UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Large Scale Reaction Profiling and Determination of Permeability-Relevant Lipophilicity for Macrocyclic Peptides in Complex Mass-Encoded Mixtures and DNA-Encoded Libraries

Permalink

https://escholarship.org/uc/item/3pk2v2r0

Author

Koch, Grant Andrew

Publication Date

2024

Supplemental Material

https://escholarship.org/uc/item/3pk2v2r0#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

Large Scale Reaction Profiling and Determination of Permeability-Relevant Lipophilicity for Macrocyclic Peptides in Complex Mass-Encoded Mixtures and DNA-Encoded Libraries

A dissertation submitted in partial satisfaction

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Grant Andrew Koch

December 2024

The Dissertation of Grant A. Koch is approved:

_____

Professor R. Scott Lokey, chair

_____

Professor Michael Stone

_____

Professor Jevgenij Raskatov

_____

Peter Biehl
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

List of Figures xv

List of Tables xvii

List of Schemes xvii

List of Supplementary Figures xviii

List of Supplementary Tables xx

Abstract xxi

Dedication xxiv

Acknowledgements xxiv

Abbreviations xxvii

Chapter 1: Chromatographic Determination of Lipophilic Permeability Efficiency Facilitates Rapid Identification of Passively Permeable Macrocyclic Peptide Scaffolds 1

1.1 Introduction 1

1.2 Non-linear Regression Model Between Capacity Factor and Partition Coefficients from Diverse Peptide Macrocycles 5

1.2.1 Initial investigation with heptapeptide library 5

1.2.2 Expansion to full training set and nonlinear regression model 7

1.3 Lipophilic Permeability Efficiency Determined by Chromatography 9

1.3.1 Derivation of cLPE equation from model hexapeptides 9

1.3.2 cLPE of PROTAC model compounds 12

1.3.3 Evaluation of cLPE scores across diverse peptide scaffolds 13

1.4 Relationship Between cLPE and Passive Permeability for Scaffold Diverse Thioether Decapeptides 14

1.4.1 Trends in PAMPA permeability 14

1.4.2 Trends in MDCK permeability 16

1.5 PAMPA and Log $ED_{dd/w}$ Trends Across Diverse Peptide Macrocycles 16

1.5.1 PAMPA and Log $ED_{dd/w}$ trends 16

1.5.2 AF-10b chameleonicity effect on PAMPA 18

1.6 PCA of Peptide Libraries Show Similarity to Other Permeable Peptides 19

1.7 McMD Simulations Indicate Log $ED_{dd/w}$ Captures Permeability-Relevant Conformational Changes 20

1.8 Conclusion 22

1.9 Experimental Section 24

1.9.1 Library Information 24

1.9.2 General Materials 25

1.9.3 Synthetic Methods 26

Overview of JT-7 synthesis 26

Overview of JF-10c synthesis 28

1.9.4 Analytical Procedures 32

1.9.5 Computational Procedures 38

Chapter 2: Large Scale Reaction Profiling and Determination of Permeability-Relevant Lipophilicity in DNA-Encoded Libraries 40

2.1 Introduction 40

2.2 Method Optimization with Model Compounds 45

2.2.1 Small molecule dataset 45

2.2.2 AF-Hexapeptomer Dataset 45

2.2.3 WH-Diastereomer Dataset 47

2.3 Scaffold Diverse Macrocyclic Peptide DNA-Encoded Library Design 48

2.4 DEL Chromatographic Partitioning and Sequencing of Fractions 50

2.4.1 Strategy to Obtain Sequencing Count Chromatograms 50

2.4.2 Sequencing Depth of Fractions 53

2.5 Null Informed Model to Guide Retention Time Assignment and Profile Reactivity 53

2.5.1 Identification of Synthetic Truncations Through Comparisons to Null Truncation Chromatograms 53

2.5.2 Split-Tree Visualization Reveals Trends in Library-Wide Reactivity 55

2.6 Library-Wide Assessment of Peak Assignment and Lipophilicity 56

2.6.1 Library-Wide Trends Between Capacity Factor and Peptide Size and Backbone N-Substitution 56

2.6.2 Scaffold-Based Analysis of DNA-Encoded Measurements of Lipophilicity 59

2.7 Trends Between On-DNA Capacity Factors Derived from Sequencing Counts and Off-DNA Capacity Factors and Passive Permeability in PAMPA 62

2.7.1 Resynthesis of Encoded Library Members Off-DNA and Trends Between Capacity Factors On- and Off-DNA 62

2.7.2 Trends Between On-DNA Capacity Factors and PAMPA 64

2.8 Conclusion 65

2.9 Experimental Section 67

2.9.1 General Materials 67

2.9.2 Synthetic Methods 68

Synthesis of AF Liposcan Peptomers 68

Synthesis of WH Diastereomer Peptides 71

Synthesis of Small Molecule-DNA Conjugates 74

Synthesis of AF Liposcan Peptide-DNA Conjugates 77

Synthesis of WH Diastereomer Peptide-DNA Conjugates 80

DNA-Encoded Library Synthesis 82

Resynthesis of Select DNA Encoded Library Members (LDEL001-044) 97

2.9.3 Analytical Procedures 102

LCMS of Off-DNA Peptides 102

Measurement of Capacity Factors (LogK'60) for Off-DNA Peptides 103

LCMS of DNA-Bound Compounds and DNA-Encoded Library 104

Measurement of Chromatographic Lipophilicity for DNA-Bound Compounds and DNA-Encoded Library 105

Gel Electrophoresis 106

qPCR Protocol – Checking Library Stocks 106

Shake-flask Partition Experiment 107

Parallel Artificial Membrane Permeability Assay (PAMPA) 109

Chromatographic Partitioning of DEL, Barcoding Fractions by PCR, and Sequencing 109

2.9.4 Computational Procedures 112

General 112

Library Enumeration 113

Generation of Sequencing Count Chromatograms 113

Simple Gaussian Peak Picking Model (SGPPM) 114

Classic Chrome – Null Informed Model 116

Split-Tree Null Analysis 117

Appendix A: Supplementary Figures and Discussion for Chapter 1 119

Appendix B: Supplementary Figures and Discussion for Chapter 2 141

Library Nomenclature 141

Headpiece Structure and Model Compound Tag Design 141

LDEL DNA Tag and Library Design 142

Classic Chrome Model for Peak Assignment and Reaction Profiling 145

Sequencing Count Chromatogram Cyclization Analysis 148

Appendix C: Selected-mass spectra for Chapter 1 Libraries 151

JT-7 Sublibraries 152

Library A1 152
Library A4 154
Library A7 156
Library B1 158
Library B3 160
Library B5 162

JF-10c Libraries 165

Library G1 165
Library G2 166
Library G3 166
Library G4 167

Library G5 167
Library G6 168
Library Amix 168
Library Bmix 169
Library Cmix 169

Appendix D: LCMS Traces of DNA-encoded Model Compounds 170

AF Hexapeptomers (off-DNA) 170

LCMS of compound **AF_13** off-DNA 170
LCMS of compound **AF_14** off-DNA 170
LCMS of compound **AF_15** off-DNA 171
LCMS of compound **AF_16** off-DNA 171
LCMS of compound **AF_17** off-DNA 172
LCMS of compound **AF_18** off-DNA 172
LCMS of compound **AF_19** off-DNA 173
LCMS of compound **AF_20** off-DNA 173

WH Diastereomers (off-DNA) 174

LCMS of compound **WH_21** off-DNA 174
LCMS of compound **WH_22** off-DNA 174
LCMS of compound **WH_23** off-DNA 175
LCMS of compound **WH_24** off-DNA 175

Small Molecule-DNA Conjugates 176

LCMS of compound **N_3-HP (8 bp)** 176
LCMS of compound **N_3-HP (28 bp)** 177
LCMS of compound **SM_1** 178
LCMS of compound **SM_2** 178
LCMS of compound **SM_3** 179
LCMS of compound **SM_4** 179

LCMS of compound **SM_5** 180
LCMS of compound **SM_6** 180
LCMS of compound **SM_7** 181
LCMS of compound **SM_8** 181
LCMS of compound **SM_9** 182
LCMS of compound **SM_10** 182
LCMS of compound **SM_11** 183
LCMS of compound **SM_12** 183

AF-DNA Conjugates 184

LCMS of compound **N_3-HP (48 bp)** 184
LCMS of compound **AF_13** 185
LCMS of compound **AF_14** 186
LCMS of compound **AF_15** 187
LCMS of compound **AF_16** 188
LCMS of compound **AF_17** 189
LCMS of compound **AF_18** 190
LCMS of compound **AF_19** 191
LCMS of compound **AF_20** 192
AF-**13-20** Single Compound Injections (Same Conditions as Coinjection) 193

WH-DNA Conjugates 194

LCMS of compound **WH_21** 194
LCMS of compound **WH_22** 194
LCMS of compound **WH_23** 195
LCMS of compound **WH_24** 195

Appendix E: LCMS Traces and Gels of DNA-encoded Library Synthesis 196

Library Starting Materials 196

LCMS of library starting material, **Fmoc-Ado-HP-Primer** (21 bp) 196

LCMS of library starting material, **H_2N-Ado-HP-Primer** (21 bp)..... 197
LCMS of library starting material, **Fmoc-Prg-Ado-HP-Primer** (21 bp)..... 198
LCMS of library starting material, **H_2N-Prg-Ado-HP-Primer** (21 bp)..... 199
Library Synthesis Cycle 1..... 200
Gel electrophoresis of **Cycle 1 ligations** with controls (5% agarose, 1X TAE running buffer)..... 200
Representative gel electrophoresis of **Cycle 1 ligations** (5% agarose, 1X TAE running buffer)..... 200
LCMS of representative **Cycle 1 ligation, well G6** (BB_{33} = Fmoc-Leu-D-Leu(NMe)-D-Pro-OH)..... 201
LCMS of representative **Cycle 1 coupling, well G6** (BB_{33} = Fmoc-Leu-D-Leu(NMe)-D-Pro-OH)..... 202
LCMS of **Crude Pooled Cycle 1 Fmoc-Library** (34 bp)..... 203
HPLC UV-Vis Trace of **Cycle 1 Purification**..... 203
LCMS of **Purified Pooled Cycle 1 H_2N-Library** (34 bp)..... 204
Library Synthesis Cycle 2..... 204
Gel electrophoresis of representative **Cycle 2 ligations** (5% agarose, 1X TAE running buffer)..... 204
LCMS of representative **Cycle 2 ligation, well C11** (BB_{12} = Fmoc-LA18-OH)..... 205
LCMS of representative **Cycle 2 coupling, well D5** (BB_{20} = Fmoc-D-Leu-Leu-Pro-OH)..... 206
LCMS of **Crude Cycle 2 Pooled Fmoc-Library** (47 bp)..... 207
LCMS of **Cycle 2 Pooled H_2N-Library** (47 bp)..... 208
Library Synthesis Cycle 3..... 209
Gel electrophoresis of representative **Cycle 3 ligations** (5% agarose, 1X TAE running buffer)..... 209

LCMS of representative **Cycle 3 ligation, well 2-A1** (BB_{30} = Fmoc-D-Leu-D-Leu(NMe)-Pro-OH) 209
LCMS of representative **Cycle 3 coupling, well 1-E4** (BB_{52} = Fmoc-D-Leu-LA03-Pro-OH) 210
LCMS of **Crude Pooled Cycle 3 Fmoc-Library** (60 bp) 211
HPLC UV-Vis Trace of **Cycle 3 Purification** 211
LCMS of **Pure Pooled Cycle 3 H_2N-Library** (60 bp) 212
Final Library Reactions 213
LCMS of **N_3-Library** (60 bp) 213
Gel electrophoresis of **Library ID tag ligations** 214
LCMS of **N_3-Library-LID** (72 bp) 214
LCMS of **Cyclized-Library-LID** (72 bp) 215
Gel electrophoresis of **Pool ID tag ligations** 216
LCMS of **Linear-Library-LID-PID** (114 bp) 216
LCMS of **Cyclized-Library-LID-PID** (114 bp) 217
Appendix F: LCMS Traces, HNMR, and Classic Chrome Sequencing Count Chromatograms for LDEL001-044 218
LDEL001 218
LDEL002 221
LDEL004 224
LDEL005 227
LDEL006 230
LDEL007 233
LDEL008 236
LDEL011 239
LDEL012 242
LDEL015 245
LDEL016 248

LDEL019..... 251
LDEL020..... 254
LDEL023..... 257
LDEL024..... 260
LDEL025..... 263
LDEL027..... 266
LDEL028..... 269
LDEL029..... 272
LDEL030..... 275
LDEL032..... 278
LDEL034..... 281
LDEL035..... 284
LDEL037..... 287
LDEL038..... 290
LDEL040..... 293
LDEL041..... 296
LDEL042..... 299
LDEL043..... 302
LDEL044..... 305

Appendix G: LCMS Traces and HNMR of Custom Tripeptide Building Blocks... 308

BB019 308
BB020 310
BB021 312
BB022 314
BB023 316
BB024 318
BB025 320
BB026 322
BB027 324

BB028 326
BB029 328
BB030 330
BB031 332
BB032 334
BB033 336
BB034 338
BB035 340
BB036 342
BB037 344
BB038 346
BB039 348
BB040 350
BB041 352
BB042 354
BB044 358
BB046 362
BB047 364
BB048 366
BB049 368
BB050 370
BB051 372
BB052 374
BB053 376
BB054 378
BB055 380
BB056 382
BB057 384
BB058 386

References 388

List of Figures

Figure 1: Non-linear regression between measured shake-flask lipophilicity and chromatographically determined lipophilicity. A) Structure and property data of cyclic peptide libraries. Circles in JF-10c and AF-10a-b represent either a peptoid or a N-methylated alpha amino acid. B) Relationship between shake-flask lipophilicity and chromatographic lipophilicity. Points are colored by cyclic peptide library (1a). C) Relationship between estimated shake-flask lipophilicity (Log $ED_{dd/w}$) and measured shake-flask lipophilicity (Log $D_{dd/w}$) of test set compounds. 6
Figure 2: Lipophilic permeability efficiency determined by chromatography (cLPE) for bRo5 compounds. A) Structure of the cyclic hexapeptide library with exposed hydrogen bond donors colored in orange. B) Linear regressions of Log $ED_{dd/w}$ and ALogP for each peptide class. Open circles have no exposed HBD (class A), squares have one side-chain HBD (class B), stars have one backbone HBD (class C), and X's have two exposed HBDs (class D). Compounds are binned by ALogP (<3, blue; 3-4, green; >4 orange). C) Structure of PROTAC model compounds. D) cLPE plot comparing amide to ester matched pairs. Compounds are colored by scaffold (esters, green; amides, blue). 10
Figure 3: cLPE analysis of stereodiverse thioether-cyclized decapeptides from JF-10a. A) Structure of the library. B) Relationship between Log $ED_{dd/w}$ and ALogP. Compounds are colored by PAMPA LogPe in five evenly distributed bins of Lowest (-8.5 < LogPe < -7.8), Low (-7.8 < LogPe < -7.1), Medium (-7.1 < LogPe < -6.4), High (-6.4 < LogPe < -5.7), and Highest (-5.7 < LogPe < -5.1). C) Relationship between MDCK cell permeability and Log $ED_{dd/w}$. Compounds are colored by Log $ED_{dd/w}$ (<2, blue; >2, green). The linear regression for compounds with a Log $ED_{dd/w}$ <2 is displayed with a black line ($R^2 = 0.754$). 15
Figure 4: Relationship between PAMPA permeability and Log $ED_{dd/w}$ for scaffold-diverse decapeptides. A) Structures of decapeptide libraries. Circles in JF-10a-c represent either a peptoid or an N-methylated alpha amino acid. B) Relationship between PAMPA permeability and Log $ED_{dd/w}$. Compounds are colored in accordance with their library ID. 17
Figure 5: McMD structures of diastereotopic, thioether-cyclized matched pairs in cyclohexane. Each plot (X1-X3) represents a conformational cluster, where the intensity of the color reflects the relative occurrence of an IMHB across all structures within a cluster. The lowest RMSD frame from each X1-X3 conformational cluster is overlaid for SB60 (purple) and SB64 (orange). Dashed lines represent intramolecular hydrogen bonds which are also represented as green squares in the accompanying donor-acceptor plots. 21
Figure 6: Chromatogram of a selected ion monitored spectra of diastereotopic, thioether-cyclized peptides, SB60 and SB64 (left and right, respectively). 22

Figure 7: Chromatographic lipophilicity of various DNA-bound substrates. A) Small molecule alkynes clicked onto 28 bp azido-AOP-HP. B) Cyclic hexapeptomer liposcan clicked onto 48 bp azido-AOP-HP via propargyl tyrosine. C) Cyclic hexapeptide diastereomers clicked onto 48 bp azido-AOP-HP. On-DNA lipophilicities strongly correlate with both off-DNA partition coefficients with 1,9-decadiene and water and CACO-2 cell permeability. 46
Figure 8: DNA-encoded library design. All possible stereoisomers (highlighted in blue) were used as building blocks. N-methyl nucleophiles (BB13-18, 35-42, 55-58) were only used at cycle 3. 49
Figure 9: Chromatographic separation and sequencing of library fractions. A) Chromatographic elution of a 1:1 mixture of linear and cyclized library represented by qPCR quantification in each fraction. Library elution is shown in green and blue, with Null library in black. B) Box-and-whisker plots of deduplicated sequencing counts per compound per fraction. C) Matched pair example of sequencing count chromatograms. Two DNA-encoded library members who differ by one side chain (Val, blue; Phe, green). 51
Figure 10: Null-informed chromatogram analysis A) Scheme showing the result of null tag incorporation during library synthesis for a single compound (Val-Phe-Leu). B) Sequence count chromatogram for compound "Val-Phe-Leu" from 10-40%B library elution. The full compound (FC, black), a single truncation product (ST, green) and a double truncation product (DT, blue) are displayed. Peak overlap between FT and ST indicates incomplete product formation for "Val-Phe-Leu". C) Sequence count chromatogram for compound "LeuMe-LA03-Pro-Phe-Leu" from 10-40%B library elution. The full compound (FC, black), two single truncation products (ST1, dark green, ST2, light green), and a double truncation product (DT, blue) are displayed. Lack of overlap with any associated truncation products indicates good product formation for "LeuMe-LA03-Pro-Phe-Leu". D) Split-tree plot of full DNA-encoded library (LDEL). Each node represents a building block at a given cycle of chemistry, resulting in 32 BB1 clusters, 32 BB2 clusters, and 58 terminal nodes per BB2 cluster. Terminal nodes which are blue indicate a compound whose full compound product peak does not match any of its associated truncations, while red nodes indicate a compound which matches an associated truncation. BB3 connections and nodes are not displayed in the full LDEL split tree for visual clarity. 57
Figure 11: Assessment of library-wide retention time assignments. Box and whisker plots of the capacity factors (retention time) for every compound from the full DNA-encoded library (A) or null pruned DNA-encoded library (B), binned by both ring size and number of N-substituted residues (peptoid or N-methyl). C) Heptapeptide scaffold matched pairs who differ by stereoinversion of a single leucine residue and display a large difference in on-DNA chromatographic capacity factors for each side-chain variant. Linear regressions are displayed for each scaffold. D) Heptapeptide scaffold matched pairs who differ by stereoinversion of their proline residue and contain 16 liposcan members per scaffold. Linear regressions are displayed for each scaffold. 60

Figure 12 Assessment of on-DNA lipophilicity measurements from the DEL. A) Simplified structures of the resynthesized compound sets. The connection to the DNA has been replaced with an ethyl ester for off-DNA testing. B) Relationship between on-DNA capacity factors and off-DNA capacity factors. An excellent correlation (R^2 = 0.936) is observed across all ring sizes. C) Relationship between PAMPA permeability and calculated lipophilicity (AlogP). Vertical bands of compounds indicate isomers or stereoisomers. D) Relationship between PAMPA permeability and on-DNA capacity factors. Linear regressions for 5mers, 7mers, and 9mers are displayed directly on the plot. 63

List of Tables

Table 1: Summary of compound libraries used throughout this study 19
Table 2: List of LCMS methods used in this study. 32
Table 3: List of all equations used for Log $ED_{dd/w}$ and cLPE conversions..... 33
Table 4: List of HPLC and LCMS methods used in this study 102

List of Schemes

Scheme 1: Detailed split-and-pool library design of JT-7 library..... 27
Scheme 2: Synthetic scheme for JT-7 compounds..... 27
Scheme 3: Synthetic scheme for JF-10c compounds. (i) TFA (ii) COMU, DIPEA.. 29
Scheme 4: Solid phase peptide synthesis and cyclization of AF Liposcan Peptomers 68
Scheme 5: Installation of alkyne handle for AF Liposcan Peptides 70
Scheme 6: Solid phase peptide synthesis and cyclization of WH Diastereomer Peptides 71
Scheme 7: Installation of alkyne handle for WH Diastereomer Peptides 72
Scheme 8: Synthesis of DNA-Headpiece starting material, N_3-AOP-HP-T1..... 74
Scheme 9: Synthesis of small molecule (SM)-DNA conjugates..... 75
Scheme 10: Synthesis of DNA-headpiece starting material, N_3-HP-T1-T2 77
Scheme 11: Synthesis of Peptide Liposcan(AF)-DNA conjugates..... 78
Scheme 12: Synthesis of Peptide Diastereomer-DNA conjugates..... 80
Scheme 13: Synthesis of DNA-Encoded Library starting material, H_2N-Prg-Ado-HP-Primer..... 82
Scheme 14: Cycle 1 ligations and couplings. Exact masses represent the average for the library 84
Scheme 15: Cycle 2 ligations and couplings. Exact masses represent the average for the library 87
Scheme 16: Cycle 3 ligations and couplings. Exact masses represent the average for the library 90
Scheme 17: Library coupling to azido acetic acid. Exact masses represent the average for the library 93
Scheme 18: Copper-mediated azide-alkyne macrocyclization. Exact masses represent the average for the library 94

Scheme 19: Library ID Ligations. Exact masses represent the average for the library 95
Scheme 20: Closing Primer Ligations. Exact masses represent the average for the library 96
Scheme 21: Loading of trityl alcohol lanterns for synthesis of off-DNA library members. Trityl alcohol lanterns were affixed with an RFID tag (green line) for compound tracking and identification during parallel compound synthesis. 97
Scheme 22: Solid phase peptide synthesis of off-DNA peptides from LDEL on RFID-encoded lanterns 98

List of Supplementary Figures

Figure S1: Summary of all compound libraries used in this study. Individual compound information can be found in the associated spreadsheets included in the supporting information. Colored dots correspond to the coloration implemented in **Figure 1** and **4** found in the main text. Undefined stereocenters signifies the use of both diastereomers at that position. 119
Figure S2: Comparison of columns and methods for capacity factor measurements (LogK) which reflect decadiene-water partition coefficient values for the JT-7 library. Method and column descriptions are listed above each plot. The linear regression equations and coefficients of determination are displayed directly on each plot. 121
Figure S3: Comparison of linear and exponential regression model fits for a long gradient and an isocratic method on a Hamilton 2.1x250 mm PRP-C18 column. A) Exponential regression fit for isocratic 60%B. B) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the isocratic 60%B method. C) Exponential regression fit for 20-100%B. D) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the 20-100%B method. E) Linear regression fit for isocratic 60%B. F) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the isocratic 60%B method. G) Linear regression fit for 20-100%B. H) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the 20-100%B method. 123
Figure S4: Comparison of linear and exponential regression model fits for a long gradient and an isocratic method on an Acclaim™ Vanquish RSLC C18 2.2 µm 2.1 x 250 mm column. A) Exponential regression fit for isocratic 60%B. B) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the isocratic 60%B method. C) Exponential regression fit for 20-100%B. D) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the 20-100%B method. E) Linear regression fit for isocratic 60%B. F) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the isocratic 60%B method. G) Linear regression fit for 20-100%B. H) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the 20-100%B method. 124

Figure S5: Comparison of exponential regression model fits for low flow rate (0.2 mL/min) and high flow rate (1.0 mL/min) on a Hamilton 2.1x250mm PRP-C18 column using an isocratic method. A) Exponential regression fit for isocratic 60%B with a 0.2 mL/min flow rate. B) Comparison of predicted vs experimental shake-flask values using low flow rate method. C) Exponential regression fit for isocratic 60%B with a 1.0 mL/min flow rate D) Comparison of predicted vs experimental shake-flask values using the high flow rate method. 125
Figure S6: Comparison of capacity factors for MN-6a-d compounds measured on either the long Hamilton PRP-C18 column (2.1x250 mm) with compound mixtures or on a shorter but otherwise identical Hamilton PRP-C18 column (2.1x33 mm) with single compound injections. Open circles have no exposed HBD (class A), squares have one side-chain HBD (class B), stars have one backbone HBD (class C), and X's have two exposed HBDs (class D). Compounds are colored by ALogP (<3, blue; 3-4, green; >4 orange). 127
Figure S7: Comparison of Log $ED_{dd/w}$ derived from PRP-C18 a) LogK'60 or b) LogK'20-100 values and AlogP for library MN-6a-d. cLPE equations derived from the minimized slopes and intercepts are displayed above the plots. 128
Figure S8: Comparison of LogK'60 for acidic and neutral pH in the aqueous phase. Structures of libraries, with ionizable centers if applicable, are displayed directly to the left of the plots. 129
Figure S9: Comparison of the correlation between PAMPA and Log $D_{dd/w}$ (orange) or Log $ED_{dd/w}$ (blue) for the MN-6a-d hexapeptides. Observed solubility cutoff indicated with a dotted black line (Log $D_{dd/w}$ = 1, Log $ED_{dd/w}$ = 2). Linear regressions are plotted with solid black lines for compounds within the observed soluble regime. Coefficients of determination for the linear regressions displayed directly on each plot. 130
Figure S10: cLPE analysis of individual liposcan libraries for Log $ED_{dd/w}$ measured on PRP-C18 with an isocratic method. Each linear correlation is defined by Log $ED_{dd/w}$ = 1.01*ALogP – b_{avg}, for compounds with Log $ED_{dd/w}$ < 2. RMSD between measured values and projected y-values derived from the linear correlation are reported. cLPEs are directly calculated by subtracting the y-intercept of MN6-d from the y-intercept of each liposcan scaffold (cLPE = $y_{scaffold} - y_{MN6\text{-}d}$). 131
Figure S11: Trend between PAMPA LogPe and Log $ED_{dd/w}$ for each JF-10a thioether stereoisomer group (defined by compounds having identical AlogP, **Figure 3**). Linear regression lines are displayed in black with the regression equation and coefficient of determination displayed above. 134
Figure S12: PAMPA permeability plotted against Log $ED_{dd/w}$ for AF-10a (squares), AF-10b (circles), and Cyclosporine A (diamond). Compounds are colored by ChamelogK value (darkest, 1.362-1.213; dark, 1.212-1.063; medium, 1.062-0.914; light, 0.913-0.765; lightest, 0.764-0.616) with averages for each scaffold displayed directly on the plot. Tabulated values are displayed to the right with conditional formatting applied (highest values dark green, lowest values white). Compounds with red Chamelogk values in the table were not observed in one of the four isocratic

LogK measurements and were omitted from the plot. Carbamazepine (Cmz) and Cyclosporin A (Csa) were used as standards. 137
Figure S13: Principal component analysis comparing all macrocyclic peptides used in this study (dark blue) with all compounds found on CycPeptMPDB (light blue). PCA was conducted using either a) molecular descriptors or b) Morgan fingerprints as features..... 140
Figure S14: Structure of headpiece and linkers used in this study 141
Figure S15: Structure and sequences of DNA tags for model compounds **1-24**. 142
Figure S16: Illustration of tag design used in LDEL. 142
Figure S17: Physicochemical properties of LDEL. The library diversity reported in the table does not include diversity from null encoding. 144
Figure S18: Full compound and null-truncation chromatograms for LDEL compound Val-Phe-Leu. The full compound peak closely matches the compound peak associated with ST1, Phe-Leu, indicating the full compound is a truncated byproduct and a failed Val coupling in the third cycle of chemistry. 145
Figure S19: Full compound and null-truncation chromatograms for LeuMe-LA03-Pro-Phe-Leu. The full compound peak does not align with the assigned compound peaks for any associated Null truncation, indicating a successful synthesis of this encoded library member. 146
Figure S20: Sequencing count chromatograms of an 11mer matched pair which differs only by the N-methylation about their terminal amine (prior to final coupling with azido acetic acid, highlighted in yellow). 148
Figure S21: Comparison between compounds which end Cycle 3 with an N-terminal substitution (methyl or R) and compounds which end with an N-terminal NH. Cyclized compounds generally exhibited higher longer retention times than their linear counterparts. 150

List of Supplementary Tables

Table S1 Linear regression analyses of Log $D_{dd/w}$ and LogK using several chromatographic conditions and columns. 122
Table S2: Comparison of m_{lipo} between shake-flask and chromatographic lipophilicity measurements for liposcans of 12 different peptide scaffolds and PROTACs 133
Table S3: Thioether stereoisomer matched pairs with significant differences in PAMPA permeability which are reflected by increases in Log $ED_{dd/w}$. All matched pairs are the result of a single or double stereochemical inversion in the backbone, highlighted in yellow. Matched pair "BDL" (highlighted in blue) was featured in the main text and was subject to modeling by McMD (Figure 5 and 6). 135

Abstract

Title: Investigation of Chromatographic Lipophilicity and Passive Cell Permeability for Macrocyclic Peptides in Complex Mass Encoded Mixtures and DNA-Encoded Libraries

Author: Grant Andrew Koch

Lipophilicity is one of the most important physicochemical parameters to control during early-stage drug development. Optimization of lipophilicity is critical especially for larger compounds, as the margin for achieving both acceptable passive cell permeability and aqueous solubility thins tremendously as molecular weight increases. Larger drug leads have recently become more highly pursued due to the advent of encoded library technologies (DEL, mRNA display) and for their unique ability to bind to extended protein surfaces and disrupt protein-protein interactions. To ensure that early-stage lead compounds beyond the rule-of-5 are optimized to simultaneously exhibit acceptable aqueous solubility and passive cell permeability, a rapid and accurate empirical measurement of lipophilicity is desirable. The gold standard of lipophilicity is shake-flask partition coefficients, which are prone to sensitivity issues and experimental error. However, chromatographic assessments of lipophilicity which reflect hydrocarbon/water partition coefficients have significant potential to rapidly assess passive cell permeability and aqueous solubility profiles for early-stage peptide macrocycles in development. By applying the combinatorial and encoding power of DNA-encoded libraries, chromatographic measurements of lipophilicity could rapidly translate DEL hits into lead compounds by selecting

compounds with the highest potential for passive cell permeability and aqueous solubility.

Chapter 1 is an investigation of the potential to use chromatographic measurements of lipophilicity to reproduce 1,9-decadiene/water partition coefficients and predict passive cell permeability for complex, mass-encoded libraries of peptide macrocycles. We constructed a non-linear regression model from over 200 peptide macrocycles which can estimate 1,9-decadiene/water partition coefficients (Log $D_{dd/w}$) from a single capacity factor measurement. This model predicts Log $D_{dd/w}$ values with high accuracy for a diverse set of peptide macrocycles. These estimated values, called Log $ED_{dd/w}$, can be used to determine lipophilic permeability efficiency (cLPE), a quantitative metric that is descriptive of permeability-relevant lipophilicity and is directly related to the amount of exposed hydrogen bond donors. Log $ED_{dd/w}$ correlates strongly with PAMPA and MDCK permeability for peptide macrocycles in the soluble regime. This chromatographic lipophilicity assay is sensitive to minute structural differences, such as inversion of a single stereocenter, and MDCK simulations demonstrate that it can capture desirable conformational features for peptide macrocycles, such as intramolecular hydrogen bonding.

In Chapter 2, chromatographic conditions for assessing individual compound purity and measuring permeability-relevant lipophilicity were developed for DNA-encoded libraries (DELs). Optimized conditions were developed on model compounds, featuring a series of hexapeptide diastereomers conjugated to 48bp DNA whose on-DNA capacity factors reflect their off-DNA CACO-2 cell permeabilities. A 62,581

member DNA-encoded library of peptide macrocycles was synthesized which features significant structural diversity in the peptide backbone. A novel method was developed to partition the library chromatographically and sequence the collected library fractions to produce sequencing count chromatograms which can be used to assess purity and calculate retention times for individual library members. Additionally, an analytical framework was developed utilizing encoded null truncations to identify synthetic truncations from the library synthesis and determine the relative purity of each DNA-encoded library member. The on-DNA retention times from the sequencing count chromatograms strongly correlated with off-DNA retention times and PAMPA permeability for 32 resynthesized library members spanning four macrocyclic ring sizes. This novel strategy is an ultra-high throughput method to profile reactivity in DNA-encoded libraries. Furthermore, it provides purity data about each individual encoded library member and measures permeability-relevant lipophilicity at an unprecedented scale. This chromatographic method can be applied to inform the design of future DELs or used to guide hit resynthesis efforts from DEL protein selection experiments.

Dedication

I dedicate this body of work first and foremost to my wife Leila and soon-to-be newborn daughter Sally. I love you two with all my heart and soul. Thank you for all the support, patience, and encouragement. I couldn't have done it without you! Leila, you are my rock, my best friend, and my love. Song of Solomon 5:16. I can't wait to start this next chapter of our lives!

Acknowledgements

To my advisor, Scott Lokey, who supported me every step of the way, afforded me a lot of freedom to grow independently as a scientist, and taught me an incredible amount about science, hard work, and life. Thank you, Scott.

To the Lokey Lab, thank you all for your support and love. First and foremost, I'd like to thank Alex Engstrom. It was a great pleasure to work on the chromatography paper with you. Your attitude, work ethic, and insights on complex, abstract topics will carry you far. We do great science but make even better music. Thank you to Akshar Lohith for teaching me everything I know about excel and keeping my mental health in check throughout the years. Thank you to Sean Carrick for keeping my spirits high in the home stretch. A very special thank you to Panpan Zhang, it was a great pleasure to work on DELs with you and figure out techniques and new chemistries together. I'd like to thank Justin Faris, who motivated and inspired me with his incredible work ethic and intellect. To Kevin Yang, who never failed to make me laugh but also taught me a lot about hard work and good science. Thank you to Victoria Klein, Colin Kelly, and

Tannia Lau for their friendship and tutelage early in my grad school career. Also, a special thank you to Jaru Taechalertpaisarn and Matt Naylor for all the training, teachings, guidance on manuscripts, and putting up with a lot of questions.

Thank you to my small army of undergraduate student researchers: Mari, Conor, Sarah Beth, and William. It was such a pleasure to work with you all and a joy to teach you the wonders of peptide synthesis.

To Jack Lee for his help and support in the NMR facility.

To Qiangli (Li) Zhang for her support and maintenance of the department LC/MS. You're the best Li!

To Satoshi Ono at Mitsubishi Tanabe Pharma for his McMD calculations and help interpreting the results.

To our awesome collaborators at Anagenex; Nicolas Tilmans, Meghan Lawler, LaShadric Grady, Joe Franklin, Neil Carlson, and others. Special thank you to Meghan who taught me a lot about DELs and make many significant contributions to the project.

Finally, thank you to the rest of my family. To Mom and Dad for always answering my calls, encouraging me, loving me, and always supporting me. I couldn't have asked for

better parents. To Hayden and Steff, for motivating me, challenging me, supporting me, and helping me keep my eyes on the prize. To Sean, Troy, Bobby, and Brandon, brothers, thanks for always being there for me. To my grandparents Larry and Gretchen, thank you so much for all the love and support. To Carl and Christy, my new family, thanks for welcoming me in with open arms and loving me like I'm one of your own.

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

Koch, Grant; Engstrom, Alex; Taechalertpaisarn, Jaru; Faris, Justin; Naylor, Matthew R.; Lokey, R. Scott. Chromatographic Determination of Lipophilic Permeability Efficiency Facilitates Rapid Identification of Passively Permeable Macrocyclic Peptide Scaffolds. *Accepted to J. Med. Chem.* **2024**.

This work was a co-first authorship, shared between Grant Koch and Alex Engstrom. Grant Koch and Alex Engstrom shared most of the responsibility and work on this project, including writing the manuscript, designing the project, performing all chromatography, performing shake-flask partition experiments, performing PAMPA, and doing computations. Grant Koch wrote much of the supplementary discussion and produced figures **S9, S11, S12,** and **Table 3**, as well as produced the code for processing ChamelogK data. Alex Engstrom performed the non-linear regression fits in Egor, the LPE minimization calculations, and the PCA, producing figures **S3-5, S7,**

S10, and **S12**. The manuscript was written in its entirety by Grant Koch and Alex Engstrom with revisions by Matt Naylor and Scott Lokey.

Abbreviations

Ado, 4,7,10,13-tetraoxapentadecanoic acid; ACN, acetonitrile; ALogP, atomistic calculated octanol/water partition coefficient; Aop, 3,6,9-trioxaundecanoic acid; b, y-intercept; Bp, base pair; bRo5, beyond rule-of-5; C18 or ODS, octadecylsilane; cLogP, calculated octanol/water partition coefficient; CMS, compact mass spectrometer; COMU, (1-Cyano-2-ethoxy-2- oxoethylidenaminooxy)dimethylaminomorpholino-carbenium hexafluorophosphate; cLPE, lipophilic permeability efficiency determined by chromatography; CTC, chlorotrityl chloride; CuAAC, copper-catalyzed azide-alkyne cycloaddition; CV, column volume; DCM, dichloromethane; DEL, DNA-encoded library; DIC, diisopropylcarbodiimide; DIPEA, diisopropylethylamine; DMF, N,N-dimethylformamide; DMSO, dimethylsulfoxide; DMTMM, 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride; DNA, deoxyribonucleic acid; DT, double truncation; EDC, 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide; ELSD, evaporative light scanning detector; EtOH, ethanol; FC, full compound; Fmoc, 9-fluorenylmethoxycarbonyl; HATU, 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate, *N*-[(Dimethylamino)-1*H*-1,2,3-triazolo-[4,5-*b*]pyridin-1-ylmethylene]-*N*-methylmethanaminium hexafluorophosphate *N*-oxide; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HFIP, 1,1,1,3,3,3-Hexafluoro-2-propanol; HOAt, 1- hydroxy-7- azabenzotriazole; HP, headpiece; HPLC, high-

performance liquid chromatography; HRMS, high resolution mass spectrometry; Hyp, hydroxyproline; LA, Lokey amine; LCMS, liquid chromatography mass spectrometry; LDEL, Lokey DNA-encoded library; Log D, distribution coefficient at pH 7.; Log $D_{dd/w}$, distribution coefficient at pH 7.4 with 1,9-decadiene/water; Log $ED_{dd/w}$, estimated distribution coefficient at pH 7.4 with 1,9-decadiene/water; Log K, logarithm of the capacity factor; Log P, partition coefficient; LPE, lipophilic permeability efficiency; m, slope; McMD, multi-canonical molecular dynamics; MDCK, Madin-Darby canine kidney; MeOH, methanol; MOPS, 3-(N-morpholino)propanesulfonic acid; MW, molecular weight; MWCO, molecular weight cut-off; NaDTC, sodium diethyldithiocarbamate; NMR, nuclear magnetic resonance; NGS, next generation sequencing; PAMPA, parallel artificial membrane permeability assay; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; Pe, effective permeability; PEG, polyethylene glycol; PROTAC, proteolysis targeting chimera; PRP, polystyrene reverse phase; PyBOP, benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate; qPCR, quantitative polymerase chain reaction; RFID, radio frequency identification; RMSD, root mean squared deviation; RNA, ribonucleic acid; RP-HPLC, reverse phase high performance liquid chromatography; ROX, carboxy-X-rhodamine; RT, retention time **or** room temperature; SMILES, simplified molecular input line entry system; SPPS, solid phase peptide synthesis; ST, single truncation; TAE, tris acetate EDTA; TBTA, Tris((1-benzyl-4-triazolyl)methyl)amine; TE, tris EDTA; TEAA, triethylammonium acetate; TES, triethylsilane; TFA, trifluoroacetic

acid; THF, tetrahydrofuran Tle, *tert*-leucine; TT, triple truncation; UPLC, ultra performance liquid chromatography; UV, ultraviolet.

Chapter 1: Chromatographic Determination of Lipophilic Permeability Efficiency Facilitates Rapid Identification of Passively Permeable Macrocyclic Peptide Scaffolds

1.1 Introduction

Membrane permeability is a prerequisite for accessing intracellular targets and is a desirable property in drug candidates intended for oral delivery. Most oral drugs are thought to passively diffuse across the cell membrane, a process driven by physicochemical properties such as molecular size and lipophilicity.[1, 2] In general, increasing lipophilicity leads to enhanced permeability; however, for larger molecules in particular, the lipophilicity required for achieving permeability tends to result in poor aqueous solubility.[3] Thus, medicinal chemists require accurate, high-throughput assessments of lipophilicity as one of the key properties to monitor in early drug discovery programs.

The pursuit of challenging, previously “undruggable” targets has expanded lead discovery efforts toward larger molecules that often violate Lipinski's Rule of 5 for predicting oral bioavailability. Macrocyclic peptides are prototypical molecules in this "beyond Rule of 5" (bRo5) chemical space, offering promising leads against challenging targets enabled by advancements in encoded library technologies such as mRNA display and DNA-encoded libraries.[4-6] Despite their potent binding affinities, these leads are often impermeable and require significant optimization due to the

inherent polarity of the peptide backbone and all but a few of the natural amino acid side chains.[7]

In contrast, many cyclic peptide natural products demonstrate passive cell permeability without exceeding lipophilicity limits by sequestering hydrogen bond donors, particularly backbone NH groups, in the membrane's low dielectric environment through steric occlusion or intramolecular hydrogen bonding.[8] This observation has led to a growing number of empirical[9, 10] and computational screening[11, 12] approaches for discovering novel backbone geometries, or scaffolds, that exploit intramolecular interactions to achieve membrane permeability. Since lipophilicity is conformation-dependent and can be difficult to predict computationally, especially for larger molecules, these approaches can benefit from experimental assessments of lipophilicity and membrane permeability that are robust, reproducible, and relatively high-throughput.

A compound's lipophilicity can be assessed empirically by its partition coefficient in a biphasic shake-flask system, such as *n*-octanol/water, expressed as the logarithm of the partition coefficient (Log $P_{oct/w}$) or the distribution coefficient at a given pH (Log $D^{7.4}_{oct/w}$). Although algorithms for predicting Log $P_{oct/w}$, such as cLogP and ALogP, have gained widespread use as calculated 2-dimensional lipophilicity descriptors, in terms of membrane permeability, the octanol-water system under-penalizes solvent-exposed hydrogen bond donors (HBD) and thus often overestimates permeability.[13] Instead, purely hydrocarbon solvents (e.g. 1,9-decadiene, hexadecane, and cyclohexane) have been noted for their ability to better capture the desolvation

penalty associated with exposed HBD.[14] Since this desolvation penalty can be highly dependent on a compound's low dielectric conformation, experimental hydrocarbon-water partition coefficients are preferable to algorithmically derived 2-dimensional descriptors. This is especially true for larger compounds and macrocycles in particular, which exhibit complex conformations.

Although shake-flask methods are the generally considered the gold standard measurement of lipophilicity and are amenable to some degree of miniaturization and cassette-sampling, chromatographic lipophilicity methods enable even greater throughput with simple automation.[15] Most established chromatographic tools for determining lipophilicity, such as ElogD[16] and ChromLogD,[17, 18] as well as more recently published methods such as AlphaLogD,[19] focus on correlating retention times of small molecule drugs to their shake-flask octanol-water partition coefficients. Other chromatographic systems have been developed to provide more biomimetic models of membrane transport, including the use of immobilized artificial membranes (IAM)[20] for the stationary phase and the use of multiple isocratic conditions to investigate environment-dependent conformational effects (Chamelogk).[21] Recent progress has been made to develop conditions to replicate hydrocarbon-derived shake flask lipophilicity, including a study by Caron et al. which demonstrated that an isocratic method using a column with a pure polystyrene-divinylbenzene matrix can produce capacity factors strongly correlated with algorithmically determined toluene-water partition coefficients (Log $P_{tol/w}$).[22] Additionally, Jensen, et al. estimated hydrocarbon-water partition coefficients using a

variety of columns and both isocratic and gradient techniques.[23] These studies primarily focused on classic Ro5 small molecule drugs which, in general, lack the conformational heterogeneity exhibited by macrocycles and other bRo5 compounds.

Previously, we observed strong correlations between 1,9-decadiene-derived shake-flask lipophilicities (Log $D_{dd/w}$) and passive permeability (artificial or cell-based) for a variety of bRo5 drug and peptide model systems,[24-26] highlighting the measurement's sensitivity to the strong permeability impact of solvent-exposed H-bond donors. Experimental Log $D_{dd/w}$ values, when combined with the calculated ALogP "bulk lipophilicity" descriptor, allowed derivation of a metric termed lipophilic permeability efficiency (LPE),[24] in which the permeability-relevant lipophilicity (Log $D_{dd/w}$) is compared to the solubility-relevant lipophilicity (ALogP) to capture the efficiency with which a particular compound utilizes its lipophilicity to achieve passive permeability. Increases in LPE for compounds with sufficient aqueous solubility yield opportunities for greater permeability in both PAMPA and cell-based permeability assays such as Caco-2 and MDCK. A LPE derivation from chromatographic measurements (cLPE) would yield improvements in convenience, throughput, and dynamic range over the shake-flask methodology, and this has already proven to be feasible in a recent study done by Watanabe, et al.[27] Here we expand upon the work of Watanabe, et al., by introducing a more general cLPE model which can be applied across multiple peptide scaffolds and other bRo5 compounds. Based on a dataset of cyclic peptides and other bRo5 compounds, we conclude that Log $D_{dd/w}$ and LPE can be predicted with high accuracy using chromatographic

methods and that the resulting values accurately predict membrane-relevant lipophilicity for a variety of bRo5 systems.

1.2 Non-linear Regression Model Between Capacity Factor and Partition Coefficients from Diverse Peptide Macrocycles

1.2.1 Initial investigation with heptapeptide library

To investigate the relationship between chromatographically determined lipophilicity and hydrocarbon shake-flask lipophilicity, we synthesized a library of diverse cyclic heptapeptides (**Figure 1a**, JT-7**)** in which both stereochemistry and side-chain lipophilicity were varied. The library was synthesized using split-and-pool solid phase peptide synthesis methods and was mass-encoded to facilitate multiplexed library screening (**Figure S1, Scheme S1-2**). JT-7 spans a wide ALogP range and contains 64 unique stereoisomeric scaffolds. Hydrocarbon shake-flask partition coefficients for each compound were determined between 1,9-decadiene and PBS (pH 7.4). From this library, we observed 188 compounds with high-fidelity Log $D_{dd/w}$ data as determined by the ability to accurately quantify their presence in both the aqueous and organic phases by LCMS.

Using JT-7, we first set out to determine the conditions that produced the strongest correlation between retention time and hydrocarbon shake-flask values. We determined chromatographic lipophilicities (LogK') using two isocratic and two gradient methods with different C18 stationary phases (**Figure S2**), including

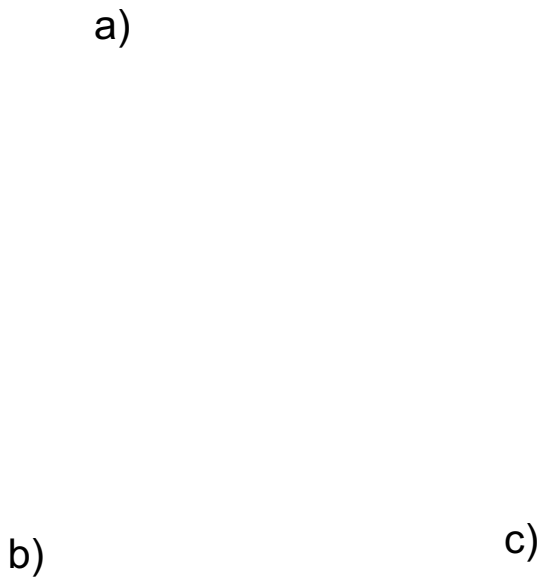


Figure 1: Non-linear regression between measured shake-flask lipophilicity and chromatographically determined lipophilicity. A) Structure and property data of cyclic peptide libraries. Circles in JF-10c and AF-10a-b represent either a peptoid or a N-methylated alpha amino acid. B) Relationship between shake-flask lipophilicity and chromatographic lipophilicity. Points are colored by cyclic peptide library (1a). C) Relationship between estimated shake-flask lipophilicity (Log $ED_{dd/w}$) and measured shake-flask lipophilicity (Log $D_{dd/w}$) of test set compounds.

traditional silica-backed columns as well as a polystyrene-backed, fully apolar C18 matrix (PRP-C18). Although silica-backed columns have been used extensively to estimate octanol-water partition coefficients, there is evidence to suggest that polystyrene-backed column matrices would afford measurements of lipophilicity that more closely resemble hydrocarbon-water partition coefficients due to the lack of silanol groups in the polymer matrix.[23, 28, 29]

We observed a strong correlation between retention time and Log $D_{dd/w}$ values for all methods and columns tested ($R^2 > 0.88$) (**Table S1**). Similar performance was achieved between the two silica-C18 columns and the PRP-C18 column, with the silica-C18 yielding marginally better correlations in some cases. Interestingly, at higher lipophilicities (Log $D_{dd/w} > 1$), we noted a deviation from linearity in the correlation for isocratic methods, despite maintaining a high R^2.

1.2.2 Expansion to full training set and nonlinear regression model

To increase the structural diversity of the training set, two cyclic decapeptide libraries (JF-10c, AF-10a-b) were included in addition to JT-7 (**Figure 1a**). AF-10a-b were previously synthesized and studied in our lab[30] whereas JF-10c was synthesized specifically for this study with traditional split-and-pool solid phase peptide synthesis methods (**Scheme 2**). These decapeptides feature unique cyclization patterns and are, on average, larger and more lipophilic than the heptapeptides in JT-7. Combined, the three libraries totaled 273 cyclic peptides, and 28 compounds (~10% from each library) were left out of the model to be used as a test set (see Supporting Information for full compound information). LogK' values were measured for all training and test set compounds using the PRP-C18 and one silica-C18 column under both isocratic and gradient conditions. Shake-flask Log $D_{dd/w}$ values for JF-10c and AF-10a-b were measured in triplicate. For the training set compounds, the relationships between LogK'60 or LogK'20-100 and Log $D_{dd/w}$ using either PRP-C18 (**Figure S3**) or silica-C18 (**Figure S4**) columns were fit to both exponential and linear functions. The exponential fit of the isocratic 60% method (**Figure 1b**) resulted in the best

predictions of Log $D_{dd/w}$ from LogK' values for the test set compounds (R^2 = 0.97 PRP, 0.96 silica). This model predicts 1,9-decadiene/water shake-flask partition coefficients via **Equation 1,**

(1) Log $ED_{dd/w}$ = 4.44- $8.01e^{-(\frac{\text{LogK}'60+0.186}{1.003})}$

where Log $ED_{dd/w}$ is the estimated distribution coefficient derived from chromatographically determined lipophilicity. This model accurately estimates Log $D_{dd/w}$ values for the test set (RMSD = 0.356) from a single chromatographic measurement (**Figure 1c**).

The PRP-C18 column was selected for further experiments in this study based on its affordability and relatively low back pressure, resulting in a longer column lifetime. Additionally, we pursued the isocratic 60% method because column equilibration is not required between runs, making this method more suitable for higher-throughput applications. Due to the low observed back-pressure, we investigated utilizing higher flow rates on PRP-C18 to reduce analysis times (**Figure S5**) and obtained an analogous exponential regression model with nearly identical predictive power of Log $D_{dd/w}$ (R^2 = 0.96). Under higher flow rate conditions, the most lipophilic peptides in this study eluted within 30 minutes. Pairing higher flow rates with the analysis of mass-unique mixtures facilitated rapid assessments of lipophilicity, with more diverse mixtures resulting in over 60 Log $ED_{dd/w}$ measurements per hour. As medicinal chemistry campaigns tend to be driven by pure compound investigations, we also compared our multiplexed capacity factor measurements on the long column to single compound injections on a short column

under identical isocratic conditions (**Figure S6**). We observed very strong correlations ($R^2 = 0.986$), suggesting that this model can be adapted for facile analysis of pure compounds. While variation of these chromatographic conditions (column length, flow rate, isocratic or gradient program) results in significantly different capacity factors, transformation to Log $ED_{dd/w}$ produces convergent values regardless of the methodology. This allows one to tailor the chromatographic method to the property of the analyte and produces lipophilicity values which are standardized between methods (comparable to pre-existing shake-flask values).

1.3 Lipophilic Permeability Efficiency Determined by Chromatography

1.3.1 Derivation of cLPE equation from model hexapeptides

We next sought to determine the ability of this chromatographic technique to produce meaningful LPE scores that are both reflective of known structural differences (e.g. absence or steric occlusion of exposed HBD) and increases in passive membrane permeability. Watanabe et al. recently reported a fast-gradient method for the chromatographic determination of LPE on a library of thioether cyclized peptide macrocycles.[27] A good correlation was observed between cLPE and passive permeability in pig kidney epithelial cells. The cLPE values reported in their work were obtained by transforming Log $D_{dd/w}$ in the original LPE equation to LogK' via a simple linear regression of 11 structurally similar peptides with the same extent of N-methylation in the backbone. Despite the reported strong correlations between Log $D_{dd/w}$ and LogK', it is still unclear if these methods penalize exposed hydrogen

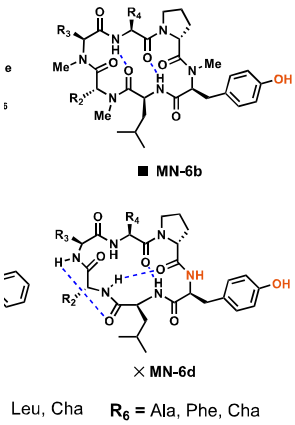


Figure 2: Lipophilic permeability efficiency determined by chromatography (cLPE) for bRo5 compounds. A) Structure of the cyclic hexapeptide library with exposed hydrogen bond donors colored in orange. B) Linear regressions of Log $ED_{dd/w}$ and ALogP for each peptide class. Open circles have no exposed HBD (class A), squares have one side-chain HBD (class B), stars have one backbone HBD (class C), and X's have two exposed HBDs (class D). Compounds are binned by ALogP (<3, blue; 3-4, green; >4 orange). C) Structure of PROTAC model compounds. D) cLPE plot comparing amide to ester matched pairs. Compounds are colored by scaffold (esters, green; amides, blue).

donors to the same degree and if LPE and cLPE scores are comparable across diverse peptide structures.

To investigate the ability of our chromatographic system to reproduce LPE scores derived from 1,9-decadiene/water shake-flask partition coefficients, we measured the LogK'60 of the hexapeptide libraries from which the original LPE model was developed (**Figure 2a**).[24] These four hexapeptide scaffolds (MN-6a-d) vary in their number of hydrogen bond donors that are exposed in low dielectric

media, with all other NH groups sequestered in IMHB or sterically occluded.[31, 32] Log $ED_{dd/w}$ and ALogP displayed strong linear correlations for all compounds within each scaffold set. Additionally, class separation, as reflected by difference in y-intercept, was observed between the scaffolds which generally reflected their previously determined LPE relationships (**Figure 2b**).

A minimization was performed as previously described[24] to calculate a common slope (1.01) between each of the hexapeptide classes and standardize the y-intercepts within classes. Next, the intercept of the linear regression equation for the class with the most exposed hydrogen bond donors (MN-6d, 2 HBD) was determined to produce **Equation 2**.

$$(2)\ cLPE = \text{Log } ED_{dd/w} - 1.01*ALogP + 3.98$$

In this equation, cLPE is defined as the lipophilic permeability efficiency determined chromatographically, with the lowest cLPE class from this dataset (MN-6d) fixed to zero. The constants for this equation were consistent between the isocratic and gradient methods tested (**Figure S7**).

The 1,9-decadiene/water partition coefficient estimations for this dataset were in agreement with measured shake-flask values for MN-6a (RMSD = 0.403) and MN-6c (RMSD = 0.522) but were less accurate for both MN-6b (RMSD = 1.153) and MN-6d (RMSD = 1.313). The Log $ED_{dd/w}$ values for these poorly predicted sub-libraries were consistently much higher than their respective Log $D_{dd/w}$ values. This result may suggest that side-chain phenols, or hydrogen bond donor moieties in general, may incur a much higher desolvation penalty in the shake-flask experiment

with 1,9-decadiene, whereas solid-phase partitioning into the C18 is more accommodated. It is unlikely that this difference could also be attributed to the relative acidity of the side-chain phenol (pKa = ~10) and the difference in buffer operating pH between shake-flask measurements (7.4) and chromatography (~2.5). In fact, we observed no measurable difference in retention time for these compounds when we performed the chromatography at pH 7 (see **Figure S8**). On the contrary, hydrogen bond donors in the peptide backbone (MRN-6a and MRN-6c) seem to have a similar effect on measured lipophilicities on both assays. Despite the noted differences, our chromatographic model produces cLPE values which are both representative of differences in exposed hydrogen bond donors and strongly predictive of passive permeability in PAMPA (see **Figure S9**).

1.3.2 cLPE of PROTAC model compounds

We next investigated a second class of bRo5 compounds, PROTACs, whose structure-permeability relationships have been investigated previously.[26] This compound set (VK-P) features two series of five lipophilicity-scanning model compounds, each bearing a VH032 warhead with either an amide or ester linkage, resulting in a difference of one hydrogen bond donor between the two scaffolds. We measured the LogK'60 of each compound and transformed these values to Log $ED_{dd/w}$ (**Figure 2c**). Encouragingly, for the matched pairs in this dataset, we obtained an average difference in cLPE (ΔcLPE) of 1.5 between the esters and the amides (**Figure 2d**), which is identical to our previously reported LPE difference derived from shake-flask values (ΔLPE = 1.5). Additionally, this ΔcLPE is very similar to the

cLPE difference observed for addition or removal of a backbone HBD in the MN6-a-d dataset (ΔcLPE = 1.7). The absolute cLPE values of the PROTACs are more reflective than the shake-flask derived LPE values of the predicted solvent exposure of two HBD (the -OH of the Hyp residue and the C-terminal -NH, with the -NH of the Tle being sterically occluded as previously shown). This discrepancy could be related to the protonation of the ionizable thiazole nitrogen in our chromatographic system (pKa = 4.06), however we observed no difference in retention time when using a buffered aqueous solvent (**Figure S8**).

1.3.3 Evaluation of cLPE scores across diverse peptide scaffolds

Accurate comparisons of cLPE across diverse systems require a consistent slope between Log $ED_{dd/w}$ and ALogP (m_{lipo}) for a given series of compounds with the same degree of HBD exposure (for the hexapeptide model system used to derive the original LPE equation, $m_{lipo} = 1.01$) (**Equation 2**). To determine whether this relationship holds true for larger macrocycles, we calculated the m_{lipo} for 12 distinct peptide scaffolds and the PROTACs for various chromatographic systems and shake-flask partition coefficients (see **Table S2**). Encouragingly, the m_{lipo} was consistent across all 12 peptide scaffolds and PROTACs for both chromatographic (s.d. = 0.204) and shake-flask (s.d. = 0.160) methods. Additionally, we calculated projected Log $ED_{dd/w}$ values for each of the liposcan scaffolds (except those from MN-6a-d libraries) utilizing the m_{lipo} from the cLPE equation (1.01) and the averaged y-intercepts of each scaffold (**Figure S10**). The projected Log $ED_{dd/w}$ values for all 8

scaffolds were in strong agreement with their measured values ($RMSD_{avg}$ = 0.178), indicating that cLPE can be accurately compared across diverse peptide scaffolds.

1.4 Relationship Between cLPE and Passive Permeability for Scaffold Diverse Thioether Decapeptides

1.4.1 Trends in PAMPA permeability

We next investigated whether chromatographically determined Log $ED_{dd/w}$ values could be used to predict passive cell permeability in a library of stereodiverse thioether decapeptides, JF-10a (**Figure 3a**).[33] This library consisted of 96 unique scaffolds with a liposcan element built into the peptoid position, resulting in seven stereoisomeric clusters. We were unable to obtain 1,9-decadiene/water shake-flask lipophilicity data on these compounds due to the propensity of the thioether to oxidize in the 1,9-decadiene layer, but LogK'60 data was collected on the entire library using the methods described above. As expected, we observed a wide range of lipophilicity across the seven sets of stereoisomers, identified by vertical groupings of compounds with identical ALogP values (**Figure 3b**). Gratifyingly, as Log $ED_{dd/w}$ increased within a set of stereoisomers, PAMPA permeability generally increased. This relationship becomes non-linear and even negative when extending the lipophilicity of the analytes outside the soluble limits of the permeability assay, which yields an inverse parabolic relationship. To investigate whether this trend was consistent across all stereoisomeric clusters, the relationship between PAMPA permeability and Log $ED_{dd/w}$ was examined in closer detail (**Figure S11**). All seven stereoisomeric clusters

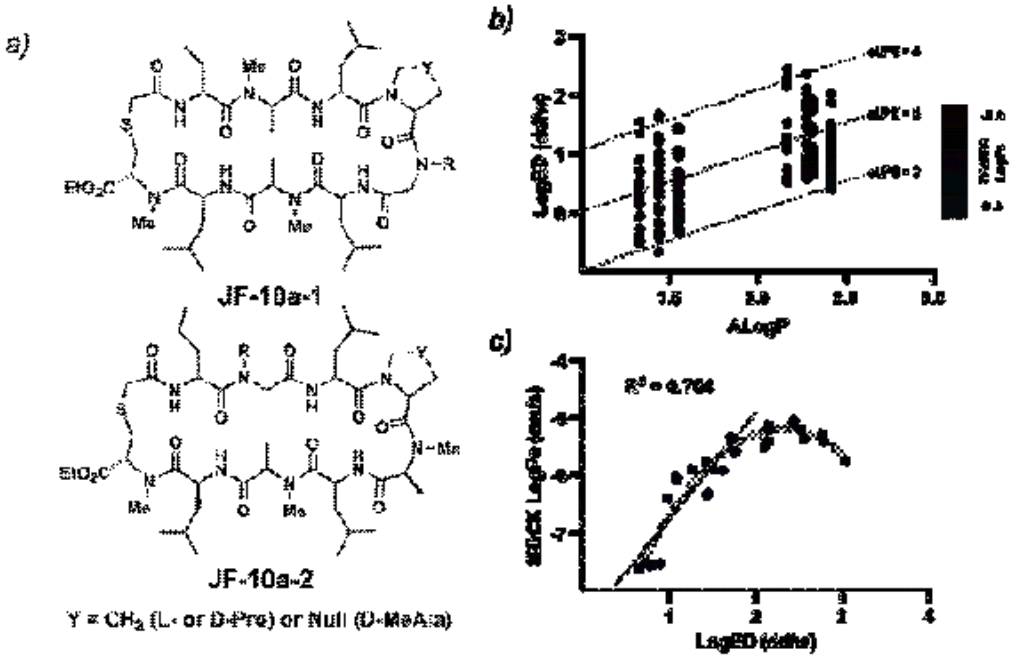


Figure 3: cLPE analysis of stereodiverse thioether-cyclized decapeptides from JF-10a. A) Structure of the library. B) Relationship between Log $ED_{dd/w}$ and ALogP. Compounds are colored by PAMPA LogPe in five evenly distributed bins of Lowest (-8.5 < LogPe < -7.8), Low (-7.8 < LogPe < -7.1), Medium (-7.1 < LogPe < -6.4), High (-6.4 < LogPe < -5.7), and Highest (-5.7 < LogPe < -5.1). C) Relationship between MDCK cell permeability and Log $ED_{dd/w}$. Compounds are colored by Log $ED_{dd/w}$ (<2, blue; >2, green). The linear regression for compounds with a Log $ED_{dd/w}$ <2 is displayed with a black line (R^2 = 0.754).

had a strong linear correlation between Log $ED_{dd/w}$ and PAMPA LogPe (average R^2 = 0.815) with an average ΔLogPe/ΔLogED of 1.13 (s.d. = 0.223). These results indicate that increases in Log $ED_{dd/w}$ produce similar increases in PAMPA permeability for compounds well-within the bounds of aqueous solubility.

Within this dataset, we identified several matched pair peptides which differ by only 1-2 stereocenters and exhibit a 50 to 100-fold increase in PAMPA permeability, which is reflected by a comparable increase in Log $ED_{dd/w}$ (see **Table S3**). The selected matched pairs highlight the ability for cLPE to identify significant

conformational changes which arise from seemingly small structural alterations, such as a stereochemical inversion of a single amino acid, which remains difficult to predict in silico.[11]

1.4.2 Trends in MDCK permeability

To investigate if these trends hold true for cell-based permeability assays, we compared our Log $ED_{dd/w}$ values to reported MDCK permeability values for selected compounds from the thioether library (**Figure 3c**). These data produced the expected reverse-parabolic relationship that has been previously observed in a variety of systems.[9, 24, 34] We observed an excellent correlation ($R^2 = 0.754$) between MDCK permeability and Log $ED_{dd/w}$ for this set of compounds within the observed soluble regime (Log $ED_{dd/w} < 2$). The ΔLogPe/ΔLogED for these data is 1.385, which is slightly higher than the average increase we observed in PAMPA. This dataset features a matched pair which differs by inversion of a single stereocenter yet shows a nearly 10-fold difference in MDCK permeability which is reflected by their Log $ED_{dd/w}$ values. These findings underscore the strong influence of cLPE values on cell-based permeability, indicating that early cLPE assessment could greatly assist in prioritizing lead compounds during drug discovery.

1.5 PAMPA and Log $ED_{dd/w}$ Trends Across Diverse Peptide Macrocycles

1.5.1 PAMPA and Log $ED_{dd/w}$ trends

For shake-flask measurements, we often observe a positive correlation between Log $D_{dd/w}$ and PAMPA permeability for compounds residing in the soluble

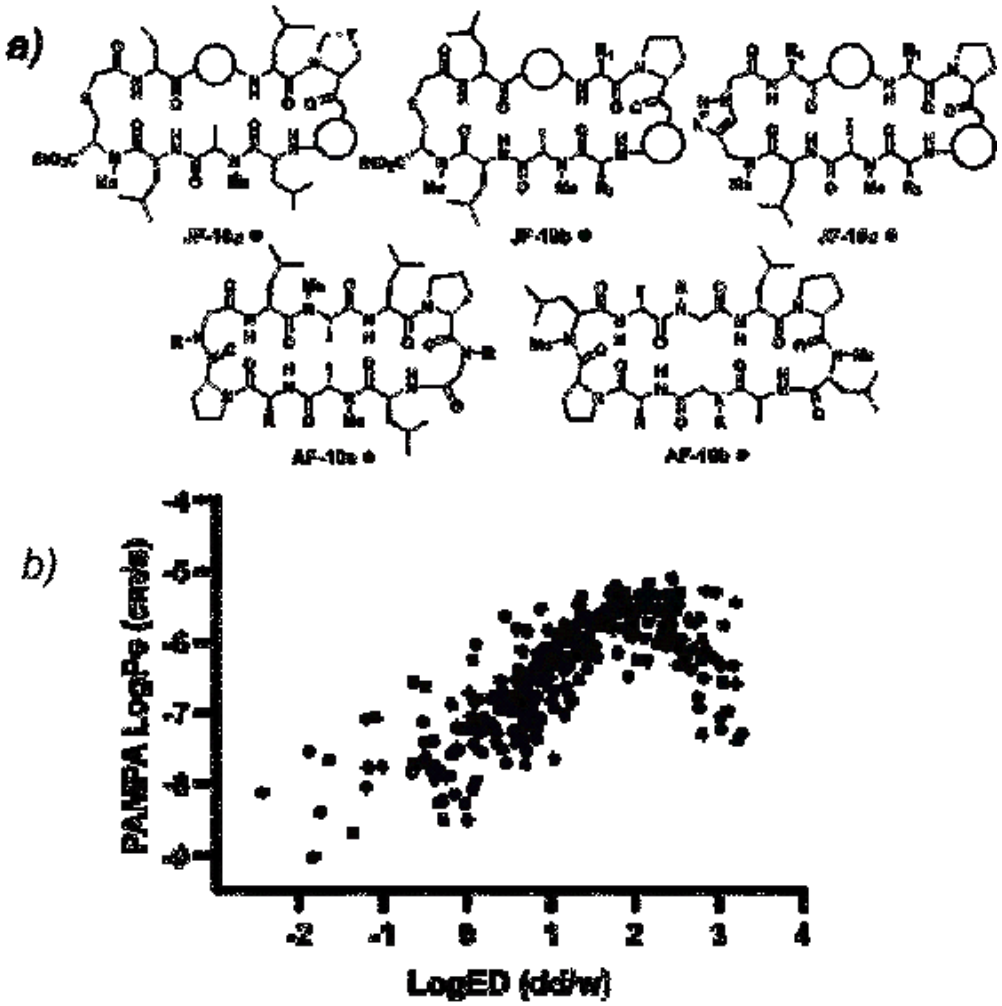


Figure 4: Relationship between PAMPA permeability and Log $ED_{dd/w}$ for scaffold-diverse decapeptides. A) Structures of decapeptide libraries. Circles in JF-10a-c represent either a peptoid or an N-methylated alpha amino acid. B) Relationship between PAMPA permeability and Log $ED_{dd/w}$. Compounds are colored in accordance with their library ID.

regime of lipophilicity. Furthermore, PAMPA and cell-based permeabilities have been shown to correlate well for a wide variety of bRo5 compounds ($R^2 > 0.70$).[2, 9, 25] To analyze the relationship between Log $ED_{dd/w}$ and aqueous solubility constraints in PAMPA, we compared the PAMPA LogPe and Log $ED_{dd/w}$ for 280 different cyclic decapeptides from four distinct peptide libraries featuring different cyclization

schemes (**Figure 4a**). These peptides were designed based on the decapeptide scaffold introduced by Fouche et al.[35] and altogether comprise 104 unique scaffolds. The parabolic relationship between LogPe and Log $ED_{dd/w}$ was consistent across most libraries analyzed, with the positive linear portion of the curve existing between Log $ED_{dd/w}$ -2 and 2. PAMPA permeability tends to level off and start decreasing around Log $ED_{dd/w}$ values of 2, presumably due to decreasing aqueous solubility or other recovery losses (**Figure 4b**). The triazole-containing compounds in the JF-10c library were more permeable than their thioether congeners in the JF-10a library at the same Log $ED_{dd/w}$ (and Log $D_{dd/w}$), despite good correlations within each compound class. The difference in PAMPA behavior between the two chemotypes, despite their comparable lipophilicities, suggests that correlations between Log $ED_{dd/w}$ and permeability may need to be determined independently for different chemical series in a medicinal chemistry program.

1.5.2 AF-10b chameleonicity effect on PAMPA

Library AF-10b maintains higher PAMPA permeability at high Log $ED_{dd/w}$ values compared to its congener, Library AF-10a, consistent with the known differences in their chameleonic behavior[30]. Whereas both compounds exhibit a fully hydrogen-bonded, cross-beta conformation in low-dielectric solvents, the location of a peptoid residue in one of the beta strands of Library AF-10b destabilizes the folded state of this scaffold in aqueous solution, giving rise to its chameleonic behavior. In contrast, Library AF-10a, with a peptoid in one of the beta turns, is fully folded in both polar and apolar media.

To further investigate the hypothesis that Libraries AF-10a and AF-10b differ with respect to their chameleonicity, we subjected them to an adapted form of a recently reported chromatographic technique for investigating chameleonicity, ChamelogK (**Figure S12**).[21] Notably, we utilized the PRP-C18 column from this study instead of the reported PLRP-S column. Library AF-10b has a higher average ChamelogK (1.081) than AF-10a (0.816), in support of our previous observations[30]. This additional flexibility may be the driving force behind AF-10b having a "right shifted" permeability curve, as the peak permeability for this scaffold is achieved at higher Log $ED_{dd/w}$ values. These results highlight the potential for chromatographic techniques to accurately assess both permeability-relevant lipophilicity and molecular flexibility, which is critical for rational design of large peptides macrocycles.[3]

Table 1: Summary of compound libraries used throughout this study

Library	Subclass	Cyclization	N	ALogP	MW
AF-10a	Macrocycle	Amide	18	0.39-4.72	987-1197
AF-10b	Macrocycle	Amide	18	0.39-4.72	987-1197
JF-10a	Macrocycle	Thioether	192	1.33-2.42	967-1028
JF-10b	Macrocycle	Thioether	36	1.19-4.26	965-1075
JF-10c	Macrocycle	Triazole	54	0.36-5.30	914-1104
JT-7	Macrocycle	Amide	188	0.22-5.04	645-1011
MN-6a-d	Macrocycle	Amide	33	0.66-6.18	578-792
VK-P	PROTAC	-	10	1.20-6.60	615-796

1.6 PCA of Peptide Libraries Show Similarity to Other Permeable Peptides

We wanted to determine the extent to which the compounds in this study (**Table 1, Figure S1**) represent the broader chemical space defined by published cyclic peptides of known membrane permeability. Therefore, we calculated seven physicochemical

descriptors (MW, LogP, TPSA, number of HBD, HBA, rotatable bonds and aromatic rings) and Morgan fingerprints for the compounds herein and compared them to those of the Cyclic Peptide Membrane Permeability Database (CycPeptMPD),[36] which consists of 7344 cyclic peptides from 47 publications. Based on a principal component analysis of these collected features, there was very good overlap between the feature space represented by our compounds and those in the database (**Figure S13**), suggesting that our method should be applicable across the larger class of previously studied cyclic peptides.

1.7 McMD Simulations Indicate Log $ED_{dd/w}$ Captures Permeability-Relevant Conformational Changes

To further validate that measured cLPE differences are due to differences in exposed polar surface area or exposure of hydrogen bond donors in low dielectric solvent, we conducted McMD simulations in explicit cyclohexane (a more convenient molecule for simulations with electronic character grossly similar to 1,9-decadiene) for two thioether decapeptide matched pairs (from data set JF-10a) which differ by one stereocenter. Both exhibit a cross-beta conformation and feature transannular hydrogen bonding typically observed in beta sheets. However, SB60 experiences a slight twist in the turn of the beta sheet due to the difference in stereochemistry about leucine 6 (**Figure 5,** 6d). This twist in the backbone appears to disrupt the transannular hydrogen bond between the Leu-NH (6d) and the carbonyl of the leucine across the ring (3a). The top three populated clusters of SB60 conformations are

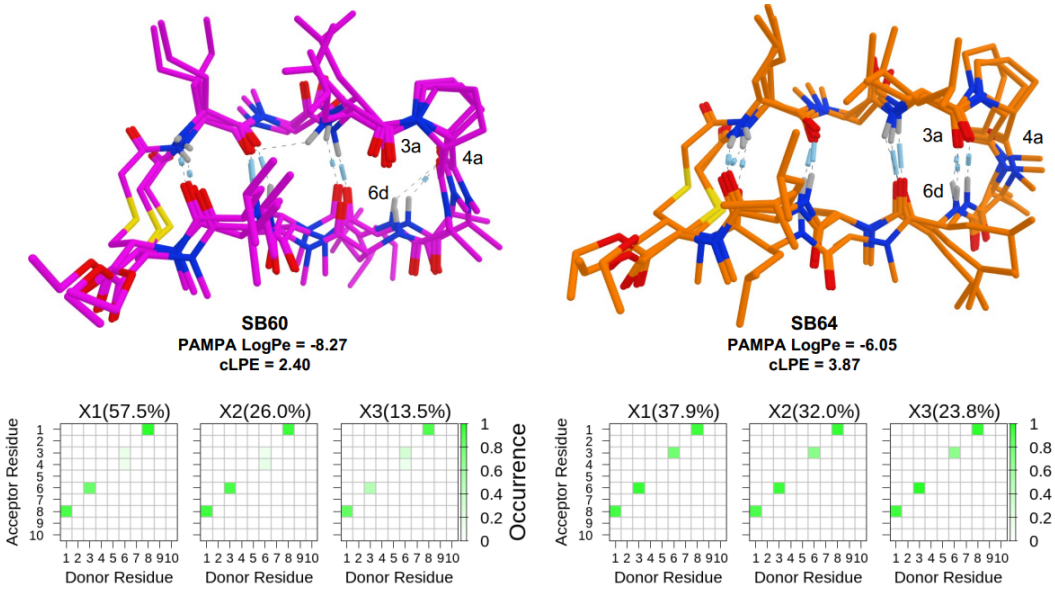


Figure 5: McMD structures of diastereotopic, thioether-cyclized matched pairs in cyclohexane. Each plot (X1-X3) represents a conformational cluster, where the intensity of the color reflects the relative occurrence of an IMHB across all structures within a cluster. The lowest RMSD frame from each X1-X3 conformational cluster is overlaid for SB60 (purple) and SB64 (orange). Dashed lines represent intramolecular hydrogen bonds which are also represented as green squares in the accompanying donor-acceptor plots.

instead predicted to contain weak, long-range hydrogen bonds from the Leu-NH (6d) to both the leucine (3a) and proline (4a) carbonyl. In contrast, across all conformational clusters, SB64 consistently exhibits a strong, short range 6d-3a hydrogen bond. SB64 likely manages to achieve a higher effective lipophilicity by adopting conformations which consistently occlude the 6d Leu-NH in a strong, stable transannular hydrogen bond. Careful control of hydrogen bond donors is essential to optimize bRo5 peptide macrocycles and is often modulated by N-methylating amide protons in the peptide backbone. However, these backbone NH protons can play significant roles in target binding and excessive N-methylation can lead to poor

solubility. Log $ED_{dd/w}$ values obtained under these chromatographic conditions can identify privileged scaffolds which manage to occlude exposed backbone NHs.

1.8 Conclusion

Despite the simplicity of standard shake-flask methods, these chromatographic methods increase ease, throughput, reproducibility and automatability while still permitting multiplexed analysis of complex mass-encoded compound mixtures. The Log $ED_{dd/w}$ values derived from our model are strongly predictive of measured Log $D_{dd/w}$ values for a diverse test set of peptide macrocycles, representative of the broader chemical space generally occupied by this class of compounds (**Figure S13**). Additionally, we have also shown that various chromatographic methods, including isocratic and gradient approaches, can generate predictive regression models for 1,9-decadiene/water partition coefficients of structurally varied peptide macrocycles. This enables comparisons of Log $ED_{dd/w}$ values across different chromatographic techniques. The ΔcLPE values produced from this model are comparable to ΔLPE values previously published which reflect known structural changes, chiefly exposure of hydrogen bond donors. As previously discussed[24], the above correlation and cLPE

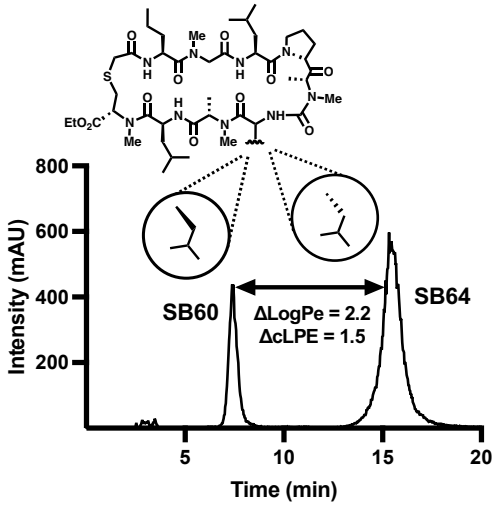


Figure 6: Chromatogram of a selected ion monitored spectra of diastereotopic, thioether-cyclized peptides, SB60 and SB64 (left and right, respectively).

derivation from well-understood macrocycle analytes places a convenient "flagpost" within bRo5 chemical space, but the approach herein is easily applied to other compound collections as needed. Acquiring cLPE scores early in drug discovery campaigns identifies scaffolds with opportunities for increased permeability to access intracellular targets or oral delivery.

Analyzing stereoisomer matched pair peptides from our dataset effectively demonstrates the value of cLPE. SB60 and SB64 are two closely related diastereomer scaffolds in which a single stereochemical inversion results in a significant difference in both chromatographic lipophilicity and PAMPA permeability (**Figure 6**). McMD simulations of these diastereomers predict disruption of the beta sheet in SB60, significantly reducing the strength of one of the four IMHBs. The difference in cLPE (1.47) between the scaffolds is analogous to previously observed differences for both the hexapeptide and PROTAC model systems, which feature scaffolds that differ by one exposed hydrogen bond donor.

In comparison to the shake-flask assay, chromatographic lipophilicity has significantly better precision between measurements and can more easily be reproduced between different users. Shake-flask partition coefficients are often plagued by sensitivity issues of the analyte in the less concentrated solvent (often times in the low nM range),[37] which restricts the dynamic range of this assay without extremely sensitive instrumentation (e.g. high resolution mass spectrometry) and can lead to low precision between measurements for compounds with very high or low values.

Although we obtained similar Log $ED_{dd/w}$ for isocratic and gradient methods and these values are reflective of low-dielectric conformational differences as predicted by McMD, more work must be done to elucidate the effect of the mobile phase composition on the distribution of cyclic peptide conformations while partitioning through the chromatographic system. Furthermore, the compounds in this study feature exclusively non-ionizable side chains, reflecting the composition of natural products known to be passively permeable.[8] More work is needed to evaluate the utility of our method towards the analysis of bRo5 molecules with ionizable side chains. Chromatographic lipophilicity and subsequent estimation of 1,9-decadiene/water shake-flask values represent a rapid and practical framework for assessment of lipophilicity, solubility, and potential for passive permeability in early-stage drug development for compounds beyond the traditional bounds of drug-likeness. In the next chapter, we will investigate the potential of applying this framework to DNA-encoded libraries.

1.9 Experimental Section

1.9.1 Library Information

Seven different compound libraries were used in this study (see **Figure S1**). Five of these seven libraries (JF-10a,[33] JF-10b,[33] AF-10a-b,[30] MN-6a-d,[24] TK-P[26]) have been featured in previous studies from our lab and were used directly from their stored DMSO stock solutions. MN-6a-d and TK-P were previously synthesized as individual compounds, and each compound was proven to be ≥ 95% pure by UV absorbance and

HNMR. AF-10a-b were initially synthesized as mixtures of 18 compounds and individual compounds were identified by UPLC-MS and selected-mass monitoring. A small subset of these libraries was resynthesized as pure compounds and their purities were confirmed ≥ 95% pure by UV absorbance and HNMR. JF-10a was initially synthesized in mixtures in 6 sets of 6 compounds. JF-10b was also initially synthesized in mixtures in 12 sets of 16 compounds. Individual compounds from both libraries were identified by UPLC-MS and selected-mass monitoring. From each of these two libraries (JF-10a and JF-10b), 12 compounds were resynthesized as pure constructs for passive cell permeability analysis by MDCK (**Figure 3**). The purity of each of these was confirmed to be ≥ 95% pure by UV absorbance and HNMR. The TK-P PROTACs were synthesized as pure compounds, and their purities were confirmed to be ≥ 95% pure by UV absorbance and HNMR.

Two of the seven libraries (JT-7 and JF-10c) were synthesized for this study. JT-7 was synthesized in mixtures in 16 sets of 100 compounds. JF-10c was synthesized in mixtures in 9 sets of 6 compounds. The compounds were purified as mixtures and used in all studies as mixtures. Individual compounds from these libraries were identified by UPLC-MS and selected mass monitoring. Selected mass monitoring for each compound in this study can be found in the appendix.

1.9.2 General Materials

Solvents and reagents, including dry solvents, were purchased from Fisher Scientific. Reagents were purchased at a high commercial quality (typically 97% or higher) and used without further purification, unless otherwise stated. Fmoc-protected amino

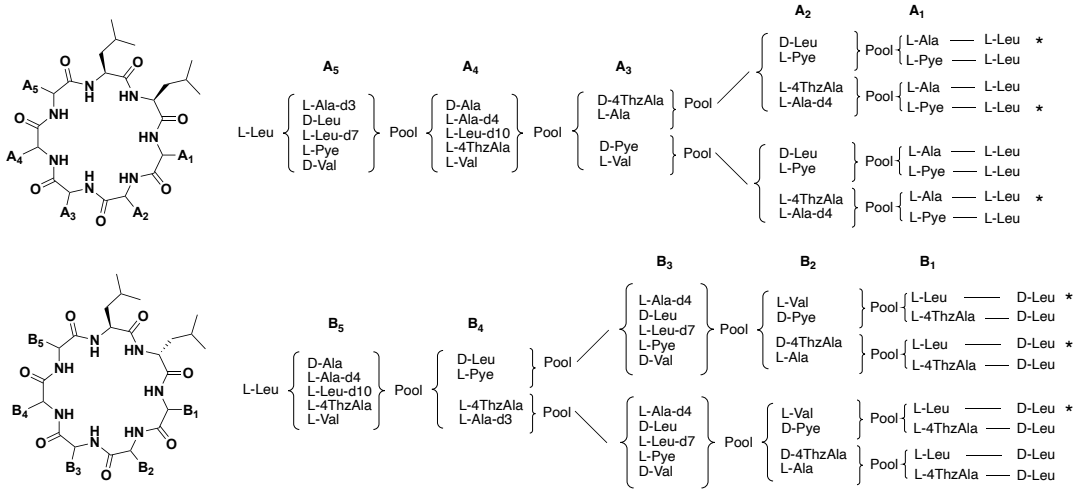
acids and coupling agents were purchased from Combi-Blocks, Oakwood, or Chem-Impex. Deuterated amino acids were purchased from Cambridge Isotope Laboratories. HPLC solvents were Optima grade, purchased from Fisher Scientific. 2-Chlorotrityl chloride polystyrene resin was purchased from Rapp-Polymere. Dodecane and soy lecithin used in PAMPA were purchased from Alfa Aesar. 1,9-Decadiene for shake-flask partition assay was purchased from Combi-Blocks. Columns were purchased from Hamilton® and Thermo Fischer Scientific. Three columns were used in this study, see Analytical Procedures section for details. The purities of synthesized compounds and all LogK chromatographic measurements were determined on a Thermo Scientific Ultimate 3000 UPLC system with a Thermo Scientific OrbiTrap VelosPro mass spectrometer. The solvent A mobile phase was composed of either 0.1% v/v formic acid in Milli-Q water (pH = 2.5) or 10 mM ammonium formate (pH = 7) in Milli-Q water. The solvent B mobile phase consisted of 0.1% formic acid in CH_3CN or pure CH_3CN.

1.9.3 Synthetic Methods

Overview of JT-7 synthesis

*The JT-7 series of compounds were made by Jaru Taechalertpaisarn

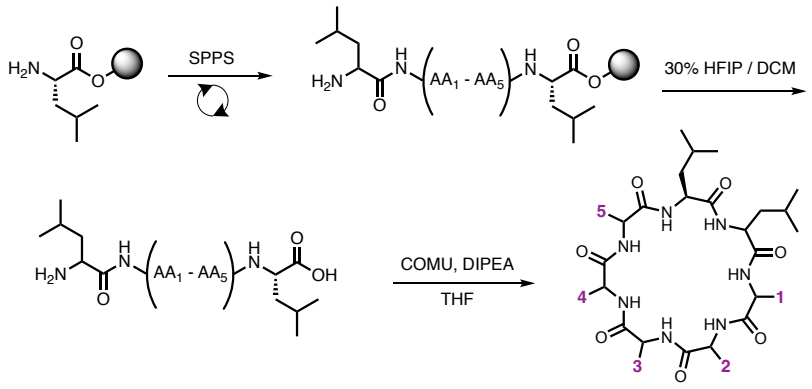
The JT-7 libraries were synthesized in a split-and-pool fashion (**Scheme 1**) resulting in 16 sub libraries with 100 compounds per sub library. For our analysis, we analyzed 6 select sub libraries (marked with *). After filtering out compounds due to mass-redundancy, we obtained high-quality Log $D_{dd/w}$ and LogK data for 188 unique cyclic heptapeptides.



Scheme 1: Detailed split-and-pool library design of JT-7 library

Manual peptide synthesis

Pre-loaded 2-Chlorotrityl resin (L-Leucine loaded – 0.65 mmol/g) was used (**Scheme 2**). Fmoc deprotections were carried out with 20% piperidine solution in DMF for 20 min twice. Couplings were performed using Fmoc-protected amino acids (4 eq), HATU (4 eq), and DIPEA (8 eq) in DMF for 1 hour at room temperature. Between coupling and deprotection steps, the resin was washed with DMF (5x), DCM (3x), and DMF (5x). Coupling efficiency was evaluated by negative ninhydrin test. Linear peptides were cleaved in a 30% HFIP-DCM solution and cyclized with COMU (2 eq) and DIPEA



Scheme 2: Synthetic scheme for JT-7 compounds

(4 eq) in THF relative to the mass of cleaved peptide and average molecular weight in each sublibrary. Cyclic peptides were purified by short C18 plug to remove excess coupling reagent and byproducts.

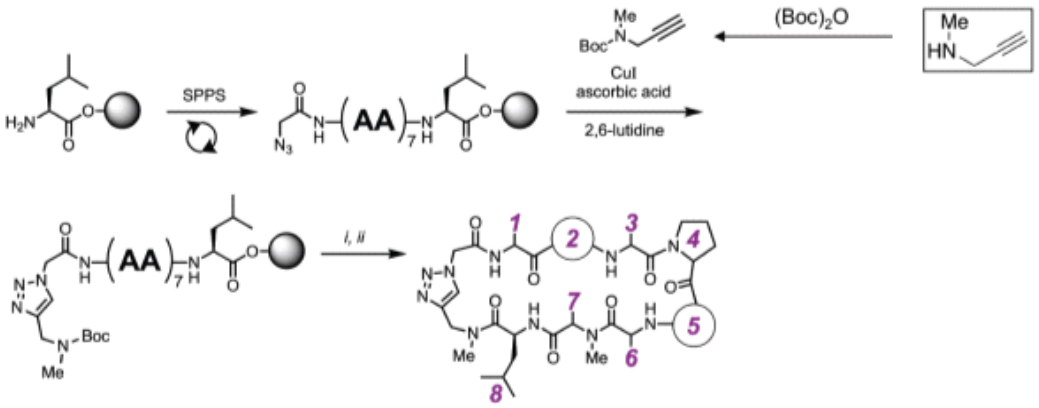
Overview of JF-10c synthesis

*The JF-10c series of compounds were synthesized by Justin Faris

The JF-10c library was synthesized in multiplex fashion using routine split-pool Fmoc-SPPS, installing the 1,4-disubstituted-1,2,3-triazole moiety as the final residue in the linear synthesis before simultaneous resin-cleavage and N-terminal Boc-deprotection (**Scheme 3**). Subsequently, the crude linear peptides were cyclized in-solution and purified to yield the peptide mixture. Synthesis was carried out to yield nine separate mixtures of six compounds each with a unique mass.

Synthesis of Boc-N-methylpropargylamine

To a 250 mL round-bottomed flask charged with a stir bar was added methanol (60 mL) and N-methylpropargylamine (3.05 mL, 36.2 mmol). While the solution was stirring, di-tert-butyl decarbonate (8.40 mL, 36.6 mmol, 1.01 eq) was added dropwise to the solution and the reaction stirred at RT overnight. The next day, the reaction was concentrated in vacuo to yield a dark brown liquid which was placed in the vacuum desiccator overnight and used without any further purification.



Scheme 3: Synthetic scheme for JF-10c compounds. (i) TFA (ii) COMU, DIPEA.

Azidoacetylation

Azidoacetic acid was coupled to the deprotected N-terminus of the on-resin linear peptide. In a separate vessel, azidoacetic acid (3 eq) was dissolved in enough DMF to eventually fully swell and submerge the resin. DIC (3 eq) was added to the solution and then allowed to rock on a linear shaker for 20 minutes. To the pre-swelled Fmoc-deprotected linear sequence on-resin was added this mixture and the suspension was allowed to rock at room temperature for 1-1.5 hr. Upon completion the resin was washed with DMF (3x), DCM (3x), and a final time with DMF. The reaction was monitored by HPLC-MS and repeated if necessary.

Synthesis of linear precursors

Commercially purchased pre-loaded 2-Chlorotrityl resin (Leucine loaded – 0.65 mmol/g) was used in 0.2 mmol portions. Fmoc deprotections were carried out with 2% DBU and 2% piperidine solution in DMF for 15 min. Couplings were performed using Fmoc-protected amino acids (4 eq), HATU (3.8 eq), HOAt (3.8 eq) and DIPEA

(6 eq) in DMF (0.1 M with respect to amino acid) for 1 hour at 50 °C or overnight at room temperature. Between coupling and deprotection steps, the resin was washed with DMF (3x), DCM (3x), and one final time with DMF. Coupling reactions were monitored by HPLC-MS and repeated until the starting material was no longer observed. Complete linear peptides were cleaved off resin using 10% TFA in DCM for 1 hour (2x) with a DCM wash equivalent to 2 resin volumes in between repetitions. Solvent was removed under a flow of nitrogen followed by dissolution in acetone and evaporation again under nitrogen. The crude linear product was stored under vacuum until cyclization.

Peptoid addition

Peptoid residues were added in-sequence to the Fmoc-deprotected amino acid by initial bromoacetylation followed by amine substitution. Bromoacetic acid (1 M) was dissolved in DMF in a separate vessel and followed by addition of DIC (0.5 M). This solution was allowed to rock on a linear shaker for 20 minutes. To the pre-swelled Fmoc-deprotected linear sequence on-resin was added this mixture in sufficient volume to cover the resin and the suspension was allowed to rock at room temperature for 45 min. Upon completion the resin was washed with DMF (3x), DCM (3x), and a final time with DMF. To the resin was then added a 1M solution of the corresponding amine in sieve-dried DMF and allowed to rock overnight at room temperature. The reaction was monitored by HPLC-MS and the amine addition repeated if necessary.

Peptide macrocyclization & purification

Unpurified linear peptides were dissolved in 10 mL of argon purged ACN with 4 eq of DIPEA and added dropwise to a solution of argon purged ACN containing 2 eq of COMU, for a total concentration of 0.002 M respective to the peptide. Reactions were stirred for 12-24 hours under an argon atmosphere until complete cyclization was achieved as monitored by HPLC-MS. The reaction was reduced in vacuo prior to purification which was carried out via trap-and-elute method utilizing Biotage Isolute C18 column.

Click (CuAAC) reaction conditions

In preparation for the CuAAC reaction, a 20% (v/v) solution of 2,6-lutidine in DMF was prepared fresh. Copper (I) iodide (5 eq) and ascorbic acid (15 eq) was added to a separate plastic vessel and the powders mixed thoroughly. Then the 20% (v/v) solution of 2,6-lutidine (20 mL for 0.2 mmol scale) was added to the powders and sonicated until complete dissolution (adding additional DMF if the solution was not dissolved after 10 min). This solution was then added to the pre-swelled resin followed by the addition of N-Boc-N-methylpropargylamine (2 eq) and the final suspension sparged briefly (< 1 min) with argon. Then the vessel was capped and shaken at RT overnight. Upon completion, the resin was drained and washed once with DMF. Next the resin was washed with a “click-wash” solution (5% sodium diethyldithiocarbamate & 5% DIPEA in DMF; 3-5x), DMF (3x), DCM (3x), and a final time with DMF. Completion of the reaction was verified by HPLC-MS.

1.9.4 Analytical Procedures

Table 2: List of LCMS methods used in this study.

Method	Column	Buffer A	Buffer B	Flow Rate (mL/min)	Gradient Program	Purpose
1	PRP Long	Water (+0.1% FA)	ACN (+0.1% FA)	0.2	60%B, 60-180 min[a]	LogED (dd/w) methodology (**Figure S2-5, Table S1**)
2	C18(1)					
3	C18(2)					
4	PRP Long			1	60%B, 30 min	
5	PRP Long			0.2	80%B, 60 min	
6	C18(1)					
7	C18(2)					
8	PRP Long			0.2	20-100%B, 60 min	
9	C18(1)					
10	C18(2)					
11	PRP Long			0.2	40-80%B, 60 min	
12	C18(1)					
13	C18(2)					
14	PRP Short	Water (+0.1% FA)	ACN (+0.1% FA)	0.2	60%B, 60-180 min[a]	Short column, pure compounds (**Figure S6**)
15	PRP Long	10 mM Ammonium Formate (pH 6.9)	ACN	0.2	60%B, 60-180 min[a]	Testing pH effect on ionizable groups (**Figure S8**)
16	PRP Long	Water (+0.1% FA)	ACN (+0.1% FA)	0.2	50%B, 60-180 min[a]	ChamelogK (**Figure S12**)
					70%B. 60-180 min[a]	
					100%B, 60-180 min[a]	

[a]These isocratic runs would vary in length, ranging between 60 minutes and 180 minutes depending on the lipophilicity of the analytes. Isocratic run times were extended as needed. All gradient programs were run for 60 minutes.

Information on Columns

PRP Long - Hamilton PRP-C18, 5 µm, 2.1x250 mm

PRP Short – Hamilton PRP-C18, 5 µm, 2.1x33 mm

C18(1) – Acclaim™ RSLC 120 C18, 2.2 µm, 120 Å, 2.1x250 mm

C18(2) – Acclaim™ Vanquish C18, 2.2 µm, 120 Å, 2.1x250 mm

Both C18 columns used in this study are standard end-capped ODS C18 from Thermo Fischer Scientific. Of note is the small particle size (2.2 µm) which may reduce

exposure to any backbone silanol groups which could contribute to more accurate comparison to decadiene-water partition coefficients.

The PRP column used is a polystyrene backed C18 column from Hamilton and is known to produce good correlations with hydrocarbon-water partition coefficients.[23]

After **Figure S2**, all silica C18 measurements were performed on **C18(2)**.

Table 3: List of all equations used for Log $ED_{dd/w}$ and cLPE conversions

LogED Linear Regression Equation (PRP, Iso 60)	LogED Exponential Regression Equation (PRP, Iso 60)	cLPE Equation (PRP, Iso 60, Exponential Fit)
$LogED(dd/w) = 3.804 * LogK'60 - 1.929$	$LogED(dd/w) = 4.4379 - 8.0061e^{-(\frac{LogK'60+0.186}{1.0028})}$	$cLPE = LogED(dd/w) - 1.01 * AlogP + 3.98$
LogED Linear Regression Equation (PRP, 20-100)	**LogED Exponential Regression Equation (PRP, 20-100)**	**LPE Equation (RDKit, PRP, Iso 60, Exponential Fi**
$LogED(dd/w) = 17.28 * LogK'20{-}100 - 20.82$	$LogED(dd/w) = 25.197 - 29.07e^{-(\frac{LogK'20-100-0.9936}{1.453})}$	$cLPE = LogED(dd/w) - 1.04 * AlogP + 2.64$
LogED Linear Regression Equation (SiC18, Iso 60)	**LogED Exponential Regression Equation (SiC18, Iso 60)**	**cLPE Equation (PRP, 20-100, Exponential Fit)**
$LogED(dd/w) = 5.118 * LogK'60 - 3.292$	$LogED(dd/w) = 4.0933 - 7.6848e^{-(\frac{LogK'60+00.027}{0.97429})}$	$cLPE = LogED(dd/w) - 1.02 * AlogP + 4.08$
LogED Linear Regression Equation (SiC18, 20-100)	**LogED Exponential Regression Equation (SiC18, 20-100)**	**LPE Equation (RDKit, PRP, 20-100, Exponential F**
$LogED(dd/w) = 21.81 * LogK'20{-}100 - 27.81$	$LogED(dd/w) = 995.07 - 998.69e^{-(\frac{LogK'20-100-1.099}{48.125})}$	$cLPE = LogED(dd/w) - 1.03 * AlogP + 2.07$
		Original LPE Equation
		$LPE = LogD(dd/w) - 1.06 * AlogP + 5.47$

Summarized Methodology to Calculate Log $ED_{dd/w}$ and cLPE

To calculate Log $ED_{dd/w}$:

1. Measure t_r and t_0 to generate capacity factors (K') and transform these to LogK' values by taking the base 10 logarithm of the capacity factor. t_0 can be measured by injecting an unretained analyte (e.g. citric acid) or observing baseline perturbation. In this study, t_r was measured at the center of each peak obtained by selected ion monitoring.
2. Transform LogK' to Log $ED_{dd/w}$ using appropriate equation from columns 1 or 2 in Table S5 based on the method used to obtain LogK'. This study provides regression equations for Methods 1, 3, 8, and 10 (see Table S4).

To calculate cLPE:

1. Calculate ALogP (available on RDKit as rdkit.Chem.Crippen.MolLogP)
2. Calculate cLPE with your measured Log $ED_{dd/w}$ values and calculated AlogP values by using the appropriate equation from column 3 in Table S5.

LogK Assay

DMSO stock solutions of each mixture of analytes were diluted in 1 mL of 40:60 H_2O/ACN (1 – 10 μM/compound, 0 – 5% v/v DMSO). 10 μL of this solution were injected to be analyzed by LC-MS. The LCMS solvents used were water (+0.1% formic acid) and ACN (+0.1% formic acid) or 10 mM ammonium formate (pH = 7) and ACN. Using a flowrate of 0.2 mL/min or 1.0 mL/min, and a column oven temperature of 35°C the retention time (rt) of each analyte was measured. The rt was transformed to a LogK value using the following equation:

$$LogK' = Log\left(\frac{rt - rt_0}{rt_0}\right)$$

where rt_0 = the observed solvent front of the column measured by injection of citric acid.

Samples were run as either single measurements, duplicate, or triplicate measurements with carbamazepine and cyclosporin A used as standards.

For the short column comparison injections, samples were prepared in an identical manner and the LCMS solvents were water (+0.1% formic acid) and ACN (+0.1% formic acid). A flow rate of 0.4 mL/min and a column oven temperature of 35°C was used.

Log $D_{dd/w}$ Shake-flask Partition Coefficient Assay

Before the assay, 1,9-decadiene was saturated with H_2O by vortexing an equal volume of PBS buffer (pH 7.4) and 1,9-decadiene allowing the emulsion to completely separate. 6 µL DMSO stock was added to either 1) a 1.5 mL centrifuge tube or 2) Simport T100 1.2 mL cluster tubes, to yield a final concentration per compound of 1 µM (unless otherwise specified). 600 µL of organic solvent and 600 µL of PBS (pH 7.4) were added (for a final concentration of 0.5% DMSO), sealed carefully, and agitated with vortexing (20 min) and sonication (20 min). The emulsions were then centrifuged for 10 min at 16,000 x g. 150 µL of each layer was carefully separated to minimize contamination from the pipette tip, especially when collecting the lower aqueous layer: 1) expel an air while passing through organic layer to prevent capillary wicking of organic solvent, 2) gently wipe off organic solvent adhered to tip after collecting and before dispensing. These layers were collected in either a 1) 1.5 mL centrifuge tube or 2) 96-well 300 µL conical-bottom plate and evaporated overnight in a Genevac centrifugal evaporator (60 °C). 150 µL of DMSO was reintroduced into each sample, mixed, sealed, and sonicated for 30 min. After a final centrifugation for 10 min at 700 x g, the top 150 µL was collected

and quantified via UPLC-MS in 20 µL injections (see above). Data was collected as single point. Partitioning standards were often included in these experiments at 1 µM: carbamazepine (Log $D_{dd/w}$ = -0.4) to help identify any experimental errors.

Parallel Artificial Membrane Permeability Assay (PAMPA)

PAMPA was used to determine the passive permeability of these compounds with modifications and calculations as described in Naylor *et al*.[24] and here. Briefly, pure compound stocks (150 µL, 1 µM, 5% DMSO in phosphate-buffer saline (PBS) pH 7.4) were added to a 96-well donor plate with 0.45 µm hydrophobic Immobilon-P membranes (Millipore MAIPNTR10) pre-soaked with a 1% solution of lecithin in n-dodecane. The donor plate was loaded into a 96-well Teflon acceptor plate (Millipore MSSACCEPTOR) with 300 µL of PBS with 5% DMSO. After a ~15 h, the acceptor and donor plates were separated, and 50 µL was collect from each well. Samples were diluted with 50 µL methanol and analyzed by LC/MS as described above. Samples were run in triplicate with 1 µM propranolol as a control. Exact time in seconds was used in the calculations. Error was calculated as standard deviation.

UPLC-MS analysis

All UPLC-MS analysis was performed on a Thermo Scientific Ultimate 3000 UPLC system equipped with a Thermo Scientific OrbiTrap Velos Pro mass spectrometer. PAMPA mixture analysis was performed with a Thermo Hypersil GOLD C18 (30

mm x 2.1 mm, 1.9 µm particle size) column. Purity analysis was performed on a Thermo Scientific Acclaim RSLC 120 C18 (2.1 mm x 250 mm, 2.2 µm, 120 Å) column. All samples were run eluting in water/acetonitrile with 0.1% formic acid. Purity assessment was quantified by UV absorbance at 200 nm.

Log $D_{dd/w}$ and PAMPA analysis conditions:

Flow rate: 1.0 mL/min

Temperature: 50 ºC

Gradient: 0-0.5min (05%ACN), 0.5-0.6min (05-95%ACN), 0.6-3.5min (95%ACN), 3.5-3.6min (95-05%ACN), 3.6 min-5min (05%ACN)

Source ionization: off

MS mode: FTMS

Resolution: 30000

Polarity: (+)

Data type: centroid

MS^n: Collision-induced dissociation

Normalized collision energy: 35.0

Isolation width: 1.0 m/z

Activation Q: 0.250

Activation time: 10.00 ms

UV: off

1.9.5 Computational Procedures

McMD - System Setups

The initial conformer was built using the Molecular Operating Environment (MOE)[38] protein builder, and LowModeMD implemented in MOE was used to search for the most stable conformer in vacuo. PACKMOL[39] was used to solvate the system, where the numbers of cyclohexane molecules were 300. The AMBER ff03 force field was used for amino acids[40] and ForceField_NCAA was used for N-methylated alanine, Sar, and Nva.[41] The angle parameter c-c3-ss was adopted for MeCys residue from the general amber force field (GAFF).[42] For the atomic charges of thioether residue and tail CO_2Et, RESP ESP charges from R.E.D. version III.52[43] were used. The initially solvated system was

minimized by steepest descent, followed by 50,000 steps of NVT calculation. The system was then equilibrated by NPT calculation at atmospheric pressure by applying the Berendsen barostat for 500,000 steps. A positional restraint was imposed on Cα atoms during equilibration. The resultant system was used as an initial structure for the following simulations.

McMD - Simulation details

*McMD simulations were performed by our collaborator Satoshi Ono at Mitsubishi Tanabe Pharma Corporation.

Virtual-system-coupled Trivial Trajectory Parallelization of Multicanonical Molecular Dynamics (TTP-V-McMD)[44, 45] simulations were adopted to effectively sample conformations of cyclic peptides. A total of 288 pre-TTP-V-McMD runs were initiated with random velocities for each atom at T = 300 K. Then the systems were heated to T = 1525 K with 5,000 steps, followed by 495,000 steps at T = 1525 K to randomize the initial structures. Flat potential energy distributions between T=280 K and 1525 K were obtained by iterating TTP-V- McMD simulations for ten times with eight virtual states. The cut-off distance for Coulombic and van der Waals interactions was 1.0 nm, and PME was used to calculate the long-range electrostatic interaction. The NVT ensemble was used for all TTP-V-McMD simulations using the velocity rescaling method (Bussi thermostat).[46] The LINCS algorithm was used to constrain the bonds with a hydrogen atom, enabling a time step of 2.0 fs. For each solvent, 1 × 107 steps × 288 production runs (aggregating 5.76 μs) were performed. The structure and potential energy were stored every 2 ps. The virtual states were exchanged in every 5,000 steps. All 2.88 million conformers were considered to determine the free energy landscape by a potential of mean force (PMF) calculation; $W = -kBT\ln\rho$, where kB is the Boltzmann constant, T = 300 K and ρ is the density of state. 5,000 Conformers at T = 300 K were used for the cluster analysis. An in-house implemented TTP-V-McMD using GROMACS version 5.1.4 was used for the simulations.

Chapter 2: Large Scale Reaction Profiling and Determination of Permeability-Relevant Lipophilicity in DNA-Encoded Libraries

2.1 Introduction

Combinatorial encoded library technologies have recently emerged as a powerhouse strategy for the discovery of potent ligands against challenging protein targets.[47, 48] These techniques accelerate the discovery of novel ligands for challenging protein targets by coupling combinatorial drug libraries containing unparalleled chemical diversity with simple affinity-based selections on immobilized protein targets. Display technologies, such as mRNA display[49, 50] and phage display,[51, 52] rely on translation-based machinery to produce libraries of peptides or antibodies. Although these display technologies traditionally produce larger peptides with excessive polarity, recent advances in mRNA display have significantly improved the drug-likeness of these libraries.[53] Unlike display technologies, DNA-encoded libraries (DELs) do not rely on enzymatic translation are instead produced through iterative rounds of DNA-encoding and chemical synthesis performed in a split-and-pool manner.[54, 55] Non-reliance on enzymatic machinery and charged tRNAs vastly increases the diversity of available building blocks in DEL. Additionally, the chemical transformations available to DELs expand far beyond simple carboxylic acid activation and amide bond formation,[56-58] with an ever-increasing toolbox of reactions[59-61] available to create combinatorial encoded libraries of non-peptidic scaffolds.

Although DNA-encoded libraries are currently widely pursued in both academia and the pharmaceutical industry,[62] there still remain challenges in effectively translating hits from encoded library selections to lead compounds. A significant issue plaguing DELs is the prevalence of inefficient chemical transformations during synthesis,[63] which naturally arise due to the synthesis constraints in aqueous media and the vast diversity of building block combinations used throughout library synthesis. As synthetic truncations compound throughout each round of synthesis, properly synthesized constructs are typically poorly represented in the final encoded library mixture. For a four-cycle library with an average 70% reaction conversion rate, fully encoded compounds would constitute a meager 24% of the total material, and attempts to quantify this suggest average yields may be much lower.[64] To combat this issue, DEL chemists regularly perform extensive building block rehearsals using a representative library constructs to gauge conversion rates and judge which building blocks should be included in a synthesis. Recently, several promising strategies have been reported in an effort to improve the purity of DNA-encoded libraries by eliminating synthetic truncations.[65, 66] There have also been reported strategies for DEL reaction monitoring, such as a solid-phase technique to gauge the cyclization efficiency of scaffold-diverse DNA-encoded thioether macrocycles,[67] and strategies for identifying problematic chemistry which may affect the amplification efficiency of the DNA tag.[68] However, despite general improvements to the synthesis of DELs,[69] methodology to improve DEL purity, and efforts to analyze purification outputs to inform resynthesis efforts,[70] there still

remains no direct method to assess the quality of synthesized encoded library members wholesale.

Another significant obstacle found in DELs and other encoded library techniques is the propensity for affinity selections to enrich library members bearing polar moieties, which results in enriched hits with poor pharmacokinetic properties like low passive cell permeability.[5, 71] This selection bias towards highly polar library members is especially deleterious for libraries of larger encoded constructs well beyond the "rule of 5", where a fine-tuning of lipophilicity is critical. Although algorithmic predictions of lipophilicity are readily available, these 2D atomistic calculations fail to capture complex conformational effects exhibited by larger molecules like peptide macrocycles[24] and therefore an empirical assessment of lipophilicity is highly desirable. Despite over a decade of harnessing the encoding power of DELs to perform affinity selections, there have been no reports of measuring the physical properties of the pendent encoded molecule across an entire library. A chromatographic assay represents a significant potential strategy which could simultaneously determine compound purity and lipophilicity for individual members in a DNA-encoded library.

Although no such chromatographic assay has been reported to date, there is precedent for efficient separation of oligonucleotides based on chemical modifications which affect analyte polarity.[72] This phenomenon is well known to DNA-encoded library chemists, as coupling DNA-bound substrates with sufficiently hydrophobic building blocks (e.g. Fmoc-amino acids) produces large retention time

increases under the conditions typically employed during HPLC purifications and LCMS of DNA substrates and DELs.[69, 73] To achieve separation based on the warhead instead of the DNA, ion-pair reverse-phase chromatography conditions are employed. These conditions typically a buffer consisting of a tertiary amine bearing aliphatic groups (e.g. triethylammonium acetate, tributylammonium acetate). This buffer adsorbs into the stationary phase of the column and allows for electrostatic interactions between the C18 column and the polyanionic oligonucleotide tag while still allowing for non-electrostatic interactions between alkyl chains of the stationary phase and the molecule pendent to the DNA. This ion-paired system is widely used in oligonucleotide purifications and has been demonstrated to be extremely selective and efficient at separation of minor structural changes.[74] These conditions are typically used to purify truncations and incorrect chemical modifications made during the synthesis of single stranded oligonucleotides.[75] As DNA-encoded libraries typically feature double stranded DNA tags all at the same length, these ion-paired reverse-phase chromatography conditions may be applied to selectively partition DNA-encoded libraries based on the lipophilicity of their pendent encoded compounds. Property-based chromatographic partitioning and subsequent sequencing of a DNA-encoded library would allow for a direct assessment of individual compound purity and lipophilicity, which could vastly improve hit resynthesis efforts.

Herein, we provide a chromatographic approach to simultaneously analyze the purity and lipophilicity of individual molecules in a DNA-encoded library. This approach utilizes simple ion-paired reverse phase chromatography and an apolar

column media to chromatographically separate DELs based on the lipophilicity of the warhead. The validity of this strategy was first tested on DNA-conjugates of both simple small molecules and peptide macrocycles. Afterwards, we synthesized a proof-of-concept library of 60,000 peptide macrocycles and partitioned the library chromatographically. We collected individual library fractions, encoded them with PCR, and sequenced each fraction altogether. The deduplicated sequencing counts in each fraction were then used to produce functional chromatograms of each encoded library member. This strategy simultaneously provides purity data on individual library members and a measure of permeability-relevant lipophilicity via chromatographic capacity factors. We confirmed the on-DNA purities and lipophilicities by resynthesizing thirty library members. Within each ring size, we observed excellent correlations between on-DNA library lipophilicity and off-DNA passive permeability in the Parallel Artificial Membrane Permeability Assay (PAMPA).

This chromatographic assay is the first report of a technique to assess both the purity and lipophilicity of individual encoded library members. This strategy can be used to improve hit translation efforts as the individual compound chromatograms reveal compound purity through both identification of synthetic truncations and assessment of peptide macrocyclization. Additionally, this strategy has the potential to significantly optimize hit-to-lead efforts as it provides a direct measurement of compound lipophilicity which has strong bearings on passive permeability.

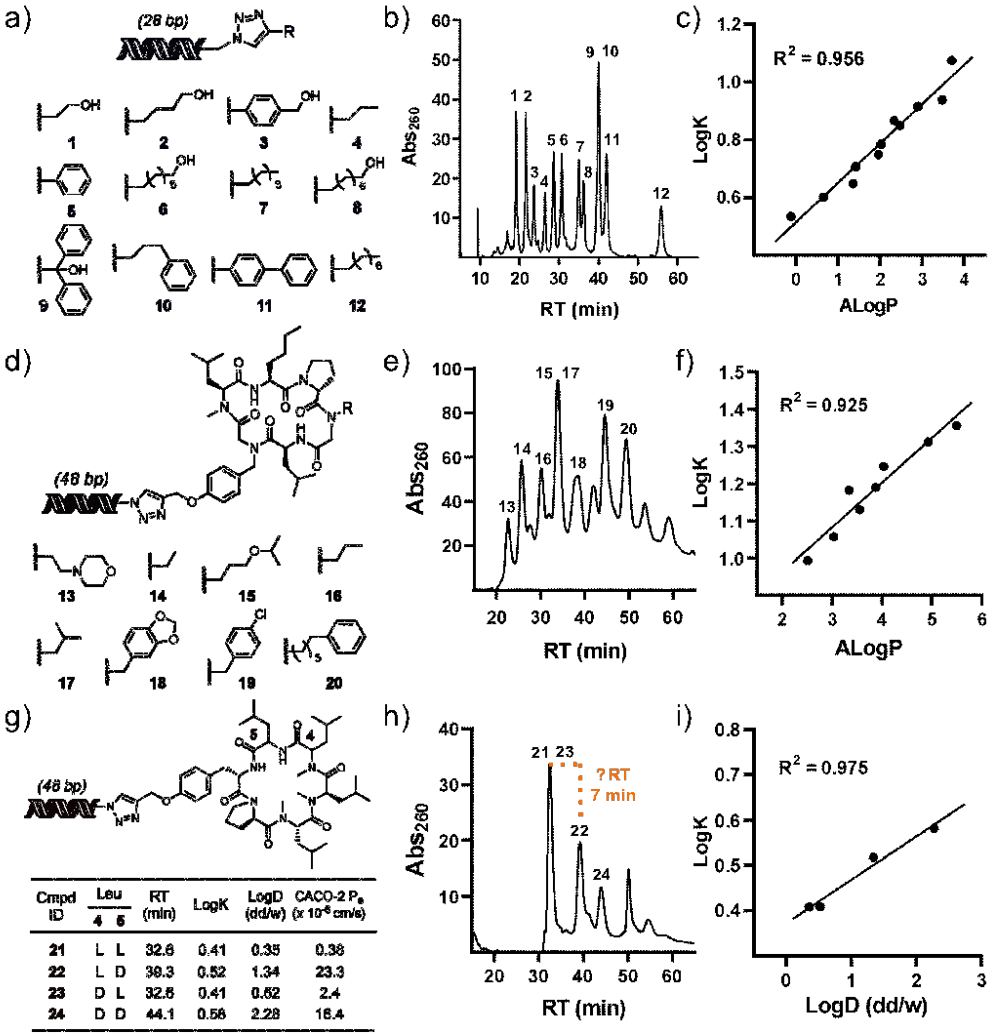
2.2 Method Optimization with Model Compounds

2.2.1 Small molecule dataset

To determine if chromatographic lipophilicity could be accurately measured for DNA-bound substrates, we first synthesized a series of simple, small molecule-DNA conjugates on 28 bp DNA. DNA headpiece was first acylated with azido-PEG3-acid and then ligated to a 20 base pair tag to extend the duplex. Next, twelve alkynes spanning a wide range of lipophilicities were coupled to the DNA with the standard copper-mediated Huisgen cyclization to create **1-12** (**Figure 7a**). These test substrates were subject to a wide variety of chromatographic conditions to assess if their lipophilicities could be accurately measured despite being bound to a DNA tag which is over 100 times the size of the small molecule warhead. Utilizing a Hamilton PRP-C18 column (2.1 x 250 mm) and the ion-pairing buffer conditions typically employed in purifications of DNA and DNA encoded libraries (mobile phase A; 50 mM TEAA, pH = 7, mobile phase B; ACN) produced great peak shapes and an excellent correlation to algorithmically predicted lipophilicity (AlogP) (**Figure 7b** and **7c**). Remarkably, small changes in the small molecule warheads yielded large differences in on-DNA retention time, such as the addition of a single methylene (**6** and **8**) resulting in a 6-minute retention time difference.

2.2.2 AF-Hexapeptomer Dataset

Encouraged by the preliminary results of the small molecule data set, we next investigated a series of cyclic peptomers previously studied in our lab[76] by



Cmpd ID	Leu 4	Leu 5	RT (min)	LogK	LogD (dd/w)	CACO-2 P_e (x 10^{-6} cm/s)
21	L	L	32.6	0.41	0.35	0.38
22	L	D	39.3	0.52	1.34	23.3
23	D	L	32.5	0.41	0.62	2.4
24	D	D	44.1	0.58	2.28	16.4

Figure 7: Chromatographic lipophilicity of various DNA-bound substrates. A) Small molecule alkynes clicked onto 28 bp azido-AOP-HP. B) Cyclic hexapeptomer liposcan clicked onto 48 bp azido-AOP-HP via propargyl tyrosine. C) Cyclic hexapeptide diastereomers clicked onto 48 bp azido-AOP-HP. On-DNA lipophilicities strongly correlate with both off-DNA partition coefficients with 1,9-decadiene and water and CACO-2 cell permeability.

conjugating them to even longer duplex DNA (**Figure 7d**). The eight hexapeptides are all composed of the same peptide scaffold and differ by substitution at a single peptoid position. We prepared azido-PEG3-headpiece with a 48 base pair tag by ligating an additional 20 base pair tag onto our previous construct. We then

synthesized the peptides with traditional solid-phase synthesis methods and installed an alkyne handle by propargylating the tyrosine phenol and clicked each peptide onto the DNA with similar copper-click conditions. All eight peptide DNA conjugates were subject to the previously established chromatographic conditions. We observed an excellent correlation between on-DNA measured chromatographic lipophilicity and algorithmically predicted lipophilicity (**Figure 7e and 7f**). Similar chromatographic separation was observed between close matched-pairs for this class of compounds, however the observed peaks had slightly worse resolution, possibly due to the increased size of the duplex DNA. For simple modulations of molecular lipophilicity, (e.g. the small molecule and cyclic peptomer datasets) 2D calculations of molecular lipophilicity (AlogP) are strongly predictive of empirical lipophilicity.[24] However, the lipophilicity of peptide macrocycles which exhibit complex confirmational dynamics is impossible to accurately predict with simple 2D calculations and would require significant computational modeling.[11, 12]

2.2.3 WH-Diastereomer Dataset

To investigate if differences in lipophilicity due to three dimensional conformational effects can be accurately measured for DNA bound peptides, we synthesized four cyclic peptide diastereomers bound to 48 bp DNA. These peptides have been previously studied in our lab[32] and exhibit a remarkable difference in lipophilicity and passive cell permeability dependent on the stereochemistry on a specific leucine side-chain (**Figure 7g**). When Leucine 5 on these scaffolds bears a D stereochemistry, the off-DNA shake-flask partition coefficients with 1,9-

decadiene/water increases dramatically. This increase in lipophilicity is accompanied by a marked increase in passive cell permeability in CACO-2 cells.

The cyclic hexapeptides were synthesized, cyclized, and a click handle was installed similarly by propargylating the tyrosine. Each alkyne-bearing peptide was clicked onto the 48 base pair azido-PEG3-headpiece and their on-DNA lipophilicities were measured as described previously. The DNA-bound diastereomer peptides display a large difference in chromatographic lipophilicity (**Figure 7h**) which corresponds to both the off-DNA shake-flask lipophilicities (**Figure 7i**) and CACO-2 passive cell permeabilities of their parent compounds (bearing an -OH on the tyrosine). These results strongly indicate that the lipophilicity of DNA bound substrates can be accurately measured with chromatography and these measurements are sensitive to the conformation of the peptide warhead.

2.3 Scaffold Diverse Macrocyclic Peptide DNA-Encoded Library Design

Inspired by these results, we synthesized a DNA encoded library of scaffold diverse cyclic peptides (**Figure 8**) to assess if the lipophilicity of DNA-bound compounds can still be measured in the context of a full encoded library. The DNA encoded library was synthesized with 32 building blocks in cycles 1 and 2 and 58 building blocks in cycle 3, resulting in a 59,392-member library. The building blocks were selected to prioritize scaffold diversity and span a wide range of lipophilicity. Building blocks consisted of either Fmoc amino acids (BB1-8, BB13-18), β-homoamino acids (BB9-10), Fmoc-N-alkyl glycines (BB11-12), or tripeptides (BB19-58). All stereochemical variants of each amino acid, N-methyl amino acid, β-

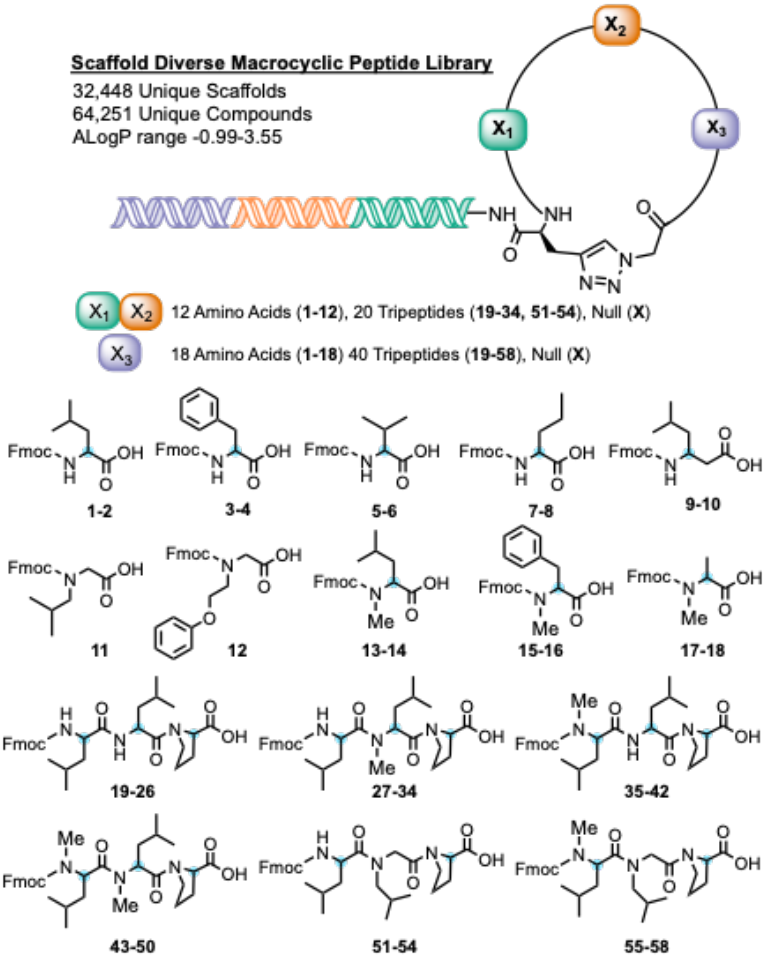


Figure 8: DNA-encoded library design. All possible stereoisomers (highlighted in blue) were used as building blocks. N-methyl nucleophiles (BB13-18, 35-42, 55-58) were only used at cycle 3.

homoamino acid and tripeptide were included to maximize the range of backbone geometries within the library. The aliphatic side-chains were selected to feature a variety of branching patterns which may produce different effects of hydrophobic shielding dependent on the relative position. Additionally, incorporation of both amino acids and tripeptides at each position yielded four unique ring sizes.

Controls were built into the library to facilitate compound sequencing and deconvolute sequence count chromatograms. First, each building block at each position was encoded with three unique DNA tags at each cycle of chemistry, resulting in 27 unique encoding tag sequences for each compound. Additionally, the library was closed with a partially single stranded tag bearing both a priming region and an N12 unique molecular identifier (UMI). Combined, the tagging in triplicate and UMI resulted in a theoretical ~453 million unique sequences per compound. Another critical control was the inclusion of three null tag sequences at each cycle of chemistry. Incorporating nulls at each cycle of chemistry engineers encoded truncations into the library, which could provide additional chromatographic data for comparison and aid in product identification. Finally, after final acylation with azido acetic acid, half of the library was left linear while the other half was macrocyclized with a copper mediated cycloaddition. Accounting for the null compound diversity, the overall theoretical diversity for either the linear or cyclized library was 64,251.

2.4 DEL Chromatographic Partitioning and Sequencing of Fractions

2.4.1 Strategy to Obtain Sequencing Count Chromatograms

Previously, we relied on LCMS and UV to accurately identify the retention times of our DNA bound substrates. To accomplish this in an encoded library format, a novel strategy was developed. A 5 μM stock of 1:1 linear:cyclized library (2.5 μM in each) was prepared and a 10 μL aliquot (50 pmol) was injected and separated chromatographically under similar conditions as described previously. The long

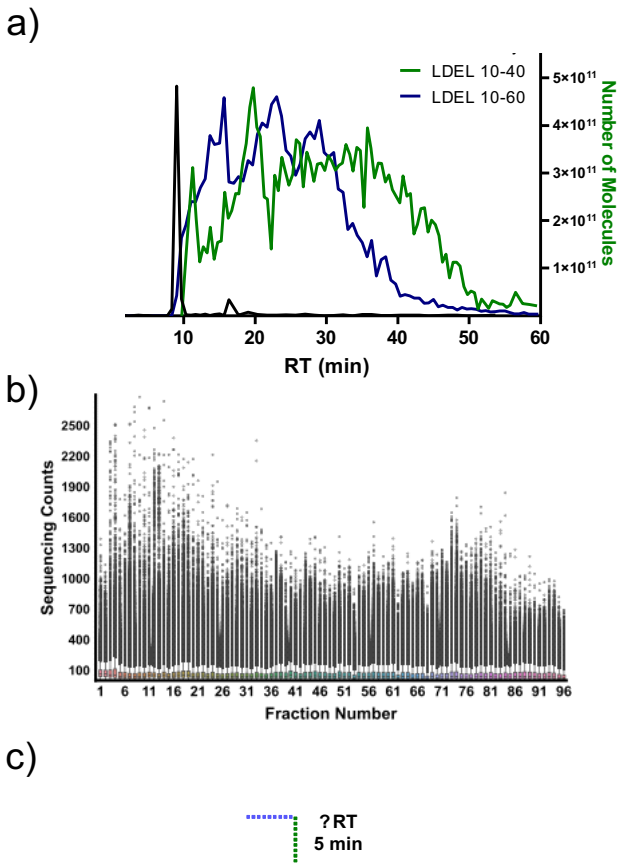


Figure 9: Chromatographic separation and sequencing of library fractions. A) Chromatographic elution of a 1:1 mixture of linear and cyclized library represented by qPCR quantification in each fraction. Library elution is shown in green and blue, with Null library in black. B) Box-and-whisker plots of deduplicated sequencing counts per compound per fraction. C) Matched pair example of sequencing count chromatograms. Two DNA-encoded library members who differ by one side chain (Val, blue; Phe, green).

PRP-C18 column (2.1x250 mm) was employed with 50 mM TEAA buffer and acetonitrile as the mobile phase. The library was collected in 96 different fractions and the amount of library in each fraction was determined by qPCR (**Figure 9a**). This experiment was also performed on a Null library (black trace), which contained a similar number of encoding tags but lacked any pendent warheads. Encouragingly, the Null library elutes primarily in a few early fractions, indicating that the full encoded tag length (114 bp) and partially single-stranded closing tag containing a UMI won't exhibit sequence-dependent retention time differences and significantly effect observed retention. By contrast, the encoded library displayed a broad elution profile that responded strongly to the gradient program employed (blue and green traces). The 10-40 gradient program was selected for sequencing due to the consistency of library elution throughout the run, as a broader eluting library will yield a larger relative chromatographic separation between library members.

Due to the relatively even elution of the library throughout the run, equivolume aliquots from each fraction were taken into PCR and were all run for the same number of amplification cycles. Each fraction was individually barcoded with unique Illumina adapters during the PCR in preparation for high through sequencing. The barcoded fractions were combined and sequenced all together in a Next Seq 2000 run with 1 billion reads. Due to the 1:1 mixture of linear:cyclized library, the overall theoretical library diversity was 128,502, which resulted in potentially 81 unique sequencing reads per compound per fraction (assuming a perfectly even distribution of each compound in each fraction).

2.4.2 Sequencing Depth of Fractions

The deduplicated sequencing counts of each compound in each fraction were determined (**Figure 9b**). The total sequencing reads were 7.6E8 counts (maximum of 1.2E9) and the deduplicated sequencing reads were 5.1E8 counts (67% unique reads). Additionally, the average total deduplicated sequencing counts per fraction were 5.3±1.2E6, indicating good representation of each fraction in sequencing. Generally, every compound in the library was observed in all fractions, with each molecule having an average median sequencing count in each fraction of 25.4, with fractions earlier in the run having slightly higher counts. However, each fraction contains significant sequencing count outliers. These large spikes in sequencing count are indicative of compound elution, with the average maximum sequencing count of each molecule in any fraction being 247.8, a ~10-fold ratio above baseline elution. Plotting the deduplicated sequencing counts against the median retention time of each fraction produces chromatograms of each compound which reveals both quality of compound synthesis and compound lipophilicity via retention times (**Figure 9c**).

2.5 Null Informed Model to Guide Retention Time Assignment and Profile Reactivity

2.5.1 Identification of Synthetic Truncations Through Comparisons to Null Truncation Chromatograms

To extract retention times for each compound from these sequencing count chromatograms, a simple model was initially employed in which peaks were picked and fit to gaussian curves. The latest-eluting peak of sufficient amplitude was

assumed to be the full synthetic product, as truncations should generally elute earlier than full compounds due to the aliphatic nature of the building blocks in this library. However, this simple strategy proved to be insufficient for library-wide analysis as poor coupling efficiency impeded accurate peak assignment in many cases.

To improve overall peak picking fidelity, we developed an analytical framework which compares all observed peaks within a cyclized compound chromatogram to each of its associated null compound chromatograms. By including null tags in each cycle of library synthesis, singly and doubly truncated structures of each compound exist as encoded molecules within the library (**Figure 10a**). For a compound with three unique building blocks, this results in three different single truncation products (ST), three different double truncation products (DT), and a full null construct (TT, triple truncation). Each single truncation and double truncation construct has three unique representations within the library (e.g. Leu-Null-Null, Null-Leu-Null, Null-Null-Leu), resulting in triplicate chromatograms for each null construct. By comparing full compound (FC) chromatograms to each of its associated nulls, truncated chemistry can be identified and quantified. In some cases, the product peak in a FC chromatogram will perfectly match the product peak in the null chromatogram (**Figure 10b**), indicating that chemistry failed at a particular step and the DNA-encoded product was not properly synthesized. In other cases, the FC product peak does not match any of its associated nulls (**Figure 10c**), indicating full product formation. For compounds with suspected product formation, comparisons to

all associated null truncation products can still be utilized to identify possible truncations and gauge the extent of product conversion.

2.5.2 Split-Tree Visualization Reveals Trends in Library-Wide Reactivity

Although adding null encoding tags to DELs is a common strategy to identify potential synthetic truncations that may be surreptitiously binding to the target in affinity selections, comparison to null structures with chromatographic resolution allows for a clearer assessment of the quality of synthesis for each DNA-encoded library member. Applying this analysis to the entire DNA-encoded library identifies truncated compounds and tracks reactivity across all cycles of chemistry. Visualizing these data in a split-tree plot (**Figure 10d**) provides an assessment of coupling efficiency for each building block pair in every step of the synthesis. In the full split tree, each cluster represents an individual building block in the first round of library synthesis. Each of these clusters is connected to 33 additional clusters corresponding to the 32 building blocks and null tag used during cycle 2. As 58 building blocks and a null were used in cycle 3, each of these cycle two clusters terminates in 59 nodes, where each node represents a unique DNA-encoded library member. Red nodes indicate compounds in which the product peak in the FC chromatogram matches the product peak of an associated null, whereas black nodes indicate compounds that were synthesized successfully and thus do not match any associated null product peak.

At the BB1 level, an all-red cluster would indicate either failure of that building block to couple to the library starting material or failure of that building

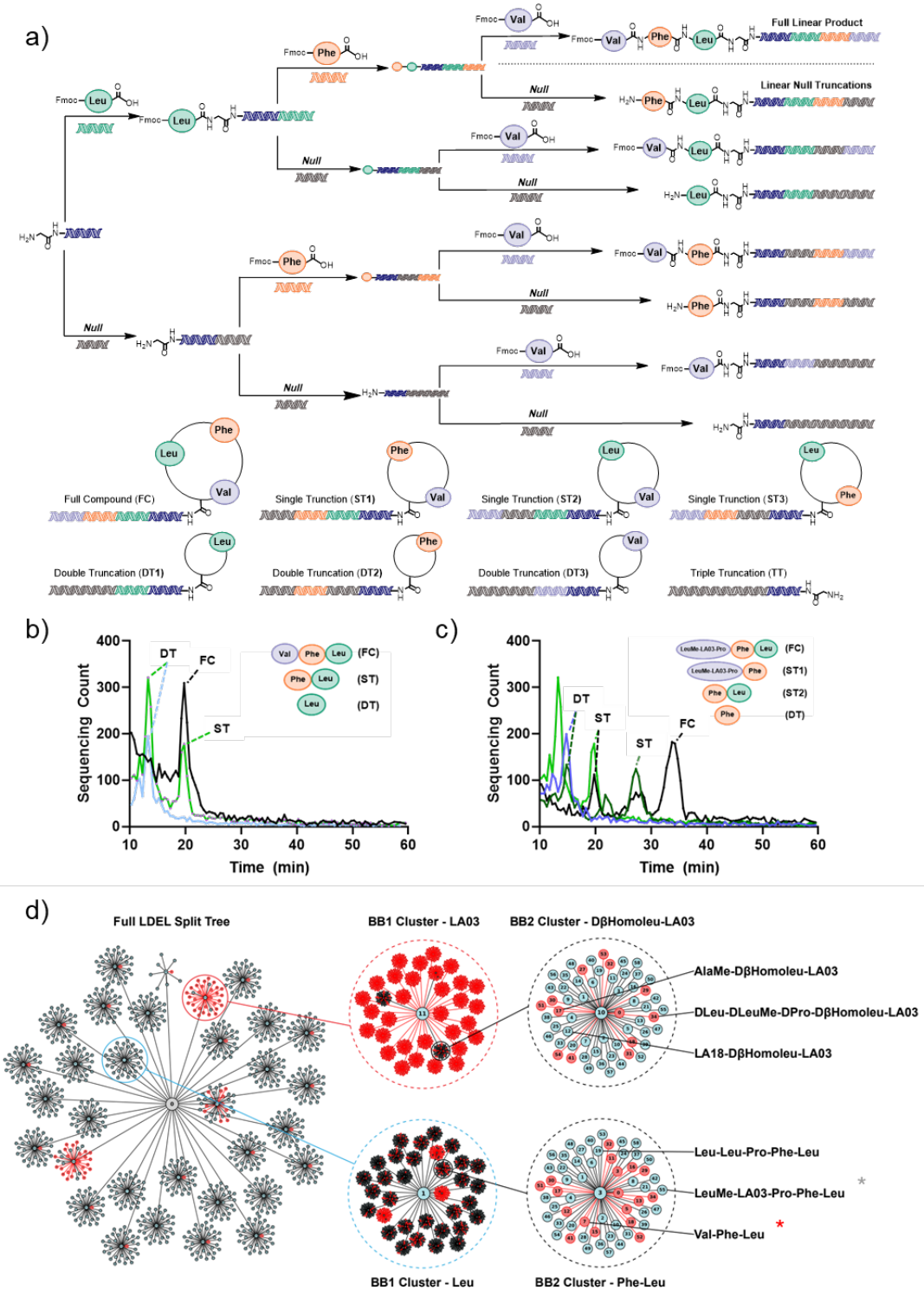
block to act as a nucleophile and couple to all 32 building blocks in cycle 2. Strikingly, two BB1 clusters, which correspond to the two peptoid building blocks (BB11 and 12), have all-red connections to each BB2 cluster except for two specific building blocks. The two Cycle 2 building blocks which successfully coupled to the peptoid nucleophile are the two β-homo amino acids (BB9 and 10), the only two building blocks other than the Fmoc-peptoids which feature a secondary carbon and freely rotatable bond alpha to the carbonyl. Interestingly, the Cycle 1 peptoid building blocks failed to couple to their own respective Fmoc-peptoid building blocks in Cycle 2. These results demonstrate that the null compound analytical framework is not only useful for pruning truncated structures from the dataset, but also for identifying trends in chemistry during DEL synthesis. Data from this analysis can be used to guide future synthesis of DELs by giving data on each individual building block combination, whereas building block validations performed prior to library synthesis typically only gauge conversion to the library starting material or a small set of building blocks.

2.6 Library-Wide Assessment of Peak Assignment and Lipophilicity

2.6.1 Library-Wide Trends Between Capacity Factor and Peptide Size and Backbone N-Substitution

Pruning truncated compounds from our dataset significantly improved the accuracy of retention time assignments and facilitated library-wide assessments.

