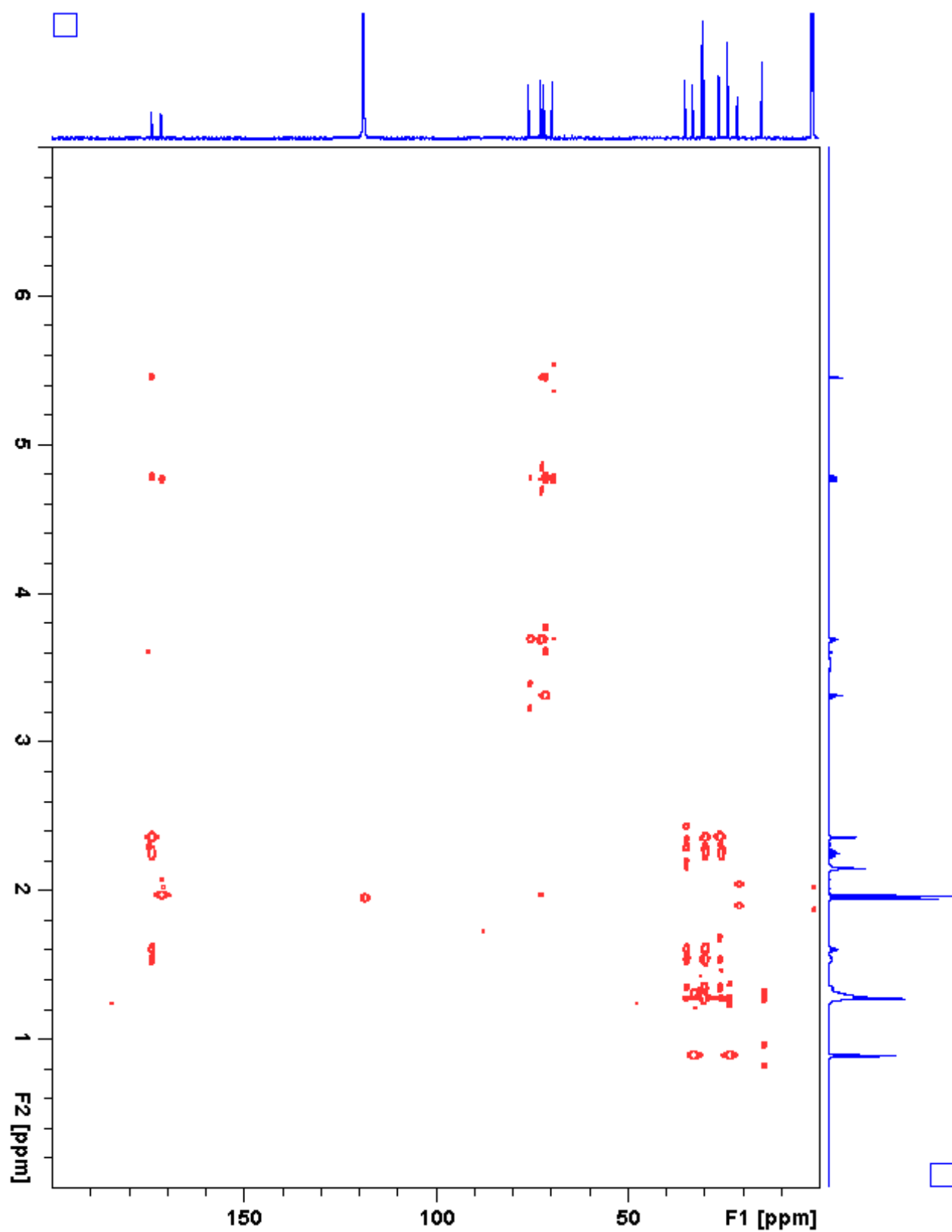
**Figure 3.82.** I3:24:0(2,10,12) gHMBC
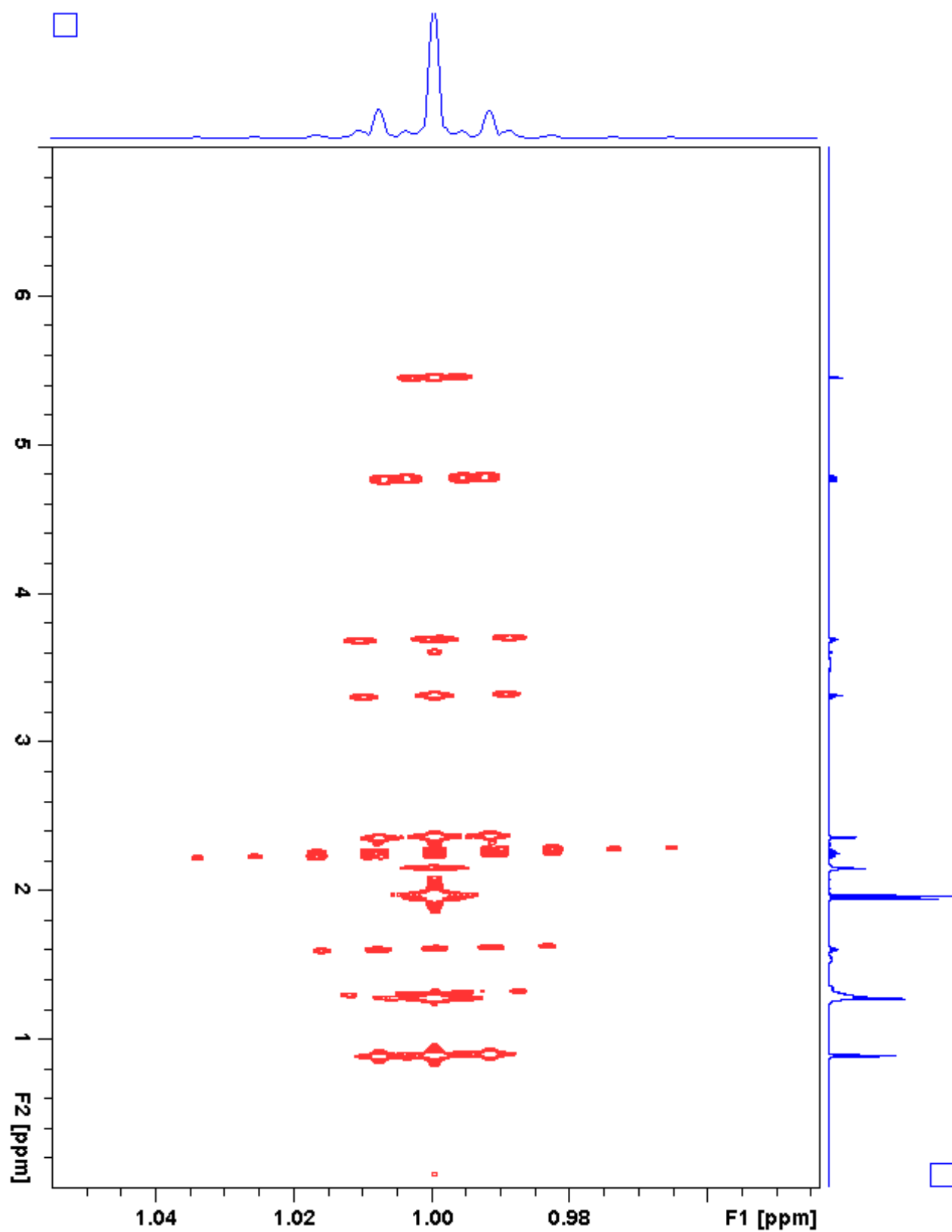
**Figure 3.83.** I3:24:0(2,10,12) *J*-resolved
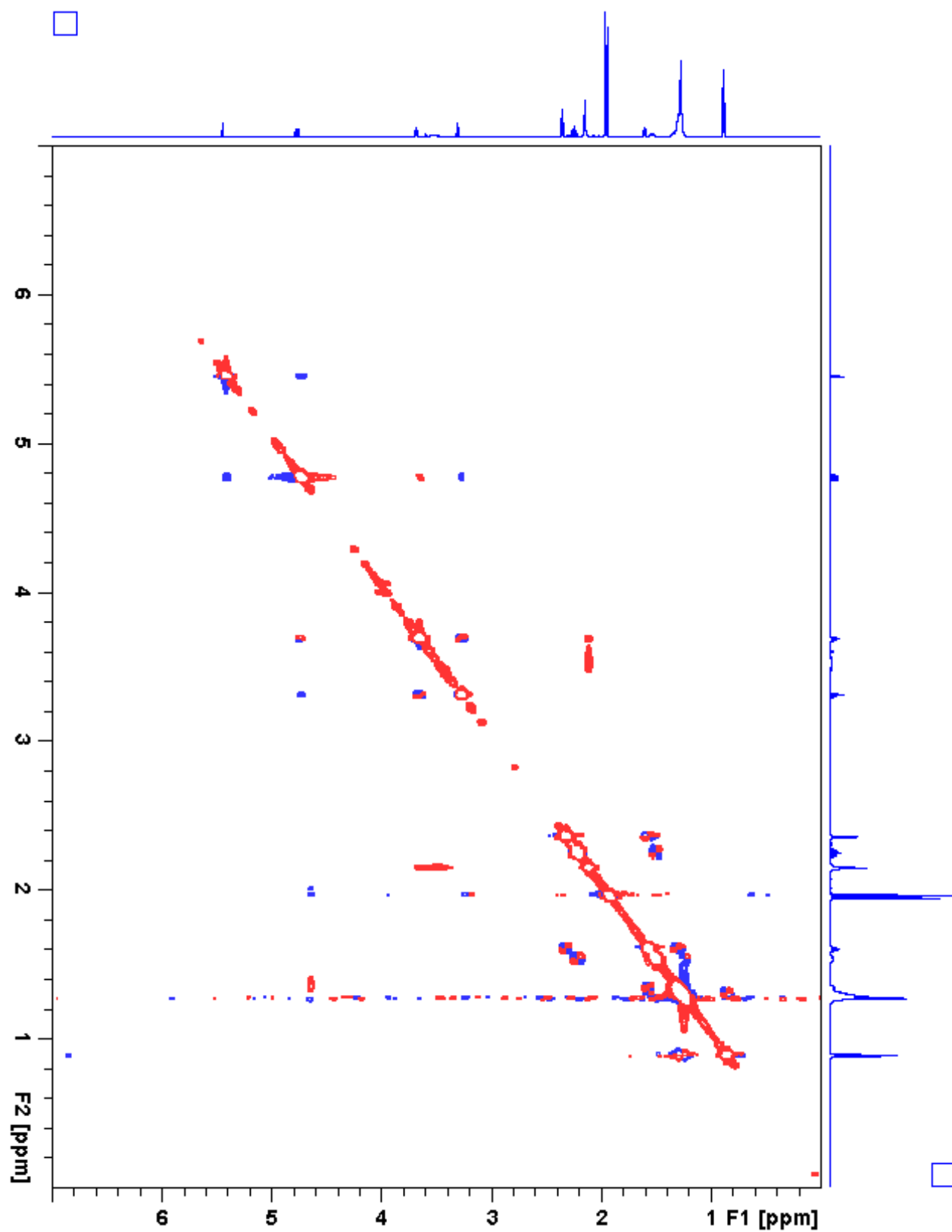
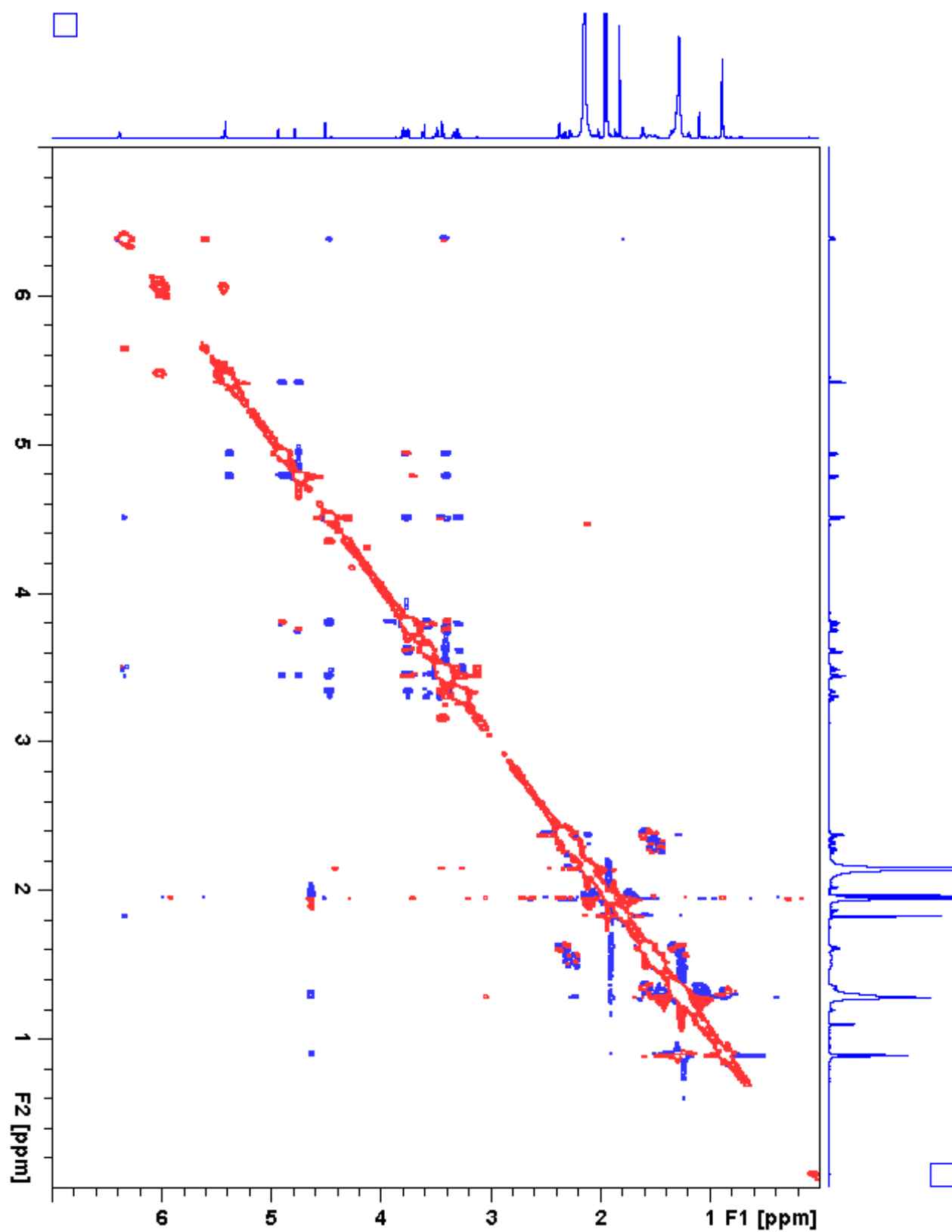**Figure 3.84.** I3:24:0(2,10,12) ROESY

**Table 3.24.** I4:26:0(2,2,10,12) Chemical shifts and coupling constants
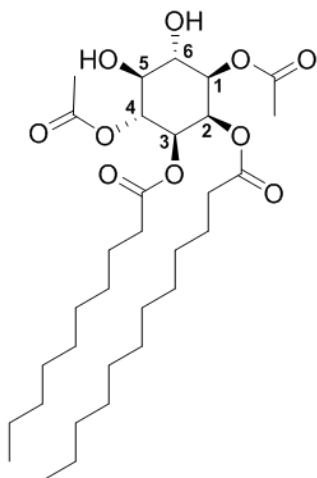


**Molecular Formula:** $C_{32}H_{56}O_{10}$

**110 min Retention Time (ESI-):** 80.27 mins
**HRMS:** (ESI-) $m/z$ calculated for $C_{33}H_{57}O_{12}^-$
([M+HCOO]$^-$): 645.3856, found: 645.3857

**110 min Retention Time (ESI+):** 80.26 mins
**HRMS:** (ESI+) $m/z$ calculated for $C_{32}H_{60}NO_{10}^+$ ([M+NH$_4$]$^+$):
618.4212, found: 618.4181

**Instrument:** Bruker Avance 900 MHz NMR
**Fraction:** #53
**Sample mass for NMR analysis:** 1.5 mg
**NMR Solvent:** D$_3$CN
**InChi Key:** RAWFQSLFZCSMTH-YBYHVSLFSA-N

| Carbon # (group) | $^1$H (ppm) | $^{13}$C (ppm) |
|---|---|---|
| **1**(CH) | 4.81 (dd, $J$ = 10.2, 3.0 Hz) | 72.48 |
| -1(CO) | | 171.46 |
| -2(CH$_3$) | 1.97 | 21.36 |
| **2**(CH) | 5.49 (t, $J$ = 2.9 Hz) | 69.42 |
| -1(CO) | | 174.11 |
| -2(CH$_2$) | 2.38 (m) | 35.07 |
| -3(CH$_2$) | 1.62 (p, $J$ = 7.3 Hz) | 26.36 |
| -4 to 11(CH$_2$) | 1.37-1.21 (m) | 30.8-30.1[a], 33.0[b], 23.8[c] |
| -12(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.77[d] |
| **3**(CH) | 4.97 (dd, $J$ = 10.6, 3.0 Hz) | 70.36 |
| -1(CO) | | 173.80 |
| -2(CH$_2$) | 2.18 (m) | 35.00 |
| -3(CH$_2$) | 1.48 (p, $J$ = 7.5 Hz) | 25.89 |
| -4 to 9(CH$_2$) | 1.37-1.21 (m) | 30.8-30.1[a], 33.0[b], 23.8[c] |
| -10(CH$_3$) | 0.88 (t, $J$ = 7.2 Hz) | 14.77[f] |
| **4**(CH) | 5.21 (dd, $J$ = 10.8, 9.4 Hz) | 72.67 |
| -1(CO) | | 171.14 |
| -2(CH$_3$) | 2.01 | 21.54 |
| **5**(CH) | 3.50 (t, $J$ = 9.5 Hz) | 73.52 |
| **6**(CH) | 3.77 (dd, J = 10.5, 9.0 Hz) | 71.87 |

a - $^{13}$C signals for CH$_2$ carbon positions 4 to 9, 4 to 7 (30.78, 30.75, 30.64, 30.51, 30.47, 30.44, 30.41, 30.05, 30.05, 30.05 ppm)

b - Two $^{13}$C signals for CH$_2$ carbon position 10 or 8 (33.04, 32.99 ppm)

c - Two $^{13}$C signals for CH$_2$ carbon position 11 or 9 (23.78, 23.77 ppm)

