UNIVERSITY OF CALIFORNIA

SANTA CRUZ

USE OF SYNTHETIC LIBRARIES TO SURVEY PERMEABILITY IN ATYPICAL CYCLIC PEPTIDE CHEMICAL SPACE

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Colin N. Kelly

December 2021

The Dissertation of Colin N Kelly
is approved:

_____
Professor R. Scott Lokey, Advisor

_____
Professor Seth Rubin, Chair

_____
Professor Michael Stone

_____
Peter Biehl

Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

List of Schemes

Abstract

Title: Use of synthetic libraries to survey permeability in atypical cyclic peptide chemical space

Author: Colin N Kelly

The passive membrane permeability of cyclic peptides continues to astonish and inspire a growing number of researchers. While the literature has become increasingly dense and the methods of studying cyclic peptides have become more powerful, the problem of turning these molecules into drugs remains. Although cyclic peptides thoroughly violate the previously embraced rules of drug discovery, synthetic investigation leading to de novo discovery in the chemical space beyond traditional small molecule drugs has concentrated on head-to-tail cyclic peptides. Many bioactive peptide natural products with intracellular targets are not head-to-tail cyclized, but rather side chain to tail cyclized resulting in cyclic/linear hybrid peptides known as lariat peptides. We pursued two strategies to study structure-permeability relationships in lariat peptides. Chapter 1 describes the use of synthetic libraries to explore a generic lariat peptide chemical space. We generated a library of scaffolds using stable isotopes to encode stereochemistry and determined the passive membrane permeability of over 1000 novel lariat peptide scaffolds with molecular weights around 1000. Many lariats were surprisingly permeable, comparable to many known orally bioavailable drugs. Passive permeability was strongly dependent on N-methylation, stereochemistry, and ring topology. A variety of structure-permeability

trends were observed including a relationship between alternating stereochemistry and high permeability, as well as a set of highly permeable consensus sequences. Chapter 2 describes our investigation of xentrivalpeptides, a class of lariat peptide natural products. These lariats are composed entirely of simple lipophilic amino acids, indicating potential passive permeability. These compounds are heavily enriched in valine, suggesting a possible role of β-branching. We established high permeability for xentrivalpeptide A and prepared several analogues of xentrivalpeptide A to probe the roles of β-branching and stereochemistry. Our results reveal that β-branching is not important for permeability while stereochemistry does influence permeability. Obtaining low-dielectric NMR data proved difficult for these lariats and, having established that lariat peptides can achieve passive permeability, we directed our efforts towards the library approach discussed in chapter 1.

Acknowledgements

The text of this dissertation includes portions of the following previously published material:

Kelly, Colin N., Chad E. Townsend, Ajay N. Jain, Matthew R. Naylor, Cameron R. Pye, Joshua Schwochert, and R. Scott Lokey. “Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability.” *Journal of the American Chemical Society* 143, no. 2 (January 20, 2021): 705–14.

Professor R. Scott Lokey directed and supervised the research described in this dissertation.

I would like to thank the following organizations for financial support:

University of California, UCSC Chemistry Department, US National Institutes of Health

I would like to thank my advisor, Scott Lokey, for his unwavering patience and encouragement, and my other committee members, Seth Rubin and Michael Stone, for their support.

I would like to thank my collaborators and coauthors:

Chad Townsend, Matthew Naylor, Cameron Pye, Joshua Schwochert, Ajay Jain, Akihiro Furukawa

I would like to thank the following individuals for logistical and research support:

Qiangli Zhang, Jack Lee, Walter Bray, Karen Meece, Katie Cramton

I would like to thank past and present members of the Lokey lab:

Scott Lokey, Matt Naylor, Joshua Schwochert, Chad Townsend, Tannia Lau, Victoria Klein, Grant Koch, Akihiro Furukawa, Akshar Lohith, Justin Faris, Kevin Yang, Sean Copley, Panpan Zhang, Jaru Taechalertpaisarn, Rachel Snelling, Christin Cudahy, Nadeeya Amjadi, Sarah Beth Avila, Andrew Ly, William Rosenberg, Alejandro Cabrera-Cortez, Ryan Noland, Yongtonxian Lao, Quinn Edmonson, Jessica Ochoa, Daniel Ramos, Savannah Conlon, Alexandra Turmon, Eva Jason, Mason Handford

Abbreviations

Chapter 1: PPI, protein-protein interaction; Ro5, Lipinski's Rule of 5; MW, molecular weight; AlogP, atomistic logP; bRo5, beyond Rule of 5; DNA, deoxyribonucleic acid; DnaN, Beta sliding clamp; K_D, dissociation constant; eEF-1A, eukaryotic elongation factor 1A; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; IMHB, intramolecular hydrogen bond; HBD, hydrogen bond donor; MS, mass spectrometry; Ala, alanine; Pro, proline; MS^2, tandem mass spectrometry; LC, liquid chromatography; Fmoc, fluorenylmethoxycarbonyl; Thr, threonine; Leu, leucine; HATU, 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate; PAMPA, parallel artificial membrane permeability assay; Me, methyl; 3D, 3-dimensional; LCMS, liquid chromatography-mass spectrometry; NMR, nuclear magnetic resonance; MDCK, Madin-Darby canine kidney; P_{app}, apparent permeability; ROESY, Rotating Frame Overhauser Enhancement spectroscopy; ppb/K, parts per billion / Kelvin; HeLa cells, Henrietta Lacks immortalized human cancer cell line; CP, cytological profiling; anti-pH3-Ab, Phospho-Histone H3 antibody; EdU, 5-Ethynyl-2'-deoxyuridine; HPLC, high-performance liquid chromatography; DCM, dichloromethane; SPE, solid-phase extraction; DIPEA, diisopropylethylamine; MeOH, methanol; DMF, N,N-dimethylformamide; DBU, 1,8-diazabicycloundec-7-ene; COMU, (1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate; DIC, diisopropylcarbodiimide; DMAP, 4-dimethylaminopyridine; HFIP, 1,1,1,3,3,3-hexafluoroisopropanol; THF, tetrahydrofuran; ACN, acetonitrile; TFA, trifluoroacetic acid; UPLC, ultra performance liquid chromatography; DMSO, dimethylsulfoxide; PBS, phosphate buffer solution; H_2O, water; TPGS-750M, DL-α-Tocopherol methoxypolyethylene glycol succinate; BSA, bovine serum albumin; TMR, tetramethylrhodamine; MTs, microtubule stain; $CDCl_3$, chloroform-d; Abu, 2-aminobutyric acid; Nva, norvaline; Phe, phenylalanine; FPhe, 4-fluorophenylalanine; Cha, cyclohexylalanine; F2Phe, 3,4-difluorophenylalanine; 1-Nal, 1-naphthylalanine; hPhe, alpha-homophenylalanine; MeOPhe, 4-methoxyphenylalanine; NPhe, 4-nitrophenylalanine; Avg, average (mean); Rt, retention time; UV, ultra-violet light; m/z, mass / charge ratio; HZ, hertz.

Chapter 2: XvA, xentrivalpeptide A; AlogP, atomistic prediction of octanol/water partition coefficient; 3D, 3-dimensional; Thr, threonine; Val, valine; Fmoc, fluorenylmethoxycarbonyl; HATU, 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate; TSTU, N,N,N′,N′-

Tetramethyl-O-(N-succinimidyl)uronium tetrafluoroborate; XvI, xentrivalpeptide I; Nva, norvaline; Phe, phenylalanine; PAMPA, parallel artificial membrane permeability assay; RRCK, Ralph Russ canine kidney; PVDF, polyvinylidene fluoride; MDCK, Madin-Darby canine kidney; P_{app}, apparent permeability; tBuGly, α-tertbutylglycine; NMR, nuclear magnetic resonance; $CDCl_3$, chloroform-d; HMBC, heteronuclear multiple bond correlation; TOCSY, total correlation spectroscopy; H-bond, hydrogen bond; DMF, dimethylformamide; DCM, dichloromethane; DIPEA, diisopropylethylamine; HBTU, N,N,N′,N′-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene; DMAP, dimethylaminopyridine; DIC, diisopropylcarbodiimide; HFIP, 1,1,1,3,3,3-hexafluoroisopropanol; COMU, (1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate; ACN, acetonitrile; TFA, trifluoroacetic acid; HPLC, high-performance liquid chromatography; DMSO, dimethylsulfoxide; PBS, phosphate buffer solution; LC-MS, liquid chromatography-mass spectrometry.

Chapter 1: Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability

1.1. Introduction

Targeting protein-protein interactions (PPIs) with small molecules remains challenging. Typical drug-like small molecules are too small to bind large protein interfaces, while larger molecules suffer from poor membrane permeability and thus have limited access to intracellular targets. Increasing attention is being devoted to the chemical space beyond what is traditionally considered 'drug-like' based on Lipinski's Rule of 5 (Ro5), which predicts poor absorption and permeation for molecules with more than five H-bond donors, ten H-bond acceptors, molecular weight (MW) exceeding 500, and AlogP greater than 5.[1] This interest in larger molecules is further fueled by the existence of known membrane-permeable molecules that grossly violate these rules. Macrocycles are the most promising structural class for targeting intracellular PPIs. Macrocycles are well-represented among orally bioavailable beyond Rule of 5 (bRo5) drugs.[2-4] An analysis of bRo5 drugs and clinical candidates has revealed that high-MW macrocyclic molecules are more likely to bind flat binding sites than non-macrocyclic bRo5 molecules.[5] Such binding sites, which are typical of PPIs, are currently among the most difficult to target with small molecules.[6]

A large portion of the macrocycles in current clinical use and clinical trials are cyclic peptides. However, almost all these cyclic peptides are administered parenterally,

being rendered impermeable by polar side chains.[4] Oral bioavailability is generally seen as a key limitation of peptide drugs. However, the striking permeability of cyclosporin A (Figure 1-1), a cyclic undecapeptide with MW of 1203, clearly demonstrates the potential of cyclic peptides to target intracellular PPIs. Evidence that cyclosporin A exhibits solvent-dependent conformation and extensive intramolecular hydrogen-bonding in low-dielectric environment gave rise to the theory that these features may explain its passive membrane permeability.[7] Research by Rezai et al. has established that the tendency of some cyclic peptides to exhibit conformational flexibility, adopting low-dielectric conformations in which the peptides exhibit preference for the membrane environment and high-dielectric conformations in which the peptides prefer the aqueous environment, explains the permeability of some cyclic peptides.[8] The extent to which the amide N-H groups are shielded from solvent in the low-dielectric conformation is a determining factor in membrane permeability. Since the publication of this study in 2006, numerous examples of membrane permeable cyclic peptides have been reported.[9] The role of N-methylation is well-established.[10] Cyclic peptides containing non-peptidic structural elements have been studied, including statines,[11] heterocycles,[12-13] and peptoids.[14] In many cases, these investigations were inspired by naturally occurring cyclic peptides containing these features. In addition to these natural product-inspired investigations, cyclic peptide model systems consisting of only α-amino acids have been studied.[8, 15-17] Cyclic peptides with MW over 1000 have been

designed and optimized for increased permeability based on principles and structural motifs established in cyclic peptide model systems.[17-18]

Unlike head-to-tail cyclized peptides, lariat peptides, in which cyclization occurs between a terminus and a side chain resulting in a partially linear peptide, have not been studied with respect to permeability. However, a survey of the Natural Products Atlas[19] revealed that roughly 30% of cyclic peptide natural products are cyclized between a terminal residue and a side chain (Figure 1-2). The majority of these "lariat" natural products are cyclized from the C-terminal carboxylic acid onto a side-chain hydroxyl group to form an ester (i.e., depsipeptide) linkage. Lariat depsipeptide natural products have a variety of ring sizes and tail lengths, and many contain non-peptidic polyketide elements in their backbones. Nonpolar members of this class are known to inhibit a variety of intracellular molecular targets. Griselimycin (Figure 1-3), for example, is a lariat depsipeptide with an 8-residue macrocycle and 2-residue tail which potently inhibits the polymerase sliding clamp DnaN of *Mycobacterium tuberculosis* (K_D: 1.0×10^{-10}).[20] The crystal structure of griselimycin bound to DnaN shows the cyclic portion of the molecule bound in one subsite with the lariat tail extended to engage a separate subsite.[20] An optimized analogue of grislemycin, cyclohexylgriselimycin, exhibited an oral bioavailabity in mice of 89% (vs. 48% for griselimycin) and in vivo efficacy against tuberculosis.[20] Didemnin B (Figure 1-3) is a potent inhibitor of the eukaryotic translation elongation factor eEF-1A,[21] with cytotoxic, immunosuppressive, antiviral, and antihepatotoxic effects.[22-23] Although didemnin B was poorly tolerated in clinical trials, an analogue of didemnin B,

plitidepsin (Aplidin), has been approved for multiple myeloma in Australia[24] and was evaluated in clinical trials against the novel coronavirus SARS-CoV-2.[25] Both griselimyicin and didemnin B have molecular weights above 1000, and although 3D structural data exist on both griselimycin and didemnin B,[21, 26] the relationship between structure and membrane permeability has not been established for either compound. As the effects of both natural products are mediated by intracellular targets, and both compounds are lipophilic and N-methylated, we considered these compounds as a structural basis for the further exploration of permeability in the lariat peptide space.

The conformational hypothesis of passive permeability posits that the ability of bRo5 molecules to adopt conformations in which polar functional groups are shielded from solvent is necessary for passive permeability. The major question to be answered with regards the lariats is how the linear tail of the molecule will interact with the cyclic body of the molecule, whether the linear portion is more likely to behave as an asset or a liability with respect to permeability. The greater flexibility of the linear segment could conceivably increase the probability of favorable IMHBs (intramolecular hydrogen bonds) with backbone HBDs (hydrogen bond donors). However, the flexibility of the tail could produce an entropic penalty to IMHB. A crystal structure of didemnin B reveals IMHB both within the lariat tail and between the lariat tail and cyclic backbone.[27] Meanwhile, a low-dielectric solution structure of methylgriselimycin shows that the lariat tail does not engage the cyclic core of the molecule.[26] The influence of the lariat tail on solubility is another consideration, as

the solubility of lipophilic compounds may be an impediment to their development as drugs, and the inclusion of a lariat tail may affect solubility by increasing the degrees of freedom.

Given the current lack of knowledge concerning the permeability of lariat peptides, we sought to determine the landscape of membrane permeability in structurally diverse lariat depsipeptides inspired by natural products. In a previous study, our lab used a split-pool library approach to explore structure-permeability relationships in cyclic hexapeptides.[28-29] Stereochemistry and N-methylation were varied in this study to sample a diverse conformational space. By identifying individual scaffolds differing only in relative stereochemistry with dramatic differences in permeability, this work confirmed the impact of stereochemistry on permeability through its effect on conformation. Identification of individual cyclic peptides was accomplished by resynthesis, limiting data recovery. In a recent investigation in the Lokey lab, Chad Townsend developed a cyclic peptide sequencing program, referred to as CycLS, that determines the sequence based on fragments generated during mass spectroscopy and applied it to the analysis of a cyclic peptide library.[30] CycLS allows for the generation of large data sets for the development of structure-permeability relationships. Here we describe the synthesis and permeability data on a library of 4096 lariat peptides, in which stereochemistry was encoded using isotopic labels and tandem mass spectrometry (MS^2) used to sequence over 1000 of the library members. The data reveal specific patterns of stereochemistry and N-methylation that facilitate

permeability and highlight the impact of the flexible tail on permeability across a wide range of loop geometries.

1.2. Library 1 design and synthesis

The design of Library 1 was inspired by the structures of griselimycin and didemnin B, which feature a linkage between the C-terminus and an internal Thr residue, and an N-acylated, two-residue tail (Figure 1-4). Orally bioavailable compounds exceeding MW 1000 are quite rare.[3] Library 1 samples a chemical space extending slightly beyond this limit. Lipophilicity was kept within a range characteristic of orally bioavailable macrocycles, while the number of hydrogen bond donors was varied.[31] Leucine features prominently in the design of Library 1 as a generic lipophilic amino acid. The library was designed to sample diverse lariat peptide backbone geometries by permuting stereochemistry as well as the number and pattern of amide N-methyl groups within the macrocycle. For Library 1, we included only simple aliphatic side chains to focus specifically on the effect of backbone geometry on membrane permeability. All compounds with the same number of N-methyl groups have the same molecular weight and predicted lipophilicities (as captured by the calculated octanol-water partition coefficient AlogP) (Table 1-1); therefore, variation in permeability within an isomeric series must be the result of conformational differences associated with stereochemistry and N-methyl position. Leucine was selected as the predominant amino acid due to its prevalence among passively permeable natural products, including lariats, with a single Ala residue included to adjust lipophilicity to a range optimal for permeability.[29, 32] Two Pro residues were

included, one in the tail and one in the macrocycle, similar to didemnin B and griselimycin. The average molecular weight of the library was 1016, slightly smaller than griselimycin and didemnin B but significantly larger than model systems previously reported in studies of passive permeability.

The 4096-member Library 1 was prepared as 16 sub-libraries, each theoretically containing a mixture of 256 members (Figure 1-5). The number of compounds in each sub-library was limited by the necessity of chromatographic separation of compounds with the same parent mass during analysis as required for optimal MS^2-based sequencing using CycLS, a program developed previously in our lab.[33] The stereochemistry of residues 1, 2, 8, and 9 are specific to each sub-library. The stereochemistry of residues 4-7 were encoded using stable isotope labelling to allow for identification by LC-MS^2. The number of N-methyl groups in Library 1 ranges from one to five and the number of hydrogen-bond donors from two to six (Table 1-1). The N-methyl group on the lariat tail is constant for all library members while the degree of N-methylation among the mass-encoded residues ranges from zero to four.

The synthesis of Library 1 was devised to avoid potential challenges associated with lariat peptides (Scheme 1-1). Cyclization efficiency often determines the overall efficiency of cyclic peptide synthesis; therefore, we avoided macrolactonization as the cyclization step and instead opted to form the ester linkage on-resin prior to cyclization. However, since continuing Fmoc peptide synthesis on the Thr hydroxyl group could lead to diketopiperazine formation after attachment and deprotection of

the second residue, we chose the Pro^9-Leu^8 connection as the site for cyclization. This strategy has been successful in several lariat depsipeptide natural product syntheses.[34-36] Leu^8 was not N-methylated to prevent premature cleavage from the 2-chlorotrityl resin caused by diketopiperazine formation during Fmoc deprotection.[37] The Thr hydroxyl was not protected during the synthesis.[34-35] Couplings with N-methylated amino acids are known to be challenging. We chose HATU as the coupling agent for library synthesis due to its established use in the coupling of N-methylated amino acids.[38] For the formation of the ester linkage, we adapted conditions applied previously in the total synthesis of theonellapeptolide 1d.[34]

1.3. Library 1 structure-permeability relationships

Overall permeability of Library 1

Permeability was acquired using the parallel artificial membrane permeability assay (PAMPA), a high-throughput method which correlates well to cell-based passive permeability methods and even oral absorption.[39] Permeabilities among library members varied widely, ranging from below 0.01×10^{-6} to over 10×10^{-6} cm/s. In total, 656 compounds in the Library 1 dataset have permeabilities exceeding 1×10^{-6} cm/s. To benchmark the permeability of the lariat peptide library, 1NMe3 (Figure 1-6), a membrane permeable cyclic hexapeptide developed previously,[10] was added to each sub-library. The average permeability of 1NMe3 in sub-libraries 1-16 was 8.1×10^{-6} cm/s (Table 1-2). In total, 29 lariats in Library 1 exceeded this permeability and may be considered highly permeable.

Effect of N-methylation

On average, compounds with more N-Me groups were more permeable (Figure 1-7b), as previously reported in a variety of head-to-tail cyclized peptides.[10, 15, 28, 40] Most of the compounds with no N-Me groups in the macrocycle had negligible permeability. Although the most permeable compounds with only one N-Me group in the macrocycle were among the most permeable in the library, the proportion of impermeable compounds and the variance of $\log P_{app}$ decreased with increasing degree of N-methylation (Table 1-3). To identify features that could be associated with increased permeability when comparing isomeric compounds, we investigated the relationship between the pattern of N-methylation in the macrocycle and permeability. The positions of the N-methyl groups had a marked effect on permeability (Figure 3-2 c, d). Lariats with a single N-methyl group in the macrocycle at R^5, R^6, or R^7 had higher permeability than unmethylated lariats, while N-methylation at R^4 did not significantly increase permeability (Figure 1-7c, d). Similarly, for compounds with two N-methyl groups in the macrocycle, those with N-methylation at R^4 were less permeable than those with the N-methyl groups elsewhere in the ring (Figure 1-7d). No compounds with N-methylation at Leu^4 appear among the 25 most permeable compounds that contain a single macrocyclic N-methyl group. Therefore, although both the number and position of N-methyl groups give rise to strong permeability trends within the library, the large variation in permeability

among closely related sequences highlights the subtle dependence of conformation (and therefore, permeability) on backbone geometry.

Effect of stereochemistry

While the degree and position of N-methylation had a profound effect on permeability, we also observed strong stereochemical effects. We found that the number of heterochiral residue junctions, defined as adjacent residues having opposite stereochemical configurations, correlated strongly with permeability (Figure 1-7f). Most adjacent residue pairs in the macrocycle contributed to this effect, except for MeLeu^{2}/Thr^{3} and Leu^{6}/Ala^{7} (Figure 1-7e).

The general association between heterochirality and permeability occurs for 1-3 degrees of N-methylation but appears strongest for lariats with two N-methyl groups in the macrocycle (Figure 1-8). The effect appears absent at 4 degrees of N-methylation, although the low representation of compounds with 4 degrees of N-methylation and low heterochirality may obscure the effect. For each pair of adjacent stereocenters, the effect of heterochirality is mostly consistent for each degree of N-methylation (Table 1-4). The stereochemistry at Leu^{8} had a clear effect on permeability (Figure 1-9), with higher permeability observed for compounds featuring L stereochemistry at Leu^{8}. This increase in permeability only occurred when either Ala^{7} or Pro^{9} had D stereochemistry (Figure 1-10) and was highest when both Ala^{7} and Pro^{9} had D stereochemistry (Figure 1-11). The impact of individual

stereocenters on permeability was generally consistent for all degrees of N-methylation and does not appear related to variation in representation among the different classes (Tables 1-5 and 1-6). Even in the context of extensive variation in both stereochemistry and N-methylation, certain stereochemical patterns favor permeability, suggesting that some structural elements are consistently instrumental in producing favorable membrane-associated conformations.

Consensus stereochemistry of highly permeable compounds

Polar contacts between macrocycles and their targets are dominated by interactions involving the amide backbone of the macrocycle.[2] Therefore, we were interested in further investigating the factors associated with high permeability among compounds with fewer N-methyl groups. Stereochemical consensus among highly permeable compounds with specific N-methylation patterns indicates which stereocenters are critical for allowing permeable conformations and which may be varied without compromising permeability. A cursory glance revealed that the three most permeable compounds with one N-methyl group in the macrocycle were N-methylated at Ala^{7} and only varied with respect to two stereocenters. To identify more N-methylation and stereochemical patterns associated with high permeability, we examined the 25 most permeable compounds with one or two N-methyl groups in the macrocycle. Among the 25 most permeable compounds with one macrocyclic N-methyl group, 12 were N-methylated at Leu^{6} and showed strong stereochemical consensus at Leu^{4},

Leu^5, Leu^8, and Pro^9 (Figure 1-12, Scaffold A). On the other hand, 17 of the 25 most permeable lariats with two N-methyl groups in the macrocycle were N-methylated at Leu^5 and Ala^7, although there was no stereochemical consensus among these compounds (Figure 1-12, Scaffold B). This N-methylation pattern is clearly exceptional in its ability to accommodate stereochemical variation and thus support diverse permeable 3D geometries.

1.4. Permeability validation by resynthesis of Library 1 members

To validate the permeability results from Library 1, compounds were individually synthesized and evaluated in PAMPA (Figure 1-13, Table 1-8, LCMS and NMR spectra in the experimental section). Compounds were selected at random from the entire dataset and represented a range of permeabilities and N-methyl group counts. Of 11 resynthesized compounds, 9 had retention times matching the corresponding library member, indicating that two of the 11 compounds were mis-sequenced. The permeability trends were reproduced overall, although P_{app} values were roughly four times higher in the library. This effect is most likely a result of the higher total peptide concentration when assaying the library versus individual compounds. In the case of individual compounds, adhesion to the walls of the PAMPA plate and dissolution in the lipid layer may have disproportionately affected the results, whereas in the library these sinks were saturated. We found that the average recovery in the library was 67% while average recovery for the nine individual compounds was 42%, providing some support for this hypothesis. We also assayed these compounds in the MDCK cell-based assay (Figure 1-13). The MDCK assay results confirm the trend in

permeabilities. Interestingly, the P_{app} values obtained in the MDCK assay are generally higher than those obtained in the library PAMPA, in contrast to the P_{app} values from the individual PAMPA.

The occurrence of mis-sequenced entries in the data set is a potential source of error. Of 11 compounds resynthesized individually, two (18%) were mis-sequenced. This is similar to our previous study, in which the mis-sequencing rate was 23%.[33] At worst, patterns of mis-sequencing specifically affecting permeable compounds could result in false conclusions regarding the effect of molecular features on membrane permeability. However, in the absence of evidence to the contrary, mis-sequencing most likely adds noise to the data, weakening observed structure-permeability relationships. Thus, the observed structure-permeability relationships likely emerged in spite of mis-sequencing rather than because of it. Future efforts will be directed toward determining the extent of systematic bias resulting from mis-sequencing, using a much larger set of individually synthesized compounds.

1.5. Analysis of representation in Library 1

The final Library 1 dataset included 1099 unique entries that were sequenced with high confidence using a conservative scoring function based on the original application of CycLS in the identification of cyclic hexa- and heptapeptides. Although the library design describes a 4096-member library, the final data set contains 27% of the theoretical number. This prompted us to determine to what extent the observed structure-property trends were influenced by differences in

representation within the dataset among different stereochemical and N-methylation patterns. We examined representation in the library with respect to all the features from which we derived structure-permeability relationships including number of N-methyl groups, N-methylation pattern, and stereochemistry at individual stereocenters and between adjacent residues.

Chromatographic overlap of compounds with the same m/z impeded data recovery and decreased the representation of lariats in the dataset (Table 1-9). Thus, higher heterochirality (specifically among the mass-encoded positions 4-7) and intermediate numbers of N-methyl groups were associated with decreased representation due to increased mass redundancy (Figure 1-16, Table 1-9). A notable exception occurs for compounds with 4 N-methyl groups in the macrocycle. Here, the sequences with no heterochirality between positions 4-7 were poorly represented in the data. This is most likely due to low synthetic efficiency. Surprisingly, the dataset is biased in favor of lariats with higher degrees of N-methylation (Table 1-11), despite the presumed decreased synthetic efficiency of incorporating N-methylated residues. This is explained by differences in average retention time and retention time variance between each degree of N-methylation (Table 1-12). Retention times vary more for higher levels of N-methylation, resulting in reduced peak overlap. Thus, disparities in representation are related to the chromatographic gradient employed. The gradient was designed to place the bulk of the peaks in the middle of the gradient while ensuring the most lipophilic lariats eluted within the run (conditions provided in Table 1-7). Given the known relationship between reversed-phase retention time and

permeability, the data could be biased towards exclusion of less permeable compounds, although the library chromatograms show the highest peak density towards the middle of the run. The number of compounds in the dataset with each N-methylation pattern varied widely (Table 1-13) as did the number of compounds identified from each sub-library (Table 1-14). However, average permeability did not correlate with the number of compounds with a given N-methylation pattern or in each sub-library (Figure 1-17).

We also examined the effect of differences in representation of stereochemical features on permeability trends. Overall, we did not find a meaningful correlation between representation of heterochiral compounds relative to homochiral compounds and permeability for any of the degrees of N-methylation (Figure 1-18). Considering the known effect of N-methylation on permeability, we assessed the effects of stereochemistry on permeability separately for each degree of N-methylation (Tables 1-4 and 1-5). We noted four representation disparities (exceeding 20% difference), affecting Ala^7, Pro^9, and relative stereochemistry at Leu^5/Leu^6, and Leu^6/Ala^7 (Tables 1-6, 1-10, 1-16, 1-17). The overrepresentation of L-Ala^7 in the library is strongest at lower numbers of N-methyl groups, implying that our observation of higher permeability for D-Ala^7 may be an artifact of representation bias (Table 1-6). The particularly large disparity in representation for Ala^7 stereochemistry combined with that for Leu^6/Ala^7 prompted us to examine the effect of disparate representation of individually impactful stereocenters (Leu^4, Ala^7, and Leu^8) on the heterochirality results for adjacent residue pairs that contain those stereocenters. For Leu^6/Ala^7,

Ala^7/Leu^8, and Leu^8/Pro^9, the more permeable of the two heterochiral configurations was underrepresented (Tables 1-18 and 1-19). This pattern did not occur for the homochiral configurations. If the permeability trends observed here are consistent with the theoretical library space, then this disparity will cause the permeability enhancement associated with heterochirality to be artificially suppressed. To some degree, this would account for the weaker heterochirality effect observed for these three pairs compared to the other three pairs in the cyclic portion of the lariat structure. Moreover, this disparity could significantly skew the average permeability of the entire library downwards. Alternatively, the underrepresentation of the most permeable diastereomers could reflect an absence of the impermeable library members bearing those diastereomers. This appears to be the case for Leu^6/Ala^7 and Ala^7/Leu^8, in which the higher degrees of N-methylation were overrepresented for the most permeable diastereomer, skewing the perceived permeability of the affected diastereomers upwards (Table 1-20). To better understand the extent of this effect, we determined the permeability of all stereochemistries for Leu^6/Ala^7, Ala^7/Leu^8, and Leu^8/Pro^9. The effect of stereochemistry on permeability is consistent for each degree of N-methylation for Ala^7/Leu^8 and Leu^8/Pro^9, while the high permeability of L-Leu^6/D-Ala^7 only occurred for 1 and 2 degrees of N-methylation (Table 1-21). We conclude that the structure-permeability relationships discovered in Library 1 remain valid in spite of differences in representation among structural sub-classes.

1.6. Libraries 2-4: Effect of lariat topology

We investigated the effect of lariat topology on permeability by creating additional libraries with the Thr residue transposed, resulting in varied tail length and ring size (Figure 1-14). For Libraries 2 and 3, the Thr was transposed towards the N-terminus, resulting in lariats containing 8 residues in the macrocycle and one residue in the tail. Library 3 lacks an N-methyl group on the residue adjacent to the Thr to control for the deleterious effect of N-methylation at this position that was observed in the initial library. For Library 4, the Thr residue was situated at the N-terminus, resulting in a 9-residue macrocycle that lacked a lariat tail. Compounds with the same number of N-methyl groups are isomeric, allowing direct comparison of permeability without lipophilicity (ALogP) and molecular weight as confounding factors. Libraries 2-4 were prepared as single sub-libraries of 256 compounds each. The stereochemistry of sub-library 6 from Library 1 ("Library 1.6") was used for Libraries 2-4 as this was the most permeable sub-library.

Library 3 with the 8-1 macrocycle-tail topology was slightly more permeable, on average, than the parent Library 1.6, indicating that the favorable permeability observed for Library 1.6 was not specific to its 7-2 macrocycle-tail topology. Interestingly, the deleterious effect of N-methylation at the residue neighboring the Thr which was observed in Library 1 was also observed in Library 2 with the 8-1 lariat topology. The large difference in permeability between Libraries 2 and 3 further establishes that N-methyl group location is as important as the type of linkage in determining permeability. Library 4, bearing the "9-0" linkage, was significantly less

permeable than any of the other libraries, indicating a role for at least one residue in the tail in facilitating permeability across a broad range of macrocycle geometries. Although Library 4 lacks the single residue lariat tail that is theoretically sufficient to promote a γ-turn involving the Thr NH, as seen in the solution structure of **2**, the significantly diminished permeability of Library 4 may alternatively result from the addition of a second Pro residue to the macrocycle, a rigidifying element that could disfavor permeable conformations.

1.7. Library 5: Sidechain variation

Sidechain substitution is an ideal approach to combinatorial drug discovery. An effect of sidechain identity on the permeability of cyclic peptides has been observed in previous studies.[29, 41] We designed Library 5 to investigate the effect of sidechain variation on the permeability of **2**. We chose **2** as the scaffold on which to perform sidechain substitution because it is highly permeable despite having five HBD, and its relatively low ALogP allowed for substitution with bulkier side chains while keeping much of the library below the solubility threshold. Keeping stereochemistry and N-methylation constant, we varied positions 4-6 among eight nonpolar amino acids, producing a theoretical diversity of 512 library members (Figure 1-15a). At each position, all eight amino acids had unique masses to allow for identification using CycLS.[33] Amino acids without side-chain hydrogen bond donors were selected to avoid scaffold perturbation caused by potential side-chain-to-backbone hydrogen

bonding. The amino acids, ranging in size from Abu to 1-Nal, were chosen at random from Fmoc amino acids already available in the lab. We limited the scope of this library to amino acids without β-branching.

Due to the poor overall aqueous solubility observed for this library, we adapted the "sink PAMPA" conditions that have been described for the analysis of highly lipophilic compounds.[42] Encouragingly, **2** was permeable under these conditions with $P_{app} = 4.6 \times 10^{-6}$ cm/s. After data processing, the dataset contained 121 library members from a theoretical diversity of 512.

Of the 121 compounds identified in Library 5, 84 had P_{app} values above 1×10^{-6} cm/s (Figure 1-15b). As seen in previous studies in both peptidic and non-peptidic systems, the relationship between AlogP and permeability followed a bell-shaped curve,[15, 29, 43] indicative of the mutually opposing effects of lipophilicity and solubility on permeabililty along the polarity continuum.[32] Permeable library members were most abundant between AlogP of 3 and 4. Of the 57 Library 5 members with AlogP values in this range, 52 had P_{app} values above 1×10^{-6} cm/s. Although **2** was one of the more permeable compounds in Library 5, side chain substitution did not abrogate permeability. Substitution with Abu did not prevent permeability at any of the three positions, indicating that steric shielding of HBDs is not vital for allowing permeability. This result indicates that mass-encoded libraries probing stereochemistry and N-methylation may be useful for the development of permeable

libraries with diversity provided by side chain variation, at least among those with nonpolar funtionality.

1.8. Structural study of a permeable lariat peptide from Library 1

To better understand the low dielectric conformations facilitating high permeability among lariats with multiple hydrogen bond donors (HBDs), we used NMR and molecular modelling to study the solution conformation of compound **2** (Figure 1-19) in $CDCl_3$, the most permeable library member with only one macrocyclic N-methyl amide. The compound has several features associated with high permeability: high heterochirality, D stereochemistry at Leu^4, non-N-methylated Leu^4, and L stereochemistry at Leu^8. We used ROESY-derived interproton distance restraints (Table 1-22) and dihedral (ϕ-angle) restraints derived from ^3J vicinal coupling constants between Hα-HN protons (Table 4-2). These restraints were provided as input to ForceGen, an algorithm designed to model macrocycles using NMR-derived restraints.[44-45] Our implementation of this method is described in the Supporting Information. Proton NMR peak assignments are listed in Table 1-25.

A very strong ROESY crosspeak between the Hα atoms of $D\text{-}Leu^6$ and $D\text{-}MeAla^7$ provided convincing evidence that this amide bond adopts the cis conformation (ω = 0°). Strong ROESY crosspeaks between the δ protons of Pro^9 and the α-protons of Leu^8, and between the δ protons of Pro1 and the terminal acetyl protons, along with an absence of Hα-Hα crosspeaks, indicate trans geometry for these amides. The strong ROESY crosspeak between the N-methyl protons of $D\text{-}MeLeu^2$ and the a-

proton of Pro[1], along with no Hα-Hα crosspeak, indicate a trans geometry for this amide as well. These backbone amide geometries were therefore used as ω torsional restraints in the ForceGen structure calculations.

Consistent with its membrane permeability, the NMR solution conformation of **2** in chloroform is characterized by an extensive intramolecular hydrogen bonding (IMHB) network involving all five HBDs (Figure 1-19). The type VI β-turn centered about the cis-amide between D-Leu[6] and D-MeAla[7] facilitates two transannular hydrogen bonds between Leu[5] and Leu[8], flanked on one side by an inverse γ-turn centered around Leu[5]. The NH group of Leu[4] points towards the center of the macrocycle to form hydrogen bonds with the Thr[3] sidechain and Leu[8] carbonyl. In the lariat tail, the NH group of Thr[3] is sequestered from solvent by a γ-turn centered around D-MeLeu[2]. While the central ring motif is well conserved among the lowest energy conformers, the tail is quite mobile. We used variable temperature ^{1}H NMR to determine the extent of solvent exposure of each NH group. The low temperature shift coefficients (<4 ppb/K) are consistent with the NMR structure showing exclusion of all five NH groups from solvent (Table 1-24).

Some backbone features associated with permeability may be understood in the context of this 3D solution structure. Heterochirality in the region between Thr[3] and Leu[6] likely supports IMHB networks by providing a torsional space that favors turns and allows transannular orientation of HBDs.[46-48] For a transannular orientation of HBDs to be permitted, bulky leucine sidechains must face opposite directions when

adjacent, requiring heterochiral relationships. Additional solution structures of compounds from this library are likely to reveal low-dielectric conformations that may provide additional insight into the observed structure-permeability relationships.

1.9. Bioactivity of Library 1

Library 1 was designed to approximate the size and shape of bioactive lariat peptide natural products, particularly griselimycin and didemnin B. To determine the bioactivity of these lariat peptides, we performed cytological profiling in HeLa cells. In the first round of screening, the sixteen sub-libraries from Library 1, each a mixture of 256 compounds (Figure 1-20), were added to HeLa cells at a concentration of 10 μM and cellular changes examined. The eight sub-libraries with an L-Leu at position 8 produced pronounced phenotypic disturbance while those with D-Leu at position 8 had little effect (Figure CP fingerprint). The strongest effect was observed for sub-library 6 (Figure 1-20), indicated as "LDLD" in the cytological fingerprint). This sub-library produced a bizarre elongation of the microtubules, along with absence of mitotic cells (anti-pH3-Ab) and S-phase (EdU) cells (Figure 1-21). With the goal of identifying individually bioactive compounds, we fractionated sub-library 6 by HPLC and screened the fractions for bioactivity. None of the fractions were associated with significant bioactivity. We instead pursued a strategy of iterative resynthesis. First, we synthesized sub-library 6 as sixteen separate subdivisions of 16 compounds each and screened these for bioactivity (Figure 1-22). Three of the subdivisions decreased EdU staining at 3 μM, but only one decreased EdU at 1.6 μM. This subdivision, "LHLMe", produced some of the cellular effects of sub-library 6,

such as microtubule elongation and disappearance of EdU staining, but to less of an extent (Figure 1-23). We followed up on this subdivision by individually synthesizing each of the sixteen compounds comprising the subdivision. One of the compounds, compound 16, inhibited EdU intensity at lower concentrations than the other fifteen compounds, but not with sufficient potency to fully account for the effects of sub-library 6 (Figure 1-24). The low potency of this compound combined with the observation of bioactivity in multiple subdivisions leads us to conclude that the bioactivity of sub-library 6 is due to the combined action of many library members.

1.10. Conclusions

We set out to investigate the permeability landscape of lipophilic lariat peptides inspired by natural products. Using a mass-encoded library with varied stereochemistry and N-methylation but invariant, aliphatic side chains, we obtained the permeabilities of over 1000 lariat peptides with diverse backbone geometries. The striking variation in passive permeability observed among isomeric compounds that differ only in N-methyl position and stereochemistry highlight the key role of conformation in determining passive membrane permeability in this chemical space. Although we identified intriguing structure-permeability relationships from the data without using computational tools, automated data mining techniques would undoubtedly reveal deeper relationships. In addition, machine learning approaches may yield models with more predictive power that could be applied to a larger set of scaffolds. Nonetheless, the surprisingly high number of permeable scaffolds in this

dataset indicates that the landscape of passive permeability in lariat depsipeptides extends well beyond that defined by existing lariat peptide natural products.

Although the vast majority of orally bioavailable macrocycles are below MW 1000,[3] permeability above MW 1000 has been reported for synthetic macrocycles. A designed permeable cyclic decapeptide scaffold has been reported previously,[17] and further studied with side-chain substitution and peptide-peptoid substitution.[18] However, the results reported here reveal numerous scaffolds supporting drug-like permeability above MW 1000. The permeability of many compounds in this library, even some with as many as 5 HBDs, as well as the maintenance of permeability upon side chain variation of compound **2**, indicates a potentially important role for lariat peptides in future discovery efforts. The crystal structure of the DNA sliding clamp from *M. smegmatis* in complex with griselimycin, in which the cyclic part of the molecule binds one subsite and the lariat tail extends into an adjacent subsite,[20] illustrates the unique potential of lariats to target biomolecular interactions.

Our results reveal that, as with cyclic peptides, the permeability of lariat peptides is sensitive to stereochemistry as well as the number and location of N-methyl groups. The emergence of consensus features and general trends among permeable compounds illustrates how this data can be applied to the design of compounds or libraries biased towards membrane permeability in this relatively uncharted chemical space.

1.11. Synthetic methods

Loading of 2-chlorotrityl resin

The desired amount of 2-chlorotrityl chloride resin was swelled with DCM in a SPE tube for 1 h, at which time a solution of Fmoc-D-leucine (2 eq) and DIPEA (3 eq) in DCM was added. The tube was capped and inverted, and the stopcock opened to allow gas evolution. Once gas evolution subsided, the tube was shaken for 3 h. The resin was capped with a solution of 2:1:17 MeOH:DIPEA:DCM (2 x 30 min). The resin was washed with DMF (3x) followed by DCM (3x). The loading value was calculated by quantifying UV absorbance of the dibenzofulvene byproduct (300 nm) after Fmoc removal.

Library synthesis

The libraries were synthesized starting with 2-chlorotrityl resin loaded with residue 8 (Scheme 1-1). The procedures for manual amide coupling and Fmoc deprotection were used to install residues 1-7 without protection of the threonine side chain. At steps requiring transfer of resin, transfers were carried out prior to Fmoc deprotection. Residue 9 was added as Fmoc-proline by ester coupling. After Fmoc deprotection and cleavage from resin, cyclization (residue 9 – residue 8) was carried out in solution. The lariat peptides were purified by reversed-phase chromatography.

Library 1 was prepared in 16 sub-libraries at a scale of 0.025 mmol/sub-library using a split-pool strategy (Scheme S2). Sub-libraries **1-8** were prepared from L-Leu-2CT and sub-libraries **9-16** from D-Leu-2CT. Couplings were performed manually. For

the installation of residues 7, 6, 5, and 4, the resin was separated into four tubes for the addition of different amino acids and recombined and mixed prior to Fmoc deprotection.

Libraries **2-4** were prepared as single mixtures at a scale of 0.025 mmol starting from L-Leu-2CT resin and installing amino acids in the appropriate order. For Library 4, the N-terminal acetyl group was installed using the procedure for N-terminal acetylation.

Library **5** was prepared as a single mixture at a scale of 0.2 mmol starting from L-Leu-2CT resin using a split-pool strategy. For the installation of residues 6, 5, and 4, the resin was separated into eight tubes for the addition of different amino acids and recombined and mixed prior to Fmoc deprotection.

Synthesis of pure lariat peptides

Lariat peptides were synthesized starting with 2-chlorotrityl resin loaded with L-Leu or D-Leu. Further residues were added using the procedure for automated peptide synthesis yielding a linear octapeptide bound to resin with an unprotected Thr sidechain. The free hydroxyl group was then acylated with Fmoc-L-Pro or Fmoc-D-Pro by the procedure for ester coupling. After final Fmoc removal, the peptide was removed from the resin using the procedure for cleavage from 2-chlorotrityl resin. The peptide was cyclized using the procedure for cyclization and purified using the procedure for purification.

Manual solid-phase peptide synthesis

To a solution of the Fmoc amino acid (2 eq, 0.5 M in DMF) was added HATU (1.9 eq, 0.5 M in DMF) followed by DIPEA (2.5 eq). The resultant solution was swirled and allowed to stand for 5 minutes, then added to the drained resin. The resin was heated to 50°C for 2 h. After coupling, the resin was washed with DMF (3x) and DCM (3x). For Fmoc deprotection, the resin was treated with a solution containing 2% 1,8-diazabicycloundec-7-ene (DBU) and 2% piperidine in DMF for 15 min at room temperature. The resin was then washed with DMF (3x) and DCM (3x).

Automated peptide synthesis (Prelude X, Protein Technologies)

Synthesis was performed on 0.1 mmol scale using loaded 2-chlorotrityl resins. Fmoc deprotection was accomplished with 2% DBU and 2% piperidine in DMF for 1 min at 90°C. Couplings were carried out with Fmoc-protected amino acids (5 eq), COMU (4 eq), and DIPEA (6 eq) in DMF for 10 min at 90°C. Each coupling and deprotection step was followed by a wash with DMF (4x) and DCM (2x).

Ester formation using DIC

Fmoc-proline (10 eq) was dissolved in DMF/DCM (1:9, roughly 3mL/g Fmoc-proline). DMAP was added (0.25 eq) followed by DIC (10 eq). The solution was swirled rapidly until a precipitate formed. The mixture was added to the resin and the SPE tube capped. The reaction was shaken at room temperature for 3 h, then drained. Another portion of reactants (10 eq Fmoc-Pro-OH, 0.25 eq DMAP, 10 eq DIC) was immediately added without washing the resin and the resin was allowed to react

overnight at room temperature. The resin was drained and washed with DMF until the thick precipitate which formed during the reaction was removed, then washed with DCM (3x).

Cleavage from 2-chlorotrityl resin

Branched linear peptides were cleaved from the resin with 25% HFIP in DCM (2 x 30 min). The resin was rinsed with DCM between treatments. Solvent was evaporated under a stream of nitrogen. DCM was added and evaporation repeated. The residue was stored overnight in a vacuum desiccator prior to cyclization.

Cyclization

The solvent volumes in this procedure are for 0.1 mmol of peptide. The concentration during cyclization was approximately 0.001 M.

COMU (3 eq) was placed in a round-bottom flask with a stir bar, followed by THF (90 mL) and DIPEA (3 eq). In a separate vessel, the branched linear peptide was dissolved in ACN (10 mL) and DIPEA (3 eq). The peptide was added dropwise and in portions to the round-bottom flask with rapid stirring during 30 minutes. Stirring was continued for 16 h. The solution was concentrated under reduced pressure.

N-Terminal acetylation

A mixture containing acetic anhydride (6 eq), DIPEA (7.5 eq), and DMF (0.6mL) was added to the drained resin. The resin was shaken for 2 h at room temperature, drained, and washed with DMF (3x) and DCM (3x).

Lariat peptide purification

Crude cyclic peptides were purified on a Biotage Isolera Prime automated chromatography system equipped with a SNAP Bio C18 25g column eluting with water/acetonitrile containing 0.1% TFA. Libraries were purified with the following gradient (%ACN in water): 30% (50mL), 30-100% (25mL), 100% (25mL). Individual compounds were purified with the following gradient (%ACN in water): 20% (50mL), 20-80% (450mL), 100% (75mL).

Fmoc protection of free amino acids

Fmoc-protected L-Ala-d_3 and L-Leu-d_3 were prepared by Fmoc protection of the commercially available deuterated amino acids according to literature procedure.[49]

N-Methylation of Fmoc amino acids

Fmoc-N-methyl amino acids were prepared from Fmoc amino acids by reduction of the formaldehyde oxazolidinone as previously described.[50]

1.12. Analytical methods and assays

Library analysis by UPLC-MS^2

Acquisition of MS^2 data for sequencing was carried out as reported previously, with some modifications.[33] Source ionization was used, as we suspected this may facilitate fragmentation at the ester bond. Source ionization was optimized to produce the strongest signal from the M+Na mass and minimize signal from the M+H mass, which could interfere with sequencing. For libraries 1-4, the top 7 most intense non-isotopic peaks were selected for MS^2 acquisition at each MS^1 acquisition. For Library 5, the top 10 were selected. LCMS conditions are described in Table 3-6.

PAMPA Assay

The PAMPA assay was carried out and the peak volumes interpreted using a procedure utilized previously in our lab.[32]

The analyte concentration in the PAMPA assay was 250 μM for sub-libraries (roughly 1 μM per compound) and 1 μM for compounds assayed individually. Internal standards were included in the assay at a concentration of 1 μM. For sub-libraries, 1NMe3, synthesized according to published procedures,[10] was used as the internal standard and for compounds assayed individually, carbamazepine was used as the standard. The assay was run for 18 h.

A 96-well donor plate with 0.45 μm hydrophobic Immobilon-P membrane supports (Millipore MAIPNTR10) and a 96-well Teflon acceptor plate (Millipore MSSACCEPTOR) were used in the PAMPA permeability test. The acceptor plate

was prepared by adding 300 µL of 5% DMSO in 1X PBS to each well. Donor well solutions were prepared by diluting 50 µL DMSO stock solutions prepared above to a final volume of 1000 µL with PBS and mixed thoroughly. The frits were infused with 5 µL of dodecane containing 1% (w/v) soy lecithin (90%, Alfa Aesar). The membranes were allowed to equilibrate for 5 minutes before adding the donor well solution and placing on top of the acceptor well solution to begin the assay. Samples were prepared for LC-MS analysis by diluting with an equal volume of ACN. The donor wells were further diluted tenfold with 1:1 ACN/H_2O for approximately even analyte concentration in the donor and acceptor wells.

Sink PAMPA conditions

The assay was run as described above, except that the donor well contained 0.2% (v/v) polysorbate 80 and the acceptor well contained 0.2% (w/v) TPGS-750M. Donor well samples were prepared for LC-MS analysis by diluting with an equal volume of 9:1 ACN / 2% (w/v) TPGS-750M in water. Acceptor well samples were prepared for LC-MS analysis by diluting with an equal volume of 9:1 ACN / 2% (v/v) polysorbate 80 in water.

Library Data Processing

The PAMPA LC-MS data was processed using AUTOPAMPA, CycLS, RTMerge. These programs and instructions for installation and use are available on GitHub at https://github.com/LokeyLab/PAMPA-Analysis-Support-Tools.

Although the parent mass provides the degree of N-methylation and number of residues of each stereochemistry, the parent mass does not provide the order of residues. Previously we reported an algorithm for deconvoluting cyclic peptide libraries based on matching their MS^2 fragment ions to a virtual library derived from the theoretical compounds present in each sub-library.[33] After processing the raw MS^2 spectra using CycLS, we removed all data with sequencing confidence scores below 0.01 (the confidence score refers to the difference between the sequencing scores of the highest scoring sequence and the second highest scoring sequence divided by the highest score).[33] Duplicate sequences were resolved by removing the entry with lower confidence score. For data with assigned permeability values of 0, we integrated the LC-MS peaks manually wherever possible. When the peaks could not be integrated, the data were discarded.

MDCK Assay

The MDCK assay was carried out according to the procedure utilized by Furukawa et al.[18] Transcellular transport of the test compounds from apical to basal direction using MDCK II cells was investigated. For the transcellular transport assay, the culture medium on the apical side and the basal side was replaced with buffer (pH 6.5) and buffer (pH 7.4) containing BSA, respectively. The buffer on the apical side was replaced with the buffer containing 10 μM of test compounds to start the incubation. After the incubation, aliquots of the solutions were sampled from both the apical side and basal side, and the concentrations of each compound were determined by LC-MS/MS.

Apparent permeability coefficient (P_{app}) was calculated by the following equations.

$$P_{app} = (Cb \times V) / (Ca \times t \times A)$$

P_{app}: apparent permeability (10^{-6} cm/sec)

Ca: test compound concentration added to the apical side (μM)

Cb: test compound concentration on the basal side (μM)

A: surface area of cell monolayer (cm^2)

V: volume of buffer on the basal side (cm^3)

t: incubation time (sec)

Bioactivity screening

Libraries and compounds were screened for bioactivity using cytological profiling as described by Woehrmann et al, 2013. Two stain sets were used: set 1 (Hoechst (DNA), anti-pH3-Ab (mitosis), EDU (S-phase)) and set 2 (Hoechst (DNA), TMR-phalloidin (actin), anti-tubulin Ab (MTs)).

NMR Solution Structure Generation for Compound 2

A 2D ROESY spectrum of compound **2** was obtained at 277K in chloroform-*d* with a mixing time of 300ms, which was in the linear range in cross-relaxation ROESY buildup curve as determined by performing separate 1D ROESY experiments with mixing times of 100, 200, 300, 400, 500, and 600 ms. Cross peaks were classified as strong, medium, or weak by visual inspection. The following distance restraints were applied:

Strong	1.7 – 2.5 Å
Medium	2.5 – 3.5 Å
Weak	3.5 – 4.5 Å

We avoided ROESY crosspeaks involving leucine sidechain atoms when selecting distance restraints due to the high level of peak overlap in the upfield region of the spectrum. In total, thirteen distance restraints were applied to the agnostic conformer pool (Table 1-22).

The amide resonances were sufficiently sharp to obtain HN-Hα *J*-coupling values for calculating dihedral restraints. The Karplus relationship was used to obtain estimates of the ϕ dihedral angles. Theses ϕ angle values are listed in Table 1-23.

To determine the solution structure of **2**, we used the ForceGen approach.[45] Beginning with an initial set of NMR constraints, some of which might be degenerate and some of which might be incorrect, an initial conformer pool is produced without the use of NMR data. That pool is profiled against the full set of NMR constraints, allowing for selection of a subset of non-degenerate constraints that are consistent with some of the conformers in the pool. In an iterative process, new conformer pools are produced using a set of NMR constraints that explores the space of additions to the prior set of constraints where the new constraints are shown to be feasible from the prior conformer pool. In cases where multiple possibilities exist for a particular constraint, all of which are feasible, they are all explored. The process ends when all choices from among degenerate constraints are made and when no non-degenerate constraints are feasible to add. In this case, all non-degenerate constraints were selected for the final set, and a single choice was made for each degenerate constraint.

1.13. References

1. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings 1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997). *Advanced Drug Delivery Reviews* **2001,** *46* (1-3), 3-26.

2. Villar, E. A.; Beglov, D.; Chennamadhavuni, S.; Porco, J. A., Jr.; Kozakov, D.; Vajda, S.; Whitty, A., How proteins bind macrocycles. *Nat Chem Biol* **2014,** *10* (9), 723-31.

3. Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J., Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem Biol* **2014,** *21* (9), 1115-42.

4. Giordanetto, F.; Kihlberg, J., Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? *Journal of Medicinal Chemistry* **2014,** *57* (2), 278-295.

5. Doak, B. C.; Zheng, J.; Dobritzsch, D.; Kihlberg, J., How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. *Journal of Medicinal Chemistry* **2016,** *59* (6), 2312-2327.

6. Blundell, T. L.; Sibanda, B. L.; Montalvão, R. W.; Brewerton, S.; Chelliah, V.; Worth, C. L.; Harmer, N. J.; Davies, O.; Burke, D., Structural biology and bioinformatics in drug design: opportunities and challenges for target identification

and lead discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2006,** *361* (1467), 413-423.

7. El Tayar, N.; Mark, A. E.; Vallat, P.; Brunne, R. M.; Testa, B.; Van Gunsteren, W. F., Solvent-dependent conformation and hydrogen-bonding capacity of cyclosporin A: evidence from partition coefficients and molecular dynamics simulations. *Journal of Medicinal Chemistry* **1993,** *36* (24), 3757-3764.

8. Rezai, T.; Yu, B.; Millhauser, G. L.; Jacobson, M. P.; Lokey, R. S., Testing the Conformational Hypothesis of Passive Membrane Permeability Using Synthetic Cyclic Peptide Diastereomers. *Journal of the American Chemical Society* **2006,** *128* (8), 2510-2511.

9. Nielsen, D. S.; Shepherd, N. E.; Xu, W.; Lucke, A. J.; Stoermer, M. J.; Fairlie, D. P., Orally Absorbed Cyclic Peptides. *Chemical Reviews* **2017,** *117* (12), 8094-8128.

10. White, T. R.; Renzelman, C. M.; Rand, A. C.; Rezai, T.; McEwen, C. M.; Gelev, V. M.; Turner, R. A.; Linington, R. G.; Leung, S. S.; Kalgutkar, A. S.; Bauman, J. N.; Zhang, Y.; Liras, S.; Price, D. A.; Mathiowetz, A. M.; Jacobson, M. P.; Lokey, R. S., On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat Chem Biol* **2011,** *7* (11), 810-7.

11. Bockus, A. T.; Lexa, K. W.; Pye, C. R.; Kalgutkar, A. S.; Gardner, J. W.; Hund, K. C.; Hewitt, W. M.; Schwochert, J. A.; Glassey, E.; Price, D. A.; Mathiowetz, A. M.; Liras, S.; Jacobson, M. P.; Lokey, R. S., Probing the Physicochemical Boundaries of Cell Permeability and Oral Bioavailability in

Lipophilic Macrocycles Inspired by Natural Products. *J Med Chem* **2015,** *58* (11), 4581-9.

12. Bockus, A. T.; Schwochert, J. A.; Pye, C. R.; Townsend, C. E.; Sok, V.; Bednarek, M. A.; Lokey, R. S., Going Out on a Limb: Delineating The Effects of beta-Branching, N-Methylation, and Side Chain Size on the Passive Permeability, Solubility, and Flexibility of Sanguinamide A Analogues. *J Med Chem* **2015,** *58* (18), 7409-18.

13. Nielsen, D. S.; Hoang, H. N.; Lohman, R. J.; Hill, T. A.; Lucke, A. J.; Craik, D. J.; Edmonds, D. J.; Griffith, D. A.; Rotter, C. J.; Ruggeri, R. B.; Price, D. A.; Liras, S.; Fairlie, D. P., Improving on nature: making a cyclic heptapeptide orally bioavailable. *Angew Chem Int Ed Engl* **2014,** *53* (45), 12059-63.

14. Schwochert, J.; Turner, R.; Thang, M.; Berkeley, R. F.; Ponkey, A. R.; Rodriguez, K. M.; Leung, S. S.; Khunte, B.; Goetz, G.; Limberakis, C.; Kalgutkar, A. S.; Eng, H.; Shapiro, M. J.; Mathiowetz, A. M.; Price, D. A.; Liras, S.; Jacobson, M. P.; Lokey, R. S., Peptide to Peptoid Substitutions Increase Cell Permeability in Cyclic Hexapeptides. *Org Lett* **2015,** *17* (12), 2928-31.

15. Wang, C. K.; Northfield, S. E.; Swedberg, J. E.; Colless, B.; Chaousis, S.; Price, D. A.; Liras, S.; Craik, D. J., Exploring experimental and computational markers of cyclic peptides: Charting islands of permeability. *European Journal of Medicinal Chemistry* **2015,** *97*, 202-213.

16. Pye, C. R.; Hewitt, W. M.; Schwochert, J.; Haddad, T. D.; Townsend, C. E.; Etienne, L.; Lao, Y.; Limberakis, C.; Furukawa, A.; Mathiowetz, A. M.; Price, D. A.;

Liras, S.; Lokey, R. S., Nonclassical Size Dependence of Permeation Defines Bounds for Passive Adsorption of Large Drug Molecules. *J Med Chem* **2017,** *60* (5), 1665-1672.

17. Fouche, M.; Schafer, M.; Berghausen, J.; Desrayaud, S.; Blatter, M.; Piechon, P.; Dix, I.; Martin Garcia, A.; Roth, H. J., Design and Development of a Cyclic Decapeptide Scaffold with Suitable Properties for Bioavailability and Oral Exposure. *ChemMedChem* **2016,** *11* (10), 1048-59.

18. Furukawa, A.; Schwochert, J.; Pye, C. R.; Asano, D.; Edmondson, Q. D.; Turmon, A.; Klein, V.; Ono, S.; Okada, O.; Lokey, R. S., Drug-like properties in macrocycles above MW 1000: Backbone rigidity vs. side-chain lipophilicity. *Angewandte Chemie International Edition* **2020**.

19. van Santen, J. A. J., G.; Leen Singh, A.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Carnevale Neto, F.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.; Cleary Little, J. L.; Delgadillo, D. A.; Dorrestein, P. C.; Duncan, K. R.; Egan, J. M.; Galey, M. M.; Haeckl, F. P. J.; Hua, A.; Hughes, A. H.; Iskakova, D.; Khadilkar, A.; Lee, J.-H.; Lee, S.; LeGrow, N.; Liu, D. Y.; Macho, J. M.; McCaughey, C. S.; Medema, M. H.; Neupane, R. P.; O'Donnell, T. J.; Paula, J. S.; Sanchez, L. M.; Shaikh, A. F.; Soldatou, S.; Terlouw, B. R.; Tran, T. A.; Valentine, M.; van der Hooft, J. J. J.; Vo, D. A.; Wang, M.; Wilson, D.; Zink, K. E.; Linington, R. G., The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Central Science* **2019,** *5* (11), 1824-1833.

20. Kling, A.; Lukat, P.; Almeida, D. V.; Bauer, A.; Fontaine, E.; Sordello, S.; Zaburannyi, N.; Herrmann, J.; Wenzel, S. C.; Konig, C.; Ammerman, N. C.; Barrio, M. B.; Borchers, K.; Bordon-Pallier, F.; Bronstrup, M.; Courtemanche, G.; Gerlitz, M.; Geslin, M.; Hammann, P.; Heinz, D. W.; Hoffmann, H.; Klieber, S.; Kohlmann, M.; Kurz, M.; Lair, C.; Matter, H.; Nuermberger, E.; Tyagi, S.; Fraisse, L.; Grosset, J. H.; Lagrange, S.; Muller, R., Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **2015,** *348* (6239), 1106-1112.

21. Marco, E.; Martín-Santamaría, S.; Cuevas, C.; Gago, F., Structural Basis for the Binding of Didemnins to Human Elongation Factor eEF1A and Rationale for the Potent Antitumor Activity of These Marine Natural Products. *Journal of Medicinal Chemistry* **2004,** *47* (18), 4439-4452.

22. Chun, H.; Davies, B.; Hoth, D.; Suffness, M.; Plowman, J.; Flora, K.; Grieshaber, C.; Leyland-Jones, B., Didemnin B. *Investigational New Drugs* **1986,** *4* (3).

23. Hetherington, A. M.; Sawyez, C. G.; Sutherland, B. G.; Robson, D. L.; Arya, R.; Kelly, K.; Jacobs, R. L.; Borradaile, N. M., Treatment with didemnin B, an elongation factor 1A inhibitor, improves hepatic lipotoxicity in obese mice. *Physiological Reports* **2016,** *4* (17), e12963.