Figure 10: Null-informed chromatogram analysis A) Scheme showing the result of null tag incorporation during library synthesis for a single compound (Val-Phe-Leu). B) Sequence count chromatogram for compound "Val-Phe-Leu" from 10-40%B library elution. The full compound (FC, black), a single truncation product (ST, green) and a double truncation product (DT, blue) are displayed. Peak overlap between FT and ST indicates incomplete product formation for "Val-Phe-Leu". C) Sequence count chromatogram for compound "LeuMe-LA03-Pro-Phe-Leu" from 10-40%B library elution. The full compound (FC, black), two single truncation products (ST1, dark green, ST2, light green), and a double truncation product (DT, blue) are displayed. Lack of overlap with any associated truncation products indicates good product formation for "LeuMe-LA03-Pro-Phe-Leu". D) Split-tree plot of full DNA-encoded library (LDEL). Each node represents a building block at a given cycle of chemistry, resulting in 32 BB1 clusters, 32 BB2 clusters, and 58 terminal nodes per BB2 cluster. Terminal nodes which are blue indicate a compound whose full compound product peak does not match any of its associated truncations, while red nodes indicate a compound which matches an associated truncation. BB3 connections and nodes are not displayed in the full LDEL split tree for visual clarity.


a)
Full Linear Product
Linear Null Truncations
Null
Full Compound (FC)
Single Trunction (ST1)
Single Trunction (ST2)
Single Trunction (ST3)
Double Truncation (DT1)
Double Truncation (DT2)
Double Truncation (DT3)
Triple Truncation (TT)
b)
Sequencing Count
Time (min)
DT
FC
ST
(FC)
(ST)
(DT)
c)
Sequencing Count
Time (min)
DT
ST
ST
FC
(FC)
(ST1)
(ST2)
(DT)
d)
Full LDEL Split Tree
BB1 Cluster - LA03
BB2 Cluster - DβHomoleu-LA03
AlaMe-DβHomoleu-LA03
DLeu-DLeuMe-DPro-DβHomoleu-LA03
LA18-DβHomoleu-LA03
BB1 Cluster - Leu
BB2 Cluster - Phe-Leu
Leu-Leu-Pro-Phe-Leu
LeuMe-LA03-Pro-Phe-Leu
Val-Phe-Leu

As the tripeptides used in the library were completely comprised of peptoid, N-methyl, and stereo-variants of Leu-Leu-Pro, the average lipophilicity of library members increases as ring size increases (see **Figure S17**). Additionally, lipophilicity is expected to increase significantly as the amide protons in the backbone are removed due to N-methylation or N-alkylation in the peptoid residues. Therefore, as both ring size and the extent of backbone N-substitution increases, retention times are also expected to increase. For the unpruned library (**Figure 11a**), where retention time for each compound was assigned via the simple gaussian model described above, binning the library by these metrics and comparing the capacity factor of each compound within each bin results in weak trends. In the unpruned model, many compounds were assigned unreasonably low retention times and higher levels of N-substitution did not produce significant increases in retention times. This N-substitution effect is likely driven primarily by truncated compounds resulting from failed peptoid couplings in Cycle 2 (see discussion above). The pruned library yields much stronger trends with N-substitution (**Figure 11b**) and contains much fewer compounds with unexpectedly low capacity factors. Of the original 59,932 compounds, 42,095 remained after pruning, indicating that 29.2% of the library had an assigned product peak which matched a Null truncation.

2.6.2 Scaffold-Based Analysis of DNA-Encoded Measurements of Lipophilicity

We also evaluated the quality of our retention time assignments with the Null-informed model through a scaffold-based analysis. For compounds on the same peptide scaffold who differ only by their aliphatic side-chain residues, algorithmic

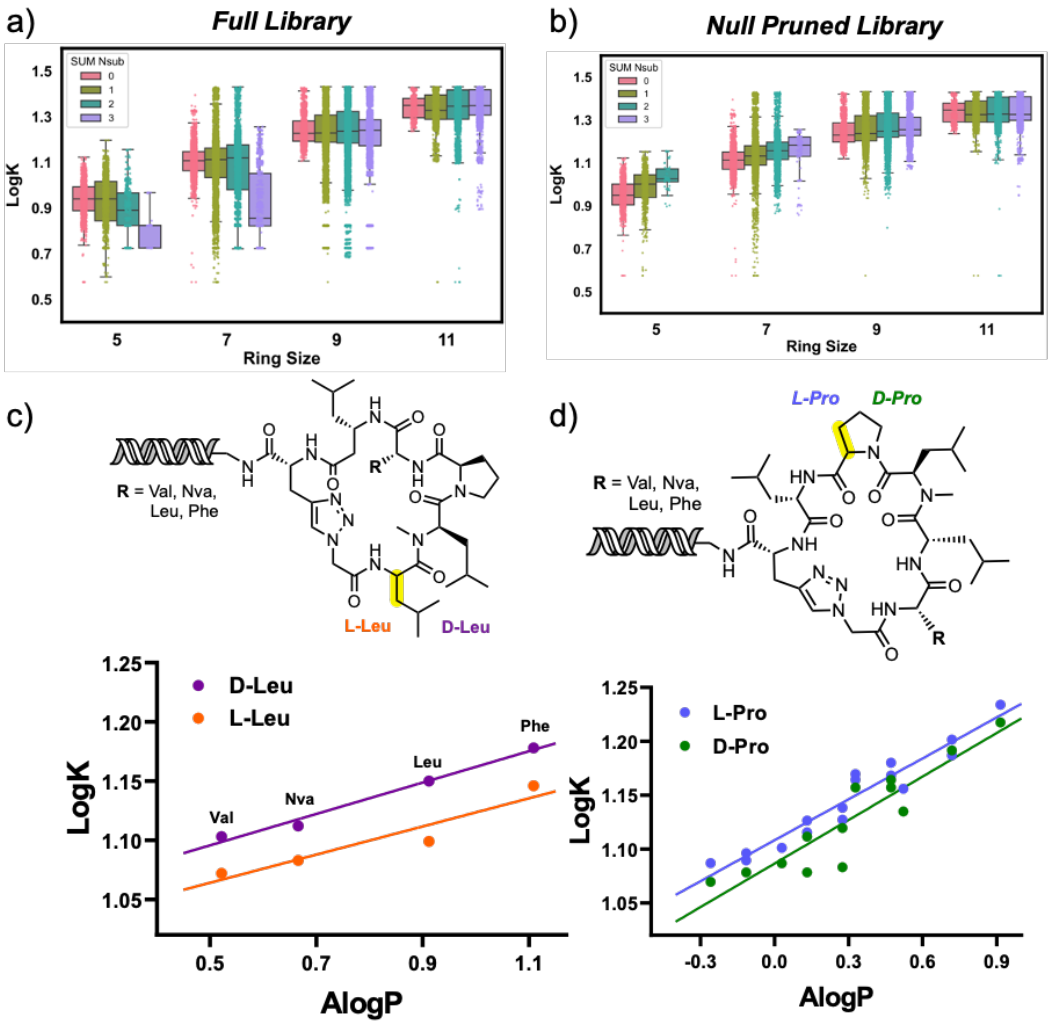


Figure 11: Assessment of library-wide retention time assignments. Box and whisker plots of the capacity factors (retention time) for every compound from the full DNA-encoded library (A) or null pruned DNA-encoded library (B), binned by both ring size and number of N-substituted residues (peptoid or N-methyl). C) Heptapeptide scaffold matched pairs who differ by stereoinversion of a single leucine residue and display a large difference in on-DNA chromatographic capacity factors for each side-chain variant. Linear regressions are displayed for each scaffold. D) Heptapeptide scaffold matched pairs who differ by stereoinversion of their proline residue and contain 16 liposcan members per scaffold. Linear regressions are displayed for each scaffold.

predictions of lipophilicity (ClogP, AlogP) yield strong linear correlations to empirical measurements such as shake-flask partition coefficients or, in this case, capacity factors (assuming those side chains don't significantly affect the overall peptide structure, such as by occluding a backbone amide proton).[24] Modulating the peptide scaffold in a manner that produces conformational changes which change the bulk lipophilicity of the inherent scaffold will break this linear correlation, resulting in a y-intercept difference between the two scaffolds with each scaffold maintaining its own linear correlation. Generally, we observed excellent correlations between capacity factor and AlogP for liposcans on individual peptide scaffolds. In our analysis, we discovered matched pair heptapeptides which differ by a single leucine stereocenter (**Figure 11c**, highlighted in yellow) in which the D-leucine stereoisomer displayed a ~0.05 increase in capacity factor (~2-minute RT difference) across all four liposcan variants. This result is similar to the phenomenon observed in the model compounds **21-24** in this study, wherein the inversion of a stereocenter about a leucine results in a large difference in chromatographic lipophilicity. This scaffold-dependent difference in lipophilicity was commonly observed during analysis of the library and is maintained even on matched-pair scaffolds which contain many more liposcan members (**Figure 11d**). These heptapeptide matched pair scaffolds differ by stereoinversion about their proline residue and contain 16 liposcan members each. These compound sets exhibit a smaller but consistent difference in capacity factors between scaffolds. These scaffolds contain isomers with common yet switched building blocks at positions R1 and R2. Interestingly, these building block matched

pairs occasionally display large differences in capacity factors, indicating that side-chain identity at particular positions may have a large influence on overall peptide structure and lipophilicity.

2.7 Trends Between On-DNA Capacity Factors Derived from Sequencing Counts and Off-DNA Capacity Factors and Passive Permeability in PAMPA

2.7.1 Resynthesis of Encoded Library Members Off-DNA and Trends Between Capacity Factors On- and Off-DNA

As the library-wide and focused scaffold-based analyses both yielded encouraging results, we next selected 44 compounds to resynthesize off-DNA to validate on-DNA capacity factors and investigate trends between these capacity factors and passive permeability. Eight pentapeptides, fourteen heptapeptides, ten nonapeptides, and twelve undecapeptides were selected for resynthesis based on the quality of their on-DNA chromatograms and to cover a diverse spread of peptide structures within each ring size. The compounds were synthesized with traditional SPPS methods in parallel on RFID encoded lanterns. Macrocyclizations were performed on-resin and peptides were cleaved in ethanolic HCl to yield ethyl esters at the C-terminus (**Figure 12a**). After synthesis and purification, thirty peptides were successfully isolated with sufficient purity. The fourteen peptides which were unsuccessfully isolated exhibited extremely low cyclization efficiencies in the on-resin cyclization. The capacity factor of each pure peptide was measured on PRP-C18 using an isocratic-60 method. The on-DNA capacity factors derived from the

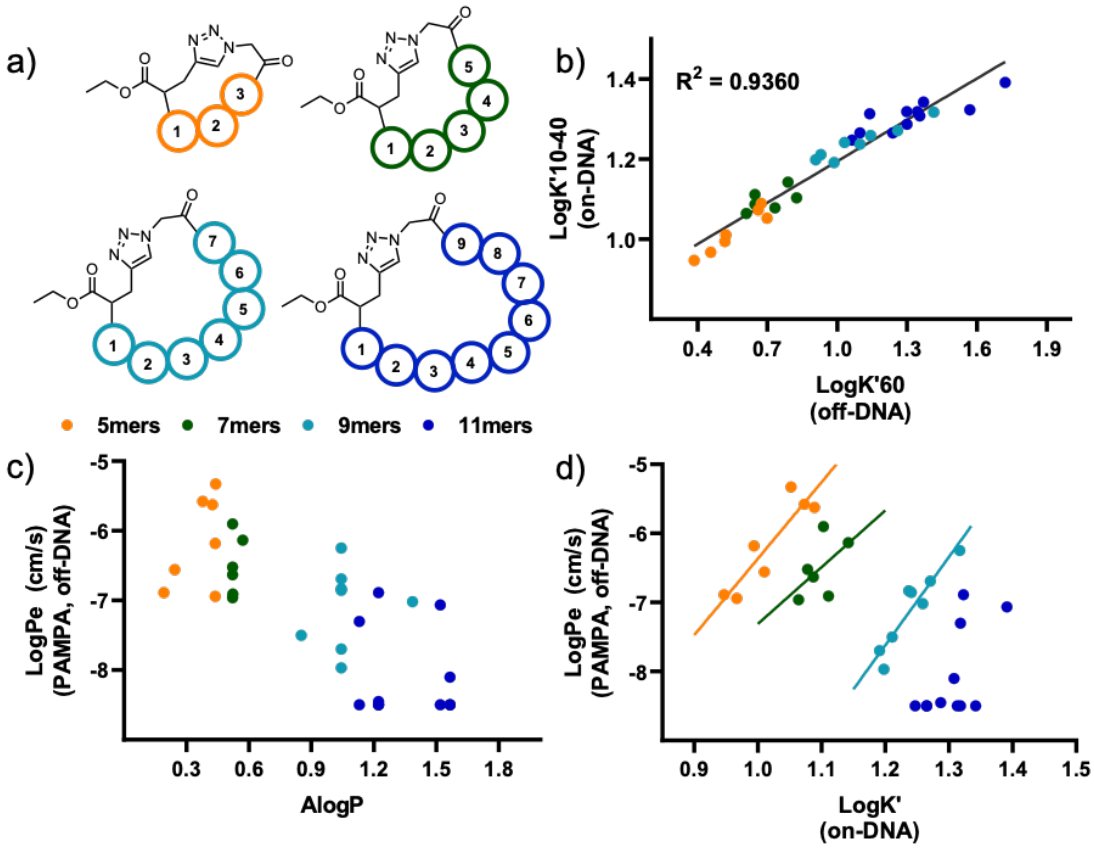


Figure 12 Assessment of on-DNA lipophilicity measurements from the DEL. A) Simplified structures of the resynthesized compound sets. The connection to the DNA has been replaced with an ethyl ester for off-DNA testing. B) Relationship between on-DNA capacity factors and off-DNA capacity factors. An excellent correlation (R^2 = 0.936) is observed across all ring sizes. C) Relationship between PAMPA permeability and calculated lipophilicity (AlogP). Vertical bands of compounds indicate isomers or stereoisomers. D) Relationship between PAMPA permeability and on-DNA capacity factors. Linear regressions for 5mers, 7mers, and 9mers are displayed directly on the plot.

sequencing count chromatograms had an excellent correlation with the off-DNA capacity factors across all ring sizes (**Figure 12b**) (R^2 = 0.936) (see supporting information for all on- and off-DNA chromatograms). The excellent correlation indicates that the assigned retention times from our pruned model are accurate and reflective of the peptides off-DNA chromatographic lipophilicity. This suggests that

not only is the chromatographic separation driven primarily by the lipophilicity of the pendent peptide, but that the sequencing count-based strategy we developed is able to produce accurate chromatograms for each compound. Furthermore, the on-DNA chromatography appears to be sensitive to conformational differences, as this resynthesis dataset features extensive variation in stereochemistry and other backbone elements (N-methyls, peptoids, β-homo amino acids).

2.7.2 Trends Between On-DNA Capacity Factors and PAMPA

To investigate trends in cell permeability, the passive permeability of each peptide was measured in PAMPA. Our lab[2, 25] and others[9] have previously demonstrated the ability for PAMPA to act as a functional surrogate for cell-based permeability assays for bRo5 compounds ($R^2 > 0.70$). Comparing the PAMPA LogPe to AlogP yields extremely poor correlations (**Figure 12c**), with many compounds in this dataset being isomers or stereoisomers which bear the same AlogP, despite having significantly different chromatographic lipophilicities both on- and off-DNA. Comparing PAMPA permeability with on-DNA capacity factors yields strong correlations for each ring size (**Figure 12d**). The slope of each correlation, which is an indication of the gains in PAMPA permeability per increase in capacity factor is consistent between each dataset, indicating similar increases in permeability across ring sizes, except for the undecamers. In this dataset, the lipophilicity required to achieve high PAMPA permeability increases as peptide ring size increases. This trend is expected, as there is a well-established negative relationship between molecular size and passive permeability in PAMPA,[2] and it is likely that the 11mers are

approaching the size threshold for passive permeability in PAMPA. Additionally, the relationship between permeability and lipophilicity is well-known to be parabolic due to loss of aqueous solubility or adhesion to the vessel as compound lipophilicity increases past acceptable thresholds. The resynthesized compounds in this study do not reach this threshold on lipophilicity and display only linear correlations. The permeability of these compounds could be improved by increasing their bulk lipophilicity via side-chain modulation. These results demonstrate that in addition to a direct assessment of compound purity, this chromatographic assay also provides a direct measure of compound lipophilicity that has strong implications for passive permeability.

2.8 Conclusion

DNA-encoded libraries are a powerful drug discovery tool whose potential has been limited from low resynthesis hit rates due to false positives from poor chemistry. This work introduces a novel strategy to get synthesis quality data for individual members of a DNA encoded library, which can significantly improve off-DNA hit confirmations by informing discovery chemists about both successful synthesis of target compounds and the extent and identity of synthetic truncations. This chromatographic assay can also be used to track reaction efficiency for all building block combinations, which represents a powerful tool for encoded library chemists to both identify poorly performing building block combinations and to aid in the design of future libraries with high coupling efficiencies. Although demonstrated with simple amide bond formations here, this strategy could also be applied to more

complex chemistries and be used for building block screening *en masse* across several rounds of chemistry.

In addition to compound purity and reaction efficiency data, this chromatographic assay provides an empirical measurement of lipophilicity for each encoded library member that is sensitive to the 3D conformation of the pendent molecules. These on-DNA chromatographic lipophilicities correlate strongly to off-DNA chromatographic lipophilicities and passive cell permeabilities in PAMPA. Obtaining a direct measurement of lipophilicity for each encoded library member prior to resynthesis would significantly expedite hit-to-lead efforts from protein selection experiments because compounds with optimal lipophilicities can be prioritized for resynthesis and further investigation. Lipophilicity is the most important parameter to control during drug development and is especially important to optimize in the effort to design passively cell-permeable and bioavailable peptide macrocycles which are well beyond the rule-of-five. Empirical measurements of lipophilicity are desirable for conformationally complex molecules such as peptide macrocycles. Traditional chromatographic measurements of lipophilicity[17, 22] can take between 5 to 15 minutes for pure compounds, which would require 10,000-30,000 hours of continuous run time for a library this size.

Although this was performed on a proof-of-concept library with limited chemical diversity, this assay can potentially be scaled up to low million-member libraries with sufficient reads per fraction based on the counts and signal-to-noise obtained in this study. More work is necessary to determine the maximum library

diversity this technique can handle and if this assay is sensitive enough to be performed solely on enriched fractions from affinity selections. Additionally, it is unclear if this strategy will work for an RNA/DNA heteroduplex, and more work must be done to determine the potential to apply this to mRNA display. We are currently investigating application of this chromatographic method to extensively map building block reaction efficiencies and analyze library output from protein selection experiments.

2.9 Experimental Section

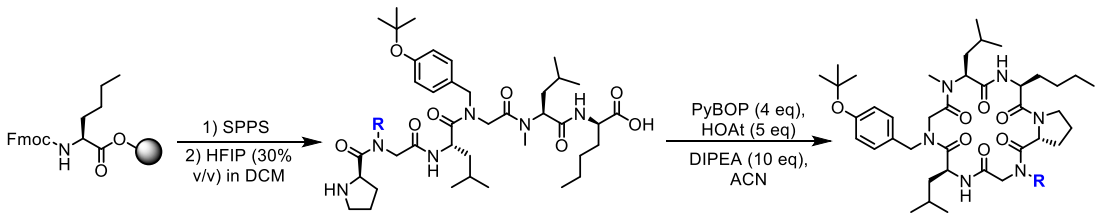
2.9.1 General Materials

Solvents and reagents, including dry solvents, were purchased from Fisher Scientific. Reagents were purchased at a high commercial quality (typically 97% or higher) and used without further purification. Fmoc-protected amino acids and coupling agents were purchased from Combi-Blocks, Oakwood, or Chem-Impex. HPLC solvents for mass spectrometry were Optima grade, purchased from Fisher Scientific. 2-Chlorotrityl chloride polystyrene resin was purchased from Rapp-Polymere. Trityl alcohol Synphase lanterns (D-series, 35 μmol) were purchased from Mimotopes. Dodecane and soy lecithin used in PAMPA were purchased from Alfa Aesar. 1,9-Decadiene for the shake-flask partition assay was purchased from Combi-Blocks. DNA headpiece was purchased from LGC Biosearch Technologies. Oligonucleotide tags were purchased from Integrated DNA Technologies. T4 DNA Ligase, Ligase buffer, and ATP were purchased from New England Biolabs.

Amicon® Ultra centrifugal filters were purchased from Sigma-Aldrich. Columns were purchased from Hamilton, Waters, and Thermo Fischer Scientific. Centrifugal pelleting of DNA was performed on an Eppendorf 5430R. High-field NMR spectra were recorded with a Bruker Avance III HD spectrometer at 500 MHz or a Bruker Ascend spectrometer with a cryoprobe at 800 MHz. NMR spectra were calibrated using residual non-deuterated solvent as an internal reference ($CDCl_3$: 1 H NMR = 7.26).

2.9.2 Synthetic Methods

Synthesis of AF Liposcan Peptomers



Scheme 4: Solid phase peptide synthesis and cyclization of AF Liposcan Peptomers

The AF Liposcan Peptomers were prepared with traditional solid-phase synthesis methods. 2-CTC Resin (Rapp Polymere, 1.6 mmol/g, 3.75g) was swelled in DCM and drained. A solution of Fmoc-Nle-OH (3 eq, 18 mmol) and DIPEA (6 eq, 36 mmol) in DCM (60 mL) was prepared and then administered directly to the resin. The reaction was allowed to proceed overnight, and then the resin was drained and rinsed with DMF (3x) and DCM (3x) and dried under vacuum. Resin loading values were determined by removing a discrete amount of dried resin (~1 mg), removing the Fmoc protecting group with 2% v/v piperidine and 2% v/v DBU in DMF, and then

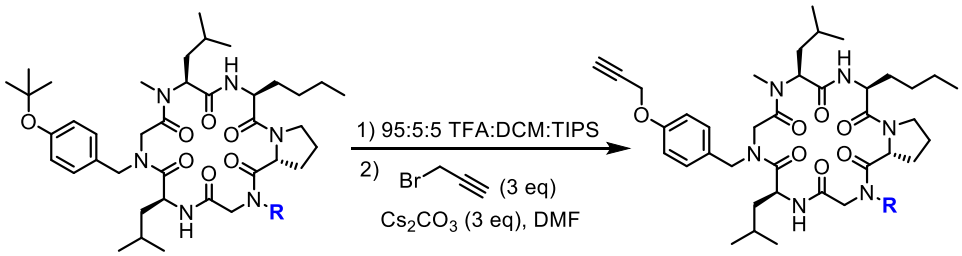
quantifying the UV absorbance of the solution via nanodrop (λ = 299 nm, ε = 6.024 L/mmol*cm) (0.75 mmol/g).

Linear peptides were prepared with repeated amino acid couplings and Fmoc deprotections. Fmoc deprotections were performed by incubating the resin in a solution of 2% v/v piperidine and 2% v/v DBU in DMF for 30 minutes. Amino acid couplings were performed by preparing a solution of Fmoc-amino acid (3 eq), HATU (3 eq), and DIPEA (6 eq) in DMF, and then administering that solution directly to the resin. Couplings were allowed to proceed for 2 hours at room temperature. Peptoid couplings were performed by preparing a solution of bromoacetic acid (5 eq) and DIC (5 eq) in DMF, and then administering that solution directly to the resin after 5 minutes of pre-activation. After 1 hour at room temperature, the solution was completely drained, the resin was washed with DMF (2x), and a solution of the amine in DMF (1 M) was added to the resin. The amine solution and resin were allowed to react overnight at room temperature. The resin was rinsed thoroughly after each deprotection and coupling step, typically with DMF (3x), followed by DCM (2x), and finishing with DMF (1x).

After the third coupling to Fmoc-Leu-OH, the resin was deprotected, washed, dried under vacuum, and then split evenly by mass into 8 tubes. Each peptide was kept separately after this point. Eight different amines were used individually in the follow peptoid coupling step, creating eight separate peptomers. After the final coupling step onto Fmoc-D-Proline-OH and subsequent deprotection, the resin was cleaved twice with 30% v/v HFIP in DCM into tared scintillation vials and then dried

under a stream of air to yield light yellow, dry oils. The crude mass of each peptide was recorded, and all crude linear peptides were analyzed by LCMS.

Acetonitrile (10 mL) and DIPEA (5 eq) were added to each cleaved linear peptide and the solutions were sonicated for 5 minutes. In separate 100 mL round bottom flasks, PyBOP (4 eq), HOAt (5 eq) and DIPEA (5 eq) were combined in ACN (70 mL) with stirring. The solutions of linear peptide and DIPEA in ACN were added dropwise to the stirring solutions of PyBOP, HOAt, and DIPEA in ACN. After the full volume of the linear peptide was administered, the reactions were allowed to proceed at room temperature. Each reaction was checked by LCMS to confirm completeness. Upon reaction completion, the solvent was removed in vacuo, and then each peptide was purified by C18 plug Briefly, each peptide was redissolved in 80:20 H_2O:ACN and fully loaded onto an Agilent SampliQ endcapped C18 column (500 mg). The C18 was washed 3x with 80:20 H2O:ACN and then peptides were eluted with 100% acetonitrile. The purified solutions of cyclized AF Liposcan Peptomers were analyzed by LCMS and then dried in vacuo.



Scheme 5: Installation of alkyne handle for AF Liposcan Peptides

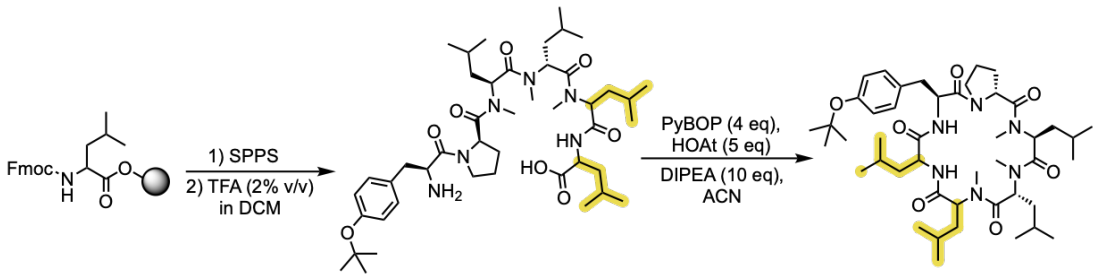
After purification by C18 plug, the dried peptides were redissolved in 90:5:5 TFA:DCM:TES. After 1 hour, the solutions were evaporated under a stream of air to

dryness. Each peptide was checked by LCMS to confirm full removal of the t-Butyl protecting group.

The dried peptides were then dissolved in DMF (to 20 mM) and $CsCO_3$ (3 eq) was added with vigorous stirring. Finally, propargyl bromide (80% w/v in toluene, 3 eq) was added. Each reaction was stirred at room temperature overnight. A liquid-liquid extraction was performed on each reaction using a mixture of ethyl acetate and water. The organic fractions were combined and dried *in vacuo*.

Each peptide was purified by flash chromatography on a Biotage Isolera Prime using a Biotage Sfär Bio C18 column (10g, 300 Å, 20 µm) with H_2O (+0.1% v/v TFA) and ACN (+0.1% v/v TFA) as running buffers. A 30-100%B gradient was applied over 10 column volumes and product fractions were determined by direct inject LCMS. Product containing fractions were frozen over dry ice and the solvent was removed by lyophilization to yield white solids. Each peptide was analyzed by LCMS and a portion was allocated to create 10 mM DMSO stock solutions.

Synthesis of WH Diastereomer Peptides

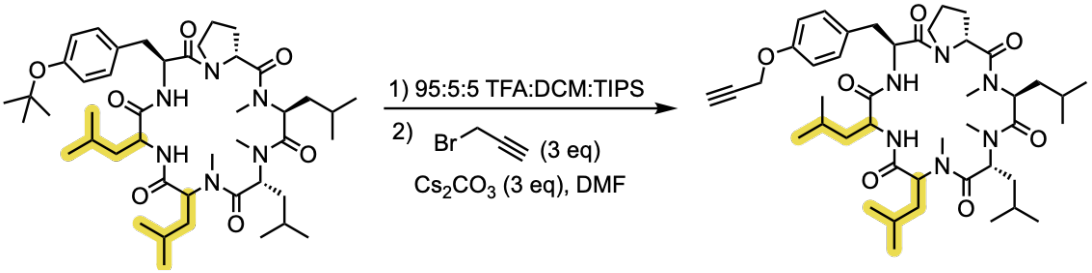


Scheme 6: Solid phase peptide synthesis and cyclization of WH Diastereomer Peptides

The WH Diastereomer Peptides were prepared with solid-phase synthesis methods as described above.

2-CTC Resin (Rapp Polymere, 1.6 mmol/g, 3.75g) was loaded with both Fmoc-L-Leu-OH and Fmoc-D-Leu-OH. Each resin was split equally in half by mass. Linear peptides were prepared as described above. Each linear peptide was synthesized starting from either 2-CTC functionalized with L-Leucine or D-Leucine (0.1 mmol starting scale). After the final coupling step onto Fmoc-D-Proline-OH and subsequent deprotection, the resin was cleaved twice with 30% v/v HFIP in DCM into tared scintillation vials and then dried under a stream of air to yield light yellow, dry oils. The crude mass of each peptide was recorded, and all crude linear peptides were analyzed by LCMS.

Cyclizations were performed as described above. After confirmation of reaction completion by LCMS, the solvent was removed *in vacuo*, and then each peptide was purified by reverse phase flash chromatography using a Biotage Sfär Bio C18 column (10g, 300 Å, 20 µm) with H_2O (+0.1% v/v TFA) and ACN (+0.1% v/v TFA) as running buffers. A 20-100%B gradient was applied over 10 column volumes and product fractions were determined by direct inject LCMS. The solvent from the



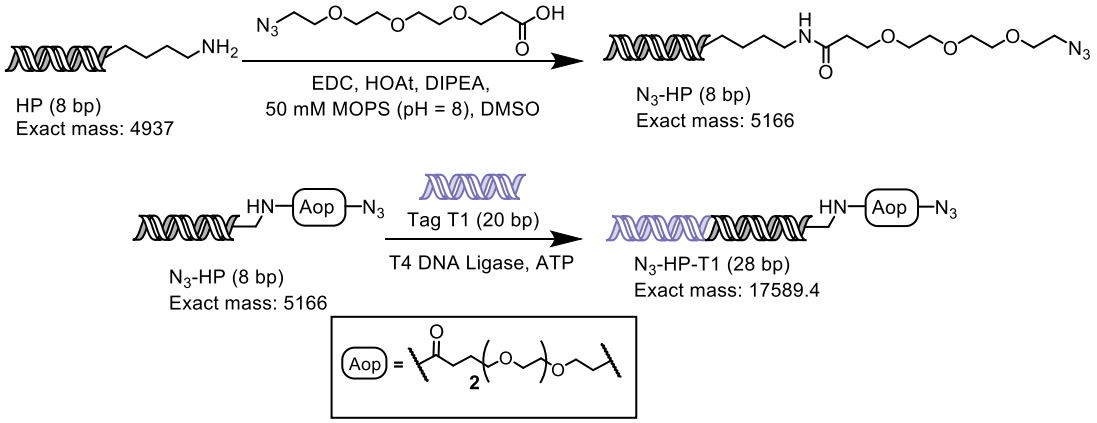
Scheme 7: Installation of alkyne handle for WH Diastereomer Peptides

product containing fractions was removed *in vacuo* to yield clear, dry oils (average of 16.5% recovery).

After purification, the dried peptides were redissolved in 90:5:5 TFA:DCM:TES. After 1 hour, the solutions were evaporated under a stream of air to dryness. Each peptide was checked by LCMS to confirm full removal of the t-Butyl protecting group.

The dried peptides were then dissolved in DMF (to 30 mM) and $CsCO_3$ (6 eq) was added with vigorous stirring. Finally, propargyl bromide (80% w/v in toluene, 6 eq) was added. Each reaction was stirred at room temperature overnight. A liquid-liquid extraction was performed on each reaction using a mixture of ethyl acetate and water. The organic fractions were combined and dried in vacuo to yield clear oils. Each peptide was taken up in 50:50 H_2O:ACN, frozen over dry ice, and lyophilized to yield fluffy white solids (average 10.2% recovery). Each peptide was analyzed by LCMS and a portion was allocated to create 10 mM DMSO stock solutions.

Synthesis of Small Molecule-DNA Conjugates

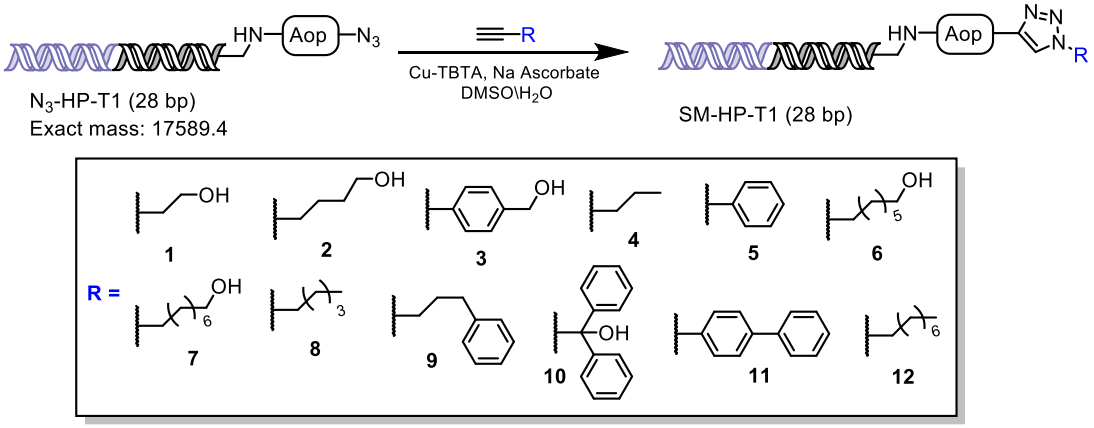


Scheme 8: Synthesis of DNA-Headpiece starting material, N_3-AOP-HP-T1

MOPS Buffer (50 mM, 0.5M NaCl, pH = 8, 200 µL) and water (1600 µL) were added to a solution of Headpiece in water (5 mM, 200 µL) and mixed by pipet. Fresh solutions of azide-PEG3-COOH (300 mM), EDC (300 mM), HOAt (60 mM), and DIPEA (300 mM) were prepared in DMSO. A pre-activated solution of azide-PEG3-COOH was created by mixing solutions of azide-PEG3-COOH (100 eq, 333.3 µL), EDC (75 eq, 250 µL), HOAt (15 eq, 250 µL), and DIPEA (75 eq, 250 µL) with DMSO (916.7 µL). After 5 minutes of pre-activation, the entire solution was added to the Headpiece solution and mixed by pipet. The reaction (0.25 mM, 4 mL) was mixed by inversion at room temperature for 2 hours. The DNA was precipitated by adding a cold aqueous solution of NaCl (3.0 M, 400 µL) and cold EtOH (12 mL). The solution was vortexed thoroughly and stored in a -20°C freezer for 1 hour. The DNA was pelleted by centrifugation for 1 hour at 4°C (18213 x g) and the EtOH solution was completely removed to yield a white pellet. The pellet was redissolved in Milli-Q

water (1 mL), quantified by nanodrop, and analyzed by LCMS (93% conversion by UV).

To an aliquot of the N_3-HP solution (1 mM, 500 nmol) was added water (228.8 µL), Tag T1 (1.2 eq, 1.09 mM, 546.9 µL), NEB® 10X Ligation buffer (142.86 µL) and NEB® T4 DNA Ligase (500 U/µL, 10 µL). The reaction was allowed to proceed at room temperature overnight and then the reaction was checked by gel electrophoresis. The DNA was precipitated by adding a cold aqueous solution of NaCl (3.0 M, 143 µL) and cold EtOH (4.3 mL). The solution was vortexed thoroughly at maximum speed and stored in a -20°C freezer for 1 hour. The DNA was pelleted by centrifugation for 1 hour at 4°C (18213 x g) and the EtOH solution was completely removed to yield a white pellet. The pellet was redissolved in Milli-Q water (500 µL), quantified by nanodrop, and analyzed by LCMS.



Scheme 9: Synthesis of small molecule (SM)-DNA conjugates

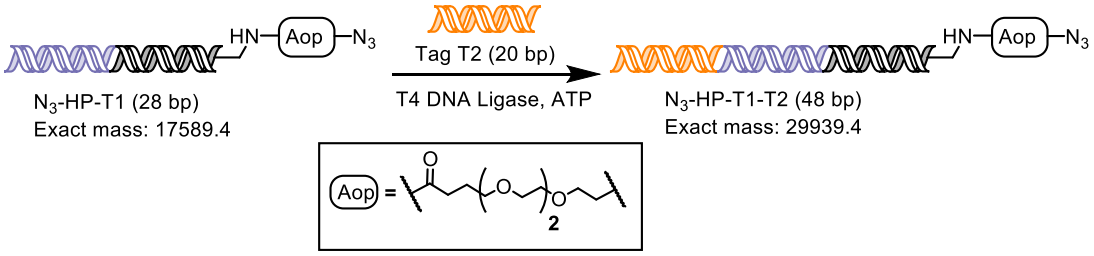
The click reaction conditions were adapted from the click chemistry protocol published from Lumiprobe. A mixed solution of $CuSO_4$ (9.5 mM) and TBTA (10.4 mM) was prepared in 55:45 DMSO:H_2O. Additionally, an aqueous solution of

sodium ascorbate was prepared (10 mM) as well as alkynes 1-12 (20 mM) in DMSO. Twelve aqueous solutions of N_3-HP-T1 (1 mM, 20 nmol) were aliquoted into 1.5 mL Eppendorf tubes and TEAA Buffer (2 M, pH =7, 10 µL) and DMSO (30 µL) were added. The solution was vortexed, and then each alkyne **1-12** was individually added (10 eq, 10 µL) to a separate tube, setting up 12 reactions. Each solution was briefly vortexed, and then sodium ascorbate (2.5 eq, 10 µL) was added. Each tube was vortexed, and then degassed by bubbling with an argon balloon. $CuSO_4$-TBTA (5 eq, 10 µL) was added and the solution was degassed with argon again and vortexed thoroughly. Each reaction (0.2 mM, 100 µL) was mixed by inversion at room temperature overnight. The DNA was precipitated by adding a cold solution of NaCl (10 µL) and cold EtOH (300 µL). The solution was vortexed, stored at -20°C for 1 hour, and then centrifuged at 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (40 µL).

To the SM-HP-T1 solutions were added an aqueous solution of sodium diethyldithiocarbamate (NaDTC) (100 mM, 3.2 µL) to a final concentration of 8 mM. The solution was mixed for 1 hour at room temperature, and then the solution was precipitated by adding a solution of cold NaCl (3.0 M, 5 µL) and cold EtOH (150 µL). The solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (40 µL). The EtOH precipitation step was repeated until the resulting solution was clear (~3x). The final pellet was redissolved in Milli-Q water (40 µL), quantified by nanodrop, and analyzed by LCMS.

Each reaction was purified with RP-HPLC on a Shimadzu SCL-10a VP system using an XBridge C18 OBD Prep Column (10 x 150 mm, 5 µm, 110 Å). The running buffers were 50 mM TEAA (pH = 7) (A) and 50:50 Buffer A:ACN (B) with a flow rate of 5 mL/min. A 20-70%B gradient was implemented over 20 minutes. All library fractions were frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. Purified SM-HP-T1 compounds (**1-12**) were redissolved in Milli-Q H_2O (500 µL) and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters. Each reaction was quantified by Nanodrop and analyzed by LCMS (average recovery of 10.3%).

Synthesis of AF Liposcan Peptide-DNA Conjugates

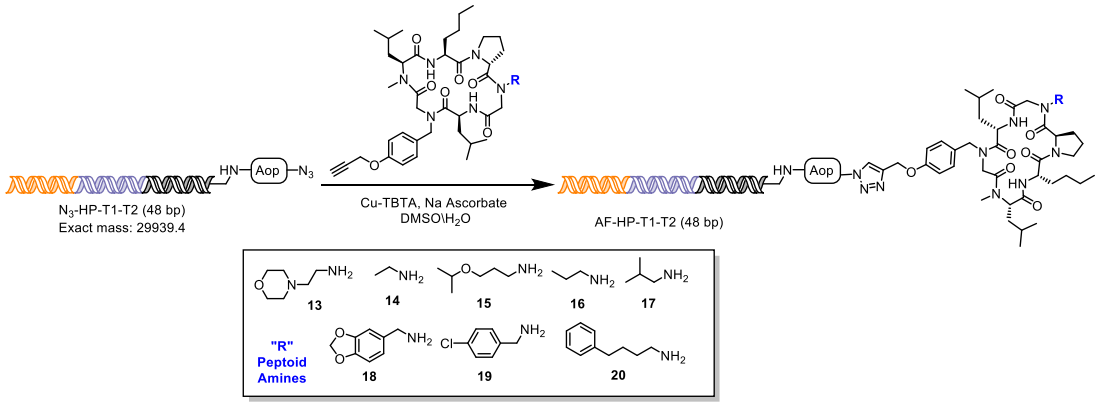


Scheme 10: Synthesis of DNA-headpiece starting material, N_3-HP-T1-T2

To an aliquot of the N_3-HP-T1 solution (0.753 mM, 200 nmol) was added water (77.09 µL), Tag T2 (1.2 eq, 1.056 mM, 227.27 µL), NEB® 10X Ligation buffer (63.49 µL) and NEB® T4 DNA Ligase (500 U/µL, 2 µL). The reaction (0.315 mM, 634.9 µL) was allowed to proceed at room temperature overnight and was checked by gel electrophoresis. The DNA was precipitated by adding a cold aqueous solution of NaCl (3.0 M, 63.5 µL) and cold EtOH (1.9 mL). The solution was vortexed thoroughly at maximum speed and stored in a -20°C freezer for 1 hour. The

DNA was pelleted by centrifugation for 1 hour at -4°C (18213 x g) and the EtOH solution was completely removed to yield a white pellet. The pellet was redissolved in Milli-Q water (200 µL), quantified by nanodrop, and analyzed by LCMS.

The reaction was purified with RP-HPLC using an XBridge C18 OBD Prep Column (10 x 150 mm, 5 µm, 110 Å). The running buffers were 50 mM TEAA (pH = 7) (A)



Scheme 11: Synthesis of Peptide Liposcan(AF)-DNA conjugates

and 50:50 Buffer A:ACN (B) with a flow rate of 5 mL/min. A 05-60%B gradient was implemented over 20 minutes. All library fractions were frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. Purified SM-HP-T1-T2 was redissolved in Milli-Q H_2O (2 mL) and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters. The reaction was then quantified by Nanodrop and analyzed by LCMS (50% recovery).

A mixed solution of $CuSO_4$ (9.5 mM) and TBTA (10.4 mM) was prepared in 55:45 DMSO:H_2O. Additionally, an aqueous solution of sodium ascorbate was prepared (10 mM) as well as peptides **AF13-20** (20 mM) in DMSO. Eight aqueous solutions of N_3-HP-T1-T2 (1.087 mM, 3 nmol) were aliquoted into 0.2 mL tubes and

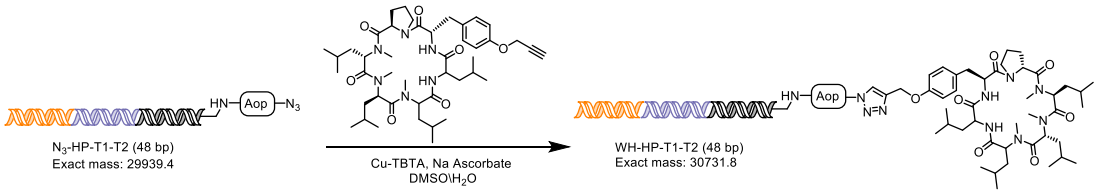
TEAA Buffer (2 M, pH =7, 2 μL) and DMSO (5.24 μL) were added. Each solution was vortexed, and then each peptide **AF13-20** was individually added (40 eq, 3.6 μL) to a separate tube, setting up 8 reactions. Each solution was briefly vortexed, and then sodium ascorbate (2.5 eq, 2 μL) was added. Each tube was vortexed, and then degassed by bubbling with an argon balloon. $CuSO_4$-TBTA (5 eq, 2 μL) was added and the solution was degassed with argon again and vortexed thoroughly. Each reaction (0.15 mM, 20 μL) was mixed by inversion at room temperature overnight. The DNA was precipitated by adding a cold solution of NaCl (3M, 2 μL) and cold EtOH (60 μL). Each solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (10 μL).

To each click reaction was added an aqueous solution of NaDTC (100 mM, 0.8 μL) to a final concentration of 8 mM. The solutions were mixed for 1 hour at room temperature, and then the solution was precipitated by adding a solution of cold NaCl (3.0 M, 5 μL) and cold EtOH (150 μL). The solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed and the resulting pellet was redissolved in Milli-Q water (20 μL). The EtOH precipitation step was repeated until the resulting solution was clear (~3x). The final pellet was redissolved in Milli-Q water (10 μL), quantified by nanodrop, and analyzed by LCMS.

Each reaction was purified with RP-HPLC using an XBridge C18 OBD Prep Column (10 x 150 mm, 5 μm, 110 Å). The running buffers were 50 mM TEAA (pH =

7) (A) and 50:50 Buffer A:ACN (B) with a flow rate of 5 mL/min. A 20-90%B gradient was implemented over 20 minutes. All library fractions were frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. Purified AF-HP-T1-T2 compounds (13-20) were redissolved in Milli-Q H_2O (1 mL) and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters. Each reaction was quantified by Nanodrop and analyzed by LCMS (average recovery of 66.9%).

Synthesis of WH Diastereomer Peptide-DNA Conjugates



Scheme 12: Synthesis of Peptide Diastereomer-DNA conjugates

A mixed solution of $CuSO_4$ (9.5 mM) and TBTA (10.4 mM) was prepared in 55:45 DMSO:H_2O. Additionally, an aqueous solution of sodium ascorbate was prepared (10 mM) as well as peptides **WH21-24** (1 mM) in DMSO. Four aqueous solutions of N_3-HP-T1-T2 (0.44 mM, 10 nmol) were aliquoted into 1.5 mL Eppendorf tubes and TEAA Buffer (2 M, pH =7, 6.67 µL) and DMSO (23.97 µL) were added. Each solution was vortexed, and then each peptide **21-24** was individually added (1.5 eq, 15 µL) to a separate tube, setting up 4 reactions. Each solution was briefly vortexed, and then sodium ascorbate (80 eq, 6.7 µL) was added. Each tube was vortexed, and then degassed by bubbling with an argon balloon. $CuSO_4$-TBTA (40

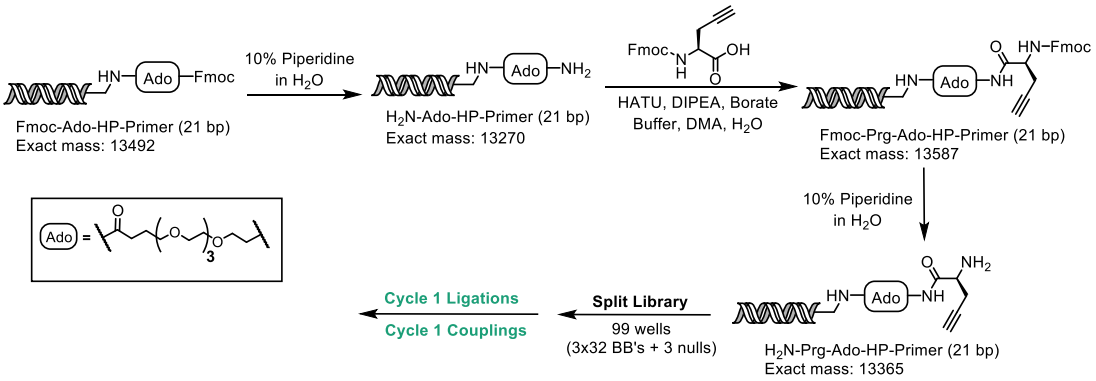
eq, 6.67 µL) was added and the solution was degassed with argon again and vortexed thoroughly. Each reaction (0.15 mM, 66.7 µL) was mixed by inversion at room temperature overnight. The DNA was precipitated by adding a cold solution of NaCl (3M, 6.7 µL) and cold EtOH (200 µL). Each solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (20 µL).

To the click reactions were added an aqueous solution of NaDTC (100 mM, 1.6 µL) to a final concentration of 8 mM. The solution was mixed for 1 hour at room temperature, and then the solution was precipitated by adding a solution of cold NaCl (3.0 M, 2 µL) and cold EtOH (60 µL). The solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (20 µL). The EtOH precipitation step was repeated until the resulting solution was clear (~3x). The final pellet was redissolved in Milli-Q water (20 µL), quantified by nanodrop, and analyzed by LCMS (average 55% recovery).

Each reaction was purified with RP-HPLC using an XBridge C18 OBD Prep Column (10 x 150 mm, 5 µm, 110 Å). The running buffers were 50 mM TEAA (pH = 7) (A) and 50:50 Buffer A:ACN (B) with a flow rate of 5 mL/min. A 05-95%B gradient was implemented over 30 minutes. All library fractions were frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. Purified WH-HP-T1-T2 compounds (**21-24**) were redissolved in Milli-Q H_2O (1 mL) and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters. Each

reaction was quantified by Nanodrop and analyzed by LCMS (average recovery of 18.6%).

DNA-Encoded Library Synthesis



Scheme 13: Synthesis of DNA-Encoded Library starting material, H_2N-Prg-Ado-HP-Primer

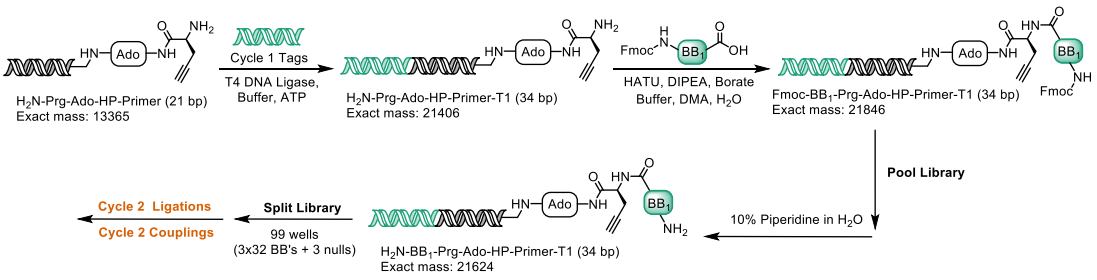
A dry pellet of Fmoc-Ado-HP-Primer (10 µmol) was dissolved in Milli-Q water (10 mL, 1 mM). The solution was made 10% v/v in piperidine by adding neat piperidine (1.1 mL) and then was mixed by inversion at room temperature.

After 1 hour of mixing, a solution of cold NaCl in water was added (3.0 M, 1.1 mL) followed by cold EtOH (30 mL). The solution was vortexed thoroughly and stored in a -20°C freezer for 1 hour. The DNA was pelleted by centrifugation for 30 minutes at -4°C (7197 x g) and the EtOH solution was completely removed to yield an off-white, oily pellet. The pellet was dried under nitrogen and redissolved in Milli-Q water (10 mL, 1mM). The solution was subject to spin filtration with Amicon® Ultra 3 kDa MWCO Centrifugal Filters. The solution was then quantified by nanodrop (10.4 µmol, 104%) and analyzed by LCMS.

The solution of H_2N-Ado-Primer in water (10 μmol, 1.395 mM) was split evenly into four separate 50 mL tubes (1.792 mL/tube). To each tube was added 500 mM borate buffer (2.5 mL) and water (1.604 mL). Fresh solutions of Fmoc-Prg-OH (200 mM), HATU (200 mM), and DIPEA (200 mM) in DMA were prepared. In a separate 50 mL falcon tube, the solutions of Fmoc-Prg-OH (100 eq, 5 mL), HATU (100 eq, 5 mL) and DIPEA (200 eq, 10 mL) were combined with additional DMA (6.416 mL) and mixed by vortex. The preactivated solution of Fmoc-Prg-OH in DMA (26.416 mL) was split evenly (6.604 mL/tube) between the four tubes containing H_2N-Ado-Primer in water (0.2 mM, 12.5 mL/tube). The tubes were mixed by inversion at room temperature for 2 hours and then subject to ethanol precipitation. To each tube was added a solution of cold NaCl (3.0 M, 1.25 mL) and cold EtOH (31.25 mL). The solutions were vortexed thoroughly and stored in a -20° freezer for 1 hour. The DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g) and the EtOH solution was completely removed to yield white pellets. The pellets were dried under nitrogen and then all pellets were carefully combined in 10 mL of Milli-Q water. The resulting solution was then analyzed by LCMS (93.8% conversion by UV).

The resultant solution containing Fmoc-Prg-Ado-HP-Primer was made 10% v/v in piperidine by adding neat piperidine (1.1 mL) and then was mixed by inversion at room temperature. After 1 hour of mixing, a solution of cold NaCl in water was added (3.0 M, 1.1 mL) followed by cold EtOH (30 mL). The solution was vortexed thoroughly at maximum speed and stored in a -20°C freezer for 1 hour. The DNA

was pelleted by centrifugation for 1 hour at -4°C (7197 x g) and the EtOH solution was completely removed to yield an off-white, oily pellet. The pellet was dried under nitrogen and redissolved in Milli-Q water (10 mL, 1mM). The solution was subject to spin filtration with Amicon® Ultra 3 kDa MWCO Centrifugal Filters. The solution was then quantified by nanodrop (10.28 µmol, 103%) and analyzed by LCMS. This solution containing H_2N-Prg-Ado-HP-Primer, which represents the Library Starting Material, was split evenly into a 2.0 mL 96 deep-well plate and 3 separate 1.5 mL Eppendorf tubes (101 µL/well).



Scheme 14: Cycle 1 ligations and couplings. Exact masses represent the average for the library

A ligation master mix was created by mixing water (25.19 mL) with NEB® 10X Ligation buffer (5.5 mL) and NEB® T4 DNA Ligase (500 U/µL, 110 µL). To each well containing Library Starting Material was added Master mix (280 µL), followed by Cycle 1 Tags (1 mM, 120 µL). 96 encoding tags (32 building blocks x 3) and 3 Null Tags were used in this cycle. Throughout the synthesis, the Null Tags were treated identically to the encoding tags, unless otherwise noted. The plate was sealed, and all reactions (0.2 mM, 500 µL) were mixed gently by vortex. Reactions

were allowed to proceed at room temperature overnight. Each well was analyzed by gel electrophoresis. Select wells were analyzed by LCMS.

To each well was added a solution of cold NaCl (3.0 M, 50 μL) and cold EtOH (1.25 mL). The solutions were mixed thoroughly by pipet after the addition of EtOH. The plate was carefully sealed and mixed again by vortex. The plate was then stored at -20°C for 1 hour and DNA was pelleted by centrifugation for 1 hour at -4°C (2204 x g). The EtOH solution was removed by inversion and the plate was allowed to dry inverted on a paper towel for 2 minutes. The plate was covered with a paper towel and allowed to dry out in a fume hood overnight, then Milli-Q water (100 μL) was added to each well to create 1 mM solutions of Cycle 1 Ligation Product. Borate buffer (500 mM, 100 μL) and DMA (125 μL) were added to each well containing Cycle 1 Ligation Product. Fresh solutions of Building Blocks 01-32 (200 mM), HATU (200 mM), DIPEA (200 mM) were prepared. Pre-activated building block solutions were prepared in a separate 2.0 mL 96-well plate by combining the DMA solutions of Building Block (100 eq, 50 μL), HATU (100 eq, 50 μL), and DIPEA (150 eq, 75 μL) and mixing by pipet. After 5 minutes of pre-activation, the Building Block solutions were administered directly into the wells containing Cycle 1 Ligation Product. The coupling reactions (0.2 mM, 500 μL) were mixed by pipet, sealed carefully, and then allowed to proceed overnight at room temperature with gentle lateral mixing on Scilogex SK-L 180E shaker at 100 rpm. The Null Ligations were subject to similar conditions with pure DMA used instead of Building Block. Each reaction was then analyzed by LCMS.

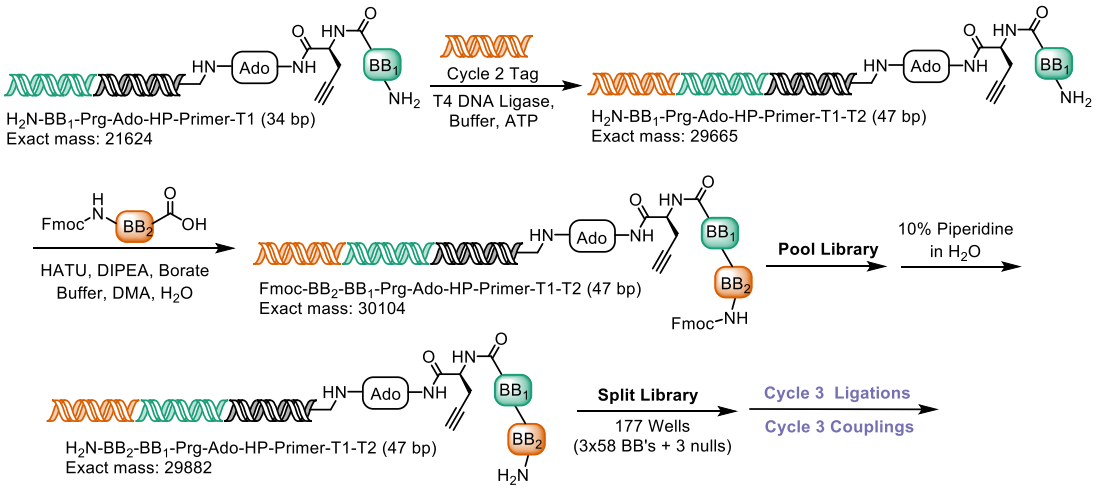
A reservoir containing 50 mL of cold ethanol was prepared. The full volume of each well (500 µL) containing Cycle 1 Coupling Product was pipetted directly into the reservoir. Each well was rinsed with a solution of 50:50 DMA:H_2O (50 µL/well), followed by water (50 µL/well), and finally by DMA (50 µL/well). The solution was carefully transferred into 6 separate tubes (19.06 mL/tube). A solution of cold NaCl (3.0 M, 1.07 mL) and cold EtOH (21.16 mL) was added to each tube. Each tube was mixed by vortex, stored at -20°C for 1 hour and the DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g) to yield a white pellet.

The pellet was dissolved in 20 mL of Milli-Q water and was filtered through Celite, then filtered through a 0.2 µm nylon syringe filter, and then finally filtered and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters to yield 7.5 mL of Pooled Cycle 1 Product. The library was then analyzed by LCMS.

The library was purified in batches with RP-HPLC using a Gemini LC C18, 21.2x100mm, 5 µm, 110 Å column. The running buffers were 50 mM TEAA (pH = 7) (A) and ACN (+ 1% H2O) (B) with a flow rate of 25 mL/min. A 05-60%B gradient was implemented over 20 minutes. All library fractions were pooled, frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. All recovered Cycle 1 Library Product was redissolved in 32 mL of H_2O and concentrated through Amicon® Ultra 3 kDa MWCO Centrifugal Filters to yield 5 mL of pure Pooled Cycle 1 Product.

At this point, the Null Ligations were mixed back into the library. The Pooled Cycle 1 Library was made 10% v/v in piperidine by adding neat piperidine (0.5 mL).

The solution was mixed by inversion for 1 hour at room temperature, and then a solution of cold NaCl in water (3.0 M, 0.7 mL) and cold EtOH (15 mL) were added. The solution was mixed by vortex, incubated at -20°C for 1 hour, and the DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g) to yield an off-white, oily



Scheme 15: Cycle 2 ligations and couplings. Exact masses represent the average for the library

pellet. The pellet was dried under nitrogen and dissolved in 3.1 mL of Milli-Q water. The Pooled Cycle 1 Library was then quantified by Nanodrop (3.78 µmol, 37.8% recovery) and analyzed by LCMS. The solution was then split evenly into a 2.0 mL 96 deep-well plate and 3 separate 1.5 mL Eppendorf tubes (31.3 µL/well).

A ligation master mix was created by mixing water (9.984 mL) with NEB® 10X Ligation buffer (2.035 mL) and NEB® T4 DNA Ligase (500 U/µL, 110 µL). To each well containing Cycle 1 Library Product was added Master mix (280 µL), followed by Cycle 2 Tags (1 mM, 44.5 µL). 96 encoding tags (32 building blocks x 3) and 3 Null Tags were used in this cycle. The plate was sealed, and all reactions (0.2

mM, 185 μL) were mixed gently by vortex. Reactions were allowed to proceed at room temperature overnight. A cross-section of the plate was analyzed by gel electrophoresis to confirm complete ligations. Select wells were also analyzed by LCMS.

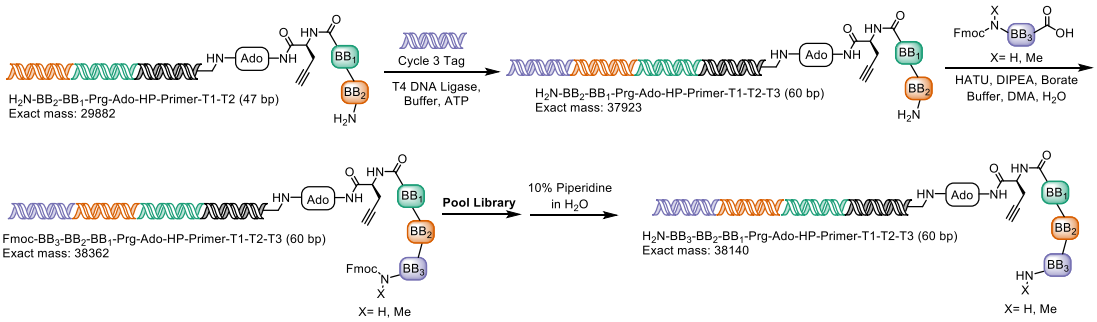
To each well was added a solution of cold NaCl (3.0 M, 18.5 μL) and cold EtOH (555 μL). The solutions were mixed thoroughly by pipet after the addition of EtOH. The plate was carefully sealed and mixed again by vortex. The plate was then stored at -20°C for 1 hour and the DNA was pelleted by centrifugation for 1 hour at -4°C (2204 x g). The EtOH solution was removed by inversion and the plate was allowed to dry inverted on a paper towel for 2 minutes. The plate was covered with Kim Wipes and allowed to dry out in a fume hood overnight, then Milli-Q water (37 μL) was added to each well to create 1 mM solutions of Cycle 2 Ligation Product. Borate buffer (500 mM, 37 μL) and DMA (46.25 μL) were added to each well containing Cycle 2 Ligation Product. Fresh solutions of Building Blocks 01-32 (200 mM), HATU (200 mM), DIPEA (200 mM) were prepared. Pre-activated building block solutions were prepared in a separate 2.0 mL 96-well plate by combining the DMA solutions of Building Block (100 eq, 18.5 μL), HATU (100 eq, 18.5 μL), and DIPEA (150 eq, 27.75 μL) and mixing by pipet. After 5 minutes of pre-activation, the Building Block solutions were administered directly into the wells containing Cycle 2 Ligation Product. The coupling reactions (0.2 mM, 185 μL) were mixed by pipet, sealed carefully, and then allowed to proceed overnight at room temperature with

gentle lateral mixing at 100 rpm. Each Cycle 2 Coupling was then analyzed by LCMS.

A reservoir containing 50 mL of cold ethanol was prepared. The full volume of each well containing Cycle 2 Coupling Product (185 µL) was pipetted directly into the reservoir. Each well was rinsed with a solution of 50:50 DMA:H_2O (50 µL/well), followed by water (50 µL/well), and finally by DMA (50 µL/well). The solution was carefully transferred into 4 separate tubes (20.54 mL/tube). A solution of cold NaCl (3.0 M, 0.804 mL) and cold EtOH (11.6 mL) were added to each tube. Each tube was mixed by vortex, stored at -20°C for 1 hour and the DNA was pelleted by centrifugation for 1 hour at -4°C (7719 x g). to yield white pellets.
The pellets were dissolved in 20 mL of Milli-Q water and then filtered through Celite and a 0.2 µm nylon syringe filter, and then finally filtered and concentrated through Amicon® Ultra 10 kDa MWCO Centrifugal Filters to yield 4 mL of Pooled Cycle 2 Product. The library was then analyzed by LCMS.

At this point, the Null Ligations were mixed back into the library. The Pooled Cycle 2 Library was made 10% v/v in piperidine by adding neat piperidine (0.4 mL). The solution was mixed by inversion for 1 hour at room temperature, and then a solution of cold NaCl in water (3.0 M, 0.4 mL) and cold EtOH (12 mL) were added. The solution was mixed by vortex, incubated at -20°C for 1 hour, and the DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g) to yield an off-white, oily pellet. The pellet was dried under nitrogen and dissolved in 5 mL of Milli-Q water. The Pooled Cycle 2 Library was then quantified by Nanodrop (3.43 µmol, 90.9%

recovery) and analyzed by LCMS. The Pooled Cycle 2 Library (0.67 mM) was then split into 177 wells, with 174 reactions spread evenly between two 2.0 mL 96 well plates and 3 reactions in separate 1.5 mL Eppendorf tubes (27 µL/well).



Scheme 16: Cycle 3 ligations and couplings. Exact masses represent the average for the library

A ligation master mix was created by mixing water (6.09 mL) with NEB® 10X Ligation buffer (1.729 mL) and NEB® T4 DNA Ligase (500 U/µL, 190 µL). To each well containing Cycle 2 Library Product was added Master mix (42.2 µL), followed by Cycle 3 Tags (1 mM, 21.84 µL). 174 encoding tags (58 building blocks x 3) and 3 Null Tags were used in this cycle. The plate was sealed, and all reactions (0.2 mM, 91 µL) were mixed gently by vortex. Reactions were allowed to proceed at room temperature overnight. A cross-section of the plate was analyzed by gel electrophoresis to confirm complete ligations. Select wells were also analyzed by LCMS.

To each well was added a solution of cold NaCl (3.0 M, 9.1 µL) and cold EtOH (273 µL). The solutions were mixed thoroughly by pipet after the addition of EtOH. The plate was carefully sealed and mixed again by vortex. The plate was then stored at -20°C for 1 hour and the DNA was pelleted by centrifugation for 1 hour at -4°C (2204 x g). The EtOH solution was removed by inversion and the plate was allowed to dry inverted on a paper towel for 2 minutes. The plate was covered with Kim Wipes and allowed to dry out in a fume hood overnight, then Milli-Q water (18.2 µL) was added to each well to create 1 mM solutions of Cycle 3 Ligation Product.

Borate buffer (500 mM, 18.2 µL) and DMA (22.1 µL) were added to each well containing Cycle 2 Ligation Product. Fresh solutions of Building Blocks 01-58 (200 mM), HATU (200 mM), DIPEA (200 mM) were prepared. Pre-activated building block solutions were prepared in separate 0.2 mL 96-well plates by combining the DMA solutions of Building Block (100 eq, 9.1 µL), HATU (100 eq, 9.1 µL), and DIPEA (150 eq, 13.65 µL) and mixing by pipet. After 5 minutes of pre-activation, the Building Block solutions were administered directly into the wells containing Cycle 3 Ligation Product. The coupling reactions (0.2 mM, 91 µL) were mixed by pipet, sealed carefully, and then allowed to proceed overnight at room temperature with gentle lateral mixing at 100 rpm. One replicate of each Cycle 3 Coupling was then analyzed by LCMS.

A reservoir containing 50 mL of cold ethanol was prepared. The full volume of each well (91 µL) containing Cycle 3 Coupling Product was pipetted directly into

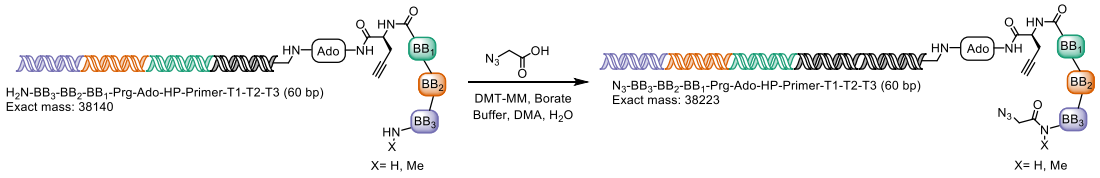
the cold ethanol. Each well was rinsed with a solution of 50:50 DMA:H_2O (50 µL/well), followed by water (50 µL/well), and finally by DMA (50 µL/well). The solution was carefully transferred into 4 separate tubes (22.98 mL/tube). A solution of cold NaCl (3.0 M, 1.05 mL) and cold EtOH (18.95 mL) were added to each tube. Each tube was mixed by vortex, stored at -20°C for 1 hour and the DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g). to yield a white pellet.

The pellet was dissolved in 20 mL of Milli-Q water and was filtered through Celite, then filtered through a 0.2 µm nylon syringe filter, and then finally filtered and concentrated through Amicon® Ultra 10 kDa MWCO Centrifugal Filters to yield 10 mL of Pooled Cycle 3 Product. The library was then analyzed by LCMS.
The library was purified in batches with RP-HPLC using a Gemini LC C18, 21.2x100mm, 5 µm, 110 Å column. The running buffers were 50 mM TEAA (pH = 7) (A) and ACN (+ 1% H2O) (B) with a flow rate of 25 mL/min. A 05-70%B gradient was implemented over 20 minutes. All library fractions were pooled, frozen over dry ice, and the solvent was removed by lyophilization to yield pure white solids. All recovered Cycle 3 Library Product was redissolved in 30 mL of H_2O and concentrated through Amicon® Ultra 10 kDa MWCO Centrifugal Filters to yield 12.35 mL of pure Pooled Cycle 3 Product.

At this point, the Null Ligations were mixed back into the library. The Pooled Cycle 3 Library was made 10% v/v in piperidine by adding neat piperidine (1.24 mL). The solution was mixed by inversion for 1 hour at room temperature, and then a solution of cold NaCl in water (3.0 M, 1.24 mL) and cold EtOH (37 mL) were added.

The solution was mixed by vortex, incubated at -20°C for 1 hour, and the DNA was pelleted by centrifugation for 1 hour at -4°C (7197 x g) to yield an off-white, oily pellet. The pellet was dried under nitrogen and dissolved in 2 mL of Milli-Q water. The Pooled Cycle 3 Library was then quantified by Nanodrop (1.531 µmol, 44.6% recovery) and analyzed by LCMS.

At this point, the Pooled Cycle 3 Library was aliquoted into 10 separate Eppendorf tubes (200 µL, 0.7657 mM) and stored at -80°C for future use.



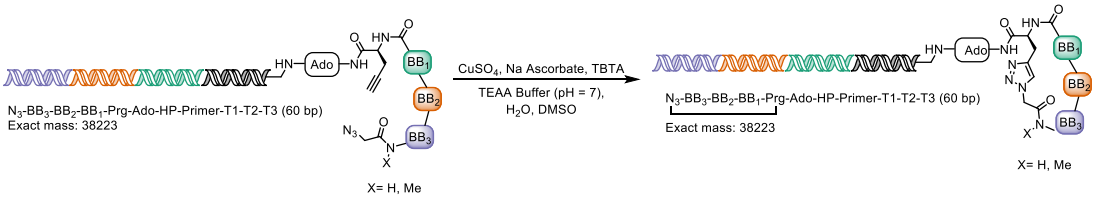
Scheme 17: Library coupling to azido acetic acid. Exact masses represent the average for the library

Two Pooled Cycle 3 Library aliquots (300 nmol, 0.7657 mM) were combined with borate buffer (500 mM, 150 µL) in a 2 mL Eppendorf tube. Fresh solutions of azido acetic acid in DMA (200 mM) and DMTMM in water (200 mM) were prepared. In a separate tube, azido acetic acid was preactivated by mixing the DMA solution of azido acetic acid (100 eq, 150 µL) with the aqueous solution of DMTMM (100 eq, 150 µL) and DMA (658 µL). After 5 minutes of pre-activation, the entire azido acetic acid solution was delivered to the Pooled Cycle 3 Library (0.2 mM, 1500 µL) and the reaction was mixed by inversion at room temperature for 2 hours.

The reaction was precipitated by adding a solution of cold NaCl (3.0 M, 150 µL) and cold EtOH (4.5 mL), followed by thorough mixing by vortex, a 1-hour

incubation at -20°C, and centrifugation at for 1 hour at -4°C (7197 x g). The resulting pellet was dried under nitrogen and dissolved in 391.8 µL of Milli-Q water.

The coupling was repeated exactly as described above (double coupled). The coupling was mixed by inversion at room temperature overnight, and then

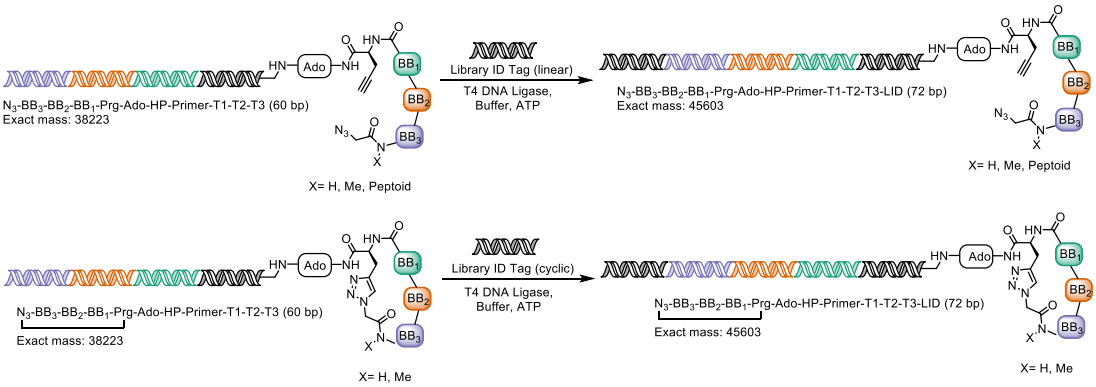


Scheme 18: Copper-mediated azide-alkyne macrocyclization. Exact masses represent the average for the library

precipitated exactly as described above. The resulting pellet was dried under nitrogen and redissolved in Milli-Q water (1 mM, 300 µL).

Fresh solutions of $CuSO_4 \cdot 5H_2O$ (200 mM) and sodium ascorbate (200 mM) were prepared in water. The azido-capped library was split evenly (150 nmol, 150 µL) into two separate 1.5 mL Eppendorf tubes. To one aliquot was added borate buffer (500 mM, 600 µL), followed by the solution of $CuSO_4 \cdot 5H_2O$ (40 eq, 30 µL) and the solution of sodium ascorbate (40 eq, 30 µL). The final cyclization reaction (0.1 mM, 1500 µL) was heated at 60°C for 1 hour. After cooling to room temperature, the DNA was precipitated by adding a cold solution of NaCl (3.0 M, 150 µL) and cold EtOH (4.5 mL). The solution was vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (7197 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (1.5 mL).

To the Cyclized Library solution was added an aqueous solution of NaDTC (100 mM, 120 µL) to a final concentration of 8 mM. The solution was mixed for 1 hour at room temperature, and then the solution was precipitated by adding a solution of cold NaCl (3.0 M, 160 µL) and cold EtOH (4.8 mL). The solution was vortexed,

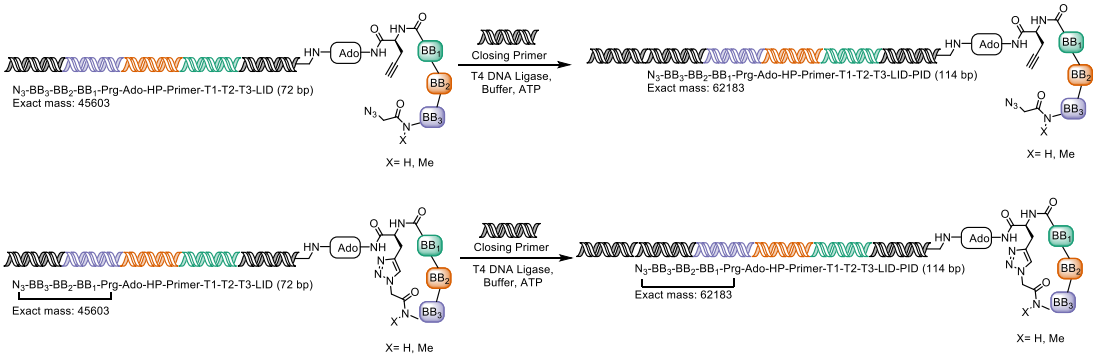


Scheme 19: Library ID Ligations. Exact masses represent the average for the library stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (7197 x g). The EtOH was removed, and the resulting pellet was redissolved in Milli-Q water (1.5 mL). The EtOH precipitation step was repeated until the resulting solution was clear (~3x). The final pellet was redissolved in Milli-Q water (430 µL) and quantified by nanodrop (71 nmol, 47.3% recovery).

To the Liner and Cyclized Library solutions were added water, NEB® 10X Ligation buffer, NEB® T4 DNA Ligase (500 U/µL, 3 µL), and a Library ID tag (1.2 eq). The Linear and Cyclized libraries each received different Library ID tags. The Library ID ligations (0.1 mM) were allowed to proceed at room temperature overnight and then checked by gel electrophoresis.

The reactions were precipitated by adding a solution of cold NaCl (3.0 M, up to 10% v/v) and cold EtOH (3x reaction volume). The solutions were vortexed, stored

at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellets were redissolved in Milli-Q water to 0.5 mM.



Scheme 20: Closing Primer Ligations. Exact masses represent the average for the library

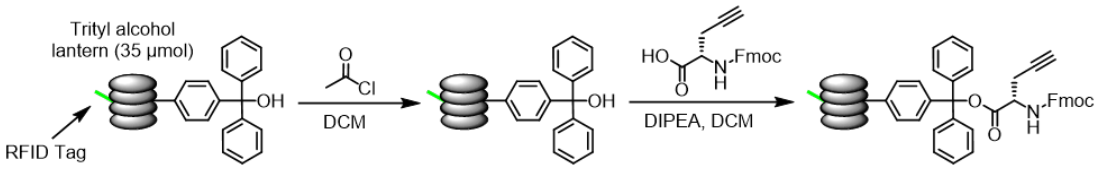
To the Liner and Cyclized Library solutions (5 nmol each) were added water, NEB® 10X Ligation buffer, NEB® T4 DNA Ligase (500 U/µL, 1.5 µL), and a Closing Primer tag (1.5 eq). The Linear and Cyclized libraries received the same Closing Primer tag. The Library ID ligations (0.1 mM) were allowed to proceed at room temperature overnight and then checked by gel electrophoresis.

The reactions were precipitated by adding a solution of cold NaCl (3.0 M, up to 10% v/v) and cold EtOH (3x reaction volume). The solutions were vortexed, stored at -20°C for 1 hour, and then centrifuged for 1 hour at -4°C (18213 x g). The EtOH was removed, and the resulting pellets were redissolved in Milli-Q water (0.05 mM,

100 µL). Each reaction was filtered through Amicon® Ultra 30 kDa MWCO Centrifugal Filters.

The filtered Linear and Cyclized Library (LDEL) solutions were quantified by qPCR and analyzed by LCMS. An equimolar mixture of these libraries (amplifiable amount determined by qPCR) was taken forward into the chromatography experiment.

Resynthesis of Select DNA Encoded Library Members (LDEL001-044)

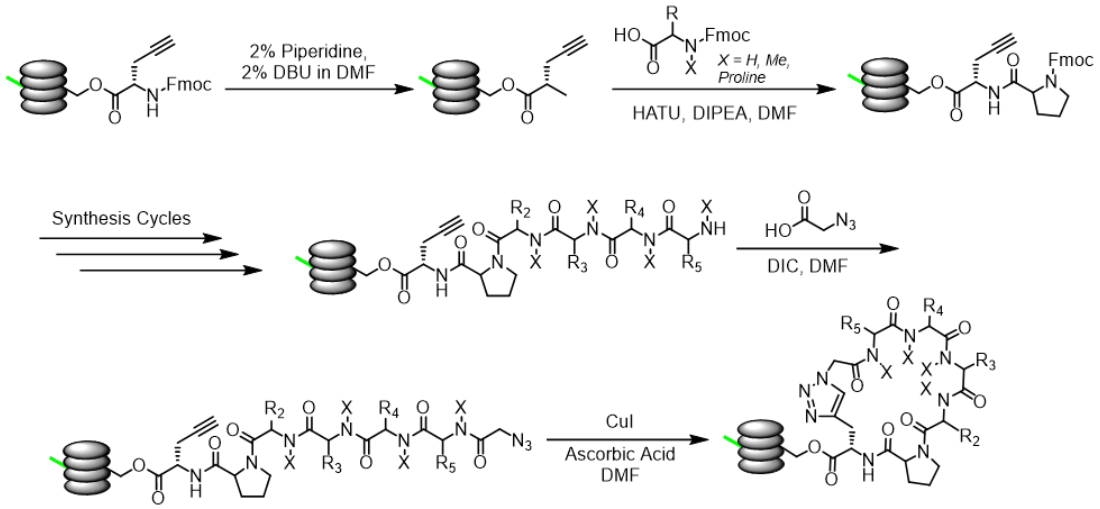


Scheme 21: Loading of trityl alcohol lanterns for synthesis of off-DNA library members. Trityl alcohol lanterns were affixed with an RFID tag (green line) for compound tracking and identification during parallel compound synthesis.

Trityl alcohol Synphase lanterns (Mimotopes, D series, 35 µmol) were affixed with an RFID tag by directly inserting it into the gaps in the polystyrene in each lantern. Each lantern was scanned, and the resulting RFID code was assigned to a peptide (**LDEL001-044**, 44 lanterns, 1.54 mmol total). The lanterns were placed in in a 250 mL round bottom flask and soaked in DCM for 15 minutes. The DCM was decanted and 30 mL of a solution of 25% v/v acetyl chloride in DCM was added to the lanterns. The lanterns were stirred in this solution for two hours and then the solution was decanted. The lanterns were washed twice with DCM, and then a solution of Fmoc-Prg-OH (3 eq, 4.62 mmol, 1.549 g) and DIPEA (6 eq, 9.24 mmol, 1.61 mL) in DCM (30 mL) was added to the lanterns. The lanterns were stirred in this

solution at room temperature overnight, and then the solution was drained, and the lanterns were washed with DMF (3x) and DCM (3x). The lanterns were then dried in vacuo overnight.

Once the lanterns were dry, one full lantern was removed and was added to 1 mL of a solution of 2% v/v piperidine and 2% v/v DBU in DMF. The lantern was shaken in this solution for 1 hour and then the full 1 mL and lantern were removed



Scheme 22: Solid phase peptide synthesis of off-DNA peptides from LDEL on RFID-encoded lanterns

and added to 9 mL of EtOH. The amount of Fmoc in solution was quantified by UV-Vis absorbance via nanodrop (λ = 299 nm, ε = 6.024 L/mmol*cm). The loading value was determined to be 30 µmol/lantern.

Forty-four linear peptides were prepared with repeated amino acid couplings and Fmoc deprotections. Fmoc deprotections were performed by incubating the lanterns in a solution of 2% v/v piperidine and 2% v/v DBU in DMF for 30 minutes. Reactions were performed in either standard polypropylene solid phase synthesis tubes or, for larger lantern quantities, in a round bottom flask. Amino acid couplings

were performed by preparing a solution of Fmoc-amino acid (3 eq), HATU (3 eq), and DIPEA (6 eq) in DMF (~0.05 M in Fmoc-amino acid), and then administering that solution directly to the lanterns. Couplings were allowed to proceed for 2 hours at room temperature. Peptoid couplings were prepared by preparing a solution of bromoacetic acid (5 eq) and DIC (5 eq) in DMF (~0.05 M), and then administering that solution directly to the lanterns after ~5 minutes of pre-activation. After 1 hour at room temperature, the solution was completely drained, the lanterns were washed, and a solution of the amine in DMF (1 M) was added to the lanterns. The amines were allowed to react overnight at room temperature. The lanterns were rinsed thoroughly with DCM (3x) and DMF (3x) after each deprotection and coupling step.

After completion of linear peptides, all the lanterns were grouped and were deprotected with a solution of 2% v/v piperidine and 2% v/v DBU in DMF for 30 minutes. Afterwards, they were washed thoroughly with DMF (3x), DCM (2x), and DMF (1x) and then reacted with a solution of azido acetic acid (5 eq) and DIC (5 eq) in DMF (~0.05M) for 2 hours at room temperature. This coupling step was then repeated (double coupled). The lanterns were rinsed thoroughly with DMF (3x) and DCM (3x), dried in vacuo, and analyzed by LCMS.

The peptides were washed with DMF (2x) and then cyclized on resin by adding a solution of 20% v/v piperidine in DMF containing copper iodide (1.5 eq), followed by a separate solution of 20% v/v piperidine in DMF containing ascorbic acid (5 eq) (final reaction was ~0.05 M in copper iodide). The reactions were wrapped in foil and allowed to proceed overnight at room temperature. After

overnight reaction (~16 hr), the copper-solutions were decanted and the resin was rinsed with DMF (2x), DCM (3x), and DMF (1x). The solution was then rinsed with a copper-chelating solution made up of 5% w/v NaDTC in DMF. This solution was introduced to the lanterns at room temperature and allowed to shake for 30 minutes, producing a dark brown color. This solution was then drained, the lanterns were rinsed with DMF (3x), and the process was repeated until no color was produced upon addition of NaDTC (typically 3-4 total washes).

After cyclization and removal of copper, the peptides were rinsed with DCM (3x). A cleavage solution consisting of 0.3M HCl in EtOH was prepared by combining a solution of HCl in dioxane (0.33M, 45 mL) with EtOH (neat, 5 mL). This solution was added to each lantern in separate containers and was allowed to react overnight at room temperature. Afterwards, the solutions were drained into tared scintillation vials and the lanterns were rinsed with a solution of 2% v/v 2,6-lutidine in DCM (5x) and combined with the initial cleavage solution. The cleaved peptide solutions were blown down to dryness. The resulting solids were first redissolved in 95:5 H_2O:ACN and were loaded onto a C18 plug (Agilent SampliQ). The C18 was rinsed with 95:5 H_2O:ACN twice to remove 2,6-lutidine chloride salts. Peptides were eluted from the plug with 4 column volumes of ACN and dried under vacuum to yield clear oils. Each peptide was then reconstituted in a minimum amount of MeOH (~0.6 mL). Each peptide was purified by reverse-phase flash chromatography on a Biotage Isolera Prime using a Biotage Sfär Bio C18 column (10g, 300 Å, 20 μm). The mobile phase A was water (+0.1% TFA) and the mobile phase B was ACN (+0.1% TFA)

using a flow rate of 25 mL/min. A general gradient program was employed to purify all the compounds consisting of a 2 CV equilibration at 25%B followed by a 12 CV gradient from 25-70%B followed by a 4 CV hold at 100%B. Pure peptides were identified by LCMS of individual fractions and solvent of the pure compound fractions was removed by lyophilization to yield off white powders.

Each pure peptide was analyzed by LCMS and ^{1}HNMR and purities were assessed by ELSD. From the initial forty-four linear peptides, thirty peptides were successfully purified and isolated for use in this study. Most failed peptides resulted from extremely low cyclization efficiencies during the on-resin cyclization. A 10 mM stock solution of each peptide was prepared in DMSO for chromatographic lipophilicity and permeability assays.

2.9.3 Analytical Procedures

Method	Purpose	Instrument	Column	Buffer A	Buffer B	Flow Rate (mL/min)	Gradient Program
1	Peptide Purity Checks (ELSD)	Advion	Phenomenex Kinetex C18, 2.1x30 mm, 2.6 µm, 100 Å	Water (+0.1% Formic Acid)	ACN (+0.1% Formic Acid)	0.4	X-100%B; 5 mins (Wash run - 100%B; 1 min, X%B; 2 mins) X = 20, 30, 40
2	Peptide Purity Checks (HRMS)	Orbitrap	Phenomenex Kinetex C18, 2.1x30 mm, 2.6 µm, 100 Å	Water (+0.1% Formic Acid)	ACN (+0.1% Formic Acid)	0.4	X-100%B; 5 mins (Wash run - 100%B; 1 min, X%B; 2 mins) X = 20, 30, 40
3	Peptide and Building Block Purity Checks	LTQ, Advion	Phenomenex Kinetex C18, 2.1x30 mm, 2.6 µm, 100 Å	Water (+0.1% Formic Acid)	ACN (+0.1% Formic Acid)	0.4	X-100%B; 5 mins, 100%B; 2mins (Wash run - 100%B; 1 min, X%B; 2 mins) X = 20, 30, 40
4	LogK'60 for Off-DNA Peptides	Orbitrap	Hamilton PRP C18, 2.1x33 mm, 5 µm, 130 Å	Water (+0.1% Formic Acid)	ACN (+0.1% Formic Acid)	1	60%B; 5 mins
5	Log $D_{(dd/w)}$, PAMPA, and MDCK	Orbitrap	Thermo Hypersil GOLD C18 2.1x30 m, 1.9 µm	Water (+0.1% Formic Acid)	ACN (+0.1% Formic Acid)	1	5%B; 0.5 mins, 5-95%B; 0.25 mins, 95%B; 2.25 mins, 95-5%B; 0.01 mins, 5% B; 0.5 mins
6	DEL Reaction and Purity Checks	LTQ, Advion, Orbitrap	Xbridge OST BEH C18 4.6x50 mm, 5 µm, 130 Å	8.6 mM TEA, 200 mM HFIP (pH = 8.0)	MeOH	0.8	15-60%B; 7 mins (Wash run - 60%B; 2 mins, 15%B; 3 mins)
7	DEL Lipophilicity - Model Compound Mixtures or Library	Advion	Hamilton PRP C18, 2.1x250 mm, 5 µm, 130 Å	50 mM TEAA (pH = 7.0)	ACN	0.2	Small molecules: 10-40%B; 60 mins. Peptides: 20-55%B; 60 mins. Library: 10-40%B, 60mins.

Table 4: List of HPLC and LCMS methods used in this study

LCMS of Off-DNA Peptides

Peptides were diluted to approximately 0.5 mg/mL in optima grade MeOH. Peptides were analyzed on either a (1) Thermo Finnigan Surveyor HPLC system and LTQ mass spectrometer, an (2) Advion AVANT HPLC-expression CMS system, or a (3) Thermo Scientific Ultimate 3000 UPLC System and Orbitrap VelosPro (FTMS mode).

For purity assessments (**Table 4**, Methods 1-3), peptides were analyzed on a Phenomenex Kinetex C18, 2.1x30 mm, 2.6 µm, 100 Å column using Milli-Q water (+0.1% formic acid) (A) and acetonitrile (+0.1% formic acid) (B) as running buffers. Peptides were analyzed via a 0.4 mL/min gradient elution starting at either 20, 30, or 40%B and ramping to 100%B over 5 minutes. Wash runs were performed between each analysis which consisted of a 1-minute hold at 100%B followed by a 2-minute equilibration hold at the mobile phase composition to be used in the following run. For LCMS to quantify samples for Log $D_{(dd/w)}$, PAMPA, and MDCK (Table 1, Method 5), peptides were analyzed on a (3) Thermo Scientific Ultimate 3000 UPLC System and Orbitrap VelosPro (FTMS mode) using a Thermo Hypersil GOLD C18 2.1x30 mm, 1.9 µm column with Milli-Q water (+0.1% formic acid) (A) and acetonitrile (+0.1% formic acid) (B) as running buffers. A 1 mL/min flow rate was used with the gradient program consisting of a 0.5 min hold at 5%B with mass spec diversion (to divert PBS salts), a 0.25 min ramp from 5%B to 95%B, a 2.25 min hold at 95%B, followed by a drop back to 5%B and a 0.5 min hold at 5%B.

Measurement of Capacity Factors (LogK'60) for Off-DNA Peptides

Peptides were diluted to approximately 0.5 mg/mL in optima grade MeOH. For capacity factor measurements (**Table 4**, Method 4), peptides were analyzed on a (3) Thermo Scientific Ultimate 3000 UPLC System and Orbitrap VelosPro (FTMS mode). A Hamilton PRP C18, 2.1x33 mm, 5 µm, 130 Å column was employed with a 1 mL/min flow rate using Milli-Q water (+0.1% formic acid) (A) and acetonitrile

(+0.1% formic acid) (B) as running buffers. The column was equilibrated at 60%B and samples were run for 5 minutes at 60%B. Retention times were picked by selected ion monitoring and were transformed to capacity factors (LogK' values) using the following equation:

$$(3)\ LogK = Log\left(\frac{T_R - T_0}{T_0}\right)$$

Where T_R is the gradient retention time and T_0 is the dead volume time. The dead volume time was determined by measuring T_R of citric acid. Capacity factors were measured in duplicate.

LCMS of DNA-Bound Compounds and DNA-Encoded Library

DNA analytes were diluted to approximately 1-10 μM in Milli-Q water. LCMS analysis of DNA-bound compounds (model compounds, DNA-encoded library) (**Table 4**, Method 6), was performed on either an a (1) Thermo Finnigan Surveyor HPLC system and LTQ mass spectrometer, an (2) Advion AVANT HPLC-expression CMS system, or a (3) Thermo Scientific Ultimate 3000 UPLC System and Orbitrap VelosPro (FTMS mode). All DNA analytes were analyzed on a Waters XBridge OST BEH C18 column (4.6 x 50 mm). Mobile phase A consisted of 8.6 mM, 200 mM HFIP in water (pH = 8.0) and mobile phase B consisted of Optima grade methanol. Compounds generally were analyzed via a 0.8 mL/min gradient elution starting at 15% B and ramping to 60% B over 6 minutes. Wash runs were performed between each analysis which consisted of a 1-minute hold at 60% B followed by a 2-minute equilibration hold at 15% B.

Measurement of Chromatographic Lipophilicity for DNA-Bound Compounds and DNA-Encoded Library

Determination of chromatographic lipophilicity for DNA compounds (**Table 4**, Method 7) was performed on an (2) Advion AVANT HPLC-expression CMS system. A Hamilton PRP C18, 2.1x250 mm, 5 µm column was used with running buffers consisting of 50 mM TEAA (pH = 7) (A) and either pure acetonitrile or 50:50 Buffer A:ACN (B). Flow was diverted away from the mass spectrometer and, for the DNA encoded library, all outflow was collected into a 96 well plate (see more details below). Compounds were typically eluted under shallow gradient conditions (e.g. 10-40%ACN) over 60 minutes, depending on the lipophilicity of the DNA-bound analytes.

The conditions used for all lipophilic measurements of DNA-bound molecules (model compounds, library elution) were 50 mM TEAA (pH = 7) (A) and ACN (B) with a Hamilton PRP C18 column. These specific conditions produced the best correlations to 1,9-decadiene/water partition coefficients of off-DNA peptides. Compounds were first identified by mass spectrometry using the LCMS-compatible conditions detailed in the previous section (TEA-HFIP buffer, **Table 4**, Method 6) and then were subsequently identified by UV at 260 nm while using the 50 mM TEAA buffer (not compatible with mass spectrometry). The identities of compounds in mixtures were validated by comparison to single compound injections under identical or similar conditions. Retention times were recorded by measuring the

centroid of each compound peak in the UV-Vis spectra and were transformed to LogK' values using the capacity factor equation described in the section above.

Gel Electrophoresis

Agarose gels (5% w/v) were prepared by combining agarose (6 g) with 1X TAE buffer (120 mL) in a 250 mL Erlenmeyer flask. The mixture was heated by microwave irradiation (750 W) until fully solubilized, then allowed to cool before casting. Gels were run in 1X TAE buffer (200 mL) mixed with ethidium bromide (neat, 50 µL). Samples were prepared by combining DNA analyte in water (500 ng, 5 µL) with Tritrack 6x loading buffer from Thermo Fischer Scientific (1 µL). GenerulerUltra Low Range DNA Ladder from Thermo Fischer Scientific was included in each run for comparison. Gels were run at 110V until completion and then visualized by UV using a Biorad Gel Doc XR+.

qPCR Protocol – Checking Library Stocks

qPCR was run after completion of library synthesis to determine the amount of amplifiable DNA. Reaction Master Mix was prepared by combining an aqueous solution of forward primer (6.7 µM, 0.3 µL), reverse primer (10.6 µM, 0.3 µL), Applied Biosystems 2X PowerUP SYBR Green Master Mix (10 µL) and water (7.4 µL). Typically, enough Reaction Master Mix would be prepared for three more reactions than planned to account for transfer loss. qPCR reactions were set up in Applied Biosystems MicroAmp Fast Optical 96-Well Reaction Plates (0.1 mL) by

adding DNA template (2 µL) followed by Reaction Master Mix (18 µL). Each reaction was mixed thoroughly by pipet, sealed with Applied Biosystems MicroAmp Optical Adhesive Film, and then briefly centrifuged at room temperature (18213 x g). qPCR was run in an Applied Biosystems Viia7 qPCR Real-Time qPCR Instrument using ROX for baseline fluorescence. The pre-amplification protocol consisted of a 5-minute hold at 95°C (2.1°C ramp rate). The amplification protocol consisted of 45 rounds of a 10 second hold at 95°C (2.1°C ramp rate), followed by a 20 second hold at 60°C (2.0°C ramp rate), followed by a 10 second hold at 72°C (2.1°C ramp rate). A final cooling stage of a 10 second hold at 40°C was also performed. qPCR standards were included in all runs. All samples were run in duplicate at three different concentrations.

Shake-flask Partition Experiment

The shake-flask partition coefficient was performed as described by Naylor et. al.[24] 1,9-Decadiene and 1X PBS Buffer (pH = 7.4) were saturated by mixing equal volumes of the other. The mixture was vortexed at max speed for 2 minutes and then briefly centrifuged to separate the emulsion. Analytes were diluted to 1 mM in DMSO. Analytes (8 µL) in DMSO were added to 2 mL Eppendorf tubes followed by saturated 1,9-decadiene (800 µL) and saturated PBS buffer (800 µL) to yield a final concentration of 5 µM with 0.5% v/v DMSO. Tubes were sealed with parafilm and mixed by vortex at max speed for 30 minutes. Tubes were then sonicated for 30 minutes and centrifuged for 10 minutes at room temperature (18213 x g). Four

aliquots (150 μL) of each sample were removed from each layer and transferred to a fresh 0.2 mL 96-well plate. Special care was taken to avoid contamination from the pipette tip, especially when collected from the lower aqueous layer. Importantly, air was expelled while passing through the organic layer to prevent capillary wicking of organic solvent to the inside of the pipet tip. Plates containing quadruplicate aliquots of 1,9-decadiene and PBS for each sample were evaporated overnight in a Genevac centrifugal evaporator at 60°C. Each sample was then resuspended in MeOH (150 μL), mixed by pipet, sealed, and sonicated for 15 minutes. After a final centrifugation for 10 minutes at 700 x g, samples were quantified by UPLC-MS by integrating selected-ion monitored peaks for each sample. Briefly, each replicate was injected (10 μL) into a (3) Thermo Scientific Ultimate 3000 UPLC System equipped with an Orbitrap VelosPro (FTMS mode). A Thermo Hypersil GOLD C18 (2.1 x 30 mm, 1.9 μm) column was used with a 1 mL/min flow rate. The running buffers were Milli-Q water (+0.1% formic acid) and acetonitrile (+0.1% formic acid). The gradient program consisted of a 30 second hold at 5% B, followed by fast 15 second gradient ramp from 5-95% B, followed by a 2 minute and 15 second hold at 95% B, and finished with a 30 second hold at 5% B. The partition coefficient Log $D_{(dd/w)}$ was calculated by taking the logarithm (base 10) of the quotient obtained from dividing the average integration from the decadiene layer and the average integration from the PBS layer.

Parallel Artificial Membrane Permeability Assay (PAMPA)

PAMPA was performed as previously described by Naylor et al.[24] Briefly, compound stocks were prepared (1000 µL, 50 µM, 5% DMSO in 1X phosphate-buffer saline (PBS) pH 7.4). Aliquots of each compound stock (150 µL) were added in quadruplicate to a 96-well donor plate with 0.45 µm hydrophobic Immobilon-P membranes (Millipore MAIPNTR10) pre-soaked with a 1% (w/v) solution of lecithin in n-dodecane. The donor plate was loaded into a 96-well Teflon acceptor plate (Millipore MSSACCEPTOR) with 300 µL of 1X PBS with 5% DMSO. After a ~16 h, the acceptor and donor plates were separated, and 50 µL was collect from each well. Samples were diluted with 50 µL methanol and analyzed by LCMS as described in the shake-flask protocol above. Each sample was run with 10 µM carbamazepine as a control. Exact assay run time in seconds was used in the calculations.

Chromatographic Partitioning of DEL, Barcoding Fractions by PCR, and Sequencing

The library was partitioned into fractions via ion-paired reverse phase chromatography under the conditions described above (**Table 4**, Method 7). The library was collected into 96 fractions in a 96 well plate with 0.5 mL round-bottom wells, beginning at 10 minutes with 92 fractions collected in 30 second intervals followed by 4 fractions collected in 1-minute intervals. After full elution of the library, the plate was sealed, frozen under dry ice, and lyophilized overnight. The dried fractions were resuspended in 100 µL of water and a 1:10 dilution of each fraction was quantified by qPCR.

The concentration of each well was assessed via qPCR (LightCycler 480 II). To a 384-white PCR plate, 2 µL of each collected fraction, 5 µL of SYBR Green 1 Master, 0.5 µL of each forward and reverse primer, and 2 µL of water were added for a 10 µL reaction per well. Each fraction was run in duplicate. A set of 6 concentration standards (from 1e4 to 1e9 molecules per µL) with a TE buffer control were added to the plate in duplicate. The plate was sealed with an optically clear adhesive plate seal, and centrifuged (1500 rpm, 1 min). Thermal cycle programming proceeded as follows: 1 cycle of pre-incubation (4.8°C/s, 5 min, 95°C), 45 cycles of amplification (4.8°C/s, 10 sec, 95°C; 2.0°C/s, 20 sec, 60°C; 4.8°C/s, 10 sec, 72°C), 1 cycle melting curve (4.8°C/s, 5sec, 95°C; 2.5°C/s, 60 sec, 65°C; 97°C), final cooling (2.0°C/s, 10 sec, 40°C). A single measurement was taken during each amplification cycle during the 72°C stage. Two analyses were performed using the instrumentation software. Second derivative max with in-run standard curve analysis assessed the success of the run based on slope and efficiency. The concentration for each fraction was assessed with Fit Points Analysis.

PCR and installation of barcoding Illumina adapters was performed by Anagenex. Each fraction was subsequently diluted 1:40 with molecular biology grade water. To 2.5 µL of each sample was added, 5 µL 10x Master Mix, 1 µL 10 mM dNTPs, 1 µL FastStart Taq Polymerase, 1.2 µL of 50 µM forward and reverse primers, 8 µL 25 mM $MgCl_2$, and water to a total volume of 50 µL in each well. Thermal cycle programming proceeded as follows: 4 minutes at 95°C initial denaturation, followed by x cycles of 30 seconds at 95°C for denaturation, 30 seconds

at 63°C annealing, 45 seconds at 75°C for elongation, with one final extension for 7 minutes at 75°C. Progression of PCR was assessed 4% 48-well agarose E-gel.
Each PCR reaction was cleaned via AMPure XP Reagent from Beckman Coulter. Following a 1.1:1 sample to beads ratio, the mixture was incubated at RT for 5 minutes and then captured on a 96-well magnetic plate. After 3 washes with 70% EtOH, the beads were air dried for 5 minutes and then resuspended in molecular biology grade water. Qubit 1X dsDNA High Sensitivity Assay reagent and a Qubit 4 were used to assess the concentration of each cleaned sample.

Prior to NGS, a portion of each sample was diluted to 2 nM. A pool was created with an equal volume of each 2 nM sample. After dilution of PhiX to the same concentration as the pool, 0.020 pmol of the pool and PhiX were denatured using an equivolume of fresh 0.1 M NaOH per Illumina protocol. The denatured pool and PhiX were combined at a 4:1 ratio and loaded onto a 150 cycle MiniSeq High Output Reagent Cartridge. After addition of custom indexes and primers, the MiniSeq was run with 120 read 1 cycles, 8 index 1 cycles in a single-end run setting.
The 2 nM pool was diluted to 600 pM in provided RSB+Tween-20 with PhiX similarly diluted. The pool and PhiX were combined at a 85:15% ratio. The custom index 1 (3.6 uL at 100 µM) and read 1 (1.8 µL at 100 µM) primers were combined with HT1 buffer to 600 µL total. Th2 pool:PhiX mixture, index 1 primer mixture, and read 1 primer mixture were each added to the reagent cartridge after thawing per manufacturer's protocol. The sequencing run was completed on a P3 100 cycle reagent kit with 120 read 1 cycles, 8 index 1 cycles in a single-end run setting.

The fastq files produced for each of the sequencing runs were analyzed using Anagenex's sequence translation and quantification software, Seq2Struct. Seq2Struct assigns each sequence to its source experiment and condition, then it converts these sequences to the structures they represent and reports unique binding events for each of those structures. These experiments were designed so that the most appropriate metric to use for analysis would be the direct comparison of these absolute counts of unique binding events for each molecule across fractions, so no further normalization nor additional score methods were applied to these results. Sequences were organized based on their corresponding fraction and each sequence was deconvoluted to its corresponding peptide. The sequencing counts and deduplicated sequencing counts for each encoded library member in each fraction were summed. These data were then organized into a .csv where each encoded library member has associated sequencing counts and deduplicated sequencing counts for each fraction. Fractions were expressed numerically as their median retention time in seconds.

2.9.4 Computational Procedures

General

UV-Vis chromatographic data and qPCR chromatograms were plotted in Graphpad Prism. Linear regression analysis was performed in Graphpad Prism. Fraction count summations were performed in Python using pandas and numpy. Box and Whisker plots were generated in Python using Seaborn.

Library Enumeration

Library enumeration was performed by Anagenex. Peptide structures were assembled by combining the SMILES strings for each building block in each cycle of chemistry. The resulting linear and cyclized peptide structures were organized in a .csv with the associated sequencing data. Physicochemical data for each compound (exact mass, AlogP, number of HBA and HBD, etc.) was calculated from SMILES string using RDKit. Each compound was assigned a library identifier.

Generation of Sequencing Count Chromatograms

Sequencing count chromatograms for each DNA-encoded library member were generated with matplotlib.pyplot. From the NextSeq data, a delimited dataset for each compound was obtained which contained the compound's sequencing count and deduplicated sequencing count in each fraction in the form of time1:count;deduplicated count, time2:count;deduplicated count, etc. (e.g. 615:200;180, 645:180;160,…). From this dataset, the times and deduplicated count were extracted and sorted by time. The time was converted from seconds to minutes. For each compound, the deduplicated count was plotted against the time in minutes to generate sequence count chromatograms (Figure 3c). Both linear and cyclized forms of each compound were typically coplotted.

Simple Gaussian Peak Picking Model (SGPPM)

The Simple Gaussian Peak Picking Model (SGPPM) was developed initially to identify peaks from sequencing count chromatograms which potentially represent the product of interest. The model processes chromatogram data in the form of sequencing counts vs time and attempts to identify the peak corresponding to the compound of interest. The signal baseline is first corrected by calculating a sliding window minimum baseline and subtracting this baseline from the original signal, while setting any resulting negative values to zero. Next, peaks are identified using find_peaks from scipy.signal using a minimum height threshold. Peak widths are then calculated using the peak_widths function from the scipy.signal module and each peak is fit to a Gaussian curve using the following function:

$$(4)\ \ g(t) = y_{corr}(t_h)\ \cdot \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$$

where:

$y_{corr}(t_h)$ is the corrected signal height at the peak

μ is the mean (representing the peak position in minutes)

σ is the standard deviation (related to the peak width)

The curve fitting was performed using the curve_fit function from scipy.optimize. Centroid retention times are then determined from the gaussian fits of each peak. Finally, the peak most likely to correspond to full encoded product is selected based on multiple criteria:

1. Peak height must exceed an absolute sequencing count threshold

2. Peak height must exceed a relative threshold (30% of the maximum signal of the chromatogram in our implementation)
3. The fitted Gaussian's σ must be below a threshold (σ = 3 in our implementation)
4. Select the peak which passes 1-3 and has the largest centroid retention time.

This model manages to effectively capture product peaks for many compounds in this DNA-encoded library, although it cannot differentiate between true product peaks and those belonging to a synthetic truncation, which spurned us to develop the Classic Chrome model which is informed by associated Null compound chromatograms (see below). Additionally, this technique can be used to identify compounds which did not properly cyclize by direct comparison to the sequencing count chromatogram of the matching linear compound. One major source of inaccurate peak picking from this model is major, late-eluting byproducts from cyclization (see discussion above).

Additionally, it is important to note that selection of the latest eluting major peak which passes the aforementioned criteria is reasonable for this DNA-encoded library specifically because successful couplings consistently increase retention time due to the aliphatic nature of the building blocks. It is unclear if this peak-selecting strategy would be viable if the library building blocks consisted of much more polar residues which may cause a decrease in retention time upon successful coupling.

Classic Chrome – Null Informed Model

The Classic Chrome model was developed to identify and analyze product peaks from sequencing count chromatograms in DNA-encoded libraries which have built in null building block truncations. This method also processes chromatogram data in the form of sequencing counts vs. time and attempts to identify the peak corresponding to the compound of interest. Classic Chrome's approach builds upon the SGPPM by incorporating more sophisticated analysis techniques and a hierarchical examination of compound truncations. Classic Chrome utilizes a peak detection algorithm that smooths the chromatogram, applies baseline correction, and performs adaptive multi-pass peak detection. Each detected peak is scored based on its prominence, width, area, symmetry, and retention time.

A key feature of Classic Chrome is its truncation analysis, which examines the respective double truncations (having two null building blocks) and single truncations (have one null building block) of a given compound to understand the contributions of individual building blocks to the overall retention time. The method proceeds through several stages of analysis, starting with the full null compound to establish a baseline, then analyzing double and single truncations before examining the full compound. This hierarchical approach allows for a comprehensive understanding of how each building block contributes to the final compound's chromatographic behavior. The method then combines information from peak detection and truncation analysis to formulate a hypothesis about the success of the full compound synthesis and individual coupling reactions.

One of the major advantages of Classic Chrome is its ability to differentiate between true product peaks and those belonging to synthetic truncations by incorporating information from associated truncation chromatograms. It is also more robust in handling compounds that did not properly cyclize. The method's effectiveness is particularly suited, like SGPPM, for libraries constructed with primarily aliphatic residues. Its applicability to libraries with more polar residues would require further investigation. Classic Chrome method represents a significant improvement in the chromatogram peak picking over SGPPM, offering a more comprehensive and nuanced approach to peak identification and compound synthesis verification. Two key limitations with the Classic Chrome model are the requirement for good chromatographic resolution of null-truncated library members, which limits the gradient program employed during chromatography, and the inherent assumption of the model that addition of building blocks increases retention time, which is only true for libraries of purely aliphatic building blocks. A more sophisticated peak detection and prediction algorithm will need to be developed to handle building blocks which may produce difficult-to-predict changes to retention time.

Split-Tree Null Analysis

The Split-Tree Analysis (**Figure 10d**) was developed as a complementary approach to analyze and visualize the hierarchical structure of compound synthesis in peptide libraries. This method focuses on organizing and representing the combinatorial nature of the library by taking into account the success or failure of

individual coupling reactions. The Split-Tree Analysis builds upon the results of chromatogram peak picking methods like SGPPM and Classic Chrome, providing a higher-level view of the library's composition and synthesis outcomes.

At its core, the Split-Tree Analysis creates a tree-like structure (an arborescence based on a Golden Fractal layout algorithm) that represents the building blocks and their combinations in the library. The root of the tree represents the null compound, with subsequent levels corresponding to the addition of each building block. The model uses data from successful and unsuccessful coupling reactions, as determined prior by product peak picking methods, to construct this tree. A custom script for this analysis was written using the Python library NetworkX.

The Split-Tree Analysis offers several advantages in understanding and optimizing DNA-encoded libraries. A key feature of the analysis is its ability to "prune" branches of the tree that represent failed reactions or undesirable compounds, allowing for a clear visualization of the successfully synthesized portion of the library. It provides a comprehensive view of the library's composition, highlighting successful building block combinations and revealing patterns in synthesis failures. This visualization can help identify problematic building blocks or combinations that consistently lead to failed syntheses. Additionally, the model supports the export of various datasets, including lists of verified compounds, pruned compounds, and detailed statistics on synthesis success rates at different levels of the tree. These outputs are valuable for refining library design, optimizing synthesis conditions, and selecting the most promising compounds for further study.

Appendix A: Supplementary Figures and Discussion for Chapter 1

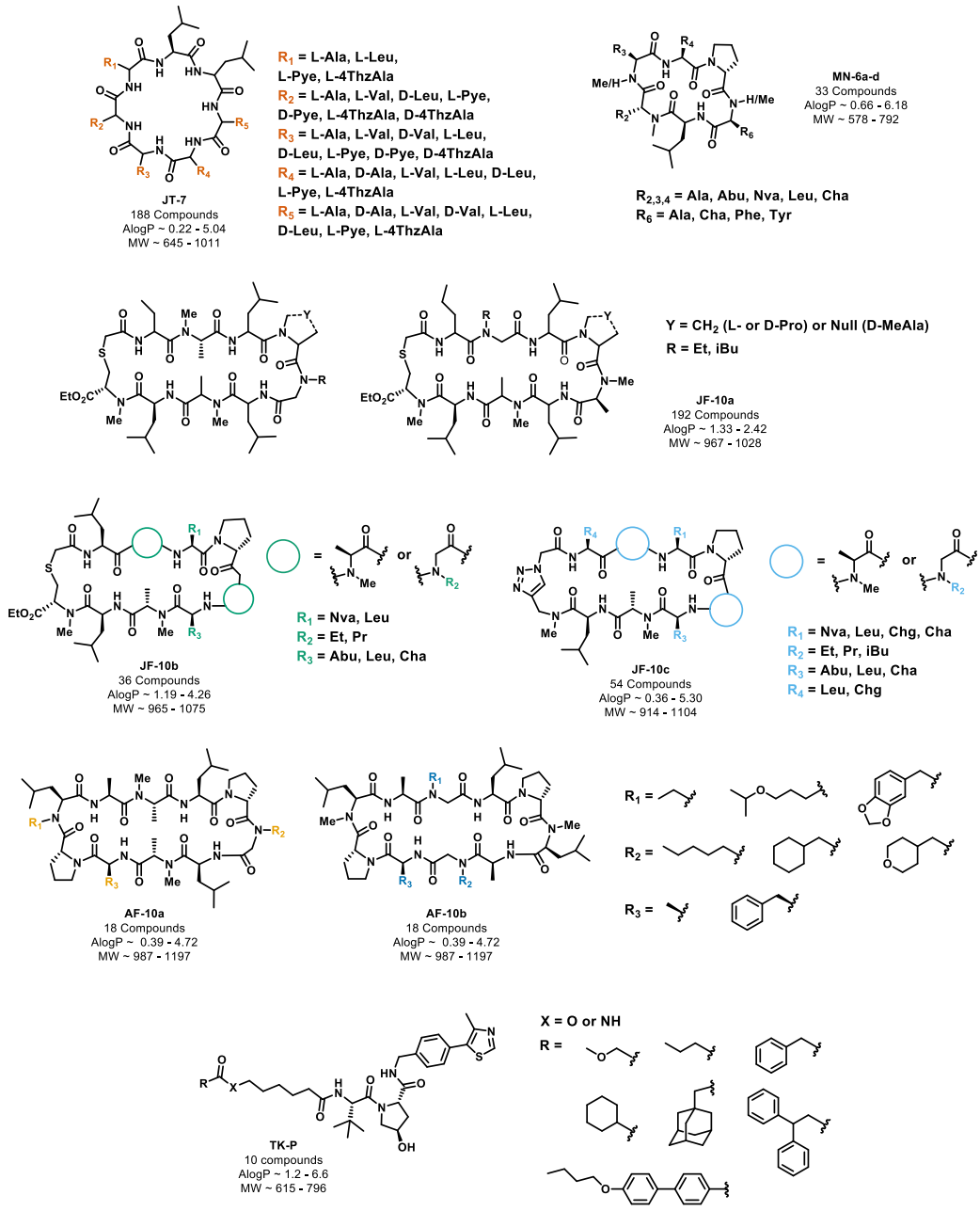


Figure S1: Summary of all compound libraries used in this study. Individual compound information can be found in the associated spreadsheets included in the supporting information. Colored dots correspond to the coloration implemented in **Figure 1** and **4** found in the main text. Undefined stereocenters signifies the use of both diastereomers at that position.

Seven different libraries of compounds were used in this study (**Figure S1**). Five of the seven libraries have been featured in publications from our lab[24, 26, 30, 33]. Two libraries (JF-10c and JT-7) have not been disclosed previously in the literature and additional details about their synthesis can be found below.

JT-7, JF-10c, and AF-10a-b[30] were utilized as a training and test set to build the regression models to predict Log $D_{dd/w}$ values (Log $ED_{dd/w}$) from LogK'60 values due to their diverse properties and high-quality shake-flask data (**Figure 1, Figure S2-4**). MN-6a-d[24] and TK-P[26] were utilized to assess the ability of chromatographic lipophilicity to distinguish between compounds with established differences in LPE due to known differences in physical properties (e.g. hydrogen bond donors). The MN-6a-d libraries specifically facilitated the determination of a new LPE equation determined by chromatography, cLPE (**Figure 2, Figure S7**). MN-6a-d, JF-10b-c, and AF-10a-b were used to assess the consistency of m_{lipo} for Log $ED_{dd/w}$ across diverse peptide scaffolds (**Table S2, Figure S10**) due to their library compositions consisting of well-defined liposcan scaffolds featuring diverse side chains. JF-10a-c and AF-10a-b were utilized to assess the relationship between PAMPA LogPe and Log $ED_{dd/w}$ (**Figure 4**) due to their high-quality PAMPA data and diverse structures. JF-10a was primarily used to investigate the relationship between Log $ED_{dd/w}$, cLPE, PAMPA LogPe, and MDCK LogPe (**Figure 3**). Although this library has been studied extensively in our lab and is featured in a recent publication,[33] the library

trends between lipophilicity and permeability were not previously investigated due to issues with oxidation in the shake-flask assay.

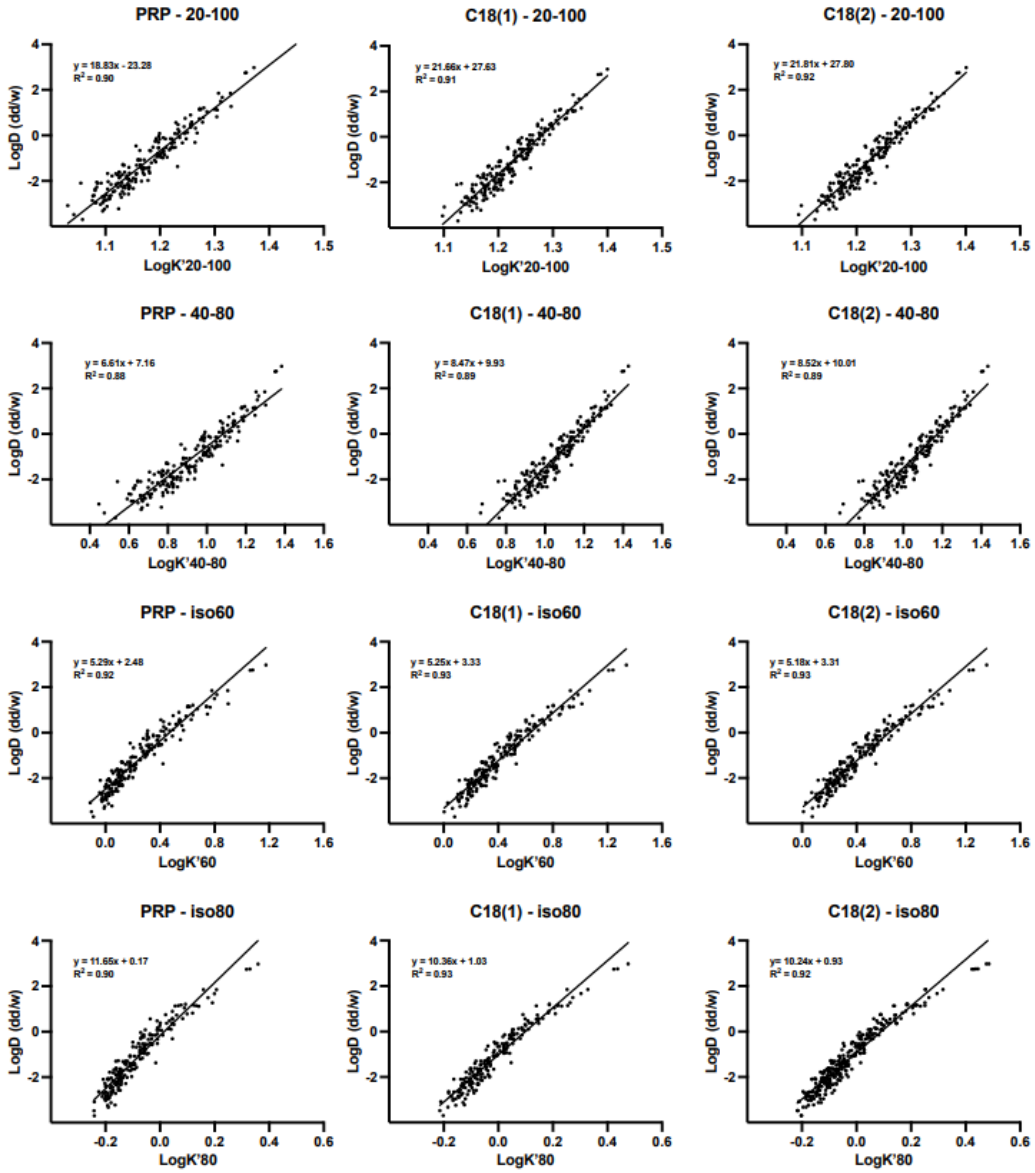


Figure S2: Comparison of columns and methods for capacity factor measurements (LogK) which reflect decadiene-water partition coefficient values for the JT-7 library. Method and column descriptions are listed above each plot. The linear regression equations and coefficients of determination are displayed directly on each plot.

Initially, we sought to investigate a variety of columns and methods to determine which methods would both produce LogK values which are predictive of measured decadiene-water partition coefficients and be amenable to multiplexing with complex mass-encoded mixtures. Our initial investigation consisted of two gradient methods and two isocratic methods across three different columns (backed with either PRP-C18 or silica-C18) and was limited only to compounds from the JT-7 library (188 heptapeptides with diverse sidechains and high variation in backbone stereochemistry). In general, all methods used produced a strong linear correlation between capacity factor and partition coefficients (**Figure S2**, **Table S1**). We noticed a slight curvature to the data, specifically in the isocratic methods. The observed nonlinearity of the data prompted us to expand our dataset beyond the JT-7 heptapeptides and to evaluate the predictive performance of non-linear regression fits to the data.

Table S1 Linear regression analyses of Log $D_{dd/w}$ and LogK using several chromatographic conditions and columns.

	LogK'60			LogK'80			LogK'20-100			LogK'40-80		
	PRP[a]	C18(1)[b]	C18(2)[c]	PRP[a]	C18(1)[b]	C18(2)[c]	PRP[a]	C18(1)[b]	C18(2)[c]	PRP[a]	C18(1)[b]	C18(2)[c]
R^2	0.916	0.934	0.930	0.900	0.928	0.925	0.900	0.911	0.919	0.882	0.891	0.893
m	0.173	0.178	0.180	0.077	0.090	0.092	0.048	0.042	0.042	0.133	0.105	0.105
b	0.451	0.621	0.624	0.005	0.091	0.080	1.230	1.271	1.270	1.064	1.158	1.162

[a]Hamilton™ PRP-C18 5 µm 2.1 x 250 mm [b]Acclaim™ RSLC C18 2.2 µm 2.1 x 250 mm
[c]Acclaim™ Vanquish RSLC C18 2.2 µm 2.1 x 250 mm

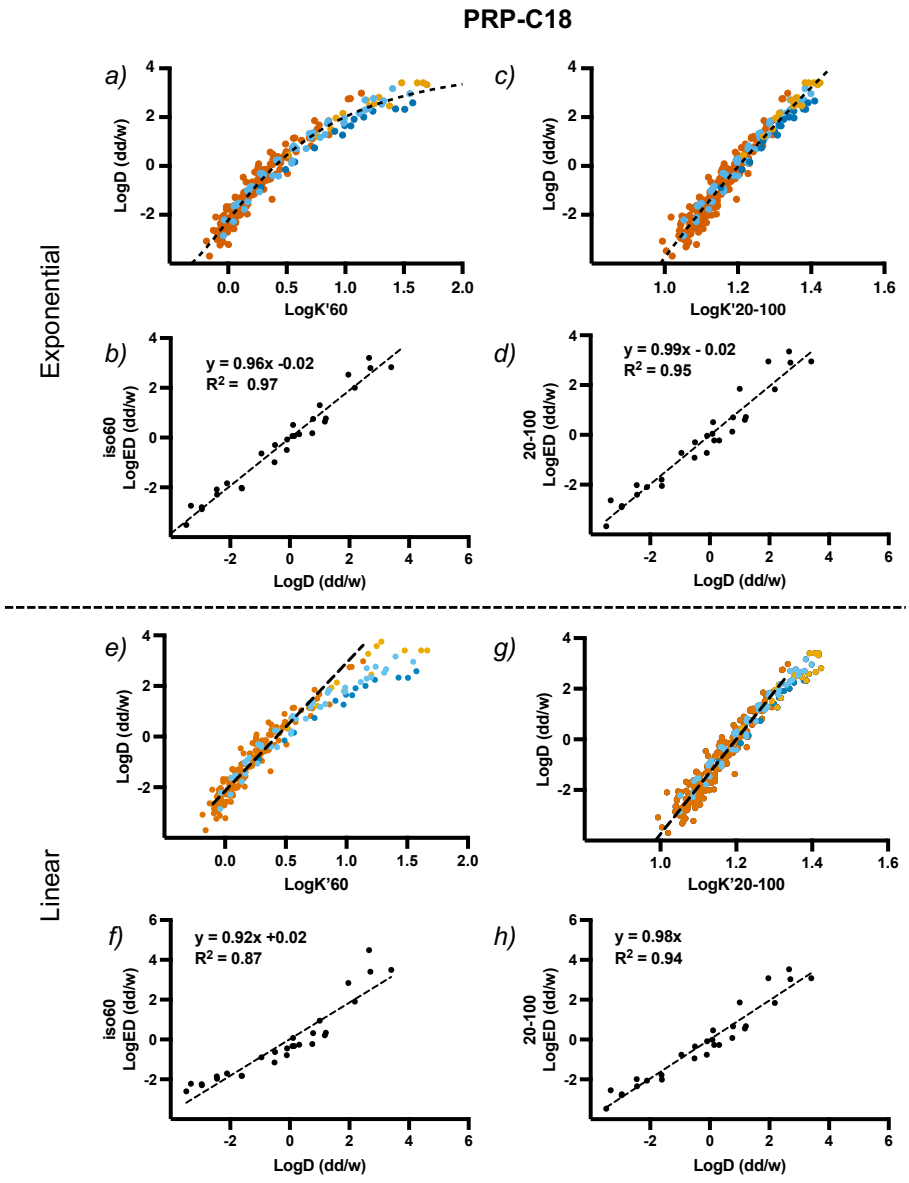


Figure S3: Comparison of linear and exponential regression model fits for a long gradient and an isocratic method on a Hamilton 2.1x250 mm PRP-C18 column. A) Exponential regression fit for isocratic 60%B. B) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the isocratic 60%B method. C) Exponential regression fit for 20-100%B. D) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the 20-100%B method. E) Linear regression fit for isocratic 60%B. F) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the isocratic 60%B method. G) Linear regression fit for 20-100%B. H) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the 20-100%B method.

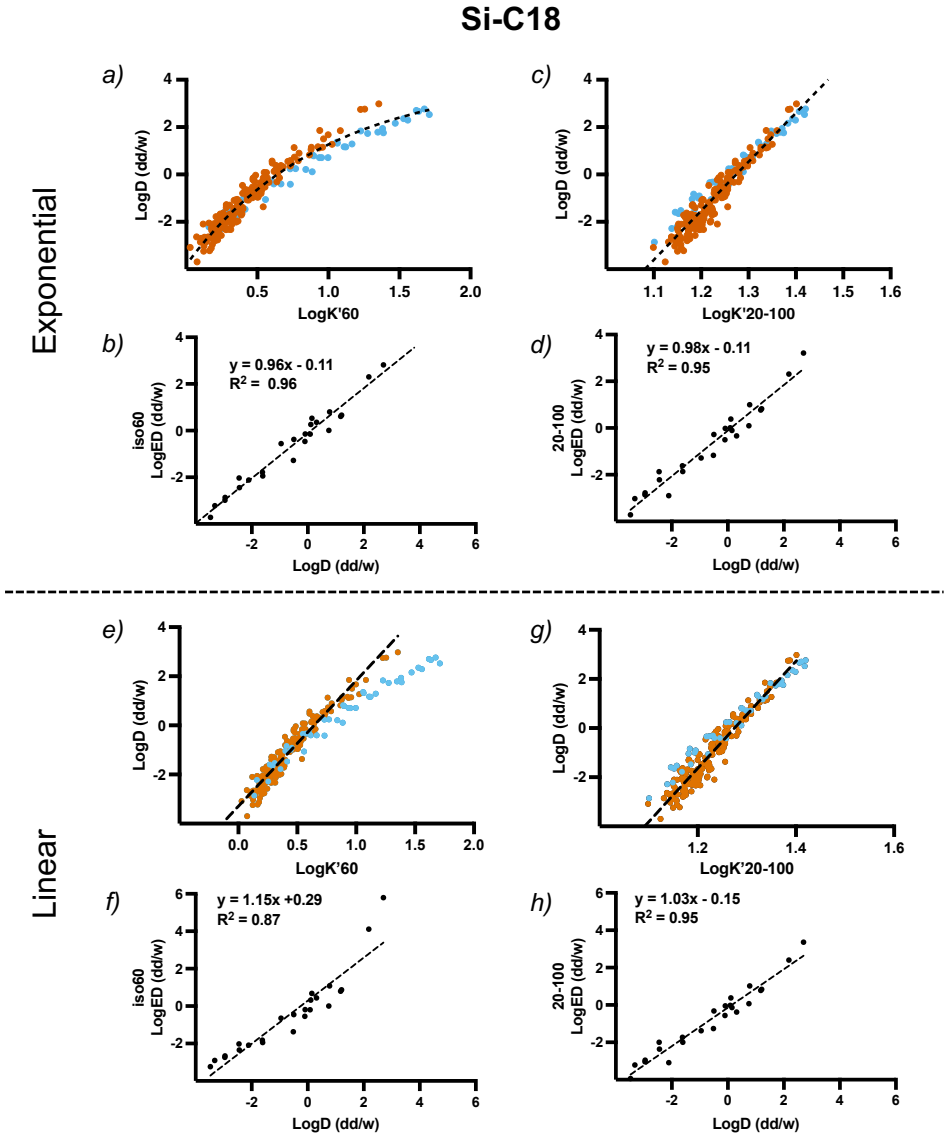


Figure S4: Comparison of linear and exponential regression model fits for a long gradient and an isocratic method on an Acclaim™ Vanquish RSLC C18 2.2 µm 2.1 x 250 mm column. A) Exponential regression fit for isocratic 60%B. B) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the isocratic 60%B method. C) Exponential regression fit for 20-100%B. D) Comparison of predicted vs experimental shake-flask values using the exponential regression equation from the 20-100%B method. E) Linear regression fit for isocratic 60%B. F) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the isocratic 60%B method. G) Linear regression fit for 20-100%B. H) Comparison of predicted vs experimental shake-flask values using the linear regression equation from the 20-100%B method.

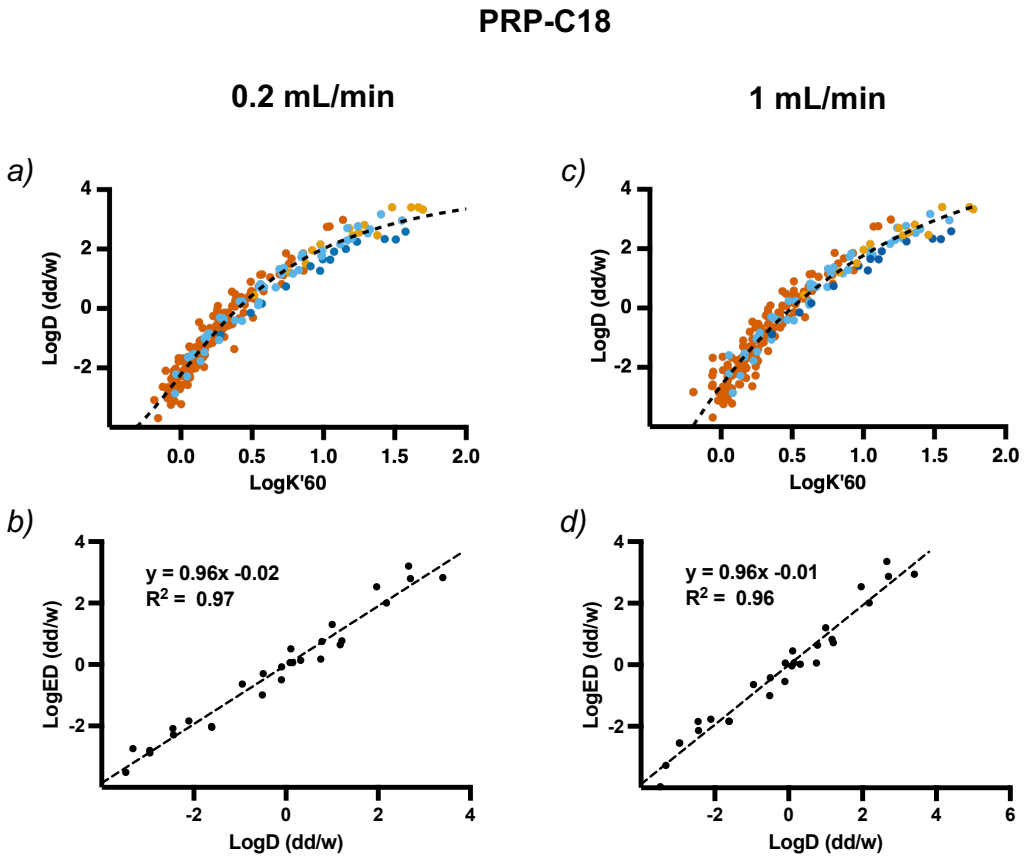


Figure S5: Comparison of exponential regression model fits for low flow rate (0.2 mL/min) and high flow rate (1.0 mL/min) on a Hamilton 2.1x250mm PRP-C18 column using an isocratic method. A) Exponential regression fit for isocratic 60%B with a 0.2 mL/min flow rate. B) Comparison of predicted vs experimental shake-flask values using low flow rate method. C) Exponential regression fit for isocratic 60%B with a 1.0 mL/min flow rate D) Comparison of predicted vs experimental shake-flask values using the high flow rate method.

We assessed the accuracy of linear and exponential regression fits between Log $D_{dd/w}$ and LogK from various chromatographic methods (**Figure S3-S5**). Linear models were fit with a simple y = mx + b equation whereas exponential models were fit using Igor Pro 8 (v8.04) with the 'exp_XOffset' function.

All columns and methods tested resulted in similar Log $ED_{dd/w}$ values, with the exponential fits generally resulting in better Log $D_{dd/w}$ predictions, especially for the

isocratic methods. All compounds behaved similarly, although data was not obtained for the AF-a-b libraries on the SiC18 column.

Encouragingly, higher flow rates did not compromise quality of the lipophilic correlation, which facilitates much faster analysis times for compounds with sufficient lipophilicity.

All equations for linear and exponential regressions for all columns and methods can be found above (**Table 3**).

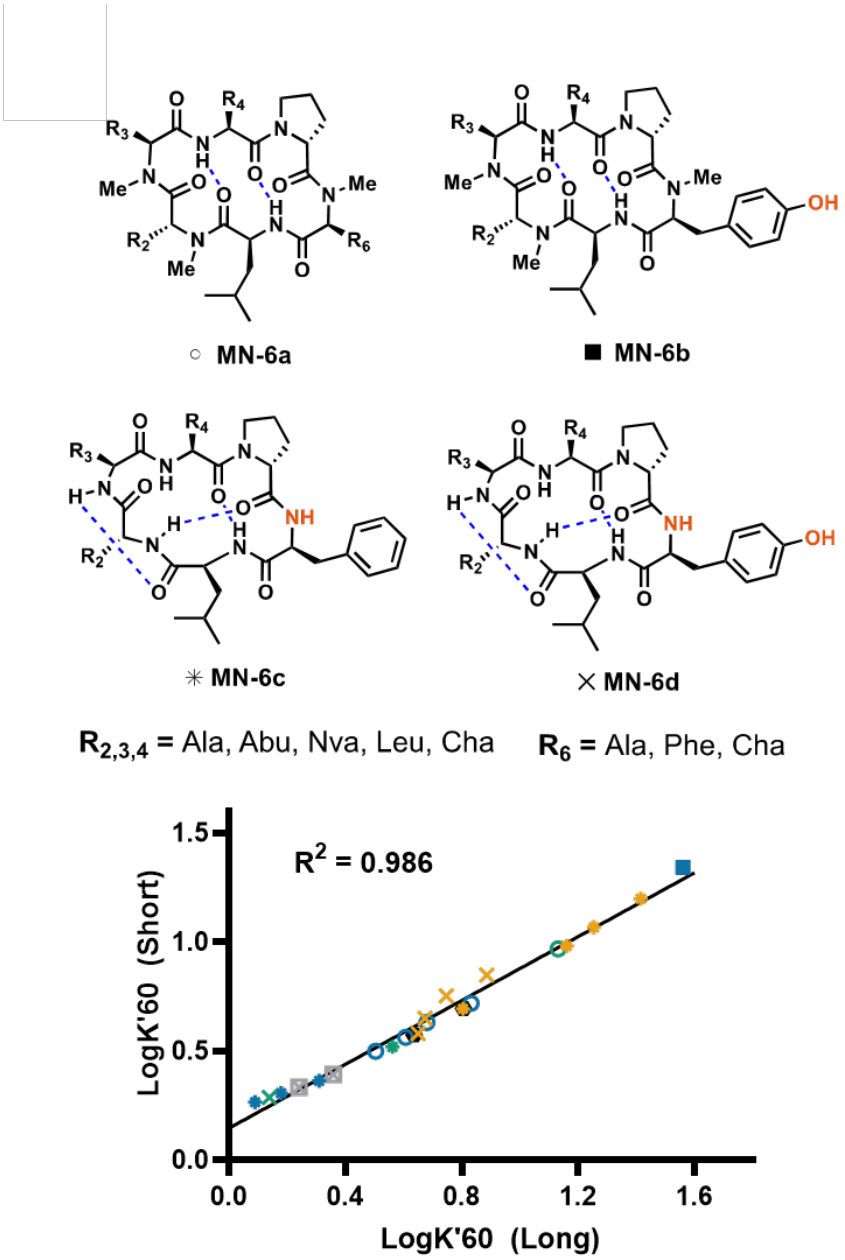


Figure S6: Comparison of capacity factors for MN-6a-d compounds measured on either the long Hamilton PRP-C18 column (2.1x250 mm) with compound mixtures or on a shorter but otherwise identical Hamilton PRP-C18 column (2.1x33 mm) with single compound injections. Open circles have no exposed HBD (class A), squares have one side-chain HBD (class B), stars have one backbone HBD (class C), and X's have two exposed HBDs (class D). Compounds are colored by ALogP (<3, blue; 3-4, green; >4 orange).