d - Overlapping $^{13}$C signals for CH$_3$ carbon position 12 or 10

**Figure 3.85.** I4:26:0(2,2,10,12) [^1]H NMR

386
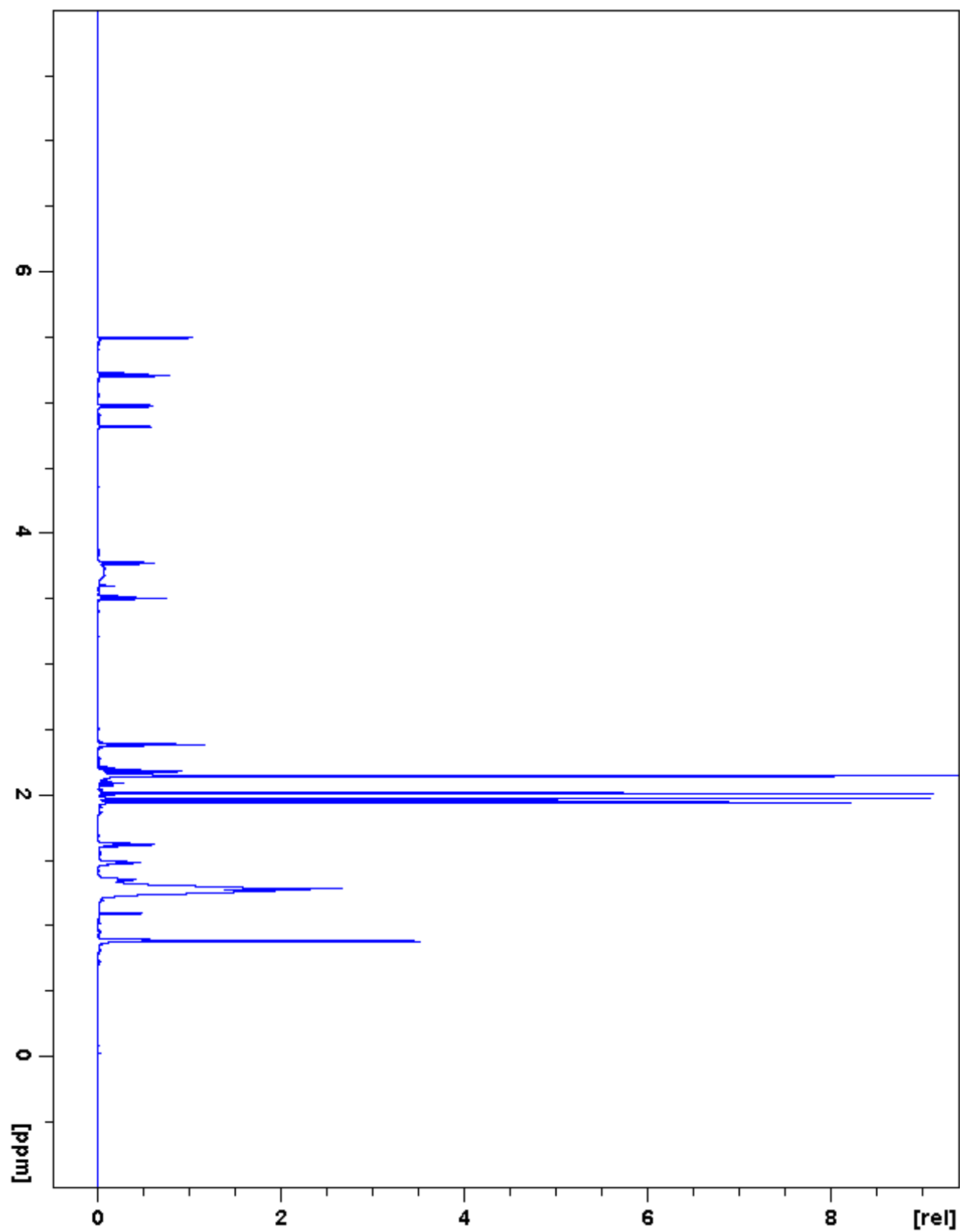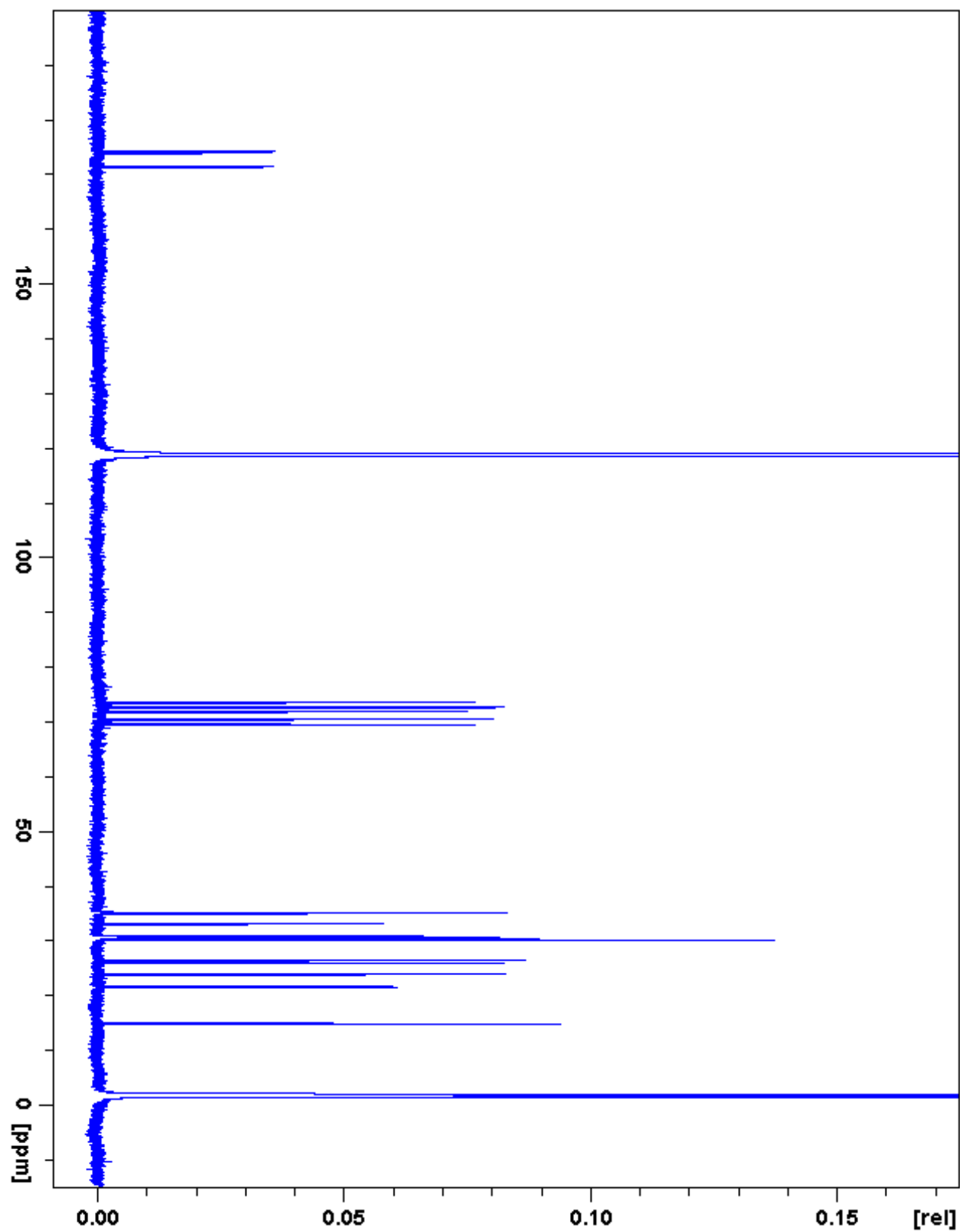
**Figure 3.86.** I4:26:0(2,2,10,12) $^{13}$C NMR
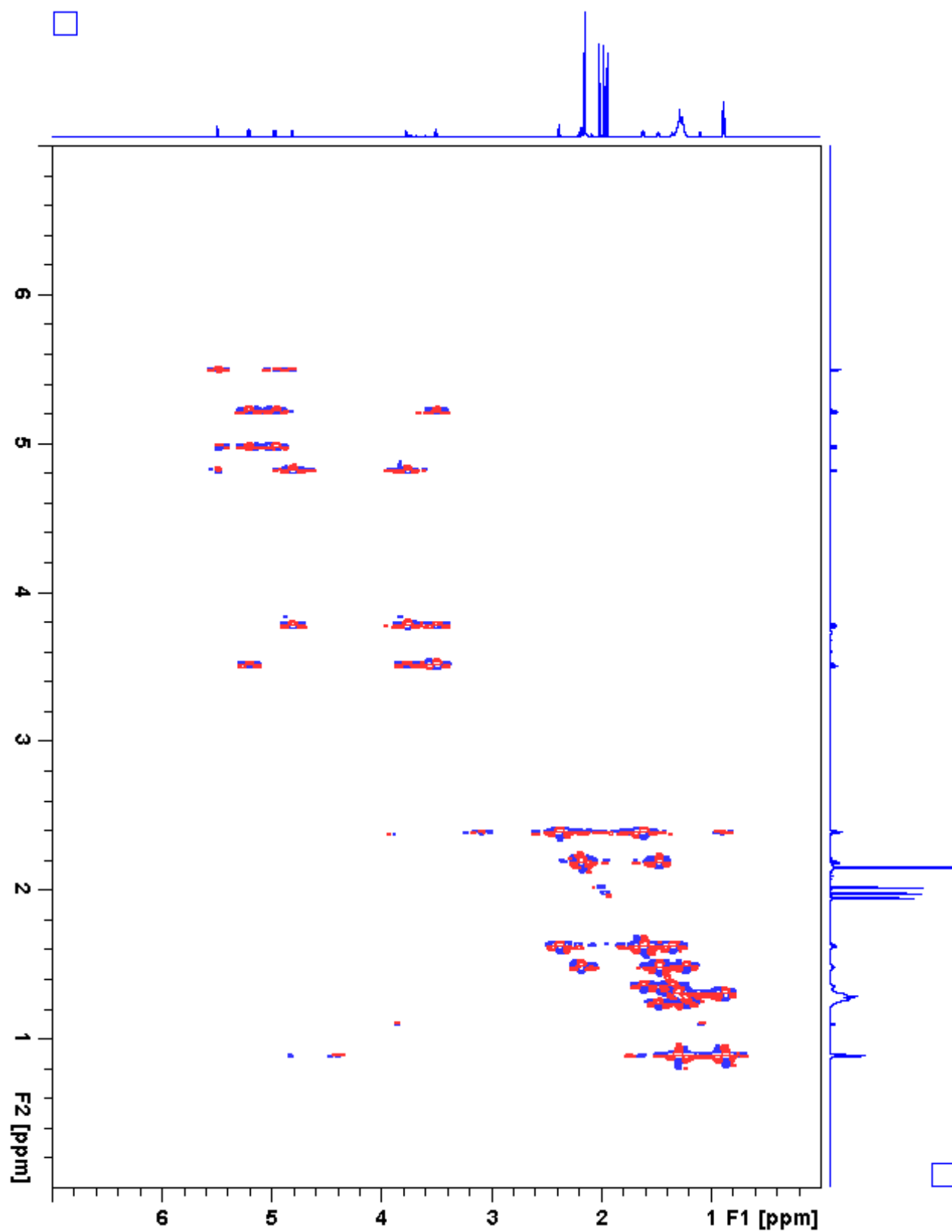
**Figure 3.87.** I4:26:0(2,2,10,12) $^1$H-$^1$H gCOSY
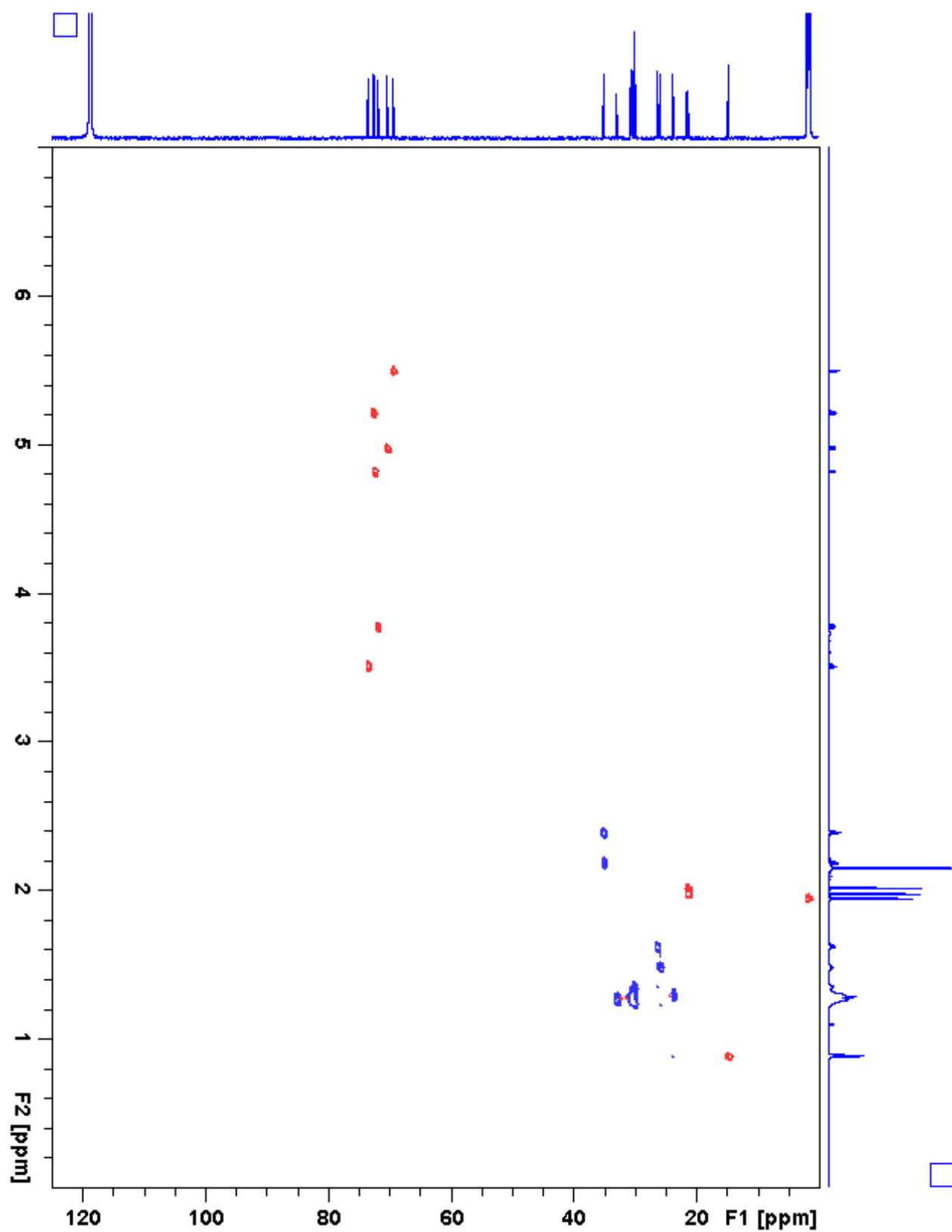
**Figure 3.88.** I4:26:0(2,2,10,12) gHSQC
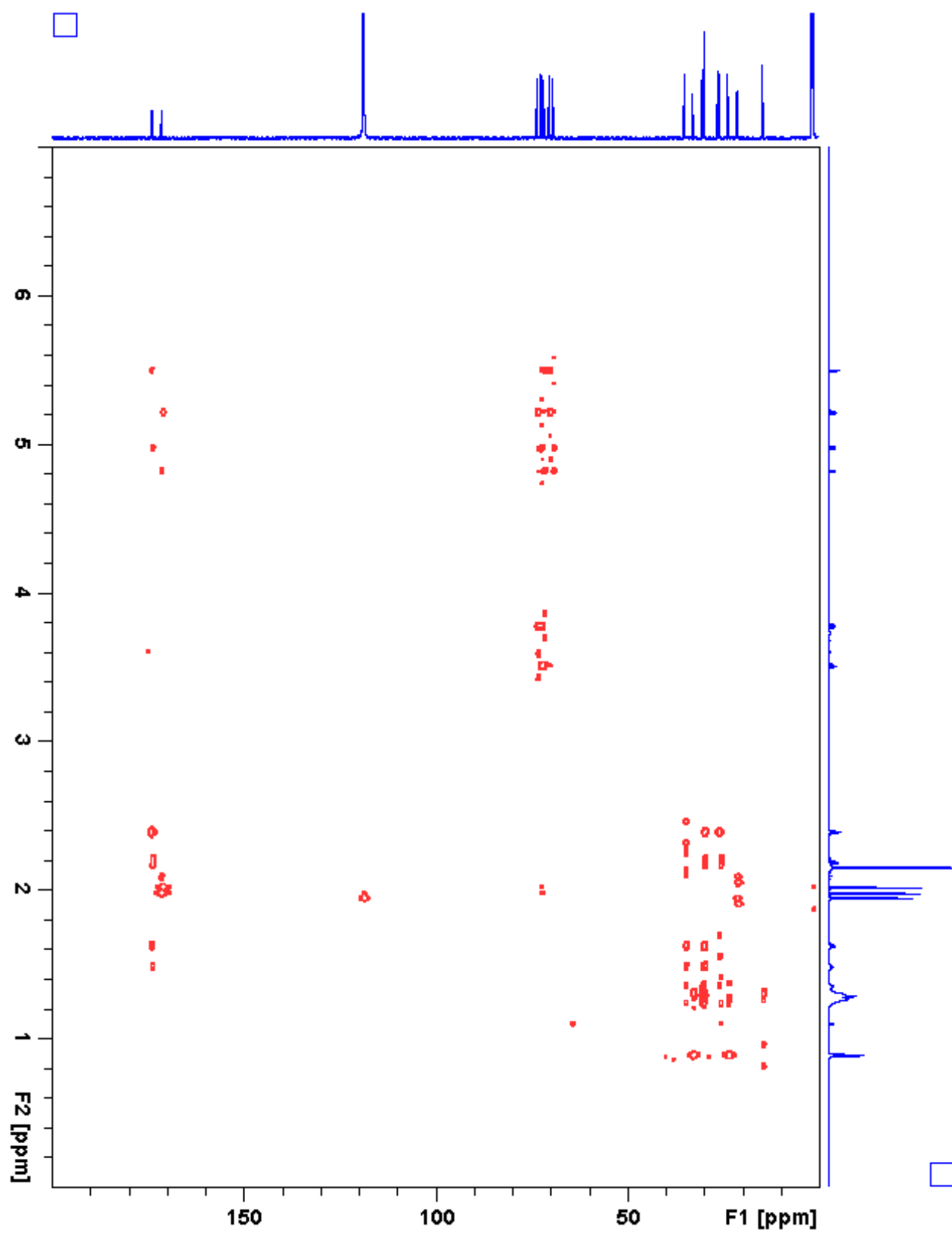
**Figure 3.89.** I4:26:0(2,2,10,12) gHMBC
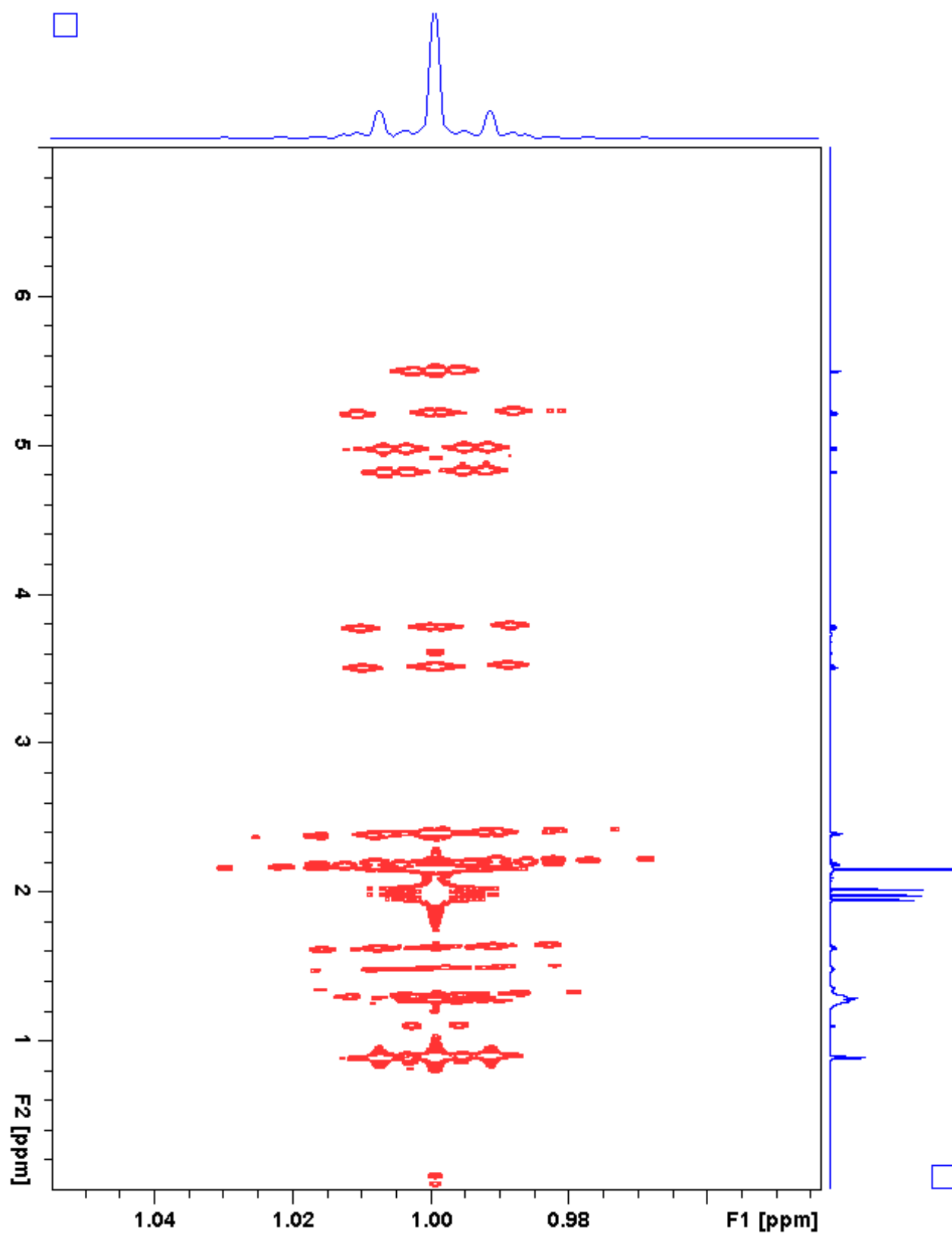
**Figure 3.90.** I4:26:0(2,2,10,12) *J*-resolved
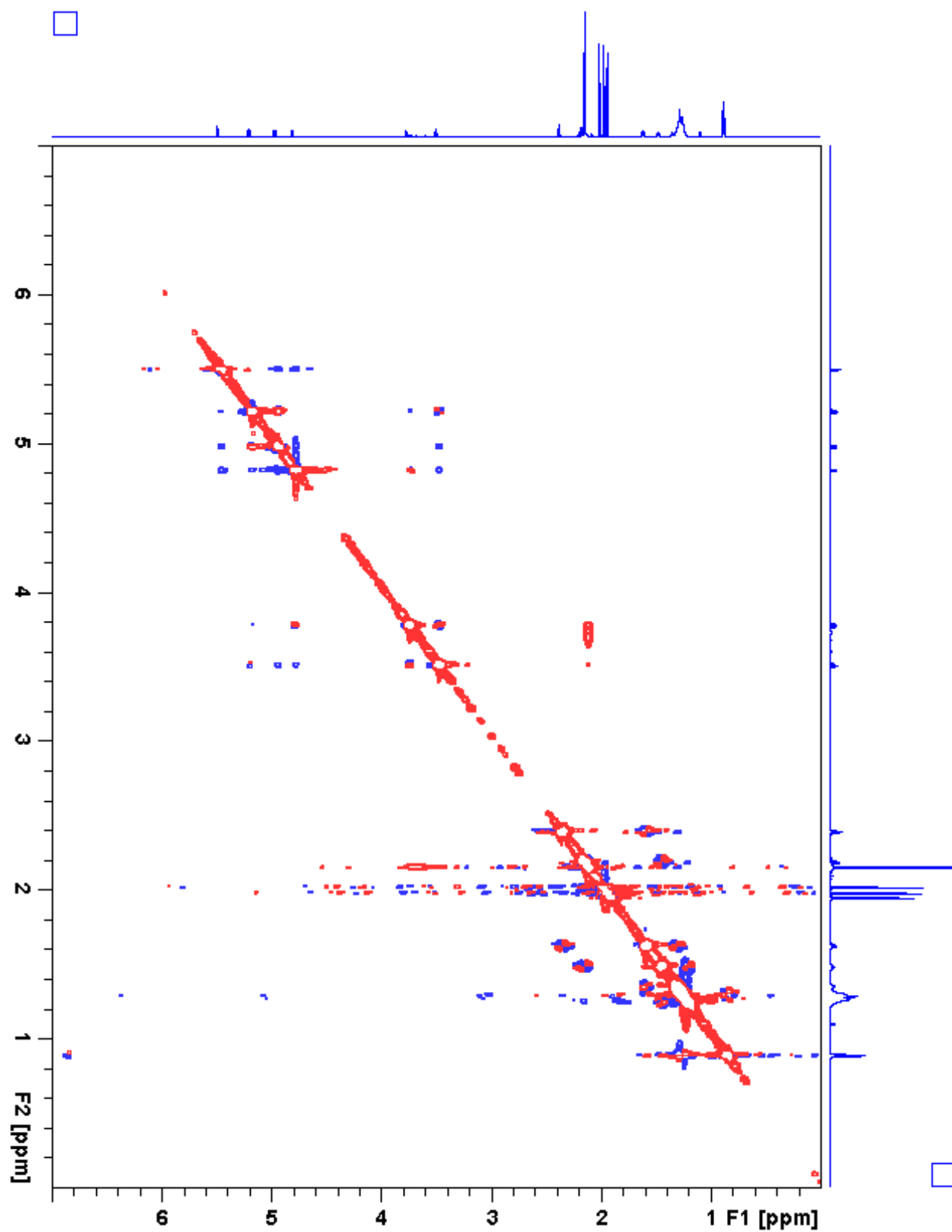
**Figure 3.91.** I4:26:0(2,2,10,12) ROESY
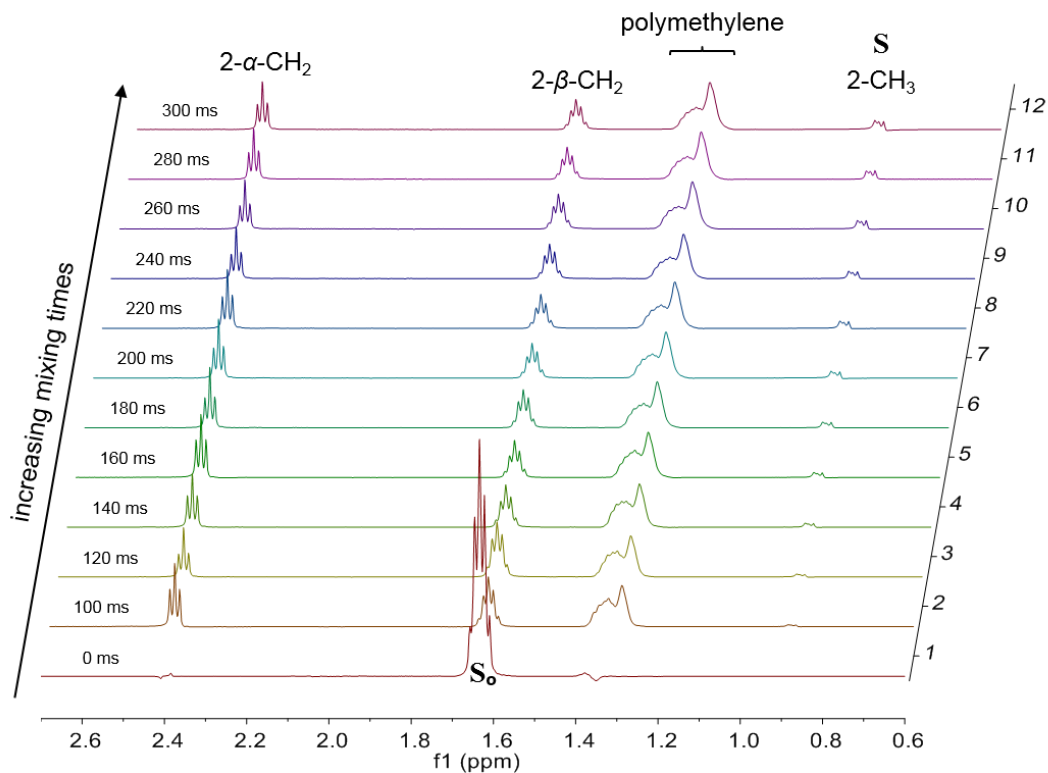
**Figure 3.92.** I4:24:0(2,2,10,10) 1D-TOCSY transfer spectra 2-β-CH₂ excitation at 1.62 ppm (generated using Varian Inova 600 MHz spectrometer).

**Table 3.25.** I4:24:0(2,2,10,10) 1D-TOCSY transfer 2-β-CH₂ (S$_o$) and 2-CH₃ (S) normalized integrals.

| mixing time (ms) | 2-β-CH₂ excitation at 1.62 ppm | 2-CH₃ S/S$_o$ at 0.88 ppm |
|---|---|---|
| 0 | 1.0000 | -0.0023 |
| 100 | 0.1992 | 0.0065 |
| 120 | 0.2397 | 0.0092 |
| 140 | 0.1706 | 0.0134 |
| 160 | 0.1441 | 0.0179 |
| 180 | 0.1607 | 0.0213 |
| 200 | 0.1445 | 0.0243 |
| 220 | 0.1380 | 0.0265 |
| 240 | 0.1427 | 0.0278 |
| 260 | 0.1498 | 0.0332 |
| 280 | 0.1330 | 0.0337 |
| 300 | 0.1249 | 0.0326 |

393
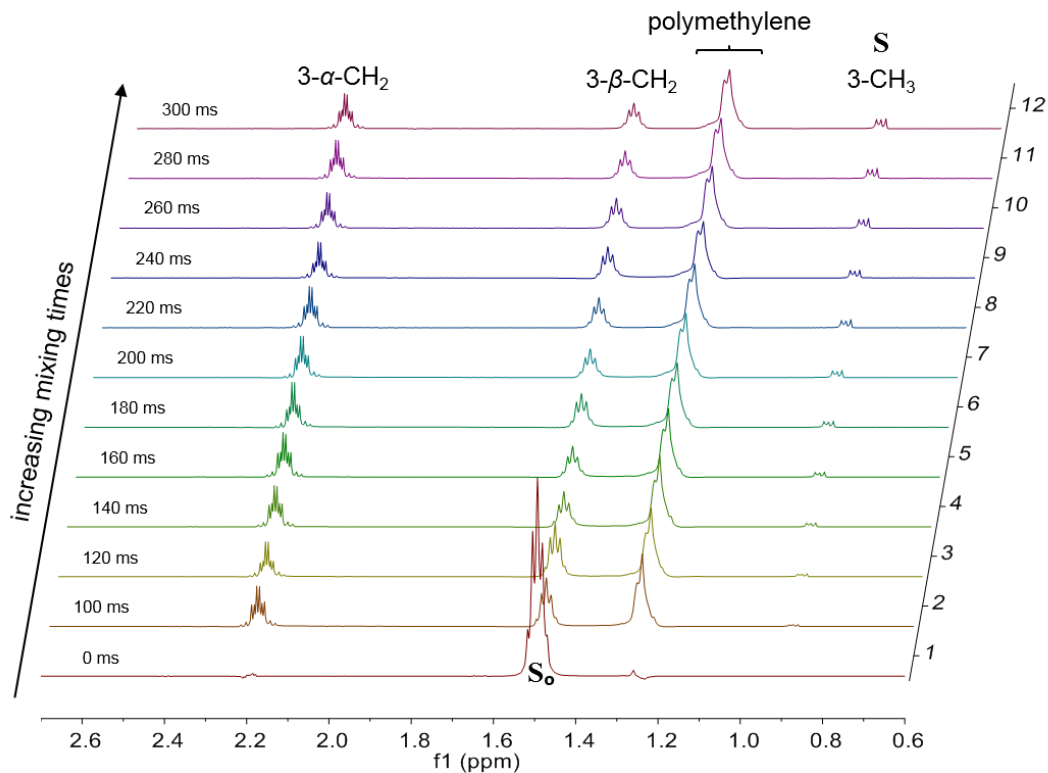
**Figure 3.93.** I4:24:0(2,2,10,10) 1D-TOCSY transfer spectra 3-$\beta$-CH$_2$ excitation at 1.48 ppm (generated using Varian Inova 600 MHz spectrometer).

**Table 3.26.** I4:24:0(2,2,10,10) 1D-TOCSY transfer 3-$\beta$-CH$_2$ (S$_o$) and 3-CH$_3$ (S) normalized integrals.

| mixing time (ms) | 3-$\beta$-CH$_2$ excitation at 1.48 ppm | 2-CH$_3$ S/S$_0$ at 0.88 ppm |
|---|---|---|
| 0 | 1.0000 | 0.0001 |
| 100 | 0.2283 | 0.0056 |
| 120 | 0.2651 | 0.0093 |
| 140 | 0.1761 | 0.0129 |
| 160 | 0.1518 | 0.0155 |
| 180 | 0.1816 | 0.0199 |
| 200 | 0.1482 | 0.0227 |
| 220 | 0.1433 | 0.0253 |
| 240 | 0.1640 | 0.0251 |
| 260 | 0.1498 | 0.0309 |
| 280 | 0.1376 | 0.0321 |
| 300 | 0.1239 | 0.0319 |
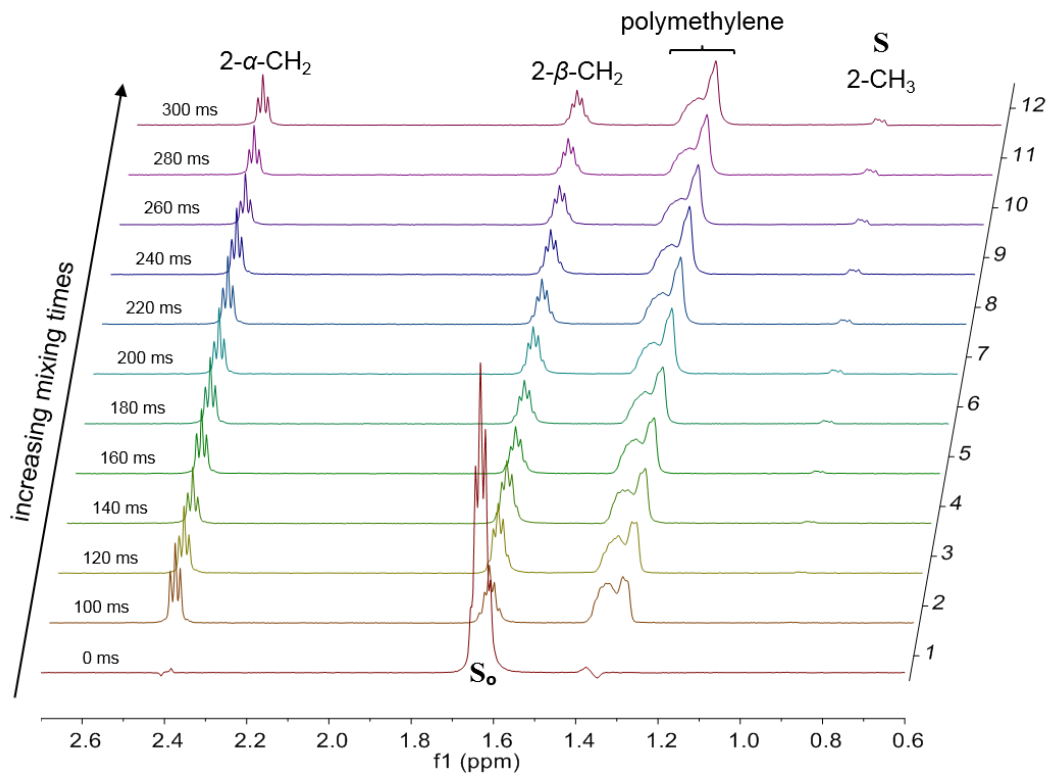
**Figure 3.94.** I4:26:0(2,2,10,12) 1D-TOCSY transfer spectra 2-$\beta$-CH$_2$ excitation at 1.62 ppm (generated using Varian Inova 600 MHz spectrometer).

**Table 3.27.** I4:26:0(2,2,10,12) 1D-TOCSY transfer 2-$\beta$-CH$_2$ (S$_o$) and 2-CH$_3$ (S) normalized integrals.

| mixing time (ms) | 2-$\beta$-CH$_2$ excitation at 1.62 ppm | 2-CH$_3$ S/S$_o$ at 0.88 ppm |
|---|---|---|
| 0 | 1.0000 | 0.0008 |
| 100 | 0.1673 | 0.0025 |
| 120 | 0.2161 | 0.0035 |
| 140 | 0.2020 | 0.0071 |
| 160 | 0.1435 | 0.0081 |
| 180 | 0.1366 | 0.0092 |
| 200 | 0.1532 | 0.0132 |
| 220 | 0.1395 | 0.0144 |
| 240 | 0.1391 | 0.0155 |
| 260 | 0.1275 | 0.0167 |
| 280 | 0.1147 | 0.0162 |
| 300 | 0.1081 | 0.0194 |
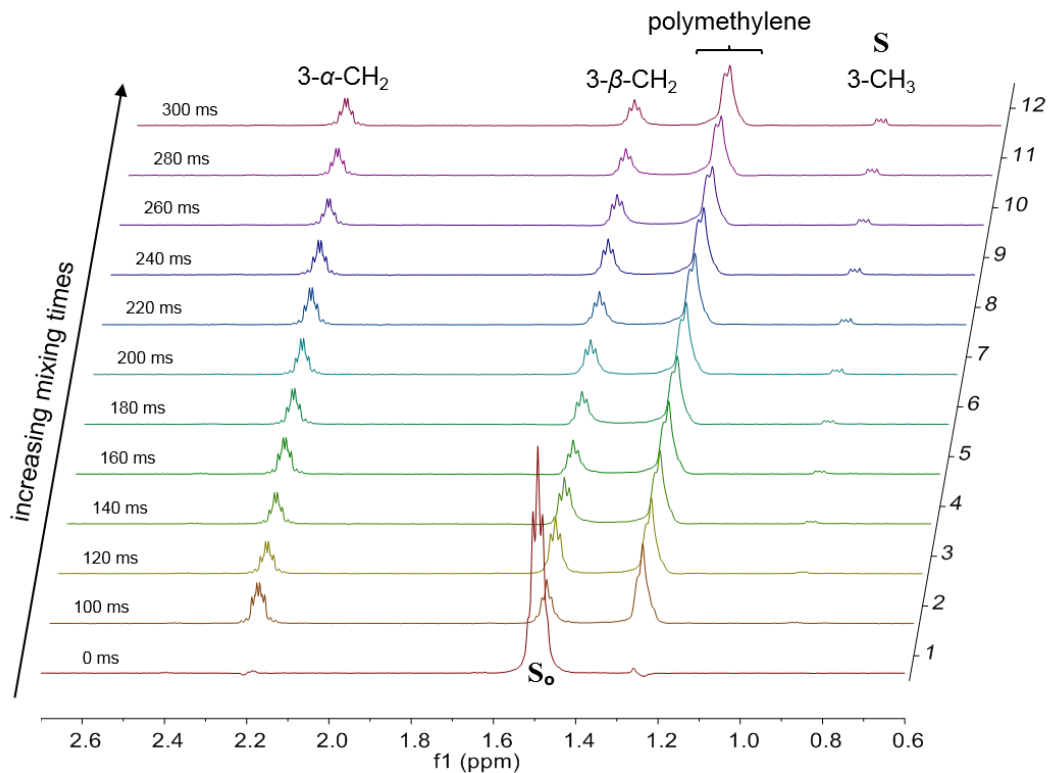
**Figure 3.95.** I4:26:0(2,2,10,12) 1D-TOCSY transfer spectra 3-$\beta$-CH$_2$ excitation at 1.48 ppm (generated using Varian Inova 600 MHz spectrometer).

**Table 3.28.** I4:26:0(2,2,10,12) 1D-TOCSY transfer spectra 3-$\beta$-CH$_2$ excitation at 1.48 ppm (generated using Varian Inova 600 MHz spectrometer).

| mixing time (ms) | 3-$\beta$-CH$_2$ excitation at 1.48 ppm | 2-CH$_3$ S/S$_0$ at 0.88 ppm |
|---|---|---|
| 0 | 1.0000 | -0.0009 |
| 100 | 0.1753 | 0.0043 |
| 120 | 0.2411 | 0.0072 |
| 140 | 0.2113 | 0.0109 |
| 160 | 0.1482 | 0.0127 |
| 180 | 0.1505 | 0.0151 |
| 200 | 0.1592 | 0.0199 |
| 220 | 0.1437 | 0.0206 |
| 240 | 0.1580 | 0.0224 |
| 260 | 0.1425 | 0.0222 |
| 280 | 0.1227 | 0.0236 |
| 300 | 0.1148 | 0.0248 |

# REFERENCES

# REFERENCES

1.  Moghe, G.D., et al., *Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway.* Elife, 2017. **6**: p. e28486.

2.  Ghosh, B., T.C. Westbrook, and A.D. Jones, *Comparative structural profiling of trichome specialized metabolites in tomato (Solanum lycopersicum) and S. habrochaites: acylsugar profiles revealed by UHPLC/MS and NMR.* Metabolomics, 2014. **10**(3): p. 496-507.

3.  King, R.R., et al., *Characterization of 2,3,4,3'-tetra-O-acylated sucrose esters associated with the glandular trichomes of Lycopersicon-typicum.* Journal of Agricultural and Food Chemistry, 1993. **41**(3): p. 469-473.

4.  Liu, X.X., et al., *Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of Petunia species.* Metabolomics, 2017. **13**(7): p. 85.

5.  Schilmiller, A.L., A.L. Charbonneau, and R.L. Last, *Identification of a BAHD acetyltransferase that produces protective acyl sugars in tomato trichomes.* Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16377-82.

6.  Fan, P., et al., *In vitro reconstruction and analysis of evolutionary variation of the tomato acylsucrose metabolic network.* Proc Natl Acad Sci U S A, 2016. **113**(2): p. E239-48.

7.  Schilmiller, A.L., et al., *Functionally divergent alleles and duplicated Loci encoding an acyltransferase contribute to acylsugar metabolite diversity in Solanum trichomes.* Plant Cell, 2015. **27**(4): p. 1002-17.

8.  Nadakuduti, S.S., et al., *Characterization of trichome-expressed BAHD acyltransferases in petunia axillaris reveals distinct acylsugar assembly mechanisms within the Solanaceae.* Plant Physiol, 2017. **175**(1): p. 36-50.

9.  Schilmiller, A., et al., *Mass spectrometry screening reveals widespread diversity in trichome specialized metabolites of tomato chromosomal substitution lines.* Plant J, 2010. **62**(3): p. 391-403.

10. Kroumova, A.B., D. Zaitlin, and G.J. Wagner, *Natural variability in acyl moieties of sugar esters produced by certain tobacco and other Solanaceae species.* Phytochemistry, 2016. **130**: p. 218-27.

11. Forero, D.P., et al., *Chemical and sensory comparison of fresh and dried lulo (Solanum quitoense Lam.) fruit aroma.* Food Chem, 2015. **169**: p. 85-91.

12. Gancel, A.L., et al., *Identifying carotenoids and phenolic compounds in naranjilla (Solanum quitoense Lam. var. Puyo hybrid), an Andean fruit.* J Agric Food Chem, 2008. **56**(24): p. 11890-9.

13. Forero, D.P., et al., *Spermidine derivatives in lulo (Solanum quitoense Lam.) fruit: sensory (taste) versus biofunctional (ACE-Inhibition) properties.* J Agric Food Chem, 2016. **64**(26): p. 5375-83.

14. Osorio, C., C. Duque, and F. Batista-Viera, *Studies on aroma generation in lulo (Solanum quitoense): enzymatic hydrolysis of glycosides from leaves.* Food Chemistry, 2003. **81**(3): p. 333-340.

15.     Treutler, H., et al., *Discovering regulated metabolite families in untargeted metabolomics studies.* Anal Chem, 2016. **88**(16): p. 8082-90.

16.     Plumb, R.S., et al., *UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation.* Rapid Commun Mass Spectrom, 2006. **20**(13): p. 1989-94.

17.     Sumner, L.W., et al., *Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics, 2007. **3**(3): p. 211-221.

18.     Ghosh, B. and A.D. Jones, *Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar esters).* Analytical Methods, 2017. **9**(6): p. 892-905.

19.     Ekanayaka, E.A., M.D. Celiz, and A.D. Jones, *Relative mass defect filtering of mass spectra: a path to discovery of plant specialized metabolites.* Plant Physiol, 2015. **167**(4): p. 1221-32.

20.     Burke, B.A., G. Goldsby, and J.B. Mudd, *Polar epicuticular lipids of Lycopersicon pennellii.* Phytochemistry, 1987. **26**(9): p. 2567-2571.

21.     Matsuzaki, T., et al., *Leaf surface glycolipids from Nicotiana-acuminata and Nicotiana-pauciflora.* Agricultural and Biological Chemistry, 1991. **55**(5): p. 1417-1419.

22.     Loewus, F.A. and P.P.N. Murthy, *myo-inositol metabolism in plants.* Plant Science, 2000. **150**(1): p. 1-19.

23.     Duus, J., C.H. Gotfredsen, and K. Bock, *Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations.* Chem Rev, 2000. **100**(12): p. 4589-614.

24.     Sachleben, J.R., et al., *Aliphatic chain length by isotropic mixing (ALCHIM): determining composition of complex lipid samples by (1)H NMR spectroscopy.* J Biomol NMR, 2014. **59**(3): p. 161-73.

25.     Herrera-Salgado, Y., et al., *Myo-inositol-derived glycolipids with anti-inflammatory activity from Solanum lanceolatum.* J Nat Prod, 2005. **68**(7): p. 1031-6.

26.     Luckwill, L.C., *The genus Lycopersicon. An historical, biological and taxonomic survey of the wild and cultivated tomatoes.* PhD thesis, Department of Botany, University of Aberdeen, UK, 1943.

27.     Glas, J.J., et al., *Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores.* Int J Mol Sci, 2012. **13**(12): p. 17077-103.

**Chapter 4: *S*-alkyl glutathione retention indexing and column performance evaluation standards for improved annotation, identification and dereplication of metabolite discovery**

**4.1 Introduction**

Annotation of novel metabolites remains as the greatest obstacle to understanding the mechanisms responsible for metabolite accumulation and the functional significance of new metabolites. The combination of mass spectrometry (MS) with high-performance liquid chromatography (HPLC) separations provides the foundation for chemical analysis of non-volatile small molecules including drugs, pesticides, endogenous metabolites and other complex mixtures, yet only a small fraction of substances detected by LC/MS have been identified. LC/MS approaches exploit relationships between chromatographic retention times (RT), mass-to-charge (*m/z*) measurements and fragment ion spectra for identification and quantification, or for determination of unknown elemental compositions and partial structural information [1].