24. PharmaMar, PharmaMar announces the approval of Aplidin in Australia for the treatment of multiple myeloma. 2018.

25. Hodgson, J., The pandemic pipeline. *Nature biotechnology* 20 March 2020, 2020.

26. Fredersdorf, M.; Kurz, M.; Bauer, A.; Ebert, M. O.; Rigling, C.; Lannes, L.; Thiele, C. M., Conformational Analysis of an Antibacterial Cyclodepsipeptide Active against Mycobacterium tuberculosis by a Combined ROE and RDC Analysis. *Chemistry* **2017,** *23* (24), 5729-5735.

27. J., H. M. B. V. d. H. D. A. J. S. G. M. S. S. K. W. A., Crystal and molecular structure of didemnin B, an antiviral and cytotoxic depsipeptide. *Proc Natl Acad Sci U S A* **1988,** *85*, 4118-4122.

28. Hewitt, W. M.; Leung, S. S.; Pye, C. R.; Ponkey, A. R.; Bednarek, M.; Jacobson, M. P.; Lokey, R. S., Cell-permeable cyclic peptides from synthetic libraries inspired by natural products. *J Am Chem Soc* **2015,** *137* (2), 715-21.

29. Furukawa, A.; Townsend, C. E.; Schwochert, J.; Pye, C. R.; Bednarek, M. A.; Lokey, R. S., Passive Membrane Permeability in Cyclic Peptomer Scaffolds Is Robust to Extensive Variation in Side Chain Functionality and Backbone Geometry. *J Med Chem* **2016,** *59* (20), 9503-9512.

30. Townsend, C.; Jason, E.; Naylor, M. R.; Pye, C. R.; Schwochert, J.; Edmonson, Q. D.; Lokey, R. S., The Passive Permeability Landscape Around Geometrically Diverse Hexa- and Heptapeptide Macrocycles. *ChemRxiv archive* **2020**.

31. Over, B.; Matsson, P.; Tyrchan, C.; Artursson, P.; Doak, B. C.; Foley, M. A.; Hilgendorf, C.; Johnston, S. E.; Lee, M. D.; Lewis, R. J.; McCarren, P.; Muncipinto, G.; Norinder, U.; Perry, M. W. D.; Duvall, J. R.; Kihlberg, J., Structural and

conformational determinants of macrocycle cell permeability. *Nature Chemical Biology* **2016,** *12* (12), 1065-1074.

32. Naylor, M. R.; Ly, A. M.; Handford, M. J.; Ramos, D. P.; Pye, C. R.; Furukawa, A.; Klein, V. G.; Noland, R. P.; Edmondson, Q.; Turmon, A. C.; Hewitt, W. M.; Schwochert, J.; Townsend, C. E.; Kelly, C. N.; Blanco, M. J.; Lokey, R. S., Lipophilic Permeability Efficiency Reconciles the Opposing Roles of Lipophilicity in Membrane Permeability and Aqueous Solubility. *J Med Chem* **2018,** *61* (24), 11169-11182.

33. Townsend, C.; Furukawa, A.; Schwochert, J.; Pye, C. R.; Edmondson, Q.; Lokey, R. S., CycLS: Accurate, whole-library sequencing of cyclic peptides using tandem mass spectrometry. *Bioorg Med Chem* **2018,** *26* (6), 1232-1238.

34. Kuranaga, T.; Enomoto, A.; Tan, H.; Fujita, K.; Wakimoto, T., Total Synthesis of Theonellapeptolide Id. *Org Lett* **2017,** *19* (6), 1366-1369.

35. Yao, G.; Wang, W.; Ao, L.; Cheng, Z.; Wu, C.; Pan, Z.; Liu, K.; Li, H.; Su, W.; Fang, L., Improved Total Synthesis and Biological Evaluation of Coibamide A Analogues. *J Med Chem* **2018,** *61* (19), 8908-8916.

36. Seo, H.; Lim, D., Total Synthesis of Halicylindramide A. *The Journal of Organic Chemistry* **2009,** *74* (2), 906-909.

37. Teixidó, M.; Albericio, F.; Giralt, E., Solid-phase synthesis and characterization of N-methyl-rich peptides. *The Journal of Peptide Research* **2008,** *65* (2), 153-166.

38. Yvonne M. Angell, C. G.-E., Daniel H. Rich, Comparative Studies of the Coupling of N-Methylated, Sterically Hindered Amino Acids During Solid-Phase Peptide Synthesis. *Tetrahedron Letters* **1994,** *35*, 5981-5984.

39. Kansy, M.; Senner, F.; Gubernator, K., Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J Med Chem* **1998,** *41* (7), 1007-10.

40. Wang, C. K.; Northfield, S. E.; Colless, B.; Chaousis, S.; Hamernig, I.; Lohman, R.-J.; Nielsen, D. S.; Schroeder, C. I.; Liras, S.; Price, D. A.; Fairlie, D. P.; Craik, D. J., Rational design and synthesis of an orally bioavailable peptide guided by NMR amide temperature coefficients. *Proceedings of the National Academy of Sciences* **2014,** *111* (49), 17504-17509.

41. Rand, A. C.; Leung, S. S.; Eng, H.; Rotter, C. J.; Sharma, R.; Kalgutkar, A. S.; Zhang, Y.; Varma, M. V.; Farley, K. A.; Khunte, B.; Limberakis, C.; Price, D. A.; Liras, S.; Mathiowetz, A. M.; Jacobson, M. P.; Lokey, R. S., Optimizing PK properties of cyclic peptides: the effect of side chain substitutions on permeability and clearance(). *Medchemcomm* **2012,** *3* (10), 1282-1289.

42. Oh, M. H.; Lee, H. J.; Jo, S. H.; Park, B. B.; Park, S.-B.; Kim, E.-Y.; Zhou, Y.; Jeon, Y. H.; Lee, K., Development of Cassette PAMPA for Permeability Screening. *Biological & Pharmaceutical Bulletin* **2017,** *40* (4), 419-424.

43. Sawada, G. A.; Barsuhn, C. L.; Lutzke, B. S.; Houghton, M. E.; Padbury, G. E.; Ho, N. F. H.; Raub, T. J., Increased Lipophilicity and Subsequent Cell Partitioning

Decrease Passive Transcellular Diffusion of Novel, Highly Lipophilic Antioxidants. *J Pharm Exp Ther* **1999,** *288* (3), 1317-1326.

44. Cleves, A. E.; Jain, A. N., ForceGen 3D structure and conformer generation: from small lead-like molecules to macrocyclic drugs. *Journal of Computer-Aided Molecular Design* **2017,** *31* (5), 419-439.

45. Jain, A. N.; Cleves, A. E.; Gao, Q.; Wang, X.; Liu, Y.; Sherer, E. C.; Reibarkh, M. Y., Complex macrocycle exploration: parallel, heuristic, and constraint-based conformer generation using ForceGen. *Journal of Computer-Aided Molecular Design* **2019,** *33* (6), 531-558.

46. Haque TS, L. J., Gellman SH, Stereochemical Requirements for ?beta-Hairpin Formation: Model Studies with Four-Residue Peptides and Depsipeptides. *Journal of the American Chemical Society* **1996,** *118* (29), 6975-6985.

47. Ghosh, D.; Lahiri, P.; Verma, H.; Mukherjee, S.; Chatterjee, J., Engineering β-sheets employing N-methylated heterochiral amino acids. *Chemical Science* **2016,** *7* (8), 5212-5218.

48. Hoang, H. N.; Hill, T. A.; Ruiz-Gómez, G.; Diness, F.; Mason, J. M.; Wu, C.; Abbenante, G.; Shepherd, N. E.; Fairlie, D. P., Twists or turns: stabilising alpha vs. beta turns in tetrapeptides. *Chemical Science* **2019,** *10* (45), 10595-10600.

49. Sigler, G. F.; Fuller, W. D.; C., N., Formation of oligopeptides during the synthesis of 9-fluorenylmethyloxycarbonyl amino acid derivatives. *Biopolymers* **1983,** *22*, 2157-2162.

50. Freidinger, R. M.; Hinkle, J. S.; Perlow, D. S., Synthesis of 9-fluorenylmethyloxycarbonyl-protecte N-alkyl amino acids by reduction of oxazolidinones. *The Journal of Organic Chemistry* **1983,** *48*, 77-81.

1.14. Figures

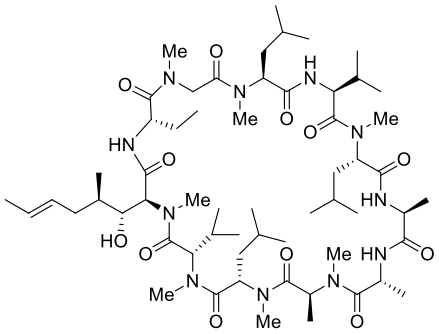


Figure 1-1. Cyclosporin A

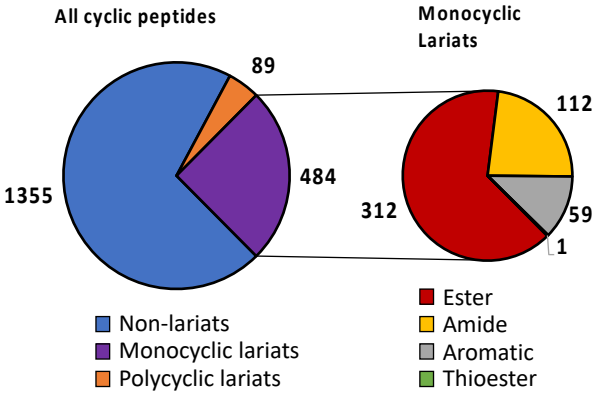


Figure 1-2. Natural Products Atlas analysis

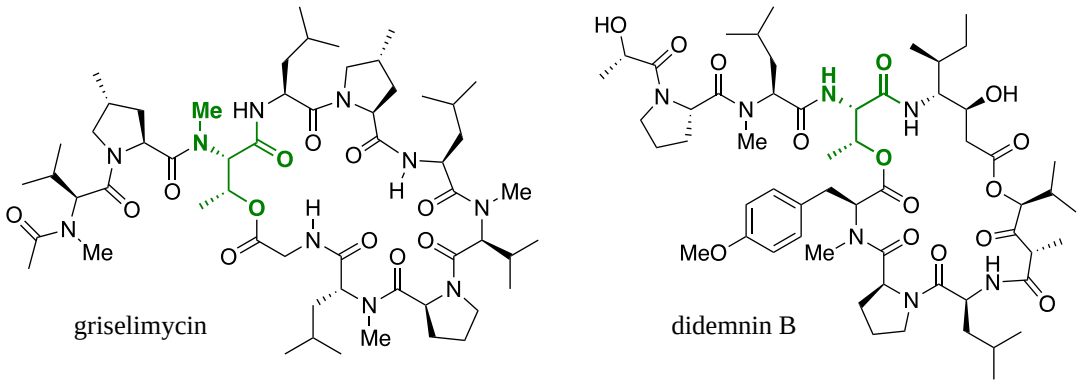


Figure 1-3. Lariat peptide natural products inspiring this investigation

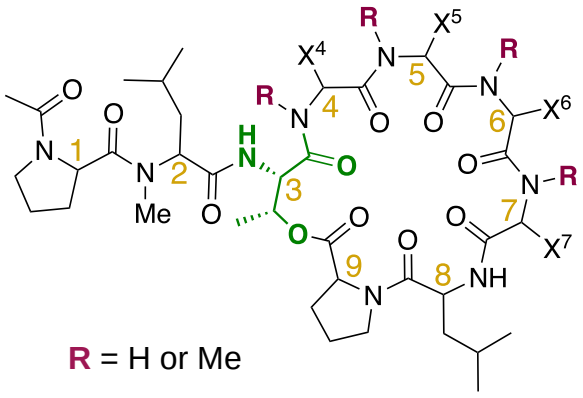


Figure 1-4. Design of Library 1

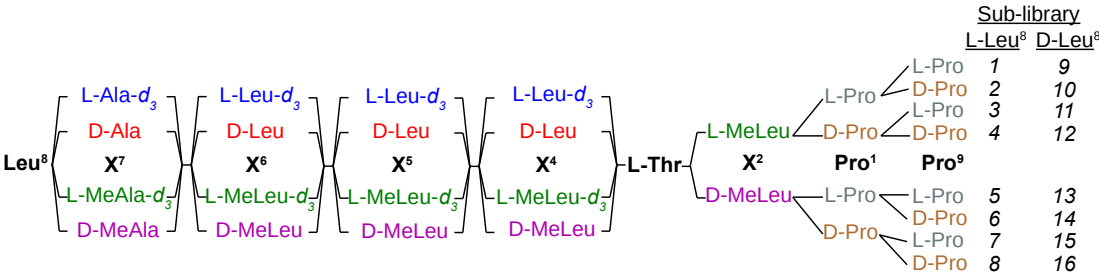


Figure 1-5. Split-pool library preparation of Library 1

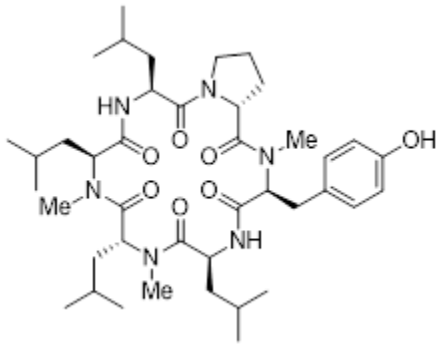


Figure 1-6. 1NMe3 used as a permeability benchmark

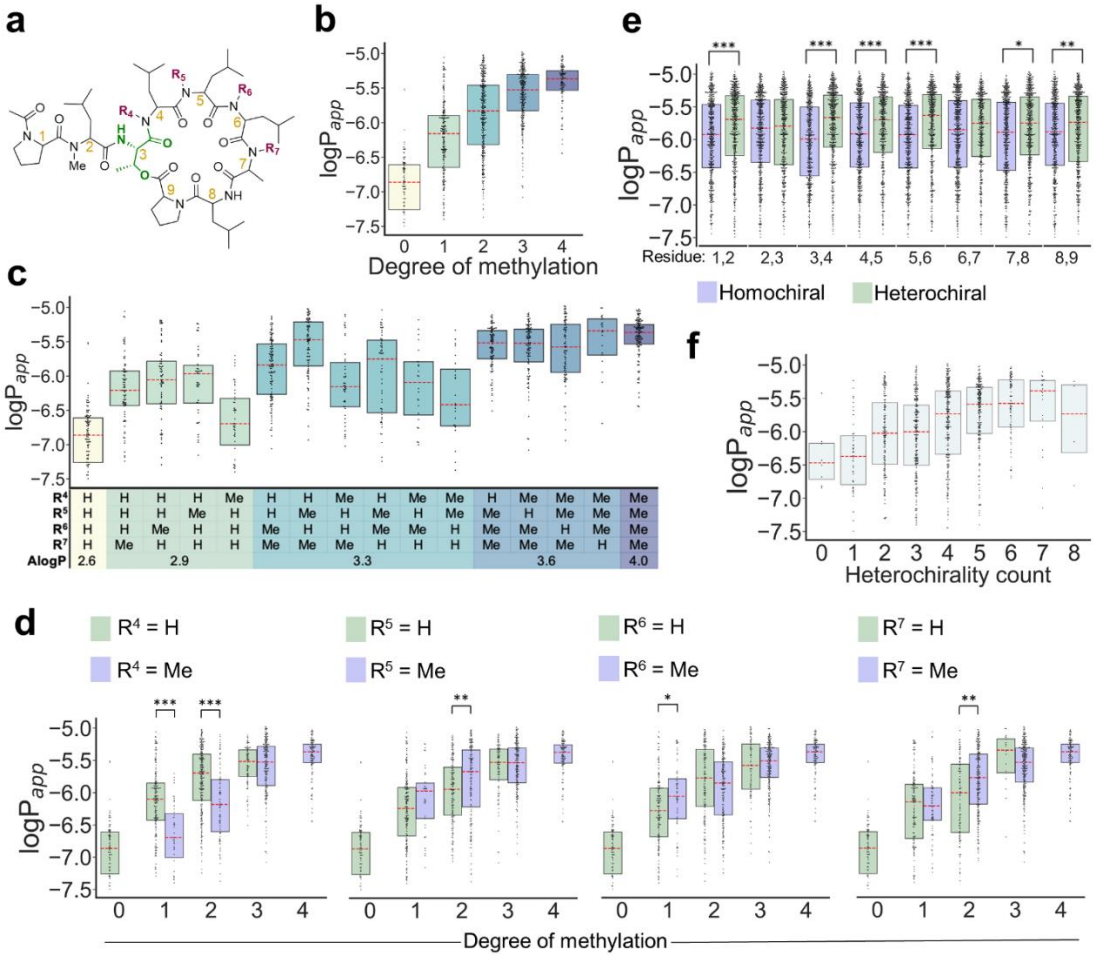


Figure 1-7. Effect of structural features on Library 1 permeability

(a) Skeletal structure of Library 1. (b) Effect of number of N-methyl groups on permeability. (c) Effect of N-methyl position on permeability. (d) Effect of methylation on permeability for each variable position. (e) Higher number of heterochiral relationships between adjacent residues is associated with higher permeability. (f) Effect of relative stereochemistry between adjacent residues on permeability. The red dashed lines represent medians and the boxes represent

quartiles. Statistics are as follows: Mann-Whitney U test; ***$P < 0.0001$, **$P < 0.001$, *$P < 0.01$.

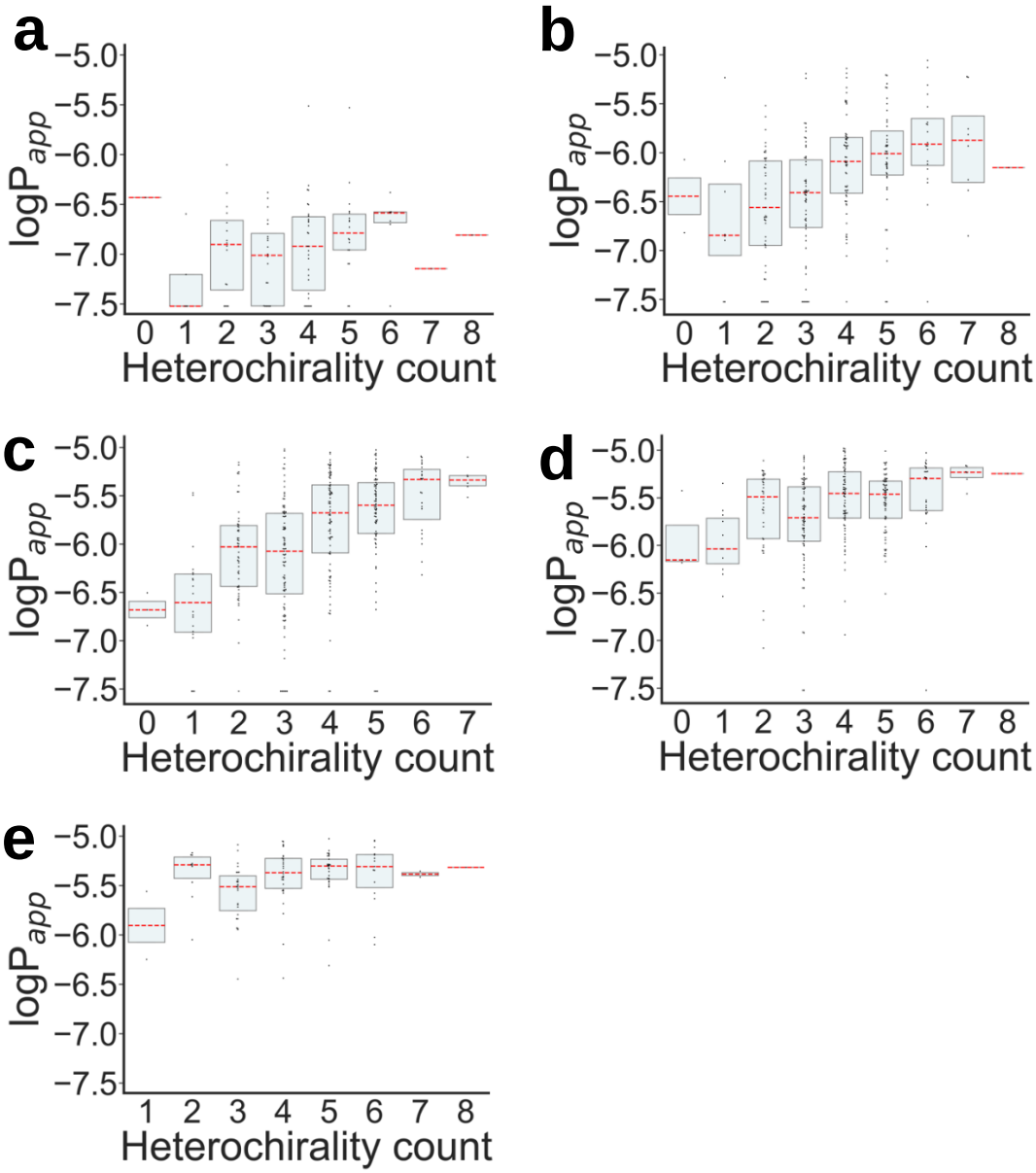


Figure 1-8. Effect of heterochirality on permeability by degree of N-methylation

(a) 0 N-methyl groups in the macrocycle, (b) 1 N-methyl group in the macrocycle, (c) 2 N-methyl groups in the macrocycle, (d) 3 N-methyl groups in the macrocycle, (e) 4 N-methyl groups in the macrocycle. The red dashed lines represent medians and the boxes represent quartiles. LogP$_{app}$ values below -7.5 were not included in the swarm plot but were used to calculate median and quartiles.

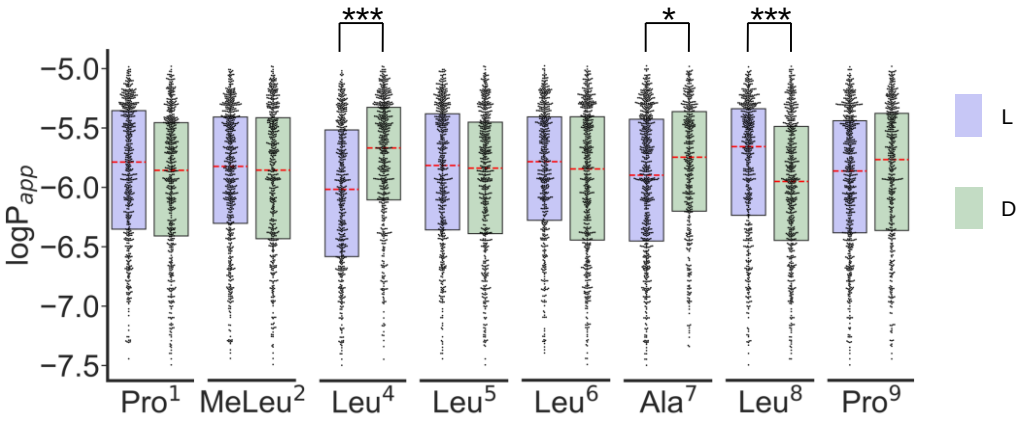


Figure 1-9. Effect of stereochemistry at each position on permeability of compounds in Library 1

The red dashed lines represent medians and the boxes represent quartiles. Statistics are as follows: Mann-Whitney U test; ***$P < 0.0001$, **$P < 0.001$, *$P < 0.01$. LogP$_{app}$ values below -7.5 were not included in the swarm plot but were used to calculate median and quartiles.

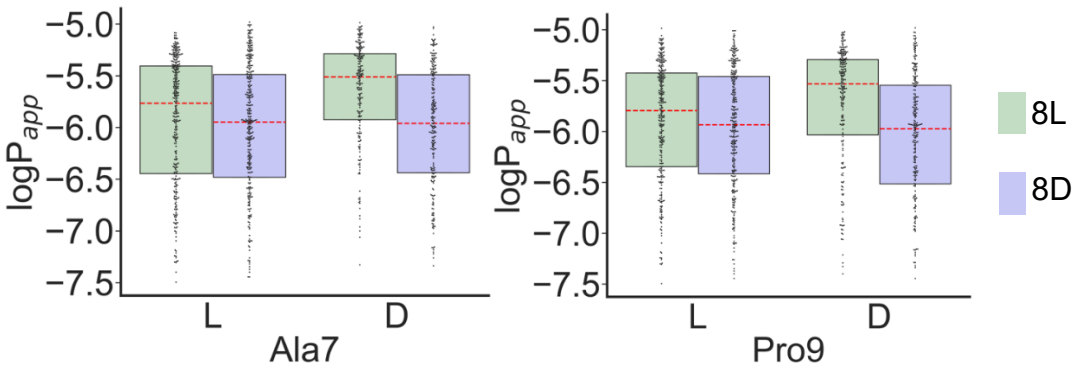


Figure 1-10. Effect of relative stereochemistry between Leu8 and adjacent residues on permeability

The red dashed lines represent medians and the boxes represent quartiles. LogP*app* values below -7.5 were not included in the swarm plot but were used to calculate median and quartiles.

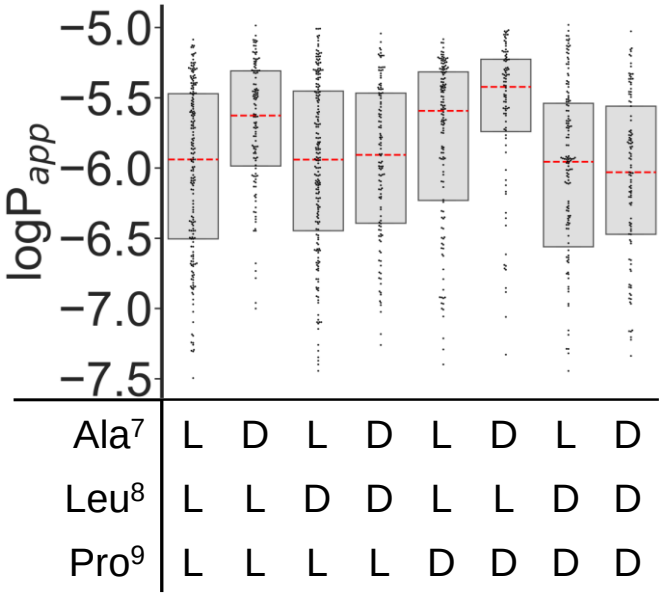


Figure 1-11. Effect of stereochemistry at residues 7-9

Considering the stereochemistry of residues 7-9, three of eight configurations have higher permeability than the others. All three of these have Leu8=L and at least one instance of heterochirality with an adjacent residue. The red dashed lines represent medians and the boxes represent quartiles. LogP_{app} values below -7.5 were not included in the swarm plot but were used to calculate median and quartiles.

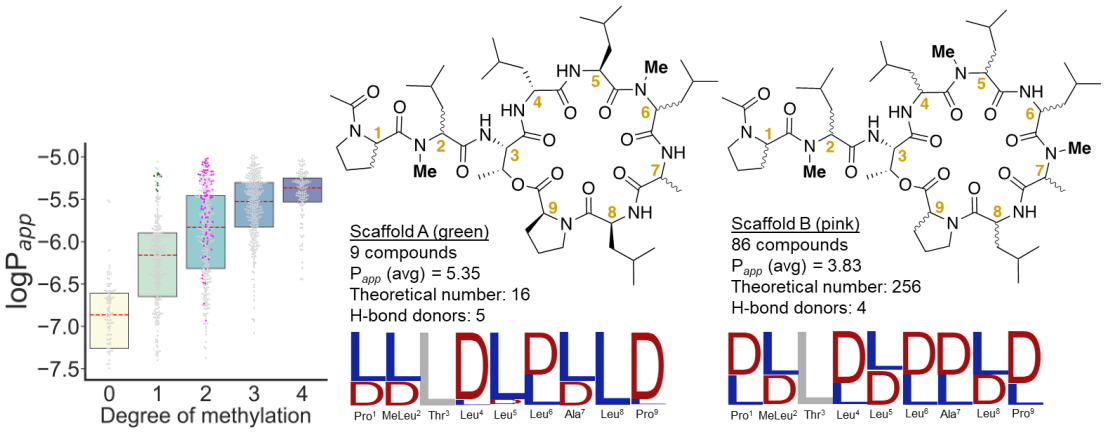


Figure 1-12. Scaffolds derived from the stereochemical consensus of permeable compounds with specific N-methylation patterns

(a) The compounds of scaffolds A (green) and B (magenta) relative to the rest of the library. (b) The structures of the scaffolds with sequence logos. The relative height of the letter symbolizes the representation of that stereochemistry among compounds with specific N-methylation patterns from among the top 25 most permeable compounds with that degree of N-methylation. The number of compounds contributing to the consensus is listed along with the average permeability in units of 10^{-6} cm/s.

Cmpd	Pro[1]	MeLeu[2]	Leu[4]	Leu[5]	Leu[6]	Ala[7]	Leu[8]	Pro[9]	P_{app} (10^{-6} cm/s)		
									Library	Individual	MDCK
1	D	D	D-Me	L-Me	L-Me	L-H	D	L	9.8	2.2 ± 0.8	7.6 ± 0.7
2	L	D	D-H	L-H	D-H	D-Me	L	L	8.8	2.2 ± 0.6	13.8 ± 1.2
3	L	L	D-Me	L-Me	L-H	L-Me	D	D	8.7	4.9 ± 0.8	26.6 ± 4.2
4	L	D	D-H	L-Me	D-H	D-Me	L	D	7.9	2.2 ± 0.6	10.8 ± 0.5
5	L	D	L-Me	L-H	D-Me	D-Me	L	D	4.7	2.1 ± 0.7	4.8 ± 0.6
6	L	D	D-H	L-Me	D-H	D-H	L	L	1.2	0.4 ± 0.1	0.9 ± 0.1
7	L	L	L-Me	L-H	L-Me	L-H	D	D	0.1	0.1 ± 0.0	0.5 ± 0.1
8	D	L	D-H	D-H	L-H	L-H	L	D	0.1	0.1 ± 0.0	0.5 ± 0.2
9	D	D	D-H	D-H	L-H	L-H	L	D	0.0	0.0 ± 0.0	0.4 ± 0.1

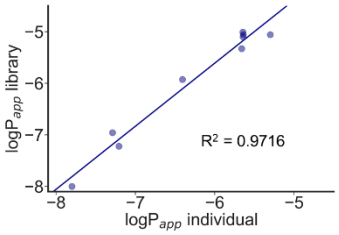


Figure 1-13. Resynthesized compound permeability

These compounds were individually synthesized and tested for PAMPA permeabilities. Permeabilities of compounds assayed individually in PAMPA are plotted against permeabilities from the library. The relationship is linear (R^2 = 0.9716).

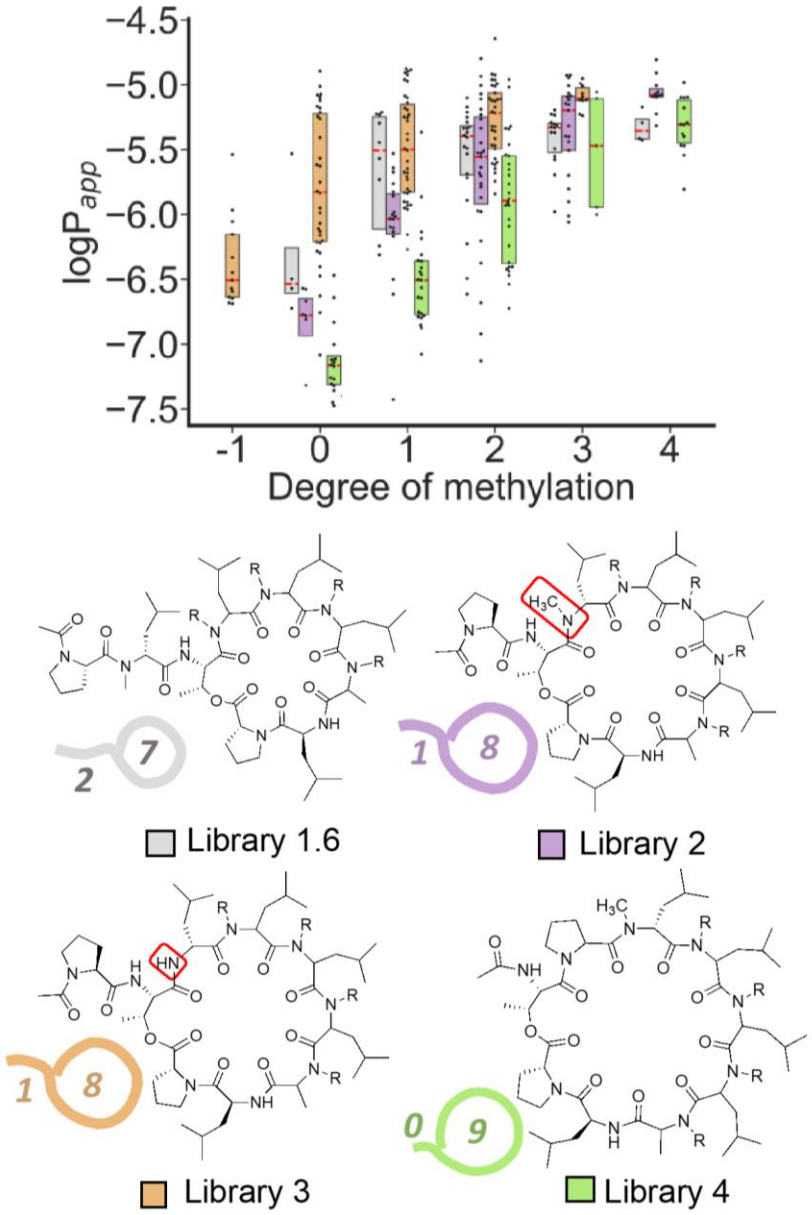


Figure 1-14. Effect of threonine position on permeability

The molecular weight and AlogP are constant between these libraries. Library 3 has one fewer N-methyl group than the other three libraries. To maintain consistency with library 1, degree of methylation refers to the total number of N-methyl groups

minus one. Within each degree of methylation, all compounds are isomeric. The red dashed lines represent medians and the boxes represent quartiles.

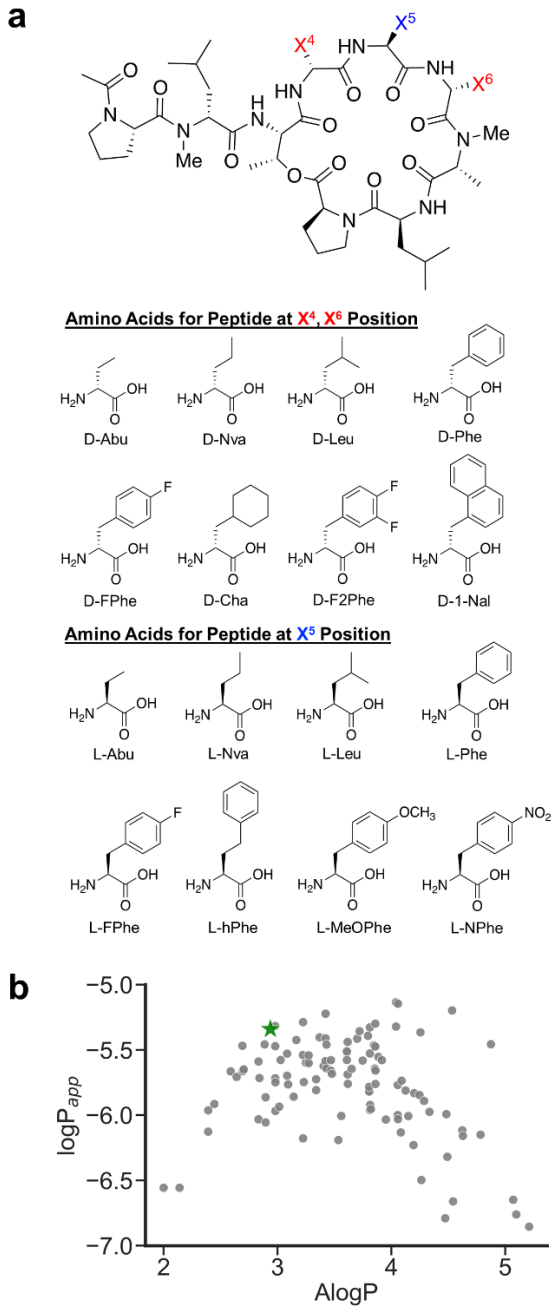


Figure 1-15. Library 5

(a) Design of Library 5. (b) logP_{app} values plotted against AlogP values. Compound **2** is indicated by a green star.

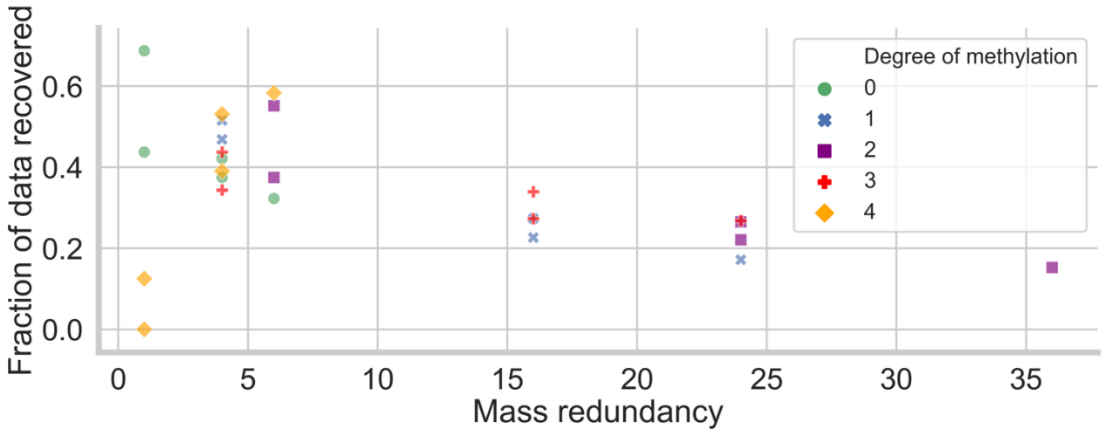
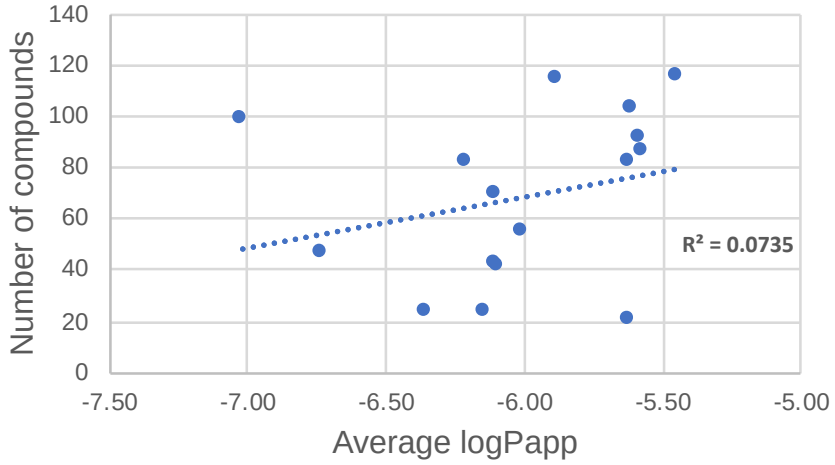


Figure 1-16. Recovery of data vs mass redundancy in Library 1

Each point represents a particular mass. Fraction of data recovered = (number of identified compounds in Library 1 dataset with a given mass) / (theoretical number of compounds with that mass). The list of masses may be found in Table 1-9.

a



b

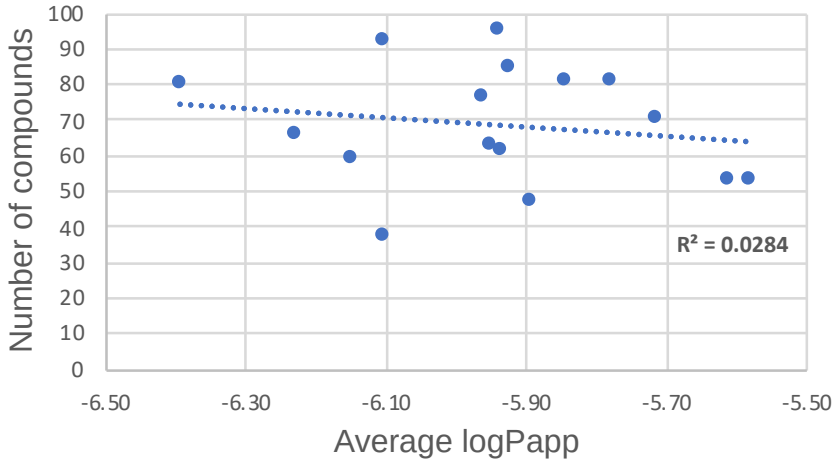


Figure 1-17. Number of identified compounds in data set vs. average $logP_{app}$ from each N-methylation pattern and sub-library

(a) methylation pattern and (b) sub-library

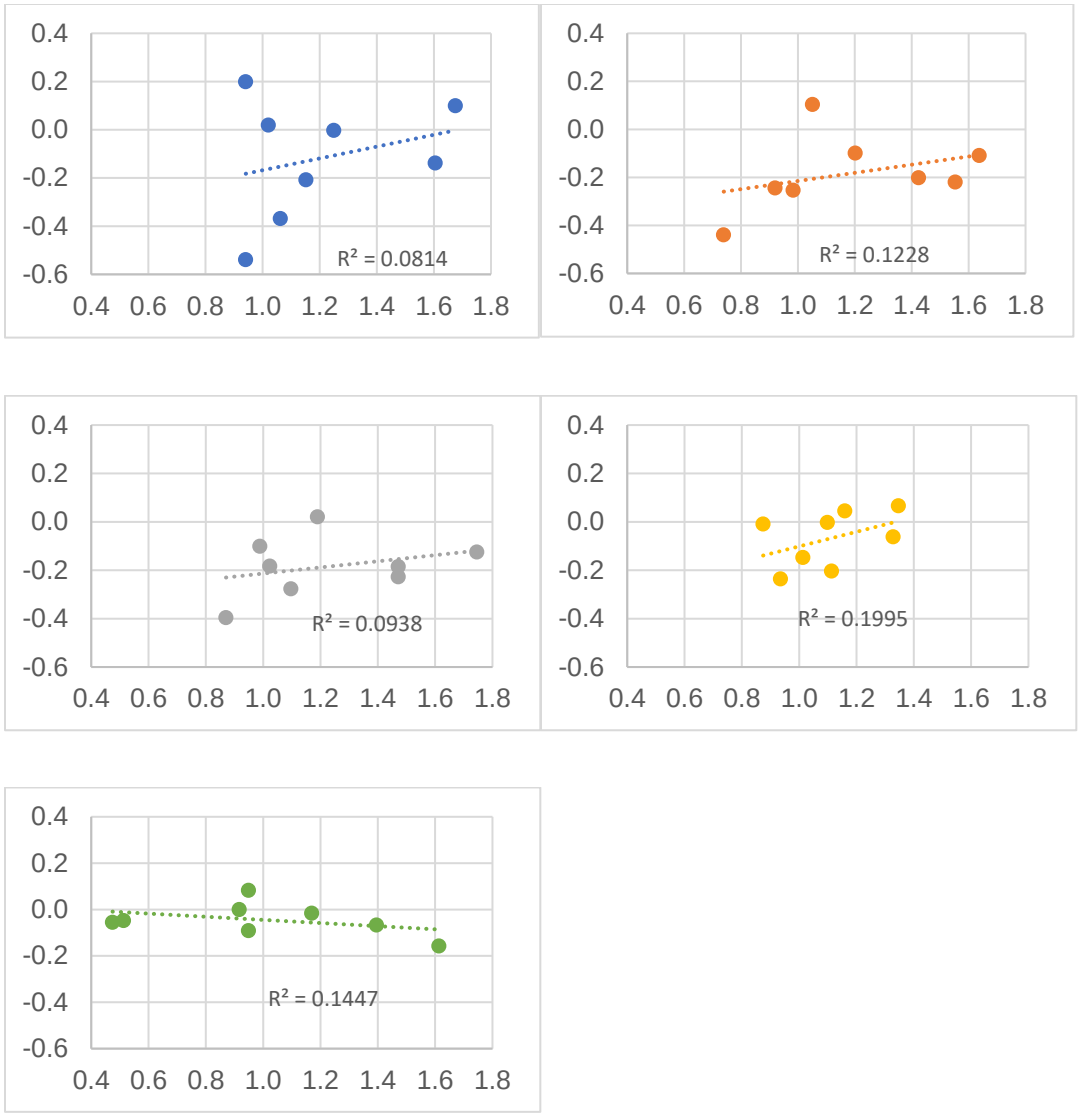


Figure 1-18. Correlation between permeability of compounds with heterochiral stereochemistry and number of compounds with heterochiral stereochemistry

Each point represents an adjacent stereochemical pair for a given number of N-methyl groups. The y-axis values correspond to the values in Table 1-4 (permeability effect). The x-axis values correspond to the values in Table 1-10 (representation disparity). Blue: 0 N-methyl groups in the macrocycle, orange: 1 N-methyl group in the

macrocycle, gray: 2 N-methyl groups in the macrocycle, yellow: 3 N-methyl groups in the macrocycle, green: 4 N-methyl groups in the macrocycle.

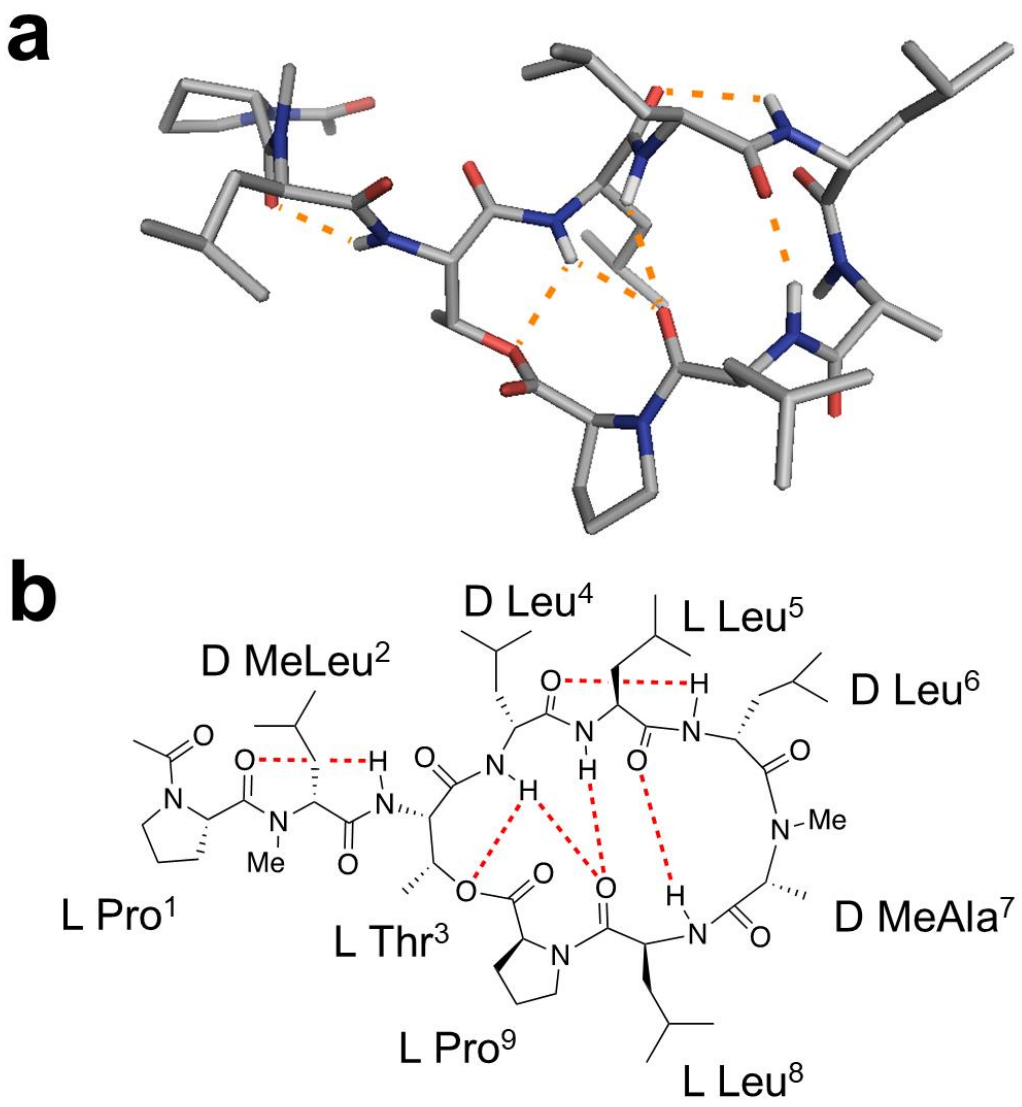


Figure 1-19. 3D structure of compound **2**

(a) 3D representation of the low-dielectric solution structure of compound **2.** (b) Schematic structure of **2** showing hydrogen bonding observed in low-dielectric solution structure.

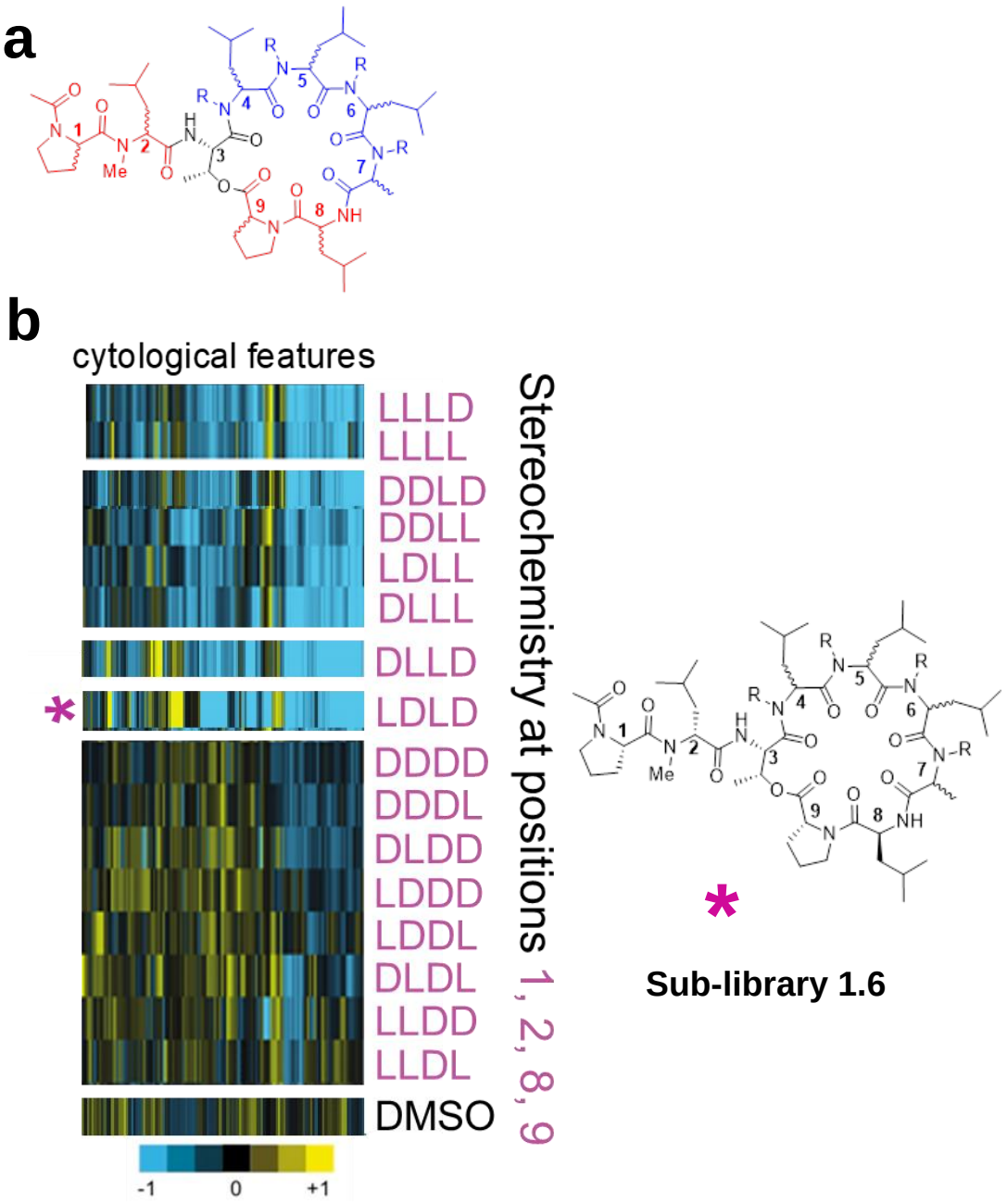


Figure 1-20. Cytological fingerprint of the sixteen Library 1 sub-libraries

(a) In each mixture, the blue residues are undefined while the red residues are defined. (b) Each row depicts the cytological fingerprint of a sub-library. The strongest bioactivity was observed for sub-library 1.6 (“LDLD”).

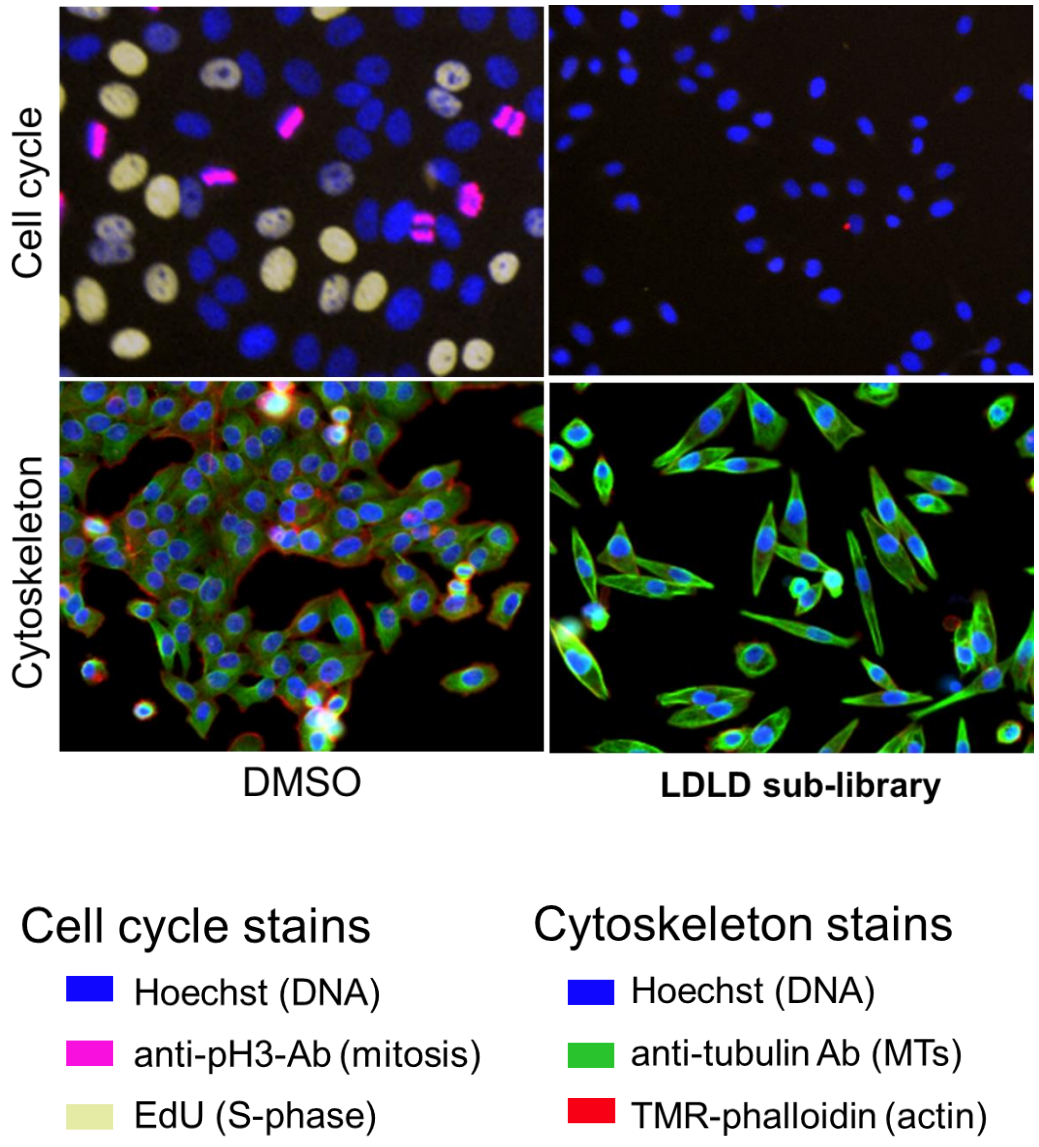


Figure 1-21. Images of cells treated with sub-library 6

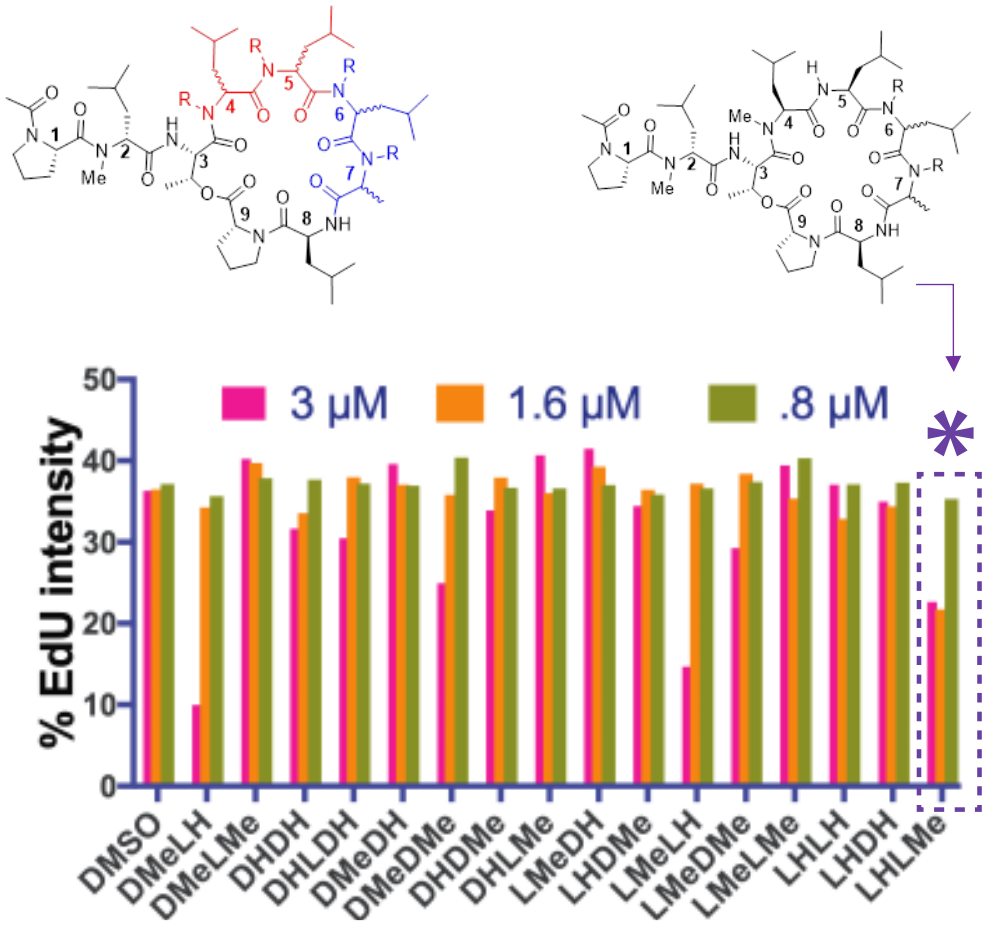


Figure 1-22. Effect of sub-library 6 subdivisions on EdU intensity

Subdivisions are labelled in the order: residue 5, residue 4. For example, the subdivision with the strongest EdU suppression, LHLMe has L-Leu at position 5 and L-MeLeu at position 4. The residues colored blue in the top left structure are undefined while the red residues are defined by the code indicated below the chart.