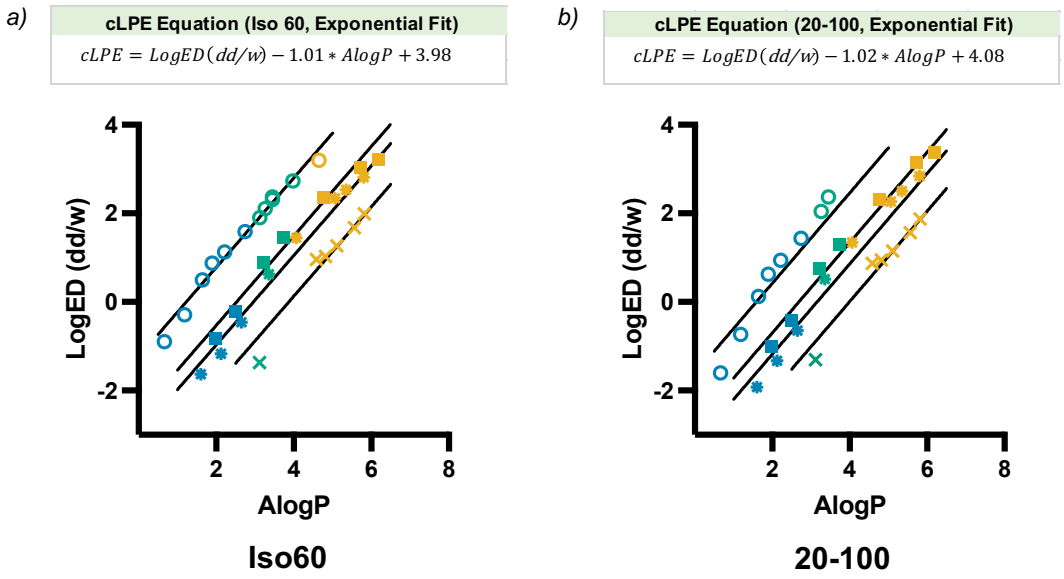


Figure S7: Comparison of Log $ED_{dd/w}$ derived from PRP-C18 a) LogK'60 or b) LogK'20-100 values and AlogP for library MN-6a-d. cLPE equations derived from the minimized slopes and intercepts are displayed above the plots.

Chromatographic analysis of MN-6a-d produced comparable cLPE equations independent of the chromatographic method (**Figure S7**). The similarity of the resultant cLPE equation corroborates the notion that Log $ED_{dd/w}$ values derived from different chromatographic methods are comparable if proper regression models are constructed from accurate shake-flask and chromatographic measurements. One notable difference between these data is the lack of high Log $ED_{dd/w}$ datapoints in the 20-100 dataset. These gradient methods were run over 60 minutes and the high Log $ED_{dd/w}$ did not elute by the end of the run. One benefit to the isocratic method is the ability to extend analysis run times to capture more lipophilic members. However, highly retained compounds in isocratic runs tended to suffer from extensive peak broadening.

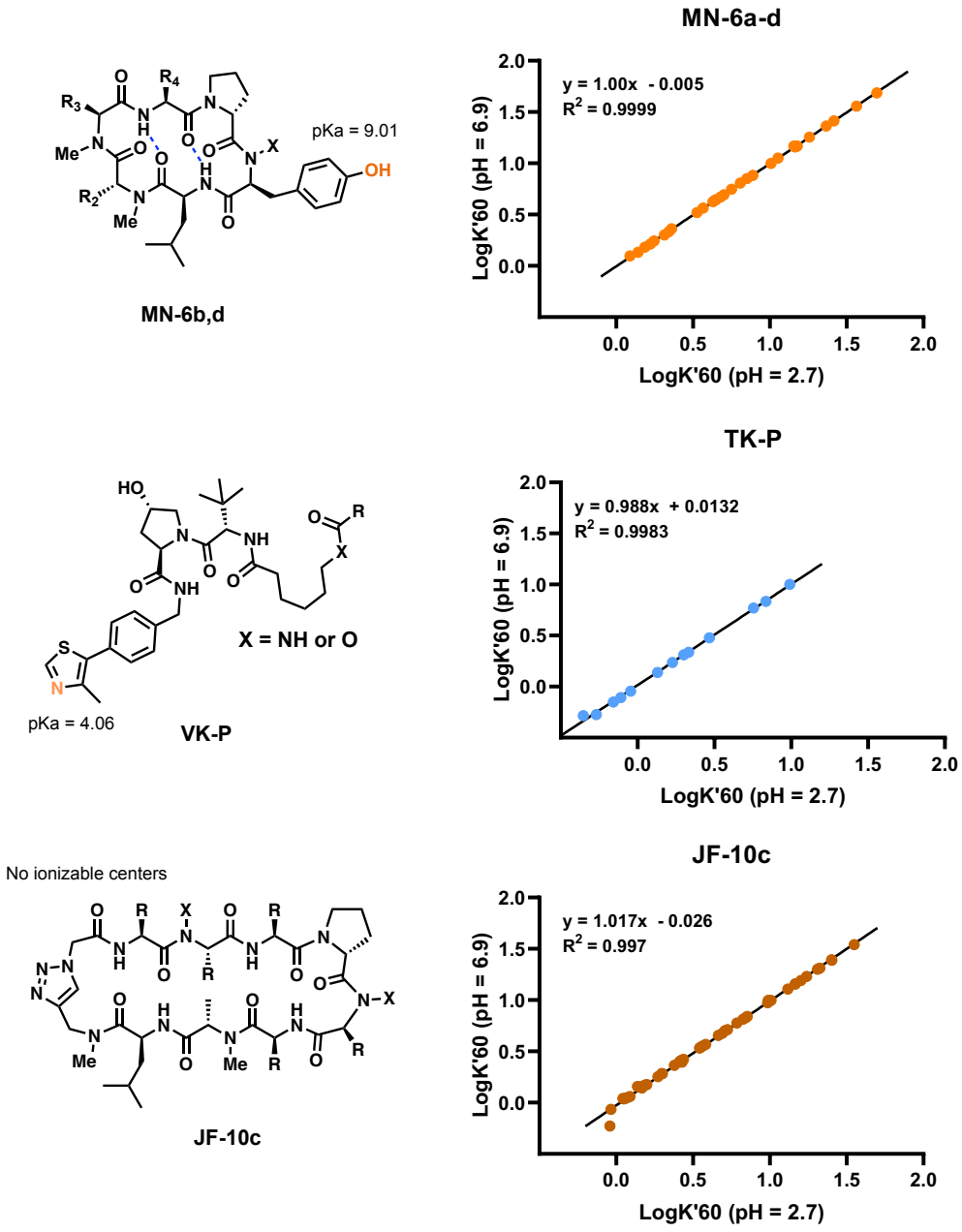


Figure S8: Comparison of LogK'60 for acidic and neutral pH in the aqueous phase. Structures of libraries, with ionizable centers if applicable, are displayed directly to the left of the plots.

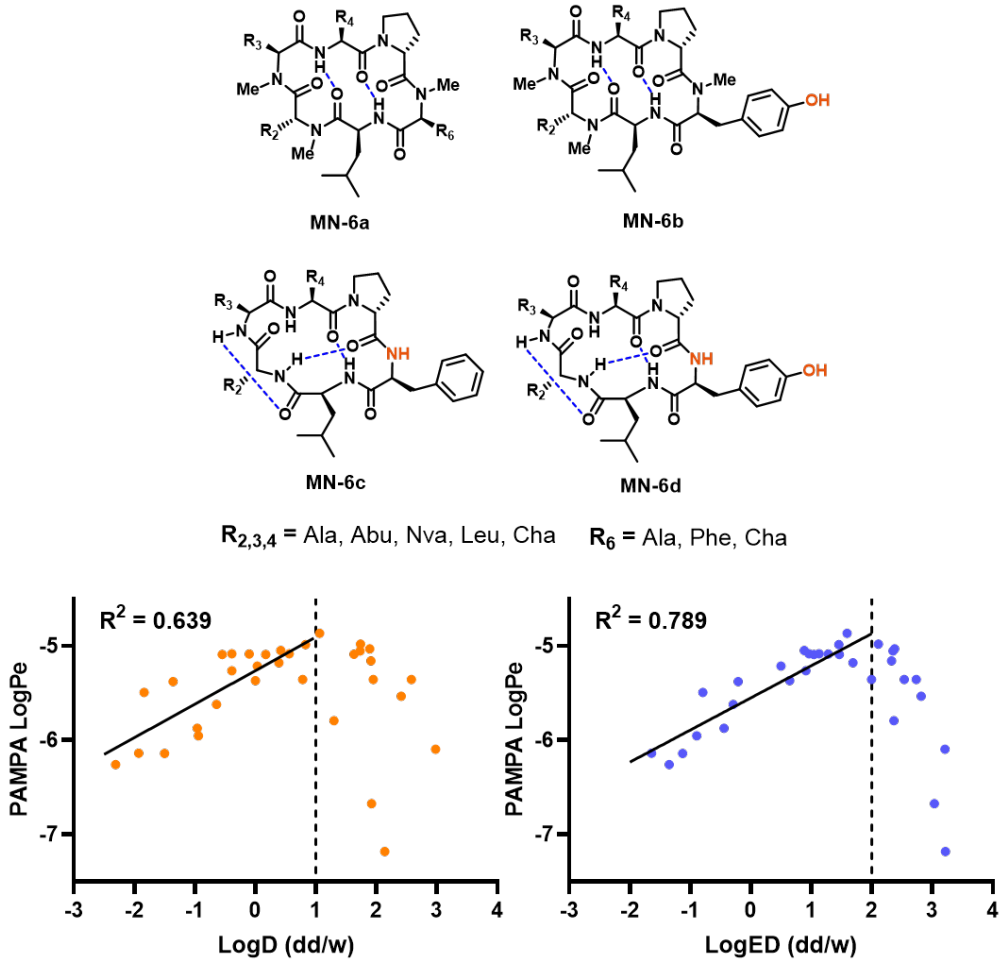


Figure S9: Comparison of the correlation between PAMPA and Log $D_{dd/w}$ (orange) or Log $ED_{dd/w}$ (blue) for the MN-6a-d hexapeptides. Observed solubility cutoff indicated with a dotted black line (Log $D_{dd/w}$ = 1, Log $ED_{dd/w}$ = 2). Linear regressions are plotted with solid black lines for compounds within the observed soluble regime. Coefficients of determination for the linear regressions displayed directly on each plot.

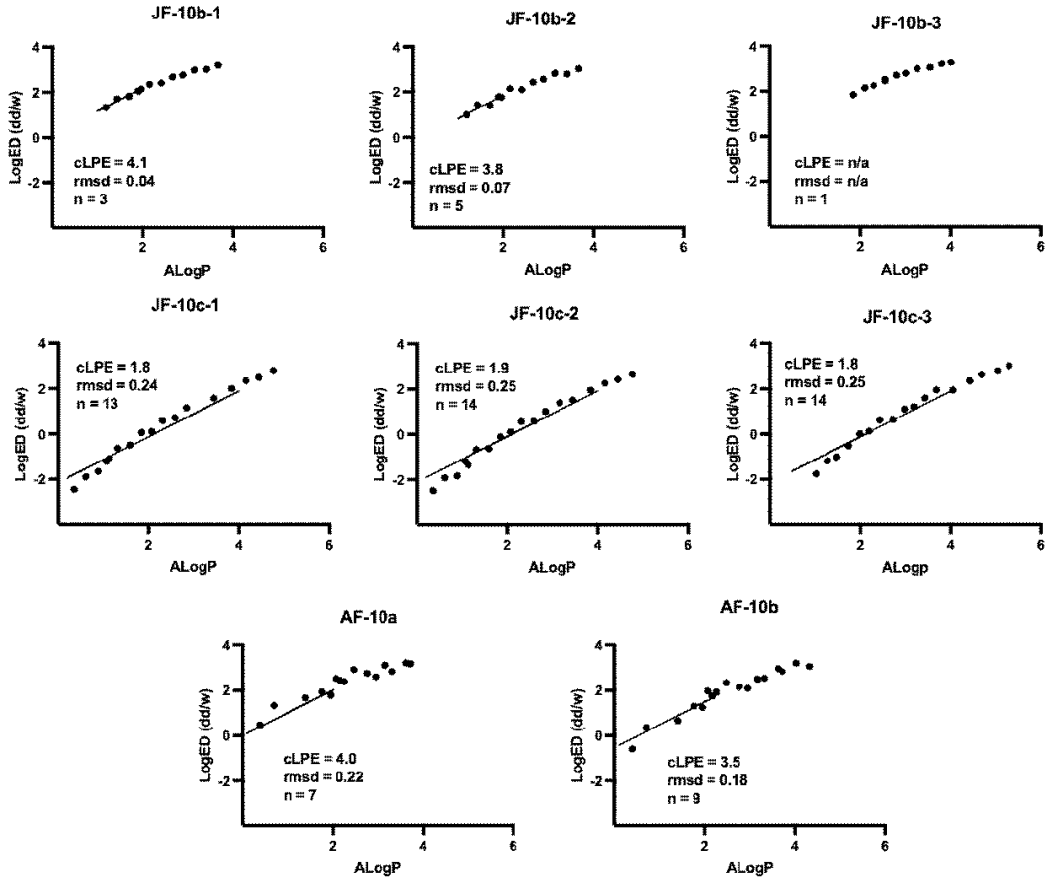


Figure S10: cLPE analysis of individual liposcan libraries for Log $ED_{dd/w}$ measured on PRP-C18 with an isocratic method. Each linear correlation is defined by Log $ED_{dd/w}$ = 1.01*ALogP – b_{avg}, for compounds with Log $\underline{ED_{dd/w}}$ < 2. RMSD between measured values and projected y-values derived from the linear correlation are reported. cLPEs are directly calculated by subtracting the y-intercept of MN6-d from the y-intercept of each liposcan scaffold (cLPE = $y_{scaffold} - y_{MN6\text{-}d}$).

To determine if the cLPE equation derived from the MN6a-d compounds and our model can be fairly applied across diverse peptide scaffolds, we assessed the quality of the linear regression fit across 8 unique peptide scaffolds via liposcans (**Figure S10**). Increases in measured lipophilicity (LogK/Log $D_{dd/w}$) due to simple modulation of aliphatic peptide side chains (i.e. liposcans) should be reflected by consistent increases in calculated lipophilicity (ALogP, CLogP) regardless of the

geometry of the backbone (assuming those side chains don't directly affect the overall peptide conformation or availability of exposed HBDs). A consistent linear relationship between AlogP and Log $ED_{dd/w}$ is necessary to accurately compare cLPE between different scaffolds. To assess this, we calculated the RMSD between Log $ED_{dd/w}$ projected from the cLPE equation to their measured values.

Briefly, from the hexapeptide model (MN-6a-d), the slope-relationship between ALogP and Log $ED_{dd/w}$ (m_{lipo}) was calculated to be 1.01 (**Figure 2b**, **Figure S7**). This slope was used to calculate the y-intercepts using ALogP and observed Log $ED_{dd/w}$ ($b = \text{Log } ED_{dd/w} - 1.01*\text{AlogP}$) for compounds with Log $ED_{dd/w} < 2$. These values were averaged (b_{avg}) and used to generate cLPE values for each liposcan class ($cLPE = b_{avg} - b_{MN\text{-}6d}$) (**Figure S7**) and predicted Log $ED_{dd/w}$. The RMSD was then calculated between the predicted and measured values. This process allows for the assessment of the fit of the generalized cLPE m_{lipo} (1.01) to various liposcan scaffolds where lower RMSDs indicate a better fit.

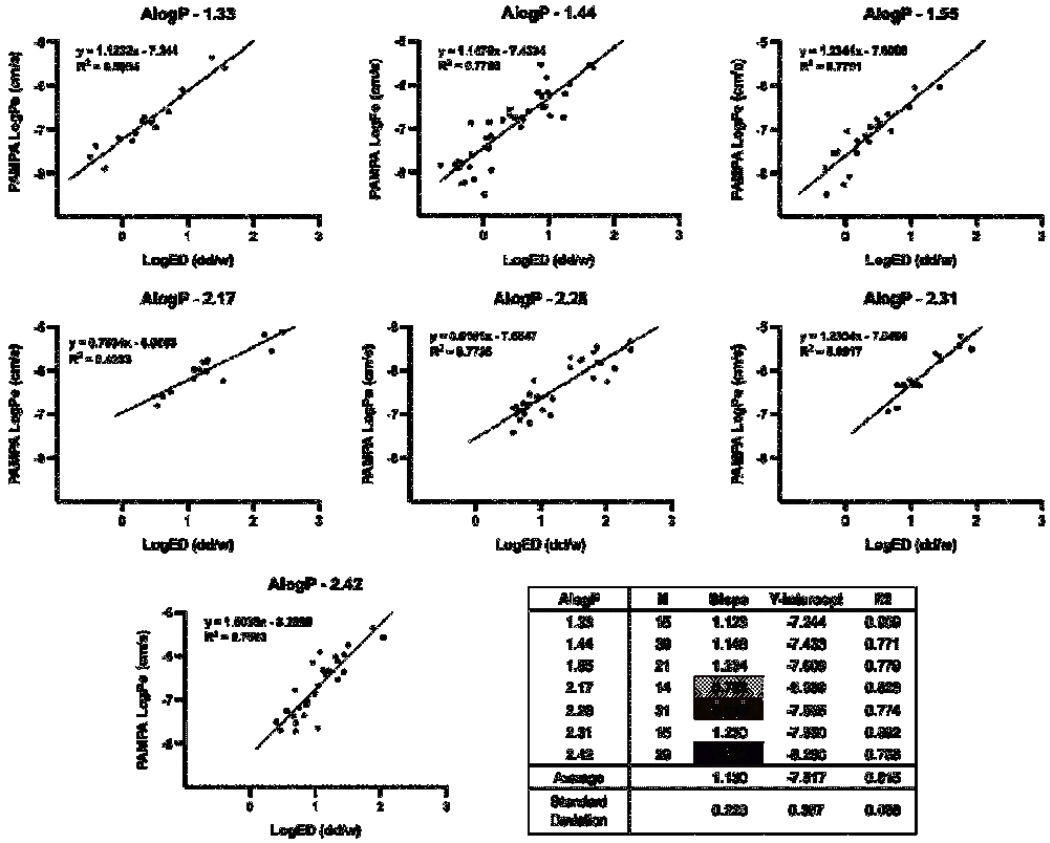
The m_{lipo} values of each dataset are a useful metric to assess the generalizability of the cLPE equation across diverse peptide scaffolds (**Table S2**). Ideally, all Log $ED_{dd/w}$ vs ALogP slope relationships should be very close to the minimized value from the equation ($m_{lipo} = 1.01$). Calculations of cLPE become less reliable as the deviation from the minimized m_{lipo} becomes more significant. The observed relationship between Log $ED_{dd/w}$ and ALogP deviates more significantly for compounds at the extreme ends of lipophilicity. Notably, the JF-10b-1-3 libraries have significantly lower m_{lipo} than any other liposcan scaffold. These scaffolds are

relatively much more lipophilic (cLPE ~4), with most compounds having Log $ED_{dd/w}$ values greater than 2. In the shake-flask assay, this observed flattening of the correlation that occurs at high AlogP is traditionally due to limits of detection in the aqueous layer. However, the observed non-linearity at high lipophilicity in our chromatographic system is likely due to the exponential fit of our non-linear regression equation, which attenuates Log $ED_{dd/w}$ predictions for high capacity factor measurements, although all the shake-flask partition coefficient data included in our model were well above the limit of detection in the aqueous layer. The measured m_{lipo} of the JF-10b libraries would likely coalesce towards the minimized value (1.01) if the polarity of the compounds were extended to capture compounds with Log $ED_{dd/w}$ well below 2.

Table S2: Comparison of m_{lipo} between shake-flask and chromatographic lipophilicity measurements for liposcans of 12 different peptide scaffolds and PROTACs

Chemist	Library ID	*mlipo*				
		LogD (dd/w)[a]	LogD (dd/w)[b]	$LogED_{60,linear}$ (dd/w)[c]	$LogED_{60,exp}$ (dd/w)[d]	$LogED_{20\text{-}100,exp}$ (dd/w)[e]
Faris	JF-10b-1	-	-	1.384	0.725	-
	JF-10b-2	-	-	1.350	0.798	-
	JF-10b-3	-	-	1.420	0.646	-
	JF-10c-1	1.055	1.271	1.203	1.155	1.215
	JF-10c-2	0.723	1.311	1.170	1.163	1.189
	JF-10c-3	0.675	1.178	1.272	1.078	1.142
Furukawa	AF-10a	0.594	0.863	1.499	0.740	0.825
	AF-10b	0.650	0.975	1.518	0.892	0.929
Naylor	MN-6a	1.051	1.232	1.380	1.034	1.366
	MN-6b	1.039	1.309	1.353	0.959	1.052
	MN-6c	1.074	1.186	1.244	1.100	1.163
	MN-6d	1.111	1.016	1.180	1.221	1.155
Klein	Ester	1.122	n/a	1.499	1.274	1.424
	Amide	0.711	n/a	1.518	1.165	1.541
Average		0.891	1.149	1.356	0.996	1.182
Standard Deviation		0.215	**0.160**	0.126	**0.204**	0.208

[a]Shake-flask data from Furukawa et al., Naylor et al., and Klein et al. from recent publications from our lab. The shake-flask data from Faris et al. has not been previously reported. [b]Shake-flask data measured by Engstrom and Koch. Full data on replicates can be found in the .csv in the Supporting Information. [c]Estimated partition coefficients from the PRP isocratic 60 linear model. [d]Estimated partition coefficients from the PRP-isocratic exponential model. [e]Estimated partition coefficients from the PRP-20-100 exponential model.



AlogP	N	Slope	Y-Intercept	R2
1.33	15	1.123	-7.244	0.859
1.44	39	1.149	-7.433	0.771
1.55	21	1.334	-7.609	0.779
2.17	14	[illegible]	[illegible]	[illegible]
2.28	31	[illegible]	-7.595	0.774
2.31	15	1.230	-7.650	0.892
2.42	29	[illegible]	[illegible]	[illegible]
Average		1.130	-7.517	0.815
Standard Deviation		0.223	0.307	0.058

Figure S11: Trend between PAMPA LogPe and Log $ED_{dd/w}$ for each JF-10a thioether stereoisomer group (defined by compounds having identical AlogP, **Figure 3**). Linear regression lines are displayed in black with the regression equation and coefficient of determination displayed above.

Grouping the JF-10a library by AlogP results in compound clusters comprised of stereoisomers. Assessing the trend between PAMPA permeability and chromatographic lipophilicity (Log $\underline{ED_{dd/w}}$) for each of these groups reveals that measurable increases in chromatographic lipophilicity result in similar increases in PAMPA permeability, especially if all plotted compounds reside well beneath the threshold for aqueous solubility in the PAMPA assay. Notably, clusters 2.17 and 2.28 feature compounds with Log $ED_{dd/w}$ values well beyond 2, our general observed cutoff for aqueous solubility in PAMPA (**Figure 4**). These compound clusters have

significantly lower slopes than every other cluster. Generally, excellent correlations were obtained (average $R^2 = 0.815$) with an average $\Delta LogPe/\Delta LogED_{dd/w}$ of 1.130 across all 167 compounds.

Table S3: Thioether stereoisomer matched pairs with significant differences in PAMPA permeability which are reflected by increases in Log $ED_{dd/w}$. All matched pairs are the result of a single or double stereochemical inversion in the backbone, highlighted in yellow. Matched pair "BDL" (highlighted in blue) was featured in the main text and was subject to modeling by McMD (Figure 5 and 6).

Library	Name	Mass	AlogP	LogK'60	LogED (dd/w)	PAMPA Log Pe	Pos1	Pos3	Pos6	Pos7
A2D	SA05	974.622	1.33	0.347	-0.267	-7.906	D	D	L	D
A2D	SA03	977.640	1.33	0.773	1.362	-5.358	D	L	D	D
ALL	SA91	993.666	1.44	0.409	0.013	-8.508	L	L	D	L
ALL	SA81	979.578	1.44	0.738	1.253	-6.206	L	D	D	D
BDL	SB60	993.666	1.55	0.401	-0.020	-8.271	L	L	D	L
BDL	SB64	1000.709	1.55	0.799	1.440	-6.052	L	L	L	L
ALD	SA68	1017.672	2.28	0.547	0.583	-7.416	D	L	D	D
ALD	SA66	1007.609	2.28	0.974	1.920	-5.811	D	D	D	D
BDD	SB37	1014.653	2.42	0.516	0.462	-7.699	D	D	L	D
BDD	SB41	1011.634	2.42	0.820	1.503	-5.742	D	D	D	L

Several matched pair thioether decapeptides which differ by inversion of one or two stereocenters exhibit large differences in chromatographic lipophilicity which directly translate to PAMPA permeability. We emphasize these matched pairs (**Table S3**) to demonstrate the most significant and useful application of LPE in early stage bRo5 drug discovery. Calculated estimations of lipophilicity (ClogP, AlogP) are typically built off octanol-water partition coefficients from small molecule datasets and are unable to capture the complex structural dynamics found in most macrocyclic peptides or properly penalize exposed HBDs due to reduced desolvation penalty into the organic phase. The simple chromatographic model described in this study can identify significant structural differences between stereoisomers which aren't able to be predicted by algorithmic models for lipophilicity. Additionally, increases in chromatographic lipophilicity seem to correspond directly to increases in PAMPA

permeability and MDCK permeability as long as those compounds reside within the soluble regime of lipophilicity.

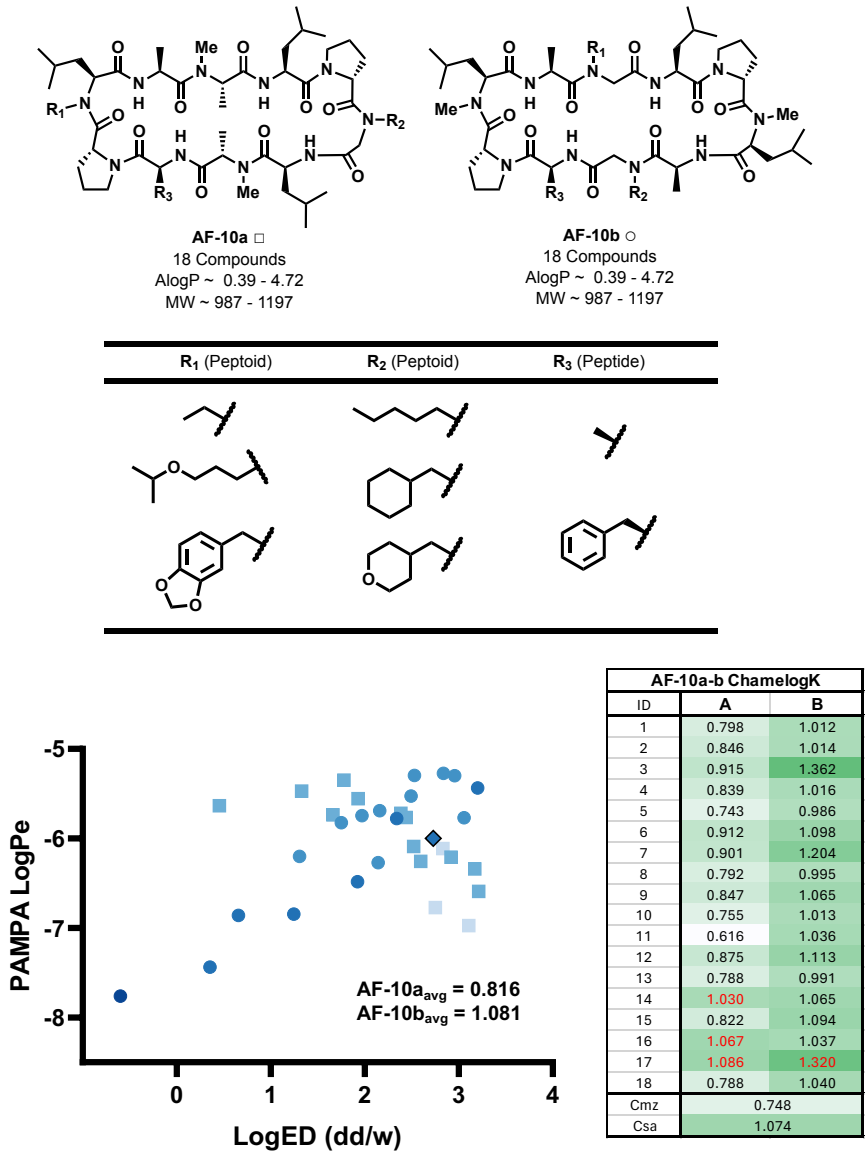


Figure S12: PAMPA permeability plotted against Log $ED_{dd/w}$ for AF-10a (squares), AF-10b (circles), and Cyclosporine A (diamond). Compounds are colored by ChamelogK value (darkest, 1.362-1.213; dark, 1.212-1.063; medium, 1.062-0.914; light, 0.913-0.765; lightest, 0.764-0.616) with averages for each scaffold displayed directly on the plot. Tabulated values are displayed to the right with conditional formatting applied (highest values dark green, lowest values white). Compounds with red Chamelogk values in the table were not observed in one of the four isocratic LogK measurements and were omitted from the plot. Carbamazepine (Cmz) and Cyclosporin A (Csa) were used as standards.

AF-10a and AF-10b are two decapeptide libraries recently studied in our lab.[30] Based on shake-flask, PAMPA, and NMR data, we have previously suggested that AF-10b exhibits a high degree of chameleonic behavior. AF-10b has a "right-shifted" permeability curve, where its peak PAMPA permeability is achieved at AlogP values relatively higher than AF-10a. Additionally, NMR and McMD simulations both suggest that AF-10b exhibits many stable conformations in water, whereas the aqueous conformations of AF-10a are much more congruent.

When assessing the relationship between PAMPA permeability and Log $ED_{dd/w}$ (**Figure 4**), we still observed a "right-shifted" permeability curve for AF-10b. For this scaffold, the trend between PAMPA permeability and Log $ED_{dd/w}$ values is still positive at the generally observed cutoff for Log $ED_{dd/w}$ in PAMPA. We hypothesize that AF-10b is able to achieve high PAMPA permeability at high measured lipophilicities due to its chameleonic nature.

To validate our assumptions about the inherent chameleonicity of both AF-10a and AF-10b, we subject both of these scaffolds to a recently reported chromatographic technique for quantifying chameleonicity, ChamelogK (**Figure S12**).[21] We adapted the reported technique exactly, but utilized our Hamilton PRP-C18 (2.1x250mm) long column. Carbamazepine (Cmz) and Cyclosporin A (Csa) were used as standards. Our measured ChamelogK value for Csa (1.074) was significantly lower than the reported value by Caron et al. (1.25), which may be due to the difference in column length and composition.

AF-10b had higher ChamelogK values than AF-10a on average ($\Delta ChamelogK_{AF\text{-}10b\ -\ AF\text{-}10a} = 0.265$), which was expected based on our previous data suggesting AF-10b exhibits chameleonic character. Interestingly, some AF-10b compounds, such as B03 and B07, boasted a much higher ChamelogK value than the scaffold average. These values even far exceed Cyclosporin A, arguably the most commonly referenced chameleonic cyclic peptide. Despite generalized differences between the scaffolds, there also seems to be intra-scaffold differences between compounds (Scaffold B_{High}-Scaffold $B_{Low} = 0.371$), which would directly be a result of side-chain substitution. This result highlights that side-chains can also bear a significant impact on scaffold conformations and ultimately chameleonicity, although a more extensive study is necessary to reveal if any underlying trends exist.

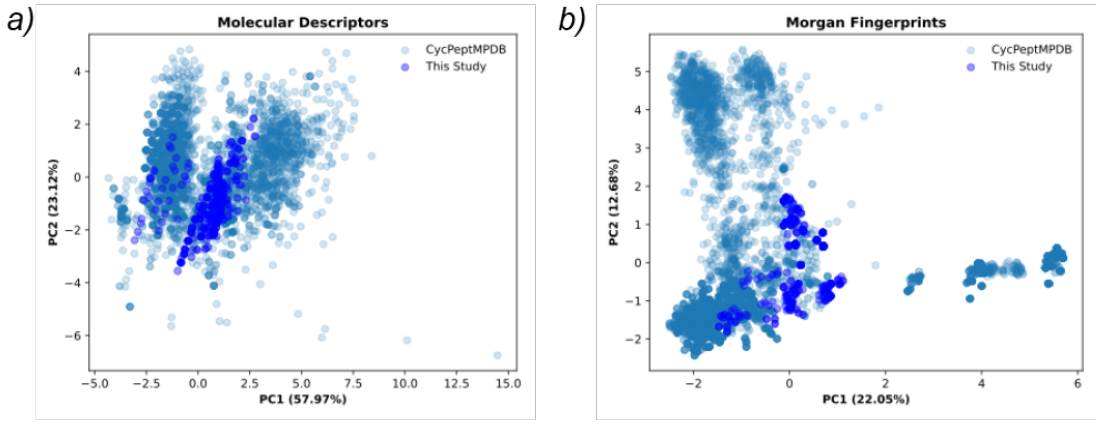


Figure S13: Principal component analysis comparing all macrocyclic peptides used in this study (dark blue) with all compounds found on CycPeptMPDB (light blue). PCA was conducted using either a) molecular descriptors or b) Morgan fingerprints as features.

To assess how representative the peptides studied herein are of their broader chemical space, a principal component analysis (PCA) was conducted. All compounds from CycPeptMPDB were downloaded (accessed 2024-09-24) and compared to the macrocyclic peptides (PROTACs omitted) included in this study. Compounds were analyzed based on the following features generated from RDKit: Molecular Weight, AlogP, HBD Count, HBA Count, Rotatable Bond Count, Aromatic Ring Count, and TPSA.

We observed a strong overlap between our compounds and those of the cyclic peptide database (**Figure S13**) suggesting a similarity between the two datasets. Notably, the furthest peptide from the origin (PC1 = -3.07, PC2 = -2.39; PC1 = -1.47, PC2 = -1.68) was still well within the bounds of the database peptides for analyses using molecular properties and Morgan fingerprints features, respectively. We conclude that our macrocyclic peptides are indeed representative of other peptides that are designed to achieve passive membrane permeability.

Appendix B: Supplementary Figures and Discussion for Chapter 2

Library Nomenclature

Throughout the main text and supporting information, compounds from the library are listed from their N-terminus to C-terminus. In terms of the DNA-encoded library cycles, that means that compounds are listed BB3-BB2-BB1. The compound Val-Phe-Leu was coupled in leucine in cycle 1, followed by phenylalanine in cycle 2, and then finally valine in cycle 3.

Headpiece Structure and Model Compound Tag Design

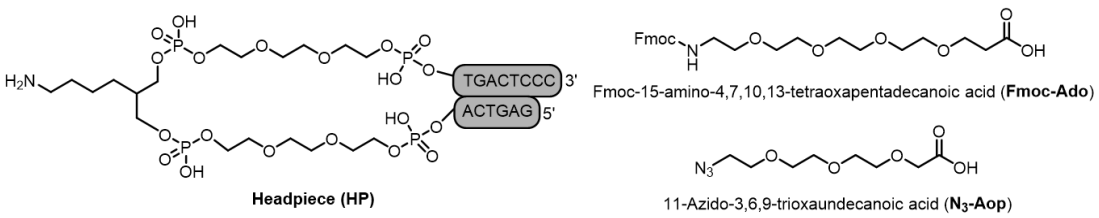
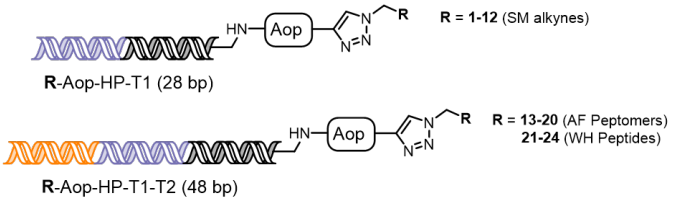


Figure S14: Structure of headpiece and linkers used in this study

DNA headpiece was purchased from Biosearch Technologies. For the synthesis of model compounds, an azide-containing linker was utilized (azido-PEG3-Acid (**N_3-Aop**), purchased from BroadPharm). For DNA-encoded library synthesis, an Fmoc-amine PEG linker (**Fmoc-Ado**) was coupled onto headpiece prior to synthesis.



Name	Duplex Sequences Aligned (5' --- 3')
HP	TGACTCCC
	ACTGAG
T-1	TCTTGGCTATTCACACCGAG
	GGAGAACCGATAAGTGTGGC
T-2	CGGCCTCGTTCATAACACGT
	TCGCCGGAGCAAGTATTGTG

Figure S15: Structure and sequences of DNA tags for model compounds **1-24**.

For synthesis of the model compounds, two 20 bp tags (T1 and T2) were utilized to extend the DNA duplex (**Figure S15**). The resulting constructs were 28 bp for the small molecules (SM**1-12**) and 48 bp for the peptides (AF**13-20** and WH**21-24**). The tags were kept short relative to a full-length DNA-encoded library to facilitate identification of products by LCMS. In these cases, the DNA tag was between 50-200x the molecular weight of the warhead.

LDEL DNA Tag and Library Design

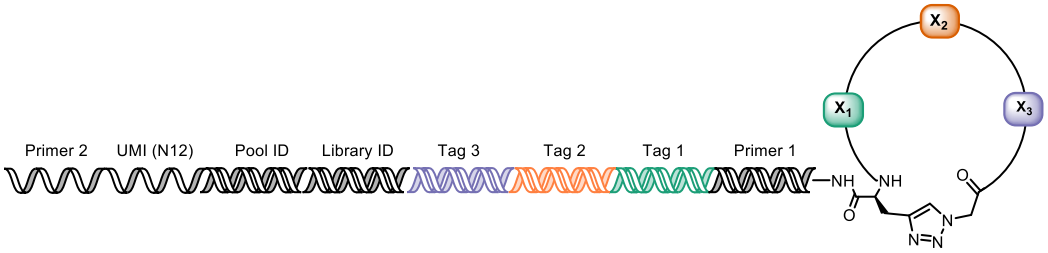
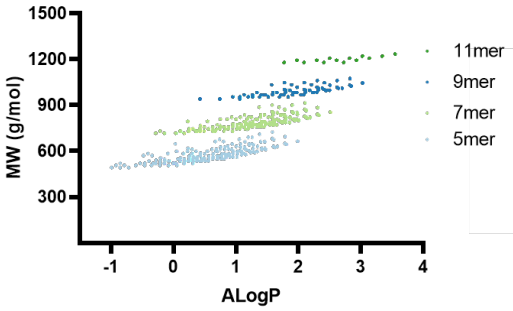


Figure S16: Illustration of tag design used in LDEL.

The DNA-encoded library in this study (LDEL) was synthesized over 3 rounds of encoded split-and-pool chemistry. The full-length DNA tag was 114 bp, with the first 82 bp being duplex DNA and the final 32 bp (N12 and Primer 2) being single stranded DNA (**Figure S16**). Strikingly, this single stranded portion seemingly does not affect the relative elution of each library member under the chromatographic

conditions employed in this study (see **Figure 9a**, null library).The N12 region, consisting of 12 randomized nucleobases (25:25:25:25 A:T:C:G), was included to assist in identification of PCR artefacts during sequencing. This N12 region confers each encoded library member with potentially 1.7E7 unique encoding sequences. Additionally, each cycle of chemistry (Tag 1, Tag 2, Tag 3) was tagged in triplicate, resulting in 27 unique tag combinations for each library member, driving the theoretical tag diversity up to 4.5E8 per library member. This significant amount of tag diversity ensures that sequences are included due to an increase in population of that library member and not due to sequence bias during the PCR experiment. Null tags were incorporated to facilitate identification of truncated structures during deconvolution of the sequence count chromatograms. The Library ID tag was used to differentiate between linear and cyclized libraries. The Pool ID tag was not used for additional encoding in this study (the fractions from the lipophilicity selection experiment were barcoded during the subsequent PCR step). Sequencing counts used for sequencing count chromatograms were deduplicated, with identical sequences resulting from PCR amplification being assigned as a single count regardless of the quantity.

The three cycle library (32 BB1 x 32 BB2 x 58 BB3) resulted in a theoretical library diversity of 59,392. With incorporation of nulls at each cycle of chemistry, the null-incorporated library diversity was 64,251. An equimolar portion of linear and cyclized library was combined and taken forward into the chromatographic lipophilicity experiment, resulting in a total theoretical library diversity of 128,502.



Number of Tripeptide Building Blocks	Ring Size	Number of Scaffolds	Number of Compounds
0	5	175	2592
1	7	2400	14400
2	9	10800	26400
3	11	16000	16000

Figure S17: Physicochemical properties of LDEL. The library diversity reported in the table does not include diversity from null encoding.

Considering differences in stereochemistry, N-methylation or N-alkylation, constricting residues such as proline, or backbone extending residues such as β-homoamino acids, LDEL contained significant scaffold diversity, especially at larger ring sizes.

Incorporation of tripeptides at each cycle of chemistry resulted in four ring sizes. Including the propargyl glycine and azido acetic acid residues, the resultant ring sizes were 5, 7, 9, and 11. Liposcan elements were included as single amino acid building blocks (Nva, Val, Leu, Phe), however the tripeptide building blocks were all derived from a consistent Leu-Leu-Pro motif with varied by stereochemistry, N-methylation, and, in some cases, N-alkylation (peptoid) with isobutylamine for the central residue. This resulted in a broad range in predicted lipophilicity (**Figure S17**) for the smaller ring size and a smaller breadth of lipophilicity for the 11mers, with the only differences arising from different extents of backbone N-methylation or N-alkylation.

Classic Chrome Model for Peak Assignment and Reaction Profiling

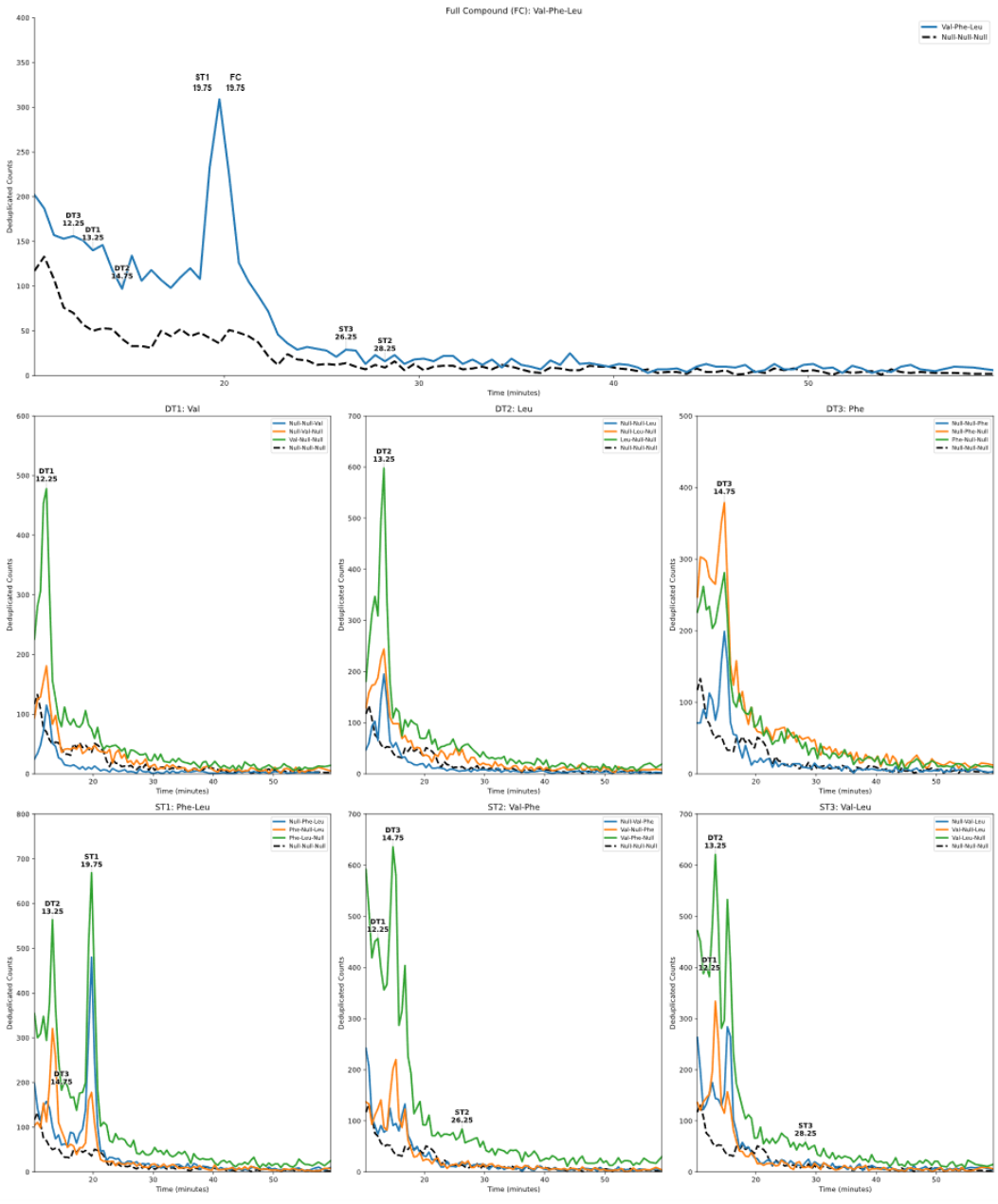


Figure S18: Full compound and null-truncation chromatograms for LDEL compound Val-Phe-Leu. The full compound peak closely matches the compound peak associated with ST1, Phe-Leu, indicating the full compound is a truncated byproduct and a failed Val coupling in the third cycle of chemistry.

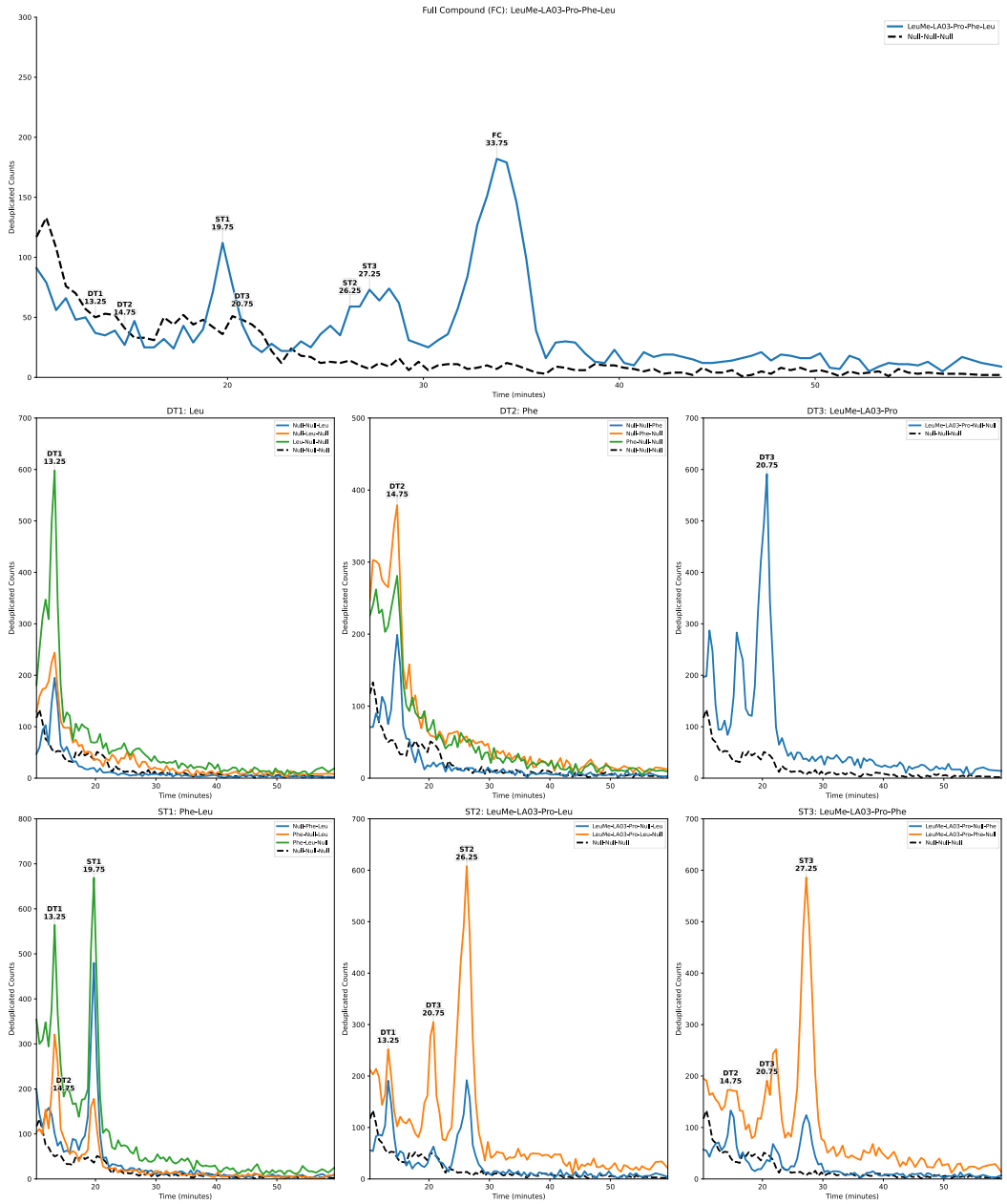


Figure S19: Full compound and null-truncation chromatograms for LeuMe-LA03-Pro-Phe-Leu. The full compound peak does not align with the assigned compound peaks for any associated Null truncation, indicating a successful synthesis of this encoded library member.

To improve peak-picking accuracy and to identify trends in reactivity throughout the library, we developed a null-informed peak picking model, Classic Chrome. This model directly compares every compound in a DNA-encoded library to its associated Null truncations, which can be displayed visually with a series of chromatograms (**Figure S18-19**). In these chromatograms, the top row represents the Full Compound sequence count chromatogram which is annotated based on available associated Null truncation chromatograms. In the row beneath the Full Compound chromatogram are the associated double truncation products (DT), and in the row beneath that are the associated single truncation products (ST). The amount of associated truncation products depends on the diversity of building blocks within each compound. For example, compound "Leu-Leu-Leu" will only have only one ST product (Leu-Leu) and one DT product (Leu).

In the first example provided above (**Figure S18**, also featured in **Figure 10**), the full compound (FC) chromatogram closely matches the compound peak associated with ST1, Phe-Leu. This suggests that the cycle 3 coupling to valine failed, and the full encoded compound is a truncated byproduct, Phe-Leu. In comparison, the Classic Chrome output for a related compound LeuMe-LA03-Pro-Phe-Leu (substituting Val for the tripeptide building block LeuMe-LA03-Pro) strongly suggests formation of the intended full product (**Figure S19**). The peak assigned to the suspected product in the FC chromatogram does not align with any associated null truncation, indicating successful coupling of the tripeptide in the third cycle of chemistry.

Classic chrome chromatograms are provided for each of the resynthesis compounds (**LDEL001-044**) featured in **Figure 12**, see the Analytical Spectra section below. Additional annotation for these compounds has been provided in red font.

Sequencing Count Chromatogram Cyclization Analysis

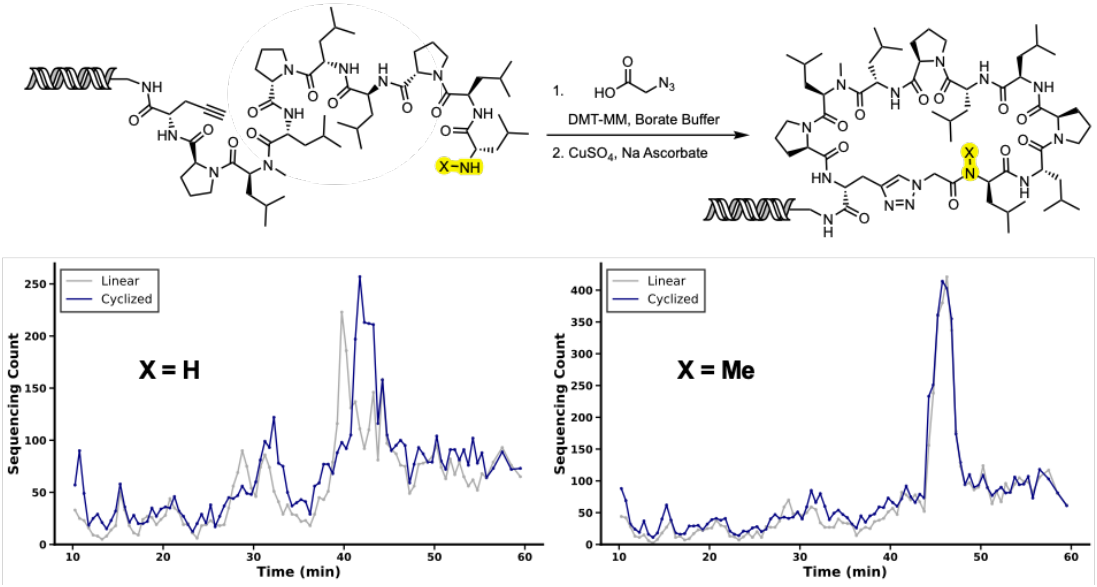


Figure S20: Sequencing count chromatograms of an 11mer matched pair which differs only by the N-methylation about their terminal amine (prior to final coupling with azido acetic acid, highlighted in yellow).

After all the encoding rounds of split-and-pool synthesis, the final reactions of the library were performed altogether on the pooled library. The library was double coupled with azido acetic acid and then split in half. Half of the library remained linear, and the other half was cyclized with copper sulfate pentahydrate and sodium ascorbate. An equimolar mixture of linear:cyclized library (quantified by qPCR) was separated chromatographically and sequenced. Comparison of linear and cyclized

sequencing count chromatograms of the same compound allows for direct assessment of cyclization efficiency.

Analysis of LDEL compounds which contain an N-terminal N-methyl or N-substituted residue (**Figure S20**) generally yielded nearly identical sequence count chromatograms between their linear and cyclized forms. This result may suggest that scaffolds terminating with a secondary amine exhibited poor cyclization efficiency, however it is far more likely that this result indicates that azido acetic acid coupled poorly to secondary amine nucleophiles in the final coupling step. We confirmed this result by performing couplings onto pure headpiece constructs bearing a secondary amine warhead, which resulted in extremely low coupling efficiencies regardless of coupling conditions (data not shown). In contrast, compounds which terminated in a primary amine typically yield significant differences between their linear and cyclized chromatograms.

Although it is difficult to predict how cyclization may affect the 3-dimensional conformation and whether that may increase or decrease bulk compound lipophilicity, cyclization tended to produce measurable differences in retention time for compounds in this library. Direct comparison of linear and cyclized sequence count chromatograms assisted in accurate peak assignment and facilitated assessments cyclization efficiency, as shown in the matched pair example above (**Figure S20**).

Generally, macrocyclic products eluted later than their linear counterparts, with the difference in retention time ranging from seconds to several minutes (**Figure**

S21). Meanwhile, DNA-encoded library members which had an N-terminal substitution (methyl or R) had nearly identical elution profiles between linear and cyclic versions of the same compound, likely due to inefficient acylation with azido acetic acid prior to cyclization.

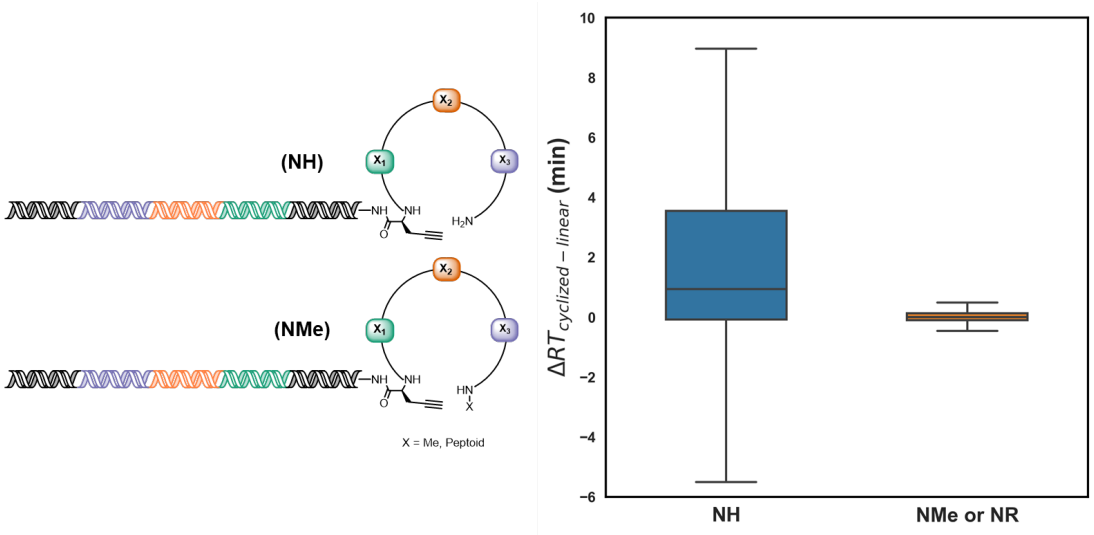


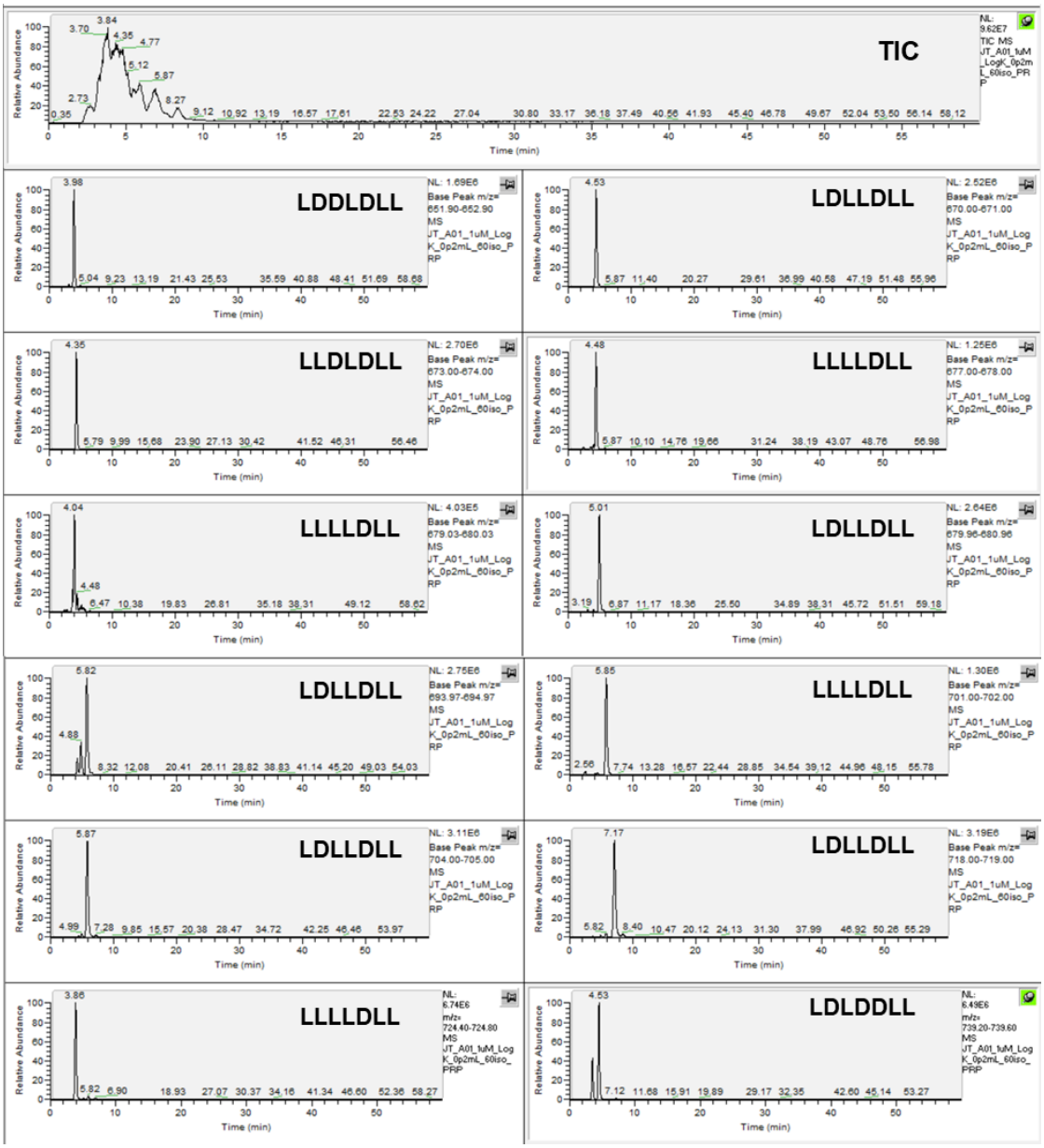
Figure S21: Comparison between compounds which end Cycle 3 with an N-terminal substitution (methyl or R) and compounds which end with an N-terminal NH. Cyclized compounds generally exhibited higher longer retention times than their linear counterparts.

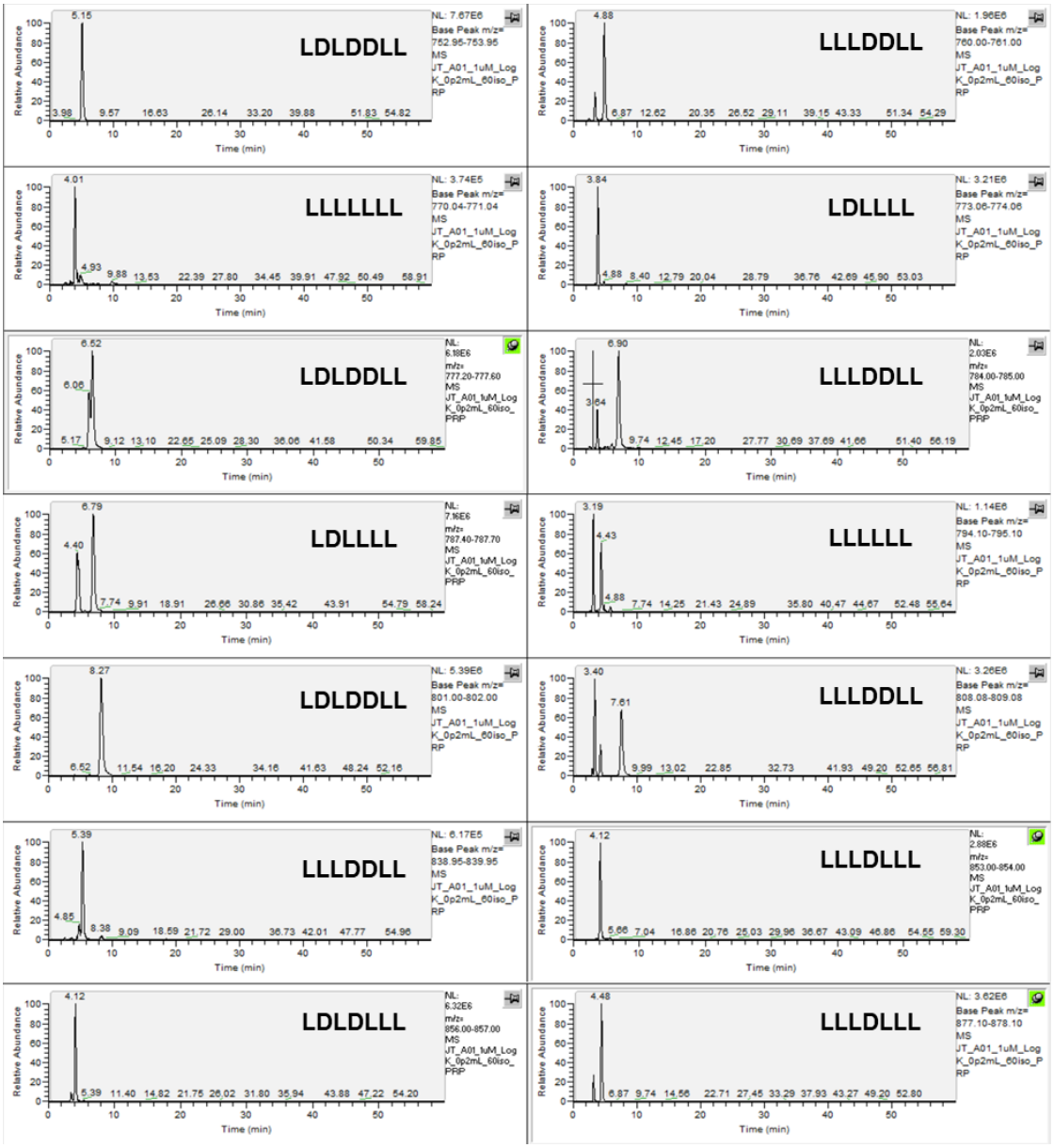
Appendix C: Selected-mass spectra for Chapter 1 Libraries

Selected ion monitoring (SIM) for each compound synthesized for this study (JT-7 and JF-10c) is displayed below. Each spectrum was obtained on a Thermo Scientific Ultimate 3000 UPLC system with a Thermo Scientific OrbiTrap VelosPro mass spectrometer using a 2.1x250 mm, 5 µm PRP-C18 column from Hamilton. Data was collected with A:Water (+0.1% formic acid) and B:ACN (+0.1% formic acid) at 60%B using a 0.2 mL/min flow rate, unless otherwise noted. Each SIM is annotated with a compound identifier found in the LogEDdata.xlsx in the supporting information of the associated publication (see acknowledgements).

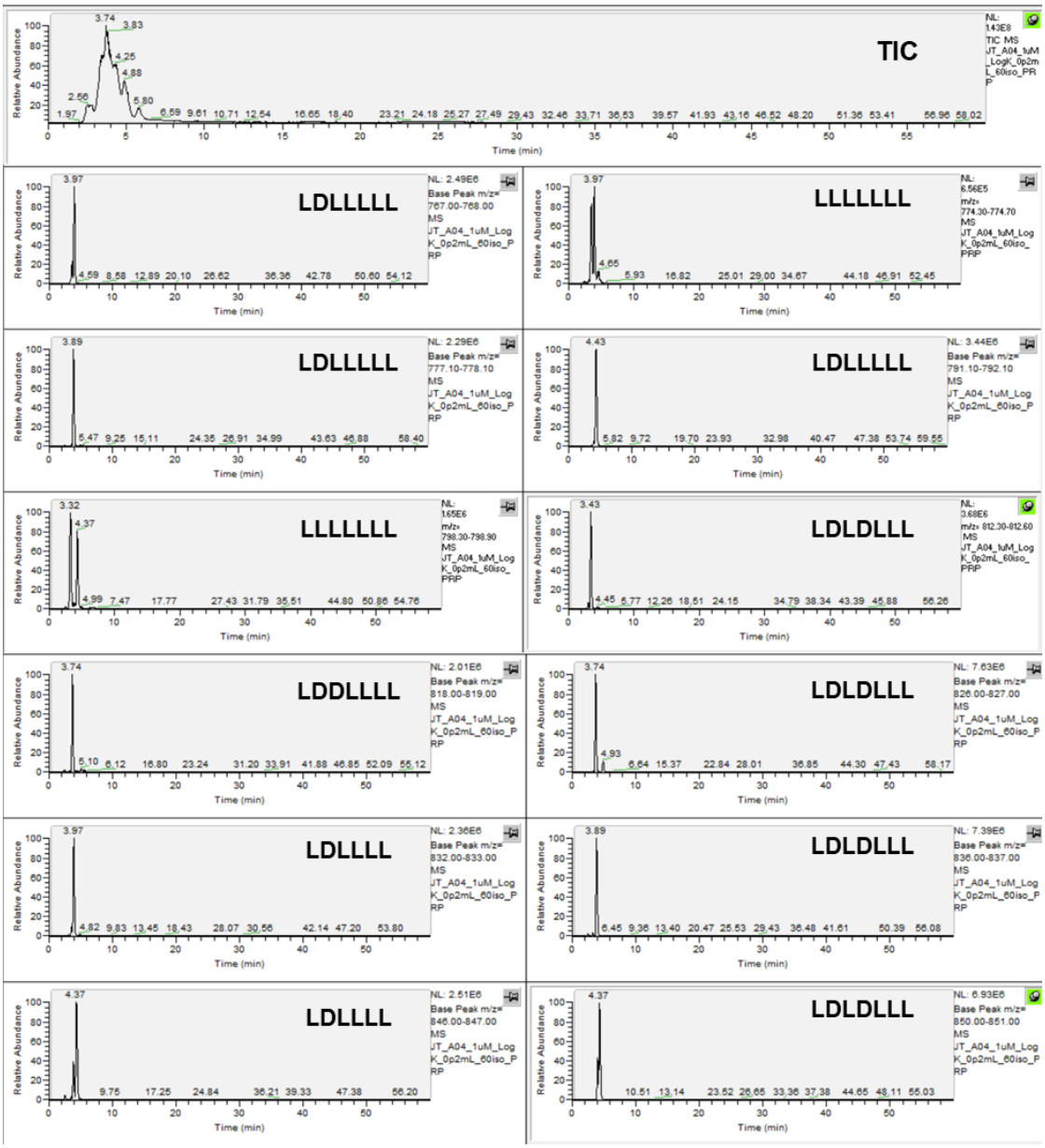
JT-7 Sublibraries

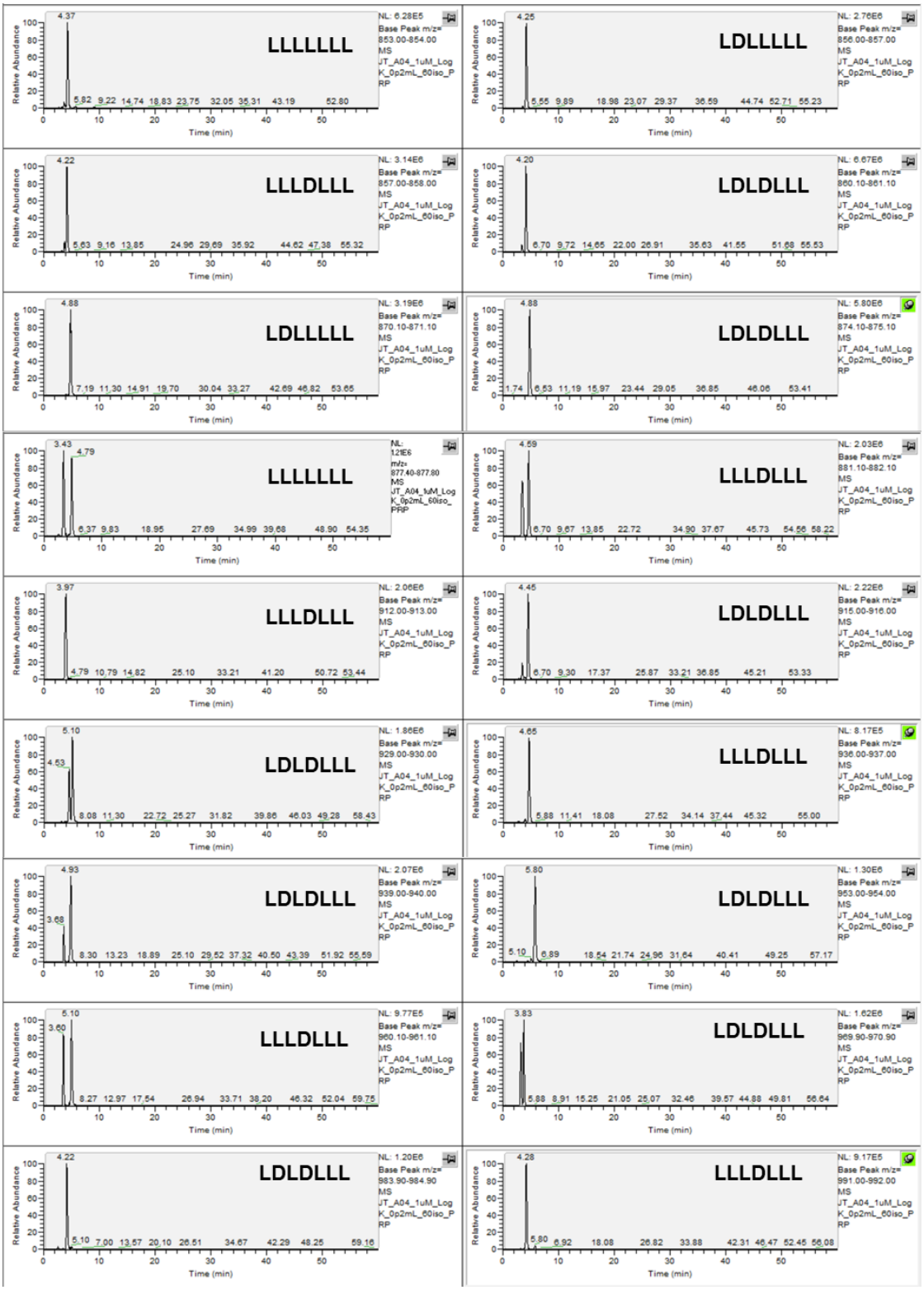
Library A1



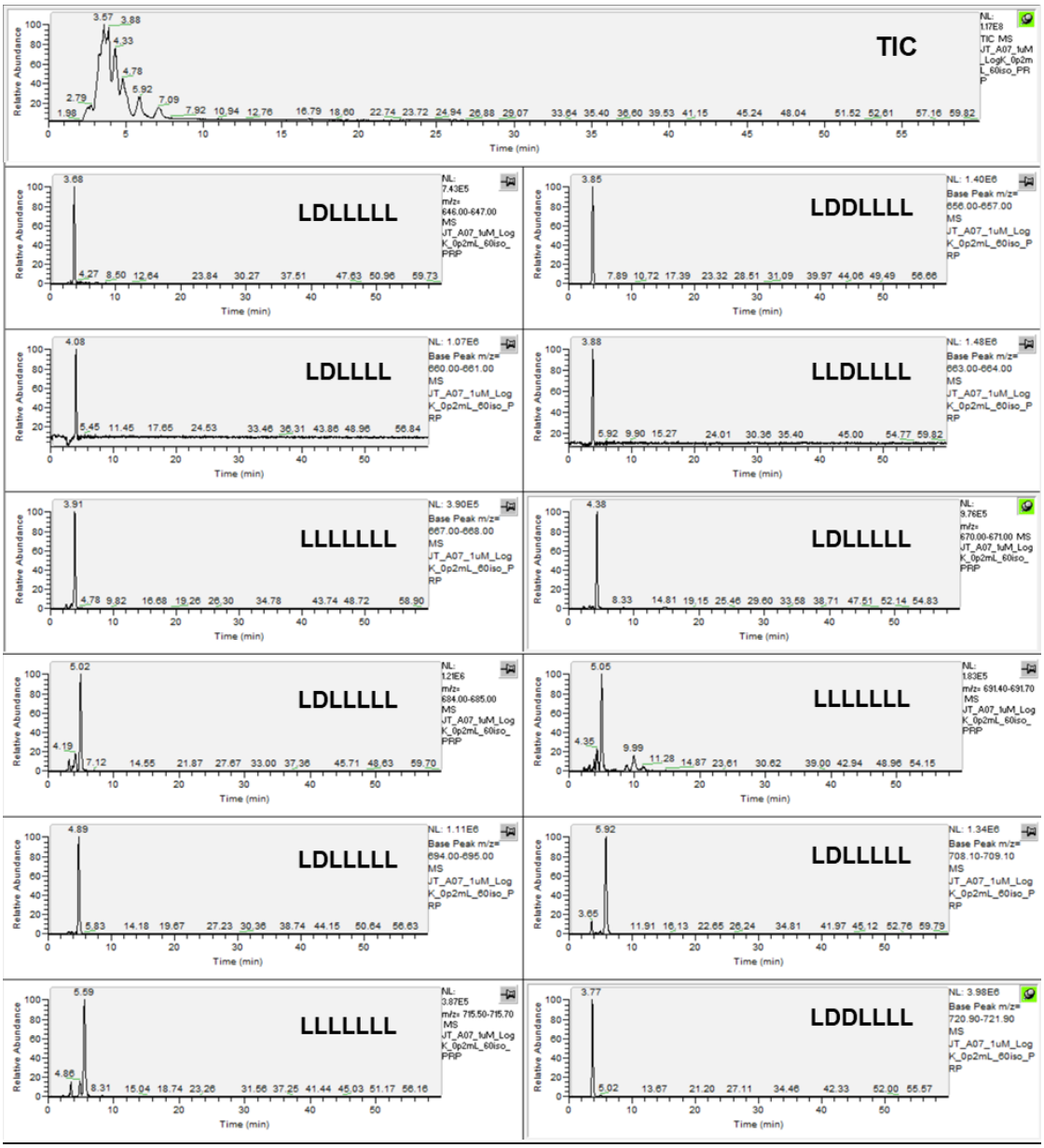

LDLDDLL
LLLDDLL
LLLLLLL
LDLLLL
LDLDDLL
LLLDDLL
LDLLLL
LLLLLL
LDLDDLL
LLLDDLL
LLLDDLL
LLLDLLL
LDLDLLL
LLLDLLL
Relative Abundance
Time (min)

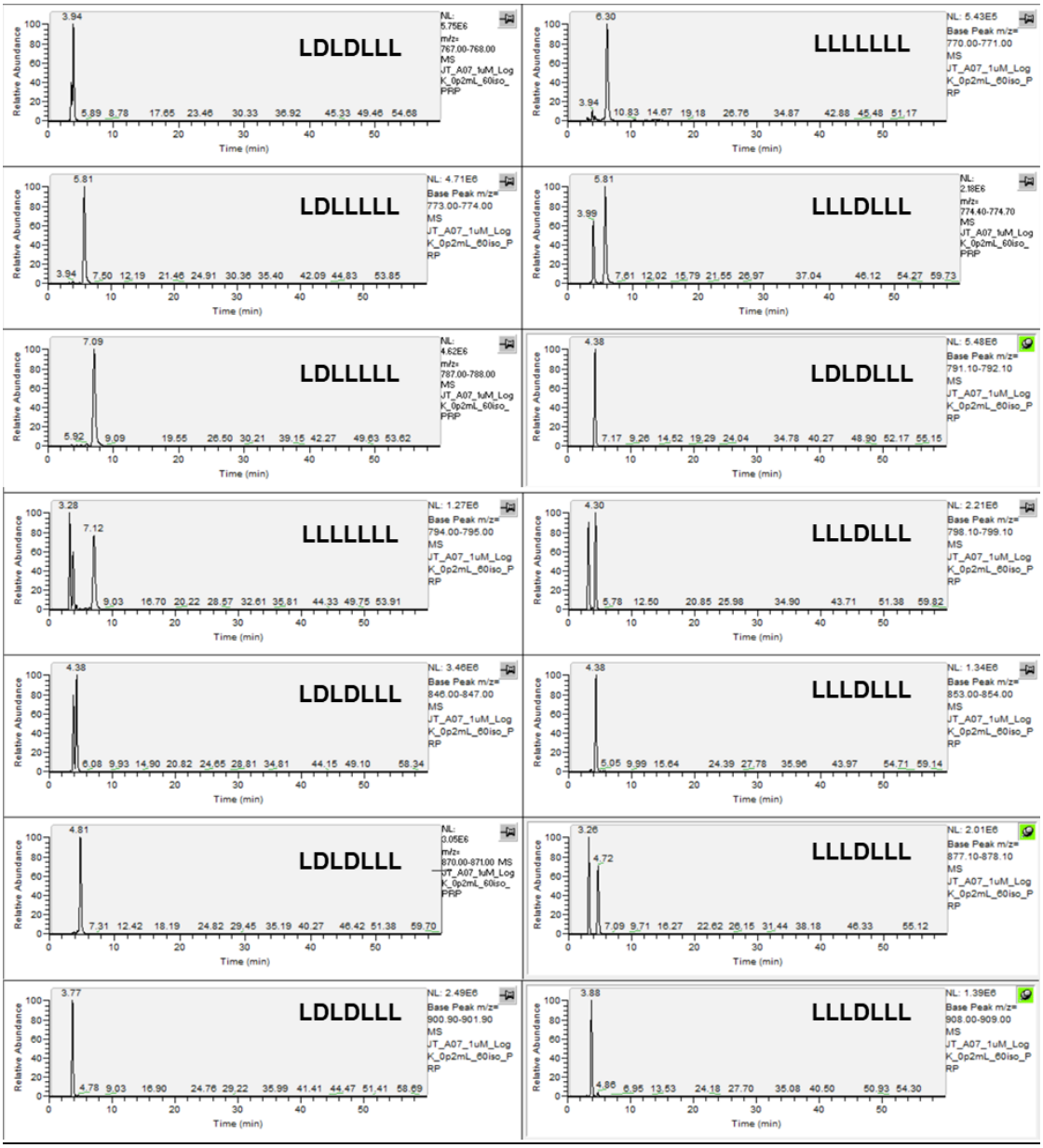
Library A4


TIC
LDLLLLL
LLLLLLL
LDLLLLL
LDLLLLL
LLLLLLL
LDLDLLL
LDDLLLL
LDLDLLL
LDLLLL
LDLDLLL
LDLLLL
LDLDLLL
Relative Abundance
Time (min)

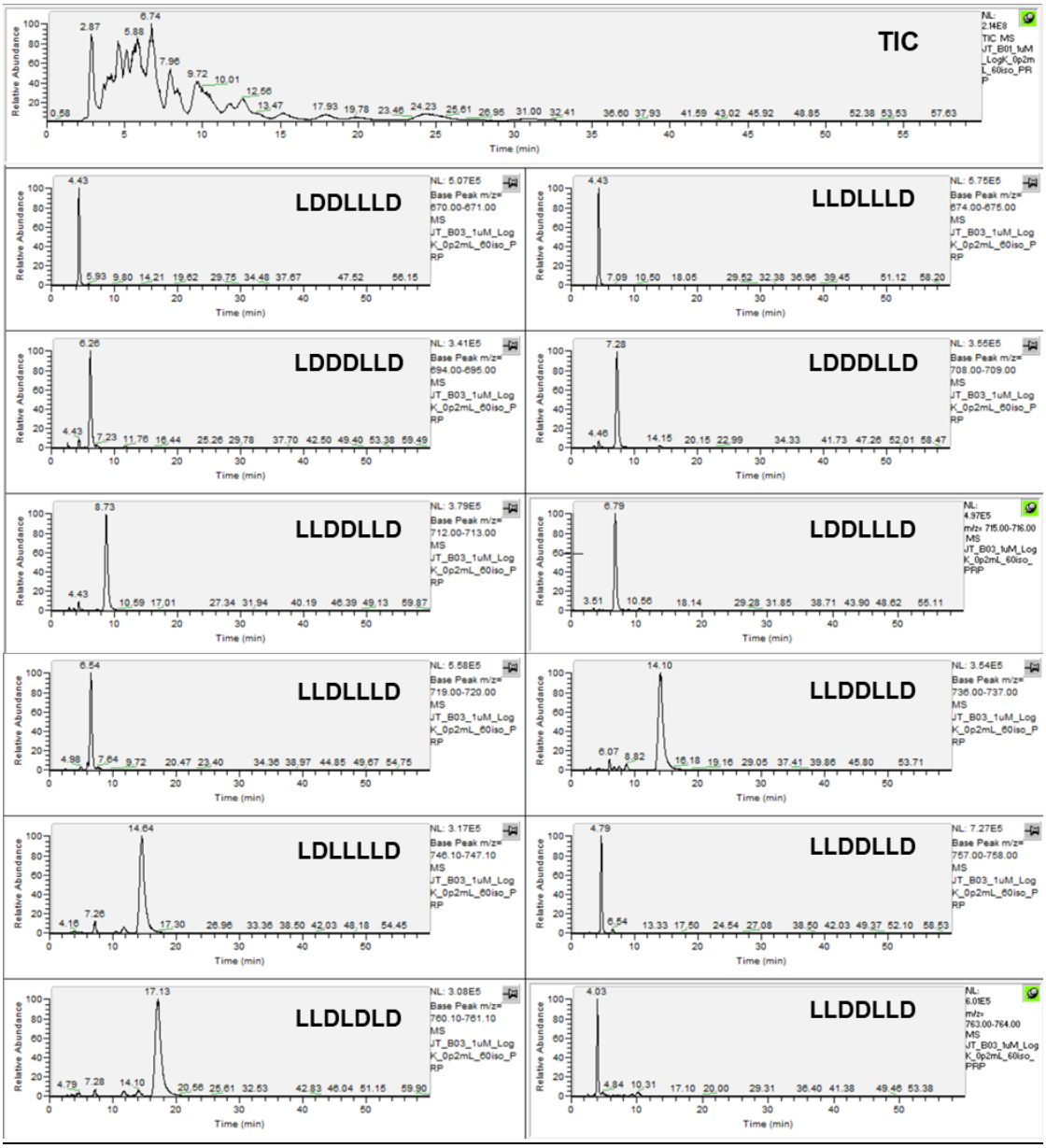

LLLLLLL
LDLLLLL
LLLDLLL
LDLDLLL
LDLLLLL
LDLDLLL
LLLLLLL
LLLDLLL
LLLDLLL
LDLDLLL
LDLDLLL
LLLDLLL
LLLDLLL
LDLDLLL
LDLDLLL
LLLDLLL
LDLDLLL
LLLDLLL
Relative Abundance
Time (min)

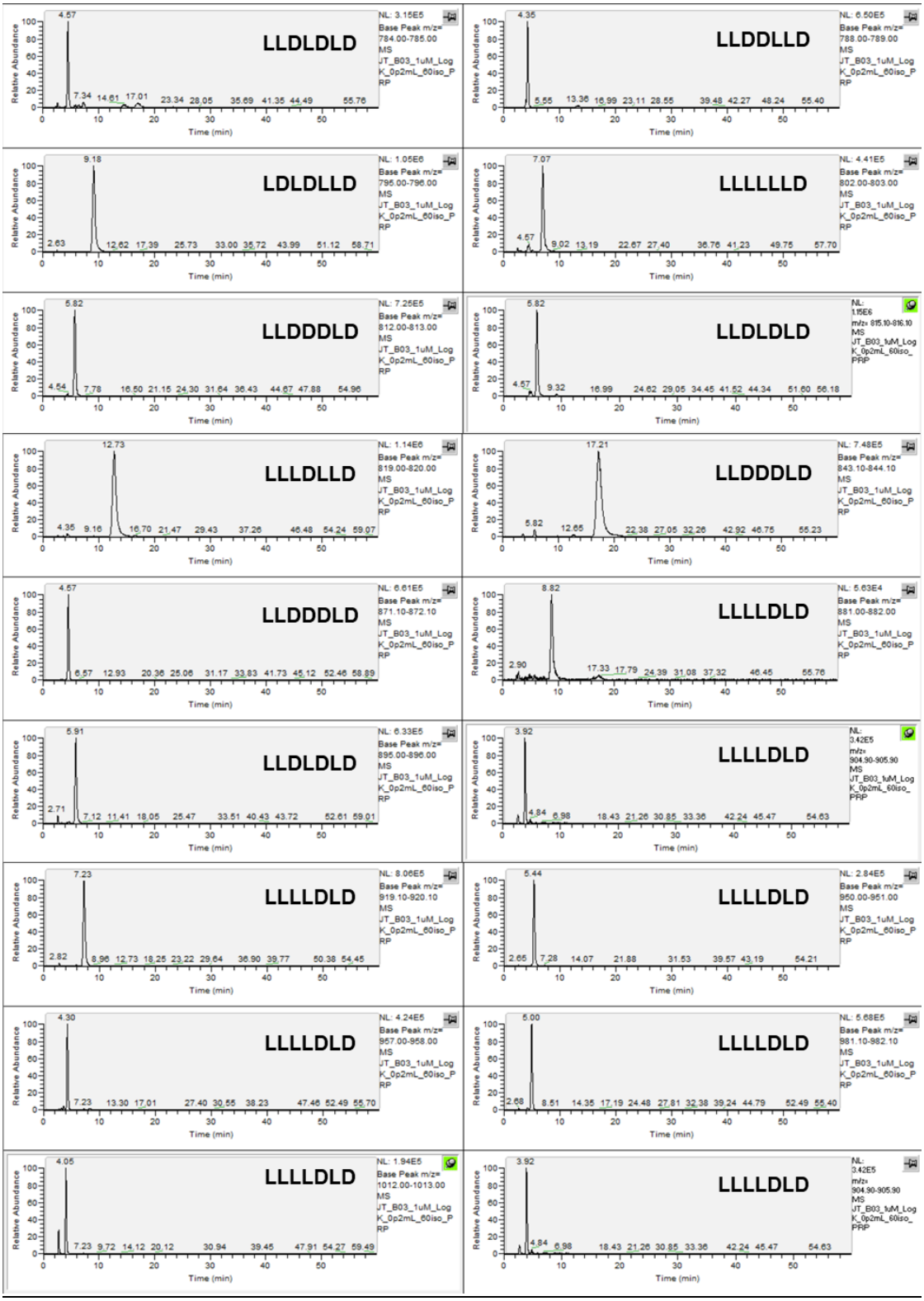
Library A7


TIC
LDLLLLL
LDDLLLL
LDLLLL
LLDLLLL
LLLLLLL
LDLLLLL
LDLLLLL
LLLLLLL
LDLLLLL
LDLLLLL
LLLLLLL
LDDLLLL
Relative Abundance
Time (min)

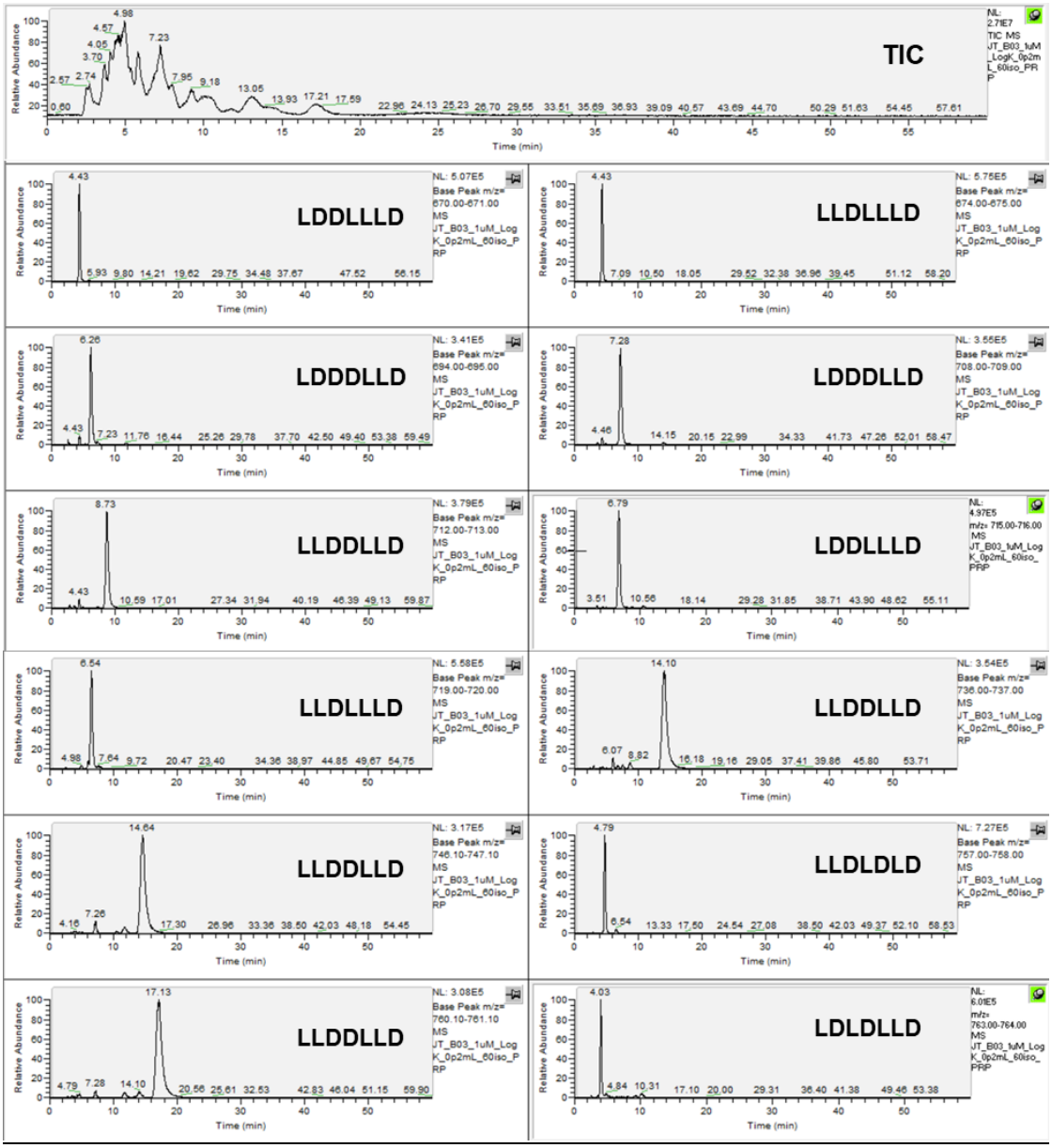

LDLDLLL
LLLLLLL
LDLLLLL
LLLDLLL
LDLLLLL
LDLDLLL
LLLLLLL
LLLDLLL
LDLDLLL
LLLDLLL
LDLDLLL
LLLDLLL
LDLDLLL
LLLDLLL
Relative Abundance
Time (min)

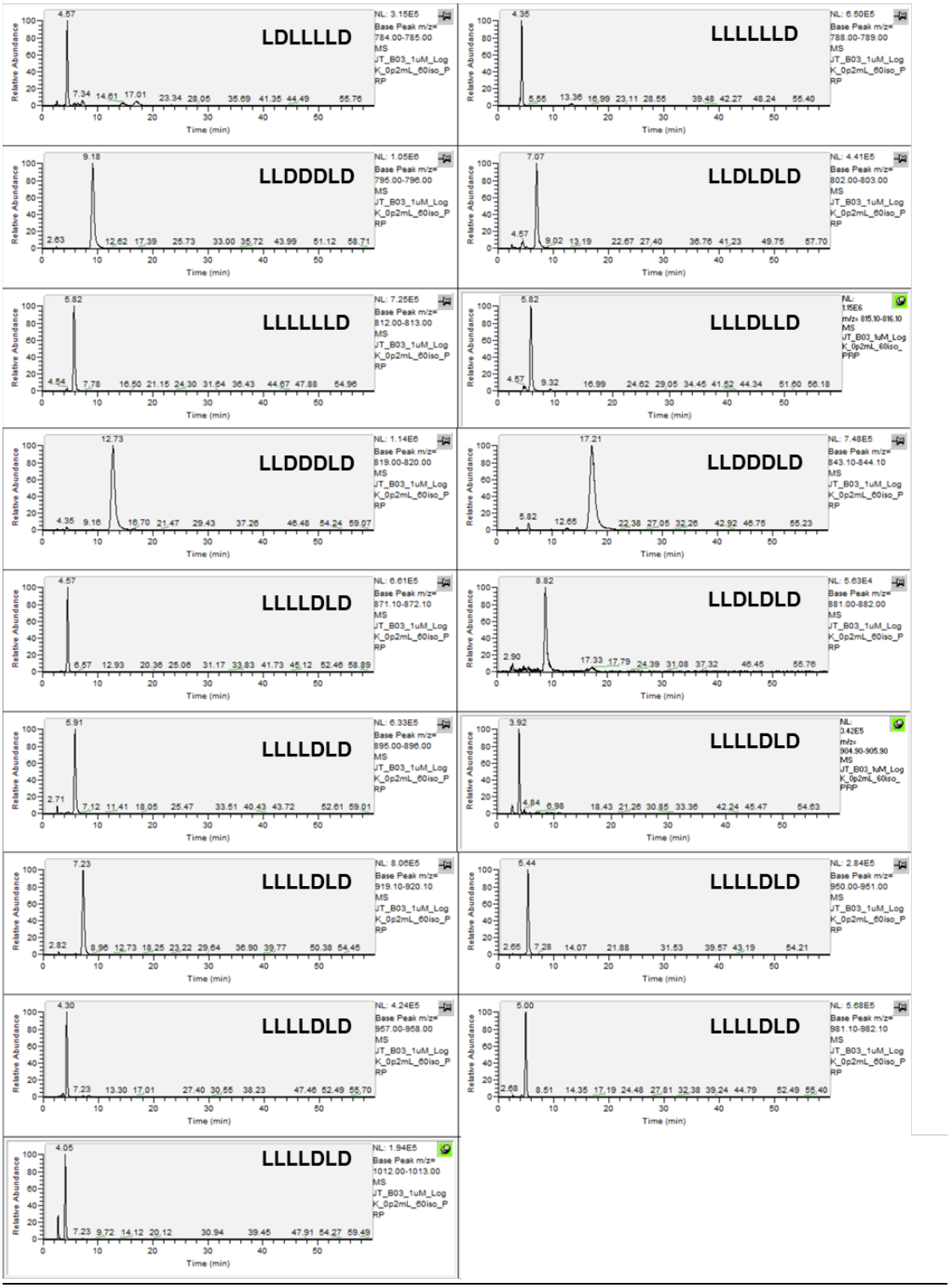
Library B1


TIC
LDDLLLD
LLDLLLD
LDDDLLD
LDDDLLD
LLDDLLD
LDDLLLD
LLDLLLD
LLDDLLD
LDLLLLD
LLDDLLD
LLDLDLD
LLDDLLD
Relative Abundance
Time (min)

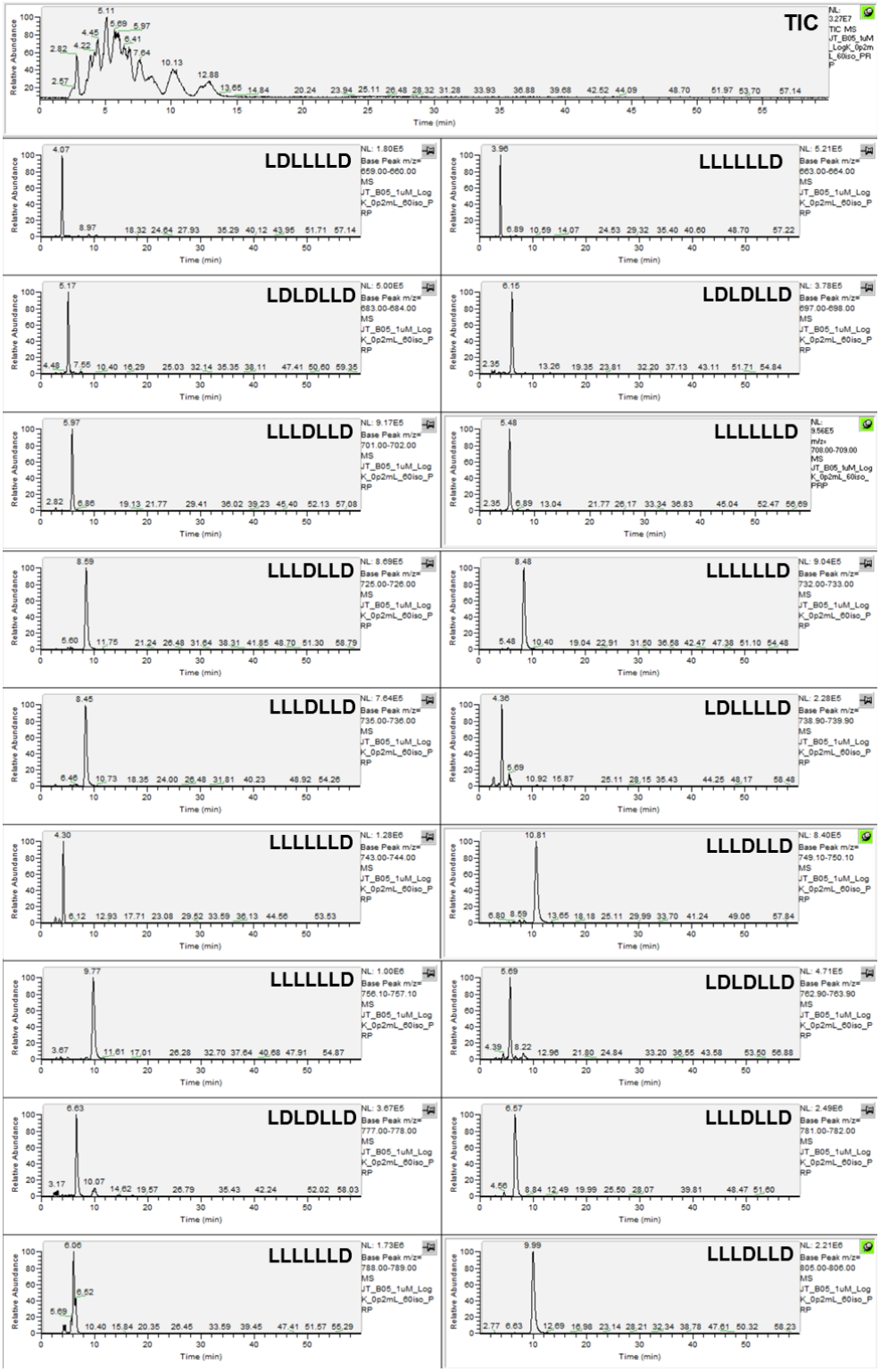

LLDLDLD
LLDDLLD
LDLDLLD
LLLLLLD
LLDDDLD
LLDLDLD
LLLDLLD
LLDDDLD
LLDDDLD
LLLLDLD
LLDLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
Relative Abundance
Time (min)

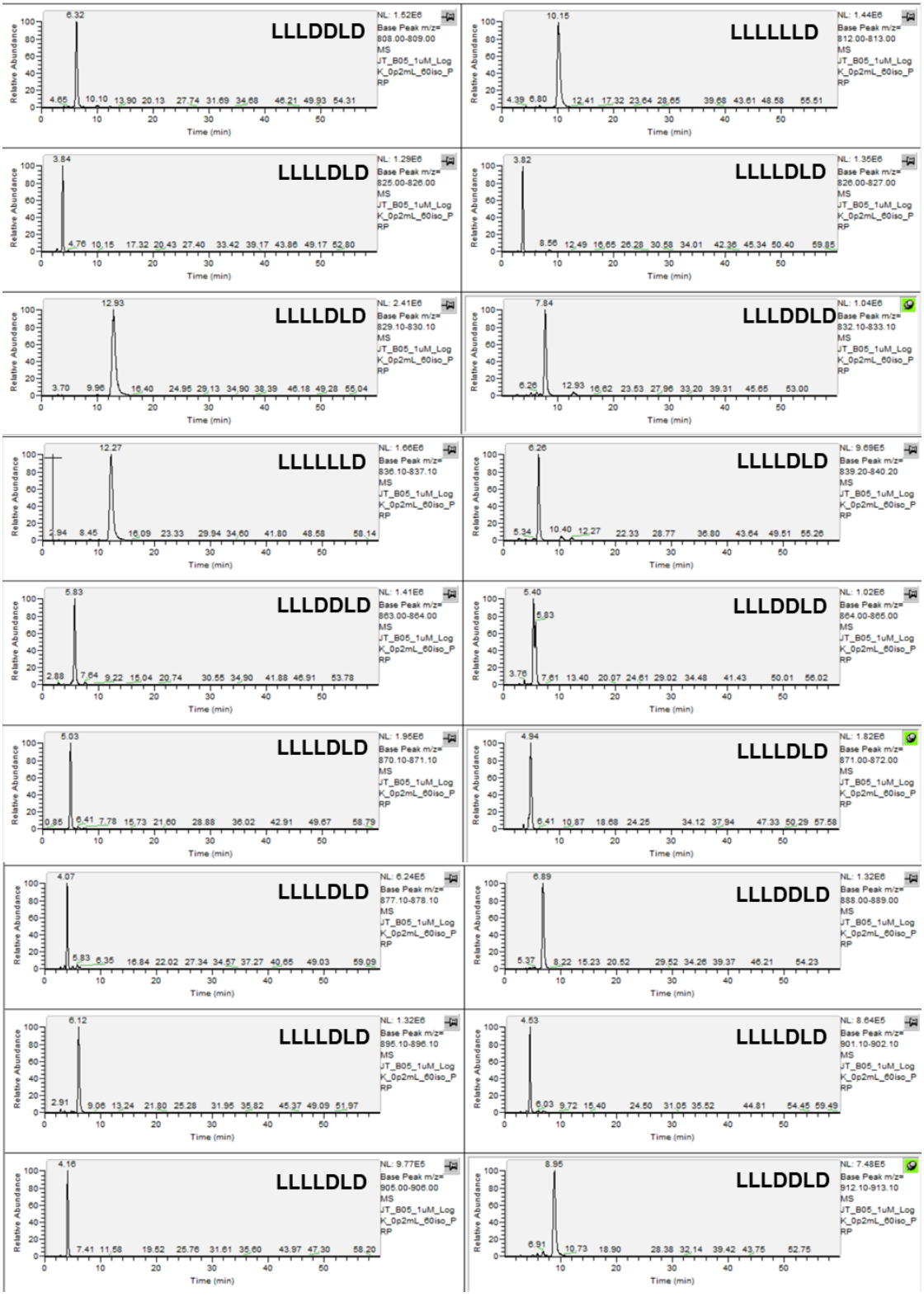
Library B3

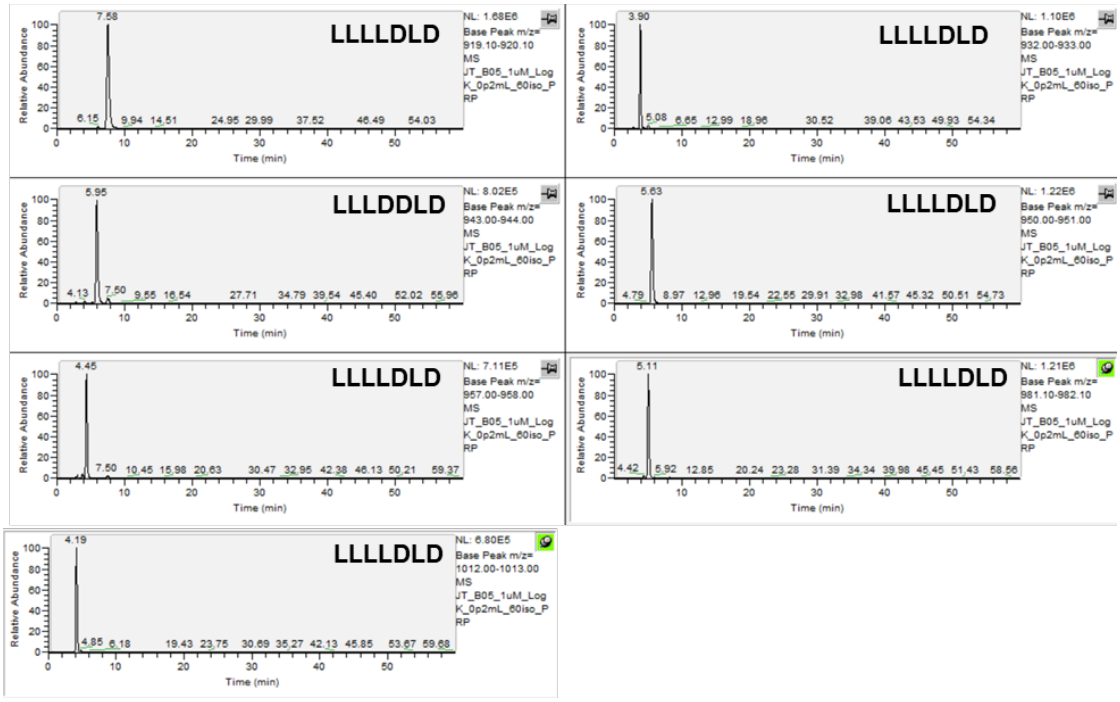

TIC
LDDLLLD
LLDLLLD
LDDDLLD
LDDDLLD
LLDDLLD
LDDLLLD
LLDLLLD
LLDDLLD
LLDDLLD
LLDLDLD
LLDDLLD
LDLDLLD
Relative Abundance
Time (min)


LDLLLLD
LLLLLLD
LLDDDLD
LLDLDLD
LLLLLLD
LLLDLLD
LLDDDLD
LLDDDLD
LLLLDLD
LLDLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLLDLD
Relative Abundance
Time (min)

Library B5

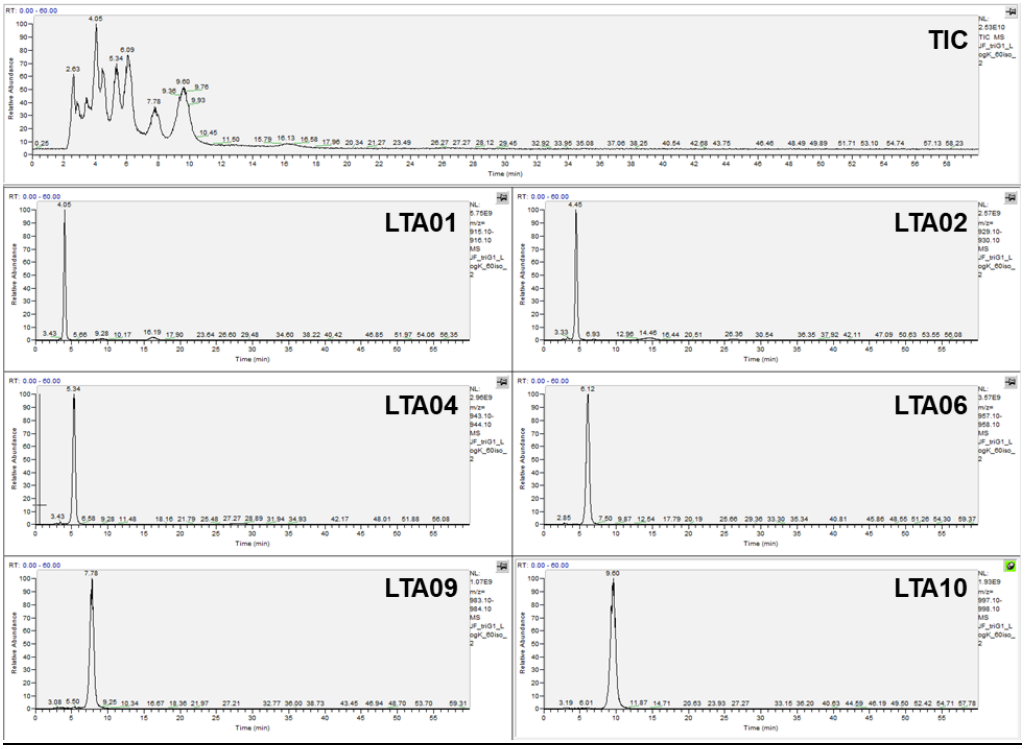

TIC
LDLLLLD
LLLLLLD
LDLDLLD
LDLDLLD
LLLDLLD
LLLLLLD
LLLDLLD
LLLLLLD
LLLDLLD
LDLLLLD
LLLLLLD
LLLDLLD
LLLLLLD
LDLDLLD
LDLDLLD
LLLDLLD
LLLLLLD
LLLDLLD
Relative Abundance
Time (min)


LLLDDLD
LLLLLLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLDDLD
LLLLLLD
LLLLDLD
LLLDDLD
LLLDDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLDDLD
LLLLDLD
LLLLDLD
LLLLDLD
LLLDDLD
Relative Abundance
Time (min)

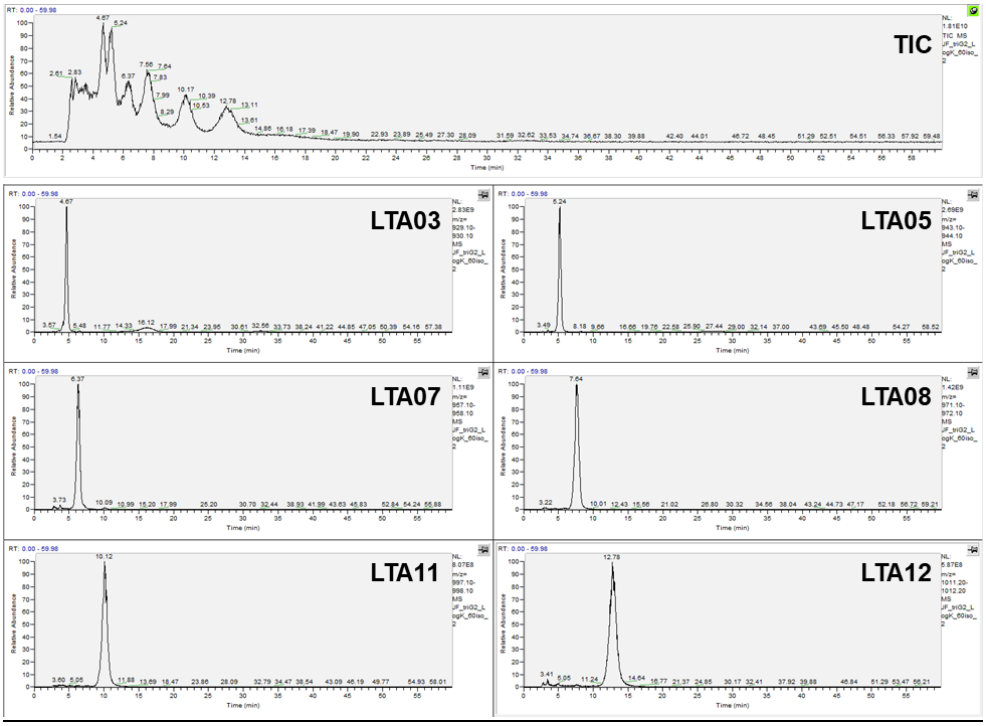

LLLLDLD
NL: 1.68E6
Base Peak m/z= 919.10-920.10
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
7.58
6.15 9.94 14.51 24.95 29.99 37.52 46.49 54.03
LLLLDLD
NL: 1.10E6
Base Peak m/z= 932.00-933.00
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
3.90
5.08 6.65 12.99 18.96 30.52 39.06 43.53 49.93 54.34
LLLDDLD
NL: 8.02E5
Base Peak m/z= 943.00-944.00
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
5.95
4.13 7.50 9.55 16.54 27.71 34.79 39.54 45.40 52.02 55.96
LLLLDLD
NL: 1.22E6
Base Peak m/z= 950.00-951.00
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
5.63
4.79 8.97 12.96 19.54 22.55 29.91 32.98 41.57 45.32 50.51 54.73
LLLLDLD
NL: 7.11E5
Base Peak m/z= 957.00-958.00
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
4.45
7.50 10.45 15.98 20.63 30.47 32.95 42.38 46.13 50.21 59.37
LLLLDLD
NL: 1.21E6
Base Peak m/z= 981.10-982.10
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
5.11
4.42 5.92 12.85 20.24 23.28 31.39 34.34 39.98 45.45 51.43 58.56
LLLLDLD
NL: 6.80E5
Base Peak m/z= 1012.00-1013.00
MS JT_B05_1uM_Log K_0p2mL_60iso_PRP
4.19
4.85 6.18 19.43 23.75 30.69 35.27 42.13 45.85 53.67 59.68
Relative Abundance
Time (min)

JF-10c Libraries

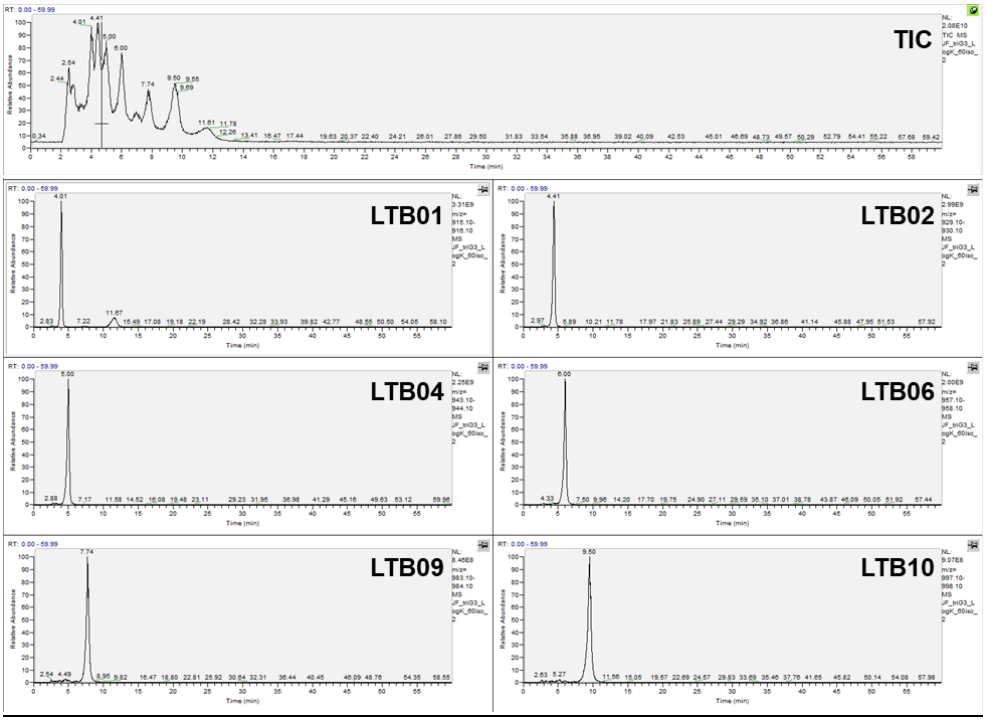
Library G1



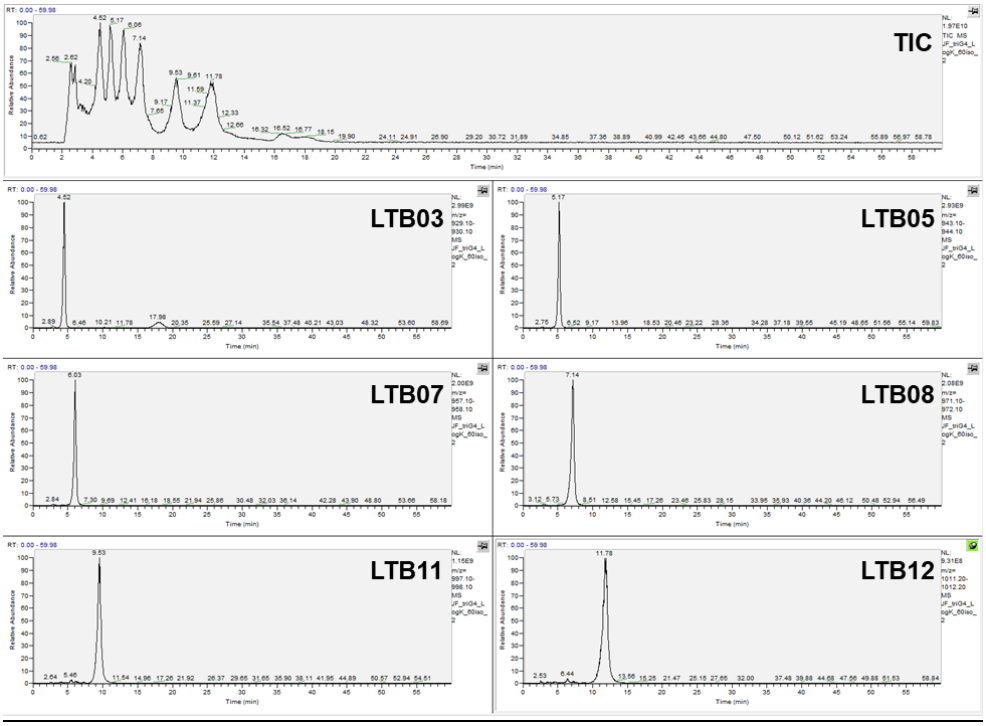
Library G2



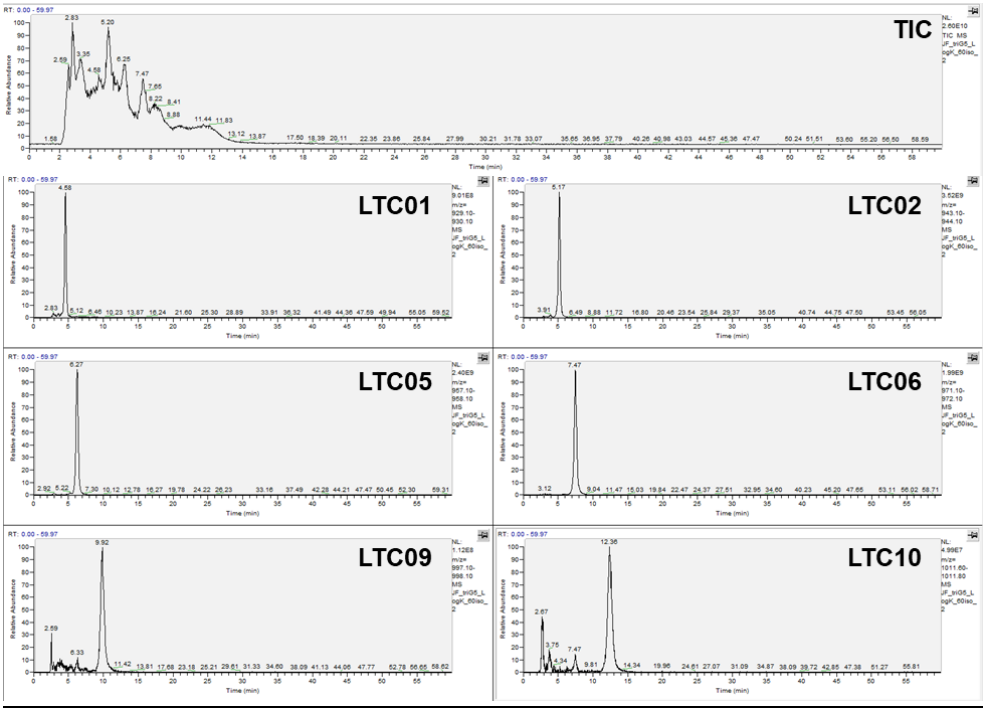
Library G3



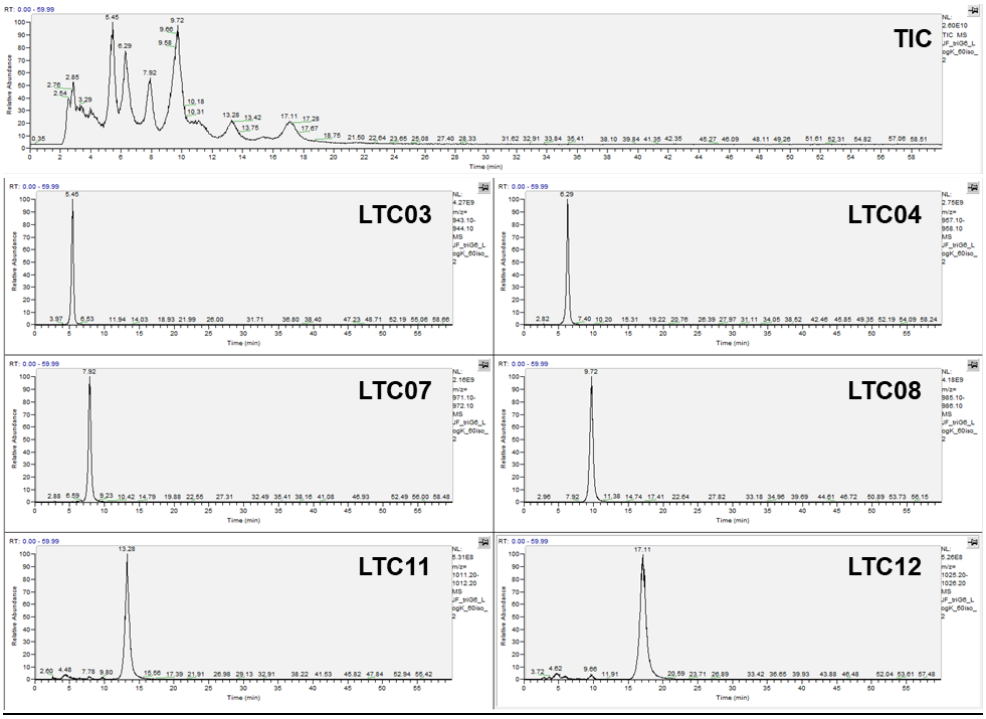
Library G4



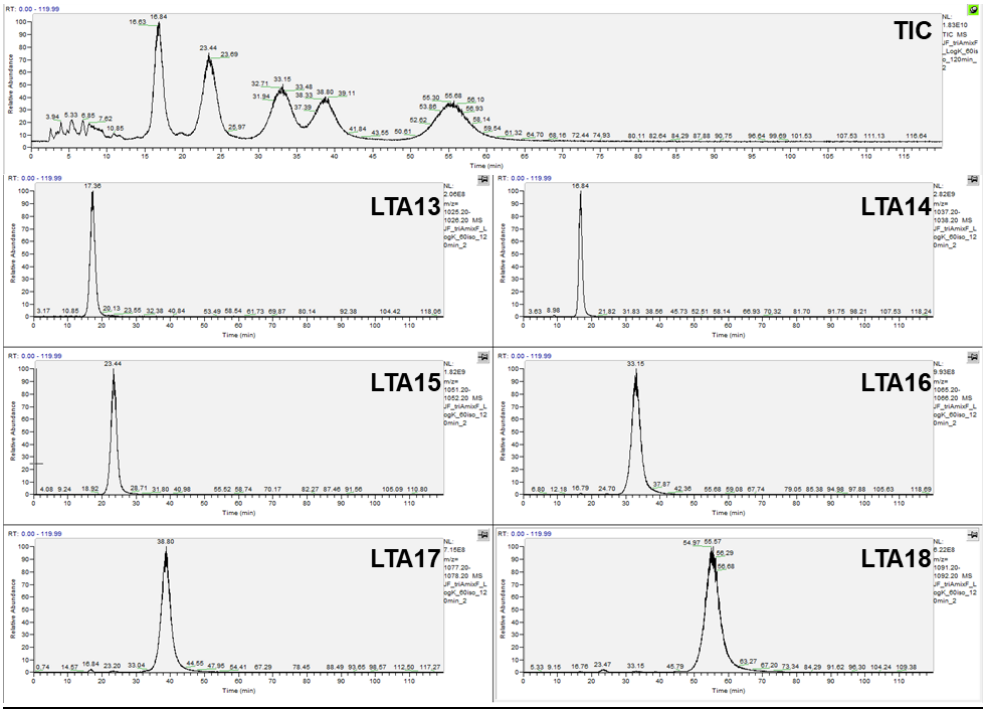
Library G5



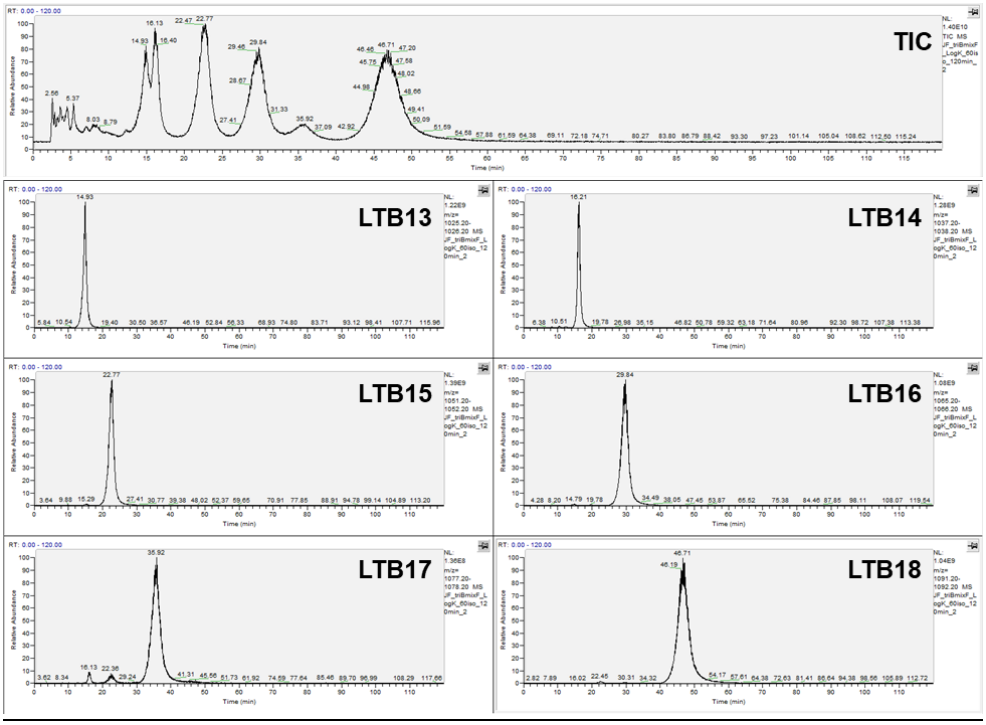
Library G6



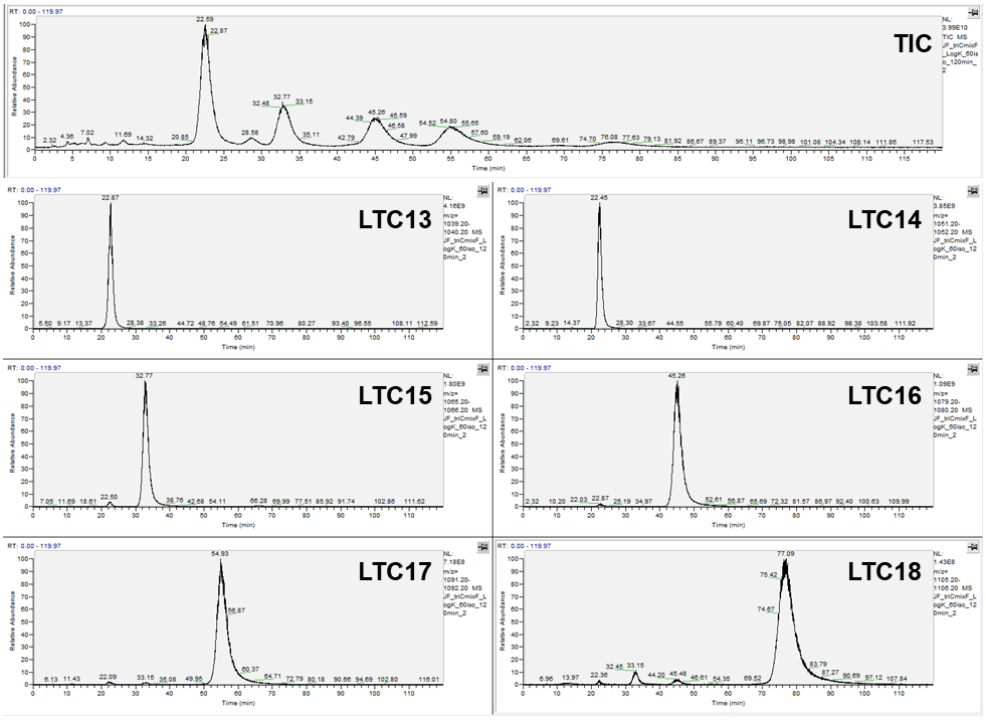
Library Amix



Library Bmix



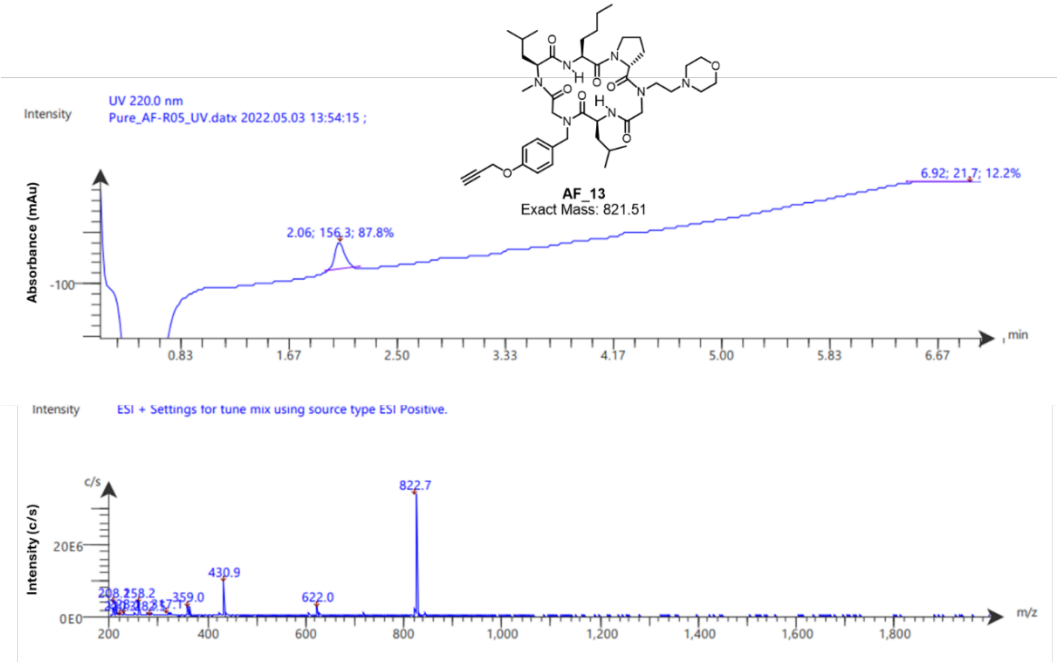
Library Cmix



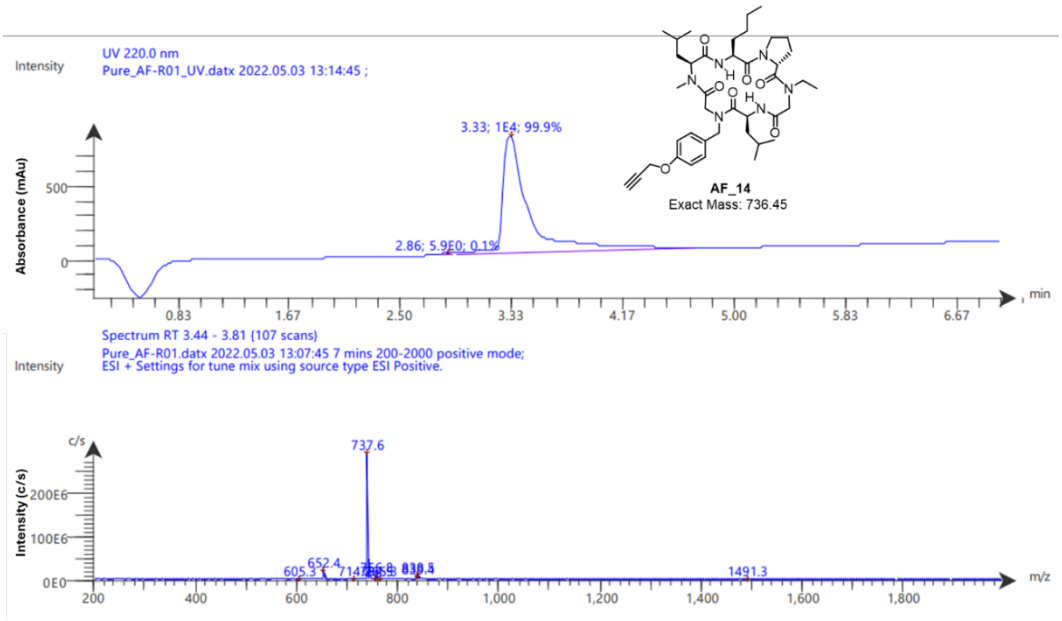
Appendix D: LCMS Traces of DNA-encoded Model Compounds

AF Hexapeptomers (off-DNA)