Though today's LC/MS technologies are advanced, offering ultra-high performance LC (UHPLC) separations with phenomenal chromatographic resolution and peak capacities [2], the number of metabolites detected in some untargeted metabolite analyses far exceeds chromatographic peak capacities. Coupled with high resolution mass analyzers, modern UHPLC/MS instruments provide high acquisition speeds, mass accuracy and sensitive detection [3]. Despite great advancements in LC/MS technologies, annotation and identification of unknowns continues to present a bottleneck. Perhaps in no other fields is this more evident than in natural product discovery and metabolomics research, where the immense array of chemical diversity in living organisms presents formidable challenges [4].

Our ability to measure metabolites within complex extracts far exceeds our ability to generate unambiguous structure information about most of them. For mass spectral interpretation, GC/MS electron ionization (EI) and LC/MS tandem (MS/MS) collision induced dissociation (CID) mass spectral database searches and molecular networking approaches aim to match mass spectra to a library or to provide structural insights

from compounds that nearly match [1, 5]. A preferred method for identifying metabolites involves matching

mass spectra and retention time data using authentic reference materials. However, it not practical or

possible to have reference standards available for all metabolites, and of course, these do not exist for novel

metabolites not described before. All too often, mass spectra alone are not enough to fully characterize

and/or identify metabolites, and final structure confirmation is performed using independent approaches

including one- and two-dimensional (1D and 2D) NMR spectroscopy or x-ray crystallography, and these

techniques work best after compounds have been purified from mixtures. While NMR spectroscopy

provides more detailed information about connections between individual atoms than MS, it is frequently

avoided because it requires time-consuming purifications, adequately pure samples and far greater sample

quantities than MS for analysis (~80% purity, ~1 mg). In addition, as more metabolite structures are

determined, there is a growing chance that researchers will purify and identify compounds that are already

known but not in mass spectrum libraries. In view of this, dereplication of metabolite discovery by LC/MS

approaches is essential for improving the efficiency of metabolite identification.

Mass spectral databases are widely used for LC/MS [6, 7], yet the addition of RT data to MS/MS libraries

has found limited success because RTs often vary across laboratories, instruments, and experimental

protocols. In GC/MS analyses, this problem has been addressed through Kováts retention indexing (RI),

which converts gas chromatographic RTs into system-independent constants by normalizing RTs to those

of a homologous series of *n*-alkanes, generating RIs that are comparable across platforms [8]. RI values are

calculated by linear interpolation between retention standards according to Equation 4.1, where, $t_r$ is the

retention time, and **n** and **N** are the number of carbon atoms in the shorter and longer *n*-alkyl standards

bracketing the peak of interest respectively.

**Equation 4.1.** Calculation of RI values.

$$RI = 100 \times \left[ n + (N - n) \frac{t_{r(unknown)} - t_{r(n)}}{t_{r(N)} - t_{r(n)}} \right]$$

By comparison to GC/MS databases, LC/MS approaches have yet to find widespread use. In part, this is because MS/MS spectral databases are less complete, and LC protocols vary, leading to retention being less reproducible across laboratories, and there are no widely-accepted RP-HPLC RI standards [1]. A few reports of RI standards have consisted of homologous series of 2-ketoalkanes, alkyl aryl ketones, and 1-nitroalkanes with ultraviolet spectroscopic detection [9-11], but most lack functional groups that facilitate their detection using mass spectrometry. Since electrospray ionization (ESI) is often the first approach for metabolite discovery by LC/MS, indexing standards that ionize well using both positive- and negative-ion modes would be preferable. This way, the same standard series could be used in both ionization modes. Two recent reports use combinations of *N*-alkyl acetamides and *N*-alkylbenzamides homologous standards to characterize HPLC gradients and project retention times across laboratories and instruments. However, it is not clear whether those standards are suitable for detection by both ionization modes [12, 13]. A recent patent by *Quilliam* (U.S. Patent No. US9594063B2) documents 1-alkylpyridinesulfonic acids that are ionizable using both modes, however, those retention standards have not been described in peer-reviewed publication [14, 15]. To address these issues, a homologous set of *S*-alkyl derivatives of the tripeptide glutathione (GS-*n*-alkyl RI standards) featuring normal saturated chain lengths (1-24 carbons) was synthesized.

Experimental factors that influence RI values for 16 previously identified acylsucrose metabolites from *Salpiglossis sinuata* (outlined in Chapter 2) were examined by varying chromatographic conditions including brands of RP C18 columns, solvent delivery systems, which may differ in delays in delivering mixed solvents to columns, mobile phase pH, column temperature, steepness of LC gradients, and variation in organic solvent modifiers. Acylsugars are diverse sugar polyester metabolites that accumulate in many plants, especially within the family Solanaceae [16]. Hundreds of unique acylsugars have been detected and resolved by UHPLC/MS [17-19]. Moreover, they exist in numerous isomeric and isobaric forms in a single plant extract, many of which are not distinguished using MS/MS, making them suitable to test the performance of LC chromatographic retention indexing to distinguish isomers.

402

### 4.1.1 LC column performance

Silica and silica hybrid supports with bonded groups based on octadecyl (C18) groups are the most widely used materials for RP-HPLC applications [20]. Compounds that exhibit ideal chromatographic behavior are expected to elute in a Gaussian profile that results from statistical behavior of individual molecules traveling through the column. However, it is common that chromatographic elution appears in the form of asymmetric, or tailing, peaks. Columns based on silica supports often exhibit severe peak tailing of basic compounds, limited pH range compatible with column stability, and irreproducibility of retention for the same column chemistries [21]. There are several factors that influence column reproducibility, including particle shape and size, pore diameter, surface area, stationary phase loading, bonded phase density, metal impurities, endcapping, and more [20]. Approaches to column performance troubleshooting have identified four main causes of poor chromatographic performance: 1) blockage of the frit of the column, 2) void formation, 3) adsorbed sample impurities that alter analyte partitioning between mobile and stationary phases, or 4) chemical attack on the silica or bonded phase [22]. Symptoms of column-related problems manifest as increased column backpressure, peak fronting/tailing, reduced column plate number, and changes in retention including selectivity.

For reversed-phase alkyl silica columns, hydrophobic factors dominate analyte partitioning between mobile and stationary phases and retention, particularly for neutral analytes. However, contributions from other solute-column interactions described by the Hydrophobic Subtraction Model are widely considered responsible for column selectivity and peak tailing [23]. Principally, underivatized silanol groups, which are always present on silica column surfaces, play a key role, and can interact with retained solutes [21]. For example, protonated bases can interact with ionized silanol groups ($-Si-O^-$), and neutral silanol groups ($-Si-OH$) can hydrogen bond with proton acceptor solutes. End-capping of silanol groups, often with trimethylsilyl (TMS) groups, reduces the number of underivatized silanol groups and decreases the silanol accessibility because of steric hindrance between underivatized silanol and TMS and/or C18 groups, but endcapping fails to achieve complete inactivation of silanols [21]. Hydrolysis of the bonded phase may

occur as the column ages, or more aggressively due to operating at very low pH (usually pH < 2). Under moderately basic conditions (usually pH > 8), the silica support reacts with hydroxide ion at more substantial rates and may dissolve, reducing the structural integrity of the silica support. Dissolution of the silica support may release the bonded stationary phase and create voids in the column due to particle collapse. In any case, these processes give rise to more underivatized silanol groups, changes in peak shape, retention and selectivity [24].

## 4.2 Materials and Methods

### 4.2.1 Synthesis of S-alkyl glutathione standards

Preparation of GS-*n*-alkyl standards was performed by modifications to methods previously reported [25]. An abbreviated nomenclature of *S*-alkylglutathiones is suggested as GS-n, where n equals the number of carbon atoms in a linear alkyl group. Authentic standards of GS-1 (used for preparation of mixed stock solution), GS-6 and GS-8 were purchased from Sigma-Aldrich and had matching HPLC retention times and fragment ion spectra generated by nonselective collision-induced dissociation compared to synthesis products. The purity of reaction products was qualitatively assessed using LC/MS.

### 4.2.1.1 Synthesis of GS-2 and GS-3

To a 250-mL round bottom flask, ~10 mmoles reduced glutathione (GSH, Sigma-Aldrich) was added. A 10-mL volume of MilliQ water was added, followed by addition of 10 mL of 2.0 M aqueous NaOH solution. Dissolution of the glutathione was promoted by stirring using a magnetic stir bar at room temperature. Once the glutathione had dissolved, 75 mL of 95% ethanol was added, and the solution turned cloudy. An equimolar quantity of ethyl iodide or propyl iodide (Sigma-Aldrich) was added. The vessel was sealed with a rubber septum, then purged by bubbling $N_2$ gas through a syringe and venting while mixing for ~10 mins. The reaction was left to proceed overnight (12-16 h) at room temperature (~ 25˚C). The pH was adjusted to ~4.5 by dropwise addition of 12 M HCl and the solution concentrated under rotary evaporation (~40°C) to a volume of ~10 mL where precipitates were observed. The mixture was further recrystallized by cooling in an ice bath, followed by vacuum filtration to recover precipitates. The precipitates were washed with

methanol and dried. The standards were further purified by repeating the recrystallization in water, recovered by vacuum filtration and washed with methanol and dried overnight in a 40°C drying cabinet. Purity was checked by LC/MS. The yield was approximately 10% and 25% for GS-2 and GS-3 respectively (most material was lost during recrystallization and washing steps).

### 4.2.1.2 Synthesis of GS-4 to GS-9

Syntheses of homologs GS-4 to GS-9 were performed according to Vince et. al. (1971) with minor deviations. To a 40-mL glass vial, ~2.0 mmoles of GSH was added. A 2.0-mL volume of MilliQ water was added, followed by addition of 2.0 mL of 2.0 M NaOH solution. Dissolution of the glutathione was promoted by stirring using a magnetic stir bar at room temperature. After dissolution, ~15 mL of 95% ethanol was added until the solution turned cloudy. An equimolar quantity of alkyl halide RX (X = iodine for butyl through octyl; X = bromine for nonyl; Sigma-Aldrich) was added, vials were sealed with Teflon lined caps and the reaction was left to proceed 12-16 h at about 25°C. Precipitated solids were often visible within a few hours. The pH was adjusted to ~3.5 by dropwise addition of 12 M HCl and the vials were placed in a -20°C freezer for 12-16 h. Insoluble products were recovered by vacuum filtration, with alternating washes using water and hexanes, and then dried overnight in a 40°C drying cabinet. The yields ranged from 55-80%.

### 4.2.1.3 Synthesis of GS-10 to GS-20 and GS-22

To a 100-mL round bottom flask, ~0.4 mmoles of alkyl bromide (from Sigma-Aldrich, except docosyl bromide from TCI America) was added. A 45-mL volume of 95% ethanol and stir bar was added. The mixture was sealed with a rubber septum, then purged by bubbling $N_2$ gas through a syringe with venting (~10 min) while heating in a sand bath at ~55°C. In a separate vial, a 1.1 molar equivalent of GSH was added to a centrifuge tube, along with 2.0 mL of MilliQ water. A volume of 2.0 M NaOH corresponding to 2:1 molar ratio of NaOH to GSH was added to the vial and vortexed to dissolve GSH. Its contents were added dropwise to the round bottom flask via a syringe while mixing and purging with $N_2$ gas. The $N_2$ gas line was removed from the septum, and the sealed flask was placed in an orbital mixer at 55°C and 140

rpm. The reaction was left to proceed for 12-16 h. To stop the reaction, the flask was cooled to room temperature and pH was adjusted to ~3.5 with dropwise addition of 12 M HCl. The flask was placed in a -20°C freezer overnight. The insoluble products were recovered by vacuum filtration, with alternating washes using water and hexanes, then dried overnight in a 40°C drying cabinet. The yields were ~65%.

### 4.2.1.4 Synthesis of GS-21, GS-23 and GS-24

Synthesis of GS-21, GS-23 and GS-24 was performed using the same method described in Section 4.2.1.3. However, longer chain 1-bromoalkanes were not readily available for purchase and were synthesized from alcohol analogues: 1-heneicosanol (Ultra Scientific), 1-tricosanol (Ultra Scientific) and 1-tetracosanol (TCI America). Their synthesis is described below.

All glassware was dried in an oven >100 °C overnight before use. To a 100-mL three neck round bottom flask, ~200 mg of long chain alcohol was added. The reactor was sealed with rubber septa, placed in a sand bath at ~60 °C and purged with $N_2$ gas. Twenty-five mL of toluene: tetrahydrofuran (1:1, v:v) dried with 3 Å molecular sieves was added to the reactor. A 10% molar excess of phosphorous tribromide ($PBr_3$) was added and the reaction left overnight. The reaction was quenched with 25 mL of 0.01 M aqueous KBr solution. The mixture was added to a 125-mL separatory funnel and the water layer removed. The organic layer was washed three times with ~20 mL MilliQ water. The organic layer was added to a 100-mL three neck round bottom reactor and dried under $N_2$ gas. Colorless solids were visible. The crude product was placed in an oven (> 100 °C) and dried overnight. The crude product underwent a second treatment with $PBr_3$ described above to raise the yield. After washing, the organic layer was dried under $N_2$ gas. The products were dissolved in 2.0 mL hexanes and purified using 500 mg - 3 mL Avantor Bakerbond silica gel solid phase extraction columns. The columns were conditioned with 12 mL of hexanes, followed by the addition of 2.0 mL of crude reaction products, then eluted with 6.0 mL of hexanes. The eluate was added to a 100-mL round bottom flask and dried under $N_2$ gas leaving white solids. Identities of 1-bromoalkane synthesis products were confirmed by analysis using GC-MS. Typical yields were ~50%.

**4.2.2 Preparation of retention index standards and their application in LC/MS analyses**

**4.2.2.1 Preparation of GS-*n*-alkyl mixed stock solution**

Individual GS-*n*-alkyl stocks were prepared by weighing ~60 µmoles of each standard into 40-mL glass vials (see Table 4.1 for stock preparation details). For vials containing GSH to GS-9, a 22.5-mL volume of 3:3:2 acetonitrile: isopropanol: water (AcN:IPA:H$_2$O, v:v:v) was added to each. To vials containing GS-10 to GS-24, an 18.0-mL volume of 3:3:2 AcN:IPA:H$_2$O was added to each. To all vials, 75 µL of ammonium hydroxide (NH$_4$OH, ACS, EMD Millipore, assay percentage range 28-30%) was added. Finally, 4.5 mL of chloroform was added to vials containing GS-10 to GS-24. In most cases the standards had not dissolved completely (especially for longer chain lengths). The vials were placed on an orbital shaker at 55 °C and 150 rpm. The stock solutions were occasionally removed, vigorously shaken by hand and visually examined for undissolved material. When all standard material had dissolved completely, an aliquot of each stock was combined to a mixed stock solution (according to the values listed in Table 4.1) such that each standard was present at 107.5 µM. The final volume of the mixed stock was adjusted to 100 µM each by addition of 1.834 mL of 3:3:2 AcN:IPA:H$_2$O solution. The stock solutions and mixed stock were stable when stored in a -20 °C freezer for more than 6 months and showed no apparent degradation of the GS-*n*-alkyl standards based on LC/MS analyses. However, reduced GSH in the mixed stock was mostly converted to the oxidized form (GSSG) when tested after a few months.

**Table 4.1.** Preparation of GS-*n*-alkyl Mixed Stock Solution

| GS-# | Mass Weighed (g) | Vol. 3:3:2 (mL) | Vol. CHCl$_3$ (mL) | Vol NH$_4$OH (mL) | Total Vol. Stock 1 (mL) | Stock 1 Conc. (µM) | Vol. from Stock 1 (mL) | Mix Stock Conc. (µM) |
|---|---|---|---|---|---|---|---|---|
| H | 0.0186 | 22.5 | --- | 0.075 | 22.575 | 2681 | 0.981 | 107.5 |
| 1 | 0.0219 | 22.5 | --- | 0.075 | 22.575 | 3019 | 0.871 | 107.5 |
| 2 | 0.0201 | 22.5 | --- | 0.075 | 22.575 | 2655 | 0.990 | 107.5 |
| 3 | 0.0213 | 22.5 | --- | 0.075 | 22.575 | 2700 | 0.974 | 107.5 |
| 4 | 0.0223 | 22.5 | --- | 0.075 | 22.575 | 2718 | 0.967 | 107.5 |
| 5 | 0.0228 | 22.5 | --- | 0.075 | 22.575 | 2676 | 0.983 | 107.5 |
| 6 | 0.0232 | 22.5 | --- | 0.075 | 22.575 | 2625 | 1.002 | 107.5 |
| 7 | 0.0237 | 22.5 | --- | 0.075 | 22.575 | 2589 | 1.016 | 107.5 |
| 8 | 0.0260 | 22.5 | --- | 0.075 | 22.575 | 2745 | 0.958 | 107.5 |
| 9 | 0.0272 | 22.5 | --- | 0.075 | 22.575 | 2779 | 0.946 | 107.5 |
| 10 | 0.0261 | 18.0 | 4.5 | 0.075 | 22.575 | 2583 | 1.018 | 107.5 |
| 11 | 0.0280 | 18.0 | 4.5 | 0.075 | 22.575 | 2687 | 0.979 | 107.5 |
| 12 | 0.0288 | 18.0 | 4.5 | 0.075 | 22.575 | 2682 | 0.980 | 107.5 |
| 13 | 0.0291 | 18.0 | 4.5 | 0.075 | 22.575 | 2632 | 0.999 | 107.5 |
| 14 | 0.0299 | 18.0 | 4.5 | 0.075 | 22.575 | 2629 | 1.000 | 107.5 |
| 15 | 0.0309 | 18.0 | 4.5 | 0.075 | 22.575 | 2644 | 0.995 | 107.5 |
| 16 | 0.0321 | 18.0 | 4.5 | 0.075 | 22.575 | 2674 | 0.983 | 107.5 |
| 17 | 0.0327 | 18.0 | 4.5 | 0.075 | 22.575 | 2654 | 0.991 | 107.5 |
| 18 | 0.0340 | 18.0 | 4.5 | 0.075 | 22.575 | 2690 | 0.977 | 107.5 |
| 19 | 0.0377 | 18.0 | 4.5 | 0.075 | 22.575 | 2910 | 0.904 | 107.5 |
| 20 | 0.0345 | 18.0 | 4.5 | 0.075 | 22.575 | 2600 | 1.011 | 107.5 |
| 21 | 0.0356 | 18.0 | 4.5 | 0.075 | 22.575 | 2620 | 1.004 | 107.5 |
| 22 | 0.0376 | 18.0 | 4.5 | 0.075 | 22.575 | 2704 | 0.972 | 107.5 |
| 23 | 0.0375 | 18.0 | 4.5 | 0.075 | 22.575 | 2637 | 0.997 | 107.5 |
| 24 | 0.0398 | 18.0 | 4.5 | 0.075 | 22.575 | 2738 | 0.960 | 107.5 |

Stock 2 Vol (mL) = 24.458

Volume 3:3:2 solution added to adjust **final Conc. to 100 µM** (mL)= 1.834

### 4.2.2.2 Preparation of *S. sinuata* retention indexing stock solution

To evaluate the performance of GS-*n*-alkyls for retention indexing, the mixed stock solution was added to achieve a concentration of 1.0 µM to a diluted extract from *Salpiglossis sinuata* containing acylsucrose metabolites. The sample was prepared by diluting 1.00 mL of *S. sinuata* bulk extract (described in Chapter 2, Section 2.2.1) into 8.80 mL of 3:3:2 AcN:IPA:H$_2$O. Followed by a 100 µL addition of GS-*n*-alkyl mixed stock solution (warmed to 50 °C and shaken vigorously prior to addition). Finally, an equimolar concentration of formic acid relative to NH$_4$OH in the solution was added (100 µL of 0.046 M).

**4.2.3 UHPLC/MS methods and experimental conditions**

**4.2.3.1 LC/MS instrument configurations**

Retention indexing was performed using two Waters Xevo G2-XS quadrupole time-of-flight mass spectrometers equipped with either Shimadzu or Waters LC systems. The Shimadzu LC system was equipped with LC-20AD pumps, CTO-20A column oven and SIL-5000 autosampler. The Waters LC system consisted of Acquity UPLC I-Class Binary Solvent Manager and 2777C Sample Manager. The instruments were operated using electrospray ionization in positive- and negative-ion modes (ESI+ and ESI-) and "sensitivity mode" mass analyzer parameters. All retention indexing experiments were performed in triplicate, with two analyses by ESI- and one by ESI+. Nonselective CID spectra were acquired over *m/z* 50-1500 in centroid format using four quasi-simultaneous collision potential functions (0, 10, 20, 40 V, each with 0.1 s acquisition times) to generate fragment ions. Lockspray mass correction was applied using Leu-enkephalin as a lock mass reference standard (0.1 s acquisition time, 10 s scan frequency).

**4.2.3.2 Retention indexing peak detection parameters**

Chromatographic retention times were determined using TargetLynx software (Waters). For GS-*n*-alkyls, $[M-H]^-$ and $[M+H]^+$ monoisotopic ions were detected. For acylsucrose metabolites, formate ($[M+HCOO]^-$) and ammonium ($[M+NH_4]^+$) monoisotopic adduct ions were detected, except when acylsucrose isomers were unresolved. For those instances, distinguishing fragment ions from high energy CID spectra were used (see Table 4.2). Peak integrations were adjusted manually as needed, the integration parameters are as follows: 1) smoothing method – mean; 2) smoothing iterations – 5; 3) smoothing width – 3; 4) chromatographic mass window – 0.02 Da. Retention time data were exported and processed using Microsoft Excel.

**Table 4.2.** LC/MS RT and RI results for 16 *S. sinuata* acylsucrose metabolites with structural identifications (column SAE-A, temperature 50°C, aqueous 10 mM ammonium formate pH 2.8 and linear gradient 1-100% acetonitrile, slope 1% B min$^{-1}$). Analyses were performed in triplicate (two by ESI- and one by ESI+) on four separate dates (n=12 total analyses). Acylsucrose ID's in bold are three S4:20:0 isomers (peaks #2-4) described below in Figure 4.3. Pseudomolecular [M+HCOO]$^-$ and [M+NH$_4$]$^+$ ions were detected by ESI- and ESI+ ion modes, except for italicized characteristic fragment ions generated at elevated collision potentials (20 V CID by ESI- and 10 V CID by ESI+).

| Acylsucrose # | Acylsucrose ID | Ion detected by ESI- (*m/z*) | Ion detected by ESI+ (*m/z*) | Avg Ret. Time (min) | Ret. Time SD (min) | Ret. Time RSD (%) | RI Value | RI Value SD | RI Value RSD (%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | S4:19:0(3,5,5,6) | *579.266* | *401.217* | 50.26 | 0.384 | 0.76 | **1446** | 0.6 | 0.04 |
| 2 | S4:19:0(2,5,6,6) | *621.313* | *443.264* | 50.41 | 0.386 | 0.77 | **1451** | 0.6 | 0.04 |
| 3 | S5:20:0(2,2,5,5,6) | 737.324 | 710.359 | 51.54 | 0.392 | 0.76 | **1487** | 0.4 | 0.03 |
| 4 | **S4:20:0(4,5,5,6)** | 723.345 | 696.380 | 52.85 | 0.398 | 0.75 | **1529** | 0.4 | 0.03 |
| 5 | **S4:20:0(3,5,6,6)** | *579.266* | *415.233* | 53.44 | 0.401 | 0.75 | **1548** | 0.4 | 0.03 |
| 6 | **S4:20:0(2,6,6,6)** | *635.328* | *457.280* | 53.64 | 0.402 | 0.75 | **1555** | 0.3 | 0.02 |
| 7 | S5:22:1(2,5,5,5,5$^T$) | 763.339 | 736.375 | 54.54 | 0.404 | 0.74 | **1583** | 0.3 | 0.02 |
| 8 | S5:21:0(2,2,5,6,6) | 751.343 | 724.375 | 54.75 | 0.406 | 0.74 | **1590** | 0.4 | 0.03 |
| 9 | S4:21:0(5,5,5,6) | 737.360 | 710.396 | 55.47 | 0.411 | 0.74 | **1613** | 0.4 | 0.02 |
| 10 | S5:25:4(2,5,5,5,8$^P$) | 799.339 | 772.375 | 56.18 | 0.416 | 0.74 | **1635** | 0.3 | 0.02 |
| 11 | S5:22:0(2,2,6,6,6) | 765.355 | 738.391 | 57.94 | 0.424 | 0.73 | **1691** | 0.2 | 0.01 |
| 12 | S4:22:0(5,5,6,6) | 751.376 | 724.411 | 58.46 | 0.424 | 0.73 | **1708** | 0.3 | 0.02 |
| 13 | S5:23:0(2,5,5,5,6) | 779.372 | 752.406 | 60.65 | 0.412 | 0.68 | **1776** | 0.4 | 0.02 |
| 14 | S4:23:0(5,6,6,6) | 765.391 | 738.427 | 61.54 | 0.412 | 0.67 | **1804** | 0.4 | 0.02 |
| 15 | S5:24:0(2,5,5,6,6) | 793.389 | 766.422 | 63.53 | 0.408 | 0.64 | **1865** | 0.5 | 0.03 |
| 16 | S6:25:0(2,2,5,5,5,6) | 821.381 | 794.417 | 64.70 | 0.411 | 0.64 | **1901** | 0.3 | 0.02 |

## 4.2.3.3 Chromatographic columns

All separations were performed using C18 columns with dimensions 10 cm × 2.1 mm. Three column types from leading manufacturers were chosen for comparison: 1) three Supelco Ascentis Express C18 (SAE; 2.7 µm particle size, superficially porous particles with monofunctional C18 bonding chemistry and endcapping) from separate lots and differing stages of use were measured (column SAE-A had Bin Lot: S17138, column SAE-B had BL: S17156 and column SAE-C had BL: S15013). Columns SAE-A and -B were new columns, while column SAE-C had been heavily used before evaluation; 2) Waters Acquity BEH C18 (WBEH; Lot No. 0308372931, 1.7 µm particle size, fully porous particle, 18% carbon loading,

trifunctional bonding chemistry with endcapping); and 3) Agilent Zorbax Eclipse Plus (AZEP; L.N. B17300, 1.8 µm particle size, fully porous particle, 9% carbon loading, monofunctional bonding chemistry with endcapping).

### 4.2.3.4 Column performance and LC systems evaluations

Comparisons of columns and LC systems were performed using linear gradient elution, aqueous 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid (Solvent A) and acetonitrile (Solvent B). Solvents were not degassed. The elution profile used for GS-*n*-alkyl retention indexing of acylsugars consisted of a 110-min method: hold at 1% B from 0-1 min, linear gradient from 1-100% B over 1-100 min (slope 1% Solvent B min$^{-1}$), hold at 100% B at 100-105 min, linear 100-1% B over 105-106 min, and hold at 1% B from 106-110 min. Injections of 5 µL (using a 5 µL sample loop) were performed with mobile phase flow rate of 0.3 mL min$^{-1}$ and column oven temperature at 50 °C. The method described herein was used for all measurements unless otherwise noted.

### 4.2.3.5 Mobile phase pH dependence of acylsugar metabolite retention index values

To assess pH dependence of GS-*n*-alkyl standards and acylsucrose RI values, the pH of aqueous mobile phase Solvent A (starting with 10 mM aqueous ammonium hydroxide) was adjusted from pH 2.5-4.0, in steps of 0.3 pH units with formic acid (Solvent A). Column SAE-A was used for all analyses of pH dependence. All other LC conditions are as outlined in Section 4.2.3.4.

### 4.2.3.6 Column temperature dependence of acylsugar retention index values

To assess column temperature dependence of GS-*n*-alkyl standards and acylsucrose RI values, the column temperature was adjusted from 30 to 60 °C in steps of 5 °C. Column SAE-A was used for all experiments. All other LC conditions are as outlined in Section 4.2.3.4.

### 4.2.3.7 Dependence of acylsugar retention index values on mobile phase gradient slope

To assess gradient slope dependence of GS-*n*-alkyl standards and acylsucrose RI values, the slope of the linear gradient was adjusted to values of 1, 10/9, 5/4, 10/7, 5/3 and 2% Solvent B min$^{-1}$ (wash and

equilibration times are described in Section 4.2.3.4; total run times: 110, 100, 90, 80, 70 and 60 mins respectively). Column SAE-A was used for all experiments. All other LC conditions are as outlined in Section 4.2.3.4.

### 4.2.3.8 Acylsugar retention index values using methanol as organic mobile phase component

To examine the effect of organic component dependence of GS-*n*-alkyl standards and acylsucrose RI values, methanol (Solvent B) was substituted for acetonitrile. Column SAE-A was used for all experiments. All other LC conditions are as outlined in Section 4.2.3.4.

### 4.2.3.9 MS/MS spectra of GS-*n*-alkyl standards

MS/MS spectra of GS-*n*-alkyl products were generated using Waters SONAR data acquisition platform using a 10-40 V collision potential ramp, 10 Da quadrupole transmission bin widths and 0.5 s acquisition times. The LC method was as described in Section 4.2.3.4. However, methanol was used as organic component instead of acetonitrile. Measurements were performed in ESI- and ESI+ modes using *S. sinuata* bulk extract containing 10 µM GS-*n*-alkyl standards. Ion abundance measurements were determined using Waters MS$^E$ Data Viewer Software (Version 1.4). Peak detection parameters were as follows: 1) chromatographic peak width – auto; 2) MS resolution – 0.05 Da; 3) Lock mass enabled Leu-Enkephalin (*m/z* 554.262); 3) low energy filtering threshold – 100 counts; 4) high energy filtering threshold – 10 counts; 5) chromatographic peak width factor – 1.0; 6) quadrupole time peak width factor – 1.0; and XIC window for high energy data – 0.05 Da.

### 4.3 Results and discussion

The performance of using *S*-alkylglutathione homologs for retention indexing was examined by adding equimolar GS-*n*-alkyl standards from a mixed stock solution to an extract of *S. sinuata* containing neutral acylsucrose metabolites and performing LC/MS analyses using linear gradient elution with acetonitrile as organic mobile phase component (Figure 4.1). The experimental conditions used for data presented herein are given in Section 4.2.3.4, unless noted otherwise (Ascentis Express C18 Column A (SAE-A), aqueous 10 mM ammonium formate pH 2.8 and linear gradient 1-100% acetonitrile, slope 1% B min$^{-1}$). RI values

were evaluated under different chromatographic conditions using linear gradient elution and by altering columns, solvent delivery systems, mobile phase pH, column temperature, linear gradient slope and organic mobile phase component.