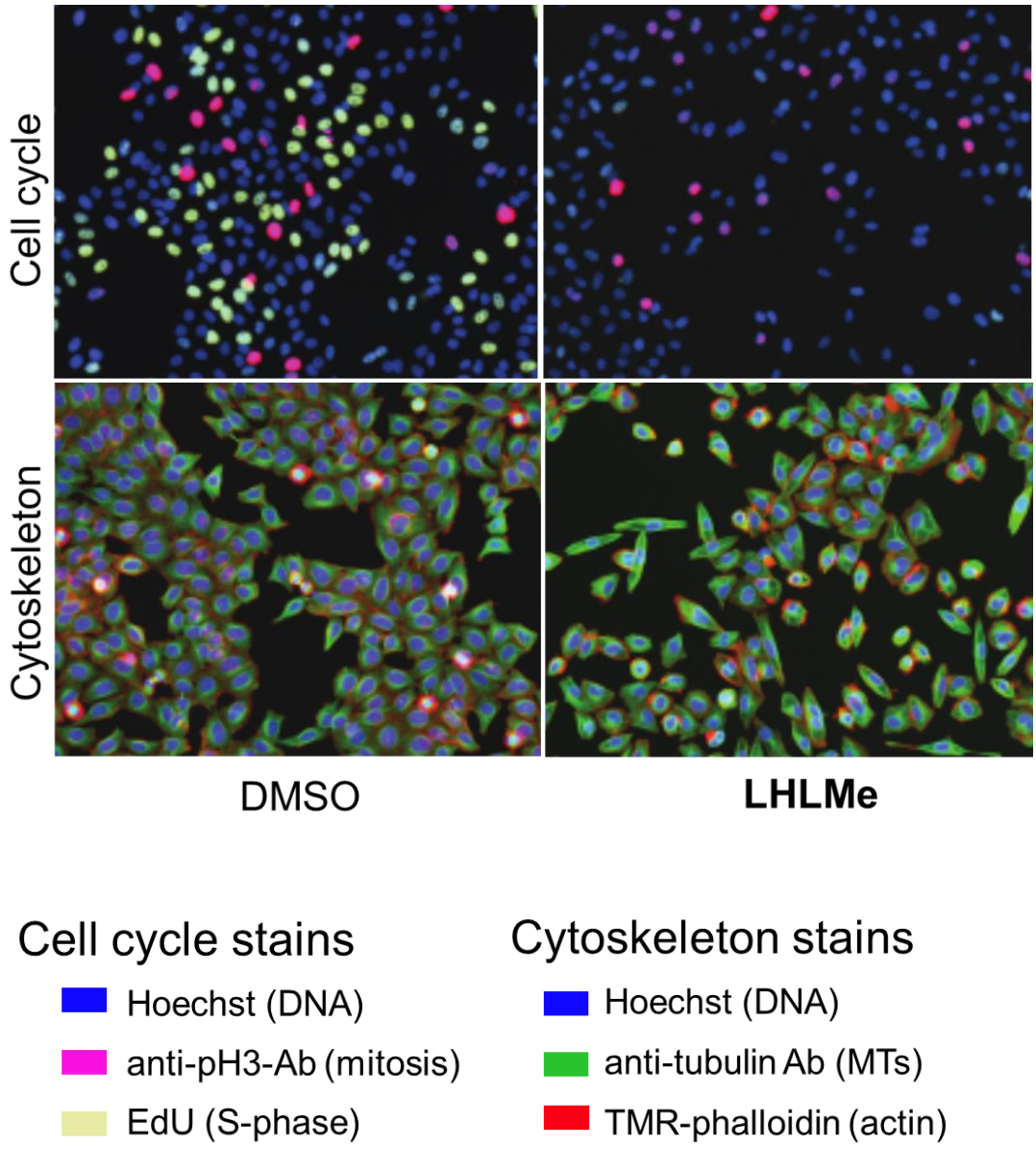


Figure 1-23. Images of cells treated with LHLMe subdivision of sub-library 6

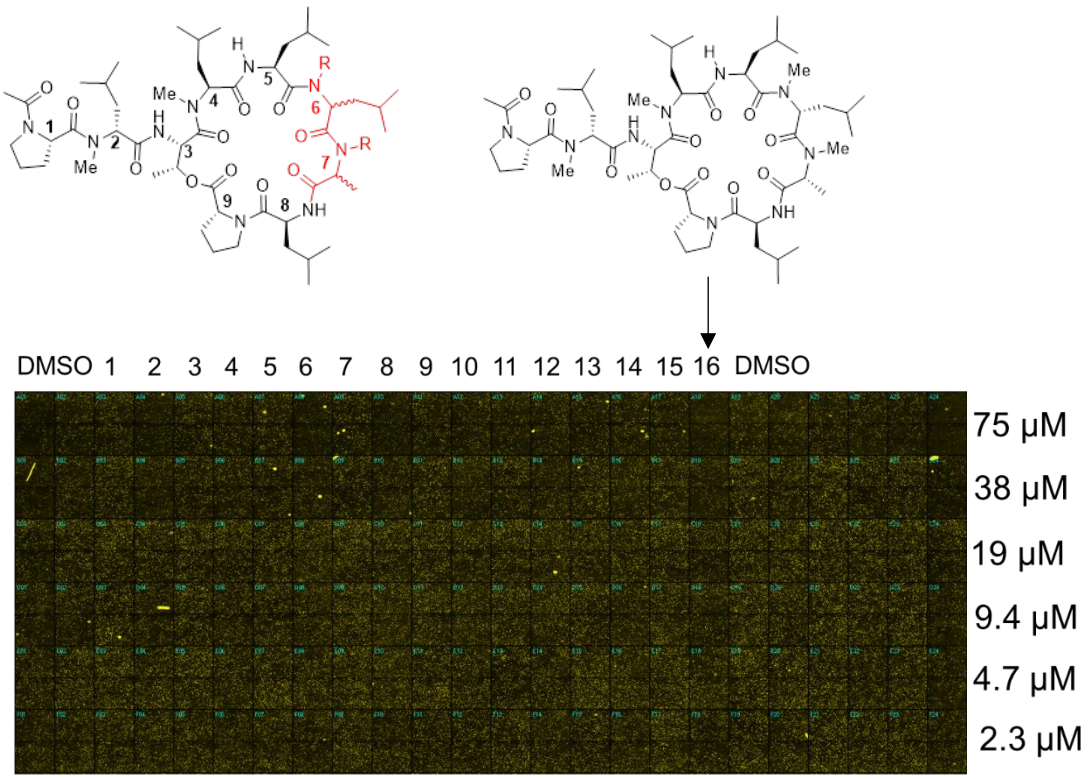


Figure 1-24. Cell images showing EdU intensity for the sixteen individually synthesized compounds from the LDLD-LHLMe subdivision

Compound 16 showed the strongest suppression of EdU, although the effect is not strong. EdU suppression is visually discernable down to around 10 µM.

1.15. Tables

Table 1-1. Key characteristics of Library 1 (table) (MW, AlogP, # H-bond donors)

Degree of Methylation (#R = Me)	MW (non-deuterated)	H-bond donors	Number of compounds (expected)	AlogP
0	987.64	6	256	2.59
1	1001.65	5	1024	2.94
2	1015.67	4	1536	3.28
3	1029.67	3	1024	3.62
4	1043.70	2	256	3.96

Table 1-2. Permeability of 1NMe3 standard in each Library 1 sub-library

Library	P_{app} (10^{-6} cm/s) ± SD	Pro[1]	MeLeu[2]	Leu[8]	Pro[9]	1/2	8/9
1.1	10.27	L	L	L	L	Hom.	Hom.
1.2	10.17	L	L	L	D	Hom.	Het.
1.3	7.74	D	L	L	L	Het.	Hom.
1.4	8.93	D	L	L	D	Het.	Het.
1.5	7.42	L	D	L	L	Het.	Hom.
1.6	6.31	L	D	L	D	Het.	Het.
1.7	6.63	D	D	L	L	Hom.	Hom.
1.8	7.44	D	D	L	D	Hom.	Het.
1.9	11.24	L	L	D	L	Hom.	Het.
1.1	6.75	L	L	D	D	Hom.	Hom.
1.11	6.79	D	L	D	L	Het.	Het.
1.12	6.61	D	L	D	D	Het.	Hom.
1.13	7.94	L	D	D	L	Het.	Het.
1.14	8.09	L	D	D	D	Het.	Hom.
1.15	7.59	D	D	D	L	Hom.	Het.
1.16	10.10	D	D	D	D	Hom.	Hom.
Average	**8.13 ± 1.55**						
Avg L		8.52	8.56	8.11	8.20		
Avg D		7.73	7.69	8.14	8.05		
Avg hom.						8.77	7.95
Avg het.						7.48	8.30
Ratio		**1.10**	**1.11**	**1.00**	**1.02**	**1.17**	**0.96**

Hom. = homochiral, Het. = heterochiral

Table 1-3. Basic permeability statistics for Library 1

	# of P_{app} > 5 × 10^{-6} cm/s	Mean $logP_{app}$	Median $logP_{app}$	Variance of $logP_{app}$
Full library	181 (16%)	-5.95	-5.83	0.47
DoM = 0	0 (0%)	-7.02	-6.86	0.45
DoM = 1	13 (5%)	-6.26	-6.16	0.37
DoM = 2	54 (15%)	-5.90	-5.83	0.32
DoM = 3	72 (24%)	-5.61	-5.53	0.19
DoM = 4	42 (37%)	-5.44	-5.37	0.09

DoM = degree of methylation

Table 1-4. Effect of heterochirality for each degree of N-methylation on permeability

Degree of methylation	Pro^1/ $MeLeu^2$	$MeLeu^2$/Thr^3	Thr^3/ Leu^4	Leu^4/ Leu^5	Leu^5/ Leu^6	Leu^6/ Ala^7	Ala^7/ Leu^8	Leu^8/ Pro^9
0	-0.539	0.019	-0.003	-0.208	-0.138	0.099	0.199	-0.369
1	-0.244	0.104	-0.440	-0.202	-0.220	-0.109	-0.099	-0.254
2	-0.277	0.021	-0.396	-0.227	-0.185	-0.125	-0.101	-0.183
3	-0.203	-0.062	-0.236	-0.003	-0.148	0.066	-0.009	0.045
4	-0.092	-0.067	-0.158	-0.048	-0.056	0.082	-0.016	-0.001
All	-0.233	0.048	-0.282	-0.195	-0.265	-0.074	-0.051	-0.128

The numbers correspond to the differences between median log(P_{app}) for homochiral vs. heterochiral, with positive numbers indicating higher log(P_{app}) for homochiral stereochemistry.

Table 1-5. Effect of stereochemistry for each degree of N-methylation on permeability

Degree of methylation	Pro[1]	MeLeu[2]	Thr[3]	Leu[4]	Leu[5]	Leu[6]	Ala[7]	Leu[8]	Pro[9]
0	0.134	0.019	N/A	-0.003	0.013	0.083	-0.021	0.274	0.507
1	0.112	0.104	N/A	-0.440	0.138	-0.021	-0.006	0.319	0.078
2	0.122	0.021	N/A	-0.396	0.033	-0.096	-0.116	0.166	-0.047
3	-0.079	-0.062	N/A	-0.236	-0.035	-0.010	-0.053	0.207	-0.088
4	0.056	-0.067	N/A	-0.158	-0.038	0.061	-0.078	-0.027	0.029
All	0.077	0.048	N/A	-0.282	0.039	0.080	-0.161	0.224	0.015

The numbers correspond to the differences between median log(P_{app}) for L stereochemistry vs. D stereochemistry, with positive numbers indicating higher log(P_{app}) for L stereochemistry.

Table 1-6. Relative number of identified compounds in the library with L vs. D stereochemistry at each stereocenter

Degree of methylation	Pro[1]	MeLeu[2]	Thr[3]	Leu[4]	Leu[5]	Leu[6]	Ala[7]	Leu[8]	Pro[9]
0	0.980	1.020	N/A	1.250	0.707	0.547	2.667	0.833	1.250
1	0.935	1.051	N/A	0.739	1.087	0.655	1.857	0.805	1.500
2	1.000	1.190	N/A	0.870	1.097	0.912	1.602	1.023	1.247
3	1.000	1.328	N/A	0.935	0.987	1.328	1.275	0.961	1.129
4	1.231	1.367	N/A	1.636	0.731	0.902	1.071	0.933	1.522
All	1.005	1.194	N/A	0.945	0.980	0.898	1.550	0.928	1.290

The numbers correspond to the ratio of the number of compounds with L stereochemistry to the number of compounds with D stereochemistry. Larger numbers indicate a preponderance of compounds with L stereochemistry at that stereocenter for library members with the specified number of N-methyl groups in the macrocycle.

Table 1-7. LCMS conditions for library analysis

Liquid chromatography	Thermo Scientific Dionex UltiMate system (RS pump, RS autosampler, RS column compartment)
Mass spectrometer	Thermo Scientific Velos Pro mass spectrometer
Eluent	H_2O/ACN containing 0.1% FA (all solvents were Optima® grade)
Column	Thermo Scientific Acclaim™ RSLC 120 C18 (2.2µm 120Å, 2.1 X 250 mm) (product# 074812, serial# 001101)
Flow rate	0.4mL/min
Temperature	50°C
Gradient for Libraries 1-4 (%ACN)	0-45min: 52-67, 45-55min: 100, 55-60min: 52
Gradient for Library 5 (%ACN)	0-45min: 50-80, 45-55min: 100, 55-60min: 50
Source ionization voltage	80V
MS1	FTMS
MS2	ITMS
MS^n activation	Collision-induced dissociation
Normalized collision energy	35.0
Isolation width	2.0 m/z
Activation Q.	0.250
Activation time	10.0 ms

Table 1-8. Data on individual compounds

Number	MW	AlogP	Rt (library)	Rt (resynthesized)	P_{app} (library)	P_{app} (resynthesized)	P_{app} (MDCK)	Purity (UV 200nm)
1	1029.68	3.62	38.62	38.51	9.81	2.26	7.6	95%
2	1002.31	2.94	32.45	32.11	8.77	2.29	13.8	95%
3	1029.68	3.62	42.37	42.23	8.72	5.02	26.6	96%
4	1016.34	3.28	37.17	36.94	7.92	2.28	10.8	96% (94%*)
5	1029.68	3.62	29.52	29.53	4.67	2.19	4.8	91%
6	1002.31	2.94	19.92	19.8	1.18	0.39	0.9	99%
7	1016.34	3.28	13.62	13.83	0.11	0.05	0.5	99%
8	988.28	2.59	12.51	12.57	0.06	0.06	0.5	87%
9	988.28	2.59	10.57	10.58	0.01	0.02	0.4	97%
10	1016.34	3.28	**17.33**	**22.38**	NA	0.06	NA	85%*
11	988.28	2.59	**20.72**	**13.37**	NA	0.03	NA	93%*

P_{app} values are in units of 10^{-6} cm/s. Retention times differing between library and resynthesized are indicated in red.

*Purities based on the total ion chromatogram. For compounds **10** and **11**, UV data was not collected.

Table 1-9. Representation of each molecular weight in Library 1

Degree of methylation	MW	Theoretical number	Total number in data set	Efficiency of data recovery (% of theoretical)
0	987.64	16	7	44
0	990.66	64	27	42
0	993.67	96	30	31
0	996.69	64	24	38
0	999.71	16	11	69
1	1001.65	64	33	52
1	1004.67	256	54	21
1	1007.69	384	63	16
1	1010.71	256	62	24
1	1013.73	64	28	44
2	1015.67	96	30	31
2	1018.69	384	97	25
2	1021.71	576	85	15
2	1024.72	384	83	22
2	1027.74	96	51	53
3	1029.68	64	21	33
3	1032.70	256	66	26
3	1035.72	384	98	26
3	1038.74	256	85	33
3	1041.76	64	28	44
4	1043.70	16	2	13
4	1046.72	64	25	39
4	1049.74	96	56	58
4	1052.76	64	33	52
4	1055.78	16	0	0

Table 1-10. Relative number of identified compounds in the library with homochirality vs. heterochirality between each pair of adjacent residues

Degree of methylation	Pro[1]/ MeLeu[2]	MeLeu[2]/ Thr[3]	Thr[3]/ Leu[4]	Leu[4]/ Leu[5]	Leu[5]/ Leu[6]	Leu[6]/ Ala[7]	Ala[7]/ Leu[8]	Leu[8]/ Pro[9]
0	0.941	1.020	1.250	1.152	1.605	1.676	0.941	1.063
1	0.920	1.051	0.739	1.424	1.553	1.637	1.202	0.983
2	1.097	1.190	0.870	1.471	1.471	1.746	0.989	1.023
3	1.113	1.328	0.935	1.099	1.014	1.346	0.874	1.159
4	0.949	1.396	1.614	0.513	0.474	0.949	1.170	0.917
All	1.031	1.194	0.945	1.181	1.207	1.498	1.009	1.043

The numbers correspond to the ratio of the number of compounds with homochiral stereochemistry between two residues to the number of compounds with heterochiral stereochemistry between the two residues. Larger numbers indicate a preponderance of compounds with homochiral stereochemistry between those stereocenters for library members with the specified number of N-methyl groups in the macrocycle.

Table 1-11. Number of identified compounds in Library 1 dataset with each degree of methylation

Degree of methylation	Theoretical number	Number in dataset	Data recovery (% of theoretical)
0	256	99	39
1	1024	240	23
2	1536	346	23
3	1024	298	29
4	256	116	45

Table 1-12. Average retention time and retention time variance by degree of N-methylation

Degree of N-methylation	Average retention time (min)	Variance (min)
0	12.95	9.24
1	18.85	41.82
2	24.22	64.57
3	28.71	55.03
4	32.54	26.37

Table 1-13. Number of identified compounds in Library 1 dataset with each methylation pattern

Methylation pattern (Leu4, Leu5, Leu6, Ala7)	Number in dataset
NNNN	99
NNNY	82
NNYN	70
NYNN	41
YNNN	47
NNYY	115
NYNY	86
NYYN	55
YNNY	42
YNYN	24
YYNN	24
NYYY	92
YNYY	103
YYNY	82
YYYN	21
YYYY	116

Theoretical number is 256 in all cases.

Table 1-14. Number of identified compounds in Library 1 dataset from each sub-library

Sub-library	Number of compounds in dataset
1	95
2	53
3	81
4	81
5	70
6	53
7	59
8	37
9	76
10	80
11	85
12	47
13	61
14	63
15	92
16	66

Theoretical number is 256 in all cases.

Table 1-15. Effect of heterochirality for each degree of N-methylation on permeability

Degree of methylation	Pro[1]/ MeLeu[2]	MeLeu[2]/ Thr[3]	Thr[3]/ Leu[4]	Leu[4]/ Leu[5]	Leu[5]/ Leu[6]	Leu[6]/ Ala[7]	Ala[7]/ Leu[8]	Leu[8]/ Pro[9]
0	-0.539	0.019	-0.003	-0.208	-0.138	0.099	0.199	-0.369
1	-0.244	0.104	-0.440	-0.202	-0.220	-0.109	-0.099	-0.254
2	-0.277	0.021	-0.396	-0.227	-0.185	-0.125	-0.101	-0.183
3	-0.203	-0.062	-0.236	-0.003	-0.148	0.066	-0.009	0.045
4	-0.092	-0.067	-0.158	-0.048	-0.056	0.082	-0.016	-0.001
All	-0.233	0.048	-0.282	-0.195	-0.265	-0.074	-0.051	-0.128

The numbers correspond to the differences between median log(P_{app}) for homochiral vs. heterochiral, with positive numbers indicating higher log(P_{app}) for homochiral stereochemistry.

Table 1-16. Number of identified compounds in the Library 1 dataset with L and D stereochemistry at each stereocenter

Stereocenter	L	D
Pro^{1}	551	548
$MeLeu^{2}$	598	501
Thr^{3}	1099	0
Leu^{4}	534	565
Leu^{5}	544	555
Leu^{6}	520	579
Ala^{7}	668	431
Leu^{8}	529	570
Pro^{9}	619	480

Table 1-17. Number of identified compounds in the Library 1 dataset with homochiral and heterochiral diastereochemistry for each adjacent residue pair

Stereocenters	Homochiral	Heterochiral
$Pro^{1}/MeLeu^{2}$	558	541
$MeLeu^{2}/Thr^{3}$	598	501
Thr^{3}/Leu^{4}	534	565
Leu^{4}/Leu^{5}	595	504
Leu^{5}/Leu^{6}	601	498
Leu^{6}/Ala^{7}	659	440
Ala^{7}/Leu^{8}	552	547
Leu^{8}/Pro^{9}	561	538

Table 1-18. Representation of adjacent stereochemical configurations

	L,L	L,D	D,L	D,D
Pro[1]/MeLeu[2]	1.11	0.90	1.07	0.92
MeLeu[2]/Thr[3]	1.09		0.91	
Thr[3]/Leu[4]	0.97	1.03		
Leu[4]/Leu[5]	1.04	0.90	0.94	1.12
Leu[5]/Leu[6]	1.03	0.95	0.86	1.16
Leu[6]/Ala[7]	1.36	0.53	1.07	1.04
Ala[7]/Leu[8]	1.18	1.25	0.74	0.83
Leu[8]/Pro[9]	1.11	0.82	1.14	0.93

The values are calculated as (# library members found / # library members expected). In cases involving Thr[3], the expected number of library members was equal to half the total number of library members in Library 1. Otherwise, the expected number of library members was one quarter of the total library members in Library 1.

Table 1-19. Average permeability of adjacent stereochemical configurations

	L,L	L,D	D,L	D,D
Pro[1]/MeLeu[2]	-6.00	-5.81	-5.86	-6.15
MeLeu[2]/Thr[3]	-5.93		-5.98	
Thr[3]/Leu[4]	-6.10	-5.82		
Leu[4]/Leu[5]	-6.16	-6.02	-5.68	-5.93
Leu[5]/Leu[6]	-6.02	-5.84	-5.78	-6.12
Leu[6]/Ala[7]	-6.01	-5.67	-6.03	-5.95
Ala[7]/Leu[8]	-5.95	-6.08	-5.66	-6.03
Leu[8]/Pro[9]	-5.90	-5.75	-5.99	-6.15

The values are mean log(P_{app}) for each configuration.

Table 1-20. Representation of each diastereomer for Leu^6/Ala^7, Ala^7/Leu^8, and Leu^8/Pro^9 at each degree of N-methylation

Leu^6/Ala^7	L,L	L,D	D,L	D,D
All	374	146	294	285
0 Me	1.04	0.00	1.40	1.05
1 Me	0.98	0.47	1.18	1.11
2 Me	1.07	0.85	0.94	1.05
3 Me	1.04	1.64	0.78	0.85
4 Me	0.71	1.75	1.03	0.96

Ala^7/Leu^8	L,L	L,D	D,L	D,D
All	325	343	204	227
0 Me	1.13	1.26	0.65	0.73
1 Me	1.08	1.05	0.67	1.09
2 Me	1.05	0.98	1.06	0.91
3 Me	0.87	0.97	1.25	1.01
4 Me	0.90	0.80	1.16	1.29

Leu^8/Pro^9	L,L	L,D	D,L	D,D
All	305	224	314	256
0 Me	0.95	0.94	1.03	1.08
1 Me	0.98	0.86	1.15	0.97
2 Me	1.02	1.09	0.95	0.96
3 Me	1.00	1.04	0.88	1.11
4 Me	1.03	0.97	1.12	0.85

Each field contains the number of library members with the indicated stereochemistry at each degree of N-methylation. The color corresponds to the number of library

members found relative to the number of library members expected at each degree of N-methylation based on the total number of library members with the indicated stereochemistry. Green indicates higher than expected abundance while red indicates lower.

Table 1-21. Permeability deviation for each diastereomer of Leu^6/Ala^7, Ala^7/Leu^8, and Leu^8/Pro^9 at each degree of N-methylation

Leu^6/Ala^7	L,L	L,D	D,L	D,D
All	-0.05	0.29	-0.08	0.00
0 Me	0.05		-0.06	0.02
1 Me	-0.05	0.18	0.05	-0.03
2 Me	-0.11	0.14	0.05	0.04
3 Me	0.00	-0.01	-0.06	0.07
4 Me	0.03	0.03	-0.10	0.05

Ala^7/Leu^8	L,L	L,D	D,L	D,D
All	0.01	-0.13	0.29	-0.07
0 Me	0.17	-0.15	0.09	-0.05
1 Me	0.08	-0.08	0.42	-0.23
2 Me	-0.02	-0.07	0.24	-0.10
3 Me	0.07	-0.10	0.14	-0.10
4 Me	-0.06	-0.02	0.04	0.04

Leu^8/Pro^9	L,L	L,D	D,L	D,D
All	0.05	0.21	-0.04	-0.20
0 Me	0.19	0.09	0.25	-0.56
1 Me	0.06	0.35	0.01	-0.36
2 Me	-0.03	0.22	-0.01	-0.17
3 Me	0.07	0.15	-0.17	-0.04
4 Me	0.00	-0.03	0.02	0.00

Each field contains the difference between the average permeability ($logP_{app}$) of library members with the indicated stereochemistry and the average permeability of all stereoisomers at each degree of N-methylation.

Table 1-22. ROE-distance restraints of Compound 2 in $CDCl_3$. Distances are presented as ranges

Atom 1	Atom 2	Experimental Atomic Distance (Å)	Predicted Atomic Distance (Å)	Violation
Leu4.NH	Leu5.NH	2.5-3.5	2.4192	0.127
Leu4.NH	Thr3.HA	2.5-3.5	2.4172	0.0207
Leu4.NH	Pro9.HA	3.5-4.5	3.1653	0.1485
Thr3.NH	Leu2.HA	2.5-3.5	2.7316	0.0945
Leu8.NH	Ala7.HA	1.7-2.5	2.2205	0
Leu6.NH	Leu5.HA	3.5-4.5	2.5213	0.7418
Ala7.HA	Leu6.HA	1.7-2.5	1.9413	0
Leu5.NH	Thr3.HA	3.5-4.5	3.4473	0.0888
Leu2.HA	Leu2.QNMe	3.5-4.5	3.4802	0.5238
Pro1.HA	Leu2.QNMe	2.5-3.5	2.8647	0.3992
Leu8.HA	Pro9.HD	1.7-2.5	2.1925	0
Pro1.HD	Ac.CH3	2.5-3.5	3.5565	0.0184
Ala7.QB	Ala7.QNMe	1.7-2.5	2.8425	0.0925

Table 1-23. ^3J vicinal coupling constants between Hα-HN protons and derived ϕ dihedral restraints for Compound 2

	^{3}J coupling constant (Hz)	Estimated allowed ϕ angles	Predicted ϕ angles
Thr^3	7.14	-158, -82, 50, 70	-92
Leu^4	8.64	94, 146	101
Leu^5	8.22	-149, -90	-101
Leu^6	9.24	101, 139	110
Leu^8	7.92	-152, -88	-95

Table 1-24. Amide NH temperature shift coefficients for Compound 2

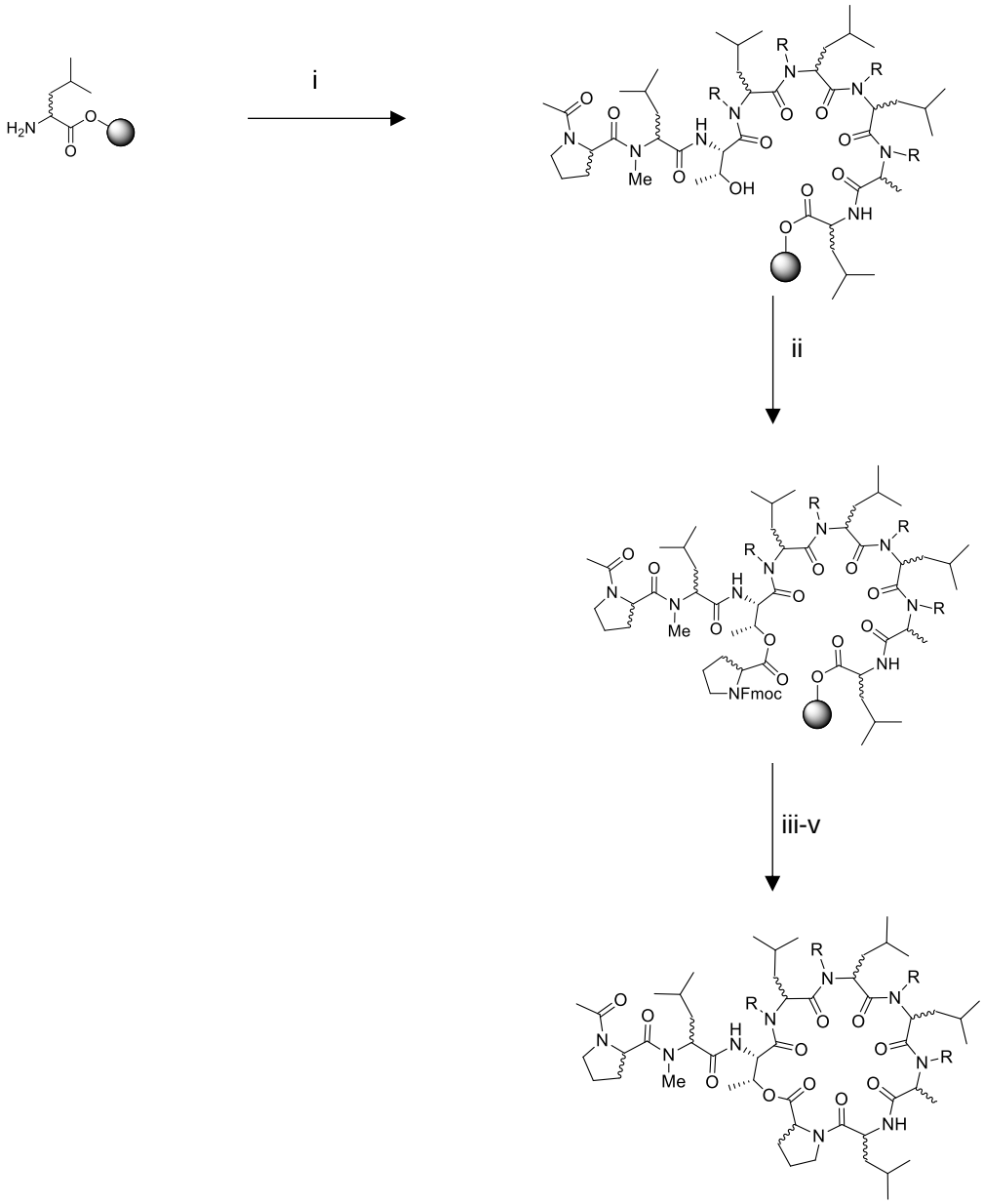
	Thr^3	Leu^4	Leu^5	Leu^6	Leu^8
Δδ/ΔT (ppb/K)	-3.2	-0.8	-1.6	-3.2	-2.4

Table 1-25. Peak assignments for Compound 2

Amino acid	Atom group	Proton (^{1}H) δ (ppm)	Carbon (^{13}C) δ (ppm)
Pro1	Carbonyl-C-Acetyl		169.27
	Acetyl-Methyl	2.01	22.37
	α	4.66	56.13
	β	2.17	28.89
	β	1.95	28.83
	γ	2.19	25.53
	γ	1.96	25.53
	δ	3.65	48.11
	δ	3.56	48.14
	Carbonyl-C		173.49
MeLeu2	N-methyl	3.04	31.32
	α	5.35	54.90
	β	1.87	35.97
	β	1.55	35.96
	γ	1.38	24.88
	δ	0.90	?
	δ	0.87	?
	Carbonyl-C		170.57
Thr3	NH	7.55	
	α	4.60	55.49
	β	4.99	70.91
	γ	1.32	14.79
	Carbonyl-C		168.65
Leu4	NH	7.94	
	α	4.56	52.01
	β	1.71	40.78
	β	1.31	40.78
	γ	1.75	24.70
	δ	0.92	?
	δ	0.88	?
	Carbonyl-C		173.92
Leu5	NH	6.91	
	α	4.46	50.15
	β	1.83	37.92
	β	1.54	37.79
	γ	1.59	24.68
	δ	0.90	?
	δ	0.88	?
	Carbonyl-C		173.18
Leu6	NH	7.42	
	α	4.99	45.79
	β	1.72	40.78
	β	1.58	41.13
	γ	1.48	24.76
	δ	0.92	?

	δ	0.88	?
	Carbonyl-C		171.12
MeAla7	N-methyl	2.67	28.09
	α	5.10	54.44
	β	1.31	15.28
	Carbonyl-C		169.18
Leu8	NH	7.82	
	α	4.69	48.73
	β	1.70	40.55
	β	1.52	40.44
	γ	1.77	24.70
	δ	1.01	23.30
	δ	0.96	21.87
	Carbonyl-C		172.03
Pro9	α	4.11	59.41
	β	2.25	28.93
	β	2.03	28.89
	γ	2.25	25.45
	γ	2.03	25.35
	δ	4.08	47.48
	δ	3.68	47.55
	Carbonyl-C		171.53

1.16. Schemes



Scheme 1-1. Synthesis of Library 1

(i) solid-phase peptide synthesis; (ii). Fmoc-Pro-OH (10eq), DIC (10eq), DMAP (0.25eq) [2x treatment]; (iii) piperidine, DBU; (iv) HFIP; (v) COMU, DIPEA

Chapter 2: Xentrivalpeptides

2.1. Introduction

Natural products have evolved to perform many functions, some of which require passive permeability. We therefore considered lariat peptide natural products as a potential source of permeable lariat peptides. We searched the literature for isoacylthreonine lariat depsipeptides (peptides cyclized from the C-terminus onto a threonine side chain) as candidates for structure-permeability relationship study. Threonine was preferred for the central amino acid over other trifunctional amino acids due to its presence in several bioactive lariat peptides with intracellular targets. The ester linkage with threonine is more stable than the ester linkage with a serine or tyrosine and, unlike an amide, does not contribute an extra H-bond donor. The threonine linkage occurs in several notable lariat peptides with strong bioactivity at intracellular targets, including didemnin B and griselimycin. We searched for structurally simple lipophilic lariat peptides lacking the amino acids present in most lariat peptides. As this investigation would require synthesis of the natural product along with numerous structural analogues, synthetic feasibility was a key criterion.

The xentrivalpeptides are a class of lariat peptides produced by bacteria of the genus *Xenorhabdus*, one of several classes of depsipeptides isolated from this genus. A 2012 study identified a novel class of lariat peptides in a *Xehorhabdus* species, which they termed xentrivalpeptides. Although seventeen were isolated, xentrivalpeptide A (XvA) was the only lariat peptide recovered in quantities sufficient for the

determination of stereochemistry.[1] Sixteen of the xentrivalpeptides have seven amino acids. Interestingly, nine of them (xentrivalpeptides A-I) differ only in the lariat tail residues. The xentrivalpeptides are rich in valine, with XvA containing four valine residues. Aside from the N-terminal cap, the xentrivalpeptides are composed exclusively of non-N-methylated, lipophilic, alpha amino acids. Xentrivalpeptide A (XvA, Figure 2.1) has a molecular weight of 860.07 and contains six H-bond donors. In spite of its high number of H-bond donors, the AlogP of XvA (3.62) is ideal for passive permeability and we therefore hypothesized that this lariat peptide may be passively permeable and that the 3D structure in low-dielectric solvent would offer clues to the structure-permeability relationships of lariat peptides.

2.2. Synthesis

The synthesis was devised to avoid potential challenges associated with lariat peptides (Scheme 2.1). Cyclization efficiency often determines the overall efficiency of cyclic peptide synthesis; therefore, we avoided macrolactonization as the cyclization step and instead opted to form the ester linkage on-resin prior to cyclization. However, since continuing peptide synthesis on the Thr hydroxyl group could lead to diketopiperazine formation after attachment and deprotection of the second residue, we chose the Val^7-Val^6 connection as the site for cyclization. This strategy has been successful in several lariat depsipeptide natural product syntheses.[2-4] The Thr hydroxyl was not protected during the synthesis.[2-3] Beginning with commercially available preloaded L-Val 2-chlorotrityl polystyrene resin, coupling of the first three amino acids was accomplished using the classical Fmoc strategy with

HATU as coupling agent. Acylation of the unprotected Thr hydroxyl group was problematic during subsequent steps. During Fmoc-L-Thr(OH) coupling, acylation of the Thr hydroxyl group occurred using HATU. We therefore opted for TSTU as coupling agent and reduced the Fmoc amino acid equivalency from 4 to 2. Coupling of the subsequent L-Val residue was accomplished using HBTU, and phenylacetic acid coupled to the N-terminus using TSTU. For the formation of the ester linkage, we used DIC/DMAP. Epimerization during ester formation was problematic and required careful separation of the resultant diastereomers. This synthesis successfully produced XvA and was therefore applied to the synthesis of all XvA analogues in this study.

2.3. Structure-permeability relationships

We investigated three structural features of XvA through the synthesis of analogues. First, in the lariat tail, we substituted the extra-annular amino acid and N-terminal cap. The cap was reduced in size and flexibility by substitution of the phenylacetic acid for isobutyric acid (XvA-iBu). This modification significantly decreased the lipophilicity of the peptide. We therefore synthesized XvA and XvA-iBu with Val1 replaced by isoleucine (XvI and XvI-iBu) to restore lipophilicity to XvA-iBu. Xentrivalpeptide I (XvI) was discovered in *Xenorhabdus* alongside XvA in much smaller quantities.

Previous work has indicated a possible role of β-branching in enhancing the permeability of cyclic peptides. Thus, we were interested in the disproportionate

inclusion of valine in this natural product. To study this feature, we individually substituted Val1, Val4, and Val5 with norvaline (Nva) (XvA2, XvA4, and XvA5), and prepared an isomer with Val1, Val4, and Val5 replaced by Nva (XvA6) (Figure 2.2).

Finally, XvA contains a single D amino acid, D-Phe3. Most permeable cyclic peptides discovered to date contain at least one D-amino acid and stereochemistry has been consistently found to influence permeability. To determine the importance of stereochemistry at D-Phe3 to permeability, we prepared an analogue of XvA with the D-Phe replaced by L-Phe3.

The permeability of these analogues was measured in the parallel artificial membrane permeability assay (PAMPA) and in the RRCK cell assay (Figure 2.3). PAMPA is an artificial system comprised of two aqueous compartments separated by a PVDF frit infused with a hydrocarbon to emulate the non-polar environment of a cell membrane. Compared to a cell membrane, this non-polar layer is extremely thick. RRCK cells are a variant of MDCK cells with very low expression of Pgp transporter. RRCK cell permeability is considered a good cell-based indicator of passive permeability.

XvA was moderately permeable with a PAMPA P_{app} of 4.16. The permeabilities of the N-terminal-modified XvA derivatives generally correlated with lipophilicity. XvA-iBu (AlogP: 3.05) had lower permeability while XvI (AlogP: 4.28) had higher permeability, and the permeability of XvI-iBu (AlogP: 3.51) was similar to that of

XvA (3.62). The Nva-substituted XvA analogues had broadly similar permeabilities to XvA, including XvA6 (three of four valines replaced). XvA2, with the lariat tail valine replaced by norvaline, was highly permeable in PAMPA (P_{app} = 9.48). On the other hand, inversion of the D-Phe residue of XvA reduced PAMPA permeability by 60%. Of the three structural features investigated, stereochemistry is the only one that is clearly important to permeability. Modification of the N-terminal acyl group did not impact permeability when AlogP was adjusted by simultaneous modification of the Val1 side chain. The high permeability of valine-to-norvaline modified XvA derivatives indicates a lack of importance of β-branching in facilitating membrane permeability.

The reason for the high valine content of the xentrivalpeptides remains unknown. A structure-permeability study of cyclic heptapeptides suggested that β-branched sidechains improve permeability by steric shielding of N-H groups based on observed improvement in permeability following substitution of Ala by tBuGly.[5] However, evidence for an effect of β-branched sidechains on permeability is mixed. A 2015 structure-permeability relationship study of sanguinamide A found that replacing leucine with valine while keeping AlogP constant had little effect on permeability.[6] Alternatively, β-branched residues have been found to have a role in stabilizing specific conformations of cyclic peptides.[7] Although conformational flexibility has a clear effect on permeability in many cases, conformational stability may also confer increased resistance to enzymatic degradation and decrease the entropic penalty of protein binding. On the other hand, this study confirms the importance of the D

amino acid in XvA in facilitating permeability. Ultimately, this study provided no insight into the effect of the lariat tail on passive permeability, as modification of the lariat tail had no effect on permeability in this case.

2.4. NMR study

NMR has been used extensively to establish relationships between structure and permeability of cyclic peptides. Most often, the NMR structure is solved in a low dielectric solvent, typically $CDCl_3$, to study the mechanism by which the polar backbone amide N-H groups are sequestered from solvent to allow for partitioning into the non-polar environment of the cell membrane. First, we confirmed that the synthesized material was identical to the natural product. Having verified the proposed structure of the natural product, we sought to determine the low-dielectric structure of XvA in $CDCl_3$ with the goal of understanding the behavior of the lariat tail in a low-dielectric environment and the role of the lariat tail in sequestering backbone amide groups. Unfortunately, a lack of amide N-H to carbonyl carbon coupling in the HMBC spectrum combined with a high number of valine residues indistinguishable by TOCSY made peak assignment impossible. The proton spectrum reveals significant broadening of the amide peaks indicating conformational flexibility in $CDCl_3$. Next, we tried collecting NMR data in an even less polar environment using cyclohexane-d_{12} as the solvent. The proton spectrum of XvA in cyclohexane-d_{12} provided sharp N-H peaks. However, the solubility of XvA in cyclohexane was too low to gather HMBC data of sufficient quality to assign the

valine residues. Thus, the use of NMR to study the structure of XvA in a low-dielectric medium proved impossible.

2.5. Conclusions

While this research on the xentrivalpeptides was being carried out, rapid progress was made on the isotope-encoded lariat peptide library project described in Chapter 1. The setbacks described here were not insurmountable, but the opportunity costs outweighed the benefits of continuing to study the xentrivalpeptides. Nevertheless, a few conclusions can be made. XvA has a molecular weight of 860 g/mol and six H-bond donors, five in the cyclic core and one significantly removed from the macrocycle in the flexible lariat tail. Together, this makes permeability seem unlikely. Because four of the seven amino acids are valine, it is reasonable to predict that steric shielding by β-branched residues is implicated in this surprising membrane permeability. We disproved this hypothesis by substituting three of the four valines for structurally generic norvalines without loss of permeability. Although this does not demonstrate the scope of sidechain tolerance for this scaffold, the tolerance of this scaffold to sidechain substitution may indicate that this is a viable space for drug discovery. Our results also reaffirm the importance of tuning lipophilicity (as predicted by AlogP) to optimize permeability for the first time in lariat peptides. Perhaps most importantly, we demonstrated that a non-N-methylated lariat peptide with five H-bond donors in the cyclic portion of the molecule along with a H-bond donor positioned in the lariat tail, a significant distance from the macrocyclic portion of the molecule, is passively permeable. Given the importance of sequestering H-

bond donors to passive permeability, the low-dielectric structure would certainly prove fascinating and unrestrained molecular dynamics simulation could be a rewarding future direction for the study of XvA.

2.6. Methods

Peptides were synthesized on commercially available L-Val preloaded 2-chlorotrityl polystyrene resin (0.85mmol/g). All reagents used were purchased and used without further purification.

Amide coupling procedures (A, C, D, and E in Scheme 2.1)

Method A

To the Fmoc amino acid (4eq) was added HATU (3.8eq, 0.5 M in DMF) followed by DIPEA (5eq). The resultant solution was sonicated and allowed to stand for 5 minutes, then added to the drained resin. The resin was heated to 50°C for 60min. After coupling, the resin was washed with DMF (3x) and DCM (3x).

Method C

To Fmoc-L-Thr(OH) (2eq) was added TSTU (1.9eq, 0.5 M in DMF) followed by DIPEA (2.5eq). The resultant solution was sonicated and allowed to stand for 5 minutes, then added to the drained resin. The resin was heated to 50°C for 60min. After coupling, the resin was washed with DMF (3x) and DCM (3x).

Method D

To Fmoc-L-Val(OH) (2eq) was added HBTU (1.9eq, 0.5 M in DMF) followed by DIPEA (2.5eq). The resultant solution was sonicated and allowed to stand for 5 minutes, then added to the drained resin. The resin was heated to 50°C for 60min. After coupling, the resin was washed with DMF (3x) and DCM (3x).

Method E

To phenylacetic acid (2eq) was added TSTU (1.9eq, 0.5 M in DMF) followed by DIPEA (2.5eq). The resultant solution was sonicated and allowed to stand for 5 minutes, then added to the drained resin. The resin was heated to 50°C for 60min. After coupling, the resin was washed with DMF (3x) and DCM (3x).

Fmoc deprotection (B in Scheme 2.1)

The resin was first washed DMF to remove residual DCM from previous steps. The resin was then treated with a solution containing 2% DBU and 2% piperidine in DMF for 15min at room temperature. The resin was then washed with DMF (3x) and DCM (3x).

Ester formation using DIC (F in Scheme 2.1)

Fmoc-L-Val (3 eq) was dissolved in DMF/DCM (1:9, roughly 3mL/g Fmoc-L-Val). DMAP (0.1 eq) was added followed by DIC (3 eq). The solution was swirled rapidly and allowed to stand for 5 minutes. The mixture was added to the resin and the SPE

tube capped. The reaction was shaken at room temperature overnight. The resin was drained and washed with DMF (3x) and DCM (3x).

Cleavage from 2-chlorotrityl resin (G in Scheme 2.1)

Branched linear peptides were cleaved from the resin with 25% HFIP in DCM (2 x 30 min). The resin was rinsed with DCM between treatments. Solvent was evaporated under a stream of nitrogen. DCM was added and evaporation repeated. The residue was stored overnight in a vacuum desiccator prior to cyclization.

Cyclization (H in Scheme 2.1)

The solvent volumes in this procedure are for 0.1 mmol of peptide. The concentration during cyclization was approximately 0.001 M.

COMU (3 eq) was placed in a round-bottom flask with a stir bar, followed by THF (90 mL) and DIPEA (3eq). In a separate vessel, the branched linear peptide was dissolved in ACN (10mL) and DIPEA (3 eq). The peptide was added dropwise and in portions to the round-bottom flask with rapid stirring during 30 minutes. Stirring was continued for 16 h. The solution was concentrated under reduced pressure.

Peptide purification

Crude cyclic peptides were purified on a Biotage Isolera Prime automated chromatography system equipped with a SNAP Bio C18 25g column eluting with water/acetonitrile containing 0.1% TFA. If further purification was required, a Waters HPLC system was utilized. Eluting with water/acetonitrile containing 0.1% formic acid.

PAMPA Assay

The PAMPA assay was carried out and the peak volumes interpreted using a procedure utilized previously in our lab.[32]

The analyte concentration in the PAMPA assay was 1 μM. Internal standards were included in the assay at a concentration of 1 μM. Carbamazepine (1 μM) was used as the internal standard. The assay was run for 18 h.

A 96-well donor plate with 0.45 μm hydrophobic Immobilon-P membrane supports (Millipore MAIPNTR10) and a 96-well Teflon acceptor plate (Millipore MSSACCEPTOR) were used in the PAMPA permeability test. The acceptor plate was prepared by adding 300 μL of 5% DMSO in 1X PBS to each well. Donor well solutions were prepared by diluting 50 μL DMSO stock solutions prepared above to a final volume of 1000 μL with PBS and mixed thoroughly. The frits were infused with 5 μL of dodecane containing 1% (w/v) soy lecithin (90%, Alfa Aesar). The membranes were allowed to equilibrate for 5 minutes before adding the donor well solution and placing on top of the acceptor well solution to begin the assay. Samples were prepared for LC-MS analysis by diluting with an equal volume of ACN. The donor wells were further diluted tenfold with 1:1 ACN/H_2O for approximately even analyte concentration in the donor and acceptor wells.

2.7. References

1. Zhou, Q.; Dowling, A.; Heide, H.; Wohnert, J.; Brandt, U.; Baum, J.; Ffrench-Constant, R.; Bode, H. B., Xentrivalpeptides A-Q: depsipeptide diversification in Xenorhabdus. *J Nat Prod* **2012,** *75* (10), 1717-22.

2. Kuranaga, T.; Enomoto, A.; Tan, H.; Fujita, K.; Wakimoto, T., Total Synthesis of Theonellapeptolide Id. *Org Lett* **2017,** *19* (6), 1366-1369.

3. Yao, G.; Wang, W.; Ao, L.; Cheng, Z.; Wu, C.; Pan, Z.; Liu, K.; Li, H.; Su, W.; Fang, L., Improved Total Synthesis and Biological Evaluation of Coibamide A Analogues. *J Med Chem* **2018,** *61* (19), 8908-8916.

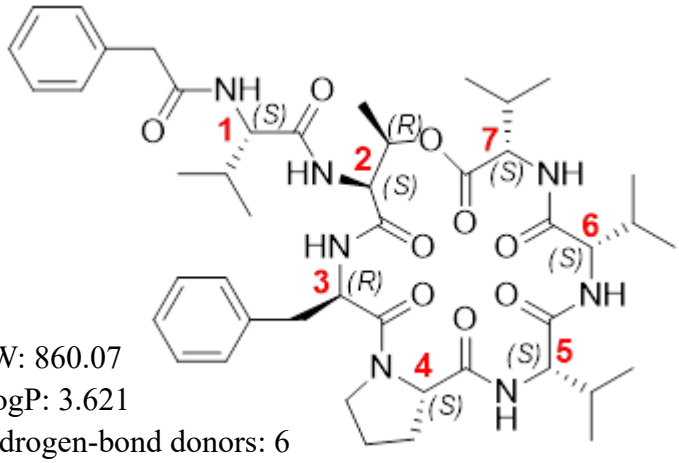
4. Seo, H.; Lim, D., Total Synthesis of Halicylindramide A. *The Journal of Organic Chemistry* **2009,** *74* (2), 906-909.

5. Nielsen, D. S.; Hoang, H. N.; Lohman, R.-J.; Hill, T. A.; Lucke, A. J.; Craik, D. J.; Edmonds, D. J.; Griffith, D. A.; Rotter, C. J.; Ruggeri, R. B.; Price, D. A.; Liras, S.; Fairlie, D. P., Improving on Nature: Making a Cyclic Heptapeptide Orally Bioavailable. *Angewandte Chemie International Edition* **2014,** *53* (45), 12059-12063.

6. Bockus, A. T.; Schwochert, J. A.; Pye, C. R.; Townsend, C. E.; Sok, V.; Bednarek, M. A.; Lokey, R. S., Going Out on a Limb: Delineating The Effects of beta-Branching, N-Methylation, and Side Chain Size on the Passive Permeability, Solubility, and Flexibility of Sanguinamide A Analogues. *J Med Chem* **2015,** *58* (18), 7409-18.

7. Cummings, A. E.; Miao, J.; Slough, D. P.; McHugh, S. M.; Kritzer, J. A.; Lin, Y.-S., β-Branched Amino Acids Stabilize Specific Conformations of Cyclic Hexapeptides. *Biophysical Journal* **2019,** *116* (3), 433-444.

2.8. Figures



MW: 860.07
AlogP: 3.621
Hydrogen-bond donors: 6

Figure 2-1. Structure of Xentrivalpeptide A (XvA)

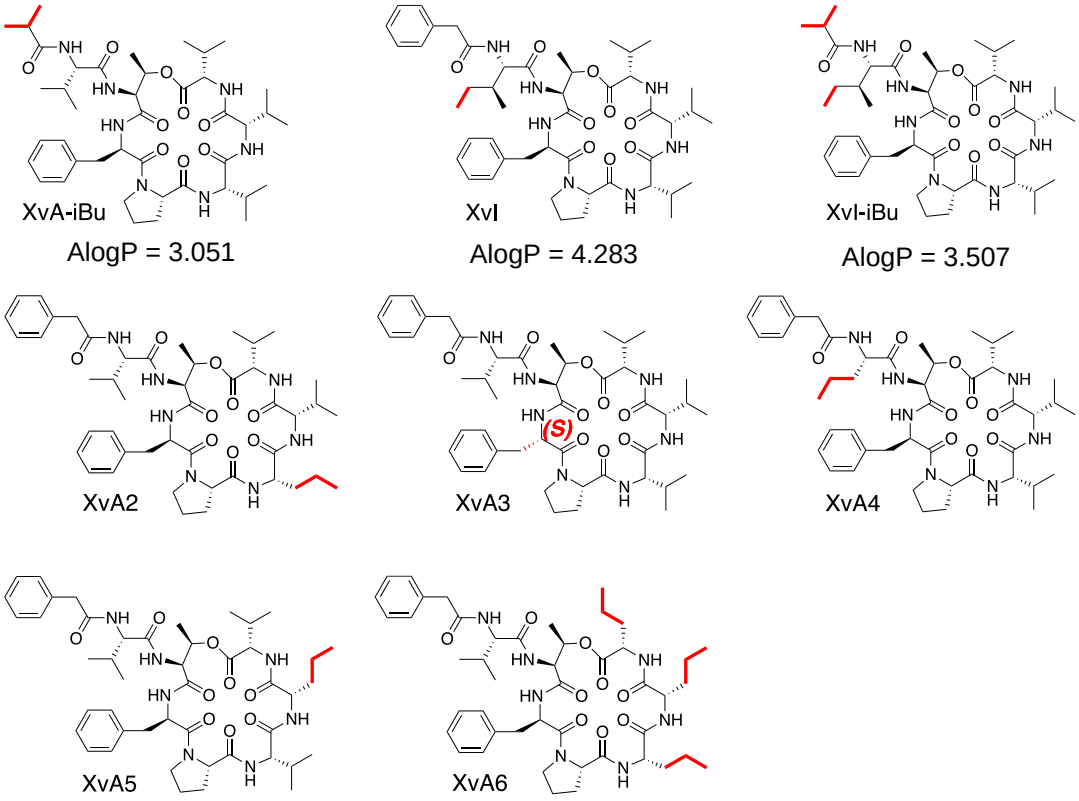


Figure 2-2. Structures of synthesized XvA derivatives

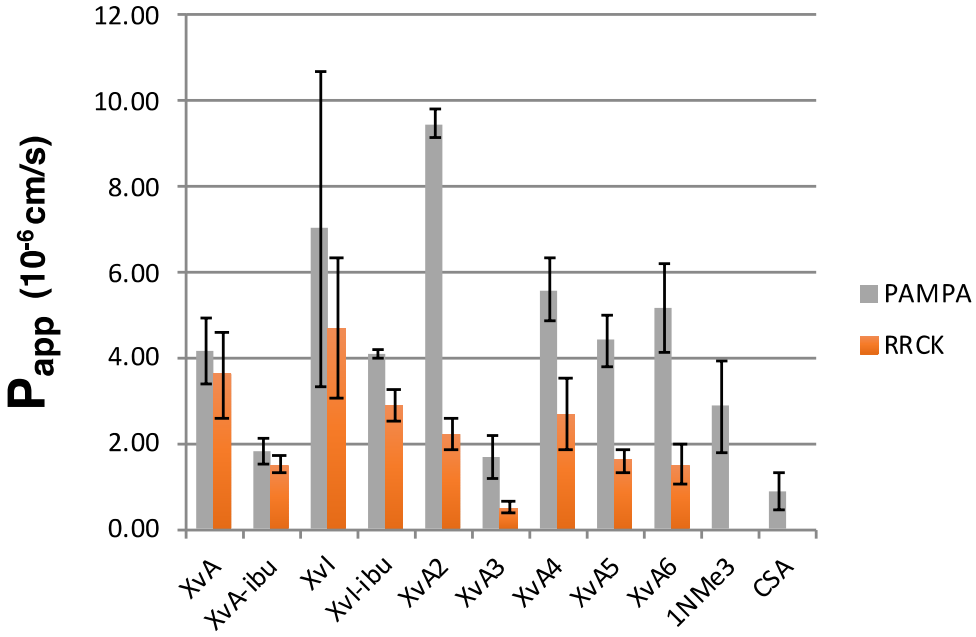
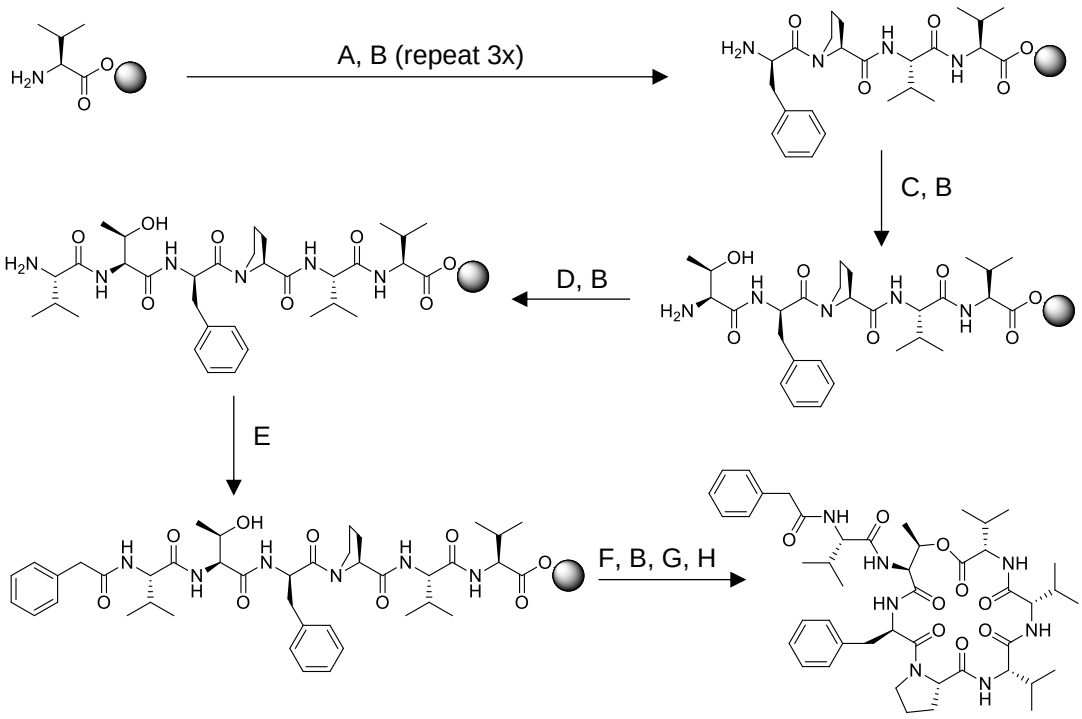


Figure 2-3. Permeabilities of XvA analogues in PAMPA and RRCK assay

2.9. Schemes



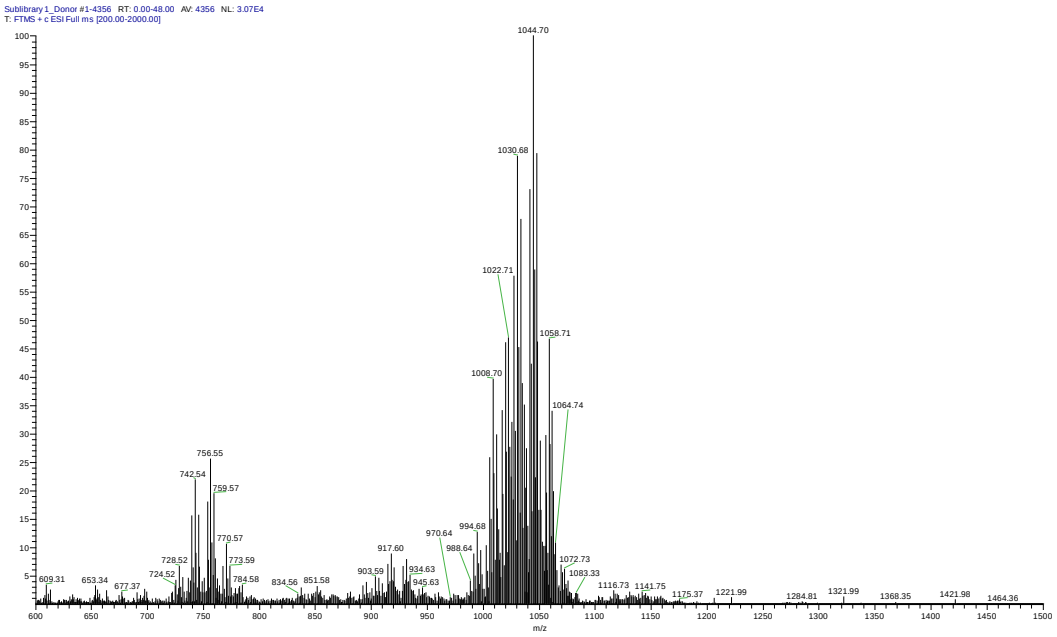
Scheme 2-1. Synthesis of xentrivalpeptide A

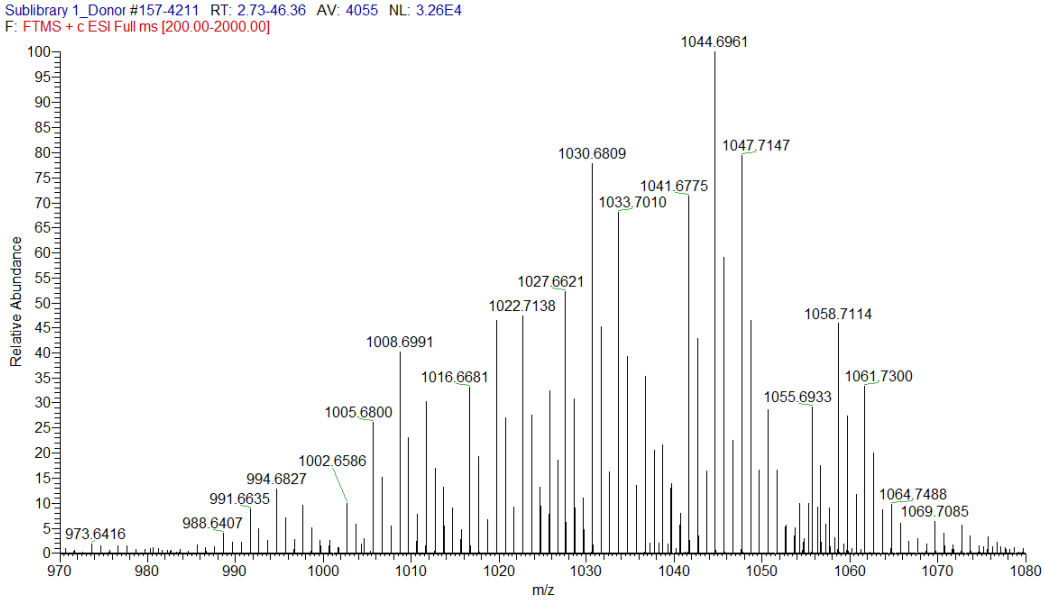
These conditions were applied to the synthesis of all XvA derivatives:

(A) Fmoc-aa (4eq), HATU (3.8eq), DIPEA (5eq)
(B) piperidine, DBU
(C) Fmoc-L-Thr(OH) (2eq), TSTU (1.9eq), DIPEA (2.5eq)
(D) Fmoc-L-Val (2eq), HBTU (1.9eq), DIPEA (2.5eq)
(E) Phenylacetic acid (2eq), TSTU (1.9eq), DIPEA (2.5eq)
(F) Fmoc-L-Val (3eq), DIC (3eq), DMAP (0.1eq)
(G) HFIP
(H) COMU (3eq), DIPEA (6eq)

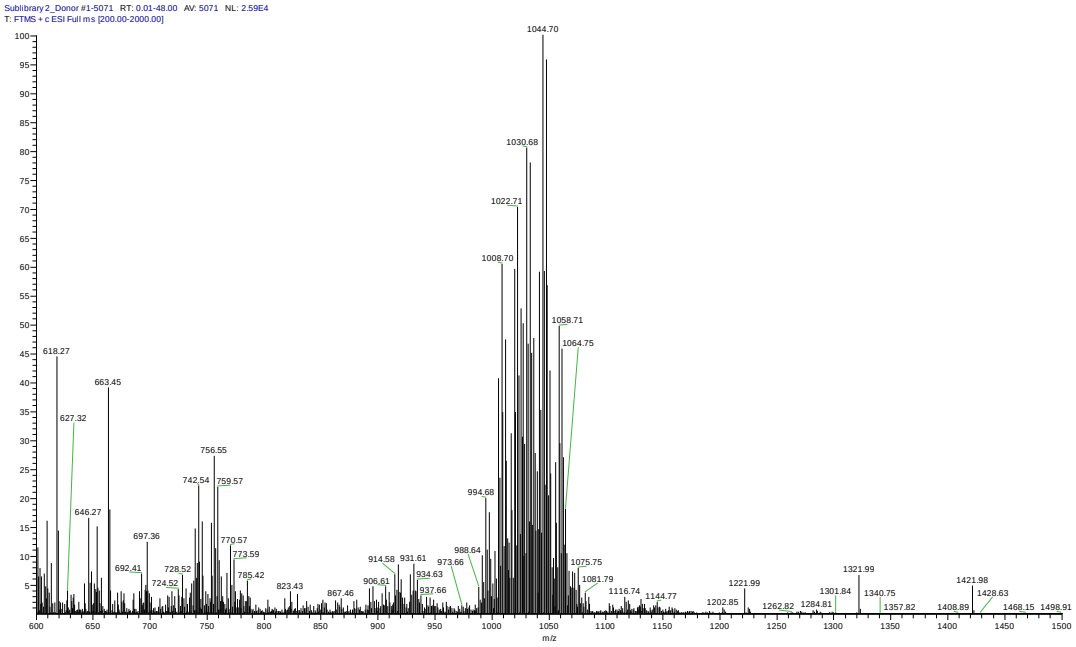
APPENDIX A Chapter 1 Mass spectra of Libraries 1-5