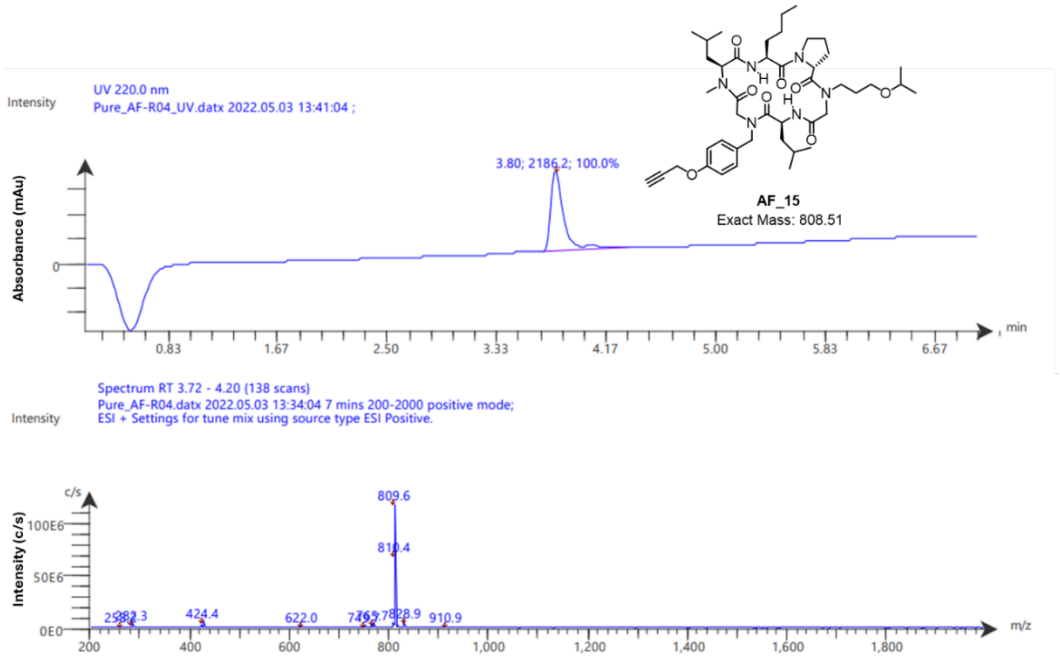
LCMS of compound **AF_13** off-DNA



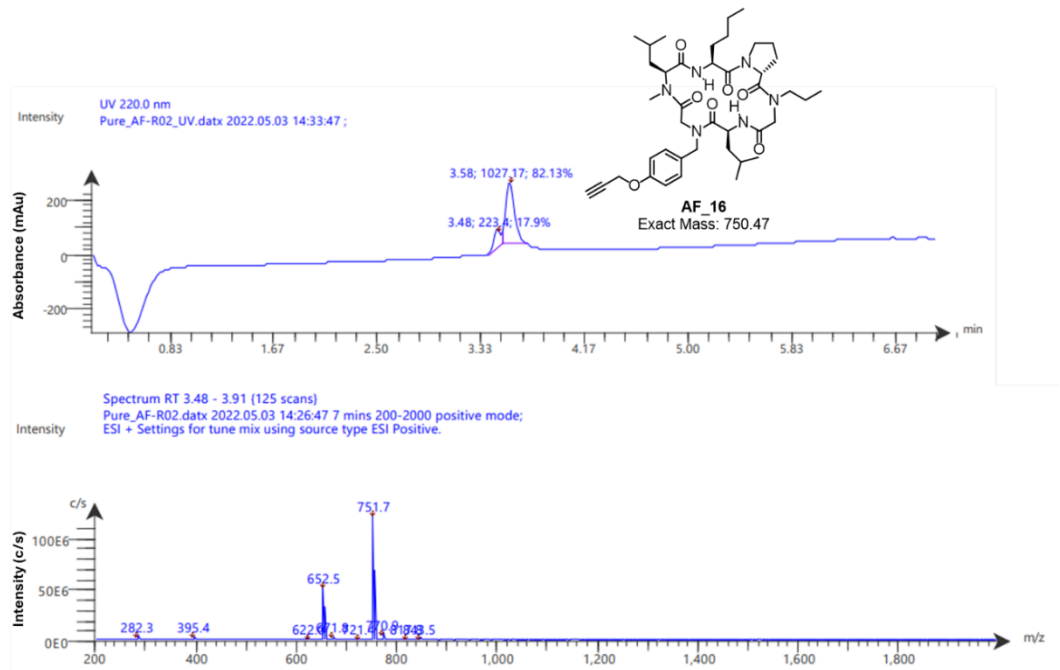
LCMS of compound **AF_14** off-DNA



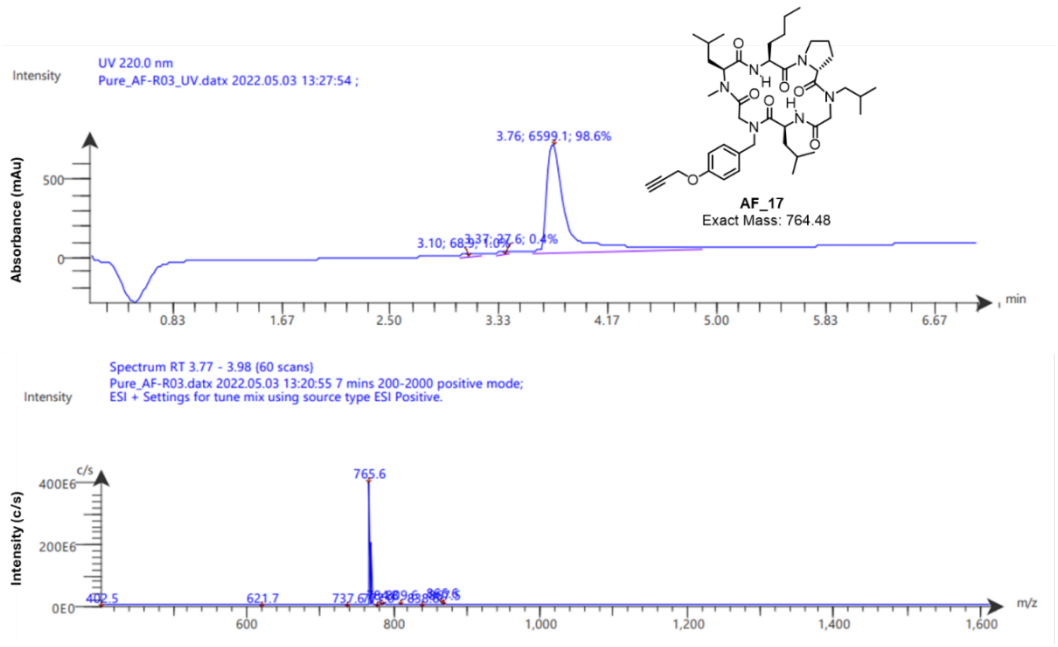
LCMS of compound **AF_15** off-DNA



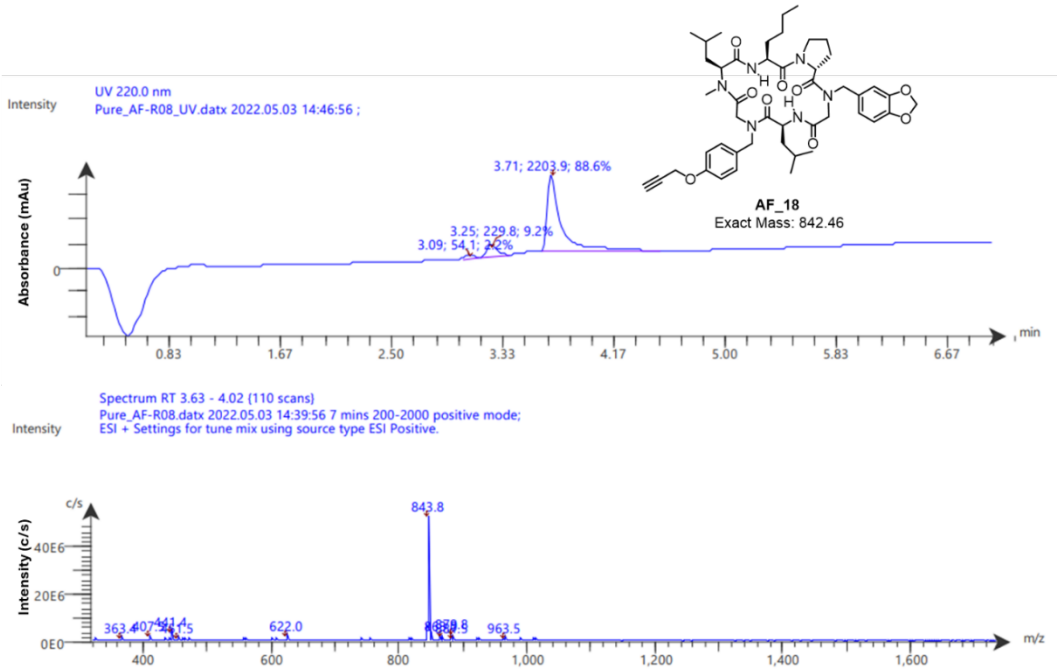
LCMS of compound **AF_16** off-DNA



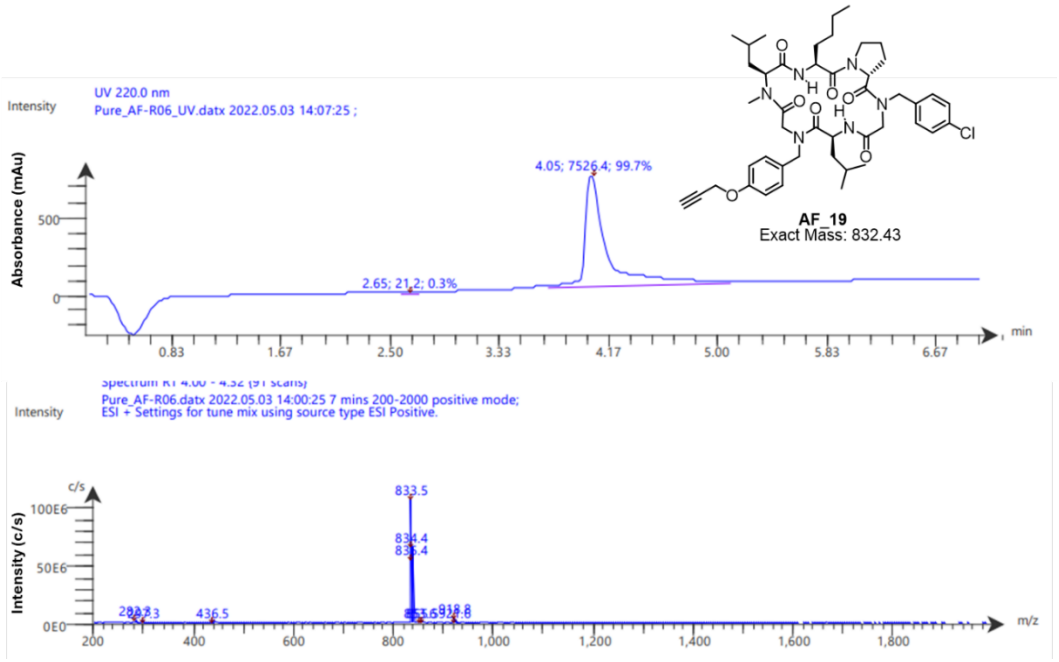
LCMS of compound **AF_17** off-DNA



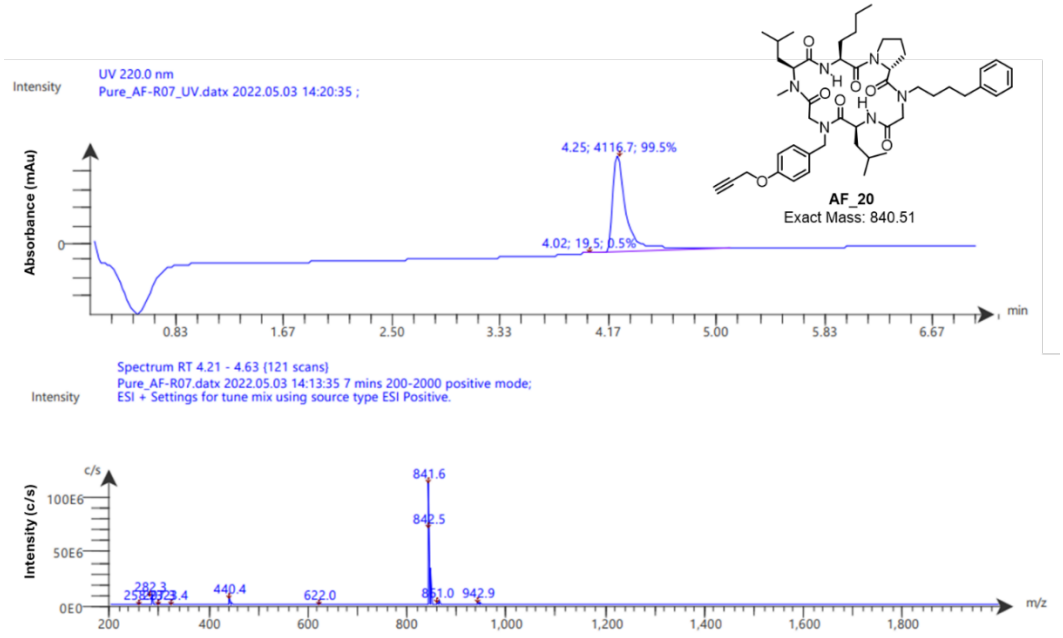
LCMS of compound **AF_18** off-DNA



LCMS of compound **AF_19** off-DNA

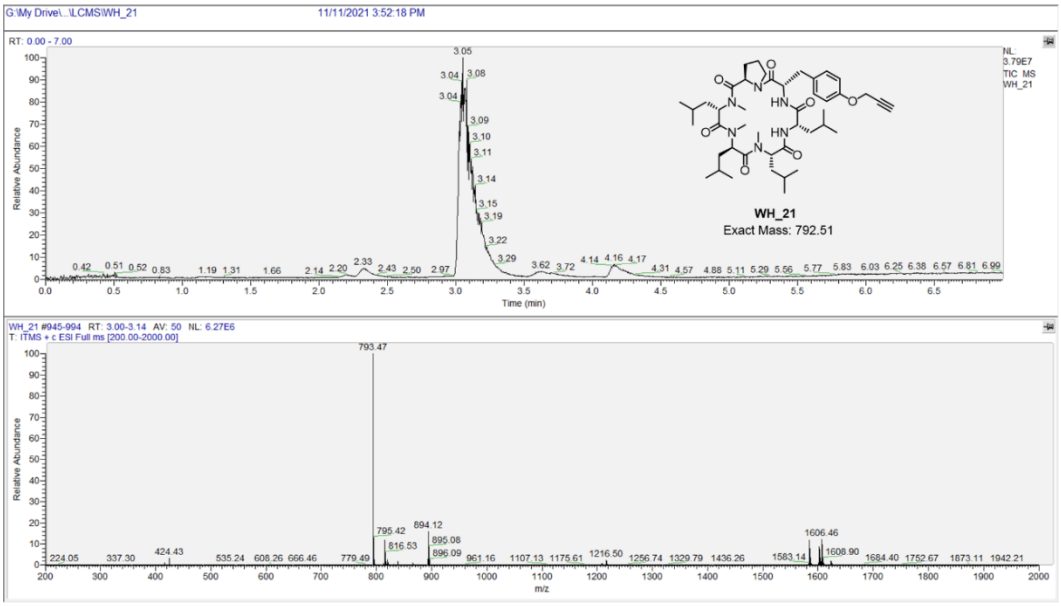


LCMS of compound **AF_20** off-DNA

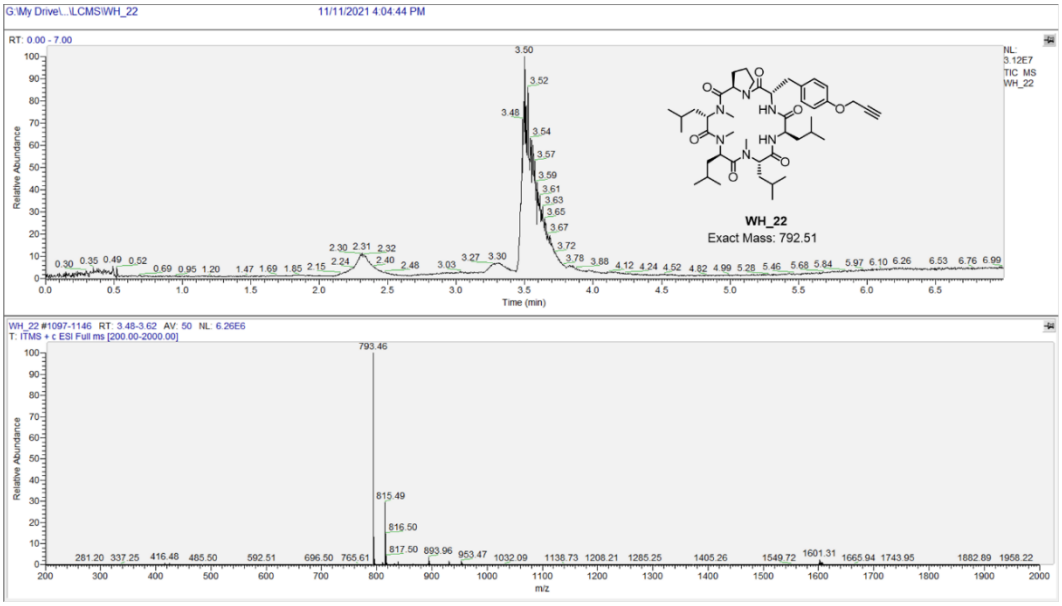


WH Diastereomers (off-DNA)

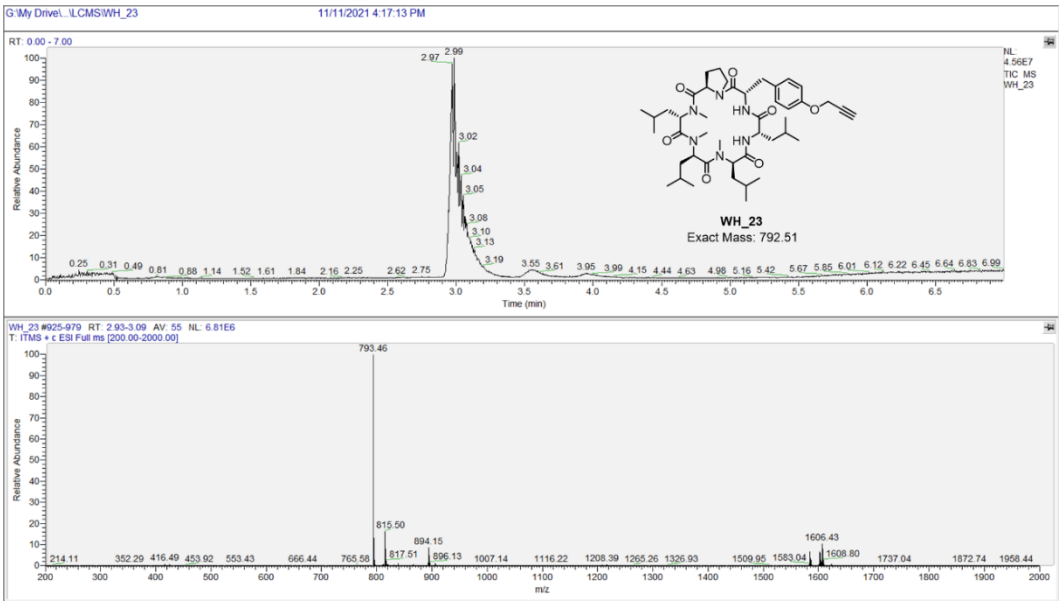
LCMS of compound **WH_21** off-DNA



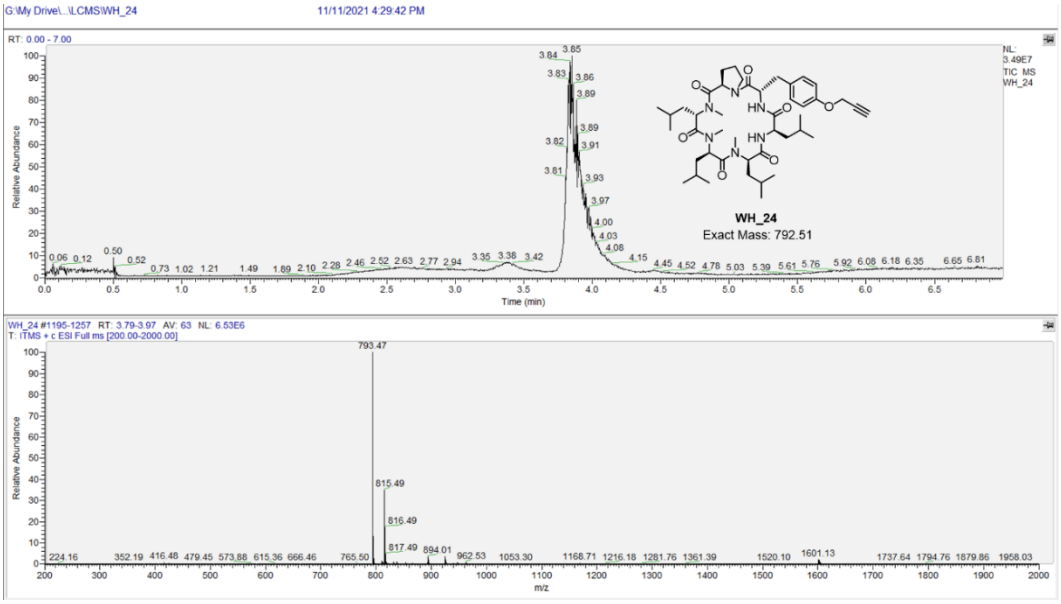
LCMS of compound **WH_22** off-DNA



LCMS of compound **WH_23** off-DNA

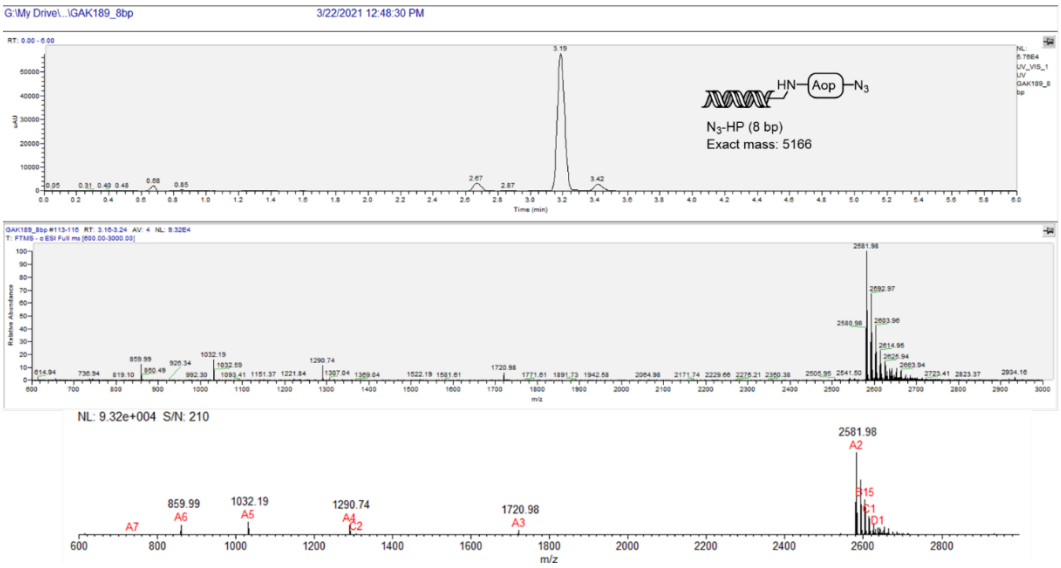


LCMS of compound **WH_24** off-DNA



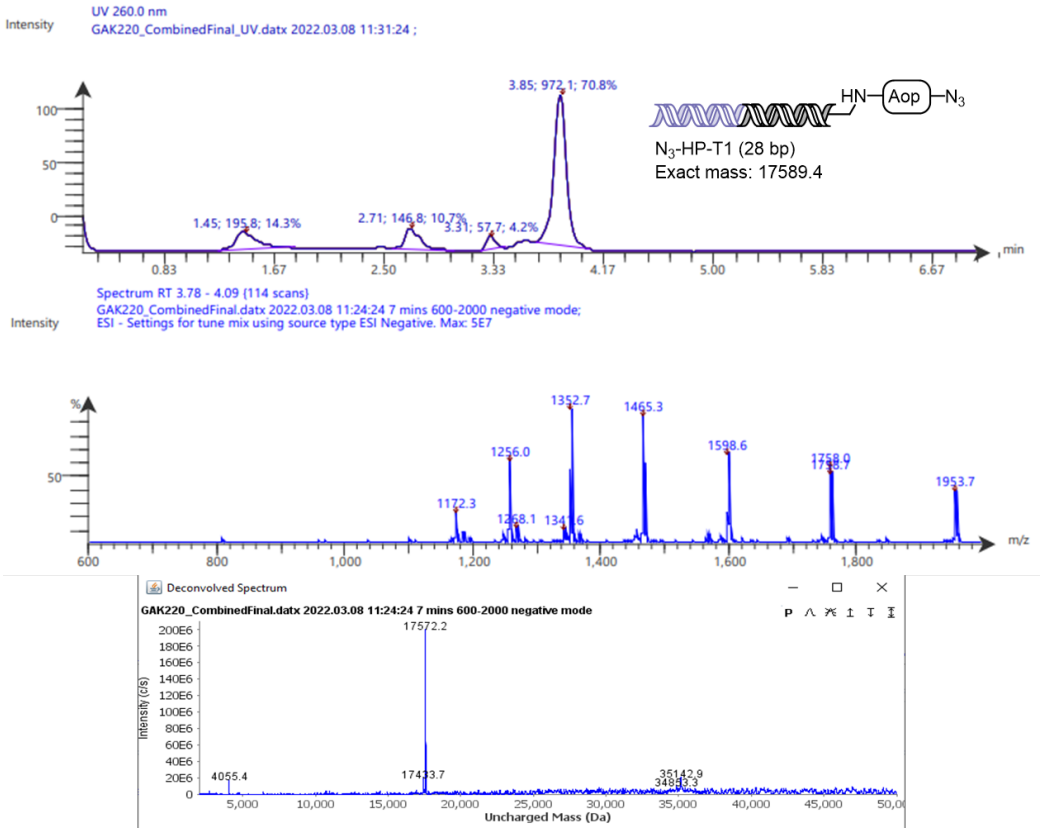
Small Molecule-DNA Conjugates

LCMS of compound **N_3-HP (8 bp)**



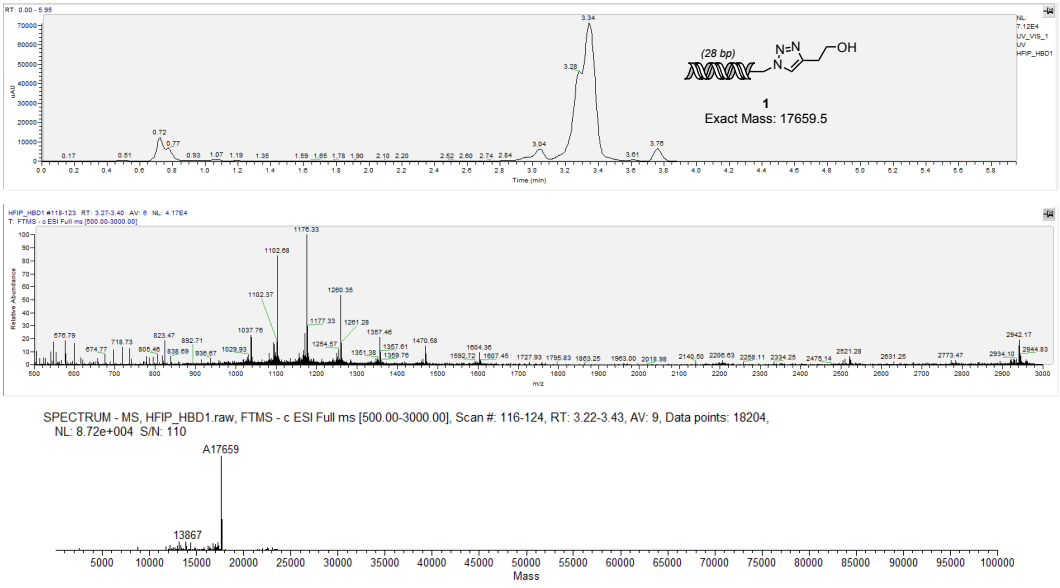
Expected mass: 5166; **Observed mass**: 5165

LCMS of compound **N_3-HP (28 bp)**



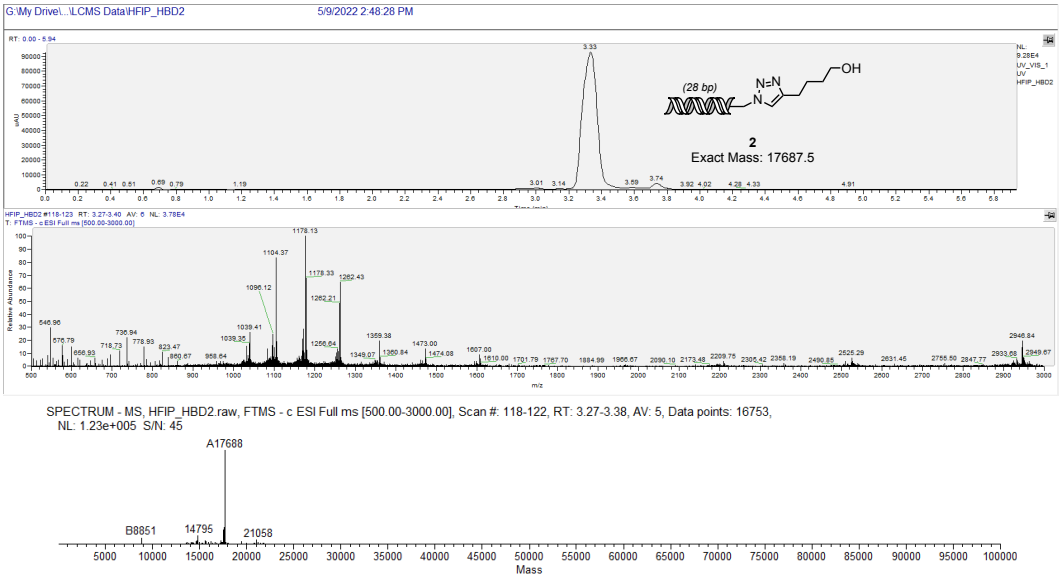
Expected mass: 17589.4; **Observed mass**: 17572.2

LCMS of compound **SM_1**



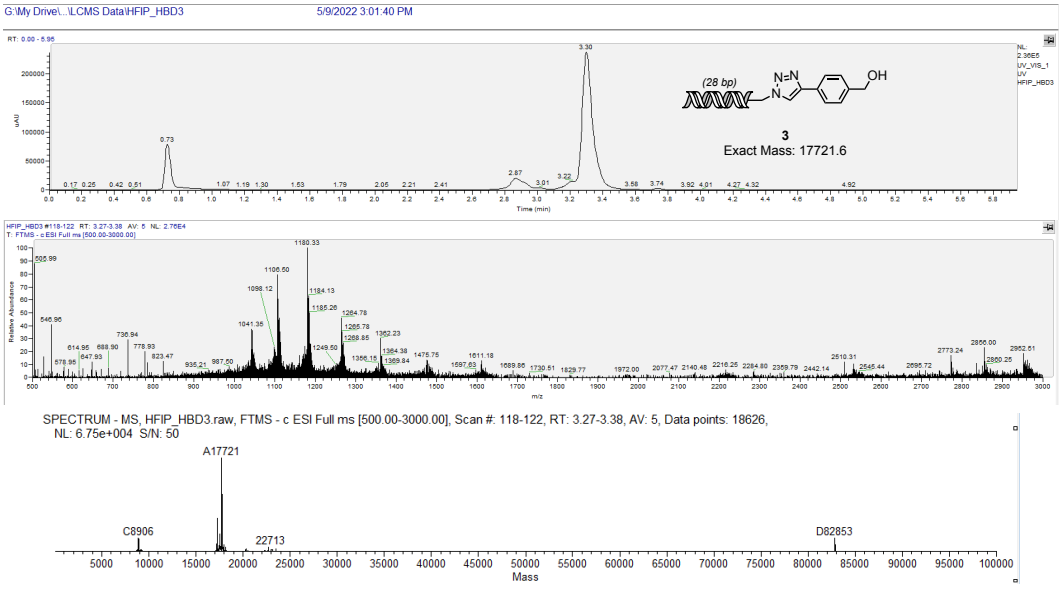
Expected mass: 17659.5; **Observed mass**: 17659

LCMS of compound **SM_2**



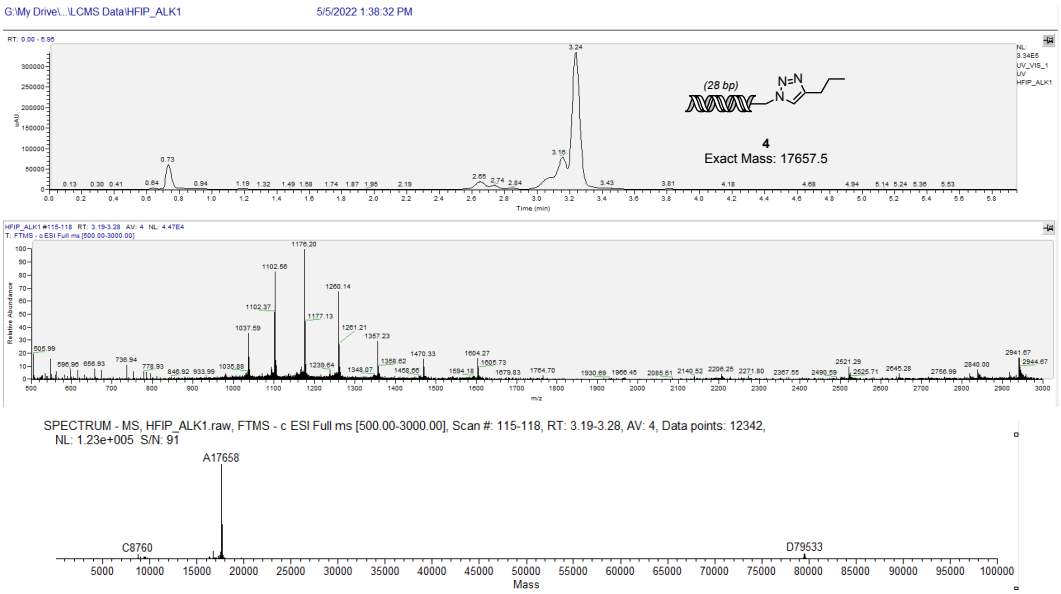
Expected mass: 17687.5; **Observed mass**: 17688

LCMS of compound **SM_3**



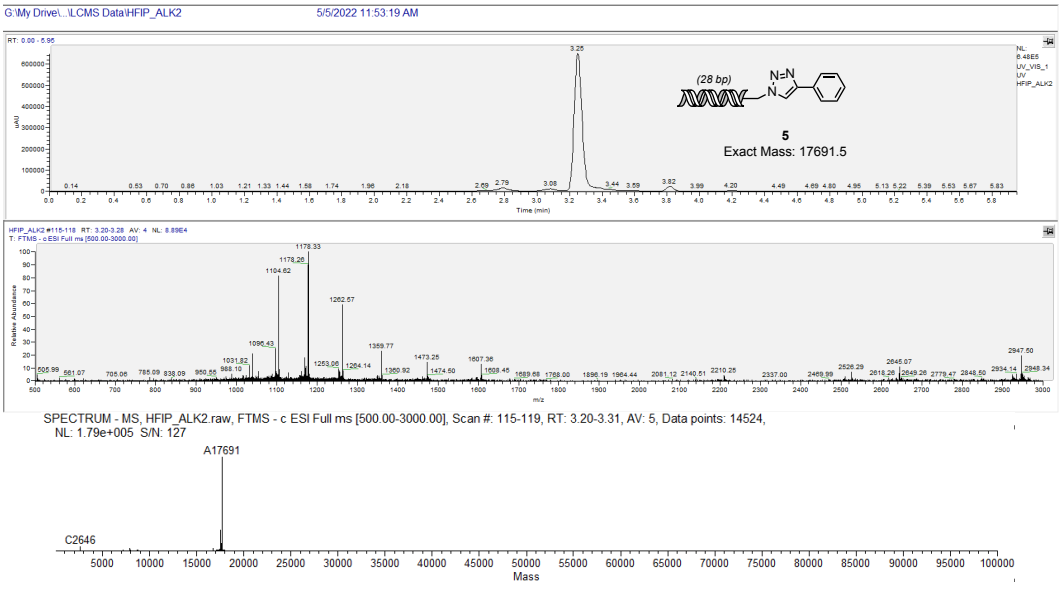
Expected mass: 17721.6; **Observed mass**: 17721

LCMS of compound **SM_4**



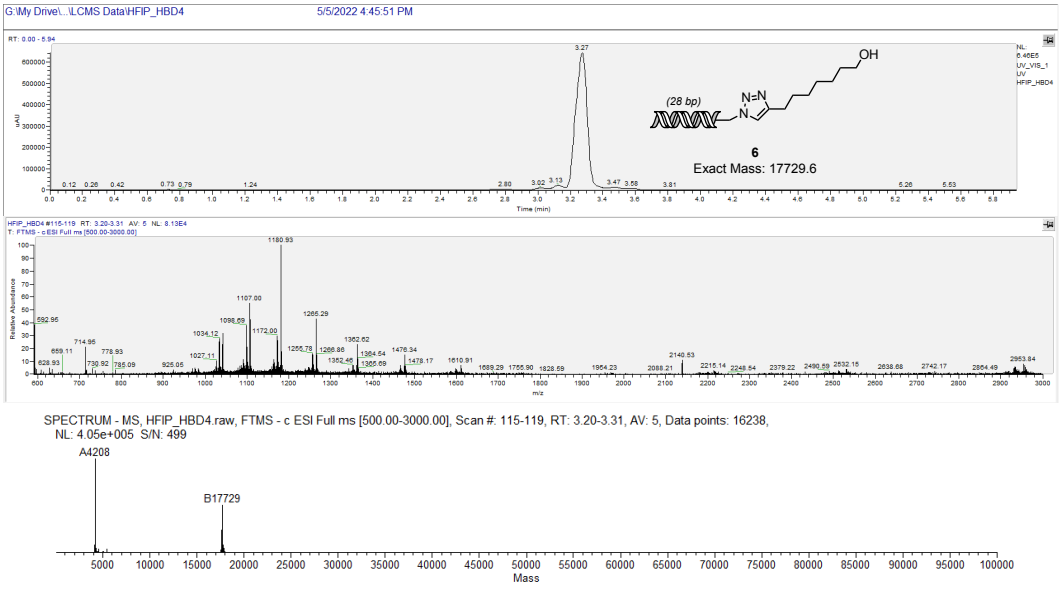
Expected mass: 17657.5; **Observed mass**: 17658

LCMS of compound **SM_5**



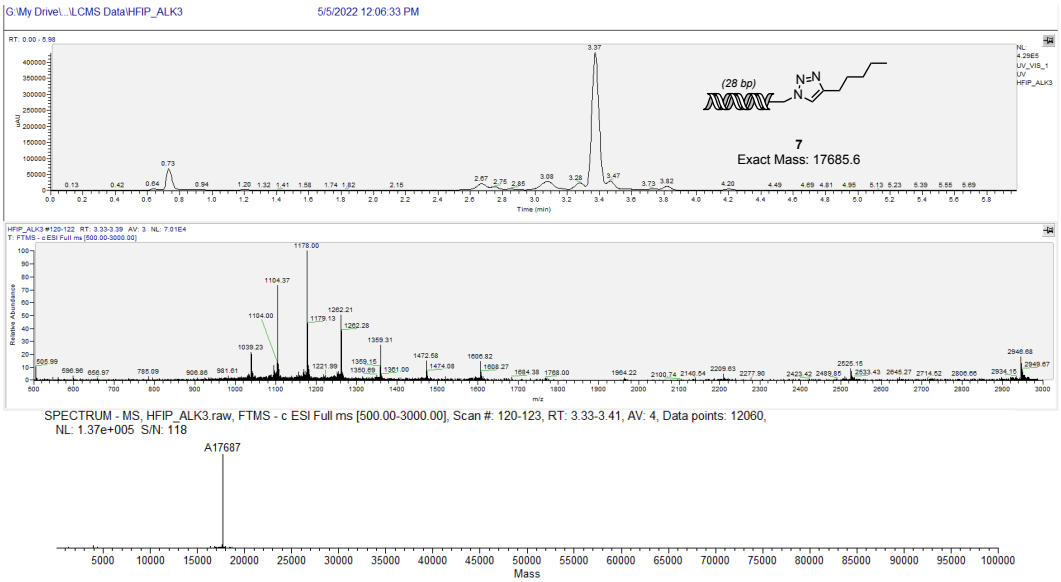
Expected mass: 17691.5; **Observed mass**: 17691

LCMS of compound **SM_6**



Expected mass: 17729.6; **Observed mass**: 17729

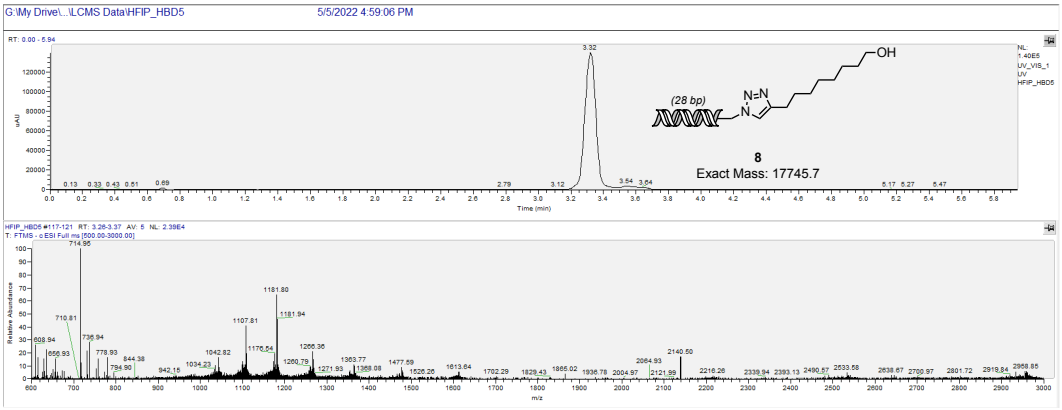
LCMS of compound **SM_7**



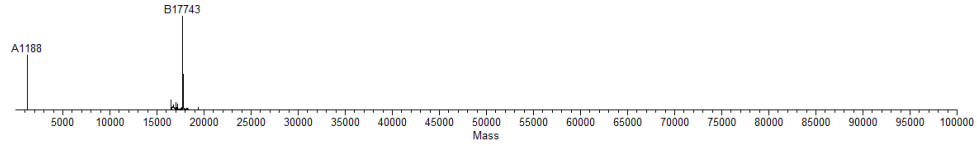
SPECTRUM - MS, HFIP_ALK3.raw, FTMS - c ESI Full ms [500.00-3000.00], Scan #: 120-123, RT: 3.33-3.41, AV: 4, Data points: 12060, NL: 1.37e+005 S/N: 118

Expected mass: 17685.6; **Observed mass**: 17687

LCMS of compound **SM_8**



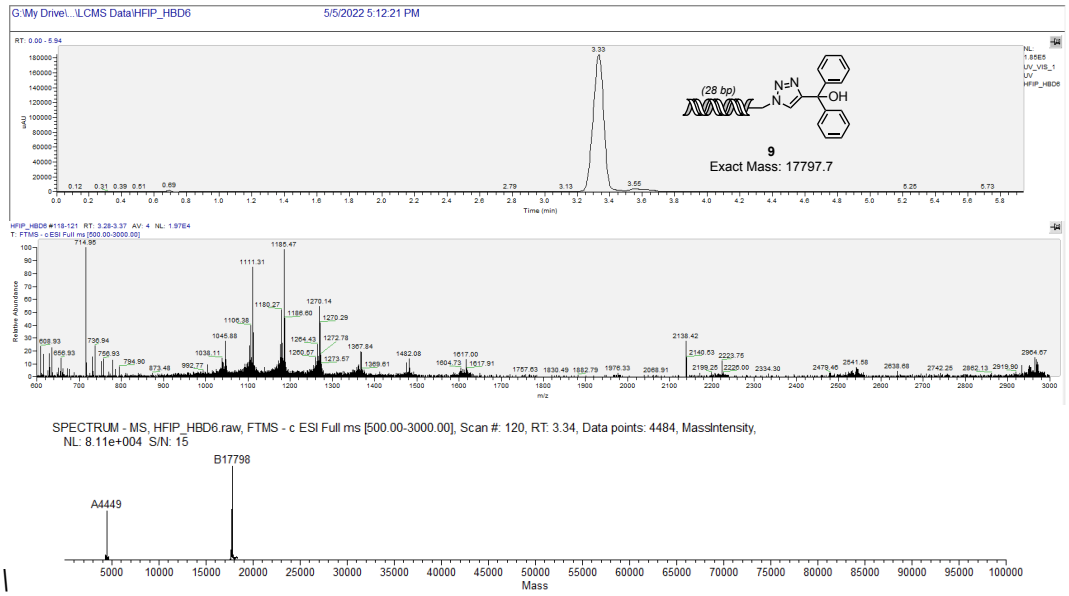
SPECTRUM - MS, HFIP_HBD5.raw, FTMS - c ESI Full ms [500.00-3000.00], Scan #: 119, RT: 3.31, Data points: 4558, MassIntensity, NL: 6.07e+004 S/N: 13



\

Expected mass: 17745.7; **Observed mass**: 17743

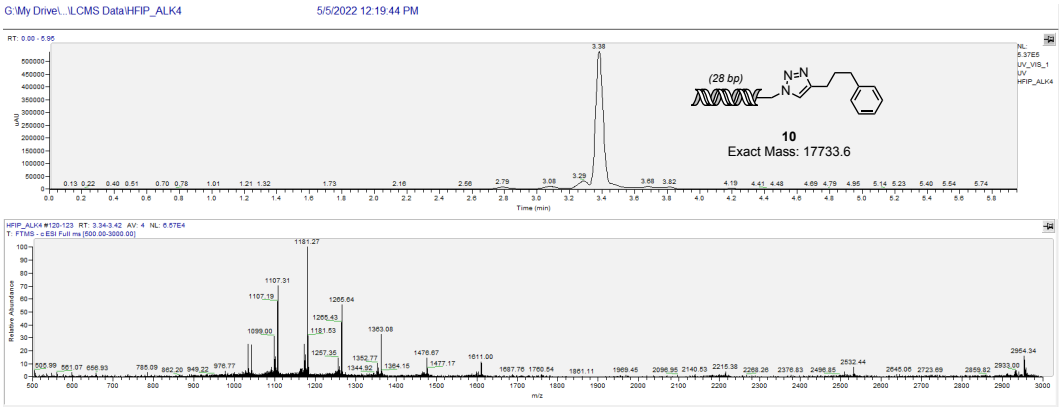
LCMS of compound **SM_9**

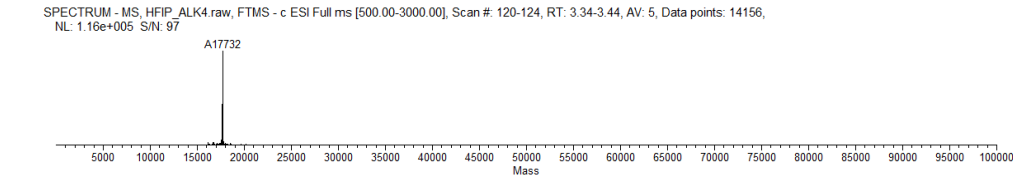


\

Expected mass: 17797.7; **Observed mass**: 17798

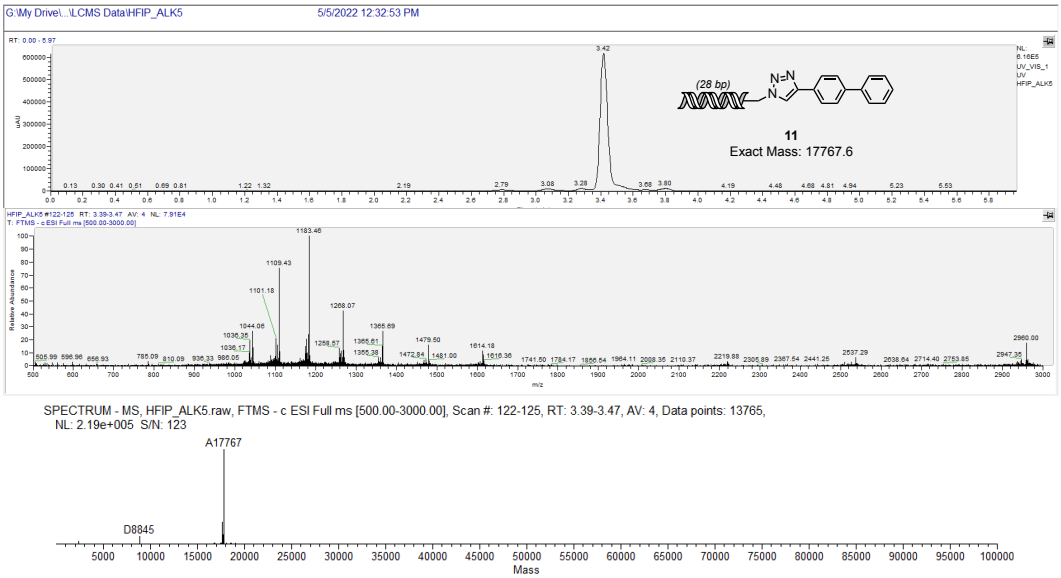
LCMS of compound **SM_10**





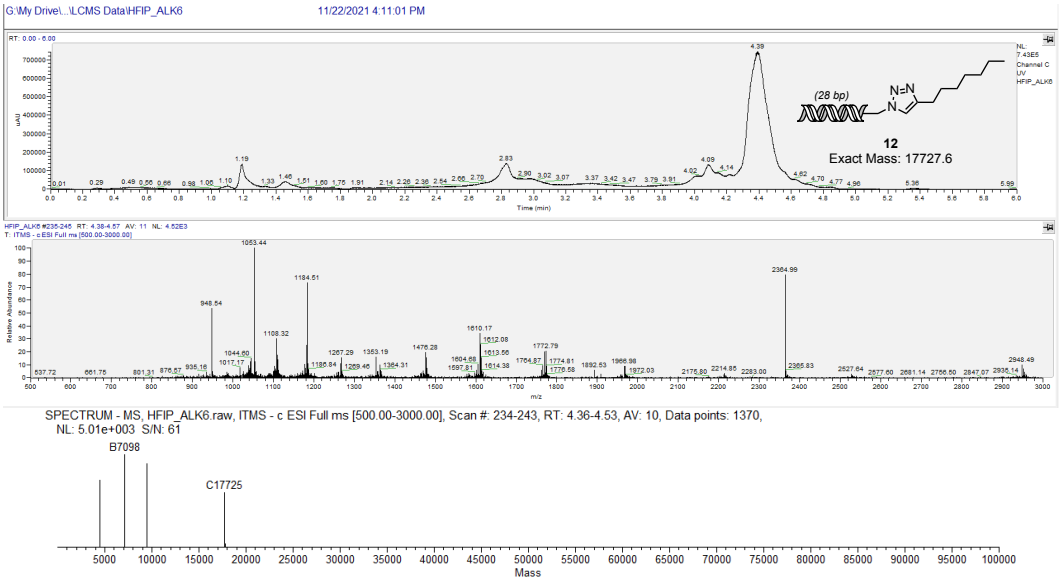
Expected mass: 17733.6; **Observed mass**: 17732

LCMS of compound **SM_11**



Expected mass: 17767.6; **Observed mass**: 17767

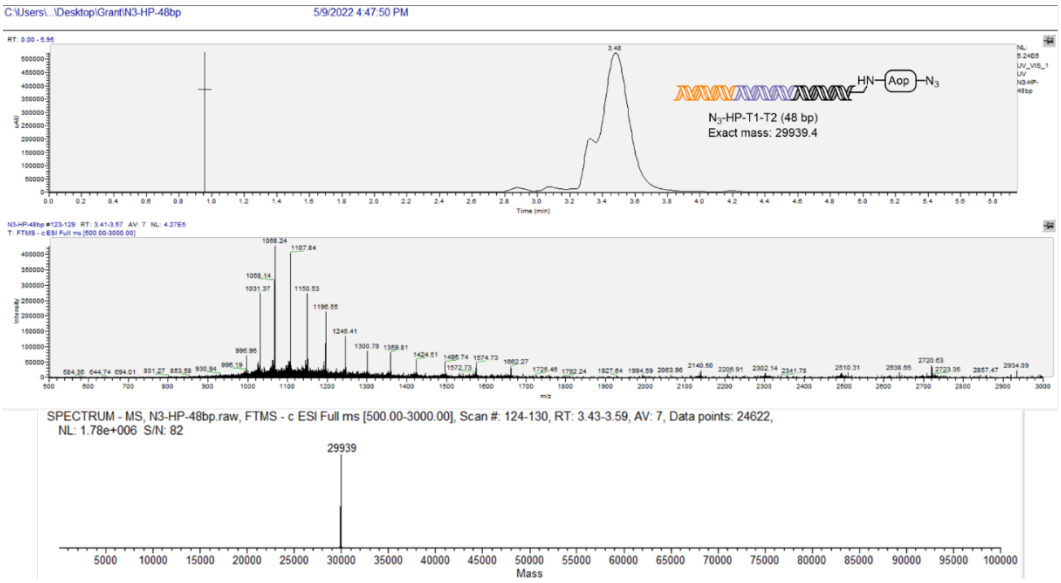
LCMS of compound **SM_12**



Expected mass: 17727.6; **Observed mass**: 17725

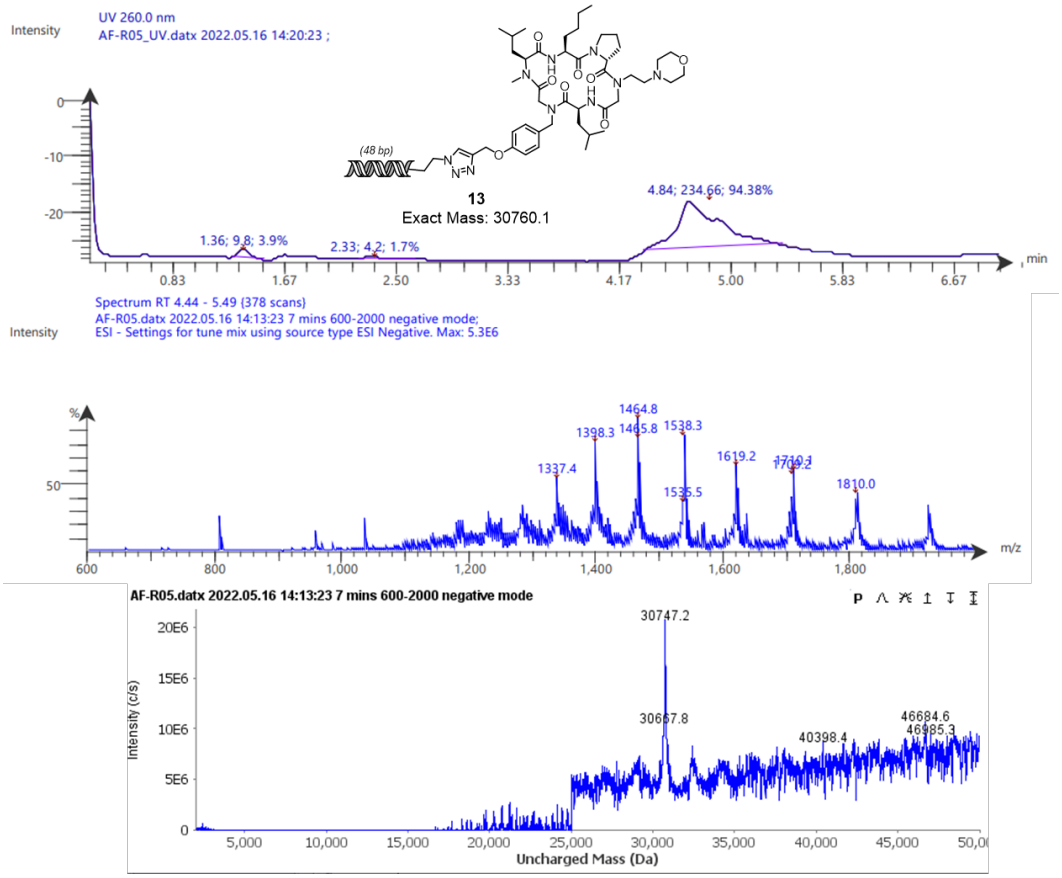
AF-DNA Conjugates

LCMS of compound **N_3-HP (48 bp)**



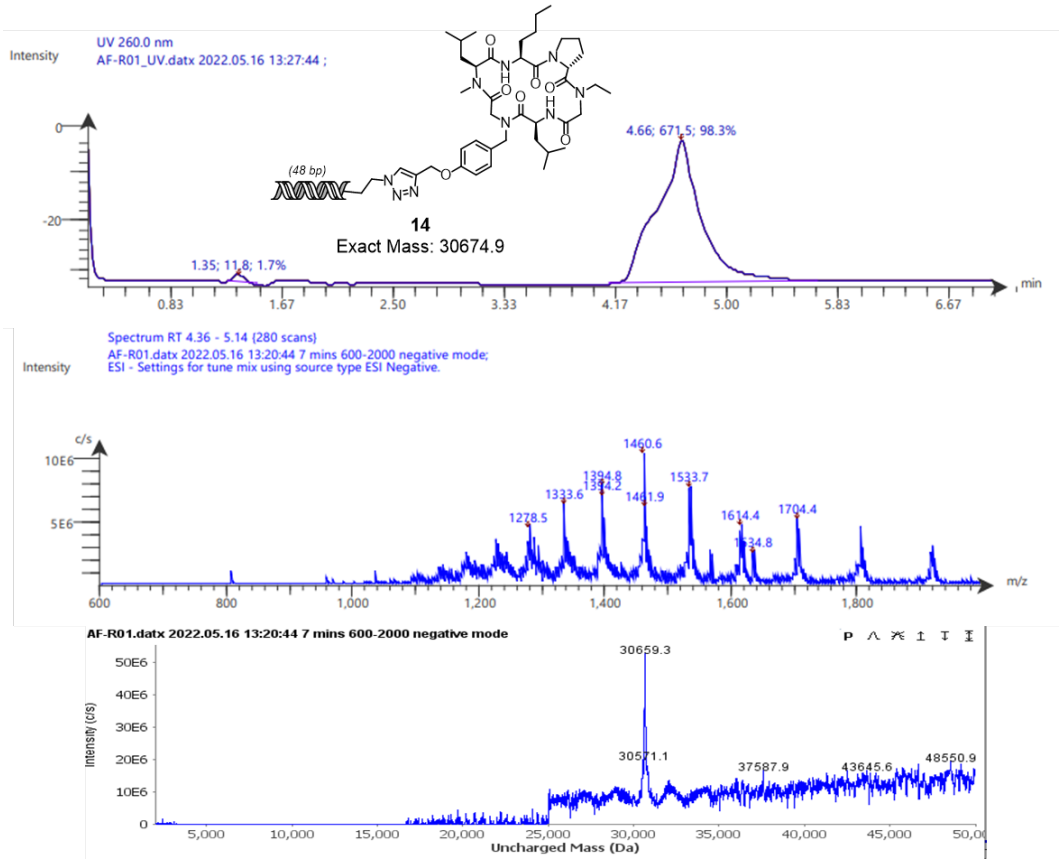
Expected mass: 29939.4; **Observed mass**: 29939

LCMS of compound **AF_13**



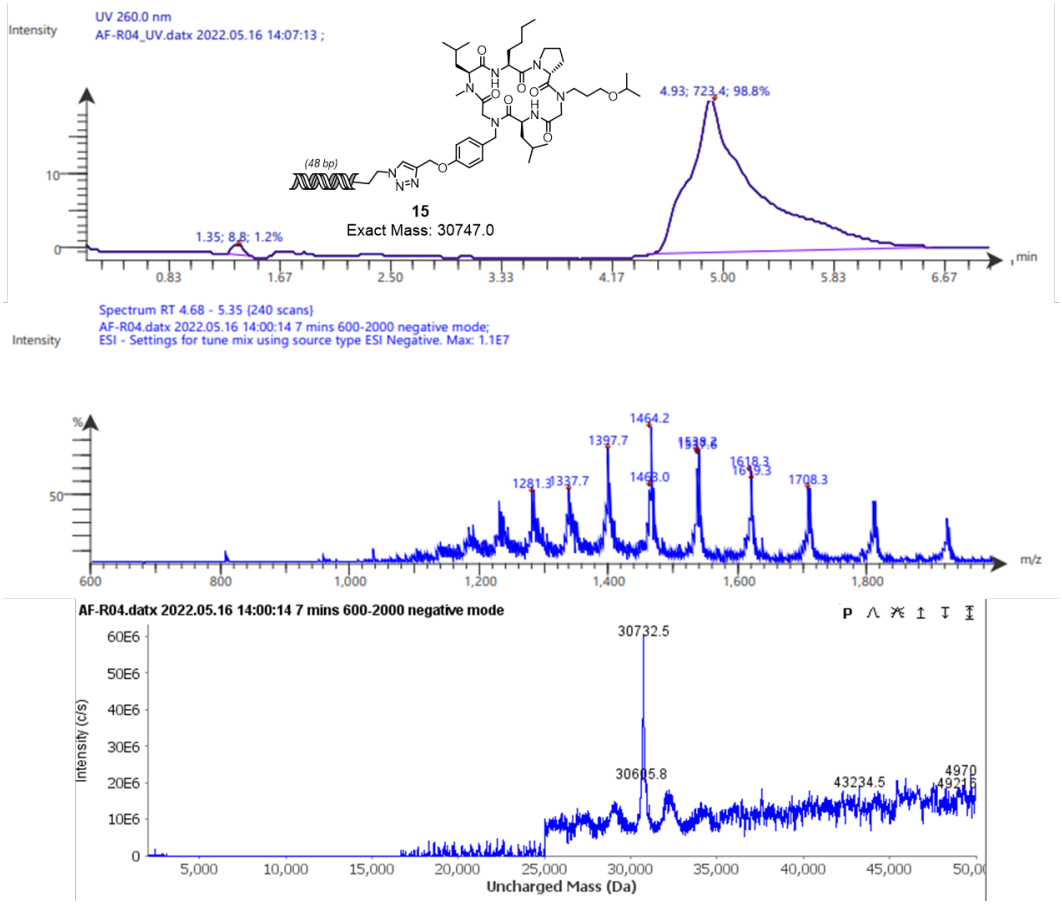
Expected mass: 30760.1; **Observed mass**: 30747.2

LCMS of compound **AF_14**



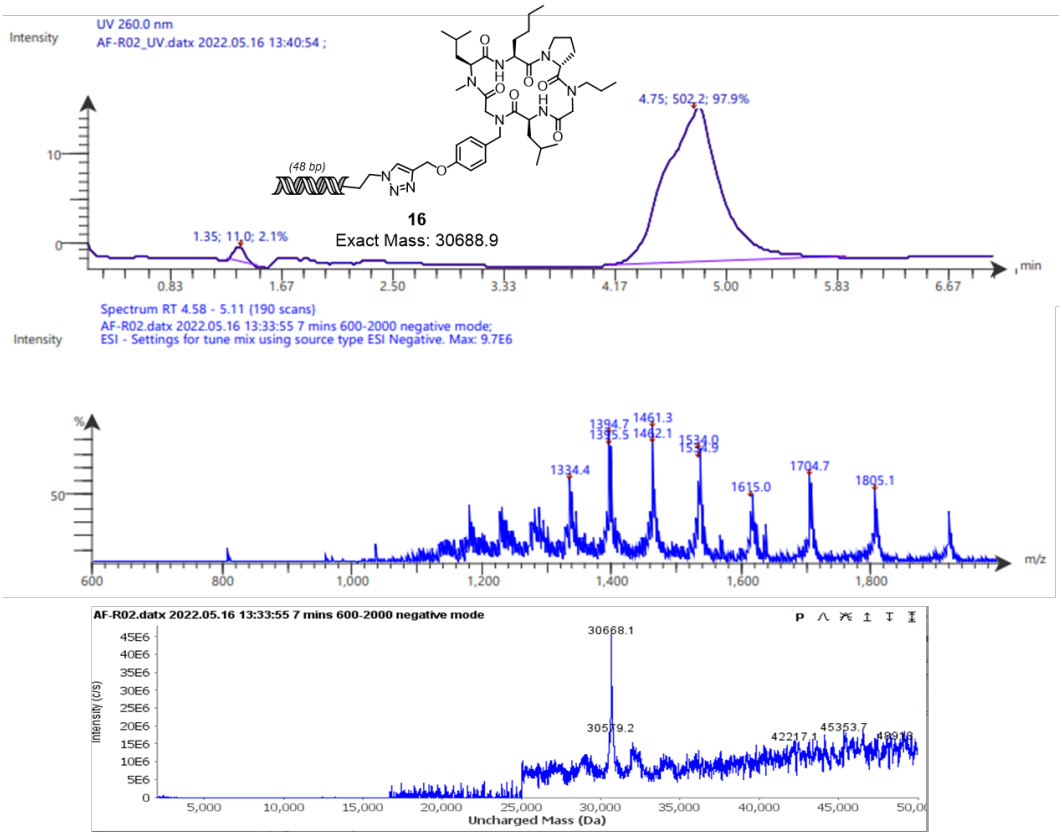
Expected mass: 307674.9; **Observed mass**: 30659.3

LCMS of compound **AF_15**



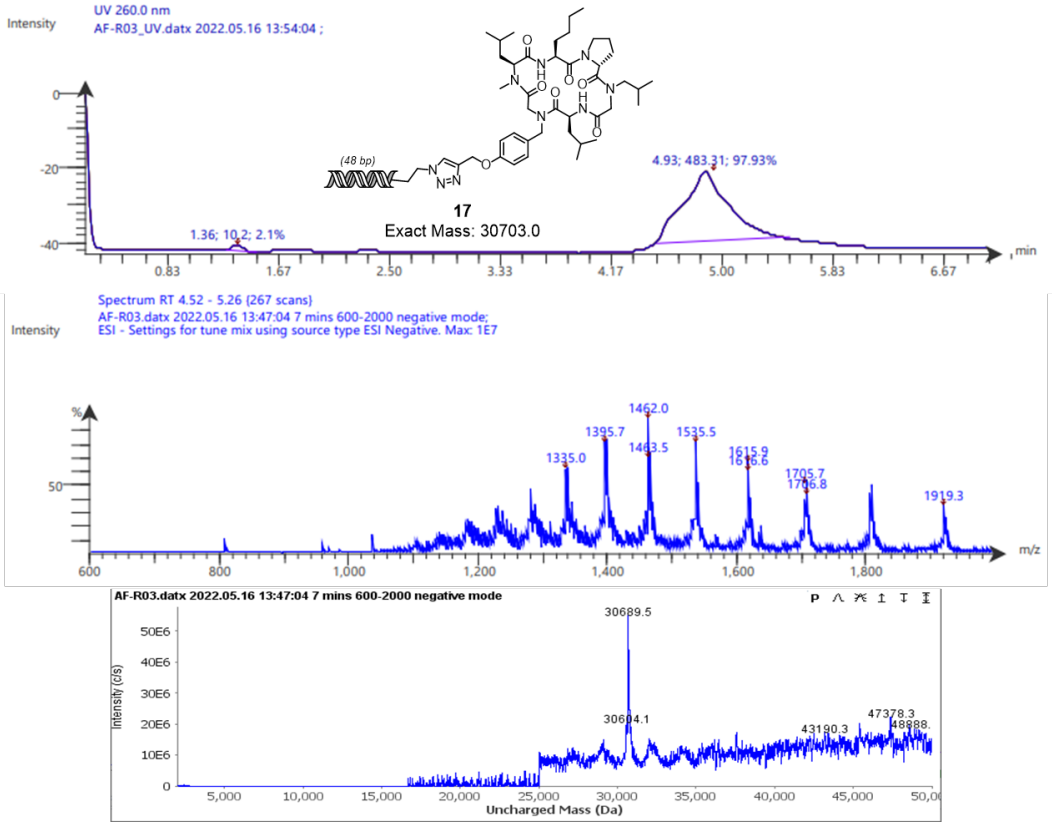
Expected mass: 30747.0; **Observed mass**: 30732.5

LCMS of compound **AF_16**



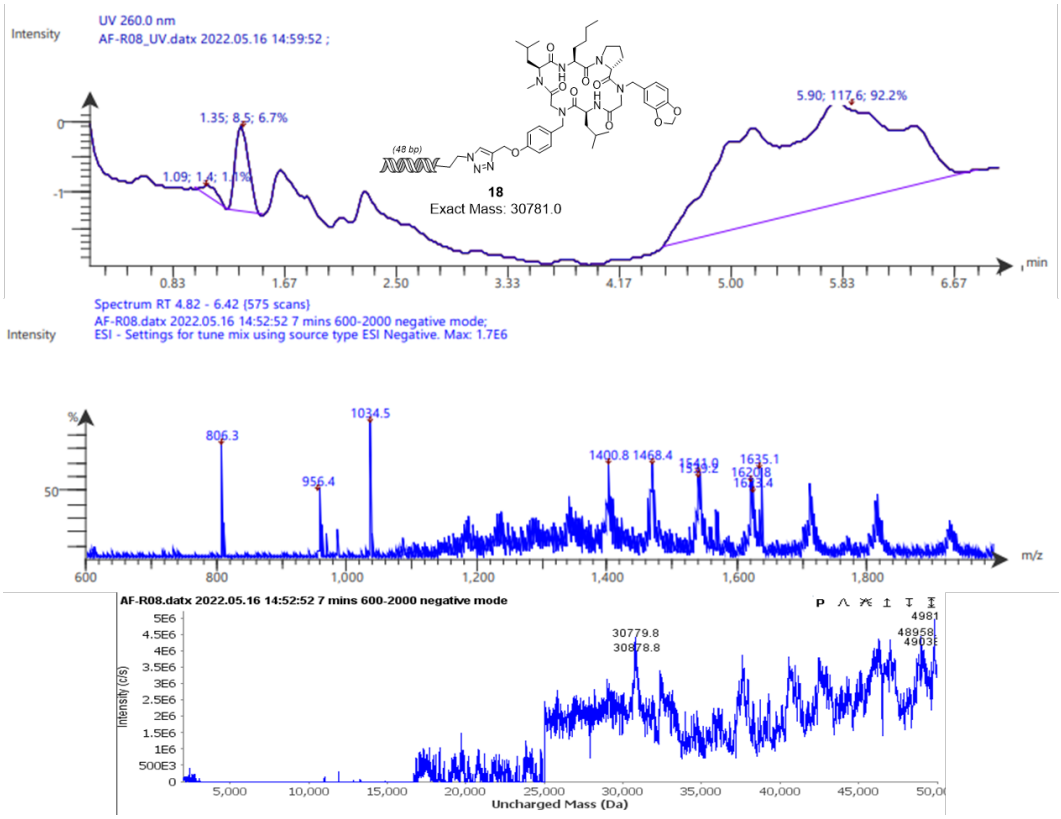
Expected mass: 30688.9; **Observed mass**: 30668.1

LCMS of compound **AF_17**



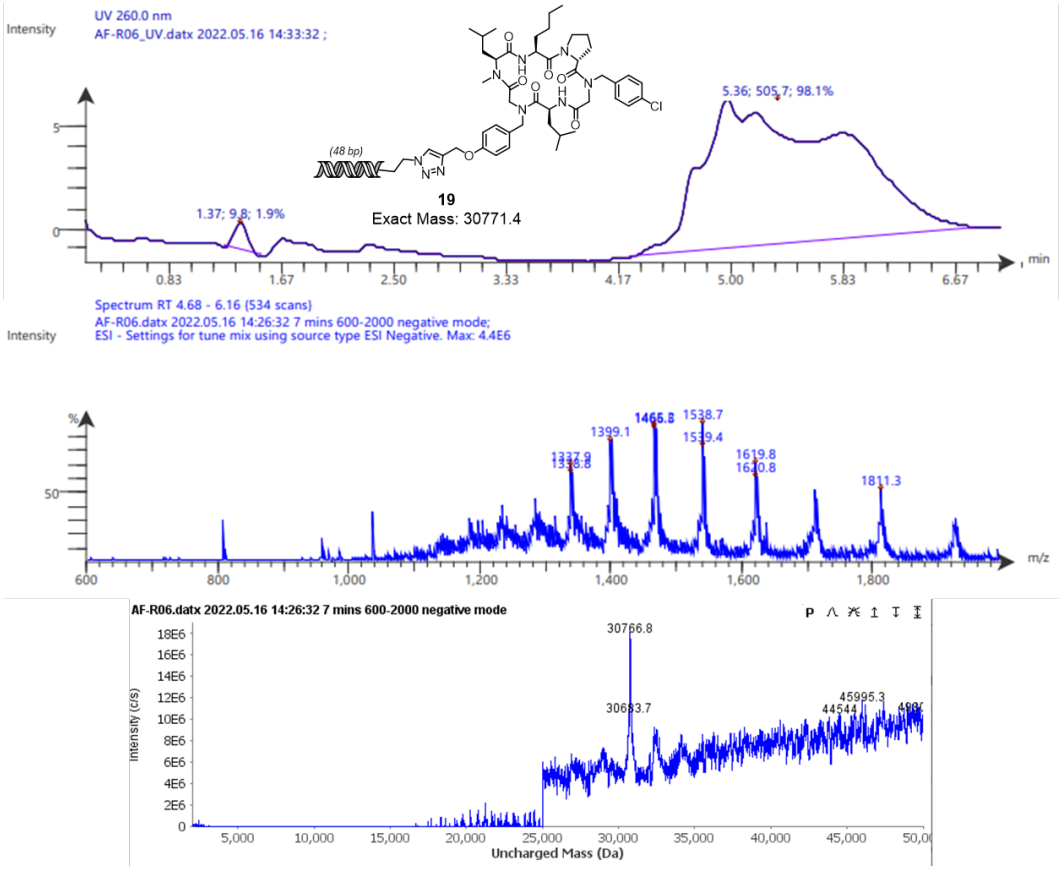
Expected mass: 30703.0; **Observed mass**: 30689.5

LCMS of compound **AF_18**



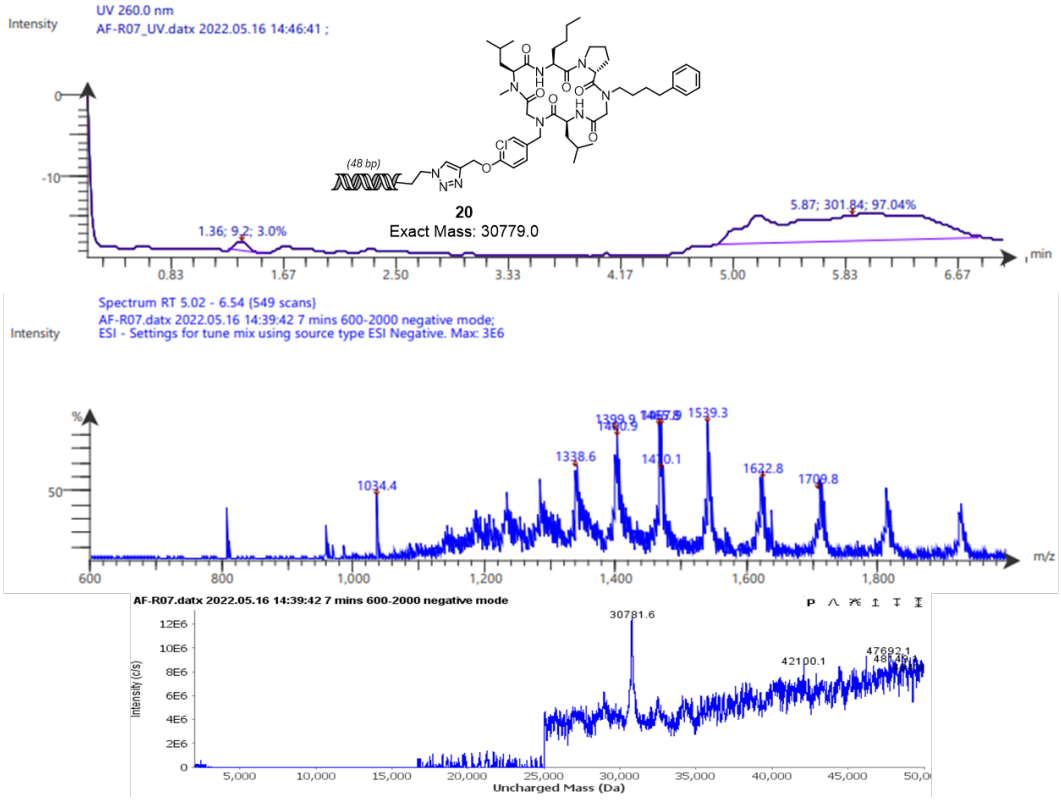
Expected mass: 30781.0; **Observed mass**: 30779.8

LCMS of compound **AF_19**



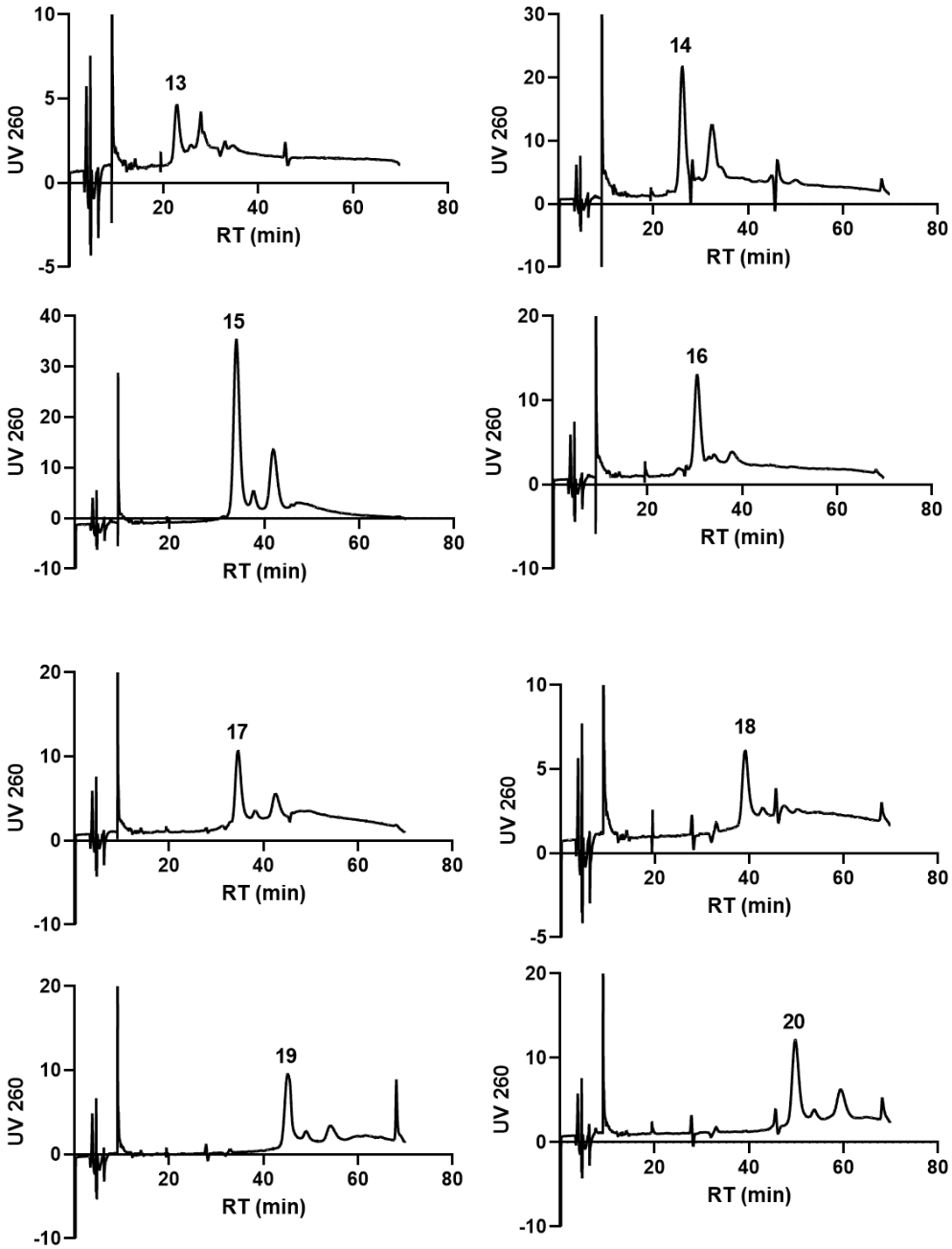
Expected mass: 30771.4; **Observed mass**: 30766.8

LCMS of compound **AF_20**



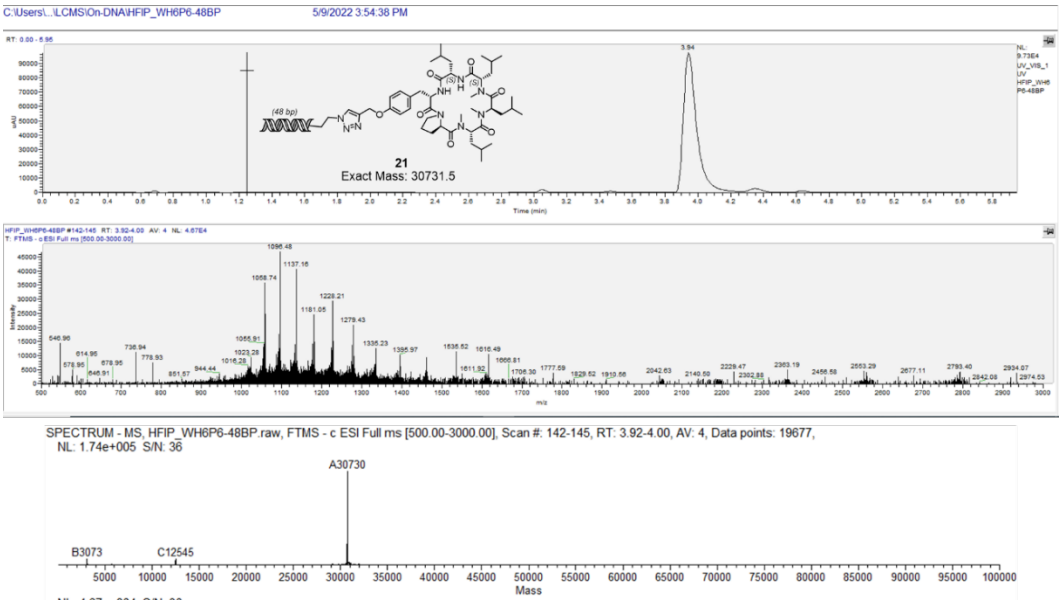
Expected mass: 30779.0; **Observed mass**: 30781.6

AF-**13-20** Single Compound Injections (Same Conditions as Coinjection)



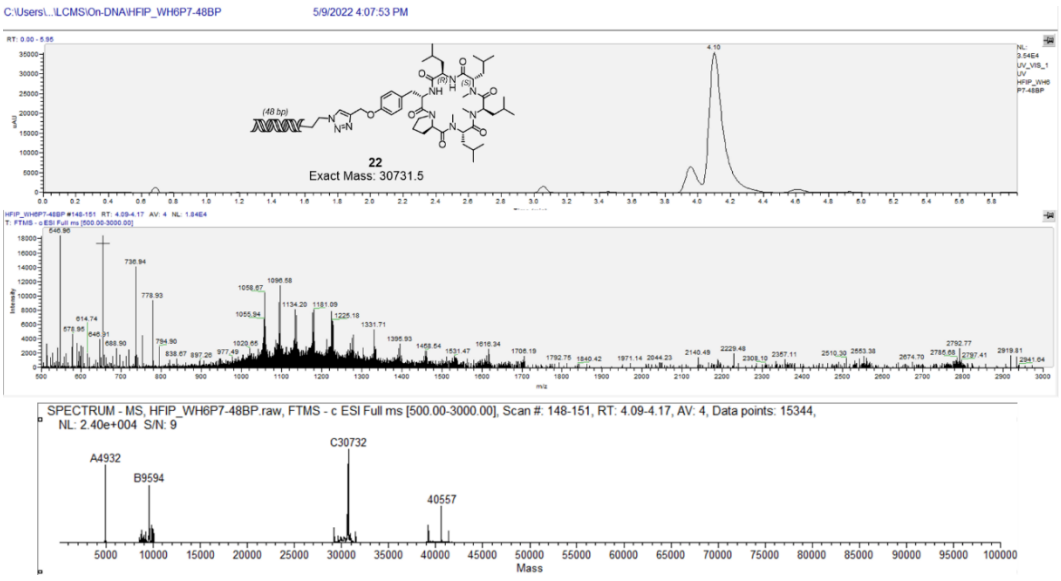
WH-DNA Conjugates

LCMS of compound **WH_21**



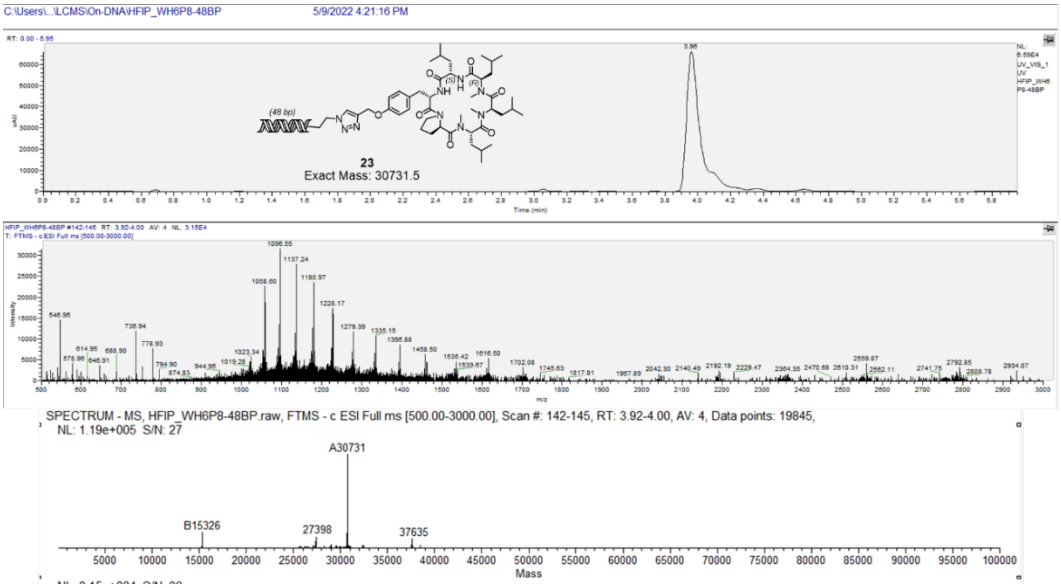
Expected mass: 30731.5; **Observed mass**: 30730

LCMS of compound **WH_22**



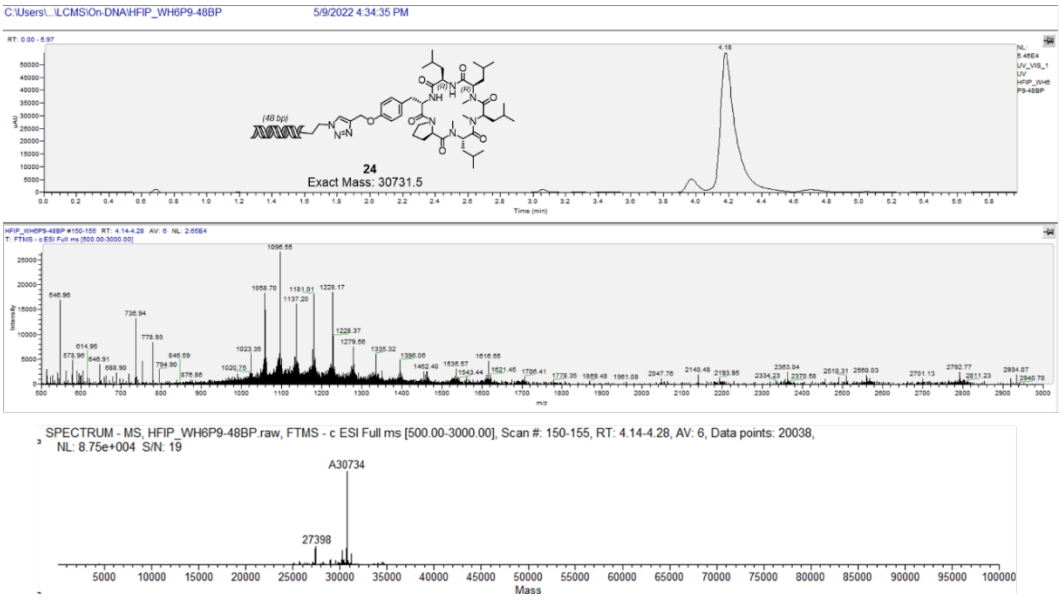
Expected mass: 30731.5; **Observed mass**: 30732

LCMS of compound **WH_23**



Expected mass: 30731.5; **Observed mass**: 30731

LCMS of compound **WH_24**

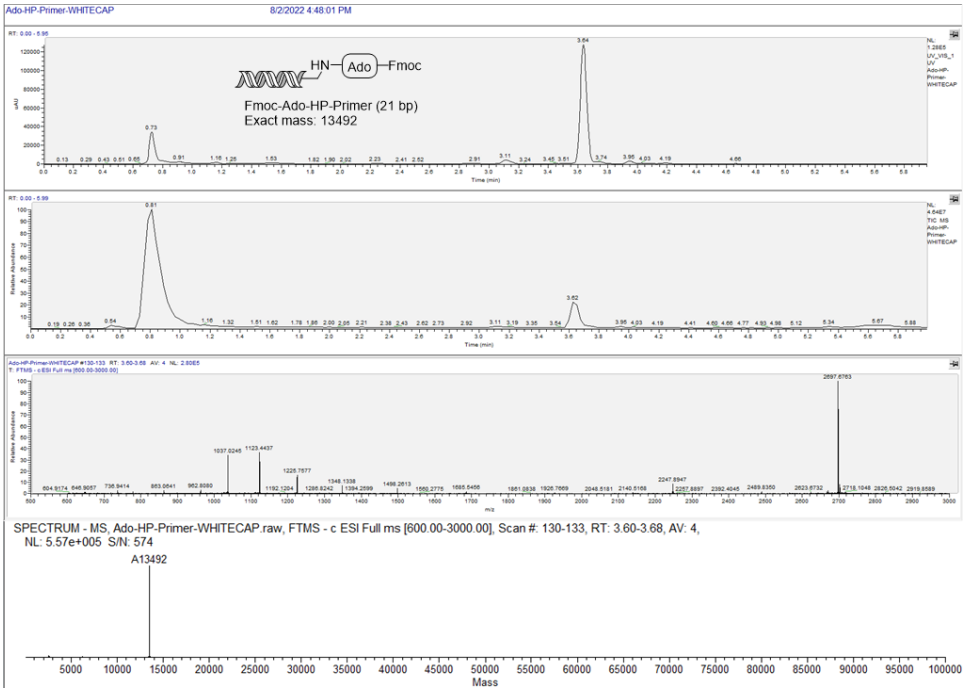


Expected mass: 30731.5; **Observed mass**: 30734

Appendix E: LCMS Traces and Gels of DNA-encoded Library Synthesis

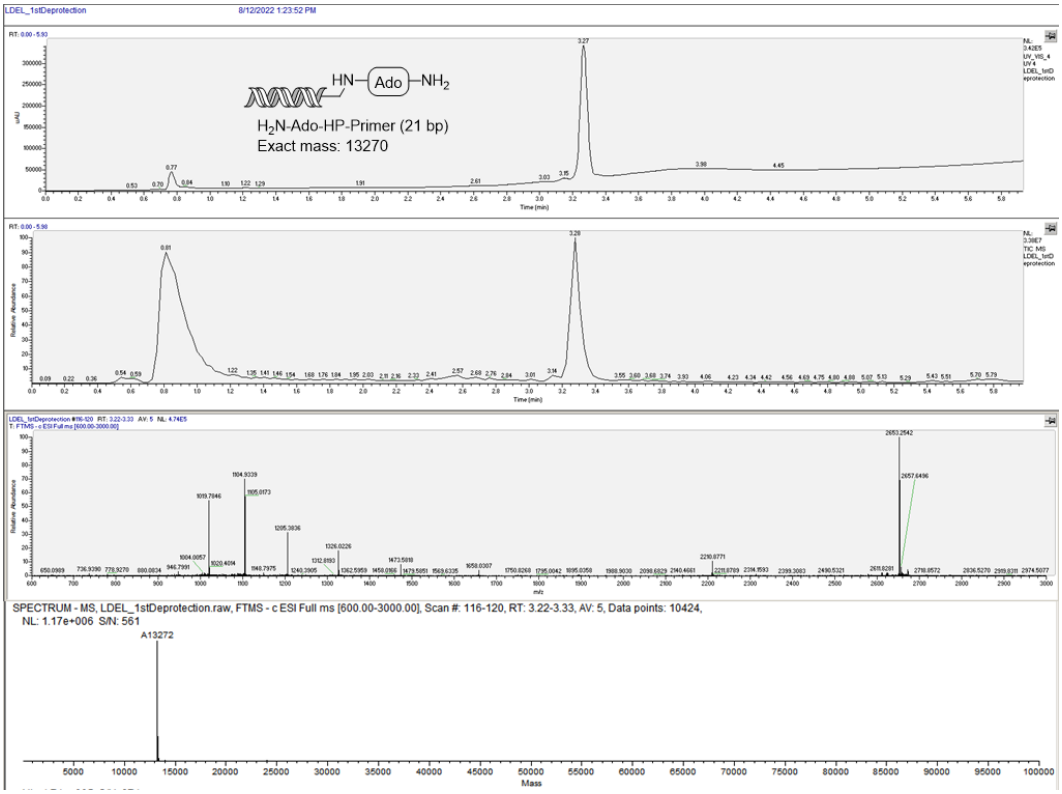
Library Starting Materials

LCMS of library starting material, **Fmoc-Ado-HP-Primer** (21 bp)



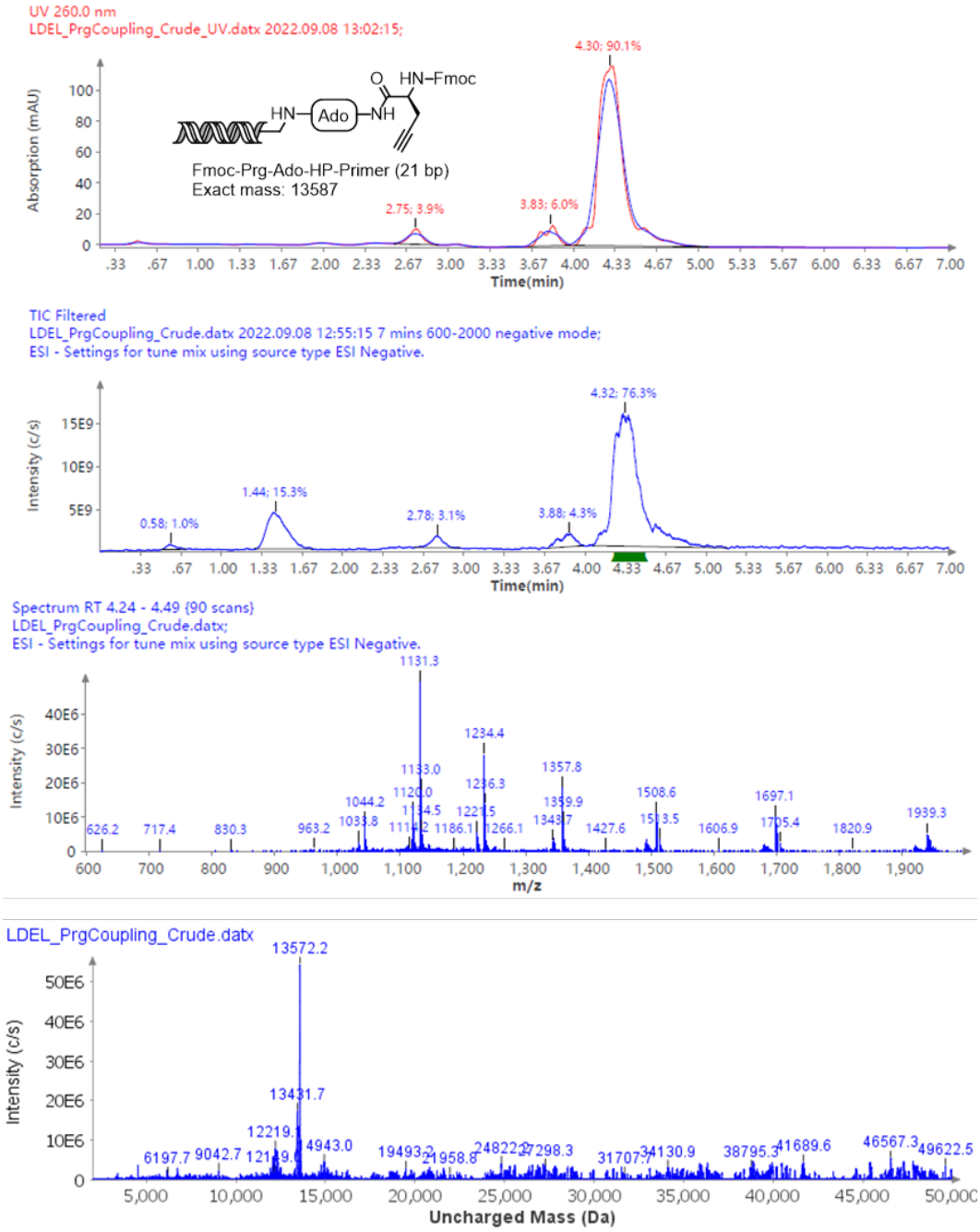
Expected mass: 13492; **Observed mass**: 13492

LCMS of library starting material, **H_2N-Ado-HP-Primer** (21 bp)



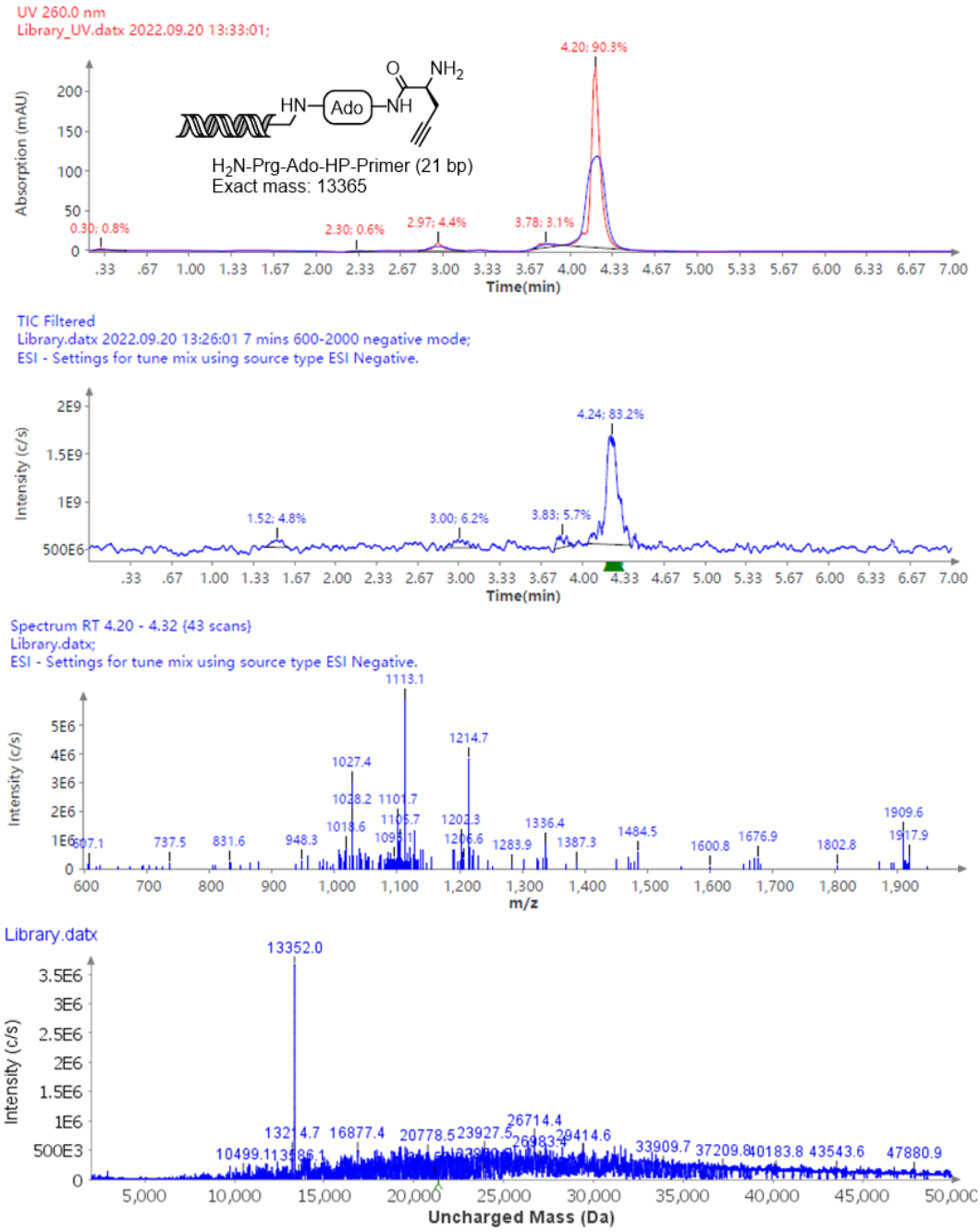
Expected mass: 13270; **Observed mass**: 13272

LCMS of library starting material, **Fmoc-Prg-Ado-HP-Primer** (21 bp)



Expected mass: 13587; **Observed mass**: 13572.2

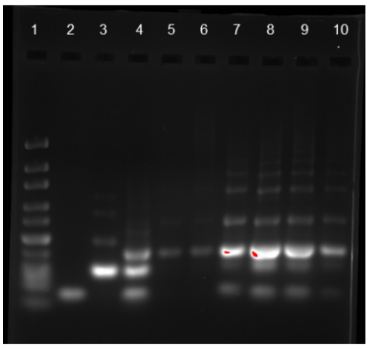
LCMS of library starting material, **H_2N-Prg-Ado-HP-Primer** (21 bp)



Expected mass: 13365; **Observed mass**: 13552

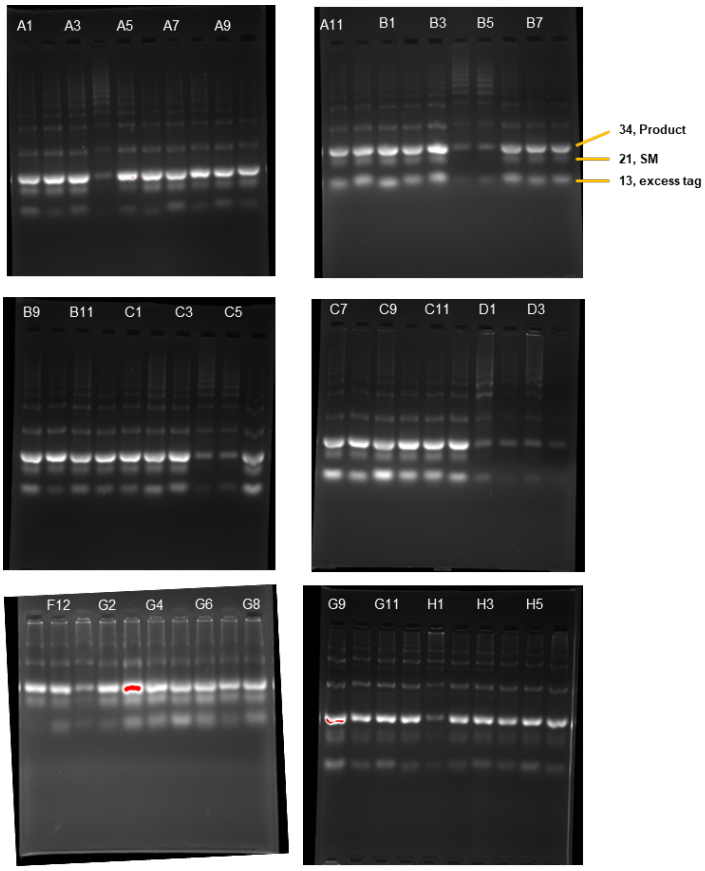
Library Synthesis Cycle 1

Gel electrophoresis of **Cycle 1 ligations** with controls (5% agarose, 1X TAE running buffer)

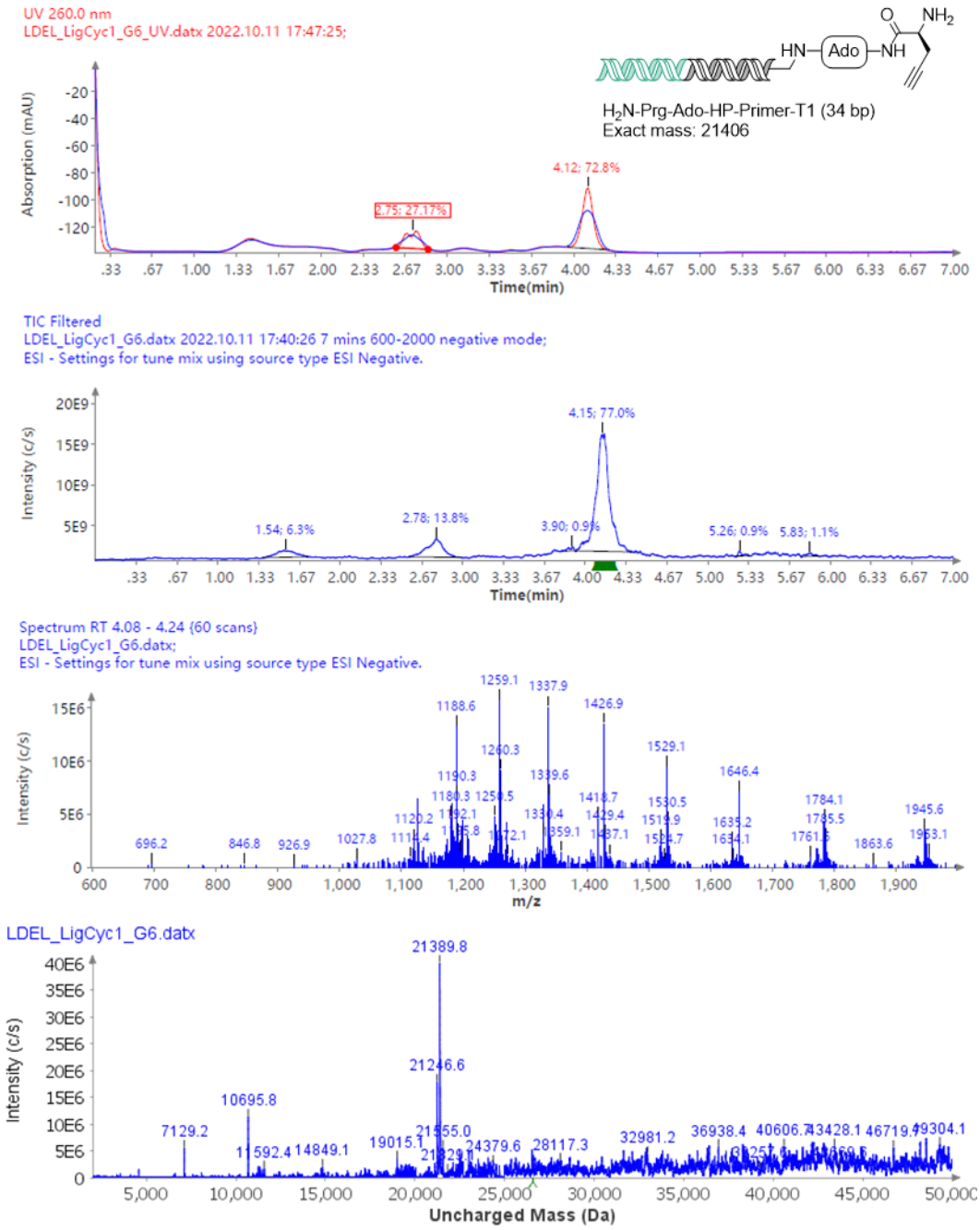


Lane	ID
1	Ladder
2	13 bp (Cycle 1 Tag)
3	21 bp (Prg-HP-Primer)
4	Null 1
5	D4
6	D5
7	D6
8	D7
9	D8
10	D9

Representative gel electrophoresis of **Cycle 1 ligations** (5% agarose, 1X TAE running buffer)

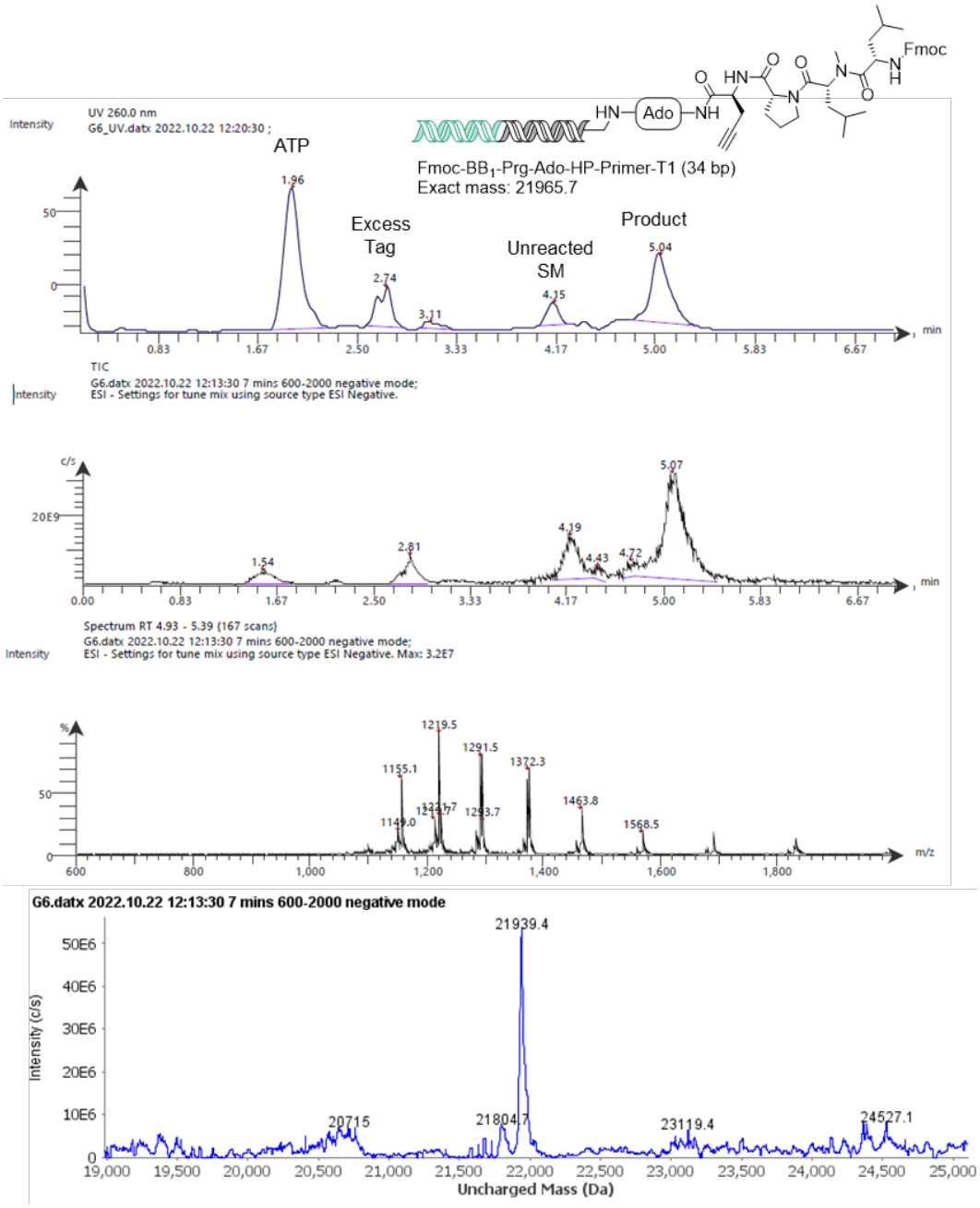


LCMS of representative **Cycle 1 ligation, well G6** (BB_{33} = Fmoc-Leu-D-Leu(NMe)-D-Pro-OH)



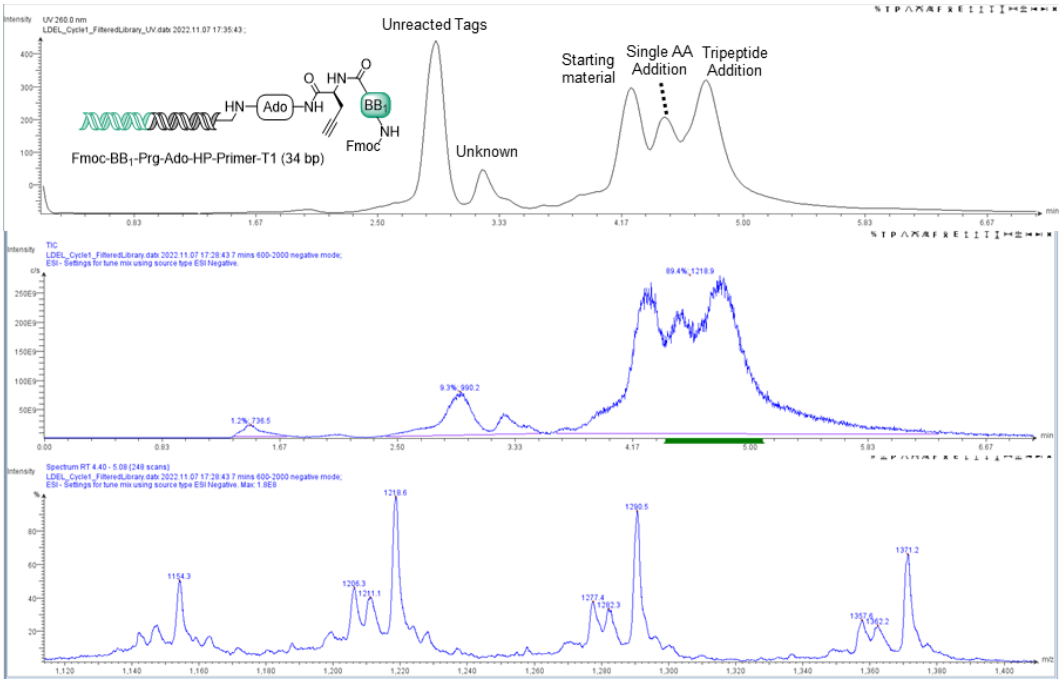
Expected mass: 21406; **Observed mass**: 21389.8

LCMS of representative **Cycle 1 coupling, well G6** (BB_{33} = Fmoc-Leu-D-Leu(NMe)-D-Pro-OH)

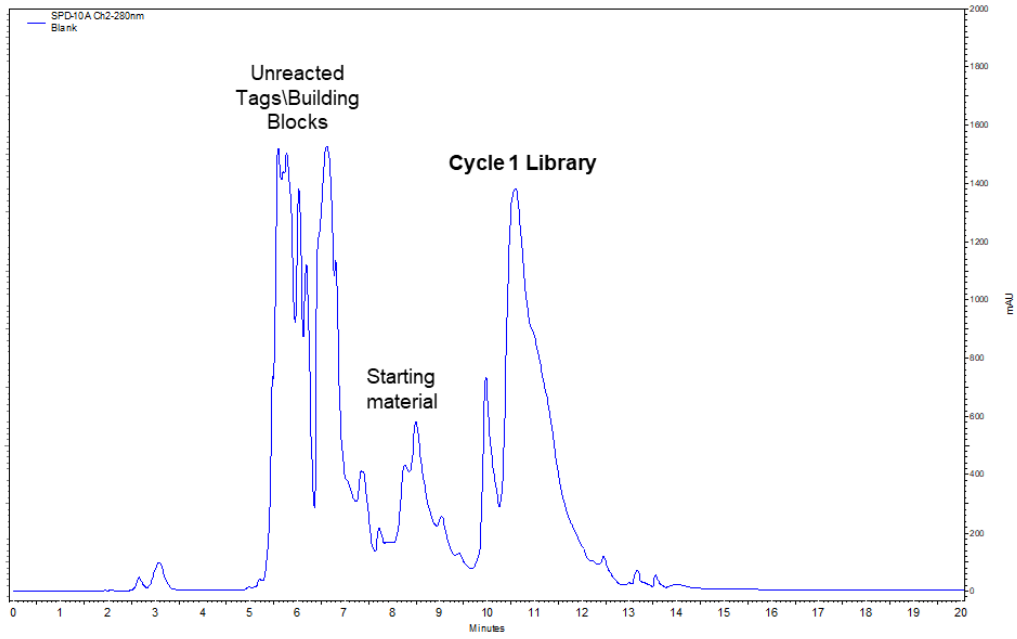


Expected mass: 21965.7; **Observed mass**: 21939.4

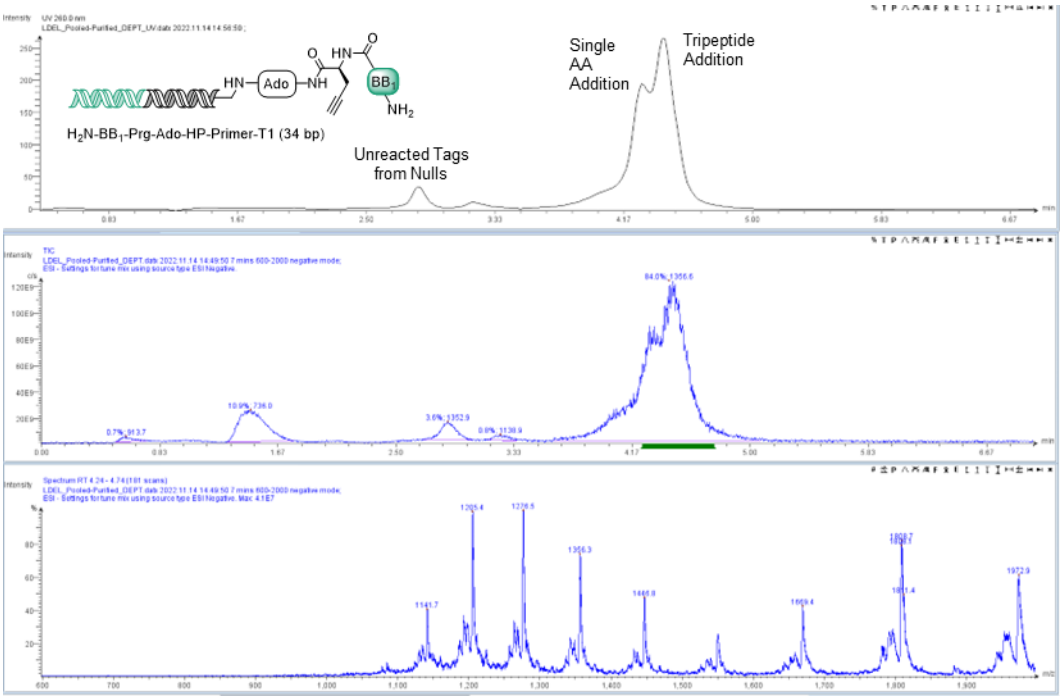
LCMS of **Crude Pooled Cycle 1 Fmoc-Library** (34 bp)



HPLC UV-Vis Trace of **Cycle 1 Purification**

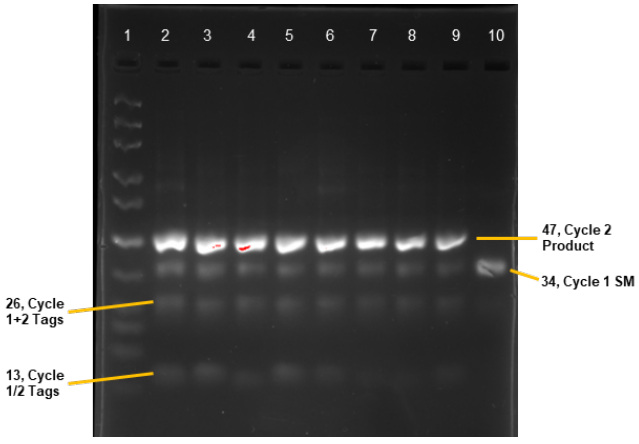


LCMS of **Purified Pooled Cycle 1 H_2N-Library** (34 bp)



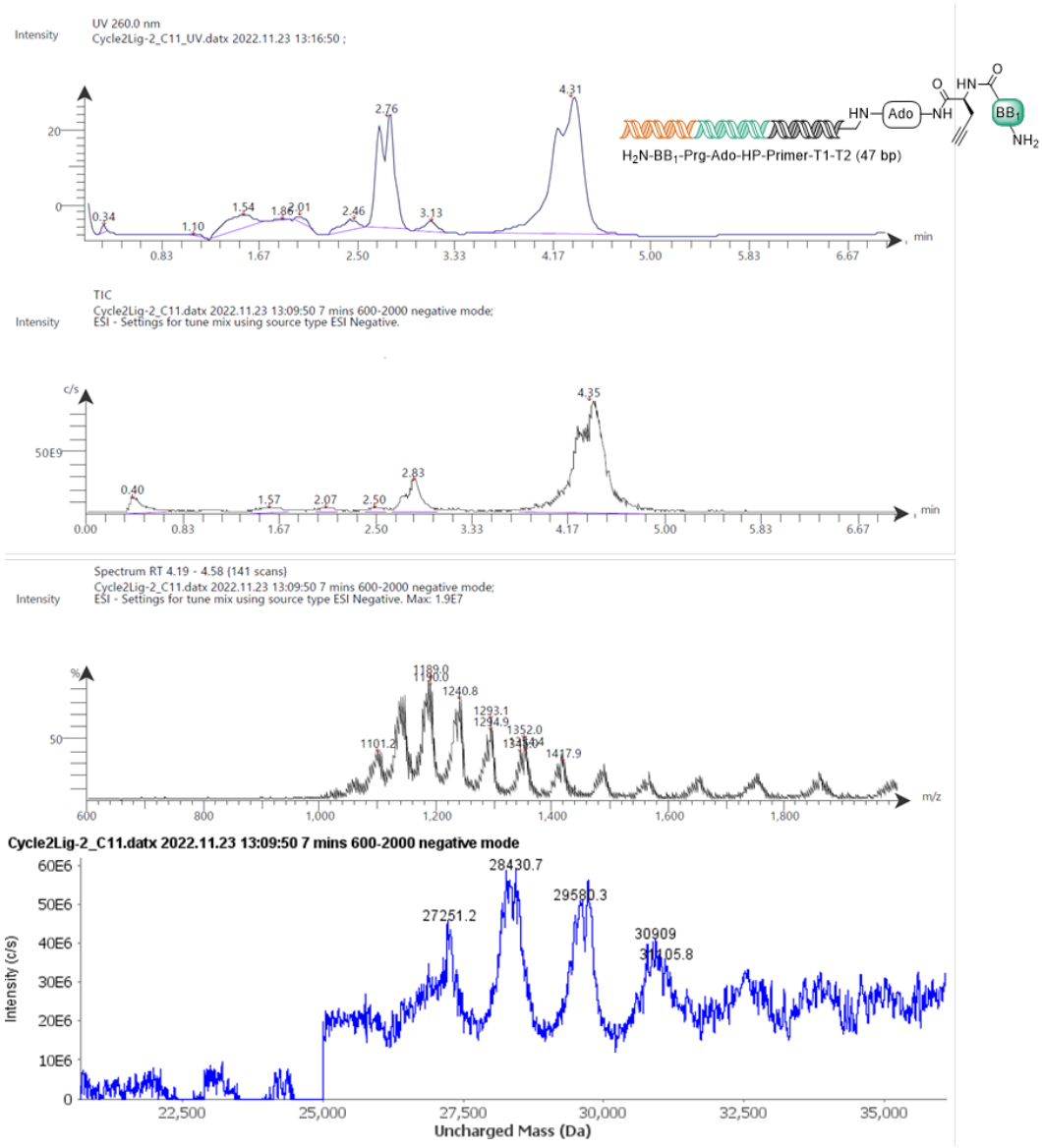
Library Synthesis Cycle 2

Gel electrophoresis of representative **Cycle 2 ligations** (5% agarose, 1X TAE running buffer)



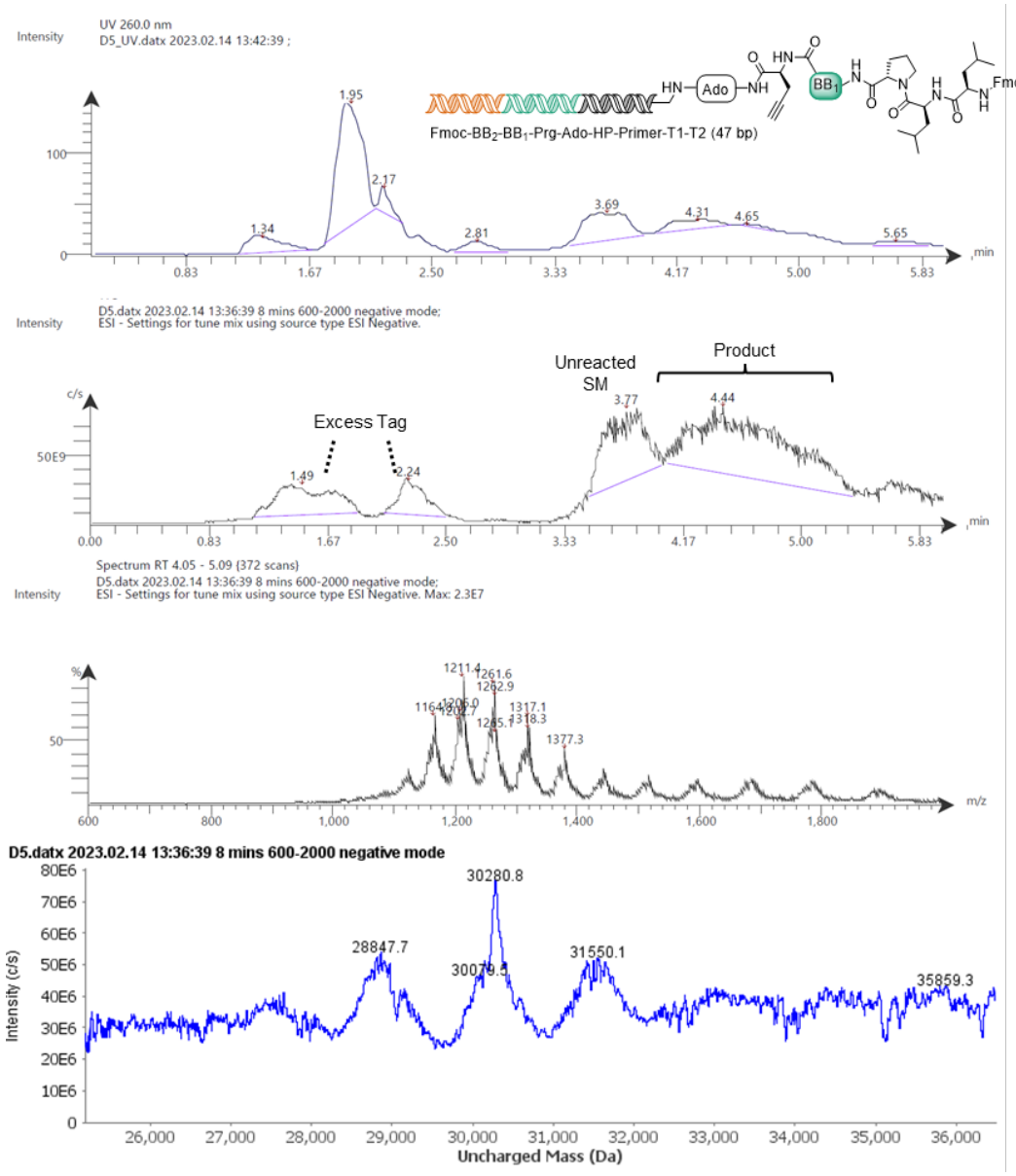
Lane	ID
1	Ladder
2	A1
3	B2
4	C3
5	D4
6	E5
7	F6
8	G7
9	H8
10	34 bp (from previous cycle)

LCMS of representative **Cycle 2 ligation, well C11** (BB_{12} = Fmoc-LA18-OH)



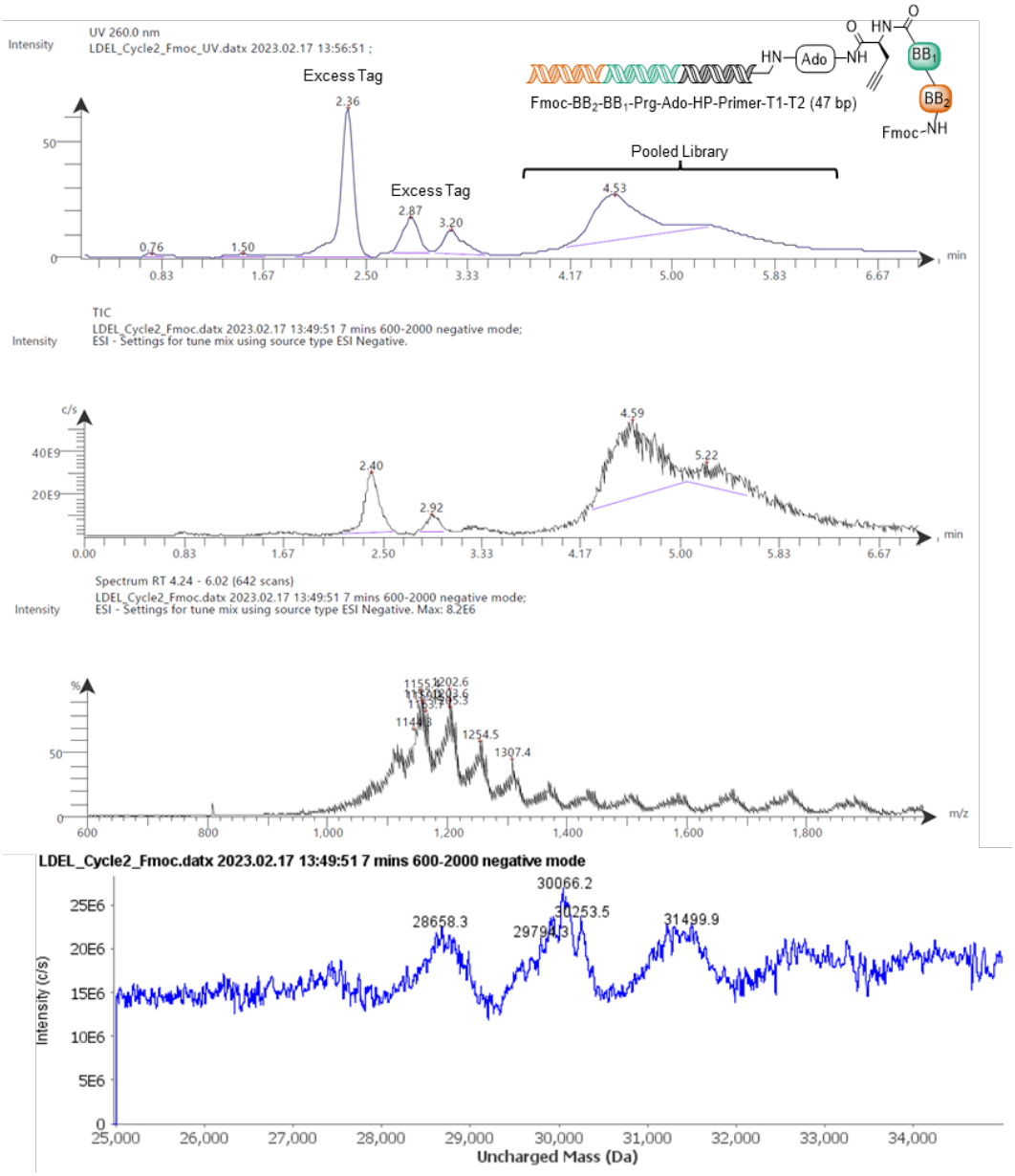
Expected mass: 29644 (BB average); **Observed mass**: 29580

LCMS of representative **Cycle 2 coupling, well D5** (BB_{20} = Fmoc-D-Leu-Leu-Pro-OH)



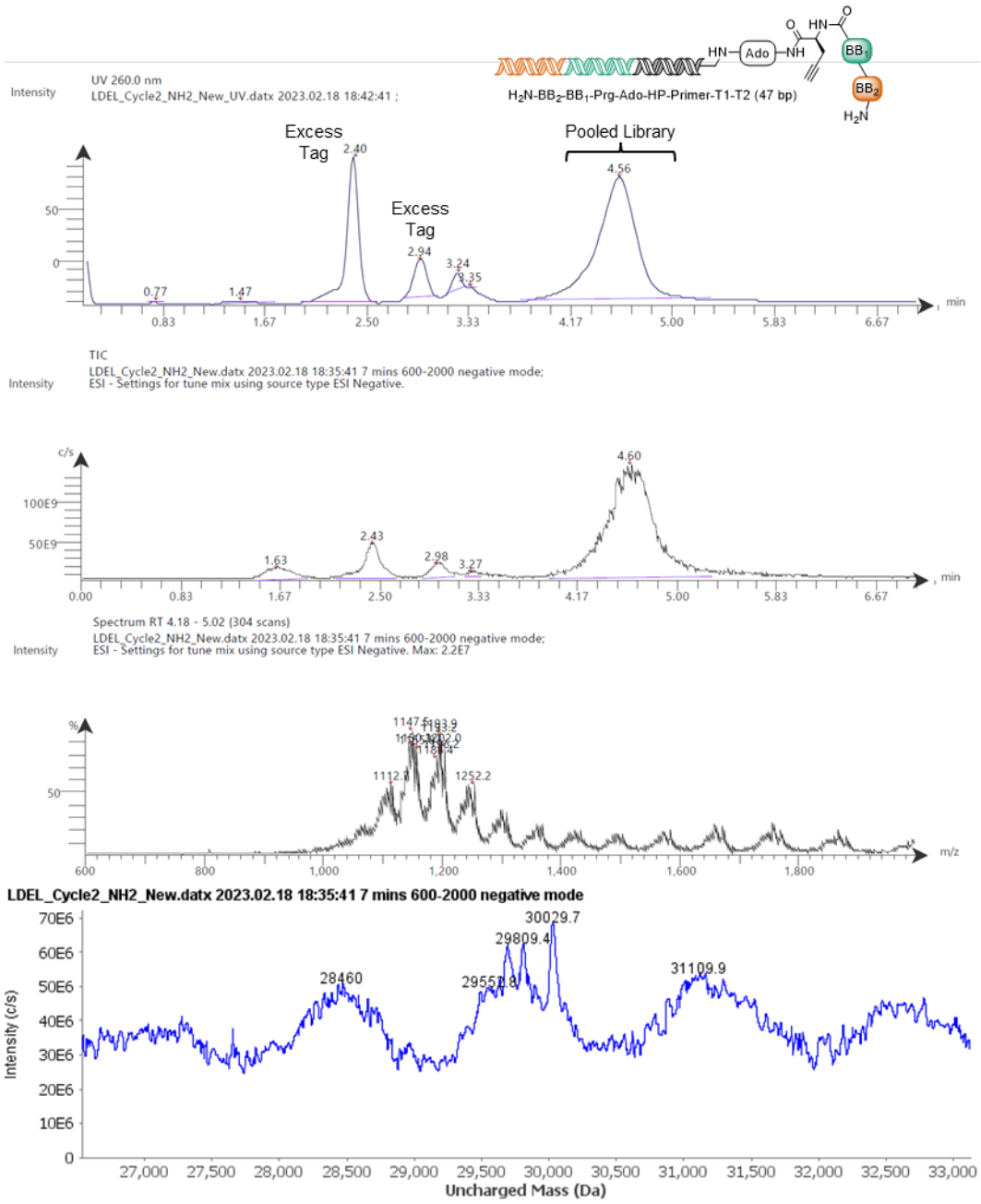
Expected mass: 30104 (BB average); **Observed mass**: 30280.8

LCMS of **Crude Cycle 2 Pooled Fmoc-Library** (47 bp)



Expected mass: 30103.9 (BB average); **Observed mass**: 30066.2

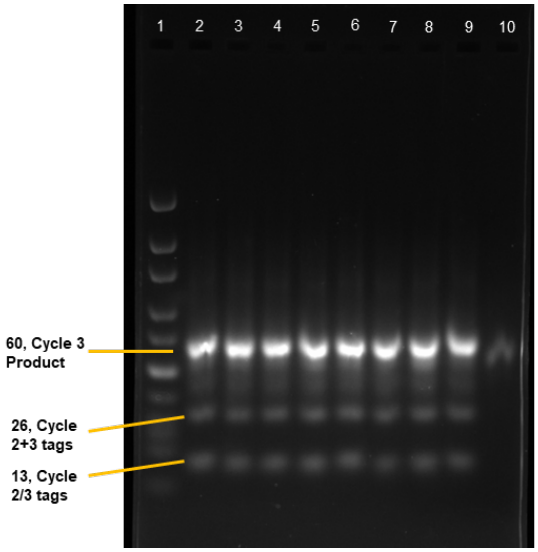
LCMS of **Cycle 2 Pooled H_2N-Library** (47 bp)



Expected mass: 29881.9 (BB average); **Observed mass**: 29809.4

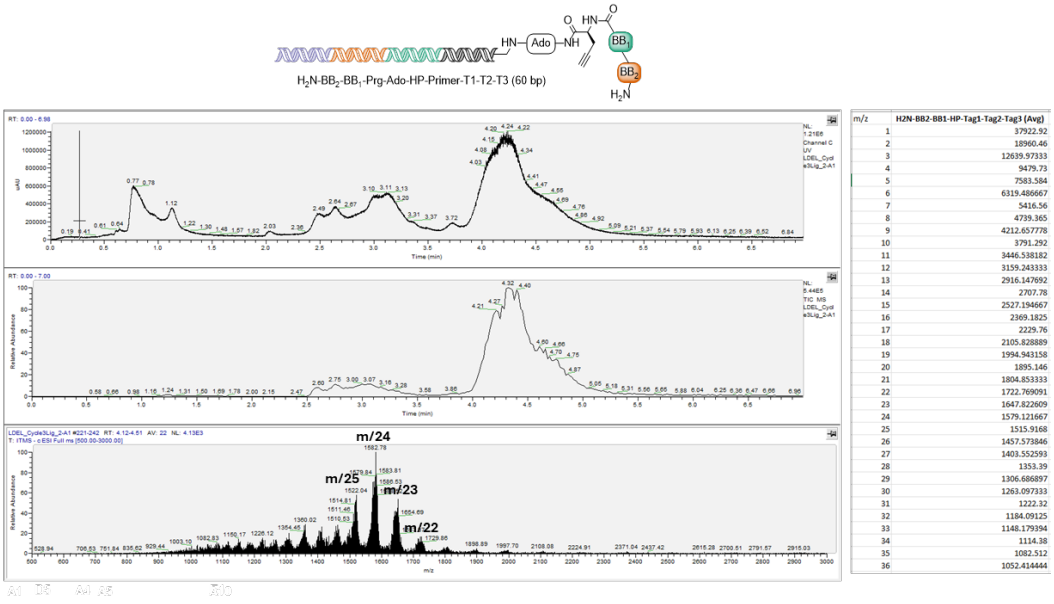
Library Synthesis Cycle 3

Gel electrophoresis of representative **Cycle 3 ligations** (5% agarose, 1X TAE running buffer)

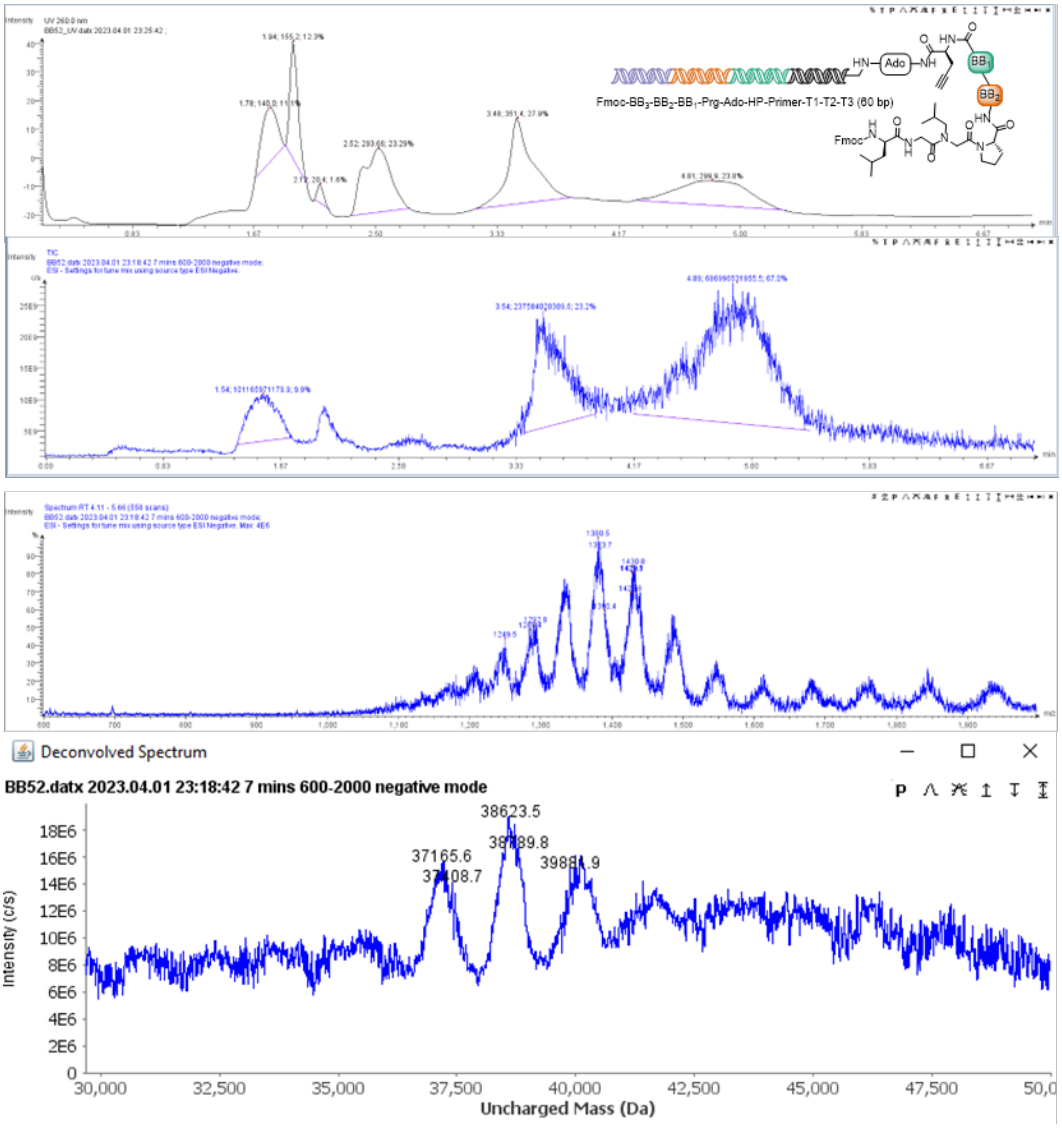


Lane	ID
1	Ladder
2	1-H2
3	1-G3
4	1-F4
5	1-E5
6	2-H2
7	2-G3
8	2-F4
9	2-E5
10	47 bp (from previous cycle)