**4.3.1 LC/MS of GS-*n*-alkyl standards**

GS-*n*-alkyl RI standards encompass a wide RP-UHPLC retention range and are ionized by ESI in positive- and negative-ion modes (yielding $[M+H]^+$ and $[M-H]^-$ ions). Extracted ion chromatograms (XICs) for characteristic fragment ions such as *m/z* 162.022 in positive-ion mode ($C_5H_8NO_3S^+$, Figure 4.1B) and *m/z* 254.078 ($C_{10}H_{12}N_3O_5^-$) in negative-ion mode allow for their selective detection using a single narrow *m/z* window for each polarity, simplifying reporting of standard RTs when they have been added to a sample. GSH and GS-1 standards were not retained and eluted as unretained solutes. GS-2 was poorly retained and displayed peaks that eluted at the solvent front and a smaller peak that eluted at ~3 mins (when detected, the later peak was reported). This behavior is attributed to having the standard dissolved in a strong solvent that contributed to double chromatographic peaks. Standards GS-3 to GS-5 showed peak fronting/splitting that was attributed to the sample's solvent strength and/or partial ionization of GS-*n*-alkyl functional groups at injection (the later/larger peak was reported). Standards < GS-6 carbon chain length showed lower signal intensity than the longer chain homologs, a portion of which was attributed to sampling of broadened/split peaks and lower ionization efficiency when the aqueous mobile phase was the dominant mobile phase component at those elution times [26]. GS-6 to GS-24 showed single peaks that broadened with more tailing as the series chain length increased. This may be attributed to reduced solvation of charged functional groups at high acetonitrile solvent compositions. Figure 4.2 displays a polynomial fit for standards GSH (GS-0) to GS-7, relating retention times to corresponding RI values, while standards GS-7 to GS-24 exhibit a linear relationship (~3.3 mins per GS-*n*-alkyl standard) between retention time and RI value. Thus, the GS-*n*-alkyl elution order is predictable.
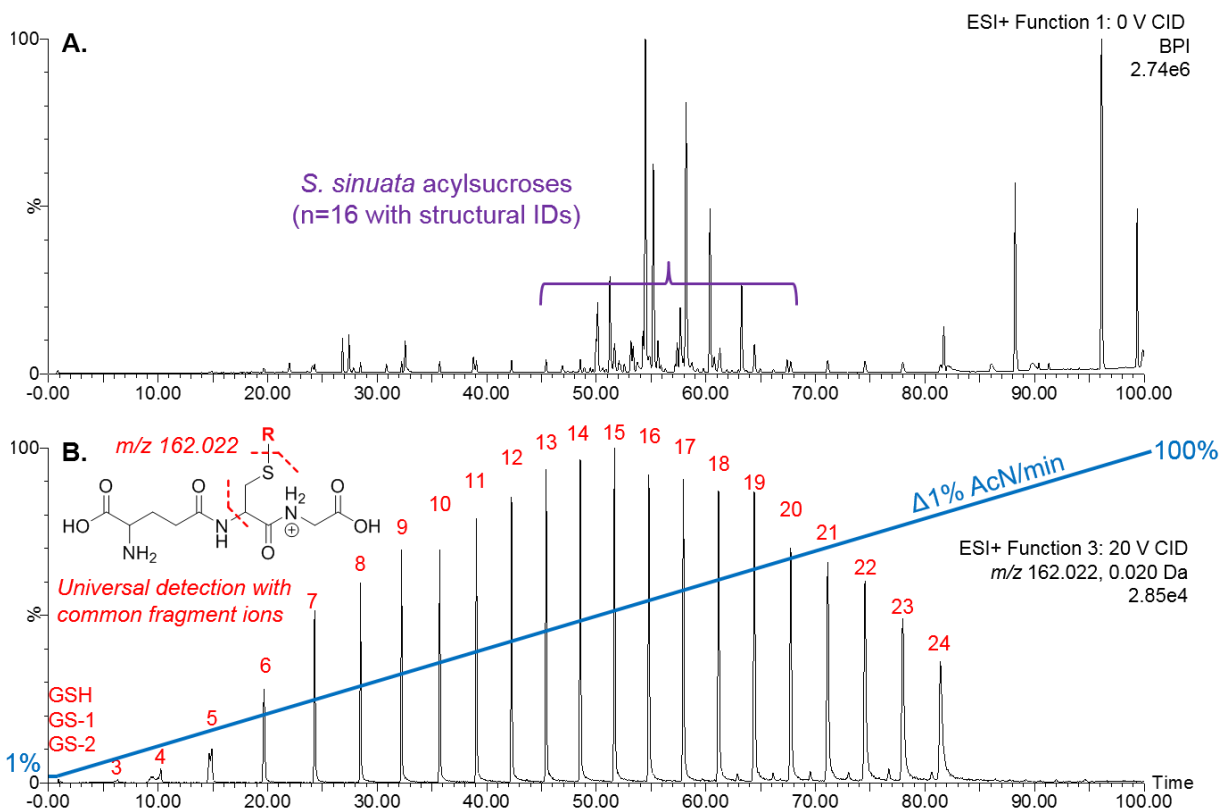
**Figure 4.1.** LC/MS chromatograms of *S. sinuata* extract containing acylsucroses plus 1.0 µM each GS-*n*-alkyl RI standards analyzed by ESI+ mode (column SAE-A, temperature 50°C, aqueous 10 mM ammonium formate pH 2.8 and linear gradient 1-100% acetonitrile, slope 1% B min$^{-1}$). **(A)** BPI chromatogram of *S. sinuata* extract containing acylsucroses . **(B)** Extracted ion chromatogram (XIC) of common fragment ion *m/z* 162.022 (C$_5$H$_8$NO$_3$S$^+$) generated at 20 V collision potential. Illustration of GS-*n*-alkyl standard chemical structure and proposed fragmentation positions.
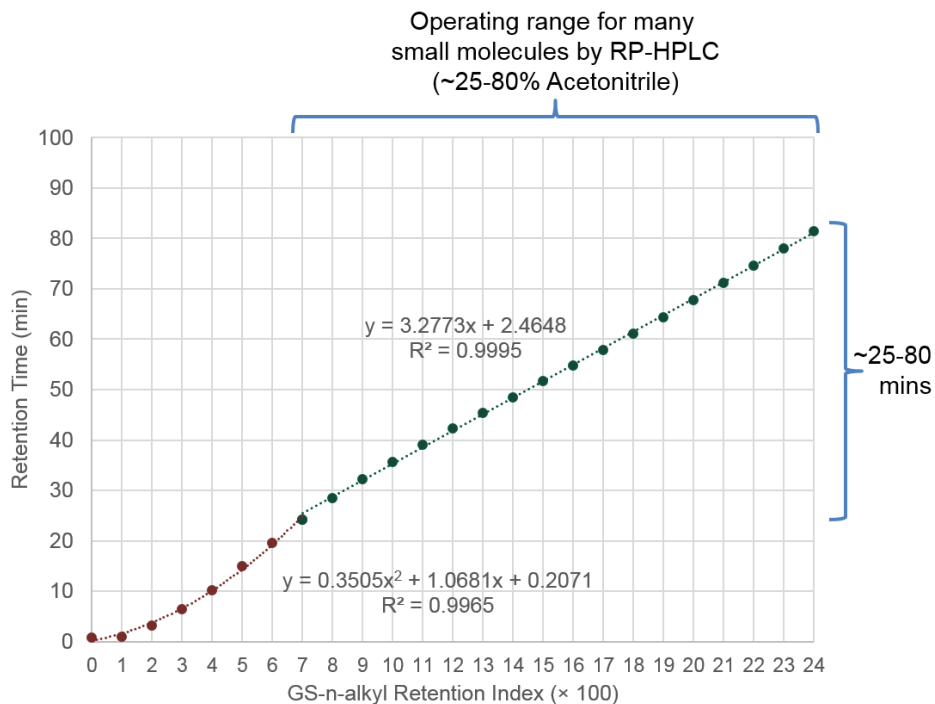
**Figure 4.2.** GS-n-alkyl HPLC retention times on a Supelco Ascentis Express C18 column using aqueous ammonium formate/acetonitrile gradient as a function of alkyl chain length (RI value) (n=3 replicates).

### 4.3.2 RI corrects day-to-day chromatographic variation

Even when chromatographic protocols are followed (aiming to standardize LC methods, instrument, column and solvents), day-to-day RT variations compromise RT comparisons across multiple samples, particularly when multiple isomeric forms are detected. Figure 4.3A-D illustrates chromatographic retention drift over the course of a 7-week period. Extracted ion chromatograms (XICs) for a group of S4:20:0 *S. sinuata* acylsucrose isomers labeled peaks #1-4 were measured on four separate dates ($C_{32}H_{54}O_{15}$, detected as $[M+NH_4]^+$ at $m/z$ 696.380, peaks #2-4 have structural identifications outlined in Chapter 2). Inspection of overlaid XICs for adjacently eluting GS-*n*-alkyl standards, GS-15 and GS-16 ($C_{25}H_{47}N_3O_6S$ and $C_{26}H_{49}N_3O_6S$, $[M+H]^+$ at $m/z$ 518.326 and 532.341), shows parallel RT drift of standards and analytes. For instance, acylsucrose peak #4 (S4:20:0(2,6,6,6)) exhibits significant RT variability. Relative to the first measurement (Figure 4.3A), RTs varied relative to the first measurement by -0.95, -0.83 and -0.38 mins (Figure 4.3B-D), or -1.7, -1.5 and -0.7% different respectively. However, the absolute percent deviations of RI values for peak #4 from the first RI measurement were < 0.03%.

| Col. SAE-A Shimadzu | RI #2 (% Diff) | RI #3 (% Diff) | RI #4 (% Diff) |
|---|---|---|---|
| 20180203 | 1529 | 1548 | 1555 |
| 20180209 | 1530 (0.03) | 1549 (0.03) | 1555 (-0.03) |
| 20180303 | 1529 (-0.02) | 1548 (-0.01) | 1555 (-0.004) |
| 20180322 | 1529 (-0.03) | 1548 (-0.02) | 1554 (-0.01) |
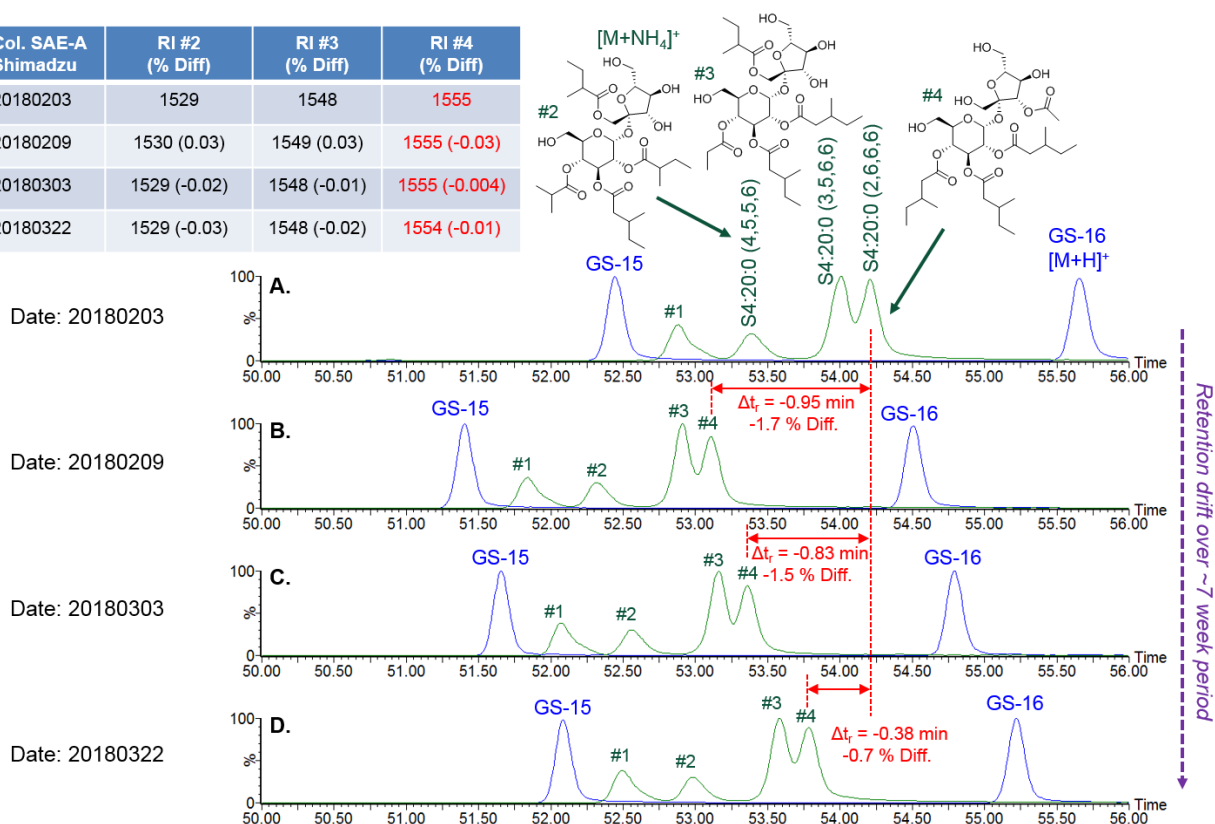
**Figure 4.3.** RT drift and application of RI correction applied to a group of S4:20:0 *S. sinuata* acylsucrose isomers ([M+NH4]+, green) with overlaid GS-*n*-alkyl standards ([M+H]+, blue) using column SAE-A. Peaks #2-4 have structural identifications based on NMR spectra. Dates when analyses were performed are displayed in year, month, day format. **(A)** 20180203, **(B)** 20180209, **(C)** 20180303, **(D)** 20180322.

RTs of the GS-*n*-alkyl standards and 16 *S. sinuata* acylsucrose metabolites were determined and the RI values calculated for each of the acylsucrose metabolites (Table 4.2). Values of standard deviation (SD) in RT of twelve measurements were approximately ±0.4 mins, or ~0.7% relative standard deviation (RSD), while the SD of RI values was approximately ±0.4, or ~0.02% RSD. To illustrate the day-to-day differences among the 16 acylsucroses, Figure 4.4A-B displays RT and RI percent differences for each of four triplicate analyses, with results displayed relative to the first date of analysis. This example shows that the differences in acylsucrose RI values are much smaller compared to variability of RT and demonstrates how GS-*n*-alkyl retention indexing has capacity for reducing day-to-day and/or sample batch-to-batch chromatographic variability.
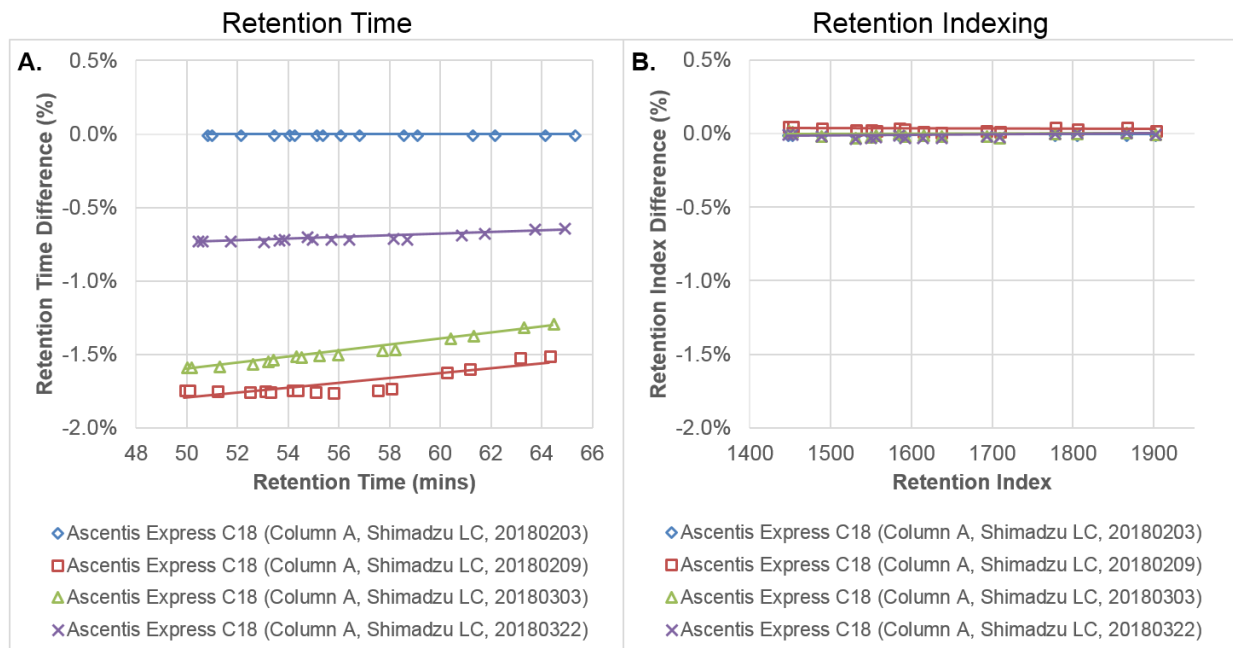
**Figure 4.4.** Percent difference comparisons of RT and RI values applied to 16 *S. sinuata* acylsucroses (each marker represents one of the acylsucrose metabolites) relative to the first set of measurements (column SAE-A, temperature 50°C, aqueous 10 mM ammonium formate pH 2.8 and linear gradient 1-100% acetonitrile, slope 1% B min⁻¹). Results are plotted against RT and RI value for visual purposes. Dates of analyses are indicated in the legend using year, month, day format. **(A)** RT percent differences relative to the first analysis. **(B)** RI percent differences relative to the first analysis.

### 4.3.3 Potential for cross-platform RI application

GS-*n*-alkyl standards show improved capacity for comparing retention data across laboratories. To demonstrate RI application across laboratory conditions, the *S. sinuata* extract was analyzed using two different LC solvent delivery systems (Shimadzu LC-20AD and Waters Acquity I-class pumps; same column SAE-A), and by performing measurements on the same make and model columns (SAE) produced from separate lots and at different stages of use (same Shimadzu LC system; columns SAE-A and SAE-B were new columns, while column SAE-C had been heavily used). For comparison, all results are referenced to column SAE-A measured on the Shimadzu LC system. Figure 4.5A-B demonstrates a significant shift to earlier RTs when the same column (SAE-A) is transferred to the Waters LC system, consistent with lower extracolumn dead volumes or differences in mixing of solvents in gradient formation. For instance, Peak #4 had RTs that differed by -1.62 mins, or -3.1% different. However, when RI was applied (values of 1555 and 1552 respectively for a single acylsucrose peak #4 (S4:20:0(2,6,6,6)), the difference is reduced to only -0.1%. Figures 4.5A and 4.5C compare two SAE columns prepared from separate lots (SAE-A and -

417

B) measured using the Shimadzu LC system. Peak #4 had RTs that differed by -0.72 min, or -1.3%, while the RI value (1559) only differed by 0.3%.
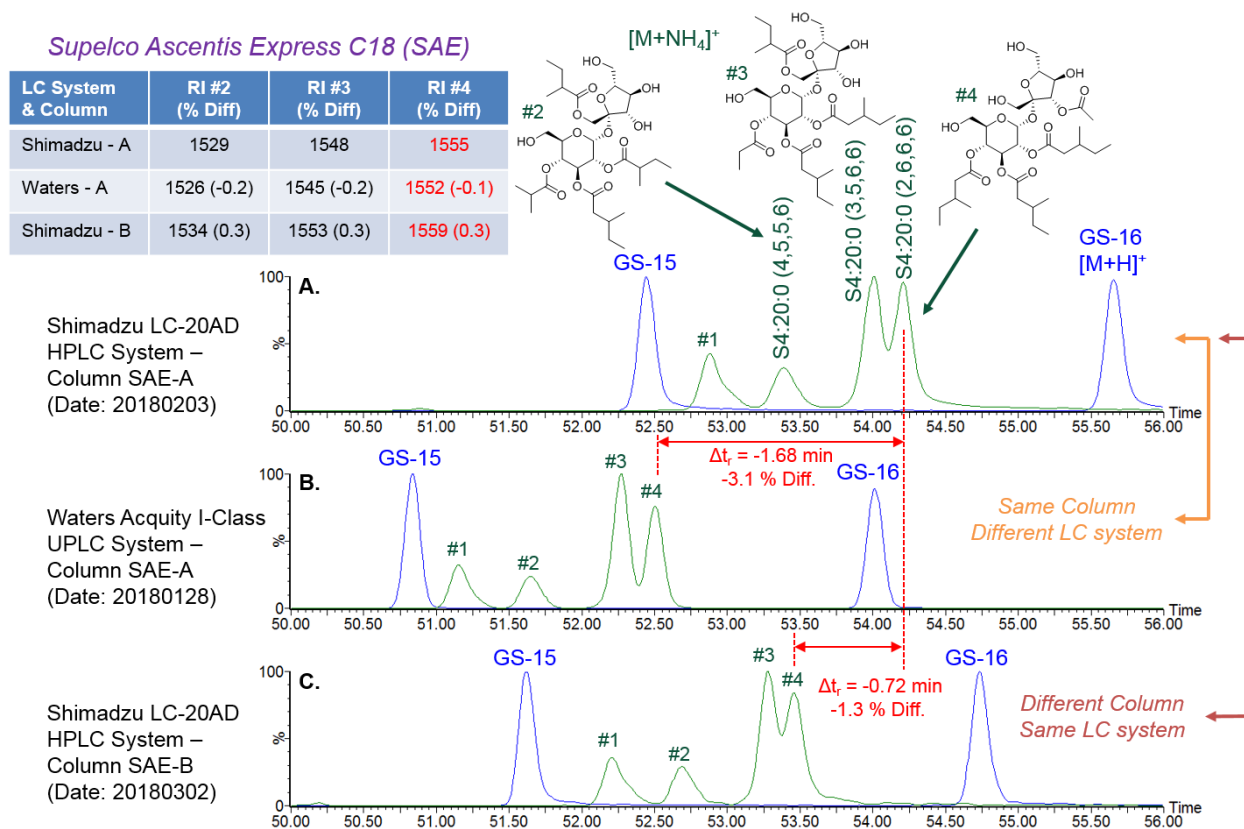


**Figure 4.5.** LC system and column dependence assessment using SAE columns. The group of S4:20:0 *S. sinuata* acylsucrose isomers peaks #1-4 is shown ([M+NH$_4$]$^+$, green) with overlaid GS-15 and GS-16 ([M+H]$^+$, blue). **(A)** Column SAE-A, measured using Shimadzu LC system. **(B)** Column SAE-A measured using the Waters LC system. **(C)** Column SAE-B measured using Shimadzu LC system.

RI values on different LC systems and columns were calculated for the 16 acylsucrose metabolites. Figure 4.6A-B illustrates percent differences relative to column SAE-A, measured using the Shimadzu LC system. Acylsucrose RTs showed substantial differences across LC systems, with as much as -3.5% difference for earlier eluting acylsucroses (Figure 4.6A). In contrast, their RI values differed over a range of -0.2 to 0.05% (Figure 4.6B). When the column was changed from SAE-A to SAE-B, RTs differed by ≥ 1.0%. When RI was applied, their values ranged from 0.3 to 0.7% different. The greater differences observed for columns SAE-A to SAE-B are attributed to batch-to-batch column variability. The RI results from a third column SAE-C, showed reduced capacity for RI correction. These findings are discussed further in Section 4.3.4.
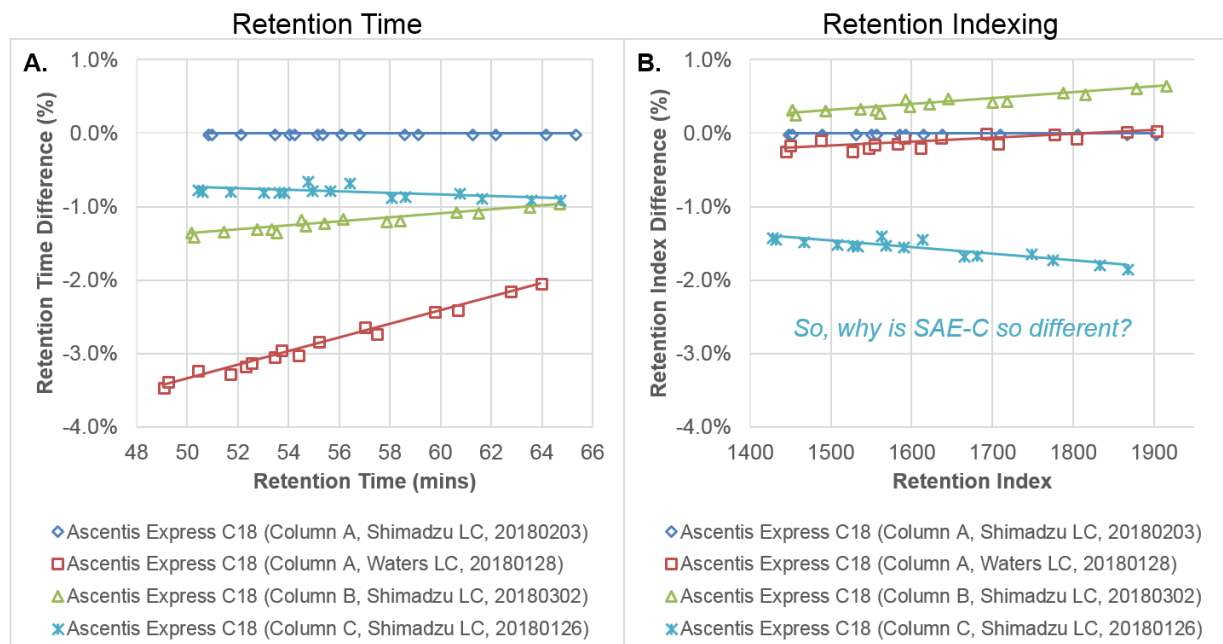
**Figure 4.6.** LC system and SAE columns RT and RI value percent difference comparison when applied to 16 acylsucroses from *S. sinuata* (percent differences are relative to column SAE-A measured using Shimadzu LC system). Columns, LC systems and dates of analyses are indicated in the legend. **(A)** RT percent difference **(B)** RI percent difference

### 4.3.4 GS-*n*-alkyls for use in column performance evaluation

Silica-based LC columns eventually show reduced performance with time and use owing to dissolution of the silica support [27], which dissolves slightly in the range of pH 2-7 and more substantially at higher pH [28]. Symptoms of column-related problems manifest as increased column backpressure, peak asymmetry, reduced column plate number, retention selectivity changes and retention time changes. To illustrate this, Figures 4.7A-B compare the group of S4:20:0 *S. sinuata* acylsucrose isomers analyzed with SAE columns at various stages of use. Column SAE-A was new (<15 hours use at the time of analysis), while column SAE-C had been heavily used (estimated ~900 hours), involving analysis of hundreds of acidic and basic sample injections (pH < 2 and pH > 8) and had been briefly operated at low pH while at temperatures above manufacturer-recommended conditions (aqueous phase pH 2.5, up to 70°C for a period > 20 hours).

Inspection of peaks #1-4 in Figure 4.7 reveals lower chromatographic resolution, peak fronting, as well as changes in selectivity for the heavily-used column SAE-C (notably, column backpressures of SAE-A, -B and -C were roughly the same). Without this direct comparison to column SAE-A, the reduced column

performance of column SAE-C could be overlooked. A surprising finding was that acylsucroses had RTs that were more consistent than RI values (peak #4 RTs showed -0.8% difference, while RI values were -1.5% different). These results are further illustrated in Figures 4.6A-B, where the 16 acylsucrose metabolites had RTs that differed by -0.7 to -0.9%, while RI values differed by -1.4 to -1.8%. In this instance, GS-*n*-alkyls RI values revealed evidence of column degradation in the form of changed RI values. In this context, GS-*n*-alkyl standards demonstrate value as a column performance test mixture. Comparison of SAE-A and SAE-C high energy XICs for GS-*n*-alkyl common fragment ion *m/z* 162.022 in Figure 4.8A-B reveals distinguishing peak shape differences for each column. Column SAE-C shows substantial peak broadening/tailing and later RTs with increasing GS-*n*-alkyl chain length (Figure 4.8B). Interestingly, this outcome was most apparent for standards with alkyl chain lengths ≥ GS-19. It may be notable that these standards had alkyl groups that exceeded the length of the octadecyl-C18 bonded phase. The peak width of standard GS-24 demonstrates this most clearly. It had $w_{50\%}$ of ~0.25 mins for column SAE-A, while column SAE-C was ~2.0 mins (Figure 4.8A-B). In addition, its RTs differed by ~5.5 mins. By comparison, RT as a function of RI value for column SAE-C reveals an upward inflection starting near GS-19 (Figure 4.8C). This feature was not observed for column SAE-A (or SAE-B, not shown). These results suggest that GS-*n*-alkyl standard peak shapes and RTs could be used as tool for evaluating column performance and column equivalency and identifying when a column should be replaced.
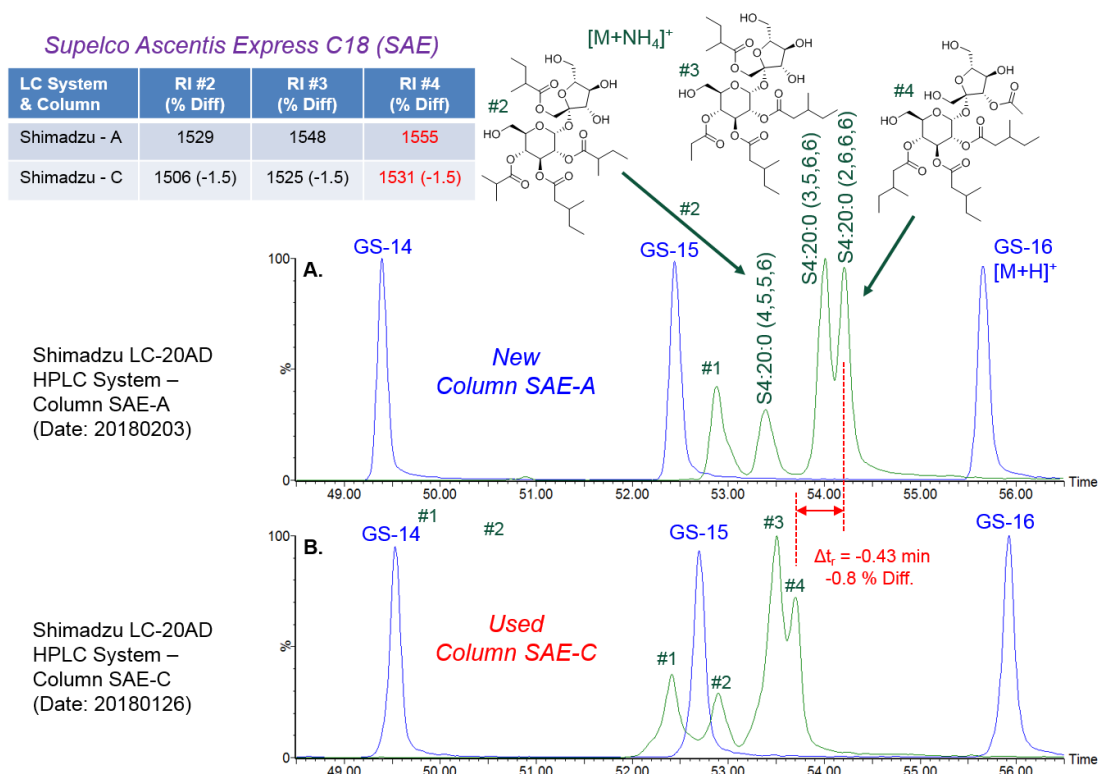
**Figure 4.7.** Column performance comparison. The group of *S. sinuata* S4:20:0 acylsucrose isomers is shown ([M+NH₄]⁺, green) with overlaid standards GS-14, -15, and -16 ([M+H]⁺, blue). (**A**) Analysis using column SAE-A. (**B**) Analysis using heavily used column SAE-C.
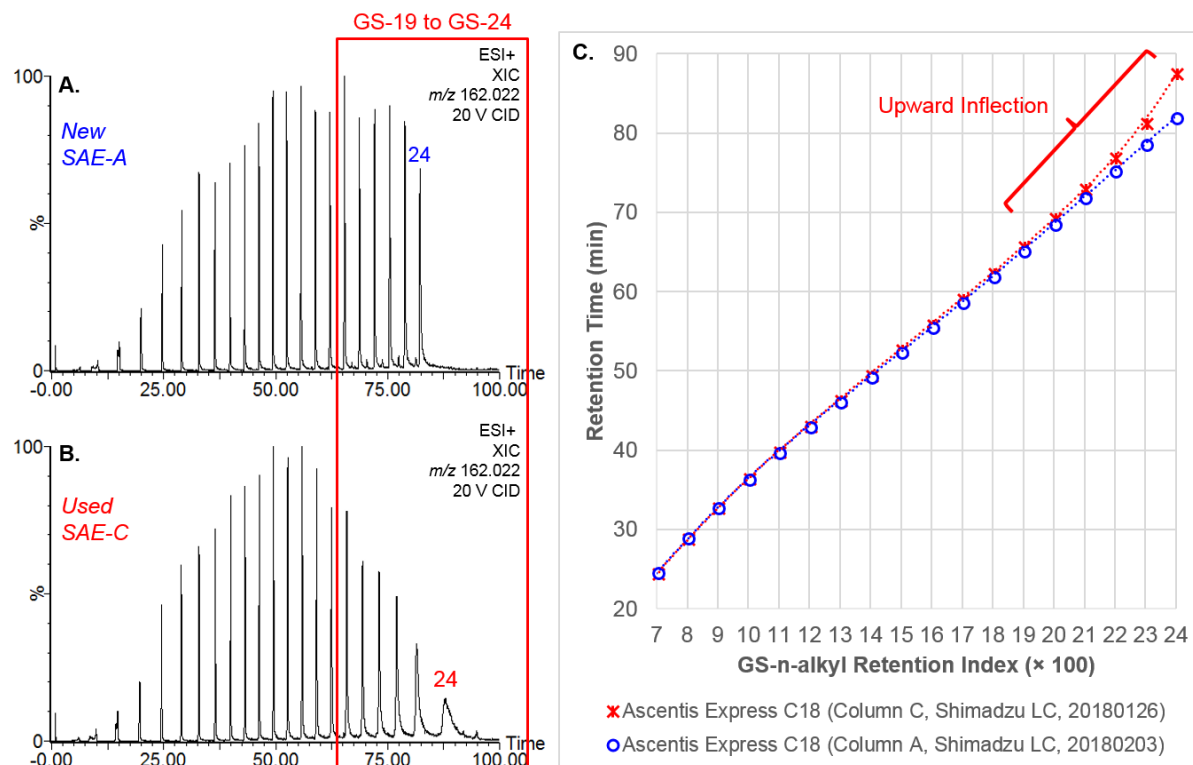
**Figure 4.8.** Column performance comparison using XICs for GS-*n*-alkyl common fragment ion *m/z* 162.022 (generated at 20 V collision energy). GS-19 to GS-24 are boxed in red. **(A)** Column SAE-A. **(B)** Column SAE-C. **(C)** Plot of RT against RI value using each column.