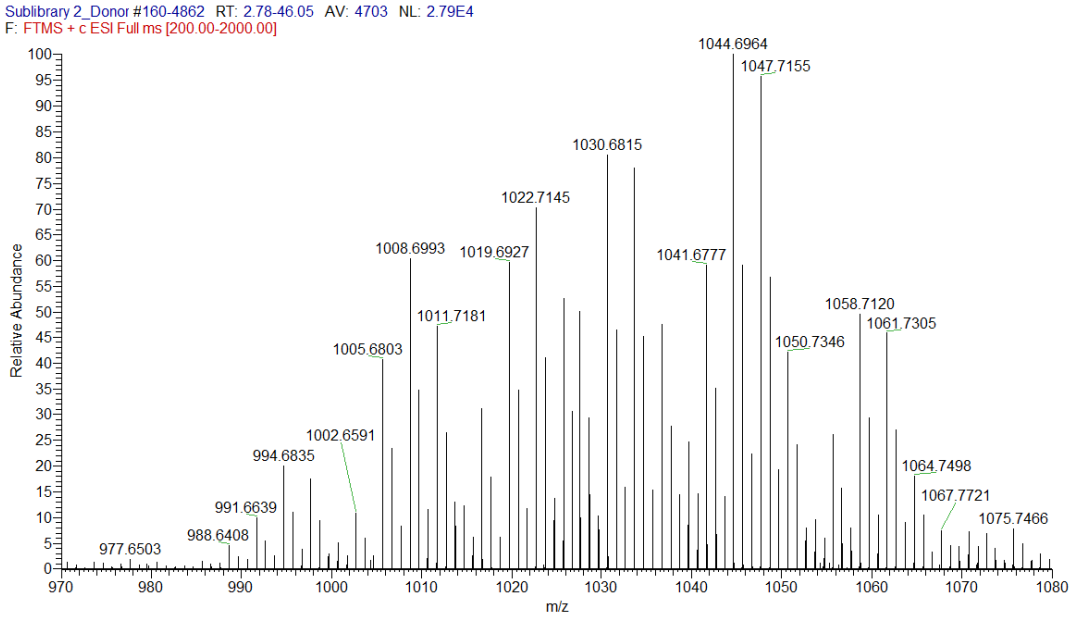
Mass spectrum of sub-library 1



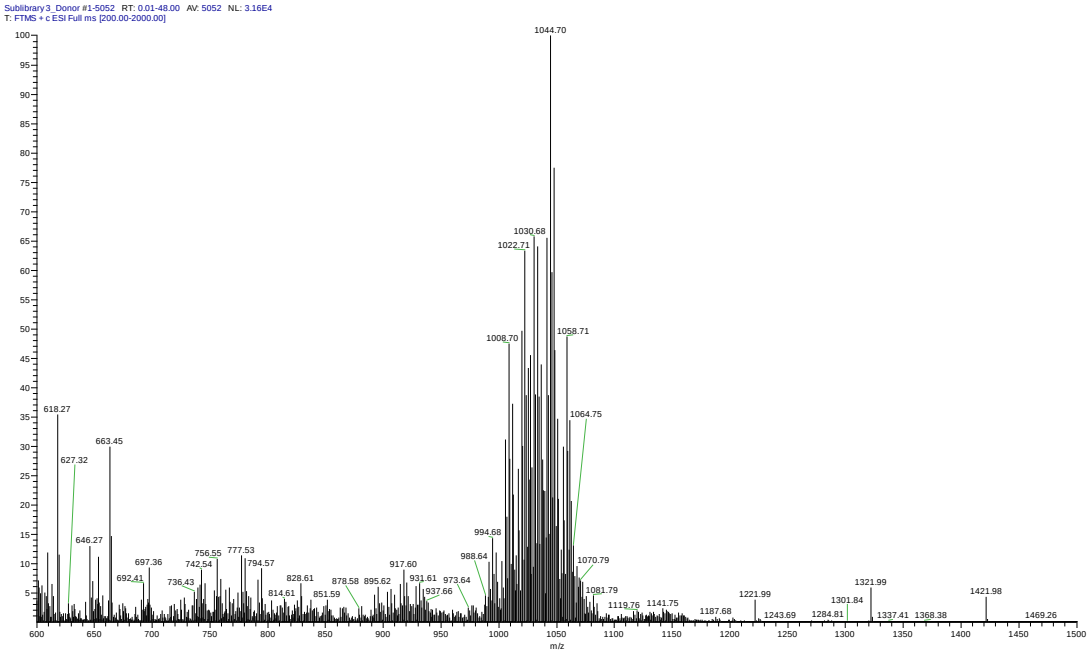


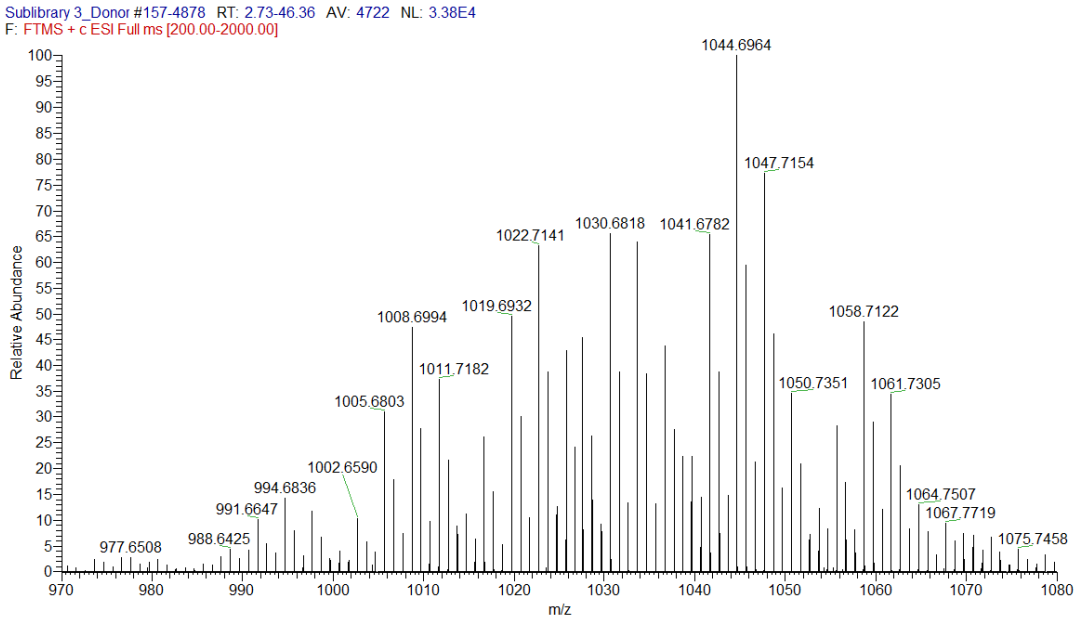
Mass spectrum of sub-library 2



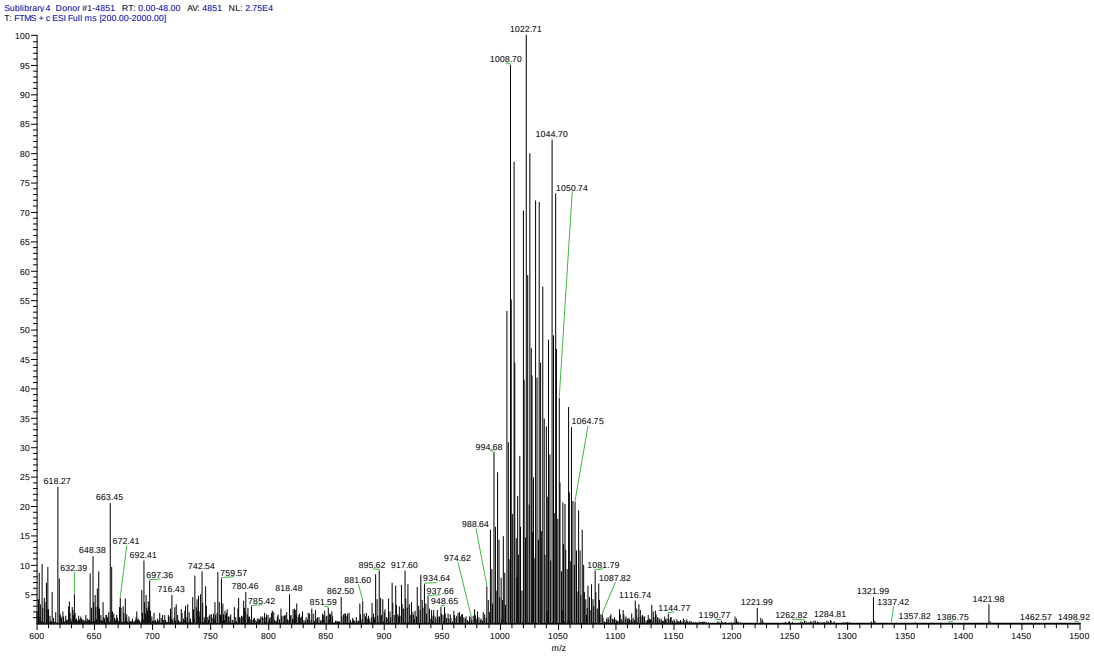


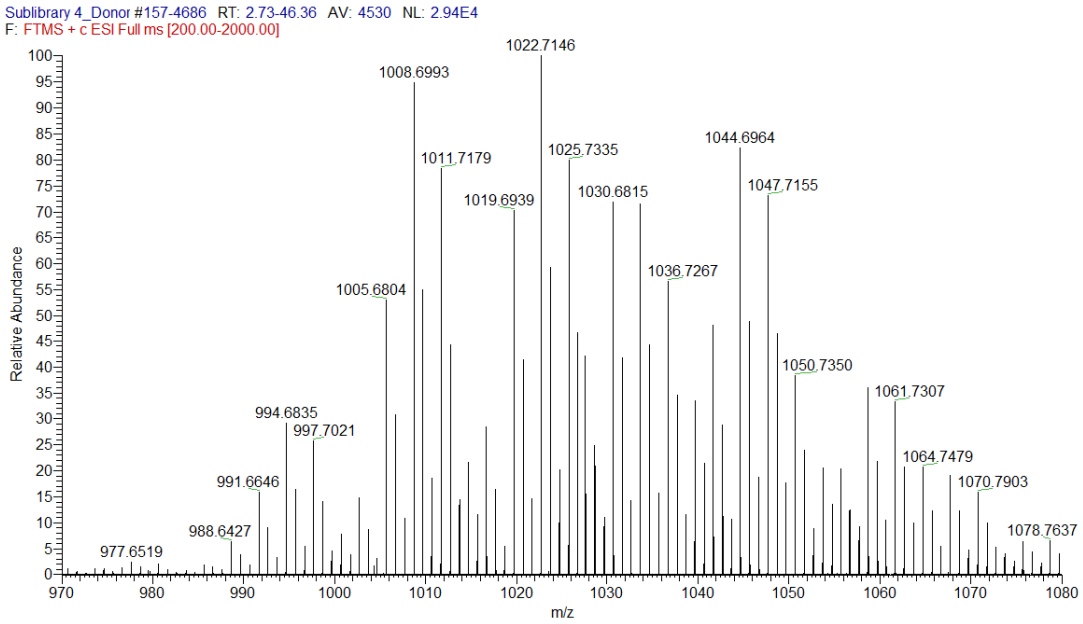
Mass spectrum of sub-library 3



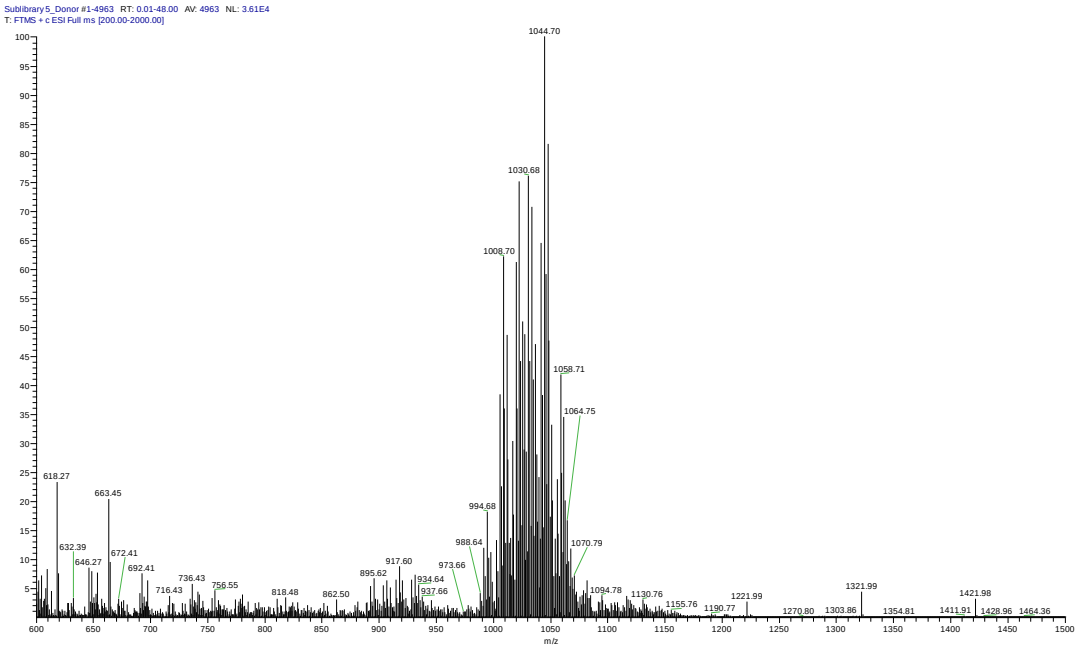


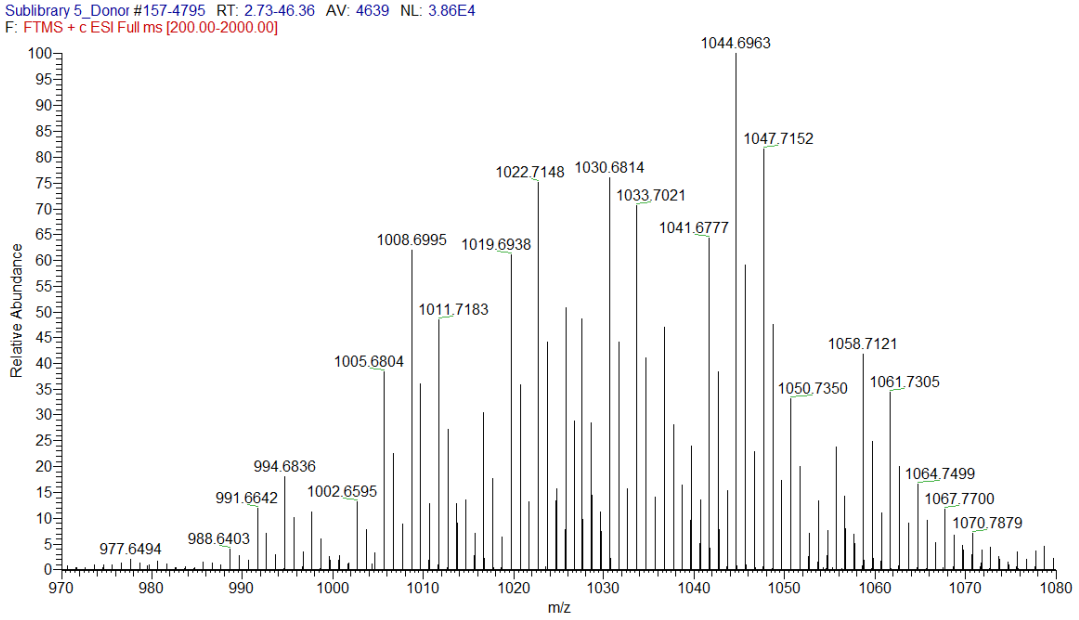
Mass spectrum of sub-library 4



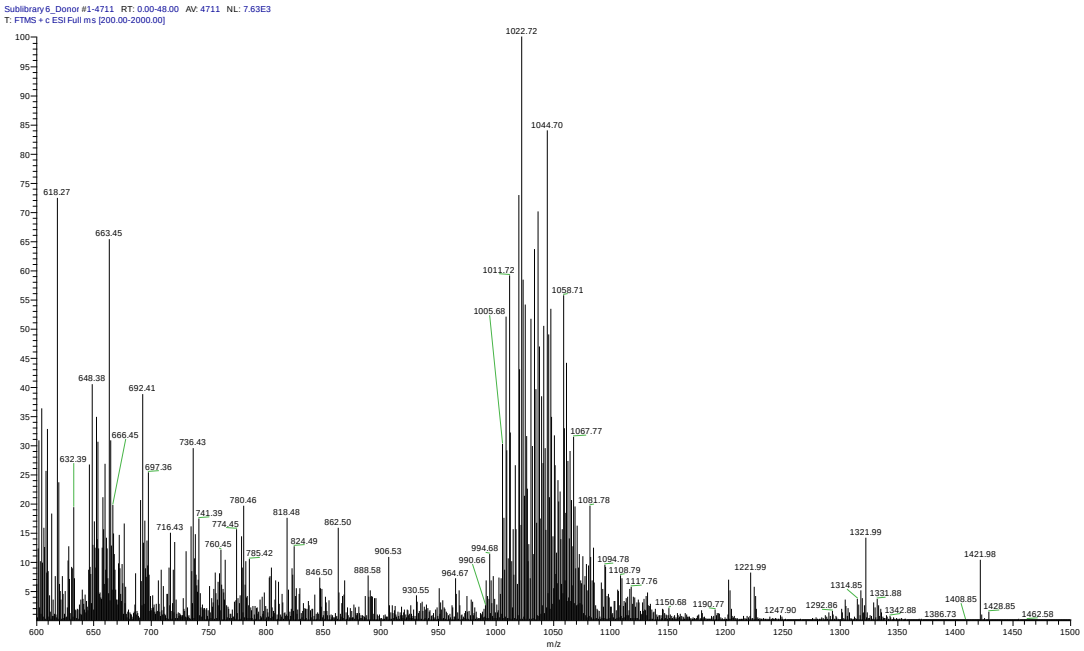


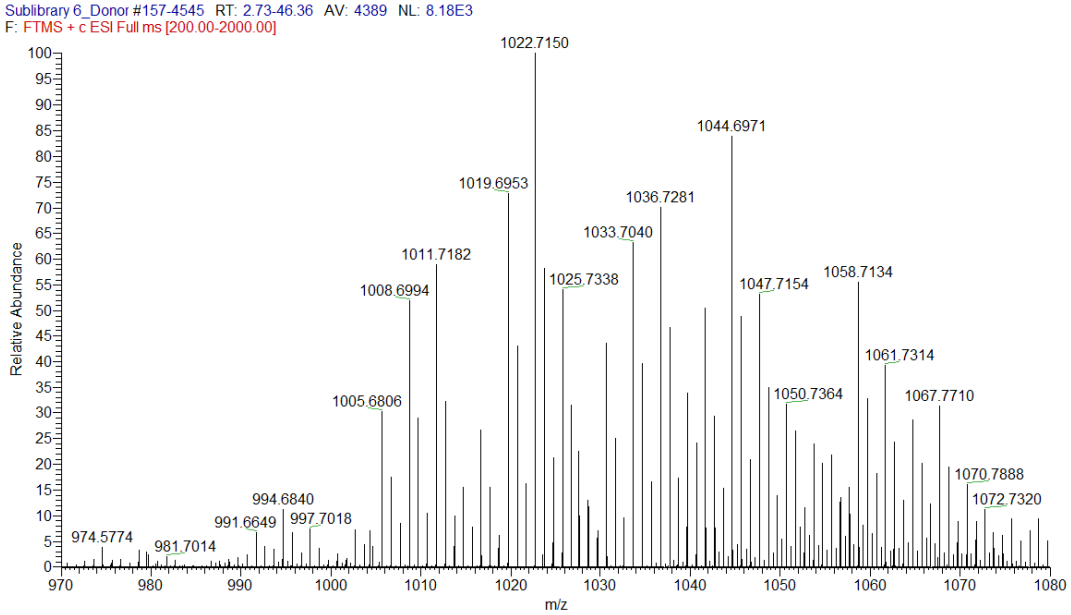
Mass spectrum of sub-library 5



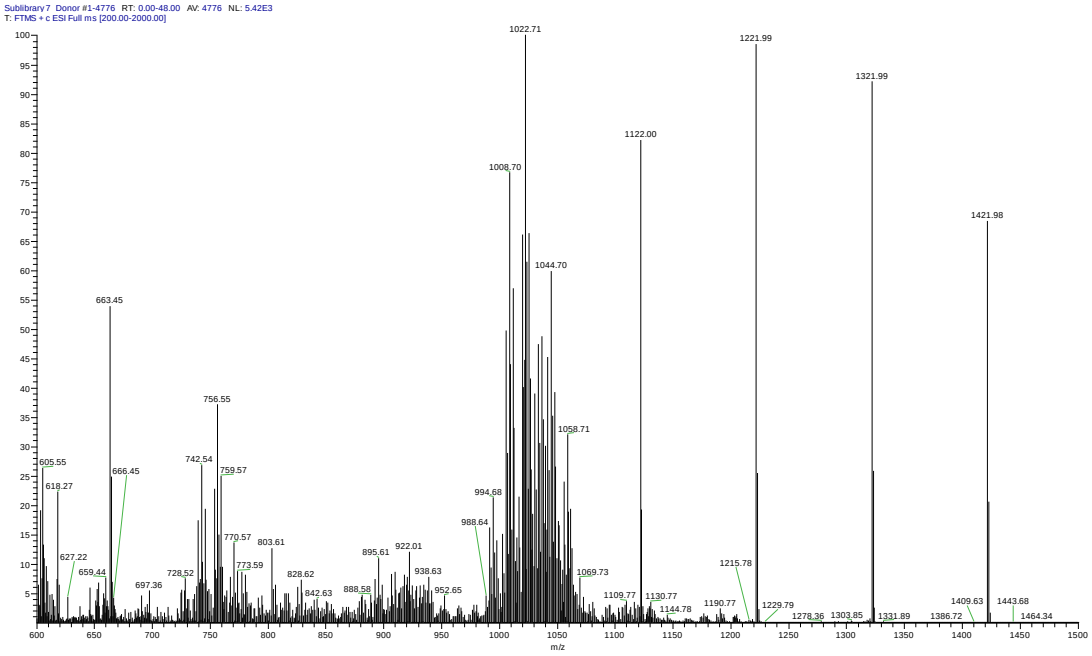


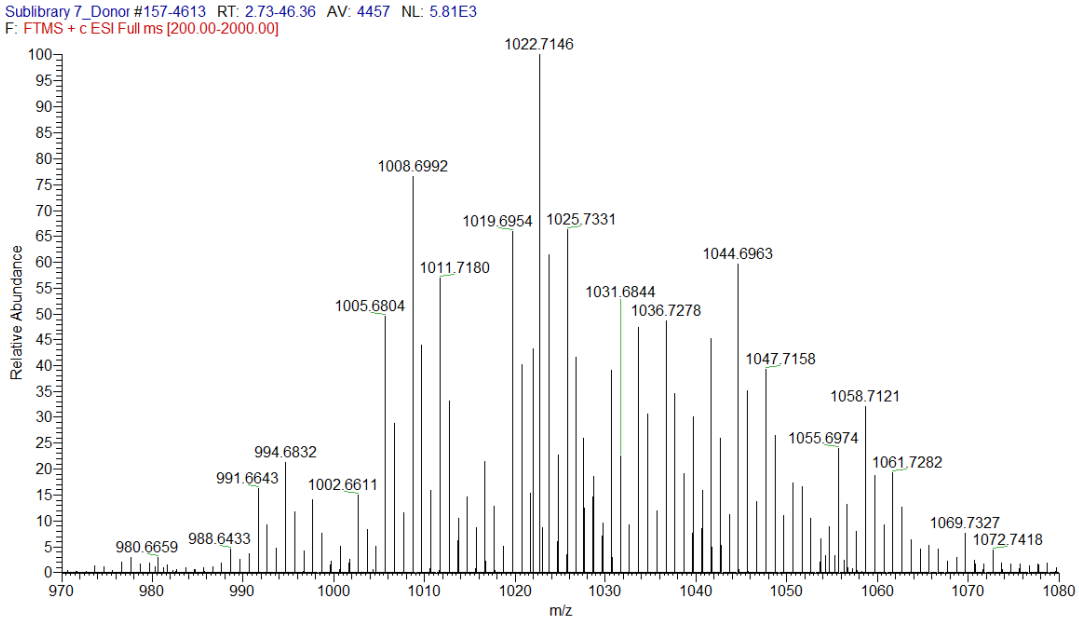
Mass spectrum of sub-library 6



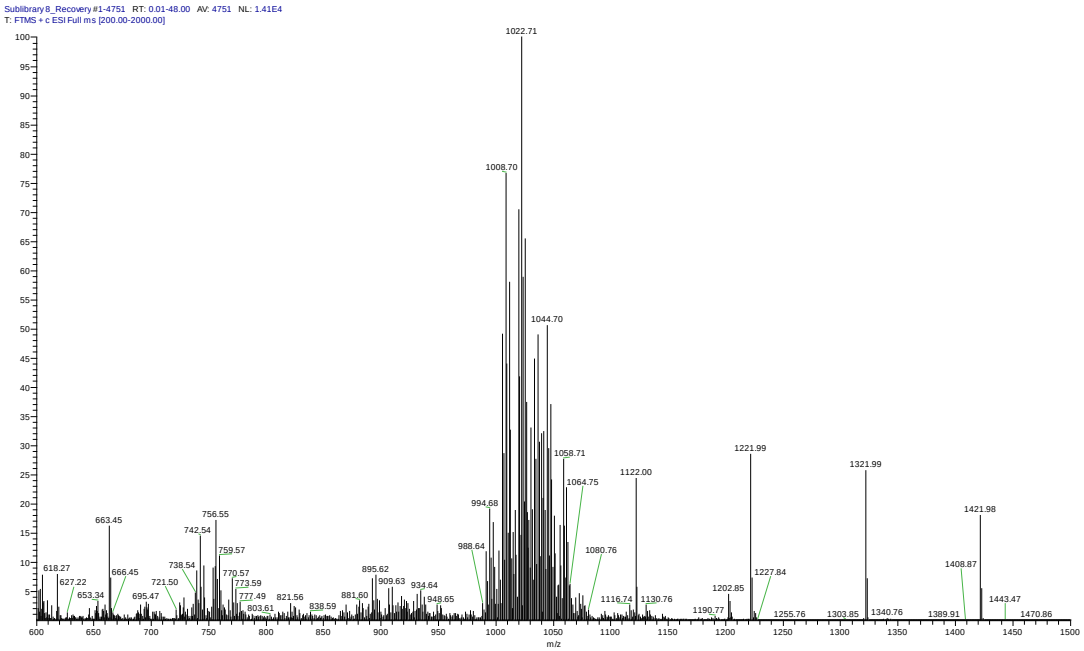


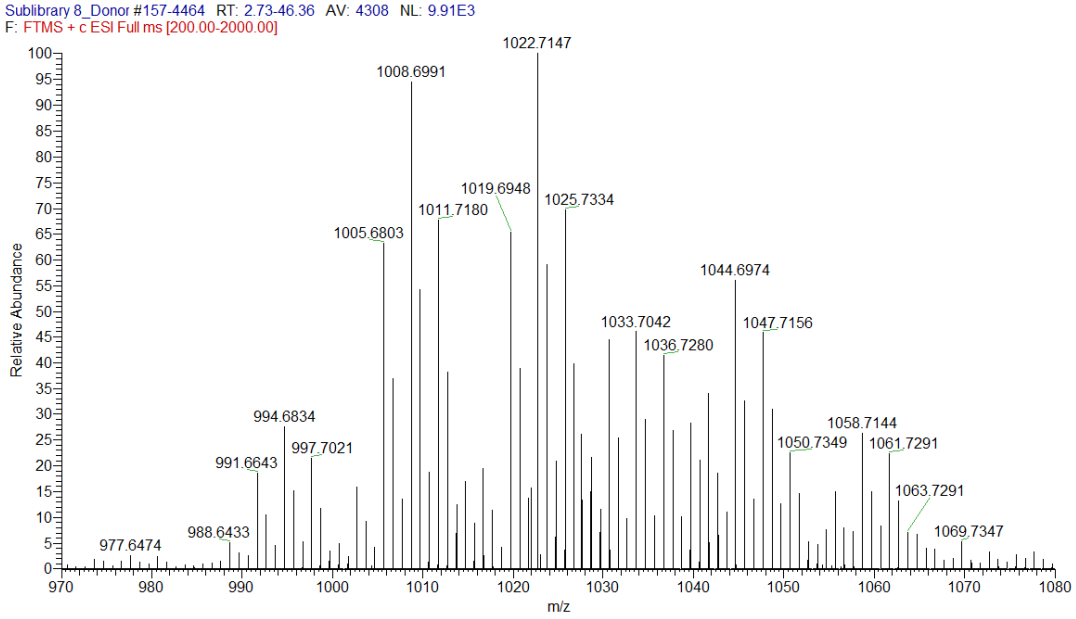
Mass spectrum of sub-library 7



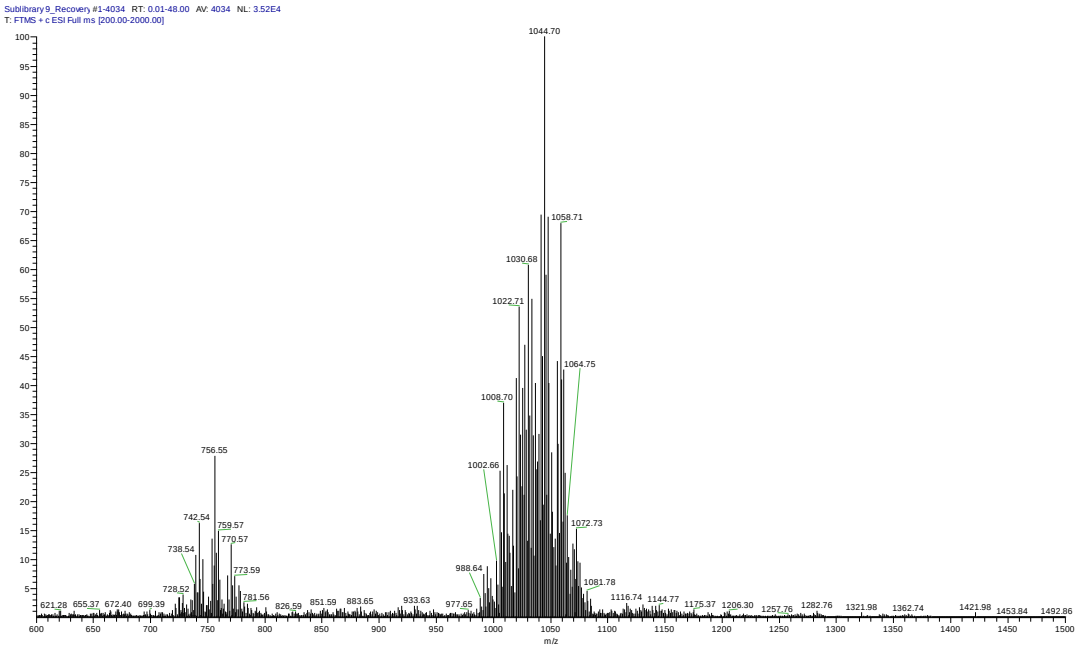


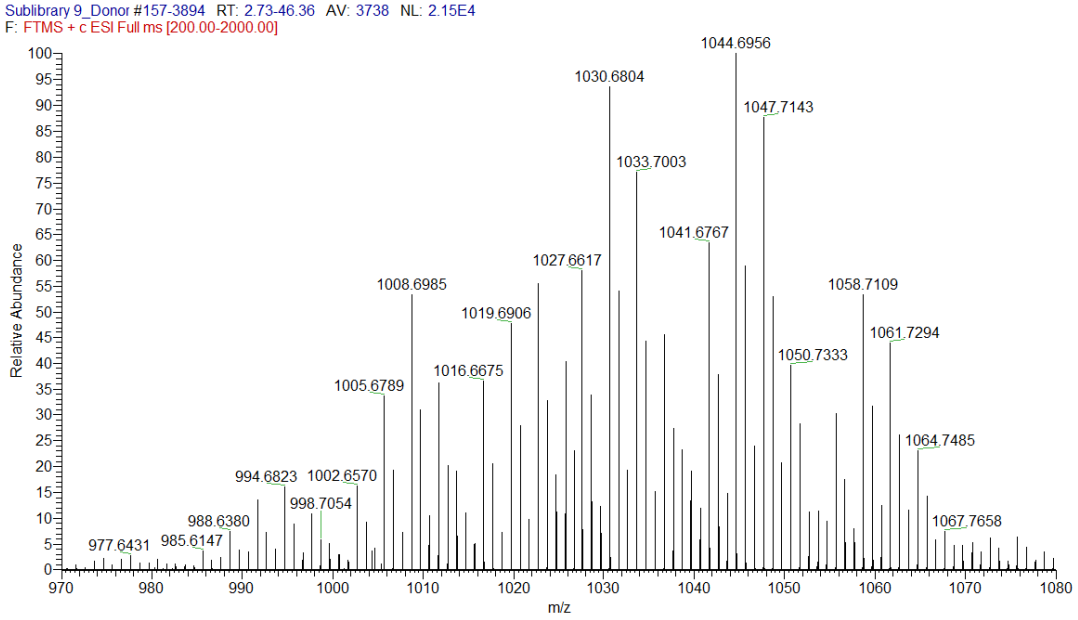
Mass spectrum of sub-library 8



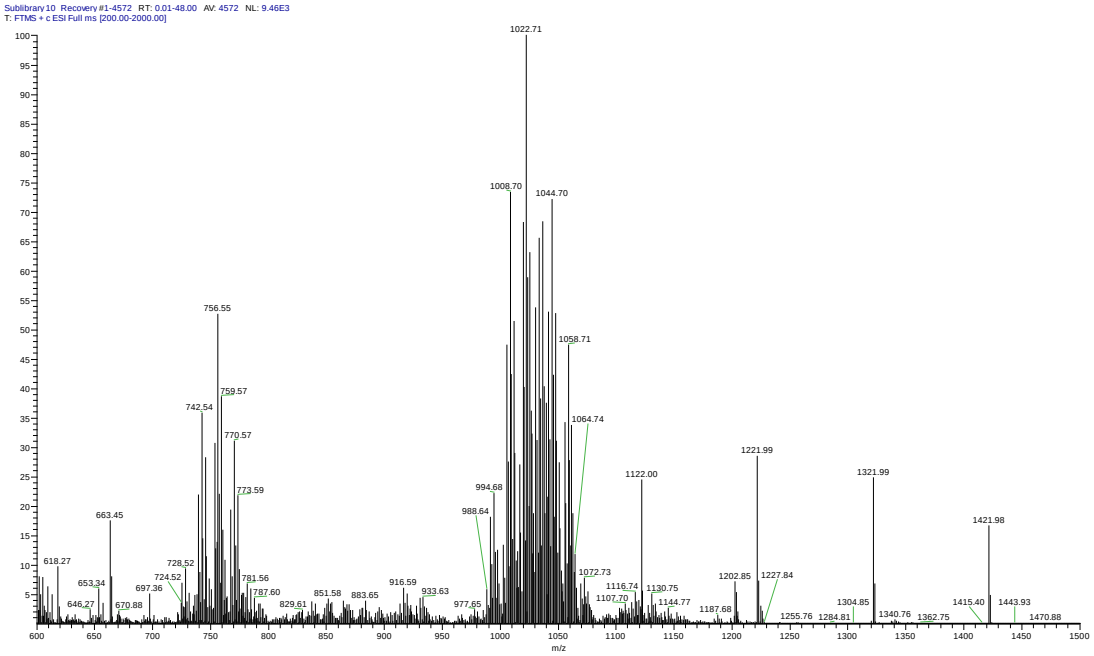


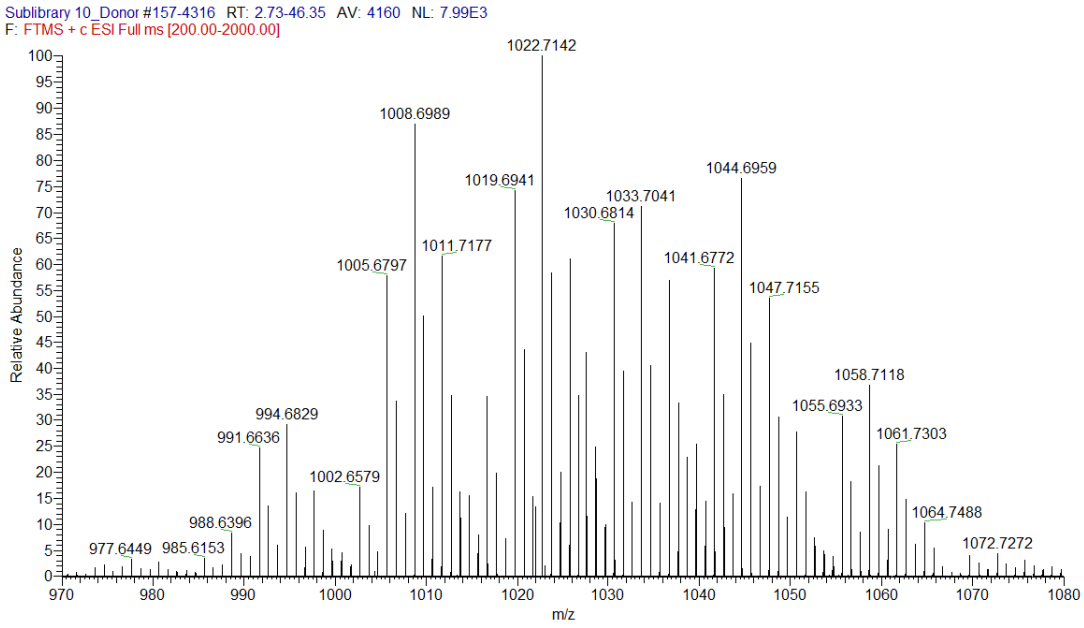
Mass spectrum of sub-library 9



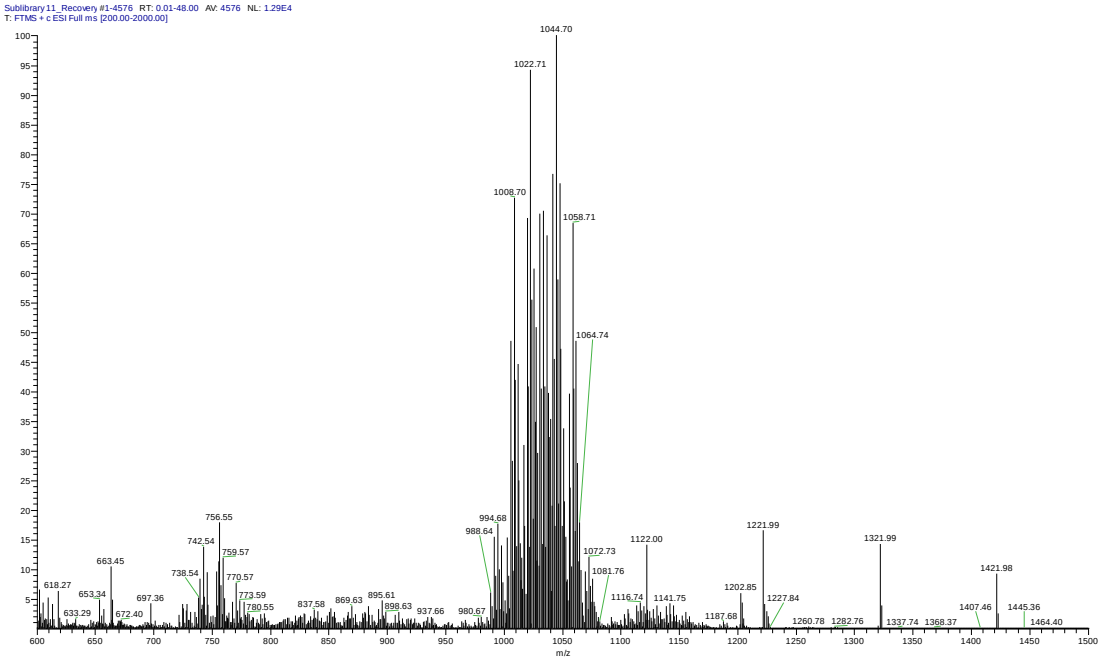


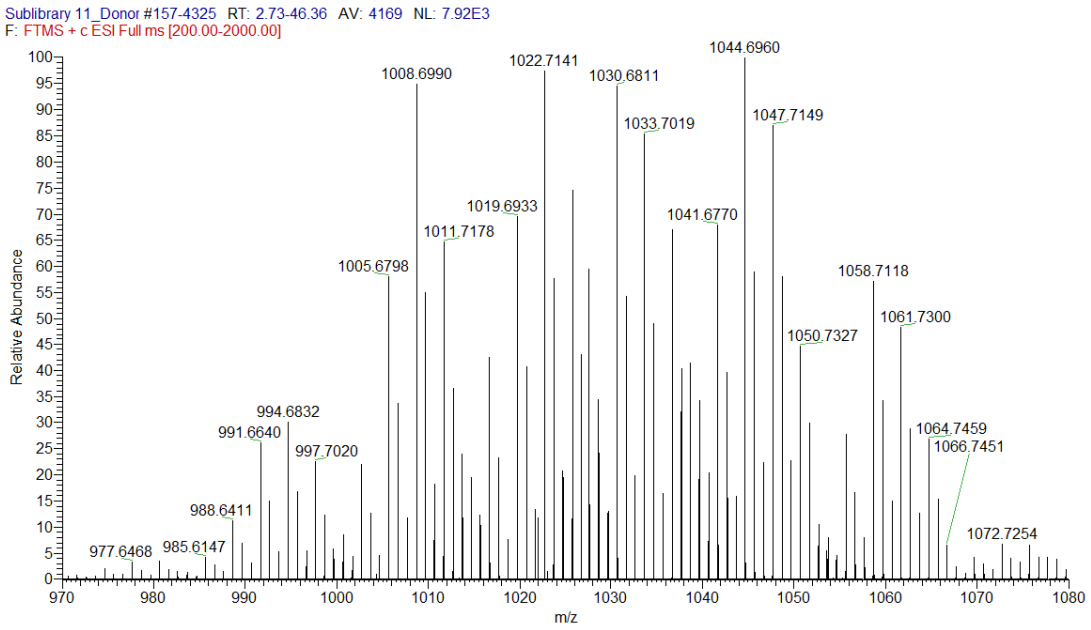
Mass spectrum of sub-library 10



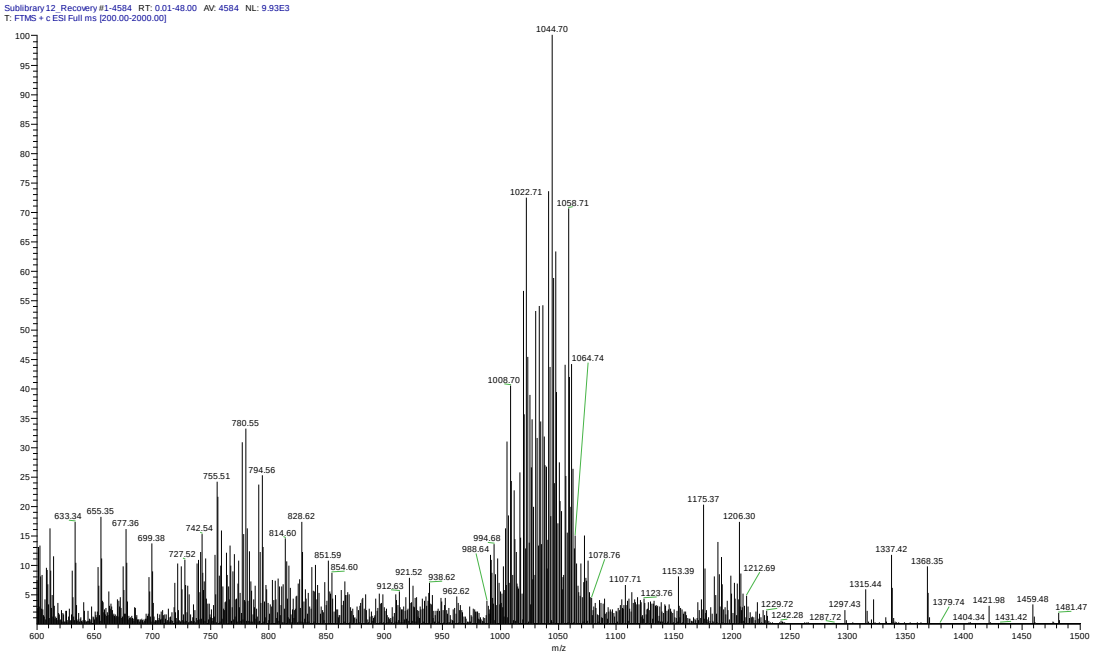


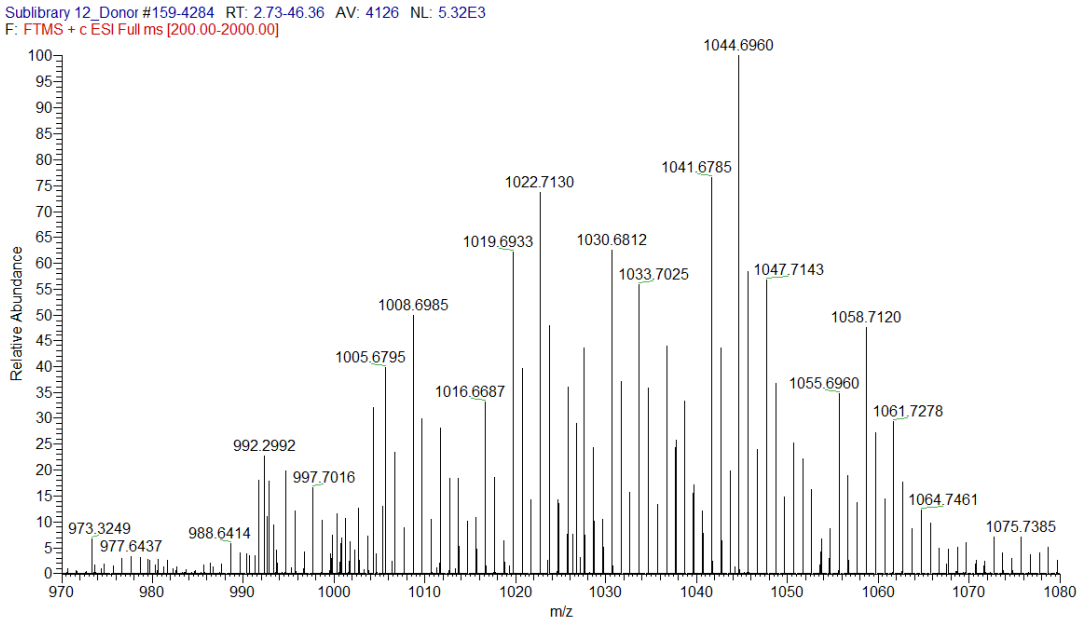
Mass spectrum of sub-library 11



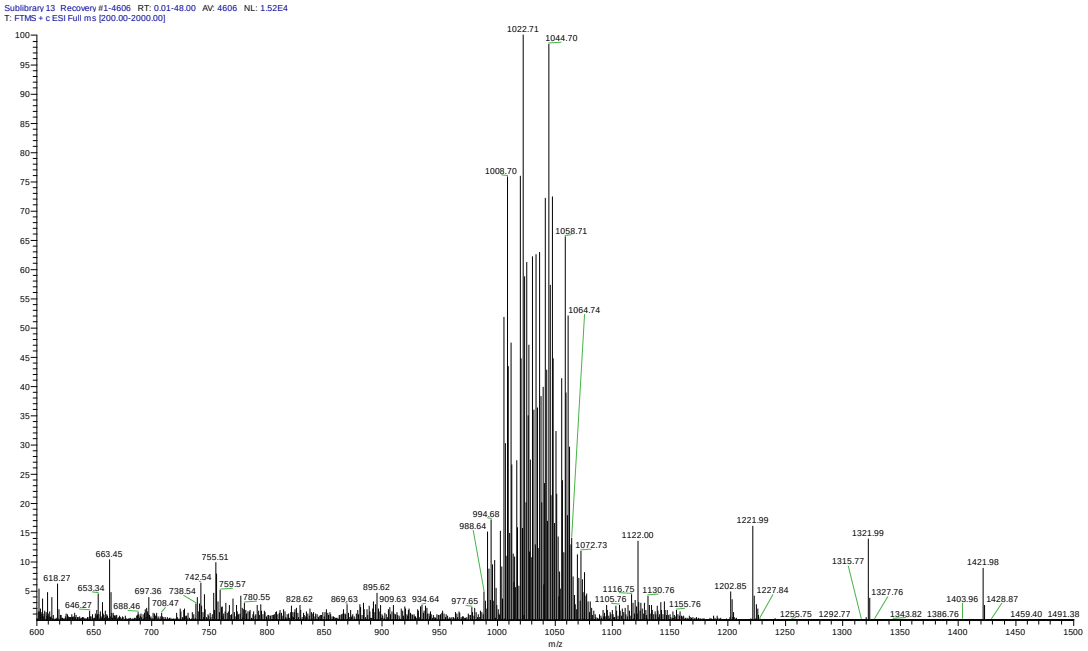


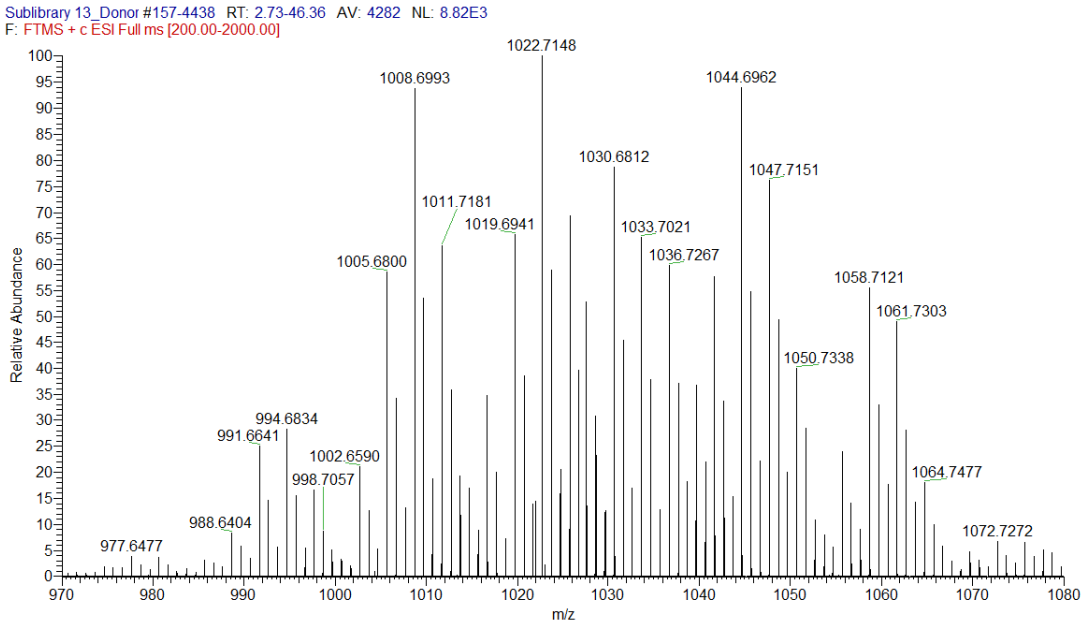
Mass spectrum of sub-library 12



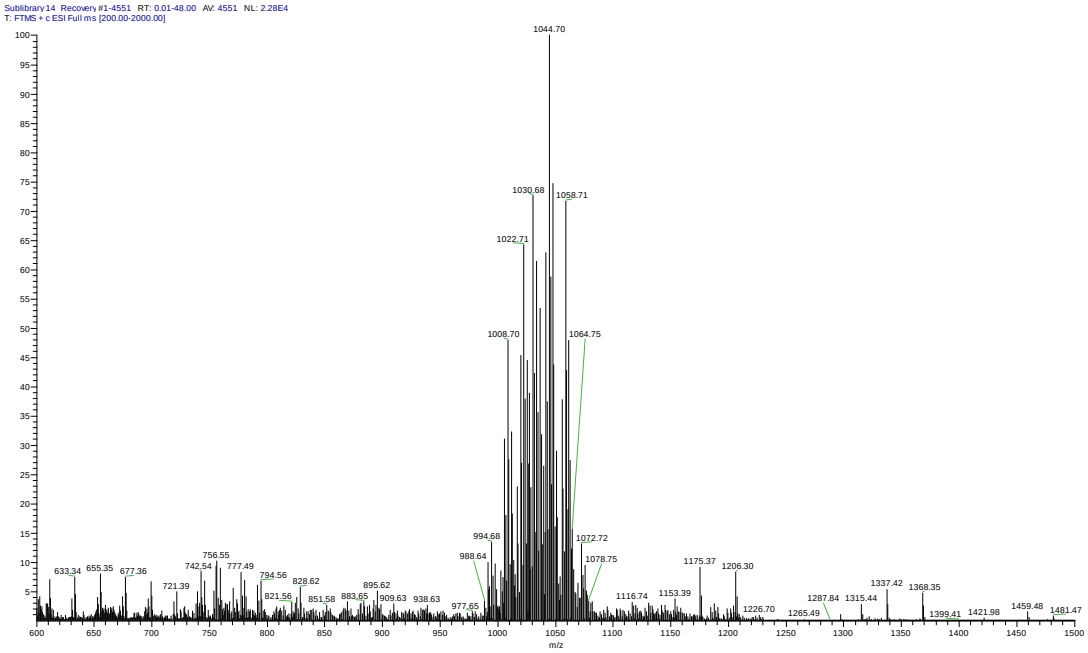


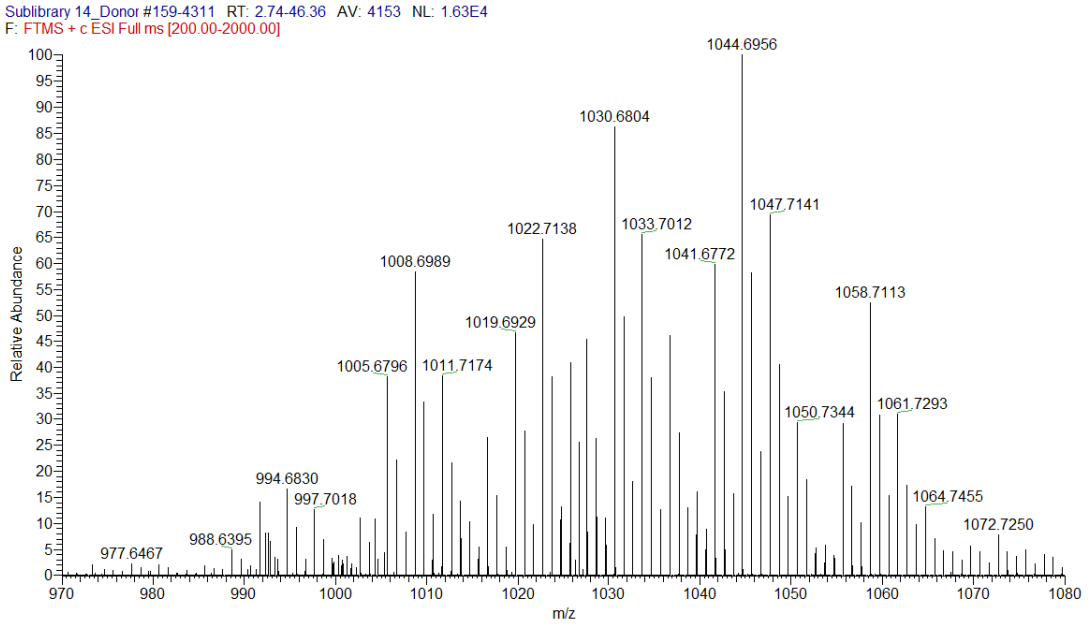
Mass spectrum of sub-library 13



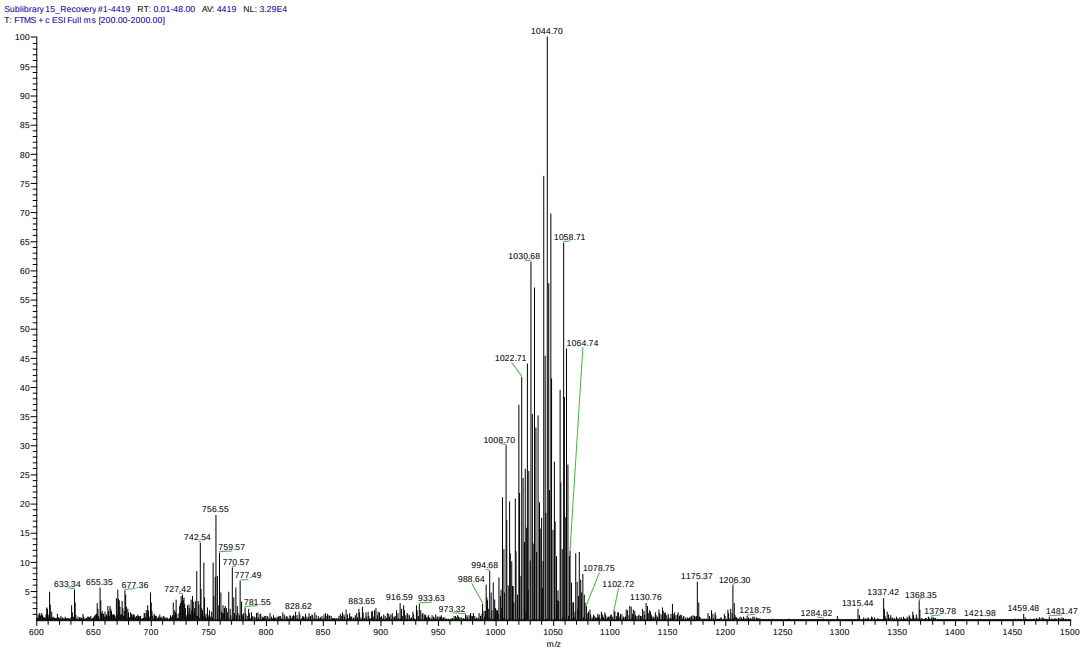


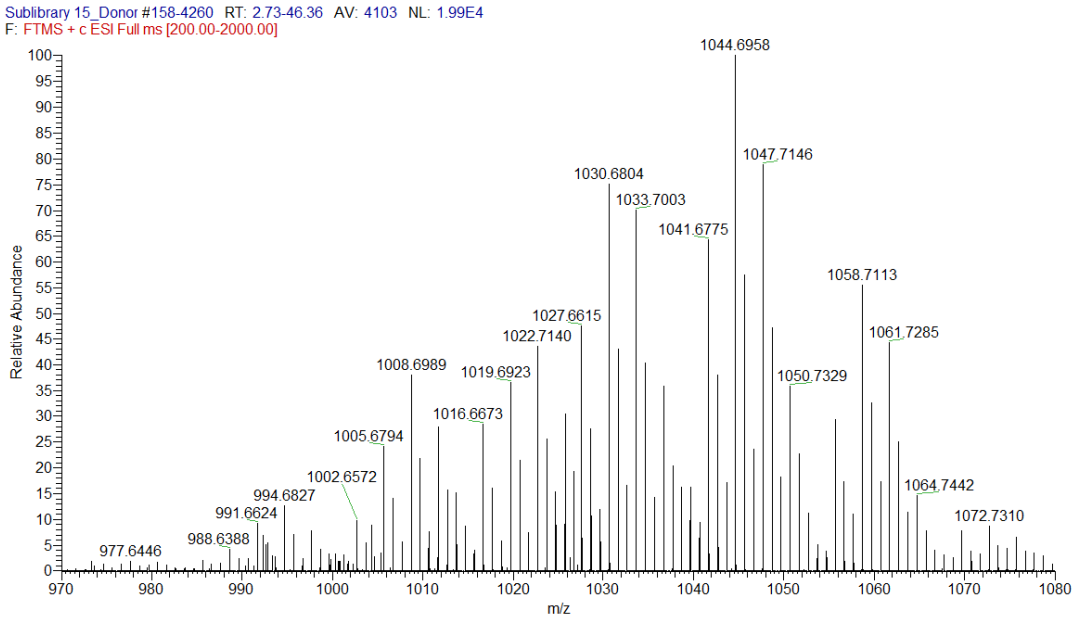
Mass spectrum of sub-library 14



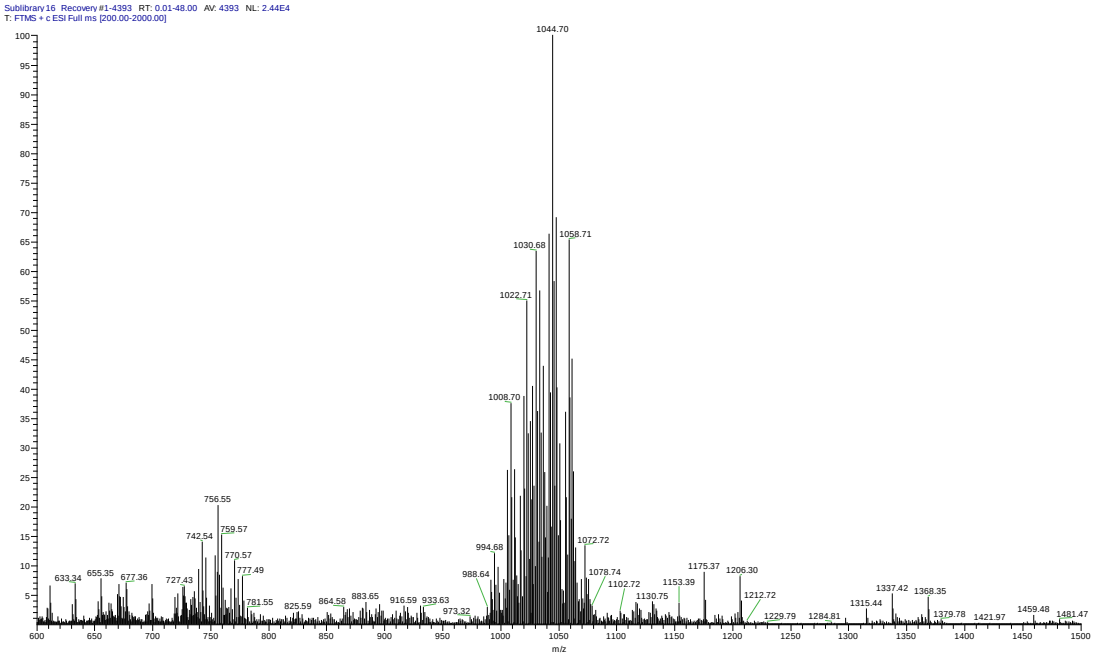


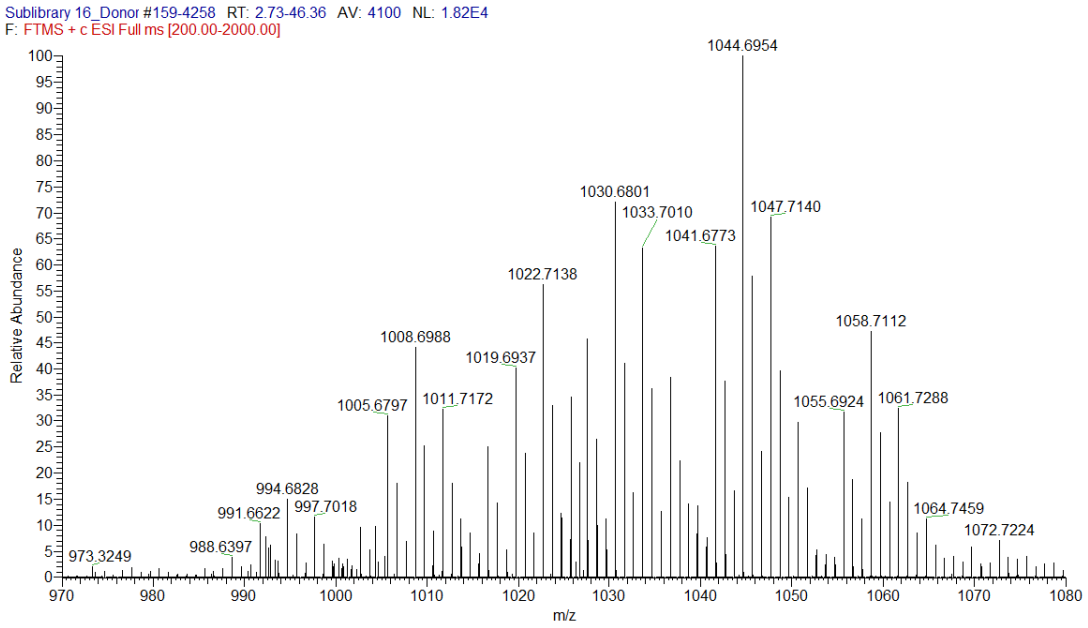
Mass spectrum of sub-library 15



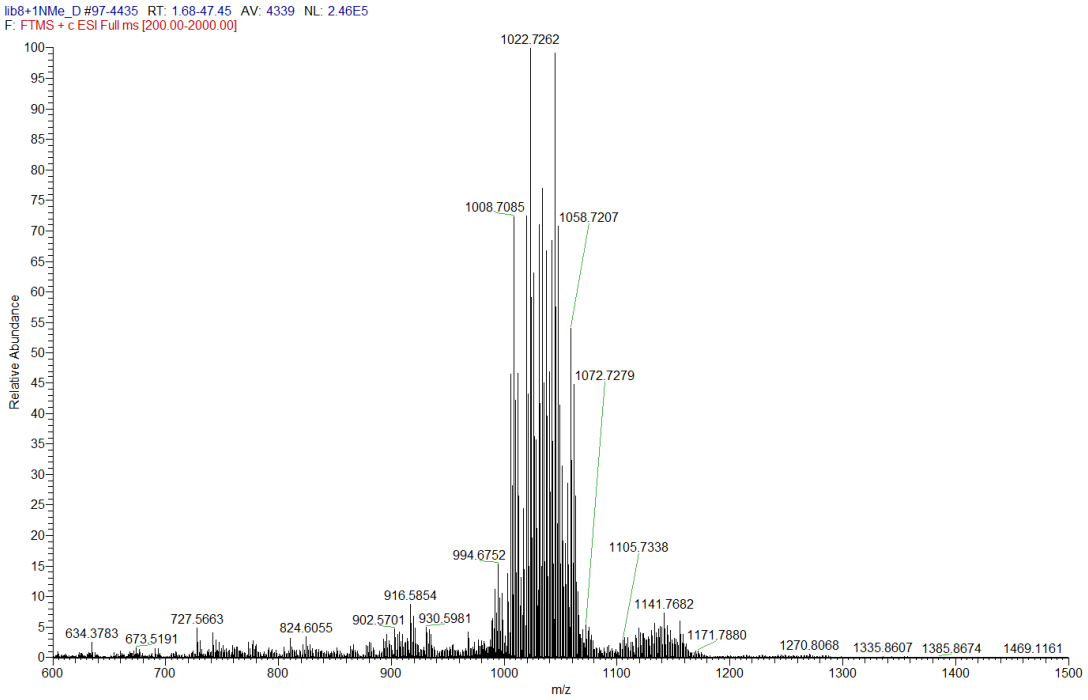


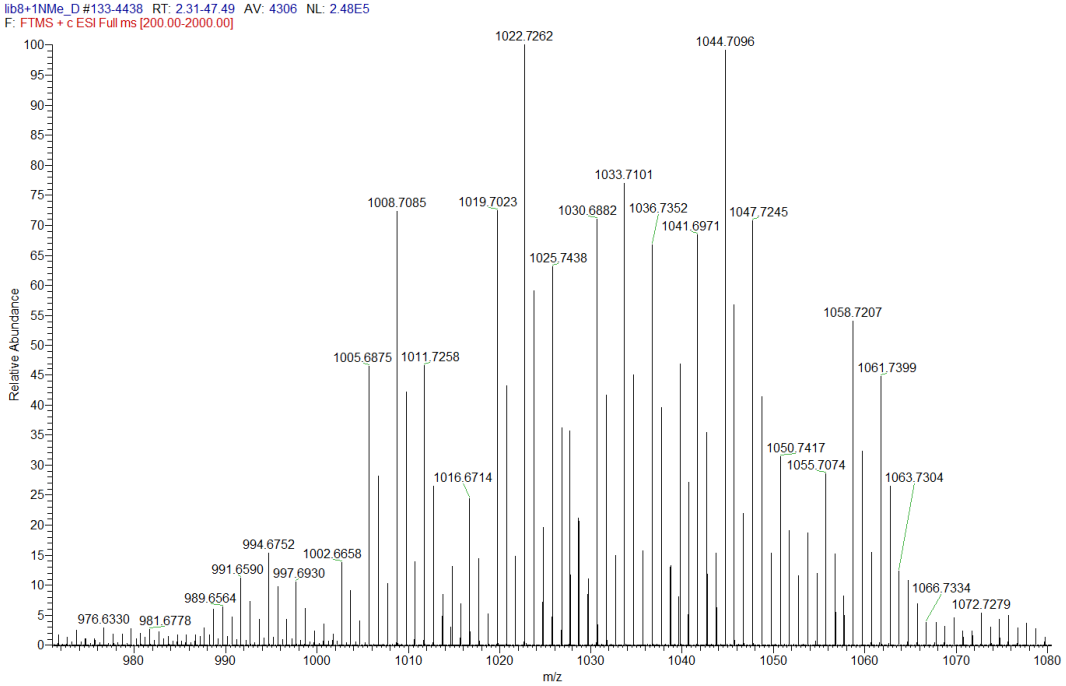
Mass spectrum of sub-library 16



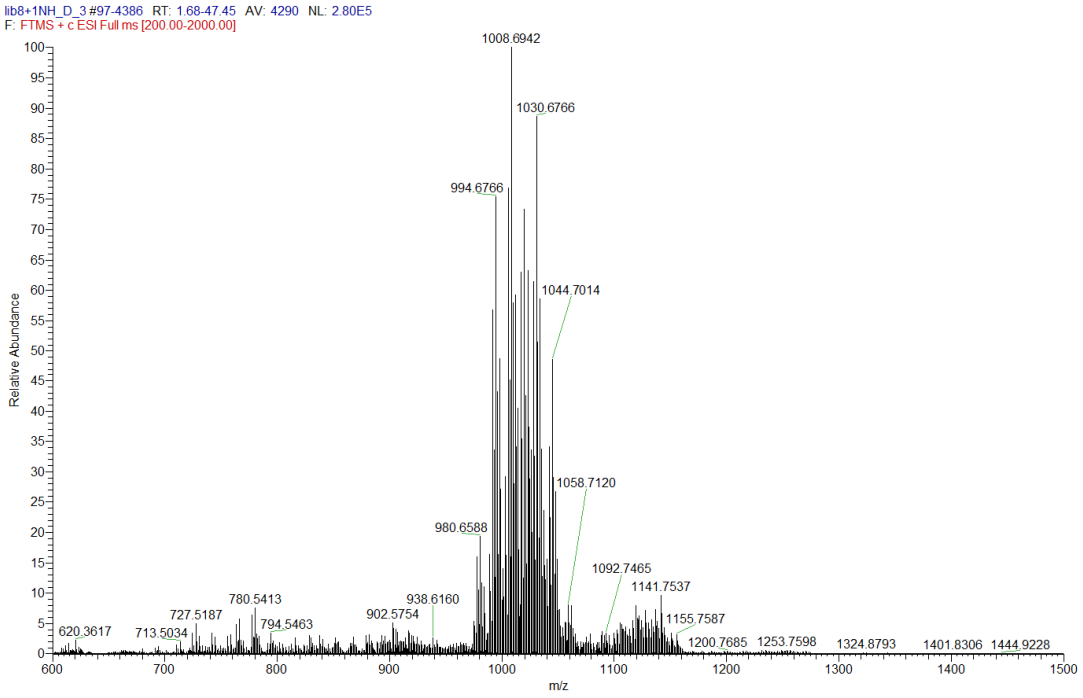


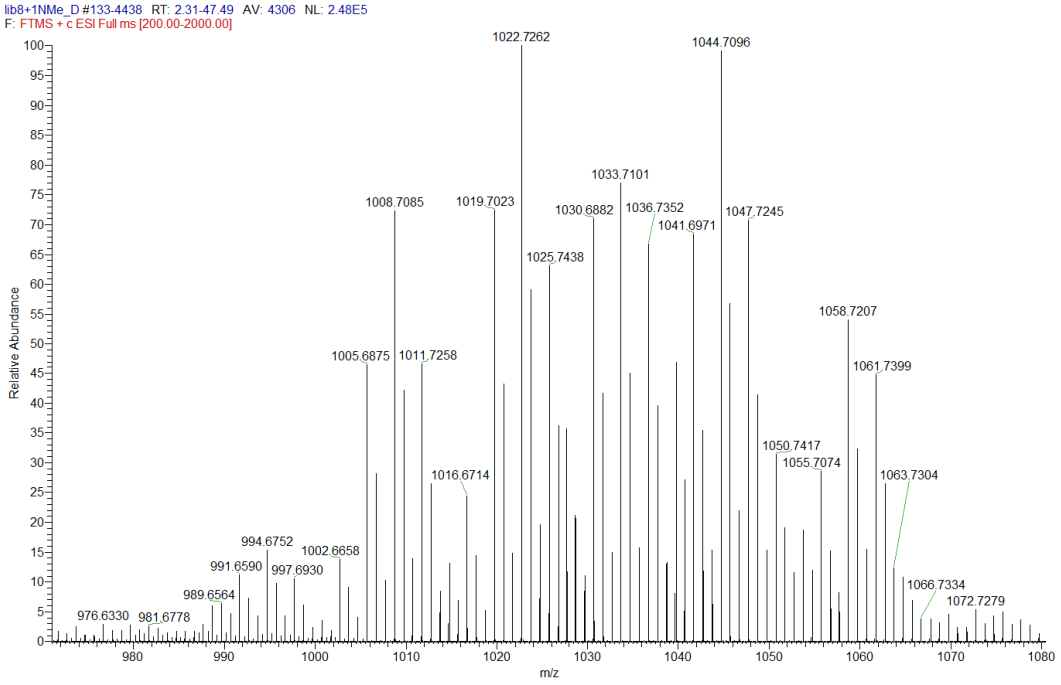
Mass spectrum of Library 2



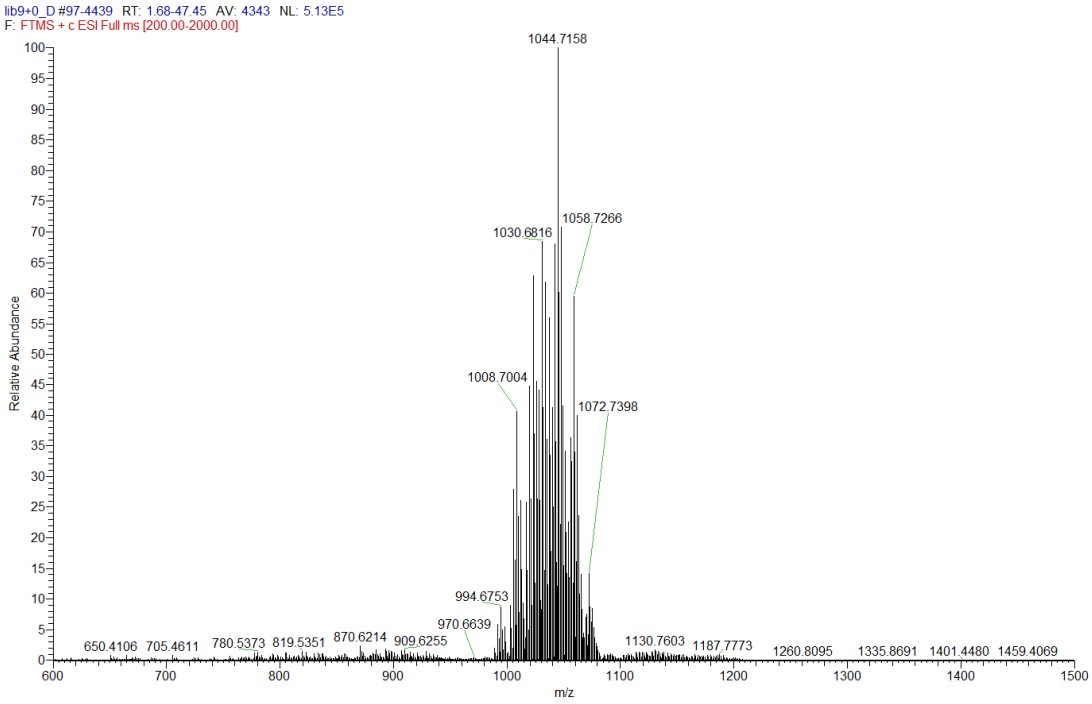


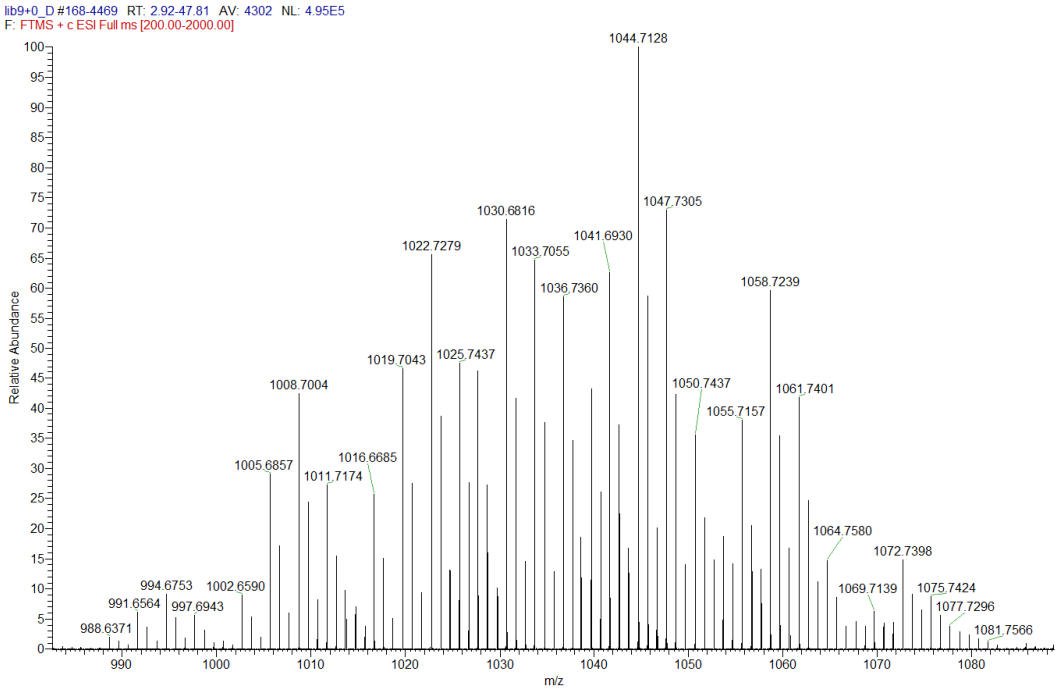
Mass spectrum of Library 3



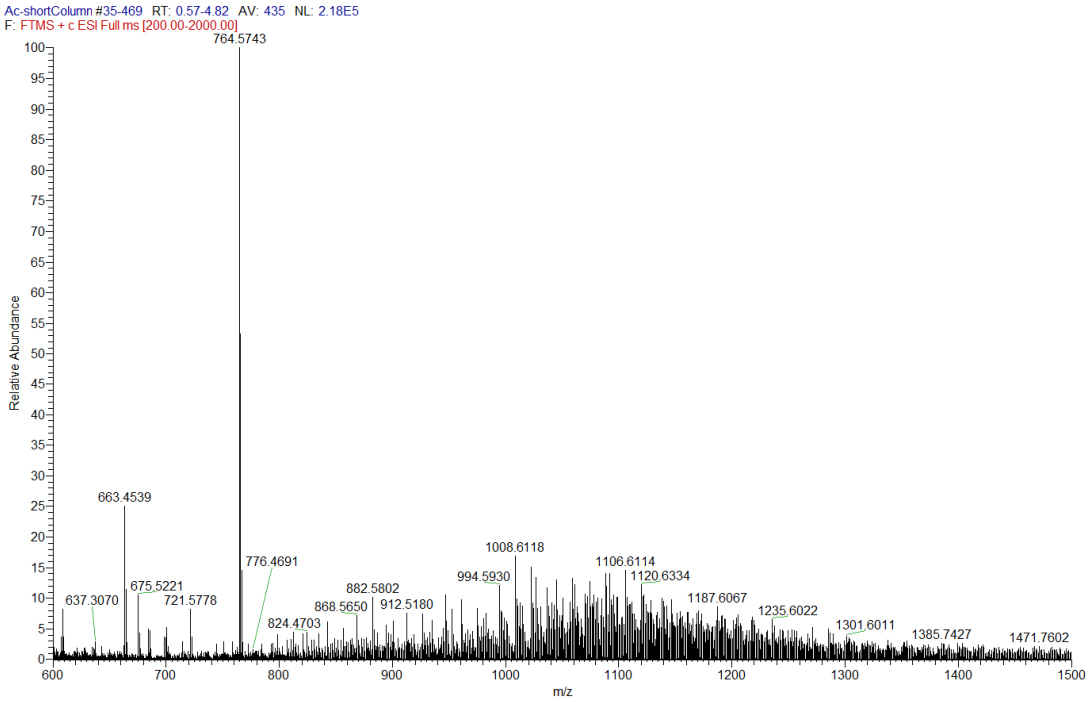