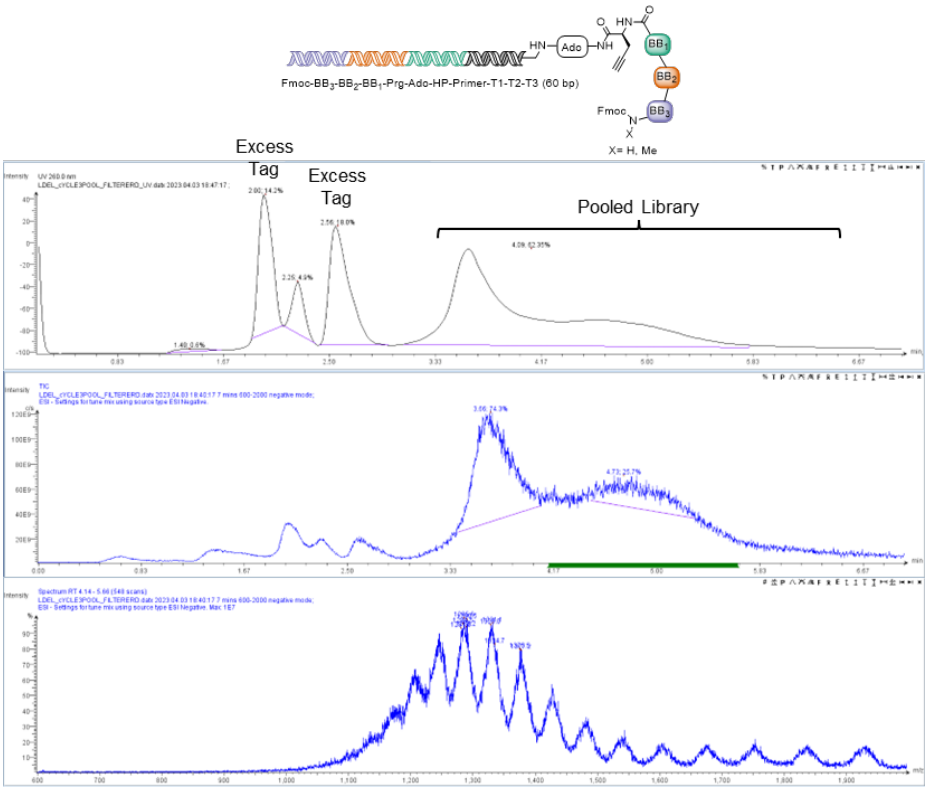
LCMS of representative **Cycle 3 ligation, well 2-A1** (BB_{30} = Fmoc-D-Leu-D-Leu(NMe)-Pro-OH)



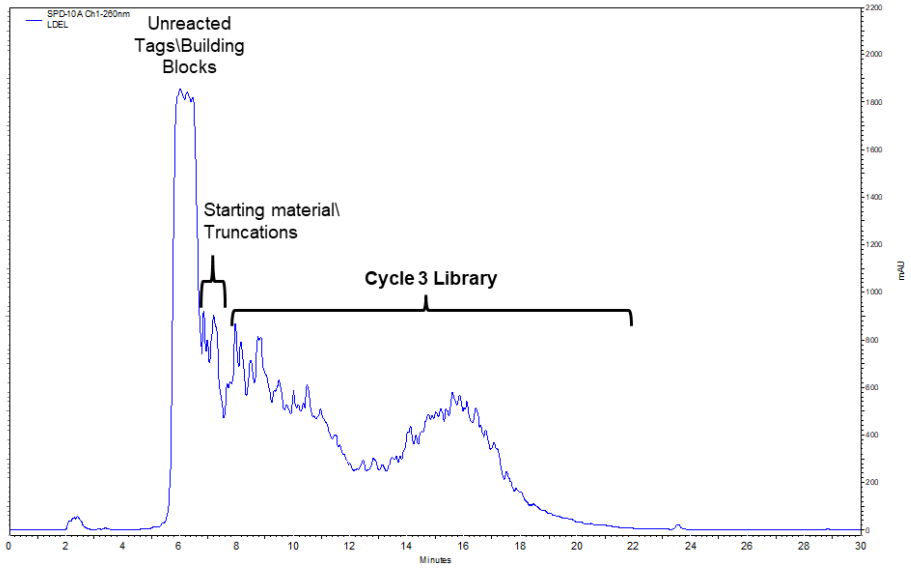
LCMS of representative **Cycle 3 coupling, well 1-E4** (BB_{52} = Fmoc-D-Leu-LA03-Pro-OH)



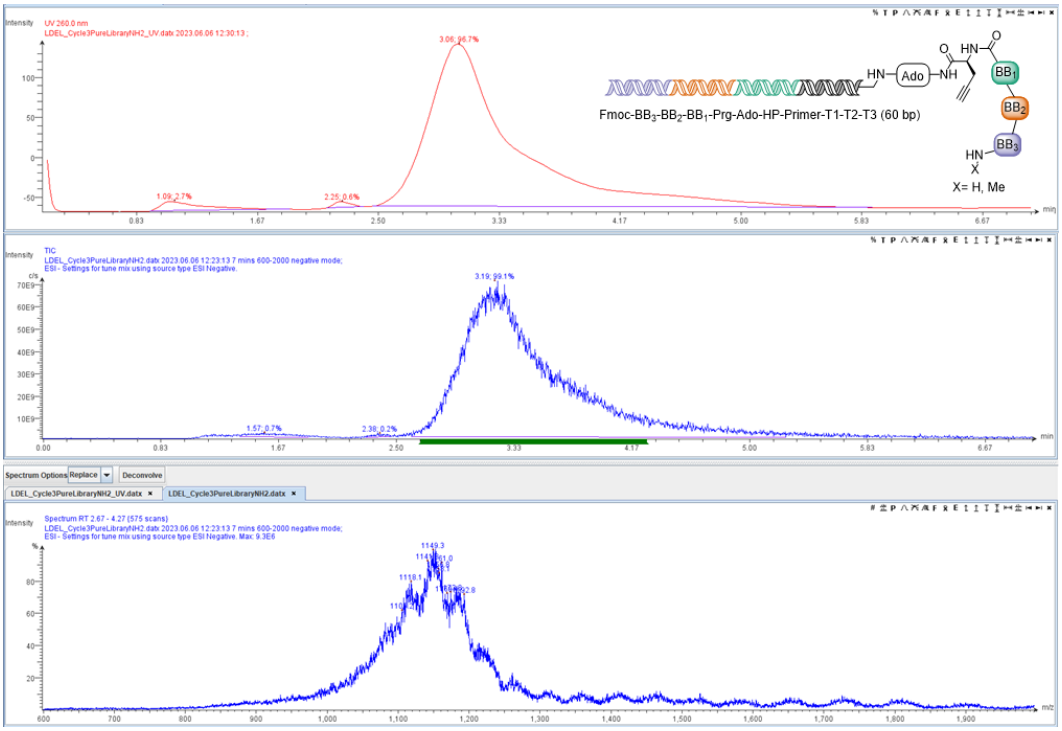
LCMS of **Crude Pooled Cycle 3 Fmoc-Library** (60 bp)



HPLC UV-Vis Trace of **Cycle 3 Purification**

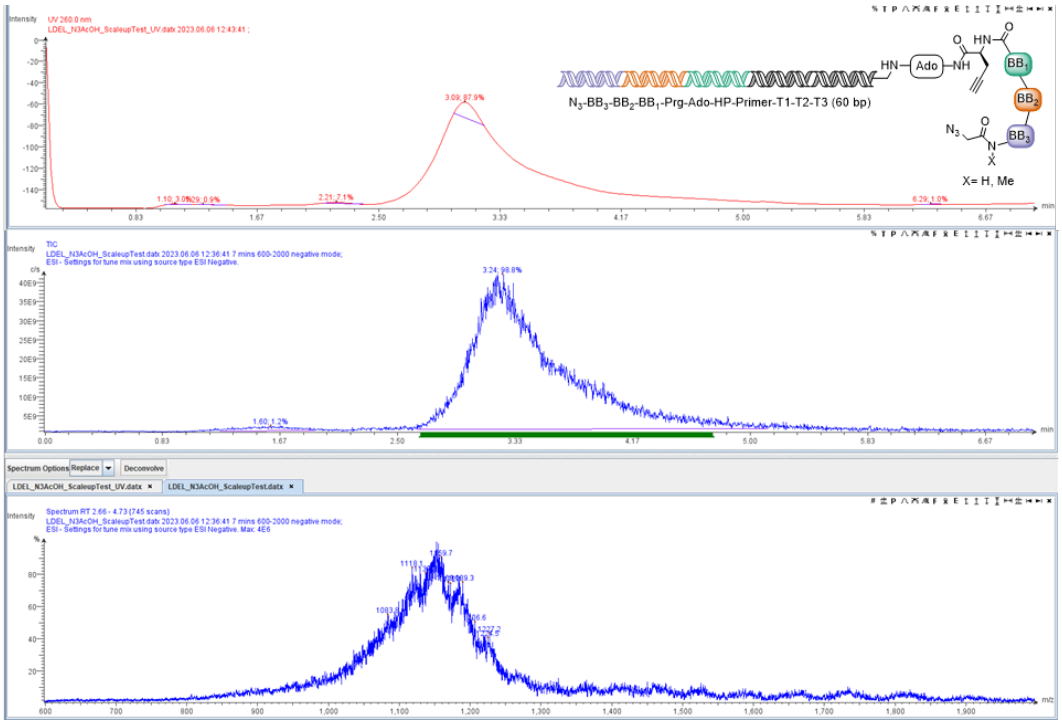


LCMS of **Pure Pooled Cycle 3 H_2N-Library** (60 bp)

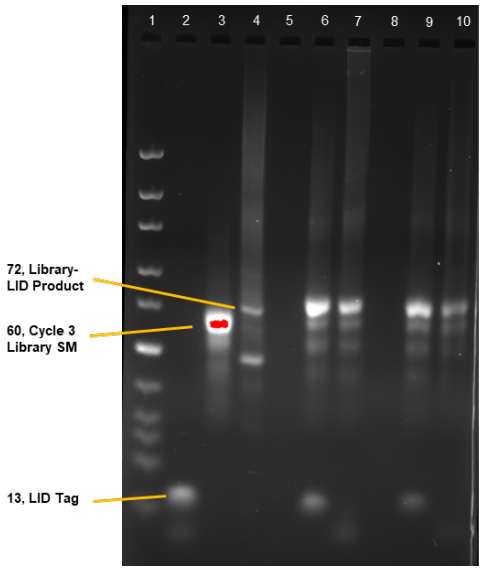


Final Library Reactions

LCMS of **N_3-Library** (60 bp)

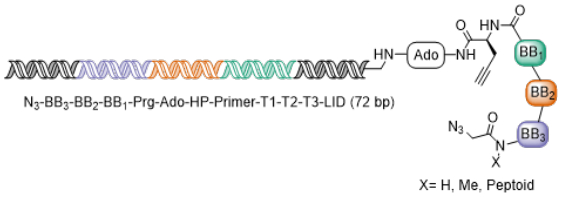


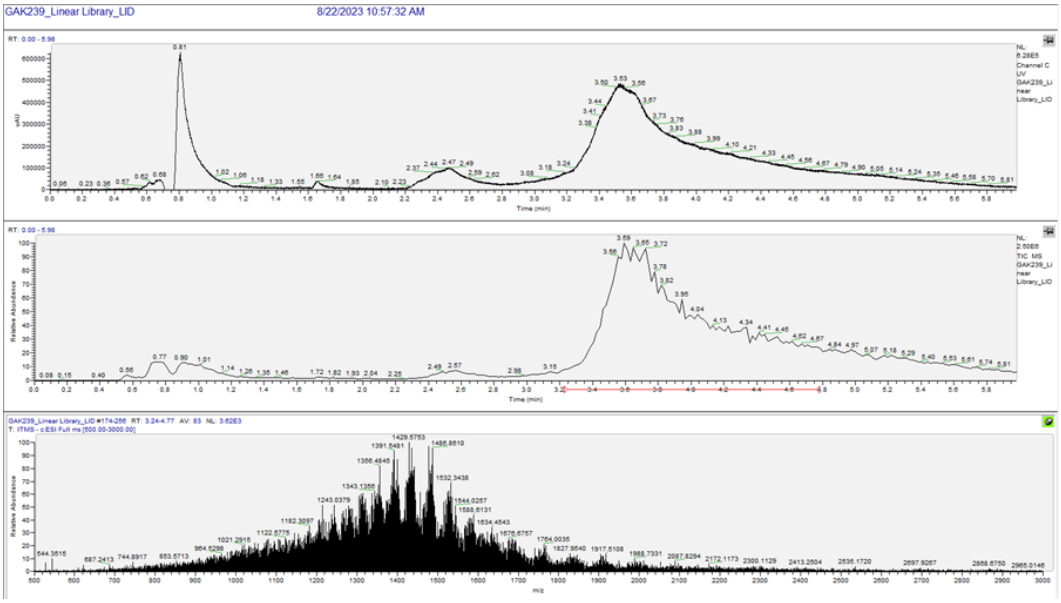
Gel electrophoresis of **Library ID tag ligations**



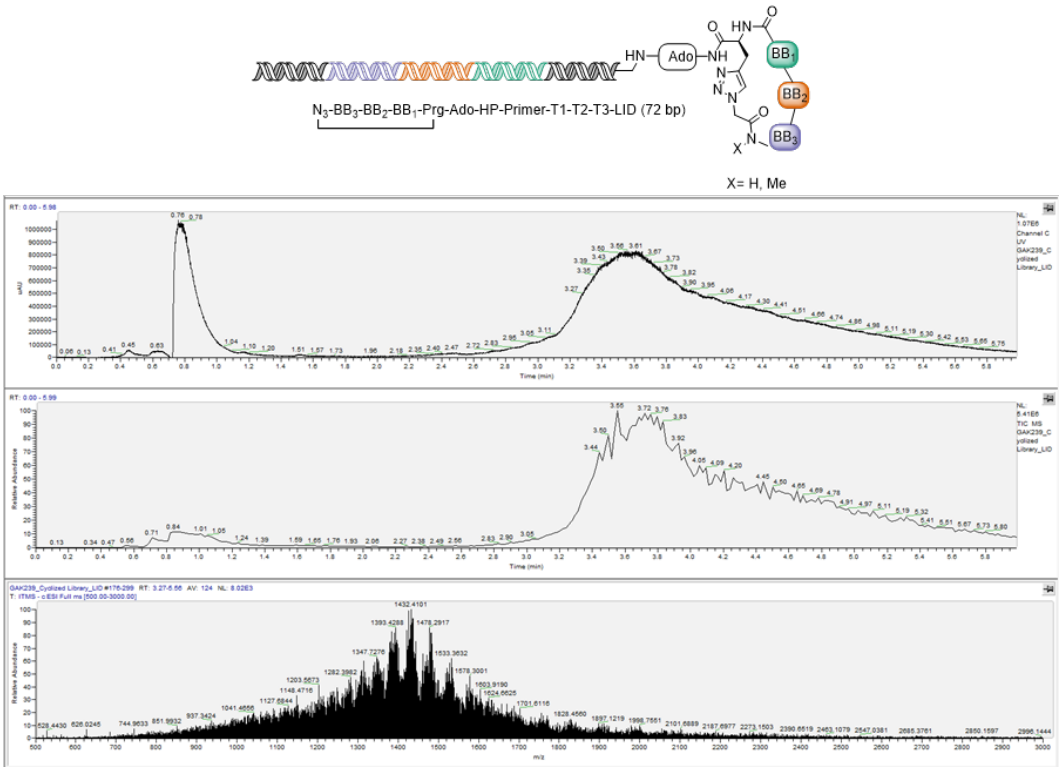
Lane	ID
1	Ladder
2	Library ID Tag (13 bp)
3	N_3-Library (60 bp)
4	Null Library (72 bp)
5	Empty
6	N_3-Library-LID (72 bp)
7	N_3-Library-LID (72 bp)
8	Empty
9	Cyclized Library-LID (72 bp)
10	Cyclized Library-LID (72 bp)

LCMS of **N_3-Library-LID** (72 bp)

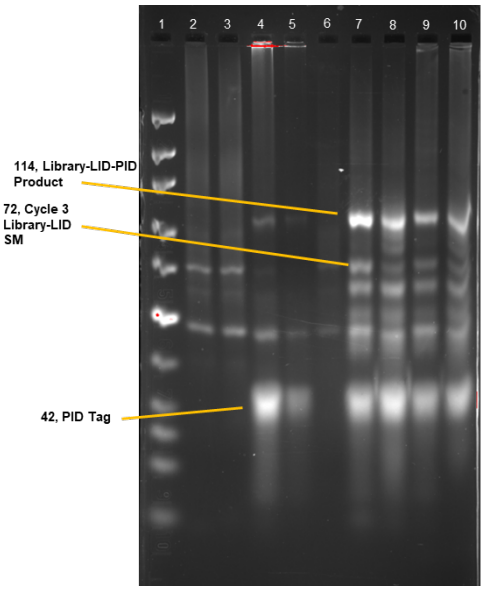




LCMS of **Cyclized-Library-LID** (72 bp)

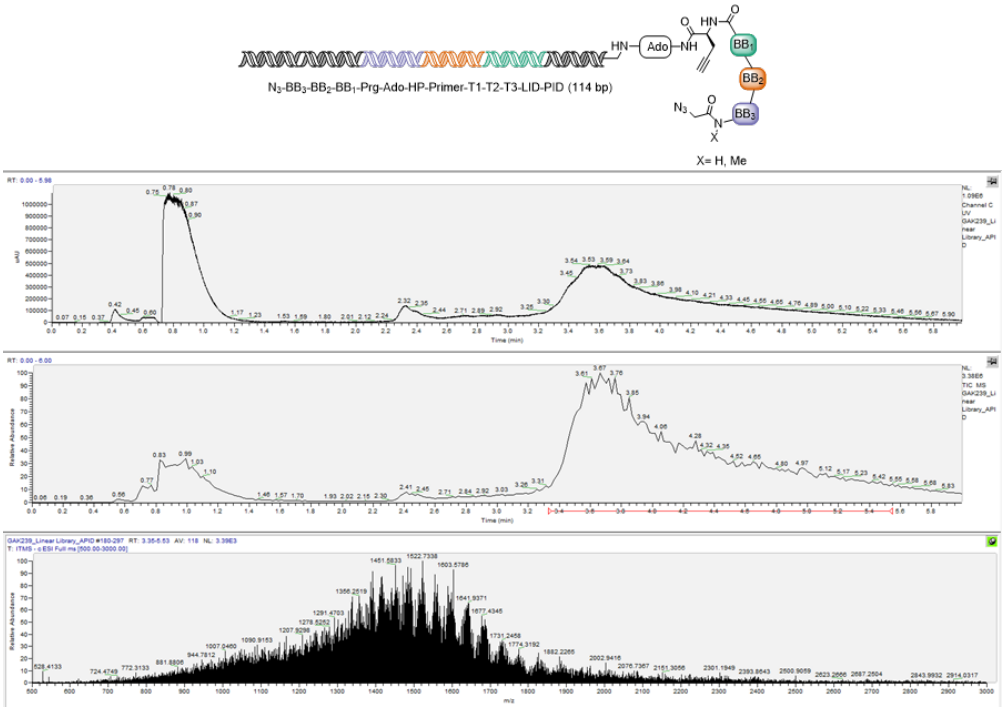


Gel electrophoresis of **Pool ID tag ligations**

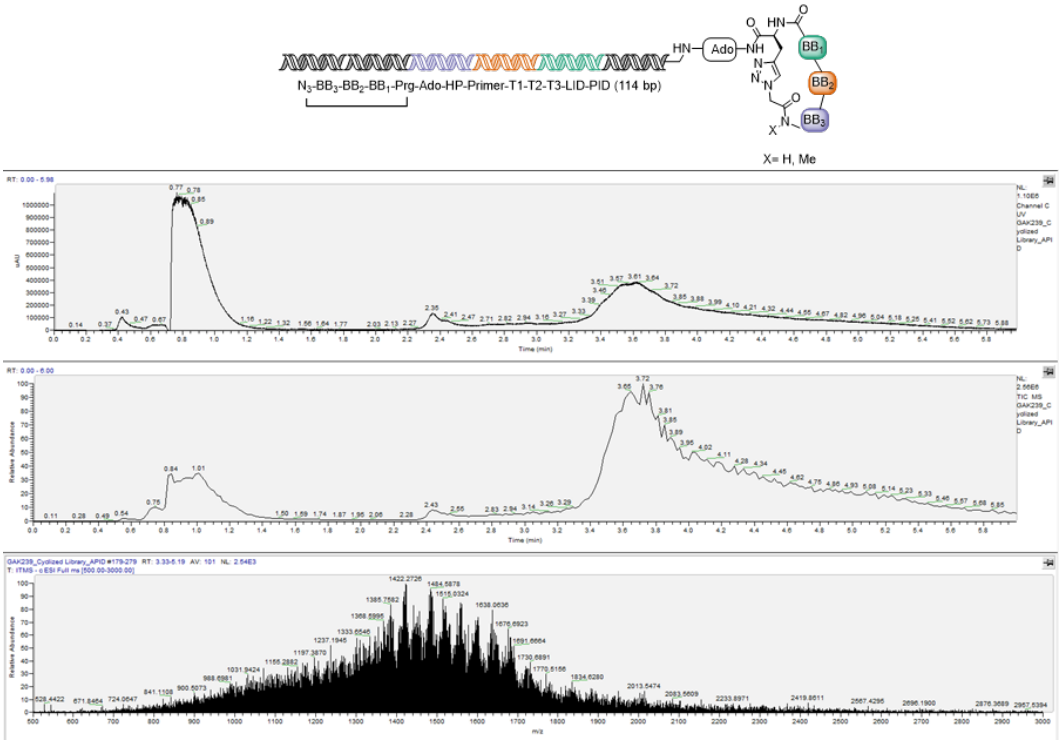


Lane	ID
1	Ladder
2	Null Library (72 bp)
3	Null Library (72 bp)
4	Null Library-PID (114 bp)
5	Null Library-PID (114 bp)
6	Null Library (72 bp)
7	N_3-Library-LID (114 bp)
8	N_3-Library-LID (114 bp)
9	Cyclized Library-LID (114 bp)
10	Cyclized Library-LID (114 bp)

LCMS of **Linear-Library-LID-PID** (114 bp)



LCMS of **Cyclized-Library-LID-PID** (114 bp)

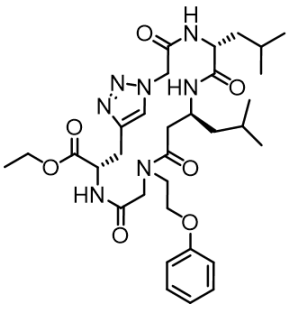


Appendix F: LCMS Traces, HNMR, and Classic Chrome Sequencing Count Chromatograms for LDEL001-044

Compounds were identified by HRMS (**Table 4**, Method 1). Compound purity was assessed by ELSD (**Table 4**, Method 2). ^{1}HNMR for each compound was collected on a Bruker Ascend spectrometer at 800 MHz.

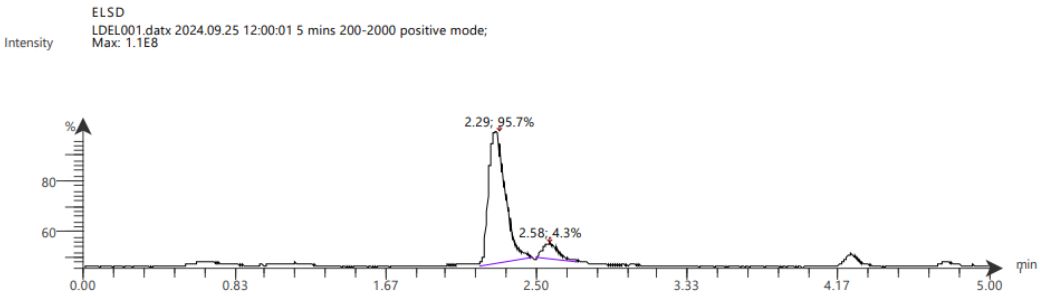
A "Classic Chrome Sequence Count Chromatogram" is also provided for each compound. See supplementary discussion above for additional details.

LDEL001

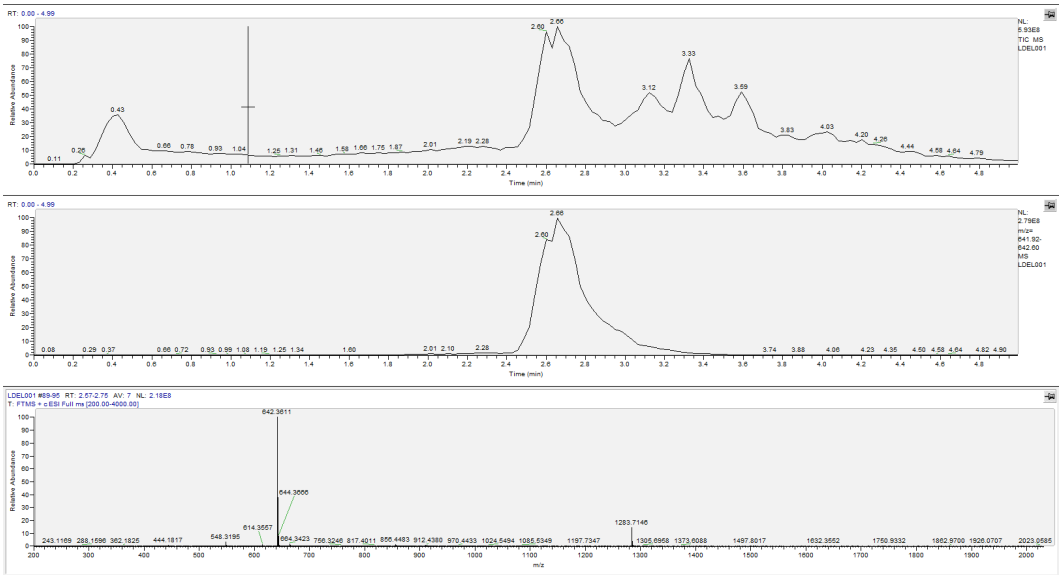


LDEL001
DLeu-ßHomoleu-LA18
Exact Mass: 641.3537

ELSD of **LDEL001**

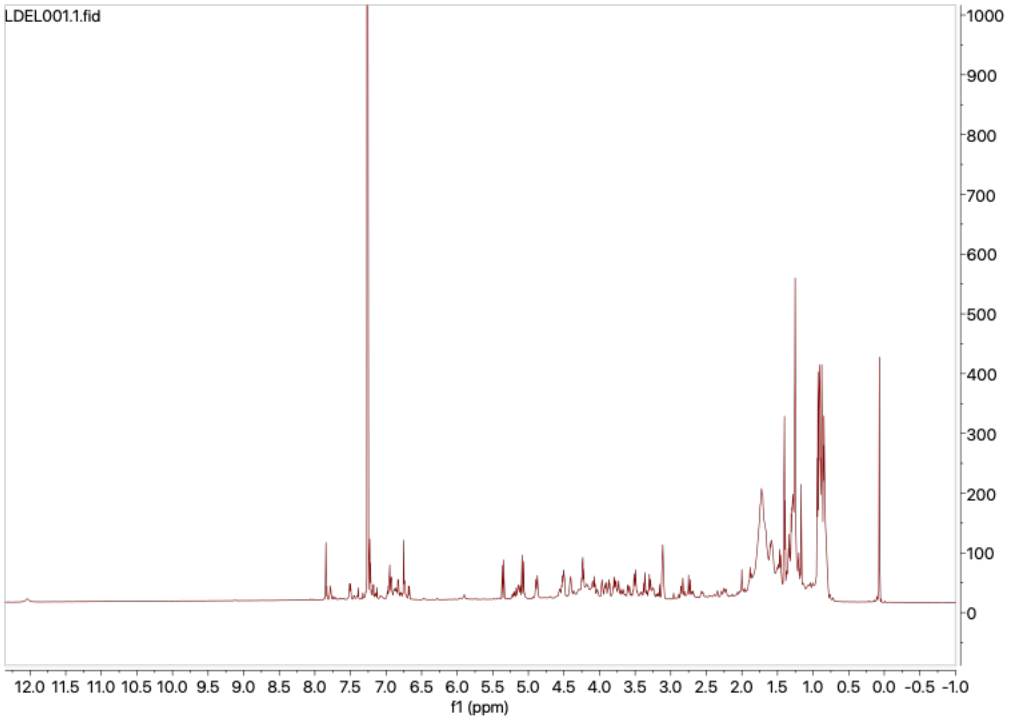


HRMS of **LDEL001**

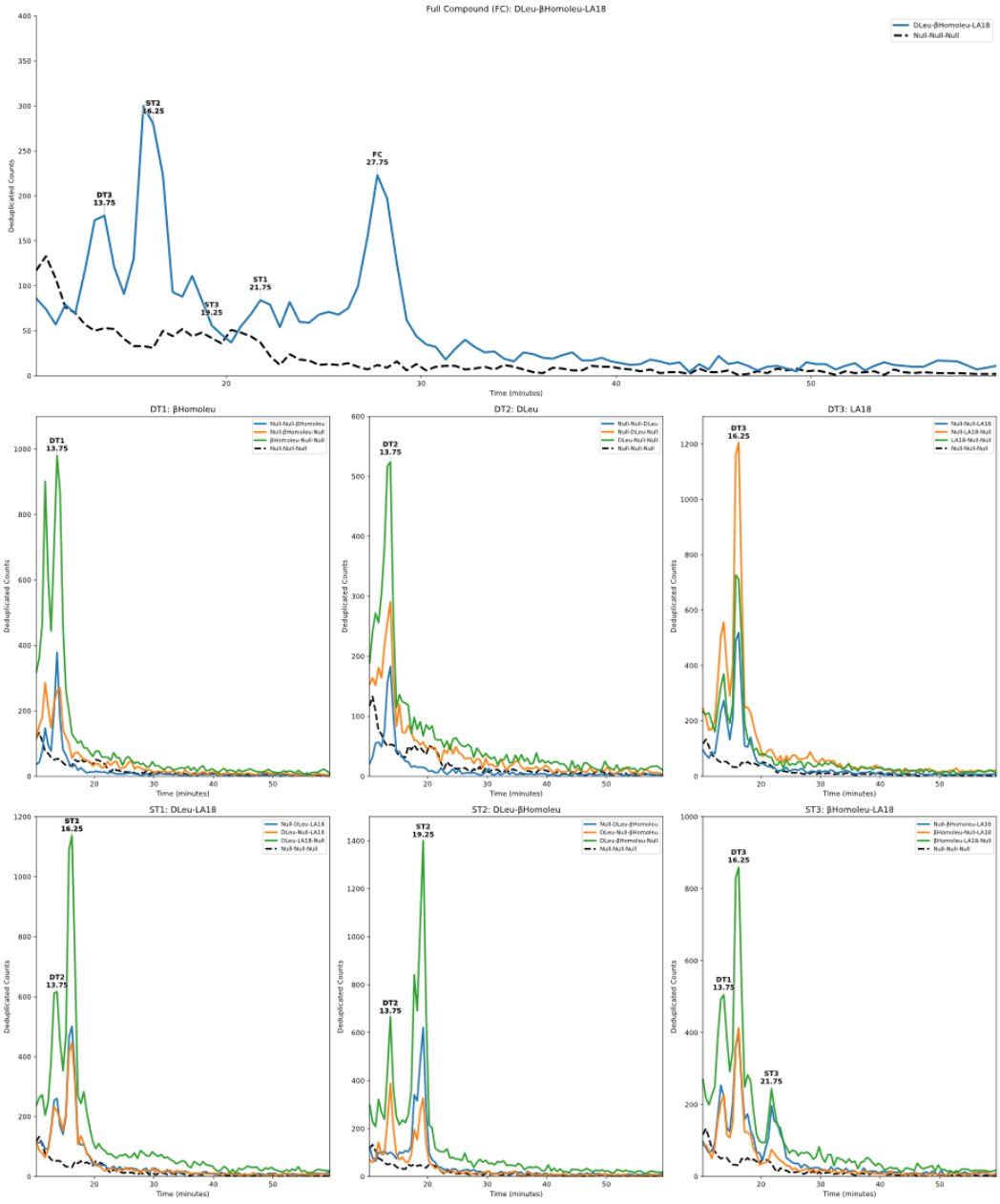


HRMS (ESI) m/z calculated for $C_{32}H_{47}N_7O_7^+$ ($M+H^+$) 642.3571; found 642.3611

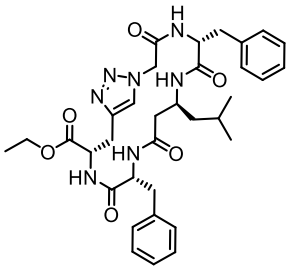
^{1}HNMR of **LDEL001**



Classic Chrome Sequence Count Chromatogram of **LDEL001**

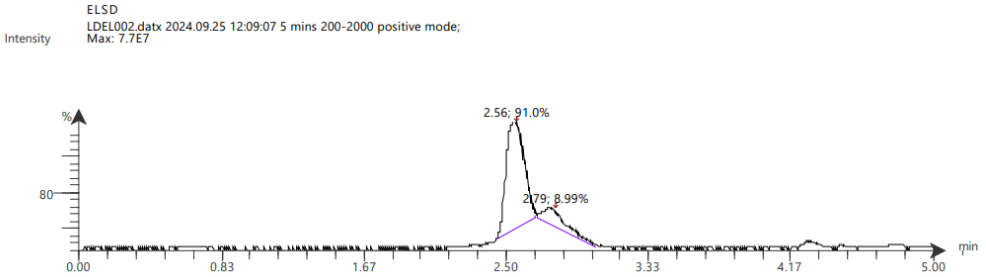


LDEL002

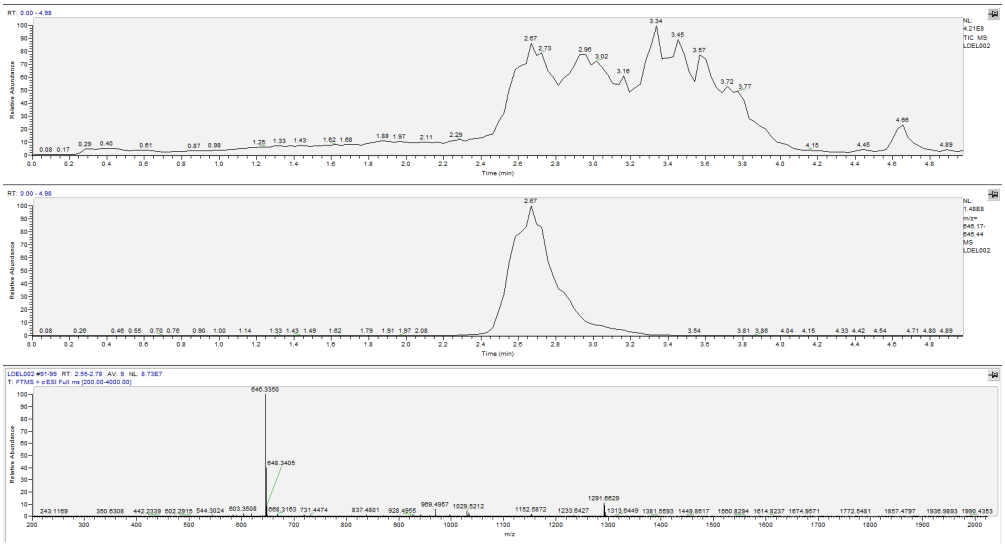


LDEL002
DPhe-ßHomoleu-DPhe
Exact Mass: 645.3275

ELSD of **LDEL002**

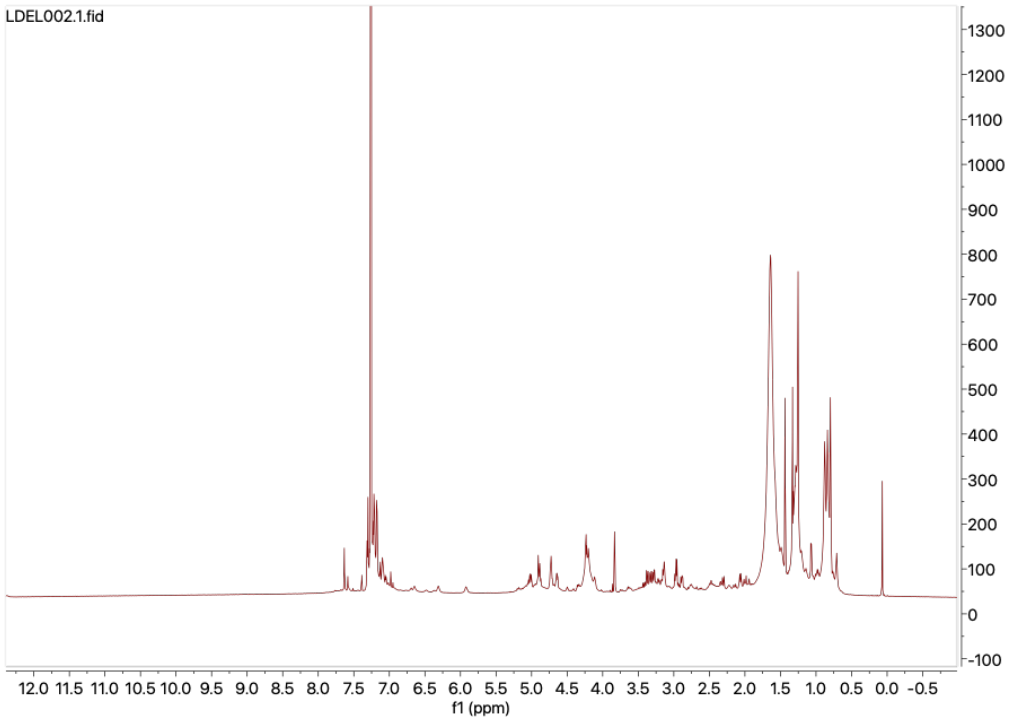


HRMS of **LDEL002**

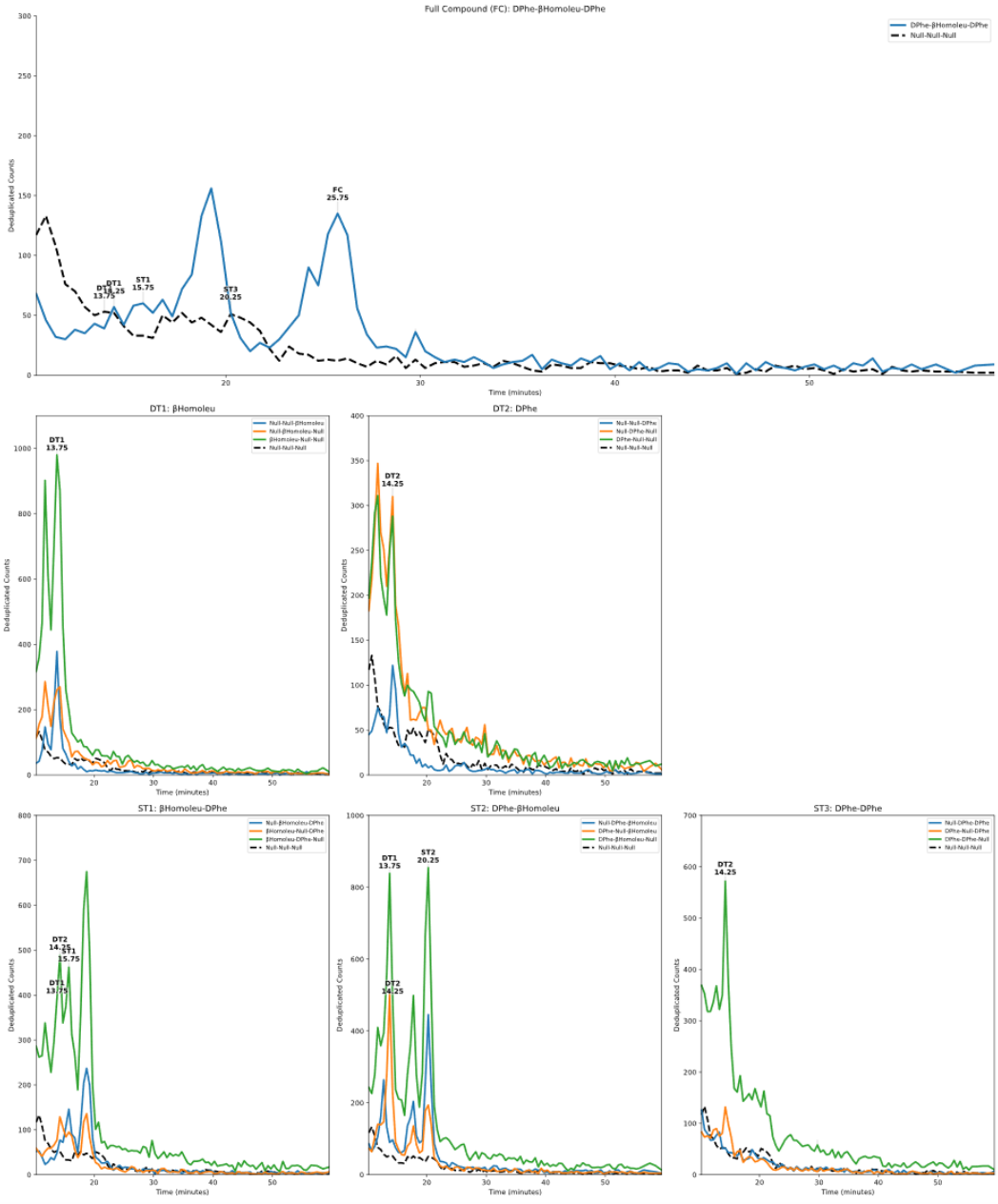


HRMS (ESI) m/z calculated for $C_{34}H_{43}N_7O_6^+$ ($M+H^+$) 646.3308; found 646.3350

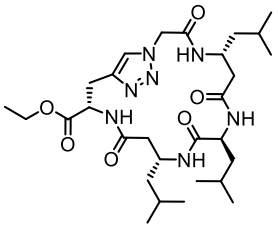
^{1}HNMR of **LDEL002**



Classic Chrome Sequence Count Chromatogram of **LDEL002**

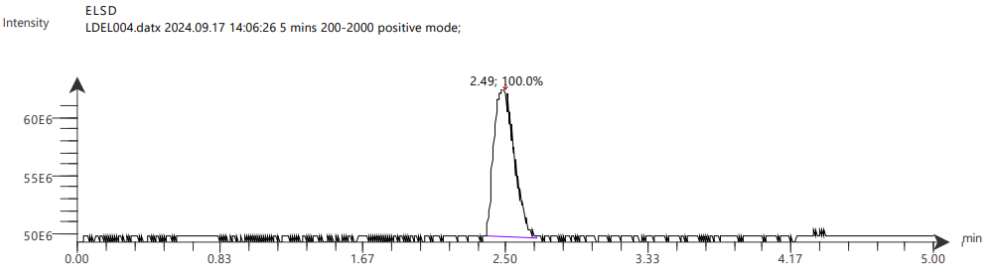


LDEL004

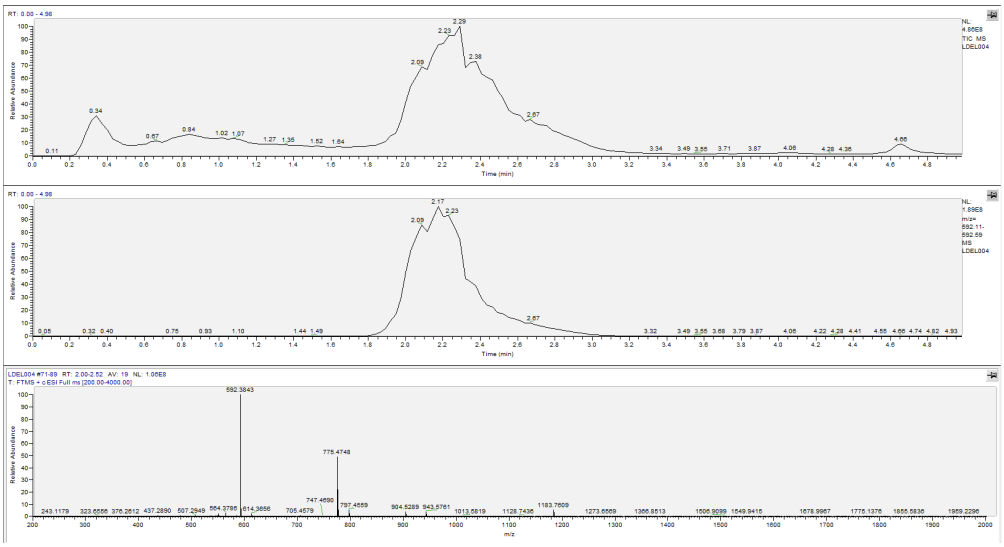


LDEL004
DßHomoleu-Leu-DßHomoleu
Exact Mass: 591.3744

ELSD of **LDEL004**

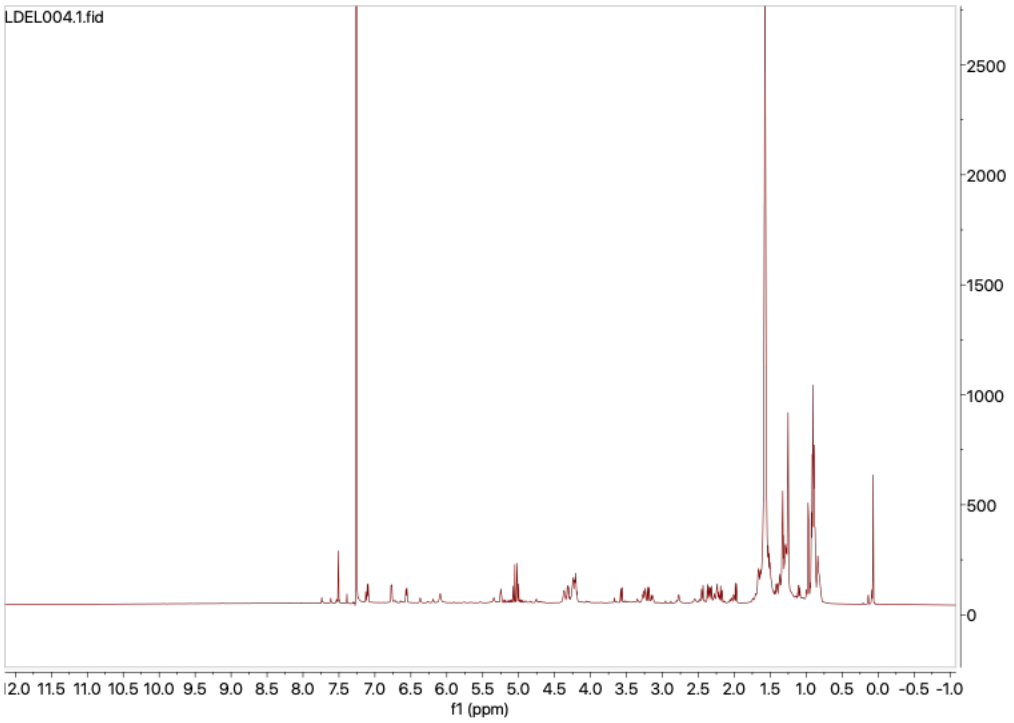


HRMS of **LDEL004**

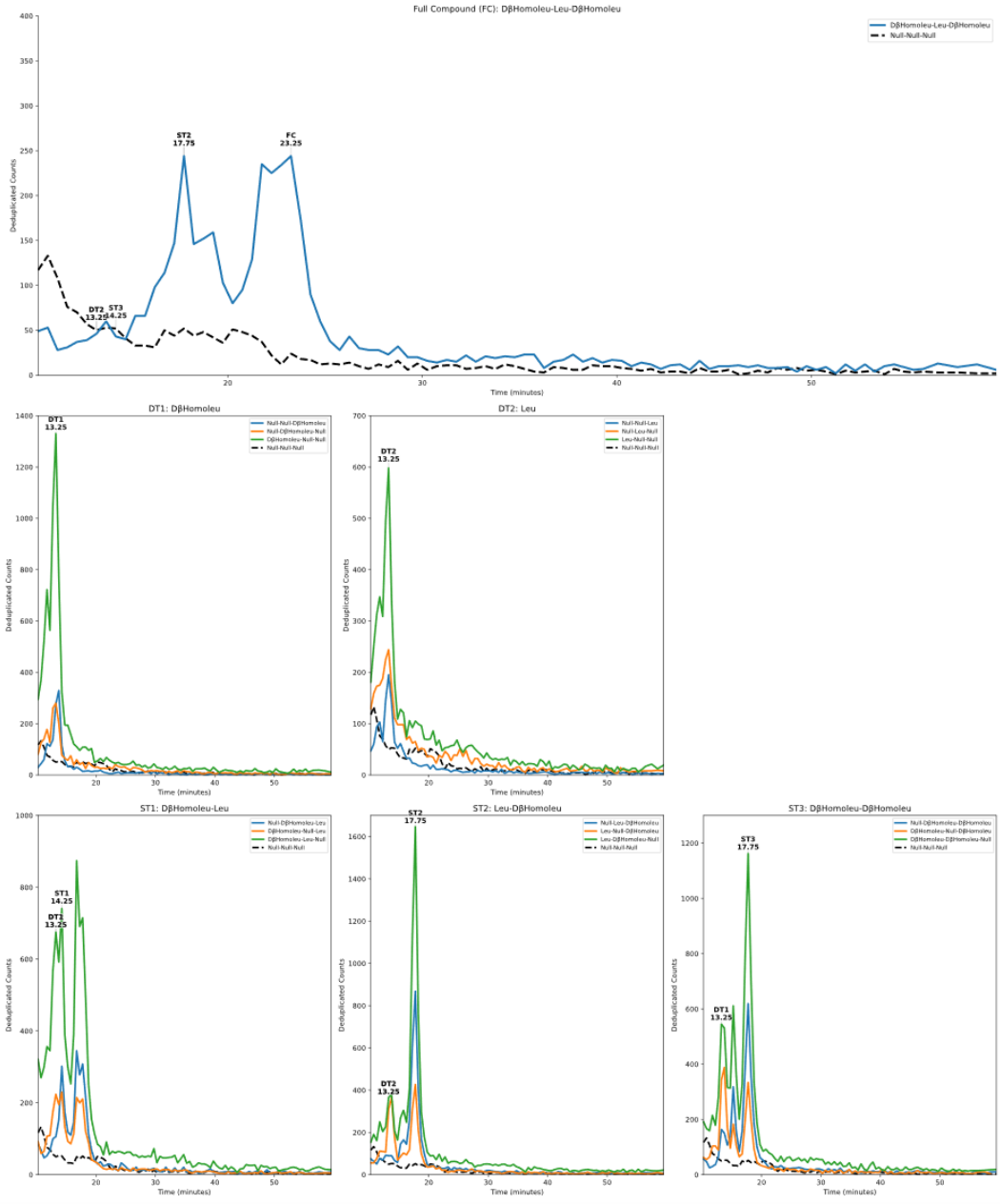


HRMS (ESI) m/z calculated for $C_{29}H_{49}N_7O_6^+$ ($M+H^+$) 592.3778; found 592.3843

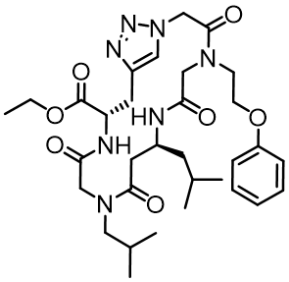
^{1}HNMR of **LDEL004**



Classic Chrome Sequence Count Chromatogram of **LDEL004**

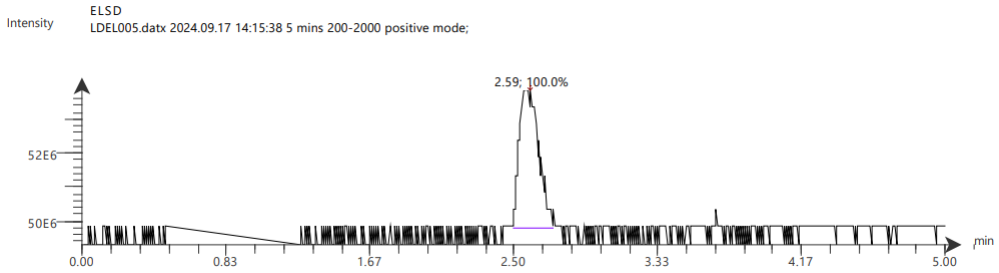


LDEL005

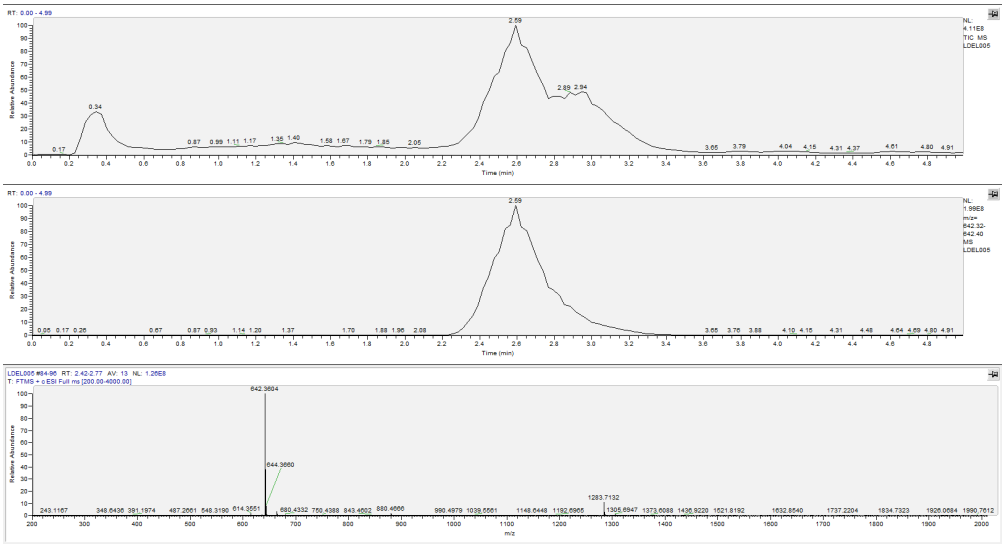


LDEL005
LA18-ßHomoleu-LA03
Exact Mass: 641.3537

ELSD of **LDEL005**

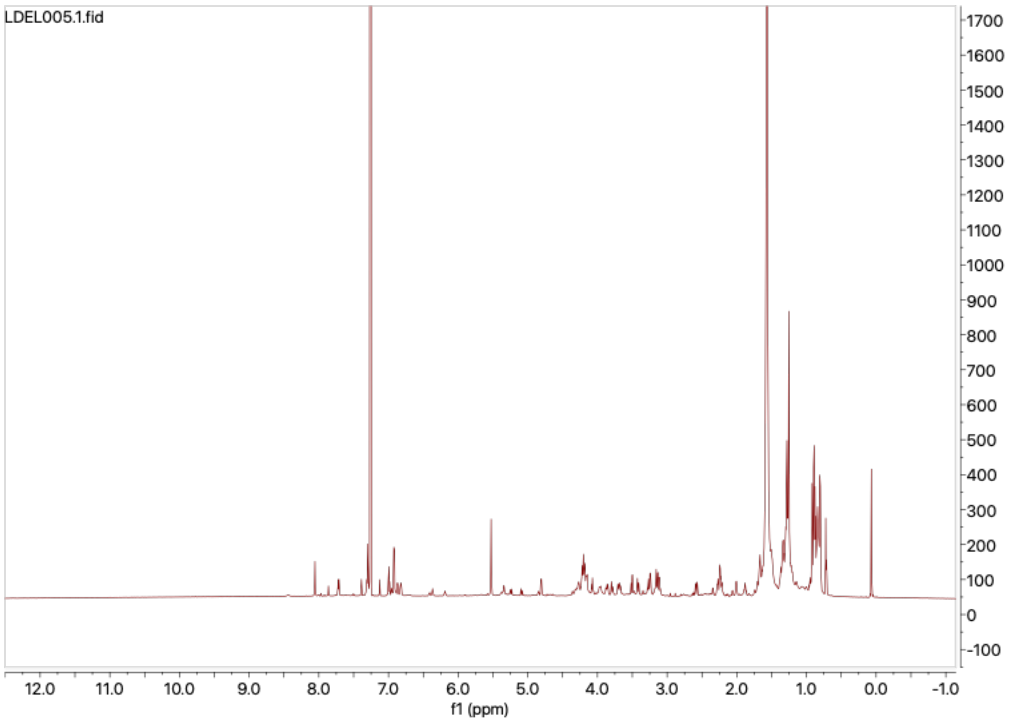


HRMS of **LDEL005**

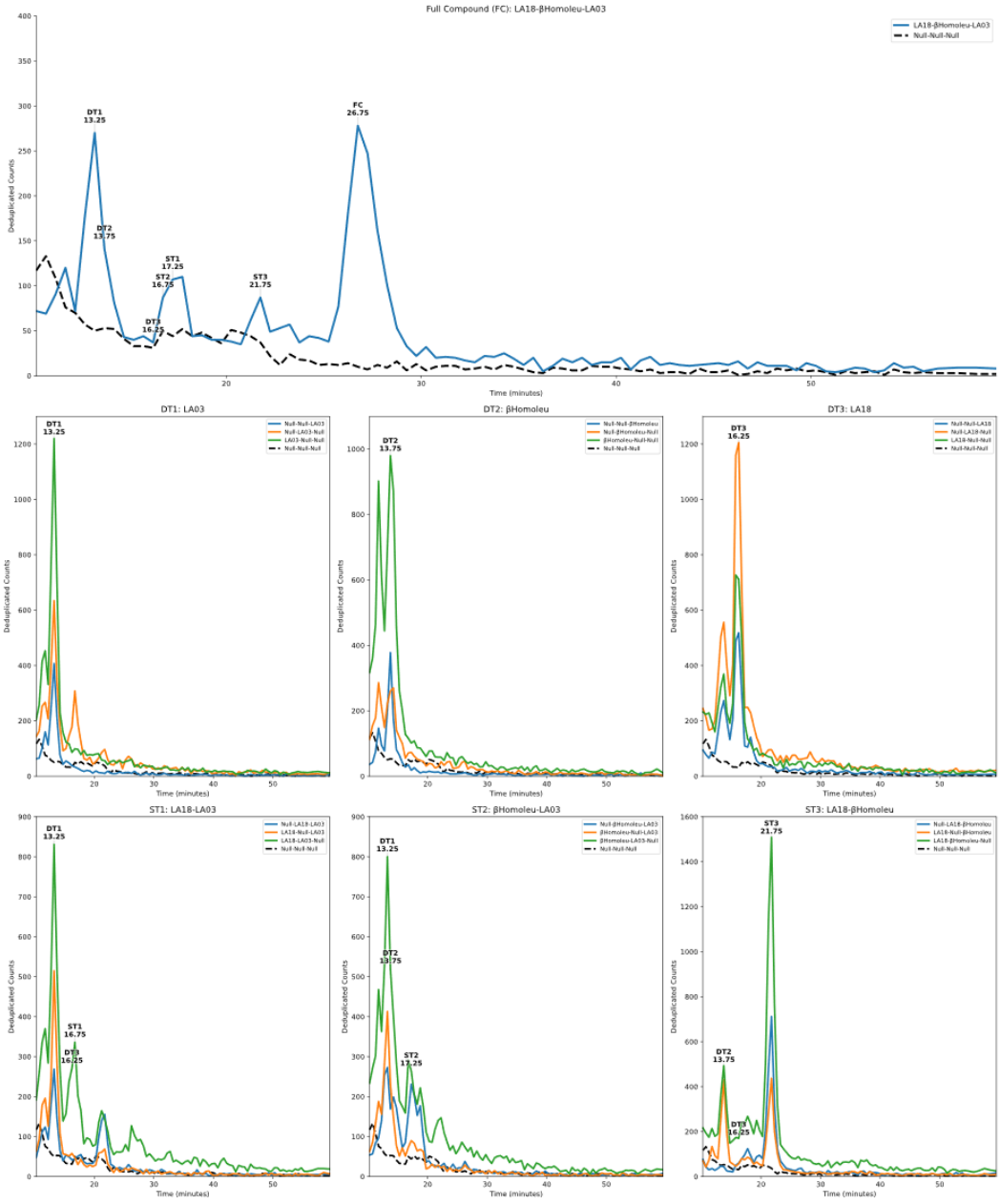


HRMS (ESI) m/z calculated for $C_{32}H_{47}N_7O_7^+$ ($M+H^+$) 642.3571; found 642.3602

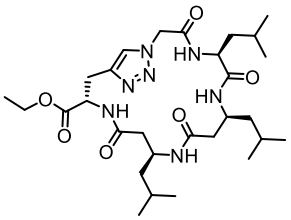
^{1}HNMR of **LDEL005**



Classic Chrome Sequence Count Chromatogram of **LDEL005**

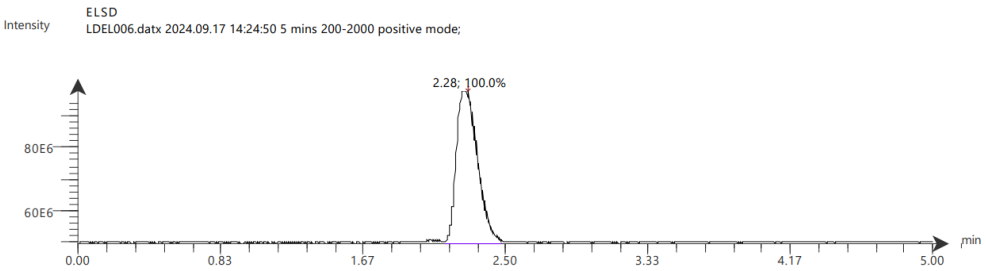


LDEL006

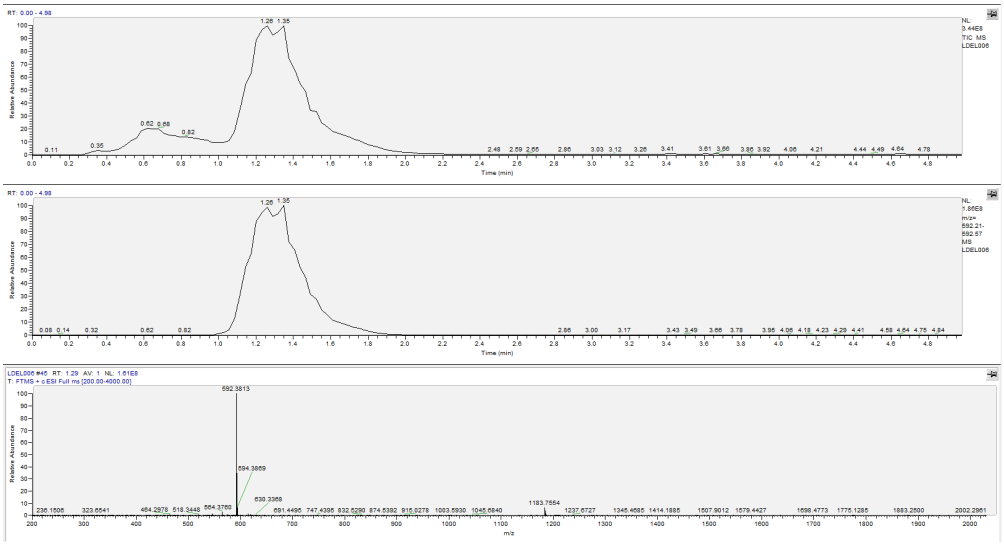


LDEL006
Leu-ßHomoleu-ßHomoleu
Exact Mass: 591.3744

ELSD of **LDEL006**

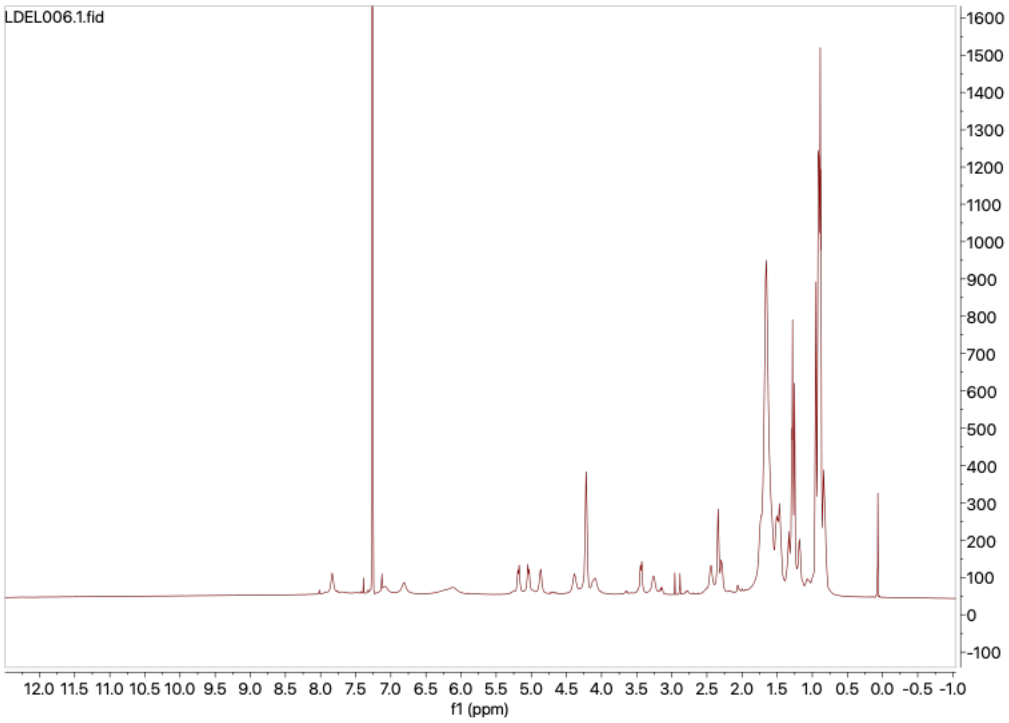


HRMS of **LDEL006**

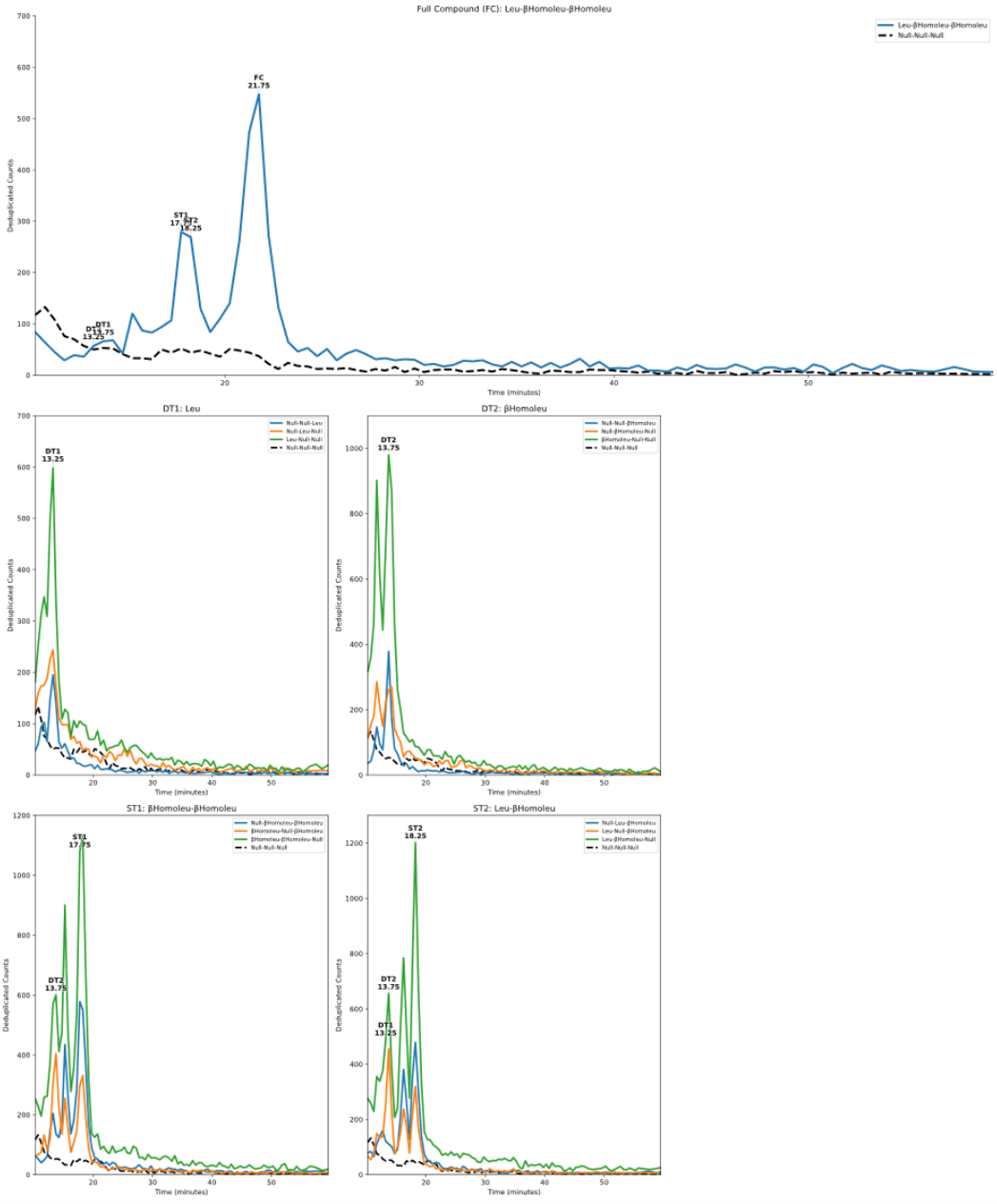


HRMS (ESI) m/z calculated for $C_{29}H_{49}N_7O_6^+$ ($M+H^+$) 592.3778; found 592.3813

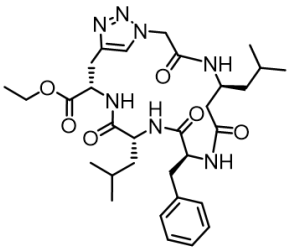
^{1}HNMR of **LDEL006**



Classic Chrome Sequence Count Chromatogram of **LDEL006**

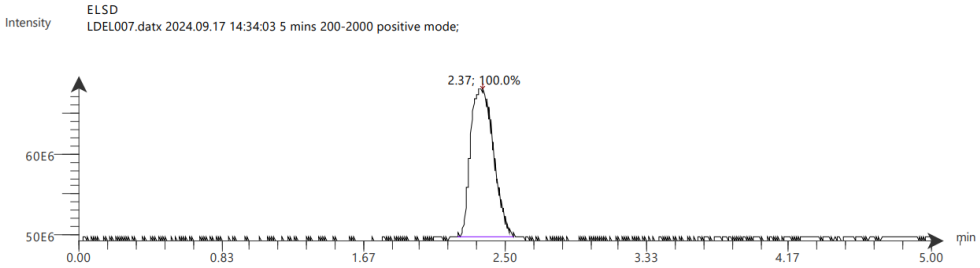


LDEL007

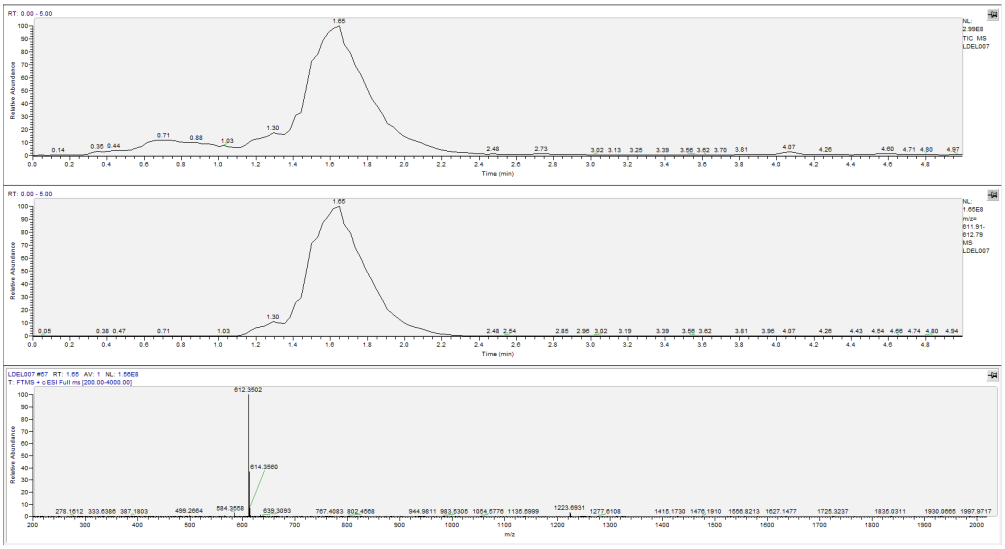


LDEL007
ßHomoleu-Phe-DLeu
Exact Mass: 611.3431

ELSD of **LDEL007**

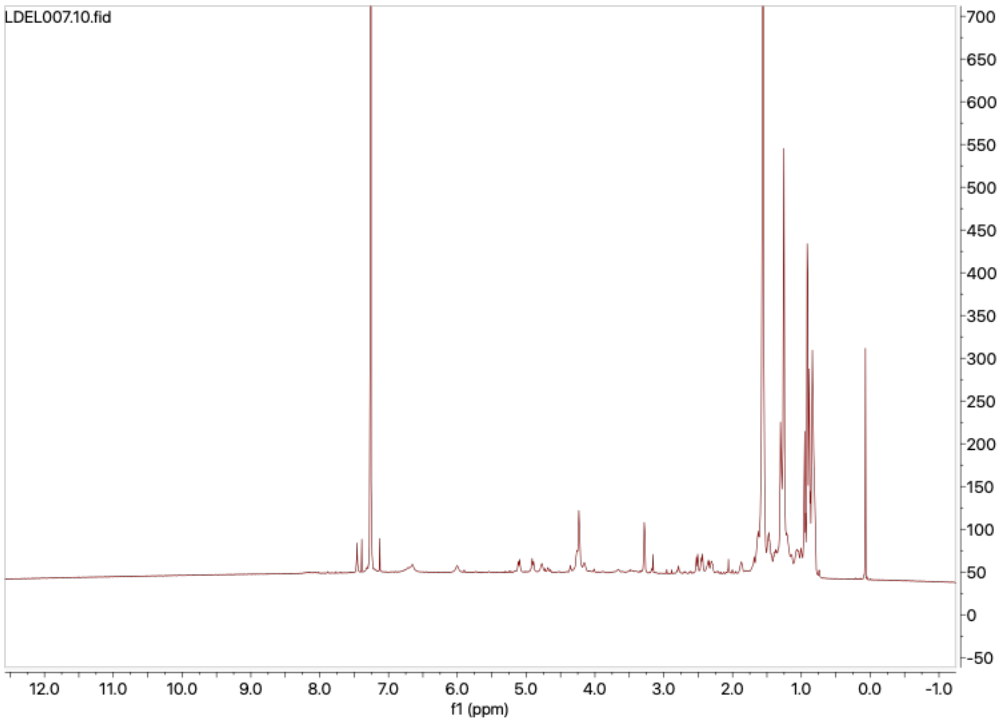


HRMS of **LDEL007**

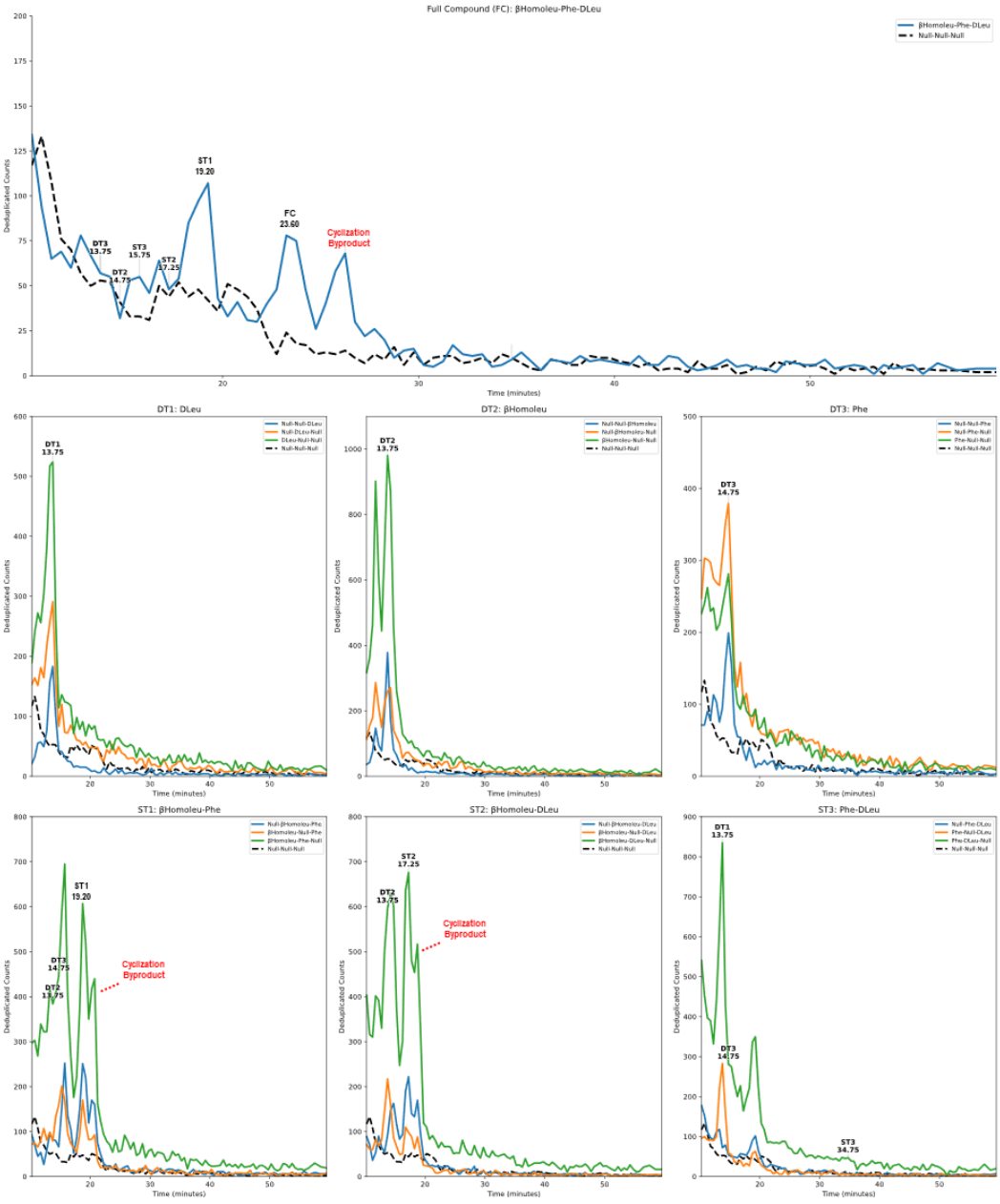


HRMS (ESI) m/z calculated for $C_{31}H_{45}N_7O_6^+$ ($M+H^+$) 612.3465; found 612.3502

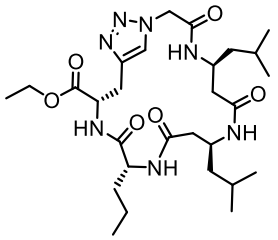
^{1}HNMR of **LDEL007**



Classic Chrome Sequence Count Chromatogram of **LDEL007**

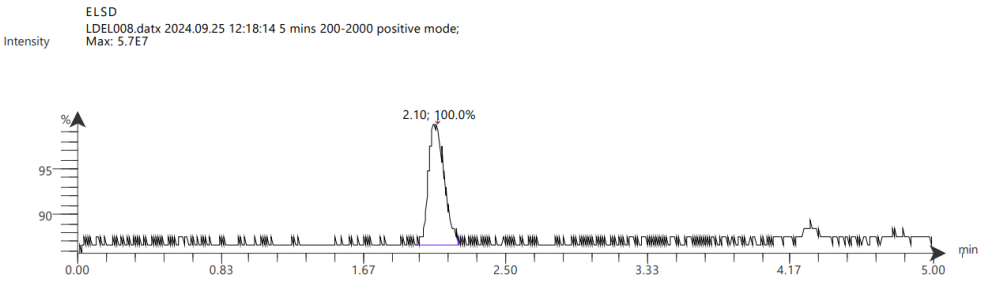


LDEL008

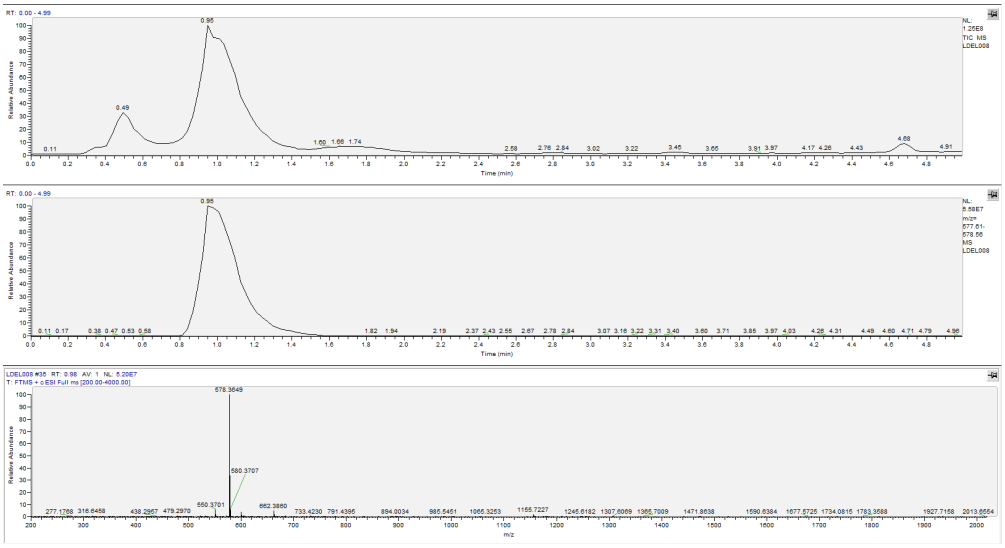


LDEL008
ßHomoleu-ßHomoleu-DNvl
Exact Mass: 577.3588

ELSD of **LDEL008**

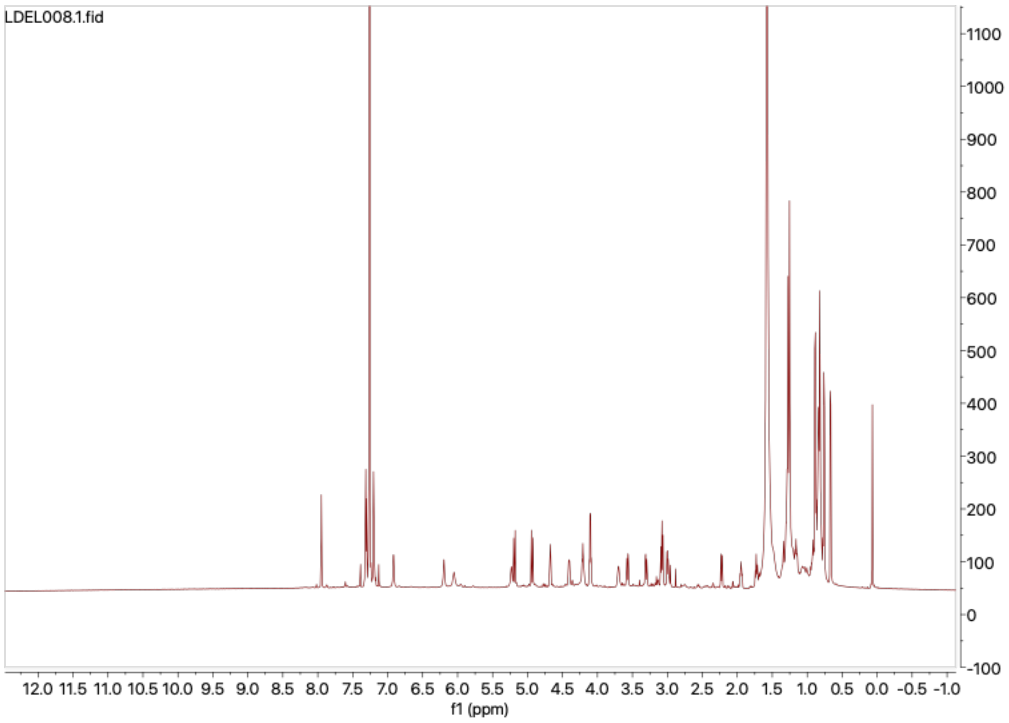


HRMS of **LDEL008**

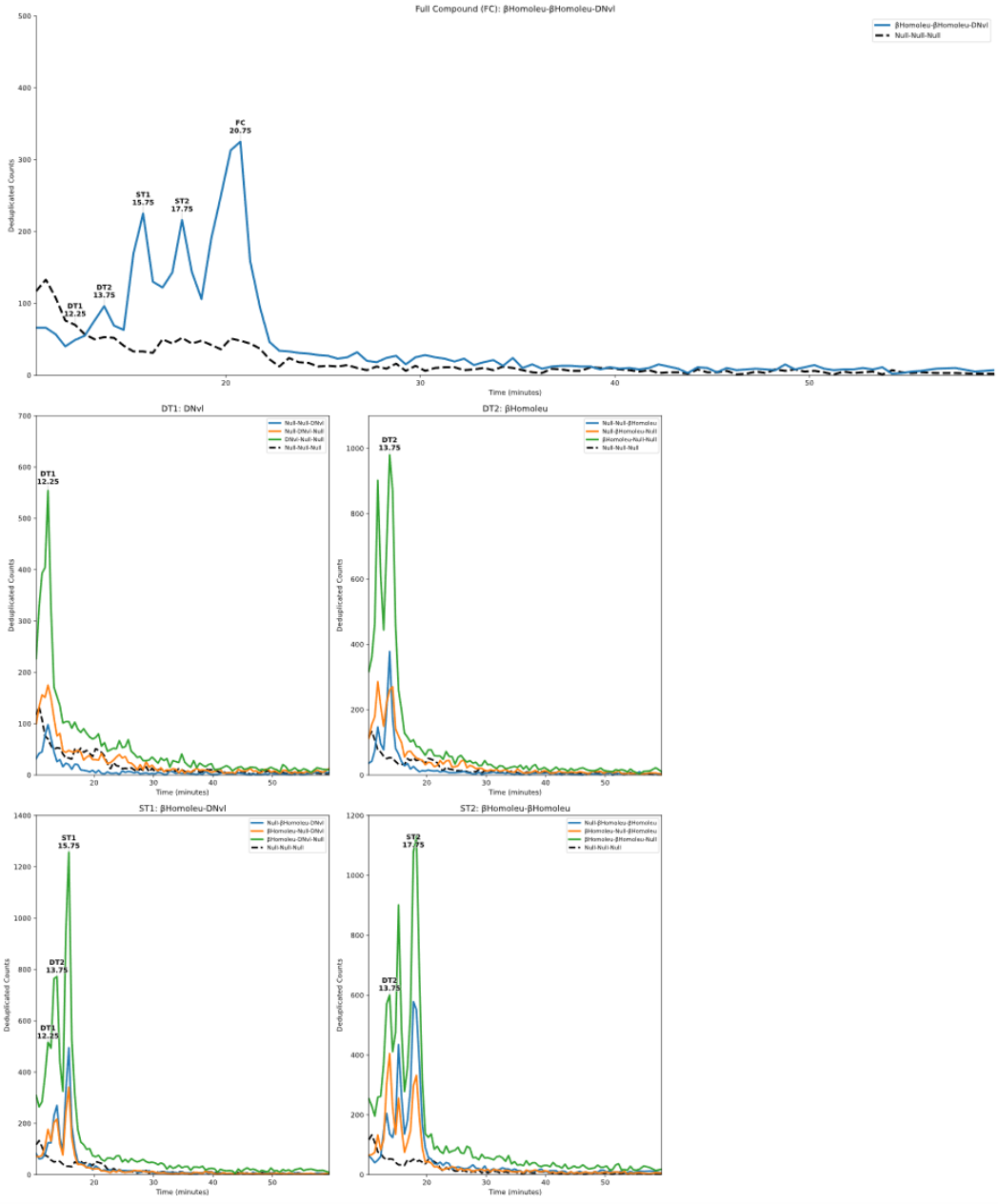


HRMS (ESI) m/z calculated for $C_{28}H_{47}N_7O_6^+$ ($M+H^+$) 578.3621; found 578.3649

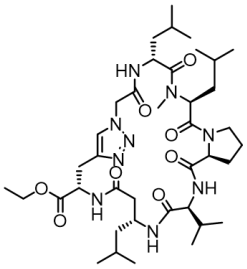
[1]HNMR of **LDEL008**



Classic Chrome Sequence Count Chromatogram of **LDEL008**

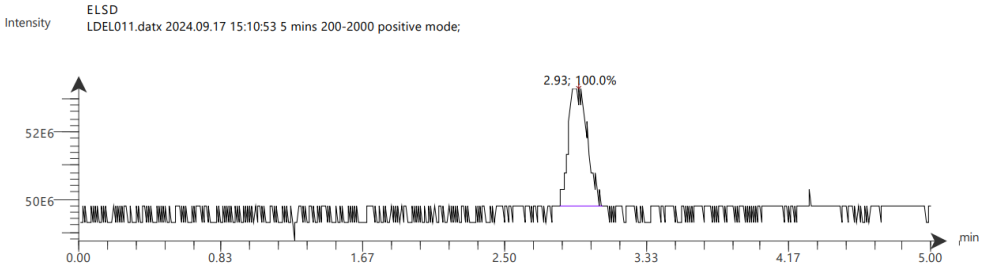


LDEL011

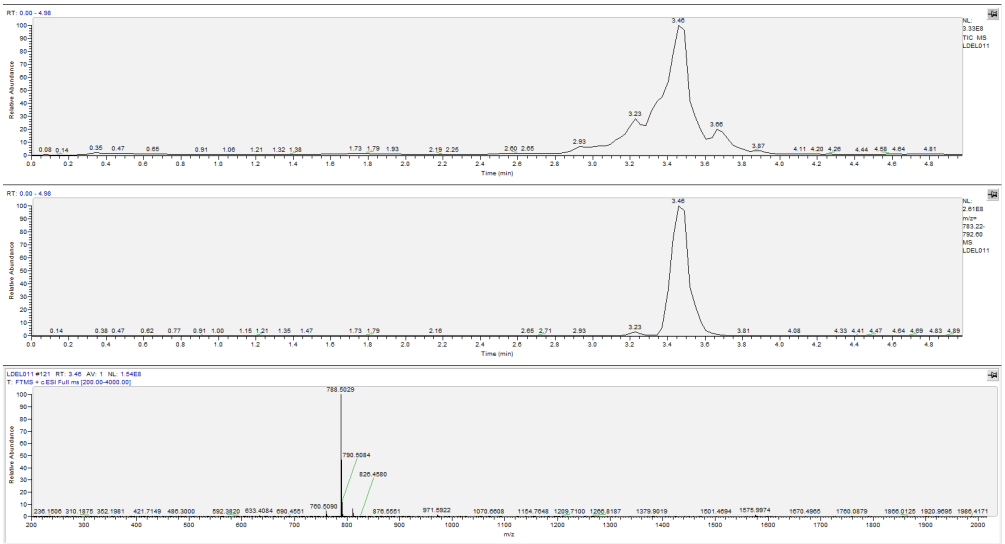


LDEL011
DLeu-LeuMe-Pro-Val-DßHomoleu
Exact Mass: 787.4956

ELSD of **LDEL011**

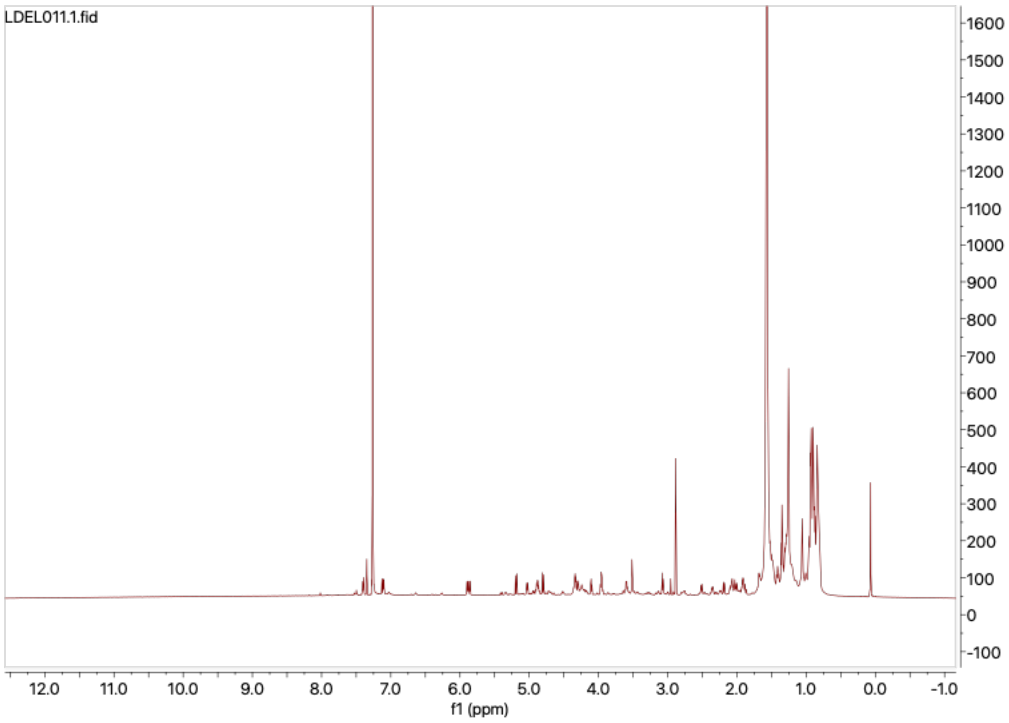


HRMS of **LDEL011**

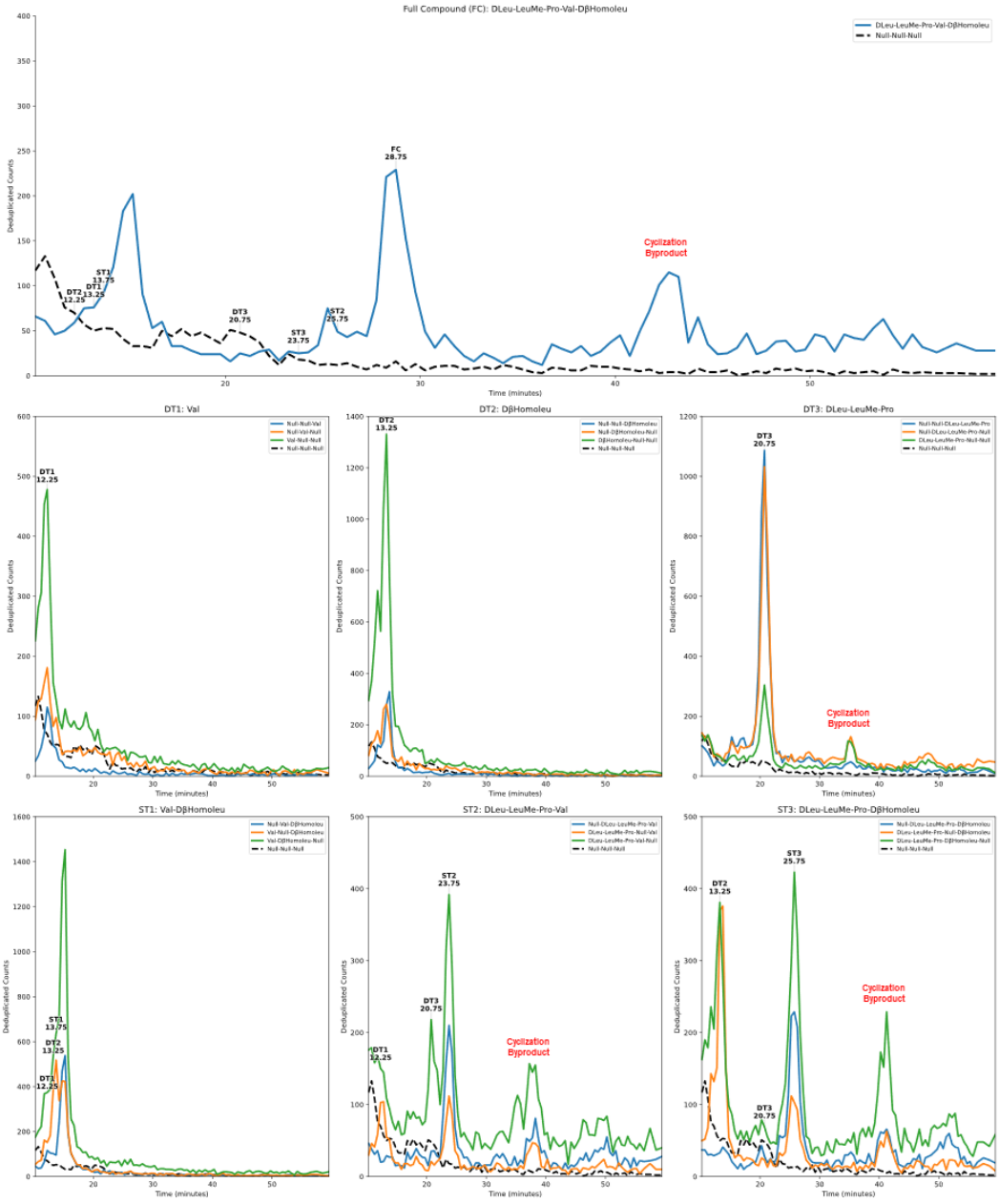


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.5029

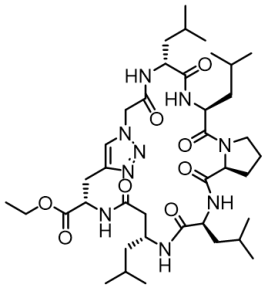
[1]HNMR of **LDEL011**



Classic Chrome Sequence Count Chromatogram of **LDEL011**

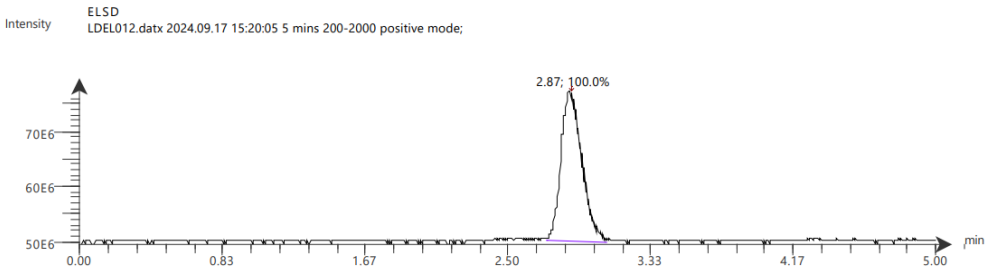


LDEL012

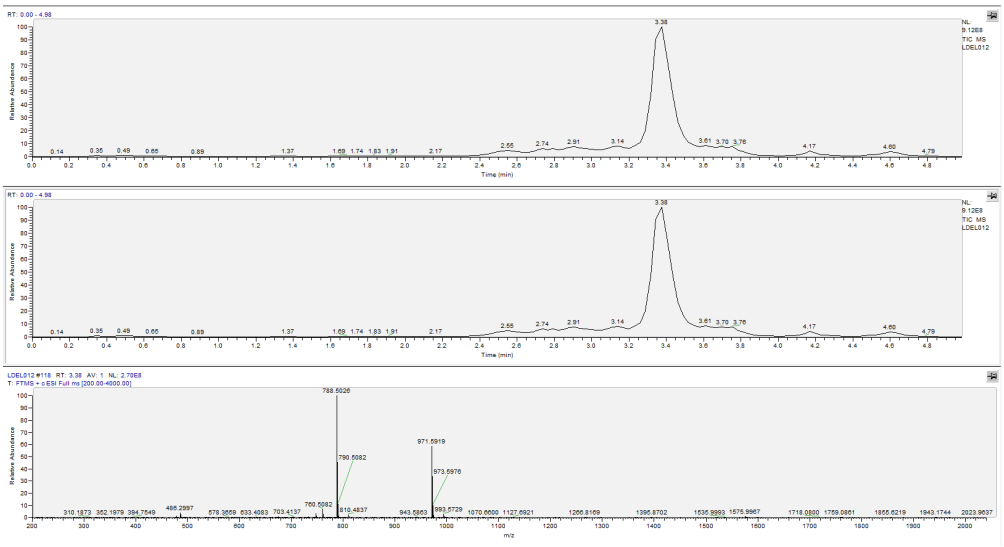


LDEL012
DLeu-Leu-Pro-Leu-DßHomoleu
Exact Mass: 787.4956

ELSD of **LDEL012**

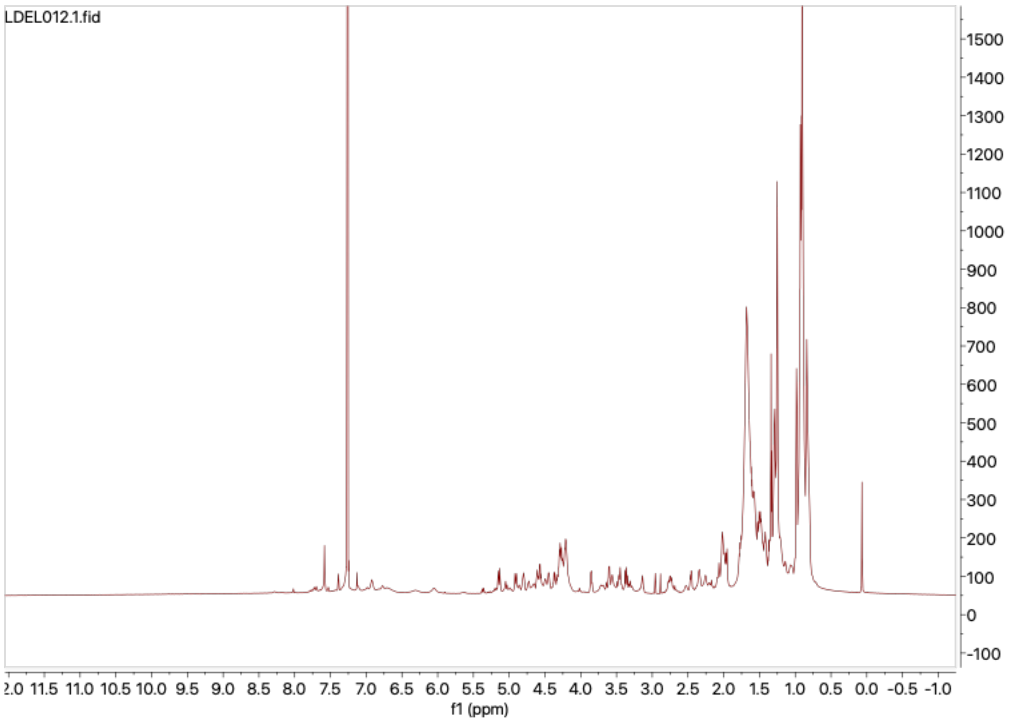


HRMS of **LDEL012**

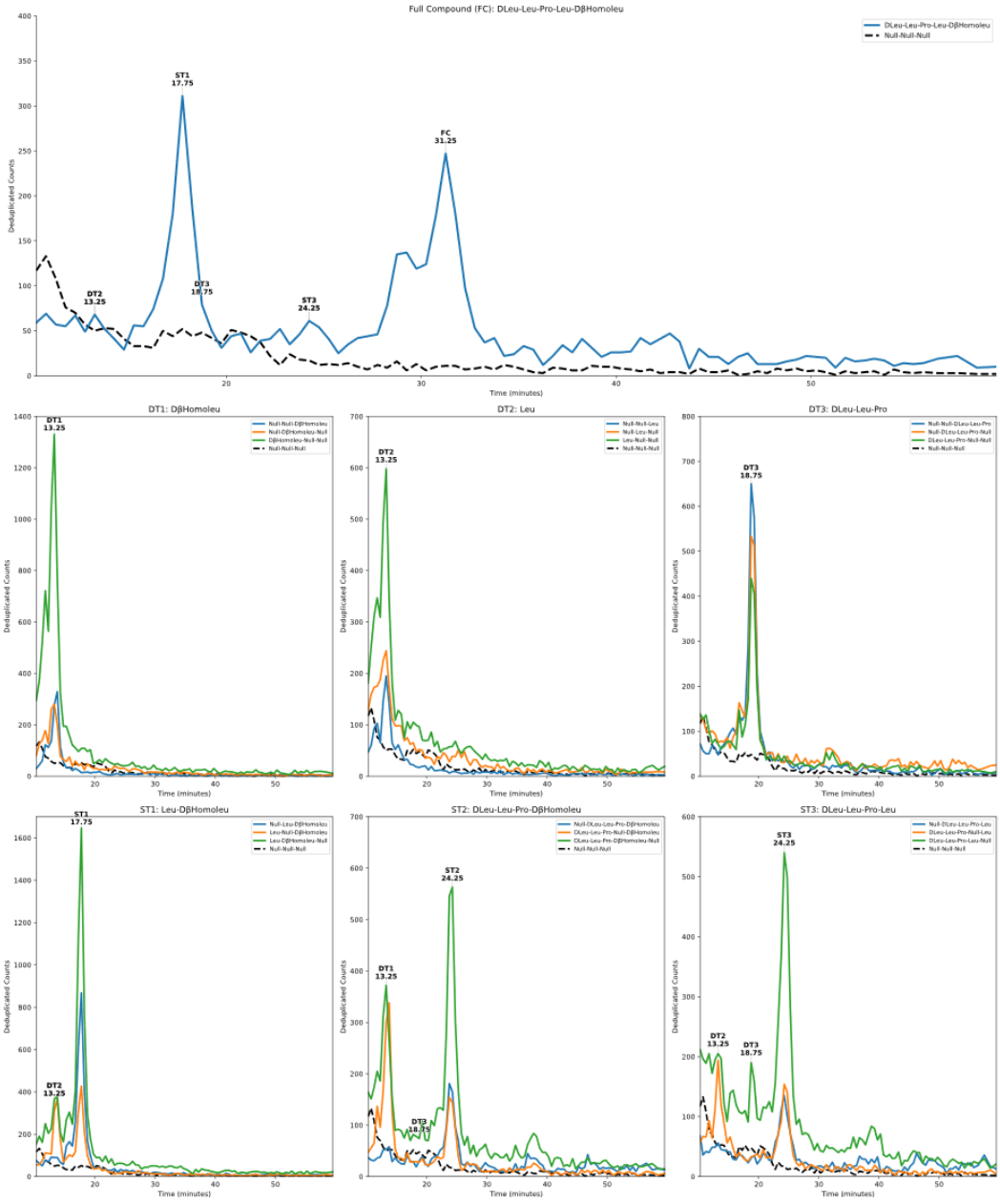


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.5026

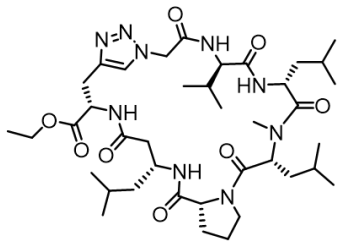
^{1}HNMR of **LDEL012**



Classic Chrome Sequence Count Chromatogram of **LDEL012**

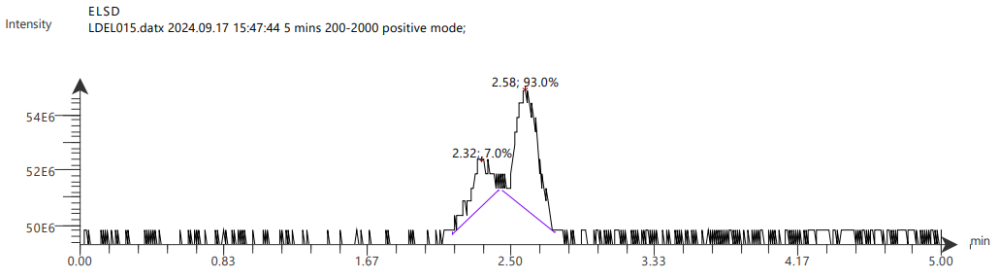


LDEL015

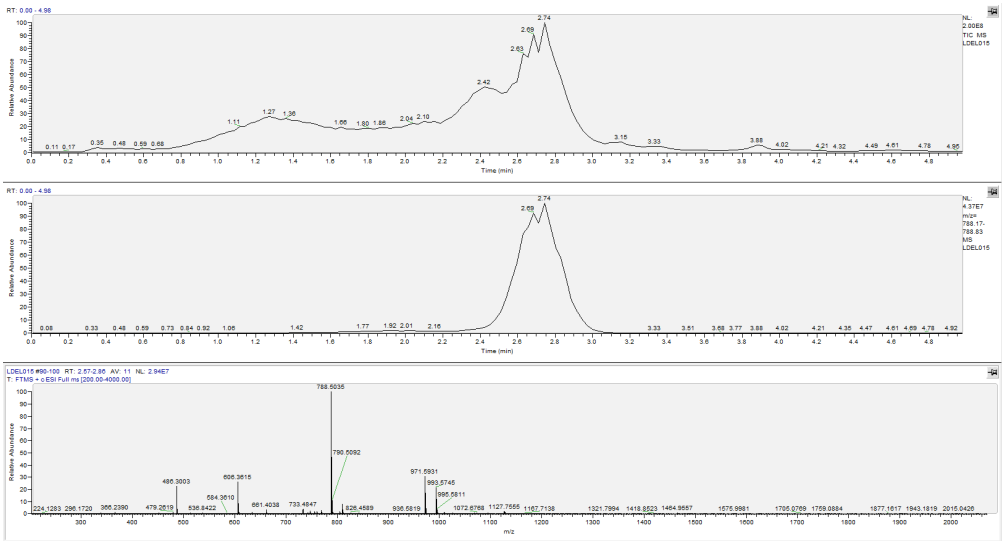


LDEL015
DVal-DLeu-DLeuMe-DPro-DßHomoleu
Exact Mass: 787.4956

ELSD of **LDEL015**

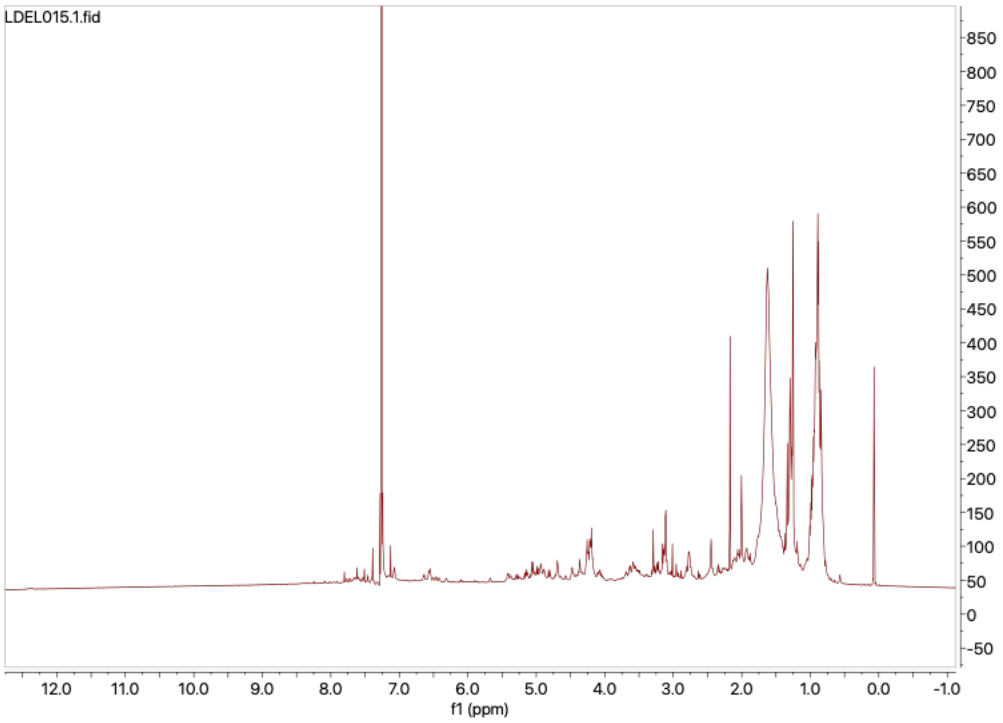


HRMS of **LDEL015**

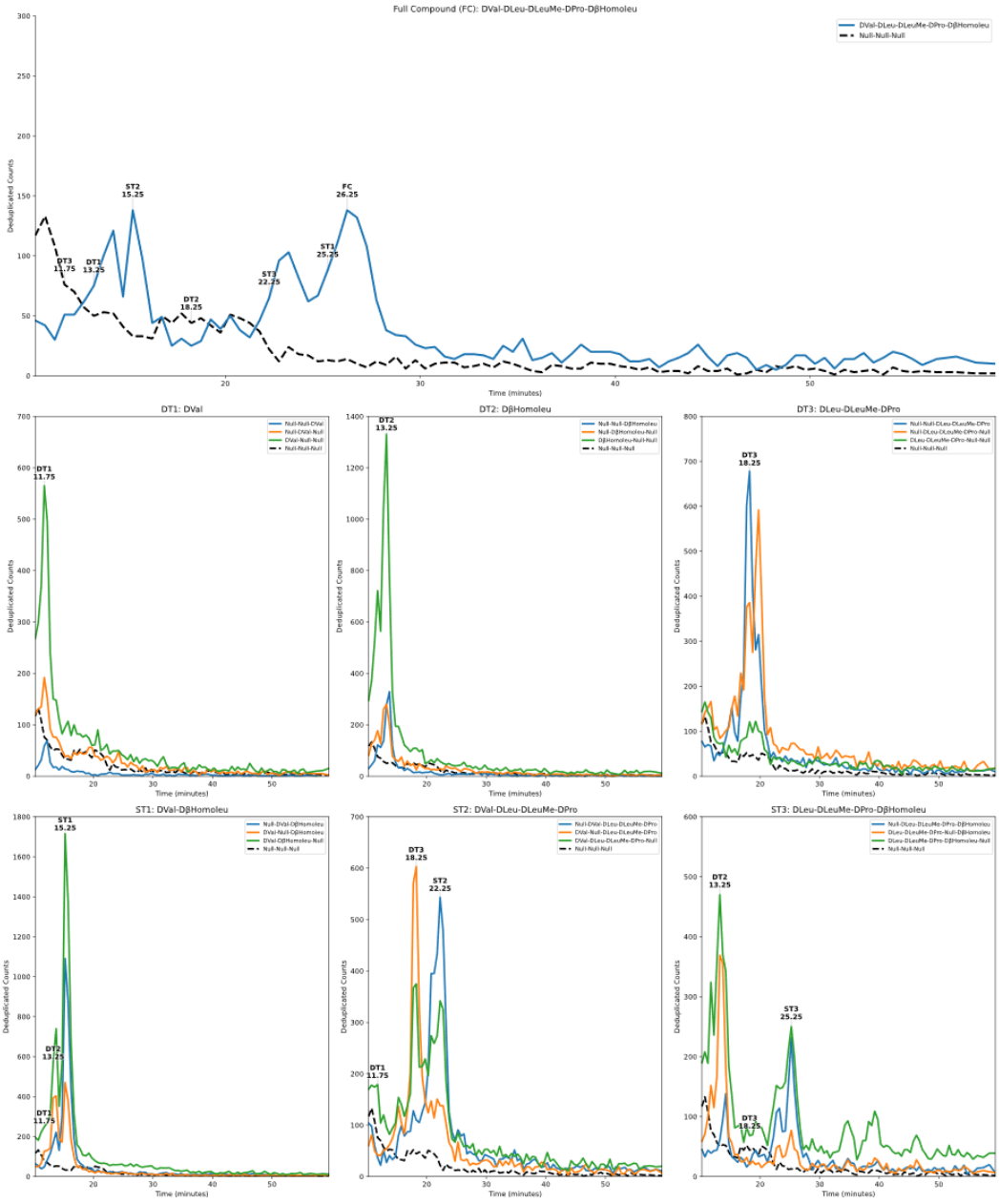


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.5035

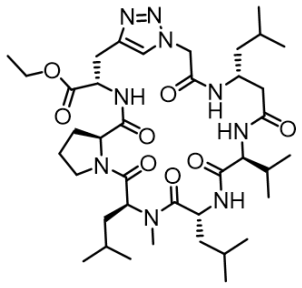
^{1}HNMR of **LDEL015**



Classic Chrome Sequence Count Chromatogram of **LDEL015**

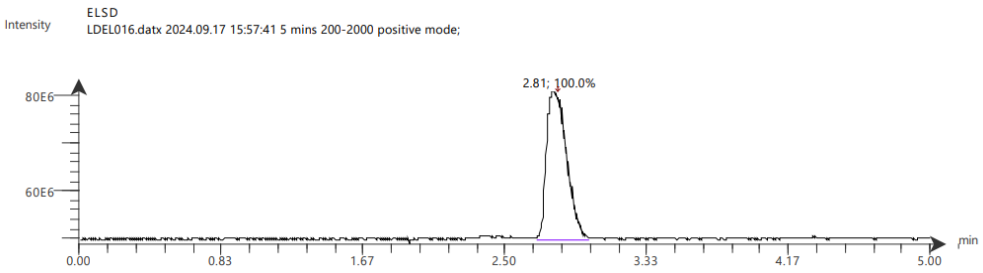


LDEL016

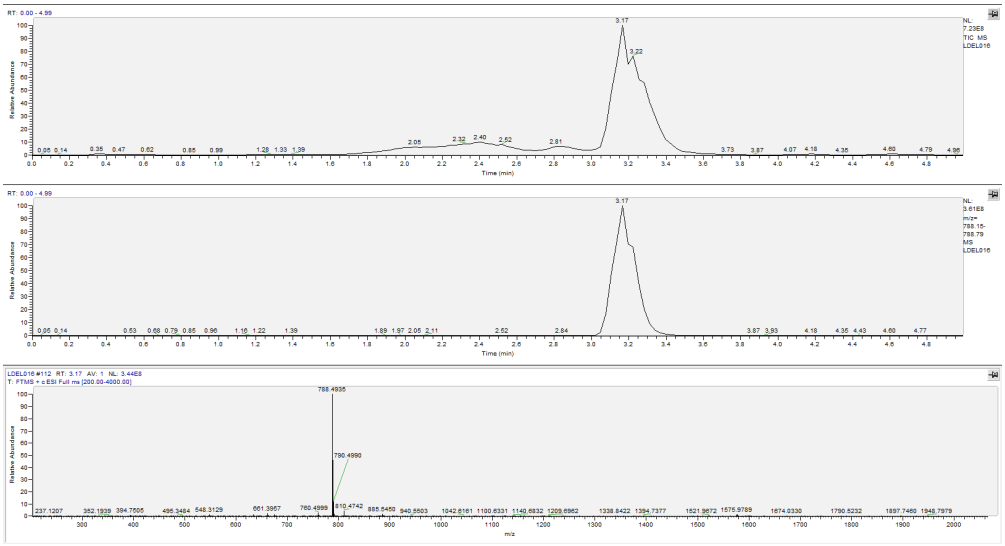


LDEL016
DßHomoleu-Val-DLeu-LeuMe-Pro
Exact Mass: 787.4956

ELSD of **LDEL016**

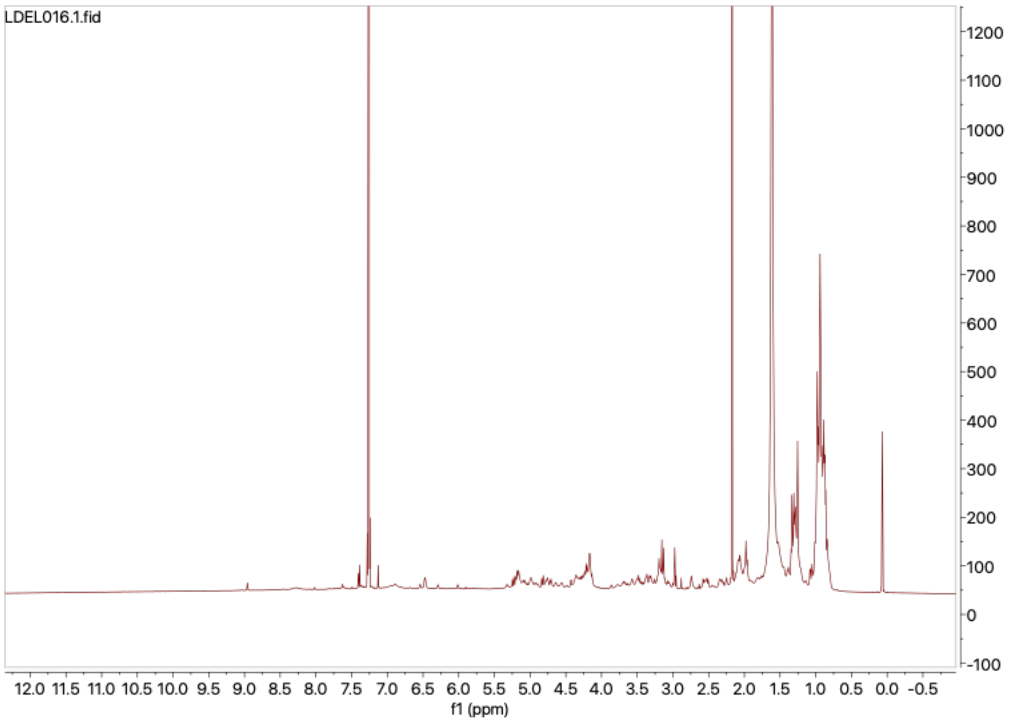


HRMS of **LDEL016**

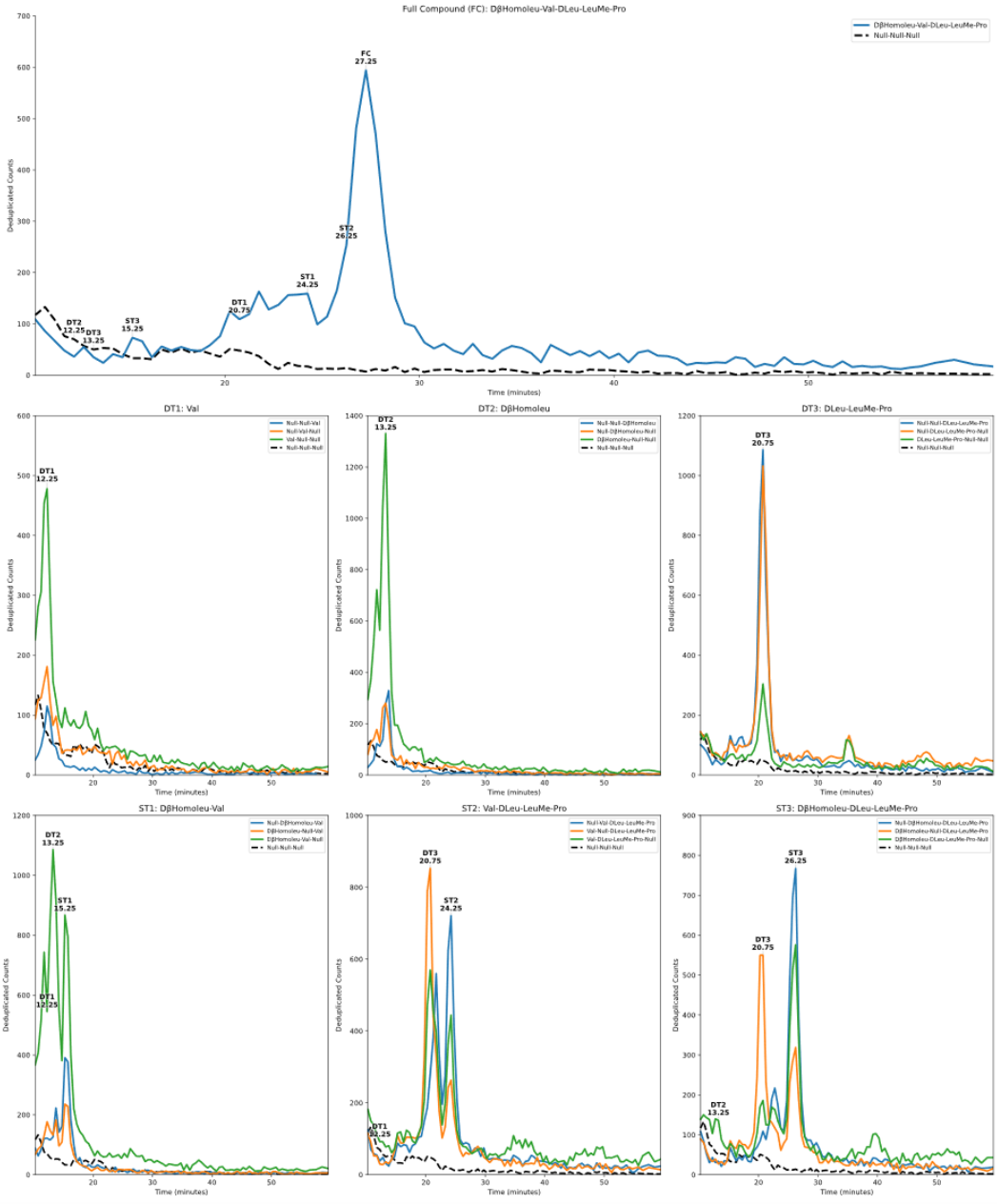


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.4935

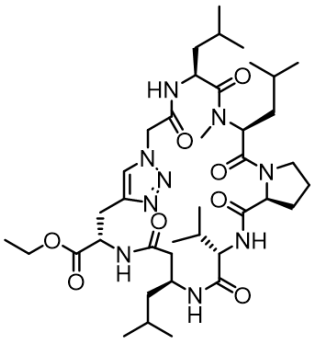
^{1}HNMR of **LDEL016**



Classic Chrome Sequence Count Chromatogram of **LDEL016**

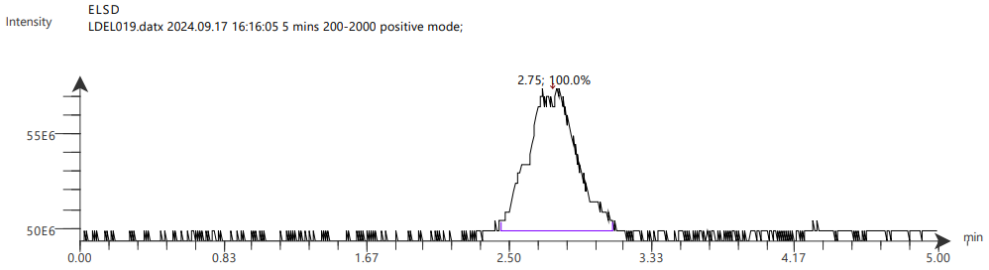


LDEL019

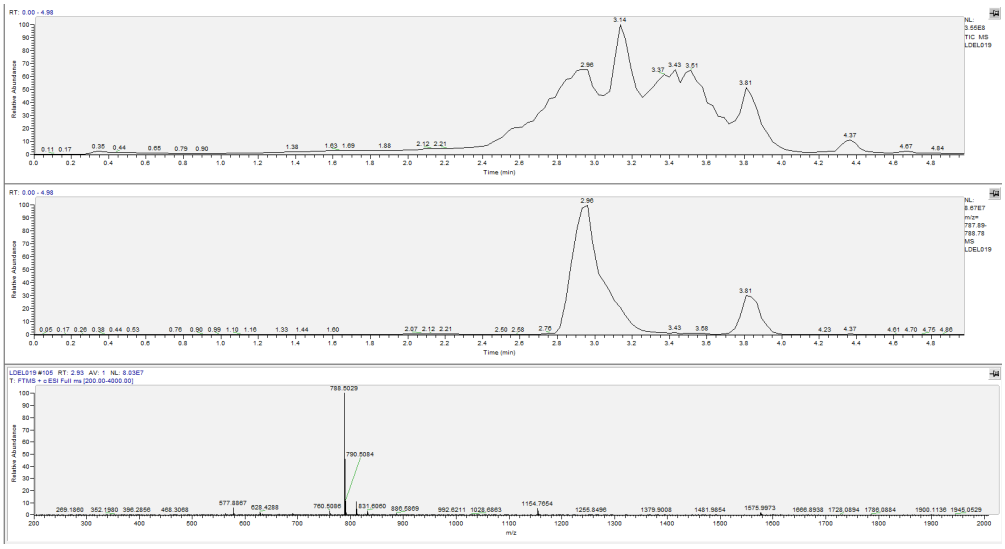


LDEL019
Leu-LeuMe-Pro-Val-ßHomoleu
Exact Mass: 787.4956

ELSD of **LDEL019**

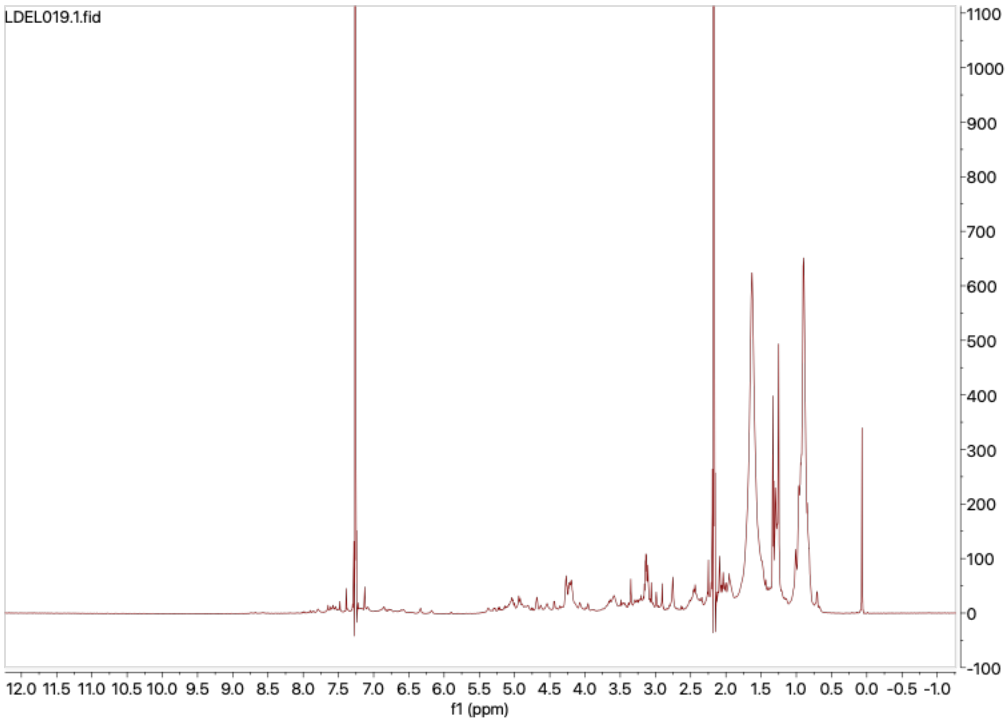


HRMS of **LDEL019**

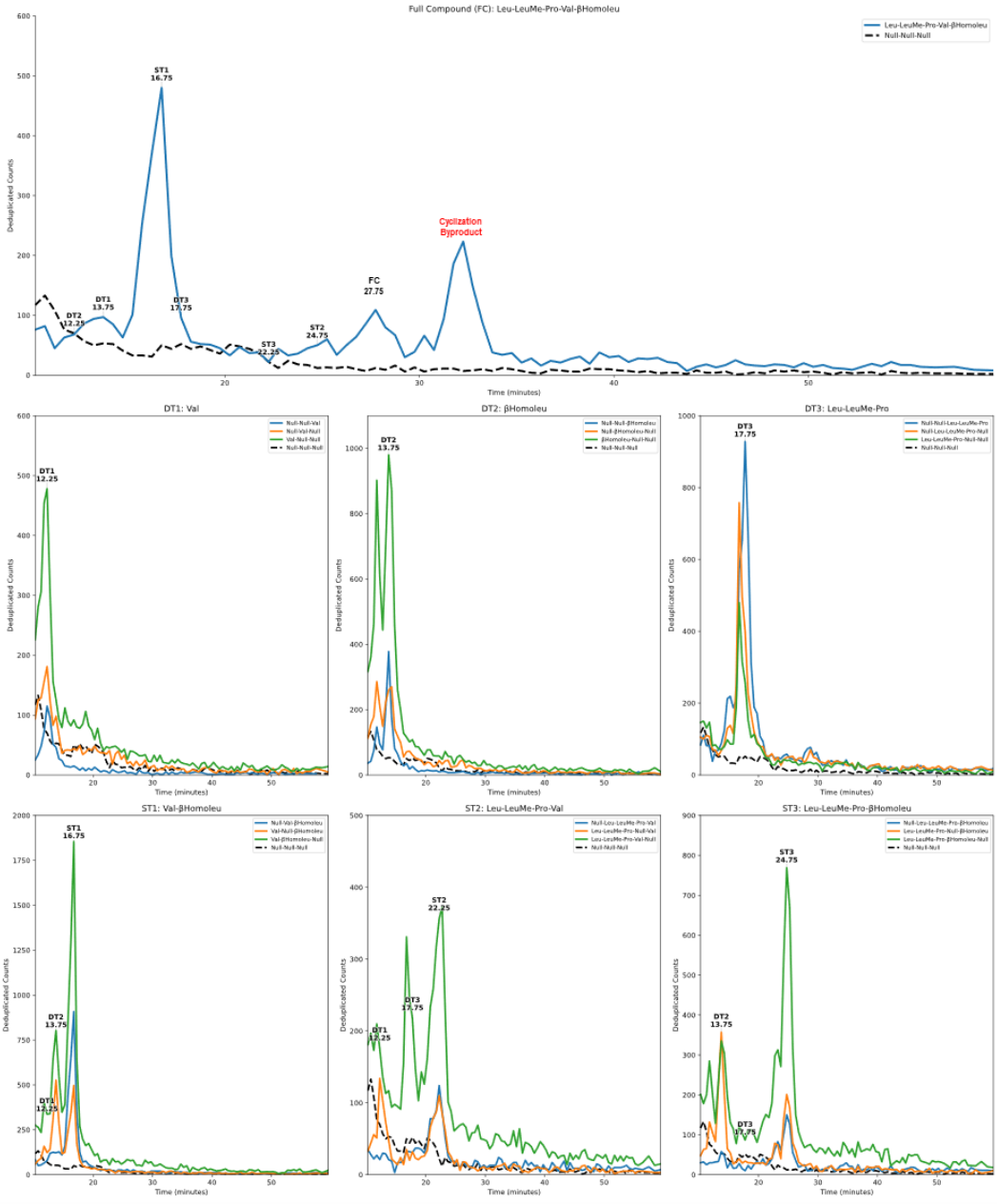


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.5029

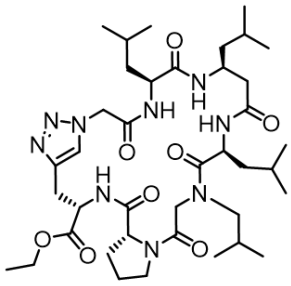
^{1}HNMR of **LDEL019**



Classic Chrome Sequence Count Chromatogram of **LDEL019**

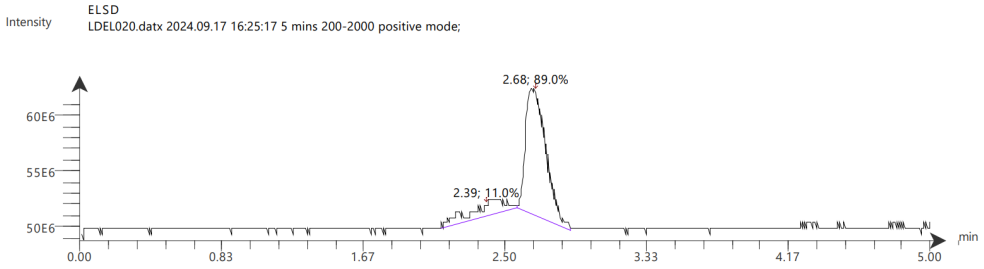


LDEL020

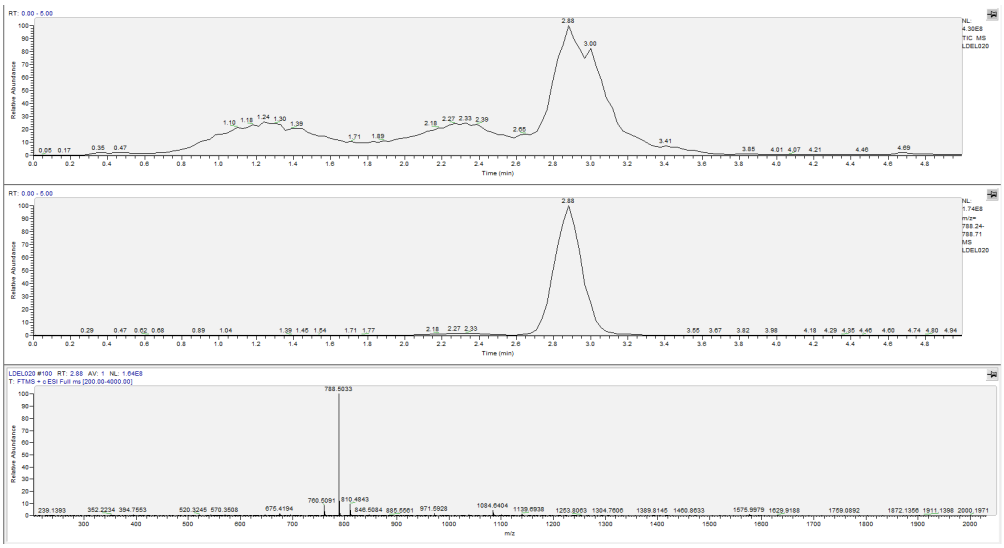


LDEL020
Leu-ßHomoleu-Leu-LA03-DPro
Exact Mass: 787.4956

ELSD of **LDEL020**

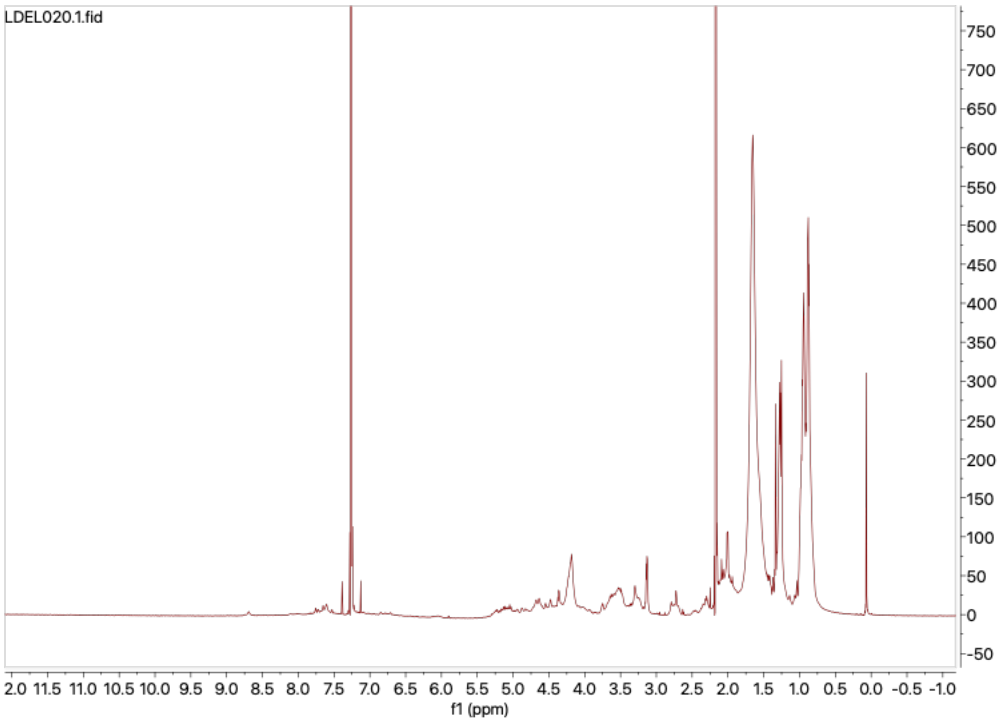


HRMS of **LDEL020**

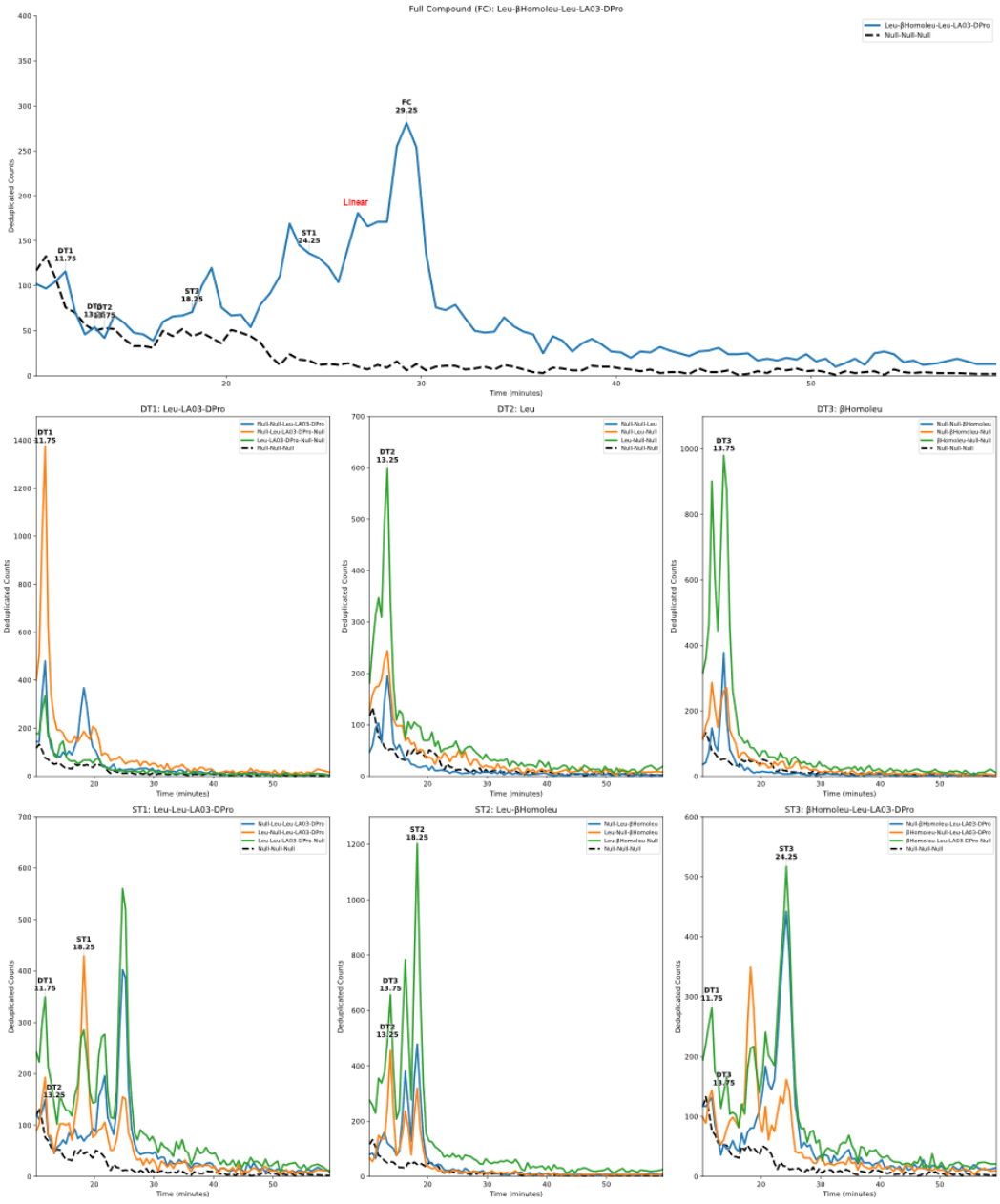


HRMS (ESI) m/z calculated for $C_{39}H_{65}N_9O_8^+$ ($M+H^+$) 788.4990; found 788.5033