Considering these results, it is recommended that researchers establish a set of column equivalency acceptance parameters before making comparisons between GS-*n*-alkyl RI values, even when the same LC method conditions and column make/model is used. Exploration of predefined limits for column performance parameters such as RT, theoretical plate number (N) or $w_{50\%}$, peak asymmetry factor ($A_s$) and tailing factor (TF) may be used to establish column lifetime and/or batch-to-batch column equivalency. To illustrate this, Figure 4.9A-C displays XICs for GS-22 ($C_{32}H_{61}N_3O_6S$, $[M+H]^+$ at *m/z* 616.435) measured using columns SAE-A, -B and -C respectively. Columns SAE-A and -B had ~4.5 times more theoretical plates than the heavily-used column SAE-C. Perhaps the most perilous evidence for reduced column performance using column SAE-C are values of $A_s$ and TF. Column SAE-A had values of $A_s$ = 1.56 and TF = 1.58, while heavily-used column SAE-C had $A_s$ = 3.10 and TF = 2.49 respectively.
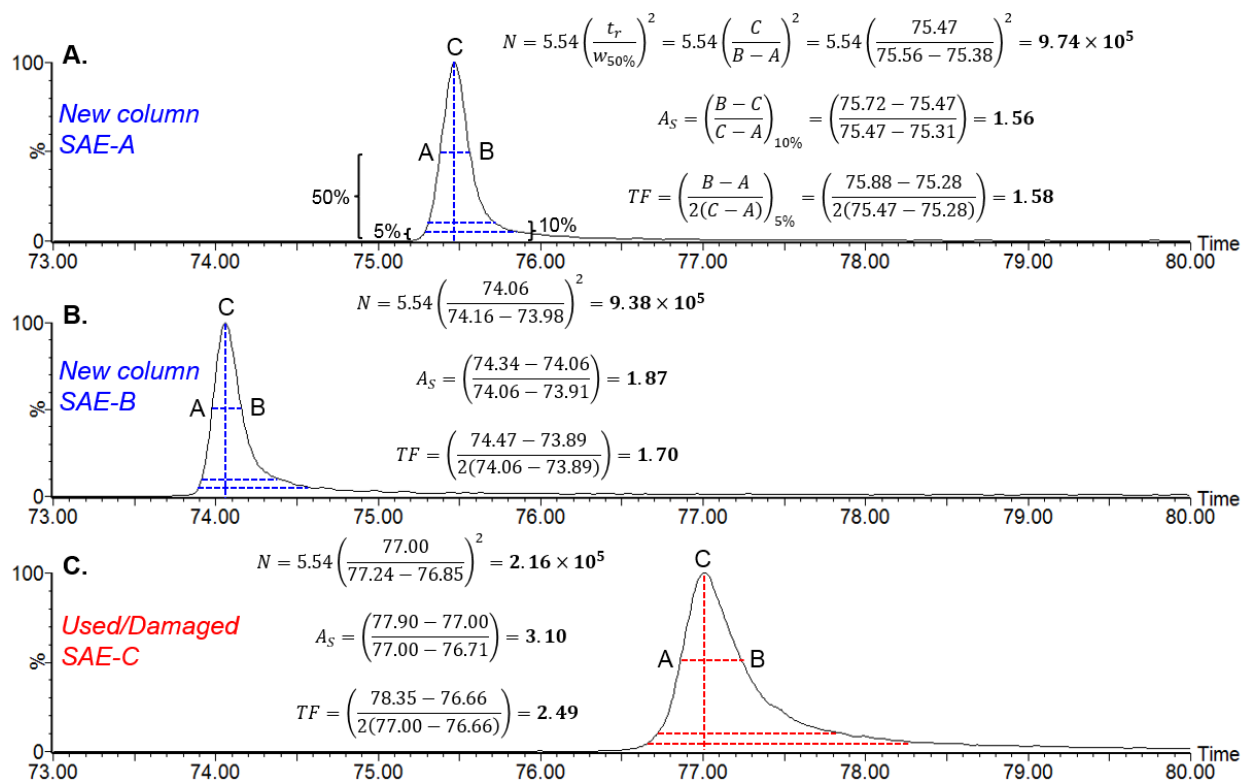
**Figure 4.9.** Analysis of column performance test parameters using standard GS-22 [M+H]⁺ as example. **(A)** Column SAE-A. **(B)** Column SAE-B. **(C)** Column SAE-C.

The increased peak asymmetry and tailing of GS-*n*-alkyl standards using column SAE-C is believed to be the product of interactions of the standard with underivatized silanol groups. These include cation-exchange interactions between the positively-charged amino group with negatively-charged ionized silanol groups (Figure 4.10). Furthermore, vicinal silanol groups, comprised of adjacently bonded silanols may contribute to bidentate hydrogen bonding with the carboxylic acids of the GS-*n*-alkyl standards (Figure 4.10). These and other interactions (such as adsorbed impurities) produce changes in retention and selectivity with column age and use [23]. Accordingly, the examination of RT and peak shape parameters of GS-*n*-alkyl standards offers a new approach for evaluating column performance, column lifetime and manufacturer batch-to-batch column equivalency.
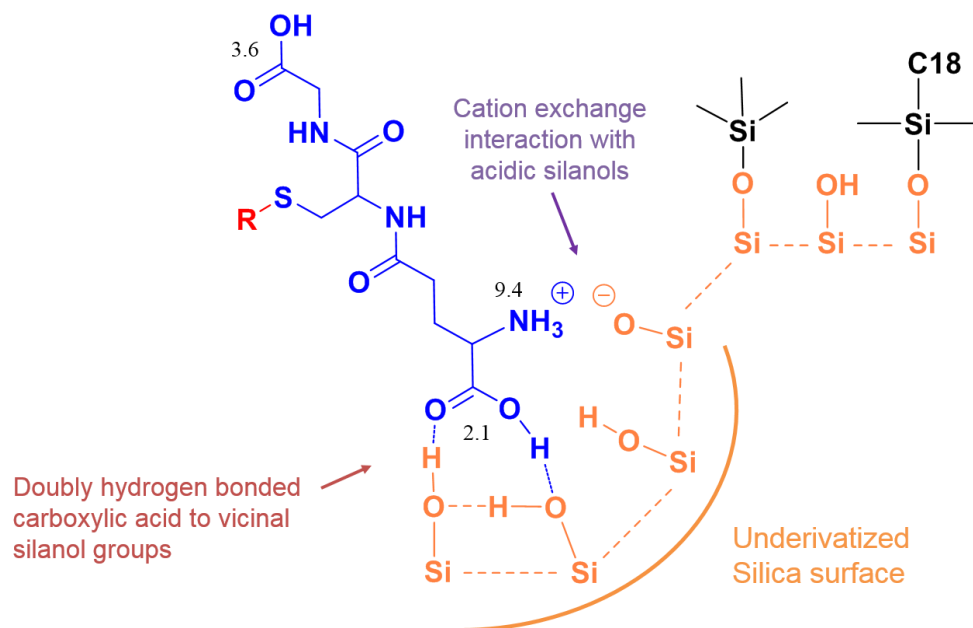
**Figure 4.10.** Predicted underivatized silanol group interactions. Approximate pKa values were calculated using ChemDraw Professional Software (Version 16.0.1.4).

### 4.3.5 C18 columns differ in retention selectivity

While hydrophobic solute-column and solute-solvent interactions dominate the C18 RP-HPLC retention process, other solute-column interactions contribute to differences in retention and selectivity across columns [23]. Differences among column chemistries, such as bonded phase chemistry (monomeric versus polymeric), silica supports (silica versus hybrid silica), ligand density ($\mu$moles$\cdot$m$^{-2}$), particle pore diameter, extent of end-capping (or residual silanols) and particle packing all play important roles in column retention and selectivity. Consequently, it is improbable that a single RI standard series could be developed that would yield identical RI values on all types of reversed-phase columns, even when the column bonded phase is of common alkyl length (e.g. C18).

To illustrate this, three C18 column types from leading manufacturers were chosen for side-by-side evaluation of RI reproducibility. Columns were chosen for their differences in particle technology, carbon loading levels (percent by weight of carbon on the stationary phase) and C18 bonding chemistry. SAE superficially porous columns (pore size 90 Å, surface area 120 m$^2\cdot$g$^{-1}$) were chosen because they can function at similar pressures to more traditional HPLC columns, while achieving comparable separation

efficiencies for a 2.7 µm particle size as UHPLC separations performed on sub-2 µm particles. An Agilent

Zorbax Eclipse Plus (AZEP) UHPLC column was selected because it is manufactured using fully porous

silica particles with monomeric C18 bonded phase but with lower carbon loading levels of 9% (pore size

95 Å, surface area 160 $m^2 \cdot g^{-1}$). Finally, a Waters Acquity BEH (WBEH) UPLC column was selected

because it is manufactured using fully porous bridged ethylene hybrid (BEH) silica particles with polymeric

(trifunctional) C18 bonded phase and has higher carbon loading levels of 17.7% (pore size 130 Å, surface

area 185 $m^2 \cdot g^{-1}$).
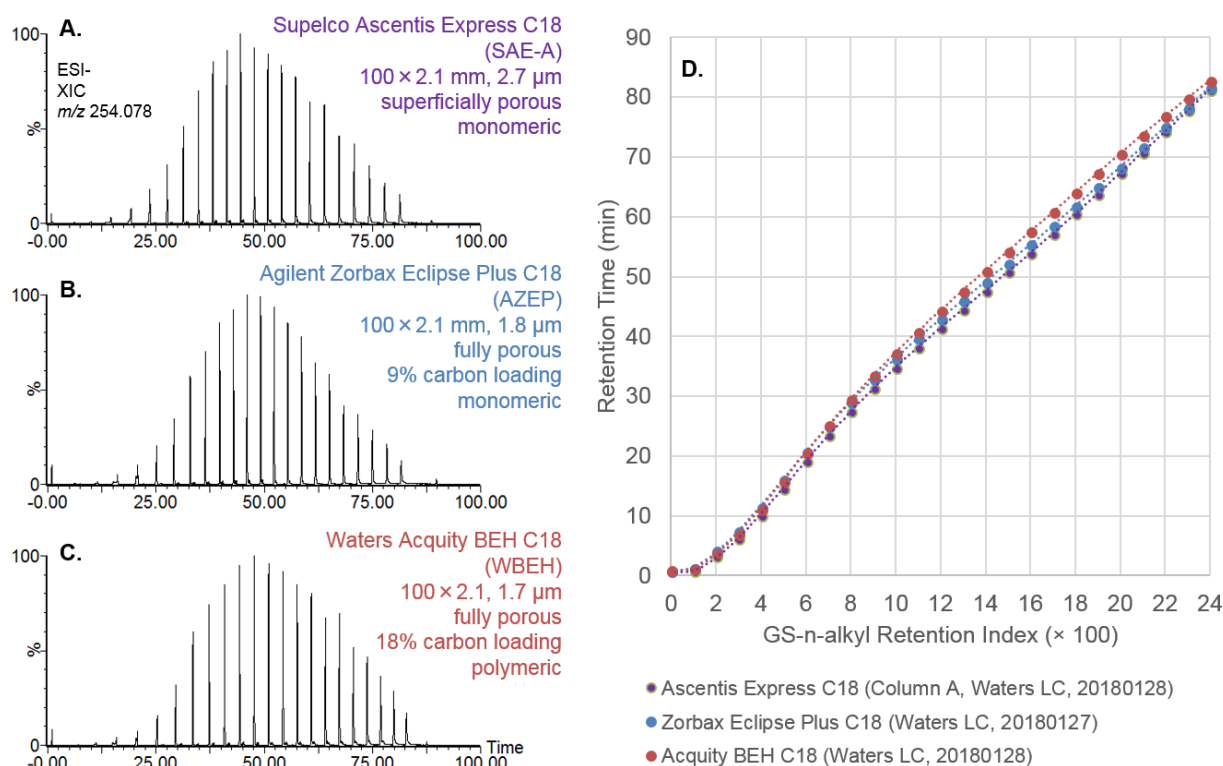


**Figure 4.11.** LC/MS chromatograms and retention of GS-*n*-alkyl standards using C18 columns from three leading manufacturers (measured using Waters LC system). XICs for common fragment ion *m/z* 254.078 ($C_{10}H_{12}N_3O_5^-$) generated at 20 V collision energy in ESI- mode. **(A)** Column SAE-A. **(B)** Column AZEP. **(C)** Column WBEH. **(D)** Plot of RT against RI value using each column.

Measurements using each column were performed in sequence on the same day (to reduce day-to-day

chromatographic variance) using the Waters Acquity LC system with identical LC conditions. Figure

4.11A-C demonstrates universal detection of the GS-*n*-alkyl series with ESI- mode common fragment ion

*m/z* 254.078 ($C_{10}H_{12}N_3O_5^-$) measured using columns SAE-A, AZEP and WBEH respectively. The RTs for

each column are plotted against RI value in Figure 4.11D. Notably, no other columns had an upward

inflection like the heavily used column SAE-C (Figure 4.9C). While RTs differed, the GS-*n*-alkyl standards had predictable elution order and similar peak shape for all three column manufacturers. Figure 4.12A-C compares the retention indexing of the group of S4:20:0 *S. sinuata* acylsucrose isomers analyzed on each column. Like the GS-*n*-alkyl standards, the peak shape and width of resolved acylsucroses were comparable. However, the selectivity is noticeably different for each column. By comparison, isomeric peaks #3 and #4 had reduced retention selectivity when column AZEP was used (Figure 4.12B), but were baseline resolved using column WBEH (Figure 4.12C). By comparison, it appears there are upwards of seven isomers detected using column WBEH, where columns SAE-A and AZEP resolved only four isomers. Accordingly, RI values were noticeably different across the columns. For instance, peak #4 had RI values of 1552 (SAE-A), 1583 (AZEP) and 1498 (WBEH) respectively. Thus, is not appropriate to assume that all C18 columns will have the same selectivity and RI values generated from different column manufacturers may not be equivalent.
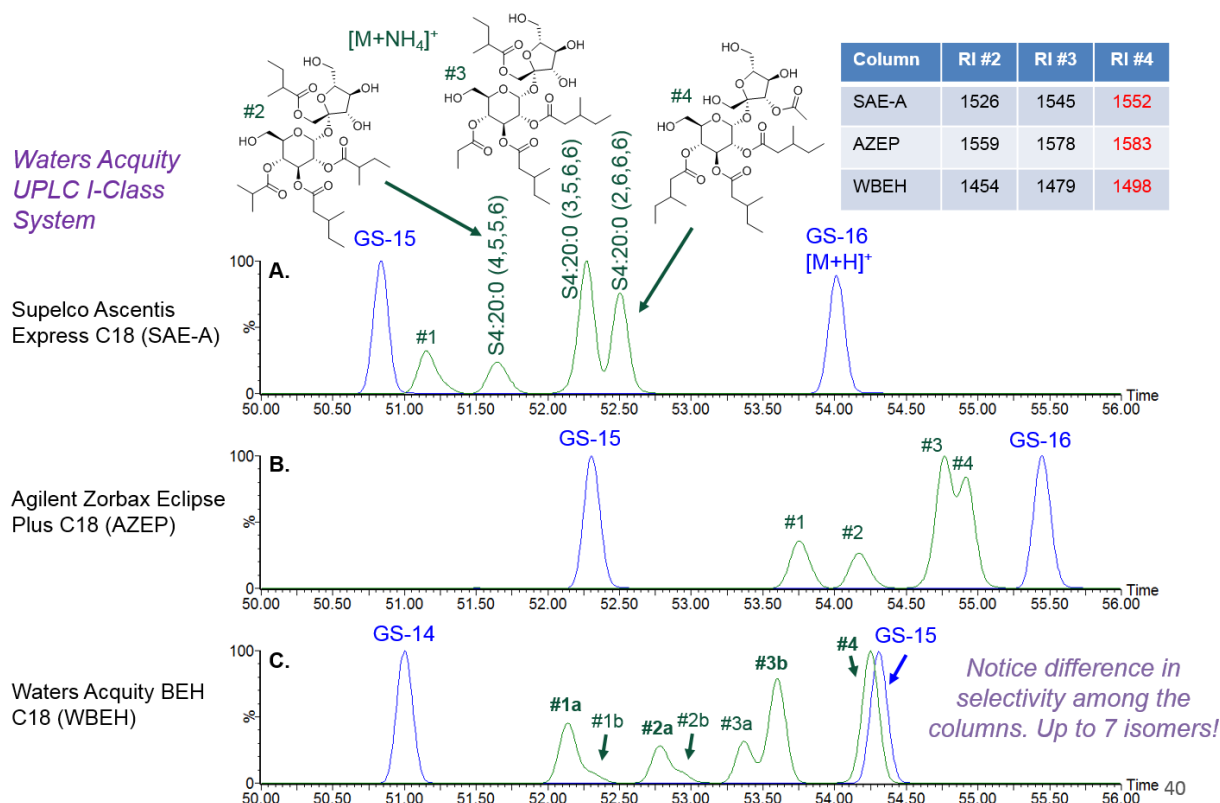
**Figure 4.12.** Column manufacturer selectivity and RI dependence. The group of S4:20:0 *S. sinuata* acylsucrose isomers is shown ([M+NH$_4$]$^+$, green) with overlaid GS-14, -15, and -16 ([M+H]$^+$, blue). (**A**) Column SAE-A. (**B**) Column AZEP. (**C**) Column WBEH.

## 4.3.6 RI dependence on mobile phase pH, column temperature and LC gradient

The pH of the mobile phase, column temperature and LC gradient play important roles for retention and selectivity of analytes by RP-HPLC. Therefore, RI reproducibility is likely to be dependent on alteration of these chromatographic variables. As the mobile phase pH increases, silanol ionization (pKa ~3.5-4.5) and the negative charge on the column increases [23]. Furthermore, the mobile phase pH may influence the ionization state of solutes. As the column temperature is increased, the analyte specific rate of exchange between the stationary and mobile phase is increased and RTs may be shortened. Furthermore, RP-HPLC analyses are used to analyze a wide range of compounds by changing aqueous and organic mobile phase components during gradient elution. As the mobile phase solvent strength increases with higher organic content, analyte specific interactions with the stationary phase may be reduced, solubility in the mobile phase increases, and the analyte advances through the column.

In this study, RI values were calculated for neutral acylsucrose metabolites, while GS-*n*-alkyl standards are comprised of ionizable amine and carboxylic acid functional groups that may change ionization state depending on mobile phase pH (pKa values of GS-*n*-alkyl standards were predicted from ChemDraw Software; 2.1, 3.6 and 9.4 respectively, see Figure 4.10). To evaluate the pH dependence of acylsucrose RI values when using ammonium formate aqueous mobile phase component (buffer range ~2.8-4.8), the pH was altered from 2.5-4.0 by 0.3 units (Appendix Figure 4.18, other LC conditions were held constant as outlined in Section 4.2.3.4). Figure 4.13A demonstrates the RI dependence on pH when applied to the group of S4:20:0 *S. sinuata* acylsucrose isomers (peaks #2-4). RI values exhibit a polynomial fit that rises sharply above pH 3.1, but appear to be pH independent $\leq 2.8$. RI values of all 16 acylsucrose metabolites followed a similar trend (Appendix Figure 4.19). The increase in RI values at pH>3 is due to earlier elution of the standard series (Figure 4.14A) relative to the acylsucrose metabolites. These observations are attributed to increased ionization of the residues of the GS-*n*-alkyl standard series as the mobile phase pH is increased, producing more negatively charged and water-soluble forms. These results show that mobile phase pH is an important consideration when attempting to reproduce RI values. Consequently, RI values were generated at pH 2.8 where the retention of the homologous standard series exhibited less pH dependence.
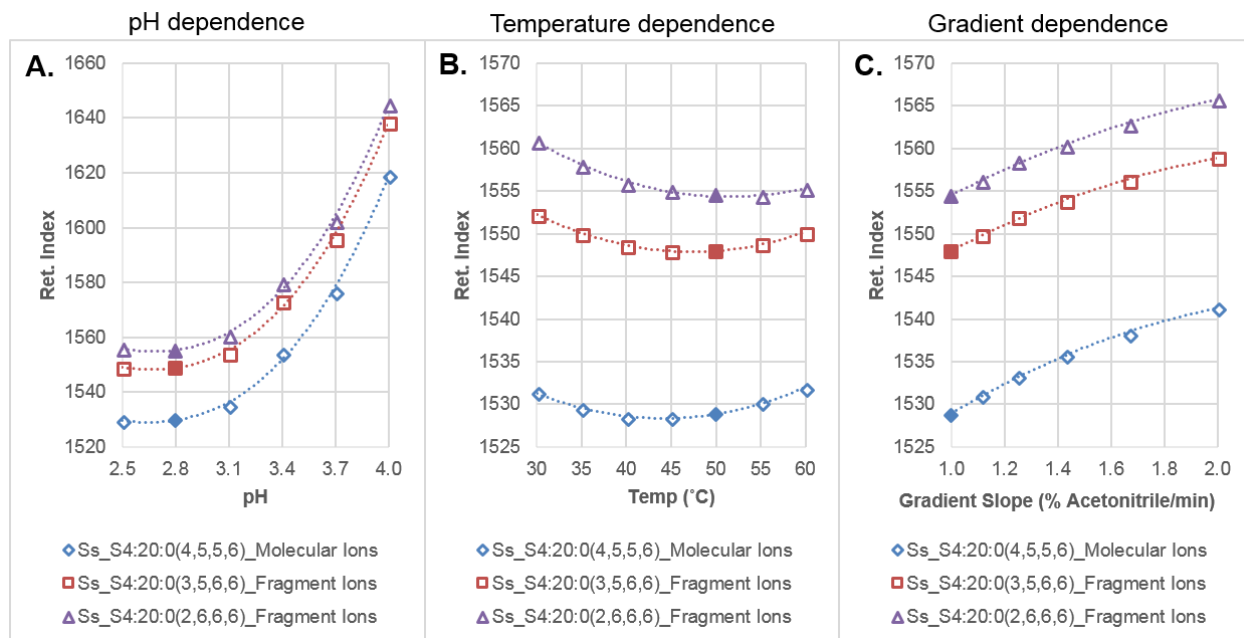
**Figure 4.13.** RI dependence of the group of S4:20:0 *S. sinuata* acylsucrose isomers when chromatographic conditions are altered using column SAE-A. Filled markers are equivalent analyses measured on separate dates. Standard deviations were too small to display error bars (see Appendix Figures 4.19, 4.21, 4.23) (**A**) Aqueous mobile phase pH 2.5-4.0, column temp. 50°C, gradient slope 1% acetonitrile·min$^{-1}$ (**B**) Column temperature 30-60°C, aqueous mobile phase pH 2.8, gradient slope 1% acetonitrile·min$^{-1}$ (**C**) Linear gradient slope 1, 10/9, 5/4, 10/7, 5/3, and 2% acetonitrile·min$^{-1}$, aqueous mobile phase pH 2.8, column temp. 50°C.

Examination of the GS-*n*-alkyl standards mobile phase pH dependence also provides useful information for evaluating symptoms of reduced column performance. Figure 4.14A-B compares the elution profile of the standard series using columns SAE-A and SAE-C. For column SAE-A, earlier RTs were observed for GS-*n*-alkyl standards as the mobile phase pH was increased (Figure 4.14A). However, column SAE-C showed excessive peak broadening/tailing and later RTs for GS-*n*-alkyl standards that was exacerbated by increased pH (Figure 4.14B). Surprisingly, as the Solvent A pH was increased to 4.0, the last standard to elute as a recognizable peak was GS-22, while standards GS-23 and GS-24 no longer eluted as resolved peaks. These results further illustrate reduced performance for column SAE-C, attributed to greater roles of underivatized silanol groups.
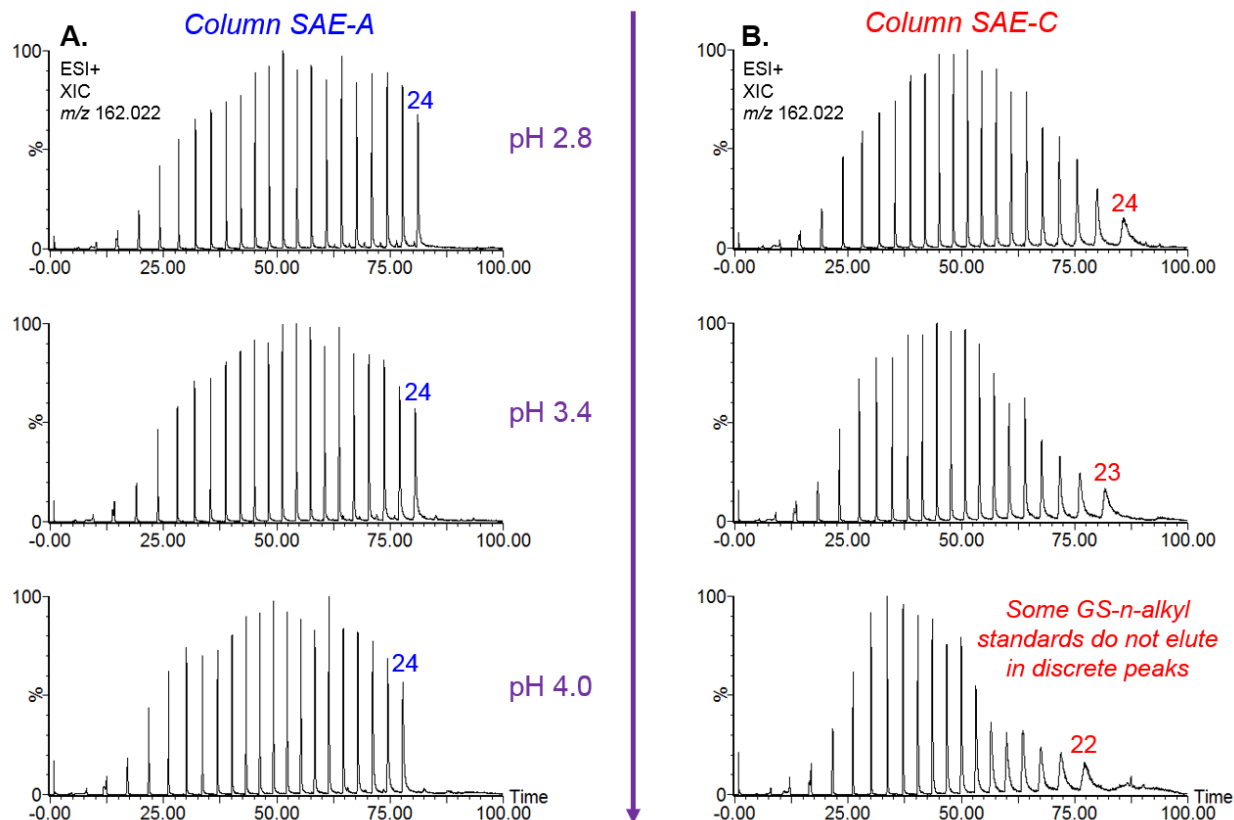
**Figure 4.14.** GS-*n*-alkyl pH dependence comparisons (XICs of common fragment ion *m/z* 162.022 by ESI+) using columns: **(A)** SAE-A and **(B)** SAE-C.

To evaluate the dependence of acylsucrose RI values on column temperature, the column temperature was varied from 30-60°C, in steps of 5°C (Appendix Figure 4.20, other LC conditions were held constant as outlined in Section 4.2.3.4). Figure 4.13B shows that RI values display unique U-shaped (concave up) behavior for the group of S4:20:0 *S. sinuata* acylsucrose isomers (peaks #2-4), reaching minima between ~45-50°C. These results show that RI values have analyte-specific dependence on column temperature. Interestingly, each of the 16 acylsucrose metabolites in this study displayed unique polynomial temperature dependence (Appendix Figure 4.21), suggesting two or more temperature-dependent retention processes (mixed retention modes) influence acylsucrose selectivity (and their RI values), such as hydrophobic and silanophilic interactions and changes in conformation of acylsucroses interacting at the surface of the stationary phase [29]. We expect the pKa of ionizable groups on GS-*n*-alkyl standards to be lowered by increased temperature producing more negatively charged and water-soluble forms [30]. However, this

does not account for the reversal of RI values with temperature for some acylsucroses, but not others. For instance, all tetra-esters exhibited U-shape behavior reaching minima between 40-55°C, while penta- and hexa-esters displayed monotonic decreases as temperature increased (Appendix Figure 4.21).

The temperature dependence behavior of RI values offers a novel approach for metabolite annotation that ought to be further explored. Table 4.3 summarizes the pH and temperature-dependence of RI values for the 16 acylsucrose metabolites in this study by applying a quadratic fit regression of experimental retention data (using 50°C (323 K) as reference). The pH and temperature-dependence illustrate how much (or how little) RI change occurs with varied conditions and different compounds. For instance, a 10°C change in column temperature (column temperature at 40°C) drives relatively small differences in RI for the tetra-esters, but were somewhat larger changes for penta- and hexa-esters.

For most RP-HPLC analyses of complex mixtures, gradients are preferred over isocratic elution because they offer superior peak shapes, peak capacity and ability to resolve analytes that have varied partition behavior within a reasonable time frame. To explore the gradient dependence of acylsucrose RI values, the gradient slope was varied using values of 1, 10/9, 5/4, 10/7, 5/3, and 2% acetonitrile·min$^{-1}$ (Appendix Figure 4.22, other LC conditions were held constant as outlined in Section 4.2.3.4). Figure 4.13C demonstrates RI gradient dependence for the group of S4:20:0 *S. sinuata* acylsucrose isomers (peaks #2-4). RI values for the set of isomers rise with increasing gradient slope. However, this feature is gradually transformed from a monotonic increase in RI, to an upside-down U-shape (concave down), to a monotonic decrease in the order of later eluting acylsucroses (Appendix Figure 4.23). These results demonstrate the gradient dependence of RI values, and are in accordance with investigations of gradient elution that have demonstrated the rate of change of analyte retention with gradient elution is solute-dependent [31].

**Table 4.3.** Coefficients for calculating acylsucrose retention index values as a function of aqueous mobile phase pH and column temperature using solvent A = 10 mM aqueous ammonium formate adjusted to pH with formic acid and solvent B = acetonitrile, 1% acetonitrile/min gradient. Column = Ascentis Express C18, 2.1 x 100 mm.  Solvent gradient slope at 1% acetonitrile/min.  Coefficients were determined by quadratic fit regression of experimental retention data, using 50°C (323 K) as reference.

| Acylsugar # | Number of acyl groups | Acylsugar annotation | pH dependence $RI_{(323K)} = A(pH)^2+B(pH)+C$ | | | Calculated $RI_{(323K)}$ at pH 2.8, 50°C (323K) | Temperature dependence $RI_{(T, K)}/RI_{(323K)} = D(T)^2+E(T)+F$ | | | Calculated RI at pH 2.8, 40°C (313K) | $\Delta RI^a$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | A | B | C | | D | E | F | | |
| 1 | 4 | Ss_S4:19:0(3,5,5,6) | 53.327 | -290.82 | 1841.4 | 1445.2 | 1.303E-05 | -8.139E-03 | 2.269 | 1442.3 | -2.8 |
| 2 | 4 | Ss_S4:19:0(2,5,6,6) | 53.947 | -295.09 | 1853.3 | 1450.0 | 7.900E-06 | -5.018E-03 | 1.797 | 1450.4 | 0.5 |
| 3 | 5 | Ss_S5:20:0(2,2,5,5,6) | 53.654 | -291.17 | 1880.6 | 1486.0 | 1.426E-05 | -9.492E-03 | 2.578 | 1491.9 | 5.9 |
| 4 | 4 | Ss_S4:20:0(4,5,5,6) | 55.568 | -303.39 | 1942.0 | 1528.2 | 9.412E-06 | -5.971E-03 | 1.947 | 1528.4 | 0.2 |
| 5 | 4 | Ss_S4:20:0(3,5,6,6) | 56.279 | -307.92 | 1968.2 | 1547.3 | 1.022E-05 | -6.544E-03 | 2.047 | 1547.2 | -0.1 |
| 6 | 4 | Ss_S4:20:0(2,6,6,6) | 56.872 | -311.83 | 1981.0 | 1553.8 | 8.296E-06 | -5.399E-03 | 1.878 | 1555.0 | 1.3 |
| 7 | 5 | Ss_S5:22:1(2,5,5,5,5$^T$) | 55.879 | -303.14 | 1992.9 | 1582.2 | 1.640E-05 | -1.075E-02 | 2.762 | 1588.3 | 6.1 |
| 8 | 5 | Ss_S5:21:0(2,2,5,6,6) | 56.429 | -307.03 | 2006.3 | 1589.0 | 1.372E-05 | -9.215E-03 | 2.545 | 1596.6 | 7.5 |
| 9 | 4 | Ss_S4:21:0(5,5,5,6) | 58.067 | -317.77 | 2046.5 | 1612.0 | 1.073E-05 | -6.889E-03 | 2.105 | 1611.9 | -0.1 |
| 10 | 5 | Ss_S5:25:4(2,5,5,5,8$^P$) | 57.311 | -311.00 | 2055.8 | 1634.3 | 1.561E-05 | -1.035E-02 | 2.715 | 1641.9 | 7.6 |
| 11 | 5 | Ss_S5:22:0(2,2,6,6,6) | 59.123 | -322.21 | 2128.9 | 1690.2 | 1.262E-05 | -8.535E-03 | 2.441 | 1700.1 | 9.8 |
| 12 | 4 | Ss_S4:22:0(5,5,6,6) | 60.358 | -330.68 | 2159.6 | 1706.9 | 1.071E-05 | -6.904E-03 | 2.113 | 1709.1 | 2.2 |
| 13 | 5 | Ss_S5:23:0(2,5,5,5,6) | 59.992 | -325.95 | 2217.1 | 1774.8 | 1.336E-05 | -8.882E-03 | 2.476 | 1783.2 | 8.4 |
| 14 | 4 | Ss_S4:23:0(5,6,6,6) | 62.015 | -339.55 | 2267.5 | 1803.0 | 1.026E-05 | -6.657E-03 | 2.080 | 1805.6 | 2.7 |
| 15 | 5 | Ss_S5:24:0(2,5,5,6,6) | 62.097 | -338.11 | 2323.7 | 1863.8 | 1.172E-05 | -7.774E-03 | 2.288 | 1869.2 | 5.3 |
| 16 | 6 | Ss_S6:25:0(2,2,5,5,5,6) | 62.301 | -338.22 | 2358.1 | 1899.5 | 1.168E-05 | -7.910E-03 | 2.338 | 1911.6 | 12.1 |

$^a$Difference between calculated retention index values at 40°C and 50°C.