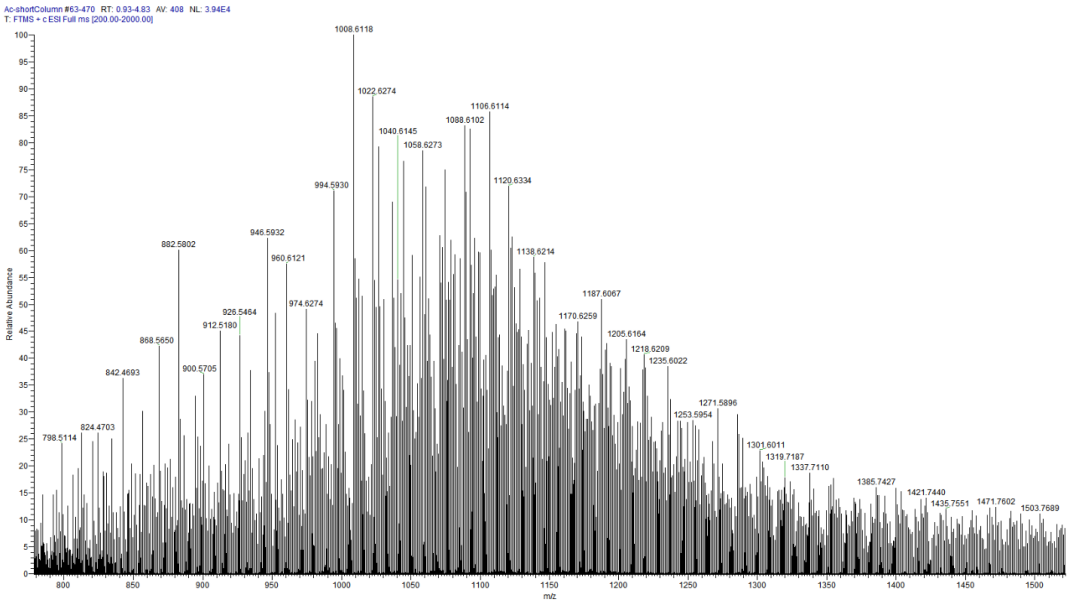
Mass spectrum of Library 4





Mass spectrum of Library 5

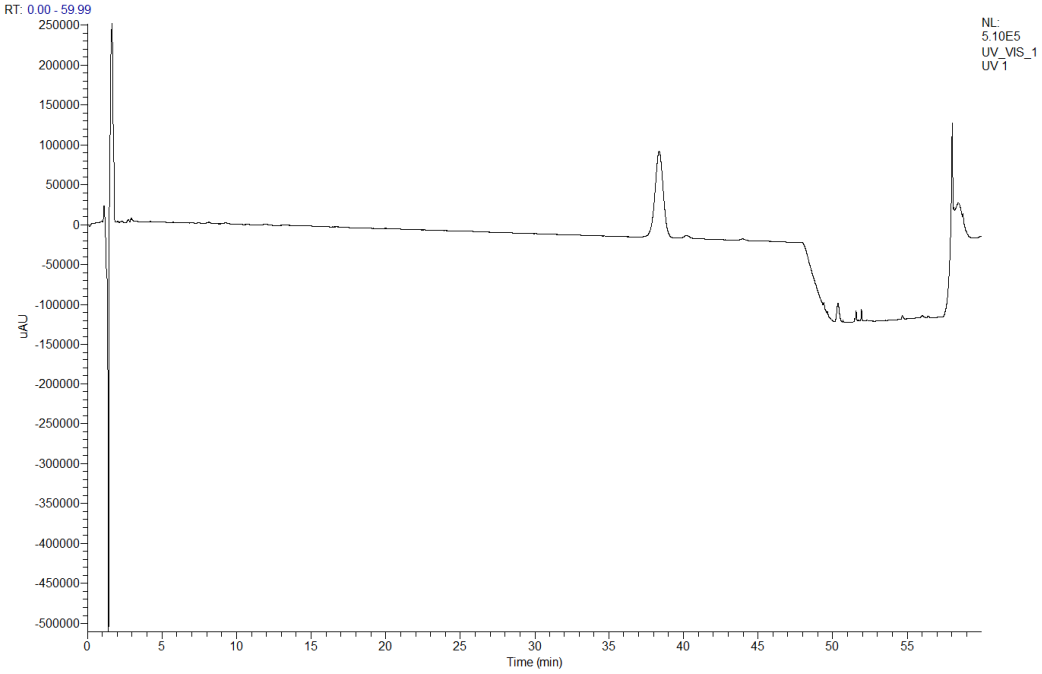


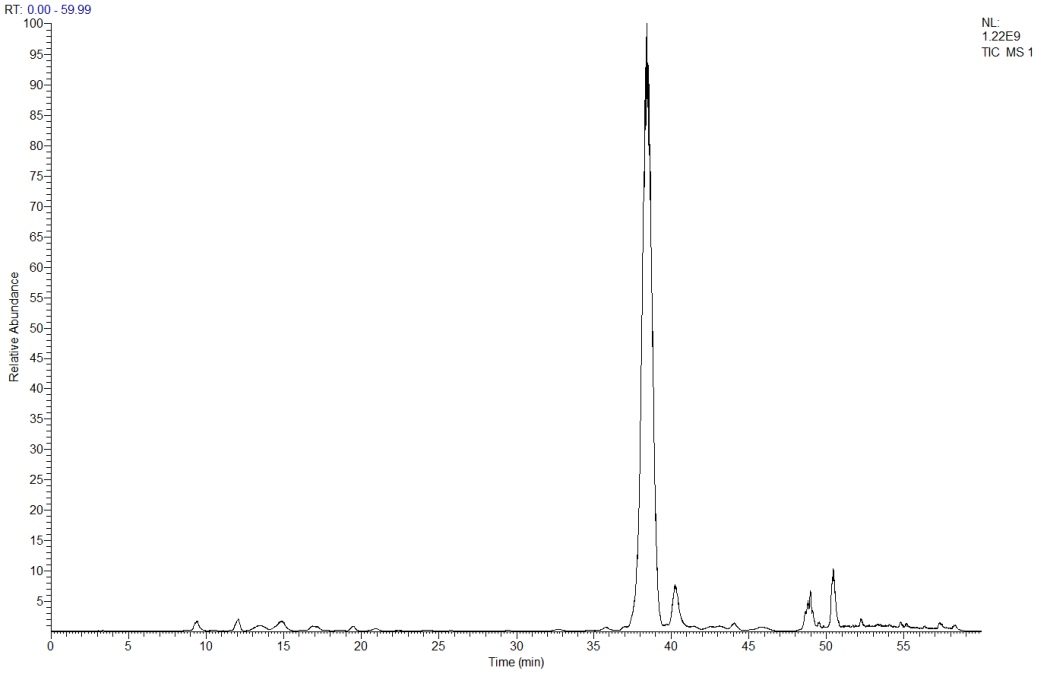


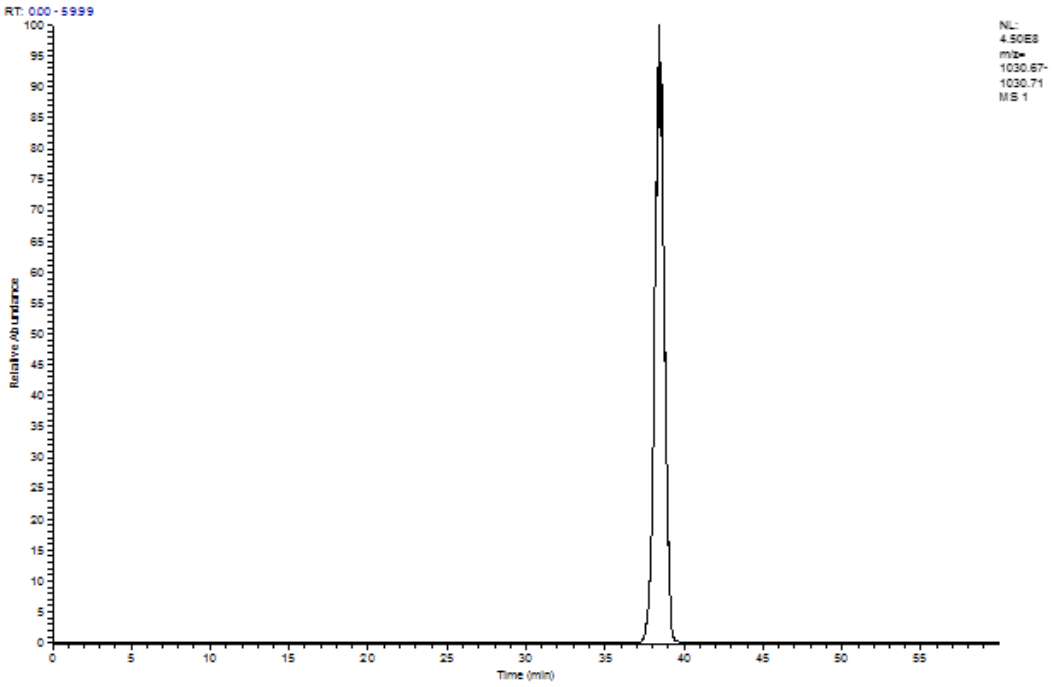
APPENDIX B **Chapter 1**

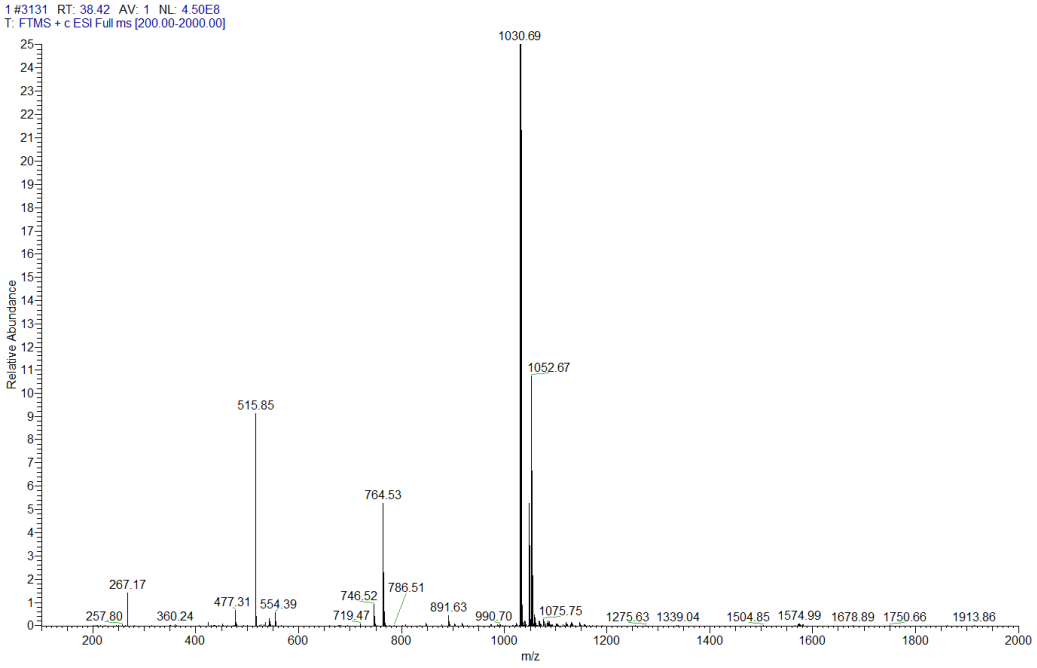
Analytical data of individual compounds (UV, TIC, selected ion, MS, H1 NMR)

LCMS data for **1**

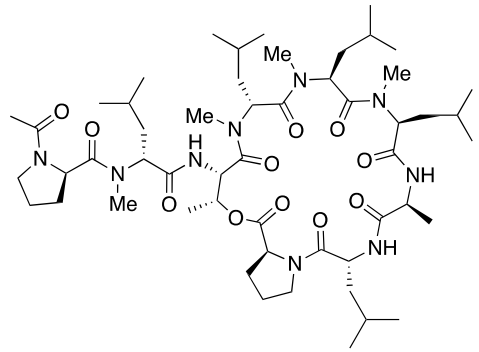


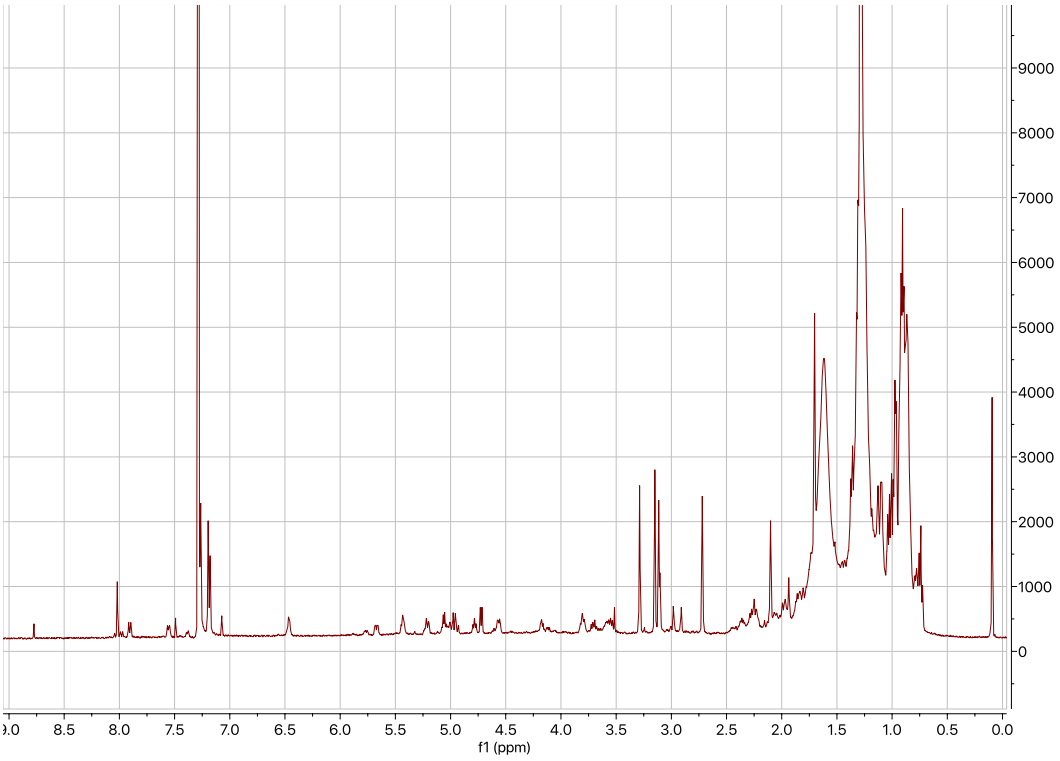



RT: 0.00 - 59.99
NL:
4.50E8
m/z=
1030.67-
1030.71
MS 1
Relative Abundance
Time (min)

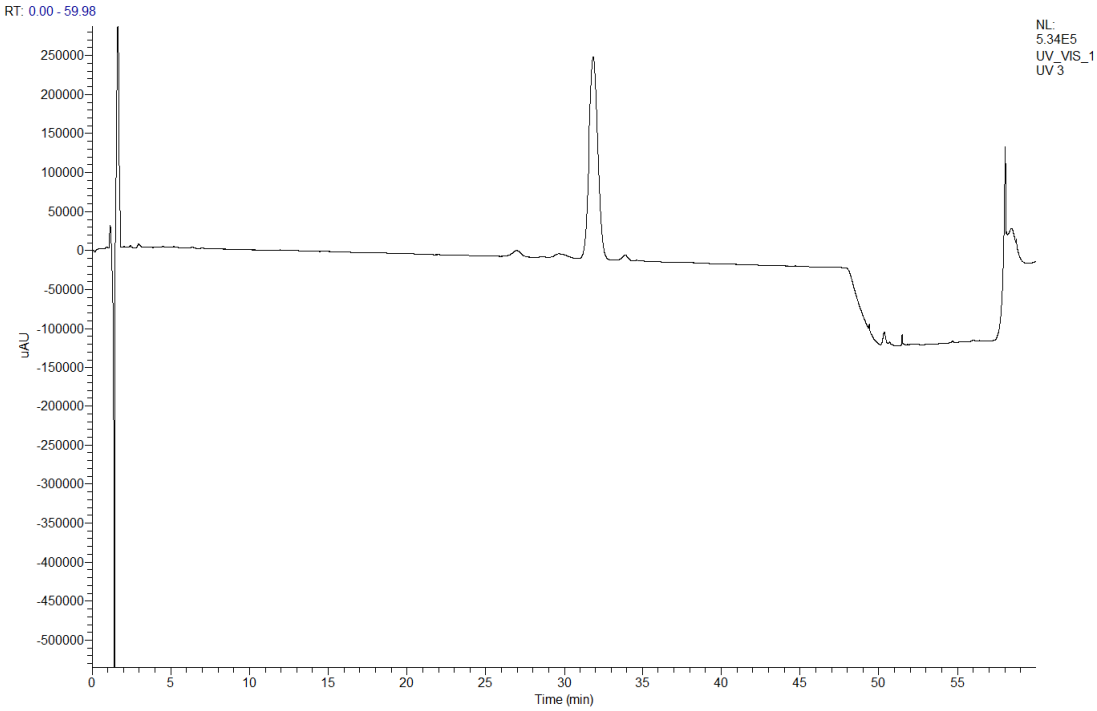

1 #3131 RT: 38.42 AV: 1 NL: 4.50E8
T: FTMS + c ESI Full ms [200.00-2000.00]
Relative Abundance
m/z
1030.69
1052.67
515.85
764.53
267.17
786.51
746.52
477.31
554.39
891.63
257.80
360.24
719.47
990.70
1075.75
1275.63
1339.04
1504.85
1574.99
1678.89
1750.66
1913.86

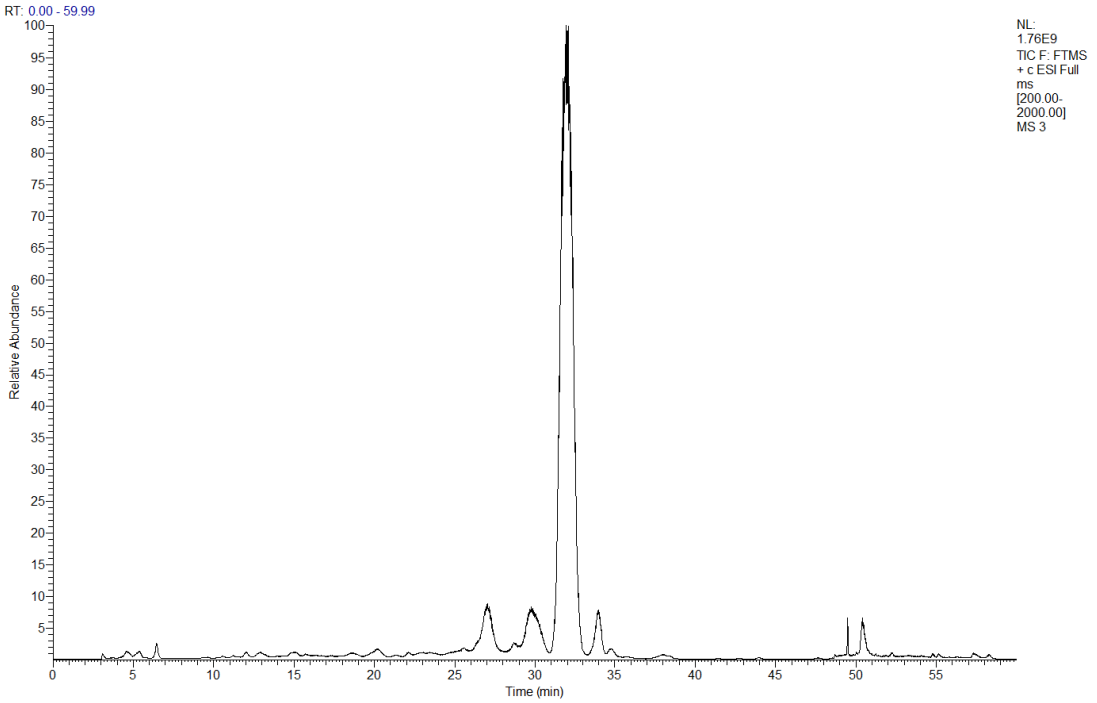
H1 NMR spectrum of **1**

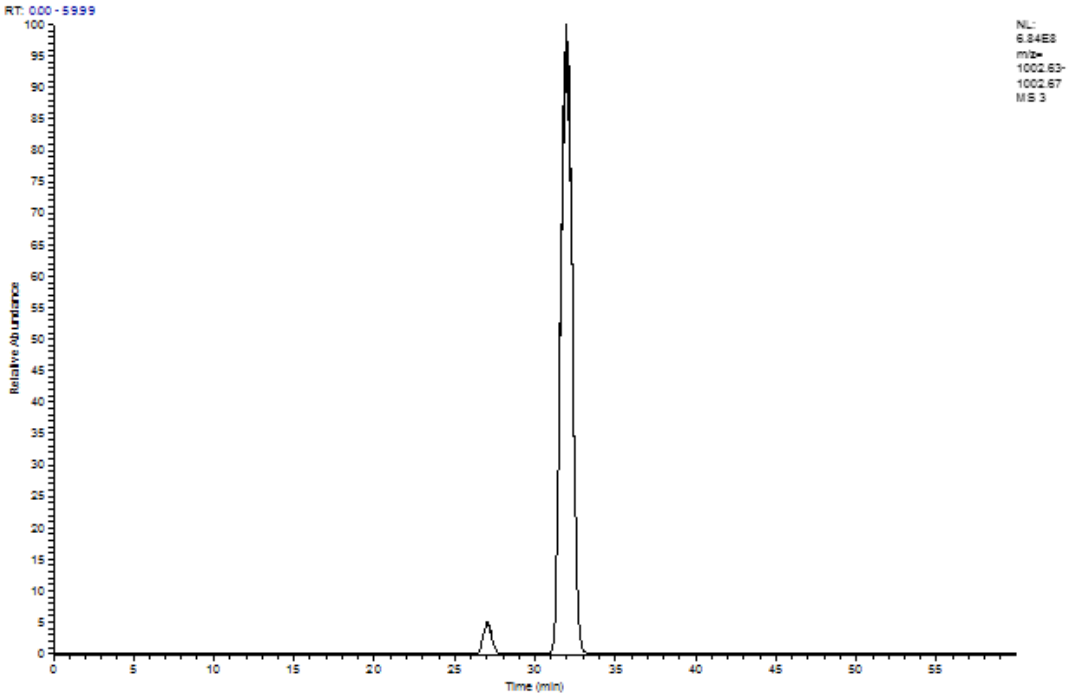


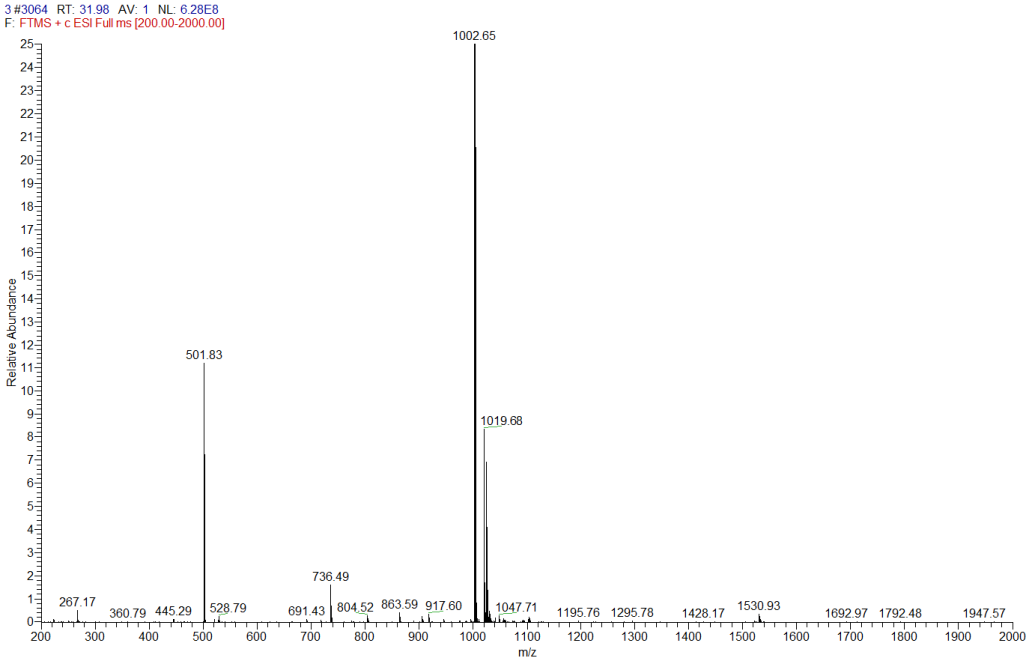


LCMS data for **2**

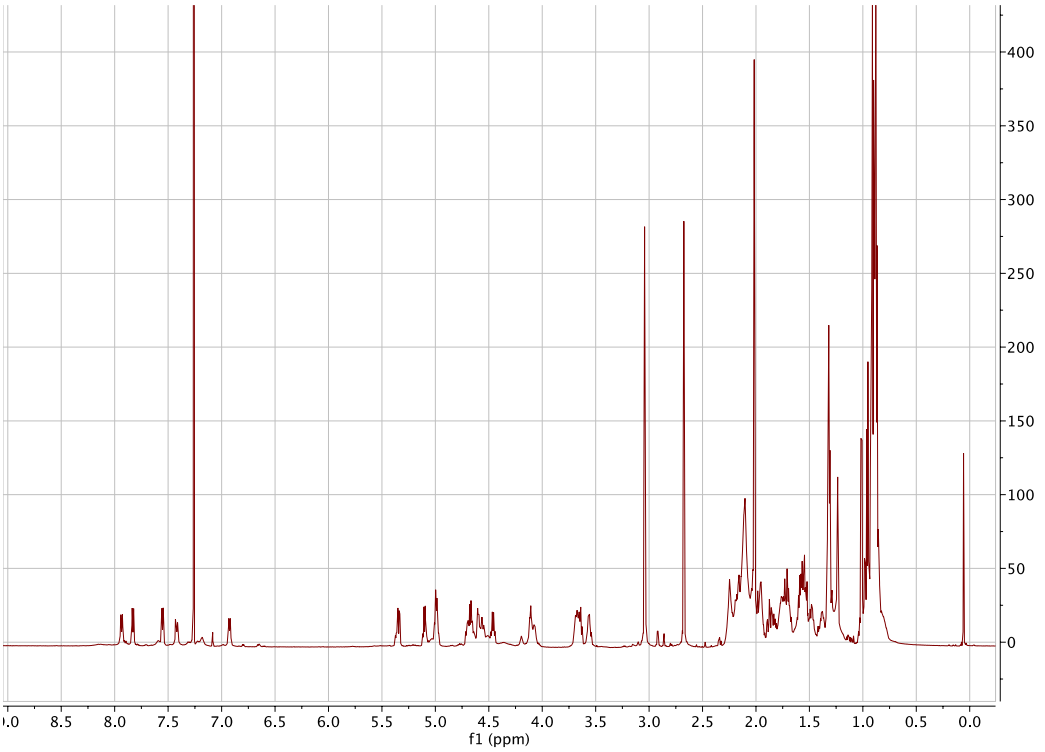


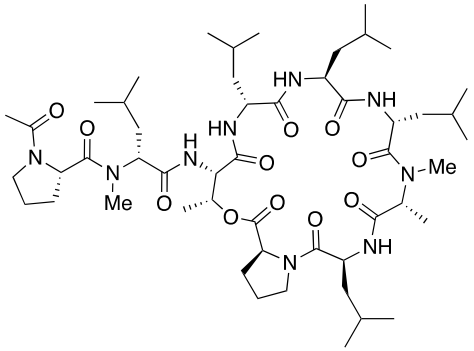



RT: 0.00 - 59.99
NL:
6.84E8
m/z=
1002.63-
1002.67
MS 3
Relative Abundance
Time (min)

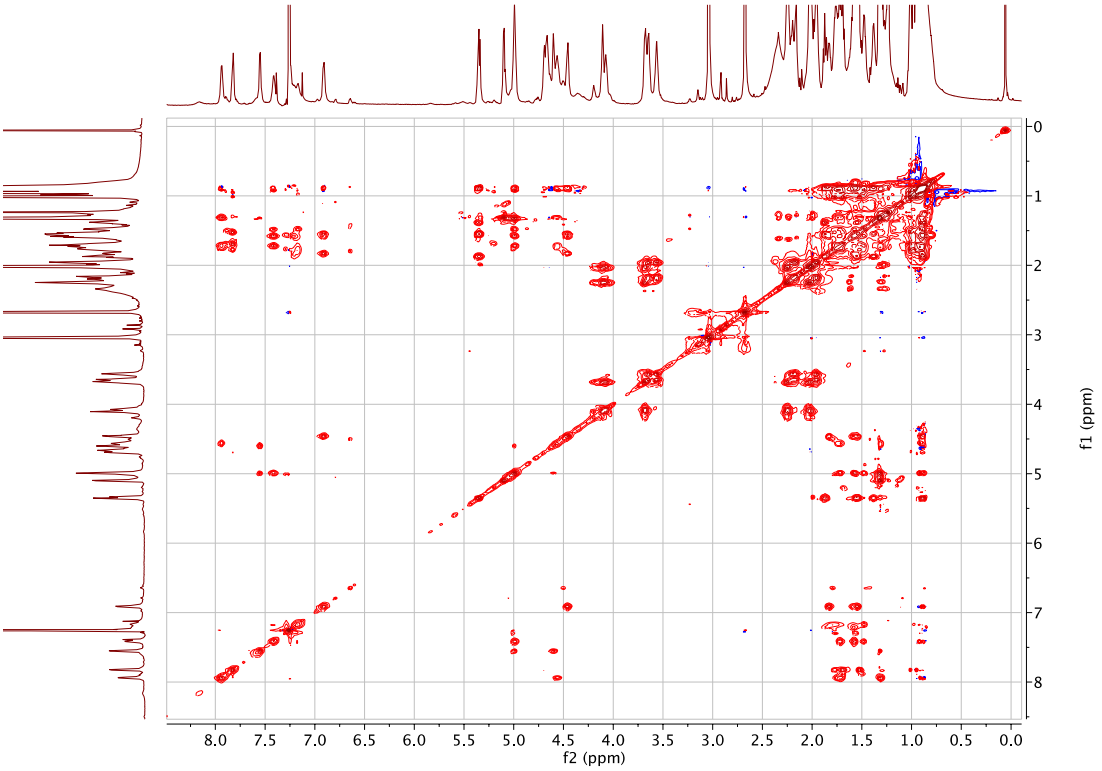

3 #3064 RT: 31.98 AV: 1 NL: 6.28E8
F: FTMS + c ESI Full ms [200.00-2000.00]
1002.65
501.83
1019.68
736.49
267.17
360.79
445.29
528.79
691.43
804.52
863.59
917.60
1047.71
1195.76
1295.78
1428.17
1530.93
1692.97
1792.48
1947.57
Relative Abundance
m/z

H1 NMR spectrum of **2**



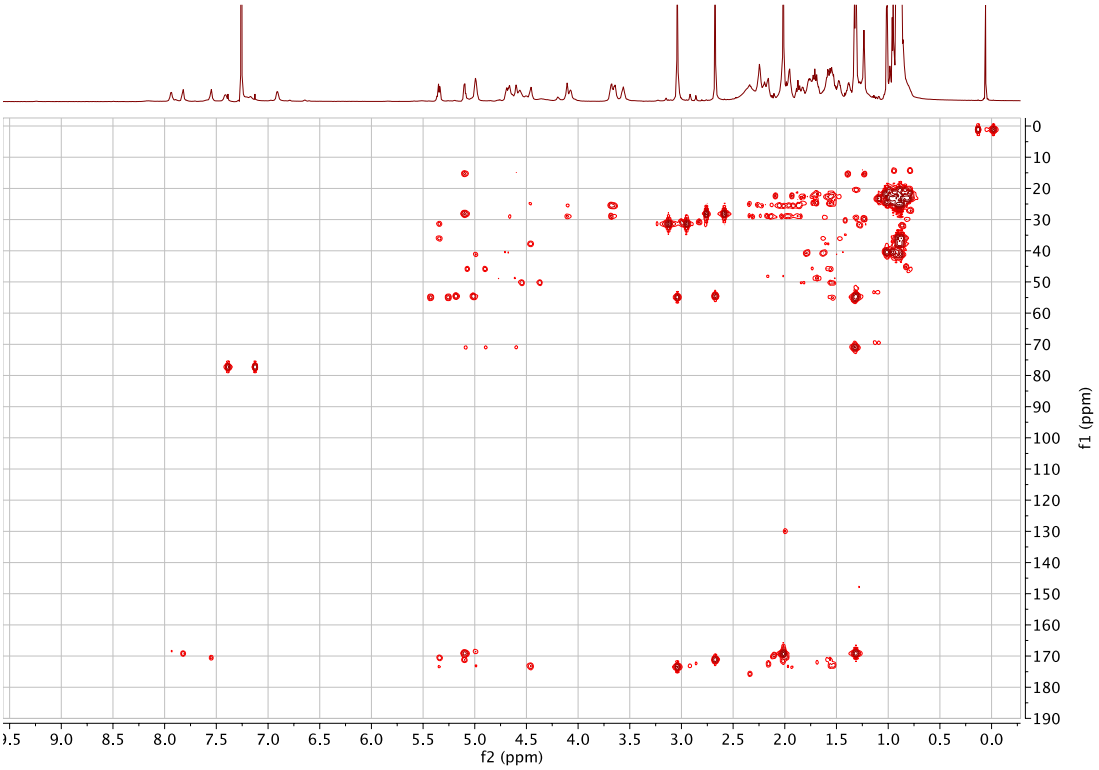


TOCSY spectrum of **2**

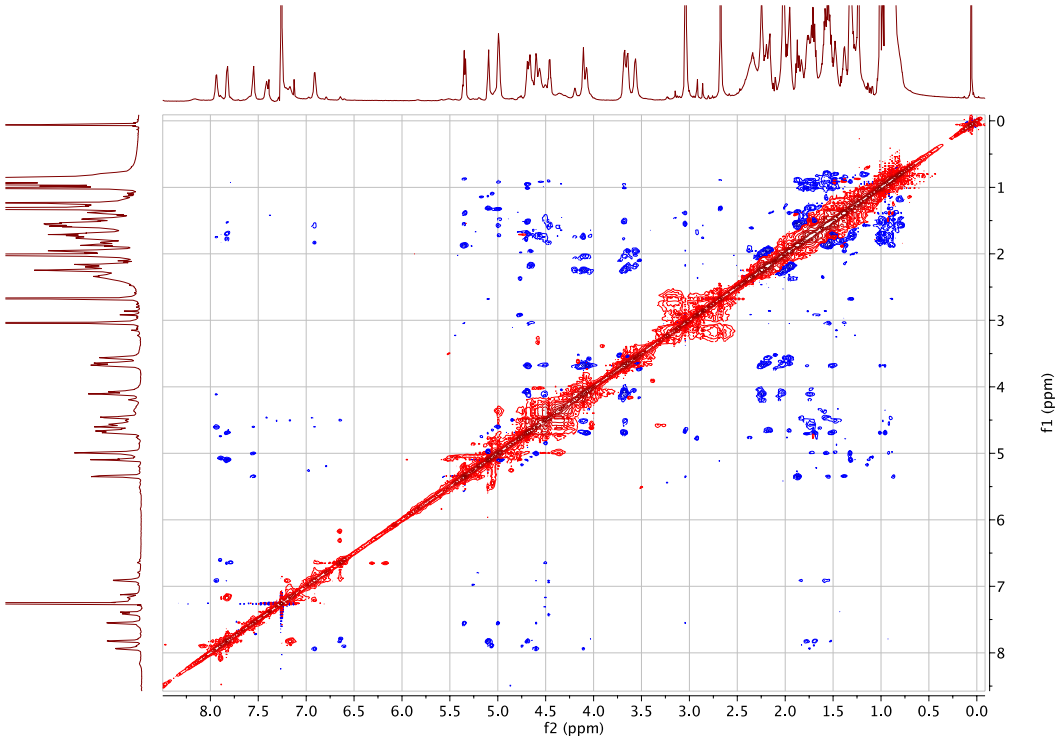


HSQC spectrum of **2**

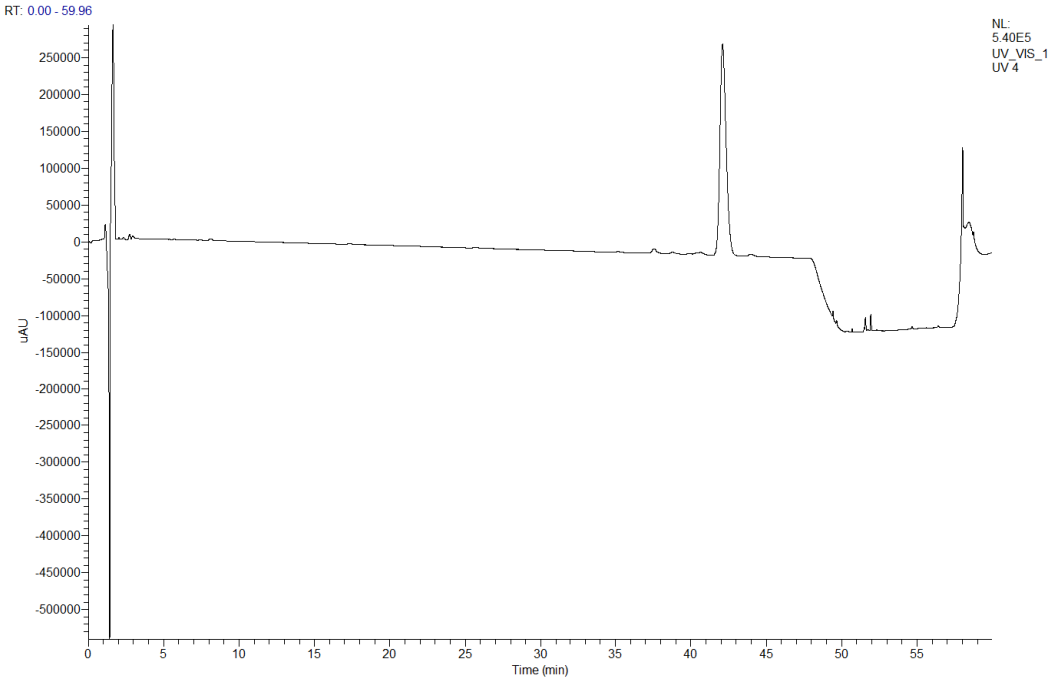
HMBC spectrum of **2**

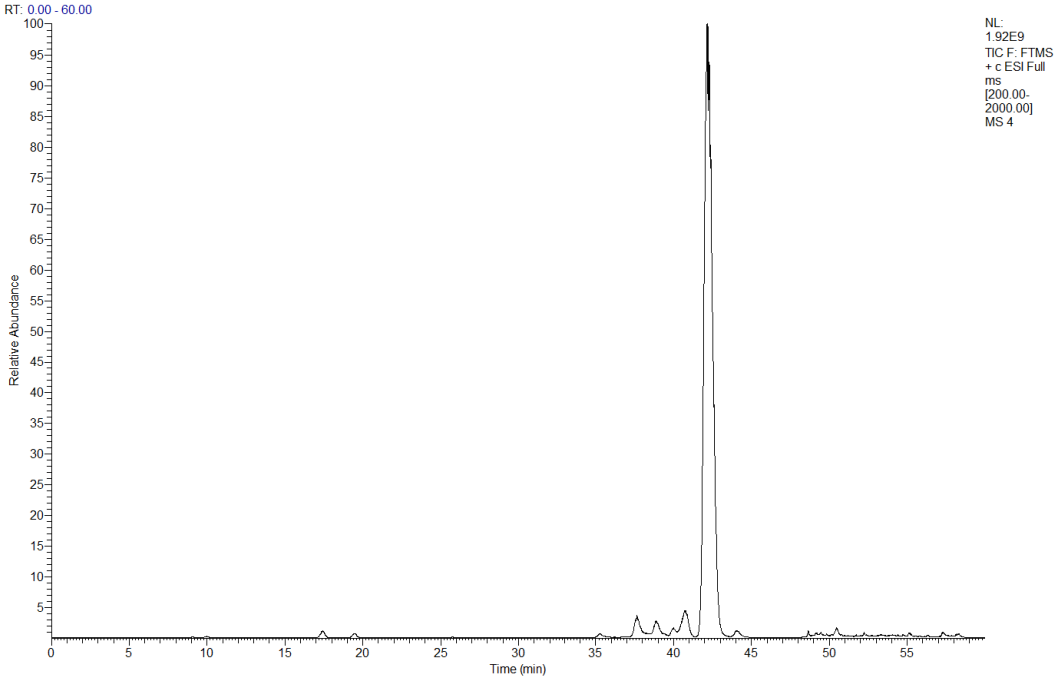


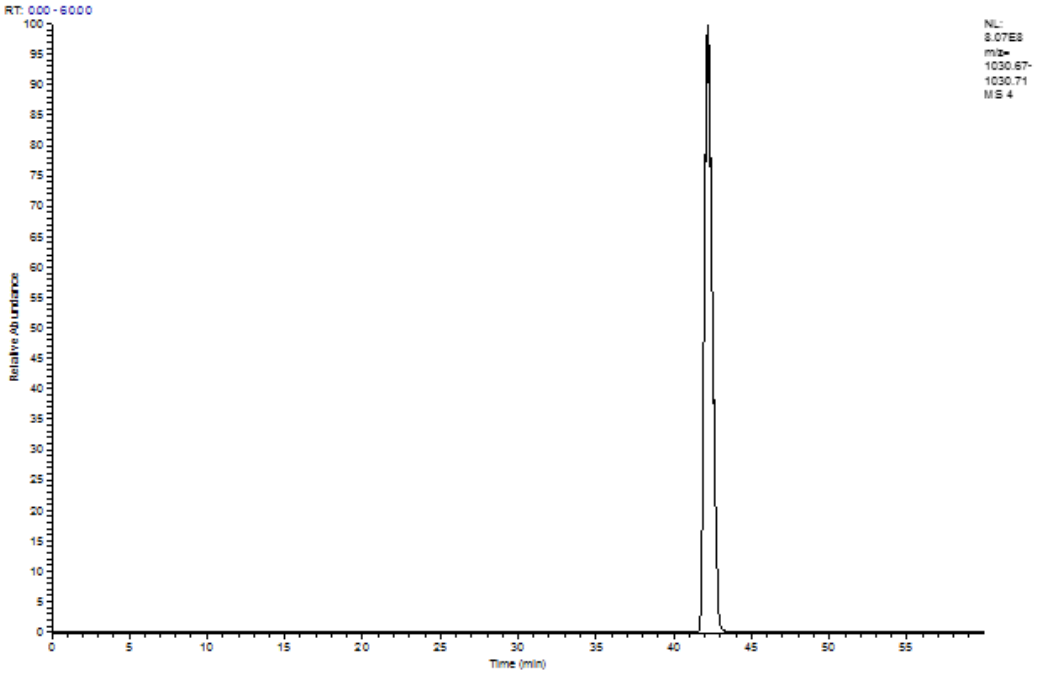
ROESY spectrum of **2**

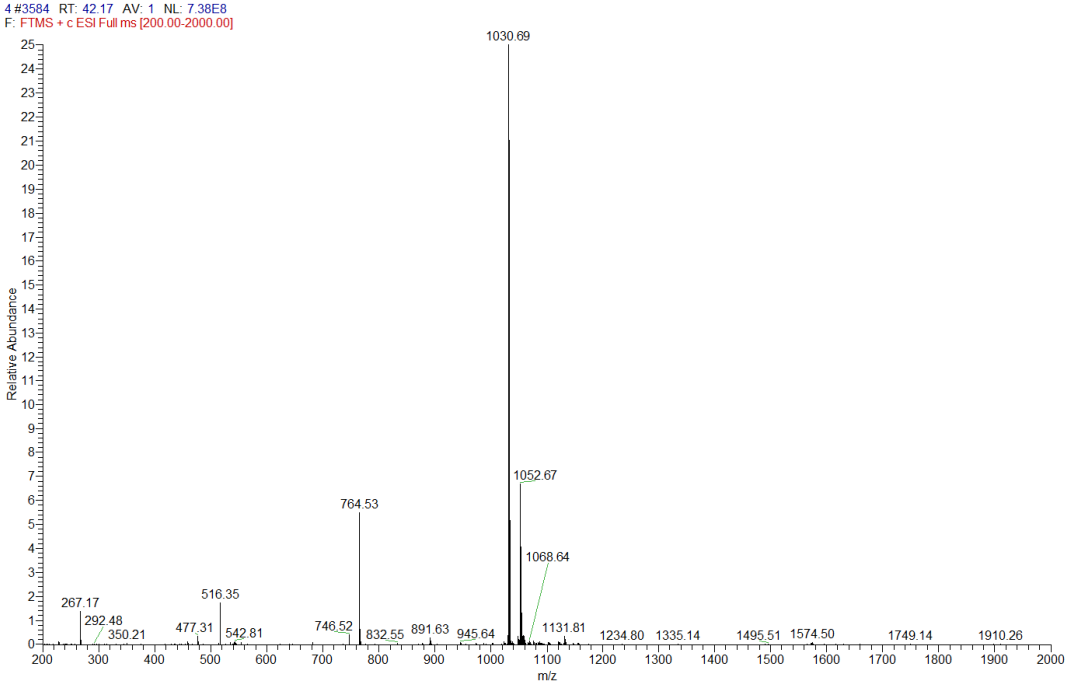


LCMS data for **3**

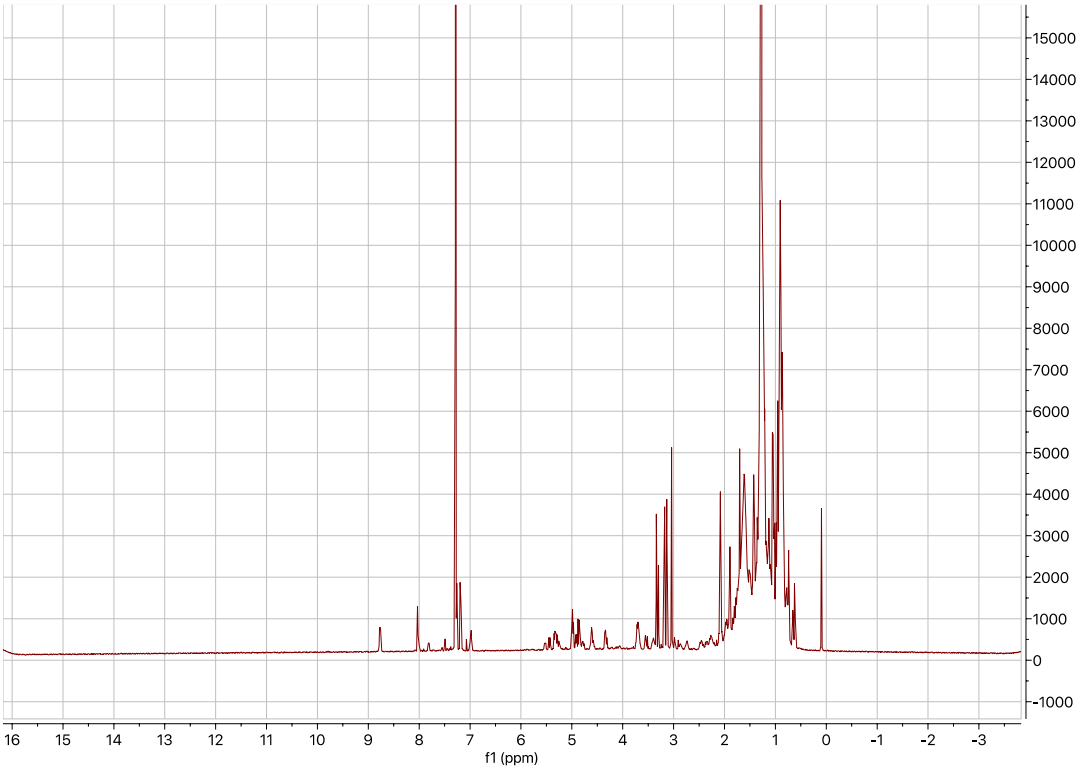


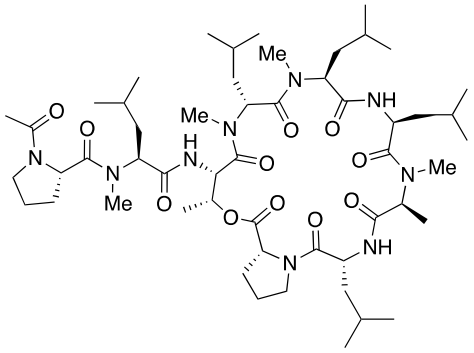



RT: 0.00 - 60.00
NL:
8.07E8
m/z=
1030.67-
1030.71
MS 4
Relative Abundance
Time (min)