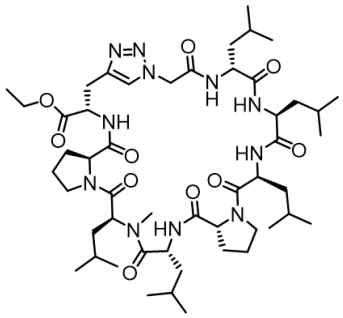
^{1}HNMR of **LDEL020**



Classic Chrome Sequence Count Chromatogram of **LDEL020**



LDEL023

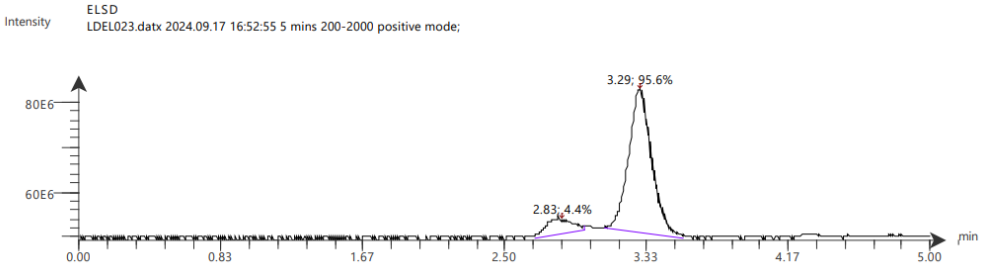


LDEL023

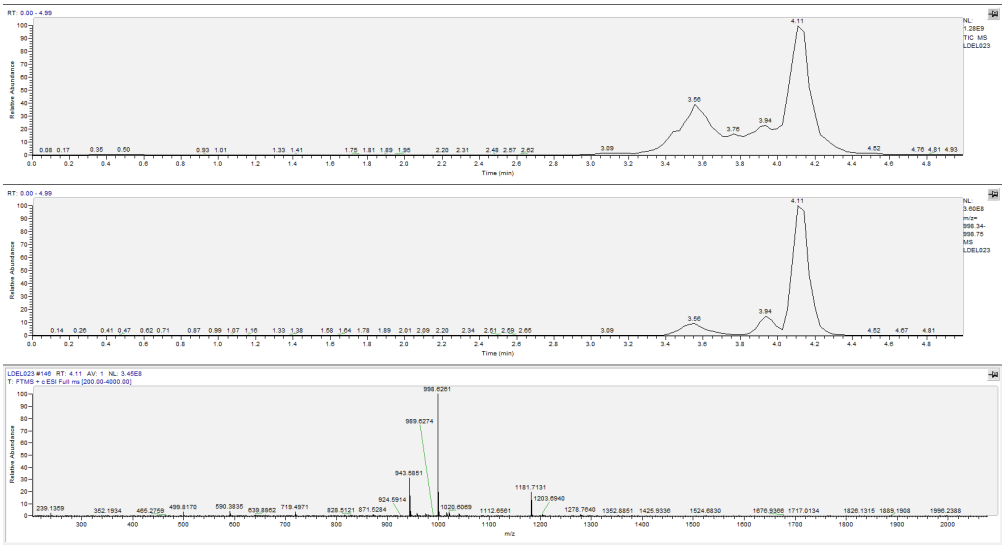
DLeu-Leu-Leu-DPro-DLeu-LeuMe-Pro

Exact Mass: 997.6324

ELSD of **LDEL023**

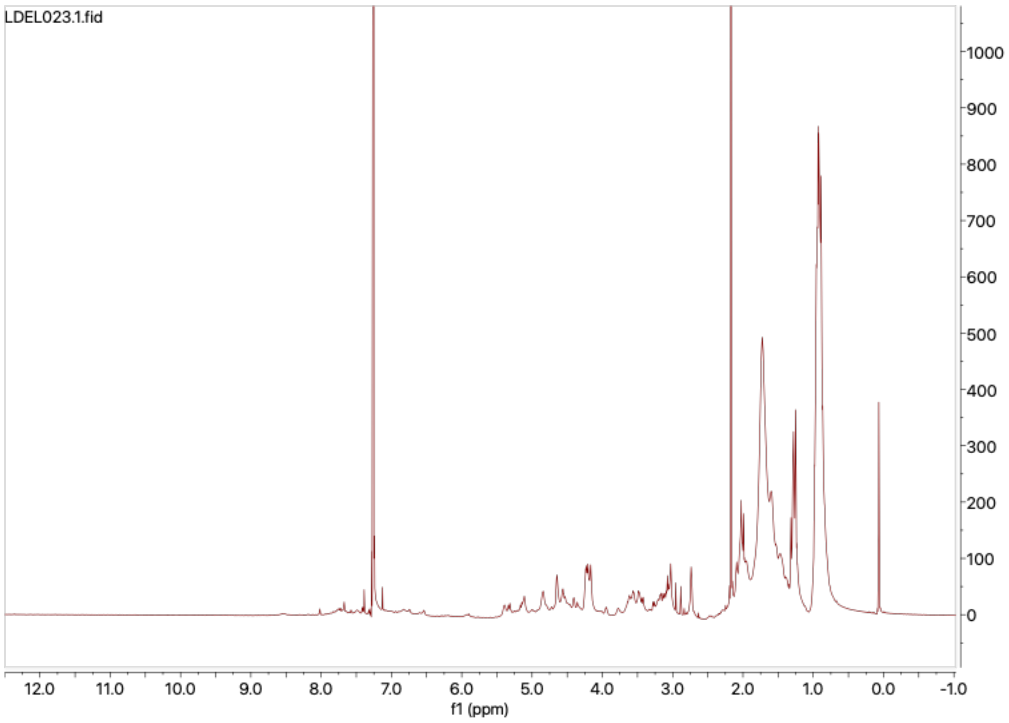


HRMS of **LDEL023**



HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 998.6358; found 998.6261

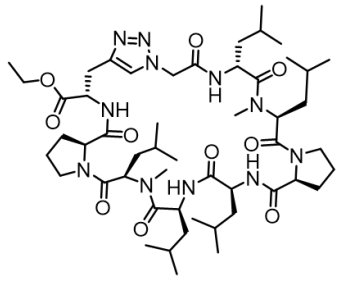
^{1}HNMR of **LDEL023**



Classic Chrome Sequence Count Chromatogram of **LDEL023**



LDEL024

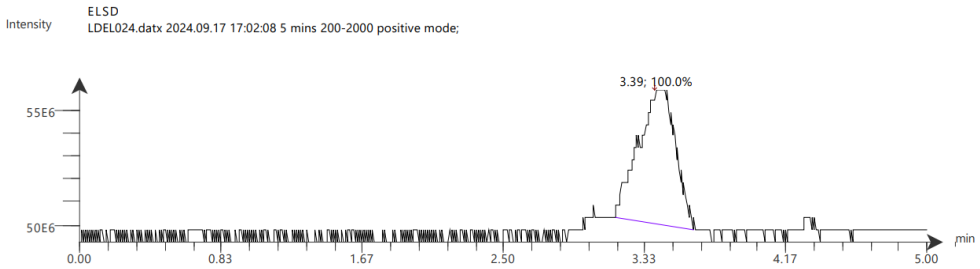


LDEL024

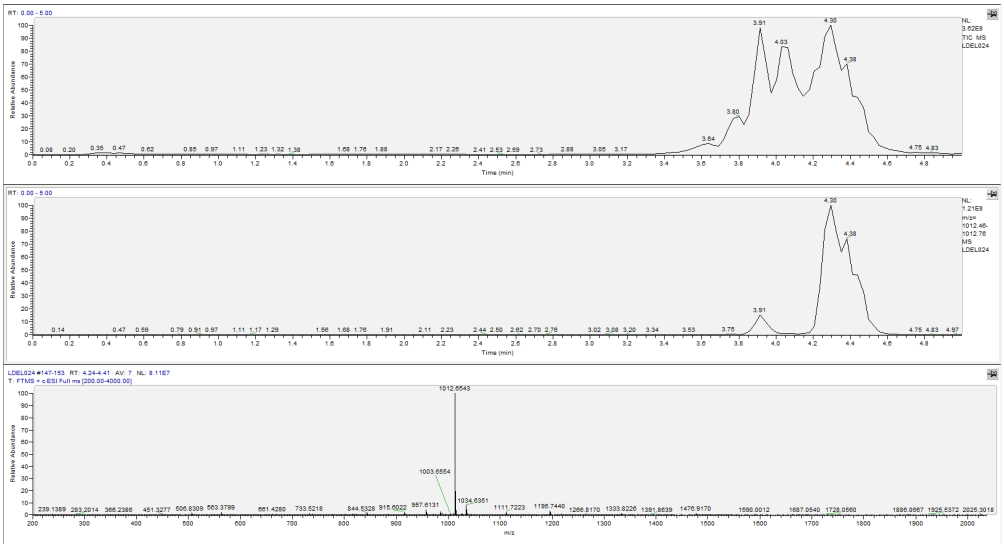
DLeu-LeuMe-Pro-Leu-Leu-DLeuMe-Pro

Exact Mass: 1011.6481

ELSD of **LDEL024**

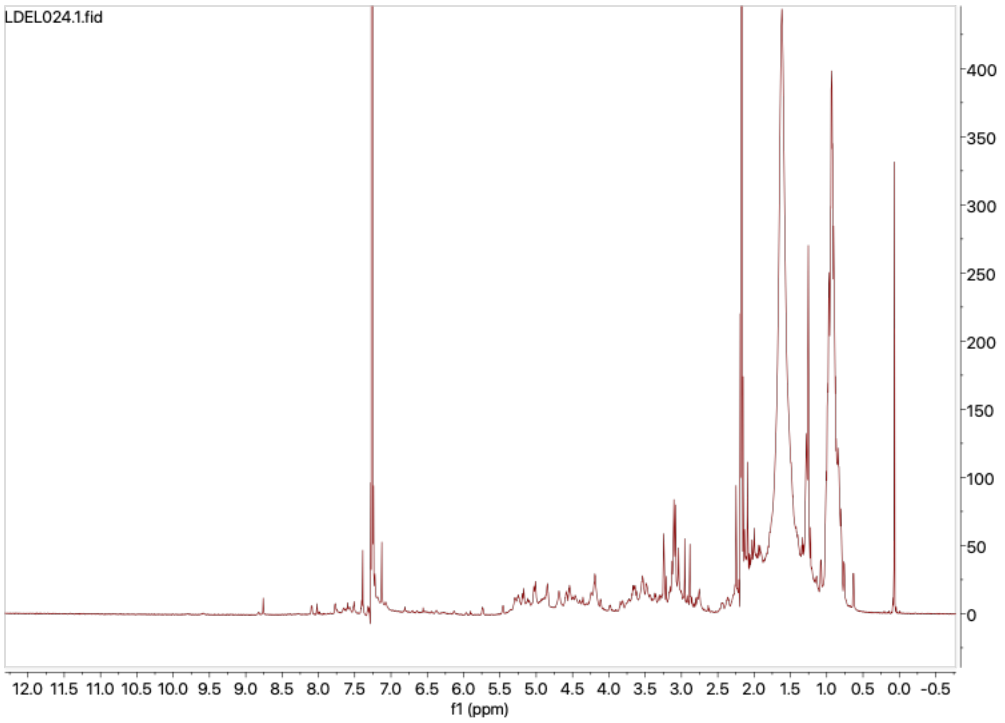


HRMS of **LDEL024**

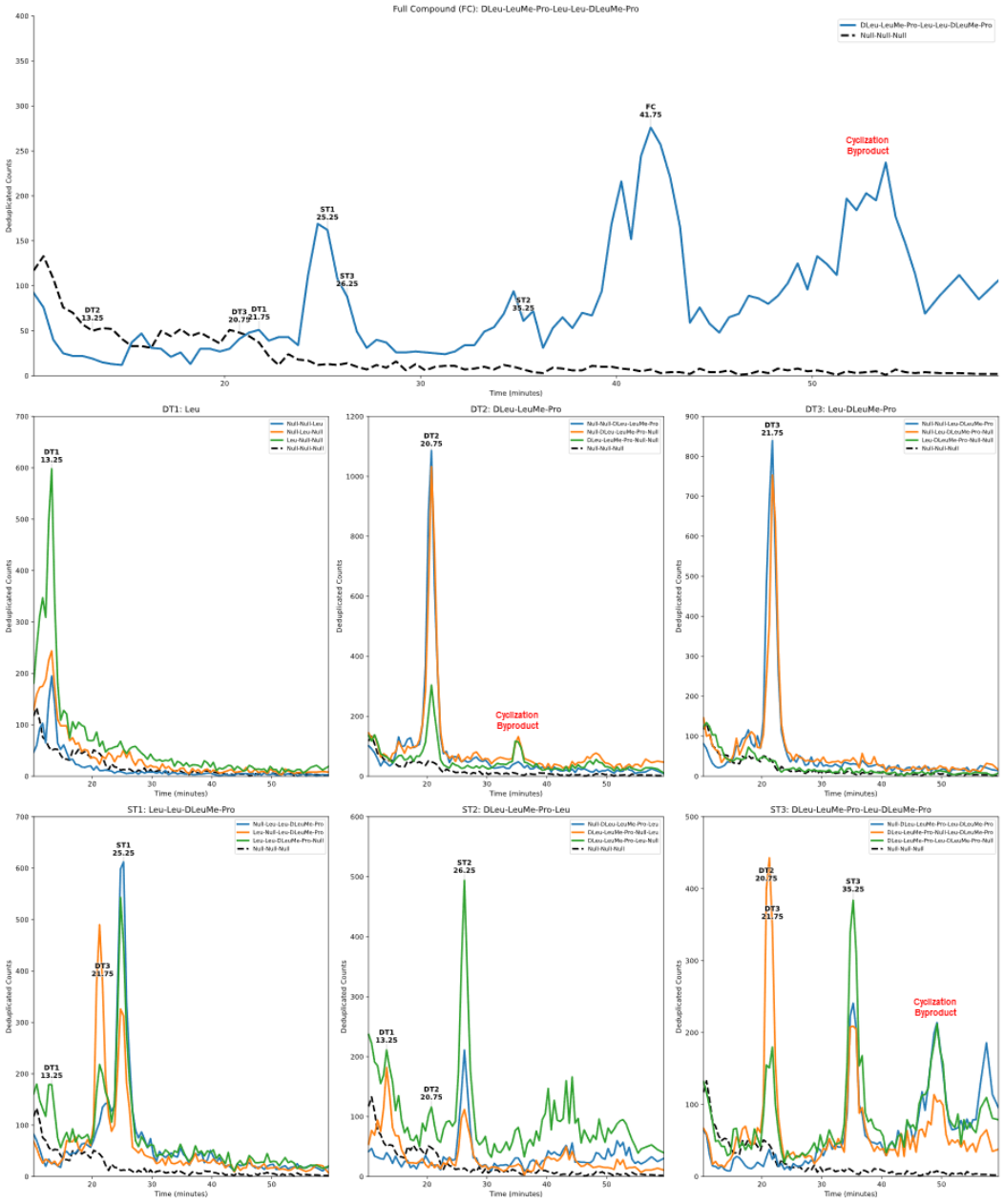


HRMS (ESI) m/z calculated for $C_{51}H_{85}N_{11}O_{10}^{+}$ ($M+H^{+}$) 1012.6514; found 1012.6543

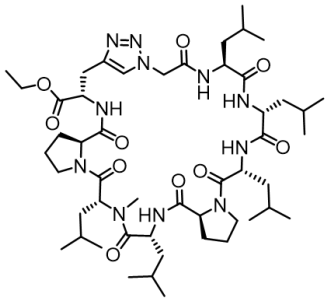
^{1}HNMR of **LDEL024**



Classic Chrome Sequence Count Chromatogram of **LDEL024**



LDEL025

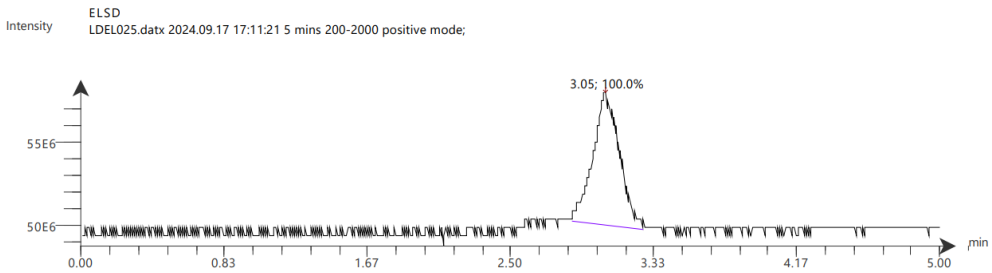


LDEL025

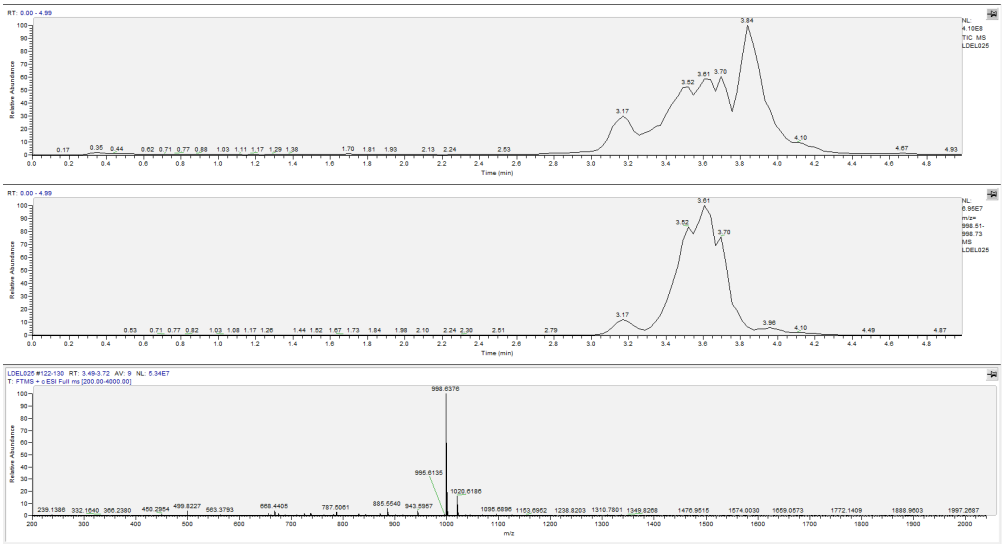
Leu-DLeu-DLeu-Pro-DLeu-DLeuMe-Pro

Exact Mass: 997.6324

ELSD of **LDEL025**

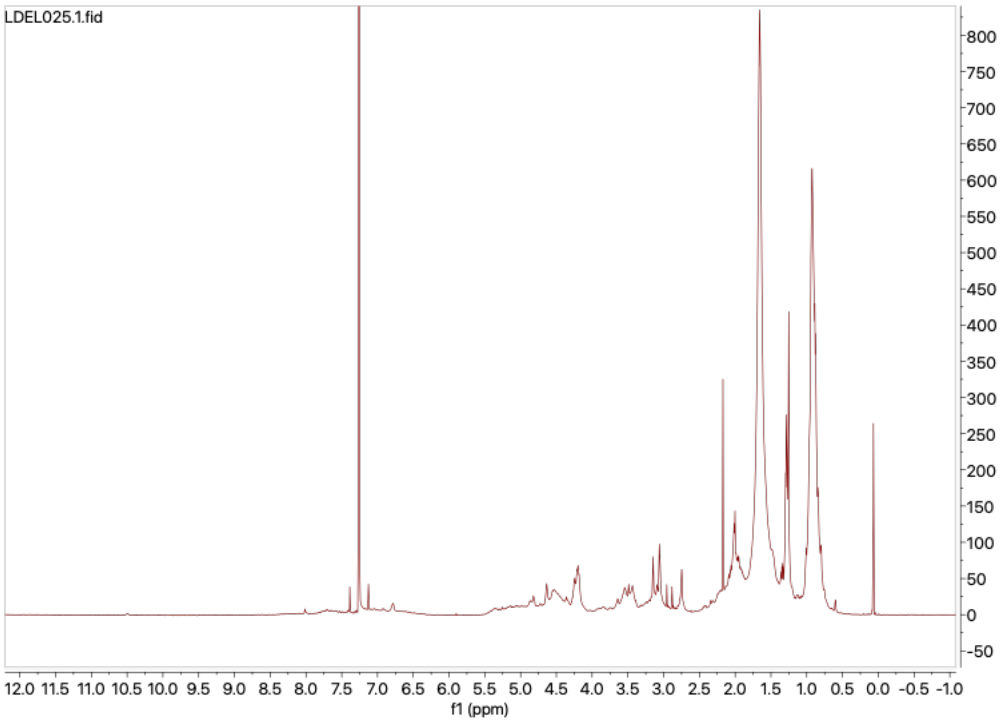


HRMS of **LDEL025**

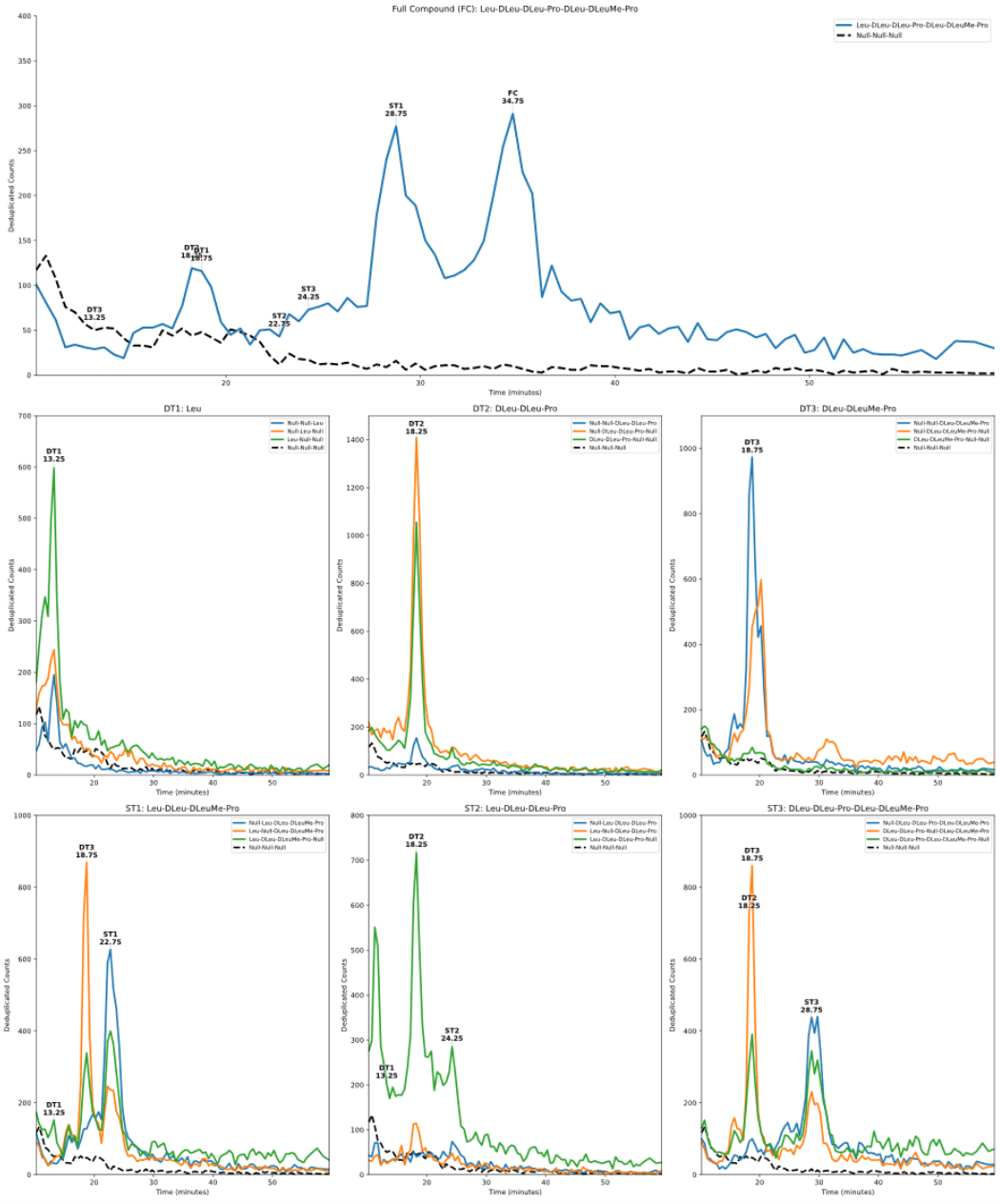


HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 998.6358; found 998.6376

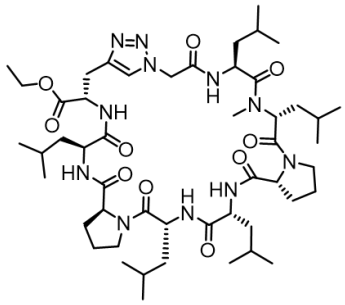
^{1}HNMR of **LDEL025**



Classic Chrome Sequence Count Chromatogram of **LDEL025**



LDEL027

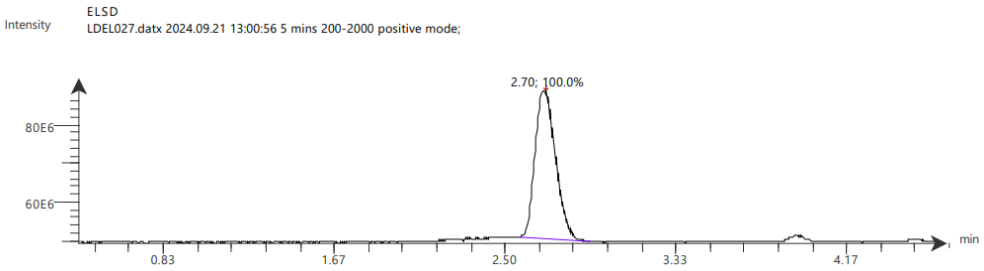


LDEL027

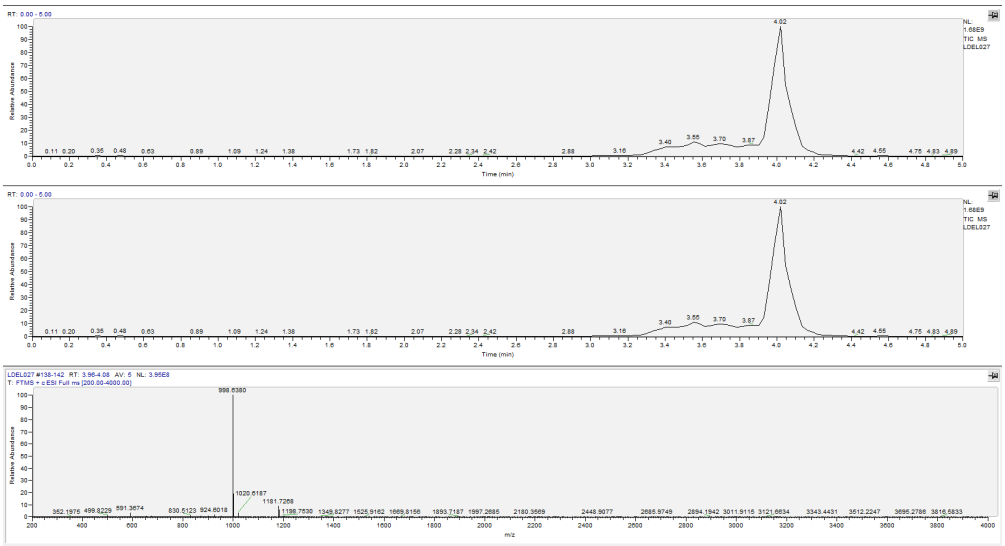
Leu-DLeuMe-DPro-DLeu-DLeu-Pro-Leu

Exact Mass: 997.6324

ELSD of **LDEL027**

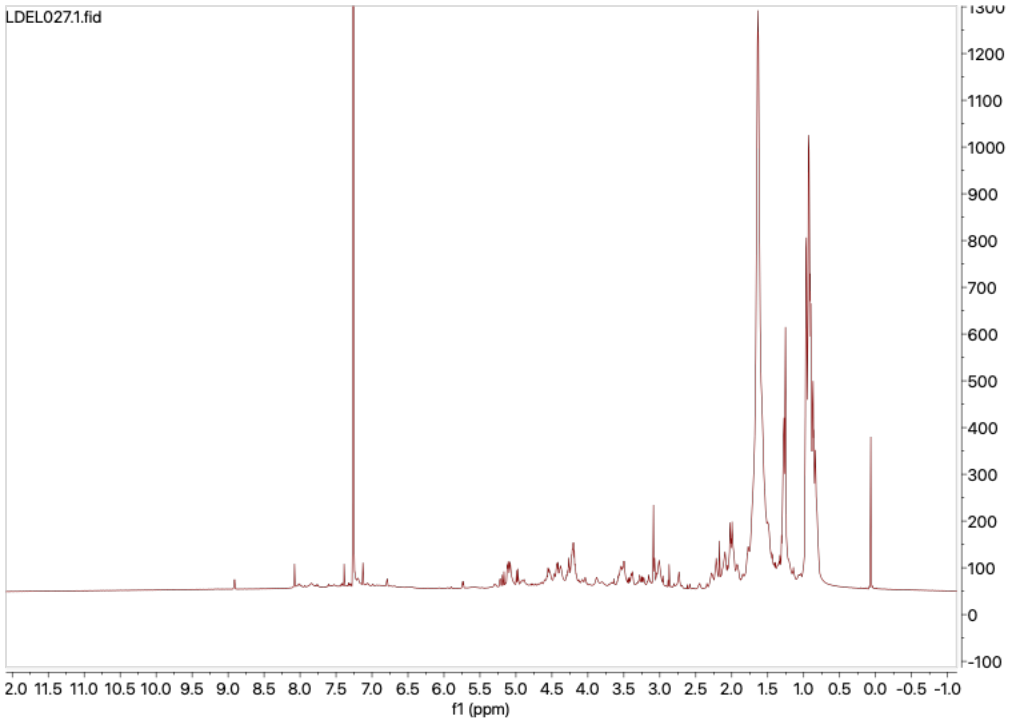


HRMS of **LDEL027**

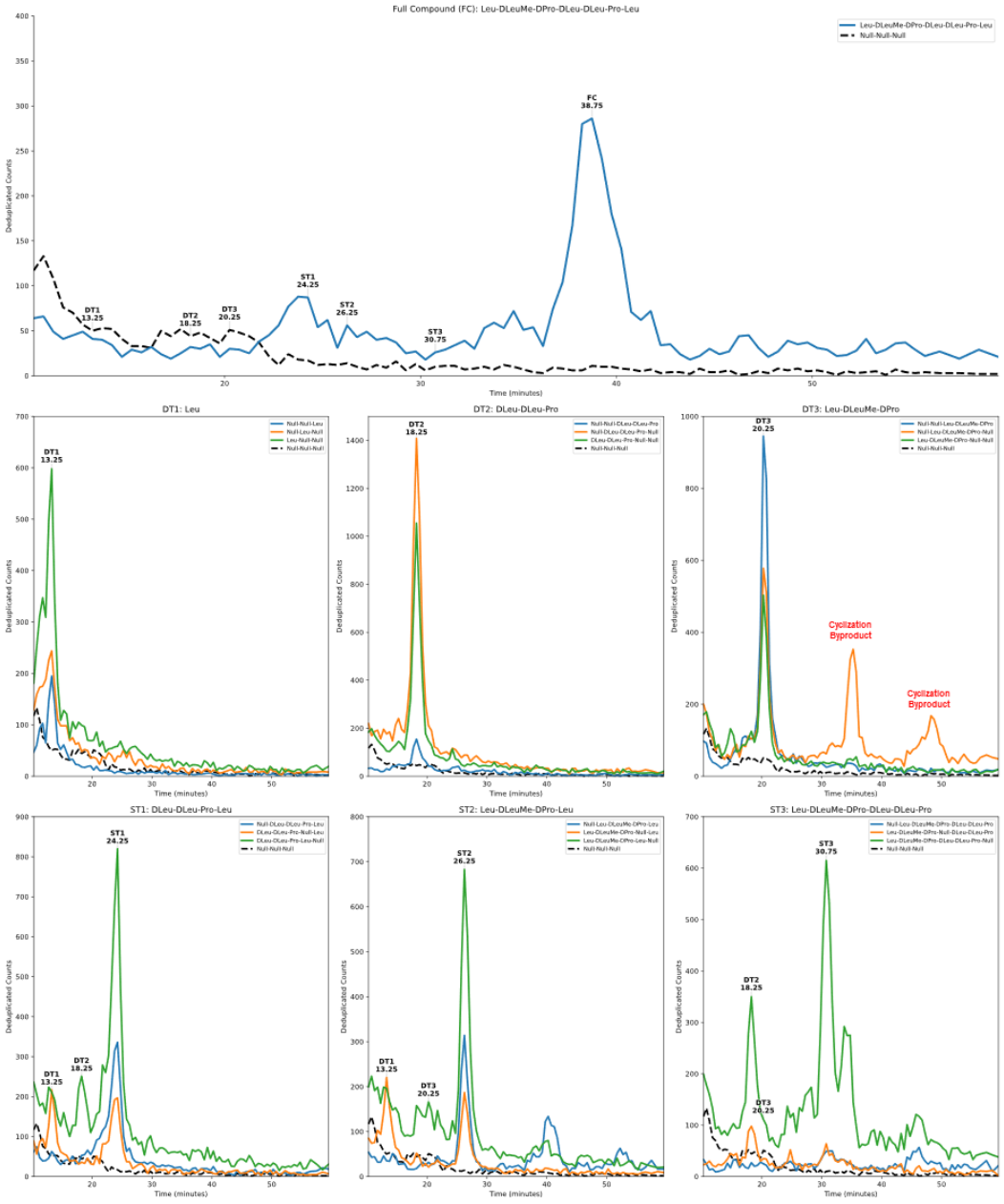


HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 998.6358; found 998.6380

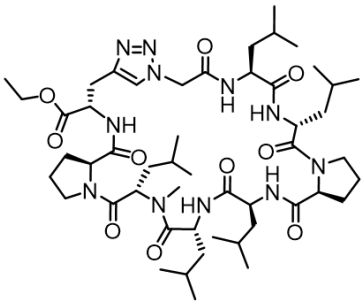
^{1}HNMR of **LDEL027**



Classic Chrome Sequence Count Chromatogram of **LDEL027**



LDEL028

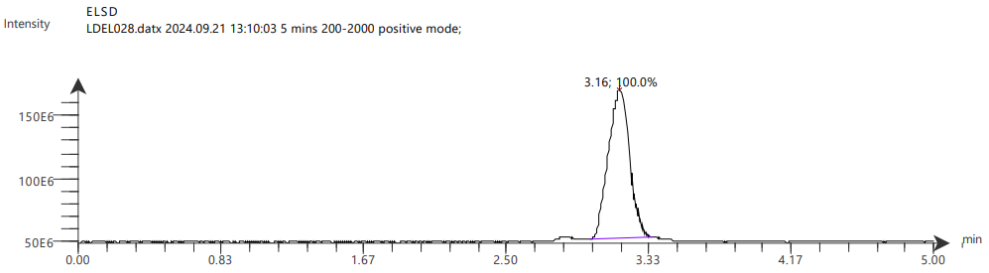


LDEL028

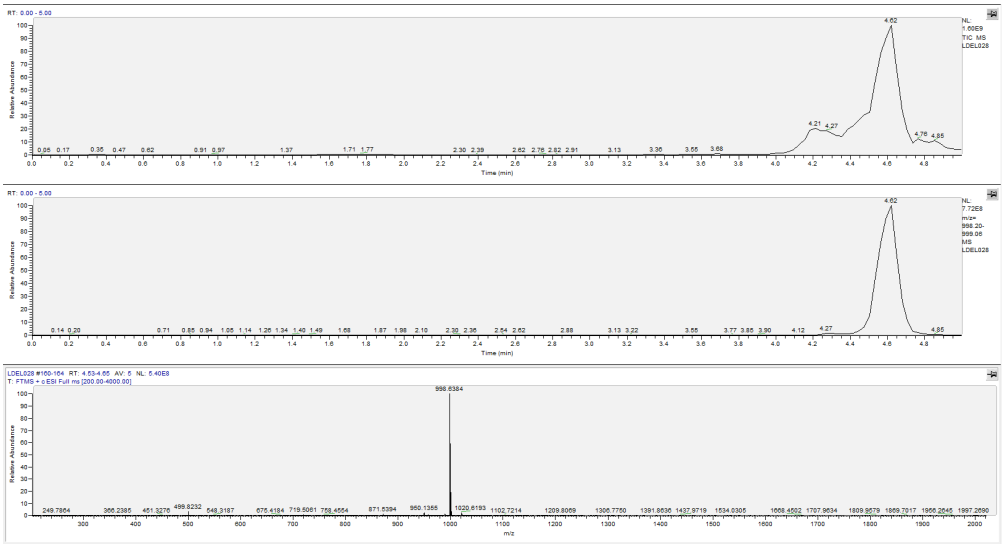
Leu-DLeu-Pro-Leu-DLeu-LeuMe-Pro

Exact Mass: 997.6324

ELSD of **LDEL028**

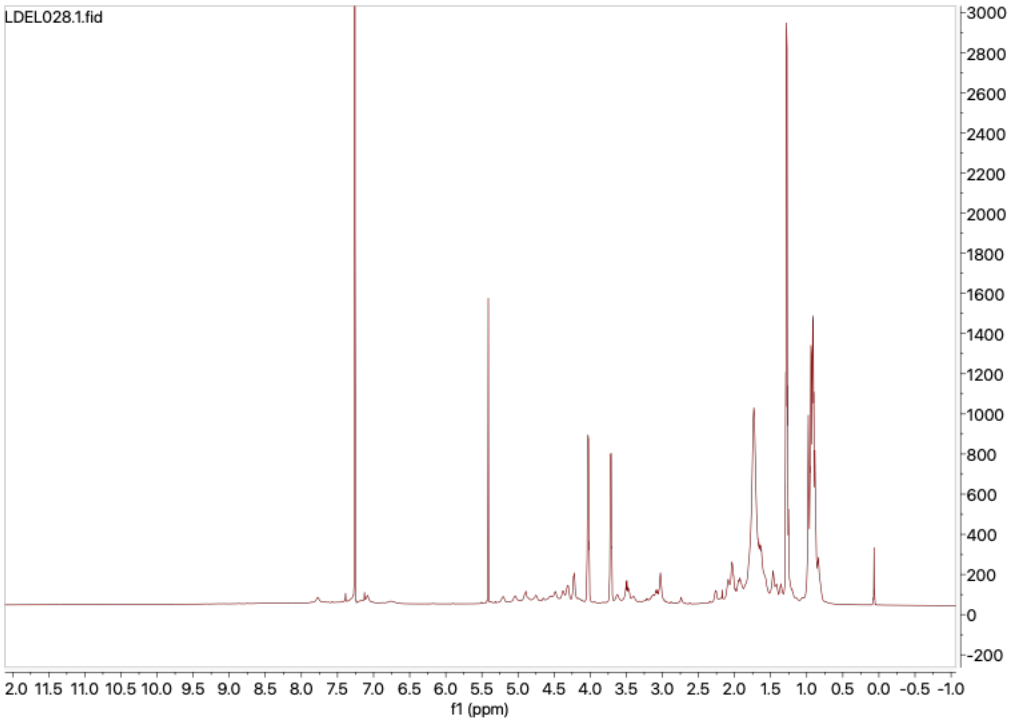


HRMS of **LDEL028**

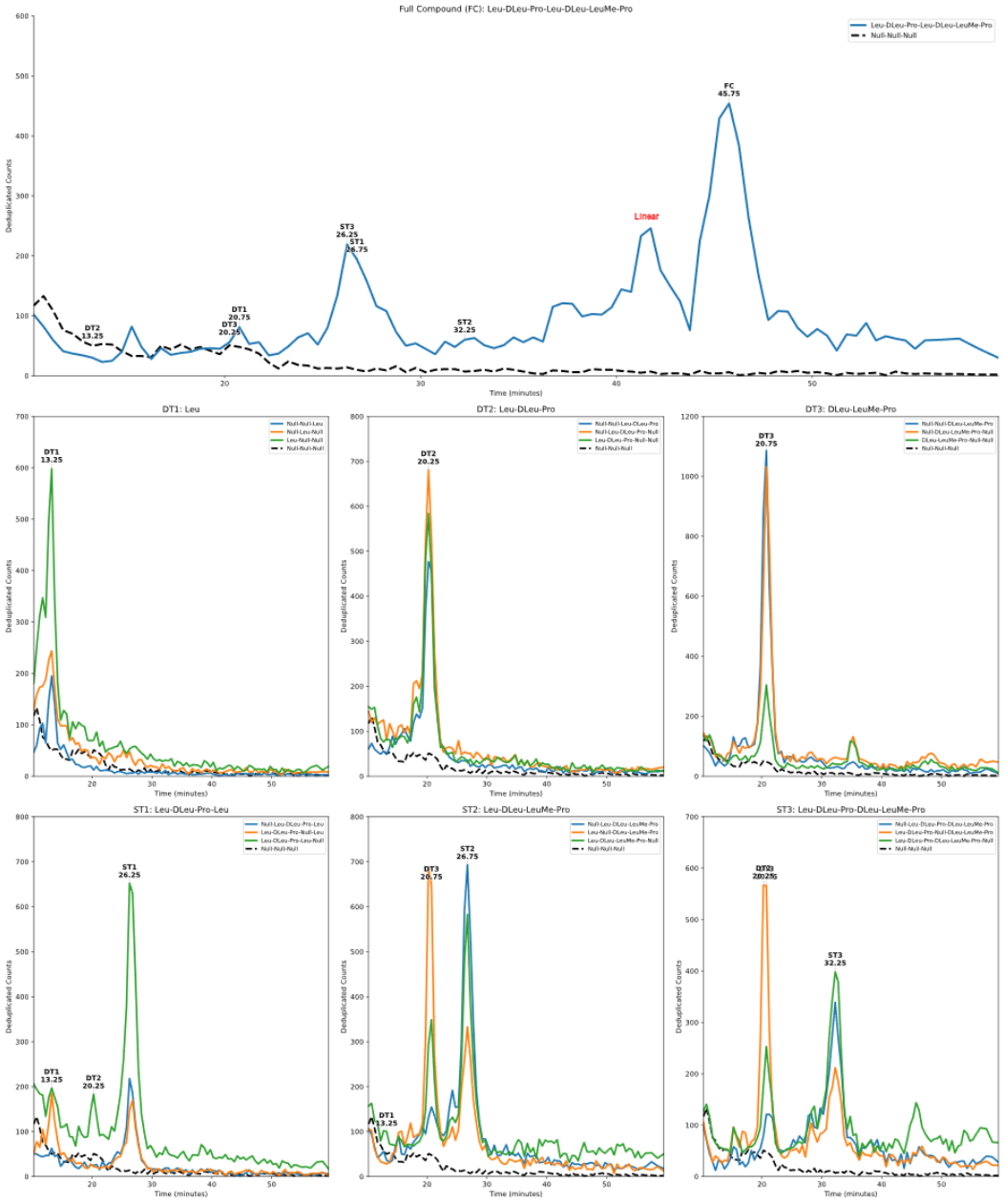


HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 988.6358; found 988.6384

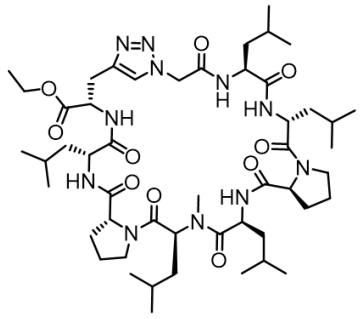
^{1}HNMR of **LDEL028**



Classic Chrome Sequence Count Chromatogram of **LDEL028**



LDEL029

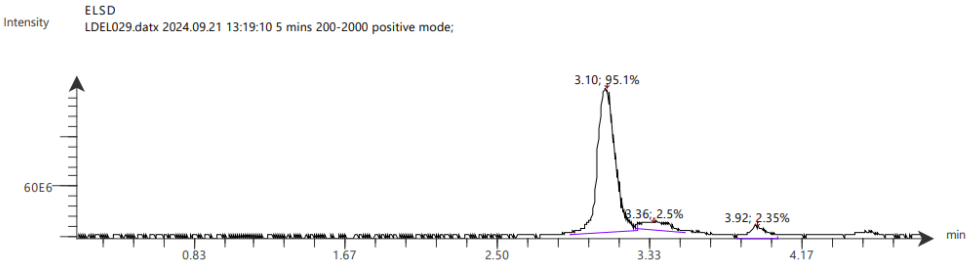


LDEL029

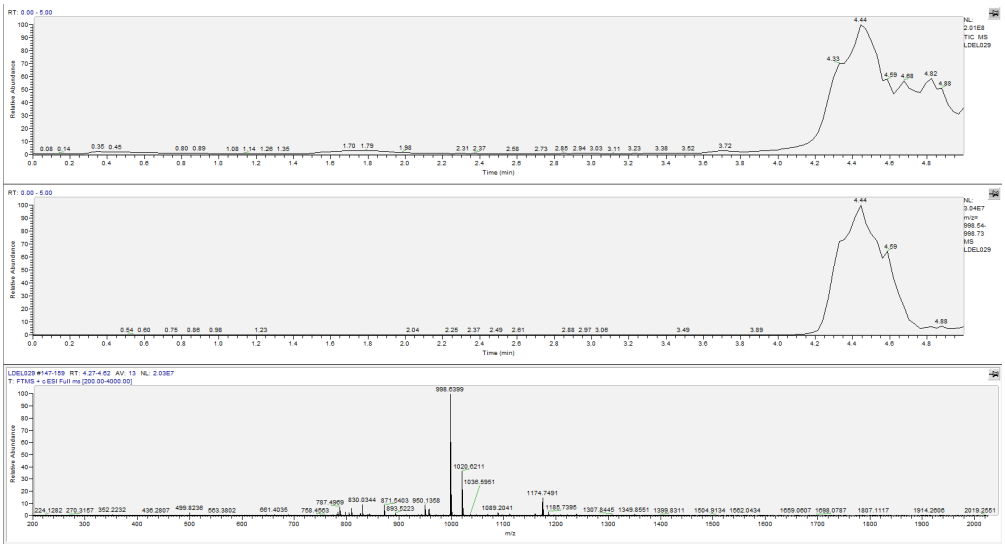
Leu-DLeu-Pro-Leu-LeuMe-DPro-DLeu

Exact Mass: 997.6324

ELSD of **LDEL029**

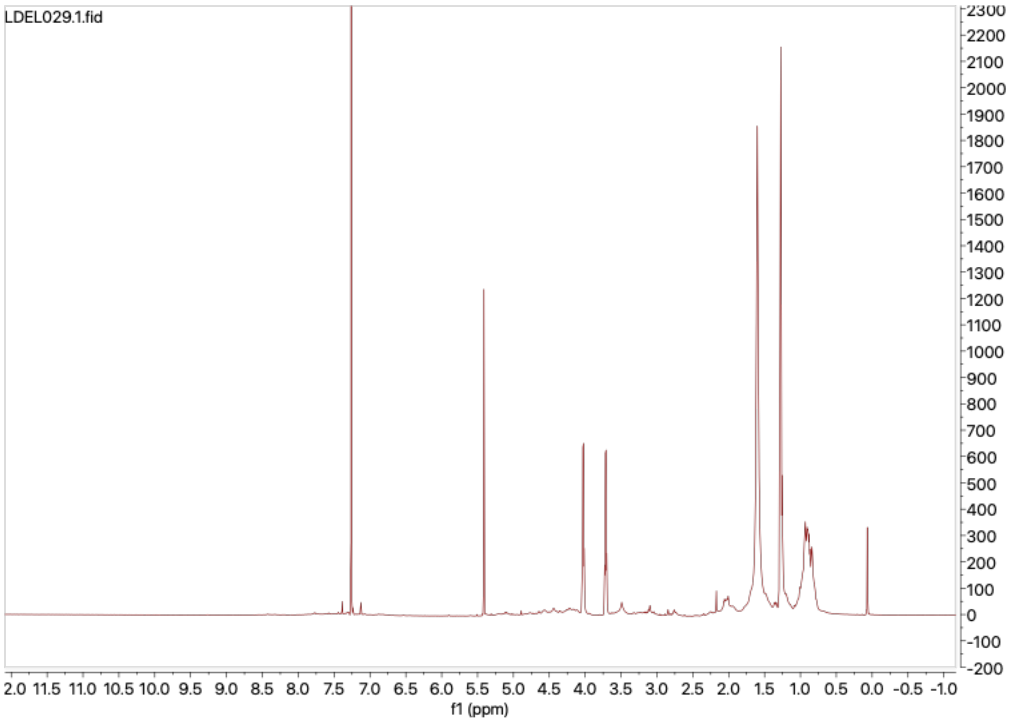


HRMS of **LDEL029**

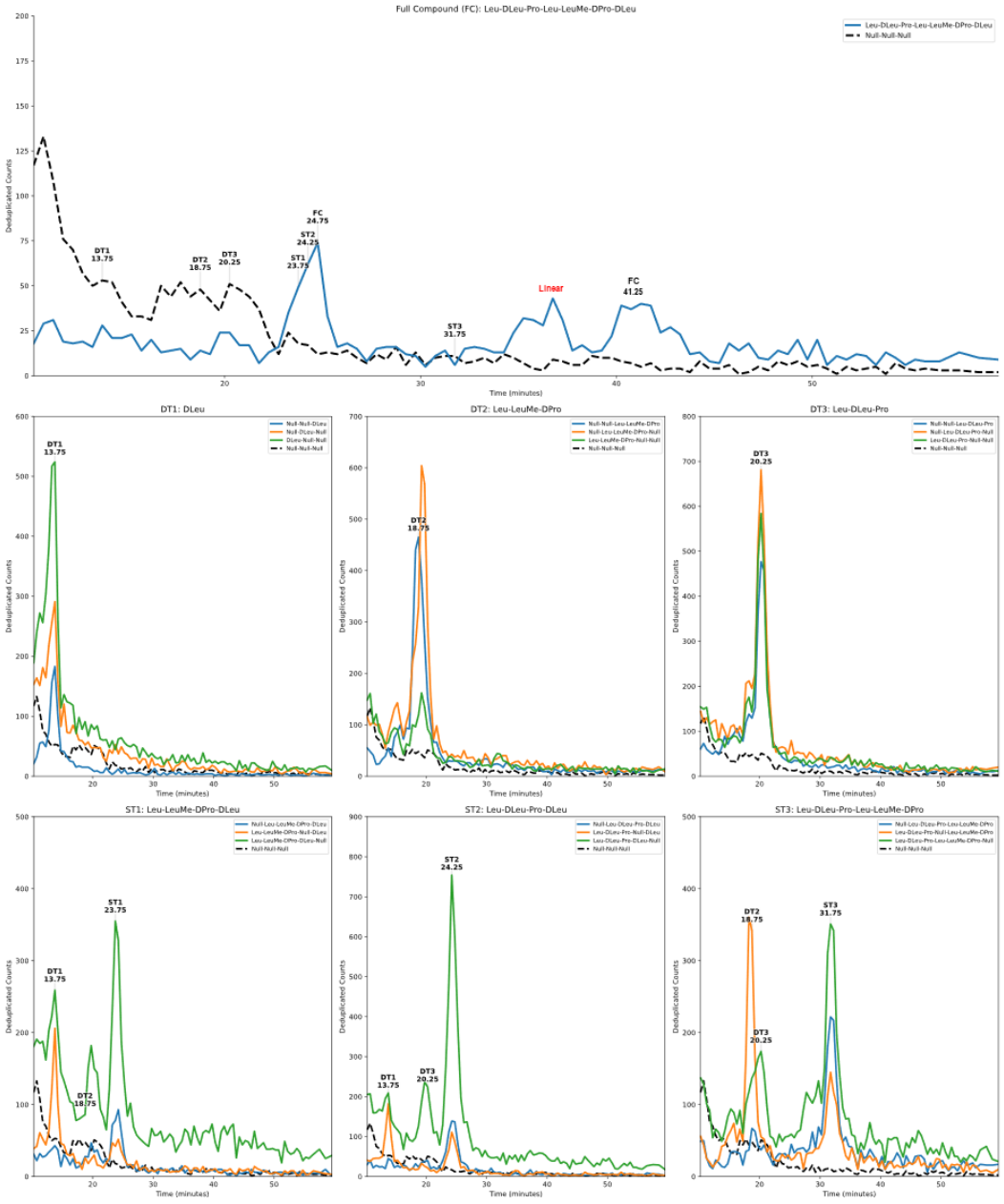


HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 998.6358; found 998.6399

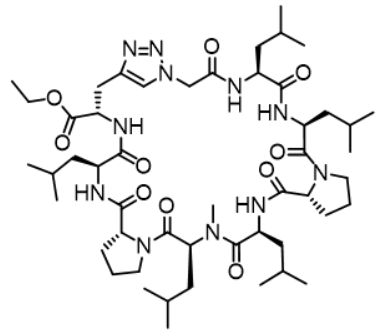
^{1}HNMR of **LDEL029**



Classic Chrome Sequence Count Chromatogram of **LDEL029**



LDEL030

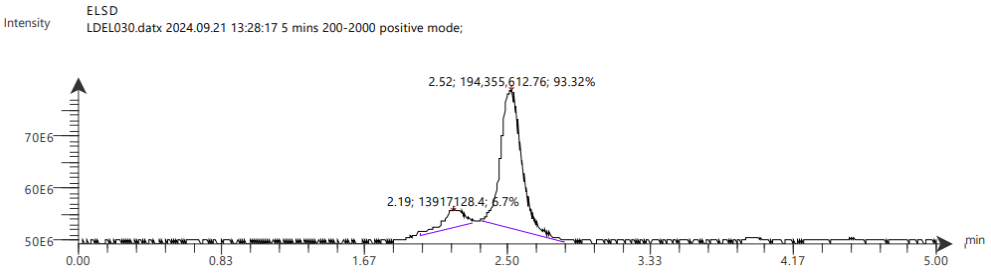


LDEL030

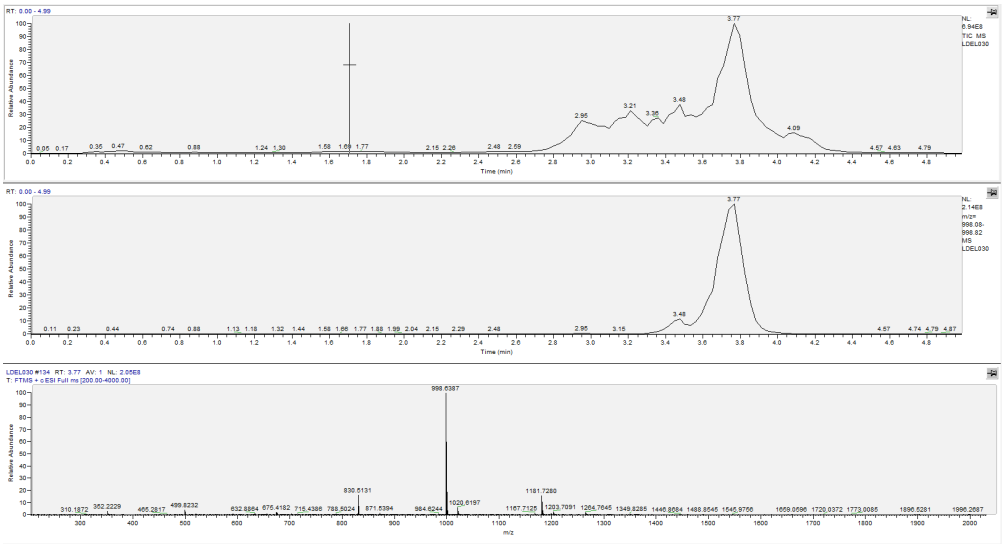
Leu-Leu-DPro-Leu-LeuMe-DPro-Leu

Exact Mass: 997.6324

ELSD of **LDEL030**

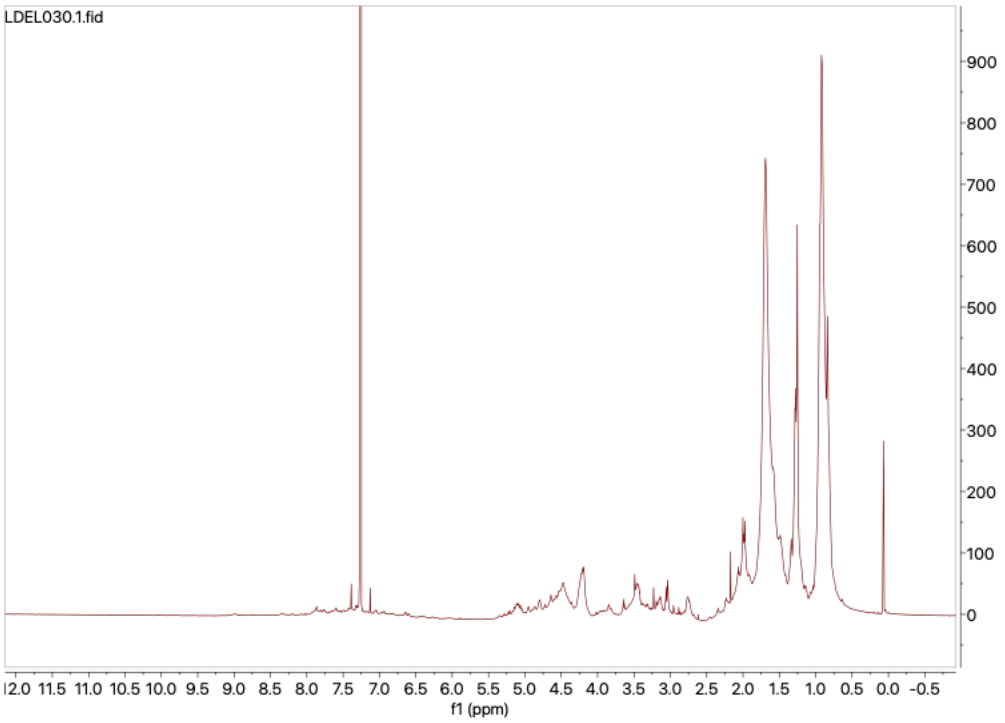


HRMS of **LDEL030**

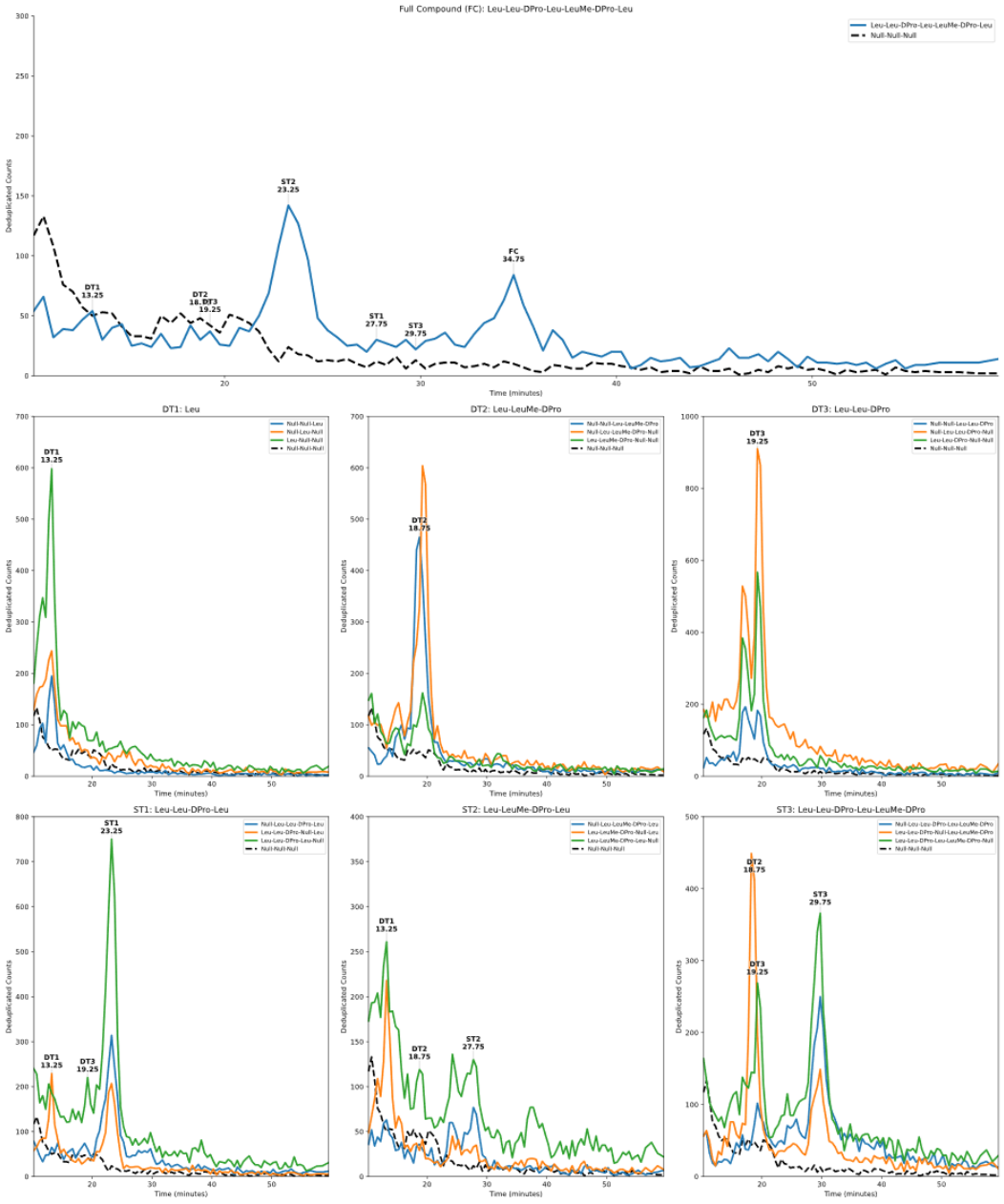


HRMS (ESI) m/z calculated for $C_{50}H_{83}N_{11}O_{10}^{+}$ ($M+H^{+}$) 998.6358; found 998.6387

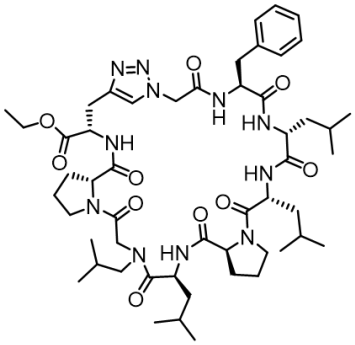
^{1}HNMR of **LDEL030**



Classic Chrome Sequence Count Chromatogram of **LDEL030**



LDEL032

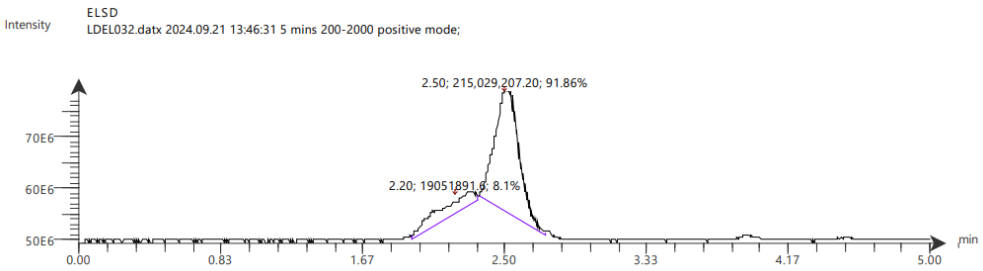


LDEL032

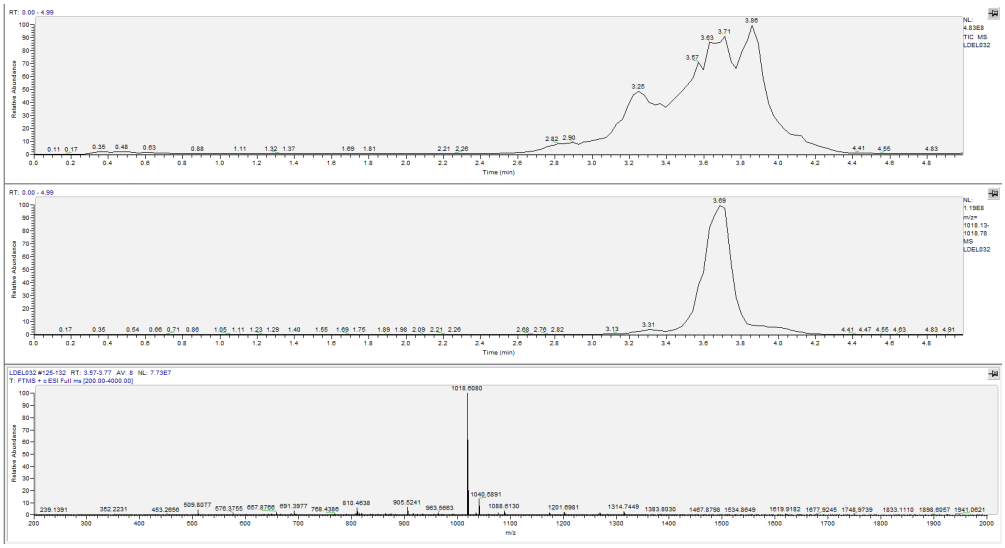
Phe-DLeu-DLeu-Pro-Leu-LA03-DPro

Exact Mass: 1017.6011

ELSD of **LDEL032**

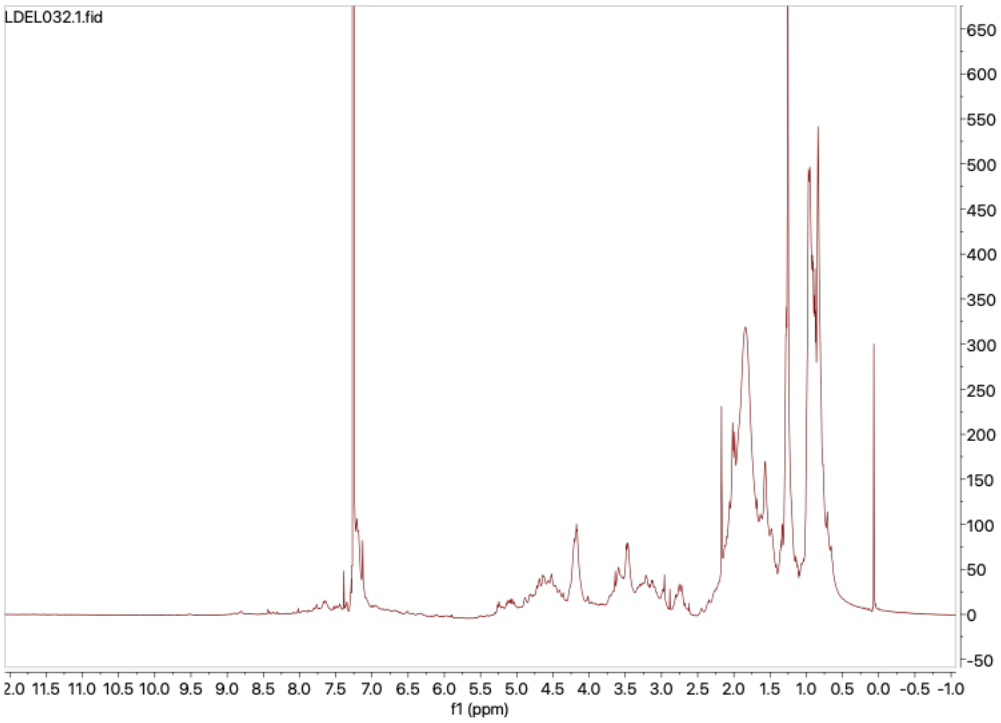


HRMS of **LDEL032**

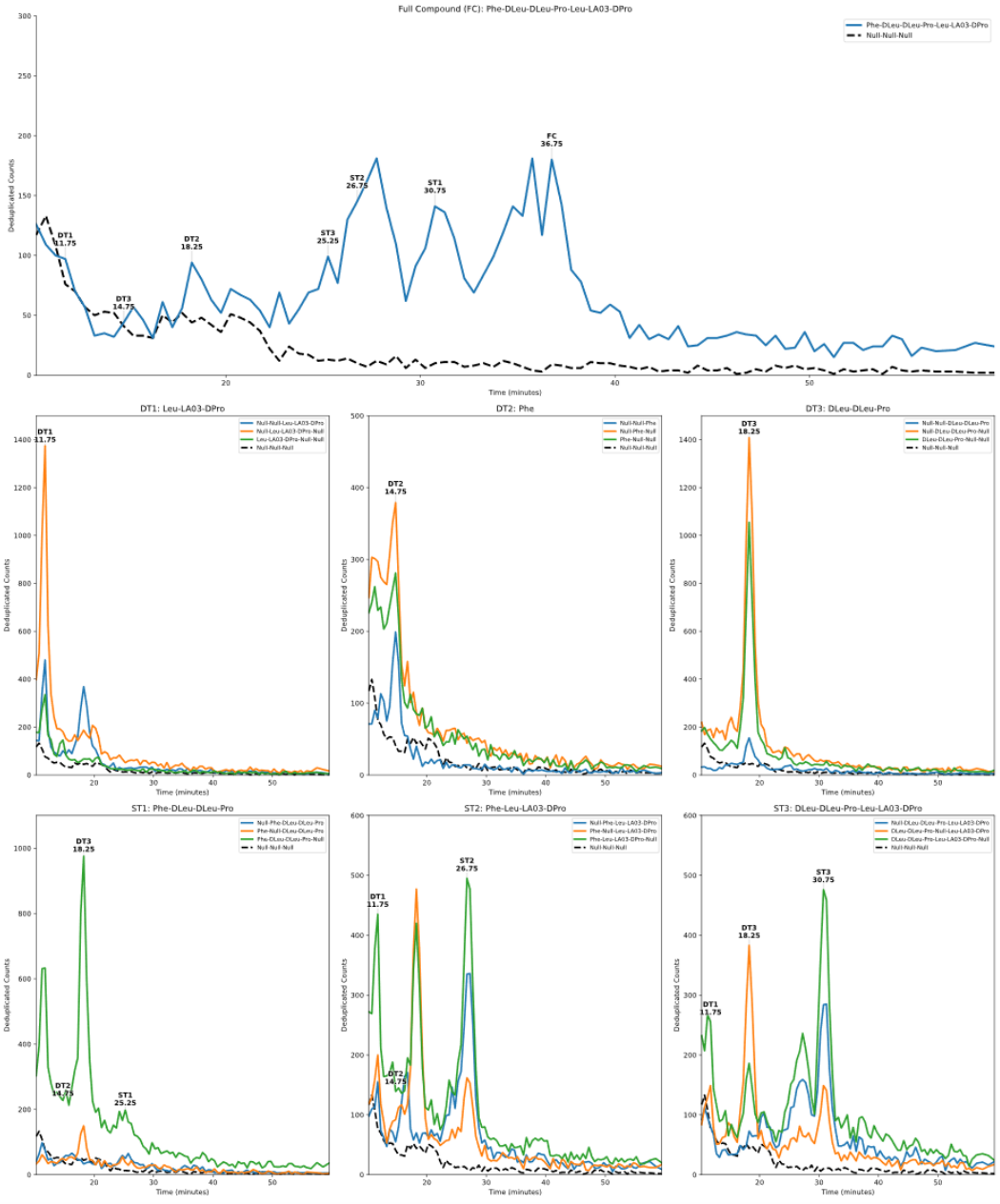


HRMS (ESI) m/z calculated for $C_{52}H_{79}N_{11}O_{10}^{+}$ ($M+H^{+}$) 1018.6045; found 1018.6080

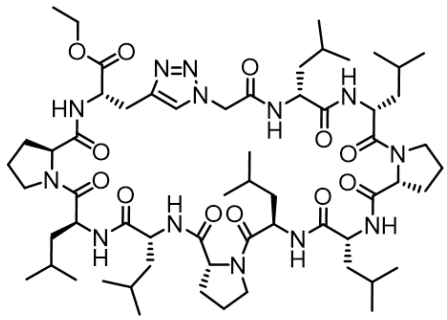
^{1}HNMR of **LDEL032**



Classic Chrome Sequence Count Chromatogram of **LDEL032**



LDEL034

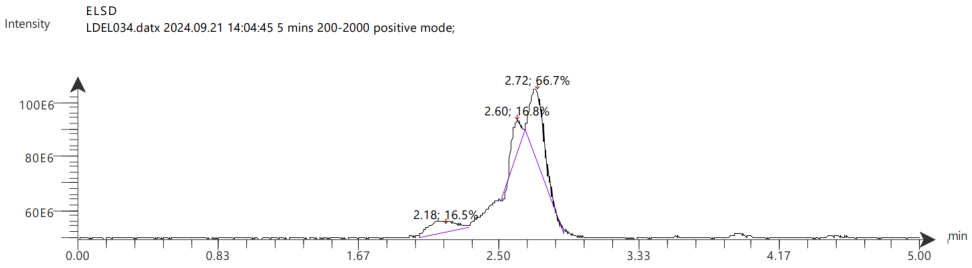


LDEL034

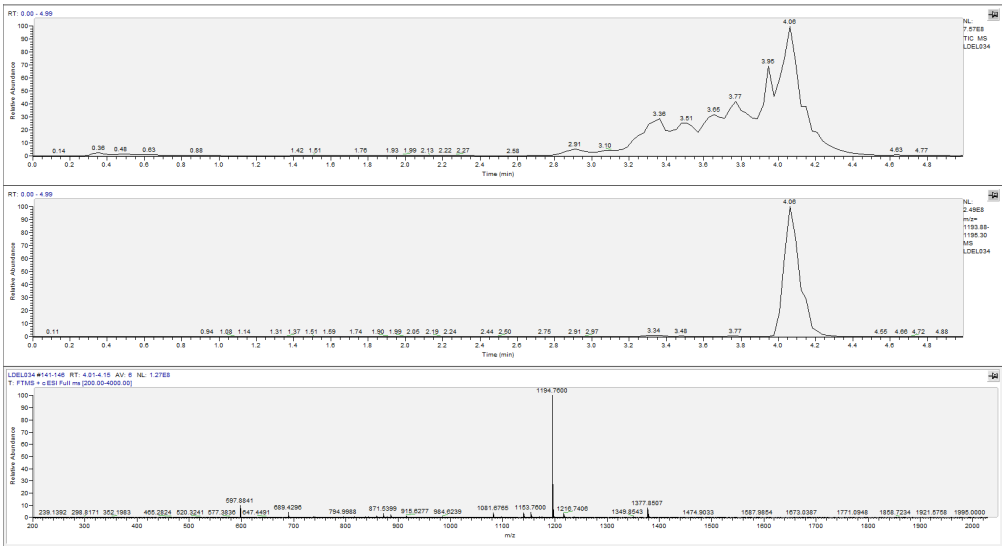
DLeu-DLeu-DPro-DLeu-DLeu-DPro-DLeu-Leu-Pro

Exact Mass: 1193.7536

ELSD of **LDEL034**

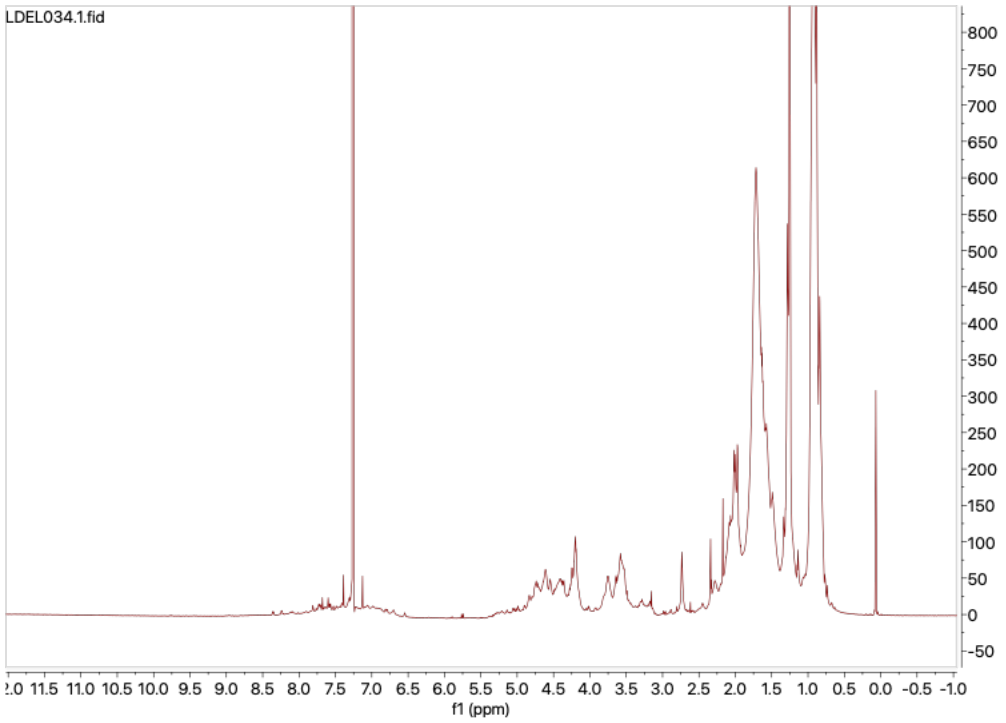


HRMS of **LDEL034**

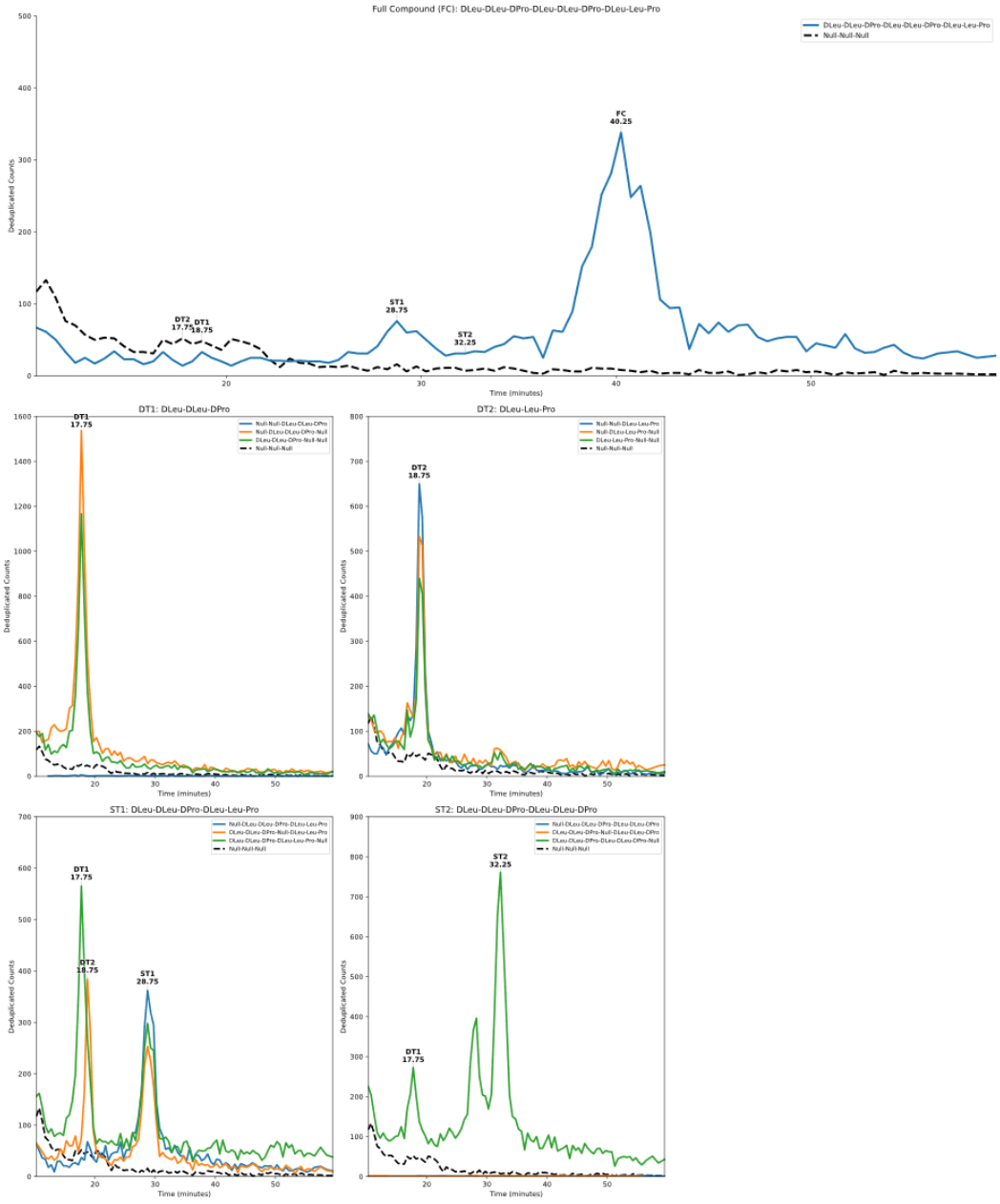


HRMS (ESI) m/z calculated for $C_{60}H_{99}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1194.7570; found 1194.7600

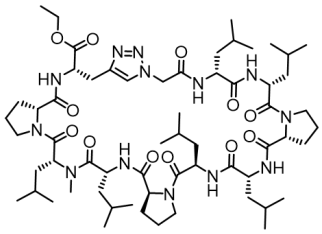
^{1}HNMR of **LDEL034**



Classic Chrome Sequence Count Chromatogram of **LDEL034**



LDEL035

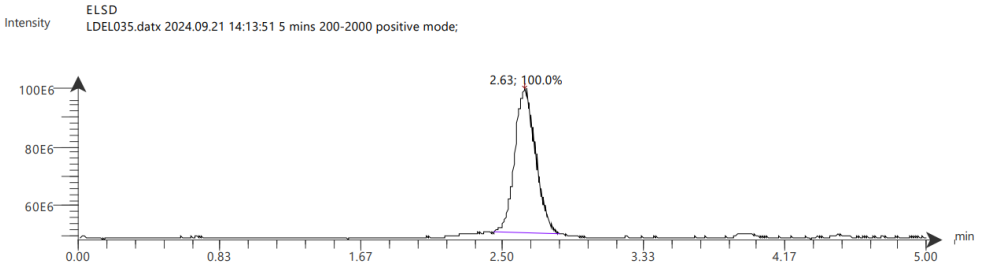


LDEL035

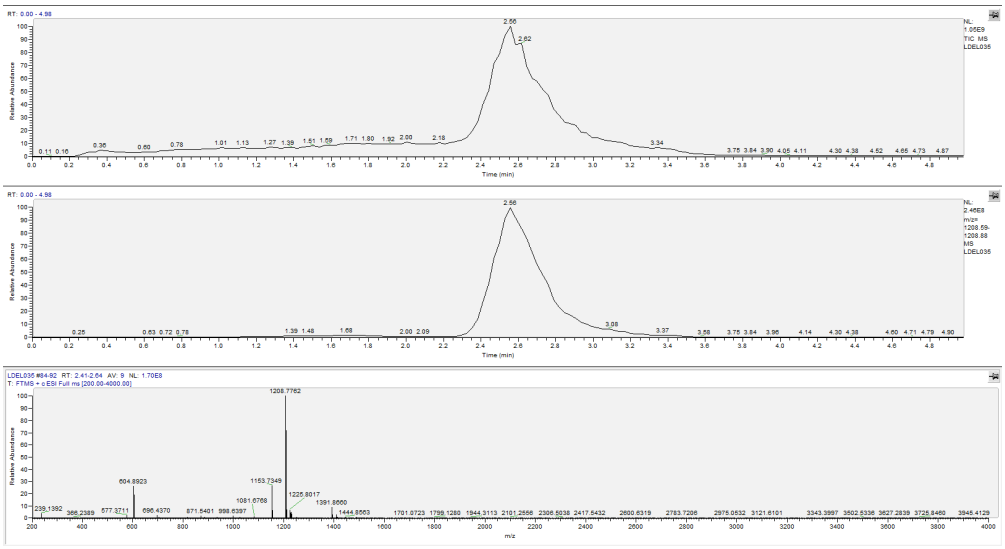
DLeu-DLeu-DPro-DLeu-DLeu-Pro-DLeu-DLeuMe-DPro

Exact Mass: 1207.7693

ELSD of **LDEL035**

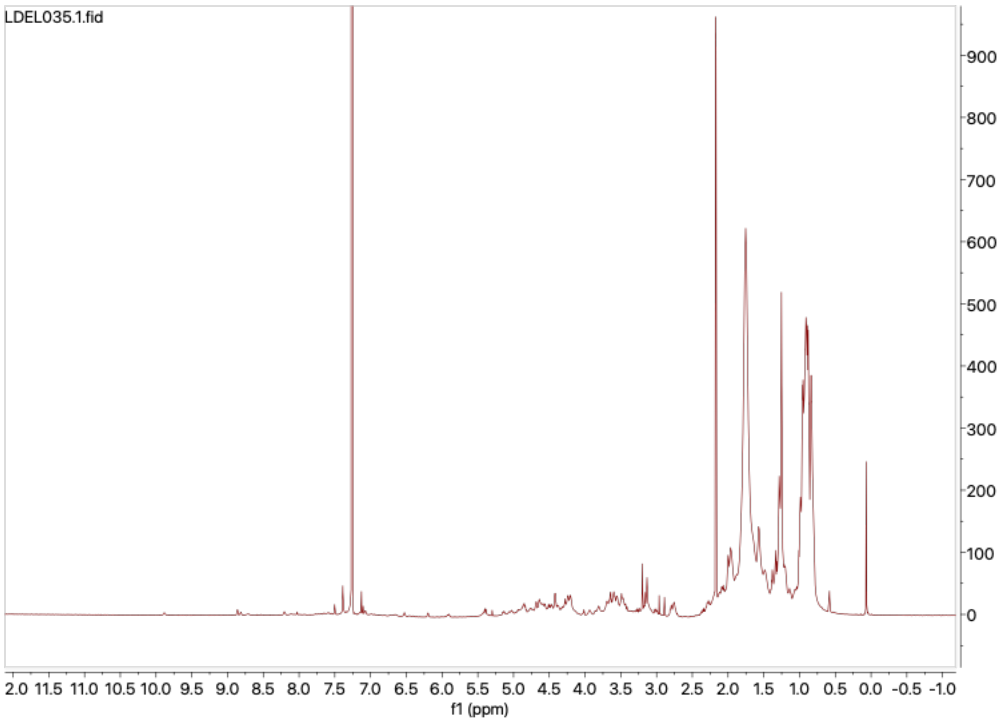


HRMS of **LDEL035**

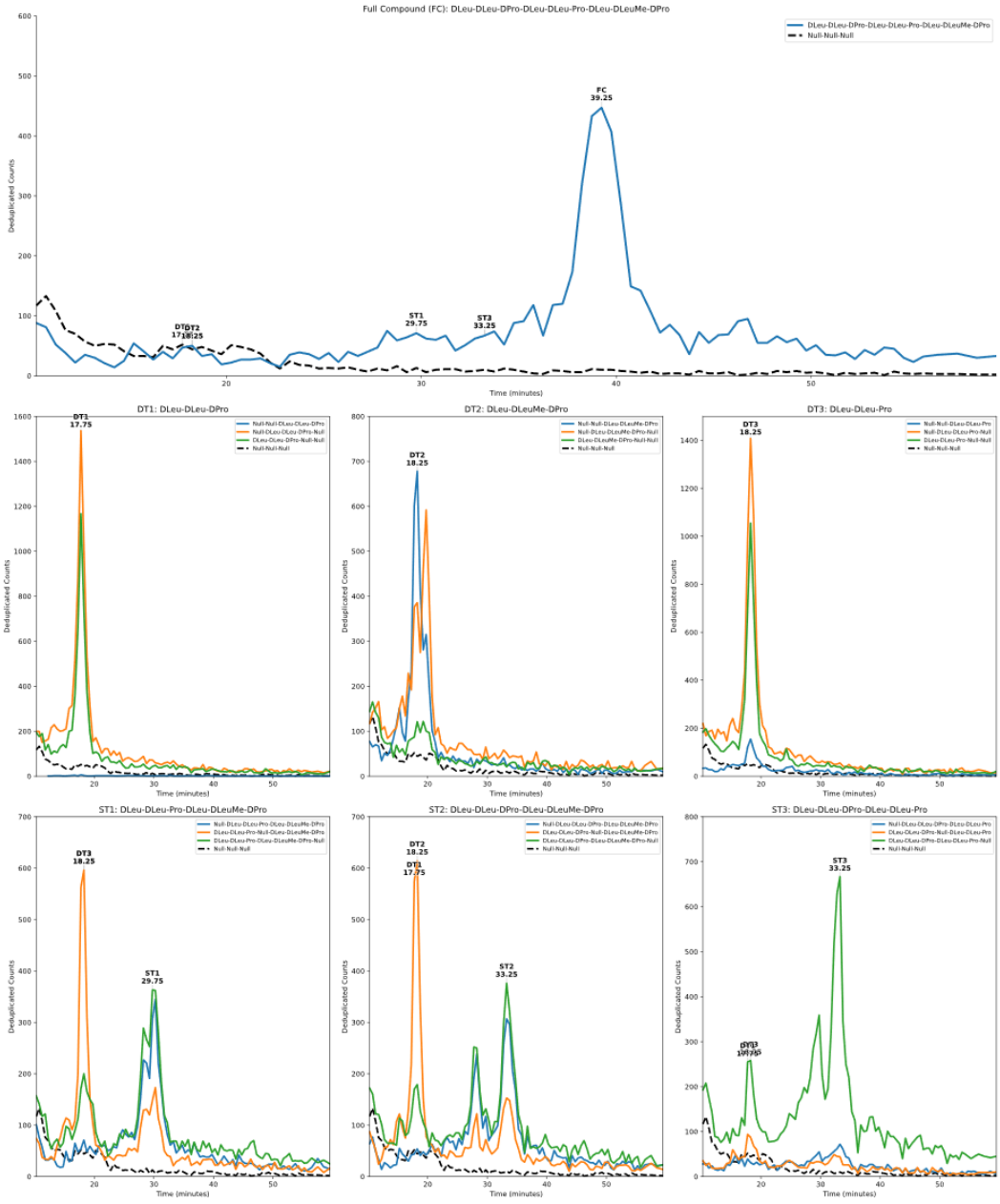


HRMS (ESI) m/z calculated for $C_{61}H_{101}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1208.7726; found 1208.7762

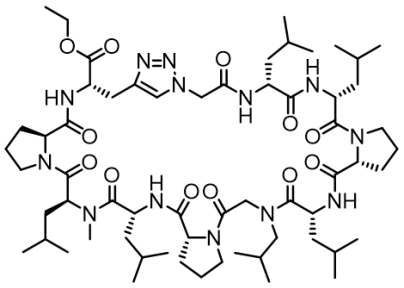
HNMR of **LDEL035**



Classic Chrome Sequence Count Chromatogram of **LDEL035**



LDEL037

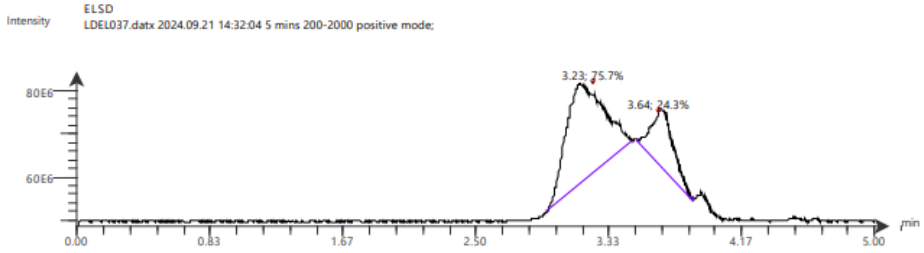


LDEL037

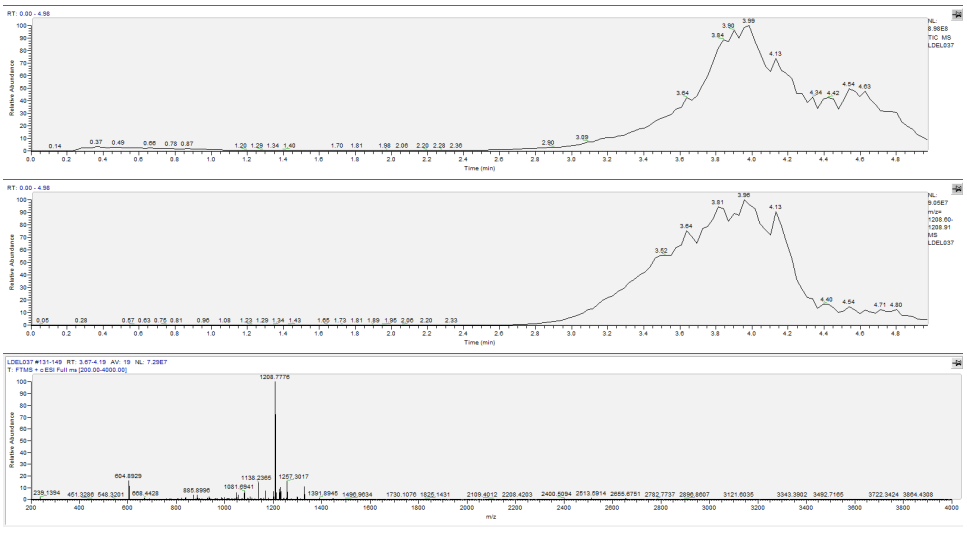
DLeu-DLeu-DPro-DLeu-LA03-DPro-DLeu-LeuMe-Pro

Exact Mass: 1207.7693

ELSD of **LDEL037**

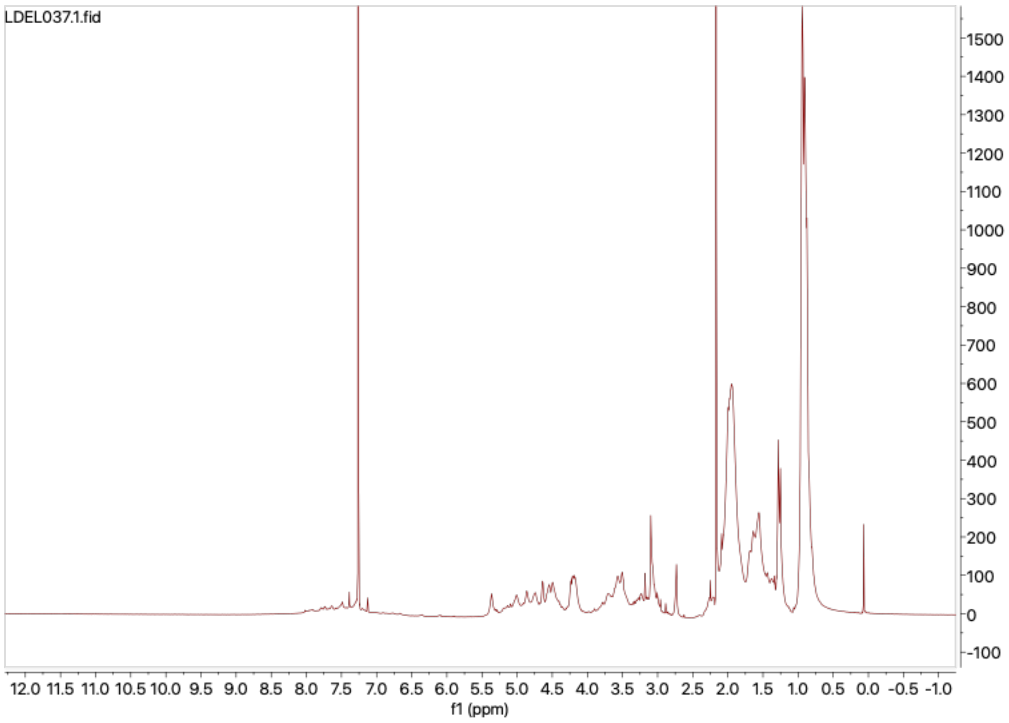


HRMS of **LDEL037**

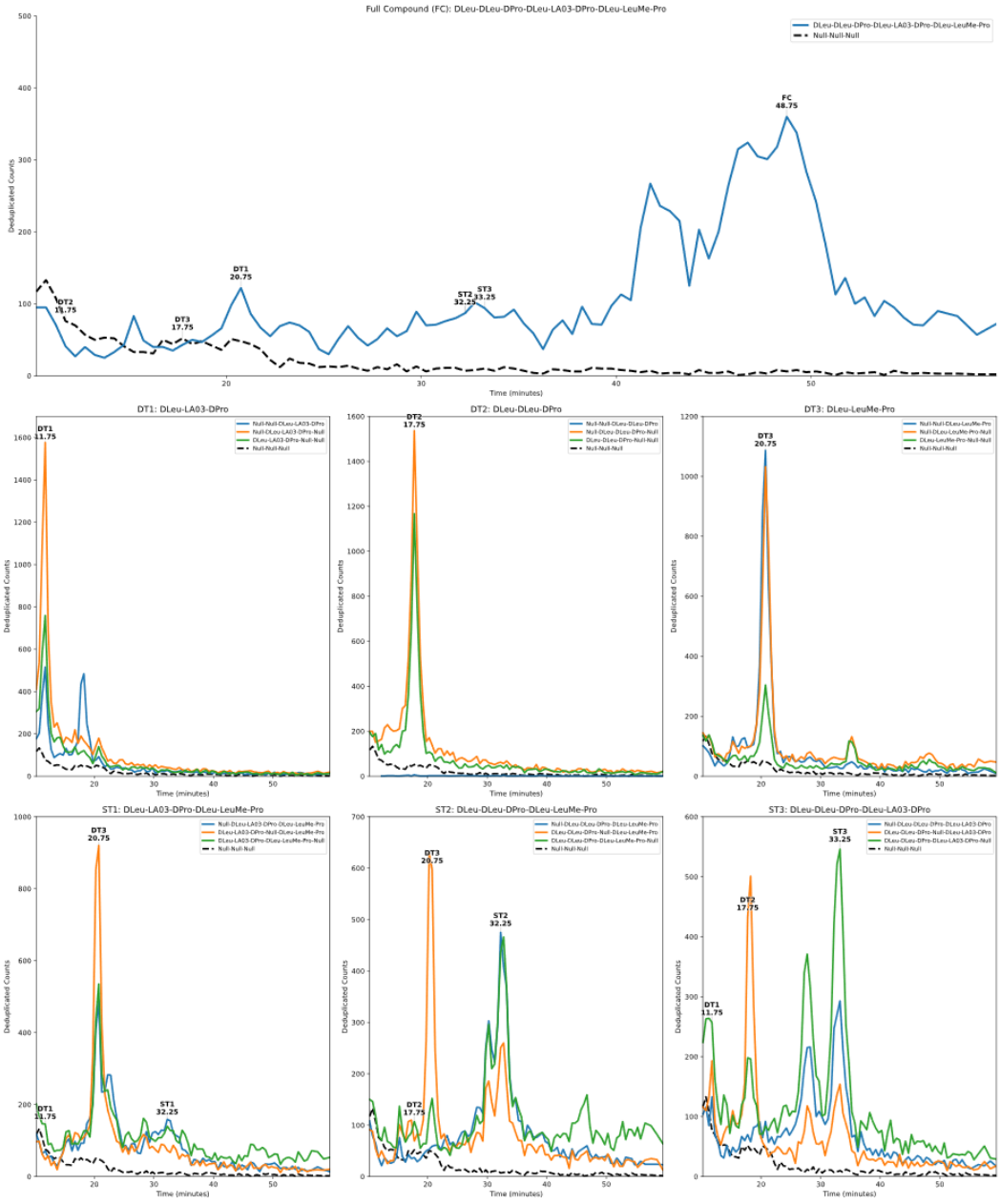


HRMS (ESI) m/z calculated for $C_{61}H_{101}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1208.7726; found 1208.7776

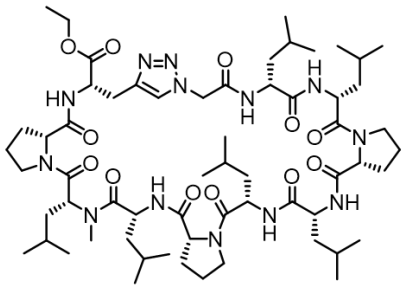
^{1}HNMR of **LDEL037**



Classic Chrome Sequence Count Chromatogram of **LDEL037**



LDEL038

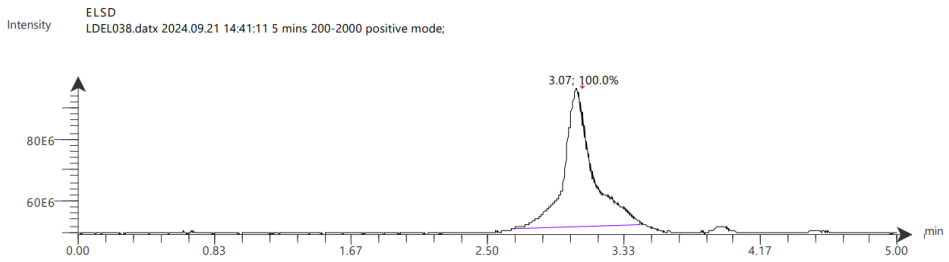


LDEL038

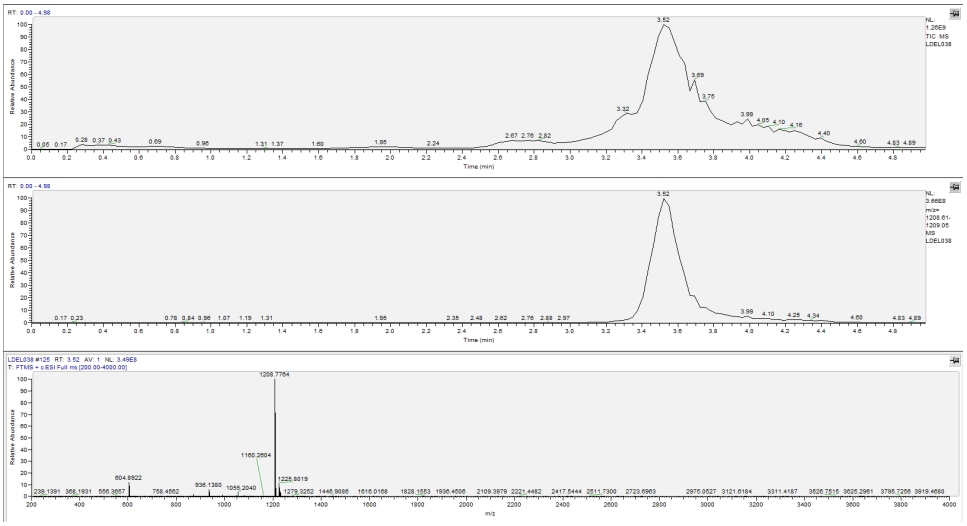
DLeu-DLeu-DPro-DLeu-Leu-DPro-DLeu-DLeuMe-DPro

Exact Mass: 1207.7693

ELSD of **LDEL038**

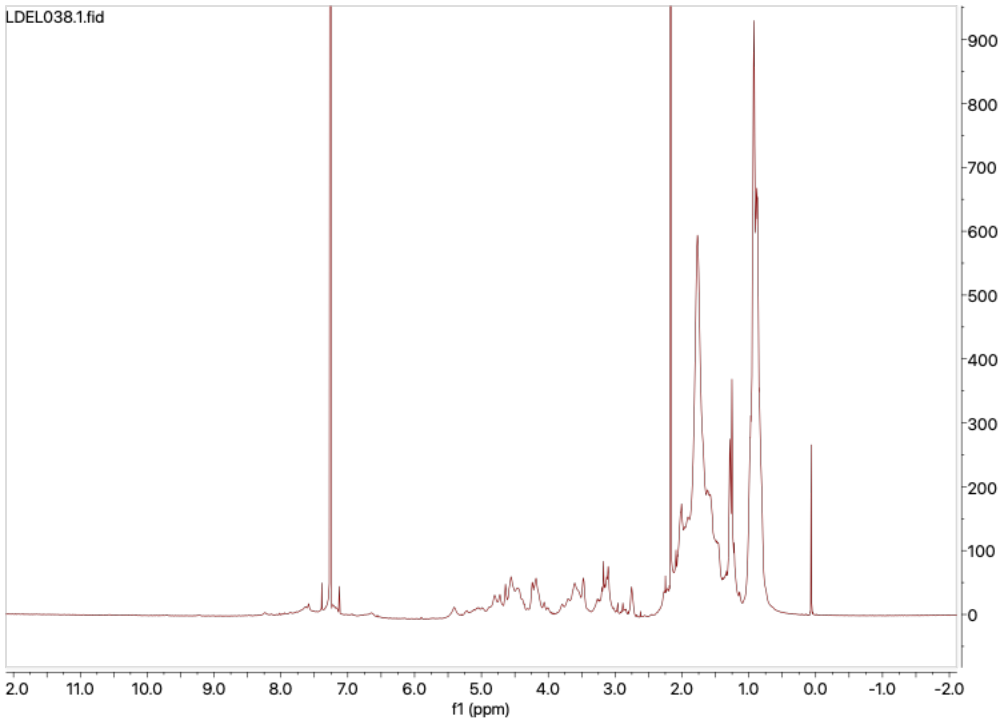


HRMS of **LDEL038**

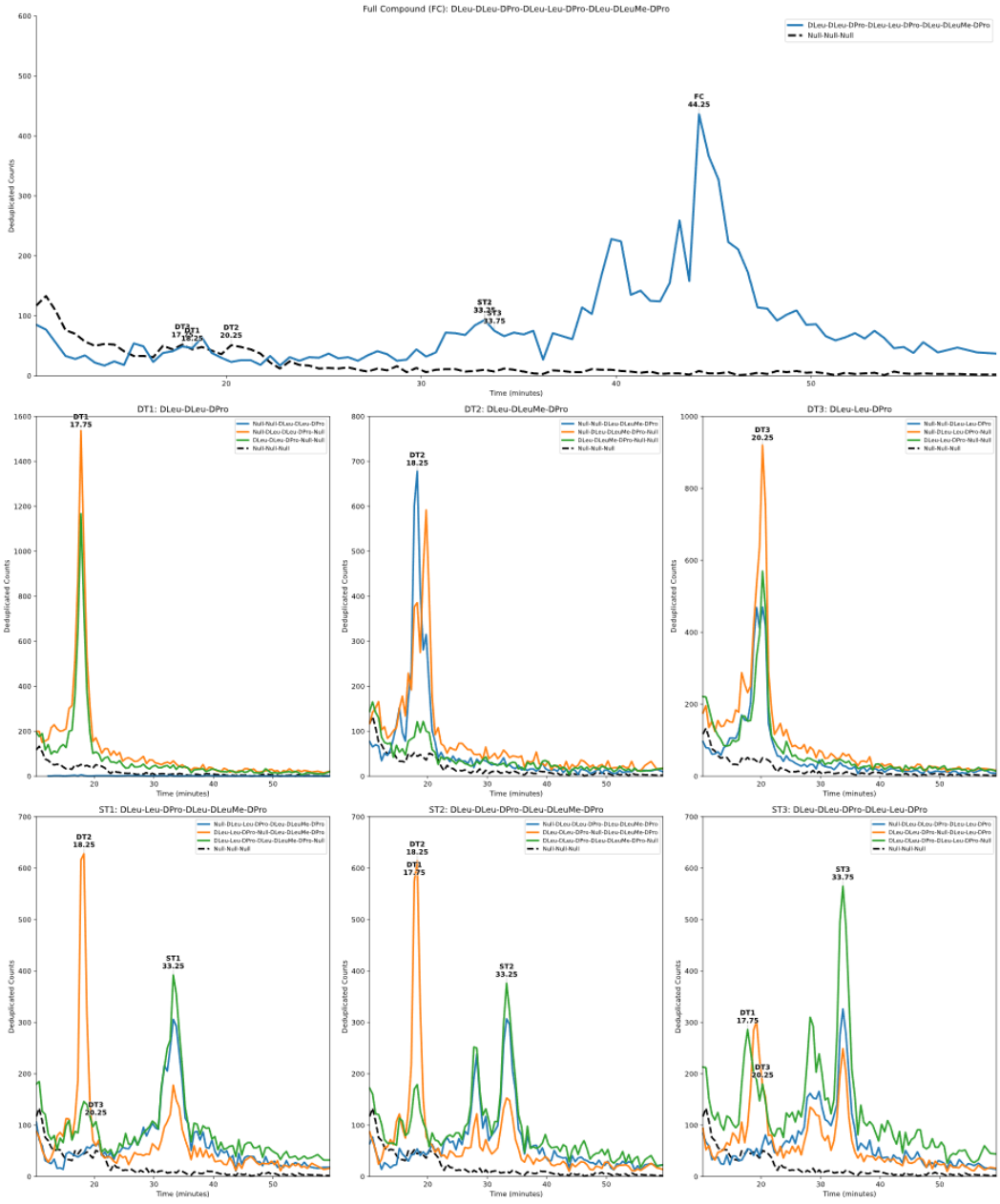


HRMS (ESI) m/z calculated for $C_{61}H_{101}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1208.7726; found 1208.7770

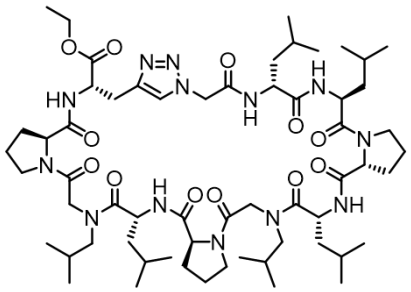
^{1}HNMR of **LDEL038**



Classic Chrome Sequence Count Chromatogram of **LDEL038**



LDEL040

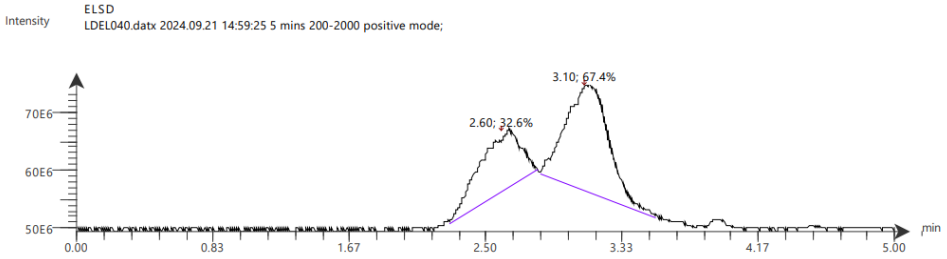


LDEL040

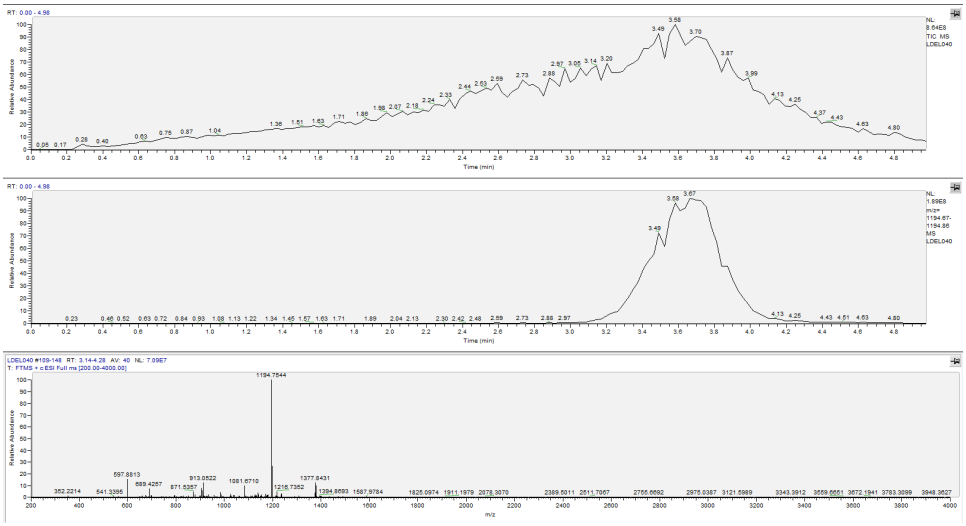
DLeu-Leu-DPro-DLeu-LA03-Pro-DLeu-LA03-Pro

Exact Mass: 1193.7536

ELSD of **LDEL040**

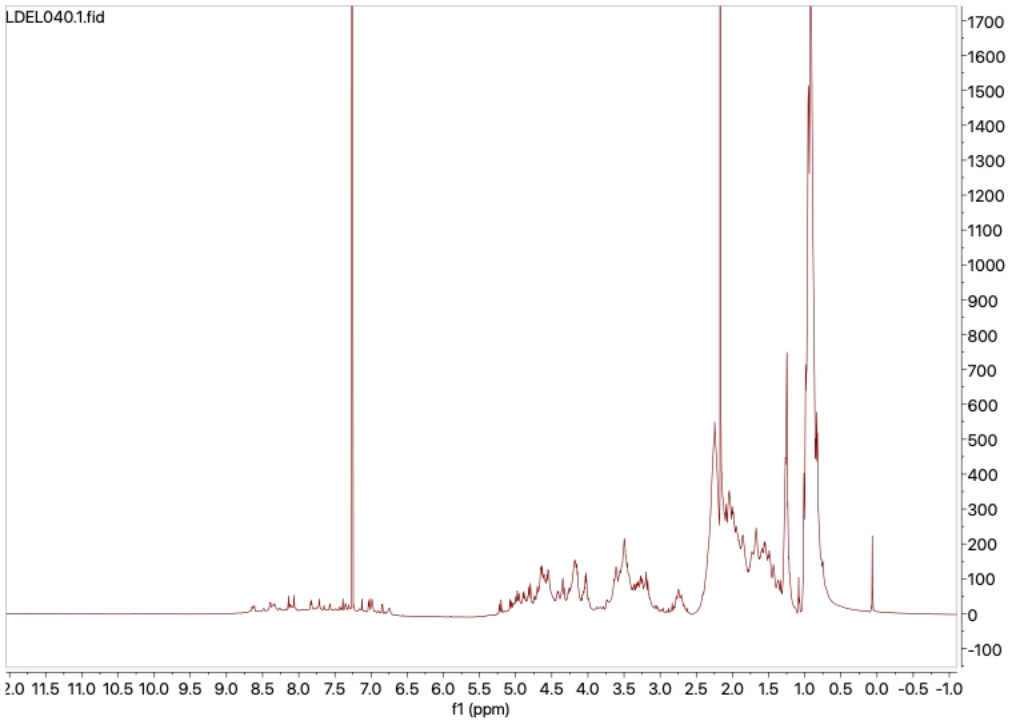


HRMS of **LDEL040**

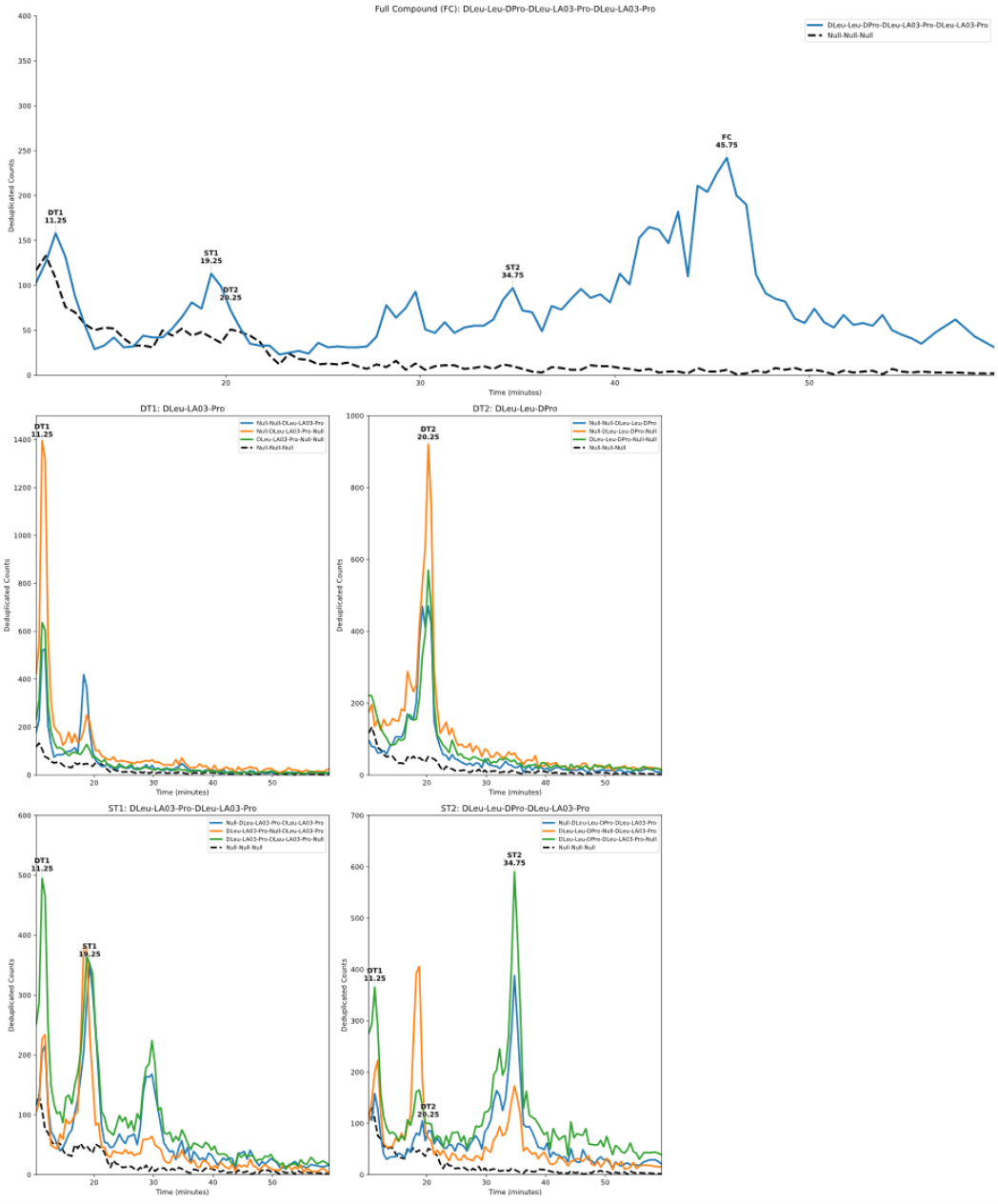


HRMS (ESI) m/z calculated for $C_{60}H_{99}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1194.7570; found 1194.7544

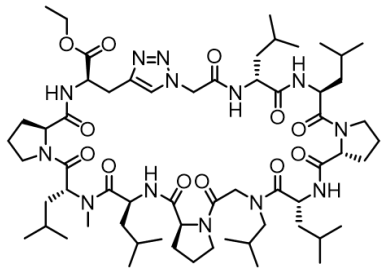
^{1}HNMR of **LDEL040**



Classic Chrome Sequence Count Chromatogram of **LDEL040**



LDEL041

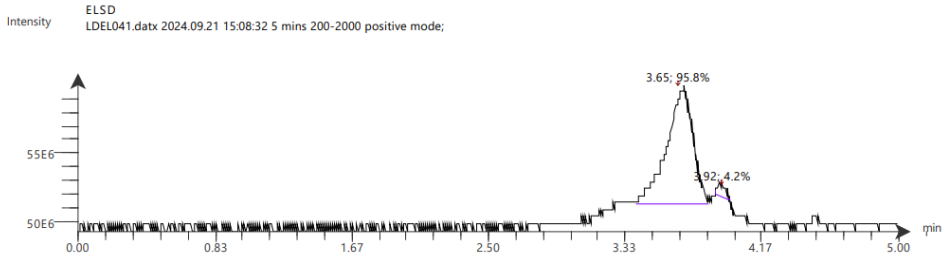


LDEL041

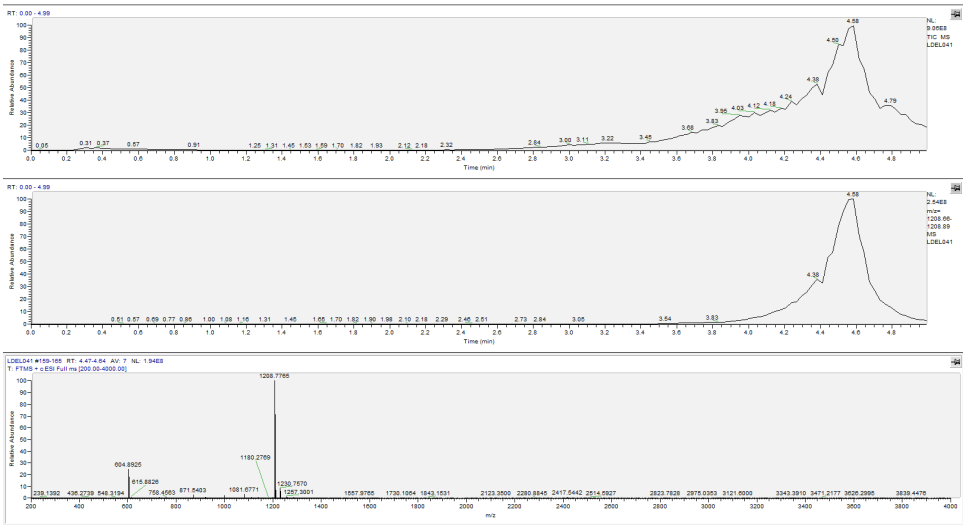
DLeu-Leu-DPro-DLeu-LA03-Pro-Leu-DLeuMe-Pro

Exact Mass: 1207.7693

ELSD of **LDEL041**

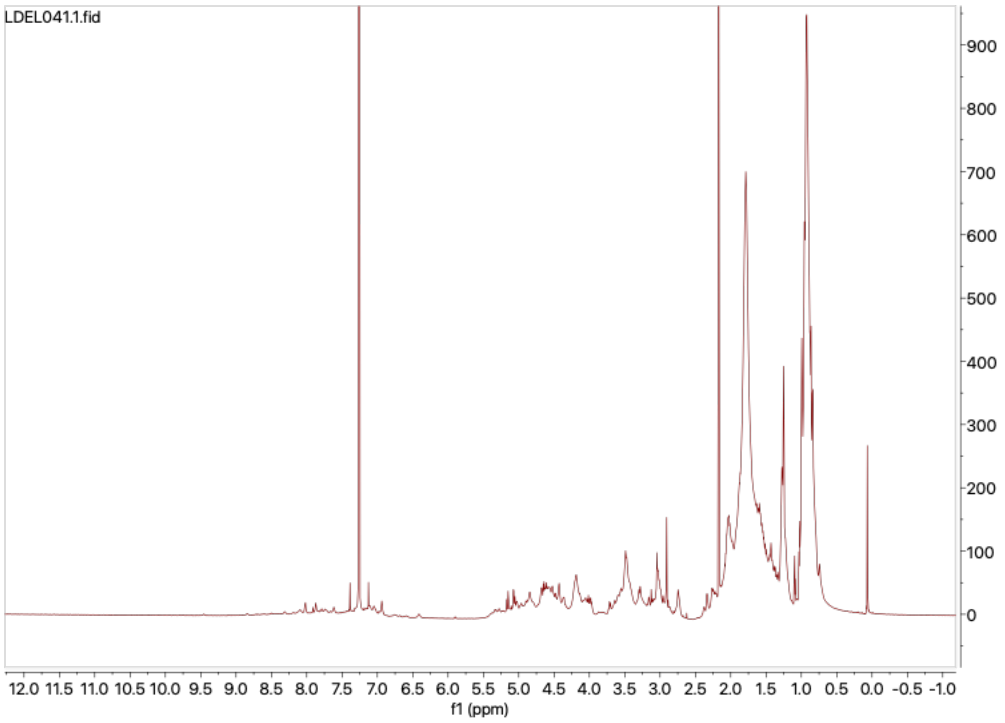


HRMS of **LDEL041**

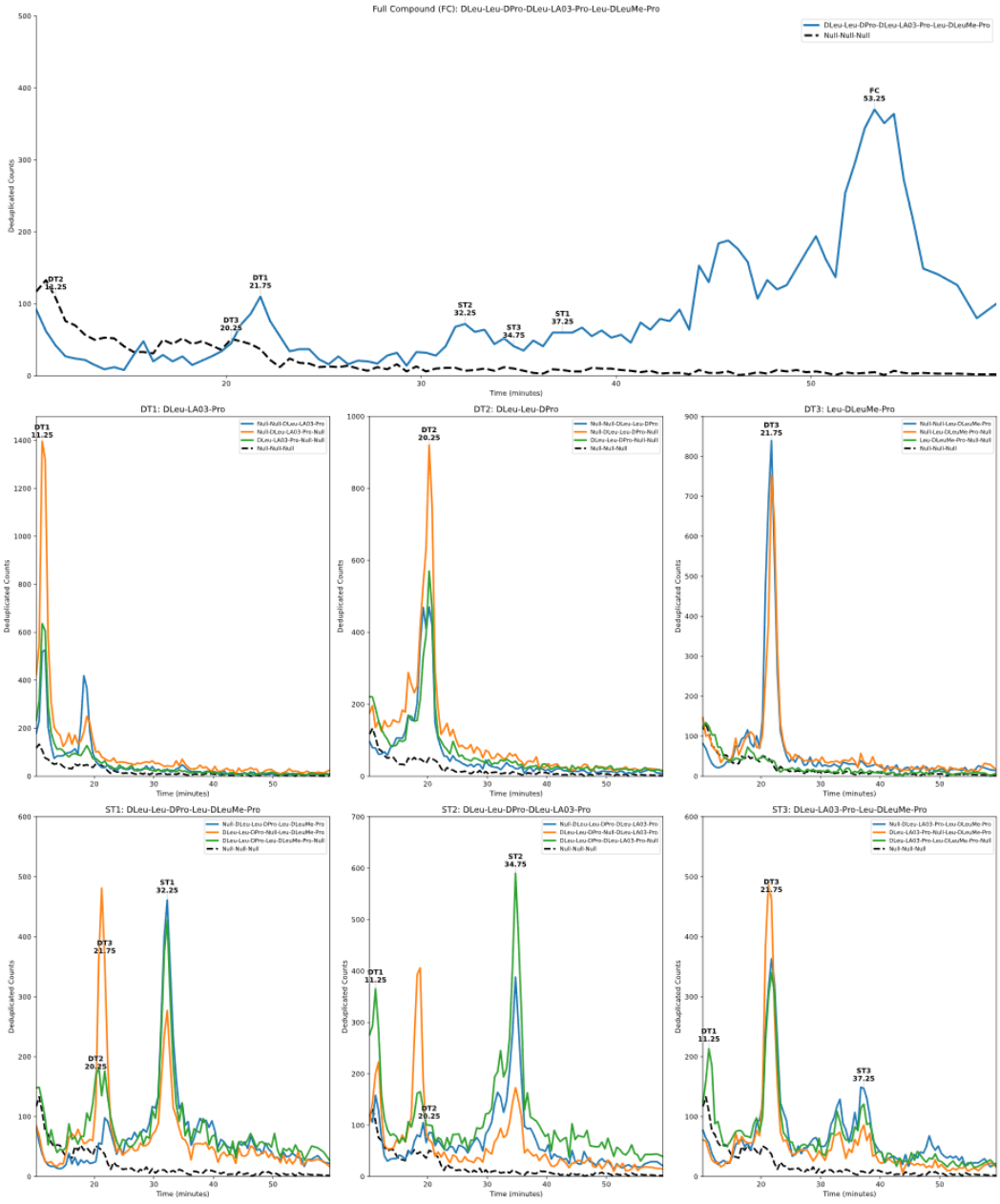


HRMS (ESI) m/z calculated for $C_{61}H_{101}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1208.7726; found 1208.7765

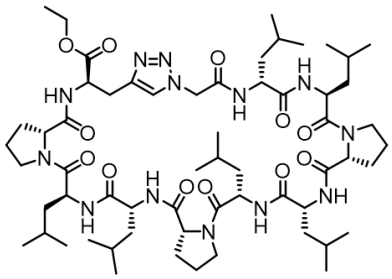
^{1}HNMR of **LDEL041**



Classic Chrome Sequence Count Chromatogram of **LDEL041**



LDEL042

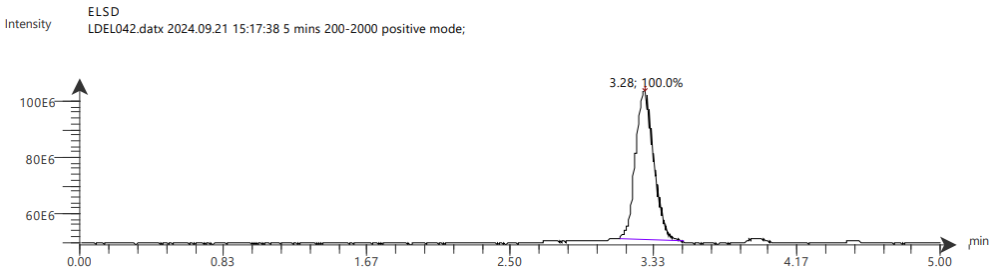


LDEL042

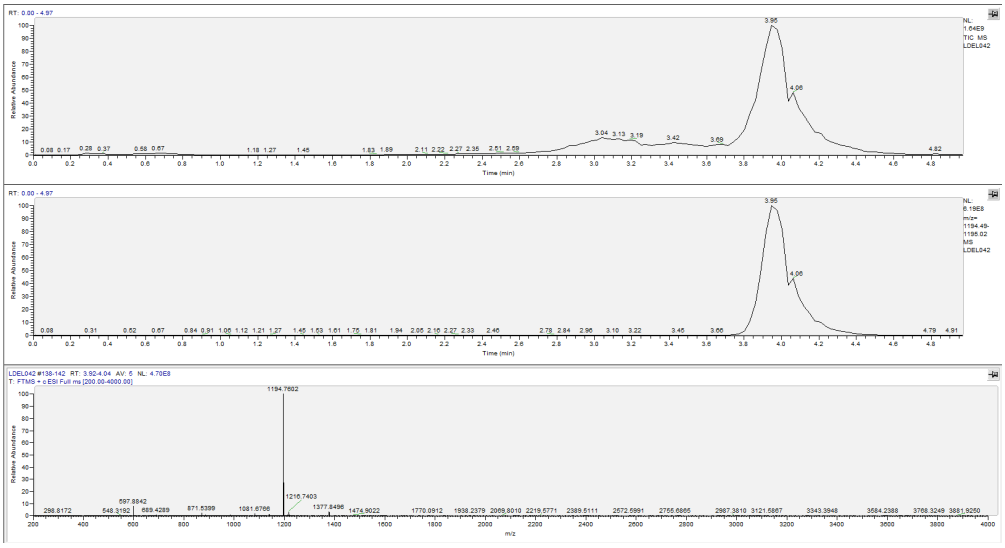
DLeu-Leu-DPro-DLeu-Leu-DPro-DLeu-Leu-DPro

Exact Mass: 1193.7536

ELSD of **LDEL042**

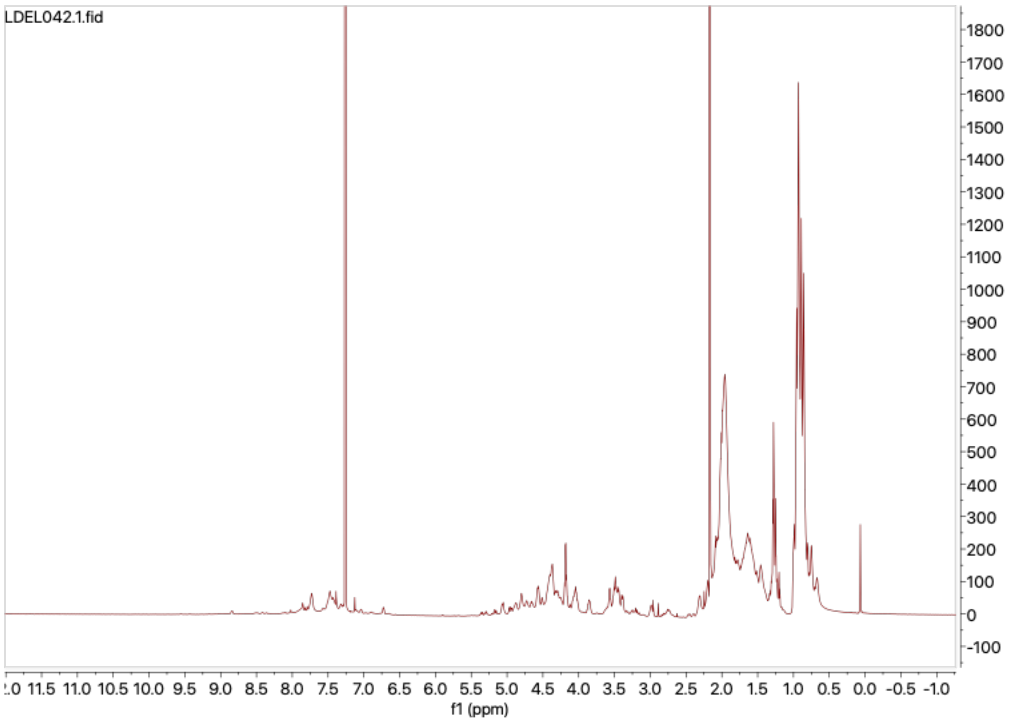


HRMS of **LDEL042**

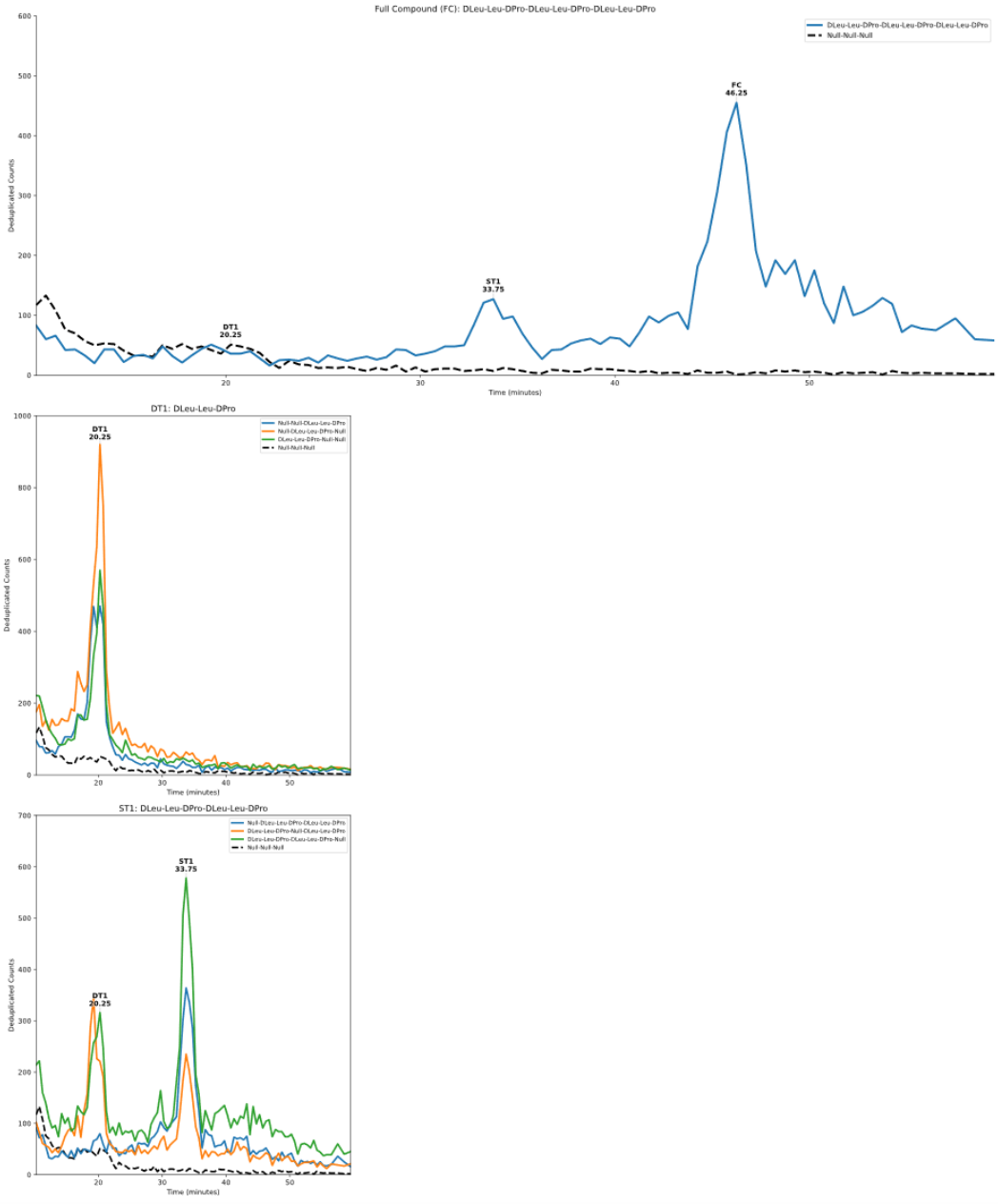


HRMS (ESI) m/z calculated for $C_{60}H_{99}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1194.7570; found 1194.7662