432

### 4.3.7 RI dependence with methanol organic component

While acetonitrile is a widely used RP-HPLC organic solvent component, other solvents such as methanol may provide alternatives in retention and selectivity. Figure 4.15A-B demonstrates LC/MS analysis of the *S. sinuata* extract using methanol organic component instead of acetonitrile (Appendix Figure 4.24, other LC conditions were held constant as outlined in Section 4.2.3.4). Once more, the GS-*n*-alkyl standard series encompasses a wide RP-UHPLC retention range (Figure 4.15B). In contrast to the mostly linear elution profile observed for acetonitrile (Figure 4.2), GS-*n*-alkyl standards showed a sigmoidal relationship between chain length and retention time, with large gaps in RT for earlier eluting members of the standard series (Figure 4.15C). The gaps in RT increased to an inflection point near GS-7, after which RT density increased as the standard series alkyl chain length increased. RI values calculated for the set of 16 acylsucrose metabolites had SD in RT of approximately ±0.2 mins, or ~0.25% RSD, while the SD of RI values was approximately ±1.4, or ~0.1% RSD (Table 4.4).
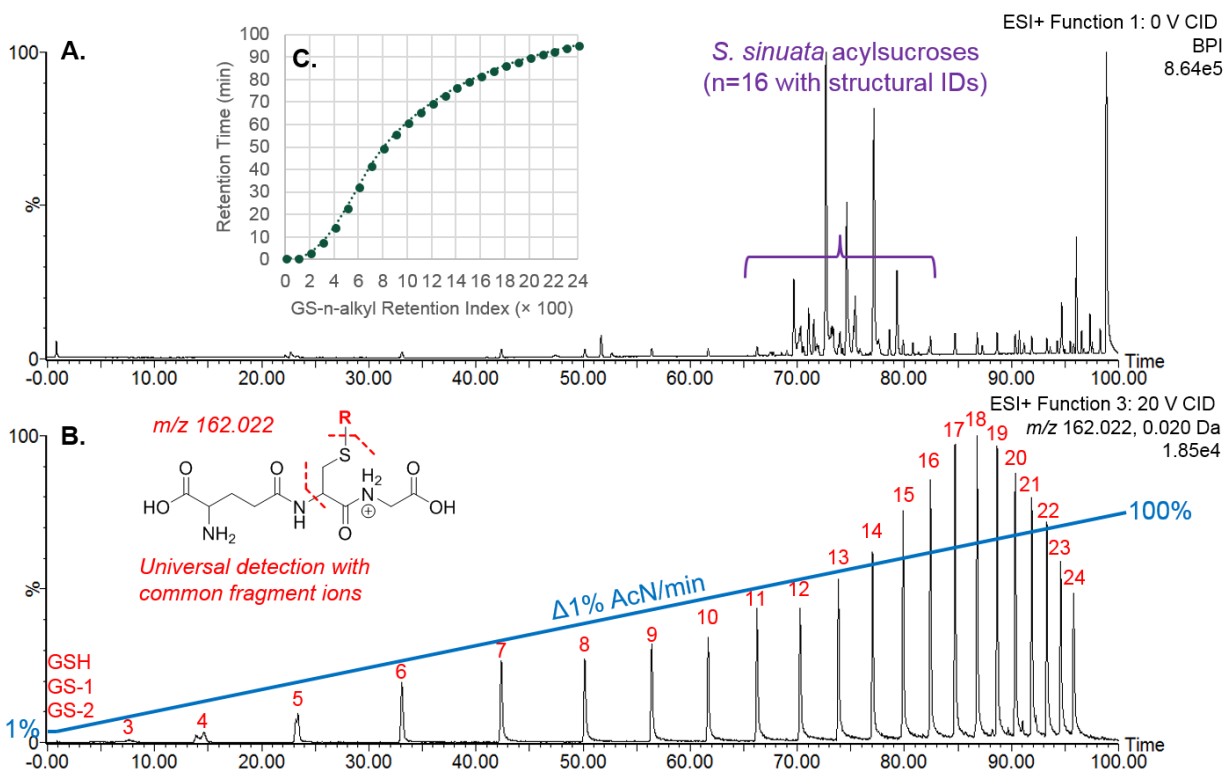


**Figure 4.15.** LC/MS of *S. sinuata* acylsucrose sample analyzed by ESI+ mode with methanol organic component. **(A)** BPI chromatogram of *S. sinuata* acylsucrose extract sample. **(B)** XIC of common fragment ion *m/z* 162.022 ($C_5H_8NO_3S^+$) generated at 20 V collision energy. **(C)** RT as a function of RI value.

RTs and selectivity of the 16 *S. sinuata* acylsucroses showed significant differences compared to acetonitrile. Acylsucrose RTs were shifted longer by 13.7-20.5 mins using methanol (Table 4.3). Additionally, the relative elution order of several acylsucroses had changed. Although the acylsucrose isomers in this study did not change elution order, several acylsucroses that had different numbers of ester groups changed elution order. Interestingly, the differences in retention and selectivity between the two organic components offers yet another approach for RI annotation. Relative to acetonitrile measurements, RI values were shifted to lower values that ranged from 228-445 difference (Table 4.4).

**Table 4.4.** LC/MS retention time and retention index values for 16 *S. sinuata* acylsucrose metabolites with structural identifications using methanol organic component (other LC conditions were held constant as outlined in Section 4.2.3.4). Analyses were performed on an Ascentis Express C18 column in triplicate (two by ESI- and one by ESI+) on four separate dates (n=12 total analyses). Acylsucrose ID's in bold are three S4:20:0 isomers (peaks #2-4) shown in Figure 4.3. The same ions listed in Table 4.2 were used for detection. Acylsucrose numbering is in order of elution by acetonitrile using methanol organic modifier (Table 4.2)

| Acylsucrose # | Acylsucrose ID | Avg Ret. Time ± SD (min) | Ret. Time RSD (%) | Methanol RI Value ± SD | RI Value RSD (%) |
|---|---|---|---|---|---|
| 1 | S4:19:0(3,5,5,6) | 70.13 ± 0.17 | 0.25 | **1202 ± 1.2** | 0.10 |
| 2 | S4:19:0(2,5,6,6) | 70.87 ± 0.18 | 0.25 | **1223 ± 1.2** | 0.10 |
| 3 | S5:20:0(2,2,5,5,6) | 69.50 ± 0.17 | 0.25 | **1186 ± 1.0** | 0.08 |
| 4 | **S4:20:0(4,5,5,6)** | 72.31 ± 0.17 | 0.24 | **1263 ± 1.2** | 0.10 |
| 5 | **S4:20:0(3,5,6,6)** | 73.17 ± 0.18 | 0.24 | **1287 ± 1.2** | 0.09 |
| 6 | **S4:20:0(2,6,6,6)** | 73.82 ± 0.18 | 0.25 | **1306 ± 1.7** | 0.13 |
| 7 | S5:22:1(2,5,5,5,5$^T$) | 71.36 ± 0.18 | 0.25 | **1236 ± 1.2** | 0.10 |
| 8 | S5:21:0(2,2,5,6,6) | 72.50 ± 0.18 | 0.24 | **1268 ± 1.2** | 0.09 |
| 9 | S4:21:0(5,5,5,6) | 74.42 ± 0.19 | 0.26 | **1325 ± 1.4** | 0.11 |
| 10 | S5:25:4(2,5,5,5,8$^P$) | 73.77 ± 0.19 | 0.25 | **1305 ± 1.5** | 0.11 |
| 11 | S5:22:0(2,2,6,6,6) | 75.22 ± 0.19 | 0.25 | **1350 ± 1.4** | 0.10 |
| 12 | S4:22:0(5,5,6,6) | 76.96 ± 0.20 | 0.26 | **1405 ± 1.6** | 0.11 |
| 13 | S5:23:0(2,5,5,5,6) | 76.92 ± 0.20 | 0.26 | **1404 ± 1.6** | 0.11 |
| 14 | S4:23:0(5,6,6,6) | 79.18 ± 0.20 | 0.25 | **1483 ± 1.5** | 0.10 |
| 15 | S5:24:0(2,5,5,6,6) | 79.12 ± 0.20 | 0.26 | **1481 ± 1.6** | 0.11 |
| 16 | S6:25:0(2,2,5,5,5,6) | 78.40 ± 0.20 | 0.26 | **1456 ± 1.6** | 0.11 |

Acylsucrose RI dependence on chromatographic parameters was also examined by altering mobile phase pH, column temperature and LC gradient. Figure 4.16A-C demonstrates their RI dependence when applied to the group of S4:20:0 *S. sinuata* isomers. RI values displayed polynomial U-shaped pH dependence with minima between pH 3.1-3.4 (Figure 4.16A). All 16 *S. sinuata* acylsucroses followed a similar trend (Appendix Figure 4.25). Unlike the RI determinations for varied column temperatures using acetonitrile, which displayed polynomial dependence, methanol RI values for all 16 *S. sinuata* acylsucroses exhibited linear relationship between column temperature and RI (Figure 4.16C), varying in slopes and y-intercept values (Appendix Figure 4.26). Finally, RI values declined with increasing gradient slope for the set of 16 acylsucroses (Figure 4.16C, Appendix Figure 4.27).

| pH dependence | Temperature dependence | Gradient dependence |
|---|---|---|
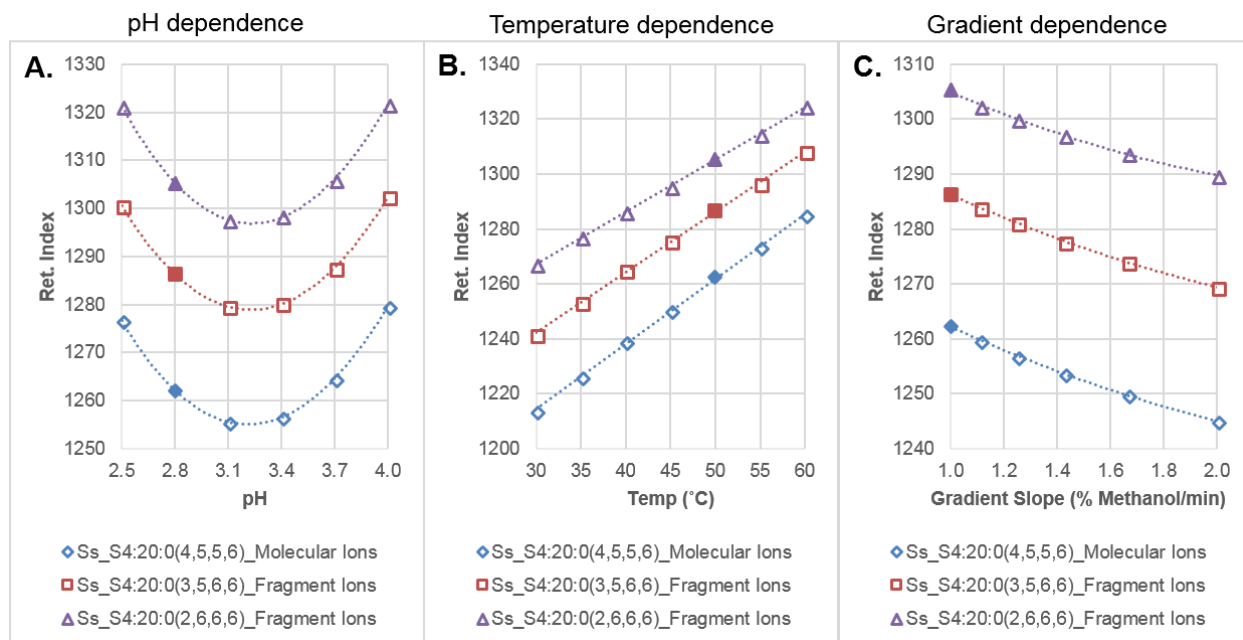
**Figure 4.16.** Retention index dependence of the group of S4:20:0 *S. sinuata* acylsucrose isomers when chromatographic conditions were altered using column SAE-A and methanol was the organic modifier. Filled markers are equivalent analyses measured on separate dates. Standard deviations were too small to display error bars (see Appendix Figures 4.25-27) **(A)** Aqueous mobile phase varied from pH 2.5-4.0, column temp. 50°C, gradient slope 1% methanol·min$^{-1}$ **(B)** Column temperature varied from 30-60°C, aqueous mobile phase pH 2.8, gradient slope 1% methanol·min$^{-1}$ **(C)** Linear gradient slope 1, 10/9, 5/4, 10/7, 5/3, and 2% methanol·min$^{-1}$, aqueous mobile phase pH 2.8, column temp. 50°C.

### 4.3.8 GS-*n*-alkyl standards further applications

In addition to retention indexing and column performance evaluations, GS-*n*-alkyl standards offer additional LC/MS applications. The standards may be used as mass axis calibration standards, or to examine LC/MS system performance and instrument sensitivity. Furthermore, RI values may be combined with MS/MS spectral databases to improve confidence in metabolite annotations, particularly when MS/MS spectra fail to distinguish isomeric compounds. Interestingly, MS/MS spectra of the standards exhibit common fragment ions that differ in relative parent and product ion abundances according to their alkyl chain length (Appendix Figures 4.28-33). For example, Figure 4.17A displays ESI- mode MS/MS spectra of standards GS-5, -10, -15 and -20 using a 10-40 V CID ramp (generated using Waters SONAR data acquisition platform, 10 Da bins widths, 0.5 s acquisition time). As the GS-*n*-alkyl chain length is increased, the product ion abundances are increased relative to the [M-H]$^-$ ion. Such relationship runs in contrast to the center-of-mass collision energies, which dictate that the maximum amount of energy deposited by a single ion-molecule collision should decrease as the precursor ion mass increases [32]. An illustration of

436

the relationship between ion abundances and chain length is presented in Figure 4.17B, which shows an increasing relationship (linear regression $R^2 = 0.9626$) when the ratio of common product ion $[C_{10}H_{12}N_3O_5]^-$ at $m/z$ 254.078 and $[M-H]^-$ ion abundances are plotted as a function of GS-$n$-alkyl chain length (Appendix Table 4.5). These results suggest that the standard set could operate as a fragmentation energy index for standardizing collision energy conditions or normalizing MS/MS fragment ion abundances generated across instruments (comparable to the robust 70 eV electron ionization spectral libraries used by GC/MS) [33, 34]. Comparison of the ratio of common product ion $[C_5H_8NO_3S]^+$ at $m/z$ 162.022 and $[M+H]^+$ ion abundances as a function of GS-$n$-alkyl chain length also showed an increasing relationship with chain length (Appendix Table 4.6 and Figure 4.34). The observation of increased relative product ion abundance is attributed to greater collision cross-section for longer chain lengths and higher probability for ion-molecule collision in the collision cell. As the GS-$n$-alkyl alkyl chain length increases so does the collisional cross section. The potential predictability of GS-$n$-alkyl collisional cross sections also suggests that this group of standards may be suitable as ion mobility reference standards [35].
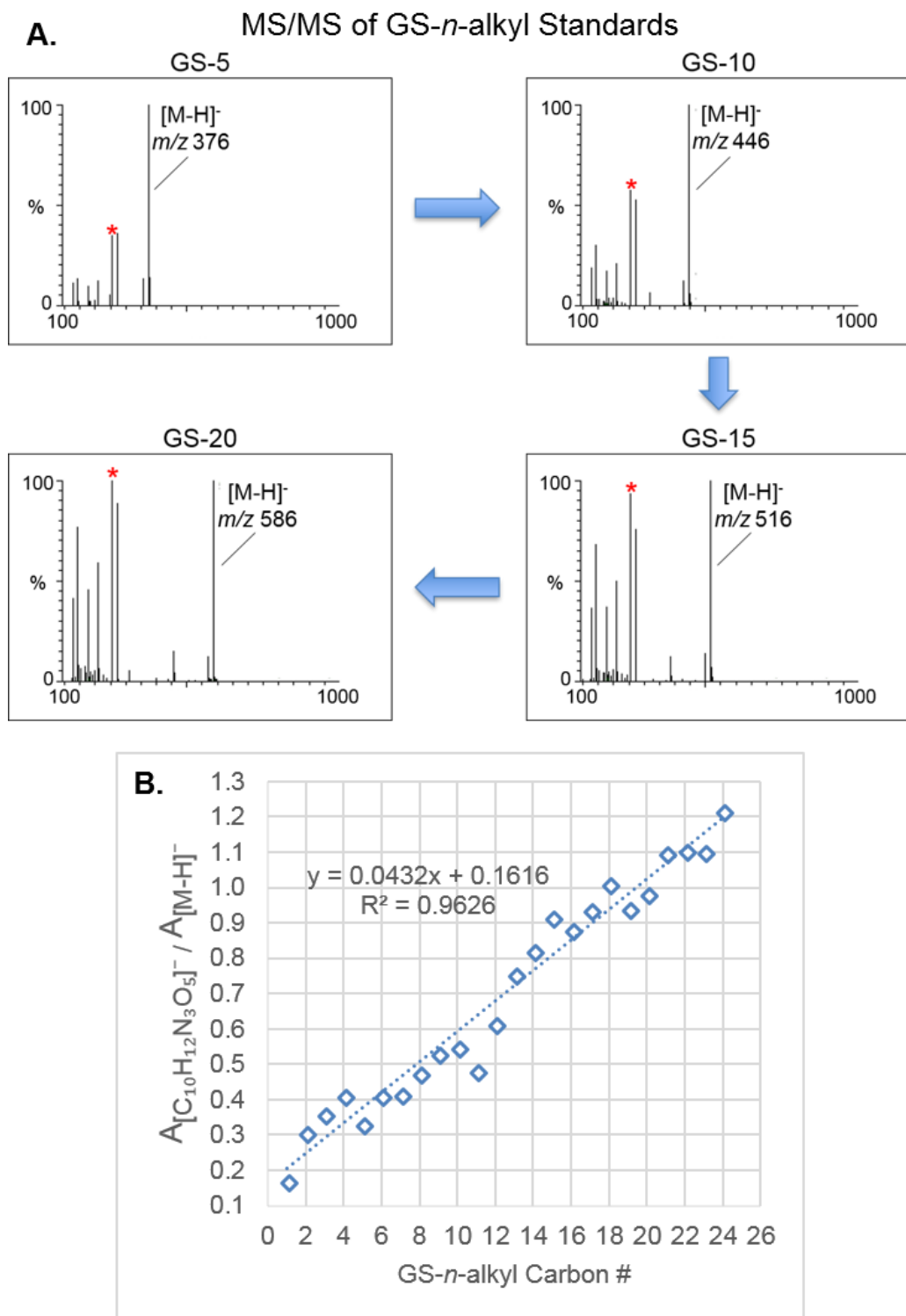
**Figure 4.17.** MS/MS product ion analysis of GS-*n*-alkyl standards. **(A)** Selected MS/MS product ion spectra of [M-H]⁻ ions. **(B)** Ratio of common fragment $[C_{10}H_{12}N_3O_5]^-$ = 254.078 (labeled by an asterisk in Figure 4.17A) and [M-H]⁻ ion abundances as a function of GS-*n*-alkyl chain length.

**4.4 Conclusions**

GS-*n*-alkyl retention indexing standards offer an increased level of confidence for metabolite identification and discovery. The standards encompass a wide RP-UHPLC retention range, are ionized by ESI in positive- and negative-ion modes and have characteristic fragment (product) ions that allow for their recognition using narrow window XICs for a single *m/z* value of a given polarity, simplifying reporting of standard RTs when added directly to a sample with minimal interference by sample constituents. This RI approach shows capacity for reducing day-to-day and/or sample batch-to-batch chromatographic variability and displays improved capacity for archiving retention data across laboratories.

In this study, RI values were calculated for acylsucrose metabolites. However, this RI approach is applicable for other analytes as well. The application of retention indexing across laboratories is highly dependent on the column, its stage of use and the manufacturer's ability to reduce batch-to-batch column variance. It is not appropriate to assume that all C18 columns will have the same selectivity. Thus, RI values generated from different column manufacturers and models may not be equivalent. Using the same manufacturer and model column is recommended for RI comparisons. However, even when the same manufacturer/model of LC column is used, it may be beneficial to evaluate GS-*n*-alkyl retention and peak shape parameters before attempting to draw accurate comparisons between RI values. In this regard, GS-*n*-alkyl standard LC/MS analysis shows promise for evaluating RP-HPLC column performance, batch-to-batch column reproducibility and column lifetime.

This report presents a thorough investigation of RI dependence by gradient elution using GS-*n*-alkyl standards while altering several important chromatographic experiment variables. Mobile phase pH, column temperature, LC gradients and organic solvent component all influenced RI values. We acknowledge that there are other factors that will influence RI values, including mobile phase ionic strength, buffer composition, mobile phase components, column length, flow rates and more. The ability of the analyst to reproduce and document LC experimental variables as accurately as possible will improve RI reproducibility. While RI values are not accurate enough to provide unambiguous determination of

structures on their own, they may act as refinement filter for comprehensive structural elucidation workflows, dereplicating metabolite discovery. Lastly, we hypothesize that GS-*n*-alkyl standards have applications that are not limited to retention indexing, such as fragmentation energy index standards and as ion mobility reference standards.

**APPENDIX**
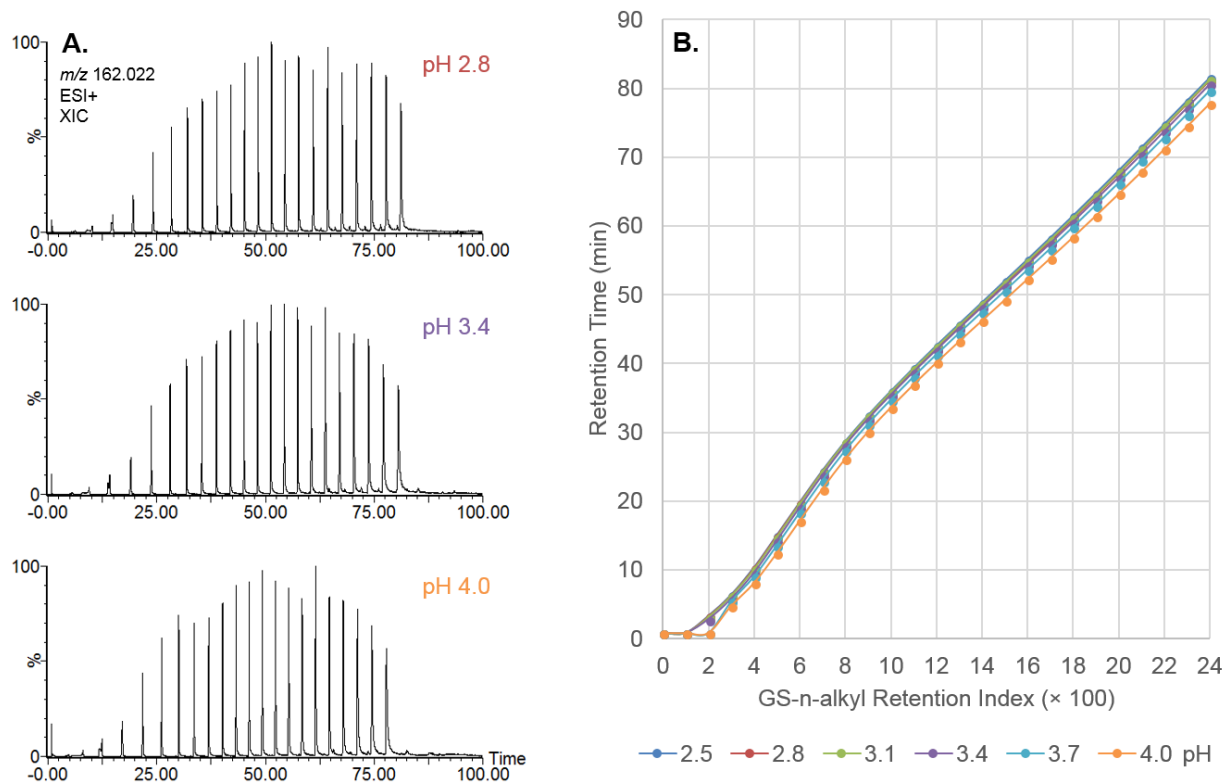
**Figure 4.18.** LC/MS chromatograms showing GS-*n*-alkyl pH dependence (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted with formic acid, column temperature 50°C, gradient slope 1% acetontrile·min$^{-1}$, column SAE-A) **(A)** XICs of common fragment ion *m/z* 162.022 by ESI+. **(B)** RT as a function of RI value.
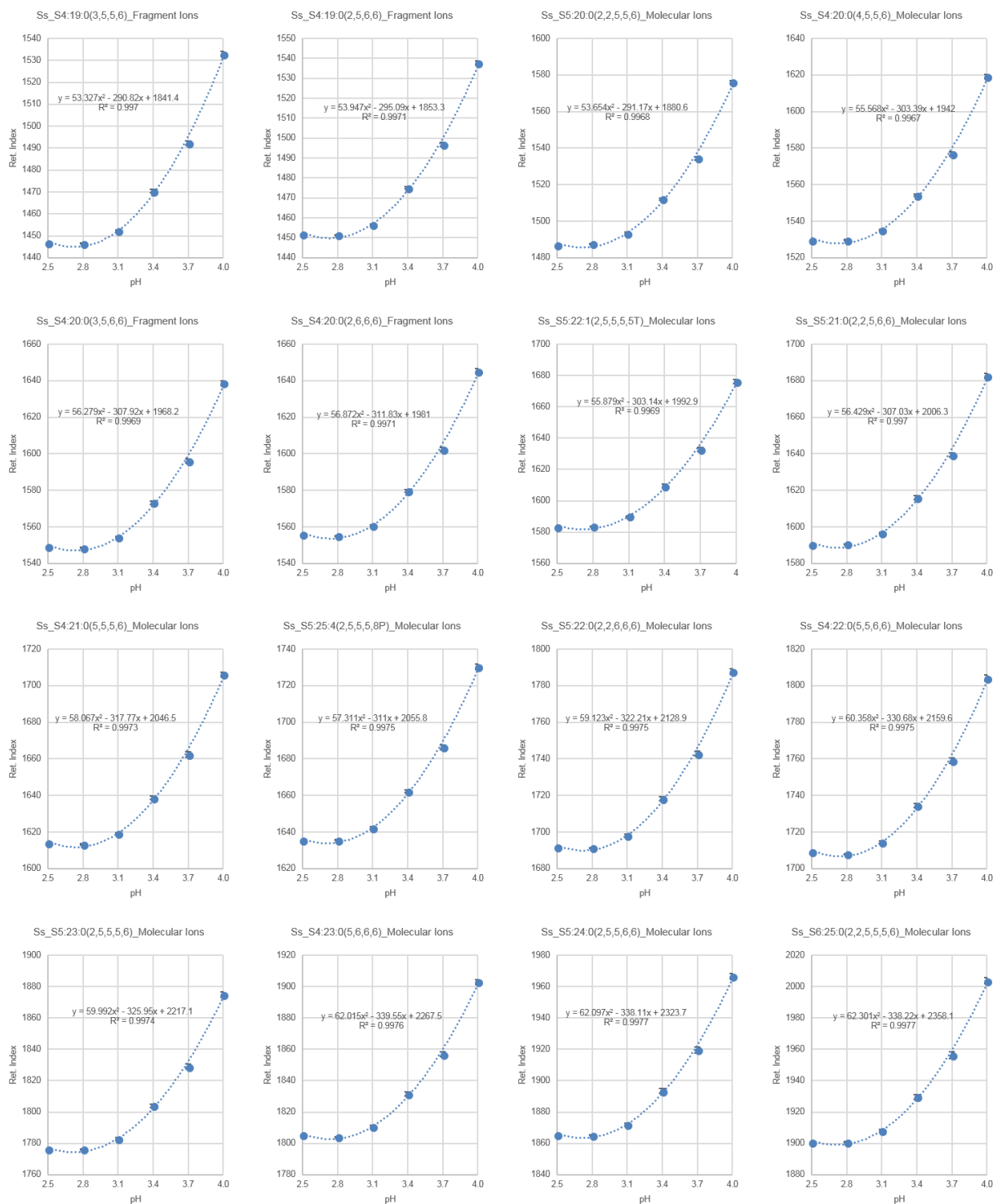
**Figure 4.19.** RI value pH dependence of 16 *S. sinuata* acylsucrose metabolites (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted with formic acid, column temperature 50°C, gradient slope 1% acetontrile·min⁻¹, column SAE-A).
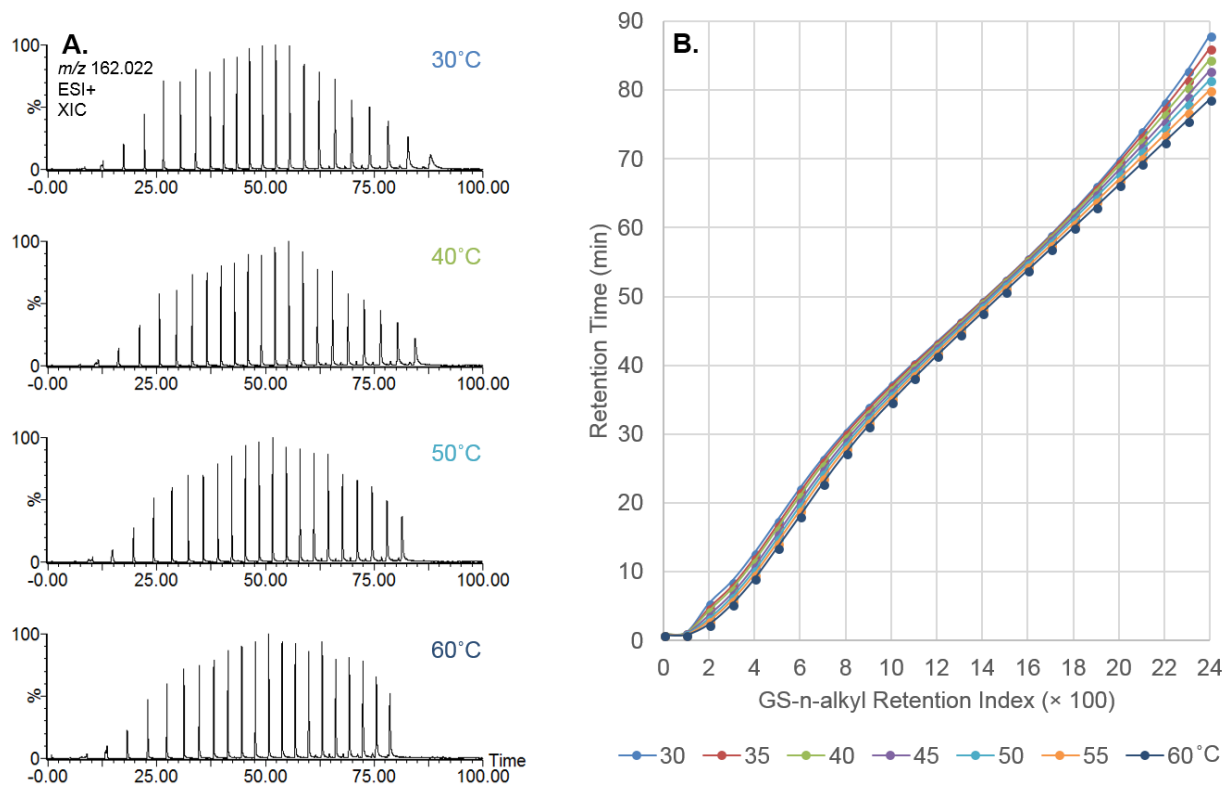
**Figure 4.20.** LC/MS chromatograms showing GS-*n*-alkyl temperature dependence (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, gradient slope 1% acetontrile·min$^{-1}$, column SAE-A). (**A**) XICs of common fragment ion *m/z* 162.022 by ESI+. (**B**) RT as a function of RI value.