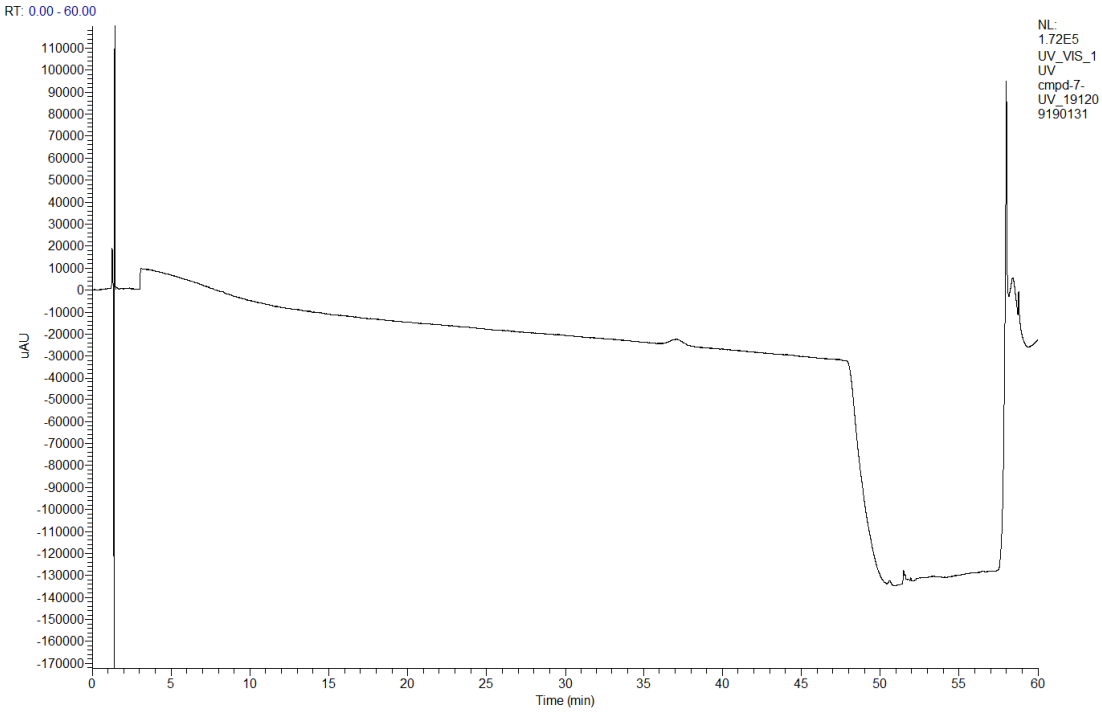

4 #3584 RT: 42.17 AV: 1 NL: 7.38E8
F: FTMS + c ESI Full ms [200.00-2000.00]
1030.69
1052.67
764.53
1068.64
516.35
267.17
292.48
350.21
477.31
542.81
746.52
832.55
891.63
945.64
1131.81
1234.80
1335.14
1495.51
1574.50
1749.14
1910.26
Relative Abundance
m/z

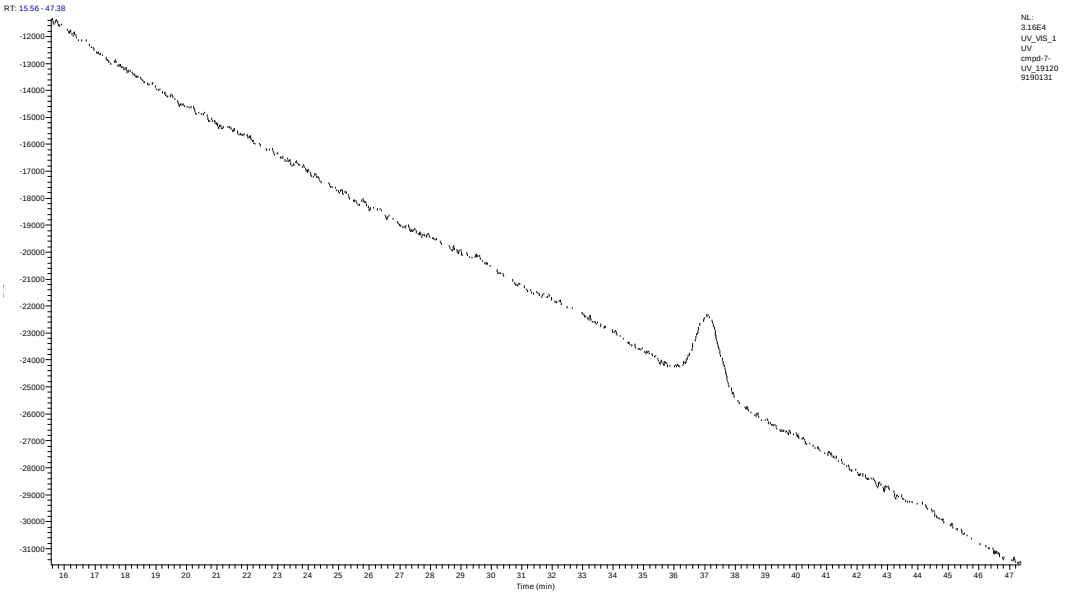
1H NMR spectrum of **3**



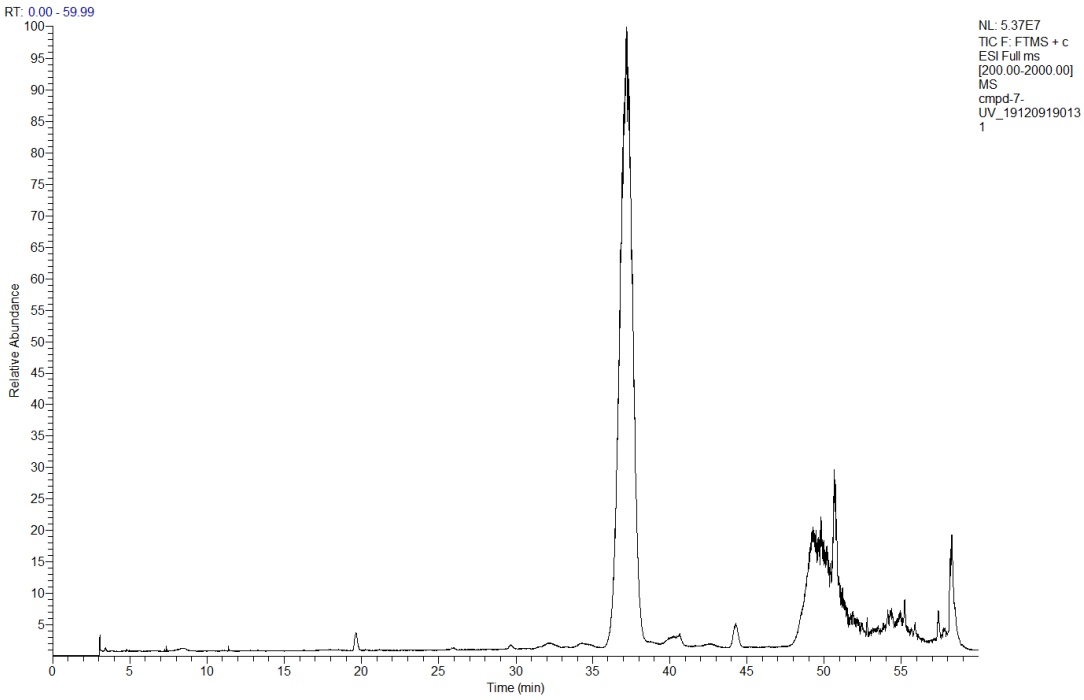


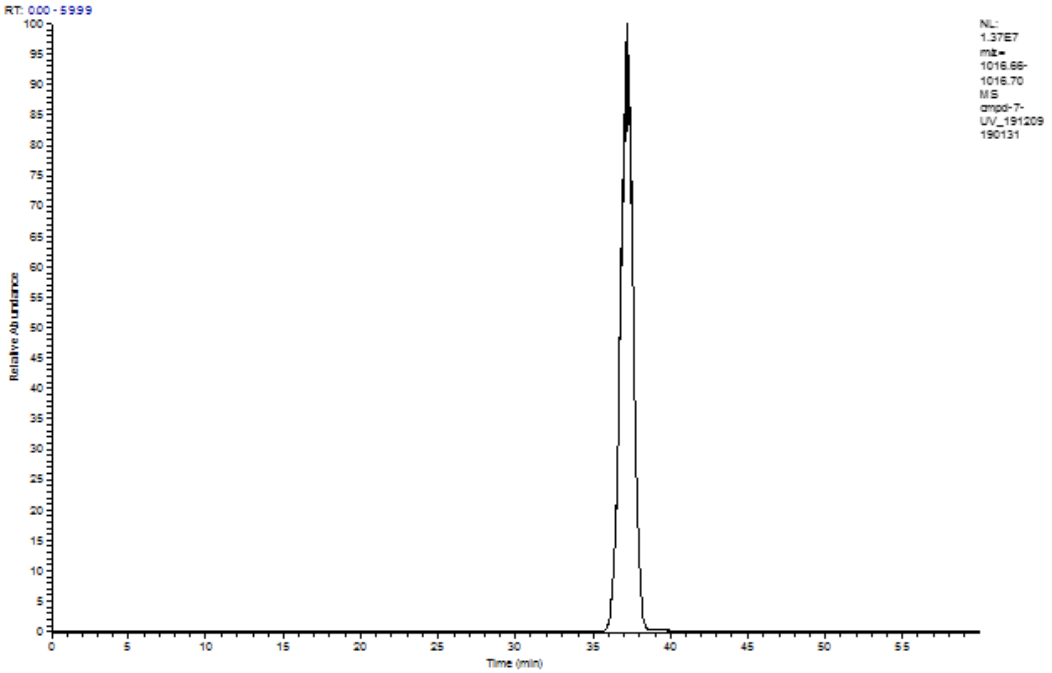
LCMS data for **4**

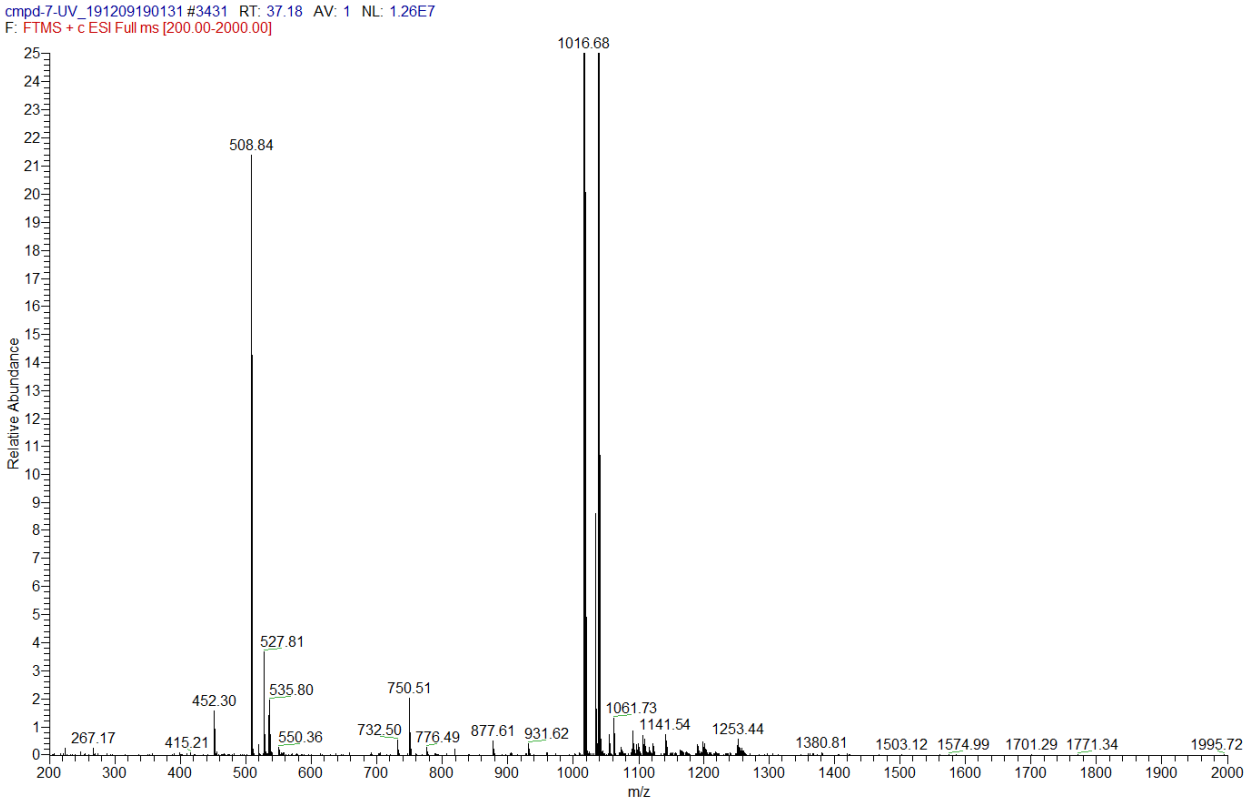




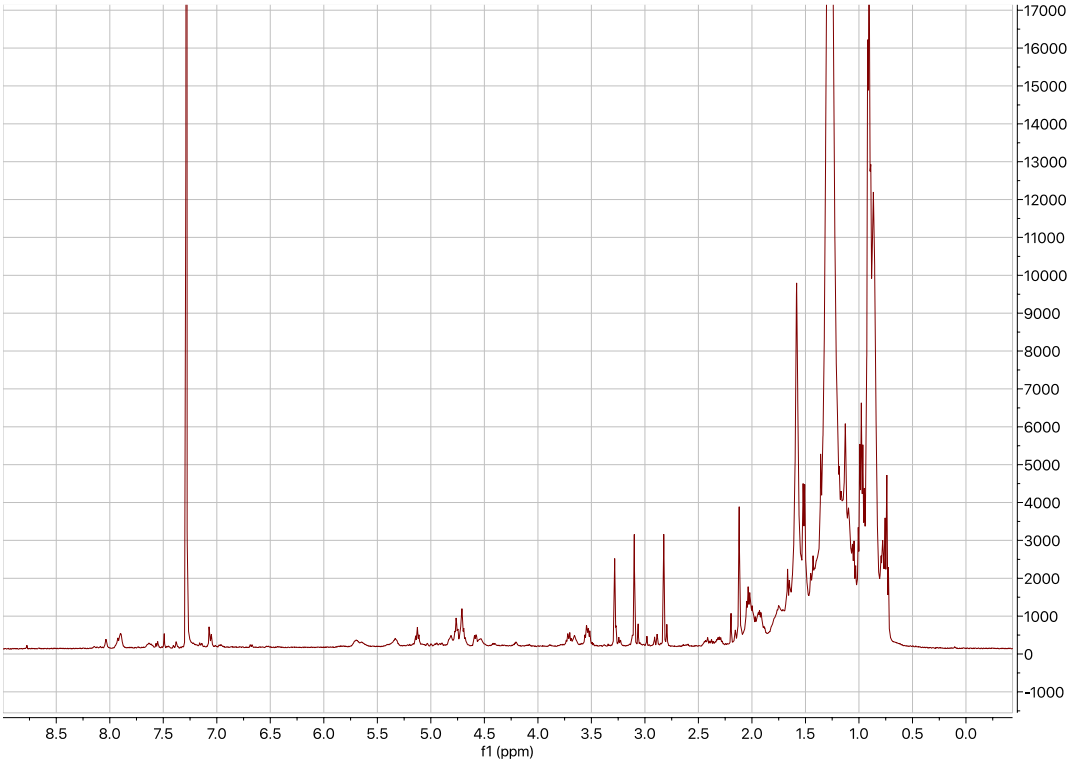
UV, 16min – 47min

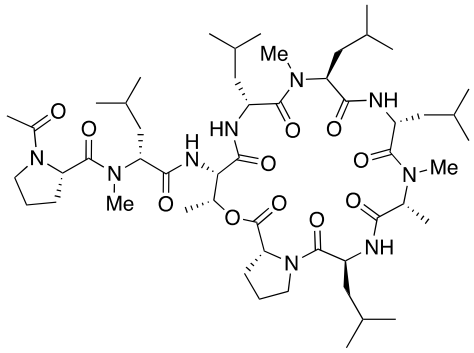

RT: 0.00 - 59.99
NL: 5.37E7
TIC F: FTMS + c
ESI Full ms
[200.00-2000.00]
MS
cmpd-7-
UV_19120919013
1
Relative Abundance
Time (min)


RT: 0.00 - 59.99
NL:
1.37E7
m/z=
1016.66-
1016.70
MS
cmpd-7-
UV_191209
190131
Relative Abundance
Time (min)

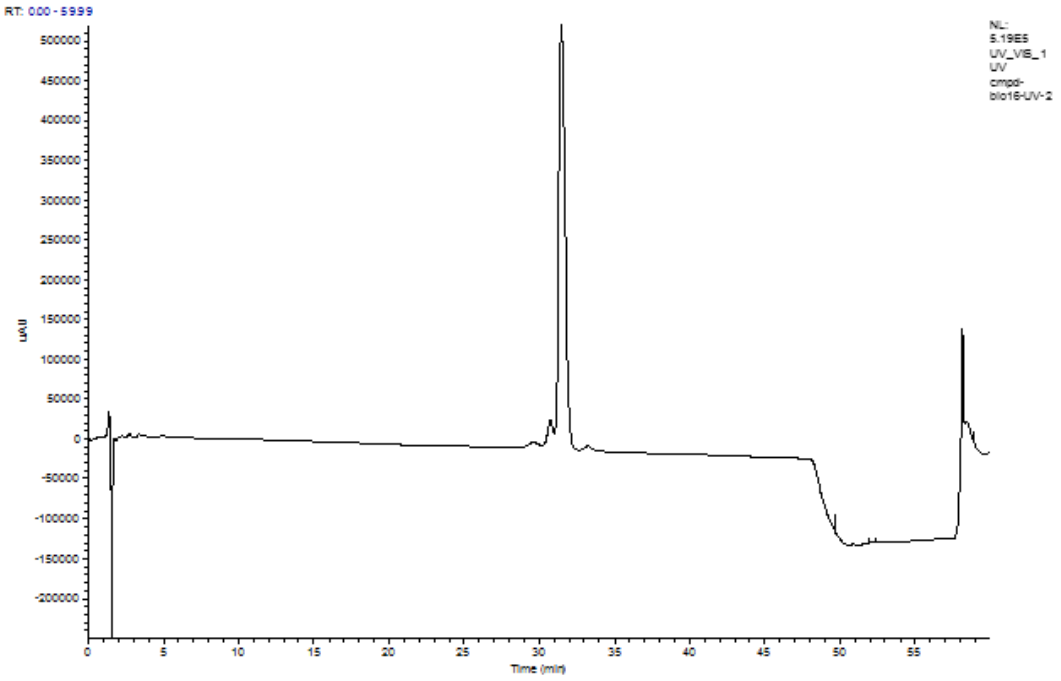

cmpd-7-UV_191209190131 #3431 RT: 37.18 AV: 1 NL: 1.26E7
F: FTMS + c ESI Full ms [200.00-2000.00]
1016.68
508.84
527.81
535.80
452.30
750.51
1061.73
1141.54
1253.44
267.17
415.21
550.36
732.50
776.49
877.61
931.62
1380.81
1503.12
1574.99
1701.29
1771.34
1995.72
Relative Abundance
m/z

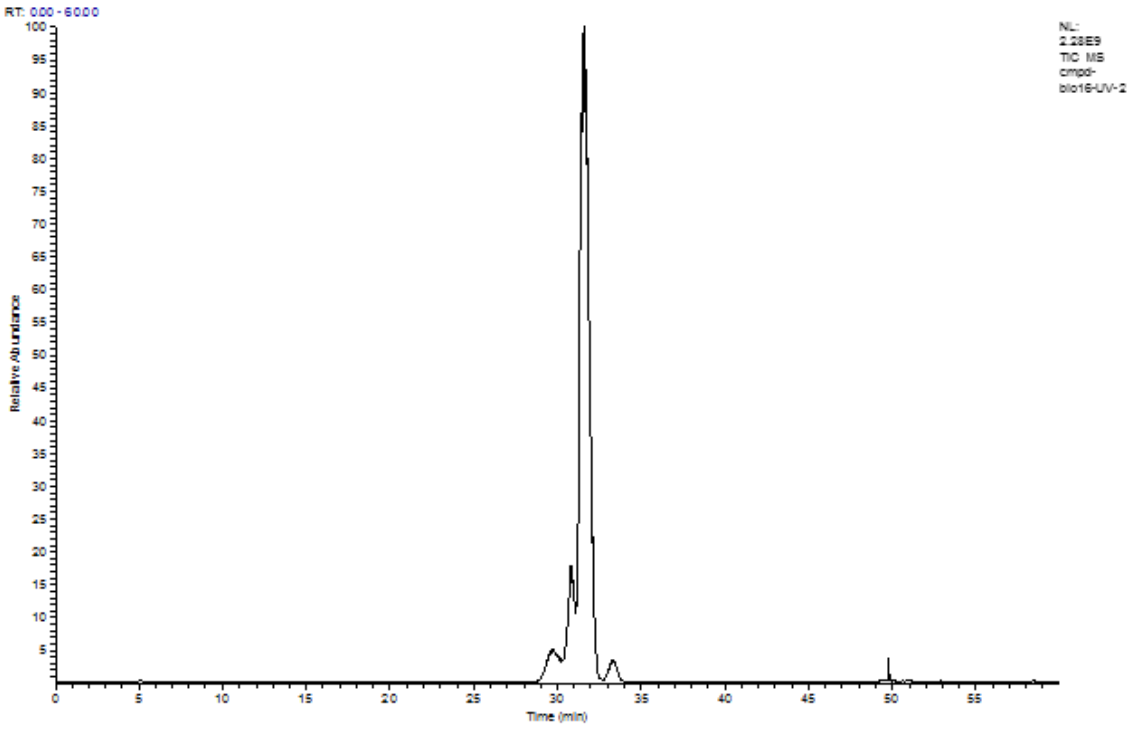
1H NMR spectrum of **4**

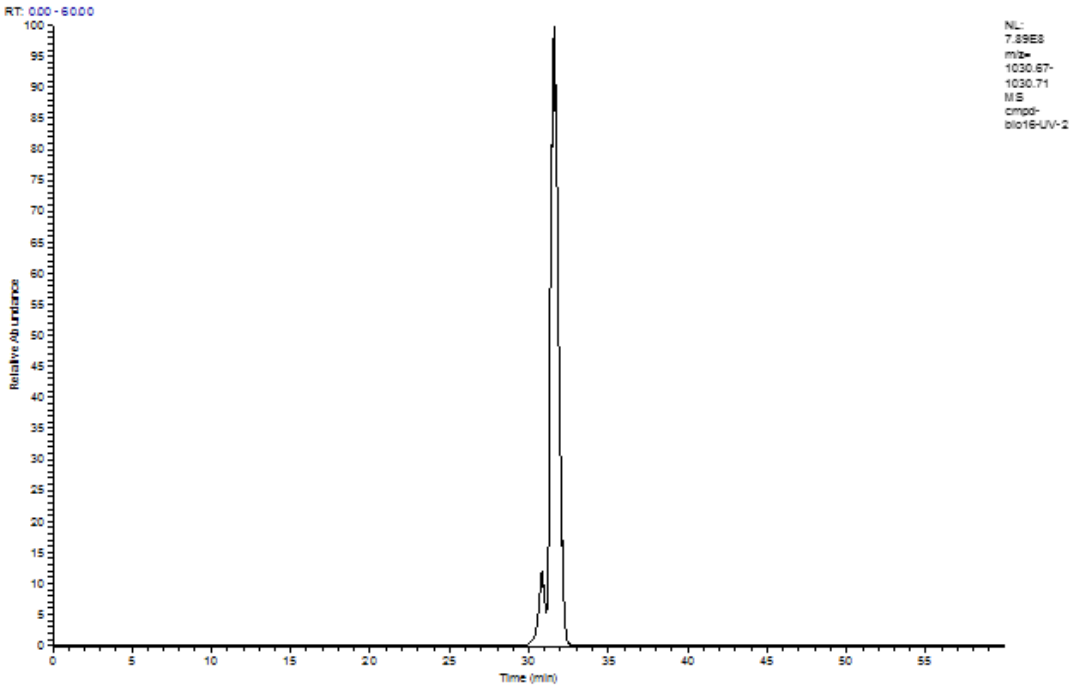


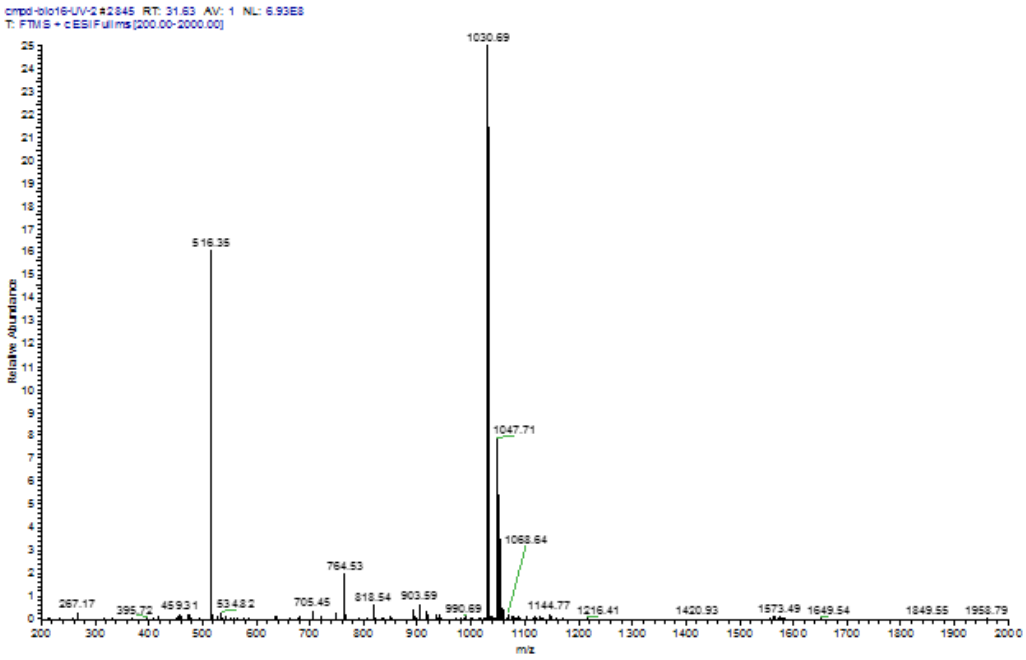


LCMS data for **5**

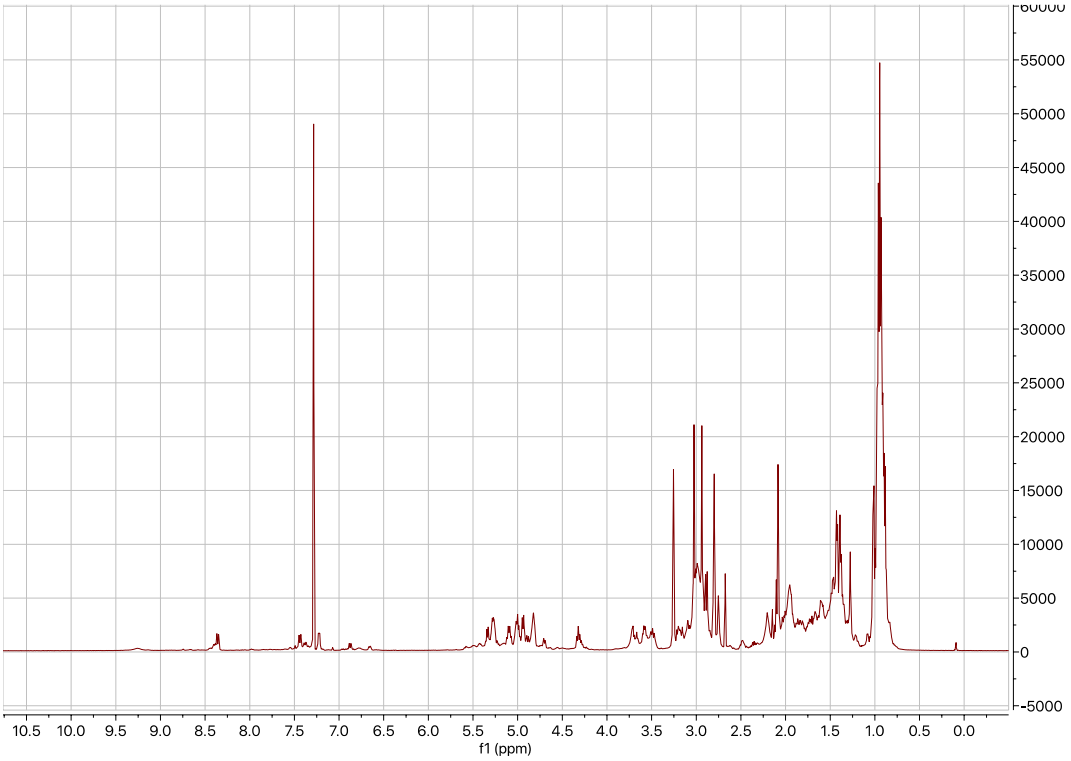


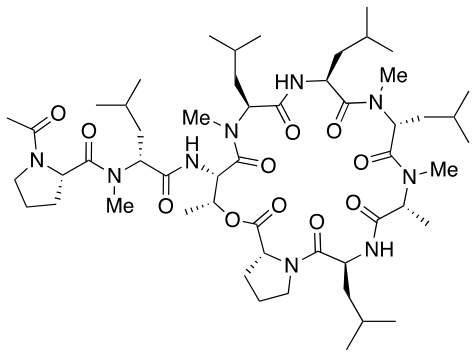



RT: 0.00 - 60.00
NL:
7.89E8
m/z=
1030.67-
1030.71
MS
cmpd-
blo16-UV-2
Relative Abundance
Time (min)

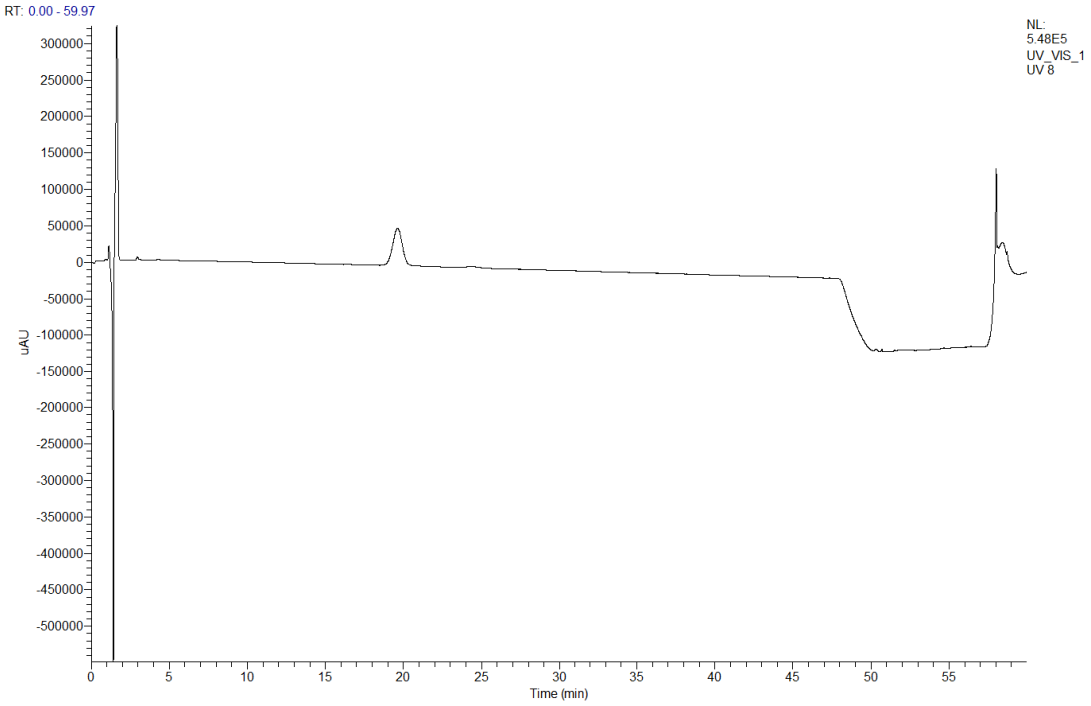

cmpd-blo16-UV-2 #2845 RT: 31.63 AV: 1 NL: 6.93E8
T: FTMS + c ESI Full ms [200.00-2000.00]
1030.69
516.35
1047.71
1068.64
764.53
287.17
395.72
459.31
534.82
705.45
813.54
903.59
990.69
1144.77
1216.41
1420.93
1573.49
1649.54
1849.55
1958.79
Relative Abundance
m/z

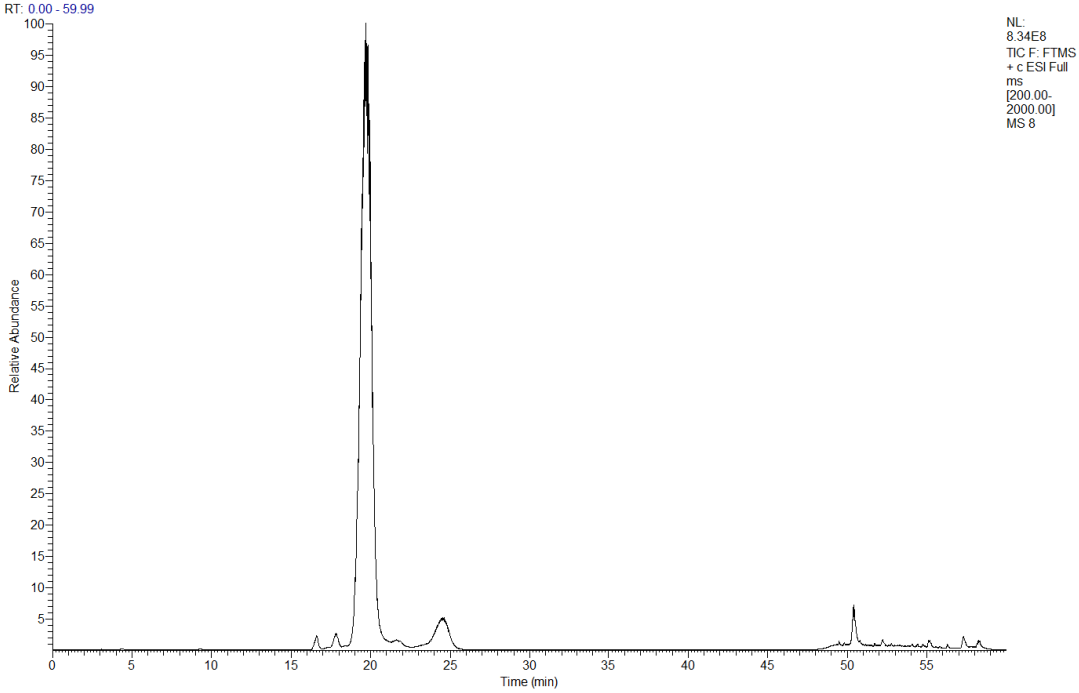
H1 NMR spectrum of **5**

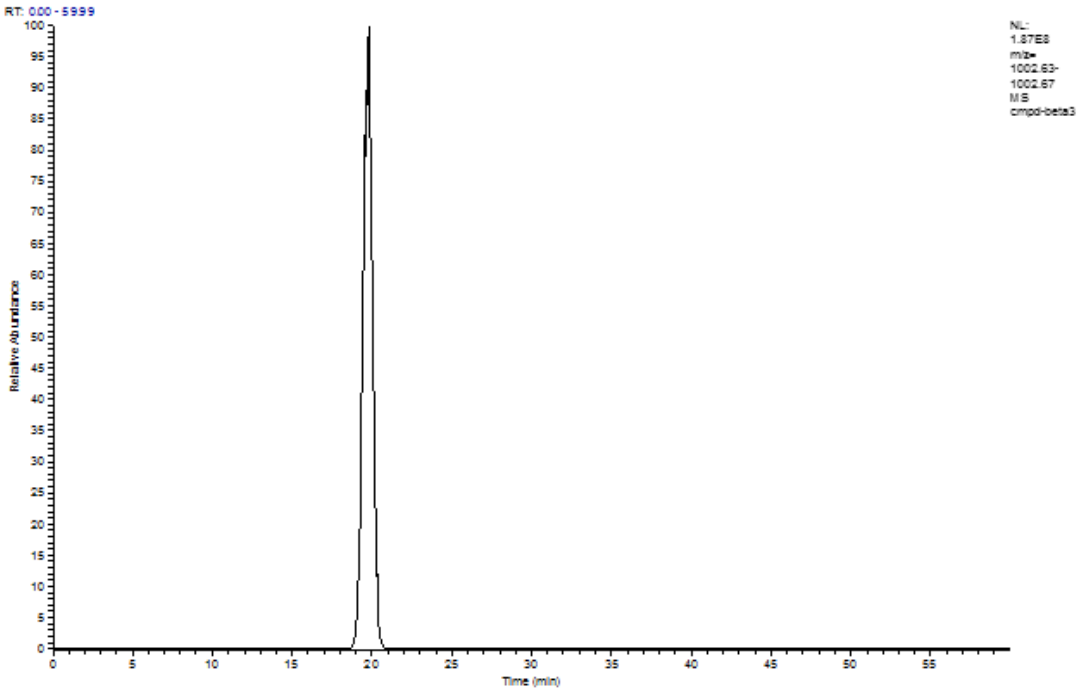


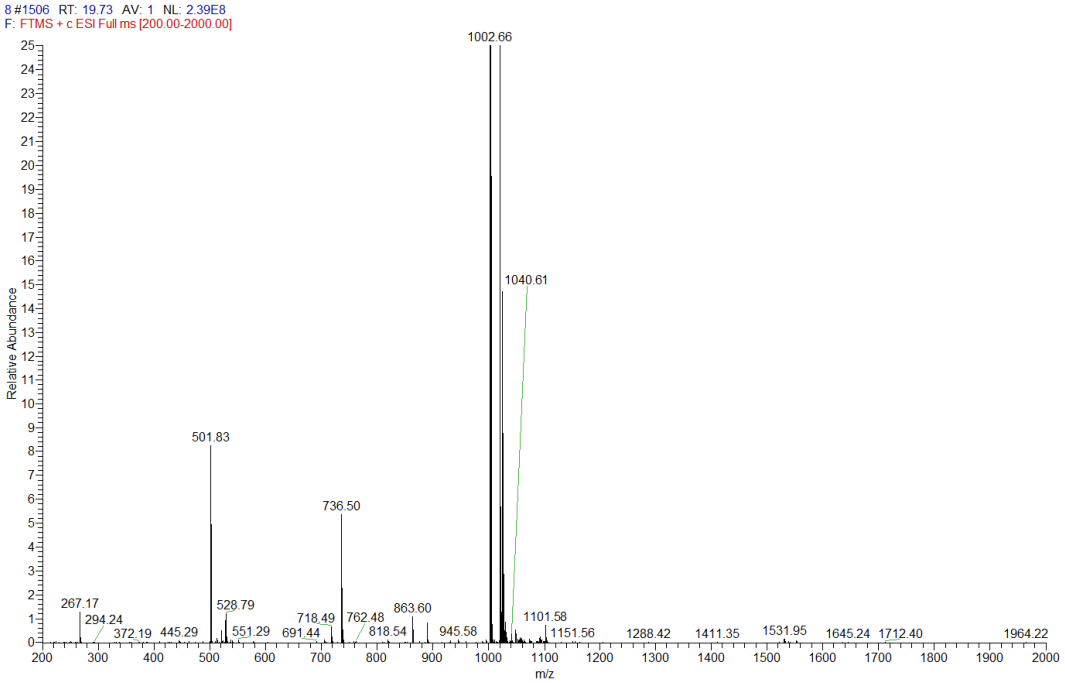


LCMS data for **6**

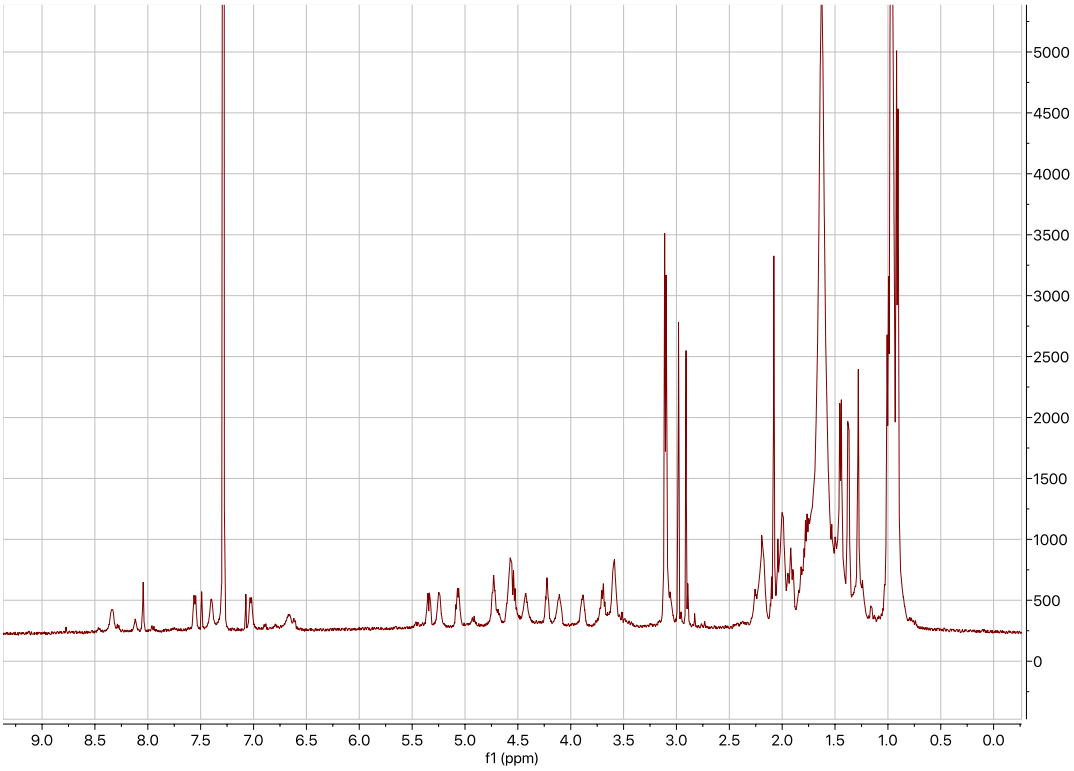


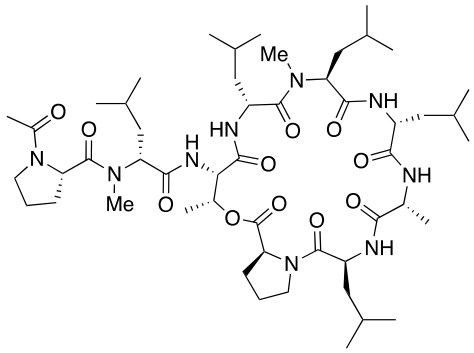



RT: 0.00 - 59.99
NL:
1.87E8
m/z=
1002.63-
1002.67
M:S
cmpd-beta3
Relative Abundance
Time (min)

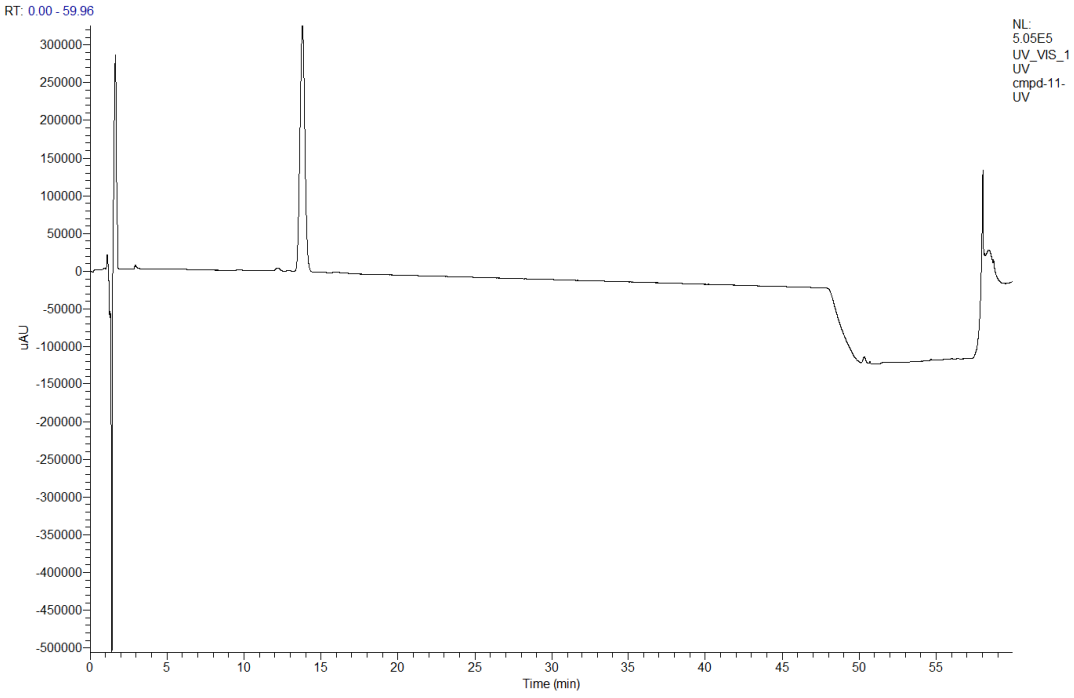

8 #1506 RT: 19.73 AV: 1 NL: 2.39E8
F: FTMS + c ESI Full ms [200.00-2000.00]
1002.66
1040.61
501.83
736.50
267.17
294.24
372.19
445.29
528.79
551.29
691.44
718.49
762.48
818.54
863.60
945.58
1101.58
1151.56
1288.42
1411.35
1531.95
1645.24
1712.40
1964.22
Relative Abundance
m/z

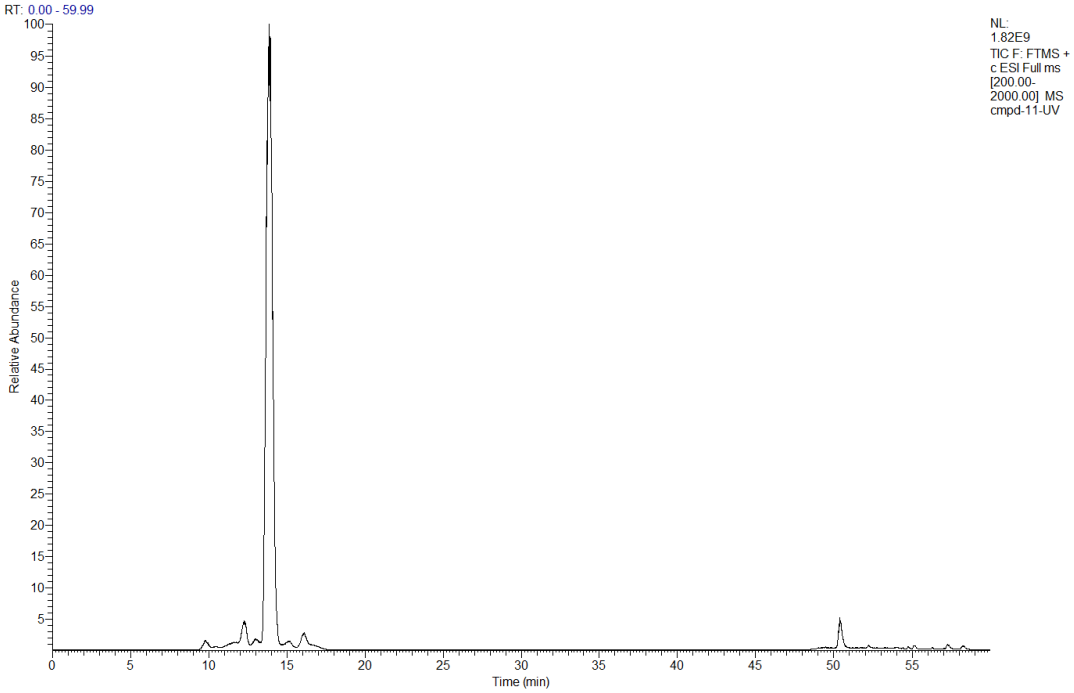
H1 NMR spectrum of **6**

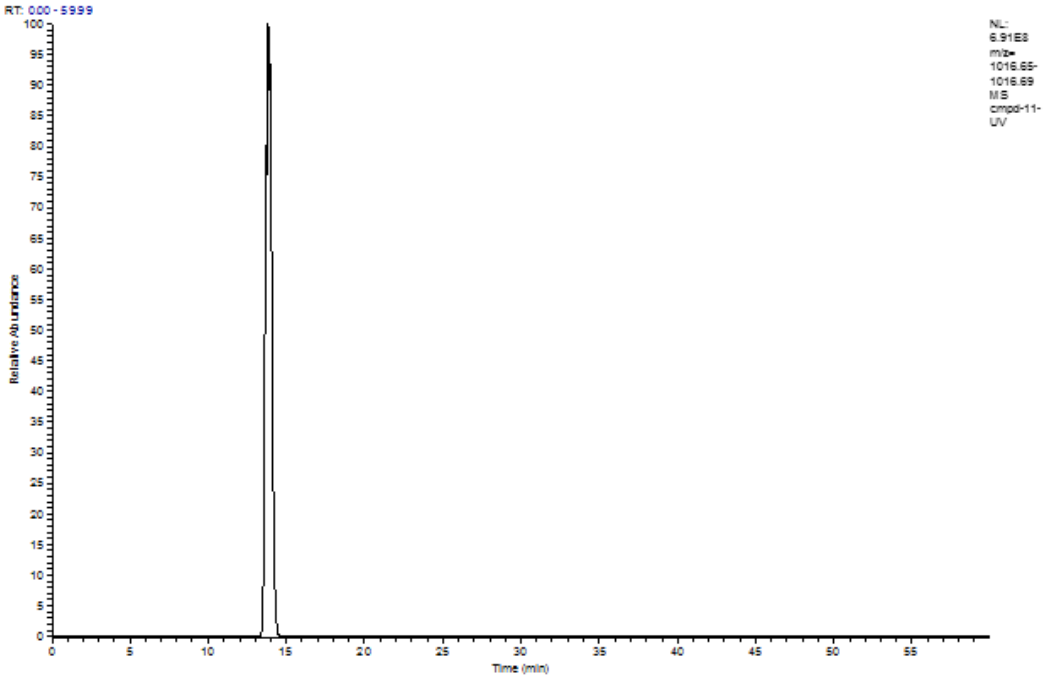


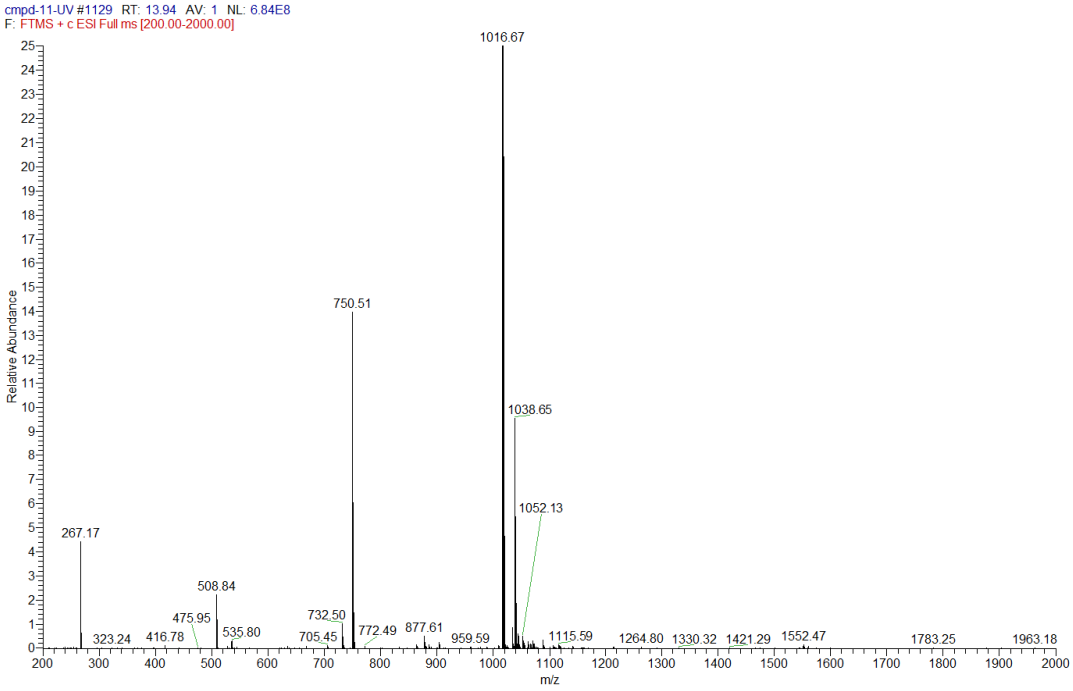


LCMS data for **7**

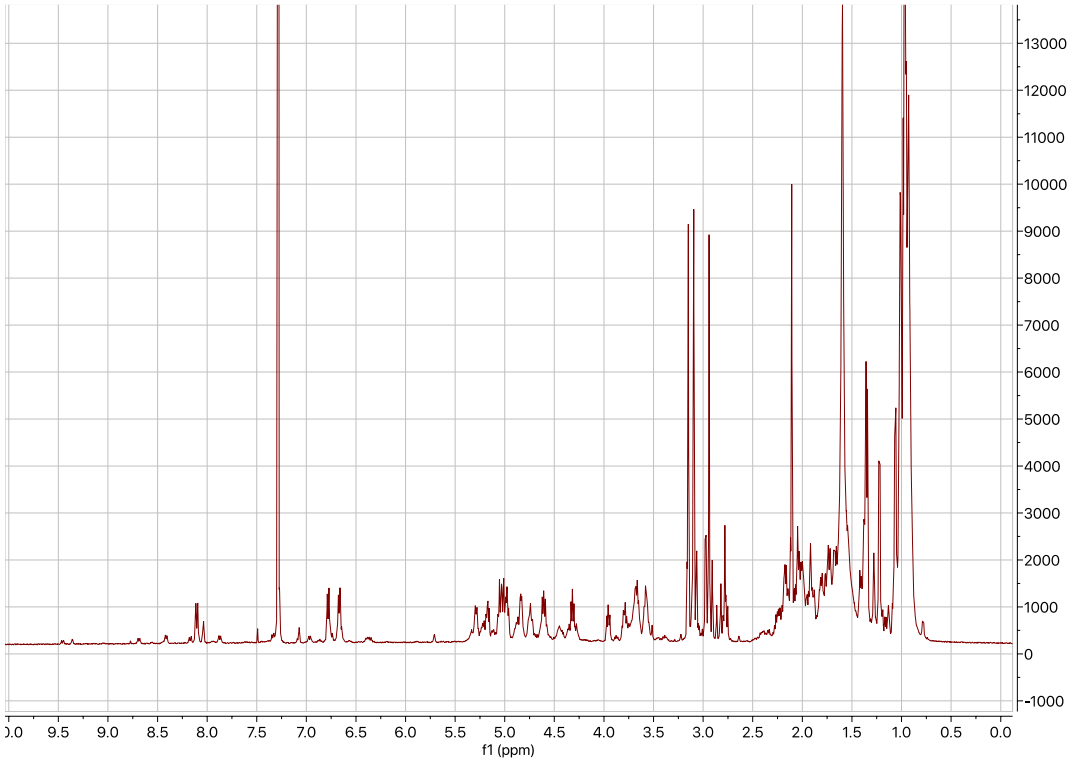


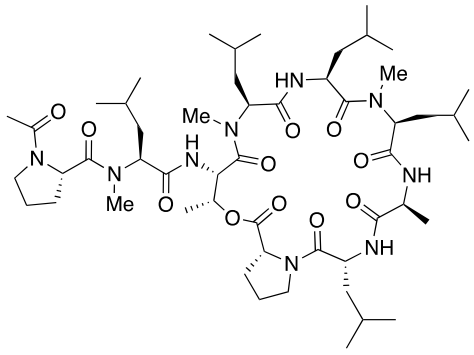



RT: 0.00 - 59.99
NL:
6.91E8
m/z=
1016.65-
1016.69
MS
cmpd-11-
UV
Relative Abundance
Time (min)