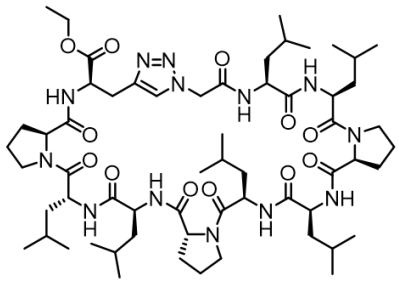
^{1}HNMR of **LDEL042**



Classic Chrome Sequence Count Chromatogram of **LDEL042**



LDEL043

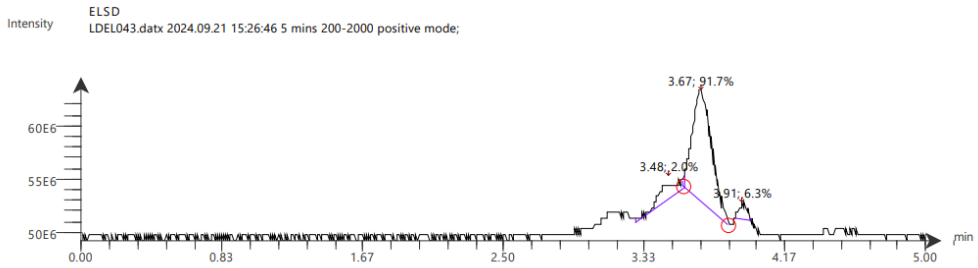


LDEL043

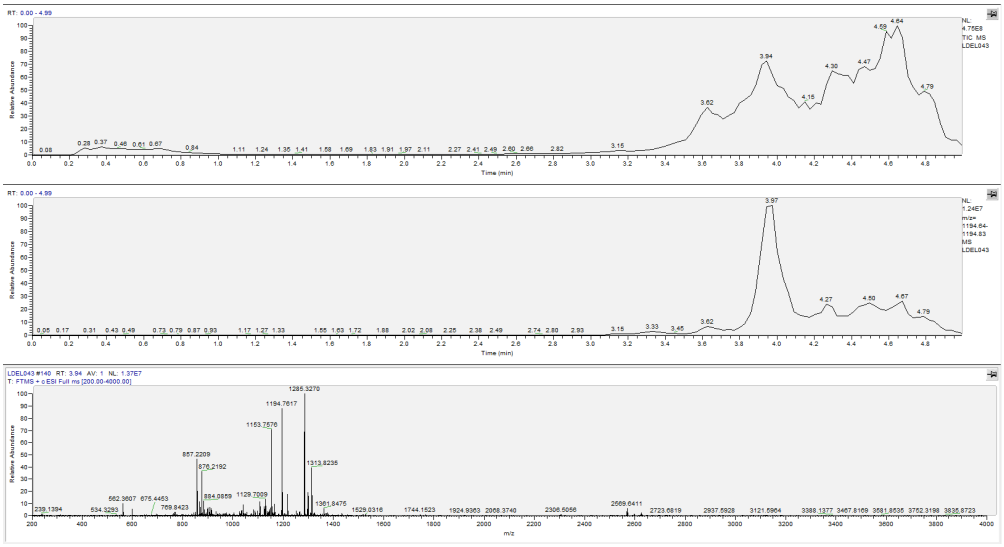
Leu-Leu-Pro-Leu-DLeu-DPro-Leu-DLeu-Pro

Exact Mass: 1193.7536

ELSD of **LDEL043**

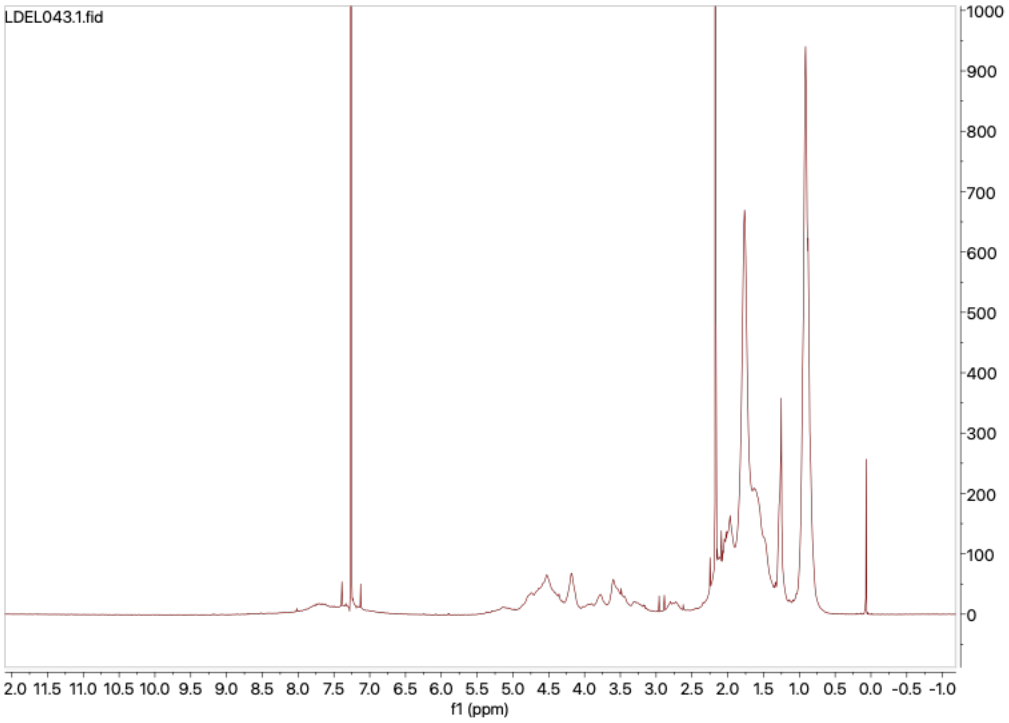


HRMS of **LDEL043**

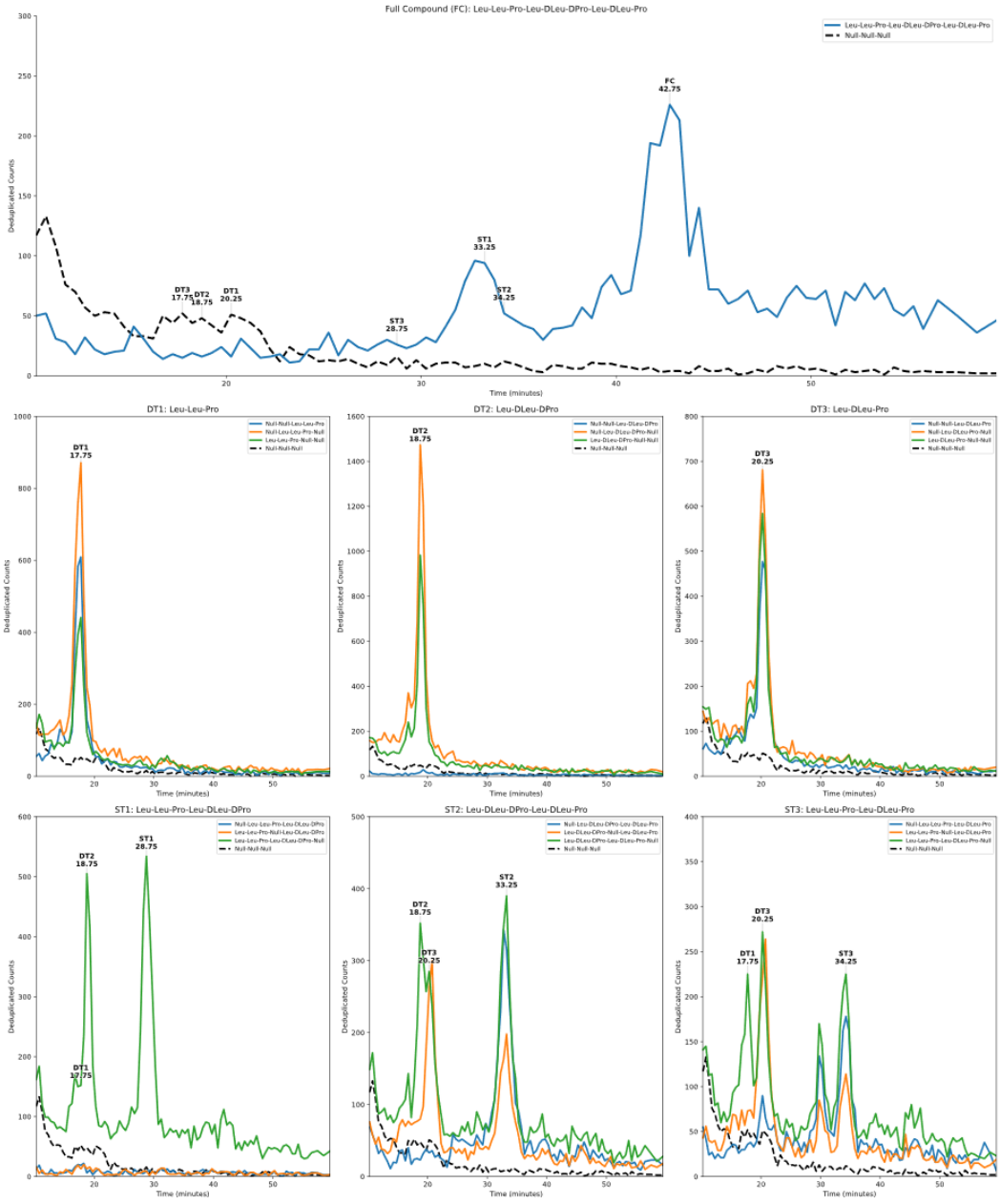


HRMS (ESI) m/z calculated for $C_{60}H_{99}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1194.7570; found 1194.7617

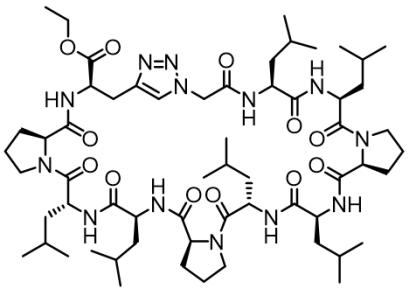
^{1}HNMR of **LDEL043**



Classic Chrome Sequence Count Chromatogram of **LDEL043**



LDEL044

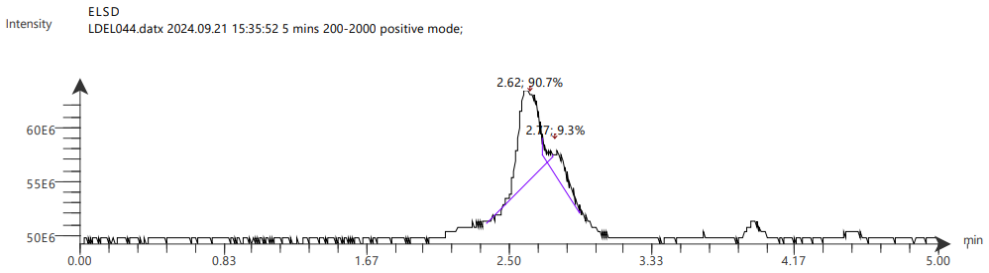


LDEL044

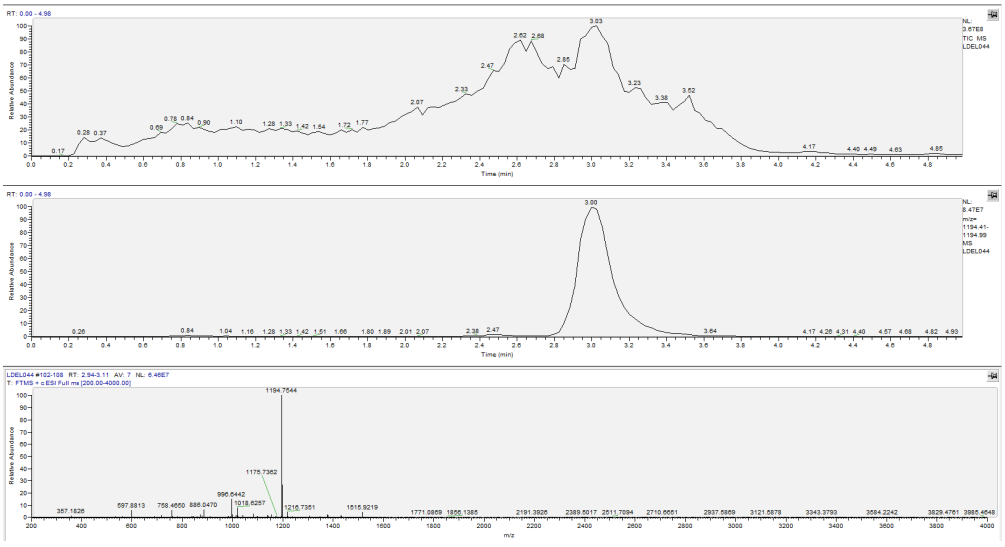
Leu-Leu-Pro-Leu-Leu-Pro-Leu-DLeu-Pro

Exact Mass: 1193.7536

ELSD of **LDEL044**

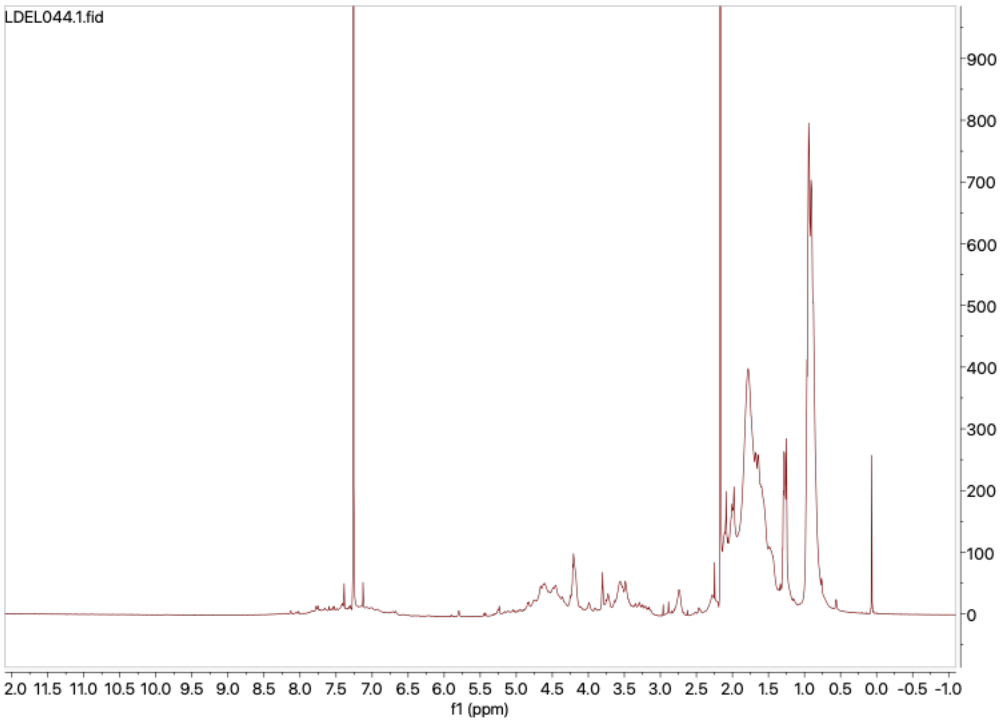


HRMS of **LDEL044**

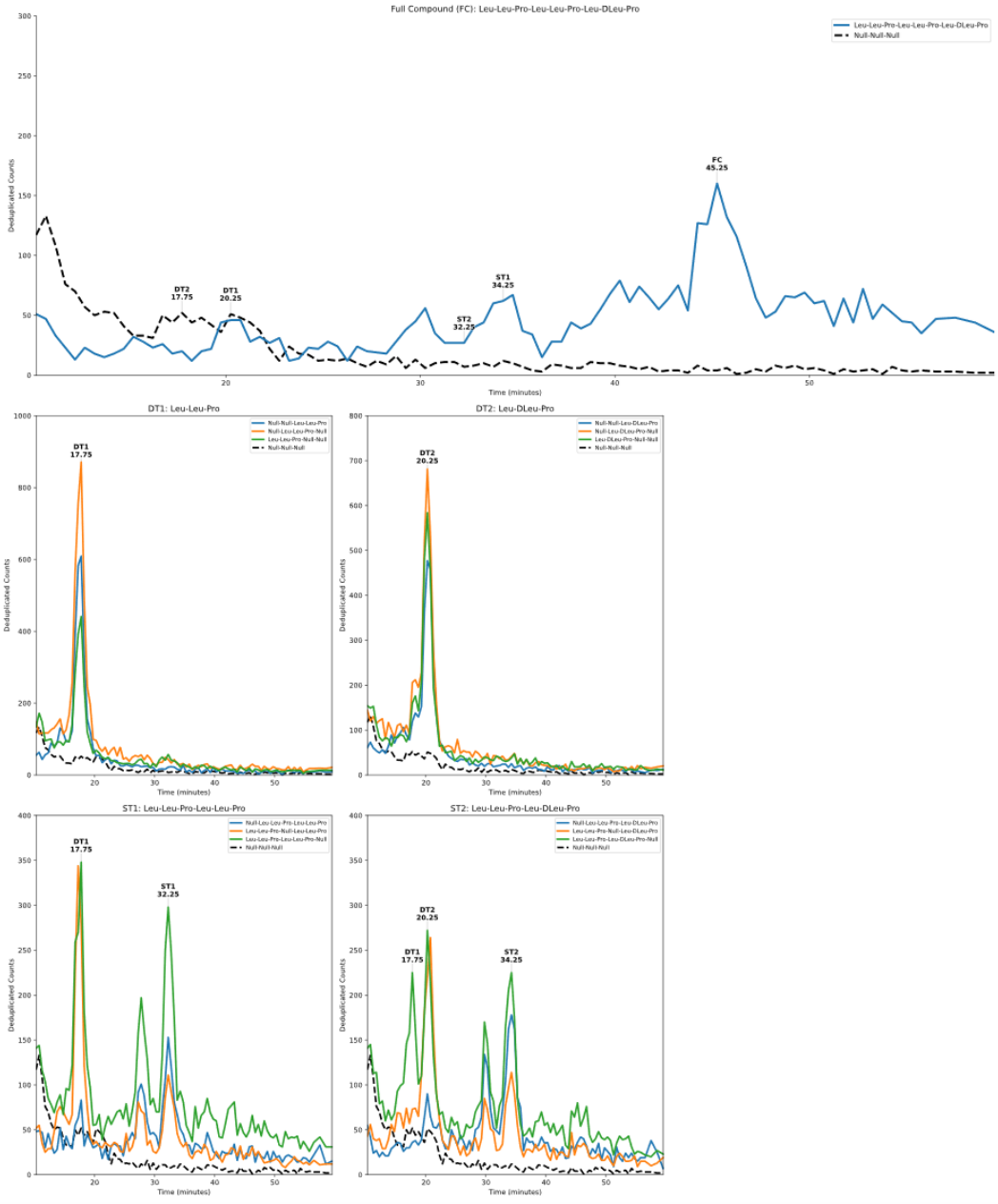


HRMS (ESI) m/z calculated for $C_{60}H_{99}N_{13}O_{12}^{+}$ ($M+H^{+}$) 1194.7570; found 1194.7544

^{1}HNMR of **LDEL044**



Classic Chrome Sequence Count Chromatogram of **LDEL044**

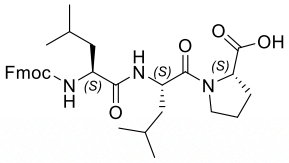


Appendix G: LCMS Traces and HNMR of Custom Tripeptide Building Blocks

Tripeptide building blocks experienced significant in-source fragmentation about the C-terminal proline; the largest mass fragment is often the highest intensity peak in these spectra

BB019

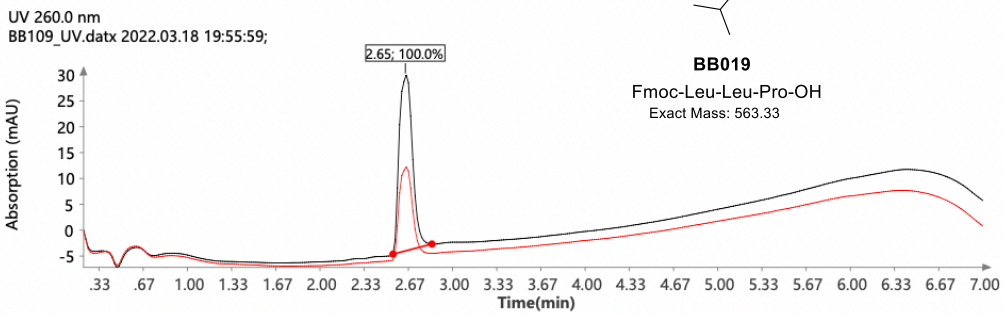
LCMS of Building Block 019

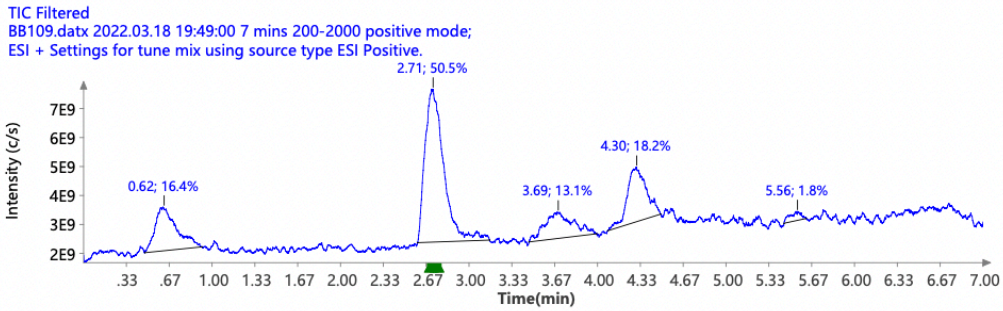


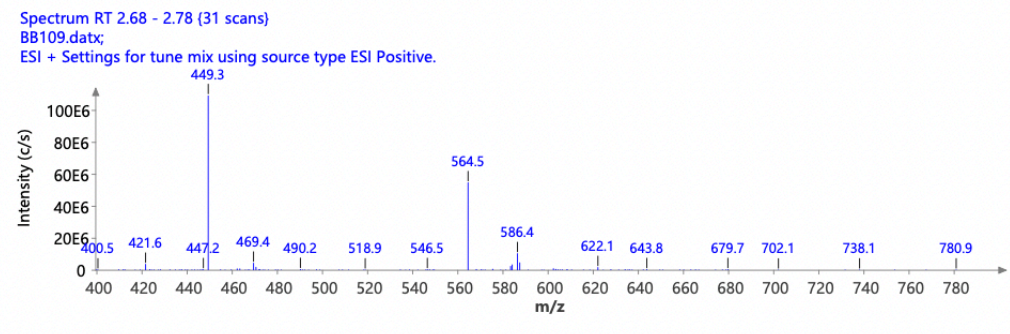
BB019

Fmoc-Leu-Leu-Pro-OH

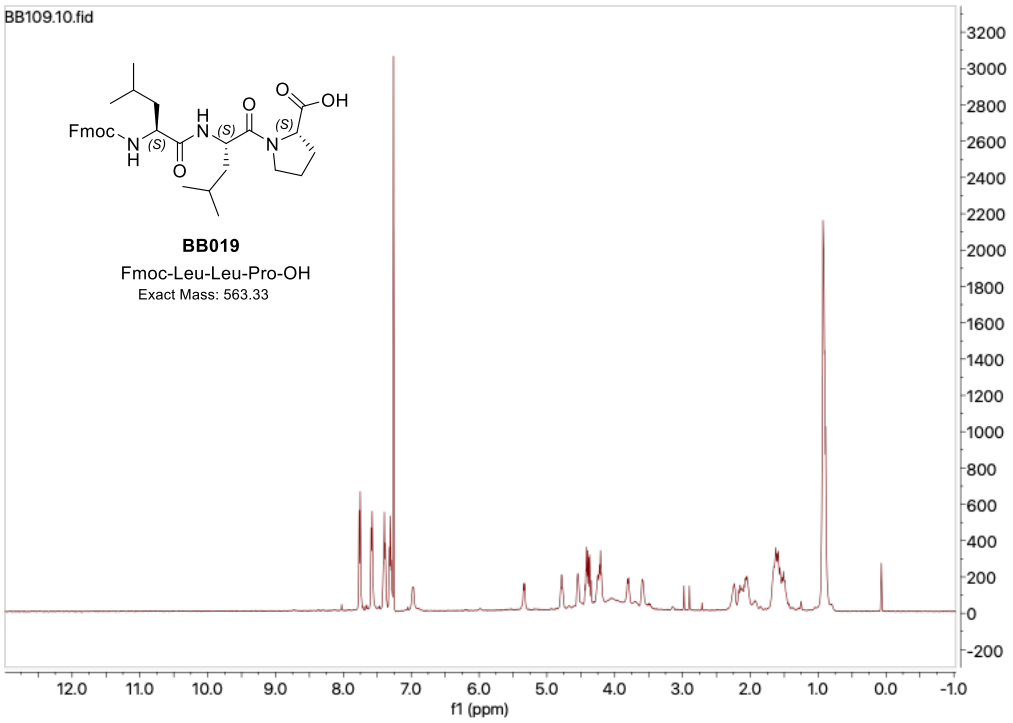
Exact Mass: 563.33





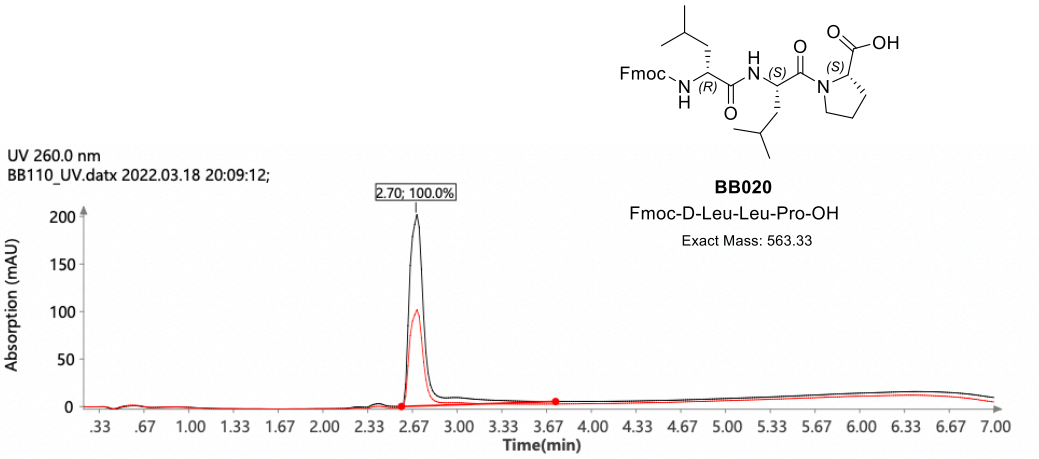


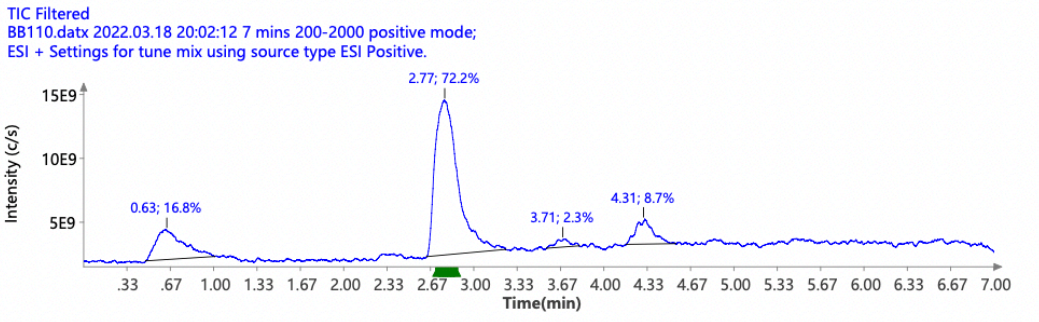
^{1}H-NMR of Building Block 019

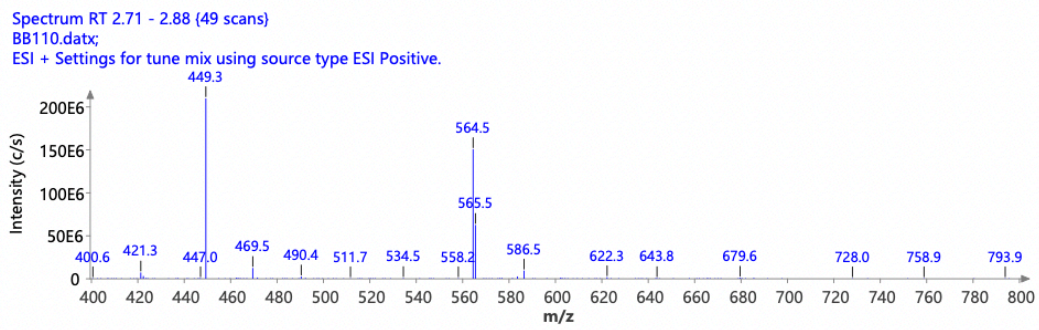


BB020

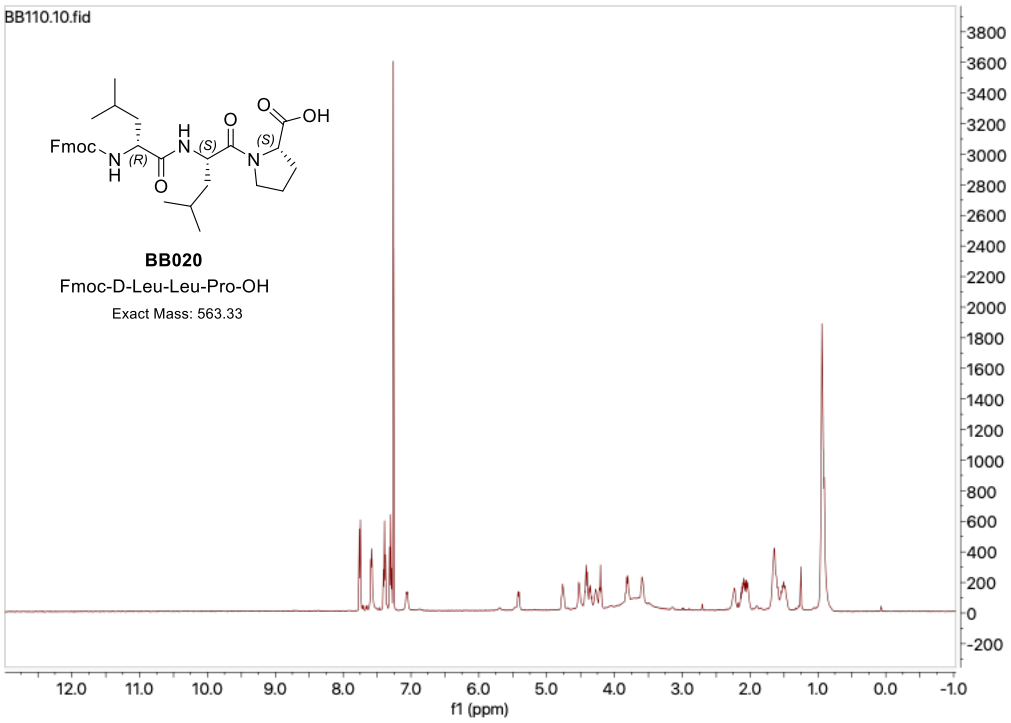
LCMS of Building Block 020





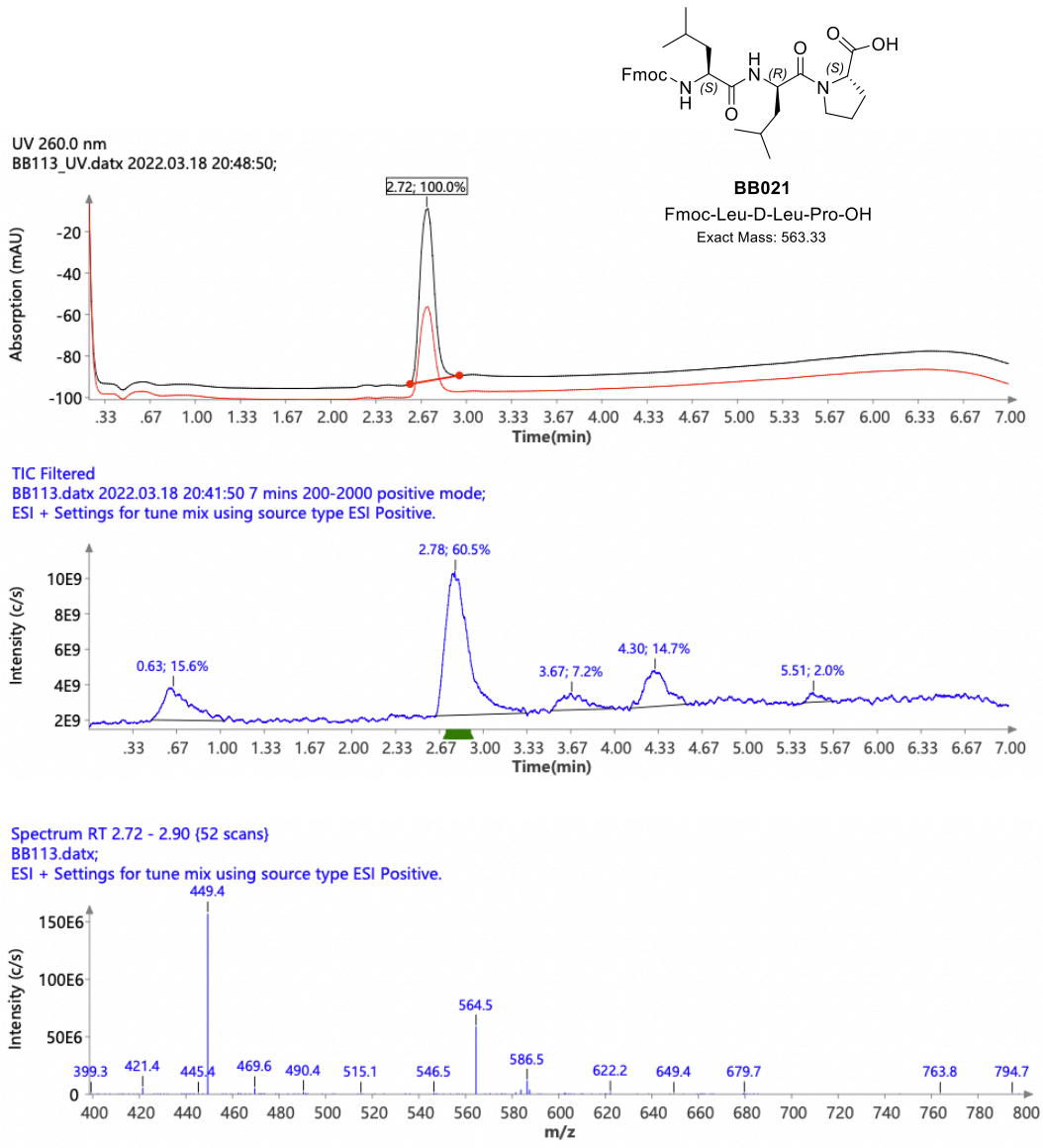


[1]H-NMR of Building Block 020

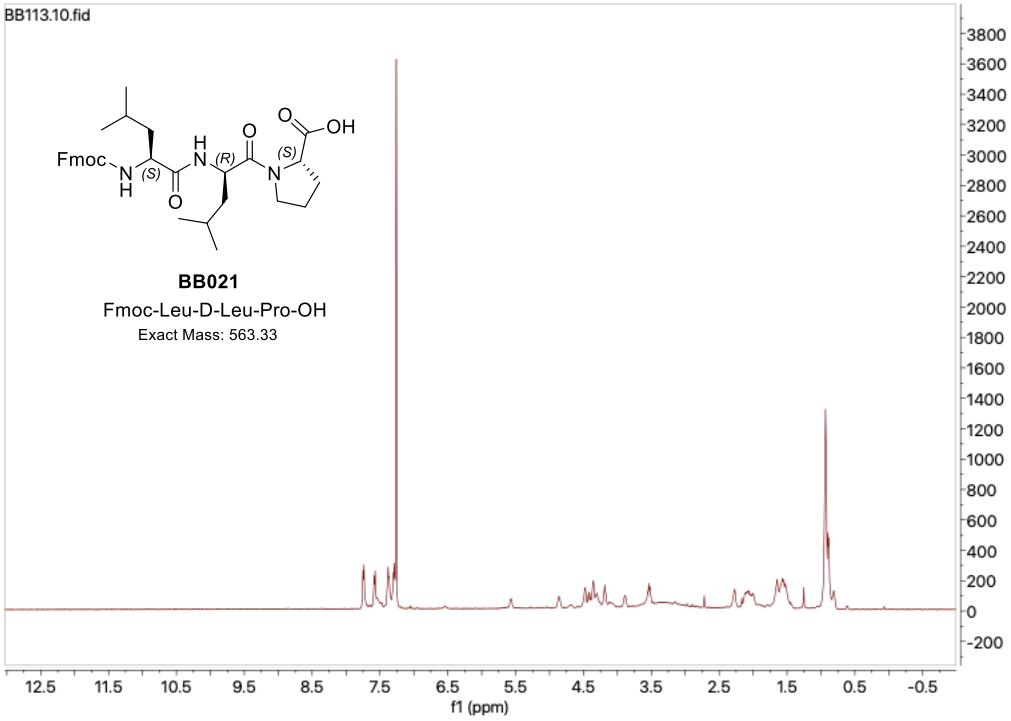


BB021

LCMS of Building Block 021

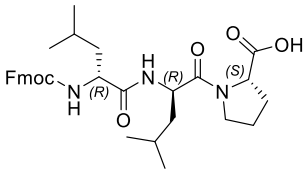


^{1}H-NMR of Building Block 021



BB022

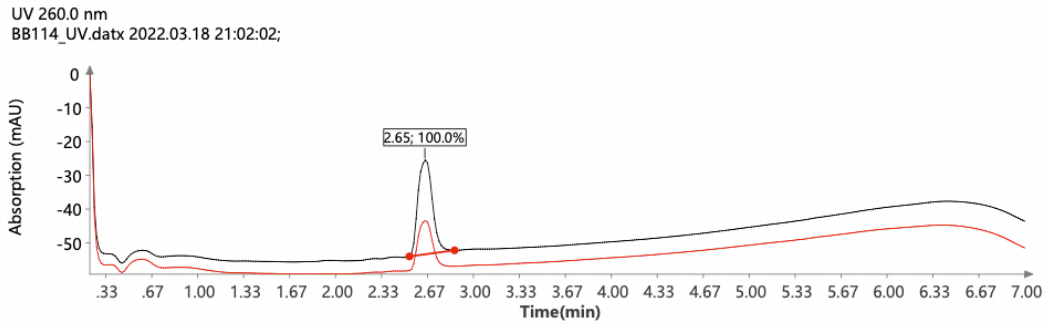
LCMS of Building Block 022

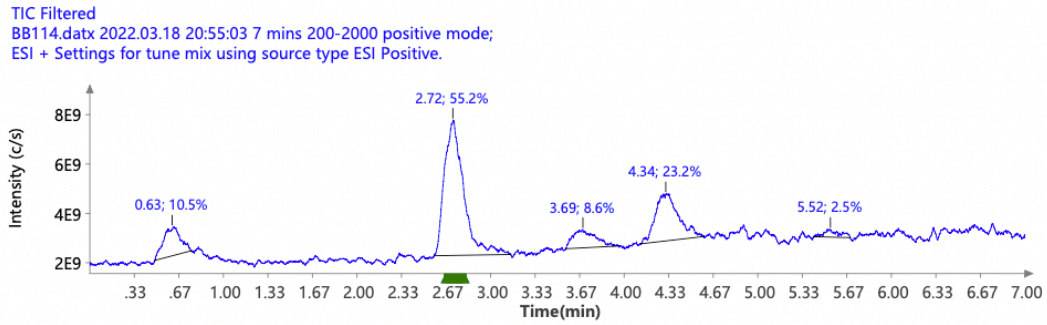


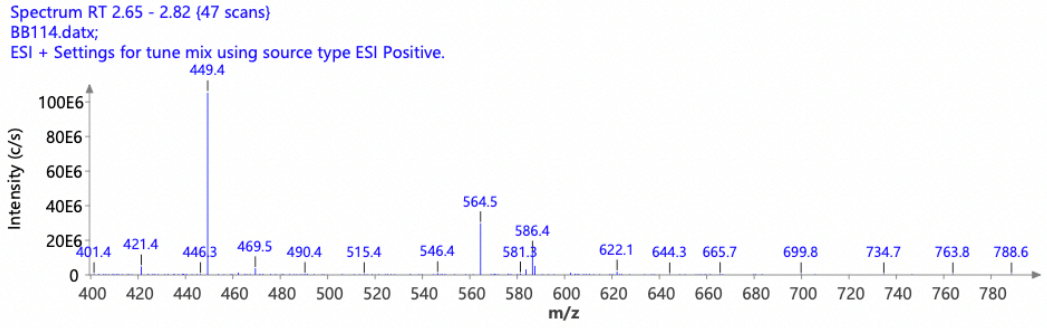
BB022

Fmoc-D-Leu-D-Leu-Pro-OH

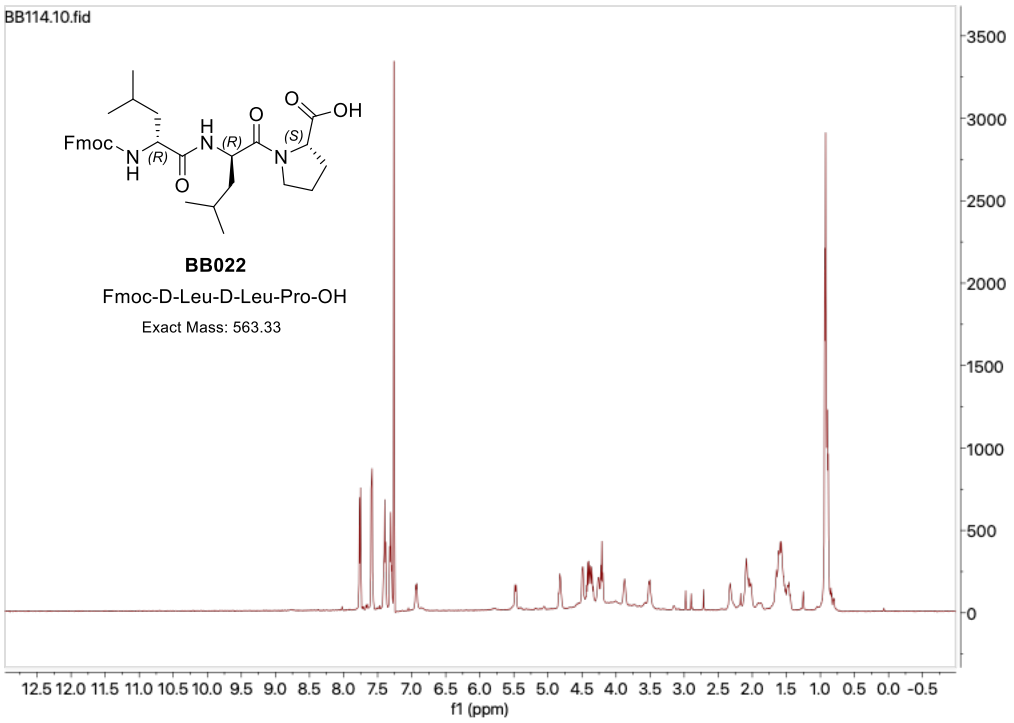
Exact Mass: 563.33





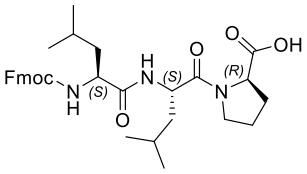


^{1}H-NMR of Building Block 022

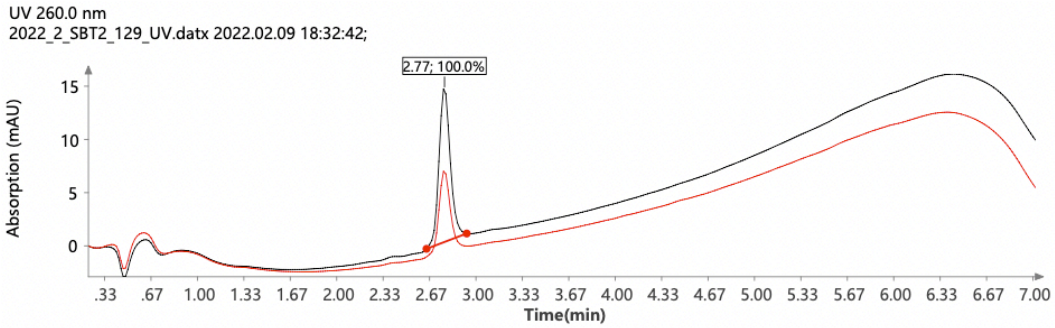


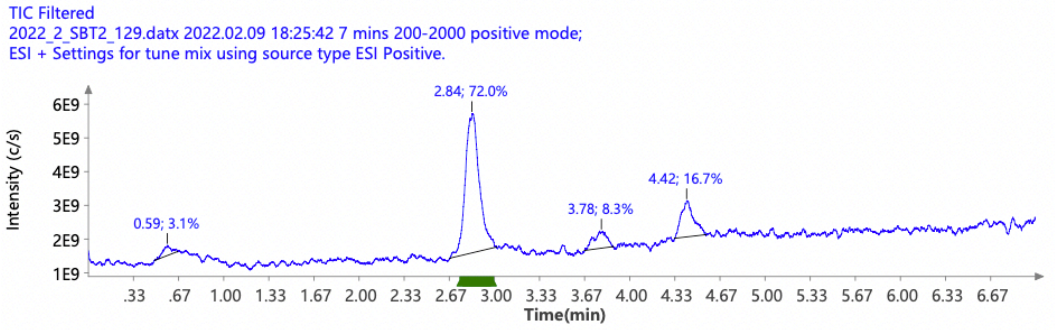
BB023

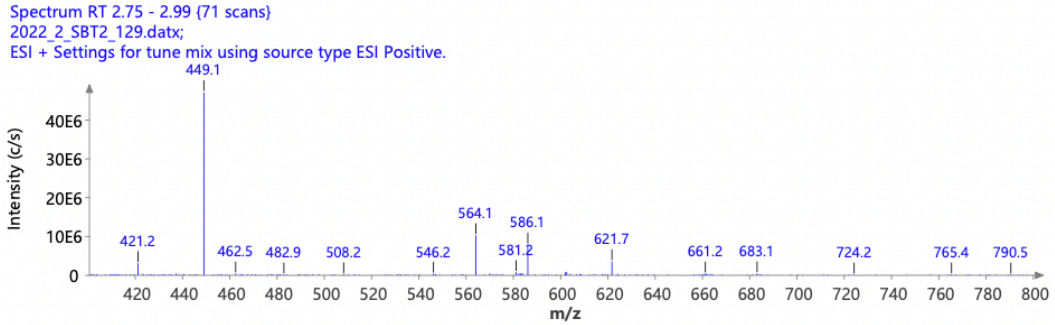
LCMS of Building Block 023



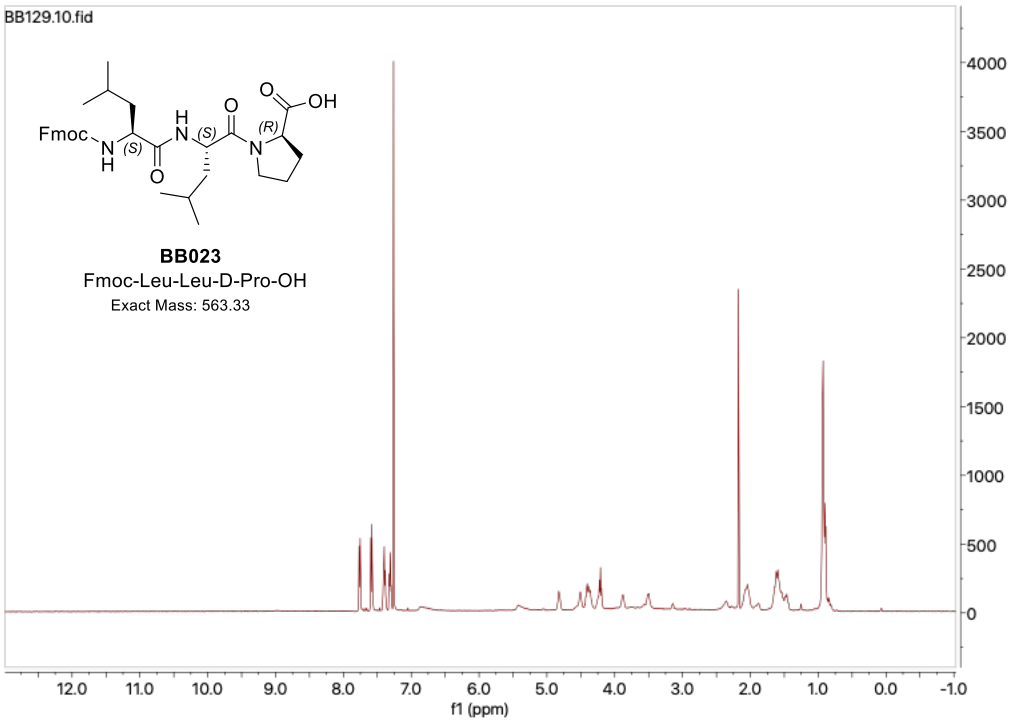
BB023
Fmoc-Leu-Leu-D-Pro-OH
Exact Mass: 563.33





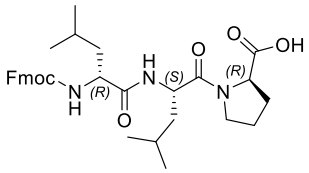


[1]H-NMR of Building Block 023

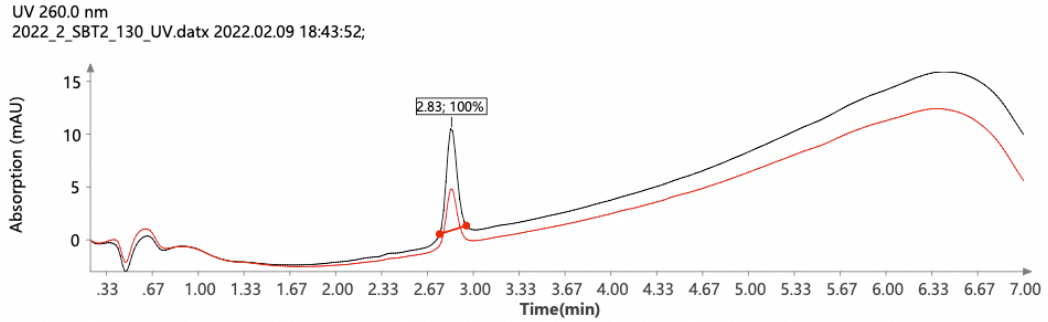


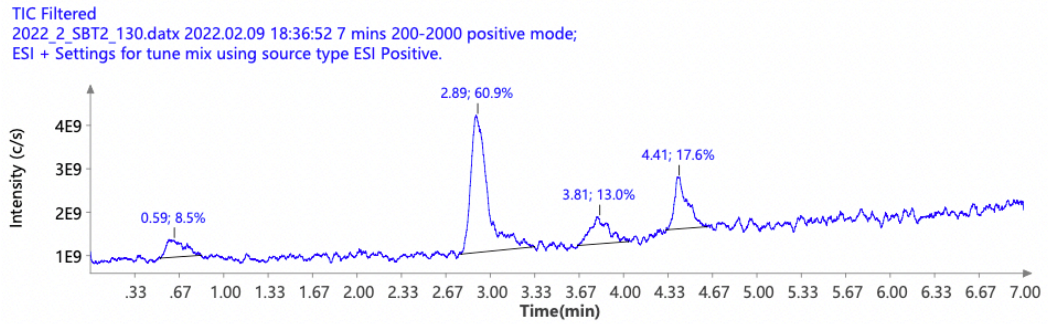
BB024

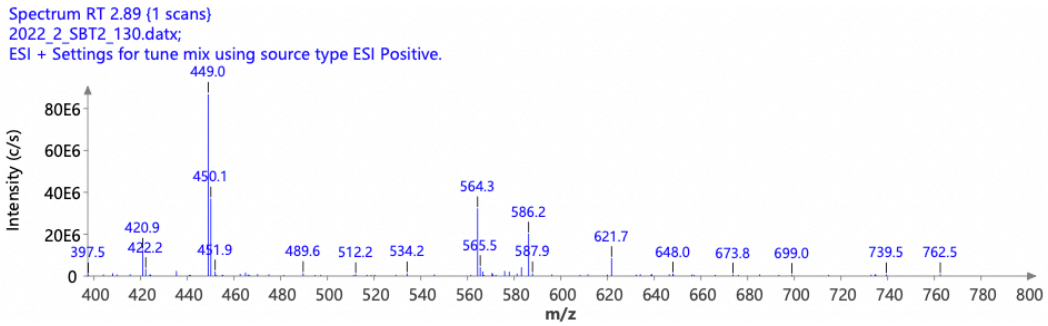
LCMS of Building Block 024



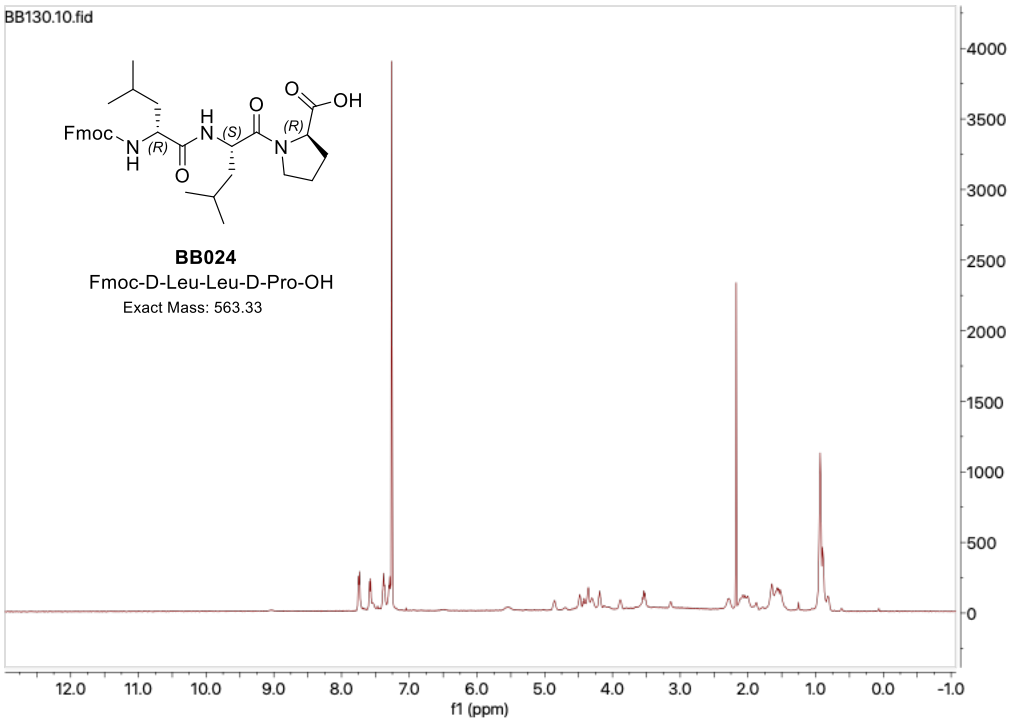
BB024
Fmoc-D-Leu-Leu-D-Pro-OH
Exact Mass: 563.33





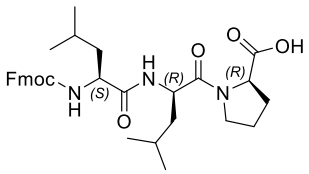


^{1}H-NMR of Building Block 024



BB025

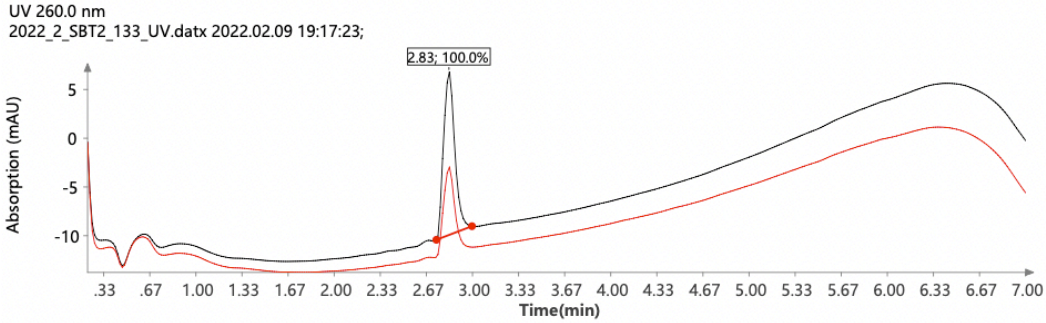
LCMS of Building Block 025

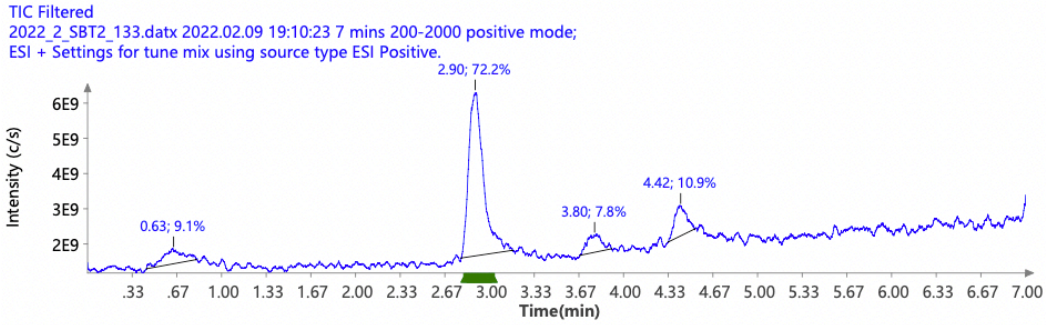


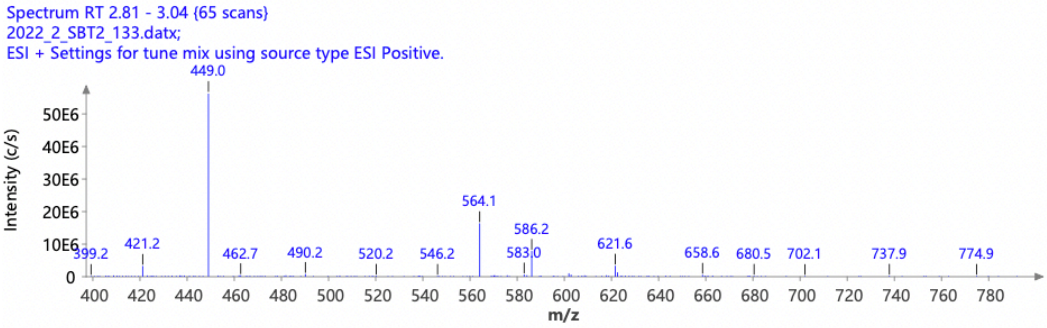
BB025

Fmoc-Leu-D-Leu-D-Pro-OH

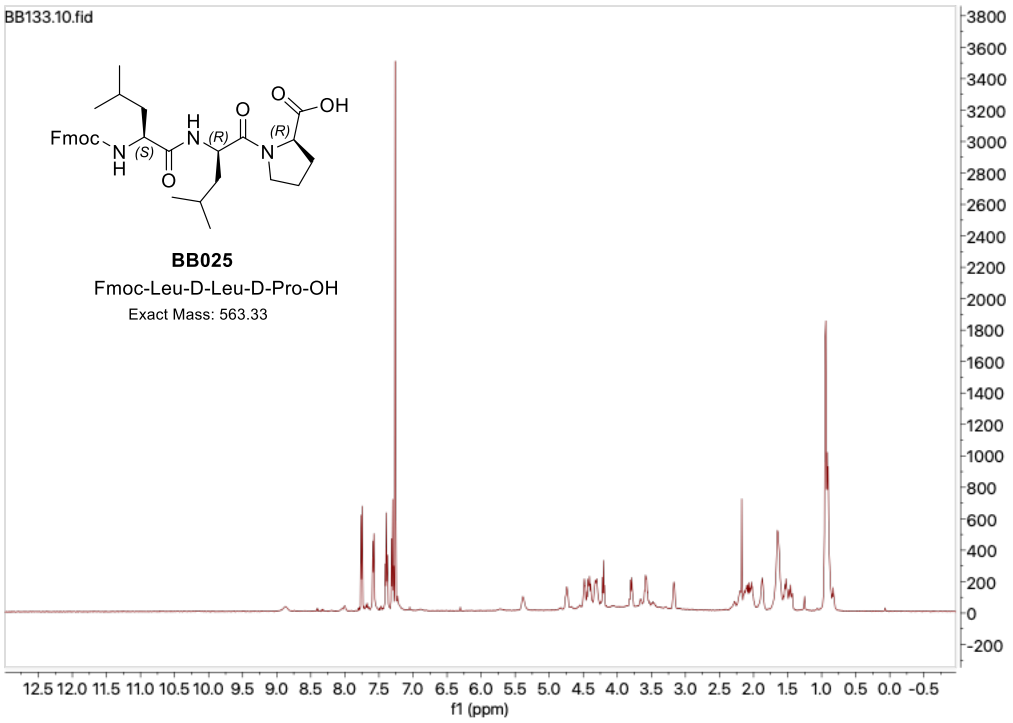
Exact Mass: 563.33





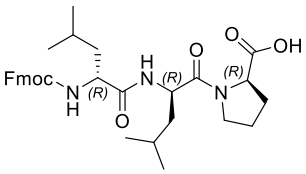


^{1}H-NMR of Building Block 025



BB026

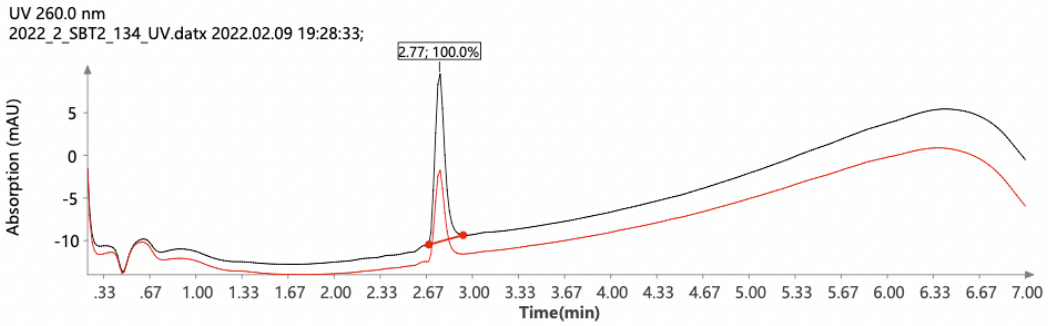
LCMS of Building Block 026

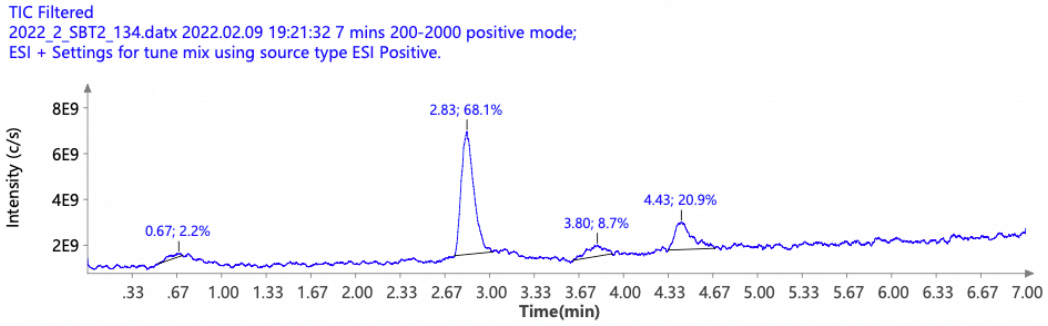


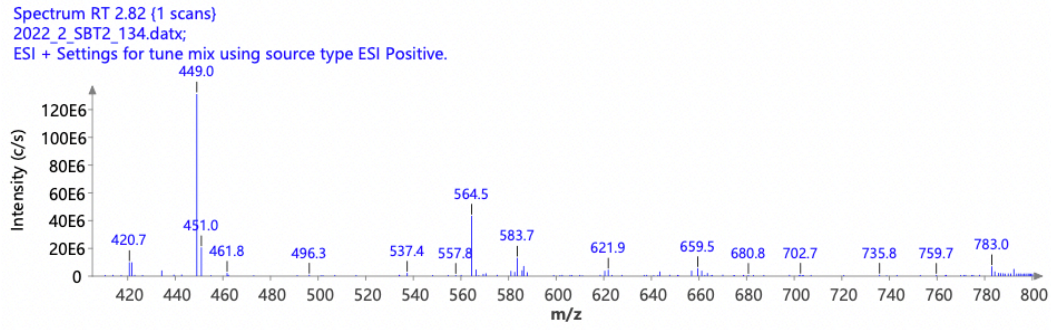
BB026

Fmoc-D-Leu-D-Leu-D-Pro-OH

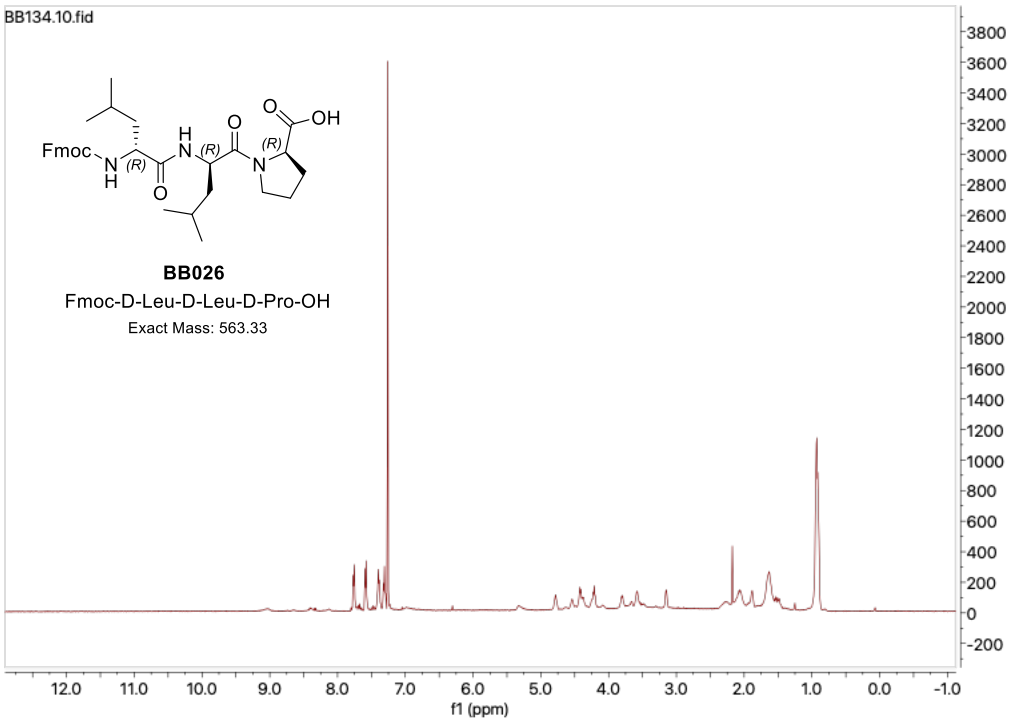
Exact Mass: 563.33





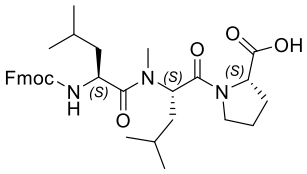


[1]H-NMR of Building Block 026



BB027

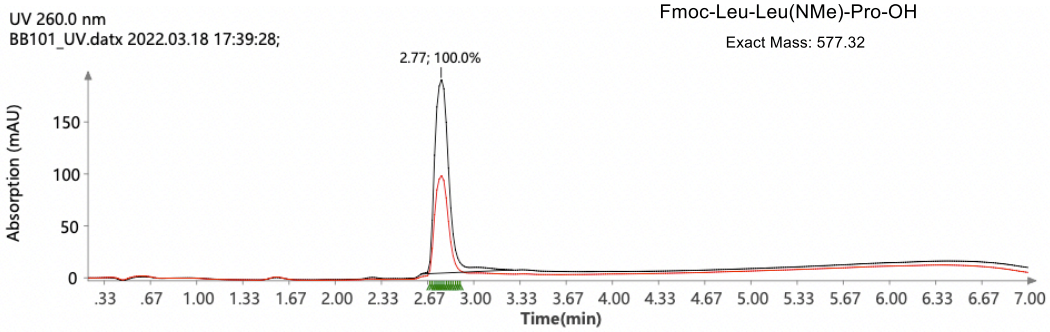
LCMS of Building Block 027

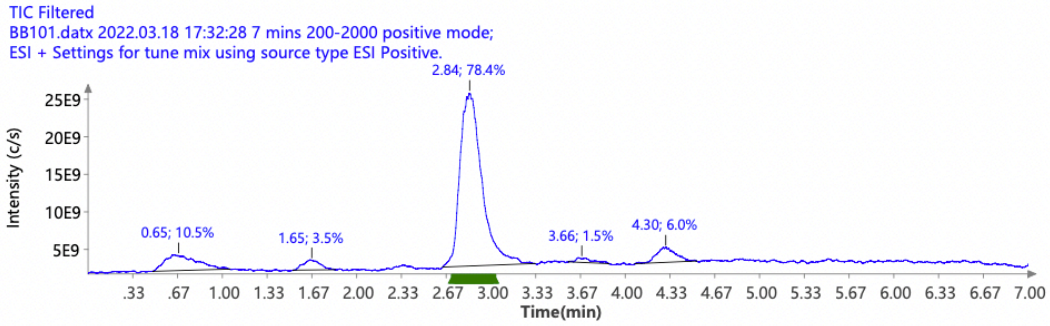


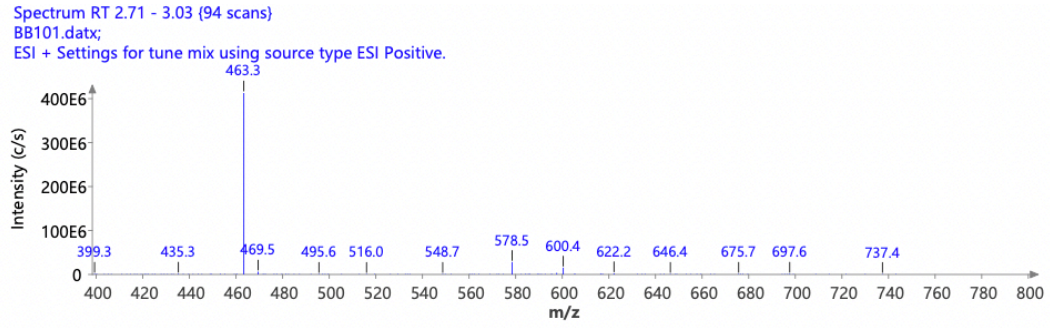
BB027

Fmoc-Leu-Leu(NMe)-Pro-OH

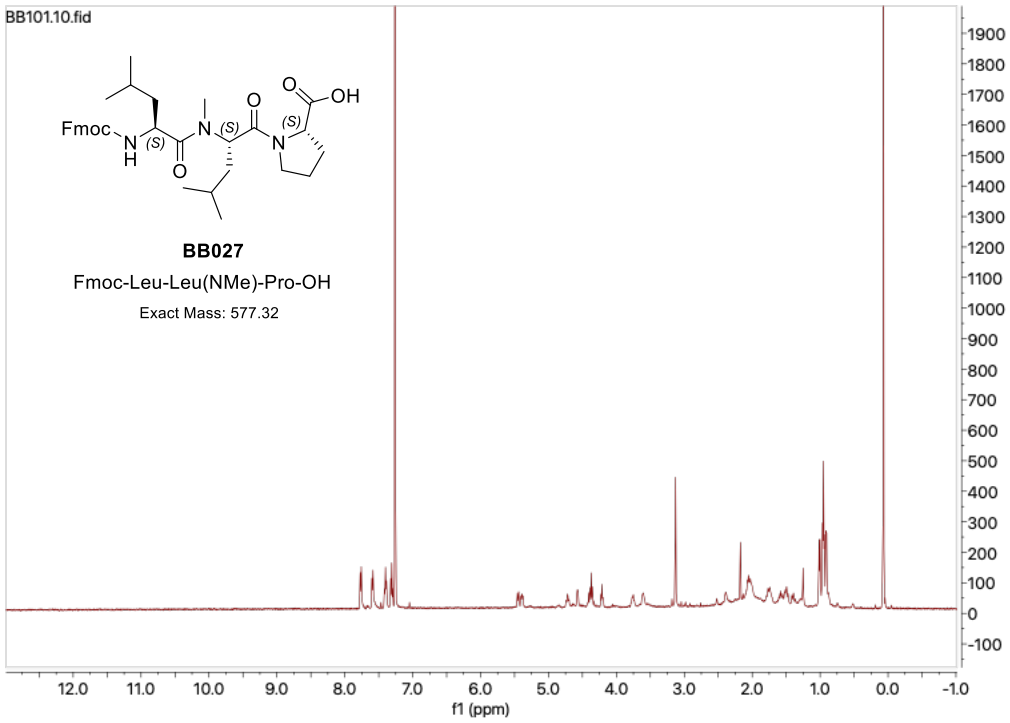
Exact Mass: 577.32





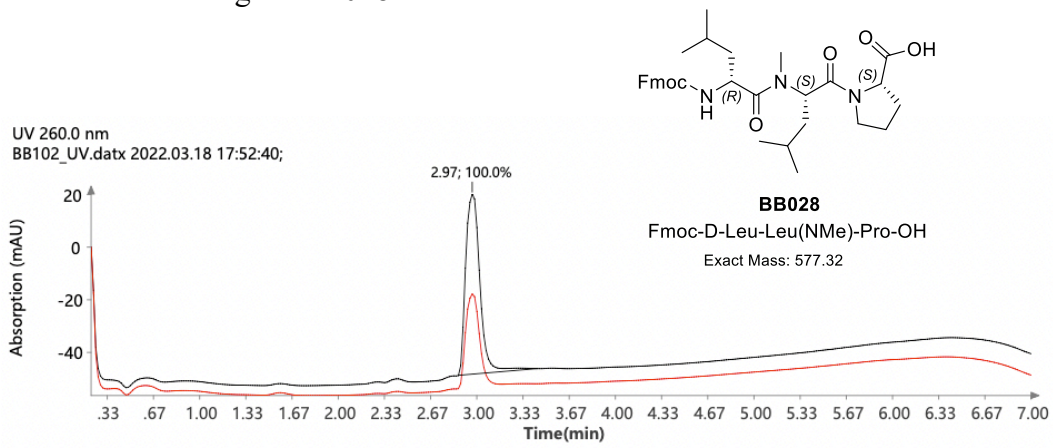


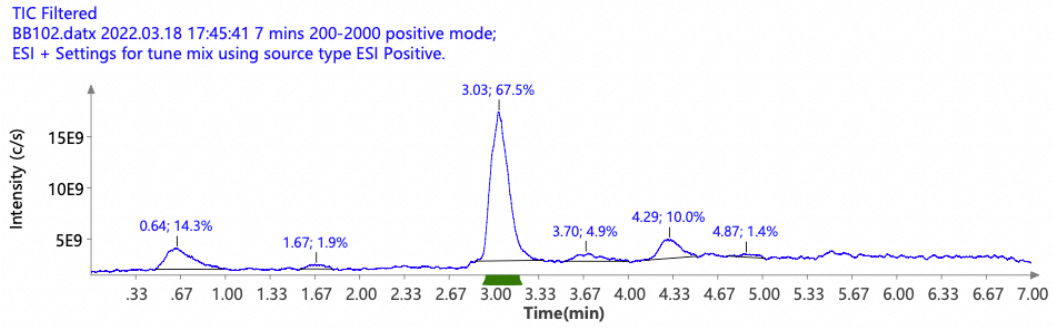
[1]H-NMR of Building Block 027

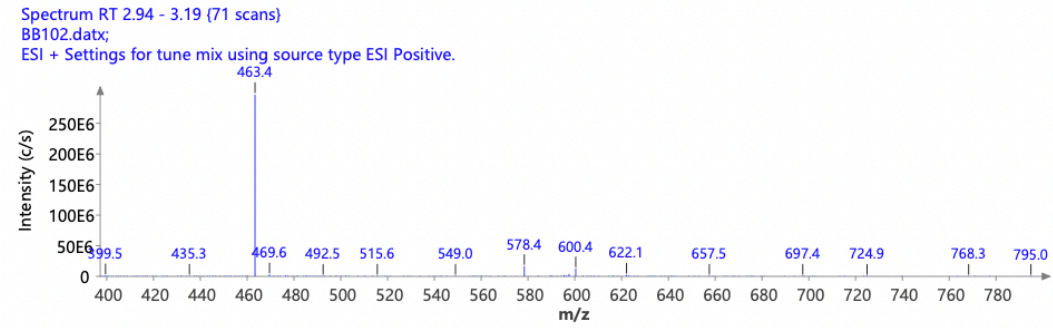


BB028

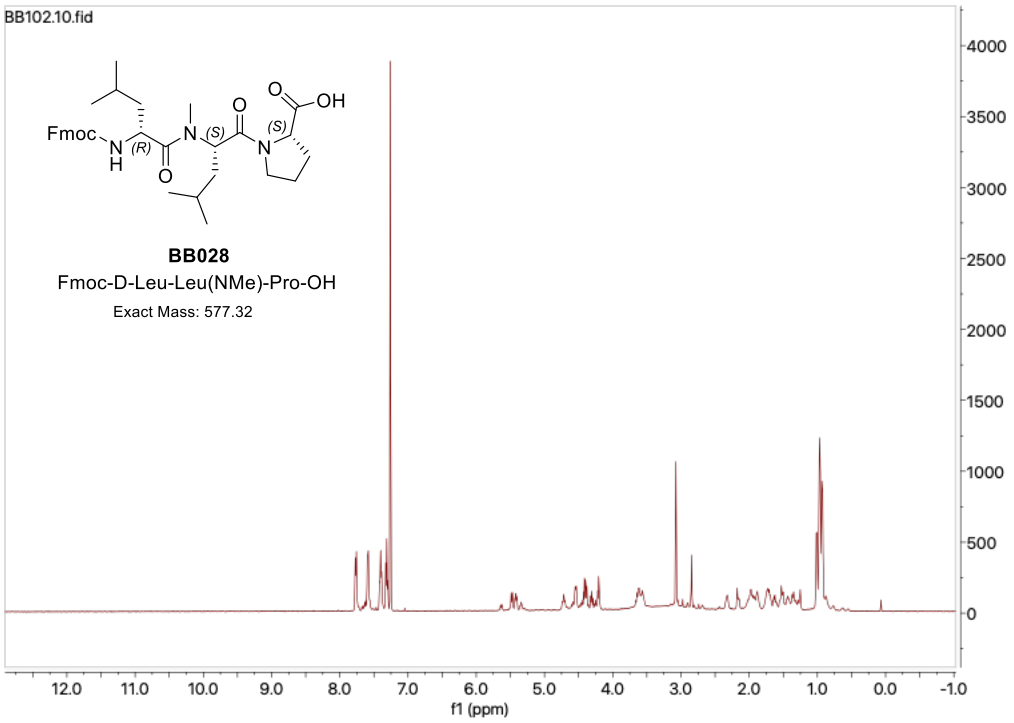
LCMS of Building Block 028





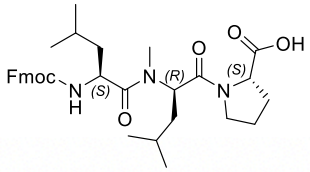


[1]H-NMR of Building Block 028



BB029

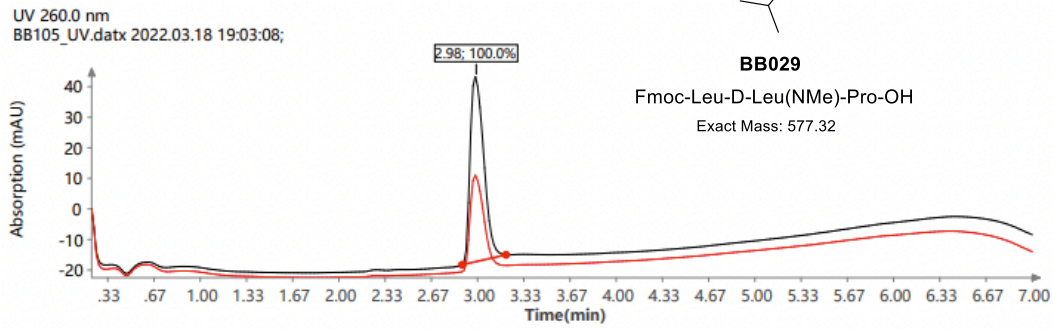
LCMS of Building Block 029

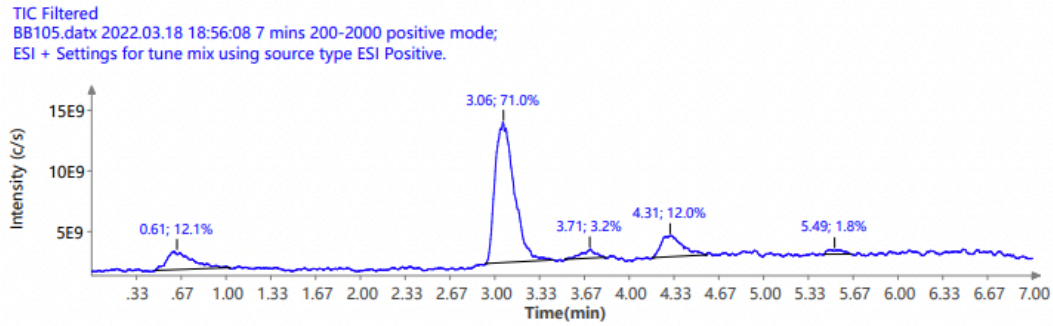


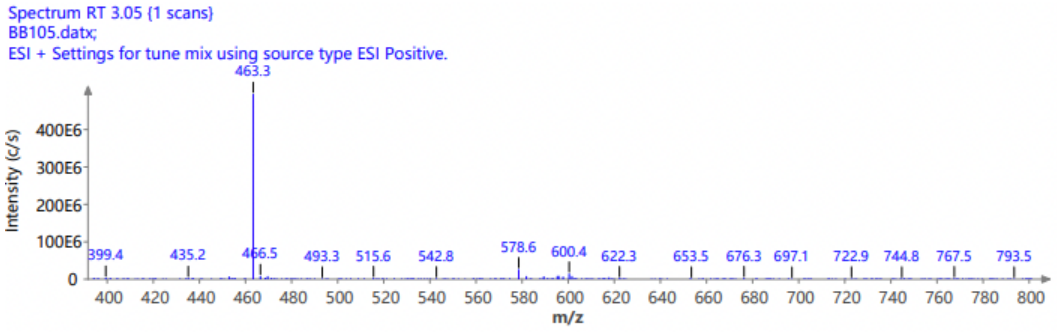
BB029

Fmoc-Leu-D-Leu(NMe)-Pro-OH

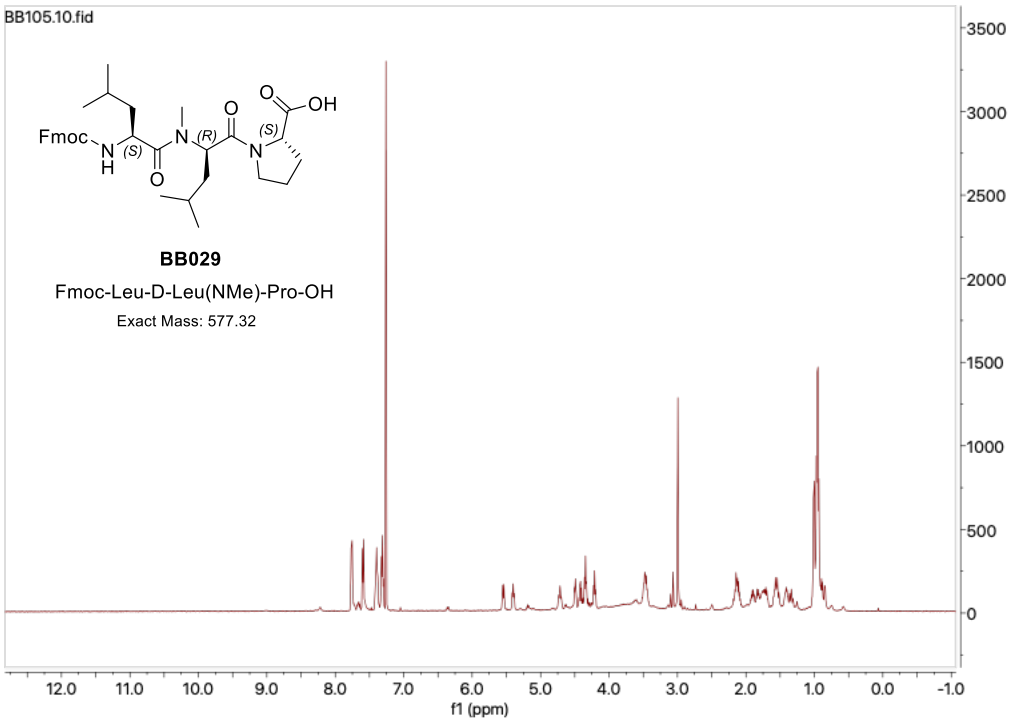
Exact Mass: 577.32





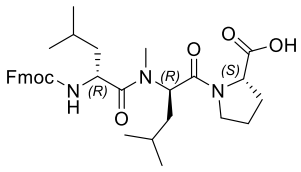


[1]H-NMR of Building Block 029



BB030

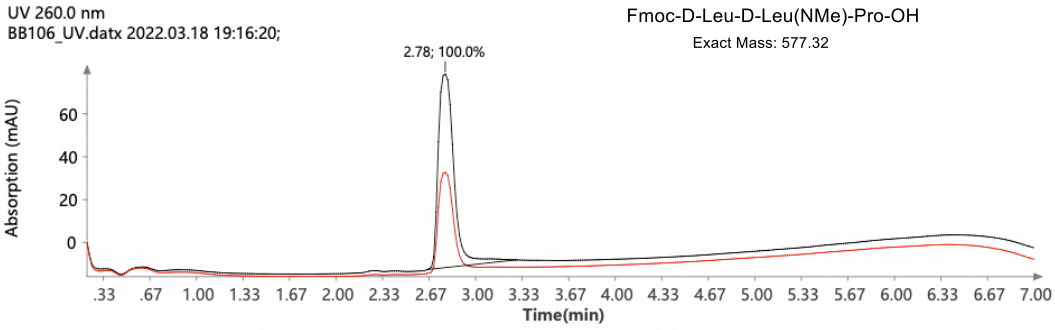
LCMS of Building Block 030

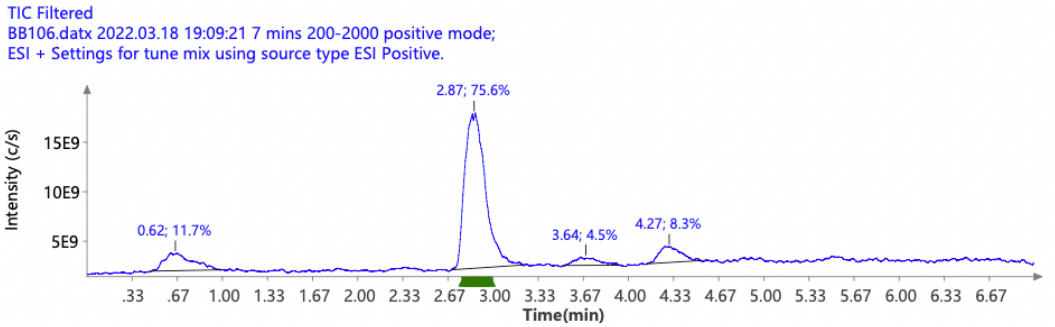


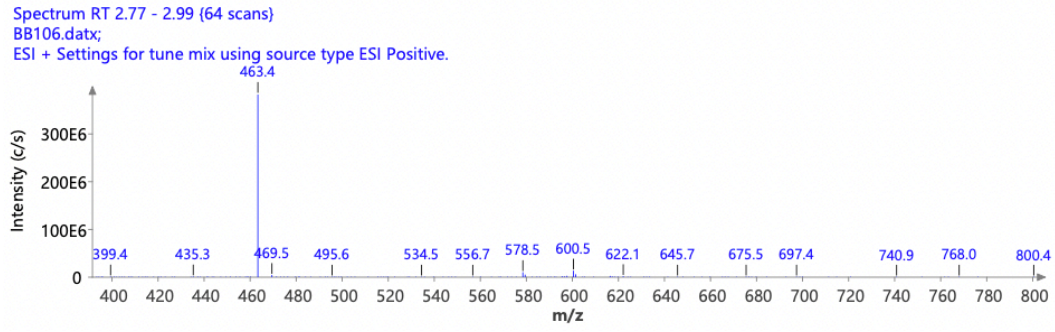
BB030

Fmoc-D-Leu-D-Leu(NMe)-Pro-OH

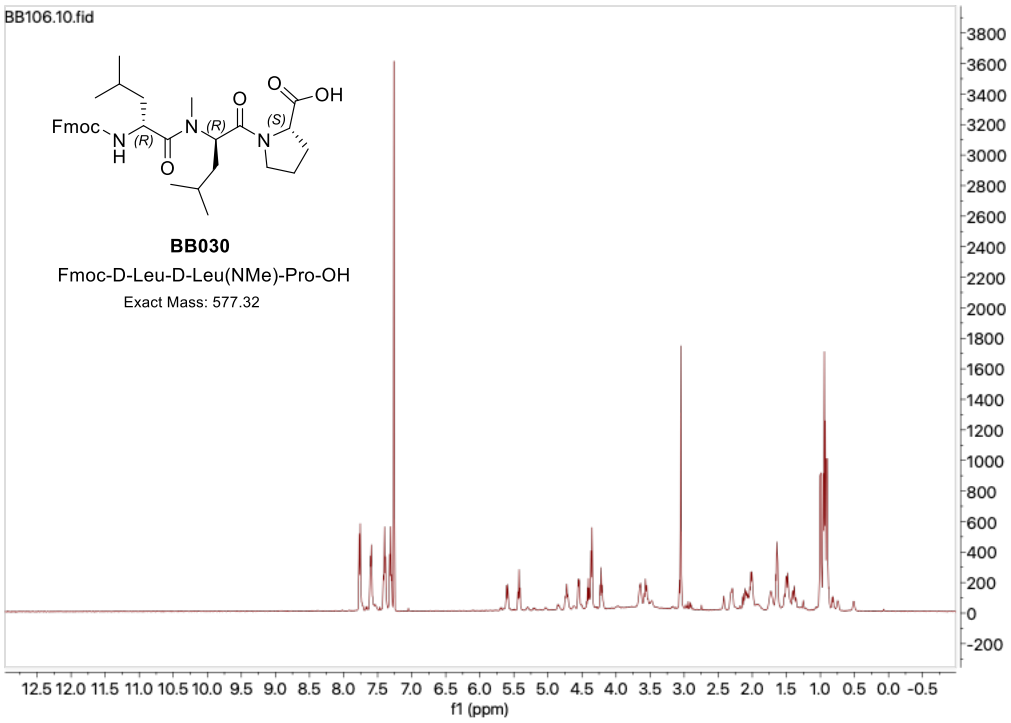
Exact Mass: 577.32





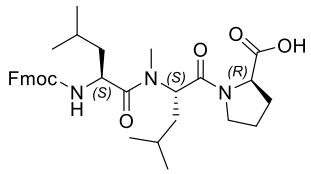


[1]H-NMR of Building Block 030



BB031

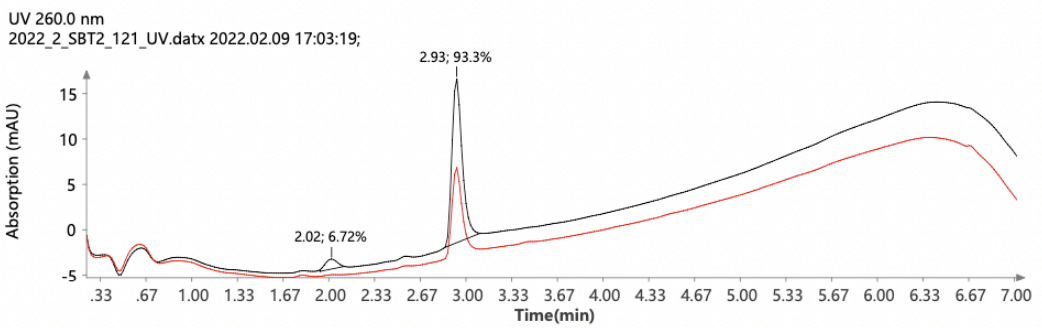
LCMS of Building Block 031

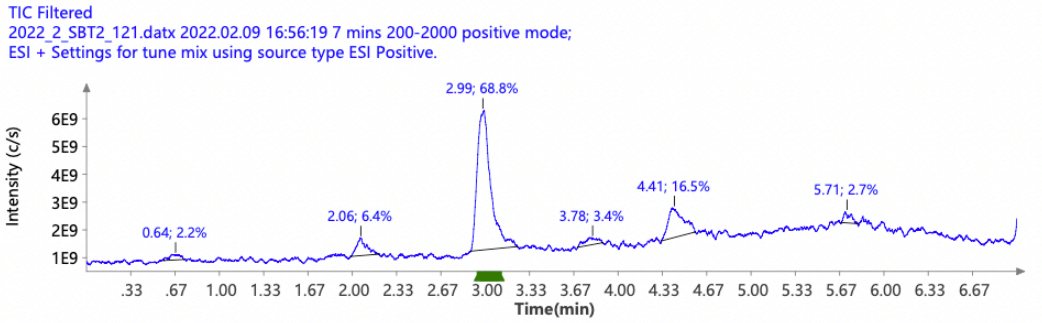


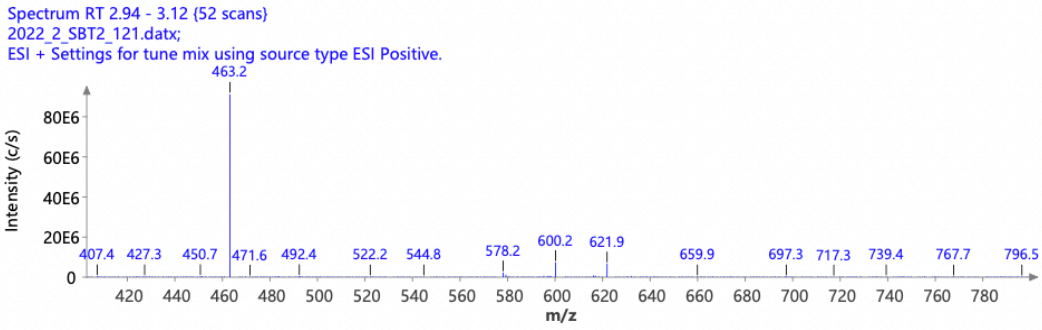
BB031

Fmoc-Leu-Leu(NMe)-D-Pro-OH

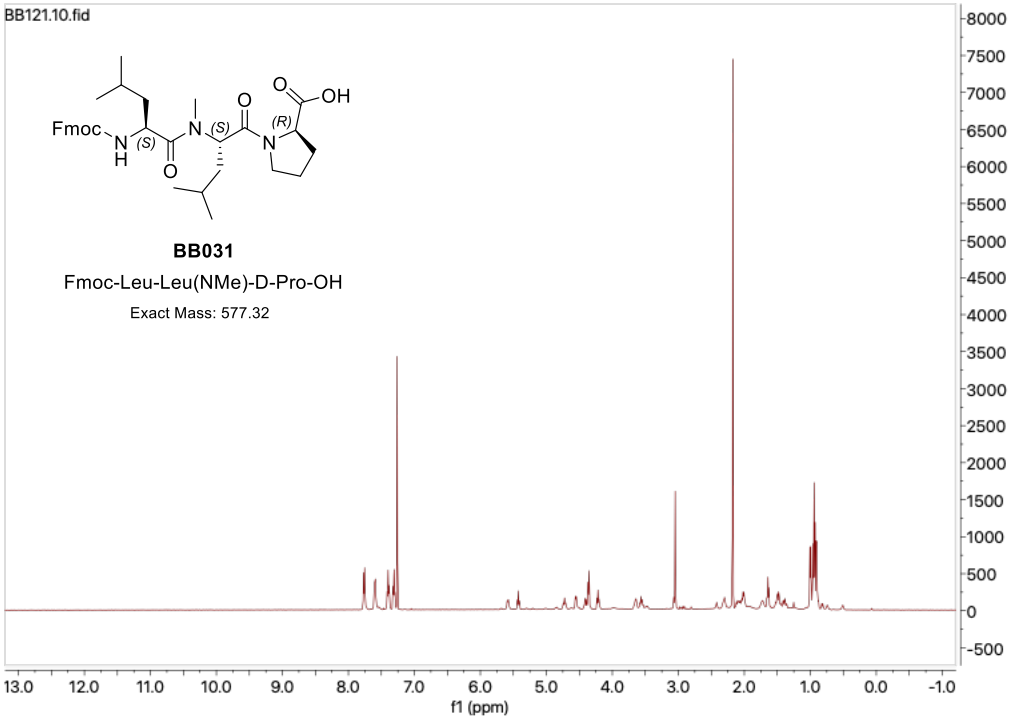
Exact Mass: 577.32





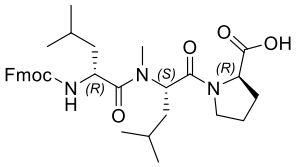


^{1}H-NMR of Building Block 031

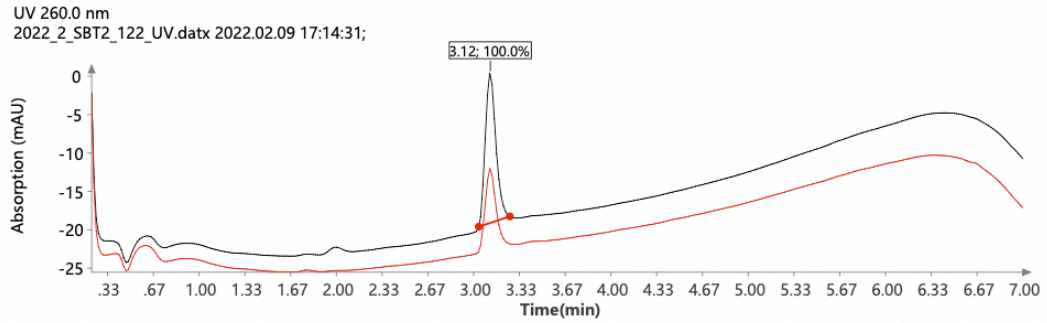


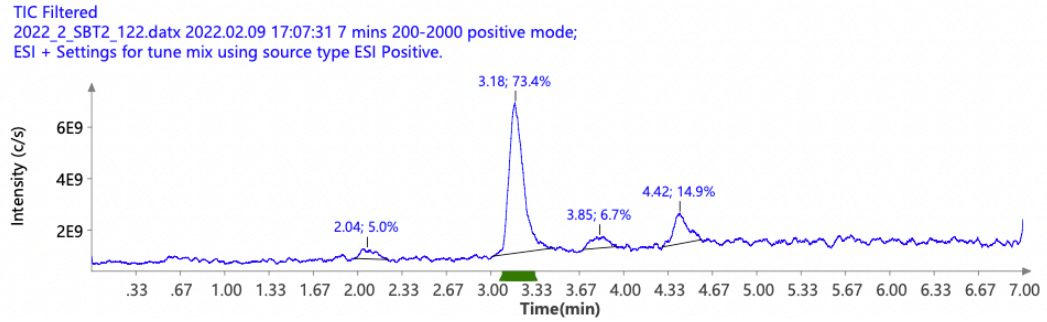
BB032

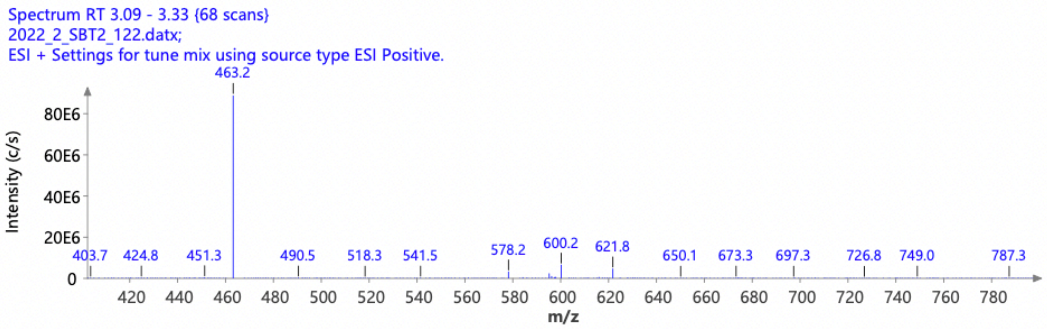
LCMS of Building Block 032



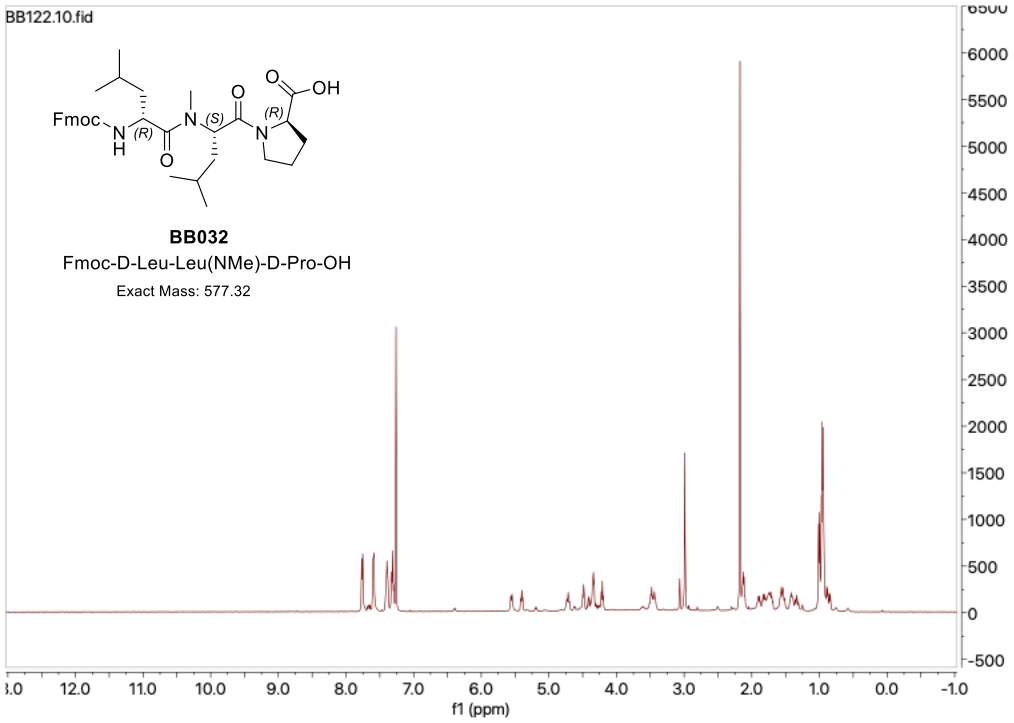
BB032
Fmoc-D-Leu-Leu(NMe)-D-Pro-OH
Exact Mass: 577.32





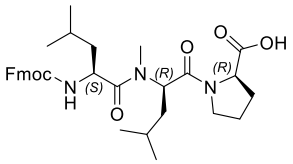


[1]H-NMR of Building Block 032



BB033

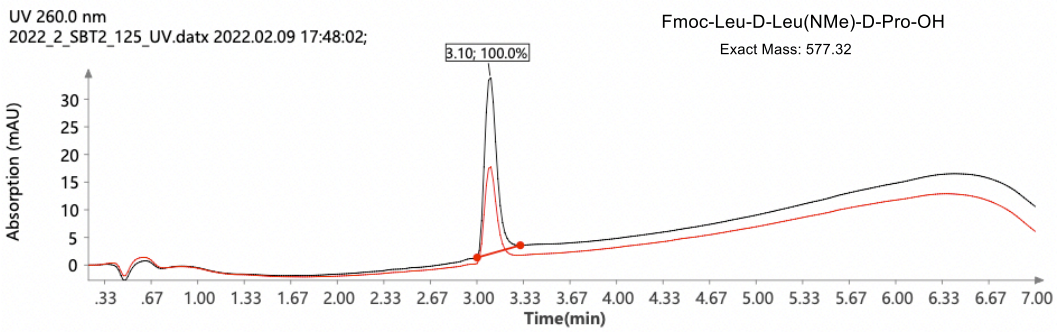
LCMS of Building Block 033

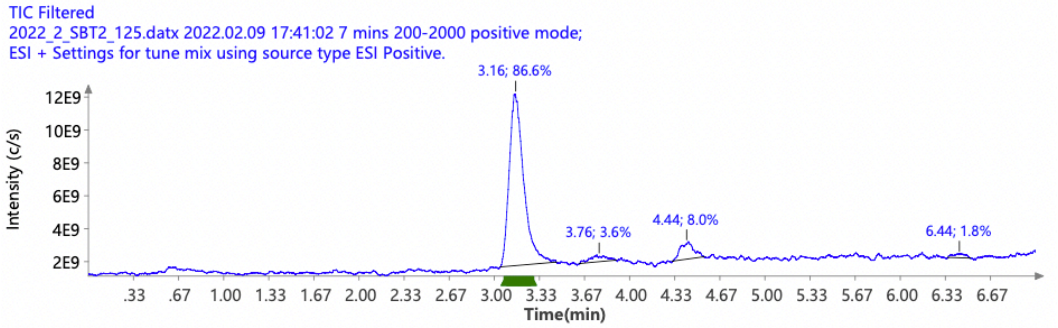


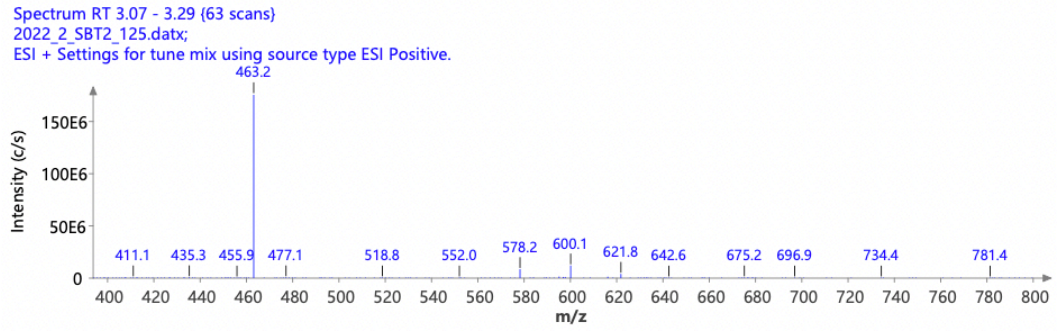
BB033

Fmoc-Leu-D-Leu(NMe)-D-Pro-OH

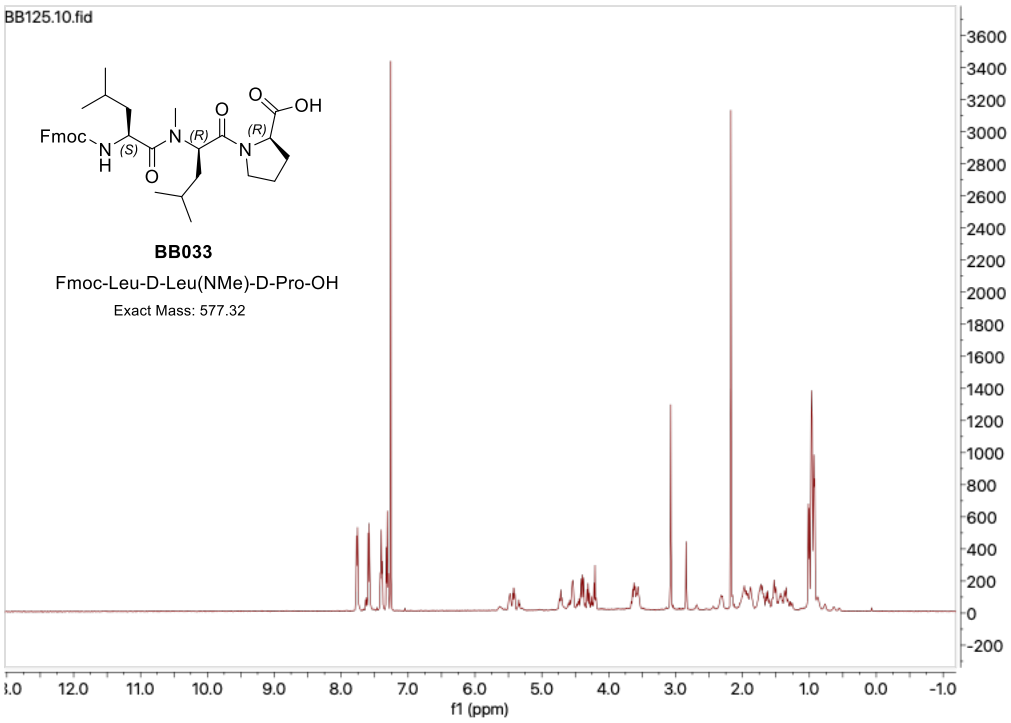
Exact Mass: 577.32





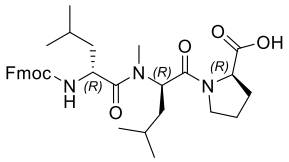


[1]H-NMR of Building Block 033



BB034

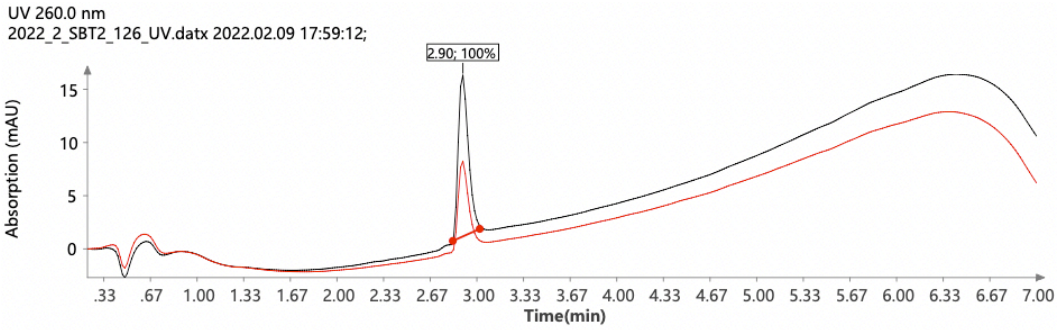
LCMS of Building Block 034

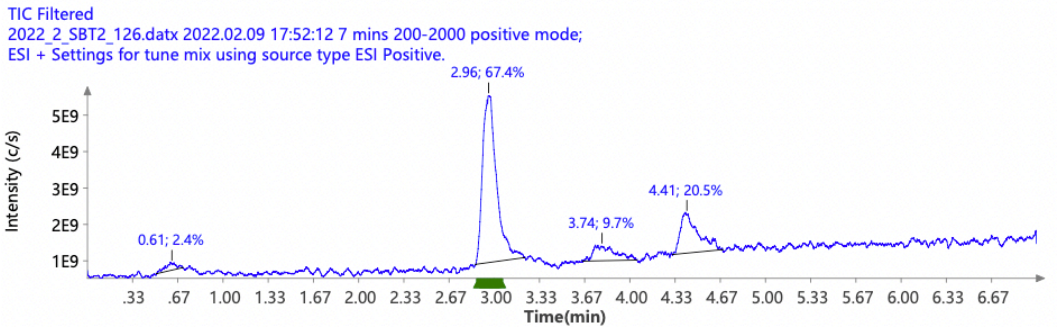


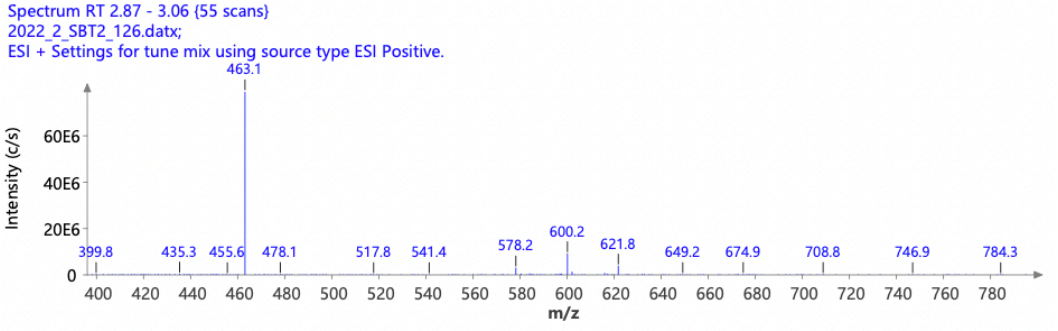
BB034

Fmoc-D-Leu-D-Leu(NMe)-D-Pro-OH

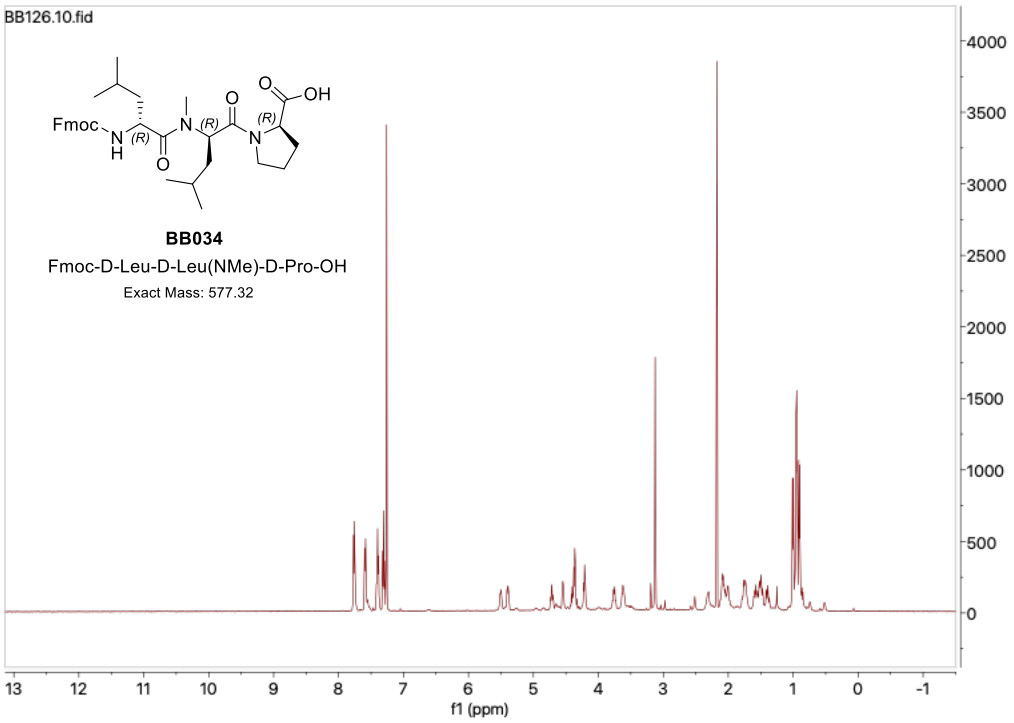
Exact Mass: 577.32





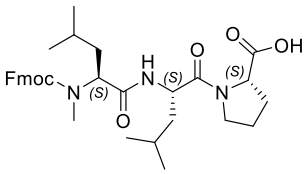


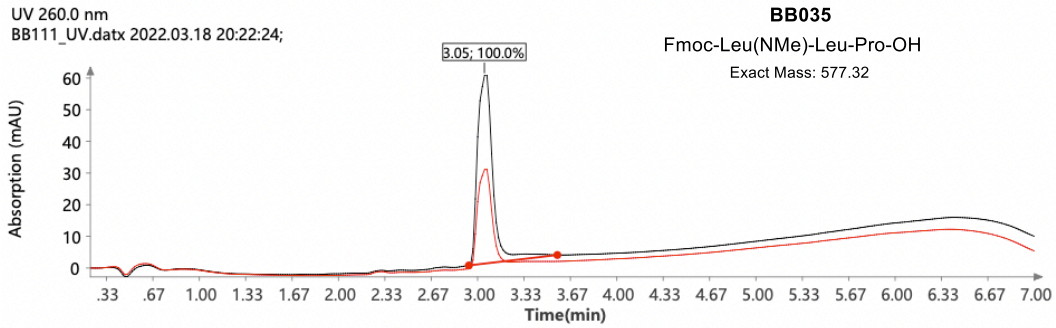
[1]H-NMR of Building Block 034

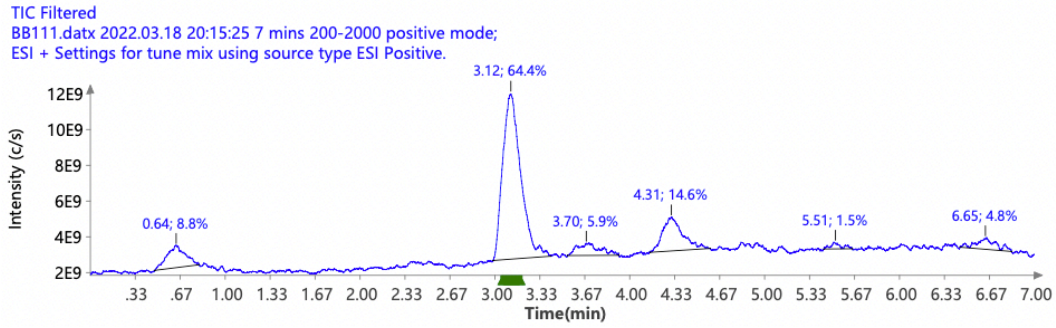


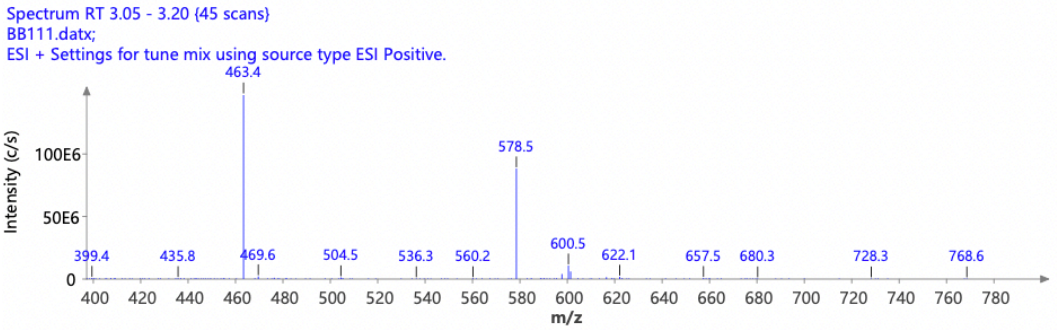
BB035

LCMS of Building Block 035

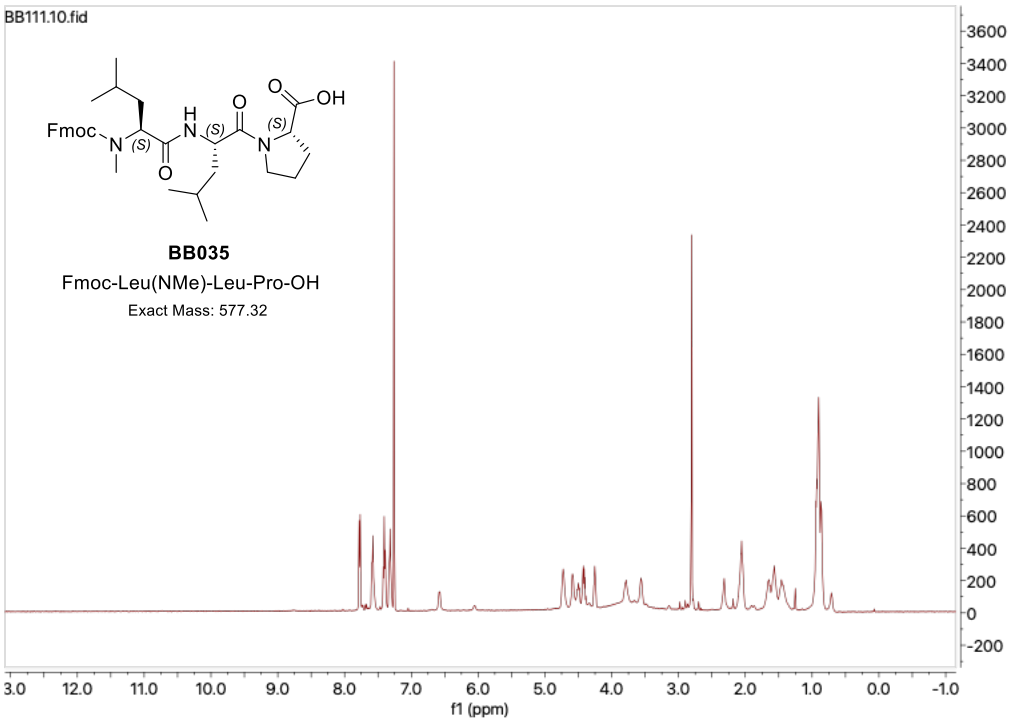






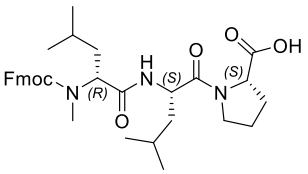


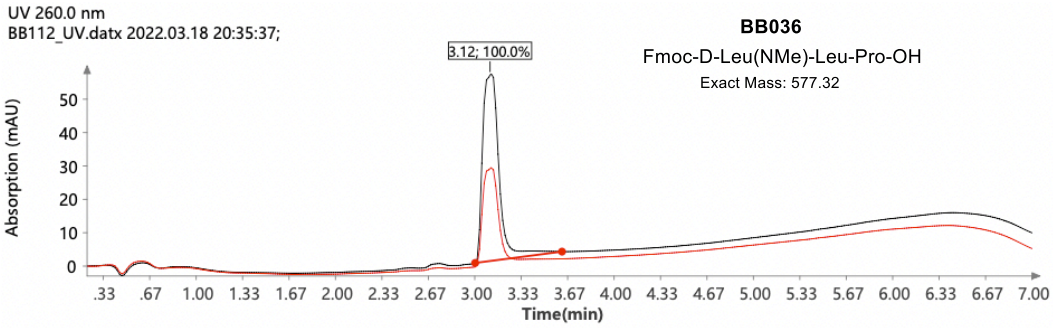
^{1}H-NMR of Building Block 035

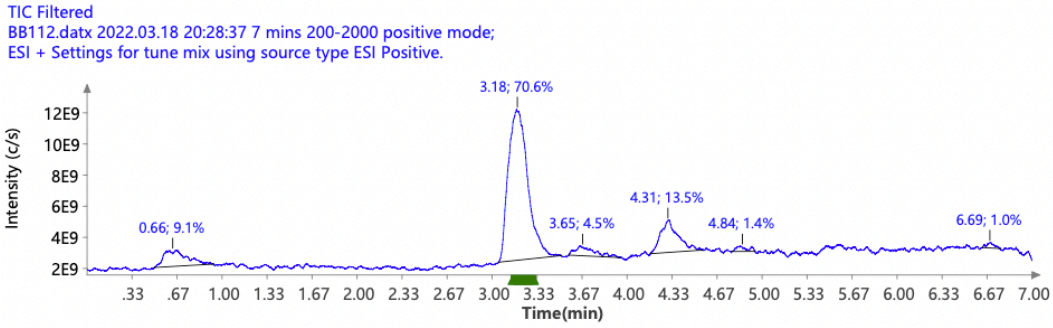


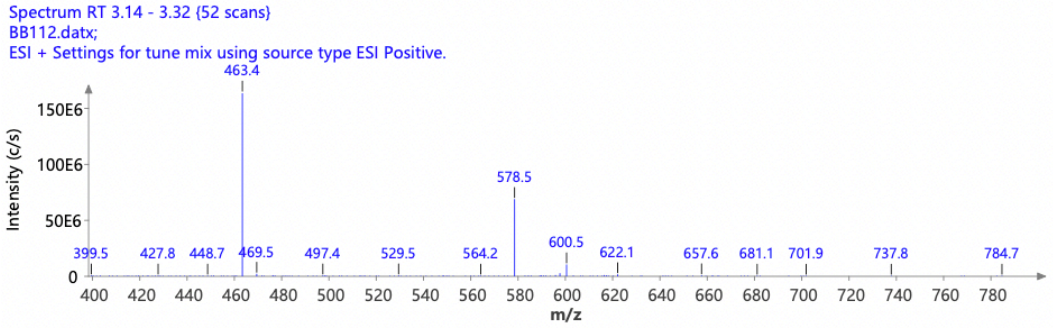
BB036

LCMS of Building Block 036

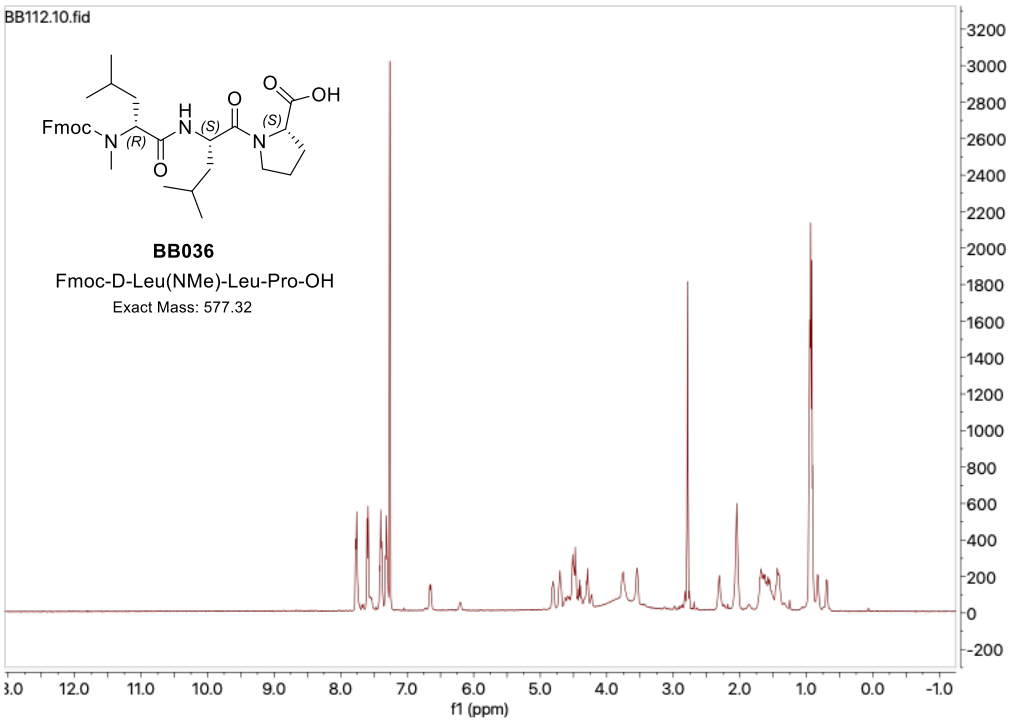






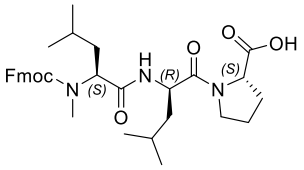


^{1}H-NMR of Building Block 036



BB037

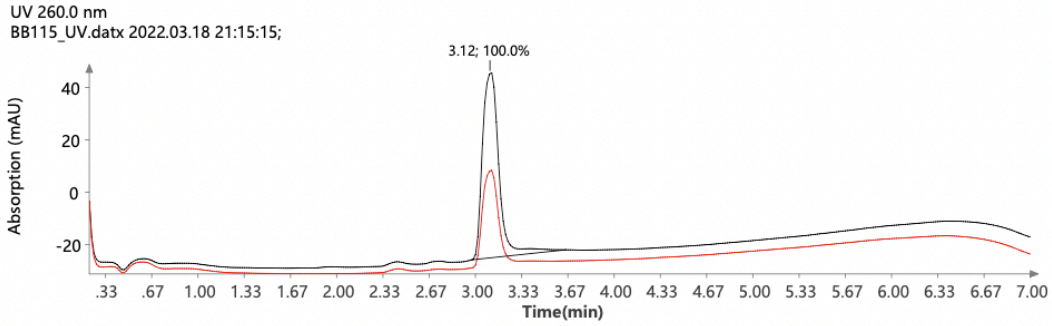
LCMS of Building Block 037

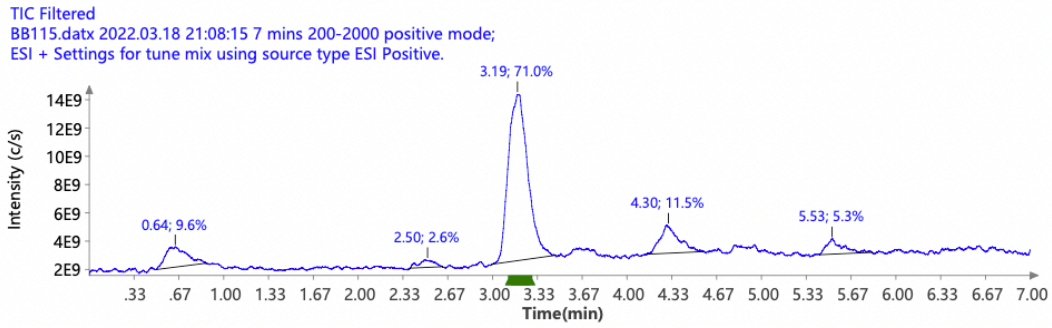


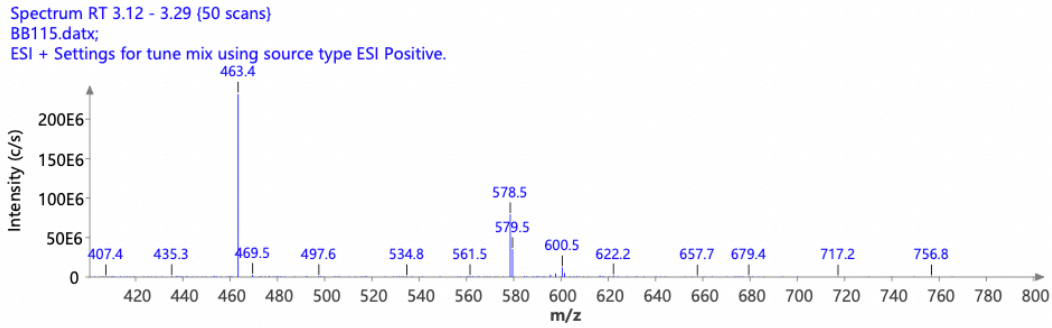
BB037

Fmoc-Leu(NMe)-D-Leu-Pro-OH

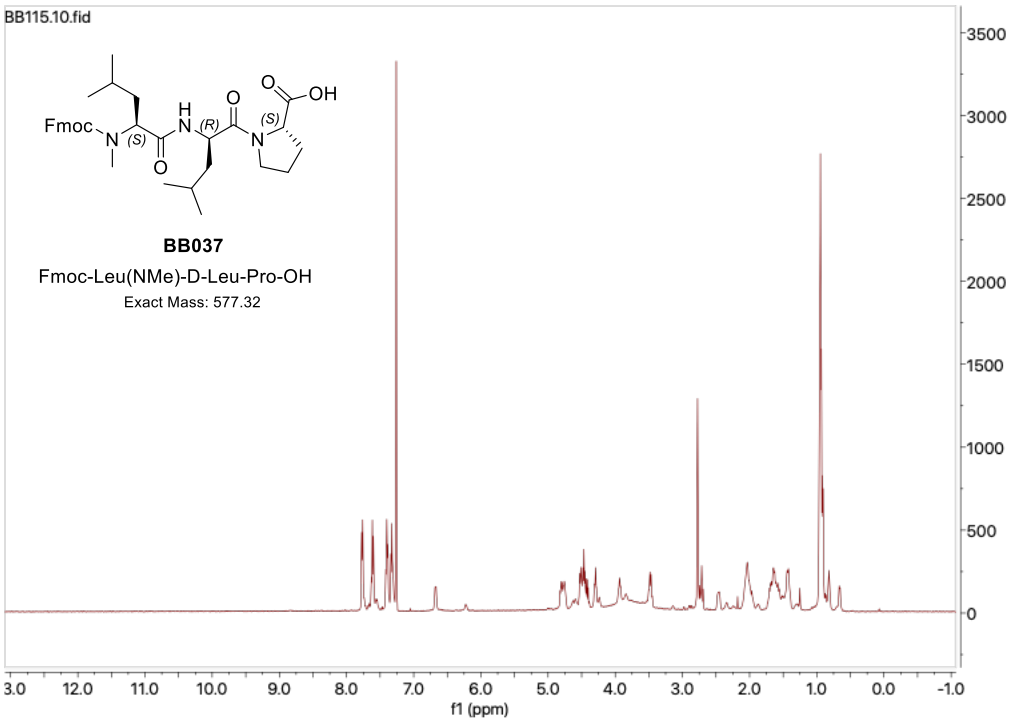
Exact Mass: 577.32





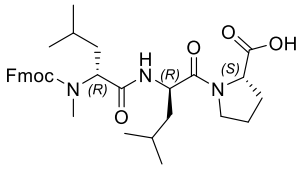


^{1}H-NMR of Building Block 037



BB038

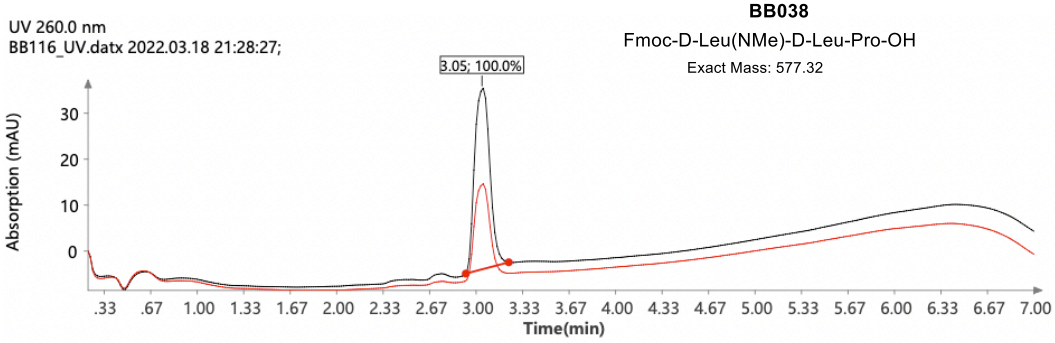
LCMS of Building Block 038

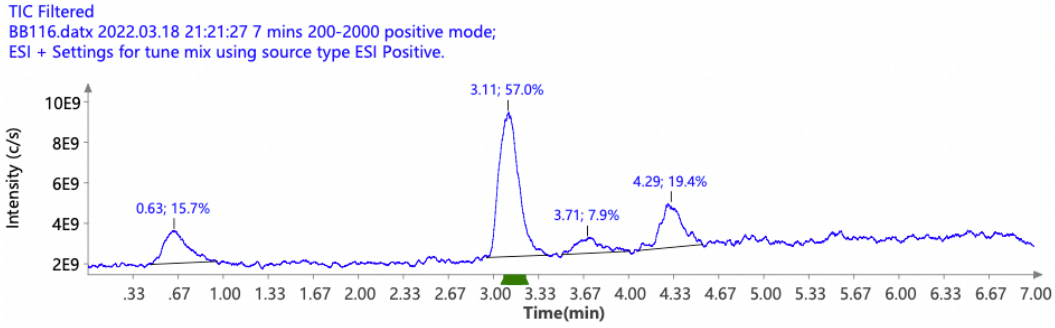


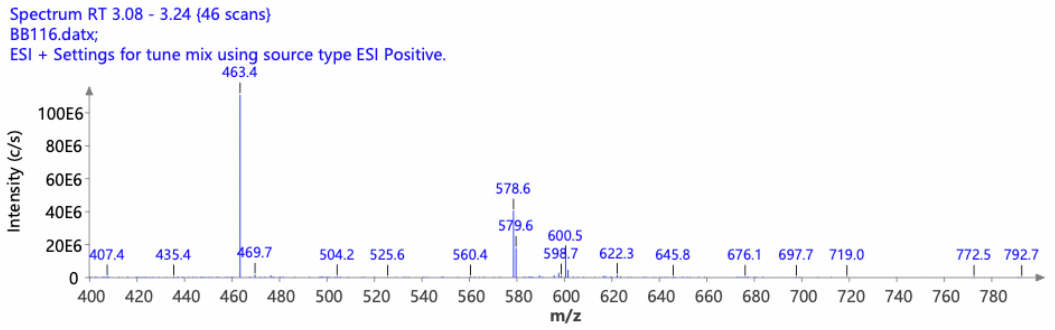
BB038

Fmoc-D-Leu(NMe)-D-Leu-Pro-OH

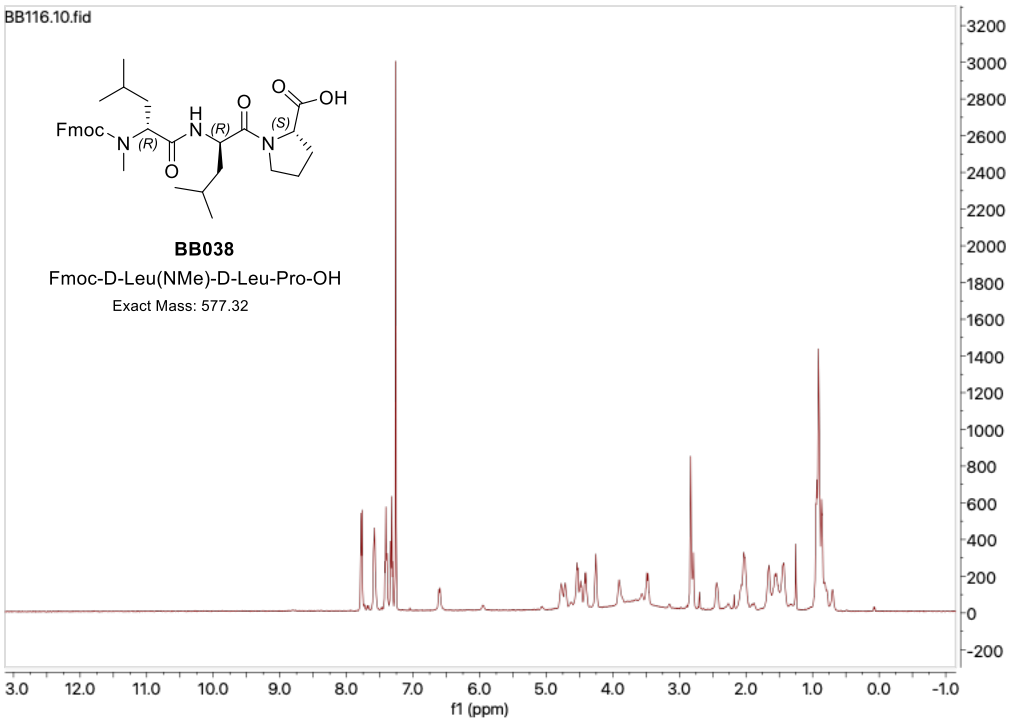
Exact Mass: 577.32





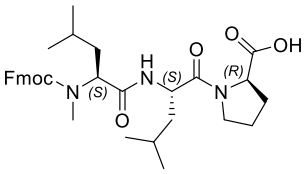


[1]H-NMR of Building Block 038

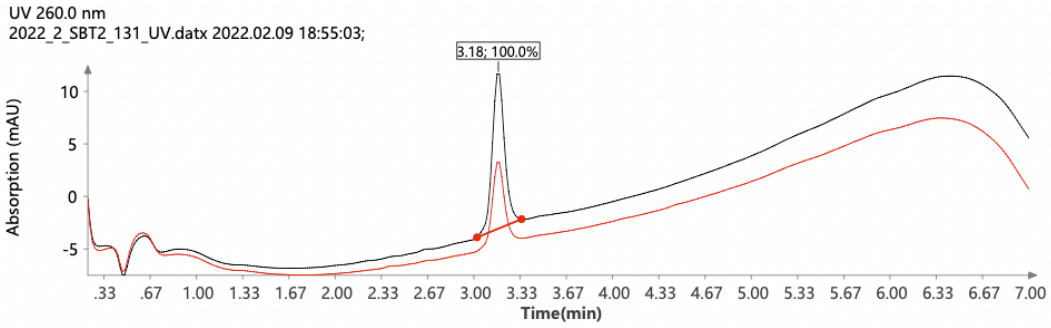


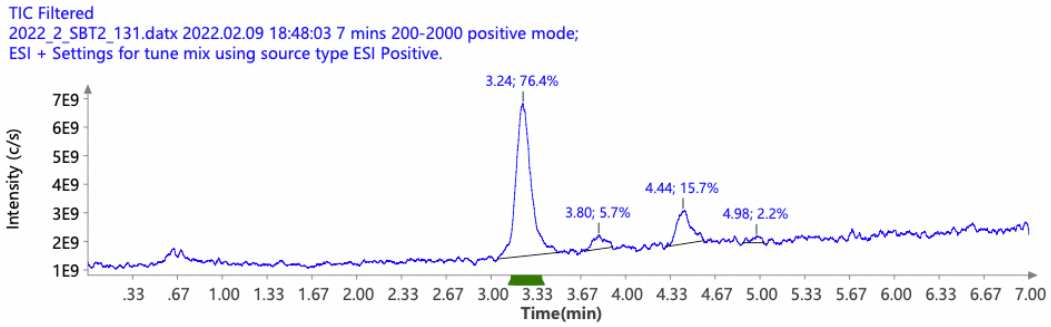
BB039

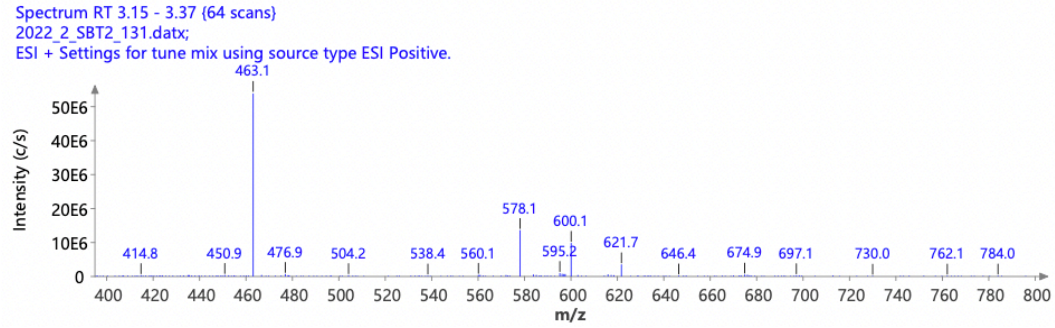
LCMS of Building Block 039



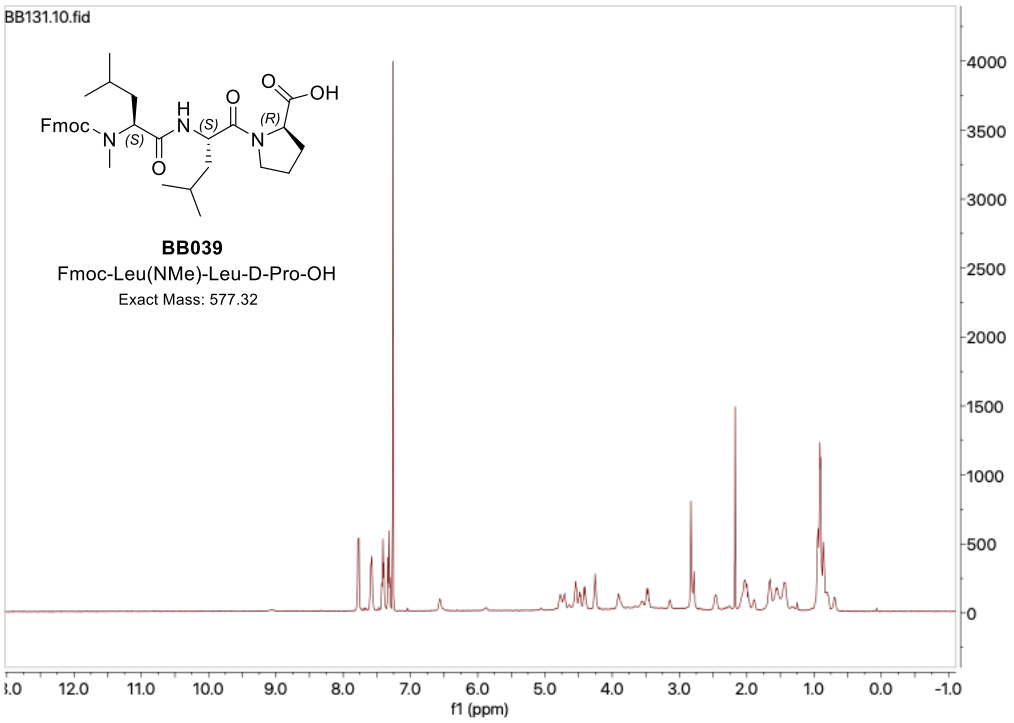
BB039
Fmoc-Leu(NMe)-Leu-D-Pro-OH
Exact Mass: 577.32