444
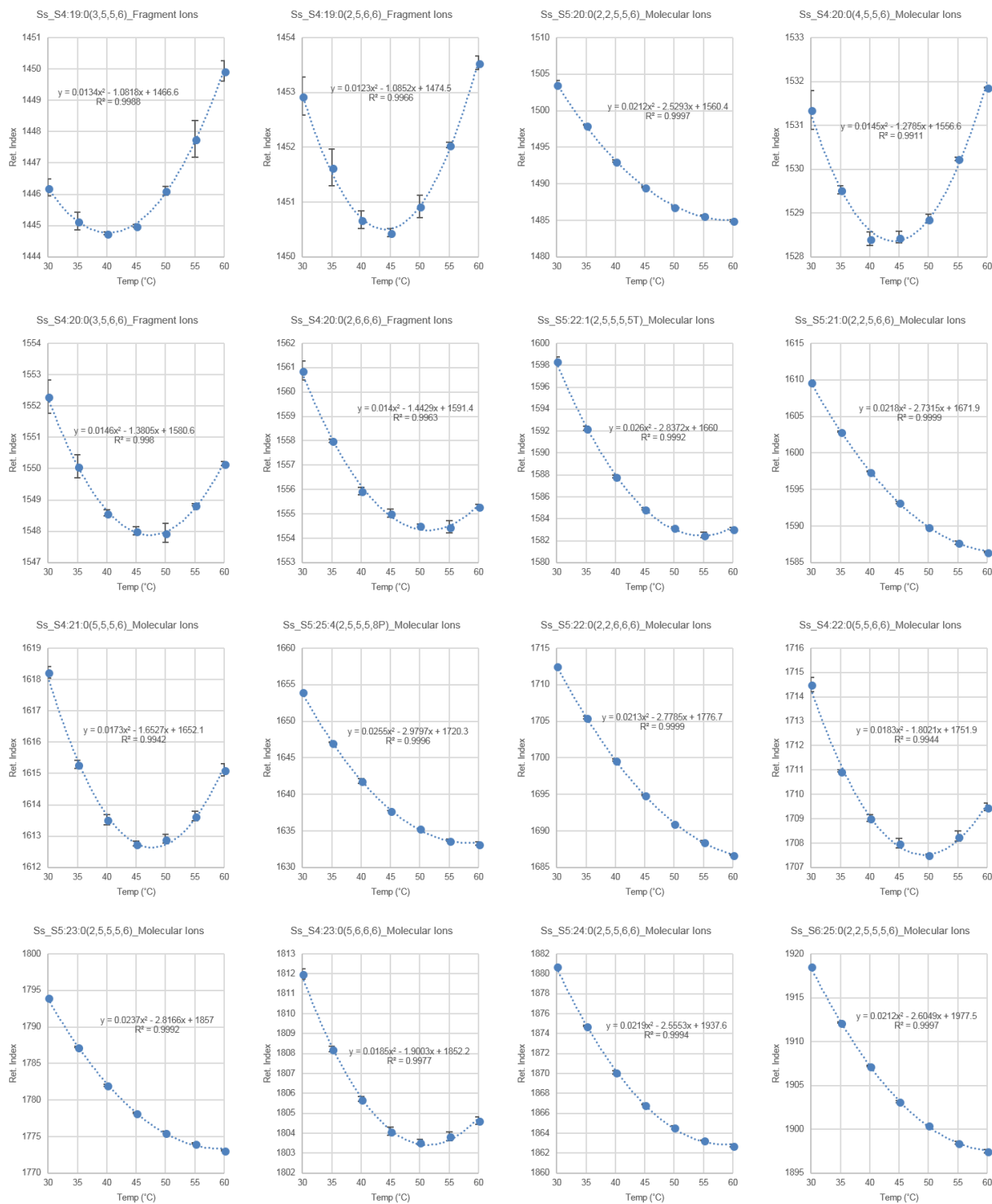
**Figure 4.21.** RI value column temperature dependence of 16 *S. sinuata* acylsucrose metabolites (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, gradient slope 1% acetontrile·min⁻¹, column SAE-A).

445

**Figure 4.22.** GS-*n*-alkyl gradient slope dependence (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, column temperature 50°C, column SAE-A). **(A)** XICs of common fragment ion *m/z* 254.078 by ESI-. **(B)** RT as a function of RI value.

446

**Figure 4.23.** RI value gradient dependence of 16 *S. sinuata* acylsucrose metabolites (acetonitrile organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, column temperature 50°C, gradient slope 1, 10/9, 5/4, 10/7, 5/3, and 2% acetontrile·min⁻¹, column SAE-A).

447
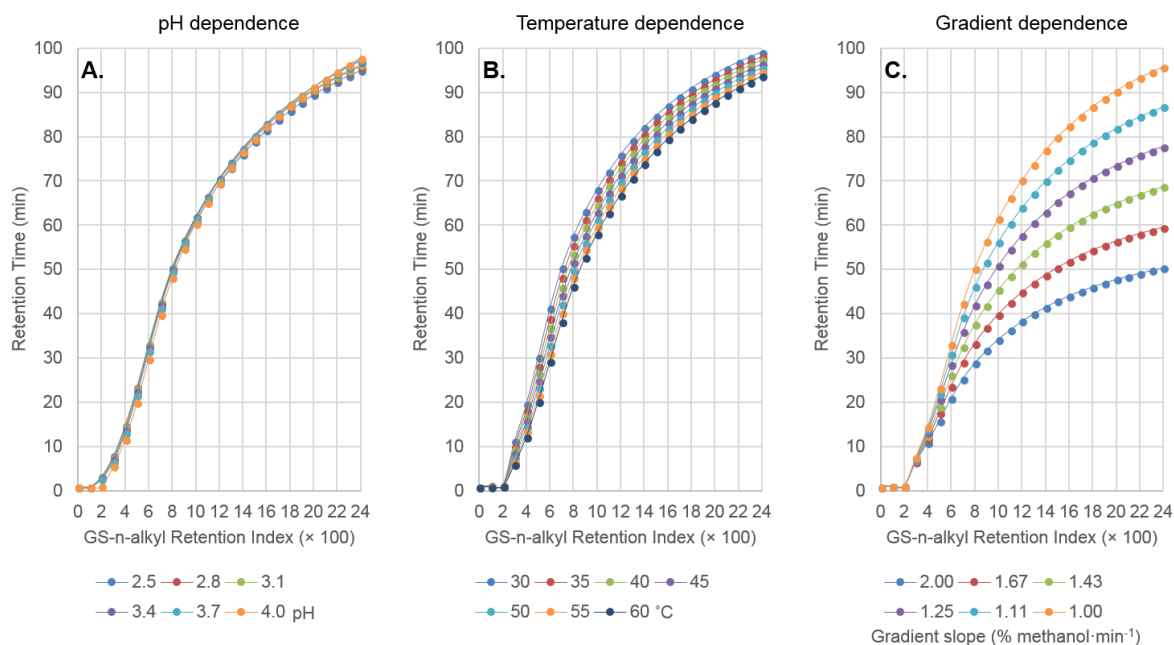
**Figure 4.24.** GS-*n*-alkyl dependence with altered chromatographic conditions using methanol organic component (column SAE-A). **(A)** RT as a function of RI value when pH is altered (aqueous component 10 mM ammonium hydroxide adjusted with formic acid, column temperature 50°C, gradient slope 1% acetontrile·min$^{-1}$). **(B)** RT as a function of RI value when column temperature is altered (aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, gradient slope 1% acetontrile·min$^{-1}$). **(C)** RT as a function of RI value when the LC gradient is altered (aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, column temperature 50°C, gradient slope 1, 10/9, 5/4, 10/7, 5/3, and 2% acetontrile·min$^{-1}$).
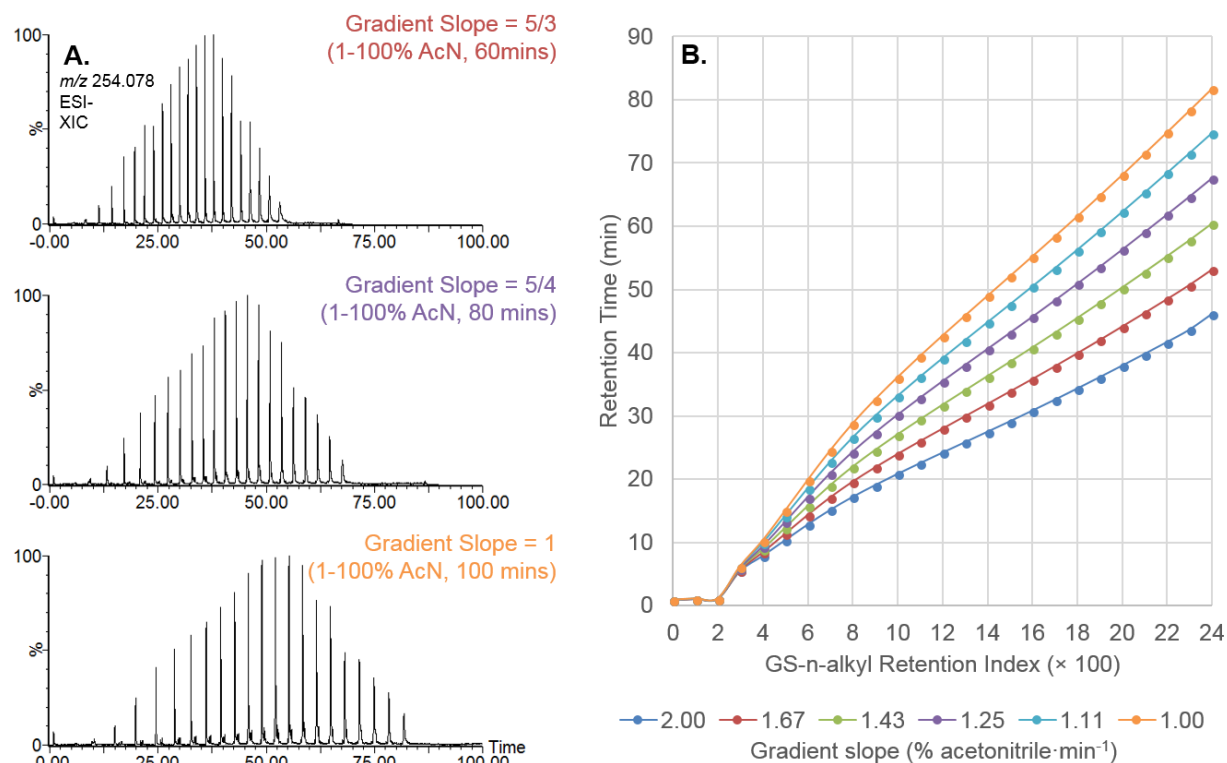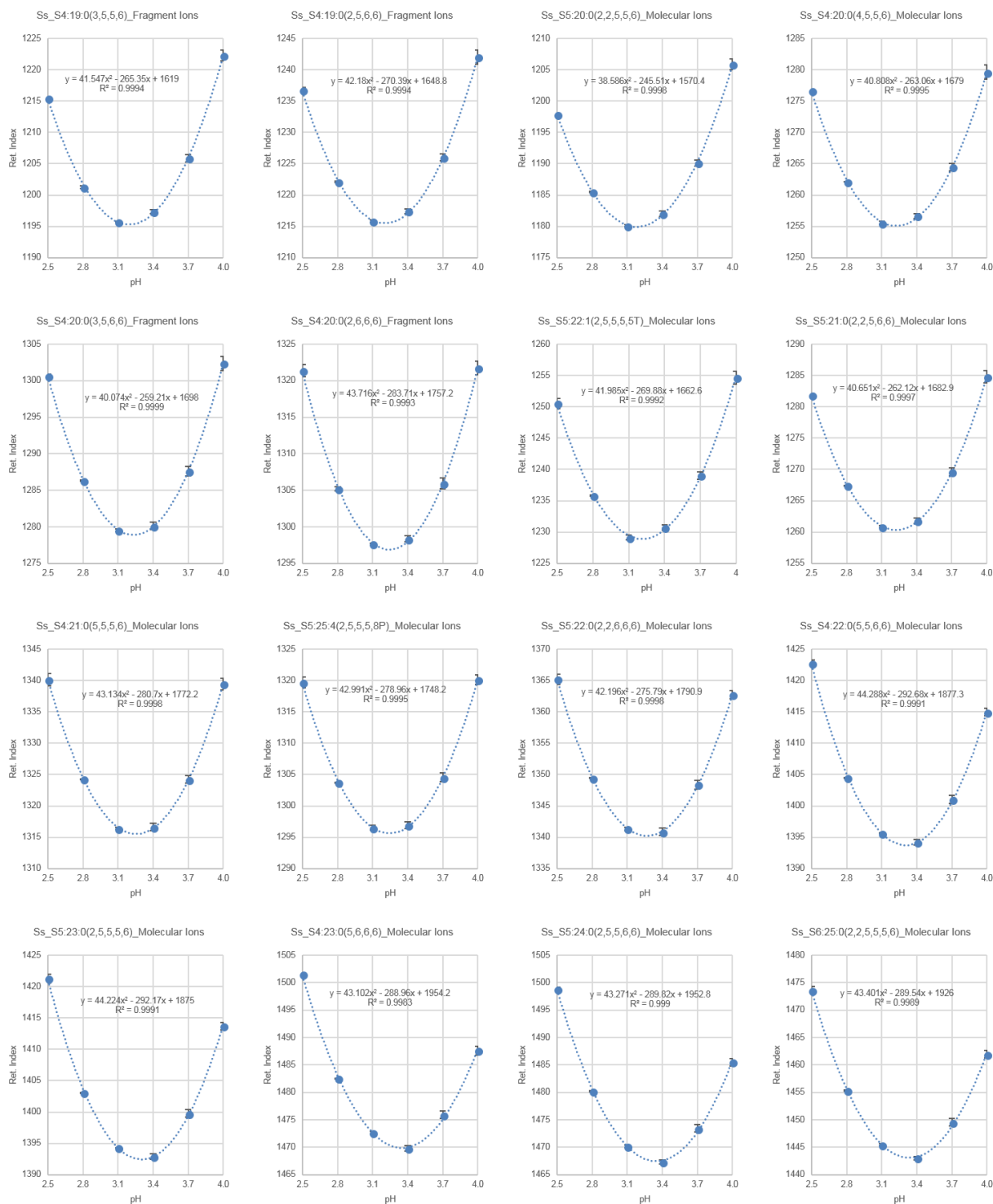
**Figure 4.25.** RI value pH dependence of 16 *S. sinuata* acylsucrose metabolites (methanol organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted with formic acid, column temperature 50°C, gradient slope 1% acetontrile·min⁻¹, column SAE-A).
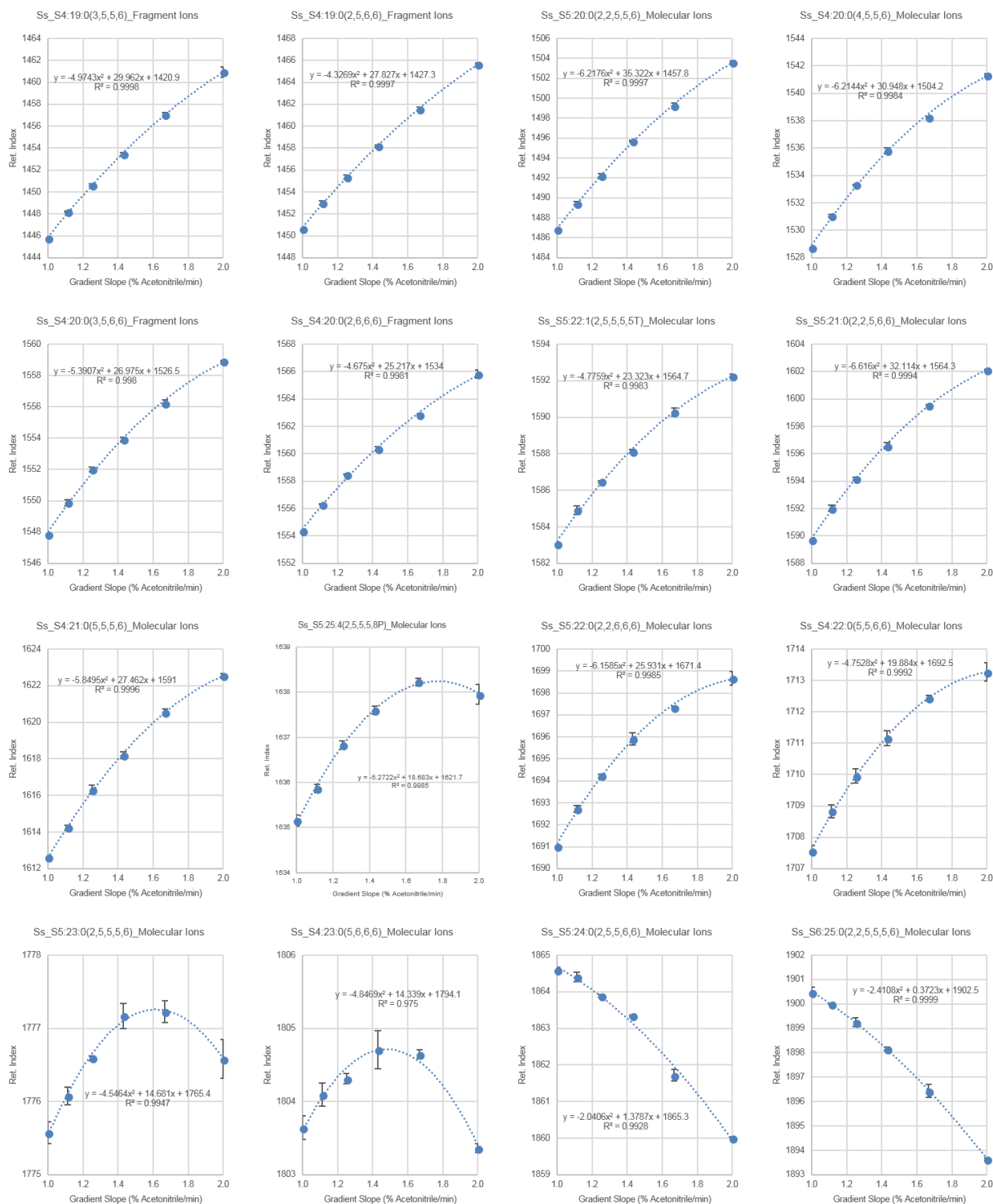
449
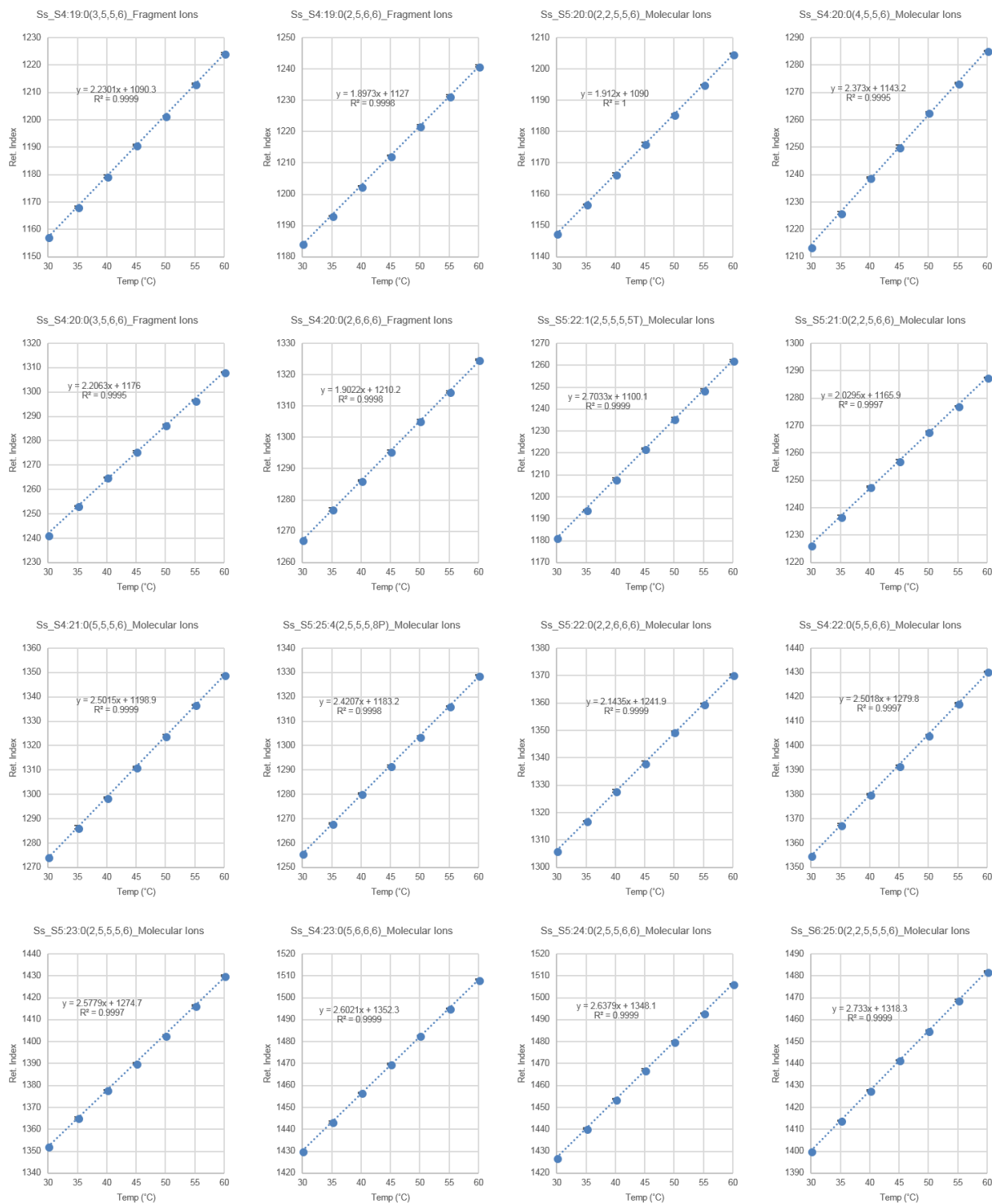
**Figure 4.26.** RI value column temperature dependence of 16 *S. sinuata* acylsucrose metabolites (methanol organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, gradient slope 1% acetontrile·min⁻¹, column SAE-A).
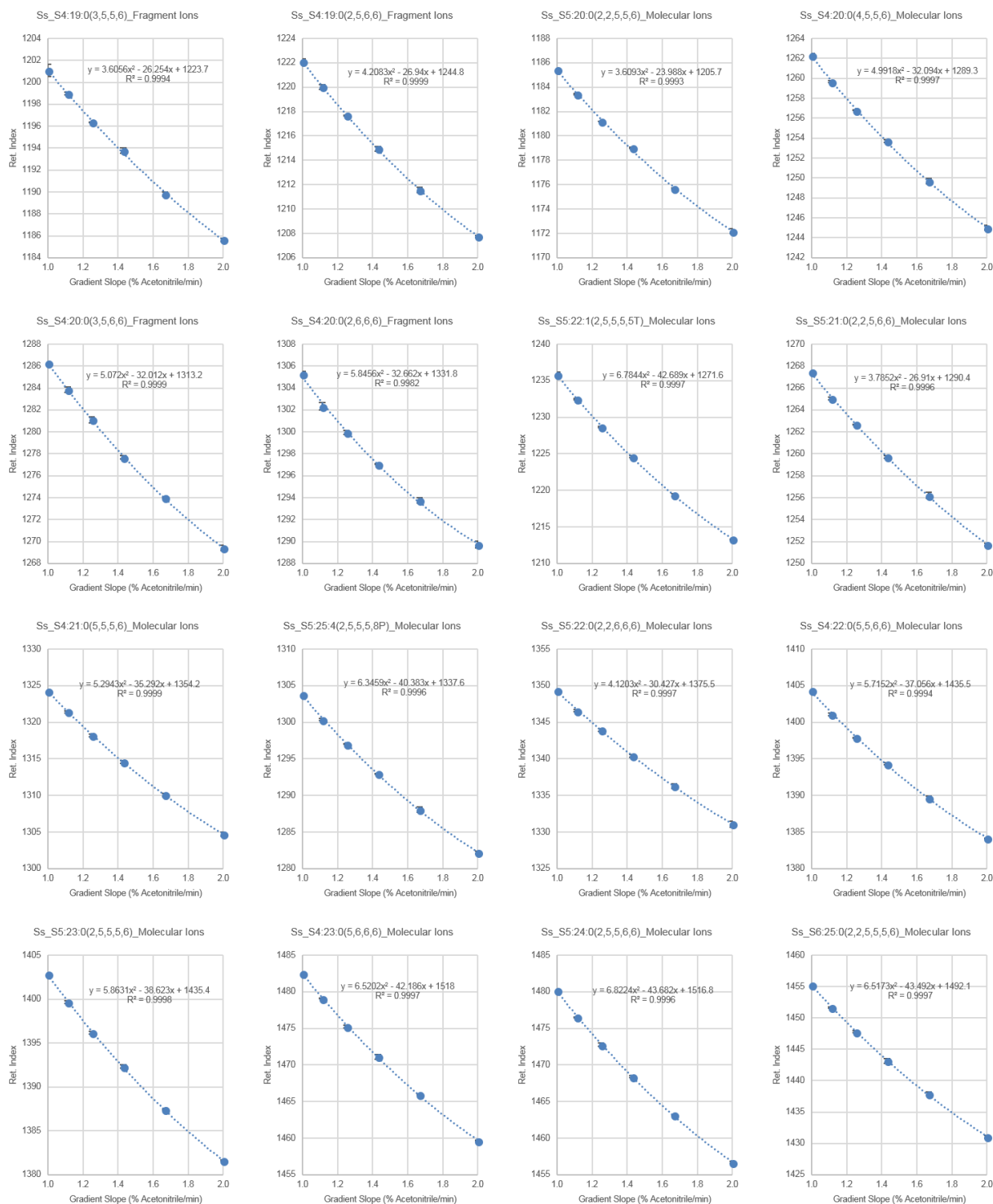
**Figure 4.27.** RI value gradient dependence of 16 *S. sinuata* acylsucrose metabolites (methanol organic mobile phase component, aqueous component 10 mM ammonium hydroxide adjusted to pH 2.8 with formic acid, column temperature 50°C, gradient slope 1, 10/9, 5/4, 10/7, 5/3, and 2% acetontrile·min$^{-1}$, column SAE-A).

451
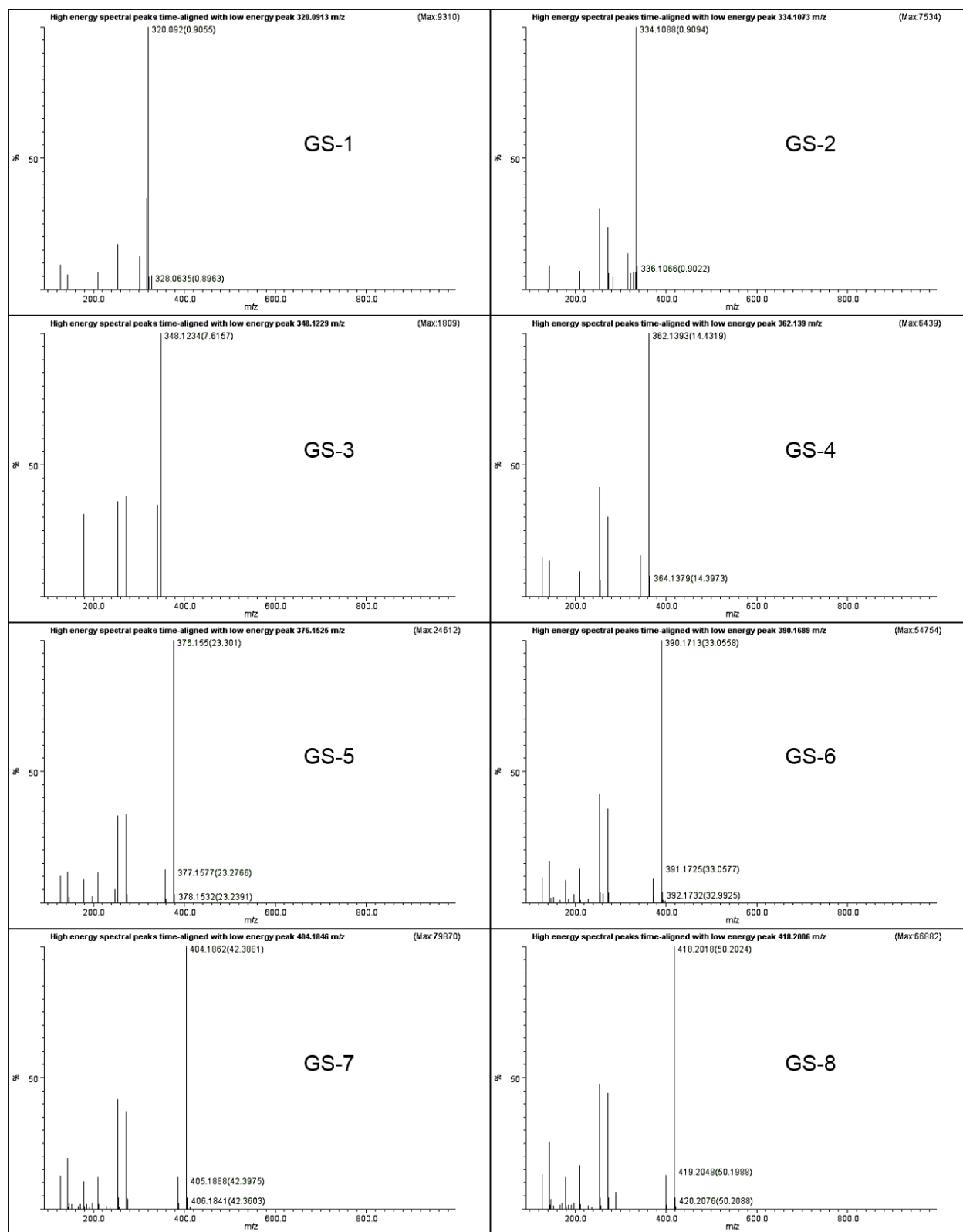
**Figure 4.28.** ESI negative mode MS/MS spectra of GS-*n*-alkyl standards [M-H]⁻ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MS$^E$ Data Viewer).
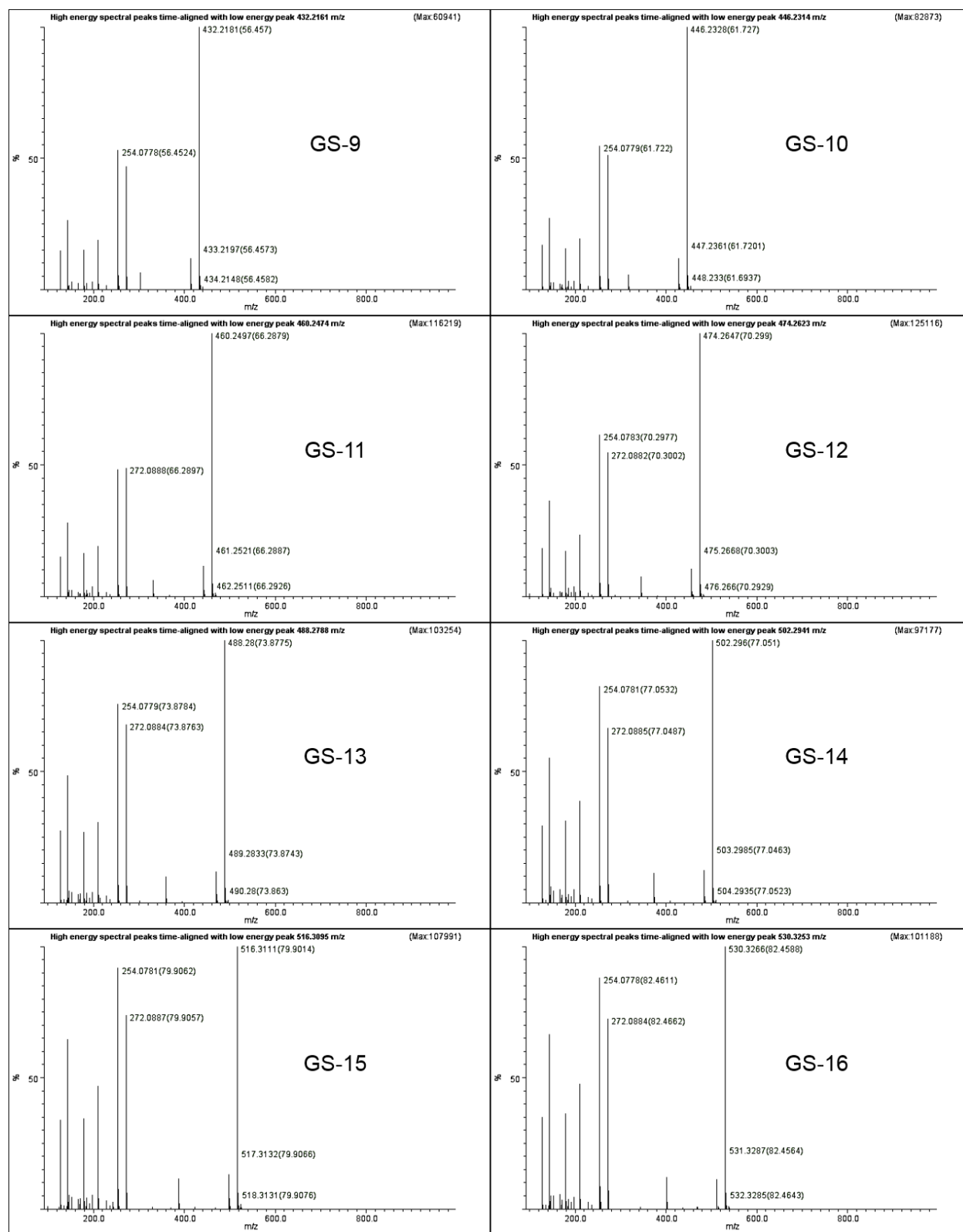
**Figure 4.29.** ESI negative mode MS/MS spectra of GS-*n*-alkyl standards [M-H]⁻ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MS$^E$ Data Viewer).