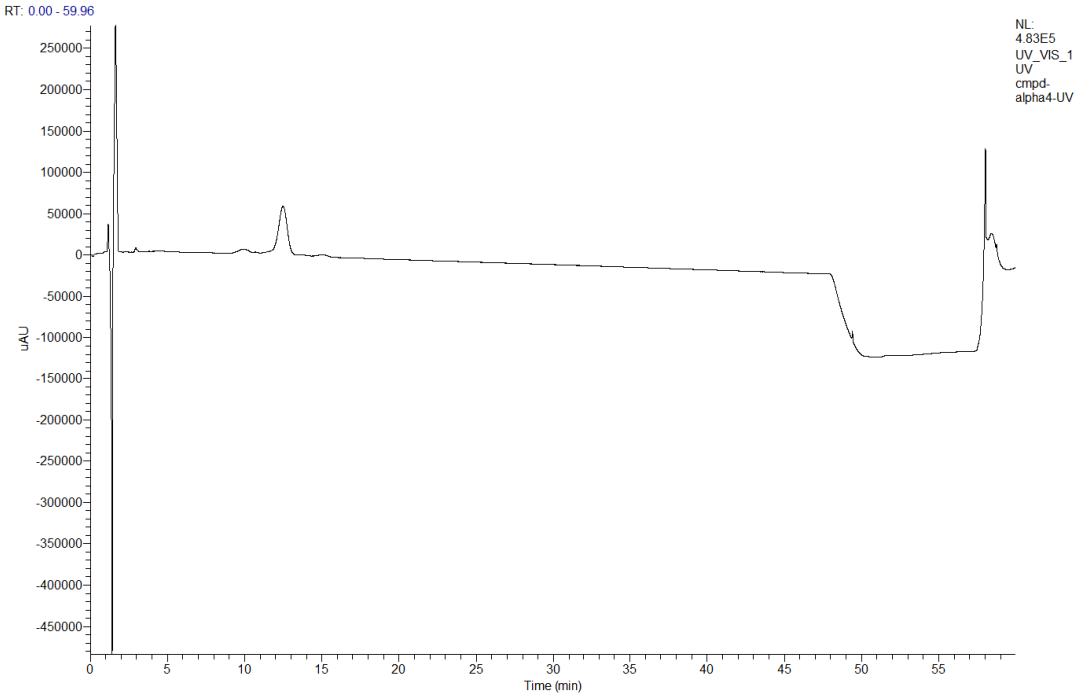

cmpd-11-UV #1129 RT: 13.94 AV: 1 NL: 6.84E8
F: FTMS + c ESI Full ms [200.00-2000.00]
1016.67
750.51
1038.65
1052.13
267.17
508.84
475.95
323.24
416.78
535.80
732.50
705.45
772.49
877.61
959.59
1115.59
1264.80
1330.32
1421.29
1552.47
1783.25
1963.18
Relative Abundance
m/z

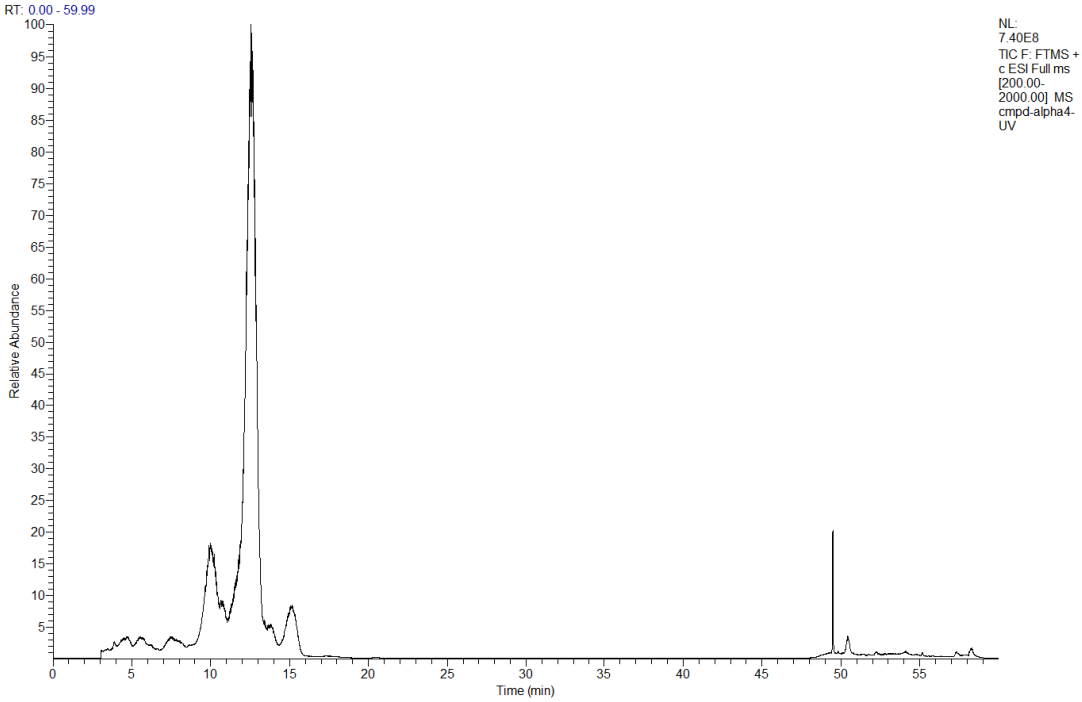
H1 NMR spectrum of **7**

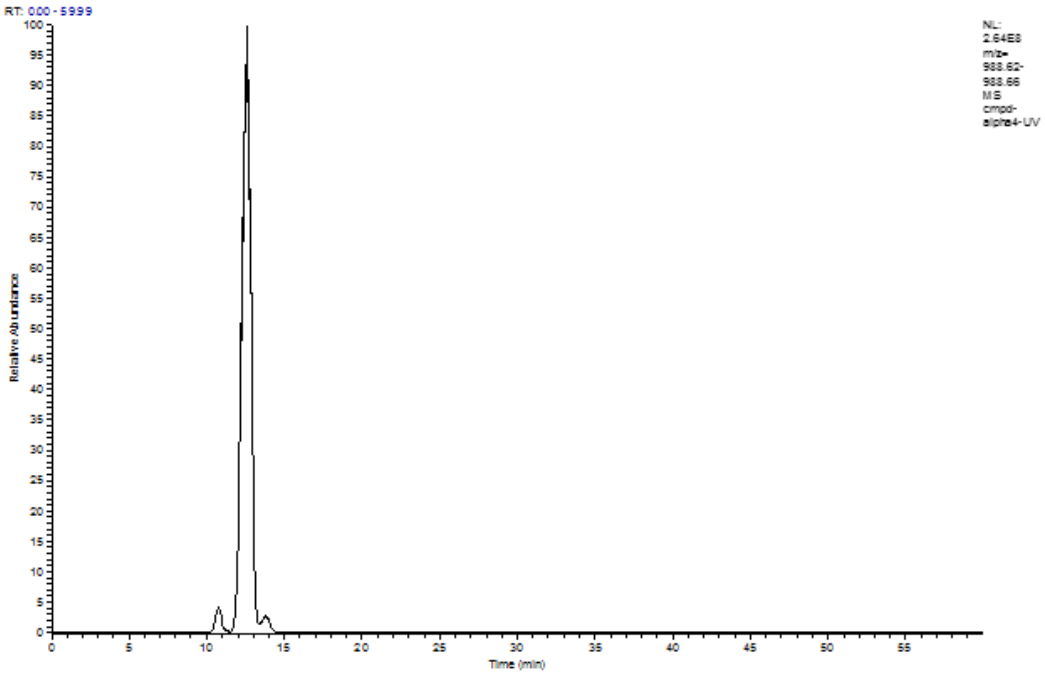


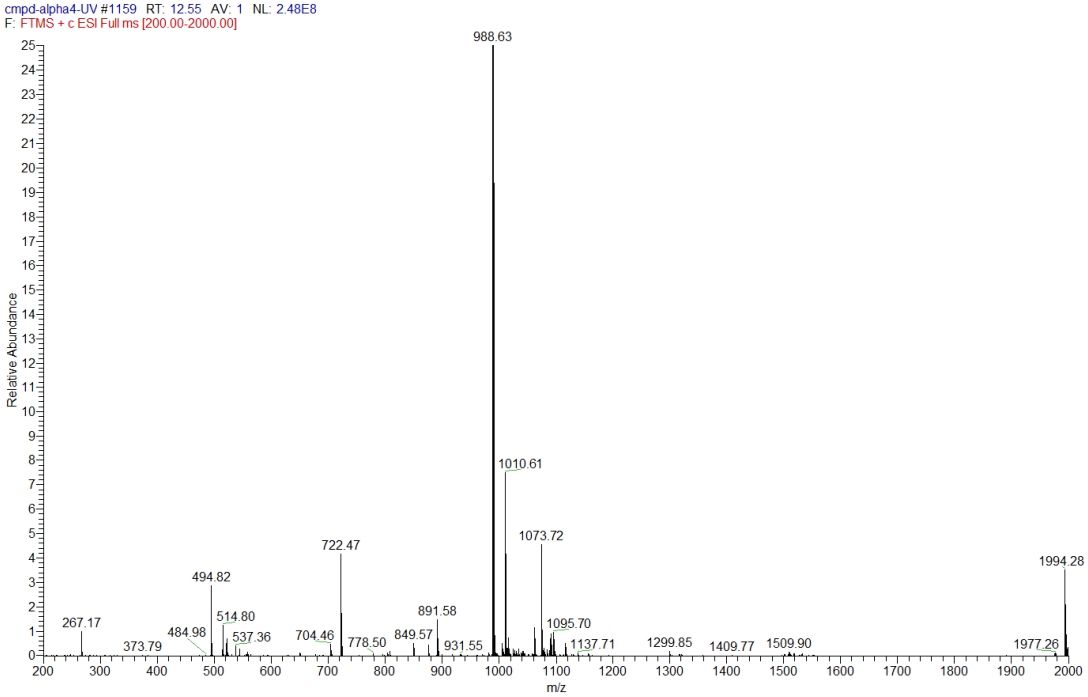


LCMS data for **8**

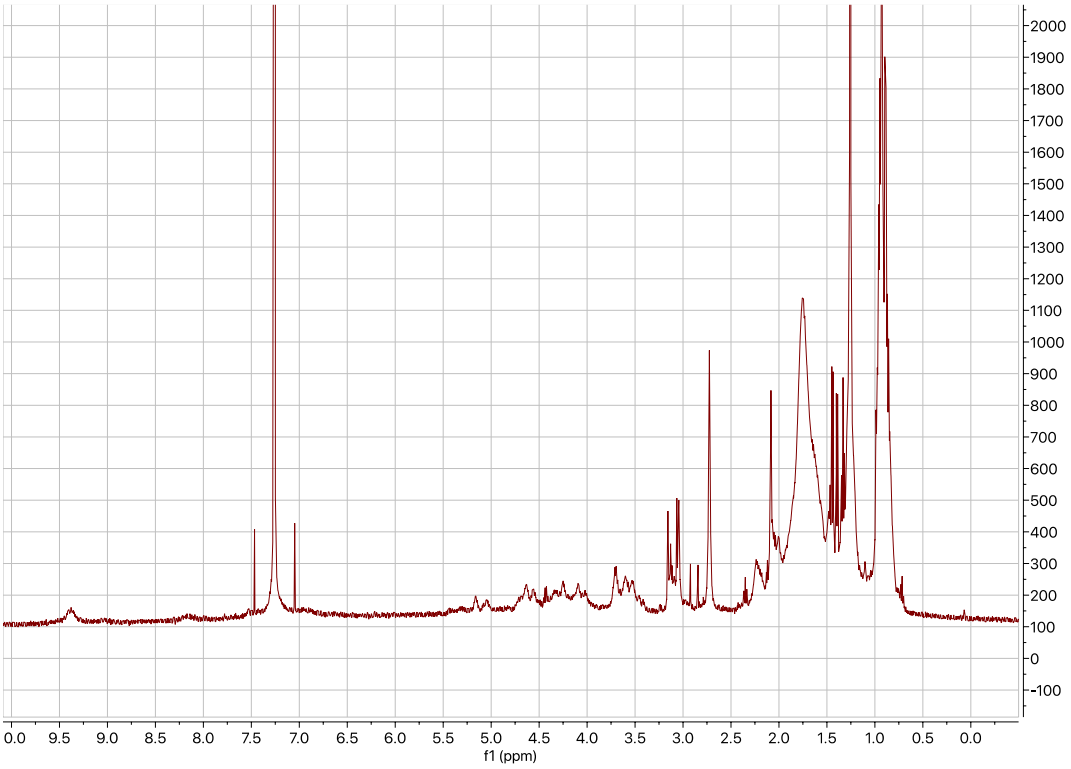


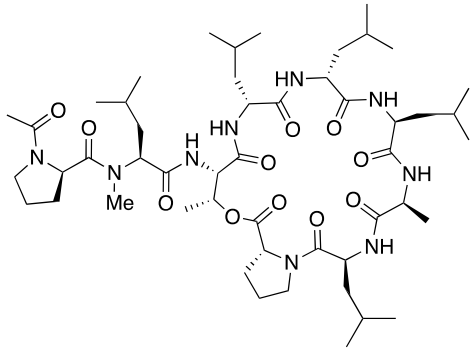



RT: 0.00 - 59.99
NL:
2.64E8
m/z=
988.62-
988.66
MS
cmpd-
alpha4-UV
Relative Abundance
Time (min)

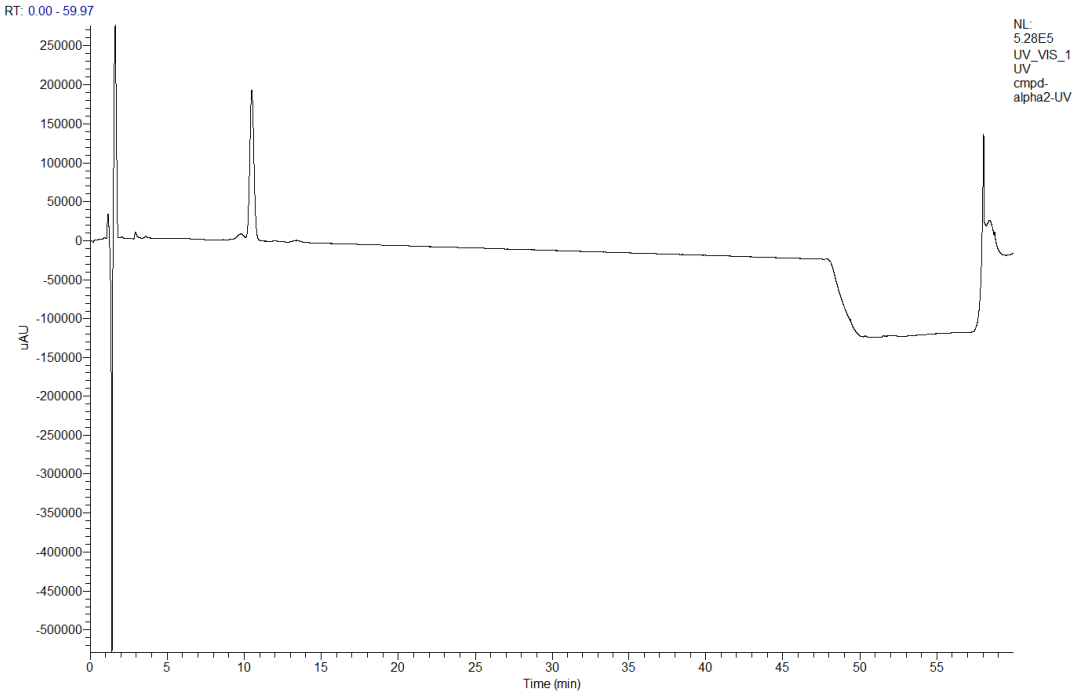

cmpd-alpha4-UV #1159 RT: 12.55 AV: 1 NL: 2.48E8
F: FTMS + c ESI Full ms [200.00-2000.00]
Relative Abundance
m/z
988.63
1010.61
1073.72
722.47
494.82
514.80
891.58
1095.70
1994.28
267.17
484.98
537.36
373.79
704.46
778.50
849.57
931.55
1137.71
1299.85
1409.77
1509.90
1977.26

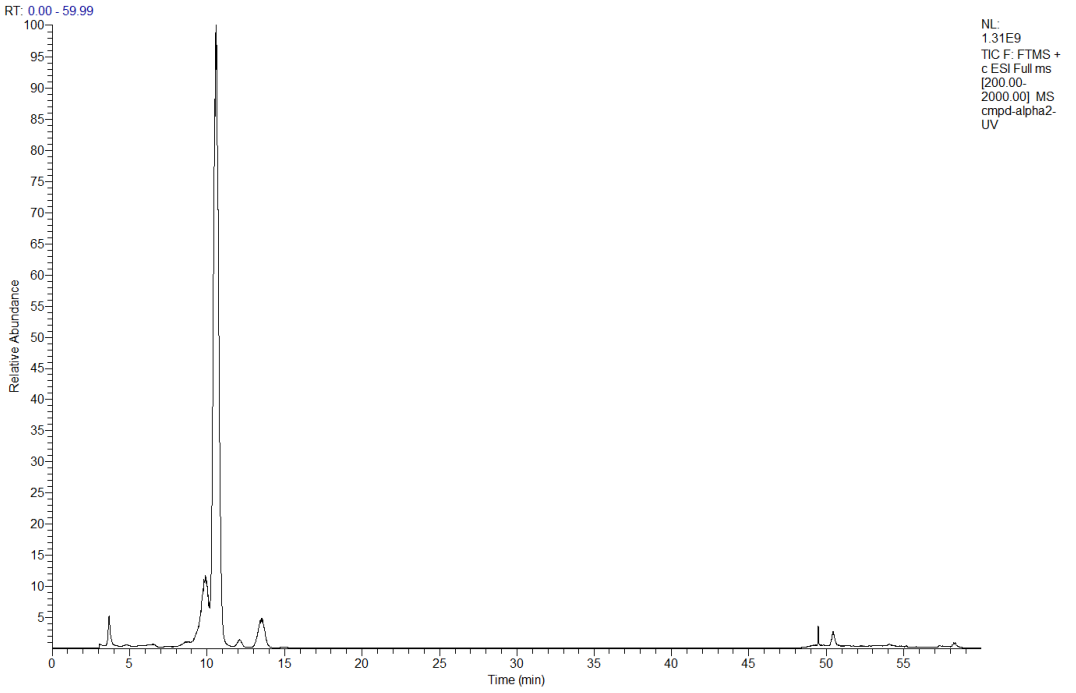
H1 NMR spectrum of **8**

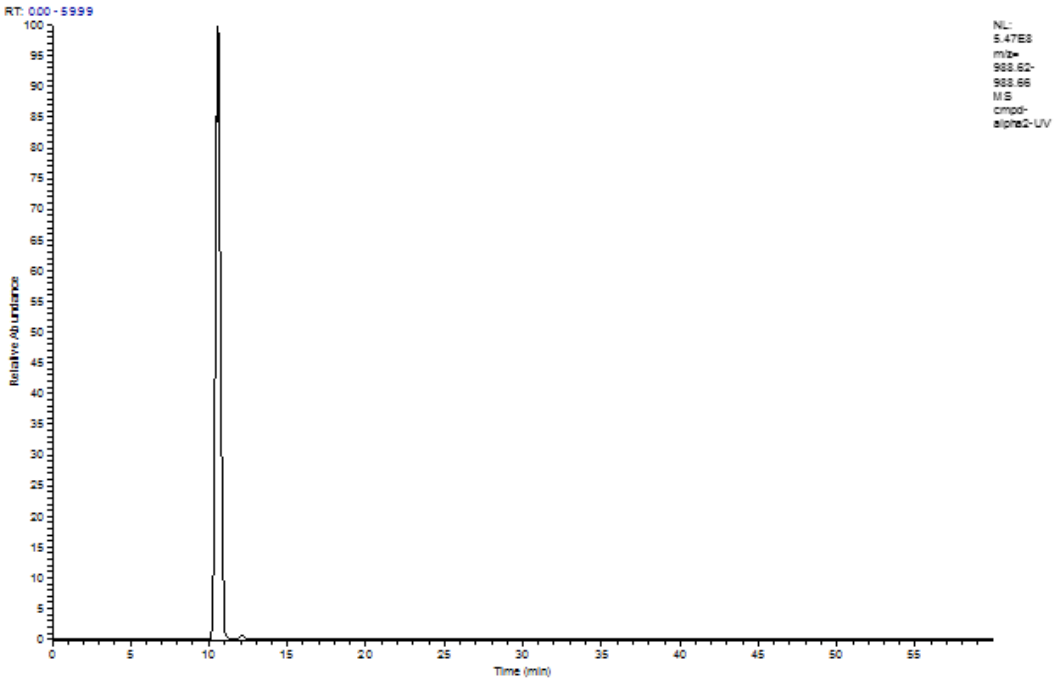


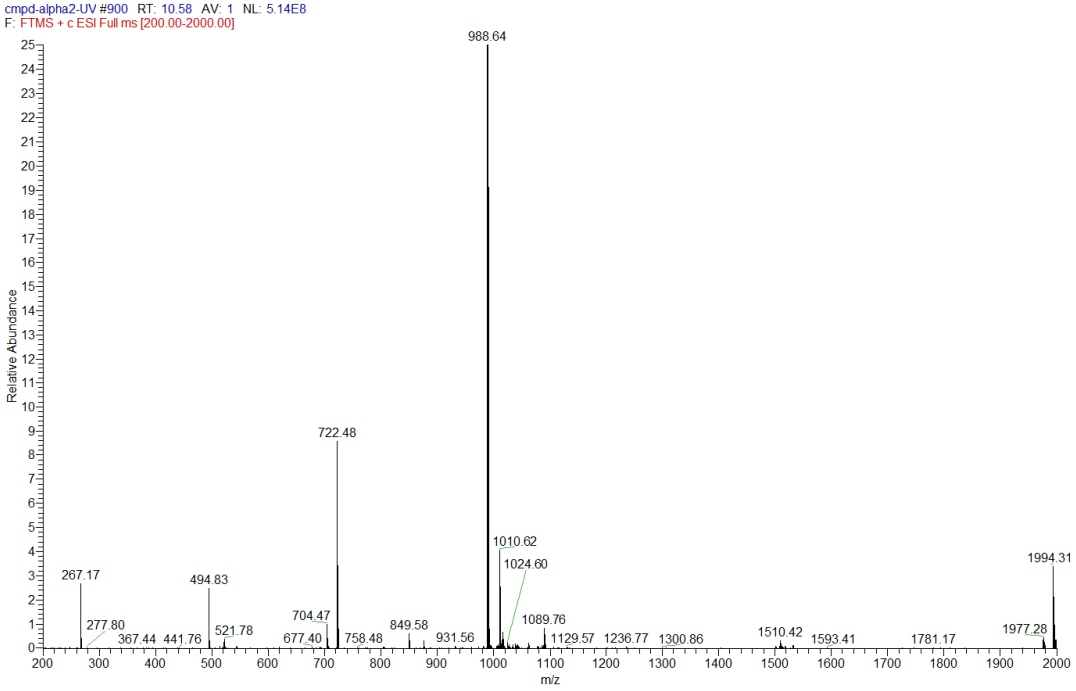


LCMS data for **9**

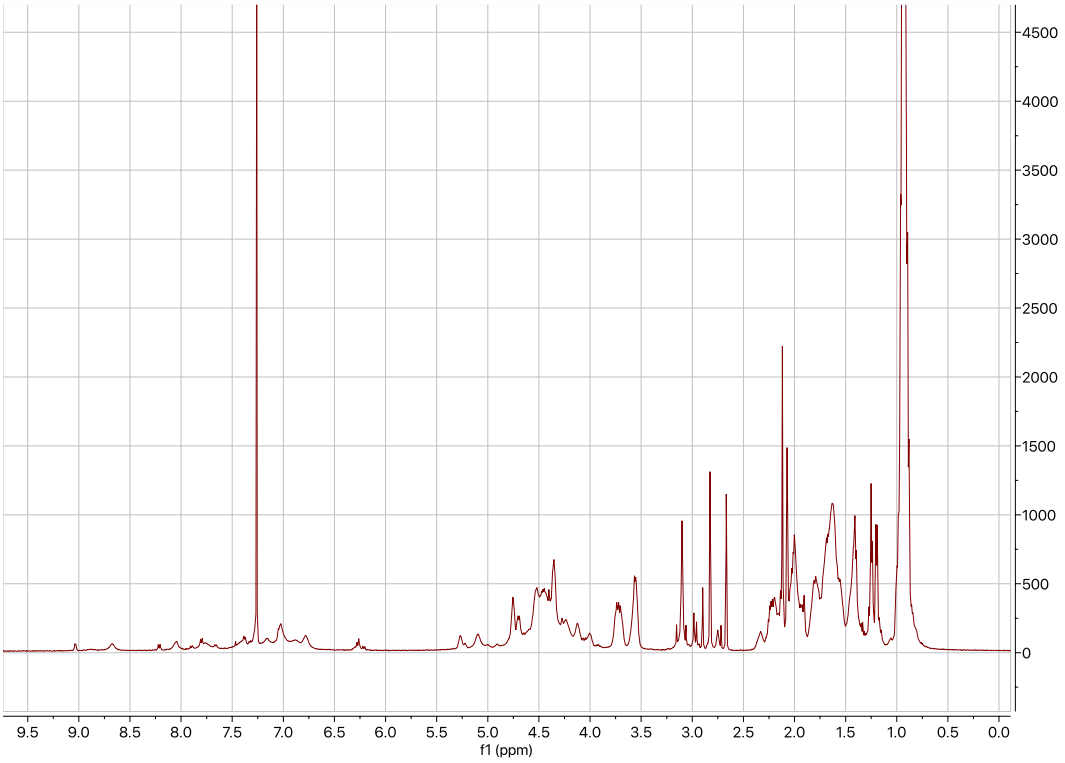


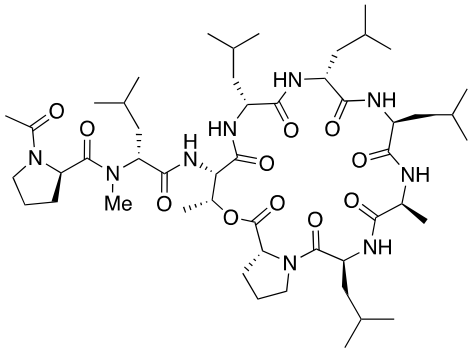



RT: 0.00 - 59.99
NL:
5.47E8
m/z=
988.62-
988.66
MS
cmpd-
alpha2-UV
Relative Abundance
Time (min)

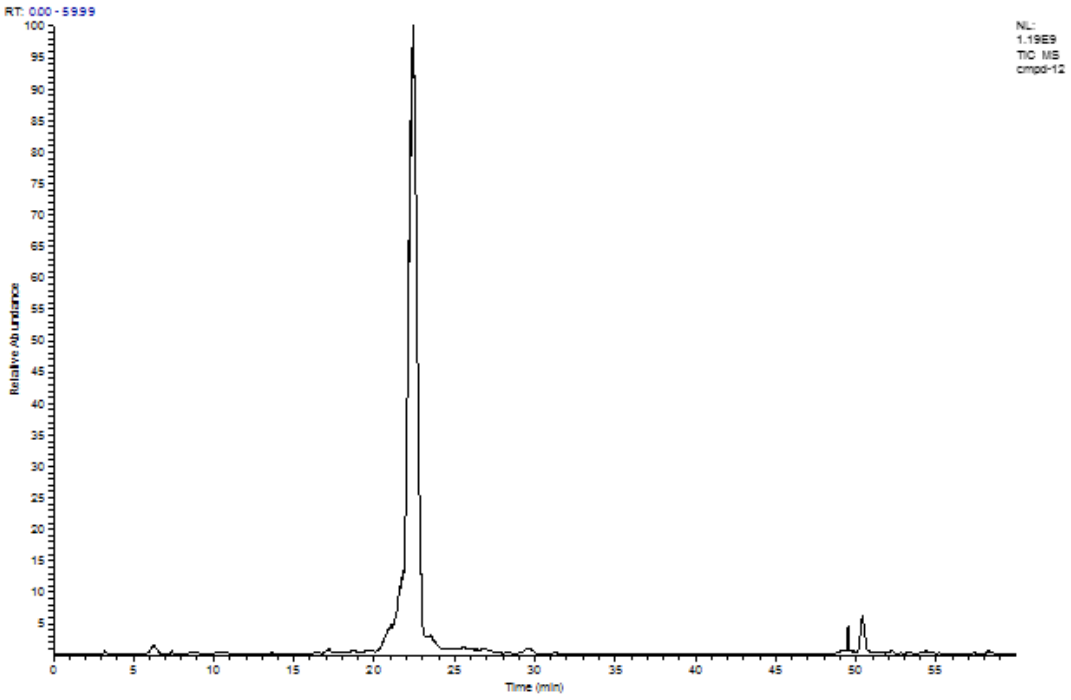

cmpd-alpha2-UV #900 RT: 10.58 AV: 1 NL: 5.14E8
F: FTMS + c ESI Full ms [200.00-2000.00]
988.64
722.48
1010.62
1024.60
1994.31
267.17
494.83
704.47
277.80
367.44
441.76
521.78
677.40
758.48
849.58
931.56
1089.76
1129.57
1236.77
1300.86
1510.42
1593.41
1781.17
1977.28
Relative Abundance
m/z

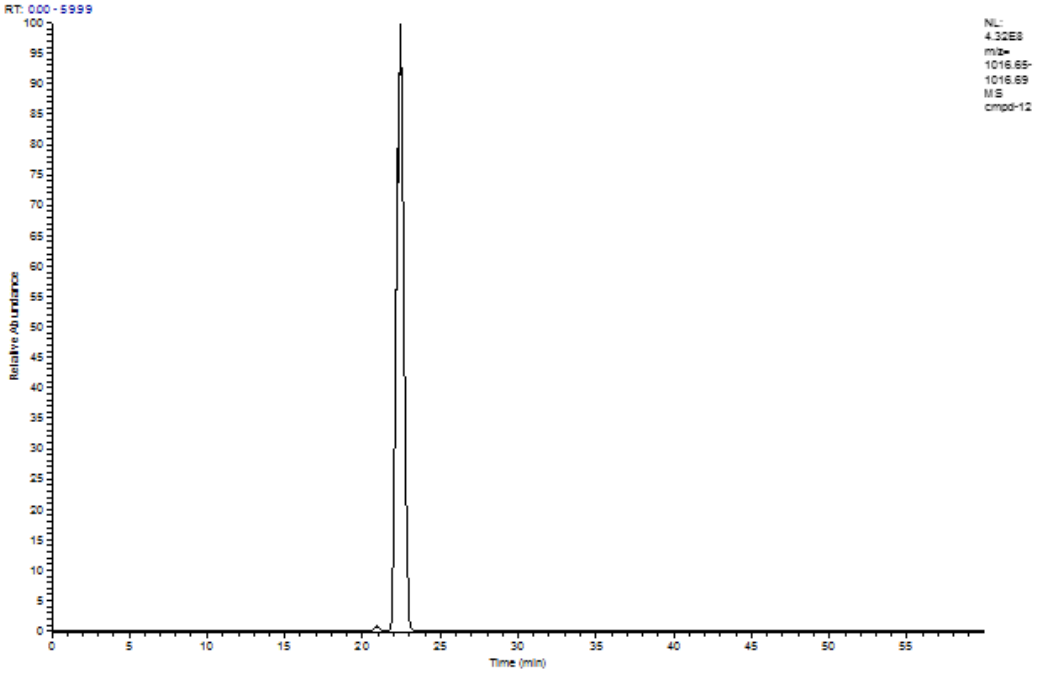
H1 NMR spectrum of **9**

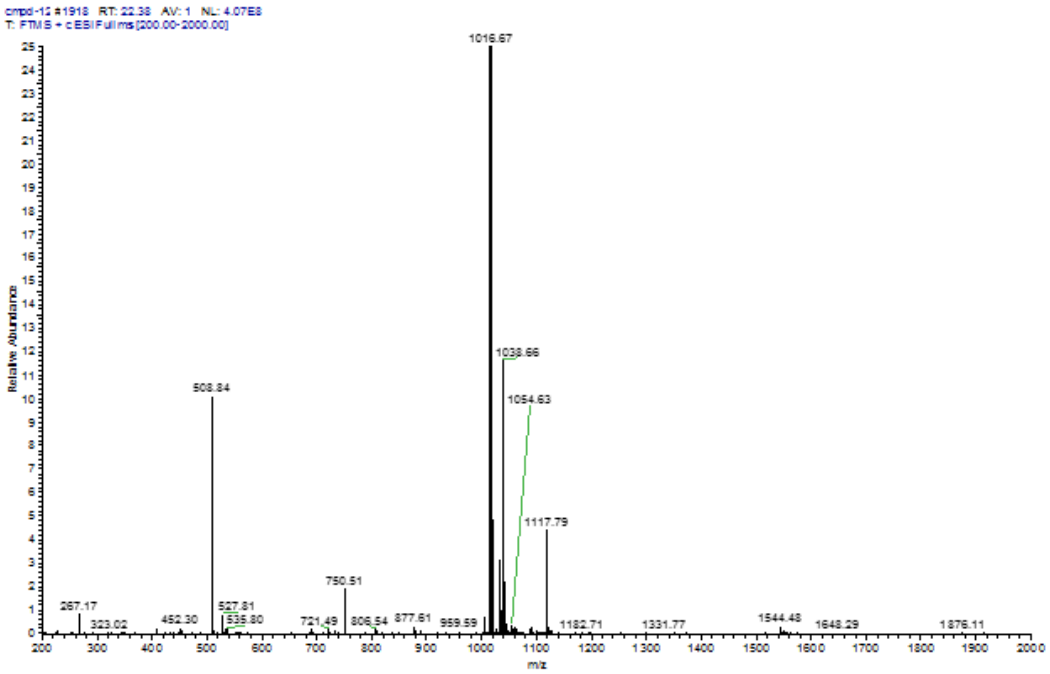




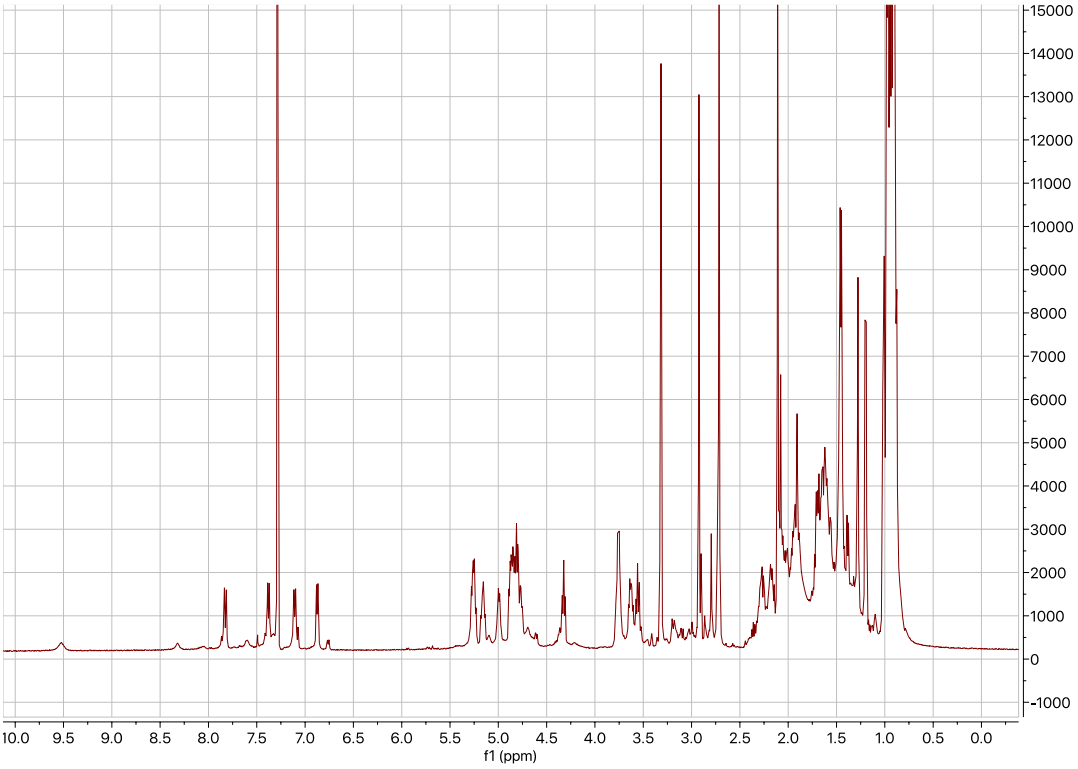
LCMS data for **10**

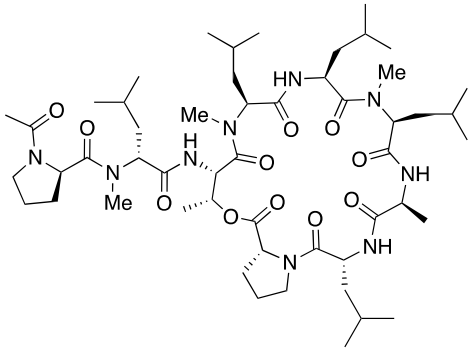




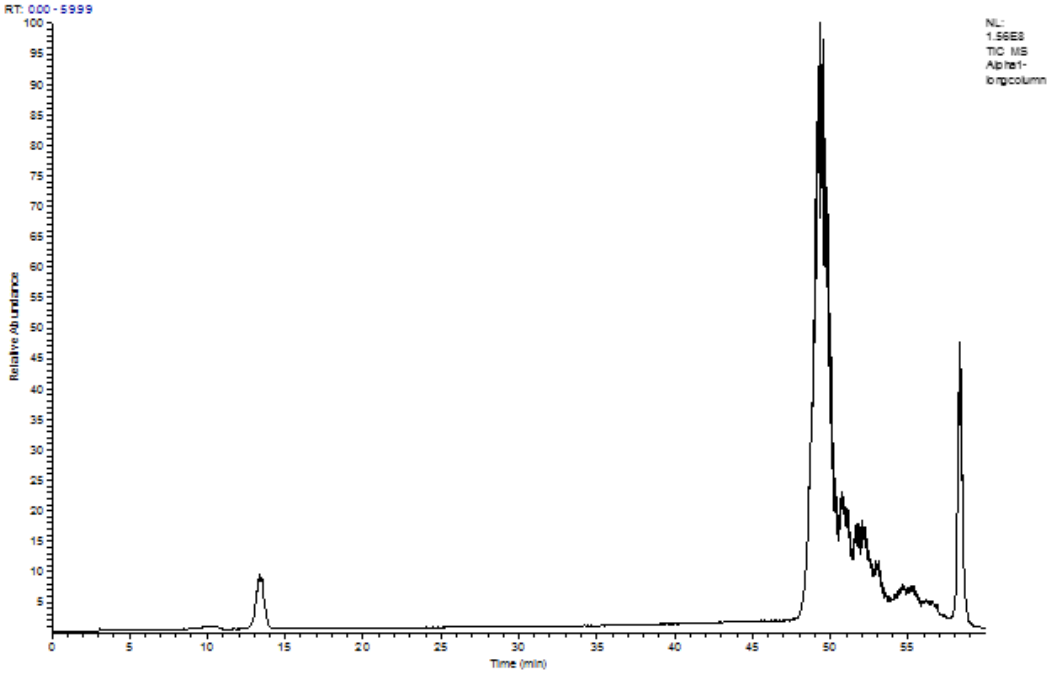

cmpd-12 #1918 RT: 22.38 AV: 1 NL: 4.07E8
T: FTMS + c ESI Full ms [200.00-2000.00]
Relative Abundance
m/z
1016.67
1038.66
1054.63
1117.79
508.84
750.51
267.17
323.02
452.30
527.81
535.80
721.49
806.54
877.61
959.59
1182.71
1331.77
1544.48
1648.29
1876.11

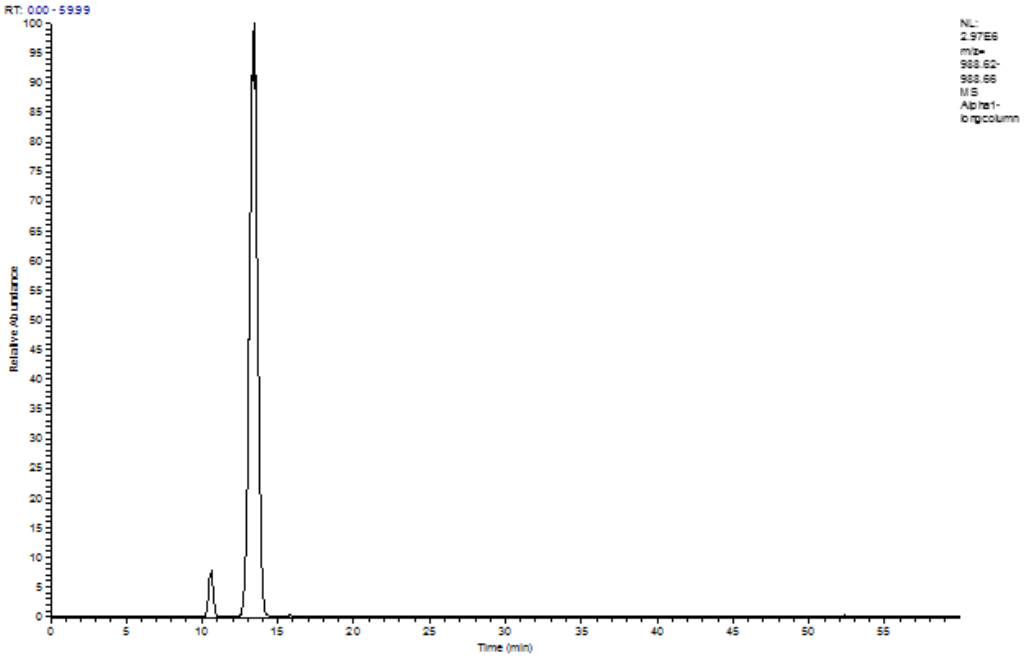
H1 NMR spectrum of **10**

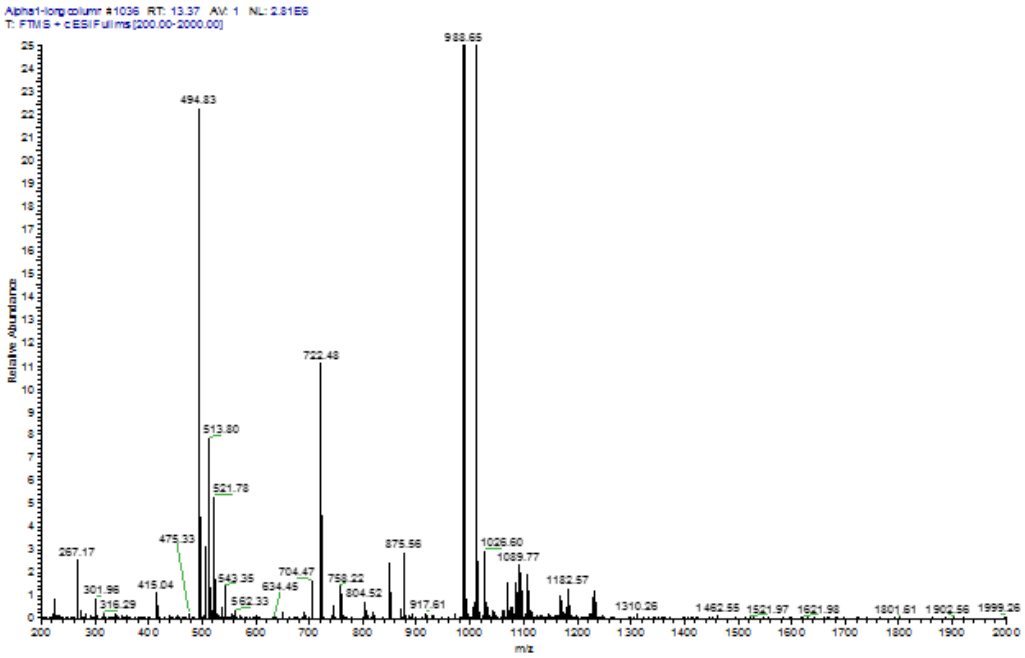




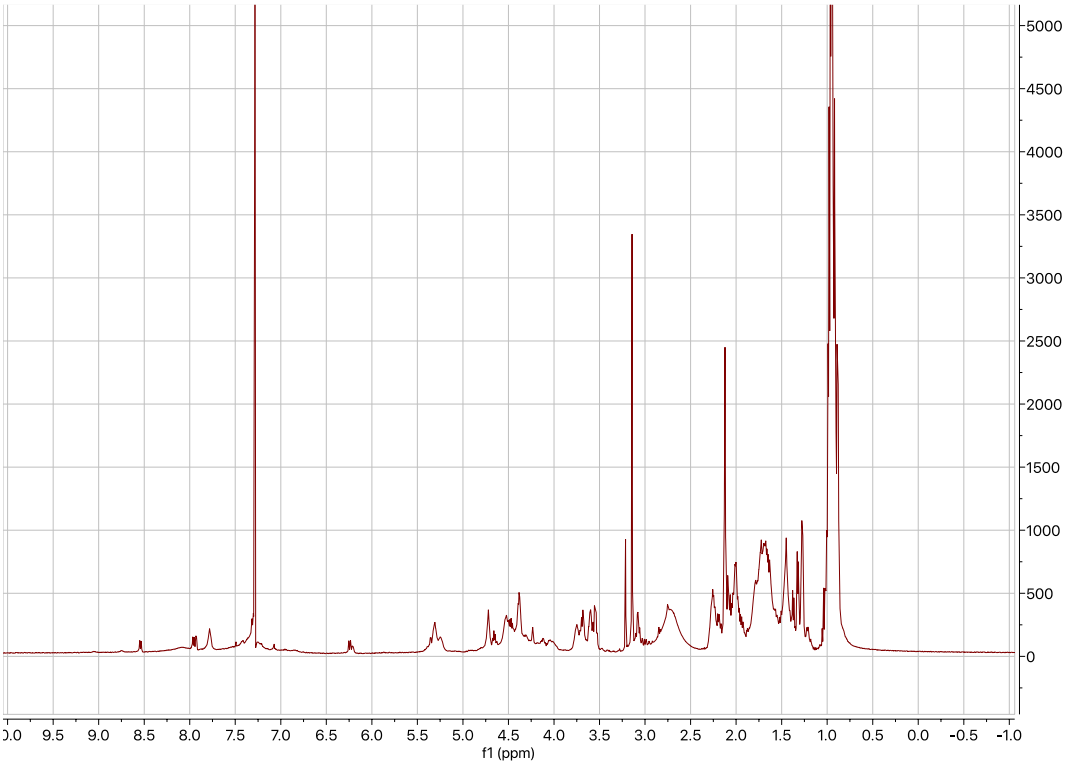
LCMS data for **11**

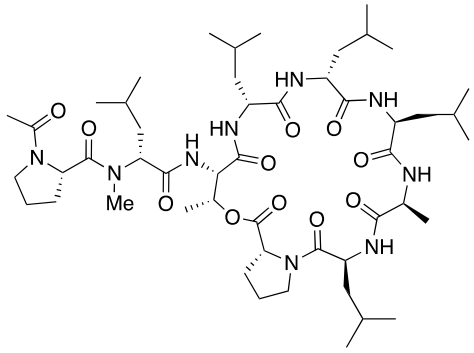


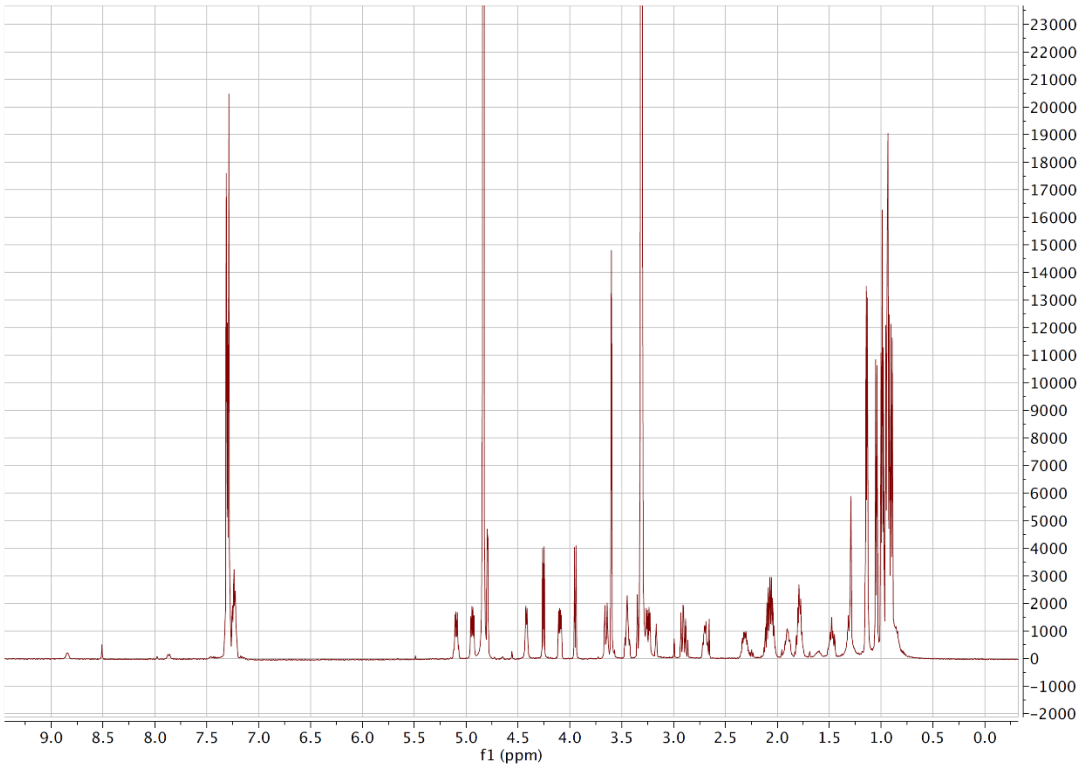



Alpha1-long column #1036 RT: 13.37 AV: 1 NL: 2.81E6
T: FTMS + c ESI Full ms [200.00-2000.00]
Relative Abundance
m/z
988.65
494.83
722.48
513.80
521.78
475.33
267.17
875.56
1026.60
1089.77
704.47
543.35
634.45
758.22
804.52
415.04
301.96
316.29
562.33
917.61
1182.57
1310.26
1462.55
1521.97
1621.98
1801.61
1902.56
1999.26

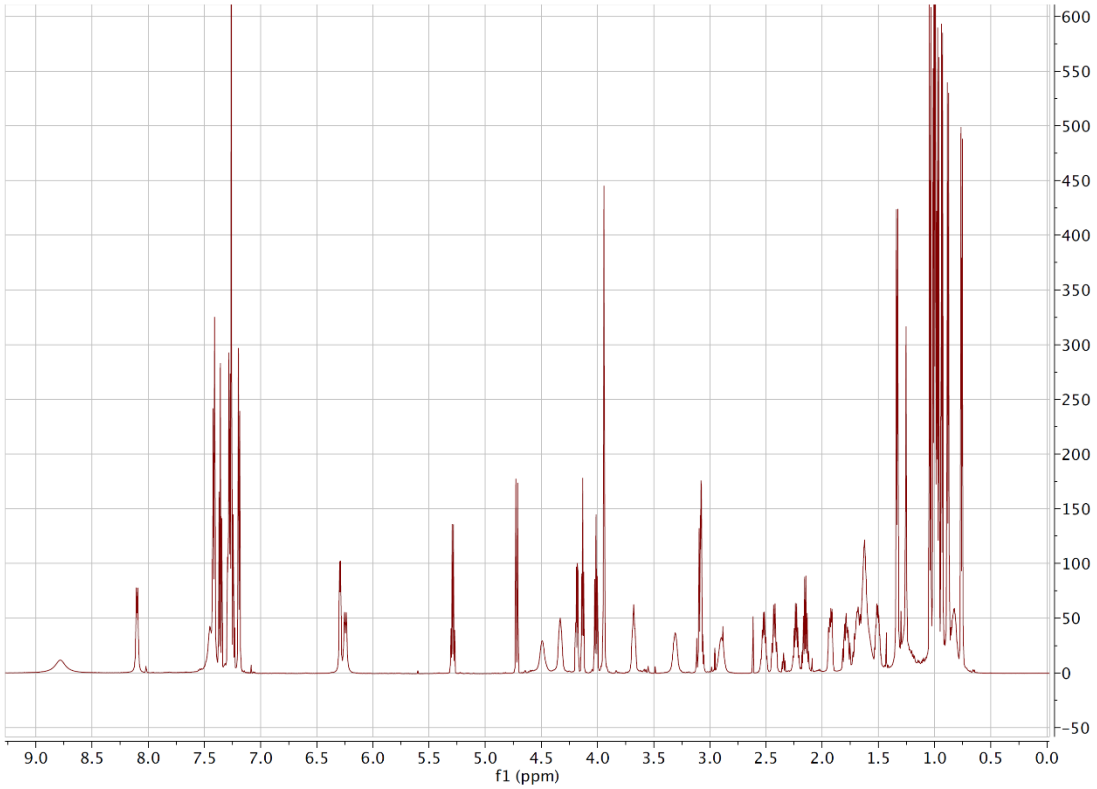
H1 NMR spectrum of **11**



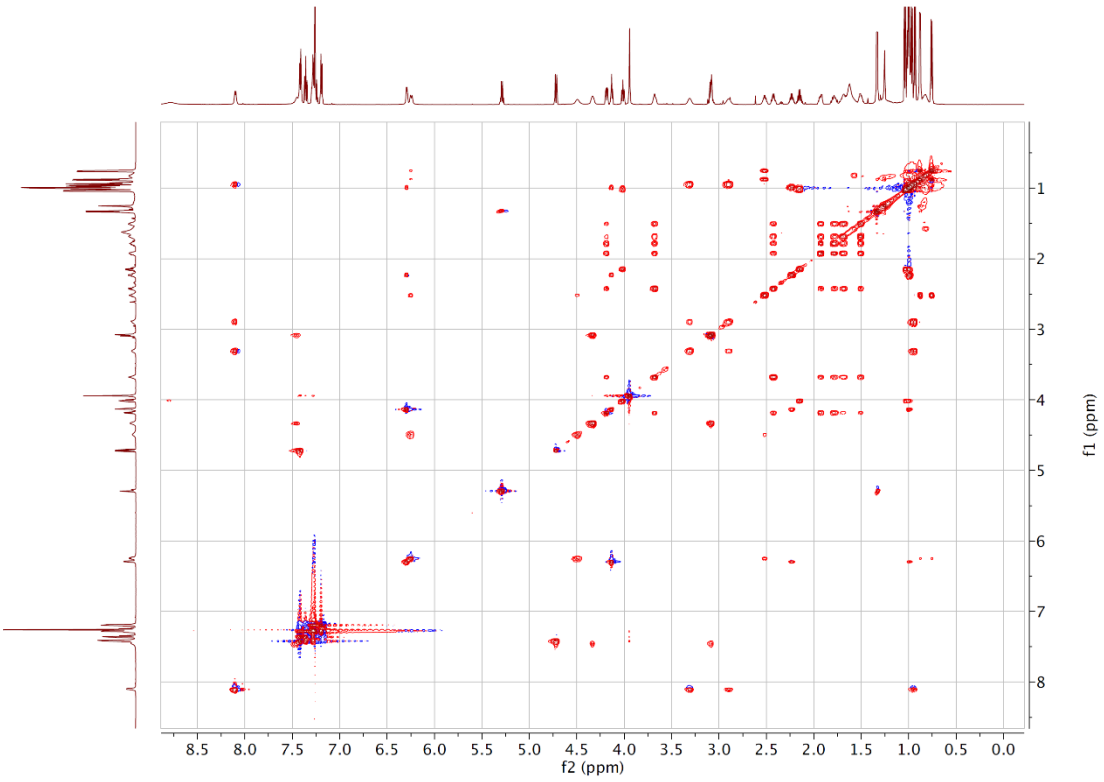




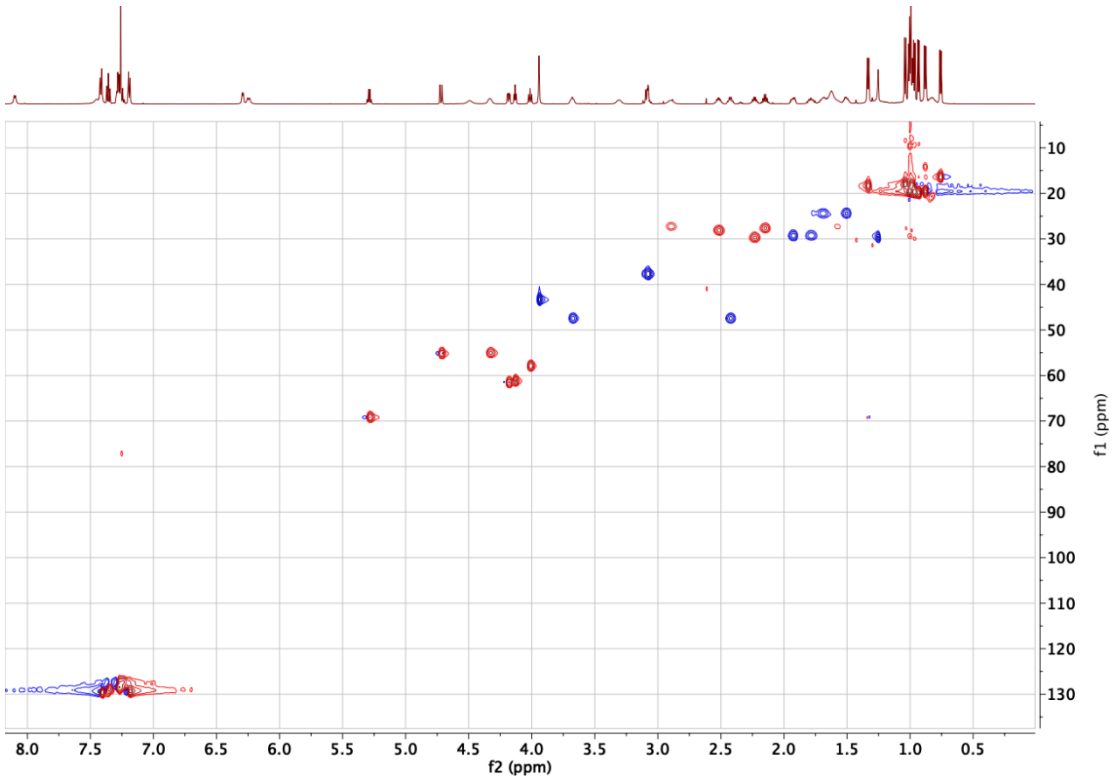
Proton spectrum of XvA in methanol-d_4



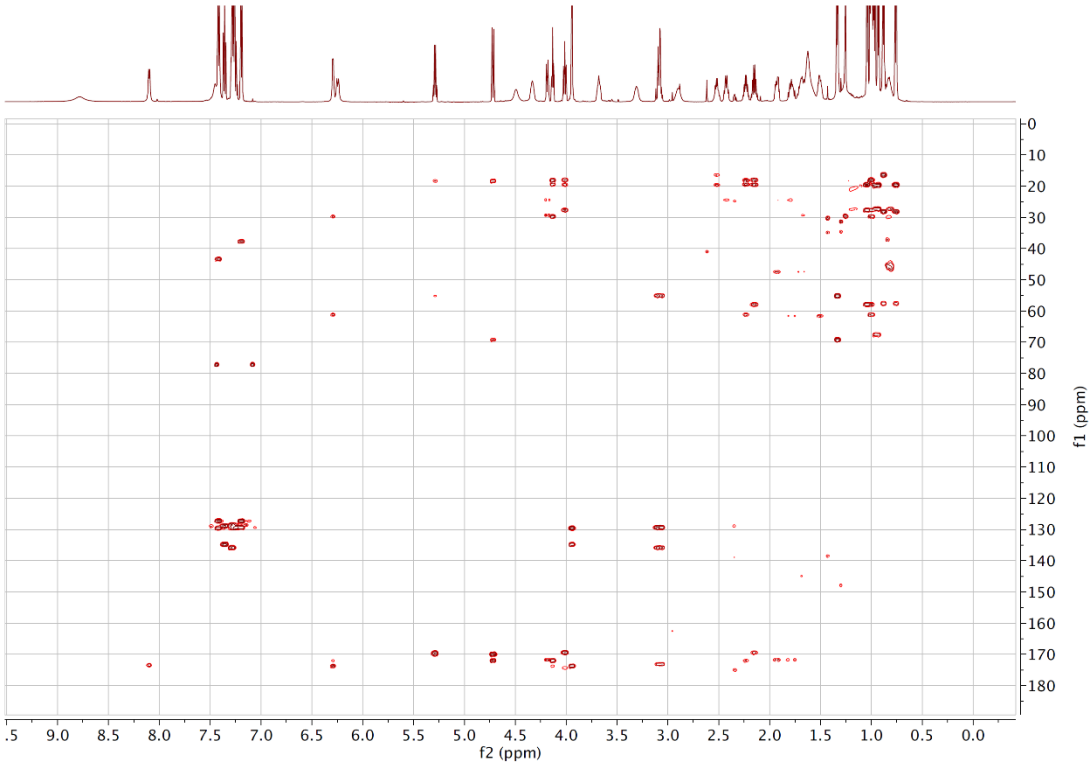
Proton spectrum of XvA in $CDCl_3$



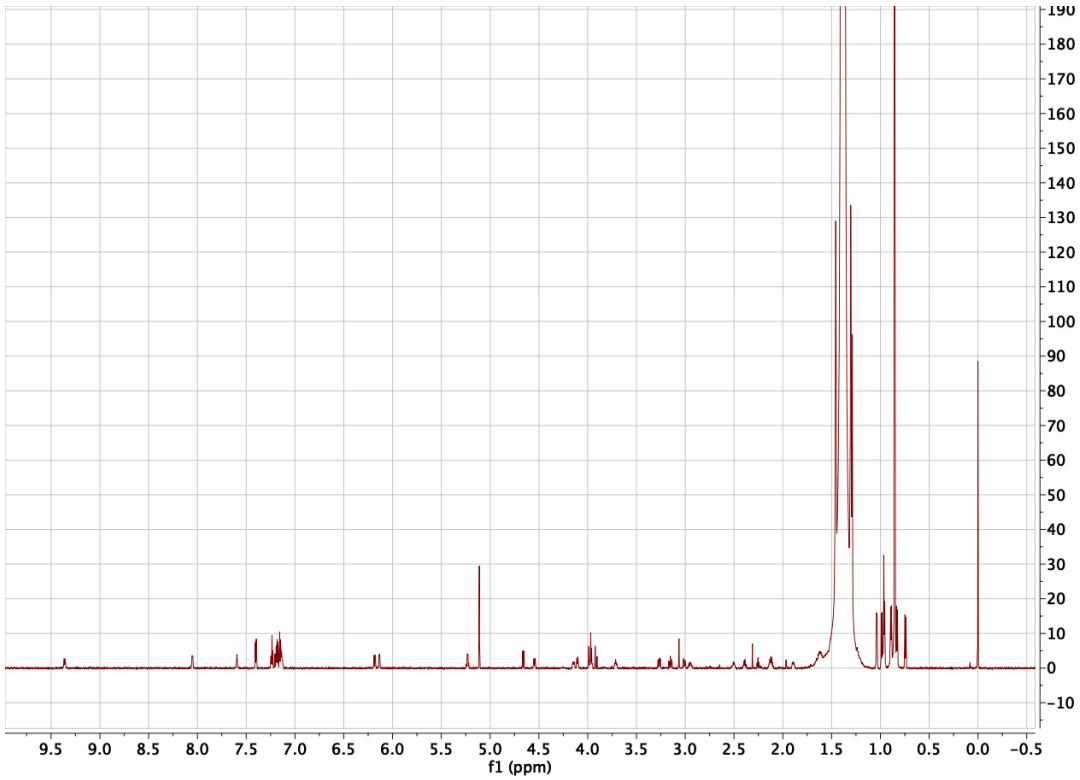
TOCSY spectrum of XvA in $CDCl_3$



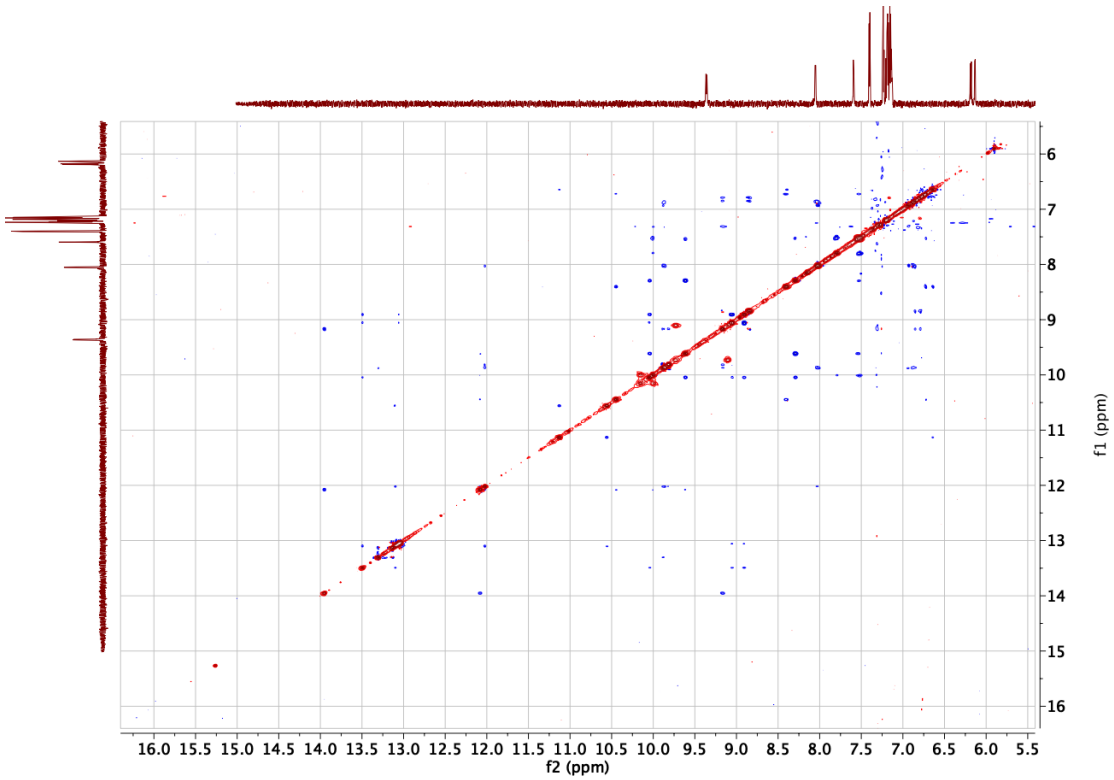
HSQC spectrum of XvA in $CDCl_3$



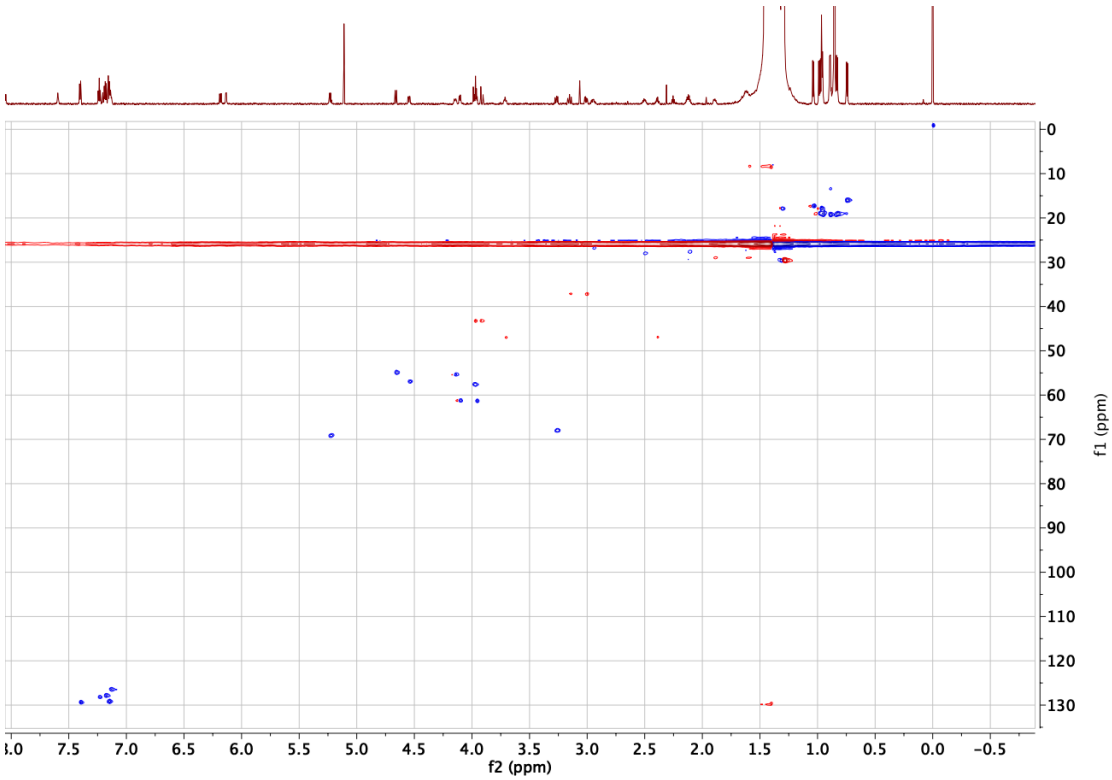
HMBC spectrum of XvA in $CDCl_3$



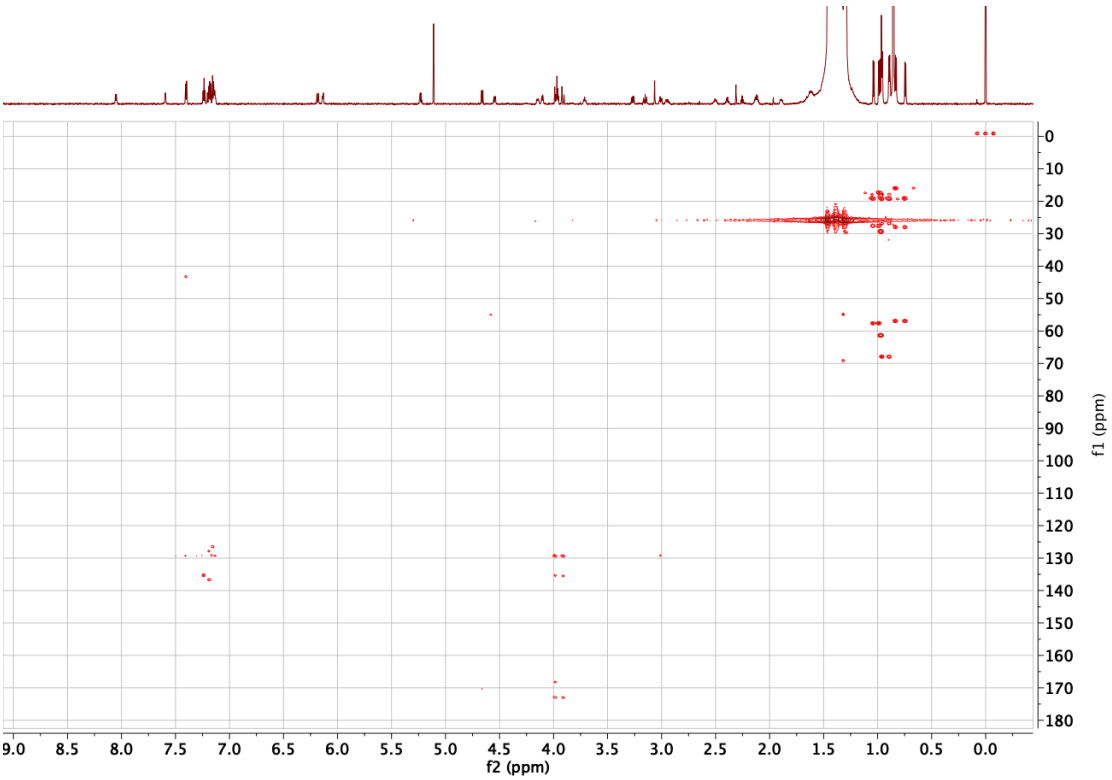
Proton spectrum of XvA in cyclohexane-d_{12}