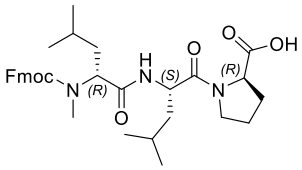


^{1}H-NMR of Building Block 039

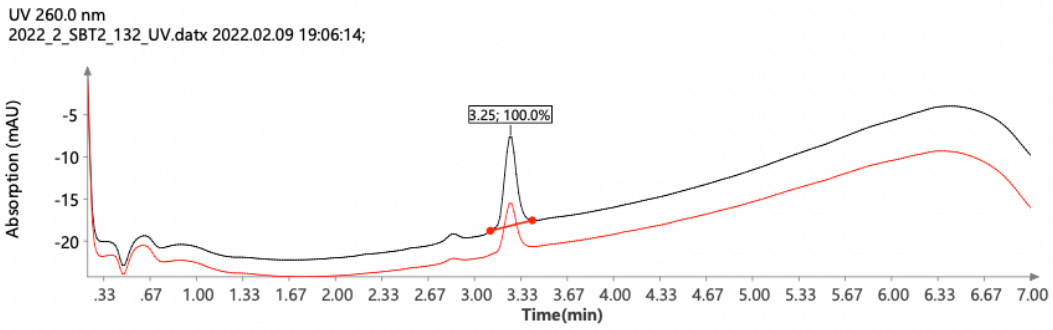


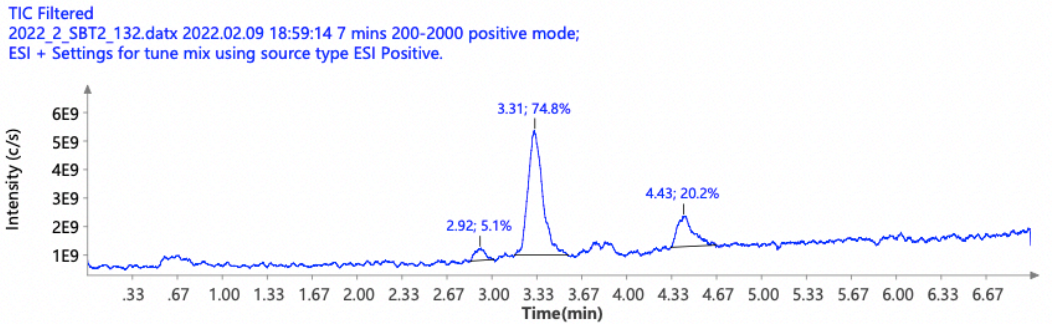
BB040

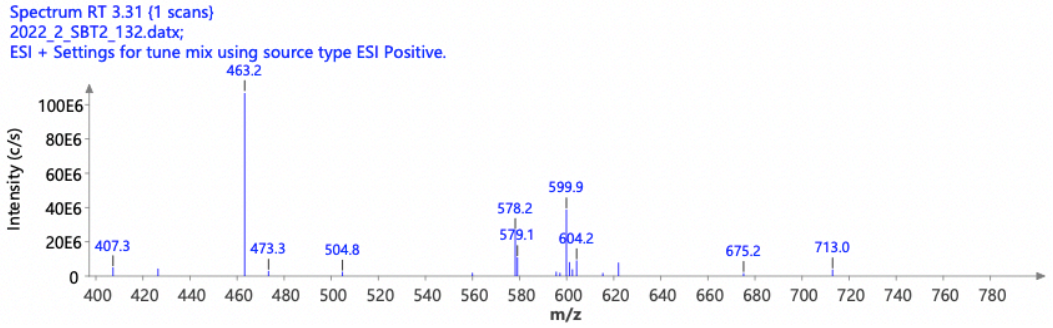
LCMS of Building Block 040



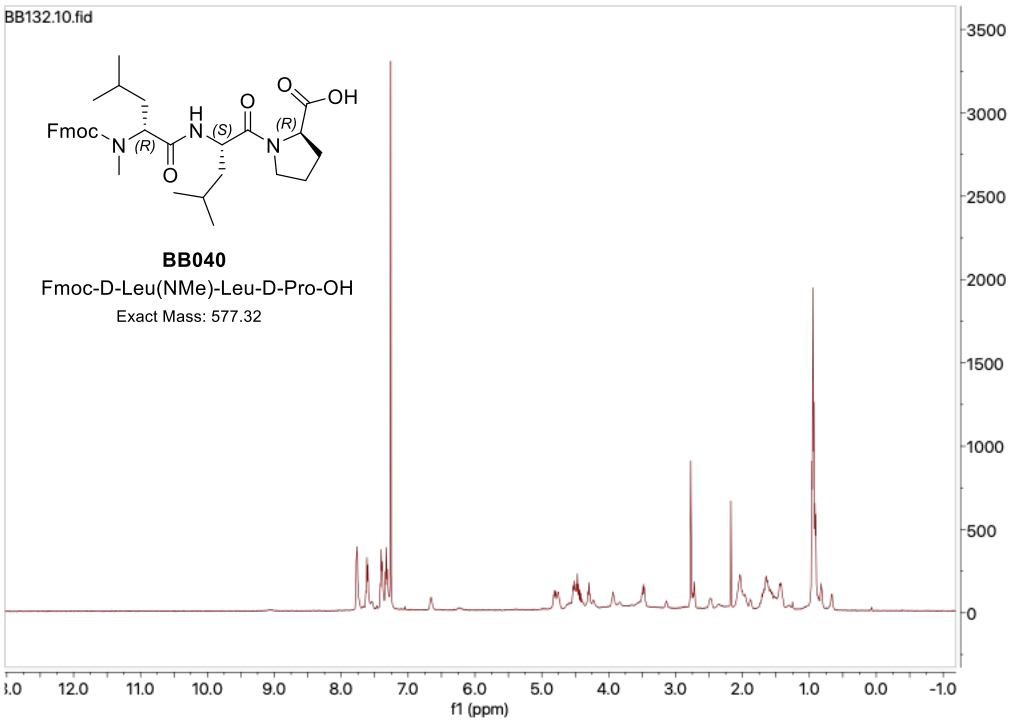
BB040
Fmoc-D-Leu(NMe)-Leu-D-Pro-OH
Exact Mass: 577.32





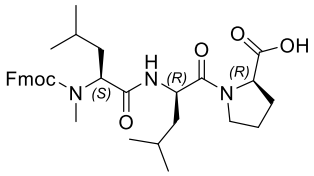


^{1}H-NMR of Building Block 040

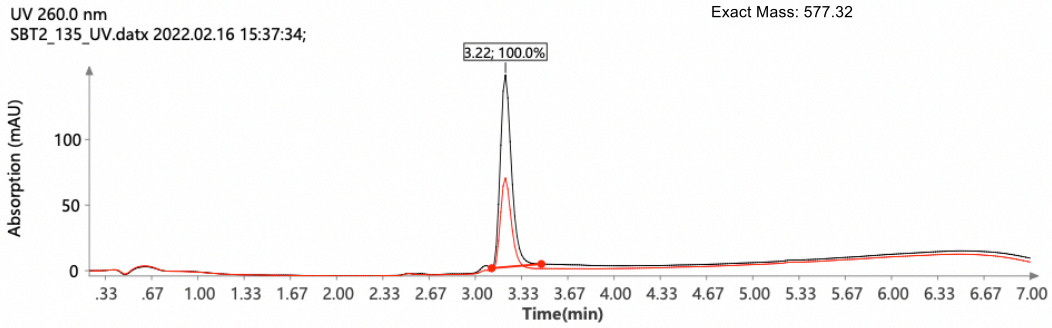


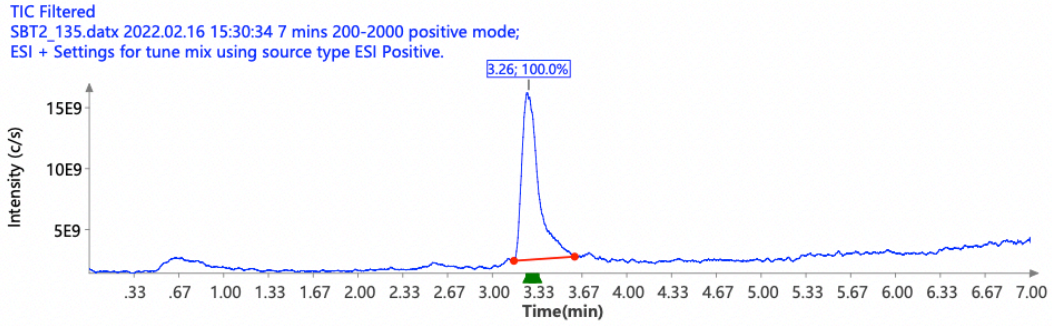
BB041

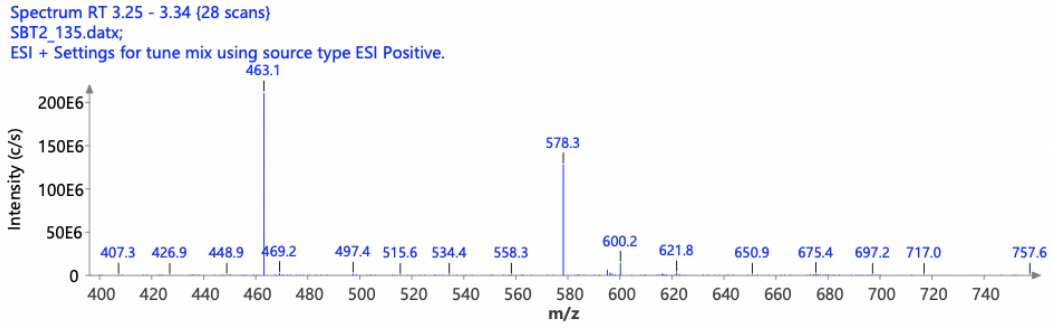
LCMS of Building Block 041



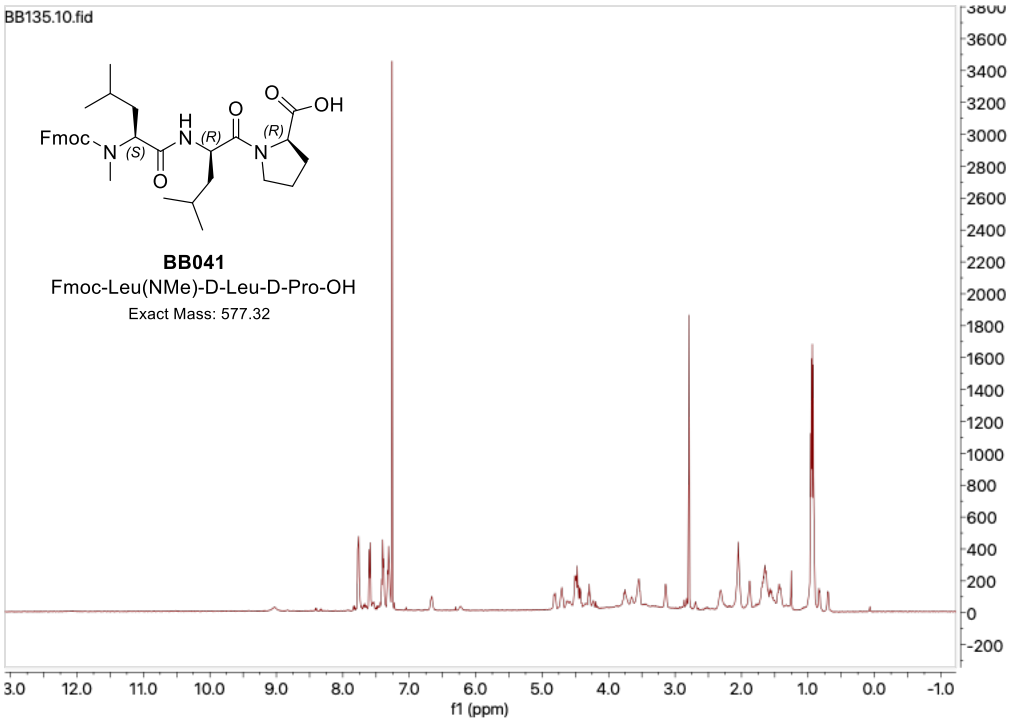
BB041
Fmoc-Leu(NMe)-D-Leu-D-Pro-OH
Exact Mass: 577.32





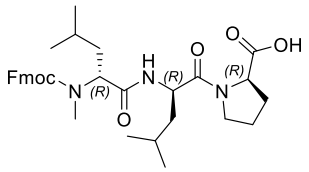


^{1}H-NMR of Building Block 041

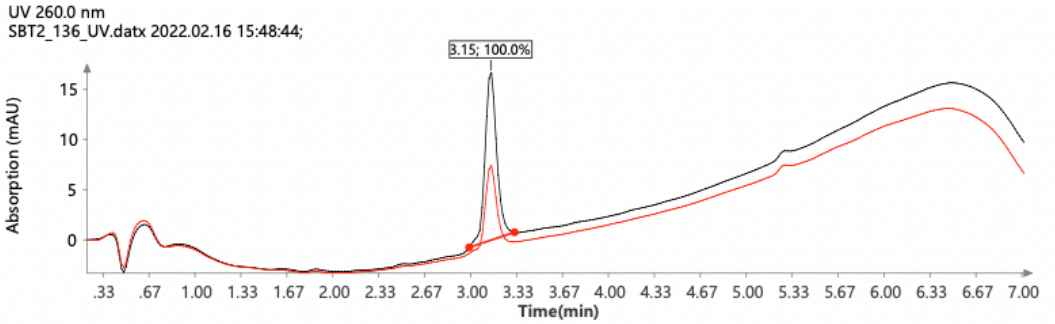


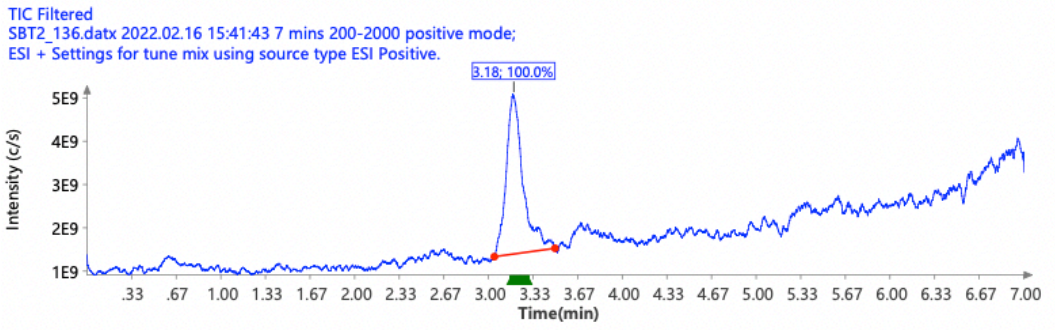
BB042

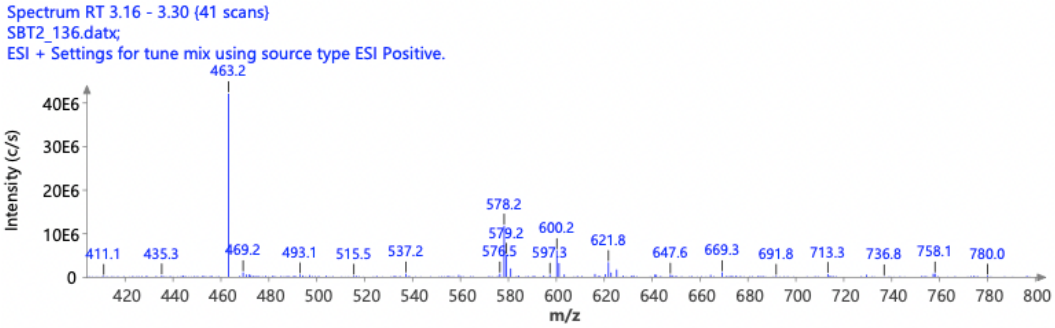
LCMS of Building Block 042



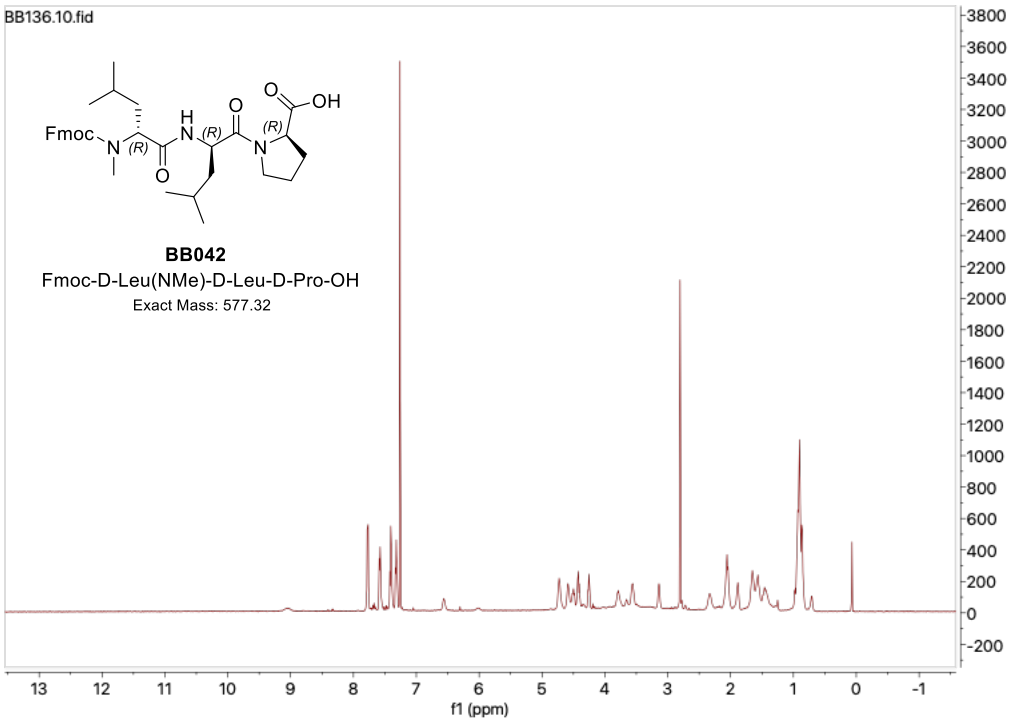
BB042
Fmoc-D-Leu(NMe)-D-Leu-D-Pro-OH
Exact Mass: 577.32





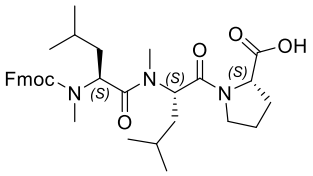


[1]H-NMR of Building Block 042



BB043

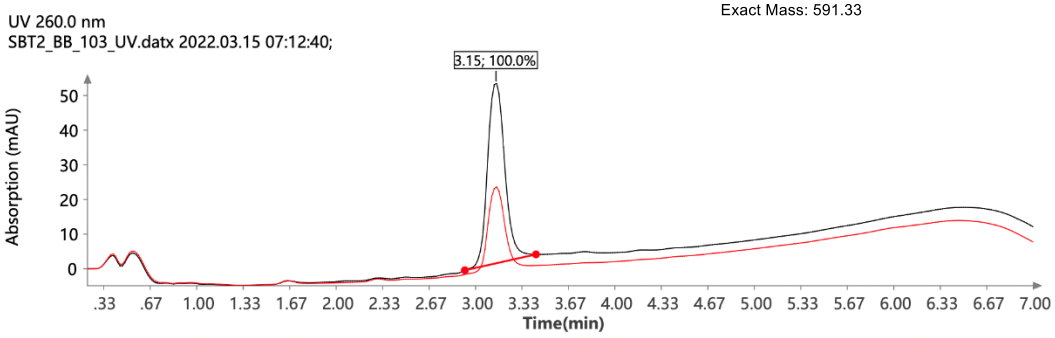
LCMS of Building Block 043

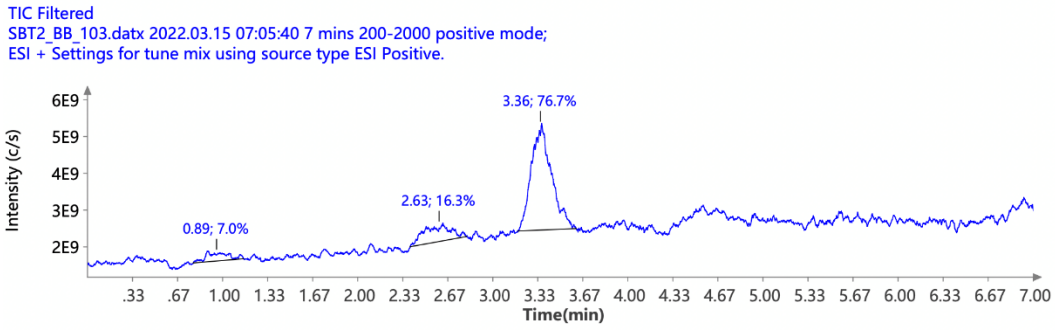


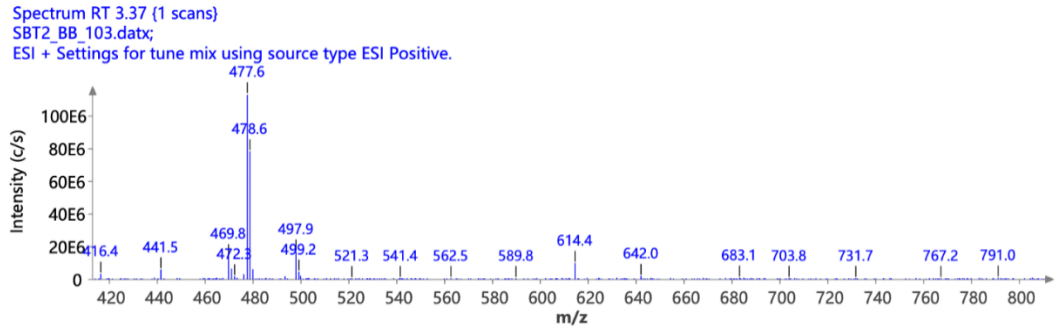
BB043

Fmoc-Leu(NMe)-Leu(NMe)-Pro-OH

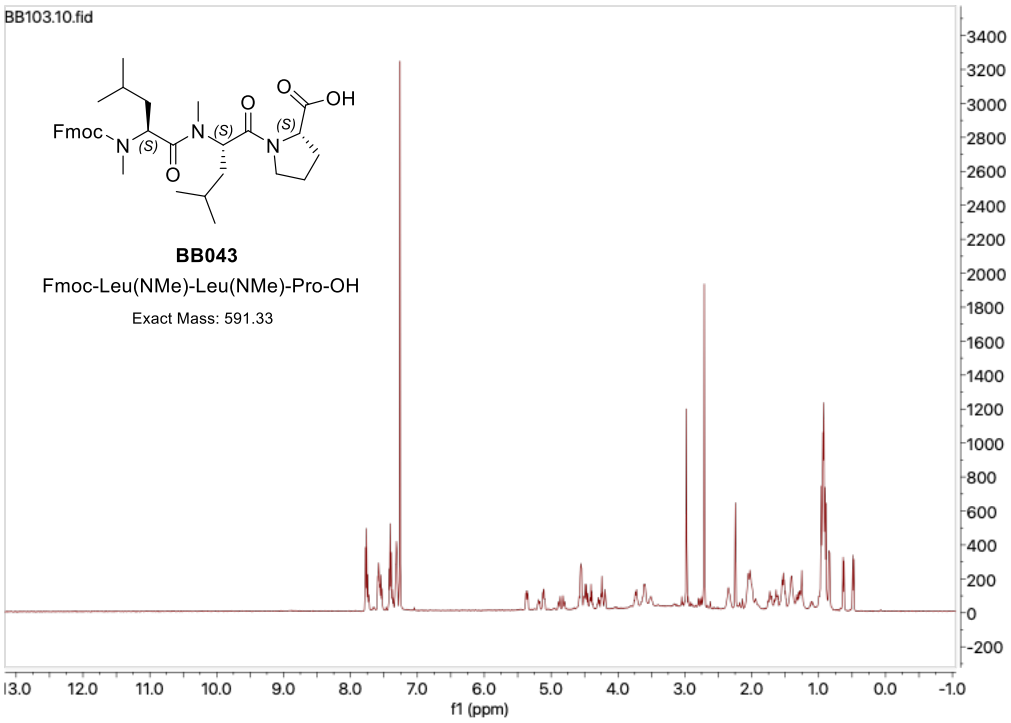
Exact Mass: 591.33





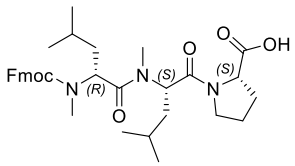


^{1}H-NMR of Building Block 043



BB044

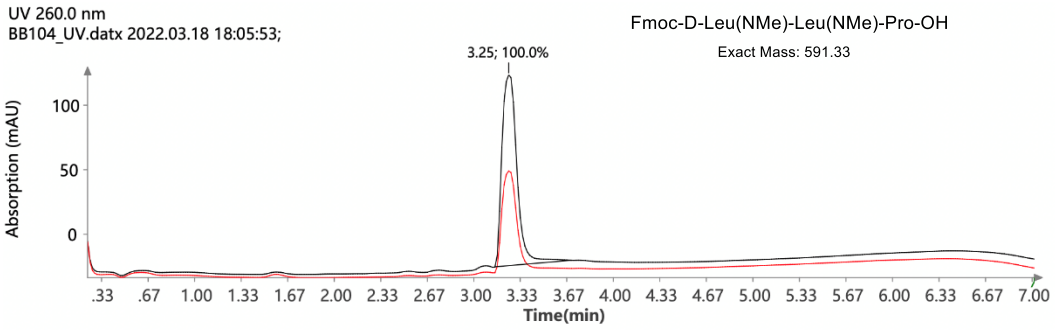
LCMS of Building Block 044

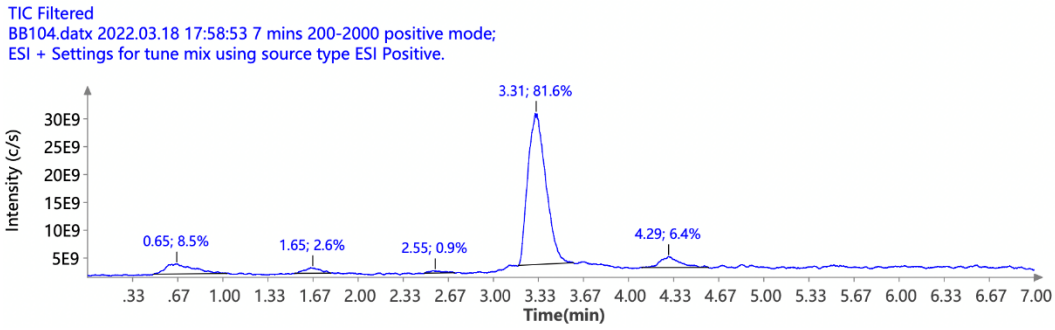


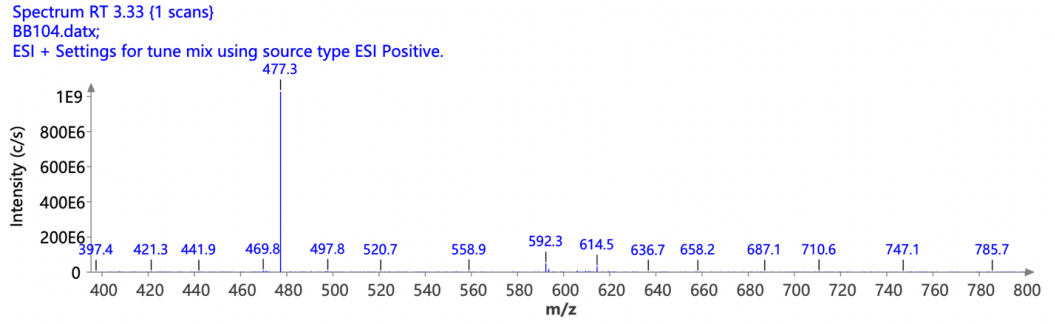
BB044

Fmoc-D-Leu(NMe)-Leu(NMe)-Pro-OH

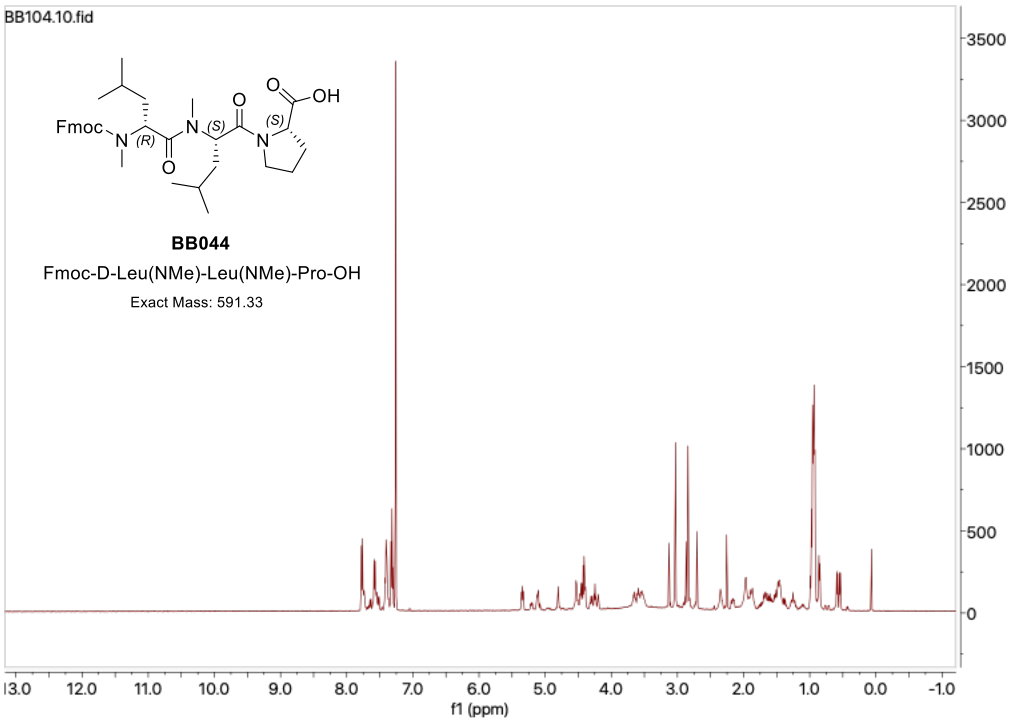
Exact Mass: 591.33





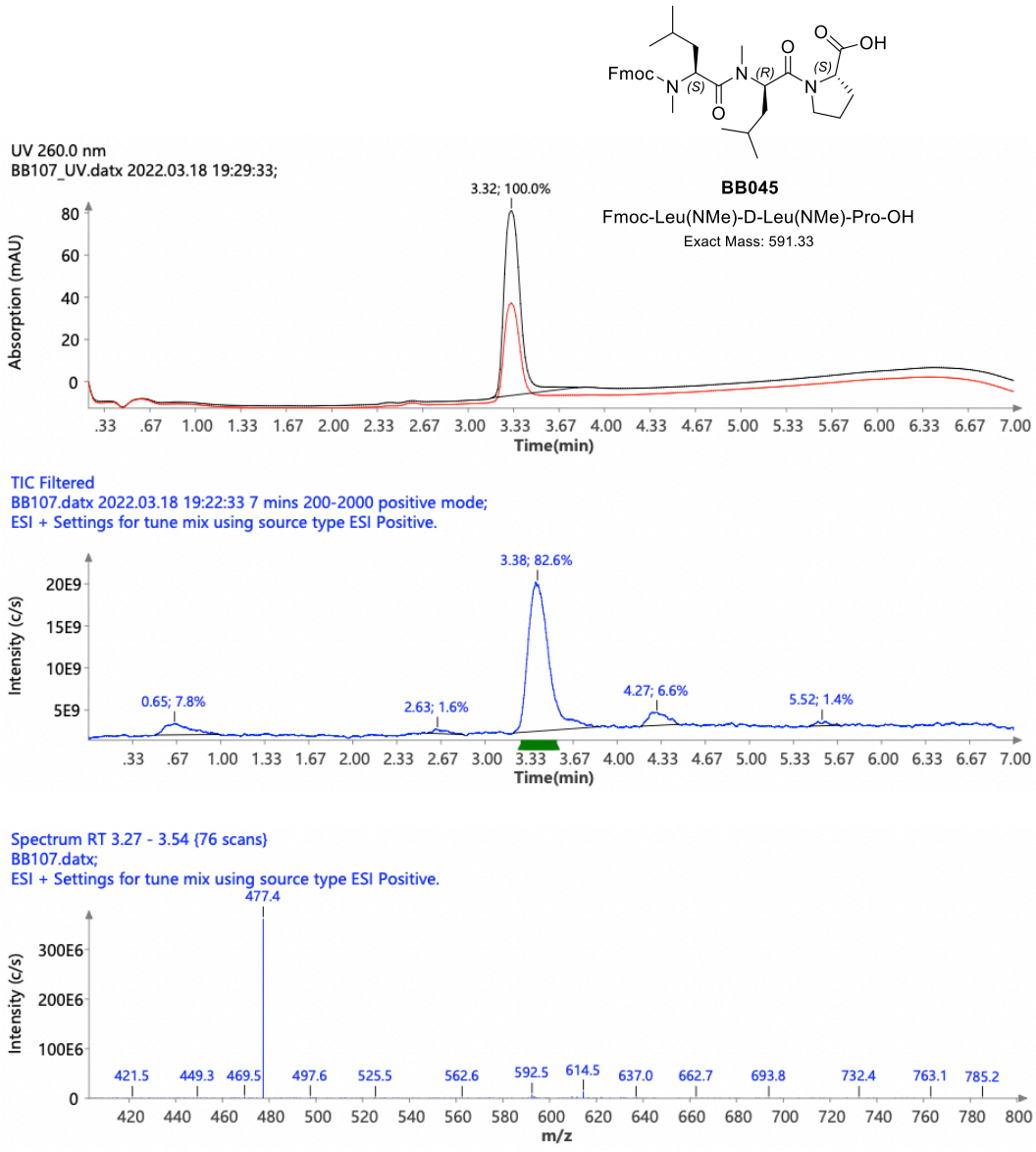


^{1}H-NMR of Building Block 044

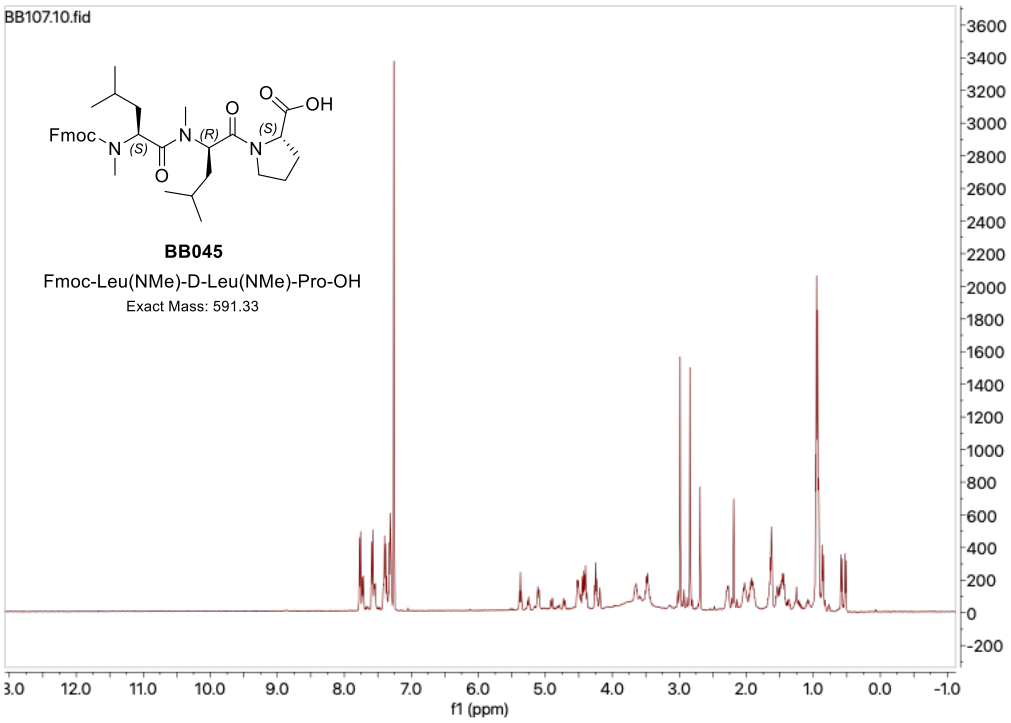


BB045

LCMS of Building Block 045

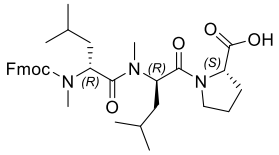


[1]H-NMR of Building Block 045



BB046

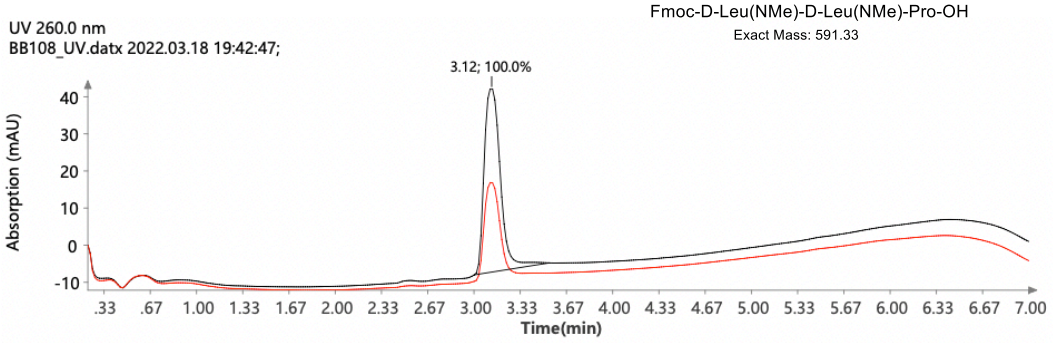
LCMS of Building Block 046

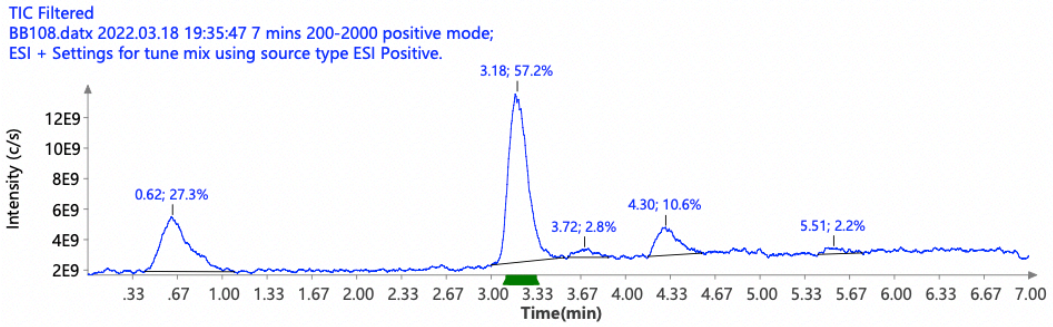


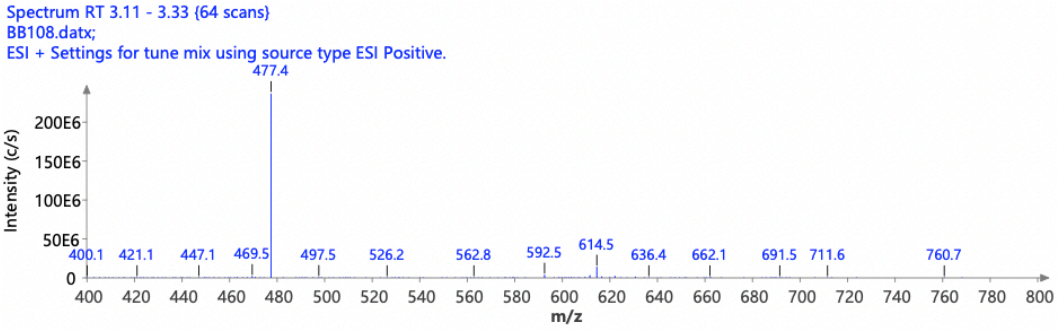
BB046

Fmoc-D-Leu(NMe)-D-Leu(NMe)-Pro-OH

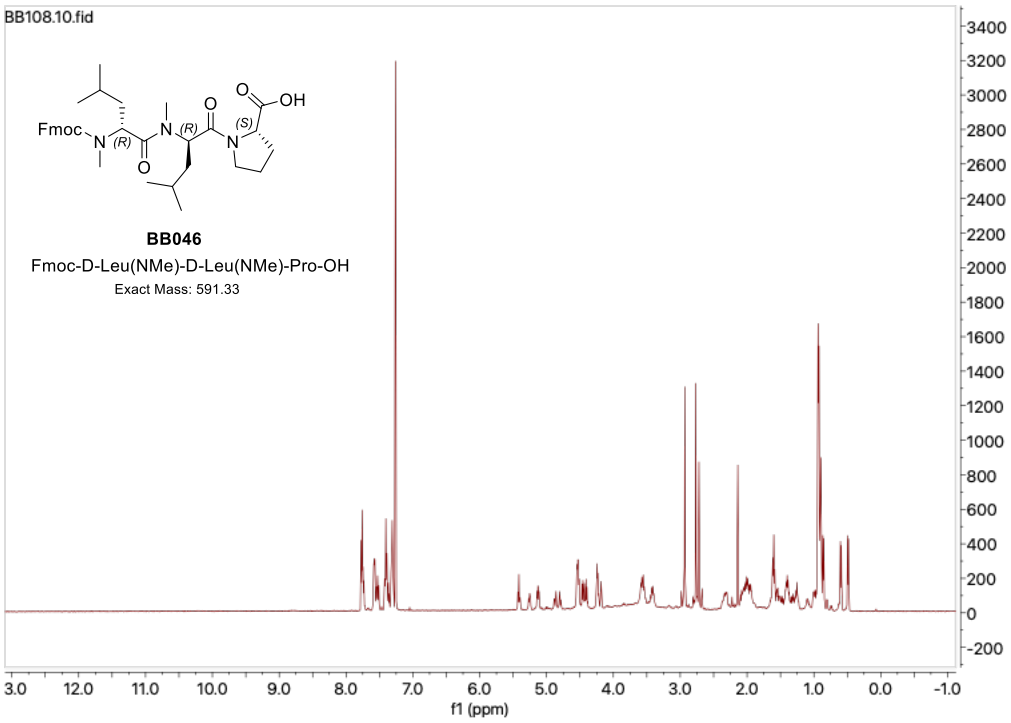
Exact Mass: 591.33





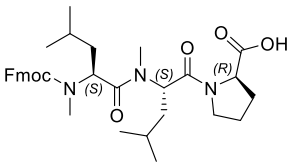


^{1}H-NMR of Building Block 046

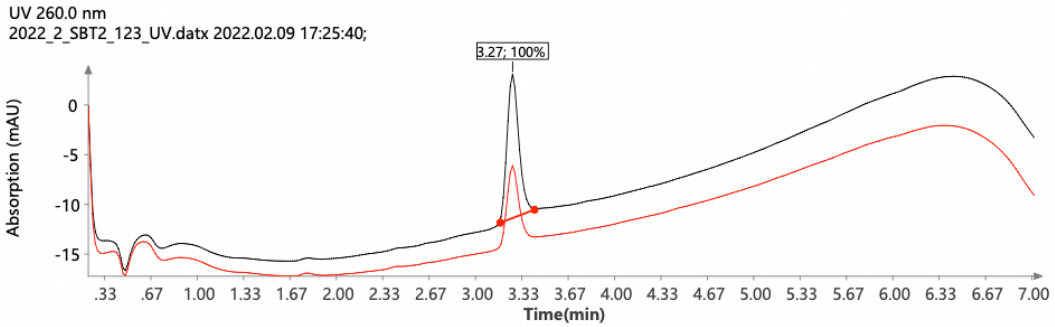


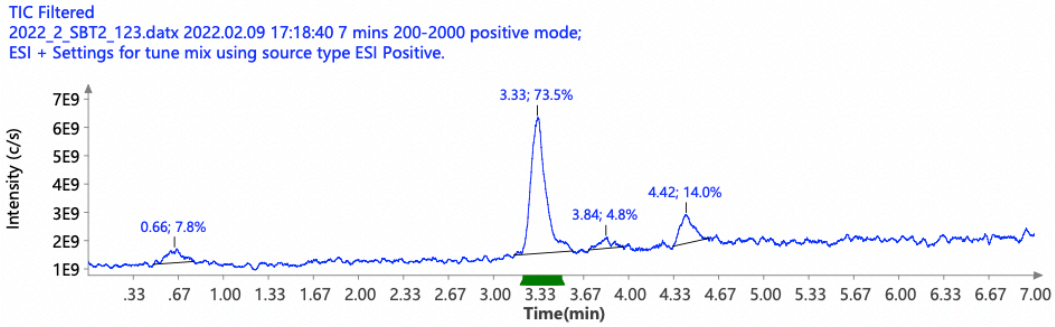
BB047

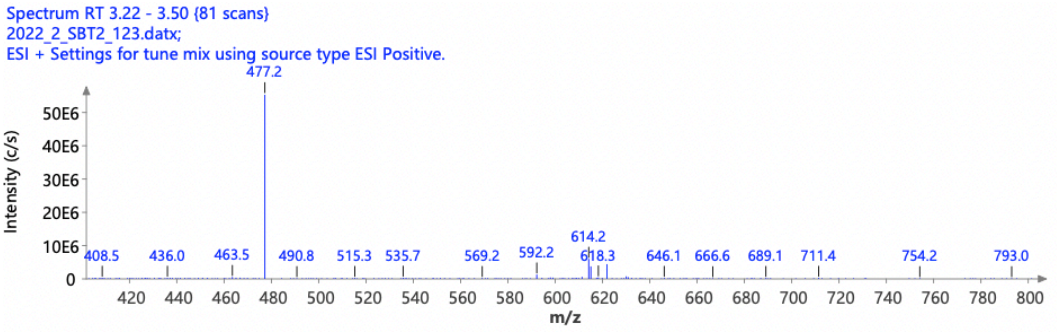
LCMS of Building Block 047



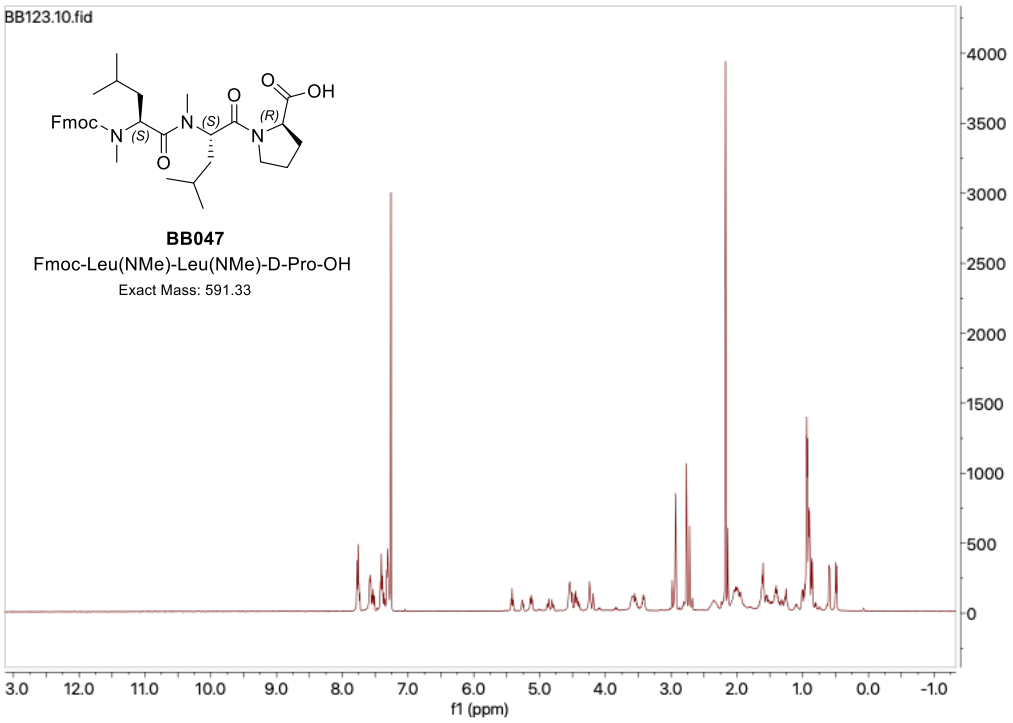
BB047
Fmoc-Leu(NMe)-Leu(NMe)-D-Pro-OH
Exact Mass: 591.33





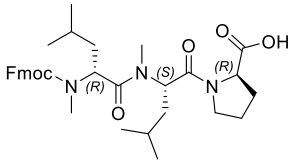


^{1}H-NMR of Building Block 047



BB048

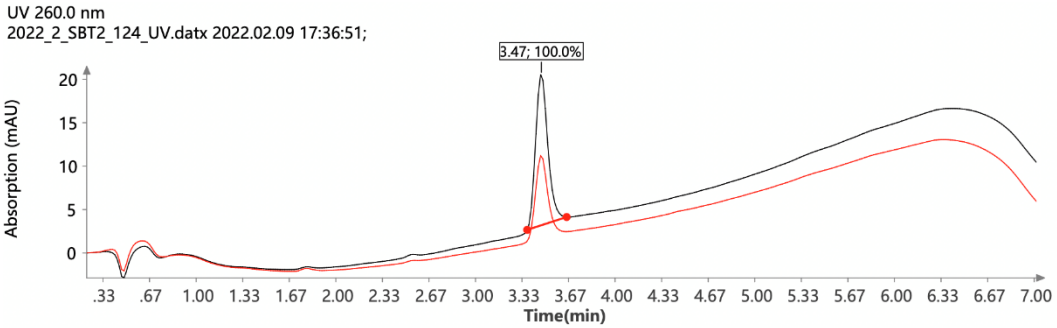
LCMS of Building Block 048

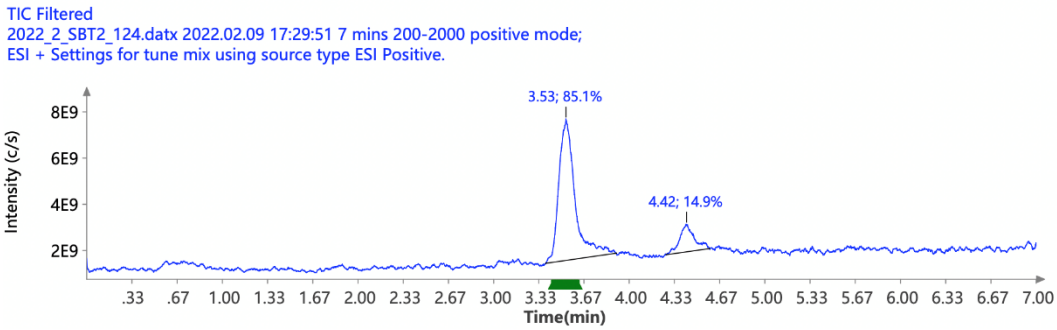


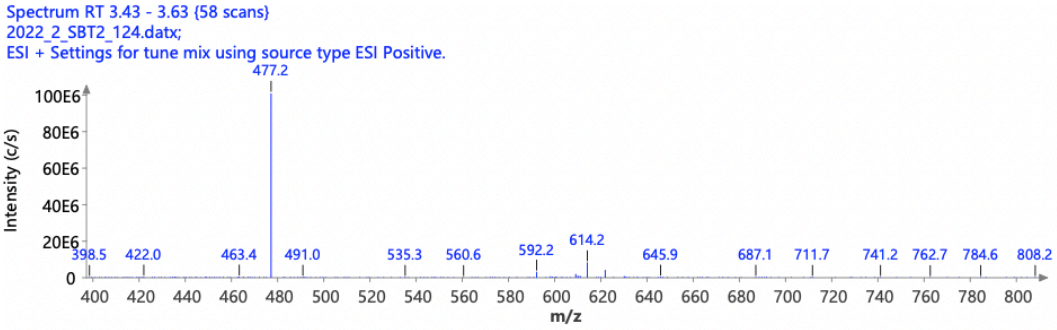
BB048

Fmoc-D-Leu(NMe)-Leu(NMe)-D-Pro-OH

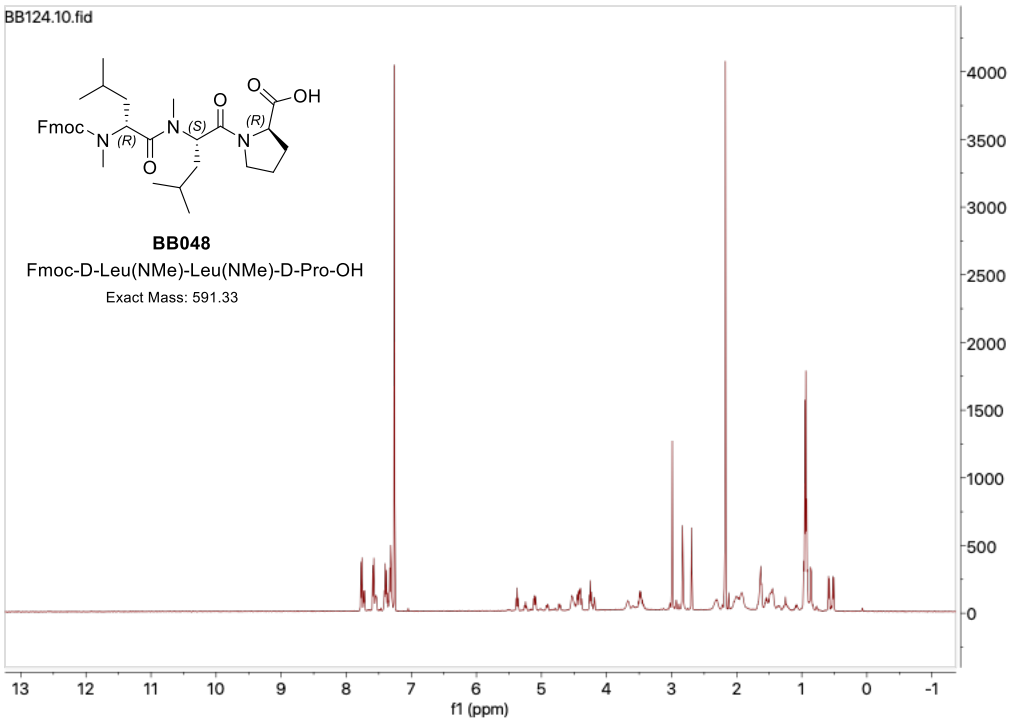
Exact Mass: 591.33





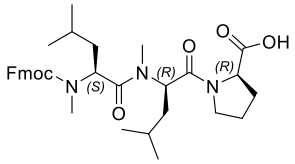


[1]H-NMR of Building Block 048



BB049

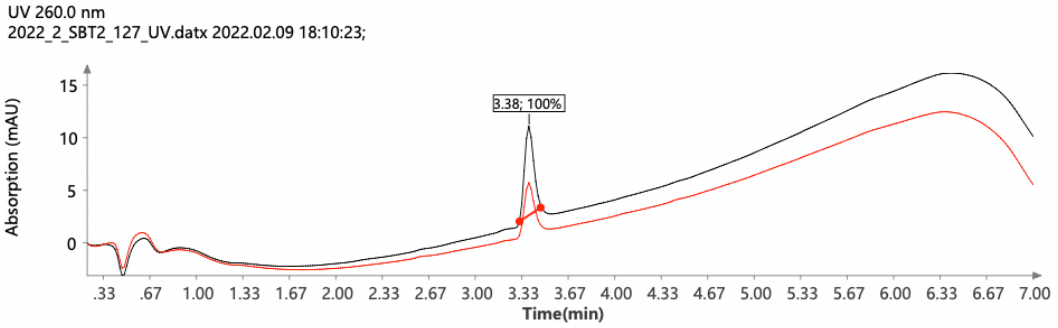
LCMS of Building Block 049

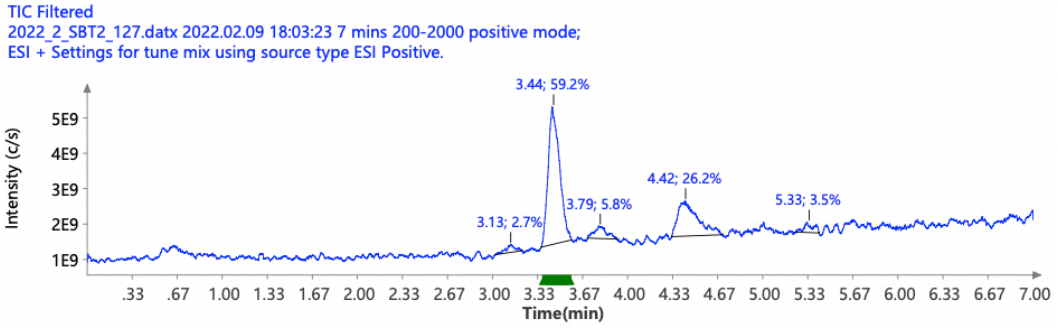


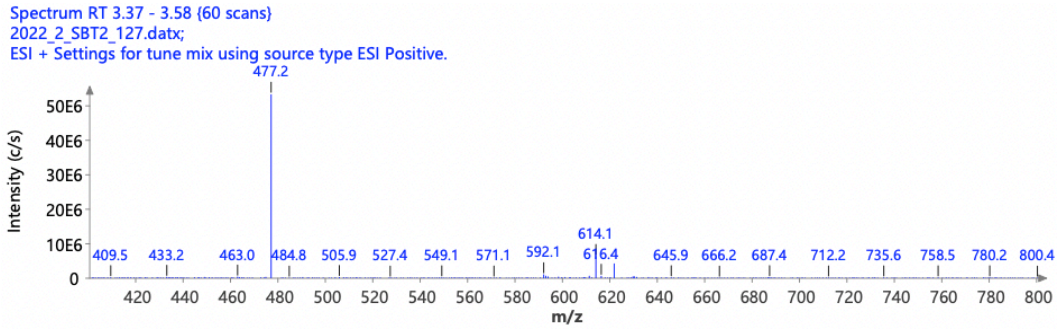
BB049

Fmoc-Leu(NMe)-D-Leu(NMe)-D-Pro-OH

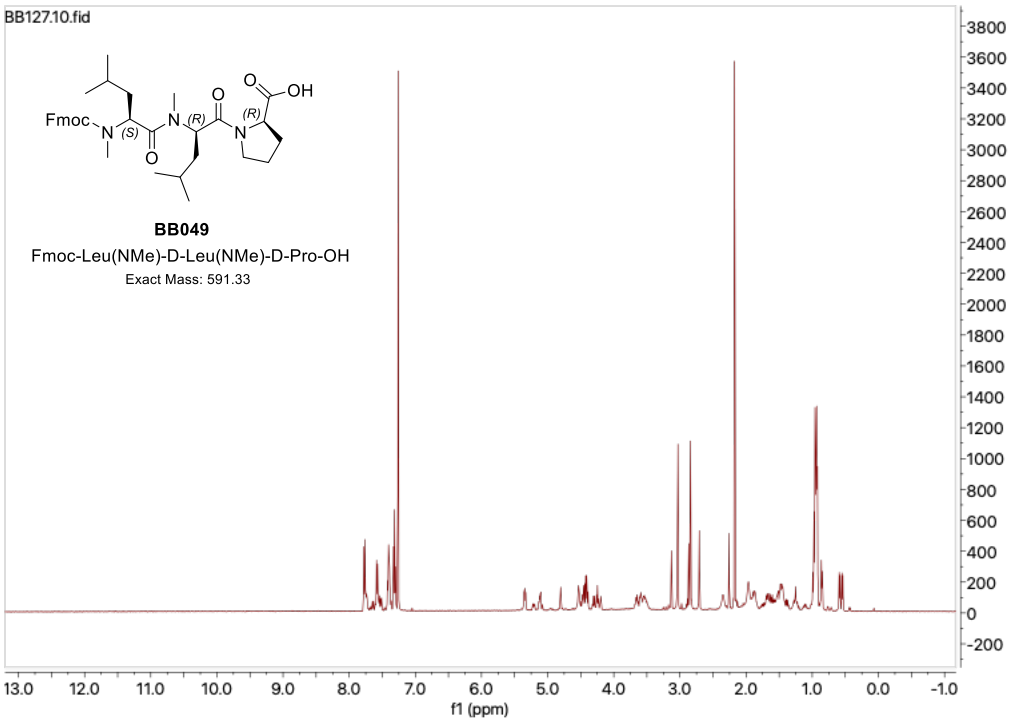
Exact Mass: 591.33





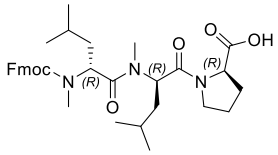


[1]H-NMR of Building Block 049

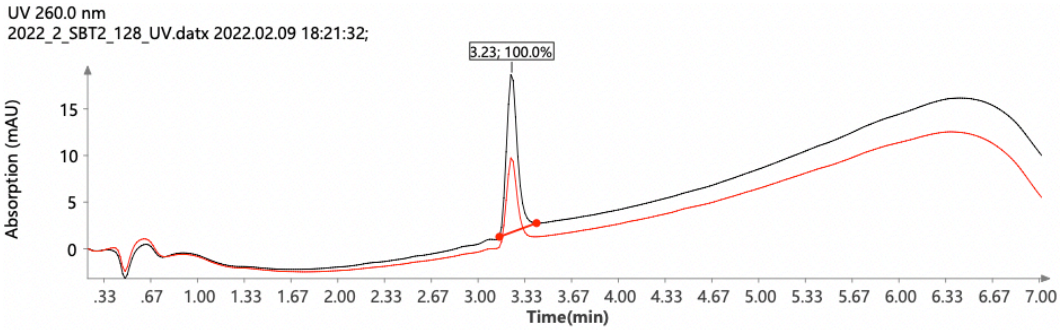


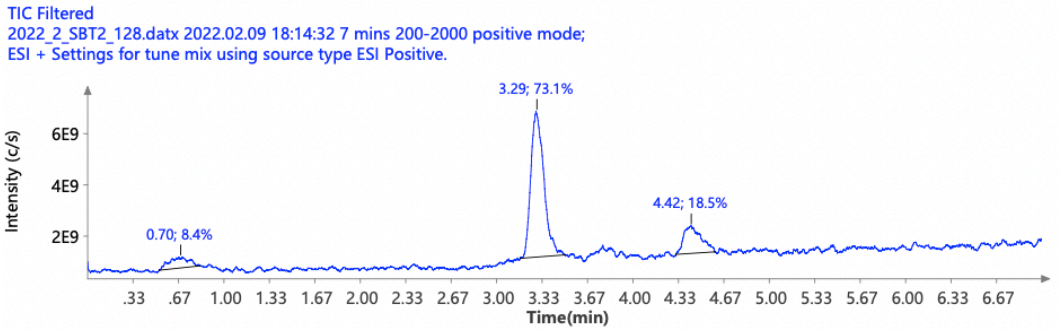
BB050

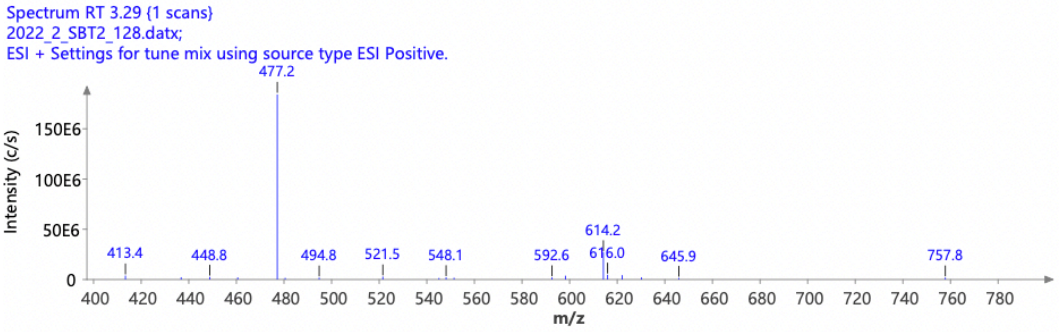
LCMS of Building Block 050



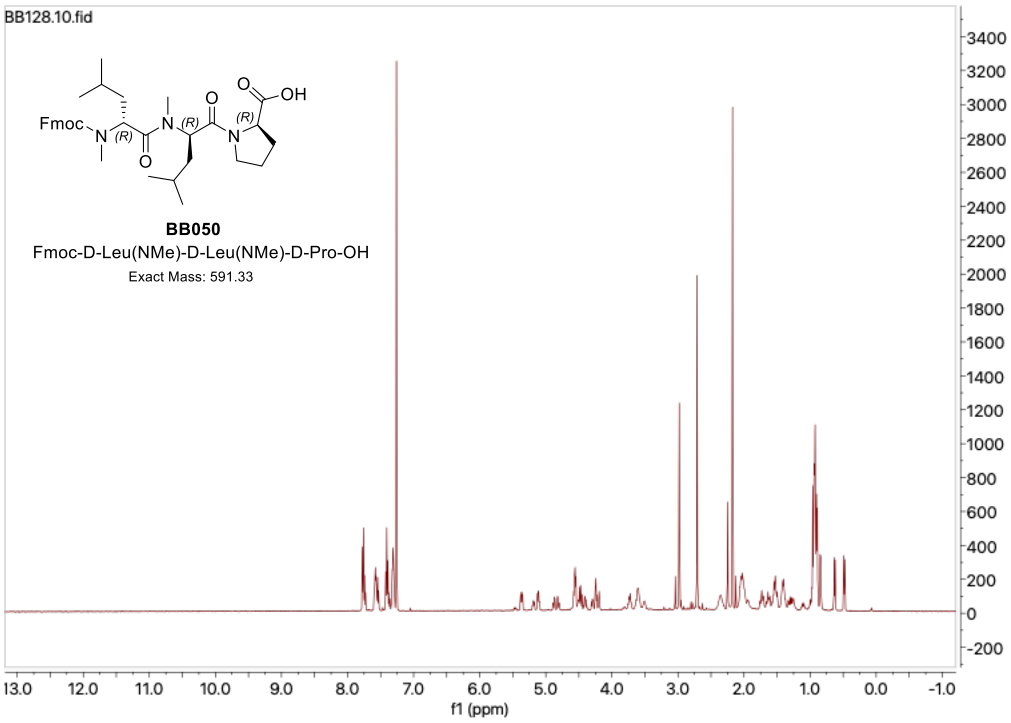
BB050
Fmoc-D-Leu(NMe)-D-Leu(NMe)-D-Pro-OH
Exact Mass: 591.33





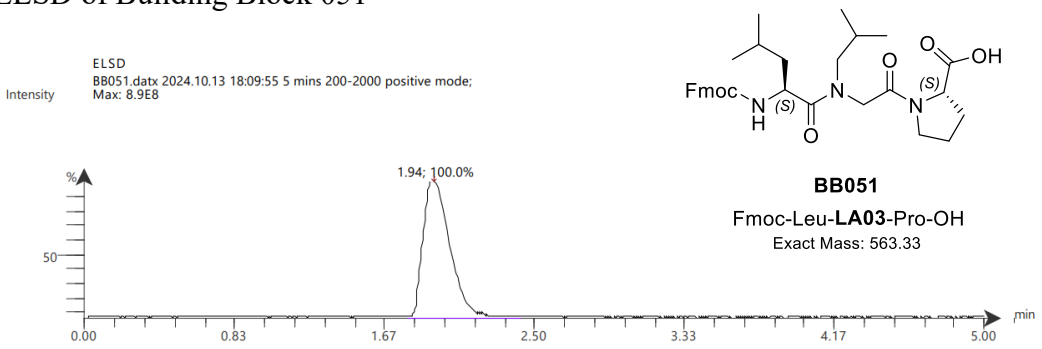


[1]H-NMR of Building Block 050

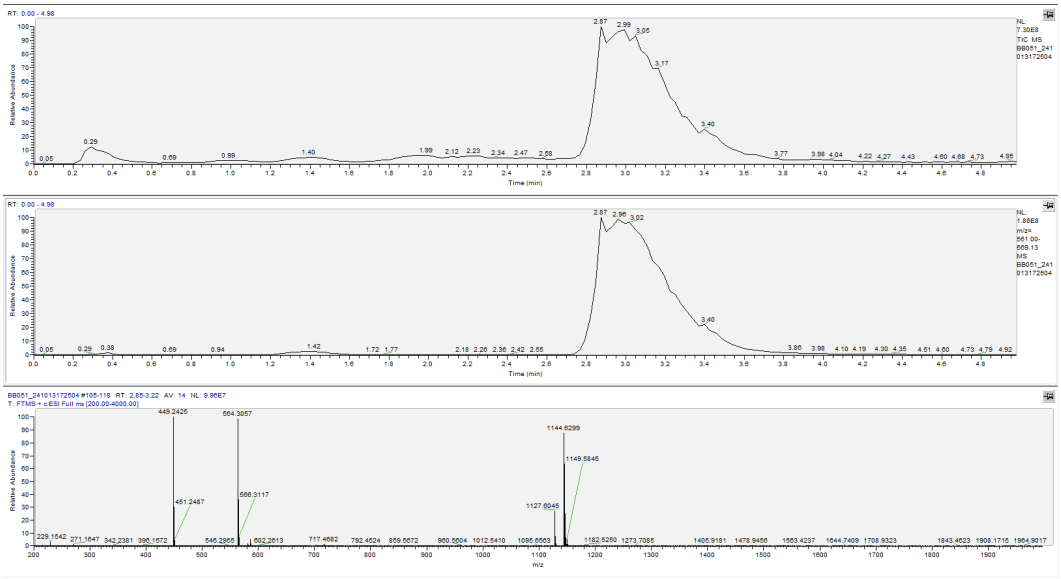


BB051

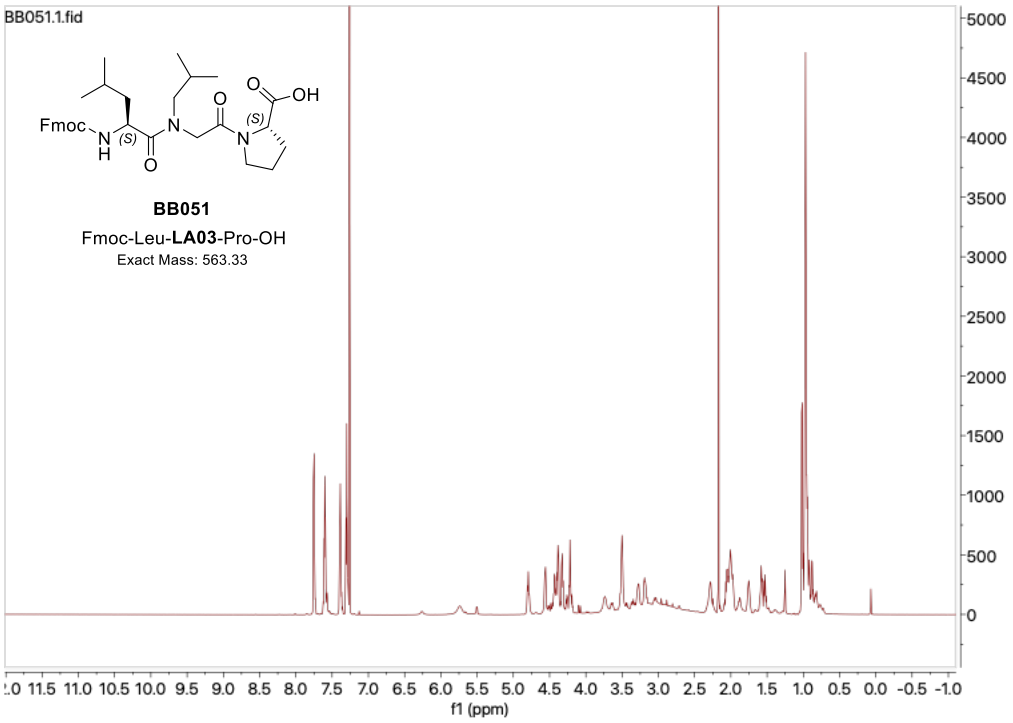
ELSD of Building Block 051



HRMS of Building Block 051

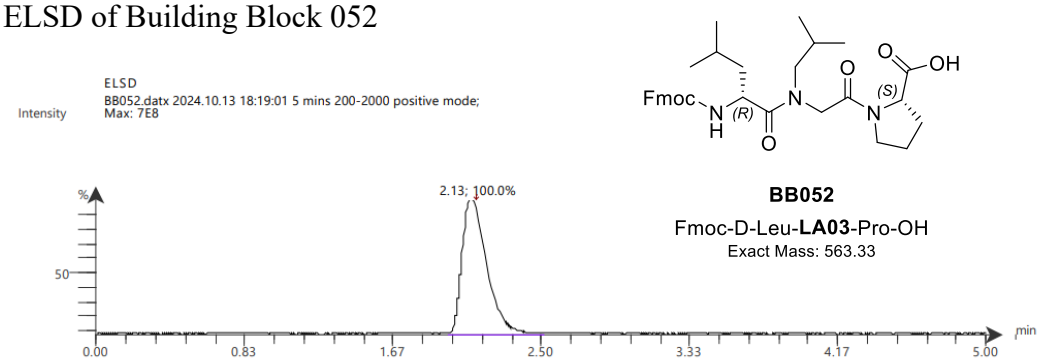


^{1}H-NMR of Building Block 051

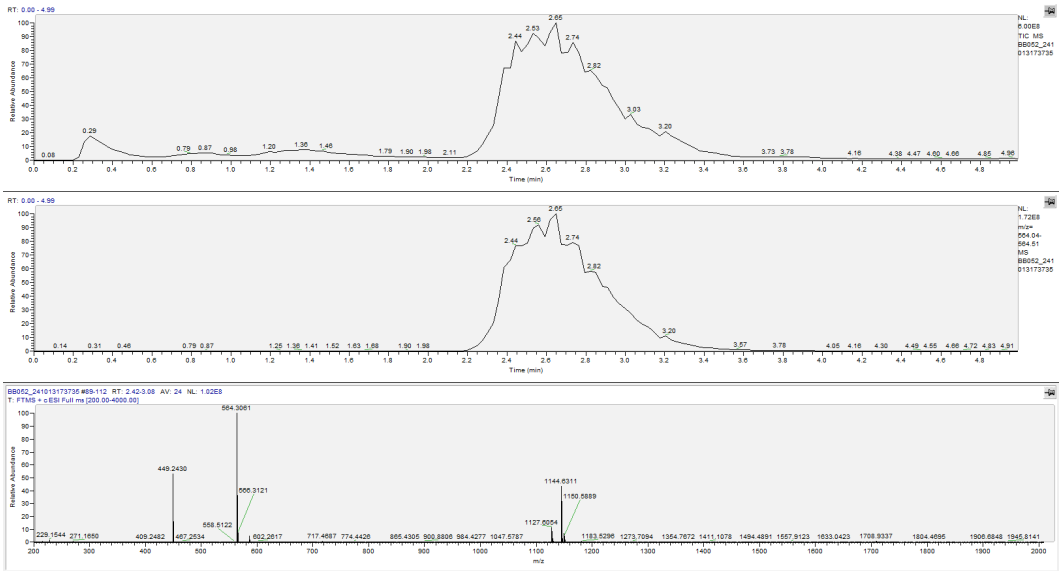


BB052

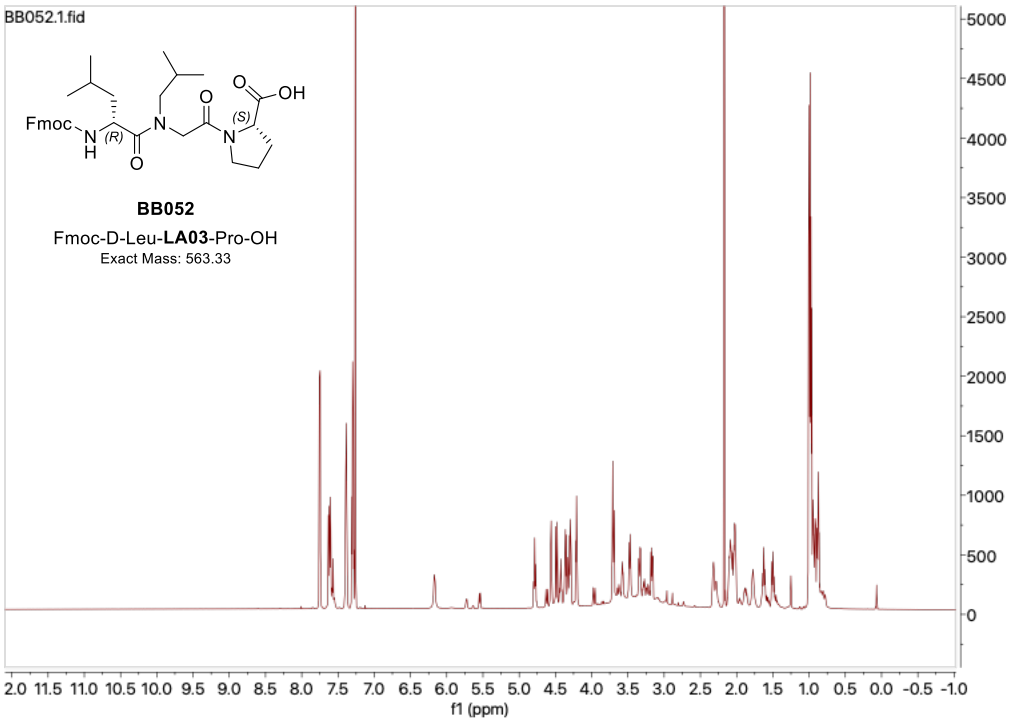
ELSD of Building Block 052



HRMS of Building Block 052

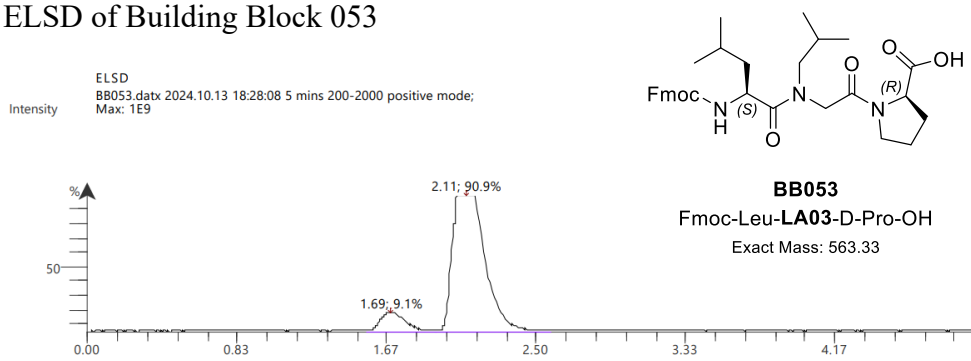


^{1}H-NMR of Building Block 052

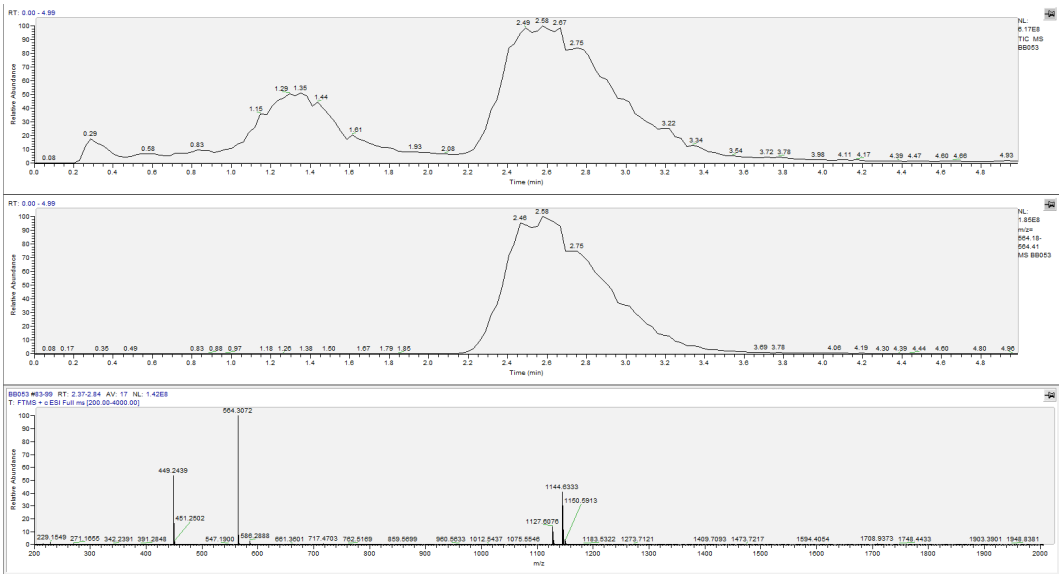


BB053

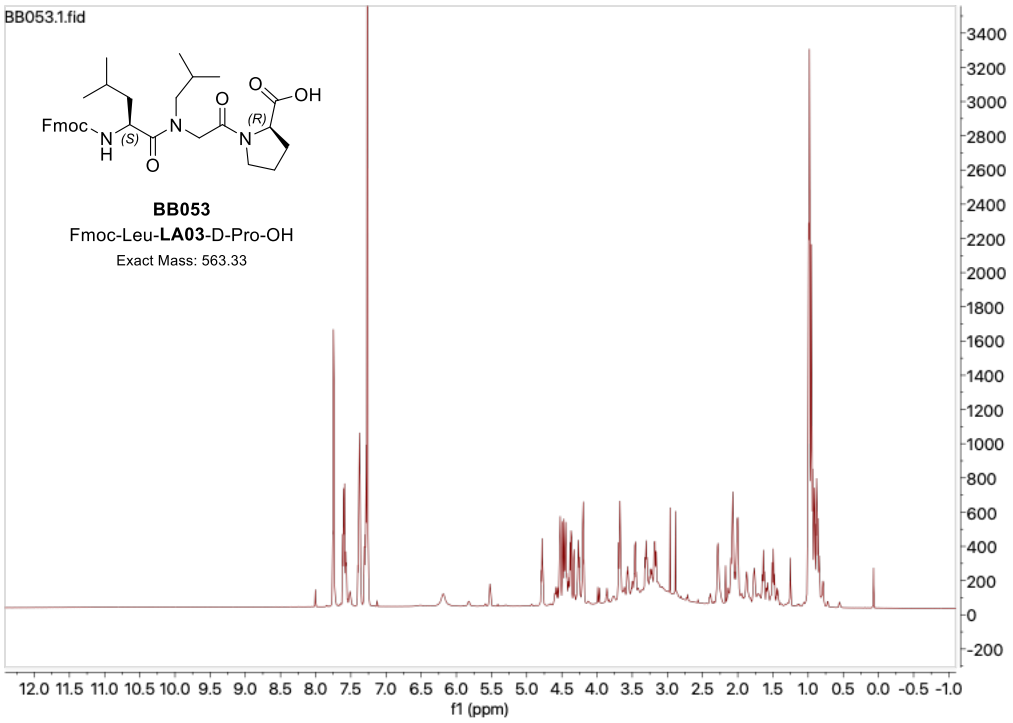
ELSD of Building Block 053



HRMS of Building Block 053

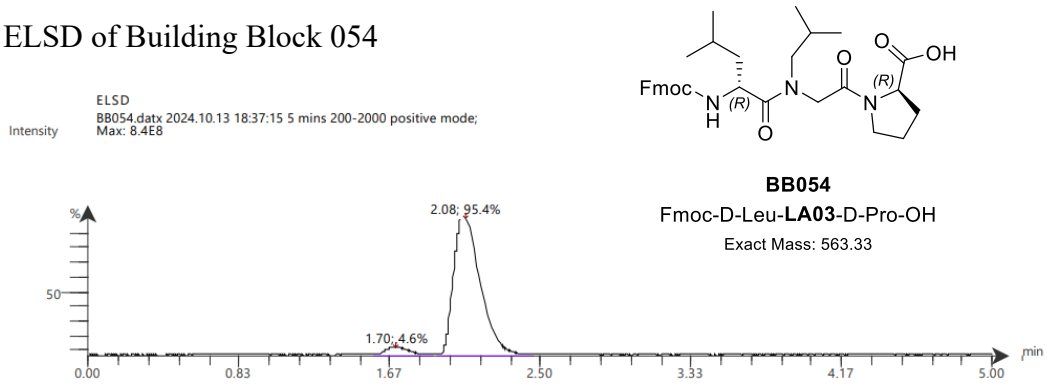


^{1}H-NMR of Building Block 053

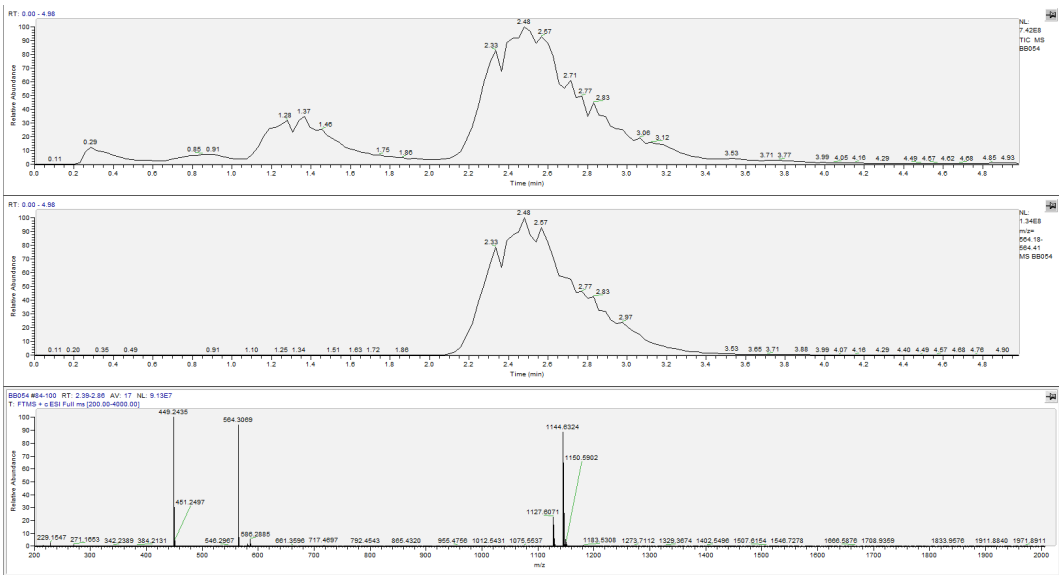


BB054

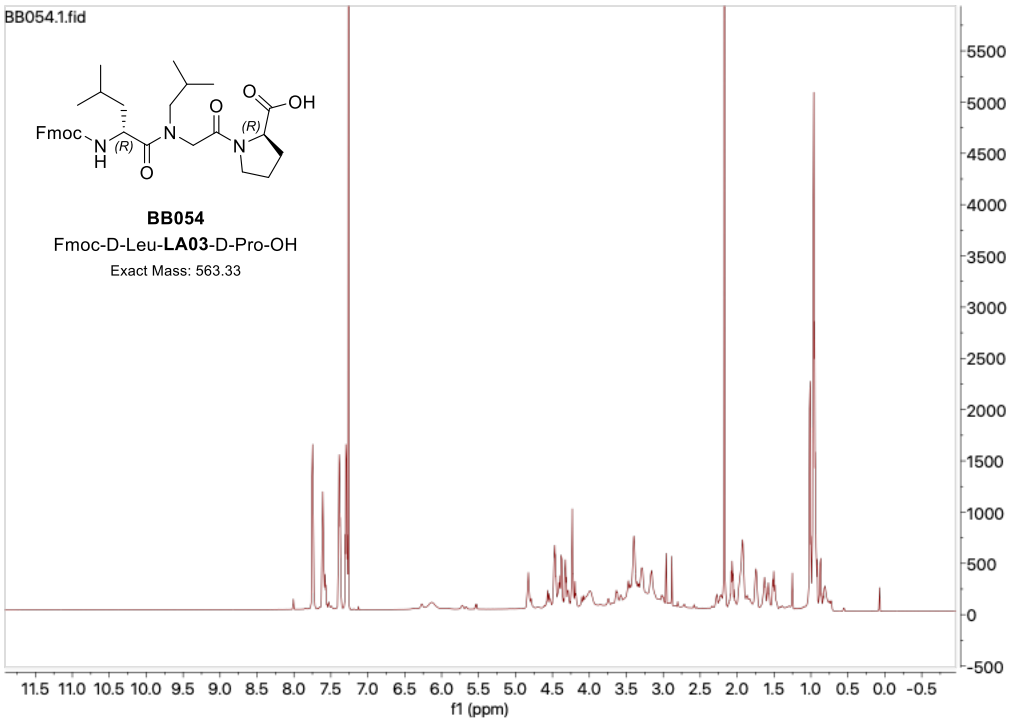
ELSD of Building Block 054



HRMS of Building Block 054

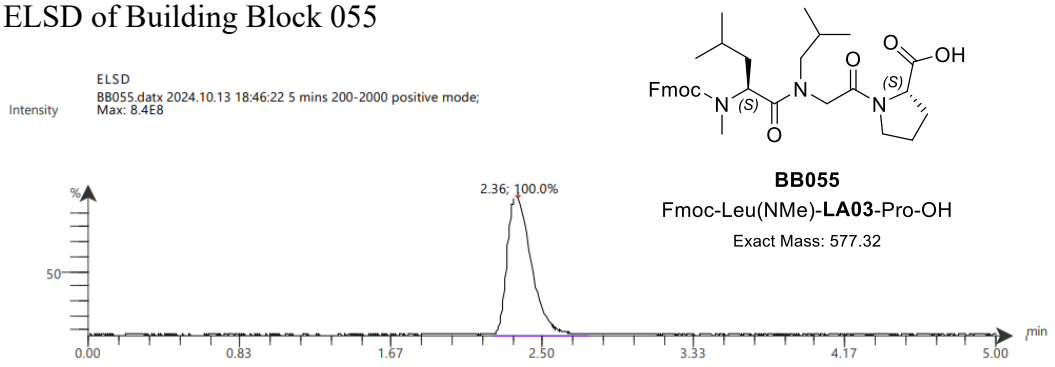


[1]H-NMR of Building Block 054

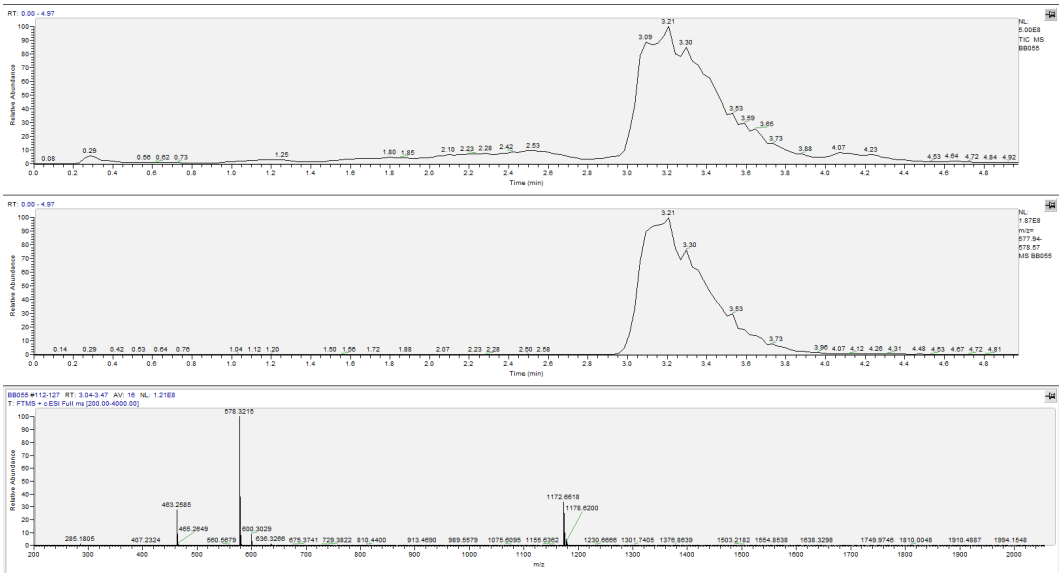


BB055

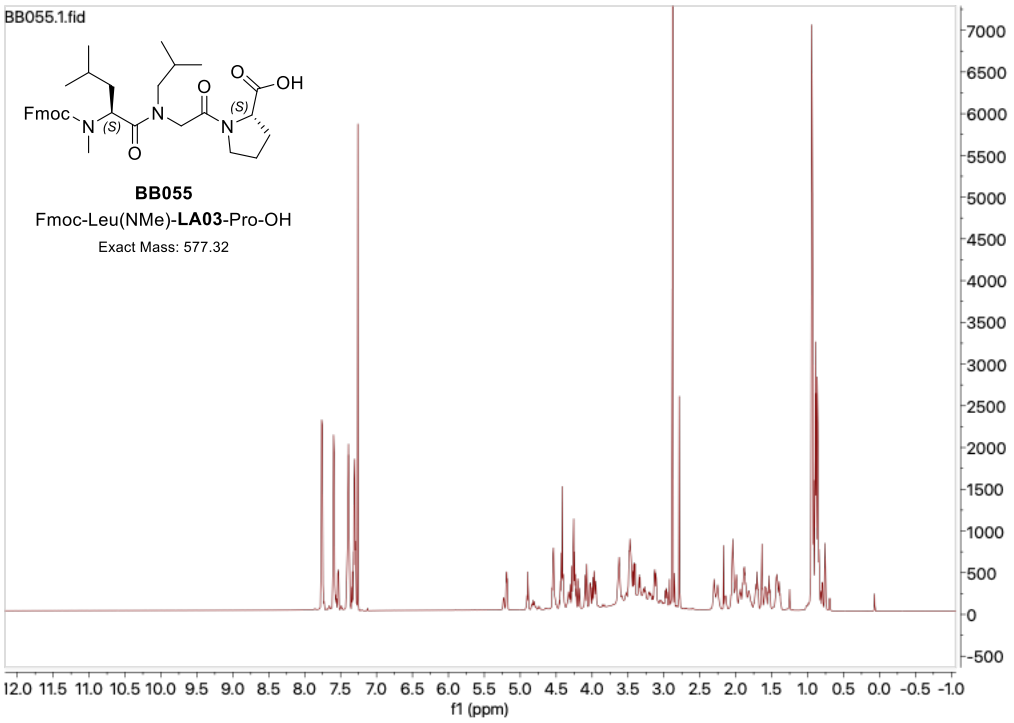
ELSD of Building Block 055



HRMS of Building Block 055

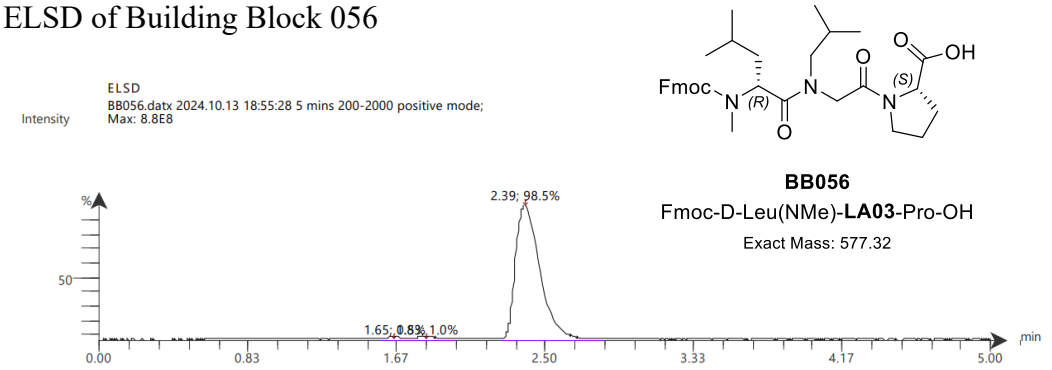


^{1}H-NMR of Building Block 055

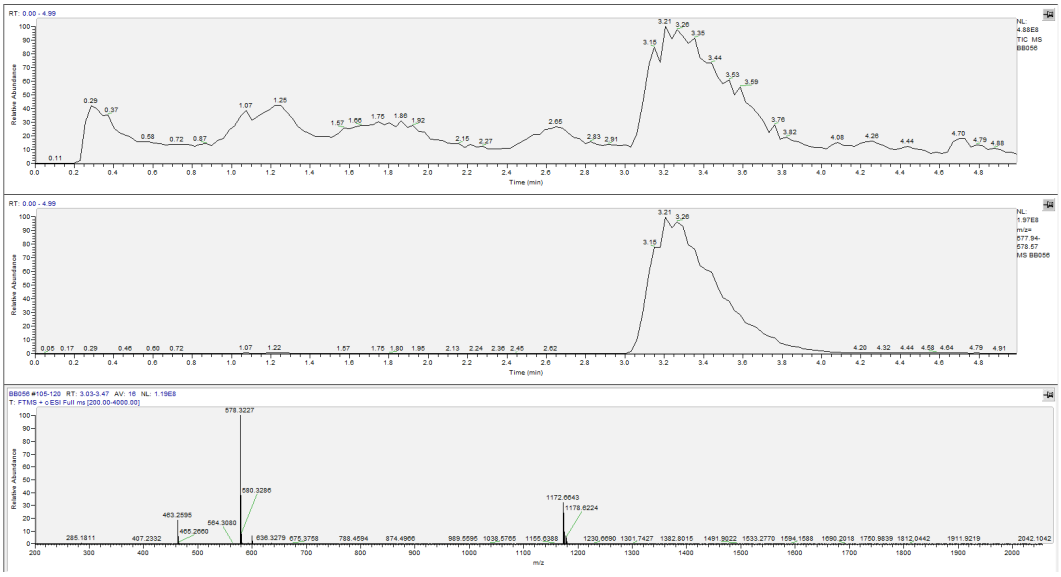


BB056

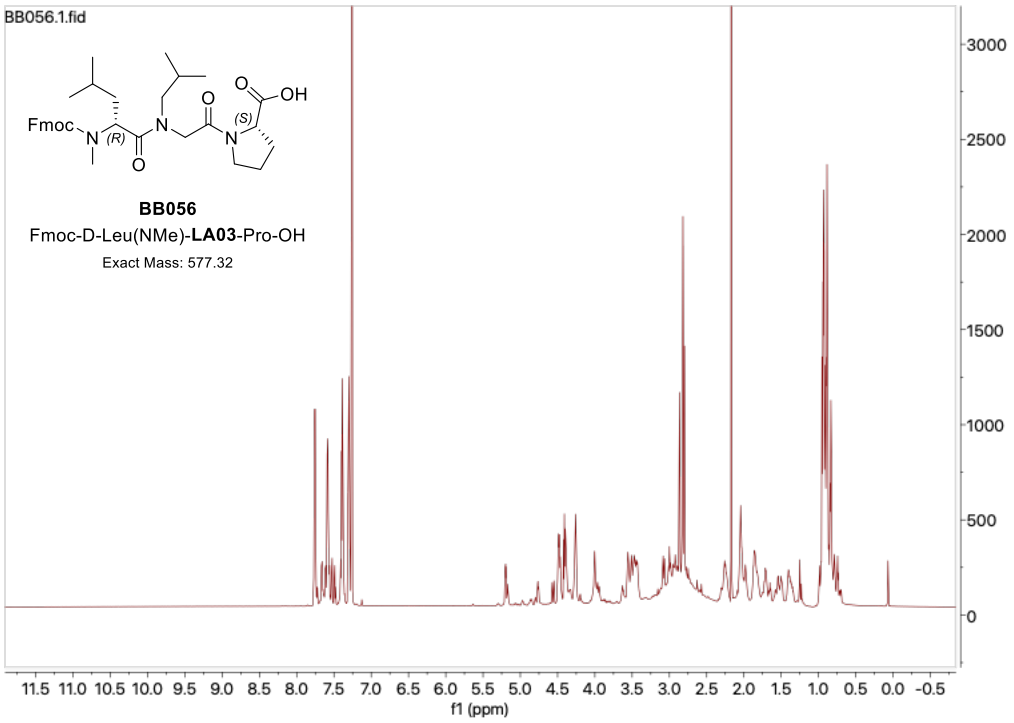
ELSD of Building Block 056



HRMS of Building Block 056

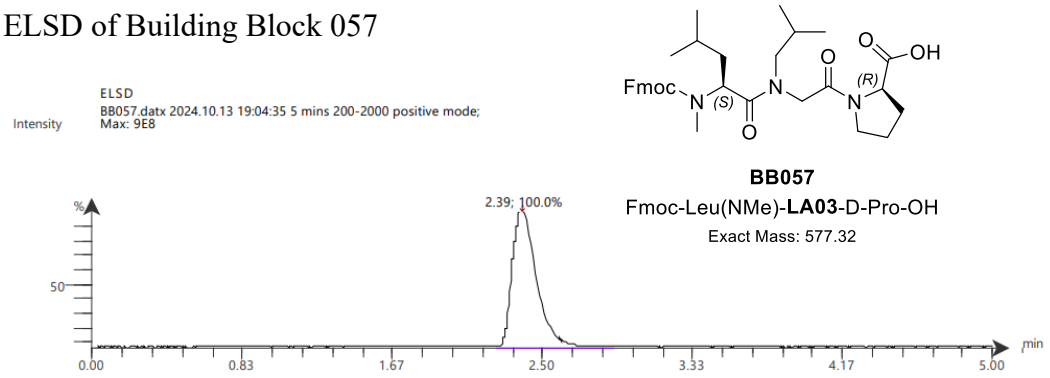


^{1}H-NMR of Building Block 056

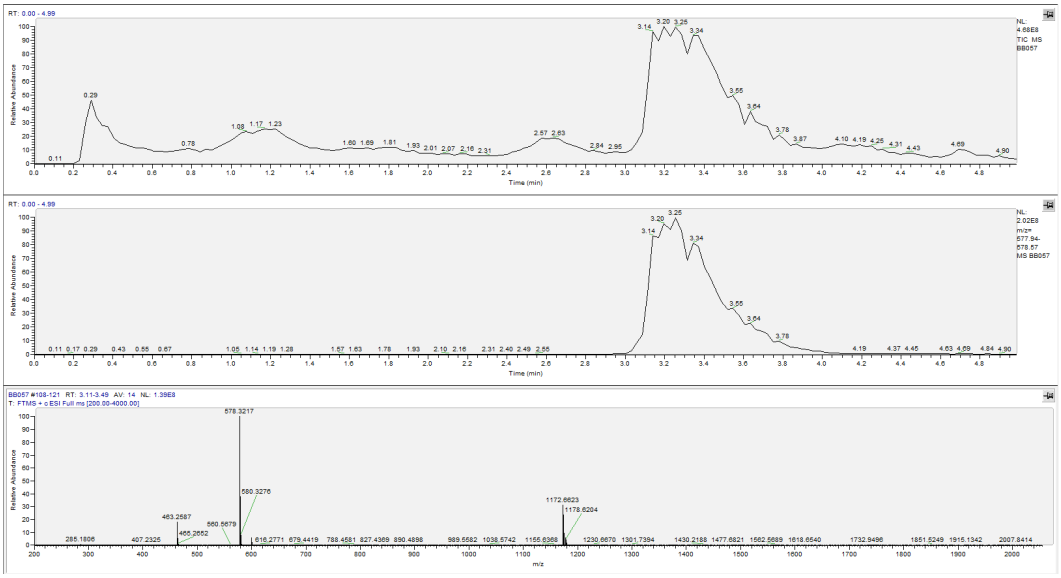


BB057

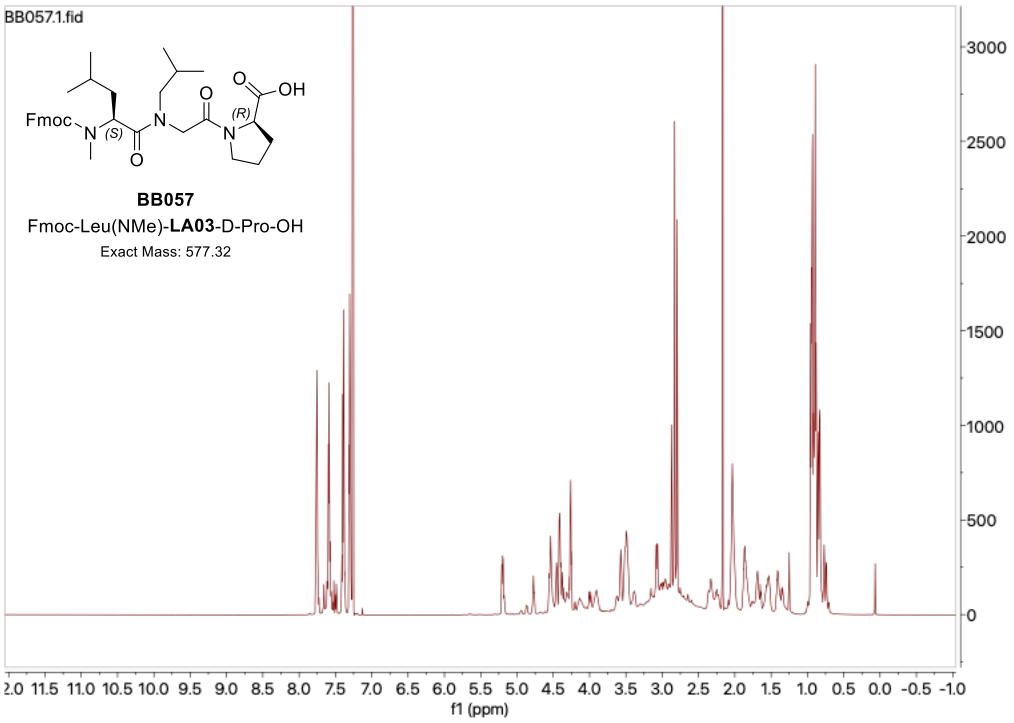
ELSD of Building Block 057



HRMS of Building Block 057

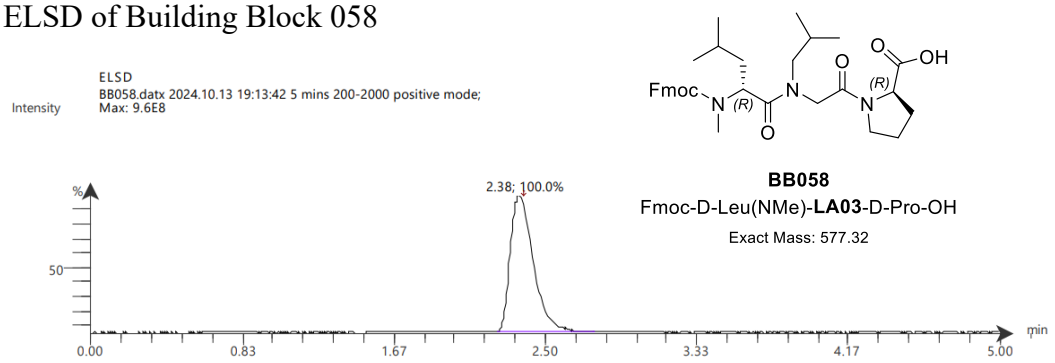


^{1}H-NMR of Building Block 057

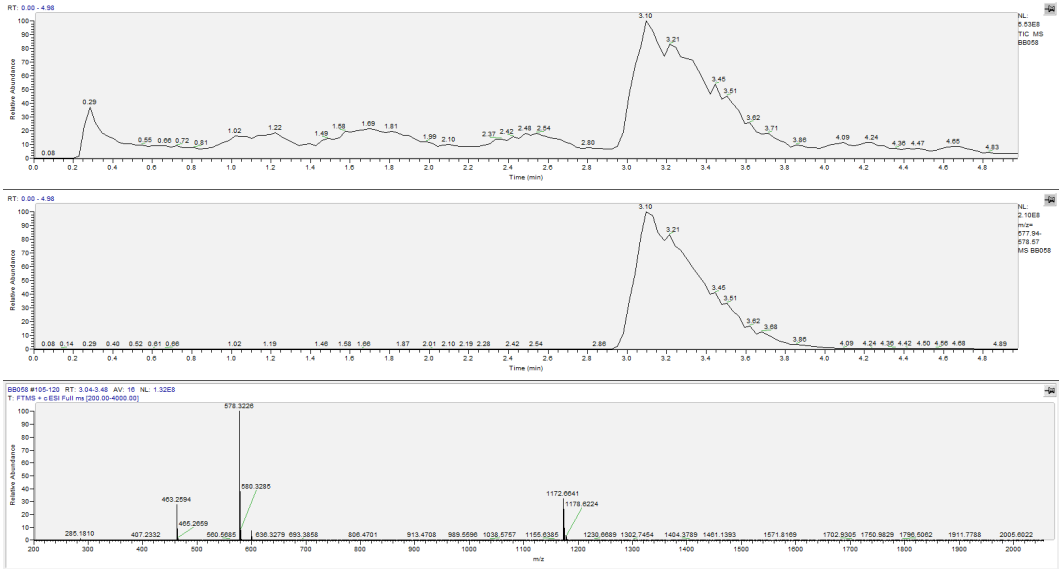


BB058

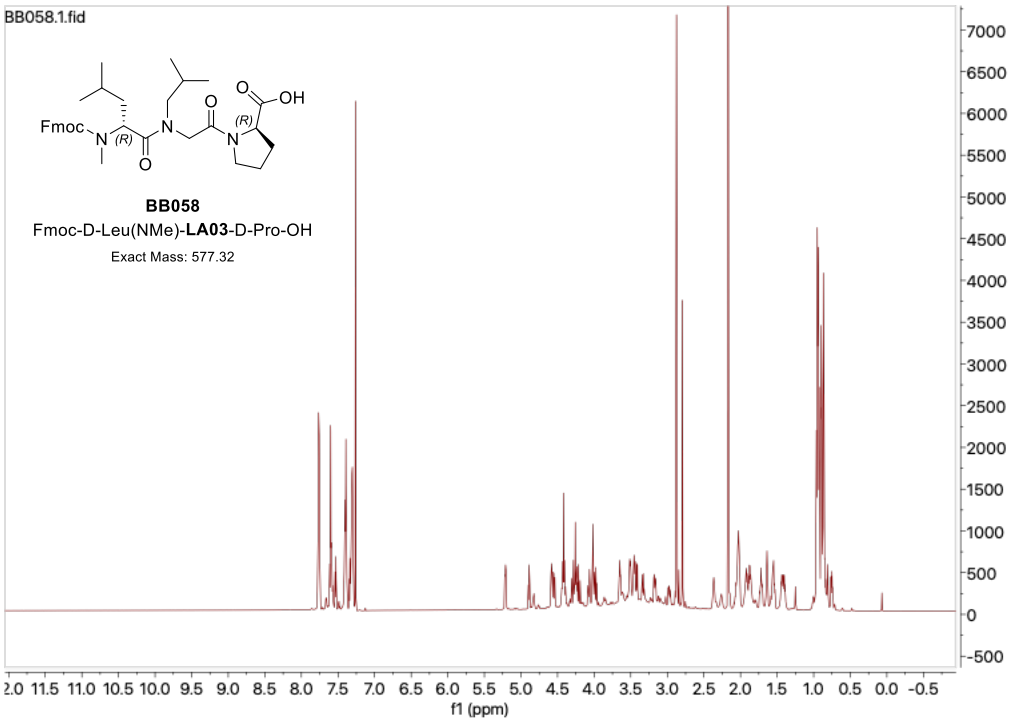
ELSD of Building Block 058



HRMS of Building Block 058



^{1}H-NMR of Building Block 058



References

(1) Waring, M. J. Defining Optimum Lipophilicity and Molecular Weight Ranges for Drug Candidates-Molecular Weight Dependent Lower LogD Limits Based on Permeability. *Bioorg. Med. Chem. Lett.* **2009**, *19* (10), 2844-2851.

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

(3) Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S. Quantifying the Chameleonic Properties of Macrocycles and Other High-Molecular-Weight Drugs. *Drug. Discov. Today* **2016**, *21* (5), 712-717.

(4) Tanada, M.; Tamiya, M.; Matsuo, A.; Chiyoda, A.; Takano, K.; Ito, T.; Irie, M.; Kotake, T.; Takeyama, R.; Kawada, H.; et al. Development of Orally Bioavailable Peptides Targeting an Intracellular Protein: From a Hit to a Clinical KRAS Inhibitor. *J. Am. Chem. Soc.* **2023**, *145* (30), 16610-16620.

(5) Silvestri, A. P.; Zhang, Q.; Ping, Y.; Muir, E. W.; Zhao, J.; Chakka, S. K.; Wang, G.; Bray, W. M.; Chen, W.; Fribourgh, J. L.; et al. DNA-Encoded Macrocyclic Peptide Libraries Enable the Discovery of a Neutral MDM2-p53 Inhibitor. *ACS Med. Chem. Lett.* **2023**, *14* (6), 820-826.

(6) Zhu, Z.; Shaginian, A.; Grady, L. C.; O'Keeffe, T.; Shi, X. E.; Davie, C. P.; Simpson, G. L.; Messer, J. A.; Evindar, G.; Bream, R. N.; et al. Design and

Application of a DNA-Encoded Macrocyclic Peptide Library. *ACS Chem. Biol.* **2018**, *13* (1), 53-59.

(7) Tucker, T. J.; Embrey, M. W.; Alleyne, C.; Amin, R. P.; Bass, A.; Bhatt, B.; Bianchi, E.; Branca, D.; Bueters, T.; Buist, N.; et al. A Series of Novel, Highly Potent, and Orally Bioavailable Next-Generation Tricyclic Peptide PCSK9 Inhibitors. *J. Med. Chem.* **2021**, *64* (22), 16770-16800.

(8) Ahlbach, C. L.; Lexa, K. W.; Bockus, A. T.; Chen, V.; Crews, P.; Jacobson, M. P.; Lokey, R. S. Beyond cyclosporine A: conformation-dependent passive membrane permeabilities of cyclic peptide natural products. *Future Med. Chem.* **2015**, *7* (16), 2121-2130.

(9) 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. *Eur. J. Med. Chem.* **2015**, *97*, 202-213.

(10) Kelly, C. N.; Townsend, C. E.; Jain, A. N.; Naylor, M. R.; Pye, C. R.; Schwochert, J.; Lokey, R. S. Geometrically Diverse Lariat Peptide Scaffolds Reveal an Untapped Chemical Space of High Membrane Permeability. *J. Am. Chem. Soc.* **2021**, *143* (2), 705-714.

(11) Bhardwaj, G.; O'Connor, J.; Rettie, S.; Huang, Y. H.; Ramelot, T. A.; Mulligan, V. K.; Alpkilic, G. G.; Palmer, J.; Bera, A. K.; Bick, M. J.; et al. Accurate De Novo Design of Membrane-Traversing Macrocycles. *Cell* **2022**, *185* (19), 3520-3532.

(12) Linker, S. M.; Schellhaas, C.; Kamenik, A. S.; Veldhuizen, M. M.; Waibl, F.; Roth, H.-J.; Fouché, M.; Rodde, S.; Riniker, S. Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes. *J. Med. Chem.* **2023**, *66* (4), 2773-2788.

(13) Kenny, P. W. Hydrogen-Bond Donors in Drug Design. *J. Med. Chem.* **2022**, *65* (21), 14261-14275.

(14) Mayer, P. T.; Anderson, B. D. Transport Across 1,9-Decadiene Precisely Mimics the Chemical Selectivity of the Barrier Domain in Egg Lecithin Bilayers. *J. Pharm. Sci.* **2002**, *91* (3), 640-646.

(15) Valkó, K. L. Lipophilicity and Biomimetic Properties Measured by HPLC to Support Drug Discovery. *J. Pharm. Biomed. Anal.* **2016**, *130*, 35-54.

(16) Lombardo, F.; Shalaeva, M. Y.; Tupper, K. A.; Gao, F. ElogDoct: A Tool for Lipophilicity Determination in Drug Discovery. 2. Basic and Neutral Compounds. *J. Med. Chem.* **2001**, *44* (15), 2490-2497.

(17) Valkó, K.; Bevan, C.; Reynolds, D. Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to Log P/Log D. *Anal. Chem.* **1997**, *69* (11), 2022-2029.

(18) Bunally, S. B.; Luscombe, C. N.; Young, R. J. Using Physicochemical Measurements to Influence Better Compound Design. *SLAS Discov.* **2019**, *24* (8), 791-801.

(19) Katz, D.; Fike, K.; Longenberger, J.; Placko, S.; Philippe-Venec, L.; Chervenak, A. AlphaLogD determination: An Optimized Reversed-Phase Liquid

Chromatography Method to Measure Lipophilicity on Neutral and Basic Small and Beyond-Rule-of-Five Compounds. *J. Chromatogr. A* **2022**, *1674*, 1-11.

(20) Grumetto, L.; Carpentiero, C.; Barbato, F. Lipophilic and Electrostatic Forces Encoded in IAM-HPLC Indexes of Basic Drugs: Their Role in Membrane Partition and Their Relationships with BBB Passage Data. *Eur. J. Pharm. Sci.* **2012**, *45* (5), 685-692.

(21) Garcia Jimenez, D.; Vallaro, M.; Rossi Sebastiano, M.; Apprato, G.; D'Agostini, G.; Rossetti, P.; Ermondi, G.; Caron, G. Chamelogk: A Chromatographic Chameleonicity Quantifier to Design Orally Bioavailable Beyond-Rule-of-5 Drugs. *J. Med. Chem.* **2023**, *66* (15), 10681-10693.

(22) Caron, G.; Vallaro, M.; Ermondi, G.; Goetz, G. H.; Abramov, Y. A.; Philippe, L.; Shalaeva, M. A Fast Chromatographic Method for Estimating Lipophilicity and Ionization in Nonpolar Membrane-Like Environment. *Mol. Pharm.* **2016**, *13* (3), 1100-1110.

(23) Jensen, D. A.; Gary, R. K. Estimation of Alkane-Water LogP for Neutral, Acidic, and Basic Compounds Using an Alkylated Polystyrene-Divinylbenzene High-Performance Liquid Chromatography Column. *J. Chromatogr. A* **2015**, *1417*, 21-29.

(24) 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.; et al. Lipophilic Permeability Efficiency Reconciles the Opposing Roles of Lipophilicity in Membrane Permeability and Aqueous Solubility. *J. Med. Chem.* **2018**, *61* (24), 11169-11182.

(25) Taechalertpaisarn, J.; Ono, S.; Okada, O.; Johnstone, T. C.; Lokey, R. S. A New Amino Acid for Improving Permeability and Solubility in Macrocyclic Peptides through Side Chain-to-Backbone Hydrogen Bonding. *J. Med. Chem.* **2022**, *65* (6), 5072-5084.

(26) Klein, V. G.; Bond, A. G.; Craigon, C.; Lokey, R. S.; Ciulli, A. Amide-to-Ester Substitution as a Strategy for Optimizing PROTAC Permeability and Cellular Activity. *J. Med. Chem.* **2021**, *64* (24), 18082-18101.

(27) Watanabe, A.; Uehara, S.; Akazawa, T.; Fujiu, M. Optimization of Cyclic Peptide Property Using Chromatographic Capacity Factor on Permeability of Passive Cell Membrane and Human Induced Pluripotent Stem Cell-Derived Intestinal Membrane. *J. Pharm. Sci.* **2022**, *111* (7), 1879-1886.

(28) Abraham, M. H.; Chadha, H. S.; Leitao, R. A. E.; Mitchell, R. C.; Lambert, W. J.; Kaliszan, R.; Nasal, A.; Haber, P. Determination of Solute Lipophilicity, As Log P(octanol) and Log P(alkane) Using Poly(styrene–divinylbenzene) and Immobilised Artificial Membrane Stationary Phases in Reversed-Phase High-Performance Liquid Chromatography. *J. Chromatogr. A* **1997**, *766* (1-2), 35-47.

(29) Lambert, W. J.; Wright, L. A.; Stevens, J. K. Development of a Preformulation Lipophilicity Screen Utilizing a C-18 Deriviatzed Polystyrene-Divinylbenzene High-Performance Liquid Chromatographic (HPLC) Column. *Pharm. Res.* **1990**, *7* (6), 577-586.

(30) Furukawa, A.; Schwochert, J.; Pye, C. R.; Asano, D.; Edmondson, Q. D.; Turmon, A. C.; Klein, V. G.; Ono, S.; Okada, O.; Lokey, R. S. Drug-Like Properties

in Macrocycles above MW 1000: Backbone Rigidity versus Side-Chain Lipophilicity. *Angew. Chem. Int. Ed. Engl.* **2020**, *59* (48), 21571-21577.

(31) 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.; et al. On-Resin N-Methylation of Cyclic Peptides for Discovery of Orally Bioavailable Scaffolds. *Nat. Chem. Biol.* **2011**, *7* (11), 810-817.

(32) 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), 715-721.

(33) Faris, J. H.; Adaligil, E.; Popovych, N.; Ono, S.; Takahashi, M.; Nguyen, H.; Plise, E.; Taechalertpaisarn, J.; Lee, H. W.; Koehler, M. F. T.; et al. Membrane Permeability in a Large Macrocyclic Peptide Driven by a Saddle-Shaped Conformation. *J. Am. Chem. Soc.* **2024**, *146* (7), 4582-4591.

(34) Kubinyi, H. Quantitative Structure-Activity Relationships. 7. The Bilinear Model, a New Model for Nonlinear Dependence of Biological Activitiy on Hydrophobic Character. *J. Med. Chem.* **1976**, *20* (5), 625-629.

(35) 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), 1048-1059.

(36) Li, J.; Yanagisawa, K.; Sugita, M.; Fujie, T.; Ohue, M.; Akiyama, Y. CycPeptMPDB: A Comprehensive Database of Membrane Permeability of Cyclic Peptides. *J .Chem. Inf. Model.* **2023**, *63* (7), 2240-2250.

(37) Andres, A.; Roses, M.; Rafols, C.; Bosch, E.; Espinosa, S.; Segarra, V.; Huerta, J. M. Setup and Validation of Shake-Flask Procedures for the Determination of Partition Coefficients (logD) From Low Drug Amounts. *Eur. J. Pharm. Sci.* **2015**, *76*, 181-191.

(38) Molecular Operating Environment (MOE), C. C. G. U., 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, **2024**.

(39) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157-2164.

(40) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999-2012.

(41) Khoury, G. A.; Smadbeck, J.; Tamamis, P.; Vandris, A. C.; Kieslich, C. A.; Floudas, C. A. Forcefield_NCAA: Ab Initio Charge Parameters to Aid in the Discovery and Design of Therapeutic Proteins and Peptides with Unnatural Amino Acids and Their Application to Complement Inhibitors of the Compstatin Family. *ACS Synth. Biol.* **2014**, *3* (12), 855-869.

(42) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157-1174.

(43) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839.

(44) Higo, J.; Kamiya, N.; Sugihara, T.; Yonezawa, Y.; Nakamura, H. Verifying trivial parallelization of multicanonical molecular dynamics for conformational sampling of a polypeptide in explicit water. *Chem. Phys. Lett.* **2009**, *473* (4-6), 326-329.

(45) Higo, J.; Umezawa, K.; Nakamura, H. A virtual-system coupled multicanonical molecular dynamics simulation: Principles and applications to free-energy landscape of protein–protein interaction with an all-atom model in explicit solvent. *J. Chem. Phys.* **2013**, *138* (184106), 1-11.

(46) Bussi, G. D., Davide; Parrinello, Michele. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (014101), 1-8.

(47) Fan, Y.; Feng, R.; Zhang, X.; Wang, Z. L.; Xiong, F.; Zhang, S.; Zhong, Z. F.; Yu, H.; Zhang, Q. W.; Zhang, Z.; et al. Encoding and display technologies for combinatorial libraries in drug discovery: The coming of age from biology to therapy. *Acta. Pharm. Sin. B* **2024**, *14* (8), 3362-3384.

(48) Gironda-Martinez, A.; Donckele, E. J.; Samain, F.; Neri, D. DNA-Encoded Chemical Libraries: A Comprehensive Review with Succesful Stories and Future Challenges. *ACS Pharmacol. Transl. Sci.* **2021**, *4* (4), 1265-1279.

(49) Huang, Y.; Wiedmann, M. M.; Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **2018**, *119* (17), 10360-10391.

(50) Iskandar, S. E.; Bowers, A. A. mRNA Display Reaches for the Clinic with New PCSK9 Inhibitor. *ACS Med. Chem. Lett.* **2022**, *13* (9), 1379-1383.

(51) Jaroszewicz, W.; Morcinek-Orlowska, J.; Pierzynowska, K.; Gaffke, L.; Wegrzyn, G. Phage display and other peptide display technologies. *FEMS Microbiol. Rev.* **2022**, *46* (2), 1-25.

(52) Hampton, J. T.; Liu, W. R. Diversification of Phage-Displayed Peptide Libraries with Noncanonical Amino Acid Mutagenesis and Chemical Modification. *Chem. Rev.* **2024**, *124* (9), 6051-6077.

(53) Ohta, A.; Tanada, M.; Shinohara, S.; Morita, Y.; Nakano, K.; Yamagishi, Y.; Takano, R.; Kariyuki, S.; Iida, T.; Matsuo, A.; et al. Validation of a New Methodology to Create Oral Drugs beyond the Rule of 5 for Intracellular Tough Targets. *J. Am. Chem. Soc.* **2023**, *145* (44), 24035-24051.

(54) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; et al. Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nat. Chem. Biol.* **2009**, *5* (9), 647-654.

(55) Brenner, S.; Lerner, R. A. Encoded Combinatorial Chemistry. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381-5383.

(56) Ma, P.; Zhang, S.; Huang, Q.; Gu, Y.; Zhou, Z.; Hou, W.; Yi, W.; Xu, H. Evolution of chemistry and selection technology for DNA-encoded library. *Acta. Pharm. Sin. B* **2024**, *14* (2), 492-516.

(57) Li, L.; Su, M.; Lu, W.; Song, H.; Liu, J.; Wen, X.; Suo, Y.; Qi, J.; Luo, X.; Zhou, Y. B.; et al. Triazine-Based Covalent DNA-Encoded Libraries for Discovery of Covalent Inhibitors of Target Proteins. *ACS Med. Chem. Lett.* **2022**, *13* (10), 1574-1581.

(58) Škopić, M. K.; Bugain, O.; Jung, K.; Onstein, S.; Brandherm, S.; Kalliokoski, T.; Brunschweiger, A. Design and synthesis of DNA-encoded libraries based on a benzodiazepine and a pyrazolopyrimidine scaffold. *MedChemComm* **2016**, *7* (10), 1957-1965.

(59) Phelan, J. P.; Lang, S. B.; Sim, J.; Berritt, S.; Peat, A. J.; Billings, K.; Fan, L.; Molander, G. A. Open-Air Alkylation Reactions in Photoredox-Catalyzed DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* **2019**, *141* (8), 3723-3732.

(60) Ruff, Y.; Martinez, R.; Pelle, X.; Nimsgern, P.; Fille, P.; Ratnikov, M.; Berst, F. An Amphiphilic Polymer-Supported Strategy Enables Chemical Transformations under Anhydrous Conditions for DNA-Encoded Library Synthesis. *ACS Comb. Sci.* **2020**, *22* (3), 120-128.

(61) Siripuram, V. K.; Sunkari, Y. K.; Nguyen, T. L.; Flajolet, M. DNA-Compatible Suzuki-Miyaura Cross-Coupling Reaction of Aryl Iodides With (Hetero)Aryl Boronic Acids for DNA-Encoded Libraries. *Front. Chem.* **2022**, *10* (894603), 1-7.

(62) Yuen, L. H.; Franzini, R. M. Achievements, Challenges, and Opportunities in DNA-Encoded Library Research: An Academic Point of View. *Chembiochem* **2017**, *18* (9), 829-836. DOI: 10.1002/cbic.201600567 From NLM Medline.

(63) Satz, A. L. Simulated Screens of DNA Encoded Libraries: The Potential Influence of Chemical Synthesis Fidelity on Interpretation of Structure-Activity Relationships. *ACS Comb. Sci.* **2016**, *18* (7), 415-424.

(64) Satz, A. L.; Hochstrasser, R.; Petersen, A. C. Analysis of Current DNA Encoded Library Screening Data Indicates Higher False Negative Rates for Numerically Larger Libraries. *ACS Comb. Sci.* **2017**, *19* (4), 234-238.

(65) Keller, M. P., Dimitar; Gloger, Andreas; Dietschi, Bastien; Jobin, Kilian; Gradinger, Timon; Martinelli, Adriano; Plais, Louise; Onda, Yuichi; Neri, Dario; Scheuermann, Jörg. Highly pure DNA-encoded chemical libraries by dual-linker solid-phase synthesis. *Science* **2024**, *384*, 1259-1265.

(66) Franzini, R. M.; Biendl, S.; Mikutis, G.; Samain, F.; Scheuermann, J.; Neri, D. "Cap-and-Catch" Purification for Enhancing the Quality of Libraries of DNA Conjugates. *ACS Comb. Sci.* **2015**, *17* (7), 393-398.

(67) Roy, A.; Koesema, E.; Kodadek, T. High-Throughput Quality Control Assay for the Solid-Phase Synthesis of DNA-Encoded Libraries of Macrocycles*. *Angew. Chem. Int. Ed. Engl.* **2021**, *60* (21), 11983-11990.

(68) Ratnayake, A. S.; Flanagan, M. E.; Foley, T. L.; Smith, J. D.; Johnson, J. G.; Bellenger, J.; Montgomery, J. I.; Paegel, B. M. A Solution Phase Platform to Characterize Chemical Reaction Compatibility with DNA-Encoded Chemical Library Synthesis. *ACS Comb. Sci.* **2019**, *21* (10), 650-655.

(69) Monty, O. B. C.; Simmons, N.; Chamakuri, S.; Matzuk, M. M.; Young, D. W. Solution-Phase Fmoc-Based Peptide Synthesis for DNA-Encoded Chemical Libraries: Reaction Conditions, Protecting Group Strategies, and Pitfalls. *ACS Comb. Sci.* **2020**, *22* (12), 833-843.

(70) Rama-Garda, R.; Amigo, J.; Priego, J.; Molina-Martin, M.; Cano, L.; Dominguez, E.; Loza, M. I.; Rivera-Sagredo, A.; de Blas, J. Normalization of DNA encoded library affinity selection results driven by high throughput sequencing and HPLC purification. *Bioorg. Med. Chem.* **2021**, *40*, 116178.

(71) Passioura, T.; Liu, W.; Dunkelmann, D.; Higuchi, T.; Suga, H. Display Selection of Exotic Macrocyclic Peptides Expressed under a Radically Reprogrammed 23 Amino Acid Genetic Code. *J. Am. Chem. Soc.* **2018**, *140* (37), 11551-11555.

(72) Sharma, V. K.; Glick, J.; Vouros, P. Reversed-phase ion-pair liquid chromatography electrospray ionization tandem mass spectrometry for separation, sequencing and mapping of sites of base modification of isomeric oligonucleotide adducts using monolithic column. *J. Chromatogr. A.* **2012**, *1245*, 65-74.

(73) Li, Y.; Gabriele, E.; Samain, F.; Favalli, N.; Sladojevich, F.; Scheuermann, J.; Neri, D. Optimized Reaction Conditions for Amide Bond Formation in DNA-Encoded Combinatorial Libraries. *ACS Comb. Sci.* **2016**, *18* (8), 438-443.

(74) Fornstedt, T.; Enmark, M. Separation of therapeutic oligonucleotides using ion-pair reversed-phase chromatography based on fundamental separation science. *Journal of Chromatography Open* **2023**, *3*. DOI: 10.1016/j.jcoa.2023.100079.

(75) Enmark, M.; Harun, S.; Samuelsson, J.; Ornskov, E.; Thunberg, L.; Dahlen, A.; Fornstedt, T. Selectivity limits of and opportunities for ion pair chromatographic separation of oligonucleotides. *J. Chromatogr. A* **2021**, *1651*, 462269.

(76) 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.