453
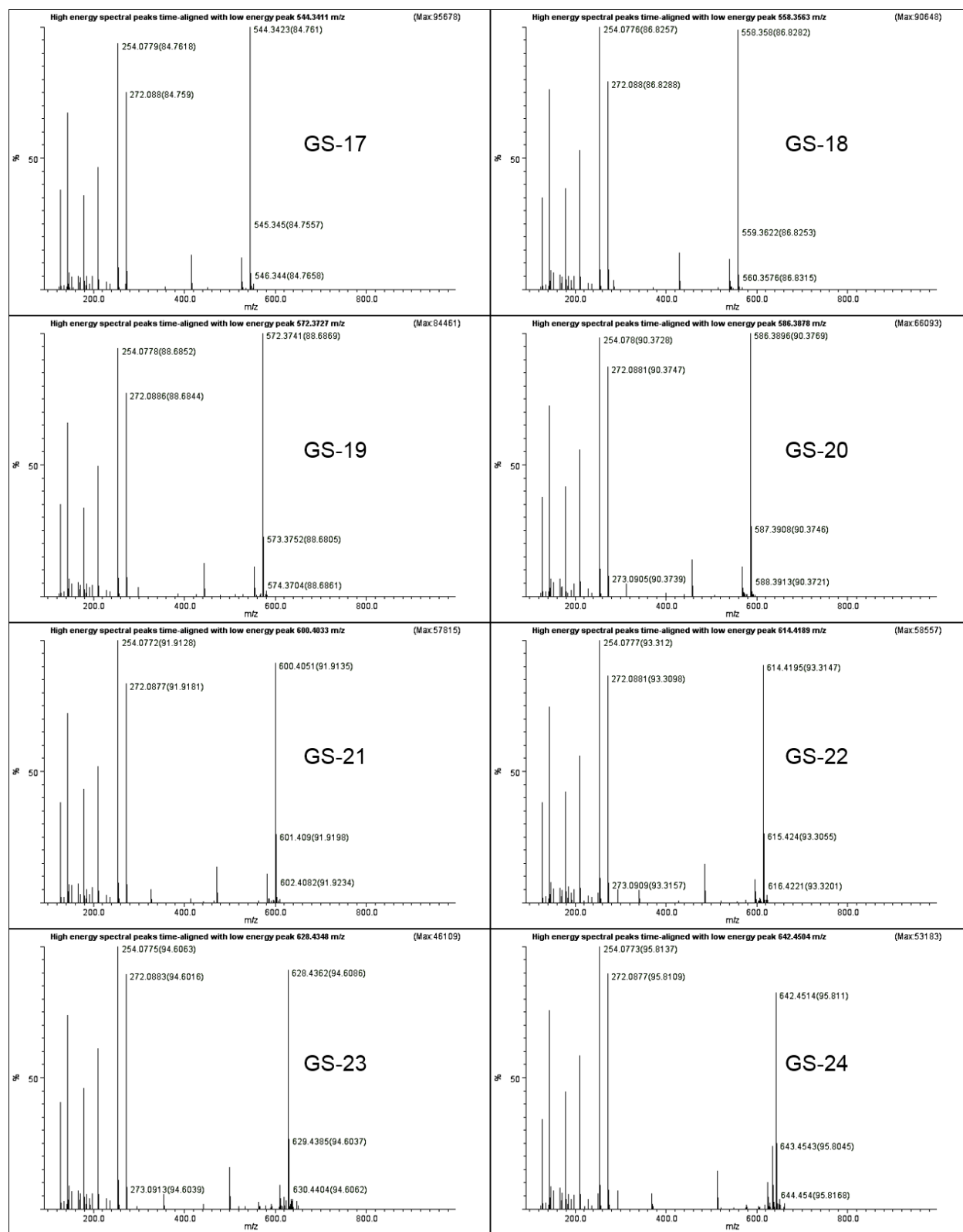
**Figure 4.30.** ESI negative mode MS/MS spectra of GS-*n*-alkyl standards [M-H]⁻ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MS^E Data Viewer).
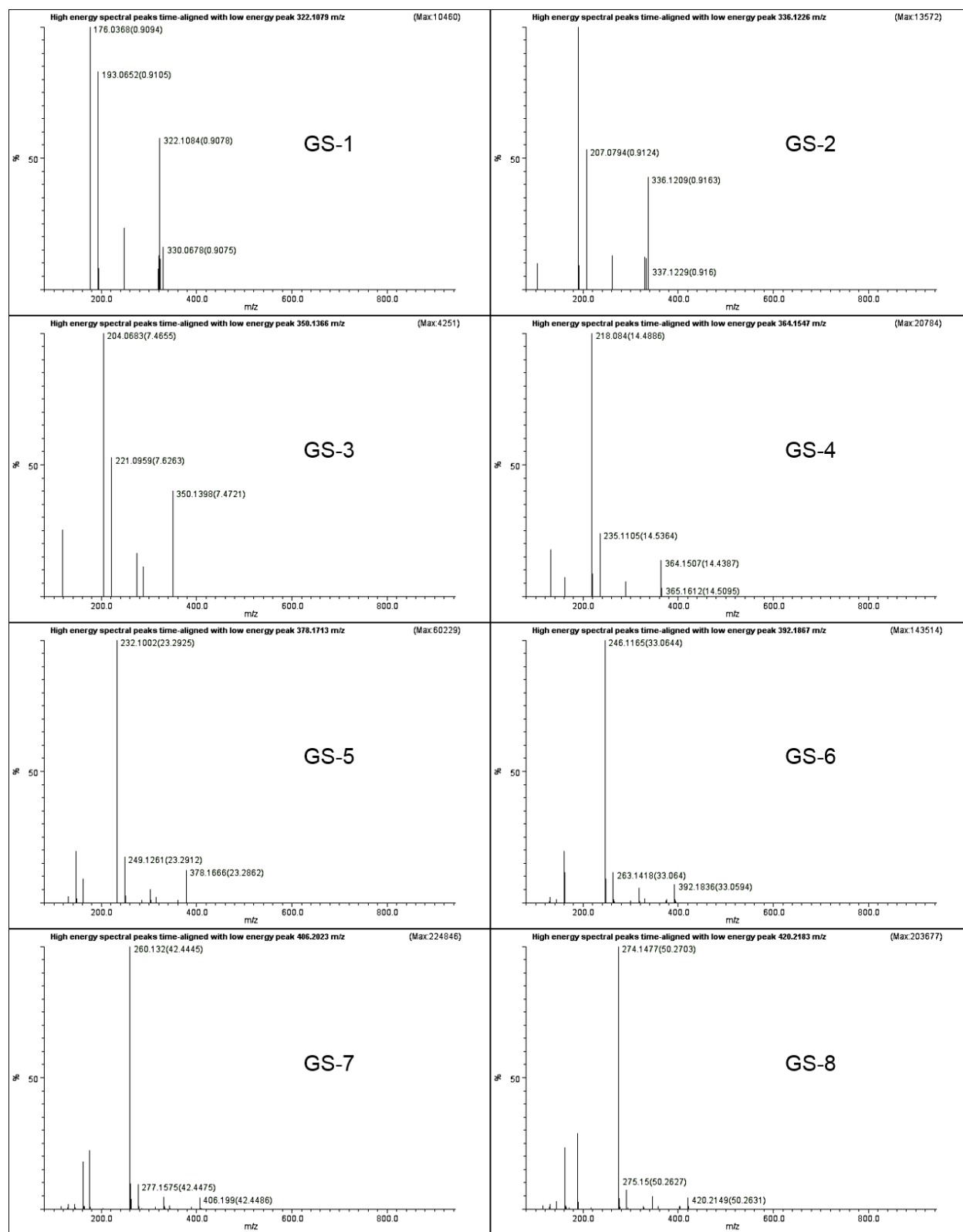
**Figure 4.31.** ESI positive mode MS/MS spectra of GS-*n*-alkyl standards [M+H]$^+$ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MS$^E$ Data Viewer).

455
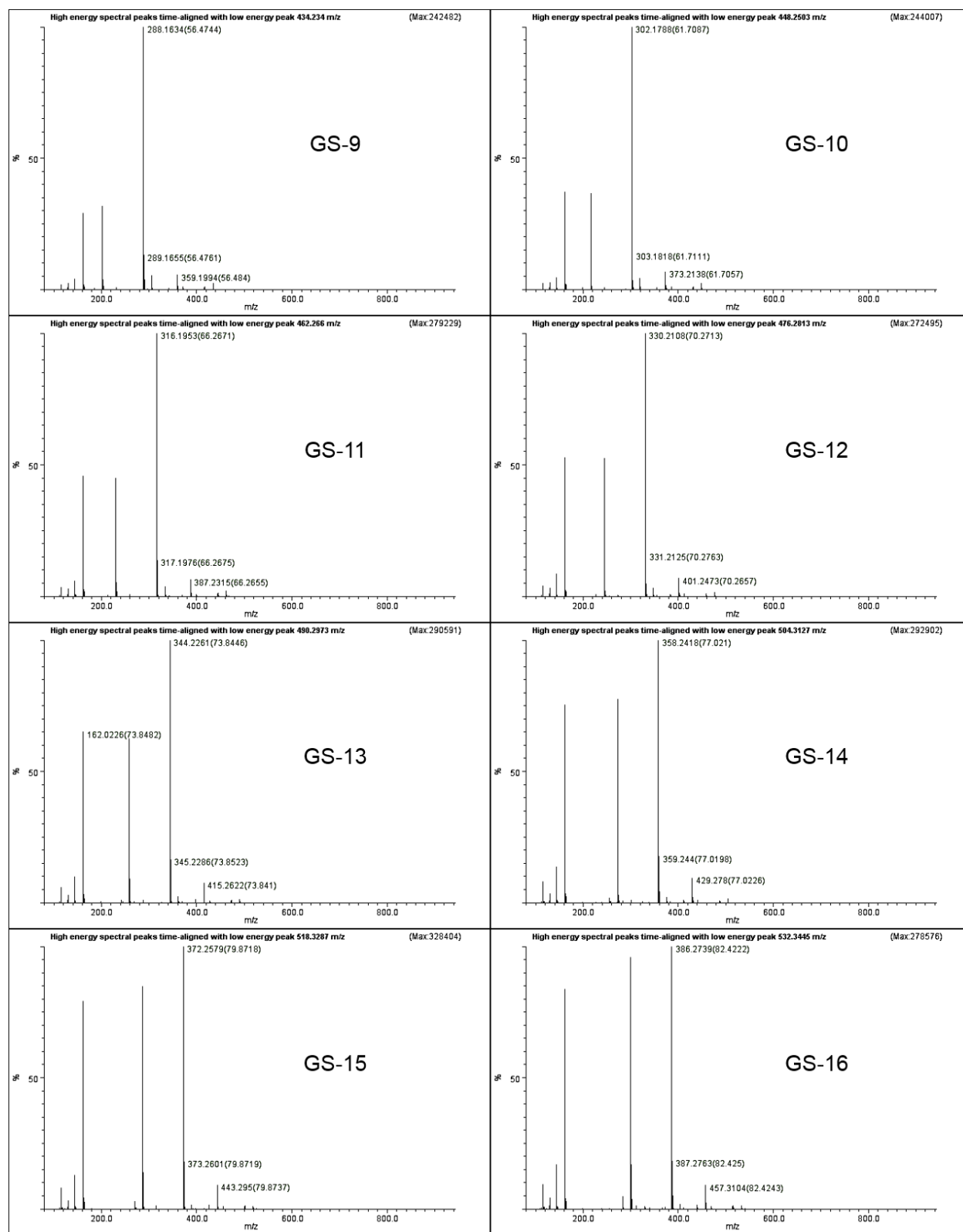
**Figure 4.32.** ESI positive mode MS/MS spectra of GS-*n*-alkyl standards [M+H]+ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MSE Data Viewer).
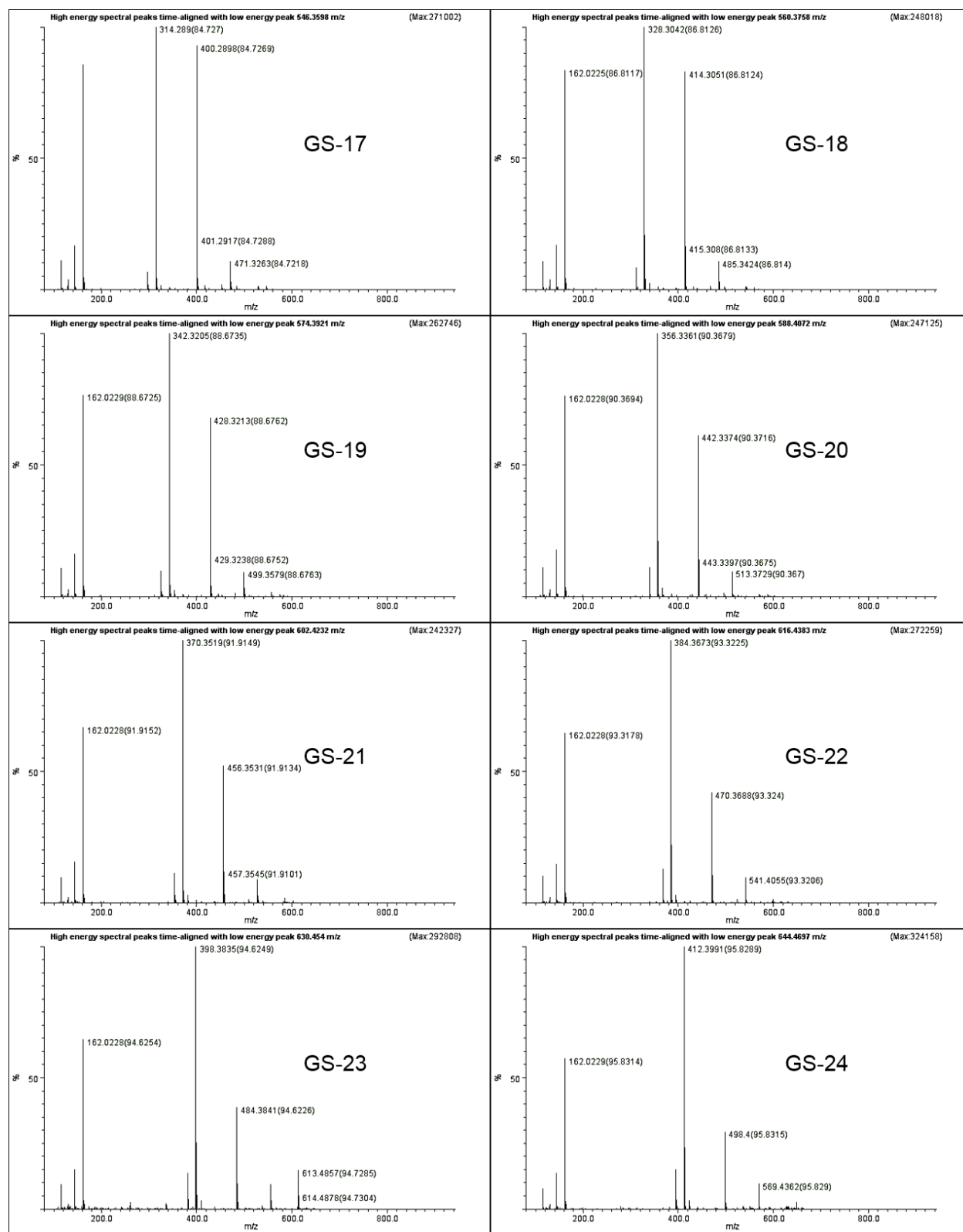
**Figure 4.33.** ESI positive mode MS/MS spectra of GS-*n*-alkyl standards [M+H]⁺ (generated using Waters SONAR data acquisition platform, 10-40 V CID ramp, 0.5 s acquisition time, 10 Da bin widths, spectra processed using Waters MSᴱ Data Viewer).

**Table 4.5.** MS/MS product ion analysis of GS-*n*-alkyl standards. Ratio of common fragment $[C_{10}H_{12}N_3O_5]^- = 254.078$ and $[M-H]^-$ ion intensities measured using 10-40 V CID ramp (0.5 s acquisition time).

| GS-*n*-alkyl Carbon # | $[M-H]^-$ Peak Abundance | $[C_{10}H_{12}N_3O_5]^- = 254.078$ Peak Abundance | $A_{[C_{10}H_{12}N_3O_5]^-} / A_{[M-H]^-}$ |
|---|---|---|---|
| 1 | 9310 | 1599 | 0.172 |
| 2 | 7534 | 2315 | 0.307 |
| 3 | 1809 | 650 | 0.359 |
| 4 | 6439 | 2660 | 0.413 |
| 5 | 24612 | 8149 | 0.331 |
| 6 | 54754 | 22600 | 0.413 |
| 7 | 79870 | 33296 | 0.417 |
| 8 | 66882 | 31786 | 0.475 |
| 9 | 60941 | 32356 | 0.531 |
| 10 | 82873 | 45326 | 0.547 |
| 11 | 116219 | 55879 | 0.481 |
| 12 | 125116 | 76766 | 0.614 |
| 13 | 103254 | 78101 | 0.756 |
| 14 | 97177 | 79983 | 0.823 |
| 15 | 107991 | 98996 | 0.917 |
| 16 | 101188 | 89185 | 0.881 |
| 17 | 95678 | 89548 | 0.936 |
| 18 | 89609 | 90648 | 1.012 |
| 19 | 84461 | 79556 | 0.942 |
| 20 | 66093 | 64989 | 0.983 |
| 21 | 52703 | 57815 | 1.097 |
| 22 | 52975 | 58557 | 1.105 |
| 23 | 41919 | 46109 | 1.100 |
| 24 | 43747 | 53183 | 1.216 |

**Table 4.6.** MS/MS product ion analysis of GS-*n*-alkyl standards. Ratio of common fragment $[C_5H_8NO_3S]^+ = 162.022$ and $[M+H]^+$ ion intensities measured using 10-40 V CID ramp (0.5 s acquisition time).

| GS-*n*-alkyl Carbon # | $[M+H]^+$ Peak Abundance | $[C_5H_8NO_3S]^+ = 162.022$ Peak Abundance | $A_{[C_5H_8NO_3S]^+} / A_{[M+H]^+}$ |
|---|---|---|---|
| 1 | 6021 | ND | --- |
| 2 | 5788 | ND | --- |
| 3 | 1114 | ND | --- |
| 4 | 2812 | 1464 | 0.521 |
| 5 | 7400 | 5434 | 0.734 |
| 6 | 9971 | 16624 | 1.667 |
| 7 | 9401 | 40365 | 4.294 |
| 8 | 8550 | 47700 | 5.579 |
| 9 | 5558 | 70572 | 12.70 |
| 10 | 5374 | 90319 | 16.81 |
| 11 | 5979 | 127259 | 21.28 |
| 12 | 4195 | 143271 | 34.15 |
| 13 | 3697 | 188873 | 51.09 |
| 14 | 4420 | 220424 | 49.87 |
| 15 | 3053 | 259937 | 85.14 |
| 16 | 3214 | 233103 | 72.53 |
| 17 | 2964 | 231604 | 78.14 |
| 18 | 1964 | 206838 | 105.3 |
| 19 | 1807 | 200967 | 111.2 |
| 20 | 1786 | 188049 | 105.3 |
| 21 | 1375 | 161634 | 117.6 |
| 22 | 1299 | 176077 | 135.5 |
| 23 | 1143 | 188927 | 165.3 |
| 24 | 922 | 185657 | 201.4 |
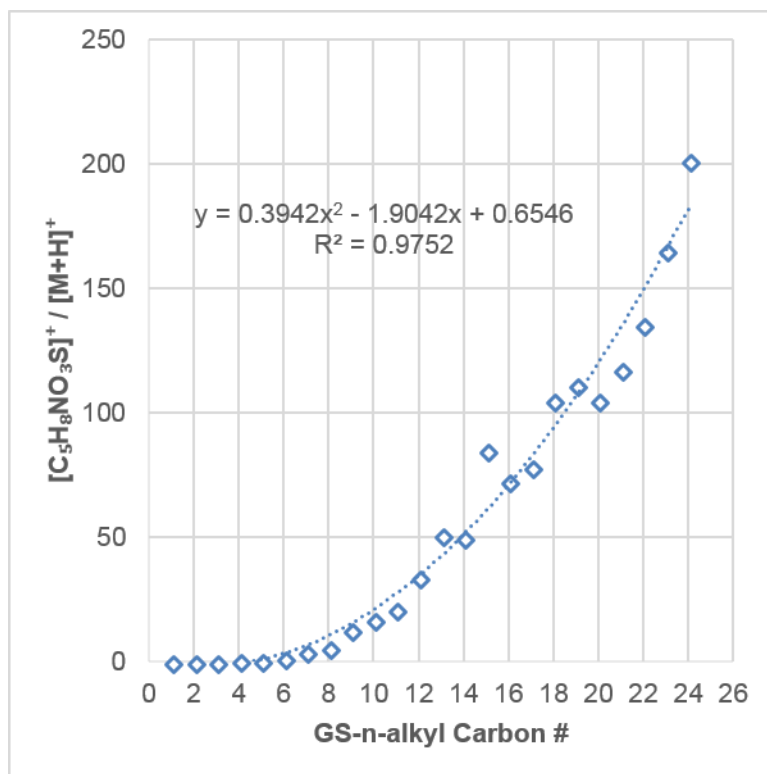
**Figure 4.34.** MS/MS product ion analysis of GS-*n*-alkyl standards. Ratio of common fragment $[C_5H_8NO_3S]^+ = 162.022$ and $[M+H]^+$ ion intensities as a function of GS-n-alkyl chain length (measured using 10-40 V CID ramp, 0.5 s acquisition time)

**REFERENCES**

# REFERENCES

1.   Kind, T. and O. Fiehn, *Advances in structure elucidation of small molecules using mass spectrometry.* Bioanal Rev, 2010. **2**(1-4): p. 23-60.

2.   Forcisi, S., et al., *Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling.* J Chromatogr A, 2013. **1292**: p. 51-65.

3.   Xian, F., C.L. Hendrickson, and A.G. Marshall, *High resolution mass spectrometry.* Anal Chem, 2012. **84**(2): p. 708-19.

4.   Last, R.L., A.D. Jones, and Y. Shachar-Hill, *Towards the plant metabolome and beyond.* Nat Rev Mol Cell Biol, 2007. **8**(2): p. 167-74.

5.   Nguyen, D.D., et al., *MS/MS networking guided analysis of molecule and gene cluster families.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. E2611-20.

6.   Johnson, S.R. and B.M. Lange, *Open-access metabolomics databases for natural product research: present capabilities and future potential.* Front Bioeng Biotechnol, 2015. **3**: p. 22.

7.   Vinaixa, M., et al., *Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects.* Trac-Trends in Analytical Chemistry, 2016. **78**: p. 23-35.

8.   Kováts, E., *Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone.* Helvetica Chimica Acta, 1958. **41**(7): p. 1915-1932.

9.   Baker, J.K. and C.Y. Ma, *Retention index scale for liquid-liquid chromatography.* Journal of Chromatography, 1979. **169**(Feb): p. 107-115.

10.  Smith, R.M., *Alkylarylketones as a retention index scale in liquid-chromatography.* Journal of Chromatography, 1982. **236**(2): p. 313-320.

11.  Bogusz, M. and R. Aderjan, *Improved standardization in seversed-phase high-performance liquid-chromatography using 1-nitroalkanes as a retention index scale.* Journal of Chromatography, 1988. **435**(1): p. 43-53.

12.  Magee, M.H., et al., *"Measure Your Gradient": a new way to measure gradients in high performance liquid chromatography by mass spectrometric or absorbance detection.* J Chromatogr A, 2014. **1369**: p. 73-82.

13.  Abate-Pella, D., et al., *Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods.* J Chromatogr A, 2015. **1412**(Supplement C): p. 43-51.

14.  Quilliam, M., et al., *A new retention index system for liquid chromatography- mass spectrometry*, in *42nd International Symposium on High Performance Liquid Phase Separations and Related Techniques (HPLC2015).* 2015: Geneva, Switzerland.

15.     Quilliam, M., *Retention index standards for liquid chromatography*, U.S. Patent, Editor. 2017, National Research Council of Canada.

16.     Ghosh, B. and A.D. Jones, *Profiling, characterization, and analysis of natural and synthetic acylsugars (sugar esters).* Analytical Methods, 2017. **9**(6): p. 892-905.

17.     Treutler, H., et al., *Discovering regulated metabolite families in untargeted metabolomics studies.* Anal Chem, 2016. **88**(16): p. 8082-90.

18.     Ghosh, B., T.C. Westbrook, and A.D. Jones, *Comparative structural profiling of trichome specialized metabolites in tomato (Solanum lycopersicum) and S. habrochaites: acylsugar profiles revealed by UHPLC/MS and NMR.* Metabolomics, 2014. **10**(3): p. 496-507.

19.     Liu, X.X., et al., *Profiling, isolation and structure elucidation of specialized acylsucrose metabolites accumulating in trichomes of Petunia species.* Metabolomics, 2017. **13**(7): p. 85.

20.     Poole, C.F., *Chromatographic test methods for characterizing alkylsiloxane-bonded silica columns for reversed-phase liquid chromatography.* J Chromatogr B Analyt Technol Biomed Life Sci, 2018. **1092**: p. 207-219.

21.     Nawrocki, J., *The silanol group and its role in liquid chromatography.* Journal of Chromatography A, 1997. **779**(1-2): p. 29-71.

22.     Dolan, J.W., *Detective work part 1: simplify the choices.* LCGC Europe, 2015. **28**(11): p. 612-615.

23.     Snyder, L.R., J.W. Dolan, and P.W. Carr, *The hydrophobic-subtraction model of reversed-phase column selectivity.* J Chromatogr A, 2004. **1060**(1-2): p. 77-116.

24.     Dolan, J.W., *Detective work, part 4: chemical problems with the column: chemical attack.* LCGC Europe, 2016. **29**(2): p. 82-84.

25.     Vince, R., S. Daluge, and W.B. Wadd, *Studies on the inhibition of glyoxalase I by S-substituted glutathiones.* J Med Chem, 1971. **14**(5): p. 402-4.

26.     Kostiainen, R. and T.J. Kauppila, *Effect of eluent on the ionization process in liquid chromatography-mass spectrometry.* J Chromatogr A, 2009. **1216**(4): p. 685-99.

27.     Kirkland, J.J., M.A. van Straten, and H.A. Claessens, *High pH mobile-phase eEffects on silica-based reversed-phase high-performance liquid-chromatographic columns.* Journal of Chromatography A, 1995. **691**(1-2): p. 3-19.

28.     Claessens, H.A. and M.A. van Straten, *Review on the chemical and thermal stability of stationary phases for reversed-phase liquid chromatography.* J Chromatogr A, 2004. **1060**(1-2): p. 23-41.

29.     Li, J., *Effect of temperature on selectivity in reversed-phase liquid chromatography, a thermodynamic analysis.* Analytica Chimica Acta, 1998. **369**(1): p. 21-37.

30.     Dolan, J.W., *Temperature selectivity in reversed-phase high performance liquid chromatography.* J Chromatogr A, 2002. **965**(1-2): p. 195-205.

31.     Shoenmakers, P.J., H.A.H. Billiet, and L. De Galan, *Influence of organic modifiers on the rentention behaviour in reversed-phase liquid chromatography and its consequences for gradient elution.* Journal of Chromatography A, 1979. **185**: p. 179-195.

32.     Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry.* J Mass Spectrom, 2004. **39**(10): p. 1091-112.

33.     Bristow, A.W., et al., *Reproducible product-ion tandem mass spectra on various liquid chromatography/mass spectrometry instruments for the development of spectral libraries.* Rapid Commun Mass Spectrom, 2004. **18**(13): p. 1447-54.

34.     Hopley, C., et al., *Towards a universal product ion mass spectral library - reproducibility of product ion spectra across eleven different mass spectrometers.* Rapid Commun Mass Spectrom, 2008. **22**(12): p. 1779-86.

35.     Paglia, G., et al., *Ion mobility derived collision cross sections to support metabolomics applications.* Anal Chem, 2014. **86**(8): p. 3985-93.

464

**Chapter 5: Closing Thoughts**

Plants produce a truly enormous array of structurally complex and varied metabolites, and annotation and discovery of their chemical structures will continue to present formidable challenges. Comprehensive metabolite profiling requires multiple chromatographic and mass spectrometric platforms to separate and detect their chemical complexity, while the important relationship between mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy will continue to be central to *de novo* structural elucidation of metabolites.

The analytical challenges of metabolomics are well-demonstrated in the case of acylsugars. In Chapter 2, more than 400 acylsucroses were manually annotated from LC/MS of *Salpiglossis sinuata*, while slightly more than ~ 300 peaks were detected using automated processing by collaborator Dr. Gaurav Moghe. A point of significance is that these numbers are from a single plant genotype, at a specific stage of development, in a single set of environmental conditions, and from a single tissue type (trichomes). These results suggest that the acylsugar metabolome is far more complex than imagined. Even by manual annotation, no matter what we do it seems we are still underestimating the range of acylsugars that can be produced by a single plant. For instance, acylsucrose mono- and di-esters must be made to form tri-, tetra-, penta-esters, and so on. However, the early pathway precursors were not detected and are likely at very low concentrations, but must be intermediates in the formation of more heavily acylated forms. Furthermore, column performance comparisons in Chapter 4 showed that reversed-phase C18 columns from different manufacturers differ in acylsucrose retention selectivity, revealing coeluting isomers that we would never know existed (four isomers resolved on one column, and seven isomers for the same extract analyzed on another column).

Plant metabolic complexity is even more multifaceted when we consider that individual acyltransferase enzymes can use several different acyl-CoAs as substrates, and that mutations can alter their sugar acyl acceptor substrates. To our knowledge, there are 25 documented acylsugar ester groups (Chapter 1), including two new ester groups (tiglyl and phenylacetyl) outlined in Chapter 2. With eight positions

available on sucrose for acylation, there are billions of possible combinations for the compound class of acylsucroses alone. Many of these metabolites are expected to exist as isomeric forms and/or may be of minor abundance. Furthermore, plants have capacity to make far more acyl-CoA groups, so it would not be surprising to find new acylsugar chemistries as new plants are investigated. For instance, where previous reports have mostly identified acylglucoses and acylsucroses from the Solanaceae, Chapter 3 showed a group of acylated *myo*-inositols and *myo*-inositol glycosides discovered in *Solanum quitoense*, revealing the carbohydrate core can also vary. There are certainly many other carbohydrates that exist in nature that could be acylated too. Moreover, those sugar groups also have potential to exist as isomers, differing by stereochemical configuration and linkage positions. Therefore, the challenges to characterize acylsugars remain to be enormous. In this regard, it is not feasible to purify and perform NMR on all metabolites, and no combination of current analytical technologies could be used to resolve them all. Therefore, the chemical complexity of nature drives a need for continued improvements in analytical technologies, particularly for those that can be used to resolve and identify isomeric metabolites. The further advancement of pre-fractionation, two-dimensional LC and ion mobility mass spectrometry separations offers opportunities to improve analytical resolving capacity to advance future discovery of isomeric metabolites.

Though application of GS-*n*-alkyl retention indexing outlined in Chapter 4 does not improve instrumental resolution of metabolites, it improves confidence in annotation, identification and dereplication of discovery of specialized metabolites and offers a second level of information about metabolites, particularly for those isomeric metabolites that are not readily differentiated by their mass spectra but exhibit differential changes in retention behavior (e.g. retention index temperature-dependence). Liquid chromatography retention indexing has yet to find widespread use, and for retention indexing to become more commonly adopted, researchers must incorporate approaches that embrace the idea that more experimental variables influence liquid chromatographic retention than is the case for gas chromatography. The development of a computational framework that models the effects of experimental parameters on retention index values of individual compounds and/or classes has potential to improve recognition of these effects and enhance

466

confident annotations of unknowns. As a follow-up to this work, we hope to make the GS-*n*-alkyl standards available to the community to encourage more metabolomics researchers to adopt retention indexing systems as additional parameters in open-access metabolite databases.