ROESY spectrum of XvA in cyclohexane-d_{12}

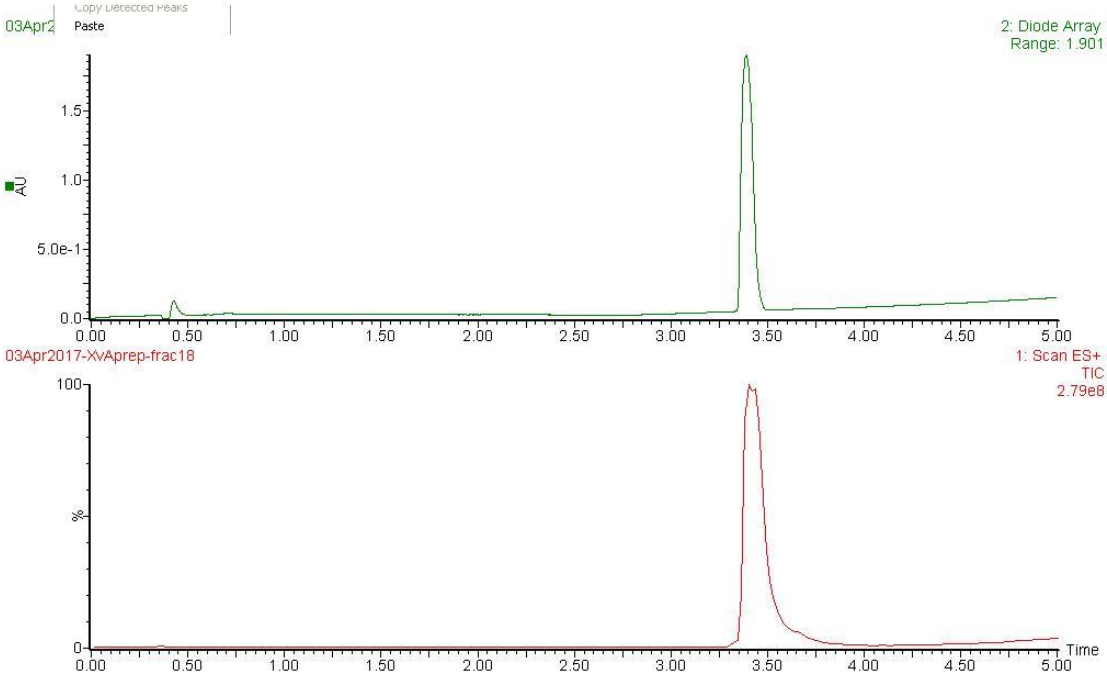


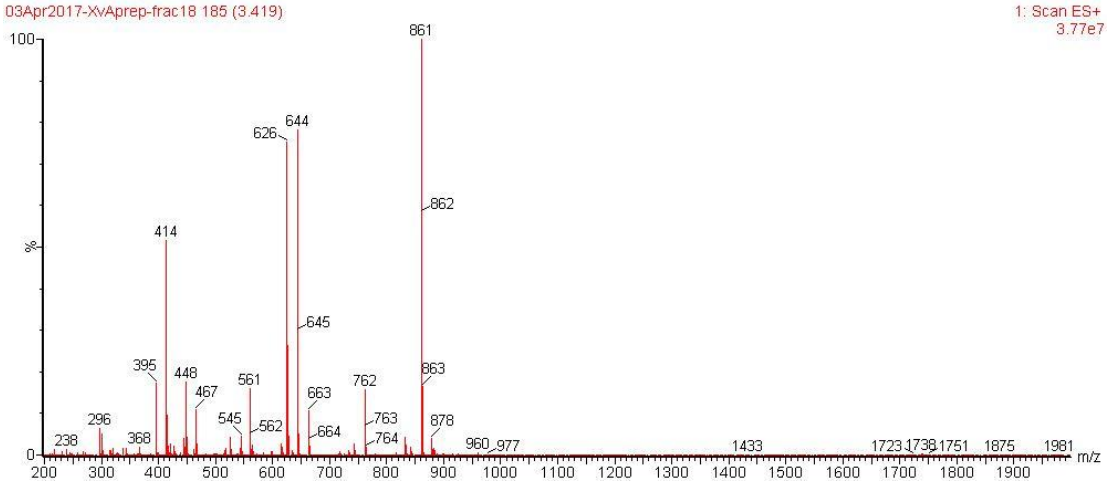
HSQC spectrum of XvA in cyclohexane-d_{12}



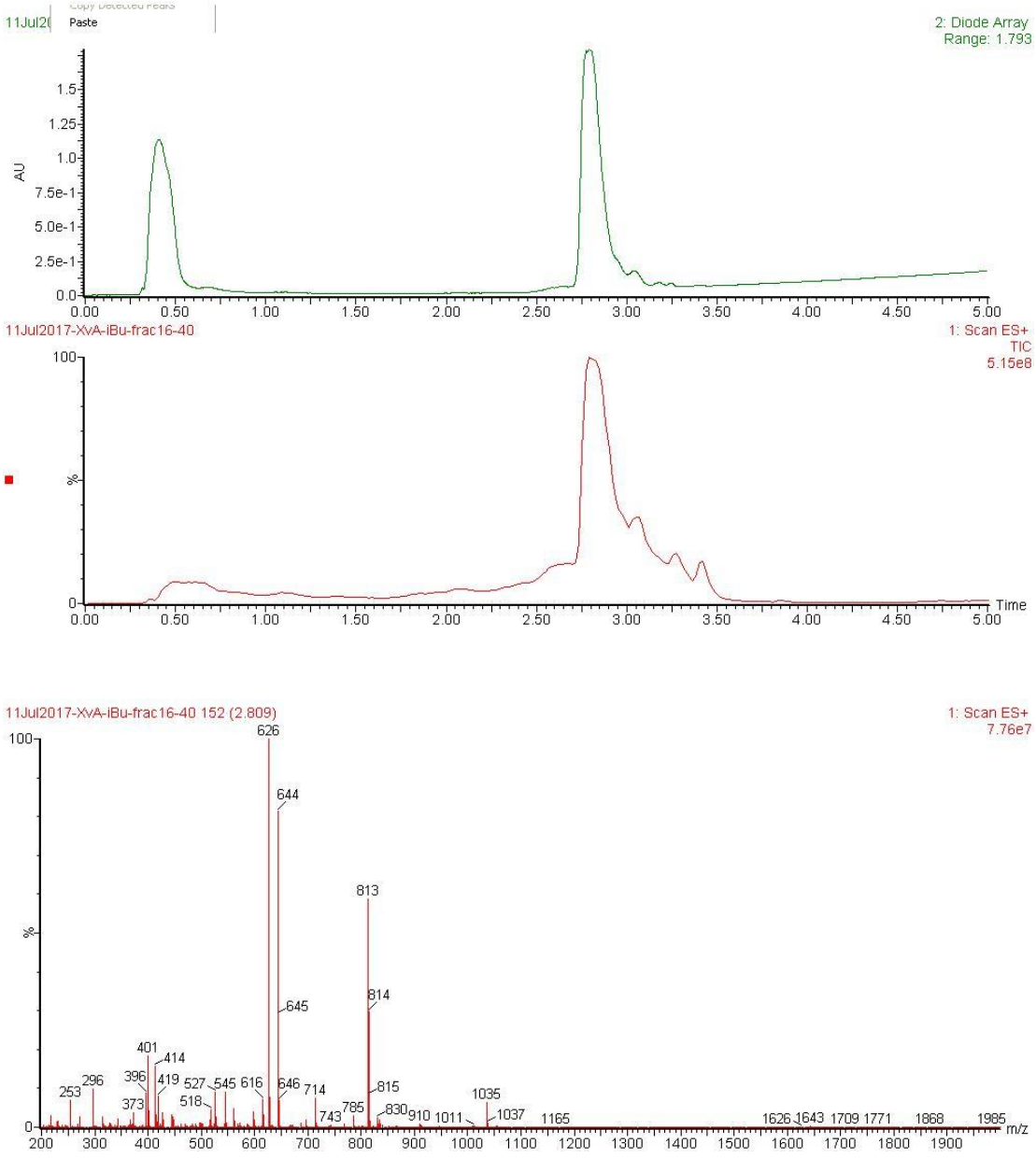
HMBC spectrum of XvA in cyclohexane-d_{12}

LCMS data for XvA

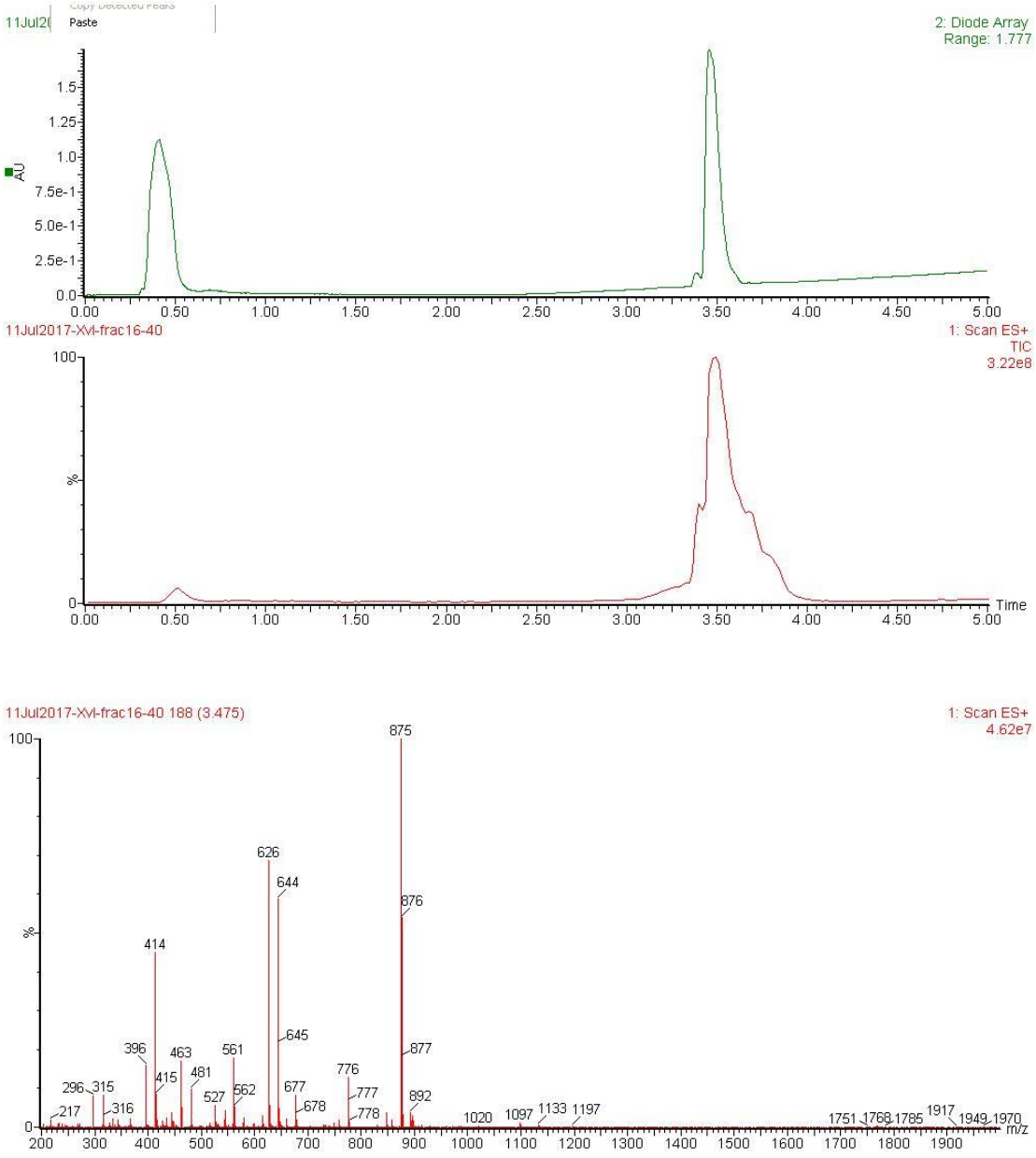




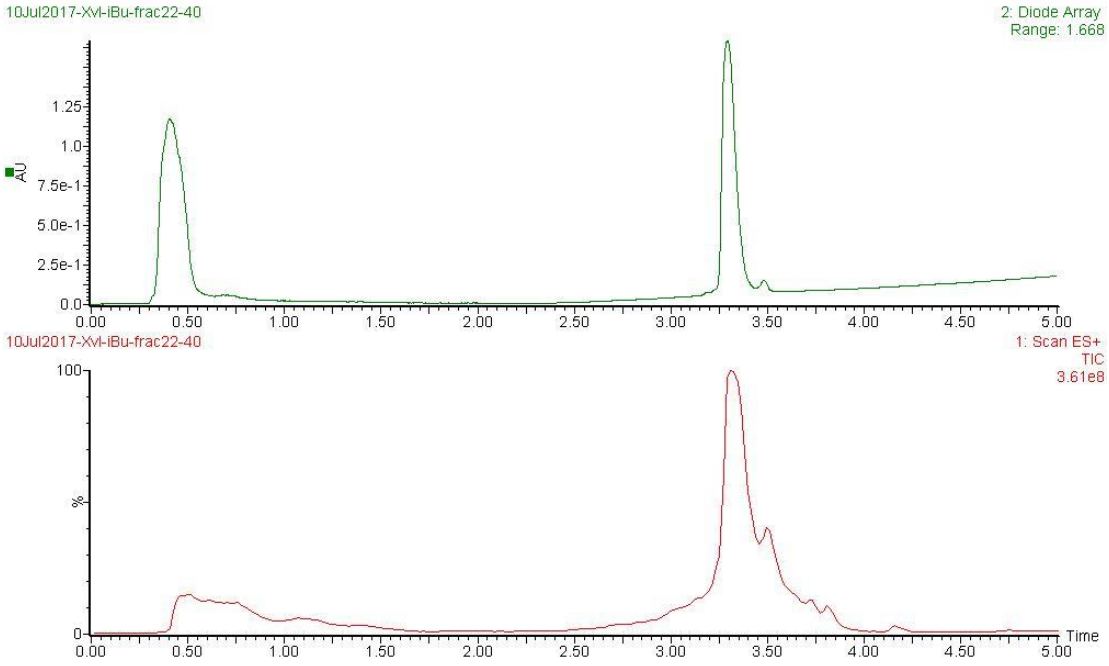
LCMS data for XvA-iBu

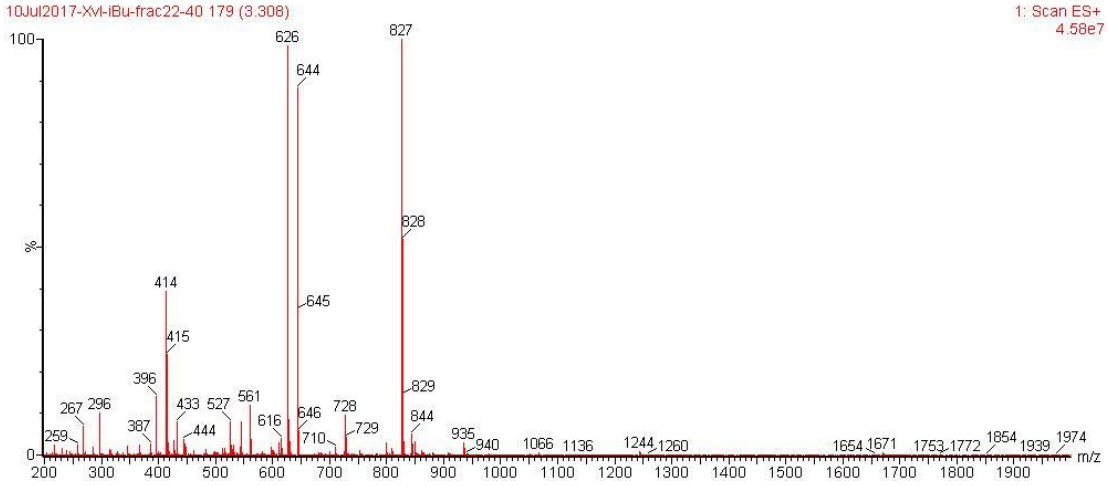


LCMS data for XvI

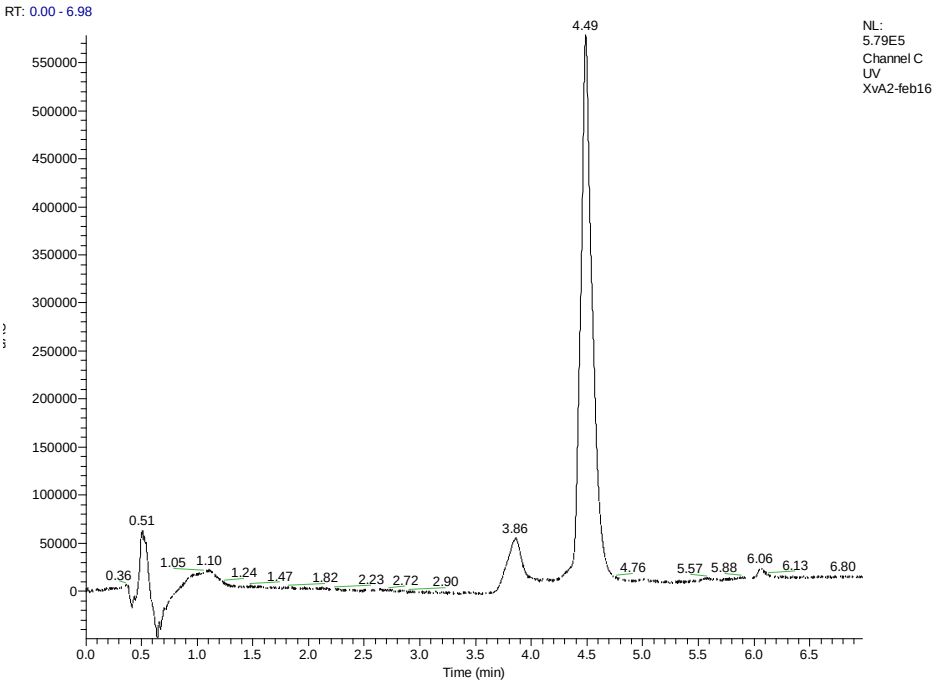


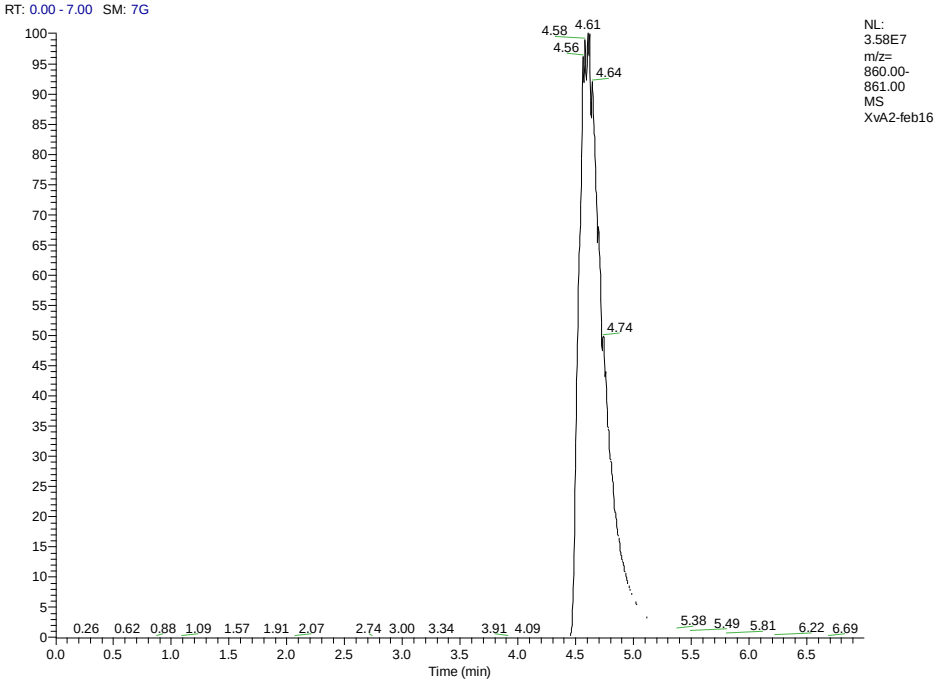
LCMS data for XvI-iBu

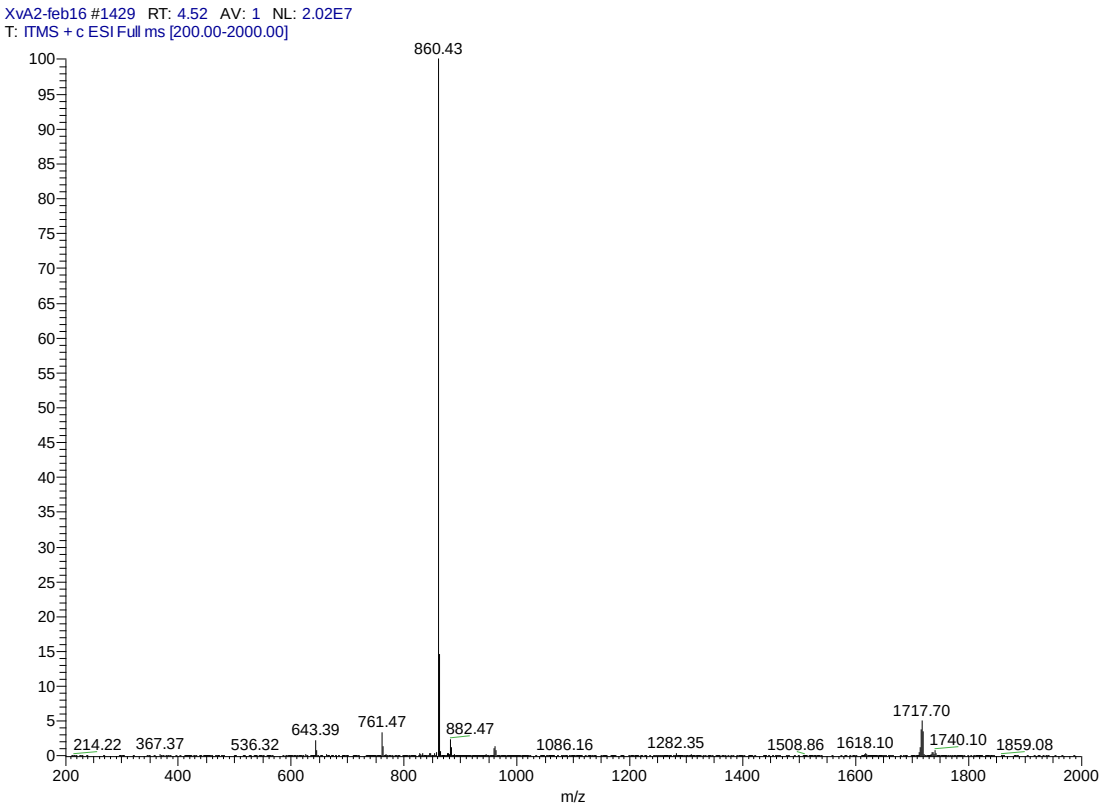




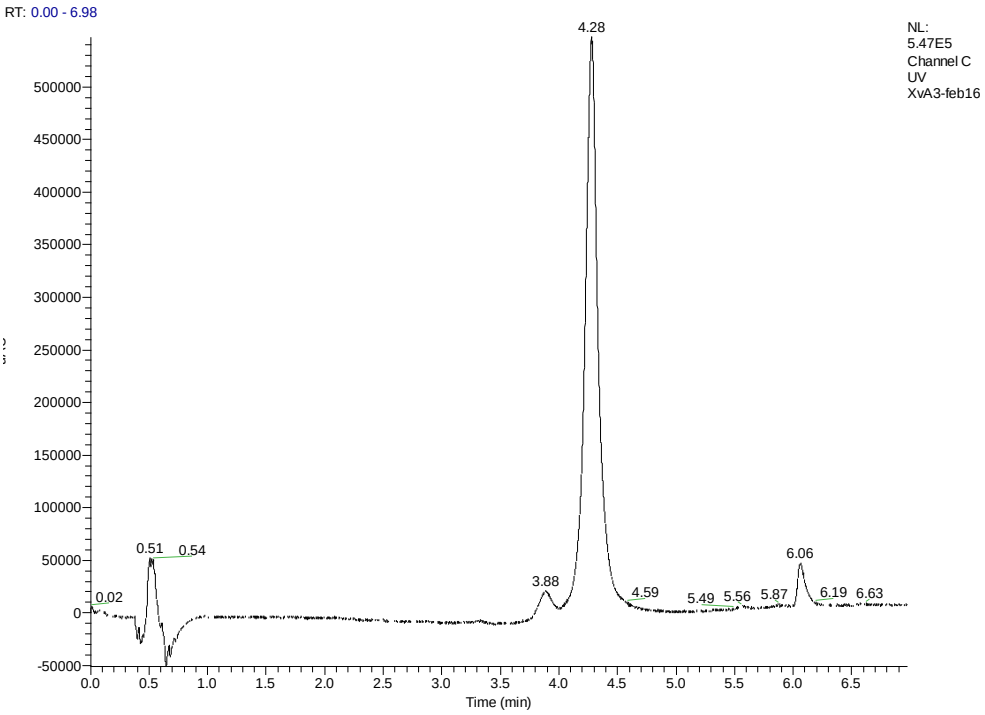
LCMS data for XvA2

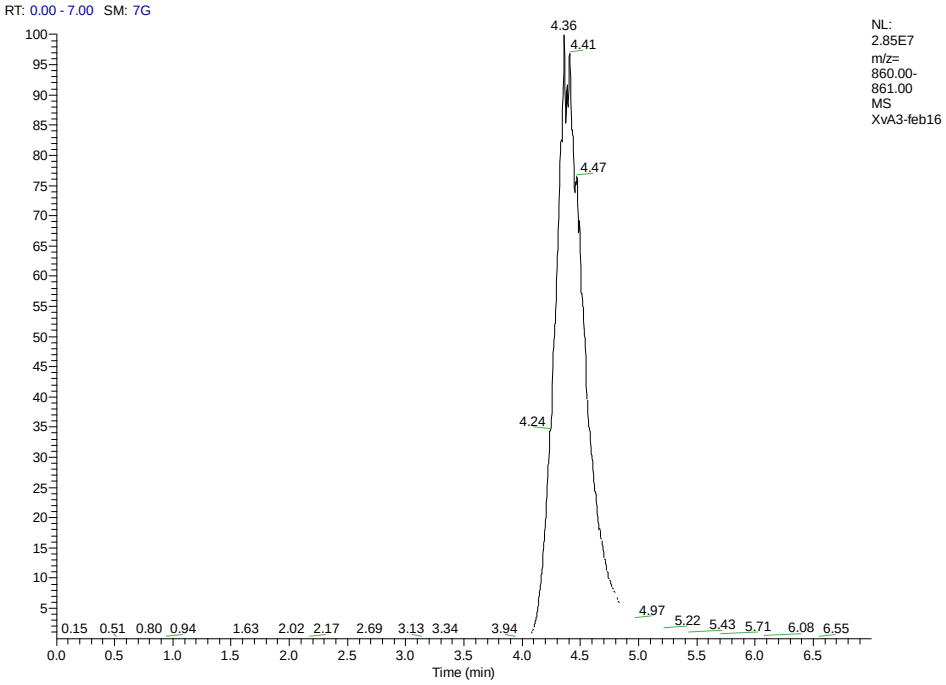


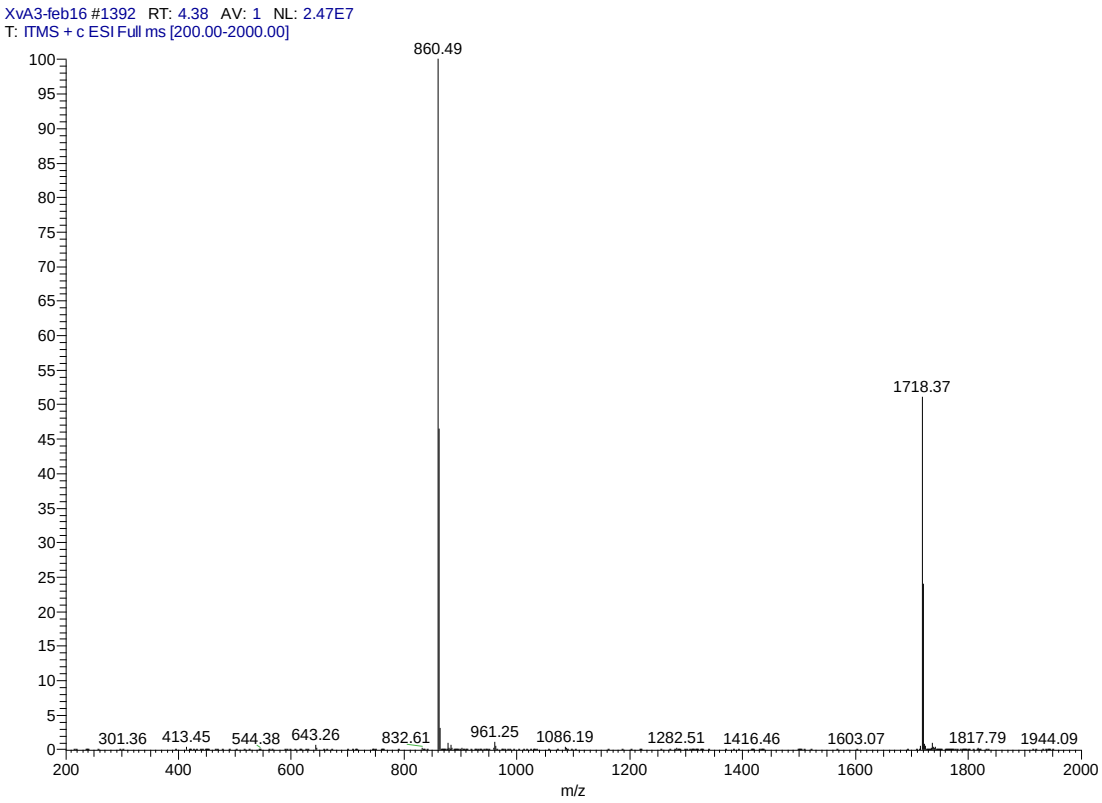



XvA2-feb16 #1429 RT: 4.52 AV: 1 NL: 2.02E7
T: ITMS + c ESI Full ms [200.00-2000.00]
860.43
214.22
367.37
536.32
643.39
761.47
882.47
1086.16
1282.35
1508.86
1618.10
1717.70
1740.10
1859.08
m/z

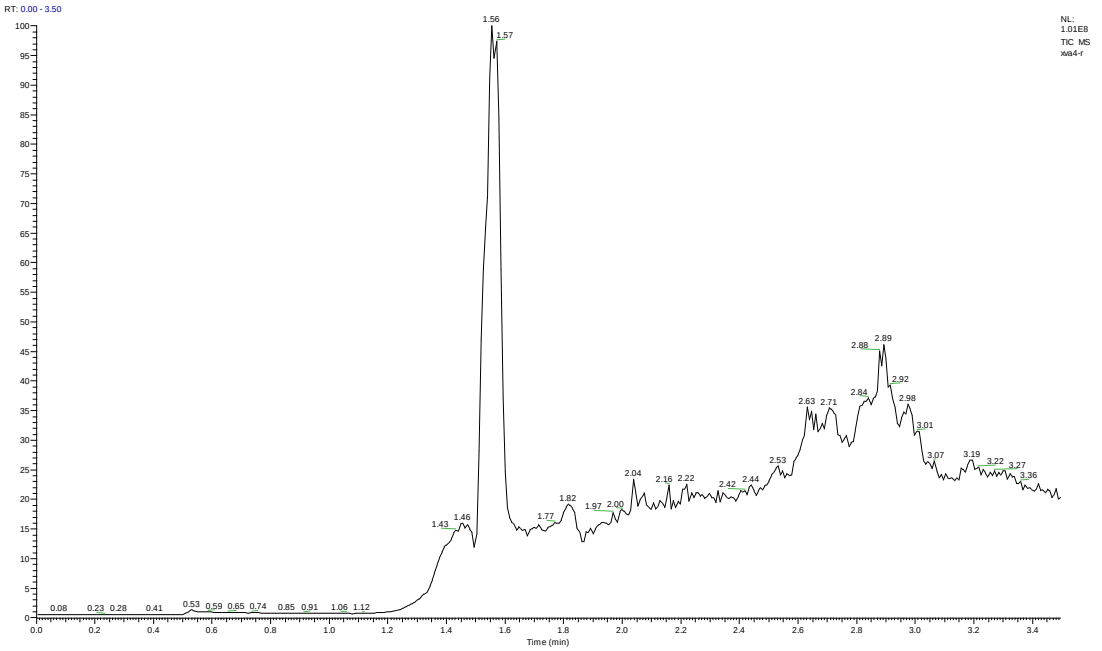
LCMS data for XvA3

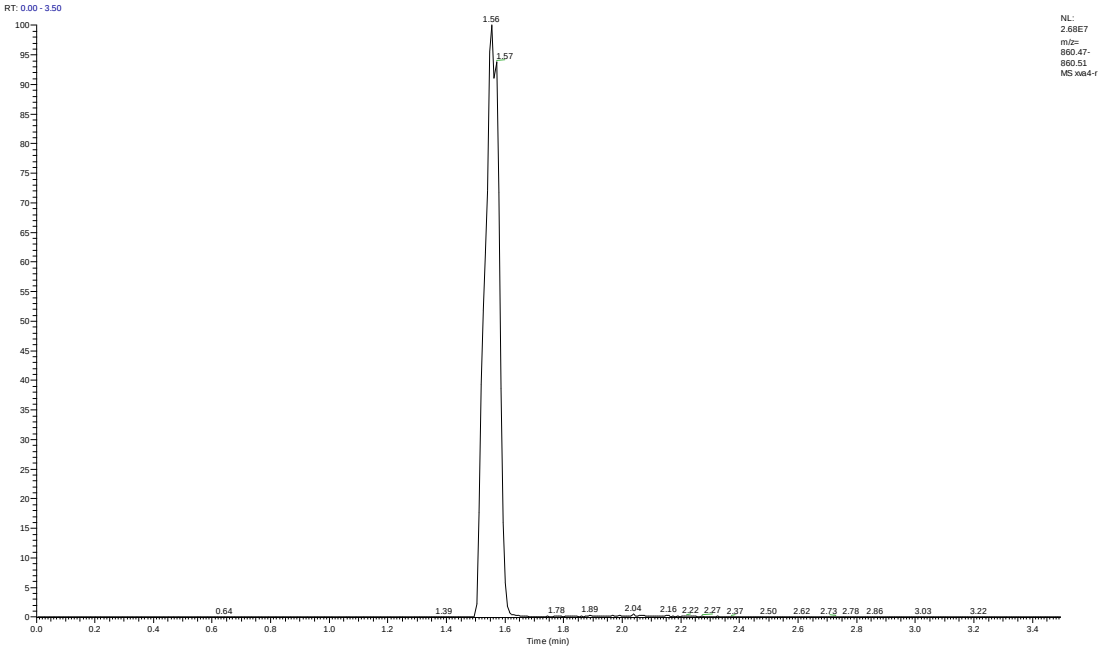


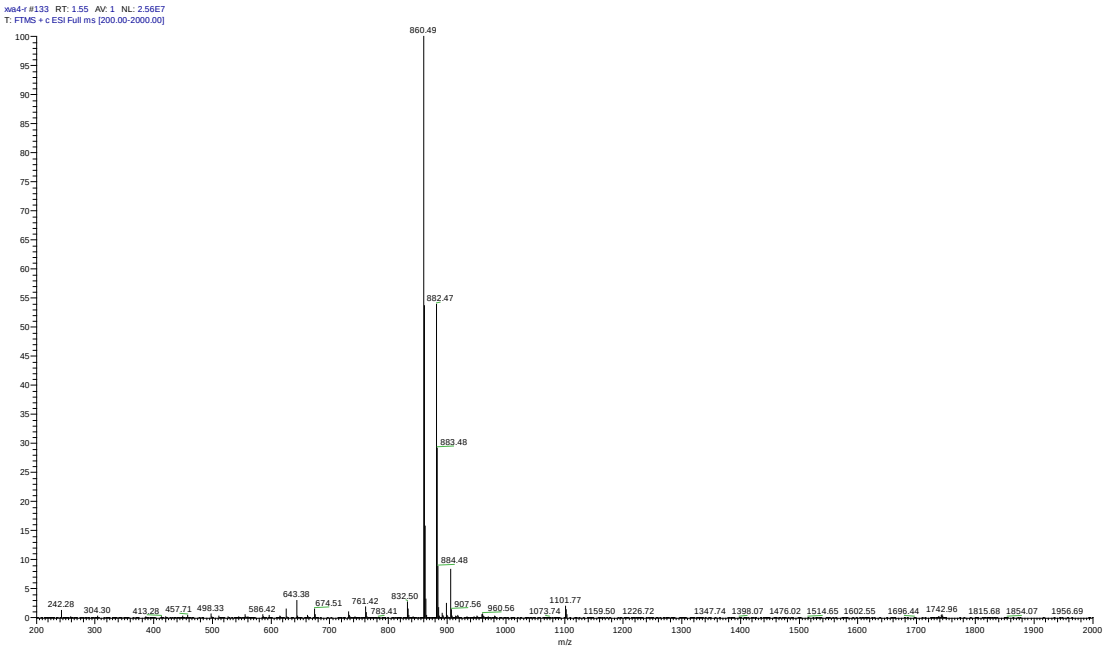



XvA3-feb16 #1392 RT: 4.38 AV: 1 NL: 2.47E7
T: ITMS + c ESI Full ms [200.00-2000.00]
860.49
1718.37
301.36
413.45
544.38
643.26
832.61
961.25
1086.19
1282.51
1416.46
1603.07
1817.79
1944.09
m/z

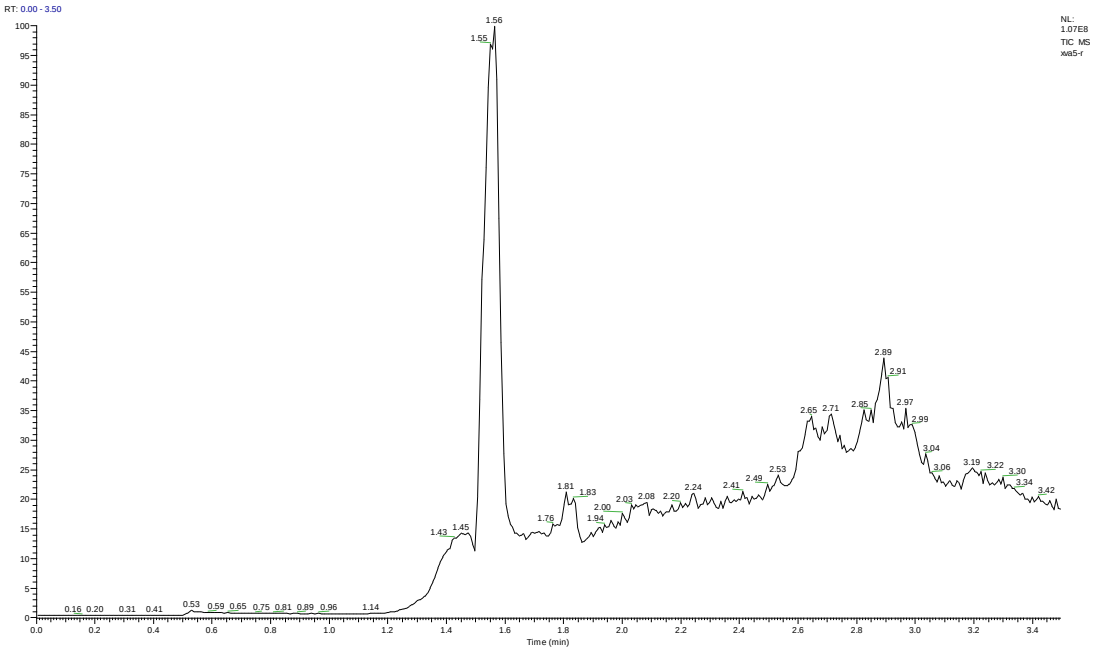
LCMS data for XvA4

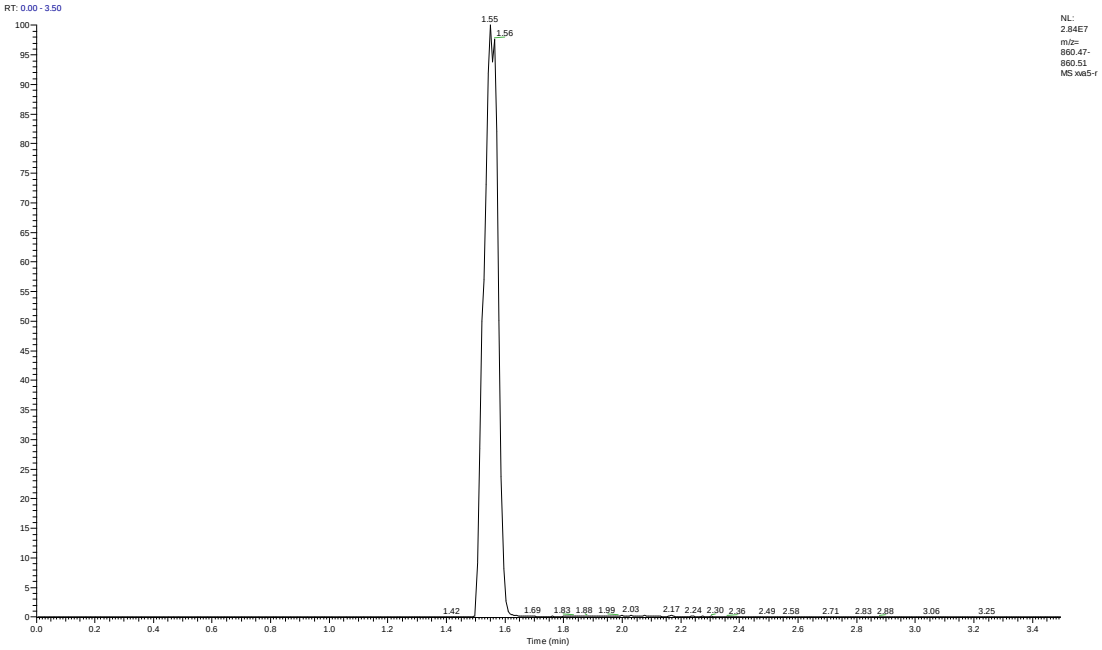


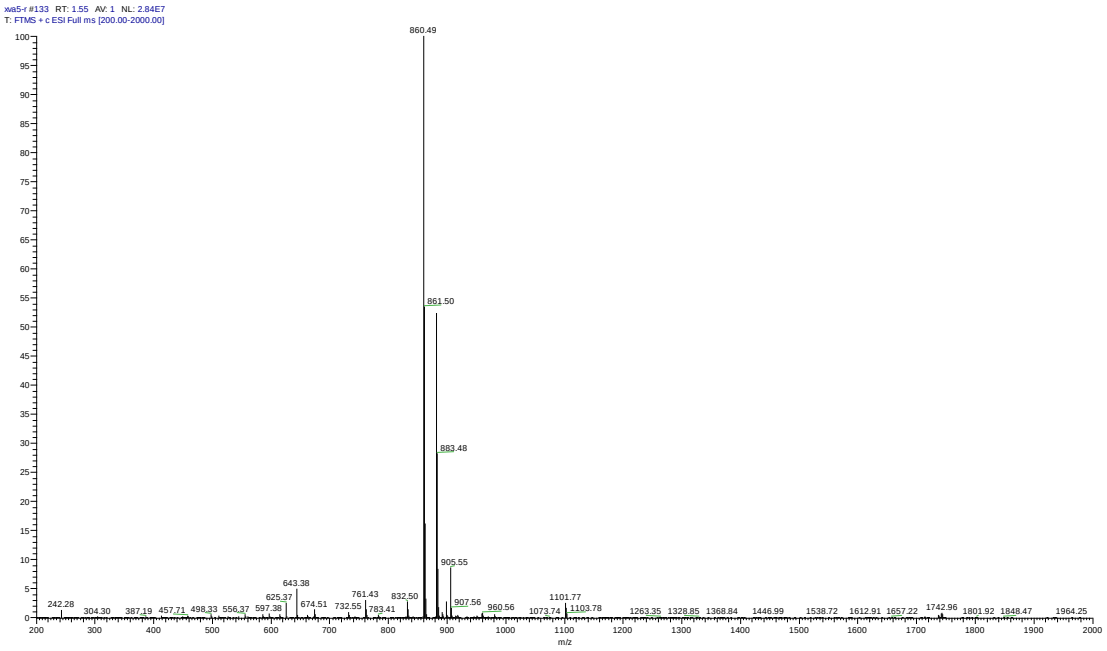



xwa4-r #133 RT: 1.55 AV: 1 NL: 2.56E7
T: FTMS + c ESI Full ms [200.00-2000.00]
860.49
882.47
883.48
884.48
643.38
674.51
761.42
783.41
832.50
907.56
960.56
1101.77
242.28
304.30
413.28
457.71
498.33
586.42
1073.74
1159.50
1226.72
1347.74
1398.07
1476.02
1514.65
1602.55
1696.44
1742.96
1815.68
1854.07
1956.69
m/z

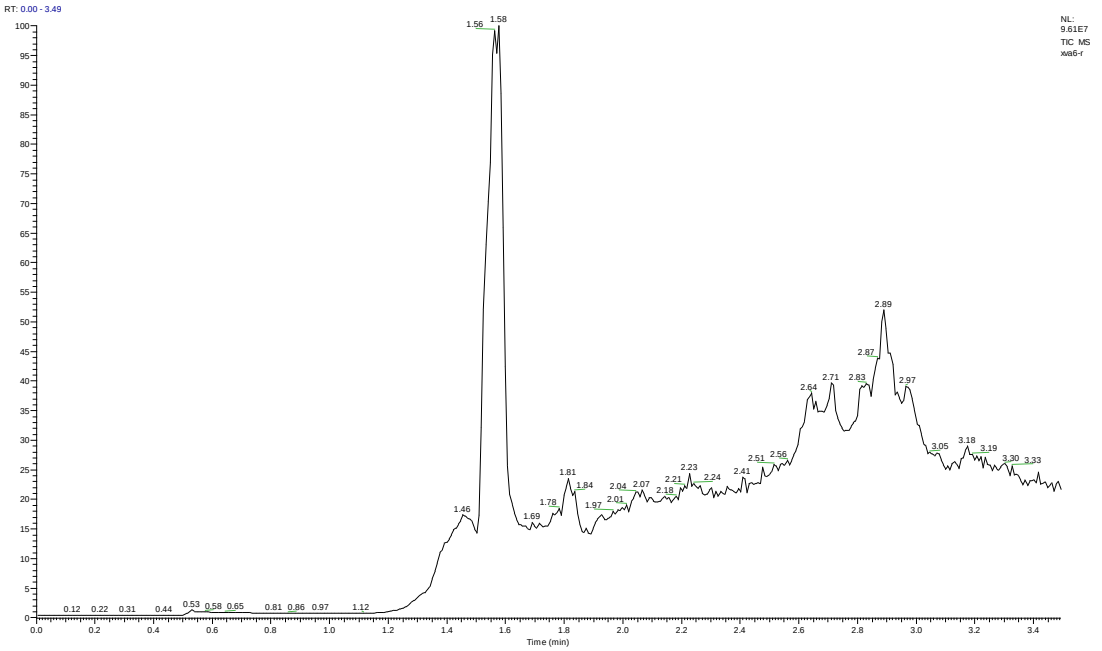
LCMS data for XvA5

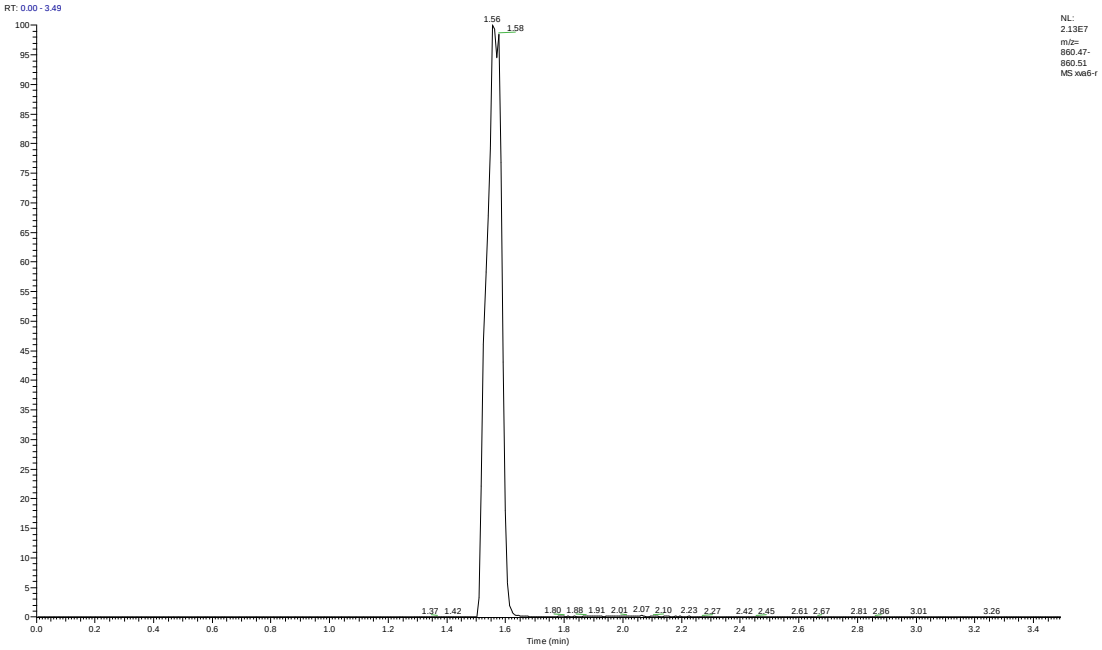


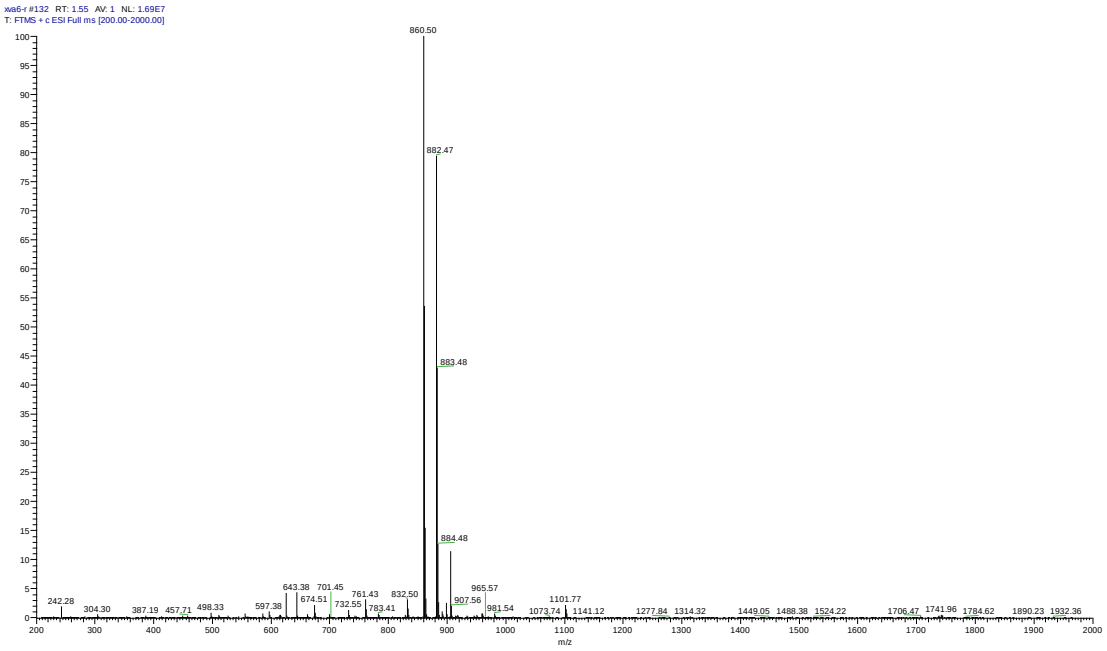



xwa5-r #133 RT: 1.55 AV: 1 NL: 2.84E7
T: FTMS + c ESI Full ms [200.00-2000.00]
860.49
861.50
883.48
905.55
643.38
625.37
761.43
832.50
907.56
960.56
1101.77
1103.78
242.28
304.30
387.19
457.71
498.33
556.37
597.38
674.51
732.55
783.41
1073.74
1263.35
1328.85
1368.84
1446.99
1538.72
1612.91
1657.22
1742.96
1801.92
1848.47
1964.25
m/z

LCMS data for XvA6






xwa6-r #132 RT: 1.55 AV: 1 NL: 1.69E7
T: FTMS + c ESI Full ms [200.00-2000.00]
860.50
882.47
883.48
884.48
643.38
701.45
674.51
761.43
732.55
783.41
832.50
907.56
965.57
981.54
1101.77
1073.74
1141.12
1277.84
1314.32
1449.05
1488.38
1524.22
1706.47
1741.96
1784.62
1890.23
1932.36
242.28
304.30
387.19
457.71
498.33
597.38
m/z